

Workload Optimization Manager 3.13.0 Full Documentation

THE SPECIFICATIONS AND INFORMATION REGARDING THE PRODUCTS REFERENCED IN THIS DOCUMENTATION ARE SUBJECT TO CHANGE WITHOUT NOTICE. EXCEPT AS MAY OTHERWISE BE AGREED BY CISCO IN WRITING, ALL STATEMENTS, INFORMATION, AND RECOMMENDATIONS IN THIS DOCUMENTATION ARE PRESENTED WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED.

The Cisco End User License Agreement and any product specific license terms govern your use of any Cisco software, including this product documentation, and are located at: http://www.cisco.com/go/eula. Cisco product warranty information is available at http://www.cisco.com/go/warranty. US Federal Communications Commission Notices are found here http://www.cisco.com/c/en/us/products/us-fcc-notice.html.

IN NO EVENT SHALL CISCO OR ITS SUPPLIERS BE LIABLE FOR ANY INDIRECT, SPECIAL, CONSEQUENTIAL, OR INCIDENTAL DAMAGES, INCLUDING, WITHOUT LIMITATION, LOST PROFITS OR LOSS OR DAMAGE TO DATA ARISING OUT OF THE USE OR INABILITY TO USE THIS MANUAL, EVEN IF CISCO OR ITS SUPPLIERS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Any products and features described in this document as in development or available at a future date remain in varying stages of development and will be offered on a when-and if-available basis. Any such product or feature roadmaps are subject to change at the sole discretion of Cisco, and Cisco will have no liability for delay in the delivery or failure to deliver any products or feature roadmap items that may be set forth in this document.

Any Internet Protocol (IP) addresses and phone numbers used in this document are not intended to be actual addresses and phone numbers. Any examples, command display output, network topology diagrams, and other figures included in the document are shown for illustrative purposes only. Any use of actual IP addresses or phone numbers in illustrative content is unintentional and coincidental.

The documentation set for this product strives to use bias-free language. For the purposes of this documentation set, bias-free is defined as language that does not imply discrimination based on age, disability, gender, racial identity, ethnic identity, sexual orientation, socioeconomic status, and intersectionality. Exceptions may be present in the documentation due to language that is hardcoded in the user interfaces of the product software, language used based on RFP documentation, or language that is used by a referenced third-party product.

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com go trademarks. Third-party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1721R)

© 2025 Cisco Systems, Inc. All rights reserved.

Contents

Documentation overview....22

Release notes....23

- What's new....23
- Feature updates and notices....27
- Versioning explanation....31
- Configuration requirements....32
- Known issues....35
- REST API notes....41
  - REST API changes....41
  - API deprecation history....41
  - API deprecation contract....48

Installation....49

- Minimum requirements....49
- Installing on a virtual machine image....50
  - OVA: Installing the vCenter image for on-prem environments....51
  - VHD: Installing the Microsoft Hyper-V image....53
  - Deploying the Workload Optimization Manager components....54
- General configuration tasks....57
  - (Required) Synchronizing time....58
  - (Important) Verifying your MariaDB version....59
  - (Optional) Configuring a remote database....61
  - (Optional) Increasing available disk space....74
  - (Optional) Enforcing secure access by using LDAP....77
  - (Optional) Adding a certificate for securing the Workload Optimization Manager UI....79
  - (Optional) Adding additional CA certificates for probes....83
  - (Optional)Modifying the certificates for Cluster Manager....85
  - (Optional) Enabling the reporting Add-on....87
  - (Optional) Enabling the Data Exporter....87
  - (Optional) Changing the IP address of the platform node....91
  - (Optional) Redirecting audit and container logs....93
  - (Optional) Enabling and disabling probe components....95
  - (Optional) Disabling component health monitoring....98
  - (Optional) Overriding default ports....99
- License installation and first-time login....100
- Single Sign-On authentication....100
  - Typical settings for an IdP....100
  - Setting Up SAML authentication....102

Setting Up OpenID authentication....106
Disabling Single Sign-On....111
Reference: FIPS cipher suites....112
Reference: Step-wise platform deployment....113
Reference: Upgrading from MySQL 5.7 to 8.0....117
Reference: Migrating from CentOS to Rocky Linux....118
Reference: Working with YAML files....125
Reference: Image listing....127

Updating Workload Optimization Manager....132
Checking before updating....133
External databases and Workload Optimization Manager updates....134
Offline update....137
Reference: Step-wise offline update....140

Product overview....143
How Workload Optimization Manager works....143
The desired state....144
The market and virtual currency....144
Risk index....145
The Workload Optimization Manager supply chain....146
Sustainability features....147

Getting started....151
Logging in to Workload Optimization Manager....151
The Home Page....152
Application view....152
On-prem view....153
Cloud view....154
Supply chain of entities....159
Working with a scoped view....160
Scoping the Workload Optimization Manager session....161
Overview charts....164
Details view....165
Scope policies....167
List of entities....170
Navigating with the supply chain....171
Viewing cluster headroom....172

Target configuration....173
Configuring targets....176
Applications and databases targets....180

Apache Tomcat....180
Cisco AppDynamics....183
Datadog....189
Dynatrace....194
IBM WebSphere....200
Instana....203
JBoss....210
JVM application....212
MySQL....215
New Relic....219
Oracle....223
Oracle WebLogic....227
SQL server....230
Container platform targets....234
Kubeturbo deployment requirements....234
Connecting to Kubernetes clusters....244
Connecting to Red Hat OpenShift clusters....277
Container platform monitored resources....286
Container platform actions....289
Updating Kubeturbo....295
Kubeturbo references....302
Fabric targets....310
Cisco UCS Manager....310
HPE OneView....313
Guest OS processes targets....317
WMI....317
SNMP....321
Hyperconverged targets....326
Cisco HyperFlex....326
Nutanix....329
Hypervisor targets....333
IBM PowerVM....334
Microsoft Hyper-V....339
vCenter Server....346
IT Management targets....354
Flexera....354
ServiceNow....359
Orchestrator targets....360
Action Script....361
Private cloud targets....367

Microsoft Virtual Machine Manager....368
Public cloud targets....373
Amazon Web Services....373
Google Cloud....404
Microsoft Azure....426
Storage targets....484
EMC VMAX....485
EMC XtremIO....487
EMC ScaleIO....489
EMC VPLEX....491
HPE 3PAR....493
IBM FlashSystem....496
NetApp....500
Pure Storage....507
Virtual desktop infrastructure targets....509
VMware Horizon....509
Prometheus....512
Prometurbo deployment requirements....513
Deploying Prometurbo through OperatorHub....515
Deploying Prometurbo through YAML....517
Enabling metrics collection for Prometurbo....519
Updating Prometurbo....522
Data Ingestion Framework....522
Deploying Data Ingestion Framework....523
DIF schema and example files....526
Topology object....535
Entity object....535
Metrics entry object....539
Port configuration....540

User interface reference....543
Entity types - applications....543
Business application....544
Business transaction....546
Service....549
Application component....551
Application map....557
Entity types - container platform....566
Container platform business application....567
Container platform service....569
Container....577

Container spec....580
Workload controller....591
Container pod....604
Namespace....613
Container platform cluster....617
Virtual machine (container platform node)....624
Container platform CPU metrics....632
Entity types - cloud infrastructure....634
Virtual machine (cloud)....635
App component spec....669
Virtual Machine spec....670
Database server (cloud)....684
Database (cloud)....703
Data warehouse....719
Document collection....722
Volume (cloud)....726
Zone....736
Region....738
Entity types - on-prem infrastructure....740
Virtual machine (on-prem)....741
Database server (on-prem)....767
Volume (on-prem)....772
Virtual data center (private cloud)....774
Business user....778
Desktop pool....781
View pod....785
Host....786
Chassis....795
Data center....796
Storage....798
Logical pool....807
Disk array....809
Storage controller....813
IO module....815
Switch....815
Workload Optimization Manager actions....817
Working with Action Center....818
Action details....825
Actions by entity type....835
Action categories....844

Action types....845
Action acceptance modes....848
Plans: looking to the future....849
Plan management....850
Setting Up plan scenarios....851
Plan scenarios and types....857
Configuring nightly plans....941
Park: stop or start cloud resources....942
Place: reserve workload resources....952
Creating a reservation....954
Managing reservations....956
Dashboards: focused views....957
Built-in dashboards....958
Creating and editing custom dashboards....963
Creating and editing chart widgets....966
Chart types....969
Creating groups....1039
Working with policies....1042
Placement policies....1043
Automation policies....1049
Parking policies....1087
Working with schedules....1091
Managing calendar schedules....1092
Managing parking schedules....1097
Templates: resource allocations for new entities....1100
Creating templates....1100
VM template settings....1102
Host template settings....1102
HCI host template settings....1104
Storage template settings....1106
Billing and costs....1107
AWS price override....1113
AWS billing family recognition....1114
Administrative tasks....1115
Managing user accounts....1116
The Maintenance page....1127
License configuration....1129
Email settings....1131

API reference....1133
API guide revision history....1133

API guide revision history ... 1134
API change history ... 1135
Getting started with the Workload Optimization Manager REST API ... 1139
Workload Optimization Manager REST API Swagger documentation ... 1142
Workload Optimization Manager markets ... 1143
The real-time market ... 1143
Plan markets ... 1143
Running plans ... 1145
Working with policies in the API ... 1148
Placement policies requests ... 1148
Automation policies requests ... 1151
Parking policies requests ... 1157
Calculating reservations and workload placement ... 1157
Workload Optimization Manager REST API endpoints ... 1161
Actions endpoint ... 1161
Admin endpoint ... 1179
Authorization endpoint ... 1182
Businessunits endpoint ... 1185
Entities endpoint ... 1195
Groups endpoint ... 1222
Licenses endpoint ... 1257
Logs requests ... 1258
Markets endpoint ... 1258
Notifications endpoint ... 1280
Parking endpoint ... 1282
Policies endpoint ... 1285
Price lists utility requests ... 1287
Price lists requests ... 1288
Reservations endpoint ... 1288
Reserved instances endpoint ... 1296
Roles requests ... 1298
Scenarios endpoint ... 1298
Schedules endpoint ... 1338
Search endpoint ... 1351
Settings endpoint ... 1363
Settings policies endpoint ... 1366
Statistics endpoint ... 1372
Supplychains endpoint ... 1389
Tags endpoint ... 1392
Targets endpoint ... 1396

Templates endpoint....1406
Topology definitions endpoint....1415
Users endpoint....1418
Workflows endpoint....1434
Workload Optimization Manager cookbook....1438
Authenticating with the API....1439
Authenticating OAuth 2.0 clients with the API....1441
Plan results cookbook....1445
Creating a static group....1451
Creating a dynamic group....1457
Reference: Internal Workload Optimization Manager API requests....1461
Reference: Scenario peak-of-peak full scenario input DTO....1462
Reference: Tables....1478
Reference: DTO index....1484
AcceleratorCardApiDTO....1484
ActionApiDTO....1485
ActionApiInputDTO....1489
ActionDetailsApiDTO....1492
ActionExecutionApiDTO....1493
ActionExecutionAuditApiDTO....1493
ActionExecutionCharacteristicApiDTO....1494
ActionExecutionCharacteristicInputApiDTO....1494
ActionExecutionInputApiDto....1495
ActionResourceImpactStatApiInputDTO....1495
ActionResourceImpactStatInput....1497
ActionScheduleApiDTO....1498
ActionScopesApiInputDTO....1499
ActionScopesResourceImpactStatApiInputDTO....1500
ActionStateChangeApiDTO....1500
ActionWorkflowsApiDTO....1501
ActiveDirectoryApiDTO....1502
ActiveDirectoryGroupApiDTO....1502
AddObjectApiDTO....1503
AggregateCommodityFieldApiDTO....1504
AggregatedHealthResponseDTO....1505
AppliedScopeApiDTOObject....1506
AppliedScopeApiDTO....1506
AutomatedEntityDefinitionData....1506
AzureVirtualMachineScaleSetDTO....1507
AzureVirtualMachineScaleSetImageReferenceDTO....1508

AzureVirtualMachineScaleSetUpgradePolicyDTO..... 1509
BaseApiDTO..... 1509
BaseCommodityApiDTO..... 1509
BaseGroupApiDTO..... 1510
BasicActionApiDTO..... 1511
BillingFamilyApiDTO..... 1512
BooleanConditionApiDTO..... 1512
BooleanFieldValueApiDTO..... 1512
BulkActionRequestApiDTO..... 1513
BulkActionRequestInputDTO..... 1513
BusinessUnitApiDTO..... 1514
BusinessUnitApiInputDTO..... 1516
BusinessUnitPriceAdjustmentApiDTO..... 1517
BusinessUserEntityAspectApiDTO..... 1517
BusinessUserSessionApiDTO..... 1517
ChangePasswordApiDTO..... 1518
ClientConfigurationApiDTO..... 1518
ClientConfigurationEndpoint..... 1518
ClientInputDTO..... 1519
ClientNetworkRawTokenApiDTO..... 1519
ClientNetworkRawTokenMetadata..... 1520
ClientNetworkSitesMetadataApiDTO..... 1520
ClientNetworkTokenApiDTO..... 1520
ClientNetworkTokenApiOutputDTO..... 1521
ClientNetworkTokenInputDTO..... 1521
ClientNetworkTokensMetadataApiDTO..... 1521
ClientServiceApiDTO..... 1522
CloudAspectApiDTO..... 1523
CloudCommitmentAspectApiDTO..... 1525
CloudCommitmentCapacityApiDTO..... 1526
CloudCommitmentPricingApiDTO..... 1527
CloudCommitmentScopeDTO..... 1527
CloudCommitmentScopeInfoApiDTO..... 1527
CloudFamilyReferenceApiDTO..... 1528
CloudProvisionActionDetailsApiDTO..... 1528
CloudResizeActionDetailsApiDTO..... 1529
CloudServicePriceAdjustmentApiDTO..... 1530
CloudServicePricingModelApiDTO..... 1531
CloudSuspendActionDetailsApiDTO..... 1532
CommodityDetailsApiDTO..... 1532

CommodityFieldApiDTO..... 1533
ComputeTierAspectApiDTO..... 1534
ConditionApiDTO..... 1535
ConfigChangesApiDTO..... 1536
ConstraintApiDTO..... 1537
ConstraintApiInputDTO..... 1537
ContainerPlatformContextAspectApiDTO..... 1538
CostInputApiDTO..... 1539
CpuModelApiDTO..... 1539
CriteriaOptionApiDTO..... 1540
DBEntityAspectApiDTO..... 1540
DatabaseServerEntityAspectApiDTO..... 1543
DatabaseServerTierAspectApiDTO..... 1544
DatabaseServerTierInfoApiDTO..... 1545
DatabaseTierAspectApiDTO..... 1545
DemandEntityInfoDTO..... 1546
DemandReservationApiDTO..... 1546
DemandReservationApiInputDTO..... 1548
DemandReservationParametersDTO..... 1549
DeploymentInfoDTO..... 1549
DeploymentParametersDTO..... 1550
DeploymentProfileApiDTO..... 1550
DeploymentProfileParamApiDTO..... 1551
DeploymentProfileProviderApiDTO..... 1551
DeploymentProfileTargetApiDTO..... 1552
DesktopPoolEntityAspectApiDTO..... 1552
DetailDataApiDTO..... 1553
DiscountCoverageFilterApiDTO..... 1553
DiscoveredEntityApiDTO..... 1554
EffectiveCostFilterApiDTO..... 1554
EntityActionsApiDTO..... 1555
EntityAspect..... 1555
EntityCostApiDTO..... 1556
EntityCountRequestApiDTO..... 1556
EntityDetailsApiDTO..... 1557
EntityMetadataRequestApiDTO..... 1557
EntityPriceDTO..... 1557
EntityQueryApiDTO..... 1558
EntityStatsApiDTO..... 1558
EntityUptimeApiDTO..... 1559

EnumFieldValueApiDTO..... 1560
ExecutorInfoApiDTO..... 1560
ExecutorScheduleApiDTO..... 1561
ExecutorUserApiDTO..... 1561
FailedResourceApiDTO..... 1561
FamilyScopeApiDTO..... 1562
FieldApiDTO..... 1562
FieldValueApiDTO..... 1562
FieldValueTypeApiDTO..... 1563
FilterApiDTO..... 1563
FrontdoorIdpApiDTO..... 1564
GroupApiDTO..... 1564
GroupCountRequestApiDTO..... 1567
GroupMetadataRequestApiDTO..... 1567
GroupQueryApiDTO..... 1568
HealthCategoryResponseDTO..... 1568
HttpProxyDTO..... 1569
IEntityDefinitionData..... 1570
IManualConnectionsData..... 1570
IncludedCouponsApiDTO..... 1570
InclusionConditionApiDTO..... 1571
InputFieldApiDTO..... 1571
InstanceTypeFilterApiDTO..... 1573
InstanceTypeIdsApiDTO..... 1573
IntegerConditionApiDTO..... 1573
IntegerFieldValueApiDTO..... 1574
KubeturboConfigInputDTO..... 1574
KubeturboDeployOutputDTO..... 1576
LicenseApiDTO..... 1576
LinkNotificationCriteria..... 1578
LinkNotificationRequestApiDTO..... 1579
LinkRelation..... 1579
Link..... 1579
LoadChangesApiDTO..... 1580
LogEntryApiDTO..... 1581
LoggingApiDTO..... 1582
ManualDynamicConnections..... 1583
ManualEntityDefinitionData..... 1583
ManualGroupConnections..... 1583
ManualGroupsConnections..... 1584

ManualStaticConnections.... 1584
MarketApiDTO.... 1584
MasterImageEntityAspectApiDTO.... 1586
MaxUtilizationApiDTO.... 1586
MemberFieldApiDTO.... 1587
MigrateObjectApiDTO.... 1588
MultiTextFieldValueApiDTO.... 1589
NameValueInputDTO.... 1589
NetworkEntityAspectApiDTO.... 1590
NoSqlEntityAspectApiDTO.... 1590
NumberConditionApiDTO.... 1591
NumberFieldValueApiDTO.... 1592
OAuth2ClientApiDTO.... 1592
OAuth2ClientInputAccessToken.... 1593
OAuth2ClientInputTokenSettings.... 1593
OAuth2ClientOutputAccessToken.... 1594
OAuth2ClientOutputTokenSettings.... 1594
OAuth2CreateClientApiDTO.... 1594
OAuth2CreateClientApiInputDTO.... 1595
OAuthDataApiDTO.... 1596
OnDemandCostFilterApiDTO.... 1597
OnPremCostsManagerApiDTO.... 1597
OnPremResizeActionDetailsApiDTO.... 1598
OrderByApiDTO.... 1599
PMDiskAspectApiDTO.... 1599
PMDiskGroupAspectApiDTO.... 1599
PMEntityAspectApiDTO.... 1600
PaginationApiDTO.... 1602
ParkableEntityApiDTO.... 1602
ParkableEntityDetailsApiDTO.... 1605
ParkableEntityEnforcementPauseApiDTO.... 1608
ParkableEntityEnforcementPauseInputDTO.... 1609
ParkableEntityGroupApiDTO.... 1609
ParkableEntityInputDTO.... 1610
ParkableEntityRestrictionApiDTO.... 1611
ParkableEntityScheduleApiDTO.... 1611
ParkableEntitySmartParkingRecommendationApiDTO.... 1611
ParkableEntityStatsInputDTO.... 1612
ParkableEntityUUIDSetDTO.... 1612
ParkableEntityUnparkableRestrictionApiDTO.... 1612

ParkableVirtualMachineSpecApiDTO..... 1613
ParkableVirtualMachineSpecConfigApiDTO..... 1613
ParkableVirtualMachineSpecInstanceApiDTO..... 1614
ParkingEntitySchedulerApiDTO..... 1615
ParkingItemApiDTO..... 1615
ParkingPolicyApiDTO..... 1616
ParkingPolicyImpactsApiDTO..... 1616
ParkingPolicyStatsInputDTO..... 1617
ParkingPolicyTimeSpansApiDTO..... 1617
PatchedTargetDataApiDTO..... 1617
PatchedTargetFieldApiDTO..... 1618
PeakBaselineApiDTO..... 1618
PlacementInfoDTO..... 1618
PlacementOptionApiDTO..... 1619
PlacementOptionApiInputDTO..... 1620
PlacementParametersDTO..... 1620
PlanDestinationApiDTO..... 1621
PolicyApiDTO..... 1622
PolicyApiInputDTO..... 1624
PortChannelApiDTO..... 1625
PortsAspectApiDTO..... 1626
PowerVmPMEntityAspectDTO..... 1627
PowerVmVMEntityAspectDTO..... 1627
PriceAdjustmentDTO..... 1629
PrimitiveFieldApiDTO..... 1629
ProbeApiDTO..... 1629
ProbeDeploymentApiDTO..... 1630
ProbePropertyApiDTO..... 1631
ProbePropertyNameValuePairApiDTO..... 1631
ProductCapabilityDTO..... 1631
ProductVersionDTO..... 1632
PurchaseProfileDTO..... 1634
QueryInputApiDTO..... 1634
RIBuyActionDetailsApiDTO..... 1635
RIPurchaseProfilesSettingApiDTO..... 1635
RangeApiDTO..... 1636
RangeInputApiDTO..... 1637
Recommendation..... 1637
ReconfigureActionDetailsApiDTO..... 1637
RecurrenceApiDTO..... 1638

RegionAspectApiDTO..... 1639
RelatedActionApiDTO..... 1640
RelatedEntityFieldApiDTO..... 1640
RelatedGroupFieldApiDTO..... 1641
RelatedStatsApiDTO..... 1641
RelievePressureObjectApiDTO..... 1641
RemoveConstraintApiDTO..... 1642
RemoveObjectApiDTO..... 1643
ReplaceObjectApiDTO..... 1643
ReplaceWebhookBodyRequestApiDTO..... 1643
ReplaceWebhookBodyResponseApiDTO..... 1644
ReplicaApiDTO..... 1644
RequestHeader..... 1645
ReservationConstraintApiDTO..... 1645
ReservationFailureInfoDTO..... 1646
ReservationInvalidInfoApiDTO..... 1647
ReservedInstanceApiDTO..... 1647
ReservedInstancePurchaseSettingsDTO..... 1650
ResizeActionDetailsApiDTO..... 1650
ResourceApiDTO..... 1651
ResourceGroupApiDTO..... 1651
RoleApiDTO..... 1652
SAMLIdpApiDTO..... 1652
STEntityAspectApiDTO..... 1653
SaaSReportingConfigApiDTO..... 1655
SavingsInvestmentsApiDTO..... 1656
SavingsInvestmentsApiInputDTO..... 1657
SavingsInvestmentsFilterApiDTO..... 1658
SavingsInvestmentsSnapshotApiDTO..... 1658
SavingsInvestmentsTagApiDTO..... 1659
ScenarioApiDTO..... 1659
ScheduleApiDTO..... 1660
ScheduleEntityResponseApiDTO..... 1662
ScheduleItemApiDTO..... 1662
ScheduleTimeSpansApiDTO..... 1663
ScheduleTimeSpansInputApiDTO..... 1663
ScopeUuidsApiInputDTO..... 1664
SearchAllQueryApiDTO..... 1664
SearchCountRecordApiDTO..... 1665
SearchQueryRecordApiDTO..... 1665

SelectAllApiDTO ... 1666
SelectEntityApiDTO ... 1666
SelectGroupApiDTO ... 1667
ServiceEntityApiDTO ... 1668
SettingActivePolicyApiDTO ... 1670
SettingApiDTOSerializable ... 1671
SettingApiDTOString ... 1673
SettingApiDTO ... 1675
SettingOptionApiDTO ... 1676
SettingsManagerApiDTO ... 1677
SettingsPolicyApiDTO ... 1677
SiteInformationApiDTO ... 1678
SkippedActionApiDTO ... 1679
StatApiDTO ... 1679
StatApiInputDTO ... 1681
StatCommoditySourceApiDTO ... 1682
StatFilterApiDTO ... 1682
StatHistUtilizationApiDTO ... 1683
StatPeriodApiInputDTO ... 1683
StatScopesApiInputDTO ... 1684
StatSnapshotApiDTO ... 1684
StatValueApiDTO ... 1685
StringListFilterApiDTO ... 1686
StringSettingApiDTO ... 1686
SupplyChainStatsApiInputDTO ... 1687
SupplychainApiDTO ... 1687
SupplychainEntryDTO ... 1688
SustainabilityInputApiDTO ... 1689
TagApiDTO ... 1690
TagKeyApiDTO ... 1690
TargetApiDTO ... 1690
TargetErrorDetailsApiDTO ... 1692
TargetHealthApiDTO ... 1693
TargetHealthSummaryApiDTO ... 1694
TargetOperationStageApiDTO ... 1694
TargetOperationStageStatusApiDTO ... 1695
TargetPropertiesApiDTO ... 1695
TargetThirdPartyErrorDetailsApiDTO ... 1696
TelemetryDTO ... 1697
TemplateApiDTO ... 1697

TemplateApiInputDTO....1699
TemplatePriceAdjustmentDTO....1701
TemplateValidationStatusApiDTO....1702
TextConditionApiDTO....1702
TextFieldValueApiDTO....1703
TimeBasedTopologyChangesApiDTO....1703
TimeSpanApiDTO....1704
TimeSpanGroupApiDTO....1704
TopologyChangesApiDTO....1704
TopologyDataDefinitionApiDTO....1705
TopologyScopeApiDTO....1706
TunnelComponents....1707
UnplacementDetailsApiDTO....1707
UnplacementReasonApiDTO....1707
UserApiDTO....1708
UtilizationApiDTO....1709
VMEntityAspectApiDTO....1710
VirtualDiskApiDTO....1713
VirtualDisksAspectApiDTO....1715
VirtualMachineSpecAspectApiDTO....1715
VirtualVolumeEntityAspectApiDTO....1717
WebhookApiDTO....1717
WeekDayTimeSpansApiDTO....1719
WhereApiDTO....1719
WidgetApiDTO....1720
WidgetElementApiDTO....1721
WidgetsetApiDTO....1722
WorkflowApiDTO....1723
WorkflowAspect....1724
WorkflowOperationRequestApiDTO....1724
WorkflowOperationResponseApiDTO....1725
WorkloadControllerAspectApiDTO....1725
Reference: Entity statistics....1726
ApplicationComponent statistics....1726
BusinessApplication statistics....1726
BusinessTransaction statistics....1727
BusinessUser statistics....1727
Chassis statistics....1728
ContainerPlatformCluster statistics....1728
ContainerPod statistics....1729

ContainerSpec statistics..... 1730
Container statistics..... 1730
DataCenter statistics..... 1731
DatabaseServer statistics..... 1734
Database statistics..... 1735
DesktopPool statistics..... 1736
DiskArray statistics..... 1737
IOModule statistics..... 1738
LogicalPool statistics..... 1738
Namespace statistics..... 1739
PhysicalMachine statistics..... 1740
Service statistics..... 1743
StorageController statistics..... 1743
Storage statistics..... 1743
Switch statistics..... 1744
ViewPod statistics..... 1744
VirtualDataCenter statistics..... 1745
VirtualMachine statistics..... 1745
VirtualVolume statistics..... 1749
WorkloadController statistics..... 1749

Integration - Embedded reporting..... 1751
Enabling embedded reporting..... 1751
Navigating to the Embedded Reports page..... 1759
(Optional) Report editing..... 1761
Embedded reports storage requirement estimates..... 1763
Custom reports..... 1765
Importing reports..... 1765
Creating reports..... 1765
Exploring SQL queries..... 1766
Schema tables..... 1766
action_group table..... 1766
cloud_service_cost table..... 1767
completed_action Table..... 1767
Entity table..... 1769
entity_cost table..... 1770
entity_savings table..... 1771
entity_types_with_cost table..... 1771
File table..... 1771
historical_entity_attrs table..... 1772
Metric table..... 1775

pending_action table..... 1778
pending_action_stats table..... 1779
run_health_checks table..... 1779
Scope table..... 1780
scope_to table..... 1782
scope_to_types table..... 1782
topology_stats table..... 1782
wasted_file table..... 1782
JSON data objects..... 1783
ActionCommodity JSON..... 1783
ActionEntity JSON..... 1783
ActionImpactedCost JSON..... 1784
ActionImpactedCosts JSON..... 1784
ActionImpactedEntity JSON..... 1784
BuyRiInfo JSON..... 1785
CommodityChange JSON..... 1785
CommodityPercentileChange JSON..... 1786
DeleteInfo JSON..... 1786
entity.attrs JSON..... 1786
EntitySettings JSON..... 1787
group.attrs JSON..... 1788
ImpactedMetric JSON..... 1789
MoveChange JSON..... 1789
Target JSON..... 1789
action.attrs JSON..... 1789
Enums..... 1790
AttachmentState Enum..... 1790
OSType Enum..... 1790
action_category Enum..... 1791
action_mode Enum..... 1791
action_state Enum..... 1791
action_type Enum..... 1791
attr_type Enum..... 1792
cost_category Enum..... 1792
cost_source Enum..... 1792
entity_state Enum..... 1793
entity_type Enum..... 1793
environment_type Enum..... 1795
file_type Enum..... 1795
metric_type Enum..... 1795

savings_type Enum..... 1799
severity Enum.....1800
terminal_state Enum..... 1800
Configuring an external TimescaleDB.....1800
Creating an external TimescaleDB.....1801
Manual Provisioning of users and databases.....1801
Automatic provisioning of users and databases..... 1804
Reference: Working with YAML files.....1806

Documentation overview

Documentation for this product release includes the following general topics.

- Release notes *(on page 23)* – Describes new features, improvements, deprecated/removed features, and known issues
- Installation *(on page 49)* – Helps you install Workload Optimization Manager in your on-prem environment
- Updating Workload Optimization Manager *(on page 132)* – Helps you perform a Workload Optimization Manager update
- Product overview *(on page 143)* – Provides an overview of the platform and its underlying architecture
- Getting Started *(on page 151)* – Describes login steps, the Home Page, actions, and policies
- Target configuration *(on page 173)* – Provides a list of targets that the product can monitor, and describes how to configure each target properly
- User interface reference *(on page 543)* – Provides a list entities discovered from targets, and describes how to configure plans, charts, and administrative settings
- API reference *(on page 1133)* – Helps you use the REST API as you script interactions with the product
- Integration - Embedded Reporting *(on page 1751)* – Describes how to set up Embedded Reporting, an add-on that stores a history of your managed environment and presents this history via dashboards and reports

Release notes

Release Date: January 24, 2025

These release notes describe the following for Workload Optimization Manager 3.13.0:

- New features
- Deprecated or removed features
- Improvements
- Known issues
- REST API changes

For any questions, contact your support representative.

What's new

NOTE:
Frequent changes to the product or third-party targets require that some features are updated, deprecated, removed, or no longer supported. For more information about these features, see Feature Updates and Notices *(on page 27)*.

Version 3.13.0

Application performance management

- **Support for monitoring multiple New Relic accounts**: Workload Optimization Manager now lists all New Relic accounts that are associated with the target user key on the **New Relic Accounts** tab on the target view page. With this new capability, you can also filter entities by New Relic account on the Search page. You can now view scoped New Relic accounts in a manner similar to groups and observe their associated supply chain.
- **Configuration of multiple Datadog targets for the same host**: You can now configure more than one Datadog target for the same host URL by using the new **Display name** configuration field to uniquely identify each Datadog target. For existing Datadog targets, the display name field is set to the target hostname by default. If you need to make changes to the target, you must confirm that this display name contains only alphanumeric, space, or hyphen characters. If the default display name does not meet these constraints, you must update the name before you can edit the target. For more information, see the Datadog target configuration documentation *(on page 189)*.

Container platform resource management

- **Versioned Helm chart for Kubeturbo**: You can now use a Helm chart that is hosted in a Helm chart repository in IBM GitHub Pages to deploy Kubeturbo. This chart delivery mechanism supports versioning, which enables seamless access to Kubeturbo updates and allows you to roll back Kubeturbo to a previous version if needed. It also simplifies your deployment process by eliminating the need to clone the Kubeturbo repository. For more information, see Deploying Kubeturbo through a Helm chart in a Helm chart repository *(on page 253)*.
- **OAuth 2.0 client credentials for Kubeturbo**: This release introduces support for OAuth 2.0 client credentials for Kubeturbo connections to your Workload Optimization Manager instance. These credentials are more secure than the local account credentials created from the Workload Optimization Manager user interface.

  If you previously set up local account credentials for Kubeturbo, switch to OAuth 2.0 client credentials by downloading the `oauth2_probe_migration.sh` script and then running it in the cluster where Kubeturbo is deployed.
- **New Kubeturbo deployment parameters**: You can now configure the following parameters in the Custom Resource (CR) for your Kubeturbo deployment to control where Kubeturbo runs and allow application workloads to run on dedicated nodes.
  - `kubeturboPodScheduling.affinity`
  - `kubeturboPodScheduling.nodeSelector`
  - `kubeturboPodScheduling.tolerations`

  For more information, see Kubeturbo Custom Resource values *(on page 249)*.
- **Go-based operator for Kubeturbo**: Kubeturbo now uses a Go-based operator to enable automatic changes to the Kubeturbo deployment. With this improvement, you no longer need to update or delete resources manually.
  - For Kubeturbo deployments in Red Hat OpenShift OperatorHub, the Go-based operator is available as the certified edition in Kubeturbo's stable channel. Kubeturbo automatically switches from the Helm-based operator to the Go-based operator if your approval strategy is Automatic. If your approval strategy is Manual, OperatorHub notifies you that an operator update is available. Approve the update when you see the notification.
  - For Kubeturbo deployments that use a Helm-based operator before version 3.12.3:
    - Switch to the Go-based operator by downloading the Kubeturbo operator YAML bundle and then running it in the same namespace where your Kubeturbo operator is deployed. For more information, see Downloading and applying the YAML bundle *(on page 272)*.
    - Run the following command to install the latest Kubeturbo operator CRD. Installing the latest CRD prevents a runtime error that may occur when updating the operator.

      ```
      kubectl apply -f
       https://raw.githubusercontent.com/turbonomic/kubeturbo-deploy/refs/heads/staging/conf
      ig/crd/bases/charts.helm.k8s.io_kubeturbos.yaml
      ```
- **Resize up actions for container vCPU and vMem requests**: Workload Optimization Manager can now recommend resize up actions for container vCPU and vMem requests to improve performance and mitigate performance risks. Currently, you can manually execute these actions in the Workload Optimization Manager user interface or directly in your container platform cluster. Automated action execution will be supported in a future release.

  To control the generation of resize up actions for container vCPU and vMem requests, configure analysis properties in container spec policies. Any analysis properties that you previously configured in container spec policies to control the generation of resize down actions are overridden by the new default analysis properties that apply to both resize up and resize down actions. For example, the target utilization midpoint for vCPU and vMem requests is now 110% to keep requests close to the average utilization. Review these container spec policies and make adjustments as necessary.

  For more information, see Resizing resource limits and requests *(on page 593)*.

Cloud resource management

- **Support for Amazon Redshift**: Workload Optimization Manager now discovers Amazon Redshift provisioned clusters and presents them as data warehouse entities in the supply chain. If there are no connections to a provisioned cluster for at least 24 hours, Workload Optimization Manager recommends suspending the cluster as a cost-saving measure. For more information, see Data warehouse *(on page 719)*.

To take advantage of this feature, add the following permissions to the policies for the IAM user that you set up for Workload Optimization Manager.

  - `redshift:DescribeClusters`
  - `redshift:DescribeTags`
  - `redshift:PauseCluster`

- **Multi-account AWS target**: You can now configure an AWS target to access multiple accounts. This feature eliminates the need to configure several targets to access these accounts. For example, if your AWS environment has a management account and five member accounts, you can now configure an AWS target to access all of your accounts, or only the management account and several member accounts.

  You can continue to use the AWS targets that you added before version 3.12.4. If you want to consolidate your existing targets to take advantage of this feature, delete the targets first and then configure a new multi-account AWS target to access the same accounts that were previously managed by the deleted targets.

  For more information about setting up a new multi-account AWS target and consolidating existing targets, see Setting up a multi-account AWS target *(on page 383)*.

- **Historically informed action generation for Azure workloads**: Workload Optimization Manager can now generate actions informed by 14 days of historical utilization data immediately after a new Azure service principal target is added. To take advantage of this feature, add at least one Azure service principal target.

  Currently, actions are generated for virtual machines, volumes, and SQL databases (DTU and vCore) in both global Azure and Azure Government accounts.

  The `metrics-adapter-azure-monitormetrics` component runs after you add an Azure service principal target.

- **New requirement for discovering Azure savings plans**: To enable the discovery of Azure savings plans, assign the built-in Savings Plan Reader role to the Azure service principal for Workload Optimization Manager. For more information, see Enabling discovery of Azure savings plans *(on page 445)*.

  NOTE:
  If you previously added the `Microsoft.BillingBenefits/savingsPlanOrders/savingsPlans/read` permission to the Azure service principal for Workload Optimization Manager, you can now remove the permission as it is no longer required for discovering savings plans.

- **Cloud VM scaling improvements**: This release introduces the following improvements to scale actions for cloud VMs:
  - **Scale actions for powered off cloud VMs**: Workload Optimization Manager can now recommend scale actions for cloud VMs that are powered off (for example, during maintenance). If you accept the actions, Workload Optimization Manager scales the VMs while they are powered off so that the changes are already applied when they are powered on.
  - **Discount coverage chart in scale actions for cloud VMs**: When you review the Action Details page for an action to scale a cloud VM, you will now see a discount coverage chart that plots coverage for the given observation period. This chart and the existing resource utilization charts explain usage trends that influence Workload Optimization Manager's scaling recommendations.
  - **Discount-aware scale actions for AWS VMs**: For AWS VMs that are covered by Compute Savings Plans, Workload Optimization Manager now considers Compute Savings Plan coverage when recommending scale actions for the VMs and calculating the cost impact of the actions. This feature results in more accurate scaling recommendations and cost reporting.
  - **Cost impact of scale actions for AWS VMs covered by discounts**: For AWS VMs that are covered by Reserved Instances or Compute Savings Plans, Workload Optimization Manager now calculates the cost impact of scale actions based on the on-demand cost and effective cost.
    - On-demand cost is the uncovered usage cost that considers the AWS-published on-demand rates and the VM's uptime.
    - Effective cost considers the on-demand cost and the amortized costs for the associated Reserved Instances or Compute Savings Plans.

You can configure the **Cost basis** setting in the user interface to specify the default cost that displays in the Action Details page for a scale action. For more information, see Cost basis *(on page 1108)*.

NOTE:
Add an AWS Billing target to take advantage of this feature.

- **New sustainability charts for public cloud VMs**: Workload Optimization Manager now visualizes energy consumption and carbon footprint data for public cloud VMs. When you scope to a public cloud VM from any of the supported providers and services, and then click the **Details** tab, you can view sustainability data in the **Energy** and **Carbon Footprint** charts. For more information, see the Energy chart *(on page 987)* and Carbon Footprint chart *(on page 978)* documentation.

On-prem resource management

- **Ordered resizing for VMs in consistent resizing groups**: When you create an automation policy for VMs and select **Consistent Resizing** as a scaling constraint, you can now opt to **Run automated resizes in sequence**. This option prevents possible downtime associated with resizing multiple VMs simultaneously. When you enable it, Workload Optimization Manager resizes VMs that are in the same consistent resizing group one at a time, starting with the least utilized VM and ending with the highest utilized VM. Workload Optimization Manager compares the total utilization of the virtual CPU and virtual memory for all VMs within the consistent resizing group.
- **Support for a vendor link in the Entity Information chart for vCenter based entities**: A new **Vendor link** field was introduced in the Entity Information charts for vCenter-based hosts, virtual machines, and storage devices. This field provides an external link that you can use to navigate to the appropriate section in vCenter.
- **Availability of virtual processors as a commodity for PowerVM hosts**: In IBM PowerVM targets, Workload Optimization Manager now visualizes virtual processors (VP) in resource charts to illustrate their allocation on logical partitions (LPARs). Operations teams can now easily understand and manage VP capacity and usage across their hosts. For more information, see IBM PowerVM *(on page 334)*.
- **Scaling target utilization settings for on-prem VMs**: You can now configure the **Scaling target vCPU Utilization** and **Scaling target vMem Utilization** settings in on-prem VM policies to control the resize actions that Workload Optimization Manager generates. Earlier, these settings could be configured only in cloud VM policies to control scale actions.
- **Sustainability metrics in on-prem VM resize actions**: Workload Optimization Manager now shows the projected change in carbon footprint and energy usage in the action details for on-prem VM resize actions. This information can help you understand the sustainability benefits of resize actions and prioritize them for execution accordingly. For more information, see Action details- VM resize impact on energy usage and carbon footprint *(on page 833)*.
- **On-prem VM level fault tolerance**: Many applications are load balanced against VMs at different tiers. You can now configure fault tolerance settings on groups of VMs to protect application performance if any VM failures or interruptions occur. When you configure a fault tolerance count of 1 or more, Workload Optimization Manager modifies its analysis across your group of VMs so that recommended scaling actions consider the High Availability (HA) needs for your application. Workload Optimization Manager also reserves the capacity needed to sustain performance in case one or more VMs fail. If a failure does occur, the other VMs in the group are able to take on the application workload. When you enable consistent resizing for a group of VMs, you can set the fault tolerance configuration to allow Workload Optimization Manager to generate more accurate actions. For more information, see On-prem VM policies - Fault tolerance *(on page 752)*.

IT management

- **ServiceNow targets are included in the IT Management category**: ServiceNow targets are now included in the IT Management Workload Optimization Manager target category. IT Management targets *(on page 354)* enable Workload Optimization Manager to understand and participate in your operational environment. These targets integrate with your change approval processes and optimize license usage and technology business management. Previously, ServiceNow targets were included in the Orchestrator category. No action is required to continue using your existing ServiceNow targets.

- **ServiceNow improvements and support for the Xanadu release**: The Workload Optimization Manager ServiceNow integration is enhanced with the following changes:
  - The Workload Optimization Manager ServiceNow App now supports the Xanadu release. Users can take advantage of the Xanadu release of ServiceNow, confident in the knowledge that the decisions made by Workload Optimization Manager can be governed by a change management process as required.
  - Automatic configuration item (CI) matching is now supported for Google Cloud Platform (GCP) and Nutanix VMs.
  - Preview support is available for configuring multiple ServiceNow targets in the same Workload Optimization Manager instance. Register your interest at Turbonomic - aiops.
  - The new `DisplayName` target property enables you to define a custom name for ServiceNow targets.
  - The Basic/OAuth target authentication configuration experience is improved.
  - ServiceNow targets are now located in the IT Management Workload Optimization Manager target category.
  - The Actions API response now includes `externalActionName` and `externalActionUrl` values, which provide the action name and the ServiceNow URL for the specific Change Request (CR). Integrations can use these objects for deep linking back to ServiceNow when consuming action data from Turbonomic.

User interface management

- **Learn More link in the Create Workflow page for Action UUID**: A new **Learn More** link is available for the **Action UUID** tooltip on the **Create Workflow** page under the **Request** body page. This link provides information about finding the UUID for an action that you want to test in your workflow.

Platform management

- **Concurrent discovery overrides for all probe types**: Administrators can now specify the number of concurrent discoveries per probe type. To set up the desired concurrent discovery property for a specific probe type, configure the setting under the `topology-processor` properties section.

  The following example shows how to set the value for the vCenter probe type.

```
spec:
  properties:
    topology-processor:
      concurrentDiscoveriesByProbeType.vcenter: 20
```

  For a list of probe types, see this YAML file in GitHub. Probe types are specified in the following format.

```
              <probe_type>:
                properties:
                  enabled:
                    description: Enable <probe_type> mediation
                    type: boolean
```

  For example, the AppDynamics probe type is specified as follows:

```
              appdynamics:
                properties:
                  enabled:
                    description: Enable AppDynamics mediation
                    type: boolean
```

Feature updates and notices

Frequent changes to the product or third-party targets require that some features are updated, deprecated, removed, or no longer supported. See the following sections for more information about these features.

Updates

The updates that are listed here will be made in the noted version. Consider the details and recommended actions that are provided.

Feature	Status	Details and recommended action
Buy RI actions	Disabled in version 3.10.5	Buy RI actions are disabled in the real-time market and are no longer displayed in the Action Center. These actions will be reintroduced in a future release along with improvements to the existing analysis. You can run Buy VM Reservations plans to view new Buy RI actions.
Prometheus target	Updated in 3.10.4	Starting in version 3.10.4, new deployments of Prometurbo automatically add Prometheus targets to the Observability category. Prometurbo deployments before version 3.10.4 automatically added Data Ingestion Framework (Prometheus) targets to the Custom category. After you update to version 3.10.4 or later, existing Data Ingestion Framework (Prometheus) targets are cloned and the clones are then added to the Observability category. The original targets remain in the Custom category but are now undiscoverable. Remove them immediately to clean up your targets list and avoid seeing target discovery failures.
Pure Storage targets running Purity 6.4.4 (Pure API 1.6)	Added to version 3.10.2	Workload Optimization Manager now supports Pure Storage targets running Purity 6.4.4 (Pure API 1.6).
Memory setting in host templates	Updated in version 3.9.4	Workload Optimization Manager now uses GB as the primary unit for setting the Memory capacity for host templates. Also, the Memory default value is now 1024 GB.
Microsoft Hyper-V 2022 and Microsoft System Center Virtual Machine Manager 2022	Added to version 3.8.1	Workload Optimization Manager now supports Microsoft Hyper-V 2022 and Microsoft System Center Virtual Machine Manager 2022 targets.
MySQL 8.0	Added to version 3.8.1	Workload Optimization Manager now supports MySQL 8.0 as an optional alternative database server for MariaDB for running Workload Optimization Manager.
Legacy savings and investments charts	Added to version 3.7.5	When the enhanced savings and investments charts were released in version 3.7.3, the documentation stated that legacy savings and investments data is not available in the enhanced charts, but can be extracted from the platform by using a script. Starting in version 3.7.5, legacy data is available in the legacy savings and investments charts. The charts are titled 'Legacy' so you can distinguish them from the enhanced charts.
Sustainability features	Documentation updated in version 3.7.4	When the first set of sustainability features *(on page 147)* were released in version 3.7.2, the documentation incorrectly listed three targets that fully support these features – HPE OneView, Cisco UCS Manager, and VMware vCenter. Currently, only VMware vCenter fully supports these features. The documentation has been updated accordingly.
Azure permissions	Updated in version 3.7.1	Permissions were updated to more accurately reflect the minimum permissions for discovering workloads and executing actions. For example, permissions with wildcards (*) were replaced with the exact permissions. For details, see Azure Permissions *(on page 447)*.

Feature	Status	Details and recommended action
Rate of Resize setting in on-prem VM policies	Updated in version 3.7.1	The Rate of Resize default value has changed from 2 to 3. If you have changed your default setting to 1 or want to keep the current default setting of 2, create a new policy that is scoped to all on-prem VMs and configure the Rate of Resize to your wanted setting.

Notices

Features are labeled based on the following definitions:

- Deprecated – The feature is still supported but no longer developed or enhanced. The feature is not recommended for use and might become obsolete. Cisco might remove it in a subsequent release of the product.
- Removed – The feature is no longer available in the product.
- Unsupported – The feature is no longer supported in the product.

> **NOTE:**
> When a specific release or version of an integration partner technology reaches end-of-life (EOL) or its end of support date, Workload Optimization Manager no longer provides support for that version. Workload Optimization Manager follows integration partners' official EOL timeline for version support. Targeting a unsupported version, or one that is no longer supported by the vendor, is at your own risk.

The changes that are listed here will be made in the noted version. Consider the details and recommended actions that are provided.

Feature	Status	Details and recommended action
Azure Av1-series virtual machines	Deprecated in version 3.12.4	Microsoft has retired the Azure Av1-series virtual machines. For more information, see the Azure documentation. When you view the VM Policy page in the Workload Optimization Manager user interface and click **Cloud Instance Types for Scaling**, the category for Azure Av1-series displays "Retired".
Various charts	Deprecated in version 3.12.4	The following charts are deprecated and will be removed from the product in a future release. Delete these charts from any custom dashboard and use the replacement charts if applicable. ▪ Cost breakdown by tags This chart is replaced by the Cost breakdown chart. ▪ Legacy savings / Legacy investments These charts show savings and investments before version 3.7.3. The method for tracking savings and investments was enhanced in version 3.7.3, and data from this enhanced tracking is available in the Savings and Investments charts. ▪ Location This chart shows the cloud regions discovered in your environment. The same information is available in Search when you click **Regions**. ▪ Volume summary This chart is replaced by the Entity breakdown chart.
UCS Director target	Removed in version 3.12.2	The UCS Director target was removed in version 3.12.2 and is no longer supported.

Feature	Status	Details and recommended action
		After updating to version 3.12.2, delete any existing UCS Director target from the Target Configuration page by selecting the target in the list and then clicking **Delete**. It is recommended that you delete the `ucsdirector` and `mediation-ucsdirector` sections in the Workload Optimization Manager Custom Resource (CR) file. For details, see (Optional) Enabling and Disabling Probe Components *(on page 95)*.
New Relic REST API key and GraphQL API key	Removed in version 3.11.1	The fields are replaced by the **User Key** field on the New Relic target configuration page.
Microsoft Enterprise Agreement target	Removed in version 3.11.1	Microsoft has retired the Azure Enterprise Reporting APIs that the target uses to discover billing data. For more information, see the Azure documentation. If you added a Microsoft Enterprise Agreement target, remove it from the user interface immediately and then add an Azure Billing target. For more information, see Notice: Microsoft Enterprise Agreement *(on page 463)*.
CentOS Linux 7 based virtual machine images (OVA and VHD)	Unsupported	CentOS Linux reached end-of-life on June 30, 2024 and is no longer supported. For details, see What to know about CentOS Linux EOL. Starting with version 3.10.3, Workload Optimization Manager virtual machine images are based on Rocky Linux. For more information about migrating from CentOS to Rocky Linux based images, see the following topics: ▪ Migrating from CentOS to Rocky Linux *(on page 118)*
Recommended RI Purchases chart	Removed from the cloud dashboard in version 3.10.3	Use the Potential Savings chart to view 'Buy RI' actions for your cloud workloads. **NOTE:** The Recommended RI Purchases chart remains in plan results.
Red Hat OpenShift 3.11	Deprecated in 3.9.3	Red Hat OpenShift 3.11 reached End of Life in June 2022 and End of Extended Life in July 2024. It is no longer supported by Red Hat. As such, it will not receive any new bug fixes or security fixes. For details about the deprecation, see this Red Hat page. Workload Optimization Manager supports Red Hat OpenShift 4.x.
Application Insights target	Removed in version 3.9.1	The Application Insights target was deprecated in version 3.8.3 and removed from Workload Optimization Manager in version 3.9.1. After upgrading to version 3.9.1, delete any existing Application Insights targets from the Target Configuration page by selecting the target in the list and then clicking **Delete**. It is recommended that you delete the `appinsights` and `mediation-appinsights` sections in the Workload Optimization Manager Custom Resource (CR) file. For details, see (Optional) Enabling and Disabling Probe Components *(on page 95)*.
MySQL 5.7	Unsupported since version 3.8.4	MySQL version 5.7 reached EOL on October 21, 2023 and is no longer supported. For details, see EOL MySQL. Because it is no longer support-

Feature	Status	Details and recommended action
		ed, references to MySQL 5.7 were removed from the documentation in version 3.8.4. Workload Optimization Manager supports MySQL 8.0 as an optional alternative database server for MariaDB for running Workload Optimization Manager.
Azure unmanaged volumes	No longer discovered or monitored, if unattached	Microsoft deprecated Azure unmanaged volumes (disks). In response, Workload Optimization Manager no longer discovers or monitors unmanaged volumes that are not attached to any VM. Workload Optimization Manager continues to discover and monitor unmanaged volumes that are attached to VMs, to establish their relationship with VMs in the supply chain. Workload Optimization Manager does not recommend actions for these volumes.
`Use hypervisor VMEM for Resize` setting in on-prem VM policies	Removed in version 3.7.4	This setting is replaced by the **Collect Virtual Machine Metrics** option in the target configuration pages for APM targets.
MySQL 5.6	Removed in version 3.7.0	MySQL version 5.6 reached EOL on February 5, 2021. Because it is no longer supported, references to MySQL 5.6 were removed from the documentation in version 3.7.0. For details, see EOL MySQL.
Billing breakdown and Estimated cost breakdown charts	Removed in version 3.7.1	The charts are removed from the product and can no longer be added to dashboards. If you added these charts to your dashboards, delete them immediately and start using the replacement chart, Workload Cost Breakdown.
VMware vCenter versions 6.0, 6.5, and 6.7	Unsupported since version 3.6.3	Update your VMware vCenter version to 7.0 or 8.0, which are fully supported.

Versioning explanation

Workload Optimization Manager versioning uses V-R-M elements (Version, Release, Modification) in the version number to express the status of a release.

Numbered element	Example	Description
V - Version number	**3**.X.X	▪ Changes to platform architecture or significant changes to data models
R - Release number	X.**2**.X	▪ Major feature changes
M - Modification number is 0 (zero)	X.X.**0**	▪ A quarterly release ▪ All Private Preview features from previous bi-weekly releases are now GA ▪ No new Private Preview features in this release
M - Modification number is greater than zero (1 or higher)	X.X.**3**	▪ A bi-weekly release ▪ Can include new Private Preview features ▪ Includes fixed issues

Numbered element	Example	Description
		NOTE: For API developers, the X.X.1 release can include final implementations of deprecated API features. These final implementations can make API changes that are not compatible with an earlier version.

Configuration requirements

For this release of Workload Optimization Manager, you must satisfy the following configuration requirements.

Compute and storage requirements

The requirements for running a Workload Optimization Manager instance depend on the deployment method and the size of the environment you are managing.

Virtual machine image installations of Workload Optimization Manager version 3.7.1 or later require at least 1.5 TB for disk storage (that uses Thin provisioning or Thick provisioning), and the `/var` partition must be at least 340 GB. For more information, see Installing on a Virtual Machine Image *(on page 50)*.

NOTE:
Before updating Workload Optimization Manager, you must extend the `/var` partition to at least 340 GB. If the `/var` partition requirement is not met, your upgrade may fail, new container images may not load properly, and you can run into an ImagePullBackOff issue for some of the images. Before upgrading, be sure that the partition has sufficient disk space. If you need to free up disk space to increase the size of the partition, you can delete old images that are not in use. For details, see Increasing Available Disk Space *(on page 74)*.

For more information, see Minimum Requirements *(on page 49)*.

Workload Optimization Manager updates and operator version

Workload Optimization Manager deploys as a cloud-native application on a Kubernetes cluster. This cluster can be pre-configured on a VM that you deploy, or you can deploy Workload Optimization Manager to a Kubernetes cluster in your environment. In either case, Workload Optimization Manager uses an Operator to manage the application deployment.

For different versions of Workload Optimization Manager, the version of Operator you use changes as follows:

Product version	Operator version
3.12.6 - 3.13.0	42.71
3.12.5	42.70
3.12.4	42.69
3.12.3	42.68
3.12.2	42.67
3.12.1	42.66
3.11.6 - 3.12.0	42.65
3.11.5	42.64
3.11.4	42.63

Product version	Operator version
3.11.3	42.62
3.11.2	42.61
3.11.1	42.60
3.10.6 - 3.11.0	42.59
3.10.5	42.58
3.10.4	42.57
3.10.3	42.56
3.10.2	42.55
3.10.1	42.54
3.10.0	42.53
3.9.6	42.52
3.9.5	42.51
3.9.4	42.50
3.9.3	42.49
3.9.2	42.48
3.9.1	42.47
3.8.7 - 3.9.0	42.46
3.8.6	42.45
3.8.5	42.44
3.8.4	42.43
3.8.3	42.42
3.8.2	42.41
3.8.1	42.40
3.8.0	42.39
3.7.7	42.38
3.7.6	42.37
3.7.5	42.36
3.7.4	42.35
3.7.3	42.34
3.7.2	42.33
3.7.1	42.32
3.7.0	42.31
3.6.6	42.30
3.6.5	42.29

Product version	Operator version
3.6.4	42.28
3.6.3	42.27
3.6.2	42.25
3.6.1	42.24
3.5.6 - 3.6.0	42.23
3.5.5	42.22
3.5.4	42.21
3.5.3	42.20
3.5.2	42.19
3.5.1	42.18
3.4.6 - 3.5.0	42.17
3.4.4 - 3.4.5	42.16
3.4.3	42.15
3.4.2	42.14
3.4.1	42.13
3.4.0	42.12

When you update Workload Optimization Manager, always include the matching version of Operator in the update. Online or offline updates that were completed according to the latest installation instructions *(on page 132)* automatically include the latest Operator.

Supported MariaDB version for OVA and VHD installations

For its default historical database on Open Virtual Appliance (OVA) and Virtual Hard Disk (VHD) installations, Workload Optimization Manager currently supports MariaDB version 10.6.19. This support includes comprehensive testing and quality control for Workload Optimization Manager usage of the historical database.

If you are running Workload Optimization Manager installed as an OVA or VHD image, update to MariaDB version 10.6.19 if not already done.

For more information, see Verifying Your MariaDB Version *(on page 59)*.

External PostgreSQL and TimescaleDB version requirements for embedded reporting

If you run embedded reporting using a deployment other than OVA and use an external PostgreSQL with a TimescaleDB extension, both PostgreSQL 15.x and TimescaleDB 2.17.1 are required. If you are currently running PostgreSQL 12.x and TimescaleDB 2.10.2, you must first upgrade PostgreSQL to version 15.x, and then upgrade TimescaleDB to version 2.17.1. For more information, including upgrade links and instructions for both PostgreSQL and TimescaleDB, see Upgrade PostgreSQL in the TimescaleDB documentation.

SQL Server modes for external databases

If you deploy Workload Optimization Manager to work with an external database instead of the included historical database, then you must specify the correct SQL Server modes for the database.

For more information, see Configuring a Remote Database *(on page 61)*.

Transport Layer Security requirements

By default, Workload Optimization Manager requires Transport Layer Security (TLS) version 1.2 to establish secure communications with targets. Most targets have TLS 1.2 enabled; however, some targets do not enable TLS or they enabled an earlier version. In that case, you see handshake errors when Workload Optimization Manager tries to connect with the target service. When you go to the Target Configuration view, you see a Validation Failed status for such targets.

NetApp filers often do not enable TLS and the supported version is TLS 1.0, which causes the NetApp target to fail validation.

If target validation fails because of TLS support, you see validation errors with the following strings:

- No appropriate protocol

  Ensure that you enable the latest version of TLS that your target technology supports.
- Certificates do not conform to algorithm constraints

  Refer to the documentation for your target technology (such as the NetApp documentation) for instructions to generate a certification key with a length of 1024 or greater on your target server.

Known issues

- **Known Issue:**

  > "Reconfigure actions are not displayed for Cloud VMs in Action Center."

  Starting in 8.14.5, Reconfigure actions are not displayed for Cloud VMs in the Action Center.
- **Known Issue:**

  > "A runtime error may occur when upgrading a Kubeturbo operator from a version prior to 8.14.3 to version 8.14.3 or later."

  For Kubernetes clusters other than Red Hat OpenShift (such as AWS EKS, Azure AKS, and Google Cloud GKE clusters), if your Kubeturbo operator is earlier than version 8.14.3 and you update to version 8.14.3 or newer, you may see the following runtime error in the Kubeturbo operator logs.

  ```
  panic: runtime error: invalid memory address or nil pointer dereference
  ```

  To work around this issue, perform the following steps:

  1. Apply the new CRD by running the following command.

     ```
     kubectl apply -f
      https://raw.githubusercontent.com/turbonomic/kubeturbo-deploy/refs/heads/staging/config/crd/bases/charts.helm.k8s.io_kubeturbos.yaml&lt;
     ```

  2. Wait for the operator to apply the updated CRD. Alternatively, to apply the CRD immediately, force delete the `kubeturbo-operator` pod.
- **Known Issue:**

  > "Kubeturbo 8.13.1 and older require Workload Optimization Manager Version in deployment `configMap`."

  When following the documentation to deploy Kubeturbo version 8.13.1 or older there is a manual step required to allow Kubeturbo to be successfully deployed and add the Container Platform as a target in Workload Optimization Manager. The following line containing the major version is required in `configMap`, with 8.14 as an example (for any 8.14.x releases)

  ```
  "communicationConfig": {
  ```

  ```
                        "serverMeta": {
  ```

```
                    "version": "8.14",
```

- **Known Issue:**

  > "Using Kubeturbo agents with different versions than the running Workload Optimization Manager version may result in discovery issues or failures."

  Using Kubeturbo agents with a different version than the running Workload Optimization Manager version may result in issues including discovery failures, a decreased number of actions generated, or no actions. This can in turn result in reduced cost savings and increased performance risks in your environment.

  To avoid these discrepancies, keep your Kubeturbo agent version the same as the Workload Optimization Manager version. For more information, see Updating Kubeturbo *(on page 295)*.

- **Known Issue:**

  > "After upgrading to Workload Optimization Manager 3.9.4, you may encounter a **yum upgrade** failure."

  After upgrading to Workload Optimization Manager 3.9.4 (using either online upgrade or offline upgrade), **yum upgrade** may fail as follows:

  ```
  "GPG key retrieval failed: [Errno 14] curl#37 - "Couldn't open
   file /etc/pki/rpm-gpg/PGDG-RPM-GPG-KEY-RHEL7"
  ```

  The following workaround will address this issue by installing the latest Postgres signed keys:

  ```
  sudo yum install -y
   https://download.postgresql.org/pub/repos/yum/reporpms/EL-7-x86_64/pgdg-redhat-repo-latest.noarch.
  rpm
  ```

  You must have internet connectivity to `download.postgresql.org`.

- **Known Issue:**

  > "The Workload Cost Breakdown chart may report inaccurate on-demand costs when a VM is covered by an AWS Savings Plan."

  The Workload Cost Breakdown chart may report inaccurate on-demand costs when a VM is covered by an AWS Savings Plan. This is because the chart considers Savings Plan discounts for any covered resource as part of on-demand costs and does not categorize these discounts separately.

- **Known Issue:**

  > "The Cost Breakdown by Tag chart incorrectly handles tags with near identical values."

  When tag values have identical alphanumeric characters but different special characters, the Cost Breakdown by Tag chart displays one tag value and ignores the rest. For example, it might display `ab-1`, and ignore `ab_1` and `ab%1`.

- **Known Issue:**

  > "The Discount Coverage chart does not include on-demand usage of memory or CPU for Google Cloud C3 instances."

  The Discount Coverage chart does not include on-demand usage of memory or CPU for Google Cloud C3 instances.

- **Known Issue:**

  > "For customers going through MySQL 5.7 to MySQL 8.0 in-place or full data migrations between a MySQL 5.7 instance and a MySQL 8.0 instance."

  For customers going through MySQL 5.7 to MySQL 8.0 in-place or full data migrations between a MySQL 5.7 instance and a MySQL 8.0 instance, you may run into View related errors as follows:

```
Issues reported by 'check table x for upgrade' command

vmtdb.app_daily_ins_vw - Table 'vmtdb.app_spend_by_hour' doesn't exist

vmtdb.app_daily_ins_vw - View 'vmtdb.app_daily_ins_vw' references invalid table(s) or column(s) or
 function(s) or definer/invoker of view lack rights to use them

vmtdb.app_daily_ins_vw - Corrupt

vmtdb.app_daily_upd_vw - Table 'vmtdb.app_spend_by_hour' doesn't exist

vmtdb.app_daily_upd_vw - View 'vmtdb.app_daily_upd_vw' references invalid table(s) or column(s) or
 function(s) or definer/invoker of view lack rights to use them
```

If this is the case, the following workaround can be used:

- Connect to the MySQL database server.
- Run the following commands, with the sample output provided:

```
mysql> use vmtdb;

Database changed

mysql> SHOW FULL TABLES IN vmtdb WHERE TABLE_TYPE LIKE 'VIEW';

+---------------------------------+

Tables_in_vmtdb         Table_type

+---------------------------------+

app_daily_ins_vw       VIEW

app_daily_upd_vw        VIEW

app_monthly_ins_vw      VIEW

app_monthly_upd_vw      VIEW

service_daily_ins_vw    VIEW

service_daily_upd_vw    VIEW

service_monthly_ins_vw  VIEW

service_monthly_upd_vw  VIEW

vm_daily_ins_vw         VIEW

vm_daily_upd_vw         VIEW

vm_monthly_ins_vw       VIEW

vm_monthly_upd_vw       VIEW

+---------------------------------+
```

- Proceed to drop all of these views as follows:

```
drop view app_daily_ins_vw, app_daily_upd_vw, app_monthly_ins_vw. app_monthly_upd_vw,
 service_daily_ins_vw, service_daily_upd_vw, service_monthly_ins_vw, service_monthly_upd_vw,
 vm_daily_ins_vw, vm_daily_upd_vw, vm_monthly_ins_vw, vm_monthly_upd_vw;
```

For more information about in-place MySQL upgrades, see Configuring a Remote Database *(on page 61)*.

- **Known Issue:**

> "Upgrade issues from Workload Optimization Manager version 3.3.2 or earlier to version 3.7.2 or later."

If you upgrade from Workload Optimization Manager version 3.3.2 or earlier to version 3.7.2 or later, you might see issues such as the Workload Optimization Manager UI forcing you to reset the administrator password and seeing no targets on the Targets page.

These issues occur because the data stored in the consul is lost during the upgrade. To avoid this issue, first upgrade Workload Optimization Manager to a version before 3.7.2, then upgrade again to your preferred version.

For example, if you are running 3.2.1 and want to upgrade to 3.7.5, you can upgrade to 3.7.1 first, then upgrade to 3.7.5.

- **Known Issue:**

> "When updating to Workload Optimization Manager version 3.7.3 or later from OVAs that run older versions of Kubernetes, you might encounter an error."

When updating to Workload Optimization Manager version 3.7.3 or later from OVAs that run older versions of Kubernetes, you might encounter the following error: `/opt/local/bin/t8c-license-service.sh >>>>> ERROR - Pre requirement check: FAILED`. The License Service cannot be installed since it is not supported in older versions of Kubernetes. The upgrade of other Workload Optimization Manager components and the normal use of Workload Optimization Manager are not affected.

This error affects Kubernetes versions earlier than 1.19. To check your Kubernetes version, run the following command: `kubectl version`.

If you encounter this error and need assistance, contact your support representative.

- **Known Issue:**

> "When deploying to certain Kubernetes environments, the prometheus-server may fail."

When deploying Workload Optimization Manager on certain customer-provided Kubernetes environments, the prometheus component might enter a crash loop due to a failure to write to attached volumes. To resolve this, add a `securityContext` that specifies an `fsGroup` in the `prometheus` section of the CR file as follows:

```
spec:

  prometheus:

    server:

      securityContext:

        fsGroup: 2000
```

If the `prometheus` section is not already present in the CR file, add a new one underneath the top level `spec` section.

After modifying the CR file, apply it to activate the change.

- **Known Issue:**

> "The Top VM and Top Host widgets on the PowerVM dashboard may display incorrectly. "

The Top VM and Top Host widgets on the PowerVM dashboard might encounter an issue related to temporary groups, which are ad-hoc groups created in the following scenarios:

- When you click on any of the nodes or entity donuts in the supply chain
- When you select entities in the search and scope the selection

To work around this issue, you can do one of the following:

- Edit the Top VM or Top Host widget to add the power commodities.
- Reopen the PowerVM group.
- Edit the URL to remove the `?entityType=` query string parameter.

- **Known Issue:**

  > "For vCenter Server environments, Workload Optimization Manager does not recognize DRS rules for VM restart dependencies that are based on `VM-VM Dependency` rules."

  For vCenter Server environments, Workload Optimization Manager does not recognize DRS rules for VM restart dependencies that are based on `VM-VM Dependency` rules.

  You may be able to achieve a similar effect by expressing dependencies via the `VM-Host` rule, or cluster affinity or anti-affinity rules.

- **Known Issue:**

  > "For AppDynamics environments, the platform cannot discover Database Servers if the target authentication uses oAuth for credentials."

  For AppDynamics environments, Workload Optimization Manager cannot discover Database Servers if the target authentication uses `oAuth` for credentials.

- **Known Issue:**

  > "Changes to a policy do not immediately show up in the user interface view of the affected scope."

  When you set the scope of the Workload Optimization Manager view to a group, you can then view the automation policies that impact the given group. If you edit a policy for that group (in Settings: Policies), and then scope the view to that group again, the policy changes do not appear in the display for that group.

  The display should update within ten minutes, after the next round of incremental discovery. If the condition persists, log out of your session and log in again to update the display.

- **Known Issue:**

  > "Optimized Improvements for plans do not include hosts to provision."

  For cases where actions indicate provisioning new hosts, the Optimized Improvements chart does not include the hosts to provision in the After Plan section.

- **Known Issue:**

  > "The Optimal Improvements chart can show incorrect data for hosts to be suspended."

  In cases where actions recommend that you suspend hosts, the Optimal Improvements chart should indicate no utilization on the hosts to be suspended. Under some circumstances, the chart can show utilization on these hosts. The result is incorrectly low values for utilization on the other hosts in the current scope.

- **Known Issue:**

  > "For Workload Optimization Manager deployments where the database uses a custom (non-3306) port for MariaDB or MySQL, the Parking feature cannot start."

  For Workload Optimization Manager deployments where the database uses a custom (non-3306) port for MariaDB or MySQL, the Parking feature cannot start. In these cases, suspend should be disabled from the `charts_v1alpha1_xl_cr.yaml` resource as follows:

```
spec:
  suspend:
    enabled: false
```

- **Known Issue:**

  > "New Relic has ended support for monitoring integration with Microsoft SQL Server 2012. "

  Workload Optimization Manager no longer supports monitoring and stitching of Microsoft SQL 2012 discovered through New Relic. We recommend that you upgrade your Microsoft SQL instance to a version New Relic supports.

- **Known Issue:**

  > "For Azure, billing information for RIs can show costs as NA."

  In Azure environments, if you configure a Microsoft Enterprise Agreement target, costs for RIs can show in the user interface as NA. For example, a VM that shows 100% RI coverage can show the Reserved Compute Cost as NA. This occurs because of a known gap in the data provided by the Microsoft EA API that the target probe uses to collect data.

- **Known Issue:**

  > "Cloud VMs might scale to very large instance types even if there are smaller instance types available for scaling."

  When a cloud VM with a specific disk count and disk type applies a policy that enables 'instance store aware scaling', Workload Optimization Manager might recommend scaling the VM to a very large instance type, even if there are smaller, less expensive instance types that can adequately meet the VM's resource requirements.

  To avoid this issue, disable instance store aware scaling and restrict the VM to its current instance family. For example, if an AWS VM is currently running the `i3.2xlarge` instance type, specify the `i3` instance family as a scaling constraint for the VM.

- **Known Issue:**

  > "For New Relic MySQL 8.0, DB Cache Hit Rate values are incorrect."

  For New Relic MySQL 8.0, DB Cache Hit Rate values are incorrect in Workload Optimization Manager. Because the percentage of queries that are retrieved from the cache (`db.qCacheHitRatio`) is not supported in New Relic MySQL version 8.0 and higher, the DB Cache Hit Rate values are no longer displayed in Workload Optimization Manager.

  For more information, see the New Relic documentation.

- **Known Issue:**

  > "The user interface shows incomplete Azure billed costs and metrics because of an issue with the Azure API."

  When attempting to discover Azure billing targets configured with partitioned cost exports on the first day of any month, Workload Optimization Manager is unable to find the directory containing the export files within a customer's Azure Storage account. The export files exist, but the Azure API does not return the correct storage location of that day's cost export files. As a result, the Workload Optimization Manager user interface may not reflect any billed cost that Azure exported on the first of the month, and the metrics dependent on that billed cost. The billed cost, including costs for the first day of the month, will start appearing on the second day of the month. Note that the billed cost is not limited to expenses incurred on the first day of the month.

- **Known Issue:**

  > "In a Nutanix environment, a Replace Hosts plan can fail to place VMs."

  You can configure a Replace Hosts plan on a Nutanix cluster to replace the hosts with HCI templates. However, the plan will fail to create the HCI hosts, and will result in unplaced VMs.

- **Known Issue:**

  > "Hardware Refresh to replace hosts with HCI templates can fail to place workloads."

  When running a Hardware Replace plan, the plan can fail to place workloads onto HCI hosts. The plan correctly places workloads if the plan scope is in a hyperconverged environment. If the scope is *not* in a hyperconverged environment, then you must scope the plan to an entire cluster, and you must configure the plan to replace all the hosts in the cluster with HCI templates.

- **Known Issue:**

  > "For Kubernetes environments with nodes running Linux with cgroup v2 enabled, you must use Kubernetes version 1.23.2 or later."

  For Kubernetes environments with nodes running Linux with cgroup v2 enabled, an issue in earlier versions of Kubernetes prevents Workload Optimization Manager from collecting CPU utilization data for the affected nodes. To collect CPU utilization from the cgroup v2 nodes, you must run Kubernetes version 1.23.2 or later.

- **Known Issue:**

  > "With Executed Actions charts, some data is missing for actions on entities that have been removed from the environment."

  When you view Executed Actions charts or export data from them, some data is missing for actions on entities that have been removed from the environment. For example, assume an action was executed on a storage volume, and that volume has later been removed from the environment. In that case, the exported data for that action will not include values that describe the removed volume.

REST API notes

This section describes issues that result in changes to the REST API. It is a running list of changes that accumulates until the next release of the API Guide. New versions of the guide include these changes. At that time, this list will start with new changes to the API.

Because the API is a full reflection of the Workload Optimization Manager product, some changes to the product necessarily cause changes in the API. This can include bug fixes and improvements to the product. For example, new commodities that you can get for a given entity type may be added, or a bug fix might change the data that is returned by an API request. IBM strives to make these changes backward-compatible whenever possible.

In some instances, changes that are not backward-compatible must be made. In this case, the change is deprecated for a period of time that should be sufficient for you to understand the change and implement an alternative in your scripts.

REST API changes

For this release (version 3.13.0), there are no new API changes to report. API changes from the previous version family are incorporated in the API Guide.

API deprecation history

The following endpoints, requests, DTOs, and/or parameters have been deprecated in the Workload Optimization Manager API. The changes were finalized and removed from the API in the listed release.

NOTE:
See the instructions for alternative usage for deprecated items.

3.12.1

No items were scheduled for deprecation in version 3.12.1.

3.11.1

API component	Changes
Endpoints	None
Requests	None
Request Parameters	None
DTOs	None
DTO Parameters	DTO: `RequestHeader` Parameter name: `updateSecret` Instruction: No replacement parameter is needed to update the secret in a header request. You need only to specify `"isSecret": true` and include the updated secret value in the header. If you do not specify a new value for the secret, the existing value is reused

3.10.1

API component	Changes
Endpoints	None
Requests	None
Request Parameters	▪ Request: `GET /search` Parameter Name: `entity_types` Instruction: This parameter was previously deprecated and is scheduled for removal in 8.12.1. The mandatory `types` or `group_type` parameters are the preferred parameters to specify an object or group type in a search request.
DTOs	▪ `BusinessUnitRelatedApiDTO` Instruction: This DTO was used with the `/pricelists` endpoints from Workload Optimization Manager Classic, which was never implemented in Workload Optimization Manager 8.x and was deleted in version 8.9.3. This DTO was never implemented or used in Workload Optimization Manager, but it is referenced in the `BusinessUnitApiDTO` DTO. Because the DTO was never implemented, no alternative is required.
DTO Parameters	▪ DTO: `ActionApiDTO` Parameter: `actionState` Enums: `RECOMMENDED` and `PENDING_ACCEPT` are removed as possible values for the `actionState` parameter. The parameter is otherwise unchanged. Instruction: Both these action *states* are replaced by the READY action state.

3.9.1

API component	Changes
Endpoints	▪ `GET /tags/<tagKey>/entities` Instruction: Use `GET /tags/entities?key=<tagKey>`. This endpoint better supports searches for certain special characters, such as a slash (/), within tag names.
Requests	None
Request Parameters	None
DTOs	None
DTO Parameters	None

3.8.1

API component	Changes
Endpoints	None
Requests	None
Request Parameters	None
DTOs	None
DTO Parameters	▪ DTO: `StatApiInputDTO` Parameter Name: `costPrice` Instruction: Use the dedicated `cost` endpoints in `/entities`, `/groups`, and `/markets` instead. **NOTE:** These are marked "In Development" currently but will reach GA soon. – `/entities/{entity_Uuid}/cost` – `/groups/{group_Uuid}/cost` – `/markets/{market_Uuid}/cost`

3.7.1

API component	Changes
Endpoints	None
Requests	None

API component	Changes
Request Parameters	None
DTOs	▪ DTO: `ScenarioApiDTO` Parameter Name: `changes` Instruction: Use the `loadChanges` or `configChanges` parameters.
DTO Parameters	▪ DTO: `StatApiInputDTO` Parameter Name: `costPrice` Instruction: Use the dedicated `cost` endpoints in `/entities`, `/groups`, and `/markets` instead. **NOTE:** These are marked "In Development" currently but will reach GA soon. – `/entities/{entity_Uuid}/cost` – `/groups/{group_Uuid}/cost` – `/markets/{market_Uuid}/cost`

3.6.1

API component	Changes
Endpoints	None
Requests	▪ `GET /search/market/` Instruction: Use `GET /markets`. ▪ `GET /search/target/` Instruction: Use `GET /targets`. ▪ `GET /groups/<groupId>/actions/<actionId>` Instruction: Do not use this call (not implemented). ▪ `GET, POST /target/<targetId>/stats` Instruction: Do not use this call (not implemented). ▪ `GET /businessunits/<busUnitId>/stats` Instruction: Use `GET /entities/<entityId>/stats` or `GET /stats/<entityId>/stats`. These calls use the business unit entity ID. ▪ `POST /businessunits/<busUnitId>/stats` Instruction: Use `POST /stats/<entityId>/stats` or `POST /stats`.
Request Parameters	None

API component	Changes
DTOs	None
DTO Parameters	▪ DTO: `ActionApiDTO` Parameter Name: `CurrentLocation` Instruction: The object in this parameter will only contain the following fields: – `links` – `uuid` – `className` – `displayName` – `discoveredBy` – `environmentType` – `vendorIDs` ▪ DTO: `ActionApiDTO` Parameter Name: `NewLocation` Instruction: The object in this parameter will only contain the following fields: – `links` – `uuid` – `className` – `displayName` – `discoveredBy` – `environmentType` – `vendorIDs` ▪ DTO: `ProductVersionDTO` Parameter Name: `buildUser` Instruction: Deprecated with no replacement. This parameter was not implemented in the 3.x product family. ▪ DTO: `GroupApiDTO` Parameter Name: `remoteId` Instruction: Deprecated with no replacement. This parameter was not implemented in the 3.x product family. ▪ DTO: `UserApiDTO` Parameter Names: – `roleUuid` – `roleName` Instruction: Deprecated. Use the `roles` parameter in `UserApiDTO`.
ENUM values	▪ `entityType:VMSpec` Instruction: Use `VirtualMachineSpec`. ▪ `entityTypes:VMSpec` Instruction: Use `VirtualMachineSpec`. ▪ `entityTypeFilter:VMSpec` Instruction: Use `VirtualMachineSpec`. ▪ `providerEntityType:VMSpec`

API component	Changes
	Instruction: Use `VirtualMachineSpec`. ▪ `relatedEntityType:VMSpec` Instruction: Use `VirtualMachineSpec`. ▪ `providerType:VMSpec` Instruction: Use `VirtualMachineSpec`.

3.5.1

API component	Changes
Endpoints	None
Requests	▪ `DELETE /entities/<entityId>/tags/<tagKey>` Instruction: Use `DELETE /entities/<entityId>/tags?key=<tagKey>`. ▪ `DELETE /groups/<groupId>/tags/<tagKey>` Instruction: Use `DELETE /groups/<groupId>/tags?key=<tagKey>`.
Request Parameters	None
DTOs	▪ `LicenseApiInputDTO` Instruction: Use `LicenseApiDTO`.
DTO Parameters	▪ DTO: `HttpProxyDTO` Parameter Name: `portNumber` Instruction: Use `proxyPortNumber`. ▪ DTO: `LogEntryApiDTO` Parameter Name: `reasonCommodity` Instruction: Use the array, `reasonCommodities`.

3.4.1

Behavior Changes:

- Pagination of Returned Data from `/search` requests:

  For searches, the API paginates the returned data. Starting with version 3.4.1, the default pagination limit is 100 entries, and the maximum pagination limit is 500. If you do not specify a pagination limit, then the API returns pages set to the default limit of 100. If you make no settings, and your search results include more than 100 entries, the results will be paginated. Your scripts should always check for a pagination cursor when processing search results.

  You can change these settings in the `charts_v1alpha1_xl_cr.yaml` resource. Find the entries for:

  – `apiPaginationDefaultLimit`
  – `apiPaginationMaxLimit`

API component	Changes
Endpoints	None
Requests	None

API component	Changes
Request Parameters	None
DTOs	▪ `ReportScheduleApiDTO` This release will remove `ReportScheduleApiDTO`.
DTO Parameters	▪ DTO: `ActionApiDTO` Parameter Name: `CurrentLocation` Instruction: The object in this parameter will only contain the following fields: – `links` – `uuid` – `className` – `displayName` – `discoveredBy` – `environmentType` – `vendorIDs` ▪ DTO: `ActionApiDTO` Parameter Name: `NewLocation` Instruction: The object in this parameter will only contain the following fields: – `links` – `uuid` – `className` – `displayName` – `discoveredBy` – `environmentType` – `vendorIDs` ▪ DTO: `ProductVersionDTO` Parameter Name: `buildUser` Instruction: Deprecated with no replacement. This parameter was not implemented in the 3.x product family. ▪ DTO: `GroupApiDTO` Parameter Name: `remoteId` Instruction: Deprecated with no replacement. This parameter was not implemented in the 3.x product family. ▪ DTO: `UserApiDTO` Parameter Names: – `roleUuid` – `roleName` Instruction: Deprecated. Use the `roles` parameter in `UserApiDTO`.

API deprecation contract

The following endpoints, requests, DTOs, and parameters are deprecated in the Workload Optimization Manager API. The changes will be finalized and removed from the API in the listed release.

NOTE:
See the instructions for alternative usage for deprecated items.

3.15.1

Currently, no items are scheduled for deprecation in version 3.15.1.

3.14.1

Currently, no items are scheduled for deprecation in version 3.14.1.

3.13.1

API component	Changes
Endpoints	None
Requests	None
Request Parameters	None
DTOs	None
DTO Parameters	DTO: `ServiceEntityApiDTO` ▪ Parameter name: `placedOn` Instruction: Use `placedOn` within `unplacementDetails` instead. ▪ Parameter name: `notPlacedOn` Instruction: Use `unplacementDetails.reasons` instead. ▪ Parameter name: `unplacedExplanation` Instruction: Use `unplacementDetails.reasons` instead. ▪ Parameter name: `remoteId` Instruction: Use `vendorIds` instead. DTO: `NetworkEntityAspectApiDTO` ▪ Parameter name: `clusterName` Instruction: Use `clusterNames` instead. DTO: `ParkableEntityApiDTO` ▪ Parameter name: `parkable` Instruction: Use `restrictions` instead. DTO: `UserApiDTO` ▪ Parameter name: `authToken` Instruction: Internal use only, no replacement required

Installation

This guide gives you information you need to install Workload Optimization Manager in your virtual environment, install your license, and get started managing your resources.

If you have any questions, contact Cisco support.

Minimum requirements

License requirements

To run Workload Optimization Manager on your environment, you must install the appropriate license. Licenses enable different sets of Workload Optimization Manager features, and they support a specified number of workloads in your environment.

User interface requirements

To display the Workload Optimization Manager user interface, you must log into the platform with a browser that can display HTML5 pages. Workload Optimization Manager currently supports the following browsers:

- Apple Safari
- Google Chrome
- Microsoft Edge
- Mozilla Firefox

Network addressing requirements

Workload Optimization Manager requires static IP addressing. Static IP setup is covered as a step when installing the Workload Optimization Manager VM image.

Compute and storage requirements

The requirements for running a Workload Optimization Manager instance depend on the deployment method and the size of the environment you are managing.

The requirements listed for each deployment method are recommendations that you should keep in mind as you plan your Workload Optimization Manager deployment.

- Virtual Machine image *(on page 50)*

Installing on a virtual machine image

You can get a download of the Workload Optimization Manager platform as a:

- VMware Open Virtual Appliance (OVA) 1.0 image
- Microsoft Hyper-V image

Minimum requirements

The requirements for running a Workload Optimization Manager instance depend on the deployment method and the size of the environment you are managing.

Workload Optimization Manager keeps a real-time representation of your environment in memory. The greater the number of entities to manage, and the more extensive the relationships between them, the more resources you need for the VM that runs Workload Optimization Manager. As the VM requirements increase, so do the requirements for the physical machine that hosts the VM.

The requirements that are listed for each deployment method are recommendations that you should keep in mind as you plan your Workload Optimization Manager deployment. After deploying Workload Optimization Manager, if you need to change memory capacity, CPU capacity, or both for the VM, you can shut it down, make changes, and then power it up again to use the new capacity.

NOTE:
The machine that hosts the Workload Optimization Manager platform must support the SSE4.2 instruction set. Support for this instruction set was introduced at different times for different chip manufacturers:

- Intel: November 2008
- AMD: October 2011

The machine that hosts Workload Optimization Manager should be newer than these dates. On a Linux system, you can run the following command to check for this support:

```
cat /proc/cpuinfo | grep sse4
```

For more information, see the glossary entry at http://www.cpu-world.com/Glossary/S/SSE4.html.

Usually you can run Workload Optimization Manager on a host that meets the following minimum requirements:

Supported VM Image technology		Storage requirements	Memory	CPUs
VMware	vCenter versions 7.0 and 8.0	1.75 TB or greater (using Thin provisioning or Thick provisioning) **NOTE:** The `/var` partition must be at least 340 GB.	- Default: 128 GB - For 10,000 VMs or less, 64 GB **NOTE:** If you plan to install a VM with 64 GB of memory, you must modify the default for VM memory.	8 vCPUs
Microsoft	Hyper-V Server 2012 R2 or later			

Cisco provides a VM image (an OVA or VHD file) which is preconfigured with two hard disk drives. A minimum of 1.80 TB is necessary to ensure that the drives have the proper amount of space for storage.

Installing the platform

You install the platform in two main steps:

1. Install the Workload Optimization Manager VM image on your network.

   This process installs and starts up the VM that hosts your instance of the Workload Optimization Manager platform.
2. Deploy the Workload Optimization Manager components on the VM.

Workload Optimization Manager VM image

Workload Optimization Manager installs as a VM that runs the Rocky Linux 9 OS. For more information about Rocky Linux 9 OS, see https://wiki.rockylinux.org/rocky/version/.

For each new version, Cisco delivers a VM image (Open Virtual Appliance (OVA) or Virtual Hard Disk (VHD)) that you install to run the product. Typically, you install this image once, and for subsequent updates to Workload Optimization Manager you execute product updates on that installed VM. This means two things:

- Product updates patch new components of the Workload Optimization Manager application stack onto the same Rocky platform that you got when you originally installed the VM image. Product updates do not affect the underlying OS.
- Over time, you might learn of important security patches for the Rocky distribution. It is your responsibility to keep the OS up to date. You can install these patches on your Workload Optimization Manager VM whenever necessary.

NOTE:
The VM image is released with the Rocky Linux 9 OS since it meets your overall security requirements. Rocky is used as long as it remains viable and secure. For more information, see https://wiki.rockylinux.org/rocky/version/.

OVA: Installing the vCenter image for on-prem environments

The first step to installing Workload Optimization Manager is to deploy the VM that will host the platform.

For vCenter Server environments, an OVA image is delivered for each quarterly release. If you want to run Workload Optimization Manager on vCenter Server, install the Quarterly Release and then update to a later point release if necessary.

NOTE:
For minimum requirements, 128 GB of memory for the VM that hosts Workload Optimization Manager is recommended. However, if you plan to manage a smaller environment (10,000 VMs or less), you can install on a VM that provides 64 GB of memory.

If you plan to install a VM with 64 GB of memory, then you must modify the default for VM memory.

To install the Workload Optimization Manager OVA:

1. Download the Workload Optimization Manager installation package.

   Navigate to the Workload Optimization Manager Software Download page (https://software.cisco.com/download/home/286328879/type/286317011/release) for links to the latest OVA image.

   The installation package includes the `cisco_cwom-<version>-<XXXXXXXXXXXXXX>.ova` file

   where <*version*> is the Workload Optimization Manager version number and <*XXXXXXXXXXXXXX*> is the timestamp.

   For example: `cisco_cwom-3.0.0-20190916164429000.ova`

   The OVA file deploys as a VM with the Workload Optimization Manager components ready for installation.
2. Import the OVA file into your datacenter.

   Use the vCenter Server client to import the OVA into your environment.
3. Deploy the Workload Optimization Manager VM.

Configure the VM that was deployed from the OVA file.

If you want to deploy a VM with 64 GB of memory, manually modify the default value for Memory:

a. Right-click the VM and select **Edit Settings**.
b. Type **64** for Memory.
c. Click **OK** to save the settings.
d. Power on the VM.

4. Open the remote console.

   For the Workload Optimization Manager VM that you just deployed:

   a. Choose the **Summary** tab.
   b. Click **Launch Remote Console**.

5. Set up the Workload Optimization Manager System Administrator account.

   a. In the remote console, log in with the following default credentials:
      - Username: `turbo`

        Do not use the account name, `root`.
      - Password: `vmturbo`

        Then, you will be prompted to enter a new password.
   b. Enter your new password.

      The new password must comply with the strong password policy (a mixture of uppercase and lowercase letters, numbers, and a symbol). Only you will know this new password.

      NOTE:
      Be sure to save the changes account credentials in a safe place. For security reasons, this is the only account that can access and configure the Workload Optimization Manager VM.

   c. Enter your new password again to verify it.

6. Update the root password.

   The platform uses the `root` account for certain processes, such as rolling up log messages in `/var/log/messages`. To ensure the account credentials are current, you must change the password:

   a. Open a SuperUser session.
      - In the remote console, enter `su -`
      - At the password prompt, enter the default password: `vmturbo`
   b. Reset a new password.

      After you log in as root with the default password, the system prompts you for a `New password`. This new password must comply with the strong password policy (a mixture of uppercase and lowercase letters, numbers, and a symbol). Only you will know this new password.

      NOTE:
      Be sure to save the root account credentials in a safe place.

   c. Exit the SuperUser session.

      Enter `exit`.

7. Perform other necessary configuration steps, and then install the Workload Optimization Manager components.

   To perform the required and important configuration steps for the Workload Optimization Manager instance, see General Configuration Tasks *(on page 57)*.

   To install the Workload Optimization Manager components, see Deploying the Workload Optimization Manager Components *(on page 113)*.

VHD: Installing the Microsoft Hyper-V image

The first step to installing Workload Optimization Manager is to deploy the VM that will host the platform.

For Hyper-V environments, Cisco delivers a Hyper-V image for each quarterly release. If you want to run Workload Optimization Manager on a Hyper-V VM, you can install the Quarterly Release, and then update to a later point release if necessary.

> **NOTE:**
> For minimum requirements, Cisco recommends 128 GB of memory for the VM that hosts Workload Optimization Manager. However, if you plan to manage a smaller environment (10,000 VMs or less), you can install on a VM that provides 64 GB of memory. (See Minimum Requirements *(on page 49)*).
>
> If you plan to install a VM with 64 GB of memory, then you must modify the default for VM memory. (See Deploy the Workload Optimization Manager VM *(on page 51)*).

To install Workload Optimization Manager:

1. Download the Workload Optimization Manager installation package.

   Navigate to the Workload Optimization Manager Software Download page (https://software.cisco.com/download/home/286328879/type/286317011/release) for links to the latest Hyper-V image.
2. Expand the .zip file and copy the contents, which includes the Virtual Machine image, to your Hyper-V server (either to your cluster shared volume or to a local hard drive).
3. Use the Import Virtual Machine Wizard in the Hyper-V Manager to import the Virtual Machine into your environment.
4. Make sure your virtual network adapter is connected to the correct virtual network.
5. Ensure the Workload Optimization Manager instance will have sufficient memory.

   Cisco recommends that you use static memory for your Workload Optimization Manager instance. However, you can specify static or dynamic memory for the instance. By default, the installation sets static memory to128 GB.
6. Start the Workload Optimization Manager appliance and record its IP address.
7. Set up the Workload Optimization Manager System Administrator account.
   a. Log into the VM's Hyper-V console with the following default credentials:
      - Username: `turbo`

        Do not use the account name, `root`.
      - Password: `vmturbo`

        Then, you will be prompted to enter a new password.
   b. Enter your new password.

      The new password must comply with the strong password policy (a mixture of uppercase and lowercase letters, numbers, and a symbol). Only you will know this new password.

      > **NOTE:**
      > Be sure to save the changes account credentials in a safe place. For security reasons, this is the only account that can access and configure the Workload Optimization Manager VM.

   c. Enter your new password again to verify it.
8. Update the root password.

   The platform uses the `root` account for certain processes, such as rolling up log messages in `/var/log/messages`. To ensure the account credentials are current, you must change the password:
   a. Open a SuperUser session.
      - In the remote console, enter `su -`
      - At the password prompt, enter the default password: `vmturbo`
   b. Reset a new password.

After you log in as root with the default password, the system prompts you for a `New password`. This new password must comply with the strong password policy (a mixture of upper- and lower-case letters, numbers, and a symbol). Only you will know this new password.

c. Exit the SuperUser session.

Enter `exit`.

9. Enable the NIC for the installed VM.

The Workload Optimization Manager instance configuration includes one NIC, but it is not enabled or connected to a network. Display the NIC in the Hyper-V Manager and enable it.

10. Perform other necessary configuration steps, and then install the Workload Optimization Manager components.

To perform the required and important configuration steps for the Workload Optimization Manager instance, see General Configuration Tasks *(on page 57)*.

To install the Workload Optimization Manager components, see Deploying the Workload Optimization Manager Components *(on page 113)*.

Deploying the Workload Optimization Manager components

After you install the Workload Optimization Manager VM that will host the platform, you can install the platform components.

First, gather the information that you need to run the installation:

- Network Time Source for Time Synchronization (optional)

  You can complete this step during installation or later. If you want to synchronize the VM's clock now, you are prompted for the Network Time Source. For more information about synchronizing the VM's clock, see Synchronizing Time *(on page 58)*.

- Your updated `root` password

  The installation script requires that you update the `root` password for the VM. If you followed the instructions in OVA: Installing the vCenter Image *(on page 51)* or in VHD: Installing the vCenter Image *(on page 53)*, then you have already completed this step.

When you are ready with the necessary information, you can run the installation script.

The IBM License Service, which collects license usage information that is used to create usage snapshots and reports, is installed automatically as part of the `t8cinstall.sh` script. For more information, see "Tracking License Usage with IBM License Service" in the *Installation Guide*.

The following steps describe the default installation process. If you want to customize your installation, then you should consider taking the steps in Stepwise Platform Deployment *(on page 113)*.

NOTE:
Starting with Workload Optimization Manager version 3.5.5, IBM Container Registry is used for all Workload Optimization Manager images for online upgrades and installations. Ensure that you have access to `https://*.icr.io` before continuing.

All OVA installs use Kubernetes v1.28.8. Docker commands no longer work. You can use `crictl` and `ctr` commands instead. For more information, see the following resources:

- Container runtime changes in Kubernetes 1.28 and beyond on the Kubernetes documentation site
- Mapping from dockercli to crictl on the Kubernetes documentation site

- crictl command reference and information on GitHub
- ctr command reference and information on GitHub

1. Start up the installation script.
   - Start a secure session (SSH) on your Workload Optimization Manager VM as the `turbo` user.
   - Run the installation script.

     ```
     /opt/local/bin/t8cInstall.sh
     ```

2. Verify that you configured a static IP address for the Workload Optimization Manager VM.

   After the components start, you will type this static IP address into a web browser to access the login page for the Workload Optimization Manager user interface.

   As a first step, the script prompts you with:

   ```
   Have you run the ipsetup script to setup networking yet? [y/n] n
   ```

   If you did not previously configure a static IP for the platform VM, the setup script will be called on so that you can configure a static IP, and continue the installation.

   If you configured a static IP for the platform VM, enter `y` to continue the installation.

   Because of dependencies between Workload Optimization Manager and the Kubernetes installation, it is not recommended to change the IP address after the Workload Optimization Manager installation. The VM must run with a static IP, regardless if this is a production, testing, or evaluation installation.

   If you followed the instructions in OVA: Installing the vCenter Image *(on page 51)* or in VHD: Installing the vCenter Image *(on page 53)*, then you already ran the `ipsetup` script to configure a static IP.

3. Wait while the script completes the installation.

   As the installation process continues, the script completes the following installation steps.
   - Configures the platform environment with the necessary certificates.
   - Configures the Kubernetes cluster on the VM.

     Configuring the Kubernetes cluster can take a few tries before it succeeds. For each try that does not succeed, you see messages similar to:

     ```
     To further debug and diagnose cluster problems, use 'kubectl cluster-info dump'.
     ```

     When the connection succeeds, the script advances to the next steps.
   - Establishes local storage for the platform.
   - Creates the Kubernetes namespace for the platform as `turbonomic`.
   - Configures authorization to access the required Kubernetes secrets.
   - Initializes the MariaDB database server to manage historical data for the platform.

     The script creates two accounts on the MariaDB that have full privileges:
     - `root@localhost`

       This account does not use a password. To connect via this account the user must be `system root`.
     - `mysql@localhost`

       This account does not use a password. To connect via this account the user must be `system mysql`.

       NOTE:
       For security reasons, Cisco recommends that you configure passwords for these accounts. You can connect with these accounts by using `sudo`. For example, `sudo mysql`. After you connect, you can then set passwords to these accounts. For more information, see the MariaDB Knowledgebase at https://mariadb.com/kb.

- Installs the Timescale database for embedded reports and the Data Exporter.
- Deploys and starts up the platform components.

As the deployment begins, the script prints the following:

```
############################
Start the deployment rollout
############################
```

After it deploys the components, it waits for the components to start:

```
The installation process is complete, waiting for all the components to start up.
** The script will wait for as long as 30 minutes. **
```

If the components all start within 30 minutes, then the installation is complete and successful.

If the components do not all start within 30 minutes, the script displays the following and then exits:

```
=============================================================================
    One or more of your deployments has not started up yet.
    ** Please give your environment another 30 minutes to stablize. **
    To check the status of your components, execute the following command:
    kubectl get pods
    If some components are still not ready, contact your support representative
    Deployments not ready:
```

The script then displays the formatted result of the `kubectl get pods` command. This shows you the status of the pods in the Workload Optimization Manager platform.

If the script exits before the components all start, give the platform another 30 minutes. To periodically check the component status, run the `kubectl get pods` command. If the components do not all start after another 30 minutes, contact your support representative.

If the installation is successful and all of the components started, the script displays a message similar to the following, where it gives the VM's static IP address:

```
################################################
Deployment Completed, please login through the UI
https://10.10.123.123
################################################
```

You can move on to the next steps.

4. Save a copy of the platform's Master Key secret.

   The installation procedure creates a Master Key secret in the Kubernetes cluster. Workload Optimization Manager uses this secret to provide access for the platform components. Save the key data to a safe location.

   To save the data:

   a. List the platform secrets.

      ```
      kubectl get secrets
      ```

      The results should include the Master Key secret.

      ```
      ...
      master-key-secret             Opaque                       1      57d
      ...
      ```

   b. Display the Master Key data.

      After you find the Master Key name, you can then display the key data.

      ```
      kubectl get secret master-key-secret -o yaml
      ```

      The output is similar to the following example:

```
apiVersion: v1
data:
  primary_key_256.out: AfnJWutxNHAduaIOdAii3DRA2fMa6lzX4rWetZxxZvc=
kind: Secret
metadata:
  creationTimestamp: "2021-06-30T02:59:19Z"
  managedFields:
  - apiVersion: v1
    fieldsType: FieldsV1
    fieldsV1:
      f:data:
        .: {}
        f:primary_key_256.out: {}
      f:type: {}
    manager: kubectl-create
    operation: Update
    time: "2021-06-30T02:59:19Z"
  name: master-key-secret
  namespace: turbonomic
  resourceVersion: "1072"
  uid: a314b2ba-2061-4b41-b844-56caf2c3728d
type: Opaque
```

The important key data to save is the `primary_key...` data. In this example, you should save the line:

```
primary_key_256.out: AfnJWutxNHAduaIOdAii3DRA2fMa6lzX4rWetZxxZvc=
```

c. Save the data to a safe place.

Write this data to a file and save it in a safe backup location. If you ever need to recover the Master Key, your support representative uses this data to perform the recovery.

5. Log in to the Workload Optimization Manager user interface using the server address and import your license key.

For more information, see License installation and first-time login *(on page 100)*.

6. Create other user accounts with various roles. For more information about user accounts and roles, see Managing user accounts *(on page 1116)*.

NOTE:
For security reasons, you can create a different account with an `Administrator` role to serve as the main administrator of your Workload Optimization Manager installation, and then delete the default `administrator` account. You must always have at least one user account with administrator privileges.

General configuration tasks

After you install the Workload Optimization Manager instance, you should perform the following configuration tasks:

- (Required) Synchronize the system clock and configure your time servers
- (Important) Verify your MariaDB version
- (Optional) Increase available disk space
- (Optional) Enforce secure access via LDAP
- (Optional) Enforce secure access via trusted certificate
- (Optional) Enable secure access for probes
- (Optional) Modify the certificates for Cluster Manager
- (Optional) Enable embedded reports
- (Optional) Enable the Data Exporter

- (Optional) Change the IP address of the platform node
- (Optional) Enable and disable probe components
- (Optional) Disable component health monitoring

(Required) Synchronizing time

It is important that you synchronize the clock on the Workload Optimization Manager instance with the other devices on the same network. By default, the Workload Optimization Manager server is configured to synchronize with any one of the following time servers:

- `0.centos.pool.ntp.org`
- `1.centos.pool.ntp.org`
- `2.centos.pool.ntp.org`
- `3.centos.pool.ntp.org`

To synchronize with these servers, your installation of Workload Optimization Manager must have access to the internet. If your environment restricts internet access, configure synchronization with a time server on your network.

In all cases, verify that the Workload Optimization Manager clock is properly synchronized.

1. Open an SSH terminal session to your Workload Optimization Manager instance.

   Log in with the System Administrator that you set up when you installed Workload Optimization Manager:

   - Username: `turbo`
   - Password: `[your_private_password]`

2. Verify your time settings.

   Run the `date` command.

   The output is similar to the following example:

   ```
   Thu Feb 2 14:25:45 UTC 2019
   ```

   Run the `timedatectl` command.

   The output is similar to the following example:

   ```
        Local time: Fri 2019-12-06 21:09:26 UTC
    Universal time: Fri 2019-12-06 21:09:26 UTC
          RTC time: Fri 2019-12-06 21:09:27
         Time zone: UTC (UTC, +0000)
       NTP enabled: yes
  NTP synchronized: yes
   RTC in local TZ: no
        DST active: n/a
   ```

   This tells you whether you have NTP enabled, and whether it is synchronized, along with other time synchronization information.

If the output is correct and your environment has access to the internet, you can assume that the system clock is synchronized.

If the output is incorrect, or if you need to configure synchronization with a time server on your network, you must configure `chrony` on the server instance.

To set up `chrony` on your Workload Optimization Manager instance:

1. Open an SSH terminal session to your Workload Optimization Manager instance.
2. Open the `chrony` configuration file.

   ```
   sudo vi /etc/chrony.conf
   ```

3. Specify the time servers that you want to use in your environment.

   The `chrony` file includes the following statements to configure time servers:

   ```
   server 0.centos.pool.ntp.org iburst
   server 1.centos.pool.ntp.org iburst
   server 2.centos.pool.ntp.org iburst
   server 3.centos.pool.ntp.org iburst
   ```

   Enter statements for the servers that you want to use. Then delete or comment out the statements that you do not want to use.

   Specify a time server by running the following command syntax:

   ```
   server <My_Time_Server_Name> iburst
   ```

4. Save the file.
5. Restart the chrony service.

   ```
   sudo systemctl restart chronyd
   ```

6. Verify that your time is correct.

   Run the `date` command.

   The output is similar to the following example:

   ```
   Fri Dec 6 21:09:26 UTC 2019
   ```

   To verify that the time is synchronized, run the `timedatectl` command.

   The output is similar to the following example:

   ```
         Local time: Fri 2019-12-06 21:09:26 UTC
     Universal time: Fri 2019-12-06 21:09:26 UTC
           RTC time: Fri 2019-12-06 21:09:27
          Time zone: UTC (UTC, +0000)
        NTP enabled: yes
   NTP synchronized: yes
    RTC in local TZ: no
         DST active: n/a
   ```

   To verify the time, compare the `date` output with the output from a known Coordinated Universal Time time server.

(Important) Verifying your MariaDB version

For its historical database, Workload Optimization Manager supports long-term maintenance stable series of MariaDB (the minimum supported version is 10.6) and MySQL 8.0.x. This support includes comprehensive testing and quality control for Workload Optimization Manager usage.

If you are running Workload Optimization Manager installed as a VM image (OVA or VHD), and using the database that is included in that image installation, then you must use version 10.6.19. If you are updating your version of Workload Optimization Manager (instead of installing it for the first time), ensure that you are using the correct version of MariaDB with your installation.

IMPORTANT:
It is a requirement that you run MariaDB version 10.6.19 or later. Workload Optimization Manager can operate with other versions of MariaDB; however, it is fully tested to operate with MariaDB version 10.6.19.

- For VM image installations, it is possible to configure the installation to use a remote database (external to the VM). See Configuring a Remote Database *(on page 61)*
- For installations on a Kubernetes cluster (not deployed as a Workload Optimization Manager VM image), you can find a download package at: `https://archive.mariadb.org/mariadb-10.6.19/yum/centos7-amd64`.

Checking your MariaDB version

Check the version of MariaDB running on your Workload Optimization Manager OVA.

1. Open an SSH terminal session to your Workload Optimization Manager instance.

   Log in with the System Administrator that you set up when you installed Workload Optimization Manager:

   - Username: `turbo`
   - Password: `[your_private_password]`
2. Check the MariaDB version.

   ```
   mysql -V
   ```

   The output is similar to the following example:

   ```
   mysql  Ver 15.1 Distrib 10.6.19-MariaDB, for Linux (x86_64) using readline 5.1
   ```

If the version is earlier than `10.6.19-MariaDB`, then you must update your database.

If your version is equal to or greater than `10.6.19-MariaDB` you do not need to update your MariaDB.

Updating your MariaDB

If you are using Workload Optimization Manager installed as a VM image, and you are using the default MariaDB that was installed with that image, you must run MariaDB version 10.6.19.

To update your MariaDB on your Workload Optimization Manager VM:

1. Open an SSH terminal session to your Workload Optimization Manager instance.

   Log in with the System Administrator that you set up when you installed Workload Optimization Manager:

   - Username: `turbo`
   - Password: `[your_private_password]`
2. Ensure that the VM is mounted on the Workload Optimization Manager update ISO image.

   NOTE:
   When you complete a Workload Optimization Manager update, the system automatically unmounts the ISO image. To update MariaDB, your Workload Optimization Manager instance must be mounted on the same ISO image that you used to update to a recent Workload Optimization Manager version. For more information, see Offline Update *(on page 137)*.

3. Run the MariaDB update script.

   Before you run the script, you need to know the MariaDB password. By default, this password is `vmturbo`.

   a. Contact your support representative for the latest version of this script.

   b. Run the database update script.

   ```
   sudo /opt/local/bin/mariadbUpgrade.sh |& tee -a $HOME/Turbonomic_MariaDB_upgrade.log
   ```

   The script updates the version of MariaDB. It also increases size limits for the allowed packets, and buffer and log sizes for the innodb. The script output includes the following (where `Total Memory` and `buffer pool size` can vary depending on your VM configuration):

   ```
   ==============================
   Update the mariadb configuration
   ==============================
   Total Memory: 128773 MB
   Changing Innodb buffer pool size to: 9216 MB
   Changing max allowed packets to: 1G
   Changing innodb log file size to: 10G
   ==============================
   ```

4. Verify the updated MariaDB version.

   When the script completes, you are running version 10.6.19. To verify, run:

   ```
   mysql -V
   ```

   The output is similar to the following example:

   ```
   mysql  Ver 15.1 Distrib 10.6.19-MariaDB, for Linux (x86_64) using readline 5.1
   ```

5. Scale up the Workload Optimization Manager platform's pods.

   To update the database, the script scales down your platform pods. When it completes, the script displays the following prompt:

   ```
   ################################################################################
   When confirmed the mariadb has been upgraded and is properly working, run:
   kubectl scale deployment --replicas=1 t8c-operator -n turbonomic
   ################################################################################
   ```

   After you verify that the correct version of MariaDB is running, scale up the platform.

   ```
   kubectl scale deployment --replicas=1 t8c-operator -n turbonomic
   ```

(Optional) Configuring a remote database

If you installed Workload Optimization Manager as a VM image (OVA or VHD), the image includes MariaDB for the historic database. While most installations use that instance, you can use an external MariaDB or MySQL database instead. In addition, you can install Workload Optimization Manager directly to a Red Hat OpenShift or Kubernetes cluster (instead of installing the VM image), and you can provide your own historical database.

For its historical database, Workload Optimization Manager supports long-term maintenance stable series of MariaDB (the minimum supported version is 10.6) and MySQL 8.0.x. This support includes comprehensive testing and quality control for Workload Optimization Manager usage.

Task overview

1. (Only applicable for Azure MySQL databases) Create the Azure Database for MySQL Flexible Server.
2. Configure a remote database server.
3. Configure Workload Optimization Manager.

Creating the Azure Database for MySQL flexible server

1. In the Azure portal, search for and select **Azure Database for MySQL flexible servers** and click **Create**.

   NOTE:
   Azure Database for MySQL servers is scheduled for retirement 16 September 2024.

2. On the **Basics** tab, specify the project and server details based on the load requirements.
   - Specify the following information:

Fields	Instructions
Subscription	Specify the Azure subscription used for billing.
Resource Group	Specify the container that holds related resources for an Azure solution.
Server Name	Specify a name for the MySQL server.
Region	Specify the Azure region where the server is created. Choose a region close to the users who will access the server.
MySQL Version	Specify to use MySQL 8.0.
Admin username	Specify the primary administrator username for the server.
Password	Specify the administrator password.

   For more information, see the Azure documentation.

   - To specify the workload type, click **Configure server** next to **Compute + storage**. You can create an Azure Database for MySQL - Flexible Server in one of three different service tiers: Burstable, General Purpose, and Business Critical. The recommended configuration is to use **General Purpose** with compute size and storage size as shown in the following table.

Topology size	Preferred Server type	Storage size
Less than 100 K	`Standard_D8ads_v5 (8 vCores, 32 GB memory, 12800 max iops)`	500 GB
100 K-1 million	`Standard_D16ads_v5 (16 vCores, 64 GB memory, 20000 max iops)`	2 TB
Over 1 million	`Standard_D32ads_v5 (32 vCores, 128 GB memory, 20000 max iops)`	3 TB

   - (Optional) If you want to automatically provision and manage a standby replica in a different availability zone within one region, select **Enable high availability**. You can then select to use the zone redundant high availability mode to improve the availability.

     NOTE:
     These settings might lead to a substantial increase in the cost. Current on-prem OVA deployments do not use high availability.

   - The default backup retention period is seven days. The backup retention is billed based on extra storage that is used for retaining backups. For more information, see the Azure documentation.

3. Click **Save**.
4. Configure networking options.

   Add any firewall and private end-point details in the Networking tab.
5. Configure security options.

   Configure any Azure key vault or user-managed identity that is used for the encryption in the Security tab. Otherwise, leave the default values.
6. Configure tags.

   Add any tags for the deployment in the Tags tab.
7. Click **Review + create** to review your Azure Database for MySQL flexible server configuration.
8. Click **Create** to provision the server. Provisioning can take a few minutes.

Configuring a remote database server

Before you install Workload Optimization Manager, you must configure a remote database server for historical data. This configuration allows for high availability, backups, and handling larger IO demands for larger environments or from remote SQL Server queries. If you are not going to use an existing database server, you must first create a new database server.

You can configure a MariaDB *(on page 63)*, MySQL *(on page 64)*, or Azure MySQL *(on page 65)* database server. Refer to the steps for the database server you are configuring.

> **NOTE:** If you want to use the fully qualified domain name (FQDN) for the remote database server, you need to use the `externalDBName` parameter. Remove the `externalDbIP` parameter from the following examples and replace it with the `externalDBName` parameter.

Configuring MariaDB

> **NOTE:**
> Use only long-term maintenance stable series of MariaDB. The minimum supported version is 10.6.

1. Create a remote database with the following settings for a MariaDB 10.6.19 instance:
   - Cluster configuration - the main/replica where Workload Optimization Manager communicates to a single replica
   - Select network and security policies that allow communication between the Kubernetes cluster and the database server; the default DB port is 3306
   - Minimum memory - 8 GB memory and 2 vCPU
   - Storage - 100 GB (set to auto-increase) or provide 1024 GB (or 1 TB)
   - Encryption (SSL) is supported
2. Open the configuration file in an editor and change the values for these system variables.

   > **NOTE:**
   > In Amazon Web Services (AWS), create a new Relational Database Service Parameter Group.

```
event_scheduler = ON
sql_mode = ERROR_FOR_DIVISION_BY_ZERO,NO_AUTO_CREATE_USER,NO_ENGINE_SUBSTITUTION
max_allowed_packet = 1GB
log_bin_trust_function_creators = 1
character-set-server = utf8mb4
collation-server = utf8mb4_unicode_ci
innodb_buffer_pool_size = <see the following note>
```

NOTE:
The `innodb_buffer_pool_size` must be at least 4096 M (or 4G). Increase the value to support larger topologies. Refer to the following table for suggested values based on system memory.

System memory (mem)	innodb_buffer_pool_size
`mem <= 32GB`	`4G`
`mem > 32GB and mem <= 64GB`	`8G`
`mem > 64GB and mem <= 128GB`	`16G`
`mem > 128GB`	`32G`

3. Save your changes.
4. Create a Workload Optimization Manager admin user that the deployment uses to create application users and grant privileges. You can define your own password, but it must be reflected in the Custom Resource (CR) files that are described in the next section.

   Using a MySQL client or CLI, run the following commands, substituting your credentials against your database server:

```
CREATE USER 'turboadmin'@'%' IDENTIFIED BY 'vmturbo';
GRANT SELECT, INSERT, UPDATE, DELETE, CREATE, DROP, RELOAD, PROCESS, REFERENCES, INDEX, ALTER,
 SHOW DATABASES, CREATE TEMPORARY TABLES, LOCK TABLES, EXECUTE, REPLICATION SLAVE, REPLICATION
 CLIENT, CREATE VIEW, SHOW VIEW, CREATE ROUTINE, ALTER ROUTINE, CREATE USER, EVENT, TRIGGER ON *.*
 TO 'turboadmin'@'%' WITH GRANT OPTION;
```

   NOTE:
   If you are using an AWS Relational Database Service (RDS) or Azure Database service, see AWS RDS or Azure Database Service Configuration *(on page 72)* for steps to create the application database users, databases (empty schema), and grant privileges.

   For information on using Kubernetes Secrets for DB app username and password credentials, see the Using Kubernetes secrets with a remote database *(on page 66)* section in this topic.

   The database server is ready to be included in your Workload Optimization Manager installation.
5. After the database is configured, configure Workload Optimization Manager *(on page 70)*.

Configuring MySQL

NOTE:
MySQL version 5.7 reached EOL on 21 October 2023 and is no longer supported. Workload Optimization Manager supports MySQL 8.0.x as an alternative database server for MariaDB for running Workload Optimization Manager.

For more information, see Reference: Upgrading from MySQL 5.7 to 8.0 *(on page 117)*.

1. Open the configuration file for editing.

   NOTE:
   By default, with Centos 7.9, the Configuration File is found at `/etc/my.cnf`. With RHEL 8.8, the Configuration File is found at `/etc/my.cnf.d/mysql-server.cnf`.

2. Modify the CR file to set the required server system variables.

```
event_scheduler = ON
sql_mode = ERROR_FOR_DIVISION_BY_ZERO,NO_ENGINE_SUBSTITUTION
```

```
max_allowed_packet = 1GB
skip-log-bin
log_bin_trust_function_creators = ON
explicit_defaults_for_timestamp = OFF
character-set-server = utf8mb4
collation-server = utf8mb4_general_ci
innodb_buffer_pool_size = <see the following note>
```

NOTE:
The `innodb_buffer_pool_size` must be at least 4096 M (or 4G). Increase the value to support larger topologies. Refer to the following table for suggested values based on system memory.

System memory (mem)	innodb_buffer_pool_size
`mem <= 32GB`	`4G`
`mem > 32GB and mem <= 64GB`	`8G`
`mem > 64GB and mem <= 128GB`	`16G`
`mem > 128GB`	`32G`

If you are using an AWS Relational Database Service (RDS) or Azure Database service, see AWS RDS or Azure Database Service Configuration *(on page 72)* for steps to create the application database users, databases (empty schema), and grant privileges.

3. Save your changes.
4. Restart MySQL for the changes to take effect.

   With Rocky 9.3, the following command restarts MySQL:

   ```
   systemctl restart mysqld.service
   ```

5. After the database is configured, configure Workload Optimization Manager *(on page 70)*.

Configuring Azure MySQL Database

1. After the Azure MySQL database is created, make the following configuration changes in the **Settings > Server Parameters** section.

   ```
   event_scheduler = ON
   sql_mode = ERROR_FOR_DIVISION_BY_ZERO,NO_ENGINE_SUBSTITUTION
   max_allowed_packet = 1073741824=
   log_bin_trust_function_creators = ON
   explicit_defaults_for_timestamp = OFF
   character_set_server = utf8mb4
   collation_server = utf8mb4_general_ci
   sql_generate_invisible_primary_key = OFF
   innodb_buffer_pool_size = <see the following note>
   ```

NOTE:
The `innodb_buffer_pool_size` must be at least 4096 M (or 4G). Increase the value to support larger topologies. Refer to the following table for suggested values based on system memory.

Azure VM Memory (mem)	innodb_buffer_pool_size
`mem = 32GB`	`4G`
`mem = 64GB`	`8G`
`mem = 128GB`	`16G`

2. Save your changes.

Using Kubernetes secrets with a remote database

If you are using a remote database, you can use a Kubernetes Secret to store the database credentials and then specify it in your Workload Optimization Manager Custom Resource (CR) file.

For its historical database, Workload Optimization Manager supports long-term maintenance stable series of MariaDB (the minimum supported version is 10.6) and MySQL 8.0.x. This support includes comprehensive testing and quality control for Workload Optimization Manager usage.

Several options exist to configure the Workload Optimization Manager deployment to use Kubernetes Secrets that store database credentials. The option that works best for you depends on how you configure access to your remote database. Complete the following steps to use Opaque Type Kubernetes Secrets to obfuscate the database credentials that are required to connect to a remote database. The Workload Optimization Manager deployment looks for the DB-credentials Secret first by default.

NOTE:
For more information about Kubernetes Secrets, see the Kubernetes documentation. For information on how to use a webhook injection method that supports Kubernetes Secrets, see the Kubernetes Webhook documentation.

Scenario 1 - Single database user with default application users

This use case is applicable to any of the following database services where the user allows the Workload Optimization Manager application to manage its own application database users.

- Allow Workload Optimization Manager to create application users with a default password
  - AWS Relational Database Service (RDS) services of MariaDB or MySQL
  - Google Cloud Storage services of MySQL
  - Bring your own MariaDB and MySQL server

In this scenario, the Workload Optimization Manager configuration needs to supply:

- DB service endpoint
- Connection properties for the MySQL dialect, secure access, and port
- Global credentials of the DB root user

  The global credentials are placed into a Kubernetes Secrets and the name of the Secret is referenced in the Workload Optimization Manager Custom Resource file.

1. Create the global user Kubernetes Secrets.

   Modify the `stringData` field of the `global-secret.yaml` to include the `db-creds`.

   The stringData field type allows you to enter the content in plain text, which is then converted to a Base64 encoded Secret when you create it on the Kubernetes cluster.

```
apiVersion: v1
kind: Secret
metadata:
  name: globaldb
type: Opaque
stringData:
  db-creds: |-
    username: <yourDBUser>
    password: <yourUncodedDBUserPassword>
```

Where

- name – the Secret name can be anything, but it must be the same as what is specified in the Workload Optimization Manager Custom Resource file.
- username – *dbRootUsername*
- password – *dbRootPassword*; the password must be plain text

2. Apply the Secret.

   Apply the Secret to the Workload Optimization Manager namespace before deploying Workload Optimization Manager.

```
kubectl create -f global-secret.yaml -n turbonomic
```

3. With a remote database, create the empty application database.

   Even with a single global database, you must first create the empty application databases when you are using a remote database. Connect to your database with a db admin user and run the following commands.

```
create database action;
create database auth;
create database clustermgr;
create database cost;
create database group_component;
create database market;
create database plan;
create database repository;
create database suspend;
create database topology_processor;
create database vmtdb;
create database actions_history;
create database oauth2;
create database sustainability;
```

4. Modify the Workload Optimization Manager Custom Resource file to use the Secret.

   Update with the Custom Resource file `yaml` file as shown in the following example:

```
global:
  externalDBName: yourDB.cloud.com
  dbSecretName: globaldb
properties:
  global:
    enableSecureDBConnection: true
    sqlDialect: MySQL
```

NOTE:
Ensure the `dbSecretName` is the same as the Secret name specified in the `global-secret.yaml`. In this example, the Secret name is `globaldb`.

5. Make any other configuration changes for your Workload Optimization Manager instance, then save the CR file.
6. Apply the changes that you made to the CR file.

   For example, run the following command for VM image installations:

```
kubectl apply -f \
  /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
```

Scenario 2 – Specifying all application users

This use case is applicable to the following scenarios where the user wants or needs to first create the application database users, and then provide this information to the Workload Optimization Manager deployment through the Custom Resource file.

- Manage the password for app users (not use default)
- Usernames that need to specify the DB instance in a multi-tenant database server
- Azure RDS services (MySQL, MariaDB) which require that the username specify the DB instance

> **NOTE:**
> Azure Database for MariaDB is scheduled for retirement by 19 September 2025. As part of this retirement, there is no extended support from Microsoft for creating new MariaDB server instances from the Azure portal beginning 19 January 2024. If you still need to create MariaDB instances to meet business continuity needs, you can use Azure CLI until 19 March 2024.
>
> For more information, see the Azure documentation.

In this scenario, the Workload Optimization Manager configuration needs to supply:

- DB service endpoint
- Connection properties for the MySQL dialect, secure access, and port
- Global credentials of the DB root user and Workload Optimization Manager admin user
- Application users for action, auth, clustermgr, cost, group, market, plan orchestrator, repository, suspend, topology processor, vmtdb (history), actions_history, oauth2 and sustainability

> **NOTE:**
> All DB users must be defined in separate Kubernetes Secrets.

1. Create the global user Kubernetes Secrets.

   Modify the `stringData` field of the `global-secret.yaml` to include the `db-creds`.

   The stringData field type allows you to enter the content in plain text, which is then converted to a Base64 encoded Secret when you create it on the Kubernetes cluster.

```
apiVersion: v1
kind: Secret
metadata:
  name: globaldb
type: Opaque
stringData:
  db-creds: |-
    username: <yourDBUser>
    password: <yourUncodedDBUserPassword>
```

   Where

   - name – the Secret name can be anything, but it must be the same as what is specified in the Workload Optimization Manager Custom Resource file.
   - username – *DBRootUsername*

NOTE:
If required, the username will contain the DB name: `turboadmin@mydb`.

- password – *`DBRootPassword`*

2. Apply the Secret.

   Apply the Secret to the Workload Optimization Manager namespace prior to deploying Workload Optimization Manager.

```
kubectl create -f global-secret.yaml -n turbonomic
```

3. Create the application user Secrets.

   Modify the `action.yaml` so that each application user has their own Secret.

```
apiVersion: v1
kind: Secret
metadata:
  name: actiondb
type: Opaque
stringData:
  db-creds: |-
    username: <yourActionDBUser>
    password: <yourUnencodedActionDBUserPassword>
```

   Where :

   - name – the Secret name can be anything, as long as it is the same as what is specified in the Workload Optimization Manager Custom Resource file.
   - username – `app component username`

     NOTE:
     If required, the username contains the db name: `action@mydb`.

   - password – `app component username`; the password must be plain text

   Define a Secret for the action, auth, clustermgr, cost, group, history, plan-orchestrator, repository, topology-processor, actions-history, oauth2 and sustainability components. In total, you create thirteen Secrets (twelve app secrets and one global secret).

4. Apply the Secrets.

   Apply all Secrets to the Workload Optimization Manager namespace before deploying Workload Optimization Manager.

```
kubectl create -f action.yaml -n turbonomic
```

5. Modify the Workload Optimization Manager Custom Resource file to use the Secrets.

   Update with the Custom Resource file `yaml` file as shown in the following example:

```
global:
  externalDBName: yourDB.cloud.com
  dbSecretName: globaldb
properties:
  global:
    enableSecureDBConnection: true
    sqlDialect: MySQL
  action-orchestrator:
    dbSecretName: actiondb
  auth:
    dbSecretName: authdb
  clustermgr:
    dbSecretName: clustermgrdb
```

```
  cost:
    dbSecretName: costdb
  group:
    dbSecretName: groupdb
  history:
    dbSecretName: historydb
  market:
    marketSecretName: marketdb
  plan-orchestrator:
    dbSecretName: plandb
  topology-processor:
    dbSecretName: topodb
  repository:
    dbSecretName: repodb
  suspend:
    dbSecretName: suspenddb
  actions-history:
    actions-history: actionhistorydb
  oauth2:
    dbSecretName: oauth2db
  sustainability:
    dbSecretName: sustainabilitydb
```

NOTE:
If you have other properties that are defined for these components, you must combine them.

6. Make any other configuration changes for your Workload Optimization Manager instance, then save the CR file.
7. Apply the changes that you made to the CR file.

   For example, run the following command for VM image installations:

```
kubectl apply -f \
  /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
```

Encoding credentials

Opaque type Secrets can take as input for data field type values that are already Base64 encoded. Encode the Secrets by running the following command in a Linux / MacOS environment and place the output in your configuration file.

```
echo -n 'yourCredential' | base64
```

Configuring Workload Optimization Manager

After your database is configured properly, the next step is to declare access to it through the Workload Optimization Manager Custom Resource (CR) file.

1. Open the CR file for editing.

   The location of the file depends on the type of Workload Optimization Manager installation that you are configuring.

   - VM Image installation of Workload Optimization Manager

     Log in to an SSH terminal session of your Workload Optimization Manager instance as the System Administrator that you set up when you installed Workload Optimization Manager.

     ```
     /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
     ```

   - Workload Optimization Manager on a Kubernetes node or node cluster

     ```
     deploy/crds/charts_v1alpha1_xl_cr.yaml
     ```

- Azure MySQL

```
kubectl edit xl -n <your Workload Optimization Manager namespace>
```

2. Modify the CR file to include the external database properties.

   Add the entry to the CR file, by using the following example as a guide. This example also contains optional parameters. For a complete list of options, refer to the Custom Resource definition for the Workload Optimization Manager operator under `APIV3Schema` in the `Validation` section of the CRD yaml here.

   NOTE:
   The `externalDBName` parameter creates an external type Kubernetes Secrets service, and therefore the DB name must be the Fully Qualified Domain Names (FQDN), not the short name. Optionally, you can use the `externalDbIP` parameter, which creates a Kubernetes Secrets endpoint and service.

   MariaDB configuration example:

```
spec:
  # Global settings
  global:
    repository: icr.io/cpopen/turbonomic
    tag: 3.13.0
    externalDbIP: <MariaDB_host_IP>
  properties:
    global:
      dbRootUsername: <MariaDB_root_username>
      dbRootPassword: <MariaDB_root_complex_password>
```

   MySQL and Azure MySQL configuration example:

```
spec:
  # Global settings
  global:
    repository: icr.io/cpopen/turbonomic
    tag: 3.13.0
    externalDbIP: <MySQL_host_IP>
  properties:
    global:
      sqlDialect: MYSQL
      enableSecureDBConnection: true
      dbRootUsername: <MySQL_root_username>
      dbRootPassword: <MySQL_root_complex_password>
```

   NOTE:
   If you are using an AWS RDS or Azure Database service, see AWS RDS or Azure Database Service Configuration *(on page 72)* for extra parameters to add to the CR file.

   Alternatively, you can specify the database username and password by using Kubernetes Secrets or by using a vault + webhook injection method. For more information, see the Using Kubernetes secrets with a remote database *(on page 66)* section in this topic or this site.

   (Azure MySQL databases) If the database IP is not available, specify the `externalDBName` instead. For example, `externalDbName: MySQL_8.0_host_Name`.

3. Save the changes to the CR.
4. Apply the changes that you made to the CR file.

NOTE:
For Azure MySQL databases, the changes apply automatically after you save the changes and exit the live edit. You do not need to perform this step.

For example, run the following command for VM image installations:

```
kubectl apply -f \
  /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
```

AWS RDS or Azure database service configuration

If you are using a supported AWS RDS (MariaDB or MySQL) or Azure Database PaaS service, you must create the application databases and users and then provide this information to the Custom Resource file. You edit the CR file when you configure the database and again when you configure Workload Optimization Manager to use the external database.

Configuring the database

Create the application users and the empty databases that Workload Optimization Manager will use. Using a MySQL client or CLI, run the following commands against your database server.

NOTE:
You can define custom passwords but do not change the username.

```
create user 'action'@'%' identified by 'vmturbo';
create user 'auth'@'%' identified by 'vmturbo';
create user 'clustermgr'@'%' identified by 'vmturbo';
create user 'cost'@'%' identified by 'vmturbo';
create user 'group_component'@'%' identified by 'vmturbo';
create user 'market'@'%' identified by 'vmturbo';
create user 'plan'@'%' identified by 'vmturbo';
create user 'repository'@'%' identified by 'vmturbo';
create user 'suspend'@'%' identified by 'vmturbo';
create user 'topology_processor'@'%' identified by 'vmturbo';
create user 'history'@'%' identified by 'vmturbo';
create user 'actions_history'@'%' identified by 'vmturbo';
create user 'oauth2'@'%' identified by 'vmturbo';
create user 'sustainability'@'%' identified by 'vmturbo';
create database action;
create database auth;
create database clustermgr;
create database cost;
create database group_component;
create database market;
create database plan;
create database repository;
create database suspend;
create database topology_processor;
create database vmtdb;
create database actions_history;
create database oauth2;
create database sustainability;
GRANT SELECT, INSERT, UPDATE, DELETE, CREATE, DROP, REFERENCES, INDEX, ALTER, CREATE TEMPORARY TABLES,
 LOCK TABLES, EXECUTE, CREATE VIEW, SHOW VIEW, CREATE ROUTINE, ALTER ROUTINE, EVENT, TRIGGER ON action.*
 TO 'action'@'%';
GRANT SELECT, INSERT, UPDATE, DELETE, CREATE, DROP, REFERENCES, INDEX, ALTER, CREATE TEMPORARY TABLES,
 LOCK TABLES, EXECUTE, CREATE VIEW, SHOW VIEW, CREATE ROUTINE, ALTER ROUTINE, EVENT, TRIGGER ON auth.* TO
 'auth'@'%';
```

```
GRANT SELECT, INSERT, UPDATE, DELETE, CREATE, DROP, REFERENCES, INDEX, ALTER, CREATE TEMPORARY TABLES,
 LOCK TABLES, EXECUTE, CREATE VIEW, SHOW VIEW, CREATE ROUTINE, ALTER ROUTINE, EVENT, TRIGGER ON
 clustermgr.* TO 'clustermgr'@'%';
GRANT SELECT, INSERT, UPDATE, DELETE, CREATE, DROP, REFERENCES, INDEX, ALTER, CREATE TEMPORARY TABLES,
 LOCK TABLES, EXECUTE, CREATE VIEW, SHOW VIEW, CREATE ROUTINE, ALTER ROUTINE, EVENT, TRIGGER ON cost.* TO
 'cost'@'%';
GRANT SELECT, INSERT, UPDATE, DELETE, CREATE, DROP, REFERENCES, INDEX, ALTER, CREATE TEMPORARY TABLES,
 LOCK TABLES, EXECUTE, CREATE VIEW, SHOW VIEW, CREATE ROUTINE, ALTER ROUTINE, EVENT, TRIGGER ON
 group_component.* TO 'group_component'@'%';
GRANT SELECT, INSERT, UPDATE, DELETE, CREATE, DROP, REFERENCES, INDEX, ALTER, CREATE TEMPORARY TABLES,
 LOCK TABLES, EXECUTE, CREATE VIEW, SHOW VIEW, CREATE ROUTINE, ALTER ROUTINE, EVENT, TRIGGER ON market.*
 TO 'market'@'%';
GRANT SELECT, INSERT, UPDATE, DELETE, CREATE, DROP, REFERENCES, INDEX, ALTER, CREATE TEMPORARY TABLES,
 LOCK TABLES, EXECUTE, CREATE VIEW, SHOW VIEW, CREATE ROUTINE, ALTER ROUTINE, EVENT, TRIGGER ON plan.* TO
 'plan'@'%';
GRANT SELECT, INSERT, UPDATE, DELETE, CREATE, DROP, REFERENCES, INDEX, ALTER, CREATE TEMPORARY TABLES,
 LOCK TABLES, EXECUTE, CREATE VIEW, SHOW VIEW, CREATE ROUTINE, ALTER ROUTINE, EVENT, TRIGGER ON
 repository.* TO 'repository'@'%';
GRANT SELECT, INSERT, UPDATE, DELETE, CREATE, DROP, REFERENCES, INDEX, ALTER, CREATE TEMPORARY TABLES,
 LOCK TABLES, EXECUTE, CREATE VIEW, SHOW VIEW, CREATE ROUTINE, ALTER ROUTINE, EVENT, TRIGGER ON suspend.*
 TO 'suspend'@'%';
GRANT SELECT, INSERT, UPDATE, DELETE, CREATE, DROP, REFERENCES, INDEX, ALTER, CREATE TEMPORARY TABLES,
 LOCK TABLES, EXECUTE, CREATE VIEW, SHOW VIEW, CREATE ROUTINE, ALTER ROUTINE, EVENT, TRIGGER ON
 topology_processor.* TO 'topology_processor'@'%';
GRANT SELECT, INSERT, UPDATE, DELETE, CREATE, DROP, REFERENCES, INDEX, ALTER, CREATE TEMPORARY TABLES,
 LOCK TABLES, EXECUTE, CREATE VIEW, SHOW VIEW, CREATE ROUTINE, ALTER ROUTINE, EVENT, TRIGGER ON vmtdb.* TO
 'history'@'%';
GRANT SELECT, INSERT, UPDATE, DELETE, CREATE, DROP, REFERENCES, INDEX, ALTER, CREATE TEMPORARY TABLES,
 LOCK TABLES, EXECUTE, CREATE VIEW, SHOW VIEW, CREATE ROUTINE, ALTER ROUTINE, EVENT, TRIGGER ON
 actions_history.* TO 'actions_history'@'%';
GRANT SELECT, INSERT, UPDATE, DELETE, CREATE, DROP, REFERENCES, INDEX, ALTER, CREATE TEMPORARY TABLES,
 LOCK TABLES, EXECUTE, CREATE VIEW, SHOW VIEW, CREATE ROUTINE, ALTER ROUTINE, EVENT, TRIGGER ON oauth2.*
 TO 'oauth2'@'%';
GRANT SELECT, INSERT, UPDATE, DELETE, CREATE, DROP, REFERENCES, INDEX, ALTER, CREATE TEMPORARY TABLES,
 LOCK TABLES, EXECUTE, CREATE VIEW, SHOW VIEW, CREATE ROUTINE, ALTER ROUTINE, EVENT, TRIGGER ON
 sustainability.* TO 'sustainability'@'%';
GRANT PROCESS on *.* TO 'history'@'%';
GRANT PROCESS on *.* TO 'cost'@'%';
GRANT PROCESS on *.* TO 'group_component'@'%';
flush privileges;
```

Configuring Workload Optimization Manager

Modify the CR file to include the external database properties. Substitute your values for `dbRootUsername` and `dbRootPassword`.

AWS RDS configuration example:

```
  global:
    externalDBName: yourDB.yourURL.com
```

```
  properties:
    global:
      enableSecureDBConnection: true
    dbRootUsername: <yourvalue>
    dbRootPassword: <yourvalue>

    action-orchestrator:
      actionDbUsername: action
      actionDbPassword: vmturbo
    ...
```

Azure configuration example:

```
global:
  externalDBName: yourDB.yourURL.com
```

```
properties:
  global:
    enableSecureDBConnection: true
  dbRootUsername: turboadmin@yourDB
  dbRootPassword: <yourvalue>

  action-orchestrator:
    actionDbUsername: action@yourDB
    actionDbPassword: vmturbo
  ...
```

(Optional) Increasing available disk space

A standard installation of Workload Optimization Manager on a VM image includes a MariaDB database server for historical data. If you enable embedded reports, the platform also uses TimescaleDB Postgres database to manage the reports data. For various reasons, you might find that the default storage capacity for your database services is not sufficient. In that case, you need to increase the available storage capacity.

A common reason to increase this capacity is to accommodate estimated needs for embedded reports. The storage requirements for embedded reports can change over time as your environment changes, or as you increase the number of targets you configure your your Workload Optimization Manager installation. For information about estimating embedded reports requirements, see Embedded reports storage requirement estimates *(on page 1763)*.

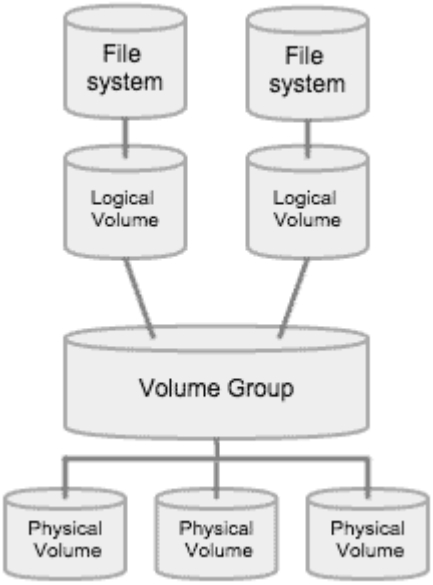
A summary of the steps you will perform is:

- Add a new disk to the VM
- Rescan the scsi devices
- Create a new LVM partition
- Create a physical volume (pv)
- Add the pv to the existing volume group (vg)
- Extend the logical volume (lv)
- Extend the file system to use the new lv
- To increase storage for embedded reports, increase the XFS quota

  To increase space for MariaDB, you do not need to perform this step.

Logical volume management for Workload Optimization Manager storage

The platform uses Logical Volume Management (LVM) to manage the VM disks. To increase database storage, you should add a new disk to the VM, and then use it to extend the LVM logical volume, `/dev/turbo/var_lib_mysql`. This logical volume serves both the historical database and the embedded reports database.



Increasing storage

To increase the storage space available to your databases:

1. Add a new disk to the VM.

   Use the steps for your VM datacenter to add a new disk to the VM. Workload Optimization Manager installs as a VMware or a Hyper-V VM. Refer to the documentation for your hypervisor for the steps to add a new disk.

2. Open an SSH terminal session to your Workload Optimization Manager instance.

   Log in with the System Administrator that you set up when you installed Workload Optimization Manager:

   - Username: `turbo`
   - Password: `[your_private_password]`

3. Rescan the scsi devices.

   To make sure the new disk is available, rescan the scsi devices and then list your block devices.

   ```
   echo "- - -" > /sys/class/scsi_disk//0\:0\:0\:0/device/rescan
   ```

   To check for the new disk, run the command:

   ```
   lsblk
   ```

   The new disk should appear with a name similar to `/dev/sdc`. If you don't see the new disk, try this alternative to force a rescan:

   - Check the number of scsi host devices that are on your VM.

     ```
     ls /sys/class/scsi_host
     ```

     You should see a list of devices, such as `host0, host1, host2... hostn`

   - Scan each device.

     For each device run the command (where *<hostn>* is a numbered host device such as `host0` or `host1`):

     ```
     echo "- - -" > /sys/class/scsi_host/host0/scan
     ```

   - List the block devices.

     Run `lsblk` again to list the block devices.

4. Create a new LVM partition.

   Assuming the new disk is named `/dev/sdc1`, run the command:

   ```
   cfdisk /dev/sdc1
   ```

   Then run the operations:

   - `new`
   - `primary`
   - `confirm size`
   - `change type to 8E`
   - `write`
   - `quit`

5. Create the Physical Volume (pv).

   Assuming the new disk is named `/dev/sdc1`, run the command:

   ```
   pvcreate /dev/sdc1
   ```

6. Add the new pv to the existing Volume Group.

   Assuming the new disk is named `/dev/sdc1`, run the command:

   ```
   vgextend /dev/ibmturbo /dev/sdc1
   ```

7. Extend the Logical Volume (lv) to use the free space in the new pv.

   First list the physical extents (PE) that are available. Run the command:

   ```
   vgdisplay
   ```

   The output is similar to the following example:

   ```
   Free  PE / Size       128000 / 500.00 GiB
   ```

   In this example, `128000` is the amount to extend the lv. For this example, run the command:

   ```
   lvextend -l +128000 /dev/ibmturbo/var_lib_mysql
   ```

8. Extend the XFS file system to use all the current lv space.

   Before you extend the XFS, view the free disk space and record the number. To verify that you have increased the available space, you will compare this value to the free space after you have extended XFS. Run the command:

   ```
   df -h
   ```

   Then extend the XFS capacity:

   ```
   xfs_growfs /dev/ibmturbo/var_lib_mysql
   ```

   Then list the updated free disk space and compare it to your original number:

   ```
   df -h
   ```

9. If you are increasing capacity for embedded reports, extend the XFS quota for the TimescaleDB.

   To increase space for MariaDB, you do not need to perform this step.

To increase capacity for the Timescale DB, you need to increase the quota for that process by the amount you want. The quota name is `Postgresql`.

For example, assume you added a 400 GB volume, and the current `Postgresql` quota is 400 GB. In that case, you could increase the quota to 800 GB. Following this example, run the command:

```
xfs_quota -x -c 'limit -p bhard=800g Postgresql' /var/lib/dbs
```

To see the current quotas set for `/var/lib/dbs`, run the command:

```
xfs_quota -xc 'report -pbih' /var/lib/dbs
```

(Optional) Enforcing secure access by using LDAP

If your company policy requires secure access, you can use a certificate with your LDAP service to set up secure access for your users. For example, you can configure Active Directory (AD) accounts to manage *External Authentication* for users or user groups. The user interface to enable AD includes a **Secure** option, which enforces certificate-based security. For more information, see Managing User Accounts *(on page 1116)*.

If your LDAP service uses a Certificate Authority (CA), then the certificate signed by that CA should support this feature as it is. Simply turn on the **Secure** option when you are setting up your AD connection.

If your LDAP service uses a self-signed certificate, then you must install that certificate on the Workload Optimization Manager authorization pod. The steps you will perform include:

- Get the certificate from your LDAP server
- Import the certificate to the platform's TrustStore
- Add the certificate to the Workload Optimization Manager platform's authorization pod
- Enable the TrustStore in the Workload Optimization Manager platform's Operator chart

Installing a self-signed certificate

1. Open an SSH terminal session to your Workload Optimization Manager instance.

   Log in with the System Administrator that you set up when you installed Workload Optimization Manager:

   - Username: `turbo`
   - Password: `[your_private_password]`
2. Download your LDAP Server certificate to the Workload Optimization Manager instance.

   Acquire a certificate from your LDAP administrator, and download it to the Workload Optimization Manager platform. For example, you can download it to the file `/tmp/ldapserver.crt`.
3. Import the `.crt` file to the Workload Optimization Manager TrustStore by using the `keytool` utility.

   NOTE:
   This step modifies the `cacerts` file on the Workload Optimization Manager platform.

   a. If an alias for an LDAP certificate already exists, delete that certificate. For example, assuming the alias `ldapcert1`, run the following command:

   ```
   keytool -delete -alias ldapcert1 -keystore cacerts -storepass changeit
   ```

   b. Import your new certificate to the `TrustStore`.

   ```
   keytool -import -alias ldapcert1 -file /tmp/ldapserver.crt -keystore cacerts \
       -deststoretype jks -storepass changeit -noprompt
   ```

4. Create an `auth` secret from the `cacerts` file.

```
base64 cacerts > auth-secrets.yaml
```

5. Open the secrets file for editing.

```
vi auth-secrets.yaml
```

6. Edit the file to make it a valid yaml file.
   a. Indent every line of the certificate by four spaces.

      When you created the file, you concatenated the contents of the certificate. The first step is to indent the certificate by four spaces. For example, in a `vi` editor, run the following command:

      ```
      :%s/^/    /g
      ```

   b. Add data fields to the secrets file.

      Add the following text to the upper section of the file:

      ```
      apiVersion: v1
      kind: Secret
      metadata:
        name: auth-secret
      data:
        cacerts: |
      ```

   c. Save your changes.

   The output is similar to the following example:

```
apiVersion: v1
kind: Secret
metadata:
  name: auth-secret
data:
  cacerts: |
    /u3+7QAAAAIAAAABAAAAAgAFY2VydDEAAAF5H2lEigAFWC41MDkAAAYQMIIGDDCCBPSgAwIBAgIT
    HAAAARHIFJdLbG90sAAAAAABETANBgkqhkiG9w0BAQUFADBcMRMwEQYKCZImiZPyLGQBGRYDY29t
    MRcwFQYKCZImiZPyLGQBGRYHdm10dXJibzEUMBIGCgmSJomT8ixkARkWBGNvcnAxFjAUBgNVBAMT
    DWNvcnAtREVMTDEtQ0EwHhcNMjEwNDA4MDM0OTEyWhcNMjIwNDA4MDM0OTEyWjAhMR8wHQYDVQQD
    ExZkZWxsMS5jb3JwLnZtdHVyYm8uY29tMIIBIjANBgkqhkiG9w0BAQEFAAOCAQ8AMIIBCgKCAQEA
    sCXuh2MTrFERyU/aKgdbgyjLezNuwF6nmZveZUhDaJDpfLHJlzhwfyYRTGfSSusVo4polJS4WqPZ
    T3Zk8f2IaX04RpfpQErq5N3uY/BxFkATWLMDiquSd0Di798k2diYXAxXvzMmfmIkBBYJta9oztum
    uXyh/42dXOGznQ5fFuxosgAksZ6CnXGDKrTBlb0bHpST1z1Pdg+fJ+f9Tq7IffOYdVbuedFTwsik
    Z0JgDCIRrmmsOJphiHdBqJ6ZLdbSeEzBIbboiQs81pAELw7V0ZZUfKV6y8+zMTACGwpVPJSFv7LX
    RlW1TWcqhXVAOmroe2WcU8KJE6XZTBxp7z7dzwIDAQABo4IDADCCAvwwLwYJKwYBBAGCNxQCBCIe
    IABEAG8AbQBhAGkAbgBDAG8AbgB0AHIAbwBsAGwAZQByMB0GA1UdJQQWMBQGCCsGAQUFBwMCBggr
    BgEFBQcDATAOBgNVHQ8BAf8EBAMCBaAweAYJKoZIhvcNAQkPBGswaTAOBggqhkiG9w0DAgICAIAw
    DgYIKoZIhvcNAwQCAgCAMAsGCWCGSAFlAwQBKjALBglghkgBZQMEAS0wCwYJYIZIAWUDBAECMAsG
    CWCGSAFlAwQBBTAHBgUrDgMCBzAKBggqhkiG9w0DBzBCBgNVHREEOzA5oB8GCSsGAQQBgjcZAaAS
    BBDswjlHut/nQZ0uK2aUglGbghZkZWxsMS5jb3JwLnZtdHVyYm8uY29tMB0GA1UdDgQWBBR6M7Hb
    BiirpjIXQ3PXXScB8LkmRDAfBgNVHSMEGDAWgBRjs9l3el7SuKUDMlrHHRhBkENgaDCB0QYDVR0f
    BIHJMIHGMIHDoIHAoIG9hoG6bGRhcDovLy9DTj1jb3JwLURFTEwxLUNBLENOPWRlbGwxLENOPUNE
    UCxDTj1QdWJsaWMlMjBLZXklMjBTZXJ2aWNlcyxDTj1TZXJ2aWNlcyxDTj1Db25maWd1cmF0aW9u
    LERDPWNvcnAsREM9dm10dXJibyxEQz1jb20/Y2VydGlmaWNhdGVSZXZvY2F0aW9uTGlzdD9iYXNl
    P29iamVjdENsYXNzPWNSTERpc3RyaWJ1dGlvblBvaW50MIHHBggrBgEFBQcBAQSBujCBtzCBtAYI
    KwYBBQUHMAKGgadsZGFwOi8vL0NOPWNvcnAtREVMTDEtQ0EsQ049QUlBLENOPVB1YmxpYyUyMEtl
    eSUyMFNlcnZpY2VzLENOPVNlcnZpY2VzLENOPUNvbmZpZ3VyYXRpb24sREM9Y29ycCxEQz12bXR1
    cmJvLERDPWNvbT9jQUNlcnRpZmljYXRlP2Jhc2U/b2JqZWN0Q2xhc3M9Y2VydGlmaWNhdGlvbkF1
    dGhvcml0eTANBgkqhkiG9w0BAQUFAAOCAQEADP6OYLONkZ2j6gaBdfdoIJtvn1g1qXTsRrtFuUcF
    C9mUxL0G5Tudr0VlyEnLH2wtj10CGsIi54+aPGYiElXijThEe1WTHaO2hklRLdNrM8KxUp3tUNb/
    cP4d+EYt297wVWgxpl9MStiND8+7M2+65daoEu5IOLtq4lC7YlCSXay19N5HdiGBHV5L07PTZ261
    qDzShSb0ZWtG7++5VkqveVEIfs3hUYdaItz0Zu6sym90aUcvn5wohV1GPPqGDvVCg5Kf50hsZfmy
```

```
    ltNlaqiiqLMnYVMa93CkpFFjoP9gmGFJky0yTfh6G8HuqbI7guddDsUqMQTT3uv3EBwSYeImOya7
    Zye5C4NnsAfnx8kOwXdsVERC
```

7. Apply this secrets file to the platform environment.

```
kubectl apply -f auth-secrets.yaml
```

8. Update the platform's Operator Chart to use the `cacerts` certificate that you created in the secrets file.
   a. Open the chart file for editing.

```
/opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
```

   b. Add the certification secret as an authorization spec for the component options.

      In the chart file, find the `spec:` section. Within that section, find the `auth:` subsection.

      This should be the second subsection in `spec:`, after `global:`. If there is no `auth:` subsection, you can add it to `spec:`.

   c. Add the certification secret to the file:

      Add the secret's path to a `javaComponentOptions:` statement within the `auth:` subsection. Add the path as a `-D` option. The `auth:` subsection should be similar to the following example, with `auth` indented by two spaces and `javaComponentOptions` indented by four spaces:

```
# Pass in the JAVA_OPTS to the auth POD to set up additional options such as
# a trustStore for AD Certificate(s) for LDAPS (Secure LDAP)
auth:
 javaComponentOptions: "-Djavax.net.ssl.trustStore=/home/turbonomic/data/helper_dir/cacerts"
```

   d. Apply the Operator Chart changes to the Workload Optimization Manager platform.

```
kubectl apply -f \
/opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
```

      The authorization component restarts so that it can use the new setting.

(Optional) Adding a certificate for securing the Workload Optimization Manager UI

If your company policy requires SSL connections that use a trusted certificate, Workload Optimization Manager installs a trusted certificate from a known certificate authority.

Requesting a certificate

The first step is to acquire a certificate. The following steps describe how to generate a certificate request.

1. Open a Secure Shell terminal session.

   Open an SSH terminal session on your Workload Optimization Manager instance. Log in as `turbo`, and use the password that you created for the administration account during installation.

   For more information, see Set up the Workload Optimization Manager System Administrator account *(on page 52)*.

2. Change to the directory where you want to store the private key file.

   If your shell session is on your Workload Optimization Manager instance, use the `/opt/turbonomic` directory.

```
cd /opt/turbonomic
```

3. Create and save the private key file.

   For this example, the private key file is named `myPrivate.key`.

   ```
   openssl genrsa -out myPrivate.key 2048
   ```

   You need this file later. If you are in a session on your Workload Optimization Manager instance, copy the file to your local machine.

4. Create a file to contain the information that generates the signed certificate request (CSR).

   ```
   vi certsignreq.cfg
   ```

5. Add the request data to the `certsignreq.cfg` file.

   In the file, insert the following code.

   ```
   [req]
    ts = 2048
    prompt = no
    default_md = sha256
    req_extensions = req_ext
    distinguished_name = dn
   ```

   ```
    [dn]
    C=<2-letter_country_code>
    L=<city>
    O=<company>
    OU=<organizational_unit_name>
    CN=<FQDN>
    emailAddress=<email_address>
   ```

   ```
    [req_ext]
    subjectAltName = @alt_names
   ```

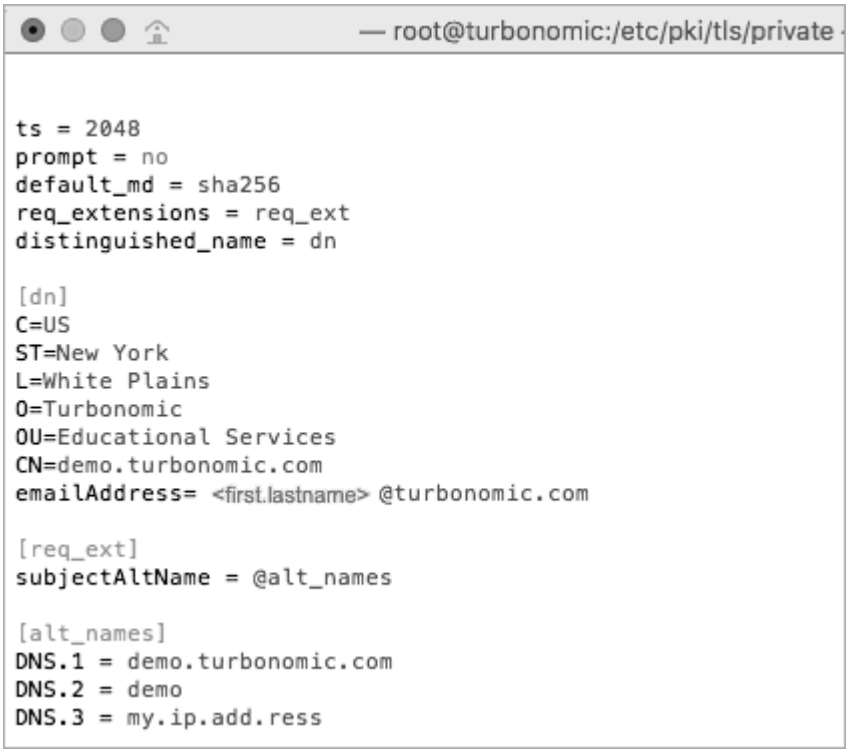
   ```
    [alt_names]
    DNS.1 = <FQDN>
    DNS.2 = <server_short_name>
    DNS.3 = <server_IP_address>
   ```

   To configure the file properly, review the following guidelines:

   - For any fields marked by angle brackets, provide the indicated value. For example, your country, city, company, and so on.
   - To avoid errors, do not include special characters, such as commas and quotes.
   - For the CN field, specify the fully qualified domain name of the Workload Optimization Manager instance.

     Alternative names are other ways to access the Workload Optimization Manager instance. In the `[alt_names]` section, the value for the `DNS.1` field is required. For `DNS.1`, specify the fully qualified domain name of the Workload Optimization Manager instance. Values for the `DNS.2` and `DNS.3` are optional. You can add more `DNS.x` fields if needed.

   For example:



6. Write and quit the file.

   Press `Esc`, type `:wq!`, and press `Enter`.

7. Generate the certificate request file.

   In this example, the file is named `myRequest.csr`.

   ```
   openssl req -new -sha256 -nodes -out myRequest.csr -key \
     myPrivate.key -config certsignreq.cfg
   ```

8. Send the generated request file to your certificate authority.

   If you generated the file on your Workload Optimization Manager instance, transfer the file to your local machine. The path to the certificate request file on your remote machine is `/opt/turbonomic/myRequest.csr`.

   Your certificate authority uses this file to create the certificate for you. If your certificate authority gives you an encoding choice between DER and Base 64, choose **Base 64**.

   > **NOTE:** Ensure that the generated certificate is not password protected, as password protected certificates are not supported in Workload Optimization Manager.

9. When you receive the certificate, save it to disk.

   If you did not receive the certificate encoded in Base 64, you must convert it from DER to Base 64.

   In this example, the certificate is named `MyCertificate.crt`.

   ```
   openssl x509 -inform der -in MyCertificate.der -out MyCertificate.crt
   ```

Installing the signed certificate in Workload Optimization Manager

After you obtain the signed certificate, you can install it on your Workload Optimization Manager instance. You use the private key and certificate files that you obtained when requesting the signed certificate:

- `myPrivate.key`
- `MyCertificate.crt`

NOTE: Ensure that the generated certificate is not password protected, as password protected certificates are not supported in Workload Optimization Manager.

To install the signed certificate:

1. Open an SSH terminal session on your Workload Optimization Manager instance.
2. Add the key and certificate data to your Workload Optimization Manager charts.yaml file.

   ```
   /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

   Find the section for `global` parameters. Under the `global` parameters, create the `ingress:secrets` section, and then create entries for `certificate`, `key`, and `name`.

   Your global parameters should be similar to the following example:

   ```
   global:
     ingress:
       secrets:
         - certificate: |
             -----BEGIN CERTIFICATE-----
             SAMPLE PUBLIC KEY
             -----END CERTIFICATE-----
           key: |
             -----BEGIN RSA PRIVATE KEY-----
             SAMPLE PRIVATE KEY
             -----END RSA PRIVATE KEY-----
           name: nginx-ingressgateway-certs
   ```

   For the fields you added:

   - `certificate:` holds the content of your `MyCertificate.crt` file. Open that file to copy its contents and paste them here.
   - `key:` holds the content of your `myPrivate.key` file. Open that file to copy its contents and paste them here.
   - `name:` Is required and the name must be `nginx-ingressgateway-certs`.
3. Apply the changes that you made to the CR file.

   ```
   kubectl apply -f \
     kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

4. Restart the `nginx` pod.

   To require a certificate for HTTPS access, you must restart the `nginx` pod.

   a. Get the full name of the pod.

      ```
      kubectl get pods -n turbonomic
      ```

      In the output, look for the entry for `nginx`. You should find an entry similar to:

      ```
      nginx-5b775f498-sm2mm                          1/1     Running   0
      ```

   b. Restart the pod.

      ```
      kubectl delete pod nginx-<UID>
      ```

      Where *<UID>* is the generated ID for the pod instance.

   After the nginx pod restarts Workload Optimization Manager requires a certificate for HTTPS access.

(Optional) Adding additional CA certificates for probes

If your target requires SSL connections that use a trusted certificate, Workload Optimization Manager installs a trusted certificate on the associated probe component.

The Workload Optimization Manager platform includes various probe components that it uses to connect to targets and discover their data. This procedure assumes setup for one component, the Dynatrace probe. You can use the same steps for other probes, providing a different Kubernetes Secret Name for each.

To install a certificate on a probe component, familiarize yourself with the Kubernetes secret name for the probe. This table lists the probes that you can configure and their secret names.

Probe	Kubernetes Secret Name

`mediation-aws`
`mediation-awscloudbilling`
`mediation-awscost`
	`aws`

`mediation-awsinfra`	`awsinfra`
`mediation-awsorganization`	`awsorganization`

`mediation-azure`
`mediation-azurecost`
`mediation-azurepricing`
`mediation-azuresp`
`mediation-azurevolumes`
	`azure`

`mediation-azurebilling`	`azurebilling`
`mediation-azureea`	`azureea`
`mediation-azureinfra`	`azureinfra`
`mediation-datadog`	`datadog`
`mediation-dynatrace`	`dynatrace`
`mediation-gcpbilling` `mediation-gcpcost`	`gcp`
`mediation-gcpinfra`	`gcpinfra`
`mediation-gcpproject`	`gcpproject`
`mediation-gcpsa`	`gcpsa`
`mediation-newrelic`	`newrelic`
`metrics-adapter-aws-cloudwatch`	`aws`
`metrics-adapter-aws-performanceinsights`	`aws`
`metrics-adapter-gcp-monitoring`	`gcpproject`

Installing the signed certificate on the probe component

This procedure assumes that you already have a valid `.crt` file. If you do not have the certificate file, ask your networking team to generate one for you.

After you obtain the signed certificate (`MyCertificate.crt`), install it on your probe instance.

1. Copy the certificate from your local machine to the Workload Optimization Manager instance.

   Use SmartCloud Provisioning to copy the `MyCertificate.crt` from your local machine to the `/tmp` directory on the instance.

2. Open an SSH terminal session on your Workload Optimization Manager instance, that uses the `turbo` user account.

3. Obtain the truststore from the probe component.

   Obtain the ID for the pod that runs the probe.

   ```
   kubectl get pods
   ```

   This lists the pods running in the Workload Optimization Manager platform, including their IDs. Record the ID of the pod that you want to configure.

   To get the CA truststore, run the following command, where <*Probe-Pod-Id*> is the ID you recorded:

   ```
   kubectl cp <Probe-Pod-Id>:etc/pki/ca-trust/extracted/java/cacerts cacerts
   ```

4. Import the certificate into the pod's keystore.

   As part of this step, you ensure that the certificate is in Base64 format and you create a yaml file that uses the K8s Secret Name for the probe. While still in the bash session, run the following commands:

   - ```
     chmod 775 cacerts
     ```

   - ```
     keytool -import -alias probe_certificate -file \
       <MyCertificate.crt> -keystore cacerts -deststoretype jks \
       -storepass changeit -no-prompt
     ```

     Where <*MyCertificate.crt*> is the name of the certificate that you acquired.

   - ```
     base64 cacerts > <Secret_Name>-secrets.yaml
     ```

     Where <*Secret_Name*>`-secrets.yaml` is the yaml file that you create that uses the K8s Secret Name for the probe.

     For example, assume that you are enabling SSL for the Dynatrace probe. In that case, the secret name is `dynatrace`, and you would create the yaml file `dynatrace-secrets.yaml`.

5. Update the <*Secret_Name*>`-secrets.yaml` file you created.

   a. While still in the bash session on the Workload Optimization Manager server, open the yaml file created in the previous step in a vi editor session.

      ```
      vi <Secret_Name>-secrets.yaml
      ```

   b. Align the base64 data to the yaml format.

      Type : to enter the command mode. For the command, type the following, where the white space token is four space characters.

      ```
      :%s/^/    /g
      ```

      Press RETURN, then save and exit the vi editor.

   c. Add the following content to the file preceding your Base64 data.

      ```
      apiVersion: v1
      kind: Secret
      metadata:
        name: <Secret_Name>
      data:
        cacerts: |
          xxxxxxxxxxxxxxxxxxxxxxxxx
      ```
      ```

```
    xxxxxxxxxxxxxxxxxxxxxxxxx
```

Your Base64 data should be in the `cacerts` section in place of the `xxx` characters in the preceding example .

6. Apply the yaml file to the Workload Optimization Manager platform.

   Run the following command, where `<Secret_Name>-secrets.yaml` is yaml you created.

   ```
   kubectl apply -f <Secret_Name>-secrets.yaml
   ```

7. For each probe that you configure with an SSL certificate, add an entry in the chart_v1alpha1_cl_cr.yaml file.

   a. With a shell session running on the Workload Optimization Manager platform, open the following file in a text editor.

   ```
   vi /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

   b. Search the file for the entry for the probe that you are configuring. Use the probe names listed in the preceding table. For example, if you are configuring the Dynatrace probe, find the entry for `mediation-dynatrace`. If the entry does not exist in the file, you can add it to the `spec:` section at the same level as `global:`, with the probe entry that is indented by two spaces and `javaComponentOptions` indented by four spaces.

   c. Underneath the probe entry, add the following entry for `javaComponentOptions`.

   ```
   javaComponentOptions: -Djavax.net.ssl.trustStore=/etc/targets/cacerts
   ```

   For example, if you are configuring the Dynatrace probe, the entry should be:

   ```
   mediation-dynatrace:
       javaComponentOptions: -Djavax.net.ssl.trustStore=/etc/targets/cacerts
       resources:
         limits:
           memory: 2Gi
   ```

   d. Save and exit the chart_v1alpha1_cl_cr.yaml file.

   e. Apply the changed file to your Workload Optimization Manager platform.

   ```
   kubectl apply -f /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

(Optional)Modifying the certificates for Cluster Manager

For installations behind a firewall, to upload diagnostics from the `clustermgr` component, you must modify its certificates.

These steps to modify the certificates on `clustermgr` assume that you have already generated the certificates that you want to add to the cluster manager.

1. Open an SSH terminal session on your Workload Optimization Manager instance.

   Log in with the System Administrator that you set up when you installed Workload Optimization Manager:

   - Username: `turbo`
   - Username: `[your_private_password]`

2. Get the full name of the clustermgr pod.

   ```
   kubectl get pods -n turbonomic | grep clustermgr
   ```

   The output is similar to the following example:

   ```
   clustermgr-5f487f58f-tf84b   0/1     Running   52         2d4h
   ```

   In this example, `clustermgr-5f487f58f-tf84b` is the full name of the pod, and `5f487f58f-tf84b` is the POD_ID.

3. Save a copy of the pod's current ca-bundle.crt file to /tmp.

   Run the following command, where <*POD_ID*> is the ID you get from the pod's full name.

   ```
   kubectl cp \
   clustermgr-<POD_ID>:etc/pki/ca-trust/extracted/pem/tls-ca-bundle.pem \
   /tmp/ca-bundle.crt
   ```

4. Add your certificates to the bundle.

   Repeat this command for each certificate, where <*MY_CERT*> is your certificate file.

   ```
   cat <MY_CERT> >> /tmp/ca-bundle.crt
   ```

5. Create a Kubernetes secret for the modified certificates.

   ```
   kubectl create secret generic clustermgr-secret --from-file=/tmp/ca-bundle.crt
   ```

6. Open the cr.yaml file for editing.

   For example:

   ```
   vi /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

7. Modify the cr.yaml file to use this secret.

   Add the following to the file:

   ```
   clustermgr:
     env:
       - name: component_type
         value: clustermgr
       - name: instance_id
         valueFrom:
           fieldRef:
             fieldPath: metadata.name
       - name: instance_ip
         valueFrom:
           fieldRef:
             fieldPath: status.podIP
       - name: serverHttpPort
         value: "8080"
       - name: kafkaServers
         value: kafka:9092
       - name: kafkaNamespace
         valueFrom:
           fieldRef:
             apiVersion: v1
             fieldPath: metadata.namespace
       - name: CURL_CA_BUNDLE
         value: /home/turbonomic/data/ca-bundle.crt
   ```

8. Save your changes and apply the cr.yaml file.

   ```
   kubectl apply -f \
   /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

If you watch the log with `grep ^clustermgr`, then you will see the appropriate curl command run whenever you send diagnostics.

You can also check the .crt file in the cluster manager pod by using the following commands, where <*POD_ID*> is the ID you got from the pod's full name:

```
kubectl exec -it clustermgr-<POD_ID> bash
vi /home/turbonomic/data/ca-bundle.crt
```

(Optional) Enabling the reporting Add-on

The Workload Optimization Manager platform includes a reporting add-on that you can choose to enable when you install the platform. Use the reporting add-on to understand trends in application resource management and to share insights with stakeholders through reports and dashboards.

The reporting add-on runs as its own component within the Workload Optimization Manager platform. This architecture enhances performance and reduces storage requirements. It stores a history of your managed environment and then presents selective snapshots of this history through a set of standard dashboards and reports. You can create your own dashboards and reports to focus on other areas of concern.

For more information, see Integration - Embedded reporting *(on page 1751)*.

(Optional) Enabling the Data Exporter

To support Data Export, Workload Optimization Manager provides an extractor component that can stream data to a standard format. You can load that data into search and analytics services such as Elasticsearch.

To enable the Data Exporter, you must:

- Enable the extractor component.

  The extractor is a component that runs as part your Workload Optimization Manager installation. The extractor is not enabled by default.
- Deploy a connector that delivers the extractor's stream to your data service.

  The extractor publishes Workload Optimization Manager data as Kafka topics. The connector enables your data service to consume the data topic. This document includes a deployment file for a sample Elasticsearch connector.

Enabling the Extractor component

The first step to enabling the Data Exporter is to enable the extractor component.

> **NOTE:**
> If you have enabled embedded reporting, then the extractor component is already enabled (set to `true`).
>
> It is possible to enable the Data Exporter without enabling embedded reports, just as it is possible to enable embedded reports without enabling the Data Exporter.

1. Open an SSH terminal session to your Workload Optimization Manager instance.

   Log in with the System Administrator that you set up when you installed Workload Optimization Manager:
   - Username: `turbo`
   - Password: `[your_private_password]`
2. Open the cr.yaml file to enable the extractor component.

   ```
   vi /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```
3. Edit the entry for the extractor component.

   Search for the `extractor` entry in the cr.yaml file.

   ```
   extractor:
       enabled: false
   ```

   Change the entry to `true`.
4. Edit the entry for the extractor properties.

   Search for the `properties: extractor` entry in the cr.yaml file.

```
properties:
    extractor:
      enableDataExtraction: false
```

Change the entry to `true`.

5. Save and apply your changes to the platform.

```
kubectl apply -f \
/opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
```

6. Verify that the extractor component is running.

   Give the platform enough time to restart the components. Then run the command:

```
kubectl get pods -n turbonomic
```

   You should see output similar to the following:

```
NAME                                          READY   STATUS    RESTARTS
...
extractor-5f41dd61c4-4d6lq                    1/1     Running   0
...
```

   Look for an entry for the `extractor` component. If the entry is present, then the extractor component is installed and running.

Deploying a connector

The extractor publishes Workload Optimization Manager data as Kafka topics. To load this data into a search and analysis service, you must deploy a connector to that service. For example, you must deploy an Easticsearch connector to load the data into Elasticsearch. You deploy the connector in the same Kubernetes node that runs the Workload Optimization Manager platform. Create a Kubernetes Deployment that declares the pods you need for the connector.

To deploy the connector, create a deployment yaml file on the same host that is running the extractor component, and run the command:

```
kubectl create -f <MyConnectorDeployment.yaml>
```

Where <*MyConnectorDeployment.yaml*> is the name of the deployment file.

Assume that the name of the deployed pod is `es-kafka-connect`. To verify that the connector is running, run the command:

```
kubectl get pods -n turbonomic
```

The output is similar to the following example:

```
NAME                                          READY   STATUS    RESTARTS
...
es-kafka-connect-5f41dd61c4-4d6lq             1/1     Running   0
...
```

After you deploy the connector, wait for a cycle of Workload Optimization Manager analysis (approximately ten minutes). Then you will see the entities and actions from your Workload Optimization Manager environment, loaded as JSON in your data service.

Connector deployment example

The following example is a sample deployment of a connector to Elasticsearch that uses Kibana with Elasticsearch to display data dashboards. In this example, you have the following set up:

- Elasticsearch is deployed to a VM on the network where you are running Workload Optimization Manager. The Elasticsearch host is visible from the Workload Optimization Manager Kubernetes node. You specify this host address in the connector deployment.
- An Elasticsearch index is set up to load the Workload Optimization Manager data. You specify this index in the connector deployment.

The following listing is a deployment that uses a Logstash image to collect the extractor data and pipe it to the Elasticsearch host. The deployment also sets up storage volumes, configures the input from the extractor, and configures output to the Elasticsearch instance. As you go over the listing, pay attention to the following:

- The location of the Elasticsearch host and the login credentials.

```
...
        env:
          - name: ES_HOSTS
            value: "<UrlToMyElasticsearchHost>"
          - name: ES_USER
            value: "<MyElasticsearchUser>"
          - name: ES_PASSWORD
            valueFrom:
              secretKeyRef:
                name: <MyES_KeyName>
                key: <MyES_Key>
...
```

  Logstash will use the following environment variables:

  - ES_HOSTS: to identify where to pipe the exported data.
  - ES_USER: to identify the user account on Elasticsearch.
  - ES_PASSWORD: for the account login. This connector example assumes that you have stored the Elasticsearch password as a Kubernetes Secret.

- The name of the Kafka topic.

```
...
  logstash.conf: |
    input {
      kafka {
        topics => ["turbonomic.exporter"]
...
```

  The Logstash input configuration expects a single topic named `turbonomic.exporter`.

- The Logstash output configuration is to the Elasticsearch server that is identified by the ES_HOSTS environment variable. You specify your own Elasticsearch index in place of *<MyElasticsearchIndex>*.

```
...
    output {
      elasticsearch {
        index => "<MyElasticsearchIndex>"
        hosts => [ "${ES_HOSTS}" ]
      }
    }
...
```

Sample Listing: Elasticsearch Connector

This listing is a sample of a deployment file that can work to create an Elasticsearch connector for the Data Exporter. You need to change some settings, such as username and password. You also might need to specify ports and other settings to make the connector comply with your specific environment.

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: elasticsearch-kafka-connect
  labels:
    app.kubernetes.io/name: elasticsearch-kafka-connect
spec:
  replicas: 1
  selector:
    matchLabels:
      app.kubernetes.io/name: elasticsearch-kafka-connect
  template:
    metadata:
      labels:
        app.kubernetes.io/name: elasticsearch-kafka-connect
    spec:
      containers:
      - name: logstash
        image: docker.elastic.co/logstash/logstash:7.10.1
        ports:
          - containerPort: 25826
        env:
          - name: ES_HOSTS
            value: "<UrlToMyElasticsearchHost>"
          - name: ES_USER
            value: "<MyElasticsearchUser>"
          - name: ES_PASSWORD
            valueFrom:
              secretKeyRef:
                name: <MyES_KeyName>
                key: <MyES_Key>
        resources:
          limits:
            memory: 4Gi
        volumeMounts:
          - name: config-volume
            mountPath: /usr/share/logstash/config
          - name: logstash-pipeline-volume
            mountPath: /usr/share/logstash/pipeline
      volumes:
      - name: config-volume
        configMap:
          name: logstash-configmap
          items:
            - key: logstash.yml
              path: logstash.yml
      - name: logstash-pipeline-volume
        configMap:
          name: logstash-configmap
          items:
            - key: logstash.conf
              path: logstash.conf
---
apiVersion: v1
kind: ConfigMap
metadata:
  name: logstash-configmap
data:
  logstash.yml: |
    http.host: "0.0.0.0"
    path.config: /usr/share/logstash/pipeline
```

```
  logstash.conf: |
    input {
      kafka {
        topics => ["turbonomic.exporter"]
        bootstrap_servers => "kafka:9092"
        client_id => "logstash"
        group_id => "logstash"
        codec => "json"
        type => "json"
        session_timeout_ms => "60000"   # Rebalancing if consumer is found dead
        request_timeout_ms => "70000"   # Resend request after 70 seconds
      }
    }
    filter {
    }
    output {
      elasticsearch {
        index => "<MyElasticsearchIndex>"
        hosts => [ "${ES_HOSTS}" ]
        user => "${ES_USER}"
        password => "${ES_PASSWORD}"
      }
    }
---
apiVersion: v1
kind: Service
metadata:
  labels:
    app: elasticsearch-kafka-connect
  name: elasticsearch-kafka-connect
spec:
  ports:
    - name: "25826"
      port: 25826
      targetPort: 25826
  selector:
    app: elasticsearch-kafka-connect
```

(Optional) Changing the IP address of the platform node

For standard installations of Workload Optimization Manager (installed as a VM image), you might need to change the platform's IP address. For example, if you must move the VM then you might need to assign it a different address. If you must change the IP address of the platform, you can use the supplied scripts.

> **NOTE:**
> Change the IP address of your Workload Optimization Manager installation as seldom as possible. Changing the IP address is a sensitive action that can impact unforeseen dependencies.
>
> Use these steps to change the IP address only for Workload Optimization Manager version 3.0.0 or later. If you must change your IP address and you cannot update to version 3.0.0 or later, contact your support representative.

To change the IP address of the Workload Optimization Manager VM:

1. Get your information ready.

   Identify the current IP address for your platform and the new IP address you use.

   Know the credentials to open a shell session on the VM and run commands.
2. Create a full snapshot of the VM.

   It is important to make a full snapshot of your installation before you try to modify its IP address.
3. Change the VM's IP address.

The Workload Optimization Manager VM includes the `ipsetup` script to complete this task.

a. Open an SSH terminal session to your Workload Optimization Manager VM.

   Use the following credentials:

   - Username: `turbo`
   - Password: Give the password that you assigned to the `turbo` account when you first installed the platform.

b. Once the session is open, run the `ipsetup` script:

```
sudo /opt/local/bin/ipsetup
```

   When the script runs, it requests the following inputs. Proxy servers are not supported with the secure client so they cannot be configured.

   NOTE:
   Provide values for these fields; otherwise, the installation can fail or your VM can be unreachable.

   - **Required:** `Do you want to use DHCP or set a static IP...`
     Choose `static`
   - **Required:** `Please enter the IP Address for this machine`
   - **Required:** `Please enter the network mask for this machine`
   - **Required:** `Please enter the Gateway address for this machine`
   - **Required:** `Enter DNS Server(s) IP Address for this machine`
   - **Required:** `Do you want to configure a proxy server? (y/n)`
     Choose `n`

   Make a note of the IP address that you provide.

c. Propagate your IP change through to the Kubernetes cluster on the VM.

   - For VM images based on Rocky Linux:

```
sudo /opt/local/bin/k8s-ip-change.sh
```

   - For VM images based on Cent OS Linux:

```
sudo /opt/local/bin/kubeNodeIPChange.sh
```

   IMPORTANT:
   - The `k8s-ip-change.sh` script is not designed for CentOS Linux distributions and might cause serious errors if it is run on CentOS VMs. If you are running CentOS Linux, you must run the `kubeNodeIPChange.sh` script, regardless of your Workload Optimization Manager version level.
   - If you are running CentOS Linux with Workload Optimization Manager version 8.12.2 or later, the `kubeNodeIPChange.sh` script is in a versioned `bin-*` folder in the `/opt/local/` directory. The version of the `bin-*` folder corresponds to the last version you ran before upgrading to 8.12.2 or later. For example, if you upgraded from 8.12.0 to 8.13.0, the script is located in the `/opt/local/bin-8.12.0/` folder.

d. Verify that the change is successful.

Log in to the Workload Optimization Manager user interface for the newly located installation, and ensure that it displays correctly. Review the Supply Chain, your groups, and your policies. Also ensure that charts show data correctly.

When you are sure that the change is successful, you can remove the snapshot that you made of the VM in its old location.

(Optional) Redirecting audit and container logs

Workload Optimization Manager deployments include a syslog service that runs in a container. The Workload Optimization Manager containers are configured to centralize their application and audit logs to the included syslog container by default. Audit logs capture a group of user or system events that are used for user analysis of user and system actions. These actions include log in, log out, account management, policy changes, actions, and authorization management events.

Viewing audit logs

All Workload Optimization Manager Server components (except for the Operator) generate logs that are sent to the `rsyslog` pod in the namespace in which you deployed Workload Optimization Manager. To view the `rsyslog` logs, run the following command if Workload Optimization Manager is deployed in the `turbonomic` namespace. If Workload Optimization Manager is deployed in another namespace, replace `turbonomic` with the correct namespace.

```
for i in `kubectl get pods -n turbonomic | grep rsyslog | awk {'print $1'}`; do kubectl -n turbonomic logs
 $i; done
```

The Workload Optimization Manager Operator writes its logs to the `t8c-operator` pod. To view the operator logs, run the following command if Workload Optimization Manager is deployed in the `turbonomic` namespace. If Workload Optimization Manager is deployed in another namespace, replace `turbonomic` with the correct namespace.

```
for i in `kubectl get pods -n turbonomic | grep t8c-operator | awk {'print $1'}`; do kubectl -n turbonomic
 logs $i; done
```

If you do not want to capture the Operator logs, add the following argument to the `operator.yaml` deployment.

```
containers:
- name: t8c-operator
  image: icr.io/cpopen/t8c-operator:<OperatorVersion>
  imagePullPolicy: Always
  args:
  - --zap-level=error
```

Enabling log redirection

You can redirect logs to your centralized logging systems for analysis and tamper resistance. Workload Optimization Manager supports the syslog protocol, which enables it to easily integrate logging with many third-party logging solutions, such as Splunk.

1. Open an SSH terminal session to your Workload Optimization Manager instance.

   Log in with the System Administrator that you set up when you installed Workload Optimization Manager.

   - Username: `turbo`
   - Password: `[your_private_password]`

2. Open the custom resource (CR) file for editing.

   ```
   /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

3. Redirect logs to an external syslog collector.
   - Container pod logs

     To redirect container pod logs, add the `externalSyslog` parameter to the `global` section and specify the syslog URL and port.

     ```
     global:
       externalSyslog: <syslog_URL>:port_number>
     ```

     For example:

     ```
     global:
       externalSyslog: rsyslog.company.com:2514
     ```

     By default, Workload Optimization Manager uses port 2514. You can specify a different port number. If you do not specify a port number, Workload Optimization Manager uses port 2514.
   - Audit logs

     To redirect audit logs, add the `externalAuditLog` parameter to the `global` section and specify the audit log URL and port.

     ```
     global:
       externalAuditLog: <audit_log_URL>:port_number>
     ```

     For example:

     ```
     global:
       externalAuditLog: rsyslog.company.com:3001
     ```

     NOTE:
     If you are using Splunk, specify the DNS of the Splunk URL and ensure that Splunk is configured to listen on the ports that Workload Optimization Manager uses to redirect logs. For more information about configuring Splunk, see the next section in this topic.

4. Apply the changes that you made to the CR file.

   ```
   kubectl apply -f \
     /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

5. Open the ports that you configured for the `rsyslog` logs.

   To check whether the ports are open, run the following command:

   ```
   ss -tnlp
   ```

   If the ports are not open, run the following command to open the ports based on the firewall service that is used in your instance:

   ```
   sudo <firewall_service> allow <port_number>
   ```

   For more information on opening ports based on your firewall service, see the wikiHow documentation.

6. Restart the rsyslog container.

   Run the following command, where *xxxxxxx* is the pod ID. For example, `rsylog-858b7f596f-rd9v5`.

   ```
   kubectl delete po -n <namespace> rsyslog-xxxxxxx
   ```

   Replace *xxxxxxx* with the pod ID.

   For example:

```
kubectl delete po -n my_namespace rsyslog-rsylog-858b7f596f-rd9v5
```

The logs now redirect to your syslog service.

Configuring Splunk

Splunk is a software platform that helps organizations search, monitor, and analyze data from any source.

NOTE:
You can use any third-party logging solutions to gather log information. Splunk is the most commonly used solution to monitor and troubleshoot problems with applications, servers, and networks.

1. Install the Enterprise edition of Splunk in your VM by using the link https://www.splunk.com/en_us/download/splunk-enterprise.html.
2. Log in to Splunk by using the admin username and password you specified during setup.
3. Select **Settings > Add Data** to configure Splunk to listen to the Workload Optimization Manager logs.
4. On the Select Source page, select **TCP/UDP** to add new TCP connections as the source data.
    - In the **Port** field, specify the port that was configured in Workload Optimization Manager. For Audit log redirection, enter the port number that you specified in the `rsyslog` parameter in your Workload Optimization Manager CR file. For Container pod log redirection, you must use port 2514.
    - Optionally specify any additional parameters.
5. Click **Next** to open the **Input Settings** page.
6. On the Input Settings page, specify the following settings:
    - **Source Type** - Select **New** and specify the name for the source data (for example, "Workload Optimization Manager_Log").
    - **App Context** - Select **Search & Reporting**.
    - **Host** - Select **DNS**.
    - **Index** - Select to create a new Index or use the default one.
7. Click **Review** to review the connection, then click **Submit**.

   A new TCP data input is created for the port that you specified. You can search for the data by selecting **Apps > Search & Reporting**.

(Optional) Enabling and disabling probe components

In Workload Optimization Manager, a probe is a platform component that connects to a target. A probe discovers the target's entities, loads the entities into the Workload Optimization Manager supply chain, and runs actions in the target environment. Workload Optimization Manager includes many probe components that you can use to connect Workload Optimization Manager with your environment.

Depending on your installation method, Workload Optimization Manager may enable probes by default. Each probe uses resources in your Workload Optimization Manager installation. If probes are not needed, consider disabling them. If there are disabled probes that you need, enable them to put them into service.

Viewing the current list of available probes

NOTE:
As Workload Optimization Manager evolves, the set of delivered probes change. Also, from one version to the next, the set of probes that are enabled by default can change. When you update to a new version, the update does not change your probe configuration. An update to a newer version does not automatically enable any new probes in your deployment. If you want to take advantage of new probes in an update, then you must enable them manually.

To enable any new probes, you must first know the internal name for the probe. To get a list of probes that are available to your current version, you can view the contents of the `values.yaml` file.

1. Open an SSH terminal session to your Workload Optimization Manager instance.

   Log in with the System Administrator that you set up when you installed Workload Optimization Manager:

   - Username: `turbo`
   - Password: `[your_private_password]`

2. Display the list of available probes.

   ```
   cat /opt/turbonomic/kubernetes/operator/helm-charts/xl/values.yaml
   ```

   The output is similar to the following example:

   ```
   customdata:
     enabled: false
   dynatrace:
     enabled: false
   gcp:
     enabled: false
   hpe3par:
     enabled: false
   ...
   ```

   This list gives the internal names of the probes. If you want to add a new probe to your list of configured probes, you must use the internal name, and set `enabled: true`.

Viewing the current list of configured probes

Your current installation of Workload Optimization Manager has a set of available probes. Some of these will be enabled, and it is likely that some probes are disabled. To view the current configuration of probes that are available, open the `cr.yaml` file for your Workload Optimization Manager installation and review the probe entries:

1. In the same SSH session, open the `cr.yaml` file for editing.

   ```
   vi /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

2. Search for the list of probes that are configured for your current installation.

   The output is similar to the following example:

   ```
   actionscript:
     enabled: true
   appdynamics:
     enabled: true
   aws:
     enabled: true
   azure:
     enabled: true
   dynatrace:
     enabled: true
   hpe3par:
     enabled: true
   horizon:
     enabled: false
   hyperflex:
     enabled: false
   ...
   ```

This list identifies all the probes that are configured for your installation, and shows whether they are enabled (`true`) or disabled (`false`).

NOTE:
This list of probes is not identical to the list of probe pods that are running in your installation. Some probes use multiple pods. Probe pod names use the following convention, where *<ProbeName>* is the probe internal name (mentioned in the preceding lists), and *<NameExtension>* is an optional extension to that name in case there are multiple pods for this probe:

```
mediation-<ProbeName><NameExtension>
```

For example, if you run `kubectl get pods -n turbonomic`, the results can show the following for the `vcenter` probe:

```
NAME                                          READY   STATUS    RESTARTS
mediation-vcenter-5bc4f5fbd4-nzm4j            1/1     Running   0
mediation-vcenterbrowsing-5c5987f66c-bfjq4    1/1     Running   0
```

Enabling or disabling probes

To enable or disable probes, edit the `cr.yaml` file. You can add new probes, change the values of the `enabled:` properties, and apply changes to reload the platform components.

1. Log in to an SSH terminal session for your Workload Optimization Manager instance and display the list of available probes.
2. In the SSH session, open the `cr.yaml` file for editing.

   ```
   vi /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

3. Edit the probe entries.

   To enable or disable currently configured probes, find the probes to edit and change the settings to `true` (enable) or `false` (disable).

   To add new probes to the list, copy the probe entry you want from the output when you used `cat` to view the available probes. Then paste that entry into the `cr.yaml` file and set `enabled: true`.
4. Save your changes to the platform.

   NOTE:
   During the online or offline upgrade process, you should not use `kubectl` to apply these changes now.

   After you save your changes, apply the changes:

   ```
   kubectl apply -f \
   /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

5. Verify that the probes have installed correctly and all the Workload Optimization Manager pods have started.

   ```
   kubectl get pods -n turbonomic
   ```

   Review the list for the mediation pods that implement your probes. All pods should display READY and STATUS states similar to:

   ```
   NAME                      READY   STATUS    RESTARTS
   [...]                     1/1     Running   0
   ```

6. View the new probe configuration in the user interface.

   Refresh your browser and go to the Target Management page. You will now see the target categories and types to match your configuration changes.

(Optional) Disabling component health monitoring

The Notification Center includes component health information to notify you of Workload Optimization Manager component issues and the steps you can take to resolve these issues. To open Notification Center, click the information icon on the main menu.

If you would rather use your own tools to monitor the Workload Optimization Manager component status, you can disable the feature in the CRD file.

To disable component health monitoring:

1. Open an SSH terminal session on your Workload Optimization Manager instance.

   Log in with the System Administrator that you set up when you installed Workload Optimization Manager:

   - Username: `turbo`
   - Username: `[your_private_password]`

2. Open the following `cr.yaml` file in a text editor:

   ```
   /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

3. Add the following to the CRD file:

   ```
   properties:
     global:
       featureFlags:
         componentHealth: false
   ```

4. When you are done editing the `cr.yaml` file, save and apply your changes.

   - Save your changes and quit the text editor.
   - To apply the changes, run the following command in your SSH terminal session:

     ```
     kubectl apply -f \
     /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
     ```

5. Delete the api and Cluster Manager pods.

   Deleting these pods triggers them to restart, which loads the changes you made.

   - Close the Workload Optimization Manager UI if it is open in your browser.
   - To get the full pod names, run the command:

     ```
     kubectl get pods
     ```

   - Find the entries for the pods that begin with `api` and `clustermgr`. For example, assume that the entries are:

     ```
     ...
             api-7df7b8c8c7-6rqgg
     ...
             clustermgr-7c6579745d-m8dwc
     ...
     ```

   - Run the following command to delete the pods that use the `api` and `clustermgr` pod names:

```
kubectl delete pods api-7df7b8c8c7-6rqgg clustermgr-7c6579745d-m8dwc
```

NOTE:
The pod names change when deleted and started again so always get the latest pods before you run the delete command.

The pods are deleted and re-created.

6. Verify the deletion and recreation of the pods.

   To verify that the pods are re-created and running, run the command:

   ```
   kubectl get pods
   ```

   After all pods start, the READY column shows `1/1`, `2/2`, and so on. The STATUS column for each pod shows `Running`.

   The output is similar to the following example:

   ```
   clustermgr-7c6579745d-m8dwc                    1/1     Running   1 (2d16h ago)    3d15h
   api-7df7b8c8c7-6rqgg                           1/1     Running   1 (2d16h ago)    3d15h
   ```

(Optional) Overriding default ports

The MariaDB or MySQL database server uses port 3306 by default and the Postgres database (for embedded reporting) uses port 5432 by default. Use the following sample configurations to override the default ports.

Sample configuration for MariaDB with IP address used for the database server with custom port (non-3306)

```
global:
    repository: icr.io/cpopen/turbonomic
    tag: 3.13.0
    externalIP: 10.10.10.10
    externalDbIP: 10.10.10.11
    externalDbPort: 3307
```

Sample configuration for MariaDB with a fully qualified domain name used for the database server with custom port (non-3306)

```
global:
    repository: icr.io/cpopen/turbonomic
    tag: 3.13.0
    externalIP: 10.10.10.10
    externalDBName: MyDatabaseServer.com
    externalDbPort: 3307
```

Sample configuration for Postgres database (for embedded reporting) with custom port (non-5432)

```
  global:
    repository: icr.io/cpopen/turbonomic
    tag: 3.13.0
    externalDbIP: 10.10.10.30
    externalTimescaleDBIP: 10.10.10.31
    externalTimescaleDBPort: 5434
```

License installation and first-time login

Before you begin, make sure you have your full or trial license key file that was sent to you in a separate email. Save the license file on your local machine so you can upload it to your Workload Optimization Manager installation.

To use Workload Optimization Manager for the first time, perform the following steps:

1. Open and browser window and then specify the address of your Workload Optimization Manager instance in the address bar.
2. In the login page, specify the required credentials.
   - **Username** – Specify `administrator`.
   - **Password** – Specify your password.

     The new password must comply with the strong password policy (a mixture of uppercase and lowercase letters, numbers, and a symbol).
   - **Repeat password** – Specify your password again.

   Record these credentials for future use.
3. Continue setting up your Workload Optimization Manager installation.

   Click **Let's Go**.
4. Open the **Enter License** fly-out.

   Click **Import License**.
5. Upload your license key file.
   a. In the Enter License fly-out, you can upload the license in one of the following ways. Be sure to upload only `.xml`, `.lic`, or `.jwt` files.
      - Drag and drop the license key file into the Enter License fly-out.
      - Browse to the license key file.

   b. Click **Save**.

Single Sign-On authentication

If your company policy supports Single Sign-On (SSO) authentication, you can configure Workload Optimization Manager to support SSO authentication via either Security Assertion Markup Language (SAML) 2.0 or OpenID Connect 1.0.

At a high-level, to do this you will:

- Create external groups or at least one external user for SSO. See Managing User Accounts *(on page 1116)*.
- Configure Workload Optimization Manager to use either SAML or OpenID to support SSO.
  - SSO via a SAML Identity Provider (IdP). See Setting Up SAML Authentication *(on page 102)*.
  - SSO via an OpenID Identity Provider. See Setting Up OpenID Authentication *(on page 106)*.

When SSO is enabled, users provide their SSO credentials to log in to the Workload Optimization Manager instance. Once SSO is enabled, users cannot give local or Active Directory (AD) credentials for to login. The Identity Provider (IdP) will perform the authentication.

Prerequisites

Before you begin, make sure the IdP is set up for SSO. You can use a proprietary or public IdP. For examples of settings for a public Okta IdP, see Typical Settings for an IdP *(on page 100)*.

Typical settings for an IdP

Before you begin configuring Single Sign-On (SSO) for OVA deployments, make sure the IdP is set up for SSO.

This topic includes typical settings for a public Okta IdP that may be useful when you set up your IdP.

General SAML settings

Setting	Value
Single sign-on URL This is also referred to as the Assertion Consumer Service (ACS).	`https://<hostname>/vmturbo/saml2/sso/<samlRegistrationID>` Where: ▪ *<hostname>* is the host for Workload Optimization Manager. ▪ *<samlRegistrationID>* is the Registration ID from your SSO provider.
Single logout URL	`https://<hostname>/vmturbo/saml2/logout/saml2/slo` Where: ▪ *<hostname>* is the host for Workload Optimization Manager.
Recipient URL	`https://<hostname>/vmturbo/saml2/sso/<samlRegistrationID>` Where: ▪ *<hostname>* is the host for Workload Optimization Manager. ▪ *<samlRegistrationID>* is the Registration ID from your SSO provider.
Destination URL	`https://<hostname>/vmturbo/saml2/sso/<samlRegistrationID>` Where: ▪ *<hostname>* is the host for Workload Optimization Manager. ▪ *<samlRegistrationID>* is the Registration ID from your SSO provider.
Audience Restriction	`urn:test:turbo:markharm`
Default Relay State	
Name ID Format	`Unspecified`
Application username	The username for the account that is managed by Okta
Response	`Signed`
Assertion Signature	`Signed`
Signature Algorithm	`RSA_SHA256`
Digital Algorithm	`SHA256`
Assertion Encryption	`Unencrypted`
SAML Single Logout	`Enabled`
SP Issuer	`turbo`
Signature Certificate	`Example.cer (CN=apollo)`
authnContextClassRef	`PasswordProtectedTransport`

Setting	Value
Honor Force Authentication	`Yes`
SAML Issuer ID	`http://www.okta.com/$(org.externalKey)`

Setting Up SAML authentication

Security Assertion Markup Language (SAML) is an XML-based open standard for exchanging authentication and authorization data between parties. To configure Workload Optimization Manager to authenticate via SAML, complete the following steps.

1. (Required) Create external groups or at least one external user for SSO.

   IMPORTANT:
   When SSO is enabled, Workload Optimization Manager only permits logins via the SSO IdP. Whenever you navigate to your Workload Optimization Manager installation, it redirects you to the SSO Identity Provider (IdP) for authentication before displaying the Workload Optimization Manager user interface.

   Before you enable SSO for your Workload Optimization Manager installation, *you must configure at least one SSO user with Workload Optimization Manager administrator privileges*. If you do not, then once you enable SSO you will not be able to configure any SSO users in Workload Optimization Manager. To authorize an SSO user as an administrator, use **EXTERNAL AUTHENTICATION** to do one of the following:

   - Configure a single SSO user with administrator authorization.

     Add an external user. The username must match an account that is managed by the IdP.
   - Configure an SSO user group with administrator authorization.

     Add an external group. The group name must match a user group on the IdP, and that group must have at least one member.

   For information about creating external groups or external users for SSO, see Managing User Accounts *(on page 1116)*.

2. (Required) Ensure that chrony is configured and the system time on your Workload Optimization Manager instance is correct.

   For instructions, see Synchronizing Time *(on page 58)*.

3. Obtain the metadata from your IdP.

   You will use this metadata to configure SSO in the Workload Optimization Manager CR file located at:

   ```
   /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

   To get the metadata:

   a. Contact your security administrator to obtain the metadata from IdP.

   b. Save the metadata file in a directory on your local machine.

      ```
      /tmp/MySamlMetadata.txt
      ```

   c. Compare your metadata to the example IdP metadata *(on page 105)*.

      If your metadata includes optional attribute tags that are not listed in the example, remove those optional attribute tags since they are not supported.

4. Obtain a certificate from IdP.

   Contact your security administrator to obtain a certificate from IdP.

5. Update the CR file with your SAML configuration.

Edit the `cr.yaml` file that configures your Workload Optimization Manager node, and then deploy or restart the node.

- Display the contents of your downloaded SAML metadata.

  For example, assuming you saved the file to this location on your local machine, run the command:

  ```
  cat /tmp/MySamlMetadata.txt
  ```

- Open the CR file for editing.

  In a shell, cd to the `deploy/crds` directory in the Workload Optimization Manager VM.

  ```
  cd /opt/turbonomic/kubernetes/operator/deploy/crds
  ```

  Open the CR file for editing. For example, to open the file in VI:

  ```
  vi charts_v1alpha1_xl_cr.yaml
  ```

  As you edit this file, refer to the metadata that you obtained from your IdP.

- In the CR file, search for or scroll to the entry for the API component.

  ```
  apiVersion: charts.helm.k8s.io/v1alpha1
  ```

- Turn on the SAML feature.

  For the first API property under `spec:properties:api:`, set the following:

  ```
  samlEnabled: true
  ```

- Set the SSO endpoint.

  In the SAML metadata, find the entry for `md:SingleSignOnService`. Within that element, find the `Location` attribute. The value of `Location` is the SSO endpoint. Using the sample metadata, make the following setting in your CR file:

  ```
  samlWebSsoEndpoint:
   https://dev-771202.oktapreview.com/app/ibmdev771202_turbo2_1/exkexl6xc9MhzqiC30h7/sso/saml
  ```

- (Optional) Set the SLO endpoint.

  If you configured a logout endpoint in IdP, you can set the SLO endpoint to enable the single logout function.

  In the SAML metadata, find the entry for `md:SingleLogoutService`. Within that element, find the `Location` attribute. The value of `Location` is the SLO endpoint. Using the sample metadata, make the following setting in your CR file:

  ```
  samlWebSloEndpoint:
   https://dev-771202.oktapreview.com/app/ibmdev771202_turbo2_1/exkexl6xc9MhzqiC30h7/slo/saml
  ```

- Set the SAML entity ID.

  In the SAML metadata, find the entry for `md:EntityDescriptor`. Within that element, find the `entityID` attribute. Using the sample metadata, make the following setting in your CR file:

  ```
  samlEntityId: http://www.okta.com/exkexl6xc9MhzqiC30h7
  ```

- Set the SAML registration.

  ```
  samlRegistrationId: simplesamlphp
  ```

- Set the SAML SP entity ID.

  ```
  samlSpEntityId: turbo
  ```

- Enter the SAML certificate.

In the metadata that you got from your IdP, find the entry for `<ds:X509Certificate>`. Copy the characters between `<ds:X509Certificate>` and `</ds:X509Certificate>`.

Create an entry for the certificate in the API properties section of the CR file. On a new line, enter:

```
samlIdpCertificate: |
```

Open a new line after the entry you just created, and paste the certificate content that you copied from your metadata file.

The finished API section of the CR file is similar to the following example:

```
apiVersion: charts.helm.k8s.io/v1alpha1
kind: Xl
metadata:
  name: xl-release
spec:
  properties:
    api:
      samlEnabled: true
      samlWebSsoEndpoint:
 https://dev-771202.oktapreview.com/app/ibmdev771202_turbo2_1/exkexl6xc9MhzqiC30h7/sso/saml
      samlEntityId: http://www.okta.com/exkfdsn6oy5xywqCO0h7
      samlRegistrationId: simplesamlphp
      samlSpEntityId: turbo
      samlIdpCertificate: |
        -----BEGIN CERTIFICATE-----
        MIIDpDCCAoygAwIBAgIGAWMnhv7cMA0GCSqGSIb3DQEBCwUAMIGSMQswCQYDVQQGEwJVUzETMBEG
        A1UECAwKQ2FsaWZvcm5pYTEWMBQGA1UEBwwNU2FuIEZyYW5jaXNjbzENMAsGA1UECgwET2t0YTEU
        MBIGA1UECwwLU1NPUHJvdmlkZXIxEzARBgNVBAMMCmRldi03NzEyMDIxHDAaBgkqhkiG9w0BCQEW
        DWluZm9Ab2t0YS5jb20wHhcNMTgwNTAzMTk0MTI4WhcNMjgwNTAzMTk0MjI4WjCBkjELMAkGA1UE
        BhMCVVMxEzARBgNVBAgMCkNhbGlmb3JuaWExFjAUBgNVBAcMDVNhbiBGcmFuY2lzY28xDTALBgNV
        BAoMBE9rdGExFDASBgNVBAsMC1NTT1Byb3ZpZGVyMRMwEQYDVQQDDApkZXYtNzcxMjAyMRwwGgYJ
        KoZIhvcNAQkBFg1pbmZvQG9rdGEuY29tMIIBIjANBgkqhkiG9w0BAQEFAAOCAQ8AMIIBCgKCAQEA
        ugxQGqHAXpjVQZwsO9n8l8bFCoEevH3AZbz7568XuQm6MK6h7/O9wB4C5oUYddemt5t2Kc8GRhf3
        BDXX5MVZ8G9AUpG1MSqe1CLV2J96rMnwMIJsKeRXr01LYxv/J4kjnktpOC389wmcy2fE4RbPoJne
        P4u2b32c2/V7xsJ7UEjPPSD4i8l2QG6qsUkkx3AyNsjo89PekMfm+Iu/dFKXkdjwXZXPxaL0HrNW
        PTpzek8NS5M5rvF8yaD+eE1zS0I/HicHbPOVvLal0JZyN/f4bp0XJkxZJz6jF5DvBkwIs8/Lz5GK
        nn4XW9Cqjk3equSCJPo5o1Msj8vlLrJYVarqhwIDAQABMA0GCSqGSIb3DQEBCwUAA4IBAQC26kYe
        LgqjIkF5rvxB2QzTgcd0LVzXOuiVVTZr8Sh57l4jJqbDoIgvaQQrxRSQzD/X+hcmhuwdp9s8zPHS
        JagtUJXiypwNtrzbf6M7ltrWB9sdNrqc99d1gOVRr0Kt5pLTaLe5kkq7dRaQoOIVIJhX9wgynaAK
        HF/SL3mHUytjXggs88AAQa8JH9hEpwG2srN8EsizX6xwQ/p92hM2oLvK5CSMwTx4VBuGod70EOwp
        6Ta1uRLQh6jCCOCWRuZbbz2T3/sOX+sibC4rLIlwfyTkcUopF/bTSdWwknoRskK4dBekFcvN9N+C
        p/qaHYcQd6i2vyor888DLHDPXhSKWhpG
        -----END CERTIFICATE-----
```

6. Save your changes to the CR file.
7. Apply the modified cr.yaml file.

```
kubectl apply -f /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
```

8. Restart the API component to load the new spec.
   a. Open an SSH terminal session to your Workload Optimization Manager instance.
   b. Restart the API component.

      ```
      kubectl delete pod api-<API_POD_ID>
      ```

      To auto-fill the pod ID, type `api-` and then press TAB.
9. Verify that the configuration is successful.
   a. Navigate to the Workload Optimization Manager User Interface.

      You are automatically redirected to your IdP for authentication.

b. Log in with the username that is a member of the external group or external user that you previously configured.

c. Verify that the system time on your Workload Optimization Manager instance is correct.

   If the time is not synchronized, this might cause an HTTP `Status 401 -authentication failed` exception in the browser.

d. If the configuration is not successful, look for an HTTP `Status 500` exception in the product log. If this exception exists, review your CR file for invalid entries.

Example of IdP metadata

The following example of IdP metadata may be useful when you are examining the optional attributes in your metadata.

If your metadata includes optional attribute tags that are not listed in the example, remove those optional attribute tags since they are not supported.

```
   <?xml version="1.0" encoding="UTF-8"?>
        <md:EntityDescriptor xmlns:md="urn:oasis:names:tc:SAML:2.0:metadata"
        entityID="http://www.okta.com/exkexl6xc9MhzqiC30h7">
        <md:IDPSSODescriptor WantAuthnRequestsSigned="false"
        protocolSupportEnumeration="urn:oasis:names:tc:SAML:2.0:protocol">
        <md:KeyDescriptor use="signing">
        <ds:KeyInfo xmlns:ds="http://www.w3.org/2000/09/xmldsig#">
        <ds:X509Data>
        <ds:X509Certificate>
        MIIDpDCCAoygAwIBAgIGAWMnhv7cMA0GCSqGSIb3DQEBCwUAMIGSMQswCQYDVQQGEwJVUzETMBEG
        A1UECAwKQ2FsaWZvcm5pYTEWMBQGA1UEBwwNU2FuIEZyYW5jaXNjbzENMAsGA1UECgwET2t0YTEU
        MBIGA1UECwwLU1NPUHJvdmlkZXIxEzARBgNVBAMMCmRldi03NzEyMDIxHDAaBgkqhkiG9w0BCQEW
        DWluZm9Ab2t0YS5jb20wHhcNMTgwNTAzMTk0MTI4WhcNMjgwNTAzMTk0MjI4WjCBkjELMAkGA1UE
        BhMCVVMxEzARBgNVBAgMCkNhbGlmb3JuaWExFjAUBgNVBAcMDVNhbiBGcmFuY2lzY28xDTALBgNV
        BAoMBE9rdGExFDASBgNVBAsMC1NTT1Byb3ZpZGVyMRMwEQYDVQQDDApkZXYtNzcxMjAyMRwwGgYJ
        KoZIhvcNAQkBFg1pbmZvQG9rdGEuY29tMIIBIjANBgkqhkiG9w0BAQEFAAOCAQ8AMIIBCgKCAQEA
        ugxQGqHAXpjVQZwsO9n8l8bFCoEevH3AZbz7568XuQm6MK6h7/O9wB4C5oUYddemt5t2Kc8GRhf3
        BDXX5MVZ8G9AUpG1MSqe1CLV2J96rMnwMIJsKeRXr01LYxv/J4kjnktpOC389wmcy2fE4RbPoJne
        P4u2b32c2/V7xsJ7UEjPPSD4i8l2QG6qsUkkx3AyNsjo89PekMfm+Iu/dFKXkdjwXZXPxaL0HrNW
        PTpzek8NS5M5rvF8yaD+eE1zS0I/HicHbPOVvLal0JZyN/f4bp0XJkxZJz6jF5DvBkwIs8/Lz5GK
        nn4XW9Cqjk3equSCJPo5o1Msj8vlLrJYVarqhwIDAQABMA0GCSqGSIb3DQEBCwUAA4IBAQC26kYe
        LgqjIkF5rvxB2QzTgcd0LVzXOuiVVTZr8Sh57l4jJqbDoIgvaQQrxRSQzD/X+hcmhuwdp9s8zPHS
        JagtUJXiypwNtrzbf6M7ltrWB9sdNrqc99d1gOVRr0Kt5pLTaLe5kkq7dRaQoOIVIJhX9wgynaAK
        HF/SL3mHUytjXggs88AAQa8JH9hEpwG2srN8EsizX6xwQ/p92hM2oLvK5CSMwTx4VBuGod70EOwp
        6Ta1uRLQh6jCCOCWRuZbbz2T3/sOX+sibC4rLIlwfyTkcUopF/bTSdWwknoRskK4dBekFcvN9N+C
        p/qaHYcQd6i2vyor888DLHDPXhSKWhpG
        </ds:X509Certificate>
        </ds:X509Data>
        </ds:KeyInfo>
        </md:KeyDescriptor>
        <md:NameIDFormat>urn:oasis:names:tc:SAML:1.1:nameid-format:unspecified</md:NameIDFormat>
        <md:NameIDFormat>urn:oasis:names:tc:SAML:1.1:nameid-format:emailAddress</md:NameIDFormat>
        <md:SingleSignOnService Binding="urn:oasis:names:tc:SAML:2.0:bindings:HTTP-POST"

Location="https://dev-771202.oktapreview.com/app/ibmdev771202_turbo2_1/exkexl6xc9MhzqiC30h7/sso/saml"/>
        <md:SingleSignOnService Binding="urn:oasis:names:tc:SAML:2.0:bindings:HTTP-Redirect"

Location="https://dev-771202.oktapreview.com/app/ibmdev771202_turbo2_1/exkexl6xc9MhzqiC30h7/sso/saml"/>
        </md:IDPSSODescriptor>
        </md:EntityDescriptor>
```

Setting Up OpenID authentication

According to the OpenID Foundation, "OpenID Connect 1.0 is a simple identity layer on the OAuth 2.0 protocol". OpenID Connect enables clients to verify user identity via a given authentication server. Workload Optimization Manager supports OpenID authentication through the following providers:

- Google
- IBM-MCM
- Okta

Logging in to Workload Optimization Manager with OpenID

When you configure OpenID in Workload Optimization Manager, the platform registers the OpenID clients that you specify. To log in through of these OpenID clients, you manually navigate to a URL that tells Workload Optimization Manager which client to use. It then redirects to the OpenID login screen for that given client.

The URL that you provide is in the form:

```
https://<hostname>/vmturbo/oauth2/login/code/<openIdClients>
```

Where:

- *<hostname>* is the host address for your installation of Workload Optimization Manager
- *<openIdClients>* is the is the client name of the OpenID provider you want to use

  You specify this as the `openIdClients` property when you configure OpenID in Workload Optimization Manager.

NOTE:
This URL must also be set in the **Authorized direct URIs** section of your provider's OpenID configuration.

For example, assume your Workload Optimization Manager host address is 10.10.12.34, and you configured an Okta OpenID client. In that case, when the Workload Optimization Manager login screen appears, you would navigate to:

```
https://10.10.12.34/vmturbo/oauth2/login/code/okta
```

After you navigate to that URL, the browser redirects to the OpenID login screen, where you can enter credentials for a single user or a user group.

NOTE:
To authenticate a user group, the group must be configured on the OpenID provider, and also on Workload Optimization Manager. The group name must be identical in both configurations.

On the OpenID provider, the client that you are using should include groups scope values that give specific names for user groups. Contact your OpenID administrator to get the group names. Then, on Workload Optimization Manager you should create user groups that use the same names.

For example, assume the OpenID ID token includes the following groups claim:

```
{
  "sub": "1234567890",
  "name": "My_User_Name",
  "iat": "12121212",
  "groups": "My_Special_User_Group"
}
```

To use the group `My_Special_User_Group` for authentication, you must create a user group in Workload Optimization Manager with the name `My_Special_User_Group`. Any members of that group will then get the role you have assigned to that user group.

Configuring OpenID on Workload Optimization Manager

To configure Workload Optimization Manager to authenticate via OpenID:

1. (Required) Ensure that chrony is configured and the system time on your Workload Optimization Manager instance is correct.

   For instructions, see Synchronizing Time *(on page 58)*.

2. Obtain the necessary data from your OpenID provider.

   Contact your security administrator to obtain the data from the provider. You will use this data to configure SSO in the Workload Optimization Manager CR file located at:

   ```
   /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

The data you need and the properties you declare in the CR file will differ depending on the OpenID provider you want to use:

- Google:

CR fields	Description
`openIdClients`	`google` The name of the OpenID client you are using to perform authentication.
`openIdClientId`	The OAuth2 Client Identifier for the OpenID client that you are using.
`openIdClientSecret`	The OAuth2 Client Secret for the OpenID client that you are using.

- IBM-MCM:

CR fields	Description
`openIdClients`	`ibm` The name of the OpenID client you are using to perform authentication.
`openIdClientAuthentica-tion`	`post` The client authentication method.
`openIdUserAuthentication`	`form` The user authentication method.
`openIdClientId`	The OAuth2 Client Identifier for the OpenID client that you are using.
`openIdClientSecret`	The OAuth2 Client Secret for the OpenID client that you are using.
`openIdAccessTokenUri`	The URI the login process will use to get an Access Token.
`openIdUserAuthorization-Uri`	The URI to the Authorization Endpoint for OpenID Connect.
`openIdUserInfoUri`	The URI to the OpenID Connect UserInfo endpoint.
`openIdJwkSetUri`	The URI to get the JSON Web Key set that can verify the Access Token.
`openIdExternalGroupTag`	The name of a custom group to use for authentication.

- Okta:

CR fields	Description
`openIdClients`	`okta` The name of the OpenID client you are using to perform authentication.
`openIdClientId`	The OAuth2 Client Identifier for the OpenID client that you are using.
`openIdClientSecret`	The OAuth2 Client Secret for the OpenID client that you are using.
`openIdAccessTokenUri`	The URI the login process will use to get an Access Token.
`openIdUserAuthorization-Uri`	The URI to the Authorization Endpoint for OpenID Connect.
`openIdUserInfoUri`	The URI to the OpenID Connect UserInfo endpoint.
`openIdJwkSetUri`	The URI to get the JSON Web Key set that can verify the Access Token.

3. Update the Workload Optimization Manager CR file with your configuration data.

You now have the data that you need to configure SSO via OpenID. You will edit the `cr.yaml` file that configures your Workload Optimization Manager node, and then deploy or restart the node.

- Open the CR file for editing.

  ```
  cd /opt/turbonomic/kubernetes/operator/deploy/crds
  ```

  Then open the CR file for editing. For example, to open the file in VI:

  ```
  vi charts_v1alpha1_xl_cr.yaml
  ```

  As you edit this file, you will refer to the data that you obtained from your authentication provider.

- In the CR file, navigate to the entry for the API component.

  In the CR file search for or scroll to the entry:

  ```
  apiVersion: charts.helm.k8s.io/v1alpha1
  ```

  You will make changes to this component spec, under `spec:properties:api:`

- Turn on the OpenID feature.

  For the first API property, set the following:

  ```
  openIdEnabled: true
  ```

  The output is similar to the following example:

  ```
  apiVersion: charts.helm.k8s.io/v1alpha1
  kind: Xl
  metadata:
    name: xl-release
  spec:
    properties:
      api:
        openIdEnabled: true
  ```

- Enter the relevant OpenId data for your authentication provider. The CR file should be similar to these examples, depending on which provider you use.
  - Google:

    ```
    apiVersion: charts.helm.k8s.io/v1alpha1
    kind: Xl
    metadata:
      name: xl-release
    spec:
      properties:
        api:
          openIdEnabled: true
          openIdClients: google
          openIdClientId: xxxx-4vinrdgllag5p84jjebc6xxxxxx5u.apps.googleusercontent.com
          openIdClientSecret: xxxxxhGcdFEjQa-xxxxxxxx
    ```

  - IBM-MCM:

    ```
    apiVersion: charts.helm.k8s.io/v1alpha1
    kind: Xl
    metadata:
      name: xl-release
    spec:
      properties:
        api:
          openIdEnabled: true
          openIdClients: ibm
    ```

```
     openIdClientAuthentication: post
     openIdUserAuthentication: form
     openIdClientId: turbonomic-mcm-demo
     openIdClientSecret: "xxxxxxvZ2ZscDhtOFVxxxxxxU3d6cXR4cTZhb2xxxxxxRT0K"
     openIdAccessTokenUri:
https://icp-console.apps.blue-13.dev.multicloudops.io/idprovider/v1/auth/token
     openIdUserAuthorizationUri:
https://icp-console.apps.blue-13.dev.multicloudops.io/idprovider/v1/auth/authorize
     openIdUserInfoUri:
https://icp-console.apps.blue-13.dev.multicloudops.io/v1/auth/userInfo
     openIdJwkSetUri:
https://icp-console.apps.blue-13.dev.multicloudops.io/oidc/endpoint/OP/jwk
```

- Okta

```
apiVersion: charts.helm.k8s.io/v1alpha1
kind: Xl
metadata:
  name: xl-release
spec:
  properties:
    api:
      openIdEnabled: true
      openIdClients: okta
      openIdClientId: xxxxxxxxxh1xhQnSKxxxx
      openIdClientSecret: xxxxxxxxxxtIhVCIRUnhq4xxxxxxxDdhLdqx0
      openIdAccessTokenUri: https://vmturbo.okta.com/oauth2/v1/token
      openIdUserAuthorizationUri: https://vmturbo.okta.com/oauth2/v1/authorize
      openIdUserInfoUri: https://vmturbo.okta.com/oauth2/v1/userinfo
      openIdJwkSetUri: https://vmturbo.okta.com/oauth2/v1/keys
```

4. Save your changes to the CR file.
5. Apply the modified cr.yaml file.

```
kubectl apply -f /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
```

6. Restart the API component to load the new spec.
   a. Open an SSH terminal session to your Workload Optimization Manager instance.
   b. Restart the API component.

      ```
      kubectl delete pod api-<API_POD_ID>
      ```

      To auto-fill the pod ID, type `api-` and then press TAB.
7. Verify that the configuration is successful.
   a. Navigate to the Workload Optimization Manager User Interface.

      You will be automatically redirected to your authentication provider for authentication.
   b. Log in with the username that is a member of the external group or external user that you previously configured.
   c. Verify that the system time on your Workload Optimization Manager instance is correct.

      If the time is not synchronized, this might cause an `HTTP Status 401 -authentication failed` exception in the browser.
   d. If the configuration is not successful, look for an `HTTP Status 500` exception in the product log. If this exception exists, review your CR file for invalid entries.

Disabling Single Sign-On

If for some reason you no longer want to use SSO, you can disable it for your Workload Optimization Manager installation. To disable Single Sign-On, perform these steps:

1. Update the SSO configuration to disable it.
    a. Open an SSH terminal session to your Workload Optimization Manager instance.
    b. Open the CR file for editing.

    In a shell, `cd` to the `deploy/crds` directory in the Workload Optimization Manager VM:

    ```
    cd /opt/turbonomic/kubernetes/operator/deploy/crds
    ```

    Then open the CR file for editing. For example, to open the file in VI:

    ```
    vi charts_v1alpha1_xl_cr.yaml
    ```

    c. In the CR file, navigate to the entry for the API component.

    In the CR file search for or scroll to the entry:

    ```
    apiVersion: charts.helm.k8s.io/v1alpha1
    ```

    You will make changes to this component spec, under `spec:properties:api:`

    d. Turn off the SSO feature.

    The entry to set to false is different depending on whether you use SAML or OpenID authentication:

    - SAML Authentication:

      Set the `samlEnabled:` property to `false`.

      ```
      samlEnabled: false
      ```

    - OpenID Authentication:

      Set the `openIdEnabled:` property to `false`.

      ```
      openIdEnabled: false
      ```

    e. Save your changes to the CR file.
2. Apply the modified cr.yaml file.

    ```
    kubectl apply -f /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
    ```

3. Restart the API component.

    In the same SSH terminal session that you opened to edit the CR file:

    a. Use sudo as root.

    ```
    sudo bash
    ```

    b. Restart your API component.

    ```
    kubectl delete pod api-<API_POD_ID>
    ```

    NOTE:
    To auto-fill the pod ID, type `api-` and then press TAB.

4. Verify that the configuration is successful.
    a. Navigate to the Workload Optimization Manager User Interface.

       You will no longer be redirected to your IdP for authentication. You will be redirected to the default Workload Optimization Manager login screen.
    b. Log in with a local account or an Active Directory (AD) account.

Reference: FIPS cipher suites

NOTE:
The process that is described here is applicable only to the OVA deployment model.

The Federal Information Processing Standard (FIPS) is in place to ensure the cryptographic strength of secure connections. By default, Workload Optimization Manager ships with the following TLS protocol versions and cipher suites already enabled:

- TLSv1.2:
  - `TLS_ECDHE_RSA_WITH_CHACHA20_POLY1305_SHA256`
  - `TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384`
  - `TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256`
- TLSv1.3:
  - `TLS_CHACHA20_POLY1305_SHA256`
  - `TLS_AES_256_GCM_SHA384`
  - `TLS_AES_128_GCM_SHA256`

Modifying the protocols and cipher suites

To modify the cipher suites and protocols used by Workload Optimization Manager, edit or add the `nginx` section of the `Xl` Custom Resource, as described in the following example:

```
apiVersion: charts.helm.k8s.io/v1
kind: Xl
metadata:
  name: xl-release
spec:
...
  nginx:
    # This variable contains a space separated list of TLS protocol versions
    # to support.
    # The value found here will be inserted into an 'ssl_protocols' directive
    # of the nginx configuration for turbonomic
    tlsProtocols: TLSv1.2 TLSv1.3
```

```
    # This variable contains a colon separated list of cipher suites
    # that should be supported for TLSv2.
    # The value placed here will be inserted into the 'ssl_ciphers' directive
    # of the nginx configuration for turbonomic
    tlsv2Ciphers: ECDHE-RSA-AES128-GCM-SHA256:ECDHE-RSA-AES256-GCM-SHA384:ECDHE-RSA-CHACHA20-POLY1305
```

```
    # This variable contains a colon separated list of cipher suites
    # that should be supported for TLSv3.
    # The value placed here will be inserted into the
    # 'ssl_conf_command Ciphersuites' directive of the nginx
    # configuration for turbonomic
    tlsv3Ciphers: TLS_AES_128_GCM_SHA256:TLS_AES_256_GCM_SHA384:TLS_CHACHA20_POLY1305_SHA256
```

Modifying the protocols and cipher suite on a Workload Optimization Manager OVA

If necessary, you can modify the cipher suites to comply with your internal policies.

1. Open an SSH terminal session on your Workload Optimization Manager instance.

   Log in with the System Administrator credentials that you set up when you installed Workload Optimization Manager:

   - Username: `turbo`
   - Password: `[your_private_password]`

2. In the SSH session, open the `cr.yaml` file for editing. For example:

   ```
   vi /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

3. Edit the cipher suite.

   a. Search for the `nginx` section in the file or create the section if not present.

   b. Modify the `nginx` section (as described in the previous section of this topic) and then save the file.

4. Apply your changes to the platform.

   ```
   kubectl apply -f \
   /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

Reference: Step-wise platform deployment

NOTE:
The process that is described here is applicable only to the OVA deployment model.

After you install the Workload Optimization Manager VM that will host the platform, you can install the platform components, as follows:

1. (Optional) Configure Single Sign-On Authentication (SSO) for this installation.

   If you plan to use SSO to authenticate your Workload Optimization Manager users, edit the `charts_v1alpha1_xl_cr.yaml` file before you complete the installation, or edit it later and restart the affected components. For more information, see Single Sign-On Authentication *(on page 100)*.

2. Deploy Workload Optimization Manager Kubernetes nodes.

   When you deploy Workload Optimization Manager on Kubernetes, you deploy one Kubernetes node as a VM that hosts pods to run the Workload Optimization Manager components. The script to deploy and initialize the Kubernetes node also deploys the Kubernetes pods that make up the Workload Optimization Manager application.

   Start a secure session (SSH) on your Workload Optimization Manager VM as the turbo user and complete the following steps:

   a. Initialize the Kubernetes node and deploy the pods.

      ```
      sudo /opt/local/bin/t8cInstall.sh
      ```

      The script takes up to 20 minutes to complete.

   b. Verify that the deployment succeeded.

      At the end of the script output, in the summary section, verify that no errors are reported. If any errors are reported, contact Workload Optimization Manager Support.

   c. Verify that Workload Optimization Manager installed correctly.

      ```
      kubectl get pods -n turbonomic
      ```

After all pods start, the READY column shows `1/1`, `2/2`, and so on. The STATUS column for each pod shows `Running`.

The output is similar to the following example:

```
NAME                                          READY   STATUS    RESTARTS
action-orchestrator-b6454c9c8-mfl85           1/1     Running   0
api-7887c66f4b-shndq                          1/1     Running   0
auth-5b86976bc8-vxwz4                         1/1     Running   0
clustermgr-85548678d9-r5wb8                   1/1     Running   0
consul-7f684d8cb8-6r677                       1/1     Running   0
cost-5f46dd66c4-6d6cb                         1/1     Running   0
extractor-5f41dd61c4-4d6lq                    1/1     Running   0
group-5bfdfbc6f8-96bsp                        1/1     Running   0
history-5fc7fbc855-6zslq                      1/1     Running   0
kafka-74cc77db94-dfrbl                        1/1     Running   0
kubeturbo-785945b66c-ltpjg                    1/1     Running   0
market-5f54699447-z4wkm                       1/1     Running   0
mediation-actionscript-57b4fc6df-4lzfv        1/1     Running   0
mediation-appdynamics-6d65f8766f-kb44l        1/1     Running   0
mediation-hpe3par-d7c475c4c-v8ftc             1/1     Running   0
mediation-hyperv-6bd8c94df5-4dbzx             1/1     Running   0
mediation-netapp-7f8fc955d9-4kkdl             1/1     Running   0
mediation-oneview-7dbd7b54cf-7rfqp            1/1     Running   0
mediation-pure-58c4bd8cd9-8n256               1/1     Running   0
mediation-ucs-6f4bb9889-9rnqk                 1/1     Running   0
mediation-vcenter-5bc4f5fbd4-nzm4j            1/1     Running   0
mediation-vcenterbrowsing-5c5987f66c-bfjq4    1/1     Running   0
mediation-vmax-6c59969b89-28t9j               1/1     Running   0
mediation-vmm-9c4878cf9-rfxnl                 1/1     Running   0
nginx-5b775f498-sm2mm                         1/1     Running   0
plan-orchestrator-6dffc4c9b6-p5t5n            1/1     Running   0
prometheus-mysql-exporter-5574d587c6-vnqx9    1/1     Running   0
prometheus-server-64f6954bbf-xhvbb            2/2     Running   0
reporting-b44fbdfb4-8fjv5                     1/1     Running   0
repository-6d555bb4bf-fxldh                   1/1     Running   0
rsyslog-fd694878c-5tb2c                       1/1     Running   0
t8c-operator-558bcc758d-5h8mp                 1/1     Running   0
topology-processor-b646b786b-9skp7            1/1     Running   0
zookeeper-5f65b5bf69-nnmbt                    1/1     Running   0
```

d. Synchronize the system clock.

   To ensure the correct display of data, and to support Single Sign-On (SSO) authentication, you need to synchronize the system clock. For more information, see Synchronizing Time *(on page 58)* and Single Sign-On Authentication *(on page 100)*.

e. Verify that the Load Balancer installed correctly.

```
kubectl get services -n turbonomic | grep LoadBalancer
```

   The output is similar to the following example:

```
nginx  LoadBalancer   10.10.10.10   10.10.10.11  443:32669/TCP,80:32716/TCP  17h
```

f. Configure mediation.

   The installation script automatically enables a default set of mediation probes. After installation completes, you can change the set of enabled mediation probes (see Enabling and Disabling Probe Components *(on page 95)*).

For Workload Optimization Manager to manage your IT environment, it must attach to targets in your environment so it can perform discovery and run actions. The combination of the processes of discovery and action execution is *mediation*. This release of Workload Optimization Manager supports mediation through the following targets.

- Applications and Databases
  - Apache Tomcat 7.x, 8.x, or 8.5.x
  - Cisco AppDynamics 4.1+
  - Datadog
  - Dynatrace 1.1+
  - IBM WebSphere Application Server 8.5+
  - Instana release 242 or later
  - JBoss Application Server 6.3+
  - JVM 6.0+
  - SQL Server 2012, 2014, 2016, 2017, 2019, or 2022
  - MySQL 8.0
  - New Relic
  - Oracle 19c or 21c
  - Oracle WebLogic 12c
- Container Platform
  - Kubernetes 1.27 up to the latest supported GA version, including (but not limited to) Amazon Elastic Kubernetes Service (EKS), Azure Kubernetes Service (AKS), Google Kubernetes Engine (GKE), IBM Cloud Kubernetes Service (IKS), and Rancher
  - Red Hat supported GA versions of Red Hat OpenShift 4.1x, including (but not limited to) Red Hat OpenShift Kubernetes Service (ROKS), Red Hat OpenShift on AWS (ROSA), Red Hat OpenShift on Azure (ARO), and Red Hat OpenShift on Google
- Custom
  - Data Ingestion Framework
- Fabric and Network
  - Cisco UCS Manager 3.1+
  - HPE OneView 3.00.04
- Guest OS Processes
  - SNMP
  - WMI: Windows 7, 8, 8.1, 10, 2008 R2, 2012, 2012 R2, 2016, or 2019
- Hyperconverged
  - Cisco HyperFlex 3.5
  - Nutanix Community Edition
  - VMware vSAN
- Hypervisors
  - IBM PowerVM
    - HMC 10: Power8, Power9, or Power10
    - HMC 9: Power8 or Power9 only; Power7 LPARs and Systems are ignored.
  - Microsoft Hyper-V 2012 R2, 2016, 2019, or 2022
  - VMware vCenter 7.0 or 8.0

- IT Management
  - Flexera
    - Flexera One SaaS
    - FlexNet Manager Suite 2023 R2 or above
  - IBM Apptio
  - ServiceNow

    Certified for ServiceNow versions:
    - Utah
    - Vancouver
    - Washington DC
- Observability
  - Prometheus
- Orchestrator
  - Action Script
- Private Cloud
  - VMM for Microsoft System Center 2012 R2, System Center 2016, System Center 2019, or System Center 2022
- Public Cloud
  - Amazon Web Services (AWS)
  - AWS Billing
  - Azure Service Principal
  - Azure Billing
  - Microsoft Enterprise Agreement
  - Google Cloud
  - Google Cloud Billing
- Storage
  - EMC ScaleIO 2.x or 3.x
  - EMC VMAX using SMI-S 8.1+
  - EMC VPLEX Local Architecture with 1:1 mapping of virtual volumes and LUNs
  - EMC XtremIO XMS 4.0+
  - HPE 3PAR InForm OS 3.2.2+, 3PAR SMI-S, 3PAR WSAPI
  - IBM FlashSystem running on Spectrum Virtualize 8.3.1.2 or later (8.4.2.0 or later recommended)
  - NetApp Cluster Mode using ONTAP 8.0+ (excluding AFF and SolidFire)
  - Pure Storage F-series and M-series arrays running Purity 5.3.6 and 6.4.4 (Pure API 1.6)
- Virtual Desktop Infrastructure
  - Horizon 7.0 or higher

For more information, see Target Configuration *(on page 173)*.

IMPORTANT:
Wait until all the platform components start, are running, and are fully ready before your first log in. If you try to add a license or add a target to the platform before the components are all ready, the platform can fail to initialize correctly. After the components start, in your web browser, type the static IP address of your Workload Optimization Manager VM. Your browser redirects the login page for Workload Optimization Manager users.

3. Log in to the Workload Optimization Manager user interface using the server address and import your license key.

   For more information, see License installation and first-time login *(on page 100)*.

4. Create other user accounts with various roles. For more information about user accounts and roles, see Managing user accounts *(on page 1116)*.

   NOTE:
   For security reasons, you can create a different account with an `Administrator` role to serve as the main administrator of your Workload Optimization Manager installation, and then delete the default `administrator` account. You must always have at least one user account with administrator privileges.

Reference: Upgrading from MySQL 5.7 to 8.0

NOTE:
MySQL version 5.7 reached EOL on 21 October 2023 and is no longer supported. Workload Optimization Manager supports MySQL 8.0.x as an alternative database server for MariaDB for running Workload Optimization Manager.

If you utilized MySQL 5.7 as your external database and you need to upgrade to MySQL 8.0, follow the process outlined to perform an in-place upgrade.

NOTE:
The following process is an example of how to upgrade in-place by using Ubuntu 22.04.

1. Run the `upgrade-precheck.sh` script.

   You can use this script to make sure that your current installation of Workload Optimization Manager is ready to update. It is recommended that you run this script before updating your installation (see Checking Before Updating *(on page 133)*).

   NOTE:
   Any table corruptions must be resolved with IBM support before you upgrade.

2. Update and upgrade all installations.

```
apt-get update
```

```
apt-get upgrade
```

3. Remove the existing MySQL 5.7 installation.

```
apt-get remove mysql-server
apt-get autoremove
dpkg -l | grep mysql | grep ii
apt-get remove mysql-client mysql-common mysql-community-client mysql-community-server
```

4. Edit your Configuration File to include MySQL 8.0 configurations.

   Open the Configuration File (`/etc/mysql/my.cnf`) in an editor, change the values for these system variables, and then save your changes.

```
event_scheduler = ON
sql_mode = ERROR_FOR_DIVISION_BY_ZERO,NO_ENGINE_SUBSTITUTION
max_allowed_packet = 1GB
skip-log-bin
```

```
log_bin_trust_function_creators = ON
explicit_defaults_for_timestamp = OFF
character-set-server = utf8mb4
collation-server = utf8mb4_general_ci
```

5. Backup the updated MySQL Configuration Files in preparation for MySQL 8.0.

```
cp /etc/mysql/my.cnf /etc/mysql/mysql.conf.d/mysqld.cnf
```

```
cp /etc/mysql/my.cnf /etc/mysql/mysql.cnf
```

6. Install the latest version of MySQL 8.0.

```
apt-get install mysql-server
```

7. Verify the MySQL 8.0 service status.

```
systemctl status mysql.service
```

8. Reconfigure Workload Optimization Manager to use the MySQL 8.0 instance.

   Open the Custom Resource (CR) file for editing.

```
kubectl edit /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
```

   Modify the CR file to replace MySQL 5.7 values with 8.0, and include the `enableSecureDBConnection: true` property.

```
global:
  repository: icr.io/cpopen/turbonomic
  tag: 3.13.0
  externalIP: nginx_host_IP
  externalDbIP: MySQL_8.0_host_IP
  externalTimescaleDBIP: nginx_host_IP
```

```
spec:
  properties:
    global:
      enableSecureDBConnection: true
      dbRootUsername: MySQL_8.0_root_username
      dbRootPassword: MySQL_8.0_complex_password
```

9. Apply the changes that you made to the CR file.

```
kubectl apply -f \
  /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
```

10. Restart auth and all other DB-based components.

```
kubectl scale --replicas=0 deployment auth history action-orchestrator clustermgr cost group market
 plan-orchestrator suspend repository server-power-modeler topology-processor
```

Reference: Migrating from CentOS to Rocky Linux

CentOS Linux reached end-of-life on June 30, 2024 and is no longer supported. If you are running CentOS Linux 7, migrate to Rocky Linux to continue receiving security updates, fixes, and new features. For more information, see What to know about CentOS Linux EOL.

Complete the following steps to migrate from a CentOS Linux virtual machine (VM) image to a Rocky Linux virtual machine image. This process maintains the full integrity of the database and all groups, settings, policies, and historical data. This process takes into account any Workload Optimization Manager licenses, secrets,network settings, and certificates. Other customizations must be manually ported.

Prerequisites

You must first upgrade Workload Optimization Manager to the version on which you will deploy Rocky Linux.

The migration requires two scripts - a backup script (`xl-backup.sh`) and a restore script (`xl-restore.sh`) - which are located at `/opt/turbonomic/kubernetes/migration`. For CentOS-based installs, the scripts become available after you upgrade Workload Optimization Manager.

> **NOTE:** New scripts for renewing certificates (`k8s-certs-renew.sh`) and changing the IP address (`k8s-ip-change.sh`) are available for Rocky Linux. These scripts are designed only for Rocky Linux and might cause major errors if run on CentOS VMs. Scripts for renewing certificates (`kubeNodeCertUpdate.sh`) and changing the IP address (`kubeNodeIPChange.sh`) for VMs running CentOS Linux with Workload Optimization Manager 8.12.2 or later are in a versioned `bin-*` folder in the `/opt/local/` directory. This version of the note is consistent with how the scripts are discussed in the following topics:
>
> - Checking before updating *(on page 133)*
> - (Optional) Changing the IP address of the platform node *(on page 91)*

Obtain the CentOS VM image network information

To set up Rocky Linux VM image so that it is identical to the CentOS Linux VM image, run the following commands on the CentOS Linux VM image to obtain the network information.

- Hostname - `uname -n`
- IP address - `ifconfig eth0`
- Subnet mask - `ifconfig eth0`
- Gateway - `ip route`
- DNS Server (preferable DNS server and backup DNS server - separated by a space) - `cat/etc/resolv.conf`
- DNS domain - `cat /etc/resolv.conf`

The output is similar to the following example:

- Hostname - `xlhost`
- IP address - `10.16.173.8`
- Subnet mask - `255.255.254.0`
- Gateway - `10.16.172.0`
- DNS1/DNS2: `10.16.150.2 10.16.150.3`
- DNS domain - `demo.turbonomic.com`

During the backup, you must set the VM hard disk size in vCenter. For most installations, setting the size to `500 GB` is sufficient but there may be some customers who need more space. To determine the size of your installation, run the following command:

```
df -h | grep turbo-var_lib_mysql
```

The output is similar to the following example:

```
Filesystem                          Size  Used Avail Use% Mounted on
/dev/mapper/turbo-var_lib_mysql     800G   22G  779G   3% /var/lib/dbs
```

If the `Used` value (in this example, `22G`) is larger than `500 GB`, contact Workload Optimization Manager Support.

Preparing the CentOS Linux virtual machine image for backup

The following process backs up the existing CentOS Linux VM image to a `*.tar.xz` file on an attached drive (for example, `backup-2024-05-06.tar.xz`). The drive can then be detached from the origin VM and reattached to the destination VM. This process assumes that you have vCenter access to complete the necessary actions.

1. Add the drive to the virtual machine for the backup destination.
   a. In vCenter, right-click on your origin virtual machine and select **Edit Settings**.
   b. Click **Add New Device** and select **Hard Disk** from the menu.
   c. Expand **New Hard Disk** and set the size to `500 GB` (or the value provided by Workload Optimization Manager Support) and set **Disk Provisioning** to `Thin Provision` to save on non-used space for your host. Configure other options based on your environment, then click **OK**.
   d. The **Total hard disks** field increases in number to reflect that you added a new hard disk. Select the **Total hard disks** value to view the attached disks.
   e. Expand the new hard disk to obtain the name of the **Disk File**.

      Save the name and path to use in the restoration process.
2. Mount the drive and set drive ownership.

   You must have both an administrator (or similar) account in Workload Optimization Manager and root access to the virtual machine that uses an SSH terminal.
   a. Open an SSH terminal session to the existing CentOS VM image.

      Log in with the System Administrator that you set up when you installed Workload Optimization Manager.
   b. Create the `/opt/backup` directory, if it does not already exist. This directory is used as the mounted directory for your backup.

      ```
      mkdir /opt/backup
      ```
   c. Scan for the newly added drive to ensure that it is present.

      ```
      sudo su -

      echo "- - -" > /sys/class/scsi_host/host0/scan

      echo "- - -" > /sys/class/scsi_host/host1/scan

      echo "- - -" > /sys/class/scsi_host/host2/scan

      ls /dev/sdc
      ```
   d. Configure the disk partitions by using the following command and commonly-used answers to the follow-up prompts.

      ```
      fdisk /dev/sdc

      n

      p

      <press enter>

      <press enter>

      w
      ```
   e. Create a disk partition on the newly added drive.

```
mkfs.ext4 /dev/sdc1
```

f. Mount the drive that was attached in vCenter to the `/opt/backup` directory you created.

```
sudo mount /dev/sdc1 /opt/backup
```

g. Display the current mounts to your device to ensure that the drive is properly mounted to your VM.

You can add a `| grep` to the command to limit the results for easier readability.

```
df -h | grep backup
```

h. Change the ownership of the `/opt/backup` directory to the Workload Optimization Manager user.

This command requires root privileges.

```
sudo chown -R turbo.turbo /opt/backup
```

i. Exit `sudo`.

```
exit
```

3. (Optional) Test file permissions.
   a. Create a file in your backup directory.

```
touch /opt/backup/x
```

   b. Check that the file `x` exists and belongs to the Workload Optimization Manager user.

```
ls -al /opt/backup
```

   c. Delete the test file.

```
rm /opt/backup/x
```

Run the backup script

1. Change directories to the location of the backup script.

```
cd /opt/turbonomic/kubernetes/migration
```

2. Make the script executable.

```
chmod +x xl-backup.sh
```

3. Run the backup script and output the progress in a log file.

   First, run the Linux `screen` command, then run the backup script. For more information, see this website.

```
screen -S xl-backup
```

```
./xl-backup.sh | tee -a /<path>/<filename>.log
```

   If you want to watch the rolling log, open a new SSH session to the same terminal and run the command: `tail -f /<path>/<filename>.log`

   A file ending in `.tar.xz` (for example, `backup-2024-05-06.tar.xz`) is written to the `/opt/backup` directory. This backup file requires restoration on the new Workload Optimization Manager node.

4. Unmount the drive.
   - Detach the drive from the source node.

```
sudo umount /opt/backup
```

- Check to ensure that the drive is no longer mounted.

```
df -h | grep backup
```

5. Detach the drive in vCenter.

   Switch back over to vCenter and detach the drive that was unmounted from the file system.

   a. Right-click on your source virtual machine and select **Edit Settings**.
   b. Select the value to the right of **Total hard disks**.
   c. Hover over the rightmost side to display an **X** icon to detach the drive. Click the **X** icon.
   d. Expand the new hard disk to view the various aspects of the drive.

      NOTE:
      Ensure that the Disk File path and name are saved for later use. Do not select the option to delete the files from data store or else the drive and backup file contents are lost.

      Click **OK**.
   e. Power down the virtual machine once the drive is detached so that you can reuse the IP address on the destination virtual machine.
   f. Rename the virtual machine in case that it needs to be powered back on.

      You may need to power on the virtual machine if an issue occurs during the restoration process. If the virtual machine needs to be powered on, ensure to not power on the VM at the same time as the destination VM to avoid conflicting IP addresses.

Preparing the Rocky Linux VM image

1. Create a new VM within vCenter using the Workload Optimization Manager image.

   You should have this image ready locally to build the destination VM by using the same build as the source VM.

   a. Open an SSH terminal session to the new Rocky Linux VM image.
   b. Configure the Workload Optimization Manager user password.
   c. Configure the `sudo` password.

```
su -
```

   d. Run `ipsetup` to configure the IP address.

      Set the IP address to the original source VM network details that you previously captured, similar to the following example.

      - Hostname: `xlhost`
      - IP address: `10.16.173.8`
      - Subnet mask: `255.255.254.0`
      - Gateway: `10.16.172.0`
      - DNS1/DNS2: `10.16.150.2 10.16.150.3`
      - DNS domain: `demo.turbonomic.com`

   e. Restart the VM.
2. Change to the directory where the installer is stored.

```
cd /opt/local/bin
```

3. Run the Workload Optimization Manager installation.

```
t8cInstall.sh
```

4. Add the drive to the VM for the restoration destination.
    a. In vCenter, right-click on your origin VM and select **Edit Settings**.
    b. Click **Add New Device** and select **Existing Hard Disk** from the menu.
    c. Navigate to and select the directory that contains the hard disk path that you previously saved, then click **OK**.
    d. Verify that the path to the disk is correct, then click **OK**.
5. Mount the drive and set drive ownership.

    You must have both an administrator (or similar) account in Workload Optimization Manager and root access to the virtual machine by using an SSH terminal.
    a. Open an SSH terminal session to the Rocky VM image.

        Log in with the System Administrator that you set up when you installed Workload Optimization Manager.
    b. Create the `/opt/backup` directory, which is used as the mounted directory for your backup.

        ```
        mkdir /opt/backup
        ```
    c. Scan for the newly added drive to ensure that it is present.

        ```
        sudo su -

        echo "- - -" > /sys/class/scsi_host/host0/scan

        echo "- - -" > /sys/class/scsi_host/host1/scan

        echo "- - -" > /sys/class/scsi_host/host2/scan

        ls /dev/sdb
        ```
    d. Mount the drive that was attached in vCenter to the `/opt/backup` directory you created.

        ```
        mount /dev/sdb1 /opt/backup
        ```
    e. Display the current mounts to your device to ensure that the drive is properly mounted to your VM.

        You can add a `| grep` to the command to limit the results for easier readability.

        ```
        df -h | grep backup
        ```
    f. Change the ownership of the `/opt/backup` directory to the Workload Optimization Manager user.

        This command requires root privileges.

        ```
        chown -R turbo.turbo /opt/backup
        ```
    g. Exit `root`.

        ```
        exit
        ```
6. (Optional) Test file permissions.
    a. Create a file in your backup directory.

        ```
        touch /opt/backup/x
        ```
    b. Check that the file `x` exists and belongs to the Workload Optimization Manager user.

        ```
        ls -al /opt/backup
        ```

c. Delete the test file.

```
rm /opt/backup/x
```

Run the restore script

1. Change directories to the location of the backup file.

```
cd /opt/turbonomic/kubernetes/migration
```

2. Make the script executable.

```
chmod +x xl-restore.sh
```

3. Run the restore script and output the progress a log file.

   First, run the Linux `screen` command, then run the restore script.

```
screen -S xl-restore
```

```
./xl-restore.sh | tee -a /<path>/<filename>.log
```

   If you want to watch the rolling log, open a new SSH session to the same terminal and run the command: `tail -f /<path>/<filename>.log`

   A file ending in `*.tar.xz` (for example, `restore-2024-05-06.tar.xz`) is written to the `/opt/backup` directory.

4. Verify that the restore worked as expected.
   a. Check the status of the pods coming up.

```
watch kubectl get pods
```

      Wait for the pods to come into a ready state.

   b. Log in to the UI and validate targets, virtual machines, policies, historical data, and so on to ensure that the wanted data is present. The abstraction initially appears all green, which is normal behavior.

   c. Let Workload Optimization Manager run through a few market cycles to catch up and start displaying actions and the health indicators.

5. (Optional) Unmount the drive.
   - Detach the drive from the source node.

```
sudo umount /opt/backup
```

   - Check to ensure that the drive is no longer mounted.

```
df -h | grep backup
```

6. Detach the drive in vCenter.

   Switch back over to vCenter and detach the drive that was unmounted from the file system.
   a. Right-click on your source virtual machine and select **Edit Settings**.

   b. Select the value to the right of **Total hard disks**.

   c. Hover over the rightmost side to display an **X** icon to detach the drive. Click the **X** icon.

   d. Expand the new hard disk to view the various aspects of the drive.

NOTE:
Ensure that the Disk File path and name are saved for later use. Do not select the option to delete the files from data store or else the drive and backup file contents are lost.

Click **OK**.

Reference: Working with YAML files

YAML is the primary file format to create and configure resources on kubernetes, including everything to do with the Workload Optimization Manager platform. The Custom Resource YAML provides a convenient, single place to define the majority of configuration details for Workload Optimization Manager. When editing YAML files, you must be careful to respect the file syntax, including indents in the file. General rules for edits include:

- Always uses spaces, not tabs, for all indentation. Since spacing and indentations matter, and can yield an invalid result or a parameter completely skipped over, you should work with YAML files with an editor that supports using vertical lines associated with indentations to visually spot a misalignment. If your editor of choice makes this difficult, you can use the linux `expand` utility when you're done, to convert tabs to equivalent spaces.
- Indention uses two spaces per level.
- Be careful to keep the same indentation for all properties in a given section.
- Never use the same property name twice in the same section. Doing this will render the YAML file invalid, though you will likely not see any notification of a problem. Rather, all but one of the property definitions will be silently ignored.

Spacing matters

For the Workload Optimization Manager Custom Resource, indentation defines where parameters are applied (globally or to specific components), so ensure you line up the text appropriately. The following example shows examples of specifications that are applied at a global level (`{“spec”:{“global”:[{“tag”:”8.6.4”}]}}`) which sets the container image tag for all instances. Then indented you see properties that are global for the remote database (`{“spec”:{“properties”:{"global":[{“dbPort”:”6033”}]}}}`) describes a property of `dbPort` that would be set for remote DB connections.

Each line is indented two spaces from the higher level.

```
spec:
  global:
    repository: turbonomic
    tag: 3.13.0
  properties:
    global:
      dbPort: 6033
  kubeturbo:
    enabled: true
  aws:
    enabled: true
```

Combine properties correctly

A YAML file is read beginning to end, and if there are different parameters that apply to the same component, they need to be combined. The following example shows a YAML where properties for the `ui` component of image tag *and* memory limit resources.

```
spec:
  global:
    repository: turbonomic
    tag: 3.13.0
  ui:
```

```
    image:
      tag: 3.13.0
  properties:
    global:
      dbPort: 6033
  kubeturbo:
    enabled: true
  aws:
    enabled: true
  ui:
    resources:
      limits:
        memory: 4Gi
```

This YAML will not set both image tag and memory limit resources for the `ui` component because they are set in two different sections. The memory limits would be applied since that is the last section to be read, and it would overwrite the first set of properties with the image tag. The following YAML will set both image tag and memory limit resources for the `ui` component:

```
spec:
  global:
    repository: turbonomic
    tag: 3.13.0
  ui:
    image:
      tag: 3.13.0
    resources:
      limits:
        memory: 4Gi
  properties:
    global:
      dbPort: 6033
  kubeturbo:
    enabled: true
  aws:
    enabled: true
```

To avoid having the same property name twice in the same section, search for an existing instance of the key you want to add. With the complex embedded structure in many k8s resources, this can be quite challenging; however, you can use an online validator that does not accept duplicate keys to check your entire CR file (such as YAMLint – The YAML Validator).

Finding a property

This documentation refers to specific properties in the CR file using a "path" expression. For example, assume the path `/spec/global/repository` to designate a `repository` property in the file. You can find the property in the file as follows:

1. Find a line that says `spec:` with no indentation at all.
2. Between that line and the next unindented line (not counting comments, which start with #), find a line that says `global:` and is at the next level of indentation.
3. Between that line and the next line with the same indentation, find a line that starts with `repository:`. That line is where the addressed property is defined.

In this example, find the `repository` property specified by `/spec/global/repository`:

```
apiVersion: charts.helm.k8s.io/v1alpha1
kind: Xl
metadata:
  name: xl-release
spec:
  properties:
    global:
      repository:        # This is NOT the correct property
        ...
```

```
# Global settings
global:
  repository:          # This is the one you want
```

The first `repository` property is not at `/spec/global/repository`, but at `/spec/properties/global/repository`.

Reference: Image listing

Workload Optimization Manager images are located at https://icr.io/cpopen/turbonomic, except for the operators which are located at https://icr.io/cpopen.

> **NOTE:**
> Starting with Workload Optimization Manager version 3.5.5, IBM Container Registry is used for all Workload Optimization Manager images for online upgrades and installations. Ensure that you have access to `https://*.icr.io` before continuing.

All of the required Workload Optimization Manager application images are listed in the following table. The total size of the required images is approximately 12GB. Adding additional components, such as probes or the secure client, increases the total size. Pull the image with the required `tag`, which is the version to deploy.

Image name	Component type	Required	Comment
`action-orchestrator`	Server	Required	Core (default) component
`api`	Server	Required	Core (default) component
`auth`	Server	Required	Core (default) component
`cloud-optimizer`	Server	Required	Core (default) component
`clustermgr`	Server	Required	Core (default) component
`consul`	Server	Required	Core (default) component
`cost`	Server	Required	Core (default) component
`group`	Server	Required	Core (default) component
`history`	Server	Required	Core (default) component
`kafka`	Server	Required	Core (default) component
`kube-state-metrics`	Server	Required	Used by `prometheus-kube-state-metrics`, which is a core component
`market`	Server	Required	Core (default) component
`metadata`	Server	Required	Used by metadata, which is a core component
`nginx`	Server	Required	Core (default) component
`plan-orchestrator`	Server	Required	Core (default) component
`prometheus`	Server	Required	Used by `prometheus-server`, which is a core component

Image name	Component type	Required	Comment
prometheus-config-manager	Server	Required	Used by prometheus-server, which is a core component
prometheus-kafka-adapter	Server	Required	Used by prometheus-server, which is a core component
redis	Server	Required	Used by suspend, which is a core component
repository	Server	Required	Core (default) component
rsyslog	Server	Required	Core (default) component
server-power-modeler	Server	Required	Core (default) component
suspend	Server	Required	Core (default) component
sustainability	Server	Optional	Enabled by default; used to calculate energy usage and carbon footprint of public cloud VMs
t8c-operator	Server	Required	See Workload Optimization Manager updates and operator version *(on page 32)*
topology-processor	Server	Required	Core (default) component
ui	Server	Required	Core (default) component
zookeeper	Server	Required	Core (default) component, used by Kafka
db	Server	Optional	Used for containerized database
kinesis-kafka-connect	Server	Optional	Used by the Telemetry component
telemetry-scrubber	Server	Optional	Used by Prometheus-server when telemetry is enabled
extractor	Embedded Reporting	Optional	Used for embedded reporting, SaaS Reporting, and telemetry
grafana	Embedded Reporting	Optional	Used for embedded reporting
timescaledb	Embedded Reporting	Optional	
kubeturbo	Probe (Kubeturbo)	Optional	Used to support all Kubernetes related targets
kubeturbo-operator	Probe (Kubeturbo)	Optional	
cpufreqgetter	Probe (Kubeturbo)	Optional	
mediation-action-script	Probe	Optional	

Image name	Component type	Required	Comment
mediation-action-stream-kafka	Probe	Optional	
mediation-appdynamics	Probe	Optional	
mediation-aws	Probe	Optional	
mediation-awscloud-billing	Probe	Optional	
mediation-awscost	Probe	Optional	
mediation-awsorganization	Probe	Optional	
mediation-azure	Probe	Optional	
mediation-azure-billing	Probe	Optional	
mediation-azurecost	Probe	Optional	
mediation-azureea	Probe	Optional	
mediation-azureinfra	Probe	Optional	
mediation-azurepricing	Probe	Optional	
mediation-azuresp	Probe	Optional	
mediation-azurevolumes	Probe	Optional	
mediation-datadog	Probe	Optional	
mediation-dynatrace	Probe	Optional	
mediation-flexera	Probe	Optional	
mediation-gcpbilling	Probe	Optional	
mediation-gcpcost	Probe	Optional	
mediation-gcpinfra	Probe	Optional	
mediation-gcpproject	Probe	Optional	
mediation-gcpsa	Probe	Optional	
mediation-horizon	Probe	Optional	
mediation-hpe3par	Probe	Optional	
mediation-hyperflex	Probe	Optional	
mediation-hyperv	Probe	Optional	
mediation-ibmstorage-flashsystem	Probe	Optional	
mediation-instana	Probe	Optional	
mediation-jboss	Probe	Optional	

Image name	Component type	Required	Comment
mediation-jvm	Probe	Optional	
mediation-mssql	Probe	Optional	
mediation-netapp	Probe	Optional	
mediation-newrelic	Probe	Optional	
mediation-nutanix	Probe	Optional	
mediation-oneview	Probe	Optional	
mediation-oracle	Probe	Optional	
mediation-powervm	Probe	Optional	
mediation-pure	Probe	Optional	
mediation-scaleio	Probe	Optional	
mediation-servicenow	Probe	Optional	
mediation-snmp	Probe	Optional	
mediation-tomcat	Probe	Optional	
mediation-ucs	Probe	Optional	
mediation-udt	Probe	Optional	Enabled by default; used for the creation and discovery of application map entities
mediation-vcenter	Probe	Optional	
mediation-vcenter-browsing	Probe	Optional	
mediation-vmax	Probe	Optional	
mediation-vmm	Probe	Optional	
mediation-vplex	Probe	Optional	
mediation-webhook	Probe	Optional	
mediation-weblogic	Probe	Optional	
mediation-websphere	Probe	Optional	
mediation-wmi	Probe	Optional	
mediation-xtremio	Probe	Optional	
prometurbo	Probe	Optional	
prometurbo-operator	Probe	Optional	
prometheus-server	Probe	Optional	
turbodif	Probe	Optional	
winrm-proxy	Probe	Optional	
hydra	Secure client	Optional	

Image name	Component type	Required	Comment
kibitzer	Secure client	Optional	Used to help in retrieving debug information for components such as history
rsyslog-courier	Secure client	Optional	
skupper-config-sync	Secure client	Optional	
skupper-router	Secure client	Optional	
skupper-service-controller	Secure client	Optional	
skupper-site-controller	Secure client	Optional	
t8c-client-operator	Secure client	Optional	
t8c-client-operator-bundle	Secure client	Optional	
t8c-client-operator-catalog	Secure client	Optional	
tsc-site-resources	Secure client	Optional	

NOTE:

- Required server images are the minimal images that are needed to run the Workload Optimization Manager Server.
- Reporting requires `postgresql` and `timescale db`, which are pulled from other repositories. See this chart for image details.
- Kubeturbo, which should be deployed separately, provides robust container platform monitoring and management, including self-optimization of Workload Optimization Manager. For more information, see this topic *(on page 234)*.
- If you want a copy of a sample deployment yaml file you will need access to GitHub to copy or clone the `t8c-install` project.

Updating Workload Optimization Manager

When a new version is available, it is important to properly update your existing installed instance. When you first installed Workload Optimization Manager, you put into place sophisticated data collection and analysis processes, and your database retains performance data from across your virtual environment. Workload Optimization Manager uses this historical data for optimization, projecting trends, and other analysis. This means that the database is important to Workload Optimization Manager and becomes more so over time. Properly updating your installation of Workload Optimization Manager preserves the database for continued use.

OVA updates

Before you begin the update procedure:

- Review What's New *(on page 23)* and Release Notes *(on page 23)* to see what is new for this release.

  NOTE:
  As Workload Optimization Manager evolves, the set of delivered probes change. Also, from one version to the next, the set of probes that are enabled by default can change. When you update to a new version, the update does not change your probe configuration. An update to a newer version does not automatically enable any new probes in your deployment. If you want to take advantage of new probes in an update, then you must enable them manually.

- Make sure you have the email that Cisco sent to you with links to the Workload Optimization Manager 0VA file and to the ISO image.
- For on-prem installations, make sure that the physical machine hosting the VM meets the minimum requirements (see Minimum Requirements *(on page 49)*).

  NOTE:
  Before updating Workload Optimization Manager, you must extend the `/var` partition to at least 340 GB. If the `/var` partition requirement is not met, your upgrade may fail, new container images may not load properly, and you can run into an ImagePullBackOff issue for some of the images. Before upgrading, be sure that the partition has sufficient disk space. If you need to free up disk space to increase the size of the partition, you can delete old images that are not in use. For details, see Increasing Available Disk Space *(on page 74)*.

- Ensure you are running the correct version of the historical database.

  For its historical database, Workload Optimization Manager supports long-term maintenance stable series of MariaDB (the minimum supported version is 10.6) and MySQL 8.0.x. This support includes comprehensive testing and quality control for Workload Optimization Manager usage.

  For more information, see Verifying your MariaDB Version *(on page 59)*.
- Run the `upgrade-precheck.sh` script.

This script ensures that your current installation of Workload Optimization Manager is ready to update and it must be run before going on to update your installation (see Checking Before Updating *(on page 133)*).

- Execute an offline update, via a downloaded ISO image. For more information, see Offline Update *(on page 137)*.

Checking before updating

Before you perform an update of your Workload Optimization Manager instance, you should run the `upgrade-precheck.sh` script. This script inspects your installation to check for the following:

- Free disk space is sufficient.

  For example, the `/var` partition must be at least 340 GB in size. This is the recommended size for deployments starting in version 3.7.1. If you need to free up disk space to increase the size of the partition, you can delete old images that are not in use. For more information, see Increasing Available Disk Space *(on page 74)*.
- (For online updates) The required endpoints can be accessed.
- The MariaDB service is running.

  This check is for the default installation of the MariaDB service only. The script does not check an external installation of MySQL or MariaDB. If you deployed an external historical database, the script indicates that your MariaDB service is not running. This is a normal result.
- The Kubernetes service is running.
- The necessary Kubernetes certificates are valid.

  If the certificates are not valid, you can correct the issue by running a script located on your installation.
  - For VM images based on Rocky Linux, use `/opt/local/bin/k8s-certs-renew.sh`.
  - For VM images based on CentOS Linux, use `/opt/local/bin/kubeNodeCertUpdate.sh`.

    IMPORTANT:
    - The `k8s-certs-renew.sh` script is not designed for CentOS Linux distributions and might cause serious errors if it is run on CentOS VMs. If you are running CentOS Linux, you must run the `kubeNodeCertUpdate.sh` script, regardless of your Workload Optimization Manager version level.
    - If you are running CentOS Linux with Workload Optimization Manager version 8.12.2 or later, the `kubeNodeCertUpdate.sh` script is in a versioned `bin-*` folder in the `/opt/local/` directory. The version of the `bin-*` folder corresponds to the last version you ran before upgrading to 8.12.2 or later. For example, if you upgraded from 8.12.0 to 8.13.0, the script is located in the `/opt/local/bin-8.12.0/` directory.
    - Before you run either script, save a snapshot of your current Workload Optimization Manager VM. This provides a reliable restore point you can turn to in the event that trouble occurs during the update.

      Before updating, shut down (not power off) the Workload Optimization Manager VM.

      ```
      sudo init 0
      ```

      Then, perform a snapshot (or clone the VM). After you have the snapshot, bring the VM back online and run the appropriate script for your Linux distribution.

- Root password is not set to expire.
- Time sync is enabled and current (if running).
- All Workload Optimization Manager pods are running.

To run this script:

1. Open an SSH terminal session to your Workload Optimization Manager instance.

   Log in with the System Administrator that you set up when you installed Workload Optimization Manager:

   - Username: `turbo`
   - Password: `[your_private_password]`

2. Change to the scripts directory.

   ```
   cd /opt/local/bin
   ```

3. Download the latest version of the script.

   a. Navigate to the following Cisco web page:

   https://software.cisco.com/download/home/286328879/type.

   b. Under **Select a Software Type**, click **Workload Optimization Manager**.

   c. From the menu, select the desired Workload Optimization Manager version.

   d. Click the download button for the following file:

   ```
   upgrade-precheck-X.X.X.zip
   ```

   e. When prompted, login using your Cisco account.

   f. After the download completes, unzip the downloaded file.

4. Make the script executable.

   ```
   chmod +x upgrade-precheck.sh
   ```

5. Run the script.

   ```
   ./upgrade-precheck.sh
   ```

   As the script executes, it identifies any issues that you should address before you run an update.

External databases and Workload Optimization Manager updates

If you deployed Workload Optimization Manager with an external database server, for some updates you might need to manually create a new database and user for that deployment. This is important if your external database server is multi-tenant, or if your deployment does not grant administrative privileges to Workload Optimization Manager.

> **NOTE:**
> If your external database server is multi-tenant, or if your database server does not grant administrative privileges to Workload Optimization Manager, then you must continue with this configuration requirement.
>
> Azure database services are multi-tenant. If you deployed an external database on Azure, this configuration requirement applies to you.
>
> If you deployed your database server in a way that grants Workload Optimization Manager privileges to create new databases and new users, then a product update automatically creates the database. This configuration requirement does not apply to you and you do not need to create the database.

For some Workload Optimization Manager updates, the updated version includes new databases on the historical database server. If you are updating to one of these versions, then you must create the new database and a user account with privileges to access that database.

New databases

This table lists the Workload Optimization Manager versions that require new databases. If you are updating from an earlier version, you must create the indicated new databases. For example, if you are updating from version 3.0.1 to 3.0.5, then you must create the `api` database.

Workload Optimization Manager Version	New databases	Notes
3.0.5	`api`	If you are updating from a version earlier than 3.0.5, you must create a new database that is named `api`, and a user account named `api`.

Creating new databases

To create the databases and users:

- Manually create each required database.

  Create the database in your DB instance, create a user to access the database, and grant privileges to the user.

- Manually add each required database to your custom resource (CR) yaml file.

  The CR yaml file declares entries for each component database. Each entry names the component, and gives the user and password that the component can use to access that database. Add an entry for each new database.

To create a new database:

1. Connect to your external DB with a global account.

   The account must have privileges to create databases and users. If you have specified `dbRootUsername` in the CR yaml file, you can use that account.

2. Create the database, where *<New_Database>* matches the database name in the previous table:

   ```
   create database <New_Database>;
   ```

   For example, to create a new `api` database, run:

   ```
   create database api;
   ```

3. Create the account that Workload Optimization Manager uses to access the database where *<New_Database>* matches the database name in the previous table:

   ```
   create user '<New_Database>'@'%' identified by 'vmturbo';
   ```

   For example, to create a user for the `api` database, run:

   ```
   create user 'api'@'%' identified by 'vmturbo';
   ```

   NOTE:
   The value `vmturbo` is the default password that Workload Optimization Manager uses for all component database accounts. If you manually created accounts with different credentials, you can do so for this database as well.

4. Set the user account privileges for the new user account, where <*New_Database*> matches the database name in the previous table:

```
GRANT SELECT, INSERT, UPDATE, DELETE, CREATE, DROP, REFERENCES, INDEX, ALTER, CREATE TEMPORARY
 TABLES, LOCK TABLES, EXECUTE, CREATE VIEW, SHOW VIEW, CREATE ROUTINE, ALTER ROUTINE, EVENT,
 TRIGGER ON <New_Database>.* TO '<New_Database>'@'%';
```

   For example, to set account privileges for the `api` user, run:

```
GRANT SELECT, INSERT, UPDATE, DELETE, CREATE, DROP, REFERENCES, INDEX, ALTER, CREATE TEMPORARY
 TABLES, LOCK TABLES, EXECUTE, CREATE VIEW, SHOW VIEW, CREATE ROUTINE, ALTER ROUTINE, EVENT,
 TRIGGER ON api.* TO 'api'@'%';
```

5. Flush privileges to make the privileges take effect.

```
flush privileges;
```

Now that the new database is created in your external DB service, you must declare access to it in the Workload Optimization Manager CR yaml resource.

1. Open the CR file for editing.

   The location of the file depends on the type of Workload Optimization Manager installation that you are configuring.

   - VM Image installation of Workload Optimization Manager

     Log in to an SSH terminal session of your Workload Optimization Manager instance as the System Administrator that you set up when you installed Workload Optimization Manager.

     ```
     /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
     ```

   - Workload Optimization Manager on a Kubernetes node or node cluster

     ```
     deploy/crds/charts_v1alpha1_xl_cr.yaml
     ```

2. Add the credentials for the matching pod to access the new database.

   Add the entry to the `properties:` section of the CR yaml file, where <*vmturbo*> is the user account password, <*yourDB*> is the qualified name of your external DB or your multi-tenant DB partition, and <*New_Database*> is the name of your new database.

```
  <New_Database>:
    <New_Database>DbUsername: <New_Database>@<yourDB>
    <New_Database>DbPassword: <vmturbo>
```

   For example, if you added the `api` database, the resulting CR yaml file is similar to the following example:

```
properties:
    global:
      enableSecureDBConnection: true
      sqlDialect: MYSQL
      dbRootPassword: yourAdminPassword
      dbRootUsername: xladmin@yourDB
      #dbUserPassword:
      #dbUsername:
    action-orchestrator:
      actionDbUsername: action@yourDB
      actionDbPassword: yourPassword
    auth:
      authDbUsername: auth@yourDB
      authDbPassword: yourPassword
    clustermgr:
      clustermgrDbUsername: clustermgr@yourDB
      clustermgrDbPassword: yourPassword
    cost:
      costDbUsername: cost@yourDB
```

```
    costDbPassword: yourPassword
  group:
    groupComponentDbUsername: group_component@yourDB
    groupComponentDbPassword: yourPassword
  history:
    historyDbUsername: history@yourDB
    historyDbPassword: yourPassword
  plan-orchestrator:
    planDbUsername: plan@yourDB
    planDbPassword: yourPassword
  topology-processor:
    topologyProcessorDbUsername: topology_processor@yourDB
    topologyProcessorDbPassword: yourPassword
  repository:
    repositoryDbUsername: repository@yourDB
    repositoryDbPassword: yourPassword
  market:
    marketDbUsername: market@yourDB
    marketDbPassword: yourPassword
  api:
    apiDbUsername: api@yourDB
    apiDbPassword: yourPassword
  actions-history:
    ahDbUsername: ah@yourDB
    ahDbPassword: yourPassword
  oauth2:
    oauth2DbUsername: oauth2@yourDB
    oauth2DbPassword: yourPassword
  sustainability:
    sustainabilityDbUsername: sustainability@yourDB
    sustainabilityDbPassword: yourPassword
  api:
    apiDbUsername: api@yourDB
    apiDbPassword: yourPassword
```

After you add the database, you can update to the latest version of Workload Optimization Manager.

NOTE:
Upgrading applies changes to the version information in this file.

Offline update

If you do not allow internet access, then you must complete offline updates of Workload Optimization Manager. When you complete the update, the system automatically unmounts the ISO image. To update MariaDB, your Workload Optimization Manager instance must be mounted on the same ISO image that you used to update it to version 8.3.5 or later.

Performing an offline update

1. Run the `upgrade-precheck.sh` script.

   This script ensures that your current installation of Workload Optimization Manager is ready to update and it must be run before going on to update your installation (see Checking Before Updating *(on page 133)*).

2. Save a snapshot of your current Workload Optimization Manager VM.

   Before updating, shut down (not power off) the Workload Optimization Manager VM.

   ```
   sudo init 0
   ```

   Then, perform a snapshot (or clone the VM). This provides a reliable restore point you can turn to in the event that trouble occurs during the update. After you have the snapshot, bring the VM back online.

3. Optionally, enable new probes in your environment.

   NOTE:
   As Workload Optimization Manager evolves, the set of delivered probes change. Also, from one version to the next, the set of probes that are enabled by default can change. When you update to a new version, the update does not change your probe configuration. An update to a newer version does not automatically enable any new probes in your deployment. If you want to take advantage of new probes in an update, then you must enable them manually.

   For steps to enable new probes in your updated version, see Enabling and Disabling Probe Components *(on page 95)*. Use these steps to edit the platform's cr.yaml file.

   NOTE:
   During the online or offline upgrade process, you should not apply these changes at this time.

4. Download the ISO image.
   a. Navigate to the following Cisco web page:

      https://software.cisco.com/download/home/286328879/type.
   b. Under **Select a Software Type**, click **Workload Optimization Manager**.
   c. From the menu, select the desired Workload Optimization Manager version.
   d. Click the download button for `update64_package-X.X.X.iso`.
   e. When prompted, login using your Cisco account.
5. Save the ISO image to a location that is available to the VM that runs Workload Optimization Manager. Then mount the image as a CD drive.

   For example, if you run the Workload Optimization Manager VM in vCenter Server do the following:
   a. In vCenter, navigate to the Workload Optimization Manager VM.
   b. Right-click the VM and choose **Edit Settings**.
   c. In the CD/DVD Drive drop-down menu, select **Datastore ISO**, then browse to the Workload Optimization Manager update ISO image and choose it.
   d. Ensure that **Connect at power on** is selected.
6. Open an SSH terminal session to your Workload Optimization Manager instance.

   Log in with the System Administrator that you set up when you installed Workload Optimization Manager:
   - Username: `turbo`
   - Password: `[your_private_password]`
7. Get the `isoUpdate.sh` script for your update version.
   a. Navigate to the following Cisco web page:

      https://software.cisco.com/download/home/286328879/type.
   b. Under **Select a Software Type**, click **Workload Optimization Manager**.
   c. From the menu, select the desired Workload Optimization Manager version.
   d. Click the download button for `isoUpdate-X.X.X.zip`.
   e. When prompted, login using your Cisco account.
   f. After the download completes, unzip the downloaded file.
8. Upload the script to your Workload Optimization Manager instance.

   Initiate a file transfer from your local machine to the Workload Optimization Manager server. Save the script to `/opt/local/bin/` on the VM that runs Workload Optimization Manager.

9. Make the script executable.

```
chmod +x /opt/local/bin/isoUpdate.sh
```

10. Run the offline installation.

```
/opt/local/bin/isoUpdate.sh
```

As the script runs, it:

- Backs up the old scripts in your installation.
- Updates the configuration and code assets in your installation.
- Updates the platform to the new version.
- Updates custom resources.
- Updates the MariaDB configuration (but this does not update the MariaDB version).
- If you have enabled embedded reports or Data Export, installs the embedded reports and data export database (Postgres and TimescaleDB).
- Scales down the t8c-operator and the Workload Optimization Manager components.
- Executes the final updates for this version.
- Scales up the t8c-operator, which then restarts the Workload Optimization Manager components.

After the script is finished updating your platform, give it enough time for all the components to start up again.

11. Verify that Workload Optimization Manager installed correctly.

```
kubectl get pods -n turbonomic
```

After all pods start, the READY column shows `1/1`, `2/2`, and so on. The STATUS column for each pod shows `Running`.

The output is similar to the following example:

```
NAME                                          READY   STATUS    RESTARTS
action-orchestrator-b6454c9c8-mfl85           1/1     Running   0
api-7887c66f4b-shndq                          1/1     Running   0
auth-5b86976bc8-vxwz4                         1/1     Running   0
clustermgr-85548678d9-r5wb8                   1/1     Running   0
...
```

12. Verify that you are running the correct version of MariaDB.

For this version of the product, Workload Optimization Manager supports MariaDB, version 10.6.19. Even after updating to this Workload Optimization Manager version, it is possible that your installation is running an earlier version of MariaDB.

a. While still in the SSH session, check the MariaDB version.

```
mysql -V
```

The output is similar to the following example:

```
mysql  Ver 15.1 Distrib 10.6.19-MariaDB, for Linux (x86_64) using readline 5.1
```

b. If the MariaDB version is earlier than 10.6.19, update your MariaDB.

Before you run the script, you need to know the MariaDB password. By default, this password is `vmturbo`.

```
sudo /opt/local/bin/mariadbUpgrade.sh |& tee -a $HOME/Turbonomic_MariaDB_upgrade.log
```

The script updates the version of MariaDB. It also increases size limits for the allowed packets, and buffer and log sizes for the innodb. The script output includes the following (where `Total Memory` and `buffer pool size` can vary depending on your VM configuration):

```
==============================
Update the mariadb configuration
==============================
Total Memory: 128773 MB
Changing Innodb buffer pool size to: 9216 MB
Changing max allowed packets to: 1G
Changing innodb log file size to: 10G
==============================
```

c. When the script completes, verify that the correct version of MariaDB is running.

```
mysql -V
```

13. Scale up the Workload Optimization Manager platform's pods.

    To update the database, the script scales down your platform pods. When it completes, the script displays the following prompt:

```
################################################################################
When confirmed the mariadb has been upgraded and is properly working, run:
kubectl scale deployment --replicas=1 t8c-operator -n turbonomic
################################################################################
```

    After you verify that the correct version of MariaDB is running, scale up the platform.

```
kubectl scale deployment --replicas=1 t8c-operator -n turbonomic
```

14. Clear your browser data and refresh your browser.

    After clearing the browser data and refreshing your browser, you have full access to Workload Optimization Manager features. However, features that rely on current analysis data will not be available until after a full market cycle — usually 10 minutes. For example, the Pending Actions charts will not show any actions until after a full market cycle.

15. Notify other users to clear their browser data and refresh their Workload Optimization Manager browser sessions.

Reference: Step-wise offline update

NOTE:
The process that is described here is applicable only to the OVA deployment model.

Complete the following steps to perform a stepwise offline update of your Workload Optimization Manager installation.

1. Save a snapshot of your current Workload Optimization Manager VM.

   Before updating, shut down (not power off) the Workload Optimization Manager VM.

```
sudo init 0
```

   Then, perform a snapshot (or clone the VM). This provides a reliable restore point you can turn to in the event that trouble occurs during the update. After you have the snapshot, bring the VM back online.

2. Download and attach the ISO image to the VM that runs Workload Optimization Manager.

   Refer to the email you received from Cisco for links to the Workload Optimization Manager 0VA file and to the ISO image.

3. Mount the ISO image by logging in to vCenter.
   a. In vCenter, navigate to the Workload Optimization Manager VM.
   b. Right-click the VM and choose **Edit Settings**.

c. In the CD/DVD Drive drop-down menu, select **Datastore ISO**, then browse to the Workload Optimization Manager update ISO image and choose it.

d. Ensure that **Connect at power on** is selected.

4. Open an SSH terminal session to your Workload Optimization Manager instance.

   Log in with the System Administrator that you set up when you installed Workload Optimization Manager:

   - Username: `turbo`
   - Password: `[your_private_password]`

5. If you are prompted to override the CD-ROM lock, select **Yes** and click **OK** to disconnect anyway and override the lock.

6. Mount the ISO image.

```
sudo mount /dev/cdrom /mnt/iso
```

7. Verify the correct version of the ISO image is mounted.

```
ls /mnt/iso
```

   Verify that the ISO image contains the correct version for your update.

8. Load the latest images.

```
sudo /mnt/iso/turboload.sh
```

   This script loads all the images to the Workload Optimization Manager instance.

   The output is similar to the following example:

```
The t8c upgrade iso has been mounted
Image check:
============
**************************
All images have been loaded
**************************
```

   If the load does not succeed, the script lists any images that did not load, along with instructions to load them manually.

9. Run the following command as a single line command to update Workload Optimization Manager.

```
/mnt/iso/turboupgrade.sh | tee \
-a $HOME/Turbonomic_MariaDB_upgrade.log
```

   Wait until the script is finished.

10. Verify that you are running the correct version of MariaDB.

    For this version of the product, Workload Optimization Manager supports MariaDB, version 10.6.19. Even after updating to this Workload Optimization Manager version, it is possible that your installation is running an earlier version of MariaDB.

    a. While still in the SSH session, check the MariaDB version.

```
mysql -V
```

       The output is similar to the following example:

```
mysql  Ver 15.1 Distrib 10.6.19-MariaDB, for Linux (x86_64) using readline 5.1
```

    b. If the MariaDB version is earlier than 10.6.19, update your MariaDB.

       Before you run the script, you need to know the MariaDB password. By default, this password is `vmturbo`.

```
sudo /opt/local/bin/mariadbUpgrade.sh |& tee -a $HOME/Turbonomic_MariaDB_upgrade.log
```

The script updates the version of MariaDB. It also increases size limits for the allowed packets, and buffer and log sizes for the innodb. The script output includes the following (where `Total Memory` and `buffer pool size` can vary depending on your VM configuration):

```
================================
Update the mariadb configuration
================================
Total Memory: 128773 MB
Changing Innodb buffer pool size to: 9216 MB
Changing max allowed packets to: 1G
Changing innodb log file size to: 10G
================================
```

c. When the script completes, verify that the correct version of MariaDB is running.

```
mysql -V
```

11. Unmount the ISO image.

```
sudo umount /dev/cdrom
```

12. Optionally, enable new probes in your environment.

NOTE:
As Workload Optimization Manager evolves, the set of delivered probes change. Also, from one version to the next, the set of probes that are enabled by default can change. When you update to a new version, the update does not change your probe configuration. An update to a newer version does not automatically enable any new probes in your deployment. If you want to take advantage of new probes in an update, then you must enable them manually.

For steps to enable new probes in your updated version, see Enabling and Disabling Probe Components *(on page 95)*. Use these steps to edit the platform's cr.yaml file.

NOTE:
During the online or offline upgrade process, you should not apply these changes at this time.

13. Verify that Workload Optimization Manager installed correctly.

```
kubectl get pods -n turbonomic
```

After all pods start, the READY column shows `1/1`, `2/2`, and so on. The STATUS column for each pod shows `Running`.

The output is similar to the following example:

```
NAME                                           READY   STATUS    RESTARTS
action-orchestrator-b6454c9c8-mfl85            1/1     Running   0
api-7887c66f4b-shndq                           1/1     Running   0
auth-5b86976bc8-vxwz4                          1/1     Running   0
clustermgr-85548678d9-r5wb8                    1/1     Running   0
...
```

14. Clear your browser data and refresh your browser.

After clearing the browser data and refreshing your browser, you have full access to Workload Optimization Manager features. However, features that rely on current analysis data will not be available until after a full market cycle — usually 10 minutes. For example, the Pending Actions charts will not show any actions until after a full market cycle.

15. Notify other users to clear their browser data and refresh their Workload Optimization Manager browser sessions.

Product overview

Workload Optimization Manager is the premier solution for Application Resource Management (ARM), a hierarchical, application-driven approach that continuously analyzes applications' resource needs and generates fully automatable actions to ensure applications always get what they need to perform. It runs 24/7/365 and scales with the largest, most complex environments.

To perform Application Resource Management, Workload Optimization Manager represents your environment holistically as a *supply chain* of resource *buyers* and *sellers*, all working together to meet application demand. By empowering buyers (such as VMs) with a budget to seek the resources that applications need to perform, and sellers (such as hosts) to price their available CPU, memory, storage, and other resources based on utilization in real time, Workload Optimization Manager keeps your applications in an optimal state.

Workload Optimization Manager is a microservices architected platform that runs in your network or in the public cloud. It discovers and monitors your application environment through targets. It then performs analysis, anticipates risks to performance or efficiency, and recommends actions to avoid problems before they occur.

How Workload Optimization Manager works

To keep your infrastructure in the desired state, Workload Optimization Manager performs Application Resource Management. This is an ongoing process that solves the problem of assuring application performance while simultaneously achieving the most efficient use of resources and respecting environment constraints to comply to business rules.

This is not a simple problem to solve. Application Resource Management has to consider many different resources and how they are used in relation to each other, and numerous control points for each resource. As you grow your infrastructure, the factors for each decision increase exponentially. Moreover, the environment is constantly changing — to stay in the desired state, you are constantly trying to hit a moving target.

To perform Application Resource Management, Workload Optimization Manager models the environment as a *market* made up of *buyers* and *sellers*. These buyers and sellers make up a *supply chain* that represents tiers of entities in your inventory. This supply chain represents the flow of resources from the datacenter, through the physical tiers of your environment, into the virtual tier and out to the cloud. By managing relationships between these buyers and sellers, Workload Optimization Manager provides closed-loop management of resources, from the datacenter, through to the application.

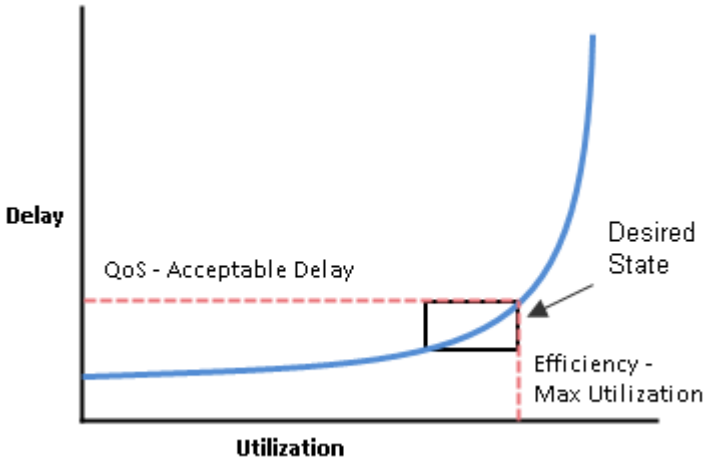
See Supply Chain of Entities *(on page 159)* for a visual layout of the buyer and seller relationships.

Workload Optimization Manager uses *Virtual Currency* to give a budget to buyers and assign cost to resources. This virtual currency assigns value across all tiers of your environment, making it possible to compare the cost of application transactions with the cost of space on a disk or physical space in a data center.

The price that a seller charges for a resource changes according to the seller's supply. As demand increases, prices increase. As prices change, buyers and sellers react. Buyers are free to look for other sellers that offer a better price, and sellers can duplicate themselves (open new storefronts) to meet increasing demand. Workload Optimization Manager uses its *Economic Scheduling Engine* to analyze the market and make these decisions. The effect is an invisible hand that dynamically guides your IT infrastructure to the optimal use of resources.

To get the most out of Workload Optimization Manager, you should understand how it models your environment, the kind of analysis it performs, and the desired state it works to achieve.

The desired state



The goal of Application Resource Management is to assure performance while maintaining efficient use of resources. When performance and efficiency are both maintained, the environment is in the desired state. You can measure performance as a function of delay, where zero delay gives the ideal QoS for a given service. Efficient use of resources is a function of utilization where 100% utilization of a resource is the ideal for the most efficient utilization.

If you plot delay and utilization, the result is a curve that shows a correlation between utilization and delay. Up to a point, as you increase utilization, the increase in delay is slight. There comes a point on the curve where a slight increase in utilization results in an unacceptable increase in delay. On the other hand, there is a point in the curve where a reduction in utilization doesn't yield a meaningful increase in QoS. The desired state lies within these points on the curve.

You could set a threshold to post an alert whenever the threshold limit is crossed. In that case, you would never react to a problem until delay has already become unacceptable. To avoid that late reaction you could set the threshold to post an alert before the threshold limit is crossed. In that case, you guarantee QoS at the cost of over-provisioning – you increase operating costs and never achieve efficient utilization.

Instead of responding *after* a threshold is crossed, Workload Optimization Manager analyzes the operating conditions and constantly recommends actions to keep the entire environment within the desired state. If you execute these actions (or let Workload Optimization Manager execute them for you), the environment will maintain operating conditions that assure performance for your customers, while ensuring the lowest possible cost thanks to efficient utilization of your resources.

The market and virtual currency

To perform Application Resource Management, Workload Optimization Manager models the environment as a market, and uses market analysis to manage resource supply and demand. For example, bottlenecks form when local workload demand exceeds the local capacity – in other words, when demand exceeds supply. By modeling the environment as a market, Workload Optimization Manager can use economic solutions to efficiently redistribute the demand or increase the supply.

Workload Optimization Manager uses two sets of abstraction to model the environment:

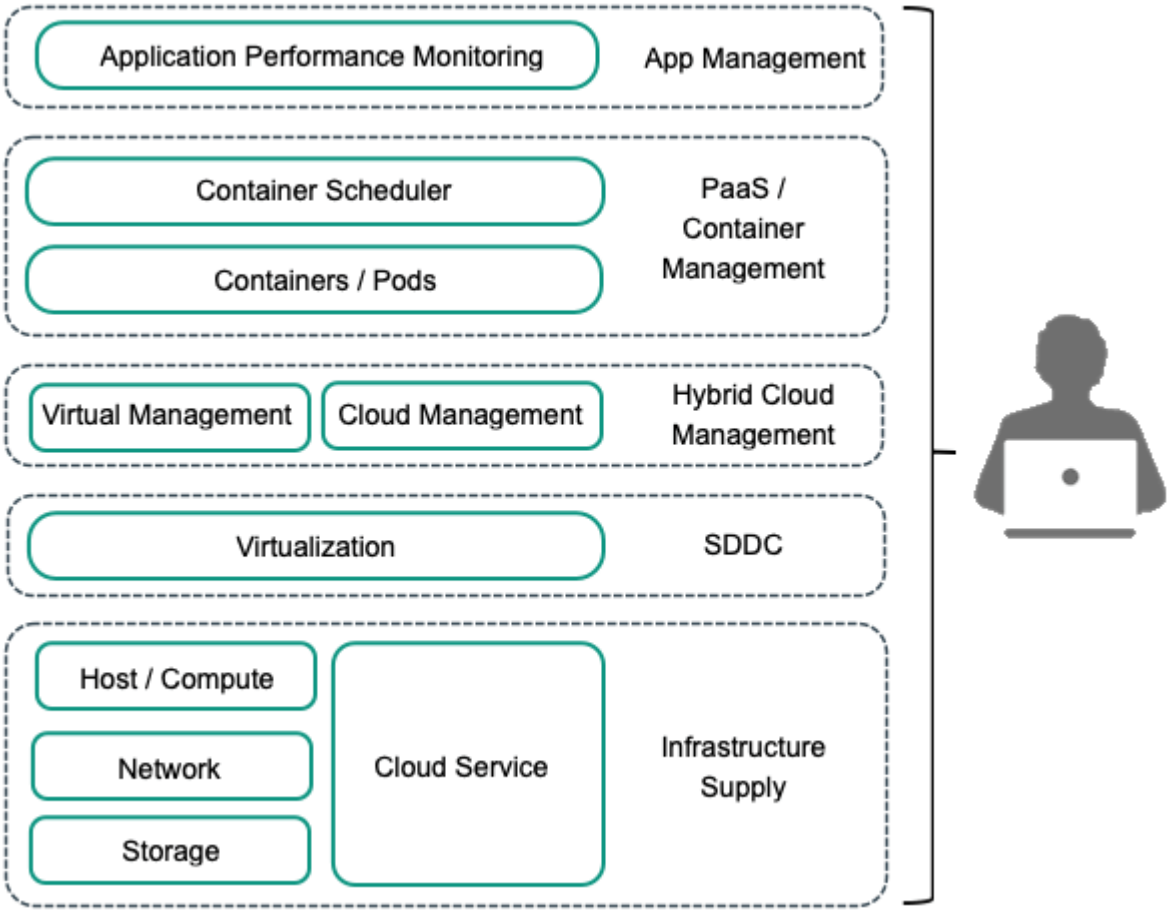
- Modeling the physical and virtual IT stack as a service supply chain

  The supply chain models your environment as a set of managed entities. These include applications, VMs, hosts, storage, containers, availability zones (cloud), and data centers. Every entity is a buyer, a seller, or both. A host machine buys physical space, power, and cooling from a data center. The host sells resources such as CPU cycles and memory to VMs. In turn, VMs buy host services, and then sell their resources (VMem and VCPU) to containers, which then sell resources to applications.

  See Supply Chain of Entities *(on page 159)* for a visual layout of the buyer and seller relationships.

- Using virtual currency to represent delay or QoS degradation, and to manage the supply and demand of services along the modeled supply chain

  The system uses virtual currency to value these buy/sell transactions. Each managed entity has a running budget — the entity adds to its budget by providing resources to consumers, and the entity draws from its budget to pay for the resources it consumes. The price of a resource is driven by its utilization — the more demand for a resource, the higher its price.



These abstractions open the whole spectrum of the environment to a single mode of analysis — market analysis. Resources and services can be priced to reflect changes in supply and demand, and pricing can drive resource allocation decisions. For example, a bottleneck (excess demand over supply) results in rising prices for the given resource. Applications competing for the same resource can lower their costs by shifting their workloads to other resource suppliers. As a result, utilization for that resource evens out across the environment and the bottleneck is resolved.

Risk index

Workload Optimization Manager tracks prices for resources in terms of the risk index. The higher this index for a resource, the more heavily the resource is utilized, the greater the delay for consumers of that resource, and the greater the risk to your QoS. Workload Optimization Manager constantly works to keep the risk index within acceptable bounds.

You can think of risk index as the cost for a resource — Workload Optimization Manager works to keep the cost at a competitive level. This is not simply a matter of responding to threshold conditions. Workload Optimization Manager analyzes the full range of buyer/seller relationships, and each buyer constantly seeks out the most economical transaction that is available.

This last point is crucial to understanding Workload Optimization Manager. The virtual environment is dynamic, with constant changes to workload that correspond with the varying requests your customers make of your applications and services. By examining each buyer/seller relationship, Workload Optimization Manager arrives at the optimal workload distribution for the current state of the environment. In this way, it constantly drives your environment toward the desired state.

NOTE:
The default Workload Optimization Manager configuration is ready to use in many environments. However, you can fine-tune the configuration to address special services and resources in your environment. Workload Optimization Manager provides a full range of policies that you can set to control how the software manages specific groups of entities. Before you make such policy changes, you should understand default Workload Optimization Manager operation. For more information about policies, see Working With Policies *(on page 1042)*.

The Workload Optimization Manager supply chain

Workload Optimization Manager models your environment as a market of buyers and sellers. It discovers different types of entities in your environment via the targets you have added, and then maps these entities to the supply chain to manage the workloads they support. For example, for a hypervisor target, Workload Optimization Manager discovers VMs, the hosts and datastores that provide resources to the VMs, and the applications that use VM resources. The entities in your environment form a chain of supply and demand where some entities provide resources while others consume the supplied resources. Workload Optimization Manager *stitches* these entities together.

For information about specific members of the supply chain, see Supply Chain of Entities *(on page 159)*.

Supply Chain Terminology

Workload Optimization Manager introduces specific terms to express IT resources and utilization in terms of supply and demand. These terms are largely intuitive, but you should understand how they relate to the issues and activities that are common for IT management.

Term	Definition
Commodity	The basic building block of Workload Optimization Manager supply and demand. All the resources that Workload Optimization Manager monitors are commodities. For example, the CPU capacity or memory that a host can provide are commodities. Workload Optimization Manager can also represent clusters and segments as commodities. When the user interface shows *commodities*, it's showing the resources a service provides. When the interface shows *commodities bought*, it's showing what that service consumes.
Composed Of	The resources or commodities that make up the given service. For example, in the user interface you might see that a certain VM is *composed of* commodities such as one or more physical CPUs, an Ethernet interface, and physical memory. Contrast *Composed Of* with *Consumes*, where consumption refers to the commodities the VM has bought. Also contrast *Composed Of* with the commodities a service offers for sale. A host might include four CPUs in its composition, but it offers CPU Cycles as a single commodity.
Consumes	The services and commodities a service has bought. A service *consumes* other commodities. For example, a VM consumes the commodities offered by a host, and an application consumes commodities from one or more VMs. In the user interface you can explore the services that provide the commodities the current service consumes.
Entity	A buyer or seller in the market. For example, a VM or a datastore is an entity.
Environment	The totality of data center, network, host, storage, VM, and application resources that you are monitoring.
Inventory	The list of all entities in your environment.
Risk Index	A measure of the risk to Quality of Service (QoS) that a consumer will experience. The higher the Risk Index on a provider, the more risk to QoS for any consumer of that provider's services.

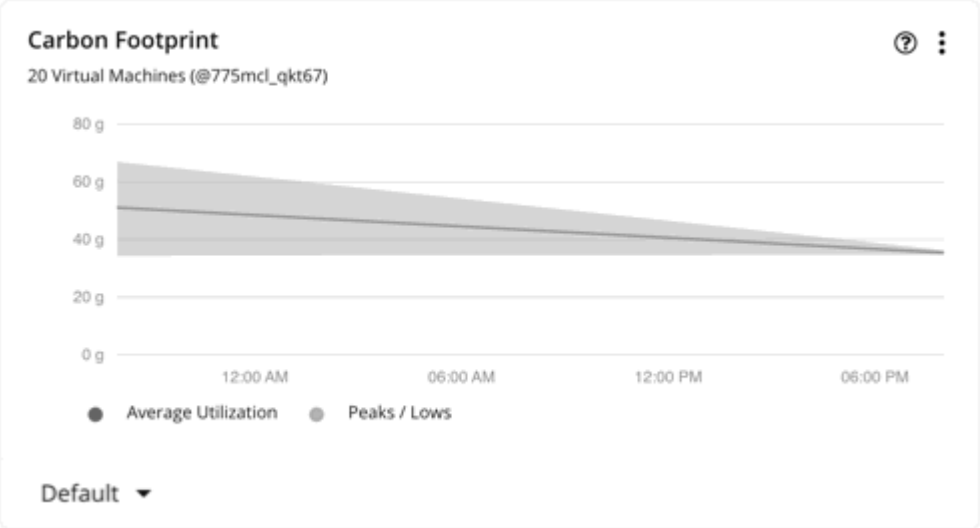
Term	Definition
	For example, a host provides resources to one or more VMs. The higher the Risk Index on the provider, the more likely that the VMs will experience QoS degradation. In most cases, for optimal operation the Risk Index on a provider should not go into double digits.

Sustainability features

Workload Optimization Manager helps companies reduce their energy usage and carbon footprint by optimizing IT infrastructure while ensuring that applications get the resources that they need to perform optimally. This core functionality, along with visibility into sustainability data, is crucial to setting and achieving sustainability goals. To visualize this data, Workload Optimization Manager provides the following features.

Sustainability charts

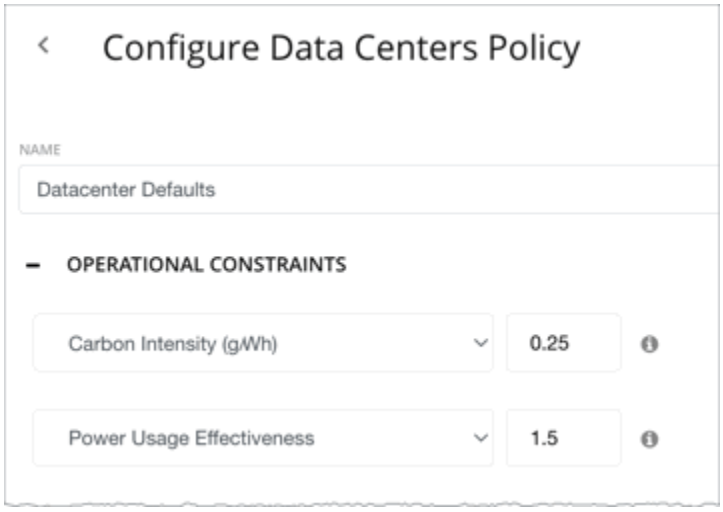
Workload Optimization Manager monitors the energy usage and carbon footprint of your IT infrastructure, and then displays relevant data in charts. Currently, these capabilities are supported for on-prem hosts and VMs. When you set the scope to any of these entities and then click the **Details** tab, you can view sustainability data in the **Energy** and **Carbon Footprint** charts.



For details, see Energy Chart *(on page 987)* and Carbon Footprint Chart *(on page 978)*.

Data center policies

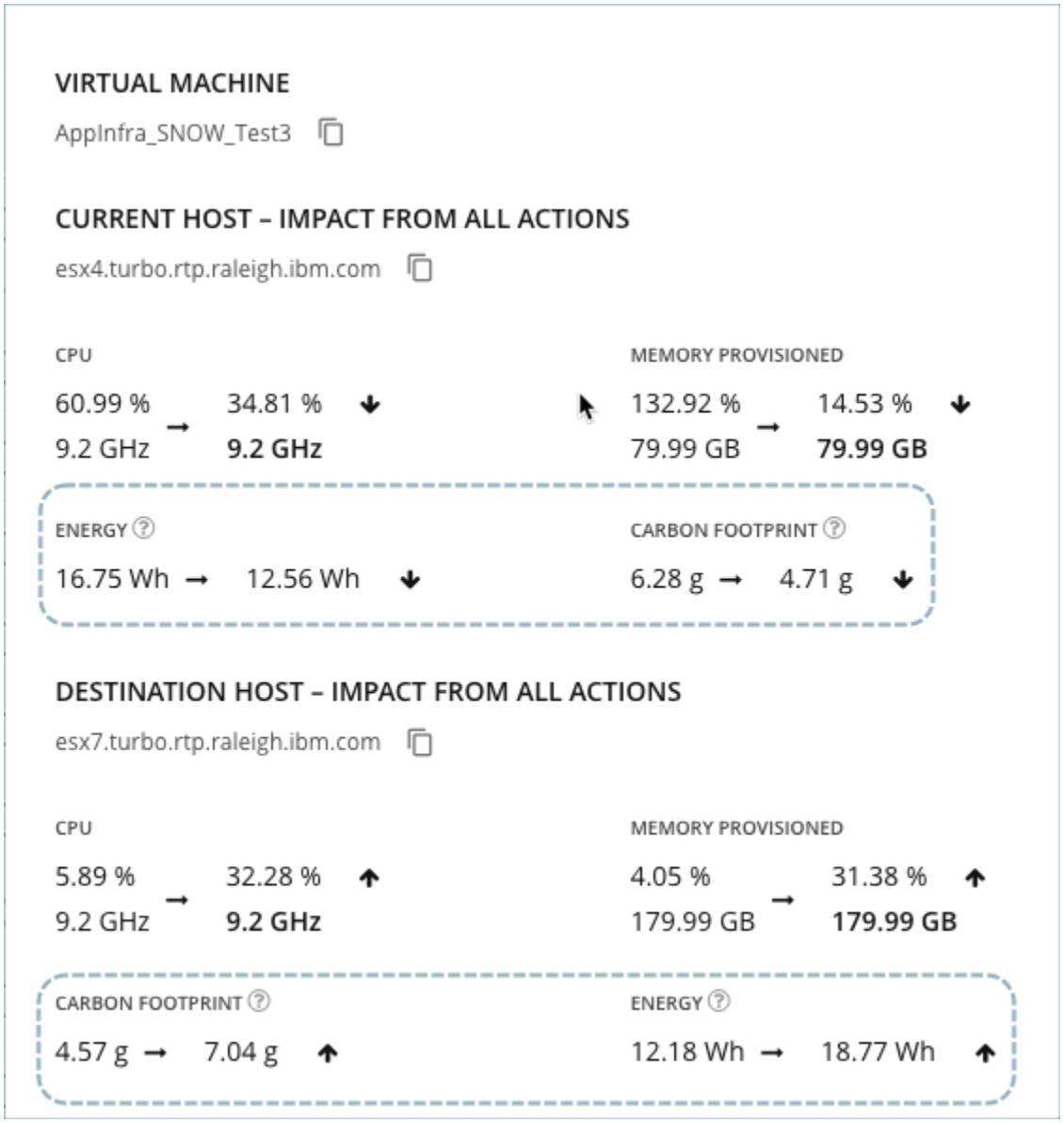
Workload Optimization Manager calculates *(on page 978)* carbon footprint using industry standards that take into account energy consumption, datacenter efficiency, and carbon intensity data. You can create Data Center policies to adjust the calculations according to the requirements of your data centers. For example, a data center in a particular location might have different requirements than data centers in other locations. After you adjust the calculations via policies, Workload Optimization Manager can accurately report your organization's carbon footprint.



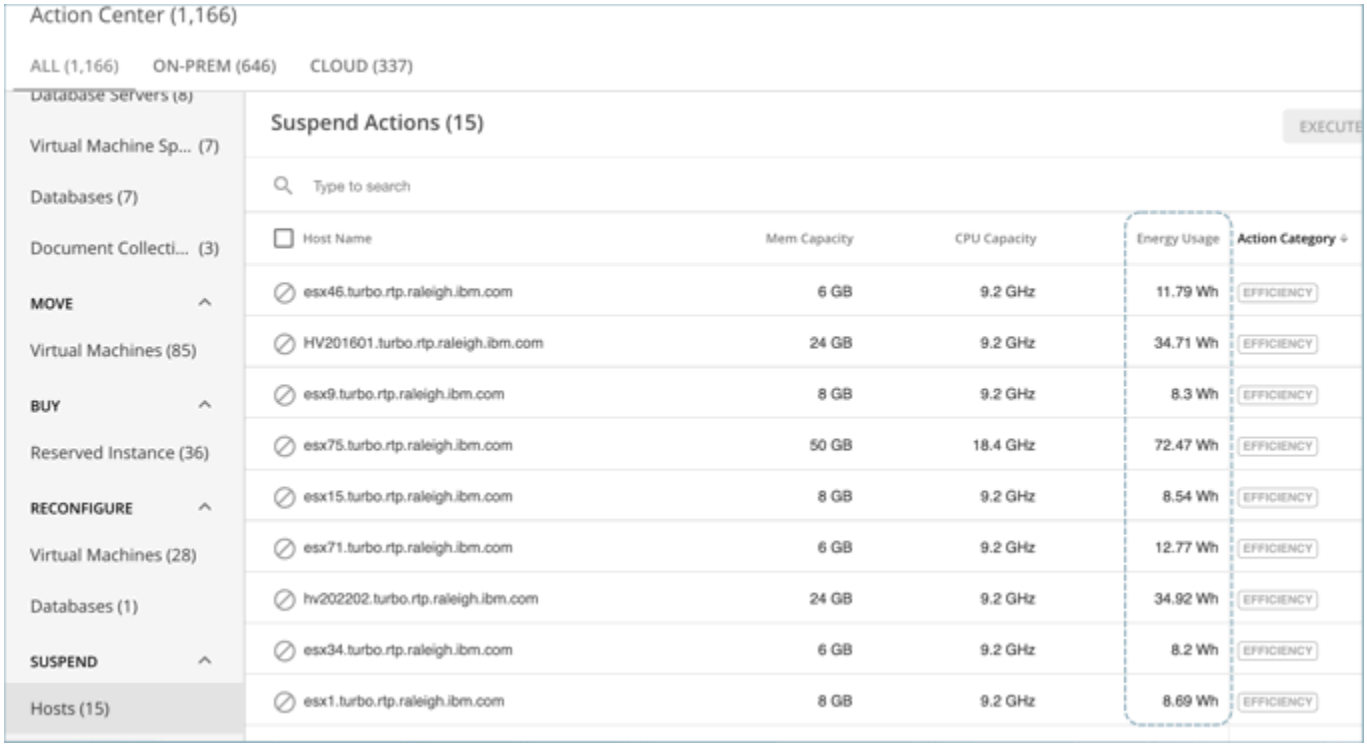
For details, see Data Center Policies *(on page 797)*.

Sustainability metrics in VM move, VM resize, and host suspend actions

Workload Optimization Manager displays the projected changes in carbon footprint and energy usage that are associated with VM move, VM resize, and host suspend actions on the **Action Details** page for those actions. You can use these metrics to understand the reduction in carbon footprint and energy usage that are associated with moving VMs and suspending hosts.



The current **Energy Usage** for each host is also displayed in the **Action Center** summary for host suspend actions.



For more information about these metrics and how they are calculated, see the following resources:

- Action details- host suspend impact on energy usage and carbon footprint *(on page 830)*
- Action details- VM move impact on energy usage and carbon footprint *(on page 832)*
- Action details- VM resize impact on energy usage and carbon footprint *(on page 833)*

Getting started

To get started with the platform, open a web browser to your Workload Optimization Manager installation. The Workload Optimization Manager platform serves the user interface to your browser, where you can log in and get started managing your environment. In this way, you can access the unique capabilities of Workload Optimization Manager from any internet connection.

Logging in to Workload Optimization Manager

To get started with the platform, open a web browser to your Workload Optimization Manager installation. The Workload Optimization Manager platform serves the user interface to your browser, where you can log in and get started managing your environment. In this way, you can access the unique capabilities of Workload Optimization Manager from any internet connection.

Before you can log in, your enterprise must have a valid Workload Optimization Manager account, or an instance of Workload Optimization Manager must be installed in your environment. To get the IP address of your Workload Optimization Manager installation, contact your system administrator.

To log in to Workload Optimization Manager:

1. Navigate your Web browser to the Workload Optimization Manager installation.

   For the URL, provide the IP address or machine name for the installation. This URL opens the Workload Optimization Manager Login page. You should bookmark this URL for future use.

2. Provide the user name and password for your account.

   Your system administrator creates user accounts. Contact your system administrator for login information.

After you log in, the browser opens to the Home Page *(on page 152)*. This page is your starting point for sessions with the Workload Optimization Manager platform. From the Home Page you can see the overviews of your environment.

To display this information, Workload Optimization Manager communicates with *target services* such as hypervisors, storage controllers, and public cloud accounts. Note that your Workload Optimization Manager administrator sets up the target configuration. For information about supported targets and how to configure them, see Target Configuration *(on page 173)*.

The Home Page

When you launch Workload Optimization Manager, the **Home Page** is the first view you see. From there you can:

- Choose a View to see overviews of your environment:
  - APPLICATION – See your environment in the context of your Business Applications *(on page 544)*.
  - ON-PREM – See details for the on-prem environment. Notice that the Supply Chain excludes cloud entities and only shows the entities that are on-prem.
  - CLOUD – See details for the cloud environment, including pending actions, a listing of your cloud accounts by cost, the locations of cloud datacenters that you are using, estimated costs, and other cost-related information.
- Use the Supply Chain Navigator to inspect lists of entities

  Click an entity tier in the Supply Chain to see a list of those entities. For example, click Virtual Machine to see a list of all the VMs in your environment.
- Navigate to other Workload Optimization Manager pages, including:
  - Search – Set the session scope to drill down to details about your environment
  - Plan – Run what-if scenarios
  - Place – Use Workload Optimization Manager to calculate the best placement for workloads, and execute the placement at the time you specify
  - Dashboards – Set up custom views with charts that focus on specifics in your environment
  - Settings – Configure Workload Optimization Manager to set up business rules and policies, configure targets, define groups, and perform other administrative tasks

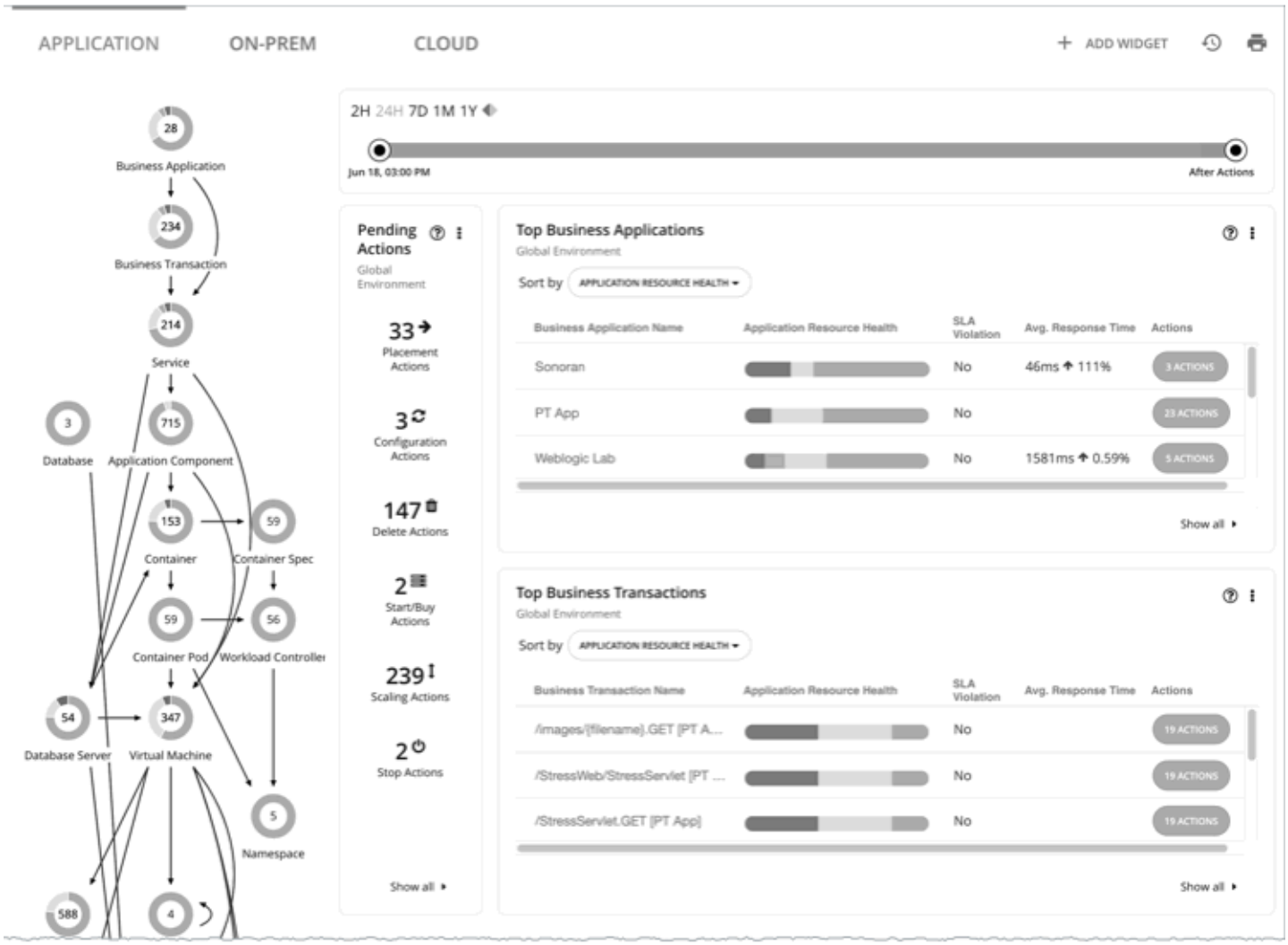
Getting home



Wherever you are in your Workload Optimization Manager session, you can always click the Home icon to return to the **Home Page**.

Application view

The Application view presents your environment in the context of your Business Applications *(on page 544)*. See the overall health of your applications, examine any performance and compliance risks, and execute the actions that Workload Optimization Manager recommends to address these risks.

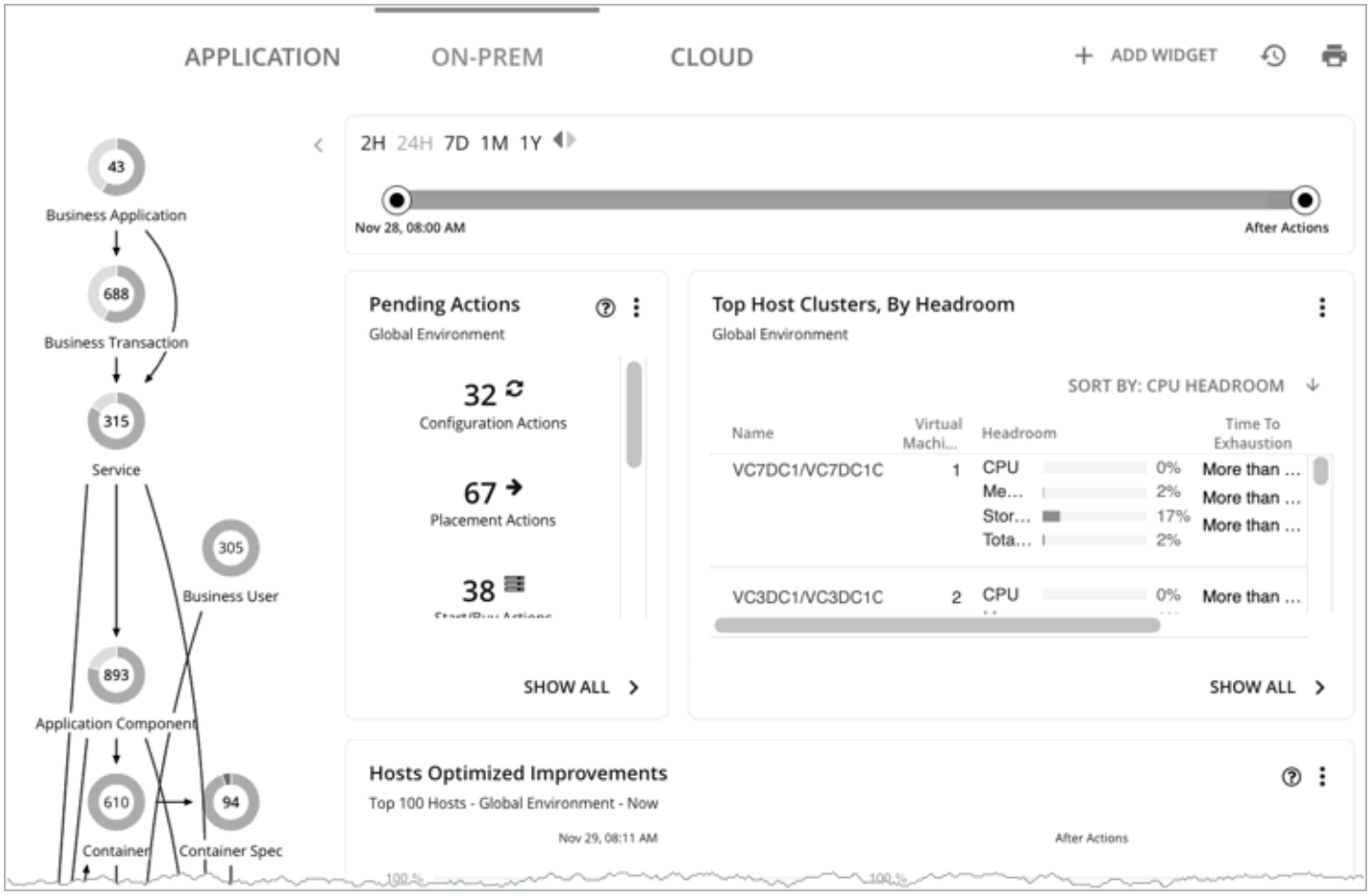


This view also shows the Business Transactions *(on page 546)* and Services *(on page 549)* that make up your Business Applications. You can see finer details and set SLOs at these levels of the application model.

NOTE:
If certain application entities do not stitch into the supply chain infrastructure for some reason, Workload Optimization Manager displays them in both the On-prem and the Cloud views. Once Workload Optimization Manager can stitch them into the infrastructure, it classifies them according to the class of the infrastructure and displays them in the correct views.

On-prem view

When you set your session to the Global Scope, you can then select the On-prem view. This shows an overview of your on-prem environment. If you don't have any workload on the public cloud, then you should use this as your starting point for a Workload Optimization Manager session. If you have a hybrid environment (on-prem and on the public cloud), then you can refer to this view to see a detailed on-prem overview.



The Supply Chain shows all the on-prem entities in your environment. The charts show details about your environment, including:

- Overviews of pending actions

  When appropriate, the overview includes estimated one-time savings or costs associated with the actions.

- Top Host Cluster utilization

  See a list of the most utilized clusters. The chart shows these clusters, along with a count of actions for each. To drill down into the cluster details, click the cluster name. To see and execute the specific actions, click the **ACTIONS** button for that cluster. To see all the clusters in your environment, click **SHOW ALL**.

- Optimized Improvements

  Compare current resource utilization with the utilization you would see if you choose to execute all the pending actions.

- Action history

  You can see a history of all actions that have been recommended and executed, or of just the actions that have been accepted and executed.

Cloud view

When you set your session to the Global Scope, you can then select the Cloud view. This shows an overview of your cloud environment. If all your workload is on the public cloud, then you should use this as your starting point for a Workload Optimization Manager session. If you have a hybrid environment (on-prem and on the public cloud), then you can refer to this view to see a detailed cloud overview.

To view cloud cost information, you must have one or more public cloud targets set up in your Workload Optimization Manager installation. For information about setting up public cloud targets, see Cloud Targets *(on page 373)*.

In this view, the Supply Chain shows all the cloud entities in your environment. The charts show details about your cloud environment, including:

- Overviews of pending actions

  The overview includes the estimated monthly savings or cost associated with those actions.

- Top Accounts utilization

  See a list of the most utilized public cloud accounts. The chart shows these accounts, along with an estimate of the monthly cost for each. To see all the cloud accounts in your environment, click **SHOW ALL**.

- Necessary Investments and Potential Savings

  For the current set of pending actions, these charts show the impact in dollar value. Necessary Investments are from actions to provision more workloads or to resize workloads up. Potential Savings are from actions to resize down, or to purchase discounts and put them into active use.

- Charts that show your current discounts. For details, see Discounts *(on page 157)*.

- Billed Cost by Service

  This chart shows costs over time for each cloud service that you use in your cloud accounts. For example, you can see the cost for AWS CloudWatch, compared to the cost for AWS S3 storage.

Tracking cloud cost

Workload Optimization Manager tracks your cloud costs based on the cost information it discovers from targets (for example, accounts, billing reports, and on-demand or discount costs) and price adjustments *(on page 1109)*.

> **NOTE:**
> It is possible for Workload Optimization Manager to report negative amounts. For example, when discounts are larger than costs, the result is a negative amount. Currently, these amounts are not shown directly in cost-focused charts (such as the Expenses charts). To check for any negative amounts, hover on a data point in a chart and then review the data in the tooltip.

Cost for services

Workload Optimization Manager uses the billing reports from your cloud service providers, as they are associated with your cloud targets. Workload Optimization Manager parses these reports to get cost breakdowns by service, service provider, Azure Resource Group, and cloud account. You can see cost data in the Expenses charts and Cost Breakdown by Tag charts.

Workload expenses

Workloads are the VMs running in your environment, or other hosted processes such as database servers and containers. Workload Optimization Manager tracks the following expenses for your workloads:

- Compute

  For compute expenses Workload Optimization Manager uses hourly expense per template as specified in the associated public cloud account.

- Storage

  Workload Optimization Manager discovers the storage tier that supports a given workload, and uses the tier pricing to calculate storage cost.

- License

  For AWS environments, Workload Optimization Manager can calculate OS costs. To calculate the OS cost for a VM, Workload Optimization Manager subtracts the template cost from the published workload cost. It assumes the difference is the license cost for that workload. If the OS is open source, then there will be no difference, and license cost is zero. Analysis does not consider AWS Marketplace costs.

  For Azure environments, Workload Optimization Manager can track OS costs for existing VMs. For actions to purchase reservations, Workload Optimization Manager does not include the OS cost. Analysis considers the base OS cost, but does not consider additional costs for support or other add-on features that are bundled with the OS. The affected OS types are Ubuntu PRO, SUSE 24/7, and RHEL with HA.

Workload Optimization Manager uses this cost information when making scaling decisions, both in real time and in plans. You can see this information in Expenses charts and in the results of Migrate to Cloud plans.

Costs for dedicated tenancy on AWS

When you create VMs on AWS, you can specify their tenancy. When you specify Dedicated Tenancy (DT), the VMs you create are Amazon EC2 instances running on hardware that is dedicated to a single customer. To understand DT in the context of Workload Optimization Manager, you should consider:

- For AWS, the Workload Optimization Manager supply chain shows an Availability Zone as a Host. The supply chain does not indicate whether certain VMs have tenancy dedicated to specific resources in the given availability zone. Also, Workload Optimization Manager does not discover or show the costs for dedicated hosting of your workloads.
- Pricing for DT workloads is different than pricing for Shared Tenancy. Workload Optimization Manager does not discover that difference, and uses Shared Tenancy cost for the DT workloads. In action descriptions, the listed savings or investments will be based on Shared Tenancy costs.
- Workload Optimization Manager discovers the true costs of RIs for DT workloads. However, because the on-demand VM costs are based on Shared Tenancy, Workload Optimization Manager can overstate the savings you would get for purchasing and using RI capacity. In most cases, recommendations to purchase RIs will be correct. However, the time to achieve ROI could take longer than action descriptions and charts indicate.
- Some instance types that are valid for Shared Tenancy are not valid for DT. To see which instance types are valid for your DT VMs, consult the AWS documentation or your AWS representative.
- Under some circumstances Workload Optimization Manager can recommend changing a workload to a valid instance type for the tenant, even though the current type is already valid. This can happen when the instance type is not included in the Offer File for the tenancy. For example, assume the t3a template family does not support dedicated tenancy. However, assume that the user created a t3a instance with dedicated tenancy in the EC2 console. In that case, Workload Optimization Manager will see this as a misconfiguration and recommend changing to a different instance type.

To address these issues, you can create groups that set a scope to your DT workloads. For example, you can use naming conventions, tagging, or other means to identify your DT workloads. Then you can create dynamic groups based on those

indicators. With those groups, you can create policies and dashboards that correspond to the differences you see in your DT environment. Use this approach to address issues for:

- Available Instance Types

  To resize a workload, Workload Optimization Manager generates an action to change that workload to a different instance type. Because Workload Optimization Manager does not discover the difference between instance types that are valid for DT and for Shared Tenancy, it can recommend scaling a DT workload to an unavailable instance type. To avoid this, create a policy for the DT group, and exclude the unavailable instance types.

- Displaying Costs

  Workload Optimization Manager charts show the costs for your environment. If the scope includes Dedicated Tenancy workloads, then the calculated cost will be incomplete. For example, since AWS does not return pricing data for converted RIs (that is, RIs that have been exchanged at least once) that are on *All Upfront* payment plans, Workload Optimization Manager does not include such RIs in its calculations of RI utilization or cost.

  Use scope to minimize this effect. You can create separate dashboards for your DT and Shared Tenancy workloads.

Discounts

Workload Optimization Manager analysis takes advantage of cloud provider discounts to calculate optimal workload placement and to arrive at the best possible costs for your deployments on the cloud.

The following table describes the current level of support for the different cloud provider discounts. Discounts that are not listed are not supported.

Discount	Discovered	Considered in analysis
AWS EC2 Reserved Instances (RIs)	Yes	Yes
AWS Compute Savings Plans	Yes	Yes
AWS RDS Reserved DB Instances	Yes	No
Azure reservations	Yes	Yes
Azure savings plans	Yes	No
Google Cloud committed use discounts	Yes	Yes

The Cloud View in the Homepage includes the following charts that show discount data:

- Discount Inventory *(on page 1028)*

  This chart lists the cloud provider discounts that you have purchased for your workloads.

- Discount Utilization *(on page 1032)*

  This chart shows how well you have utilized your current discount inventory *(on page 1028)*. The desired goal is to maximize the utilization of your inventory and thus take full advantage of the discounted pricing offered by your cloud provider.

- Discount Coverage *(on page 1026)*

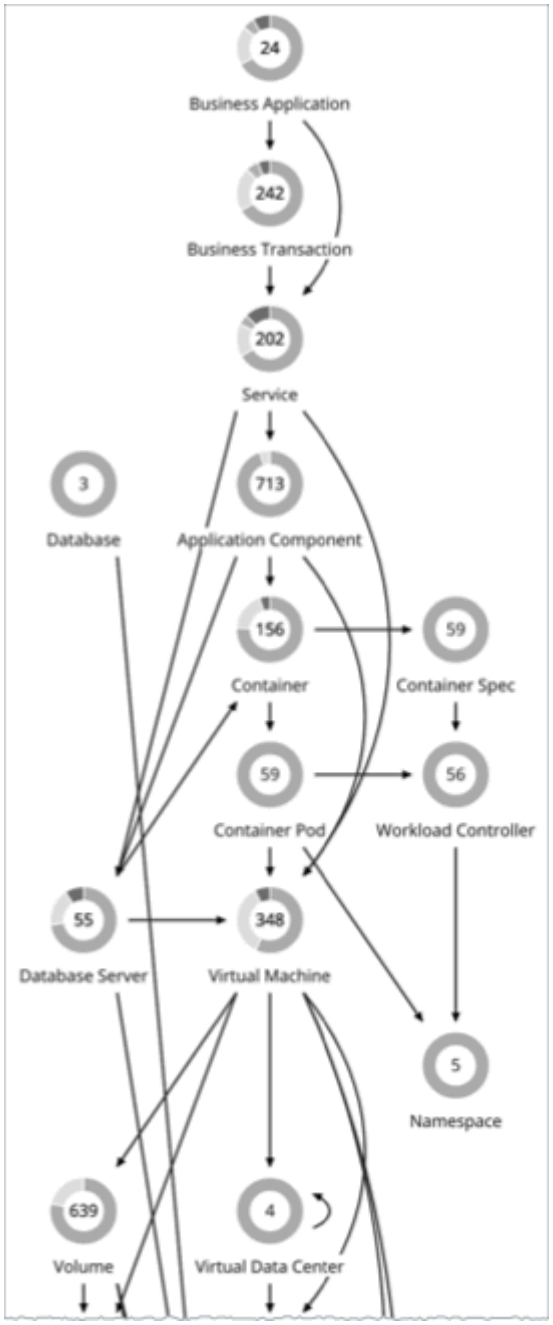
  This chart shows the percentage of cloud workloads (VMs and RDS database servers) covered by discounts. For VMs covered by discounts, you can reduce your costs by increasing coverage. To increase coverage, you scale VMs to instance types that have existing capacity.

- Potential Savings

  Workload Optimization Manager can recommend purchasing instance types at a discounted rate to help you increase the percentage of VMs covered by discounted pricing and reduce on-demand costs. This chart shows your pending purchases. Download the list of purchases and then send it your cloud provider or representative to initiate the purchase process.

  NOTE:
  Currently, Workload Optimization Manager can recommend purchasing AWS EC2 RIs and Azure reservations, but only in plans. Purchase actions are not generated in the real-time market and therefore do not show in Action Center.

Supply chain of entities



To perform Application Resource Management, Workload Optimization Manager models your environment as a market of buyers and sellers linked together in a supply chain. This supply chain represents the flow of resources from the datacenter, through the physical tiers of your environment, into the virtual tier and out to the cloud. By managing relationships between these buyers and sellers, Workload Optimization Manager provides closed-loop management of resources, from the datacenter, through to the application.

Reading the supply chain

By looking at the Supply Chain, you can see:

- How many entities you have on each tier

  Each entry in the supply chain gives a count of entities for the given type.

- The overall health of entities in each tier

  The ring for each entry indicates the percentage of pending actions for that tier in the datacenter. Ring colors indicate how critical the actions are - Green shows the percentage of entities that have no actions pending. To get actual counts of pending actions, hover on a ring to more details.

- The flow of resources between tiers

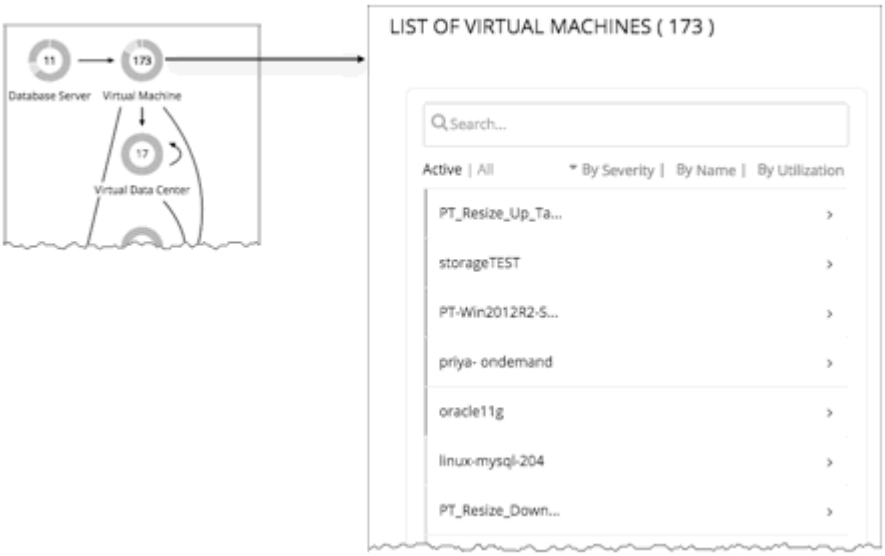
  The arrow from one entry to another indicates the flow of resources. For example, the Virtual Machine entry has arrows to Hosts and to Storage. If the VMs are running in a Virtual Data Center, it will have another arrow to that as well. This means that your VMs consume resources from hosts, storage, and possibly from VDCs.

Listing Entities From the Home Page

The Supply Chain shows the relationships of entities in your environment. When you're on the **Home Page** with a global scope, the supply chain filters its display according to the view you have chosen:

- APPLICATIONS – All your Business Applications *(on page 544)*
- ON-PREM – All your on-prem entities
- CLOUD – All your entities on the public cloud

To see a list of entities, click an entity tier in the Supply Chain.



Working with a scoped view

By default, the **Home Page** shows a global view of your environment. To drill down into specifics of your environment, you can set a scope to your Workload Optimization Manager session. A scoped view shows details about the specific entities in that scope.

Once you have set a scope, you can use the supply chain to zoom in on a related tier to see details about the entities on that tier.

If you find the current scope to be useful, you can save it as a named group. Using named groups is an easy way to return to different scopes that you have saved.

Things you can do

- Scoping the Workload Optimization Manager Session *(on page 161)*
- Navigating with the supply chain *(on page 171)*
- Viewing cluster headroom *(on page 172)*

Scoping the Workload Optimization Manager session

The default scope for the **Home Page** shows an overview of the global environment. What if you want to focus on less than the global environment? Assume you are responsible for a subset of workloads in your environment. This could be:

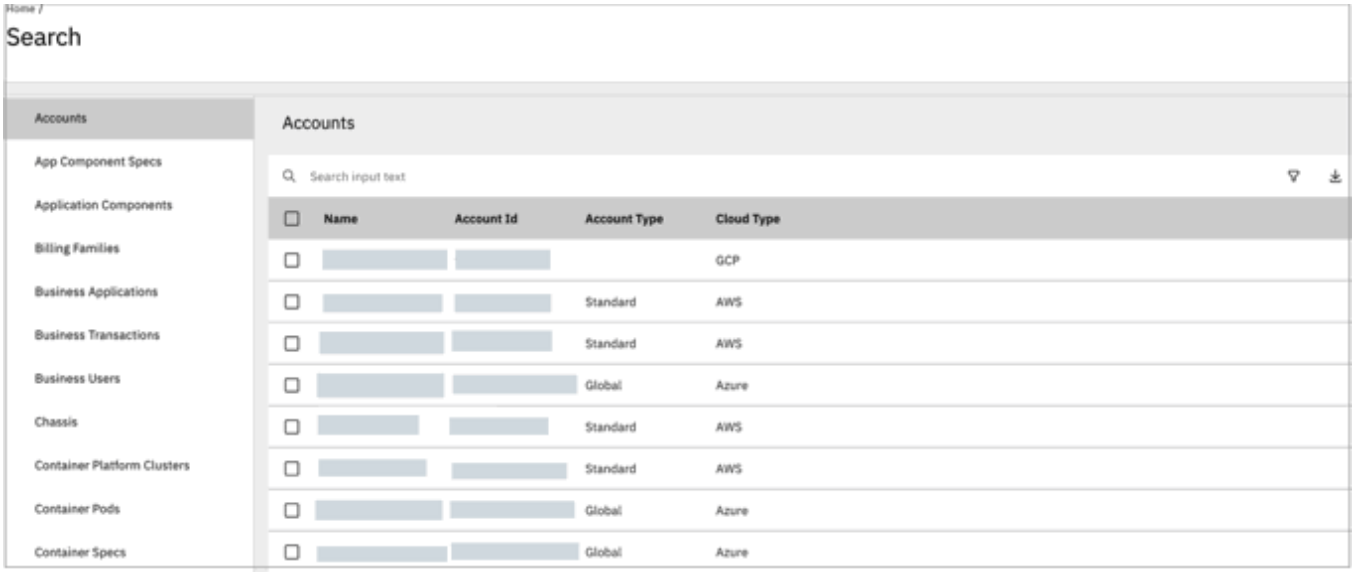
- Workloads managed on a single host cluster
- The workloads in a single datacenter
- A custom group of workloads you have created in Workload Optimization Manager

It's easy to set the session scope so that Workload Optimization Manager zooms in on the part of the environment that you want to inspect. Once you set the scope, you can get a quick picture of system health for that scope. If you find a certain scope to be useful, you can save it as a named group that you can return to later.

1. Navigate to the Search Page.

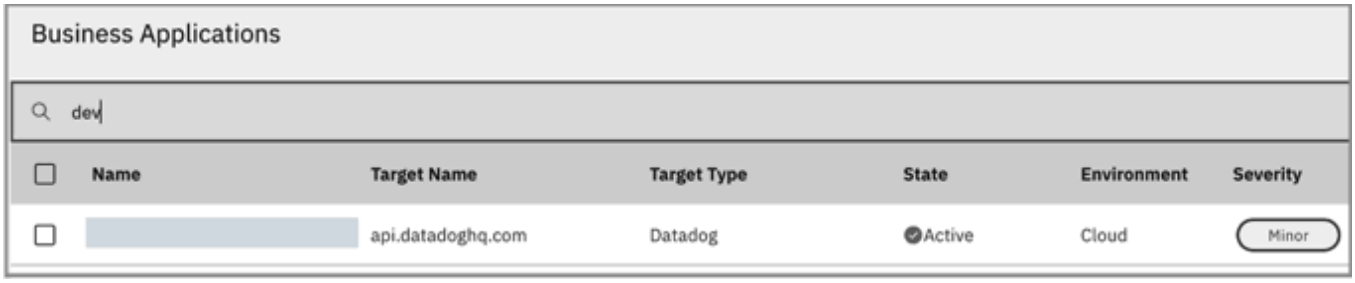
   Click to navigate to the Search Page. This is where you can choose the scope you want.
2. Choose the type of entities to search.

   In the Search Page, choose a category that you want to search through. You can focus on entities by type, by groups, or by clusters. When you select an entity type, the page updates to show all entities of that type.

   You can download a list of entities.
3. Use **Search** to filter the listing.

   For example, if you're showing **Business Applications** and you search for "Dev", then you will see all Business Applications with "Dev" in their names.

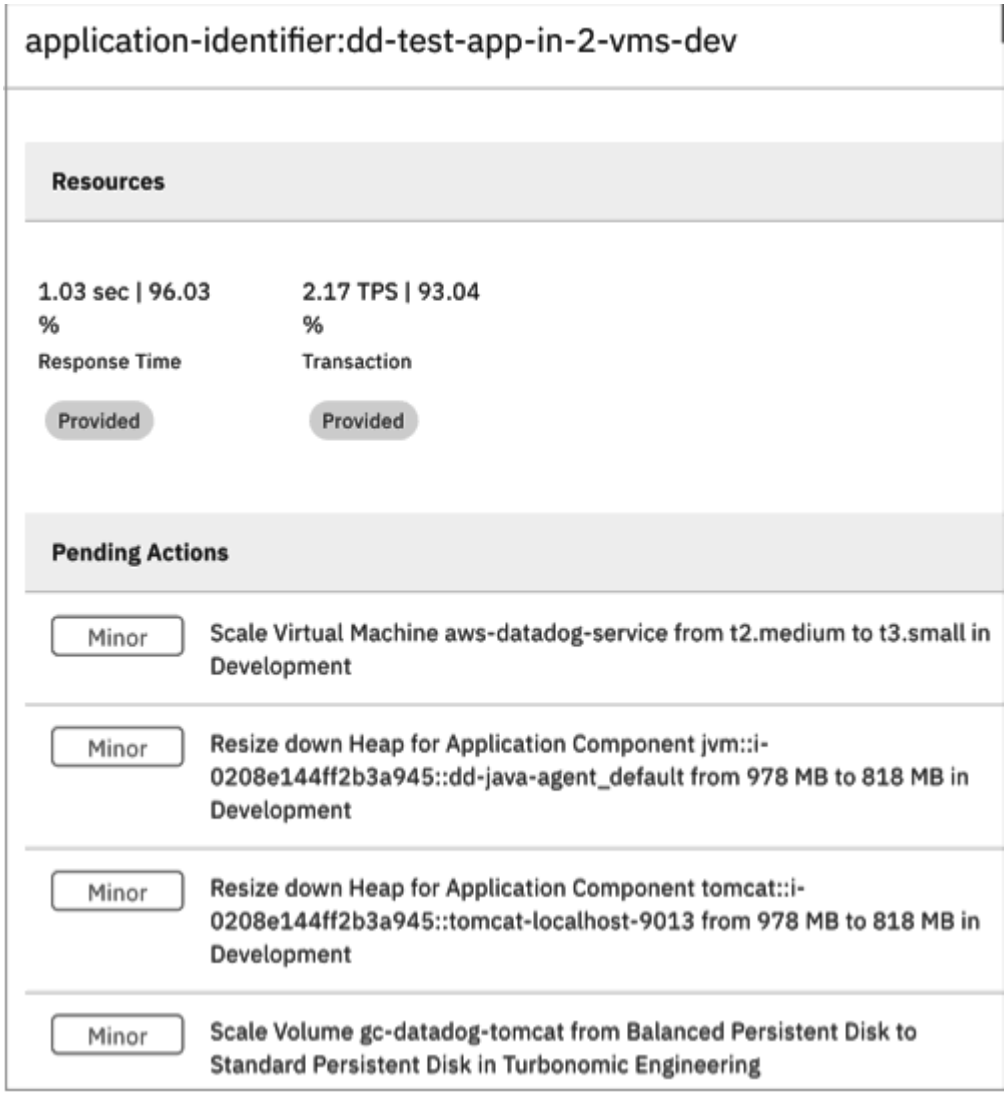


4. Expand an entry to see details.

Select the **Show more details** button ( ) to see details.

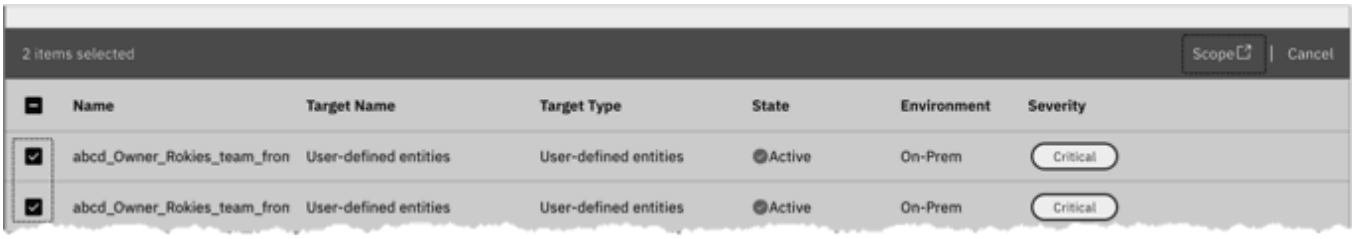
For example, details of a group or an entity include utilization details and pending actions.

NOTE:
For hosts in the public cloud, utilization and capacity for host and datacenter resources don't affect Workload Optimization Manager calculations. When you expand an entry for a public cloud host, the details do not include information for these resources.



5. Select one or more entries to set the focus of the **Home Page**.
   - Choose an entity type, and set the scope to one or more of those entities.
   - For different types of groups, click to set a single group as your scope.

   If you choose a category of entities to limit the list, then you can select one or more of the entities for your session scope. After you select the entities you want to include in your scope, click **SCOPE TO SELECTION** to set the session scope to those entities. Click **SCOPE** to set the session scope to those entities.



   If you choose groups or clusters, then you can select a single entry to set the scope for your session. When you select an entry in the list, that sets the focus of the **Home Page**. For example, if you select a host cluster in the **Search** listing, you set the **Home Page** focus to that cluster. Use the **Home Page** bread crumbs to set a different scope, or you can return to **Search** and set a different scope from there.

Overview charts



The Overview Charts show your environment's overall operating health for the current session scope. A glance at the Overview gives you insights into service performance health, overall efficiency of your workload distribution, projections into the future, and trends over time.

The charts in this view show data for the current scope that you have set for the Workload Optimization Manager session. For the global scope, the charts roll up average, minimum, and peak values for the whole environment. When you reduce the scope (for example, set the scope to a cluster), the charts show values for the entities in that scope.

Some charts included in this view are:

- Pending Actions

  See all the actions that are pending for the current scope.
- Health

  Quickly see the health of the entities in this scope– How many entities have risks, and how critical the risks are.
- Optimized Improvements

  A comparison of utilization in your environment before executing the pending actions, and then after.
- Capacity and Usage

  This chart lists resources that are used by the current scope of entities, showing utilization as a percentage of the capacity that is currently in use.
- Multiple Resources

  See the utilization over time of various resources that are used by the current scope of entities.
- Top Entities

  For example, Top Virtual Machines. These charts list the top consumer entities in the current scope.
- Risks Avoided

Each action addresses one or more identified risks or opportunities in your environment. This chart shows how many risks have been addressed by the executed actions.

- Accepted Actions

  This chart shows how many actions have been executed or ignored, and whether they have been executed manually or automatically.

What you can do:

- Set scope: See Scoping the Workload Optimization Manager Session *(on page 161)*
- Create new charts: See Creating and Editing Chart Widgets *(on page 966)*

Setting chart focus

The charts update to reflect the focus that you have set for your viewing session. While viewing the Overview Charts, you can set the focus in different ways:

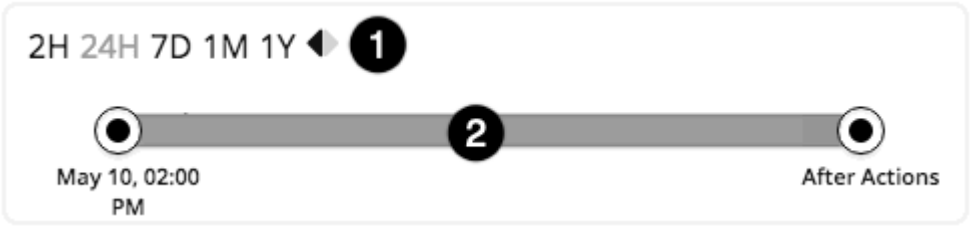
- Set supply chain focus

  Choose a tier in the supply chain to set the view focus - see Navigating With the Supply Chain *(on page 171)*

- Set scope

  Use Search to set the scope of the viewing session - see Scoping the Workload Optimization Manager Session *(on page 161)*

Chart time frame



You can set a time frame from recent hours to the past year (1), and set that to the charts in the view. Use the Time Slider to set specific start and end times within that range (2). The green section in the slider shows that you can set the time range to include a projection into the future. For this part of the time range, charts show the results you would see after you execute the current set of pending actions.

For most charts, you can also configure the chart to hard-code the time range. In that case, the chart always shows the same time scale, no matter what scale and range you set for the given view.

Note that Workload Optimization Manager stores historical data in its database. As you run Workload Optimization Manager in your environment for more time, then you can set a time range to show more history.

Details view

The Details View shows more details about the entities in your session scope. These charts focus on the utilization of resources by these entities, so you can get a sense of activity in that scope over time.



What you can do:

- Set scope: See Scoping the Workload Optimization Manager Session *(on page 161)*
- Create new charts: See Creating and Editing Chart Widgets *(on page 966)*

Setting chart focus

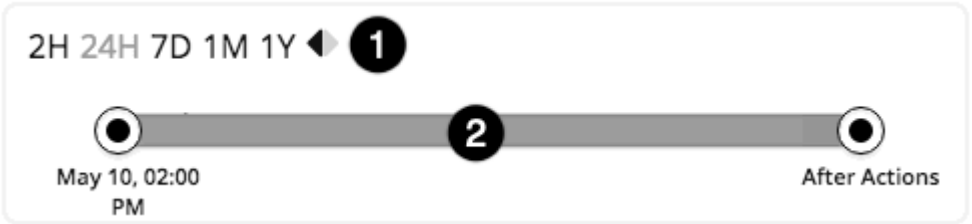
The charts update to reflect the focus that you have set for your viewing session. While viewing the Overview Charts, you can set the focus in different ways:

- Set supply chain focus

  Choose a tier in the supply chain to set the view focus - see Navigating With the Supply Chain *(on page 171)*
- Set scope

  Use Search to set the scope of the viewing session - see Scoping the Workload Optimization Manager Session *(on page 161)*

Chart time frame

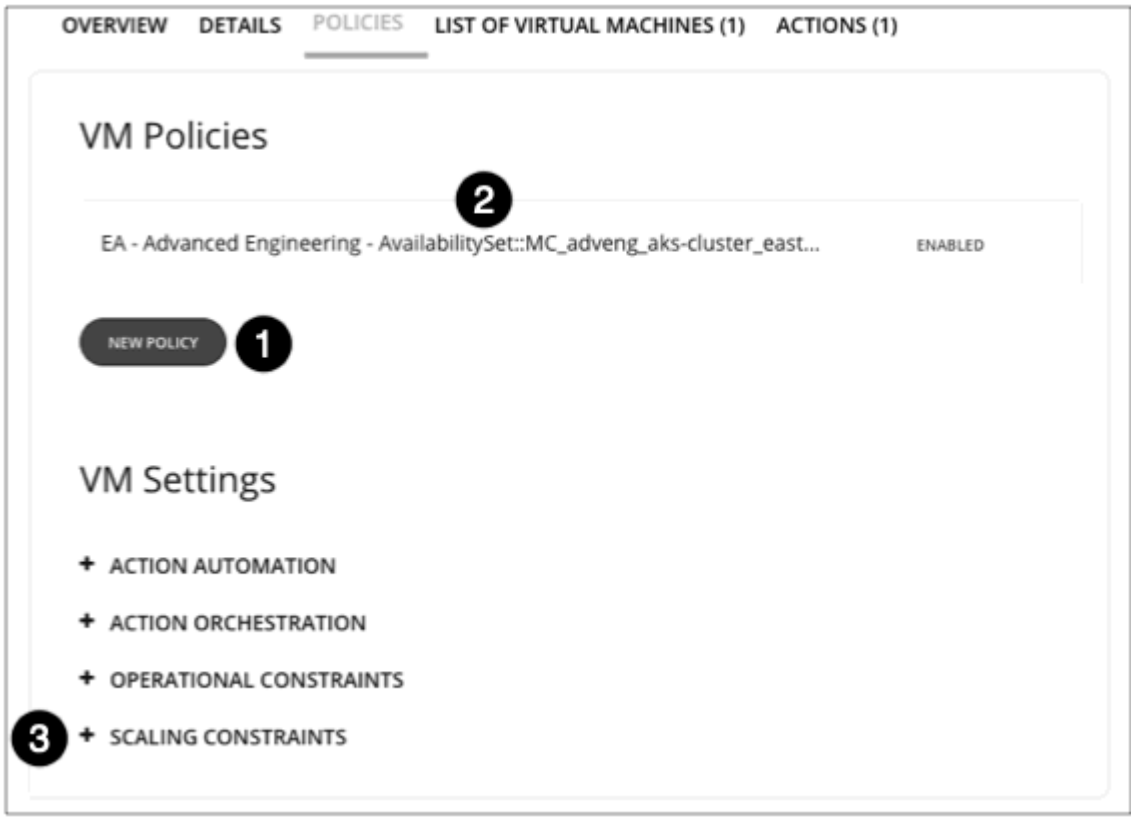


You can set a time frame from recent hours to the past year (1), and set that to the charts in the view. Use the Time Slider to set specific start and end times within that range (2). The green section in the slider shows that you can set the time range to include a projection into the future. For this part of the time range, charts show the results you would see after you execute the current set of pending actions.

For most charts, you can also configure the chart to hard-code the time range. In that case, the chart always shows the same time scale, no matter what scale and range you set for the given view.

Note that Workload Optimization Manager stores historical data in its database. As you run Workload Optimization Manager in your environment for more time, then you can set a time range to show more history.

Scope policies



The Policy View gives you a look at the Automation Policies that are set for the entities in the current scope. For each policy, you can see whether it is enabled. In addition,

1. To create a policy, click **New Policy**. When you create a new policy, it automatically includes the current scope. You can add other groups to the policy scope.

   NOTE:
   You can enable more than one policy for the same scope. If two policies apply different values for the same setting, then the most conservative value takes effect.

2. To edit a policy, click the policy name. You can then change the policy settings and whether the policy is enabled.
3. To see the current policy settings, expand a settings category. For each setting, you can see which policy determines the value – either the default policy or a custom policy that is applied to this scope.

For more information, see Automation Policies *(on page 1049)*.

Entity placement constraints



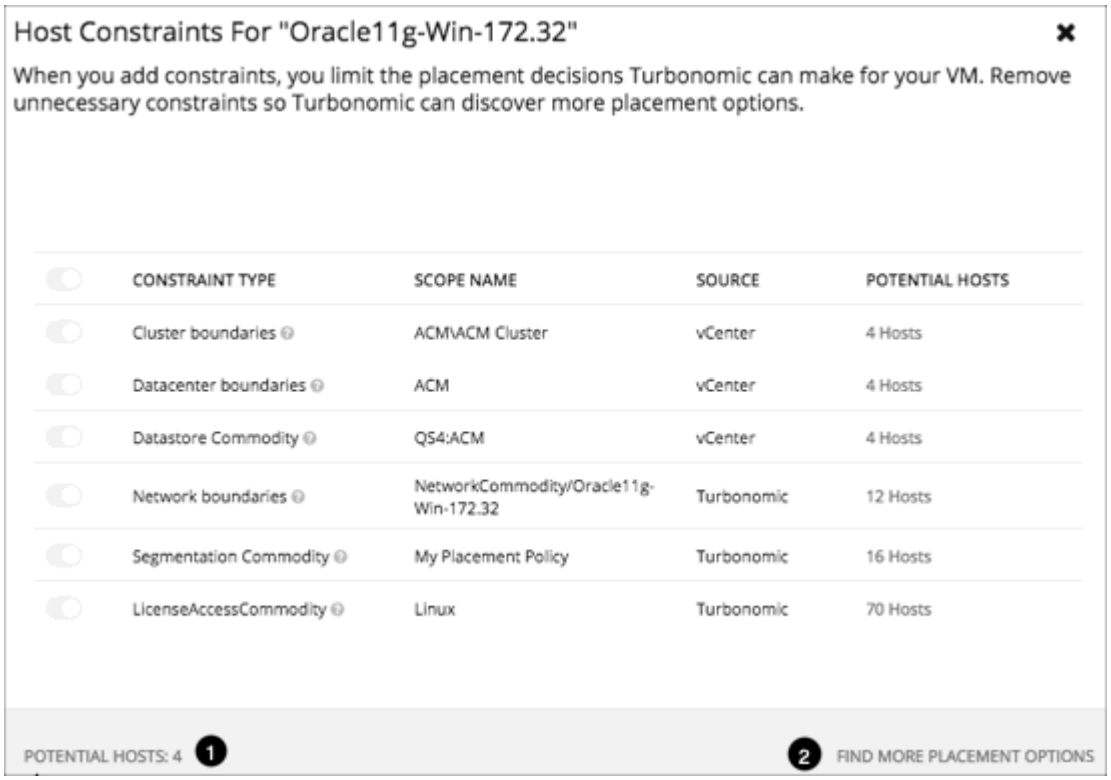
When you drill down to a single entity, you can see details about the entity's relationships in the supply chain. This shows you which entities provide resources to this entity. When considering providers for this entity, you can see the name of each current provider, and how many alternative providers Workload Optimization Manager can choose from if the current one becomes overutilized.

Reviewing the constraints on an entity helps you understand the actions that Workload Optimization Manager recommends. If an action seems questionable to you, then you should look at the constraints on the affected entities. It's possible that some policy or constraint is in effect, and it keeps Workload Optimization Manager from recommending a more obvious action.

Experimenting with placement constraints

For each provider or consumer in the list, you can open a **Constraints** fly-out that gives more details about limits on the current element's supply chain relationships.

For example, assume the **PROVIDERS** list shows your VM's **CURRENT PLACEMENT** is on Host A, and for **OTHER POTENTIAL PLACEMENT** you see that Workload Optimization Manager can choose from 4 hosts. When you click **Constraints**, the flyout displays a list of host constraints that currently result in the four potential hosts (1) for this VM.



The list information includes:

- CONSTRAINT TYPE

  Most constraints are boundaries that are inherent in your environment such as a cluster boundaries or a networks, or the can be constraint rules such as discovered HA or DRS rules authored Workload Optimization Manager placement policies (sometimes called *segments*)

- SCOPE NAME

  For a given rule or constraint, the scope to which it was applied.

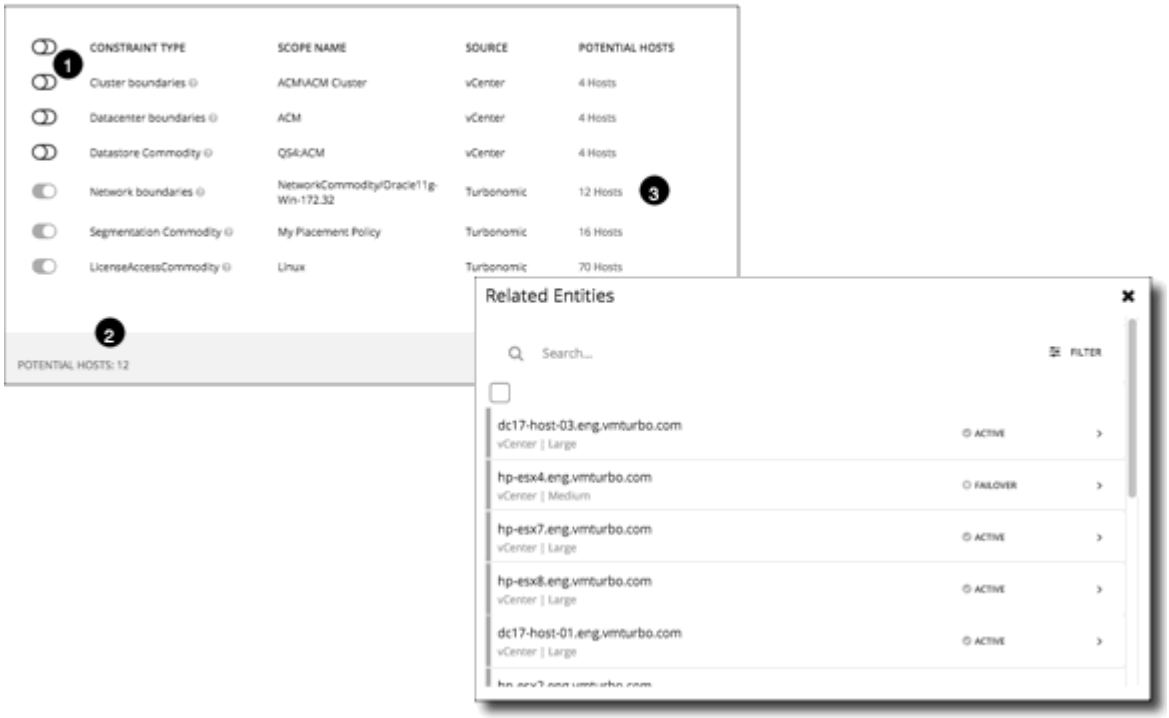
- SOURCE

  If this is a discovered constraint, the source shows the type of target that imposes this constraint. For example, for a DRS rule the source will be vCenter.

- POTENTIAL HOSTS

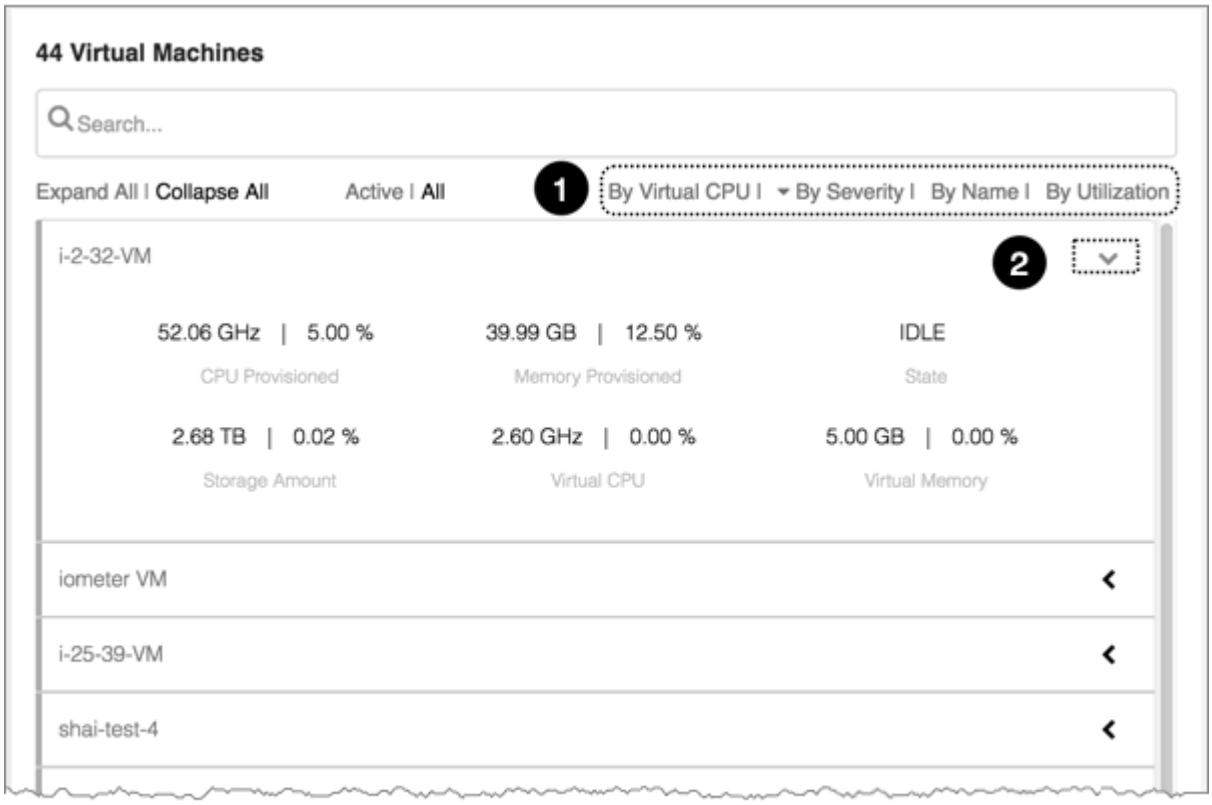
  For the given constraint, how many hosts that constraint allows. To see a list of the potential hosts, click the POTENTIAL HOSTS value.

To dig deeper into how these constraints affect your entity, click **FIND MORE PLACEMENT OPTIONS** (2). This puts you into a simulation mode that you can use to experiment with changing the effective constraints. For example, you might see that a cluster boundary is limiting your placement possibilities, and you would like the option to place the current VM on other clusters. Armed with this information, you could navigate to Policies and create a Merge Cluster policy.



In this mode you can enable and disable different combinations of constraints (1). As you do, the **POTENTIAL HOSTS** (2) label updates to show how many hosts are available to your entity. For example, by turning off the 4-Host constraints, you have 12 potential hosts for this VM. To see the resulting list of hosts, click the **POTENTIAL HOSTS** label (3).

List of entities

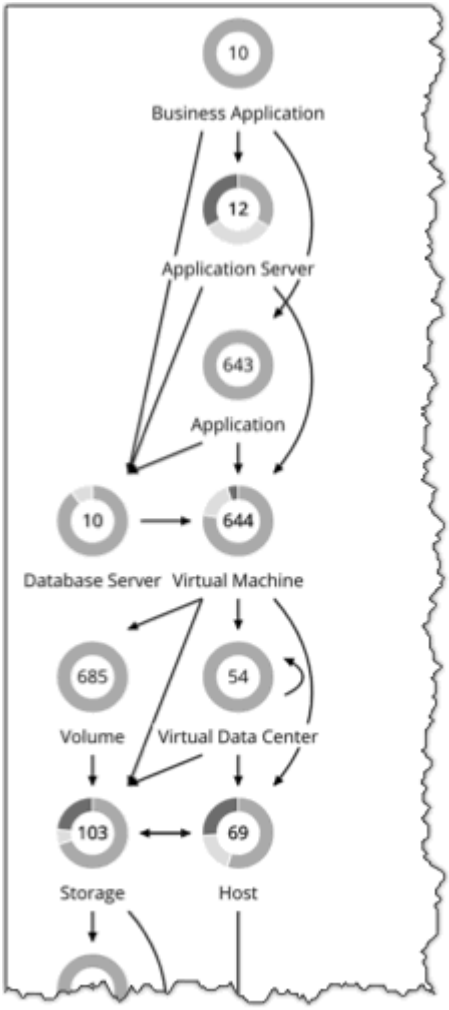


The list of entities is a quick way to drill down to details about your environment.

1. You can sort the list of entities by various categories.
2. You can see specifics about resource consumption or state. For example, you can see the amount of capacity that has been assigned to a VM that is currently idle.

This list always updates to reflect the focus you have selected in the Supply Chain Navigator. When you select an entity type in the supply chain, the entities list updates to show the entities of that type for your current scope. For example, select Host to see a list of hosts in the current scope. For more information, see Navigating with the Supply Chain *(on page 171)*.

Navigating with the supply chain



After you have set the scope of your Workload Optimization Manager session, you can use the supply chain to change the focus of the main view, and see details about different types of entities within the current scope.

Drilling down in a scoped session

When you set a scope to your Workload Optimization Manager session, the **Home Page** shows information about your environment, including:

- Overview

  Charts and lists to give you an overview of your environment for the current scope. This overview corresponds to all the entities in scope.
- Details – Charts that give you a more detailed look at your environment for the given scope
- Policies – Any policies that are defined for the entities in the current scope
- Entity Lists – Details about the entities in the current scope
- Pending Actions – Actions that are pending for any entities in the current scope

The supply chain shows the currently selected tier of entities. The change the focus of the scoped view, select different tiers in the supply chain. The Policies, Entities List, and Pending Actions tabs update to focus on the tier you selected. These tabs show information for all the entities of that type that are in the current scope. For example, if you click the Host tier, these tabs update to show information about the hosts in your current scope.

To zoom in on a specific entity, you can click its name in the Entities List. This sets the scope to that specific entity. To return to the previous scope, use the browser's **Back** button.

Viewing cluster headroom



Cluster headroom shows you how much extra capacity your clusters have to host workloads. When you set the scope to a cluster, the **Home Page** then includes charts that show headroom for that cluster, as well as time to exhaustion of the cluster resources.

To view host cluster headroom:

1. Navigate to the Search page.
2. Choose the **Host Clusters** category.
3. Select the cluster you want to view.
4. When the **Home Page** displays, scroll down to show the headroom charts.

   Make sure you have selected the Host tier in the supply chain navigator.

To calculate cluster capacity and headroom, Workload Optimization Manager runs nightly plans that take into account the conditions in your current environment. The plans identify the optimal workload distribution for your clusters. This can include moving your current VMs to other hosts within the given cluster, if such moves would result in a more desirable workload distribution. The result of the plan is a calculation of how many more VMs the cluster can support.

To calculate VM headroom, the plan simulates adding VMs to your cluster. The plan assumes a certain capacity for these VMs, based on a specific VM template. For this reason, the count of VMs given for the headroom is an approximation based on that VM template.

To specify the templates these plans use, you can configure the nightly plans for each cluster. For more information, see Configuring Nightly Plans *(on page 941)*

Target configuration

A target is an integration partner technology that provides resources or workload management services in your virtual environment. For example, Amazon Web Services (AWS) and VMware vCenter Server are supported targets.

Target configuration specifies the credentials that Workload Optimization Manager uses to connect to targets. Workload Optimization Manager connects through the management protocol that it exposes, such as REST API, SMI-S, XML, or some other management transport. It uses this connection to discover resources, monitor resource utilization, and recommend actions.

After a target is added, Workload Optimization Manager discovers and processes resources and utilization metrics every 10 minutes. This mechanism is needed to generate resource optimization actions in response to continuous changes in resources and utilization over time.

> **NOTE:**
> When a specific release or version of an integration partner technology reaches end-of-life (EOL) or its end-of-support date, Workload Optimization Manager no longer provides support for that version. Workload Optimization Manager follows integration partners' official EOL timeline for version support. Targeting a non-supported version, or one that is no longer supported by the integration partner, is at your own risk.

Supported targets

Workload Optimization Manager supports the following targets.

- Applications and Databases
  - Apache Tomcat 7.x, 8.x, or 8.5.x
  - Cisco AppDynamics 4.1+
  - Datadog
  - Dynatrace 1.1+
  - IBM WebSphere Application Server 8.5+
  - Instana release 242 or later
  - JBoss Application Server 6.3+
  - JVM 6.0+
  - SQL Server 2012, 2014, 2016, 2017, 2019, or 2022
  - MySQL 8.0
  - New Relic

  - Oracle 19c or 21c
  - Oracle WebLogic 12c
- Container Platform
  - Kubernetes 1.27 up to the latest supported GA version, including (but not limited to) Amazon Elastic Kubernetes Service (EKS), Azure Kubernetes Service (AKS), Google Kubernetes Engine (GKE), IBM Cloud Kubernetes Service (IKS), and Rancher
  - Red Hat supported GA versions of Red Hat OpenShift 4.1x, including (but not limited to) Red Hat OpenShift Kubernetes Service (ROKS), Red Hat OpenShift on AWS (ROSA), Red Hat OpenShift on Azure (ARO), and Red Hat OpenShift on Google
- Custom
  - Data Ingestion Framework
- Fabric and Network
  - Cisco UCS Manager 3.1+
  - HPE OneView 3.00.04
- Guest OS Processes
  - SNMP
  - WMI: Windows 7, 8, 8.1, 10, 2008 R2, 2012, 2012 R2, 2016, or 2019
- Hyperconverged
  - Cisco HyperFlex 3.5
  - Nutanix Community Edition
  - VMware vSAN
- Hypervisors
  - IBM PowerVM
    - HMC 10: Power8, Power9, or Power10
    - HMC 9: Power8 or Power9 only; Power7 LPARs and Systems are ignored.
  - Microsoft Hyper-V 2012 R2, 2016, 2019, or 2022
  - VMware vCenter 7.0 or 8.0
- IT Management
  - Flexera
    - Flexera One SaaS
    - FlexNet Manager Suite 2023 R2 or above
  - IBM Apptio
  - ServiceNow

    Certified for ServiceNow versions:
    - Utah
    - Vancouver
    - Washington DC
- Observability
  - Prometheus
- Orchestrator
  - Action Script
- Private Cloud
  - VMM for Microsoft System Center 2012 R2, System Center 2016, System Center 2019, or System Center 2022

- Public Cloud
  - Amazon Web Services (AWS)
  - AWS Billing
  - Azure Service Principal
  - Azure Billing
  - Microsoft Enterprise Agreement
  - Google Cloud
  - Google Cloud Billing
- Storage
  - EMC ScaleIO 2.x or 3.x
  - EMC VMAX using SMI-S 8.1+
  - EMC VPLEX Local Architecture with 1:1 mapping of virtual volumes and LUNs
  - EMC XtremIO XMS 4.0+
  - HPE 3PAR InForm OS 3.2.2+, 3PAR SMI-S, 3PAR WSAPI
  - IBM FlashSystem running on Spectrum Virtualize 8.3.1.2 or later (8.4.2.0 or later recommended)
  - NetApp Cluster Mode using ONTAP 8.0+ (excluding AFF and SolidFire)
  - Pure Storage F-series and M-series arrays running Purity 5.3.6 and 6.4.4 (Pure API 1.6)
- Virtual Desktop Infrastructure
  - Horizon 7.0 or higher

Transport Layer Security requirements

Workload Optimization Manager requires Transport Layer Security (TLS) version 1.2 to establish secure communications with targets. Most targets should have TLS 1.2 enabled. However, some targets might not have TLS enabled, or they might have enabled an earlier version. In that case, you will see handshake errors when Workload Optimization Manager tries to connect with the target service. When you go to the Target Configuration view, you will see a Validation Failed status for such targets.

If target validation fails because of TLS support, you might see validation errors with the following strings:

- `No appropriate protocol`

  To correct this error, ensure that you have enabled the latest version of TLS that your target technology supports. If this does not resolve the issue, contact Cisco Technical Support.
- `Certificates do not conform to algorithm constraints`

  To correct this error, refer to the documentation for your target technology for instructions to generate a certification key with a length of 2048 or greater on your target server. If this does not resolve the issue, contact Cisco Technical Support.

Enabling targets

In Workload Optimization Manager, a probe is a platform component that connects to a target. A probe discovers the target's entities, loads the entities into the Workload Optimization Manager supply chain, and runs actions in the target environment. Workload Optimization Manager includes many probe components that you can use to connect Workload Optimization Manager with your environment.

Depending on your installation method, Workload Optimization Manager may enable probes by default. Each probe uses resources in your Workload Optimization Manager installation. If probes are not needed, consider disabling them. If there are disabled probes that you need, enable them to put them into service.

To enable or disable probes, edit the `cr.yaml` file. You can add new probes, change the values of the `enabled:` properties, and apply changes to reload the platform components.

For more information, see (Optional) Enabling and Disabling Probe Components *(on page 95)*.

Duplicate targets

Workload Optimization Manager does not support duplicate instances of the same target. When configuring targets, do not configure two or more targets to the same address in your environment. For example, do not configure two different targets to the same AWS account or vCenter Server instance.

If you configured duplicate targets, actions can fail to execute with an error that begins: `Analysis Exception occurred...`

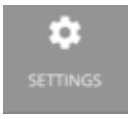
To resolve this issue, identify the duplicate targets, and delete them until you have only one target for each address.

Configuring targets

To configure a target, choose the target type and specify the required settings. Workload Optimization Manager then discovers and validates the target, and then updates the supply chain with the entities that the target manages.

Configuring a target

1. Navigate to the Settings Page.

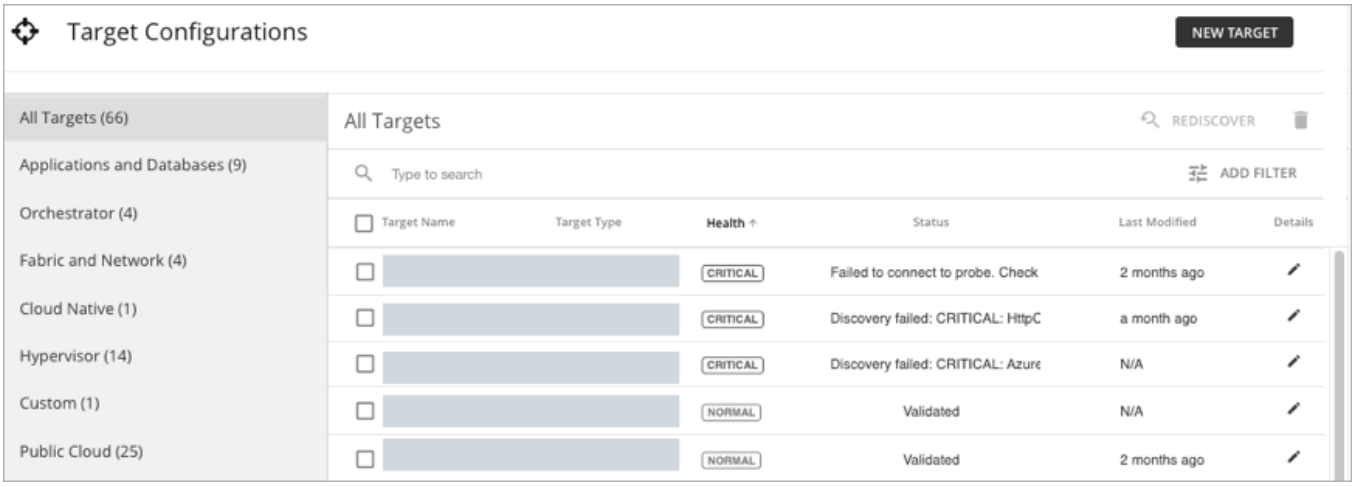
   Click to navigate to the Settings Page. From there, you can perform a variety of Workload Optimization Manager configuration tasks.

2. Choose Target Configuration.

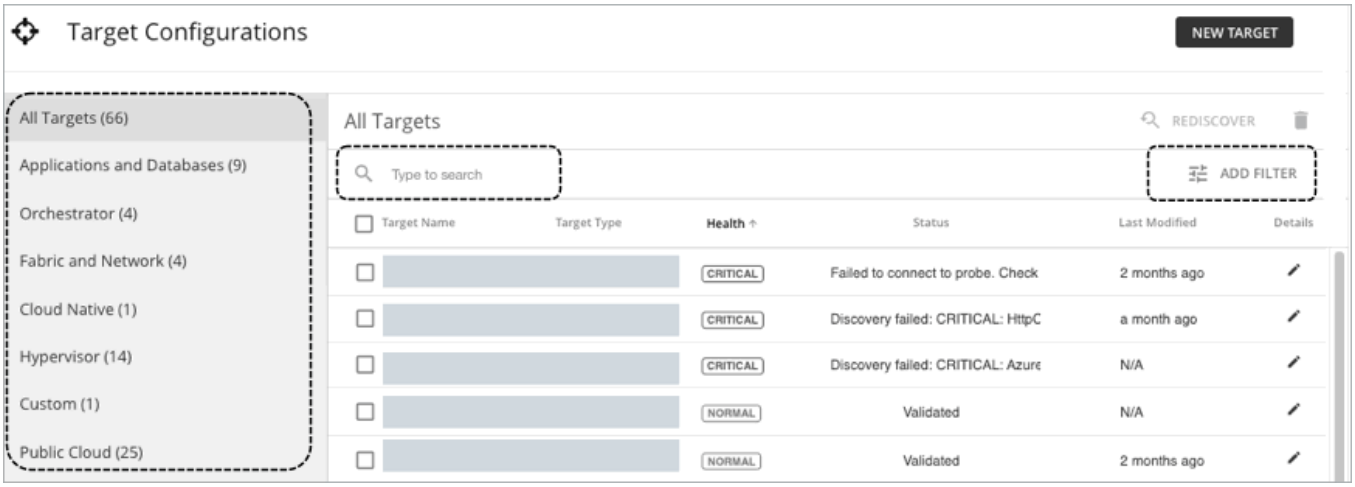
   Click to navigate to the Target Configuration Page.

3. Review the list of targets.

   This page lists all the targets that you currently have configured for Workload Optimization Manager in a sortable table. The table is sorted by target health by default. You can inspect or edit these targets, or add a new target.

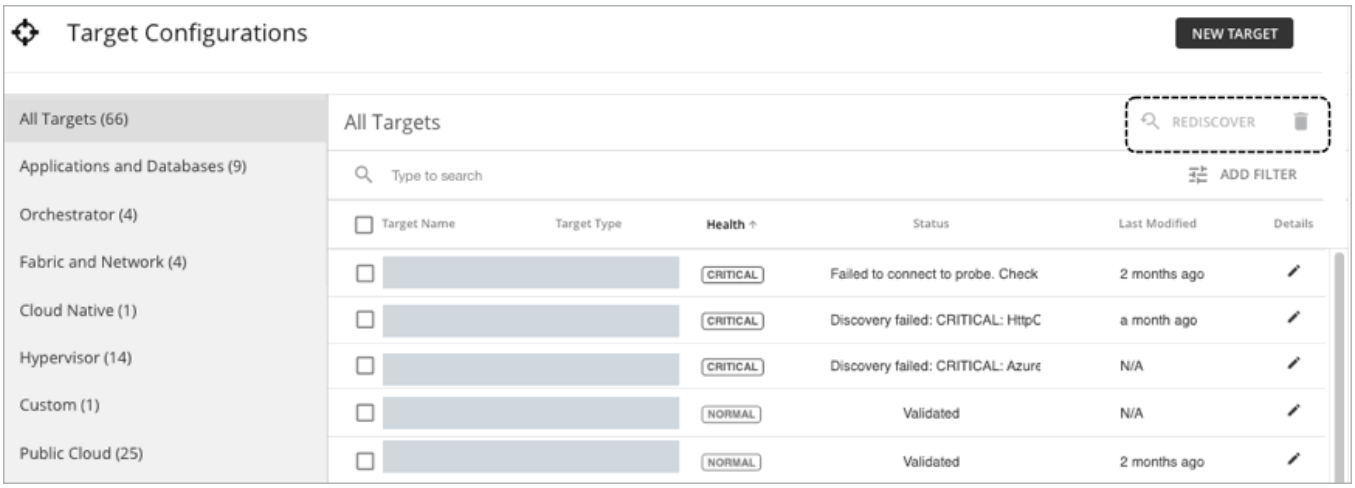
4. Filter the list of targets.



For a long list of targets, you can:

- Filter targets by target type.
- Use Search to filter targets by Target Name using a text string (partial matching is supported).
- Use Filter to filter targets by status (for example, only show validated targets). You can also use Filter to filter targets by target type or health.

5. Select one or more targets to work with.



When you select a target you can:

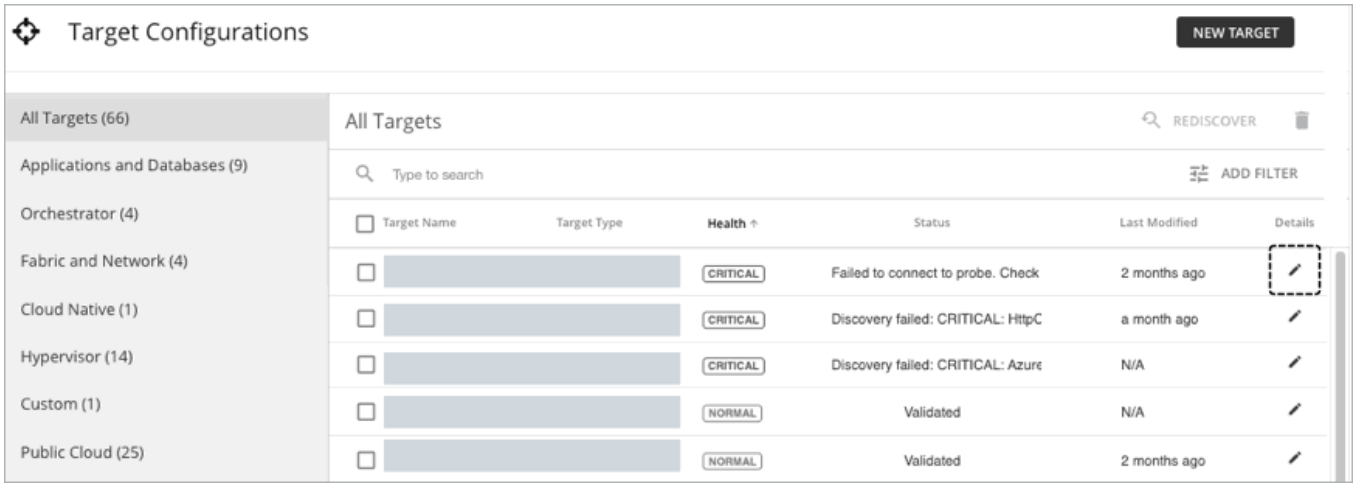
- Rediscover

  Direct Workload Optimization Manager to fully discover the entities that this target manages. This will rebuild the topology that is associated with this target.

- Delete

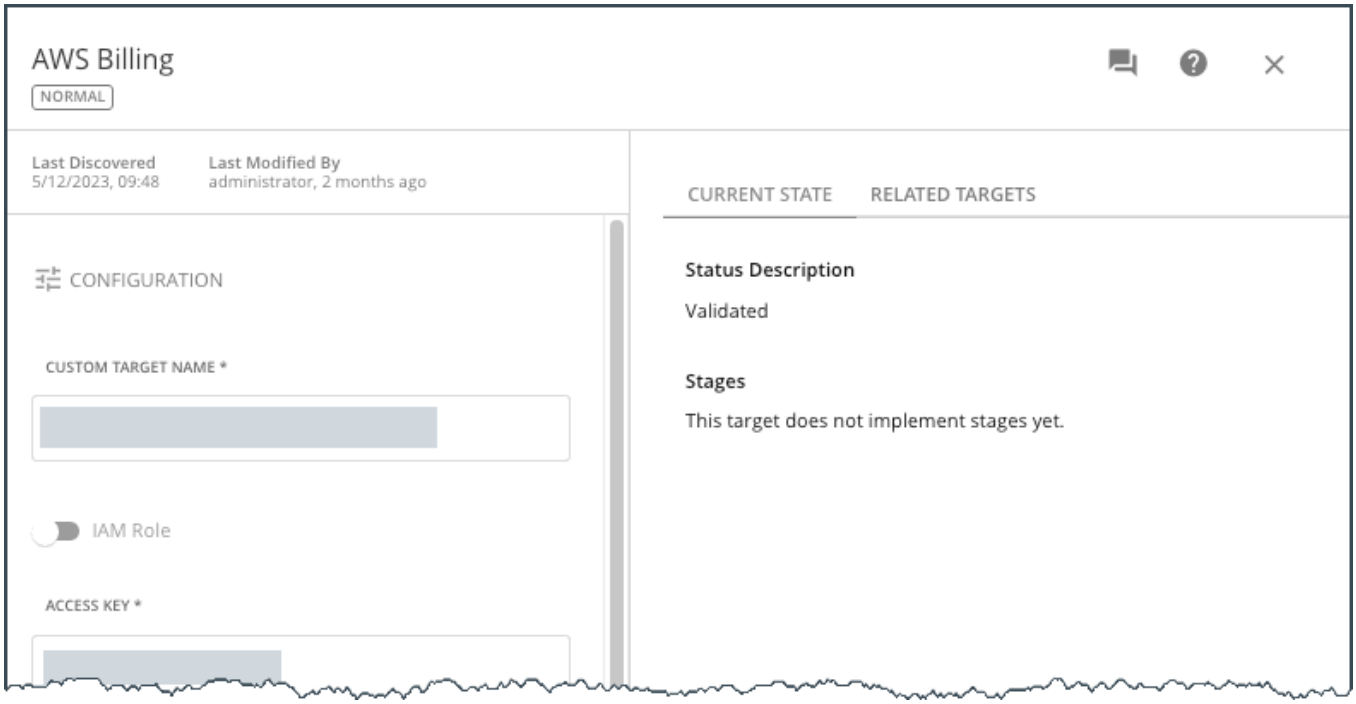
  When you delete a target, Workload Optimization Manager removes all the associated entities from the supply chain.

6. View or edit the target details by clicking the icon under the **Details** column.

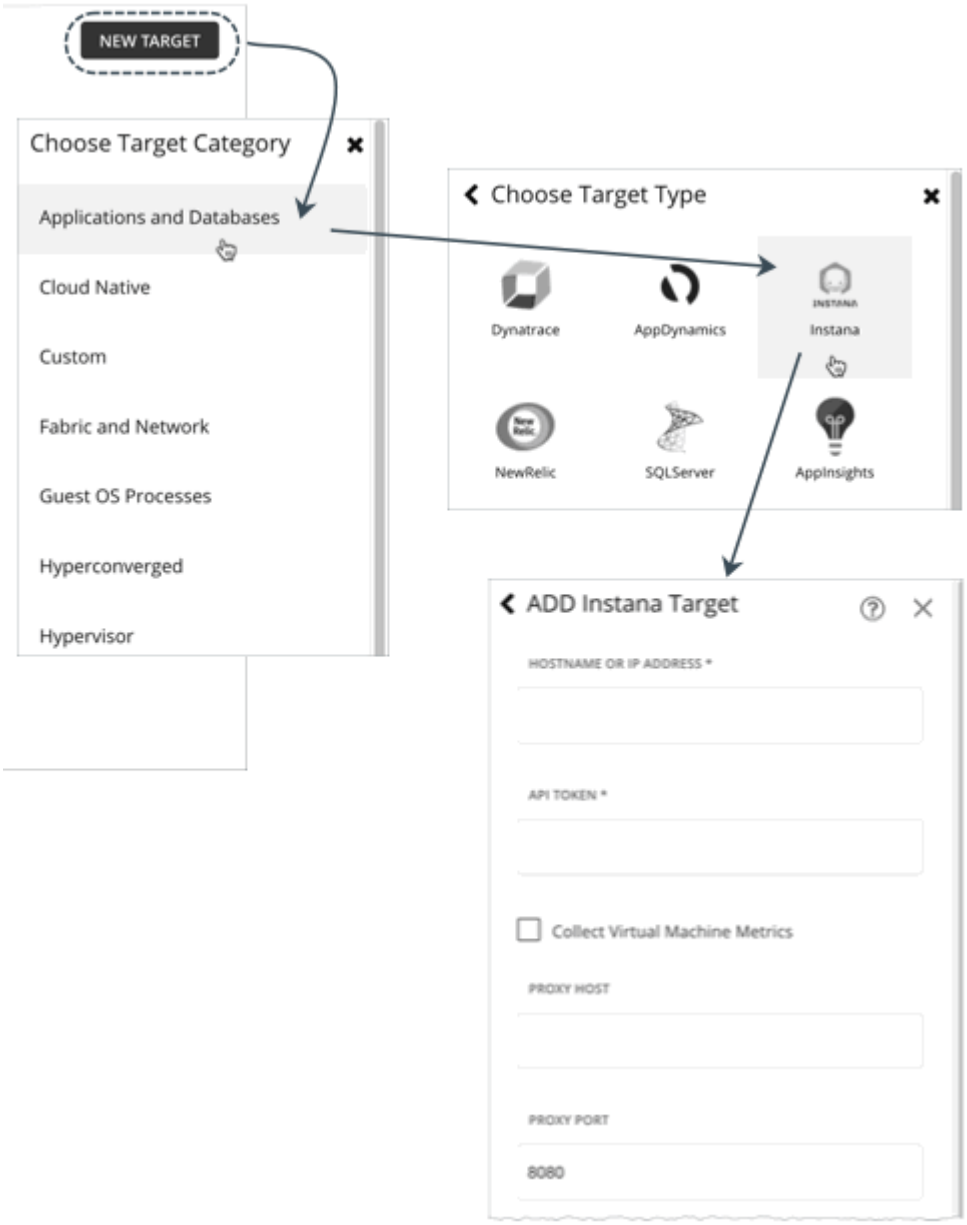


Use the target details page to view or edit target details. In addition to target details, this page shows the last discovered and last modified dates, target health, state, and any related targets.

NOTE:
A value of "N/A" in the Last Modified column indicates that the target is a derived target, and therefore the details cannot be edited.



7. Create a new target and add it to Workload Optimization Manager.



Click **New Target**, select the target category and type, and then provide the address and credentials for that target.

After you add the target, the Target Configuration page refreshes to show the current validation status.

- Validating

  Validation is in progress.

- Validated

  Validation was successful. Workload Optimization Manager can now monitor the target and will start discovering the entities that the target manages.

- Validation Failed

  Validation was unsuccessful. Expand the target to see additional information.

  Workload Optimization Manager regularly checks the status of your targets. If target discovery or validation fails, the Target Configuration page updates the status. Under some circumstances, the target can become discoverable or valid again, but the status does not update. In this case, select the target and then click **Rediscover** or **Validate**.

Applications and databases targets

Applications and Databases targets support domains of particular application servers that are controlled by management servers. For such managed domains you will add the management server as a target, and Workload Optimization Manager will discover the managed application servers.

> **NOTE:**
> As it manages your applications environment, Workload Optimization Manager discovers connected application components to stitch them into a supply chain of entities. For connections that are made by name and not IP address, Workload Optimization Manager makes DNS calls to resolve these names to IP addresses. This can happen during repeated discovery cycles.

Supply chain

Applications and Databases targets add Business Application, Business Transaction, Service, Application Component, Application Server, and Database Server entities to the supply chain. You can navigate to the associated target page to see how these entities map to the target nomenclature.

Apache Tomcat

Workload Optimization Manager supports connecting to individual Apache Tomcat targets. Workload Optimization Manager connects to the Tomcat process as a remote client via remote JMX access. Target configuration includes the port used by the JMX/RMI registry.

Prerequisites

- Apache Tomcat 7.x, 8.x, or 8.5.x
- A valid JMX user account for the Tomcat server

  If Tomcat security is enabled, this must be a Tomcat JMX user with a `readonly` role.
- For VMware environments, VMware Tools must be installed on the VM that hosts the Tomcat server. For Hyper-V environments, Hyper-V Integration Services must be installed

  This ensures that the VM hosting the Tomcat server can get its IP address.
- Remote JMX access is enabled through a port that is opened to the firewall
- Discovered infrastructure

  Workload Optimization Manager discovers Tomcat servers that are running on VMs or containers. The hosting VM or container must already be in your Workload Optimization Manager inventory.

  To set the target for a server running on a VM, you must have first discovered the hosting VM through a hypervisor target. To set the target for a server running in a container, you must have configured container discovery for Tomcat applications.

Configuring JMX remote access

Workload Optimization Manager monitors and controls the Tomcat server via JMX Remote access. You must configure a JMX Remote port.

Note that to work with a firewall you should also set the RMI Server port — If you don't set an RMI port, then JMX sets an arbitrary *ephemeral port*, and you can't guarantee that the port will be open to your firewall.

There are two ways to set JMX Remote port on Linux platforms:

- Ports specified as system properties

  You can set the port via the system property, `com.sun.management.jmxremote.port`. For example:

```
com.sun.management.jmxremote.port=8050
```

A common way to set this property is to declare it in the CATALINA_OPTS system variable – You can set this in the `setenv.sh` script. For example:

```
CATALINA_OPTS="$CATALINA_OPTS

-Dcom.sun.management.jmxremote

-Dcom.sun.management.jmxremote.port=8050"

export CATALINA_OPTS
```

Note that this sets the JMX Remote port, but it does not set the RMI Server port – Tomcat startup will specify an ephemeral port for the RMI server.

- Ports specified in a JMX Remote Lifecycle Listener

  This listener component fixes the ports used by the JMX/RMI Server. When you configure the listener, you specify both the JMX Remote port and the RMI Server port. This is the preferred method when working with a firewall. For more information, see the Apache Tomcat documentation.

On Windows, the typical installation is with Tomcat as a service. There are two ways to set the JMX Remote port:

- Via setenv.bat

  Add the property to the CATALINA_OPTS environment variable:

  ```
  set "CATALINA_OPTS=%CATALINA_OPTS% -Dcom.sun.management.jmxremote.port=8050"
  ```

- Use the Tomcat configuration utility (tomcat7w or tomcat8w)

  Set the port with the following command:

  ```
  -Dcom.sun.management.jmxremote.port=8050"
  ```

To discover the JMX port that is set to an already running Tomcat, you can look in the following locations:

- For Linux platforms, look in the configuration files – Either:
  - `setenv.sh` – Assuming you configured the port by adding it to the CATALINA_OPTS environment variable
  - `$CATALINA_HOME/conf/server.xml` – Assuming you configured a JMX Remote Lifecycle Listener in this file
- For Windows platforms, look in:
  - `setenv.bat` – Assuming you configured the port by adding it to the CATALINA_OPTS environment variable
  - The Windows registry – Assuming you installed Tomcat as a Windows service using the Tomcat Configuration utility

Adding a Tomcat target

You can add an individual Tomcat server as a target, or you can add all matching servers within a given scope.

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **Tomcat**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Display name

     Specify a name that uniquely identifies this connection.

This name is for display purposes only and does not need to match any name in Tomcat.

- Username

  Specify the username of an account with the Admin role.

- Password

  Specify the password of an account with the Admin role.

- Scope

  Specify the scope Workload Optimization Manager uses for application discovery.

  The scope is a group of applications that are stitched to the underlying VMs when the VMs are discovered as part of a separate Workload Optimization Manager target.

  If you set a scope, Workload Optimization Manager searches for virtual machines in the selected group.

  Workload Optimization Manager can monitor up to 500 virtual machines in a group. If you have more than 500 virtual machines in your environment, split them across smaller groups and then add those groups as individual targets.

- JMX remote port

  Specify a JMX port that is set to an already running Tomcat process.

- Authenticate all database servers

  If you select this option, Workload Optimization Manager attempts to authenticate all database servers in the selected scope. If Workload Optimization Manager is unable to authenticate a database server, the target is not added and no data is collected.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Application Component
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
  - Heap

    Heap is the portion of a VM or container's memory allocated to individual applications.
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
  - Threads

    Threads is the measurement of thread capacity utilized by applications.
  - Connection

    Connection is the measurement of database connections utilized by applications.
  - Remaining GC capacity

    Remaining GC capacity is the measurement of Application Component uptime that is *not* spent on garbage collection (GC).
- Virtual Machine
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.

- Virtual CPU (vCPU)

  Virtual CPU is the measurement of CPU that is in use.

Actions

Workload Optimization Manager supports the following actions:

- Application Component (Tomcat Application)
  - Resize Heap

    This action can only be executed outside Workload Optimization Manager.
  - Resize Thread Pool

    This action can only be executed outside Workload Optimization Manager.
  - Resize Connection Capacity

    This action can only be executed outside Workload Optimization Manager.
- Virtual Machine
  - Provision additional resources (VMem, VCPU)
  - Move Virtual Machine
  - Move Virtual Machine Storage
  - Reconfigure Storage
  - Reconfigure Virtual Machine
  - Suspend VM
  - Provision VM

Cisco AppDynamics

Workload Optimization Manager supports workload management of the application infrastructure that is monitored by Cisco AppDynamics by adding the AppDynamics instance to Workload Optimization Manager as a target.

The Workload Optimization Manager integration with AppDynamics provides a full-stack view of your environment, from application to physical hardware. With information obtained from AppDynamics, Workload Optimization Manager is able to make recommendations and take actions to both assure performance and drive efficiency with the full knowledge of the demands of each individual application.

In its default configuration, the AppDynamics target collects up to 1100 AppDynamics nodes within the default collection period when a proxy is used, and up to 5000 nodes when no proxy is required. Larger AppDynamics environments are expected to take longer than one cycle to collect complete data.

NOTE:
For container platform environments, Workload Optimization Manager stitches New Relic, AppDynamics, Instana, and Dynatrace containerized application components into the supply chain to provide a unified view of your applications. For more information, see Container Platform Targets *(on page 234)*.

Prerequisites

- Cisco AppDynamics 4.1+
- A valid AppDynamics user account

  For all types of application instances, the service account must have the `Read Only User` role. For monitoring database instances, this user must also have the `DB Monitoring User` role.

NOTE:
In newer versions of AppDynamics where these roles are available, they should be used instead:

- Applications and Dashboards Viewer
- DB Monitoring User
- Server Monitoring

To use a custom role, ensure that the role has the `View Server Visibility` permission for both applications and databases.

AppDynamics database servers

AppDynamics also monitors database servers. For your database servers to be correctly stitched to the rest of your environment, you must:

- Enable enhanced metric collection.

  For Hyper-V hosts, you must install Hyper-V Integration Services on the target VM hosting the database. For more information, refer to the following integration services TechNet article:

  `https://technet.microsoft.com/en-us/library/dn798297%28v=ws.11%29.aspx`

  For VMware hosts, you must install VMware Tools on the target VMs.
- Ensure that the database name in AppDynamics is resolvable to an IP address by the Workload Optimization Manager instance.

  You may need to make changes to your DNS or the file `/etc/resolv.conf` on the Workload Optimization Manager instance.

Adding an AppDynamics target

NOTE:
It is possible to monitor certain applications or database servers with both AppDynamics and Workload Optimization Manager, but this must be avoided as it causes the entities to appear duplicated in the market.

If an application is monitored by AppDynamics, do not add it as a separate Workload Optimization Manager application target.

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **AppDynamics**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Hostname or IP address

     Specify the hostname or IP address of the AppDynamics controller instance.
   - Port number

     Specify the port that Workload Optimization Manager uses to connect to the AppDynamics controller. By default, the HTTP port is 80 and the HTTPS port is 443. For SaaS-based AppDynamics instances, use port 443.
   - Username or API client name

     Specify the username or account ID with the Read Only User and DB Monitoring User permissions. For OAuth authentication, the username must be user-defined as an API Client.

     The username can be found on the **License > Account** page in AppDynamics.

The format must be *Central ID@tenant*, where *Central ID* can be either a username or an email ID and *tenant* is not a domain. If the Central ID is an email ID (such as *user@domain*), you must type *%40* instead of "@" in front of the domain since Central IDs must be URL encoded. The "@" symbol is accepted as a delimiter between the Central ID and tenant only.

Examples: *<username>@tenant* or *<user%40domain>@tenant*.

NOTE:
The Central ID cannot contain any of the following special characters:

```
\ / " [ ] : | < > + = ; , ? * , ' tab space @
```

For Central IDs containing the "@" symbol, URL encode the "@" character as %40.

- Password or client secret

  Specify the password for the account used to connect to the AppDynamics instance. For OAuth authentication, specify the client secret key.

  NOTE:
  The password cannot contain any of the following special characters:

  ```
  \ / " [ ] : | < > + = ; , ? * , ' tab space @
  ```

  For passwords containing the "@" symbol, URL encode the "@" character as %40.

- Collect virtual machine metrics

  If you select this option, Workload Optimization Manager uses the virtual machine metrics that it collects from this target, instead of the metrics that can be collected from related infrastructure targets.

- Connect using HTTPS

  If you select this option, Workload Optimization Manager connects to the target servers using HTTPS. Make sure that the required certificate is configured for use on the host.

  This option is selected by default.

- Use API Client (OAuth)

  If you select this option, Workload Optimization Manager enables Open Authorization (OAuth) token-based authentication for the target connection. For more information, see the AppDynamics documentation.

- Validate server certificates

  If you select this option, Workload Optimization Manager verifies the target certificate and proxy, if in use.

  This option is selected by default.

- Proxy configuration

  Specify proxy information only if you connect through a proxy.

  - Proxy hostname or IP address

    Specify the address of the proxy used for this target.

  - Proxy port

    Specify the port to use with the proxy used for this target. By default, the port is 8080.

  - Proxy username

    Specify the username to use with the proxy.

  - Proxy password

    Specify the password to use with the proxy.

  - Connect using HTTPS

    Select this option if Workload Optimization Manager will connect to the proxy through HTTPS.

For more information about creating API client users, see the AppDynamics Documentation.

Entity mapping

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

AppDynamics	Workload Optimization Manager
Business Application	Business Application
Business Transaction	Business Transaction
Tier	Service
Node	Application Component
Database	Database Server
Machine (when the machine type is Container)	Container
Server	Virtual Machine

Monitored resources

Workload Optimization Manager monitors the following resources:

NOTE:
The exact resources that are monitored will differ based on application type. This list includes all of the resources that you may see.

- Application Component
  - Connection

    Connection is the measurement of database connections utilized by applications.
  - Heap

    Heap is the portion of a VM or container's memory allocated to individual applications.

    This commodity applies to Java, .NET, and Node.js only.
  - Remaining GC capacity

    Remaining GC capacity is the measurement of Application Component uptime that is *not* spent on garbage collection (GC).
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
  - Threads

    Threads is the measurement of thread capacity utilized by applications.
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.

    This commodity applies to Java, .NET, and Node.js only.
  - Virtual memory (vMem)

Virtual memory (vMem) is the measurement of memory that is in use.

This commodity applies to Java, .NET, and Node.js only.

- Business application
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
- Business transaction
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
- Database server
  - Connection

    Connection is the measurement of database connections utilized by applications.

    This commodity applies to MongoDB only.
  - DB cache hit rate

    DB cache hit rate is the measurement of Database Server accesses that result in cache hits, measured as a percentage of hits versus total attempts. A high cache hit rate indicates efficiency.

    This commodity applies to SQL and Oracle only.
  - Transaction log

    Transaction log is the measurement of storage capacity utilized by a Database Server for transaction logging.

    This commodity applies to SQL only.
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

    This commodity applies to SQL, MySQL, and Oracle only.
- Service
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

    For container platform environments, this is the desired *weighted average* response time of all Application Component replicas associated with a Service.
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

    For container platform environments, this is the maximum number of transactions per second that each Application Component replica can handle.
- Virtual machine
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.

NOTE:
To collect data, a machine agent must be present and database hardware monitoring must be enabled.

- Virtual memory (vMem)

  Virtual memory (vMem) is the measurement of memory that is in use.

  NOTE:
  To collect data, a machine agent must be present and database hardware monitoring must be enabled.

For a VM, the resources you see depend on how the VM is discovered, and whether the VM provides resources for an application that is discovered by this target:

- If the VM hosts an application that is discovered through this target, then you see VM metrics that are discovered through this target.
- If the VM is discovered through a different target, and it does not host any application discovered through this target, you see VM metrics that are discovered through that different target.
- If the VM is discovered through this target, but it does not host any application that is discovered through this target, then Workload Optimization Manager does not display metrics for the VM.

Actions

NOTE:
The specific actions that Workload Optimization Manager recommends can differ, depending on the processes that Workload Optimization Manager discovers.

For other application components, Workload Optimization Manager can recommend actions based on the resources it can discover for the application. For example, Node.js® applications report CPU usage, so Workload Optimization Manager can generate vCPU resize actions and display them in the user interface.

Workload Optimization Manager supports the following actions:

- Application component
  - Resize Heap

    This action can only be executed outside Workload Optimization Manager.
- Database server
  - Resize Connections

    This action can only be executed outside Workload Optimization Manager.
  - Resize Database Memory (DBMem)

    This action can only be executed outside Workload Optimization Manager.

NOTE:
For different types of Database Servers, the AppDynamics target returns different metrics. This affects Workload Optimization Manager actions as follows:

- MySQL:

  For MySQL database servers, analysis does not generate resize actions for DB Memory, Connections, or Transaction Log. The target does not discover DB Cache Hit Rate, DB Memory, Connections, or Transaction Log.

- SQL Server:

  For SQL database servers, analysis does not generate resize actions for DB Memory, Connections, or Transaction Log. The target does not discover DB Memory or Connections.

- MongoDB:

  For MongoDB database servers, analysis does not generate resize actions for DB Memory, Connections, or Transaction Log. The target does not discover DB Cache Hit Rate, DB Memory, Transactions, or Transaction Log.

- Oracle:

  For Oracle database servers, analysis does not generate resize actions for DB Memory, Connections, or Transaction Log. The target does not discover DB Memory, Connections, or Transaction Log.

Datadog

The Workload Optimization Manager-Datadog integration combines application performance data with infrastructure data that Workload Optimization Manager collects from Datadog to generate full-stack aware, automatable actions that optimize application performance and reduce cost.

By using Workload Optimization Manager analytics and automation capabilities, you can dynamically allocate resources to ensure optimal application performance, while you minimize infrastructure costs and comply with licensing and other operational policies.

Prerequisites

- Set up a valid Datadog user account with read-only API and application key access.
- To discover tags as Business Applications, configure Datadog tags *(on page 193)*.

Adding a Datadog target

NOTE:
If an application is monitored by Datadog, do not add it as a separate Workload Optimization Manager application target.

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **Datadog**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Display name

     Specify a name that uniquely identifies this connection.

     This name is for display purposes only and does not need to match any name in Datadog.

     The name field can include only alphanumeric, space, or hyphen characters and must have a unique name when multiple Datadog targets are added.

IMPORTANT: If your Datadog target was created before Workload Optimization Manager version 3.12.2, the display name field was set to the target hostname by default. Because most hostnames include at least one period ("."), the display names field for most targets that were created before Workload Optimization Manager version 3.12.2 must be edited before you can make any further changes to the target configuration. You can replace the periods with alphanumeric, space, or hyphen characters, or replace the display name entirely with an option that satisfies these constraints.

- Hostname

  Specify the address of the Datadog server. For more information, see the Datadog documentation.
- API key

  Specify the API key provided by the Datadog platform. For more information, see the Datadog documentation.
- Application key

  Specify the Application key provided by the Datadog platform. For more information, see the Datadog documentation.
- Business application Datadog tag

  Specify the tags that you configured in Datadog to identify Business Applications.

  To configure tags, follow the steps outlined in this topic *(on page 193)*.

  The tag name must begin and end with alphabetical characters and can contain hyphens, dots, and underscores.

  Custom tag values that have a matching tag name with the target configuration value are identified as a valid Business Application.
- Collect virtual machine metrics

  If you select this option, Workload Optimization Manager uses the virtual machine metrics that it collects from this target, instead of the metrics that can be collected from related infrastructure targets.
- Collect tag information

  If you select this option, Workload Optimization Manager collects tags assigned to virtual machines discovered from this target.

  Tags are key-value pairs that you define in the target and then assign to different entities.

  You can use these tags in Workload Optimization Manager when you search for, filter, or group entities discovered from the target. To view the collected tags, set the scope to a particular entity and then check the **Related Tag Information** chart.
- Proxy configuration

  Specify proxy information only if you connect through a proxy.
  - Proxy hostname or IP address

    Specify the address of the proxy used for this target.
  - Proxy port

    Specify the port to use with the proxy used for this target. By default, the port is 8080.
  - Proxy username

    Specify the username to use with the proxy.
  - Proxy password

    Specify the password to use with the proxy.
  - Connect using HTTPS

    Select this option if Workload Optimization Manager will connect to the proxy through HTTPS.

Supported applications

Workload Optimization Manager discovers the following application types and associated commodities from the Datadog target:

Application type	Commodities
Stand-Alone Java Application	Virtual CPU, Virtual Memory, Heap, Remaining GC Capacity
Tomcat	Virtual CPU, Virtual Memory, Response Time, Transactions, Heap, Remaining GC Capacity

Entity mapping

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

Datadog	Workload Optimization Manager
Tag	Business Application
Resource	Business Transaction
Service	Service
App	Application Component
Container	Container
Host	Virtual Machine

Entities may not appear in the supply chain if the following requirements are not met:

- To discover Business Applications and Services and stitch them with the underlying entities in the supply chain, you must configure tags *(on page 193)* in Datadog.
- Workload Optimization Manager-Datadog support is focused on servlet-based applications and services. If your application components utilize technologies other than servlet-based ones (such as Tomcat, WebSphere, or WebLogic), they are not visible in the supply chain. Similarly, Business Applications that rely on technologies other than servlet-based ones will appear orphaned in the Workload Optimization Manager supply chain.
- Workload Optimization Manager-Datadog support extends solely to containers deployed through the Kubernetes deployment method. Containers deployed through alternate methods (like Rancher, Docker Swarm, and so on) will not appear in the Workload Optimization Manager supply chain.
- Virtual Machines in Workload Optimization Manager are discovered by cloud targets (AWS, Azure, Google Cloud) or on-prem targets (such as VMware or Microsoft Hyper-V). Targets that host the Virtual Machines need to be added as targets in Workload Optimization Manager in order to appear in the supply chain after the Datadog target is discovered.
- When VM entities identified by the Datadog target are stitched with VM entities discovered by the cloud or on-prem target, you can view the link between the VMs and the application components within the supply chain.

Monitored resources

Workload Optimization Manager monitors the following resources:

NOTE:
The exact resources that are monitored will differ based on application type. This list includes all of the resources that you may see.

- Business Application
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

- Business Transaction
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
- Service
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
- Application Component
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.

    This commodity applies only to JVM applications.
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.

    This commodity applies only to Tomcat applications.
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
  - Heap

    Heap is the portion of a VM or container's memory allocated to individual applications.
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
  - Remaining GC capacity

    Remaining GC capacity is the measurement of Application Component uptime that is *not* spent on garbage collection (GC).
- Container
  - Virtual memory (vMem)

    Virtual memory (vMem) is the virtual memory utilized by a container against the memory limit. If no limit is set, node capacity is used.
  - vCPU

    vCPU is the virtual CPU (in mCores) utilized by a container against the CPU limit. If no limit is set, node capacity is used).
- Virtual Machine
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.

- Virtual memory (vMem)

  Virtual memory (vMem) is the measurement of memory that is in use.

Actions

Workload Optimization Manager supports the following actions:

NOTE:
The actions that Workload Optimization Manager recommends might differ, depending on the application type that Workload Optimization Manager discovers and supports for the Datadog target.

- Application Component
  - Resize Heap

    This action can only be executed outside Workload Optimization Manager.

    Actions to resize heap are driven by the application type. For Tomcat applications, analysis only generates resize actions for the Remaining GC Capacity commodity. For JVM applications, analysis only generates resize actions for the Heap commodity.

    For Workload Optimization Manager to generate actions to resize heap, cloud hosting platforms (AWS, Azure, and Google Cloud) or on-prem targets (such as VMware and Microsoft Hyper-V) that host the Virtual Machines must be included as targets in Workload Optimization Manager.

Configuring Datadog tags

For Workload Optimization Manager to discover Business Applications and Services and link them with the underlying entities in the supply chain, configure tags in your Datadog environment.

Configuring tags as business applications

1. Configure application tags at the host level in the main Datadog configuration file (`/etc/datadog-agent/datadog.yaml`).

   For more information about configuring application tags, see the Datadog documentation.
2. If your Datadog environment integrates with Azure, you can import Azure tags into Datadog or set up Azure-related tags in Datadog.
   a. Log in to Datadog.

      https://app.datadoghq.com
   b. Click **Integrations** in the navigation bar and then click the **Integrations** tab.
   c. Select **Azure** from the list of integrations and then click **Azure VM**.
   d. Click the **Configuration** tab. If Azure VM Integration is not installed, follow the installation instructions. Otherwise, click the **main Azure Integration** link to configure the Datadog Agent on your Virtual Machine.
   e. Create a new app registration by clicking **Add New App Registration** or reuse the app registration that you set up for the Azure Service Principal target in Workload Optimization Manager.

      NOTE:
      For more information about app registration and credentials for the Azure Service Principal target, see Setting Up an Azure Service Principal *(on page 432)*.

      The app registration process generates the required credentials, including the Tenant ID, Client ID, and Client Secret Value. Specify these credentials and turn on the **Enable Resource Collection** option.
   f. Click **Create Configuration**.

When you add a Datadog target *(on page 189)* in the Workload Optimization Manager user interface, configure the **Business application Datadog tag** field by specifying the application tags that you configured in Datadog. The value assigned to the application tag serves as the name of the Business Application.

Configuring tags as services

1. Configure `service` and `env` tags.

   You can configure tags at the host level in the main Datadog configuration file (`/etc/datadog-agent/datadog.yaml`) or at the application level.

   For a Tomcat application, a dedicated configuration file for Tomcat can be found at `/etc/datadog-agent/conf.d/tomcat.d/conf.yaml`.

Dynatrace

Workload Optimization Manager supports discovery of applications that are managed by the Dynatrace platform. Workload Optimization Manager includes the discovered information about these applications in its calculations for VM actions.

NOTE:
For container platform environments, Workload Optimization Manager stitches New Relic, AppDynamics, Instana, and Dynatrace containerized application components into the supply chain to provide a unified view of your applications. For more information, see Container Platform Targets *(on page 234)*.

Prerequisites

- Dynatrace 1.1+
- A Dynatrace Server instance

  This instance must be configured to monitor applications that are running in your environment.

  Workload Optimization Manager supports both SaaS and on-prem Dynatrace server installations.
- Managed VMs that host applications managed by Dynatrace

  For Workload Optimization Manager to discover applications through Dynatrace, the applications must be running on VMs in your environment. Also, Workload Optimization Manager targets such as hypervisors or public cloud targets must manage these VMs.
- An API access token with the proper scopes

  Workload Optimization Manager uses the API token to authenticate its calls to the Dynatrace API. This token must have permission to run GET methods using the Dynatrace API, both Version 1 and Version 2. Generate a new generic access token with these scopes:

Workload Optimization Manager Functionality	Required Permissions
Monitoring	– API V1 scopes: • `Access problem and event feed, metrics, and topology` – API V2 scopes: • `Read entities` • `Read metrics`

NOTE:
If you are updating to Workload Optimization Manager version 3.4.2 or later, from a version that is earlier than 3.4.2, you must generate a new API token for each existing Dynatrace target. Then you must enter that token in the target configuration, and validate the target.

If the target still fails to validate after you update the access token, take note of your configuration settings, delete the target, and configure the target again. Be sure to use the new API token that you have generated.

- Custom calculated service metrics

  For Workload Optimization Manager to discover Response Time and Transaction metrics for Dynatrace Application Component, you must configure custom calculated service metrics in Dynatrace. For details, see Dynatrace Custom Calculated Service Metrics *(on page 198)*.

Adding a Dynatrace target

NOTE:
You can manage certain applications or database servers with both Dynatrace and Workload Optimization Manager. Avoid such a configuration because it can cause Workload Optimization Manager to generate duplicate entities in the market.

If you manage an application by using a Dynatrace server, and you configure that Dynatrace server as a Workload Optimization Manager target, ensure you have not added that application as a separate application target in Workload Optimization Manager.

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **Dynatrace**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Hostname or IP address

     Specify the host name or IP of the server for on-prem installations, or the SaaS URL.

     For an on-prem installation of Dynatrace, specify the hostname or IP address with endpoint, separated by a slash. For example, `10.10.10.10/e/b70e3eb2-e82b-4c13-a5a4-560d9865841r`.

     For a SaaS installation, specify the URL without the `{IP_ADDRESS}/e/` statement in the path. For example, `ayz12745.live.dynatrace.com`
   - API token

     Specify the token that Workload Optimization Manager uses to authenticate calls to the Dynatrace API.

     This token must have permission to run GET methods using the Dynatrace API V1 and V2.

     Refer to the Prerequisites *(on page 194)* section for more information.
   - Collect virtual machine metrics

     If you select this option, Workload Optimization Manager uses the virtual machine metrics that it collects from this target, instead of the metrics that can be collected from related infrastructure targets.
   - Collect tag information

     If you select this option, Workload Optimization Manager collects tags that are assigned to application components, business applications, containers, services, and virtual machines discovered from this target.

     Tags are key-value pairs that you define in the target and then assign to different entities.
   - Validate server certificates

     If you select this option, Workload Optimization Manager verifies the target certificate and proxy, if in use.

This option is selected by default.

- Proxy configuration

  Specify proxy information only if you connect through a proxy.

  - Proxy hostname or IP address

    Specify the address of the proxy used for this target.

  - Proxy port

    Specify the port to use with the proxy used for this target. By default, the port is 8080.

  - Proxy username

    Specify the username to use with the proxy.

  - Proxy password

    Specify the password to use with the proxy.

  - Connect using HTTPS

    Select this option if Workload Optimization Manager will connect to the proxy through HTTPS.

Entity mapping

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

Dynatrace	Workload Optimization Manager
Application	Business Application **NOTE:** For Dynatrace Applications, Workload Optimization Manager displays Business Application entities in the supply chain when they are active for at least the past 10 minutes.
Service	Service
Process	Application Component, Database Server
NA	Container
Host	Virtual Machine

Monitored resources

Workload Optimization Manager monitors the following resources:

NOTE:
The exact resources that are monitored will differ based on application type. This list includes all of the resources that you may see.

- Application Component
  - Heap

    Heap is the portion of a VM or container's memory allocated to individual applications.

    This commodity applies only to Java applications.

  - Remaining GC capacity

Remaining GC capacity is the measurement of Application Component uptime that is *not* spent on garbage collection (GC).

This commodity applies only to Java applications.

- Response time

  Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

- Transaction

  Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

- Virtual CPU (vCPU)

  Virtual CPU is the measurement of CPU that is in use.

- Virtual memory (vMem)

  Virtual memory (vMem) is the measurement of memory that is in use.

- Business Application
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
- Container
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.

  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
- Database Server

  For Database Server applications, Workload Optimization Manager discovers metrics for MySQL and SQL Server databases only.

  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.

  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.

  - Database memory (DBMem)

    Database memory (or DBMem) is the measurement of memory that is utilized by a Database Server.

    This commodity applies only to SQL and MySQL databases.

    Actions to resize database memory are driven by data on the Database Server, which is more accurate than data on the hosting VM.

  - DB cache hit rate

    DB cache hit rate is the measurement of Database Server accesses that result in cache hits, measured as a percentage of hits versus total attempts. A high cache hit rate indicates efficiency.

    This commodity applies only to SQL databases.

  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

    This commodity applies only to SQL databases.

- Service
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
- Virtual Machine
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.

  For a VM, the resources you see depend on how the VM is discovered, and whether the VM provides resources for an application that is discovered by this target:
  - If the VM hosts an application that is discovered through this target, then you see VM metrics that are discovered through this target.
  - If the VM is discovered through a different target, and it does not host any application discovered through this target, you see VM metrics that are discovered through that different target.
  - If the VM is discovered through this target, but it does not host any application that is discovered through this target, then Workload Optimization Manager does not display metrics for the VM.

Actions

Workload Optimization Manager supports the following actions:

- Application Component
  - Resize Heap

    This action can only be executed outside Workload Optimization Manager.
- Database Server
  - Resize Database Memory (DBMem)

    This action can only be executed outside Workload Optimization Manager.

    This commodity applies to MySQL only.
- Workload Controller
  - Scale

    Actions associated with a workload controller scale replicas horizontally. This is a natural representation of these actions because the parent controller's container specs and number of replicas are modified. The workload controller then rolls out the changes in the running environment.

    For details, see Workload Controller Scale Actions *(on page 598)*.

Dynatrace custom calculated service metrics

Dynatrace collects important metrics for services with no additional configurations. However, you might need more business or technical metrics that are specific to your application. These metrics can be calculated and derived based on the captured metric data with Dynatrace.

For Workload Optimization Manager to discover Dynatrace application component level metrics (such as Response Time and Transaction), you must configure custom calculated service metrics in Dynatrace.

NOTE:
Custom calculated service metrics use Dynatrace DDU (Davis data units).

Creating an auto tag for service entity types

Before you can create custom calculated metrics, you must first create an auto tag for the service entity types.

Dynatrace Service entities are tagged with `service-autotag:service.custom.tag` after you create the auto tag.

1. Log in to your Dynatrace account.
2. Select **Settings > Tags > Automatically applied tags**.
3. Click **Create tag**.

   Specify the following properties for the tag:
   - **Tag name** - `service-autotag`
4. Click **Add a new rule**.

   Specify the following properties for the rule:
   - **Optional tag value** - `service.custom.tag`
   - **Value Normalization** - `Leave text as-is`
   - **Rule type** - `Entity selector`
   - **Entity selector** - `type(SERVICE)`
5. Click **Save Changes**.

   The tag now appears in the **Settings > Tags > Automatically applied tags** view.
6. Verify that the tag was created.
   - Select any Dynatrace Service.
   - In the **Properties and tags** section, verify that the new tag `service-autotag:service.custom.tag` is listed under **Tags**.

Creating calculated service metrics for response time

NOTE:
Custom calculated service metrics use Dynatrace DDU (Davis data units).

Follow the steps to create the custom service metric `app.responsetime` to track Response time. The dimension `{Service:Instance}` named `service-instance` splits the values of this metric.

1. Log in to your Dynatrace account.
2. Select **Settings > Server-side service monitoring > Calculated service metrics**.
3. Click **Create new metric**.

   Specify the following properties for the metric:
   - **Metric name -** `app.responsetime`
   - **Metric source -** `Response time`
4. In the **Conditions** section, click **Add Condition**.
   Specify the following properties for the condition:

   ```
   Service tag equals service-autotag service.custom.tag
   ```
5. Click **Add**.

   The selected services display in the **Preview** section.

6. Turn on the **Split by dimension** toggle to add dimensions for the metric.

   Specify the following properties:

   - **Dimension value pattern –** `{Service:Instance}`
   - **Dimension name –** `service-instance`

7. Click **Save Metric**.

   The metric now appears in the **Settings > Server-side service monitoring > Calculated service metrics** view.

Creating calculated service metrics for request count

NOTE:
Custom calculated service metrics use Dynatrace DDU (Davis data units).

Follow the steps to create the custom service metric `app.requestcount` to track Request count. The dimension `{Service:Instance}` named `service-instance` splits the values of this metric.

1. Log in to your Dynatrace account.
2. Select **Settings > Server-side service monitoring > Calculated service metrics**.
3. Click **Create new metric**.

   Specify the following properties for the metric:

   - **Metric name –** `app.requestcount`
   - **Metric source –** `Request count`

4. In the **Conditions** section, click **Add Condition**.
   Specify the following properties for the condition:

   ```
   Service tag equals service-autotag service.custom.tag
   ```

5. Click **Add**.

   The selected services display in the **Preview** section.

6. Turn on the **Split by dimension** toggle to add dimensions for the new metric.

   Specify the following properties:

   - **Dimension value pattern –** `{Service:Instance}`
   - **Dimension name –** `service-instance`

7. Click **Save Metric**.

   The metric now appears in the **Settings > Server-side service monitoring > Calculated service metrics** view.

IBM WebSphere

The typical IBM® WebSphere deployment is a cell of WebSphere servers, controlled by a Deployment Manager. A cell makes up a managed domain that incorporates multiple VMS that host managed application servers. The Deployment Manager is a WebSphere instance that provides a single point of entry for the managed domain.

NOTE:
When adding a WebSphere Deployment Manager as a target, you must ensure that the name of each WebSphere node is resolvable to an IP address by the Workload Optimization Manager instance.

You may need to make changes to your DNS or the file `/etc/resolv.conf` on the Workload Optimization Manager instance to make it aware of the domain names in use in your environment.

To configure the WebSphere installation, you can use the WebSphere Integrated Solutions Console. This is a client that exposes configuration settings including the SOAP port and the PMI settings.

To manage the servers in an installation, WebSphere uses the Performance Monitoring Infrastructure (PMI). Each WebSphere server runs a PMI service that collects performance data from the various application server components. Workload Optimization Manager uses PMI for monitoring and control of the WebSphere installation.

Prerequisites

- IBM WebSphere Application Server 8.5+
- A service user account

  To execute actions the service account must have an Administrator role. For read-only monitoring and analysis, you can set the target with a more restricted role (Monitor), but then you will have to execute all recommended actions manually, through the WebSphere interface.

  NOTE:
  You cannot use Active Directory (AD) accounts to target WebSphere.

- The PMI service set to monitor at the Basic level or greater
- Discovered infrastructure

  Workload Optimization Manager discovers WebSphere servers that are running on VMs or containers. The hosting VM or container must already be in your Workload Optimization Manager inventory.

  To set the target for a server running on a VM, you must have first discovered the hosting VM through a Hypervisor target. To set the target for a server running in a container, you must have configured container discovery for WebSphere applications.

Finding the SOAP connector address

To configure a WebSphere target, you need to know the port that the server listens on for administrative communications. Launch the WebSphere Administration Console:

- Navigate to System **Administration > Deployment Manager**
- Under **Additional Properties**, click **Ports**

  The entry for SOAP_CONNECTOR_ADDRESS gives the currently set port number.

Adding a WebSphere target

You can add an individual WebSphere server as a target, or you can add all matching targets within a given scope.

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **WebSphere**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Display name

     Specify a name that uniquely identifies this connection.

     This name is for display purposes only and does not need to match any name in WebSphere.
   - Username

     Specify the username of an account with the Admin role.
   - Password

     Specify the password of an account with the Admin role.
   - Scope

The scope is a group of applications that are stitched to the underlying VMs when the VMs are discovered as part of a separate Workload Optimization Manager target.

If you set a scope, Workload Optimization Manager searches for virtual machines in the selected group.

Workload Optimization Manager can monitor up to 500 virtual machines in a group. If you have more than 500 virtual machines in your environment, split them across smaller groups and then add those groups as individual targets.

- Port number

  Specify the WebSphere remote port number. By default, the port is 8879.

- Authenticate all database servers

  If you select this option, Workload Optimization Manager attempts to authenticate all database servers in the selected scope. If Workload Optimization Manager is unable to authenticate a database server, the target is not added and no data is collected.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Service
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
- Application Component
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
  - Heap

    Heap is the portion of a VM or container's memory allocated to individual applications.
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
  - Threads

    Threads is the measurement of thread capacity utilized by applications.
  - Connection

    Connection is the measurement of database connections utilized by applications.
  - Remaining GC capacity

    Remaining GC capacity is the measurement of Application Component uptime that is *not* spent on garbage collection (GC).

Actions

Workload Optimization Manager supports the following actions:

- Service

  Workload Optimization Manager does not recommend actions to perform on the Service itself, but it does recommend actions to perform on the Application Components and hosting VMs. For example, assume a Service that manages three SQL databases. If a surge in requests degrades performance across all three databases, then Workload Optimization

Manager can start a new application component to run another instance of the database application, and bind it to the service. On the other hand, if SQL requests drop off so that the load balancer only forwards requests to two of the databases, Workload Optimization Manager can suspend the dormant database and unbind it.

- Application Component
  - Resize Heap

    This action can only be executed by Workload Optimization Manager if the entity is running in a domain controller. Actions for standalone entities can only be executed outside Workload Optimization Manager.
  - Resize Connection Capacity

    This action can only be executed outside Workload Optimization Manager.
- Virtual Machine
  - Provision additional resources (VMem, VCPU)
  - Move Virtual Machine
  - Reconfigure Storage
  - Suspend VM

Instana

Workload Optimization Manager supports discovery of applications that are managed by the Instana platform. Workload Optimization Manager includes the discovered information about these applications in its calculations for environment health.

NOTE:
For container platform environments, Workload Optimization Manager stitches New Relic, AppDynamics, Instana, and Dynatrace containerized application components into the supply chain to provide a unified view of your applications. For more information, see Container Platform Targets *(on page 234)*.

Prerequisites

- Instana release 242 or later
- An on-prem or SaaS instance of Instana

  This instance must be configured to monitor applications running in your environment.

  For an on-prem instance, you must also install BeeInstana. If BeeInstana is not installed, Workload Optimization Manager will return a `NullPointerException` and will not be able to collect metrics for Application Components and the underlying entities in the supply chain.
- Virtual machines and/or containers that are running applications managed by Instana

  For Workload Optimization Manager to discover applications through Instana, the applications must be running on virtual machines or containers in your environment. Also, these entities must be managed by Workload Optimization Manager hypervisor, public cloud, or container platform targets.
- A valid Instana API token

  This must be an API token created with the default permissions.

Adding an Instana target

NOTE:
It is possible to monitor certain applications or database servers with both Instana and another Workload Optimization Manager target. You should avoid such a configuration because it can cause Workload Optimization Manager to generate duplicate entities in the supply chain.

If you monitor an application via an Instana server, and you configure that Instana server as a Workload Optimization Manager target, then be sure you have not configured a separate application target in Workload Optimization Manager for that same application.

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **Instana**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Hostname or IP address

     Specify the hostname or IP address of the Instana server.
   - API token

     Specify the Instana API token.
   - Collect virtual machine metrics

     If you select this option, Workload Optimization Manager uses the virtual machine metrics that it collects from this target, instead of the metrics that can be collected from related infrastructure targets.
   - Collect tag information

     If you select this option, Workload Optimization Manager collects tags assigned to virtual machines that have been specified in the `configuration.yaml` file for Instana.

     The full directory is: `<instana-agent-dir>/etc/instana/configuration.yaml.`

     Instana tags do not use key:value pairs. Tags are defined in the Instana agent configuration files as:

     ```
     tags:
     - 'production'
     - 'app1'
     ```

     You can manually define a `key:value` pair by separating two strings with a colon. If you have not defined a `key:value` pair, the key is used as the value. For example:

     ```
     tags:
     - 'IamATag'
     - 'IamAKey:IamAValue'
     ```

     In Workload Optimization Manager, this example displays as:

     ```
     IamATag IamATag
     ```

     ```
     IamAKey IamAValue
     ```

     You can use these tags in Workload Optimization Manager when you search for, filter, or group entities discovered from the target. To view the collected tags, set the scope to a particular entity and then check the **Related Tag Information** chart.
   - Validate server certificates

     If you select this option, Workload Optimization Manager verifies the target certificate and proxy, if in use.

     This option is selected by default.
   - Proxy configuration

Specify proxy information only if you connect through a proxy.

- Proxy hostname or IP address

  Specify the address of the proxy used for this target.
- Proxy port

  Specify the port to use with the proxy used for this target. By default, the port is 8080.
- Proxy username

  Specify the username to use with the proxy.
- Proxy password

  Specify the password to use with the proxy.
- Connect using HTTPS

  Select this option if Workload Optimization Manager will connect to the proxy through HTTPS.

Entity mapping

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

Instana	Workload Optimization Manager
Application	Business Application
Database	Database Server
Endpoint	Business Transaction
Service	Service
Process	Application Component
Docker / Crio Container	Container
Host	Virtual Machine

Instana permissions

Workload Optimization Manager Functionality	Required permissions
Monitoring	▪ Default Workload Optimization Manager uses the default API token for monitoring access.

Monitored resources

Workload Optimization Manager monitors the following resources:

NOTE:
The exact resources that are monitored will differ based on application type. This list includes all of the resources that you may see.

- Business Application
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

  - Service level indicator (SLI)

    A service level indicator (SLI) is the defined quantitative measure of one characteristic of the level of service that is provided to a customer. Common examples of such indicators are error rate or response latency of a service.

    NOTE:
    Currently, SLI data is available in the Entity Information chart and Response Time chart.

- Business Transaction
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

  - Service level indicator (SLI)

    A service level indicator (SLI) is the defined quantitative measure of one characteristic of the level of service that is provided to a customer. Common examples of such indicators are error rate or response latency of a service.

    NOTE:
    Currently, SLI data is available in the Entity Information chart and Response Time chart.

- Service
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

  - Service level indicator (SLI)

    A service level indicator (SLI) is the defined quantitative measure of one characteristic of the level of service that is provided to a customer. Common examples of such indicators are error rate or response latency of a service.

    NOTE:
    Currently, SLI data is available in the Entity Information chart and Response Time chart.

- Application Component
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.

  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.

  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

  - Remaining GC capacity

Remaining GC capacity is the measurement of Application Component uptime that is *not* spent on garbage collection (GC).

This commodity applies only to Java applications.

  - Heap

    Heap is the portion of a VM or container's memory allocated to individual applications.

    This commodity applies only to Java applications.

- Database Server
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.

    This commodity applies only to Oracle, MySQL, and SQL databases.
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.

    This commodity applies only to Oracle, MySQL, and SQL databases.
  - Database memory (DBMem)

    Database memory (or DBMem) is the measurement of memory that is utilized by a Database Server.

    This commodity applies only to MySQL and SQL databases.
  - DB cache hit rate

    DB cache hit rate is the measurement of Database Server accesses that result in cache hits, measured as a percentage of hits versus total attempts. A high cache hit rate indicates efficiency.

    This commodity applies only to Oracle, MySQL, and SQL databases.
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

    This commodity applies only to Oracle databases.
  - Connection

    Connection is the measurement of database connections utilized by applications.

    This commodity applies only to Oracle databases.
- Container
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.

    This commodity applies only to Docker containers.
- Virtual Machine
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.

Actions

Workload Optimization Manager supports the following actions:

- Application Component
  - Resize Heap

    This action can only be executed outside Workload Optimization Manager.
- Database Server
  - Resize Database Memory (DBMem) - MySQL and SQL Server only

    This action can only be executed outside Workload Optimization Manager.

    Actions to resize database memory are driven by data on the Database Server, which is more accurate than data on the hosting VM.
  - Resize Connection Capacity - Oracle only

    This action can only be executed outside Workload Optimization Manager.

Workload Optimization Manager does not recommend actions for Instana Business Applications, Business Transactions, or Services. However, it considers resource utilization by these applications when recommending actions for the underlying virtual machines or containers. Workload Optimization Manager propagates this underlying information upward in the supply chain to show the impact of infrastructure risks on the Business Applications, Transactions, and Services.

Enabling the Instana-Workload Optimization Manager two-way integration

You can maximize the benefits of Workload Optimization Manager and Instana Observability (Instana) by setting up a two-way integration. To set up the two-way integration, you need to configure Workload Optimization Manager to fetch data from Instana, and then configure Instana by using the Workload Optimization Manager sensor to receive data (or actions) from Workload Optimization Manager. After you install the Instana agent, you must enable the Workload Optimization Manager sensor in the agent configuration file. Then, you can view performance *(on page 845)*-related actions *(on page 817)* from Workload Optimization Manager in the Instana UI.

The Workload Optimization Manager sensor polls Workload Optimization Manager APIs every 10 minutes to retrieve the executed and recommended actions. The sensor uses a unique key (SnapshotID) to link Instana entities with the corresponding Workload Optimization Manager actions.

> **NOTE:**
> The sensor needs the Instana automation action framework to automate the data retrieval and correlation process.

Supported operating systems

The supported operating systems of the Workload Optimization Manager sensor are consistent with host agent requirements. For more information, see the supported operating systems section of each host agent installation topic, such as Supported operating systems for Unix.

Prerequisites

Before you enable the Workload Optimization Manager sensor to set up a two-way integration between Workload Optimization Manager and Instana, complete the following steps:

- Instana must be added as a target in Workload Optimization Manager. For more information, see Instana *(on page 203)*.
- Enable the Instana automation actions framework for your Instana backend. For more information, see Managing actions in the Instana documentation.

Configuring Instana

By default, the Workload Optimization Manager sensor is disabled. To enable the sensor, update the agent configuration file as shown in the following example:

```
# Turbonomic Sensor
com.instana.plugin.turbonomic:
 enabled: true
 host: '<INSERT_TURBONOMIC_SERVER_HERE>'  # Turbonomic Server Host or IP address
 target: '<INSERT_TURBONOMIC_TARGET_NAME>' # Name of the target (Instana Server) as configured in
 Turbonomic
 auth:
   username: <INSERT_USERNAME_OF_TURBONOMIC_SERVER_HERE>  # User must have administrator, deployer, or
 automator authority.
   password: <INSERT_PASSWORD_OF_TURBONOMIC_SERVER_HERE>
```

NOTE:
You must have administrator, deployer, or automator authority to be able to use the Workload Optimization Manager API to collect actions.

Viewing Workload Optimization Manager actions in Instana

After you enable the Workload Optimization Manager sensor, you can view the executed and recommended actions from Workload Optimization Manager in the Instana UI. These actions are aimed at improving the performance of the entities that are monitored by Instana.

For more information, see Viewing Workload Optimization Manager actions in the Instana documentation.

Executed actions

You can view the executed actions on any of the following dashboards in the Instana UI.

Application perspective dashboard

1. In the sidebar of the Instana UI, select **Applications**.
2. Click an application perspective. On the **Summary** tab, you can see the actions that are executed by Turbonomic in the Actions swimlane with the application KPIs to indicate the impact of the action on the KPIs.
3. Click the action in the Actions swimlane. The Action History dialog is displayed with the details of the executed action.

Infrastructure dashboard

1. In the sidebar of the Instana UI, select **Infrastructure**.
2. On the **Map** tab, click a tower specific to the monitored host.
3. Click **Open Dashboard**. You can see the actions that are executed by or from Turbonomic in the Actions swimlane with the entity KPIs to indicate the impact of the action on the KPIs.
4. Click the action in the Actions swimlane. The Action History dialog is displayed with the details of the action.

Automation dashboard

1. In the sidebar of the Instana UI, select **Automation**.
2. Click the **Action History** tab. You can view a list of executed actions.
3. Click an executed action from the list to view its details.

Recommended actions

Recommended actions are proposed by Workload Optimization Manager for performance improvements on the entities that are monitored by Instana. These actions are retrieved by Instana and correlated against the ongoing incidents.

1. In the sidebar of the Instana UI, go to **Events** > **Issues**.
2. Click an issue.
3. Click the **Recommended Actions** tab. You can see all recommended actions with the Incident context.

> **NOTE:**
> When you click any recommended action from the list, you're directed to the Entity dashboard in the Workload Optimization Manager UI.

JBoss

Workload Optimization Manager supports connecting to JBoss targets running in the Managed Domain or Standalone operation modes.

- Managed Domain

  A collection of JBoss servers in a domain, with a single Domain Controller process acting as the central management control point. In this case, the VM that hosts the Domain Controller will be the target. You configure a Domain Controller via the `domain-controller` entry in the `host.xml` file on the JBoss machine.
- Standalone

  A single JBoss server. The VM that hosts the server is the target.

Prerequisites

- JBoss Application Server 6.3+
- The target VM is properly configured as Domain Controller or standalone, depending on the operation mode
- Hosting VM is discovered through a hypervisor target

  Workload Optimization Manager discovers JBoss servers that are running on VMs, or on containers that are deployed on VMs. To set the target for a JBoss server, you must have first discovered the hosting VM through a hypervisor target.

  For information about hypervisor targets, see Hypervisor Targets *(on page 333)*.
- JBoss server must define a thread pool in its configuration files

  Workload Optimization Manager monitors the Threads resource in application servers to track utilization of thread pool capacity. To monitor threads in JBoss, each JBoss server must define a thread pool in its configuration files. For instructions on specifying thread pools, see the JBoss documentation.
- JBoss supports Role-Based Access Control (RBAC) as a way to specify a set of permissions for user accounts

  JBoss disables RBAC by default. If you enable RBAC, Workload Optimization Manager requires the `Monitor` role for the target JBoss account in order to discover JBoss servers.

Adding a JBoss target

You can add an individual JBoss server as a target, or you can add all matching targets within a given scope.

1. Click **Settings > Target Configuration**.
2. Click **New Target > Applications and Databases**.
3. Select **JBoss**.
4. Configure the following settings:
   - Target Name

     Specify a name that uniquely identifies this connection.

This name is for display purposes only and does not need to match any name in JBoss.

- Username

  Specify the username of the account Workload Optimization Manager uses to connect to the target.

  Username should not include the Active Directory domain.

- Password

  Specify the password of the account Workload Optimization Manager uses to connect to the target.

- Scope

  Specify the scope Workload Optimization Manager uses for application discovery.

  The scope is a group of virtual machines that contain the databases that are discovered as part of a separate Workload Optimization Manager target.

  If you set a scope, Workload Optimization Manager searches for virtual machines in the selected group.

  Workload Optimization Manager can monitor up to 500 virtual machines in a group. If you have more than 500 virtual machines in your environment, split them across smaller groups and then add those groups as individual targets.

- JBoss Webservice port

  Specify the port for connections to the JBoss server. By default, the HTTP port is 9990.

- Authenticate all database servers

  If you select this option, Workload Optimization Manager attempts to authenticate all database servers in the selected scope. If Workload Optimization Manager is unable to authenticate a database server, the target is not added and no data is collected.

- Connect using HTTPS

  If you select this option, Workload Optimization Manager connects to the target servers using HTTPS. Make sure that the required certificate is configured for use on the host.

- Proxy configuration

  Specify proxy information only if you connect through a proxy.

  - Proxy hostname or IP address

    Specify the address of the proxy used for this target.

  - Proxy port

    Specify the port to use with the proxy used for this target. By default, the port is 8080.

  - Proxy username

    Specify the username to use with the proxy.

  - Proxy password

    Specify the password to use with the proxy.

  - Connect using HTTPS

    Select this option if Workload Optimization Manager will connect to the proxy through HTTPS.

Application component names

Workload Optimization Manager displays discovered JBoss servers in the user interface. These names indicate whether they're standalone or in a managed domain.

The name is divided into three sections:

- Domain Controller name or “STANDALONE”
- The server name
- The name or IP address of the VM hosting the JBoss server

For example:

```
STANDALONE:acm-jboss73 [ACM-JBoss7.3-171.47]
```

Monitored resources

Workload Optimization Manager monitors the following resources:

- Application Component
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
  - Heap

    Heap is the portion of a VM or container's memory allocated to individual applications.
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
  - Threads

    Threads is the measurement of thread capacity utilized by applications.
  - Connection

    Connection is the measurement of database connections utilized by applications.
  - Remaining GC capacity

    Remaining GC capacity is the measurement of Application Component uptime that is *not* spent on garbage collection (GC).

Actions

Workload Optimization Manager supports the following actions:

- Application Component
  - Resize Heap

    This action can only be executed outside Workload Optimization Manager.
  - Resize Thread Pool

    This action can only be executed outside Workload Optimization Manager.
  - Resize Connection Capacity

    This action can only be executed outside Workload Optimization Manager.

JVM application

Workload Optimization Manager supports connecting to individual JVM Applications as targets. Workload Optimization Manager connects to the JVM process as a remote client via remote JMX access. Target configuration includes the port used by the JMX/RMI registry.

Prerequisites

- JVM 6.0+
- A valid JMX user account for the JVM application

  If JMX security is enabled this must be a JMX user with a `readonly` role
- The application should run on JVM version 6.0 or higher
- For VMware environments, VMware Tools must be installed on the VM that hosts the application

  This ensures that the VM hosting the application can get the application's IP address
- Remote JMX access is enabled through a port that is opened to the firewall

- Discovered infrastructure

  Workload Optimization Manager discovers JVM applications that are running on VMs or containers. The hosting VM or container must already be in your Workload Optimization Manager inventory.

  To set the target for a server running on a VM, you must have first discovered the hosting VM through a hypervisor target. To set the target for a server running in a container, you must have configured container discovery for JVM applications.

Adding JVM application targets

When you configure JVM targets, you declare a given scope and add all matching applications within that given scope. To do this, specify:

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **JVM Application**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Display name

     Specify a name that uniquely identifies this connection.
   - Username

     Specify the username of an account with the Admin role.

     Credentials must match the credentials that you specify for the JMX login configuration when you start up the application. If you disable authentication on the application, you must still provide arbitrary values for Username and Password.
   - Password

     Specify the password of an account with the Admin role.

     Credentials must match the credentials that you specify for the JMX login configuration when you start up the application. If you disable authentication on the application, you must still provide arbitrary values for Username and Password.
   - Scope

     Specify the scope Workload Optimization Manager uses for application discovery.

     The scope is a group of applications that are stitched to the underlying VMs when the VMs are discovered as part of a separate Workload Optimization Manager target.

     If you set a scope, Workload Optimization Manager searches for virtual machines in the selected group.

     Workload Optimization Manager can monitor up to 500 virtual machines in a group. If you have more than 500 virtual machines in your environment, split them across smaller groups and then add those groups as individual targets.
   - JMX remote port

     Specify the JMX remote port number.

     To disable JMX authentication, use the following flags in the command line as you start the application:

     ```
     -Dcom.sun.management.jmxremote.authenticate=false
     ```

     ```
     -Dcom.sun.management.jmxremote.ssl=false
     ```

   - Authenticate all database servers

     If you select this option, Workload Optimization Manager attempts to authenticate all database servers in the selected scope. If Workload Optimization Manager is unable to authenticate a database server, the target is not added and no data is collected.

Configuring JMX remote access

Workload Optimization Manager monitors and controls JVM applications via JMX Remote access. You must configure a JMX Remote port.

Note that to work with a firewall you should also set the RMI Server port — If you don't set an RMI port, then JMX sets an arbitrary *ephemeral port*, and you can't guarantee that the port will be open to your firewall.

To set the JMX Remote port, pass in the port at the command line when you start your application. For example, to set the port to `8090`, start your application with the following options:

```
-Dcom.sun.management.jmxremote -Dcom.sun.management.jmxremote.port=8090
```

Multiple JVM targets on single VM

Note that you can specify targets with different ports, but that run on the same VM (use the same IP address). You can also specify targets via the same scope, but with different ports — This is another way to assign applications running on the same VM to different ports. To do this:

To do this, add the targets in two separate steps. For example, assume you want to add two JVM application targets, and they both run on the VM at `10.10.123.45`. One application is on port `123`, and the other application is on port `456`. To specify these two targets:

- Specify the first target with the following parameters:
  - Scope: `VMs_myCluster.mycorp.com`
  - Port number: `123`
  - Username: `AppUser`
  - Password: `**********`

  Click **ADD**.
- Specify the second target with the following parameters:
  - Scope: `VMs_myCluster.mycorp.com`
  - Port number: `456`
  - Username: `OtherAppUser`
  - Password: `**********`

  Click **ADD**.

Monitored resources

Workload Optimization Manager monitors the following resources:

- **Application Component** (JVM Application)
  - Heap

    Heap is the portion of a VM or container's memory allocated to individual applications.
  - Remaining GC capacity

    Remaining GC capacity is the measurement of Application Component uptime that is *not* spent on garbage collection (GC).

    Data is collected if JVM profiler is enabled.
- Virtual Machine
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.

Actions

Workload Optimization Manager supports the following actions:

- **Application Component** (JVM Application)
  - Resize Heap

    This action can only be executed outside Workload Optimization Manager.
  - Resize Thread Pool

    This action can only be executed outside Workload Optimization Manager.
  - Resize Connection Capacity

    This action can only be executed outside Workload Optimization Manager.
  - Suspend VM

    This action can only be executed by Workload Optimization Manager if a VM is hosted in a vCenter environment. Actions for applications running on other hypervisors can only be executed outside Workload Optimization Manager.
  - Provision VM

    This action can only be executed outside Workload Optimization Manager.
- Virtual Machine
  - Provision additional resources (VMem, VCPU)
  - Move Virtual Machine
  - Move Virtual Machine Storage
  - Reconfigure Storage
  - Reconfigure Virtual Machine
  - Suspend VM
  - Provision VM

MySQL

To manage MySQL databases, Workload Optimization Manager can connect to one or more database servers within a defined scope.

Prerequisites

- MySQL 8.0
- User permissions are enabled on the MySQL Server

  For more information, see Enabling user permissions on MySQL *(on page 217)*.

Adding a MySQL database target

You can add all matching targets within a given scope.

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **MySQL**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Display name

     Specify a name that uniquely identifies this connection.

     This name is for display purposes only and does not need to match any name in MySQL.

- Username

  Specify the username of the account Workload Optimization Manager uses to connect to the target.

- Password

  Specify the password of the account Workload Optimization Manager uses to connect to the target.

- Scope

  Specify the scope Workload Optimization Manager uses for application discovery.

  The scope is a group of applications that are stitched to the underlying VMs when the VMs are discovered as part of a separate Workload Optimization Manager target.

  If you set a scope, Workload Optimization Manager searches for virtual machines in the selected group.

  Workload Optimization Manager can monitor up to 500 virtual machines in a group. If you have more than 500 virtual machines in your environment, split them across smaller groups and then add those groups as individual targets.

- Port number

  Specify the MySQL remote port number. By default, the port is 3306.

- Authenticate all database servers

  If you select this option, Workload Optimization Manager attempts to authenticate all database servers in the selected scope. If Workload Optimization Manager is unable to authenticate a database server, the target is not added and no data is collected.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Database server
  - Database memory (DBMem)

    Database memory (or DBMem) is the measurement of memory that is utilized by a Database Server.
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
  - Connection

    Connection is the measurement of database connections utilized by applications.

    A database connection is a physical communication pathway that holds database sessions, which are logical entities in the database instance memory that represent the state of a current user login to a database. Connections should be managed properly.
  - DB cache hit rate

    DB cache hit rate is the measurement of Database Server accesses that result in cache hits, measured as a percentage of hits versus total attempts. A high cache hit rate indicates efficiency.
- Virtual machine
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.
  - Virtual storage

    Virtual ssorage is the measurement of virtual storage capacity that is in use.

- Storage access (IOPS)

  Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

- Latency

  Latency is the measurement of storage latency.

Actions

Workload Optimization Manager supports the following actions:

- Database server

  Resize

  - Database memory (DBMem)

    Actions to resize database memory are driven by data on the Database Server, which is more accurate than data on the hosting VM. Workload Optimization Manager uses database memory and cache hit rate data to decide whether resize actions are necessary.

    A high cache hit rate value indicates efficiency. The optimal value is 100% for on-prem (self-hosted) Database Servers, and 90% for cloud Database Servers. When the cache hit rate reaches the optimal value, no action generates even if database memory utilization is high. If utilization is low, a resize down action generates.

    When the cache hit rate is less than the optimal value but database memory utilization remains low, no action generates. If utilization is high, a resize up action generates.

- Virtual machine

  Resize

  Resize resource capacity, reservation, or limit to improve performance.

Enabling user permissions on MySQL server

Follow the following steps to enable appropriate user permissions on a MySQL Server.

1. Edit the MySQL server's configuration file to grant user permissions.
   a. Open a Secure Shell session on the server and open the `.conf` file on the MySQL server in an editor.

      Depending on the platform your MySQL is running on, you find the file at different locations.

      - Debian Linux

        `/etc/mysql/my.cnf`
      - Red Hat Linux (Fedora or Rocky)

        `/etc/my.cnf`
      - FreeBSD Linux

        Create the file at `/var/db/mysql/my.cnf`

   b. Make the following changes in the `[mysqld]` section:

      - Comment out the following line to enable remote connections over TCP/Is:

        `skip-networking`
      - Add the following line to bind your MySql server address:

        `bind-address=<MySQL_IP_Address>`
      - Add the following line to enable the collection of Transaction metrics:

        ```
        innodb_monitor_enable = trx_rw_commits, trx_nl_ro_commits,
        trx_ro_commits, trx_rollbacks
        ```

      For example, if your MySQL server has the address, `123.45.66.77`, after you bound the IP address and enabled Transaction metrics, the section of the `.conf` file should look like the following example:

```
[mysqld]
user                = mysql
pid-file            = /var/run/mysqld/mysqld.pid
socket              = /var/run/mysqld/mysqld.sock
port                = 3306
basedir             = /usr
datadir             = /var/lib/mysql
tmpdir              = /tmp
language            = /usr/share/mysql/English
bind-address        = 123.45.66.77
# skip-networking
# Uncomment the following line for MySQL versions 5.6+
innodb_monitor_enable = trx_rw_commits, trx_nl_ro_commits, trx_ro_commits, trx_rollbacks
....
```

c. Save the `.conf` file.

NOTE:
Some MySQL installations use multiple configuration files. If a setting you made does not have the wanted effect, make sure that a different configuration file is not overwriting the value.

2. Enable collection of Response Time metrics.

   Run the following command to log in to to the MySQL server:

```
$mysql -u root -p mysql
```

   Then run the following SQL commands:

```
UPDATE performance_schema.setup_instruments SET ENABLED = 'YES' WHERE NAME LIKE 'statement/sql%';
```

```
UPDATE performance_schema.setup_instruments SET TIMED = 'YES' WHERE NAME LIKE 'statement/sql%';
```

   NOTE:
   If you want these changes to take effect each time you restart the MySQL server, add these statements to a file, and start the server with the `--init-file` option. For example, if you name the file `MyInit.txt`, then start the MySQL server with the following option:

```
--init-file=MyInit.txt
```

3. Give your Workload Optimization Manager server remote access to the database.

   If you are not already logged in to the MySQL server, run the following command:

```
$mysql -u root -p mysql
```

   Then run the following commands.

   Assume a user named USER_NAME with a password PWD_STRING. Then assume that your Workload Optimization Manager has an IP address of `10.10.123.45`. The following command grants privileges to that Workload Optimization Manager, if it connects with the specified user account:

```
GRANT SELECT ON performance_schema.* TO 'USER_NAME'@'10.10.123.45' IDENTIFIED BY 'PWD_STRING';
            GRANT PROCESS ON *.* TO 'USER_NAME'@'10.10.123.45' IDENTIFIED BY 'PWD_STRING';
            FLUSH PRIVILEGES;
```

   The `FLUSH PRIVILEGES` command causes MySQL to retain these settings upon restart.

   When you are finished running these SQL commands, log out of MySQL.

New Relic

Workload Optimization Manager supports workload management of the application infrastructure monitored by New Relic, from application instance to host. With information obtained from New Relic, Workload Optimization Manager can make recommendations and take actions to both assure performance and drive efficiency to address the demands of each individual application.

> **NOTE:**
> For container platform environments, Workload Optimization Manager stitches New Relic, AppDynamics, Instana, and Dynatrace containerized application components into the supply chain to provide a unified view of your applications. For more information, see Container Platform Targets *(on page 234)*.

Prerequisites

- A valid New Relic user account that includes both APM and infrastructure monitoring

Entity mapping

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

New Relic	Workload Optimization Manager
APM: Key Transactions	Business Transaction
APM: Application / Service (New Relic One)	Service
APM: Application Instance	Application Component
Infra: Database	Database Server
Infra: Host	Virtual Machine

Supported applications

Workload Optimization Manager discovers the following application types (and associated commodities) via the New Relic target:

Application Type	Commodities
.NET	Virtual CPU, Virtual Memory, Response Time, Transactions
GO	Virtual CPU, Virtual Memory, Response Time, Transactions
Java	Virtual CPU, Virtual Memory, Response Time, Transactions, Heap, Collection Time, Threads
Node.js	Virtual CPU, Virtual Memory, Response Time, Transactions, Heap, Collection Time
PHP	Virtual CPU, Virtual Memory, Response Time, Transactions
Python	Virtual CPU, Virtual Memory, Response Time, Transactions

Supported databases

Workload Optimization Manager supports the following Database types and commodities:

NOTE: Database commodities are exposed only if the New Relic account used to connect to Workload Optimization Manager has a `New Relic Infrastructure Pro` subscription.

Database	Commodities
SQL	Cache Hit Rate, Virtual Memory, Transactions
MySQL	Cache Hit Rate, Transactions **NOTE:** Workload Optimization Manager does not support MySQL version 8.0 or higher. For more information, see the New Relic documentation.
OracleDB	Cache Hit Rate, Transactions, Response Time
MongoDB	Virtual Memory, Connections

Adding a New Relic target

NOTE:
If an application is monitored by New Relic, do not add it as a separate Workload Optimization Manager application target.

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **New Relic**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Display name

     Specify a name that uniquely identifies this connection.

     This name is for display purposes only and does not need to match any name in New Relic.

     The name field can include only alphanumeric, space, or hyphen characters and must have a unique name when multiple New Relic targets are added.
   - User key

     Specify the user key that is provided by the New Relic platform.
     - The user key is used for the New Relic GraphQL API.
     - The user key is assigned to a New Relic user and can access all New Relic accounts that are available to that user.
     - The availability of data that the user key can collect is defined by the user's permissions in New Relic.
     - It is recommended that the user has access to both New Relic domains: APM and Infrastructure.
     - For more information about the New Relic user access model, see Important user management concepts in the New Relic documentation.

     For more information, see the New Relic documentation.
   - Connect to EU region data centers

     If you select this option, Workload Optimization Manager uses the EU API endpoints.
   - Collect virtual machine metrics

If you select this option, Workload Optimization Manager uses the virtual machine metrics that it collects from this target, instead of the metrics that can be collected from related infrastructure targets.

- Collect tag information

  If you select this option, Workload Optimization Manager collects tags that are assigned to containers, database servers, and virtual machines discovered from this target.

  Tags are key-value pairs that you define in the target and then assign to different entities.

  You can use these tags in Workload Optimization Manager when you search for entities discovered from the target. You can also use tags to filter or group Containers and Virtual Machines. To view the collected tags, set the scope to a particular entity and then check the **Related Tag Information** chart.

- Proxy configuration

  Specify proxy information only if you connect through a proxy.

  - Proxy hostname or IP address

    Specify the address of the proxy used for this target.

  - Proxy port

    Specify the port to use with the proxy used for this target. By default, the port is 8080.

  - Proxy username

    Specify the username to use with the proxy.

  - Proxy password

    Specify the password to use with the proxy.

  - Connect using HTTPS

    Select this option if Workload Optimization Manager will connect to the proxy through HTTPS.

Monitored resources

Workload Optimization Manager monitors the following resources:

NOTE:
The exact resources that are monitored will differ based on application type. This list includes all of the resources that you may see.

- Application Component
  - Heap

    Heap is the portion of a VM or container's memory allocated to individual applications.

  - Remaining GC capacity

    Remaining GC capacity is the measurement of Application Component uptime that is *not* spent on garbage collection (GC).

  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

  - Threads

    Threads is the measurement of thread capacity utilized by applications.

  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.

  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.

- Database

  NOTE:
  Refer to the supported Database types and commodities *(on page 219)*.

  - Connection

    Connection is the measurement of database connections utilized by applications.
  - Database memory (DBMem)

    Database memory (or DBMem) is the measurement of memory that is utilized by a Database Server.
  - DB cache hit rate

    DB cache hit rate is the measurement of Database Server accesses that result in cache hits, measured as a percentage of hits versus total attempts. A high cache hit rate indicates efficiency.
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
- Business Transaction
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
- Service
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
- Container
  - vCPU

    vCPU is the virtual CPU (in mCores) utilized by a container against the CPU limit. If no limit is set, node capacity is used).
  - Virtual memory (vMem)

    Virtual memory (vMem) is the virtual memory utilized by a container against the memory limit. If no limit is set, node capacity is used.
- Virtual Machine
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.

- Virtual memory (vMem)

  Virtual memory (vMem) is the measurement of memory that is in use.

For a VM, the resources you see depend on how the VM is discovered, and whether the VM provides resources for an application that is discovered by this target:

- If the VM hosts an application that is discovered through this target, then you see VM metrics that are discovered through this target.
- If the VM is discovered through a different target, and it does not host any application discovered through this target, you see VM metrics that are discovered through that different target.
- If the VM is discovered through this target, but it does not host any application that is discovered through this target, then Workload Optimization Manager does not display metrics for the VM.

Actions

NOTE:
The specific actions that Workload Optimization Manager recommends can differ, depending on the processes that Workload Optimization Manager discovers.

For other application components, Workload Optimization Manager can recommend actions based on the resources it can discover for the application. For example, Node.js® applications report CPU usage, so Workload Optimization Manager can generate vCPU resize actions and display them in the user interface.

Workload Optimization Manager supports the following actions:

- Application Component
  - Resize Heap

    This action can only be executed outside Workload Optimization Manager.
  - Resize Threads

    This action can only be executed outside Workload Optimization Manager.
- Virtual Machine
  - Provision VM

    This action can only be executed outside Workload Optimization Manager.
  - Suspend VM

    This action can only be executed outside Workload Optimization Manager.

New Relic accounts

When you select a New Relic target and click **View** in the **Details** menu, Workload Optimization Manager lists all New Relic accounts that are associated with the target user key on the **New Relic accounts** tab. You can also filter entities by New Relic account on the Search page. When you select **New Relic accounts** from the filter menu on the Search page, you can select individual scoped New Relic accounts to observe their associated supply chains.

Oracle

To manage Oracle databases, Workload Optimization Manager can connect to one or more database servers within a defined scope.

To connect to an Oracle database, you will:

- Add a Dynamic Performance view to the Oracle database
- Configure a service account on the database that Workload Optimization Manager can use to log on
- Find the Service Name and port for the database

Prerequisites

- Oracle 19c or 21c
- User permissions that grant access to Workload Optimization Manager through a specific user account

  See Creating a Service User Account in Oracle *(on page 226)*.
- Dynamic Performance View (V$) must be enabled

  For more information, see Adding a Dynamic Performance View *(on page 224)*.
- Access through the firewall to the Oracle database port that you specify for the Workload Optimization Manager target connection

Adding a dynamic performance view

In order to collect data from the Oracle database, Workload Optimization Manager uses the Dynamic Performance View (referred to as V$). V$ is not enabled by default. You must run a script to build the tables and views that are necessary to enable V$. In some environments only the DBA has privileges to run this script.

To enable V$:

- Open a secure shell session (ssh) on the database host as a system user or a user with the `sysdba` role
- In the shell session enter the following commands:

```
sqlplus /nolog
connect /as sysdba
CREATE USER My_Username IDENTIFIED BY My_Password container=all;
GRANT CONNECT TO My_Username container=all;
GRANT sysdba TO My_Username container=all;
```

NOTE:
If security or other practices prohibit assigning SYSDBA to this user, you can use the following command to provide access to all V$ views:

```
GRANT select any dictionary TO My_Username;
```

This creates a user account named `My_Username` with full privileges to access the V$ Dynamic Performance view.

Adding an Oracle database to Workload Optimization Manager

You can add an individual database server as a target, or you can add all matching targets within a given scope.

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **Oracle**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Display name

     Specify a name that uniquely identifies this connection.

     This name is for display purposes only and does not need to match any name in Oracle.
   - Username

     Specify the username of the account Workload Optimization Manager uses to connect to the target.

     For Workload Optimization Manager to execute actions, the account must have administrator privileges. You must have enabled user permissions to this user account, including remote access from the Workload Optimization Manager server.
   - Password

Specify the password of the account Workload Optimization Manager uses to connect to the target.

- Scope

  Specify the scope Workload Optimization Manager uses for application discovery.

  The scope is a group of virtual machines that contain the databases that are discovered as part of a separate Workload Optimization Manager target.

  If you set a scope, Workload Optimization Manager searches for virtual machines in the selected group.

  Workload Optimization Manager can monitor up to 500 virtual machines in a group. If you have more than 500 virtual machines in your environment, split them across smaller groups and then add those groups as individual targets.

  NOTE:
  All database servers in the scope must share the same service name, credentials, and port. For databases that have a different value for any of these, you must create a separate target using those values.

- Oracle TCP Port

  Specify the port that connects to the database. You must open the firewall on the database server to allow access through this port. By default, the port is 1521.

  Any firewall on the database must allow access through this port. To find the port, open an SSH session (as a system or sysdba user) on the database's host, run `lsnrctl status`, and then `check PROTOCOL=tcp`.

- Oracle service name

  Specify the service name for the database that you are connecting to Workload Optimization Manager.

  To find the service name, open an SSH session on the database's host and run the following commands:

  ```
  sqlplus /no log
  connect /as sysdba
  SELECT SYS_CONTEXT ('userenv', 'db_name') FROM dual;
  ```

- Authenticate all database servers

  If you select this option, Workload Optimization Manager attempts to authenticate all database servers in the selected scope. If Workload Optimization Manager is unable to authenticate a database server, the target is not added and no data is collected.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Database server
  - Database memory (DBMem)

    Database memory (or DBMem) is the measurement of memory that is utilized by a Database Server.

    Actions to resize database memory are driven by data on the Database Server, which is more accurate than data on the hosting VM.
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
  - Connection

    Connection is the measurement of database connections utilized by applications.

A database connection is a physical communication pathway that holds database sessions, which are logical entities in the database instance memory that represent the state of a current user login to a database. Connections should be managed properly.

  - Transaction log

    Transaction log is the measurement of storage capacity utilized by a Database Server for transaction logging.

  - DB cache hit rate

    DB cache hit rate is the measurement of Database Server accesses that result in cache hits, measured as a percentage of hits versus total attempts. A high cache hit rate indicates efficiency.

- Virtual machine
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.

  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.

  - Virtual storage

    Virtual ssorage is the measurement of virtual storage capacity that is in use.

  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

  - Latency

    Latency is the measurement of storage latency.

Actions

Workload Optimization Manager supports the following actions:

- Database server

  Resize

  Actions to resize database memory are driven by data on the Database Server, which is more accurate than data on the hosting VM. Workload Optimization Manager uses database memory and cache hit rate data to decide whether resize actions are necessary.

  A high cache hit rate value indicates efficiency. The optimal value is 100% for on-prem (self-hosted) Database Servers, and 90% for cloud Database Servers. When the cache hit rate reaches the optimal value, no action generates even if database memory utilization is high. If utilization is low, a resize down action generates.

  When the cache hit rate is less than the optimal value but database memory utilization remains low, no action generates. If utilization is high, a resize up action generates.

- Virtual machine

  Resize

  Resize resource capacity, reservation, or limit to improve performance.

Creating a service user account in Oracle

To collect data from the Oracle database, Workload Optimization Manager requires a service account that has privileges to access the V$ Dynamic Performance view. To create this account:

- Open a secure shell session (ssh) on the database host as a system user or a user with the `sysdba` role
- In the shell session enter the following commands:

```
sqlplus /nolog
connect /as sysdba
```

```
CREATE USER My_Username IDENTIFIED BY My_Password container=all;
GRANT CONNECT TO My_Username container=all;
GRANT sysdba TO My_Username container=all;
```

This creates a user account named My_Username with full privileges to access the V$ Dynamic Performance view.

NOTE:
The preceding example uses a fictitious username. To comply with Oracle 12C norms, the username should include a prefix of c##.

Some enterprises don't allow accounts with sysdba access. Cisco recommends using sysdba, according to the Oracle documentation. However, you can work with your Oracle DBA staff to provide read access to the following views, which are the ones that Workload Optimization Manager needs:

- V$INSTANCE
- V$LOG
- V$LOGFILE
- V$PARAMETER
- V$PGASTAT
- V$RESOURCE_LIMIT
- V$SGASTAT
- V$SYS_TIME_MODEL
- V$SYSMETRIC
- V$SYSSTAT

Oracle WebLogic

Oracle WebLogic deployments can include clusters to distribute workload across multiple WebLogic servers. Workload Optimization Manager recommended actions respect the cluster architecture. For example, if you have enabled horizontal scaling for your WebLogic servers, then Workload Optimization Manager can recommend provisioning new servers for a given cluster.

The typical WebLogic deployment is a managed domain with one Administration Server that provides a single point of entry for administration and management of the domain. The domain can include other WebLogic Servers which are the Managed Servers. You set the WebLogic Administration Server as a Workload Optimization Manager target.

For a standalone WebLogic deployment, the single server acts as its own Administration Server — You can set the standalone server as a Workload Optimization Manager target.

Prerequisites

- Oracle WebLogic 12c
- A service user account

  To execute actions the service account must have an Admin role. For read-only monitoring and analysis, you can set the target with a more restricted role, but then you will have to execute all recommended actions manually, through the WebLogic interface.
- The Workload Optimization Manager IP address and all WebLogic server IP addresses must be resolvable by your local DNS server
- Discovered infrastructure

  Workload Optimization Manager discovers WebLogic servers that are running on VMs or containers. The hosting VM or container must already be in your Workload Optimization Manager inventory.

To set the target for a server running on a VM, you must have first discovered the hosting VM through a hypervisor target. To set the target for a server running in a container, you must have configured container discovery for WebLogic applications.

- For information about container platform targets, see this topic *(on page 234)*.
- For information about hypervisor targets, see this topic *(on page 333)*.

Finding the T3 listen port

To configure a WebLogic target, you need to know the port that the server listens on for administrative communications. Launch the WebLogic Administration Console:

- Navigate to Domain Structure and display the domain you're interested in
- Navigate to **Environment > Servers** and select the Domain Administration Server you're setting as a target

  The console displays configuration information for the server, including the T3 listen port.

Adding a WebLogic target

You can add one or more WebLogic targets in a given scope, including both standalone servers and domain managers.

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **WebLogic**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Display name

     Specify a name that uniquely identifies this connection.

     This name is for display purposes only and does not need to match any name in WebLogic.
   - Username

     Specify the username of an account with the Admin role.
   - Password

     Specify the password of an account with the Admin role.
   - Scope

     Specify the scope Workload Optimization Manager uses for application discovery.

     The scope is a group of applications that are stitched to the underlying VMs when the VMs are discovered as part of a separate Workload Optimization Manager target.

     If you set a scope, Workload Optimization Manager searches for virtual machines in the selected group.

     Workload Optimization Manager can monitor up to 500 virtual machines in a group. If you have more than 500 virtual machines in your environment, split them across smaller groups and then add those groups as individual targets.
   - Port number

     Specify the WebLogic remote port number. By default, the port is 7001.
   - Authenticate all database servers

     If you select this option, Workload Optimization Manager attempts to authenticate all database servers in the selected scope. If Workload Optimization Manager is unable to authenticate a database server, the target is not added and no data is collected.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Service
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
- Application Component
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
  - Heap

    Heap is the portion of a VM or container's memory allocated to individual applications.
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
  - Threads

    Threads is the measurement of thread capacity utilized by applications.
  - Connection

    Connection is the measurement of database connections utilized by applications.
  - Remaining GC capacity

    Remaining GC capacity is the measurement of Application Component uptime that is *not* spent on garbage collection (GC).

Actions

Workload Optimization Manager supports the following actions:

- Service

  Workload Optimization Manager does not recommend actions to perform on the Service itself, but it does recommend actions to perform on the Application Components and hosting VMs. For example, assume a Service that manages three SQL databases. If a surge in requests degrades performance across all three databases, then Workload Optimization Manager can start a new application component to run another instance of the database application, and bind it to the service. On the other hand, if SQL requests drop off so that the load balancer only forwards requests to two of the databases, Workload Optimization Manager can suspend the dormant database and unbind it.
- Application Component
  - Resize Heap

    This action can only be executed by Workload Optimization Manager if the entity is running in a domain controller. Actions for standalone entities can only be executed outside Workload Optimization Manager.
  - Resize Connection Capacity

    This action can only be executed outside Workload Optimization Manager.
- Virtual Machine
  - Provision additional resources (VMem, VCPU)
  - Move Virtual Machine

  - Reconfigure Storage
  - Suspend VM

SQL server

Workload Optimization Manager discovers both standalone and clustered SQL Servers, and represents them as Database Server entities in the supply chain.

Prerequisites

- SQL Server 2012, 2014, 2016, 2017, 2019, or 2022
- A user account with SQL permissions including `Connect SQL` and `View Server State` on the database
- The Net.Tcp Port Sharing Service and Net.Tcp Listener Adapter services must be running, and set to enabled
- TCP/IP is enabled on the port used for Workload Optimization Manager discovery

Creating a service user account

The user account that Workload Optimization Manager uses for its service login must include the following:

- The account must exist in the Security folder within the SQL Server Object Explorer, with the following properties:
  - Enable **SQL Server Authentication**
  - Disable **Enforce password policy**
- The account's security properties must include:
  - Permission to connect to the database through SQL
  - Permission to view the server state

Adding a SQL server target

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **SQLServer**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Display name

     Specify a name that uniquely identifies this connection.

     This name is for display purposes only and does not need to match any name in SQLServer.
   - Username

     Specify the username of the account Workload Optimization Manager uses to connect to the target.

     Username must not include the Active Directory domain.
   - Password

     Specify the password of the account Workload Optimization Manager uses to connect to the target.

     Password must not include the Active Directory domain.
   - Active Directory Domain

     Specify the Active Directory domain used by Workload Optimization Manager in conjunction with the username for authentication. Leave blank for local accounts.
   - Discovery path

     Specify the hostname or IP address or scope when performing the discovery process.

Workload Optimization Manager discovers SQL instances through the SQL Server's hostname/IP address or your selected scope.

If you change your selection, the value associated with the deselected option is automatically removed.

  - Hostname or IP address

    Specify the hostname or IP address of your MSSQL environment configuration. Workload Optimization Manager scans the targeted hostname or IP address and tries to connect to the target using the specified port. Workload Optimization Manager adds any instances of the target it finds as entities from which metrics are retrieved.

  - Scope

    Specify the scope Workload Optimization Manager uses for application discovery.

    The scope is a group of applications that are stitched to the underlying VMs when the VMs are discovered as part of a separate Workload Optimization Manager target.

    If you set a scope, Workload Optimization Manager searches for virtual machines in the selected group.

    Workload Optimization Manager can monitor up to 500 virtual machines in a group. If you have more than 500 virtual machines in your environment, split them across smaller groups and then add those groups as individual targets.

- Browsing service port

  Specify the MS SQL Browsing Service port for the browsing service that listens for incoming connections to the SQL instances running on the SQL Server. This port overrides the SQLServer port. By default, the port is 1434.

  If the browsing service is reachable via the specified port, Workload Optimization Manager discovers the SQL instances used by the VM group that you defined as your scope, as well as the listening ports on those SQL instances.

  If the service is unreachable, or if you did not specify a MS SQL Browsing Service port, Workload Optimization Manager uses the TCP port that you specified in the SQLServer Port field to discover SQL instances.

- SQL Server port

  Specify the TCP port of the MS SQL Server. By default, the port is 1433.

  Workload Optimization Manager uses this port if the browsing service is unreachable, or if you did not specify a browsing service port.

- Authenticate all database servers

  If you select this option, Workload Optimization Manager attempts to authenticate all database servers in the selected scope. If Workload Optimization Manager is unable to authenticate a database server, the target is not added and no data is collected.

Standalone and clustered SQL servers

Workload Optimization Manager discovers both standalone and clustered SQL Servers, and represents them as Database Server entities in the supply chain.

- When you set the scope to a SQL Server entity and view the Entity Information chart, the **Server Configuration** field indicates whether that entity is a standalone server or is part of a cluster.
- When you search for Database Servers or create groups of Database Servers, use the `Server Configuration` filter to get a list of standalone or clustered servers.

For clustered environments:

- Workload Optimization Manager represents each SQL Server instance in the cluster as a Database Server entity, and automatically creates a group for these instances.
- To see all the auto-created groups, go to **Settings > Groups** and then search for group names starting with `MSSQL:Cluster:`. Click a group name to set the scope to that group. In the resulting supply chain, the Database Server entity shows the number of SQL Server instances in the cluster. This entity is stitched to the Virtual Machine entity, which represents the corresponding SQL Server nodes. Click **Database Server** to see a list of instances, and identify which instance is currently active and which ones are idle (redundant). Workload Optimization Manager only monitors resources for the active instance, and shows resource metrics when you set the scope to that instance.

- If you have several clusters, you can use filters to identify all the active/idle instances in those clusters. In Search, select **Database Servers**, and then set the filter as follows:
  - Active instances

    `Server Configuration=Clustered` and `State=ACTIVE`
  - Idle (redundant) instances

    `Server Configuration=Clustered` and `State=IDLE`

Monitored resources

Workload Optimization Manager monitors the following resources:

- Database server

  NOTE:
  For clustered environments, Workload Optimization Manager only monitors resources for the currently active SQL Server instance, and shows resource metrics when you set the scope to that instance.

  - Connection

    Connection is the measurement of database connections utilized by applications.

    A database connection is a physical communication pathway that holds database sessions, which are logical entities in the database instance memory that represent the state of a current user login to a database. Connections should be managed properly.
  - Database memory (DBMem)

    Database memory (or DBMem) is the measurement of memory that is utilized by a Database Server.

    Actions to resize database memory are driven by data on the Database Server, which is more accurate than data on the hosting VM.
  - DB cache hit rate

    DB cache hit rate is the measurement of Database Server accesses that result in cache hits, measured as a percentage of hits versus total attempts. A high cache hit rate indicates efficiency.
  - Response time

    Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Transaction log

    Transaction log is the measurement of storage capacity utilized by a Database Server for transaction logging.
  - Transaction

    Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
  - Virtual storage

    Virtual ssorage is the measurement of virtual storage capacity that is in use.
- Virtual machine
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
  - Virtual CPU (vCPU)

Virtual CPU is the measurement of CPU that is in use.

- Virtual storage

  Virtual ssorage is the measurement of virtual storage capacity that is in use.

- Storage access (IOPS)

  Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

Actions

Workload Optimization Manager supports the following actions:

- Database server

  Resize

  - Connections

    Workload Optimization Manager uses connection data to generate memory resize actions for on-prem Database Servers.

  - Database memory (DBMem)

    Actions to resize database memory are driven by data on the Database Server, which is more accurate than data on the hosting VM. Workload Optimization Manager uses database memory and cache hit rate data to decide whether resize actions are necessary.

    A high cache hit rate value indicates efficiency. The optimal value is 100% for on-prem (self-hosted) Database Servers, and 90% for cloud Database Servers. When the cache hit rate reaches the optimal value, no action generates even if database memory utilization is high. If utilization is low, a resize down action generates.

    When the cache hit rate is less than the optimal value but database memory utilization remains low, no action generates. If utilization is high, a resize up action generates.

  - Transaction log

    Resize actions based on the transaction log resource depend on support for virtual storage in the underlying hypervisor technology.

    Currently, Workload Optimization Manager does not support resize actions for Oracle and Database Servers on the Hyper-V platform (due to the lack of API support for virtual storage).

- Virtual machine
  - Provision additional resources (VMem, VCPU)
  - Move Virtual Machine
  - Move Virtual Machine Storage
  - Reconfigure Storage
  - Reconfigure Virtual Machine
  - Suspend VM
  - Provision VM

NOTE:
Without separate targets to discover Guest OS Processes or Application Servers, Workload Optimization Manager does not generate actions on applications. Instead, it generates resize actions on the host VMs. For on-prem environments, if host utilization is high enough on the host running the application VM, Workload Optimization Manager can also recommend provisioning a new host.

To retrieve the IP address and DNS name of the node inside the SQL Server cluster, Workload Optimization Manager might enable advanced tools and `xp_cmdshell` options and run commands against `nslookup`.

Container platform targets

Workload Optimization Manager supports Kubernetes and Red Hat OpenShift. It connects to these container platforms through an agent called Kubeturbo, which you must deploy to each cluster that you want Workload Optimization Manager to manage. After the deployment, Workload Optimization Manager discovers, monitors, and generates actions to optimize the containerized workloads in your clusters.

NOTE:
In the supply chain, Workload Optimization Manager can link your containerized workloads with managed applications discovered from certain APM targets. For example, the supply chain can show the full application stack if your container platform environment includes applications managed by any Instana *(on page 203)*, Cisco AppDynamics *(on page 183)*, Dynatrace *(on page 194)*, or New Relic *(on page 219)* target that you added to Workload Optimization Manager.

Getting started with Kubeturbo

Review the following topics to get started with Kubeturbo:

- Connecting to Kubernetes clusters *(on page 244)*
- Connecting to Red Hat OpenShift clusters *(on page 277)*

Kubeturbo deployment requirements

Workload Optimization Manager collects information from Kubernetes or Red Hat OpenShift through an agent called Kubeturbo, which you must deploy to each cluster that you want Workload Optimization Manager to manage. Kubeturbo is a single deployment that runs on one worker node in a given cluster. It is not a DaemonSet.

NOTE:
Each Workload Optimization Manager release includes a corresponding Kubeturbo version. For example, Workload Optimization Manager 3.13.0 includes Kubeturbo 8.15.0. The Kubeturbo version must map to the corresponding Workload Optimization Manager version to be in a supported configuration. If the versions do not map, there can be discovery issues and target failures. For more information about version mapping, see this reference *(on page 302)*.

Be sure to update the Kubeturbo version *(on page 295)* in your deployment after your Workload Optimization Manager instance updates.

Minimum requirements

Requirement	Details
Container platform version	The following versions are supported: ▪ Kubernetes 1.27 up to the latest supported GA version, including (but not limited to) Amazon Elastic Kubernetes Service (EKS), Azure Kubernetes Service (AKS), Google Kubernetes Engine (GKE), IBM Cloud Kubernetes Service (IKS), and Rancher ▪ Red Hat supported GA versions of Red Hat OpenShift 4.1x, including (but not limited to) Red Hat OpenShift Kubernetes Service (ROKS), Red Hat OpenShift on AWS (ROSA), Red Hat OpenShift on Azure (ARO), and Red Hat OpenShift on Google For more information, see this topic *(on page 236)*.

Requirement	Details
Workload Optimization Manager instance and credentials	Your Workload Optimization Manager instance must run with a valid trial or premium license, and should be up and running. Kubeturbo can use the following credentials to connect to your Workload Optimization Manager instance. ▪ (Recommended) OAuth 2.0 `clientId` and `clientSecret` ▪ (Not recommended) Local user account username and password (in Base64 in a secret) For more information, see this topic *(on page 236)*.
Container image repository	The node to which Kubeturbo deploys must have internet access to pull the Kubeturbo container images from IBM Container Registry (`icr.io`). You can configure a private repository if you do not want to pull from IBM Container Registry. For more information, see this topic *(on page 242)*.
Network	▪ The Kubeturbo pod requires access to the `apiserver` and the `kubelet` on every node. The `kubelet` network is `https + port=10250` (default port). ▪ Kubeturbo requires HTTPS/TCP port 443 to communicate with Workload Optimization Manager. Kubeturbo uses WebSocket Secure (WSS) over HTTPS with TLS 1.2+. ▪ If you use a proxy server to establish a connection to Workload Optimization Manager, allow WebSocket Secure (WSS) protocol communication through the proxy.
Cluster roles	To deploy Kubeturbo to a cluster, you must have the `cluster-admin` role in the given cluster. This role has sufficient privileges to create a namespace and cluster role binding for the service account. By default, Kubeturbo deploys to your cluster with the `cluster-admin` role. This role has full control over every resource in the cluster. You can assign a custom role during deployment. For more information, see this topic *(on page 239)*.

Kubeturbo resources

Kubeturbo can run as a single pod deployment or can be deployed through an operator with the following resources:

- Namespace or project. The default is `turbo`.
- Service account
- Cluster role binding
- Custom Resource Definition (CRD)
- `ConfigMap` with information to connect to Workload Optimization Manager
- Kubeturbo `deployment` (YAML, Operator, OperatorHub, or Helm)

The following resources are optional:

- Operator
- Secret
- Cluster role

By default, Kubeturbo deploys without limits or requests. If you prefer to set limits or requests, the amount of resources required for Kubeturbo is related to the number of workloads and pods that it manages. For more information, see this topic *(on page 242)*.

Custom name for the Kubeturbo namespace

By default, Kubeturbo deploys to a namespace called `turbo`. If you want to specify a custom name during deployment, be sure to provide a name that meets the following requirements:

- The name must consist of lower case alphanumeric characters or the dash symbol (`-`), and must start and end with an alphanumeric character. For example, `my-name` and `123-abc` are valid names.
- For a production cluster, consider not using the `default` namespace.
- Avoid creating namespaces with the prefix `kube-`. This prefix is reserved for Kubernetes system namespaces.

Kubeturbo supported platforms

Kubeturbo supports the following container platforms:

- Kubernetes 1.27 up to the latest supported GA version, including (but not limited to) Amazon Elastic Kubernetes Service (EKS), Azure Kubernetes Service (AKS), Google Kubernetes Engine (GKE), IBM Cloud Kubernetes Service (IKS), and Rancher
- Red Hat supported GA versions of Red Hat OpenShift 4.1x, including (but not limited to) Red Hat OpenShift Kubernetes Service (ROKS), Red Hat OpenShift on AWS (ROSA), Red Hat OpenShift on Azure (ARO), and Red Hat OpenShift on Google

Supported platforms

Kubeturbo supports x86, IBM Power (ppc64le), IBM LinuxONE (OS/390), and Linux on IBM Z platforms. A Linux node is required to run Kubeturbo.

By default, Kubeturbo deploys on any non-control plane node that can be scheduled, such as worker, app, agent, and compute nodes.

NOTE:
Container images are no longer compatible with Dockershim container runtimes.

Workload Optimization Manager instance and credentials for Kubeturbo

Kubeturbo connects to Workload Optimization Manager to manage containerized workloads in your clusters. For seamless connection, be sure that Workload Optimization Manager is deployed with a valid trial or premium license, and is up and running.

Workload Optimization Manager instance information

One Kubeturbo agent can communicate with only one Workload Optimization Manager instance. If your environment has multiple instances, configure one Kubeturbo agent per Workload Optimization Manager instance. Multiple Kubeturbo agents can occupy the same namespace or separate namespaces.

Record the following information for your Workload Optimization Manager instance. You will need this information when you deploy Kubeturbo.

- Workload Optimization Manager instance URL, such as `https://10.1.1.1` or `https://myinstance.com`
- Workload Optimization Manager version

  Deployment requires the Kubeturbo image tag version.

  To determine the image tag version to specify, open the Workload Optimization Manager user interface, click the Help icon in the navigation menu, and then record the version number that displays. This version number has a corresponding image tag version. For example, if your Workload Optimization Manager instance version is 3.13.0, the corresponding image tag version is 8.15.0. For the latest Workload Optimization Manager and image tag version mapping, see this reference *(on page 302)*.
- Workload Optimization Manager instance credentials (see the next section)

Workload Optimization Manager instance credentials

Kubeturbo can use the following credentials to connect to your Workload Optimization Manager instance.

- (Recommended) OAuth 2.0 `clientId` and `clientSecret`
- (Not recommended) Local user account username and password (in Base64 in a secret)

Using OAuth 2.0 as credentials

Use the Workload Optimization Manager API to create and manage OAuth 2.0 client credentials. These credentials are more secure than the local account credentials created from the Workload Optimization Manager user interface.

To create the OAuth 2.0 client credentials, perform the following steps:

1. Log in to your Workload Optimization Manager instance.
2. In your browser's address bar, change the URL to `https://{your_instance_address}/swagger/#/Authorization/createClient`.

   For example, change the URL to `https://my-instance.com/swagger/#/Authorization/createClient`.
3. Click **Try it out**.
4. In the **body** section, replace the sample request with the following request. This request has all the required parameters for generating OAuth 2.0 credentials for Kubeturbo.

```
{
  "clientName": "kubeturbo",
  "grantTypes": [
    "client_credentials"
  ],
  "clientAuthenticationMethods": [
    "client_secret_post"
  ],
  "scopes": [
    "role:PROBE_ADMIN"
  ],
  "tokenSettings": {
    "accessToken": {
      "ttlSeconds": 600
    }
  }
}
```

5. Click **Execute** and then scroll to the **Server response** section. If the credentials were generated successfully, a response with a code of `200` displays.
6. In the response, find and record the `clientID` and `clientSecret` credentials. These credentials cannot be retrieved after you close the API so it is important that you record them.

The credentials that you created for Kubeturbo must be converted to Base64.

In Linux, you can run the following command to convert each credential to Base64.

```
echo {credential} | base64
```

NOTE:
If your credential fails to convert to Base64, it might have invalid or unsupported characters and must be changed before it can be converted to Base64 successfully.

Be sure to record the Base64 credentials for use in your Kubeturbo deployment.

Using local account username and password as credentials

NOTE:
Support for these credentials will be discontinued in a future release.

You can set up a local user account in the Workload Optimization Manager user interface. As part of setting up the account, you configure a username and password as credentials. When you deploy Kubeturbo, you can use a secret that stores the username and password, or specify the username and password in Base64.

Guidelines and instructions:

- Before creating a user account

  Review the following requirements before creating the user account that Kubeturbo will use to connect to your Workload Optimization Manager instance.

  - Kubeturbo requires a Workload Optimization Manager local user account with the site administrator or administrator role. This level of privilege allows Kubeturbo to automatically add itself as container platform target in the Workload Optimization Manager user interface. SSO user accounts are not supported.
  - If you manage Workload Optimization Manager user accounts through Active Directory (LDAP), the account can be a local or Active Directory user. For an Active Directory user, the format is `//`.
  - If you use SSO or Multi-Factor Authentication (MFA), the user account must be a local user and not an SSO user.
- Creating a user account

  For general instructions on creating the required user account, see this topic *(on page 1116)*. After you create the user account, record the plain text username and password.
- Converting the plain text username and password to Base64

  The credentials that you created for Kubeturbo must be converted to Base64.

  In Linux, you can run the following command to convert each credential to Base64.

  ```
  echo {credential} | base64
  ```

  NOTE:
  If your credential fails to convert to Base64, it might have invalid or unsupported characters and must be changed before it can be converted to Base64 successfully.

  Be sure to record the Base64 username and password. You will need this information when you deploy Kubeturbo.
- User account credentials accepted in Kubeturbo

  When you deploy Kubeturbo, you have the following options:

  - Use a secret that stores the username and password.

    Workload Optimization Manager recommends using a secret because it is more secure.

    The steps for configuring the secret vary by deployment method. Review the topic for a particular deployment method to see the relevant steps.

By default, Kubeturbo looks for and uses the credentials in a secret, if the secret exists. If a secret does not exist, Kubeturbo uses the plain text username and password.

– Specify the username and password in plain text. This option is not recommended because it is not secure.

Cluster roles for Kubeturbo

A cluster role specifies the permissions and privileges that are required to perform the following operations.

Operation	Supported role
Deploy Kubeturbo	To deploy Kubeturbo to a cluster, you must have the `cluster-admin` role in the given cluster. This role has sufficient privileges to create a namespace and cluster role binding for the service account.
Monitor and optimize workloads in your cluster	The role that you choose for Kubeturbo determines its level of access to your cluster. By default, Kubeturbo deploys to your cluster with the `cluster-admin` role. This role has full control over every resource in the cluster. If you prefer a custom role, you must explicitly set that role when you deploy Kubeturbo. The following custom roles are supported: ▪ `turbo-cluster-admin` custom role This custom role specifies the minimum permissions that Kubeturbo needs to monitor your workloads and execute the actions that Workload Optimization Manager generated to optimize these workloads. For more information on adding a reference to the YAML with the permissions, see this YAML resource. For more information on the `turbo-cluster-admin` custom role, see the next section in this topic. ▪ `turbo-cluster-reader` custom role This custom role is the least privileged role. It specifies the minimum permissions that Kubeturbo needs to monitor your workloads. Actions that Workload Optimization Manager generated to optimize these workloads can only be executed outside of Workload Optimization Manager (for example, in your cluster). For more information on adding a reference to the YAML with the permissions, see this YAML resource. For more information on `turbo-cluster-reader` custom role, see the last section in this topic.

Workload Optimization Manager cluster admin custom role

apiGroups	Resources	Verbs	Description
“” batch	pods jobs	"*"	Needed to take automated actions on all pods and jobs.
“” apps apps.openshift.io extensions turbonomic.com	deployments replicasets replicationcontrollers statefulsets	get list watch update patch	Needed to take automated resize actions on all of the resources in the list.

apiGroups	Resources	Verbs	Description
devops.turbonomic.io redis.redis.opstree-labs.in charts.helm.k8s.io	daemonsets deploymentconfigs resourcequotas operatorresourcemap-pings operatorresourcemap-pings/status redis xls		
“” apps batch extensions policy app.k8s.io argoproj.io apiextensions.k8s.io config.openshift.io policy.turbonomic.io	nodes services endpoints namespaces limitranges persistentvolumes persistentvolume-claims poddisruptionbudget cronjobs applications customresourcedefin-itions clusterversions slohorizontalscales containerverti-calscales policybindings	get list watch	Needed to discover all off the resources in the list.
machine.openshift.io	machines machinesets	get list update	Needed to automate node provision and suspend in Red Hat OpenShift using machinesets.
“”	nodes/spec nodes/stats nodes/metrics nodes/proxy nodes/log	get	Needed to discover all off the resources in the list.
security.openshift-.io	securitycontextcon-straints	list use	Needed in Red Hat OpenShift to use scc for au-tomated action.
“”	serviceaccounts	create delete impersonate	Needed to create, delete and impersonate service ac-counts for use in automated actions. Kubeturbo also cre-ates a separate service ac-count for every SCC in use

apiGroups	Resources	Verbs	Description
			for automating actions with resources using those cc.
rbac.authorization-.k8s.io	roles rolebindings clusterroles clusterrolebindings	create delete update	Needed to create the required resources in the cluster to automate actions. Kubeturbo will autocreate such resources based on the role and update them overtime as needed.
“”	secrets	get watch	Needed for Kubeturbo to read the secret that stores the details on how to connect to .

Workload Optimization Manager cluster reader custom role

apiGroups	Resources	Verbs	Description
“” apps app.k8s.io apps.openshift.io batch extensions turbonomic.com devops.turbonomic.io policy.turbonomic.io config.openshift.io	nodes pods deployments replicasets replicationcon-trollers services endpoints namespaces limitranges resourcequotas persistentvolumes persistentvolume-claims applications jobs cronjobs statefulsets daemonsets deploymentconfigs operatorresourcemap-pings clusterversions slohorizontalscales containerverti-calscales policybindings	get watch list	Discover/read all of the resources.
machine.openshift.io	machines machinesets	get list	Discover/read all of the machinesets in Red Hat OpenShift.

apiGroups	Resources	Verbs	Description
""	nodes/spec nodes/stats nodes/metrics nodes/proxy	get	Discover/read all of the nodes details.

Container image repository

The node to which Kubeturbo deploys must have internet access to pull the Kubeturbo container images from IBM Container Registry. You can configure a private repository if you prefer to pull from this repository.

Kubeturbo container images

Workload Optimization Manager gathers information from your clusters through the Kubeturbo container images that are available through a public repository in IBM Container Registry (`icr.io`).

The following list describes the different Kubeturbo container images.

- Container image for Kubeturbo

  This is the primary image and is pulled from `icr.io/cpopen/turbonomic/kubeturbo:<version>`.
- Container image for the CPU frequency getter job

  The `cpufreqgetter` image (also known as BusyBox) runs a job on every node to collect CPU speed. This image is pulled from `icr.io/cpopen/turbonomic/cpufreqgetter`.

  For more information about the parameters associated with this job, see this topic *(on page 304)*.
- (Operator method only) Container image for `kubeturbo-operator`

  This image is pulled for `icr.io/cpopen/kubeturbo-operator:<version>`

Working with a private repository

You can configure a private repository to store the Kubeturbo images in your environment. The steps for configuring a private repository are described in the respective deployment topics.

If you are using a private repository, be aware that IBM Container Registry provides multi-architecture container images. Your private repository needs to support loading these images. For example, multi-architecture container images are supported in Artifactory version 7.2+, but not in Artifactory version 6. Optionally, pull only the architecture version that you need. Specify the `--platform` parameter with `docker pull` and use `docker inspect` to confirm the architecture.

```
docker pull --platform linux/arm64 icr.io/cpopen/turbonomic/kubeturbo:8.15.0
docker inspect 49a61c21c3a1 | grep architecture
                "architecture": "aarch64",
```

Kubeturbo resource limits

By default, Kubeturbo deploys without limits or requests. If you prefer to set limits or requests, the amount of resources required for Kubeturbo is related to the number of workloads and pods that it manages. A workload is defined as a unique workload controller, such as `deployment foo` and `statefulset bar`.

After you deploy Kubeturbo, you can view the number of workloads in the Workload Optimization Manager supply chain. Set the scope to a single container platform cluster and then view the Workload Controller entity.

> **NOTE:**
> Kubeturbo discovers, collects, and processes metrics for container platform resources *(on page 286)* every one minute and then sends the data to Workload Optimization Manager for analysis every 10 minutes. Since data is collected every one minute, Kubeturbo holds the data in memory consistently.

Use the following table as a guide for setting memory limits:

Number of pods	Number of workload controllers	Recommended memory limit (Gi)
5,000	2,500	4
5,000	5,000	4
10,000	5,000	6
10,000	10,000	6.5
20,000	10,000	9.2
20,000	20,000	12
30,000	15,000	13
30,000	30,000	16

Points to consider:

- To avoid CPU throttling, do not set a CPU limit.
- If needed, memory requests can be set to 1 GB.
- If needed, CPU requests can be set to 1 core (1000 millicores).
- If Kubeturbo memory usage is very high and consistently growing over time, set a limit that is lower than the node memory capacity to limit the memory usage. Refer to the preceding table for guidance.
- To configure Kubeturbo container spec limits and requests, see the next sections.

YAML configuration: Kubeturbo container spec limit and request

1. Edit the deployment.

   ```
   kubectl edit deployment kubeturbo -n turbo
   ```

2. In the `containers` section, find and update the `resources: {}` section as shown in the following example:

   ```
   resources:
     limits:
       memory: 16Gi
     requests:
       memory: 1Gi
   ```

   If the section does not exist, create it as described.

3. Save the changes.

   The `kubeturbo` pod restarts with the changes applied automatically.

4. Verify the changes.

   ```
   kubectl get pods -n turbo -o
    custom-columns='NAME:.metadata.name,MEMORY_REQ:spec.containers[].resources.requests.memory,MEM_LIM
   :spec.containers[].resources.limits.memory'
   ```

   The output should be similar to the following example. This output confirms that the memory limit and request was set correctly for `kubeturbo`.

   ```
   NAME MEMORY_REQ MEM_LIM
   kubeturbo-6cbfc74b68-dfvtk 1Gi 16Gi
   ```

Operator configuration: Kubeturbo container spec limit and request

1. Edit the deployment.

   ```
   kubectl edit kubeturbo kubeturbo-release -n turbo
   ```

2. In the `spec` section, find and update the `resources: {}` section as shown in the following example.

```
resources:
  limits:
    memory: 16Gi
  requests:
    memory: 1Gi
```

   If the section does not exist, create it as described.

3. Save the changes.

   The `kubeturbo-release` pod restarts with the changes applied automatically.

4. Verify the changes.

```
kubectl get pods -n turbo -o
 custom-columns='NAME:.metadata.name,MEMORY_REQ:spec.containers[].resources.requests.memory,MEM_LIM
:spec.containers[].resources.limits.memory'
```

   Verify that the output is similar to the following example. This output confirms that the memory limit and request was set correctly for `kubeturbo-release`.

```
NAME MEMORY_REQ MEM_LIM
kubeturbo-release-6cbfc74b68-dfvtk 1Gi 16Gi
```

Connecting to Kubernetes clusters

Kubernetes is a platform that manages containerized applications. Workload Optimization Manager discovers these containerized applications through the **Kubeturbo** agent that you deploy to each Kubernetes cluster.

> If you are running Red Hat OpenShift 4.1x on Linux (x86, IBM Power, or IBM LinuxONE platforms), the best practice is to deploy Kubeturbo using OperatorHub. For OperatorHub deployment instructions, see this topic *(on page 277)*.

After the deployment, Kubeturbo automatically adds itself as a Kubernetes target in the Workload Optimization Manager user interface, in **Settings > Target Configuration**. You do not need to configure any settings in the user interface.

Kubeturbo deployment requirements

Kubeturbo supports Kubernetes 1.27 up to the latest supported GA version, including (but not limited to) Amazon Elastic Kubernetes Service (EKS), Azure Kubernetes Service (AKS), Google Kubernetes Engine (GKE), IBM Cloud Kubernetes Service (IKS), and Rancher.

Before deploying Kubeturbo, be sure to:

- Review the general requirements *(on page 234)*.
- Set up and record the credentials *(on page 236)* for your Workload Optimization Manager instance. You will specify these credentials when you deploy Kubeturbo.

Kubeturbo deployment methods

You can deploy Kubeturbo through YAML, Helm, or an operator. For deployment instructions, see the following topics:

- Deploying Kubeturbo through YAML *(on page 245)*
- Deploying Kubeturbo through Helm *(on page 253)*
- Deploying Kubeturbo through an operator *(on page 268)*

Monitoring and optimization of Kubernetes clusters

After connecting to your Kubernetes clusters, Workload Optimization Manager monitors and optimizes the resources that it discovered. See the following topics for more information:

- Container Platform Monitored Resources *(on page 286)*
- Container Platform Actions *(on page 289)*

Deploying Kubeturbo through YAML

This topic describes the single YAML resource that specifies the configurations for your Kubeturbo deployment. Update the required parameters in this YAML resource before deploying Kubeturbo.

Individual YAML resources with specific configurations are available here for your reference.

Deployment requirements

Before deploying Kubeturbo, be sure to:

- Review the general requirements *(on page 234)*.
- Set up and record the credentials *(on page 236)* for your Workload Optimization Manager instance. You will specify these credentials when you deploy Kubeturbo.

Task overview

To deploy Kubeturbo, perform the following tasks:

1. Download the YAML resource for the Kubeturbo role.
2. Update the YAML resource with the minimum required parameters.
3. Deploy the YAML resource to your cluster.

   After you perform this task, Kubeturbo deploys to your cluster and the Kubeturbo pod starts.
4. Validate the deployment.

NOTE:
Workload Optimization Manager gathers information from your clusters through the Kubeturbo container images. By default, these images are pulled from IBM Container Registry (`icr.io`). If you prefer to pull these images from your private repository, configure that repository before deploying Kubeturbo. For more information about private repositories, see this topic *(on page 252)*.

Downloading the YAML resource for the Kubeturbo role

The role that you choose for Kubeturbo determines its level of access to your cluster.

Choose from the following roles and then download the corresponding YAML resource.

- Option 1: Default role

  By default, Kubeturbo deploys to your cluster with the `cluster-admin` role. This role has full control over every resource in the cluster.

  To download the YAML resource for this role, run the following command:

  ```
  curl -O
   https://raw.githubusercontent.com/turbonomic/kubeturbo-deploy/master/deploy/kubeturbo_yamls/turbo_
  kubeturbo_full.yaml
  ```

- Option 2: `turbo-cluster-admin` custom role

This custom role specifies the minimum permissions that Kubeturbo needs to monitor your workloads and execute the actions that Workload Optimization Manager generated to optimize these workloads.

To download the YAML resource for this role, run the following command:

```
curl -O
 https://raw.githubusercontent.com/turbonomic/kubeturbo-deploy/master/deploy/kubeturbo_yamls/kubeturbo_least_admin_full.yaml
```

- Option 3: `turbo-cluster-reader` custom role

  This custom role is the least privileged role. It specifies the minimum permissions that Kubeturbo needs to monitor your workloads. Actions that Workload Optimization Manager generated to optimize these workloads can only be executed outside of Workload Optimization Manager (for example, in your cluster).

  To download the YAML resource for this role, run the following command:

```
curl -O
 https://raw.githubusercontent.com/turbonomic/kubeturbo-deploy/master/deploy/kubeturbo_yamls/kubeturbo_reader_full.yaml
```

Updating the YAML resource with the minimum required parameters

Update the YAML resource that you downloaded. For example, you can update the YAML resource in VS Code or `vi`. Be sure to set up and record the credentials *(on page 236)* for your Workload Optimization Manager instance before performing this task.

IMPORTANT:
The YAML resource that you downloaded specifies the following default values:

- Namespace – `turbonomic`
- Secret – `turbonomic-credentials`

It is recommended that you leave these values unchanged because the same values are specified in other Kubeturbo resources. Changing the values in the YAML resource but not in the other Kubeturbo resources could result in deployment issues.

1. Update the `turboServer` parameter.

```
                "turboServer": "{your_instance_address}"
            },
```

   Specify the address of your Workload Optimization Manager instance, such as `https://10.1.1.1` or `https://myinstance.com`.

2. Update the `targetName` parameter.

```
        "targetConfig": {
            "targetName":"{your_cluster_name}"
        },
```

   Specify the cluster name that will display in the Workload Optimization Manager user interface. Spaces are not allowed.

3. Update the `image` parameter.

```
        image: icr.io/cpopen/turbonomic/kubeturbo:{your_image_tag_version}
```

   Specify the image tag version.

   To determine the image tag version to specify, open the Workload Optimization Manager user interface, click the Help icon in the navigation menu, and then record the version number that displays. This version number has a corresponding image tag version. For example, if your Workload Optimization Manager instance version is 3.13.0, the corresponding image tag version is 8.15.0. For the latest Workload Optimization Manager and image tag version mapping, see this reference *(on page 302)*.

4. Update the Kubeturbo credentials. Choose from the following options:
   - (Recommended) OAuth 2.0 `clientId` and `clientSecret`

     Use the Workload Optimization Manager API to create and manage OAuth 2.0 client credentials. These credentials are more secure than the local account credentials created from the Workload Optimization Manager user interface.

     To create the OAuth 2.0 client credentials, perform the following steps:

     a. Log in to your Workload Optimization Manager instance.

     b. In your browser's address bar, change the URL to `https://{your_instance_address}/swagger/#/Authorization/createClient`.

        For example, change the URL to `https://my-instance.com/swagger/#/Authorization/createClient`.

     c. Click **Try it out**.

     d. In the **body** section, replace the sample request with the following request. This request has all the required parameters for generating OAuth 2.0 credentials for Kubeturbo.

```
{
  "clientName": "kubeturbo",
  "grantTypes": [
    "client_credentials"
  ],
  "clientAuthenticationMethods": [
    "client_secret_post"
  ],
  "scopes": [
    "role:PROBE_ADMIN"
  ],
  "tokenSettings": {
    "accessToken": {
      "ttlSeconds": 600
    }
  }
}
```

     e. Click **Execute** and then scroll to the **Server response** section. If the credentials were generated successfully, a response with a code of `200` displays.

     f. In the response, find and record the `clientID` and `clientSecret` credentials. These credentials cannot be retrieved after you close the API so it is important that you record them.

     The credentials that you created for Kubeturbo must be converted to Base64.

     In Linux, you can run the following command to convert each credential to Base64.

```
echo {credential} | base64
```

     NOTE:
     If your credential fails to convert to Base64, it might have invalid or unsupported characters and must be changed before it can be converted to Base64 successfully.

     After the conversion, specify the Base64 `clientId` and `clientSecret` in the YAML file that you are updating, as shown in the following example.

```
data:
  clientid: {Base64_client_ID}
  clientsecret: {Base64_client_secret}
```

   - (Not recommended) Local user account username and password (in Base64 in a secret)

NOTE:
Support for these credentials will be discontinued in a future release.

Comment the `clientId` and `clientSecret` parameters, and then specify the `username` and `password` in Base64.

```
data:
  username: {Base64_username}
  password: {Base64_password}
```

5. If you want to enable move actions for pods with persistent volumes, uncomment the following parameter.

```
#- --fail-volume-pod-moves=false
```

For example:

```
- --fail-volume-pod-moves=false
```

6. (Optional) If you need to adjust or specify other parameters, see the last section in this topic, 'Kubeturbo custom resource values'.

Deploying the YAML resource to your cluster

1. Log in to your cluster through the command line.
2. Deploy the YAML resource.

```
kubectl apply -f {YAML_name}
```

Replace *{YAML_name}* with the YAML resource that you downloaded and updated. Choose from the following:

- `turbo_kubeturbo_full.yaml`
- `kubeturbo_reader_full.yaml`
- `kubeturbo_least_admin_full.yaml`

After you run the command, Kubeturbo deploys to your cluster and the Kubeturbo pod starts.

Validating the deployment

1. Verify the status of the Kubeturbo pod.

```
kubectl get pods -n turbo
```

The following example result indicates that the pod was deployed and is currently running.

```
NAME                    READY   STATUS    RESTARTS   AGE
kubeturbo-asdf1234asd3  1/1     Running   0          37m
```

2. If you use OAuth 2.0 credentials for Kubeturbo, verify the that the credentials are configured correctly. In the Kubeturbo logs, the following log example indicates that the credentials are configured correctly.

```
I0723 20:21:46.659559       1 tap_service.go:114] Secure credentials is provided, target
 Kubernetes-ocp-414 could be auto-registered if the server is running in secure mode and secure
 connection is established
```

3. Open the Workload Optimization Manager user interface and navigate to **Settings > Target Configuration**. If the deployment was successful, a new container platform target appears in the list.

If you do not see the target, there might be a deployment issue that you need to resolve.

- To start troubleshooting an issue, review the Kubeturbo logs. You can retrieve logs by running the following command.

  ```
  kubectl logs kubeturbo-{kubeturbo_pod_ID} -n turbo
  ```

  For example:

  ```
  kubectl logs kubeturbo-asdf1234asd3 -n turbo
  ```

- If you need further assistance, contact your Workload Optimization Manager representative.

Kubeturbo custom resource values

The following table describes the values that you can specify in the Kubeturbo custom resource to configure your deployment.

Refer to the Kubeturbo CRD for the schema.

Parameter	Default value	Changes to default value	Parameter type
`args.cleanupSccImpersonationResources`	`true`	Optional	Clean up the resources for SCC impersonation by default. For details about SCC impersonation, see this topic *(on page 609)*.
`args.discovery-interval-sec`	None	Optional	Number in seconds
`args.discovery-sample-interval`	None	Optional	Number in seconds
`args.discovery-samples`	None	Optional	Number
`args.discovery-timeout-sec`	None	Optional	Number in seconds
`args.discovery-workers`	None	Optional	Number
`args.garbage-collection-interval`	None	Optional	Number in seconds
`args.kubelethttps`	`true`	Optional, change to `false` if using Kubernetes 1.10 or older	Boolean
`args.kubeletport`	`10250`	Optional, change to `10255` if using Kubernetes 1.10 or older	Number
`args.logginglevel`	`2`	Optional	Number. A high value increases logging.
`args.pre16k8sVersion`	`false`	Optional	If Kubernetes version is older than 1.6, then add another `arg` for move/resize actions.
`args.sccsupport`	None	Optional	Enabled by default in clusters that use SCC.

Parameter	Default value	Changes to default value	Parameter type
`args.stitchuuid`	`true`	Optional, change to `false` if IaaS is VMM or Hyper-V	Boolean
`daemonPodDetectors.daemonPodNamespaces1` and `daemonPodNamespaces2`	`daemonSet` kinds are allowed for node suspension by default. Adding this parameter changes the default.	Optional but required to identify pods in the namespace to be ignored for cluster consolidation.	Regular expression used, values are in quotes and are comma-separated, such as `"kube-system","kube-service-catalog","openshift-.*"`.
`daemonPodDetectors.daemonPodNamePatterns`	`daemonSet` kinds are allowed for node suspension by default. Adding this parameter changes the default.	Optional but required to identify pods matching this pattern to be ignored for cluster consolidation.	Regular expression used `.*ignorepod.*`
`HANodeConfig.nodeRoles`	Any value for label key value `node-role.kubernetes.io/` for master and others	Optional. This is used to automate policies that keep nodes of the same role limited to one instance per ESX host or availability zone.	Values in `values.yaml` or `cr.yaml` use escapes or quotes, or are comma-separated. Master nodes are by default. Other roles populated by `nodeRoles:""foo","bar""`
`image.busyboxRepository`	`icr.io/cpopen/turbonomic/cpufreqgetter:1.0`	Optional	Full path to the repository, image, and tag.
`image.cpufreqgetterRepository`	`icr.io/cpopen/turbonomic/cpufreqgetter`	Optional	Repository used to get node `cpufrequency`.
`image.imagePullSecret`	None	Optional	Defines the secret used to authenticate to the container image registry
`image.pullPolicy`	`IfNotPresent`	Optional	
`image.repository`	`icr.io/cpopen/turbonomic/kubeturbo`	Optional	Path to the repository. Must be used with `image.tag`.
`image.tag`	Depends on product version	Optional	Kubeturbo tag. Must be used with `image.repository`.
`kubeturboPodScheduling.affinity`	None	Optional	Specify `affinity` pod scheduling constraint in the cluster. For examples, see the Kubernetes documentation
`kubeturboPodScheduling.nodeSelector`	None	Optional	Specify `nodeSelector` pod scheduling constraint in the cluster. For examples, see the Kubernetes documentation.
`kubeturboPodScheduling.tolerations`	None	Optional	Specify `tolerations` pod scheduling constraint in the cluster. For examples, see

Parameter	Default value	Changes to default value	Parameter type
			the Kubernetes documentation.
masterNodeDetectors.nodeLabels	Any value for label key value node-role.kubernetes.io/master	Deprecated. Previously used to avoid suspending masters identified by node label key value pair. The value ignored if there is no match.	Regular expression used, specify the key as masterNodeDetectors.nodeLabelsKey (such as node-role.kubernetes.io/master) and the value as masterNodeDetectors.nodeLabelsValue (such as .*).
masterNodeDetectors.nodeNamePatterns	Node name includes .*master.*	Deprecated. Previously used to avoid suspending masters identified by node name. The value is ignored if there is no match.	String, regular expression used. For example: .*master.*
restAPIConfig.opsManagerPassword	None	Optional	Administrator password.
restAPIConfig.opsManagerUserName	None	Optional	Local or Active Directory user with site administrator role.
restAPIConfig.turbonomicCredentialsSecretName	turbonomic-credentials	Required only if using secret and not taking the default secret name	Secret that contains the Workload Optimization Manager site administrator username and password in Base64.
roleBinding	turbo-all-binding	Optional	Specify the name of clusterrolebinding
roleName	cluster-admin	Optional	Specify custom turbo-cluster-reader or turbo-cluster-admin role instead of the default cluster-admin role.
serverMeta.proxy	None	Optional	Format of http://username:password@proxyserver:proxyport or http://proxyserver:proxyport
serverMeta.turboServer	None	Required	HTTPS URL to log in to Workload Optimization Manager
serverMeta.version	None	Optional	Number x.y
serviceAccountName	turbo-user	Optional	Specify the name of the service account.
targetConfig.targetName	{Your_cluster}	Optional but required for multiple clusters	String that indicates how you want to identify your cluster.

Configuring a private repository (YAML)

Workload Optimization Manager gathers information from your clusters through the Kubeturbo container images that are available through a public repository in IBM Container Registry (`icr.io`).

The following list describes the different Kubeturbo container images.

- Container image for Kubeturbo

  This is the primary image and is pulled from `icr.io/cpopen/turbonomic/kubeturbo:<version>`.

- Container image for the CPU frequency getter job

  The `cpufreqgetter` image (also known as BusyBox) runs a job on every node to collect CPU speed. This image is pulled from `icr.io/cpopen/turbonomic/cpufreqgetter`.

  For more information about the parameters associated with this job, see this topic *(on page 304)*.

If you are using a private repository, be aware that IBM Container Registry provides multi-architecture container images. Your private repository needs to support loading these images. For example, multi-architecture container images are supported in Artifactory version 7.2+, but not in Artifactory version 6. Optionally, pull only the architecture version that you need. Specify the `--platform` parameter with `docker pull` and use `docker inspect` to confirm the architecture.

```
docker pull --platform linux/arm64 icr.io/cpopen/turbonomic/kubeturbo:8.15.0
docker inspect 49a61c21c3a1 | grep architecture
               "architecture": "aarch64",
```

To use a private repository, you need to define additional parameters in the `deployment` resource, as shown in the following example:

```
kind: Deployment
metadata:
  name: kubeturbo
  namespace: turbo
spec:
  replicas: 1
  selector:
     matchLabels:
       app.kubernetes.io/name: kubeturbo
  template:
    spec:
      containers:
      - name: kubeturbo
        image: {your_kubeturbo_repository}/icr.io/cpopen/turbonomic/kubeturbo:8.15.0
        args:
          - --turboconfig=/etc/kubeturbo/turbo.config
          - --v=2
          - --kubelet-https=true
          - --kubelet-port=10250
          -
 --cpufreqgetter-image={your_frequency_getter_repository}/icr.io/cpopen/turbonomic/cpufreqgetter
          # If using busybox parameter, uncomment to override default, and specify your own location
          #- --busybox-image=docker.io/busybox
          # or uncomment the following to pull from RHCC
          #- --busybox-image=registry.access.redhat.com/ubi8/ubi-minimal
          # Uncomment to specify the secret name which holds the credentials to busybox OR cprfreqgetter
 image
          #- --busybox-image-pull-secret=<secret-name>
```

Update the sample command with the following parameters:

- `image: {your_kubeturbo_repository}/icr.io/cpopen/turbonomic/kubeturbo`

  Specify your private Kubeturbo repository, such as `myrepo.com`.

- `--cpufreqgetter-image={your_frequency_getter_repository}/icr.io/cpopen/turbonomic/cpufreqgetter`

  Specify your private CPU frequency getter repository, such as `mygetterrepo.com`.

Deploying Kubeturbo through a Helm chart in a Helm chart repository

Helm manages the charts that package all resources associated with an application. For more information, see the Helm documentation.

You can deploy Kubeturbo using a Helm chart that is hosted in a Helm chart repository in IBM GitHub Pages. This chart delivery mechanism supports versioning, which enables seamless access to Kubeturbo updates and allows you to roll back Kubeturbo to a previous version if needed. It also simplifies your deployment process by eliminating the need to clone the Kubeturbo repository.

Deployment requirements

- General deployment requirements

  Before deploying Kubeturbo, be sure to:

  - Review the general requirements *(on page 234)*.
  - Set up and record the credentials *(on page 236)* for your Workload Optimization Manager instance. You will specify these credentials when you deploy Kubeturbo.

- Helm

  For instructions on installing Helm, see the Helm documentation.

  This topic assumes that you are familiar with Helm usage and chart repositories.

Task overview

To deploy Kubeturbo, perform the following tasks:

1. Add the Helm chart repository for Kubeturbo.
2. Create the YAML resource for the Kubeturbo namespace and credentials.
3. Deploy Kubeturbo.

   Deployment creates the following resources in the cluster:

   - Service account and binding to the Kubeturbo cluster role
   - Updated `configMap` containing the required information for Kubeturbo to connect to Workload Optimization Manager
   - Deployment of Kubeturbo
4. Validate the deployment.

NOTE:
Workload Optimization Manager gathers information from your clusters through the Kubeturbo container images. By default, these images are pulled from IBM Container Registry (`icr.io`). If you prefer to pull these images from your private repository, configure that repository before deploying Kubeturbo. For more information about private repositories, see this topic *(on page 268)*.

Adding the Helm chart repository for Kubeturbo

Run the following command to add the Helm chart repository for Kubeturbo to your local Helm registry.

```
helm repo add turbonomic https://ibm.github.io/turbonomic-container-platform/
```

You can run the following commands for additional options.

- Retrieve the latest versions of the packages.

  ```
  helm repo update
  ```

- See the available charts and their version numbers.

```
helm search repo turbonomic
```

NOTE:
If you need to uninstall the chart later, run `helm delete my-kubeturbo`.

Creating the YAML resource for the Kubeturbo namespace and credentials

This YAML resource specifies the Kubeturbo namespace and Workload Optimization Manager instance credentials. Be sure to set up and record the credentials *(on page 236)* before performing this task.

1. Download the YAML resource.

```
curl -O
 https://raw.githubusercontent.com/turbonomic/kubeturbo-deploy/master/deploy/kubeturbo_yamls/kubeturbo_namespace_turbo_credentials_secret.yaml
```

2. Update the Kubeturbo credentials in the YAML resource that you downloaded. For example, you can update the YAML resource in VS Code or `vi`.

   IMPORTANT:
   The YAML resource that you downloaded specifies the following default values:

   - Namespace – `turbonomic`
   - Secret – `turbonomic-credentials`

   It is recommended that you leave these values unchanged because the same values are specified in other Kubeturbo resources. Changing the values in the YAML resource but not in the other Kubeturbo resources could result in deployment issues.

   Choose from the following options:

   - (Recommended) OAuth 2.0 `clientId` and `clientSecret`

     Use the Workload Optimization Manager API to create and manage OAuth 2.0 client credentials. These credentials are more secure than the local account credentials created from the Workload Optimization Manager user interface.

     To create the OAuth 2.0 client credentials, perform the following steps:

     a. Log in to your Workload Optimization Manager instance.

     b. In your browser's address bar, change the URL to `https://{your_instance_address}/swagger/#/Authorization/createClient`.

        For example, change the URL to `https://my-instance.com/swagger/#/Authorization/createClient`.

     c. Click **Try it out**.

     d. In the **body** section, replace the sample request with the following request. This request has all the required parameters for generating OAuth 2.0 credentials for Kubeturbo.

```
{
  "clientName": "kubeturbo",
  "grantTypes": [
    "client_credentials"
  ],
  "clientAuthenticationMethods": [
    "client_secret_post"
  ],
  "scopes": [
    "role:PROBE_ADMIN"
```

```
  ],
  "tokenSettings": {
    "accessToken": {
      "ttlSeconds": 600
    }
  }
}
```

e. Click **Execute** and then scroll to the **Server response** section. If the credentials were generated successfully, a response with a code of `200` displays.

f. In the response, find and record the `clientID` and `clientSecret` credentials. These credentials cannot be retrieved after you close the API so it is important that you record them.

The credentials that you created for Kubeturbo must be converted to Base64.

In Linux, you can run the following command to convert each credential to Base64.

```
echo {credential} | base64
```

NOTE:
If your credential fails to convert to Base64, it might have invalid or unsupported characters and must be changed before it can be converted to Base64 successfully.

After the conversion, specify the Base64 `clientId` and `clientSecret` in the YAML file that you are updating, as shown in the following example.

```
data:
  clientid: {Base64_client_ID}
  clientsecret: {Base64_client_secret}
```

- (Not recommended) Local user account username and password (in Base64 in a secret)

NOTE:
Support for these credentials will be discontinued in a future release.

Comment the `clientId` and `clientSecret` parameters, and then specify the `username` and `password` in Base64.

```
data:
  username: {Base64_username}
  password: {Base64_password}
```

3. Deploy the YAML resource.
   a. Log in to your cluster through the command line.
   b. Deploy the YAML resource.

```
kubectl apply -f kubeturbo_namespace_turbo_credentials_secret.yaml
```

Deploying Kubeturbo

1. (Optional) Perform a dry run to test your deployment.

```
helm install --dry-run \
--debug kubeturbo kubeturbo/deploy/kubeturbo \
--namespace turbo \
--set serverMeta.turboServer={your_instance_address} \
--set targetConfig.targetName={your_cluster_name}
```

Update the following parameters:

- `--set serverMeta.turboServer={your_instance_address}`

  Specify the address of your Workload Optimization Manager instance, such as `https://10.1.1.1` or `https://myinstance.com`.

- `--set targetConfig.targetName={your_cluster_name}`

  Specify the cluster name that will display in the Workload Optimization Manager user interface. Spaces are not allowed.

- If you want to enable move actions for pods with persistent volumes, add the following parameter.

```
--set args.failVolumePodMoves=false
```

Be sure to resolve any errors before proceeding to the next step.

2. Deploy Kubeturbo with a specific role by choosing one of the following options. The role that you choose for Kubeturbo determines its level of access to your cluster.

   - Option 1: Default role

     By default, Kubeturbo deploys to your cluster with the `cluster-admin` role. This role has full control over every resource in the cluster.

     If you want to use the default role, run the following command. This command does not explicitly specify a role parameter, which means that the default role will be used.

```
helm install kubeturbo kubeturbo/deploy/kubeturbo \
--namespace turbo \
--set serverMeta.turboServer={your_instance_address} \
--set targetConfig.targetName={your_cluster_name}
```

     For details about the parameters that you need to configure, see the previous step.

   - Option 2: `turbo-cluster-admin` custom role

     This custom role specifies the minimum permissions that Kubeturbo needs to monitor your workloads and execute the actions that Workload Optimization Manager generated to optimize these workloads.

     If you want to use this role, specify the role parameter, as shown in the last line of the following command.

```
helm install kubeturbo kubeturbo/deploy/kubeturbo \
--namespace turbo \
--set serverMeta.turboServer={your_instance_address} \
--set targetConfig.targetName={your_cluster_name} \
--set roleName=turbo-cluster-admin
```

     For details about the other parameters that you need to configure, see the previous step.

   - Option 3: `turbo-cluster-reader` custom role

     This custom role is the least privileged role. It specifies the minimum permissions that Kubeturbo needs to monitor your workloads. Actions that Workload Optimization Manager generated to optimize these workloads can only be executed outside of Workload Optimization Manager (for example, in your cluster).

     If you want to use this role, specify the role parameter, as shown in the last line of the following command.

```
helm install kubeturbo kubeturbo/deploy/kubeturbo \
--namespace turbo \
--set serverMeta.turboServer={your_instance_address} \
--set targetConfig.targetName={your_cluster_name} \
--set roleName=turbo-cluster-reader
```

     For details about the other parameters that you need to configure, see the previous step.

3. (Optional) If you need to adjust or specify other parameters, see the last section in this topic, 'Kubeturbo custom resource values'.

The deployment starts after you run the command.

Validating the deployment

1. Review the output in Helm. The following example indicates that the deployment was successful.

```
NAME: kubeturbo
LAST DEPLOYED: Thu Aug 10 15:42:16 2023
NAMESPACE: turbo
STATUS: deployed
REVISION: 1
TEST_SUITE: None
```

2. Verify the status of the Kubeturbo pod.

```
kubectl get pods -n turbo
```

The following example result indicates that the pod was deployed and is currently running.

```
NAME                    READY   STATUS    RESTARTS   AGE
kubeturbo-asdf1234asd3  1/1     Running   0          37m
```

3. If you use OAuth 2.0 credentials for Kubeturbo, verify the that the credentials are configured correctly. In the Kubeturbo logs, the following log example indicates that the credentials are configured correctly.

```
I0723 20:21:46.659559       1 tap_service.go:114] Secure credentials is provided, target
 Kubernetes-ocp-414 could be auto-registered if the server is running in secure mode and secure
 connection is established
```

4. Open the Workload Optimization Manager user interface and navigate to **Settings > Target Configuration**. If the deployment was successful, a new container platform target appears in the list.

   If you do not see the target, there might be a deployment issue that you need to resolve.

   - To start troubleshooting an issue, review the Kubeturbo logs. You can retrieve logs by running the following command.

     ```
     kubectl logs kubeturbo-{kubeturbo_pod_ID} -n turbo
     ```

     For example:

     ```
     kubectl logs kubeturbo-asdf1234asd3 -n turbo
     ```

   - If you need further assistance, contact your Workload Optimization Manager representative.

Reference: Kubeturbo custom resource values

The following table describes the values that you can specify in the Kubeturbo custom resource to configure your deployment.

Refer to the Kubeturbo CRD for the schema.

Parameter	Default value	Changes to default value	Parameter type
`args.cleanupSccImpersonationResources`	`true`	Optional	Clean up the resources for SCC impersonation by default. For details about SCC impersonation, see this topic *(on page 609)*.
`args.discovery-interval-sec`	None	Optional	Number in seconds
`args.discovery-sample-interval`	None	Optional	Number in seconds
`args.discovery-samples`	None	Optional	Number

Parameter	Default value	Changes to default value	Parameter type
`args.discovery-time-out-sec`	None	Optional	Number in seconds
`args.discovery-work-ers`	None	Optional	Number
`args.garbage-collec-tion-interval`	None	Optional	Number in seconds
`args.kubelethttps`	`true`	Optional, change to `false` if using Kubernetes 1.10 or older	Boolean
`args.kubeletport`	`10250`	Optional, change to `10255` if using Kubernetes 1.10 or older	Number
`args.logginglevel`	`2`	Optional	Number. A high value increases logging.
`args.pre16k8sVersion`	`false`	Optional	If Kubernetes version is older than 1.6, then add another `arg` for move/resize actions.
`args.sccsupport`	None	Optional	Enabled by default in clusters that use SCC.
`args.stitchuuid`	`true`	Optional, change to `false` if IaaS is VMM or Hyper-V	Boolean
`daemonPodDetectors.daemonPodNamespaces1` and `daemonPodNamespaces2`	`daemonSet` kinds are allowed for node suspension by default. Adding this parameter changes the default.	Optional but required to identify pods in the namespace to be ignored for cluster consolidation.	Regular expression used, values are in quotes and are comma-separated, such as `"kube-system","kube-service-catalog","openshift-.*"`.
`daemonPodDetectors.daemonPodNamePatterns`	`daemonSet` kinds are allowed for node suspension by default. Adding this parameter changes the default.	Optional but required to identify pods matching this pattern to be ignored for cluster consolidation.	Regular expression used `.*ignorepod.*`
`HANodeConfig.nodeRoles`	Any value for label key value `node-role.kubernetes.io/` for master and others	Optional. This is used to automate policies that keep nodes of the same role limited to one instance per ESX host or availability zone.	Values in `values.yaml` or `cr.yaml` use escapes or quotes, or are comma-separated. Master nodes are by default. Other roles populated by `nodeRoles:""foo","bar""`
`image.busyboxRepository`	`icr.io/cpopen/turbonomic/cpufreqgetter:1.0`	Optional	Full path to the repository, image, and tag.
`image.cpufreqgetterRepository`	`icr.io/cpopen/turbonomic/cpufreqgetter`	Optional	Repository used to get node `cpufrequency`.
`image.imagePullSecret`	None	Optional	Defines the secret used to authenticate to the container image registry

Parameter	Default value	Changes to default value	Parameter type
`image.pullPolicy`	`IfNotPresent`	Optional	
`image.repository`	`icr.io/cpopen/turbonomic/kubeturbo`	Optional	Path to the repository. Must be used with `image.tag`.
`image.tag`	Depends on product version	Optional	Kubeturbo tag. Must be used with `image.repository`.
`kubeturboPodScheduling.affinity`	None	Optional	Specify `affinity` pod scheduling constraint in the cluster. For examples, see the Kubernetes documentation
`kubeturboPodScheduling.nodeSelector`	None	Optional	Specify `nodeSelector` pod scheduling constraint in the cluster. For examples, see the Kubernetes documentation.
`kubeturboPodScheduling.tolerations`	None	Optional	Specify `tolerations` pod scheduling constraint in the cluster. For examples, see the Kubernetes documentation.
`masterNodeDetectors.nodeLabels`	Any value for label key value `node-role.kubernetes.io/master`	Deprecated. Previously used to avoid suspending masters identified by node label key value pair. The value ignored if there is no match.	Regular expression used, specify the key as `masterNodeDetectors.nodeLabelsKey` (such as `node-role.kubernetes.io/master`) and the value as `masterNodeDetectors.nodeLabelsValue` (such as `.*`).
`masterNodeDetectors.nodeNamePatterns`	Node name includes `.*master.*`	Deprecated. Previously used to avoid suspending masters identified by node name. The value is ignored if there is no match.	String, regular expression used. For example: `.*master.*`
`restAPIConfig.opsManagerPassword`	None	Optional	Administrator password.
`restAPIConfig.opsManagerUserName`	None	Optional	Local or Active Directory user with site administrator role.
`restAPIConfig.turbonomicCredentialsSecretName`	`turbonomic-credentials`	Required only if using secret and not taking the default secret name	Secret that contains the Workload Optimization Manager site administrator username and password in Base64.
`roleBinding`	`turbo-all-binding`	Optional	Specify the name of `clusterrolebinding`

Parameter	Default value	Changes to default value	Parameter type
`roleName`	`cluster-admin`	Optional	Specify custom `turbo-cluster-reader` or `turbo-cluster-admin` role instead of the default `cluster-admin` role.
`serverMeta.proxy`	None	Optional	Format of `http://username:password@proxyserver:proxyport` or `http://proxyserver:proxyport`
`serverMeta.turboServer`	None	Required	HTTPS URL to log in to Workload Optimization Manager
`serverMeta.version`	None	Optional	Number `x.y`
`serviceAccountName`	`turbo-user`	Optional	Specify the name of the service account.
`targetConfig.targetName`	`{Your_cluster}`	Optional but required for multiple clusters	String that indicates how you want to identify your cluster.

Deploying Kubeturbo through a Helm chart in a GitHub repository

Helm manages the charts that package all resources associated with an application. For more information, see the Helm documentation.

Deployment requirements

- General deployment requirements

  Before deploying Kubeturbo, be sure to:

  - Review the general requirements *(on page 234)*.
  - Set up and record the credentials *(on page 236)* for your Workload Optimization Manager instance. You will specify these credentials when you deploy Kubeturbo.

- Helm

  For instructions on installing Helm, see the Helm documentation.

  This topic assumes that you are familiar with Helm usage.

- Git

  You need Git to clone the Kubeturbo repository locally. For instructions on installing Git, see the GitHub documentation.

Task overview

To deploy Kubeturbo, perform the following tasks:

1. Clone the Kubeturbo repository.
2. Create the YAML resource for the Kubeturbo namespace and credentials.
3. Deploy Kubeturbo.

Deployment creates the following resources in the cluster:

- Service account and binding to the Kubeturbo cluster role
- Updated `configMap` containing the required information for Kubeturbo to connect to Workload Optimization Manager
- Deployment of Kubeturbo

4. Validate the deployment.

NOTE:
Workload Optimization Manager gathers information from your clusters through the Kubeturbo container images. By default, these images are pulled from IBM Container Registry (`icr.io`). If you prefer to pull these images from your private repository, configure that repository before deploying Kubeturbo. For more information about private repositories, see this topic *(on page 268)*.

Cloning the Kubeturbo repository

Clone the Kubeturbo repository to your local environment. The deployment will not work if you do not clone the repository.

```
git clone https://github.com/turbonomic/kubeturbo-deploy.git
```

Creating the YAML resource for the Kubeturbo namespace and credentials

This YAML resource specifies the Kubeturbo namespace and Workload Optimization Manager instance credentials. Be sure to set up and record the credentials *(on page 236)* before performing this task.

1. Download the YAML resource.

```
curl -O
 https://raw.githubusercontent.com/turbonomic/kubeturbo-deploy/master/deploy/kubeturbo_yamls/kubeturbo_namespace_turbo_credentials_secret.yaml
```

2. Update the Kubeturbo credentials in the YAML resource that you downloaded. For example, you can update the YAML resource in VS Code or `vi`.

IMPORTANT:
The YAML resource that you downloaded specifies the following default values:

- Namespace – `turbonomic`
- Secret – `turbonomic-credentials`

It is recommended that you leave these values unchanged because the same values are specified in other Kubeturbo resources. Changing the values in the YAML resource but not in the other Kubeturbo resources could result in deployment issues.

Choose from the following options:

- (Recommended) OAuth 2.0 `clientId` and `clientSecret`

  Use the Workload Optimization Manager API to create and manage OAuth 2.0 client credentials. These credentials are more secure than the local account credentials created from the Workload Optimization Manager user interface.

  To create the OAuth 2.0 client credentials, perform the following steps:

  a. Log in to your Workload Optimization Manager instance.

  b. In your browser's address bar, change the URL to `https://{your_instance_address}/swagger/#/Authorization/createClient`.

  For example, change the URL to `https://my-instance.com/swagger/#/Authorization/createClient`.

c. Click **Try it out**.

d. In the **body** section, replace the sample request with the following request. This request has all the required parameters for generating OAuth 2.0 credentials for Kubeturbo.

```
{
  "clientName": "kubeturbo",
  "grantTypes": [
    "client_credentials"
  ],
  "clientAuthenticationMethods": [
    "client_secret_post"
  ],
  "scopes": [
    "role:PROBE_ADMIN"
  ],
  "tokenSettings": {
    "accessToken": {
      "ttlSeconds": 600
    }
  }
}
```

e. Click **Execute** and then scroll to the **Server response** section. If the credentials were generated successfully, a response with a code of `200` displays.

f. In the response, find and record the `clientID` and `clientSecret` credentials. These credentials cannot be retrieved after you close the API so it is important that you record them.

The credentials that you created for Kubeturbo must be converted to Base64.

In Linux, you can run the following command to convert each credential to Base64.

```
echo {credential} | base64
```

NOTE:
If your credential fails to convert to Base64, it might have invalid or unsupported characters and must be changed before it can be converted to Base64 successfully.

After the conversion, specify the Base64 `clientId` and `clientSecret` in the YAML file that you are updating, as shown in the following example.

```
data:
  clientid: {Base64_client_ID}
  clientsecret: {Base64_client_secret}
```

- (Not recommended) Local user account username and password (in Base64 in a secret)

NOTE:
Support for these credentials will be discontinued in a future release.

Comment the `clientId` and `clientSecret` parameters, and then specify the `username` and `password` in Base64.

```
data:
  username: {Base64_username}
  password: {Base64_password}
```

3. Deploy the YAML resource.
   a. Log in to your cluster through the command line.
   b. Deploy the YAML resource.

```
kubectl apply -f kubeturbo_namespace_turbo_credentials_secret.yaml
```

Deploying Kubeturbo

1. (Optional) Perform a dry run to test your deployment.

```
helm install --dry-run \
--debug kubeturbo kubeturbo/deploy/kubeturbo \
--namespace turbo \
--set serverMeta.turboServer={your_instance_address} \
--set image.tag={your_image_tag_version} \
--set targetConfig.targetName={your_cluster_name}
```

   Update the following parameters:

   - `--set serverMeta.turboServer={your_instance_address}`

     Specify the address of your Workload Optimization Manager instance, such as `https://10.1.1.1` or `https://myinstance.com`.

   - `--set image.tag={your_image_tag_version}`

     Specify the image tag version.

     To determine the image tag version to specify, open the Workload Optimization Manager user interface, click the Help icon in the navigation menu, and then record the version number that displays. This version number has a corresponding image tag version. For example, if your Workload Optimization Manager instance version is 3.13.0, the corresponding image tag version is 8.15.0. For the latest Workload Optimization Manager and image tag version mapping, see this reference *(on page 302)*.

   - `--set targetConfig.targetName={your_cluster_name}`

     Specify the cluster name that will display in the Workload Optimization Manager user interface. Spaces are not allowed.

   - If you want to enable move actions for pods with persistent volumes, add the following parameter.

```
--set args.failVolumePodMoves=false
```

   Be sure to resolve any errors before proceeding to the next step.

2. Deploy Kubeturbo with a specific role by choosing one of the following options. The role that you choose for Kubeturbo determines its level of access to your cluster.

   - Option 1: Default role

     By default, Kubeturbo deploys to your cluster with the `cluster-admin` role. This role has full control over every resource in the cluster.

     If you want to use the default role, run the following command. This command does not explicitly specify a role parameter, which means that the default role will be used.

```
helm install kubeturbo kubeturbo/deploy/kubeturbo \
--namespace turbo \
--set serverMeta.turboServer={your_instance_address} \
--set image.tag={your_image_tag_version} \
--set targetConfig.targetName={your_cluster_name}
```

     For details about the parameters that you need to configure, see the previous step.

   - Option 2: `turbo-cluster-admin` custom role

     This custom role specifies the minimum permissions that Kubeturbo needs to monitor your workloads and execute the actions that Workload Optimization Manager generated to optimize these workloads.

     If you want to use this role, specify the role parameter, as shown in the last line of the following command.

```
helm install kubeturbo kubeturbo/deploy/kubeturbo \
--namespace turbo \
--set serverMeta.turboServer={your_instance_address} \
--set image.tag={your_image_tag_version} \
--set targetConfig.targetName={your_cluster_name} \
--set roleName=turbo-cluster-admin
```

For details about the other parameters that you need to configure, see the previous step.

- Option 3: `turbo-cluster-reader` custom role

  This custom role is the least privileged role. It specifies the minimum permissions that Kubeturbo needs to monitor your workloads. Actions that Workload Optimization Manager generated to optimize these workloads can only be executed outside of Workload Optimization Manager (for example, in your cluster).

  If you want to use this role, specify the role parameter, as shown in the last line of the following command.

  ```
  helm install kubeturbo kubeturbo/deploy/kubeturbo \
  --namespace turbo \
  --set serverMeta.turboServer={your_instance_address} \
  --set image.tag={your_image_tag_version} \
  --set targetConfig.targetName={your_cluster_name} \
  --set roleName=turbo-cluster-reader
  ```

  For details about the other parameters that you need to configure, see the previous step.

3. (Optional) If you need to adjust or specify other parameters, see the last section in this topic, 'Kubeturbo custom resource values'.

The deployment starts after you run the command.

Validating the deployment

1. Review the output in Helm. The following example indicates that the deployment was successful.

   ```
   NAME: kubeturbo
   LAST DEPLOYED: Thu Aug 10 15:42:16 2023
   NAMESPACE: turbo
   STATUS: deployed
   REVISION: 1
   TEST_SUITE: None
   ```

2. Verify the status of the Kubeturbo pod.

   ```
   kubectl get pods -n turbo
   ```

   The following example result indicates that the pod was deployed and is currently running.

   ```
   NAME                    READY   STATUS    RESTARTS   AGE
   kubeturbo-asdf1234asd3  1/1     Running   0          37m
   ```

3. If you use OAuth 2.0 credentials for Kubeturbo, verify the that the credentials are configured correctly. In the Kubeturbo logs, the following log example indicates that the credentials are configured correctly.

   ```
   I0723 20:21:46.659559       1 tap_service.go:114] Secure credentials is provided, target
    Kubernetes-ocp-414 could be auto-registered if the server is running in secure mode and secure
    connection is established
   ```

4. Open the Workload Optimization Manager user interface and navigate to **Settings > Target Configuration**. If the deployment was successful, a new container platform target appears in the list.

If you do not see the target, there might be a deployment issue that you need to resolve.

- To start troubleshooting an issue, review the Kubeturbo logs. You can retrieve logs by running the following command.

  ```
  kubectl logs kubeturbo-{kubeturbo_pod_ID} -n turbo
  ```

  For example:

  ```
  kubectl logs kubeturbo-asdf1234asd3 -n turbo
  ```

- If you need further assistance, contact your Workload Optimization Manager representative.

Reference: Kubeturbo custom resource values

The following table describes the values that you can specify in the Kubeturbo custom resource to configure your deployment.

Refer to the Kubeturbo CRD for the schema.

Parameter	Default value	Changes to default value	Parameter type
`args.cleanupSccImpersonationResources`	`true`	Optional	Clean up the resources for SCC impersonation by default. For details about SCC impersonation, see this topic *(on page 609)*.
`args.discovery-interval-sec`	None	Optional	Number in seconds
`args.discovery-sample-interval`	None	Optional	Number in seconds
`args.discovery-samples`	None	Optional	Number
`args.discovery-timeout-sec`	None	Optional	Number in seconds
`args.discovery-workers`	None	Optional	Number
`args.garbage-collection-interval`	None	Optional	Number in seconds
`args.kubelethttps`	`true`	Optional, change to `false` if using Kubernetes 1.10 or older	Boolean
`args.kubeletport`	`10250`	Optional, change to `10255` if using Kubernetes 1.10 or older	Number
`args.logginglevel`	`2`	Optional	Number. A high value increases logging.
`args.pre16k8sVersion`	`false`	Optional	If Kubernetes version is older than 1.6, then add another `arg` for move/resize actions.
`args.sccsupport`	None	Optional	Enabled by default in clusters that use SCC.

Parameter	Default value	Changes to default value	Parameter type
`args.stitchuuid`	`true`	Optional, change to `false` if IaaS is VMM or Hyper-V	Boolean
`daemonPodDetectors.daemonPodNamespaces1` and `daemonPodNamespaces2`	`daemonSet` kinds are allowed for node suspension by default. Adding this parameter changes the default.	Optional but required to identify pods in the namespace to be ignored for cluster consolidation.	Regular expression used, values are in quotes and are comma-separated, such as `"kube-system","kube-service-catalog","openshift-.*"`.
`daemonPodDetectors.daemonPodNamePatterns`	`daemonSet` kinds are allowed for node suspension by default. Adding this parameter changes the default.	Optional but required to identify pods matching this pattern to be ignored for cluster consolidation.	Regular expression used `.*ignorepod.*`
`HANodeConfig.nodeRoles`	Any value for label key value `node-role.kubernetes.io/` for master and others	Optional. This is used to automate policies that keep nodes of the same role limited to one instance per ESX host or availability zone.	Values in `values.yaml` or `cr.yaml` use escapes or quotes, or are comma-separated. Master nodes are by default. Other roles populated by `nodeRoles:""foo","bar""`
`image.busyboxRepository`	`icr.io/cpopen/turbonomic/cpufreqgetter:1.0`	Optional	Full path to the repository, image, and tag.
`image.cpufreqgetterRepository`	`icr.io/cpopen/turbonomic/cpufreqgetter`	Optional	Repository used to get node `cpufrequency`.
`image.imagePullSecret`	None	Optional	Defines the secret used to authenticate to the container image registry
`image.pullPolicy`	`IfNotPresent`	Optional	
`image.repository`	`icr.io/cpopen/turbonomic/kubeturbo`	Optional	Path to the repository. Must be used with `image.tag`.
`image.tag`	Depends on product version	Optional	Kubeturbo tag. Must be used with `image.repository`.
`kubeturboPodScheduling.affinity`	None	Optional	Specify `affinity` pod scheduling constraint in the cluster. For examples, see the Kubernetes documentation
`kubeturboPodScheduling.nodeSelector`	None	Optional	Specify `nodeSelector` pod scheduling constraint in the cluster. For examples, see the Kubernetes documentation.
`kubeturboPodScheduling.tolerations`	None	Optional	Specify `tolerations` pod scheduling constraint in the cluster. For examples, see

Parameter	Default value	Changes to default value	Parameter type
			the Kubernetes documentation.
masterNodeDetectors.nodeLabels	Any value for label key value node-role.kubernetes.io/master	Deprecated. Previously used to avoid suspending masters identified by node label key value pair. The value ignored if there is no match.	Regular expression used, specify the key as masterNodeDetectors.nodeLabelsKey (such as node-role.kubernetes.io/master) and the value as masterNodeDetectors.nodeLabelsValue (such as .*).
masterNodeDetectors.nodeNamePatterns	Node name includes .*master.*	Deprecated. Previously used to avoid suspending masters identified by node name. The value is ignored if there is no match.	String, regular expression used. For example: .*master.*
restAPIConfig.opsManagerPassword	None	Optional	Administrator password.
restAPIConfig.opsManagerUserName	None	Optional	Local or Active Directory user with site administrator role.
restAPIConfig.turbonomicCredentialsSecretName	turbonomic-credentials	Required only if using secret and not taking the default secret name	Secret that contains the Workload Optimization Manager site administrator username and password in Base64.
roleBinding	turbo-all-binding	Optional	Specify the name of clusterrolebinding
roleName	cluster-admin	Optional	Specify custom turbo-cluster-reader or turbo-cluster-admin role instead of the default cluster-admin role.
serverMeta.proxy	None	Optional	Format of http://username:password@proxyserver:proxyport or http://proxyserver:proxyport
serverMeta.turboServer	None	Required	HTTPS URL to log in to Workload Optimization Manager
serverMeta.version	None	Optional	Number x.y
serviceAccountName	turbo-user	Optional	Specify the name of the service account.
targetConfig.targetName	{Your_cluster}	Optional but required for multiple clusters	String that indicates how you want to identify your cluster.

Configuring a private repository (Helm)

Workload Optimization Manager gathers information from your clusters through the Kubeturbo container images that are available through a public repository in IBM Container Registry (`icr.io`).

The following list describes the different Kubeturbo container images.

- Container image for Kubeturbo

  This is the primary image and is pulled from `icr.io/cpopen/turbonomic/kubeturbo:<version>`.

- Container image for the CPU frequency getter job

  The `cpufreqgetter` image (also known as BusyBox) runs a job on every node to collect CPU speed. This image is pulled from `icr.io/cpopen/turbonomic/cpufreqgetter`.

  For more information about the parameters associated with this job, see this topic *(on page 304)*.

If you are using a private repository, be aware that IBM Container Registry provides multi-architecture container images. Your private repository needs to support loading these images. For example, multi-architecture container images are supported in Artifactory version 7.2+, but not in Artifactory version 6. Optionally, pull only the architecture version that you need. Specify the `--platform` parameter with `docker pull` and use `docker inspect` to confirm the architecture.

```
docker pull --platform linux/arm64 icr.io/cpopen/turbonomic/kubeturbo:8.15.0
docker inspect 49a61c21c3a1 | grep architecture
                "architecture": "aarch64",
```

To use a private repository, define additional parameters in the `values.yaml` resource, as shown in the following example (see a sample YAML here).

```
image:
  repository: {your_kubeturbo_repository}/icr.io/cpopen/turbonomic/kubeturbo
  tag: 8.15.0
  cpufreqgetterRepository: {your_frequency_getter_repository}/icr.io/cpopen/turbonomic/cpufreqgetter
#  busyboxRepository: busybox
#  imagePullSecret: ""
#  cpufreqgetterRepository: icr.io/cpopen/turbonomic/cpufreqgetter
```

Update the sample command with the following parameters:

- `repository: {your_kubeturbo_repository}/icr.io/cpopen/turbonomic/kubeturbo`

  Specify your private Kubeturbo repository, such as `myrepo.com`.

- `cpufreqgetterRepository: {your_frequency_getter_repository}/icr.io/cpopen/turbonomic/cpufreqgetter`

  Specify your private CPU frequency getter repository, such as `mygetterrepo.com`.

Deploying Kubeturbo through an operator

You can deploy Kubeturbo by using a Go-based operator and configuring a custom resource.

> If you are running Red Hat OpenShift 4.1x on Linux (x86, IBM Power, or IBM LinuxONE platforms), the best practice is to deploy Kubeturbo using OperatorHub. For OperatorHub deployment instructions, see this topic *(on page 277)*.

Deployment requirements

Before deploying Kubeturbo, be sure to:

- Review the general requirements *(on page 234)*.
- Set up and record the credentials *(on page 236)* for your Workload Optimization Manager instance. You will specify these credentials when you deploy Kubeturbo.

Task overview

To deploy Kubeturbo, perform the following tasks:

1. Deploy the Kubeturbo Custom Resource Definition (CRD).
2. Download the YAML resource for the Kubeturbo role.
3. Update the YAML resource with the minimum required parameters.
4. Deploy the YAML resource to your cluster.

   After you perform this task, Kubeturbo deploys to your cluster and the Kubeturbo pod starts.
5. Download and apply the Kubeturbo operator YAML bundle.
6. Validate the deployment.

NOTE:
Workload Optimization Manager gathers information from your clusters through the Kubeturbo container images. By default, these images are pulled from IBM Container Registry (`icr.io`). If you prefer to pull these images from your private repository, configure that repository before deploying Kubeturbo. For more information about private repositories, see this topic *(on page 275)*.

Deploying the Kubeturbo Custom Resource Definition (CRD)

The CRD ensures the validity of your Kubeturbo deployment.

1. Deploy the Kubeturbo CRD.

```
kubectl create -f
 https://raw.githubusercontent.com/turbonomic/kubeturbo-deploy/refs/heads/staging/config/crd/bas
es/charts.helm.k8s.io_kubeturbos.yaml
```

Downloading the YAML resource for the Kubeturbo role

The role that you choose for Kubeturbo determines its level of access to your cluster.

Choose from the following roles and then download the corresponding YAML resource.

- Option 1: Default role

  By default, Kubeturbo deploys to your cluster with the `cluster-admin` role. This role has full control over every resource in the cluster.

  To download the YAML resource for this role, run the following command:

```
curl -O
 https://raw.githubusercontent.com/turbonomic/kubeturbo-deploy/master/deploy/kubeturbo_yamls/turbo_
kubeturbo_operator_full.yaml
```

- Option 2: `turbo-cluster-admin` custom role

  This custom role specifies the minimum permissions that Kubeturbo needs to monitor your workloads and execute the actions that Workload Optimization Manager generated to optimize these workloads.

  To download the YAML resource for this role, run the following command:

```
curl -O
 https://raw.githubusercontent.com/turbonomic/kubeturbo-deploy/master/deploy/kubeturbo_yamls/turbo_
kubeturbo_operator_least_admin_full.yaml
```

- Option 3: `turbo-cluster-reader` custom role

  This custom role is the least privileged role. It specifies the minimum permissions that Kubeturbo needs to monitor your workloads. Actions that Workload Optimization Manager generated to optimize these workloads can only be executed outside of Workload Optimization Manager (for example, in your cluster).

To download the YAML resource for this role, run the following command:

```
curl -O
 https://raw.githubusercontent.com/turbonomic/kubeturbo-deploy/master/deploy/kubeturbo_yamls/turbo_
kubeturbo_operator_reader_full.yaml
```

Updating the YAML resource with the minimum required parameters

Update the YAML resource that you downloaded. For example, you can update the YAML resource in VS Code or `vi`. Be sure to set up and record the credentials *(on page 236)* for your Workload Optimization Manager instance before performing this task.

IMPORTANT:
The YAML resource that you downloaded specifies the following default values:

- Namespace – `turbonomic`
- Secret – `turbonomic-credentials`

It is recommended that you leave these values unchanged because the same values are specified in other Kubeturbo resources. Changing the values in the YAML resource but not in the other Kubeturbo resources could result in deployment issues.

1. Update the `turboServer` parameter.

   ```
   serverMeta:
     turboServer: "{your_instance_address}"
   ```

   Specify the address of your Workload Optimization Manager instance, such as `https://10.1.1.1` or `https://myinstance.com`.
2. Update the `targetName` parameter.

   ```
     targetName: "{your_cluster_name}"
   ```

   Specify the cluster name that will display in the Workload Optimization Manager user interface. Spaces are not allowed.
3. Update the `image` parameter.

   ```
           image: icr.io/cpopen/kubeturbo-operator:{your_image_tag_version}
   ```

   Specify the image tag version.

   To determine the image tag version to specify, open the Workload Optimization Manager user interface, click the Help icon in the navigation menu, and then record the version number that displays. This version number has a corresponding image tag version. For example, if your Workload Optimization Manager instance version is 3.13.0, the corresponding image tag version is 8.15.0. For the latest Workload Optimization Manager and image tag version mapping, see this reference *(on page 302)*.
4. Update the Kubeturbo credentials.
   - Choose from the following options:
     - (Recommended) OAuth 2.0 `clientId` and `clientSecret`

       Use the Workload Optimization Manager API to create and manage OAuth 2.0 client credentials. These credentials are more secure than the local account credentials created from the Workload Optimization Manager user interface.

       To create the OAuth 2.0 client credentials, perform the following steps:

       a. Log in to your Workload Optimization Manager instance.

       b. In your browser's address bar, change the URL to `https://{your_instance_address}/swagger/#/Authorization/createClient`.

          For example, change the URL to `https://my-instance.com/swagger/#/Authorization/createClient`.

c. Click **Try it out**.

d. In the **body** section, replace the sample request with the following request. This request has all the required parameters for generating OAuth 2.0 credentials for Kubeturbo.

```
{
  "clientName": "kubeturbo",
  "grantTypes": [
    "client_credentials"
  ],
  "clientAuthenticationMethods": [
    "client_secret_post"
  ],
  "scopes": [
    "role:PROBE_ADMIN"
  ],
  "tokenSettings": {
    "accessToken": {
      "ttlSeconds": 600
    }
  }
}
```

e. Click **Execute** and then scroll to the **Server response** section. If the credentials were generated successfully, a response with a code of `200` displays.

f. In the response, find and record the `clientID` and `clientSecret` credentials. These credentials cannot be retrieved after you close the API so it is important that you record them.

The credentials that you created for Kubeturbo must be converted to Base64.

In Linux, you can run the following command to convert each credential to Base64.

```
echo {credential} | base64
```

NOTE:
If your credential fails to convert to Base64, it might have invalid or unsupported characters and must be changed before it can be converted to Base64 successfully.

After the conversion, specify the Base64 `clientId` and `clientSecret` in the YAML file that you are updating, as shown in the following example.

```
data:
  clientid: {Base64_client_ID}
  clientsecret: {Base64_client_secret}
```

– (Not recommended) Local user account username and password (in Base64 in a secret)

NOTE:
Support for these credentials will be discontinued in a future release.

Comment the `clientId` and `clientSecret` parameters, and then specify the `username` and `password` in Base64.

```
data:
  username: {Base64_username}
  password: {Base64_password}
```

5. If you want to enable move actions for pods with persistent volumes, uncomment the `args` section and then add a new parameter, as shown in the following example.

```
  args:
    failVolumePodMoves: 'false'
```

6. (Optional) If you need to adjust or specify other parameters, see the last section in this topic, 'Kubeturbo custom resource values'.

Deploying the YAML resource to your cluster

Deploy the YAML resource.

```
kubectl apply -f {YAML_name}
```

Replace *{YAML_name}* with the YAML resource that you downloaded and updated. Choose from the following:

- `turbo_kubeturbo_operator_full.yaml`
- `turbo_kubeturbo_operator_reader_full.yaml`
- `turbo_kubeturbo_operator_least_admin_full.yaml`

After you run the command, Kubeturbo deploys to your cluster and the Kubeturbo pod starts.

Downloading and applying the Kubeturbo operator YAML bundle

Kubeturbo uses a Go-based operator to enable automatic changes to the Kubeturbo deployment. This operator removes the need to update or delete resources manually. The YAML bundle specifies the parameters needed by the Go-based operator.

1. Download and apply the Kubeturbo operator YAML bundle.

   ```
   curl
    https://raw.githubusercontent.com/turbonomic/kubeturbo-deploy/refs/heads/staging/deploy/kubeturbo_
   operator_yamls/operator-bundle.yaml | sed "s/: turbonomic$/: <target_namespace>/g" | kubectl apply
    -f -
   ```

   Replace *<target_namespace>* with the namespace where the Kubeturbo operator is deployed. For example, change the namespace to `turbo`.

Validating the deployment

1. Verify the status of the Kubeturbo pod.

   ```
   kubectl get pods -n turbo
   ```

   The following example result indicates that the pod was deployed and is currently running.

   ```
   NAME                    READY   STATUS    RESTARTS   AGE
   kubeturbo-asdf1234asd3  1/1     Running   0          37m
   ```

2. If you use OAuth 2.0 credentials for Kubeturbo, verify the that the credentials are configured correctly. In the Kubeturbo logs, the following log example indicates that the credentials are configured correctly.

   ```
   I0723 20:21:46.659559       1 tap_service.go:114] Secure credentials is provided, target
    Kubernetes-ocp-414 could be auto-registered if the server is running in secure mode and secure
    connection is established
   ```

3. Open the Workload Optimization Manager user interface and navigate to **Settings > Target Configuration**. If the deployment was successful, a new container platform target appears in the list.

   If you do not see the target, there might be a deployment issue that you need to resolve.

   - To start troubleshooting an issue, review the Kubeturbo logs. You can retrieve logs by running the following command.

     ```
     kubectl logs kubeturbo-{kubeturbo_pod_ID} -n turbo
     ```

For example:

```
kubectl logs kubeturbo-asdf1234asd3 -n turbo
```

- If you need further assistance, contact your Workload Optimization Manager representative.

Reference: Kubeturbo custom resource values

The following table describes the values that you can specify in the Kubeturbo custom resource to configure your deployment.

Refer to the Kubeturbo CRD for the schema.

Parameter	Default value	Changes to default value	Parameter type
`args.cleanupSccImpersonationResources`	`true`	Optional	Clean up the resources for SCC impersonation by default. For details about SCC impersonation, see this topic *(on page 609)*.
`args.discovery-interval-sec`	None	Optional	Number in seconds
`args.discovery-sample-interval`	None	Optional	Number in seconds
`args.discovery-samples`	None	Optional	Number
`args.discovery-timeout-sec`	None	Optional	Number in seconds
`args.discovery-workers`	None	Optional	Number
`args.garbage-collection-interval`	None	Optional	Number in seconds
`args.kubelethttps`	`true`	Optional, change to `false` if using Kubernetes 1.10 or older	Boolean
`args.kubeletport`	`10250`	Optional, change to `10255` if using Kubernetes 1.10 or older	Number
`args.logginglevel`	`2`	Optional	Number. A high value increases logging.
`args.pre16k8sVersion`	`false`	Optional	If Kubernetes version is older than 1.6, then add another `arg` for move/resize actions.
`args.sccsupport`	None	Optional	Enabled by default in clusters that use SCC.
`args.stitchuuid`	`true`	Optional, change to `false` if IaaS is VMM or Hyper-V	Boolean
`daemonPodDetectors.daemonPodNamespaces1` and `daemonPodNamespaces2`	`daemonSet` kinds are allowed for node suspension by default. Adding this parameter changes the default.	Optional but required to identify pods in the namespace to be ignored for cluster consolidation.	Regular expression used, values are in quotes and are comma-separated, such as `"kube-system","kube-ser-`

Parameter	Default value	Changes to default value	Parameter type
			vice-catalog","open-shift-.*".
daemonPodDetectors.daemonPodNamePatterns	daemonSet kinds are allowed for node suspension by default. Adding this parameter changes the default.	Optional but required to identify pods matching this pattern to be ignored for cluster consolidation.	Regular expression used .*ignorepod.*
HANodeConfig.nodeRoles	Any value for label key value node-role.kubernetes.io/ for master and others	Optional. This is used to automate policies that keep nodes of the same role limited to one instance per ESX host or availability zone.	Values in values.yaml or cr.yaml use escapes or quotes, or are comma-separated. Master nodes are by default. Other roles populated by nodeRoles:""foo","bar""
image.busyboxRepository	icr.io/cpopen/turbonomic/cpufreqgetter:1.0	Optional	Full path to the repository, image, and tag.
image.cpufreqgetterRepository	icr.io/cpopen/turbonomic/cpufreqgetter	Optional	Repository used to get node cpufrequency.
image.imagePullSecret	None	Optional	Defines the secret used to authenticate to the container image registry
image.pullPolicy	IfNotPresent	Optional	
image.repository	icr.io/cpopen/turbonomic/kubeturbo	Optional	Path to the repository. Must be used with image.tag.
image.tag	Depends on product version	Optional	Kubeturbo tag. Must be used with image.repository.
kubeturboPodScheduling.affinity	None	Optional	Specify affinity pod scheduling constraint in the cluster. For examples, see the Kubernetes documentation
kubeturboPodScheduling.nodeSelector	None	Optional	Specify nodeSelector pod scheduling constraint in the cluster. For examples, see the Kubernetes documentation.
kubeturboPodScheduling.tolerations	None	Optional	Specify tolerations pod scheduling constraint in the cluster. For examples, see the Kubernetes documentation.
masterNodeDetectors.nodeLabels	Any value for label key value node-role.kubernetes.io/master	Deprecated. Previously used to avoid suspending masters identified by node label key value pair. The value ignored if there is no match.	Regular expression used, specify the key as masterNodeDetectors.nodeLabelsKey (such as node-role.kubernetes.io/master) and

Parameter	Default value	Changes to default value	Parameter type
			the value as `masterNodeDetectors.nodeLabelsValue` (such as `.*`).
`masterNodeDetectors.nodeNamePatterns`	Node name includes `.*master.*`	Deprecated. Previously used to avoid suspending masters identified by node name. The value is ignored if there is no match.	String, regular expression used. For example: `.*master.*`
`restAPIConfig.opsManagerPassword`	None	Optional	Administrator password.
`restAPIConfig.opsManagerUserName`	None	Optional	Local or Active Directory user with site administrator role.
`restAPIConfig.turbonomicCredentialsSecretName`	`turbonomic-credentials`	Required only if using secret and not taking the default secret name	Secret that contains the Workload Optimization Manager site administrator username and password in Base64.
`roleBinding`	`turbo-all-binding`	Optional	Specify the name of `clusterrolebinding`
`roleName`	`cluster-admin`	Optional	Specify custom `turbo-cluster-reader` or `turbo-cluster-admin` role instead of the default `cluster-admin` role.
`serverMeta.proxy`	None	Optional	Format of `http://username:password@proxyserver:proxyport` or `http://proxyserver:proxyport`
`serverMeta.turboServer`	None	Required	HTTPS URL to log in to Workload Optimization Manager
`serverMeta.version`	None	Optional	Number `x.y`
`serviceAccountName`	`turbo-user`	Optional	Specify the name of the service account.
`targetConfig.targetName`	`{Your_cluster}`	Optional but required for multiple clusters	String that indicates how you want to identify your cluster.

Configuring a private repository (operator)

Workload Optimization Manager gathers information from your clusters through the Kubeturbo container images that are available through a public repository in IBM Container Registry (`icr.io`).

The following list describes the different Kubeturbo container images.

- Container image for Kubeturbo

  This is the primary image and is pulled from `icr.io/cpopen/turbonomic/kubeturbo:<version>`.
- Container image for the CPU frequency getter job

The `cpufreqgetter` image (also known as BusyBox) runs a job on every node to collect CPU speed. This image is pulled from `icr.io/cpopen/turbonomic/cpufreqgetter`.

For more information about the parameters associated with this job, see this topic *(on page 304)*.

- (Operator method only) Container image for `kubeturbo-operator`

  This image is pulled for `icr.io/cpopen/kubeturbo-operator:<version>`

If you are using a private repository, be aware that IBM Container Registry provides multi-architecture container images. Your private repository needs to support loading these images. For example, multi-architecture container images are supported in Artifactory version 7.2+, but not in Artifactory version 6. Optionally, pull only the architecture version that you need. Specify the `--platform` parameter with `docker pull` and use `docker inspect` to confirm the architecture.

```
docker pull --platform linux/arm64 icr.io/cpopen/turbonomic/kubeturbo:8.15.0
docker inspect 49a61c21c3a1 | grep architecture
               "architecture": "aarch64",
```

To use a private repository, you need to define additional parameters.

1. Define the parameters in the operator instance or custom resource you have configured (default is `kubeturbo-release`), as shown in the following example (see a sample YAML here):

   ```
   spec:
     image:
       # Supply your private repo and specific product version here
       # With v8.7.5 and newer, kubeturbo is available via IBM Container Registry
       repository: {your_kubeturbo_repository}/icr.io/cpopen/turbonomic/kubeturbo
       tag: 8.15.0
       # Specify your cpu frequency job container image.
       cpufreqgetterRepository: {your_frequency_getter_repository}/cpopen/turbonomic/cpufreqgetter
       # Note cpufreqgetter will use the same pull secret as set for the kubeturbo container image
       #imagePullSecret: secretName
   ```

   ```
     #Rest of CR will be retained
     ...
   ```

   Update the sample command with the following parameters:

   - `repository: {your_kubeturbo_repository}/icr.io/cpopen/turbonomic/kubeturbo`

     Specify your private Kubeturbo repository, such as `myrepo.com`.

   - `cpufreqgetterRepository: {your_frequency_getter_repository}/icr.io/cpopen/turbonomic/cpufreqgetter`

     Specify your private CPU frequency getter repository, such as `mygetterrepo.com`.

2. Define your Kubeturbo operator image path in the operator's `deployment` resource (see a sample YAML here).

   ```
   kind: Deployment
   metadata:
     name: kubeturbo-operator
     namespace: turbo
   spec:
     replicas: 1
     selector:
        matchLabels:
          app.kubernetes.io/name: kubeturbo-operator
     template:
       spec:
         containers:
         - name: kubeturbo-operator
           image: YOUR-REPO.COM/icr.io/cpopen/turbonomic/kubeturbo-operator:8.15.0
   ```

Connecting to Red Hat OpenShift clusters

Red Hat OpenShift is a platform that manages containerized applications. Workload Optimization Manager discovers these containerized applications through the **Kubeturbo** agent that you deploy to each Red Hat OpenShift cluster.

After the deployment, Kubeturbo automatically adds itself as a Red Hat OpenShift target in the Workload Optimization Manager user interface, in **Settings > Target Configuration**. You do not need to configure any settings in the user interface.

Kubeturbo deployment requirements

Kubeturbo supports Red Hat supported GA versions of Red Hat OpenShift 4.1x, including (but not limited to) Red Hat OpenShift Kubernetes Service (ROKS), Red Hat OpenShift on AWS (ROSA), Red Hat OpenShift on Azure (ARO), and Red Hat OpenShift on Google.

Before deploying Kubeturbo, be sure to:

- Review the general requirements *(on page 234)*.
- Set up and record the credentials *(on page 236)* for your Workload Optimization Manager instance. You will specify these credentials when you deploy Kubeturbo.

Kubeturbo deployment method

Deploy Kubeturbo through the OperatorHub in Red Hat OpenShift. For deployment instructions, see the following topic:

- Deploying Kubeturbo Through OperatorHub *(on page 277)*

Monitoring and optimization of Red Hat OpenShift clusters

After connecting to your Red Hat OpenShift clusters, Workload Optimization Manager monitors and optimizes the resources that it discovered. See the following topics for more information:

- Container Platform Monitored Resources *(on page 286)*
- Container Platform Actions *(on page 289)*

Deploying Kubeturbo through OperatorHub

If you are running Red Hat OpenShift on Linux (x86, IBM Power, or IBM LinuxONE platforms), you can deploy Kubeturbo by using the Go-based certified Kubeturbo operator in Red Hat OpenShift OperatorHub and configuring a custom resource that creates a single instance of Kubeturbo per cluster. You do not need to manually create the service account, role, or Cluster Role Binding (CRB). The operator automatically creates these items. This is the main advantage of using the operator to deploy Kubeturbo.

> **NOTE:**
> If you are running Red Hat OpenShift in a disconnected network and want to deploy any certified operator, first follow the steps outlined in the Red Hat OpenShift documentation based on the Red Hat OpenShift version that you are running.

Deployment requirements

Before deploying Kubeturbo, be sure to:

- Review the general requirements *(on page 234)*.
- Set up and record the credentials *(on page 236)* for your Workload Optimization Manager instance. You will specify these credentials when you deploy Kubeturbo.

Task overview

To deploy Kubeturbo in Red Hat OpenShift OperatorHub UI, perform the following tasks:

1. Create a project in the Red Hat OpenShift cluster where you will deploy Kubeturbo.
2. Deploy the Kubeturbo operator from OperatorHub.
3. Create an instance using the Form or YAML method.
4. Validate the deployment.

NOTE:
Workload Optimization Manager gathers information from your clusters through the Kubeturbo container images. By default, these images are pulled from IBM Container Registry (`icr.io`). If you prefer to pull these images from your private repository, configure that repository before deploying Kubeturbo. For more information about private repositories, see this topic *(on page 285)*.

Creating a project in the Red Hat OpenShift cluster

Use cluster administrator privileges to create a project in the Red Hat OpenShift cluster where you will deploy Kubeturbo.

1. Log in to your Red Hat OpenShift console and navigate to **Home > Projects**.
2. Click **Create Project**.
3. Specify the project name. Optionally, specify a display name and description.
4. Click **Create**.

Setting up Kubeturbo credentials

Kubeturbo can use the following credentials to connect to your Workload Optimization Manager instance.

- OAuth 2.0 `clientId` and `clientSecret`

  For any instance of Workload Optimization Manager (for example, OVA or Red Hat OpenShift), you can use OAuth 2.0 client credentials.

  Use the Workload Optimization Manager API to create and manage OAuth 2.0 client credentials. These credentials are more secure than the local account credentials created from the Workload Optimization Manager user interface.

  To create the OAuth 2.0 client credentials, perform the following steps:

  1. Log in to your Workload Optimization Manager instance.
  2. In your browser's address bar, change the URL to `https://{your_instance_address}/swagger/#/Authorization/createClient`.

     For example, change the URL to `https://my-instance.com/swagger/#/Authorization/createClient`.
  3. Click **Try it out**.
  4. In the **body** section, replace the sample request with the following request. This request has all the required parameters for generating OAuth 2.0 credentials for Kubeturbo.

```
{
  "clientName": "kubeturbo",
  "grantTypes": [
    "client_credentials"
  ],
  "clientAuthenticationMethods": [
    "client_secret_post"
  ],
  "scopes": [
    "role:PROBE_ADMIN"
  ],
  "tokenSettings": {
```

```
    "accessToken": {
      "ttlSeconds": 600
    }
  }
}
```

5. Click **Execute** and then scroll to the **Server response** section. If the credentials were generated successfully, a response with a code of `200` displays.
6. In the response, find and record the `clientID` and `clientSecret` credentials. These credentials cannot be retrieved after you close the API so it is important that you record them.

After creating the OAuth 2.0 client credentials, perform the following steps.

1. In the project that you created in the previous task, navigate to **Workload > Secrets**.
2. Click **Create** and select **Key/value secret**.
3. Configure the following settings:
   - Secret name

     Specify the default name `turbonomic-credentials`.

     IMPORTANT: It is recommended that you use the default name because the same name is specified in other Kubeturbo resources. Using a different name could result in deployment issues.

   - Key/value pair for `clientID`
     - Key

       Specify **clientid**.
     - Value

       Specify the `clientID`. Red Hat OpenShift automatically converts the value into Base64 after you create the secret.
   - Key/value pair for `clientSecret`
     - Key

       Specify **clientsecret**.
     - Value

       Specify the `clientSecret`. Red Hat OpenShift automatically converts the value into Base64 after you create the secret.
4. Click **Create**.

Deploying the Kubeturbo operator from OperatorHub

1. Navigate to **Operators > OperatorHub**.
2. Search for Kubeturbo Operator in the search bar.
3. Select the **Certified** edition of Kubeturbo Operator and confirm the image version that you want to use.

   Kubeturbo uses a Go-based operator to enable automatic changes to the Kubeturbo deployment. This operator removes the need to update or delete resources manually.

   You can change the Kubeturbo image version after deployment. For more information, see this topic *(on page 297)*.

   NOTE:
   Do not use the **Community** or **Marketplace** edition.

4. Click **Install**. Basic install is the only option available for the operator.

5. Configure the following settings:
   - Installation mode

     Choose **A specific namespace on the cluster**.
   - Installed Namespace

     Choose the project that you created for Kubeturbo in the previous task.
   - Approval strategy

     Choose **Automatic** to update Kubeturbo automatically when a new Workload Optimization Manager version is released. Otherwise, choose **Manual**.

     NOTE:
     If you choose **Manual**, be sure to update the operator when prompted by OperatorHub.

   NOTE:
   The stable channel is the generally available product. The beta channel is a pre-release candidate and should not be used unless instructed.

   The operator installs.

   NOTE:
   Wait until the installation completes to see the following navigation tabs - Details, YAML, Subscription, Events, Kubeturbo Operator. You need to click the **Details** tab to create an instance (see the next section).

6. After the installation, click **View Operator** to create an instance. This instance is the Kubeturbo agent that will monitor and manage workloads in the given cluster.

Creating an instance

Be sure to set up and record the credentials *(on page 236)* for your Workload Optimization Manager instance before performing this task.

1. In the Kubeturbo Operator page, open the **Details** tab and click **Create instance**.
2. Select **Form view** or **YAML view** and then specify the minimum required parameters described in the next steps.

   Use **YAML view** for full control of editing/updating the custom resource.

3. Specify the address of your Workload Optimization Manager instance, such as `https://10.1.1.1` or `https://myinstance.com`.

YAML view	Form view
`serverMeta:` `  turboServer: {your_instance_address}`	Expand the **serverMeta** section and update the **turboServer** field.

4. Specify the cluster name that will display in the Workload Optimization Manager user interface. Spaces are not allowed.

YAML view	Form view
`targetConfig:` `  targetName: {your_cluster_name}`	Expand the **targetConfig** section and update the **targetName** field.

The following example shows a YAML resource with the minimum required parameters for Kubeturbo.

```
spec:
  restAPIConfig:
  serverMeta:
    turboServer: https://my-instance.com
  targetConfig:
    targetName: my_cluster_name
```

5. (Optional) Specify a custom role for Kubeturbo.

   By default, Kubeturbo deploys to your cluster with the `cluster-admin` role. This role has full control over every resource in the cluster.

   If you prefer a custom role, choose from the following roles:

   - `turbo-cluster-admin` role

     This custom role specifies the minimum permissions that Kubeturbo needs to monitor your workloads and execute the actions that Workload Optimization Manager generated to optimize these workloads.

YAML view	Form view
`spec:` `  roleName: turbo-cluster-admin`	In the **roleName** field, specify `turbo-cluster-admin`.

     To enable move actions for pods with persistent volumes, add the following parameter.

YAML view	Form view

`spec:`
`  args:`
`  failVolumePodMoves: 'false'`
	Expand the **args** section and specify **false** in the **failVolumePodMoves** field.

   - `turbo-cluster-reader` role

     This custom role is the least privileged role. It specifies the minimum permissions that Kubeturbo needs to monitor your workloads. Actions that Workload Optimization Manager generated to optimize these workloads can only be executed outside of Workload Optimization Manager (for example, in your cluster).

YAML view	Form view
`spec:` `  roleName: turbo-cluster-reader`	In the **roleName** field, specify `turbo-cluster-reader`.

6. (Optional) Enable the stitching of Red Hat OpenShift Kubernetes Service (ROKS) deployed in Azure through IBM Cloud Satellite.

YAML view	Form view

`spec:`
`  args:`
`  satelliteLocationProvider: azure`
	Expand the **args** section and specify **Azure** in the **satelliteLocationProvider** field.

7. (Optional) If you need to adjust or specify other parameters, see the last section in this topic, 'Kubeturbo custom resource values'.
8. Click **Create**.

Validating the deployment

1. Verify that there are two deployments and two running pods in the namespace. One is the operator and the other is Kubeturbo (release).
2. Open the Workload Optimization Manager user interface and navigate to **Settings > Target Configuration**. If the deployment was successful, a new container platform target appears in the list.

   If you do not see the target, there might be a deployment issue that you need to resolve.

   - To start troubleshooting an issue, review the Kubeturbo logs. You can retrieve logs by running the following command.

     ```
     kubectl logs kubeturbo-{kubeturbo_pod_ID} -n turbo
     ```

     For example:

     ```
     kubectl logs kubeturbo-asdf1234asd3 -n turbo
     ```

   - If you need further assistance, contact your Workload Optimization Manager representative.
3. Review the logs for the `kubeturbo-release` pod to verify that Kubeturbo successfully connected and registered with Workload Optimization Manager, and that a full discovery has occurred.

   To retrieve logs, perform any of the following steps:

   - Run the following command.

     ```
     oc logs kubeturbo-release{pod_ID} -n turbo
     ```

     For example:

     ```
     oc logs kubeturbo-release-12aga3jasd -n turbo
     ```

   - Select the `kubeturbo-release` pod and then click the **Logs** tab.

Reference: Kubeturbo custom resource values

The following table describes the values that you can specify in the Kubeturbo custom resource to configure your deployment.

Refer to the Kubeturbo CRD for the schema.

Parameter	Default value	Changes to default value	Parameter type
`args.cleanupSccImpersonationResources`	`true`	Optional	Clean up the resources for SCC impersonation by default. For details about SCC impersonation, see this topic *(on page 609)*.
`args.discovery-interval-sec`	None	Optional	Number in seconds
`args.discovery-sample-interval`	None	Optional	Number in seconds
`args.discovery-samples`	None	Optional	Number
`args.discovery-timeout-sec`	None	Optional	Number in seconds
`args.discovery-workers`	None	Optional	Number
`args.garbage-collection-interval`	None	Optional	Number in seconds

Parameter	Default value	Changes to default value	Parameter type
`args.kubelethttps`	`true`	Optional, change to `false` if using Kubernetes 1.10 or older	Boolean
`args.kubeletport`	`10250`	Optional, change to `10255` if using Kubernetes 1.10 or older	Number
`args.logginglevel`	`2`	Optional	Number. A high value increases logging.
`args.pre16k8sVersion`	`false`	Optional	If Kubernetes version is older than 1.6, then add another `arg` for move/resize actions.
`args.sccsupport`	None	Optional	Enabled by default in clusters that use SCC.
`args.stitchuuid`	`true`	Optional, change to `false` if IaaS is VMM or Hyper-V	Boolean
`daemonPodDetectors.daemonPodNamespaces1` and `daemonPodNamespaces2`	`daemonSet` kinds are allowed for node suspension by default. Adding this parameter changes the default.	Optional but required to identify pods in the namespace to be ignored for cluster consolidation.	Regular expression used, values are in quotes and are comma-separated, such as `"kube-system","kube-service-catalog","openshift-.*"`.
`daemonPodDetectors.daemonPodNamePatterns`	`daemonSet` kinds are allowed for node suspension by default. Adding this parameter changes the default.	Optional but required to identify pods matching this pattern to be ignored for cluster consolidation.	Regular expression used `.*ignorepod.*`
`HANodeConfig.nodeRoles`	Any value for label key value `node-role.kubernetes.io/` for master and others	Optional. This is used to automate policies that keep nodes of the same role limited to one instance per ESX host or availability zone.	Values in `values.yaml` or `cr.yaml` use escapes or quotes, or are comma-separated. Master nodes are by default. Other roles populated by `nodeRoles:""foo","bar""`
`image.busyboxRepository`	`icr.io/cpopen/turbonomic/cpufreqgetter:1.0`	Optional	Full path to the repository, image, and tag.
`image.cpufreqgetterRepository`	`icr.io/cpopen/turbonomic/cpufreqgetter`	Optional	Repository used to get node `cpufrequency`.
`image.imagePullSecret`	None	Optional	Defines the secret used to authenticate to the container image registry
`image.pullPolicy`	`IfNotPresent`	Optional	
`image.repository`	`icr.io/cpopen/turbonomic/kubeturbo`	Optional	Path to the repository. Must be used with `image.tag`.
`image.tag`	Depends on product version	Optional	Kubeturbo tag. Must be used with `image.repository`.

Parameter	Default value	Changes to default value	Parameter type
`kubeturboPodScheduling.affinity`	None	Optional	Specify `affinity` pod scheduling constraint in the cluster. For examples, see the Kubernetes documentation
`kubeturboPodScheduling.nodeSelector`	None	Optional	Specify `nodeSelector` pod scheduling constraint in the cluster. For examples, see the Kubernetes documentation.
`kubeturboPodScheduling.tolerations`	None	Optional	Specify `tolerations` pod scheduling constraint in the cluster. For examples, see the Kubernetes documentation.
`masterNodeDetectors.nodeLabels`	Any value for label key value `node-role.kubernetes.io/master`	Deprecated. Previously used to avoid suspending masters identified by node label key value pair. The value ignored if there is no match.	Regular expression used, specify the key as `masterNodeDetectors.nodeLabelsKey` (such as `node-role.kubernetes.io/master`) and the value as `masterNodeDetectors.nodeLabelsValue` (such as `.*`).
`masterNodeDetectors.nodeNamePatterns`	Node name includes `.*master.*`	Deprecated. Previously used to avoid suspending masters identified by node name. The value is ignored if there is no match.	String, regular expression used. For example: `.*master.*`
`restAPIConfig.opsManagerPassword`	None	Optional	Administrator password.
`restAPIConfig.opsManagerUserName`	None	Optional	Local or Active Directory user with site administrator role.
`restAPIConfig.turbonomicCredentialsSecretName`	`turbonomic-credentials`	Required only if using secret and not taking the default secret name	Secret that contains the Workload Optimization Manager site administrator username and password in Base64.
`roleBinding`	`turbo-all-binding`	Optional	Specify the name of `clusterrolebinding`
`roleName`	`cluster-admin`	Optional	Specify custom `turbo-cluster-reader` or `turbo-cluster-admin` role instead of the default `cluster-admin` role.
`serverMeta.proxy`	None	Optional	Format of `http://username:password@proxyserver:proxyport`

Parameter	Default value	Changes to default value	Parameter type
			or `http://proxyserv-er:proxyport`
`serverMeta.turbo-Server`	None	Required	HTTPS URL to log in to Workload Optimization Manager
`serverMeta.version`	None	Optional	Number `x.y`
`serviceAccountName`	`turbo-user`	Optional	Specify the name of the service account.
`targetConfig.target-Name`	`{Your_cluster}`	Optional but required for multiple clusters	String that indicates how you want to identify your cluster.

Configuring a private repository (OperatorHub)

Workload Optimization Manager gathers information from your clusters through the Kubeturbo container images that are available through a public repository in IBM Container Registry (`icr.io`).

The following list describes the different Kubeturbo container images.

- Container image for Kubeturbo

  This is the primary image and is pulled from `icr.io/cpopen/turbonomic/kubeturbo:<version>`.

- Container image for the CPU frequency getter job

  The `cpufreqgetter` image (also known as BusyBox) runs a job on every node to collect CPU speed. This image is pulled from `icr.io/cpopen/turbonomic/cpufreqgetter`.

  For more information about the parameters associated with this job, see this topic *(on page 304)*.

- (Operator method only) Container image for `kubeturbo-operator`

  This image is pulled for `icr.io/cpopen/kubeturbo-operator:<version>`

By default, the operator deploys a Kubeturbo image that matches the deployed operator version, and pulls directly from IBM Container Registry. While it is important that the Kubeturbo version matches your Workload Optimization Manager version, you might be using a different version or a private repository.

If you are using a private repository, be aware that IBM Container Registry provides multi-architecture container images. Your private repository needs to support loading these images. For example, multi-architecture container images are supported in Artifactory version 7.2+, but not in Artifactory version 6. Optionally, pull only the architecture version that you need. Specify the `--platform` parameter with `docker pull` and use `docker inspect` to confirm the architecture.

```
docker pull --platform linux/arm64 icr.io/cpopen/turbonomic/kubeturbo:8.15.0
docker inspect 49a61c21c3a1 | grep architecture
               "architecture": "aarch64",
```

To use a private repository, you need to define additional parameters.

1. Define the parameters in the operator instance or custom resource you have configured (default is `kubeturbo-release`), as shown in the following example (see a sample YAML here):

```
spec:
  image:
    # Supply your private repo and specific product version here
    # With v8.7.5 and newer, kubeturbo is available via IBM Container Registry
    repository: {your_kubeturbo_repository}/icr.io/cpopen/turbonomic/kubeturbo
    tag: 8.15.0
    # Specify your cpu frequency job container image.

 cpufreqgetterRepository: {your_frequency_getter_repository}/icr.io/cpopen/turbonomic/cpufreqgetter
    # Note cpufreqgetter will use the same pull secret as set for the kubeturbo container image
    #imagePullSecret: secretName
```

```
#Rest of CR will be retained
...
```

Update the sample command with the following parameters:

- `repository: {your_kubeturbo_repository}/icr.io/cpopen/turbonomic/kubeturbo`

  Specify your private Kubeturbo repository, such as `myrepo.com`.

- `cpufreqgetterRepository: {your_frequency_getter_repository}/icr.io/cpopen/turbonomic/cpufreqgetter`

  Specify your private CPU frequency getter repository, such as `mygetterrepo.com`.

2. Pull the Kubeturbo operator and specify the operator's path and digest ID in the `ClusterServerVersion` resource.

   For information on how to work with operators or OLMs in a disconnected or restricted network (also referred to as an air-gapped environment) and how to stage certified operators, see the Red Hat OpenShift documentation. The Kubeturbo certified operator bundle is available in this GitHub repository. Be sure to work with the bundle version that matches the Kubeturbo product version that you need.

After you apply your configuration, an instance or custom resource called `kubeturbo-release` is created, with a Kubeturbo deployed at either the same version as the operator, or a user-specified version.

Container platform monitored resources

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

Container platform	Workload Optimization Manager
Service	Service *(on page 549)*
Container	Container *(on page 577)*
A container's spec	Container spec *(on page 580)*
Pod	Container Pod *(on page 604)*
Controller	Workload Controller *(on page 591)*
Namespace	Namespace *(on page 613)*
Cluster	Container platform cluster *(on page 617)*
Node	Virtual Machine *(on page 624)*
Persistent Volume (PV)	Volume **NOTE:** If a container pod is attached to a volume *(on page 726)*, Workload Optimization Manager discovers it as a Persistent Volume (PV), and shows which pods are connected to the PV.

Monitored resources for services

Workload Optimization Manager monitors the following resources:

- Response time

  Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

  For LLM inference workloads, response time is the turnaround time for each request, including both queuing time and service time. When there is no request, response time is unavailable.

- Transaction

Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

For LLM inference workloads, Transaction is the total number of tokens per second, which includes both input tokens and generated tokens. When there is no request, Transaction is zero.

- Number of replicas

  Number of replicas is the number of Application Component replicas running over a given time period.

- Concurrent queries

  For LLM inference workloads, concurrent queries is the number of concurrent queries to a workload. When there is no request, concurrent queries is zero.

- Queueing time

  For LLM inference workloads, queueing time is the amount of time that a request spends in a queue before it is processed. When there is no request, queueing time is zero.

- Service time

  For LLM inference workloads, service time SLO is the amount of processing time needed to generate the next token. This metric is relatively stable for a given model and GPU resource. When there is no request, service time is unavailable.

Monitored resources for containers

Workload Optimization Manager monitors the following resources:

- Virtual memory (vMem)

  Virtual memory (vMem) is the virtual memory utilized by a container against the memory limit. If no limit is set, node capacity is used.

- vMem request

  If applicable, vMem request is the virtual memory utilized by a container against the memory request.

- vCPU

  vCPU is the virtual CPU (in mCores) utilized by a container against the CPU limit. If no limit is set, node capacity is used).

- vCPU request

  If applicable, vCPU request is the virtual CPU (in mCores) utilized by a container against the CPU request.

- vCPU throttling

  vCPU throttling is the throttling of container virtual CPU that could impact response time, expressed as the percentage of throttling for all containers associated with a Container Spec. In the Capacity and Usage chart for containers, *used* and *utilization* values reflect the actual throttling percentage, while *capacity* value is always 100%.

Monitored resources for container specs

Workload Optimization Manager monitors the historical usage of any instance of a container running for the workload (assuming the workload name stays the same). Charts show the trend of usage even with restarts or redeployments.

Monitored resources for container pods

Workload Optimization Manager discovers pods with the following Status in Kubernetes and Red Hat OpenShift and matches them to a specific State in Workload Optimization Manager.

Kubernetes Status	Workload Optimization Manager State
Running	Active
`ImagePullBackOff`	Unknown
`CrashLoopBackoff`	Unknown
Error	Unknown

It is expected that the total number of pods in the cluster will not match in Workload Optimization Manager due to this. As Workload Optimization Manager also does not discover Pending pods or Completed job pods.

Workload Optimization Manager monitors the following resources:

- vMem

  vMem is the virtual memory utilized by a pod against the node physical capacity.
- vMem request

  vMem request is the virtual memory request allocated by a pod against the node allocatable capacity.
- vCPU

  vCPU is the virtual CPU (in mCores) utilized by a pod against the node physical capacity.
- vCPU request

  vCPU request is the virtual CPU request (in mCores) allocated by a pod against the node allocatable capacity.
- vMem request quota

  If applicable, vMem request quota is the amount of virtual memory request a pod has allocated against the namespace quota.
- vCPU request quota

  If applicable, vCPU request quota is the amount of virtual CPU request (in mCores) a pod has allocated against the namespace quota.
- vMem limit quota

  If applicable, vMem limit quota is the amount of virtual memory limit a pod has allocated against the namespace quota.
- vCPU limit quota

  If applicable, vCPU limit quota is the amount of virtual CPU limits (in mCores) that a pod allocated against the namespace quota.

Monitored resources for workload controllers

Workload Optimization Manager monitors quotas (limits and requests) for vCPU and vMem, and associates how much each Workload Controller is contributing to a quota based on all replicas. This allows Workload Optimization Manager to generate rightsizing decisions, and manage the quota as a constraint to rightsizing. Metrics on resource consumption are shown in the Container Spec, Container, and Container Pod views.

Monitored resources for namespaces

Workload Optimization Manager monitors the following resources:

- vMem request quota

  vMem request quota is the total amount of virtual memory request for all pods allocated to the namespace against the namespace quota.
- vCPU request quota

  vCPU request quota is the total amount of virtual CPU request (in mCores) for all pods allocated to the namespace against the namespace quota.
- vMem limit quota

  vMem limit quota is the total amount of virtual memory limit for all pods allocated to the namespace against the namespace quota.
- vCPU limit quota

  vCPU limit quota is the total amount of virtual CPU limits (in mCores) for all pods allocated to the namespace against the namespace quota.

Monitored resources for container platform clusters

Workload Optimization Manager monitors resources for the containers, pods, nodes (VMs), and volumes in a cluster.

Monitored resources for nodes (VMs)

Workload Optimization Manager monitors the following resources for nodes that host pods. These resources are monitored along with the resources from the infrastructure probes, such as vCenter or a public cloud mediation probe.

- vMem

  vMem is the virtual memory currently used by all containers on the node. The capacity for this resource is the Node Physical capacity.

- vCPU

  vCPU is the virtual CPU currently used by all containers on the node. The capacity for this resource is the Node Physical capacity.

- Memory request allocation

  Memory request allocation is the memory available to the node to support the `ResourceQuota` request parameter for a given Kubernetes namespace or Red Hat OpenShift project.

- CPU request allocation

  CPU request allocation is the CPU available to the node to support the `ResourceQuota` request parameter for a given Kubernetes namespace or Red Hat OpenShift project.

- Virtual memory request

  Virtual memory request is the memory currently guaranteed by all containers on the node with a memory request. The capacity for this resource is the Node Allocatable capacity, which is the amount of resources available for pods and can be less than the physical capacity.

- Virtual CPU request

  Virtual CPU request is the CPU currently guaranteed by all containers on the node with a CPU request. The capacity for this resource is the node allocatable capacity, which is the amount of resources available for pods and can be less than the physical capacity.

- Memory allocation

  Memory allocation is the memory `ResourceQuota` limit parameter for a given Kubernetes namespace or Red Hat OpenShift project.

- CPU allocation

  CPU allocation is the CPU `ResourceQuota` limit parameter for a given Kubernetes namespace or Red Hat OpenShift project.

Container platform actions

Workload Optimization Manager monitors the state and performance of your containerized workloads and then recommends actions to optimize these workloads.

Actions for services

Scale (through workload controllers)

Workload Optimization Manager recommends actions to scale the replicas that back the container platform services. Set the scope to the associated Workload Controller to view and execute the actions.

For details, see Workload Controller Scale Actions *(on page 598)*.

Actions for containers

Resize

Workload Optimization Manager generates container resize actions for Bare Pods where the workload has no controller.

Actions for container specs

Resize (through workload controllers)

A container spec retains the historical utilization data of ephemeral containers. Workload Optimization Manager uses this data to make resize decisions that assure optimal utilization of resources. By default, all replicas of the same container for the same workload type resize consistently.

For details, see Workload Controller Resize Actions *(on page 593)*.

Actions for container pods

- Move

  Move a pod that is controlled by a ReplicationController, ReplicaSet, or Deployment (through a ReplicaSet) to another node (VM) to address performance issues or improve infrastructure efficiency. For example, if a particular node is congested for CPU, you can move pods to a node with sufficient capacity. If a node is underutilized and is a candidate for suspension, you must first move the pods before you can safely suspend the node.

  For details, see Container Pod Move Actions *(on page 606)*.

- Provision or suspend

  When recommending node provision or suspend actions, Workload Optimization Manager will also recommend provisioning pods (based on demand from DaemonSets) or suspending the related pods.

  For details, see Container Pod Provision and Suspension Actions *(on page 610)*.

Actions for workload controllers

- Resize

  Actions associated with a workload controller resize container specs vertically to assure optimal utilization of resources. This is a natural representation of these actions because the parent controller's container specs for vCPU limits/requests and vMem limits/requests are modified. The workload controller then rolls out the changes in the running environment.

  For details, see Workload Controller Resize Actions *(on page 593)*.

  Resize actions for certain types of workloads may be blocked or set to recommend-only, which means that they are currently not executable in Workload Optimization Manager. These actions usually require a dependent action or an external configuration to be completed first. For example, a workload controller resize up action may be temporarily blocked if there is insufficient resources in the corresponding node or namespace. In this case, a node provision or namespace quota resize action must be completed first. For more information, see this topic *(on page 291)*.

- Scale

  Actions associated with a workload controller scale replicas horizontally to maintain Service Level Objectives (SLOs) for your applications. This is a natural representation of these actions because the parent controller's number of replicas are modified. The workload controller then rolls out the changes in the running environment.

  For details, see Workload Controller Scale Actions *(on page 598)*.

Actions for namespaces

Resize quota

Workload Optimization Manager treats quotas defined in a namespace as constraints when making resize decisions. If existing actions would exceed the namespace quotas, Workload Optimization Manager recommends actions to resize up the affected namespace quota.

Note that Workload Optimization Manager does not recommend actions to resize *down* a namespace quota. Such an action reduces the capacity that is already allocated to an application. The decision to resize down a namespace quota should include the application owner.

For details, see Namespace Actions *(on page 616)*.

Actions for container platform clusters

None

Workload Optimization Manager does not recommend actions for a container platform cluster. Instead, it recommends actions for the containers, pods, nodes (VMs), and volumes in the cluster. Workload Optimization Manager shows all of these actions when you scope to a container platform cluster and view the Pending Actions chart.

Actions for nodes (VMs)

A node (cloud or on-prem) is represented as a Virtual Machine entity in the supply chain.

> **NOTE:**
> For nodes in the public cloud, Workload Optimization Manager reports the cost savings or investments attached to these actions.

For details, see Node Actions *(on page 626)*.

Special cases for container platform actions

This topic describes the action handling mechanism for certain types of container platform workloads.

Disabled move actions for pods with topology spread constraints

Currently, Workload Optimization Manager does not support moving pods with topology spread constraints that use the `DoNotSchedule` for the `whenUnsatisfiable` setting. These pods apply an auto-generated policy in your container platform that disables pod move actions.

Until support is available, Workload Optimization Manager prevents automating pod move actions in the following ways.

- Workload Optimization Manager adds the affected pods in each cluster to a group called `TopologySpreadConstraint Pods`. This group applies a Workload Optimization Manager read-only policy called `TopologySpreadConstraint Pods Move Disabled`, which sets the action acceptance mode for pod move actions to *Disabled*.
- For the nodes that the affected pods run on, Workload Optimization Manager adds the nodes to a group called `TopologySpreadConstraint Nodes`. This group applies a Workload Optimization Manager read-only policy called `TopologySpreadConstraint Nodes Suspend Disabled`, which sets the action acceptance mode for node suspension actions to *Disabled*.

Recommend-only move actions for certain controlled pods

For a controlled pod (non-bare pod) to be movable, it must be controlled by one of the following built-in APIs for managing workloads:

- `Deployment`
- `DeploymentConfig`
- `ReplicaSet`
- `ReplicationController`

For pods that are controlled in other ways, Workload Optimization Manager adds the affected pods in each cluster to a group called `Unmovable ContainerPods`. This group applies a Workload Optimization Manager read-only policy called `Unmovable ContainerPods Move Recommend Only`, which sets the action acceptance mode for pod move actions to *Recommend*. This means that you can only execute pod move actions outside of Workload Optimization Manager.

Recommend-only move actions for StatefulSet pods

Currently, Workload Optimization Manager does not support moving StatefulSet pods. These pods apply an auto-generated policy in your container platform that sets the action acceptance mode for pod move actions to recommend-only.

Until support is available, Workload Optimization Manager prevents automating pod move actions by adding the affected pods in each cluster to a group called `StatefulSet/All Pods`. This group applies a Workload Optimization Manager read-only policy called `StatefulSet Pods Move Recommended Only`, which sets the action acceptance mode for pod move actions to *Recommend*. This means that you can only execute pod move actions outside of Workload Optimization Manager.

Recommend-only resize actions for sidecar container specs

A container platform service might include sidecar container specs to provide additional services to a running pod, such as security or logging services. Sidecars injected at pod creation cannot be updated from the parent workload controller because limits or requests are controlled by another workload controller. Because the sidecar specification does not exist, a resize action fails.

To prevent the execution of resize actions on injected sidecars, Workload Optimization Manager adds the sidecars in each a cluster to a group called `Injected Sidecars/All ContainerSpecs`. This group applies a Workload Optimization Manager read-only policy called `Injected Sidecars/All ContainerSpecs Resize Recommend Only [cluster_name]`, which sets the action acceptance mode for resizes to *Recommend*. This means that you can only execute resizes outside of Workload Optimization Manager. The parent workload controller will continue to resize non-sidecar container specs as usual.

Recommend-only resize actions for container specs that are unresizable in Workload Optimization Manager

For a container spec to be resizable in Workload Optimization Manager, it must be controlled by one of the following built-in APIs for managing workloads:

- `DaemonSet`
- `Deployment`
- `DeploymentConfig`
- `ReplicaSet`
- `ReplicationController`
- `StatefulSet`

For container specs that are controlled by other APIs (for example by `CronJob` or `Job`), Workload Optimization Manager adds the affected container specs in each cluster to a group called `Unresizable ContainerSpecs`. This group applies a Workload Optimization Manager read-only policy called `Unresizable ContainerSpecs Resize Recommend Only`, which sets the action acceptance mode for container spec resize actions to *Recommend*. This means that you can only execute resize actions outside of Workload Optimization Manager.

NOTE:
For container specs that are controlled by Custom Resources (or operators), see the next section.

Recommend-only resize actions for operator-controlled workloads

For workloads in a cluster that are controlled by an operator, Kubeturbo automatically creates a group called `Operator Controlled ContainerSpecs`. This group applies a Workload Optimization Manager read-only policy called `Operator Controlled ContainerSpecs Resize Recommend Only [cluster_name]`, which sets the action acceptance mode for resizes to *Recommend*. This means that you can only execute resizes outside of Workload Optimization Manager.

Workload Optimization Manager prevents or blocks the creation of a conflicting policy that automates resize actions.

When an ORM is deployed, you can execute resize actions for specific operator-controlled workloads by adding these workloads to an exclusion list that is defined in a `configmap`.

For OperatorHub or operator deployments of Kubeturbo, update the `kubeturbo-release` Custom Resource (CR).

1. Edit `kubeturbo-release`.

   ```
   kubectl edit kubeturbo kubeturbo-release
   ```

2. In the `exclusionDetectors` section, use one or both of the following parameters to exclude workloads:
   - `operatorControlledWorkloadsPatterns` – Specify workload name patterns.
   - `operatorControlledNamespacePatterns` – Specify namespace name patterns.

```
spec:
  exclusionDetectors:
    operatorControlledWorkloadsPatterns:
    - turbon.*
    - testing-.*
    operatorControlledNamespacePatterns:
    - turbonomic
    - gke-.*
```

3. Verify that `configmap` was updated.

   There are two ways to verify the updates.

   - Check the `configmap`.

     ```
     kubectl get configmap turbo-config-kubeturbo-release -o yaml
     ```

     In the output, the `turbo-autoreload.config` section shows the updates.

     ```
     apiVersion: v1
     data:
       turbo-autoreload.config: |-
         {
          ...
          },
          "exclusionDetectors": {
            "operatorControlledWorkloadsPatterns": ["turbon.*","testing-.*"],
            "operatorControlledNamespacePatterns": ["turbonomic","gke-.*"]
          }
        }
     ```

   - Check the Kubeturbo logs.

     ```
     kubectl logs kubeturbo-release-7cb4b886c5-glv9b | grep exclusion
     ```

     Log entries shows the updates.

     ```
     I1122 14:43:07.603989 dynamic_config.go:119] Operator controlled workload exclusion set to:
      []
     I1122 14:43:07.604013 dynamic_config.go:125] Operator controlled namespace exclusion set to:
      [turbonomic]
     I1122 19:14:52.314970 dynamic_config.go:119] Operator controlled workload exclusion set to:
      [turbon.* testing-.*]
     I1122 19:14:52.315096 dynamic_config.go:125] Operator controlled namespace exclusion set to:
      [turbonomic gke-.*]
     ```

Disabled resize actions for system namespace workloads

Workload Optimization Manager disables resize actions for workloads in well-known system namespaces, such as those starting with `kube-.*`, `openshift-.*`, or `cattle.*`. A setting in the Kubeturbo `configmap` automatically creates a group called `System Namespaced ContainerSpecs` in each cluster. This group applies a Workload Optimization Manager read-only policy called `System Namespaced ContainerSpecs Resize Disabled [`*`cluster_name`*`]`, which sets the action acceptance mode for resizes to *Disabled*.

For OperatorHub or operator deployments of Kubeturbo, update the `kubeturbo-release` Custom Resource (CR).

1. Edit `kubeturbo-release`.

   ```
   kubectl edit kubeturbo kubeturbo-release
   ```

2. In the `namespacePatterns` section, add name patterns for namespaces.

   ```
   spec:
     systemWorkloadDetectors:
   ```

```
    namespacePatterns:
    - kube-.*
    - openshift-.*
    - cattle.*
```

For matching namespaces, Workload Optimization Manager adds all the discovered workloads into the `System Namespaced ContainerSpecs` group.

3. Verify that `configmap` was updated.

   There are two ways to verify the updates.

   - Check the `configmap`.

     ```
     kubectl get configmap turbo-config-kubeturbo-release -o yaml
     ```

     In the output, the `turbo-autoreload.config` section shows the updates.

     ```
     apiVersion: v1
     data:
       turbo-autoreload.config: |-
         {
          ...
          },
          "systemWorkloadDetectors": {
            "namespacePatterns": ["kube-.*","openshift-.*","cattle.*"]
          }
        }
     ```

   - Check the Kubeturbo logs.

     ```
     kubectl logs kubeturbo-release-7cb4b886c5-glv9b | grep -i "namespace det"
     ```

     Log entries shows the updates.

     ```
     I1122 14:43:07.603966 dynamic_config.go:113] System Namespace detectors set to: [kube-.*]
     I1122 19:14:52.314883 dynamic_config.go:113] System Namespace detectors set to: [kube-.*
      openshift-.* cattle.*]
     ```

Minimum and maximum nodes for node pools and MachineSets

Workload Optimization Manager manages cluster resources based on changing demand and recommends actions to provision *(on page 628)* and suspend nodes in a node pool. These actions have a default minimum node count of 1 and a maximum of 100, as defined in the `configmap`.

```
data:
  turbo-autoreload.config: |-
    {
      "nodePoolSize": {
        "min": 1,
        "max": 1000
```

You can adjust the default node counts for all discovered Red Hat OpenShift MachineSets in the Red Hat OpenShift cluster.

NOTE:
Workload Optimization Manager does not increase the number of nodes from 0 and assumes that you have a way of creating the first node in a MachineSet.

1. Edit `kubeturbo-release`.

   ```
   oc edit kubeturbo kubeturbo-release
   ```

2. In the `nodePoolSize` section, update the values, as shown in the following example. You can specify a minimum of 0.

```
spec:
  nodePoolSize:
    max: 100
    min: 2
```

3. Verify that `configmap` was updated.

```
oc get configmap turbo-config-kubeturbo-release -o yaml
```

In the output, the `turbo-autoreload.config` section shows the updates.

```
data:
  turbo-autoreload.config: |-
    {
      "nodePoolSize": {
        "min": 2,
        "max": 100
```

Updating Kubeturbo

Update the following Kubeturbo information as needed.

- Kubeturbo image tag version

  Update the image tag version of your Kubeturbo deployment whenever your Workload Optimization Manager instance updates to a new version or if you are instructed to do so.

- Kubeturbo role

  Update the role if you deployed Kubeturbo with a specific role and now want to use a different role. For example, you may want to change the role from `turbo-cluster-reader` to `turbo-cluster-admin` to execute actions in Workload Optimization Manager.

Updating the Kubeturbo version through YAML

Update the image tag version of your Kubeturbo deployment whenever your Workload Optimization Manager instance updates to a new version or if you are instructed to do so.

To determine the image tag version to specify, open the Workload Optimization Manager user interface, click the Help icon in the navigation menu, and then record the version number that displays. This version number has a corresponding image tag version. For example, if your Workload Optimization Manager instance version is 3.13.0, the corresponding image tag version is 8.15.0. For the latest Workload Optimization Manager and image tag version mapping, see this reference *(on page 302)*.

There are several ways to update the Kubeturbo deployment.

- Update your manifest.
- Update the running Kubeturbo `deployment` resource.

Updating your manifest

1. In the `deployment` YAML that you used to create the Kubeturbo `deployment` resource, specify the image tag version in the `image:` parameter.

```
image: icr.io/cpopen/turbonomic/kubeturbo:{new_image_tag_version}
```

2. Apply the update.

```
kubectl apply -f {yourDeployment}.yaml -n turbo
```

Repeat these steps for every cluster with a Kubeturbo deployment.

Updating the running Kubeturbo `deployment` resource

1. Update the deployment through the Kubernetes/Red Hat OpenShift dashboard, or the command line (by using the `kubectl` or `oc edit deployment` command).

   The following example shows the `kubectl` command.

   ```
   kubectl edit deployment kubeturbo -n turbo
   ```

2. Update the following parameter.

   ```
   image: icr.io/cpopen/turbonomic/kubeturbo:{new_image_tag_version}
   ```

   After you save your changes, the Kubeturbo pod redeploys.

Repeat these steps for every cluster with a Kubeturbo deployment.

Updating the Kubeturbo version through Helm

Update the image tag version of your Kubeturbo deployment whenever your Workload Optimization Manager instance updates to a new version or if you are instructed to do so.

To determine the image tag version to specify, open the Workload Optimization Manager user interface, click the Help icon in the navigation menu, and then record the version number that displays. This version number has a corresponding image tag version. For example, if your Workload Optimization Manager instance version is 3.13.0, the corresponding image tag version is 8.15.0. For the latest Workload Optimization Manager and image tag version mapping, see this reference *(on page 302)*.

1. Update the Kubeturbo deployment.

   ```
   helm upgrade {deployment_name} {helm_chart_location} \
   --namespace turbo \
   --set serverMeta.turboServer={your_instance_address} \
   --set serverMeta.version={your_instance_version} \
   --set image.tag={new_image_tag_version} \
   --set targetConfig.targetName={your_cluster_name}
   ```

2. Verify that the Kubeturbo pod restarted and has been updated with the new values.

Repeat these steps for every cluster with a Kubeturbo deployment.

Updating the Kubeturbo version through an operator

Update the image tag version of your Kubeturbo deployment whenever your Workload Optimization Manager instance updates to a new version or if you are instructed to do so.

To determine the image tag version to specify, open the Workload Optimization Manager user interface, click the Help icon in the navigation menu, and then record the version number that displays. This version number has a corresponding image tag version. For example, if your Workload Optimization Manager instance version is 3.13.0, the corresponding image tag version is 8.15.0. For the latest Workload Optimization Manager and image tag version mapping, see this reference *(on page 302)*.

There are several ways to update the Kubeturbo deployment.

- Update your manifest.
- Update the running Kubeturbo `deployment` resource.

Updating your manifest

1. Edit the original custom resource used to deploy Kubeturbo. For example, if you downloaded the YAML resource for the `cluster-admin` role, edit `deploy/kubeturbo_yamls/turbo_kubeturbo_operator_full.yaml`.

   Update the following parameter:

   ```
   image: icr.io/cpopen/kubeturbo-operator:{new_image_tag_version}
   ```

2. (Optional) If you are using a private repository, update the following parameter:

```
tag: {new_image_tag_version}
```

3. Apply the change to the operator.

```
kubectl apply -f deploy/kubeturbo_yamls/{YAML_resource}
```

For example:

```
kubectl apply -f deploy/kubeturbo_yamls/turbo_kubeturbo_operator_full.yaml
```

Repeat these steps for every cluster with a Kubeturbo deployment.

Updating the running Kubeturbo `deployment` resource

1. If you do not have the original custom resource used to deploy Kubeturbo, you can edit Kubeturbo directly by running the following command:

```
kubectl edit deployment kubeturbo-operator
```

2. Specify the image tag version in the `image:` parameter.

```
image: icr.io/cpopen/kubeturbo-operator:{new_image_tag_version}
```

3. (Optional) If you are using a private repository, update the following parameter:

```
tag: {new_image_tag_version}
```

After you save your changes, the Kubeturbo pod redeploys.

Repeat these steps for every cluster with a Kubeturbo deployment.

Updating when using a non-default cluster role

If you are updating from Kubeturbo 8.9.6 or earlier and are using a non-default cluster role, such as `turbo-cluster-reader` or `turbo-cluster-admin`, perform the following steps after you update Workload Optimization Manager to ensure that your deployment uses the updated cluster role names and to avoid errors.

1. Delete all cluster roles that start with the names `turbo-cluster-reader` and `turbo-cluster-admin`. There could be four cluster roles with such names.
2. Delete any cluster role binding that starts with the name `turbo-all-binding-kubeturbo`.

After a few minutes, the kubeturbo-operator automatically recreates the required non-default cluster role and cluster role binding using the new names.

Updating the Kubeturbo version through OperatorHub

Update the image tag version of your Kubeturbo deployment whenever your Workload Optimization Manager instance updates to a new version or if you are instructed to do so.

To determine the image tag version to specify, open the Workload Optimization Manager user interface, click the Help icon in the navigation menu, and then record the version number that displays. This version number has a corresponding image tag version. For example, if your Workload Optimization Manager instance version is 3.13.0, the corresponding image tag version is 8.15.0. For the latest Workload Optimization Manager and image tag version mapping, see this reference *(on page 302)*.

1. Open the installed operator and update the Kubeturbo operator based on your subscription. You can enable automatic updates or require update approvals. The Kubeturbo operator is given the same version as the Kubeturbo product version.

   If your subscription is set to automatic, Kubeturbo automatically updates when a new version of Kubeturbo is released. There are no manual steps for you to perform.

2. If your Kubeturbo release was installed with default values, check if the `kubeturbo-release` pod has restarted and verify the new version in the Workload Optimization Manager user interface (in **Settings > Target Configuration**, open the target that corresponds to the Red Hat OpenShift cluster).

   Repeat this step for every cluster with a Kubeturbo deployment.

3. If you are using a non-default cluster role and your version of the Kubeturbo operator was deployed before 8.9.5, review the following guidelines and additional steps.

   - As of Kubeturbo operator 8.9.5, there were changes introduced to have a new unique cluster role binding name. Additionally, in version 8.9.6, there were changes introduced to have a new unique cluster role name. These changes were introduced to support deploying multiple operator-based Kubeturbo in a single cluster. As such, some manual changes are needed to ensure that your operator-based deployment of Kubeturbo continues to work and uses these new unique cluster role and cluster role binding names. Since the Kubeturbo operator is Helm-based, these manual steps are required to delete the old cluster role and cluster role binding names.
   - There are a few different scenarios where you will be required to perform the following manual steps to use these new cluster role and cluster role binding names.
     - Your operator or OperatorHub deployment of Kubeturbo uses a custom cluster role, such as `turbo-cluster-reader`, and you now want to change it to `turbo-cluster-admin` to allow Kubeturbo to execute actions automatically.
     - Your operator or OperatorHub deployment of Kubeturbo uses a custom cluster role, such as `turbo-cluster-reader` or `turbo-cluster-admin`, and you upgraded a version older than 8.9.5 to a later version.

   Perform these additional steps.

   a. Delete all cluster roles that start with the names `turbo-cluster-reader` and `turbo-cluster-admin`. There could be four cluster roles with such names.

   b. Delete any cluster role binding that starts with the name `turbo-all-binding-kubeturbo`.

   After a few minutes, the Kubeturbo operator automatically recreates the required non-default cluster role and cluster role binding using the new names.

4. If you have explicitly specified a Kubeturbo version in the `kubeturbo-release` custom resource (`kubeturbo-release` is default), update the `tag:` parameter to upgrade your installation, as shown in the following example..

```
spec:
  image:
    #supply your private repo and specific product version here
    repository: "icr.io/cpopen/turbonomic/kubeturbo"
    tag: {new_image_tag_version}
  #rest of CR will be retained
  ...
```

5. Apply the updates.

Repeat these steps for every cluster with a Kubeturbo deployment.

Updating the Kubeturbo role through YAML

Update the role if you deployed Kubeturbo with a specific role and now want to use a different role. For example, you may want to change the role from `turbo-cluster-reader` to `turbo-cluster-admin` to execute actions in Workload Optimization Manager.

1. Delete the Cluster Role Binding (CRB) that starts with the naming convention `turbo-all-binding-kubeturbo`.

   For example, to delete `turbo-all-binding-kubeturbo-turbo`, run the following command:

```
kubectl delete clusterrolebinding turbo-all-binding-kubeturbo-turbo
```

2. Create a new role for Kubeturbo.
   - To use the `turbo-cluster-admin` role, run the following command.

     ```
     kubectl apply -f
      https://raw.githubusercontent.com/turbonomic/kubeturbo-deploy/master/deploy/kubeturbo_yam
     ls/turbo-admin.yaml
     ```

   - To use the `turbo-cluster-reader` role, run the following command.

     ```
     kubectl apply -f
      https://raw.githubusercontent.com/turbonomic/kubeturbo-deploy/master/deploy/kubeturbo_yam
     ls/turbo-reader.yaml
     ```

     NOTE:
     For the default `cluster-admin` role, there is no YAML resource to apply. Kubeturbo will automatically switch to this role after you apply the corresponding CRB (see the next step).

3. Create a new CRB for Kubeturbo.
   - To use the `turbo-cluster-admin` CRB, run the following command.

     ```
     kubectl apply -f
      https://raw.githubusercontent.com/turbonomic/kubeturbo-deploy/master/deploy/kubeturbo_yam
     ls/step3_turbo_serviceAccountRoleBinding_turbo-cluster-admin_sample.yaml
     ```

   - To use the `turbo-cluster-reader` role, run the following command.

     ```
     kubectl apply -f
      https://raw.githubusercontent.com/turbonomic/kubeturbo-deploy/master/deploy/kubeturbo_yam
     ls/step3_turbo_serviceAccountRoleBinding_turbo-cluster-reader_sample.yaml
     ```

   - To use the `cluster-admin` CRB, run the following command.

     ```
     kubectl apply -f
      https://raw.githubusercontent.com/turbonomic/kubeturbo-deploy/master/deploy/kubeturbo_yam
     ls/step3_turbo_serviceAccountRoleBinding_admin_sample.yaml
     ```

4. Delete the Kubeturbo pod.
   a. Get a list of pods.

      ```
      kubectl get pods -n turbo
      ```

   b. Find the Kubeturbo pod. The following example shows a Kubeturbo pod named `kubeturbo-asdf1234asd3`.

      ```
      NAME                    READY   STATUS    RESTARTS   AGE
      kubeturbo-asdf1234asd3  1/1     Running   0          37m
      ```

   c. Delete the Kubeturbo pod.

      ```
      kubectl delete pod kubeturbo-{kubeturbo_pod_ID} -n turbo
      ```

      For example:

      ```
      kubectl delete pod kubeturbo-asdf1234asd3 -n turbo
      ```

Updating the Kubeturbo role through Helm

Update the role if you deployed Kubeturbo with a specific role and now want to use a different role. For example, you may want to change the role from `turbo-cluster-reader` to `turbo-cluster-admin` to execute actions in Workload Optimization Manager.

1. Delete the Cluster Role Binding (CRB) that starts with the naming convention `turbo-all-binding-kubeturbo`.

   For example, to delete `turbo-all-binding-kubeturbo-turbo`, run the following command:

   ```
   kubectl delete clusterrolebinding turbo-all-binding-kubeturbo-turbo
   ```

2. Re-clone the Kubeturbo repository to your local environment. The deployment will not work if you do not re-clone the repository.

   ```
   git clone https://github.com/turbonomic/kubeturbo.git
   ```

3. Update the Kubeturbo deployment.

   - Option 1: Change to the default role

     By default, Kubeturbo deploys to your cluster with the `cluster-admin` role. This role has full control over every resource in the cluster.

     If you want to use the default role, run the following command. This command does not explicitly specify a role parameter, which means that the default role will be used.

     ```
     helm upgrade kubeturbo kubeturbo/deploy/kubeturbo \
     --namespace turbo \
     --set serverMeta.turboServer={your_instance_address} \
     --set serverMeta.version={your_instance_version} \
     --set image.tag={your_image_tag_version} \
     --set targetConfig.targetName={your_cluster_name}
     ```

   - Option 2: Change to the `turbo-cluster-admin` custom role

     This custom role specifies the minimum permissions that Kubeturbo needs to monitor your workloads and execute the actions that Workload Optimization Manager generated to optimize these workloads.

     If you want to use this role, specify the role parameter, as shown in the last line of the following command.

     ```
     helm upgrade kubeturbo kubeturbo/deploy/kubeturbo \
     --namespace turbo \
     --set serverMeta.turboServer={your_instance_address} \
     --set serverMeta.version={your_instance_version} \
     --set image.tag={your_image_tag_version} \
     --set targetConfig.targetName={your_cluster_name} \
     --set roleName=turbo-cluster-admin
     ```

   - Option 3: Change to the `turbo-cluster-reader` custom role

     This custom role is the least privileged role. It specifies the minimum permissions that Kubeturbo needs to monitor your workloads. Actions that Workload Optimization Manager generated to optimize these workloads can only be executed outside of Workload Optimization Manager (for example, in your cluster).

     If you want to use this role, specify the role parameter, as shown in the last line of the following command.

     ```
     helm upgrade kubeturbo kubeturbo/deploy/kubeturbo \
     --namespace turbo \
     --set serverMeta.turboServer={your_instance_address} \
     --set serverMeta.version={your_instance_version} \
     --set image.tag={your_image_tag_version} \
     --set targetConfig.targetName={your_cluster_name} \
     --set roleName=turbo-cluster-reader
     ```

4. Verify that the Kubeturbo pod restarted and has been updated with the new values.

Updating the Kubeturbo role through an operator

Update the role if you deployed Kubeturbo with a specific role and now want to use a different role. For example, you may want to change the role from `turbo-cluster-reader` to `turbo-cluster-admin` to execute actions in Workload Optimization Manager.

There are several ways to update the Kubeturbo deployment.

- Update your manifest.
- Update the running Kubeturbo `deployment` resource.

Updating your manifest

1. Update the YAML resource that you downloaded during the initial Kubeturbo deployment. For example, update `turbo_kubeturbo_operator_reader_full.yaml`.

   Update the following parameter:

   ```
   roleName: {new_role_name}
   ```

   For example, if the role initially assigned to Kubeturbo is the `turbo-cluster-reader` role and you want to change this to the `turbo-cluster-admin` role, update the parameter as follows.

   ```
   roleName: turbo-cluster-admin
   ```

2. Apply the change to the operator.

   ```
   kubectl apply -f deploy/kubeturbo_yamls/{YAML_resource}
   ```

   For example:

   ```
   kubectl apply -f deploy/kubeturbo_yamls/turbo_kubeturbo_operator_reader_full.yaml
   ```

3. Delete the Cluster Role Binding (CRB) that starts with the naming convention `turbo-all-binding-kubeturbo`.

   For example, to delete `turbo-all-binding-kubeturbo-release-turbo`, run the following command:

   ```
   kubectl delete clusterrolebinding turbo-all-binding-kubeturbo-release-turbo
   ```

Updating the running Kubeturbo `deployment` resource

1. If you do not have the original custom resource used to deploy Kubeturbo, you can edit Kubeturbo directly by running the following command:

   ```
   kubectl edit kubeturbo kubeturbo-release
   ```

2. Update the following parameter.

   ```
   roleName: {new_role_name}
   ```

   For example, if the role initially assigned to Kubeturbo is the `turbo-cluster-reader` role and you want to change this to the `turbo-cluster-admin` role, update the parameter as follows.

   ```
   roleName: turbo-cluster-admin
   ```

After you save your changes, the Kubeturbo pod redeploys.

Updating the Kubeturbo role through OperatorHub

Update the role if you deployed Kubeturbo with a specific role and now want to use a different role. For example, you may want to change the role from `turbo-cluster-reader` to `turbo-cluster-admin` to execute actions in Workload Optimization Manager.

Updating the role

1. In the Kubeturbo Operator page, open the **Kubeturbo operator** tab.
2. Select your Kubeturbo deployment. For example, select **kubeturbo-release**.
3. Select **YAML view**.
4. Update the role for Kubeturbo.
   - Change to the default `cluster-admin` role

     To change to this role, delete the existing role. When this field is left empty, Kubeturbo will automatically switch to the default `cluster-admin`.
   - Change to the `turbo-cluster-admin` role

     This custom role specifies the minimum permissions that Kubeturbo needs to monitor your workloads and execute the actions that Workload Optimization Manager generated to optimize these workloads.

     ```
     spec:
       roleName: turbo-cluster-admin
     ```
   - Change to the `turbo-cluster-reader` role

     This custom role is the least privileged role. It specifies the minimum permissions that Kubeturbo needs to monitor your workloads. Actions that Workload Optimization Manager generated to optimize these workloads can only be executed outside of Workload Optimization Manager (for example, in your cluster).

     ```
     spec:
       roleName: turbo-cluster-reader
     ```
5. Save your changes.

Deleting the existing CRB for Kubeturbo

1. Navigate to **User Management > RoleBindings**.
2. Search for `turbo-all-binding-kubeturbo-release`.
3. Click the three dots icon for the CRB and then select **Delete ClusterRoleBinding**.

The operator automatically creates the CRB based on the new role that you configured.

Kubeturbo references

Review the following topics for additional information on configuring and troubleshooting Kubeturbo.

- Node CPU frequency container image *(on page 304)*
- Kubeturbo ConfigMap values *(on page 309)*
- Kubeturbo logging *(on page 306)*

Workload Optimization Manager version and image tag mapping

When you deploy the Kubeturbo or Prometurbo agent to your container platform cluster, you specify the image tag version for specific resources. The image tag version depends on the version of your Workload Optimization Manager instance.

To determine the image tag version to specify, open the Workload Optimization Manager user interface, click the Help icon in the navigation menu, and then record the version number that displays. This version number has a corresponding image tag version. For example, if your Workload Optimization Manager instance version is 3.13.0, the corresponding image tag version is 8.15.0.

Use the following version mapping table as a reference.

Workload Optimization Manager instance version	Image tag version
3.13.0	8.15.0

Workload Optimization Manager instance version	Image tag version
3.12.6	8.14.6
3.12.5	8.14.5
3.12.4	8.14.4
3.12.3	8.14.3
3.12.2	8.14.2
3.12.1	8.14.1
3.12.0	8.14.0
3.11.6	8.13.6
3.11.5	8.13.5
3.11.4	8.13.4
3.11.3	8.13.3
3.11.2	8.13.2
3.11.1	8.13.1
3.11.0	8.13.0
3.10.6	8.12.6
3.10.5	8.12.5
3.10.4	8.12.4
3.10.3	8.12.3
3.10.2	8.12.2
3.10.1	8.12.1
3.10.0	8.12.0
3.9.6	8.11.6
3.9.5	8.11.5
3.9.4	8.11.4
3.9.3	8.11.3
3.9.2	8.11.2
3.9.1	8.11.1
3.9.0	8.11.0
3.8.7	8.10.7
3.8.6	8.10.6
3.8.5	8.10.5
3.8.4	8.10.4
3.8.3	8.10.3
3.8.2	8.10.2

Workload Optimization Manager instance version	Image tag version
3.8.1	8.10.1
3.8.0	8.10.0
3.7.7	8.9.7
3.7.6	8.9.6
3.7.5	8.9.5
3.7.4	8.9.4
3.7.3	8.9.3
3.7.2	8.9.2
3.7.1	8.9.1
3.7.0	8.9.0
3.6.6	8.8.6
3.6.5	8.8.5
3.6.4	8.8.4
3.6.3	8.8.3
3.6.2	8.8.2
3.6.1	8.8.1
3.6.0	8.8.0

Node CPU frequency container image

Workload Optimization Manager represents CPUs in terms of their actual CPU frequencies to keep its calculations uniform across various platforms. To do this, Workload Optimization Manager needs the CPU frequency of each node from the container platform cluster.

Workload Optimization Manager runs a lightweight CPU frequency getter job on each node each time the Kubeturbo pod starts/restarts or when a new node is started/detected. The job queries the CPU frequency and then exits. In most environments this job only takes a maximum of 30 seconds to run on each node. By default, the job includes minimally defined CPU and memory limits and requests.

```
resources:
  limits:
    cpu: 10m
    memory: 50Mi
  requests:
    cpu: 10m
    memory: 50Mi
```

You can control how this job runs.

- Image

  To keep the job lightweight, Workload Optimization Manager uses the `cpufreqgetter` image (`icr.io/cpopen/turbonomic/cpufreqgetter:latest`) based on Alpine. You can use Universal Base Image 8 Minimal images from Red Hat (`registry.access.redhat.com/ubi8/ubi-minimal`).

To configure the image:

- YAML method

  Add the following to the Kubeturbo deployment, under `args`:

  ```
  --busybox-image=yourRepo
  ```

- Operator method

  Add the following to the `kubeturbo-release` custom resource in Red Hat OpenShift:

  ```
  --busybox-image=registry.access.redhat.com/ubi8/ubi-minimal
  ```

  NOTE:
  Any image that can provide a bash shell is sufficient.

- Node selection

  You can choose to exclude certain nodes from running this job. Workload Optimization Manager then uses a default CPU speed value of 2800 MHz per core. If you are running a heterogeneous CPU speed environment, it is recommended that you do not disable this job. Otherwise, CPU speed values may be inaccurate.

  To configure node selection:

  - YAML method

    Add the following to the Kubeturbo deployment, under `args`:

    ```
    --cpufreq-job-exclude-node-labels=kubernetes.io/key=value
    ```

  - Operator method

    Update the `kubeturbo-release` custom resource in Red Hat OpenShift as follows:

    ```
    busyboxExcludeNodeLabels: "kubernetes.io/key=value"
    ```

  - Helm method

    Add the following to the `values.yaml` file and upgrade the chart.

    ```
    busyboxExcludeNodeLabels: "kubernetes.io/key=value"
    ```

The following examples show a modified job on the Kubeturbo `deployment` resource.

- YAML method

```
spec:
  template:
    spec:
      containers:
        args:
          # Uncomment to override default of icr.io/cpopen/turbonomic/cpufreqgetter, and specify
 your own location
          #- --busybox-image=docker.io/busybox
          # or uncomment the following to pull from RHCC
          #- --busybox-image=registry.access.redhat.com/ubi8/ubi-minimal
          # Uncomment to specify the secret name which holds the credentials to busybox image
          #- --busybox-image-pull-secret=<secret-name>
          # Specify nodes to exclude from cpu frequency getter job.
          # Note kubernetes.io/os=windows and/or beta.kubernetes.io/os=windows labels will be
 automatically excluded by default.
          # If specified all the labels will be used to select the node ignoring the default.
          #- --cpufreq-job-exclude-node-labels=kubernetes.io/key=value
```

- Operator method

```
 # Uncomment next lines to specify a repository and image tag for kubeturbo
 #image:
 #  repository: registry.connect.redhat.com
 #  tag: 8.4.0
 # Uncomment to use an image from RHCC for cpu-frequency getter job - predefined in OCP Operator
Hub version. Must have `image` uncommented
 #  busyboxRepository: registry.access.redhat.com/ubi8/ubi-minimal
 # Note busy box will use the same pull secret as set for the kubeturbo container image
 #  imagePullSecret: secretName
```

```
 # Uncomment out `args` and the parameters you want to control the cpu frequency job (string
input)
 #args:
 #  busyboxExcludeNodeLabels: 'kubernetes.io/key=value'
```

NOTE:
To control the namespace to which this job runs, use the downward API as an environment variable. See the next section for an example.

Working with older deployments and configuring Kubeturbo job's namespace

If you have an older deployment of Kubeturbo (version 8.4.x and older), you may find jobs started in the `default` namespace. You can redeploy Kubeturbo with more current YAMLs or update your deployment with the environment values described in this section.

Kubeturbo sample YAMLs and other installation methods use the downward API to detect the namespace that Kubeturbo is running in. The Kubeturbo `deployment` spec shows the following snippet:

```
        env:
          - name: KUBETURBO_NAMESPACE
            valueFrom:
              fieldRef:
                fieldPath: metadata.namespace
```

Kubeturbo injects the `env` variable named KUBETURBO_NAMESPACE to identify its own namespace and ensure that any additional resources that runs for each node (for example, the CPU frequency getter job) are deployed in the same namespace. If this variable is omitted, Kubeturbo use the namespace named `default` to create and manage these resources. This variable can be configured to any other namespace, as shown in the following snippet:

```
        env:
          - name: KUBETURBO_NAMESPACE
            value: <namespace-name>
```

Kubeturbo logging

By default, Kubeturbo writes logs to a file in the container local storage at `/var/log` if the location is mounted as a writable file system within the container. If `/var/log` is not writable, Kubeturbo writes logs to `/tmp`.

The default logging level for Kubeturbo is 2 and should only be changed if instructed by your Workload Optimization Manager representative. Be sure to restore the default level of 2 after any temporary increase in logging.

When the Kubeturbo pod (operator or release) is restarted, you lose all historical logs because Kubeturbo does not use persistent storage.

Collecting and configuring Kubeturbo logs

Kubeturbo logs capture useful information even when logging is set to 2 (default). To get the logs, use the `kubectl logs` or `oc logs` command.

```
kubectl logs -n {namespace} {podName}
```

You can save the logs to a text file and then send the file to your Workload Optimization Manager representative.

```
kubectl logs -n {namespace} {podName} > kubeturbo.log
```

If you deployed Kubeturbo through an operator or OperatorHub, include logs for both the `kubeturbo-operator` and `kubeturbo-release` pods. Note that the OCP console does not always show the full log. Use the CLI or select the **Raw** option in the console to copy the full log content.

Increasing the logging level

Kubeturbo supports changing the logging level without a pod restart. It takes around one minute for the change to take effect.

Your Workload Optimization Manager representative may instruct you to update your deployment to increase the logging level from the default logging level of 2 to as high as 5. Wait for a full discovery to complete and, if possible, recreate the error before collecting logs.

1. Set the logging level.
   - YAML method
     a. Update the Kubeturbo `ConfigMap`.

     ```
     kubectl edit configmap turbo-config-kubeturbo -n turbo
     ```

     b. Update the `"level":` value from 2 to the temporary logging level. In the following example, the temporary logging level is 5.

     ```
     apiVersion: v1
     data:
       turbo-autoreload.config: |-
         {
           "logging": {
             "level": 5
           },
     ```

     The pod automatically discovers the updated logging level. There is no need to restart the pod.
   - Operator method
     a. Update the `kubeturbo-release` custom resource.

     ```
     kubectl edit kubeturbo kubeturbo-release -n turbo
     ```

     b. Add or update `logging:level` parameter. In the following example, the temporary logging level is 5.

     ```
     apiVersion: charts.helm.k8s.io/v1
     kind: Kubeturbo
     metadata:
       name: kubeturbo-release
     spec:
       logging:
         level: 5
     ...
     ```

   - Helm method

     Upgrade your Kubeturbo `release`.

     ```
     helm upgrade {helmChart} {chart location or repository} –namespace turbo --set
      logging.level={logging_level}
     ```

     Be sure to replace *`{logging_level}`* with the temporary logging level, such as 5.
2. Check the Kubeturbo `configMap` and verify that the logging level changed.

```
kubectl get configmap turbo-config-kubeturbo-release -o yaml
```

```
apiVersion: v1
data:
  turbo-autoreload.config: |-
    {
        "logging": {
           "level": 5
        }
    }
```

3. Restore the default logging level of 2 when you are done.

Changing the logging output

Kubeturbo sends its log stream to the standard error (`stderr`) file descriptor of the container, which is generally available through `kubectl logs` or to the logging pipeline if configured within the container platform cluster. To disable writing logs to the file system and to ensure that the logs are sent only to `stderr`, update the Kubeturbo command line parameters.

```
--logtostderr=true
--alsologtostderr=false
```

For operator-based deployments, update the `args:` section in the `spec` of the operator custom resource.

```
args:
  logtostderr: true
  alsologtostderr: false
```

Affinity and anti-affinity related logs

Kubeturbo can log details about pod-to-node and pod-to-pod affinities and anti-affinities. It can log statistical information, such as the number of pods with affinities and the number of pods with pod-to-pod affinities. At a higher log level, it can also log additional details, such as the list of pods that carry affinity rules, the exact rule strings, and topologies.

Logging level	Available information
3 or higher	▪ Total pods with all affinities and anti-affinities, including pod-to-node, pod-to-pod, and affinities carried from persistent volumes associated with pods ▪ Total pods with node affinities and anti-affinities ▪ Total pods with pod affinities and anti-affinities ▪ Total pods with volumes that specify node anti-affinities ▪ Total unique affinity terms by rule string across all pods ▪ Total unique anti-affinity terms by rule string across all pods ▪ Total unique label key=value pairs on pods ▪ Total unique label key=value pairs (topologies) on nodes
5 or higher	▪ List of all pods with affinities and anti-affinities ▪ List of Pods with node affinities and anti-affinities ▪ List of pods with pod-to-pod affinities and anti-affinities ▪ List of pods with volumes that specify node affinities and anti-affinities

Logging level	Available information
	▪ List of unique affinity terms by rule string ▪ List of unique anti-affinity terms by rule string ▪ List of unique label pairs on pods ▪ List of unique label pairs (topologies) on nodes

Kubeturbo ConfigMap values

The following table describes the values that you can specify in your Kubeturbo `configMap`.

Parameter	Description	Required?	Default value	Accepted values
`serverMeta.version`	Workload Optimization Manager version	No	None	`x.x`
`serverMeta.turboServer`	Workload Optimization Manager address (URL)	Yes, for all versions	None	`https://{your_instance}`
`serverMeta.proxy`	Workload Optimization Manager proxy address (URL)	No	None	`http://username:password@proxyserver.proxyport` or `http://proxyserver.proxyport`
`restAPIConfig.opsManagerUserName`	Workload Optimization Manager login username	No, optional and not recommended	None	`Plain text username`
`restAPIConfig.opsManagerPassword`	Workload Optimization Manager login password	No, optional and not recommended	None	`Plain text password`
`targetConfig.targetName`	Uniquely identifies container platform clusters	Yes, for all versions	`"Your_Cluster_Name"`	Upper or lower case, limited special characters "-" or "_"
`HANodeConfig.nodeRoles`	Automates policies to keep nodes of the same role limited to one ESX host or availability zone	Yes, for all versions	`node-role.kubernetes-.io/master`	Any value for label key `node-role.kubernetes.io`, values in quotes and comma-separated values, `"master"`, `"worker"`, `"app"`
`annotationWhitelist`	Provides a mechanism for discovering annotations for objects. To collect annotations, provide a regular expression for each entity type with annotations.	No	None	Regular expressions that accept the RE2 syntax (except for `\C`) as defined in this Google RE2 page
`logging.level` (in the `turbo-autoreload.config` section)	Changes the logging level without restarting pods	No	2	Integer value that is greater than 0

Fabric targets

A fabric target is a service that unites compute, network and storage access into a cohesive system. When you connect Workload Optimization Manager to fabric targets, it monitors the performance and resource consumption of your fabric interconnects, IO modules, chassis, and physical machines to assure application performance and utilize resources as efficiently as possible.

Once connected, Workload Optimization Manager discovers the blade servers that host the VMs, the chassis and datastores that provide resources to the blade servers, the IO modules and fabric interconnects that provide network resources, and the virtual datastores that provide storage resources to the VMs.

As part of this process, Workload Optimization Manager will stitch information from the fabric target and connected hypervisor targets to provide more granular data and information related to the applications and VMs running on the hypervisor-stitched blade servers. Combined with other targets, this information will support a hierarchical, application-driven approach to managing your environment.

For example:

When Workload Optimization Manager discovers that blade servers housed in a particular chassis have been designated as vCenter hosts, the supply chain stitches the blade servers and chassis to the corresponding vCenter data center to establish their relationship. When you set the scope to that data center and view the Health chart, you will see the blade servers in the list of hosts. In addition, when the data center is included in a merge policy (a policy that merges data centers for the purpose of VM placement), the VMs in the blade servers apply the policy, allowing them to move between data centers as necessary.

When you add application server targets, your applications and their individual components and services are discovered, enabling a view of your infrastructure from an individual application service to the physical hardware. Adding public cloud targets also allow for workloads to potentially migrate from your UCS infrastructure to the cloud, based on cost or available resources.

Supply chain

Fabric targets add IO Module, Fabric Interconnect, Domain, and Chassis entities to the supply chain. The Chassis entities host physical machines (blade servers) — The physical machines also consume network connection commodities from IO Modules. The Fabric Interconnect supplies connectivity to the overall network, and also hosts the UCS Manager for UCS Targets. The Domain serves as the bottom-level pool of network resource, supplying the Fabric Interconnect.

Cisco UCS Manager

The Cisco Unified Computing System (UCS) Manager is a management solution that participates in server, fabric, and storage provisioning, device discovery, inventory, configuration, diagnostics, monitoring, fault detection, auditing, and statistics collection.

UCS integrates all of these resources in a scalable multi-chassis platform to converge administration onto a single point. Managing these various entities on a network fabric with Workload Optimization Manager enables automation at the hardware level, including automated provisioning of hosts.

Prerequisites

- Cisco UCS Manager 3.1+
- A service account Workload Optimization Manager can use to connect to Cisco UCS Manager

Adding UCS targets

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **Cisco UCS Manager**.
4. In the side panel, review the connection requirements and then click **Connect Target**.

5. Configure the following settings:
   - Address

     Specify the IP address of the Cisco UCS Manager. Workload Optimization Manager connects to the Cisco UCS Manager using the HTTPS protocol by default.

     NOTE:
     By default, the IP address you specify in this field is used as the prefix for the display name of the physical machine (for example, `IPaddress:sys/rack-unit-4`, where `sys/rack-unit-4` is a distinguishing name for the physical machine). Alternatively, you can use the `name` or `usrlbl` attribute of the physical machine as the prefix for the display name by setting the `PMDisplayNamePrefix` probe property for the Cisco UCS Manager probe to `name` or `usrlbl` in an API editor.

   - Username

     Specify the username of the account Workload Optimization Manager uses to connect to the target.

     Specifying the IP address and credentials gives access to the Fabric Manager on the interconnect. Workload Optimization Manager discovers the fabric interfaces associated with that manager.

     NOTE:
     If the account is managed in Active Directory, include the case-sensitive domain name as part of the username. For example,`MyDomain\john` is not the same as `mydomain\john`. For local user accounts, provide the username.

   - Password

     Specify the password of the account Workload Optimization Manager uses to connect to the target.
   - Port number

     Specify the port Workload Optimization Manager uses to connect to the target. By default, the port is 443.
   - Connect using HTTPS

     If you select this option, Workload Optimization Manager connects to the target servers using HTTPS. Make sure that the required certificate is configured for use on the host.

     This option is selected by default.

Entity mapping

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

UCS	Workload Optimization Manager
Server / Blade / Rack Unit	Host
Chassis	Chassis
Datacenter	Datacenter
IO Module	IO Module
Fabric Interconnect	Switch
Network	Network

Fabric targets add IO Module, Fabric Interconnect (Switch), and Chassis entities to the supply chain. Hosts consume resources from Chassis entities, and network connection commodities from IO Modules. The Fabric Interconnect supplies connectivity to the overall network, and also hosts the UCS Manager. The Domain serves as the bottom-level pool of network resource, supplying the Fabric Interconnect. Be sure that all the FC, Ether, and Physical Ports are properly configured with suitable roles in UCS so that the supply chain is populated accurately without disjoints.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Host
  - Power

    Power is the measurement of electricity consumed by a given entity, expressed in watts.
  - Memory (Mem)

    Memory is the measurement of memory that is reserved or in use.
  - CPU

    CPU is the measurement of CPU that is reserved or in use.
  - IO

    IO is the utilization of a host's IO adapters.
  - Net

    Net is the utilization of data through the host's network adapters.
  - Swap

    Swap is the measurement of a host's swap space that is in use.
  - Balloon

    Balloon is the measurement of memory that is shared by VMs running on a host.

    This commodity applies to ESX only.
  - CPU ready

    CPU ready is the measurement of a host's ready queue capacity that is in use.

    This commodity applies to ESX only.
- Chassis
  - Power

    Power is the measurement of electricity consumed by a given entity, expressed in watts.
  - Cooling

    Cooling is the percentage of the acceptable temperature range that is utilized by the entity. As the temperature nears the high or low running temperature limits, this percentage increases.
- I/O Module
  - Net throughput

    Net throughput is the rate of message delivery over a port.
- Switch
  - Net throughput

    Net throughput is the rate of message delivery over a port.
  - PortChannel

    `PortChannel` is the amalgamation of ports with a shared net throughput and utilization.

Actions

Workload Optimization Manager supports the following actions:

- Host
  - Start
  - Provision
  - Suspend

- Chassis
  - Provision
- Switch
  - Add Port to Port Channel
  - Remove Port from Port Channel
  - Add Port

UCS blade provisioning

When managing a UCS fabric target, Workload Optimization Manager can provision any blade servers that are installed in a chassis but not currently in operation. If the workload demands more physical compute resources, Workload Optimization Manager can automatically direct UCS to provision a blade, or it can recommend that you provision a blade and you can execute the action from the To Do list. To enable this capability, you must perform two basic steps:

- Configure the way UCS and vCenter Server manage information as blades are provisioned

  To enable Workload Optimization Manager to perform automatic provisioning of UCS blades, you must configure UCS to work with vCenter Server as they both manage resources such as server pools, policies, and Service Profile Templates. This ensures that as Workload Optimization Manager directs the UCS Manager to provision a new blade, vCenter Server will recognize that the new physical host is available. Once vCenter Server can recognize the new blade, Workload Optimization Manager can direct vCenter Server to move workloads onto the server.

  Workload Optimization Manager provisions new blades based on the service profiles of operating blades. To enable this, the configuration must include Service Profile Templates, and the operating blades must be bound to these templates.

- Set the Host Provision action to Automate or Manual for the blade servers. By default, Workload Optimization Manager sets the Host Provision action to Recommend.

  For any hosts other than blade servers managed by UCS, Workload Optimization Manager cannot provision hosts automatically or manually. Instead, it recommends that you provision a host, and you then install the physical machine and wire it to the network.

  In a UCS deployment you can have blade servers installed in the chassis and ready to provision. In that case, Workload Optimization Manager can direct UCS to provision a new blade to meet workload demands. For these servers, you can set the Host Provision action to Automatic or Manual.

  NOTE:
  It's important that you only set Automatic or Manual host provisioning to UCS blades. If you set Host Provision to Automatic for other types of hosts, Workload Optimization Manager will attempt to perform the action and fail. As a result, you might never see the recommendation to provision a new host of that type.

  Workload Optimization Manager groups blade servers by chassis. To restrict Automatic or Manual settings to blade servers, use this group. You can set the action acceptance mode for all blade servers in your environment, or you can set the mode differently for individual chassis.

HPE OneView

HPE OneView is a management solution that streamlines provisioning and lifecycle management across compute, storage, and fabric. Through a unified API, infrastructure can be configured, monitored, updated, and re-purposed.

HPE OneView integrates all of these resources in a scalable multi-enclosure platform to converge administration onto a single point. Managing these various entities on a network fabric with Workload Optimization Manager enables automation at the hardware level, including automated provisioning of hosts.

Prerequisites

- HPE OneView 3.00.04
- A service account Workload Optimization Manager can use to connect to HPE OneView
- The **Banner Page** and **Require Acknowledgment** options for the user account should be disabled in the HPE OneView user interface.

Adding HPE OneView targets

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **HPE OneView**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
    - Address

      Specify the IP address of the HPE OneView target. Workload Optimization Manager uses the HTTPS protocol by default. To force the HTTP protocol, specify the address as `http://8.8.8.8` or `8.8.8.8:80`.

      This gives access to the Fabric Manager that resides on the VM.
    - Username

      Specify the username of the account Workload Optimization Manager uses to connect to the target.

      Specify the IP address and credentials for HPE OneView. Workload Optimization Manager discovers the fabric interfaces associated with that instance.

      NOTE:
      If the account is managed in Active Directory, include the case-sensitive domain name as part of the username. For example, `MyDomain@john` is not the same as `mydomain@john`. For local user accounts, just provide the username.

    - Password

      Specify the password of the account Workload Optimization Manager uses to connect to the target.

Entity mapping

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

HPE OneView	Workload Optimization Manager
IO Module	IO Module
Fabric Interconnect	Switch
Domain	Domain
Chassis	Physical Machines

Fabric targets add IO Module, Fabric Interconnect (Switch), Domain, and Chassis entities to the supply chain. The Chassis entities host physical machines – The physical machines also consume network connection commodities from IO Modules. The Fabric Interconnect supplies connectivity to the overall network. The Domain serves as the bottom-level pool of network resource, supplying the Fabric Interconnect.

NOTE:
For HPE OneView targets, the "Fabric Interconnect" entity exists as a false "Switch", and only as a pass-through for network resources. Unlike other fabric targets, such as UCS, there is no physical hardware that serves this function.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Virtual machine

  NOTE:
  Workload Optimization Manager only monitors VM resources if HPE OneView is stitched to a hypervisor in the supply chain.

  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.
  - Virtual storage

    Virtual ssorage is the measurement of virtual storage capacity that is in use.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - Latency

    Latency is the measurement of storage latency.
- Host
  - Power

    Power is the measurement of electricity consumed by a given entity, expressed in watts.
  - Memory (Mem)

    Memory is the measurement of memory that is reserved or in use.
  - CPU

    CPU is the measurement of CPU that is reserved or in use.
  - IO

    IO is the utilization of a host's IO adapters.
  - Net

    Net is the utilization of data through the host's network adapters.
  - Swap

    Swap is the measurement of a host's swap space that is in use.
  - Balloon

    Balloon is the measurement of memory that is shared by VMs running on a host.
  - CPU ready

    CPU ready is the measurement of a host's ready queue capacity that is in use.
- Chassis
  - Power

    Power is the measurement of electricity consumed by a given entity, expressed in watts.
  - Cooling

    Cooling is the percentage of the acceptable temperature range that is utilized by the entity. As the temperature nears the high or low running temperature limits, this percentage increases.

- Storage
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

    NOTE:
    When it generates actions, Workload Optimization Manager does not consider IOPS throttling that it discovers on storage entities. Analysis uses the IOPS it discovers on Logical Pool or Disk Array entities.
  - Latency

    Latency is the measurement of storage latency.
- I/O Module
  - Net throughput

    Net throughput is the rate of message delivery over a port.
- Switch
  - Net throughput

    Net throughput is the rate of message delivery over a port.
  - PortChannel

    `PortChannel` is the amalgamation of ports with a shared net throughput and utilization.

Actions

Workload Optimization Manager supports the following actions:

- Virtual machine
  - Provision additional resources (VMem, VCPU)
  - Move Virtual Machine
  - Move Virtual Machine Storage
  - Reconfigure Storage
  - Reconfigure Virtual Machine
  - Suspend VM
  - Provision VM
- Host
  - Start
  - Provision
  - Suspend
- Switch
  - Add Port to Port Channel
  - Remove Port from Port Channel
  - Add Port

Guest OS processes targets

Using WMI (Windows Management Instrumentation) or SNMP, Workload Optimization Manager can discover OS processes that are running on your VMs. Analysis creates an Application Component entity for each process.

Workload Optimization Manager does not generate actions for these discovered Application Components. For many types of applications, you can a target dedicated to that application. You should use Guest OS Process targets for cases where there is no other target you can use for the application. For more information, see Applications and Databases Targets *(on page 180)*.

WMI

Workload Optimization Manager discovers application and operating system resources using WMI (Windows Management Instrumentation) in your entire environment or a portion of your environment.

Prerequisites

- WMI: Windows 7, 8, 8.1, 10, 2008 R2, 2012, 2012 R2, 2016, or 2019
- Underlying VM host targets added to Workload Optimization Manager (for discovery)
- A WMI user account *(on page 319)* (administrator or non-administrator belonging to certain local user groups)
- WMI enabled on target VMs *(on page 320)*

NOTE:
For Hyper-V hosts, you must install Hyper-V Integration Services on the target VMs. For more information, refer to the following integration services TechNet article: https://technet.microsoft.com/en-us/library/dn798297(v=ws.11).aspx.

For VMware hosts, you must install VMware Tools on the target VMs.

Adding WMI targets

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **WMI**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
    - Display name

      Specify a name that uniquely identifies this connection.

      This name is for display purposes only and does not need to match any name in WMI.
    - Username

      Specify the username of the account Workload Optimization Manager uses to connect to the WinRM service on the Virtual Machine hosting the application. This should not include the Active Directory domain.
    - Password

      Specify the password of the account Workload Optimization Manager uses to connect to the WinRM service on the Virtual Machine hosting the application.
    - Scope

      Specify the scope Workload Optimization Manager uses for application discovery.

      If you set a scope, Workload Optimization Manager searches for Windows applications on the virtual machines found within the set scope. This scope can be set to the entire environment, clusters, or virtual machines.

      Workload Optimization Manager can monitor up to 500 virtual machines in a group. If you have more than 500 virtual machines in your environment, split them across smaller groups and then add those groups as individual targets.

- Active Directory domain

  Specify the Active Directory domain used by Workload Optimization Manager in conjunction with the username for authentication. Leave blank for local accounts.

- Connect using HTTPS

  If you select this option, Workload Optimization Manager connects to the target servers using HTTPS. Make sure that the required certificate is configured for use on the host.

- Use NTLM for domain accounts

  If you select this option, Workload Optimization Manager uses the NTLM protocol instead of the Kerberos protocol for domain accounts. NTLM is always used for local accounts.

  This option is selected by default.

- Authenticate all database servers

  If you select this option, Workload Optimization Manager attempts to authenticate all database servers in the selected scope. If Workload Optimization Manager is unable to authenticate a database server, the target is not added and no data is collected.

Enabling guest load metrics

NOTE:
For environments that use WMI or SNMP probes, you must configure the Workload Optimization Manager installation to enable guest load metrics. Without this configuration, the WMI and SNMP probes will not display discovered guest load data.

To enable guest load metrics, you must edit the CR file and restart the platform:

1. Open the cr.yaml file for editing.

   ```
   /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

2. Enable guest metrics in the `properties` section of the file:

   ```
   properties:
     repository:
       showGuestLoad: true
   ```

3. Save your changes to the file.
4. Apply your changes.

   Run the command:

   ```
   kubectl apply -f /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

5. To restart the component, delete the repository pod.

   Run the command (where **{POD_ID}** is the ID that is appended to the pod in the `get pods` listing):

   ```
   kubectl delete pod -n cwom repository-{POD_ID}
   ```

Monitored resources

Workload Optimization Manager monitors the following resources:

- WMI/SNMP Application
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.

  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.

- Virtual Machine
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
  - Virtual storage

    Virtual ssorage is the measurement of virtual storage capacity that is in use.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - Latency

    Latency is the measurement of storage latency.

Actions

NOTE:
The specific actions that Workload Optimization Manager recommends can differ, depending on the processes that Workload Optimization Manager discovers.

For other application components, Workload Optimization Manager can recommend actions based on the resources it can discover for the application. For example, Node.js® applications report CPU usage, so Workload Optimization Manager can generate vCPU resize actions and display them in the user interface.

Workload Optimization Manager supports the following actions:

- Virtual Machine
  - Provision additional resources (VMem, VCPU)
  - Move Virtual Machine
  - Move Virtual Machine Storage
  - Reconfigure Storage
  - Reconfigure Virtual Machine
  - Suspend VM
  - Provision VM

NOTE:
Without separate targets to discover Guest OS Processes or Application Servers, Workload Optimization Manager does not generate actions on applications. Instead, it generates resize actions on the host VMs. For on-prem environments, if host utilization is high enough on the host running the application VM, Workload Optimization Manager can also recommend provisioning a new host.

Creating a WMI user account

The service account Workload Optimization Manager uses to connect to WMI should be an Active Directory domain account.

Some enterprises require that the service account does not grant full administrator rights. In that case, you can create a restricted service account for WMI:

1. Add the account to each of the following local groups:
   - `WinRMRemoteWMIUsers__` (or `Remote Management Users`)
   - `Performance Monitor Users`

   NOTE:
   These groups are standard Windows Server 2012 security groups. If you are using an earlier version of Windows Server and do not see these groups, contact Cisco Support for assistance.

2. Grant permissions to the account.

   In the WMI Management console, grant the `Enable Account` and `Remote Enable` advanced security permissions to the service account:
   - Open the WMI Management console (`wmimgmt`).
   - Right-click **WMI Control (Local)** and choose **Properties**.
   - Go to the **Security** tab and then click **Security** to display the **Security for Root** dialog.
   - Click **Advanced**, select the service account, and click **Edit**.
   - Confirm that **This namespace and subnamespace** is selected.
   - Select **Enable Account** and **Remote Enable** and click **OK**.

Enabling WMI

To enable Workload Optimization Manager discovery of Windows-based Guest Processes, you must configure and enable WMI with WinRM, usually by using a group policy for AD Domains.

When local credentials are provided to Workload Optimization Manager, NTLM is the authentication mechanism. When domain credentials are provided, either NTLM or Kerberos can be specified.

NOTE:
If Kerberos is specified, the IP address of the target must be resolvable to a hostname using DNS.

While these steps will change slightly between Windows Server versions, the general instructions are:

1. Open an elevated PowerShell prompt on the virtual machines to be discovered as a WMI Target.
2. Enable the WS-Management protocol and set the default configuration for remote management.

   Run the command:

   ```
   winrm quickconfig
   ```

3. Enable the WinRM *Negotiate* authentication scheme.

   This policy is enabled by default. To enable this policy if it is disabled, run the command:

   ```
   winrm set winrm/config/service/auth '@{Negotiate="true"}'
   ```

4. Set the WinRM `AllowUnencrypted` property to `true` if non-SSL connections are preferred.

   This property must be set on both the server and the client. Note that setting this value to `true` does not mean that WMI passes sensitive data in an unencrypted form. It will send only the content of the SOAP messages as plain text:
   - Server setting:

     ```
     winrm set winrm/config/service '@{AllowUnencrypted="true"}'
     ```

- Client setting:

```
winrm set winrm/config/client '@{AllowUnencrypted="true"}'
```

5. Restart the Remote Registry service.

After restarting the Remote Registry service, Workload Optimization Manager will discover the WMI targets.

SSL connections

To connect using SSL, then a certificate must be assigned to WinRM, and an SSL listening socket enabled. Workload Optimization Manager will accept self-signed certificates. Complete details can be found in the official WinRM documentation.

Some example SSL-related commands:

- List certificates on the system.

```
ls cert://localmachine/my
```

- Find Thumbprint for the default certificate.

```
7B56C33F029E7605D5C58E5597D0A077FE1D7F1C CN=winsql-server1.corp.mycorp.com
```

- Enable SSL listener.

```
winrm create winrm/config/listener?Address=*+Transport=HTTPS
 @{Hostname="winsql-server1.corp.mycorp.com";CertificateThumbprint="7B56C33F029E7605D5C58E5597D0A07
7FE1D7F1C";Port="5986"}
```

SNMP

NOTE:
After reviewing telemetry data and conducting interviews with customers to determine use of this capability, this target was deemed appropriate to deprecate. If you replaced this technology with something that is currently not supported, submit an idea at the Cloud Management and AIOps Idea Portal. The last Workload Optimization Manager version that will contain this target is 3.13.0.

Workload Optimization Manager will discover application and operating system resources using SNMP (Simple Network Management Protocol) on Linux hosts within the specified scope of your environment.

Prerequisites

- Underlying VM host targets added to Workload Optimization Manager (for discovery)
- SNMP service enabled/configured on target VMs to allow access to the OIDs required. The credentials used must allow access to the following OIDs:
  - 1.3.6.1.2.1.1.1 - sysDescr
  - 1.3.6.1.2.1.25.4.2.1.2 - hrSWRunName
  - 1.3.6.1.2.1.25.5.1.1 - hrSWRunPerfEntry
  - 1.3.6.1.2.1.25.3.3.1.2 - hrProcessorLoad
  - 1.3.6.1.4.1.2021.4.5 - memTotalReal
  - 1.3.6.1.4.1.2021.4.6 - memAvailReal
  - 1.3.6.1.4.1.2021.4.11 - memTotalFree
  - 1.3.6.1.4.1.2021.4.14 - memBuffer
  - 1.3.6.1.4.1.2021.4.15 - memCached

NOTE:
For environments that include SNMP targets, Workload Optimization Manager can discover incorrect memory values for Linux systems. This can occur for the systems that use the net-snmp package, version 5.7.2-43.el7. You should use versions less than 5.7.2-43.el7, or greater than or equal to 5.7.2-47.el7.

For Hyper-V hosts, you must install Hyper-V Integration Services on the target VMs. For more information, refer to the following integration services TechNet article:

```
https://technet.microsoft.com/en-us/library/dn798297%28v=ws.11%29.aspx
```

For VMware hosts, you must install VMware Tools on the target VMs.

Adding SNMP targets

1. Click **Settings > Target Configuration**.
2. Click **New Target > Guest OS Processes**.
3. Select **SNMP**.
4. Configure the following settings:
    - Target Name

      Specify a name that uniquely identifies this connection.

      This name is for display purposes only and does not need to match any name in SNMP.
    - Scope

      Specify the scope Workload Optimization Manager uses for application discovery.

      Workload Optimization Manager searches for SNMP applications on the Virtual Machines found within the set scope. This scope can be set to the entire environment, single or multiple clusters, or particular Virtual Machines.

      NOTE:
      A single scope (target) must not contain more than 500 Virtual Machines. Workload Optimization Manager recommends utilization of multiple SNMP targets for environments exceeding this limit.

    - Community

      Specify the SNMP v2c community string Workload Optimization Manager uses to connect to the SNMP agent on the VM.
    - Port number

      Specify the port number Workload Optimization Manager uses to connect to the SNMP agent on the VM.
    - Full Validation

      If you select this option, Workload Optimization Manager requires all database servers hosted on the VMs in the selected scope to be a valid target. If Workload Optimization Manager is unable to authenticate a database server in the scope, the target does not validate and data is not collected.
    - Enable SNMPv3

      If you select this option, Workload Optimization Manager uses SNMPv3 to connect to the virtual machines in the selected scope.
    - SNMPv3 Username/SecurityName

      Specify the username/SecurityName that Workload Optimization Manager uses to connect to the Virtual Machines hosting the application.
    - SNMPv3 Enable Privacy

      If you select this option, Workload Optimization Manager encrypts using the privacy password to keep the connection private.
    - SNMPv3 Authentication Password

Specify the Authentication Password Workload Optimization Manager uses to connect to the virtual machines in the selected scope. This allows requests to be authenticated, confirming the sender's identity.

- SNMPv3 Privacy Password

  Specify the Privacy Password Workload Optimization Manager uses to connect to the Virtual Machines in the selected scope. For requests to be encrypted, the SNMP manager and the SNMP agent must share knowledge of the privacy password associated with the username.

- SHA-1 For Authentication

  If you select this option, Workload Optimization Manager uses SHA-1 Authentication to connect to the virtual machines in the selected scope. When unchecked, Workload Optimization Manager uses MD5

- AES For Privacy

  If you select this option, Workload Optimization Manager uses AES 128 to connect to the virtual machines in the selected scope; otherwise, Workload Optimization Manager uses DES.

NOTE:
Workload Optimization Manager does not support AES 192 or 256 for SNMP targets.

Enabling guest load metrics

NOTE:
For environments that use WMI or SNMP probes, you must configure the Workload Optimization Manager installation to enable guest load metrics. Without this configuration, the WMI and SNMP probes will not display discovered guest load data.

To enable guest load metrics, you must edit the CR file and restart the platform:

1. Open the cr.yaml file for editing.

   ```
   /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

2. Enable guest metrics in the `properties` section of the file:

   ```
   properties:
     repository:
       showGuestLoad: true
   ```

3. Save your changes to the file.
4. Apply your changes.

   Run the command:

   ```
   kubectl apply -f /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

5. To restart the component, delete the repository pod.

   Run the command (where **{POD_ID}** is the ID that is appended to the pod in the `get pods` listing):

   ```
   kubectl delete pod -n cwom repository-{POD_ID}
   ```

Monitored resources

Workload Optimization Manager monitors the following resources:

- WMI/SNMP Application
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
- Virtual Machine
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
  - Virtual storage

    Virtual ssorage is the measurement of virtual storage capacity that is in use.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - Latency

    Latency is the measurement of storage latency.

Actions

NOTE:
The specific actions that Workload Optimization Manager recommends can differ, depending on the processes that Workload Optimization Manager discovers.

For other application components, Workload Optimization Manager can recommend actions based on the resources it can discover for the application. For example, Node.js® applications report CPU usage, so Workload Optimization Manager can generate vCPU resize actions and display them in the user interface.

Workload Optimization Manager supports the following actions:

- Virtual Machine
  - Provision additional resources (VMem, VCPU)
  - Move Virtual Machine
  - Move Virtual Machine Storage
  - Reconfigure Storage
  - Reconfigure Virtual Machine
  - Suspend VM
  - Provision VM

NOTE:
Without separate targets to discover Guest OS Processes or Application Servers, Workload Optimization Manager does not generate actions on applications. Instead, it generates resize actions on the host VMs. For on-prem environments, if host utilization is high enough on the host running the application VM, Workload Optimization Manager can also recommend provisioning a new host.

Enabling SNMP

Workload Optimization Manager requires that SNMP is enabled and configured in order to discover Guest Processes. While these steps will change slightly between OS versions, this topic gives you the general instructions:

Enabling SNMP v2

1. Obtain and install the SNMP server package for your Linux distribution and version. This is commonly called `net-snmp` in most package managers. Also install the `net-snmp-utils` package if it is available.
2. Configure the SNMP daemon by editing `/etc/snmp/snmpd.conf` or running `snmpconf -i` at the command line in order to verify that:
   - SNMP v2c is enabled.
   - A read-only community name has been set. This community name will be used by Workload Optimization Manager to communicate with the SNMP agent on the VM.

     NOTE:
     The community name you select must have at least read-only access to all OIDs in the system.

3. Configure the daemon to listen on a public interface. Most default installations only listen on `127.0.0.1`.
4. Start the SNMP daemon process.

These long form and short form sample `snmp.conf` files illustrate the necessary changes.

Long form, VACM:

```
# First, map the community name "mycommunity1" into a "security name"
# sec.name  source          community
    com2sec notConfigUser  default      mycommunity1
# Second, map the security name into a group name:
# groupName securityModel securityName
    group   notConfigGroup v2c           notConfigUser
# Finally, grant the group read-only access to any UUID.
# group context sec.model sec.level prefix read   write  notif
    access  notConfigGroup ""      any       noauth
    exact  all none none
```

Short form, older syntax:

```
rocommunity mycommunity1 default system
```

Verify Your SNMP Setup

Verify that your SNMP setup is successful by running the following command from a remote machine, replacing the community string and IP address:

```
snmpwalk -Os -v 2c -c COMMUNITY_STRING IP_ADDRESS iso.3.6.1.2.1.1.1
```

If successful, the command returns the kernel version of the machine (similar to the output of `uname -a`).

Enabling SNMP v3

1. Obtain and install the SNMP server package for your Linux distribution and version.

   Most package managers call this package `net-snmp`. To verify your configuration as mentioned in the final step, you should also install the `net-snmp-utils` package on the Workload Optimization Manager VM.
2. Stop the SNMP service.
3. Create the SNMPv3 user.

```
net-snmp-config --create-snmpv3-user [-ro] [-A authpass] [-X privpass] [-a MD5|SHA] [-x DES|AES]
 [username]
```

For example, the command:

```
i. net-snmp-create-v3-user -ro -A snmpv3authPass -a SHA -X snmpv3encPass -x AES snmpv3user
```

Results in output similar to the following example:

```
adding the following line to /var/lib/net-snmp/snmpd.conf:
createUser snmpv3user SHA "snmpv3authPass" AES snmpv3encPass adding the following line
 to /etc/snmp/snmpd.conf:
rouser snmpv3user
```

4. Verify your SNMP setup.

   To verify the configuration, run the following command on the Workload Optimization Manager VM:

```
snmpwalk -Os -v 2c -c COMMUNITY_STRING IP_ADDRESS iso.3.6.1.2.1.1.1
```

Hyperconverged targets

A hyper converged target is a service that unites compute, network and storage access into a cohesive system. When you connect Workload Optimization Manager to hyper converged targets, it will monitor the performance and resource consumption of your hyper converged infrastructure to maintain application performance while utilizing resources as efficiently as possible.

As part of this process, Workload Optimization Manager will stitch information from the hyper converged target to the associated hypervisor and fabric targets, supporting Application Resource Management (ARM) and providing deeper insight into the state of the hardware and information related to the entities in the supply chain. Combined with application server targets, this information will support a hierarchical, application-driven approach to managing your environment.

Cisco HyperFlex

NOTE:
After reviewing telemetry data and conducting interviews with customers to determine use of this capability, this target was deemed appropriate to deprecate. If you replaced this technology with something that is currently not supported, submit an idea at the Cloud Management and AIOps Idea Portal. The last Workload Optimization Manager version that will contain this target is 3.13.0.

Cisco HyperFlex provides a hyperconverged platform that combines the networking and compute power of UCS with the storage capabilities of the HyperFlex HX Data Platform.

With the additional and refined storage information provided by HyperFlex, Workload Optimization Manager narrows the Desired State and recommends actions using the joint compute and storage information, gaining valuable insight into the interconnected nature of your environment.

For Workload Optimization Manager to make the most informed decisions, you must also add the compute nodes and any associated hypervisors. HyperFlex environments typically include:

- Converged (HX) Nodes

  A combination of the cluster's storage devices into a single multi-tiered, object-based datastore.

- Compute Nodes

  Cisco B or C series servers that make up the compute resources of the cluster, and are typically managed by a hypervisor.

- Controller VMs

  Each HyperFlex node includes a Controller VM that intercepts and handles all the I/O from associated virtual machines. Workload Optimization Manager will not recommend actions for these VMs.

Prerequisites

- Cisco HyperFlex 3.5
- A service account on the controller VM that Workload Optimization Manager can use to connect to the HyperFlex environment

Adding HyperFlex targets

1. Click **Settings > Target Configuration**.
2. Click **New Target > Hyperconverged**.
3. Select **Cisco HyperFlex**.
4. Configure the following settings:
   - Address

     The name or IP address of the HyperFlex controller VM. If a port is appended to the IP address, it must be a secure (HTTPS) port.
   - Username/Password

     Credentials for a user account Workload Optimization Manager will use to connect to the controller VM.

     To specify a root user for the target, give the name as `local/root`. Note that earlier versions of Workload Optimization Manager accepted a username of `root`. If an existing HyperFlex target fails to validate after you have upgraded Workload Optimization Manager, then check to make sure the username for a root account is properly set to `local/root`.

     If you use vSphere SSO for authentication to the HyperFlex target, you must prepend your username with `vc-`. For example, if your username is `Administrator@vsphere.local`, you would enter it as `vc-Administrator@vsphere.local`.
   - Port Number

     Port Number Workload Optimization Manager will use to connect to the controller VM.

Entity mapping

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

HyperFlex	Workload Optimization Manager
Volume	Storage
HX Cluster	Disk Array

HyperFlex targets add Disk Array entities to the supply chain, and receive more granular information from the compute resources in your environment.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Storage
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.

  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

    NOTE:
    When it generates actions, Workload Optimization Manager does not consider IOPS throttling that it discovers on storage entities. Analysis uses the IOPS it discovers on Logical Pool or Disk Array entities.

  - Latency

    Latency is the measurement of storage latency.

- Disk Array
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - Latency

    Latency is the measurement of storage latency.

Actions

Workload Optimization Manager supports the following actions:

- Storage
  - Move

    This action can only be executed outside Workload Optimization Manager.
  - Provision

    This action can only be executed outside Workload Optimization Manager.
  - Resize Up

    This action can only be executed outside Workload Optimization Manager.
- Disk Array
  - Provision

    This action can only be executed outside Workload Optimization Manager.
  - Suspend

    This action can only be executed outside Workload Optimization Manager.
  - Resize Up

    This action can only be executed outside Workload Optimization Manager.

NOTE:

For this target, Workload Optimization Manager discovers the HX Cluster as a Disk Array. When you see a provision action on this entity, you should determine which of the following is most relevant, based on your environment:

- Add disks to converged nodes
- Add a new converged node
- Add a new HX Cluster

Nutanix

Nutanix products provide hyperconverged platforms that include VM hosting and a distributed storage fabric. The platform presents storage in two tiers — Local HDD storage and server-attached flash (hot storage).

Nutanix environments may include:

- One or more Nutanix appliances

  An appliance contains up to four server nodes.

- Nutanix nodes

  Servers that expose compute and storage resources. Each node provides local HDD and hot storage. Nodes combine to form a unified cluster that pools resources.

- Controller VMs

  Each node includes a Controller VM that manages the node's resources within the cluster pool. To minimize storage latency, the Controller VM keeps the most frequently accessed data in the hot storage.

Workload Optimization Manager supports management of Nutanix fabrics, where the supply chain treats a Nutanix Storage Pool as a disk array. Workload Optimization Manager recognizes Nutanix storage tiers when calculating placement of VMs and VStorage. In addition, Workload Optimization Manager can recommend actions to scale flash capacity up or down by adding more hosts to the cluster, or more flash drives to the hosts.

To specify a Nutanix target, provide the Cluster External IP address. This is a logical IP address that always connects to one of the active Controller VMs in the cluster. In this way, you can specify a Nutanix target without having to specify an explicit Controller VM.

NOTE:
The Controller VM must remain *pinned* to its host machine — You must not move the Controller VM to a different host. If the Nutanix cluster uses the Nutanix Acropolis OS to manage VMs, Workload Optimization Manager automatically pins the Controller VMs. However, if you use vCenter Server or Hyper-V to manage VMs on the hosts, you must configure a group to pin the Controller VMs. For more information, see Pinning Nutanix Controller VMs *(on page 332)*.

Prerequisites

- Nutanix Community Edition
- A service account with cluster administrator rights on the Nutanix clusters for action execution. For entity discovery, a minimum of READ access is required

Finding the cluster external IP address

To configure a Nutanix target, provide the Cluster External IP address for the given Nutanix cluster.

The Cluster External IP address is a logical IP that resolves to the cluster's Prism Element Leader. If the Prism Element Leader fails, then the Cluster External IP address will resolve to the newly elected Prism Element Leader.

To find this IP address, open the Web Console (the Prism Element) on the cluster and navigate to the **Cluster Details** view. In this view you can see the **Cluster External IP** address. If there is no IP address specified, you can specify the address at this time. For more information, see the Nutanix documentation.

Operating modes

A Nutanix node is a server that hosts VMs — In this sense the node functions as a hypervisor. A cluster of nodes can host VMs using the following Hypervisor technologies:

- Nutanix Acropolis

  The native Nutanix host platform, which combines software-defined storage with built-in virtualization.

- VMware ESXi
- Microsoft Hyper-V

Controller VM pinning

Each Nutanix node hosts a Controller VM that runs the Nutanix software and manages I/O for the hypervisor and all VMs running on the host. Each Controller VM must remain on its host node –The Controller VM must be *pinned* to that host, and must not be moved to any other host.

For more information about how to pin the Controller VM, see Pinning Nutanix Controller VMs *(on page 332)*.

Adding Nutanix targets

NOTE:
This describes how to add a Nutanix cluster to Workload Optimization Manager as a target. If Nutanix is not managing hosts running Acropolis as the hypervisor, you will have to add the vCenter or Hyper-V hypervisors as targets after you have added the Nutanix cluster as a target. For more information, see Hypervisor Targets *(on page 333)*.

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **Nutanix**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Cluster external IP address

     Specify the Cluster External IP address for the Nutanix cluster.
   - Username

     Specify the username of the account Workload Optimization Manager uses to connect to the target.
   - Password

     Specify the password of the account Workload Optimization Manager uses to connect to the target.
   - Port number

     Specify the listening port of the cluster. By default, the port is 9440.
   - Connect using HTTPS

     If you select this option, Workload Optimization Manager connects to the target servers using HTTPS. Make sure that the required certificate is configured for use on the host.

     This option is selected by default.

Entity mapping

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

Nutanix	Workload Optimization Manager
Virtual Disk	Volume
Container	Storage
Storage Pool	Disk Array
Nutanix Cluster	Storage Controller

Monitored resources

Workload Optimization Manager monitors the following resources:

- Volume
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Latency

    Latency is the measurement of storage latency.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

    NOTE:
    When it generates actions, Workload Optimization Manager does not consider IOPS throttling that it discovers on storage entities. Analysis uses the IOPS it discovers on Logical Pool or Disk Array entities.
- Storage
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

    NOTE:
    When it generates actions, Workload Optimization Manager does not consider IOPS throttling that it discovers on storage entities. Analysis uses the IOPS it discovers on Logical Pool or Disk Array entities.
  - Latency

    Latency is the measurement of storage latency.
- Disk Array
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - Latency

    Latency is the measurement of storage latency.
- Storage Controller

NOTE:
Not all targets of the same type provide all possible commodities. For example, some storage controllers do not expose CPU activity. When a metric is not collected, its chart in the user interface will not display data.

  - CPU

    CPU is the measurement of CPU that is reserved or in use.
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.

    The storage allocated to a storage controller is the total of all the physical space available to aggregates managed by that storage controller.

Actions

Workload Optimization Manager supports the following actions:

- Virtual Machine (Nutanix VM)
  - Move

    VMotion to hosts can be automated, but storage moves on Nutanix can only be executed outside Workload Optimization Manager.
  - Resize

    Resize actions require the VM to power down, and power back on again.
- Datastore (Storage)
  - Provision
  - Resize Up/Down
  - Suspend
  - Move

    This action can only be executed outside Workload Optimization Manager.
- Storage Controller
  - Provision

    This action can only be executed outside Workload Optimization Manager.

Pinning Nutanix Controller VMs

Each Nutanix node hosts a Controller VM that runs the Nutanix software and manages I/O for the hypervisor and all VMs running on the host. Each Controller VM must remain on its host node —The Controller VM must be *pinned* to that host, and must not be moved to any other host.

For a cluster using vCenter or Hyper-V hypervisors, you must use Workload Optimization Manager policies to pin the Controller VMs to their respective nodes. To do this, you will create a dynamic group of Nutanix Controller VMs, and then disable move actions for all members of this group.

To pin the Controller VMs:

1. Create a group of Controller VMs.

   In Workload Optimization Manager you can create dynamic groups based on VM name — All VMs with matching names automatically belong to the group. Nutanix uses the following naming convention for Control VMs:

   `NTNX-<SerialNumber>-A-CVM`, where `<SerialNumber>` is the serial number of the Controller VM.

You can create a dynamic group that automatically includes these Nutanix controller VMs. (For complete instructions on creating groups, see Creating Groups *(on page 1039)*.)

- Create a new group.

  In Workload Optimization Manager navigate to **Settings > Groups** and create a new group.
- Set the group type to **Dynamic**.
- Add a filter to match VMs by their names.

  Add a filter that uses the regular expression, `NTNX.*CVM`. This regular expression will match the Nutanix Controller VMs.

Be sure to save the group. All the Nutanix Controller VMs will automatically become members of this group.

2. Disable moves for all VMs in this group.

   To do this, create an automation policy for the group and disable actions. (For complete instructions to create these policies, see Creating Automation Policies *(on page 1053)*).

   - In Workload Optimization Manager go to the **Policy > Action > VM** view.
   - Set the scope to the group you made.

     In the **Scope** column, expand **My Groups** and select the group you just made.
   - Disable moves for this group.

     In the **Parameter** column under **Action Acceptance**, set the value to **Disabled**. This will override the global action mode.
   - Save the action acceptance mode.

     Be sure to click **Apply Settings Change**.

Hypervisor targets

A hypervisor is a service that creates and runs virtual machines (VMs), containers, or both, and provides these entities compute and storage resources. When you connect Workload Optimization Manager to hypervisor targets in your environment, Workload Optimization Manager assures application performance by using these resources as efficiently as possible.

After connected to a hypervisor target, Workload Optimization Manager discovers the VMs, containers, or physical machines that host the VMs or containers, data stores that provide storage resources to the physical machines, and virtual data stores that provide storage resources.

As more targets are added, Workload Optimization Manager discovers the resources belonging to your physical and virtual infrastructure. For example, adding the underlying hardware as part of a UCS, storage target, or both provides more visibility into the physical infrastructure of your environment. To extend the virtual infrastructure, application server or guest operating process targets can be added.

Workload Optimization Manager represents your environment holistically as a supply chain of resource buyers and sellers, all working together to meet application demand. Workload Optimization Manager maintains your environment within the desired state by empowering buyers and sellers:

- Buyers (VMs, instances, containers, and services) are given a budget to seek the resources that applications need to perform.
- Sellers can price their available resources (CPU, memory, storage, network) based on utilization in real-time.

Workload Optimization Manager does not support moves across stand-alone hosts or the merging of stand-alone hosts in any hypervisor. Workload Optimization Manager supports only VM moves across host cluster and merging host clusters.

For more information, see Application Resource Management *(on page 143)*.

Supply chain

Each hypervisor requires a physical machine (host) and one or more data stores to provide compute and storage resources. Virtual machines (VMs) or containers run on those physical resources, and the VMs in turn provide resources to applications.

At the end of the supply chain, physical machines consume resources from data centers.

If your environment includes SAN technologies such as disk arrays, then the storage consumes resources from that underlying technology. If you add these storage targets, then Workload Optimization Manager extends the supply chain analysis into the components that make up the disk array. For more information, see Storage Manager Targets *(on page 484)*.

IBM PowerVM

For IBM® Power8, Power9, and Power10 environments that use Hardware Management Console (HMC) versions 9 or 10, Workload Optimization Manager discovers and continuously decides on the optimal logical partition (LPAR) Processing Unit (PU) allocation and Virtual Processor (VP) capacity. This calculation is based on historical demand that is collected from the HMC.

IBM Power Systems and LPARs are represented as hosts and virtual machines and are visible across all HMC targets in the Workload Optimization Manager UI. Recommended LPAR resize PU and VP actions indicate optimal LPAR size, and can use existing Workload Optimization Manager workflow capabilities for external approval and execution.

Workload Optimization Manager automatically creates dynamic groups and Automation policies for the IBM Power Systems and LPARs.

- Workload Optimization Manager creates two dynamic groups, one for LPARs and one for systems. These groups are scoped to all LPARs and all Systems across all PowerVM targets that are added to the system.

  These groups are dynamic – If you allocate or remove entities, such as VMs or hosts, Workload Optimization Manager automatically adds or removes those entities in the group.

  The name of each group begins with `IBM PowerVM`. For example, `IBM PowerVM: All LPARs` and `IBM PowerVM: All Systems`.
- Workload Optimization Manager creates Automation policies to specify utilization and scaling constraints that are known to be appropriate for PowerVM environments. The name for each policy begins with the name `IBM PowerVM`. For more information, see Automation Policies *(on page 1049)*.

  Workload Optimization Manager then applies these policies to the groups that correspond to the entity type.

Prerequisites

- Supported Hardware Management Console (HMC) versions:
  - HMC 10: Power8, Power9, or Power10
  - HMC 9: Power8 or Power9 only; Power7 LPARs and Systems are ignored.
- Compute Long Term Monitor Metrics (Compute LTM) must be enabled on each system in the HMC. You can enable this function either manually by using the HMC API or in the Workload Optimization Manager user interface (UI) when you add or edit the target.
  - To manually enable Compute LTM by using the HMC for each Power System that you want Workload Optimization Manager to discover, configure the following setting by using the HMC API. You cannot enable the setting through the HMC UI.
    - Set the `ComputeLTMEnabled` property to `true` for all partitions.

      Take care not to confuse this property with the `LongTermMonitorEnabled` property. `ComputeLTMEnabled` and `LongTermMonitorEnabled` are separate properties and `ComputeLTMEnabled` is the property that you must set to `true` to enable metrics gathering.
    - For more information, see the Management Console Performance and Capacity Monitoring Preferences for your specific system.
  - Alternatively, when you add or edit the target in the Workload Optimization Manager UI, select **Automatically enable HMC Long Term Metric collection on Power Systems** to enable Compute LTM on all capable systems during discovery. This option is not enabled by default and it requires higher HMC account permissions *(on page 335)*, which are detailed in the user account prerequisites.

- Optionally, you can enable Energy Monitoring to view the power and energy metrics on Workload Optimization Manager. You can enable this function either manually by using the HMC API or in the Workload Optimization Manager UI when you add or edit the target.
  - To manually enable Energy Monitoring for each system in the HMC, enable performance collection from HMC UI:
    - The value for the `EnergyMonitoringCapable` attribute must be `true`. This read-only attribute determines whether the managed system supports energy data collection.
    - Enable energy monitoring by setting the `EnergyMonitorEnabled` property to `true`.
    - For more information, see the Management Console Performance and Capacity Monitoring Preferences for your specific system.
  - Alternatively, when you add or edit the target in the Workload Optimization Manager UI, select **Automatically enable energy monitoring on energy monitoring capable Power Systems** to enable energy monitoring on all capable systems during discovery. This option is not enabled by default and it requires higher HMC account permissions *(on page 335)*, which are detailed in the user account prerequisites.
- An HMC user account that Workload Optimization Manager can use to connect to the PowerVM target.

  The HMC user account must have the following roles:
  - The `hmcviewer` role is required to discover and view metrics. This role cannot edit HMC settings.
  - To use the **Automatically enable HMC Long Term Metric collection on Power Systems** or **Automatically enable energy monitoring on energy monitoring capable Power Systems** options in the Workload Optimization Manager UI, the user account must have one of the following two permissions:
    - The `hmcoperator` role
    - The `Manage Partition Utilization Data` task role

  The user account must also be enabled for remote access by selecting the **Allow remote access via the web** option in the HMC user properties.

  To enable this option, select **HMC Management** in the HMC navigation panel and click **Manage User Profiles and Access**. Select the user profile and click the **User, Modify**, menu option. In the **Modify User** dialog, click the **User Properties...** button and select the **Allow remote access via the web** checkbox. Click **OK**, **OK** to save the change. For more information, see HMC Version 7 and later: How to Access the HMC GUI Remotely.
- The HTTPS TCP port 443 from Workload Optimization Manager to the HMC instance must allow connections from Workload Optimization Manager to the HMC.

IBM PowerVM target configuration details

Before you configure an IBM PowerVM target, review the following considerations:

- Workload Optimization Manager recommends Resize LPAR actions only.
- Stitching to entities above the LPAR/VM or below the System/Host is not currently supported.
- Where multiple HMCs are managing the same Power Systems (Clustered/HA configuration) add only one of the HMC nodes as a target to Workload Optimization Manager.
- Plans are not currently supported. Planning features are being considered in a future release.
- Processor pools are not currently supported.
- Reporting that uses ThoughtSpot is not currently supported.
- The **Details** tab requires manual configuration for static group scopes.
- VIOS LPARs are not discovered currently.
- In a few cases, actions to resize LPAR PU and VP resources might result in recommendations for PU allocations above VP capacity and vice versa.

Adding IBM PowerVM targets

NOTE:
If you do not see the IBM PowerVM target, you must enable the PowerVM probe before you add the target. To enable the probe, set the following in the `charts_v1alpha1_xl_cr.yaml` resource:

```
spec:
  powervm:
    enabled: true
```

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **IBM PowerVM**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Display name

     Specify a name that uniquely identifies this connection.

     The name that you provide here is also used for the Data Center and Cluster name to which Systems managed by this HMC belong.
   - Address

     Specify the IP/Domain of the PowerVM HMC.
   - Port number

     Specify the port number that Workload Optimization Manager uses to connect to remote management. By default, the port is 443 for HTTPS.
   - Username

     Specify the username of the account Workload Optimization Manager uses to connect to the target.
   - Password

     Specify the password of the account Workload Optimization Manager uses to connect to the target.
   - Enable HMC Long Term Metric collection on Power Systems

     If you select this option, Workload Optimization Manager automatically enables Compute-Long Term Monitor metrics (LTM) on each power system that the HMC manages. By default, Workload Optimization Manager does not make any changes on HMC; however, Compute-LTM must be enabled on each Power System either automatically (enable this option) or manually.

     Compute-LTM is a prerequisite for metrics collection.

     To enable this option, the user account that is used by Workload Optimization Manager must have `Manage Partition Utilization Data` as a Task Role in HMC.
   - Enable energy monitoring on energy monitoring capable Power Systems

     If you select this option, Workload Optimization Manager automatically enables energy monitoring for hosts that are energy monitoring capable and enabled on each power system that the HMC manages.

Entity mapping

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

HMC	Workload Optimization Manager
Logical Partition (LPAR)	Virtual Machine (VM)
System	Host

HMC	Workload Optimization Manager
-	Data Center; skeleton entity used to complete the topology This entity uses the target Display Name. One Data Center entity exists for each PowerVM target.

The following table describes the mapping between HMC systems states and Workload Optimization Manager.

HMC System State	Workload Optimization Manager Power State
INITIALIZING	UNKNOWN
OPERATING	ACTIVE
STANDBY	UNKNOWN
POWER_OFF	IDLE

Monitored resources

Workload Optimization Manager monitors the following resources:

- Virtual Machine
  - Processing unit (PU)

    For PowerVM, processing unit is a unit of measurement for shared processing power across one or more virtual processors. One shared processing unit on one virtual processor accomplishes approximately the same work as one dedicated processor.

    PUs can be a decimal value in a VM setting. A VM that has a capacity of 0.5 PU means that 0.5 PU is always readily available for that VM to use, and that amount can only be used for that VM.
  - Virtual processor (VP)

    For PowerVM, virtual processor (VP) is the guest OS representation of the capacity of the VM. VPs have a one-to-one correlation with PUs. Therefore ,1 VP that is utilized in a VM equates to 1 PU being used. However, VPs do not have to be equal to PUs on a system. VPs are a virtual representation. Therefore, they can be overprovisioned.

    For example, a VM can have 0.5 PUs but 3 VPs. In this scenario, a VM's capacity is 0.5 PU but it can use more than its assigned capacity; the max it can use is the number of VPs. The number of VPs must never exceed 20x the value of PU, therefore the highest utilization a VM can be at is 2000% of its PU capacity.

    NOTE:
    Only shared-uncapped VM utilization can go up to 2000% of its PU capacity; the utilization for all other VM types cannot go above 100% of its PU capacity.

- Host
  - Energy

    Energy is the measurement of electricity used by a given entity over a period of time, expressed in watt-hours (Wh).
  - Carbon footprint

    Carbon footprint is the measurement of carbon dioxide equivalent (CO2e) emissions for a given entity. Workload Optimization Manager measures carbon footprint in grams.
  - Power

    Power is the measurement of electricity consumed by a given entity, expressed in watts.
  - Memory (Mem)

    Memory is the measurement of memory that is reserved or in use.

- Virtual processor (VP)

  For PowerVM, virtual processor (VP) is the guest OS representation of the capacity of the VM. VPs have a one-to-one correlation with PUs. Therefore ,1 VP that is utilized in a VM equates to 1 PU being used. However, VPs do not have to be equal to PUs on a system. VPs are a virtual representation. Therefore, they can be overprovisioned.

  For example, a VM can have 0.5 PUs but 3 VPs. In this scenario, a VM's capacity is 0.5 PU but it can use more than its assigned capacity; the max it can use is the number of VPs. The number of VPs must never exceed 20x the value of PU, therefore the highest utilization a VM can be at is 2000% of its PU capacity.

  NOTE:
  Only shared-uncapped VM utilization can go up to 2000% of its PU capacity; the utilization for all other VM types cannot go above 100% of its PU capacity.

- Datacenter

  The Datacenter entity is not monitored but is used in the supply chain to represent each PowerVM HMC and the Hosts and VMs associated with it.

  - Power

    Power is the measurement of electricity consumed by a given entity, expressed in watts.

Actions

Workload Optimization Manager supports the following actions:

- Virtual Machine

  Resize PU or VP (IBM PowerVM)

  Resize PU (processing units) or VP (virtual processor) to optimize LPAR processing unit allocation and virtual processor capacity based on historical demand collected from the HMC.

  When Workload Optimization Manager generates an action, the action details include a list of policies that constrain the action recommendation. In this list, policy names that begin with `IBM PowerVM:` indicate policies that implement the constraints suitable for PowerVM. You can click to navigate to the policy definition. You can see the policy definition, but you cannot edit it.

  The policies that Workload Optimization Manager generates affect Workload Optimization Manager actions in the following ways:

Entity	Policy Impact on Actions
Virtual Machine	Workload Optimization Manager does not support native execution of Resize PU (processing units) and Resize VP (virtual processor) actions. However, actions can be executed using external automation platforms through automation workflows such as Action Scripts or Webhooks. The following policy settings affect the resize actions that Workload Optimization Manager generates: – Min/Max observation period Observation period settings only apply to Resize VP actions. For details, see Aggressiveness and Observation Periods *(on page 752)*. – Increment constant for processing unit Change the increment constant value to use for VM PU resize actions. PUs can be a decimal value in a VM setting. For details, see Increment Constant for Processing Unit *(on page 749)*. – Increment constant for virtual processors

Entity	Policy Impact on Actions
	Change the increment constant value to use for VM VP resize actions. VPs must be an integer value in a VM setting. For details, see Increment Constant for Virtual Processors *(on page 750)*.
Host	– Memory The memory that is given to a VM from the host system; what is provided by the host is all that the VM can use. • Memory Overprovisioned Percentage Change the percentage of the existing capacity that Workload Optimization Manager considers to be 100% of capacity. For details, see Utilization Constraints *(on page 791)*.

Microsoft Hyper-V

If you have a few Microsoft Hyper-V hosts in your environment, you can add them individually as Workload Optimization Manager targets. Also, if you have deployed the Hyper-V hosts in a clustered domain (for example as a failover cluster), you can specify one Hyper-V host as a target and Workload Optimization Manager automatically adds the other members of that cluster.

For large Hyper-V environments, it's typical to manage the hosts via System Center Virtual Machine Manager (VMM). You can specify the VMM server as a target and Workload Optimization Manager will use it to discover and manage its child Hyper-V hosts. If you use VMM, you should not add individual Hyper-V hosts as targets. For information about adding VMM targets, see Adding Virtual Machine ManagerTargets *(on page 368)*.

> **NOTE:**
> Even if you manage Hyper-V using a VMM target, you must still configure remote management on each Hyper-V server. This Hyper-V topic includes instructions to configure remote management – see Enabling Windows Remote Management *(on page 343)*.

For accurate Server Message Block (SMB) storage calculations, Workload Optimization Manager requires a VMM target.

Prerequisites

- Microsoft Hyper-V 2012 R2, 2016, 2019, or 2022
- Create a service user account in Hyper-V *(on page 342)*
- Configure remote management on each Hyper-V server *(on page 343)*
- Sync the time on each Hyper-V host with the rest of the managed Hyper-V environment
- Ensure that your Hyper-V environment does not use Server Message Block (SMB) storage

  To manage SMB storage, Workload Optimization Manager requires a VMM target, and that VMM instance must manage the Hyper-V hypervisors and the SMB storage that they use.

  Managing a Hyper-V plus SMB environment via Hyper-V targets will result in incorrect data collection for SMB storage.

Adding Hyper-V targets

Once you've enabled remote management, you can add your Hyper-V hosts as targets.

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **Hyper-V**.
4. In the side panel, review the connection requirements and then click **Connect Target**.

5. Configure the following settings:
   - Address

     Specify the fully qualified domain name of the Hyper-V host. If you are using the “Discover Host Cluster” option to add an entire cluster, enter the name of any one of the Hyper-V hosts in the cluster.

     NOTE:
     You can enter an IP address for the host, but you must first configure an SPN on the host. Cisco recommends that you use the FQDN in this field.

   - Username

     Specify the username of the account Workload Optimization Manager uses to connect to the target. If you select "Discover Host Cluster”, use an account that is valid for all Hyper-V hosts in that cluster.
   - Fully qualified domain name

     Specify the fully qualified domain name of the cluster to which the host belongs.
   - Password

     Specify the password of the account Workload Optimization Manager uses to connect to the target.
   - Port number

     Specify the port number that Workload Optimization Manager uses to connect to remote management. By default, this is port 5985 for HTTP and port 5986 for HTTPS.
   - Discover host cluster

     If you select this option, Workload Optimization Manager discovers and adds all Hyper-V hosts in the named cluster. Each server must be configured to allow remote management.

     You may find it helpful to configure WinRM using a GPO so new servers are configured automatically (see Enabling WinRM Via a GPO *(on page 344)*).
   - Connect using HTTPS

     If you select this option, Workload Optimization Manager connects to the target servers using HTTPS. Make sure that the required certificate is configured for use on the host.
   - Use Kerberos message-level encryption over HTTP

     If this option is selected, Workload Optimization Manager uses Kerberos message-level encryption to discover the environment. This value is ignored if "Secure connection" is enabled.

NOTE:
If your Hyper-V hosts are running in a Nutanix environment, you must understand pinning a Nutanix Controller VM. For more information, see Pinning Controller VMs in Generic Hypervisor Mode *(on page 330)*.

Exporting Hyper-V virtual machines

In Hyper-V environments, you must be sure that all VMs have unique IDs.

Hyper-V supports the export of a VM, so that you can create exact copies of it by importing those exported files. The `Copy` import type creates a new unique ID for the imported VM. When importing VMs in your environment, you should always use the `Copy` import type.

Workload Optimization Manager uses the unique ID to discover and track a VM. If your environment includes multiple VMs with the same ID, then discovery assumes they are the same VM. As a result, the counts for VMs are incorrect.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Virtual Machine
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.
  - Virtual storage

    Virtual ssorage is the measurement of virtual storage capacity that is in use.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - Latency

    Latency is the measurement of storage latency.
- Host
  - Memory (Mem)

    Memory is the measurement of memory that is reserved or in use.
  - CPU

    CPU is the measurement of CPU that is reserved or in use.
  - IO

    IO is the utilization of a host's IO adapters.
  - Net

    Net is the utilization of data through the host's network adapters.
  - Swap

    Swap is the measurement of a host's swap space that is in use.
- Storage
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

    NOTE:
    When it generates actions, Workload Optimization Manager does not consider IOPS throttling that it discovers on storage entities. Analysis uses the IOPS that it discovers on Logical Pool or Disk Array entities.

  - Latency

    Latency is the measurement of storage latency.

- Datacenter

  NOTE:
  For datacenter entities, Workload Optimization Manager does not monitor resources directly from the datacenter, but from the hosts in the datacenter. See host monitored resources for details.

Actions

Workload Optimization Manager supports the following actions:

- Virtual Machine
  - Start
  - Move
  - Suspend
  - Resize Up/Down
  - Terminate

    This action can only be executed outside Workload Optimization Manager.
  - Provision

    This action can only be executed outside Workload Optimization Manager.
  - Reconfigure

    This action can only be executed outside Workload Optimization Manager.
- Host
  - Start
  - Suspend
  - Terminate

    This action can only be executed outside Workload Optimization Manager.
  - Provision

    This action can only be executed outside Workload Optimization Manager.
- Storage
  - Provision

    This action can only be executed outside Workload Optimization Manager.

Creating a service user account

The service account Workload Optimization Manager uses to connect to a Hyper-V host must be an Active Directory domain account. The account must have full access to the cluster. To create such an account, execute the following command at a PowerShell prompt:

```
Grant-ClusterAccess <domain>\<service_account> -Full
```

Additionally, the service account must have specific local access rights on each host. The easiest way to grant Workload Optimization Manager the access it requires is to add the domain account to the `Local Administrators` group on each Hyper-V server.

Some enterprises require that the service account does not grant full administrator rights. In that case, you can create a restricted service account on every Hyper-V host.

NOTE:
Workload Optimization Manager does not support Restricted User Accounts on Windows 2012 Hyper-V nodes.

To create a restricted service account on your Hyper-V hosts:

1. Add the service account to each of the following local groups:
   - `WinRMRemoteWMIUsers__` or `Remote Management Users`
   - `Hyper-V Administrators`
   - `Performance Monitor Users`

   NOTE:
   These groups are examples only. If your version of Windows Server does not include these groups, contact Technical Support for assistance.

2. Grant permissions to the service account.

   In the WMI Management console, grant the following permissions to the service account:
   - `Enable Account`
   - `Remote Enable`
   - `Act as Operating System` (for Windows 2016)
3. Configure the WinRM security descriptor to allow access by the service account:
   - At a PowerShell prompt, run the command:

     ```
     winrm configSDDL default
     ```

   - In the Permissions for Default dialog box, grant the service account Read and Execute access.

Enabling Windows Remote Management

Workload Optimization Manager communicates with your Hyper-V servers using Web Services Management (WS-Management), which is implemented on Microsoft platforms using Windows Remote Management (WinRM). The following steps show how to enable WinRM on a single host, using the command line.

1. Ensure Windows Firewall is running on the host.

   For you to configure WinRM successfully, Windows Firewall must be running on the host. For more information, see the Microsoft Knowledge Base article #2004640 (http://support.microsoft.com/kb/2004640).
2. Set up a Service Principal Name (SPN) for the host machine.

   The machine must have a SPN of the form, `protocol/host_address`. For example, `WSMAN/10.99.9.2`.

   To get a list of SPNs for the machine, run the following in the command window:

   ```
   setspn -l <vmm-server-name>
   ```

   If there is no valid SPN in the list, create one by running the command:

   ```
   setspn -A protocol/host-address:port
   ```

   where `port` is optional. For example:

   ```
   setspn -A WSMAN/10.99.9.2:VMM-02
   ```
3. Set up the Windows Remote Management (WinRM) service to run on startup.

Run the `quickconfig` utility to set up the WinRM service. The `quickconfig` utility:

- Configures the WinRM service to auto-start
- Configures basic authentication and disables unencrypted traffic
- Creates a firewall exception for the current user profile
- Configures a listener for HTTP and HTTPS on any IP address
- Enables remote shell access

To run `quickconfig`, log into a command window as Administrator on the host machine. Then execute the following commands:

```
winrm quickconfig
```

Enter `y` to accept the `quickconfig` changes

4. Set permissions on the host machine.

   Execute the following commands in the command window to modify the settings made by `quickconfig`:

   - To set the memory capacity for remote shells:

     ```
     winrm set winrm/config/winrs @{MaxMemoryPerShellMB="1024"}
     ```

   - To set up an unsecured HTTP connection:

     ```
     winrm set winrm/config/service @{AllowUnencrypted="true"}
     ```

     ```
     winrm set winrm/config/service/Auth @{Basic="true"}
     ```

These steps showed you how to enable WinRM for a single host. Some users find the following methods useful for enabling WinRM on multiple hosts:

- EnablingWinRmViaGlobal Policy Objects *(on page 344)*
- EnablingWinRMViaPowerShell *(on page 345)*

Enabling Windows Remote Management through global policy objects

You can configure Windows Remote Management (WinRM) for all of your Hyper-V targets by creating and linking a Global Policy Object (GPO) within the Hyper-V domain and applying the GPO to all servers.

Follow the steps to enable WinRM for your Hyper-V targets.

1. On the AD domain controller, open the Group Policy Management Console (GPMC). If the GPMC is not installed, see https://technet.microsoft.com/en-us/library/cc725932.aspx.
2. Create a new Global Policy Object:
   a. In the GPMC tree, right-click **Group Policy Objects** within the domain containing your Hyper-V servers.
   b. Choose **Create a GPO in this domain,** and link it here.
   c. Enter a name for the new GPO and click **OK**.
3. Specify the computers that need access:
   a. Select the new GPO from the tree.
   b. On the **Scope** tab, under **Security Filtering**, specify the computer or group of computers you want to grant access. Make sure you include all of your Hyper-V targets.
4. Right-click the new GPO and choose **Edit** to open the Group Policy Management Editor.
5. Configure the WinRM Service:
   a. In the Group Policy Management Editor, select **Computer Configuration > Policies > Administrative Templates > Windows Components > Windows Remote Management (WinRM) > WinRM Service**.
   b. Double-click each of following settings and configure as specified:

Setting	Value
Allow automatic configuration of listeners (“Allow remote server management through WinRM” on older versions of Windows Server)	Enabled IPv4 filter: *
Allow Basic authentication	Enabled
Allow unencrypted traffic	Enabled

6. Configure the WinRM service to run automatically:
   a. In the Group Policy Management Editor, expand **Computer Configuration > Preferences > Control Panel Settings**.
   b. Under Control Panel Settings, right-click Services and choose **New > Service**.
   c. In the New Service Properties window, configure the following settings:

Setting	Value
Startup	Automatic
Service name	WinRM
Service option	Service start

7. Enable Windows Remote Shell:
   a. In the Group Policy Management Editor, select **Computer Configuration > Policies > Administrative Templates > Windows Components > Windows Remote Shell**.
   b. Double-click the following setting and configure as specified:

Setting	Value
Allow Remote Shell Access:	Enabled

8. Add a Windows Firewall exception:
   a. In the Group Policy Management Editor, open **Computer Configuration > Windows Settings > Security Settings > Windows Firewall > Windows Firewall**.
   b. Under Windows Firewall, right-click **Inbound Rules** and choose **New > Rule**.
   c. In the New Inbound Rule Wizard, select **Predefined: Windows Remote Management and Allow the connection**.

The new group policy will be applied during the next policy process update. To apply the new policy immediately, run the following command at a Powershell prompt:

```
gpupdate /force
```

Enabling Windows Remote Management through PowerShell

Using PsExec, you can run `quickconfig` on all your Hyper-V servers and change the default settings remotely. PsExec is a component of PsTools, which you can download from https://technet.microsoft.com/en-us/sysinternals/bb897553.aspx.

1. Create a text file containing the Hyper-V hostnames, for example:

```
hp-vx485
hp-vx486
```

2. Since Cisco requires changes to the default quickconfig settings, create a batch file containing the following command:

```
@echo off Powershell.exe Set-WSManQuickConfig -Force Powershell.exe Set-Item
 WSMan:\localhost\Shell\MaxMemoryPerShellMB 1024
```

NOTE:
If you are connecting via HTTP, you must include the following command:

```
Powershell.exe Set-Item WSMan:\localhost\Service\AllowUnencrypted -Value $True
```

3. Use PsExec to enable WinRM on the remote servers:

```
.\PsExec.exe @<hosts_file_path> -u <username> -p <password> -c <batch_file_path>
```

NOTE:
If you get an error message when executing this command, add the `-h` option (`.\PsExec.exe -h`).

vCenter Server

VMware vCenter Server provides a centralized management platform for VMware hypervisors. To manage your VMware environment with Workload Optimization Manager, you specify a vCenter Server instance as a target. Workload Optimization Manager discovers the infrastructure that target manages, and links it into a supply chain to deliver application performance management.

Prerequisites

- VMware vCenter 7.0 or 8.0
- Create a user account in vCenter *(on page 352)*

  Workload Optimization Manager uses this user account to connect to your vCenter Server and execute actions.

General considerations

Before you configure a vCenter Server target, you should consider the following:

- Linked vCenters

  For linked vCenters, you must add each vCenter Server separately so Workload Optimization Manager can communicate with each vCenter Server through a separate API endpoint.
- Restricting Workload Optimization Manager Access to Specific Clusters

  When you add a vCenter Server target, Workload Optimization Manager discovers all of the connected entities that are visible, based on the target account that it uses to connect to the vCenter Server target. If you have clusters or other entities you want to exclude from discovery, you can use the vSphere management client to the role of the Workload Optimization Manager account to `No access` for the given entities.
- Shared Datastores

  If you add more than one vCenter Server target that manages the same datastore, you can enable or disable datastore browsing to discover wasted files on the shared datastore:
  - Enable datastore browsing:

    To properly enable browsing, you must turn on the **Enable Datastore Browsing** option in the target configuration for each vCenter Server target that manages the shared datastore.
  - Disable datastore browsing:

    If you don't want datastore browsing over shared datastores, you must turn *off* the **Enable Datastore Browsing** option in the target configuration for each vCenter Server target that manages the shared datastore.

  If set **Enable Datastore Browsing** differently for separate targets that manage the same datastore, datastore browsing can give inconsistent results for active and wasted files.
- VSAN Permissions

In order to enable VSAN support and discover groups based on storage profiles, you must ensure that the user role Workload Optimization Manager is assigned has the `Profile-driven storage view` permission enabled. This permission is *disabled* in the built-in `readonly` role.

- Guest Metrics

  Enable Guest Metrics is enabled by default for all vCenter Server targets added to Workload Optimization Manager.

  > **NOTE:**
  > If your vCenter Server target was added prior to 8.7.3, this option is enabled as part of the update.

  This target option instructs that Workload Optimization Manager request from the VC instance that it start collecting and reporting enhanced memory metrics from each virtual machine, using the VMware Tools installed on each VM. Without this option, Workload Optimization Manager only has access to the "Active" memory for each virtual machine, which is a less accurate representation of the memory that the VM is really using.

  For accurate metrics use vCenter Server 6.5 U3 or higher with VMware Tools version 10.3.2 or higher on guest VMs. If you run vCenter Server version 7.0 or greater, then it can automatically update your VMware Tools through the user interface. For other versions of vCenter Server, you must manually update your VMware Tools deployments.

  For the permissions required for Guest Metrics, see vCenter Permissions *(on page 352)*.

- VMware Cloud Deployments

  VMware Cloud (VMC) is a managed cloud offering that provides dedicated VMware vSphere-based Software Defined Data Centers (SDDC). VMC on AWS hosts SDDC on AWS facilities. For on-premise VMC, SDDC runs on specific hardware that meets VMware specifications.

  VMware Cloud on AWS, Azure VMware Services, Google Cloud VMware Engine, IBM Cloud for VMware and other cloud hosted VMware solutions are treated like any other VMware integration. The same familiar Workload Optimization Manager experience can be expected when using the VMware vCenter Server target integration against any vCenter Server target whether it is cloud hosted or not, provided you pay close attention to satisfying all of the vCenter target prerequisites, especially permissions and roles. For example, VMware Cloud bare metal host instances may be modeled in Workload Optimization Manager as Hyperconverged templates, assigned a cost, and used to plan the migrating from existing on-premises host clusters to cloud hosted VMware solutions. Once in the cloud, you can use Workload Optimization Manager to plan migrations from one cloud hosted VMware solution to another. Provision and Suspend Host actions generated by Workload Optimization Manager may be more readily actionable on cloud hosted VMware solutions and automated using custom Action Scripts if required. If you have any requirements or improvements that relate specifically to cloud hosted VMware solutions, submit an enhancement request through IBM Ideas for future consideration.

  For VMC environments, you can use the Workload Optimization Manager vCenter Server target to manage workloads, but you should be aware of the following concerns:

  - DRS Settings

    VMC SDDC requires DRS to be set to migration level 3. You cannot change this setting.

  - Datacenter Names

    VMC SDDC datacenter names are always SDDC `Datacenter`. This is true even for multiple VMC targets. You can use the `Target Name` filter to search and group VMs across the set of VMC SDDC datacenters and clusters that share this name.

  - Discovery

    - Workload Optimization Manager does not discover VMC service costs or subscription details.
    - The user interface does not show cost savings or investments.

  - Migration Plans

    Plans do not discover VMC bare metal hosts. You can use Workload Optimization Manager HCI Host Templates to represent the bare metal instances. Then the plan can determine the number of VMC instances you need to support your existing on-premise workloads.

    Migrate to Cloud plans do not support VMC.

Adding vCenter Server targets

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **vCenter**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Address

     Specify the fully qualified domain name or IP address of the vCenter Server.
   - Username

     Specify the username of the account Workload Optimization Manager uses to connect to the target.

     Include the domain if required (`<domain>\<username>`).
   - Password

     Specify the password of the account Workload Optimization Manager uses to connect to the target.

     NOTE:
     The password cannot contain any of the following special characters: `\ / " [ ] : | < > + = ; , ? * , ' tab space @`
   - Discover unused storage

     If this option is selected, Workload Optimization Manager discovers unused storage.
   - Enable guest metrics (vSphere 6.5U3+)

     If this option is selected, Workload Optimization Manager collects advanced guest memory metrics that increase the accuracy of the VMEM data that Workload Optimization Manager uses for analysis of virtual machines. Guest metrics are only supported on vCenter Server Server version 6.5U3 or later. Guest VMs must run VMware Tools 10.3.2 Build 10338 or later.

     To enable guest metrics, ensure the following:
     - VMware Tools is installed and running on the target VMs.
     - The user account has the `Performance.Modify Intervals` performance privilege.

     For more information, see vCenter Performance Privileges.
   - Certificate authority

     Specify the root certificate for validating the SSL connection to vCenter storage arrays.

vCenter Server imported settings

In addition to discovering entities managed by the hypervisor, Workload Optimization Manager also imports a wide range of vSphere settings, such as Host DRS rules, annotations, Resource Pools, and DRS HA settings (See Other Information Imported From vCenter *(on page 353)*).

NOTE:
Workload Optimization Manager does not import Storage DRS rules at this time.

VMware vSphere 6.0 introduced the ability to move VMs between vCenters. If you enabled this feature in your VMware environment, you can configure Workload Optimization Manager to include cross vCenter Server vMotions in its recommendations. You must create a Workload Placement Policy that merges the data centers on the different vCenters, and then another policy to merge the given clusters.

For vCenter environments, you can create placement policies that merge data centers to support cross-vCenter moves. In this case, where a data center corresponds to a vCenter target, the merged clusters can be in different data centers. In this case, you must create two merge policies; one to merge the affected data centers, and another to merge the specific clusters.

To create a Merge policy:

1. In the Policy Management tab, select **Placement Policy**.
2. For `Type`, select **Merge**.
3. For `Merge`, choose the merge type, and click **Select**.
4. Choose the specific data centers or clusters to merge in this policy, then click **Select**.
5. Click **Save Policy**.

NOTE:
Since Workload Optimization Manager can only execute vMotions between clusters that use the same switch type (VSS or VDS), make sure any clusters you merge use the same switch type. Although Workload Optimization Manager will not initiate VSS → VDS vMotions, vSphere may do so. If this happens, Workload Optimization Manager displays a compliance violation notification.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Virtual Machine
  - Energy

    Energy is the measurement of electricity used by a given entity over a period of time, expressed in watt-hours (Wh).
  - Carbon footprint

    Carbon footprint is the measurement of carbon dioxide equivalent (CO2e) emissions for a given entity. Workload Optimization Manager measures carbon footprint in grams.
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
  - Virtual storage

    Virtual ssorage is the measurement of virtual storage capacity that is in use.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - Latency

    Latency is the measurement of storage latency.
- Host
  - Energy

    Energy is the measurement of electricity used by a given entity over a period of time, expressed in watt-hours (Wh).
  - Power

    Power is the measurement of electricity consumed by a given entity, expressed in watts.
  - Carbon footprint

    Carbon footprint is the measurement of carbon dioxide equivalent (CO2e) emissions for a given entity. Workload Optimization Manager measures carbon footprint in grams.
  - Memory (Mem)

    Memory is the measurement of memory that is reserved or in use.
  - CPU

CPU is the measurement of CPU that is reserved or in use.

- IO

  IO is the utilization of a host's IO adapters.

- Net

  Net is the utilization of data through the host's network adapters.

- Swap

  Swap is the measurement of a host's swap space that is in use.

- Balloon

  Balloon is the measurement of memory that is shared by VMs running on a host.

  This commodity applies to ESX only.

- CPU ready

  CPU ready is the measurement of a host's ready queue capacity that is in use.

  This commodity applies to ESX only.

- Storage
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.

  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.

  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

    NOTE:
    When it generates actions, Workload Optimization Manager does not consider IOPS throttling that it discovers on storage entities. Analysis uses the IOPS it discovers on Logical Pool or Disk Array entities.

  - Latency

    Latency is the measurement of storage latency.

  - Unattached files

    Unattached files are files that are not currently connected to a virtual machine.

- Datacenter

  NOTE:
  For datacenter entities, Workload Optimization Manager does not monitor resources directly from the datacenter, but from the hosts in the datacenter. See host monitored resources for details.

- Provider Virtual Datacenter
  - Memory (Mem)

    Memory is the measurement of memory that is reserved or in use.

  - CPU

    CPU is the measurement of CPU that is reserved or in use.

  - Storage

    Storage is the utilization of the storage attached to the entity.

- Consumer Virtual Datacenter
  - Memory (Mem)

    Memory is the measurement of memory that is reserved or in use.
  - CPU

    CPU is the measurement of CPU that is reserved or in use.
  - Storage

    Storage is the utilization of the storage attached to the entity.

Actions

In order to execute cross-vCenter migrations as a non-admin user, you must have the following permissions enabled for the user account in both the current and destination vCenter servers:

Entity	Permissions
Virtual Machine	Edit Inventory, Create From Existing (Move, Register, Remove, Unregister sub-options), Create New
Datacenter	Reconfigure Datacenter
Network	Assign Network

Workload Optimization Manager supports the following actions:

- Virtual Machine
  - Start
  - Move
  - Move VM Storage
  - Suspend
  - Resize Up/Down
  - Terminate

    This action can only be executed outside Workload Optimization Manager.
  - Provision

    This action can only be executed outside Workload Optimization Manager.
  - Reconfigure

    This action can only be executed outside Workload Optimization Manager.
- Host
  - Start
  - Suspend
  - Terminate

    This action can only be executed outside Workload Optimization Manager.
  - Provision

    This action can only be executed outside Workload Optimization Manager.
- Storage
  - Delete unattached file

    Any file that has not been modified in the period defined in the storage policy. The default behavior is to generate an action after 15 consecutive days of non-use.
  - Provision

    This action can only be executed outside Workload Optimization Manager.

Creating a service user account in vCenter

The service account you use must have specific permissions on the vCenter. The easiest way to grant Workload Optimization Manager the access it requires is to grant full administrator rights.

Some enterprises require that the service account does not grant full administrator rights. In that case, you can create a restricted service account that grants the following permissions to enable the required Workload Optimization Manager activities. For any environment implementing merge policies, these permissions need to be enabled at the vCenter level. The list of permissions is defined in the following table.

Requirements and permissions for the vCenter target

Workload Optimization Manager Functionality	Requirements and Permissions
Monitor workloads	▪ Read-only role for all entity types Assign either Global permissions or permissions for the given vCenter Server instance to the target user or user group. ▪ Requirement to monitor VSAN and storage profiles In order to enable VSAN support and discover groups based on storage profiles, you must enable the `Profile-driven storage view` permission. This permission is *disabled* in the built-in `readonly` role.
Recommend actions	▪ Read-only role for all entity types Assign either Global permissions or permissions for the given vCenter Server instance to the target user or user group.
Report wasted storage	▪ `Datastore > Browse Datastore`
Execute delete files	▪ `Datastore > Low level file operations`
Execute host suspension	The host should be configured to support standby mode using Wake-on-Lan or the Intelligent Platform Management Interface Protocol (IPMI) protocol. The Evacuate parameter is valid only when the host is a part of a DRS-enabled cluster. For more information, see the vCenter documentation.
Execute VM move	▪ `Resources > Assign VM to Resource Pool` ▪ `Resources > Migrate Powered Off VMs` ▪ `Resources > Migrate Powered On VMs` ▪ `Resources > Modify Resource Pool` ▪ `Resources > Query Vmotion` ▪ `Network > Assign` Assign either Global permissions or permissions for the given vCenter Server instance to the target user or user group.
Execute VM storage move	▪ `Datastore > Allocate Space` ▪ `Datastore > Browse Datastore` ▪ `Resources > Assign VM to Resource Pool` ▪ `Resources > Migrate` ▪ `Resources > Modify Resource Pool` ▪ `Resources > Move Resource Pool` ▪ `Resources > Query VMotion`

Workload Optimization Manager Functionality	Requirements and Permissions
	▪ `Virtual Machine > Change Configuration > Change resource` ▪ `Virtual Machine > Change Configuration > Change Swapfile placement`
Execute VM resize	▪ `Virtual Machine > Change Configuration > Change CPU count` ▪ `Virtual Machine > Change Configuration > Change Memory` ▪ `Virtual Machine > Change Configuration > Change resource` ▪ `Virtual Machine > Interaction > Reset` ▪ `Virtual Machine > Interaction > Power Off` ▪ `Virtual Machine > Interaction > Power On`
Discover tags	▪ `Global > Global tag` You must also open ports `10443` and `7443` on the target server
Collect guest metrics	▪ `Performance > Modify Intervals`

Other information imported from vCenter

In addition to discovering entities managed by the vSphere hypervisors and their resources, Workload Optimization Manager:

- Imports any vSphere Host DRS rules when DRS is enabled, and displays them on the **Policy > Workload Placement** view under **Imported Placement Policies**. Imported rules are enabled by default, but you can disable them in Workload Optimization Manager.

  NOTE:
  In vCenter environments, Workload Optimization Manager does not import DRS rules if DRS is disabled on the hypervisor. Further, if Workload Optimization Manager did import an enabled DRS rule and somebody subsequently disables that DRS rule, then Workload Optimization Manager will discover that the rule was disabled and will remove the imported placement policy.

- Imports any custom annotations and displays related groupings in the **Inventory > Groups** tree view, under **VC Annotations**. The service account must enable the **Global > Global tag** privilege.
- For vCenter Server versions 5.5 and later, discovers Virtual Machine Storage Profiles and displays them as groups anywhere that you can set scope. The groups appear under **VC Storage Profiles**. You can use these discovered storage profiles the same as any other groups — For example, to scope dashboards, or to set the scope for specific action policies.
- Discovers resource pools and displays them as folders in the Inventory tree and as components in the Supply Chain Navigator. If you have the Cloud Control Module license, Workload Optimization Manager manages resource pools as Virtual Datacenters (VDCs) and can recommend resize actions. Root resource pools appear as Provider VDCs in the supply chain, whereas child resource pools appear as Consumer VDCs.
- Imports vSphere HA cluster settings and translates them into CPU and memory utilization constraints. These are displayed as cluster-level overrides under **Folders** on the **Policy > Analysis > Host** view.
- Discovers the GPUs installed in your ESXi hosts, the virtual GPUs allocated to your Virtual Machines, and GPUs in pass-through mode. GPU information is shown in the **Entity Information** widget.

IT Management targets

IT Management targets enable Workload Optimization Manager to understand and participate in your operational environment. These targets integrate with your change approval processes and optimize license usage and technology business management.

Workload Optimization Manager currently supports IBM Apptio, Flexera (FlexNet Manager Suite 2023 R2 (on-prem), Flexera One (SaaS)), and ServiceNow IT Management targets.

Flexera

Workload Optimization Manager integrates its management of workloads with Flexera Software License Asset Management. To do this, Workload Optimization Manager discovers the licenses and entitlements that you have configured in your Flexera environment. Workload Optimization Manager then creates groups and policies to represent these licenses and entitlements so that Workload Optimization Manager can optimize license usage and deliver continuous compliance.

- For each license, Workload Optimization Manager creates a dynamic group.

  These groups contain the entities that are affected by a Flexera license. For example, Workload Optimization Manager might create one group of all the VMs affected by a license, and another group of all the host machines affected by that license.

  These groups are dynamic. If you allocate or remove entities (workloads, hosts, and so on) under a given license, Workload Optimization Manager adds or removes those entities in the given group. Even if you have a license with no entities allocated to it, Workload Optimization Manager creates an empty group for that license.

  The name of each group begins with `Flexera`. The group name then includes the name of the license, plus the entity type. For example, you might see the groups, `Flexera: Microsoft SQL Server 2016 Standard VMs` and `Flexera: Microsoft SQL Server 2016 Standard Hosts`.

- For supported licenses, Workload Optimization Manager creates Placement and Automation policies.

  A Flexera license can include placement constraints, or resource utilization constraints. Workload Optimization Manager can create policies to comply with those constraints:

  – Placement Policies

    When calculating workload placement, Workload Optimization Manager uses placement policies to comply with constraints such as keeping workloads on licensed hosts, or clearing unlicensed workloads off of licensed hosts when necessary.

    For more information, see Placement Policies *(on page 1043)*.

  – Automation Policies

    Among other settings, these policies can specify operational, utilization, and scaling constraints to match the constraints in your Flexera license policies.

    For more information, see Automation Policies *(on page 1049)*.

  The name for each policy begins with the name `Flexera`. The policy name then includes the name of the license and constraint. For example, you might see a policy named `Flexera: Microsoft SQL Server 2016 Standard restrict to hosts`.

  Workload Optimization Manager then applies these policies to the groups that correspond to the entity type and license name.

Prerequisites

For Flexera One:

- Flexera One SaaS address or fully qualified domain name (FQDN) and Organization ID
- OAuth API refresh token created in Flexera One for Workload Optimization Manager use

For FlexNet Manager:

- FlexNet Manager Suite 2023 R2 or above
- Address or fully qualified domain name (FQDN) of FlexNet Manager Suite instance
- Local Windows Server username and password of an account Workload Optimization Manager can use

Workload Optimization Manager policies for Flexera licenses

Workload Optimization Manager generates policies to comply with the following types of Flexera licenses, where **Usage Rights** lists the configurations Workload Optimization Manager considers when creating policies:

License Types	Example Product	Usage Rights	Summary of auto-created policy	Effect of policy
Microsoft Server Processor	Microsoft Windows Server 2012 RedHat Enterprise Linux	Datacenter Edition `"coverInstallsOnVirtualMachines": true, "limitNumberOfVirtualInstalls": false, "limitVirtualInstallsIncludesHost": true, "numberOfAllowedVirtualInstalls": 1`	Placement policy that limits move actions for Microsoft Server Processor Datacenter licensed VM to licensed hosts	The hosts are licensed based on their number of **processors** and all licensed VMs should run on the licensed hosts.
		Standard Edition `"coverInstallsOnVirtualMachines": true, "limitNumberOfVirtualInstalls": true, "limitVirtualInstallsIncludesHost": false, "numberOfAllowedVirtualInstalls": 1`	Automation policy to disable move actions for Microsoft Server Processor Standard licensed VMs that have a coverage count set	The number of VMs and the number host processors determine the license consumption.
Microsoft Server/Management Core	Windows Server 2016 and later	Datacenter Edition `"coverInstallsOnVirtualMachines": true, "limitNumberOfVirtualInstalls": false, "limitVirtualInstallsIncludesHost": false`	Placement policy that limits on-prem VM moves to the licensed host.	The hosts are licensed based on their number of cores and all licensed VMs should run on the licensed hosts.
		Standard Edition `"coverInstallsOnVirtualMachines": true, "limitNumberOfVirtualInstalls": true, "limitVirtualInstallsIncludesHost": false, "minimumNumberOfProcessors": 2`	Automation policy that disables VM move actions.	The number of VMs and the number host cores determine the license consumption.
Microsoft Server Core	Microsoft SQL Server 2016	`"coverInstallsOnVirtualMachines": false, "limitNumberOfVirtualInstalls": false,`	Workload Optimization Manager analysis assumes licensing is based on the host.	If licensing is by host, the hosst are licensed based on a combination of processors/cores and all VMs that are running

License Types	Example Product	Usage Rights	Summary of auto-created policy	Effect of policy
		<pre>"limitVirtualInstallsIncludesHost": true, "minimumNumberOfThreadsPerVM": 4</pre>	If the licenseType is Microsoft Server Core and `virtualization.calculationType` is either `unlimitedVMs` or `fullCapacity`, a placement policy that limit on-prem VM moves to the licensed host is created. If `minimumNumberOfThreadsPerVM` is set, an automation policy is created to enforce that minimum vCPU threshold for VMs.	on those hosts are covered. If licensed by individual VMs, the VMs are licensed based on the number of cores. A minimum number of cores for each VM are assumed in the calculation.

Oracle Processor	Oracle Database Enterprise	<pre>"rules": {
		"virtualization": {
		"calculationType": "limitedVMs",
		"hosts": {
		"virtualMachines": {
		"coverage": {
		"count": 20,
		"includeHost": false
		},
		"cores": {
		"minCount": 16,
		"minCountPerProcessor": 8
		}
		}
		}
		}
		}</pre>
			If `licenseType` is Oracle Processor and `coverage.count` is not set,a placement policy is created that limits VM move actions on the **licensed** hosts. Otherwise, the policy is not created.	The number of VMs and the number host processors determine the license consumption.

Flexera permissions

The account that Workload Optimization Manager uses to connect with a Flexera target must be a Read-Only account with the following roles:

- `View IT assets`
- `call APIs`

Adding Flexera targets

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **FlexNet Manager** or **FlexeraONE**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:

For Flexera One:

- Display name

  Specify a name that uniquely identifies this connection.

  This name is for display purposes only and does not need to match any name in Flexera One.
- Address

  Specify the fully qualified domain name or IP address of the Flexera One platform.
- Organization ID

  Specify the unique identifier that defines your organization within Flexera One.
- Refresh token

  Specify the API refresh token that Workload Optimization Manager uses to generate access tokens for OAuth authentication. Learn more
- Connect using HTTPS

  If you select this option, Workload Optimization Manager connects to the target servers using HTTPS. Make sure that the required certificate is configured for use on the host.

  HTTPS is enabled by default and cannot be modified.
- Proxy configuration

  Specify proxy information only if you connect through a proxy.
  - Proxy hostname or IP address

    Specify the address of the proxy used for this target.
  - Proxy port

    Specify the port to use with the proxy used for this target. By default, the port is 8080.
  - Proxy username

    Specify the username to use with the proxy.
  - Proxy password

    Specify the password to use with the proxy.
  - Connect using HTTPS

    Select this option if Workload Optimization Manager will connect to the proxy through HTTPS.

For FlexNet Manager:

- Display name

  Specify a name that uniquely identifies this connection.

  This name is for display purposes only and does not need to match any name in FlexNet Manager.
- Address

  Specify the fully qualified domain name or IP address of the FlexNet Manager Suite platform.
- Username

  Specify the local Windows Server username. For example, `domain\user.name`.
- Password

  Specify the password of the account Workload Optimization Manager uses to connect to the target.
- Connect using HTTPS

  If you select this option, Workload Optimization Manager connects to the target servers using HTTPS. Make sure that the required certificate is configured for use on the host.

  This option is selected by default.
- Validate server certificates

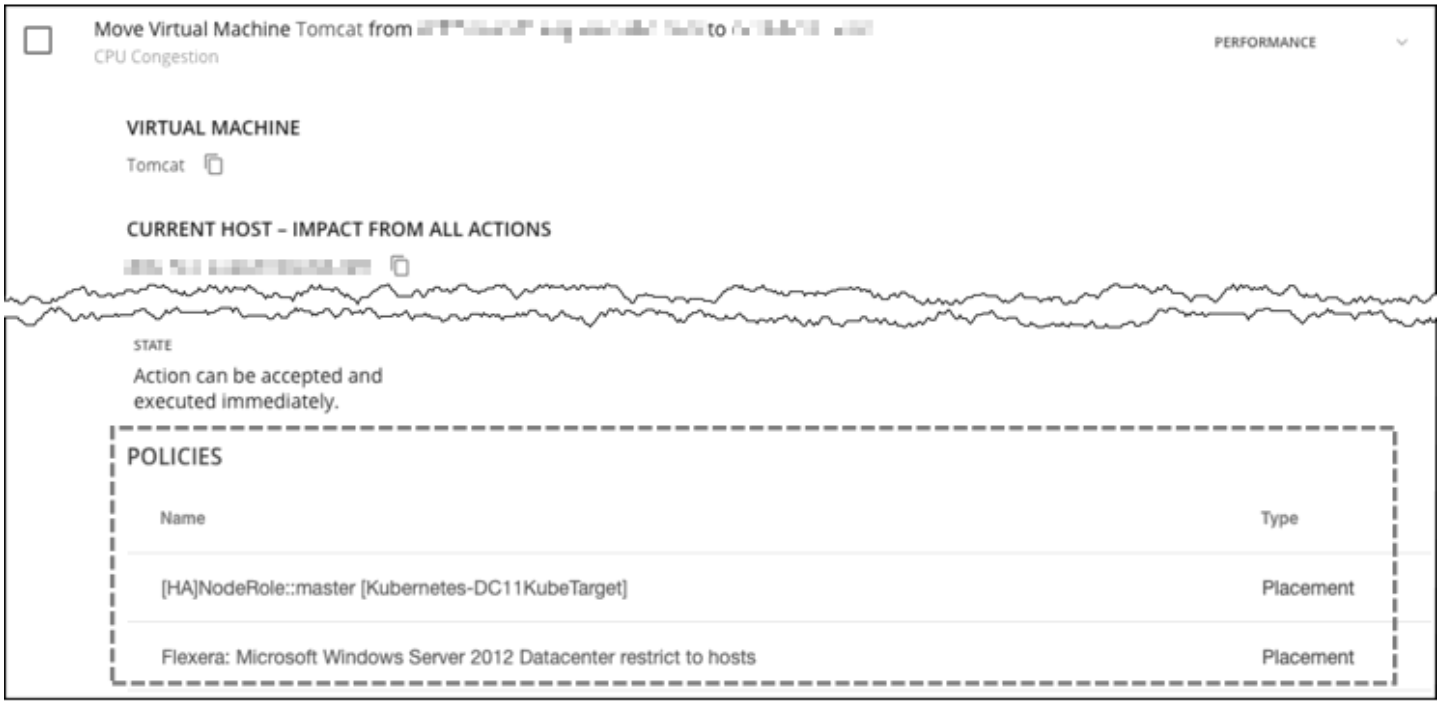
If you select this option, Workload Optimization Manager verifies the target certificate and proxy, if in use.

This option is selected by default.

Actions

Workload Optimization Manager does not generate unique actions for entities it discovers through a Flexera target. Instead, it uses the license specifications it discovers in Flexera to prevent actions that do not comply with license restrictions. It can also modify or generate actions to ensure license compliance.

When Workload Optimization Manager generates an action, the action details include a list of policies that constrain the action recommendation. In this list, policy names that begin with `Flexera:` indicate policies that implement the license constraints Workload Optimization Manager has discovered for Flexera. You can click to navigate to the policy definition. Note that you can see the policy definition, and you can only enable and disable it. The status remains the same across target and policy rediscoveries, and disabled policies do not affect actions.



The policies that Workload Optimization Manager generates affect Workload Optimization Manager actions in the following ways:

Entity	Policy Impact on Actions
Virtual Machine	▪ Resize vCPU Limit Resize actions to stay above a specified minimum. Workload Optimization Manager will not recommend a resize that falls below a specified minimum. However, if a VM is already below the specified minimum and shows no performance problems, it will not recommend a resize up. ▪ Move Workload Optimization Manager restricts licensed VMs to their matching licensed Hosts. If the host has capacity, it can move unlicensed VMs onto a licensed

Entity	Policy Impact on Actions
	host. However, if the host sees bottlenecks, Workload Optimization Manager will move unlicensed VMs off of the host first. Move actions can also consolidate license consumption on hosts. Assume you have one host that supports two licenses (LA and LB), and another host that only supports one license (LA). Also assume a mix of workloads, some that require both LA and LB, and some that only require LA. Move actions can consolidate the LA workloads on the LA host, leaving more resources for workloads that also require LB.
Host	▪ Provision Host Workload Optimization Manager can recommend to provision a new licensed host. For example, assume your environment includes hosts that support two different licenses (LA and LB). Now assume you add more workloads that require only LB licenses. Workload Optimization Manager can recommend that you provision a new host that supports the LB license. ▪ Reconfigure Host Workload Optimization Manager can recommend Host Reconfigure actions to change the licensing on a host. Assume Workload Optimization Manager has executed actions to consolidate LB-only workloads on a host with LA and LB licenses. Workload Optimization Manager can recommend that you reconfigure that host to remove LA, and only support the LB license.

ServiceNow

You can configure Workload Optimization Manager policies that log Workload Optimization Manager actions in your ServiceNow instance, and that submit actions for approval in ServiceNow workflows.

For more information concerning orchestration, see Automation Workflow *(on page 1063)*.

NOTE:
When creating the action orchestration policy as explained in the previous section, the scope of the policy must match the scope of the ServiceNow target.

Prerequisites

- A ServiceNow user with the `web_service_admin` role and the custom role `x_turbo_turbonomic.user` that is created during installation that can communicate with Workload Optimization Manager via the REST API

Adding ServiceNow targets

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **ServiceNow**.
4. In the side panel, review the connection requirements and then click **Connect Target**.

5. Configure the following settings:
    - Display name

      Specify a name that uniquely identifies this connection.
    - Address

      Specify the hostname of the ServiceNow instance without the `http` or `https` protocols. For example, `dev-env-266.service-now.com`.
    - Username
      Specify the username for the account Workload Optimization Manager uses to connect to the ServiceNow instance.
    - Password

      Specify the password for the account Workload Optimization Manager uses to connect to the ServiceNow instance.
    - Use OAuth to communicate with the ServiceNow target

      If this option is checked, Workload Optimization Manager uses OAuth to communicate with the ServiceNow target. If this option is not checked, Workload Optimization Manager uses basic authentication.

      Currently, Workload Optimization Manager supports OAuth 2.0 when this protocol is enabled in the ServiceNow target.
    - Client ID

      Specify the client ID Workload Optimization Manager uses if `Use OAuth to communicate with the ServiceNow target` is checked.
    - Client secret

      Specify the password Workload Optimization Manager uses if `Use OAuth to communicate with the ServiceNow target` is checked.
    - Port number

      Specify the port used to access the ServiceNow instance. By default, the port number is 443.
    - Proxy configuration

      Specify proxy information only if you connect through a proxy.
        - Proxy hostname or IP address

          Specify the address of the proxy used for this target.
        - Proxy port

          Specify the port to use with the proxy used for this target. By default, the port is 8080.
        - Proxy username

          Specify the username to use with the proxy.
        - Proxy password

          Specify the password to use with the proxy.
        - Connect using HTTPS

          Select this option if Workload Optimization Manager will connect to the proxy through HTTPS.

ServiceNow integration

In order to complete target addition, see the Workload Optimization Manager Actions for ServiceNow documentation.

Orchestrator targets

With orchestrator targets, you can integrate Workload Optimization Manager actions with the orchestrator's application management process. Workload Optimization Manager supports Action Script orchestrator targets.

About orchestrators

Orchestration targets assign workflows that execute multiple actions to make changes in your environment. Workload Optimization Manager discovers workflows that you have defined on the orchestrator. You can then set up an automation policy that maps workflows to actions. If the action acceptance mode is *Manual* or *Automated*, then when Workload Optimization Manager recommends the action, it will direct the orchestrator to use the mapped workflow to execute it.

Action Script

Workload Optimization Manager uses remote servers to execute Action Script orchestration. Managing the processes remotely means that you do not install custom code on the Workload Optimization Manager server. This eliminates associated security risks at that point. However, you are responsible to maintain the security of your Action Script Server, and to ensure the integrity of your custom code.

Resource requirements for the server

The remote server can be a VM or a container. The capacity you configure for the server depends entirely on the processes you intend to run on it. Workload Optimization Manager does not impose any special resource requirements on the server.

Configuring command execution

To support execution of your scripts, you must install any software that is necessary to run the scripts, including libraries, language processors, or other processes that your scripts will call.

Workload Optimization Manager calls the scripts as commands on the server. The server must run an SSH service that you configured to support command execution and SFTP operations. Currently, Cisco tested action scripts with the OpenSSH sshd daemon.

The standard port for SSH is 22. You can configure a different port, and provide that for admins who configure the server as an Action Script target.

> **NOTE:**
> An action script can invoke any process you deployed on the remote server. You do not have to run scripts; however, you must be able to invoke the processes from the command line. The script manifest gives Workload Optimization Manager the details it needs to build the command line invocation of each script.

Configuring the action script user account

To run the scripts on your server, Workload Optimization Manager logs on using a user account that is authorized to run the scripts from the command line. You provide the user credentials when you configure the Action Script target. To support this interaction, the user account must meet the following requirements:

- Public Key

  The user must have an RSA public key in the `.ssh/authorized_keys` file. When you configure the Action Script target, you provide this as the Private Token for the target.

  > **NOTE:**
  > For the Action Script target to validate, the permissions on the `authorized_keys` file must be set to `600`.

- Security for the `.ssh` Directory

  The Action Script User should be the only user with authorized access. Set file permissions to `700`.

- Supported Shells

  The Action Script User shell can be either the Bourne shell (usually at `/bin/sh`) or the Bourne-Again shell (usually at `/bin/bash`). Workload Optimization Manager passes parameters as it invokes your scripts. At this time it only supports script execution through these shells.

Action script manifest

The Action Script Manifest is a file that declares an array of Script Objects for each script you want to expose to Workload Optimization Manager. You can create the manifest as either a JSON or a YAML file. You can save the Manifest file to any location on your server, so long as the Scripts User has access to that location, and has read and execute privileges. The location of the manifest is provided as part of the Action Script target configuration.

After Workload Optimization Manager validates the target, it then discovers these scripts and presents them in the Orchestration Policy user interface.

For details, see Deploying the Action Script Manifest *(on page 1073)*.

Handling action script timeouts

Workload Optimization Manager limits script execution to 30 minutes. If a script exceeds this limit, Workload Optimization Manager sends a `SIGTERM` to stop the execution of the process.

Note that Workload Optimization Manager does not make any other attempt to end a process. For example, you might implement the script so it traps the `SIGTERM` and continues to run. The process should end at the soonest safe opportunity. However, if the process does not end, then you must implement some way to stop it outside of Workload Optimization Manager. A runaway process continues to use its execution thread, which can block other processes (action scripts or primary processes) if there are no more threads in the pool.

Adding Action Script targets

The Action Script target can be added through the UI or the API. For details on adding the target using the API, see Adding Action Script Targets using the API *(on page 363)*.

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **Action Script**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Action Script server address

     Specify the IP address or FQDN of the script execution server.
   - Script path

     Specify the path on the script execution server to the Action Script manifest file.
   - SSH port

     Specify the port for the SSH connection. By default, the port is 22.
   - SSH user ID

     Specify the username for SSH access into the server.
   - SSH private token

     Specify the private key for the user that has the public key set up in the `.ssh/authorized_keys` file on the remote server.
   - Public host key

     Specify the public key that is presented by the SSH server for host authentication.

     If this value is not provided, the presented key is accepted into the target definition for future operations.

Adding Action Script targets using the API

The Action Script target can be added through the UI or the API. This topic outlines the steps for adding the target using the API. Refer to the Action Script Server *(on page 361)* topic for configuration information.

Obtaining private key string

In order to pass the private key as a string, newlines must be inserted appropriately with the `\n` character, as the JSON format will not preserve these required newlines. One method to insert these lines on *nix systems is the `awk` command. This command should follow the following format:

```
awk 'ORS="\\n"' path/to/keyfile > path/to/modified/keyfile
```

For example:

```
awk 'ORS="\\n"' ~/.ssh/id_rsa > /tmp/privkey.txt
```

You can then `cat` or `vi/vim` the file to copy the text and insert it into the API call.

Example private key file, before modification:

```
-----BEGIN RSA PRIVATE KEY-----
MIIJKQIBAAKCAgEAxeKUk6WLFMf8seI2KztL4DoKKw0d6AG/39SH6or5SiNbAGkM
Umv8ABA0aCQqMF01j9V3kzGn71DIWWTxLOMBH06I01dvd9XYPVRJotGjpxus+xmV
GbeXxJsdHfXrbtMeLnpkwGC7gIu6W2o5Ez6YuLDKf+CeFZhGMOtVLiB2NWtzn33D
KBh96tIom9EqKseovtQAj5VYDwPeuTaWX1RGMTl1tsNGyezEc4Z2sppT4VFVhsUV
L+voDZzrtMfFiHUIgjPH4pK9c/iPj0um7YzmyMyARftyCsUg7iUZH4NlNLpikXVy
xKiro6Quu7SZ4kcvr3ZF3970CON0hHzC3mFsvZvRFgYKzgc8/BcZ8ZzEUFfqBqY4
dejCU+jo86dYFL2Q8RfF2zsJenG0WbVRLtLGkc7RbMdBGjtt5roV5VAp+TyR9Hdd
Dc5FYvt4fAAMJtcmTIzdEAPSbVzDhkSajxRhOXvezJn+2Ddr34/9ZvFE7nJNgP/8
KOGJ5H1dgG5mc7H+Ly3x5dXzPhcVNSUdNGgKnEvzFgMJaGByAg5AZc5HOmiIwFKs
KSE/m8dSRoHrHmenGGmk1IMUnvCC1mKkbUEJ2V6Rl7x8zmM4DYNRSTwcSL
SZ7U/yDuwpHwzeozNeoekGKeC23+jgyzr9/mzgmPF08q5F+wVg/7nlDRSXsCAwEA
AQKCAgBPHMPbt25AM6gCEfcoF3/BOFMCXS8cpz5npRdlnIQHsNTb+CN1/Df2aoQZ
MoQva8RTKbuxfEOweLJwktuc33B+bK5lF+q3lSXrdvwtfPSZhYYFqXejnXOupXjh
loS2fO7vE/2dr86mmLY7f98ro3X5OKzsrQdmiMdfHWajl1TLDJdgJy5TpF8K91RG
3zuleMcs66MBxGo9LoXntDwwBog9e8Rk+fMeHWt46EQiKXLpfNPdmoNa8kyGzLqr
PsbpRCjPrRX3rkTRzYjPeCZXiGJn2uJZmCpCxgbhTrGZmLVofC+WCceFeg2M0cHh
R1XSUjVtR7KtyvvzZcQah6p8r6SCASQFeVbb13FJVkBd0J8Hzk5e4N5Qk3bv9V4W
F7CC8dRV66kcAkf68ytLGXkQAsgo5RPBDYSp8E2JDLF4m+9w6aouPlg//QxG5+XA
wqGJseTipiHbiALcIGsCpc5QCqab2SNddaDtIIpnp2Zmo5E1OjutGeUAE3k+cnJ1
JIFCkRL0vsPk0xODSXbV2PMaUjMT65ue7uBkDqCtgHiSBXTk6DiWBjduDnoZlAmC
I8M4o42mSPs2uoKcngJ2fboY5iSsXLQ8VT5Tix8fk7VzBFmOvchBUSFlyx6o3/0l
0SV3Qrqz7tb1Xx+axQapIUEVbLV14OkOjepvxF3HVELXRQd2eQKCAQEA8EzXro5j
N/VgDQJ9pGphwUyQ9mNBC1c1jviAHHskh8gwa4BxhBLzZ+aUUoIx0/mAYK0rihN+
QJLI84c9tA62Au6Tq+CT60ad4SwxQ5m9mQZjq888T3dRssmY+f+5kYBzmaPHvl8I
Sohh2tHqNh68T1KmLbCABahc0ppqJTjP0Oi0kR7XUk5gEeDNNb65oDOtuNytiVd4
cNEq/cO8rqk2rTupLLjhVKzP9lgVXL/SBw/HBjpMyJ3DMrcPXH84JeMshJay/OnW
eY3Y+hsghw37zRxeuUtn5to0wVYayAmEhyXVing7EVV5wsyUaJQE25ZZWLjjiBHP
ZpLsSRc++uugnwKCAQEA0tBSQrDdOH+O6Ydu9Tno1IUeilTQ8+VGx3esFlKtiDTW
cSEZSfZKbf/GQgHKbxLVXjCLhbfPbCvCcojbmGSmtWRWxlySYPSW4tBiYJSk6Fo8
J2hEYW3HQzyj23RjXIghCUqY5wNWj1brU4CRxHtMyV1eYJWGmTaBfDm8VoBghu9g
FAyPkJs4WZUkUI7Sbg2vcJyLLP4XETPlfACMptLrPs8RCyUjJAJYxpSdaSjLC7r9
9gBusQOlIGpClrx+/86NYhdKDWmnd0lyjXNvMVwPIpoSVblXTZK7xRBgcNTsd2eh
h1HLuhLXZ76SplvtNH6MQP46WoblounEx7ONnsddpQKCAQEA1lVD339DQktLo+7/
0CHPFZHMjueSwaHEtMQzyLORrsbBYXGMTqh0kM0aqMlV1mNpvVFYT2z5pHEIaibL
w4mHkO/pqVnDtgwjkA9R1aEQeqg3SKysd5jOtzCXkQ0mIIC5G2X5KHUtOQWSepHX
XPRnND1c2c3fMqysNnDiklatDgIx5Grk1ZXwnKGth+2ZNipCHmw/IQxy8fWvae8n
oYZurW9Ugv5YhaUqTOjeDP9PP1s0QvrHN/13eQnM2szicPDC6XRdK7cKxRrfWJ3P
kobHfHh8iLuYblFc0sGRW7EZk04MvNhmnxxlYXoGOY0lFBNzl+45jEmOe0g7Fypz
021NswKCAQAHx/6xTXvs4PByeDhFPuz0IzG8Y2MPDZBuPmGQtV6IX4LmzBT14Oxg
YQd9WjmRnf1YK6yKs/0o/fiAZ8LLPjPVEC4Pxkz03TJQjuaCsBniJaqAvd2ARyAK
```

```
wEtAB2y/dprZj/JS3JGuWsP24hD/UpnD+PlJ0VsVfqiK4EquWrC1alJ+ivQln0UW
MWYUvyDPRRwLqQuODpblYKgAyrhmU/Sw7c2wPp+sbZW/Rq8Euae8+Br1oJB1YjAP
ITzx+WUqpgIb88Yc9qBiy8g5pmAdT0uTuGIi3eN2Dot1u6QDpihQQ/f3WbKOXLGs
t6EMGOqJfH5Ky9d1jfMd70j/IgaBLNStAoIBAQCKqJNuH7FSta/KQ5fJzFrD/Mil
Q6yfB9D92UEabBlEt22Zr7Ry5WAv4DRcYWVlHrBJQBOw4Gc5WZcEnE1HvhkX519w
hccp1P+tyRf+h6ppYqdl/YaW2vEMTR1gvvtV8+2aMYTNRQJuTjJIsk1vWvX42IhQ
xqKM9xqmKGEsW+6e7UheaXY8afkjf20/euGgzTuEPoy+IA8vKrSAnLLWm4xDYjMV
i9PFbuSyRV6tJqQZxZ58dS2ykL5YpTLM+ZFyZmhAqSJWULpsKAxxE2fnmniXOYVF
E4VxEPsgNYH1k2eqtGRQwYR0ML/fihHbJIlQnMr6o7/tKljx+3PCZeRaVR4/
-----END RSA PRIVATE KEY-----
```

The same private key, after running the preceding command:

```
-----BEGIN RSA PRIVATE KEY-----\n
MIIJKQIBAAKCAgEAxeKUk6WLFMf8seI2KztL4DoKKw0d6AG/39SH6or5SiNbAGkM\n
Umv8ABA0aCQqMF01j9V3kzGn71DIWWTxLOMBH06I01dvd9XYPVRJotGjpxus+xmV\n
GbeXxJsdHfXrbtMeLnpkwGC7gIu6W2o5Ez6YuLDKf+CeFZhGMOtVLiB2NWtzn33D\n
KBh96tIom9EqKseovtQAj5VYDwPeuTaWX1RGMTl1tsNGyezEc4Z2sppT4VFVhsUV\n
L+voDZzrtMfFiHUIgjPH4pK9c/iPj0um7YzmyMyARftyCsUg7iUZH4NlNLpikXVy\n
xKiro6Quu7SZ4kcvr3ZF3970CON0hHzC3mFsvZvRFgYKzgc8/BcZ8ZzEUFfqBqY4\n
dejCU+jo86dYFL2Q8RfF2zsJenG0WbVRLtLGkc7RbMdBGjtt5roV5VAp+TyR9Hdd\n
Dc5FYvt4fAAMJtcmTIzdEAPSbVzDhkSajxRhOXvezJn+2Ddr34/9ZvFE7nJNgP/8\n
KOGJ5H1dgG5mc7H+Ly3x5dXzPhcVNSUdNGgKnEvzFgMJaGByAg5AZc5HOmiIwFKs\n
KSE/m8dSRoHrHmenGGmk1IMUnvCC1mKkbUEJ2V6Rl7x8zmM4DYNRSTwcSL\n
SZ7U/yDuwpHwzeozNeoekGKeC23+jgyzr9/mzgmPF08q5F+wVg/7nlDRSXsCAwEA\n
AQKCAgBPHMPbt25AM6gCEfcoF3/BOFMCXS8cpz5npRdlnIQHsNTb+CN1/Df2aoQZ\n
MoQva8RTKbuxfEOweLJwktuc33B+bK5lF+q3lSXrdvwtfPSZhYYFqXejnXOupXjh\n
loS2fO7vE/2dr86mmLY7f98ro3X5OKzsrQdmiMdfHWajl1TLDJdgJy5TpF8K91RG\n
3zuleMcs66MBxGo9LoXntDwwBog9e8Rk+fMeHWt46EQiKXLpfNPdmoNa8kyGzLqr\n
PsbpRCjPrRX3rkTRzYjPeCZXiGJn2uJZmCpCxgbhTrGZmLVofC+WCceFeg2M0cHh\n
R1XSUjVtR7KtyvvzZcQah6p8r6SCASQFeVbb13FJVkBd0J8Hzk5e4N5Qk3bv9V4W\n
F7CC8dRV66kcAkf68ytLGXkQAsgo5RPBDYSp8E2JDLF4m+9w6aouPlg//QxG5+XA\n
wqGJseTipiHbiALcIGsCpc5QCqab2SNddaDtIIpnp2Zmo5E1OjutGeUAE3k+cnJ1\n
JIFCkRL0vsPk0xODSXbV2PMaUjMT65ue7uBkDqCtgHiSBXTk6DiWBjduDnoZlAmC\n
I8M4o42mSPs2uoKcngJ2fboY5iSsXLQ8VT5Tix8fk7VzBFmOvchBUSFlyx6o3/0l\n
0SV3Qrqz7tb1Xx+axQapIUEVbLV14OkOjepvxF3HVELXRQd2eQKCAQEA8EzXro5j\n
N/VgDQJ9pGphwUyQ9mNBC1c1jviAHHskh8gwa4BxhBLzZ+aUUoIx0/mAYK0rihN+\n
QJLI84c9tA62Au6Tq+CT60ad4SwxQ5m9mQZjq888T3dRssmY+f+5kYBzmaPHvl8I\n
Sohh2tHqNh68T1KmLbCABahc0ppqJTjP0Oi0kR7XUk5gEeDNNb65oDOtuNytiVd4\n
cNEq/cO8rqk2rTupLLjhVKzP9lgVXL/SBw/HBjpMyJ3DMrcPXH84JeMshJay/OnW\n
eY3Y+hsghw37zRxeuUtn5to0wVYayAmEhyXVing7EVV5wsyUaJQE25ZZWLjjiBHP\n
ZpLsSRc++uugnwKCAQEA0tBSQrDdOH+O6Ydu9Tno1IUeilTQ8+VGx3esFlKtiDTW\n
cSEZSfZKbf/GQgHKbxLVXjCLhbfPbCvCcojbmGSmtWRWxlySYPSW4tBiYJSk6Fo8\n
J2hEYW3HQzyj23RjXIghCUqY5wNWj1brU4CRxHtMyV1eYJWGmTaBfDm8VoBghu9g\n
FAyPkJs4WZUkUI7Sbg2vcJyLLP4XETPlfACMptLrPs8RCyUjJAJYxpSdaSjLC7r9\n
9gBusQOlIGpClrx+/86NYhdKDWmnd0lyjXNvMVwPIpoSVblXTZK7xRBgcNTsd2eh\n
h1HLuhLXZ76SplvtNH6MQP46WoblounEx7ONnsddpQKCAQEA1lVD339DQktLo+7/\n
0CHPFZHMjueSwaHEtMQzyLORrsbBYXGMTqh0kM0aqMlV1mNpvVFYT2z5pHEIaibL\n
w4mHkO/pqVnDtgwjkA9R1aEQeqg3SKysd5jOtzCXkQ0mIIC5G2X5KHUtOQWSepHX\n
XPRnND1c2c3fMqysNnDiklatDgIx5Grk1ZXwnKGth+2ZNipCHmw/IQxy8fWvae8n\n
oYZurW9Ugv5YhaUqTOjeDP9PP1s0QvrHN/13eQnM2szicPDC6XRdK7cKxRrfWJ3P\n
kobHfHh8iLuYblFc0sGRW7EZk04MvNhmnxxlYXoGOY0lFBNzl+45jEmOe0g7Fypz\n
021NswKCAQAHx/6xTXvs4PByeDhFPuz0IzG8Y2MPDZBuPmGQtV6IX4LmzBT14Oxg\n
YQd9WjmRnf1YK6yKs/0o/fiAZ8LLPjPVEC4Pxkz03TJQjuaCsBniJaqAvd2ARyAK\n
wEtAB2y/dprZj/JS3JGuWsP24hD/UpnD+PlJ0VsVfqiK4EquWrC1alJ+ivQln0UW\n
MWYUvyDPRRwLqQuODpblYKgAyrhmU/Sw7c2wPp+sbZW/Rq8Euae8+Br1oJB1YjAP\n
ITzx+WUqpgIb88Yc9qBiy8g5pmAdT0uTuGIi3eN2Dot1u6QDpihQQ/f3WbKOXLGs\n
t6EMGOqJfH5Ky9d1jfMd70j/IgaBLNStAoIBAQCKqJNuH7FSta/KQ5fJzFrD/Mil\n
Q6yfB9D92UEabBlEt22Zr7Ry5WAv4DRcYWVlHrBJQBOw4Gc5WZcEnE1HvhkX519w\n
hccp1P+tyRf+h6ppYqdl/YaW2vEMTR1gvvtV8+2aMYTNRQJuTjJIsk1vWvX42IhQ\n
xqKM9xqmKGEsW+6e7UheaXY8afkjf20/euGgzTuEPoy+IA8vKrSAnLLWm4xDYjMV\n
i9PFbuSyRV6tJqQZxZ58dS2ykL5YpTLM+ZFyZmhAqSJWULpsKAxxE2fnmniXOYVF\n
```

```
        E4VxEPsgNYH1k2eqtGRQwYR0ML/fihHbJIlQnMr6o7/tKljx+3PCZeRaVR4/\n
        -----END RSA PRIVATE KEY-----\n
```

Adding Action Script targets

The TargetApiInputDTO for this target has the following parameters:

- `nameOrAddress`

  IP or FQDN of the script execution server.
- `port`

  Port used to connect to the script execution server.
- `userid`

  Username used to connect to the script execution server.
- `privateKeyString`

  Path to the SSH private token corresponding to the user used to connect to the script execution server.
- `manifestPath`

  Path to the Action Script manifest file on the script execution server.

Use the `POST https://10.10.10.10/api/v3/targets` request to add this target. Send the constructed TargetApiInputDTO as the body of the request.

Example Input:

```
{
  "category":"Orchestrator",
  "inputFields":[
    {
      "name":"nameOrAddress",
      "value":"10.10.10.10"
    },
    {
      "name":"port",
      "value":"22"
    },
    {
      "name":"userid",
      "value":"shepp"
    },
    {
      "name":"privateKeyString",
      "value":"-----BEGIN RSA PRIVATE
 KEY-----\nMIIJKQIBAAKCAgEAxeKUk6WLFMf8seI2KztL4DoKKw0d6AG/39SH6or5SiNbAGkM\nUmv8ABA0aCQqMF01j9V3kzGn71DIW
WTxLOMBH06I01dvd9XYPVRJotGjpxus+xmV\nGbeXxJsdHfXrbtMeLnpkwGC7gIu6W2o5Ez6YuLDKf+CeFZhGMOtVLiB2NWtzn33D\nKBh
96tIom9EqKseovtQAj5VYDwPeuTaWX1RGMTl1tsNGyezEc4Z2sppT4VFVhsUV\nL
+voDZzrtMfFiHUIgjPH4pK9c/iPj0um7YzmyMyARftyCsUg7iUZH4NlNLpikXVy\nxKiro6Quu7SZ4kcvr3ZF3970CON0hHzC3mFsvZvRF
gYKzgc8/BcZ8ZzEUFfqBqY4\ndejCU+jo86dYFL2Q8RfF2zsJenG0WbVRLtLGkc7RbMdBGjtt5roV5VAp+TyR9Hdd\nDc5FYvt4fAAMJtc
mTIzdEAPSbVzDhkSajxRhOXvezJn+2Ddr34/9ZvFE7nJNgP/8\nKOGJ5H1dgG5mc7H+Ly3x5dXzPhcVNSUdNGgKnEvzFgMJaGByAg5AZc5
HOmiIwFKs\nKSE/m8dSRoHrHmenGGm5TtA4lu1IMUnvCC1mKkbUEJ2V6Rl7x8zmM4DYNRSTwcSL\nSZ7U/yDuwpHwzeozNeoekGKeC23+j
gyzr9/mzgmPF08q5F+wVg/7nlDRSXsCAwEA\nAQKCAgBPHMPbt25AM6gCEfcoF3/BOFMCXS8cpz5npRdlnIQHsNTb+CN1/Df2aoQZ\nMoQ
va8RTKbuxfEOweLJwktuc33B+bK5lF+q3lSXrdvwtfPSZhYYFqXejnXOupXjh\nloS2fO7vE/2dr86mmLY7f98ro3X5OKzsrQdmiMdfHWa
jl1TLDJdgJy5TpF8K91RG\n3zuleMcs66MBxGo9LoXntDwwBog9e8Rk+fMeHWt46EQiKXLpfNPdmoNa8kyGzLqr\nPsbpRCjPrRX3rkTRz
YjPeCZXiGJn2uJZmCpCxgbhTrGZmLVofC+WCceFeg2M0cHh\nR1XSUjVtR7KtyvvzZcQah6p8r6SCASQFeVbb13FJVkBd0J8Hzk5e4N5Qk
3bv9V4W\nF7CC8dRV66kcAkf68ytLGXkQAsgo5RPBDYSp8E2JDLF4m+9w6aouPlg//
QxG5+XA\nwqGJseTipiHbiALcIGsCpc5QCqab2SNddaDtIIpnp2Zmo5E1OjutGeUAE3k+cnJ1\nJIFCkRL0vsPk0xODSXbV2PMaUjMT65u
e7uBkDqCtgHiSBXTk6DiWBjduDnoZlAmC\nI8M4o42mSPs2uoKcngJ2fboY5iSsXLQ8VT5Tix8fk7VzBFmOvchBUSFlyx6o3/0l
\n0SV3Qrqz7tb1Xx+axQapIUEVbLV14OkOjepvxF3HVELXRQd2eQKCAQEA8EzXro5j\nN/VgDQJ9pGphwUyQ9mNBC1c1jviAHHskh8gwa4
BxhBLzZ+aUUoIx0/mAYK0rihN+\nQJLI84c9tA62Au6Tq+CT60ad4SwxQ5m9mQZjq888T3dRssmY+f
```

```
+5kYBzmaPHvl8I\nSohh2tHqNh68T1KmLbCABahc0ppqJTjP0Oi0kR7XUk5gEeDNNb65oDOtuNytiVd4\ncNEq/cO8rqk2rTupLLjhVKzP
9lgVXL/SBw/HBjpMyJ3DMrcPXH84JeMshJay/OnW
\neY3Y+hsghw37zRxeuUtn5to0wVYayAmEhyXVing7EVV5wsyUaJQE25ZZWLjjiBHP\nZpLsSRc+
+uugnwKCAQEA0tBSQrDdOH+O6Ydu9Tno1IUeilTQ8+VGx3esFlKtiDTW\ncSEZSfZKbf/GQgHKbxLVXjCLhbfPbCvCcojbmGSmtWRWxlyS
YPSW4tBiYJSk6Fo8\nJ2hEYW3HQzyj23RjXIghCUqY5wNWj1brU4CRxHtMyV1eYJWGmTaBfDm8VoBghu9g\nFAyPkJs4WZUkUI7Sbg2vcJ
yLLP4XETPlfACMptLrPs8RCyUjJAJYxpSdaSjLC7r9\n9gBusQOlIGpClrx+/86NYhdKDWmnd0lyjXNvMVwPIpoSVblXTZK7xRBgcNTsd
2eh\nh1HLuhLXZ76SplvtNH6MQP46WoblounEx7ONnsddpQKCAQEA1lVD339DQktLo+7/\n0CHPFZHMjueSwaHEtMQzyLORrsbBYXGMTqh
0kM0aqMlV1mNpvVFYT2z5pHEIaibL\nw4mHkO/pqVnDtgwjkA9R1aEQeqg3SKysd5jOtzCXkQ0mIIC5G2X5KHUtOQWSepHX\nXPRnND1c2
c3fMqysNnDiklatDgIx5Grk1ZXwnKGth+2ZNipCHmw/IQxy8fWvae8n\noYZurW9Ugv5YhaUqTOjeDP9PP1s0QvrHN/13eQnM2szicPDC6
XRdK7cKxRrfWJ3P\nkobHfHh8iLuYblFc0sGRW7EZk04MvNhmnxxlYXoGOY0lFBNzl+45jEmOe0g7Fypz\n021NswKCAQAHx/6xTXvs4PB
yeDhFPuz0IzG8Y2MPDZBuPmGQtV6IX4LmzBT14Oxg\nYQd9WjmRnf1YK6yKs/0o/fiAZ8LLPjPVEC4Pxkz03TJQjuaCsBniJaqAvd2AR
yAK\nwEtAB2y/dprZj/JS3JGuWsP24hD/UpnD+PlJ0VsVfqiK4EquWrC1alJ+ivQln0UW\nMWYUvyDPRRwLqQuODpblYKgAyrhmU/Sw7c2
wPp+sbZW/Rq8Euae8+Br1oJB1YjAP\nITzx+WUqpgIb88Yc9qBiy8g5pmAdT0uTuGIi3eN2Dot1u6QDpihQQ/f3WbKOXLGs\nt6EMGOqJf
H5Ky9d1jfMd70j/IgaBLNStAoIBAQCKqJNuH7FSta/KQ5fJzFrD/Mil
\nQ6yfB9D92UEabBlEt22Zr7Ry5WAv4DRcYWVlHrBJQBOw4Gc5WZcEnE1HvhkX519w\nhccp1P+tyRf+h6ppYqdl/YaW2vEMTR1gvvtV8+
2aMYTNRQJuTjJIsk1vWvX42IhQ\nxqKM9xqmKGEsW+6e7UheaXY8afkjf20/euGgzTuEPoy+IA8vKrSAnLLWm4xDYjMV\ni9PFbuSyRV6t
JqQZxZ58dS2ykL5YpTLM+ZFyZmhAqSJWULpsKAxxE2fnmniXOYVF\nE4VxEPsgNYH1k2eqtGRQwYR0ML/fihHbJIlQnMr6o7/tKljx+3PC
ZeRaVR4/\n-----END RSA PRIVATE KEY-----\n"
    },
    {
      "name":"manifestPath",
      "value":"/home/shepp/action-script/manifest.json"
    }
  ],
  "type":"Action Script"
}
```


The API will return the TargetApiDTO for the Action Script target upon success, or an error message upon failure.

Example Response:


```
{
  "uuid": "73356329749216",
  "displayName": "10.10.10.10-/home/shepp/action-script/manifest.json",
  "category": "Orchestrator",
  "inputFields": [
    {
      "displayName": "Name or Address",
      "name": "nameOrAddress",
      "value": "10.10.10.10",
      "isMandatory": true,
      "isSecret": false,
      "isTargetDisplayName": false,
      "valueType": "STRING",
      "description": "IP or FQDNS for the Script Execution Server",
      "verificationRegex": ".*"
    },
    {
      "displayName": "Port",
      "name": "port",
      "value": "22",
      "defaultValue": "22",
      "isMandatory": false,
      "isSecret": false,
      "isTargetDisplayName": false,
      "valueType": "STRING",
      "description": "Port to use for the Script Execution Server",
      "verificationRegex": ".*"
    },
    {
      "displayName": "User ID",
      "name": "userid",
```

```
      "value": "shepp",
      "isMandatory": true,
      "isSecret": false,
      "isTargetDisplayName": false,
      "valueType": "STRING",
      "description": "Userid to use to execute command on the Script Execution Server",
      "verificationRegex": ".*"
    },
    {
      "displayName": "Private Token",
      "name": "privateKeyString",
      "isMandatory": true,
      "isSecret": true,
      "isTargetDisplayName": false,
      "valueType": "STRING",
      "description": "SSH Private Token corresponding to the Userid",
      "verificationRegex": ".*"
    },
    {
      "displayName": "Script Path",
      "name": "manifestPath",
      "value": "/home/shepp/action-script/manifest.json",
      "isMandatory": true,
      "isSecret": false,
      "isTargetDisplayName": false,
      "valueType": "STRING",
      "description": "File Path to the ActionScript manifest file on the Execution Server",
      "verificationRegex": ".*"
    },
    {
      "displayName": "Public Host Key",
      "name": "hostKey",
      "isMandatory": false,
      "isSecret": false,
      "isTargetDisplayName": false,
      "valueType": "STRING",
      "description": "Public key presented by the SSH server for host authenticaion; if not provided, the
 presented key will be accepted and integrated into the target definition for future operations",
      "verificationRegex": ".*"
    }
  ],
  "lastValidated": "2020-07-28T20:46:22Z",
  "status": "Validated",
  "type": "Action Script",
  "readonly": false
}
```


```
Response headers
```

```
cache-control: no-cache, no-store, max-age=
```

Once the success response is received, the target is validated and ready for use.

Private cloud targets

A private cloud manages resources in pools to support multi-tenancy and self-service provisioning of virtual workloads. Workload Optimization Manager manages these resource pools in real time as demand fluctuates. This includes demand across resource pools, virtual datacenters (VDCs), and tenants.

Adding a Private Cloud target to Workload Optimization Manager aids in the goal of Application Resource Management (ARM) by preventing bottlenecks within and between your managed entities, resizing virtual machines and containers, prioritizing

managed workloads, and reducing storage latency throughout your private cloud, ensuring that applications running in your private cloud always have the resources they need to perform, when they need to perform.

On the private cloud, you can use Workload Optimization Manager to:

- Set up charge-back and show-back for private cloud or service-provider scenarios
- For service-providers, set up scoped views to limit exposure to the customer base
- Plan hardware requirements — the planning scenarios takes cloud architectures into account

Supply chain

For private clouds, Workload Optimization Manager discovers resource partitions that are managed by the cloud manager, the workloads running on these partitions (the VMs and applications), and the supply that hosts workload (the physical machines and storage). Workload Optimization Manager represents these partitions as the following types of Virtual Datacenters (VDCs):

- **Provider VDC**: A collection of physical resources (PMs and datastores) within a private cloud. The cloud administrator has access to these resources, and defines the datacenter members. Administrators allocate Provider VDCs to manage resources that will be allocated to external customers through one or more Consumer VDCs.
- **Consumer VDC**: A collection of resources that are available for customers to perform self-service management of workload through the cloud. It is an environment customers can use to store, deploy, and operate virtual systems. Consumer VDCs use the resources supplied by a Provider VDC.

NOTE:
Different targets use different names to refer to virtual data centers. In the Workload Optimization Manager supply chain, these entities are all represented by Consumer and Provider VDCs, as follows:

Workload Optimization Manager	vCenter Server	VMM
Consumer VDC	Resource Pool (Child)	Tenant or TenantQuota
Provider VDC	Resource Pool (Root)	Cloud

Microsoft Virtual Machine Manager

In a Microsoft Virtual Machine Manager (VMM) environment, the VMM management server processes commands and controls communications with Hyper-V hosts. To manage VMM, add the management server as a target. Workload Optimization Manager communicates with that target and the associated Hyper-V hosts. You do not need to add individual Hyper-V hosts as targets.

Prerequisites

- VMM for Microsoft System Center 2012 R2, System Center 2016, System Center 2019, or System Center 2022
- VMM Dynamic Optimization disabled (recommended) or set to Low aggressiveness, with a frequency of at least 60 minutes.
- Configure remote management on the VMM management Hyper-V server. See Enabling Windows Remote Management *(on page 343)*.
- Grant access on every host in the cluster that Workload Optimization Manager will manage to the following WMI namespaces and subspaces:
  - `root/virtualization`
  - `root/virtualization/v2`
  - `root/microsoft/windows/storage`
  - `root/microsoft/windows/smb`
  - `root/wmi`

  - `root/cimv2`
  - `root/mscluster`
- Apply necessary hot fixes on the VMM host.

  For example, if you are running VMM Server on a Windows Server version earlier than Windows Server 2012 R2, you must apply the hot-fix referenced in the Microsoft Knowledge Base article #2842230 (http://support.microsoft.com/kb/2842230).
- Apply time synchronization.

  The VM that hosts Workload Optimization Manager must be synchronized with each target VMM management server. For instructions on synchronizing the clock on the Workload Optimization Manager server, see Synchronizing Time *(on page 58)*.
- Enable PowerShell execution on the VMM management server.
- Configure port access.

  WinRM uses ports 5985 and 5986 for standard and secure communications, respectively. The firewall on your VMM server must open these ports.
- Delete preexisting Hyper-V targets, as needed.

  If any of the hosts that make up the VMM target were added separately as Hyper-V targets (and you do not plan to exclude these hosts), these targets must be deleted. Failure to do so creates duplicate entities in the market, which negatively impacts Workload Optimization Manager performance.

Adding VMM targets

Workload Optimization Manager uses the address and credentials you provide to discover the VMM target. From the VMM target, Workload Optimization Manager gets the list of managed Hyper-V instances. It then uses that list to discover each Hyper-V instance. The Hyper-V credentials you provide must be valid for all of these machines.

Workload Optimization Manager will also import your Availability Sets, representing them as placement policies for the affected infrastructure.

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **VMM**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Hostname or IP address

     Specify the IP address or hostname of the VMM management server.
   - Username

     Specify the username of a user account Workload Optimization Manager uses to connect to the VMM management server.
   - Password

     Specify the password for account used.
   - Fully qualified domain name

     Specify the fully qualified domain name for the user account. This domain name identifies the user account for Active Directory.
   - Port number

     Specify the port to use for the WSMan connection. By default, the HTTP port is 5985 and the HTTPS port is 5986.
   - Connect using HTTPS

If this option is selected, Workload Optimization Manager uses a secure connection (HTTPS). If you enable a secure connection, then you must configure a certificate, and you must configure Workload Optimization Manager to communicate over HTTPS. For more information, see Secure Setup for WSMan *(on page 372)*.

NOTE:
Setting a secure connection for VMM does not also set secure connections for the underlying Hyper-V hosts. Any communications between Workload Optimization Manager and VMM will be secure. To configure secure connections to the underlying Hyper-V hosts, you must specify secure connections on each one.

- Use Kerberos message-level encryption over HTTP

  If this option is selected, Workload Optimization Manager uses Kerberos message-level encryption to discover the environment. This value is ignored if "Secure connection" is enabled.

- Hyper-V username and password

  Specify the login credentials for the Hyper-V servers that are managed by the VMM target. If you leave the Hyper-V credentials blank, Workload Optimization Manager uses the same credentials that VMM uses. If you provide Hyper-V credentials, it uses that service account to log into every Hyper-V that is managed by the VMM.

  NOTE:
  The service account Workload Optimization Manager uses to log into a Hyper-V host must satisfy certain requirements. For more information, see Creating a Service User Account in Hyper-V *(on page 342)*.

- Exclude hosts

  Specify the hosts in the VMM target that you would like to exclude from Workload Optimization Manager.

  Enter multiple hosts in a comma-separated list. This field accepts both hostnames and fully qualified domain names. It also recognizes the * and ? wild-card characters. Workload Optimization Manager does not discover or manage any hosts you enter in this field. These hosts do not participate in the market, and Workload Optimization Manager does not generate actions that involve these hosts.

Microsoft VMM and Hyper-V VLANs

Windows Server Hyper-V provides support for VLANs on host and VM partitions. If your Hyper-V environment makes use of this VLAN support, then your VM moves must be sensitive to which hosts provide networking access to your defined VM networks. If a VM is a member of a given VM network, then any move of that VM must be to a host that has access to the same network.

For Hyper-V targets in a VMM environment, the Cloud Control Module is aware of the VM networks, and ensures that a move is to a host that provides connectivity over the given VM network. For example, if Host 1 and 2 provides connectivity to VM Network A, and Host 3 and 4 provide connectivity to VM Network B, Workload Optimization Manager will never recommend a VM with access to Network A, residing on Host 1, to move to Host 3. That would render the VM unable to communicate on Network A.

Configuring Server Message Block (SMB) 3.0 file shares discovery

With VMM, Workload Optimization Manager can discover Server Message Block (SMB) 3.0 shares as data stores, assuming these shares have been properly added to your VMM service center. When you add shares to your VMM environment, be sure to:

- Use the Fully Qualified Domain Name (FQDN) of the file server.

  As you the associated file server to your VMM environment (via the Add Storage Devices Wizard), be sure to specify the FQDN of the file server on the Specify Discovery Scope page of the wizard. Do not use the file server's IP address.

- Ensure that file server names are unique.

  Do not specify file servers with the same name, even if they belong to different domains. Workload Optimization Manager requires the file server names to be unique.

For information about setting up SMB 3.0 shares, see the Microsoft documentation.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Virtual machine
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.
  - Virtual storage

    Virtual ssorage is the measurement of virtual storage capacity that is in use.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - Latency

    Latency is the measurement of storage latency.
- Host
  - Memory (Mem)

    Memory is the measurement of memory that is reserved or in use.
  - CPU

    CPU is the measurement of CPU that is reserved or in use.
  - IO

    IO is the utilization of a host's IO adapters.
  - Net

    Net is the utilization of data through the host's network adapters.
  - Swap

    Swap is the measurement of a host's swap space that is in use.
  - Balloon

    Balloon is the measurement of memory that is shared by VMs running on a host.
  - CPU ready

    CPU ready is the measurement of a host's ready queue capacity that is in use.
- Datacenter

  NOTE:
  For datacenter entities, Workload Optimization Manager does not monitor resources directly from the datacenter, but from the hosts in the datacenter. See host monitored resources for details.

- Provider virtual datacenter
  - Memory (Mem)

    Memory is the measurement of memory that is reserved or in use.
  - CPU

    CPU is the measurement of CPU that is reserved or in use.
  - Storage

    Storage is the utilization of the storage attached to the entity.

- Consumer virtual datacenter
  - Memory (Mem)

    Memory is the measurement of memory that is reserved or in use.
  - CPU

    CPU is the measurement of CPU that is reserved or in use.
  - Storage

    Storage is the utilization of the storage attached to the entity.

Actions

Workload Optimization Manager supports the following actions:

- Virtual machines
  - Provision additional resources (VMem, VCPU)
  - Move Virtual Machine
  - Move Virtual Machine Storage
  - Reconfigure Storage
  - Reconfigure Virtual Machine
  - Suspend VM
  - Provision VM
- Host
  - Start
  - Provision
  - Suspend
- Consumer Virtual Data Centers
  - Resize Consumer vDC
  - Provision Consumer vDC

Secure setup of WSMan

Workload Optimization Manager provides a secure option for Hyper-V/VMM Targets which requires that WSMan be set up securely. Use PowerShell to generate a self-signed certificate, and create an HTTPS WinRM listener.

> **NOTE:**
> For clustered Hyper-V targets, you do not need to create a listener on each host. Only create a listener on the host that is being added to the "Address" field in the Target Configuration.

To enable secure WSMan on your Hyper-V host:

1. Generate a self-signed certificate using the following command:

   ```
   New-SelfSignedCertificate -CertstoreLocation Cert:\LocalMachine\My -DnsName "myhost.example.org"
   ```

2. Find the thumbprint for the certificate for the host:

   ```
   Get-childItem cert:\LocalMachine\My
   ```

3. Create an HTTPS WinRM listener for the host with the thumbprint you've found:

   ```
   winrm create winrm/config/Listener?Address=*+Transport=HTTPS '@{Hostname="myhost.example.org";
    CertificateThumbprint="THUMBPRINT_YOU_FOUND"}'
   ```

4. Verify the presence of configured listeners:

```
Get-WSManInstance –ResourceURI winrm/config/listener –Enumerate
```

Public cloud targets

The public cloud provides compute, storage, and other resources on demand. Workload Optimization Manager can analyze the performance of workloads running on the public cloud, and scale workloads as demand requires.

With public cloud targets, you can use Workload Optimization Manager to scale workloads at the lowest possible cost, or reduce costs by purchasing discounts, deleting unattached volumes, or stopping workloads temporarily.

Amazon Web Services

Amazon Web Services (AWS) is Amazon's cloud computing platform. Workload Optimization Manager discovers your AWS resources through IAM identities and then optimizes these resources to assure performance at the lowest possible cost.

To connect to AWS, follow the steps and guidelines outlined in the following topics:

- Connecting to AWS *(on page 373)*
- AWS permissions *(on page 393)*
- AWS Billing permissions *(on page 393)*

After connecting to AWS, Workload Optimization Manager monitors and optimizes the resources that it discovered. See the following topics for more information:

- AWS Monitored Resources *(on page 393)*
- AWS Actions *(on page 403)*

Connecting to AWS

To connect Workload Optimization Manager to your AWS environment, perform the following tasks:

1. Configure firewall and proxy access. *(on page 373)*

   If you run Workload Optimization Manager behind a firewall, configure firewall and proxy access.
2. Set up IAM identities in AWS *(on page 375)*.

   Workload Optimization Manager discovers and monitors your AWS workloads through IAM identities that you set up in AWS.
3. Set up a data export in AWS. *(on page 386)*

   Workload Optimization Manager uses a data export stored in an S3 bucket to visualize historical cloud expenses, and discover discounts and billing family relationships.
4. Add an AWS Billing target in Workload Optimization Manager. *(on page 389)*

   Authorize a secure connection to your billing data.

   To authorize the connection, add an AWS Billing target in the Workload Optimization Manager user interface.
5. Add an AWS target in Workload Optimization Manager. *(on page 391)*

   Authorize a secure connection to your workloads.

   To authorize the connection, add an AWS target in the Workload Optimization Manager user interface.

Configuring firewall and proxy access for AWS

If you run Workload Optimization Manager behind a proxy or firewall, configure Workload Optimization Manager to allow unrestricted access to the following URLs to enable the full set of features for AWS targets.

NOTE:
If your firewall performs IP based routing, be aware that Amazon reserves the right to change the DNS names on these endpoints over time. Also note, the AWS SDK is designed to use dynamic endpoints. It is possible that Amazon will change the endpoints the SDK depends on over time. As a result, the names of endpoints you must access might be different than the names listed here.

- Autoscaling
  `autoscaling.{region-id}.amazonaws.com`
- AWS Bill
  `{bucket-name}.s3.{region-containing-report}.amazonaws.com`
- CloudWatch
  `monitoring.{region-id}.amazonaws.com`
- CloudWatch Events
  `events.{region-id}.amazonaws.com`
- CloudWatch Logs
  `logs.{region-id}.amazonaws.com`
- Cost Explorer
  `ce.us-east-1.amazonaws.com`
- EC2
  `ec2.{region-id}.amazonaws.com`
- Elastic Load Balancing
  `elasticloadbalancing.{region-id}.amazonaws.com`
- IAM
  `iam.amazonaws.com`
- Organizations
  `organizations.{region-id}.amazonaws.com`
- Performance Insights
  `pi.{region-id}.amazonaws.com`
- Price List
  `pricing.us-east-1.amazonaws.com`
- Relational Database Service (RDS)
  `rds.{region-id}.amazonaws.com`
- Resource Groups
  `resource-groups.{region-id}.amazonaws.com`
- Savings Plans (Optional: Required if you use Savings Plans in your environment)
  `savingsplans.amazonaws.com`
- Service Catalog
  `servicecatalog.{region-id}.amazonaws.com`
- S3
  `s3.{region-id}.amazonaws.com`
- Storage Gateway
  `storagegateway.{region-id}.amazonaws.com`

Next step

In AWS, set up an access key or IAM role that Workload Optimization Manager will use to connect to your AWS environment. For details, see this topic *(on page 375)*.

Overview of required AWS IAM identities for Workload Optimization Manager

Workload Optimization Manager connects to your AWS environment through IAM identities. This topic provides guidance on the IAM identities that you need to set up.

Multi-account and single-account targets

You can configure the AWS target in Workload Optimization Manager to access multiple accounts or a single account.

- Multi-account target

  In this scenario, an AWS target monitors and optimizes workloads in multiple accounts. This eliminates the need to manually add several targets in the Workload Optimization Manager user interface.

  For example, your AWS environment might have a management account and five member accounts. You can configure an AWS target to access all of these accounts, or only the management account and several member accounts.

  For a multi-account target, you must configure two IAM identities in AWS.

  - An **IAM role** or **IAM user** for the management account, depending on your Workload Optimization Manager deployment. See the next section in this topic for details.
  - A **cross-account IAM role** that the AWS target uses to discover existing member accounts. You can enable automatic deployment for the role so that the AWS target automatically discovers new member accounts as they are added to the management account.

- Single-account target

  In this scenario, an AWS target monitors and optimizes workloads in a particular account. This account can be a management account or a member account. For example, if you configure a management account, Workload Optimization Manager only monitors and optimizes workloads in the management account, but not in member accounts.

  For a single-account target, you must configure an **IAM role** or **IAM user** for the management account, depending on your Workload Optimization Manager deployment. See the next section in this topic for details.

It is possible to configure multiple targets. For example, you can configure a multi-account target to access the management account and all but one member accounts, and a single-account target to access the member account that is not included in the multi-account target. When you configure multiple targets, be sure that no account exists in more than one target. Otherwise, multiple targets will report the same account and you will need to either update or delete targets until only one of them is accessing that account.

NOTE:
For best practices on managing IAM identities, see the AWS documentation.

Required IAM identities for a multi-account target

Refer to the following table for guidance on the IAM identities that you need to set up based on your Workload Optimization Manager deployment.

Workload Optimization Manager deployment	IAM identities to set up
Deployed on premises using an OVA or VHD image	▪ IAM user *(on page 376)* for the management account ▪ Cross-account IAM role *(on page 380)* for member account discovery
Deployed to a properly configured Kubernetes cluster in AWS through Amazon EKS or Red Hat OpenShift Service on AWS (ROSA) **NOTE:** For these deployments, cluster configurations must support an OIDC provider, webhooks, and role assumption through an annotated service account.	▪ IAM user *(on page 376)* or IAM role *(on page 378)* for the management account ▪ Cross-account IAM role *(on page 380)* for member account discovery
Deployed to Red Hat OpenShift on IBM Cloud	▪ IAM user *(on page 376)* for the management account ▪ Cross-account IAM role *(on page 380)* for member account discovery
Deployed to Kubernetes or Red Hat OpenShift on other cloud providers	▪ IAM user *(on page 376)* for the management account ▪ Cross-account IAM role *(on page 380)* for member account discovery

Required IAM identity for a single-account target

Refer to the following table for guidance on the IAM identity that you need to set up based on your Workload Optimization Manager deployment.

Workload Optimization Manager deployment	IAM identity to set up
Deployed on premises using an OVA or VHD image	IAM user *(on page 376)*
Deployed to a properly configured Kubernetes cluster in AWS through Amazon EKS or Red Hat OpenShift Service on AWS (ROSA) **NOTE:** For these deployments, cluster configurations must support an OIDC provider, webhooks, and role assumption through an annotated service account.	IAM role *(on page 378)* or IAM user *(on page 376)*
Deployed to Red Hat OpenShift on IBM Cloud	IAM user *(on page 376)*
Deployed to Kubernetes or Red Hat OpenShift on other cloud providers	IAM user *(on page 376)*

Setting up an AWS IAM user

Perform the following tasks to set up an IAM user for use with Workload Optimization Manager.

For any deployment of Workload Optimization Manager, you can set up an IAM user that the AWS target uses to connect to one of the following accounts:

- Management account for a multi-account AWS target

  If you want an AWS target in Workload Optimization Manager to manage multiple accounts, follow the instructions in this topic to set up an IAM user for the management account.

  For the member accounts, you need to set up an IAM role called cross-account IAM role *(on page 380)*.

- Management or member account for a single-account AWS target

  If you want an AWS target in Workload Optimization Manager to manage only a single account (either a management or member account), follow the instructions in this topic to set up an IAM user for the account.

Guidelines

- Workload Optimization Manager recommends setting up an IAM user group that has the necessary permissions and then adding the IAM user to that group.
- If the IAM user that you are setting up also grants Workload Optimization Manager access to your billing data, the IAM user requires access to the S3 bucket that contains your data export. Billing access is not required.

  Workload Optimization Manager uses a data export stored in an S3 bucket to visualize historical cloud expenses, and discover discounts and billing family relationships.

  NOTE:
  You will set up a data export in a later task.

Task overview

To set up an IAM user, perform the following tasks in the AWS Management Console:

1. Create IAM policies that specify the permissions that Workload Optimization Manager needs to connect to AWS.
2. Create an IAM user and then assign the policies that you created to that user.
3. Generate an access key for the IAM user.

Creating IAM policies

1. Sign in to the AWS Management Console and open the IAM console.

   https://console.aws.amazon.com/iam/

2. In the navigation pane , choose **Policies**.
3. Choose **Create policy**.
4. In the **Policy editor** section, choose **JSON**.
5. Download the JSON file that contains the Workload Optimization Manager minimum permissions and then copy the permissions to the **JSON** field.

   The JSON file that you need to download depends on the account that the AWS target manages (see the introduction in this topic for more information), and the operations that are allowed on the workloads in the given account. Choose one of the following JSON files to download.

Account	Workload Optimization Manager operations	Download link
Management account for a multi-account AWS target	Monitor workloads and automatically execute actions for the workloads	Minimum permissions
Management account for a multi-account AWS target	Only monitor workloads. Actions for the workloads are executed in AWS.	Minimum permissions
Management or member account for a single-account AWS target	Monitor workloads and automatically execute actions for the workloads	Minimum permissions

Account	Workload Optimization Manager operations	Download link
Management or member account for a single-account AWS target	Only monitor workloads. Actions for the workloads are executed in AWS.	Minimum permissions

To retrieve billing data in your data export, use the minimum permissions for billing data monitoring.

6. Resolve any security warnings, errors, or general warnings generated during policy validation, and then choose **Next**.
7. In the **Review and create** page, type a **Policy Name** and a **Description** (optional) for the policy that you are creating.
8. Choose **Create policy**.

Creating an IAM user

1. In the navigation pane of the IAM console, select **Users** and then choose **Create user**.
2. Specify your preferred user name and then choose **Next**.
3. Select **Attach policies directly**, select the policy or policies that you created in the previous task, and then choose **Next**.
4. Review the user details and then choose **Create user**.

Generating an access key for the IAM user

1. In the navigation pane of the IAM console, select **Users** and then choose the user that you created in the previous task.
2. Choose **Security credentials**, scroll to the **Access keys** section, and then choose **Create access key**.
3. Choose **Third-party service** and then choose **Next**.
4. (Optional) Set a description tag value to describe the purpose of the access key.
5. Choose **Create access key**.
6. Record the access key ID and secret access key. You will need this information later when you add an AWS target in the Workload Optimization Manager user interface.

Next step

If the IAM user that you just set up is for:

- The management account in a multi-account target, set up a cross-account IAM role for the member accounts. For details see this topic *(on page 380)*.
- A management or member account in a single-account target, you do not need to set up a cross-account IAM role. The next step is to set up a data export for use with the AWS Billing target. For details, see this topic *(on page 386)*.

Setting up an AWS IAM role for Red Hat OpenShift or Amazon EKS deployments

Connection to AWS through an IAM role is supported when a Workload Optimization Manager instance is deployed to a Kubernetes cluster in AWS through Red Hat OpenShift Service on AWS (ROSA) or Amazon EKS. For these deployments, cluster configurations must support an OIDC provider, webhooks, and role assumption through an annotated service account.

IMPORTANT:
Workload Optimization Manager requires specific IAM identities depending on the AWS account that it can access. You may need to set up other IAM identities in addition to the one described in this topic. For guidance on the IAM identities that you need to set up, see this topic *(on page 375)*.

Setting up an AWS IAM role

1. Follow the instructions provided by Red Hat OpenShift on AWS (ROSA) or Amazon Elastic Kubernetes Service (EKS) to make sure you have the required configurations to support the Web Identity provider method, leveraging the AWS webhook and an OIDC provider.
   - Introducing EKS Granular IAM Roles for Pods via Service Accounts and OIDC Providers
   - Overview of setting up IAM Role for EKS WebIdentity Provider
   - Creating a Service Account with an IAM Role: EKS example
   - Red Hat OpenShift support for Granular IAM Roles
   - Grant users and groups access to AWS resources
     - Creating a new role
     - Editing the trust relationship for an existing role

   NOTE:
   Follow the instructions from AWS in the previously provided links to set up your cluster's OIDC provider as an IAM Identity Provider (Web Identity Provider - OIDC URL) in the AWS account that you will be targeting.

2. Configure a Kubernetes service account in the Workload Optimization Manager namespace that will assume an IAM role.
   - (Best Practice) Manually create a separate service account for the AWS Mediation pods to use. You must then modify the Custom Resource YAML to specify this service account to the AWS Mediation components. For example:

     ```
     spec:
       mediation-aws:
         serviceAccountName: t8c-iam-role
       mediation-awscloudbilling:
         serviceAccountName: t8c-iam-role
       mediation-awscost:
         serviceAccountName: t8c-iam-role
     ```

     After you update the Custom Resource YAML, apply the updated Custom Resource and ensure the three Mediation pods restarted.
   - Use the Workload Optimization Manager default service account with which the AWS Mediation pods are running.

     NOTE:
     Unless specified in the Workload Optimization Manager Custom Resource, the AWS Mediation pods will run with the default service account called "default" in the namespace. If you modify the default "default" account, there is nothing more to do except restart the AWS Mediation pods.

3. Configure the IAM role in AWS.
   - Workload Optimization Manager AWS IAM role requirements, including cross account access if required, are described here. IAM policy definition must use the `sts:AssumeRoleWithWebIdentity` role.
   - Using the Web identity provider set up, and the service account you will use, update the Trust Relationships in the IAM role. See the instructions here.
   - The IAM role for Workload Optimization Manager to target your AWS account requires the following policies:

     ```
     AmazonRDSReadOnlyAccess
     AmazonEC2ReadOnlyAccess
     AmazonS3ReadOnlyAccess
     AWSOrganizationsReadOnlyAccess
     ```

4. Annotate the service account you will use with the IAM role. See this article for an example.

```
apiVersion: v1
kind: ServiceAccount
#use the name of the SA that will contain the annotation
name: default
metadata:
  annotations:
    eks.amazonaws.com/role-arn: arn:aws:iam::<AWS_ACCOUNT_ID>:role/<IAM_ROLE_NAME>
```

5. For the Kubernetes service account that will have the ability to assume a role, you have the following options:
   - Use the Workload Optimization Manager default service account that the AWS mediation pods run with.

     NOTE:
     In the namespace, the AWS mediation pods will run with the default service account called "default". If you modify this default service account, restart the AWS mediation pods.

   - Manually create a separate service account that will be used only by the AWS mediation pods. This is an AWS best practice. You will need to supply this service account to the Workload Optimization Manager deployment via the custom resource (CR).

     If you choose this option, modify the CR yaml to specify this service account to the AWS mediation components. See a the following sample YAML.

     ```
     spec:
       mediation-aws:
         env:
         - name: AWS_REGION
           value: US_EAST_1
         serviceAccountName: t8c-iam-role
       mediation-awscloudbilling:
         env:
         - name: AWS_REGION
           value: US_EAST_1
         serviceAccountName: t8c-iam-role
       mediation-awscost:
         env:
         - name: AWS_REGION
           value: US_EAST_1
         serviceAccountName: t8c-iam-role
     ```

Next step

In AWS, set up a data export for use with the AWS Billing target. For details, see this topic *(on page 386)*.

Setting up an AWS cross-account IAM role

In Workload Optimization Manager, you can configure a multi-account target, which is one AWS target that has access to multiple AWS accounts. For example, your AWS environment might have a management account and five member accounts. You can configure a multi-account target to access all of these accounts, or only the management account and several member accounts.

For a multi-account target, you must set up two IAM identities in AWS.

- For any deployment of Workload Optimization Manager, you can set up an IAM user *(on page 376)* for the management account. Be sure to set up an IAM user for the management account before performing the tasks outlined in this topic.
- A cross-account IAM role that the AWS target uses to discover member accounts. This topic describes the steps to set up the cross-account IAM role.

NOTE:
Skip this topic if you want to set up a single-account target (that is, a target that accesses only a single account, which can be a management account or a member account). For guidance on the IAM identity that you need to set up for the single-account target, see this topic *(on page 376)*.

Setting up a cross-account IAM role for a management account that uses an IAM user

Create an AWS CloudFormation StackSets to set up the cross-account IAM role. To create StackSets successfully, you must have sufficient CloudFormation permissions and must not be restricted by service control policies. These service control policies might prevent you from accessing certain member accounts, which could then result in failures to discover these member accounts in Workload Optimization Manager.

Before proceeding, download the CloudFormation template.

- To monitor workloads and execute actions within Workload Optimization Manager, download this template.
- To monitor workloads only, download this template. Actions for these workloads can be executed in AWS.

1. Sign in to the AWS CloudFormation console.
   https://console.aws.amazon.com/cloudformation/
2. In the page banner, find the dropdown menu of regions and then select the region for the stack set.
3. In the navigation pane , choose **StackSets**.
4. Choose **Create StackSet**.
5. Configure settings for **Choose a template**.

Setting	Instructions
Permissions	Create a stack set using service-managed or self-managed permissions. ▪ (Recommended) Service-managed permissions With service-managed permissions, you can deploy stack instances to accounts managed by your AWS organization without the need to create the necessary IAM roles. StackSets creates the IAM roles on your behalf. You can also turn on automatic deployments to accounts that you add to your organization in the future. ▪ Self-managed permissions With self-managed permissions, you create the IAM roles required by StackSets and deploy the roles across accounts and regions. These roles are necessary to establish a trusted relationship between the accounts for the stack set and the stack instances. Using this permission model, StackSets can deploy to any AWS account in which you have permissions to create an IAM role.
Prerequisite - Prepare template	Choose **Template is ready**.
Specify template	Choose **Upload a template file** and then upload the CloudFormation template that you downloaded in a previous task.

6. Configure settings for **Specify StackSet details**.

Setting	Instructions
StackSet name	Specify your preferred name.
StackSet description	Specify your preferred description.
Parameters	The parameters are pre-populated based on the template that you uploaded. Update the parameters as described in the next sections.

- Workload Optimization Manager IAM role name

Setting	Instructions
Workload Optimization Manager IAM role name	The default IAM role name is `Turbonomic_Org_Access` and is specified in the CloudFormation template that you downloaded. You can change the default name in the StackSets that you are configuring, but be sure to use that same name in the Workload Optimization Manager user interface when you add an AWS target. If the names in StackSets and in the Workload Optimization Manager user interface do not match, there could be issues discovering member accounts. A valid name must include only alphanumeric characters and/or the following symbols: + = , . @ _ -

- Management account user ARN

Setting	Instructions
Management account user ARN	Specify the ARN of the IAM user *(on page 376)* that you previously set up for the management account.

7. Configure settings for **Configure StackSet options**.

Setting	Instructions
Tags - optional	Specify any tags to apply to resources in your stack set.
Execution configuration	Choose **Active** so that StackSets performs non-conflicting operations concurrently and queues conflicting operations. After conflicting operations finish, StackSets starts queued operations in request order.

8. Configure settings for **Set deployment options**.

Setting	Instructions
Add stacks to stack set	Choose **Deploy new stacks**.
Accounts	Choose from the following options: - **Deploy to organization**: Workload Optimization Manager discovers all member accounts in your organization. - **Deploy to organizational units (OUs)**: Workload Optimization Manager discovers the member accounts in OUs.

Setting	Instructions
	If you choose this option, be sure to set up OUs. For setup instructions, see the AWS documentation. When the OUs are ready, choose **Add an OU** and then specify the OU ID. Repeat this step for each OU.
Automatic deployment	Workload Optimization Manager recommends enabling automatic deployment so it can automatically monitor and optimize workloads in member accounts that are added to the organization or OUs in the future. If you enable automatic deployment, choose an **Account removal behavior** option. Your selection determines whether stack resources are retained or deleted when an account is removed from the organization or OU. If you choose to delete stack resources, Workload Optimization Manager stops monitoring and optimizing workloads in the removed accounts.
Specify regions	Since the CloudFormation template only specifies global resources, you can select `us-east-1` or any of the given regions.
Deployment options	▪ **Maximum concurrent accounts**: Configure the maximum concurrent accounts as needed. Increase the value to accelerate the creation of member accounts IAM roles across a large number of AWS accounts. ▪ **Failure tolerance**: Configure a failure tolerance as needed. ▪ **Region concurrency**: Choose a region concurrency option as needed. Since you are deploying global resources using only one region, **Region concurrency** is irrelevant. ▪ **Concurrency mode**: Choose a concurrency mode option as needed.

9. In the Review page, verify that the settings are correct and then choose **Create StackSet**.

Next step

In AWS, set up a data export for use with the AWS Billing target. For details, see this topic *(on page 386)*.

Reference: Setting up a multi-account AWS target

Starting with version 3.12.4, you can configure a single AWS target to access multiple accounts (either all accounts or a subset of accounts). This feature is also referred to as 'multi-account target' and is intended to eliminate the need to configure several targets to access the same accounts.

1. (Recommended) If you have existing AWS targets and you added them before version 3.12.4 to discover individual accounts, you can consolidate them starting with version 3.12.4 for efficiency.

   NOTE:
   Skip this step if your existing targets are for AWS accounts that are not included in the multi-account target that you are setting up in the next step.

To consolidate existing AWS targets, it is recommended that you delete them before adding a multi-account target. Otherwise, you will see a notification about duplicate targets when you view the details for the multi-account target.

Navigate to **Settings > Target configuration** and then delete the targets individually.

To ensure a seamless transition to a multi-account target, Workload Optimization Manager persists data for the deleted targets, as long as the data is not scheduled for deletion based on your data retention settings *(on page 1128)*. To prevent any potential data loss, it is recommended that you set up a multi-account target immediately after deleting your existing targets.

2. Set up a multi-account target.
   a. In AWS, set up an IAM user *(on page 376)* for the management account that manages member accounts.

      After you set up the IAM identity, record the following information. You need this information when you add the multi-account target in the Workload Optimization Manager user interface.

      - For an IAM user, record the access key ID and secret access key.

   b. In AWS, set up a cross-account IAM role *(on page 380)* for the member accounts. The multi-account target in Workload Optimization Manager uses this role to discover the member accounts.

      After you set up the role, record the IAM role name (not the ARN). You need this information when you add the multi-account target. The default name is `Turbonomic_Org_Access`.

   c. In the Workload Optimization Manager user interface, navigate to **Settings > Target configuration** and then click **Add target**.

   d. Select **AWS** and then click **Connect Target**.

   e. In the Connection overview page, specify the display name and select **Multiple accounts**.

   f. In the Management account access page, click the **IAM User** tab.

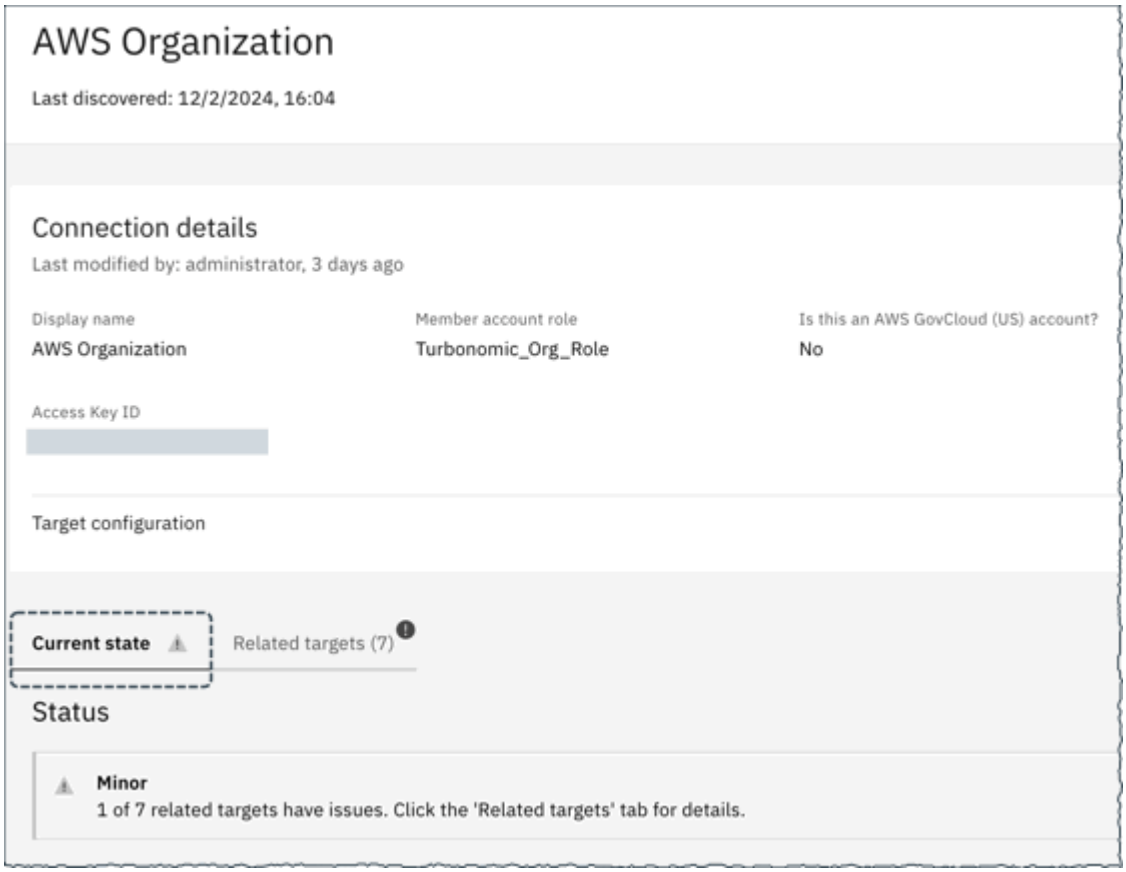
      - For an IAM user, specify the access key ID and secret access key that you recorded in a previous step.

   g. In the Member account access page, specify the cross-account IAM role name that you recorded in a previous step.

   h. Complete the target addition steps and click **Add target**.

i. In the Target configuration page, find the multi-account target that you just added and then click **View**.

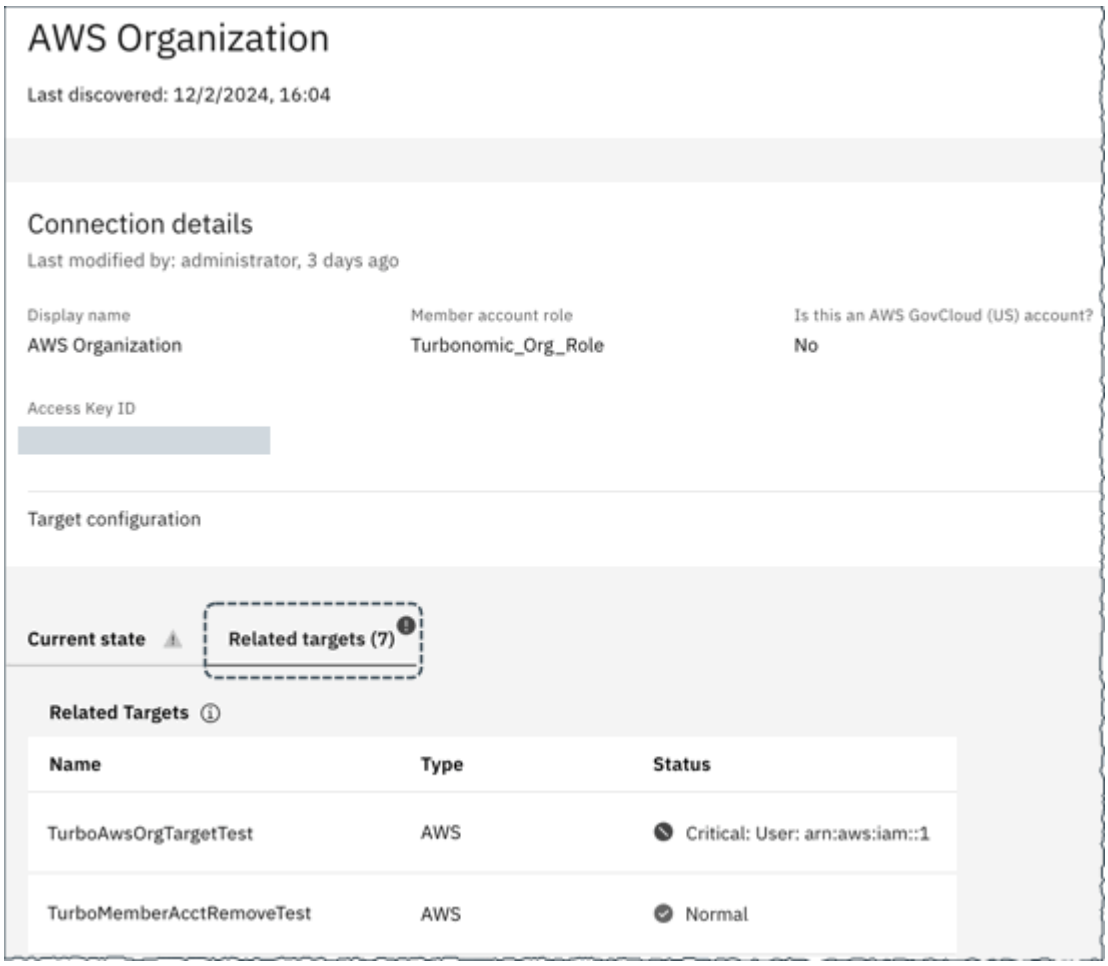
- The **Current state** tab shows the status of the multi-account target. The status shows a warning if there are issues discovering the accounts that you configured for the target.



If you did not delete individual targets as described at the beginning of this topic, the **Current state** tab also shows a notification that informs you of duplicate targets. It is recommended that you delete these individual targets since the AWS accounts they manage are now discovered as part of the multi-account target that you just set up. The notification stops displaying after the targets are deleted.

NOTE:
It is possible to configure a multi-account target along with a single-account target. For example, you can configure a multi-account target to access the management account and all but one member accounts, and a single-account target to access the member account that is not included in the multi-account target. When you configure multiple targets, be sure that no account exists in more than one target. Otherwise, the same notification about duplicate targets displays since there are multiple targets reporting the same account. You will need to either update or delete targets until only one of them is accessing that account.

- The **Related targets** tab shows the number of accounts that the multi-account target manages and a warning icon if there are issues with any of the accounts. Click the tab to review the details for the accounts.



Setting up an AWS data export

Workload Optimization Manager uses a data export stored in an S3 bucket to visualize historical cloud expenses, and discover discounts and billing family relationships.

Guidelines

- Workload Optimization Manager supports only the following data exports:
  - Standard data export (CUR 2.0)
  - Legacy CUR export

  All other data exports available in AWS are not supported.
- Set up the data export exactly as described in this topic to comply with Workload Optimization Manager requirements. Any additional configuration may not work with the AWS Billing target *(on page 389)* that is associated with the data export.
- The IAM user or role that you set up for Workload Optimization Manager requires access to the S3 bucket that contains your data export. Billing access is not required.

- Workload Optimization Manager supports a data export created at the management account, but not member accounts.
- AWS publishes the daily data export twice a day, but it may take up to 24 hours for AWS to deliver the data export to the S3 bucket. Charts in Workload Optimization Manager, such as the Workload Cost Breakdown chart, will start to show data 1 to 2 hours after data export delivery.

Setting up a standard data export (CUR 2.0)

1. Sign in to the Billing and Cost Management console.

   https://console.aws.amazon.com/billing
2. In the page banner, find the dropdown menu of regions and then select the region for the S3 bucket.
3. In the navigation pane, find the **Cost Analysis** section and choose **Data Exports**.
4. Choose **Create**.
5. Choose **Standard data export** and configure the following settings.
   - Export name

Setting	Instructions
Export name	Specify your preferred name.

   - Data table content settings

Setting	Instructions
Additional export content	Select **Include resource IDs**.
Time granularity	Choose **Daily**.

   - Column selection

     By default, all columns are selected. To include only the columns that Workload Optimization Manager currently requires, select the following items:

     - `bill_payer_account_id`
     - `identity_line_item_id`
     - `line_item_currency_code`
     - `line_item_line_item_type`
     - `line_item_normalization_factor`
     - `line_item_normalized_usage_amount`
     - `line_item_operation`
     - `line_item_product_code`
     - `line_item_resource_id`
     - `line_item_unblended_cost`
     - `line_item_usage_account_id`
     - `line_item_usage_amount`
     - `line_item_usage_end_date`
     - `line_item_usage_start_date`
     - `line_item_usage_type`
     - `pricing_public_on_demand_cost`
     - `pricing_term`
     - `pricing_unit`
     - `product`
     - `product_instance_type`

  - `product_product_family`
  - `product_sku`
  - `reservation_reservation_a_r_n`
  - `savings_plan_savings_plan_a_r_n`
  - `savings_plan_savings_plan_effective_cost`
- Data export delivery options

Setting	Instructions
Compression type and file format	Choose **gzip - text/csv**.
File versioning	Choose **Create new data export file**.

- Data export storage settings

Setting	Instructions
S3 bucket	Click **Configure** and then select an existing bucket or create a new one. If you created a new one, the region that you selected in a previous step is automatically specified. Note that the `AmazonS3FullAccess` and `Billing` permissions are required to set up the bucket. For detailed setup instructions, see the AWS documentation. Be sure to verify the bucket policy that AWS applies.
S3 path prefix	Specify your preferred prefix, such as `daily`.

6. Review your settings and record the following information. You will need this information later when you add the AWS Billing target in the Workload Optimization Manager user interface.
   - S3 bucket name, such as `turbodataexport`
   - S3 path prefix, such as `daily`
   - S3 bucket region, such as `us-east-1`
7. Click **Create**.

   AWS can now deliver the data export to the S3 bucket. It could take up to 24 hours for AWS to deliver the first data export.

Setting up a Legacy CUR export

1. Sign in to the Billing and Cost Management console.

   https://console.aws.amazon.com/billing
2. In the page banner, find the dropdown menu of regions and then select the region for the S3 bucket.
3. In the navigation pane, find the **Cost Analysis** section and choose **Data Exports**.
4. Choose **Create**.
5. Choose **Legacy CUR export** and configure the following settings.
   - Export name

Setting	Instructions
Export name	Specify your preferred name.

   - Export content

Setting	Instructions
Additional export content	Select **Include resource IDs**.

- Data export delivery options

Setting	Instructions
Report data time granularity	Choose **Daily**.
Report versioning	Choose **Create new report version**.

- Data export storage settings

Setting	Instructions
S3 bucket	Click **Configure** and then select an existing bucket or create a new one. If you created a new one, the region that you selected in a previous step is automatically specified. Note that the `AmazonS3FullAccess` and `Billing` permissions are required to set up the bucket. For detailed setup instructions, see the AWS documentation. Be sure to verify the bucket policy that AWS applies.
S3 path prefix	Specify your preferred prefix, such as `daily`.

6. Review your settings and record the following information. You will need this information later when you add the AWS Billing target in the Workload Optimization Manager user interface.
   - S3 bucket name, such as `turbodataexport`
   - S3 path prefix, such as `daily`
   - S3 bucket region, such as `us-east-1`
7. Click **Create report**.

   AWS can now deliver the data export to the S3 bucket. It could take up to 24 hours for AWS to deliver the first data export.

Next step

In the Workload Optimization Manager user interface, add an AWS Billing target using the data export settings and the IAM user or role that you set up. For details, see this topic *(on page 389)*.

Adding an AWS Billing Target

The AWS Billing target grants Workload Optimization Manager access to billing data from a data export stored in an S3 bucket. Workload Optimization Manager uses this data to visualize historical cloud expenses, and discover discounts and billing family relationships.

AWS member accounts and standalone accounts are not supported.

You can add multiple billing targets. Data for these targets will be aggregated and shown when you set the scope to your global environment.

NOTE:
The AWS Billing target only supports adding a data export associated with a management account. Member accounts or stand alone accounts are not supported.

Prerequisites

Before performing this task, set up IAM identities *(on page 375)* and a data export *(on page 386)* in AWS.

Adding an AWS Billing target using an IAM role

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **AWS Billing**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
    - Display name

      Specify a name that uniquely identifies this connection.

      This name is for display purposes only and does not need to match any name in AWS.
    - IAM Role ARN

      Specify the ARN (Amazon Resource Name) for your IAM role.

      ARN authorizes a secure connection that grants Workload Optimization Manager the specific permissions that it needs to manage resources without sharing your security credentials.
    - S3 bucket name

      Specify the full name of the S3 bucket that contains the data export.
    - S3 path prefix

      Specify the S3 path prefix for the data export.
    - S3 bucket region

      Specify the region of the S3 bucket that contains the data export.
    - Proxy configuration

      Specify proxy information only if you connect through a proxy.
        - Proxy hostname or IP address

          Specify the address of the proxy used for this target.
        - Proxy port

          Specify the port to use with the proxy used for this target. By default, the port is 8080.
        - Proxy username

          Specify the username to use with the proxy.
        - Proxy password

          Specify the password to use with the proxy.
        - Connect using HTTPS

          Select this option if Workload Optimization Manager will connect to the proxy through HTTPS.

Adding an AWS Billing target using an IAM user

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **AWS Billing**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
    - Display name

      Specify a name that uniquely identifies this connection.

This name is for display purposes only and does not need to match any name in AWS.

- Access Key ID

  Specify the access key ID associated with the IAM user.

  Access keys are long-term credentials for an IAM user or the AWS account root user.

  View security best practices in IAM
- Secret Access Key

  Specify the secret access key associated with the IAM user.
- S3 bucket name

  Specify the full name of the S3 bucket that contains the data export.
- S3 path prefix

  Specify the S3 path prefix for the data export.
- S3 bucket region

  Specify the region of the S3 bucket that contains the data export.
- Proxy configuration

  Specify proxy information only if you connect through a proxy.
  - Proxy hostname or IP address

    Specify the address of the proxy used for this target.
  - Proxy port

    Specify the port to use with the proxy used for this target. By default, the port is 8080.
  - Proxy username

    Specify the username to use with the proxy.
  - Proxy password

    Specify the password to use with the proxy.
  - Connect using HTTPS

    Select this option if Workload Optimization Manager will connect to the proxy through HTTPS.

Next step

In the Workload Optimization Manager user interface, add an AWS target. For details, see this topic *(on page 391)*.

Adding an AWS target

Add an AWS target in the Workload Optimization Manager user interface to monitor and optimize workloads in your AWS environment. This target specifies the IAM identities that Workload Optimization Manager will use to connect to AWS.

Before performing this task, be sure to set up the required IAM identities in AWS. For information on the required IAM identities, see this topic *(on page 375)*.

Adding an AWS target

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **AWS**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure **Connection overview** settings.
   - Display name

     Specify a name that uniquely identifies this connection.

This name is for display purposes only and does not need to match any name in AWS.

- Account scope

  Choose from the following:

  - Multiple accounts

    This target has access to multiple AWS accounts. Before proceeding, be sure to set up an IAM user *(on page 376)* for the management account and a cross-account IAM role *(on page 380)* for member accounts.

  - Single account

    This target has access to a single AWS account. Before proceeding, be sure to set up an IAM user *(on page 376)* for the account.

- AWS GovCloud (US) account toggle

  When enabled, Workload Optimization Manager identifies the target as an AWS GovCloud *(on page 397)* account.

6. Configure **Account access** settings.
   - If you chose **Multiple accounts** or **Single account** and set up an IAM user:
     a. Click the **IAM User** tab.
     b. Specify the access key ID associated with the IAM user.

        Access keys are long-term credentials for an IAM user or the AWS account root user.

        View security best practices in IAM

     c. Specify the secret access key associated with the IAM user.
7. Configure **Member accounts access** settings.

   These settings apply only if you selected **Multiple accounts** in the Connection overview page.

   a. Specify the IAM role name on this page. Use only alphanumeric characters and/or the following characters:

      + = , . @ _ -

      Specify the name (not the ARN) of the cross-account IAM role that Workload Optimization Manager uses to automatically discover member accounts in your organization. The name must be unique and should not be used in any IAM roles assigned to your member accounts.

   b. Review and select a policy that meets your permission requirements.
   c. Click **Download template** to download the CloudFormation template based on your selected policy.
   d. Click **Launch CloudFormation** to create a StackSet that deploys IAM roles in each member account. For detailed instructions, see this topic *(on page 381)*.
8. Proxy configuration

   Specify proxy information only if you connect through a proxy.

   - Proxy hostname or IP address

     Specify the address of the proxy used for this target.

   - Proxy port

     Specify the port to use with the proxy used for this target. By default, the port is 8080.

   - Proxy username

     Specify the username to use with the proxy.

   - Proxy password

     Specify the password to use with the proxy.

   - Connect using HTTPS

     Select this option if Workload Optimization Manager will connect to the proxy through HTTPS.

After adding an AWS target

You have completed the required tasks for connecting to AWS.

- Workload Optimization Manager can now monitor your AWS workloads and billing data, and recommend actions to optimize these workloads at the lowest possible cost.

  Review the following topics for more information:

  - AWS Monitored Resources *(on page 393)*
  - AWS Actions *(on page 403)*
- If you have price adjustments for workloads based on your agreement with AWS, manually configure price adjustments in the Workload Optimization Manager user interface, in **Settings > Billing and Costs > Price Adjustments**. If price adjustments are not set, Workload Optimization Manager will use on-demand pricing, which could result in incorrect cost information in Workload Optimization Manager. For details, see this topic *(on page 1109)*.

Reference: AWS permissions

The IAM identities that you set up in AWS must specify the permissions that Workload Optimization Manager needs to discover and monitor your AWS workloads and billing data. Permissions to execute actions from Workload Optimization Manager are optional but recommended.

Click the links in the following table to view the minimum permissions for both the AWS and AWS Billing targets.

Target	IAM policy
AWS Billing	Choose one: ▪ Data export retrieval, workload monitoring, and action execution (JSON) ▪ Data export retrieval and workload monitoring (JSON)
AWS	▪ For a multi-account target (the target has access to multiple AWS accounts) – For the management account, choose one: • Workload monitoring and action execution (JSON) • Workload monitoring (JSON) – For member accounts, choose one: • Workload monitoring and action execution (YAML) • Workload monitoring (YAML) ▪ For a single-account target (the target has access to one AWS account), choose one: – Workload monitoring and action execution (JSON) – Workload monitoring (JSON)

AWS monitored resources

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

AWS	Workload Optimization Manager
Elastic Compute Cloud (EC2)	Virtual machine
EC2 Auto Scaling group	Virtual machine spec

AWS	Workload Optimization Manager
Relational Database Service (RDS)	Database server
Elastic Block Store (EBS)	Volume
Redshift provisioned cluster	Data warehouse
Availability zone	Zone
Region	Region

Points to consider:

- Workload Optimization Manager supports discovery and management of entities in certain AWS regions, including AWS GovCloud (US) regions.

  For details, see Supported AWS Regions *(on page 395)* and Support for AWS GovCloud *(on page 397)*.

Monitored resources for virtual machines (EC2)

Workload Optimization Manager monitors the following resources:

- Virtual memory (vMem)

  Virtual memory (vMem) is the measurement of memory that is in use.

  It is highly recommended that you enable collection of metrics in your environment. Enabling metrics allows Workload Optimization Manager to generate scale actions to optimize VM resource usage. For Workload Optimization Manager to collect metrics, you must enable the collection of these metrics on the VMs in your environment.

  For details, see AWS Memory Metrics Collection *(on page 399)*.
- Virtual CPU (vCPU)

  Virtual CPU is the measurement of CPU that is in use.
- Storage amount

  Storage amount is the measurement of storage capacity that is in use.
- Storage access (IOPS)

  Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
- I/O throughput

  I/O throughput is the measurement of an entity's throughput to the underlying storage.
- Net throughput

  Net throughput is the rate of message delivery over a port.

Monitored resources for virtual machine specs (Auto Scaling groups)

Workload Optimization Manager does not monitor resources for Auto Scaling groups, but allows you to park (stop or start) these groups on demand or according to a schedule.

Monitored resources for database servers

Workload Optimization Manager monitors the following resources:

- Virtual memory (vMem)

  Virtual memory (vMem) is the measurement of memory that is in use.
- Virtual CPU (vCPU)

  Virtual CPU is the measurement of CPU that is in use.
- Storage amount

Storage amount is the measurement of storage capacity that is in use.

- Storage access (IOPS)

  Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

- DB cache hit rate

  DB cache hit rate is the measurement of Database Server accesses that result in cache hits, measured as a percentage of hits versus total attempts. A high cache hit rate indicates efficiency.

- Connection

  Connection is the measurement of database connections utilized by applications.

Monitored resources for data warehouses (Redshift provisioned clusters)

Workload Optimization Manager monitors the following resources:

- Number of replicas

  Number of replicas is the total number of nodes (leader node and compute nodes) in a Redshift provisioned cluster.

- Storage amount

  Storage amount is the measurement of storage capacity that is in use.

  NOTE:
  Workload Optimization Manager calculates storage amount. The calculated amount may not match the data returned by Redshift due to an AWS issue. For this reason, you may see a slight difference in the storage amounts shown in Workload Optimization Manager and in the AWS console.

- Connection

  Connection is the measurement of database connections utilized by applications.

Monitored resources for volumes

Workload Optimization Manager monitors the following resources:

- Storage amount

  Storage Amount is the storage capacity (disk size) of a volume.

  Workload Optimization Manager discovers Storage Amount, but does not monitor utilization.

  For a Kubeturbo (container) deployment that includes volumes, Kubeturbo monitors Storage Amount utilization for the volumes. You can view utilization information in the Capacity and Usage chart.

- Storage access (IOPS)

  Storage access, also known as IOPS, is the measurement of IOPS capacity that is in use.

- I/O throughput

  I/O throughput is the measurement of I/O throughput capacity that is in use.

Supported AWS regions

Workload Optimization Manager supports discovery and management of entities in the following AWS regions:

Region code	Region name	Notes
af-south-1	Africa (Cape Town)	Requires opt-in within the AWS console
ap-east-1	Asia Pacific (Hong Kong)	Requires opt-in within the AWS console

Region code	Region name	Notes
ap-northeast-1	Asia Pacific (Tokyo)	
ap-northeast-2	Asia Pacific (Seoul)	
ap-northeast-3	Asia Pacific (Osaka)	
ap-south-1	Asia Pacific (Mumbai)	
ap-south-2	Asia Pacific (Hyderabad)	
ap-southeast-1	Asia Pacific (Singapore)	
ap-southeast-2	Asia Pacific (Sydney)	
ap-southeast-3	Asia Pacific (Jakarta)	
ap-southeast-4	Asia Pacific (Melbourne)	
ca-central-1	Canada (Central)	
ca-west-1	Canada West (Calgary)	
eu-central-1	Europe (Frankfurt)	
eu-central-2	Europe (Zurich)	
eu-north-1	Europe (Stockholm)	
eu-south-1	Europe (Milan)	Requires opt-in within the AWS console
eu-south-2	Europe (Spain)	
eu-west-1	Europe (Ireland)	
eu-west-2	Europe (London)	
eu-west-3	Europe (Paris)	
il-central-1	Israel (Tel Aviv)	
me-central-1	Middle East (UAE)	
me-south-1	Middle East (Bahrain)	Requires opt-in within the AWS console
sa-east-1	South America (São Paulo)	
us-east-1	US East (N. Virginia)	
us-east-2	US East (Ohio)	
us-gov-east-1	AWS GovCloud (US-East)	Restricted Access (GovCloud)
us-gov-west-1	AWS GovCloud (US-West)	Restricted Access (GovCloud)
us-west-1	US West (N. California)	
us-west-2	US West (Oregon)	

Unreachable AWS regions

When Workload Optimization Manager fails to reach one or more AWS regions as part of discovering your AWS environment, then AWS discovery will fail for that target.

There may be policy decisions that prevent Workload Optimization Manager from reaching all AWS regions. For example, if you operate Workload Optimization Manager behind a firewall, you might not be able to reach all the regions that are available to your AWS account. In that case, you need to specify which regions you want Workload Optimization Manager to discover.

For information about how to specify the regions that you want Workload Optimization Manager to discover, contact your Workload Optimization Manager representative.

Support for AWS GovCloud

AWS GovCloud (US) provides dedicated regions for US government customers and their partners to architect secure cloud solutions and meet regulatory and compliance requirements.

Workload Optimization Manager can monitor workloads in AWS GovCloud (US) regions, including VMs, auto-scaling groups, volumes, database servers, and spot instances. Workload Optimization Manager recommends actions on VMs, volumes, and database servers to address performance issues and optimize costs.

Guidelines

Review the following guidelines before you specify a GovCloud account as an AWS target in Workload Optimization Manager.

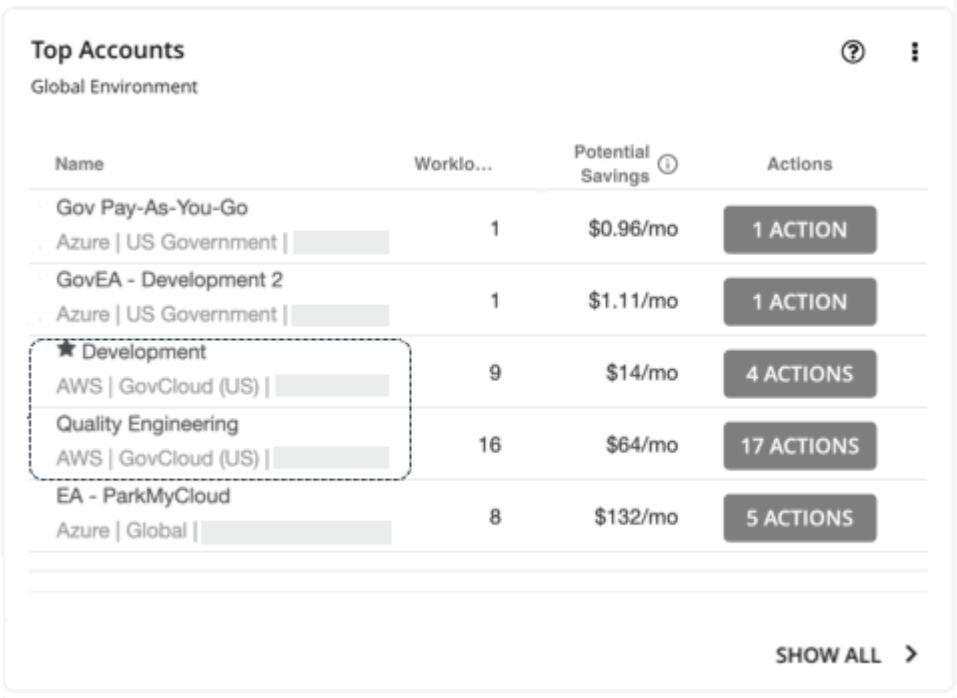
- The GovCloud account must be a management account with organization access, or a member account belonging to a management account that you already added to Workload Optimization Manager as a target. In addition, each account within the organization must have a unique name.
- When you add an AWS target *(on page 391)* in Workload Optimization Manager, be sure to turn on the **GovCloud** toggle to indicate that the target is a GovCloud account.
- To discover RIs in GovCloud, add an AWS Billing target *(on page 389)* with access to the AWS organization and GovCloud billing data.
- Optionally, you can add the AWS standard account that manages invoicing for a GovCloud account as an AWS target. When you add this standard account, it appears in the Top Accounts chart and shows the total projected billed cost for the GovCloud account and the standard account itself.

Information in charts

Use the following charts to view information about your government accounts and workloads.

- Top Accounts chart

  Use the Top Accounts chart as a starting point.

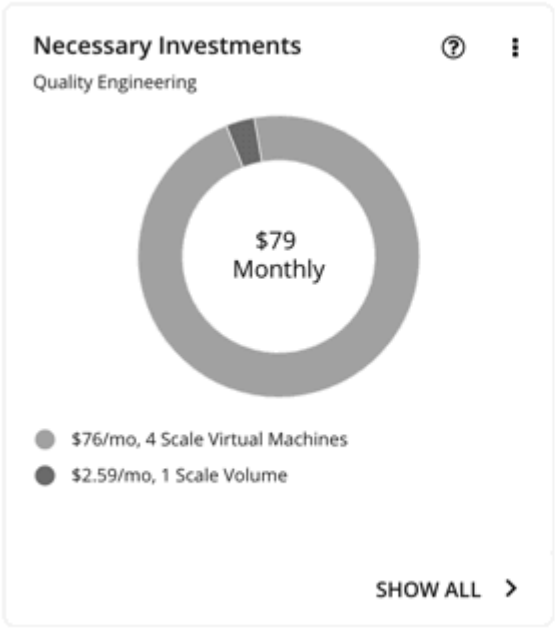


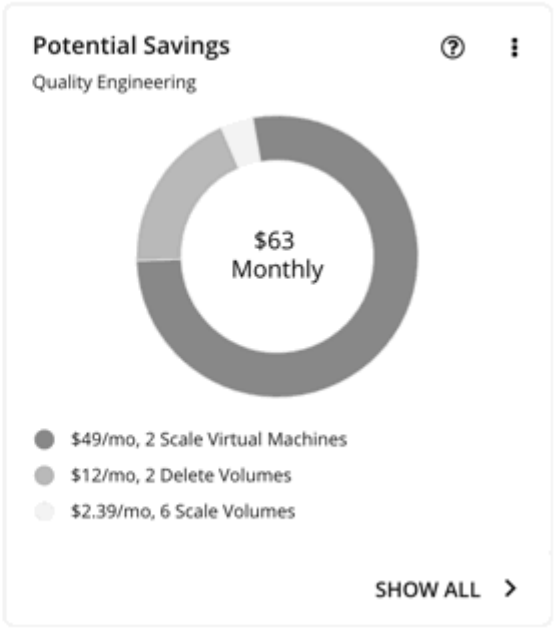
The chart shows the GovCloud management and member accounts that you have added as targets. Accounts with a star symbol are management accounts.

Be aware that the 30-day cost is shown as **N/A** since invoicing for that account is managed through an associated AWS *standard account*. Adding this standard account as a target is optional. When added, it appears in the Top Accounts chart and shows the total billed cost for the GovCloud account and the standard account itself.

- Necessary Investments and Potential Savings charts

  Set the scope to a government account or subscription, and then see the Necessary Investments and Potential Savings charts to evaluate the costs you would incur or save if you execute all the pending actions for your government workloads.





- Discount Inventory chart

The government accounts that you added as targets enable Workload Optimization Manager to gain full insight into the discounts *(on page 157)* that you have purchased for your government workloads. Even as you selectively add secondary targets, Workload Optimization Manager remains aware of all discounts, and how they are utilized across the board. This increases the accuracy of the allocation and purchase recommendations that Workload Optimization Manager generates for your government workloads.

Workload planning

You can run an Optimize Cloud plan to identify performance and efficiency opportunities for existing government workloads, or a Migrate to Cloud plan to migrate government VM groups to another cloud provider.

For on-prem clusters, you can run a Migrate to Cloud plan to see how you can safely migrate the VMs in these clusters to a government account/subscription and region.

AWS metrics collection

It is highly recommended that you enable collection of metrics in your environment. Enabling metrics allows Workload Optimization Manager to generate scale actions to optimize VM resource usage. For Workload Optimization Manager to collect metrics, you must enable the collection of these metrics on the VMs in your environment.

This topic describes the collection of the following metrics:

- Standard memory for AWS VMs
- NVIDIA GPU metrics for AWS VMs running EC2 GPU instance types *(on page 638)*

Metrics collection requirements

To enable metrics collection, you must meet the following requirements. Some requirements may be different depending on whether your VM runs Linux or Windows.

- The VM image must have an SSM agent installed.
  - Linux VMs

    By default, Linux AMIs dated 2017.09 and later include an installed SSM agent.
  - Windows VMs

    You must install the SSM agent on the VMs. For more information, see the AWS documentation.
- Access to the CloudWatch service

  Your AWS instance must have internet access or direct access to CloudWatch to push data to CloudWatch.
- Access from Workload Optimization Manager

  For Workload Optimization Manager to access metrics, the account that it uses to connect to the AWS target must include the correct permissions. If you configured the AWS target through an AWS key (not an IAM role), include the permissions as specified in the section for configuring an AWS target.

  If you use an IAM role for the Workload Optimization Manager connection, that role must include the following minimum permissions:
  - `AmazonEC2ReadOnlyAccess`
  - `AmazonS3ReadOnlyAccess`
  - `AmazonRDSReadOnlyAccess`

Enabling metrics collection

To enable metrics collection for your VMs, perform the following steps:

1. Attach an IAM role to each VM instance.

   Each EC2 instance must have an attached IAM role that grants CloudWatch access. To grant that access, include the `AmazonSSMFullAccess` policy in the role.

   Use the AWS System Manager to attach the necessary roles to your VMs.

NOTE:
If you want to grant the role lesser access, you can use the `AmazonEC2RoleforSSM` policy. This is a custom policy that allows the action `ssm:GetParameter` to access the resource, `arn:aws:ssm:*:*:parameter/*`.

2. Install the CloudWatch agent on your Linux VMs.

   Navigate to the AWS System Manager service for the account and region that you want to configure. In the service, navigate to the **Run Command** screen and set up the **AWS-ConfigureAWSPackage** command to install **AmazonCloudWatchAgent** on your VMs. Use the following required values to install **AmazonCloudWatchAgent**:

Item	Configuration
Action	`Install`
Name	`CloudWatchAgent`

3. Create a parameter store.
   a. In the AWS System Manager, navigate to **Parameter Store** and create a parameter. Copy and paste the JSON object configuration (shown in the next step) into the parameter **Value** field.
   b. Specify a parameter name, such as `AmazonCloudWatch-MyMemoryParam`. You can use a different name, but according to the Amazon documentation, the name must begin with `AmazonCloudWatch`.
   c. In the **value** field copy and paste the JSON configuration (shown in the next step) that corresponds to your VM.
4. Create configuration data for the CloudWatch agent.

   The configuration data is a JSON object that you add as a parameter to the Parameter Store. The object must include the following configuration, depending on whether the object is for a Linux or a Windows VM instance.

   - Linux configuration for standard memory

```
{
   "agent":{
      "metrics_collection_interval":60,
      "logfile":"/opt/aws/amazon-cloudwatch-agent/logs/amazon-cloudwatch-agent.log"
   },
   "metrics":{
      "metrics_collected":{
         "mem":{
            "measurement":[
               {
                  "name":"mem_used",
                  "rename":"MemoryUsed"
               },
               {
                  "name":"mem_available_percent",
                  "rename":"MemoryAvailablePercent"
               },
               {
                  "name":"mem_used_percent",
                  "rename":"MemoryUsedPercent"
               },
               {
                  "name":"mem_available",
                  "rename":"MemoryAvailable"
               }
            ]
         }
      },
      "append_dimensions":{
         "InstanceId":"${aws:InstanceId}",
          "ImageId":"${aws:ImageId}",
```

```
          "InstanceType":"${aws:InstanceType}",
          "AutoScalingGroupName":"${aws:AutoScalingGroupName}"
      },
      "aggregation_dimensions":[
         [
            "AutoScalingGroupName"
         ]
      ]
   }
}
```

- Linux configuration for standard memory and NVIDIA GPU card/memory utilization

```
{
  "agent":{
   "metrics_collection_interval":60,
   "logfile":"/opt/aws/amazon-cloudwatch-agent/logs/amazon-cloudwatch-agent.log"
  },
  "metrics":{
   "namespace": "CWAgent",
   "metrics_collected":{
      "nvidia_gpu": {
         "measurement": [
             "utilization_gpu",
             "memory_used"
          ]
      },
     "mem":{
      "measurement":[
        {
         "name":"mem_available", "rename":"MemoryAvailable", "unit": "Bytes"
        }
      ]
     }
   },
   "append_dimensions":{
     "AutoScalingGroupName": "${aws:AutoScalingGroupName}",
     "ImageId": "${aws:ImageId}",
     "InstanceId": "${aws:InstanceId}",
     "InstanceType": "${aws:InstanceType}"
   }
  }
}
```

- Linux configuration for NVIDIA GPU metrics (DCGM)

  Run the `setup_aws_dcgm_exporter.py` script to automate the collection of NVIDIA GPU metrics through Data Center GPU Manager (DCGM). Certain prerequisites must be met before you run the script. For more information, see this GitHub page.

  If you need assistance with the script, contact your Workload Optimization Manager representative.

- Windows configuration for standard memory

```
{
   "metrics":{
      "namespace":"CWAgent",
      "append_dimensions":{
         "InstanceId":"${aws:InstanceId}",
         "AutoScalingGroupName":"${aws:AutoScalingGroupName}"
      },
      "aggregation_dimensions":[
         [
            "InstanceId"
         ],
         [
```

```
            "AutoScalingGroupName"
         ]
      ],
      "metrics_collected":{
         "Memory":{
            "measurement":[
               {
                  "name":"Available Bytes",
                  "rename":"MemoryAvailable",
                  "unit":"Bytes"
               }
            ],
            "metrics_collection_interval":60
         },
         "Paging File":{
            "measurement":[
               {
                  "name":"% Usage",
                  "rename":"paging_used"
               }
            ],
            "metrics_collection_interval":60,
            "resources":[
               "*"
            ]
         }
      }
   }
}
```

Note that you can configure optional parameters for the CW namespace and region. However, if you configure more metrics for CloudWatch to collect, these metrics do not affect Workload Optimization Manager analysis and they do not show up in the user interface.

5. Create the parameter and record the parameter name for use in a later step.
6. Deploy the CloudWatch parameter to your VMs.
   a. In AWS System Manager, navigate to the **Run Command** screen to configure and run the **AmazonCloudWatch-ManageAgent** command. The configuration should include the following items:

Item	Configuration
Action	`configure`
Mode	`ec2`
Optional Configuration Source	`ssm`
Optional Configuration Location	Specify the name of the parameter that you created earlier.
Optional Restart	`yes` This configuration restarts the CloudWatch Agent, not the VM instance.
Targets	Specify the VMs to which you deploy the Cloud-Watch configuration.

   b. Run the command to enable metrics collections for your instances.
7. Verify that metrics collection is enabled.
   a. Navigate to the CloudWatch page and open **Metrics** in the **CWAgent** namespace.
   b. Inspect the instances by ID. You should see `MemoryAvailable` or `utilization_gpu` and `memory_used` metrics if metrics collection is enabled.

AWS actions

Workload Optimization Manager monitors the state and performance of your workloads and then recommends actions to optimize these workloads at the lowest possible cost.

> **NOTE:**
> Use the Potential Savings and Necessary Investments charts to view pending actions and evaluate their impact on your cloud expenditure.

Historically-informed action generation for AWS workloads

Workload Optimization Manager can generate actions informed by 14 days of historical utilization data immediately after a new AWS target is added. To take advantage of this feature, add at least one AWS target. The IAM role or user for this target must be configured with the `pi:ListAvailableResourceMetrics` permission and all the permissions listed here.

The following components run after you add an AWS target:

- `metrics-processor`
- `metrics-adapter-aws-cloudwatch`
- `metrics-adapter-aws-performance-insights`

Currently, this feature does not support workloads in AWS GovCloud accounts.

Actions for virtual machines (EC2)

Workload Optimization Manager supports the following actions:

- Stop and start (also known as 'parking' actions)

  Stop a VM for a given period of time to reduce your cloud expenses, and then start it at a later time.
- Scale

  Change the VM instance to use a different instance type or tier to optimize performance and costs.
- Discount-related actions

  If you have a high percentage of on-demand VMs, you can reduce your monthly costs by increasing RI or Savings Plans coverage. To increase coverage, you scale VMs to instance types that have existing capacity.

For details, see Actions for AWS VMs *(on page 637)*.

Actions for virtual machine specs (EC2 Auto Scaling groups)

Workload Optimization Manager supports the following actions:

Stop and start (also known as 'parking' actions)

Stop an Auto Scaling group for a given period of time to reduce your cloud expenses, and then start it at a later time.

For details, see Parking: Stop or Start Cloud Resources *(on page 942)*.

Actions for database servers

Workload Optimization Manager supports the following actions:

- Stop and start (also known as 'parking' actions)

  Stop a Database Server for a given period of time to reduce your cloud expenses, and then start it at a later time.
- Scale

  Scale compute and storage resources to optimize performance and costs.

For details, see Cloud Database Server Actions *(on page 685)* and Parking: Stop or Start Cloud Resources *(on page 942)*.

Actions for data warehouses (Redshift provisioned cluster)

Workload Optimization Manager supports the following actions:

Suspend

Suspend a Redshift provisioned cluster to reduce compute costs.

For details, see Suspend actions for Redshift provisioned clusters *(on page 720)*.

Actions for volumes

Workload Optimization Manager supports the following actions:

- Scale

  Scale attached volumes to optimize performance and costs.

- Delete

  Delete unattached volumes as a cost-saving measure. Workload Optimization Manager generates an action immediately after discovering an unattached volume.

For details, see Cloud Volume Actions *(on page 728)*.

Google Cloud

Google Cloud is Google's cloud computing platform. Workload Optimization Manager discovers your Google Cloud resources through a service account, and then optimizes these resources to assure performance at the lowest possible cost.

To connect to Google Cloud, follow the steps and guidelines outlined in the following topics:

- Connecting to Google Cloud *(on page 404)*
- Google Cloud permissions *(on page 418)*
- Google Cloud Billing permissions *(on page 422)*

After connecting to Google Cloud, Workload Optimization Manager monitors and optimizes the resources that it discovered. See the following topics for more information:

- Google Cloud Monitored Resources *(on page 423)*
- Google Cloud Actions *(on page 425)*

Connecting to Google Cloud

To connect Workload Optimization Manager to your Google Cloud environment, perform the following tasks:

1. Enable the required APIs in Google Cloud. *(on page 405)*

   When enabled, these APIs allow Workload Optimization Manager to discover your Google Cloud resources and billing data.

2. Configure firewall and proxy access. *(on page 406)*

   If you run Workload Optimization Manager behind a firewall, configure firewall and proxy access.

3. Set up a service account for workload monitoring in Google Cloud. *(on page 406)*

   Workload Optimization Manager discovers and monitors your Google Cloud workloads through a service account that you set up in Google Cloud. You can set up a service account that accesses your entire organization, individual projects, or individual folders.

4. Add a Google Cloud target in Workload Optimization Manager. *(on page 413)*

   Authorize a secure connection through your service account.

   To authorize the connection, add a Google Cloud target in the Workload Optimization Manager user interface.

5. Set up a billing export in Google Cloud. *(on page 414)*

Workload Optimization Manager uses billing data from a billing export to BigQuery to discover Committed Use Discounts and visualize historical cloud expenses.

6. Set up a service account for billing data monitoring in Google Cloud. *(on page 414)*

   Workload Optimization Manager retrieves billing data through a service account that you set up in Google Cloud. You can use the service account that you previously set up for workload monitoring, or create a new one specifically for billing data monitoring.

7. Add a Google Cloud Billing target in Workload Optimization Manager. *(on page 417)*

   Authorize a secure connection to BigQuery.

   To authorize the connection, add a Google Cloud Billing target in the Workload Optimization Manager user interface.

Enabling required Google Cloud APIs

For Workload Optimization Manager to discover your Google Cloud environment and billing data, you must enable the following APIs:

- Cloud Resource Manager API

  Creates, reads, and updates metadata for resource containers.

- Compute Engine API

  Creates VMs and volumes.

- Cloud Billing API

  Enables developers to manage billing for their projects programmatically.

- BigQuery API

  A data platform for customers to create, manage, share, and query data.

To enable these APIs:

1. Navigate the Google Cloud Console to the library of APIs.

   On the Google Cloud Console home page, navigate to **APIs & Services > Library**.

2. Search for the API you want to enable.

   In the API Library **Search** box, enter the name of the API you want to enable. Then press **Enter** to execute the search. Repeat these steps for each of:

   - Cloud Resource Manager API
   - Compute Engine API
   - Cloud Billing API
   - BigQuery API

3. Enable the given API.

   In the list that appears, click the API name to navigate to that API page. If the API is not already enabled, click **Enable**.

   After you enable the given API, the console displays a details page for that API.

4. Navigate to the console Home page.

   For each API you want to enable, navigate back to the home page and repeat these steps.

Next step

If you run Workload Optimization Manager behind a firewall, configure firewall and proxy access (see this topic *(on page 406)*). Otherwise, set up a service account for workload monitoring in Google Cloud (see this topic *(on page 406)*).

Configuring firewall and proxy access for Google Cloud

If you run Workload Optimization Manager behind a firewall, configure Workload Optimization Manager to allow unrestricted access to the following Google Cloud URLs:

- https://cloudresourcemanager.googleapis.com
- https://cloudbilling.googleapis.com
- https://compute.googleapis.com
- https://monitoring.googleapis.com
- https://bigquery.googleapis.com
- https://oauth2.googleapis.com

Next step

In Google Cloud, set up a service account for workload monitoring. For details, see this topic *(on page 406)*.

Setting up a Google Cloud service account for workload monitoring

This topic describes the steps to set up a valid service account that Workload Optimization Manager will use to connect to your Google Cloud environment. The access that you grant this service account determines the Google Cloud workloads that Workload Optimization Manager discovers, monitors, and optimizes.

Task overview

To set up a service account, perform the following tasks in Google Cloud:

1. Create a service account for workload monitoring.
2. Create a custom role and then assign the role to the service account.

Creating a service account for workload monitoring

Create a service account and generate a key file for the account. The key file is required when adding a Google Cloud target in the Workload Optimization Manager user interface.

For seamless monitoring of your Google Cloud workloads, create the service account in a project that does not typically hit the rate limits enforced by Google Cloud, such as a non-production project.

1. In the project that will host the new service account, open a `gcloud` CLI session.
2. Create a service account.

   ```
   gcloud iam service-accounts create <SERVICE_ACCOUNT_NAME>
   ```

   Where:

   <SERVICE_ACCOUNT_NAME> is the internal name of the new service account. The name must be between 6 and 30 characters in length.
3. Record the following information for later use.
   - *<SERVICE_ACCOUNT_NAME>*
   - *<PROJECT_ID>*

     NOTE:
     <PROJECT_ID> identifies the project that hosts the service account. This information is needed if you need to review or edit the service account later.
4. Generate a key file for the service account.

```
gcloud iam service-accounts keys create <KEY_FILE_NAME> \
  --iam-account=<SERVICE_ACCOUNT_NAME>@<PROJECT_ID>.iam.gserviceaccount.com
```

Where:

- *<KEY_FILE_NAME>* is your preferred name for the key file.
- *<SERVICE_ACCOUNT_NAME>* is the name of the service account that you created.
- *<PROJECT_ID>* is the project that hosts the service account.

5. Download the key file (JSON file) to your local machine. You will use the key file later when you add a Google Cloud target in the Workload Optimization Manager user interface.

```
cloudshell download <KEY_FILE_NAME>
```

Where:

<KEY_FILE_NAME> is the key file name that you specified in the previous step.

Overview of custom roles

You can assign the following custom roles to the service account that you created.

- (Required) Custom role for workload monitoring

  This custom role specifies the permissions that Workload Optimization Manager needs to discover and monitor workloads in your entire organization, or in individual folders or projects.

Google Cloud resources to monitor	Task
Organization	Create a custom role at the organization level.
Individual folders	Create a custom role at the organization level. It is not possible to create custom roles at the folder level.
Individual projects	Create a custom role at the project level.

- (Optional) Custom role for action execution

  To execute actions from Workload Optimization Manager, create a role that specifies the required permissions for executing actions.

Creating and assigning custom roles (organization level)

NOTE:
Skip to the next section if you want Workload Optimization Manager to monitor individual projects.

1. Create a custom role for workload monitoring.

```
gcloud iam roles create <ROLE_ID> --organization=<ORGANIZATION_ID> \
  --title='CWOM Role: Access - Organization' \
  --description='Minimum permissions to manage the Google Cloud organization' \
  --permissions="billing.accounts.list,\
billing.resourceAssociations.list,\
compute.commitments.list,\
compute.disks.get,\
compute.disks.list,\
compute.diskTypes.list,\
compute.instances.get,\
compute.instances.list,\
compute.instanceGroupManagers.get,\
compute.instanceGroupManagers.list,\
compute.instanceGroups.get,\
compute.instanceGroups.list,\
compute.machineTypes.get,\
compute.machineTypes.list,\
compute.regions.list,\
compute.zones.list,\
container.clusters.get,\
logging.logEntries.list,\
logging.views.get,\
logging.views.list,\
monitoring.services.get,\
monitoring.services.list,\
monitoring.timeSeries.list,\
resourcemanager.folders.get,\
resourcemanager.folders.list,\
resourcemanager.organizations.get,\
resourcemanager.projects.get,\
resourcemanager.projects.list,\
serviceusage.services.get" --stage=ALPHA
```

   Where:

   - *<ROLE_ID>* is your preferred ID for the custom role.
   - *<ORGANIZATION_ID>* is the organization that Workload Optimization Manager will monitor.

2. Assign the custom role to the service account.

```
gcloud organizations add-iam-policy-binding <ORGANIZATION_ID> \
  --member=serviceAccount:<SERVICE_ACCOUNT_NAME>@<PROJECT_ID>.iam.gserviceaccount.com
 --role=<ROLE_NAME>
```

   Where:

   - *<ORGANIZATION_ID>* is the organization that Workload Optimization Manager will monitor.
   - *<SERVICE_ACCOUNT_NAME>* is the name of the service account that you created.
   - *<PROJECT_ID>* is the project that hosts the service account.
   - *<ROLE_NAME>* is the complete path for the role name, expressed as follows:

```
organizations/<ORGANIZATION_ID>/roles/<ROLE_ID>
```

NOTE:
<ROLE_ID> was created in a previous step.

3. Add the predefined **Billing Account Viewer** role to the service account.

```
gcloud organizations add-iam-policy-binding <ORGANIZATION_ID> \
  --member=serviceAccount:<SERVICE_ACCOUNT_NAME>@<PROJECT_ID>.iam.gserviceaccount.com \
  --role=roles/billing.viewer
```

Where:

- *<ORGANIZATION_ID>* is your Google Cloud organization.
- *<SERVICE_ACCOUNT_NAME>* is the name of the service account that you created.
- *<PROJECT_ID>* is the project that hosts the service account.

4. (Optional) Create a custom role to execute actions from Workload Optimization Manager.

```
gcloud iam roles create <ROLE_ID_ACTION> --organization=<ORGANIZATION_ID> \
  --title='CWOM Role: Org Action Execution' \
  --description='Minimum permissions to execute actions in the Google Cloud organization' \
  --permissions="compute.disks.create,\
compute.disks.createSnapshot,\
compute.disks.delete,\
compute.disks.resize,\
compute.disks.setLabels,\
compute.disks.update,\
compute.disks.use,\
compute.disks.useReadOnly,\
compute.globalOperations.get,\
compute.instanceGroupManagers.update,\
compute.instanceGroups.get,\
compute.instanceGroups.list,\
compute.instanceGroups.use,\
compute.instances.attachDisk,\
compute.instances.detachDisk,\
compute.instances.setLabels,\
compute.instances.setMachineType,\
compute.instances.start,\
compute.instances.stop,\
compute.instances.useReadOnly,\
compute.instanceTemplates.list,\
compute.instantSnapshots.list,\
compute.regionOperations.get,\
compute.reservations.list,\
compute.resourcePolicies.use,\
compute.snapshots.create,\
compute.snapshots.delete,\
compute.snapshots.get,\
compute.snapshots.list,\
compute.snapshots.useReadOnly,\
compute.zoneOperations.get,\
container.clusters.update,\
iam.serviceAccounts.actAs" --stage=ALPHA
```

Where:

- *<ROLE_ID_ACTION>* is your preferred ID for the custom role for action execution.
- *<ORGANIZATION_ID>* is the organization that Workload Optimization Manager will monitor.

5. (Optional) Assign the custom role for action execution to the service account.

```
gcloud projects add-iam-policy-binding <ORGANIZATION_ID> \
  --member=serviceAccount:<SERVICE_ACCOUNT_NAME>@<PROJECT_ID>.iam.gserviceaccount.com
 --role=<ROLE_NAME_ACTION>
```

Where:

- *<ORGANIZATION_ID>* is the organization that Workload Optimization Manager will monitor.
- *<SERVICE_ACCOUNT_NAME>* is the name of the service account that you created.
- *<PROJECT_ID>* is the project that hosts the service account.
- *<ROLE_NAME_ACTION>* is the complete path for the role name, expressed as follows:
  `organizations/<ORGANIZATION_ID>/roles/<ROLE_ID_ACTION>`

  NOTE:
  <ROLE_ID_ACTION> was created in a previous step.

Creating and assigning custom roles (project level)

NOTE: See the previous section if you want Workload Optimization Manager to monitor individual folders or your entire organization.

1. Create a custom role for workload monitoring.

```
gcloud iam roles create <ROLE_ID> --project=<PROJECT_ID_MONITOR> \
  --title='CWOM Role: Min Access - Project' \
  --description='Minimum permissions to manage the Google Cloud project' \
  --permissions="compute.commitments.list,\
compute.disks.get,\
compute.disks.list,\
compute.diskTypes.list,\
compute.instances.get,\
compute.instances.list,\
compute.instanceGroupManagers.get,\
compute.instanceGroupManagers.list,\
compute.instanceGroups.get,\
compute.instanceGroups.list,\
compute.machineTypes.get,\
compute.machineTypes.list,\
compute.regions.list,\
compute.zones.list,\
container.clusters.get,\
logging.logEntries.list,\
logging.views.get,\
logging.views.list,\
monitoring.services.get,\
monitoring.services.list,\
monitoring.timeSeries.list,\
resourcemanager.projects.get,\
serviceusage.services.get" --stage=ALPHA
```

   Where:

   - *<ROLE_ID>* is your preferred ID for the custom role.
   - *<PROJECT_ID_MONITOR>* is the project that Workload Optimization Manager will monitor.
2. Assign the custom role to the service account.

```
gcloud projects add-iam-policy-binding <PROJECT_ID_MONITOR> \
  --member=serviceAccount:<SERVICE_ACCOUNT_NAME>@<PROJECT_ID>.iam.gserviceaccount.com
 --role=<ROLE_NAME>
```

Where:

- *<PROJECT_ID_MONITOR>* is the project that Workload Optimization Manager will monitor.
- *<SERVICE_ACCOUNT_NAME>* is the name of the service account that you created.
- *<PROJECT_ID>* is the project that hosts the service account.
- *<ROLE_NAME>* is the complete path for the role name, expressed as follows:
  `projects/<PROJECT_ID_MONITOR>/roles/<ROLE_ID>`

  NOTE:
  <ROLE_ID> was created in a previous step.

3. Add the predefined **Billing Account Viewer** role to the service account.

```
gcloud organizations add-iam-policy-binding <ORGANIZATION_ID> \
  --member=serviceAccount:<SERVICE_ACCOUNT_NAME>@<PROJECT_ID>.iam.gserviceaccount.com \
  --role=roles/billing.viewer
```

Where:

- *<ORGANIZATION_ID>* is your Google Cloud organization.
- *<SERVICE_ACCOUNT_NAME>* is the name of the service account that you created.
- *<PROJECT_ID>* is the project that hosts the service account.

4. (Optional) Create a custom role to execute actions from Workload Optimization Manager.

```
gcloud iam roles create <ROLE_ID_ACTION> --project=<PROJECT_ID_MONITOR> \
  --title='CWOM Role: Project Action Execution' \
  --description='Minimum permissions to execute actions in the Google Cloud project' \
  --permissions="compute.disks.create,\
compute.disks.createSnapshot,\
compute.disks.delete,\
compute.disks.resize,\
compute.disks.setLabels,\
compute.disks.update,\
compute.disks.use,\
compute.disks.useReadOnly,\
compute.globalOperations.get,\
compute.instanceGroupManagers.update,\
compute.instanceGroups.get,\
compute.instanceGroups.list,\
compute.instanceGroups.use,\
compute.instances.attachDisk,\
compute.instances.detachDisk,\
compute.instances.setLabels,\
compute.instances.setMachineType,\
compute.instances.start,\
compute.instances.stop,\
compute.instances.useReadOnly,\
compute.instanceTemplates.list,\
compute.instantSnapshots.list,\
compute.regionOperations.get,\
compute.reservations.list,\
compute.resourcePolicies.use,\
compute.snapshots.create,\
compute.snapshots.delete,\
compute.snapshots.get,\
compute.snapshots.list,\
compute.snapshots.useReadOnly,\
compute.zoneOperations.get,\
container.clusters.update,\
iam.serviceAccounts.actAs" --stage=ALPHA
```

Where:

- *<ROLE_ID_ACTION>* is your preferred ID for the custom role for action execution.
- *<PROJECT_ID_MONITOR>* is the project that Workload Optimization Manager will monitor.

5. (Optional) Assign the custom role for action execution to the service account.

```
gcloud projects add-iam-policy-binding <PROJECT_ID_MONITOR> \
  --member=serviceAccount:<SERVICE_ACCOUNT_NAME>@<PROJECT_ID>.iam.gserviceaccount.com
 --role=<ROLE_NAME_ACTION>
```

Where:

- *<PROJECT_ID_MONITOR>* is the project that Workload Optimization Manager will monitor.
- *<SERVICE_ACCOUNT_NAME>* is the name of the service account that you created.
- *<PROJECT_ID>* is the project that hosts the service account.
- *<ROLE_NAME_ACTION>* is the complete path for the role name, expressed as follows:
  `projects/<PROJECT_ID_MONITOR>/roles/<ROLE_ID_ACTION>`

  NOTE:
  <ROLE_ID_ACTION> was created in a previous step.

Next step

In the Workload Optimization Manager user interface, add a Google Cloud target. For details, see this topic *(on page 413)*.

Adding a Google Cloud target

Add an Google Cloud target in the Workload Optimization Manager user interface to monitor and optimize workloads in your Google Cloud environment. This target specifies the service account that Workload Optimization Manager will use to connect to Google Cloud.

Before performing this task, be sure you have the key file for the service account that Workload Optimization Manager will use to connect to Google Cloud. If you do not have this key, follow the steps outlined in this topic *(on page 406)*.

Adding the target

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **Google Cloud**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
    - Display name

      Specify a name that uniquely identifies this connection.

      This name is for display purposes only and does not need to match any name in Google Cloud.
    - Service account details

      Upload the service account key (JSON file).

      This is the JSON for the key file that you generated in a previous task *(on page 406)*.
    - Proxy configuration

      Specify proxy information only if you connect through a proxy.
        - Proxy hostname or IP address

          Specify the address of the proxy used for this target.
        - Proxy port

          Specify the port to use with the proxy used for this target. By default, the port is 8080.
        - Proxy username

          Specify the username to use with the proxy.
        - Proxy password

          Specify the password to use with the proxy.
        - Connect using HTTPS

          Select this option if Workload Optimization Manager will connect to the proxy through HTTPS.

After adding a Google Cloud target

Workload Optimization Manager starts to discover the projects that define compute, storage, and networking resources for your workloads. It then creates a derived target for each discovered project. Derived targets are not directly modifiable within Workload Optimization Manager but can be validated like any other target.

Workload Optimization Manager discovers a broader resource hierarchy if you added a target with permissions to access folders or your entire organization.

Next step

In Google Cloud, set up a billing export. For details, see this topic *(on page 414)*.

Setting up a Google Cloud billing export

This topic describes the billing export that Workload Optimization Manager will use to monitor billing data exported to BigQuery. Without this data, Workload Optimization Manager cannot discover any cost data used for analysis.

When you export Google Cloud billing data from BigQuery, you can choose to export standard or detailed usage cost data. Workload Optimization Manager recommends detailed usage cost data because it includes granular information for VMs, which the platform uses to generate accurate VM scaling actions. This information is also reflected in cloud charts (such as the Workload Cost Breakdown chart) when you set the scope to individual VMs. Note that support for standard usage cost data will be discontinued in a future release.

Setup guidelines:

- For an overview of billing export, see this Google Cloud page.
- To set up a billing export, follow the steps in this Google Cloud page.
- Shortly after you set up a billing export, Google Cloud automatically creates the following billing data tables in the BigQuery dataset. Be sure to record these table names. You will need them later when you add a Google Cloud Billing target in the Workload Optimization Manager user interface.
  - Standard usage cost data table

    In your BigQuery dataset, this table is named *gcp_billing_export_v1_<Billing_Account_ID>*.
  - Detailed usage cost data table

    This table includes all the data fields from the standard usage cost data table, along with additional fields that provide resource-level cost data, such as a virtual machine or SSD that generates service usage. In your BigQuery dataset, this table is named *gcp_billing_export_resource_v1_<Billing_Account_ID>*.
  - Pricing data export table

    In your BigQuery dataset, this table is named *cloud_pricing_export*. However, if you have negotiated pricing or are billed in a non-US dollar currency, record the corresponding pricing export table name.

Next step

In Google Cloud, set up a service account for billing data monitoring. For details, see this topic *(on page 414)*.

Setting up a Google Cloud service account for billing data monitoring

This topic describes the steps to set up a valid service account that Workload Optimization Manager will use to connect to your Google Cloud environment and monitor billing data. Without this data, Workload Optimization Manager cannot discover any cost data used for analysis.

Task overview

To set up a service account, perform the following tasks in Google Cloud:

1. Create a service account for billing data monitoring.
2. Create a custom role and then assign the role to the service account.
3. Add the Billing Account Viewer role to the service account.

Creating a service account for billing data monitoring

Create a service account and generate a key file for the account. The key file is required when adding a Google Cloud Billing target in the Workload Optimization Manager user interface.

For seamless monitoring of your Google Cloud resources, create the service account in a project that does not typically hit the rate limits enforced by Google Cloud, such as a non-production project.

NOTE:
You can use the existing service account that you set up for workload monitoring, or create a new one specifically for billing data monitoring. Skip to the next task if you plan to use your existing service account.

1. In the project that will host the new service account, open a `gcloud` CLI session.
2. Create a service account.

   ```
   gcloud iam service-accounts create <SERVICE_ACCOUNT_NAME>
   ```

   Where:

   <SERVICE_ACCOUNT_NAME> is the internal name of the new service account. The name must be between 6 and 30 characters in length.
3. Record the following information for later use.
   - *<SERVICE_ACCOUNT_NAME>*
   - *<PROJECT_ID>*

     NOTE:
     <PROJECT_ID> identifies the project that hosts the service account. This information is needed if you need to review or edit the service account later.

4. Generate a key file for the service account.

   ```
   gcloud iam service-accounts keys create <KEY_FILE_NAME> \
     --iam-account=<SERVICE_ACCOUNT_NAME>@<PROJECT_ID>.iam.gserviceaccount.com
   ```

   Where:
   - *<KEY_FILE_NAME>* is your preferred name for the key file.
   - *<SERVICE_ACCOUNT_NAME>* is the name of the service account that you created.
   - *<PROJECT_ID>* is the project that hosts the service account.
5. Download the key file to your local machine. You will use the key file later when you add a Google Cloud target in the Workload Optimization Manager user interface.

   ```
   cloudshell download <KEY_FILE_NAME>
   ```

   Where:

   <KEY_FILE_NAME> is the key file name that you specified in the previous step.

Creating and assigning a custom role for billing data monitoring

This custom role is **required** and must be created in the project that stores billing data. The role specifies the permissions that Workload Optimization Manager needs to discover and monitor billing data.

1. Create a custom role.

```
gcloud iam roles create <ROLE_ID_BILL> --project=<PROJECT_ID_BILL> \
  --title='CWOM Billing Data Viewer Role' \
  --description='Minimum permissions to view \
  billed cost and pricing in the Google Cloud project' \
  --permissions="bigquery.jobs.create,\
bigquery.tables.get,\
bigquery.tables.getData,\
bigquery.tables.list,\
billing.accounts.get,\
billing.resourceAssociations.list,\
compute.commitments.list,\
compute.diskTypes.list,\
compute.machineTypes.list,\
compute.regions.list,\
compute.zones.list" --stage=ALPHA
```

   Where:

   - *<ROLE_ID_BILL>* is your preferred ID for the custom role.
   - *<PROJECT_ID_BILL>* is the project that stores billing data.

2. Assign the custom role to the service account.

```
gcloud projects add-iam-policy-binding <PROJECT_ID_BILL> \
  --member=serviceAccount:<SERVICE_ACCOUNT_NAME>@<PROJECT_ID>.iam.gserviceaccount.com
 --role=<ROLE_NAME_BILL>
```

   Where:

   - *<PROJECT_ID_BILL>* is the project that stores billing data.
   - *<SERVICE_ACCOUNT_NAME>* is the name of the service account that you created.
   - *<PROJECT_ID>* is the project that hosts the service account.
   - *<ROLE_NAME_BILL>* is the complete path for the role name, expressed as follows:

     `projects/<PROJECT_ID_BILL>/roles/<ROLE_ID_BILL>`

     NOTE:
     <ROLE_ID_BILL> was created in a previous step.

Adding the billing account viewer role to the service account

1. Add the predefined **Billing Account Viewer** role to the service account.

```
gcloud organizations add-iam-policy-binding <ORGANIZATION_ID> \
  --member=serviceAccount:<SERVICE_ACCOUNT_NAME>@<PROJECT_ID>.iam.gserviceaccount.com \
  --role=roles/billing.viewer
```

   Where:

   - *<ORGANIZATION_ID>* is your Google Cloud organization.
   - *<SERVICE_ACCOUNT_NAME>* is the name of the service account that you created.
   - *<PROJECT_ID>* is the project that hosts the service account.

Next step

In the Workload Optimization Manager user interface, add a Google Cloud Billing target. For details, see this topic *(on page 417)*.

Adding a Google Cloud Billing target

The Google Cloud Billing target grants Workload Optimization Manager access to billing data from a billing export to BigQuery. Workload Optimization Manager uses this data to visualize historical cloud expenses and discover Committed Use Discounts.

Points to consider:

- The Google Cloud Billing target can retrieve billing data from BigQuery data sets across all applicable regions, both inside and outside the US.
- You can add multiple billing targets. Data for these targets will be aggregated and shown when you set the scope to your global environment.

Before adding a target, be sure you have the key file for the service account that Workload Optimization Manager will use to connect to Google Cloud and discover billing data. If you do not have this key, follow the steps outlined in this topic *(on page 414)*.

Adding the target

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **Google Cloud Billing**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
    - Display name

      Specify a name that uniquely identifies this connection.

      This name is for display purposes only and does not need to match any name in Google Cloud.
    - Service account details

      Upload the service account key (JSON file).

      This is the JSON for the key file that you generated in a previous task *(on page 406)*.
    - GCP Project ID

      Specify the ID for the project that stores billing data.

      This is the unique ID assigned to the project associated with the billing account. Costs accrued to this project are charged to the billing account you are adding.
    - BigQuery billing export dataset name

      Specify the dataset name associated with your BigQuery billing export.

      This data set is for billed costs. You can find the data set name in the Google Cloud Billing dashboard under **Billing export / BIGQUERY EXPORT**.
    - Type of billing data enabled for export
        - Standard usage cost data
        - Detailed usage cost data

      When you export Google Cloud billing data from BigQuery, you can choose to export standard or detailed usage cost data. Workload Optimization Manager recommends detailed usage cost data because it includes granular information for VMs, which the platform uses to generate accurate VM scaling actions. This information is also reflected in cloud charts (such as the Workload Cost Breakdown chart) when you set the scope to individual VMs. Note that support for standard usage cost data will be discontinued in a future release.
    - BigQuery billing cost table name

      Specify the cost table name associated with the BigQuery billing export.

The cost table names in your BigQuery dataset are as follows:

- Standard usage cost data

  gcp_billing_export_v1_<Billing_Account_ID>
- Detailed usage cost data

  gcp_billing_export_resource_v1_<Billing_Account_ID>

- Billing account ID

  Specify the billing account ID associated with your project. For help finding the ID, see this Google Cloud page.

  This field is required if you configure **BigQuery Pricing Export Data Set Name** and **BigQuery Pricing Export Table Name**.
- BigQuery pricing export dataset name

  Specify the dataset name associated with your BigQuery pricing export.

  You can find the dataset name in the Google Cloud Billing dashboard under **Billing export / BIGQUERY EXPORT**.
- BigQuery pricing export table name

  Specify the table name associated with your BigQuery pricing export.

  In your BigQuery dataset, the pricing export table is named *cloud_pricing_export*.

  If you have negotiated pricing or are billed in a non-US dollar currency, set up a pricing export in Google Cloud and then specify the corresponding pricing export table name.
- Proxy configuration

  Specify proxy information only if you connect through a proxy.

  - Proxy hostname or IP address

    Specify the address of the proxy used for this target.
  - Proxy port

    Specify the port to use with the proxy used for this target. By default, the port is 8080.
  - Proxy username

    Specify the username to use with the proxy.
  - Proxy password

    Specify the password to use with the proxy.
  - Connect using HTTPS

    Select this option if Workload Optimization Manager will connect to the proxy through HTTPS.

After adding a Google Cloud billing target

You have completed the required tasks for connecting to Google Cloud. Workload Optimization Manager can now monitor your Google Cloud workloads and billing data, and recommend actions to optimize these workloads at the lowest possible cost. Review the following topics for more information:

- Google Cloud Monitored Resources *(on page 423)*
- Google Cloud Actions *(on page 425)*

Reference: Google Cloud permissions

Workload Optimization Manager requires specific permissions to monitor your Google Cloud workloads and billing data.

Permissions for the Google Cloud target

The Google Cloud target that you add to the Workload Optimization Manager user interface monitors your Google Cloud workloads. You can grant the target permissions to monitor workloads in individual projects, individual folders, or your entire organization.

Optionally, you can grant the target permissions to execute actions for your workloads automatically.

- Project-level permissions

  NOTE:
  Be sure to set up a service account *(on page 406)* in Google Cloud, create a custom role *(on page 410)*, and then assign the custom role to the service account. You need to specify these permissions when you create the custom role.

  When a service account has been properly configured for use with Workload Optimization Manager, add a Google Cloud target *(on page 413)* in the user interface.

  - (Required) Monitoring permissions
    - `compute.commitments.list`
    - `compute.disks.get`
    - `compute.disks.list`
    - `compute.diskTypes.list`
    - `compute.instances.get`
    - `compute.instances.list`
    - `compute.instanceGroupManagers.get`
    - `compute.instanceGroupManagers.list`
    - `compute.instanceGroups.get`
    - `compute.instanceGroups.list`
    - `compute.machineTypes.get`
    - `compute.machineTypes.list`
    - `compute.regions.list`
    - `compute.zones.list`
    - `container.clusters.get`
    - `logging.logEntries.list`
    - `logging.views.get`
    - `logging.views.list`
    - `monitoring.services.get`
    - `monitoring.services.list`
    - `monitoring.timeSeries.list`
    - `resourcemanager.projects.get`
    - `serviceusage.services.get`
  - (Optional) Action execution permissions
    - `compute.disks.create`
    - `compute.disks.createSnapshot`
    - `compute.disks.delete`
    - `compute.disks.resize`
    - `compute.disks.setLabels`
    - `compute.disks.update`
    - `compute.disks.use`
    - `compute.disks.useReadOnly`
    - `compute.globalOperations.get`

- compute.instanceGroupManagers.update
- compute.instanceGroups.get
- compute.instanceGroups.list
- compute.instanceGroups.use
- compute.instances.attachDisk
- compute.instances.detachDisk
- compute.instances.setLabels
- compute.instances.setMachineType
- compute.instances.start
- compute.instances.stop
- compute.instances.useReadOnly
- compute.instanceTemplates.list
- compute.instantSnapshots.list
- compute.regionOperations.get
- compute.reservations.list
- compute.resourcePolicies.use
- compute.snapshots.create
- compute.snapshots.delete
- compute.snapshots.get
- compute.snapshots.list
- compute.snapshots.useReadOnly
- compute.zoneOperations.get
- container.clusters.update
- iam.serviceAccounts.actAs

▪ Folder-level permissions

To monitor workloads at the folder level, organization-level permissions are required (see the next item).

▪ Organization-level permissions

NOTE:
Be sure to set up a service account *(on page 406)* in Google Cloud, create a custom role *(on page 408)*, and then assign the custom role to the service account. You need to specify these permissions when you create the custom role.

When a service account has been properly configured for use with Workload Optimization Manager, add a Google Cloud target *(on page 413)* in the user interface.

– (Required) Monitoring permissions

- Workload monitoring
  - billing.accounts.list
  - billing.resourceAssociations.list
  - compute.commitments.list
  - compute.disks.get
  - compute.disks.list
  - compute.diskTypes.list
  - compute.instances.get

        - `compute.instances.list`
        - `compute.instanceGroupManagers.get`
        - `compute.instanceGroupManagers.list`
        - `compute.instanceGroups.get`
        - `compute.instanceGroups.list`
        - `compute.machineTypes.get`
        - `compute.machineTypes.list`
        - `compute.regions.list`
        - `compute.zones.list`
        - `container.clusters.get`
        - `logging.logEntries.list`
        - `logging.views.get`
        - `logging.views.list`
        - `monitoring.services.get`
        - `monitoring.services.list`
        - `monitoring.timeSeries.list`
        - `resourcemanager.projects.get`
        - `serviceusage.services.get`
    - Resource hierarchy monitoring
        - `resourcemanager.folders.get`
        - `resourcemanager.folders.list`
        - `resourcemanager.organizations.get`
        - `resourcemanager.projects.get`
        - `resourcemanager.projects.list`
- (Optional) Action execution permissions
    - `compute.disks.create`
    - `compute.disks.createSnapshot`
    - `compute.disks.delete`
    - `compute.disks.resize`
    - `compute.disks.setLabels`
    - `compute.disks.update`
    - `compute.disks.use`
    - `compute.disks.useReadOnly`
    - `compute.globalOperations.get`
    - `compute.instanceGroupManagers.update`
    - `compute.instanceGroups.get`
    - `compute.instanceGroups.list`
    - `compute.instanceGroups.use`
    - `compute.instances.attachDisk`
    - `compute.instances.detachDisk`
    - `compute.instances.setLabels`
    - `compute.instances.setMachineType`

- compute.instances.start
- compute.instances.stop
- compute.instances.useReadOnly
- compute.instanceTemplates.list
- compute.instantSnapshots.list
- compute.regionOperations.get
- compute.reservations.list
- compute.resourcePolicies.use
- compute.snapshots.create
- compute.snapshots.delete
- compute.snapshots.get
- compute.snapshots.list
- compute.snapshots.useReadOnly
- compute.zoneOperations.get
- container.clusters.update
- iam.serviceAccounts.actAs

Permissions for the Google Cloud Billing target

The Google Cloud Billing target grants Workload Optimization Manager access to billing data from a billing export to BigQuery. Workload Optimization Manager uses this data to visualize historical cloud expenses and discover Committed Use Discounts.

NOTE:
Be sure to set up a service account *(on page 414)* in Google Cloud, create a custom role *(on page 416)*, and then assign the custom role to the service account. You need to specify these permissions when you create the custom role.

When a service account has been properly configured for use with Workload Optimization Manager, add a Google Cloud Billing target *(on page 417)* in the user interface.

- bigquery.jobs.create
- bigquery.tables.get
- bigquery.tables.getData
- bigquery.tables.list
- billing.accounts.get
- billing.resourceAssociations.list
- compute.commitments.list
- compute.diskTypes.list
- compute.machineTypes.list
- compute.regions.list
- compute.zones.list

Billing data exports for Google Cloud

When you export Google Cloud billing data from BigQuery, you can choose to export standard or detailed usage cost data. Workload Optimization Manager recommends detailed usage cost data because it includes granular information for VMs, which the platform uses to generate accurate VM scaling actions. This information is also reflected in cloud charts (such as the Workload Cost Breakdown chart) when you set the scope to individual VMs. Note that support for standard usage cost data will be discontinued in a future release.

If you configured your Google Cloud Billing target *(on page 417)* to retrieve standard usage cost data from BigQuery, switch to the detailed usage cost data by following these steps:

1. Export detailed usage cost data from BigQuery.
2. In the Workload Optimization Manager user interface, go to **Settings > Target Configuration** and then open your Google Cloud Billing target for editing.
3. Select **Detailed usage cost data**.
4. In the **BigQuery billing cost table name** field, specify the name of the detailed usage cost data table.

Google Cloud monitored resources

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

Google Cloud	Workload Optimization Manager
Virtual Machine (VM) Instance	Virtual Machine (VM)
Disk	Volume
Zone	Zone
Region	Region

Points to consider:

- Google Cloud *projects*, *folders*, and *billing accounts* do not appear as entities in the supply chain. Use Search to scope to these resources. In Search, projects are grouped under Accounts, folders under Folders, and billing accounts under Billing Families.
- Workload Optimization Manager supports discovery and management of entities in certain Google Cloud regions. For details, see Supported Google Cloud regions *(on page 424)*.

Monitored resources for virtual machines

> **NOTE:**
> Workload Optimization Manager discovers Google Cloud labels attached to VMs as tags. You can filter VMs by tags when you use Search or create groups. The Action Details page for a pending VM action also lists all the discovered tags.

Workload Optimization Manager monitors the following resources:

- Virtual memory (vMem)

  Virtual memory (vMem) is the measurement of memory that is in use.

  Google Cloud collects memory metrics via Ops Agent. In order for Workload Optimization Manager to retrieve these metrics, install and configure Ops Agent on each VM that it monitors. See Ops Agent installation instructions here, and configuration details here.

  > **NOTE:**
  > Google Cloud recommends using Ops Agent instead of its legacy monitoring agent.

- Virtual CPU (vCPU)

  Virtual CPU is the measurement of CPU that is in use.

  Workload Optimization Manager calculates CPU based on the normalized CPU frequency and the number of vCPUs for a given VM. Normalized CPU frequency takes into account performance variations seen in different models of a given CPU platform. Because frequency is normalized, charts might show utilization values that are slightly higher than 100% (for example, 100.03%) when capacity is fully utilized.

- Net throughput

  Net throughput is the rate of message delivery over a port.

- Storage amount

  Storage amount is the measurement of storage capacity that is in use.

- Storage access (IOPS)

  Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

- I/O throughput

  I/O throughput is the measurement of an entity's throughput to the underlying storage.

  For both Storage Access (IOPS) and I/O Throughput, Workload Optimization Manager calculates capacity or uses capacity data published by Google Cloud, depending on the VM's machine type and disk.

  - Shared-core machine types share a physical core and are used for running small, non-resource intensive apps.

    For shared-core machine types with *standard disks* or *SSDs*, Workload Optimization Manager calculates capacity using internal benchmark data, which takes into consideration the observed maximum limit that can be achieved for IOPS and I/O throughput. It then uses the calculated capacity to analyze utilization more accurately.

    The use of internal benchmark data could result in differences in the capacity data shown in Workload Optimization Manager and Google Cloud.

  - For machine types that are *not* shared-core:

    - Workload Optimization Manager uses published capacity data and assumes that I/O block size is 16KB per I/O.
    - For machine types with persistent disks, Workload Optimization Manager assumes that the published capacity for the *SSD* disk type also applies to the *balanced* and *extreme* disk types. When a VM is attached to at least one of these disk types, capacity is assumed to be the per-VM limit for the SSD disk type. When a VM is attached only to the *standard* disk type, capacity is the per-VM limit for the standard disk type.

Monitored resources for volumes

Workload Optimization Manager monitors the following resources:

- Storage amount

  Storage Amount is the storage capacity (disk size) of a volume.

  Workload Optimization Manager discovers Storage Amount, but does not monitor utilization.

  For a Kubeturbo (container) deployment that includes volumes, Kubeturbo monitors Storage Amount utilization for the volumes. You can view utilization information in the Capacity and Usage chart.

- Storage access (IOPS)

  Storage access, also known as IOPS, is the measurement of IOPS capacity that is in use.

- I/O throughput

  I/O throughput is the measurement of I/O throughput capacity that is in use.

NOTE:
Workload Optimization Manager also monitors the attachment state of volumes and then generates delete actions for unattached volumes.

Supported Google Cloud regions

Workload Optimization Manager supports discovery and management of entities in the following Google Cloud regions:

- asia-east1
- asia-east2

- asia-northeast1
- asia-northeast2
- asia-northeast3
- asia-south1
- asia-south2
- asia-southeast1
- asia-southeast2
- australia-southeast1
- australia-southeast2
- europe-central2
- europe-north1
- europe-southwest1
- europe-west1
- europe-west2
- europe-west3
- europe-west4
- europe-west6
- europe-west8
- europe-west9
- me-west1
- northamerica-northeast1
- northamerica-northeast2
- southamerica-east1
- southamerica-west1
- us-central1
- us-east1
- us-east4
- us-east5
- us-south1
- us-west1
- us-west2
- us-west3
- us-west4

Google Cloud actions

Workload Optimization Manager monitors the state and performance of your workloads and then recommends actions to optimize these workloads at the lowest possible cost.

NOTE:
Use the Potential Savings and Necessary Investments charts to view pending actions and evaluate their impact on your cloud expenditure.

Historically-informed action generation for Google Cloud virtual machines

Workload Optimization Manager can recommend VM scale actions informed by historical utilization data on the same day a new Google Cloud target is added.

To take advantage of this feature, add at least one Google Cloud target. The service account for this target requires the `logging.logEntries.list` permission and all the permissions listed in this topic *(on page 418)*.

NOTE:
Currently, VM scale actions consider historical utilization of all resources *(on page 423)*, except I/O throughput.

Actions for virtual machines

Workload Optimization Manager supports the following actions:

- Scale

  Change the VM instance to use a different instance type or tier to optimize performance and costs.

- Discount-related actions

  If you have a high percentage of on-demand VMs, you can reduce your monthly costs by increasing Committed Use Discount (CUD) coverage. To increase coverage, you scale VMs to instance types that have existing capacity.

- Stop and start (also known as 'parking' actions)

  Stop a VM for a given period of time to reduce your cloud expenses, and then start it at a later time.

- Reconfigure

  Google Cloud provides a specific set of machine types for each zone in a region. If you create a policy that restricts a VM to certain machine types and the zone it is currently on does not support all of those machine types, Workload Optimization Manager will recommend a reconfigure action as a way to notify you of the non-compliant VM.

For details, see Actions for Google Cloud VMs *(on page 646)*.

Actions for volumes

Workload Optimization Manager supports the following actions:

- Scale

  Scale attached volumes to optimize performance and costs.

  For scale actions that require a tier change or actions that scale up IOPS for extreme persistent disks, Google Cloud requires a snapshot of the volume and the creation of a new volume based on the snapshot. When a new volume is created from a snapshot, it effectively has a new identity, thus resulting in the loss of historical utilization and cost information for that volume.

- Delete

  Delete unattached volumes as a cost-saving measure. Workload Optimization Manager generates an action immediately after discovering an unattached volume.

For additional information, see Cloud Volume Actions *(on page 728)*.

Microsoft Azure

Microsoft Azure is Microsoft's cloud computing platform. Workload Optimization Manager discovers your Azure resources through a service principal, and then optimizes these resources to assure performance at the lowest possible cost.

To connect to Azure, follow the steps and guidelines outlined in the following topics:

- Connecting to Azure *(on page 427)*
- Azure permissions *(on page 447)*

After connecting to Azure, Workload Optimization Manager monitors and optimizes the resources that it discovered. See the following topics for more information:

- Azure Monitored Resources *(on page 464)*
- Azure Actions *(on page 481)*

Connecting to Azure

To connect Workload Optimization Manager to your Azure environment, perform the tasks outlined in this topic.

These tasks apply to both global Azure and Azure Government. For Azure Government, be sure to review the requirements and caveats outlined in this topic *(on page 427)*.

> **NOTE:**
> Workload Optimization Manager requires Resource Manager to monitor your Azure resources. If you are currently managing Azure resources using the classic deployment model, migrate to Resource Manager before connecting Workload Optimization Manager to Azure.

1. Configure firewall and proxy access. *(on page 430)*

   If you run Workload Optimization Manager behind a firewall, configure firewall and proxy access.
2. Register resource providers. *(on page 431)*

   Register resource providers for the Azure subscriptions that Workload Optimization Manager will manage.
3. Set up a service principal for workload monitoring in Azure. *(on page 431)*

   Workload Optimization Manager discovers and monitors your Azure workloads through a service principal that you set up in Azure.
4. Add an Azure Service Principal target in Workload Optimization Manager *(on page 434)*.

   Authorize a secure connection through your service principal.

   To authorize the connection, add an Azure Service Principal target in the Workload Optimization Manager user interface.
5. Decide whether to use the Cost Details API or a cost export. *(on page 435)*

   Workload Optimization Manager can collect billing data through the Cost Details API or a cost export that you set up in Azure. Microsoft recommends the Cost Details API for billing data that is less than 2 GB in size, and a cost export for sizes that exceed 2 GB.
6. Set up a service principal for billing data monitoring in Azure. *(on page 439)*

   Workload Optimization Manager collects billing data through a service principal that you set up in Azure. You can use the service principal that you previously set up for workload monitoring, or create a new one specifically for billing data monitoring.
7. Add an Azure Billing target in Workload Optimization Manager. *(on page 446)*

   Authorize a secure connection to your billing data.

   To authorize the connection, add an Azure Billing target in the Workload Optimization Manager user interface.

Support for Azure Government

Azure Government provides dedicated regions for US government customers and their partners to architect secure cloud solutions and meet regulatory and compliance requirements.

Workload Optimization Manager can monitor workloads and billing data in Azure Government regions. Workloads include VMs, availability/scale sets, volumes, and SQL databases. Workload Optimization Manager recommends actions on these entities to address performance issues and optimize costs.

Connecting to Azure Government

To connect to Azure Government, perform the following tasks:

1. Configure firewall and proxy access. *(on page 430)*

   If you run Workload Optimization Manager behind a firewall, configure firewall and proxy access.

   Be sure to allow unrestricted access to the Azure Government URLs.
2. Register resource providers. *(on page 431)*

   Register resource providers for the Azure Government subscriptions that Workload Optimization Manager will manage.

3. Set up a service principal for workload monitoring in Azure. *(on page 431)*

   Workload Optimization Manager discovers and monitors your Azure Government workloads through a service principal that you set up in Azure.

4. Add an Azure Service Principal target in Workload Optimization Manager *(on page 434)*.

   When you add the target, be sure to turn on the **US Government** toggle to indicate that the target is an Azure Government account.

5. Decide whether to use the Cost Details API or a cost export. *(on page 435)*

   Workload Optimization Manager can collect billing data through the Cost Details API or a cost export that you set up in Azure. Microsoft recommends the Cost Details API for billing data that is less than 2 GB in size, and a cost export for sizes that exceed 2 GB.

   If you choose to set up a cost export for Azure Government EA accounts, the cost export must have a destination storage account that is under an Azure Government subscription.

6. Set up a service principal for billing data monitoring in Azure. *(on page 439)*

   When you set up a service principal, you assign the Enrollment Reader role to the service principal to enable access to billing data.

   To assign the Enrollment Reader role, you execute an API request. For Azure Government EA accounts, it is *not* possible to execute an API request in the Microsoft Learn portal. This is because the API request in the portal is hard-coded and executes against `https://management.azure.com`, which only applies to global Azure EA accounts. The API request for Azure Government EA accounts should execute against `https://management.usgovcloudapi.net`. For more information see this topic *(on page 443)*.

7. Add an Azure Billing target in Workload Optimization Manager. *(on page 446)*

   When you add the target, be sure to turn on the **US Government** toggle to indicate that the target is an Azure Government account.

   NOTE:
   Starting in version 3.10.2, the Azure Billing target is the required target for monitoring Azure Government billing data. Any Microsoft Enterprise Agreement target that you added for Azure Government should be removed from the user interface. The Microsoft Enterprise Agreement target has reached end of support.

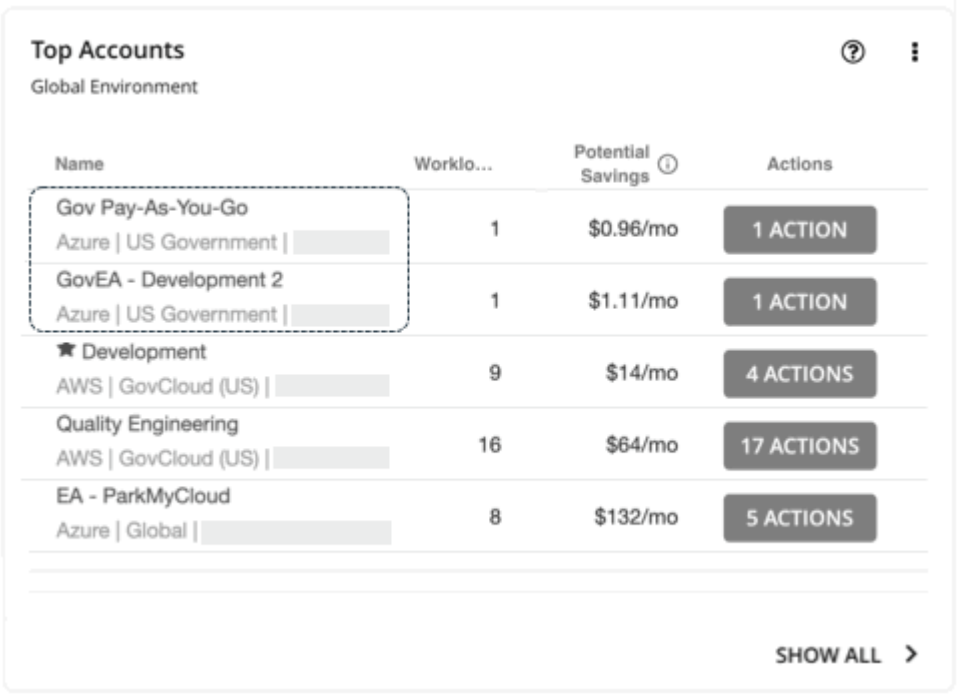
   For more information about end of support for the Microsoft Enterprise Agreement target, see this topic *(on page 463)*.

Information in charts

Use the following charts to view information about your government accounts and workloads.

- Top Accounts chart

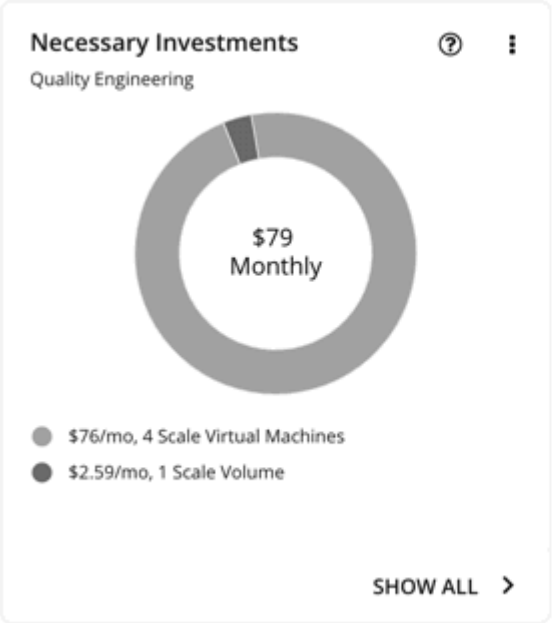
  Use the Top Accounts chart as a starting point.

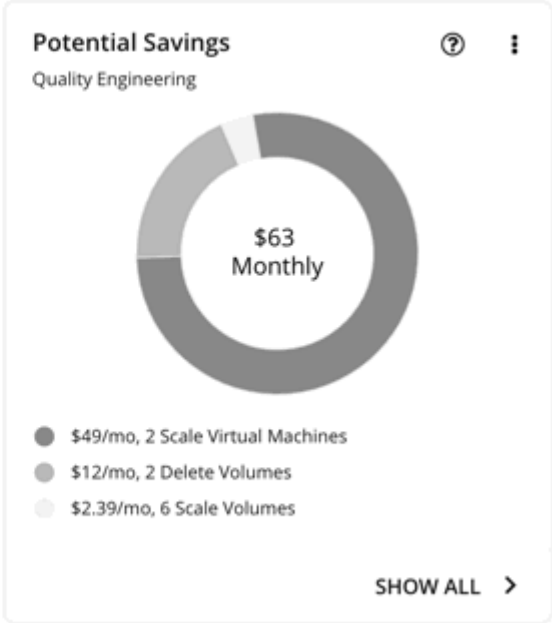


The chart shows the Azure Government subscriptions discovered via the service principal accounts that you have added as targets.

- Necessary Investments and Potential Savings charts

  Set the scope to a government account or subscription, and then see the Necessary Investments and Potential Savings charts to evaluate the costs you would incur or save if you execute all the pending actions for your government workloads.





Workload planning

You can run an Optimize Cloud plan to identify performance and efficiency opportunities for existing government workloads, or a Migrate to Cloud plan to migrate government VM groups to another cloud provider.

For on-prem clusters, you can run a Migrate to Cloud plan to see how you can safely migrate the VMs in these clusters to a government account/subscription and region.

Configuring firewall and proxy access for Azure

If you run Workload Optimization Manager behind a proxy or firewall, configure Workload Optimization Manager to allow unrestricted access to the following URLs to enable the full set of features for Azure targets.

Firewall and proxy access for the Azure service principal target

For global Azure:

- `ratecard.azure-api.net`
- `management.core.windows.net`
- `management.azure.com`
- `login.microsoftonline.com`
- `ratecard.blob.core.windows.net`
- `api.loganalytics.io`
- `graph.microsoft.com`
- `prices.azure.com`

For Azure Government:

- `ratecard.azure-api.net`
- `management.core.usgovcloudapi.net`
- `management.usgovcloudapi.net`
- `login.microsoftonline.us`
- `ratecard.blob.core.usgovcloudapi.net`
- `api.loganalytics.us`
- `graph.microsoft.com`
- `prices.azure.com`

Firewall and proxy access for the Azure Billing target

For global Azure:

- `api.loganalytics.io`
- `login.microsoftonline.com`
- `management.azure.com`
- For billing data *(on page 435)*:
  - Cost Details API

    `*.blob.core.windows.net`
  - Cost export
    - Less strict: `*.blob.core.windows.net`
    - Strict: `[STORAGE_ACCOUNT].blob.core.windows.net`

      Replace `[STORAGE_ACCOUNT]` with the name of the storage account containing the cost export.

For Azure Government:

- `{}api.loganalytics.io{}`
- `login.microsoftonline.us`
- `management.usgovcloudapi.net`

- For billing data *(on page 435)*:
  - Cost Details API

    `*.blob.core.windows.net`
  - Cost export
    - Less strict: `*.blob.core.windows.net`
    - Strict: `[STORAGE_ACCOUNT].blob.core.windows.net`

      Replace `[STORAGE_ACCOUNT]` with the name of the storage account containing the cost export.

Next step

Register Azure subscriptions with resource providers. For details, see this topic *(on page 431)*.

Registering Azure resource providers

A resource provider is a set of Azure REST operations that enable functionality for a specific Azure service. Registration configures your Azure subscriptions to work with resource providers.

Most resource providers are auto-registered and some need to be manually registered. In some cases, resource providers must be re-registered to support new locations that you may need to use.

Register the following resource providers for the Azure subscriptions that Workload Optimization Manager will manage:

- `Microsoft.Compute`
- `Microsoft.Capacity`
- `Microsoft.Sql`

To register resource providers, follow the registration steps in the Azure documentation.

> **NOTE:**
> If you do not register these resource providers, you might notice certain errors in the Workload Optimization Manager logs, as documented in this Azure page.

Next step

Set up a service principal in Azure. For details, see this topic *(on page 431)*.

Setting up an Azure service principal for workload monitoring

This topic describes the steps to set up a valid service principal that Workload Optimization Manager will use to connect to your Azure environment and discover subscriptions. Workload Optimization Manager discovers, monitors, and optimizes the workloads in these subscriptions.

Task overview

The following tasks apply to global Azure and Azure Government.

1. Register Workload Optimization Manager with Microsoft Entra ID (formerly known as Azure Active Directory).
2. Create a client secret key.
3. Configure API permissions.
4. Enable access to the subscriptions that Workload Optimization Manager will manage.

Registering Workload Optimization Manager with Microsoft Entra ID (formerly known as Azure Active Directory)

Registration automatically creates a service principal that serves as Workload Optimization Manager's identity in the Microsoft Entra tenant.

1. Sign in to the Azure portal with an administrator or co-administrator account. This level of access is only required for setting up the service principal, and not for regular Workload Optimization Manager operations.
   - For global Azure

     https://portal.azure.com
   - For Azure Government

     https://portal.azure.us
2. Browse to **Microsoft Entra ID > App registrations** and select **New registration**.
3. Configure the following settings:

Setting	Instructions
Name	Specify your preferred name, such as Workload Optimization Manager.
Supported account types	Select **Accounts in this organizational directory only**.

4. Select **Register**.

   Microsoft Entra ID creates the app registration for the service principal.
5. Record the Application (client) ID and Directory (tenant) ID. You will need this information later when you add an Azure target in the Workload Optimization Manager user interface.

Creating a client secret

1. Browse to **Microsoft Entra ID > App registrations** and then select the app registration for the service principal that you created for Workload Optimization Manager.
2. In the navigation menu, select **Certificates & secrets**.
3. Select **Client secrets** and then select **New client secret**.
4. Configure the following settings:

Setting	Instructions
Description	Specify a meaningful description.
Expires	Choose `Never`.

5. Select **Add**.
6. Record the client secret value. You will need this information later when you add an Azure service principal target in the Workload Optimization Manager user interface.

   IMPORTANT:
   The client secret value only displays once. It will no longer be available after you leave the page.

Configuring API permissions

1. Browse to **Microsoft Entra ID > App registrations** and then select the app registration for the service principal that you created for Workload Optimization Manager.
2. In the navigation menu, select **API permissions**.
3. Click **Add a permission** and select **Azure Service Management**.
4. Select **Delegated permissions** and **user_impersonation**, and then click **Add permissions**.

Enabling access to subscriptions using Azure roles

Assign roles to the service principal to enable access to subscriptions.

> To speed up the process of connecting Workload Optimization Manager to Azure, enable access at a high scope level (such as at the management group level). This automatically propagates permissions to lower levels of scope, such as subscriptions and resource groups.

You can assign the service principal either a custom role or built-in roles. If you have strict requirements for the service principal, assign a custom role.

- Custom role

  When you create a custom role, you specify the permissions that Workload Optimization Manager needs to monitor workloads in the subscription. Optionally, specify permissions to execute actions from Workload Optimization Manager.

- Built-in roles

  Built-in roles include default permissions that are sufficient for Workload Optimization Manager operations, but are more permissive than a custom role.

  - (Required) The **Reader** role is required to monitor workloads.
  - (Optional) The **Owner** or **Contributor** role is needed to execute actions from Workload Optimization Manager. The Contributor role is the least privileged role for action execution.

To assign a custom role, see the next section. To assign built-in roles, skip to Assigning Built-in Roles *(on page 434)*.

Assigning a custom role

This task assumes that when you assign a custom role to the service principal, you will upload a JSON file that specifies the required permissions. You can manually configure the permissions by following the wizard for creating custom roles, but complete instructions for running that wizard are not described in this task.

1. Create a JSON file that specifies the required permissions.
   - For permissions to monitor workloads, copy the JSON content found here *(on page 457)*.
   - For permissions to monitor workloads and execute actions, copy the JSON content found here *(on page 459)*.
2. Update the JSON file with the following information:
   - `<RoleName>` – Specify your preferred name for the custom role.
   - `<Subscription_ID>` – Specify the ID of the subscription that Workload Optimization Manager will manage.

     IMPORTANT:
     For Azure Government, remove the `Microsoft.Synapse/SKUs/read` permission from the JSON file. This permission is not required for Azure Government.

3. Browse to **Subscriptions**.
4. Select a subscription that Workload Optimization Manager will manage.
5. In the navigation bar, select **Access control (IAM)**.
6. Choose **Add > Add custom role**.
7. In the **Basics** tab:
   a. In the **Baseline permissions** field, select **Start from JSON**.
   b. In the **File** field, upload the JSON file that you created in a previous step.

   Azure notifies you if the JSON is valid.
8. Choose **Review + create**.

Assigning built-in roles

1. Browse to **Subscriptions**.
2. Select a subscription that Workload Optimization Manager will manage.
3. In the navigation bar, select **Access control (IAM)**.
4. Choose **Add > Add role assignment**.
5. In the Add role assignment page:
    a. Choose the **Role** tab.
    b. In the search bar, type `Reader` as your search keyword.
    c. Choose **Reader** from the list of built-in roles that display.

       To execute actions from Workload Optimization Manager, choose **Owner** or **Contributor**.

       Choose **Next**.
    d. In the **Members** tab, choose **+ Select members**.
    e. Search for the service principal that you set up for Workload Optimization Manager.
    f. Add the service principal to the list of members. Optionally, specify a description for this role assignment and then choose **Next**.
    g. In the **Review + Assign** tab, review your settings and then choose **Review + assign**.

Next step

Add an Azure Service Principal target in Workload Optimization Manager. For details, see this topic *(on page 434)*.

Adding an Azure Service Principal target

Add an Azure Service Principal target in the Workload Optimization Manager user interface to monitor and optimize workloads in your Azure environment. This target specifies the service principal credentials that Workload Optimization Manager will use to connect to Azure.

Before performing this task, be sure you have the Application (client) ID, Directory (tenant) ID, and client secret value for the service principal. If you do not have these credentials, follow the steps outlined in this topic *(on page 431)*.

Adding the target

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **Azure Service Principal**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
    - Display name

      Specify a name that uniquely identifies this connection.

      This name is for display purposes only and does not need to match any name in Azure.
    - Azure Government toggle

      When enabled, Workload Optimization Manager identifies the target as an Azure Government *(on page 427)* account.
    - Directory (tenant) ID

      Specify the Directory ID associated with your app registration.

      Format: 32 alphanumeric characters separated by hyphens
    - Application (client) ID

      Specify the application ID associated with your app registration.

Format: 32 alphanumeric characters separated by hyphens

- Client secret value

  Specify the client secret value associated with your app registration.

- Proxy configuration

  Specify proxy information only if you connect through a proxy.

  - Proxy hostname or IP address

    Specify the address of the proxy used for this target.

  - Proxy port

    Specify the port to use with the proxy used for this target. By default, the port is 8080.

  - Proxy username

    Specify the username to use with the proxy.

  - Proxy password

    Specify the password to use with the proxy.

  - Connect using HTTPS

    Select this option if Workload Optimization Manager will connect to the proxy through HTTPS.

After adding an Azure Service Principal target

Workload Optimization Manager starts to discover the resources in your Azure environment.

In the Target Configuration page, the service principal displays as the primary target that you can edit, while the subscriptions display as derived targets that cannot be edited but can be validated like any other target.

Next step

Decide whether to use the Cost Details API or a cost export to enable access to your Azure billing data. For details, see this topic *(on page 435)*.

Using the Cost Details API or a cost export

The Azure Billing target can collect billing data for your Microsoft Customer Agreement (MCA) or Enterprise Agreement (EA) accounts through the Azure Cost Details API or a cost export that you set up in Azure.

For either global Azure or Azure Government, your choice depends mainly on the size of your billing data.

- **Cost Details API**: Microsoft recommends the Cost Details API for billing data that is less than 2 GB in size. Larger sizes will likely result in timeouts.
- **Cost Export**: Microsoft recommends a cost export for billing data that is at least 2 GB in size. The data export is in CSV format, and contains all the cost and usage data that Azure Cost Management collects.

NOTE:

Support for the Cost Details API started in version 3.6.4. New deployments of the product will use this API by default.

For Azure Billing targets added in earlier releases of the product, any associated cost export will automatically carry over to version 3.6.4 or later. You can switch these targets to the Cost Details API any time.

Using the Cost Details API

If you choose the Cost Details API:

- There is nothing that you need to set up in Azure.
- If you run Workload Optimization Manager behind a firewall, be sure to allow unrestricted access to `*.blob.core.windows.net`.
- When you add an Azure Billing target in the Workload Optimization Manager user interface in a later task, be sure to *turn off* the **Use a cost export** option.

Next Step: Set up an Azure service principal for billing data monitoring. For details, see this topic *(on page 439)*.

Using a cost export

If you choose a cost export, set up a daily cost export of month-to-date costs in the Azure portal.

The steps for setting up a cost export are slightly different for EA and MCA accounts. The following sections describe the steps.

Setting up a cost export for EA accounts

1. Sign in to the Azure portal.
   - For global Azure

     https://portal.azure.com
   - For Azure Government

     https://portal.azure.us
2. Browse to **Cost Management + Billing**.
3. (Global Azure only) Set the billing scope.
   a. In the navigation bar, choose **Billing scopes**.
   b. Select a billing scope with a billing scope type of 'Billing account'.

      NOTE:
      Workload Optimization Manager does not support the subscription, management group, or resource group billing scope type.

4. In the navigation bar, find **Settings** and select **Exports**.
5. Select **Create**.
6. Configure the following settings:
   - Export details

Setting	Instructions
Name	Specify your preferred name.
Metric	Choose **Cost and usage details (actual)**.
Export type	Choose **Daily export of month-to-date costs**.
Start date (UTC time)	Specify the current date.

   - File partitioning

Setting	Instructions
File partitioning	Turn on file partitioning.

- Storage

Setting	Instructions
Storage	Choose **Use existing** or **Create a new one**.
Subscription	Select a subscription for the cost export. Be sure that you have added the subscription to Workload Optimization Manager as an Azure service principal target. If you have not completed this task, first set up a service principal in Azure *(on page 431)* and then add an add an Azure service principal target *(on page 434)*.
Storage account	This field provides a list of storage accounts for the selected subscription. Select a storage account for the cost export. – Be sure to provide Reader and Data Access permissions to the storage account. – If the storage account is behind a firewall, add the cluster IP range for Workload Optimization Manager to the firewall's allowlist. For more information, see the Azure documentation. – For Azure Government EA accounts, the storage account must be under an Azure Government subscription.
Container	Specify the container name for the cost export. Be sure to provide Storage Blob Data Reader permissions to the container.
Directory	Specify the directory path for the cost export.

7. Select **Create**.
8. Record the name of the cost export that you just created. You will need this information later when you add an Azure Billing target in the Workload Optimization Manager user interface.

   NOTE:
   When you add the Azure Billing target, be sure to *turn on* the **Use a cost export** option.

9. If you run Workload Optimization Manager behind a firewall, allow unrestricted access to the following URLs:
   - Less strict:
     `*.blob.core.windows.net`
   - Strict:
     `{Storage_Account}``.blob.core.windows.net`
     Replace *`{Storage_Account}`* with the name of the storage account that contains the cost export.

Next Step

Set up an Azure service principal for billing data monitoring. For details, see this topic *(on page 439)*.

Setting up a cost export for MCA accounts

1. Sign in to the Azure portal.

   https://portal.azure.com
2. Browse to **Cost Management + Billing**.
3. In the navigation bar, find **Settings** and select **Exports**.
4. Set the billing scope.
   a. Select **Scope: `{Account Name}`**.
   b. Select a billing scope with a billing scope type of 'billing profile'.

   If you previously set up a cost export with a scope type of billing account, delete the cost export and then create a new one with a scope type of billing profile. Create this cost export for each active billing profile that falls under your MCA billing account. All cost exports you create for your active billing profiles are required to have the same name.

   NOTE:
   Workload Optimization Manager does not support the subscription, management group, or resource group billing scope type.

5. Select **Create**.
6. Configure the following settings:
   - Add export

Setting	Instructions
Type of data	Choose **Cost and usage details (actual)**.
Export name	Specify your preferred name.
Dataset version	Choose the latest version.
Frequency	Choose **Daily export of month-to-date costs**.
Export description	(Optional) Specify a description.

   - Destination

Setting	Instructions
Storage type	Accept the default selection **Azure blob storage**.
Destination and storage	Choose **Use existing** or **Create a new one**.
Subscription	Select a subscription for the cost export. Be sure that you have added the subscription to Workload Optimization Manager as an Azure service principal target. If you have not completed this task, first set up a service principal in Azure *(on page 431)* and then add an add an Azure service principal target *(on page 434)*.

Setting	Instructions
Storage account	This field provides a list of storage accounts for the selected subscription. Select a storage account for the cost export. – Be sure to provide Reader and Data Access permissions to the storage account. – If the storage account is behind a firewall, add the cluster IP range for Workload Optimization Manager to the firewall's allowlist. For more information, see the Azure documentation.
Container	Specify the container name for the cost export. Be sure to provide Storage Blob Data Reader permissions to the container.
Directory	Specify the directory path for the cost export.
File partitioning	Turn on file partitioning.
Overwrite data	Turn off data overwriting.

- Review + create

  Review your settings and make changes as necessary.

7. Select **Create**.
8. Record the name of the cost export that you just created. You will need this information later when you add an Azure Billing target in the Workload Optimization Manager user interface.

   NOTE:
   When you add the Azure Billing target, be sure to *turn on* the **Use a cost export** option.

9. If you run Workload Optimization Manager behind a firewall, allow unrestricted access to the following URLs:
   - Less strict:

     `*.blob.core.windows.net`
   - Strict:

     `{Storage_Account}.blob.core.windows.net`

     Replace `{Storage_Account}` with the name of the storage account that contains the cost export.

Next Step

Set up an Azure service principal for billing data monitoring. For details, see this topic *(on page 439)*.

Setting Up an Azure service principal for billing data monitoring

This topic describes the steps to set up a valid service principal that Workload Optimization Manager will use to connect to your Microsoft Customer Agreement (MCA) or Enterprise Agreement (EA) accounts and collect billing data. Workload Optimization Manager uses this data to visualize historical expenses and discover reservations.

NOTE:
For information on the current level of support for MCA and EA accounts, see this topic *(on page 462)*.

Task overview

To set up a service principal for billing data monitoring, perform the following tasks in the Azure portal:

- Scenario 1: Use the existing service principal *(on page 431)* that you set up for workload monitoring.

  Skip the first three tasks in this topic and perform only the following tasks:

  1. Enable access to billing data.
  2. (Azure Government only) Enable access to reservations.

- Scenario 2: Create a new service principal specifically for billing data monitoring.

  Perform all the tasks in this topic.

  1. Register Workload Optimization Manager with Microsoft Entra ID (formerly known as Azure Active Directory).
  2. Create a client secret key.
  3. Configure API permissions.
  4. Enable access to billing data.
  5. (Azure Government only) Enable access to reservations.

Registering Workload Optimization Manager with Microsoft Entra ID (formerly known as Azure Active Directory)

NOTE:
Skip this task if you want to use the existing service principal *(on page 431)* that you set up for workload monitoring.

Registration automatically creates a service principal that serves as Workload Optimization Manager's identity in the Microsoft Entra tenant.

1. Sign in to the Azure portal with an administrator or co-administrator account. This level of access is only required for setting up the service principal, and not for regular Workload Optimization Manager operations.
   - For global Azure

     https://portal.azure.com
   - For Azure Government

     https://portal.azure.us
2. Browse to **Microsoft Entra ID > App registrations** and select **New registration**.
3. Configure the following settings:

Setting	Instructions
Name	Specify your preferred name, such as Workload Optimization Manager.
Supported account types	Select **Accounts in this organizational directory only**.

4. Select **Register**.

   Microsoft Entra ID creates the app registration for the service principal.
5. Record the Application (client) ID and Directory (tenant) ID. You will need this information later when you add an Azure target in the Workload Optimization Manager user interface.

Creating a client secret

NOTE:
Skip this task if you want to use the existing service principal *(on page 431)* that you set up for workload monitoring.

1. Browse to **Microsoft Entra ID > App registrations** and then select the app registration for the service principal that you created for Workload Optimization Manager.
2. In the navigation menu, select **Certificates & secrets**.
3. Select **Client secrets** and then select **New client secret**.
4. Configure the following settings:

Setting	Instructions
Description	Specify a meaningful description.
Expires	Choose `Never`.

5. Select **Add**.
6. Record the client secret value. You will need this information later when you add an Azure service principal target in the Workload Optimization Manager user interface.

IMPORTANT:
The client secret value only displays once. It will no longer be available after you leave the page.

Configuring API permissions

NOTE:
Skip this task if you want to use the existing service principal *(on page 431)* that you set up for workload monitoring.

1. Browse to **Microsoft Entra ID > App registrations** and then select the app registration for the service principal that you created for Workload Optimization Manager.
2. In the navigation menu, select **API permissions**.
3. Click **Add a permission** and select **Azure Service Management**.
4. Select **Delegated permissions** and **user_impersonation**, and then click **Add permissions**.

Enabling access to billing data

To enable access to billing data, assign the required role to the service principal that you set up for Workload Optimization Manager. Note that only users with the Global Administrator role and with elevated access can assign the role to the service principal. Elevated access grants permissions to assign roles in Azure subscriptions and management groups associated with Microsoft Entra ID (formerly known as Azure Active Directory). For instructions on elevating access for users, see the Azure documentation.

The following roles are required:

Account	Role
EA account (global Azure or Azure Government)	Enrollment Reader: This role can only be assigned using Azure API methods. It cannot be assigned from the user interface of the Azure portal.
MCA account (global Azure only)	Billing Account Reader: This role can be assigned from the user interface of the Azure portal.

NOTE:
For information about the current level of support for EA and MCA accounts, see this topic *(on page 462)*.

Assigning the Enrollment Reader role to Global Azure EA accounts

If you have global Azure EA accounts, assign the Enrollment Reader role to enable access to billing data. To assign the role, execute an API request in the Microsoft Learn portal.

1. Record the values for the following parameters. These values are required when assigning the role:

Parameter	Value
`billingAccountName`	Billing account ID: In the Azure portal, navigate to **Cost management + Billing**, open your EA account, and then select **Overview**. The screen that displays includes a field for the Billing Account ID. "
`billingRoleAssignmentName`	A unique GUID: To generate a GUID, use the New-Guid PowerShell command in Microsoft Learn or the Online GUID / UUID Generator website.
`api-version`	`2019-10-01-preview`: This is a static value.
`properties.principalId`	Object ID specified in Microsoft Entra ID: For help locating the Object ID, see the Azure documentation.
`properties.principalTenantId`	Tenant ID specified in Microsoft Entra ID: For help locating the Tenant ID, see the Azure documentation.

2. Open the Microsoft Learn portal. Sign in when prompted.
3. In the following fields, specify the values that you recorded in the previous step.

Field	Value
`billingAccountName`	*`{your_billing_account_ID}`*
`billingRoleAssignmentName`	*`{your_unique_GUID}`*
`api-version`	`2019-10-01-preview`

4. In the **Body** field, paste the following content.

```
{
  "properties": {
    "principalId": "{your_object_ID}",
    "principalTenantId": "{your_tenant_ID}",
    "roleDefinitionId":
 "/providers/Microsoft.Billing/billingAccounts/{your_billing_account_ID}/billingRoleAssignments/24f
8edb6-1668-4659-b5e2-40bb5f3a7d7e"
  }
}
```

NOTE:
`24f8edb6-1668-4659-b5e2-40bb5f3a7d7e` is the ID that Microsoft assigned to the Enrollment Reader role.

5. Replace the following variables with the values that you recorded in the previous step.
   - *{your_object_ID}*
   - *{your_tenant_ID}*
   - *{your_billing_account_ID}*
6. Review the request preview and then click **Run**.

   The Enrollment Reader role is now assigned.

Assigning the Enrollment Reader role to Azure Government EA accounts

For Azure Government EA accounts, it is not possible to execute an API request in the Microsoft Learn portal. This is because the API request in the portal is hard-coded and executes against `https://management.azure.com`, which only applies to global Azure EA accounts. The API request for Azure Government EA accounts should execute against `https://management.usgovcloudapi.net`.

This section describes how to assign the Enrollment Reader role using a script that you run in a PowerShell session in Azure Cloud Shell.

IMPORTANT:
To perform this task, you must have permissions to use Azure Cloud Shell and assign roles to billing accounts in your Azure Government environment.

If you are new to Azure Cloud Shell, follow the setup guidelines in the Azure documentation.

Before proceeding, be sure you have the Application (client) ID for the service principal that you set up for Workload Optimization Manager in a previous task. You will specify this ID when you run the script.

1. In your local machine, create a file named `Turbonomic-GovCloudBillingReader.ps1` and then copy the following script to the file.

```
### Get your Azure Authentication token and create your Azure Authentication Header
$token = (Get-AzAccessToken).Token
$headers = @{'Authorization' = "Bearer $token" }
```

```
### Get the Service Principal, Tenant, Billing Account, and Role properties
$servicePrincipalApplicationID = Read-Host "Enter your Service Principal's Application (client) ID"
$servicePrincipalApplication = (Get-AzADServicePrincipal -ApplicationID
 $servicePrincipalApplicationID)
$servicePrincipalObjectID = $servicePrincipalApplication.id
$tenant = Get-AzTenant
write-host "Your tenant ID is: $($tenant.Name) (ID: $($tenant.id))"
$billingAccount = (Get-AzBillingAccount).Name
write-host "Your Billing Account ID is: $billingAccount"
$roleDefinitionID =
 "/providers/Microsoft.Billing/billingAccounts/$billingAccount/billingRoleDefinitions/24f8edb6-1668
-4659-b5e2-40bb5f3a7d7e"
$billingRoleAssignmentName = (New-GUID).GUID
write-host "Your Application permission's unique GUID is $billingRoleAssignmentName"
$apiVersion = "api-version=2019-10-01-preview"
```

```
### Create the JSON payload used in the permissions API call
write-host "Creating the permission JSON body"
$json = @"
 {
 "properties":{
 "principalId": "$servicePrincipalObjectID",
     "principalTenantId": "$tenant",
     "roleDefinitionId": "$roleDefinitionID"
     }
 }
"@
```

```
### Create the URL path for the permissions API call
$invokeBaseURL = "https://management.usgovcloudapi.net"
$invokeBillingURL =
 "/providers/Microsoft.Billing/billingAccounts/$billingAccount/billingRoleAssignments/$($billingRol
eAssignmentName)?$apiVersion"
$invokeURL = $invokeBaseURL+$invokeBillingURL

### Print the specifics that will be used in the API call for verification
write-host "Attempting to set permissions for the following Application"
write-host "`tAzure Tenant: $($tenant.Name) (ID: $($tenant.Id))"
write-host "`tService Principal Application Name: $($servicePrincipalApplication.DisplayName)"
write-host "`tService Principal Application (client) ID: $($servicePrincipalApplicationID)"
write-host "`tService Principal Object ID: $($servicePrincipalObjectID)"
write-host "`nThe permission JSON is:"
write-host $json
write-host "`n`nContinuing will attempt to set the Billing Reader role permissions for this Service
 Principal in Azure tenant $($tenant.Name) (ID: $($tenant.Id))"
Pause

### Attempt to set the Billing Reader role permissions for the Service Principal
Invoke-RestMethod -Uri $invokeURL -headers $headers -ContentType "application/json" -body $json
 -Method PUT
```

2. Sign in to the Azure Government portal with a user account that has permissions to use Azure Cloud Shell and assign roles to billing accounts.

   https://portal.azure.us

3. In the navigation bar on top of the page, select the Cloud Shell icon.
4. In the Cloud Shell window that opens, go to the navigation bar and then perform the following steps:
   a. Select the dropdown list on the left and then choose **PowerShell** to start a PowerShell session.
   b. Select the upload/download files icon, select **Upload**, and then upload the file that you created.
5. In the command line, run the following command to execute the script.

   ```
   ./Turbonomic-GovCloudBillingReader.ps1
   ```

6. Specify the Application (client) ID for the service principal that you set up for Workload Optimization Manager.

   The Enrollment Reader role is now assigned.

Assigning the Billing Account Reader role to MCA accounts

If you have MCA accounts, assign the Billing Account Reader role to enable access to billing data.

1. Browse to **Cost Management + Billing > Access control (IAM)**.
2. Select **Add**.
3. Select **Billing account reader** and then select the app registration for the service principal that you created for Workload Optimization Manager.

Enabling discovery of Azure reservations (Azure Government only)

If you have reservations for your Azure Government workloads, enable access to these reservations by assigning the built-in **Reservations Reader** role to the service principal that you set up for Workload Optimization Manager.

1. Sign in to the Azure Government portal with a user account that has permissions to assign roles to reservations.

   https://portal.azure.us

   Be sure that you are working in the correct Azure directory where you will assign the Reservations Reader role.

2. (Recommended) To speed up the process of connecting Workload Optimization Manager to Azure, create a Microsoft Entra group that includes all the service principals that will be assigned the Reservations Reader role.

Any service principal that you add to the group in the future will automatically be assigned the Reservations Reader role.

3. Browse to the Reservations page.

   https://portal.azure.us/#view/Microsoft_Azure_Reservations/ReservationsBrowseBlade/productType/Reservations

4. In the Reservations page, choose **Role Assignment**.
5. In the Access Control page, choose **Add role assignment**.
6. In the Add role assignment page:
   a. Choose the **Role** tab.
   b. In the search bar, type `Reservations Reader` as your search keyword.
   c. Choose **Reservations Reader** from the list of built-in roles that display and then choose **Next**.
   d. In the **Members** tab, choose **+ Select members**.
   e. Search for the service principal or Microsoft Entra group (if you created one in a previous step) that you set up for Workload Optimization Manager.
   f. Add the service principal or group to the list of members. Optionally, specify a description for this role assignment and then choose **Next**.
   g. In the **Review + Assign** tab, review your settings and then choose **Review + assign**.

Enabling discovery of Azure savings plans

If you have savings plans for your Azure workloads, enable access to these savings plans by assigning the built-in Savings Plan Reader role to the service principal that you set up for Workload Optimization Manager.

1. Sign in to the Azure portal with a user account that has permissions to assign roles to savings plans.

   https://portal.azure.com

   Be sure that you are working in the correct Azure directory where you will assign the Savings Plan Reader role.

2. (Recommended) To speed up the process of connecting Workload Optimization Manager to Azure, create a Microsoft Entra group that includes all the service principals that will be assigned the Savings Plan Reader role.

   Any service principal that you add to the group in the future will automatically be assigned the Savings Plan Reader role.

3. Browse to the Savings plans page.

   https://portal.azure.com/#view/Microsoft_Azure_Reservations/ReservationsBrowseBlade/productType/SavingsPlan

4. In the Savings plans page, choose **Role Assignment**.
5. In the Access Control page, choose **Add role assignment**.
6. In the Add role assignment page:
   a. Choose the **Role** tab.
   b. In the search bar, type `savings plan reader` as your search keyword.
   c. Choose **Savings Plan Reader** from the list of built-in roles that display and then choose **Next**.
   d. In the **Members** tab, choose **+ Select members**.
   e. Search for the service principal or Microsoft Entra group (if you created one in a previous step) that you set up for Workload Optimization Manager.
   f. Add the service principal or group to the list of members. Optionally, specify a description for this role assignment and then choose **Next**.
   g. In the **Review + Assign** tab, review your settings and then choose **Review + assign**.

Next step

Add an Azure Billing target in Workload Optimization Manager. For details, see this topic *(on page 446)*.

Adding an Azure Billing Target

The Azure Billing target grants Workload Optimization Manager access to billing data from the Azure Cost Details API or a cost export that you set up in Azure. Workload Optimization Manager uses this data to visualize historical cloud expenses and discover reservations. The Azure Billing target supports only Enterprise Agreements or Microsoft Customer Agreements bought directly from azure.com.

Before performing this task, be sure you have the required credentials for the service principal. Credentials include the billing account ID, application (client) ID, directory (tenant) ID, and client secret value. You also need to set up the appropriate role to enable access to billing data. If you do not have these credentials or if you have not set up a role, follow the steps outlined in this topic *(on page 439)*.

Adding an Azure Billing Target

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **Azure Billing**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
    - Display name

      Specify a name that uniquely identifies this connection.

      This name is for display purposes only and does not need to match any name in Azure.
    - Azure Government toggle

      When enabled, Workload Optimization Manager identifies the target as an Azure Government *(on page 427)* account.
    - Directory (tenant) ID

      Specify the Directory ID associated with your app registration.

      Format: 32 alphanumeric characters separated by hyphens
    - Application (client) ID

      Specify the application ID associated with your app registration.

      Format: 32 alphanumeric characters separated by hyphens
    - Client secret value

      Specify the client secret value associated with your app registration.
    - Billing Account ID

      Specify the billing account ID.

      You can find the billing ID in the **Cost Management + Billing** section of the Azure portal. The ID is different for MCA and EA accounts.
        - MCA Billing account ID

          In the Azure portal, navigate to **Cost management + Billing**, open your MCA billing account, and then select **Properties**. The screen that displays includes a field for the Billing Account ID.
        - EA Billing account ID (Enrollment Number)

          In the Azure portal, navigate to **Cost management + Billing**, open your EA account, and then select **Overview**. The screen that displays includes a field for the Billing Account ID.
    - Type of cost details dataset

      The Azure Billing target can collect billing data for your Microsoft Customer Agreement (MCA) or Enterprise Agreement (EA) accounts through the Azure Cost Details API or a cost export that you set up in Azure.

      For more information, see this topic *(on page 435)*.
        - Cost details API
        - Cost export

Specify the cost export name.

To find the cost export name in the Azure portal, navigate to **Cost management + Billing** and then select **Exports**. A list of the cost exports you have created displays. For more information about setting up cost exports, see this topic *(on page 436)*.

- Proxy configuration

  Specify proxy information only if you connect through a proxy.

  - Proxy hostname or IP address

    Specify the address of the proxy used for this target.

  - Proxy port

    Specify the port to use with the proxy used for this target. By default, the port is 8080.

  - Proxy username

    Specify the username to use with the proxy.

  - Proxy password

    Specify the password to use with the proxy.

  - Connect using HTTPS

    Select this option if Workload Optimization Manager will connect to the proxy through HTTPS.

    NOTE:
    Currently, HTTPS connection to global Azure or Azure Government is not supported. Be sure *not* to specify a secure proxy port, such as 9080 or 9081, in the **Proxy port** field.

NOTE:
After successfully adding a target, it can take up to 48 hours for any Azure reservation and billing data to display in the Workload Optimization Manager user interface.

After adding an Azure Billing Target

You have completed the required tasks for connecting to Azure. Workload Optimization Manager can now monitor your Azure workloads and billing data, and recommend actions to optimize these workloads at the lowest possible cost. Review the following topics for more information:

- Azure Monitored Resources *(on page 464)*
- Azure Actions *(on page 481)*

Reference: Azure permissions

The service principal that you set up in Azure specifies the permissions that Workload Optimization Manager needs to discover and monitor your Azure workloads. Permissions to execute actions from Workload Optimization Manager are optional.

Minimum permissions - workload monitoring

The following minimum permissions are required to monitor Azure workloads.

Workload Optimization Manager Functionality	Required Permissions
Role validation	▪ `Microsoft.Authorization/roleAssignments/read` Validates the role assigned to the Service Principal by checking if it has the minimum required permissions

Workload Optimization Manager Functionality	Required Permissions
	▪ `Microsoft.Authorization/roleDefinitions/read` Queries the permissions list from the assigned custom role
Discovery of subscriptions	▪ `Microsoft.Resources/subscriptions/read` Gets a list of accessible subscriptions for a tenant
Discovery of resource groups, locations, and SKUs	▪ `Microsoft.Resources/subscriptions/locations/read` List all locations available for the subscriptions ▪ `Microsoft.Resources/subscriptions/resourceGroups/read` Discovers all resource groups for the subscriptions ▪ `Microsoft.Compute/skus/read` Gets a list of `Microsoft.Compute` SKUs available for your subscription
Discovery of storage accounts	▪ `Microsoft.Storage/storageAccounts/read` Gets a list of storage accounts or gets the properties of a specific storage account
Discovery of metrics for various entities	▪ `Microsoft.DatabaseWatcher/watchers/read` Reads available database watchers in an Azure subscription ▪ `Microsoft.DatabaseWatcher/watchers/targets/read` Reads Azure SQL entities that are associated with a given database watcher ▪ `Microsoft.Insights/Metrics/Read` Reads metrics for various resources from Azure Monitor ▪ `Microsoft.OperationalInsights/workspaces/read` Queries a list of Log Analytics workspaces. Certain metrics (such as VM memory) are fetched from these workspaces. ▪ `Microsoft.OperationalInsights/workspaces/query/read` Allows queries to data (such as metrics) stored in Log Analytics workspaces ▪ `Microsoft.OperationalInsights/workspaces/query/InsightsMetrics/read` Queries the `InsightsMetrics` Log Analytics table for VM memory metrics. Azure Monitor Agent must be configured to send memory metrics to the table. ▪ `Microsoft.OperationalInsights/workspaces/query/Perf/read` (Only required if the `Perf` Log Analytics table is configured for VM memory metrics, instead of the `InsightsMetrics` Log Analytics table) Queries the `Perf` Log Analytics table for VM memory metrics
Discovery of VMs	▪ `Microsoft.Compute/virtualMachines/instanceView/read` Gets the detailed runtime status of a VM and its resources ▪ `Microsoft.Compute/virtualMachines/read` Gets the properties of a VM ▪ `Microsoft.Compute/virtualMachines/extensions/read`

Workload Optimization Manager Functionality	Required Permissions
	(Only required if using storage account for VM memory configuration) Gets the properties of a VM extension, to detect diagnostics extension before fetching memory metrics from storage account
Discovery of VM scale sets and availability sets	▪ `Microsoft.Compute/virtualMachineScaleSets/read` Gets the properties of a VM scale set ▪ `Microsoft.Compute/virtualMachineScaleSets/networkInterfaces/read` Lists all the network interfaces of a VM scale set and gets their properties ▪ `Microsoft.Compute/virtualMachineScaleSets/virtualMachines/instanceView/read` Retrieves the instance view of a VM in a scale set ▪ `Microsoft.Compute/virtualMachineScaleSets/virtualMachines/read` Retrieves the properties of a VM in a scale set ▪ `Microsoft.Compute/virtualMachineScaleSets/virtualMachines/extensions/read` (Only required if VM memory is configured onto a storage account table via a diagnostics agent extension) Gets the properties of an extension for a VM in a scale set ▪ `Microsoft.Compute/availabilitySets/read` Lists all availability sets and gets their properties ▪ `Microsoft.Compute/availabilitySets/vmSizes/read` Lists available sizes for creating or updating a VM in an availability set ▪ `Microsoft.Insights/AutoscaleSettings/Read` Lists the `autoscale` settings associated with a VM scale set
Discovery of reservations	▪ `Microsoft.Capacity/reservationorders/reservations/read` Monitors reservations data and reads all reservations ▪ `Microsoft.Capacity/catalogs/read` Reads the catalog of reservations
Discovery of volumes	▪ `Microsoft.Compute/disks/read` Gets the properties of volumes ▪ `Microsoft.Storage/storageAccounts/listkeys/action` (Only required if there are unmanaged disks that are attached to VMs) Discovers or queries unmanaged attached disks (volumes) in the storage account. Unmanaged disks that are not attached to VMs are not discovered.
Discovery of SQL databases (vCore/DTU) and metrics	▪ `Microsoft.Sql/servers/read` Lists all SQL servers in this subscription and gets their details ▪ `Microsoft.Sql/servers/databases/read` Lists and gets details about all SQL databases for all SQL servers in the subscription

Workload Optimization Manager Functionality	Required Permissions
	▪ `Microsoft.Sql/servers/databases/metrics/read` Queries metrics for SQL databases
Discovery of SQL Managed Instances	▪ `Microsoft.Sql/managedInstances/read` Gets a list of SQL Managed Instances ▪ `Microsoft.Sql/managedInstances/databases/read` Gets a list of databases running on a SQL Managed Instance ▪ `Microsoft.Sql/managedInstances/metrics/read` Gets metrics for SQL Managed Instances
Discovery of resources used in Azure Synapse Analytics	▪ `Microsoft.Synapse/SKUs/read` Reads SKU details for a Synapse Analytics Service resource, such as SQL pools **NOTE:** This permission is not required for Azure Government. ▪ `Microsoft.Synapse/workspaces/read` Reads details about Synapse workspaces ▪ `Microsoft.Synapse/workspaces/keys/read` Gets details of Synapse workspace key definitions ▪ `Microsoft.Synapse/workspaces/sqlDatabases/read` Reads a list of Synapse SQL Analytics databases ▪ `Microsoft.Synapse/workspaces/sqlPools/read` Reads a list of Synapse SQL Analytics pools ▪ `Microsoft.Synapse/workspaces/sqlPools/dataWarehouseUserActivities/read` Reads user activities on Synapse SQL Analytics pools ▪ `Microsoft.Synapse/workspaces/sqlPools/extensions/read` Gets extensions for Synapse SQL Analytics pools ▪ `Microsoft.Synapse/workspaces/sqlPools/operationStatuses/read` Reads the results of asynchronous operations on Synapse SQL Analytics pools ▪ `Microsoft.Synapse/workspaces/sqlPools/usages/read` Reads usage metrics for Synapse SQL Analytics pools ▪ `Microsoft.Synapse/workspaces/sqlUsages/read` Gets usage limits available for Synapse SQL Analytics pools
Discovery of App Services (plans/app instances) and metrics	▪ `Microsoft.Relay/namespaces/HybridConnections/read` Lists all Service Bus Hybrid Connections used by web apps ▪ `Microsoft.Web/geoRegions/Read` Gets a list of available geographical regions for App Services ▪ `Microsoft.Web/serverfarms/Read`

Workload Optimization Manager Functionality	Required Permissions
	Lists and gets the properties of App Service plans ▪ `Microsoft.Web/serverfarms/sites/read` Gets a list of web apps that are part of App Service plans ▪ `Microsoft.Web/serverfarms/skus/read` Gets SKUs for App Service plans ▪ `Microsoft.Web/sites/read` Gets the properties of web apps that are part of App Service plans ▪ `Microsoft.Web/sites/slots/Read` Gets the properties of a web app deployment slot ▪ `Microsoft.Web/serverfarms/metrics/read` Queries metrics for App Services plans ▪ `Microsoft.Web/serverfarms/usages/read` Gets usage information for App Service plans ▪ `Microsoft.Web/sites/metrics/read` Gets metrics for web apps that are part of App Service plans ▪ `Microsoft.Web/sites/usages/read` Gets usage information for web apps that are part of App Service plans
Discovery of Cosmos DB resources	▪ `Microsoft.DocumentDB/databaseAccounts/read` Gets the properties of a database account ▪ `Microsoft.DocumentDB/databaseAccounts/databases/metrics/read` Queries metrics for a database account ▪ `Microsoft.DocumentDB/databaseAccounts/databases/collections/metrics/read` Queries metrics for a container ▪ `Microsoft.DocumentDB/databaseAccounts/metrics/read` Queries metrics for a database ▪ `Microsoft.DocumentDB/databaseAccounts/cassandraKeyspaces/read` Gets the properties of an Apache Cassandra keyspace ▪ `Microsoft.DocumentDB/databaseAccounts/cassandraKeyspaces/throughputSettings/read` Gets the throughput of an Apache Cassandra keyspace ▪ `Microsoft.DocumentDB/databaseAccounts/cassandraKeyspaces/tables/read` Gets the properties of an Apache Cassandra table ▪ `Microsoft.DocumentDB/databaseAccounts/cassandraKeyspaces/tables/throughputSettings/read` Gets the throughput of an Apache Cassandra table ▪ `Microsoft.DocumentDB/databaseAccounts/mongodbDatabases/read` Gets the properties of a MongoDB database

Workload Optimization Manager Functionality	Required Permissions
	▪ `Microsoft.DocumentDB/databaseAccounts/mongodbDatabases/throughputSettings/read` Gets the throughput of a MongoDB database ▪ `Microsoft.DocumentDB/databaseAccounts/mongodbDatabases/collections/read` Gets the properties of a MongoDB collection ▪ `Microsoft.DocumentDB/databaseAccounts/mongodbDatabases/collections/throughputSettings/read` Gets the throughput of a MongoDB collection ▪ `Microsoft.DocumentDB/databaseAccounts/gremlinDatabases/read` Gets the properties of an Apache Gremlin database ▪ `Microsoft.DocumentDB/databaseAccounts/gremlinDatabases/throughputSettings/read` Gets the throughput of an Apache Gremlin database ▪ `Microsoft.DocumentDB/databaseAccounts/gremlinDatabases/graphs/read` Gets the properties of an Apache Gremlin graph ▪ `Microsoft.DocumentDB/databaseAccounts/gremlinDatabases/graphs/throughputSettings/read` Gets the throughput of an Apache Gremlin graph ▪ `Microsoft.DocumentDB/databaseAccounts/tables/read` Gets the properties of an Azure table ▪ `Microsoft.DocumentDB/databaseAccounts/tables/throughputSettings/read` Gets the throughput of an Azure table ▪ `Microsoft.DocumentDB/databaseAccounts/sqlDatabases/read` Gets the properties of a NoSQL account database ▪ `Microsoft.DocumentDB/databaseAccounts/sqlDatabases/throughputSettings/read` Gets the throughput of a NoSQL account database ▪ `Microsoft.DocumentDB/databaseAccounts/sqlDatabases/containers/read` Get the properties of a NoSQL account container ▪ `Microsoft.DocumentDB/databaseAccounts/sqlDatabases/containers/throughputSettings/read` Gets the throughput of a NoSQL account container ▪ `Microsoft.DocumentDB/databaseAccounts/usages/read` Gets the storage usage for a database account
Discovery of clusters managed by Azure Kubernetes Service (AKS)	▪ `Microsoft.ContainerService/managedClusters/read` Discovers and lists managed container platform clusters

Workload Optimization Manager Functionality	Required Permissions
	▪ `Microsoft.ContainerService/managedClusters/agentPools/read` Discovers agent pools on managed container platform clusters
Discovery of desktop virtualization (VDI)	▪ `Microsoft.DesktopVirtualization/hostpools/read` Discovers Azure Desktop Virtualization host pools ▪ `Microsoft.DesktopVirtualization/hostpools/sessionhosts/read` Discovers session hosts in Azure Desktop Virtualization host pools
Discovery of network resources	▪ `Microsoft.Network/networkInterfaces/read` Lists and gets details about network interfaces for resources ▪ `Microsoft.Network/publicIPAddresses/read` Lists and gets details about public IP addresses for resources
Discovery of pricing information	▪ `Microsoft.Commerce/RateCard/read` Discovers pricing information from the pay-as-you-go rate card; also returns offer data, resource/meter metadata, and rates for the given subscription ▪ `Microsoft.Consumption/pricesheets/read` Discovers pricing information from an Enterprise Agreement price sheet, and lists the price sheets data for a subscription or a management group

Minimum permissions - action execution

The following permissions are required only if you want to execute actions for Azure workloads from Workload Optimization Manager.

Workload Optimization Manager Functionality	Required Permissions
Discovery of locks that could prevent action execution	▪ `Microsoft.Authorization/locks/read` Lists all locks for a subscription, and creates action prerequisites that may prevent action execution if locks prevent write operations
Execution of actions for VMs	▪ `Microsoft.Compute/virtualMachines/deallocate/action` Stops a VM to execute a disruptive action; powers off the VM and releases the allocated compute resources ▪ `Microsoft.Compute/virtualMachines/powerOff/action` Suspends a VM by powering it off. The VM will continue to be billed while suspended. ▪ `Microsoft.Compute/virtualMachines/start/action` Restarts a VM that was stopped to execute a disruptive action ▪ `Microsoft.Compute/virtualMachines/vmSizes/read` Lists available sizes that a VM can update to ▪ `Microsoft.Compute/virtualMachines/write` Updates a VM (for example, its size) as part of executing scale actions ▪ `Microsoft.Compute/diskEncryptionSets/read`

Workload Optimization Manager Functionality	Required Permissions
	(Only required if a VM to be scaled uses Disk Encryption Sets) Gets the properties of the disk encryption set associated with a VM ▪ `Microsoft.Network/networkInterfaces/join/action` Allows a VM or VM scale set to rejoin its network after a scale action executes, by attaching the network interface to the VM ▪ `Microsoft.KeyVault/vaults/deploy/action` (Only required if a VM to be scaled is using Azure Key Vault) Enables access to secrets in a key vault when deploying Azure resources to the VM ▪ Azure Compute Gallery images (Only required during VM scaling execution if VM images are located in a separate Azure Compute Gallery image, such as in a different subscription) – `Microsoft.Compute/galleries/images/read` Gets the properties of Azure Compute Gallery images – `Microsoft.Compute/galleries/images/versions/read` Gets the versions for Azure Compute Gallery images – `Microsoft.Compute/galleries/read` Gets Azure Compute Gallery images – `Microsoft.Compute/images/read` Gets the properties of the image
Execution of actions for VM scale sets and availability sets	▪ `Microsoft.Compute/virtualMachineScaleSets/deallocate/action` Stops a VM scale set to execute a disruptive action; powers off and releases the compute resources of the instances used by the VM scale set ▪ `Microsoft.Compute/virtualMachineScaleSets/start/action` Restarts the VMs that were stopped to execute a disruptive action on VM scale set ▪ `Microsoft.Compute/virtualMachineScaleSets/vmSizes/read` Lists available sizes for creating or updating a VM in a scale set ▪ `Microsoft.Compute/virtualMachineScaleSets/write` Updates a VM scale set as part of executing scale or parking actions ▪ `Microsoft.Insights/AutoscaleSettings/Write` Updates an `autoscale` setting as part of executing scale or parking actions
Execution of actions for Azure Kubernetes Service (AKS) nodes (VMs)	▪ `Microsoft.ContainerService/managedClusters/listClusterAdminCredential/action` Lists the `clusterAdmin` credentials of a managed cluster ▪ `Microsoft.Compute/virtualMachineScaleSets/virtualMachines/delete` Deletes a specific VM in a VM scale set

Workload Optimization Manager Functionality	Required Permissions
	▪ `Microsoft.Compute/virtualMachines/delete` Deletes a specific VM ▪ `Microsoft.ContainerService/managedClusters/write` Creates a new managed cluster or updates an existing one ▪ `Microsoft.ContainerService/managedClusters/agentPools/write` Creates or updates an agent pool in the specified managed cluster ▪ `Microsoft.OperationalInsights/workspaces/sharedkeys/read` Gets the shared keys for the Log Analytics workspace. These keys are used to connect Microsoft Operational Insights agents to the workspace. ▪ `Microsoft.OperationsManagement/solutions/write` Create a new OMS solution ▪ `Microsoft.OperationsManagement/solutions/read` Gets an exiting OMS solution
Execution of actions for volumes	▪ `Microsoft.Compute/disks/write` Resizes or changes the storage tier of volumes Executes scale actions for volumes and re-attaches volumes to VMs after scaling ▪ `Microsoft.Compute/disks/delete` Deletes unattached volumes for managed disks ▪ `Microsoft.Storage/storageAccounts/blobServices/containers/blobs/delete` Deletes unmanaged disks that became unattached after the deletion of unattached volumes ▪ `Microsoft.Storage/storageAccounts/blobServices/containers/blobs/read` Checks if a page blob exists before deleting unmanaged disks that became unattached, and returns the properties of an existing page blob ▪ `Microsoft.Storage/storageAccounts/blobServices/containers/read` Lists blob containers Finds containers hosting unmanaged disks that became unattached and are to be deleted ▪ Recovery Services Vault (Only required if volumes are used for disaster recovery) Prevents the deletion of volumes used for disaster recovery, even if they become unattached – `Microsoft.RecoveryServices/Vaults/read` Gets a list of Recovery Services Vaults containing replicated disks

Workload Optimization Manager Functionality	Required Permissions
	– `Microsoft.RecoveryServices/vaults/replicationProtectedItems/read` Reads the Disk IDs of any protected items in the vaults
Execution of actions for SQL databases (vCore and DTU)	▪ `Microsoft.Sql/servers/databases/write` Executes scale actions for DTU and vCore databases, and updates database properties ▪ `Microsoft.Sql/servers/databases/pause/action` Pauses a database as part of executing a scale or suspend action ▪ `Microsoft.Sql/servers/databases/resume/action` Resumes a paused database as part of executing a scale or suspend action
Execution of actions for SQL Managed Instances	▪ `Microsoft.Sql/managedInstances/write` Executes scale actions for SQL Managed Instances and updates database server properties
Execution of actions for dedicated SQL pools for Azure Synapse Analytics	▪ `Microsoft.Synapse/workspaces/sqlPools/pause/action` Suspends or stops a Synapse SQL Analytics pool ▪ `Microsoft.Synapse/workspaces/sqlPools/resume/action` Resumes a suspended or stopped Synapse SQL Analytics pool
Execution of actions for App Services (plans)	▪ `Microsoft.Web/serverfarms/Delete` Deletes an empty App Service plan (one that is not hosting any running apps) ▪ `Microsoft.Web/serverfarms/Write` Updates an App Service plan as part of a scale action
Execution of actions for Cosmos DB databases and document collections	▪ `Microsoft.DocumentDB/databaseAccounts/cassandraKeyspaces/delete` Deletes an Apache Cassandra keyspace ▪ `Microsoft.DocumentDB/databaseAccounts/cassandraKeyspaces/throughputSettings/write` Updates the throughput of an Apache Cassandra keyspace ▪ `Microsoft.DocumentDB/databaseAccounts/cassandraKeyspaces/tables/throughputSettings/write` Updates the throughput of an Apache Cassandra table ▪ `Microsoft.DocumentDB/databaseAccounts/mongodbDatabases/collections/throughputSettings/write` Updates the throughput of a MongoDB collection ▪ `Microsoft.DocumentDB/databaseAccounts/mongodbDatabases/delete` Deletes a MongoDB database ▪ `Microsoft.DocumentDB/databaseAccounts/mongodbDatabases/throughputSettings/write` Updates the throughput of a MongoDB database

Workload Optimization Manager Functionality	Required Permissions
	▪ `Microsoft.DocumentDB/databaseAccounts/gremlinDatabases/delete` Deletes an Apache Gremlin database ▪ `Microsoft.DocumentDB/databaseAccounts/gremlinDatabases/throughputSettings/write` Updates the throughput of an Apache Gremlin database ▪ `Microsoft.DocumentDB/databaseAccounts/gremlinDatabases/graphs/throughputSettings/write` Updates the throughput of an Apache Gremlin graph ▪ `Microsoft.DocumentDB/databaseAccounts/tables/throughputSettings/write` Updates the throughput of an Azure table ▪ `Microsoft.DocumentDB/databaseAccounts/sqlDatabases/delete` Deletes a NoSQL account database ▪ `Microsoft.DocumentDB/databaseAccounts/sqlDatabases/throughputSettings/write` Updates the throughput of a NoSQL account database ▪ `Microsoft.DocumentDB/databaseAccounts/sqlDatabases/containers/throughputSettings/write` Updates the throughput of NoSQL account container

Sample JSON - minimum permissions for workload monitoring

In Azure, you can create a custom role that specifies the permissions that Workload Optimization Manager needs to monitor workloads in your subscriptions.

When you create the role, you have the option of uploading a JSON file that specifies the permissions and settings for the role. You can copy the content in this section to the JSON file.

```
{
    "properties": {
        "roleName": "<RoleName>",
        "description": "",
        "assignableScopes": [
            "/subscriptions/<Subscription_ID>"
        ],
        "permissions": [
            {
                "actions": [
                    "Microsoft.Authorization/roleAssignments/read",
                    "Microsoft.Authorization/roleDefinitions/read",
                    "Microsoft.Capacity/catalogs/read",
                    "Microsoft.Capacity/reservationorders/reservations/read",
                    "Microsoft.Commerce/RateCard/read",
                    "Microsoft.Compute/availabilitySets/read",
                    "Microsoft.Compute/availabilitySets/vmSizes/read",
                    "Microsoft.Compute/disks/read",
                    "Microsoft.Compute/skus/read",
                    "Microsoft.Compute/virtualMachines/extensions/read",
                    "Microsoft.Compute/virtualMachines/instanceView/read",
                    "Microsoft.Compute/virtualMachines/read",
                    "Microsoft.Compute/virtualMachineScaleSets/networkInterfaces/read",
                    "Microsoft.Compute/virtualMachineScaleSets/read",
```

```
                    "Microsoft.Compute/virtualMachineScaleSets/virtualMachines/extensions/read",
                    "Microsoft.Compute/virtualMachineScaleSets/virtualMachines/instanceView/read",
                    "Microsoft.Compute/virtualMachineScaleSets/virtualMachines/read",
                    "Microsoft.Consumption/pricesheets/read",
                    "Microsoft.ContainerService/managedClusters/agentPools/read",
                    "Microsoft.ContainerService/managedClusters/read",
                    "Microsoft.DatabaseWatcher/watchers/read",
                    "Microsoft.DatabaseWatcher/watchers/targets/read",
                    "Microsoft.DesktopVirtualization/hostpools/read",
                    "Microsoft.DesktopVirtualization/hostpools/sessionhosts/read",
                    "Microsoft.DocumentDB/databaseAccounts/cassandraKeyspaces/read",
                    "Microsoft.DocumentDB/databaseAccounts/cassandraKeyspaces/tables/read",

"Microsoft.DocumentDB/databaseAccounts/cassandraKeyspaces/tables/throughputSettings/read",
                    "Microsoft.DocumentDB/databaseAccounts/cassandraKeyspaces/throughputSettings/read",
                    "Microsoft.DocumentDB/databaseAccounts/databases/collections/metrics/read",
                    "Microsoft.DocumentDB/databaseAccounts/databases/metrics/read",
                    "Microsoft.DocumentDB/databaseAccounts/gremlinDatabases/graphs/read",

"Microsoft.DocumentDB/databaseAccounts/gremlinDatabases/graphs/throughputSettings/read",
                    "Microsoft.DocumentDB/databaseAccounts/gremlinDatabases/read",
                    "Microsoft.DocumentDB/databaseAccounts/gremlinDatabases/throughputSettings/read",
                    "Microsoft.DocumentDB/databaseAccounts/metrics/read",
                    "Microsoft.DocumentDB/databaseAccounts/mongodbDatabases/collections/read",

"Microsoft.DocumentDB/databaseAccounts/mongodbDatabases/collections/throughputSettings/read",
                    "Microsoft.DocumentDB/databaseAccounts/mongodbDatabases/read",
                    "Microsoft.DocumentDB/databaseAccounts/mongodbDatabases/throughputSettings/read",
                    "Microsoft.DocumentDB/databaseAccounts/read",
                    "Microsoft.DocumentDB/databaseAccounts/sqlDatabases/containers/read",

"Microsoft.DocumentDB/databaseAccounts/sqlDatabases/containers/throughputSettings/read",
                    "Microsoft.DocumentDB/databaseAccounts/sqlDatabases/read",
                    "Microsoft.DocumentDB/databaseAccounts/sqlDatabases/throughputSettings/read",
                    "Microsoft.DocumentDB/databaseAccounts/tables/read",
                    "Microsoft.DocumentDB/databaseAccounts/tables/throughputSettings/read",
                    "Microsoft.DocumentDB/databaseAccounts/usages/read",
                    "Microsoft.Insights/AutoscaleSettings/Read",
                    "Microsoft.Insights/Metrics/Read",
                    "Microsoft.Network/networkInterfaces/read",
                    "Microsoft.Network/publicIPAddresses/read",
                    "Microsoft.OperationalInsights/workspaces/query/InsightsMetrics/read",
                    "Microsoft.OperationalInsights/workspaces/query/Perf/read",
                    "Microsoft.OperationalInsights/workspaces/query/read",
                    "Microsoft.OperationalInsights/workspaces/read",
                    "Microsoft.Relay/namespaces/HybridConnections/read",
                    "Microsoft.Resources/subscriptions/locations/read",
                    "Microsoft.Resources/subscriptions/read",
                    "Microsoft.Resources/subscriptions/resourceGroups/read",
                    "Microsoft.Sql/managedInstances/databases/read",
                    "Microsoft.Sql/managedInstances/metrics/read",
                    "Microsoft.Sql/managedInstances/read",
                    "Microsoft.Sql/servers/databases/metrics/read",
                    "Microsoft.Sql/servers/databases/read",
                    "Microsoft.Sql/servers/read",
                    "Microsoft.Storage/storageAccounts/listkeys/action",
                    "Microsoft.Storage/storageAccounts/read",
                    "Microsoft.Synapse/SKUs/read",
                    "Microsoft.Synapse/workspaces/keys/read",
                    "Microsoft.Synapse/workspaces/read",
                    "Microsoft.Synapse/workspaces/sqlDatabases/read",
                    "Microsoft.Synapse/workspaces/sqlPools/dataWarehouseUserActivities/read",
                    "Microsoft.Synapse/workspaces/sqlPools/extensions/read",
                    "Microsoft.Synapse/workspaces/sqlPools/operationStatuses/read",
                    "Microsoft.Synapse/workspaces/sqlPools/read",
```

```
                    "Microsoft.Synapse/workspaces/sqlPools/usages/read",
                    "Microsoft.Synapse/workspaces/sqlUsages/read",
                    "Microsoft.Web/geoRegions/Read",
                    "Microsoft.Web/serverfarms/metrics/read",
                    "Microsoft.Web/serverfarms/Read",
                    "Microsoft.Web/serverfarms/sites/read",
                    "Microsoft.Web/serverfarms/skus/read",
                    "Microsoft.Web/serverfarms/usages/read",
                    "Microsoft.Web/sites/metrics/read",
                    "Microsoft.Web/sites/read",
                    "Microsoft.Web/sites/slots/Read",
                    "Microsoft.Web/sites/usages/read"
                ],
                "notActions": [],
                "dataActions": [],
                "notDataActions": []
            }
        ]
    }
}
```

NOTE:
Be sure to update the following information in the JSON file:

- *<RoleName>* – Specify your preferred name for the custom role.
- *<Subscription_ID>* – Specify the ID of the subscription that Workload Optimization Manager will manage.

Sample JSON – minimum permissions for workload monitoring and action execution

In Azure, you can create a custom role that specifies the permissions that Workload Optimization Manager needs to monitor workloads in your subscriptions and execute actions for these workloads.

When you create the role, you have the option of uploading a JSON file that specifies the permissions and settings for the role. You can copy the content in this section to the JSON file.

```
{
    "properties": {
        "roleName": "<RoleName>",
        "description": "",
        "assignableScopes": [
            "/subscriptions/<Subscription_ID>"
        ],
        "permissions": [
            {
                "actions": [
                    "Microsoft.Authorization/locks/read",
                    "Microsoft.Authorization/roleAssignments/read",
                    "Microsoft.Authorization/roleDefinitions/read",
                    "Microsoft.Capacity/catalogs/read",
                    "Microsoft.Capacity/reservationorders/reservations/read",
                    "Microsoft.Commerce/RateCard/read",
                    "Microsoft.Compute/availabilitySets/read",
                    "Microsoft.Compute/availabilitySets/vmSizes/read",
                    "Microsoft.Compute/diskEncryptionSets/read",
                    "Microsoft.Compute/disks/delete",
                    "Microsoft.Compute/disks/read",
                    "Microsoft.Compute/disks/write",
                    "Microsoft.Compute/galleries/images/read",
                    "Microsoft.Compute/galleries/images/versions/read",
                    "Microsoft.Compute/galleries/read",
                    "Microsoft.Compute/images/read",
                    "Microsoft.Compute/skus/read",
```

```
                    "Microsoft.Compute/virtualMachineScaleSets/deallocate/action",
                    "Microsoft.Compute/virtualMachineScaleSets/networkInterfaces/read",
                    "Microsoft.Compute/virtualMachineScaleSets/read",
                    "Microsoft.Compute/virtualMachineScaleSets/start/action",
                    "Microsoft.Compute/virtualMachineScaleSets/virtualMachines/delete",
                    "Microsoft.Compute/virtualMachineScaleSets/virtualMachines/extensions/read",
                    "Microsoft.Compute/virtualMachineScaleSets/virtualMachines/instanceView/read",
                    "Microsoft.Compute/virtualMachineScaleSets/virtualMachines/read",
                    "Microsoft.Compute/virtualMachineScaleSets/vmSizes/read",
                    "Microsoft.Compute/virtualMachineScaleSets/write",
                    "Microsoft.Compute/virtualMachines/deallocate/action",
                    "Microsoft.Compute/virtualMachines/delete",
                    "Microsoft.Compute/virtualMachines/extensions/read",
                    "Microsoft.Compute/virtualMachines/instanceView/read",
                    "Microsoft.Compute/virtualMachines/powerOff/action",
                    "Microsoft.Compute/virtualMachines/read",
                    "Microsoft.Compute/virtualMachines/start/action",
                    "Microsoft.Compute/virtualMachines/vmSizes/read",
                    "Microsoft.Compute/virtualMachines/write",
                    "Microsoft.Consumption/pricesheets/read",
                    "Microsoft.ContainerService/managedClusters/agentPools/read",
                    "Microsoft.ContainerService/managedClusters/agentPools/write",
                    "Microsoft.ContainerService/managedClusters/listClusterAdminCredential/action",
                    "Microsoft.ContainerService/managedClusters/read",
                    "Microsoft.ContainerService/managedClusters/write",
                    "Microsoft.DatabaseWatcher/watchers/read",
                    "Microsoft.DatabaseWatcher/watchers/targets/read",
                    "Microsoft.DesktopVirtualization/hostpools/read",
                    "Microsoft.DesktopVirtualization/hostpools/sessionhosts/read",
                    "Microsoft.DocumentDB/databaseAccounts/cassandraKeyspaces/delete",
                    "Microsoft.DocumentDB/databaseAccounts/cassandraKeyspaces/read",
                    "Microsoft.DocumentDB/databaseAccounts/cassandraKeyspaces/tables/read",

"Microsoft.DocumentDB/databaseAccounts/cassandraKeyspaces/tables/throughputSettings/read",

"Microsoft.DocumentDB/databaseAccounts/cassandraKeyspaces/tables/throughputSettings/write",
                    "Microsoft.DocumentDB/databaseAccounts/cassandraKeyspaces/throughputSettings/read",
                    "Microsoft.DocumentDB/databaseAccounts/cassandraKeyspaces/throughputSettings/write",
                    "Microsoft.DocumentDB/databaseAccounts/databases/collections/metrics/read",
                    "Microsoft.DocumentDB/databaseAccounts/databases/metrics/read",
                    "Microsoft.DocumentDB/databaseAccounts/gremlinDatabases/delete",
                    "Microsoft.DocumentDB/databaseAccounts/gremlinDatabases/graphs/read",

"Microsoft.DocumentDB/databaseAccounts/gremlinDatabases/graphs/throughputSettings/read",

"Microsoft.DocumentDB/databaseAccounts/gremlinDatabases/graphs/throughputSettings/write",
                    "Microsoft.DocumentDB/databaseAccounts/gremlinDatabases/read",
                    "Microsoft.DocumentDB/databaseAccounts/gremlinDatabases/throughputSettings/read",
                    "Microsoft.DocumentDB/databaseAccounts/gremlinDatabases/throughputSettings/write",
                    "Microsoft.DocumentDB/databaseAccounts/metrics/read",
                    "Microsoft.DocumentDB/databaseAccounts/mongodbDatabases/collections/read",

"Microsoft.DocumentDB/databaseAccounts/mongodbDatabases/collections/throughputSettings/read",

"Microsoft.DocumentDB/databaseAccounts/mongodbDatabases/collections/throughputSettings/write",
                    "Microsoft.DocumentDB/databaseAccounts/mongodbDatabases/delete",
                    "Microsoft.DocumentDB/databaseAccounts/mongodbDatabases/read",
                    "Microsoft.DocumentDB/databaseAccounts/mongodbDatabases/throughputSettings/read",
                    "Microsoft.DocumentDB/databaseAccounts/mongodbDatabases/throughputSettings/write",
                    "Microsoft.DocumentDB/databaseAccounts/read",
                    "Microsoft.DocumentDB/databaseAccounts/sqlDatabases/containers/read",

"Microsoft.DocumentDB/databaseAccounts/sqlDatabases/containers/throughputSettings/read",

"Microsoft.DocumentDB/databaseAccounts/sqlDatabases/containers/throughputSettings/write",
```

```
                    "Microsoft.DocumentDB/databaseAccounts/sqlDatabases/delete",
                    "Microsoft.DocumentDB/databaseAccounts/sqlDatabases/read",
                    "Microsoft.DocumentDB/databaseAccounts/sqlDatabases/throughputSettings/read",
                    "Microsoft.DocumentDB/databaseAccounts/sqlDatabases/throughputSettings/write",
                    "Microsoft.DocumentDB/databaseAccounts/tables/read",
                    "Microsoft.DocumentDB/databaseAccounts/tables/throughputSettings/read",
                    "Microsoft.DocumentDB/databaseAccounts/tables/throughputSettings/write",
                    "Microsoft.DocumentDB/databaseAccounts/usages/read",
                    "Microsoft.Insights/AutoscaleSettings/Read",
                    "Microsoft.Insights/AutoscaleSettings/Write",
                    "Microsoft.Insights/Metrics/Read",
                    "Microsoft.KeyVault/vaults/deploy/action",
                    "Microsoft.Migrate/migrateprojects/read",
                    "Microsoft.Migrate/migrateprojects/solutions/getconfig/action",
                    "Microsoft.Migrate/migrateprojects/solutions/read",
                    "Microsoft.Network/networkInterfaces/join/action",
                    "Microsoft.Network/networkInterfaces/read",
                    "Microsoft.Network/publicIPAddresses/read",
                    "Microsoft.OperationalInsights/workspaces/query/InsightsMetrics/read",
                    "Microsoft.OperationalInsights/workspaces/query/Perf/read",
                    "Microsoft.OperationalInsights/workspaces/query/read",
                    "Microsoft.OperationalInsights/workspaces/read",
                    "Microsoft.OperationalInsights/workspaces/sharedkeys/read",
                    "Microsoft.OperationsManagement/solutions/read",
                    "Microsoft.OperationsManagement/solutions/write",
                    "Microsoft.RecoveryServices/Vaults/read",
                    "Microsoft.RecoveryServices/vaults/replicationProtectedItems/read",
                    "Microsoft.Relay/namespaces/HybridConnections/read",
                    "Microsoft.Resources/subscriptions/locations/read",
                    "Microsoft.Resources/subscriptions/read",
                    "Microsoft.Resources/subscriptions/resourceGroups/read",
                    "Microsoft.Sql/managedInstances/databases/read",
                    "Microsoft.Sql/managedInstances/metrics/read",
                    "Microsoft.Sql/managedInstances/read",
                    "Microsoft.Sql/managedInstances/write",
                    "Microsoft.Sql/servers/databases/metrics/read",
                    "Microsoft.Sql/servers/databases/pause/action",
                    "Microsoft.Sql/servers/databases/read",
                    "Microsoft.Sql/servers/databases/resume/action",
                    "Microsoft.Sql/servers/databases/write",
                    "Microsoft.Sql/servers/read",
                    "Microsoft.Storage/storageAccounts/blobServices/containers/read",
                    "Microsoft.Storage/storageAccounts/listkeys/action",
                    "Microsoft.Storage/storageAccounts/read",
                    "Microsoft.Synapse/SKUs/read",
                    "Microsoft.Synapse/workspaces/keys/read",
                    "Microsoft.Synapse/workspaces/read",
                    "Microsoft.Synapse/workspaces/sqlDatabases/read",
                    "Microsoft.Synapse/workspaces/sqlPools/dataWarehouseUserActivities/read",
                    "Microsoft.Synapse/workspaces/sqlPools/extensions/read",
                    "Microsoft.Synapse/workspaces/sqlPools/operationStatuses/read",
                    "Microsoft.Synapse/workspaces/sqlPools/pause/action",
                    "Microsoft.Synapse/workspaces/sqlPools/read",
                    "Microsoft.Synapse/workspaces/sqlPools/resume/action",
                    "Microsoft.Synapse/workspaces/sqlPools/usages/read",
                    "Microsoft.Synapse/workspaces/sqlUsages/read",
                    "Microsoft.Web/geoRegions/Read",
                    "Microsoft.Web/serverfarms/Delete",
                    "Microsoft.Web/serverfarms/Read",
                    "Microsoft.Web/serverfarms/Write",
                    "Microsoft.Web/serverfarms/metrics/read",
                    "Microsoft.Web/serverfarms/sites/read",
                    "Microsoft.Web/serverfarms/skus/read",
                    "Microsoft.Web/serverfarms/usages/read",
                    "Microsoft.Web/sites/metrics/read",
```

```
                    "Microsoft.Web/sites/read",
                    "Microsoft.Web/sites/slots/Read",
                    "Microsoft.Web/sites/usages/read"
                ],
                "notActions": [],
                "dataActions": [
                    "Microsoft.Storage/storageAccounts/blobServices/containers/blobs/delete",
                    "Microsoft.Storage/storageAccounts/blobServices/containers/blobs/read"
                ],
                "notDataActions": []
            }
        ]
    }
}
```

NOTE:
Be sure to update the following information in the file:

- *<RoleName>* – Specify your preferred name for the custom role.
- *<Subscription_ID>* – Specify the ID of the subscription that Workload Optimization Manager will manage.

Minimum permissions – uploads of migration plans to Azure

Workload Optimization Manager includes a planning feature that simulates the migration of workloads to Azure. After you run the plan, you can upload the plan results *(on page 912)* to Azure to begin the actual migration process. To upload the results, the following permissions are required.

Workload Optimization Manager Functionality	Required permissions
Uploads of plan results to Azure Migrate	▪ `Microsoft.Migrate/migrateprojects/read`: Gets the properties of a Migrate project ▪ `Microsoft.Migrate/migrateprojects/solutions/read`: Gets the properties of a Migrate project solution ▪ `Microsoft.Migrate/migrateprojects/solutions/getconfig/action`: Gets the configuration of a Migrate project solution

Reference: level of support for Azure MCA and EA accounts

Workload Optimization Manager uses a service principal to connect to your Microsoft Customer Agreement (MCA) or Enterprise Agreement (EA) accounts and discover billing data.

NOTE:
For Azure Government *(on page 427)*, Azure currently supports EA accounts only. The Azure Billing target in Workload Optimization Manager aligns with this level of support.

Currently, Workload Optimization Manager support for MCA and EA accounts is *identical* in the following areas:

- Discovery of pricing data

  Pricing data is available for VMs, volumes, SQL databases (DTU and vCore), dedicated SQL pools (for Azure Synapse Analytics), and App Service plans (web apps, logic apps, and function apps).
- Usage of pricing data when recommending actions

  All actions for Azure workloads are supported, including actions to optimize/buy reservations.
- Support for the Buy VM Reservations plan

  The plan is fully supported.

- Scope for cost exports *(on page 436)* (if used to collect billing data)

  The following scopes are *not* supported:

  - Subscription
  - Management Group
  - Resource Group

The level of support *differs* in the following areas.

Item	MCA	EA
Required role	Billing Account Reader role	Enrollment Reader role
Discovery of Azure reservations	All reservations charged under the billing account are discovered. However, the percentage of workloads covered by reservations is currently not discovered.	All reservations charged under the billing account are discovered.
Discovery of Azure reservations costs in non-US dollar currencies	Enabled by default	Not enabled by default. To enable discovery, follow the instructions in this topic *(on page 478)*.
Scope for cost exports *(on page 436)* (if used to collect billing data)	Cost exports must be created at the *Billing Profile* scope. If you previously set up a cost export at the *Billing Account* scope, you must delete the cost export and then create a new one at the Billing Profile scope. Create this cost export for each active Billing Profile that falls under your MCA Billing Account. All cost exports you create for your active Billing Profiles are required to have the same name.	Cost exports must be created at the *Billing Account* scope.

Notice: Microsoft Enterprise Agreement target

The Microsoft Enterprise Agreement target is no longer supported and has been removed from the product.

Who is impacted?

- Customers who added a Microsoft Enterprise Agreement target

  If you are one of those customers, remove the target from the user interface and then set up an Azure Billing target for use with Workload Optimization Manager.

- Customers planning to add a Microsoft Enterprise Agreement target

  If you are one of those customers, set up an Azure Billing target instead.

Why was the Microsoft Enterprise Agreement target removed from the product?

Microsoft has retired the Azure Enterprise Reporting APIs that the target uses to discover billing data. For more information, see the Azure documentation.

What happens now that the Microsoft Enterprise Agreement has been removed from the product?

- The target no longer appears in the list of targets in the Workload Optimization Manager user interface.
- Any existing target remains in the user interface but will no longer be discovered or validated.

  If you encounter issues with the target, your Workload Optimization Manager representative will advise that you remove the target and then set up an Azure Billing target.

What are the benefits of using the Azure Billing target?

The Azure Billing target leverages the new Cost Management APIs and cost exports recommended by Microsoft. Starting in version 3.10.2, the Azure Billing target is the only required target for discovering billing data for global Azure and Azure Government subscriptions.

How do I set up an Azure Billing target?

Perform the following tasks:

1. Decide whether to use the Cost Details API or a cost export. *(on page 435)*

   NOTE:
   For Azure Government:

   If you choose to set up a cost export for Azure Government EA accounts, the cost export must have a destination storage account that is under an Azure Government subscription.

2. Set up a service principal for billing data monitoring in Azure. *(on page 439)*

   NOTE:
   For Azure Government:

   When you set up a service principal, you assign the Enrollment Reader role to the service principal to enable access to billing data.

   To assign the Enrollment Reader role, you execute an API request. For Azure Government EA accounts, it is *not* possible to execute an API request in the Microsoft Learn portal. This is because the API request in the portal is hard-coded and executes against `https://management.azure.com`, which only applies to global Azure EA accounts. The API request for Azure Government EA accounts should execute against `https://management.usgovcloudapi.net`. For more information see this topic *(on page 443)*.

3. Add an Azure Billing target in Workload Optimization Manager. *(on page 446)*

   NOTE:
   For Azure Government:

   When you add the target, be sure to turn on the **US Government** toggle to indicate that the target is an Azure Government account.

Azure monitored resources

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

Azure	Workload Optimization Manager
Virtual Machine (VM)	Virtual Machine (VM)
Disk (Managed)	Volume
Virtual Machine Scale Set	Virtual Machine Spec
App Service - Plan	Virtual Machine Spec
App Service - Web App	App Component Spec
SQL Database (vCore or DTU)	Database

Azure	Workload Optimization Manager
SQL Managed Instance	Database Server
Synapse Analytics (Dedicated SQL Pool)	Database
Cosmos DB - Account	Database Server
Cosmos DB - Database	Database
Cosmos DB - Container	Document Collection
Region	Region

Points to consider:

- Workload Optimization Manager supports discovery and management of entities in certain Azure regions, including Azure Government regions. For details, see Supported Azure Regions *(on page 468)* and Support for Azure Government *(on page 427)*.
- When you first configure an Azure target, under some circumstances the target might show a `No Quotas Available` status. This means that Workload Optimization Manager cannot discover the available templates. This can happen when you initially set up the Azure account and you have not enabled any providers. If this occurs, you can install a single VM in your cloud subscription to make quotas available.
- An Azure subscription can use locked storage or locked resource groups. For such subscriptions, Workload Optimization Manager discovers incomplete data. Locked resources affect Workload Optimization Manager discovery in either of these scenarios:
  - Locked resource group

    Workload Optimization Manager discovers all the entities in the resource group, but does not discover the resource group itself. For example, in the Top Accounts chart, the Resource Groups field will show no resource groups for a subscription that has a locked resource group.
  - Locked storage

    Workload Optimization Manager discovers all the entities in the resource group except the locked storage. It also discovers the resource group.

Monitored resources for virtual machines

Workload Optimization Manager monitors the following resources:

- Virtual memory (vMem)

  Virtual memory (vMem) is the measurement of memory that is in use.

  It is highly recommended that you enable collection of metrics in your environment. Enabling metrics allows Workload Optimization Manager to generate scale actions to optimize VM resource usage. For Workload Optimization Manager to collect metrics, you must enable the collection of these metrics on the VMs in your environment.

  For details, see Azure Memory Metrics Collection *(on page 471)*.
- Virtual CPU (vCPU)

  Virtual CPU is the measurement of CPU that is in use.
- Storage amount

  Storage amount is the measurement of storage capacity that is in use.
- Storage access (IOPS)

  Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
- I/O throughput

  I/O throughput is the measurement of an entity's throughput to the underlying storage.

Monitored resources for volumes

Workload Optimization Manager monitors the following resources:

> **NOTE:**
> Workload Optimization Manager discovers and optimizes Azure *managed* volumes.
>
> According to this Microsoft article, *unmanaged* volumes are deprecated and will be fully retired in 2025. In response, Workload Optimization Manager no longer discovers or monitors unmanaged volumes that are not attached to any VM. Unmanaged volumes that are attached to VMs will continue to be discovered and displayed in the user interface for your reference, but no action will be generated for these volumes.

- Storage amount

  Storage Amount is the storage capacity (disk size) of a volume.

  Workload Optimization Manager discovers Storage Amount, but does not monitor utilization.

  For a Kubeturbo (container) deployment that includes volumes, Kubeturbo monitors Storage Amount utilization for the volumes. You can view utilization information in the Capacity and Usage chart.
- Storage access (IOPS)

  Storage access, also known as IOPS, is the measurement of IOPS capacity that is in use.
- I/O throughput

  I/O throughput is the measurement of I/O throughput capacity that is in use.

Monitored resources for virtual machine specs

- For App Service Plans:

  Workload Optimization Manager monitors the following resources:
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Number of replicas

    Number of replicas is the total number of VM instances underlying an App Service plan.
- For Virtual Machine Scale Sets:

  Workload Optimization Manager does not monitor resources for Virtual Machine Scale Sets, but allows you to park (stop or start) these groups on demand or according to a schedule.

Monitored resources for app component specs (Web App service instances)

Workload Optimization Manager monitors the following resources:

- Response time

  Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
- Virtual CPU (vCPU)

  Virtual CPU is the measurement of CPU that is in use.

Monitored resources for database servers

Workload Optimization Manager monitors the following resources:

- Cosmos DB Account
  - Request unit (RU)

    Request unit (RU) is a performance currency that abstracts CPU, IOPS, and memory that are required to perform the database operations supported by Azure Cosmos DB. Azure Cosmos DB normalizes the cost of all database operations using RUs.
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
- SQL Managed Instance

  NOTE:
  Configure Azure database watcher *(on page 475)* to enable the discovery of metrics for vMem, maximum concurrent sessions, maximum concurrent workers, and DB cache hit rate.

  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Maximum concurrent sessions

    Maximum concurrent sessions is the maximum number of database connections at a time.
  - Maximum concurrent workers

    Maximum concurrent workers is the maximum number of database processes that can handle queries at a time.
  - DB cache hit rate

    DB cache hit rate is the measurement of Database Server accesses that result in cache hits, measured as a percentage of hits versus total attempts. A high cache hit rate indicates efficiency.

Monitored resources for databases

The resources that Workload Optimization Manager can monitor depend on the pricing model in place for the given database entity.

- SQL Database – DTU Pricing Model
  - DTU

    DTU is the measurement of compute capacity for the database. DTU represents CPU, memory, and IOPS/IO Throughput bundled as a single commodity.
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
- SQL Database – vCore Pricing Model
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.
  - Storage amount

Storage amount is the measurement of storage capacity that is in use.

  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

  - I/O throughput

    I/O throughput is the measurement of an entity's throughput to the underlying storage.

- Dedicated SQL Pool (for Azure Synapse Analytics)
  - DWU

    DWU (Data Warehousing Unit) is the measurement of compute capacity for the dedicated SQL pool. DWU represents CPU, memory, and IO Throughput bundled as a single commodity.

  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.

  - Connection

    Connection is the measurement of database connections utilized by applications.

- Cosmos DB Database
  - Request unit (RU)

    Request unit (RU) is a performance currency that abstracts CPU, IOPS, and memory that are required to perform the database operations supported by Azure Cosmos DB. Azure Cosmos DB normalizes the cost of all database operations using RUs.

Monitored resources for document collections (Cosmos DB containers)

Workload Optimization Manager monitors the following resources:

- Request unit (RU)

  Request unit (RU) is a performance currency that abstracts CPU, IOPS, and memory that are required to perform the database operations supported by Azure Cosmos DB. Azure Cosmos DB normalizes the cost of all database operations using RUs.

Supported Azure regions

Workload Optimization Manager supports discovery and management of entities in the following Azure regions:

Region code	Region name	Notes
eastus	East US	
eastus2	East US 2	
centralus	Central US	
northcentralus	North Central US	
southcentralus	South Central US	
westcentralus	West Central US	
westus	West US	
westus2	West US 2	
westus3	West US 3	
canadaeast	Canada East	
canadacentral	Canada Central	

Region code	Region name	Notes
brazilsouth	Brazil South	
northeurope	North Europe	
westeurope	West Europe	
francecentral	France Central	
ukwest	UK West	
uksouth	UK South	
germanywestcentral	Germany West Central	
norwayeast	Norway East	
switzerlandnorth	Switzerland North	
eastasia	East Asia	
southeastasia	Southeast Asia	
australiaeast	Australia East	
australiasoutheast	Australia Southeast	
australiacentral	Australia Central	
centralindia	Central India	
southindia	South India	
westindia	West India	
japaneast	Japan East	
japanwest	Japan West	
koreacentral	Korea Central	
koreasouth	Korea South	
uaenorth	UAE North	
southafricanorth	South Africa North	
usgovarizona	USGov Arizona	Restricted Access (Azure Government)
usgovtexas	USGov Texas	Restricted Access (Azure Government)
usgovvirginia	USGov Virginia	Restricted Access (Azure Government)

Support for Azure App Service

Azure App Service is an HTTP-based service for hosting apps. With Azure App Service, app developers can easily create enterprise-ready apps and deploy them on a scalable and reliable cloud infrastructure.

Azure App Service offers several types of apps, including web apps, mobile apps, API apps, and logic apps. Each app runs as a set of *app instances* and is associated with a *plan* that defines compute resources (CPU, memory, and storage) available to the app.

When you add an Azure account as a target:

- Workload Optimization Manager discovers all the plans in that account, except App Service Environment v3 I4, I5, and I6. Plans appear as `Virtual Machine Spec` entities in the supply chain.
- For plans associated with *web apps*, Workload Optimization Manager discovers the related app instances. In the supply chain, app instances appear as `App Component Spec` entities. Workload Optimization Manager generates actions to scale these plans to optimize app performance.
- For plans associated with the other types of apps, Workload Optimization Manager does not generate scale actions or discover the related app instances.
- For plans that are not associated with any type of app, Workload Optimization Manager generates delete actions as a cost-saving measure.

  For details about scale and delete actions, see Virtual Machine Spec *(on page 673)*.

To discover plans and app instances, you must provide permissions to support all the actions you want to perform. For a list of permissions, see Azure Permissions *(on page 447)*.

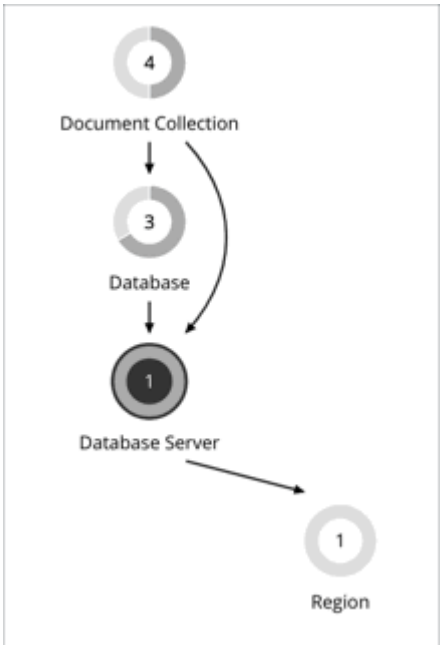
Support for Azure Cosmos DB

Azure Cosmos DB is a fully managed platform-as-a-service (PaaS) for modern app development. To deploy Cosmos DB, you create a Cosmos DB account in an Azure resource group in your subscription, and then create databases and containers within that account.

When you add Azure targets, Workload Optimization Manager discovers the accounts, databases and containers that make up your Cosmos DB deployment, and then maps them to the corresponding entities in the supply chain.

Cosmos DB	Workload Optimization Manager
Account	Database Server *(on page 702)*
Database	Database *(on page 713)*
Container	Document Collection *(on page 722)*

The following image shows how your Cosmos DB environment is represented in the supply chain.



An account has a total throughput limit that represents capacity, but actual throughput (measured in Request Units or RUs) is provisioned at the database or container (document collection) level.

Workload Optimization Manager optimizes databases and containers. It uses percentile calculations to measure Request Unit (RU) utilization for these entities and then recommends actions to scale RUs to optimize performance and costs.

Workload Optimization Manager also recommends cost-saving actions that reconfigure databases to remove unused provisioned throughput or delete databases that are not hosting document collections.

Workload Optimization Manager requires specific permissions to discover your Cosmos DB deployment. For details, see Azure Permissions *(on page 447)*.

Azure VM memory metrics collection

It is highly recommended that you enable collection of metrics in your environment. Enabling metrics allows Workload Optimization Manager to generate scale actions to optimize VM resource usage. For Workload Optimization Manager to collect metrics, you must enable the collection of these metrics on the VMs in your environment.

Workload Optimization Manager collects memory metrics for Azure VMs using the following mechanisms, in order:

1. Log Analytics workspace - `InsightsMetrics` table

   This is the preferred collection method. This mechanism is used when a VM is configured with the Azure Monitor Agent (AMA). This agent collects guest OS memory metrics and sends it to the Log Analytics workspace `InsightsMetrics` table. Workload Optimization Manager then collects available memory metrics by querying this table.

   For more information on the Azure Monitor Agent, see the Azure documentation.

2. Log Analytics workspace - `Perf` table

   This mechanism is used when a VM is configured with the Log Analytics agent (also known as MMA and OMS) and is connected to a Log Analytics workspace. This agent sends guest OS memory metrics to the Log Analytics workspace `Perf` table. Workload Optimization Manager then collects available memory metrics by querying this table.

3. Diagnostics monitoring based on the storage account SDK

   This legacy mechanism installs an Azure diagnostics extension on the VM, which sends guest OS level memory metrics to a designated storage account table. During the discovery process, Workload Optimization Manager collects VM memory from these tables. For this mechanism, Workload Optimization Manager needs the `Storage Account Contributor` role to access the storage account keys and establish a connection that can retrieve VM memory statistics.

4. Metrics REST API

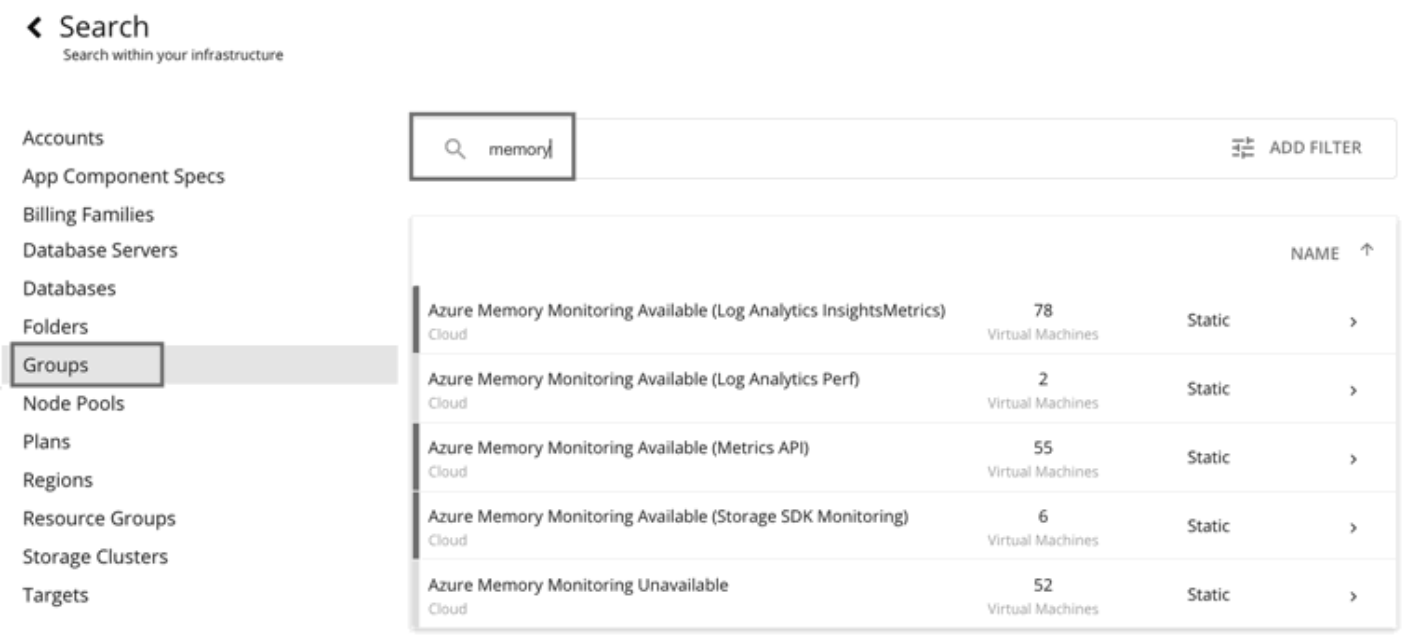
   If memory cannot be collected using the previously-mentioned mechanisms, Workload Optimization Manager tries to collect host OS based memory for the VM through REST API calls, which is the same mechanism used for collecting other VM metrics, such as CPU usage. Workload Optimization Manager collects VM memory and generates memory-driven scale actions where appropriate.

   NOTE:
   Azure provides available memory through the REST API (in preview mode) or the Azure portal (in a VM's **Monitoring > Metrics** section).

Memory source groups for Azure VMs in Workload Optimization Manager

In Workload Optimization Manager, memory source groups reflect the source of Azure VM memory metrics. To search for a specific memory source group in the Workload Optimization Manager user interface, open the Search page, choose **Groups**, and enter `memory` in the search field. Each group contains the list of VMs for which memory was collected through a specific mechanism. These are dynamic groups that are continuously updated based on each discovery.

The `Azure Memory Monitoring Unavailable` group contains VMs for which memory could not be collected using any of the collection mechanisms. This group may contain VMs that are not in a `RUNNING` state. For these VMs, metrics data is often not available and memory-driven scale actions are not generated. However, other scale actions may be generated.



Azure VM metrics collection in batch mode

Workload Optimization Manager can collect memory, CPU, and other metrics for a large number of VMs when you enable VM metric collection in batch mode.

VM metric collection in batch mode is *optional* and is *disabled* by default. When enabled, Workload Optimization Manager makes a single API call to collect metrics for 50 VMs (this number is configurable), instead of a single API call for each VM. This can reduce the likelihood of Azure throttling API requests from Workload Optimization Manager, which happens when the number of API calls exceeds a certain limit. Throttling can sometimes result in missing metrics in Workload Optimization Manager.

To enable VM metric collection in batch mode:

1. Enable the following property for the `mediation-azure` probe.

   ```
   metrics.batch.api.enabled
   ```

   In the Workload Optimization Manager Swagger API, set the property to `true` on the probe of type `Azure Subscription`.

   Contact your Workload Optimization Manager representative if you need assistance setting the probe property.

2. Assign the `Monitoring Reader` role to the subscription containing the VMs for which metrics will be collected in batches.

   Azure requires this role to enable the collection of metrics through the Azure Batch APIs. This role can view all monitoring data in a subscription but cannot modify any resources or edit any settings related to monitoring resources. For details about roles, see the Azure documentation.

Azure VM GPU metrics collection

Use NVIDIA Data Center GPU Manager (DCGM) to collect Azure VM GPU metrics. Workload Optimization Manager uses these metrics to generate scale actions that optimize performance and costs. For a list of metrics and associated scale actions, see this topic *(on page 643)*.

DCGM exporter exposes metrics in Prometheus format through the Telegraf agent with the Prometheus plugin. The plugin collects metrics from DCGM exporter and uses Azure Monitor to log metrics directly to the VM. Workload Optimization Manager then uses the Azure Monitor REST API to collect GPU metrics. This is the same API that enables the collection of standard metrics, such as vCPU. In the Azure portal, when you view a VM's details in the **Metrics** section, the GPU metrics are available in the `telegraf/prometheus` namespace.

This topic provides an example DCGM configuration to describe the configuration that is required to collect Azure VM GPU metrics. The example uses the `Standard_NC4as_T4_v3` instance type, which Workload Optimization Manager supports.

1. Verify that the GPU instance type for the VM that you plan to create is supported in the region where you will deploy the VM. Some GPU instance types may only be supported in specific Azure regions.
2. When you create a VM from the Azure portal or CLI, choose a base image that supports the GPU instance type for the VM.

   In this example, the instance type is `Standard_NC4as_T4_v3` and the corresponding base image is `NVIDIA GPU-Optimized VMI with vGPU driver - v22.08.0 - x64 Gen 2`.

   Wait for the VM to be created before proceeding.
3. In the Azure portal, browse to **Monitoring > Insights** and then click **Enable**. In the Monitoring configuration window that opens, click **Configure** to set the Azure Monitor Agent configuration. Wait for the deployment to finish.
4. Browse to **Settings > Extensions + applications**. Verify that the VM extension called **AzureMonitorLinuxAgent** is available and its status is `Provisioning Succeeded`.
5. Browse to **Connect > Native SSH** and then click **Select**. Verify that SSH is configured and record the VM's public IP address.
6. From the command line, establish an SSH connection to the new VM, using the username that you used when you created the VM. The default username is `azureuser`. Replace it as appropriate.
7. In the home directory of the user, create a `dcp-metrics-included.csv` file with the following content.

```
# Column description: DCGM FIELD, Prometheus metric type, help message
DCGM_FI_DEV_COUNT, counter, Number of Devices on the node.
DCGM_FI_DEV_GPU_UTIL, gauge, GPU utilization (in %).
DCGM_FI_DEV_FB_USED, gauge, Framebuffer memory used (in MiB).
DCGM_FI_PROF_PIPE_TENSOR_ACTIVE, gauge, Ratio of cycles the tensor (HMMA) pipe is active (in %).
DCGM_FI_PROF_DRAM_ACTIVE,        gauge, Ratio of cycles the device memory interface is active
 sending or receiving data (in %).
DCGM_FI_PROF_PIPE_FP64_ACTIVE,   gauge, Ratio of cycles the fp64 pipes are active (in %).
DCGM_FI_PROF_PIPE_FP32_ACTIVE,   gauge, Ratio of cycles the fp32 pipes are active (in %).
DCGM_FI_PROF_PIPE_FP16_ACTIVE,   gauge, Ratio of cycles the fp16 pipes are active (in %).
```

8. Run the following command to set up the DCGM exporter docker container.

```
sudo docker run --pid=host --privileged -e DCGM_EXPORTER_INTERVAL=60000 --gpus all --restart=always
 -d -v /proc:/proc -v $PWD/dcp-metrics-included.csv:/etc/dcgm-exporter/default-counters.csv -p
 9400:9400 --name dcgm-exporter nvcr.io/nvidia/k8s/dcgm-exporter:3.3.5-3.4.0-ubuntu22.04
```

9. Set up the InfluxData Telegraf agent.

```
# Add repo.
curl -s https://repos.influxdata.com/influxdb.key | sudo apt-key add -
source /etc/lsb-release
```

```
sudo echo "deb https://repos.influxdata.com/${DISTRIB_ID,,} ${DISTRIB_CODENAME} stable" | sudo
 tee /etc/apt/sources.list.d/influxdb.list
```

```
sudo curl -fsSL https://repos.influxdata.com/influxdata-archive_compat.key | sudo apt-key
 --keyring /etc/apt/trusted.gpg.d/influxdata-archive_compat.gpg add
```

```
# Install package.
sudo apt-get update
sudo apt-get install telegraf
```

```
# Generate the new Telegraf config file in the current directory.
telegraf --input-filter prometheus --output-filter azure_monitor config > azm-telegraf.conf
```

```
# Fix endpoint and version in the generated file.
sed -i 's|^\s*# metric_version = 1|  metric_version = 2|;
 s|http://localhost:9100/metrics|http://localhost:9400/metrics|' azm-telegraf.conf
```

```
# Verify telegraf configuration.
telegraf --config azm-telegraf.conf --test
```

```
# Replace the example config with the new generated config.
sudo cp azm-telegraf.conf /etc/telegraf/telegraf.conf
```

```
# Restart telegraf agent.
sudo systemctl stop telegraf
sudo systemctl enable --now telegraf
```

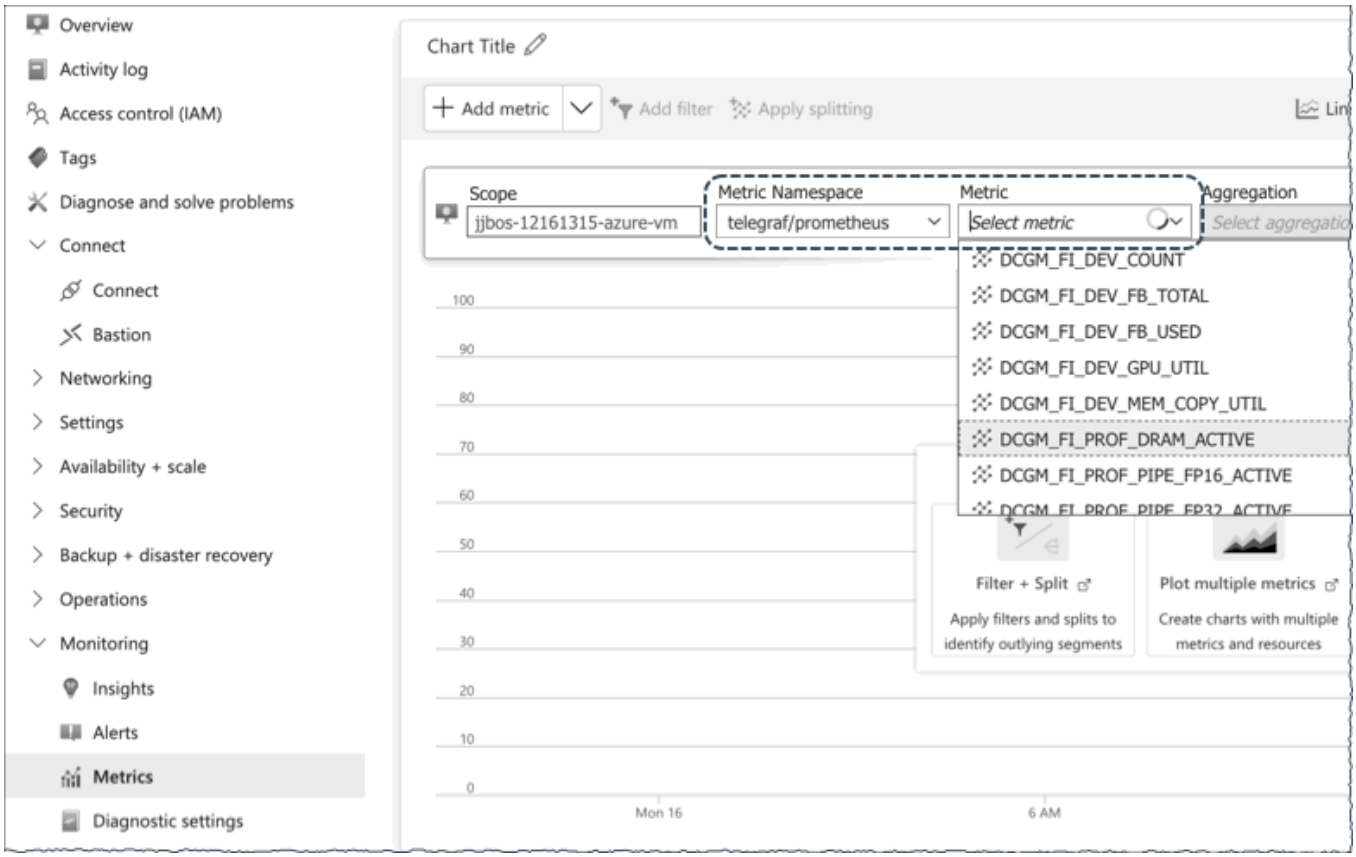
Wait a few minutes for the GPU metrics to display in the Azure portal.

10. (Optional) To view non-0 GPU metrics, run the DCGM tester tool and add some load to the GPU.

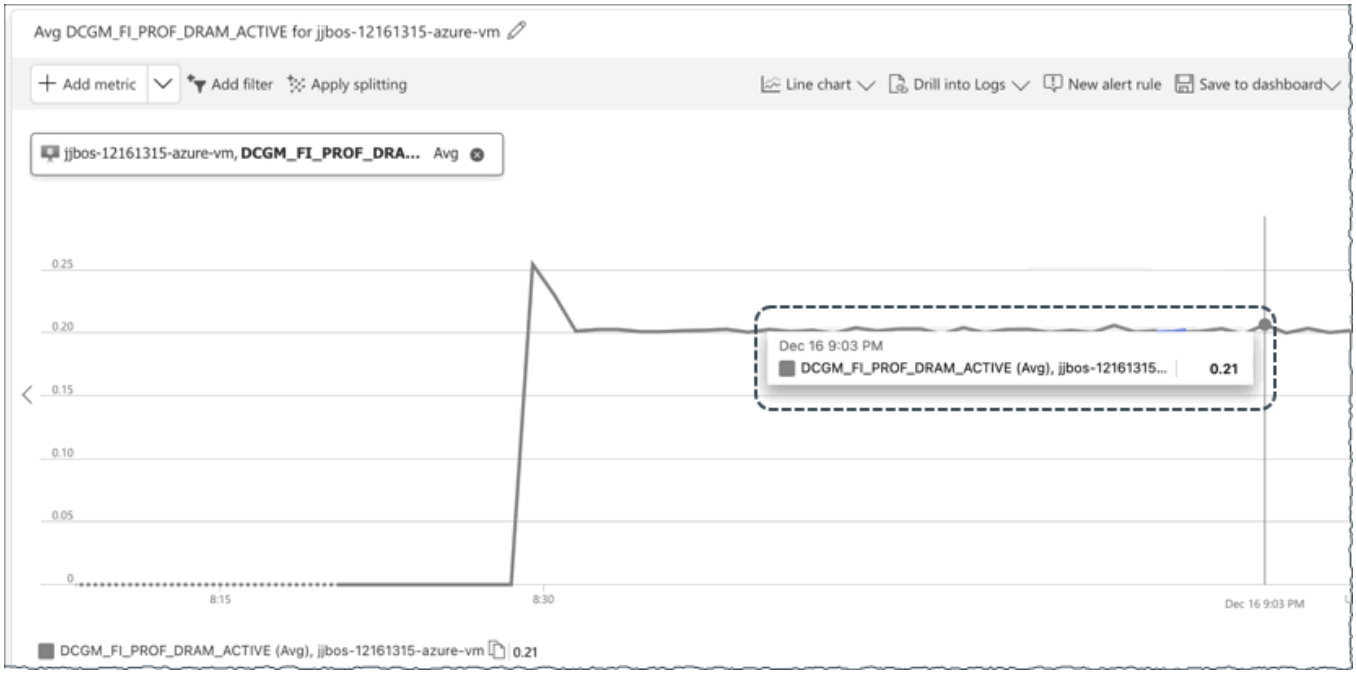
```
# Login to the DCGM exporter docker container.
docker exec -it `docker ps -aqf "name=dcgm-exporter"` /bin/bash
```

```
# Run the DCGM tester inside the container. Type Ctrl-C when you want to stop the test and exit the
 container.
dcgmproftester12 --no-dcgm-validation --target-max-value -t 1004,1005,1006,1007,1008 -d 36000
```

11. In the Azure portal, browse to **Monitoring > Metrics**, choose **telegraf/prometheus** from the **Metrics Namespace** dropdown list, and then choose a GPU metric from the **Metric** dropdown list.



If you ran the DCGM tester tool and added some load to the GPU (as described in a previous step), you will see non-0 metrics values in the chart. Otherwise, you may see 0 values.



12. Verify that GPU metrics display in the Workload Optimization Manager user interface.

    It may take between 30 minutes and an hour for GPU metrics to display in the user interface. When metrics are available to display, you can set the scope to a VM and view metrics in the Capacity and usage chart.

Capacity and usage Show all

jjbos-12161315-azure-vm

Commodity	Capacity	Used	Utilization
GPU Count	1	1	100%
GPU Tensor	65	57.457	88.4%
GPU Memory BW	320.00 GB/sec	67.21 GB/sec	21%
Virtual Memory	28.00 GB	2.44 GB	8.72%
GPU Memory	16.00 GB	0.49 GB	3.04%

For additional configuration and setup references, see the NVIDIA documentation and the Azure documentation.

Azure SQL metrics collection using database watcher

Azure database watcher is a managed monitoring solution for database services in the Azure SQL family. Workload Optimization Manager connects to Azure database watcher to collect metrics for SQL Managed Instances *(on page 691)* and then uses these metrics to generate scale actions that optimize performance and costs. Metrics include vMem, maximum concurrent sessions, maximum concurrent workers, and DB cache hit rate.

It is important that you set up database watcher to take full advantage of Workload Optimization Manager's optimization capabilities for SQL Managed Instances.

> **NOTE:**
> According to the Azure documentation, database watcher is currently in preview. In Workload Optimization Manager, metrics collection may change as Azure changes its level of support for database watcher.

Task overview

To complete the tasks outlined in this topic, you must have administrator access to the Azure portal.

1. Create an Azure database watcher.
2. Grant Workload Optimization Manager viewer access to the database watcher's data store.

Creating an Azure database watcher

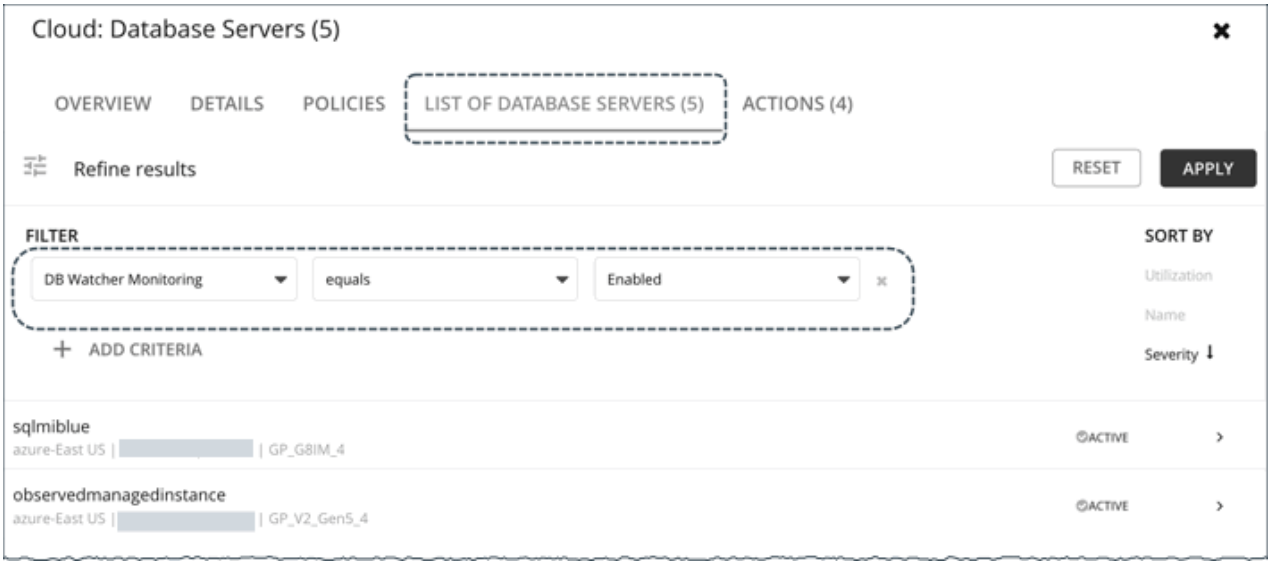
To create a database watcher, follow the instructions in the Azure documentation.

> **NOTE:**
> When you create a database watcher, you select a data store for the watcher. Currently, Workload Optimization Manager supports Azure Data Explorer only. All other data stores available in Azure are not supported.

Granting Workload Optimization Manager viewer access to the database watcher's data store

1. Browse to the Azure Data Explorer Clusters page.

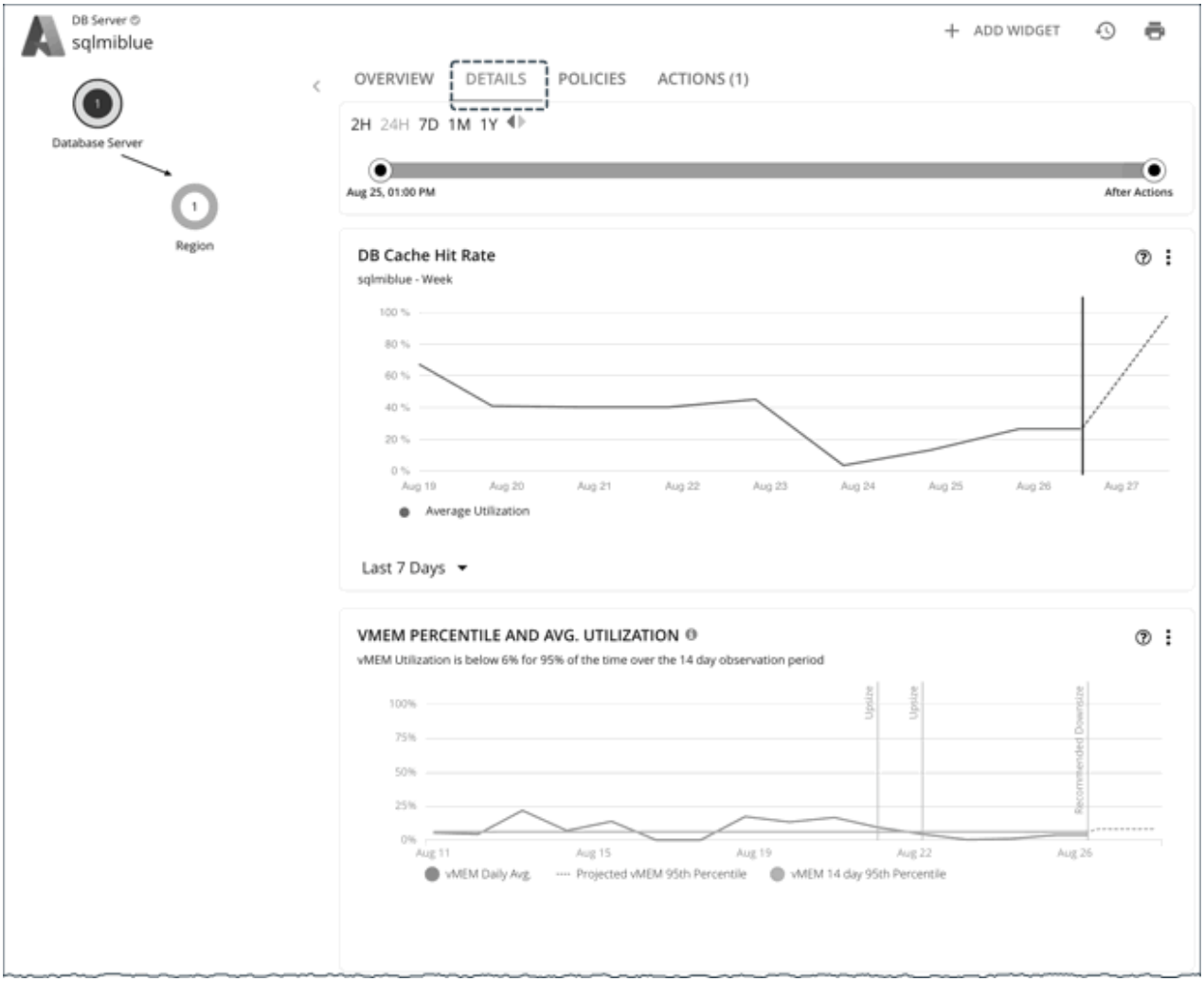
   https://portal.azure.com/#browse/Microsoft.Kusto%2Fclusters
2. Choose the data store that you configured for Azure database watcher.
3. In the window that opens, choose **Data > Databases**.
4. Choose the database for the Azure database watcher.
5. In the page that opens, choose **Overview > Permissions**.
6. Choose **Add > Viewer**.
7. Search for the service principal that you previously set up for Workload Optimization Manager and then choose **Select**.
8. Verify that Workload Optimization Manager can access Azure database watcher and collect metrics.
   a. Log in to the Workload Optimization Manager user interface.
   b. In the Home page, choose the **Cloud** tab and then choose the Database Server entity in the supply chain.
   c. In the window that opens, choose the **List of Database Servers** tab and then choose **Filter**. Set `DB Watcher Monitoring = Enabled` as the filter.



If viewer access was granted successfully, the page updates with a list of SQL Managed Instances that send metrics to Azure database watcher.

d. Choose a SQL Managed Instance.

e. In the new window that opens, choose the **Details** tab.



Charts show the metrics collected from Azure database watcher.

Discovery of Azure reservations costs in non-US dollar currencies

Workload Optimization Manager discover costs for Azure reservations through the Azure Retail API and your billing accounts. By default, the API returns cost information in US dollars (USD), while your billing accounts return cost information in your local currency.

If you use Azure Enterprise Agreement and your local currency is one of the non-US dollar currencies listed in the Azure documentation, Workload Optimization Manager might sometimes calculate unrealistic savings values.

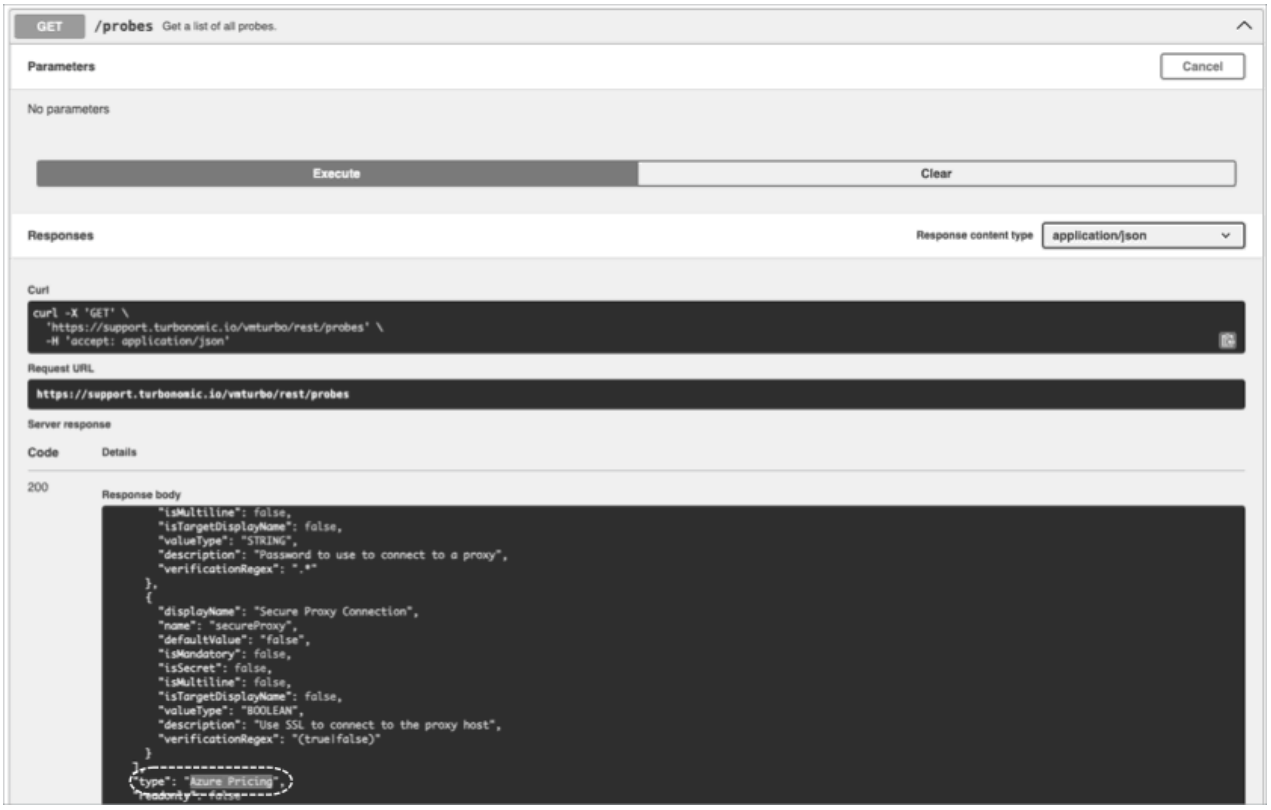
NOTE:

By default, Workload Optimization Manager uses the dollar symbol ($) when displaying the costs and savings that it discovers or calculates for your cloud workloads. You can set a different symbol to match your preferred currency. For example, if your cloud provider bills you in euros, change the currency symbol to €.

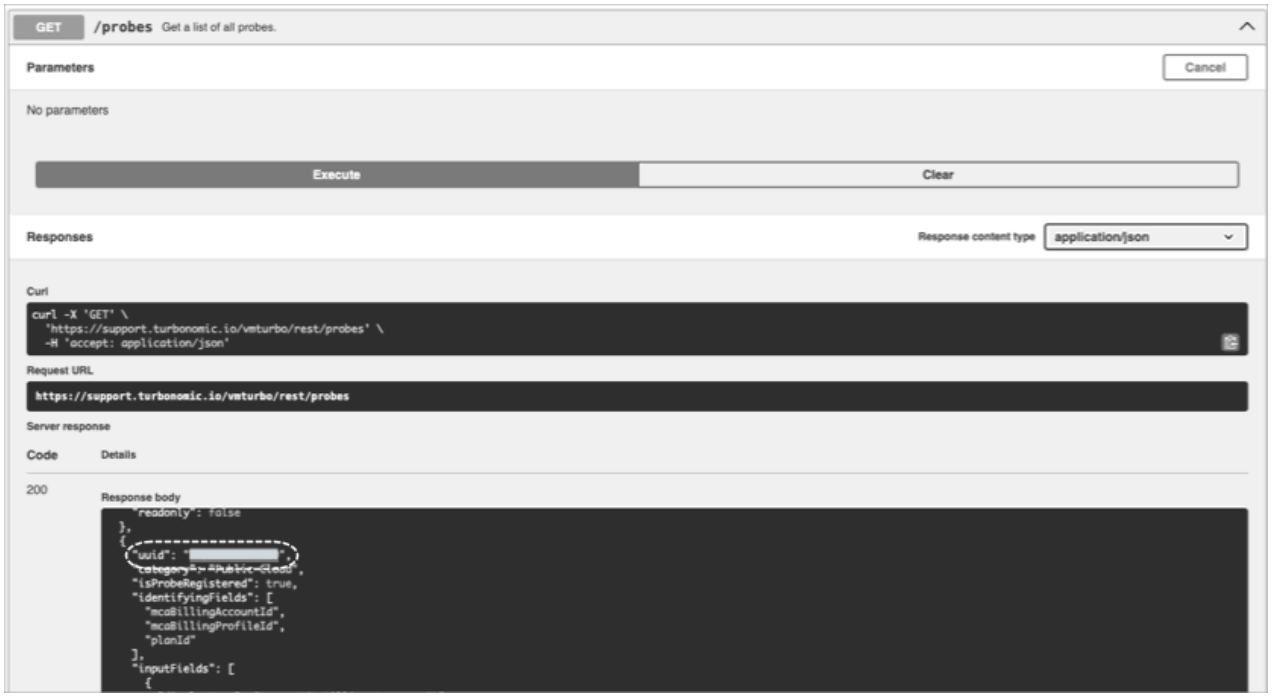
Go to **Settings > Billing and Costs > Currency** to change the symbol. Note that currency symbols are for display purposes only. Workload Optimization Manager does not convert monetary amounts when you switch symbols.

To enable the discovery of Azure reservations costs in your local currency, perform the following steps.

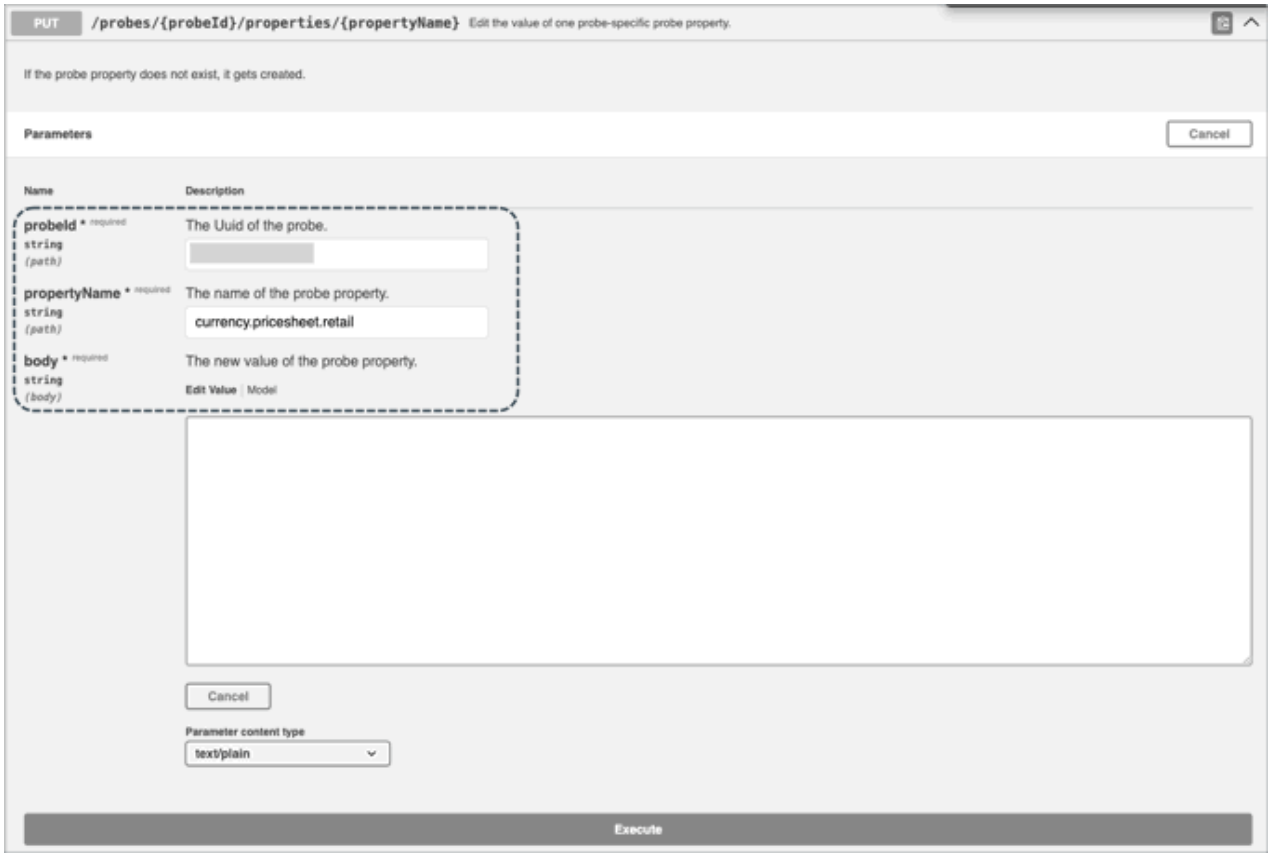
1. Get the probe UUID for Azure Pricing.
    a. Log in to the Workload Optimization Manager user interface and then navigate to `https://<your_instance_address>/apidoc`.
    b. Expand [INTERNAL USE - NOT SUPPORTED].
    c. Expand GET `/probes`, click **Try it out**, and then click **Execute**.
    d. Find the Azure Pricing probe.



e. Copy the UUID for the Azure Pricing probe.



2. Set your local currency.
   a. Expand PUT `/probes/{probeId}/properties/{propertyName}` and click **Try it out**.
   b. Configure the following settings:



- `probeid` – Specify the UUID for the Azure Pricing probe.
- `propertyName` – Specify `currency.pricesheet.retail`.
- `body` – Specify the 3 digit currency code. Codes are available in the Azure documentation.

c. Click **Execute** to apply your changes. A `200` response displays.

3. Restart the `mediation-azurepricing` pod from the backend.

```
kubectl delete pod mediation-azurepricing<tab>
```

NOTE:
After the restart, it may take Workload Optimization Manager a few hours to discover Azure reservations costs in the new currency.

Azure actions

Workload Optimization Manager monitors the state and performance of your workloads and then recommends actions to optimize these workloads at the lowest possible cost.

NOTE:
Use the Potential Savings and Necessary Investments charts to view pending actions and evaluate their impact on your cloud expenditure.

Historically-informed action generation for Azure workloads

Workload Optimization Manager can generate actions informed by 14 days of historical utilization data immediately after a new Azure service principal target is added. To take advantage of this feature, add at least one Azure service principal target.

Currently, actions are generated for virtual machines, volumes, and SQL databases (DTU and vCore) in both global Azure and Azure Government accounts.

The following components run after you add a target:

- `metrics-processor`
- `metrics-adapter-azure-monitormetrics`

Actions for virtual machines

Workload Optimization Manager supports the following actions:

- Scale

  Change the VM instance to use a different instance type or tier to optimize performance and costs.
- Discount-related actions

  If you have a high percentage of on-demand VMs, you can reduce your monthly costs by increasing Azure reservations coverage. To increase coverage, you scale VMs to instance types that have existing capacity.
- Stop and start (also known as 'parking' actions)

  Stop a VM for a given period of time to reduce your cloud expenses, and then start it at a later time.

For details, see Actions for Azure VMs *(on page 641)*.

Actions for volumes

Workload Optimization Manager discovers and optimizes Azure *managed* volumes.

According to this Microsoft article, *unmanaged* volumes are deprecated and will be fully retired in 2025. In response, Workload Optimization Manager no longer discovers or monitors unmanaged volumes that are not attached to any VM. Unmanaged volumes that are attached to VMs will continue to be discovered and displayed in the user interface for your reference, but no action will be generated for these volumes.

Workload Optimization Manager supports the following actions:

- Scale

  Scale attached volumes to optimize performance and costs.
- Delete

  Delete unattached volumes as a cost-saving measure. Workload Optimization Manager generates an action immediately after discovering an unattached volume.

For details, see Cloud Volume Actions *(on page 728)*.

Actions for virtual machine specs

Workload Optimization Manager supports the following actions:

- For Azure App Service plans:
  - Scale

    Scale Azure App Service plans to optimize app performance or reduce costs, while complying with business policies.
  - Delete

    Delete empty Azure App Service plans as a cost-saving measure. A plan is considered empty if it is not hosting any running apps.

  For details, see Virtual Machine Spec Actions *(on page 674)*.

- For Azure Virtual Machine Scale Sets:

  Stop and start (also known as 'parking' actions)

  Stop a Virtual Machine Scale Set for a given period of time to reduce your cloud expenses, and then start it at a later time.

  For details, see Parking: Stop or Start Cloud Resources *(on page 942)*.

Actions for app component specs (App service instances)

None

Workload Optimization Manager does not recommend actions for App Component Specs, but it does recommend actions for the underlying Virtual Machine Specs. For details, see Virtual Machine Spec Actions *(on page 674)*.

Actions for database servers

Workload Optimization Manager supports the following actions:

- For SQL Managed Instances:

  Scale

  Change the Managed Instance to use a different instance type or tier to optimize performance and costs.

  For details, see Actions for SQL Managed Instances *(on page 693)*.

- For Cosmos DB Accounts:

  None

  Workload Optimization Manager does not recommend actions for a Cosmos DB account but it does recommend actions for the databases *(on page 713)* and document collections *(on page 722)* in the account.

Actions for databases

Workload Optimization Manager supports the following actions:

- Scale SQL database
  - DTU pricing model

    Scale DTU and storage resources to optimize performance and costs.
  - vCore pricing model

    Scale vCPU, vMem, IOPS, throughput and storage resources to optimize performance and costs.

  For details, see Scale Actions for SQL Databases *(on page 705)*.
- Scale Cosmos DB database

  Scale request units (RUs) to optimize performance and costs.

  For details, see Scale Actions for Cosmos DB Databases *(on page 714)*.
- Reconfigure Cosmos DB database

  Remove unused provisioned throughput to reduce costs.

  For details, see Reconfigure Actions for Cosmos DB Databases *(on page 715)*.
- Delete Cosmos DB database

  Delete a database with provisioned throughput but without any underlying document collection (container) to reduce costs.

  For details, see Delete Actions for Cosmos DB Databases *(on page 715)*.
- Suspend or stop dedicated SQL pool

Suspend or stop a dedicated SQL pool (used in Azure Synapse Analytics) to reduce compute costs.

- Workload Optimization Manager analysis generates suspend actions for *idle* pools.

  NOTE:
  Currently, Workload Optimization Manager analysis does not generate actions to start a suspended pool. You can start a suspended pool from the Parking page (see the next item) or from Azure.

  For details, see Suspend Actions for Dedicated SQL Pools *(on page 712)*.

- You can use the Parking page to stop pools (running or idle), either on-demand or according to a schedule. Use the same page to start a stopped pool.

  For details, see Parking: Stop or Start Cloud Resources *(on page 942)*.

Actions for document collections (Cosmos DB containers)

Workload Optimization Manager supports the following actions:

- Scale

  Scale Request Units (RUs) to optimize performance and costs.

  For details, see Document Collection Actions *(on page 723)*.

Storage targets

Adding a storage Target enables Workload Optimization Manager to connect to your storage subsystem through a native or SMI-S provider API. Workload Optimization Manager uses the target's API to access and collect information from each of the underlying disk arrays. The information is used to set disk performance characteristics according to the type and capacity of storage, leading to improved workload placement.

Similarly, Workload Optimization Manager determines the relationships between storage controllers and disk arrays, and the location of datastores within those arrays. This information also helps optimize workload placement at a more granular level.

For on-prem applications, this optimization will enable Workload Optimization Manager to make more informed decisions about which storage devices the workloads hosting your applications run on, and assist in assuring application SLO. In the cloud, storage data is handled as part of the public cloud target.

Both virtual machines and containers benefit from this level of optimization. In the case of short-lived containers, Workload Optimization Managerwill suggest the best datastore to hold persistent data, and paired with a container or hypervisor target, will select the optimal match of compute and storage resources. For longer-lived containers and virtual machines, each workload will be continually assessed for SLA/SLO, and recommendations to move or resize storages will ensure the continued efficiency of your environment.

The following section describes the storage supply chain. For information on how to add specific storage targets, the resources Workload Optimization Manager can monitor for the various supply chain entities, and the actions it can take to optimize the environment, refer to the target configuration instructions for your specific storage type.

Entity mapping

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

EMC VMAX	EMC XtremIO	HPE 3Par	NetApp	Nutanix	Pure	Workload Optimization Manager
Volume (Regular, Thin, Meta)	Volume	Virtual Volume	Volume	Container	Volume	Storage

EMC VMAX	EMC XtremIO	HPE 3Par	NetApp	Nutanix	Pure	Workload Optimization Manager
Disk Group or Thin Pool	XTremIO Cluster	CPG	Aggregate	Storage Pool	Shelf Array	Disk Array
VMAX Array	XTremIO Cluster	Controller	Controller / Filer	Controller VM	Controller	Storage Controller

Storage targets (storage controllers) add Storage Controller and Disk Array entities to the supply chain. Disk Array entities in turn host Storage entities (datastores).

EMC VMAX

NOTE:
After reviewing telemetry data and conducting interviews with customers to determine use of this capability, this target was deemed appropriate to deprecate. If you replaced this technology with something that is currently not supported, submit an idea at the Cloud Management and AIOps Idea Portal. The last Workload Optimization Manager version that will contain this target is 3.13.0.

Workload Optimization Manager supports management of VMAX2 and 3 Series storage arrays. The VMAX series is a family of enterprise storage arrays designed for SAN environments. Workload Optimization Manager connects to VMAX storage systems via an EMC SMI-S provider that has the disk arrays added to it. A single SMI-S provider can communicate with one or more disk arrays. When you specify an SMI-S provider as a target, Workload Optimization Manager discovers all the added disk arrays.

NOTE:
Workload Optimization Manager does not utilize Unisphere. Data is collected exclusively from the SMI-S provider.

Workload Optimization Manager will create Storage Groups based on the SLO levels defined in VMAX3 Targets. By default, Storage vMotion actions will respect these SLO levels based on the configured response time.

Prerequisites

- EMC VMAX using SMI-S 8.1+
- A service account that Workload Optimization Manager can use to connect to the EMC SMI-S Provider (typically the default `admin` account)

Adding VMAX targets

1. Click **Settings > Target Configuration**.
2. Click **New Target > Storage**.
3. Select **VMAX**.
4. Configure the following settings:
    - Address

      The IP or hostname of the SMI-S provider. If the provider address begins with https, you must follow the IP with the port used to connect.
    - Use Secure Connection

      If checked, port 5989 will be used to connect. If unchecked, port 5988 will be used.
    - Username

      The Username for the SMI-S provider.

- Password

  The Password for the SMI-S provider.

Entity mapping

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

EMC VMAX	Workload Optimization Manager
Volume (Regular, Thin, Meta)	Storage
Storage Resource Pool (VMAX3) / Thick Provisioned Pool (earlier)	Disk Array
Storage Group (VMAX3) / Thin Provisioned Pool (earlier)	Logical Pool
VMAX Array	Storage Controller

Monitored resources

When calculating available storage, Workload Optimization Manager excludes disks devoted to the VMAX operating system by default. If these disks are assigned to new RAID groups or storage pools, the capacity of those disks will then be considered when calculating the capacity of the Storage Controller.

Workload Optimization Manager monitors the following resources:

- Storage
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

    NOTE:
    When it generates actions, Workload Optimization Manager does not consider IOPS throttling that it discovers on storage entities. Analysis uses the IOPS it discovers on Logical Pool or Disk Array entities.
  - Latency

    Latency is the measurement of storage latency.
- Logical Pool
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - Latency

    Latency is the measurement of storage latency.

- Disk Array
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - Latency

    Latency is the measurement of storage latency.
- Storage Controller
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.

Actions

Workload Optimization Manager supports the following actions:

- Storage
  - Provision (Clone)
  - Delete
  - Move
  - Resize (V-Volumes only)

    This action can only be executed outside Workload Optimization Manager.
- Logical Pool
  - Resize

    This action can only be executed outside Workload Optimization Manager.

EMC XtremIO

NOTE:
After reviewing telemetry data and conducting interviews with customers to determine use of this capability, this target was deemed appropriate to deprecate. If you replaced this technology with something that is currently not supported, submit an idea at the Cloud Management and AIOps Idea Portal. The last Workload Optimization Manager version that will contain this target is 3.13.0.

EMC® XtremIO® is a flash-based (SSD) storage solution, designed to push data to applications at higher speeds. The system building blocks are SAN appliances called X-Bricks. A deployment is organized into clusters of X-Bricks, and the clusters are managed by the XtremIO Management Server (XMS).

Workload Optimization Manager connects to X-Bricks through the XMS. The XMS presents a unified view of each connected X-Brick cluster, rather than exposing the individual X-Bricks within each cluster. Within Workload Optimization Manager, each X-Brick cluster displays as a single storage controller with an associated disk array.

The relationship between Storage entities and individual X-Bricks within the cluster is not exposed through the XMS — Workload Optimization Manager cannot make recommendations to move datastores from one X-Brick to another. Additionally, the X-Brick has a fixed form factor — Workload Optimization Manager does not recommend resize actions for disk array or storage controller resources.

Workload Optimization Manager recognizes XtremIO arrays as flash storage and sets the IOPS capacity on discovered arrays accordingly.

Prerequisites

- EMC XtremIO XMS 4.0+
- A service user account on XMS, typically the default `xmsadmin` account

  Workload Optimization Manager uses this account to connect to the XMS and execute commands through the XtremIO API.

Adding XtremIO targets

1. Click **Settings > Target Configuration**.
2. Click **New Target > Storage**.
3. Select **EMC XtremIO**.
4. Configure the following settings:
   - Address

     The name or IP address of the XtremIO Management Server (XMS).
   - Username/Password

     Credentials for a user account on the XMS.

Entity mapping

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

XTremIO	Workload Optimization Manager
Volume	Storage
XTremIO Cluster	Disk Array
XTremIO Cluster	Storage Controller

Storage targets (storage controllers) add Storage Controller and Disk Array entities to the supply chain. Disk Array entities then host Storage entities (datastores).

Monitored resources

When calculating available storage, Workload Optimization Manager excludes disks devoted to the VNX operating system.

Workload Optimization Manager monitors the following resources:

- Storage
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

    NOTE:
    When it generates actions, Workload Optimization Manager does not consider IOPS throttling that it discovers on storage entities. Analysis uses the IOPS it discovers on Logical Pool or Disk Array entities.

  - Latency

    Latency is the measurement of storage latency.
- Disk Array
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - Latency

    Latency is the measurement of storage latency.
- Storage Controller

    NOTE:
    Not all targets of the same type provide all possible commodities. For example, some storage controllers do not expose CPU activity. When a metric is not collected, its chart in the user interface will not display data.

  - CPU

    CPU is the measurement of CPU that is reserved or in use.
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.

    The storage allocated to a storage controller is the total of all the physical space available to aggregates managed by that storage controller.

Actions

Workload Optimization Manager supports the following actions:

- Storage
  - Provision

    This action can only be executed outside Workload Optimization Manager.
  - Resize Up

    This action can only be executed outside Workload Optimization Manager.
- Storage Controller
  - Provision

    This action can only be executed outside Workload Optimization Manager.

EMC ScaleIO

NOTE:
After reviewing telemetry data and conducting interviews with customers to determine use of this capability, this target was deemed appropriate to deprecate. If you replaced this technology with something that is currently not supported, submit an idea at the Cloud Management and AIOps Idea Portal. The last Workload Optimization Manager version that will contain this target is 3.13.0.

EMC ScaleIO is an example of Software-Defined Storage for the datacenter. It creates a Virtual SAN overlaying commodity infrastructure that consists of multiple LAN-connected Servers with locally attached commodity Storage. It presents a standard Block Storage interface to Applications accessing the Virtual SAN.

Workload Optimization Manager communicates with the EMC ScaleIO system via the REST API Gateway.

Prerequisites

- EMC ScaleIO 2.x or 3.x
- A service account that Workload Optimization Manager can use to connect to the ScaleIO Gateway.

Adding EMC ScaleIO targets

1. Click **Settings > Target Configuration**.
2. Click **New Target > Storage**.
3. Select **EMC ScaleIO**.
4. Configure the following settings:
    - Address

      The IP or hostname of the Gateway.
    - Username

      The Username for the Gateway service account.
    - Password

      The Password for the Gateway service account.

Entity mapping

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

EMC ScaleIO	Workload Optimization Manager
Volume	Storage
Storage Pool	Disk Array
Protection Domain	Storage Controller

Monitored resources

Workload Optimization Manager monitors the following resources:

- Storage

  NOTE:
  Not all targets of the same type provide all possible commodities. For example, some storage controllers do not expose CPU activity. When a metric is not collected, its chart in the user interface will not display data.

  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

- Disk Array
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

    NOTE:
    When it generates actions, Workload Optimization Manager does not consider IOPS throttling that it discovers on storage entities. Analysis uses the IOPS it discovers on Logical Pool or Disk Array entities.
  - Latency

    Latency is the measurement of storage latency.
- Storage Controller
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.

Actions

Workload Optimization Manager supports the following actions:

- Storage
  - Provision (Clone)
  - Resize (disabled by default)

    This action can only be executed outside Workload Optimization Manager.
- Disk Array
  - Resize Disk Array

    This action can only be executed outside Workload Optimization Manager.
- Protection Domain
  - Provision (Clone)

    This action can only be executed outside Workload Optimization Manager.

EMC VPLEX

NOTE:
After reviewing telemetry data and conducting interviews with customers to determine use of this capability, this target was deemed appropriate to deprecate. If you replaced this technology with something that is currently not supported, submit an idea at the Cloud Management and AIOps Idea Portal. The last Workload Optimization Manager version that will contain this target is 3.13.0.

Workload Optimization Manager supports management of EMC VPLEX virtual storage systems in a local configuration, via the VPLEX API. Currently, Workload Optimization Manager does not support Metro or Geo configurations.

VPLEX is used to aggregate and refine data collected between connected Storage and Hypervisor targets. VPLEX supports one-to-one, one-to-many, and many-to-one relationships between virtual volumes and LUNs. Only one-to-one mapping between virtual volume and LUNs is supported by Workload Optimization Manager.

Prerequisites

- EMC VPLEX Local Architecture with 1:1 mapping of virtual volumes and LUNs
- VPLEX Management Server
- Hypervisor target supported by Workload Optimization Manager
- Storage target supported by Workload Optimization Manager

> **NOTE:**
> In order for Workload Optimization Manager to make use of the information provided by VPLEX, you must also add the hypervisor and storage layered under it as targets.

VPLEX permissions

Workload Optimization Manager Functionality	Required Permissions
Monitoring	Service Account
Action Execution	Admin account

Adding EMC VPLEX targets

1. Click **Settings > Target Configuration**.
2. Click **New Target > Storage**.
3. Select **EMC VPLEX**.
4. Configure the following settings:
   - Address:

     The IP or Hostname of the VPLEX Management Server
   - Username:

     The Username for the VPLEX Management Server
   - Password:

     The Password for the VPLEX Management Server
   - Port Number:

     The port number for the remote management connection. The default port number for the VPLEX Management server is 443
   - Secure Connection:

     Select this option to use a secure connection (HTTPS)

     > **NOTE:**
     > The default port (443) uses a secure connection.

Actions

For this target, actions are generated and executed via the underlying storage targets. Workload Optimization Manager will use the enhanced visibility provided by VPLEX to make more accurate storage decisions, such as recommending storage vMotion between pools.

HPE 3PAR

HPE 3PAR StoreServ systems use controller nodes to manage pools of storage resources and present a single storage system to consumers. Workload Optimization Manager communicates with the HPE 3PAR system via both the WSAPI and SMI-S providers that are installed on the 3PAR controller node.

Prerequisites

- HPE 3PAR InForm OS 3.2.2+, 3PAR SMI-S, 3PAR WSAPI
- SMI-S Provider enabled and configured on the controller node
- WSAPI Provider enabled and configured on the controller node
- A service account on the controller node that Workload Optimization Manager can use to connect to the SMI-S and WSPAI providers

> **NOTE:**
> For discovery and monitoring, the Workload Optimization Manager service account must have the `Browse` permission on all monitored domains. To exclude domains from monitoring, the Workload Optimization Manager service account must have no permissions on those domains. For action execution, Workload Optimization Manager requires the `Edit` permission.

Adding HPE 3PAR targets

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **HPE 3PAR**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Hostname or IP address

     Specify the name or IP address of the 3PAR controller node.

     By default, the controller provides SMI-S data over port 5988 for HTTP or port 5989 for HTTPS. If your installation uses a different port for SMI-S, include the Port in the Hostname or IP address field.
   - Username

     Specify the username for a user account on the controller node that Workload Optimization Manager uses to connect to the target.
   - Password

     Specify the password for a user account on the controller node that Workload Optimization Manager uses to connect to the target.
   - Connect using HTTPS

     If you select this option, Workload Optimization Manager connects to the target servers using HTTPS. Make sure that the required certificate is configured for use on the host.

     This option is selected by default.
   - Web services API port

     For non-default configurations, specify the WSAPI port as defined in the HPE 3PAR Management Console. By default, the port is 8080.

Entity mapping

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

HPE 3PAR	Workload Optimization Manager
Virtual Volume	Storage
CPG	Disk Array
AO Configuration	Logical Pool
Controller	Storage Controller

Storage targets (storage controllers) add Storage Controller, Logical Pool and Disk Array entities to the supply chain. Logical Pool and Disk Array entities then host Storage entities (datastores).

Setting up the SMI-S provider

The HPE 3PAR SMI-S Provider should be installed on the controller node. It is disabled by default – you must ensure that it is installed properly and running on the controller node.

To enable the SMI-S provider:

1. Log into the HPE 3PAR Command Line Interface (CLI).

   Open a secure shell session (ssh) on the controller node. Default credentials are `3paradm/3pardata`.
2. Check the current status of the SMI-S provider.

   In the shell session, run the command:

   ```
   showcim
   ```
3. If the CIM service is not running, start it.

   To enable the CIM service and the SMI-S provider, run the command:

   ```
   startcim
   ```

To stop the SMI-S provider, execute the command `stopcim -f -x`.

Setting up the WSAPI provider

The HPE 3PAR WSAPI Provider should be installed on the controller node.

To enable the WSAPI provider:

1. Log into the HPE 3PAR Command Line Interface (CLI).

   Open a secure shell session (ssh) on the controller node. Default credentials are `3paradm/3pardata`.
2. Check the current status of the WSAPI provider.

   In the shell session, run the command:

   ```
   showwsapi
   ```
3. If the WSAPI service is not running, start it by runing the command:

   ```
   startwsapi
   ```

   To allow only insecure connections run the command:

   ```
   set wsapi -http enable
   ```

   Or to allow only secure connections, run the command:

   ```
   set wsapi -https enable
   ```

To stop the WSAPI provider, execute the command `stopwsapi -f`.

3Par adaptive optimization

Adaptive Optimization (AO) for HPE 3Par enables management of data storage across two or three tiers. AO places storage regions on the appropriate tier in response to periodic analysis that AO performs.

To work with the storage in an AO group, Workload Optimization Manager:

- Discovers each Common Provisioning Group (CPG) as a disk array

  In the Workload Optimization Manager user interface, these disk arrays do not host storage – They appear empty. Workload Optimization Manager will not recommend storage moves between these disk arrays, because such moves would conflict with AO block-level placement.
- Creates a single logical pool that hosts all the datastores in an AO group

  This logical pool represents the AO group, and it includes all the member CPGs. Workload Optimization Manager considers this single logical pool when it performs analysis – It can recommend moving storage into or out of the AO group. Also, Workload Optimization Manager aggregates resource capacity in this logical pool. For example, the IOPS capacity for the AO logical pool is a combination of IOPS capacity for the constituent CPGs.

You can see the AO logical pool in the Workload Optimization Manager user interface. The display name for this logical pool is the name of the AO Configuration.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Storage
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

    NOTE:
    When it generates actions, Workload Optimization Manager does not consider IOPS throttling that it discovers on storage entities. Analysis uses the IOPS it discovers on Logical Pool or Disk Array entities.
  - Latency

    Latency is the measurement of storage latency.
- Disk array
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - Latency

    Latency is the measurement of storage latency.

- Logical pool
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - Latency

    Latency is the measurement of storage latency.
- Storage controller

  NOTE:
  Not all targets of the same type provide all possible commodities. For example, some storage controllers do not expose CPU activity. When a metric is not collected, its chart in the user interface will not display data.

  - CPU

    CPU is the measurement of CPU that is reserved or in use.
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.

    The storage allocated to a storage controller is the total of all the physical space available to aggregates managed by that storage controller.

Actions

Workload Optimization Manager supports the following actions:

- Storage
  - Provision
  - Resize up/down
- Disk array
  - Provision
  - Resize up/down
- Logical pool
  - Provision

    This action can only be executed outside Workload Optimization Manager.
  - Resize up/down

    This action can only be executed outside Workload Optimization Manager.
- Storage controller
  - Provision

    This action can only be executed outside Workload Optimization Manager.

IBM FlashSystem

The IBM® FlashSystem family comprises all-Flash storage-array platforms. The platform delivers an active-active dual controller system. Assuming a pair of controller nodes, the active-active controller pair can process I/O for a specific volume through either node.

A FlashSystem control enclosure contains multiple NVMe-attached IBM FlashCore® Modules or other self-encryption NVMe-attached SSD drives. Each control enclosure contains two identical node canisters. A node canister provides host interfaces, management interfaces, and interfaces to the control enclosure. The pair of nodes within a single enclosure is known as an input/output (I/O) group.

NOTE:
For deployments that include mirrored volumes, Workload Optimization Manager discovers the Primary copy, and uses data from that for analysis and to generate actions. It does not discover the Secondary copy.

Prerequisites

- IBM FlashSystem running on Spectrum Virtualize 8.3.1.2 or later (8.4.2.0 or later recommended)

  To take advantage of significant improvements to the Spectrum Virtualize interface, it is recommended that you use Spectrum Virtualize version 8.4.2.0 or later.
- A service user account on the FlashSystem controller

  You must configure a user account on the FlashSystem with a `Monitor` role. This user account must be dedicated to a single Workload Optimization Manager instance. To manage one FlashSystem with multiple Workload Optimization Manager instances, you must create a different user account for each Workload Optimization Manager instance.

IBM FlashSystem permissions

Workload Optimization Manager functionality	Required permissions
Monitoring	Service account with a `Monitor` role

Considerations

When you configure a FlashSystem target, you should consider:

- FlashSystem devices under IBM SVC management are not currently supported.
- Workload Optimization Manager does not discover mirrored volumes or child pools.
- Clusters must consist of a single enclosure.

Adding IBM FlashSystem targets

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **IBM FlashSystem**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Hostname or management IP address

     Specify the hostname or management IP address of the FlashSystem cluster.
   - Username

     Specify the username for a dedicated Workload Optimization Manager user account on the FlashSystem instance.
   - Password

     Specify the password for the user account that Workload Optimization Manager uses to connect to the target.

Entity mapping

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

FlashSystem	Workload Optimization Manager	Description
Volume (VDisk)	Storage	A logical disk that the system presents to attached hosts. You create Volumes from the extents that are available in the pool.
Pool (MDiskGroup)	Logical Pool	An allocated amount of capacity that jointly contains all of the data for a specified set of volumes. All MDisks in a pool are split into extents of the same size.
MDisk (Array)	Disk Array	A logical unit of physical storage. MDisks are not visible to host systems.
System (Enclosure or Controller)	Storage Controller	A FlashSystem control enclosure.

IBM FlashSystem Cluster targets add Storage Controller, Disk Array, and Logical Pool entities to the supply chain. Logical Pool entities provide Storage entities, which the supply chain stitches to Hypervisor storage (for example, vCenter datastores).

Monitored resources

To discover and monitor resources, Workload Optimization Manager uses the Spectrum Virtualize REST interface, and it securely retrieves `iostats` files from the configuration node.

Workload Optimization Manager monitors the following resources:

- Storage
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

    NOTE:
    When it generates actions, Workload Optimization Manager does not consider IOPS throttling that it discovers on storage entities. Analysis uses the IOPS it discovers on Logical Pool or Disk Array entities.
  - Latency

    Latency is the measurement of storage latency.
- Logical pool
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - Latency

    Latency is the measurement of storage latency.

- Disk array
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - Latency

    Latency is the measurement of storage latency.
- Storage controller
  - CPU

    CPU is the measurement of CPU that is reserved or in use.
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.

    The storage allocated to a storage controller is the total of all the physical space available to aggregates managed by that storage controller.

Actions

Workload Optimization Manager supports the following actions:

- Storage controller
  - Provision

    This action is disabled by default.
- Disk array
  - Resize

    This action is disabled by default.
  - Move

    This action is disabled by default.
  - Provision

    This action is disabled by default.
  - Start

    This action is disabled by default.
  - Suspend

    This action is disabled by default.
- Logical pool
  - Resize (add or remove MDisk)

    This action is recommended by default.
  - Move

    This action is disabled by default.
  - Provision

    This action is disabled by default.
  - Start

    This action is disabled by default.

  - Suspend

    This action is disabled by default.
- Storage
  - Delete

    This action is manually executable by default.
  - Resize

    This action is recommended by default.
  - Move

    This action is recommended by default.
  - Provision

    This action is recommended by default.
  - Start

    This action is recommended by default.
  - Suspend

    This action is disabled by default.
- Virtual machine
  - Storage vMotion

    This action is recommended by default.

NetApp

The Storage Control Module adds support for NetApp filers running the Data ONTAP operating system. NetApp storage controllers are Storage Virtual Machines that manage storage arrays. Workload Optimization Manager connects to these storage controllers to support NetApp targets in Cluster-Mode (C-Mode).

Prerequisites

- NetApp Cluster Mode using ONTAP 8.0+ (excluding AFF and SolidFire)
- Transport Layer Security (TLS) is enabled
- A service account Workload Optimization Manager can use to connect to the NetApp target

Enabling TLS

Starting with version 5.4, by default Workload Optimization Manager requires Transport Layer Security (TLS) version 1.2 to establish secure communications with targets. NetApp filers have TLS disabled by default, and the latest version they support is TLSv1. If your NetApp target fails to validate on Workload Optimization Manager 5.4 or later, this is probably the cause.

If target validation fails because of TLS support, you might see validation errors with the following strings:

- No appropriate protocol

  To correct this error, ensure that you have enabled the latest version of TLS that your target technology supports. If this does not resolve the issue, contact Cisco Technical Support.
- Certificates does not conform to algorithm constraints

  To correct this error, refer to your NetApp documentation for instructions to generate a certification key with a length of 2048 or greater on your target server. If this does not resolve the issue, please contact Cisco Technical Support.

For information about enabling TLS, see the Data ONTAP **System Administration Guide** for sections on the SSL protocol.

Administrator role for the service user account

To discover and fully manage NetApp disk arrays, Workload Optimization Manager must have a service account that grants privileges to execute commands through the NetApp filer's OnTap API (ontapi). In most cases, you can create the administrator account through the NetApp OnCommand System Manager, or from the NetApp command line as shown in the following example.

```
security login create -role admin -username Cisco -application ontapi -authmethod password
```

If you prefer not to grant full administrator rights, see Creating Restricted Service Accounts In NetApp *(on page 503)*

Adding NetApp targets

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **NetApp**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
    - Hostname or IP address

      Specify the name or IP address of the NetApp cluster management server.
    - Username

      Specify the username for the NetApp service user account that Workload Optimization Manager uses to connect to the target.
    - Password

      Specify the password for the NetApp service user account that Workload Optimization Manager uses to connect to the target.
    - Connect using HTTPS

      If this option is selected, Workload Optimization Manager connects to the target by using SSL.

      This option is selected by default.

After validating the new target, Workload Optimization Manager discovers the connected storage entities. This table compares terms used in NetApp to those used in Workload Optimization Manager:

Entity mapping

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

NetApp	Workload Optimization Manager
Volume	Storage
Aggregate	Disk Array
Controller / Filer	Storage Controller

Storage targets (storage controllers) add Storage Controller and Disk Array entities to the supply chain. Disk Array entities then host Storage entities (datastores).

Monitored resources

Workload Optimization Manager monitors the following resources:

NOTE:
In NetApp environments, the storage controller shows 100% utilization when there are no more disks in a SPARE state that the storage controller can utilize in an aggregate. This does not indicate that the storage controller has no capacity.

- Storage
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

    NOTE:
    When it generates actions, Workload Optimization Manager does not consider IOPS throttling that it discovers on storage entities. Analysis uses the IOPS it discovers on Logical Pool or Disk Array entities.
  - Latency

    Latency is the measurement of storage latency.
- Disk array
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - Latency

    Latency is the measurement of storage latency.
- Storage controller

  NOTE:
  Not all targets of the same type provide all possible commodities. For example, some storage controllers do not expose CPU activity. When a metric is not collected, its chart in the user interface will not display data.

  - CPU

    CPU is the measurement of CPU that is reserved or in use.
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.

    The storage allocated to a storage controller is the total of all the physical space available to aggregates managed by that storage controller.

Actions

Workload Optimization Manager supports the following actions:

- Storage
  - Move
  - Provision

    This action can only be executed outside Workload Optimization Manager.
  - Resize Up

    This action can only be executed outside Workload Optimization Manager.
- Disk array
  - Resize up

    This action can only be executed outside Workload Optimization Manager.
  - Move

    This action can only be executed outside Workload Optimization Manager.
  - Provision

    This action can only be executed outside Workload Optimization Manager.
- Storage controller
  - Provision

    This action can only be executed outside Workload Optimization Manager.

Note that Workload Optimization Manager can automate moving a datastore to a disk array on the same storage controller, as well as moves to a disk array on a different storage controller.

Creating restricted service accounts in NetApp

While Workload Optimization Manager prefers a NetApp service account with administrator rights, it is possible to create an account that has limited access.

NetApp 9.x restricted service account setup

Complete the following steps to use a service account that does not have full administrator rights.

1. Log into the NetApp filer from a command shell.
2. Create a role and assign it permission to execute each of the following commands:

   For example:

   ```
   security login role create -role RoleName -cmddirname “storage aggregate show” -vserver
    Cluster-Name
   ```

   The required capabilities are listed here:

   - `cluster identity modify`
   - `cluster identity show`
   - `lun create`
   - `lun igroup create`
   - `lun igroup modify`

- `lun igroup show`
- `lun mapping create`
- `lun mapping delete`
- `lun mapping show`
- `lun modify`
- `lun show`
- `network interface create`
- `network interface delete`
- `network interface modify`
- `network interface show`
- `statistics show`
- `storage aggregate create`
- `storage aggregate modify`
- `storage aggregate show`
- `storage disk show`
- `system controller flash-cache show`
- `system node modify`
- `system node show`
- `version`
- `volume create`
- `volume modify`
- `volume move modify`
- `volume move show`
- `volume move start`
- `volume qtree create`
- `volume qtree show`

- `volume show`
- `volume snapshot create`
- `volume snapshot modify`
- `volume snapshot show`
- `vserver create`
- `vserver fcp nodename`
- `vserver iscsi nodename`
- `vserver modify`
- `vserver options`
- `vserver show`

3. For execution privileges, run the following commands for the given role, where `Role-Name` is the name of the role you are creating, and `Cluster-Name` identifies the cluster you want the role to affect. You must execute these commands individually to set privileges that affect each individual cluster:

   - `security login role create -role Role-Name -access all -cmddirname "volume offline" -vserver Cluster-Name`
   - `security login role create -role Role-Name -access all -cmddirname "volume unmount" -vserver Cluster-Name`
   - `security login role create -role Role-Name -access all -cmddirname "volume move" -vserver Cluster-Name`
   - `security login role create -role Role-Name -access all -cmddirname "volume delete" -vserver Cluster-Name`

4. Create a user that will use the newly-created role.

   For example:

   ```
   security login create -User-Name RoleUser -r Workload Optimization ManagerRole
   ```

5. Enter a password for the new user when prompted.
6. Give the user access to the `ssh` and `ontapi` applications by using the following commands, replacing `Role-Name` and `RoleUser` with the role and user you created:

   ```
   security login create -role Role-Name -username RoleUser -application ontapi
    -authmethod password
   ```

   ```
   security login create -role Role-Name -username RoleUser -application ssh
    -authmethod password
   ```

NetApp C-Mode restricted service account setup

Complete the following steps to use a service account that does not have full administrator rights.

1. Log into the NetApp filer from a command shell.
2. Create a role and assign it permission to run each of the following commands:

   - `aggr-get-iter`
   - `igroup-get-iter`
   - `cluster-identity-get`
   - `lun-map-get-iter`
   - `net-interface-get-iter`
   - `storage-disk-get-iter`
   - `system-get-node-info-iter`
   - `volume-get-iter`
   - `vserver-get-iter`
   - `fcp-node-get-name`
   - `flash-device-get-iter`
   - `iscsi-node-get-name`
   - `options-list-info`
   - `qtree-list-iter`
   - `system-get-version`
   - `lun-get-iter`
   - `snapshot-get-iter`
   - `perf-object-get-instances`
   - `volume-get-iter`
   - `volume-move-get-iter`
   - `volume-move-start`

   For example, to enable volume offline, run the following command:

```
security login role create -role Role-Name -access all -cmddirname "volume offline" -vserver
 <cluster_name>
```

3. Create a user based on the role you create.

   Give the user access to the `ssh` and `ontapi` applications. For example:

   ```
   security login create -role Role-Name -username User-Name -application ontapi -authmethod password
   ```

Pure Storage

Workload Optimization Manager supports management of the following Pure Storage technologies:

- FlashArray//C
- FlashArray//X

The following technologies are not supported:

- FlashBlade

Because of the improved performance of Pure Storage arrays, Workload Optimization Manager intelligently moves more demanding workloads to Flash-based data stores. Workload Optimization Manager analysis is also able to incorporate Pure Storage de-duplication and compression when recommending actions.

Prerequisites

- Pure Storage F-series and M-series arrays running Purity 5.3.6 and 6.4.4 (Pure API 1.6)
- A service account Workload Optimization Manager can use to connect to the FlashArray

  This account needs privileges to execute commands through the Pure Storage API – Typically the default `pureuser` administrative account.

Adding Pure Storage targets

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **Pure Storage**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
   - Hostname or IP address

     Specify the host name or IP address of the Pure Storage FlashArray.
   - Username

     Specify the username of the service account that Workload Optimization Manager uses to connect to the FlashArray.

     The username must not contain the domain. For example, `Username=jjsmith` is correct, while `Username=myDomain\jjsmith` fails to validate.
   - Password

     Specify the password of the service account that Workload Optimization Manager uses to connect to the FlashArray.
   - Connect using HTTPS

     If this option is selected, Workload Optimization Manager uses SSL to connect to the Pure target.

     Most Pure Storage installations do not accept insecure connections. If you receive an error when adding the target with secure connections disabled, try re-adding with this option enabled.

     This option is selected by default.
   - Certificate authority

     Specify the root certificate for validating the SSL connection to the FlashArray.

Entity mapping

After validating your targets, Workload Optimization Manager updates the supply chain with the entities that it discovered. The following table describes the entity mapping between the target and Workload Optimization Manager.

Pure	Workload Optimization Manager
Volume	Storage
Shelf Array	Disk Array
Controller	Storage Controller

Storage targets (storage controllers) add Storage Controller and Disk Array entities to the supply chain. Disk Array entities then host Storage entities (datastores).

Monitored resources

Workload Optimization Manager monitors the following resources:

- Storage
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

    NOTE:
    When it generates actions, Workload Optimization Manager does not consider IOPS throttling that it discovers on storage entities. Analysis uses the IOPS it discovers on Logical Pool or Disk Array entities.
  - Latency

    Latency is the measurement of storage latency.
- Disk array
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage provisioned

    Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - Latency

    Latency is the measurement of storage latency.
- Storage controller

NOTE:
Not all targets of the same type provide all possible commodities. For example, some storage controllers do not expose CPU activity. When a metric is not collected, its chart in the user interface will not display data.

  - CPU

    CPU is the measurement of CPU that is reserved or in use.
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.

    The storage allocated to a storage controller is the total of all the physical space available to aggregates managed by that storage controller.

Actions

Workload Optimization Manager supports the following actions:

- Storage
  - Resize up

    This action can only be executed outside Workload Optimization Manager.
- Storage controller
  - Provision

    This action can only be executed outside Workload Optimization Manager.
- Disk array
  - Provision

    This action can only be executed outside Workload Optimization Manager.

Pure Storage assigns all the disks managed by a storage controller to a single array, with a fixed form-factor. There are no actions to perform for an array. For example, there is no action to move a disk array from one storage controller to another. Likewise, there are no actions to move or provision volumes because of the fixed form-factor.

Virtual desktop infrastructure targets

A virtual desktop infrastructure target is a service that delivers digital workspaces to multiple application users through a single portal. When you connect Workload Optimization Manager to VDI targets, it monitors the performance and resource consumption of the business users utilizing the virtual desktops and the virtual desktop infrastructure itself. Workload Optimization Manager then performs Application Resource Management (ARM) using the enhanced data to assure application performance while also utilizing resources as efficiently as possible.

As part of this process, Workload Optimization Manager will use this information to provide deeper insight into the state of the hardware, and information related to the users and virtual desktops running in your environment.

VMware Horizon

NOTE:
After reviewing telemetry data and conducting interviews with customers to determine use of this capability, this target was deemed appropriate to deprecate. If you replaced this technology with something that is currently not supported, submit an idea at the Cloud Management and AIOps Idea Portal. The last Workload Optimization Manager version that will contain this target is 3.13.0.

VMware Horizon provides end users access to all of their virtual desktops, applications, and online services through a single digital workspace.

With the information provided by Horizon, Workload Optimization Manager efficiently utilizes the VDI infrastructure to avoid inappropriately placed or sized desktops, ensure users are matched to desktop images that meet their demand, while gaining valuable insight into the interconnected nature of your environment.

Prerequisites

- Horizon 7.0 or higher
- The version of the vCenter target associated to the Horizon target must be 5.0+

Adding Horizon targets

NOTE:
For each Horizon target, add only one connection server servicing that site. Workload Optimization Manager will use the default port of 443. Adding multiple connection servers for the same site results in inconsistent and duplicate actions.

1. Click **Settings > Target Configuration**.
2. On the Target configuration page, click **Add Target**.
3. On the Select target page, click **Horizon**.
4. In the side panel, review the connection requirements and then click **Connect Target**.
5. Configure the following settings:
    - IP address of the Horizon Connection Server

      Specify the IP address of the Horizon connection server.
    - Username

      Specify the username of the account Workload Optimization Manager uses to connect to the target.

      Specify the username in the User Principal Name (UPN) format. For example, `user.name@domain.com`.

      This account must be an Administrator (Read-Only) account.
    - Password

      Specify the password of the account Workload Optimization Manager uses to connect to the target.
    - vCenter username

      Specify the username for the user account Workload Optimization Manager uses to connect to the vCenter target related to Horizon.

      Leave this field blank if the Horizon and vCenter credentials are identical.
    - vCenter password

      Password for the user account Workload Optimization Manager uses to connect to the vCenter target related to Horizon.

      Leave this field blank if the Horizon and vCenter credentials are identical.
    - LDAP server domain name

      Specify the fully qualified domain name of the LDAP server Workload Optimization Manager uses in conjunction with the Horizon target.

      By default, Workload Optimization Manager will do a lookup based on the domain entered in the Username and select a healthy DNS server from the response.
    - LDAP server port

      Specify the port to use for the LDAP connection. By default, Workload Optimization Manager uses port 389 for unsecure connections, and port 636 for secure connections.
    - Connect securely to LDAP server using SSL

      If you select this option, Workload Optimization Manager uses SSL to communicate to the LDAP server.

This option is selected by default.

- LDAP server username

  Specify the username in the User Principal Name (UPN) format. For example, `user.name@domain.com`.

- LDAP server password

  Specify the password Workload Optimization Manager uses to connect to the LDAP server.

Monitored resources

Workload Optimization Manager monitors the following commodities of the Horizon target:

- Business User
  - ImageCPU

    ImageCPU is the measurement of CPU that is utilized by a user's desktop images, expressed as a percentage of CPU capacity.
  - ImageMem

    ImageMem is the measurement of memory that is utilized by a user's desktop images, expressed as a percentage of memory capacity.
  - ImageStorage

    ImageStorage is the measurement of storage that is utilized by a user's desktop images, expressed as a percentage of storage capacity.
- Desktop Pool
  - Pool CPU

    Pool CPU is the measurement of CPU that is utilized by active sessions in a desktop pool.
  - Pool memory

    Pool memory is the measurement of memory that is utilized by active sessions in a desktop pool.
  - Pool storage

    Pool storage is the measurement of storage that is utilized by active sessions in a desktop pool.
  - Active sessions

    Active sessions is the number of active sessions in a desktop pool, expressed as a percentage of the pool's capacity.
  - Total sessions

    Total sessions is the number of active and disconnected (non-terminated) sessions in a desktop pool, expressed as a percentage of the pool's capacity.
- View Pod
  - Active sessions

    Active sessions is the number of active sessions in a desktop pool, expressed as a percentage of the pool's capacity.
  - Total sessions

    Total sessions is the number of active and disconnected (non-terminated) sessions in a desktop pool, expressed as a percentage of the pool's capacity.

Actions

Workload Optimization Manager recommends actions for the various entities of the Horizon target as follows:

- Desktop Pool
  - Workload Optimization Manager does not recommend actions to perform on a desktop pool. Instead, it recommends actions to perform on the business users running active sessions in the pool.
- View Pod
  - Workload Optimization Manager does not recommend actions to perform on a view pod. Instead, it recommends actions to perform on the business users running active sessions in the view pod.
- Business User
  - Move Business User due to image congestion
  - Move Business User due to desktop pool congestion

Under some circumstances, you can have Business Users who need large-scale desktop images. This appears as users with high utilization of the image resources. In this case, Workload Optimization Manager can recommend moving the Business Users to a different desktop pool that serves up large-scale images.

To support moving Business Users, you must create a placement policy that merges desktop pools. Be sure to merge only desktop pools that are *similarly configured* – they should run the same operating system and applications, and differ only in allocated memory and/or CPU.

To merge desktop pools:

1. Create a new placement policy.
2. Choose **Merge** as the policy type.
3. For the consumer type to merge, choose **Desktop Pool**.
4. Choose the pools that you want to merge.
5. Save the policy.

For more information, see Creating Placement Policies *(on page 1044)*.

Prometheus

Prometheus is an open-source systems monitoring and alerting toolkit. Workload Optimization Manager collects application metrics from Prometheus through the **Prometurbo** agent and then uses these metrics to make decisions about application scaling, placement and optimization.

Deploy Prometurbo through OperatorHub or YAML. After the deployment, Prometurbo automatically adds itself as a Prometheus target in the Workload Optimization Manager user interface, in **Settings > Target Configuration**. You do not need to configure any settings in the user interface.

About Prometurbo

Prometurbo is an Application Performance Management (APM) agent that discovers entities and their corresponding metrics from Prometheus. Prometurbo leverages the Workload Optimization Manager analytics engine to provide visibility and drive actions across the entire application stack, thereby assuring performance, efficiency, and compliance.

Prometurbo deployment requirements

- Review the general requirements *(on page 513)*.

Prometurbo deployment method

Deploy Prometurbo through the OperatorHub in Red Hat OpenShift and enable metrics collection.

- Deploying Prometurbo Through OperatorHub *(on page 515)*
- Deploying Prometurbo Through YAML *(on page 517)*
- Enabling Metrics Collection for Prometurbo *(on page 519)*

Prometurbo output

Prometurbo outputs application metrics in JSON format. These metrics are served through the REST API's `/metrics` endpoint.

```
type Topology struct {
 Version    string       `json:"version"`
 Updatetime int64        `json:"updateTime"`
 Scope      string       `json:"scope"`
 Source     string       `json:"source"`
 Entities   []*DIFEntity `json:"topology"`
}
```

```
type DIFEntity struct {
 UID                string                    `json:"uniqueId"`
 Type               string                    `json:"type"`
 Name               string                    `json:"name"`
 HostedOn           *DIFHostedOn              `json:"hostedOn"`
 MatchingIdentifiers *DIFMatchingIdentifiers   `json:"matchIdentifiers"`
 PartOf             []*DIFPartOf              `json:"partOf"`
 Metrics            map[string][]*DIFMetricVal `json:"metrics"`
 partOfSet          set.Set
 hostTypeSet        set.Set
}
```

The JSON format is defined here.

Prometurbo deployment requirements

Minimum requirements

Requirement	Details
Container platform version	The following versions are supported: ▪ Kubernetes 1.27 up to the latest supported GA version, including (but not limited to) Amazon Elastic Kubernetes Service (EKS), Azure Kubernetes Service (AKS), Google Kubernetes Engine (GKE), IBM Cloud Kubernetes Service (IKS), and Rancher ▪ Red Hat supported GA versions of Red Hat OpenShift 4.1x, including (but not limited to) Red Hat OpenShift Kubernetes Service (ROKS), Red Hat OpenShift on AWS (ROSA), Red Hat OpenShift on Azure (ARO), and Red Hat OpenShift on Google
Workload Optimization Manager instance and credentials	Your Workload Optimization Manager instance must run with a valid trial or premium license, and should be up and running.

Requirement	Details
	Prometurbo connects to the instance through the username and password for a user account that you set up in the Workload Optimization Manager user interface.
Prometurbo image repository	The node to which Prometurbo deploys must have internet access to pull images from IBM Container Registry (`icr.io`). You can configure a private repository if you do not want to pull from IBM Container Registry. For more information, see this topic *(on page 514)*.
Network	▪ The Prometurbo pod requires access to the `apis-erver` and the `kubelet` on every node. The `kubelet` network is `https + port=10250` (default port). ▪ Prometurbo requires HTTPS/TCP port 443 to communicate with Workload Optimization Manager. Prometurbo uses WebSocket Secure (WSS) over HTTPS with TLS 1.2+. ▪ If you use a proxy server to establish a connection to Workload Optimization Manager, allow WebSocket Secure (WSS) protocol communication through the proxy.

Prometurbo resources

Prometurbo can run as a single pod deployment or can be deployed through an operator with the following resources:

- Namespace or project. The default is `turbo`.
- Service account
- Cluster role
- Cluster role binding
- Custom Resource Definition (CRD)
- Operator
- `ConfigMap` with information to connect to Workload Optimization Manager
- Prometurbo `deployment` (Operator, OperatorHub)

The following resource is optional:

- Secret

By default, Prometurbo deploys without limits or requests.

Prometurbo image repository

The node to which Prometurbo deploys must have internet access to pull the Prometurbo images from IBM Container Registry. You can configure a private repository if you prefer to pull from this repository.

Prometurbo container images

The following list describes the different Prometurbo images.

- Prometurbo (`prometurbo`)

  This is the main image and is pulled from `icr.io/cpopen/prometurbo`.

- Prometurbo operator (`prometurbo-operator`)

  This image is pulled for `icr.io/cpopen/prometurbo-operator:<version>`.

Working with a private repository

You can configure a private repository to store the Prometurbo images in your environment.

If you are using a private repository, be aware that IBM Container Registry provides multi-architecture container images. Your private repository needs to support loading these images. For example, multi-architecture container images are supported in Artifactory version 7.2+, but not in Artifactory version 6. Optionally, pull only the architecture version that you need. Specify the `--platform` parameter with `docker pull` and use `docker inspect` to confirm the architecture.

Deploying Prometurbo through OperatorHub

You can deploy Prometurbo by using the certified Prometurbo operator in Red Hat OpenShift OperatorHub and configuring a custom resource that creates a single instance of Prometurbo.

Deployment requirements

Review the general requirements *(on page 513)*.

Task overview

To deploy Prometurbo, perform the following tasks:

1. Install the Prometurbo operator through OperatorHub in Red Hat OpenShift.
2. Create a Prometurbo instance.
3. Validate the deployment.

Installing the Prometurbo operator through OperatorHub in Red Hat OpenShift

1. Log in to your Red Hat OpenShift console and navigate to **Home > Projects**.
2. Create a project (namespace) for your Prometurbo deployment. For example, the following YAML file creates a `turbo` namespace:

   ```
   apiVersion: v1
   kind: Namespace
   metadata:
     name: turbo
   ```

3. Navigate to **Operators > OperatorHub** and select the project that you created.
4. Search for Prometurbo in the search bar.
5. Select **Certified Prometurbo Operator**.
6. Select the `stable` channel and then install the operator.

   After the installation, the following resources are created.

Name	Kind	Status	API version
prometurbo-operator.vx.x.x	`ClusterServiceVersion`	Created	`operators.coreos.com/v1alpha1`
prometurbos.charts.helm.k8s.io	`CustomResourceDefinition`	Present	`apiextensions.k8s.io/v1`
prometurbo-operator	`ServiceAccount`	Present	`core/v1`

Name	Kind	Status	API version
prometurbo-operator.vx-.x.x-xxxxxxxx	ClusterRole	Created	rbac.authorization.k8s.io/v1
prometurbo-operator.vx-.x.x-xxxxxxxx	ClusterRoleBinding	Created	rbac.authorization.k8s.io/v1

NOTE:
A cluster role is created for `prometurbo-operator` to grant it permissions to create the necessary cluster roles for Prometurbo instances.

Creating a Prometurbo instance

1. Navigate to **Operators > Installed Operators**.
2. Select the Prometurbo operator and then click **Create Instance**. Follow the prompts to configure the instance.

   The following table lists the configuration parameters for `prometurbo`:

Parameter	Description	Default value
image.prometurboRepository	The prometurbo container image repository	icr.io/cpopen/turbonomic/prometurbo
image.prometurboTag	The prometurbo container image tag	Release version, such as 8.15.0
image.turbodifRepository	The turbodif container image repository	icr.io/cpopen/turbonomic/turbodif
image.turbodifTag	The turbodif container image tag	Release version, such as 8.15.0
image.pullPolicy	Specify either IfNotPresent or Always.	IfNotPresent
targetName	A unique name that identifies this target	Prometheus
serviceAccountName	The name of the serviceAccount used by the prometurbo pod	prometurbo
roleName	The name of the clusterrole bound to serviceAccount	prometurbo
roleBinding	The name of the clusterrolebinding that binds serviceAccount to clusterrole	prometurbo-binding
restAPIConfig.opsManagerUserName	The username to log in to Workload Optimization Manager. A username is required if you do not use a secret.	None
restAPIConfig.opsManagerPassword	The password to log in to Workload Optimization Manager. A password is required if you do not use a secret.	None
serverMeta.version	Workload Optimization Manager version	Release version, such as 8.15.0
serverMeta.turboServer	Workload Optimization Manager URL	None

Parameter	Description	Default value
`args.logginglevel`	Logging level of `prometurbo`	2
`args.ignoreCommodityIfP-resent`	Specify whether to ignore merging a commodity when a commodity of the same type already exists in Workload Optimization Manager.	`false`

The following example shows a YAML resource with the Prometurbo parameters.

```
apiVersion: charts.helm.k8s.io/v1
kind: Prometurbo
metadata:
  name: prometurbo-release
  namespace: turbo
spec:
  image:
    prometurboTag: 8.15.0
    turbodifTag: 8.15.0
    pullPolicy: Always
  restAPIConfig:
    opsManagerPassword: administrator
    opsManagerUserName: mypassword
  serviceAccountName: prometurbo-xl-ember
  roleBinding: prometurbo-binding-xl-ember
  roleName: prometurbo-xl-ember
  serverMeta:
    turboServer: 'https://10.1.1.1'
    version: 8.15.0
  targetName: OCP411-FYRE-IBM
```

Validating the deployment

1. Verify that the `prometurbo` pods are up and running.

```
oc -n turbo get po | grep prometurbo
```

The following example results indicate that the pods were deployed and are currently running.

```
prometurbo-operator-6ffc566f4c-lwbcn       1/1     Running   0          5d23h
prometurbo-release-744947bb94-kqc2b        2/2     Running   0          5d20h
```

2. Verify that the proper cluster role is created for Prometurbo. Using the example YAML resource in the previous section, a `prometurbo-xl-ember` cluster role should be created, with proper rules to access the `prometheusquerymappings` and `prometheusserverconfigs` resources.
3. Open the Workload Optimization Manager user interface and navigate to **Settings > Target Configuration**. If the deployment was successful, a new Prometheus target appears in the list.

Deploying Prometurbo through YAML

This topic describes the single YAML resource that specifies the configurations for your Prometurbo deployment. Update the required parameters in this YAML resource before deploying Prometurbo.

Deployment requirements

Review the general requirements *(on page 513)*.

It is recommended that you deploy Prometurbo to the same namespace as Kubeturbo. The default namespace is `turbo`.

Task overview

To deploy Prometurbo, perform the following tasks:

1. Download the YAML resource for Prometurbo.
2. Update the YAML resource with the minimum required parameters.
3. Deploy the YAML resource to your cluster.

   After you perform this task, Prometurbo deploys to your cluster.
4. Validate the deployment.

Downloading the YAML resource for Prometurbo

To download the YAML resource for this role, run the following command:

```
curl -O
 https://raw.githubusercontent.com/turbonomic/prometurbo/master/deploy/prometurbo_yamls/prometurbo_full.y
aml
```

Updating the YAML resource with the minimum required parameters

Update the YAML resource that you downloaded. For example, you can update the YAML resource in VS Code or `vi`. Be sure to set up and record the credentials *(on page 236)* for your Workload Optimization Manager instance before performing this task.

Update the following minimum required parameters:

- `version` and `turboServer`

  ```
            "serverMeta": {
                "version": "{your_instance_version}",
                "turboServer": "{your_instance_address}"
            },
  ```

  - `version`

    Specify the version of your Workload Optimization Manager instance.

    To get the version of your Workload Optimization Manager instance, open the Workload Optimization Manager user interface and click the Help icon in the navigation menu.
  - `turboServer`

    Specify the address of your Workload Optimization Manager instance, such as `https://10.1.1.1` or `https://myinstance.com`.
- `targetName` and `targetAddress`

  ```
        "targetConfig": {
            "targetName":"{your_cluster_name}"
            "targetAddress": "{your_Prometheus_server_address}"
        },
  ```

  - `targetName`

    Specify the cluster name that will display in the Workload Optimization Manager user interface. Spaces are not allowed.
  - `targetAddress`

    Specify the address of your Prometheus server, such as `https://127.0.0.1:8081/metrics`.
- Image tag version for `prometurbo` and `turbodif`

  ```
        image: icr.io/cpopen/turbonomic/prometurbo:{your_image_tag_version}
  ```

  ```
        image: icr.io/cpopen/turbonomic/turbodif:{your_image_tag_version}
  ```

Specify the image tag version.

To determine the image tag version to specify, open the Workload Optimization Manager user interface, click the Help icon in the navigation menu, and then record the version number that displays. This version number has a corresponding image tag version. For example, if your Workload Optimization Manager instance version is 3.13.0, the corresponding image tag version is 8.15.0. For the latest Workload Optimization Manager and image tag version mapping, see this reference *(on page 302)*.

- Workload Optimization Manager credentials

  Specify the credentials (username and password) in Base64. After the deployment, Prometurbo automatically creates a secret that stores the Base64 credentials.

```
data:
  username: {Base64_username}
  password: {Base64_password}
```

The plain text username and password for the user account that you created for Prometurbo must be converted to Base64.

In Linux, you can run the following command to convert each credential to Base64.

```
echo {credential} | base64
```

NOTE:
If your credential fails to convert to Base64, it might have invalid or unsupported characters and must be changed before it can be converted to Base64 successfully.

Deploying the YAML resource to your cluster

1. Log in to your cluster through the command line.
2. Deploy the YAML resource.

```
kubectl apply -f prometurbo_full.yaml
```

After you run the command, Prometurbo deploys to your cluster.

Validating the deployment

1. Open the Workload Optimization Manager user interface, navigate to **Settings > Target Configuration**, and set the filter to **Observability**. If the deployment was successful, a new observability target appears in the list.

   If you do not see the target, there might be a deployment issue that you need to resolve.

   - To start troubleshooting an issue, review the container pod logs for Prometurbo.
   - If you need further assistance, contact your Workload Optimization Manager representative.

Enabling metrics collection for Prometurbo

Create the necessary Custom Resources (CRs) to enable the collection of performance metrics. With CRs, you can quickly update the configurations used for metric collection without needing to restart Prometurbo.

Task overview

To enable metrics collection, perform the following tasks:

1. Create the Prometurbo CRDs.
2. Create the Prometurbo CRs.

This topic describes the CRs that expose metrics for the NVIDIA DCGM, TGI, and Istio exporters.

3. Verify metrics collection.

Creating the Prometurbo CRDs

Create the required Custom Resource Definitions (CRDs). These CRDs ensure the validity of the Prometurbo custom resources (CRs) that you will create in the next task.

```
oc create -f
 https://raw.githubusercontent.com/turbonomic/turbo-metrics/main/config/crd/bases/metrics.turbonomic.io_pr
ometheusquerymappings.yaml
```

```
oc create -f
 https://raw.githubusercontent.com/turbonomic/turbo-metrics/main/config/crd/bases/metrics.turbonomic.io_pr
ometheusserverconfigs.yaml
```

Overview of Prometurbo CRs

Prometurbo requires the following custom resources (CRs):

- `PrometheusQueryMapping`

  This CR specifies how Prometurbo maps Prometheus queries to Workload Optimization Manager entities.
- `PrometheusServerConfig`

  This CR specifies configuration options for the Prometheus server.

Creating CRs to export GPU and TGI metrics

Create the required CRs to export GPU and TGI metrics. These metrics are required to enable the scaling of LLM inference workloads *(on page 598)*.

1. Deploy a Kubernetes secret that specifies the token that is used to access the Prometheus server.

   ```
   apiVersion: v1
   data:
     authorizationToken: {authorization_token}
   kind: Secret
   metadata:
     name: ocp-thanos-authorization-token
   type: Opaque
   ```

   Replace *{authorization_token}* with the token in the `prometheus-k8s-token-#####` (service-account-token) secret in the `openshift-monitoring` namespace.
2. Create the following `PrometheusQueryMapping` CRs in the namespace of your choice, preferably in the namespace where Prometurbo is deployed.
   - `PrometheusQueryMapping for NVIDIA GPU Metrics`
   - `PrometheusQueryMapping for TGI Metrics`
3. Create the `PrometheusServerConfig` CR in the namespace where `PrometheusQueryMapping` is created.

   ```
   apiVersion: metrics.turbonomic.io/v1alpha1
   kind: PrometheusServerConfig
   metadata:
     name: {Prometheus_server_name}
   spec:
     address: {Prometheus_server_address}
     bearerToken:
       secretKeyRef:
         key: authorizationToken
         name: ocp-thanos-authorization-token
   ```

```
clusters:
- identifier:
    clusterLabels: {}
    id: {cluster_ID}
  queryMappingSelector:
    matchExpressions:
    - key: mapping
      operator: In
      values:
      - nvidia-dcgm-exporter
      - text-generation-inference
```

Update the following information:

- ```
  metadata:
    name: {Prometheus_server_name}
  ```

  Specify the name of your Prometheus server.

- ```
  spec:
    address: {Prometheus_server_address}
  ```

  Specify the address of your Prometheus server, such as `https://prometheus.us-east.containers.appdomain.cloud`.

- ```
  clusters:
  - identifier:
      clusterLabels: {}
      id: {cluster_ID}
  ```

  Specify the cluster ID. To obtain the ID, run the following command:

  ```
  `kubectl -n default get svc kubernetes -ojsonpath='{.metadata.uid}'`
  ```

Creating CRs for the Istio exporter

Create the following CRs to collect metrics exposed by the Istio exporter.

- `PrometheusQueryMapping`
- `PrometheusServerConfig`

  This CR specifies the location of the Prometheus server and a label selector to exclude `jmx-tomcat` from the Prometheus query mapping resource.

Verifying metrics collection

Check the Prometurbo logs to verify that Prometurbo started collecting metrics from the Prometheus server.

The following example verifies the collection of metrics.

```
I0328 18:42:04.003329 1 provider.go:60] Discovered 4 PrometheusQueryMapping resources.
I0328 18:42:04.007689 1 provider.go:71] Discovered 2 PrometheusServerConfig resources.
I0328 18:42:04.007903 1 serverconfig.go:19] Loading PrometheusServerConfig
 turbo-community/prometheusserverconfig-emptycluster.
I0328 18:42:04.007927 1 client.go:68] Creating client for Prometheus server:
 http://prometheus.istio-system:9090
I0328 18:42:04.007935 1 serverconfig.go:36] There are 1 PrometheusQueryMapping resources in namespace
 turbo-community
I0328 18:42:04.007943 1 serverconfig.go:19] Loading PrometheusServerConfig
 turbo/prometheusserverconfig-singlecluster.
```
```

```
I0328 18:42:04.007947 1 client.go:68] Creating client for Prometheus server:
 http://prometheus.istio-system:9090
I0328 18:42:04.007950 1 serverconfig.go:36] There are 3 PrometheusQueryMapping resources in namespace
 turbo
I0328 18:42:04.008048 1 clusterconfig.go:39] Excluding turbo/jmx-tomcat.
```

Updating Prometurbo

Update the image tag version for `prometurbo` and `turbodif` whenever your Workload Optimization Manager instance updates to a new version or if you are instructed to do so.

To determine the image tag version to specify, open the Workload Optimization Manager user interface, click the Help icon in the navigation menu, and then record the version number that displays. This version number has a corresponding image tag version. For example, if your Workload Optimization Manager instance version is 3.13.0, the corresponding image tag version is 8.15.0. For the latest Workload Optimization Manager and image tag version mapping, see this reference *(on page 302)*.

Updating Prometurbo through OperatorHub

1. Log in to your Red Hat OpenShift console and navigate to **Home > Projects**.
2. Navigate to **Operators > Installed Operators**.
3. Open the Prometurbo operator for editing and update the `image.prometurboTag` and `image.turbodifTag` parameters.

Data Ingestion Framework

Data Ingestion Framework (DIF) is a Workload Optimization Manager integration that defines custom entities and metrics in your environment that are not discovered through any of the supported targets *(on page 173)*. DIF loads these custom entities and metrics to Workload Optimization Manager for analysis.

DIF uses the patented Workload Optimization Manager analysis engine to provide visibility and control across the entire application stack.

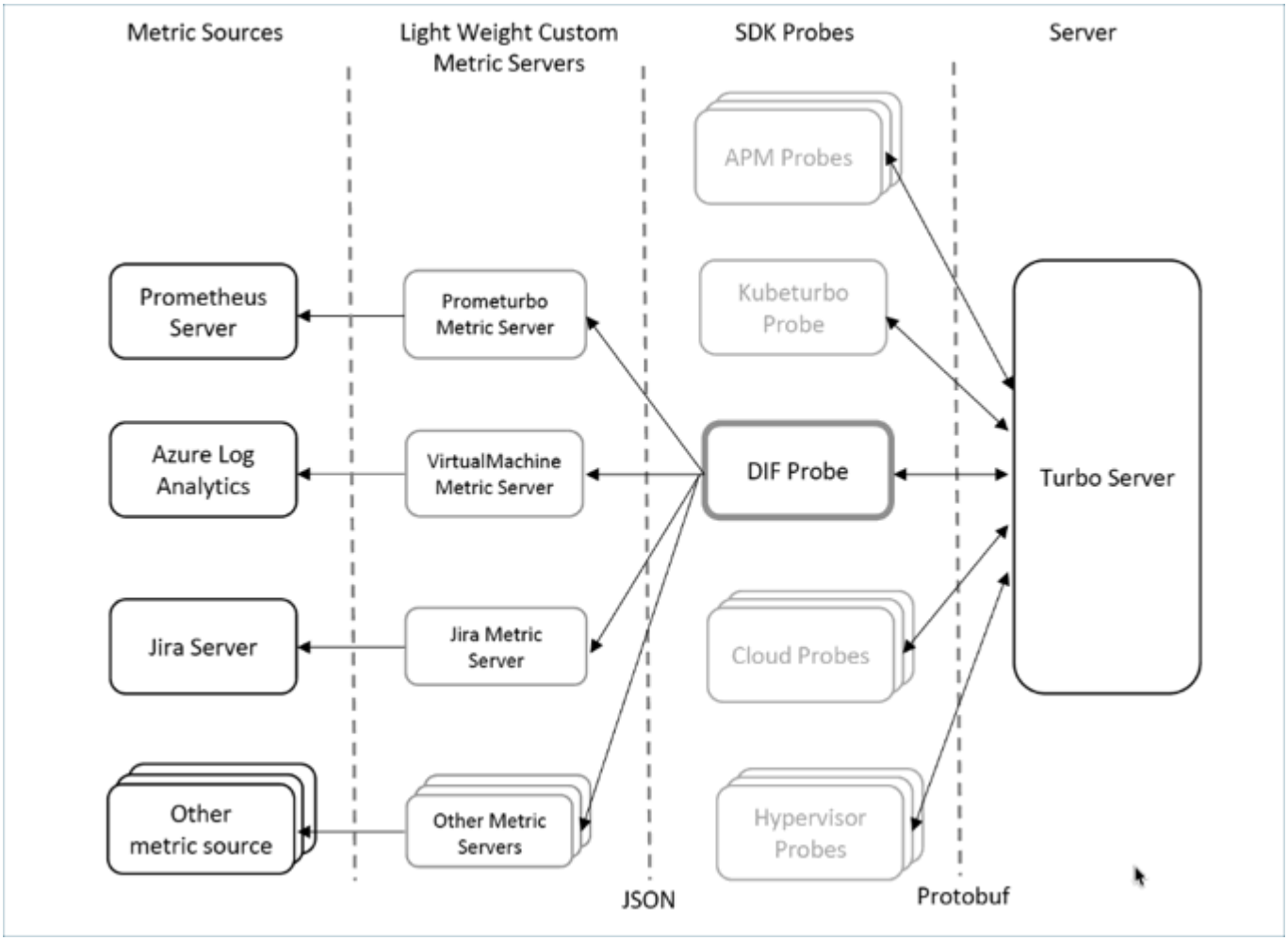
DIF consists of the following components:

- A JSON schema that declares the structure of the data that you loaded to Workload Optimization Manager
- A component that runs on the Workload Optimization Manager platform and periodically requests updated data to ingest to the platform

How DIF works

Custom entities and metrics are declared in a predefined JSON schema. DIF takes the JSON input, converts it into data structures that are known to Workload Optimization Manager, and then pushes it to Workload Optimization Manager. DIF performs supply chain and entity validations and participates in the broader mediation framework to resolve conflicts and maintain consistency.

The following diagram shows how Workload Optimization Manager uses DIF to query data from metric servers. These servers collect data from specific metric sources.



Deploying Data Ingestion Framework

This tutorial walks you through the process of deploying Data Ingestion Framework (DIF) to your environment. A successful deployment confirms that you have a valid environment for ingesting DIF data to Workload Optimization Manager.

DIF pairs a JSON schema with Workload Optimization Manager. You deploy a DIF web server and configure it as a Workload Optimization Manager target. DIF retrieves the data that your target delivers and uses it to add custom entities to the Workload Optimization Manager supply chain. This tutorial shows you how to deploy these components.

Task overview

To deploy DIF, perform the following tasks:

1. Set up a DIF web server. *(on page 524)*
2. Create the JSON data object for a custom entity that you add to the Workload Optimization Manager supply chain.
3. Configure your web server as a DIF target in Workload Optimization Manager.
4. Check the custom entity in the supply chain.

For examples of DIF JSON files and the full DIF JSON topology and entity schemas, see this topic *(on page 526)*.

Setting Up a DIF web server

A DIF web server is a data repository that can return a JSON string on request. The request must be a static URL that Workload Optimization Manager can use to retrieve DIF data at regular intervals. Be sure that the URL endpoint returns your JSON data.

For this tutorial, the JSON data is static, so you can serve it from any HTTP or HTTPS web server that Workload Optimization Manager can access.

As you set up the web server, be sure that the following requirements are met:

- Your web server must be accessible to Workload Optimization Manager directly or through a proxy.
- You must be able to create a new directory and upload a JSON file to the web server.
- Your web server must have a static address. In a real-world deployment, the static address ensures that the same web server is available for the long term.

After setting up the web server, create the `www/dif/` directory for the JSON data.

Creating the JSON data object

For this tutorial, you will add a custom Business Application named 'Hello World Tutorial' to the Workload Optimization Manager supply chain.

In the `www/dif/` directory that you created, create a file named `HelloWorld.json` and then copy the following data object to the file.

```
{
  "version": "v1",
  "updateTime": 123,
  "scope": "Tutorial",
  "source": "",
  "topology" : [
    {
      "uniqueId": "Hello_World_BusinessApp",
      "type": "businessApplication",
      "name": "Hello World Tutorial"
    }
  ]
}
```

The data object begins with basic information, such as `version`, `updateTime` (time of latest update), and other descriptors. The `updateTime` property is important to consider. In a real-world use case, you periodically update your data to capture changes in your environment. As you update your data, post the `updateTime` as well. This value might be useful for debugging if your management process stops updating the data. For this tutorial, any arbitrary value is sufficient.

> **NOTE:**
> For complete information, see the Schema Object description for Topology Object *(on page 535)*.

Specify the following properties for the supply chain topology:

- `uniqueId`

  This property identifies the custom entity among all other instances of that entity in your topology. Because the ID must be globally unique, use naming conventions that express a namespace for the topology segment that you are creating.
- `type`

  Since you will create a Business Application in this tutorial, specify `businessApplication`.

  DIF currently supports the following entity types:
  - `businessApplication`
  - `businessTransaction`

  - `service`
  - `databaseServer`
  - `application`
- `name`

  This property specifies a meaningful display name for the entity. This display name is not required to be unique. For this tutorial, specify `Hello World Tutorial`.

An entity can include other properties, such as resource utilization metrics or relationships to other entities in the supply chain topology. This tutorial does not explore any of these properties.

After you save the file, your web server must be able to serve the JSON data. To validate this functionality, run the following command:

```
curl {your_domain}/dif/HelloWorld.json
```

For example:

```
curl MyDomain.com/dif/HelloWorld.json
```

The curl output shows the data object in the JSON file.

Configuring your web server as a DIF target in Workload Optimization Manager

1. In the Workload Optimization Manager user interface, click **Settings > Target Configuration**.
2. Click **New Target > Custom**.
3. Select **Data Ingestion Framework**.
4. Configure the following settings:
   - URL

     Specify the HTTP location and file name of the JSON data for this target, such as `http://MyDomain.com/dif/HelloWorld.json`.
   - Name

     Specify a name that uniquely identifies this connection.

     This name is for display purposes only, such as `My_Custom_Application`.
   - Prometurbo version

     If this target integrates with Prometurbo to collect application performance metrics, specify the Prometurbo release version.

After adding the target, navigate to **Settings > Target Configuration**. A new target named `My_Custom_Application` appears in the list. The status for this target indicates if the validation was successful. If the validation was unsuccessful, check if the URL you specified returns the JSON data and if the data is valid for the DIF schema.

Checking the custom entity in the supply chain

Your custom Business Application appears in the supply chain within 10 minutes of adding the DIF target.

In Workload Optimization Manager, navigate to Search and select **Business Applications**. In the Search field, specify `Hello World Tutorial`. The Hello World Tutorial entity appears in the list. Click that entity to set it as your supply chain scope.



This tutorial shows the simplest possible implementation of DIF. For a real-world use case, create a larger supply chain composed of several custom entities. Consider including resource utilization metrics for these entities in the DIF data.

DIF schema and example files

DIF declares its data objects in two JSON schema files:

- dif-topology-schema.json *(on page 526)*
- dif-entity-schema.json *(on page 528)*

The following files show topology and entity objects in the context of example entities:

- Proxy Virtual Machine entity with vMem metrics *(on page 533)*
- Database Server entity hosted on a Virtual Machine *(on page 534)*

For more information about the properties in these data objects, see the following topics:

- Topology Object *(on page 535)*
- Entity Object *(on page 535)*
- Metrics Entry Object *(on page 539)*

You can also find several examples of DIF integrations in the DIF GitHub repository.

dif-topology-schema.json

```
{
  "$schema": "http://json-schema.org/draft-07/schema#",
  "$id": "http://turbonomic.com/dif-topology.json",
  "title": "DIF Topology File",
  "description": "Format for a DIF topology",
  "type": "object",
  "properties": {
    "version": {
      "description": "",
      "type": "string"
    },
    "updateTime": {
      "description": "Epoch timestamp for when files are generated",
      "type": "integer"
    },
    "topology": {
      "description": "List of topology entities",
      "type" : "array",
      "items": {
        "$ref": "#/definitions/entity"
      },
```

```
      "minItems" : 1
    }
  },
  "required": [
    "version",
    "updateTime",
    "topology"
  ],
```

```
  "definitions": {
    "entity" : {
      "description": "Format for a DIF topology entity and metrics",
      "type": "object",
      "properties": {
        "type": {
          "description": "Entity type of the entity",
          "$ref" : "dif-entity-schema.json#/definitions/entityType"
        },
        "uniqueId": {
          "description": "Unique identifier for the entity, within the scope of the target",
          "type": "string"
        },
        "name": {
          "description": "Entity name",
          "type": "string"
        },
        "matchIdentifiers": {
          "description": "Attributes used to find the entity that matches this entity",
          "type": "object",
          "properties": {
            "ipAddress": {
              "description": "IP Address of the entity used to find the matching entity",
              "type": "string"
            }
          }
        },
        "hostedOn": {
          "description": "Attributes used to find the entity that hosts this entity",
          "type": "object",
          "properties": {
            "hostType" : {
              "description": "List of entity types of the possible underlying host entity. The order in
which the entity types is specified is used to find the host entity.",
              "type": "array",
              "items": {"$ref" : "dif-entity-schema.json#/definitions/hostEntityType"}
            },
            "hostUuid": {
              "description": "Unique identifier for the host entity",
              "type": "string"
            },
            "ipAddress": {
              "description": "IP Address of the host entity",
              "type": "string"
            }
          },
          "required": [
            "hostType"
          ]
        },
        "partOf": {
          "description": "Attributes used to find all the entities that this entity is part of",
          "type": "array",
          "items": {
            "type": "object",
            "properties": {
```

```
            "entity": {
              "description": "Entity type of the parent entity ",
              "$ref": "dif-entity-schema.json#/definitions/entityType"
            },
            "uniqueId": {
              "description": "Unique identifier for the parent entity within the scope of this target",
              "type": "string"
            }
          }
        }
      },
      "metrics": {
        "description": "List of metrics and values for the entity",
        "type" : "array",
        "items": {
          "$ref": "dif-entity-schema.json#/definitions/metricsEntry"
        },
        "minItems" : 1
      }
    },
```

```
    "required": [
      "type",
      "uniqueId",
      "name"
    ]
  }
}
```

dif-entity-schema.json

```
{
  "$schema": "http://json-schema.org/draft-07/schema#",
  "$id":  "http://turbonomic.com/dif-entity.json",
  "title": "DIF Entity File",
  "description": "Format for a DIF topology entity and metrics",
  "type": "object",
  "properties": {
    "type": {
      "description": "Entity type of the entity",
      "$ref" : "#/definitions/entityType"
    },
    "uniqueId": {
      "description": "Unique identifier for the entity, within the scope of the target",
      "type": "string"
    },
    "name": {
      "description": "Entity name",
      "type": "string"
    },
    "matchIdentifiers": {
      "description": "Attributes used to find the entity that matches this entity",
      "type": "object",
      "properties": {
        "ipAddress": {
          "description": "IP Address of the entity used to find the matching entity",
          "type": "string"
        }
      }
    },
    "hostedOn": {
      "description": "Attributes used to find the entity that hosts this entity",
```

```
    "type": "object",
    "properties": {
      "hostType" : {
        "description": "List of entity types of the possible underlying host entity. The order in which
the entity types is specified is used to find the host entity.",
        "type": "array",
        "items": {"$ref" : "#/definitions/hostEntityType"}
      },
      "hostUuid": {
        "description": "Unique identifier for the host entity",
        "type": "string"
      },
      "ipAddress": {
        "description": "IP Address of the host entity",
        "type": "string"
      }
    },
    "required": [
      "hostType"
    ]
  },
  "partOf": {
    "description": "Attributes used to find all the entities that this entity is part of",
    "type": "array",
    "items": {
      "type": "object",
      "properties": {
        "entity": {
          "description": "Entity type of the parent entity ",
          "$ref": "#/definitions/entityType"
        },
        "uniqueId": {
          "description": "Unique identifier for the parent entity within the scope of this target",
          "type": "string"
        }
      }
    }
  },
```

```
  "metrics": {
    "description": "List of metrics and values for the entity",
    "type" : "array",
    "items": {
      "$ref": "#/definitions/metricsEntry"
    },
    "minItems" : 1
  }
},
"required": [
  "type",
  "uniqueId",
  "name"
],
```

```
"definitions": {
  "entityType": {
    "description": "Entity type",
    "type": "string",
    "enum": [
      "businessApplication",
      "businessTransaction",
      "service",
      "databaseServer",
      "application"
    ]
```

```
  },
  "hostEntityType": {
    "description": "Entity type of the host that underlies this entity",
    "type": "string",
    "enum": [
      "container",
      "virtualMachine"
    ]
  },
  "metricUnit" : {
    "description": "Units for different metrics",
    "type": "string",
    "enum" : ["count", "tps", "ms", "mb", "mhz", "pct"]
  },
  "metricValue": {
    "type": "object",
    "properties": {
      "average": { "type": "number" },
      "min": { "type": "number" },
      "max": { "type": "number" },
      "capacity": { "type": "number" },
      "key": { "type": "string" },
      "resizable": { "type": "boolean" },
      "unit": { "$ref": "#/definitions/metricUnit" }
    },
    "required": [
      "average", "unit"
    ]
  },
  "metricsEntry": {
    "description": "Supported metrics, TODO - add schemas for other supported metrics",
    "type": "object",
    "oneOf" : [
      { "$ref": "#/definitions/_responseTime"},
      { "$ref": "#/definitions/_transaction"},
      { "$ref": "#/definitions/_connection"},
      { "$ref": "#/definitions/_heap"},
      { "$ref": "#/definitions/_remainingGCCapacity"},
      { "$ref": "#/definitions/_kpi"},
      { "$ref": "#/definitions/_cpu"},
      { "$ref": "#/definitions/_memory"},
      { "$ref": "#/definitions/_threads"},
      { "$ref": "#/definitions/_dbCacheHitRate"},
      { "$ref": "#/definitions/_dbMem"}
    ]
  },
```

```
  "_responseTime" : {
    "description": "Specification for Response Time metrics",
    "type": "object",
    "properties": {
      "responseTime": {
        "$ref": "#/definitions/metricValue"
      }
    },
    "required": [
      "responseTime"
    ]
  },
  "_transaction" : {
    "description": "Specification for Transaction metrics",
    "type": "object",
    "properties": {
      "transaction": {
        "$ref": "#/definitions/metricValue"
```

```
      }
    },
    "required": [
      "transaction"
    ]
  },
  "_connection" : {
    "description": "Specification for Connections metrics",
    "type": "object",
    "properties": {
      "connection": {
        "$ref": "#/definitions/metricValue"
      }
    },
    "required": [
      "connection"
    ]
  },
  "_heap" : {
    "description": "Specification for Heap metrics",
    "type": "object",
    "properties": {
      "heap": {
        "$ref": "#/definitions/metricValue"
      }
    },
    "required": [
      "heap"
    ]
  },
  "_remainingGCCapacity" : {
    "description": "Specification for Collection Time metrics",
    "type": "object",
    "properties": {
      "remainingGCCapacity": {
        "$ref": "#/definitions/metricValue"
      }
    },
    "required": [
      "remainingGCCapacity"
    ]
  },
  "_threads" : {
    "description": "Specification for Threads metrics",
    "type": "object",
    "properties": {
      "threads": {
        "type": "array",
        "items": {
          "$ref": "#/definitions/metricValue"
        },
        "uniqueItems": true,
        "minItems": 1
      }
    },
    "required": [
      "threads"
    ]
  },
  "_dbCacheHitRate" : {
    "description": "Specification for DB Cache Hit Rate metrics",
    "type": "object",
    "properties": {
      "dbCacheHitRate": {
        "type": "array",
```

```
        "items": {
          "$ref": "#/definitions/metricValue"
        },
        "uniqueItems": true,
        "minItems": 1
      }
    },
    "required": [
      "dbCacheHitRate"
    ]
  },
  "_dbMem" : {
    "description": "Specification for DB Memory metrics",
    "type": "object",
    "properties": {
      "dbMem": {
        "type": "array",
        "items": {
          "$ref": "#/definitions/metricValue"
        },
        "uniqueItems": true,
        "minItems": 1
      }
    },
    "required": [
      "dbMem"
    ]
  },
  "_cpu" : {
    "description": "Specification for CPU metrics",
    "type": "object",
    "properties": {
      "cpu": {
        "$ref": "#/definitions/metricValueWithRawData"
      }
    },
    "required": ["cpu"]
  },
  "_memory" : {
    "description": "Specification for Memory metrics",
    "type": "object",
    "properties": {
      "memory": {
        "$ref": "#/definitions/metricValueWithRawData"
      }
    },
    "required": ["memory"]
  },
  "_kpi" : {
    "description": "Specification for the custom metrics",
    "type": "object",
    "properties": {
      "kpi": {
        "type": "array",
        "items": {
          "$ref": "#/definitions/metricValueWithKey"
        },
        "uniqueItems": true,
        "minItems": 1
      }
    },
    "required": ["kpi"]
  },
```

```
  "metricValueWithKey": {
```

```
      "type": "object",
      "properties": {
        "description": { "type": "string" },
        "key": { "type": "string" },
        "average": { "type": "number" },
        "capacity": { "type": "number" },
        "unit": { "$ref": "#/definitions/metricUnit" }
      },
      "required": [
        "average", "unit", "key"
      ]
    },
```

```
    "metricValueWithRawData": {
      "type": "object",
      "properties": {
        "average": { "type": "number" },
        "min": { "type": "number" },
        "max": { "type": "number" },
        "capacity": { "type": "number" },
        "unit": { "$ref": "#/definitions/metricUnit" },
        "rawData" : {
          "type": "object",
          "properties": {
            "utilization": {
              "type": "array",
              "items" : {
                "type": "object"
              },
              "minItems":1
            },
            "units:" : {"$ref": "#/definitions/metricUnit"}
          },
          "required": [
            "utilization"
          ]
        }
      },
      "required": [
        "average", "unit"
      ]
    }
  }
}
```

Proxy Virtual Machine entity with vMem metrics

```
{
  "version": "v1",
  "updateTime": 1595519486,
  "scope": "",
  "source": "",
  "topology": [
    {
      "uniqueId": "spcfq9keqj-worker-1",
      "type": "virtualMachine",
      "name": "spcfq9keqj-worker-1",
      "hostedOn": null,
      "matchIdentifiers": {
        "ipAddress": "172.23.0.5"
      },
      "partOf": [
```

```
        {
          "uniqueId": "DatabaseServer-10.10.169.38-turbonomic",
          "entity": "databaseServer"
        }
      ],
      "metrics": {
        "memory": [
          {
            "average": 1363148.8,
            "capacity": 3670016
          }
        ]
      }
    }
  ]
}
```

Database Server entity hosted on a Virtual Machine

```
{
  "version": "v1",
  "updateTime": 1595519551,
  "scope": "Prometheus",
  "source": "",
  "topology": [
    {
      "uniqueId": "DatabaseServer-10.10.169.38-turbonomic",
      "type": "databaseServer",
      "name": "DatabaseServer-10.10.169.38-turbonomic",
      "hostedOn": {
        "hostType": [
          "virtualMachine"
        ],
        "ipAddress": "10.10.169.38"
      },
      "metrics": {
        "connection": [
          {
            "average": 14,
            "capacity": 151
          }
        ],
        "dbCacheHitRate": [
          {
            "average": 100
          }
        ],
        "dbMem": [
          {
            "average": 16636512,
            "capacity": 16777216
          }
        ],
        "memory": [
          {
            "average": 16636512,
            "capacity": 16777216
          }
        ],
        "transaction": [
          {
            "average": 0.5388918827326818
          }
        ]
```

```
      }
    }
  ]
}
```

Topology object

The Topology object is an array of entities that you want to add to the overall Workload Optimization Manager supply chain of managed entities. When Workload Optimization Manager loads your DIF data, it displays the entities from your topology object in the supply chain.

The Topology object includes the following properties:

Name	Type	Description
`version`	Required, string	The current version of DIF is V1.
`scope`	Required, string	Scope is a unique identifier for the topology that you are creating. DIF can load multiple topologies from different web servers. The scope effectively declares a namespace that the probe can use to separate its processing of these different topologies.
`updateTime`	Required, integer	As a best practice, generate new data periodically to reflect changes to your environment. This property specifies the time that you generated the data.
`topology`	Required, array of entities	This property specifies the array of entities that make up your supply chain. For more information, see Entity Object *(on page 535)*.

Entity object

The Entity object describes an individual entity. You can create a standalone entity or stitch an entity into a larger supply chain topology by using relationships such as `hostedOn` or `partOf`. An entity can include specific metrics.

An entity that you create can have one of two functions in your Workload Optimization Manager supply chain:

- Topology Entity

  The entity that you create exists in the topology, and displays as a member of the supply chain. For example, assume that you want to create a business application and then create services that are members of that business application. You create these entities as Topology Entities.

- Proxy Entity

  The entity that you create serves as a placeholder in the topology data you are creating. You can assign metrics to this entity or add relationships to make it part of some other entity. You also specify the `matchIdentifiers` relationship to map this entity to an entity that exists in the environment. The match identifier is the IP address of the corresponding entity.

  When Workload Optimization Manager loads the topology, it moves your metric and relationship data from the topology file into the matched entity. The proxy entity does not appear in the Workload Optimization Manager supply chain, but the metric and relationship declarations appear in the user interface for the matched entity.

The Entity object includes the following properties:

Name	Type	Description
`type`	Required, string	The following values are valid: ▪ `businessApplication`: a Business Application *(on page 544)* entity ▪ `businessTransaction`: a Business Transaction *(on page 546)* entity ▪ `service`: a Service *(on page 549)* entity ▪ `databaseServer`: a Database Server *(on page 767)* entity ▪ `application`: an Application Component *(on page 551)* entity
`uniqueId`	Required, string	This property identifies the custom entity among all other instances of that entity in your topology. Because the ID must be globally unique, use naming conventions that express a namespace for the topology segment that you are creating.
`name`	Required, string	This property specifies a meaningful display name for the entity. This display name is not required to be unique.
`matchIdentifiers`	Optional, object	This object contains the IP address of the entity that you want to match, as follows: `{ "ipAddress" : "11.22.33.44" }` If you specify `matchIdentifiers`, then this entity is a Proxy entity. You specify the IP address of the entity that you want to modify. Then Workload Optimization Manager applies the metrics or the `partOf` specification from this proxy entity to the matching entity.
`hostedOn`	Optional, object	This object specifies the entity that is a host or provider for another entity. This object contains the following properties: ▪ (Required) `hostEntityType`: This property specifies an array of possible entity types. The probe starts with the first entity type in the list, and checks for a matching UUID or IP address. It uses the first one to succeed. This array can contain the strings, `container` and `virtualMachine`.

Name	Type	Description
		▪ (Optional) `hostUuid` This property is the unique identifier for the host. This ID can be the UUID that you declared for some other entity in your topology. Alternatively, you can use the Workload Optimization Manager REST API to get the UUID of a specific entity in the environment. If you provide a UUID, you do not need to provide an IP address. ▪ (Optional) `ipAddress` This property specifies the IP address of a host entity in your environment. If you provide an IP address, you do not need to provide a UUID for the entity.
`partOf`	Required, array of objects	This property specifies an array of entities that contains this entity. For example, one business application can contain multiple service and transaction entities. The business application would be the *container* or *parent* entity. Your entity can be contained by more than one parent. The objects that are members of this array identify the entity type and the UUID of the parent, as follows: `{ "entityType": "businessApplication", "ipAddress" : "11.22.33.44" }` The entity type can be one of: ▪ `businessApplication`: a Business Application entity ▪ `businessTransaction`: a Transaction entity ▪ `service`: a Service entity ▪ `databaseServer`: a Database Server entity ▪ `application`: an Application Component entity
`metrics`	Optional, array of `metricsEntry` objects	This property specifies a list of Metric Entry objects that define metrics and values for this entity. See Metrics Entry Object *(on page 539)*.

Commodities bought and sold by entities

Entities in a Workload Optimization Manager supply chain exist in relationship to each other. One entity can sell commodities to some entities, and it can buy commodities from other entities. A commodity is the resource that is expressed by a Metrics Entry Object *(on page 539)*.

These relationships have constraints. For example, a `businessApplication` cannot buy `dbMem` from a `businessTransaction` entity, but it can buy `dbMem` from a `databaseServer` entity.

The following table lists entity types and the commodities that they can buy and sell:

Entity type	Commodities sold	Commodities bought
`businessApplication`	▪ `kpi`	▪ Bought From: `service` – `transaction` – `responseTime` – `kpi` ▪ Bought From: `businessTransaction` – `transaction` – `responseTime` – `kpi`
`businessTransaction`	▪ `transaction` ▪ `responseTime` ▪ `kpi`	▪ Bought From: `service` – `transaction` – `responseTime` – `kpi`
`service`	▪ `transaction` ▪ `responseTime` ▪ `heap` ▪ `remainingGCCapacity` ▪ `threads` ▪ `kpi`	▪ Bought From: `application` – `transaction` – `responseTime` – `heap` – `remainingGCCapacity` – `threads` – `kpi` ▪ Bought From: `databaseServer` – `transaction` – `responseTime` – `kpi`
`application`	▪ `transaction` ▪ `responseTime` ▪ `heap` ▪ `remainingGCCapacity` ▪ `threads` ▪ `kpi`	▪ Bought From: `virtualMachine` – `cpu` – `memory` ▪ Bought From: `container` – `cpu` – `memory`
`databaseServer`	▪ `transaction` ▪ `responseTime` ▪ `connection` ▪ `dbMem`	▪ Bought From: `virtualMachine` – `cpu` – `memory`

Entity type	Commodities sold	Commodities bought
	▪ dbCacheHitRate ▪ kpi	
virtualMachine	▪ cpu ▪ memory	
container	▪ cpu ▪ memory	

Metrics entry object

The metrics entry object expresses a single metric with utilization and capacity values.

- The given metric types are all types that Workload Optimization Manager supports in its supply chain, except for kpi metrics.
- The kpi metric is for Key Performance Indicators. You can track any indicator against a capacity that you specify. Charts include this metric and show the percentage of capacity used by your entities. However, kpi metrics do not affect Workload Optimization Manager analysis or the actions it generates.
- For native metrics that you track for your entities, charts display the values. However, native metrics do not affect Workload Optimization Manager analysis or the actions it generates.
- Not all metric types apply to all entity types. For example, dbMem apply to databaseServer entities, but not to businessTransaction entities. For more information, see Commodities Bought and Sold by Entities *(on page 538)*.

Each metrics entry object is of a specific type. For each type, you express the average and capacity values in a specific metric unit. The following table describes each metric type:

Name	Metric Unit	Description
kpi	count	Key Performance Indicator (KPI) is custom metric that you can use to track entities by different criteria. For example, you can use KPI to track the number of support tickets that are open for an application. The key that you provide gives the name of the metric that you see in charts and tables in the Workload Optimization Manager user interface.
responseTime	ms	Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
transaction	tps	Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
connection	count	Connection is the measurement of database connections utilized by applications.

Name	Metric Unit	Description
heap	mb	Heap is the portion of a VM or container's memory allocated to individual applications.
remainingGCCapacity	pct	Remaining GC capacity is the measurement of Application Component uptime that is *not* spent on garbage collection (GC).
cpu	mhz	CPU is the measurement of CPU that is in use. The schema specifies a `rawData` object for this metric. The `rawData` is not used currently.
memory	mb	Memory is the measurement of memory that is in use. The schema specifies a `rawData` object for this metric. The `rawData` is not used currently.
threads	count	Threads is the measurement of thread capacity utilized by applications.
dbCacheHitRate	Percentage	DB cache hit rate is the measurement of Database Server accesses that result in cache hits, measured as a percentage of hits versus total attempts. A high cache hit rate indicates efficiency. The capacity is assumed to be 100%. You do not need to provide a capacity value.
dbMem	mb	Database memory (or DBMem) is the measurement of memory that is utilized by a Database Server.

Port configuration

To support communication between Workload Optimization Manager and your intended targets, Workload Optimization Manager must be able to reach the targets through the following ports.

Applications and databases

Target	Port
Apache Tomcat	1009
AppDynamics	443
DataDog	443
Dynatrace	443
WebSphere	8879, 8880
Instana	443

Target	Port
JBoss	9990
JVM/Java	9875
SQL Server	1434, 1433
MySQL	3306
New Relic	443
Oracle	1521
Oracle WebLogic	7001

Container platform

Target	Port
Kubernetes kubelet (with Kubeturbo)	10250 or 10255
Workload Optimization Manager (through Kubeturbo using WSS)	443

Fabric

Target	Port
HPE OneView	80, 443
Cisco UCS	80, 443

Guest OS processes

Target	Port
SNMP (Linux/UNIX)	161
WMI (Windows)	135 + dynamic ports 49152 - 65535

Hyperconverged

Target	Port
Cisco HyperFlex	433
Nutanix	9440

Hypervisors

Target	Port
Cisco PowerVM	80, 443
Microsoft Hyper-V	5985, 5986
Microsoft Hyper-V (Kerberos)	88
VMware vCenter (Monitoring)	443
VMware vCenter (Tags)	10443

Target	Port
VMware vCenter (Kerberos)	88

Orchestrator

Target	Port
Action Script	22
Flexera One	443
ServiceNow	443

Private cloud

Target	Port
Microsoft VMM	5985, 5986

Public cloud

Target	Port
All public cloud targets	443

Storage

Target	Port
EMC Storage	5988, 5989, 443
HPE 3PAR	5988, 5989, 8080
Cisco FlashSystem	7443, 443
NetApp	80, 443
Pure Storage	80, 443

Virtual desktop infrastructure

Target	Port
VMware Horizon	443

User interface reference

Cisco Workload Optimization Manager is a solution that assures application performance for any workload running in any virtualized or cloud environment.

After installing or setting up the product, you can see the results of Workload Optimization Manager analysis - actions to perform that directly improve your datacenter state — within 15 to 30 minutes.

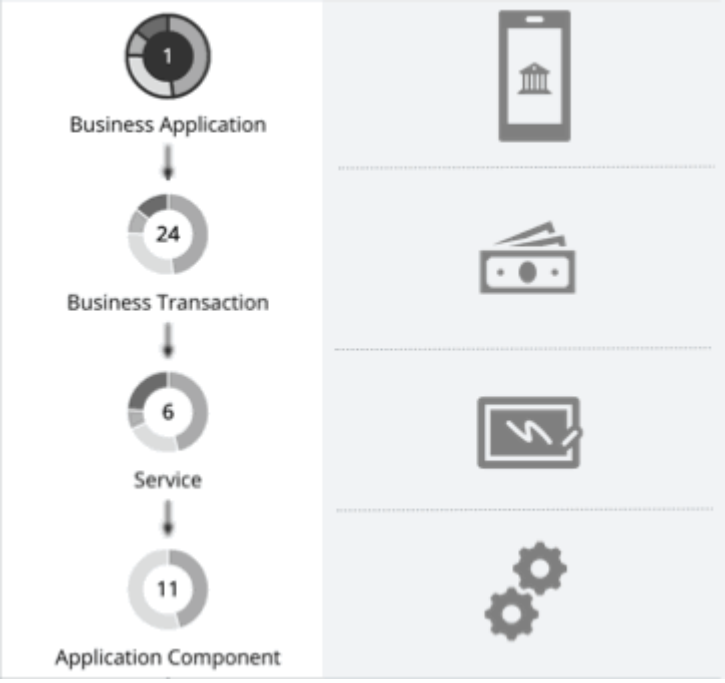
While it's true that setup is straightforward, the Workload Optimization Manager platform is rich in its coverage. Analysis takes your full stack into account, and Workload Optimization Manager recommends actions before alarms go off and the situation becomes critical. The user interface for such a product is necessarily rich with high-level information as well as fine details. You can use this interface and drill down to specific details that clarify the actions Workload Optimization Manager recommends.

Cisco also understands the need to define business rules in your environment. For example, you might need to ensure that certain applications have exclusive access to "golden" storage, while others can use less expensive resources. Or, you might want to ensure that certain workloads resize at scheduled times. Workload Optimization Manager automation supports this kind of business rule, as well as HA, affinity, discount purchase profiles, and many other rules that you expect to set up in a modern datacenter. The user interface includes tools to configure Workload Optimization Manager so that action recommendations respect the needs within your environment.

Entity types - applications

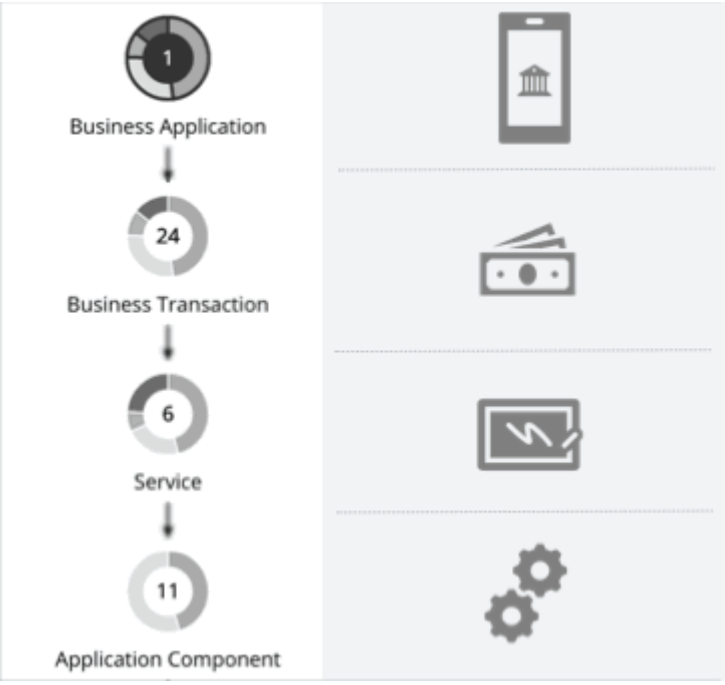
The supply chain strongly emphasizes our application-driven approach to managing your infrastructure. By showing the entity types that make up your applications at the top of the hierarchy, it is easier for you to see the health of your environment and evaluate actions from the perspective that matters - Application Performance.

For distributed applications the Workload Optimization Manager Supply Chain can include a Business Application entity, with underlying Business Transaction, Service, and Application Component entities. The application is ultimately hosted on one or more VMs or Containers.



Business application

A business application is a logical grouping of business transactions, services, application components, and other elements of the application model that work together to compose a complete application as end users would view it. For example, a mobile banking app is a business application with a business transaction that facilitates payments, a service within the business transaction that records payment information, and underlying application components (such as JVMs) that enable the service to perform its functions.



You can monitor overall performance, make resourcing decisions, and set policies in the context of your Business Applications.

Synopsis

Synopsis	
Provides:	The complete application to end users
Consumes from:	Business Transactions, Services, Application Components, Database Servers, and the underlying nodes
Discovery:	Workload Optimization Manager discovers the following: ▪ AppDynamics Business Applications ▪ Dynatrace Applications If you do not have these targets, you can create your own Business Applications using the Application Map feature. For details, see Application Map *(on page 557)*.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Response time

  Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

- Transaction

  Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

The **Response Time** and **Transaction** charts for a Business Application show average and peak/low values over time. You can gauge performance against the given SLOs. By default, Workload Optimization Manager estimates SLOs based on monitored values. You can set your own SLO values in policies.

Actions

None

Workload Optimization Manager does not recommend actions for a Business Application, but it does recommend actions for the underlying Application Components and infrastructure. The Pending Actions chart for a Business Application lists these actions, thus providing visibility into the risks that have a direct impact on the Business Application's performance.

Business application policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

None

Workload Optimization Manager does not recommend actions for a Business Application, but it does recommend actions for the underlying Application Components and infrastructure. The Pending Actions chart for a Business Application lists these actions, thus providing visibility into the risks that have a direct impact on the Business Application's performance.

Transaction SLO

Enable this SLO if you are monitoring performance through your Business Applications.

Attribute	Default setting/value
Enable Transaction SLO	Off

Attribute	Default setting/value
	Workload Optimization Manager estimates SLO based on monitored values.
Transaction SLO	None If you enable SLO, Workload Optimization Manager uses the default value of 10. You can change this to a different value.

Transaction SLO determines the upper limit for acceptable transactions per second. When the number of transactions reaches the given value, Workload Optimization Manager sets the risk index to 100%.

Response time SLO

Enable this SLO if you are monitoring performance through your Business Applications.

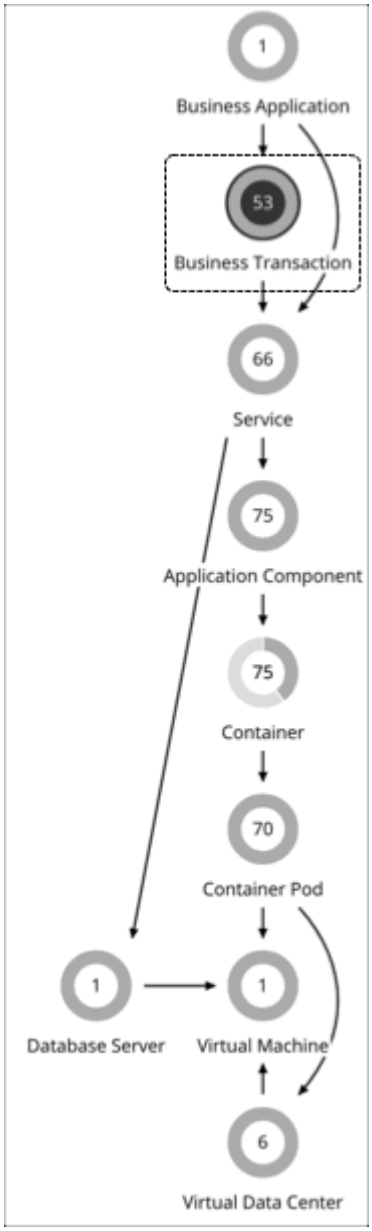
Attribute	Default setting/value
Enable Response Time SLO	Off Workload Optimization Manager estimates SLO based on monitored values.
Response Time SLO	None If you enable SLO, Workload Optimization Manager uses the default value of 2000 ms. You can change this to a different value.

Response time SLO determines the upper limit for acceptable response time (in milliseconds). If response time reaches the given value, Workload Optimization Manager sets the risk index to 100%.

Business transaction

A business transaction represents a capability within your business application that fulfills a response to a user-initiated request. Its performance directly impacts user experience. You can monitor performance as experienced by your end users in the context of business transactions.

Synopsis



Synopsis	
Provides:	Response time and transactions to Business Applications
Consumes from:	Services *(on page 549)*, Application Components *(on page 551)*, Database Servers, and the underlying nodes
Discovery:	Workload Optimization Manager discovers the following: ▪ AppDynamics Business Transactions ▪ NewRelic Key Transactions If you do not have these targets, you can create your own Business Transactions using the Application Map feature. For details, see Application Map *(on page 557)*.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Response time

  Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

- Transaction

  Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

The **Response Time** and **Transaction** charts for a Business Transaction show average and peak/low values over time. You can gauge performance against the given SLOs. By default, Workload Optimization Manager estimates SLOs based on monitored values. You can set your own SLO values in policies.

Actions

None

Workload Optimization Manager does not recommend actions for a Business Transaction, but it does recommend actions for the underlying Application Components and infrastructure. The Pending Actions chart for a Business Transaction lists these actions, thus providing visibility into the risks that have a direct impact on the Business Transaction's performance.

Business transaction policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

None

Workload Optimization Manager does not recommend actions for a Business Transaction, but it does recommend actions for the underlying Application Components and infrastructure. The Pending Actions chart for a Business Transaction lists these actions, thus providing visibility into the risks that have a direct impact on the Business Transaction's performance.

Transaction SLO

Enable this SLO if you are monitoring performance through your Business Transactions.

Attribute	Default setting/value
Enable Transaction SLO	Off Workload Optimization Manager estimates SLO based on monitored values.
Transaction SLO	None If you enable SLO, Workload Optimization Manager uses the default value of 10. You can change this to a different value.

Transaction SLO determines the upper limit for acceptable transactions per second. When the number of transactions reaches the given value, Workload Optimization Manager sets the risk index to 100%.

Response time SLO

Enable this SLO if you are monitoring performance through your Business Transactions.

Attribute	Default setting/value
Enable Response Time SLO	Off

Attribute	Default setting/value
	Workload Optimization Manager estimates SLO based on monitored values.
Response Time SLO	None If you enable SLO, Workload Optimization Manager uses the default value of 2000 ms. You can change this to a different value.

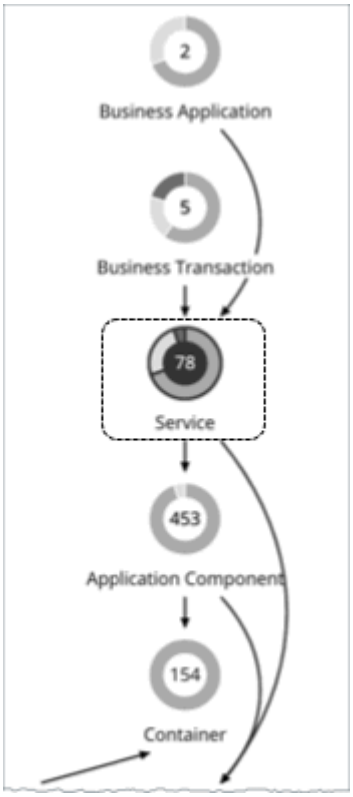
Response time SLO determines the upper limit for acceptable response time (in milliseconds). If response time reaches the given value, Workload Optimization Manager sets the risk index to 100%.

Service

A service in the supply chain represents one or several application components that perform a defined, measurable function as part of an internal or user-initiated request. Its performance is key to understanding application performance, but only indirectly impacts user experience. You can measure performance as experienced internal to the business application in the context of services.

This topic discusses Services discovered through APM targets and user-defined application maps. For details about Services discovered through container platform targets, see this topic *(on page 569)*.

Synopsis



Synopsis	
Provides:	Response time and transactions to Business Transactions *(on page 546)* and Business Applications
Consumes from:	Application Components, Database Servers, and the underlying nodes

Synopsis	
Discovery:	Workload Optimization Manager discovers the following: ▪ AppDynamics Tiers ▪ Dynatrace Services ▪ Instana Services ▪ NewRelic APM Applications / NewRelic Services (New Relic ONE) **NOTE:** If you do not have an APM target, you can create your own Services using the Application Maps feature. For details, see Application Map *(on page 557)*.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Response time

  Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

- Transaction

  Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

The **Response Time** and **Transaction** charts for a Service show average and peak/low values over time. You can gauge performance against the given SLOs. By default, Workload Optimization Manager estimates SLOs based on monitored values. You can set your own SLO values in policies.

Actions

None

Workload Optimization Manager does not recommend actions for services discovered through APM targets, but it does recommend actions for the underlying Application Components and nodes. The Pending Actions chart for services list these actions, thus providing visibility into the risks that have a direct impact on their performance.

Service policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

This topic discusses policy settings for Services discovered through APM targets. For Services discovered through container platform targets, a different set of policy settings apply. For details, see this topic *(on page 570)*.

Automation workflow

Workload Optimization Manager does not recommend actions for services discovered through APM targets, but it does recommend actions for the underlying Application Components and nodes. The Pending Actions chart for services list these actions, thus providing visibility into the risks that have a direct impact on their performance.

Response time SLO

Enable this SLO if you are monitoring performance through Services.

Attribute	Default setting/value
Enable Response Time SLO	Off

Attribute	Default setting/value
	Workload Optimization Manager estimates SLO based on monitored values.
Response Time SLO	None If you enable SLO, Workload Optimization Manager uses the default value of 2000 ms. You can change this to a different value.

Response time SLO determines the upper limit for acceptable response time (in milliseconds). If response time reaches the given value, Workload Optimization Manager sets the risk index to 100%.

Transaction SLO

Enable this SLO if you are monitoring performance through Services.

Attribute	Default setting/value
Enable Transaction SLO	Off Workload Optimization Manager estimates SLO based on monitored values.
Transaction SLO	None If you enable SLO, Workload Optimization Manager uses the default value of 10. You can change this to a different value.

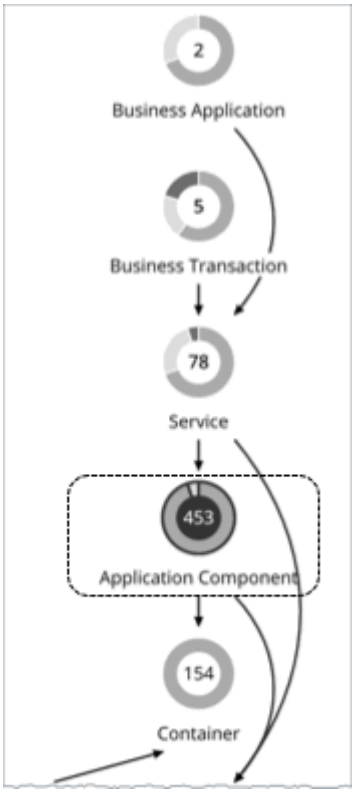
Transaction SLO determines the upper limit for acceptable transactions per second. When the number of transactions reaches the given value, Workload Optimization Manager sets the risk index to 100%.

Application component

An Application Component is a software component, application code, or a unit of processing within a Service *(on page 549)* that consumes resources to enable it to perform its function for the Business Application *(on page 544)*. For example, Apache Tomcat is a Java Servlet container that hosts a range of Java applications on the web.

Workload Optimization Manager can recommend actions to adjust the amount of resources available to Application Components.

Synopsis



Synopsis	
Provides:	▪ Response Time and Transactions to Services, Business Transactions *(on page 546)*, and Business Applications ▪ Response Time, Transactions, Heap, Remaining GC Capacity, and Threads to end users
Consumes:	Compute resources from nodes
Discovery:	Workload Optimization Manager discovers the following: ▪ Apache Tomcat ▪ AppDynamics Nodes ▪ Dynatrace Processes ▪ NewRelic APM Application Instances ▪ SNMP ▪ WMI ▪ Data Ingestion Framework *(on page 522)* metrics for container platform environments

Monitored resources

Workload Optimization Manager monitors the following resources:

NOTE:
The exact resources that are monitored will differ based on application type. This list includes all of the resources that you may see.

- Virtual CPU (vCPU)

  Virtual CPU is the measurement of CPU that is in use.
- Virtual memory (vMem)

  Virtual memory (vMem) is the measurement of memory that is in use.
- Transaction

  Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
- Heap

  Heap is the portion of a VM or container's memory allocated to individual applications.
- Response time

  Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).
- Connection

  Connection is the measurement of database connections utilized by applications.
- Remaining GC capacity

  Remaining GC capacity is the measurement of Application Component uptime that is *not* spent on garbage collection (GC).
- Threads

  Threads is the measurement of thread capacity utilized by applications.

The charts for an Application Component show average and peak/low values over time. You can gauge performance against the given SLOs. By default, Workload Optimization Manager does not enable SLOs in the default policy for Application Components. It estimates SLOs based on monitored values, but does not use these values in its analysis.

NOTE:
In container platform environments, SLOs defined in a service policy override any SLOs set in the associated Application Components to prevent conflicts. In addition, the Response Time and Transaction charts for Application Components will show SLOs specified in the service policy. For more information, see this topic *(on page 570)*.

Actions

Workload Optimization Manager supports the following actions:

Resize

Resize the following resources to maintain performance:

- Thread pool

  Workload Optimization Manager generates thread pool resize actions. These actions are recommend-only and can only be executed outside Workload Optimization Manager.
- Connections

  Workload Optimization Manager uses connection data to generate memory resize actions for on-prem Database Servers.
- Heap

Workload Optimization Manager generates Heap resize actions if an Application Component provides Heap and Remaining GC Capacity, and the underlying VM or container provides VMem. These actions are recommend-only and can only be executed outside Workload Optimization Manager.

NOTE:
Remaining GC capacity is the measurement of Application Component uptime that is *not* spent on garbage collection (GC).

The resources that Workload Optimization Manager can resize depend on the processes that it discovers from your Applications and Databases targets. Refer to the topic for a specific target to see a list of resources that can be resized.

Application component policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

For details about Application Component actions, see Application Component Actions. *(on page 553)*

Action	Default mode
Resize connections (up or down)	Recommend
Resize heap (up or down)	Recommend
Resize thread pool (up or down)	Recommend

You can use Action Scripts *(on page 1070)* for action orchestration. Third-party orchestrators (such as ServiceNow) are not supported.

Response time SLO

Enable this SLO to monitor the performance of your Application Components.

NOTE:
In container platform environments, SLOs defined in a service policy override any SLOs set in the associated Application Components to prevent conflicts. In addition, the Response Time and Transaction charts for Application Components will show SLOs specified in the service policy. For more information, see this topic *(on page 570)*.

Attribute	Default setting/value
Enable Response Time SLO	Off Workload Optimization Manager estimates SLO based on monitored values.
Response Time SLO	None If you enable SLO, Workload Optimization Manager uses the default value of 2000 ms. You can change this to a different value.

Response time SLO determines the upper limit for acceptable response time (in milliseconds). If response time reaches the given value, Workload Optimization Manager sets the risk index to 100%.

Transaction SLO

Enable this SLO to monitor the performance of your Application Components.

> **NOTE:**
> In container platform environments, SLOs defined in a service policy override any SLOs set in the associated Application Components to prevent conflicts. In addition, the Response Time and Transaction charts for Application Components will show SLOs specified in the service policy. For more information, see this topic *(on page 570)*.

Attribute	Default setting/value
Enable Transaction SLO	Off Workload Optimization Manager estimates SLO based on monitored values.
Transaction SLO	None If you enable SLO, Workload Optimization Manager uses the default value of 10. You can change this to a different value.

Transaction SLO determines the upper limit for acceptable transactions per second. When the number of transactions reaches the given value, Workload Optimization Manager sets the risk index to 100%.

Heap utilization

The Heap utilization that you set here specifies the percentage of the existing capacity that Workload Optimization Manager will consider to be 100% of capacity. For example, a value of 80 means that Workload Optimization Manager considers 80% utilization to be 100% of capacity.

Attribute	Default value
Heap Utilization (%)	80

Workload Optimization Manager uses Heap utilization and Remaining GC Capacity (the percentage of CPU time *not* spent on garbage collection) when making scaling decisions. Assume Heap utilization is at 80%, which is 100% of capacity. However, if Remaining GC Capacity is at least 90% (in other words, CPU time spent on garbage collection is only 10% or less), an 80% Heap utilization does not indicate a shortage after all. As a result, Workload Optimization Manager will not recommend Heap scaling.

If Heap utilization is low and Remaining GC Capacity is high, Workload Optimization Manager will recommend resizing down Heap. If the opposite is true, then Workload Optimization Manager will recommend resizing up Heap.

Heap scaling increment

This increment specifies how many units to add or subtract when scaling Heap for an application component.

Attribute	Default value
Heap Scaling Increment	128 MB

Do not set the increment value to be lower than what is necessary for the Application Component to operate. If the increment is too low, then it's possible there would be insufficient Heap for the Application Component to operate. When reducing allocation, Workload Optimization Manager will not leave an Application Component with less than the increment value. For example, if you use the default 128, then Workload Optimization Manager cannot reduce the Heap to less than 128 MB.

Aggressiveness and observation periods

Workload Optimization Manager uses a percentile of utilization over the specified observation period. This gives sustained utilization and ignores short-lived bursts.

Workload Optimization Manager uses these settings to calculate utilization percentiles for vCPU, vMEM, Heap, and Garbage Collection. It then recommends actions to improve utilization based on the observed values for a given time period.

- Aggressiveness

Attribute	Default value
Aggressiveness	95th Percentile

When evaluating performance, Workload Optimization Manager considers resource utilization as a percentage of capacity. The utilization drives actions to scale the available capacity either up or down. To measure utilization, the analysis considers a given utilization percentile. For example, assume a 95th percentile. The percentile utilization is the highest value that 95% of the observed samples fall below. Compare that to average utilization, which is the average of *all* the observed samples.

Using a percentile, Workload Optimization Manager can recommend more relevant actions. This is important in the cloud, so that analysis can better exploit the elasticity of the cloud. For scheduled policies, the more relevant actions will tend to remain viable when their execution is put off to a later time.

For example, consider decisions to reduce capacity. Without using a percentile, Workload Optimization Manager never resizes below the recognized peak utilization. Assume utilization peaked at 100% just once. Without the benefit of a percentile, Workload Optimization Manager will not reduce resources for that Application Component.

With **Aggressiveness**, instead of using the single highest utilization value, Workload Optimization Manager uses the percentile you set. For the above example, assume a single burst to 100%, but for 95% of the samples, utilization never exceeded 50%. If you set **Aggressiveness** to 95th Percentile, then Workload Optimization Manager can see this as an opportunity to reduce resource allocation.

In summary, a percentile evaluates the sustained resource utilization, and ignores bursts that occurred for a small portion of the samples. You can think of this as aggressiveness of resizing, as follows:

  - 99th Percentile – More performance. Recommended for critical Application Components that need maximum guaranteed performance at all times, or those that need to tolerate sudden and previously unseen spikes in utilization, even though sustained utilization is low.
  - 95th Percentile (Default) – The recommended setting to achieve maximum performance and savings. This assures performance while avoiding reactive peak sizing due to transient spikes, thus allowing you to take advantage of the elastic ability of the cloud.
  - 90th Percentile – More efficiency. Recommended for Application Components that can stand higher resource utilization.

By default, Workload Optimization Manager uses samples from the last 14 days. Use the **Max Observation Period** setting to adjust the number of days. To ensure that there are enough samples to analyze and drive resize actions, set the **Min Observation Period**.

- Max observation period

Attribute	Default value
Max Observation Period	Last 14 Days

To refine the calculation of resource utilization percentiles, you can set the sample time to consider. Workload Optimization Manager uses historical data from up to the number of days that you specify as a sample period. If the Application Component has fewer days' data then it uses all of the stored historical data.

You can make the following settings:

  - Less Elastic – Last 30 Days
  - Recommended – Last 14 Days
  - More Elastic – Last 7 Days or Last 3 Days

Workload Optimization Manager recommends an observation period of 14 days so it can recommend resize actions more often. Since Application Component resizing is minimally disruptive, resizing often should not introduce any noticeable performance risks.

- Min observation period

Attribute	Default value
Min Observation Period	None

This setting ensures historical data for a minimum number of days before Workload Optimization Manager will generate an action based on the percentile set in **Aggressiveness**. This ensures a minimum set of data points before it generates the action.

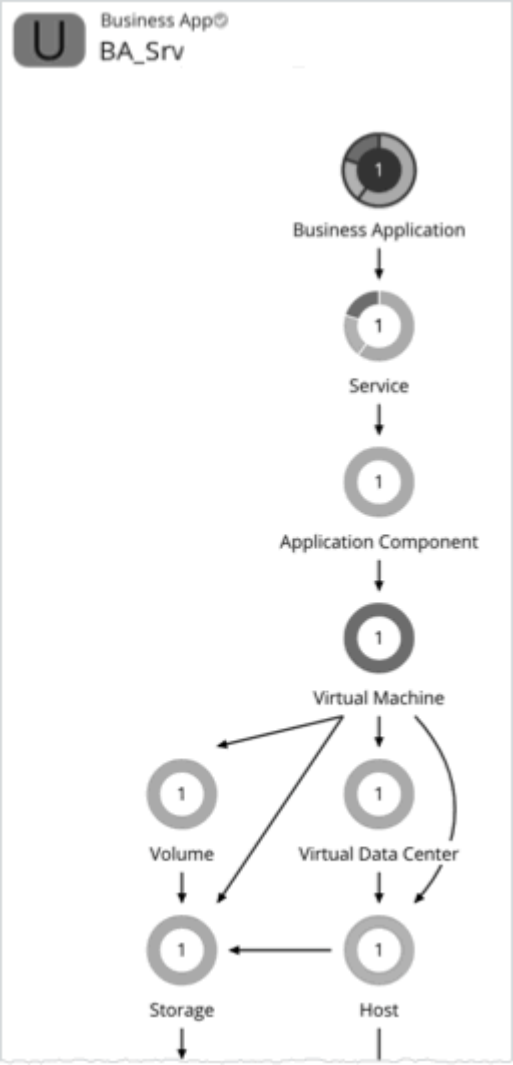
Especially for scheduled actions, it is important that resize calculations use enough historical data to generate actions that will remain viable even during a scheduled maintenance window. A maintenance window is usually set for "down" time, when utilization is low. If analysis uses enough historical data for an action, then the action is more likely to remain viable during the maintenance window.

  - More Elastic – None
  - Less Elastic – 1, 3, or 7 Days

Application map

You can create your own Business applications *(on page 544)*, Business transactions *(on page 546)*, and Services *(on page 549)* without the need to load more application data into the platform. This is especially useful in environments with gaps in the application stack that is shown in the Workload Optimization Manager supply chain. For example, in the absence of an application monitoring target such as AppDynamics or Dynatrace, you do not see Business applications in your supply chain. User-created application entities address those gaps.

Workload Optimization Manager then links them in a supply chain and represents them as a unified group. You can monitor overall performance for the group in the context of the new application entity. Drill down to the newly-created entity to monitor performance. You can also use Search to find the application entity and set it as your scope.

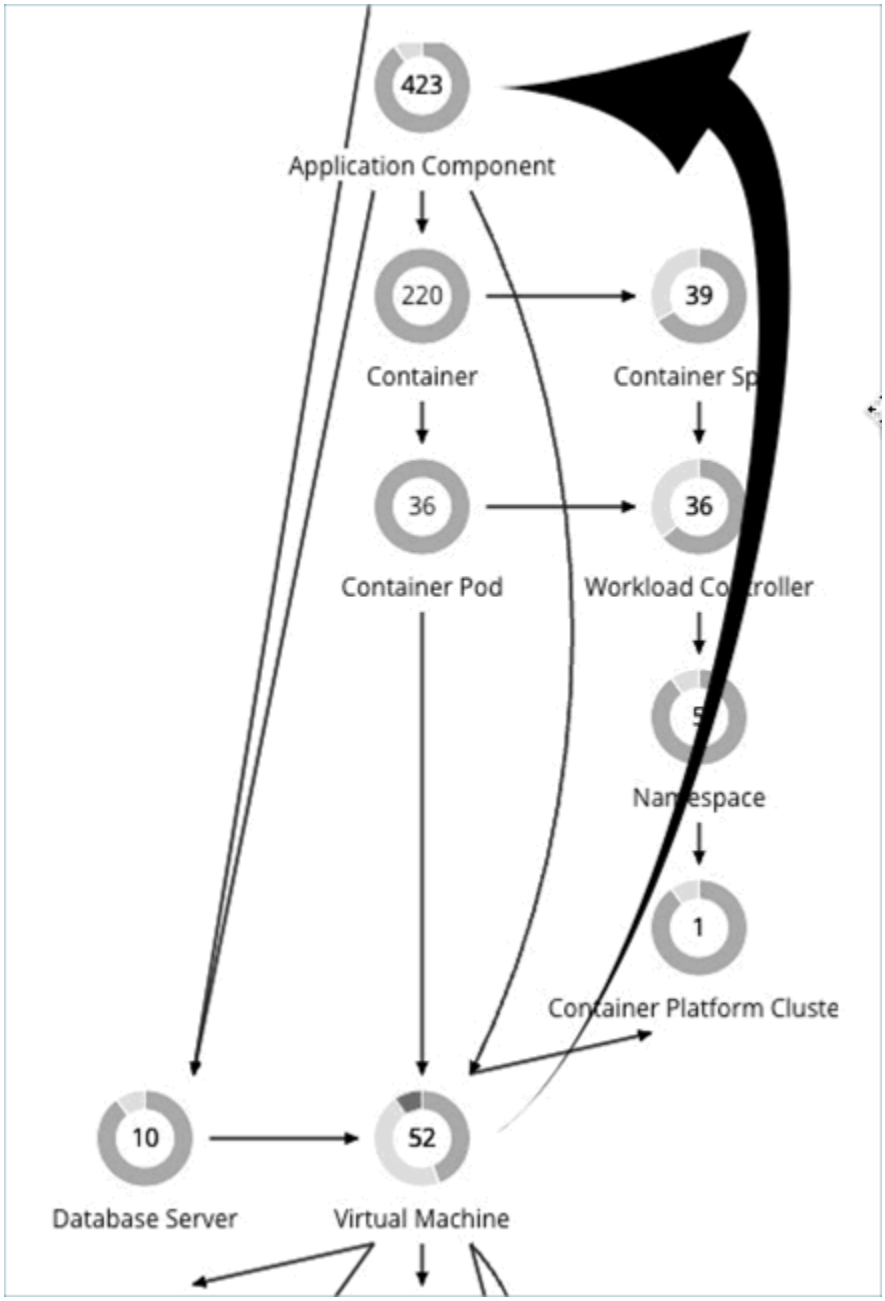


Workload Optimization Manager does not run analysis on any user-created application entity, but it aggregates the underlying risks the same way it does for auto-discovered entities. Workload Optimization Manager adds newly-created entity to the relevant charts. For example, if you created a new Service that has performance risks, it appears in the Top Services chart.

NOTE:
It might take up to 10 minutes to see newly created entities in the supply chain.

Leaf entities

Leaf entities are child nodes with no further children. To find a leaf entity, traverse upward in the supply chain from a seed entity, such as a virtual machine, until you reach a node that doesn't have any child nodes. The following diagram illustrates how this traversal is performed.

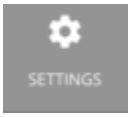


Creating application maps

You can create an application map from the Application Map Manager page or from a **Search**, **Group**, or **Supply Chain** page *(on page 561)*.

Creating application maps from the Application Map Manager page

1. Navigate to the **Settings** Page.



2. Choose **Application Map**.



Application Map

3. Click **Create application map** and then choose the creation mode: **Automatic** or **Manual**.
   - Automatic

     Create a new application entity that is composed of tagged entities. For example, create a new business application that is composed of VMs with the `Production` tag.

     Workload Optimization Manager automatically adds selected entities to the proper map.
   - Manual

     Create a new application entity that is composed of a specific set of application entities and nodes.

     Manually creating the map individually allows for more flexibility.
4. Specify the required values to create the application map based on the creation mode you selected.

   Automatic
   a. Select the entity type for the seed entity that structures your application map.
   b. Specify a name prefix.

      Workload Optimization Manager automatically creates the names for the entities. You can specify an entity name prefix to help you easily identify the application entities that Workload Optimization Manager creates for you.

      The name cannot exceed 20 characters and can contain only alphanumeric or underscore (_) characters.
   c. Click the **Virtual Machine +** button to specify the underlying entities for your application map.
   d. Click the filter icon.
   e. Select a tag option that identifies the underlying entities.
   f. Click **Save and apply**, then click **Save**.
   g. Click **Create application map**.

   Manual
   a. Select the entity type for the seed entity that structures your application map.
   b. Specify the entity name.

      The name cannot exceed 20 characters and can contain only alphanumeric or underscore (_) characters.
   c. Select the underlying entity types.
   d. Enable or disable **Direct Link**.
      - If you select **Direct Link**, Workload Optimization Manager creates an application map based solely on your selected entities. This option is ideal if you require full control of your maps. For example, you might have a requirement to limit the scope of your performance monitoring to certain entities. If you do not select **Direct Link**, Workload Optimization Manager creates a context-based definition of the application map that you are creating and automatically updates that definition as the map evolves.

        When **Direct Link** is disabled, the underlying application entities and nodes that you specified act as "seed entities" for creating the definition. Workload Optimization Manager uses these seed entities to identify the highest entity in the supply chain and any other related entities ("leaf entities"), and then creates a new context-based definition. The result is an application map that closely matches your environment.

        For example, your initial intent might be to create a new Business Application entity composed of several Services (seed entities), so you can monitor performance at the Service level. However, you might not be aware of other entities that might impact performance, making it more time-consuming to identify and resolve performance issues outside of the selected scope. With **Direct Link** disabled, Workload Optimization Manager might discover Application Components and VMs (leaf entities) that back the Services, and then show them in the supply chain. The result is a complete representation of the Business Application that shows performance risks at each level of the discovered application stack.

As the composition of the Business Application changes, Workload Optimization Manager automatically updates the definition so your supply chain view remains current.

When **Direct Link** is enabled, Workload Optimization Manager creates a definition based solely on your selected entities. The system will not include other relevant entities in the map definition. This option is ideal if you require full control of your definitions. For example, you might have a requirement to limit the scope of your performance monitoring to certain entities.

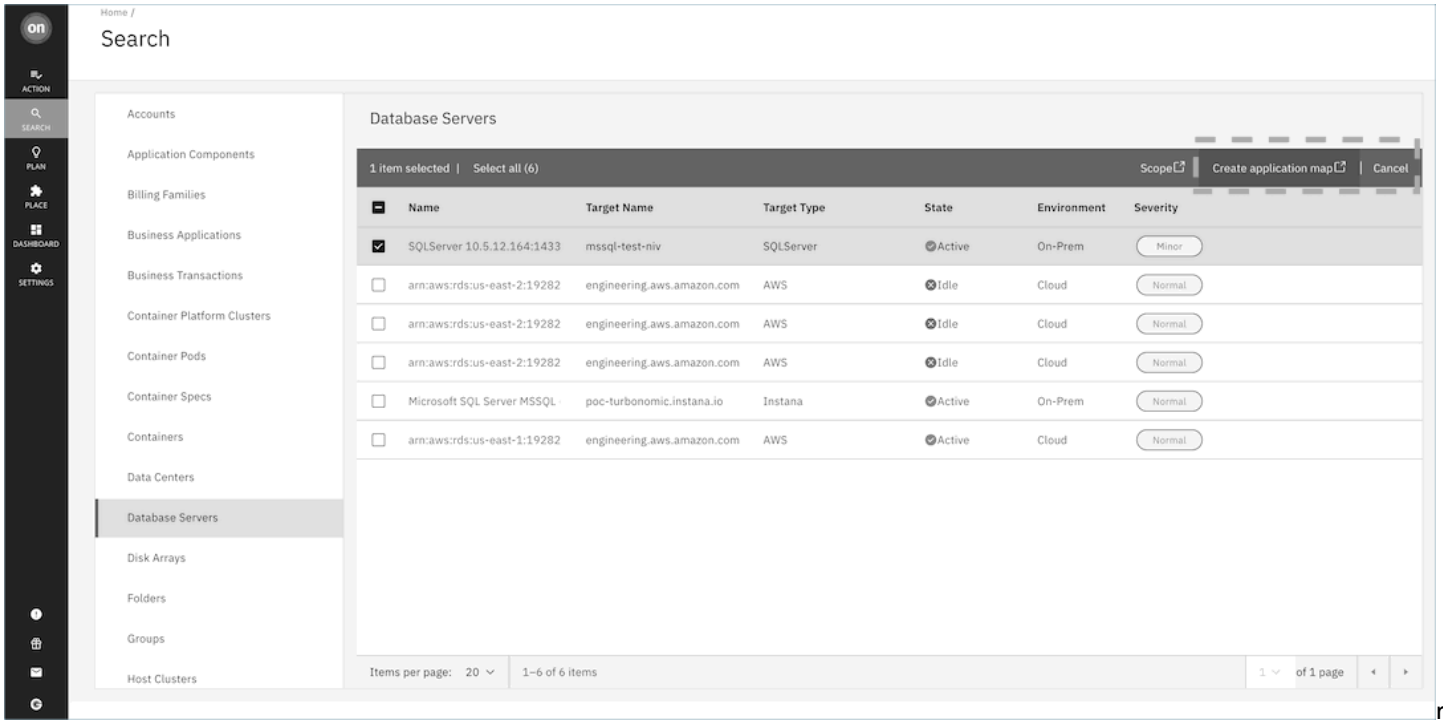
5. Click **Create application map**.

Creating application maps from Search, Group, or Supply Chain pages

You can also create an application map from the **Supply Chain** page for an entity or by selecting one or more entities of the same type from a **Group** or **Search** page. This option is only available for manual application map creation *(on page 560)*. When you select one or more entities and click **Create Application Map**, Workload Optimization Manager pre-populates these underlying entities in your application map definition.

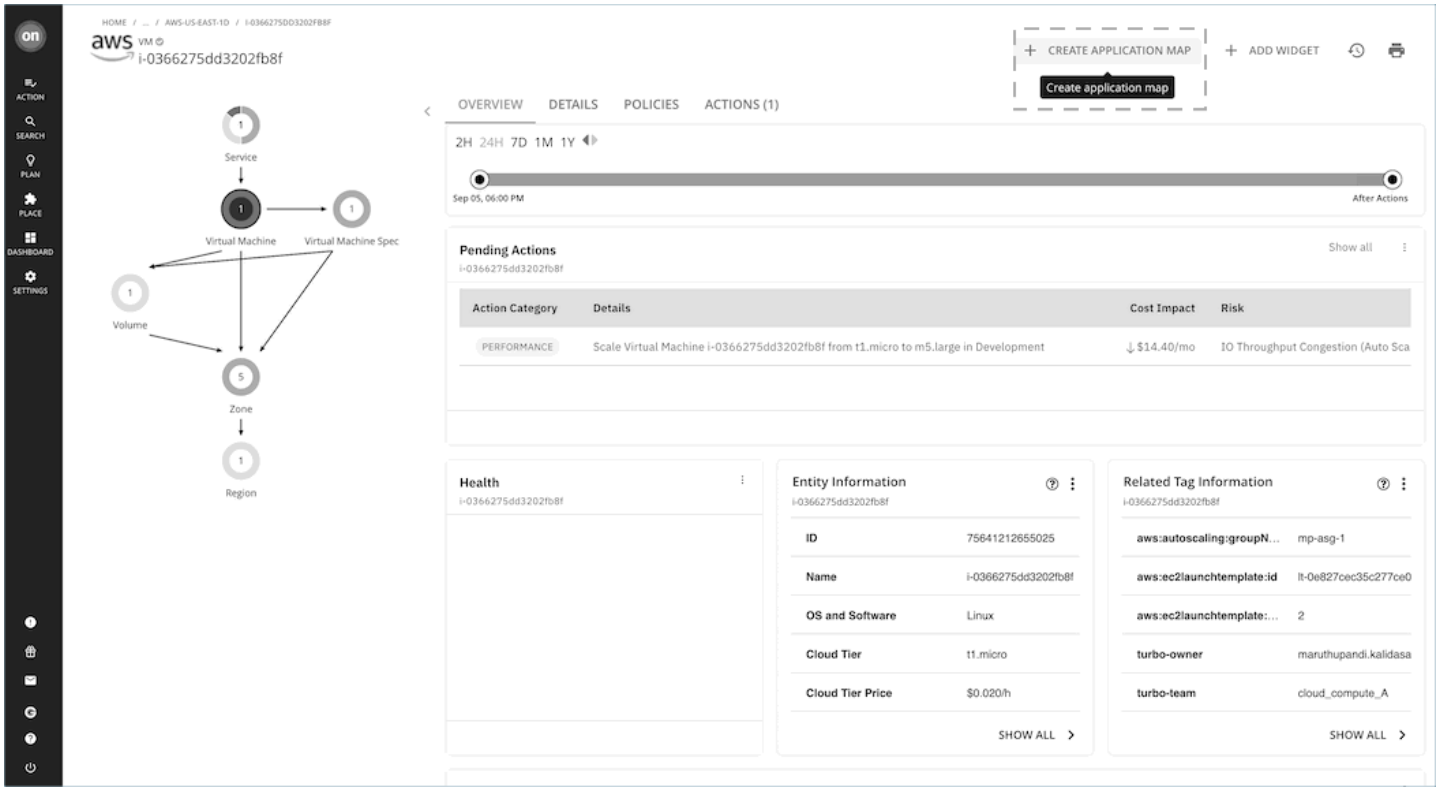
The type of underlying entity that you select determines the options for the seed entity that structures your application map. For example, if the underlying entity or entities you start with are Business Transactions, the only option for the seed entity is Business Application. If you start with a Service, the seed entity can be either a Business Transaction or a Business Application. For any other underlying entity types, you can choose from Business Application, Business Transaction, or Service for the seed entity.

The following image shows the **Create Application Map** button on a **Search** page for Database Servers. Clicking this button creates an application map for the selected server and populates the underlying entities.



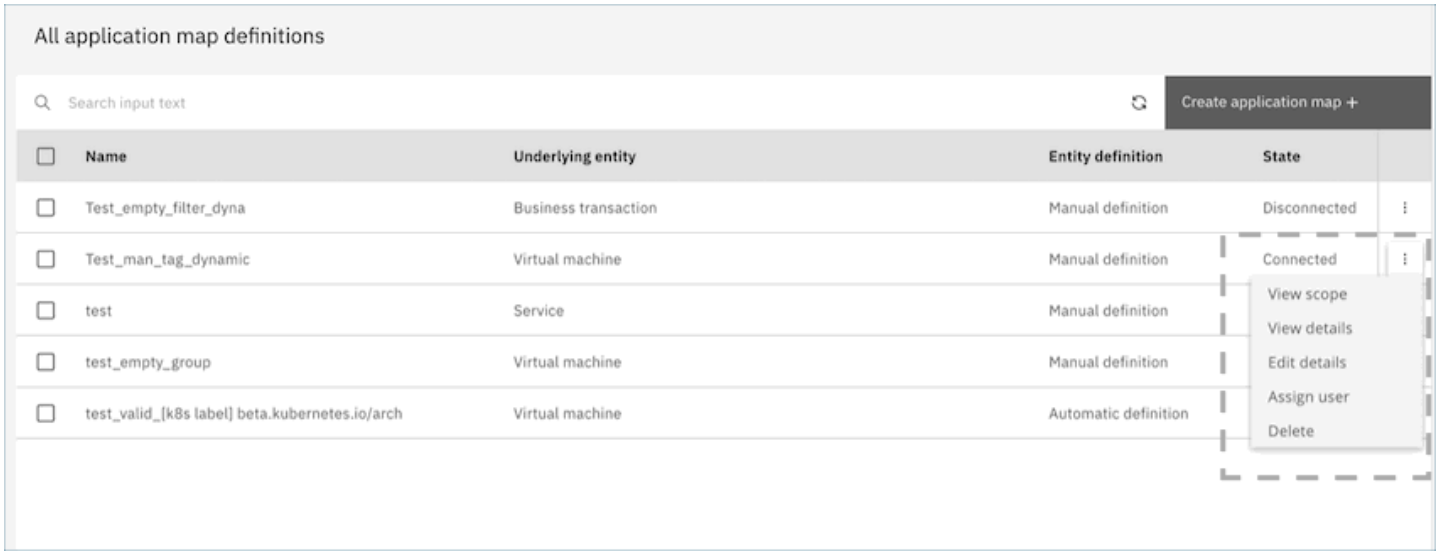
rwlp_jaxr

The following image shows the **Create Application Map** button on a **Supply Chain** page for a Virtual Machine. Clicking this button creates an application map for the Virtual Machine and populates the underlying entities.



Managing application maps from the Options menu

When you navigate to **Settings**> **Application Map** and click the three-dot **Options** menu for any application map, you can choose from the following options.



- View Scope

  Scope into the custom Business Application, Business Transaction, or Service created by the application map definition to see the supply chain it represents.

- View Details

See a brief summary of your application map definition, including:

- Entity type
- Application map state
- Application map underlying entities creation criteria
- Assigned users

- Edit Details

  Edit the application map name, underlying entities selection, and enable or disable the Direct Link option. This option is equivalent to clicking the Edit icon, which is advised when the definition is in a disconnected state.

- Assign User

  An admin can enforce role-based access control (RBAC) by assigning users with the Shared Observer role to specific application map definitions. Users with this role can view the application maps assigned to them. Admins can manage and view assigned application maps from the **User Management** page and can also edit the Shared Observer privileges.

- Delete

  Delete the application map definition from the application map manager and remove its created custom entity from the Turbonomic supply chain.

Application map states

You can view the state of an application map through the Application Manager view or by selecting **View Details** from the **Options** menu. The application map definition can be in one of three states:

- Connected

  The application map was created and is visible in the supply chain. You can access by selecting **View Scope** in the **Options** menu or from the search view.

- In Progress

  The application map is being created. It typically takes around 10 minutes for Workload Optimization Manager to create an application map.

- Disconnected

  The supply chain for the application map is detached or incomplete. This state occurs when the leaf entity (Business Application, Business Transaction, or Service) has no relationships to its underlying entities, or when the underlying entities are missing. In this case, you cannot access the **View Scope** option but you can review problem details through the **View Details** or **Edit Details** options. For more information, see Disconnected State Troubleshooting *(on page 564)*.

Troubleshooting

If you receive an error indicating that the feature has been disabled, an admin must enable the UDT probe in the YAML file. If you are not an admin or if you are a SaaS user, contact your Workload Optimization Manager representative to enable this feature.

1. Open an SSH terminal session to your Workload Optimization Manager instance.

   Log in with the System Administrator that you set up when you installed Workload Optimization Manager:

   - Username: `turbo`
   - Password: `[your_private_password]`

2. In the SSH session, open the YAML file for editing.

   ```
   /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

3. To enable or disable currently configured probes, find the probes to edit and change the settings to `true` (enable) or `false` (disable).

   Search for UDT in the YAML file and set `enabled: true`.

4. Save and apply the changes.

```
kubectl apply -f /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
```

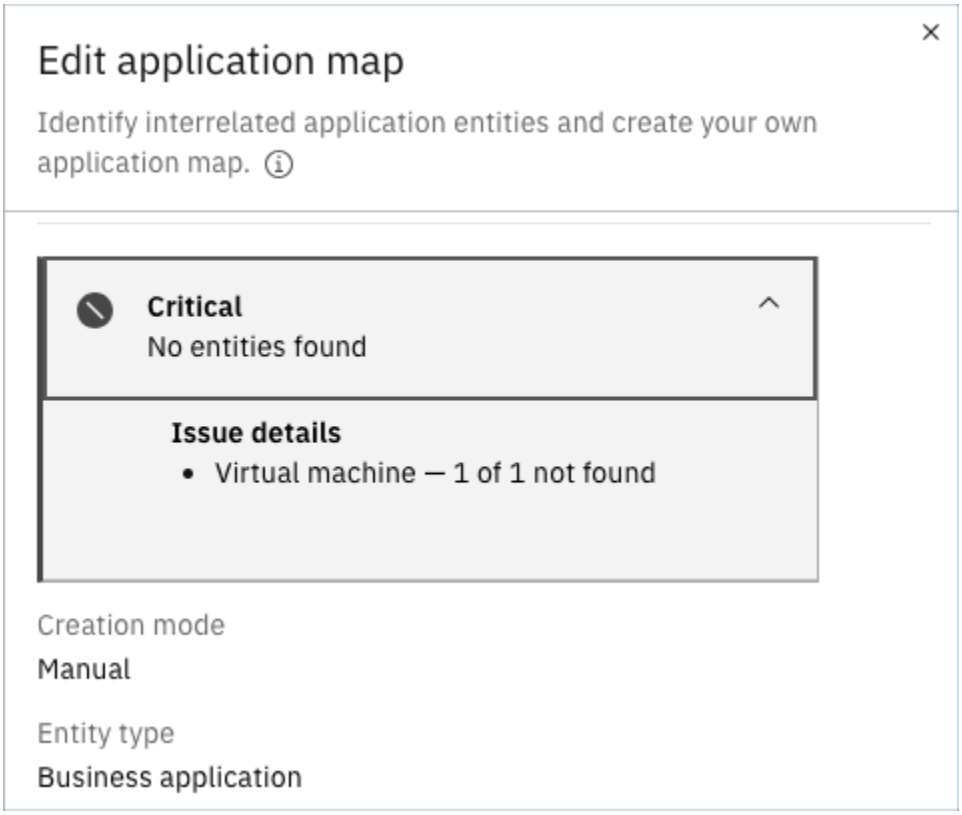
NOTE:
These changes might take up to two minutes to complete. Refresh your Workload Optimization Manager instance then attempt to create an application map.

Disconnected state troubleshooting

If your application map is in a **Disconnected** state, you are informed by a pop-up notification. You can then click the edit icon or select **Edit Details** from the **Options** menu to view more information. The following issues are common causes of disconnected state due to changes in your underlying entities:

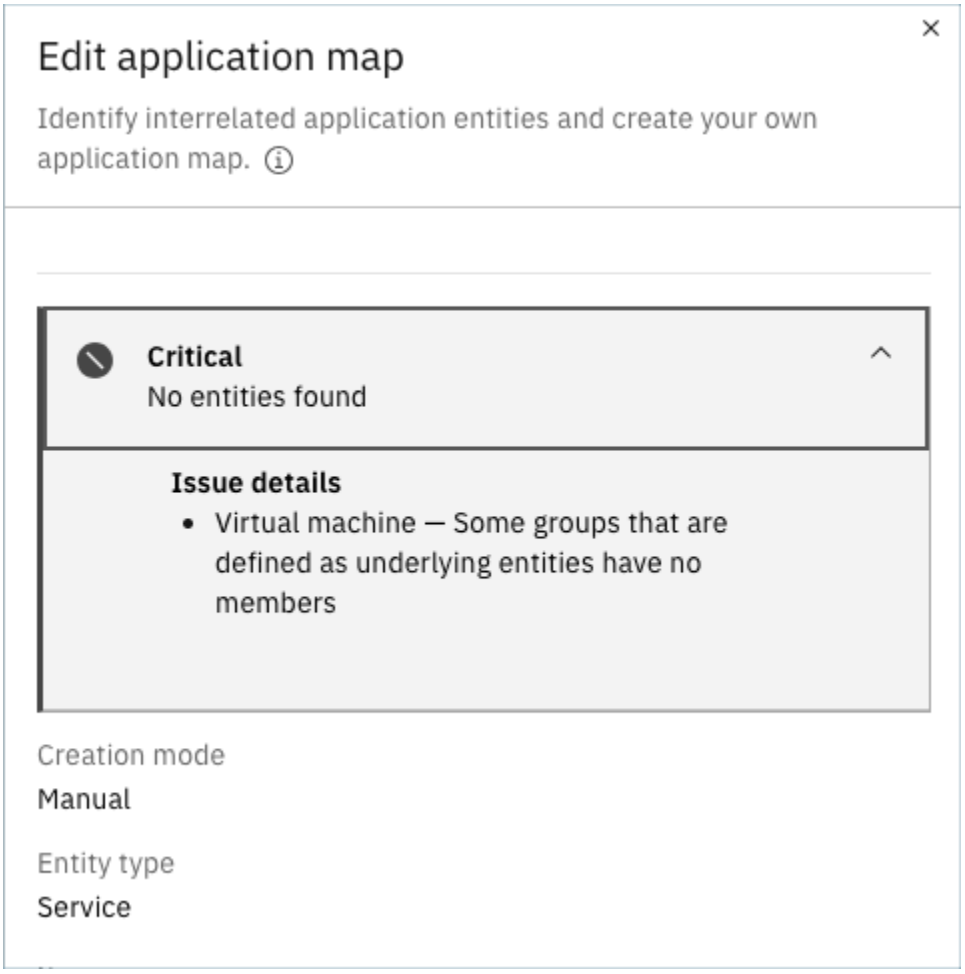
- Unavailable Underlying Entities

  Preselected underlying entities are no longer available in your environment, for example, due to environment changes or target deletions.



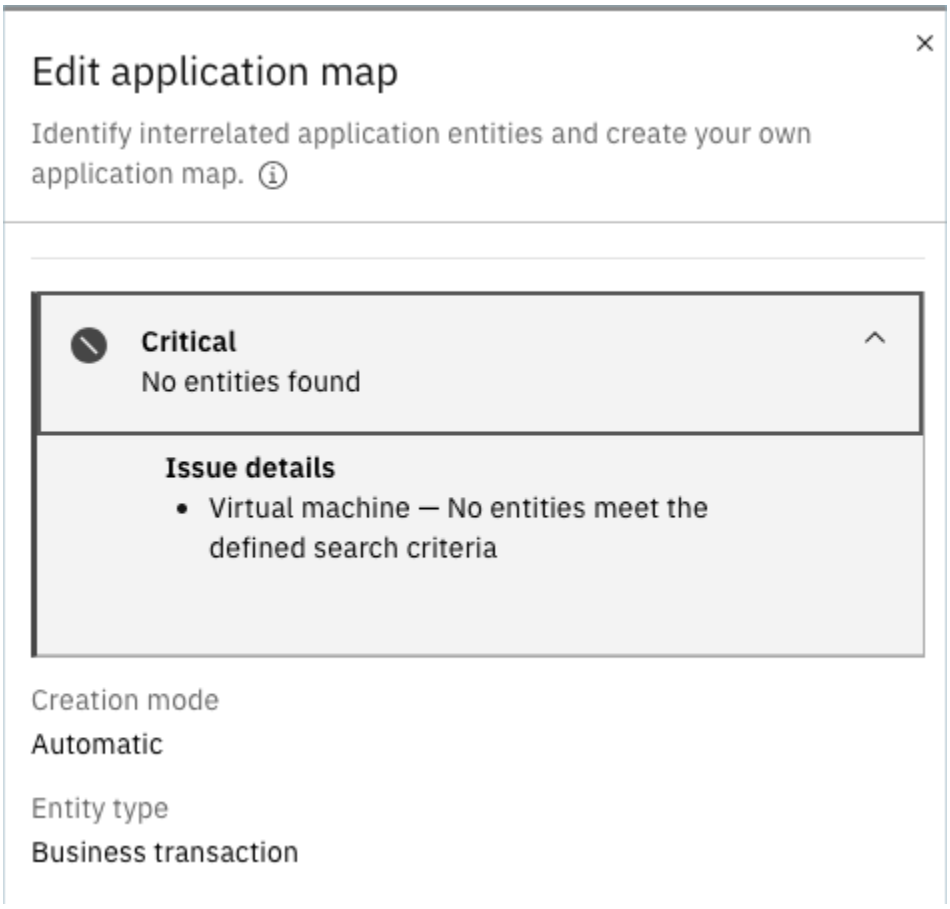
- Groups Without Members

  Some groups that are defined as underlying entities might have no members.



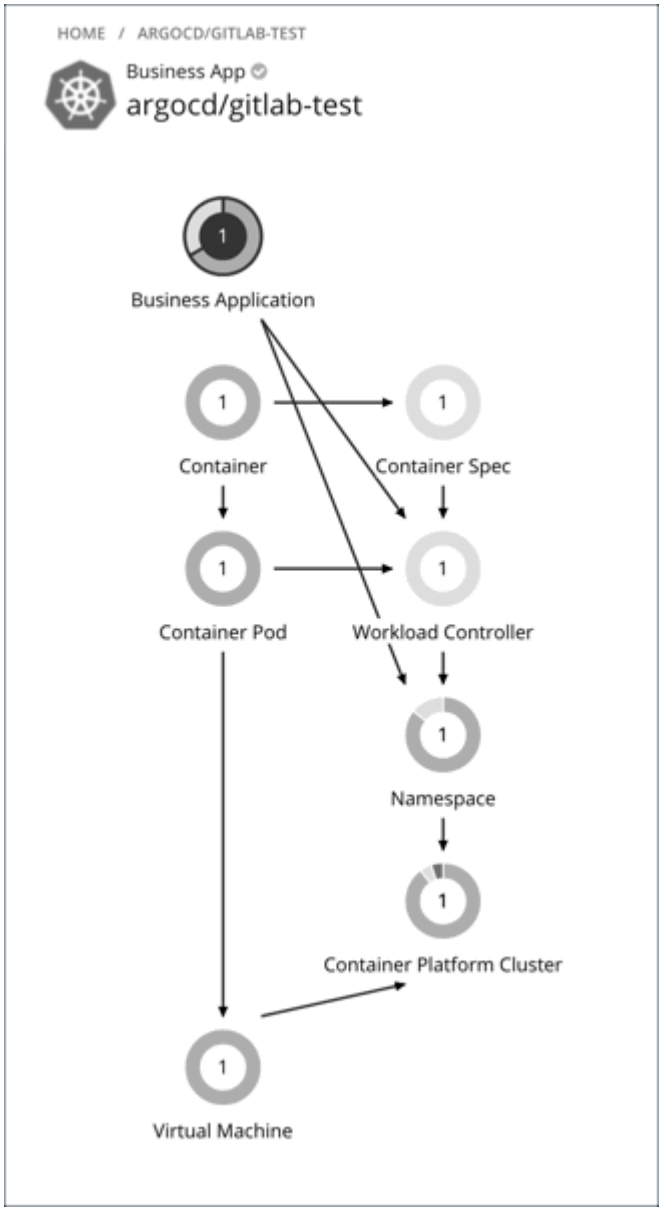
- Definition by Regex Criteria

No entities meet the defined regex criteria for the application map definition.



Entity types - container platform

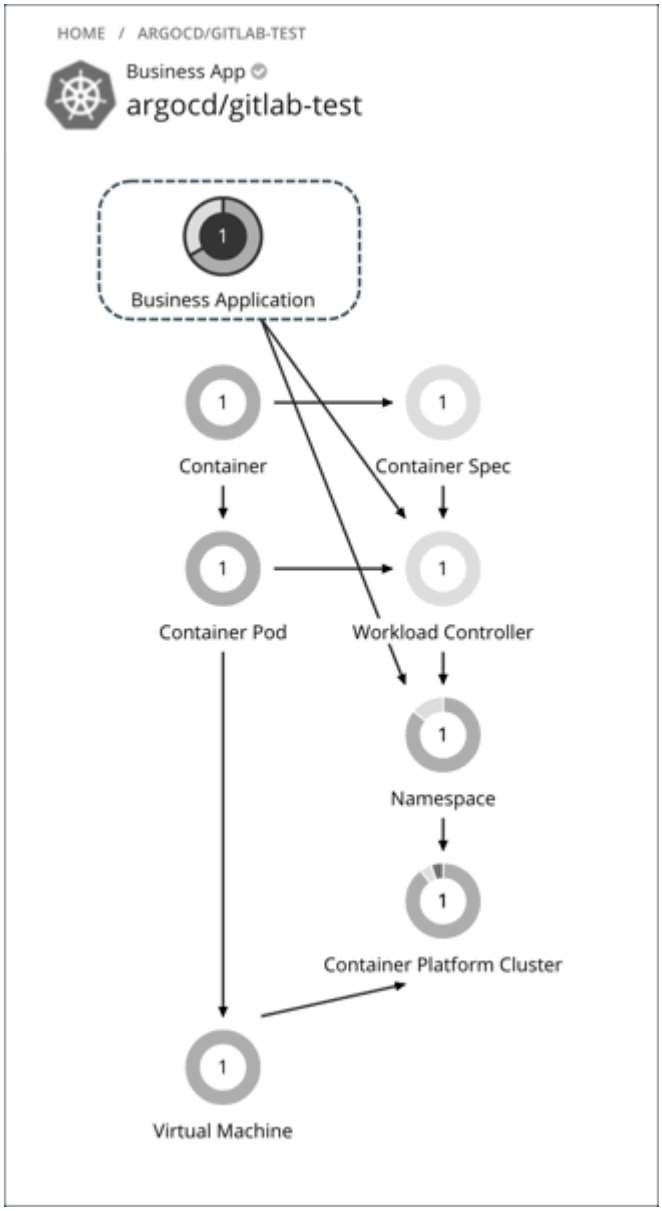
Workload Optimization Manager discovers and monitors the entities that make up your container platform, and recommends actions to assure performance for the applications that consume resources from these entities.



Container platform business application

A business application in container platform environments is a logical set of pods and workloads that represents an `Argo CD GitOps Application` or `ApplicationSet` that is configured within the cluster. While the pods and workloads that comprise the business application are ephemeral, the defined application/applicationset is persistent.

Synopsis



Synopsis	
Provides:	The complete application as defined in an `Argo CD Application` or `ApplicationSet`.
Consumes from:	Services, Application Components, Container, Container Pod, Workload Controller, Namespace, and the underlying Container Platform Cluster and Virtual Machines (Nodes).
Discovered through:	Kubeturbo agent that you deployed to your cluster.

Monitored resources

Workload Optimization Manager monitors the following resources:

- None. No resources are monitored at the business application level.

Actions

None. Workload Optimization Manager does not recommend actions for a Business Application, but it does recommend actions for the underlying workloads and infrastructure. The Pending Actions chart for a Business Application lists these actions, thus providing visibility into the risks that have a direct impact on the Business Application's performance.

GitOps integration

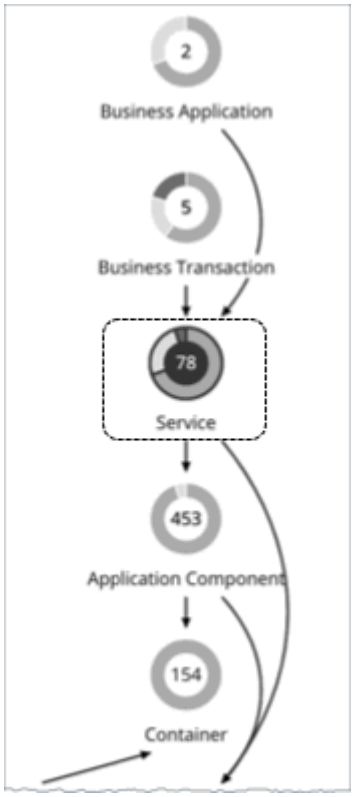
- For more information, refer to topic *(on page 620)*.

Container platform service

A service in container platform environments is a logical set of pods that represents a given application. The service exposes a single entry point for the application process. While the pods that comprise the service are ephemeral, the service is persistent. The service entity also gives historical tracking of the number of replicas that run to support the service.

> **NOTE:**
> For details about services discovered through APM targets, see this topic *(on page 549)*.

Synopsis



Synopsis	
Provides:	Response time and transactions to Business Transactions *(on page 546)* and Business Applications
Consumes from:	Container pods and the underlying nodes
Discovered through:	Kubeturbo agent that you deployed to your cluster

Monitored resources

Workload Optimization Manager monitors the following resources:

- Response time

  Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

  For LLM inference workloads, response time is the turnaround time for each request, including both queuing time and service time. When there is no request, response time is unavailable.

- Transaction

  Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

  For LLM inference workloads, Transaction is the total number of tokens per second, which includes both input tokens and generated tokens. When there is no request, Transaction is zero.

- Number of replicas

  Number of replicas is the number of Application Component replicas running over a given time period.

- Concurrent queries

  For LLM inference workloads, concurrent queries is the number of concurrent queries to a workload. When there is no request, concurrent queries is zero.

- Queueing time

  For LLM inference workloads, queueing time is the amount of time that a request spends in a queue before it is processed. When there is no request, queueing time is zero.

- Service time

  For LLM inference workloads, service time SLO is the amount of processing time needed to generate the next token. This metric is relatively stable for a given model and GPU resource. When there is no request, service time is unavailable.

Actions

Scale (through workload controllers)

Workload Optimization Manager recommends actions to scale the replicas that back the container platform services. Set the scope to the associated Workload Controller to view and execute the actions.

For details, see Workload Controller Scale Actions *(on page 598)*.

Container platform service policies

Policies are required to generate SLO-driven scale actions for workloads, including large language model (LLM) inference workloads.

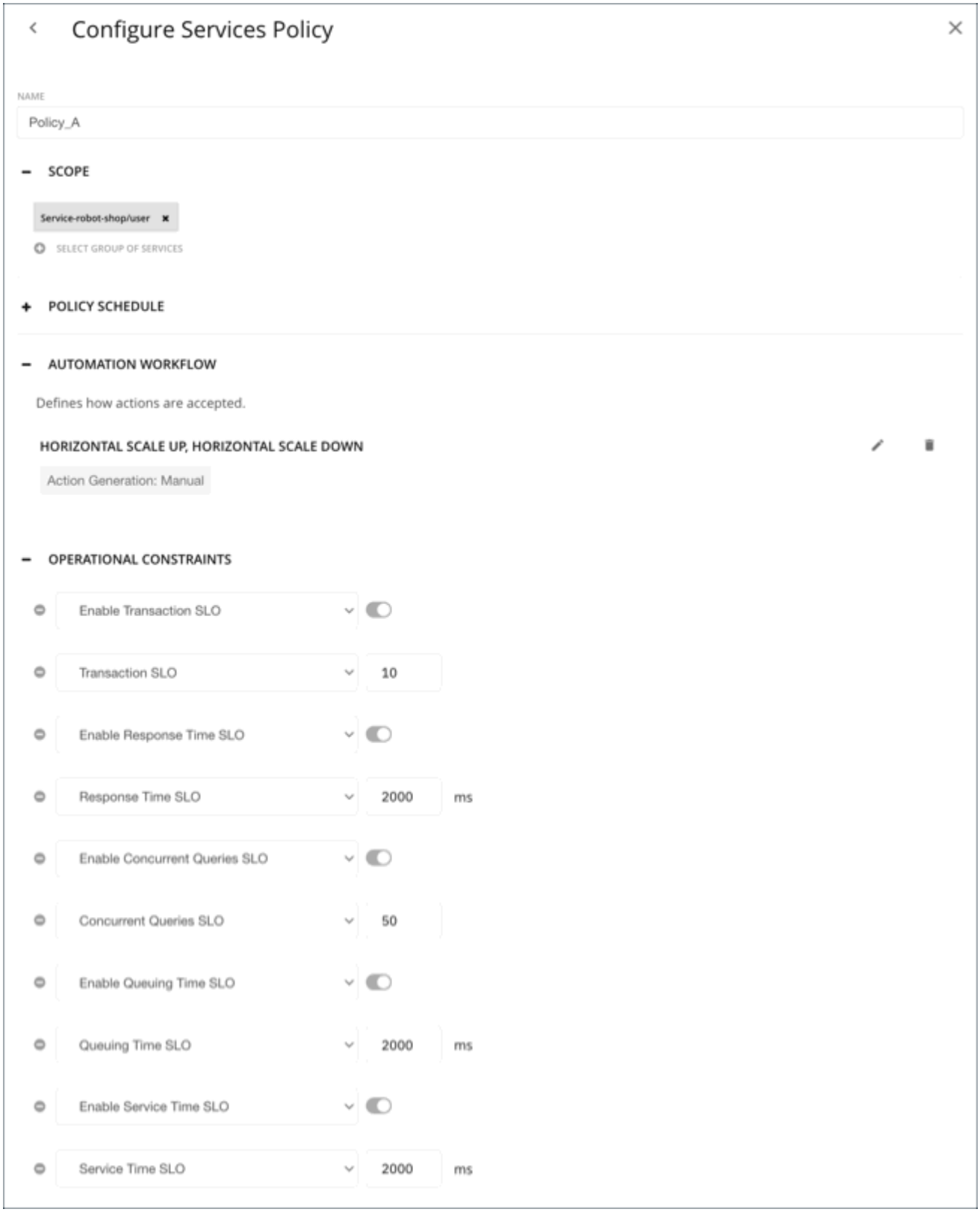
For LLM inference workloads, you can create policies from the Workload Optimization Manager user interface. For all other workloads, you can create policies from the Workload Optimization Manager user interface or declaratively by using Custom Resources in your container platform clusters.

In addition to policies, certain requirements must be satisfied in order for Workload Optimization Manager to generate SLO-driven scale actions. For details, see this topic *(on page 598)*.

NOTE:
When a group of services applies multiple conflicting policies, Workload Optimization Manager enforces the policy with the most conservative settings.

For services discovered through APM targets, a different set of policy settings apply. For details, see this topic *(on page 550)*.

Creating policies from the user interface



- Automation Workflow

Workload Optimization Manager recommends actions to scale the replicas that back the container platform services. Set the scope to the associated Workload Controller to view and execute the actions.

To generate these actions, you must turn on horizontal scaling (up and/or down) and specify your desired SLOs.

Attribute	Default setting
Horizontal Scale Down	Off
Horizontal Scale Up	Off

Action automation is recommended under the following circumstances:

- Your applications run as a set of services backed by a deployment.
- Services deploy via a namespace *without* quotas.
- Newly provisioned pods are placed on nodes with the same CPU speed.
- You do not have an upstream HPA (Horizontal Pod Autoscaling) enabled for the same workload.

It is recommended that you disable automation and manually execute actions under the following circumstances:

- Services deploy via a namespace *with* quotas.
- Newly created pods are scheduled on nodes with different CPU speeds.
- Services have non-resource constraints that could result in newly provisioned pods staying in the pending state.

- Transaction SLO

  Transaction SLO is the maximum number of transactions per second that each Application Component replica can handle.

  For LLM inference workloads, transaction SLO uses TGI metrics and is the total number of tokens per second, which includes both input tokens and generated tokens.

Attribute	Default setting/value
Enable Transaction SLO	Off
Transaction SLO	None If you enable SLO, Workload Optimization Manager uses the default value of 10. You can change this to a different value.

- Response Time SLO

  Response time SLO is the desired *weighted average* response time (in milliseconds) of all Application Component replicas associated with a Service.

  For LLM inference workloads, Response time SLO uses TGI metrics and is the turnaround time for each request, including both queuing time and service time.

Attribute	Default setting/value
Enable Response Time SLO	Off
Response Time SLO [ms]	None If you enable SLO, Workload Optimization Manager uses the default value of 2000. You can change this to a different value.

- Concurrent Queries SLO

  For LLM inference workloads, concurrent queries SLO uses TGI metrics and is the number of concurrent queries to a workload.

Attribute	Default setting/value
Enable Concurrent Queries SLO	Off
Concurrent Queries SLO	None If you enable SLO, Workload Optimization Manager uses the default value of 50. You can change this to a different value.

- Queueing Time SLO

  For LLM inference workloads, queueing time SLO uses TGI metrics and is the amount of time that a request spends in a queue before it is processed.

Attribute	Default setting/value
Enable Queueing Time SLO	Off
Queueing Time SLO [ms]	None If you enable SLO, Workload Optimization Manager uses the default value of 2000. You can change this to a different value.

- Service Time SLO

  For LLM inference workloads, service time SLO uses TGI metrics and is the amount of processing time needed to generate the next token.

Attribute	Default setting/value
Enable Service Time SLO	Off
Service Time SLO [ms]	None If you enable SLO, Workload Optimization Manager uses the default value of 2000. You can change this to a different value.

- Minimum and Maximum Replicas

  You can adjust the default values based on the characteristics of your applications or if you are planning for capacity. The maximum value also acts as a safeguard against overprovisioning of replicas.

Attribute	Default setting/value
Minimum Replicas	1
Maximum Replicas	10000

Creating policies using custom resources (CRs)

Application owners who do not have access to the Workload Optimization Manager user interface can create the following CRs in their container platform clusters:

- `ServicePolicy` CR
- `ServicePolicyBinding` CR

Kubeturbo discovers the settings in these CRs and then displays them as policies in the user interface. Policies in the user interface are read-only since the CRs are the source of truth. Policies are synced with CRs every ten minutes.

`ServicePolicy` CR

This is a YAML resource that specifies the policy settings that apply to a group of container platform services (download a sample here).

This CR is a namespaced resource, so there should be a CR in every namespace for the services, if applicable.

> **IMPORTANT:**
> Before creating the `ServicePolicy` CR, be sure to add this CRD to the container platform cluster. This CRD is mandatory and is intended to ensure the validity of the `ServicePolicy` CR.

Configure the following:

```
apiVersion: policy.turbonomic.io/v1alpha1
kind: SLOHorizontalScale
metadata:
  name: <Policy_Name>
spec:
  minReplicas: <minReplicas>
  maxReplicas: <maxRelicas>
  objectives:
  - name: ResponseTime
    value: <ResponseTimeSLO>
  - name: Transaction
    value: <TransactionSLO>
  behavior:
    scaleUp: <Action_Mode>
    scaleDown: <Action_Mode>
```

Where:

- `<Policy_Name>` is the name of the policy.

  In the user interface, this policy will be added to **Settings > Policies**.
- `<minReplicas>` and `<maxRelicas>` are values that you set based on the characteristics of your applications or if you are planning for capacity. The maximum value also acts as a safeguard against overprovisioning of replicas.
- `<ResponseTimeSLO>`

  Response time SLO is the desired *weighted average* response time (in milliseconds) of all Application Component replicas associated with a Service.

  If an SLO value is not specified, the policy will use the default value in the global policy for services.

  To disable the SLO, remove the `name-value` pair from the CR.
- `<TransactionSLO>`

  Transaction SLO is the maximum number of transactions per second that each Application Component replica can handle.

  If an SLO value is not specified, the policy will use the default value in the global policy for services.

  To disable the SLO, remove the `name-value` pair from the CR.
- `<Action_Mode>` specifies the degree of automation for the generated actions.

  Choose from the following:
  - Turn on scale up and scale down.
  - Turn on scale up, turn off scale down.
  - Turn on scale down, turn off scale up.

  To turn on, set the action acceptance mode to Automatic, Manual, or Recommend. To turn off, set the action acceptance mode to Disabled. See Action Acceptance Modes *(on page 848)* for details.

Action automation is recommended under the following circumstances:

- Your applications run as a set of services backed by a deployment.
- Services deploy via a namespace *without* quotas.
- Newly provisioned pods are placed on nodes with the same CPU speed.
- You do not have an upstream HPA (Horizontal Pod Autoscaling) enabled for the same workload.

It is recommended that you disable automation and manually execute actions under the following circumstances:

- Services deploy via a namespace *with* quotas.
- Newly created pods are scheduled on nodes with different CPU speeds.
- Services have non-resource constraints that could result in newly provisioned pods staying in the pending state.

NOTE:
If you turned off horizontal scaling but specified SLOs values in the policy, actions do not generate but SLO values will continue to display in the Response Time and Transaction charts for services, for your reference. This allows you to gauge performance against those SLOs.

The following example shows a sample CR:

```
apiVersion: policy.turbonomic.io/v1alpha1
kind: SLOHorizontalScale
metadata:
  name: Policy_A
spec:
  minReplicas: 1
  maxReplicas: 10
  objectives:
  - name: ResponseTime
    value: 300
  - name: Transaction
    value: 20
  behavior:
    scaleUp: Automatic
    scaleDown: Disabled
```

`ServicePolicyBinding` CR

This is a YAML resource that binds the policy settings specified in `ServicePolicy` to the group of container platform services specified in this CR (download a sample here).

IMPORTANT:
Before creating the `ServicePolicyBinding` CR, be sure to add this CRD to the container platform cluster. This CRD is mandatory and is intended to ensure the validity of the `ServicePolicyBinding` CR.

Configure the following:

```
apiVersion: policy.turbonomic.io/v1alpha1
kind: PolicyBinding
metadata:
  name: <Group_Name>
policyRef:
  kind: SLOHorizontalScale
  name: <Policy_Name>
targets:
- apiVersion: apps/v1
  kind: Service
  name: <Service_Name_1>
```

```
- apiVersion: apps/v1
  kind: Service
  name: <Service_Name_2>
```

Where:

- <Group_Name> is the name of the group of Services that will apply the policy. In the user interface, this group will be added to **Search > Groups** and will be shown in the format <Namespace>/<Group_Name>[Target_Name].
- <Policy_Name> is the policy name defined in the ServicePolicy CR.

  In the user interface, this policy will be added to **Settings > Policies**.
- Service_Name_1 and Service_Name_2 are the services belonging to the group. Add more entries as needed. You can specify actual names or use regular expressions.

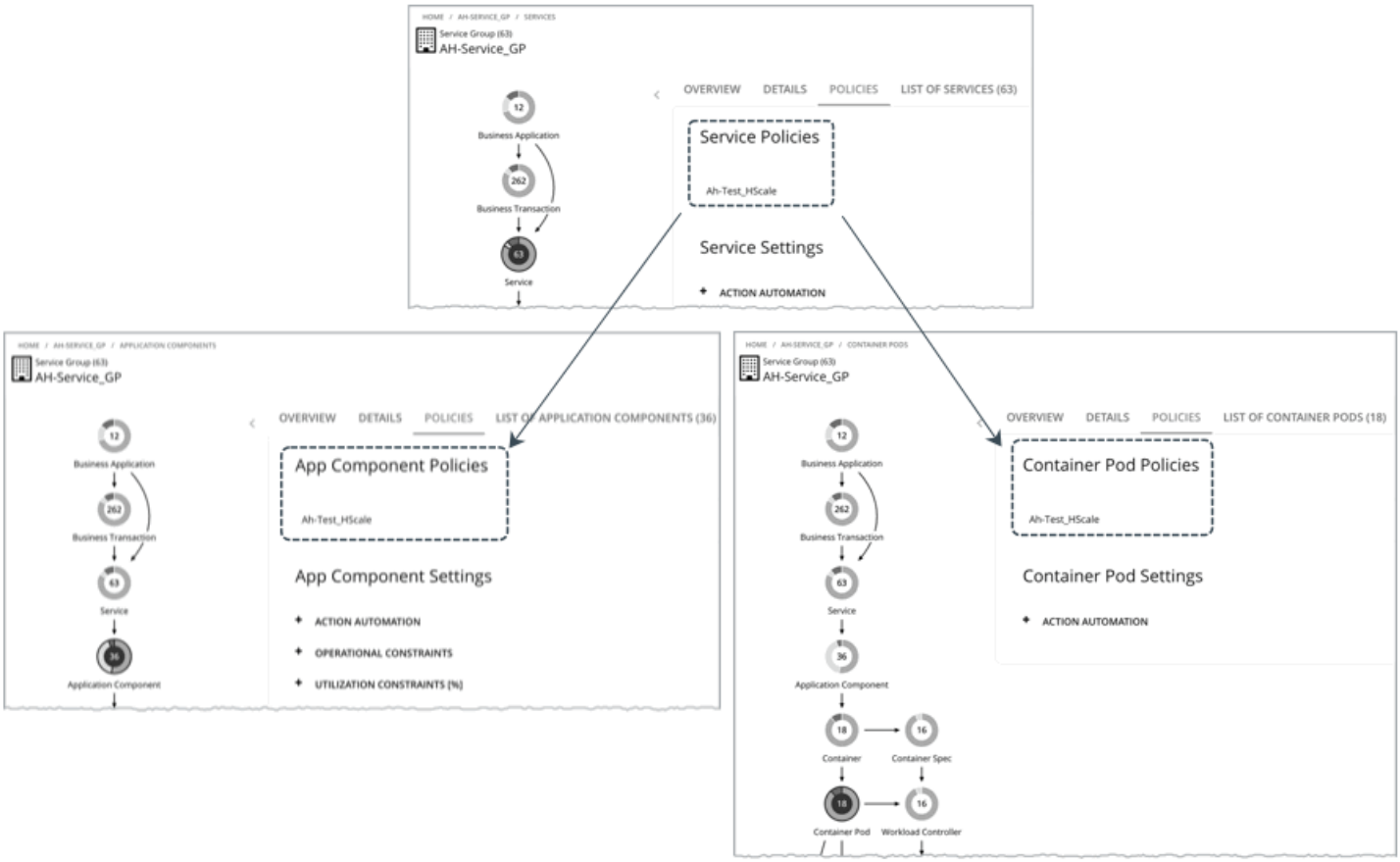
The following example shows a sample CR:

```
apiVersion: policy.turbonomic.io/v1alpha1
kind: PolicyBinding
metadata:
  name: AH-Service_GP
policyRef:
  kind: SLOHorizontalScale
  name: Policy_A
targets:
- apiVersion: apps/v1
  kind: Service
  name: cass-api
- apiVersion: apps/v1
  kind: Service
  name: cass-(cass|user)
```

Propagation of service policy settings

Settings in a service policy propagate to the associated pods and Application Components to establish their relationship and provide context.

For example, assume you created a group of services called AH-Service_GP and then applied the service policy Ah-Test_HScale to that group. When you set the scope to this group, Ah-Test_HScale displays as a policy in the entity views for services, Application Components, and Container Pods. You can click the policy name in any view to see or modify the policy settings.

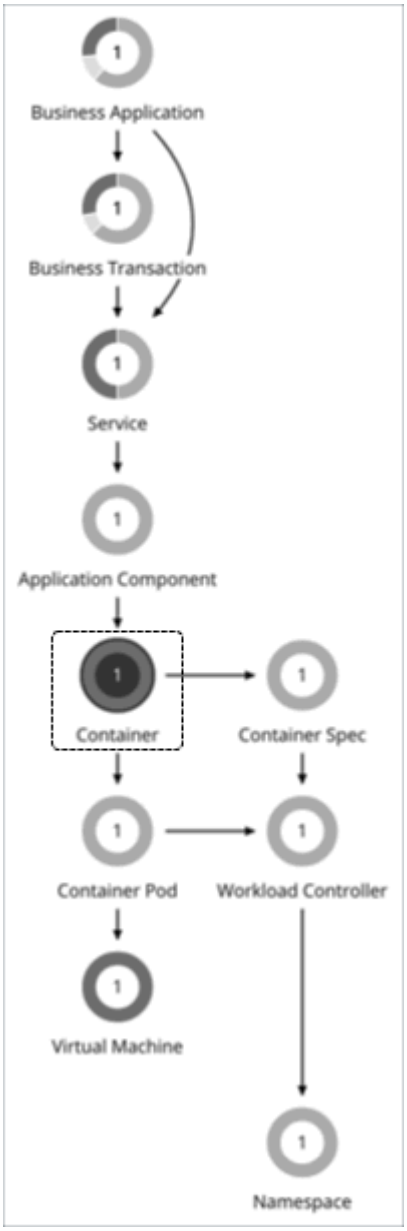


To prevent conflicts, SLO values in service policies override any SLOs set in Application Components. In addition, the Response Time and Transaction charts for Application Components show SLOs specified in the service policy.

Container

An application container is a standalone, executable image of software that includes components to host an application. Because the container instances that support an application can change at any time, containers are considered *ephemeral*.

Synopsis



Synopsis	
Provides:	Resources for the applications to use, including: ▪ Virtual CPU ▪ Virtual Memory
Consumes:	Resources from container pods, virtual machines, and virtual datacenter.
Discovered through:	For container platforms, Workload Optimization Manager discovers containers through the Kubeturbo agent that you deployed to your cluster.

Synopsis	
	For Dynatrace and AppDynamics hosted on containers: ▪ Dynatrace: Workload Optimization Manager discovers containers through the metadata of processes. ▪ AppDynamics: Workload Optimization Manager discovers containers through container objects.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Virtual memory (vMem)

  Virtual memory (vMem) is the virtual memory utilized by a container against the memory limit. If no limit is set, node capacity is used.
- vMem request

  If applicable, vMem request is the virtual memory utilized by a container against the memory request.
- Virtual CPU (vCPU)

  Virtual CPU is the measurement of CPU that is in use.
- vCPU request

  If applicable, vCPU request is the virtual CPU (in mCores) utilized by a container against the CPU request.
- vCPU throttling

  vCPU throttling is the throttling of container virtual CPU that could impact response time, expressed as the percentage of throttling for all containers associated with a Container Spec. In the Capacity and Usage chart for containers, *used* and *utilization* values reflect the actual throttling percentage, while *capacity* value is always 100%.

Actions

Resize

Workload Optimization Manager generates container resize actions for Bare Pods where the workload has no controller.

Container policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

Resize

Workload Optimization Manager generates container resize actions for Bare Pods where the workload has no controller.

Consistent resizing

- For groups in user-defined automation policies:

Attribute	Default setting
Consistent Resizing	Off

  When you create a policy for a group of containers and turn on Consistent Resizing, Workload Optimization Manager resizes all the group members to the same size, such that they all support the top utilization of each resource commodity in the group. For example, assume container A shows top utilization of CPU, and container B shows top utilization of memory. Container resize actions would result in all the containers with CPU capacity to satisfy container A, and memory capacity to satisfy container B.

For an affected resize, the Actions List shows individual resize actions for each of the containers in the group. If you automate resizes, Workload Optimization Manager executes each resize individually in a way that avoids disruption to your workloads.

- For auto-discovered groups:

  Workload Optimization Manager discovers container platform groups such as Deployments, ReplicationControllers, ReplicaSets, DaemonSets, and StatefulSets, and automatically enables Consistent Resizing in a read-only policy for each group. If you do not need to resize all the members consistently, create another policy for the group and turn off Consistent Resizing.

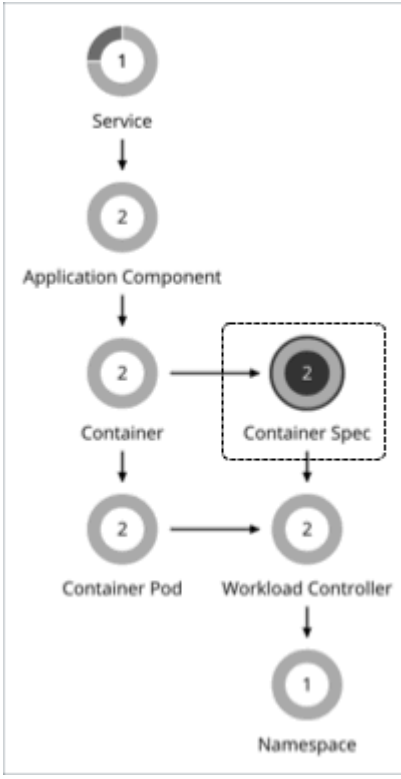
Container spec

A container spec is a shared definition for all ephemeral container replicas with like properties. It is a persistent entity that maintains historical data for its ephemeral containers, and all the replicas that have run in the past. Workload Optimization Manager uses historical data to make container sizing decisions. Data includes:

- vCPU used by all container replicas
- vCPU request capacity (if applicable)
- vMem used by all container replicas
- vMem request capacity (if applicable)

In the Workload Optimization Manager supply chain, the count of replicas maps to the count of container entities that a container spec encompasses.

Synopsis



Synopsis	
Provides:	N/A
Consumes:	N/A

Synopsis	
Discovered through:	Kubeturbo agent that you deployed to your cluster

Monitored resources

Workload Optimization Manager monitors the historical usage of any instance of a container running for the workload (assuming the workload name stays the same). Charts show the trend of usage even with restarts or redeployments.

Actions

Workload Optimization Manager supports the following actions:

Resize (through workload controllers)

A container spec retains the historical utilization data of ephemeral containers. Workload Optimization Manager uses this data to make resize decisions that assure optimal utilization of resources. By default, all replicas of the same container for the same workload type resize consistently.

For details, see Workload Controller Resize Actions *(on page 593)*.

NOTE:
For details about actions on sidecar container specs, see this topic *(on page 292)*.

Container spec policies

This policy controls the analysis of resize actions generated on container specs. To define the acceptance mode of these actions, create a workload controller policy *(on page 603)*.

Analysis properties: overview

Analysis properties specify the tolerance levels for vCPU limits and requests, vMem limits and requests, and CPU throttling that Workload Optimization Manager must consider when it generates resize actions. Resize actions outside these tolerance levels may or may not generate, depending on the action generation setting (on or off) that you specify.

For example, consider resizing vMem limits.

- As vMem demand increases, Workload Optimization Manager can generate vMem limit resize actions that fall within the capacity range that you specified for vMem limits. The action acceptance mode defined in the policy *(on page 603)* for the parent workload controller determines how resize actions are executed. For example, if the action acceptance mode is *automated*, Workload Optimization Manager executes resize actions automatically.
- If the container spec requests vMem outside the range and the action generation for out-of-range resize actions is turned on, Workload Optimization Manager generates resize actions even if they are outside the range. Note that out-of-range resize actions always generate in *recommend* mode, which means that you can view the actions in Action Center *(on page 818)*, but you can only execute them in your cluster.

Analysis properties: resize within range

Setting	Default state
Generate Actions	On
Set to Recommend mode	Off

- Generate Actions

  If you turn on this setting, resize actions generate and take into consideration the analysis properties for vCPU limits and requests, vMem limits and requests, and CPU throttling.

  If you turn off this setting, resize actions do not generate.

- Set to Recommend mode

  If you turn on this setting, resize actions generate in recommend mode, overriding the action acceptance mode configured in the policy for the parent workload controller. This means that you need to execute workload controller resize actions in your cluster. If this is not your intention, keep this setting turned off. You can turn it on for container specs that cannot be updated from their parent workload controllers, such as sidecars.

Analysis properties: vCPU limits

Setting	Default value/state
Capacity range (in mCores)	500 to 64000
Resize below minimum – Generate action in recommend mode	Off
Resize above maximum – Generate action in recommend mode	On
Target utilization	Midpoint – 70%

- Capacity range

  Workload Optimization Manager generates resize actions that fall within the specified capacity range.
- Resize below minimum / Resize above maximum

  If resize values fall outside the specified capacity range and this setting is:
  - On – Workload Optimization Manager generates resize actions in recommend mode. You can execute these actions only in your cluster.
  - Off – Workload Optimization Manager does not generates resize actions.

  For example, if the capacity range is 500 to 64000 and the resize below minimum setting is turned off, Workload Optimization Manager does not generates resize actions that are lower than 500 mCores.
- Target utilization

  Midpoint determines the range of utilization that is considered optimal. The lower and upper limits for this range are always 5% below and above the midpoint, respectively. For example, A midpoint of 70% means that optimal utilization is between 65% and 75%. Workload Optimization Manager takes this optimal utilization range (along with other resize factors) into consideration when it generates resize actions. For example, it can generate a resize action to improve performance while keeping utilization within this range.

  The slider shows 10% and 90% as the minimum and maximum midpoint values that you can specify, for your reference. These values are not configurable. This means that you cannot specify a midpoint value that is lower than 10% or higher than 90%. For example, if you set the midpoint to the minimum of 10%, the optimal utilization range is 5% to 15%.

Analysis properties: vCPU requests

Setting	Default value/state
Capacity range (in mCores)	10 to 64000
Resize below minimum – Generate action in recommend mode	On
Resize above maximum – Generate action in recommend mode	On
Target utilization	Midpoint – 110%

- Capacity range

  Workload Optimization Manager generates resize actions that fall within the specified capacity range.
- Resize below minimum / Resize above maximum

If resize values fall outside the specified capacity range and this setting is:

- On – Workload Optimization Manager generates resize actions in recommend mode. You can execute these actions only in your cluster.
- Off – Workload Optimization Manager does not generates resize actions.

For example, if the capacity range is 10 to 64000 and the resize above minimum setting is turned off, Workload Optimization Manager does not generates resize actions that are higher than 64000 mCores.

- Target utilization

  Midpoint determines the range of utilization that is considered optimal. The lower and upper limits for this range are always 5% below and above the midpoint, respectively. For example, a midpoint of 110% means that optimal utilization is between 105% and 115%. Workload Optimization Manager takes this optimal utilization range (along with other resize factors) into consideration when it generates resize actions. For example, it can generate a resize action to improve performance while keeping utilization within this range.

  NOTE:
  Requests represent the minimum guaranteed resources reserved for containers. It is possible for actual utilization to exceed 100% of requests, especially when more resources are required to maintain performance. Workload Optimization Manager uses 110% as the default value to keep requests close to the average utilization and avoid overprovisioning.

  The slider shows 10% and 190% as the minimum and maximum midpoint values that you can specify, for your reference. These values are not configurable. This means that you cannot specify a midpoint value that is lower than 10% or higher than 190%. For example, if you set the midpoint to the minimum of 10%, the optimal utilization range is 5% to 15%.

Analysis properties: vMem limits

Setting	Default value/state
Capacity range (in MB)	10 to 1048576
Resize below minimum – Generate action in recommend mode	On
Resize above maximum – Generate action in recommend mode	On
Target utilization	Midpoint – 70%

- Capacity range

  Workload Optimization Manager generates resize actions that fall within the specified capacity range.

- Resize below minimum / Resize above maximum

  If resize values fall outside the specified capacity range and this setting is:

  - On – Workload Optimization Manager generates resize actions in recommend mode. You can execute these actions only in your cluster.
  - Off – Workload Optimization Manager does not generates resize actions.

  For example, if the capacity range is 10 to 1048576 and the resize below minimum setting is turned on, Workload Optimization Manager generates resize actions that are lower than 10 MB. You can execute these actions only in your cluster.

- Target utilization

  Midpoint determines the range of utilization that is considered optimal. The lower and upper limits for this range are always 5% below and above the midpoint, respectively. For example, A midpoint of 70% means that optimal utilization is between 65% and 75%. Workload Optimization Manager takes this optimal utilization range (along with other resize

factors) into consideration when it generates resize actions. For example, it can generate a resize action to improve performance while keeping utilization within this range.

The slider shows 10% and 90% as the minimum and maximum midpoint values that you can specify, for your reference. These values are not configurable. This means that you cannot specify a midpoint value that is lower than 10% or higher than 90%. For example, if you set the midpoint to the minimum of 10%, the optimal utilization range is 5% to 15%.

Analysis properties: vMem requests

Setting	Default value/state
Capacity range (in MB)	10 to 1048576
Resize below minimum – Generate action in recommend mode	On
Resize above maximum – Generate action in recommend mode	On
Target utilization	Midpoint – 110%

- Capacity range

  Workload Optimization Manager generates resize actions that fall within the specified capacity range.
- Resize below minimum / Resize above maximum

  If resize values fall outside the specified capacity range and this setting is:
  - On – Workload Optimization Manager generates resize actions in recommend mode. You can execute these actions only in your cluster.
  - Off – Workload Optimization Manager does not generates resize actions.

  For example, if the capacity range is 10 to 1048576 and the resize above maximum setting is turned on, Workload Optimization Manager generates resize actions that are higher than 1048576 MB (or 1 TB). You can execute these actions only in your cluster.
- Target utilization

  Midpoint determines the range of utilization that is considered optimal. The lower and upper limits for this range are always 5% below and above the midpoint, respectively. For example, a midpoint of 110% means that optimal utilization is between 105% and 115%. Workload Optimization Manager takes this optimal utilization range (along with other resize factors) into consideration when it generates resize actions. For example, it can generate a resize action to improve performance while keeping utilization within this range.

  NOTE:
  Requests represent the minimum guaranteed resources reserved for containers. It is possible for actual utilization to exceed 100% of requests, especially when more resources are required to maintain performance. Workload Optimization Manager uses 110% as the default value to keep requests close to the average utilization and avoid overprovisioning.

  The slider shows 10% and 190% as the minimum and maximum midpoint values that you can specify, for your reference. These values are not configurable. This means that you cannot specify a midpoint value that is lower than 10% or higher than 190%. For example, if you set the midpoint to the minimum of 10%, the optimal utilization range is 5% to 15%.

Analysis properties: CPU throttling

Setting	Default value
CPU throttling tolerance	Midpoint – 20%

This value defines your acceptable level of throttling and directly impacts the resize actions generated on CPU limits.

A low percentage value indicates more sensitivity to throttling, while a high value indicates more tolerance for throttling and a higher risk of congestion.

Midpoint determines the range of throttling that will be tolerated. The slider shows 10% and 70% as the minimum and maximum midpoint values that you can specify, for your reference. These values are not configurable. This means that you cannot specify a midpoint value that is lower than 10% or higher than 70%.

Learn more about CPU throttling here.

Scaling constraints: increment constants

Workload Optimization Manager recommends changes in terms of the specified change constants.

Attribute	Default value
Increment constant for vCPU Limit (mCores)	100
Increment constant for vCPU Request (mCores)	10
Increment constant for vMem Limit (MB)	128
Increment constant for vMem Request (MB)	16

For example, assume the vCPU request increment constant is 100 mCores and you have requested 800 mCores for a container. Workload Optimization Manager could recommend reducing the request by 100, down to 700 mCores.

For vMem, do not set the increment value to be lower than what is necessary for the container to operate. If the vMem change value is too low, Workload Optimization Manager might allocate insufficient vMem. For a container that is underutilized, Workload Optimization Manager will reduce vMem allocation by the increment constant, but it will not leave a container with zero vMem. For example, if you set this to 128, then Workload Optimization Manager cannot reduce the vMem to less than 128 MB.

Scaling constraints: rate of resize

(For the *default* policy only)

Attribute	Default value
Rate of Resize	High

When resizing resources, Workload Optimization Manager calculates the optimal values for vCPU and vMem, but it does not necessarily make a change to that value in one action. Workload Optimization Manager uses the rate of resize setting to determine how to make the change in a single action.

- Low

  Change the value by one increment constant, only. For example, if the resize action calls for increasing vMem, and the increment constant is set at 128, Workload Optimization Manager increases vMem by 128 MB.

- Medium

  Change the value by an increment constant that is 1/4 of the difference between the current value and the optimal value. For example, if the current vMem is 2 GB and the optimal vMem is 10 GB, then Workload Optimization Manager will raise vMem to 4 GB (or as close to that as the increment constant will allow).

- High

  Change the value to be the optimal value. For example, if the current vMem is 2 GB and the optimal vMem is 8 GB, then Workload Optimization Manager will raise vMem to 8 GB (or as close to that as the increment constant will allow).

Scaling constraints: aggressiveness and observation periods

Workload Optimization Manager uses these settings to calculate utilization percentiles for vCPU and vMem. It then recommends actions to improve utilization based on the observed values for a given time period.

- Aggressiveness

Attribute	Default value
Aggressiveness	99th Percentile

When evaluating vCPU and vMem performance, Workload Optimization Manager considers resource utilization as a percentage of capacity. The utilization drives actions to scale the available capacity either up or down. To measure utilization, the analysis considers a given utilization percentile. For example, assume a 99th percentile. The percentile utilization is the highest value that 99% of the observed samples fall below. Compare that to average utilization, which is the average of *all* the observed samples.

Using a percentile, Workload Optimization Manager can recommend more relevant actions. This is important in the cloud, so that analysis can better exploit the elasticity of the cloud. For scheduled policies, the more relevant actions will tend to remain viable when their execution is put off to a later time.

For example, consider decisions to reduce the capacity for vCPU on a container. Without using a percentile, Workload Optimization Manager never resizes below the recognized peak utilization. For most containers there are moments when peak vCPU reaches high levels. Assume utilization for a container peaked at 100% just once. Without the benefit of a percentile, Workload Optimization Manager will not reduce allocated vCPU for that container.

With **Aggressiveness**, instead of using the single highest utilization value, Workload Optimization Manager uses the percentile you set. For the previous example, assume a single vCPU burst to 100%, but for 99% of the samples vCPU never exceeded 50%. If you set **Aggressiveness** to 99th Percentile, then Workload Optimization Manager can see this as an opportunity to reduce vCPU allocation for the container.

In summary, a percentile evaluates the sustained resource utilization, and ignores bursts that occurred for a small portion of the samples. You can think of this as aggressiveness of resizing, as follows:

  - 100th Percentile

    Least aggressive, recommended for critical workloads that need maximum guaranteed performance at all times
  - 99th Percentile (Default)

    Recommended setting to achieve maximum performance
  - 90th Percentile

    Most aggressive, recommended for non-production workloads that can stand higher resource utilization

- Max observation period

Attribute	Default value
Max Observation Period	Last 30 Days

To refine the calculation of resource utilization percentiles, you can set the sample time to consider. Workload Optimization Manager uses historical data from up to the number of days that you specify as a sample period. (If the database has fewer days' data then it uses all of the stored historical data.)

A shorter period means there are fewer data points to account for when Workload Optimization Manager calculates utilization percentiles. This results in more dynamic, elastic resizing, while a longer period results in more stable or less elastic resizing. You can make the following settings:

  - Less Elastic – Last 90 Days
  - Recommended – Last 30 Days
  - More Elastic – Last 7 Days

- Min observation period

Attribute	Default value
Min observation period	1 Day

This setting ensures historical data for a minimum number of days before Workload Optimization Manager will generate an action based on the percentile set in **Aggressiveness**. This ensures a minimum set of data points before it generates the action.

Especially for scheduled actions, it is important that resize calculations use enough historical data to generate actions that will remain viable even during a scheduled maintenance window. A maintenance window is usually set for "down" time, when utilization is low. If analysis uses enough historical data for an action, then the action is more likely to remain viable during the maintenance window.

- More Elastic – None
- Recommended – 1 Day
- Less Elastic – 3 or 7 Days

Container spec policies – declarative configuration

This policy controls the analysis of resize actions generated on container specs. To define the acceptance mode of these actions, create a workload controller policy *(on page 603)*.

Application owners who do not have access to the Workload Optimization Manager user interface can create the `ContainerVerticalScalePolicy` Custom Resource (CR) in their container platform clusters. This CR is a YAML resource that specifies the container spec policy settings (download a sample here).

Kubeturbo discovers the settings in this CR and then displays them as a container spec policy in the user interface. This policy is read-only (since the CR is the source of truth) and is synced with the CR every ten minutes.

IMPORTANT:
Before creating the `ContainerVerticalScalePolicy` CR, be sure to add this CRD to the container platform cluster. This CRD is mandatory and is intended to ensure the validity of the `ContainerVerticalScalePolicy` CR.

The following example shows a sample CR:

```
apiVersion: policy.turbonomic.io/v1alpha1
kind: ContainerVerticalScale
metadata:
  labels:
    app.kubernetes.io/name: containerverticalscale
    app.kubernetes.io/instance: containerverticalscale-sample
    app.kubernetes.io/part-of: turbo-policy
    app.kubernetes.io/managed-by: kustomize
    app.kubernetes.io/created-by: turbo-policy
  name: <Your_Value>
spec:
  settings:
    limits:
      cpu:
        max: <Your_Value>
        min: <Your_Value>
        recommendAboveMax: <Your_Value>
        recommendBelowMin: <Your_Value>
      memory:
        max: <Your_Value>
        min: <Your_Value>
        recommendAboveMax: <Your_Value>
        recommendBelowMin: <Your_Value>
    requests:
      cpu:
        min: <Your_Value>
        recommendBelowMin: <Your_Value>
      memory:
        min: <Your_Value>
        recommendBelowMin: <Your_Value>
```

```
    increments:
      cpu: <Your_Value>
      memory: <Your_Value>
    observationPeriod:
      min: <Your_Value>
      max: <Your_Value>
    rateOfResize: <Your_Value>
    aggressiveness: <Your_Value>
    cpuThrottlingTolerance: <Your_Value>
  behavior:
    resize: <Your_Value>
```

The following sections describe the CR settings that you need to configure.

Name

Setting	Example
`name:`	`name: container-vertical-scale-sample`

Specify the name of the policy. In the user interface, this policy will be added to **Settings > Policies**.

CPU and memory limits

Settings	Example

`limits:`
`  cpu:`
`  max:`
`  min:`
`  recommendAboveMax:`
`  recommendBelowMin:`
`  memory:`
`  max:`
`  min:`
`  recommendAboveMax:`
`  recommendBelowMin:`
	`limits:`
	`  cpu:`
	`  max: 64`
	`  min: 500m`
	`  recommendAboveMax: true`
	`  recommendBelowMin: false`
	`  memory:`
	`  max: 104857M`
	`  min: 10M`
	`  recommendAboveMax: true`
	`  recommendBelowMin: true`

Specify the capacity range (maximum and minimum thresholds) for resize limits.

- Resize CPU limits, in cores

  Add `m` if the value is in mCores.
- Resize memory limits, in GB

  Add `M` if the value is in MB.

For `recommendAboveMax` and `recommendBelowMin` values, set to either `true` or `false`.

- `true` – When resize values fall outside the capacity range, Workload Optimization Manager will post resize actions for you to review. You can only execute these actions outside Workload Optimization Manager.
- `false` – Workload Optimization Manager will not generate resize actions if resize values fall outside the normal range.

CPU and memory requests

Settings	Example
`requests:` `  cpu:`	`requests:` `  cpu:`

Settings	Example
<pre>min: recommendBelowMin: memory: min: recommendBelowMin:</pre>	<pre>min: 10m recommendBelowMin: true memory: min: 10M recommendBelowMin: true</pre>

Specify the minimum thresholds for resize requests.

- Resize CPU requests, in cores

  Add `m` if the value is in mCores.
- Resize memory requests, in GB

  Add `M` if the value is in MB.

For `recommendBelowMin` values, set to either `true` or `false`.

- `true` – When resize values fall outside the minimum thresholds, Workload Optimization Manager will post resize actions for you to review. You can only execute these actions outside Workload Optimization Manager.
- `false` – Workload Optimization Manager will not generate resize actions if resize values fall outside the minimum thresholds.

Increment constants

Settings	Example
<pre>increments: cpu: memory:</pre>	<pre>increments: cpu: 100m memory: 128M</pre>

Workload Optimization Manager recommends changes in terms of the specified change constants.

- Resize increment for CPU, in cores

  Add `m` if the value is in mCores.
- Resize increment for memory, in GB

  Add `M` if the value is in MB.

For example, assume the vCPU request increment constant is 100 mCores and you have requested 800 mCores for a container. Workload Optimization Manager could recommend reducing the request by 100, down to 700 mCores.

For vMem, do not set the increment value to be lower than what is necessary for the container to operate. If the vMem change value is too low, Workload Optimization Manager might allocate insufficient vMem. For a container that is underutilized, Workload Optimization Manager will reduce vMem allocation by the increment constant, but it will not leave a container with zero vMem. For example, if you set this to 128, then Workload Optimization Manager cannot reduce the vMem to less than 128 MB.

Rate of resize

Setting	Example
<pre>rateOfResize:</pre>	<pre>rateOfResize: high</pre>

When resizing resources, Workload Optimization Manager calculates the optimal values for vCPU and vMem, but it does not necessarily make a change to that value in one action. Workload Optimization Manager uses the rate of resize setting to determine how to make the change in a single action.

- `low`

  Change the value by one increment constant, only. For example, if the resize action calls for increasing vMem, and the increment constant is set at 128, Workload Optimization Manager increases vMem by 128 MB.

- `medium`

  Change the value by an increment constant that is 1/4 of the difference between the current value and the optimal value. For example, if the current vMem is 2 GB and the optimal vMem is 10 GB, then Workload Optimization Manager will raise vMem to 4 GB (or as close to that as the increment constant will allow).

- `high`

  Change the value to be the optimal value. For example, if the current vMem is 2 GB and the optimal vMem is 8 GB, then Workload Optimization Manager will raise vMem to 8 GB (or as close to that as the increment constant will allow).

Aggressiveness and observation period

Settings	Example

`observationPeriod:`
`  min:`
`  max:`
 	
`aggressiveness:`
	`observationPeriod:`
	`  min: 1d`
	`  max: 30d`
	
	`aggressiveness: p99`

Workload Optimization Manager uses these settings to calculate utilization percentiles for vCPU and vMem. It then recommends actions to improve utilization based on the observed values for a given time period.

Aggressiveness

When evaluating vCPU and vMem performance, Workload Optimization Manager considers resource utilization as a percentage of capacity. The utilization drives actions to scale the available capacity either up or down. To measure utilization, the analysis considers a given utilization percentile. For example, assume a 99th percentile. The percentile utilization is the highest value that 99% of the observed samples fall below. Compare that to average utilization, which is the average of *all* the observed samples.

Using a percentile, Workload Optimization Manager can recommend more relevant actions. This is important in the cloud, so that analysis can better exploit the elasticity of the cloud. For scheduled policies, the more relevant actions will tend to remain viable when their execution is put off to a later time.

For example, consider decisions to reduce the capacity for vCPU on a container. Without using a percentile, Workload Optimization Manager never resizes below the recognized peak utilization. For most containers there are moments when peak vCPU reaches high levels. Assume utilization for a container peaked at 100% just once. Without the benefit of a percentile, Workload Optimization Manager will not reduce allocated vCPU for that container.

With **Aggressiveness**, instead of using the single highest utilization value, Workload Optimization Manager uses the percentile you set. For the previous example, assume a single vCPU burst to 100%, but for 99% of the samples vCPU never exceeded 50%. If you set **Aggressiveness** to 99th Percentile, then Workload Optimization Manager can see this as an opportunity to reduce vCPU allocation for the container.

In summary, a percentile evaluates the sustained resource utilization, and ignores bursts that occurred for a small portion of the samples. You can think of this as aggressiveness of resizing, as follows:

- `p100`

  Least aggressive, recommended for critical workloads that need maximum guaranteed performance at all times

- `p99`

  Recommended setting to achieve maximum performance

- `p90`

  Most aggressive, recommended for non-production workloads that can stand higher resource utilization

Max Observation Period

To refine the calculation of resource utilization percentiles, you can set the sample time to consider. Workload Optimization Manager uses historical data from up to the number of days that you specify as a sample period. (If the database has fewer days' data then it uses all of the stored historical data.)

A shorter period means there are fewer data points to account for when Workload Optimization Manager calculates utilization percentiles. This results in more dynamic, elastic resizing, while a longer period results in more stable or less elastic resizing. You can make the following settings:

- `90d` – Less elastic
- `30d` – Recommended
- `7d` – More elastic

Min Observation Period

This setting ensures historical data for a minimum number of days before Workload Optimization Manager will generate an action based on the percentile set in **Aggressiveness**. This ensures a minimum set of data points before it generates the action.

Especially for scheduled actions, it is important that resize calculations use enough historical data to generate actions that will remain viable even during a scheduled maintenance window. A maintenance window is usually set for "down" time, when utilization is low. If analysis uses enough historical data for an action, then the action is more likely to remain viable during the maintenance window.

- None (empty) – More elastic
- `1d` – Recommended
- `3d` or `7d` – Less elastic

Max CPU throttling tolerance

Setting	Example
`cpuThrottlingTolerance:`	`cpuThrottlingTolerance: 20%`

This value defines your acceptable level of throttling and directly impacts the resize actions generated on CPU limits.

A low percentage value indicates more sensitivity to throttling, while a high value indicates more tolerance for throttling and a higher risk of congestion.

Learn more about CPU throttling here.

Resize behavior

Setting	Example
`behavior:` `  resize:`	`behavior:` `  resize: Manual`

Specify the degree of automation for the generated resize actions.

To turn on resizes, set the value to Automatic, Manual, or Recommend. To turn off, set the value to Disabled. See Action Acceptance Modes *(on page 848)* for details.

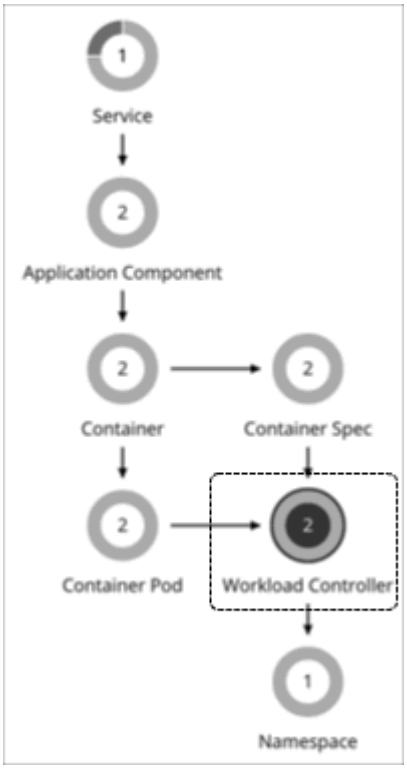
Workload controller

A workload controller is a container platform entity that watches the state of your pods and then requests changes where needed. Examples of workload controllers are `Deployments` and `StatefulSets`. A single workload controller can contain

one or more container spec entities, and it can be related to one or more running replica pods. Like container specs, workload controllers are persistent.

You can execute actions to resize container specs or scale replicas when you set the scope to a workload controller.

Synopsis



Synopsis	
Provides:	N/A
Consumes:	N/A
Discovered through:	Kubeturbo agent that you deployed to your cluster

Monitored resources

Workload Optimization Manager monitors quotas (limits and requests) for VCPU and VMem, and associates how much each workload controller is contributing to a quota based on all replicas. This allows Workload Optimization Manager to generate rightsizing decisions, and manage the quota as a constraint to rightsizing. Metrics on resource consumption are shown in the Container Spec, Container, and Container Pod views.

Workload controller actions

Workload Optimization Manager supports the following actions:

- Resize

  Actions associated with a workload controller resize container specs vertically to assure optimal utilization of resources. This is a natural representation of these actions because the parent controller's container specs for vCPU limits/requests and vMem limits/requests are modified. The workload controller then rolls out the changes in the running environment.

  For details, see Workload Controller Resize Actions *(on page 593)*.

  Resize actions for certain types of workloads may be blocked or set to recommend-only, which means that they are currently not executable in Workload Optimization Manager. These actions usually require a dependent action or an external configuration to be completed first. For example, a workload controller resize up action may be temporarily

blocked if there is insufficient resources in the corresponding node or namespace. In this case, a node provision or namespace quota resize action must be completed first. For more information, see this topic *(on page 291)*.

- Scale

  Actions associated with a workload controller scale replicas horizontally to maintain Service Level Objectives (SLOs) for your applications. This is a natural representation of these actions because the parent controller's number of replicas are modified. The workload controller then rolls out the changes in the running environment.

  For details, see Workload Controller Scale Actions *(on page 598)*.

Workload controller resize actions

Actions associated with a workload controller resize container specs vertically to assure optimal utilization of resources. This is a natural representation of these actions because the parent controller's container specs for vCPU limits/requests and vMem limits/requests are modified. The workload controller then rolls out the changes in the running environment.

Resize up actions in response to CPU throttling

For vCPU limit resizes, Workload Optimization Manager will recommend a resize up action, even if utilization percentile is low, to address slow response times associated with CPU throttling.

CPU throttling occurs when you configure a CPU limit on a container, which can inadvertently slow your applications' response time. Even if you have more than enough resources on your underlying node, your container workload will still be throttled because it was not configured properly. High response times are directly correlated to periods of high CPU throttling.

Learn more about CPU throttling here.

Especially for sudden throttling spikes, Workload Optimization Manager will persist the related resize actions so you can evaluate these actions even after the spikes have gone away, and then execute them to prevent spikes from re-occurring. As throttling drops, Workload Optimization Manager will not recommend a resize down action immediately, as this could result in subsequent back-and-forth upsize and downsize recommendations. Instead, it evaluates past throttling to decide when a resize down action is finally safe to execute. To ensure the timeliness of these actions and arrive at the optimal resize values to recommend, Workload Optimization Manager calculates fast and slow moving throttling averages, and then displays *smoothed* and *daily* averages in charts.

Smoothed average is an exponential moving average and moving variance method used on CPU throttling data. It allows analysis to generate a vCPU limit resize up action more quickly when throttling is detected, and be conservative on resize down to mitigate introducing throttling.

Resizing resource limits and requests

Workload Optimization Manager can recommend resizing vCPU/vMem limits and requests. For example, if workloads are allocated more than the requests that they actually use, Workload Optimization Manager recommends a **resize down** action to improve efficiency. If these workloads experience sustained increases in utilization, a **resize up** action is necessary to avoid performance risks.

> **NOTE:**
> For a resize up action, if the corresponding node does not have sufficient resources to execute the action, Workload Optimization Manager temporarily blocks the resize up action and generates a prerequisite action to move pods *(on page 606)* or provision a node *(on page 628)*. If the resize up action exceeds the namespace quota, the prerequisite action is to resize up the quota *(on page 616)*. After the prerequisite action is completed, the resize up action is allowed to execute.

To generate optimal resize actions, Workload Optimization Manager takes into account any analysis properties for resource limits and requests in container spec policies *(on page 581)*. Analysis properties include capacity ranges and target utilization midpoints.

Workload Optimization Manager also checks if requests and limits are specified for workloads.

- If requests and limits are specified, requests can be resized up or down based on actual utilization and target utilization. However, requests cannot exceed limits.
- If limits are specified but requests are not specified, Kubernetes sets requests to be the same value as limits when scheduling a pod. In other words, the capacities of both commodities are equal. Since requests cannot exceed limits, requests can only be resized down based on actual utilization and target utilization.
- If requests are specified but limits are not specified, a limit with a request is not necessary. Workload Optimization Manager does not recommend resizing limits.
- If requests and limits are not specified, Workload Optimization Manager does not recommend resizing requests or limits.

The following table describes the preceding scenarios in detail.

Item	Resources specified			
	Limits and requests	Limits only	Requests only	No limits or requests
Limits commodity	Created	Created	Created	Created
Capacity	From specs	From specs	Node capacity	Node capacity
Resize up	Yes	Yes	No	No
Resize down	Yes	Yes	No	No
Requests commodity	Created	Created	Created	Not created
Capacity	From specs	Requests = limits (set by Kubernetes during scheduling)	From specs	N/A
Resize up	Yes	No, as long as requests = limits	Yes	No
Resize down	Yes	Yes	Yes	No

Action propagation

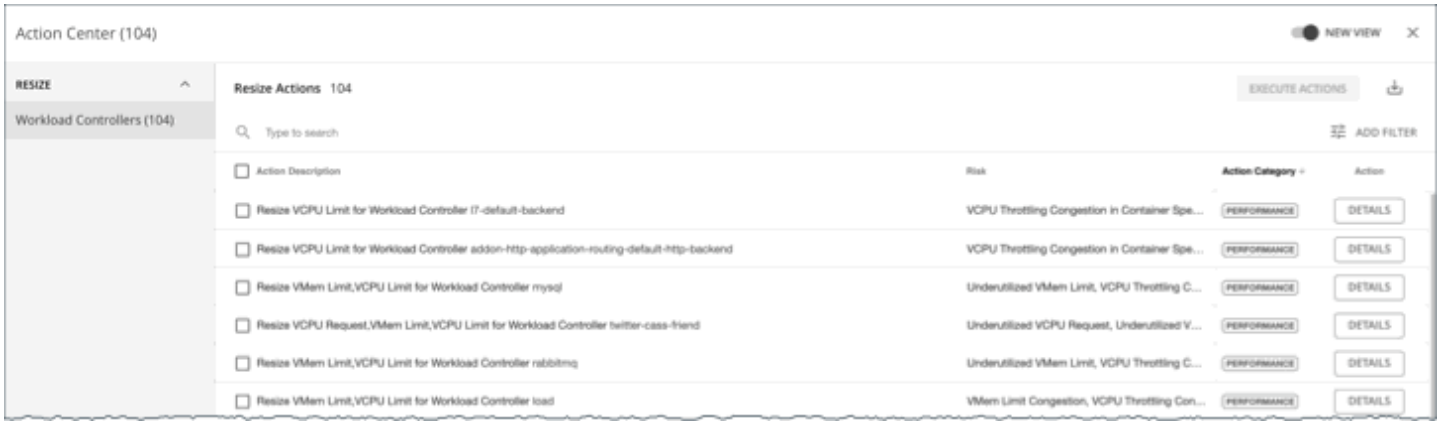
Resize actions propagate to application entities and the underlying container infrastructure to show the impact of these actions on the health of your applications and container environment.

Action merging

Executing several resize actions via workload controllers can be very disruptive since pods need to restart with each resize. For replicas of the container scale group(s) related to a single workload controller, Workload Optimization Manager consolidates resize actions into one *merged action* to minimize disruptions. When a merged action has been executed (via the associated workload controller), all resizes for all related container specifications will be changed at the same time, and pods will restart once.

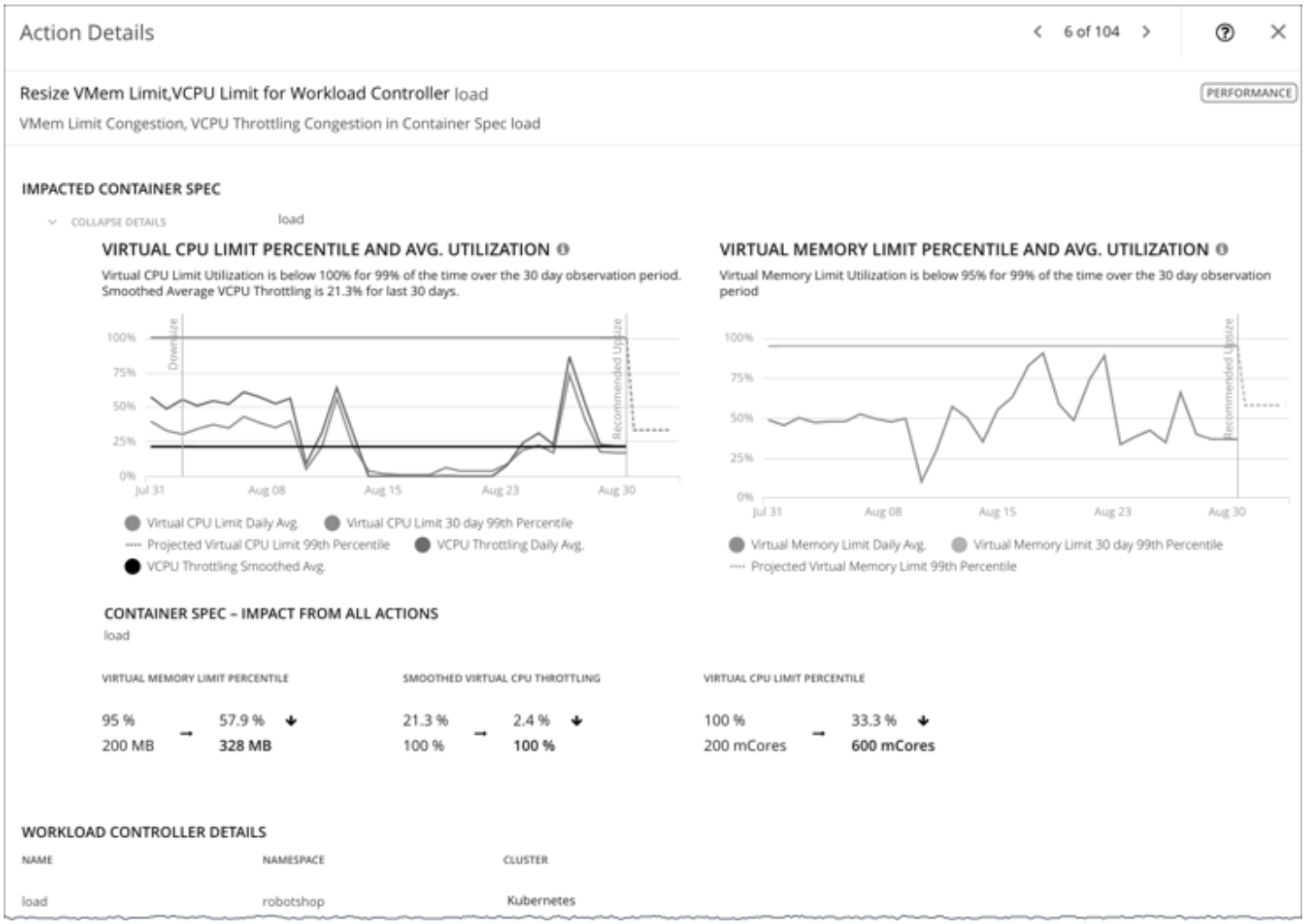
Action execution

After you set the scope to workload controllers, go to Action Center to see the full list of resize actions that you can execute. This list includes individual and merged actions. You can filter the list to focus on specific actions, such as actions to address resource congestion or vCPU throttling.



By default, resize actions are set in *Manual* mode at the workload controller level. This means that Workload Optimization Manager will not execute any action automatically, and you can manually select the actions that you want to execute. If you prefer to execute actions outside Workload Optimization Manager, create workload controller policies and set the resize action acceptance mode to *Recommend*. To automate actions, create workload controller policies and set the resize action acceptance mode to *Automatic*.

For each action, click DETAILS and expand the Details section to view time series charts that explain the reason for the action. These charts highlight *utilization percentiles* and *smoothed throttling averages* for a given observation period. Workload Optimization Manager uses percentile calculations to make accurate resize decisions.



These charts also:

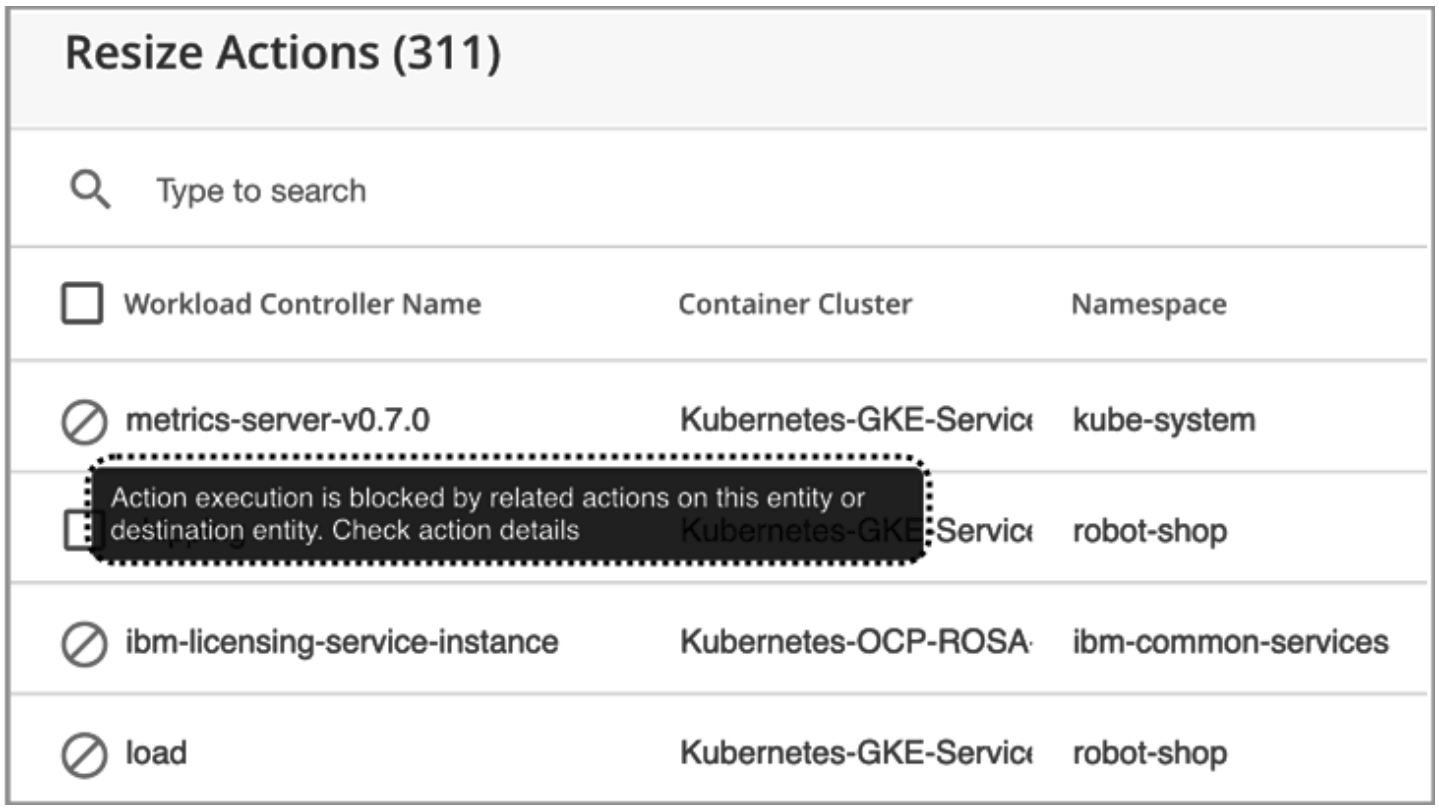
- Plot daily average percentiles and throttling, for your reference.
- Show projected percentiles after you execute the action. If you have previously executed resize actions on the same workload controller, the charts show the resulting improvements in daily average utilization.

Put together, these charts allow you to easily recognize trends that drive Workload Optimization Manager's resize recommendations.

NOTE:
You can set scaling constraints in container spec policies to refine the percentile calculations. For details, see Aggressiveness and Observation Periods *(on page 586)*.

Blocked and recommend-only actions

Resize actions for certain types of workloads may be blocked or set to recommend-only, which means that they are currently not executable in Workload Optimization Manager. These actions usually require a dependent action or an external configuration to be completed first. For example, a workload controller resize up action may be temporarily blocked if there is insufficient resources in the corresponding node or namespace. In this case, a node provision or namespace quota resize action must be completed first. For more information, see this topic *(on page 291)*.

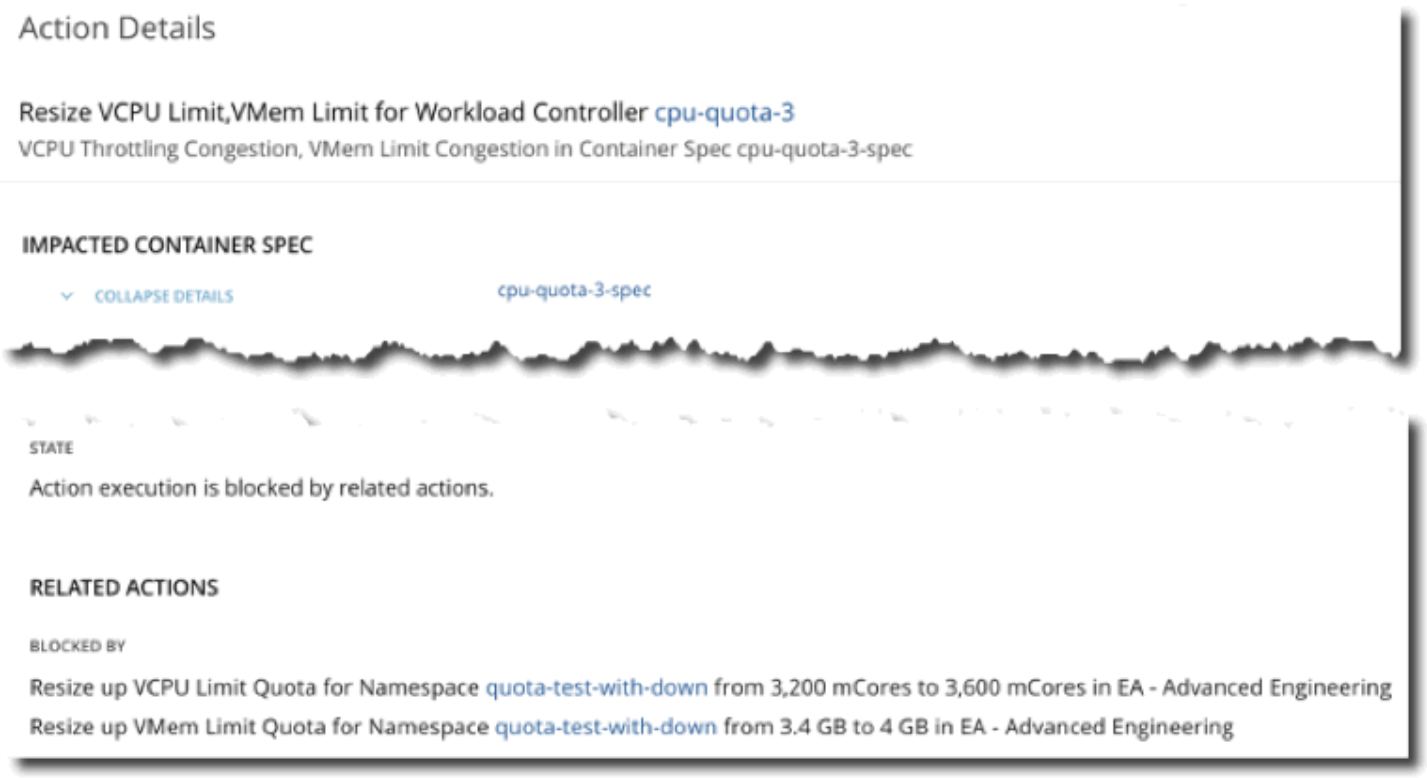


Blocking of action execution due to insufficient quotas

Workload Optimization Manager treats quotas defined in a namespace as constraints when making resize decisions. If existing actions would exceed the namespace quotas, Workload Optimization Manager recommends actions to resize up the affected namespace quota.

NOTE:
For more information about namespace quotas, see Namespace Monitored Resources *(on page 614)*.

The action details for a workload controller include descriptions of the affected container spec entities, and how the resources will change for each. If the resize exceeds current namespace quotas, then Workload Optimization Manager blocks the workload controller action. The action details list the Namespace actions that block execution of this resize in the **Related Actions** list.



Downloading data for executed resize actions

You can view and download data for executed resize actions to gain insight into workloads with performance risks due to resource limits, as well as opportunities to reclaim unused resource requests.

1. Add the Actions chart to your dashboard.
2. Configure the chart. Be sure to set the scope to workload controllers to narrow the results, select Tabular as the chart type, and then set the filter to either All Actions or Executed Actions.
3. In the chart, click **Show All**.
4. In the page that opens, click the download button. If you set All Actions as the filter, select **Executed Actions**.
5. Open the downloaded file and go to the Resize Workload Controller sheet.

   Pay attention to the following columns:

   - Namespace, Container Spec, and Container Platform Cluster columns – Data in these columns is the most up-to-date at the time of download.
   - Change column – This column highlights the impact of resize actions on the resources allocated to containers. Each value represents the difference between the Current Value and New Value columns.
   - Container Spec column – If there is more than one container spec in a workload controller, you can sort this column to easily identify resize actions on individual container specs.
   - Impacted Commodity column – This column shows if VMem or VCPU was resized. A resize action might show both commodities, which indicates that it is a *merged* action (an action that consolidates resizes related to a single workload controller, to minimize disruptions).

Workload controller scale actions

Actions associated with a workload controller scale replicas horizontally to maintain Service Level Objectives (SLOs) for your applications. This is a natural representation of these actions because the parent controller's number of replicas are modified. The workload controller then rolls out the changes in the running environment.

For example, when current response time for an application is in direct violation of SLO, Workload Optimization Manager will recommend increasing the number of replicas to improve response time. If applications can meet SLOs using less resources, Workload Optimization Manager will recommend reducing the replica count to improve infrastructure efficiency.

Action generation requirements

Workload Optimization Manager generates workload controller scale actions under the following conditions:

NOTE:
For GenAI LLM inference workloads, see the next section.

- Services are discovered by the Kubeturbo *(on page 234)* agent that you deployed to your cluster.
- Application performance metrics for services are collected through the Instana *(on page 203)* or Dynatrace *(on page 194)* target .
- You have created service policies *(on page 570)* and configured SLOs for Response Time and Transactions in those policies.

Scale actions for GenAI LLM inference workloads

For GenAI large language model (LLM) inference workloads that use GPU resources and are deployed in a Kubernetes cluster, Workload Optimization Manager generates workload controller scale actions to maintain SLOs for the following GPU metrics:

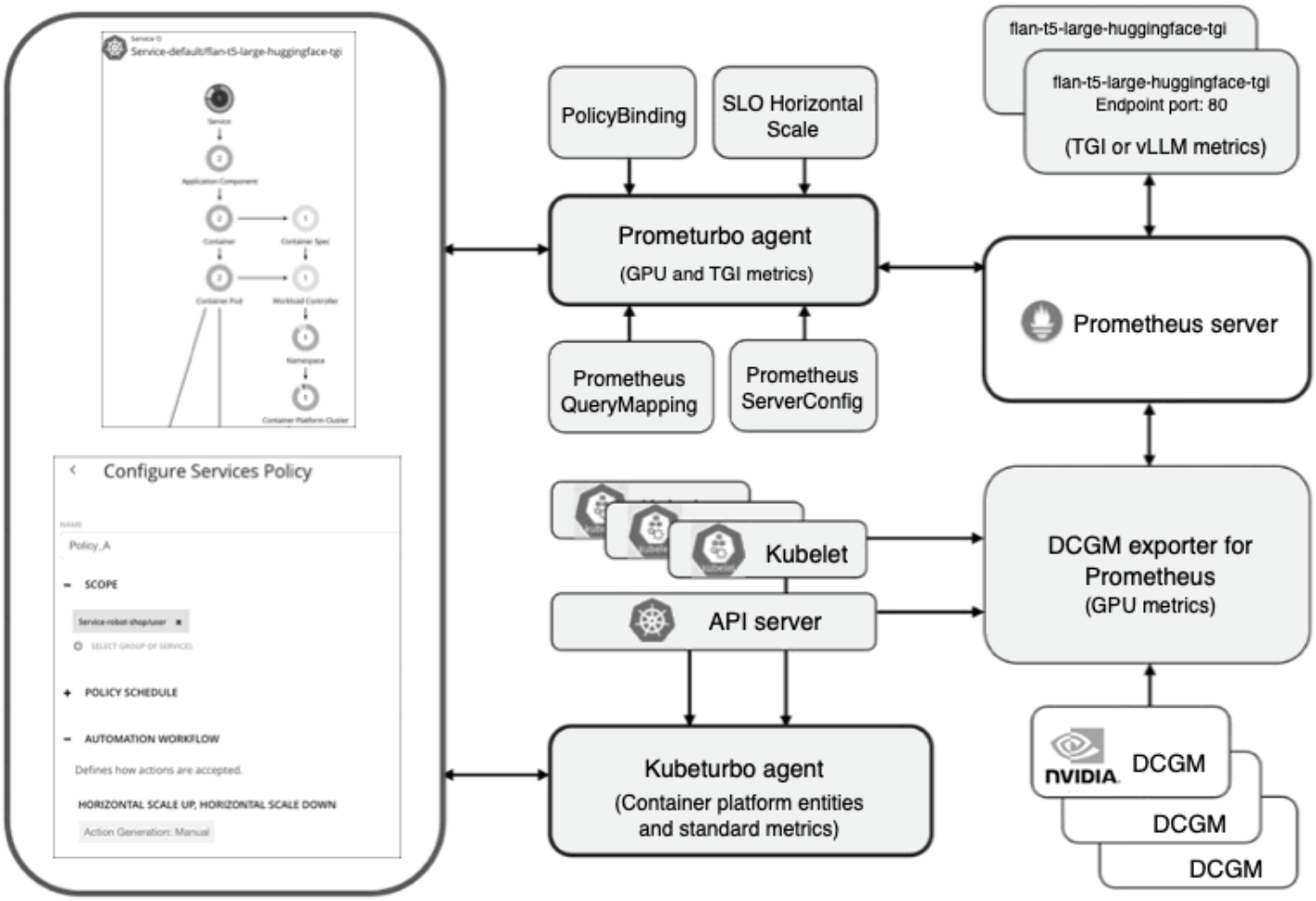
- Concurrent Queries
- Queueing Time
- Service Time
- Response Time
- Transactions

NOTE:
It is assumed that you have set up an LLM inference service on a cluster that has an NVIDIA GPU attached.

If you set up KServe in your environment, scroll to the **KServe Support** section in this topic for information about the level of support for KServe.

Requirements for generating scale actions for GenAI LLM inference Workloads

The following diagram illustrates how the components listed in the next table work together to support scale actions for LLM inference workloads. To configure these components properly, review the requirements listed in the table.



Component	Requirements
NVIDIA DCGM (Data Center GPU Manager)	NVIDIA DCGM is deployed as DaemonSet pods and collects GPU metrics. DCGM exposes these metrics as APIs.
DCGM exporter for Prometheus	DCGM exporter for Prometheus is deployed as DaemonSet pods and collects GPU metrics from DCGM. DCGM exporter exposes the data for Prometheus to scrape, connects to the Kubelet pod resources API to identify GPU devices associated with a container pod, and then appends the GPU devices to the metrics.
TGI (Text Generation Inference) or vLLM metrics	TGI or vLLM metrics are exposed directly by the LLM-serving services on predefined ports.
Prometheus server	The Prometheus server is configured to scrape both GPU and TGI metrics from DCGM exporter and TGI service endpoint. The Prometheus server makes these metrics available through PromQL queries.
Kubeturbo agent	Kubeturbo *(on page 234)* is deployed to your cluster. Kubeturbo monitors container platform entities and collects standard metrics for these entities.

Component	Requirements
Prometurbo agent	Prometurbo *(on page 515)* is deployed to your cluster and Prometurbo metrics collection *(on page 519)* is enabled. Prometurbo connects to the Prometheus server and sends PromQL queries to collect GPU and TGI metrics. Prometurbo requires these CRs: ▪ `PrometheusQueryMapping` – specifies the GPU and TGI metrics to collect ▪ `PrometheusServerConfig` – specifies settings for your Prometheus server
Workload Optimization Manager supply chain and charts	Workload Optimization Manager stitches the entities discovered from Prometurbo and Kubeturbo into the supply chain. When you set the scope to container platform entities, charts show GPU and TGI metrics. ▪ GPU metrics include GPU (utilization of Tensor cores) and GPU memory (utilization of framebuffer memory). ▪ TGI metrics include Concurrent Queries, Queueing Time, Service Time, Response Time, and Transactions. Workload Optimization Manager calculates 10-minute and 1-hour moving averages, and then uses the maximum of the two. This mechanism allows for faster generation of scale up actions, and slower or more conservative generation of scale down actions.
Workload Optimization Manager service policies	Service policies *(on page 570)* are created for the services associated with the LLM inference workloads. In these policies: ▪ Scope to the relevant services. Create a group of services from **Settings > Groups** and then specify the group as your scope. ▪ Turn on Horizontal Scale Down and Horizontal Scale Up actions. ▪ Enable SLOs for Concurrent Queries, Queueing Time, Service Time, Response Time, and Transactions, and then specify your preferred SLO values. Workload Optimization Manager generates workload controller scale actions to maintain the SLOs that you defined in the policies. See the next section for information about the generated actions.

Support for KServe

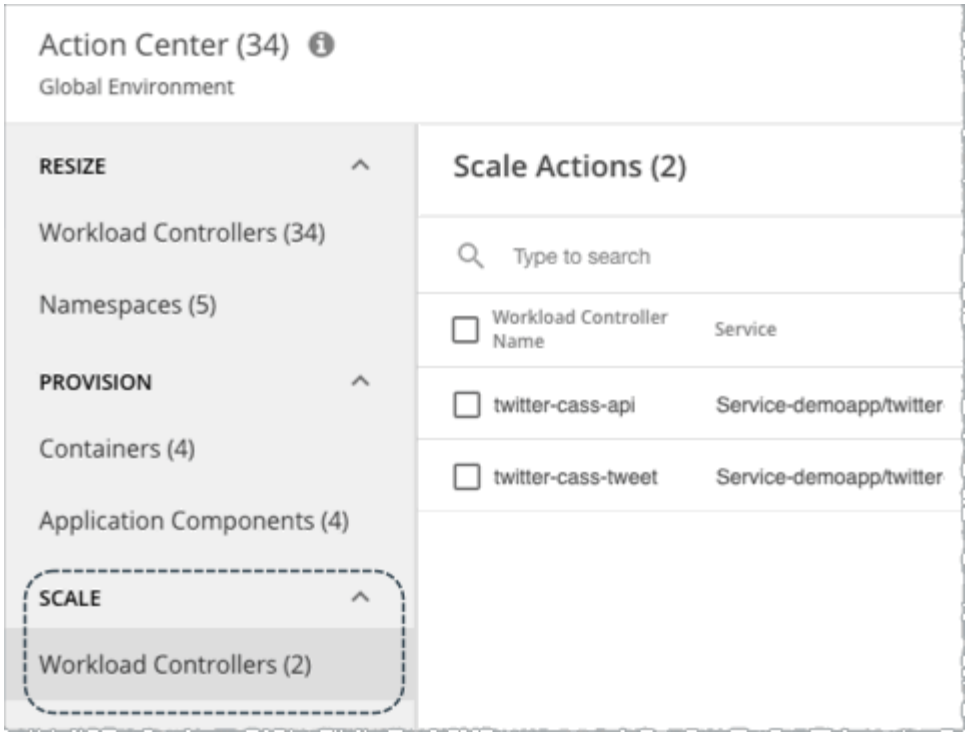
Workload Optimization Manager can automate scale actions for workloads served by KServe in Red Hat OpenShift AI. KServe is a scalable and standards-based Model Inference Platform on Kubernetes for Trusted AI. Inference services deployed by KServe are managed by `InferenceService` Custom Resources. The number of workload replicas in an `InferenceService` is dictated by the minimum and maximum replicas in the predictor, transformer, or explainer configuration in the `InferenceService` spec.

Workload Optimization Manager scale actions for `InferenceService` workloads are generated and executed as follows:

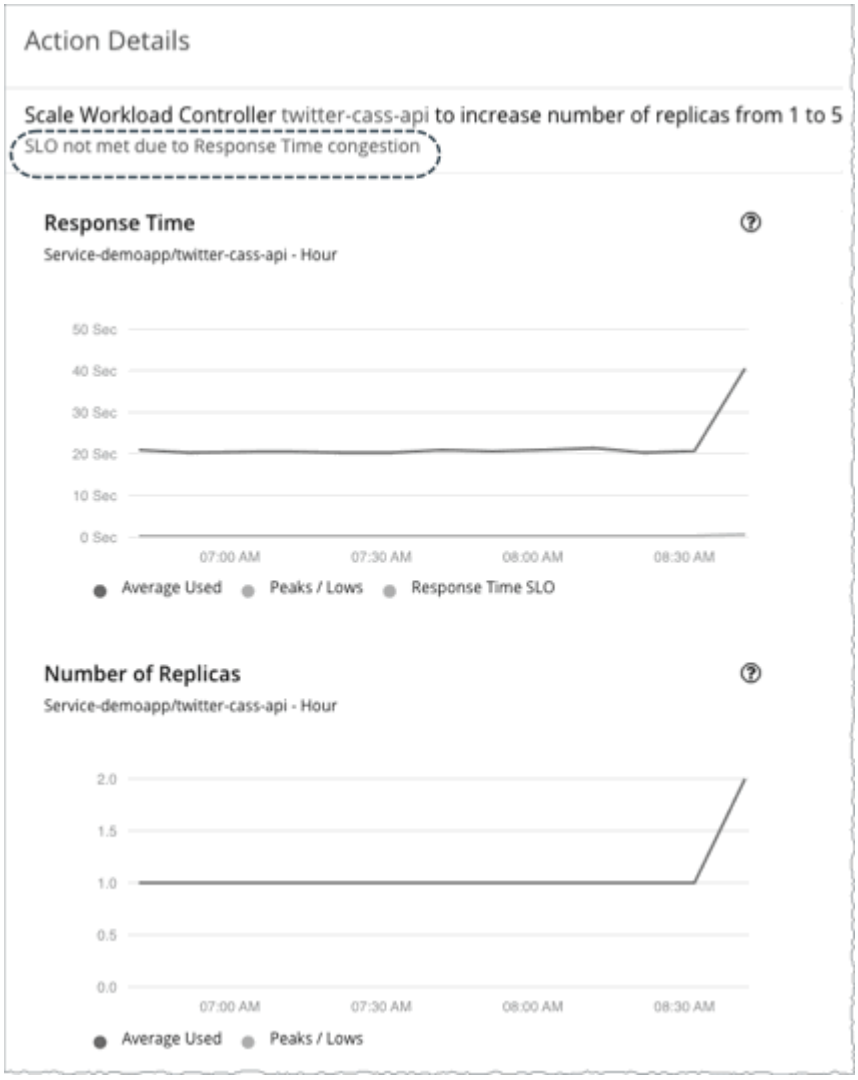
- Workload Optimization Manager automatically detects and updates the minimum and maximum replicas to the values configured in the `InferenceService` spec, instead of changing the `spec.replicas` of the Deployment directly. Since replicas are automatically updated, you do not need any additional setup or configuration in KServe.
- The KServe controller scales the workloads accordingly.

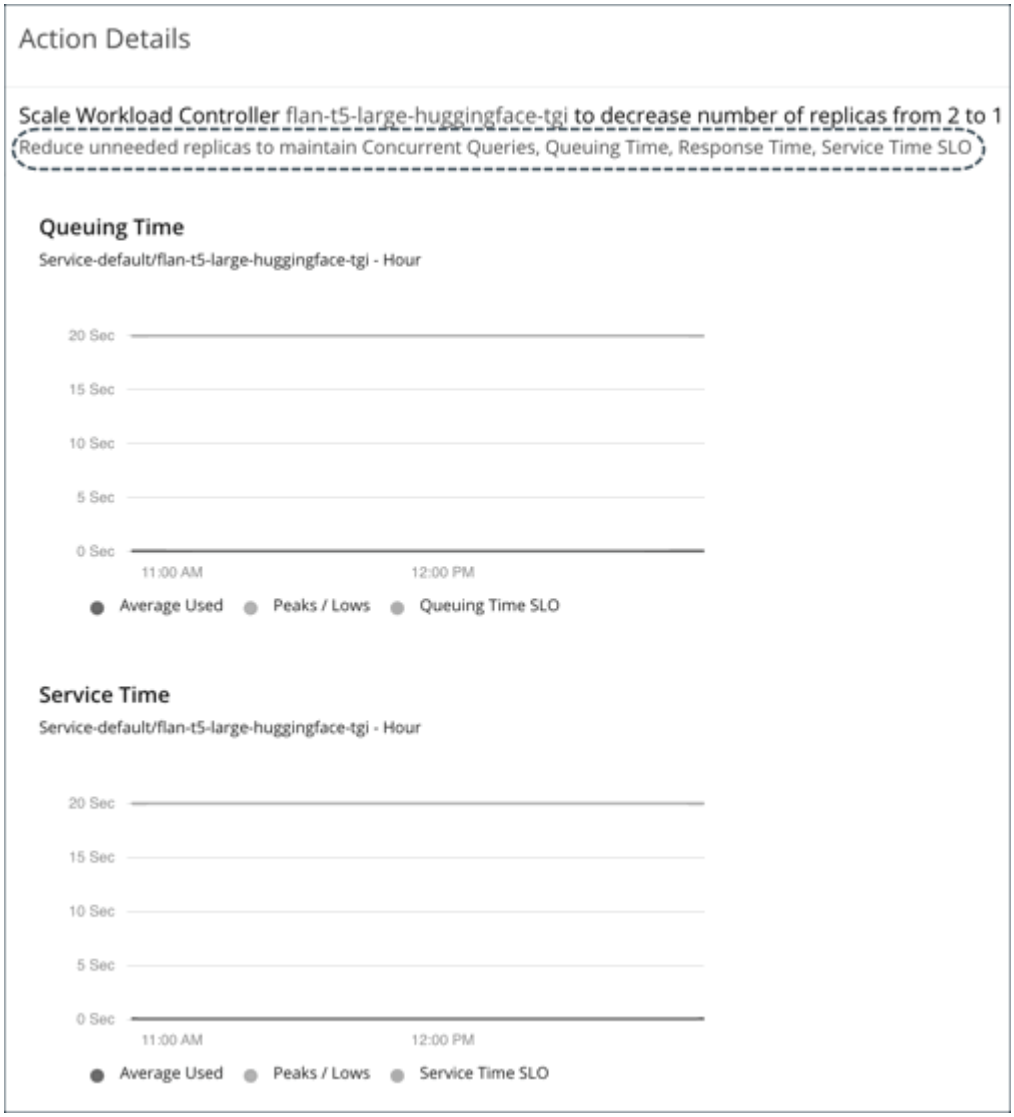
Action visibility

Workload Optimization Manager shows and executes SLO-driven scale actions through workload controllers. A single scale action represents the total number of replicas that you need to scale in or out to meet your SLOs.



When you examine an action, SLO is indicated as the reason for the action.


Action Details
Scale Workload Controller twitter-cass-api to increase number of replicas from 1 to 5
SLO not met due to Response Time congestion
Response Time
Service-demoapp/twitter-cass-api - Hour
50 Sec
40 Sec
30 Sec
20 Sec
10 Sec
0 Sec
07:00 AM
07:30 AM
08:00 AM
08:30 AM
Average Used
Peaks / Lows
Response Time SLO
Number of Replicas
Service-demoapp/twitter-cass-api - Hour
2.0
1.5
1.0
0.5
0.0
07:00 AM
07:30 AM
08:00 AM
08:30 AM
Average Used
Peaks / Lows



Workload controller policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

Actions associated with a workload controller resize container specs vertically to assure optimal utilization of resources. This is a natural representation of these actions because the parent controller's container specs for vCPU limits/requests and vMem limits/requests are modified. The workload controller then rolls out the changes in the running environment.

Points to consider:

- Scale actions

  Workload controller policies do not control scale actions. These actions are generated under the conditions described in this topic *(on page 598)*.

- Resize actions

  Container spec policies control the *generation* of resize actions, while workload controller policies control the *execution* of these actions.

Action	Default mode
vCPU Limit Resize Down	Manual (automatable)
vCPU Limit Resize Up	Manual (automatable)
vCPU Request Resize Down	Manual (automatable)
vCPU Request Resize Up	Manual (not automatable)
vMem Limit Resize Down	Manual (automatable)
vMem Limit Resize Up	Manual (automatable)
vMem Request Resize Down	Manual (automatable)
vMem Request Resize Up	Manual (not automatable)

You can control action execution based on the specific resources that you want to resize and the resize direction. For example, for vCPU limit resizes, you may want to automate resize down actions, but require reviews of resize up actions and only execute the actions that receive an approval. To enforce these rules, create a workload controller policy and then set the action mode for vCPU Limit Resize Down to *automated*, and vCPU Limit Resize Up to *manual*.

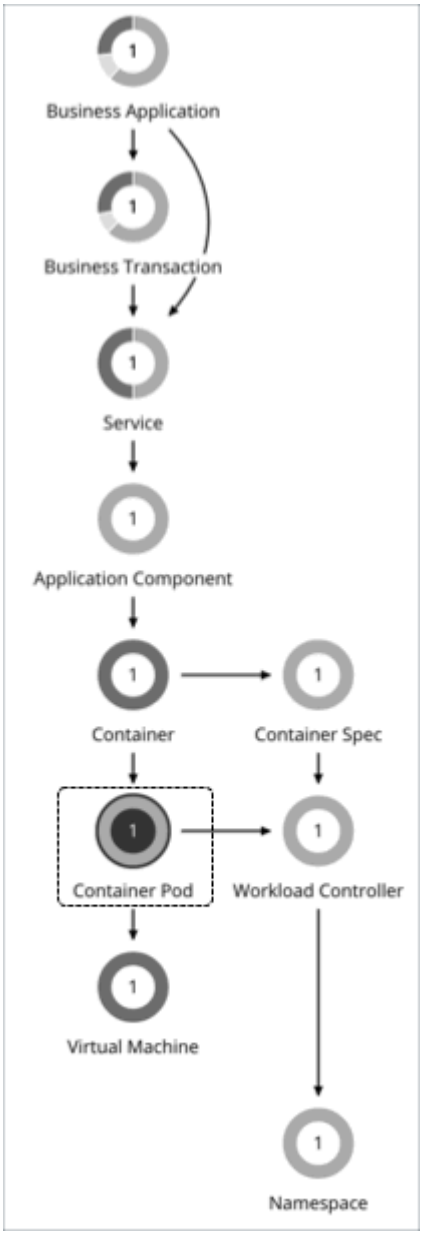
- Action orchestration is currently not supported.

Container pod

A container pod is a group of one or more containers with shared storage or network resources. It is the smallest deployable unit of computing that you can create and manage in container platform environments.

A container pod specifies how to run containers together. Like containers, a container pod is *ephemeral*.

Synopsis



Synopsis	
Provides:	Resources for containers to use, including: ▪ Virtual CPU ▪ Virtual Memory
Consumes:	Resources from virtual machines and namespaces
Discovered through:	Kubeturbo agent that you deployed to your cluster

Monitored resources

Workload Optimization Manager discovers pods with the following Status in Kubernetes and Red Hat OpenShift and matches them to a specific State in Workload Optimization Manager.

Kubernetes Status	Workload Optimization Manager State
Running	Active
`ImagePullBackOff`	Unknown
`CrashLoopBackoff`	Unknown
Error	Unknown

It is expected that the total number of pods in the cluster will not match in Workload Optimization Manager due to this. As Workload Optimization Manager also does not discover Pending pods or Completed job pods.

Workload Optimization Manager monitors the following resources:

- Virtual memory (vMem)

  Virtual memory (vMem) is the measurement of memory that is in use.
- vMem request

  vMem request is the virtual memory request allocated by a pod against the node allocatable capacity.
- Virtual CPU (vCPU)

  Virtual CPU is the measurement of CPU that is in use.
- vCPU request

  vCPU request is the virtual CPU request (in mCores) allocated by a pod against the node allocatable capacity.

Actions

Workload Optimization Manager supports the following actions:

- Move

  Move a pod that is controlled by a ReplicationController, ReplicaSet, or Deployment (through a ReplicaSet) to another node (VM) to address performance issues or improve infrastructure efficiency. For example, if a particular node is congested for CPU, you can move pods to a node with sufficient capacity. If a node is underutilized and is a candidate for suspension, you must first move the pods before you can safely suspend the node.

  For details, see Container Pod Move Actions *(on page 606)*.
- Provision or suspend

  When recommending node provision or suspend actions, Workload Optimization Manager will also recommend provisioning pods (based on demand from DaemonSets) or suspending the related pods.

  For details, see Container Pod Provision and Suspension Actions *(on page 610)*.

Container pod move actions

Move a pod that is controlled by a ReplicationController, ReplicaSet, or Deployment (through a ReplicaSet) to another node (VM) to address performance issues or improve infrastructure efficiency. For example, if a particular node is congested for CPU, you can move pods to a node with sufficient capacity. If a node is underutilized and is a candidate for suspension, you must first move the pods before you can safely suspend the node.

During a pod move, Workload Optimization Manager copies and persists all labels, except for the following labels with keys:

- `pod-template-hash`
- `deployment`
- `deploymentconfig`

If the pod move was successful, these labels are copied to the pod.

The following items impact the generation and execution of pod move actions:

Placement policies

You can create placement policies to enforce constraints for pod move actions. For example, you can have a policy that allows pods to only move to certain nodes, or a policy that prevents pods from moving to certain nodes.

For more information, see Creating Placement Policies *(on page 1044)*.

Taints and tolerations

Workload Optimization Manager considers taints and tolerations as constraints. For example, if a pod has a toleration attribute that restricts it from moving to a certain node, Workload Optimization Manager will not move that pod to the restricted node.

Node labels

Workload Optimization Manager imports node labels and treats them as constraints. For example, if a pod has a defined node label, Workload Optimization Manager will move that pod to a node with a matching label.

Auto-generated policies for pods

For the following workloads, Workload Optimization Manager auto-generates groups and policies.

- Pods with topology spread constraints that have a setting of `DoNotSchedule` for the `whenUnsatisfiable` constraint apply an auto-generated policy that disables pod move actions.
- StatefulSet pods apply an auto-generated policy that sets the action acceptance mode for pod move actions to recommend-only.

For information on the action handling mechanism for these pods in Workload Optimization Manager, see this topic *(on page 291)*.

Pod affinity and anti-affinity rules

Workload Optimization Manager recognizes pod affinity and anti-affinity rules.

Workload Optimization Manager uses a generic construct to represent entity-to-entity affinity and anti-affinity using commodities bough and sold by relevant entities. Special topologies, where the topology key used in the affinity or anti-affinity rule is *not* `hostname`, are handled through an internal provider entity that represents a group of nodes. These groups are created by processing the labels on the cluster nodes.

Workload Optimization Manager uses a proprietary mechanism to process the commodities such that the affinity and anti-affinity rules are respected while continuously driving the target environments toward the desired state. If the rules break because of changed labels or recreated pods in the Kubernetes cluster, or conflicting policies created by users in the Workload Optimization Manager user interface, Workload Optimization Manager generates move actions to alleviate the situation. To review these actions in Action Center, look for container pod move actions with an action category of **Compliance** and a risk of **Unsatisfied Affinity Constraints** or **Unsatisfied Anti-Affinity Constraints.**

The following table describes the scenarios where Workload Optimization Manager generates an action to move a source or destination pod to another provider. If there is no valid provider in the cluster, Workload Optimization Manager generates a reconfigure action.

Scenario	Kubernetes action	Workload Optimization Manager action
Source pod to destination pod affinity with the topology key `kubernetes.io/hostname` Source and destination pods are on the same node.	Originally schedule pods correctly by placing both pods on the same node.	No additional action
Source pod to destination pod affinity with the topology key `kubernetes.io/hostname`	Schedule pods correctly when the source was scheduled, but might	Generate an action to move the source or destination pod to the same node.

Scenario	Kubernetes action	Workload Optimization Manager action
Source and destination pods originally scheduled by Kubernetes are on the same node. Kubernetes restarted and scheduled the destination pod to another node without considering the source pods affinities during scheduling.	schedule pods incorrectly when only the destination pod was scheduled.	
Source pod to destination pod affinity with a topology key that is *not* `kubernetes.io/hostname` Source and destination pods are on the same topology.	Schedule pods correctly by placing both pods on node(s) on the same topology.	No additional action
Source pod to destination pod affinity with a topology key that is *not* `kubernetes.io/hostname` Source and destination pods originally scheduled by Kubernetes are on the same topology. Kubernetes restarted and scheduled the destination pod to another node without considering the source pods affinities during scheduling.	Schedule pods correctly when the source was scheduled, but might schedule pods incorrectly when only the destination pod was scheduled.	Generate an action to move the source or destination pod to nodes on the same topology.
Source pod to destination pod affinity with a topology key that is *not* `kubernetes.io/hostname` Source and destination pods originally scheduled by Kubernetes are on different nodes but are on the same topology. Labels on the nodes changed later, putting the nodes on two different topologies.	Pods remain incorrectly placed.	Generate an action to move the source or destination pod to node(s) on the same topology.
Source pod to destination pod anti-affinity with a topology key that is *not* `kubernetes.io/hostname` Source and destination pods are on different topologies.	Originally schedule pods correctly by placing pods on different topologies.	No additional action
Source pod to destination pod anti-affinity with a topology key that is *not* `kubernetes.io/hostname` Source and destination pods originally scheduled by Kubernetes are on node(s) on different topologies. Labels on the nodes changed later, putting the nodes on the same topology.	Pods remain incorrectly placed.	Generate an action to move the source or destination pod to node(s) on different topologies.
Source pod to destination pod anti-affinity with a topology key that is *not* `kubernetes.io/hostname` Source and destination pods originally scheduled by Kubernetes are on node(s) on different topologies. Labels	Pods remain incorrectly placed.	Generate a reconfigure action since there is no valid topology node that the source or destination pod can move to.

Scenario	Kubernetes action	Workload Optimization Manager action
on the nodes changed later, putting the nodes on the same topology. As a result, there is no valid topology node that the source or destination pod can move to. For example, nodes were distributed into `type=type1` and `type=type2`. The source pod is on a `type1` node while the destination pod is on a `type2` node. At some point, all `type2` nodes were changed to `type1`. This means that all cluster nodes are now on the `type=type1` topology.		

Eviction thresholds

Workload Optimization Manager considers the memory/storage eviction thresholds of the destination node to ensure that the pod can be scheduled after it moves. Eviction thresholds for `imagefs` and `rootfs` are reflected as node effective capacity in the market analysis.

Temporary quota increases

If a namespace quota is already fully utilized, Workload Optimization Manager temporarily increases the quota to allow a pod to move.

For details, see Namespace Actions *(on page 616)*.

Security context constraints (SCCs)

Red Hat OpenShift uses SCCs to control permissions for pods. This translates to permissions that users see within the containers of the pods, and the permissions for the processes running inside those pods.

When executing pod move actions, Kubeturbo normally runs with Red Hat OpenShift cluster administrator permissions to create a new pod and remove the old one. Because of this, the SCCs for the new pod are those that are available to a cluster administrator. It is therefore possible for the new pod to run with an SCC that has higher privileges than the old pod. For example, an old pod might have `restricted scc` access, while the new one might have `anyuid scc` access. This introduces a privilege escalation issue.

To prevent privilege escalation when moving pods, Kubeturbo enforces user impersonation, which carries the user-level SCCs of the old pod over to the new pod. To enforce user impersonation, Kubeturbo performs the following tasks:

- Create a user impersonation account for each SCC level.
- Create a service account and treat it as a user account for each SCC level currently running in a given cluster.
- Provide role-based access to SCCs used for impersonation via the service accounts. A service account is allowed to use only one SCC resource in the cluster.
- Create a `role binding` resource to allow service account access to a particular role.

All resources created to enforce user impersonation are removed when Kubeturbo shuts down.

Be aware that by default, an arbitrary pod running in a given cluster does not recognize the namespace it is configured to run in, which is a requirement for user impersonation enforcement. For Kubeturbo to recognize the namespaces for pods, it is recommended that you add an environment variable named KUBETURBO_NAMESPACE via the downward API. The Red Hat OpenShift standard installation methods add the following environment variable to the Kubeturbo deployment spec.

```
        env:
          - name: KUBETURBO_NAMESPACE
            valueFrom:
```

```
              fieldRef:
                fieldPath: metadata.namespace
```

With this environment variable, Kubeturbo can successfully create the resources needed to enforce user impersonation. Without this variable, Kubeturbo creates the resources in the namespace called `default`. This might cause issues if you need to run multiple instances of Kubeturbo in the same cluster. For example, one instance might run as an observer, and another as an administrator. To ensure multiple Kubeturbo instances within the same cluster do not conflict when creating and removing user impersonation resources, run the instances in separate namespaces.

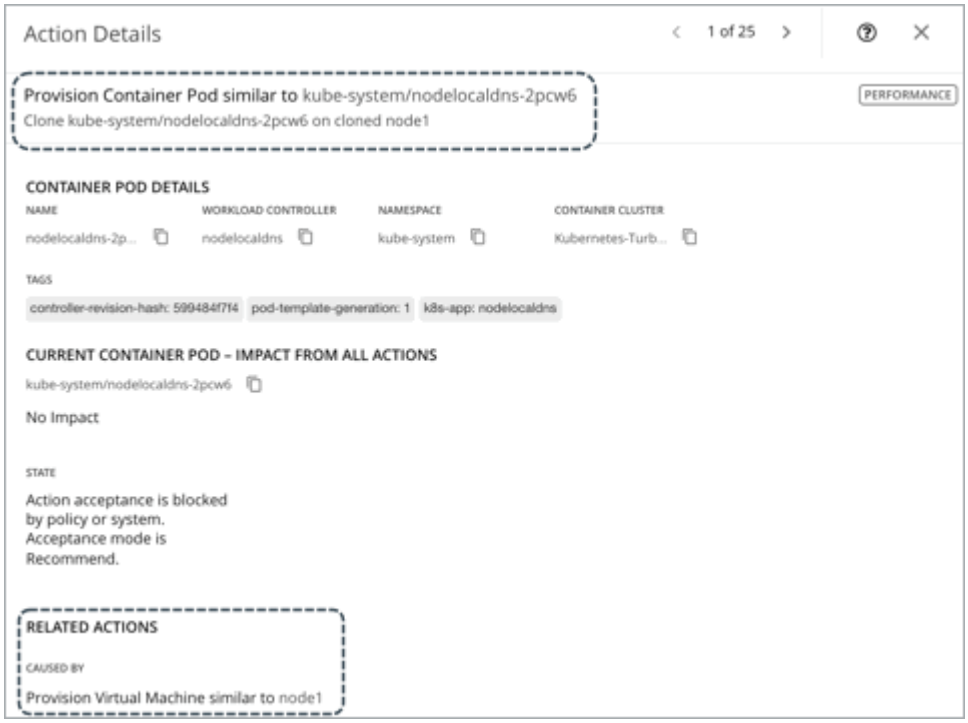
Container pod provision and suspension actions

When recommending node provision or suspend actions, Workload Optimization Manager will also recommend provisioning pods (based on demand from DaemonSets) or suspending the related pods.

Pod provision action in response to node provision

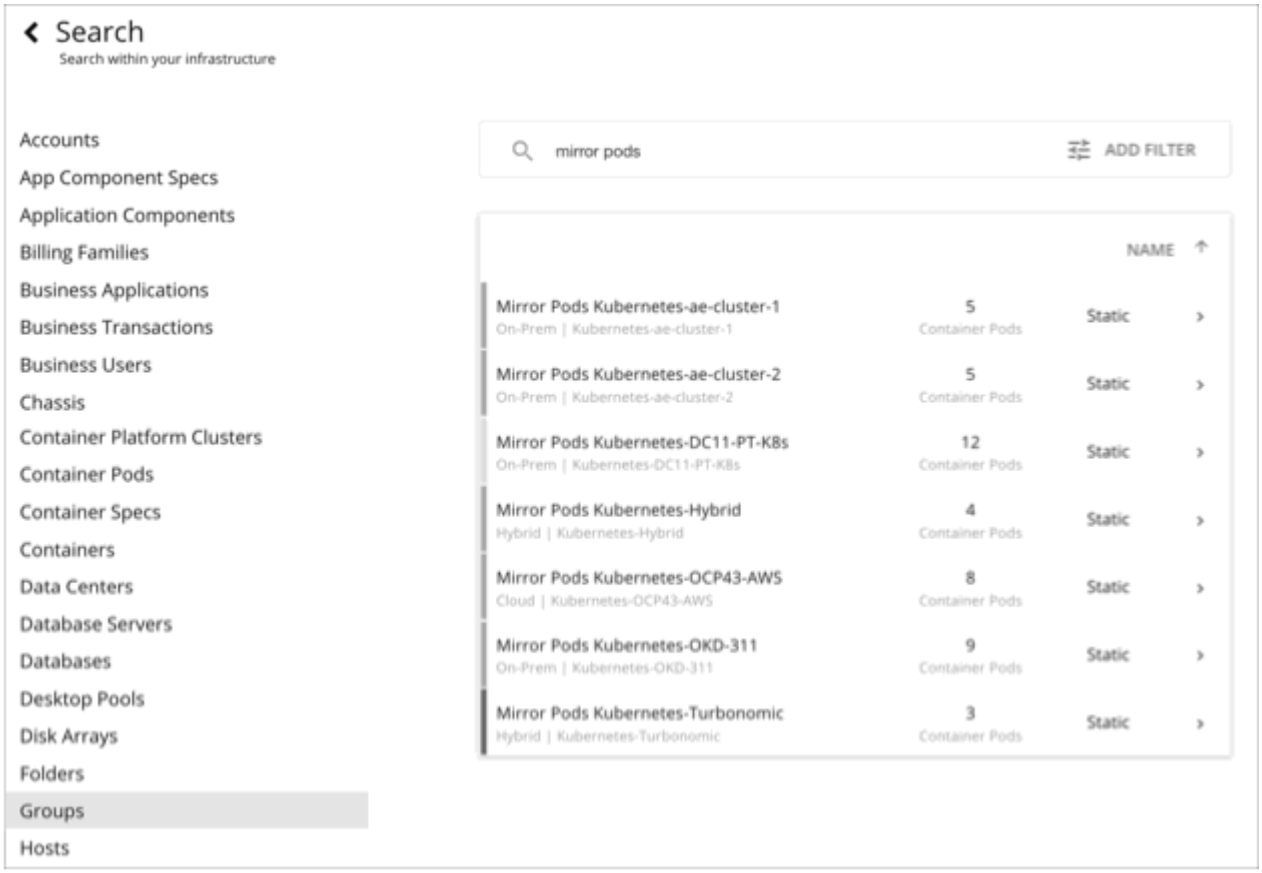
When recommending node provision actions, Workload Optimization Manager also recommends pod provision actions that reflect the projected demand from required DaemonSet pods, and respects the maximum number of pods allowed for a node. This ensures that any application workload can be placed on the new node and stay within the desired range of vMem/vCPU usage, vMem/vCPU request, and number of consumers.

The action details for a pod provision action shows the related node that you need to provision. Click the node name to set it at your scope.



Workload Optimization Manager treats static pods as DaemonSets for the purpose of provisioning nodes. Because a static pod provides a node with a specific capability, it is controlled by the node and is not accessible through the API server. If a node to be provisioned requires a static pod, Workload Optimization Manager generates actions to provision the node and the corresponding static pod.

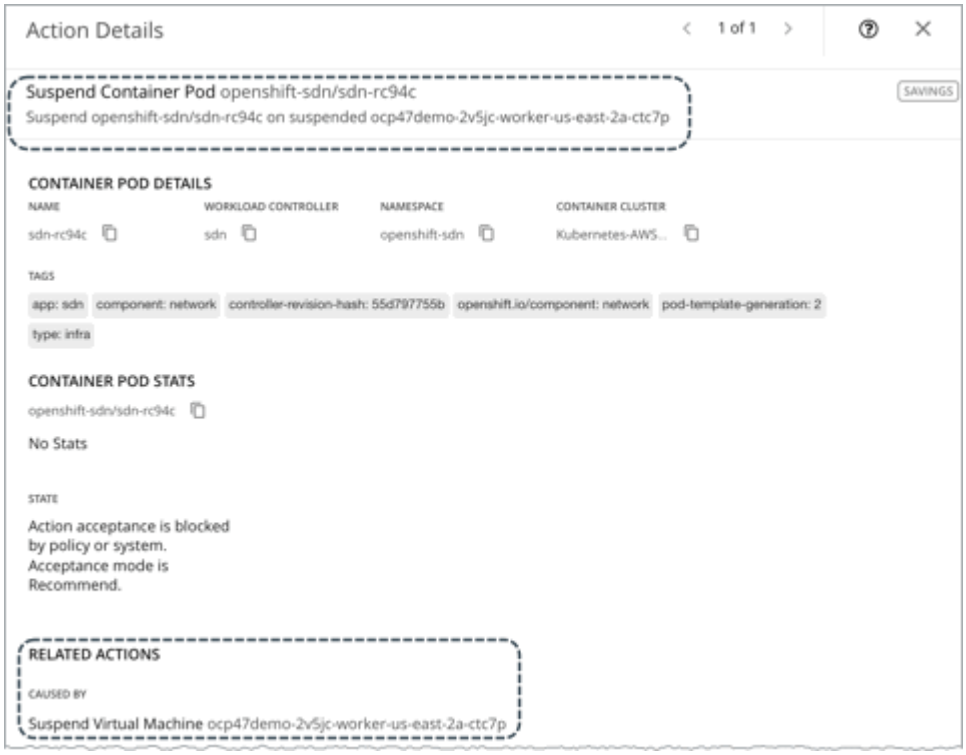
Workload Optimization Manager creates an auto-generated group of static pods when it discovers a static pod on each node in a cluster. To view all the auto-generated groups, go to Search, select Groups, and then type `mirror pods` as your search keyword.



Pod suspension action in response to node suspension

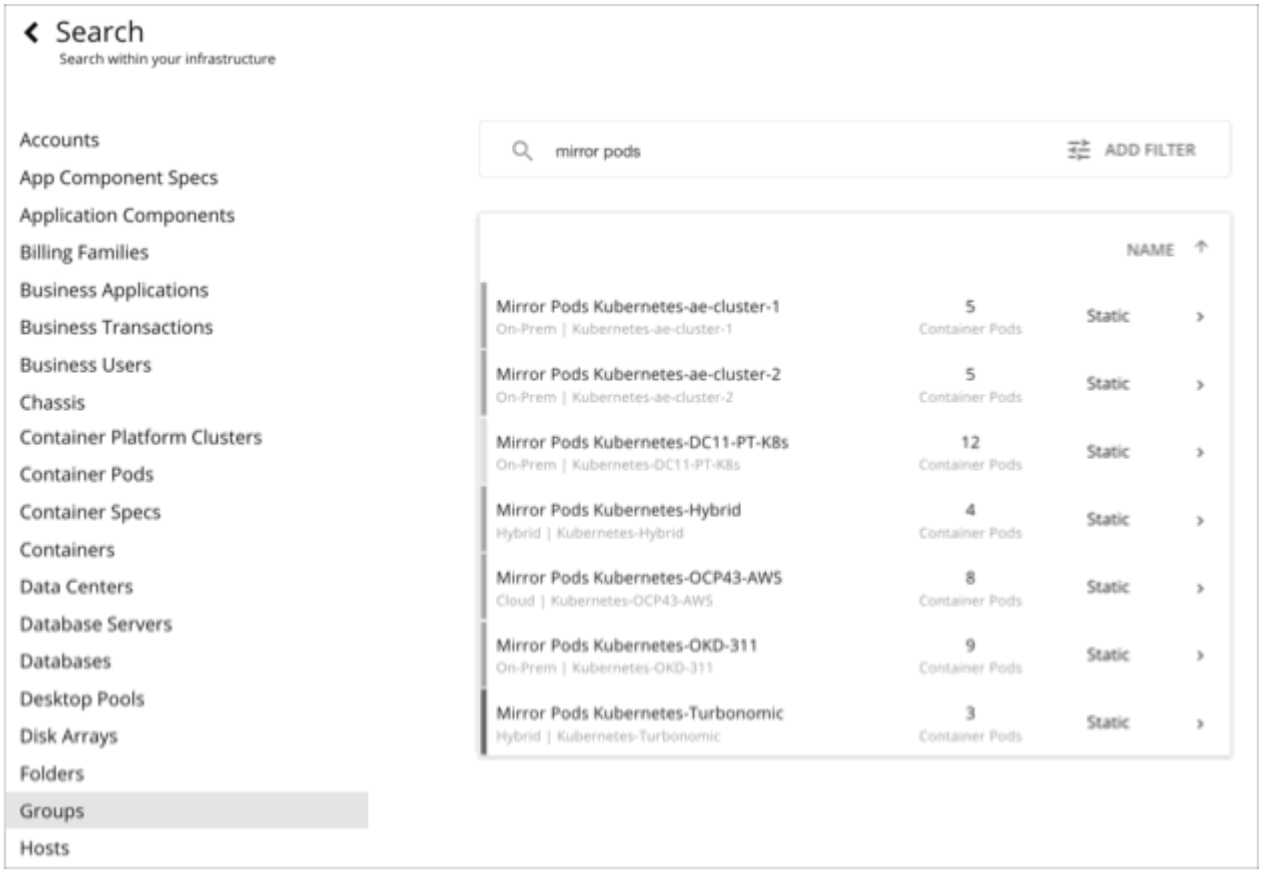
When recommending node suspension actions, Workload Optimization Manager also recommends suspending the DaemonSet pods that are no longer required to run the suspended nodes.

The action details for a pod suspension action shows the related node that you need to suspend. Click the node name to set it at your scope.



Workload Optimization Manager treats static pods as DaemonSets for the purpose of suspending nodes. Because a static pod provides a node with a specific capability, it is controlled by the node and is not accessible through the API server. If the only workload type on a node is a static pod, Workload Optimization Manager generates actions to suspend the node and the corresponding static pod.

Workload Optimization Manager creates an auto-generated group of static pods when it discovers a static pod on each node in a cluster. To view all the auto-generated groups, go to Search, select Groups, and then type `mirror pods` as your search keyword.



Container pod policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

For details about container pod actions, see Container Pod Move Actions *(on page 606)*.

Action	Default mode
Move	Manual

Action orchestration is currently not supported.

Placement policies

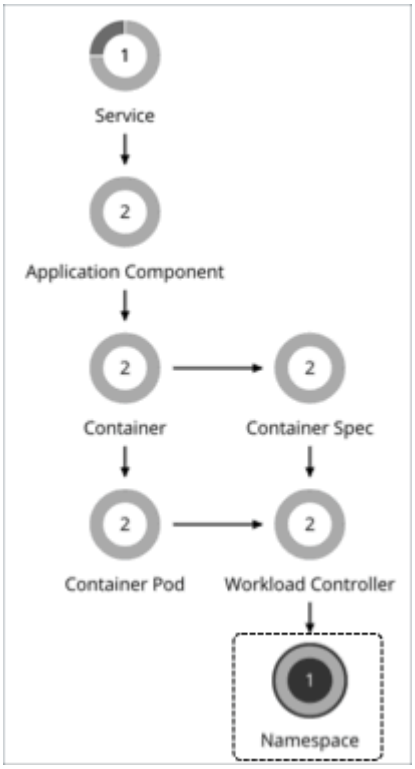
You can create placement policies to enforce constraints for pod move actions. For example, you can have a policy that allows pods to only move to certain nodes, or a policy that prevents pods from moving to certain nodes.

For more information, see Creating Placement Policies *(on page 1044)*.

Namespace

A namespace is a logical pool of resources in a container platform cluster that manages workloads based on specific requirements or business needs. For example, administrators can pool resources for different organizations within the enterprise, and assign different policies to each pool.

Synopsis



Synopsis	
Provides:	N/A
Consumes:	N/A
Discovered through:	Kubeturbo agent that you deployed to your cluster

Labels and annotations

Workload Optimization Manager discovers namespace labels and annotations as tag properties. You can filter namespaces by labels or annotations when you use Search or create Groups.

Annotations are not discovered by default. To enable discovery, configure the `annotationWhitelist` parameter in your `ConfigMap`. For more information about this parameter, see this topic *(on page 309)*.

Monitored resources

A namespace can include the following compute resource quotas:

- vMem request quota

  vMem request quota is the total amount of virtual memory request for all pods allocated to the namespace against the namespace quota.

- vCPU request quota

  vCPU request quota is the total amount of virtual CPU request (in mCores) for all pods allocated to the namespace against the namespace quota.

- vMem limit quota

  vMem limit quota is the total amount of virtual memory limit for all pods allocated to the namespace against the namespace quota.

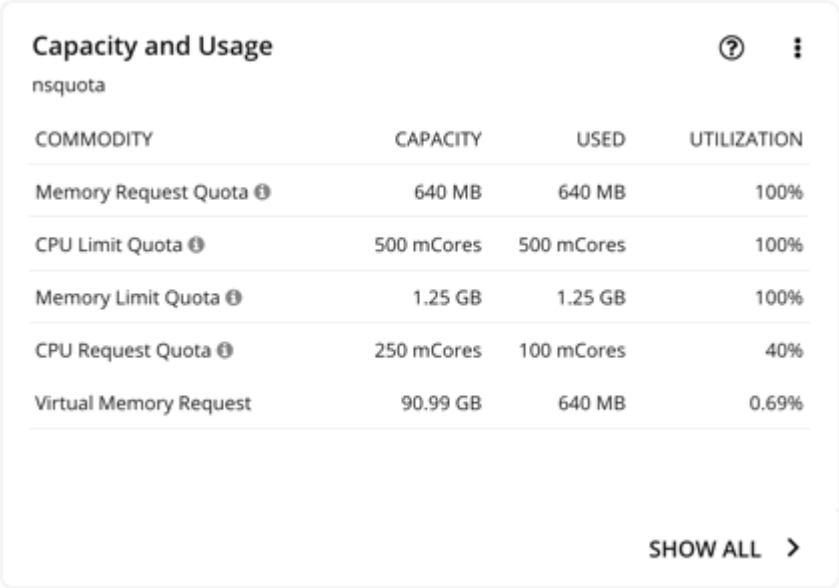
- vCPU limit quota

  vCPU limit quota is the total amount of virtual CPU limits (in mCores) for all pods allocated to the namespace against the namespace quota.

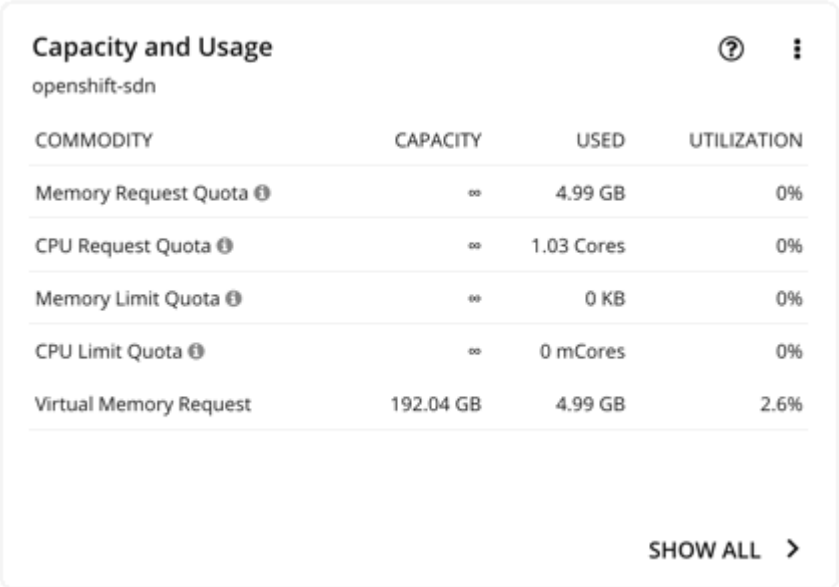
When configured, these quotas define the capacity for the given namespace. Workload Optimization Manager monitors actual utilization of these quotas against cluster capacity.

Points to consider:

- When you scope to a namespace in the supply chain, you can see utilization data in the **Capacity and Usage** and **Namespace Multiple Resources** charts. With this data, you can understand how pods running in the namespace are consuming resources.
- The Capacity and Usage chart shows *Capacity* as the namespace quotas. *Used* values are the sum of resource limits and/or requests set for all pods in the namespace.



For a namespace that does not have defined quotas, *Capacity* for the commodity is infinite (as shown in the following image). *Used* values are the sum of resource limits and/or requests set for all pods in the namespace. If these are not set, *Used* value is 0 (zero).



NOTE:
If you download the data in the chart, the downloaded file shows infinite capacities as unusually large values (for example, 1,000,000,000 cores instead of the ∞ symbol).

- To see which namespaces use the most cluster resources, set the scope to a container platform cluster and see the **Top Namespaces** chart. You can use the data in the chart for showback analysis.
- When you run Optimize Container Cluster plans, Workload Optimization Manager can calculate increased namespace quotas in the plan results. For more information, see this topic *(on page 861)*.

Namespace actions

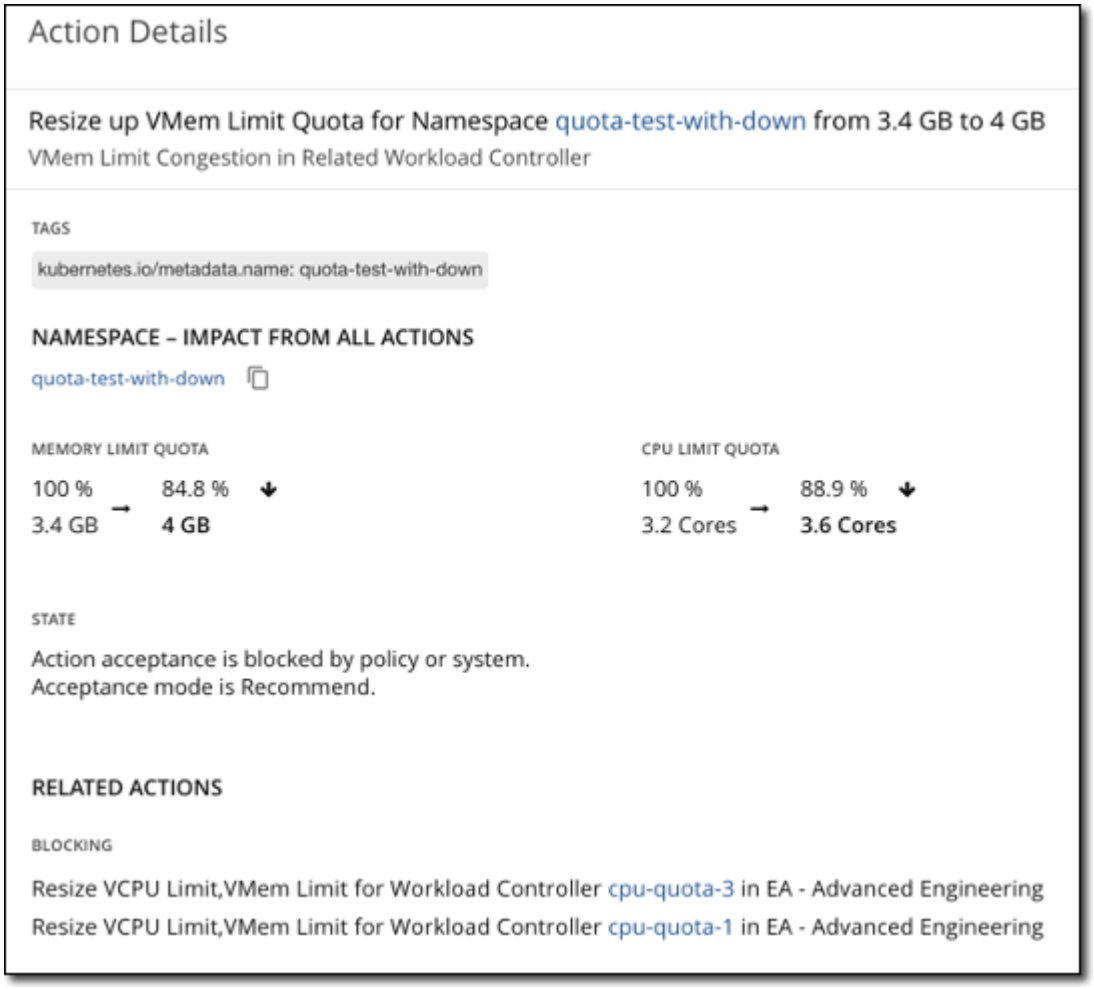
Workload Optimization Manager supports the following actions:

Resize quota

Workload Optimization Manager treats quotas defined in a namespace as constraints when making resize decisions. If existing actions would exceed the namespace quotas, Workload Optimization Manager recommends actions to resize up the affected namespace quota.

Note that Workload Optimization Manager does not recommend actions to resize *down* a namespace quota. Such an action reduces the capacity that is already allocated to an application. The decision to resize down a namespace quota should include the application owner.

When you have a recommendation to resize namespace quotas, Workload Optimization Manager blocks execution of the resize actions for the affected Workload Containers. The action details show these blocked actions in the **Related Actions** list.



NOTE:
For more information about execution of resize actions via workload controllers, see Workload Controller Resize Actions *(on page 593)*.

Temporary Increases in Namespace Quotas

Workload Optimization Manager can increase namespace quotas temporarily under the following conditions:

- Workload Optimization Manager temporarily increases the quota to allow the clone of the pod that is being moved to be scheduled.
- When executing actions to resize a workload controller in a namespace with a defined quota, Workload Optimization Manager can increase the namespace quota temporarily to accommodate new replicas.

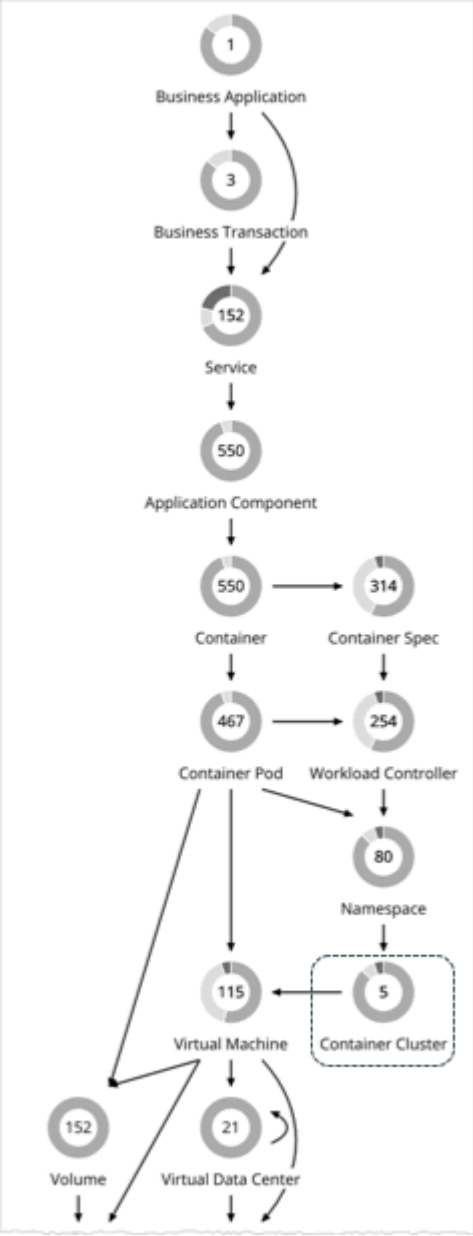
  When the execution of a resize action requires new replicas, a workload usually has a rolling update strategy defined. Even if the namespace quota is sufficient for the resource requirement of the new replicas, there is a chance that the quota is not enough to accommodate both the old and new replicas, as required by the rolling update. In this case, Kubeturbo calculates and then increases the quota based on the resources required by both the old and new replicas. Kubeturbo reverts the quota to its original size after the new replicas have been scheduled on a node.

Container platform cluster

A container platform cluster is a Kubernetes or Red Hat OpenShift cluster that Workload Optimization Manager discovers through Kubeturbo. With this entity type, Workload Optimization Manager can fully link the entire cluster with the underlying

nodes, and then present all actions in a single view. This gives you full visibility into the actions that impact the health of your cluster.

Synopsis



Synopsis	
Provides:	N/A
Consumes:	N/A
Discovered through:	Kubeturbo agent that you deployed to your cluster

GitOps integration

Workload Optimization Manager integrates with Git-based software through Argo CD, a GitOps continuous delivery tool that manages `ArgoCD Application` resources using Custom Resources (CRs). With this integration, Workload Optimization

Manager can identify the container platform cluster resources managed by Argo CD, and then execute resize actions against the Git-based version control software serving as Argo CD's 'source of truth'. Argo CD completes the action by synchronizing changes onto the running environment.

For detailed integration instructions, see this topic *(on page 620)*.

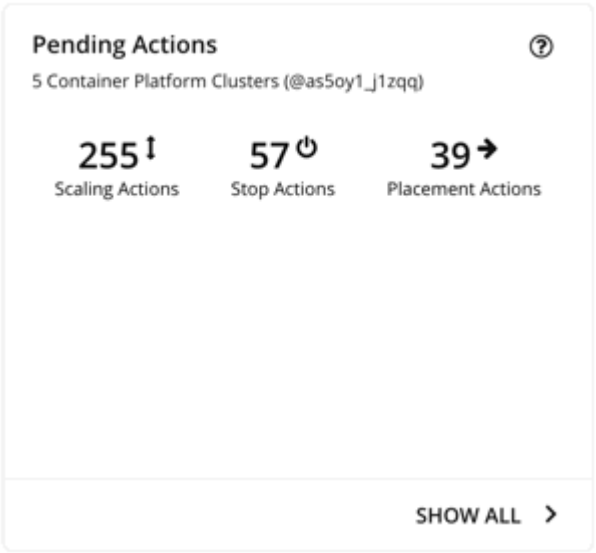
Monitored resources

Workload Optimization Manager monitors resources for the containers, pods, nodes (VMs), and volumes in a cluster.

Actions

None

Workload Optimization Manager does not recommend actions for a container platform cluster. Instead, it recommends actions for the containers, pods, nodes (VMs), and volumes in the cluster. Workload Optimization Manager shows all of these actions when you scope to a container platform cluster and view the Pending Actions chart.



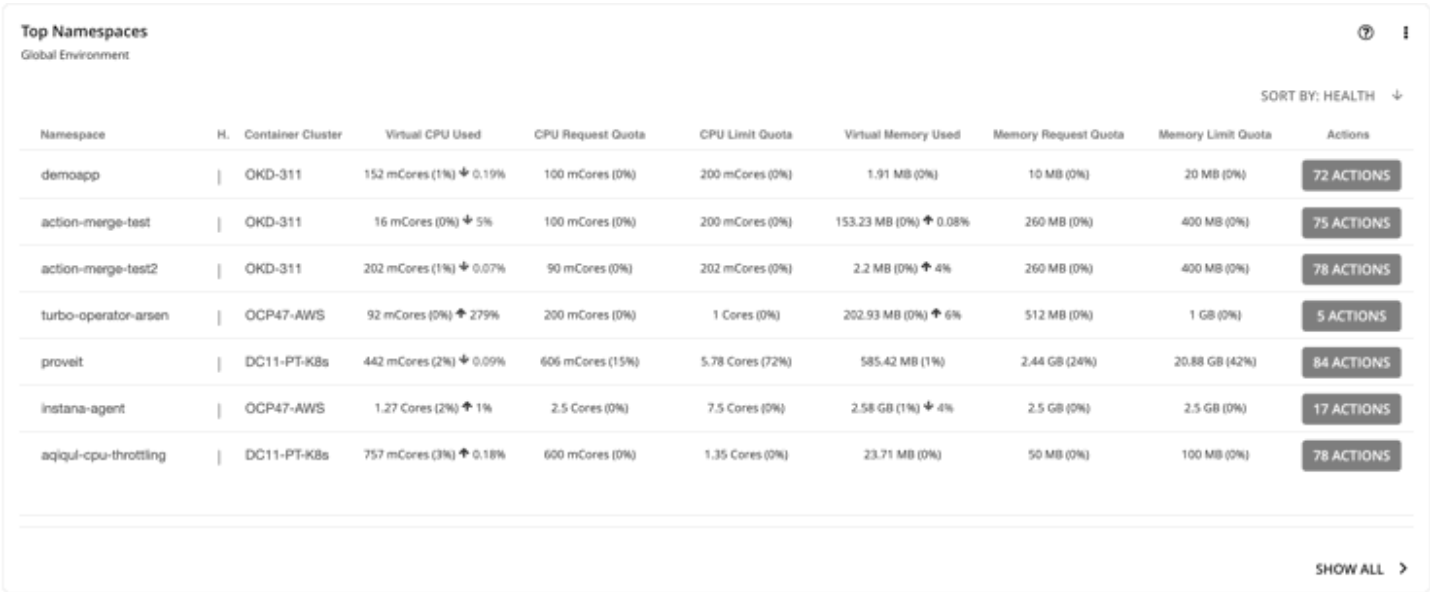
Cluster health

To assess the health of each cluster, see the **Top Container Platform Clusters** chart in the predefined Container Platform Dashboard.

For each cluster, the chart shows the sum of resources used by containers and the underlying nodes. Click the **Actions** button to see a list of pending actions.



The **Top Namespaces** chart shows the namespaces that use the most cluster resources. You can use the data in the chart for showback analysis.



GitOps integration

Workload Optimization Manager integrates with Git-based software through Argo CD, a GitOps continuous delivery tool that manages `ArgoCD Application` resources using Custom Resources (CRs). With this integration, Workload Optimization Manager can identify the container platform cluster resources managed by Argo CD, and then execute resize actions against the Git-based version control software serving as Argo CD's 'source of truth'. Argo CD completes the action by synchronizing changes onto the running environment.

Workload Optimization Manager currently supports the following definitions:

Workload definition – path to version control software	Version control software (Git source)	Application packaging
`ArgoCD Application` CR in the cluster	GitHub or GitLab	None (straight YAMLs only). Operators and Helm are not supported.
`ArgoCD ApplicationSet` CR in the cluster		

Configurations

It is assumed that Argo CD is installed and active in the cluster managed by Workload Optimization Manager. Although Workload Optimization Manager discvoers comprehensive information from the Argo CD applications that manage the resources, some details (such as Git credentials) must be configured in Kubeturbo running in the same cluster. The following sections describe these configurations.

Creating a developer access token in Git

1. In **Settings**, select **Developer Settings**.
2. Select **Personal Access Tokens**.
3. Generate a new token and grant the following permissions:
   - `delete:packages`
   - `gist`
   - `notifications`
   - `repo`
   - `write:discussion`
   - `write:packages`
4. Copy the token. You will need it in the next task.

Creating a secret to store the developer access token

Create a new secret in the namespace where Kubeturbo is deployed.

```
kind: Secret
apiVersion: v1
metadata:
  name: github
  namespace: turbo
data:
  token: {your_token_in_Base64}
type: Opaque
```

NOTE:
You will need to add a rule to your existing Kubeturbo service account role to grant permissions to get the secrets resource.

```
rules:
  # Existing rules
  - apiGroups:
      - ""
    resources:
      - secrets
    verbs:
      - get
```

Configuring Kubeturbo deployed through YAML

The following snippet shows the settings to configure in the Kubeturbo deployment. These settings will be used for all Argo CD apps deployed in the cluster.

```
--git-email string                          The email to be used to push changes to git.
--git-secret-name string                    The name of the secret which holds the git credentials.
--git-secret-namespace string               The namespace of the secret which holds the git credentials.
--git-username string                       The user name to be used to push changes to git.
--git-commit-mode                           The commit mode that should be used for git action
 executions. One of request|direct. Defaults to direct.
```

Specify these settings in the `args` section of the YAML file, as shown in the following example.

```
apiVersion: apps/v1
kind: Deployment
metadata:
  name: kubeturbo-test
....
....
spec:
  replicas: 1
  selector:
    matchLabels:
      app: kubeturbo-test
  template:
    metadata:
      annotations:
        kubeturbo.io/monitored: "false"
      labels:
        app: kubeturbo-test
    spec:
      containers:
      - args:
        - --turboconfig=/etc/kubeturbo/turbo.config
        - --v=2
        - --kubelet-https=true
        - --kubelet-port=10237
        - --git-email=youremail@gmail.com
        - --git-secret-name=your-github-secret-name
        - --git-secret-namespace=turbo
        - --git-username=yourusername
        - --git-commit-mode=direct
        image: turbonomic/kubeturbo:latest
        imagePullPolicy: Always
        name: kubeturbo-test
        resources: {}
...
...
```

Configuring Kubeturbo deployed through an operator

In the operator's CR spec, find the `args` section and configure the following settings. These settings will be used for all Argo CD apps deployed in the cluster.

```
apiVersion: charts.helm.k8s.io/v1
kind: Kubeturbo
metadata:
  name: kubeturbo-sample
spec:
  HANodeConfig:
    nodeRoles: '"master"'
  args:
```

```
    kubelethttps: true
    kubeletport: 10250
...
...
  # [ArgoCD integration] The email to be used to push changes to git.
   gitEmail: "youremail@gmail.com"
  # [ArgoCD integration] The username to be used to push changes to git.
   gitUsername: "yourusername"
  # [ArgoCD integration] The name of the secret which holds the git credentials.
   gitSecretName: "your-github-secret-name"
  # [ArgoCD integration] The namespace of the secret which holds the git credentials.
   gitSecretNamespace: "turbo"
  # [ArgoCD integration] The commit mode that should be used for git action executions. One of
 {request|direct}. Defaults to direct.
   gitCommitMode: "direct"
```

(Optional) Fine grained app specific configurations

It is possible for Argo CD apps to sync with different sources of truth (such as different GitHub repositories) within the same cluster. It is also possible that the credentials for these sources of truth are different. The app specific fine grained configuration uses a `config` Custom Resource to support different configuration information for different apps.

If you use plane YAML, first install the CRD that you can download here.

To deploy the CRD, run the following command:

```
kubectl apply -f
 https://raw.githubusercontent.com/turbonomic/turbo-gitops/main/config/crd/bases/gitops.turbonomic.io_gito
ps.yaml
```

When the type definition is available, use a configuration similar to the following example.

```
apiVersion: gitops.turbonomic.io/v1alpha1
kind: GitOps
metadata:
  name: gitops-sample
spec:
  config:
    - commitMode: direct
      credentials:
        email: test@turbonomic.comm
        secretName: gitops-secret
        secretNamespace: gitops
        username: turbo
      selector: '^turbo.*$'
      whitelist:
        - app-name-1
        - app-name-2
```

This resource is name spaced. You can specify one or several non-conflicting configurations for each namespace. If an app is selected as per the `whitelist` or the `selector`, the `credentials` and `commitMode` listed in the GitOps resource spec will be used instead of the global specs configured in Kubeturbo.

(Optional) Cluster local resource update

You may want Workload Optimization Manager to update the resources locally and directly in the cluster, rather then pushing the changes to the source of truth.

- Security

  You may want to minimize access to the credentials and the source of truth.

- Faster optimization

If the update must be pushed to the source of truth, the turnaround time for the update is longer because of the number of systems involved and the possibility of failures and retries. Workload Optimization Manager must also push a commit or a pull request (approvals might be required) to the source of truth. Even if Argo CD runs in auto-sync mode, it must pull and sync the resource into the cluster.

For faster optimization, you may want to update a resource locally in the cluster.

NOTE:
Argo CD has a default interval of 3 minutes to discover changes in the source of truth. Changes in Workload Optimization Manager can take 10 minutes or longer to apply.

- Multi-cluster scenarios

  It is possible for multiple clusters to use a unified app definition and sync from a base definition. Optimizations from one cluster might not apply to another cluster. If the updates are applied to the source of truth, then the change applies to all the clusters.

- Newer or unsupported tooling in the pipeline

  It is possible that Workload Optimization Manager does not have the capability to understand the pipeline or directly interact with the GitOps system.

To update the resources locally, configure the Argo CD app to skip the syncing of some fields from the child resources.

- `spec.ignoreDifferences` prevents out-of-sync prompts when the cluster local value changes. For details, see the Argo CD documentation.
- If the sync happened because of other field changes (as described in the Argo CD documentation), `RespectIgnoreDifferences` ignores the copying of original values into the cluster local resources.

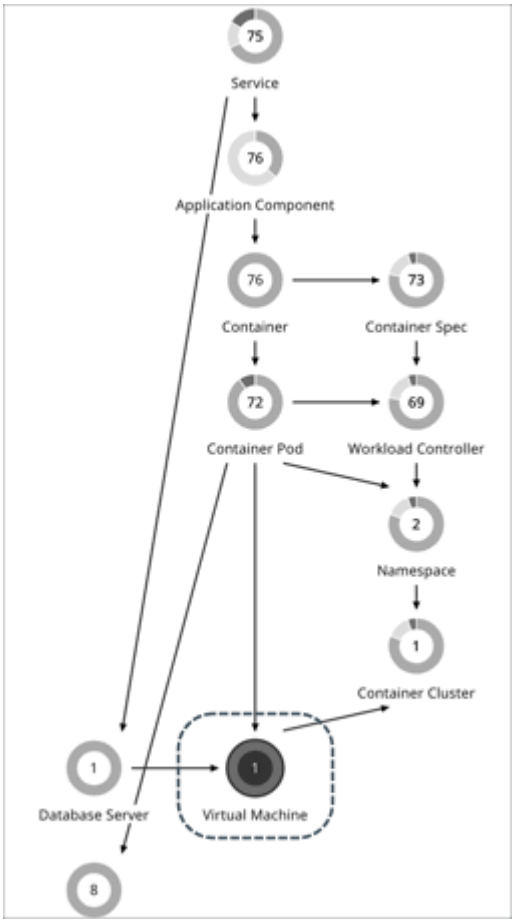
Use both options in the Argo CD applications.

Virtual machine (container platform node)

In container platform environments, a node is a virtual or physical machine that contains the services necessary to run pods. Workload Optimization Manager represents nodes as Virtual Machine entities in the supply chain.

Workload Optimization Manager can discover node roles and Master Nodes. It creates policies to keep nodes of the same role on unique host or Availability Zone providers, and policies to disable suspension of Master Nodes. Workload Optimization Manager also discovers and displays Node Pools and Red Hat OpenShift Machine Sets.

Synopsis



Synopsis	
Provides:	Resources to pods
Consumes:	Resources from container platform clusters
Discovered through:	Kubeturbo agent that you deployed to your cluster

Monitored resources

Workload Optimization Manager monitors the following resources:

> **NOTE:** These are resources for nodes that host pods. They are monitored along with the resources from the infrastructure probes, such as vCenter or a public cloud mediation probe.

- vMem

  vMem is the virtual memory currently used by all containers on the node. The capacity for this resource is the Node Physical capacity.
- vCPU

  vCPU is the virtual CPU currently used by all containers on the node. The capacity for this resource is the Node Physical capacity.
- Memory request allocation

Memory request allocation is the memory available to the node to support the `ResourceQuota` request parameter for a given Kubernetes namespace or Red Hat OpenShift project.

- CPU request allocation

  CPU request allocation is the CPU available to the node to support the `ResourceQuota` request parameter for a given Kubernetes namespace or Red Hat OpenShift project.

- Virtual memory request

  Virtual memory request is the memory currently guaranteed by all containers on the node with a memory request. The capacity for this resource is the Node Allocatable capacity, which is the amount of resources available for pods and can be less than the physical capacity.

- Virtual CPU request

  Virtual CPU request is the CPU currently guaranteed by all containers on the node with a CPU request. The capacity for this resource is the node allocatable capacity, which is the amount of resources available for pods and can be less than the physical capacity.

- Memory allocation

  Memory allocation is the memory `ResourceQuota` limit parameter for a given Kubernetes namespace or Red Hat OpenShift project.

- CPU allocation

  CPU allocation is the CPU `ResourceQuota` limit parameter for a given Kubernetes namespace or Red Hat OpenShift project.

Actions

Workload Optimization Manager supports the following actions:

- Provision

  Provision nodes to address workload congestion or meet application demand.

- Suspend

  Suspend nodes after you have consolidated pods or defragmented node resources to improve infrastructure efficiency.

- Reconfigure

  Reconfigure nodes that are currently in the `NotReady` state.

Node provision and suspension actions

For both node provision and suspension actions, review the following guidelines:

- For nodes in the public cloud, Workload Optimization Manager reports the cost savings or investments attached to node and provision actions. For example, you can see the additional costs you would incur if you provision nodes and then scale their volumes, or the savings you would realize if you suspend nodes. Note that performance and efficiency are the drivers of these actions, *not* cost. Cost information is included to help you track your cloud spend. For this reason, you will *not* see cost-optimization actions, including recommendations to re-allocate discounts or delete unattached volumes.

  To view cost information, set the scope to a node and see the Necessary Investments and Potential Savings charts. You can also set the scope to a container platform cluster *(on page 617)* or the global cloud environment to view aggregated cost information.

- For nodes that make up an Amazon Elastic Kubernetes Service (EKS), Azure Kubernetes Service (AKS), or Google Kubernetes Engine (GKE) cluster, Workload Optimization Manager can recommend node provision actions to increase the node count, and node suspension actions to reduce the node count. To manually or automatically execute these

actions in Workload Optimization Manager, be sure that Workload Optimization Manager can connect to your cluster through the following targets:

  - EKS cluster – a container platform target (Kubernetes or Red Hat OpenShift) and an AWS target
  - AKS cluster – a container platform target (Kubernetes or Red Hat OpenShift) and an Azure service principal target
  - GKE cluster – a container platform target (Kubernetes or Red Hat OpenShift) and a Google Cloud target

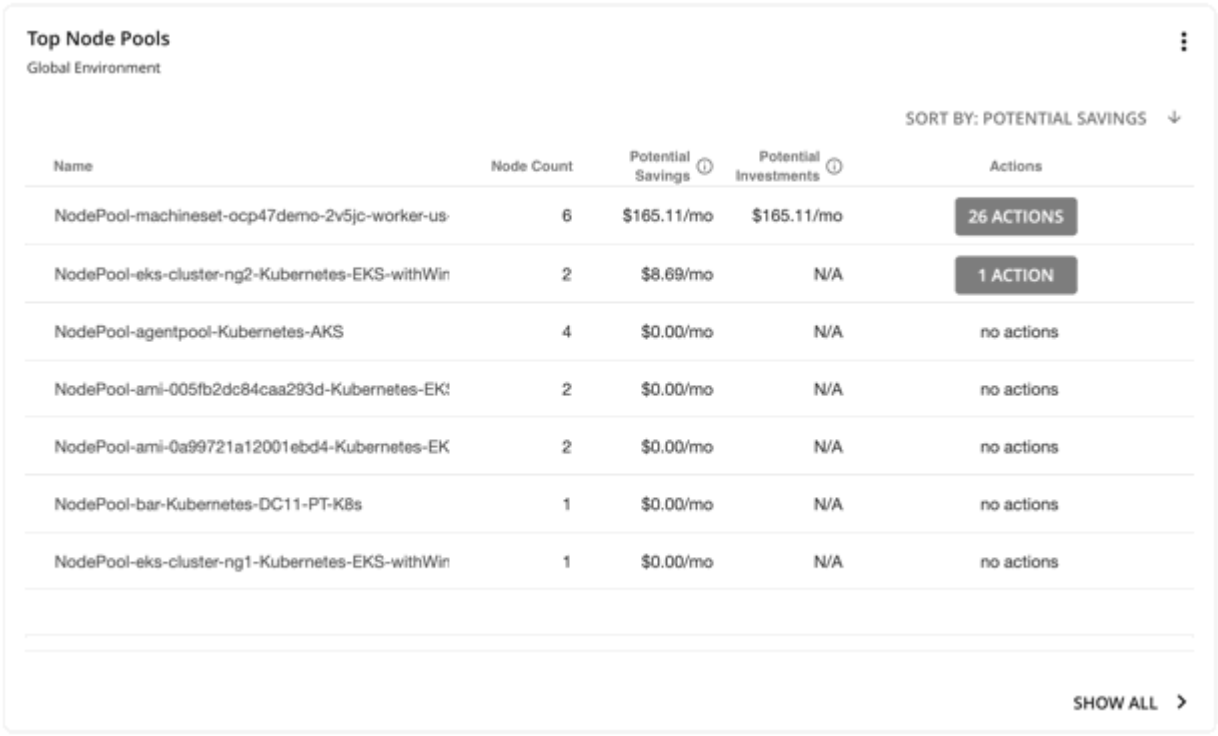
- Node pools and machine sets are ways to deploy and scale compute resources for container platform services hosted in the public cloud or Red Hat OpenShift on any infrastructure.

  For the public cloud, Workload Optimization Manager uses default labels with the following patterns to discover the node pool types within each cluster:

Node pool	Pattern
AKS	`agentpool`
EKS	`alpha.eksctl.io/nodegroup-name` `eks.amazonaws.com/nodegroup`
GKE	`cloud.google.com/gke-nodepool`

  For Red Hat OpenShift, Workload Optimization Manager creates node pools based on machine sets.

  For both discovered and auto-created node pools, Workload Optimization Manager aggregates and visualizes actions for all the nodes in a pool to help you identify performance issues and optimization opportunities at the node pool level. Use the Top Node Pools chart to see actions and detailed information. By default, this chart displays when you set the scope to your global environment and then click the Container Platform Cluster entity in the supply chain.



The chart shows the number of nodes and aggregated actions for each node pool. For node pools in the public cloud, the chart also shows the costs you would incur if you provision nodes and then scale their volumes, or the savings you would realize if you suspend nodes. To view individual actions, click the button under the Actions column. To see more details, including the full list of nodes for each pool, click the node pool name.

You can automate the execution of these actions through Workload Optimization Manager with Red Hat OpenShift Machine API Operator . You can also manually execute node actions for AKS, EKS, or GKE through the cloud provider.

NOTE:
Policies for node pools will be introduced in a future release.

Pre-execution check for node provision and suspension actions

Before executing a node provision or suspension action for a MachineSet in a Red Hat OpenShift cluster, Kubeturbo checks the node count range specified in your `ConfigMap`. If the pre-execution check determines that the node count will fall outside the range after action execution, the action executes but fails, and the failure is logged (for example, in the Executed Actions chart in the Workload Optimization Manager user interface). This mechanism ensures the overall stability and performance of the Red Hat OpenShift cluster.

NOTE:
The pre-execution check is not currently supported in AKS, EKS, GKE node pools.

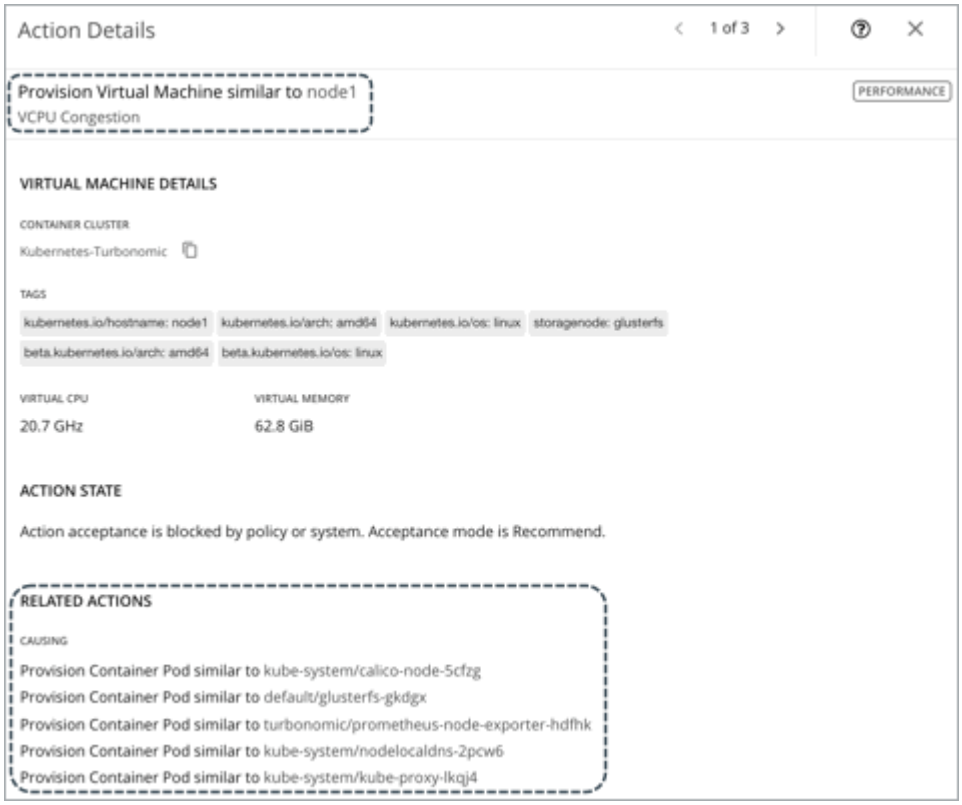
By default, the minimum node count is 1 and the maximum is 1000. You can customize these values by updating the `nodePoolSize` parameter in your Kubeturbo `ConfigMap`. This update does not require a restart of the Kubeturbo pod and takes effect after approximately one minute.

For a sample `ConfigMap` with the `nodePoolSize` parameter, see the Kubeturbo GitHub repository.

Node provision actions

When recommending node provision actions, Workload Optimization Manager also recommends pod provision actions that reflect the projected demand from required DaemonSet pods, and respects the maximum number of pods allowed for a node. This ensures that any application workload can be placed on the new node and stay within the desired range of vMem/vCPU usage, vMem/vCPU request, and number of consumers.

The action details for a node provision action show the related DaemonSet pods that are required for the node to run. Click a pod name to set it at your scope.

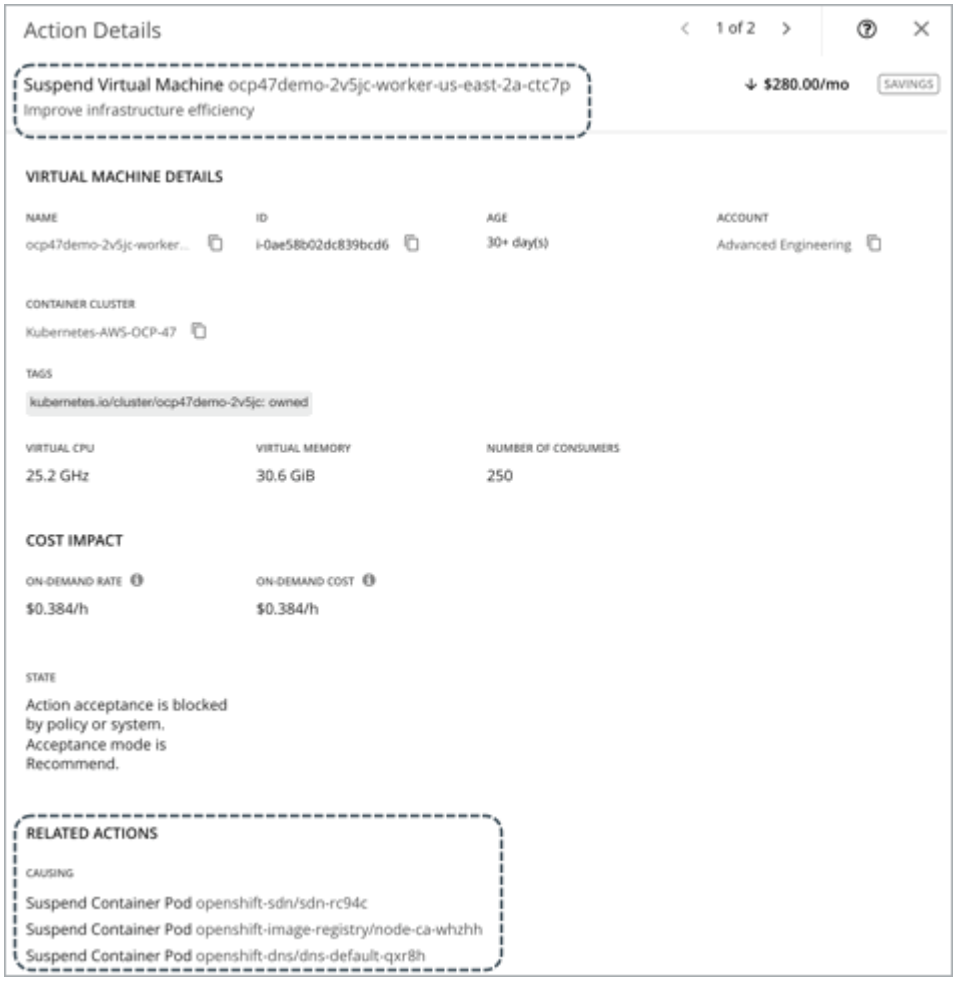


Workload Optimization Manager treats static pods as DaemonSets for the purpose of provisioning nodes. Because a static pod provides a node with a specific capability, it is controlled by the node and is not accessible through the API server. If a node to be provisioned requires a static pod, Workload Optimization Manager generates actions to provision the node and the corresponding static pod.

Node suspension actions

When recommending node suspension actions, Workload Optimization Manager also recommends suspending the DaemonSet pods that are no longer required to run the suspended nodes.

The action details for a node suspension action show the related DaemonSet pods that are no longer needed to run the suspended nodes. Click a pod name to set it at your scope.



Workload Optimization Manager treats static pods as DaemonSets for the purpose of suspending nodes. Because a static pod provides a node with a specific capability, it is controlled by the node and is not accessible through the API server. If the only workload type on a node is a static pod, Workload Optimization Manager generates actions to suspend the node and the corresponding static pod.

Nodes for Pods with Topology Spread Constraints

Currently, Workload Optimization Manager does not support moving pods with topology spread constraints. Until support is available, suspension actions for the nodes that the affected pods run on are disabled. For more information, see this topic *(on page 291)*.

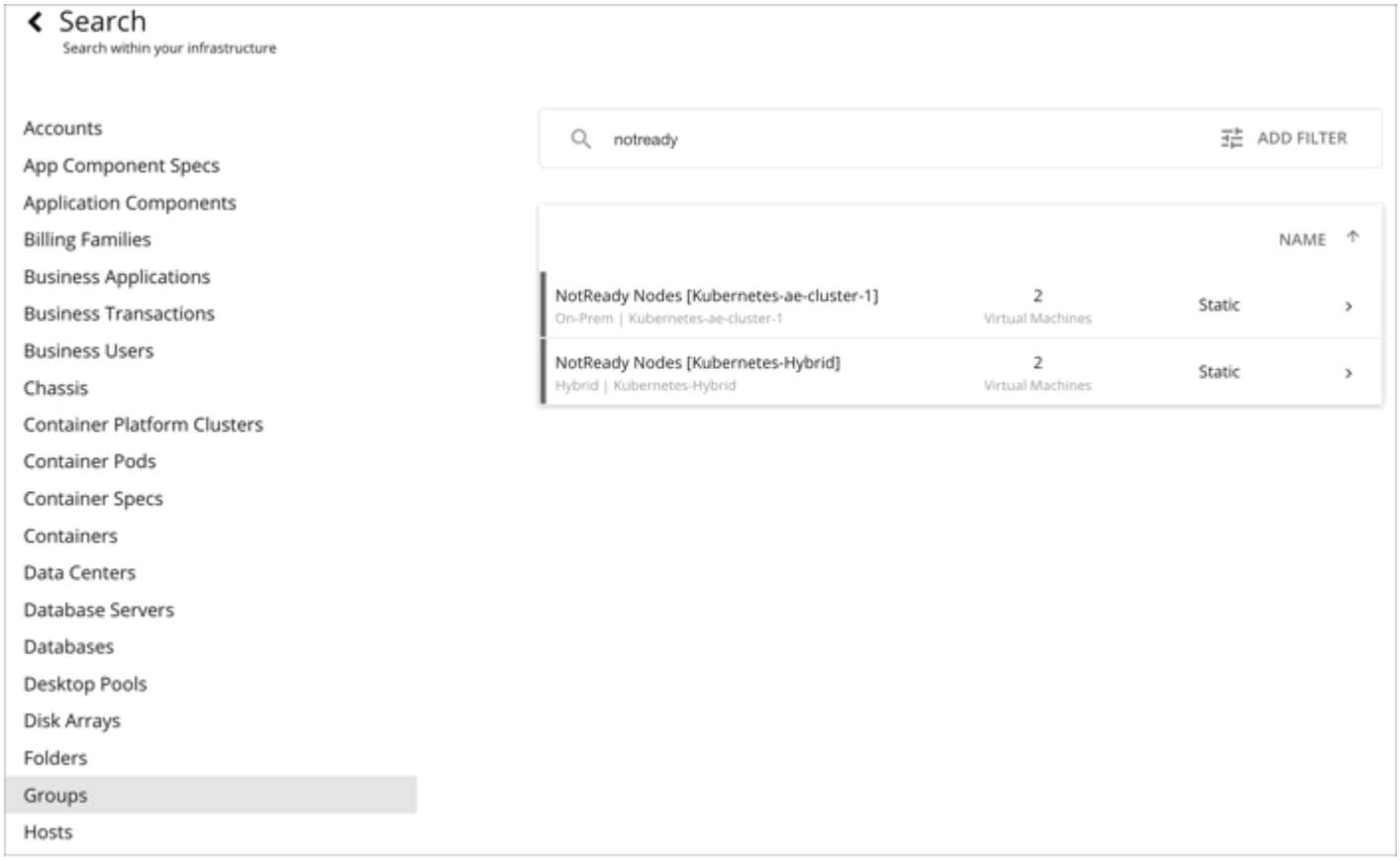
Node reconfigure actions

Workload Optimization Manager generates node reconfigure actions to notify you of nodes that are currently in the `NotReady` state.

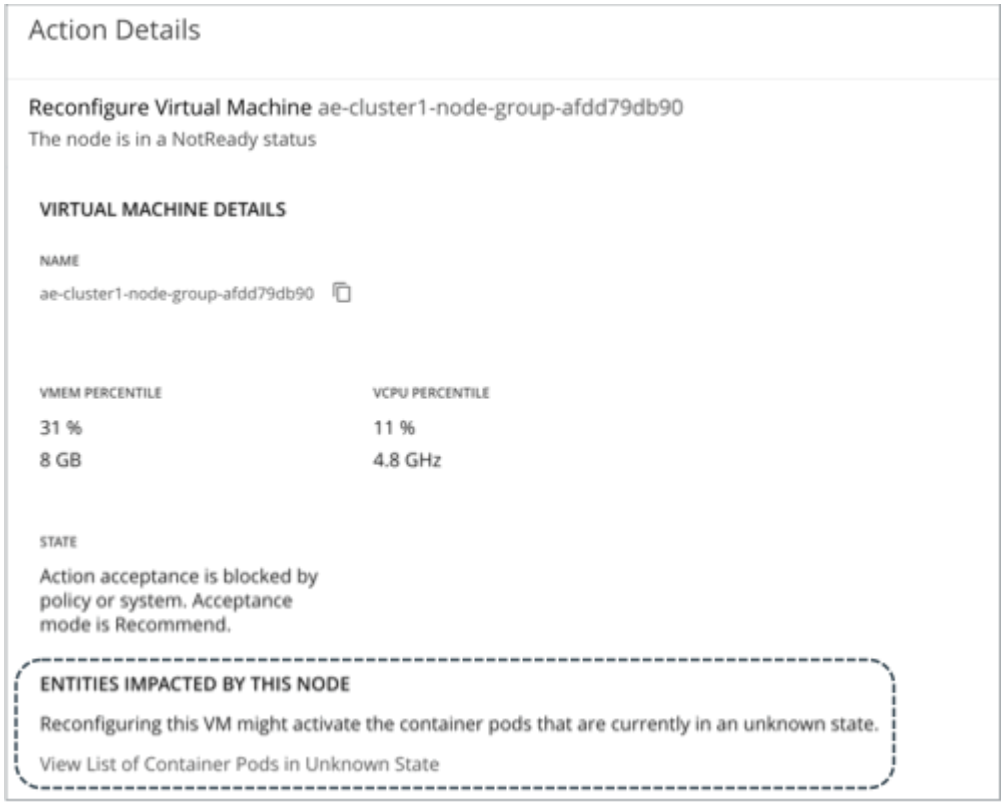
A reconfigure action is read-only in Workload Optimization Manager and must be executed directly in the container platform cluster because the action might require restarting the node or the kubelet on the node. When Workload Optimization Manager discovers that a node's state is `Ready`, it removes the reconfigure action automatically and begins to monitor the health of the node and the associated container pods.

NOTE:
Workload Optimization Manager treats a node as a VM under certain circumstances. For example, it treats a node in vCenter as a VM that can move to a different host if the current host is congested. This means that for a `NotReady` node in vCenter, it is possible to see a VM move action along with the expected node reconfigure action. Both actions are valid and safe to execute since they achieve two different and non-conflicting results.

For each container platform cluster, Workload Optimization Manager creates an auto-generated group of `NotReady` nodes. To view all the auto-generated groups, go to Search, select Groups, and then type `notready` as your search keyword. Click a group to view the individual nodes and the pending reconfigure actions.



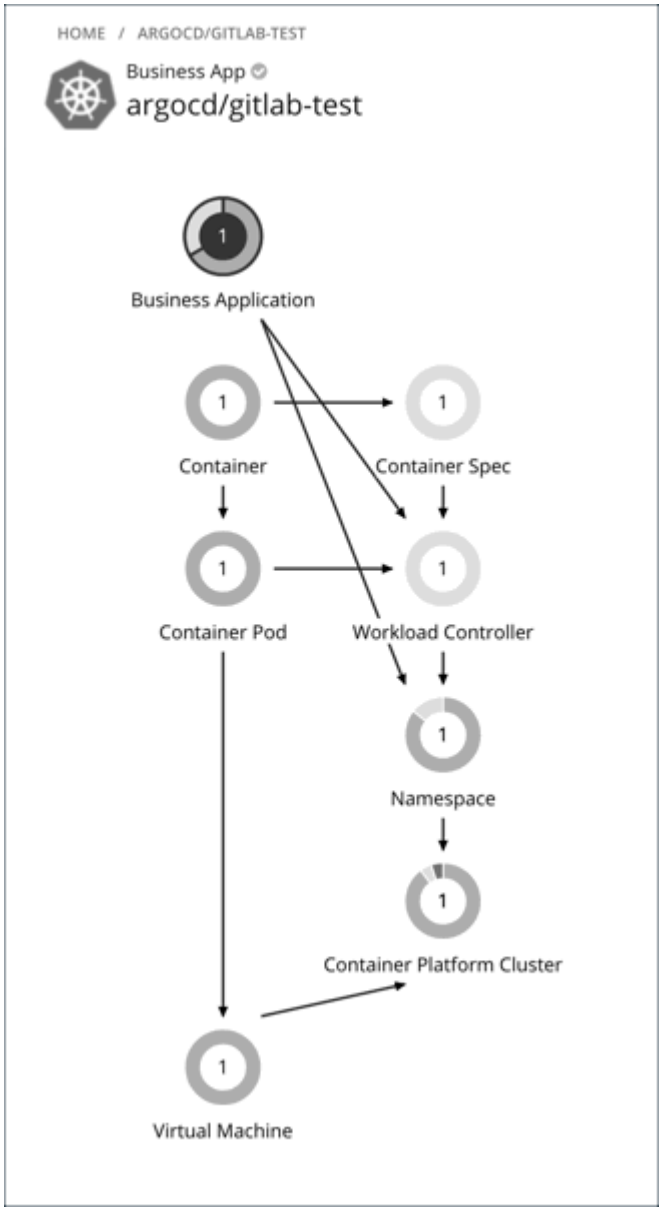
When you examine a pending reconfigure action, you can click the link in the **Entities Impacted by this Node** section to view a list of impacted pods.



These pods are in the Unknown state and are not controllable. In the supply chain and in the list of container pods, these pods display with a gray color to help you differentiate them from other pods.

Container platform CPU metrics

To meet user requirements and align with container platform specifications, Workload Optimization Manager uses millicore (mCore) as the base unit for CPU metrics for your container platform.



These include metrics for the following CPU-related commodities:

- vCPU
- vCPU Request
- vCPU Limit Quota
- vCPU Request Quota
- vCPU Throttling

Workload Optimization Manager displays these commodities in charts, actions, policies, and plans. For example:

- In the Capacity and Usage chart for container platform entities, *capacity* and *used* values for CPU-related commodities are shown in mCores.
- In the supply chain, when you scope to a Workload Controller to view pending resize actions *(on page 579)* for a container, you will see utilization and resize values in mCores.

- When you create container spec policies *(on page 581)*, resize thresholds and increment constants for CPU-related commodities are set in mCores.
- For an Optimize Container Cluster plan, the plan results *(on page 864)* for CPU-related commodities are shown in mCores.

For nodes (VMs) and Application Components:

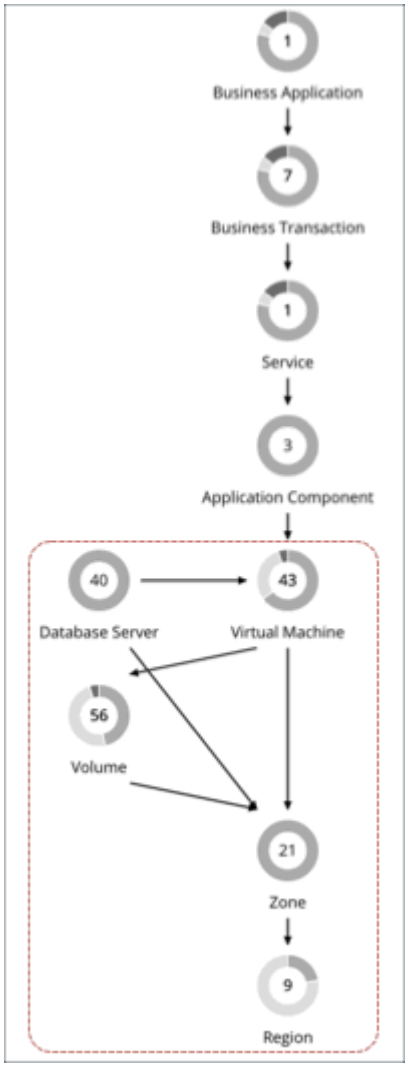
- For nodes stitched to your container platform, the base unit for *vCPU Request* is also mCore, since this commodity is provided only to the container platform.
- For both nodes and Application Components (standalone or stitched to your container platform), the base unit for *vCPU* is MHz, since this is a generic commodity. For example, when you view a pod move action, vCPU metrics for the current and destination nodes for the pod are expressed in MHz.

The following table summarizes the base units of CPU measurement that Workload Optimization Manager uses.

Entity	CPU commodity				
	vCPU	vCPU request	vCPU limit quota	vCPU request quota	vCPU throttling
Container	mCore	mCore	mCore	mCore	mCore
Container spec	mCore	mCore	N/A	N/A	mCore
Workload Controller	N/A	N/A	mCore	mCore	N/A
Container Pod	mCore	mCore	mCore	mCore	N/A
Namespace	mCore	mCore	mCore	mCore	N/A
Container Platform Cluster	mCore	mCore	N/A	N/A	N/A
Node (VM)	MHz	mCore	N/A	N/A	N/A
Application Component	MHz	N/A	N/A	N/A	N/A

Entity types - cloud infrastructure

Workload Optimization Manager discovers and monitors the entities that make up your cloud infrastructure, and recommends actions to assure application performance at the lowest possible cost.

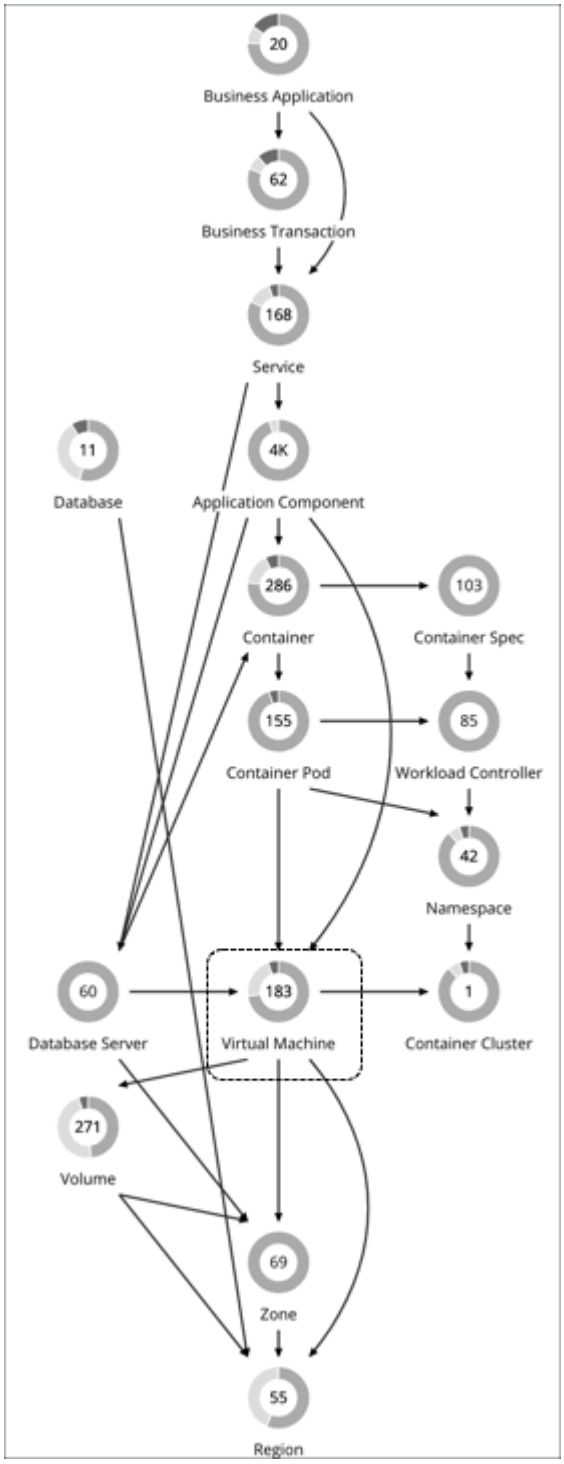


Virtual machine (cloud)

A virtual machine (VM) is a software emulation of a physical machine, including OS, virtual memory and CPUs, and network ports. VMs host applications, or they provide resources to container platforms.

NOTE:
Container platform nodes are also represented as Virtual Machines in the Workload Optimization Manager supply chain. For details, see this topic *(on page 624)*.

Synopsis



Synopsis	
Provides:	Resources for applications to use
Consumes:	Resources from cloud zones or regions
Discovered through:	Public cloud targets

Monitored resources

For details about monitored resources for cloud VMs, see the following topics:

- AWS Monitored Resources *(on page 393)*
- Azure Monitored Resources *(on page 464)*
- Google Cloud Monitored Resources *(on page 423)*

Cloud VM actions

For details about cloud VM actions, see the following topics:

- Actions for AWS VMs *(on page 637)*
- Actions for Azure VMs *(on page 641)*
- Actions for Google Cloud VMs *(on page 646)*

Actions for AWS VMs

Workload Optimization Manager supports the following actions:

- Scale

  Change the VM instance to use a different instance type or tier to optimize performance and costs.

  See additional information for scale actions.

- Discount-related actions

  If you have a high percentage of on-demand VMs, you can reduce your monthly costs by increasing RI or Savings Plans coverage. To increase coverage, you scale VMs to instance types that have existing capacity.

- Stop and start (also known as 'parking' actions)

  Stop a VM for a given period of time to reduce your cloud expenses, and then start it at a later time.

  For details, see Parking: Stop or Start Cloud Resources *(on page 942)*.

Controlling scale actions for AWS VMs

For scale actions, you can create policies to control the scale actions that Workload Optimization Manager recommends. In those policies, choose from the following options:

- Cloud scale all – execute all scaling actions
- Cloud scale for performance – only execute scaling actions that improve performance
- Cloud scale for savings – only execute scaling actions that reduce costs

The default action acceptance mode for these actions is *Manual*. When you examine the pending actions, only actions that satisfy the policy are allowed to execute. All other actions are read-only.

When policy conflicts arise, **Cloud scale all** overrides the other two scaling options in most cases. For more information, see Default and User-defined Automation Policies *(on page 1050)*.

Supported instance types for AWS VM scale actions

In the user interface, you can view the instance types that Workload Optimization Manager currently supports.

1. Navigate to **Settings > Policies**.
2. In the Policy Management page, search for and click **Virtual Machine Defaults**.
3. In the Configure Virtual Machine Policy page:
   a. Scroll to the end of the page.
   b. Click **Add Scaling Constraint**.
   c. Choose **Cloud Instance Types**.
   d. Click **Edit**.

Expand an instance family to see individual instance types and the resources allocated to them. An example of an instance family is a1.

By default, Workload Optimization Manager considers all instance types that are currently available for scaling when making scaling decisions. However, you may have set up some of your cloud VMs to only scale to certain instance types to reduce complexity and cost, improve discount utilization, or meet application demand. To limit scaling to certain instance types, create policies for the affected cloud VMs and configure an inclusion or exclusion list as a scaling constraint.

Scale actions for powered off VMs

Workload Optimization Manager can recommend scale actions for VMs that are powered off (for example, during maintenance). If you accept the actions, Workload Optimization Manager scales the VMs while they are powered off so that the changes are already applied as soon as they are powered on.

AWS Billing target and scale actions

Workload Optimization Manager uses historical billing data discovered through your AWS Billing target *(on page 389)* to generate the most optimal scaling recommendations for VMs.

When VMs are associated with AWS accounts that Workload Optimization Manager monitors through your AWS Billing target, be aware of the following caveats for VM scale actions.

- After you add a new VM to an account, it may take Workload Optimization Manager up to two days to generate scale actions for the VM because AWS does not immediately add the VM to the billing data.
- If your AWS Billing target discovers billing data using a legacy CUR export *(on page 388)*, costs that include Enterprise Discount Program (EDP) discounts are not reflected in scale actions for the affected VMs.

Scale actions for AWS VMs running EC2 GPU instance types

Currently, Workload Optimization Manager supports the following EC2 GPU instance type families with Linux AMIs.

- G3 instance family (based on NVIDIA Tesla M60 GPUs)
- G4dn instance family (based on NVIDIA T4 GPUs)
- G5 instance family (based on NVIDIA A10G Tensor Core GPUs)
- G5g instance family (based on NVIDIA T4G Tensor Core GPUs)
- P2 instance family (based on NVIDIA Kepler K80 GPUs)
- P3/P3dn instance family (based on NVIDIA Volta V100 GPUs)
- P4d instance family (based on NVIDIA A100 Tensor Core GPUs)

In the Workload Optimization Manager user interface, the default virtual machine policy shows the currently supported EC2 GPU instance type families and the resources allocated to them. These instance type families are grouped under the **Accelerated Computing** category. See the previous section for the steps to view the default virtual machine policy.

Workload Optimization Manager collects NVIDIA GPU metrics for VMs running these instance types and then uses these metrics to generate VM scale actions that optimize performance and costs.

NOTE:
For additional information about these metrics, see the NVIDIA documentation.

Metric	Description	Actions
GPU Count	Number of GPU cards in use	Scale down the number of GPU cards within the same instance type
GPU Memory	Amount of GPU memory in use	Scale GPU memory up or down within the same instance type

Metric	Description	Actions
GPU Memory BW (bandwidth)	Fraction of cycles where data was sent to or received from device memory, measured in GB/s	Scale GPU memory BW up or down This action moves the VM from its current instance type to another instance type with the same (or higher) GPU count and GPU memory capacity.
GPU FP16	Fraction of cycles the FP16 (half precision) pipe was active	Scale GPU FP16 up or down This action moves the VM from its current instance type to another instance type with the same (or higher) GPU count and GPU memory capacity.
GPU FP32	Fraction of cycles the FP32 (single precision) pipe was active	Scale GPU FP32 up or down This action moves the VM from its current instance type to another instance type with the same (or higher) GPU count and GPU memory capacity.
GPU FP64	Fraction of cycles the FP64 (double precision) pipe was active	Scale GPU FP64 up or down This action moves the VM from its current instance type to another instance type with the same (or higher) GPU count and GPU memory capacity.
GPU Tensor	Fraction of cycles the Tensor (mixed/multi-precision) pipe was active	Scale GPU Tensor up or down This action moves the VM from its current instance type to another instance type with the same (or higher) GPU count and GPU memory capacity.

NOTE:
For GPU VMs with accelerator cards, GPU card information is available when you set the scope to a VM and view the Entity Information chart. For example, the Accelerator Model field shows the GPU model.

To enable metrics discovery, configure AWS CloudWatch and NVIDIA Data Center GPU Manager (DCGM) as described in this topic *(on page 399)*.

Workload Optimization Manager can also recommend actions that scale *standard VM resources* (such as vCPU and vMem) to the supported GPU instance types and the G4ad instance family (based on AMD Radeon Pro V520 GPUs).

When generating or executing scale actions, Workload Optimization Manager considers the following policies and settings as constraints:

- Read-only tier exclusion policies

  Workload Optimization Manager automatically creates these policies and displays them in the Policy Management page (**Settings > Policies**).

  - Cross-target policies

    Cross-target policies ensure that VMs with certain GPU types only scale to an instance type with the same (or higher) GPU card count and memory per card.

    An example policy is `AWS GPU Nvidia - Cloud Compute Tier Exclusion Policy`.

  - Per-target policies

Per-target policies ensure that any VMs in GPU supported instance families do not scale out to instance families that do not support GPUs.

An example policy is `Cloud Compute Tier AWS:gpu - Cloud Compute Tier Exclusion Policy`.

- Scaling target utilization

  Workload Optimization Manager uses scaling target utilization values for GPU resources in conjunction with aggressiveness constraints to control scale actions for VMs. You can configure utilization values in automation policies for cloud VMs. For more information, see this topic *(on page 663)*.

- Ignore NVIDIA GPU compute capability constraints

  This constraint is a setting that you can choose to turn on in automation policies for cloud VMs. When turned on, scale actions that change the GPU compute capability of VMs are allowed to execute in Workload Optimization Manager. When turned off, actions are only executable in the cloud provider console (web-based UI or CLI). For more information, see this topic *(on page 668)*.

Support for AWS EC2 accelerator instance types

Workload Optimization Manager can recommend actions to scale standard VM resources (such as vCPU and vMem) to the following AWS EC2 Accelerator instance types.

- Inf1 instance family (based on AWS Inferentia chips)
- Inf2 instance family (based on AWS Inferentia2 chips)

Workload Optimization Manager also creates the appropriate read-only tier exclusion policies and displays them in the Policy Management page (**Settings > Policies**).

- Cross-target policies

  Cross-target policies ensure that AWS VMs with certain Accelerator types only scale to an instance type with the same Accelerator configuration (card count and memory per card). Policies include:

  - `AWS ML_ACCELERATOR - Inferentia1 - Cloud Compute Tier Exclusion Policy`
  - `AWS ML_ACCELERATOR - Inferentia2 - Cloud Compute Tier Exclusion Policy`

- Per-target policies

  Per-target policies ensure that any VMs in Inferentia1 instance families do not scale out to other instance families. Policies include:

  - `Cloud Compute Tier AWS:inf1 - Cloud Compute Tier Exclusion Policy`
  - `Cloud Compute Tier AWS:inf2 - Cloud Compute Tier Exclusion Policy`

Scaling prerequisites for AWS VMs

In AWS some instances require VMs to be configured in specific ways before they can scale to those instance types. If Workload Optimization Manager recommends scaling a VM that is not suitably configured onto one of these instances, then it sets the action to *Recommend*, and describes the reason. Workload Optimization Manager will not automate the action, even if you have set the action acceptance mode for that scope to *Automatic*. You can execute the action manually, after you have properly configured the instance.

Note that if you have VMs that you cannot configure to support these requirements, then you can set up a policy to keep Workload Optimization Manager from making these recommendations. Create a group that contains these VMs, and then create policy for that scope. In the policy, exclude instance types by configuring the Cloud Instance Types scaling constraint. For information about excluding instance types, see Cloud Instance Types *(on page 666)*.

The instance requirements that Workload Optimization Manager recognizes are:

- Enhanced Network Adapters

  Some VMs can run on instances that support Enhanced Networking via the Elastic Network Adapter (ENA), while others can run on instances that do not offer this support. Workload Optimization Manager can recommend scaling a VM that does not support ENA onto an instance that does. However, you must enable ENA on the VM before executing the

scaling action. If you scale a non-ENA VM to an instance that requires ENA, then AWS cannot start up the VM after the scaling action.

For information about ENA configuration, visit this page.

- Linux AMI Virtualization Type

  An Amazon Linux AMI can use ParaVirtual (PV) or Hardware Virtual Machine (HVM) virtualization. Workload Optimization Manager can recommend scaling a PV VM to an HVM instance that does not include the necessary PV drivers.

  To check the virtualization type of an instance, open the Amazon EC2 console to the Details pane, and review the Virtualization field for that instance.

- 64-bit vs 32-bit

  Not all AWS instance can support a 32-bit VMs. Workload Optimization Manager can recommend scaling a 32-bit VM to an instance that only supports a 64-bit platform.

- NVMe Block

  Some instances expose EBS volumes as NVMe block devices, but not all VMs are configured with NVMe drivers. Workload Optimization Manager can recommend scaling such a VM to an instance that supports NVMe. Before executing the action, you must install the NVMe drivers on the VM.

In addition, Workload Optimization Manager recognizes processor types that you currently use for your VM. For scale actions, Workload Optimization Manager keeps your VMs on instance types with compatible processors. For example, if your VM is on an ARM-based instance, then Workload Optimization Manager will only recommend scaling to other compatible ARM-based instance types.

Scaling storage for AWS VMs

When a VM needs more storage capacity Workload Optimization Manager recommends actions to scale its volume to an instance that provides more storage. Note that AWS supports both Elastic Block Store (EBS) and Instance storage. Workload Optimization Manager recognizes these storage types as it recommends volume actions.

If the root storage for your VM is Instance Storage, then Workload Optimization Manager will not recommend an action. This is because Instance Storage is ephemeral, and such an action would cause the VM to lose all the stored data.

If the root storage is EBS, then Workload Optimization Manager recommends volume actions. EBS is persistent, and the data will remain after the action. However, if the VM uses Instance Storage for extra storage, then Workload Optimization Manager does not include that storage in its calculations or actions.

Nodes in AWS EMR clusters

Workload Optimization Manager treats nodes in AWS EMR clusters like regular VMs. As such, it could incorrectly generate scale actions for such nodes. After a node action executes, AWS detects the action as a defect, terminates the node, and replaces it with a new instance of the initial size. To avoid this issue, disable scale actions for nodes in EMR clusters.

AWS automatically assigns system tags to EMR clusters. To disable scale actions, create a VM group that uses these tags as a filter, and then create a VM policy that disables the **Cloud Scale All** action type for the VM group.

Actions for Azure VMs

Workload Optimization Manager supports the following actions:

- Scale

  Change the VM instance to use a different instance type or tier to optimize performance and costs.

  See additional information for scale actions.

- Discount-related actions

  If you have a high percentage of on-demand VMs, you can reduce your monthly costs by increasing Azure reservations coverage. To increase coverage, you scale VMs to instance types that have existing capacity.

- Stop and start (also known as 'parking' actions)

Stop a VM for a given period of time to reduce your cloud expenses, and then start it at a later time.

For details, see Parking: Stop or Start Cloud Resources *(on page 942)*.

Controlling scale actions for Azure VMs

For scale actions, you can create policies to control the scale actions that Workload Optimization Manager recommends. In those policies, choose from the following options:

- Cloud scale all – execute all scaling actions
- Cloud scale for performance – only execute scaling actions that improve performance
- Cloud scale for savings – only execute scaling actions that reduce costs

The default action acceptance mode for these actions is *Manual*. When you examine the pending actions, only actions that satisfy the policy are allowed to execute. All other actions are read-only.

When policy conflicts arise, **Cloud scale all** overrides the other two scaling options in most cases. For more information, see Default and User-defined Automation Policies *(on page 1050)*.

Supported instance types for Azure VM scale actions

In the user interface, you can view the instance types that Workload Optimization Manager currently supports.

1. Navigate to **Settings > Policies**.
2. In the Policy Management page, search for and click **Virtual Machine Defaults**.
3. In the Configure Virtual Machine Policy page:
    a. Scroll to the end of the page.
    b. Click **Add Scaling Constraint**.
    c. Choose **Cloud Instance Types**.
    d. Click **Edit**.

Expand an instance family to see individual instance types and the resources allocated to them. An example of an instance family is `B-series`.

By default, Workload Optimization Manager considers all instance types that are currently available for scaling when making scaling decisions. However, you may have set up some of your cloud VMs to only scale to certain instance types to reduce complexity and cost, improve discount utilization, or meet application demand. To limit scaling to certain instance types, create policies for the affected cloud VMs and configure an inclusion or exclusion list as a scaling constraint.

Scale actions for powered off VMs

Workload Optimization Manager can recommend scale actions for VMs that are powered off (for example, during maintenance). If you accept the actions, Workload Optimization Manager scales the VMs while they are powered off so that the changes are already applied as soon as they are powered on.

Azure Billing target and scale actions

Workload Optimization Manager uses historical billing data discovered through your Azure Billing target *(on page 446)* to generate the most optimal scaling recommendations for VMs.

When VMs are associated with Azure subscriptions that Workload Optimization Manager monitors through your Azure Billing target, be aware of the following caveats for VM scale actions.

- After you add a new Azure Billing target, it may take Workload Optimization Manager up to 24 hours to generate scale actions for the affected VMs.
- After you add a new VM to a subscription, it may take Workload Optimization Manager up to two days to generate scale actions for the VM because Azure does not immediately add the VM to the billing data.

Scale actions for Azure VMs running GPU instance types

Currently, Workload Optimization Manager supports the following GPU instance type series with Linux source images.

- NC A100 v4-series (based on NVIDIA A100 PCIe GPUs)
- NCads H100 v5-series (based on NVIDIA H100 NVL GPUs)
- NCasT4_v3-series (based on Nvidia Tesla T4 GPUs)
- NCv3-series (based on NVIDIA Tesla V100 GPUs)
- NDv2-series (based on NVIDIA Tesla V100 GPUs)
- NVadsA10 v5-series (based on NVIDIA A10 GPUs)
- NVv3-series (based on NVIDIA Tesla M60 GPUs)

NOTE:
VMs that run NVadsA10 v5-series have partial GPUs. For these VMs, Workload Optimization Manager discovers partial GPU information and then adjusts GPU capacities (such as GPU memory) accordingly. Capacity information is available when you set the scope to an Azure GPU VM and view the Capacity and Usage chart.

In the Workload Optimization Manager user interface, the default virtual machine policy shows the currently supported GPU instance type series and the resources allocated to them. These instance series are grouped under the **Accelerated Computing** category. GPU instance series that Azure retired are grouped under the **GPU (Retired)** category. See the previous section for the steps to view the default virtual machine policy.

Workload Optimization Manager collects NVIDIA GPU metrics for VMs running these instance types and then uses these metrics to generate VM scale actions that optimize performance and costs.

NOTE:
For additional information about these metrics, see the NVIDIA documentation.

Metric	Description	Actions
GPU Count	Number of GPU cards in use	Scale down the number of GPU cards within the same instance type
GPU Memory	Amount of GPU memory in use	Scale GPU memory up or down within the same instance type
GPU Memory BW (bandwidth)	Fraction of cycles where data was sent to or received from device memory, measured in GB/s	Scale GPU memory BW up or down This action moves the VM from its current instance type to another instance type with the same (or higher) GPU count and GPU memory capacity.
GPU FP16	Fraction of cycles the FP16 (half precision) pipe was active	Scale GPU FP16 up or down This action moves the VM from its current instance type to another instance type with the same (or higher) GPU count and GPU memory capacity.
GPU FP32	Fraction of cycles the FP32 (single precision) pipe was active	Scale GPU FP32 up or down This action moves the VM from its current instance type to another instance type with the same (or higher) GPU count and GPU memory capacity.

Metric	Description	Actions
GPU FP64	Fraction of cycles the FP64 (double precision) pipe was active	Scale GPU FP64 up or down This action moves the VM from its current instance type to another instance type with the same (or higher) GPU count and GPU memory capacity.
GPU Tensor	Fraction of cycles the Tensor (mixed/multi-precision) pipe was active	Scale GPU Tensor up or down This action moves the VM from its current instance type to another instance type with the same (or higher) GPU count and GPU memory capacity.

NOTE:
For GPU VMs with accelerator cards, GPU card information is available when you set the scope to a VM and view the Entity Information chart. For example, the Accelerator Model field shows the GPU model.

To enable the discovery of GPU metrics, configure NVIDIA Data Center GPU Manager (DCGM) as described in this topic *(on page 472)*.

Workload Optimization Manager can also recommend actions that scale *standard VM resources* (such as vCPU and vMem) to the supported GPU instance types.

When generating or executing scale actions, Workload Optimization Manager considers the following policies and settings as constraints:

- Read-only tier exclusion policies

  Workload Optimization Manager automatically creates these policies and displays them in the Policy Management page (**Settings > Policies**).

  - Cross-target policies

    Cross-target policies ensure that VMs with certain GPU types only scale to an instance type with the same (or higher) GPU card count and memory per card.

    An example policy is `Azure GPU NVIDIA - Cloud Compute Tier Exclusion Policy`.

  - Per-target policies

    Per-target policies ensure that any VMs in GPU supported instance families do not scale out to instance families that do not support GPUs.

    An example policy is `Cloud Compute Tier Azure:gpu - Cloud Compute Tier Exclusion Policy`.

- Scaling target utilization

  Workload Optimization Manager uses scaling target utilization values for GPU resources in conjunction with aggressiveness constraints to control scale actions for VMs. You can configure utilization values in automation policies for cloud VMs. For more information, see this topic *(on page 663)*.

- Ignore NVIDIA GPU compute capability constraints

  This constraint is a setting that you can choose to turn on in automation policies for cloud VMs. When turned on, scale actions that change the GPU compute capability of VMs are allowed to execute in Workload Optimization Manager. When turned off, actions are only executable in the cloud provider console (web-based UI or CLI). For more information, see this topic *(on page 668)*.

Azure resource group discovery

For Azure environments that include Resource Groups, Workload Optimization Manager discovers the Azure Resource Groups and the tags that are used to identify these groups.

In the Workload Optimization Manager user interface, to search for a specific Azure Resource Group, choose **Resource Groups** in the Search Page.

You can set the scope of your Workload Optimization Manager session to an Azure Resource Group by choosing a group in the Search results and clicking **Scope To Selection**.

You can also use Azure tags as filter criteria when you create a custom Workload Optimization Manager resource group. You can choose the Azure Resource Groups that match the tag criteria to be members of the new custom group.

To find the available tags for a specific Azure Resource Group, add the Basic Info chart configured with Related Tag Information to your view or custom dashboard. See Basic Info Charts *(on page 973)*.

> **NOTE:**
> When you inspect Resource Groups, Workload Optimization Manager does not currently show the billed costs for those Resource Groups.

Azure instance requirements

In Azure environments, some instance types require workloads to be configured in specific ways, and some workload configurations require instance types that support specific features. When Workload Optimization Manager generates resize actions in Azure, these actions consider the following features:

- Accelerated Networking (AN)

  In an Azure environment, not all instance types support AN, and not all workloads on AN instances actually enable AN. Workload Optimization Manager maintains a dynamic group of workloads that have AN enabled, and it assigns a policy to that group to exclude any templates that do not support AN. In this way, if a workload is on an instance that supports AN, and that workload has enabled AN, then Workload Optimization Manager will not recommend an action that would move the workload to a non-AN instance.

- Azure Premium Storage

  Workload Optimization Manager recognizes whether a workload uses Premium Storage, and will not recommend a resize to an instance that does not support Azure Premium Storage.

In addition, Workload Optimization Manager recognizes processor types that you currently use for your workloads. If your workload is on a GPU-based instance, then Workload Optimization Manager will only recommend moves to other compatible GPU-based instance types. For these workloads, Workload Optimization Manager does not recommend resize actions.

IOPS-aware scaling for Azure VMs

Workload Optimization Manager considers IOPS utilization when making scaling decisions for Azure VMs. To measure utilization, Workload Optimization Manager takes into account a variety of attributes, such as per-disk IOPS utilization, whole VM IOPS utilization, cache settings, and IOPS capacity for the VMs. It also respects IOPS utilization and aggressiveness constraints that you set in VM policies. For details, see Aggressiveness and Observation Periods *(on page 665)*.

Analysis impacts VM scaling decisions in different ways. For example:

- If your instance experiences IOPS bottlenecks, Workload Optimization Manager can recommend scaling up to a larger instance type to increase IOPS capacity, even if you do not fully use the current VCPU or VMEM resources.
- If your instance experiences underutilization of VMEM and VCPU, but high IOPS utilization, Workload Optimization Manager might not recommend scaling down. It might keep you on the larger instance to provide sufficient IOPS capacity.
- If the instance experiences underutilization of IOPS capacity along with normal utilization of other resources, you might see an action to resize to an instance that is very similar to the current one. If you inspect the action details, you should see that you are changing to a less expensive instance with less IOPS capacity.

Actions for Google Cloud VMs

Workload Optimization Manager supports the following actions:

- Scale

  Change the VM instance to use a different instance type or tier to optimize performance and costs.

  See additional information for scale actions.

- Discount-related actions

  If you have a high percentage of on-demand VMs, you can reduce your monthly costs by increasing Committed Use Discount (CUD) coverage. To increase coverage, you scale VMs to instance types that have existing capacity.

- Stop and start (also known as 'parking' actions)

  Stop a VM for a given period of time to reduce your cloud expenses, and then start it at a later time.

  For details, see Parking: Stop or Start Cloud Resources *(on page 942)*.

- Reconfigure

  Google Cloud provides a specific set of machine types for each zone in a region. If you create a policy that restricts a VM to certain machine types and the zone it is currently on does not support all of those machine types, Workload Optimization Manager will recommend a reconfigure action as a way to notify you of the non-compliant VM.

  For example, assume Zone A does not support machine types for the M1 family. When a VM in that zone applies a policy that restricts it to M1, Workload Optimization Manager will recommend that you reconfigure the VM.

Points to consider for Google Cloud scale actions

- Workload Optimization Manager can generate scaling actions for the following VMs, but cannot execute the actions automatically:
  - VMs with local SSDs

    Workload Optimization Manager can recommend scaling to a machine type that supports local SSDs and the number of disks required by the VM, but blocks action execution due to prerequisite steps that you can only perform from Google Cloud. You can view the prerequisite steps when you examine a pending action.
  - VMs configured with a minimum CPU platform

    Google Cloud instance type families can support multiple CPU generations. A specific VM may be configured with a minimum CPU platform to prevent it from scaling to instance types with incompatible CPUs. When you examine a pending action for such a VM, verify that the recommended instance type runs a compatible CPU. Once verified, manually execute the action from Google Cloud.
- Workload Optimization Manager does not recognize prioritized attributions you may have set for CUDs. For example, if you have prioritized all your CUD allotments for a single project, Workload Optimization Manager can still recommend actions to apply CUD to other projects in your environment.
- Since all Google Cloud compute tiers have the same net throughput capacity, Workload Optimization Manager will not generate scaling actions in response to net throughput.
- For actions to scale VMs running custom machine types, Workload Optimization Manager currently supports the E2 custom (including shared-core), N1, N2, and N2D machine series.
- Workload Optimization Manager does not recommend scaling actions for:
  - Spot VMs

    NOTE:
    Workload Optimization Manager discovers spot VMs, but does not recommend actions or monitor costs for these VMs.

  - VMs running on sole-tenant nodes
  - VMs with attached GPUs
  - VMs in managed instance groups

Controlling scale actions for Google Cloud VMs

For scale actions, you can create policies to control the scale actions that Workload Optimization Manager recommends. In those policies, choose from the following options:

- Cloud scale all – execute all scaling actions
- Cloud scale for performance – only execute scaling actions that improve performance
- Cloud scale for savings – only execute scaling actions that reduce costs

The default action acceptance mode for these actions is *Manual*. When you examine the pending actions, only actions that satisfy the policy are allowed to execute. All other actions are read-only.

When policy conflicts arise, **Cloud scale all** overrides the other two scaling options in most cases. For more information, see Default and User-defined Automation Policies *(on page 1050)*.

Supported instance types for Google Cloud VM scale actions

In the user interface, you can view the instance types that Workload Optimization Manager currently supports.

1. Navigate to **Settings > Policies**.
2. In the Policy Management page, search for and click **Virtual Machine Defaults**.
3. In the Configure Virtual Machine Policy page:
    a. Scroll to the end of the page.
    b. Click **Add Scaling Constraint**.
    c. Choose **Cloud Instance Types**.
    d. Click **Edit**.

Expand an instance family to see individual instance types and the resources allocated to them. An example of an instance family is A2.

By default, Workload Optimization Manager considers all instance types that are currently available for scaling when making scaling decisions. However, you may have set up some of your cloud VMs to only scale to certain instance types to reduce complexity and cost, improve discount utilization, or meet application demand. To limit scaling to certain instance types, create policies for the affected cloud VMs and configure an inclusion or exclusion list as a scaling constraint.

Scale actions for powered off VMs

Workload Optimization Manager can recommend scale actions for VMs that are powered off (for example, during maintenance). If you accept the actions, Workload Optimization Manager scales the VMs while they are powered off so that the changes are already applied as soon as they are powered on.

Google Cloud Billing target and scale actions

Workload Optimization Manager uses historical billing data discovered through your Google Cloud Billing target *(on page 417)* to generate the most optimal scaling recommendations for VMs.

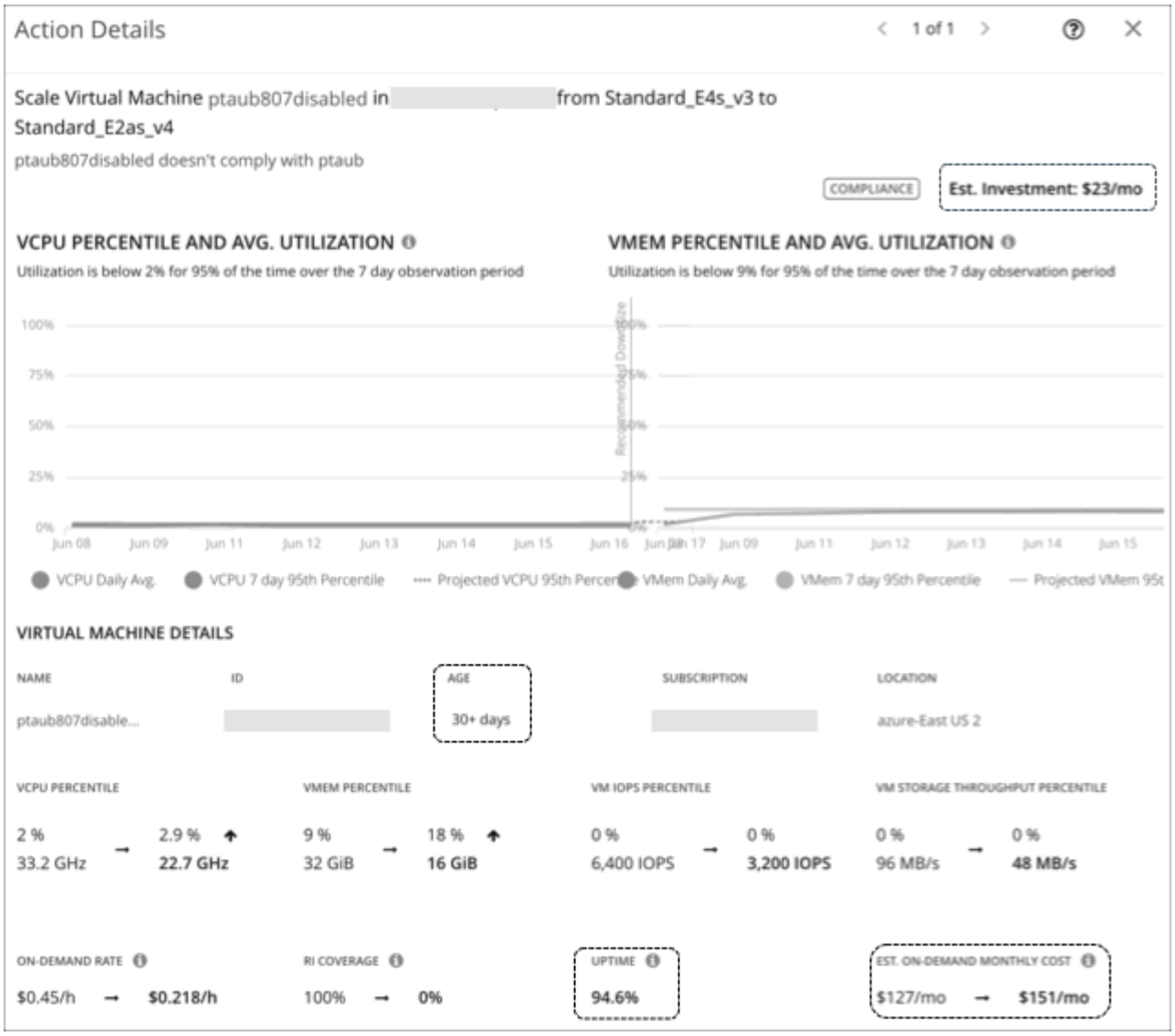
When VMs are associated with Google Cloud projects that Workload Optimization Manager monitors through your Google Cloud Billing target, be aware that it may take Workload Optimization Manager up to two days to generate scale actions for any new VM added to a project because Google Cloud does not immediately add the VM to the billing data.

Cloud VM uptime

For cloud VMs, Workload Optimization Manager includes *uptime* data in its cost calculations. This is especially important for VMs that do not run 24/7 and are charged on-demand rates. With uptime data, Workload Optimization Manager can calculate costs more accurately based on the amount of time a VM has been running.

The Action Details page shows uptime data for these VMs. Workload Optimization Manager calculates uptime based on the VM's age.

NOTE:
If you added an AWS, Azure, or Google Cloud Billing target to Workload Optimization Manager, uptime is retrieved directly from your billing data.



Key concepts

- Uptime

  A percentage value that indicates how long a VM has been running over a period of time (age)

- Age

  The number of days that a VM has existed since first discovery. For VMs older than 30 days, Workload Optimization Manager displays a value of **30+ days**, but only calculates uptime over the last 30 days.

For newly discovered VMs, age is 0 (zero) on the day of discovery. If the VM is running at the time of discovery, uptime is 100%. Otherwise, uptime is 0% and remains unchanged until the VM is powered on. Workload Optimization Manager recalculates uptime every hour and then refreshes the data shown in the user interface.

examples

- A VM that was first discovered 5 days (or 120 hours) ago and has been running for a total of 60 hours during that period has a current uptime value of 50%.
- A VM that was first discovered 2 months ago and has been running for a total of 180 hours over the last 30 days (or 720 hours) has a current uptime value of 25%.

Cost calculations using uptime data

Workload Optimization Manager uses uptime data to calculate estimated on-demand costs for your cloud VMs. For details about calculations, see Estimated On-demand Monthly Costs for Cloud VMs *(on page 651)*.

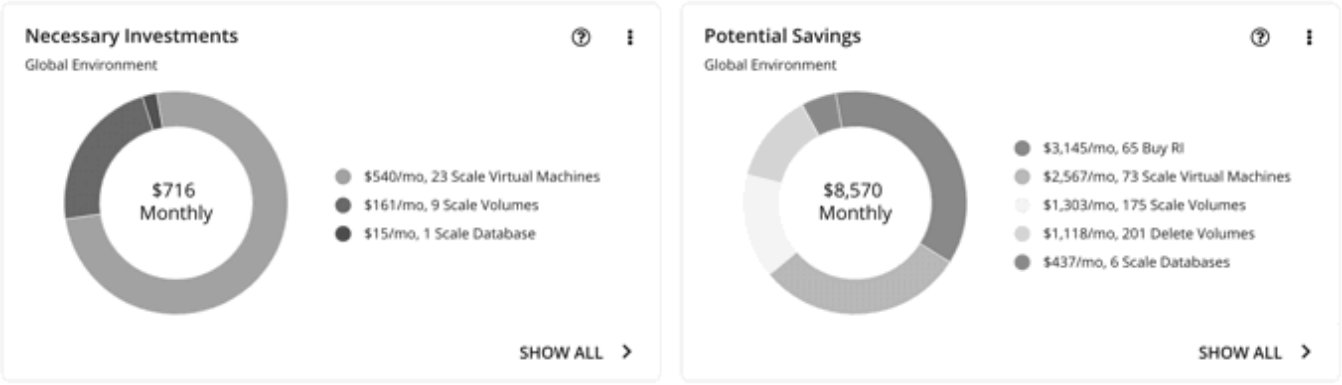
Uptime data impacts cost calculations, but not the actual scaling decisions that Workload Optimization Manager makes. These decisions rely on other factors, such as resource utilization percentiles and scaling constraints set in policies.

Uptime data in charts

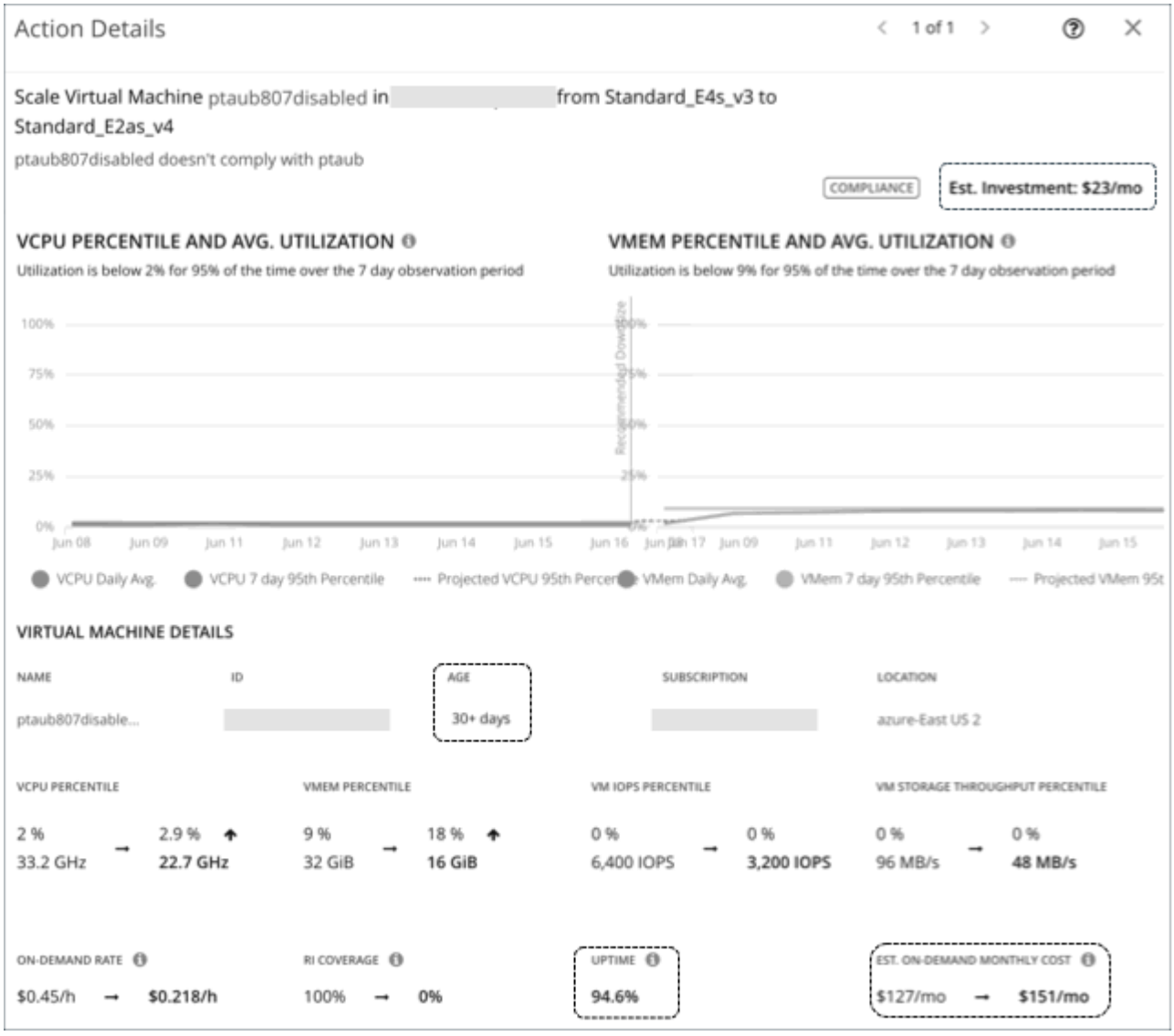
Workload Optimization Manager recalculates uptime data every hour and then updates the values shown in charts. The following charts reflect the cost impact of uptime-based calculations:

- Potential Savings and Necessary Investment charts

  The projected amounts in these charts include on-demand costs for cloud VMs.

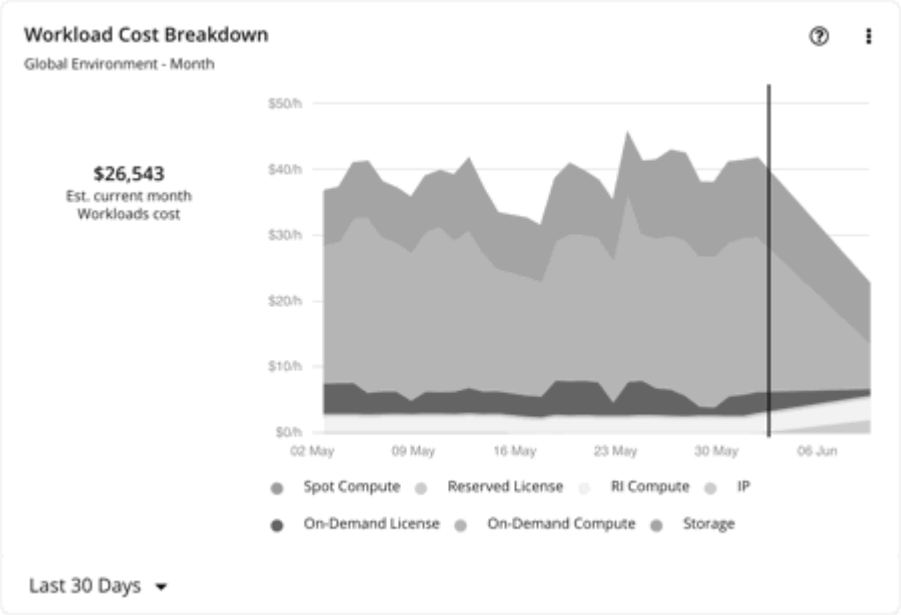


  When you click **Show All** in these charts and view details for a pending VM action, the **Action Details** page shows on-demand costs before and after you execute the action, factoring in the VM's uptime value. The page also shows the VM's age.

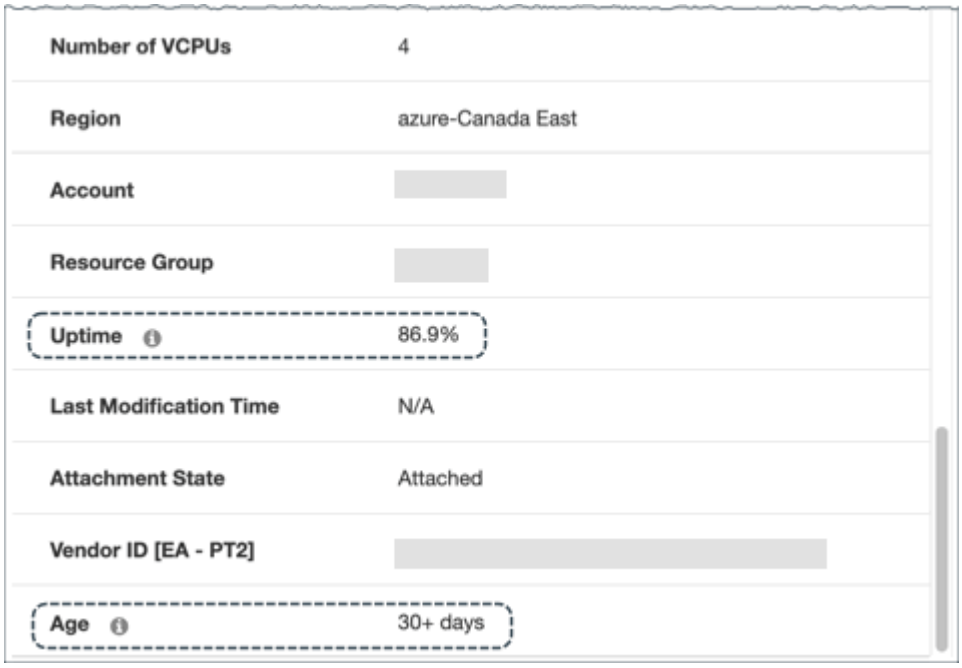


- Workload Cost Breakdown chart

  This chart shows estimated costs over time, including on-demand costs for VMs.

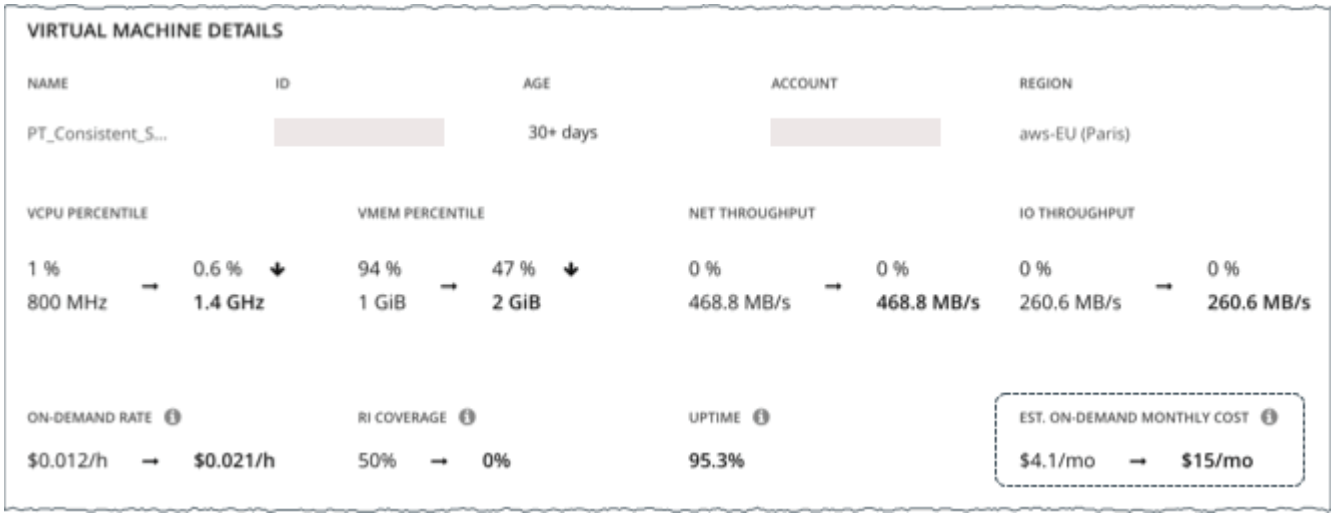


- The **Entity Information** chart shows the latest uptime and age data for a specific cloud VM.



Estimated on-demand costs for cloud VMs

Workload Optimization Manager considers a variety of factors when calculating *Estimated On-demand Monthly Cost* for a cloud VM.



Points to consider

On-demand cost is the uncovered usage cost that takes into account the AWS-published on-demand rates and the VM's uptime. Effective cost takes into account the on-demand cost and the amortized costs for the associated Reserved Instances and Compute Savings Plans.

On-demand cost is derived from your billing data. This data may sometimes not align with the calculations described in this topic for the following reasons:

- Billing data might include discounts that Workload Optimization Manager does not currently consider in its analysis. For a list of discounts that are considered in analysis, see Discounts *(on page 157)*.
- Billing data might include additional costs for VM resources that Workload Optimization Manager does not yet support. For example, your cloud provider might offer an expensive OS that Workload Optimization Manager does not yet support. In this case, billing data reflects real on-demand cost that is significantly higher than the cost calculated by Workload Optimization Manager.

Cost calculation: AWS VMs and Azure VMs without license costs

For AWS VMs and Azure VMs without license costs, the calculation for estimated on-demand monthly cost can be expressed as follows:

```
On-demand Rate * Usage Not Covered by Discounts * Uptime * 730 = Estimated On-demand Monthly Cost
```

Where:

- **On-demand Rate** is the hourly cost for a VM's instance type without discount coverage (AWS RIs/Savings Plans or Azure reservations).
  - For AWS, this rate includes all license costs, but not storage or IP. For more information about on-demand rates, see this AWS page.
  - For Azure, the rate does *not* include license costs, storage, or IP. For more information about on-demand rates, see this Azure page.

    NOTE:
    Azure VMs covered by Azure Hybrid Benefit do not have license costs.

- **Usage Not Covered by Discounts** is the percentage of hourly VM usage not covered by any discount. For example:
  - Discount Coverage = 20% (0.2)
  - Usage Not Covered by Discounts = 80% (0.8)

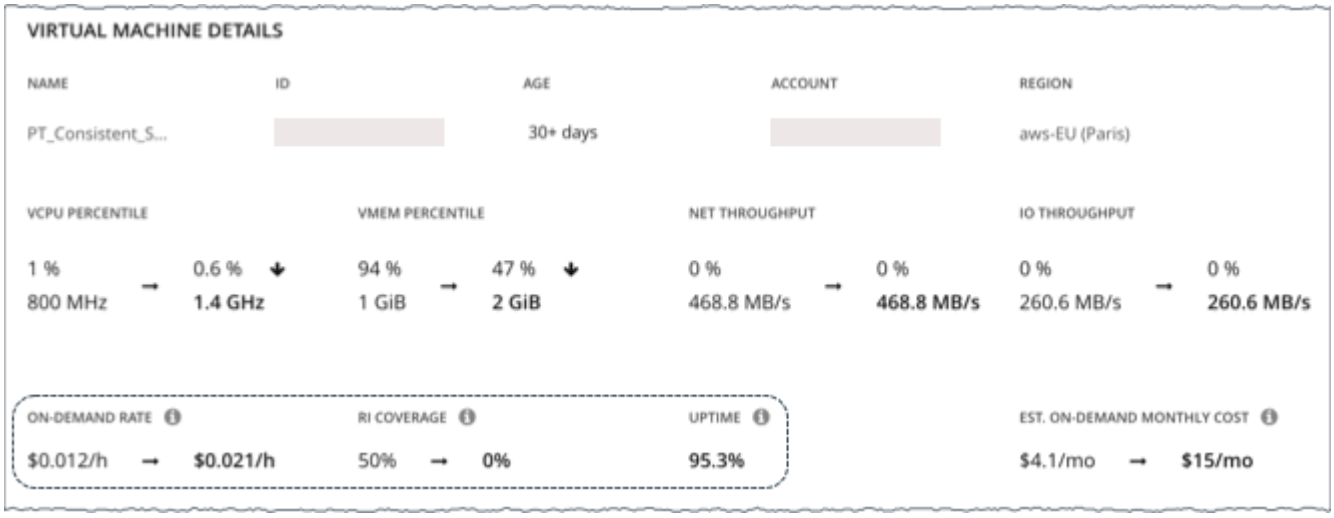
- **Uptime** is a percentage value that indicates how long a VM has been running since it was first discovered by Workload Optimization Manager. For VMs discovered more than 30 days ago, Workload Optimization Manager only calculates uptime over the last 30 days.

  To estimate monthly on-demand costs, Workload Optimization Manager projects the current uptime value into the future. It assumes that future uptime will be similar to the current uptime.
- **730** is the number of hours per month that Workload Optimization Manager uses to estimate monthly costs.

The preceding items impact cost calculations, but not the actual scaling decisions that Workload Optimization Manager makes. These decisions rely on other factors, such as resource utilization percentiles and scaling constraints set in policies.

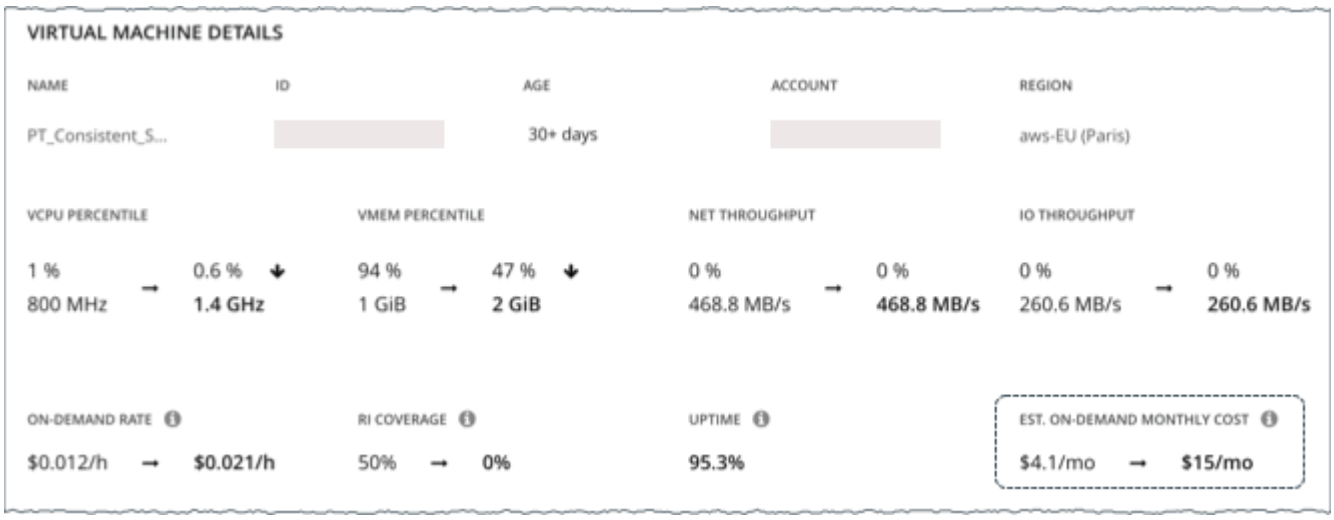
Cost calculation example: AWS VMs and Azure VMs without license costs

Assume the following data for a pending scale action for an AWS VM:



Item	Current values	Values after action execution
On-demand Rate	$0.012/hr	$0.021/hr
Discount Coverage	50% (0.5)	0% (0.0)
Usage Not Covered by Discounts (calculated based on discount coverage)	50% (0.5)	100% (1.0)
Uptime	95.3% (.953)	

Workload Optimization Manager calculates the following values:



- Current estimated on-demand monthly cost:

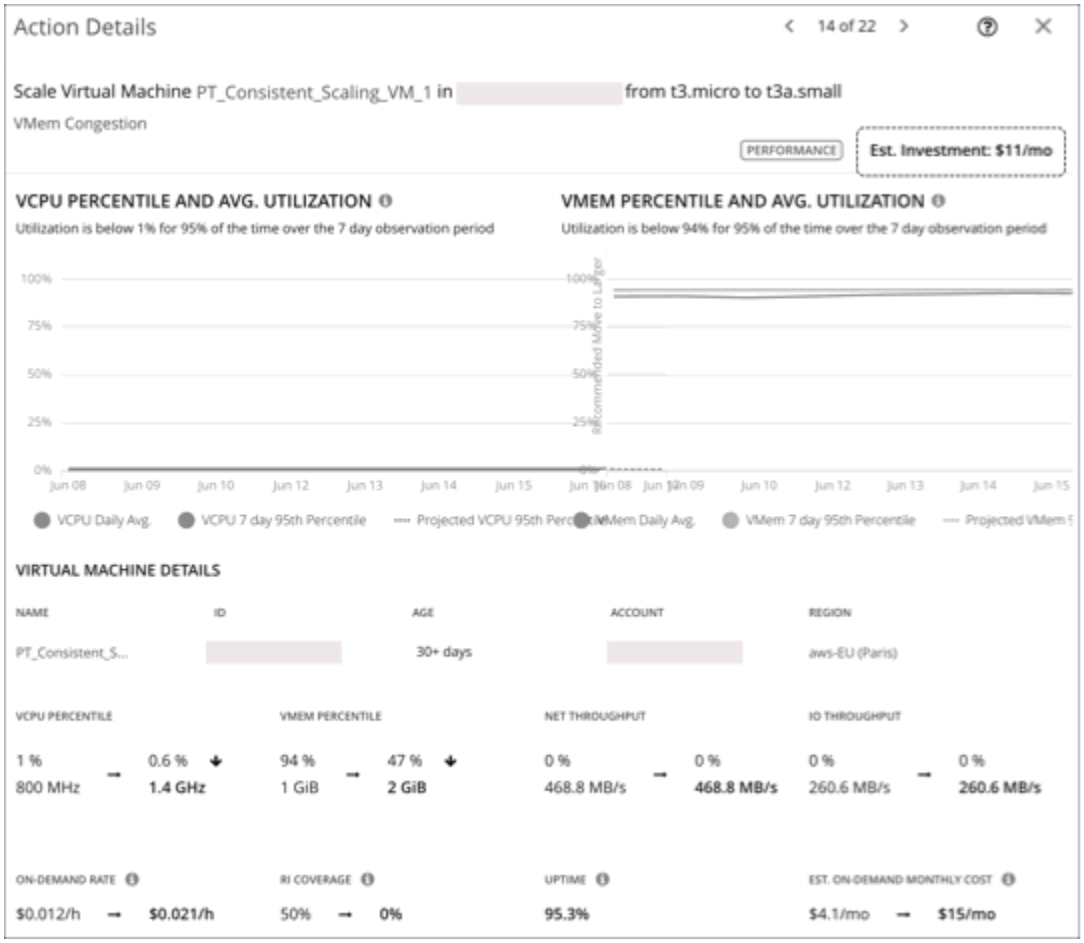
```
0.012 * 0.5 * 0.953 * 730 = 4.1
```

- Estimated on-demand monthly cost after executing the action:

```
0.021 * 1.0 * 0.953 * 730 = 15
```

NOTE:
Workload Optimization Manager rounds the calculated values that it displays in the user interface.

Since the estimated on-demand monthly cost is projected to increase from $4.1/month to $15/month, Workload Optimization Manager treats the action as an investment and shows an estimated investment of $11/month.



Cost calculation: Azure VMs with license costs

For Azure VMs with license costs, Workload Optimization Manager first calculates the on-demand compute rate, which it then uses to calculate estimated on-demand monthly costs.

The calculation for on-demand compute rate can be expressed as follows:

```
On-demand Rate - (Reserved License Cost + On-demand License Cost) = On-demand Compute Rate
```

Where:

- **On-demand Rate** is the hourly cost for a VM's instance type *without* reservation coverage. This does *not* include license costs, storage, or IP. For more information about on-demand rate, see this Azure page.
- **Reserved License Cost** and **On-demand License Cost** are the hourly costs for the VM's licenses. For more information about these costs, see this Azure page.

The on-demand compute rate and license cost (reserved and on-demand) are then used to calculate estimated on-demand monthly costs.

The calculation for on-demand monthly costs can be expressed as follows:

```
(On-demand Compute Rate * Usage Not Covered by Reservations) + License Cost * Uptime * 730 = Estimated On-demand Monthly Cost
```

Where:

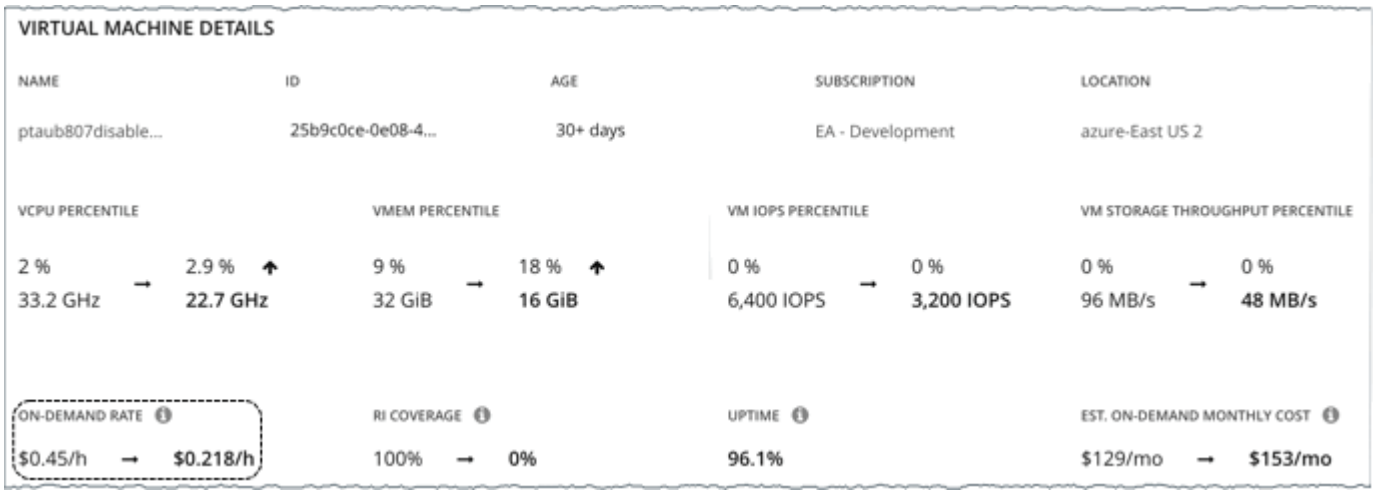
- **Usage Not Covered by Reservations** is the percentage of hourly VM usage not covered by any reservation. For example:
  - Reservation Coverage = 20% (0.2)
  - Usage Not Covered by Reservations = 80% (0.8)
- **License Cost** is the sum of On-demand License Cost and Reserved License Cost.
- **Uptime** is a percentage value that indicates how long a VM has been running since it was first discovered by Workload Optimization Manager. For VMs discovered more than 30 days ago, Workload Optimization Manager only calculates uptime over the last 30 days

  To estimate monthly on-demand costs, Workload Optimization Manager projects the current uptime value into the future. It assumes that future uptime will be similar to the current uptime.
- **730** is the number of hours per month that Workload Optimization Manager uses to estimate monthly costs.

The preceding listed items impact cost calculations, but not the actual scaling decisions that Workload Optimization Manager makes. These decisions rely on other factors, such as resource utilization percentiles and scaling constraints set in policies.

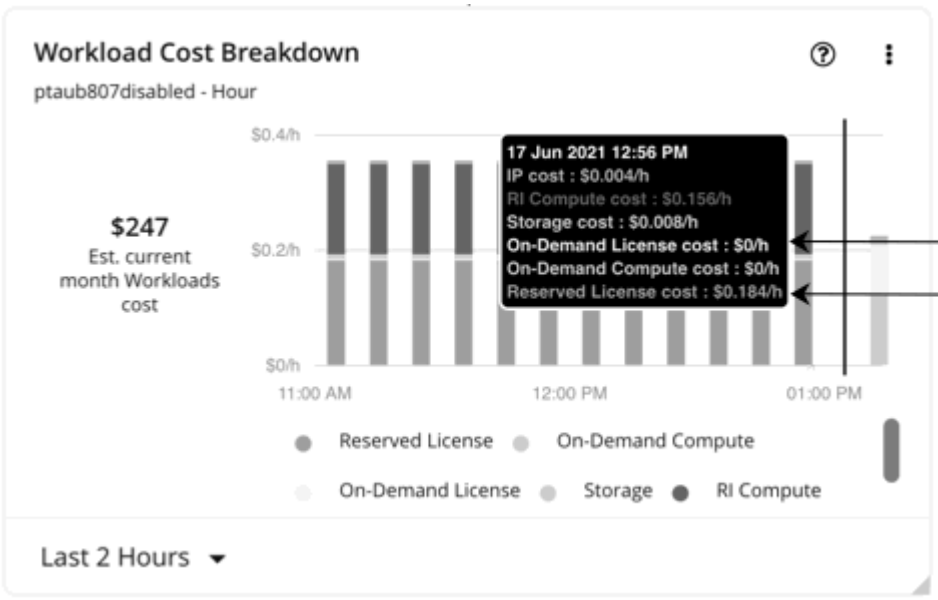
Cost calculation example: Azure VMs with license costs

Assume the following data for a pending scale action for an Azure VM with license costs:

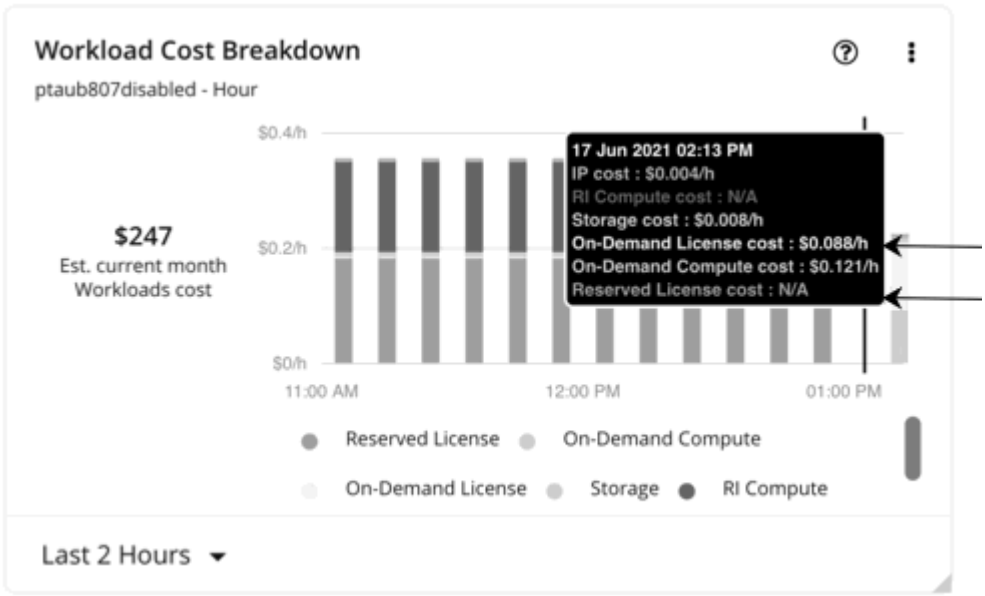


Item	Current Values	Values After Action Execution
On-demand Rate	$0.45/hr	$0.218/hr

Current license costs



License costs after action execution



Item	Current values	Values after action execution
On-demand License Cost	$0/hr	$0.088/hr
Reserved License Cost	$0.184/hr	N/A

1. Workload Optimization Manager calculates the following values:
   - Current on-demand compute rate:

     ```
     0.45 - (0.184 + 0) = 0.266
     ```

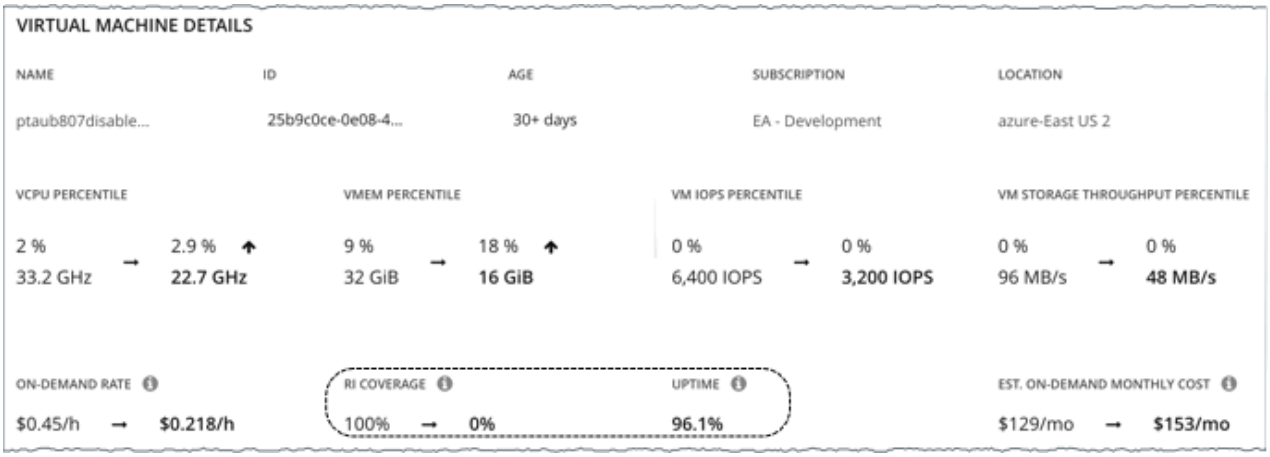
   - On-demand compute rate after executing the action:

     ```
     0.218 - (0 + 0.088) = 0.13
     ```

2. Workload Optimization Manager calculates estimated on-demand monthly cost based on the following values:
   - On-demand Compute Rate

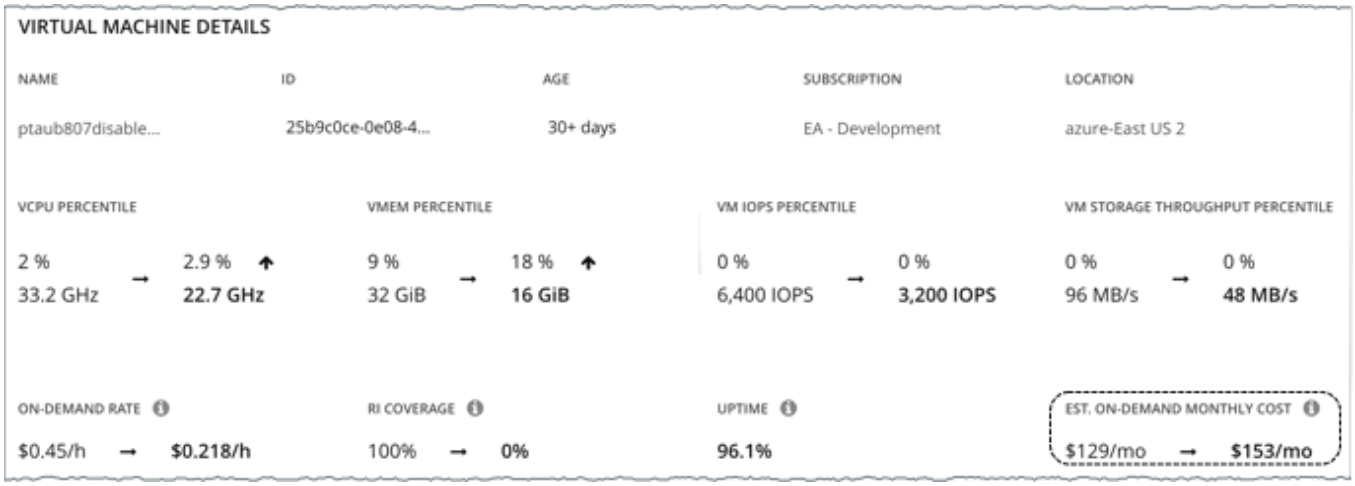
Item	Current values	Values after action execution
On-demand Compute Rate	$0.266/hr	$0.13/hr

   - Usage Not Covered by Reservations and Uptime



Item	Current values	Values after action execution
Reservation Coverage	100% (1.0)	0% (0.0)
Usage Not Covered by Reservations (calculated based on reservation coverage)	0% (0.0)	100% (1.0)
Uptime	96.1% (.961)	

Workload Optimization Manager calculates the following values:



   - Current estimated on-demand monthly cost:

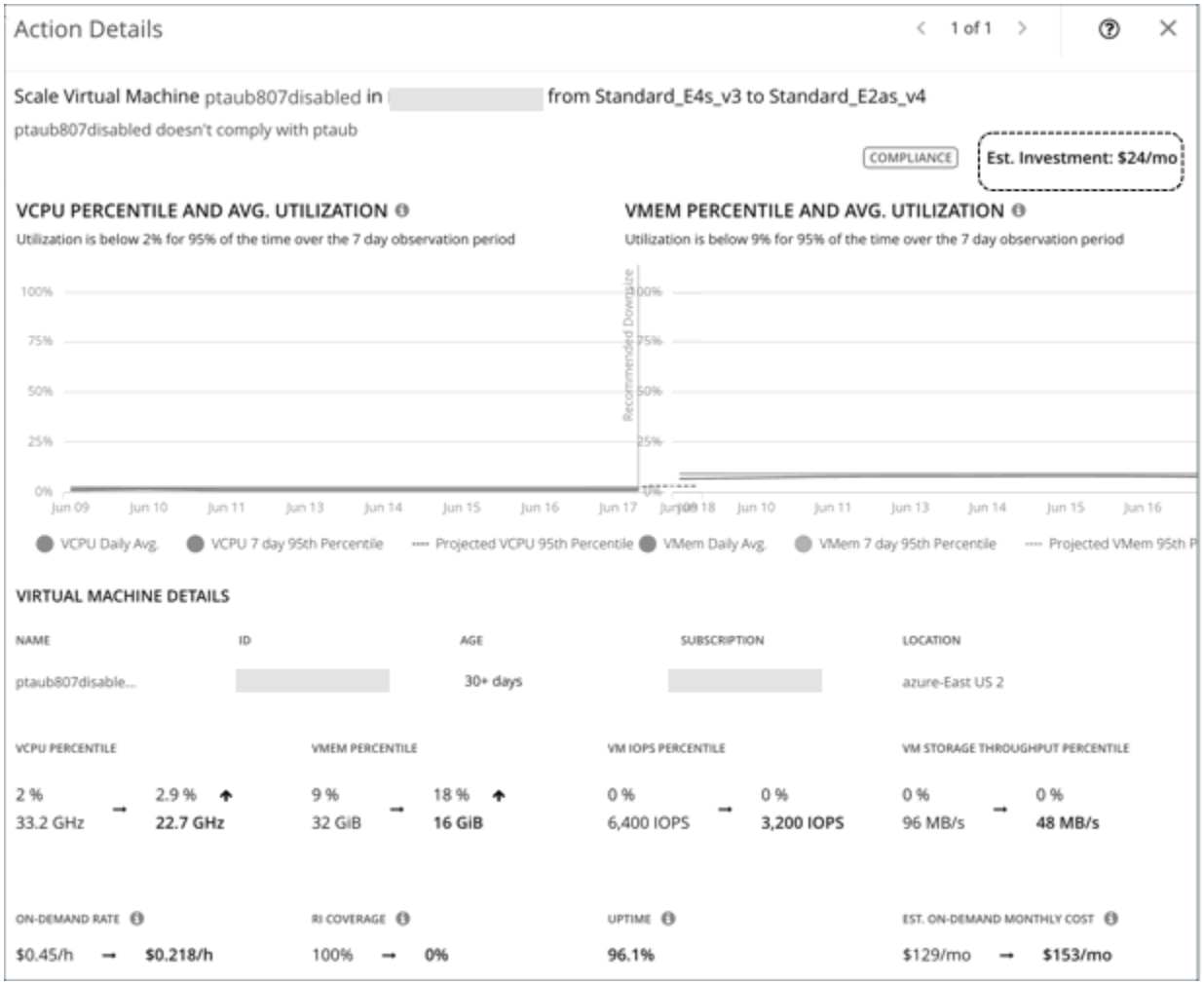
```
(0.266 * 0.0) + 0.184 * 0.961 * 730 = 129
```

- Estimated on-demand monthly cost after executing the action:

```
(0.13 * 1.0) + 0.088 * 0.961 * 730 = 153
```

NOTE:
Workload Optimization Manager rounds the calculated values that it displays in the user interface.

Since the on-demand cost is projected to increase from $129/month to $153/month, Workload Optimization Manager treats the action as an investment and shows an estimated investment of $24/month.



Effective costs for cloud VMs

Workload Optimization Manager considers a variety of factors when calculating *effective monthly cost* for a cloud VM.

Effective cost takes into account the on-demand cost and the amortized costs for the associated Reserved Instances and Compute Savings Plans.

NOTE:
Currently, Workload Optimization Manager does not calculate effective costs for cloud VMs in AWS GovCloud.

Cost calculation: AWS VMs covered by Compute Savings Plans

For AWS VMs covered by Compute Savings Plans, the calculation for effective cost can be expressed as follows where each discount type is computed separately:

On-demand hourly cost:

```
(1 - combined coverage percentage) * uptime hours * (on-demand rate)
```

Effective hourly cost:

- **Savings Plan (SP) hourly cost**: `SP coverage percentage * (SP rate)`
- **Reserved instance (RI) hourly cost**: `RI coverage percentage * (RI rate)`
- **On-demand hourly cost**: `(1- combined coverage percentage) * (on-demand rate)`

Effective monthly cost: `(onDemandCost + spCoveredCost + riCoveredCost) * uptime hours`

Where:

- **Combined coverage percentage (Discount Coverage)** is the total discount coverage which includes savings plan coverage and reserved instance coverage.
  - The calculation can be expressed as follows:

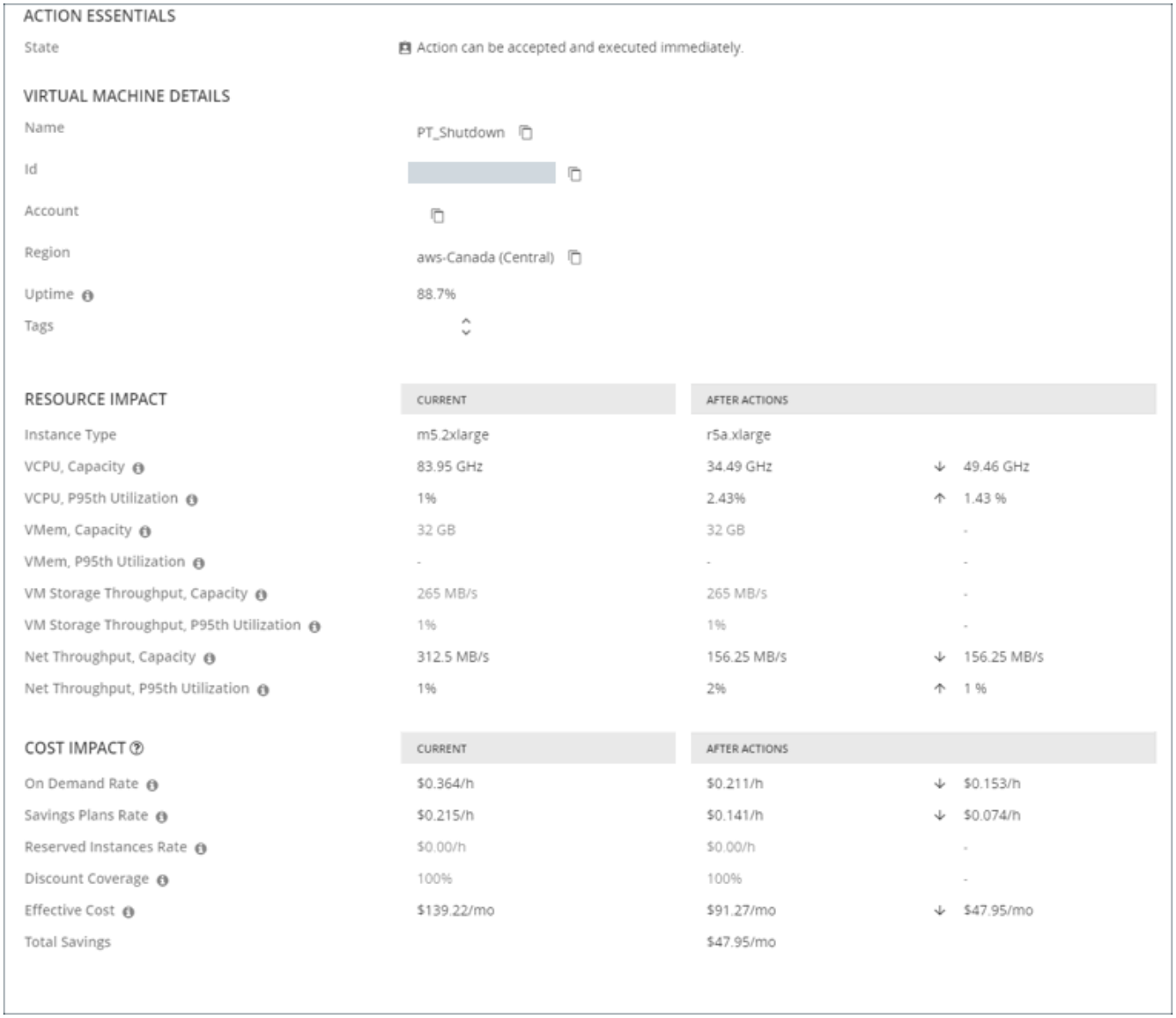
    ```
    SP coverage percentage + RI coverage percentage
    ```
- **Uptime** is a percentage value that indicates how long a VM has been running since it was first discovered by Workload Optimization Manager. For VMs discovered more than 30 days ago, Workload Optimization Manager only calculates uptime over the last 30 days.

  To estimate monthly on-demand costs, Workload Optimization Manager projects the current uptime value into the future. It assumes that future uptime will be similar to the current uptime.

  Uptime hours is the same as `uptime percentage * 730`.
- **On-demand rate** is the hourly on-demand cost for a VM's instance type. This rate includes all license costs, but not storage or IP. For more information about on-demand rates, see this AWS page.
- **730** is the number of hours per month that Workload Optimization Manager uses to estimate monthly costs.
- **SP coverage percentage** is the percentage of hourly VM usage that is covered by a savings plan.
- **RI coverage percentage** is the percentage of hourly VM usage that is covered by a reserved instance.
- **SP rate** is the discounted hourly cost for a VM covered by a compute savings plan, based on offering rates. Custom rates are not considered.
- **RI rate** is the discounted hourly cost for a VM covered by a reserved instance, based on offering rates. Custom rates are not considered.
- **onDemandCost** is the total cost of the VM's hourly usage charged at the on-demand rate.
- **spCoveredCost** is the total cost of the VM's savings plan covered hourly usage charged at the savings plan rate.
- **riCoveredCost** is the total cost of the VM's reserved instance covered hourly usage charged at the reserved instance rate.

Cost calculation example: AWS VMs covered by Compute Savings Plans

Assume the following data for a pending scale action for an AWS VM covered by Compute Savings Plans:



ACTION ESSENTIALS

State	Action can be accepted and executed immediately.

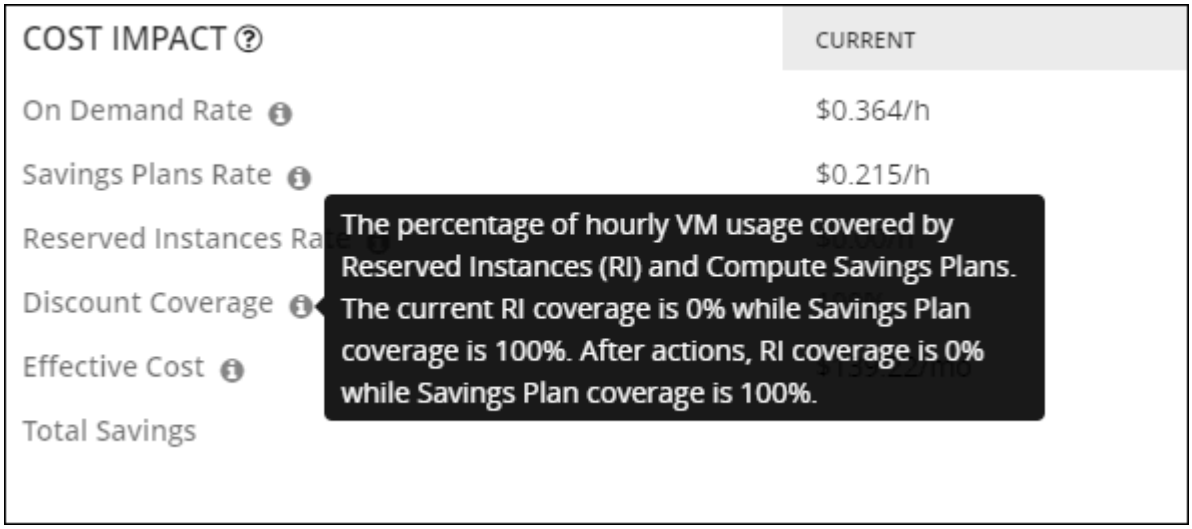
VIRTUAL MACHINE DETAILS

Name	PT_Shutdown
Id	
Account	
Region	aws-Canada (Central)
Uptime	88.7%
Tags	

RESOURCE IMPACT	CURRENT	AFTER ACTIONS	
Instance Type	m5.2xlarge	r5a.xlarge	
VCPU, Capacity	83.95 GHz	34.49 GHz	↓ 49.46 GHz
VCPU, P95th Utilization	1%	2.43%	↑ 1.43 %
VMem, Capacity	32 GB	32 GB	-
VMem, P95th Utilization	-	-	-
VM Storage Throughput, Capacity	265 MB/s	265 MB/s	-
VM Storage Throughput, P95th Utilization	1%	1%	-
Net Throughput, Capacity	312.5 MB/s	156.25 MB/s	↓ 156.25 MB/s
Net Throughput, P95th Utilization	1%	2%	↑ 1 %

COST IMPACT	CURRENT	AFTER ACTIONS	
On Demand Rate	$0.364/h	$0.211/h	↓ $0.153/h
Savings Plans Rate	$0.215/h	$0.141/h	↓ $0.074/h
Reserved Instances Rate	$0.00/h	$0.00/h	-
Discount Coverage	100%	100%	-
Effective Cost	$139.22/mo	$91.27/mo	↓ $47.95/mo
Total Savings		$47.95/mo	

The following image shows the calculation populated in the `Discount Coverage` tooltip.



Item	Current values	Values after action execution
On-demand Rate	$0.364/hr	$0.211/hr
Savings Plans Rate	$0.215/hr	$0.141/hr
Reserved Instances Rate	$0.00/hr	$0.00/h
Discount Coverage	100%	100%
SP coverage	100%	100%
RI coverage	0%	0%
Uptime	88.7% (.887)	

Workload Optimization Manager calculates the following values:

- On-demand hourly cost:

```
(1- 1) * 0.364 = $0
```

- Savings Plan (SP) hourly cost

```
(1 * 0.215) = $0.215
```

- Reserved instance (RI) hourly cost

```
(0 * 0) = $0
```

- Effective cost before executing the action:

```
(0 + 0.215 + 0) * (730 * 0.887) = $139.2
```

- Effective cost after executing the action:

```
(0 + 0.141 + 0) * (730 * 0.887) = $91.2
```

NOTE:
Workload Optimization Manager rounds the calculated values that it displays in the user interface so the results of the calculation may not match the values in the Action Details page.

Since the effective cost is projected to decrease from $139.2/month to $91.2/month, Workload Optimization Manager treats the action as a cost-saving measure and shows estimated savings of $48/month.

Cloud VM policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

> **NOTE:**
> The policy settings discussed in this topic only apply to cloud scale actions. For details about settings for parking actions (stop and start), see Parking: Stop or Start Cloud Resources *(on page 942)*.

Automation workflow

For details about cloud VM actions, see Cloud VM Actions *(on page 637)*.

Cloud Scale

Action	Default mode	AWS, Azure, and Google Cloud
Cloud Scale All	Manual	Automatable
Cloud Scale for Performance	Manual	Automatable
Cloud Scale for Savings	Manual	Automatable

Other Actions

Action	Default mode	AWS and Azure	Google Cloud
Provision container platform node (VM)	Manual	Recommend	Recommend
Suspend container platform node (VM)	Manual	Automatable	Automatable

Scaling target utilization - GPU (AWS and Azure only)

Workload Optimization Manager uses these settings in conjunction with aggressiveness constraints to control scaling actions for AWS *(on page 638)* or Azure *(on page 643)* VMs running supported NVIDIA GPU instance types. You can set the aggressiveness per the percentile of utilization, and set the length of the sample period for more or less elasticity on the cloud.

Attribute	Default value
Scaling Target GPU Count Utilization	100 The target utilization as a percentage of GPU count capacity.
Scaling Target GPU FP16 Utilization	90 The target utilization as a percentage of GPU FP16 capacity.
Scaling Target GPU FP32 Utilization	90 The target utilization as a percentage of GPU FP32 capacity.
Scaling Target GPU FP64 Utilization	90 The target utilization as a percentage of GPU FP64 capacity.
Scaling Target GPU Memory BW Utilization	90

Attribute	Default value
	The target utilization as a percentage of GPU Memory BW (bandwidth) capacity.
Scaling Target GPU Memory Utilization	90 The target utilization as a percentage of GPU memory capacity.
Scaling Target GPU Tensor Utilization	90 The target utilization as a percentage of GPU Tensor capacity.

Scaling target utilization - IOPS (Azure only)

Workload Optimization Manager uses this setting in conjunction with aggressiveness constraints to control scaling actions for VMs. You can set the aggressiveness per the percentile of utilization, and set the length of the sample period for more or less elasticity on the cloud.

Attribute	Default value
Scaling Target IOPS Utilization	70 The target percentile value Workload Optimization Manager will attempt to match.

For details on how IOPS utilization affects scaling decisions, see IOPS-aware Scaling for Azure VMs *(on page 645)*.

Scaling target utilization - vCPU, vMem, IO/Net throughput

These advanced settings determine how much you would like a scope of workloads to utilize their resources. These are fixed settings that override the way Workload Optimization Manager calculates the optimal utilization of resources. You should only change these settings after consulting with Technical Support.

While these settings offer a way to modify how Workload Optimization Manager recommends actions, in most cases you should never need to use them. If you want to control how Workload Optimization Manager recommends actions to resize workloads, you can set the aggressiveness per the percentile of utilization, and set the length of the sample period for more or less elasticity on the cloud.

Attribute	Default value
Scaling Target VCPU Utilization	70 The target utilization as a percentage of VCPU capacity.
Scaling Target VMEM Utilization	90 The target utilization as a percentage of memory capacity.
Scaling Target IO Throughput Utilization	70 The target utilization as a percentage of IO throughput (Read and Write) capacity.
Scaling Target Net Throughput Utilization	70 The target utilization as a percentage of network throughput (Inbound and Outbound) capacity.

Aggressiveness and observation periods

Workload Optimization Manager uses these settings to calculate utilization percentiles and then recommends actions to improve utilization based on the observed values for a given time period.

- Aggressiveness

Attribute	Default value
Aggressiveness	95th Percentile

When evaluating performance, Workload Optimization Manager considers resource utilization as a percentage of capacity. The utilization drives actions to scale the available capacity either up or down. To measure utilization, the analysis considers a given utilization percentile. For example, assume a 95th percentile. The percentile utilization is the highest value that 95% of the observed samples fall below. Compare that to average utilization, which is the average of *all* the observed samples.

Using a percentile, Workload Optimization Manager can recommend more relevant actions. This is important in the cloud, so that analysis can better exploit the elasticity of the cloud. For scheduled policies, the more relevant actions will tend to remain viable when their execution is put off to a later time.

For example, consider decisions to reduce the capacity for CPU on a VM. Without using a percentile, Workload Optimization Manager never resizes below the recognized peak utilization. For most VMs, there are moments when peak CPU reaches high levels, such as during reboots, patching, and other maintenance tasks. Assume utilization for a VM peaked at 100% just once. Without the benefit of a percentile, Workload Optimization Manager will not reduce allocated CPU for that VM.

With **Aggressiveness**, instead of using the single highest utilization value, Workload Optimization Manager uses the percentile you set. For the previous example, assume a single CPU burst to 100%, but for 95% of the samples CPU never exceeded 50%. If you set **Aggressiveness** to 95th Percentile, then Workload Optimization Manager can see this as an opportunity to reduce CPU allocation for the VM.

In summary, a percentile evaluates the sustained resource utilization, and ignores bursts that occurred for a small portion of the samples. You can think of this as aggressiveness of resizing, as follows:

  - 100th and 99th Percentile – More performance. Recommended for critical workloads that need maximum guaranteed performance at all times, or workloads that need to tolerate sudden and previously unseen spikes in utilization, even though sustained utilization is low.
  - 95th Percentile (Default) – The recommended setting to achieve maximum performance and savings. This assures application performance while avoiding reactive peak sizing due to transient spikes, thus allowing you to take advantage of the elastic ability of the cloud.
  - 90th Percentile – More efficiency. Recommended for non-production workloads that can stand higher resource utilization.

By default, Workload Optimization Manager uses samples from the last 30 days. Use the **Max Observation Period** setting to adjust the number of days. To ensure that there are enough samples to analyze and drive scaling actions, set the **Min Observation Period**.

- Max Observation Period

Attribute	Default value
Max Observation Period	Last 30 Days

To refine the calculation of resource utilization percentiles, you can set the sample time to consider. Workload Optimization Manager uses historical data from up to the number of days that you specify as a sample period. If the database has fewer days' data then it uses all of the stored historical data.

You can make the following settings:

  - Less Elastic – Last 90 Days
  - Recommended – Last 30 Days
  - More Elastic – Last 7 Days

Workload Optimization Manager recommends an observation period of 30 days following the monthly workload maintenance cycle seen in many organizations. VMs typically peak during the maintenance window as patching and

other maintenance tasks are carried out. A 30-day observation period means that Workload Optimization Manager can capture these peaks and increase the accuracy of its sizing recommendations.

You can set the value to 7 days if workloads need to resize more often in response to performance changes. For workloads that cannot handle changes very often or have longer usage periods, you can set the value to 90 days.

- Min Observation Period

Attribute	Default value
Min Observation Period	None

This setting ensures historical data for a minimum number of days before Workload Optimization Manager will generate an action based on the percentile set in **Aggressiveness**. This ensures a minimum set of data points before it generates the action.

Especially for scheduled actions, it is important that resize calculations use enough historical data to generate actions that will remain viable even during a scheduled maintenance window. A maintenance window is usually set for "down" time, when utilization is low. If analysis uses enough historical data for an action, then the action is more likely to remain viable during the maintenance window.

  - More Elastic – None
  - Less Elastic – 7 Days

Cloud instance types

Attribute	Default value
Cloud Instance Types	None

By default, Workload Optimization Manager considers all instance types that are currently available for scaling when making scaling decisions. However, you may have set up some of your cloud VMs to only scale to certain instance types to reduce complexity and cost, improve discount utilization, or meet application demand. Use this setting to identify those instance types. Workload Optimization Manager recommends including as many instance types as possible to give analysis more options to scale workloads to their optimal sizes.

NOTE:
When a cloud provider offers a new instance type that is meant to replace an older type, the provider may offer the new instance type at a lower cost or the same cost. If capacity and cost are equal, Workload Optimization Manager cannot ensure that it chooses the newer instance type. To work around this issue, create a policy that excludes the older instance type.

Click **Edit** to set your preferences. In the new page that displays, choose from the following options:

- Excluded

  By default, this option is selected and all instance types are deselected. This means that all instance types are considered for scaling.

  Any instance types that you select are excluded from scaling.

  Configure an exclusion list if there are fewer instance types to exclude than to include. For example, if 400 instance types are currently available for scaling and you do not want VMs to scale to 20 instance types, configure an exclusion list and select the 20 instance types.

- Included

  Any instance types that you select are considered for scaling.

  Configure an inclusion list if there are fewer instance types to include than to exclude. For example, if 400 instance types are currently available for scaling and you want VMs to scale only to 50 instance types, configure an inclusion list and select the 50 instance types.

Expand an instance family to see individual instance types and the resources allocated to them. An example of an instance family is `a1` for AWS, `B-series` for Azure, or A2 for Google Cloud.

After you save your changes, the main page refreshes to reflect your selections.

Consistent resizing

Attribute	Default setting
Consistent Resizing	Off

Consistent resizing for user-defined automation policies

When you create a policy for a group of VMs and turn on Consistent Resizing, Workload Optimization Manager resizes all the group members to the same size, such that they all support the top utilization of each resource commodity in the group. For example, assume VM A shows top utilization of CPU, and VM B shows top utilization of memory. A resize action would result in all the VMs with CPU capacity to satisfy VM A, and memory capacity to satisfy VM B.

For an affected resize, the Actions List shows individual resize actions for each of the VMs in the group. If you automate resizes, Workload Optimization Manager executes each resize individually in a way that avoids disruption to your workloads.

Use this setting to enforce the same template across all VMs in a group when resizing VMs on the public cloud. In this way, Workload Optimization Manager can enforce a rule to size all the VMs in a group equally.

Consistent resizing for auto-discovered groups

In public cloud environments, Workload Optimization Manager discovers groups that should keep all their VMs on the same template, and then creates read-only policies for them to implement Consistent Resizing. The details of this discovery and the associated policy vary depending on the cloud provider.

- Azure

  Workload Optimization Manager discovers Azure availability sets and scale sets.

  - For availability sets, Workload Optimization Manager does *not* enable Consistent Resizing, but it can recommend scale actions for individual VMs in the availability set.

    When a scale action for a VM in an availability set fails due to insufficient resources in the compute cluster, the action remains pending. When you hover over the pending action, you will see a message indicating that action execution has been temporarily disabled due to a previous execution error in the availability set. Workload Optimization Manager assumes that all other VMs in the availability set will fail to scale due to the same resource issue, so it creates a temporary policy that disables action execution for the availability set. Specifically, this policy sets the action acceptance mode for scale actions to *Recommend* and stays in effect for 730 hours (one month). This means that for the duration of the policy, Workload Optimization Manager will continue to generate read-only, non-executable scale actions for individual VMs, so you can evaluate their resource requirements and plan accordingly. You can delete this policy if you need to re-enable action execution in the availability set.
  - For scale sets, Workload Optimization Manager enables Consistent Resizing across all the VMs in the group. You execute those actions directly in Azure. If you do not need to resize all the members of a given scale set to a consistent template, create another policy for that scope and turn off Consistent Resizing.

- AWS

  Workload Optimization Manager discovers Auto Scaling Groups and automatically enables Consistent Resizing across all the VMs in each group. You can choose to execute all the actions for such a group, either manually or automatically. In that case, Workload Optimization Manager executes the resizes one VM at a time. If you do not need to resize all the members of a given Auto Scaling Group to a consistent template, create another policy for that scope and turn off Consistent Resizing.

  If you select one or all actions for the group either manually or automatically, Workload Optimization Manager will change the Launch Configuration for the Auto Scaling Group but it will not terminate the EC2 instances.

The following examples are some use cases for employing Consistent Resizing for a group.

- If you have deployed load balancing for a group, then all the VMs in the group should experience similar utilization. In that case, if one VM needs to be resized, then it makes sense to resize them all consistently.
- A common HA configuration on the public cloud is to deploy mirror VMs to different availability zones, where the given application runs on only one of the VMs at a given time. The other VMs are on standby to recover in failover events. Without Consistent Resizing, Workload Optimization Manager would tend to size down or suspend the unused VMs, which would make them unready for the failover situation.

When working with Consistent Resizing, consider these points:

- You should not mix VMs in a group that has a Consistent Resizing policy, with other groups that enable Consistent Resizing. One VM can be a member of more than one group. If one VM (or more) in a group with Consistent Resizing is also in another group that has Consistent Resizing, then both groups enforce Consistent Resizing together, for all their group members.
- If one VM (or more) is in a group with Consistent Resizing turned on, and the same VMs are in a group with Consistent Resizing turned off, the affected VMs assume the ON setting. This is true if you created both groups, or if Workload Optimization Manager created one of the groups for Azure Scale Sets or AWS Auto Scaling Groups.
- For any group of VMs that enables Consistent Resizing, you should not mix the associated target technologies. For example, one group should not include VMs that are managed on both Azure and AWS platforms.
- Charts that show actions and risks assign the same risk statement to all the affected VMs. This can seem confusing. For example, assume one VM needs to resize to address vCPU risk, and 9 other VMs are set to resize consistently with it. Then charts will state that 10 VMs need to resize to address vCPU risks.

Ignore NVIDIA GPU compute capability constraints (AWS and Azure only)

For AWS *(on page 638)* or Azure *(on page 643)* VMs running supported NVIDIA GPU instance types, Workload Optimization Manager can generate scale actions that change the GPU compute capability of VMs. By default, these actions are only executable in the cloud provider console (web-based UI or CLI). When you review a pending scale action, you will see the following message in the Execution Prerequisites section that explains why action execution in Workload Optimization Manager is blocked.

```
CUDA applications running on <GPU_VM> may need to be re-configured or recompiled
to execute on the new accelerated instance. The compute capability on the current instance
is <x>, while the new compute capability will be <y> if the action is taken.
When the pre-requisites are verified, this action must be executed manually in the Cloud Provider Console
 (Web UI or CLI).
Alternatively, you can configure a policy to ignore GPU compute capability constraints.
```

You can turn on this setting to allow scale actions that change compute capability to execute in Workload Optimization Manager.

Attribute	Default setting
Ignore NVIDIA GPU Compute Capability Constraints	Off

Ignore NVMe constraints (AWS only)

For AWS, Workload Optimization Manager recognizes when a VM instance includes an NVMe driver. To respect NVMe constraints, it will not recommend scaling to an instance type that does not also include an NVMe driver. If you ignore NVMe constraints, then Workload Optimization Manager is free to scale the instance to a type that does not include an NVMe driver.

Attribute	Default setting
Ignore NVMe Constraints	Off

Instance store aware scaling (AWS only)

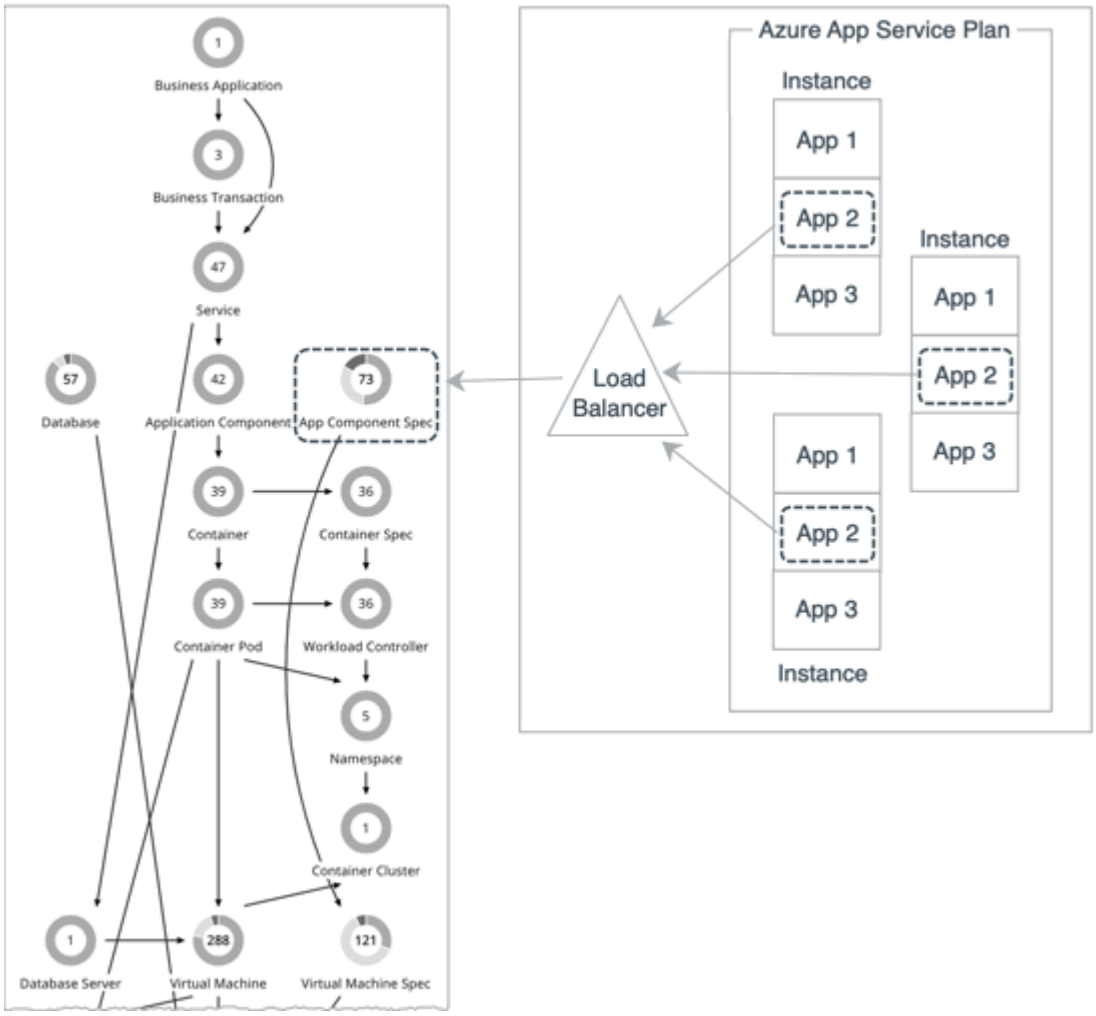
Attribute	Default setting
Instance Store Aware Scaling	Off

The template for your workload determines whether the workload can use an *instance store*, and it determines the instance store capacity. As Workload Optimization Manager calculates a resize or move action, it can recommend a new template that does not support instance stores, or that does not provide the same instance store capacity.

To ensure that resize actions respect the instance store requirements for your workloads, turn on **Instance Store Aware Scaling** for a given VM or for a group of VMs. When you turn this on for a given scope of VMs, then as it calculates move and resize actions, Workload Optimization Manager will only consider templates that support instance stores. In addition, it will not move a workload to a template that provides less instance store capacity.

App component spec

In Azure App Service *(on page 469)* deployments, an App Component Spec represents a set of app instances comprising a single web application. Workload Optimization Manager discovers App Component Specs when you add an Azure target with the necessary permissions.



NOTE:
For a list of permissions, see Azure Permissions *(on page 447)*.

Workload Optimization Manager also discovers the *plans* that provide resources to app instances. The supply chain shows these plans as Virtual Machine specs *(on page 673)* and links them with App Component Specs to establish their relationship.

Synopsis

Synopsis	
Provides:	App services to end users
Consumes:	Resources from App Service plans
Discovered through:	Azure targets

Monitored resources

Workload Optimization Manager monitors the following resources:

- Response time

  Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

- Virtual CPU (vCPU)

  Virtual CPU is the measurement of CPU that is in use.

Actions

None

Workload Optimization Manager does not recommend actions for App Component Specs, but it does recommend actions for the underlying Virtual Machine Specs. For details, see Virtual Machine Spec Actions *(on page 674)*.

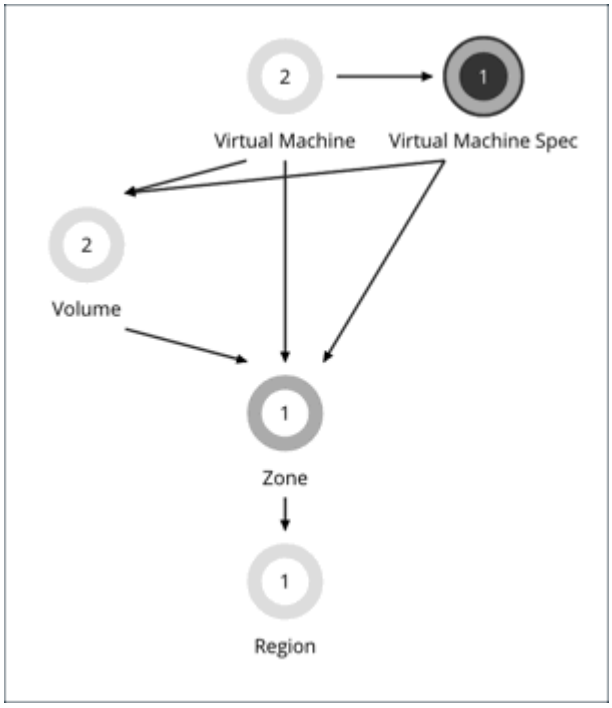
Virtual Machine spec

Workload Optimization Manager represents the following public cloud resources as Virtual Machine Spec entities in the supply chain:

- AWS Auto Scaling groups *(on page 670)*
- Azure Virtual Machine Scale Sets *(on page 672)*
- Azure App Service plans *(on page 673)*

Virtual machine spec – AWS Auto Scaling group

An Auto Scaling group is a collection of AWS EC2 instances that are treated as a logical group for the purposes of automatic scaling and management. Workload Optimization Manager discovers Auto Scaling groups through your AWS targets.



NOTE:
Currently, Auto Scaling groups in Amazon Elastic Kubernetes Service (EKS) are not discovered as Virtual Machine Specs and are not parkable.

Synopsis

Synopsis	
Provides:	Resources for applications to use
Consumes:	Resources from AWS regions
Discovered through:	AWS targets

Monitored resources

Workload Optimization Manager does not monitor resources for Auto Scaling groups, but allows you to park (stop or start) these groups on demand or according to a schedule.

Actions

Workload Optimization Manager supports the following actions:

Stop and start (also known as 'parking' actions)

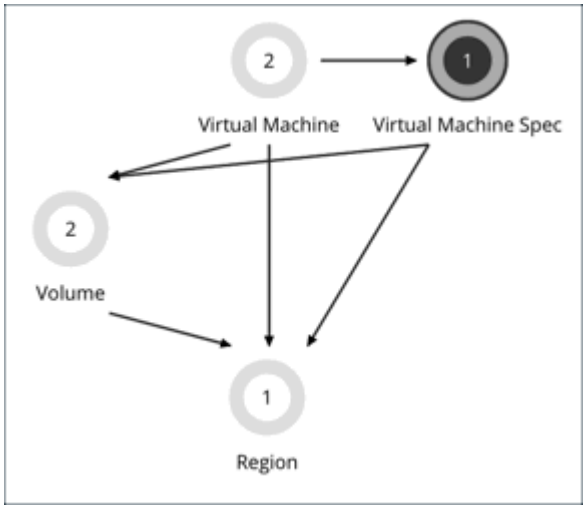
Stop an Auto Scaling group for a given period of time to reduce your cloud expenses, and then start it at a later time.

For details, see Parking: Stop or Start Cloud Resources *(on page 942)*.

NOTE:
You can create parking policies *(on page 1087)* to control the enforcement of stop and start actions. Currently, it is not possible to create automation policies or Workload Optimization Manager groups for Auto Scaling groups.

Virtual machine spec - Azure Virtual Machine Scale Set

A Virtual Machine Scale Set is a group of load-balanced VMs that you create and manage in Azure. The number of VM instances can automatically increase or decrease in response to demand or a defined schedule.



Currently, the following Virtual Machine Scale Sets are not discovered as Virtual Machine Specs and are not parkable.

- Virtual Machine Scale Sets in Azure Kubernetes Service (AKS)
- Virtual Machine Scale Sets that apply multiple autoscale profiles

Synopsis

Synopsis	
Provides:	Resources for applications to use
Consumes:	Resources from Azure regions
Discovered through:	Azure targets

Monitored resources

Workload Optimization Manager does not monitor resources for Virtual Machine Scale Sets, but allows you to park (stop or start) these groups on demand or according to a schedule.

Actions

Workload Optimization Manager supports the following actions:

Stop and start (also known as 'parking' actions)

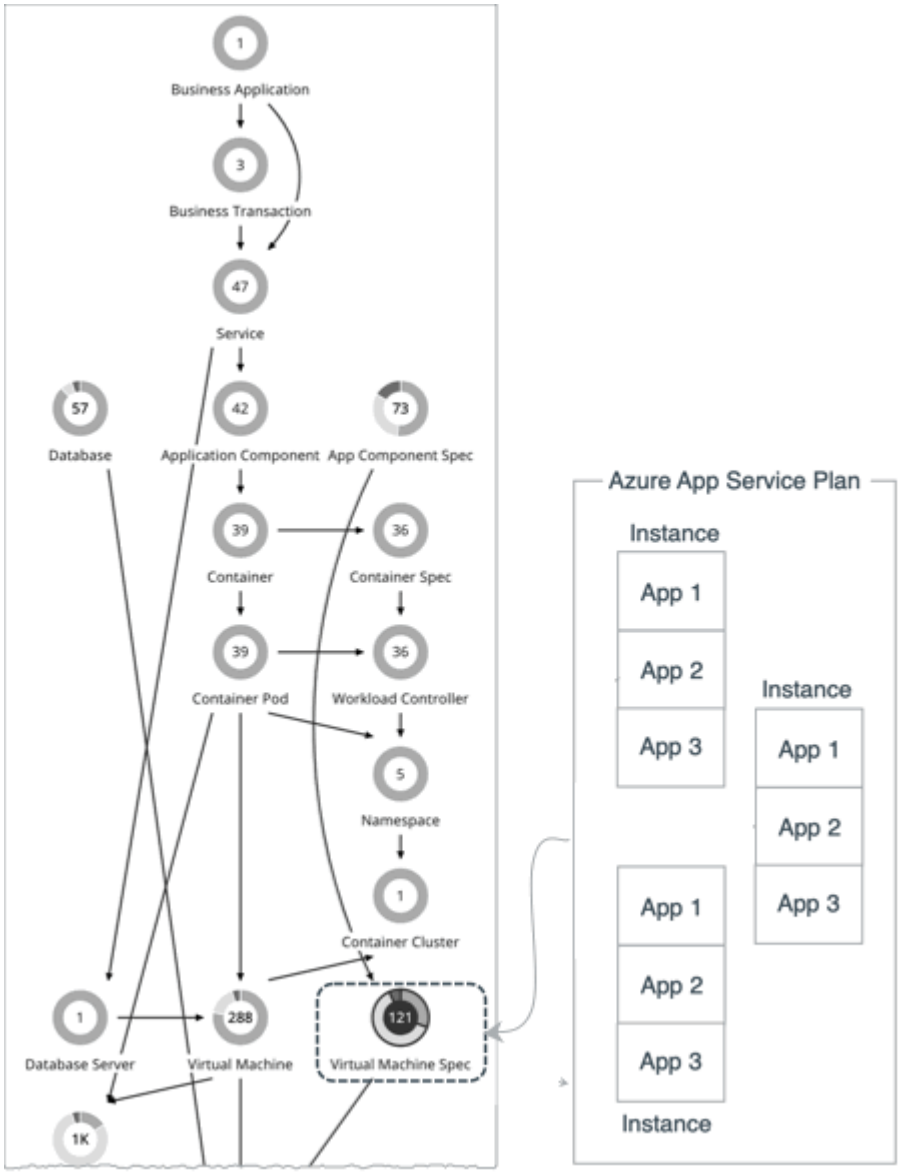
Stop a Virtual Machine Scale Set for a given period of time to reduce your cloud expenses, and then start it at a later time.

For details, see Parking: Stop or Start Cloud Resources *(on page 942)*.

> **NOTE:**
> You can create parking policies *(on page 1087)* to control the enforcement of stop and start actions. Currently, it is not possible to create automation policies or Workload Optimization Manager groups for Virtual Machine Scale Sets.

Virtual machine spec – Azure App Service plan

In Azure App Service *(on page 469)*, *plans* define CPU, memory, and storage resources that are available to VM instances to run apps. When you add an Azure target with the necessary permissions, Workload Optimization Manager discovers the plans associated with apps, and shows them as Virtual Machine Specs in the supply chain. Currently, Workload Optimization Manager discovers all plans, except App Service Environment v3 I4, I5, and I6.



NOTE:
For a list of permissions, see Azure Permissions *(on page 447)*.

Points to consider:

- Azure App Service offers several types of apps, including web apps, mobile apps, API apps, and logic apps. Workload Optimization Manager discovers the plans associated with these apps, but only recommends scale actions for plans associated with web apps. If a plan is no longer associated with any type of app, Workload Optimization Manager will recommend that you delete it.

- For web apps, Workload Optimization Manager also discovers the app instances that consume resources from a plan, and shows them as App Component Specs *(on page 669)* in the supply chain. The supply chain links App Component Specs with Virtual Machine Specs to establish their relationship.
- VM instances underlying a plan scale as a group. For this reason, Workload Optimization Manager represents these VM instances as a single Virtual Machine Spec entity and does *not* monitor them individually. The Entity Information chart for a Virtual Machine Spec shows the current number of VM instances, while resource charts (such as the Virtual CPU and Virtual Memory charts) show aggregated metrics for all VM instances.

Synopsis

Synopsis	
Provides:	Resources to apps (via App Component Specs)
Consumes:	Resources from Azure regions
Discovered through:	Azure targets

Monitored resources

Workload Optimization Manager monitors the following resources:

- Virtual memory (vMem)

  Virtual memory (vMem) is the measurement of memory that is in use.
- Virtual CPU (vCPU)

  Virtual CPU is the measurement of CPU that is in use.
- Storage amount

  Storage amount is the measurement of storage capacity that is in use.
- Number of replicas

  Number of replicas is the total number of VM instances underlying an App Service plan.

Actions

Workload Optimization Manager supports the following actions:

- Scale

  Scale Azure App Service plans to optimize app performance or reduce costs, while complying with business policies.
- Delete

  Delete empty Azure App Service plans as a cost-saving measure. A plan is considered empty if it is not hosting any running apps.

Scale actions

Workload Optimization Manager supports vertical scaling actions for provisioned App Service plans. These actions change the *size* of all VM instances underlying a plan (for example, from small to large, or large to medium). Horizontal scaling actions, which change the *number* of VM instances underlying a plan, are currently not supported.

Vertical scaling recommendations rely on a variety of factors, including:

- Resource utilization percentiles *(on page 676)*
- On-demand monthly costs *(on page 677)*
- VM instance count

  Workload Optimization Manager will only recommend vertical scaling actions on plans with six or less VM instances.

- Scaling eligibility
  - Eligible for scaling – Basic, Standard, Premium v2, Premium v3, Isolated, Isolated v2
  - Not eligible for scaling – Workflow Standard, Elastic Premium, Free, Shared, Dynamic/Serverless
- Azure-enforced constraints, including:
  - Region – Only recommend instance types in regions where they are available
  - Server rack – Only recommend instance types on server racks where they are available
  - Zone redundancy – If zone redundancy is enabled, only recommend instance types that support zone redundancy
  - Deployment slots – Only recommend instance types that support the currently configured number of deployment slots that can be added to apps
  - Hybrid connections – Only recommend scaling to instance types that support the currently configured number of hybrid connections for a plan

    NOTE:
    To see the number of deployment slots and hybrid connections configured for a plan, set the scope to the corresponding Virtual Machine Spec and then view the Entity Information chart.

- Scaling constraints that you set in Workload Optimization Manager policies *(on page 681)* for Virtual Machine Specs

  For example, you can set a constraint if you want App Service plans to *only scale to* or *avoid* certain instance types.

For scale actions, you can create policies to control the scale actions that Workload Optimization Manager recommends. In those policies, choose from the following options:

- Cloud scale all – execute all scaling actions
- Cloud scale for performance – only execute scaling actions that improve performance
- Cloud scale for savings – only execute scaling actions that reduce costs

The default action acceptance mode for these actions is *Manual*. When you examine the pending actions, only actions that satisfy the policy are allowed to execute. All other actions are read-only.

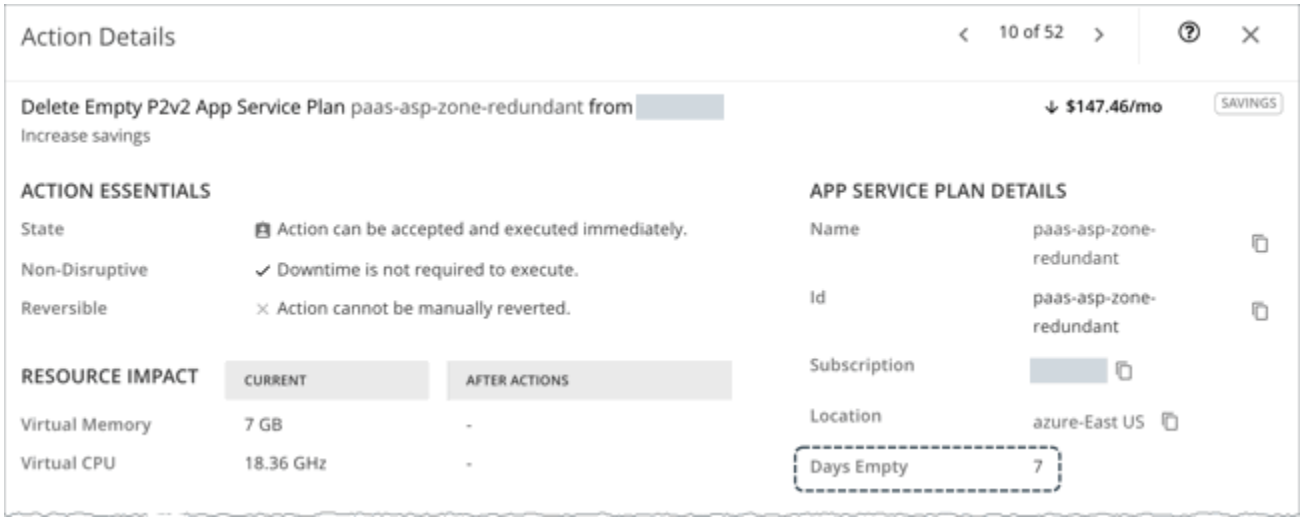
When policy conflicts arise, **Cloud scale all** overrides the other two scaling options in most cases. For more information, see Default and User-defined Automation Policies *(on page 1050)*.

Delete actions

When Workload Optimization Manager discovers an empty plan (i.e., a plan that is not hosting any running apps), it will immediately recommend that you delete the plan as a cost-saving measure. Workload Optimization Manager can recommend deleting provisioned App Service plans, as well as Elastic Premium and Workflow Standard plans.

If a currently empty plan is not deleted and is subsequently discovered as used, Workload Optimization Manager removes the delete action attached to it.

Delete actions include the 'Days Empty' information that indicates how long a plan has been empty.

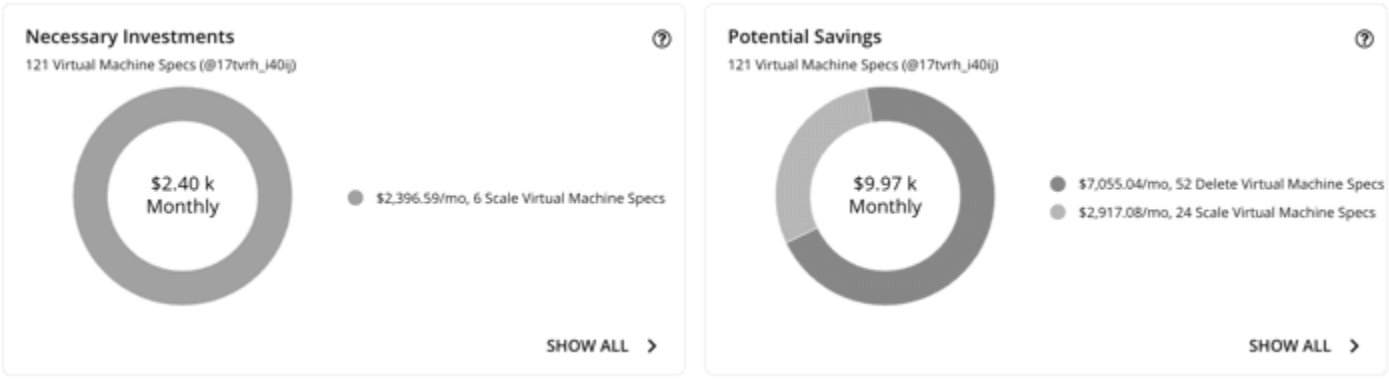


You can control the delete actions that Workload Optimization Manager recommends, based on the 'Days Empty' value that you set. For example, if you want Workload Optimization Manager to only generate delete actions for plans that have been empty for at least 5 days, perform these steps:

1. In the default policy for Virtual Machine Specs, *disable* delete actions.
2. Create a dynamic group of Virtual Machine Specs and set the 'Days Empty' filter to `Days Empty > = 5`.
3. Create a custom Virtual Machine Spec policy, set the scope to the group that you just created, and then *enable* delete actions in that policy.

Actions in charts

Use the Necessary Investments and Potential Savings charts to view pending Virtual Machine Spec actions. These charts show total monthly investments and savings, assuming you execute all the actions.



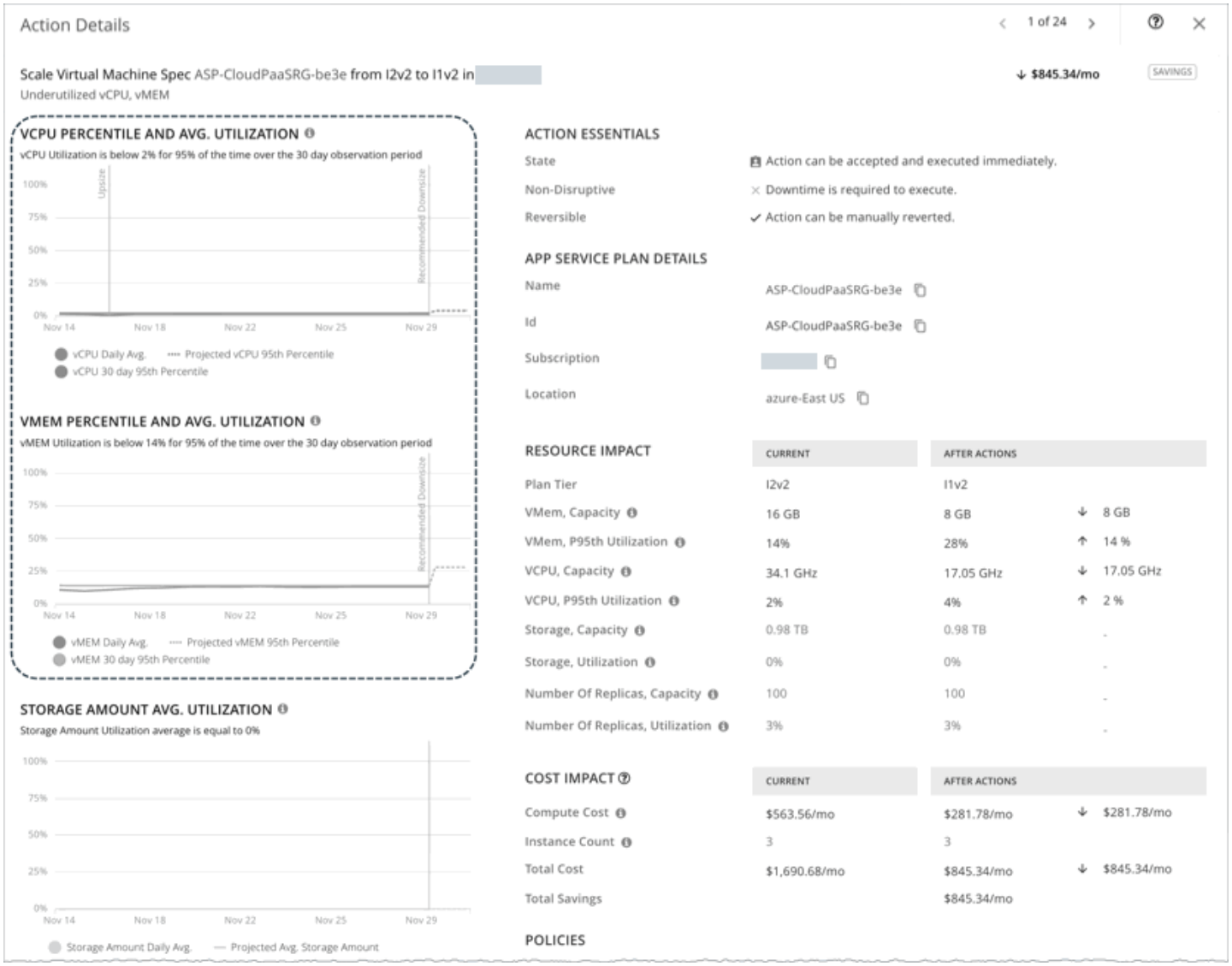
Click **Show All** for each chart to review and execute the actions.

The table shows the following:

- Actions that are pending for each Virtual Machine Spec
- Savings or investments for each Virtual Machine Spec

Utilization charts for scale actions

Workload Optimization Manager uses percentile calculations to measure resource utilization, and drive scaling actions that improve overall utilization and reduce costs. When you examine the details for a pending scaling action on an App Service plan, you will see charts that highlight resource *utilization percentiles* for a given observation period, and the projected percentiles after you execute the action.



The charts also plot *daily average utilization* for your reference. If you have previously executed scaling actions on the App Service plan, you can see the resulting improvements in daily average utilization. Put together, these charts allow you to easily recognize utilization trends that drive Workload Optimization Manager's scaling recommendations.

> **NOTE:**
> You can set scaling constraints in Virtual Machine Spec policies to refine the percentile calculations. For details, see Scaling Sensitivity *(on page 681)*.

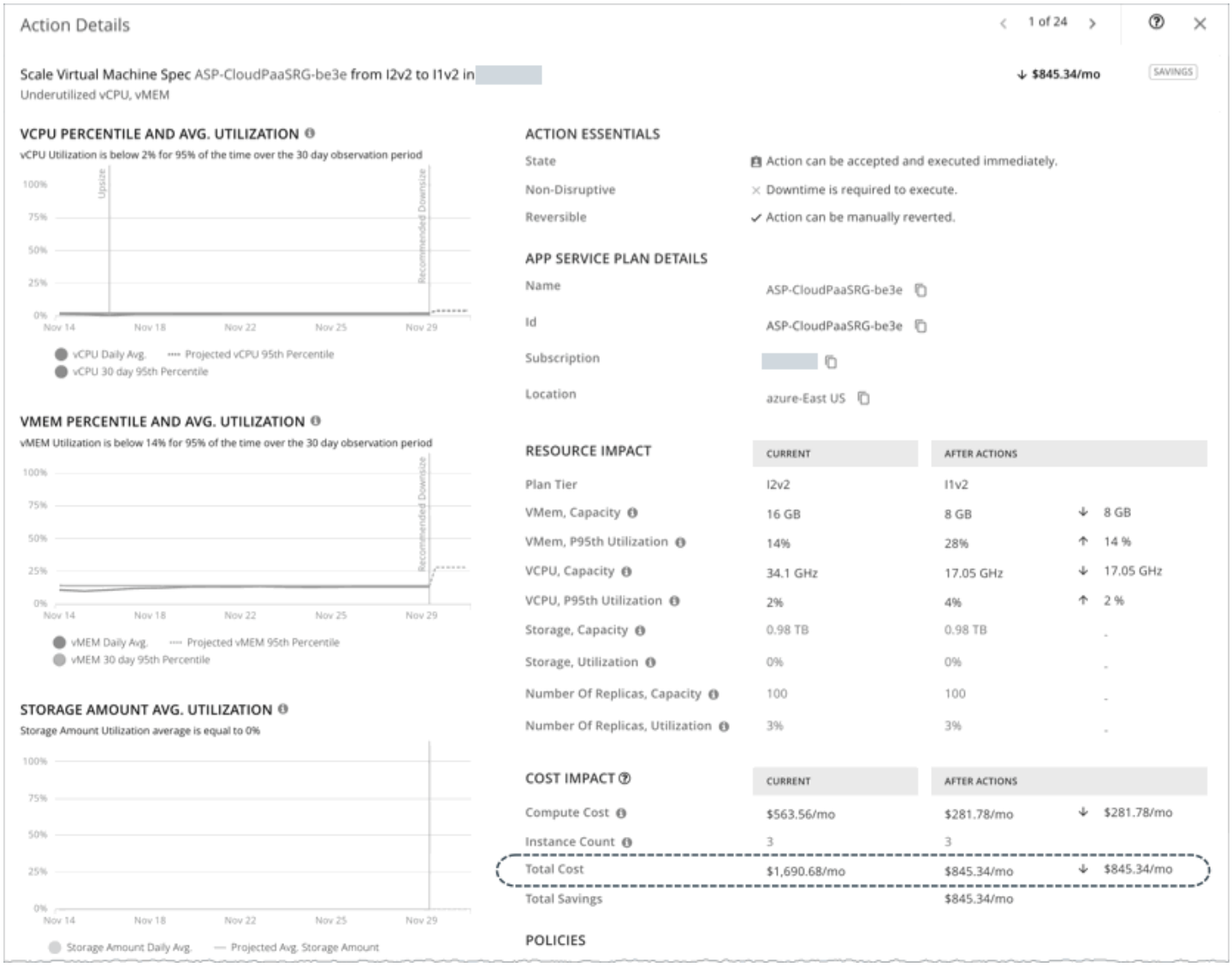
Disruptiveness and reversibility of scale actions

Workload Optimization Manager always recommends scaling to a different instance type, so all scaling actions are disruptive and require downtime. You can reverse an action by scaling an App Service plan back to its original instance type.

Estimated On-demand monthly costs for Azure App Service plans

Workload Optimization Manager considers a variety of factors when calculating estimated on-demand monthly costs for Azure App Service plans.

NOTE:
Azure App Service plans appear as Virtual Machine Spec entities in the supply chain.



The calculation for estimated on-demand monthly cost can be expressed as follows:

```
(On-demand Compute Rate * 730) * Number of Instances = Estimated On-demand Monthly Cost
```

Where:

- **On-demand Compute Rate** is the **hourly** cost for an App Service plan's instance type.
  You can obtain on-demand rates via App Service Pricing.
- **730** represents the number of hours per month that Workload Optimization Manager uses to calculate monthly costs.
- **Number of Instances** is the total number of VM instances underlying the App Service plan.

The preceding listed items impact cost calculations and the scaling decisions that Workload Optimization Manager makes. These decisions also rely on other factors, such as resource utilization percentiles and scaling constraints set in policies.

Example

Assume the following data for a pending action to scale an Azure Service plan from the I2V2 to the I1V1 instance type.

Detail	Current values	Values after action execution
On-demand Compute Rate	$0.772/hr	$0.386/hr
Number of Instances	3	3

Workload Optimization Manager calculates the following:

- *Current* estimated on-demand monthly cost:

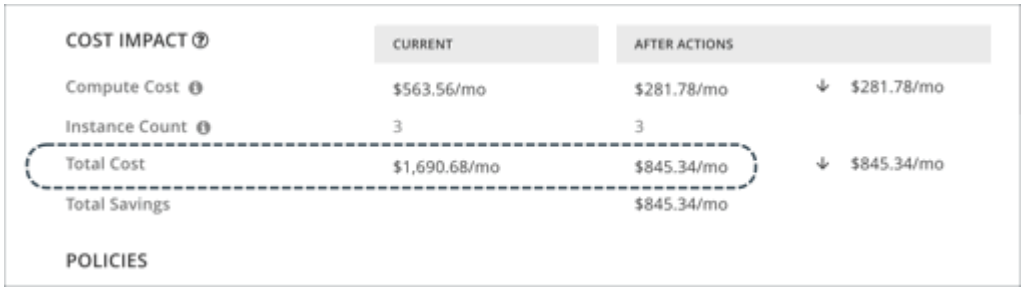
```
($0.772 * 730) * 3 = $1690.68/Mo.
```

- Estimated on-demand monthly cost *after* executing the action:

```
($0.386 * 730) * 3 = $845.34/Mo.
```

NOTE:
Workload Optimization Manager rounds the calculated values that it displays in the user interface.

The estimated on-demand monthly cost is projected to decrease from $1690.68/month to $845.34/month, as shown in the Details section of the pending action.



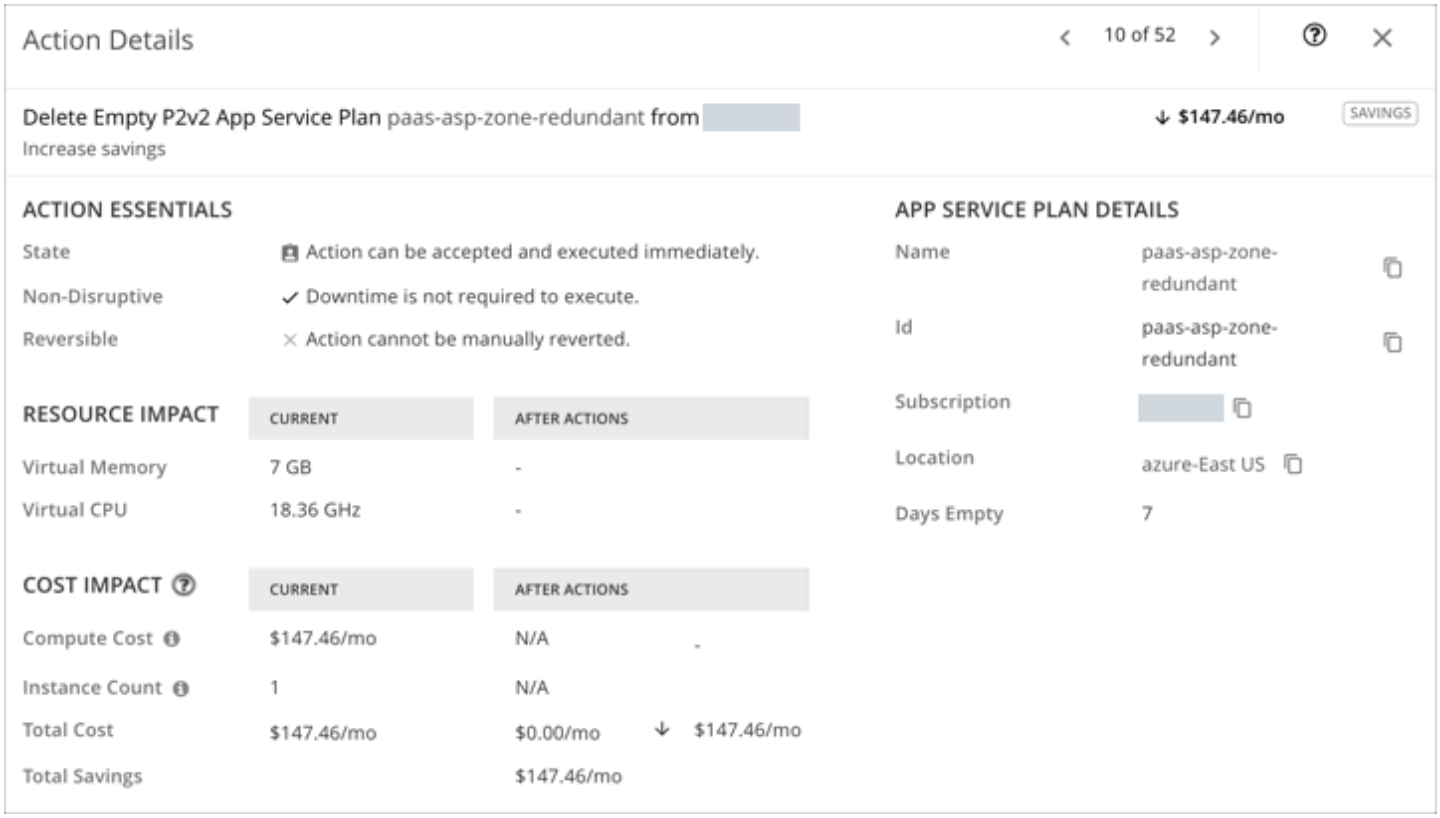
Workload Optimization Manager treats the action as a cost-saving measure, and shows total savings of $845.34/month.



Estimated on-demand monthly savings for empty Azure App Service plans

Workload Optimization Manager considers an empty App Service plan's on-demand compute rate and VM instance count when calculating the estimated on-demand monthly savings that you would realize when you delete the plan. A plan is considered empty if it is not hosting any running apps.

NOTE:
Azure App Service plans appear as Virtual Machine Spec entities in the supply chain.



Savings calculation

The calculation for estimated on-demand monthly savings can be expressed as follows:

```
(On-demand Compute Rate * 730) * Number of Instances = Estimated On-demand Monthly Savings
```

Where:

- **On-demand Compute Rate** is the **hourly** cost for an App Service plan's instance type.
  You can obtain on-demand rates via App Service Pricing.
- **730** represents the number of hours per month that Workload Optimization Manager uses to calculate monthly savings.
- **Number of Instances** is the total number of VM instances underlying the App Service plan.

Example

Assume the following data for a pending action to delete an empty Azure Service plan on the P2V2 instance type.

Detail	Current values
On-demand Compute Rate	$0.202/hr
Number of Instances	1

Workload Optimization Manager calculates savings as follows:

```
($0.202 * 730) * 1 = $147.46/Mo.
```

NOTE:
Workload Optimization Manager rounds the calculated values that it displays in the user interface.

Workload Optimization Manager shows total savings of $147.46/month.



Virtual machine spec policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

For details about Virtual Machine Spec actions, see Virtual Machine Spec Actions *(on page 674)*.

Action	Default mode
Cloud Scale All	Manual
Cloud Scale for Perfor-mance	Manual
Cloud Scale for Savings	Manual
Delete Virtual Machine Spec	Manual

You can use Action Scripts *(on page 1070)* and third-party orchestrators (such as ServiceNow) for action orchestration.

Scaling sensitivity

Workload Optimization Manager uses a percentile of utilization over the specified observation period. This gives sustained utilization and ignores short-lived bursts.

Workload Optimization Manager uses these settings to calculate utilization percentiles for vCPU and vMem. It then recommends actions to improve utilization based on the observed values for a given time period.

- Aggressiveness

Attribute	Default value
Aggressiveness	95th Percentile

  Workload Optimization Manager uses these settings to calculate utilization percentiles for vCPU and vMem.

  When evaluating performance, Workload Optimization Manager considers resource utilization as a percentage of capacity. The utilization drives actions to scale the available capacity either up or down. To measure utilization, the analysis considers a given utilization percentile. For example, assume a 95th percentile. The percentile utilization is the highest value that 95% of the observed samples fall below. Compare that to average utilization, which is the average of *all* the observed samples.

Using a percentile, Workload Optimization Manager can recommend more relevant actions. This is important in the cloud, so that analysis can better exploit the elasticity of the cloud. For scheduled policies, the more relevant actions will tend to remain viable when their execution is put off to a later time.

For example, consider decisions to reduce vCPU capacity. Without using a percentile, Workload Optimization Manager never scales below the recognized peak utilization. Assume utilization peaked at 100% just once. Without the benefit of a percentile, Workload Optimization Manager will not reduce resources for the Virtual Machine Spec.

With **Aggressiveness**, instead of using the single highest utilization value, Workload Optimization Manager uses the percentile you set. For the previous example, assume a single burst to 100%, but for 95% of the samples, utilization never exceeded 50%. If you set **Aggressiveness** to 95th Percentile, then Workload Optimization Manager can see this as an opportunity to reduce resource allocation.

In summary, a percentile evaluates the sustained resource utilization, and ignores bursts that occurred for a small portion of the samples. You can think of this as aggressiveness of resizing, as follows:

- 99th or 100th Percentile – More performance. Recommended for critical Virtual Machine Specs that need maximum guaranteed performance at all times, or those that need to tolerate sudden and previously unseen spikes in utilization, even though sustained utilization is low.
- 95th Percentile (Default) – The recommended setting to achieve maximum performance and savings. This assures performance while avoiding reactive peak sizing due to transient spikes, thus allowing you to take advantage of the elastic ability of the cloud.
- 90th Percentile – More efficiency. Recommended for Virtual Machine Specs that can stand higher resource utilization.

By default, Workload Optimization Manager uses samples from the last 14 days. Use the **Max Observation Period** setting to adjust the number of days.

- Max Observation Period

Attribute	Default value
Max Observation Period	Last 14 Days

To refine the calculation of resource utilization percentiles, you can set the sample time to consider. Workload Optimization Manager uses historical data from up to the number of days that you specify as a sample period. If the volume has fewer days' data then it uses all of the stored historical data.

Choose from the following settings:

- Less Elastic – Last 30 Days
- Recommended – Last 14 Days
- More Elastic – Last 3 or 7 Days

- Min Observation Period

Attribute	Default value
Min Observation Period	None

This setting ensures historical data for a minimum number of days before Workload Optimization Manager will generate an action based on the percentile set in **Aggressiveness**. This ensures a minimum set of data points before it generates the action.

Especially for scheduled actions, it is important that resize calculations use enough historical data to generate actions that will remain viable even during a scheduled maintenance window. A maintenance window is usually set for "down" time, when utilization is low. If analysis uses enough historical data for an action, then the action is more likely to remain viable during the maintenance window.

Choose from the following settings:

- More Elastic – None
- Less Elastic – 1, 3, or 7 Days

Cloud instance types

Attribute	Default value
Cloud Instance Types	None

By default, Workload Optimization Manager considers all instance types currently available for scaling when making scaling decisions for Virtual Machine Specs. However, you may have set up your Virtual Machine Specs to only scale to certain instance types to reduce complexity and cost, improve discount utilization, or meet application demand. Use this setting to identify those instance types.

Click **Edit** to set your preferences. In the new page that displays, choose from the following options:

- Excluded

  By default, this option is selected and all instance types are deselected. This means that all instance types are considered for scaling.

  Any instance types that you select are excluded from scaling.

  Configure an exclusion list if there are fewer instance types to exclude than to include. For example, if 400 instance types are currently available for scaling and you do not want VM Specs to scale to 20 instance types, configure an exclusion list and select the 20 instance types.

- Included

  Any instance types that you select are considered for scaling.

  Configure an inclusion list if there are fewer instance types to include than to exclude. For example, if 400 instance types are currently available for scaling and you want VM Specs to scale only to 50 instance types, configure an inclusion list and select the 50 instance types.

Expand an instance family to see individual instance types and the resources allocated to them.

After you save your changes, the main page refreshes to reflect your selections.

Scaling target utilization

The utilization that you set here specifies the percentage of the existing capacity that Workload Optimization Manager will consider to be 100% of capacity.

Attribute	Default value
VCPU	70
VMEM	90
Storage	90

These advanced settings determine how much you would like a scope of workloads to utilize their resources. These are fixed settings that override the way Workload Optimization Manager calculates the optimal utilization of resources. You should only change these settings after consulting with Technical Support.

While these settings offer a way to modify how Workload Optimization Manager recommends actions, in most cases you should never need to use them. If you want to control how Workload Optimization Manager recommends actions to resize workloads, you can set the aggressiveness per the percentile of utilization, and set the length of the sample period for more or less elasticity on the cloud.

Database server (cloud)

Workload Optimization Manager represents the following public cloud resources as Database Server entities in the supply chain:

- AWS RDS *(on page 684)*
- Azure SQL Managed Instances *(on page 691)*
- Azure Cosmos DB Accounts *(on page 702)*

NOTE:
Azure SQL databases, dedicated SQL pools, and Cosmos DB databases appear as *Database* entities in the supply chain. For details, see Database (Cloud) *(on page 703)*.

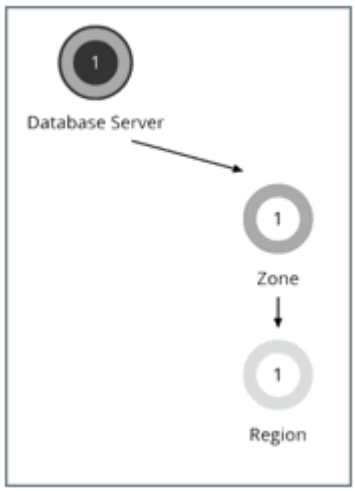
Create groups in Workload Optimization Manager so you can manage Database Servers with ease. To create groups, go to **Settings > Groups**, and then select **New Group > Database Server**. In the page that displays, choose **Dynamic** (recommended) and then set the filter as follows:

- `Cloud Service Name = Amazon RDS` – use to create a group of AWS RDS Database Servers
- `Cloud Service Name = Azure SQL Managed Instance` – use to create a group of Azure SQL Managed Instances
- `Cloud Provider = Azure` – use to create a group of Azure Cosmos DB accounts

Database server – AWS RDS

In AWS public cloud environments, a Database Server is a relational database that you have configured using AWS Relational Database Service (RDS). Workload Optimization Manager discovers RDS instances through your AWS targets, and then generates scaling actions as needed.

Synopsis



Synopsis	
Provides:	Database services to cloud applications and end users
Consumes:	Compute and storage resources in the availability zone
Discovered through:	AWS targets

Monitored resources

Workload Optimization Manager monitors the following resources:

- Virtual memory (vMem)

  Virtual memory (vMem) is the measurement of memory that is in use.
- Virtual CPU (vCPU)

  Virtual CPU is the measurement of CPU that is in use.
- Storage amount

  Storage amount is the measurement of storage capacity that is in use.
- Storage access (IOPS)

  Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
- DB cache hit rate

  DB cache hit rate is the measurement of Database Server accesses that result in cache hits, measured as a percentage of hits versus total attempts. A high cache hit rate indicates efficiency.
- Connection

  Connection is the measurement of database connections utilized by applications.

Actions

Workload Optimization Manager supports the following actions:

- Stop and start (also known as 'parking' actions)

  Stop a Database Server for a given period of time to reduce your cloud expenses, and then start it at a later time.
- Scale

  Scale compute and storage resources to optimize performance and costs.

Stop and start actions for database servers

Workload Optimization Manager supports 'parking' actions for cloud resources. These user-initiated actions *stop* your cloud resources for a given period of time to help you reduce your cloud expenses, and then *start* these resources later when you need them. You can enforce parking actions on demand, or automatically through parking schedules and policies.

For details, see Parking: Stop or Start Cloud Resources *(on page 942)*.

Scale actions for database servers

To recommend accurate scaling actions, Workload Optimization Manager analyzes resource utilization percentiles and collects relevant metrics (such as connections utilization) from AWS. It also takes into consideration constraints defined in policies *(on page 694)*.

Consider the following scenarios and actions:

- To address vCPU congestion, Workload Optimization Manager can recommend scaling a Database Server to the instance type that can adequately meet demand at the lowest possible cost. If vCPU is underutilized, it can recommend scaling to a smaller instance type.
- To address IOPS congestion, Workload Optimization Manager can recommend increasing provisioned IOPS or scaling the Database Server to a different storage type. For gp2 storage, it can recommend increasing disk size to increase provisioned IOPS. After executing these actions, Workload Optimization Manager will not recommend new actions for the next six hours, in compliance with AWS's "cooldown" period for EBS storage.
- Workload Optimization Manager analyzes DB cache hit rate before making vMem scaling decisions. To perform its analysis, it collects cache hit rate metrics for Database Servers with Performance Insights enabled.

For Database Servers with cache hit rate metrics, Workload Optimization Manager considers at least 90% cache hit rate to be optimal. This percentage value is not configurable.

- A cache hit rate value equal to or greater than 90% indicates efficiency. For this reason, Workload Optimization Manager will not recommend an action even if vMem utilization is high. If vMem utilization is low, it will recommend scaling to a smaller instance type.
- When the cache hit rate is less than 90%, Workload Optimization Manager will also not recommend an action, provided that vMem utilization remains low. If vMem utilization is high, then it will recommend scaling to a larger instance type.

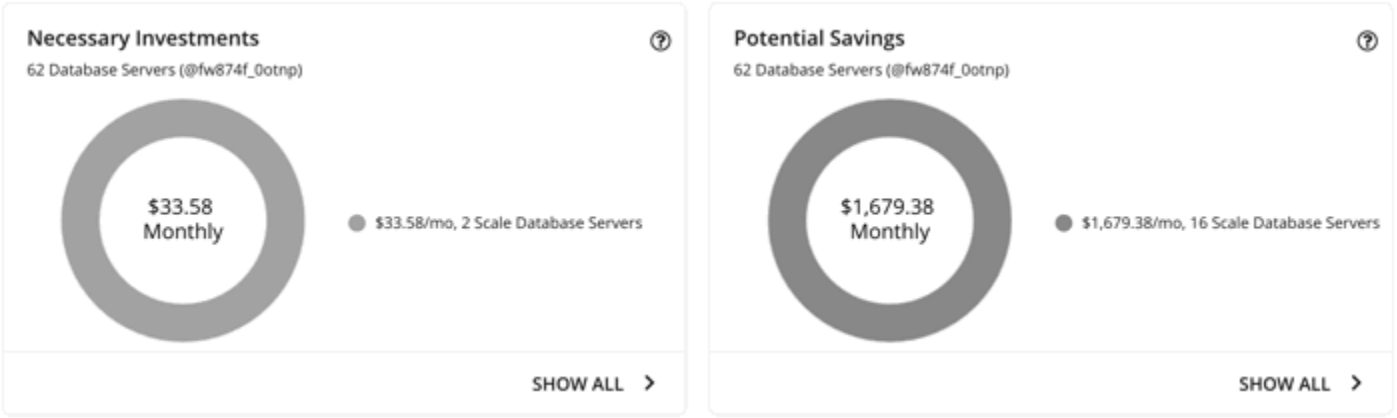
Notes on Performance Insights and cache hit rate metrics:

- Performance Insights is enabled by default on a majority of AWS Database Servers. In the Workload Optimization Manager user interface, you can use Search and then apply the Performance Insights filter to see which Database Servers have Performance Insights enabled.
- If Performance Insights is disabled or is not supported for your AWS Database Server engines or regions, Workload Optimization Manager will not have cache hit rate metrics to analyze and will therefore not generate actions in direct response to vMem utilization. For a list of supported engines and regions, see this AWS page.
- An action to scale to a different instance type in response to vCPU utilization might also include vMem changes, but vMem utilization alone (without cache hit rate metrics) will not drive actions.

Workload Optimization Manager also considers Connections utilization and capacity when making scaling decisions. It collects utilization metrics from CloudWatch and calculates capacity based on the maximum number of simultaneous connections configured for the Database Server. The maximum number varies by Database Server engine type and memory allocation, and is set in the parameter group associated with a Database Server. Workload Optimization Manager currently supports Database Servers associated with parameter groups that use default values. For example, consider a MySQL Database Server that is on a `db.t3.large` instance type with 8 GB (`8589934592 bytes`) of memory, and is associated with a parameter group that uses the default value `{DBInstanceClassMemory/12582880}`. In this case, Workload Optimization Manager calculates capacity as 682 connections (or `{8589934592/12582880}`). When Workload Optimization Manager generates an action due to vMem underutilization and sees that Connections utilization is only 15% of capacity (or roughly 100 connections), it picks a smaller instance type that is adequate for both the vMem and Connections requirements of the Database Server. For example, it could pick `db.t2.small`, which provides 2 GB of memory and a maximum of 170 connections.

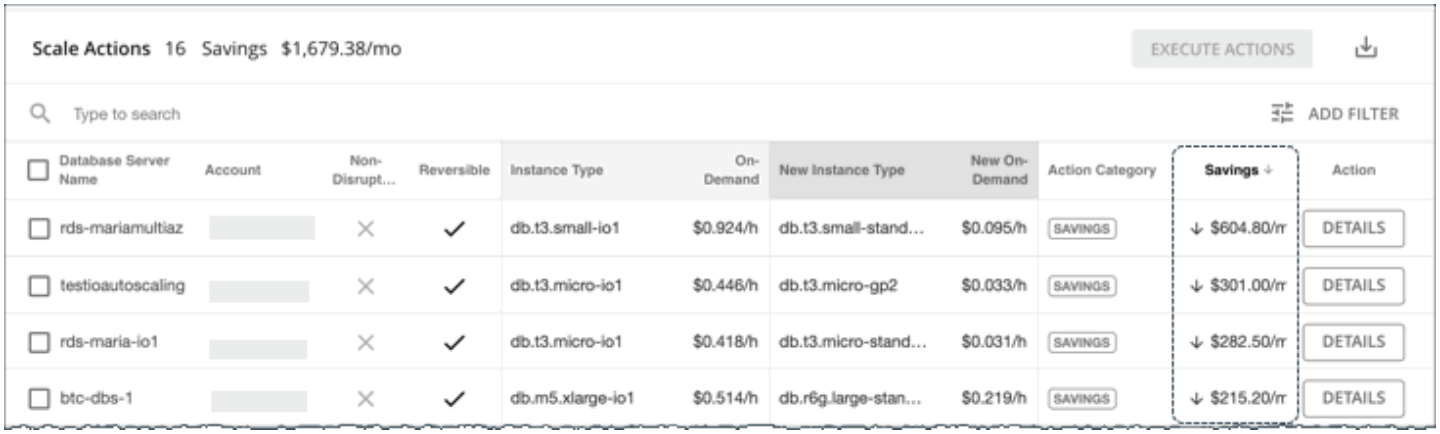
Scale actions in charts

Use the Necessary Investments and Potential Savings charts to view pending Database Server actions. These charts show total monthly investments and savings, assuming you execute all the actions.



Click **Show All** for each chart to review and execute the actions.

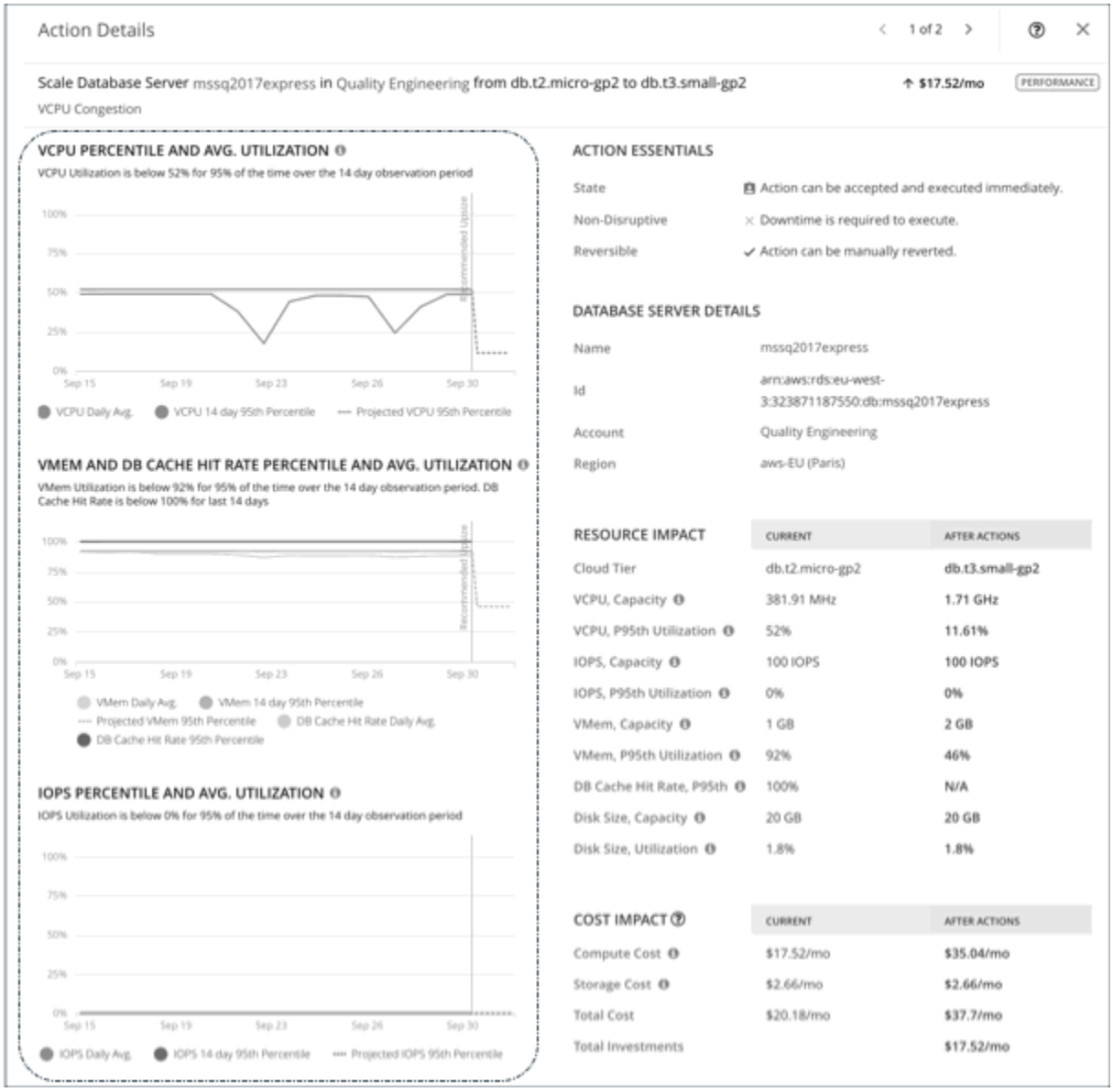
The table lists all the actions that are pending for Database Servers, and the savings or investments for each action.



For details on how Workload Optimization Manager calculates savings or investments, see Estimated On-demand Costs for Cloud Database Servers *(on page 697)*.

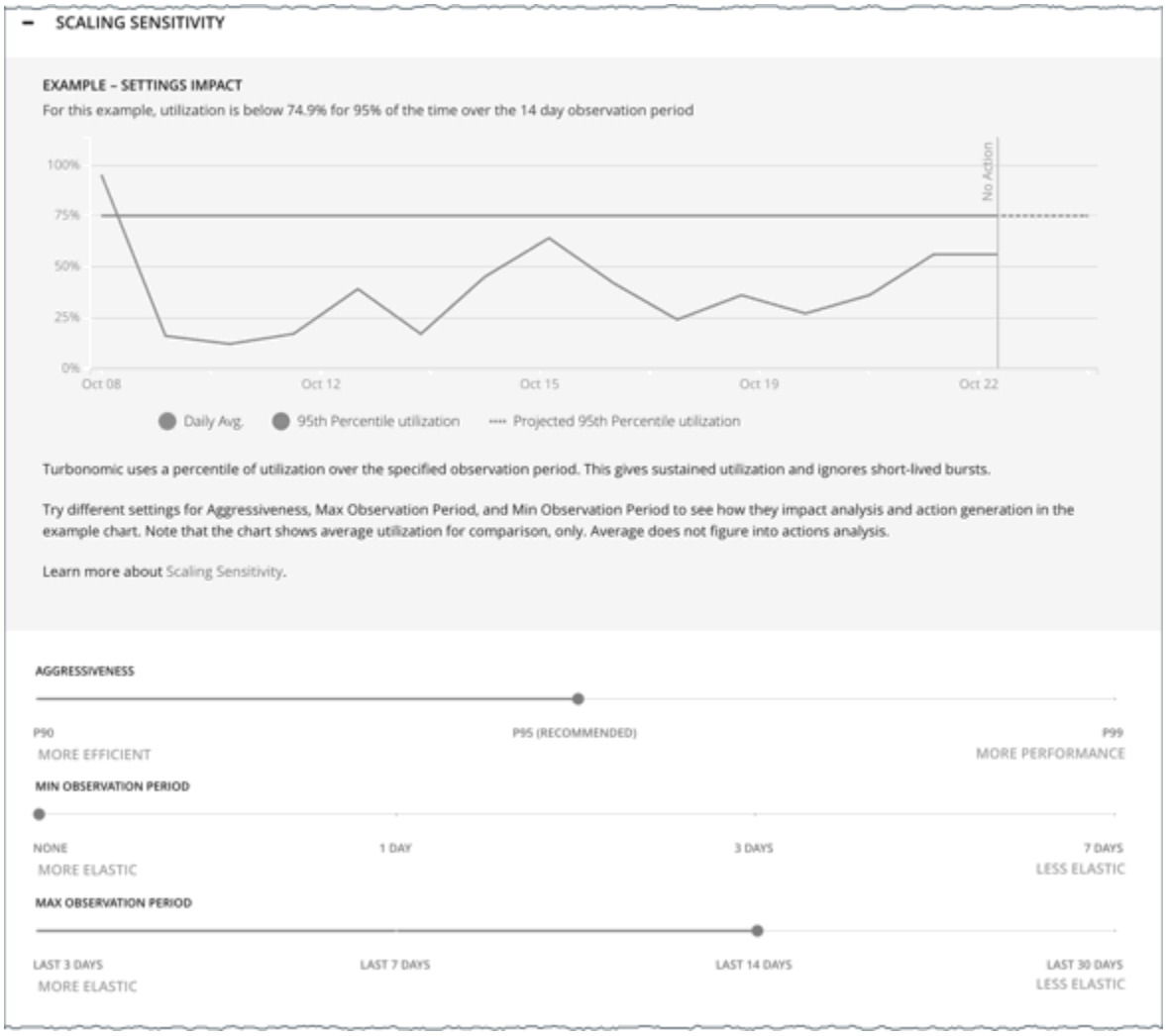
Utilization charts for scale actions

Workload Optimization Manager uses percentile calculations to measure resource utilization more accurately, and drive scaling actions that improve overall utilization and reduce costs. When you examine the details for a pending scale action on a Database Server, you will see charts that highlight *utilization percentiles* for a given observation period, and the projected percentiles after you execute the action.



The charts also plot *daily average utilization* for your reference. If you have previously executed scaling actions on the Database Server, you can see the resulting improvements in daily average utilization. Put together, these charts allow you to easily recognize utilization trends that drive Workload Optimization Manager's scaling recommendations.

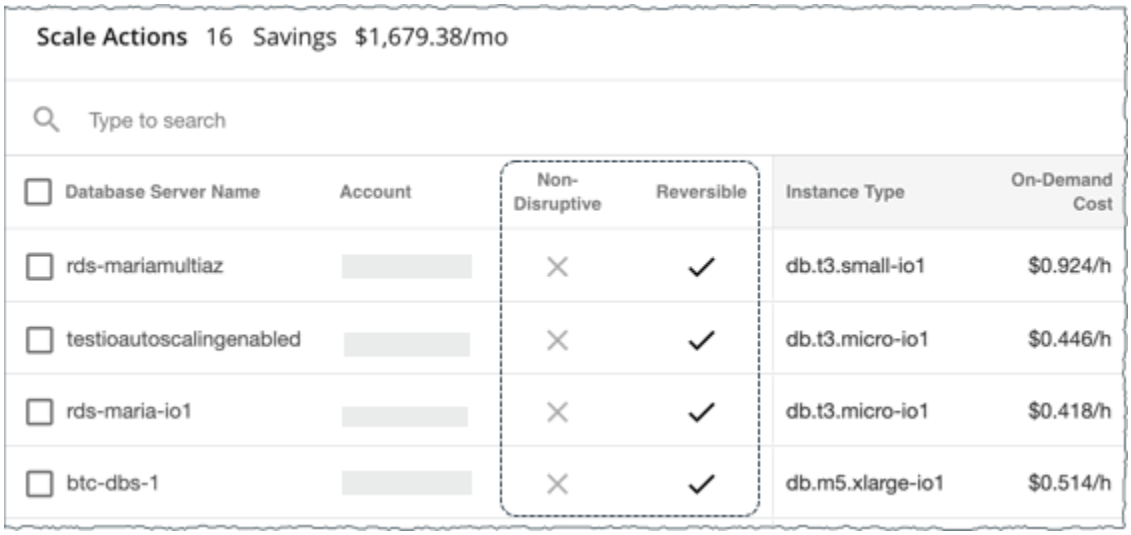
You can set scaling constraints in Database Server policies to refine the percentile calculations.



For details, see Scaling Sensitivity *(on page 695)*.

Non-disruptive and reversible scaling actions

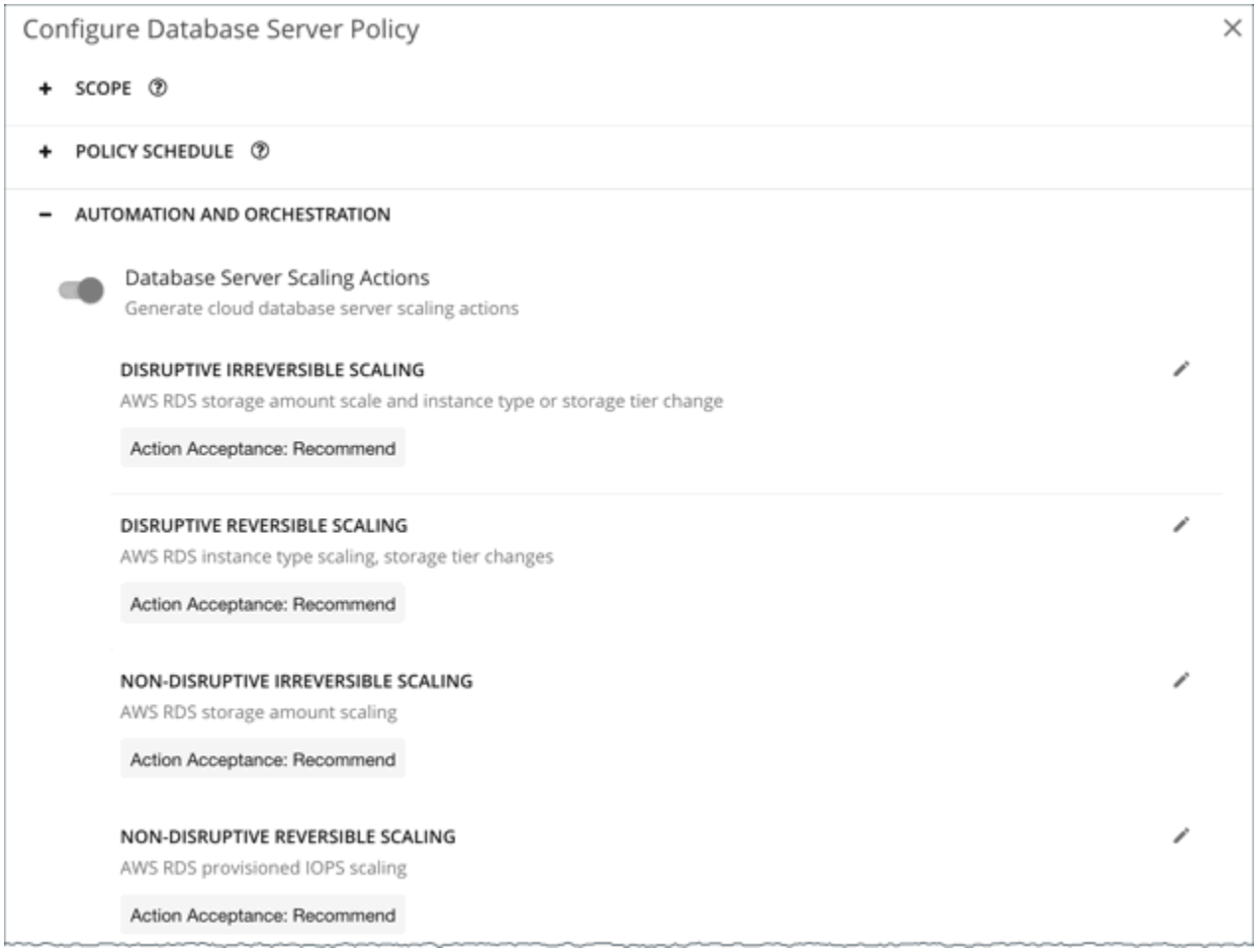
Workload Optimization Manager indicates whether a pending action is non-disruptive or reversible to help you decide how to handle the action.



The following table describes the disruptiveness and reversibility of the actions that Workload Optimization Manager recommends:

Action	Disruptive	Reversible
Scaling to a different instance type	Yes	Yes
Scaling up storage amount	No	No
Scaling up storage access (provisioned IOPS)	No	Yes
Scaling to a different storage type + Scaling up storage amount	Yes	No
Scaling to a different storage type + Scaling up storage access (provisioned IOPS)	Yes	Yes
Scaling to a different storage type + Scaling up storage amount + Scaling up storage access (provisioned IOPS)	Yes	No

You can set action acceptance modes in policies to specify the degree of automation for these actions.



Unavailable instance types

A scale action could fail if the target instance type is unavailable in the availability zone for some reason. Your AWS environment might show the instance type as available, but when the scaling action executes, the following error displays in AWS:

```
Cannot modify the instance class because there are no instances of the requested class available in the current instance's availability zone. Please try your request again at a later time.
```

> **NOTE:**
> For details about this error, see this AWS page.

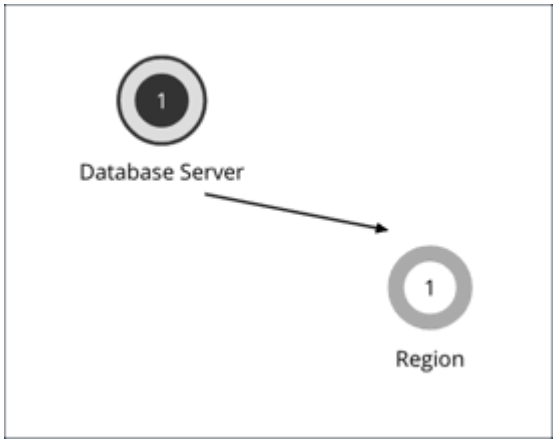
When this error occurs, Workload Optimization Manager modifies the default Database Server policy to exclude the instance type from its scaling list. When the Database Server is available again, Workload Optimization Manager adds it back to the scaling list. For details about this list, see Cloud Instance Types *(on page 696)*.

Database server – Azure SQL Managed Instance

Azure SQL Managed Instance is a scalable cloud database service with an underlying infrastructure managed by Microsoft. Workload Optimization Manager discovers SQL Managed Instances through your Azure targets, and represents them as database server entities in the supply chain.

NOTE:
Currently, SQL Managed Instances in Azure Government subscriptions are not supported.

Synopsis



Synopsis	
Provides:	Transactions to end users
Consumes:	Database server resources
Discovered through:	Azure targets

Monitored resources

NOTE:
Configure Azure database watcher *(on page 475)* to enable the discovery of metrics for vMem, maximum concurrent sessions, maximum concurrent workers, and DB cache hit rate.

Workload Optimization Manager monitors the following resources:

- Virtual CPU (vCPU)

  Virtual CPU is the measurement of CPU that is in use.
- Virtual memory (vMem)

  Virtual memory (vMem) is the measurement of memory that is in use.
- Storage amount

  Storage amount is the measurement of storage capacity that is in use.
- Maximum concurrent sessions

  Maximum concurrent sessions is the maximum number of database connections at a time.
- Maximum concurrent workers

  Maximum concurrent workers is the maximum number of database processes that can handle queries at a time.
- DB cache hit rate

  DB cache hit rate is the measurement of Database Server accesses that result in cache hits, measured as a percentage of hits versus total attempts. A high cache hit rate indicates efficiency.

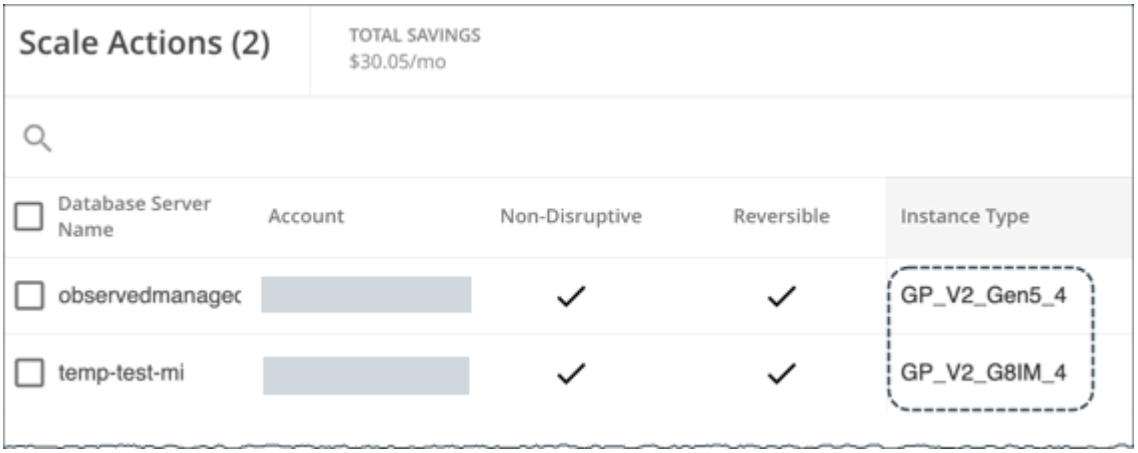
Actions

Workload Optimization Manager scaling actions aim to increase resource utilization and reduce costs while complying with business policies. Workload Optimization Manager scaling actions consider vCPU, vMem, storage amount, and IOPS utilization.

Points to consider:

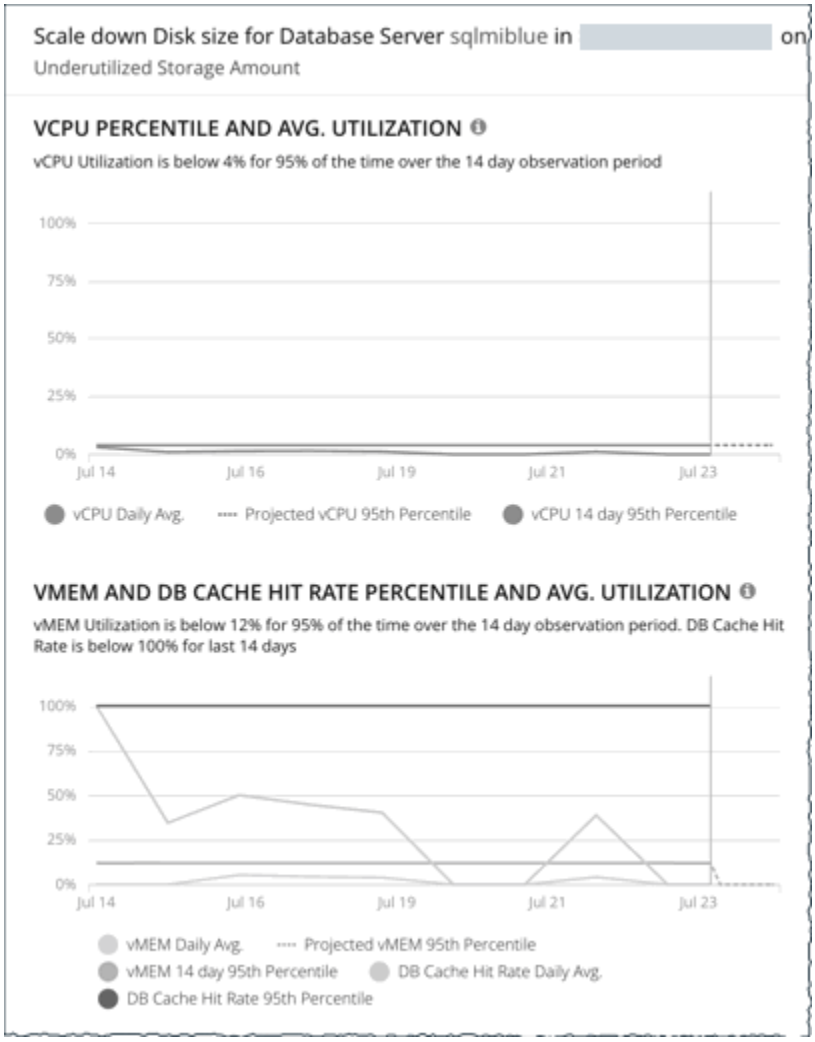
- If a SQL Managed Instance is not monitored by Azure database watcher, Workload Optimization Manager does not track vMem, maximum concurrent sessions, maximum concurrent workers, and DB cache hit rate. Consequently, Workload Optimization Manager cannot recommend vMem scaling actions, such as reducing vMem capacity.

  To configure Azure database watcher, see this topic *(on page 475)*.
- Currently, Workload Optimization Manager can recommend scaling from a General Purpose to a Next-gen General Purpose or Business Critical service tier, but not vice versa.
- For Next-gen General Purpose service tiers:
  - According to the Azure documentation, Next-gen General Purpose service tiers are currently in preview. In Workload Optimization Manager, scale actions that involve these tiers may change as Azure changes its level of support for these tiers.
  - Actions that scale up IOPS to address congestion are recommended on the Next-gen General Purpose service tier only. Workload Optimization Manager does not recommend actions to scale storage to increase the IOPS amount.
  - If a SQL Managed Instance has more than 100 user databases, Workload Optimization Manager does not recommend scaling from a Next-gen General Purpose service tier to a Business Critical service tier. This behavior ensures compliance with resource limits enforced by Azure.
  - Workload Optimization Manager adds V2 to the service tier names that display in actions and scale policies. For example, GP_V2_Gen5_4 in Workload Optimization Manager corresponds to GP_Gen5_4 in Azure.



Utilization charts for scale actions

Workload Optimization Manager uses percentile calculations to measure resource utilization, and drive scaling actions that improve overall utilization and reduce costs. When you examine the details for a pending scaling action on a database server, you will see charts that highlight resource *utilization percentiles* for a given observation period, and the projected percentiles after you execute the action.



The charts also plot *daily average utilization* for your reference. If you have previously executed scaling actions on the database server, you can see the resulting improvements in daily average utilization. Put together, these charts allow you to easily recognize utilization trends that drive Workload Optimization Manager's scaling recommendations.

> **NOTE:**
> You can set scaling constraints in database server policies to refine the percentile calculations. For details, see Scaling Sensitivity *(on page 695)*.

Non-disruptive and reversible scaling actions

All scaling actions shown in the Action Center view and Action Details page are non-disruptive and reversible.

Cloud database server policies for AWS RDS and Azure SQL Managed Instances

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

Action	Default setting/mode
Database Server Scaling Actions	On
Disruptive irreversible scaling	Manual
Disruptive reversible scaling	Manual
Non-disruptive irreversible scaling	Manual
Non-disruptive reversible scaling	Manual

You can use Action Scripts *(on page 1070)* and third-party orchestrators (such as ServiceNow) for action orchestration.

Scaling sensitivity

Workload Optimization Manager uses a percentile of utilization over the specified observation period. This gives sustained utilization and ignores short-lived bursts.

Workload Optimization Manager uses these settings to calculate utilization percentiles for vCPU, vMem, DB Cache Hit Rate, and IOPS. It then recommends actions to improve utilization based on the observed values for a given time period.

- Aggressiveness

Attribute	Default value
Aggressiveness	95th Percentile

  When evaluating performance, Workload Optimization Manager considers resource utilization as a percentage of capacity. The utilization drives actions to scale the available capacity either up or down. To measure utilization, the analysis considers a given utilization percentile. For example, assume a 95th percentile. The percentile utilization is the highest value that 95% of the observed samples fall below. Compare that to average utilization, which is the average of *all* the observed samples.

  Using a percentile, Workload Optimization Manager can recommend more relevant actions. This is important in the cloud, so that analysis can better exploit the elasticity of the cloud. For scheduled policies, the more relevant actions will tend to remain viable when their execution is put off to a later time.

  For example, consider decisions to reduce capacity. Without using a percentile, Workload Optimization Manager never resizes below the recognized peak utilization. Assume utilization peaked at 100% just once. Without the benefit of a percentile, Workload Optimization Manager will not reduce resources for that Database Server.

  With **Aggressiveness**, instead of using the single highest utilization value, Workload Optimization Manager uses the percentile you set. For the previous example, assume a single burst to 100%, but for 95% of the samples, utilization never exceeded 50%. If you set **Aggressiveness** to 95th Percentile, then Workload Optimization Manager can see this as an opportunity to reduce resource allocation.

  In summary, a percentile evaluates the sustained resource utilization, and ignores bursts that occurred for a small portion of the samples. You can think of this as aggressiveness of resizing, as follows:

  - 99th Percentile – More performance. Recommended for critical Database Servers that need maximum guaranteed performance at all times, or those that need to tolerate sudden and previously unseen spikes in utilization, even though sustained utilization is low.
  - 95th Percentile (Default) – The recommended setting to achieve maximum performance and savings. This assures performance while avoiding reactive peak sizing due to transient spikes, thus allowing you to take advantage of the elastic ability of the cloud.
  - 90th Percentile – More efficiency. Recommended for Database Servers that can stand higher resource utilization.

  By default, Workload Optimization Manager uses samples from the last 14 days. Use the **Max Observation Period** setting to adjust the number of days. To ensure that there are enough samples to analyze and drive scaling actions, set the **Min Observation Period**.

- Max observation period

Attribute	Default value
Max Observation Period	Last 14 Days

To refine the calculation of resource utilization percentiles, you can set the sample time to consider. Workload Optimization Manager uses historical data from up to the number of days that you specify as a sample period. If the Database Server has fewer days' data then it uses all of the stored historical data.

You can make the following settings:

  - Less Elastic – Last 30 Days
  - Recommended – Last 14 Days
  - More Elastic – Last 7 Days or Last 3 Days

Workload Optimization Manager recommends an observation period of 14 days so it can recommend scaling actions more often. Since Database Server scaling is minimally disruptive, scaling often should not introduce any noticeable performance risks.

- Min observation period

Attribute	Default value
Min Observation Period	None

This setting ensures historical data for a minimum number of days before Workload Optimization Manager will generate an action based on the percentile set in **Aggressiveness**. This ensures a minimum set of data points before it generates the action.

Especially for scheduled actions, it is important that resize calculations use enough historical data to generate actions that will remain viable even during a scheduled maintenance window. A maintenance window is usually set for "down" time, when utilization is low. If analysis uses enough historical data for an action, then the action is more likely to remain viable during the maintenance window.

  - More Elastic – None
  - Less Elastic – 1, 3, or 7 Days

Cloud instance types

By default, Workload Optimization Manager considers all instance types currently available for scaling when making scaling decisions for Database Servers. However, you may have set up your Database Servers to only scale to certain instance types to reduce complexity and cost, or meet demand. Use this setting to identify those instance types.

NOTE:
Workload Optimization Manager automatically discovers and enforces Database Server tier exclusions configured in your AWS environment. You do not need to configure these tier exclusions in policies. To see a list of tier exclusions that are currently enforced, set the scope to one or several Database Servers and click the **Policies** tab.

Attribute	Default value
Cloud Instance Types	None

Click **Edit** to set your preferences. In the new page that displays, choose from the following options:

- Excluded

  By default, this option is selected and all instance types are deselected. This means that all instance types are considered for scaling.

  Any instance types that you select are excluded from scaling.

Configure an exclusion list if there are fewer instance types to exclude than to include. For example, if 400 instance types are currently available for scaling and you do not want Database Servers to scale to 20 instance types, configure an exclusion list and select the 20 instance types.

- Included

  Any instance types that you select are considered for scaling.

  Configure an inclusion list if there are fewer instance types to include than to exclude. For example, if 400 instance types are currently available for scaling and you want Database Servers to scale only to 50 instance types, configure an inclusion list and select the 50 instance types.

Expand an instance family to see individual instance types and the resources allocated to them.

After you save your changes, the main page refreshes to reflect your selections.

NOTE:
This policy setting is not available in plans.

Scaling target utilization

This is the target utilization as a percentage of capacity.

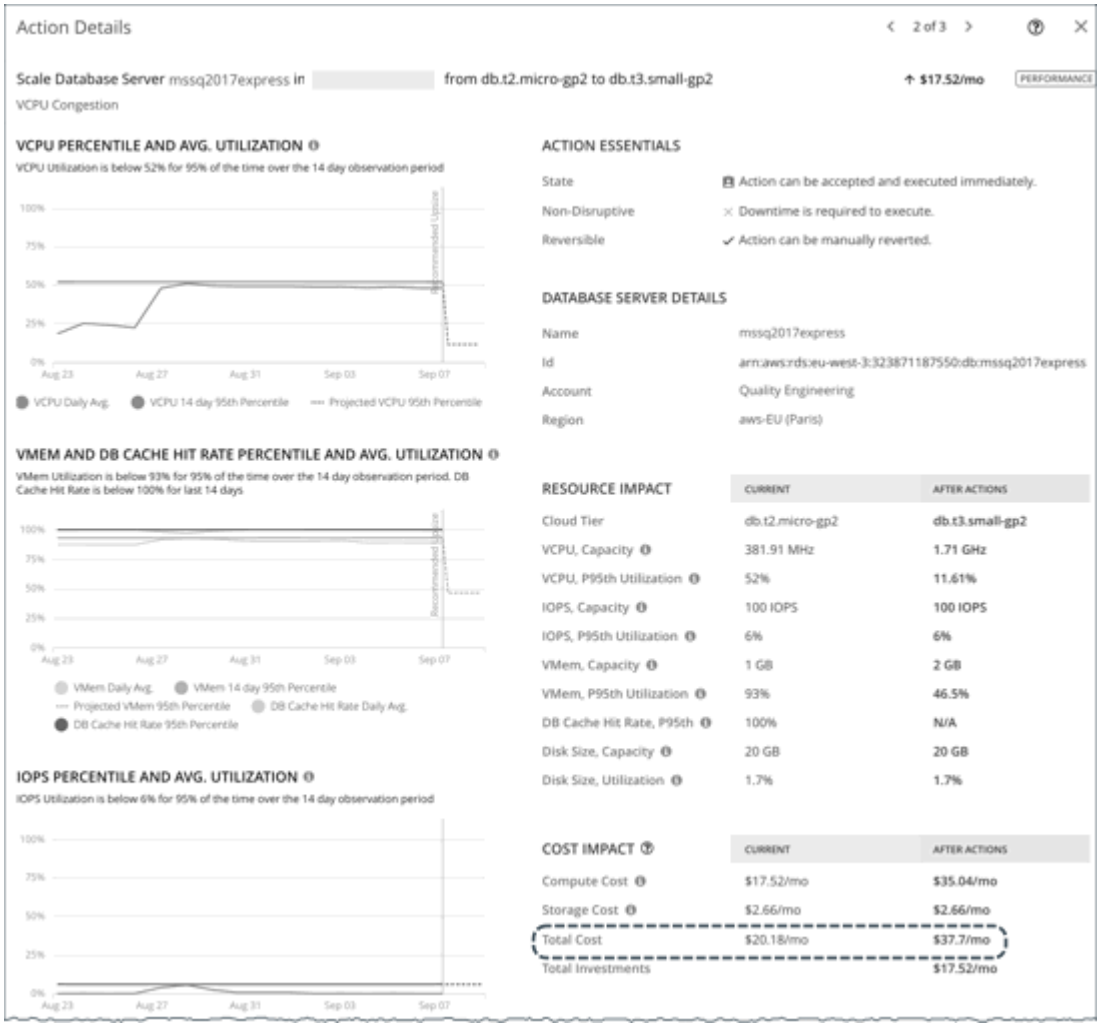
Attribute	Default value
VCPU	70
VMEM	90
IOPS	70
Storage Amount	90

These advanced settings determine how much you would like a scope of workloads to utilize their resources. These are fixed settings that override the way Workload Optimization Manager calculates the optimal utilization of resources. You should only change these settings after consulting with Technical Support.

While these settings offer a way to modify how Workload Optimization Manager recommends actions, in most cases you should never need to use them. If you want to control how Workload Optimization Manager recommends actions to resize workloads, you can set the aggressiveness per the percentile of utilization, and set the length of the sample period for more or less elasticity on the cloud.

Estimated on-demand costs for cloud database servers

Workload Optimization Manager considers a variety of factors when calculating estimated on-demand monthly cost for cloud database servers.



Azure SQL Managed Instances - cost calculation

For Azure SQL Managed Instances, the calculation for estimated on-demand monthly cost can be expressed as follows:

```
(On-demand Compute Rate * 730) + (Provisioned Database Storage Rate * (Provisioned Database Storage Amount - Included Storage Per Server)) + (Provisioned IOPS Rate * (Provisioned Database IOPS - Included IOPS Per Server)) = Estimated On-demand Monthly Cost
```

Where:

- **On-demand Compute Rate** is the hourly cost for a SQL Managed Instance.

  For more information about on-demand rates, see this Azure page.
- **730** is the number of hours per month that Workload Optimization Manager uses to estimate monthly costs.
- **Provisioned Database Storage Rate** is the hourly cost for a SQL Managed Instance's provisioned database storage.

  For more information about provisioned database storage rates, see this Azure page.
- **Provisioned Database Storage Amount** is the amount of storage provisioned for a SQL Managed Instance.
- **Included Storage Per Server** is the amount of storage that is provided at no extra cost. Included storage is always 32 GB, regardless of configuration or service tier.
- **Provisioned IOPS Rate** is the monthly cost for a SQL Managed Instance's provisioned IOPS.

Provisioned IOPS applies only to SQL Managed Instances on the Next-gen General Purpose service tier. Provisioned IOPS is either 300 IOPS or `3 * Provisioned Database Storage Amount in GB`, whichever is greater.

For more information about provisioned IOPS rates, see this Azure page.

- **Provisioned Database IOPS** is the amount of IOPS provisioned for a SQL Managed Instance.
- **Included IOPS Per Server** is the amount of IOPS that is provided at no extra cost.

  Workload Optimization Manager checks the following values and picks the greater value:

  - 300 IOPS
  - `3 * number of IOPS in GB` or the maximum amount of IOPS allowed for a given size, whichever is lower

The preceding items impact cost calculations and the scaling decisions that Workload Optimization Manager makes. These decisions also rely on other factors, such as resource utilization percentiles and scaling constraints set in policies.

Cost calculation example: Azure SQL Managed Instances

Assume the following data for a pending scale action for a SQL Managed Instance on General Purpose Premium-series. Workload Optimization Manager recommends scaling from `GP_G8IM_10` to `GP_G8IM_4` and reducing disk size to 64 GB.

Item	Current values	Values after action execution
On-demand Compute Rate	$2.760/hr	$1.104/hr
Provisioned Database Storage Rate	$0.133/GB/month	$0.133/GB/month
Included Storage Per Server	32 GB	32 GB
Provisioned Database Storage Amount	750 GB	64 GB
Provisioned IOPS Rate	$0.038/IOPS	$0.038/IOPS
Included IOPS Per Server	300	300
Provisioned IOPS Amount	7000	300
Included IOPS	2208	300

Workload Optimization Manager calculates the following values:

- Current estimated on-demand monthly cost:

```
(2.760 * 730) + (0.133 * (750 - 32) + (0.038 * (7000 - 300)) = 2364.89
```

- Estimated on-demand monthly cost after executing the action:

```
(1.104 * 730) + (0.133 * (64 - 32) + (0.038 * (300 - 300)) = 810.18
```

NOTE:
Workload Optimization Manager rounds the calculated values that it displays in the user interface.

The estimated on-demand monthly cost is projected to reduce from $2364.89/month to $810.18/month.

Workload Optimization Manager treats the action as a cost-saving measure and shows estimated savings of $1554.71/month.

AWS RDS - cost calculation

For AWS RDS, the calculation for estimated on-demand monthly cost can be expressed as follows:

```
(On-demand Compute Rate * 730) + (Provisioned Database Storage Rate * Provisioned
Database Storage Amount) + (Provisioned IOPS Rate * Provisioned IOPS Amount) = Estimated
On-demand Monthly Cost
```

Where:

- **On-demand Compute Rate** is the hourly cost for an AWS RDS's instance type.

  For more information about on-demand rates, see this AWS page.
- **730** is the number of hours per month that Workload Optimization Manager uses to estimate monthly costs.
- **Provisioned Database Storage Rate** is the hourly cost for an AWS RDS's provisioned database storage.

  For more information about provisioned database storage rates, see this AWS page.
- **Provisioned IOPS Rate** is the monthly cost for an AWS RDS's provisioned IOPS.

  Provisioned IOPS applies only to AWS RDS on Provisioned IOPS SSD (io1) storage. For more information about provisioned IOPS SSD storage, see the AWS documentation.

  For more information about provisioned IOPS rates, see this AWS page.

The preceding items impact cost calculations and the scaling decisions that Workload Optimization Manager makes. These decisions also rely on other factors, such as resource utilization percentiles and scaling constraints set in policies.

Cost calculation example: AWS RDS

Assume the following data for a pending scale action for AWS RDS SQL Server Express Edition (Single A-Z deployment):

Item	Current values	Values after action execution
On-demand Compute Rate	$0.024/hr	$0.048/hr
Provisioned Database Storage Rate	$0.133/hr	$0.133/hr
Provisioned Database Storage Amount	20 GB	20 GB
Provisioned IOPS Rate	$0.00	$0.00
Provisioned IOPS Amount	0	0

Workload Optimization Manager calculates the following values:

- Current estimated on-demand monthly cost:

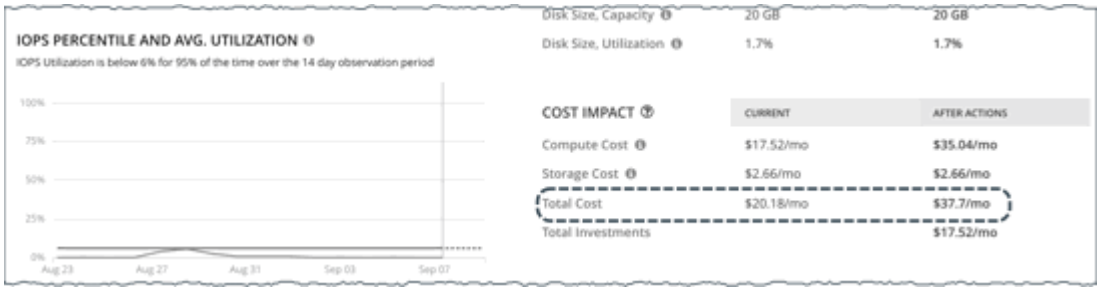
```
(0.024 * 730) + (0.133 * 20) + (0.00 * 0) = 20.18
```

- Estimated on-demand monthly cost after executing the action:

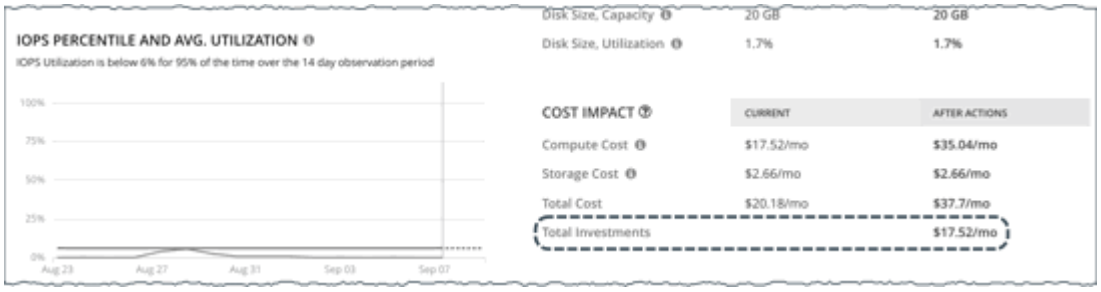
```
(0.048 * 730) + (0.133 * 20) + (0.00 * 0) = 37.7
```

NOTE:
Workload Optimization Manager rounds the calculated values that it displays in the user interface.

The estimated on-demand monthly cost is projected to increase from $20.18/month to $37.7/month, as shown in the Details section of the pending action.



Workload Optimization Manager treats the action as an investment and shows an estimated investment of $17.52/month.



AWS Aurora - cost calculation

For AWS Aurora, the calculation for estimated on-demand monthly cost can be expressed as follows:

```
(On-demand Compute Rate * 730) + (Provisioned Database Storage Rate * Provisioned
Database Storage Amount) + (I/O Request Rate * (Hourly Billed I/O Operation Count * 730))
= Estimated On-demand Monthly Cost
```

Where:

- **On-demand Compute Rate** is the hourly cost for an AWS Aurora's instance type.

  For more information about on-demand rates, see this AWS page.
- **730** is the number of hours per month that Workload Optimization Manager uses to estimate monthly costs.
- **Provisioned Database Storage Rate** is the hourly cost for an AWS Aurora's provisioned database storage.

  For more information about provisioned database storage rates, see this AWS page.
- **I/O Request Rate** is the cost per one million read/write I/O requests.

  For more information about I/O request rates, see this AWS page.
- **Hourly Billed I/O Operation Count** is the average sum of read and write I/O operations per hour over the last month.

The preceding items impact cost calculations. Except for I/O request rate, these items affect the actual scaling decisions that Workload Optimization Manager makes. These decisions also rely on other factors, such as resource utilization percentiles and scaling constraints set in policies.

Cost calculation example: AWS Aurora

Assume the following data for a pending scale action for AWS Aurora MySQL-Compatible Edition:

Item	Current values	Values after action execution
On-demand Compute Rate	$0.164/hr	$0.041/hr
Provisioned Database Storage Rate	$0.10/hr	$0.10/hr
Provisioned Database Storage Amount	100	100
I/O Request Rate	$0.20/one million requests	$0.20/one million requests
Hourly Billed I/O Operation Count	2000	2000

Workload Optimization Manager calculates the following values:

- Current estimated on-demand monthly cost:

  ```
  (0.164 * 730) + (0.10 * 100) + ((0.20 / 1000000) * (2000 * 730)) = 130.01
  ```

- Estimated on-demand monthly cost after executing the action:

  ```
  (0.041 * 730) + (0.10 * 100) + ((0.20 / 1000000) * (2000 * 730)) = 40.22
  ```

NOTE:
Workload Optimization Manager rounds the calculated values that it displays in the user interface.

Since the estimated on-demand monthly cost is projected to decrease from $130.01/month to $40.22/month, Workload Optimization Manager treats the action as a cost-saving measure and shows estimated savings of $89.79/month.

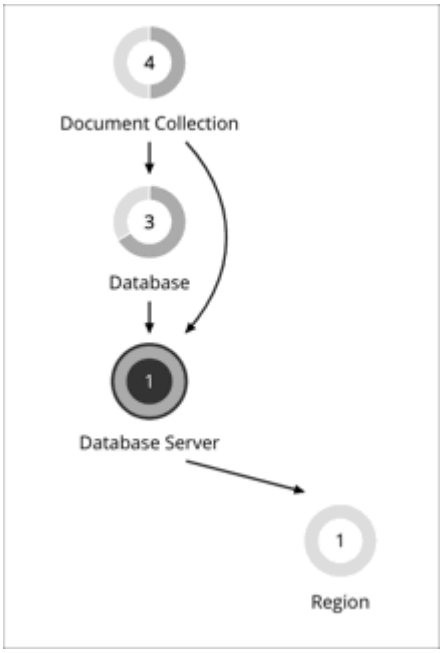
Database server - Azure Cosmos DB account

In Azure Cosmos DB *(on page 470)*, an account is an entity that manages databases *(on page 713)* and containers (document collections *(on page 722)*). An account has a total throughput limit that represents capacity, but actual throughput (measured in Request Units or RUs) is provisioned at the database or container level.

Workload Optimization Manager discovers Azure Cosmos DB accounts through your Azure targets, and represents them as database server entities in the supply chain.

NOTE:
Workload Optimization Manager does not discover Cosmos DB serverless accounts.

Synopsis



Synopsis	
Provides:	RUs to Cosmos DB databases and containers (document collections)
Consumes:	Storage resources
Discovered through:	Azure targets

Monitored resources

Workload Optimization Manager monitors the following resources:

- Request unit (RU)

  Request unit (RU) is a performance currency that abstracts CPU, IOPS, and memory that are required to perform the database operations supported by Azure Cosmos DB. Azure Cosmos DB normalizes the cost of all database operations using RUs.

- Storage amount

  Storage amount is the measurement of storage capacity that is in use.

Actions

None

Workload Optimization Manager does not recommend actions for a Cosmos DB account but it does recommend actions for the databases *(on page 713)* and document collections *(on page 722)* in the account.

Database (cloud)

Workload Optimization Manager represents the following public cloud resources as Database entities in the supply chain:

- Azure SQL databases (vCore or DTU pricing model) *(on page 703)*
- Azure dedicated SQL pool *(on page 710)*
- Azure Cosmos DB databases *(on page 713)*

NOTE:
AWS RDS, Azure SQL Managed Instances, and Azure Cosmos DB accounts appear as *Database Server* entities in the supply chain. For details, see Database Server (Cloud) *(on page 684)*.

Create groups in Workload Optimization Manager so you can manage SQL databases, dedicated SQL pools, and Cosmos DB databases with ease. When you set the scope to a particular group, you can evaluate pending actions, view metrics, and assign policies as usual. You can also view the status of individual members. For example, you can see if a dedicated SQL pool is currently active or suspended.

To create groups, go to **Settings > Groups**, and then click **New Group > Database**. In the page that displays, choose **Dynamic** (recommended) and then set the `Pricing Model` filter as follows:

- `DTU` and/or `vCore` – use to create a group of SQL databases
- `DWU` – use to create a group of dedicated SQL pools
- `RU` – use to create a group of Cosmos DB databases

Database – Azure SQL

Azure SQL database is an individual database that is managed under the DTU (Database Transaction Unit) or vCore pricing model.

- DTU Pricing Model

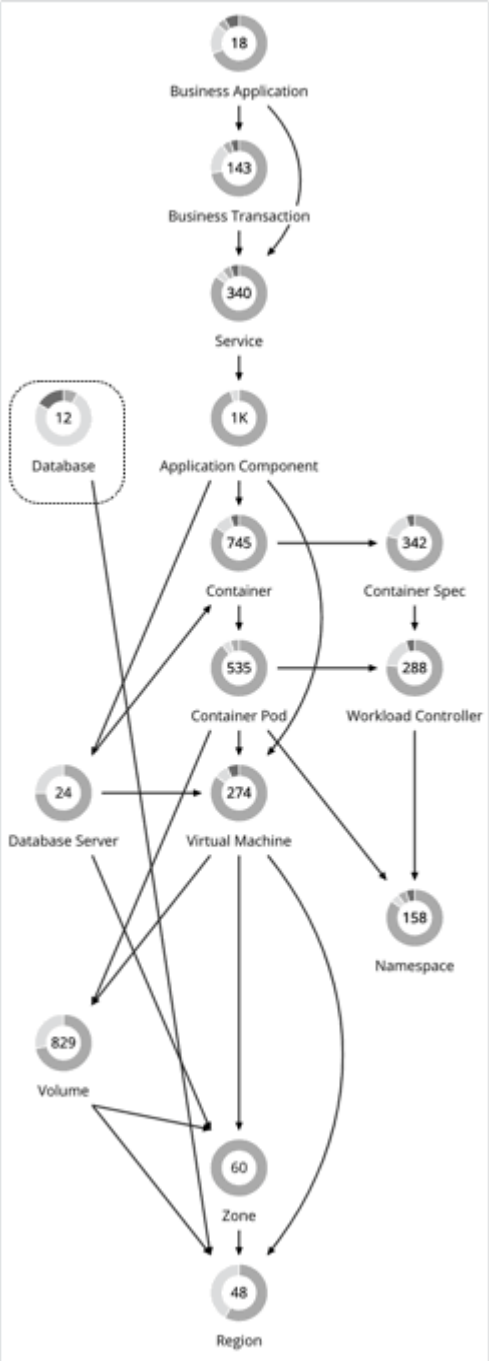
  In the DTU model, Azure bundles CPU, memory, and IOPS as a single DTU metric. Workload Optimization Manager actions on these databases consider both DTU and storage utilization.

- vCore Pricing Model

  In the vCore model, analysis can track CPU, memory, IOPS, and throughput metrics in isolation. Workload Optimization Manager actions on these databases are driven by CPU, memory, IOPS, throughput and storage utilization.

Workload Optimization Manager discovers SQL databases through your Azure targets, and represents them as Database entities in the supply chain.

Synopsis



Synopsis	
Provides:	Transactions to end users
Consumes:	▪ DTU Pricing Model: DTU and storage resources in an Azure region ▪ vCore Pricing Model: vCPU, vMem, IOPS, throughput, and storage resources in an Azure region

Synopsis	
Discovered through:	Azure targets

Monitored resources

The resources that Workload Optimization Manager can monitor depend on the pricing model in place for the given database entity.

- DTU Pricing Model
  - DTU

    DTU is the measurement of compute capacity for the database. DTU represents CPU, memory, and IOPS/IO Throughput bundled as a single commodity.
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
- vCore Pricing Model
  - Virtual memory (vMem)

    Virtual memory (vMem) is the measurement of memory that is in use.
  - Virtual CPU (vCPU)

    Virtual CPU is the measurement of CPU that is in use.
  - Storage amount

    Storage amount is the measurement of storage capacity that is in use.
  - Storage access (IOPS)

    Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
  - I/O throughput

    I/O throughput is the measurement of an entity's throughput to the underlying storage.

Scale actions for SQL databases

Workload Optimization Manager database scaling actions aim to increase resource utilization and reduce costs while complying with business policies.

Workload Optimization Manager drives scaling actions based on the utilization of resources, and treats the following limits as constraints when it makes scaling decisions:

- Maximum concurrent sessions

  Maximum concurrent sessions is the maximum number of database connections at a time.
- Maximum concurrent workers

  Maximum concurrent workers is the maximum number of database processes that can handle queries at a time.

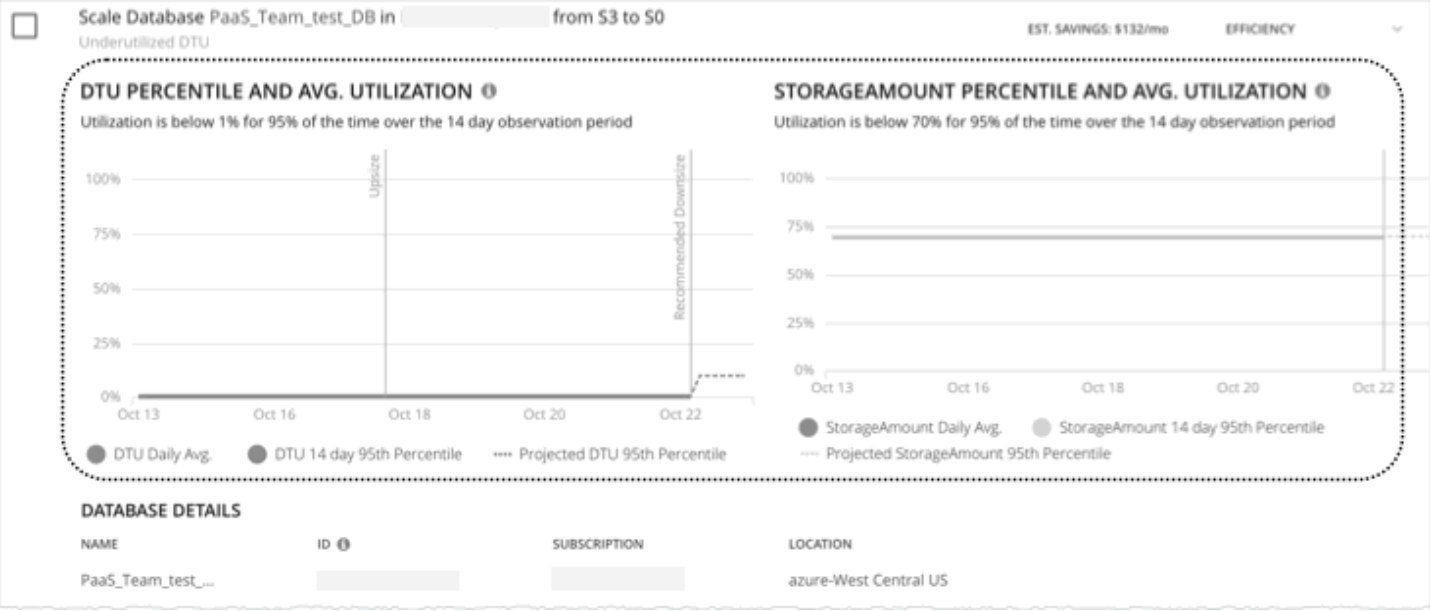
Points to consider:

- Workload Optimization Manager will *not* recommend:
  - Scaling from one pricing model to another
  - Scaling vCore databases to instance types running Gen4 hardware. This hardware generation has been retired and pricing information can no longer be retrieved via the Azure API.
  - Scaling vCore databases on the serverless compute tier
  - Scaling provisioned memory for vCore databases on the Hyperscale service tier. VMem utilization data is currently unavailable for Hyperscale due to an issue in the Azure API.
- On DTU databases, a single action can scale both DTU and storage. On vCore databases, a single action can scale vCPU, vMem, IOPS, throughput, and storage.

- In some cases, Workload Optimization Manager might recommend scaling up storage, even if there is no storage pressure on the database, to take advantage of storage provided at no extra cost. For example, Workload Optimization Manager might recommend scaling from the *S3* to the *S0* tier because of low DTU and storage utilization. Since the *S0* tier includes 250 GB of storage at no extra cost, Workload Optimization Manager will also recommend scaling up to this storage amount. If you want to scale DTU but keep the storage amount unchanged, adjust the values for aggressiveness (percentile) and observation period in your database policies.

Utilization Charts for Scale Actions

Workload Optimization Manager uses percentile calculations to measure resource utilization, and drive scaling actions that improve overall utilization and reduce costs. When you examine the details for a pending scaling action on a database, you will see charts that highlight resource *utilization percentiles* for a given observation period, and the projected percentiles after you execute the action.



The charts also plot *daily average utilization* for your reference. If you have previously executed scaling actions on the database, you can see the resulting improvements in daily average utilization. Put together, these charts allow you to easily recognize utilization trends that drive Workload Optimization Manager's scaling recommendations.

NOTE:
You can set scaling constraints in database policies to refine the percentile calculations. For details, see Aggressiveness and Observation Period *(on page 716)*.

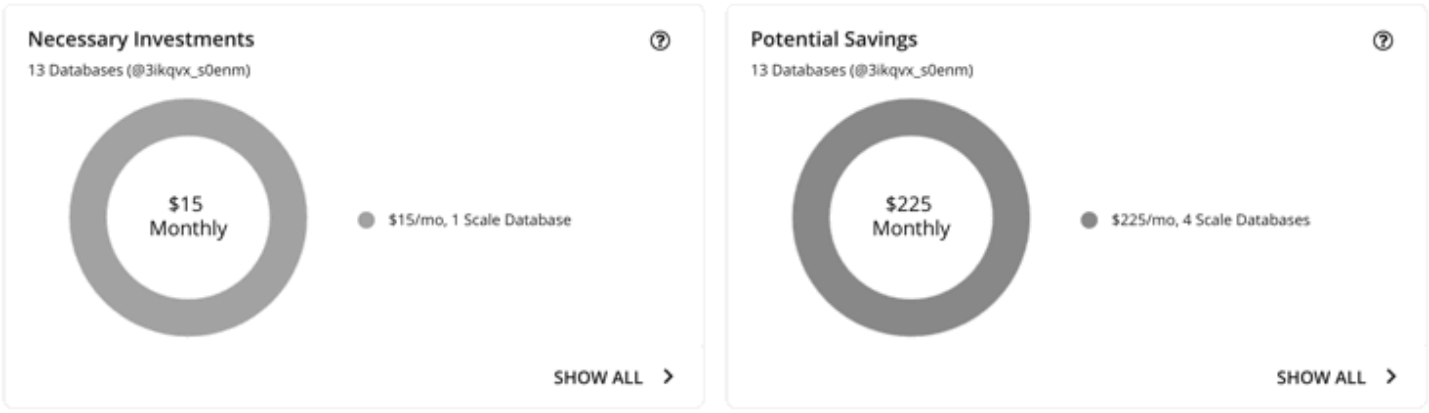
Non-disruptive and reversible scaling actions

All scaling actions shown in the Action Center view and Action Details page are non-disruptive and reversible.

For actions to scale vCore databases from General Purpose or Business Critical to Hyperscale, there are certain caveats associated with reversing such actions. To learn more, see the Azure documentation.

Actions in charts

Use the Necessary Investments and Potential Savings charts to view pending database actions. These charts show total monthly investments and savings, assuming you execute all the actions.



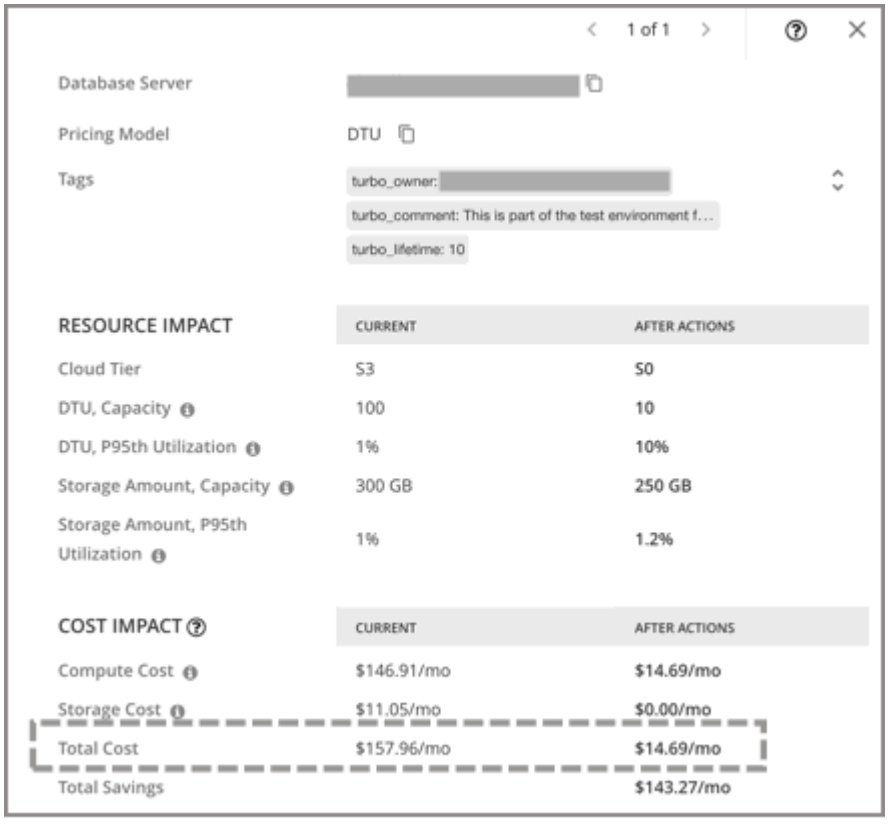
Click **Show All** for each chart to review and execute the actions.

The table shows the following:

- Actions that are pending for each database
- Savings or investments for each database

Estimated on-demand costs for Azure SQL databases

Workload Optimization Manager considers a variety of factors when calculating *Estimated On-demand Monthly Cost* for an Azure SQL Database.



Azure SQL DTU databases - cost calculation

For Azure SQL DTU databases, the calculation for estimated on-demand monthly cost can be expressed as follows:

```
(On-demand compute Rate * 730)) + (Provisioned Database Storage Rate * (Provisioned
Database Storage Amount - Performance Level Included Storage)) = Estimated On-demand
Monthly Cost
```

Where:

- **On-demand Compute Rate** is the hourly cost for a DTU database's instance type.

  For more information about on-demand rates, see this Azure page.
- **730** is the number of hours per month that Workload Optimization Manager uses to estimate monthly costs.
- **Provisioned Database Storage Rate** is the cost for 1 GB / month of a DTU database's provisioned storage.

  For more information about provisioned database storage rates, see this Azure page.
- **Performance Level Included Storage** is the storage amount included in the price of the selected Performance Level of a DTU database.

  For more information about DTU storage limits, see the Azure documentation.

The preceding items impact cost calculations and the scaling decisions that Workload Optimization Manager makes. These decisions also rely on other factors, such as resource utilization percentiles and scaling constraints set in policies.

Cost calculation example: Azure SQL DTU databases

Assume the following data for a pending scale action for an Azure SQL DTU database:

Item	Current values	Values after action execution
On-demand Compute Rate	$0.20125/hr	$0.020125/hr
Provisioned Database Storage Rate	$0.221 per 1 GB/Mo.	$0.221 per 1 GB/Mo.
Performance Level Included Storage	250 GB	250 GB
Provisioned Database Storage Amount	300 GB	250 GB

Workload Optimization Manager calculates the following values:

- Current estimated on-demand monthly cost:

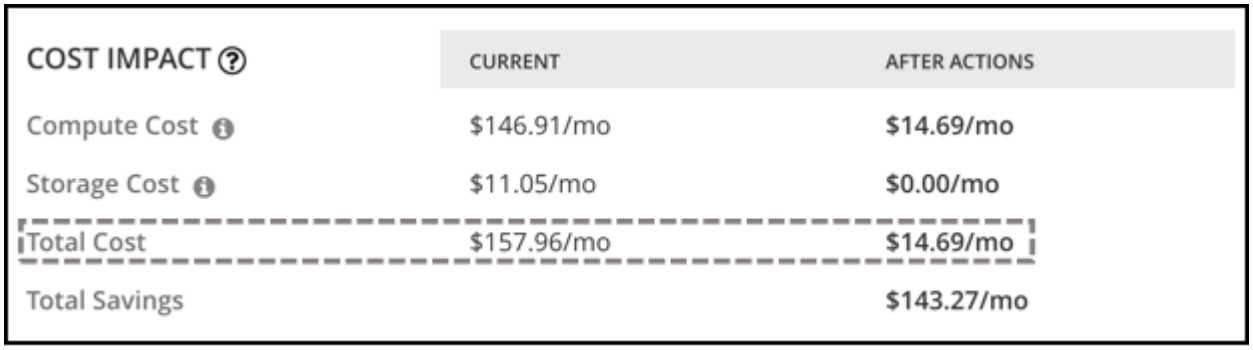
  ```
  ($0.20125 * 730) + ($0.221 * (300 - 250)) = $157.96/month
  ```

- Estimated on-demand monthly cost after executing the action:

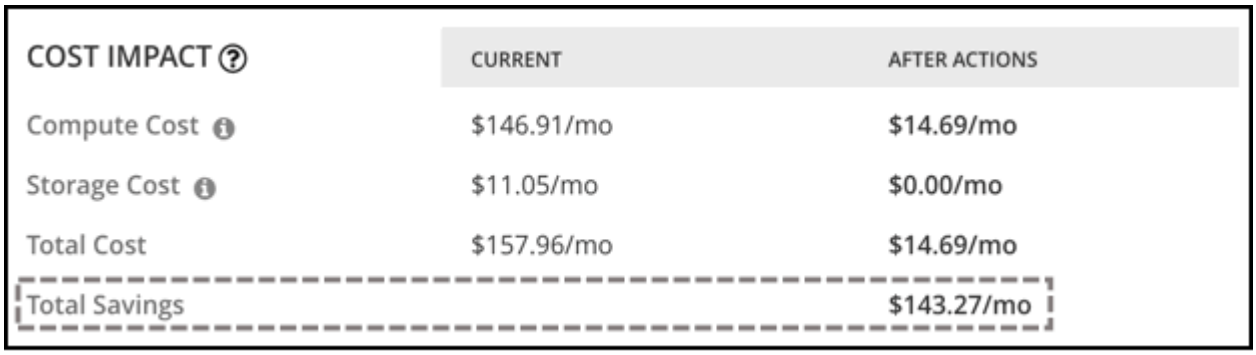
  ```
  ($0.020125 * 730) + ($0.221 * (250 - 250)) = $14.69/month
  ```

NOTE:
Workload Optimization Manager rounds the calculated values that it displays in the user interface.

The estimated on-demand monthly cost is projected to decrease from $157.96/month to $14.69/month, as shown in the Details section of the pending action.



Workload Optimization Manager treats the action as a cost-saving measure and shows estimated savings of $143.27/month.



Azure SQL vCore databases - cost calculation

For Azure SQL vCore databases, the calculation for estimated on-demand monthly cost can be expressed as follows:

```
(On-demand Compute Rate * 730) + (SQL License Rate * 730) + (Provisioned Database Storage Rate * (Provisioned Database Storage Amount + Log Space Allocated)) = Estimated On-demand Monthly Cost
```

Where:

- **On-demand Compute Rate** is the hourly cost for a vCore database's instance type.

  For more information about on-demand rates, see this Azure page.
- **730** is the number of hours per month that Workload Optimization Manager uses to estimate monthly costs.
- **SQL License Rate** is the hourly cost for a vCore database's SQL license.

  For more information about SQL license rates, see this Azure page.

  NOTE:
  In the Azure page, pay-as-you-go prices represent the sum of compute and license costs, while Azure Hybrid Benefit Price values represent compute costs only.

- **Provisioned Database Storage Rate** is the cost for 1 GB / month of a vCore database's provisioned storage.

  For more information about provisioned database storage rates, see this Azure page.
- **Log Space Allocated** is the log storage space that Azure automatically allocates to a single vCore database instance.

NOTE:
Log Storage Space is considered in database cost calculations, but not reflected in storage capacity.

For more information about provisioned database storage rates, see this Azure page.

The preceding items impact cost calculations and the scaling decisions that Workload Optimization Manager makes. These decisions also rely on other factors, such as resource utilization percentiles and scaling constraints set in policies.

Cost calculation example: Azure SQL vCore databases

Assume the following data for a pending scale action for an Azure SQL vCore database:

Item	Current values	Values after action execution
On-demand Compute Rate	$1.068/hr	$0.304/hr
SQL License Rate	$0.799728/hr	$0.199932/hr
Provisioned Database Storage Rate	$0.115/hr	$0.115/hr
Provisioned Database Storage Amount	32 GB	5 GB

Workload Optimization Manager calculates the following values:

- Current estimated on-demand monthly cost:

```
($1.068 * 730) + ($0.799728 * 730) + ($0.115 * (32 + 9.6)) = $1368.23/month
```

- Estimated on-demand monthly cost after executing the action:

```
($0.304 * 730) + ($0.199932 * 730) + ($0.115 * (5 + 1.5)) = $368.62/month
```

NOTE:
Workload Optimization Manager rounds the calculated values that it displays in the user interface.

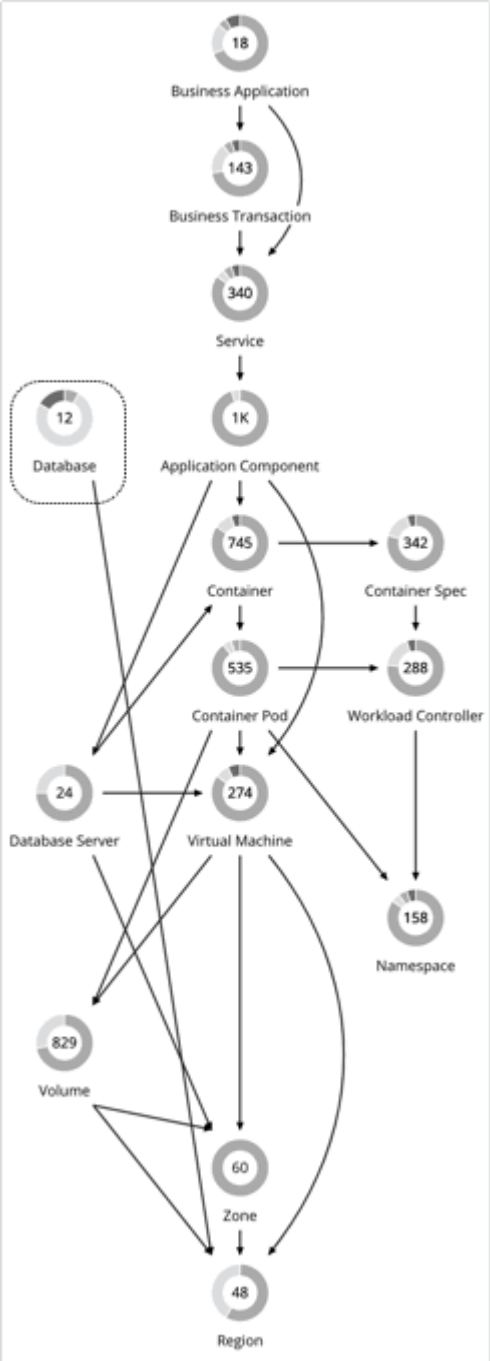
Since the estimated on-demand monthly cost is projected to decrease from $1368.23/month to $368.62/month, Workload Optimization Manager treats the action as a cost-saving measure and shows estimated savings of $999.61/month.

Database - Azure dedicated SQL pool

A dedicated SQL pool in Azure Synapse Analytics represents a collection of analytic resources that Azure provisions when you use Synapse SQL. Compute capacity for a dedicated SQL pool is expressed in Data Warehousing Units (DWU). DWU represents CPU, memory, and IO Throughput bundled as a single commodity.

Workload Optimization Manager discovers dedicated SQL pools through your Azure targets, and represents them as Database entities in the supply chain.

Synopsis



Synopsis	
Provides:	Analytic resources to Azure Synapse Analytics
Consumes:	DWU and storage resources in an Azure region
Discovered through:	Azure targets

Monitored resources

Workload Optimization Manager monitors the following resources:

- DWU

  DWU (Data Warehousing Unit) is the measurement of compute capacity for the dedicated SQL pool. DWU represents CPU, memory, and IO Throughput bundled as a single commodity.

- Storage amount

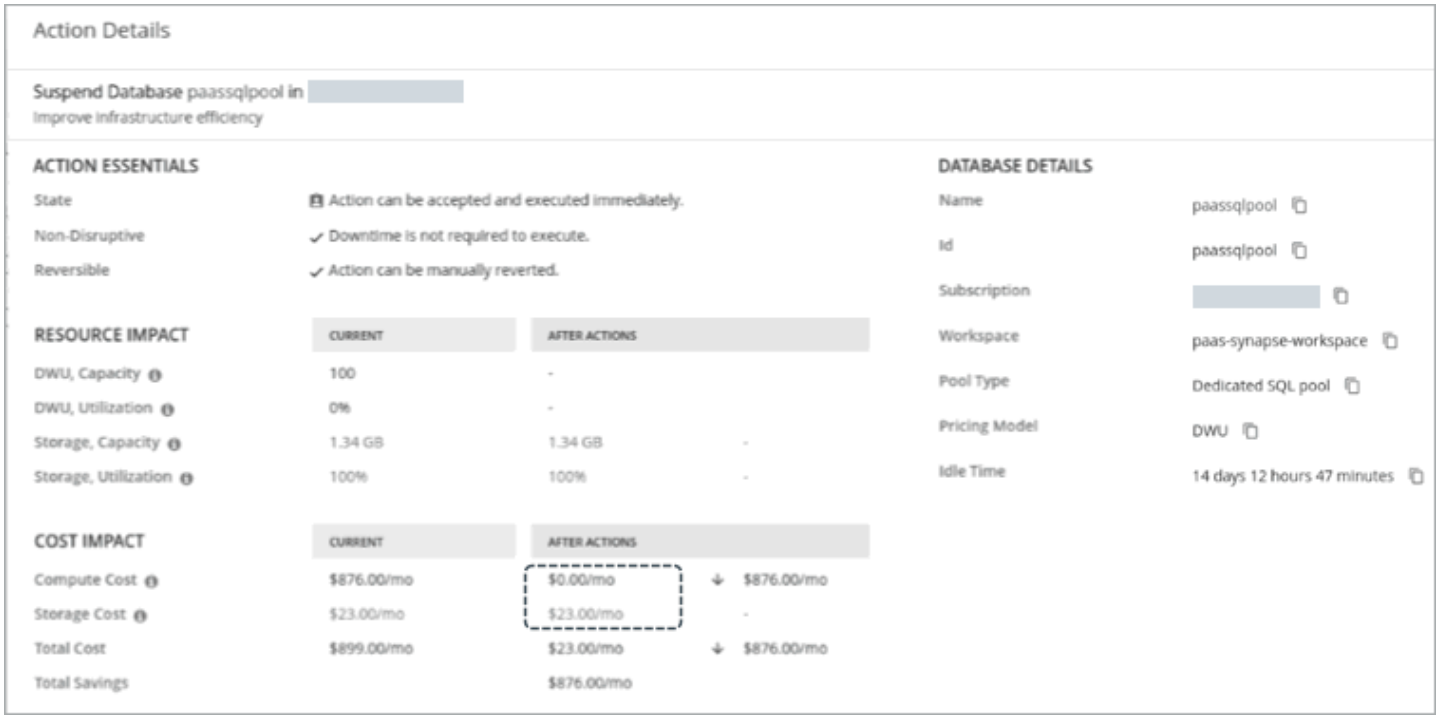
  Storage amount is the measurement of storage capacity that is in use.

- Connection

  Connection is the measurement of database connections utilized by applications.

Suspend actions for dedicated SQL pools

When Workload Optimization Manager analysis discovers an idle dedicated SQL pool (i.e., a pool with no active connections for a specific period of time), it will immediately recommend that you suspend the pool as a cost-saving measure, and show the action in Action Center. This action stops the compute instance for the SQL pool, but keeps storage running. In the Action Details page for a specific action, the compute cost is 0 (zero) *after* suspension, while storage cost remains the same.



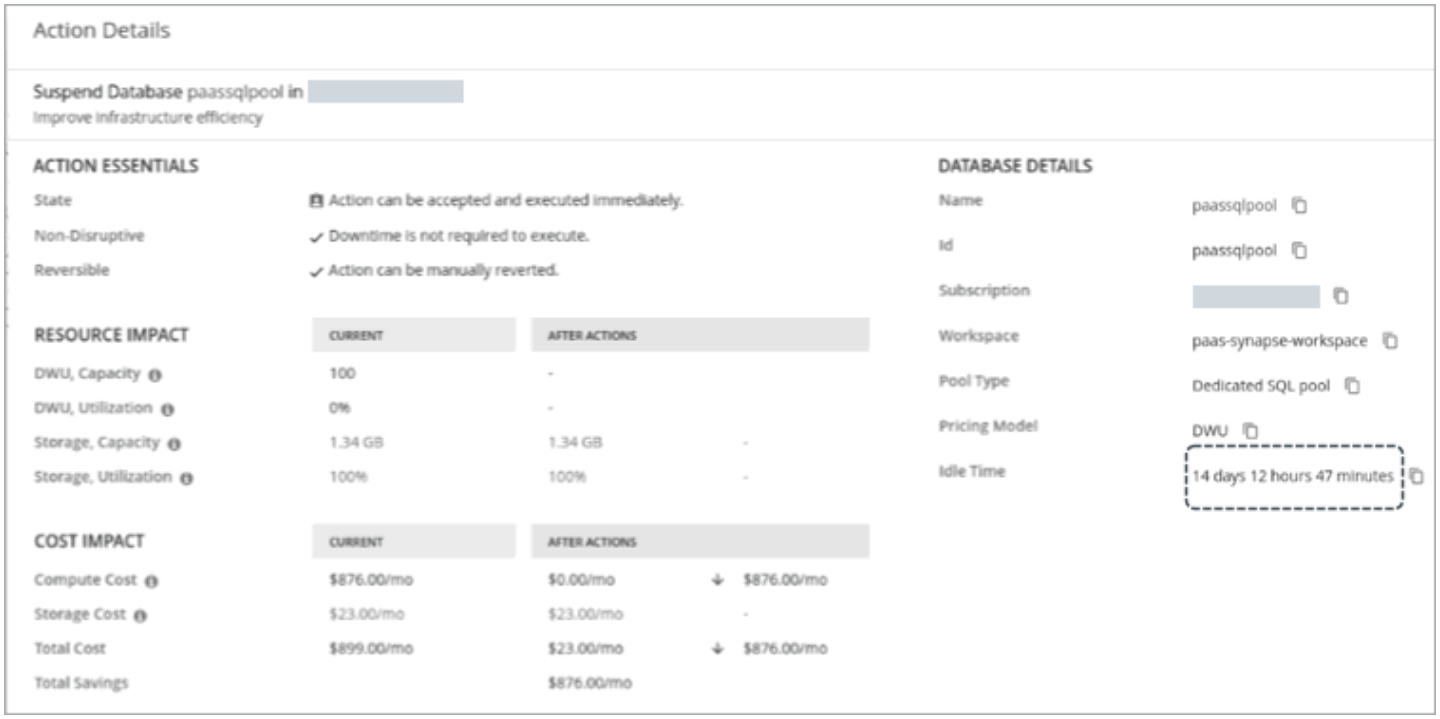
NOTE:
The Parking page *(on page 942)* lists all the dedicated SQL pools that Workload Optimization Manager discovered in your environment. From this page, you can stop a running or idle dedicated SQL pool, either on demand or according to a schedule. Stopping a pool has the same effect as suspending a pool.

The Parking page also provides options to start a stopped pool, either on demand or according to a schedule.

To calculate the current compute cost, Workload Optimization Manager retrieves the hourly on-demand cost from Azure and then multiplies the result by 730, which represents the number of hours per month that the product uses to estimate monthly costs.

If a currently idle pool is not suspended and is subsequently discovered as active, Workload Optimization Manager removes the suspend action attached to it.

Suspend actions include the 'Idle Time' information that indicates how long a dedicated SQL pool has been idle. By default, Workload Optimization Manager will generate suspend actions if dedicated SQL pools have been idle for at least one hour.



You can control the suspend actions that Workload Optimization Manager generates, based on your preferred idle time value.

For example, if you want Workload Optimization Manager to only generate suspend actions for dedicated SQL pools that have been idle for at least one day, perform these steps:

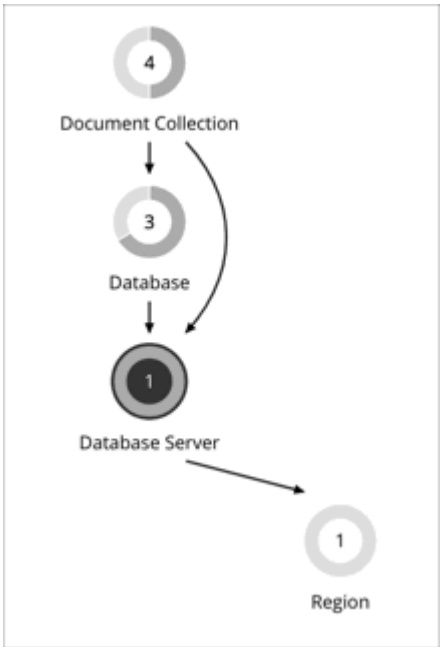
1. Create a dynamic group of Databases and set the 'Hours Idle' filter to `Hours Idle < = 24`.
2. Create a custom Database policy, set the scope to the group that you just created, and then *disable* suspend actions in that policy.

Database - Azure Cosmos DB

In Azure Cosmos DB *(on page 470)*, a database is a group of containers (document collections *(on page 722)*) and is similar to a namespace. Cosmos DB databases are managed through accounts *(on page 702)*.

Workload Optimization Manager discovers Cosmos DB databases through your Azure targets, and represents them as database entities in the supply chain.

Synopsis



Synopsis	
Provides:	Resources for database operations supported by Azure Cosmos DB
Consumes:	Request Units (RUs) from Azure Cosmos DB accounts
Discovered through:	Azure targets

Monitored resources

Workload Optimization Manager monitors the following resources:

- Request unit (RU)

  Request unit (RU) is a performance currency that abstracts CPU, IOPS, and memory that are required to perform the database operations supported by Azure Cosmos DB. Azure Cosmos DB normalizes the cost of all database operations using RUs.

Scale actions for Cosmos DB databases

Workload Optimization Manager can scale databases with provisioned throughput. It uses percentile calculations to measure RU utilization and then recommends actions to scale RUs up or down to optimize performance and costs.

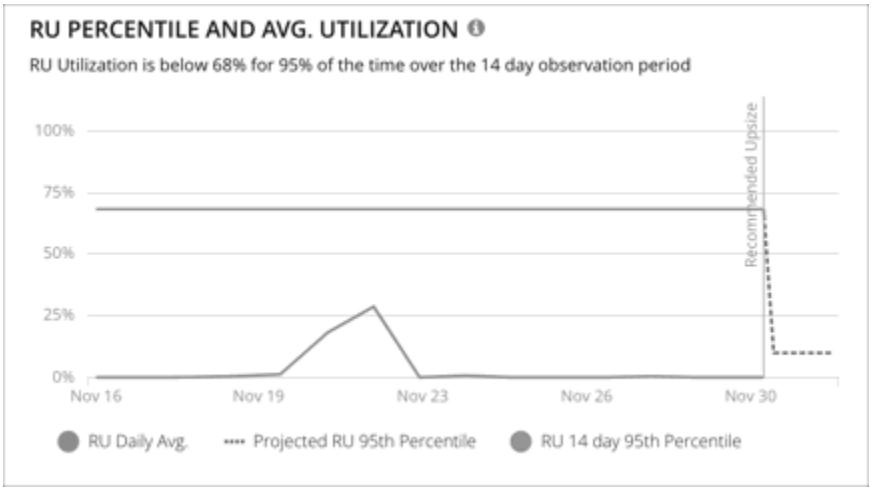
Points to consider:

- For scale down actions, Workload Optimization Manager will never recommend scaling below the minimum RU value calculated by Azure or configured in a database policy. For details, see Minimum RU Capacity for Scale Down Actions *(on page 719)*.
- Scale out actions do not apply since there is only one tier for Cosmos DB.
- Workload Optimization Manager discovers but does not scale databases without provisioned throughput.

Utilization charts for scale actions

Workload Optimization Manager uses percentile calculations to measure resource utilization, and drive scaling actions that improve overall utilization and reduce costs. When you examine the details for a pending scaling action on a database, you will

see charts that highlight resource *utilization percentiles* for a given observation period, and the projected percentiles after you execute the action.



The charts also plot *daily average utilization* for your reference. If you have previously executed scaling actions on the database, you can see the resulting improvements in daily average utilization. Put together, these charts allow you to easily recognize utilization trends that drive Workload Optimization Manager's scaling recommendations.

> **NOTE:**
> You can set scaling constraints in database policies to refine the percentile calculations. For details, see Aggressiveness and Observation Period *(on page 716)*.

Reconfigure actions for Cosmos DB databases

Workload Optimization Manager recommends removing unused throughput that is assigned to a database to help reduce your cloud expenses. To remove this resource, reconfigure the database from Azure. It is not possible to execute reconfigure actions from Workload Optimization Manager.

Workload Optimization Manager generates a reconfigure action if a database and all the underlying document collections have their own provisioned throughput. In this scenario, provisioned throughput at the database level is not used by any document collection and is therefore safe to remove. This action translates into cost savings. When you review a database reconfigure action in Action Center, the Cost Impact column indicates the cost savings that you would realize if you execute the action.

Points to consider:

- Workload Optimization Manager does not generate a reconfigure action if a database manages at least one document collection without provisioned throughput. In this scenario, the document collection is assumed to be using provisioned throughput at the database level, so it is not safe to remove that resource.
- Reconfigure actions are generated in recommend mode since it is not possible to execute these actions from Workload Optimization Manager.

  You can disable the generation of reconfigure actions in custom policies for select databases or the default database policy (for a global effect).

Delete actions for Cosmos DB databases

To help reduce your cloud expenses, Workload Optimization Manager recommends deleting a database with provisioned throughput but without any underlying document collection. When you review a database delete action in Action Center, the Cost Impact column indicates the cost savings that you would realize if you execute the action.

This action can be executed in Workload Optimization Manager manually or automatically. You can disable the generation of delete actions in custom policies for select databases or the default database policy (for a global effect).

Action disruptiveness and reversibility

The Action Center view and Action Details page indicate the disruptiveness and reversibility of actions for Cosmos DB databases.

- Scale actions

  All scale actions are non-disruptive and reversible.
- Reconfigure and delete actions

  All reconfigure and delete actions are disruptive and irreversible.

Cloud database policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

Action	Applies to	Default setting
Cloud DB scale	Azure SQL vCore and DTU	Manual
Database suspend actions	Azure dedicated SQL pool	On, manual
Database reconfigure actions	Azure Cosmos DB	On, recommend
Database delete actions	Azure Cosmos DB	On, manual

You can use Action Scripts *(on page 1070)* and third-party orchestrators (such as ServiceNow) for action orchestration.

Scaling sensitivity

Workload Optimization Manager uses a percentile of utilization over the specified observation period. This gives sustained utilization and ignores short-lived bursts.

Workload Optimization Manager uses these settings to calculate utilization percentiles for DTU, RU, and storage. It then recommends actions to improve utilization based on the observed values for a given time period.

- Aggressiveness

Attribute	Default value
Aggressiveness	95th Percentile

  When evaluating performance, Workload Optimization Manager considers resource utilization as a percentage of capacity. The utilization drives actions to scale the available capacity either up or down. To measure utilization, the analysis considers a given utilization percentile. For example, assume a 95th percentile. The percentile utilization is the highest value that 95% of the observed samples fall below. Compare that to average utilization, which is the average of *all* the observed samples.

  Using a percentile, Workload Optimization Manager can recommend more relevant actions. This is important in the cloud, so that analysis can better exploit the elasticity of the cloud. For scheduled policies, the more relevant actions will tend to remain viable when their execution is put off to a later time.

  For example, consider decisions to reduce capacity. Without using a percentile, Workload Optimization Manager never resizes below the recognized peak utilization. Assume utilization peaked at 100% just once. Without the benefit of a percentile, Workload Optimization Manager will not reduce resources for that database.

  With **Aggressiveness**, instead of using the single highest utilization value, Workload Optimization Manager uses the percentile you set. For the previous example, assume a single burst to 100%, but for 95% of the samples, utilization never exceeded 50%. If you set **Aggressiveness** to 95th Percentile, then Workload Optimization Manager can see this as an opportunity to reduce resource allocation.

In summary, a percentile evaluates the sustained resource utilization, and ignores bursts that occurred for a small portion of the samples. You can think of this as aggressiveness of resizing, as follows:

- 99th Percentile – More performance. Recommended for critical databases that need maximum guaranteed performance at all times, or those that need to tolerate sudden and previously unseen spikes in utilization, even though sustained utilization is low.
- 95th Percentile (Default) – The recommended setting to achieve maximum performance and savings. This assures performance while avoiding reactive peak sizing due to transient spikes, thus allowing you to take advantage of the elastic ability of the cloud.
- 90th Percentile – More efficiency. Recommended for databases that can stand higher resource utilization.

By default, Workload Optimization Manager uses samples from the last 14 days. Use the **Max Observation Period** setting to adjust the number of days.

- Max observation period

Attribute	Default value
Max Observation Period	Last 14 Days

To refine the calculation of resource utilization percentiles, you can set the sample time to consider. Workload Optimization Manager uses historical data from up to the number of days that you specify as a sample period. If the database has fewer days' data then it uses all of the stored historical data.

You can make the following settings:

- Less Elastic – Last 30 Days
- Recommended – Last 14 Days
- More Elastic – Last 7 Days or Last 3 Days

Workload Optimization Manager recommends an observation period of 14 days so it can recommend scaling actions more often. Since Azure SQL DB scaling is minimally disruptive, with near-zero downtime, scaling often should not introduce any noticeable performance risks.

NOTE:
For more information about Azure scaling downtimes, see the Azure documentation.

- Min observation period

Attribute	Default value
Min Observation Period	None

This setting ensures historical data for a minimum number of days before Workload Optimization Manager will generate an action based on the percentile set in **Aggressiveness**. This ensures a minimum set of data points before it generates the action.

Especially for scheduled actions, it is important that resize calculations use enough historical data to generate actions that will remain viable even during a scheduled maintenance window. A maintenance window is usually set for "down" time, when utilization is low. If analysis uses enough historical data for an action, then the action is more likely to remain viable during the maintenance window.

- More Elastic – None
- Less Elastic – 7 Days

Cloud instance types

By default, Workload Optimization Manager considers all instance types currently available for scaling when making scaling decisions for SQL databases. However, you may have set up your SQL databases to only scale to certain instance types to reduce complexity and cost, or meet demand. Use this setting to identify those instance types.

Attribute	Default value
Cloud Instance Types	None

Click **Edit** to set your preferences. In the new page that displays, choose from the following options:

- Excluded

  By default, this option is selected and all instance types are deselected. This means that all instance types are considered for scaling.

  Any instance types that you select are excluded from scaling.

  Configure an exclusion list if there are fewer instance types to exclude than to include. For example, if 400 instance types are currently available for scaling and you do not want SQL Databases to scale to 20 instance types, configure an exclusion list and select the 20 instance types.

- Included

  Any instance types that you select are considered for scaling.

  Configure an inclusion list if there are fewer instance types to include than to exclude. For example, if 400 instance types are currently available for scaling and you want SQL Databases to scale only to 50 instance types, configure an inclusion list and select the 50 instance types.

Expand an instance family to see individual instance types and the resources allocated to them.

After you save your changes, the main page refreshes to reflect your selections.

NOTE:
This policy setting is not available in plans.

Scaling target utilization

The utilization that you set here specifies the percentage of the existing capacity that Workload Optimization Manager will consider to be 100% of capacity.

The settings you make depend on the pricing model in place for the workloads in the policy scope.

- vCore pricing model

  To meet individual VCPU, VMEM, or IOPS/Throughput targets, the workloads must be charged according to the vCore pricing model.

Attribute	Default value
VCPU	70
VMEM	90
IOPS/Throughput	70
Storage Amount Utilization	90

- DTU pricing model

  To meet a target DTU utilization, the workloads must be charged according to the DTU pricing model.

Attribute	Default value
DTU Utilization	70
Storage Amount Utilization	90

- RU pricing model

To meet a target RU utilization or individual RU targets, the workloads must be charged according to the RU pricing model.

Attribute	Default value
RU Utilization	70

These advanced settings determine how much you would like a scope of workloads to utilize their resources. These are fixed settings that override the way Workload Optimization Manager calculates the optimal utilization of resources. You should only change these settings after consulting with Technical Support.

While these settings offer a way to modify how Workload Optimization Manager recommends actions, in most cases you should never need to use them. If you want to control how Workload Optimization Manager recommends actions to resize workloads, you can set the aggressiveness per the percentile of utilization, and set the length of the sample period for more or less elasticity on the cloud.

Minimum RU capacity for scale down actions

This setting applies only to Cosmos DB databases. Workload Optimization Manager checks this scaling constraint when it recommends scale down actions to ensure that Cosmos DB databases never scale below the specified value.

Attribute	Default value
Minimum RU capacity for scale down actions	400

Be sure to specify a value that is an increment of 100. By design, Azure increases or decreases the number of RUs at any time in increments or decrements of 100 RUs.

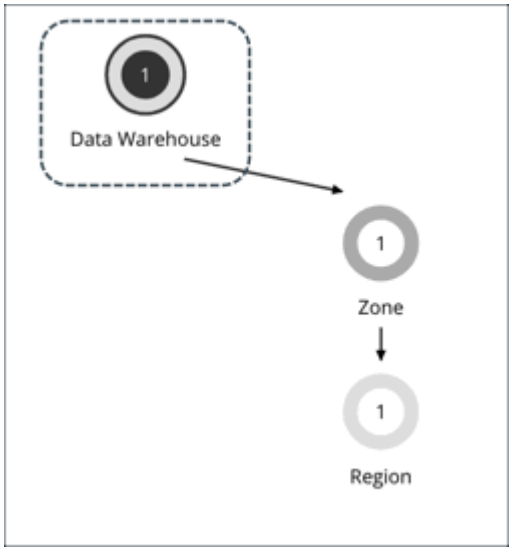
In addition to this scaling constraint, Workload Optimization Manager also checks the minimum RU limit that Azure calculated. This limit takes precedence over the value specified in the policy. For example, if the minimum RU limit calculated by Azure is 500 while the minimum RU capacity in the database policy is 400, Workload Optimization Manager will never recommend scaling below 500 RUs.

Data warehouse

In AWS, Amazon Redshift is a data warehouse service that analyzes your data using your existing business intelligence tools. Workload Optimization Manager discovers Redshift **provisioned clusters** through your AWS targets, and represents them as data warehouse entities in the supply chain.

NOTE:
Workload Optimization Manager does not discover Redshift Serverless.

Synopsis



Synopsis	
Provides:	Data warehouse services
Consumes:	Resources from availability zones
Discovered through:	AWS targets

Monitored resources

Workload Optimization Manager monitors the following resources:

- Number of replicas

  Number of replicas is the total number of nodes (leader node and compute nodes) in a Redshift provisioned cluster.
- Storage amount

  Storage amount is the measurement of storage capacity that is in use.

  NOTE:
  Workload Optimization Manager calculates storage amount. The calculated amount may not match the data returned by Redshift due to an AWS issue. For this reason, you may see a slight difference in the storage amounts shown in Workload Optimization Manager and in the AWS console.

- Connection

  Connection is the measurement of database connections utilized by applications.

Suspend actions for Redshift provisioned clusters

When Workload Optimization Manager analysis discovers an idle Redshift provisioned cluster (a cluster with no active connections for a specific period of time), it recommends suspending the cluster as a cost-saving measure and shows the action in Action Center. This action suspends the compute nodes in the cluster, but does not impact data stored on the cluster. In the Action Details page for a specific action, the compute cost is 0 (zero) after suspension. However, you may continue to incur storage cost for certain instance families. For more information about storage cost for instance families, see this AWS page.

To calculate the current compute cost, Workload Optimization Manager retrieves the hourly on-demand cost from AWS and then multiplies the result by 730, which represents the number of hours per month that the product uses to estimate monthly

costs. Note that for a multi-AZ deployment of a provisioned cluster, the compute cost is double the rate of a single-AZ deployment.

If a currently idle provisioned cluster is not suspended and is subsequently discovered as active, Workload Optimization Manager removes the suspend action attached to it.

Suspend actions include the 'Idle time' information that indicates how long a provisioned cluster has been idle. By default, Workload Optimization Manager generates suspend actions if a provisioned cluster has been idle for at least 24 hours.



You can control the suspend actions that Workload Optimization Manager generates, based on your preferred idle time value.

For example, if you want Workload Optimization Manager to only generate suspend actions for provisioned clusters that have been idle for at least two days, perform these steps:

1. Create a dynamic group of data warehouses and set the 'Hours Idle' filter to `Hours Idle < = 48`.
2. Create a custom data warehouse policy, set the scope to the group that you just created, and then *disable* suspend actions in that policy.

Action disruptiveness and reversibility

All suspend actions shown in the Action Center view and Action Details page are disruptive but reversible.

Data warehouse policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

Action	Default setting
Suspend	On, manual

You can use this setting to disable suspend actions or change the action mode.

Suspend recommendation

Action	Default setting
Suspension threshold	1 hour

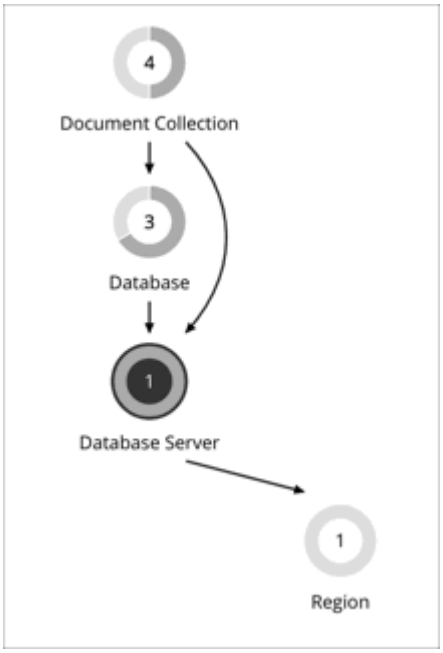
By default, Workload Optimization Manager recommends suspending Redshift provisioned clusters that have been idle for 24 hours. You can change the idle time value according to your requirements.

Document collection

In Azure Cosmos DB *(on page 470)*, containers are entities that store data on one or more partitions. Containers are grouped using databases *(on page 713)* and managed through accounts *(on page 702)*.

Workload Optimization Manager discovers Azure Cosmos DB containers through your Azure targets, and represents them as document collection entities in the supply chain.

Synopsis



Synopsis	
Provides:	Resources for database operations supported by Azure Cosmos DB
Consumes:	Request Units (RUs) from Azure Cosmos DB accounts
Discovered through:	Azure targets

Monitored resources

Workload Optimization Manager monitors the following resources:

- Request unit (RU)

  Request unit (RU) is a performance currency that abstracts CPU, IOPS, and memory that are required to perform the database operations supported by Azure Cosmos DB. Azure Cosmos DB normalizes the cost of all database operations using RUs.

Scale actions for document collections

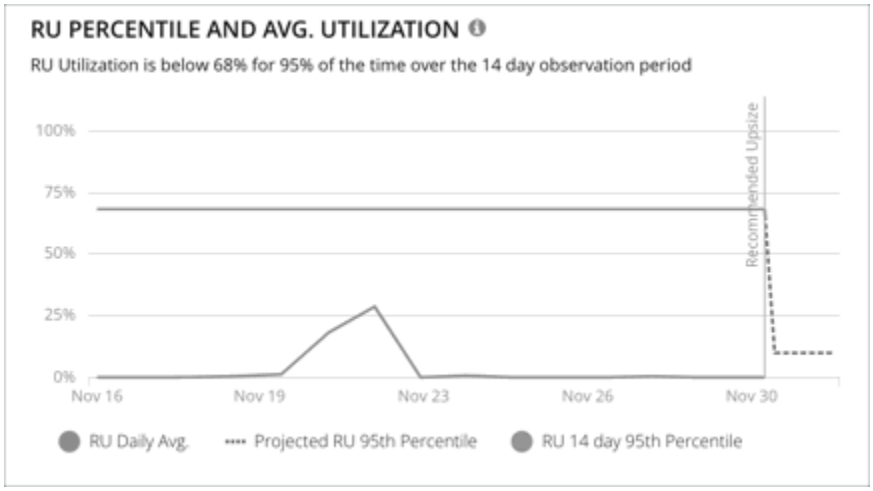
Workload Optimization Manager can scale document collections with standard (manually provisioned) dedicated throughput. It uses percentile calculations to measure RU utilization and then recommends actions to scale RUs up or down to optimize performance and costs.

Points to consider:

- For scale down actions, Workload Optimization Manager will never recommend scaling below the minimum RU value calculated by Azure or configured in a document collection policy. For details, see Minimum RU Capacity for Scale Down Actions *(on page 726)*.
- Scale out actions do not apply since there is only one tier for Cosmos DB.
- Workload Optimization Manager discovers but does not scale the following entities:
  - Document collections with auto-scaled dedicated throughput
  - Document collections with shared throughput, where throughput is provisioned at the database level and then shared by document collections within the database

Utilization charts for scale actions

Workload Optimization Manager uses percentile calculations to measure resource utilization, and drive scaling actions that improve overall utilization and reduce costs. When you examine the details for a pending scaling action on a document collection, you will see charts that highlight resource *utilization percentiles* for a given observation period, and the projected percentiles after you execute the action.



The charts also plot *daily average utilization* for your reference. If you have previously executed scaling actions on the document collection, you can see the resulting improvements in daily average utilization. Put together, these charts allow you to easily recognize utilization trends that drive Workload Optimization Manager's scaling recommendations.

> **NOTE:**
> You can set scaling constraints in document collection policies to refine the percentile calculations. For details, see Aggressiveness and Observation Period *(on page 724)*.

Non-disruptive and reversible scaling actions

All scaling actions shown in the Action Center view and Action Details page are non-disruptive and reversible.

Document collection policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Action automation and orchestration

For details about cloud document collection actions, see Document Collection Actions *(on page 723)*.

Action	Default setting
Cloud Document Collection Scale	Manual

You can use Action Scripts *(on page 1070)* and third-party orchestrators (such as ServiceNow) for action orchestration.

Scaling sensitivity

Workload Optimization Manager uses a percentile of utilization over the specified observation period. This gives sustained utilization and ignores short-lived bursts.

Workload Optimization Manager uses these settings to calculate utilization percentiles for Request Units (RUs). It then recommends actions to improve utilization based on the observed values for a given time period.

- Aggressiveness

Attribute	Default value
Aggressiveness	95th Percentile

When evaluating performance, Workload Optimization Manager considers resource utilization as a percentage of capacity. The utilization drives actions to scale the available capacity either up or down. To measure utilization, the analysis considers a given utilization percentile. For example, assume a 95th percentile. The percentile utilization is the highest value that 95% of the observed samples fall below. Compare that to average utilization, which is the average of *all* the observed samples.

Using a percentile, Workload Optimization Manager can recommend more relevant actions. This is important in the cloud, so that analysis can better exploit the elasticity of the cloud. For scheduled policies, the more relevant actions will tend to remain viable when their execution is put off to a later time.

For example, consider decisions to reduce capacity. Without using a percentile, Workload Optimization Manager never resizes below the recognized peak utilization. Assume utilization peaked at 100% just once. Without the benefit of a percentile, Workload Optimization Manager will not reduce resources for that document collection.

With **Aggressiveness**, instead of using the single highest utilization value, Workload Optimization Manager uses the percentile you set. For the previous example, assume a single burst to 100%, but for 95% of the samples, utilization never exceeded 50%. If you set **Aggressiveness** to 95th Percentile, then Workload Optimization Manager can see this as an opportunity to reduce resource allocation.

In summary, a percentile evaluates the sustained resource utilization, and ignores bursts that occurred for a small portion of the samples. You can think of this as aggressiveness of resizing, as follows:

- 99th Percentile – More performance. Recommended for critical document collections that need maximum guaranteed performance at all times, or those that need to tolerate sudden and previously unseen spikes in utilization, even though sustained utilization is low.
- 95th Percentile (Default) – The recommended setting to achieve maximum performance and savings. This assures performance while avoiding reactive peak sizing due to transient spikes, thus allowing you to take advantage of the elastic ability of the cloud.
- 90th Percentile – More efficiency. Recommended for document collections that can stand higher resource utilization.

By default, Workload Optimization Manager uses samples from the last 14 days. Use the **Max Observation Period** setting to adjust the number of days.

- Max observation period

Attribute	Default value
Max Observation Period	Last 14 Days

To refine the calculation of resource utilization percentiles, you can set the sample time to consider. Workload Optimization Manager uses historical data from up to the number of days that you specify as a sample period. If the document collection has fewer days' data then it uses all of the stored historical data.

You can make the following settings:

  - Less Elastic – Last 30 Days
  - Recommended – Last 14 Days
  - More Elastic – Last 7 Days or Last 3 Days

Workload Optimization Manager recommends an observation period of 14 days so it can recommend scaling actions more often. Since Azure SQL DB scaling is minimally disruptive, with near-zero downtime, scaling often should not introduce any noticeable performance risks.

NOTE:
For more information about Azure scaling downtimes, see the Azure documentation.

- Min observation period

Attribute	Default value
Min Observation Period	None

This setting ensures historical data for a minimum number of days before Workload Optimization Manager will generate an action based on the percentile set in **Aggressiveness**. This ensures a minimum set of data points before it generates the action.

Especially for scheduled actions, it is important that resize calculations use enough historical data to generate actions that will remain viable even during a scheduled maintenance window. A maintenance window is usually set for "down" time, when utilization is low. If analysis uses enough historical data for an action, then the action is more likely to remain viable during the maintenance window.

  - More Elastic – None
  - Less Elastic – 7 Days

Scaling target utilization

The utilization that you set here specifies the percentage of the existing capacity that Workload Optimization Manager will consider to be 100% of capacity.

To meet a target RU utilization or individual RU targets, the workloads must be charged according to the RU pricing model.

Attribute	Default value
RU Utilization	70

These advanced settings determine how much you would like a scope of workloads to utilize their resources. These are fixed settings that override the way Workload Optimization Manager calculates the optimal utilization of resources. You should only change these settings after consulting with Technical Support.

While these settings offer a way to modify how Workload Optimization Manager recommends actions, in most cases you should never need to use them. If you want to control how Workload Optimization Manager recommends actions to resize workloads, you can set the aggressiveness per the percentile of utilization, and set the length of the sample period for more or less elasticity on the cloud.

Minimum RU capacity for scale down actions

Workload Optimization Manager checks this scaling constraint when it recommends scale down actions to ensure that document collections never scale below the specified value.

Attribute	Default value
Minimum RU capacity for scale down actions	400

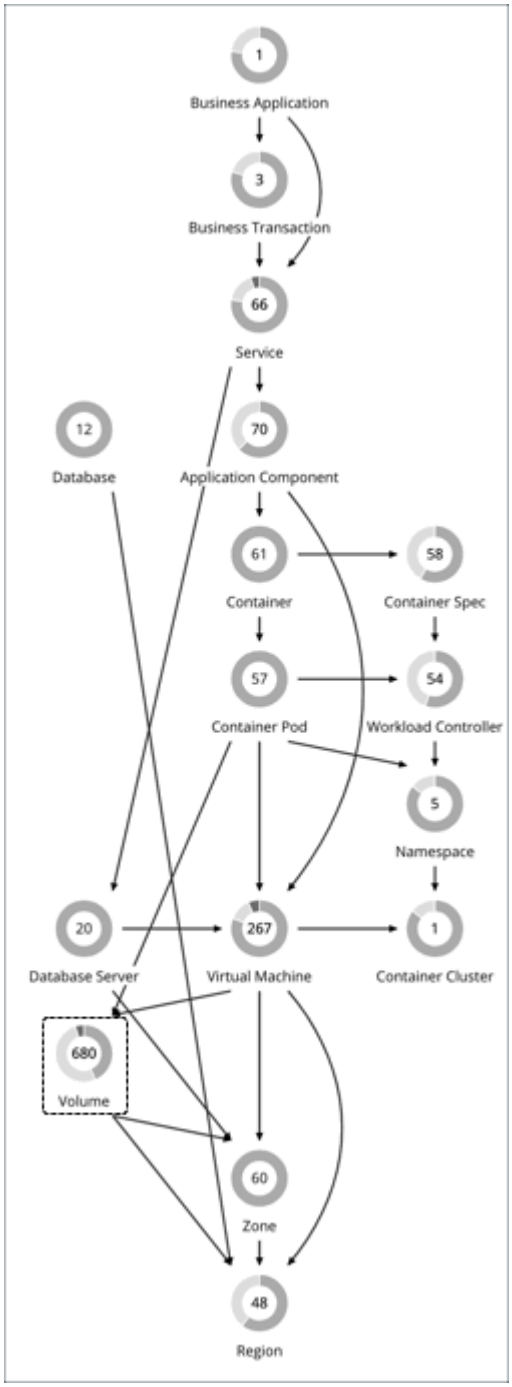
Be sure to specify a value that is an increment of 100. By design, Azure increases or decreases the number of RUs at any time in increments or decrements of 100 RUs.

In addition to this scaling constraint, Workload Optimization Manager also checks the minimum RU limit that Azure calculated. This limit takes precedence over the value specified in the policy. For example, if the minimum RU limit calculated by Azure is 500 while the minimum RU capacity in the document collection policy is 400, Workload Optimization Manager will never recommend scaling below 500 RUs.

Volume (cloud)

A cloud volume is a storage device that you can attach to a VM. You can use an attached volume the same as a physical hard drive.

Synopsis



Synopsis	
Provides:	Storage resources for VMs to use, including: ▪ Storage Access ▪ Storage Amount ▪ IO Throughput
Consumes:	Storage services provided by zones or regions

Synopsis	
Discovered through:	Public cloud targets

Monitored resources

Workload Optimization Manager monitors the following resources:

- Storage amount

  Storage Amount is the storage capacity (disk size) of a volume.

  Workload Optimization Manager discovers Storage Amount, but does not monitor utilization.

  For a Kubeturbo (container) deployment that includes volumes, Kubeturbo monitors Storage Amount utilization for the volumes. You can view utilization information in the Capacity and Usage chart.

- Storage access (IOPS)

  Storage access, also known as IOPS, is the measurement of IOPS capacity that is in use.

- I/O throughput

  I/O throughput is the measurement of I/O throughput capacity that is in use.

NOTE:
Workload Optimization Manager also monitors the attachment state of volumes and then generates delete actions for unattached volumes.

Actions

Workload Optimization Manager supports the following actions:

- Scale

  Scale attached volumes to optimize performance and costs.

- Delete

  Delete unattached volumes as a cost-saving measure. Workload Optimization Manager generates an action immediately after discovering an unattached volume.

Scale actions

Scale attached volumes to optimize performance and costs.

Workload Optimization Manager can recommend:

- Scaling within the same tier (scale up or down), or from one tier to another

  Examples:

  - An action to scale down IOPS for a high performance volume, such as Azure Managed Ultra
  - An action to scale a volume from the *io1* to the *gp2* tier

  These actions can reduce costs significantly while continuing to assure performance. In addition, they are *non-disruptive* and *reversible*.

  NOTE:
  For details about action disruptiveness and reversibility, see Non-disruptive and Reversible Scaling Actions *(on page 733)*.

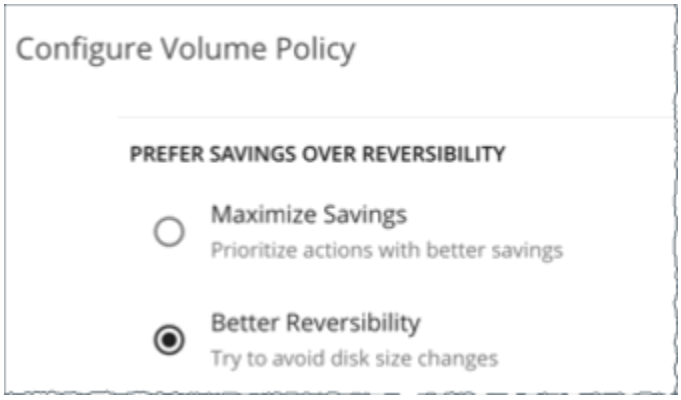
- Scaling from one tier to another and then within the new tier, in a single action

For example, to achieve higher IOPS performance for VMs that are premium storage capable, you might see an action to scale the corresponding volume from *Standard* to *Premium*, and then scale up volume capacity from *32GB* to *256 GB*. This action increases costs and is *irreversible*, but is more cost effective than scaling up within the original tier.

NOTE:
Not all VM types are premium storage capable. For details, see the Azure Documentation.

Points to consider:

- When there are multiple disks attached to a volume, every volume scale action can potentially disrupt the same VM multiple times and some of the actions may fail due to concurrency. To mitigate these issues, Workload Optimization Manager identifies all volume actions associated with a particular VM and combines them into a single unit for execution, thus reducing disruptions and the chance of failures due to concurrency. This approach applies to scale actions in *Manual* or *Automated* mode.
- You can create policies to control the scaling actions that Workload Optimization Manager generates.
  - Workload Optimization Manager includes a policy setting that lets you choose between two outcomes – better savings (default) and better reversibility. If you choose reversibility, which can increase costs, Workload Optimization Manager will prioritize actions to change tiers whenever possible.



  - Set scaling constraints if you want volumes to *only scale to* or *avoid* certain tiers. For details, see Cloud Storage Tiers *(on page 736)*.

Delete actions

Delete unattached volumes as a cost-saving measure.

Points to consider:

- If you delete an Azure volume and then later deploy a new one with an identical name, charts will include historical data from the volume that you deleted.
- For volumes that have been deleted from the cloud provider and are no longer discoverable, Workload Optimization Manager removes them from its records after 15 days.

Unattached Period for Volumes

After you add a cloud target, Workload Optimization Manager performs its initial discovery and then records the number of days volumes have been unattached (unattached period). When you review a pending volume delete action in Action Center, the **Days Unattached** column shows the unattached period. To see the last known VM attached to the volume, click **Details**.

The unattached period shown during the initial discovery varies by cloud provider.

- For Google Cloud, the 'Days Unattached' column shows the exact number of days a volume has been unattached.
- For AWS and Azure, there are limits to the amount of historical data that Workload Optimization Manager can retrieve.
  - For AWS, Workload Optimization Manager can retrieve 90 days of historical data. If a volume has been unattached for 90 days or fewer, the 'Days Unattached' column shows the exact number of days. Otherwise, the column shows 90 days.
  - For Azure, Workload Optimization Manager can retrieve 14 days of historical data. If a volume has been unattached for 14 days or fewer, the 'Days Unattached' column shows the exact number of days. Otherwise, the column shows 14 days.

NOTE:
In rare cases where Workload Optimization Manager is unable to discover the unattached period from a given cloud provider, it treats unattached volumes as newly unattached. For these volumes, the 'Days Unattached' column shows a value of 0.

After the initial discovery, Workload Optimization Manager polls your volumes with each discovery cycle and then records their state (attached or unattached) at the time of polling. It does not account for changes in state between polls. If an unattached volume in the previous poll is not deleted, there are two possible scenarios in the next poll:

- The volume remains unattached. In this case, Workload Optimization Manager updates the unattached period.
- The volume is now attached. In this case, Workload Optimization Manager updates the volume's state.

Generation of Delete Actions for Unattached Volumes

As Workload Optimization Manager discovers unattached AWS, Azure, or Google Cloud volumes, it waits an hour before generating delete actions.

NOTE:
For AWS, the waiting period can sometimes exceed one hour and can be up to 24 hours due to limitations with AWS APIs.

To change the default waiting period of one hour, contact your Workload Optimization Manager representative for assistance.

For a subset of volumes in your environment, you can create volume groups and policies to override the default waiting periods. For example, if you want Workload Optimization Manager to wait two days before generating delete actions for unattached volumes in your production environment, perform these steps:

1. Navigate to **Settings > Groups**, create a group that includes all the volumes in your production environment, and then set the 'Days Unattached' filter to `Days Unattached < 2`.
2. Navigate to **Settings > Policies > New Policy > Automation Policy > Volume**, set the scope to the group that you just created, and then disable delete actions.

Exceptions for Azure Site Recovery and Azure Backup Volumes

Workload Optimization Manager discovers Azure Site Recovery and Azure Backup volumes when you add Azure targets. Even though these volumes are always unattached, Workload Optimization Manager will never recommend deleting them because they are critical to business continuity and disaster recovery.

To identify Azure Site Recovery volumes, Workload Optimization Manager checks an Azure resource called Recovery Services vault, which includes information specific to the volumes. It also checks for the ASR-`ReplicaDisk` tag, which Azure automatically assigns to the volumes.

For Azure Backup volumes, Workload Optimization Manager checks for the RSV`aultBackup` tag.

It is important that you do not remove these tags. If these tags have been removed for some reason, create a volume policy for the affected volumes and disable delete actions in that policy.

Action execution for locked volumes

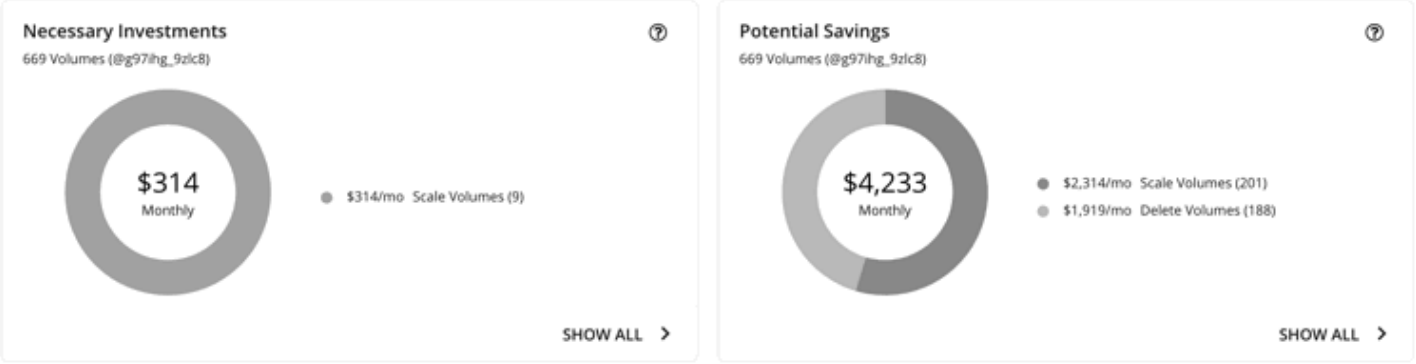
For Azure environments, Workload Optimization Manager can recommend scale and delete actions for locked volumes, but the lock level configured for the volumes may prevent some actions from executing.

- For volumes with the `ReadOnly` lock, both scale and delete actions are *not* executable.
- For volumes with the `CanNotDelete` lock, delete actions are *not* executable, but scale actions are executable.

You must sign in to Azure and then remove the locks for the affected volumes before you can execute actions. To identify the specific locks that you need to remove, open the Action Details page for a pending volume action and see the **Execution Prerequisites** section.

Volume actions in charts

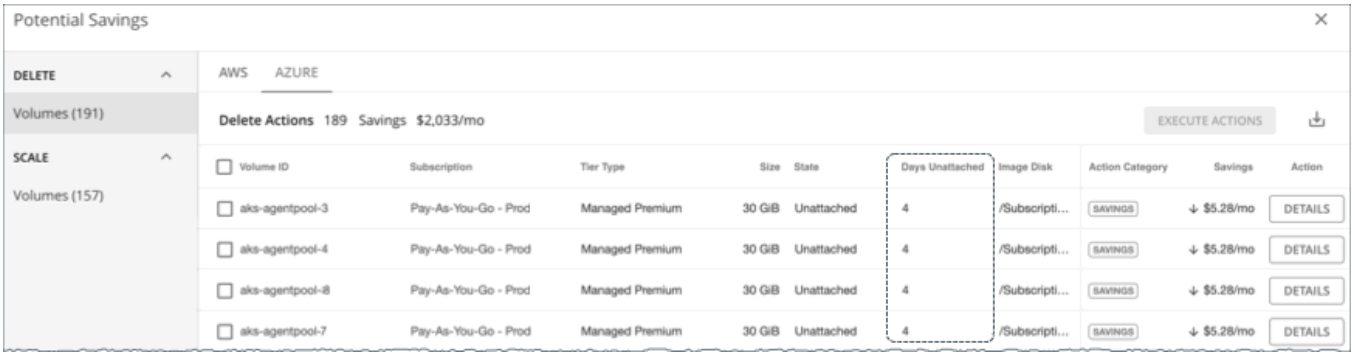
Use the Necessary Investments and Potential Savings charts to view pending volume actions. These charts show total monthly investments and savings, assuming you execute all the actions.



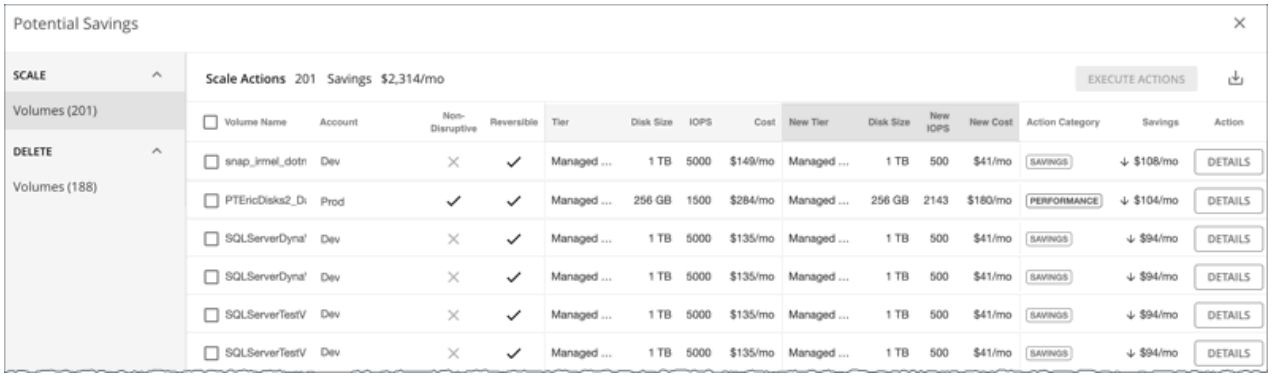
Click **Show All** for each chart to review and execute the actions.

The table shows the following:

- Actions that are pending for each volume
- Savings or investments for each volume
- For *Delete* actions in the Potential Savings chart:



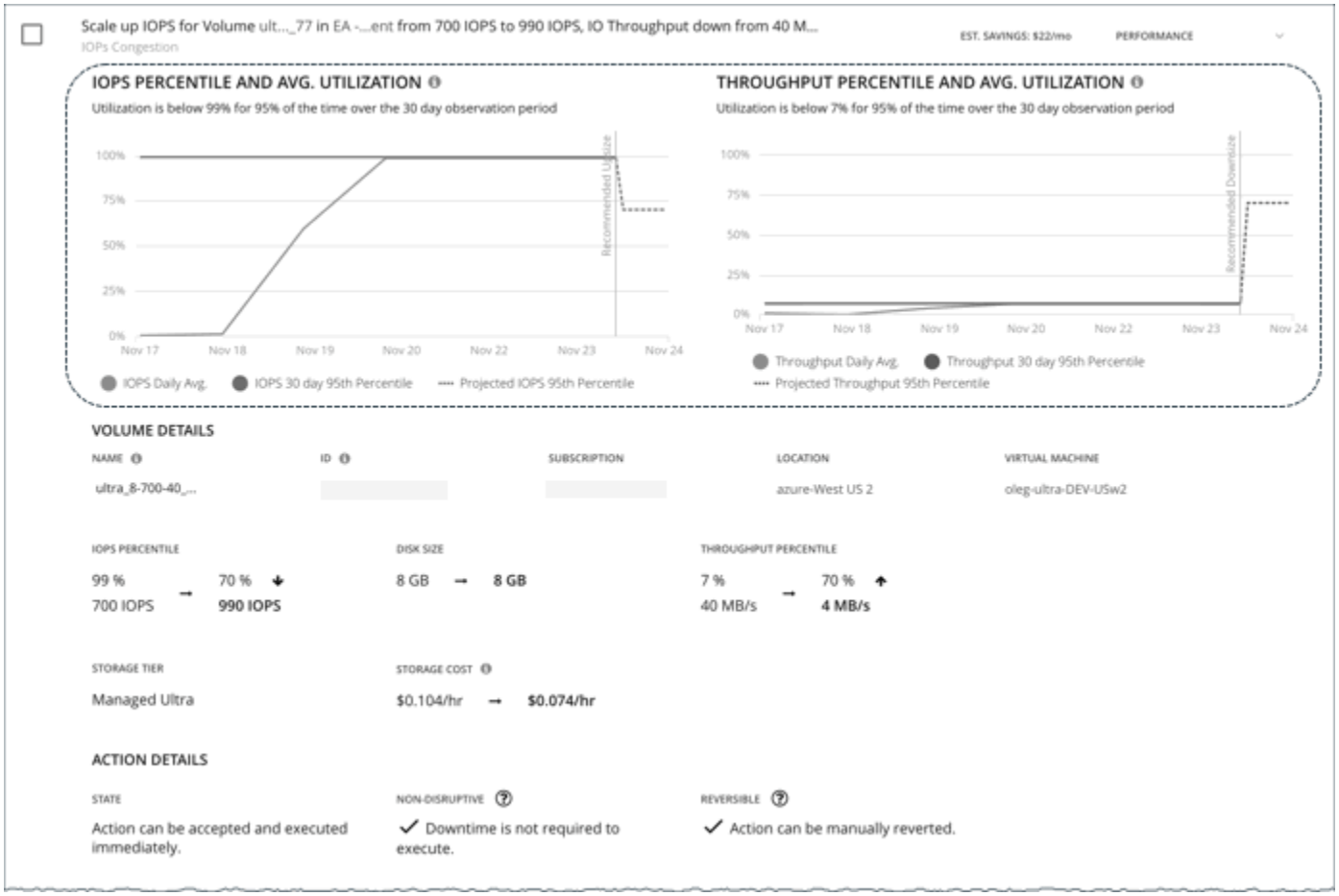
  - Number of days a volume has been unattached
  - For *Scale* actions in the Potential Savings or Necessary Investments chart:



- Whether actions are non-disruptive or reversible
- Changes the actions will effect (for example, changes in tiers and/or resource allocations)

When you click the **DETAILS** button for a scaling action, you will see utilization charts that help explain the reason for the action.

Utilization charts for volume scaling actions



Workload Optimization Manager uses percentile calculations to measure IOPS and throughput more accurately, and drive scaling actions that improve overall utilization and reduce costs. When you examine the details for a pending scaling action on a volume, you will see charts that highlight resource *utilization percentiles* for a given observation period, and the projected percentiles after you execute the action.

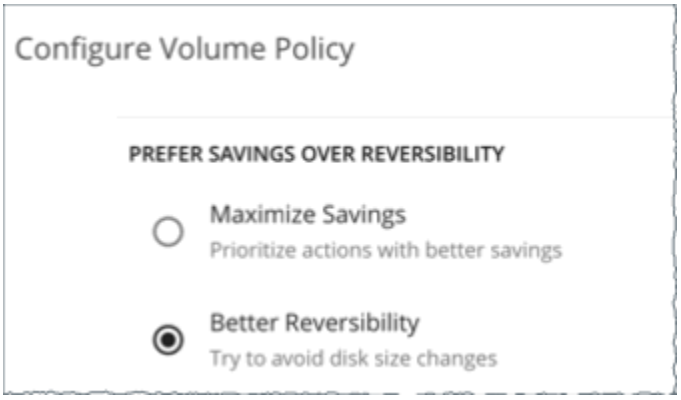
The charts also plot *daily average utilization* for your reference. If you have previously executed scaling actions on the volume, you can see the resulting improvements in daily average utilization. Put together, these charts allow you to easily recognize utilization trends that drive Workload Optimization Manager's scaling recommendations.

> **NOTE:**
> You can set scaling sensitivity in volume policies to refine the percentile calculations. For details, see Scaling Sensitivity *(on page 734)*.

Disruptiveness and reversibility of volume scaling actions

Scaling actions can sometimes be disruptive and/or irreversible.

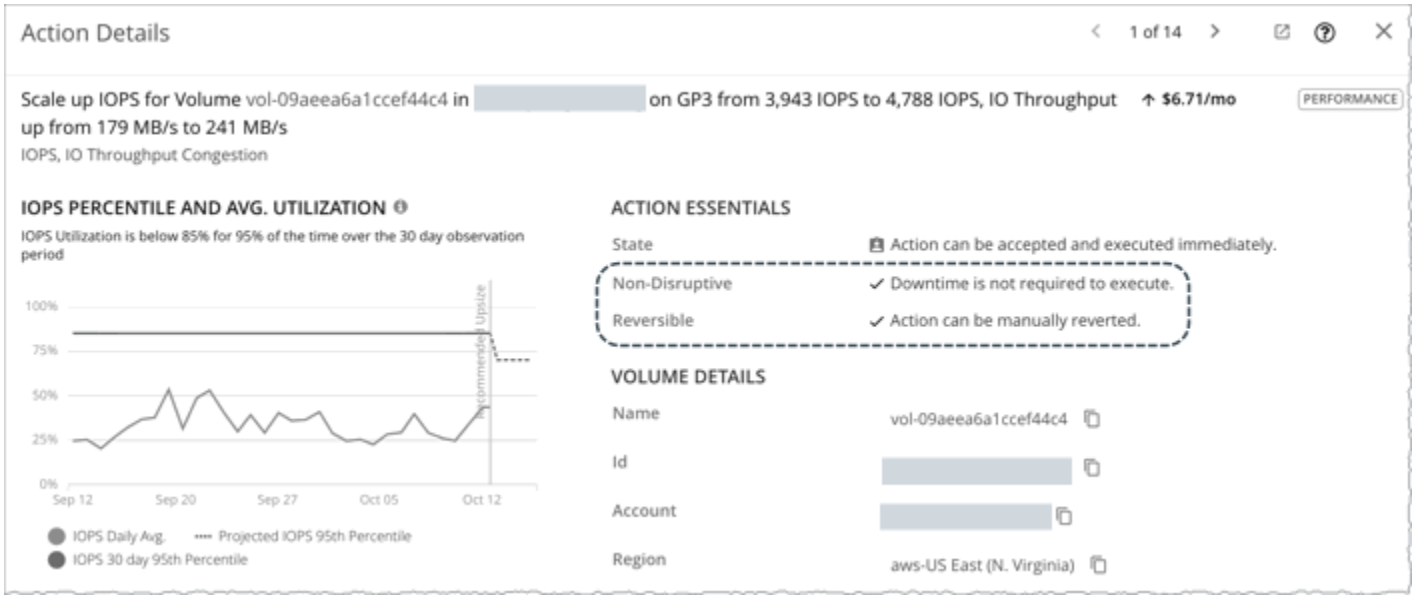
- Scaling actions are *disruptive* if the cloud provider must reboot a VM to execute a storage change. When a storage action is disruptive, expect a single reboot that usually results in 2 to 3 minutes of downtime.
- Scaling actions are irreversible if the storage must grow in size to subsequently increase IOPS or throughput capacity. In this case, shrinking the storage size later would not be possible. If you prefer reversible volume actions, create a volume policy and choose **Better Reversibility**.



The following table describes the disruptiveness and reversibility of the scaling actions that Workload Optimization Manager recommends.

Nature of Scaling Action	Applicable Actions
Non-disruptive and reversible	▪ All scaling actions for AWS storage ▪ IOPS or throughput scaling for Azure ultra disks ▪ IOPS scaling for Google Cloud Hyperdisk Extreme ▪ Throughput scaling for Google Cloud Hyperdisk Throughput
Disruptive but reversible	▪ All scaling actions for Azure storage, except IOPS or throughput scaling for ultra storage ▪ Storage tier changes for Google Cloud workloads ▪ IOPS scaling for Google Cloud extreme persistent disks
Non-disruptive but irreversible	Disk size scaling for AWS and Google Cloud storage
Disruptive and irreversible	Disk size scaling for Azure storage

The action details for a pending volume scaling action indicates whether the action is non-disruptive or reversible.



Cloud volume policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

For details about cloud volume actions, see Cloud Volume Actions *(on page 728)*.

Action	Default mode
Scale ▪ Non-disruptive reversible scaling ▪ Disruptive reversible scaling ▪ Disruptive irreversible scaling For details about action disruptiveness and reversibility, see this topic *(on page 733)*.	Manual
Delete	Manual

Prefer Savings Over Reversibility

Executing storage scaling actions can sometimes be irreversible if the volume must grow in size to subsequently increase IOPS or throughput capacity. In this case, shrinking that volume's size later would not be possible. If you prefer reversible volume actions, create a volume policy and choose **Better Reversibility**.

Scaling sensitivity

Workload Optimization Manager uses a percentile of utilization over the specified observation period. This gives sustained utilization and ignores short-lived bursts.

Workload Optimization Manager uses these settings to calculate utilization percentiles for IOPS and throughput. It then recommends actions to improve utilization based on the observed values for a given time period.

- Aggressiveness

Attribute	Default value
Aggressiveness	95th Percentile

Workload Optimization Manager uses Aggressiveness when evaluating IOPS and throughput.

When evaluating performance, Workload Optimization Manager considers resource utilization as a percentage of capacity. The utilization drives actions to scale the available capacity either up or down. To measure utilization, the analysis considers a given utilization percentile. For example, assume a 95th percentile. The percentile utilization is the highest value that 95% of the observed samples fall below. Compare that to average utilization, which is the average of *all* the observed samples.

Using a percentile, Workload Optimization Manager can recommend more relevant actions. This is important in the cloud, so that analysis can better exploit the elasticity of the cloud. For scheduled policies, the more relevant actions will tend to remain viable when their execution is put off to a later time.

For example, consider decisions to reduce capacity. Without using a percentile, Workload Optimization Manager never scales below the recognized peak utilization. Assume utilization peaked at 100% just once. Without the benefit of a percentile, Workload Optimization Manager will not reduce resources for that volume.

With **Aggressiveness**, instead of using the single highest utilization value, Workload Optimization Manager uses the percentile you set. For the previous example, assume a single burst to 100%, but for 95% of the samples, utilization never exceeded 50%. If you set **Aggressiveness** to 95th Percentile, then Workload Optimization Manager can see this as an opportunity to reduce resource allocation.

In summary, a percentile evaluates the sustained resource utilization, and ignores bursts that occurred for a small portion of the samples. You can think of this as aggressiveness of resizing, as follows:

  - 99th or 100th Percentile – More performance. Recommended for critical volumes that need maximum guaranteed performance at all times, or those that need to tolerate sudden and previously unseen spikes in utilization, even though sustained utilization is low.
  - 95th Percentile (Default) – The recommended setting to achieve maximum performance and savings. This assures performance while avoiding reactive peak sizing due to transient spikes, thus allowing you to take advantage of the elastic ability of the cloud.
  - 90th Percentile – More efficiency. Recommended for volumes that can stand higher resource utilization.

By default, Workload Optimization Manager uses samples from the last 30 days. Use the **Max Observation Period** setting to adjust the number of days.

- Max Observation Period

Attribute	Default value
Max Observation Period	Last 30 Days

To refine the calculation of resource utilization percentiles, you can set the sample time to consider. Workload Optimization Manager uses historical data from up to the number of days that you specify as a sample period. If the volume has fewer days' data then it uses all of the stored historical data.

Choose from the following settings:

  - Less Elastic – Last 90 Days
  - Recommended – Last 30 Days
  - More Elastic – Last 7 Days

- Min Observation Period

Attribute	Default value
Min Observation Period	None

This setting ensures historical data for a minimum number of days before Workload Optimization Manager will generate an action based on the percentile set in **Aggressiveness**. This ensures a minimum set of data points before it generates the action.

Especially for scheduled actions, it is important that resize calculations use enough historical data to generate actions that will remain viable even during a scheduled maintenance window. A maintenance window is usually set for "down" time, when utilization is low. If analysis uses enough historical data for an action, then the action is more likely to remain viable during the maintenance window.

Choose from the following settings:

- More Elastic – None
- Less Elastic – 1, 3, or 7 Days

Scaling target IOPS/Throughput utilization

This is the target utilization as a percentage of capacity.

Attribute	Default value
Scaling Target IOPS/Throughput Utilization	70

Cloud storage tiers

By default, Workload Optimization Manager considers all storage tiers currently available for scaling when making scaling decisions. However, you may have set up your cloud volumes to only scale to certain tiers to reduce complexity and cost, or meet demand. Use this setting to identify those storage tiers.

Attribute	Default value
Cloud Storage Tiers	None

Click **Edit** to set your preferences. In the new page that displays, choose from the following options:

- Excluded

  By default, this option is selected and all storage tiers are deselected. This means that all storage tiers are considered for scaling.

  Any storage tiers that you select are excluded from scaling.

  Configure an exclusion list if there are fewer storage tiers to exclude than to include. For example, if 400 storage tiers are currently available for scaling and you do not want cloud volumes to scale to 20 storage tiers, configure an exclusion list and select the 20 storage tiers.

- Included

  Any storage tiers that you select are considered for scaling.

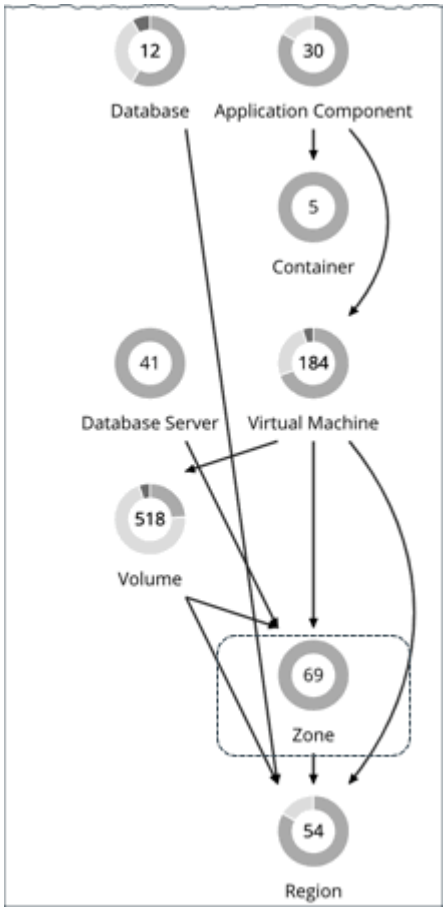
  Configure an inclusion list if there are fewer storage tiers to include than to exclude. For example, if 400 storage tiers are currently available for scaling and you want SQL Databases to scale only to 50 storage tiers, configure an inclusion list and select the 50 storage tiers.

After you save your changes, the main page refreshes to reflect your selections.

Zone

A zone represents an Availability Zone in your public cloud account or subscription. A zone is a location within a given region that serves as a datacenter to host the workloads that you run in your environment.

Synopsis



Synopsis	
Provides:	Compute and storage resources to VMs
Consumes:	Region resources
Discovered through:	Public cloud targets

Monitored resources

Workload Optimization Manager monitors the following resources:

- Templates

  The templates and template families that each zone or region delivers. This includes template capacity and cost for workload resources.

- Account Services

  These include storage modes, services the accounts offer for enhanced metrics, and services for different storage capabilities.

- (AWS only) Relational Database Services (RDS)

  The RDS capabilities each cloud account provides.

- Storage Tiers

  Workload Optimization Manager discovers the storage tier that supports your workloads, and uses the tier pricing to calculate storage cost.

- Billing

  Workload Optimization Manager discovers the billing across the zones and regions to predict costs in the future, and to track ongoing costs. This includes comparing on-demand pricing to discount billing.
- Virtual memory

  The percentage utilized of memory capacity for all the workloads in the zone.
- Virtual CPU

  The percentage utilized of vCPU capacity for all the workloads in the zone.
- Storage access

  For environments that measure storage access, the percentage utilized of access capacity for the zone.
- Storage amount

  The percentage utilized of storage capacity for the zone.
- IO throughput

  For environments that measure IO throughput, the percentage utilized of throughput capacity for the zone.
- Net throughput

  For environments that measure net throughput, the percentage utilized of throughput capacity for the zone.

Actions

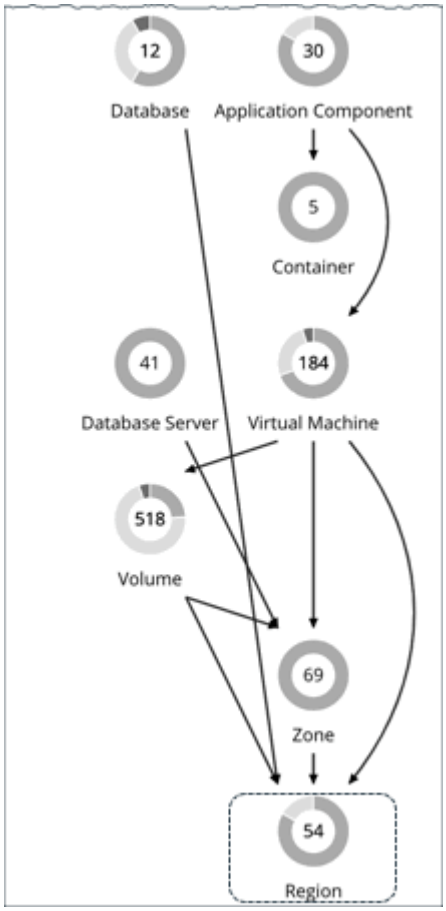
None

Workload Optimization Manager does not recommend actions for a cloud zone.

Region

A Region represents a geographical area that is home to one or more Availability Zones. Regions are often isolated from each other, and you can incur a cost for data transfer between them.

Synopsis



Synopsis	
Provides:	Host and storage resources to zones
Consumes:	N/A
Discovered through:	Public cloud targets

Monitored resources

Workload Optimization Manager does not monitor resources directly from the region, but it does monitor the following resources, aggregated for the zones in a region:

- Virtual memory

  The percentage utilized of memory capacity for workloads in the zones.
- Virtual CPU

  The percentage utilized of vCPU capacity for workloads in the zones.
- Storage access

  For environments that measure storage access, the percentage utilized of access capacity for the zones.
- Storage amount

  The percentage utilized of storage capacity for the zones.
- IO throughput

  For environments that measure IO throughput, the percentage utilized of throughput capacity for the zones.

- Net throughput

  For environments that measure Net throughput, the percentage utilized of throughput capacity for the zones.

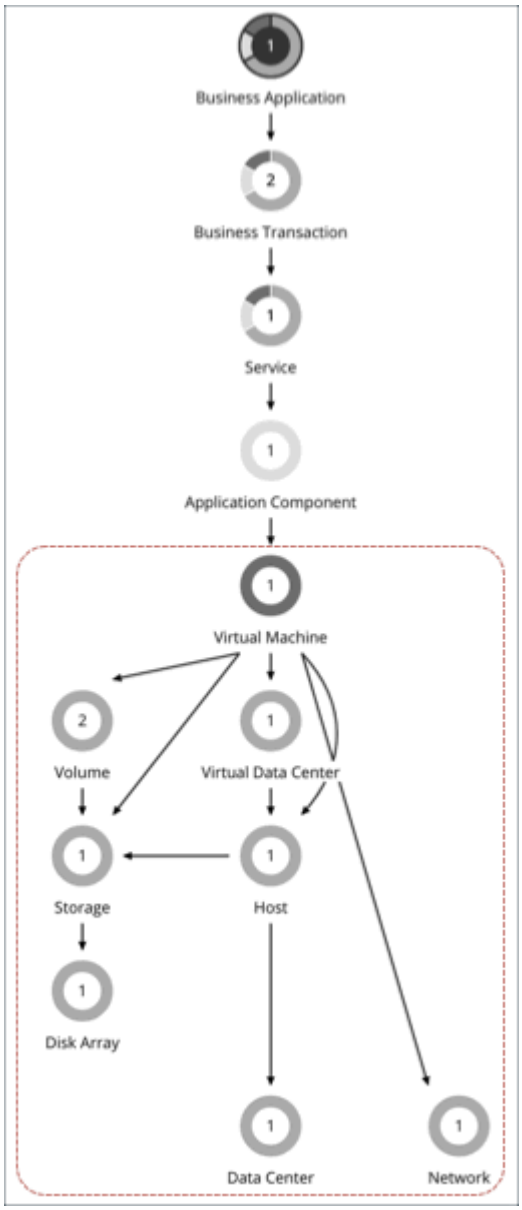
Actions

None

Workload Optimization Manager does not recommend actions for a cloud region.

Entity types - on-prem infrastructure

Workload Optimization Manager discovers and monitors the entities that make up your on-prem infrastructure, and recommends actions to assure performance for the applications that consume resources from these entities.

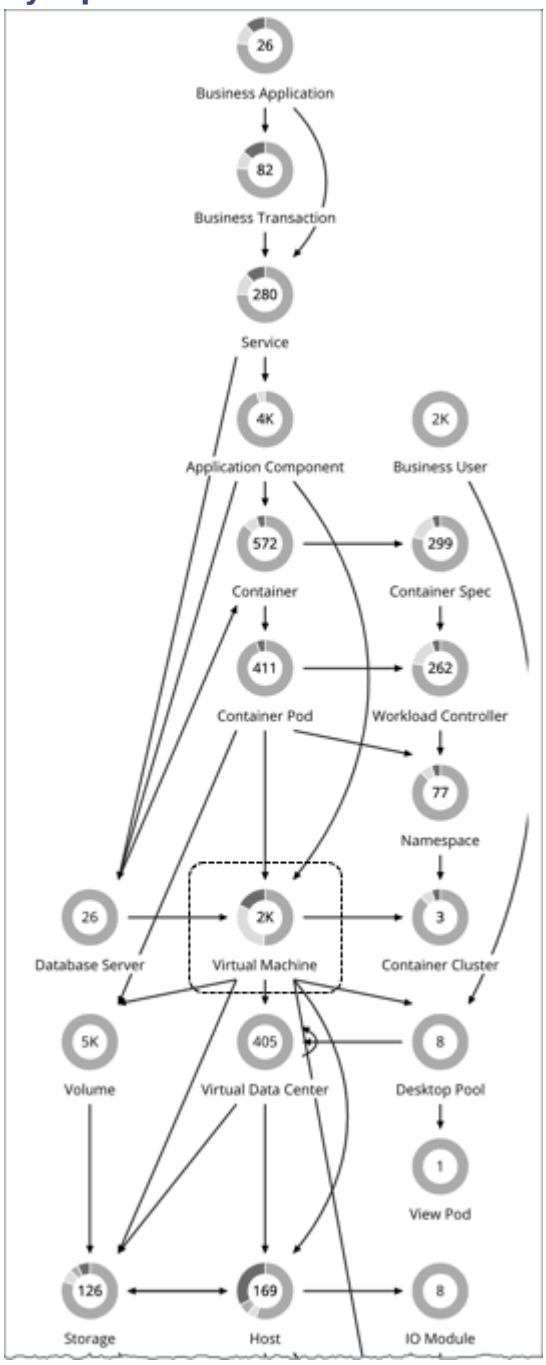


Virtual machine (on-prem)

A virtual machine (VM) is a software emulation of a physical machine, including OS, virtual memory and CPUs, and network ports. VMs host applications, or they provide resources to container platforms.

NOTE:
Container platform nodes are also represented as Virtual Machines in the Workload Optimization Manager supply chain. For details, see this topic *(on page 624)*.

Synopsis



A VM gains its budget by selling resources to the applications it hosts. If utilization is high enough, Workload Optimization Manager can allocate more resources to the VM, provision another instance, or move the VM to a host that has more resources. If utilization falls off, the VM loses budget. On the public cloud, if the budget isn't enough to pay for the host services, Workload Optimization Manager can post an action to suspend the VM.

Synopsis	
Provides:	Resources for hosted applications to use, including: ▪ VMEM ▪ VCPU ▪ Processing Units (IBM PowerVM) ▪ Virtual Processors (IBM PowerVM) ▪ VStorage ▪ IOPS ▪ Latency
Consumes:	▪ Physical host resources, including CPU and memory ▪ Storage
Discovered through:	Hypervisor targets

Monitored resources

Workload Optimization Manager monitors the following resources:

- Virtual memory (vMem)

  Virtual memory (vMem) is the measurement of memory that is in use.
- Virtual CPU (vCPU)

  Virtual CPU is the measurement of CPU that is in use.
- Processing unit (PU)

  For PowerVM, processing unit is a unit of measurement for shared processing power across one or more virtual processors. One shared processing unit on one virtual processor accomplishes approximately the same work as one dedicated processor.
- Virtual processor (VP)

  For PowerVM, virtual processor (VP) is the guest OS representation of the capacity of the VM. VPs have a one-to-one correlation with PUs. Therefore ,1 VP that is utilized in a VM equates to 1 PU being used. However, VPs do not have to be equal to PUs on a system. VPs are a virtual representation. Therefore, they can be overprovisioned.
- Virtual storage

  Virtual ssorage is the measurement of virtual storage capacity that is in use.
- Storage access (IOPS)

  Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
- Latency

  Latency is the measurement of storage latency.

For VMs discovered via vCenter, the following resources are also monitored:

- Energy

  Energy is the measurement of electricity used by a given entity over a period of time, expressed in watt-hours (Wh).
- Carbon footprint

  Carbon footprint is the measurement of carbon dioxide equivalent (CO2e) emissions for a given entity. Workload Optimization Manager measures carbon footprint in grams.

Actions

Workload Optimization Manager supports the following actions:

- Resize

  Resize resource capacity, reservation, or limit to improve performance.

  - Resize resource capacity

    Change the capacity of a resource that is allocated for the VM. For example, a resize action might recommend increasing the VMem available to a VM. Before recommending this action, Workload Optimization Manager verifies that the VM's cluster can adequately support the new size. If the cluster is highly utilized, Workload Optimization Manager will recommend a move action, taking into consideration the capacity of the new cluster and compliance with existing placement policies.

    For hypervisor targets, Workload Optimization Manager can resize vCPU by changing the VM's socket or cores per socket count. For details, see VCPU Scaling Controls *(on page 754)*.

  - Resize resource reservation

    Change the amount of a resource that is reserved for a VM. For example, a VM could have an excess amount of memory reserved. That can cause memory congestion on the host — A resize action might recommend reducing the amount reserved, freeing up that resource and reducing congestion

  - Resize resource limit

    Change the limit that is set on the VM for a resource. For example, a VM could have a memory limit set on it. If the VM is experiencing memory shortage, an action that decreases or removes the limit could improve performance on that VM.

  When multiple resize actions that relate to the same vCenter VM are accepted (manually or automated), they execute together to avoid multiple restarts for the same VM. For example, a VM has both a vCPU resize action and a vMEM resize action. When selecting both actions for execution, the actions combine into the fewest vCenter tasks to minimize the disruption on the VM instead of one set of tasks per action. Only the VMware vCenter integration supports this functionality at this time.

- Resize PU or VP (IBM PowerVM)

  Resize PU (processing units) or VP (virtual processor) to optimize LPAR processing unit allocation and virtual processor capacity based on historical demand collected from the HMC.

  Workload Optimization Manager does not support native execution of Resize PU (processing units) and Resize VP (virtual processor) actions. However, actions can be executed using external automation platforms through automation workflows such as Action Scripts or Webhooks.

  The following policy settings affect the resize actions that Workload Optimization Manager generates:

  - Min/Max observation period

    Observation period settings only apply to Resize VP actions.

    For details, see Aggressiveness and Observation Periods *(on page 752)*.

  - Increment constant for processing unit

    Change the increment constant value to use for VM PU resize actions. PUs can be a decimal value in a VM setting.

    For details, see Increment Constant for Processing Unit *(on page 749)*.

  - Increment constant for virtual processors

    Change the increment constant value to use for VM VP resize actions. VPs must be an integer value in a VM setting.

    For details, see Increment Constant for Virtual Processors *(on page 750)*.

- Move

  Move a VM due to high resource utilization on VM or host, excess IOPS or latency in VStorage, workload placement violation, underutilized host (move VM before suspending host).

- Reconfigure

Change a VM's configuration to comply with a policy.

For hypervisor targets, Workload Optimization Manager can reconfigure VMs that violate vCPU scaling policies. For details, see VCPU Scaling Controls *(on page 754)*.

Tuned scaling for on-prem VMs

For resizing VMs, Workload Optimization Manager includes tuned scaling action settings. These settings give you increased control over the action acceptance mode for various resize actions. With this feature, you can automate resize actions within a normal range (the tuned scaling range), and direct Workload Optimization Manager to take more conservative actions when resizes are outside the range.

For example, consider resizing VMs to add more memory. As memory demand increases on a VM, Workload Optimization Manager can automatically allocate more memory. If the hosted application is in a runaway state (always requesting more memory) and ultimately falls outside of the normal range, Workload Optimization Manager will not automate memory resize for the VM.

NOTE:
The VCPU Resize Min threshold setting does not generate compliance actions. Virtual machines that are already outside the threshold limits are not included in the scope of the policy.

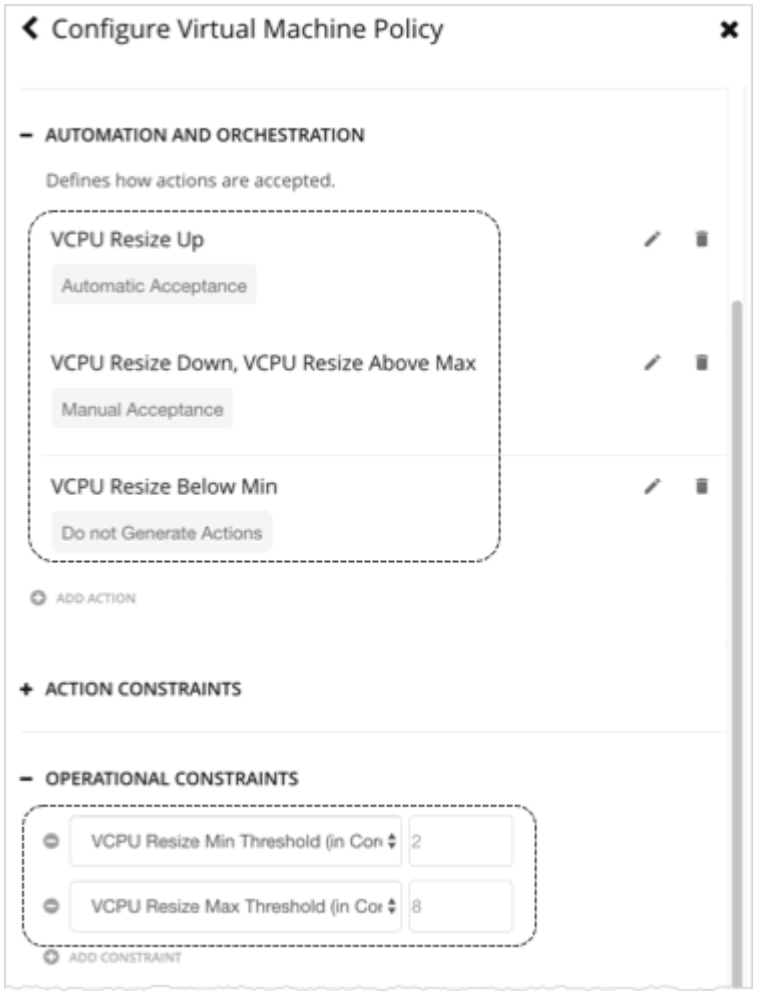
To configure tuned scaling:

1. Create a VM policy.
2. Under **Action Automation**, configure the action acceptance mode for the various resize actions.
   - VCPU Resize Up
   - VCPU Resize Down
   - VCPU Resize Above Max
   - VCPU Resize Below Min
   - VMEM Resize Up
   - VMEM Resize Down
   - VMEM Resize Above Max
   - VMEM Resize Below Min

   NOTE:
   Resize Up and **Resize Down** settings are for conditions within the tuned scaling range, while **Above Max** and **Below Min** settings are for outlying conditions.

3. Under **Operational Constraints**, specify the tuned scaling range.
   - VCPU Resize Max Threshold
   - VCPU Resize Min Threshold
   - VMEM Resize Max Threshold
   - VMEM Resize Min Threshold

For example, assume the following settings:



As VCPU utilization for a VM changes over time, Workload Optimization Manager handles resize actions as follows.

Current	Resize request	Action acceptance mode	Result
6 VCPUs	Resize up to 8 VCPUs	Automatic Since the VM will have 8 VCPUs after the requested resize, which is within the VCPU Resize Max threshold of 8, Workload Optimization Manager executes the VCPU Resize Up action automatically.	8 VCPUs
8 VCPUs	Resize up to 10 VCPUs	Manual Since the VM will have 10 VCPUs after the requested resize, which is above the VCPU Resize Max threshold of 8, Workload Optimization Manager posts the VCPU Resize Up action (as a pending action) and provides the option to execute	10 VCPUs (if you executed the pending action)

Current	Resize request	Action acceptance mode	Result
		that action through the user interface.	
10 VCPUs	Resize down to 2 VCPUs	Manual Since the VM will have 2 VCPUs after the requested resize, which is within the VCPU Resize Min threshold of 2, Workload Optimization Manager posts the VCPU Resize Down action (as a pending action) and provides the option to execute that action through the user interface.	2 VCPUs (if you executed the pending action)
2 VCPUs	Resize down to 1 VCPU	Not Generated Since the VM will have 1 VCPU after the requested resize, which is below the VCPU Resize Min threshold of 2, Workload Optimization Manager does not generate the VCPU Resize Down action to comply with the policy.	2 VCPUs

Action policies include scope to determine which entities will be affected by the given policy. It's possible for two or more policies to affect the same entities. As is true for other policy settings, tuned scaling uses the most conservative settings for the affected entities. The effective action acceptance mode will be the most conservative, and the effective tuned scaling range will be the narrowest range (the lowest MAX and highest MIN) out of the multiple policies that affect the given entities. For more information, see Default and User-defined Policies *(on page 1050)*.

You can schedule automation policies to take effect during a certain window of time. You can include tuned scaling settings in a scheduled window, the same as you can schedule other policy settings. For more information, see Policy Schedule *(on page 1061)*.

On-prem VM policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

For details about on-prem VM actions, see On-prem VM Actions *(on page 743)*.

Action	Default mode	vCenter	Hyper-V	PowerVM
Move	Manual	Automatable	Automatable	N/A
Provision (container platform nodes only)	Manual	Recommend	Recommend	N/A
Reconfigure	Recommend	Recommend	Recommend	N/A

Action	Default mode	vCenter	Hyper-V	PowerVM
Resize Down PU	Recommend	N/A	N/A	Recommend
Resize Down VP	Recommend	N/A	N/A	Recommend
Resize Up PU	Recommend	N/A	N/A	Recommend
Resize Up VP	Recommend	N/A	N/A	Recommend
Start	Manual	Automatable	Automatable	N/A
Storage Move	Recommend	Automatable	Automatable with VMM. Otherwise, Recommend.	N/A
Suspend (container platform nodes only)	Manual	Automatable	Automatable	N/A
vCPU Resize Above Max*	Recommend	Automatable	Automatable	N/A
vCPU Resize Below Min*	Recommend	Automatable	Automatable	N/A
vCPU Resize Down*	Manual	Automatable	Automatable	N/A
vCPU Resize Up*	Manual	Automatable	Automatable	N/A
vMem Resize Above Max*	Recommend	Automatable	Automatable	N/A
vMem Resize Below Min*	Recommend	Automatable	Automatable	N/A
vMem Resize Down*	Manual	Automatable	Automatable	N/A
vMem Resize Up*	Manual	Automatable	Automatable	N/A

** Workload Optimization Manager uses these settings, in conjunction with resize operational constraints (on page 749), to set up tuned scaling (on page 744) for on-prem VMs.*

You can use Action Scripts *(on page 1070)* and third-party orchestrators (such as ServiceNow) for action orchestration.

Non disruptive mode

VM actions include the modifier, **Enforce Non Disruptive Mode**. When you enable this modifier, Workload Optimization Manager ensures that a resize action in *Automatic* or *Manual* mode will not require a reboot or any other disruption to the affected VM. If the action will disrupt the VM, Workload Optimization Manager posts the action in *Recommend* mode.

Attribute	Default setting
Enforce Non Disruptive Mode	Off

This setting has no effect on actions set to *Recommend* mode. Workload Optimization Manager will continue to post those actions for you to evaluate.

You can enforce non disruptive mode in the default VM policy, and then schedule action policies to automate resize actions during downtimes. Be aware that scheduled actions do not respect the enforced non disruptive mode – Scheduled resize actions will execute during the scheduled window even if they require a reboot. This is useful for setting up certain action behaviors, but you must be aware that enforced non disruptive mode has no effect on scheduled actions.

> **NOTE:**
> When you configure a schedule window for a VM resize action, to ensure Workload Optimization Manager will execute the action during the scheduled time, you must turn off the **Enforce Non Disruptive Mode** setting for that scheduled policy. Even if you turn the setting off for the global policy, you still must turn the setting off for your scheduled policy. Otherwise Workload Optimization Manager will not execute the resize action.

Shared-nothing migration

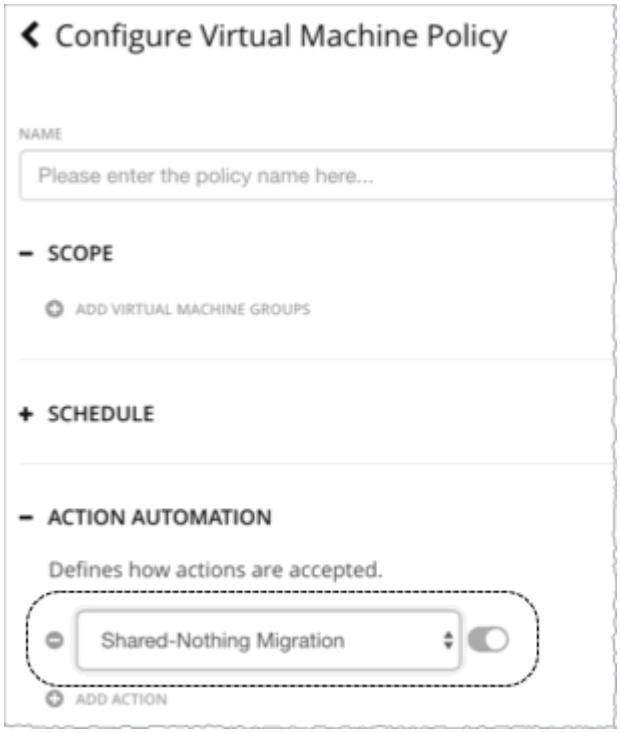
If you have enabled both storage and VM moves, Workload Optimization Manager can perform shared-nothing migrations, which move the VM and the stored VM files simultaneously. For example, assume a VM on a host also uses local storage on that host. In that case, Workload Optimization Manager can move that VM and move its data to a different datastore in a single action.

Attribute	Default setting
Shared-Nothing Migration	Off

Currently, the following targets support shared-nothing migrations:

- vSphere, versions 5.1 or greater
- VMM for Hyper-V 2012 or later

Because of this feature's potential impact on performance, it is turned off by default. Workload Optimization Manager recommends enabling it only on VMs that need it. To do this, you must first set the action acceptance mode for VM and storage moves to either *Manual* or *Automatic*, and then enable the feature in a VM policy.



If a policy that enables this feature conflicts with a more conservative policy, the latter policy wins. For example, if compute move is set to *Manual*, storage move is set to *Recommend*, and shared-nothing migration is turned on, shared-nothing migration is in effect but remains in *Recommended* state.

NOTE:
Workload Optimization Manager does not simulate shared-nothing migrations in plans.

Resize thresholds

Workload Optimization Manager uses these settings to set up **tuned scaling** actions for on-prem VMs. Tuned scaling gives you increased control over the action mode for various resize actions. With this feature, you can automate resize actions within a normal range (the tuned scaling range), and direct Workload Optimization Manager to take more conservative actions when a resize is outside of the range.

For details, see Tuned Scaling for On-prem VMs *(on page 744)*.

Attribute	Default value
vCPU Resize Max Threshold (in Cores)	64 Cores Tuned Scaling Range Upper Limit
vCPU Resize Min Threshold (in Cores)	1 Core Tuned Scaling Range Lower Limit
VMEM Resize Min Threshold	512 MB Tuned Scaling Range Lower Limit
VMEM Resize Max Threshold	131072 MB Tuned Scaling Range Upper Limit

Move enabled GPUs

Use this setting to construct a regex statement to identify which GPU models should support workload placement automation. By default, `.*_(?:a16|t4).*$` is the regex pattern used for this field.

On-prem virtual machines with vGPU types that match the specified pattern are enabled for moves. Leave this field blank to disable moves for all VMs in the selected scope.

To enable vMotion for vGPU Virtual Machines, you must change the `vgpu.hotmigrate.enabled` advanced setting. For more information, see the VMware documentation.

Increment constant for processing units

For PowerVM, processing unit is a unit of measurement for shared processing power across one or more virtual processors. One shared processing unit on one virtual processor accomplishes approximately the same work as one dedicated processor.

Workload Optimization Manager recommends VM PU resize actions in multiples of this increment constant value. For example, if Workload Optimization Manager determines that a VM with 1.0 PU needs 1.2 PU and the increment is set to 0.5, then a resize up action is not generated. Increasing this increment reduces the frequency of resize actions so that when they appear the change is more significant.

Attribute	Default value
Increment constant for Processing Units	0.1

Increment constant for virtual processors

For PowerVM, virtual processor (VP) is the guest OS representation of the capacity of the VM. VPs have a one-to-one correlation with PUs. Therefore ,1 VP that is utilized in a VM equates to 1 PU being used. However, VPs do not have to be equal to PUs on a system. VPs are a virtual representation. Therefore, they can be overprovisioned.

Workload Optimization Manager recommends VM VP resize actions in multiples of this increment constant value. For example, if Workload Optimization Manager determines that a VM with 2 VPs needs 1 VP and the increment is set to 2, then a resize down action is not generated. Increasing this increment reduces the frequency of resize actions so that when they appear, the change is more significant.

Attribute	Default value
Increment constant for Virtual Processors	1

Resize VStorage

The default setting disables resize actions. This is usually preferred because VStorage resize requires that you reformat the storage. The increment constant takes effect if you enable resizing.

Attribute	Default setting/value
Resize VStorage	Disabled
Increment constant for VStorage	None If you enable resize, Workload Optimization Manager uses the default value of 1024 GB. You can change this to a different value.

vCPU scaling controls

For details, see VCPU Scaling Controls *(on page 754)*.

Resize increments

These increments specify how many units to add or subtract when resizing the given resource allocation for a VM.

Attribute	Default value
Increment constant for VMem	1024 MB
Increment constant for VStorage	1024 GB

NOTE:
vCPU resize increments are configured in conjunction with vCPU scaling controls. For details, see VCPU Scaling Controls *(on page 754)*.

For VMem, you should not set the increment value to be less than what is necessary for the VM to operate. If the VMem increment is too low, then it's possible that Workload Optimization Manager would allocate insufficient VMem for the machine to operate. For a VM that is under utilized, Workload Optimization Manager will reduce VMem allocation by the increment amount, but it will not leave a VM with zero VMem. For example, if you set this to 1024, then Workload Optimization Manager cannot reduce the VMem to less than 1024 MB.

Rate of resize

When resizing resources for a VM, Workload Optimization Manager calculates the optimal values for VMem, VCPU, and V Storage, but it does not necessarily make a change to that value in one action. Workload Optimization Manager uses the Rate of Resize setting to determine how to make the change in a single action.

Using the Low (1) or Medium (2) value creates actions to bring the VM partially to the Desired State, resulting in more actions and possibly more total downtime. It is recommended to use the default value of High (3) to give the VM the correct resources all at once, which achieves the benefits of performance and efficiency in one action. Fewer actions results in less possible downtime.

Attribute	Default value
Rate of Resize	High (3)

- Low

  Change the value by one increment, only. For example, if the resize action calls for increasing VMem, and the increment is set at 1024, Workload Optimization Manager increases VMem by 1024 MB.

- Medium

  Change the value by an increment that is 1/4 of the difference between the current value and the optimal value. For example, if the current VMem is 2 GB and the optimal VMem is 10 GB, then Workload Optimization Manager will raise VMem to 4 GB (or as close to that as the increment constant will allow).

- High

  Change the value to be the optimal value. For example, if the current VMem is 2 GB and the optimal VMem is 8 GB, then Workload Optimization Manager will raise VMem to 8 GB (or as close to that as the increment constant will allow).

Consistent resizing

Attribute	Default Setting
Consistent Resizing	Off

When you create a policy for a group of VMs and turn on Consistent Resizing, Workload Optimization Manager resizes all the group members to the same size, such that they all support the top utilization of each resource commodity in the group. For example, if you have deployed load balancing for a group, then all the VMs in the group should experience similar utilization. In that case, if one VM needs to be resized, then it makes sense to resize them all consistently.

Assume VM A shows top utilization of CPU, and VM B shows top utilization of memory. A resize action would result in all the VMs with CPU capacity to satisfy VM A, and memory capacity to satisfy VM B.

> **NOTE:**
> If the VMs in the group have different core speeds, then CPU scaling actions might not be consistent. For example, if you set the maximum target CPU size to 2, Workload Optimization Manager might recommend resizing to more than 2 CPUs to account for the VMs with slower cores.
>
> To avoid this problem, be sure that the group only includes VMs with the same core speed.

For an affected resize, the Actions List shows individual resize actions for each of the VMs in the group. To avoid the possibility of resizing VMs disruptively at the same time, you must create automation policies with non-overlapping schedules. For example, if VMs A and B are in the same consistent resizing group, create two policies that resize the VMs at different times of the day.

- For Policy 1, set the scope to a group containing VM A and enable resize automation between, say, 01:00 and 01:45.
- For Policy 2 set the scope to a group containing VM B and enable resize automation between 02:00 and 02:45.

When working with Consistent Resizing, consider these points:

- You should not mix VMs in a group that has a Consistent Resizing policy, with other groups that enable Consistent Resizing. One VM can be a member of more than one group. If one VM (or more) in a group with Consistent Resizing is also in another group that has Consistent Resizing, then both groups enforce Consistent Resizing together, for all their group members.

- For any group of VMs that enables Consistent Resizing, you should not mix the associated target technologies. For example, one group should not include VMs that are on Hyper-V and vCenter platforms.
- Charts that show actions and risks assign the same risk statement to all the affected VMs. This can seem confusing. For example, assume one VM needs to resize to address vCPU risk, and 9 other VMs are set to resize consistently with it. Then charts will state that 10 VMs need to resize to address vCPU risks.

Fault tolerance

This setting only applies to on-prem virtual machines. The **Fault Tolerance** setting protects application performance in case of virtual machine failures or interruptions. When you configure this setting, Workload Optimization Manager reserves the capacity needed to sustain performance in case one or more virtual machines fail. If a failure does occur, the other virtual machines in the group are able to take on the application workload. For example, in a group of three virtual machines, with a **Fault Tolerance** setting of 1, the maximum utilization of any virtual machine changes to 66%. If a single VM fails, it is expected that the 66% resource utilization of the failed VM transfers to the two remaining VMs in the group. The recognition of a VM failure and load re-balancing is expected to be handled by the application.

NOTE:
It's required that the given value should be less than the total size of the group.

Nutanix: Attempt vCPU Hotplug

When this setting is enabled, Workload Optimization Manager attempts to resize vCPU for Nutanix VMs without restarting the scoped VMs. The hotplug feature is provided by Nutanix and might be subject to vendor limitations. For more information, see the Nutanix documentation.

Nutanix: Attempt vMEM Hotplug

When this setting is enabled, Workload Optimization Manager attempts to resize vMEM for Nutanix VMs without restarting the scoped VMs. The hotplug feature is provided by Nutanix and might be subject to vendor limitations. For more information, see the Nutanix documentation.

Enable hot add for vMEM and vCPU

When this setting is enabled, Workload Optimization Manager enables the vCenter hot add capability the next time it initiates a restart on the scoped VMs. After hot add is enabled, Workload Optimization Manager can resize vCPU or vMEM for vCenter VMs without restarting the scoped VMs. The hot add feature is provided by vCenter and might be subject to vendor limitations. For more information, see the vCenter documentation.

Run automated resizes in sequence

This setting only applies to on-prem virtual machines. When selected, Workload Optimization Manager runs actions that are in automatic mode one at a time, starting with the least utilized VM in the group and ending with the highest utilized VM in the group. Utilization is based on three consecutive discovery cycles.

Aggressiveness and observation periods

Workload Optimization Manager uses these settings to calculate utilization percentiles for vCPU, vMEM, IOPS, and VPs. It then recommends actions to improve utilization based on the observed values for a given time period.

- Aggressiveness

Attribute	Default value
Aggressiveness	75th Percentile

When evaluating performance, Workload Optimization Manager considers resource utilization as a percentage of capacity. The utilization drives actions to scale the available capacity either up or down. To measure utilization, the analysis considers a given utilization percentile. For example, assume a 75th percentile. The percentile utilization is the

highest value that 75% of the observed samples fall below. Compare that to average utilization, which is the average of all the observed samples.

Using a percentile, Workload Optimization Manager can recommend more relevant actions. For scheduled policies, the more relevant actions will tend to remain viable when their execution is put off to a later time.

For example, consider decisions to reduce the capacity for CPU on a VM. Without using a percentile, Workload Optimization Manager never resizes below the recognized peak utilization. For most VMs, there are moments when peak CPU reaches high levels, such as during reboots, patching, and other maintenance tasks. Assume utilization for a VM peaked at 100% just once. Without the benefit of a percentile, Workload Optimization Manager will not reduce allocated CPU for that VM.

With **Aggressiveness**, instead of using the single highest utilization value, Workload Optimization Manager uses the percentile you set. For the above example, assume a single CPU burst to 100%, but for 75% of the samples CPU never exceeded 50%. If you set **Aggressiveness** to 75th Percentile, then Workload Optimization Manager can see this as an opportunity to reduce CPU allocation for the VM.

In summary, a percentile evaluates the sustained resource utilization, and ignores bursts that occurred for a small portion of the samples.

By default, Workload Optimization Manager uses samples from the last 30 days. Use the **Max Observation Period** setting to adjust the number of days. To ensure that there are enough samples to analyze and drive scaling actions, set the **Min Observation Period**.

- Max Observation Period

Attribute	Default value
Max Observation Period	Last 90 Days

To refine the calculation of resource utilization percentiles, you can set the sample time to consider. Workload Optimization Manager uses historical data from up to the number of days that you specify as a sample period. If the database has fewer days' data then it uses all of the stored historical data.

You can make the following settings:

  - Less Elastic – Last 90 Days
  - Recommended – Last 30 Days
  - More Elastic – Last 7 Days

Workload Optimization Manager recommends an observation period of 30 days following the monthly workload maintenance cycle seen in many organizations. VMs typically peak during the maintenance window as patching and other maintenance tasks are carried out. A 30-day observation period means that Workload Optimization Manager can capture these peaks and increase the accuracy of its sizing recommendations.

You can set the value to 7 days if workloads need to resize more often in response to performance changes. For workloads that cannot handle changes very often or have longer usage periods, you can set the value to 90 days.

- Min Observation Period

Attribute	Default value
Min Observation Period	None

Especially for scheduled actions, it is important that resize calculations use enough historical data to generate actions that will remain viable even during a scheduled maintenance window. A maintenance window is usually set for "down" time, when utilization is low. If analysis uses enough historical data for an action, then the action is more likely to remain viable during the maintenance window.

  - More Elastic – None
  - Less Elastic – 7 Days

Placement policies

Workload Optimization Manager supports placement policies for on-prem VMs, as follows:

- You can create placement policies to enforce constraints for VM placements.

  For example, the VMs in a consumer group can only run on a host that is in the provider group. You can limit the number of consumers that can run on a single provider – for hosts in the provider group, only 2 instances of VMs in the consumer group can run on the same host. Or no more than the specified number of VMs can use the same storage device.
- For VMs that require paid licenses, you can create placement policies that set up certain hosts to be the VMs' preferred license providers. Workload Optimization Manager can then recommend consolidating VMs or reconfiguring hosts in response to changing demand for licenses.

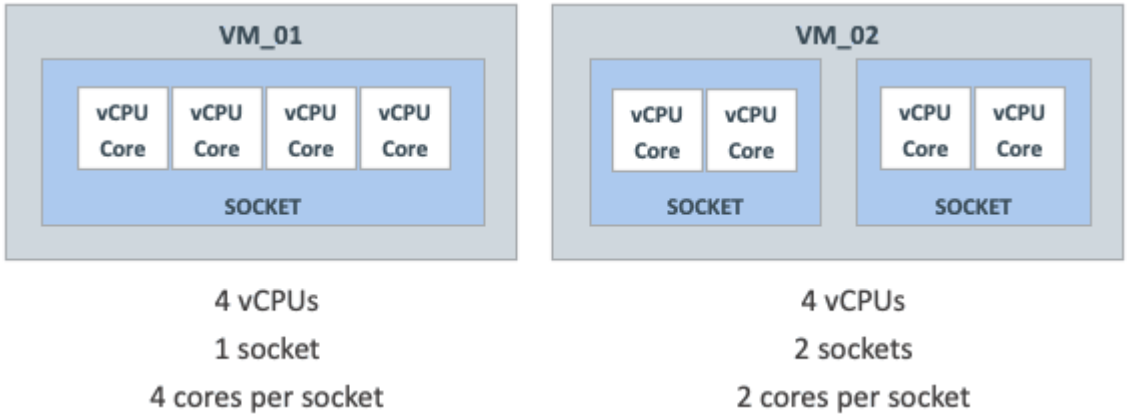
For more information, see Creating Placement Policies *(on page 1044)*.

NOTE:
For VMM targets, Cisco automatically imports your Availability Sets, representing them as placement policies for the affected infrastructure. To see these availability sets, go to the **Settings > Policies** page and click **Imported Placement Policies**.

For more information, see Importing Workload Placement Policies *(on page 1044)*.

vCPU scaling controls

Workload Optimization Manager represents the compute capacity of a VM in MHz and vCPUs. The following diagram shows how a VM with four vCPUs can be configured differently in terms of sockets and cores.



Workload Optimization Manager can resize the compute capacity by changing the number of sockets or cores per socket, depending on:

- The policy that is assigned to the VM

  On-prem VM policies *(on page 746)* include vCPU Scaling Controls that give you granular control over how VM compute resources are resized to maintain performance or reconfigured to comply with your operational policies. You can create policies for different VM groups based on their resource needs and characteristics, and decide whether to automate resize and reconfigure actions in those policies.
- The hypervisor that manages the VM

  Hypervisor targets have varying degrees of support for vCPU Scaling Controls. VMware vSphere supports all scaling controls, while Hyper-V and Nutanix AHV provide limited support. For details, see the *Hypervisor Support* section.

vCPU scaling control modes and options

Workload Optimization Manager provides **simple** and **advanced** controls to automate compute resource management actions in compliance with your policies. It also provides a **legacy** control based on units of MHz.

The controls that you choose depend on your operational policies regarding the VM configuration of sockets and cores per socket, and your choice of hypervisor. For example, your operational policies may dictate a certain VM configuration that must be respected when resizing a VM's compute resources. Changing sockets is the least disruptive, but for some workloads, it may be preferable to change cores per socket due to socket licensing or operating system constraints. For larger-size VMs where Non-Uniform Memory Access (NUMA) must be considered for performance reasons, it may be preferable to balance vCPUs across host sockets.

The following tables explain the exact operation for each mode.

Simple Controls: Simple controls change compute resources based on units of vCPU.

vCPU Scaling option	Unit	Sockets	Cores per socket	Resize action	Reconfigure action
Change virtual CPUs	vCPUs	Workload Optimization Manager decides	Reconfigured to 1 core per socket	▪ Non-disruptive if hot-add is enabled and VM sockets are increasing ▪ Disruptive if VM cores per socket does not equal 1, even if hot-add is enabled	Not generated

Advanced Controls: Use advanced controls to change sockets or cores per socket, and configure more options.

vCPU scaling option	Unit	Sockets	Cores per socket	Resize action	Reconfigure action
Change sockets	1 socket	Workload Optimization Manager decides	Preserve VM cores per socket	Non-disruptive if hot-add is enabled and VM sockets are increasing	Not generated
Change sockets	1 socket	Workload Optimization Manager decides	User-specified VM cores per socket	▪ Non-disruptive if hot-add is enabled ▪ Disruptive if VM cores per socket does not match	Disruptive if VM cores per socket does not match a user-specified value

vCPU scaling option	Unit	Sockets	Cores per socket	Resize action	Reconfigure action
				a user-specified value	
Change cores per socket	1 core per socket	Preserve VM sockets	Workload Optimization Manager decides	Disruptive	Not generated
Change cores per socket	1 core per socket	Match host sockets	Workload Optimization Manager decides	Disruptive	▪ Non-disruptive if hot-add is enabled and VM sockets are increasing ▪ Disruptive if cores per socket is changed
Change cores per socket	1 core per socket	User-specified VM sockets	Workload Optimization Manager decides	Disruptive	▪ Non-disruptive if hot-add is enabled and VM sockets are increasing ▪ Disruptive if cores per socket is changed

Legacy Controls: Legacy controls change compute resources based on units of MHz.

vCPU scaling option	Unit	Sockets	Cores per socket	Resize action	Reconfigure action
MHz legacy behavior	MHz	Workload Optimization Manager decides	▪ Assumes 1 core per socket ▪ Execution preserves actual cores per socket	Non-disruptive if hot-add is enabled and VM sockets are increasing	Not generated

Points to consider:

- If non-disruptive mode *(on page 747)* is enabled, disruptive actions are not automated and must be run manually.
- Older Guest OSes and applications may be sensitive to changes in the vCPU architecture that might result in power-on issues or kernel panics/BSODs. Some workloads require manual help with such changes, so always test certain classes of applications and guest operating systems before enabling any automation that changes the vCPU architecture. Use the Workload Optimization Manager knowledge of the application domain and Guest OS to scope them out of policies.

Scaling option: change virtual CPUs

In this scaling option, Workload Optimization Manager adds or removes compute resources in increments of vCPUs. To achieve this, it changes the number of VM sockets and enforces 1 core per socket (if not already enforced).

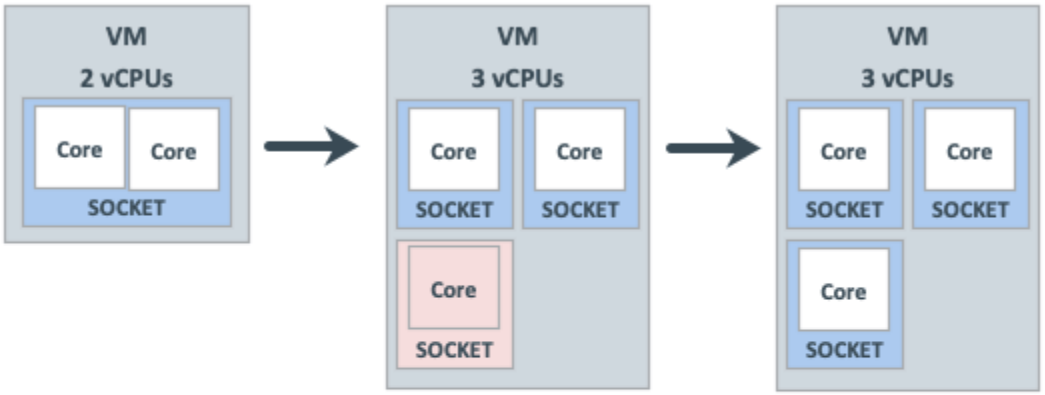
- If a VM requires a change to compute resources, Workload Optimization Manager generates a resize vCPU action that assumes 1 core per socket. If the VM currently does not have 1 core per socket, Workload Optimization Manager reconfigures it to 1 core per socket as part of action execution.
- If a VM is already optimally sized, but its current cores per socket is not 1, resize or reconfigure actions are not generated.

This scaling option is ideal under the following scenarios:

- Your environment has many small VMs where precise vCPU scaling is the priority.
- You have VMs that already have 1 core per socket and require on-demand upsizes on these VMs to be non-disruptive.

For example, a VM currently has 1 socket and 2 cores per socket, and applies a policy that changes vCPU in increments of 1.

- If Workload Optimization Manager determines that the VM needs to increase compute capacity by 1 vCPU (that is, from 2 to 3 vCPUs), a resize up action changes sockets from 2 to 3, and cores per socket from 2 to 1.

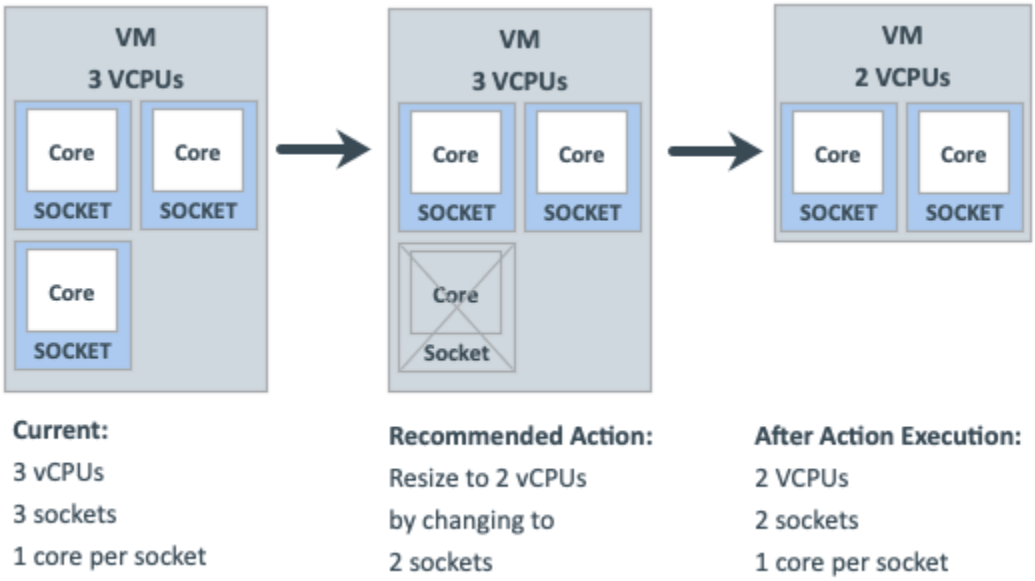


Current:
2 vCPUs
1 socket
2 cores per socket

Recommended Action:
Resize to 3 vCPUs
by changing to:
- 3 sockets
- 1 core per socket

After Action Execution:
3 vCPUs
3 sockets
1 core per socket

- When the same VM needs to reduce compute capacity by 1 vCPU (that is, from 3 to 2 vCPUs), a resize down action changes sockets from 3 to 2.



Scaling option: change sockets

In this scaling option, Workload Optimization Manager adds or removes compute resources by changing VM sockets.

- If a VM requires a change to compute resources, Workload Optimization Manager generates a resize vCPU action that considers the current cores per socket value (if the **Preserve existing VM cores per socket** option is set) or uses the user-specified cores per socket value. If the VM's current cores per socket value violates a policy (that is, does not match the user-specified value), Workload Optimization Manager reconfigures the VM's cores per socket value as part of action execution, thereby bringing the VM into compliance with the policy, while at the same time providing the required change to compute resources.
- If a VM is already optimally sized, but its current cores per socket value violates a policy (that is, does not match the user-specified value, if set), Workload Optimization Manager generates a reconfigure vCPU action to change cores per socket to the user-specified value, thereby bringing the VM into compliance with the policy.

Change Sockets and Preserve VM Cores Per Socket

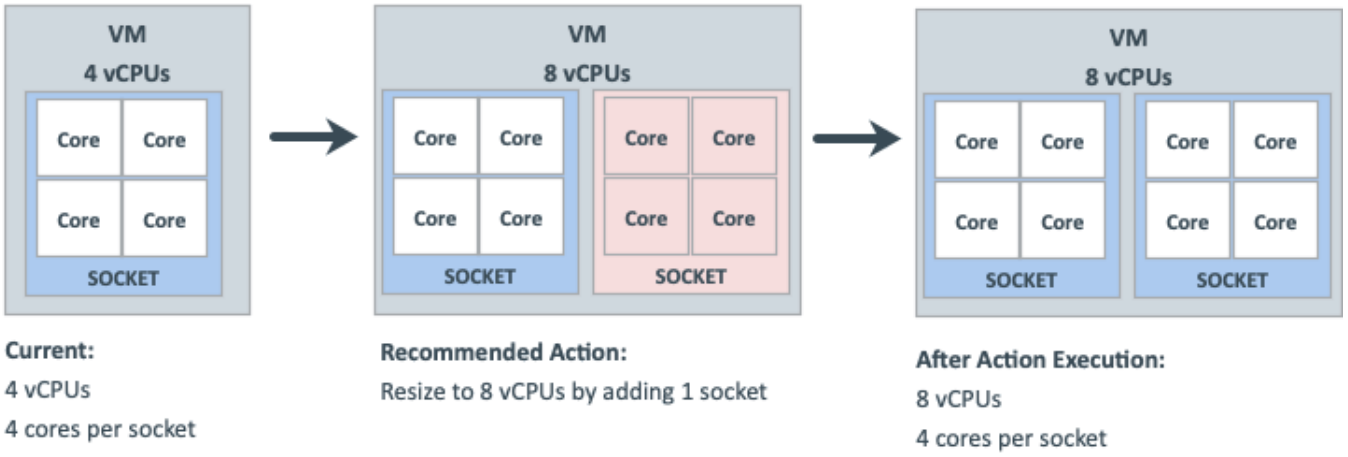
In this scaling option, Workload Optimization Manager adds or removes compute resources by changing VM sockets in increments of 1, and preserves VM cores per socket.

This scaling option is ideal under the following scenarios:

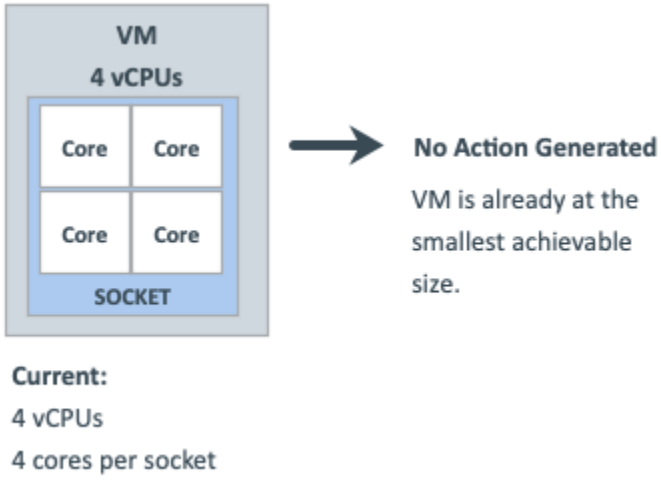
- You require Workload Optimization Manager to leave the VM cores per socket configuration unchanged for operational policy reasons (such as compliance with an application support contract policy).
- You have VMs that need to upsize non-disruptively to meet rising application demand.
- You have VMs with even numbers of cores per socket and are required to scale in even increments of vCPUs.

For example, a VM currently has 1 socket and 4 cores per socket, and applies a policy that changes sockets and preserves VM cores per socket. Workload Optimization Manager has determined that the VM requires a change in compute capacity of 1 vCPU.

- To increase compute capacity by 1 vCPU, a resize up action adds 1 socket. Because this new socket must have 4 cores to preserve VM cores per socket, the end result is 2 sockets with a total of 8 vCPUs.



- It is not possible to reduce compute capacity by 1 vCPU because the VM is already at the smallest achievable size. Therefore, no action generates.



Change Sockets and Specify Cores Per Socket

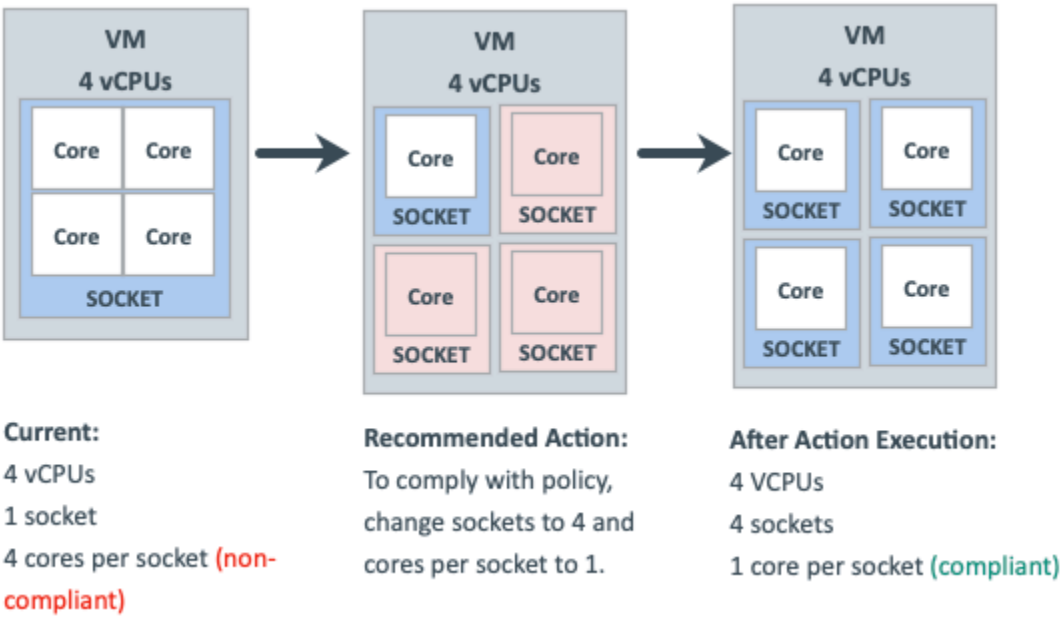
In this scaling option, Workload Optimization Manager adds or removes compute resources by changing VM sockets in increments of 1, and reconfigures VM cores per socket according to the value that you specify.

This scaling option is ideal under the following scenarios:

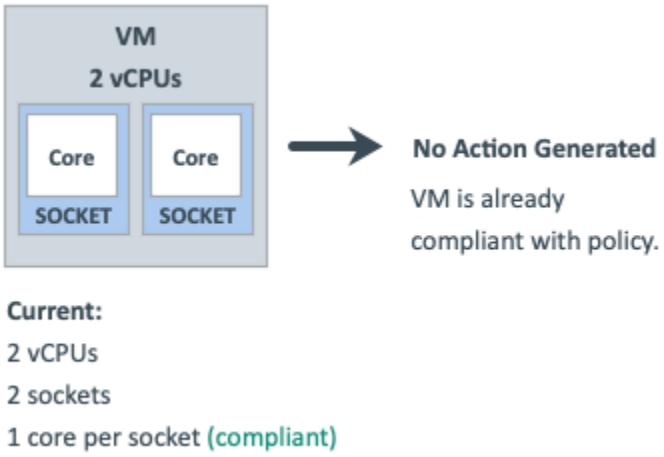
- You require any odd number of vCPUs for a VM to be an even number, by setting an even number of cores per socket.
- You want a quick, script-less bulk disruptive conversion of VMs to a specific cores per socket without negatively impacting compute capacity (vCPUs).
- You have older Guest OSes and applications that are sensitive to vCPU architecture changes that might result in power-on issues or kernel panics/BSODs. Some workloads require manual help with such changes so always test certain classes of applications and OSes before enabling any automation that changes the vCPU architecture. Use the Workload Optimization Manager knowledge of the application domain and Guest OS to scope them out of policies.

For example, a VM currently has 1 socket and 4 cores per socket, and applies a policy that changes sockets and enforces the user-specified 1 core per socket. Workload Optimization Manager has determined that the VM is already optimally sized, so a resize action is not necessary.

- Since the VM is in violation of policy, Workload Optimization Manager changes sockets from 1 to 4, and cores per socket from 4 to 1.



- When the VM is compliant with policy, no action generates.



Scaling option: change cores per socket

In this scaling option, Workload Optimization Manager adds or removes compute resources by changing the VM cores per socket.

- If a VM requires a change to compute resources, Workload Optimization Manager generates a resize vCPU action that considers the current socket value (if the **Preserve existing VM sockets** option is set), respects the user-specified socket value, or matches VM sockets to the host socket value. If the VM's current socket value violates a policy (that is, does not match the user-specified or host socket value), Workload Optimization Manager reconfigures the VM's socket

value as part of action execution, thereby bringing the VM into compliance with the policy while providing the required change to compute resources.

- If a VM is already optimally sized, but its current socket value violates a policy, Workload Optimization Manager generates a reconfigure vCPU action to change the sockets to the user-specified or host socket value, thereby bringing the VM into compliance with the policy.

Older Guest OSes and applications may be sensitive to vCPU architecture changes that might result in power-on issues or kernel panics/BSODs. Some workloads require manual help with such changes so always test certain classes of applications and OSes before enabling any automation that changes the vCPU architecture. Use the Workload Optimization Manager knowledge of the application domain and Guest OS to scope them out of policies.

Change Cores Per Socket and Preserve VM Sockets

In this scaling option, Workload Optimization Manager adds or removes compute resources by changing the VM cores per socket in increments of 1, and preserves VM sockets.

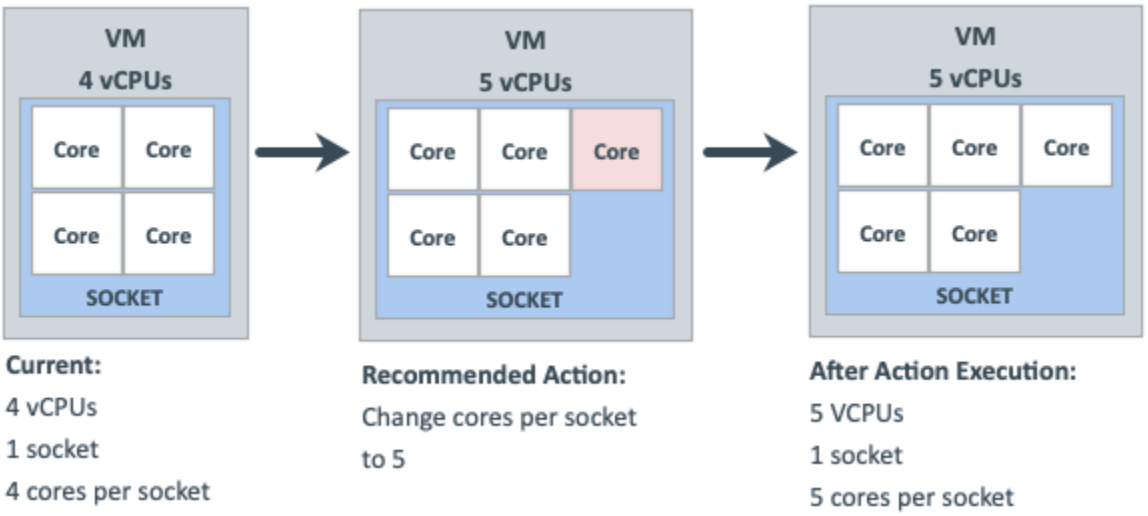
This scaling option is ideal under the following scenarios:

- You require Workload Optimization Manager to leave the VM sockets configuration unchanged for operational policy reasons (such as socket-based licensing or compliance with an application support contract policy).
- You have VDI VMs that are at their maximum Guest OS socket limitation, but require more compute resources.
- You have VMs that are configured with NUMA considerations.

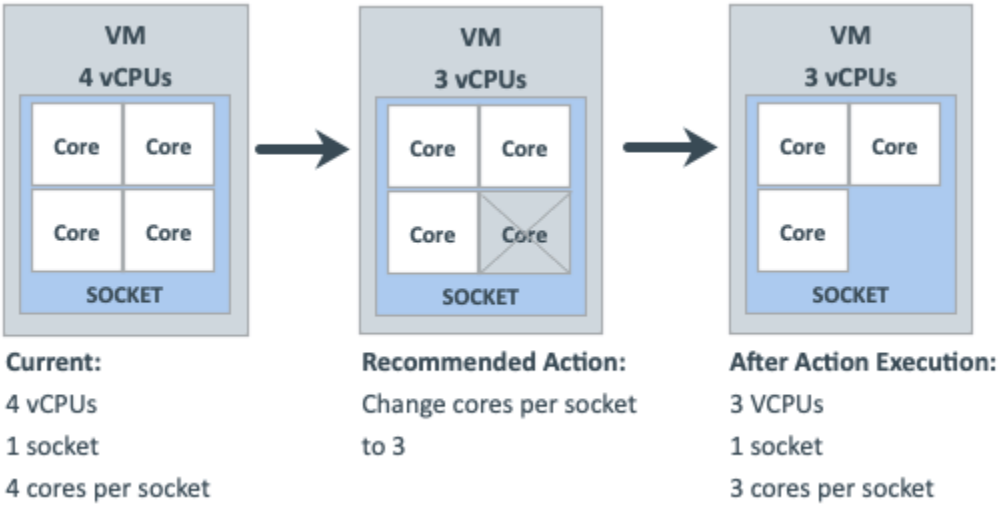
> **NOTE:**
> You can also use the **Match Host Sockets** scaling option for NUMA sensitive VMs.

For example, a VM currently has 1 socket and 4 cores per socket, and applies a policy that changes cores per socket and preserves VM sockets. Workload Optimization Manager has determined that the VM requires a change in compute capacity of 1 vCPU.

- To increase compute capacity by 1 vCPU, a resize up action changes cores per socket from 4 to 5.



- To reduce compute capacity by 1 vCPU, a resize down action changes cores per socket from 4 to 3.



Change Cores Per Socket and Match Host Sockets

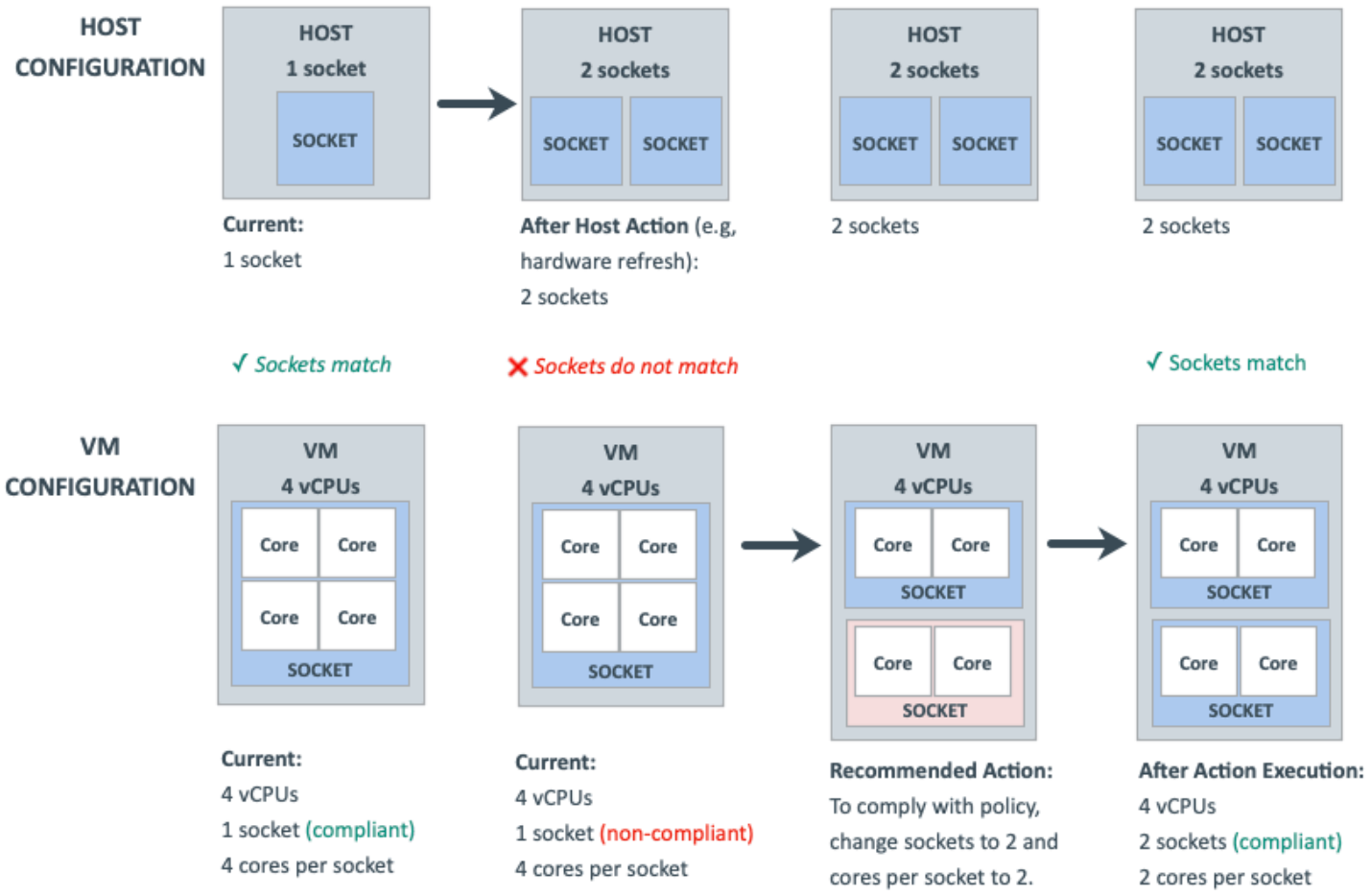
In this scaling option, Workload Optimization Manager reconfigures VM sockets to match the number of host sockets, thereby balancing vCPUs evenly across physical sockets. It also changes VM cores per socket to maintain the same compute capacity (vCPU).

This scaling option is ideal under the following scenarios:

- You have large VMs that may realize a performance benefit from reflecting the physical host CPU architecture within the Guest OS so that the application can optimize thread memory access to within a NUMA node.
- You have NUMA sensitive VMs that are migrating between hosts with different CPU architectures. Workload Optimization Manager can place the VMs on the best host and then generate an action to reconfigure the VMs to match the host sockets automatically. You can attach a schedule to the policy to automate disruptive reconfigure actions within a maintenance window.

For example, a VM currently has 1 socket and 4 cores per socket, and is on a host with 1 socket. The VM applies a policy that changes cores per socket and matches host sockets. Workload Optimization Manager has determined that the VM is already optimally sized, so a resize action is not necessary.

When the host socket value changes from 1 to 2, the VM is suddenly in violation of policy. To bring the VM into compliance while maintaining, the same vCPU capacity (since the VM is already optimally sized), Workload Optimization Manager must distribute 4 cores between 2 sockets. The result is 2 sockets and 2 cores per socket.



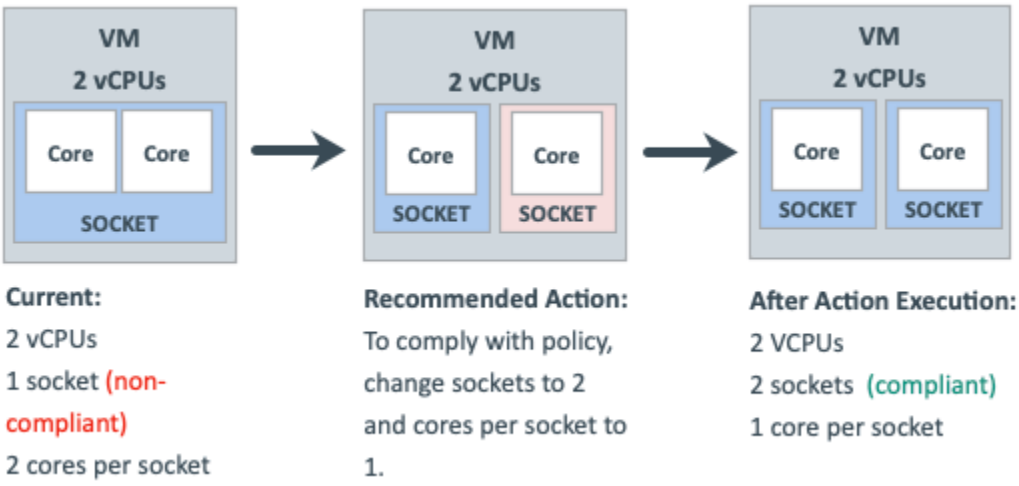
Change Cores Per Socket and Specify Sockets

In this scaling option, Workload Optimization Manager reconfigures VM sockets according to the value that you specify, and changes VM cores per socket to maintain the same compute capacity (vCPU).

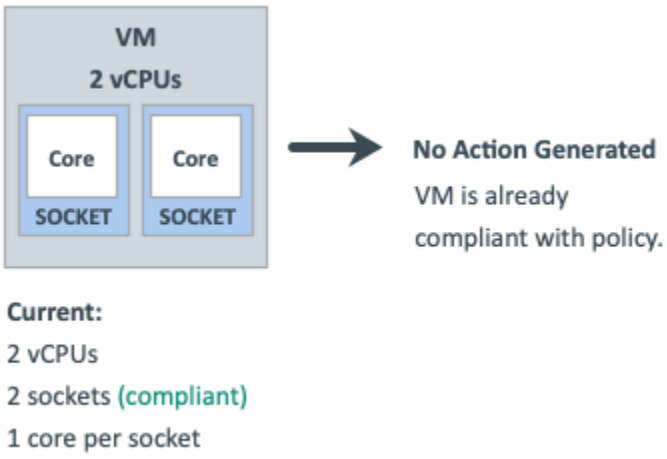
This scaling option is ideal if you have VMs that require a specific socket value for operational policy reasons (such as socket-based licensing or compliance with an application support contract policy).

For example, a VM currently has 1 socket and 2 cores per socket, and applies a policy that changes cores per socket and enforces the user-specified 2 sockets. Workload Optimization Manager has determined that the VM is already optimally sized, so a resize action is not necessary.

- Since the VM is in violation of policy, Workload Optimization Manager changes sockets from 1 to 2, and cores per socket from 2 to 1.



- When the VM is compliant with policy, no action generates.



Scaling option: change MHz legacy behavior

In this scaling option, Workload Optimization Manager adds or removes compute resources in increments of MHz (1800 MHz by default).

If a VM requires a change to compute resources, Workload Optimization Manager generates a resize vCPU action that assumes 1 core per socket, regardless of the VM's actual cores per socket.

If Workload Optimization Manager discovers the actual number of cores per socket as part of action execution, it adjusts the action.

For example, a VM currently has 4 vCPUs with 2 sockets and 2 cores per socket. Workload Optimization Manager may generate an action to resize from 4 to 5 vCPUs. However, as part of action execution, the VM socket count changes from 2 to 3, so the result is 6 vCPUs. Conversely, the same VM may have an action to resize from 4 to 3 vCPUs, but nothing changes as part of action execution.

Hypervisor support

For **VMware vSphere**, Workload Optimization Manager supports all vCPU scaling options, including changing a VM's number of sockets or cores per socket. Increasing the number of sockets is non-disruptive if CPU hot-add is enabled on a VM, while reducing the socket count always requires a restart and is therefore disruptive.

For **Hyper-V** and **Nutanix AHV**, cores per socket and hot-add features have varying degrees of support.

vCPU scaling option	vSphere	Hyper-V	Nutanix AHV (Single core)	Nutanix AHV (Multi Core)
Change virtual CPUs	Supported	Supported	Supported	Not supported by hypervisor
Change sockets – Preserve existing VM cores per socket	Supported	Supported	Supported	Supported
Change sockets – User specified cores per socket	Supported	Not supported by hypervisor	Not supported by hypervisor	Not supported by hypervisor
Change cores per socket – Preserve existing VM sockets	Supported	Not supported by hypervisor **NOTE:** Workload Optimization Manager assumes one core per socket and only changes sockets.	Not supported by Workload Optimization Manager	Not supported by Workload Optimization Manager
Change cores per socket – Match host sockets				
Change cores per socket – User specified sockets				

Tie breakers

When a single VM applies multiple conflicting policies, Workload Optimization Manager uses the following tie breakers that follow the principle of least disruptive and most conservative:

- vCPU Scaling Control

  Sockets wins over **Cores per socket** wins over **Virtual CPU** wins over **MHz legacy behavior**.

  NOTE:
  Policies that are created before the introduction of vCPU scaling controls (that is, any policy before version 3.3.7) will continue to use the **MHz legacy behavior** option but will not be enforced when policy conflicts arise. You can remove these policies or update them to use the newer scaling controls.

- Sockets setting

  Preserve existing VM cores per socket wins over **User-specified core per socket**.
- Cores Per Socket setting

  Preserve existing VM sockets wins over **User-specified socket** wins over **Match host sockets**.
- User-specified value

  The lowest value wins.
- Increment Size value

  The lowest value wins.

For example, assume a VM belonging to two groups that apply different policies. Policy A changes cores per socket and matches host sockets, while Policy B changes sockets and preserves cores per socket. In this scenario, the VM applies Policy B. Changing sockets wins over changing cores per socket because it is less disruptive.

To see which policies are in effect after the tie-break decision, set the scope to a VM or group of VMs and then click the Policies tab.

Policy cookbook

Tips:

- Use the following filters when searching for or creating VM groups:
  - Number of vCPUs
  - Number of Sockets
  - Cores per Socket
  - Target Type
  - Hot-Add Enabled
- For the least disruptive on-demand upsize of vCPU, enable hot-add on the VM and change sockets while preserving cores per socket.
- For the most precise compute resource management, change cores per socket.
- For NUMA considerations, change cores per socket and match host sockets.
- Check Guest OS application and license compatibility when changing vCPU architecture and before automating actions.

How to...

- Manage VM compute capacity by changing the number of vCPUs in increments of 2.

  A VM is reconfigured if required to use 1 core per socket, and resized by changing sockets. Actions are disruptive if the VM does not already have 1 core per socket or if hot-add is not enabled.

  1. Create a group of VMs that can have 1 core per socket and scale in sockets.
  2. Assign the group a policy with the following settings:
     - vCPU Scaling Controls
       - Change: Virtual CPU
       - Increment size: 2
     - (Optional) vCPU Resize Min/Max Threshold
- Reconfigure all odd-numbered vCPU VMs to be even-numbered, and then manage compute in even numbers of CPUs.

  A VM is reconfigured if required to use 2 cores per socket, and resized by changing sockets. Actions are disruptive if the VM does not already have 2 cores per socket or if hot-add is not enabled.

  1. Create a group of VMs that can have 2 cores per socket and scale in sockets.
  2. Assign the group a policy with the following settings:
     - vCPU Scaling Controls
       - Change: Sockets
       - User specified cores per socket: 2
     - (Optional) vCPU Resize Min/Max Threshold
- Ensure that large VMs always balance their vCPU cores across all physical host sockets (for example, NUMA VMs and Database Server VMs).

  A VM is reconfigured if its socket count does not match the host socket count. The cores per socket count may be adjusted to maintain the overall compute capacity (number of vCPUs). Resize actions are disruptive because cores per

socket change. Reconfigure actions are non-disruptive if VM sockets are increasing, hot-add is enabled, and there are no changes to cores per socket.

1. Create a group of VMs by using the filters that you require to identity typically larger-size VMs.
2. Assign the group a policy with the following settings:
   - vCPU Scaling Controls
     - Change: Cores per socket
     - Sockets: Match host sockets
   - (Optional) vCPU Resize Min/Max Threshold

- Keep VMs to 2 sockets only and manage compute by changing cores.

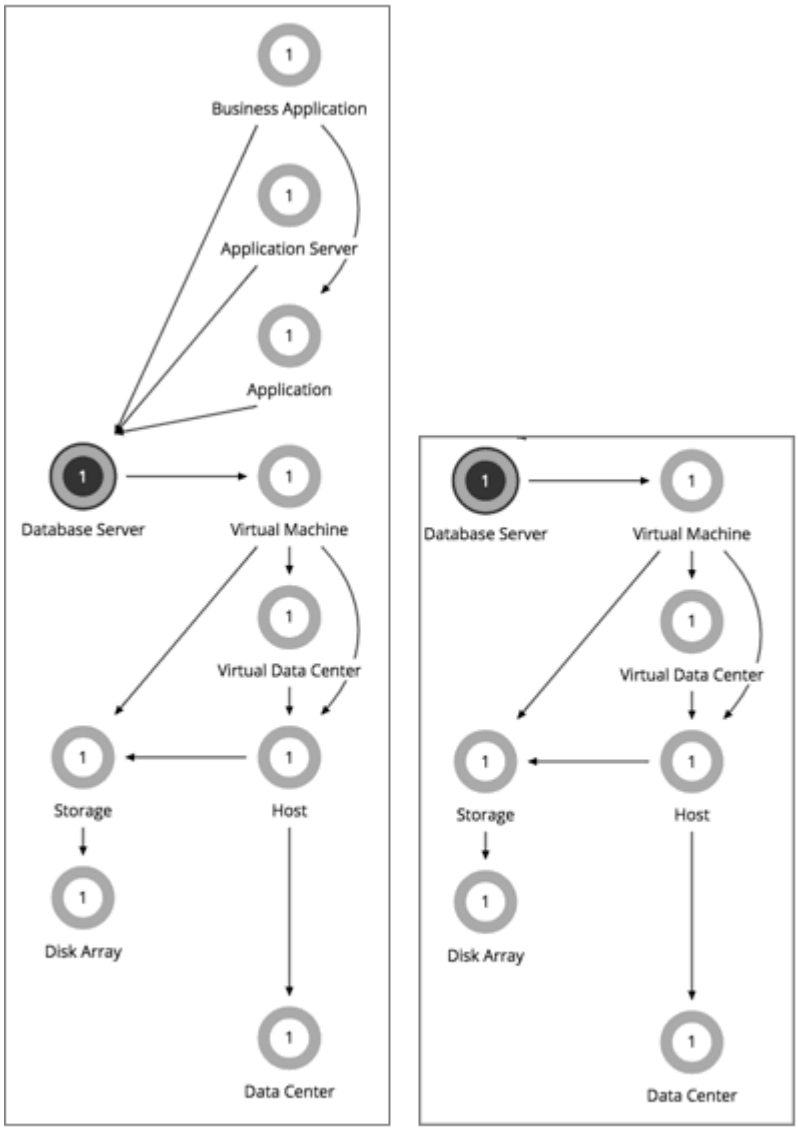
  VMs in the group will be reconfigured to 2 sockets if required, and resized by changing the cores per socket count while keeping the sockets fixed at 2, thus ensuring compliance with socket-based licensing. Resize actions are disruptive because cores per socket change. Reconfigure actions are non-disruptive if VM sockets are increasing, hot-add is enabled, and there are no changes to cores per socket.

  1. Create a VM group containing the socket-licensed VMs.
  2. Assign the group a policy with the following settings:
     - vCPU Scaling Controls
       - Change: Cores per socket
       - User specified sockets: 2
     - (Optional) vCPU Resize Min/Max Threshold

Database server (on-prem)

For on-prem, a Database Server is a database discovered through one of the associated database application targets or through APM solutions.

Synopsis



AppDynamics Database Server Target

Synopsis	
Provides:	▪ Response Time, Transactions, DBmem, Cache Hit Rate, and TransactionLog to end users ▪ Connections to Application Components
Consumes:	VM resources, including VCPU, VMem, and VStorage
Discovered through:	▪ AppDynamics targets ▪ Database Server targets ▪ Dynatrace MySQL and SQL Server processes ▪ NewRelic Infrastructure Integration (NRI): MySql, SQL Server, MongoDB, OracleDB

Monitored resources

Workload Optimization Manager monitors the following resources:

- Virtual memory (vMem)

  Virtual memory (vMem) is the measurement of memory that is in use.
- Transaction

  Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.
- Database memory (DBMem)

  Database memory (or DBMem) is the measurement of memory that is utilized by a Database Server.
- Connection

  Connection is the measurement of database connections utilized by applications.
- DB cache hit rate

  DB cache hit rate is the measurement of Database Server accesses that result in cache hits, measured as a percentage of hits versus total attempts. A high cache hit rate indicates efficiency.

Actions

Workload Optimization Manager supports the following actions:

Resize

Resize the following resources:

- Connections

  Workload Optimization Manager uses connection data to generate memory resize actions for on-prem Database Servers.
- Database memory (DBMem)

  Actions to resize database memory are driven by data on the Database Server, which is more accurate than data on the hosting VM. Workload Optimization Manager uses database memory and cache hit rate data to decide whether resize actions are necessary.

  A high cache hit rate value indicates efficiency. The optimal value is 100% for on-prem (self-hosted) Database Servers, and 90% for cloud Database Servers. When the cache hit rate reaches the optimal value, no action generates even if database memory utilization is high. If utilization is low, a resize down action generates.

  When the cache hit rate is less than the optimal value but database memory utilization remains low, no action generates. If utilization is high, a resize up action generates.
- Transaction log

  Resize actions based on the transaction log resource depend on support for virtual storage in the underlying hypervisor technology.

  Currently, Workload Optimization Manager does not support resize actions for Oracle and Database Servers on the Hyper-V platform (due to the lack of API support for virtual storage).

On-prem database server policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

Action	Default mode
Resize	Manual

Action	Default mode
Resize DBMem (Up/ Down)	Manual

You can use Action Scripts *(on page 1070)* and third-party orchestrators (such as ServiceNow) for action orchestration.

Resizing sensitivity

Workload Optimization Manager uses a percentile of utilization over the specified observation period. This gives sustained utilization and ignores short-lived bursts.

Workload Optimization Manager uses these settings to calculate utilization percentiles for DB Memory and DB Cache Hit Rate. It then recommends actions to improve utilization based on the observed values for a given time period.

- Aggressiveness

Attribute	Default value
Aggressiveness	95th Percentile

When evaluating performance, Workload Optimization Manager considers resource utilization as a percentage of capacity. The utilization drives actions to resize the available capacity either up or down. To measure utilization, the analysis considers a given utilization percentile. For example, assume a 95th percentile. The percentile utilization is the highest value that 95% of the observed samples fall below. Compare that to average utilization, which is the average of *all* the observed samples.

Using a percentile, Workload Optimization Manager can recommend more relevant actions. For scheduled policies, the more relevant actions will tend to remain viable when their execution is put off to a later time.

For example, consider decisions to reduce capacity. Without using a percentile, Workload Optimization Manager never resizes below the recognized peak utilization. Assume utilization peaked at 100% just once. Without the benefit of a percentile, Workload Optimization Manager will not reduce resources for that Application Component.

With **Aggressiveness**, instead of using the single highest utilization value, Workload Optimization Manager uses the percentile you set. For the previous example, assume a single burst to 100%, but for 95% of the samples, utilization never exceeded 50%. If you set **Aggressiveness** to 95th Percentile, then Workload Optimization Manager can see this as an opportunity to reduce resource allocation.

In summary, a percentile evaluates the sustained resource utilization, and ignores bursts that occurred for a small portion of the samples. You can think of this as aggressiveness of resizing, as follows:

  - 99th Percentile – More performance. Recommended for critical Application Components that need maximum guaranteed performance at all times, or those that need to tolerate sudden and previously unseen spikes in utilization, even though sustained utilization is low.
  - 95th Percentile (Default) – The recommended setting to achieve maximum performance and savings. This assures performance while avoiding reactive peak sizing due to transient spikes, thus allowing you to take advantage of the elastic ability of the cloud.
  - 90th Percentile – More efficiency. Recommended for Application Components that can stand higher resource utilization.

By default, Workload Optimization Manager uses samples from the last 14 days. Use the **Max Observation Period** setting to adjust the number of days. To ensure that there are enough samples to analyze and drive resize actions, set the **Min Observation Period**.

- Max observation period

Attribute	Default value
Max Observation Period	Last 14 Days

To refine the calculation of resource utilization percentiles, you can set the sample time to consider. Workload Optimization Manager uses historical data from up to the number of days that you specify as a sample period. If the Database Server has fewer days' data then it uses all of the stored historical data.

You can make the following settings:

- Less Elastic – Last 30 Days
- Recommended – Last 14 Days
- More Elastic – Last 7 Days or Last 3 Days

Workload Optimization Manager recommends an observation period of 14 days so it can recommend resize actions more often. Since Database Server resizing is minimally disruptive, resizing often should not introduce any noticeable performance risks.

- Min observation period

Attribute	Default value
Min Observation Period	None

This setting ensures historical data for a minimum number of days before Workload Optimization Manager will generate an action based on the percentile set in **Aggressiveness**. This ensures a minimum set of data points before it generates the action.

Especially for scheduled actions, it is important that resize calculations use enough historical data to generate actions that will remain viable even during a scheduled maintenance window. A maintenance window is usually set for "down" time, when utilization is low. If analysis uses enough historical data for an action, then the action is more likely to remain viable during the maintenance window.

- More Elastic – None
- Less Elastic – 1, 3, or 7 Days

Transaction SLO

Transaction SLO determines the upper limit for acceptable transactions per second. When the number of transactions reaches the given value, Workload Optimization Manager sets the risk index to 100%.

Attribute	Default setting/value
Enable Transaction SLO	Off
Transaction SLO	None If you enable SLO, Workload Optimization Manager uses the default value of 10. You can change this to a different value.

Response time SLO

Response time SLO determines the upper limit for acceptable response time (in milliseconds). If response time reaches the given value, Workload Optimization Manager sets the risk index to 100%.

Attribute	Default setting/value
Enable Response Time SLO	Off Workload Optimization Manager estimates SLO based on monitored values.
Response Time SLO [ms]	None If you enable SLO, Workload Optimization Manager uses the default value of 2000. You can change this to a different value.

DBMem scaling increment

This increment specifies how many units to add or subtract when scaling DBMem.

Attribute	Default value
DBMem Scaling Increment (MB)	128

Do not set the increment value to be lower than what is necessary for the database server to operate. If the increment is too low, then it's possible there would be insufficient DBMem. When reducing allocation, Workload Optimization Manager will not leave a database server with less than the increment value. For example, if you use the default 128, then Workload Optimization Manager cannot reduce DBMem to less than 128 MB.

DBMem utilization

The utilization that you set here specifies the percentage of the existing capacity that Workload Optimization Manager will consider to be 100% of capacity.

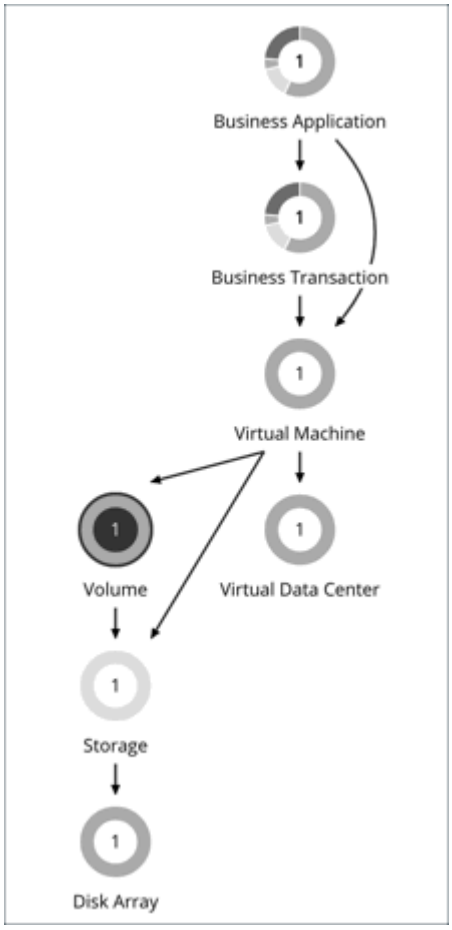
Attribute	Default value
DBMem Utilization (%)	100

For example, a value of 80 means that Workload Optimization Manager considers 80% utilization to be 100% of capacity. Workload Optimization Manager recommends actions that avoid utilization beyond the given value.

Volume (on-prem)

On-prem volumes represent VM disks discovered by hypervisor targets. A VM will have one volume for each configured disk and another volume (representing the configuration) that always moves with Disk 1.

Synopsis



Synopsis	
Provides:	Storage resources for VMs to use Set the scope to a volume and view the Entity Information chart to see a list of VM-related files (such as VMDKs) contained in the volume. Set the scope to a VM to see a list of volumes attached to the VM.
Consumes:	Datacenter resources
Discovered through:	Hypervisor targets

Actions

Workload Optimization Manager supports the following actions:

- Move

  Move a VM's volume (virtual storage) due to excess utilization of the current datastore, or for more efficient utilization of datastores in the environment.

  Points to consider:

  - The default global policy includes a setting that directs Workload Optimization Manager to use relevant metrics when analyzing and recommending actions for volumes. For details, see Enable Analysis of On-prem Volumes *(on page 1052)*.
  - Workload Optimization Manager will not recommend moving a volume to a datastore that is currently in maintenance mode. Any volume in that datastore should move to an active datastore (for example, via vMotion).

- Reconfigure

  Reconfigure a VM's volume (virtual storage) to comply with placement policies.

On-prem volume policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Placement policies

By default, all on-prem volumes associated with a storage will move together rather than independently. You can create placement policies to place individual volumes on groups of storage. To ensure successful placement, be sure to also turn on the setting `Enable Analysis of On-prem Volumes` in the default global policy.

For more information, see Creating Placement Policies *(on page 1044)* and Enable Analysis of On-prem Volumes *(on page 1052)*

Automation workflow

Action	Default mode
Move	Manual

Cloud storage tiers

This policy setting works with plans that simulate migration of on-prem volumes to the cloud. When you create the policy, be sure to set the scope to on-prem volumes and then select the cloud storage tiers that they can migrate to. Workload Optimization Manager treats these tiers as constraints when you run a Migrate to Cloud plan that includes the volumes defined in the policy.

Attribute	Default value
Cloud Storage Tiers	None

Click **Edit** to set your preferences. In the new page that displays, expand a **cloud tier** (a family of instance types, such as *Premium*) to see individual instance types.

Select your preferred instance types or cloud tiers, or clear the ones that you want to avoid. After you save your changes, the main page refreshes to reflect your selections.

Virtual data center (private cloud)

A virtual data center (vDC) is a collection or pool of resources that groups the resources around specific requirements or business needs. In private cloud environments, Workload Optimization Manager discovers the infrastructure that provides resources to the cloud, and the workloads that run on the cloud. To manage these resources, private clouds organize the infrastructure into Provider and Consumer virtual data centers.

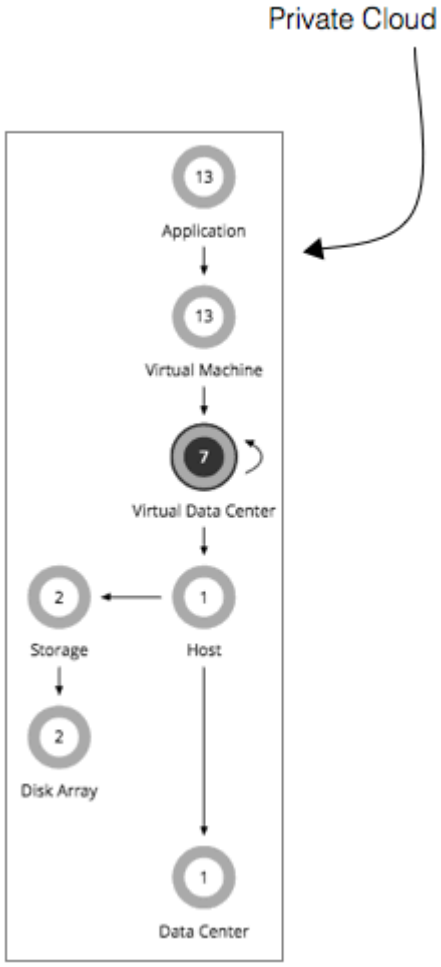
NOTE:
Different targets use different names to refer to virtual data centers. In the Workload Optimization Manager supply chain, these entities are all represented by Consumer and Provider VDCs, as follows:

Workload Optimization Manager	vCenter Server	VMM
Consumer VDC	Resource Pool (Child)	Tenant or TenantQuota
Provider VDC	Resource Pool (Root)	Cloud

Provider virtual data centers

A provider virtual data center (vDC) is a collection of physical resources (hosts and data stores) within a cloud stack. The cloud administrator has access to these resources, and defines the data center members. A Provider vDC is created to manage resources that will be allocated to external customers through one or more Consumer vDCs.

Synopsis



A Provider vDC gains its budget by selling resources to the Consumer vDCs that it hosts. If utilization falls off, the data center loses budget. Ultimately, if the budget isn't enough to pay for the services it consumes, Workload Optimization Manager will recommend decommissioning the Provider vDC.

Synopsis	
Provides:	Physical resources (hosts and datastores) to Consumer vDCs
Consumes:	Hosts and datastores from the physical infrastructure
Discovered through:	Private Cloud Stack Managers

Monitored resources

Workload Optimization Manager monitors the following resources:

- Memory (Mem)

  Memory is the measurement of memory that is reserved or in use.
- CPU

  CPU is the measurement of CPU that is reserved or in use.
- Storage

  Storage is the utilization of the storage attached to the entity.

Actions

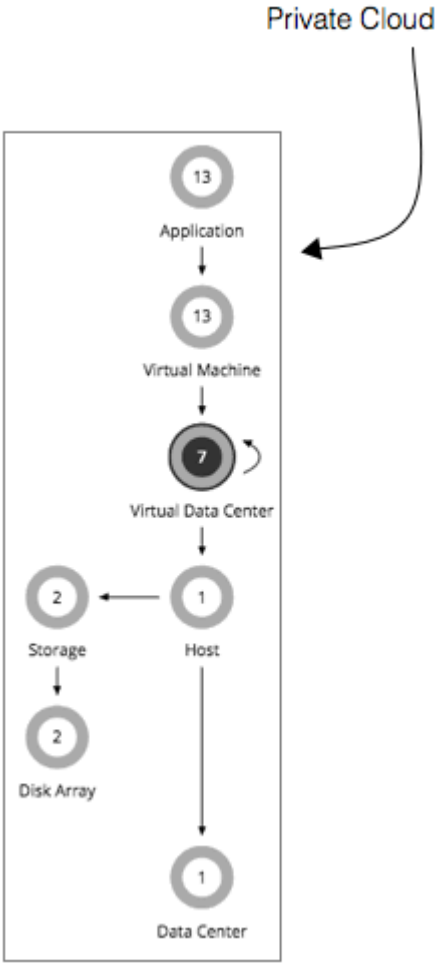
None

Workload Optimization Manager does not recommend actions for a virtual datacenter. Instead, it recommends actions for the entities that provide resources to the virtual datacenter.

Consumer virtual data centers

A Consumer Virtual Data center (vDC) is a collection of resources that are available for external customers to manage workload through the private cloud. It is an environment customers can use to store, deploy, and operate virtual systems. Consumer Data centers use the resources supplied by a Provider Data center.

Synopsis



A Consumer vDC gains its budget as a function of its activity. The higher the utilization of the vDC, the more Workload Optimization Manager assumes the vDC is selling its services to a user. If utilization is high enough on a Consumer vDC, Workload Optimization Manager can increase resources for the vDC. If utilization falls off, Workload Optimization Manager can reduce resource capacity, or ultimately recommend terminating the vDC.

Workload Optimization Manager can also resize VMs through the Consumer vDC in response to changes in VM utilization.

Synopsis	
Provides:	Resources to host virtual systems
Consumes:	Provider vDC
Discovered through:	Cloud Stack Managers

While users can see some of the physical resources that support the Consumer vDC, consumer-level users cannot modify these physical resources. Users of Consumer vDCs make changes to how the virtual devices are deployed in that environment, but they must ask the Provider vDC administrator to add more physical resources to be used by the Consumer vDC. Likewise, Workload Optimization Manager can change resources on the VMs running in the vDC, but it does not make any changes to physical resources through this vDC.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Memory (Mem)

  Memory is the measurement of memory that is reserved or in use.
- CPU

  CPU is the measurement of CPU that is reserved or in use.
- Storage

  Storage is the utilization of the storage attached to the entity.

Actions

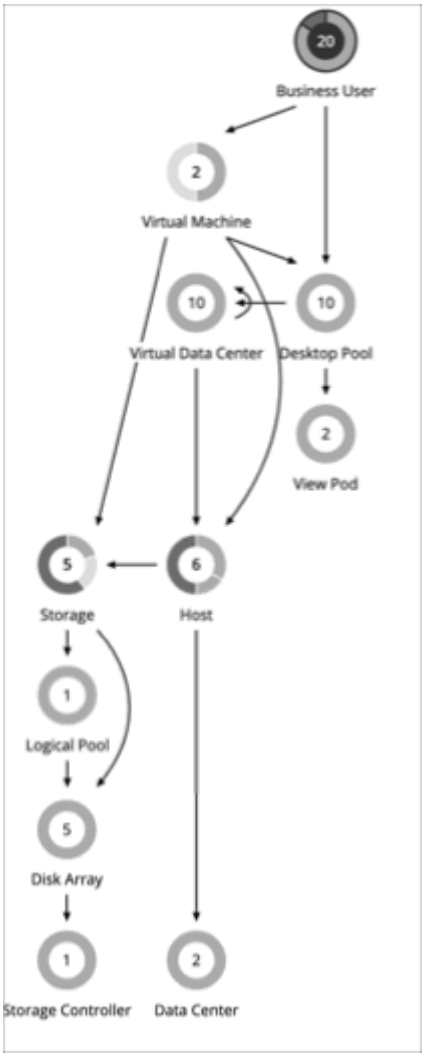
Workload Optimization Manager does not recommend actions to perform on a Consumer vDC. Instead, it recommends actions to perform on the entities running in the Provider vDC.

Business user

For Virtual Desktop Infrastructure (VDI) environments, a business user is a user account that is entitled to launch one or more active VDI sessions. As it discovers desktop pools, Workload Optimization Manager creates business user entities for each user that is entitled to a pool. One business user can be entitled to more than one desktop pool.

To properly work with business user entities, Workload Optimization Manager discovers user information through the LDAP server that manages users for the VDI environment. Note that the account Workload Optimization Manager uses to connect to the LDAP server must be trusted for the same domains as are the users in your environment.

Synopsis



The Supply Chain shows relationships of business users to Desktop Pools and also to VMs. One business user can have access to multiple Desktop Pools. When a business user has an active session, the Supply Chain shows a direct link between the user and the VM that hosts the session. However, Workload Optimization Manager does not consider this direct connection when analyzing compute resources. Instead, business users utilize Desktop Pool resources, and the Desktop Pools use compute resources from the underlying Virtual Datacenters.

Synopsis	
Provides:	N/A
Consumes:	Resources from the underlying desktop pools, including: ▪ Sessions ▪ Pool Memory ▪ Pool Storage ▪ Pool CPU When a business user has an active session, the supply chain shows it in relation to the VM that hosts the session. The business user consumes the VM's compute resources to support the session requirements for ImageCPU, ImageMem, and ImageStore resources.
Discovered through:	The LDAP server that manages these users.

Synopsis	
	You can specify the LDAP server as part of the target configuration, or Workload Optimization Manager can discover it in association with the VDI target.

Monitored resources

Workload Optimization Manager monitors the following resources:

- ImageCPU

  ImageCPU is the measurement of CPU that is utilized by a user's desktop images, expressed as a percentage of CPU capacity.
- ImageMem

  ImageMem is the measurement of memory that is utilized by a user's desktop images, expressed as a percentage of memory capacity.
- ImageStorage

  ImageStorage is the measurement of storage that is utilized by a user's desktop images, expressed as a percentage of storage capacity.

Business user actions

Workload Optimization Manager supports the following actions:

Move

Move a business user between desktop pools to address:

- Resource congestion on the image

  When utilization is consistently near capacity for image resources, Workload Optimization Manager can recommend moving a business user to a desktop pool that serves larger images.
- Resource congestion on the desktop pool

  When utilization is consistently near capacity for the desktop pool, Workload Optimization Manager can recommend moving a business user to a desktop pool that has more available resources.

NOTE:
To support moves, you must configure placement policies that merge *similarly configured* desktop pools. For details, see Desktop Pool Placement Policies *(on page 784)*.

Business user policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

For details about business user actions, see Business User Actions *(on page 780)*.

Action	Default mode
Move	Manual

Image target utilization

Workload Optimization Manager tracks utilization of desktop image resources for the business users in your Virtual Desktop Infrastructure (VDI) environment.

Attribute	Default value
Image CPU Target Utilization	70 The target utilization as a percentage of CPU capacity.
Image MEM Target Utilization	70 The target utilization as a percentage of memory capacity.
Image Storage Target Utilization	70 The target utilization as a percentage of storage capacity.

Aggressiveness and observation period

Workload Optimization Manager uses these settings to calculate utilization percentiles. It then recommends actions to improve utilization based on the observed values for a given time period.

- Aggressiveness

Attribute	Default value
Aggressiveness	95th Percentile

  When evaluating utilization of compute and storage resources, Workload Optimization Manager considers a given utilization percentile. For example, assume a 95th percentile. The maximum utilization would be the highest value that 95% of the observed samples fall below.

  Using a percentile, Workload Optimization Manager can recommend more relevant actions, so that analysis can better exploit elasticity in your environment. A percentile evaluates the sustained resource utilization, and ignores bursts that occurred for a small portion of the samples. You can think of this as aggressiveness of resizing, as follows:

  - 100th Percentile – The least aggressive, recommended for critical workloads that need maximum guaranteed performance at all times.
  - 95th Percentile (Default) – The recommended setting to achieve maximum performance and savings.
  - 90th Percentile – The most aggressive, recommended for non-production workloads that can stand higher resource utilization.

- Max observation period

Attribute	Default value
Max Observation Period	Last 30 Days

  To refine the calculation of resource utilization, you can set the sample time to consider. Workload Optimization Manager uses historical data from up to the number of days that you specify as a sample period. (If the database has fewer days' data then it uses all of the stored historical data.)

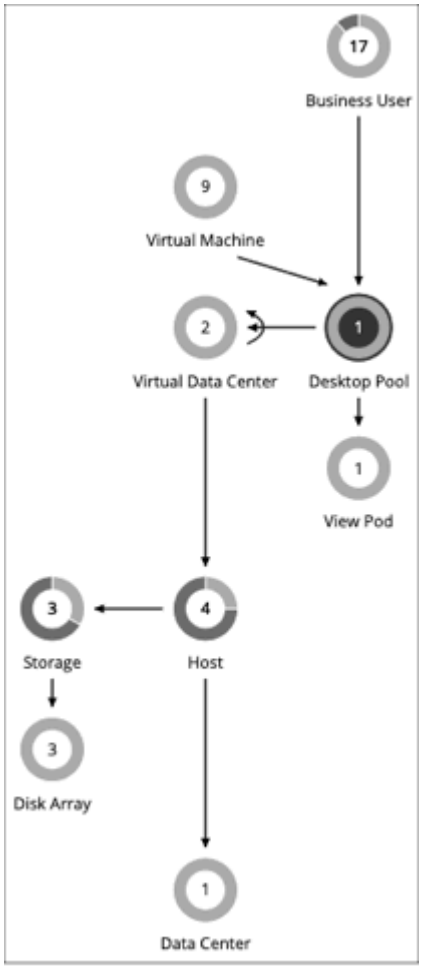
  A shorter period means there are fewer data points to account for when Workload Optimization Manager calculates utilization percentiles. This results in more dynamic, elastic moves to different Desktop Pools, while a longer period results in more stable or less elastic moves. You can make the following settings:

  - Less Elastic – Last 90 Days
  - (Recommended) More Elastic – Last 30 Days
  - Most Elastic – Last 7 Days

Desktop pool

For Virtual Desktop Infrastructure (VDI) environments, a desktop pool is a collection of desktops that users can select from. The desktop pool can provide logical grouping of desktops according to user roles, assignment type (dedicated or floating), and the source of resources (physical host or VM).

Synopsis



The desktop pool gets compute and storage resources from the underlying Virtual Datacenter. For VMware Horizon View, the VDI architecture includes one or more vCenter Server instances. When it discovers the Horizon View target, Workload Optimization Manager also discovers the supporting vCenter Server instances, and their corresponding Virtual Datacenters. These are the source of compute and storage resources for the associated desktop pools.

Synopsis	
Provides:	Resources for Business Users to use, including: ▪ Pool Memory ▪ Pool CPU ▪ Active Sessions
Consumes:	▪ Compute and storage resources from the associated Virtual Datacenters ▪ Sessions from the underlying View Pod
Discovered through:	VDI management targets For VMware Horizon View, the target is the View Connection Server.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Pool CPU

  Pool CPU is the measurement of CPU that is utilized by active sessions in a desktop pool.
- Pool memory

  Pool memory is the measurement of memory that is utilized by active sessions in a desktop pool.
- Pool storage

  Pool storage is the measurement of storage that is utilized by active sessions in a desktop pool.
- Active sessions

  Active sessions is the number of active sessions in a desktop pool, expressed as a percentage of the pool's capacity.
- Total sessions

  Total sessions is the number of active and disconnected (non-terminated) sessions in a desktop pool, expressed as a percentage of the pool's capacity.

Actions

None

Workload Optimization Manager does not recommend actions for a desktop pool. It recommends actions for the business users running active sessions in the pool.

Desktop pool policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

None

Workload Optimization Manager does not recommend actions for a desktop pool. It recommends actions for the business users running active sessions in the pool.

Observation settings

Workload Optimization Manager uses these settings to decide whether to move Business Users from one desktop pool to another.

- Daily observation windows

Attribute	Default value
Daily observation windows	3 windows per day

When evaluating utilization of pool resources, Workload Optimization Manager divides each day into different observation windows, calculates an average for each, and uses the highest value. In this way, Workload Optimization Manager can account for high-use periods in the day to base calculations off of the most representative usage of the desktop images.

Assume three observation windows:

Window	Time range	Average utilization
W1	00:00 – 08:00	10%
W2	08:00 – 16:00	80%
W3	16:00 – 24:00	40%

Average utilization for this day *without* the benefit of observation windows would be 44%. By using observation windows you can see that the representative utilization of pool resources is closer to 80%. That is because Workload Optimization Manager discovers an average utilization of 80% during the high-usage time of day.

When calculating whether to move business users from one desktop pool to another, Workload Optimization Manager averages the observation windows over the time you set for the Max Observation Period. For this reason, you should try to set up observation windows that capture the best representation of work habits amongst your business users.

- Max observation period

Attribute	Default value
Max Observation Period	Last 7 Days

To refine the calculation of resource utilization, you can set the sample time to consider. Workload Optimization Manager uses historical data from up to the number of days that you specify as a sample period. (If the Workload Optimization Manager database has fewer days' data, then it uses all of its stored historical data.)

A shorter period means there are fewer data points to account for when Workload Optimization Manager calculates utilization. This results in more dynamic, elastic resizing, while a longer period results in more stable or less elastic resizing. You can make the following settings:

  - Less Elastic – Last 30 Days
  - Recommended – Last 7 Days
  - More Elastic – Last 3 Days

Pool utilization

These settings affect the actions Workload Optimization Manager recommends as it manages business users and active accounts on the desktop pool. Workload Optimization Manager recommends actions that avoid using these resources beyond the given settings.

Attribute	Default value
Pool Memory Utilization	95
Pool Storage Utilization	95
Pool CPU Utilization	95

The values you set here specify what percentage of the existing capacity that Workload Optimization Manager will consider to be 100% of capacity. For example, setting 70 for Desktop Pool Pool CPU Utilization means that Workload Optimization Manager considers 70% utilization of that CPU to be 100% of capacity and 35% utilization to be 50% of capacity.

Placement policies

Under some circumstances, you can have Business Users who need large-scale desktop images. This appears as users with high utilization of the image resources. In this case, Workload Optimization Manager can recommend moving the Business Users to a different desktop pool that serves up large-scale images.

To support moving Business Users, you must create a placement policy that merges desktop pools. Be sure to merge only desktop pools that are *similarly configured* – they should run the same operating system and applications, and differ only in allocated memory and/or CPU.

To merge desktop pools:

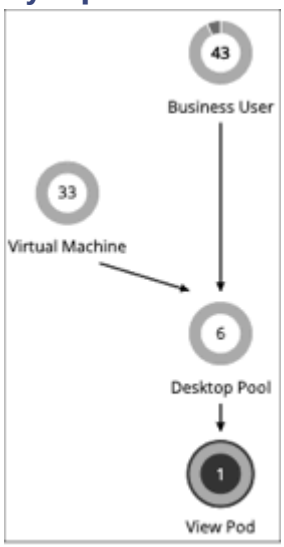
1. Create a new placement policy.
2. Choose **Merge** as the policy type.
3. For the consumer type to merge, choose **Desktop Pool**.
4. Choose the pools that you want to merge.
5. Save the policy.

For more information, see Creating Placement Policies *(on page 1044)*.

View pod

For Virtual Desktop Infrastructure (VDI) environments, a View Pod groups together a given set of Desktop Pools.

Synopsis



Synopsis	
Provides:	Active Sessions
Consumes:	N/A
Discovered through:	VDI Management targets For VMware Horizon View, the target is the View Connection Server.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Active sessions

  Active sessions is the number of active sessions in a desktop pool, expressed as a percentage of the pool's capacity.
- Total sessions

  Total sessions is the number of active and disconnected (non-terminated) sessions in a desktop pool, expressed as a percentage of the pool's capacity.

Actions

None

Workload Optimization Manager does not recommend actions for a view pod. Instead, it recommends actions for the business users that are running active sessions.

Active session capacity for view pods

Each View Pod entity has a set capacity of active sessions. By default, Workload Optimization Manager assumes a capacity of 8,000. So that Workload Optimization Manager can generate reliable actions for Business User entities, you must set this capacity to match the active session capacity that your Horizon administrator has deployed for the given view pod.

Once you know the correct active session capacity for your view pod, create an automation policy that sets the capacity. For complete information about creating automation policies, see Creating Automation Policies *(on page 1053)*. For information about view pod policies, see View Pod Policies *(on page 786)*.

1. Create a new automation policy.

   Navigate to the Settings Page and choose **Policies**. Then click **NEW AUTOMATION POLICY**, and select View Pod as the policy type. Be sure to name the new policy.

2. Set the policy scope to your view pod.

   To define its scope, you assign a group to the policy. You will have to create the group for this view pod:

   - Expand the **SCOPE** section and then click **ADD VIEW POD GROUPS**.
   - Choose the group that contains only the view pod you want to configure.

     If it has already been created, choose the group from the list. If the group does not appear, click **NEW GROUP** to create a static group that includes only the view pod you want to configure. For more information about creating groups, see Creating Groups *(on page 1039)*.

     Choose the group you want and click **SELECT**. This returns you to the Configure View Pod Policy fly-out.

3. Set the view pod capacity.

   Expand the **UTILIZATION CONSTRAINTS** section and click **ADD UTILIZATION CONSTRAINT**. From the drop-down list, choose Active Sessions Capacity. In the capacity field, enter the capacity that you have calculated for your desktop pools.

4. Save your work

   When you're done, be sure to click **SAVE AND APPLY**.

View pod policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

None

Workload Optimization Manager does not recommend actions for a view pod. Instead, it recommends actions for the business users that are running active sessions.

Active sessions capacity

This setting controls the number of active sessions a given view pod can support.

Attribute	Default value
Active Sessions Capacity	8000

For each view pod, you should set this value to match the active session capacity that has been deployed in your VDI environment for the given view pod. For more information, see Active Session Capacity for View Pods *(on page 785)*.

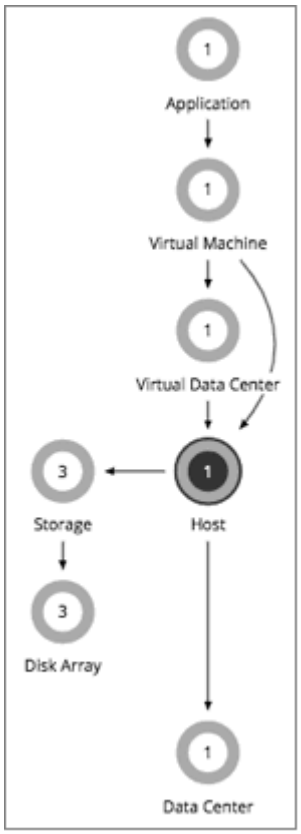
Host

For on-prem environments, a host is a server that runs processes, including hypervisor processes to host virtual workloads. Note that a host is not necessarily a physical piece of hardware. A VM can be set up as a server that runs a hypervisor, and in turn it can host other VMs within its processing space. However, it is typical to use physical hardware as your hosts.

NOTE:
To support vSAN storage in your environment, you can deploy HCI Hosts. Workload Optimization Manager discovers the vSAN as a storage entity that consumes resources from the underlying hosts. For more information, see vSAN Storage *(on page 800)*.

On the public cloud a host is an availability zone. This is where your cloud workloads run. For details, see Zone *(on page 736)*.

Synopsis



A host gains its budget by selling resources to the workloads that run on it. The more workloads running on a host, the more budget the host has to purchase storage and datacenter resources. If utilization of a host is high enough, Workload Optimization Manager can recommend that you provision a new one. If utilization falls off, the host loses budget. Ultimately, if the budget is not enough to pay for the services it consumes, Workload Optimization Manager will recommend to suspend or power off the host.

Synopsis	
Provides:	Host resources for VMs to use, including: ▪ Memory ▪ CPU ▪ IO throughput ▪ Net (network throughput) ▪ Swap ▪ Ballooning (sharing of memory among hosted VMs) ▪ CPU Ready Queue

Synopsis	
Consumes:	Datacenter resources (physical space, cooling, etc.) and storage
Discovered through:	Hypervisor targets For some hypervisor vendors, the host is the target, and for others the hosts are managed by the specified target.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Memory (Mem)

  Memory is the measurement of memory that is reserved or in use.
- CPU

  CPU is the measurement of CPU that is reserved or in use.
- IO

  IO is the utilization of a host's IO adapters.
- Net

  Net is the utilization of data through the host's network adapters.
- Swap

  Swap is the measurement of a host's swap space that is in use.
- Balloon

  Balloon is the measurement of memory that is shared by VMs running on a host.
- CPU ready

  CPU ready is the measurement of a host's ready queue capacity that is in use.

For hosts discovered via vCenter targets, the following resources are monitored:

- Energy

  Energy is the measurement of electricity used by a given entity over a period of time, expressed in watt-hours (Wh).
- Carbon footprint

  Carbon footprint is the measurement of carbon dioxide equivalent (CO2e) emissions for a given entity. Workload Optimization Manager measures carbon footprint in grams.

Actions

Workload Optimization Manager supports the following actions:

- Start

  Start a suspended host when there is increased demand for physical resources.
- Provision

  Provision a new host in the environment when there is increased demand for physical resources. Workload Optimization Manager can then move workloads to that host.
- Suspend

  When physical resources are underutilized on a host, move existing workloads to other hosts and then suspend the host.
- Reconfigure

  Workload Optimization Manager generates this action in response to changing demand for software licenses. For details, see License policy *(on page 1048)*.

NOTE:
Workload Optimization Manager discovers VMware HA configurations in clusters, and considers the reserved resources in its calculations. For tolerated host failures, or a reserved percentage of cluster resources, Workload Optimization Manager automatically sets utilization constraints for that cluster. If you configure a failover host, Workload Optimization Manager reserves that host for HA and will not move VMs to it.

DRS automation settings

Workload Optimization Manager automatically discovers DRS automation settings for vSphere hosts managed through vCenter. When you set the scope to a vSphere host and then view the Entity Information chart, the following information displays:

- Vendor Automation Mode

  The chart shows the automation mode discovered from vCenter – Not Automated, Partially Automated, or Fully Automated.

- Vendor Migration Level

  Workload Optimization Manager assigns a vendor migration level based on the migration level discovered from vCenter. The chart only shows the assigned migration level (i.e., the Workload Optimization Manager Vendor Migration Level).

Workload Optimization Manager vendor migration level	vCenter migration level
1 (Conservative)	5
2 (Less Conservative)	4
3 (Moderate)	3
4 (Less Aggressive)	2
5 (Aggressive)	1

Host policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

For details about host actions, see Host Actions *(on page 788)*.

Action	Default mode	vCenter	Hyper-V	UCS (blades only)	Power-VM
Start	Recommend	Automatable	Recommend	Automatable	N/A
Suspend	Recommend	Automatable	Recommend	Automatable	N/A
Provision	Recommend	Recommend	Recommend	Automatable	N/A
Reconfigure	Recommend	Recommend	Recommend	Recommend	N/A

You can use Action Scripts for action orchestration.

For ServiceNow:

- Host provision actions will not generate a CR.
- For host suspend actions to succeed, it must be enabled in the given hypervisor, and there must be no VMs currently running on that host.

Maintenance automation avoidance

Attribute	Default setting
Maintenance Automation Avoidance	30 minutes

The Maintenance Automation Avoidance setting applies to vCenter environments with DRS clusters. Workload Optimization Manager uses this setting when:

- Workload Optimization Manager actions to move VMs from one host to another are automated.
- The DRS automation level is *Fully Automated*, regardless of migration threshold.

  NOTE:
  Workload Optimization Manager automatically discovers DRS automation levels and migration thresholds and displays them in the Entity Information chart for hosts.

- Host maintenance is in effect.

This setting prevents action conflicts between Workload Optimization Manager and DRS.

When a host enters maintenance mode, DRS starts to move VMs on the host to other hosts to prepare for maintenance. In response, Workload Optimization Manager clears all pending actions to and from the host. For example, assume a cluster with Host_01, Host_02, and Host_03. When Host_01 enters maintenance mode, Workload Optimization Manager removes the following pending actions from the system:

- Move a VM on Host_01 to Host_02.

  This prevents a potential conflict with a DRS action that moves the VM to Host_03.
- Move a VM on Host_02 to Host_01.

  Since Host_02 and Host_03 are not in maintenance mode, Workload Optimization Manager might recommend moving the VM from Host_02 to Host_03 as an alternative action.

In addition, Workload Optimization Manager treats a host entering or in maintenance mode as uncontrollable and stops generating actions for the host. The host remains uncontrollable after it leaves maintenance mode, but is within the Maintenance Automation Avoidance period that you specified. During rolling host maintenance operations on DRS clusters, where hosts undergo maintenance on a staggered basis, this gives DRS a window (30 minutes by default) to move VMs from hosts entering maintenance to hosts that have recently left maintenance, thereby avoiding any potential conflict.

When the Maintenance Automation Avoidance period is over, Workload Optimization Manager treats the host as controllable and resumes action generation. At this stage, it is assumed that all critical DRS activities on the host have been completed, so Workload Optimization Manager actions should be safe to execute.

The following table summarizes Workload Optimization Manager's response at various stages of maintenance.

Maintenance Status	DRS activities	Host status in Workload Optimization Manager	Workload Optimization Manager pending actions	Workload Optimization Manager new actions
Host is entering maintenance mode.	Increased number of DRS activities moving VMs away from the host entering maintenance	Uncontrollable (maintenance)	Removed from the system	Not generated

Maintenance Status	DRS activities	Host status in Workload Optimization Manager	Workload Optimization Manager pending actions	Workload Optimization Manager new actions
Host is in maintenance mode.	Maintenance tasks on the host	Uncontrollable (maintenance)	N/A	Not generated
Host has left maintenance mode but is within the Maintenance Automation Avoidance window.	Increased number of DRS activities moving VMs away from other hosts entering maintenance	Uncontrollable (maintenance)	N/A	Not generated
Host has left maintenance mode and is outside the Maintenance Automation Avoidance window.	Minimal number of DRS activities on the host	Controllable	N/A	Generated

Points to consider:

- You can set a different Maintenance Automation Avoidance value that aligns with your host maintenance practices. For example, if moving VMs back to a host typically takes an hour, specify a value of 60.
- You can set a global value in the default policy for hosts, or specific values in automation policies that you create for your clusters.
- For rolling maintenance of hosts in a cluster, where hosts undergo maintenance on a staggered basis, there could be a point in the process where some or all hosts are uncontrollable. This means that Workload Optimization Manager cannot recommend actions to alleviate pressure on overburdened hosts. As such, these hosts could lose performance while they are uncontrollable.
- This setting has no effect on clusters where the DRS automation level is *Manual* or *Partially Automated*. As soon as a host enters maintenance mode, Workload Optimization Manager automates the first action to move a VM to another host, and then stops recommending actions. After the host leaves maintenance mode, Workload Optimization Manager automates actions to manage the performance of the cluster as normal.

Utilization constraints

Utilization constraints affect the actions Workload Optimization Manager recommends as it manages your environment. Workload Optimization Manager recommends actions that avoid using these resources beyond the given settings. The values you set here specify what percentage of the existing capacity that Workload Optimization Manager will consider to be 100% of capacity.

Attribute	Default value
Net Throughput	50
Memory Utilization	100
IO Throughput	50
Swapping Utilization	20
CPU Utilization	100
Ready Queue Utilization	50

For example:

- Setting 50 for Net Throughput means that Workload Optimization Manager considers 50% utilization of that throughput to be 100% of capacity and 25% utilization to be 50% of capacity.
- Setting 100 for Memory Utilization means that Workload Optimization Manager capacity reflects the physical capacity for this resource.

Desired state

The desired state for your environment is an n-dimensional sphere that encompasses the fittest conditions your environment can achieve.

Attribute	Default value
Center	70
Diameter	10

The multiple dimensions of this sphere are defined by the resource metrics in your environment. Metric dimensions include VMem, storage, CPU, etc. While the metrics on the devices in your environment can be any value, the desired state, this n-dimensional sphere, is the subset of metric values that assures the best performance while achieving the most efficient utilization of resources that is possible.

The Desired State settings define the center of the sphere as well as its diameter. This is a way for you to customize what Workload Optimization Manager considers to be the desired state.

Setting the center of the sphere chooses the priority for Workload Optimization Manager analysis. If you set the balance in favor of efficiency, Workload Optimization Manager tends to place more VMs on fewer physical hosts, and to give them storage capacity from fewer data stores. As a result, high utilization can have more impact on QoS. With a balance in favor of performance, Workload Optimization Manager tends to spread virtual loads across more physical devices. This can result in the provisioning of excess resources.

The diameter setting determines the range of deviation from the center that can encompass the desired state. If you specify a large diameter, Workload Optimization Manager will have more variation in the way it distributes workload across hosting devices.

As you move each slider, a tooltip displays the numerical value of the setting. **Center** indicates the percentage of resource utilization you want, within the range you specify as **Diameter**. For example, if you want utilization of 75%, plus or minus 10%, then you would set **Center** = 75 and **Diameter** = 20. Workload Optimization Manager recommends actions that tend toward this desired state much as possible, given the dependencies within the current environment.

> **NOTE:**
> The setting for Target Utilization can have an effect on plans that you run. If you disable provisioning and suspension for hosts and datastores, then you should always set Center and Diameter to their default values.

Over provisioning constraints

These attributes allow capacity planners to overprovision CPU and Memory resources at the cluster level, while maintaining a limit on resource usage for the hosts for compliance. Overprovisioning at the cluster level increases density and efficiency by analyzing the cluster as a whole. There are two cluster level policy settings to control overprovisioning capacity percentage (CPU Overprovisioned Percentage and Memory Overprovisioned Percentage) and two host level settings to control the maximum-allowed utilization percentage (Host CPU Overprovisioned Max Util and Host Memory Overprovisioned Max Util).

Attribute	Default value
CPU Overprovisioned Percentage	30000
Memory Overprovisioned Percentage	1000
Host CPU Overprovisioned Max Util	200
Host Memory Overprovisioned Max Util	200

- CPU Overprovisioned Percentage

  CPU Overprovisioned Percentage is a cluster level setting to control overprovisioning capacity percentage for CPU Provisioned. Setting 30000 for CPU Overprovisioned Percentage means that the CPU Provisioned capacity will be 300 times the actual CPU capacity. The calculation for CPU Provisioned can be expressed as follows:

`CPU * (CPU Overprovisioned Percentage / 100)`

Assume the following data for determining the CPU Provisioned capacity:

Attribute	Default value
CPU	38.4 GHz
CPU Overprovisioned Percentage	30000

In this example, Workload Optimization Manager calculates the CPU Provisioned capacity as follows:

```
38.4 * (30000 / 100) = 11.52 THz
```

- Memory Overprovisioned Percentage

  Memory Overprovisioned Percentage is a cluster level setting to control overprovisioning capacity percentage for Memory Provisioned. Setting 1000 for Memory Overprovisioned Percentage means that the Memory Provisioned capacity will be 10 times the actual Memory capacity. The calculation for Memory Provisioned can be expressed as follows:

  `Memory * (Memory Overprovisioned Percentage / 100)`

  Assume the following data for determining the Memory Provisioned capacity:

Attribute	Default value
Memory	287.91 GB
Memory Overprovisioned Percentage	1000

  In this example, Workload Optimization Manager calculates the Memory Provisioned capacity as follows:

```
287.91 * (1000 / 100) = 2.81 TB
```

- Host CPU Overprovisioned Max Util

  Host CPU Overprovisioned Max Util is a host level setting to control max allowed utilization percentage for CPU Provisioned. Setting 200 for Host CPU Overprovisioned Max Util means that the maximum-allowed utilization percentage for CPU Provisioned capacity is 200. This means that once the host reaches the specified percentage of utilization, no other VM can be placed over the same host. The calculation for CPU Provisioned can be expressed as follows:

  `CPU * (CPU Overprovisioned Percentage / 100)`

  Then, the calculation for Max Allowed CPU Provisioned Used can be expressed as follows:

  `CPU Provisioned * (Host CPU Overprovisioned Max Util / 100)`

  Assume the following data for determining the CPU Provisioned capacity:

Attribute	Default value
CPU	38.4 GHz
CPU Overprovisioned Percentage	30000
Host CPU Overprovisioned Max Util	105%

  In this example, Workload Optimization Manager calculates the CPU Provisioned capacity as follows:

```
38.4 * (30000 / 100) = 11.52 THz
```

  Then, Workload Optimization Manager calculates the Max Allowed CPU Provisioned Used as follows:

```
11.52 * (105 / 100) = 12.1 THz
```

- Host Memory Overprovisioned Max Util

Host Memory Overprovisioned Max Util is a host level setting to control max allowed utilization percentage for Memory Provisioned. Setting 200 for Host Memory Overprovisioned Max Util means that the maximum-allowed utilization for Memory Provisioned capacity is 200. This means that once the host reaches the specified percentage of utilization, no other VM can be placed over the same host. The calculation for Memory Provisioned can be expressed as follows:

`Memory * (Memory Overprovisioned Percentage / 100)`

Then, the calculation for Max Allowed Memory Provisioned Used can be expressed as follows:

`Memory Provisioned * (Host Memory Overprovisioned Max Util / 100)`

Assume the following data for determining the Memory Provisioned capacity:

Attribute	Default value
Memory	287.91 GB
Memory Overprovisioned Percentage	1000
Host Memory Overprovisioned Max Util	108%

In this example, Workload Optimization Manager calculates the Memory Provisioned capacity as follows:

```
287.91 * (1000 / 100) = 2.81 TB
```

Then, Workload Optimization Manager calculates the Max Allowed Memory Provisioned Used as follows:

```
2.81 * (108 / 100) = 3.03 TB
```

Points to consider:

Overprovisioning works on both the cluster and host levels. The following example uses the Memory commodity; however, the same considerations are applicable for CPU.

- Cluster level

  If you have a cluster with four hosts, each with 50 GB Memory capacity. The cluster's actual Memory capacity will be 4 * 50 = 200 GB. If the cluster Memory Memory Overprovisioned Percentage is set to 1000, then the overall cluster's Memory Provisioned will be `200 * (1000 / 100) = 2000 GB`.

- Host level

  On host level, each host is allowed to be overprovisioned to twice the cluster's Memory Overprovisioned Percentage. In this example, each host in the cluster can be overprovisioned up to `50 * (1000 / 100) * 2 = 1000 GB`, as long as the overall sum for the Memory Provisioned capacity of all the hosts still adds up to the cluster's Memory Provisioned capacity (2000 GB in this example).

In some financial institutions, customers have special groups of hosts called platinum hosts that have a policy stating that the overprovisioned commodity cannot breach the cluster's overprovisioned ratio. In other words, the overprovisioned commodity (in this example, Memory Provisioned) should not breach the 100% utilization mark on the platinum hosts. In such cases, you can set the Host Memory Overprovisioned Max Util attribute to 100, meaning that the maximum-allowed Memory Provisioned Used for each platinum host can only go up to `50 * Cluster's Memory Overprovisioned Percentage * Host Memory Overprovisioned Max Util`.

In this example, Workload Optimization Manager calculates the Memory Provisioned Used as follows:

```
50 * (1000 / 100) * (100/100) = 500 GB
```

Placement policies

You can create placement policies that merge multiple clusters into a single logical group for workload placement.

For example, you can merge three host clusters in a single provider group. This enables Workload Optimization Manager to move workload from a host in one of the clusters to a host in any of the merged clusters to increase efficiency in your environment.

For more information, see Creating Placement Policies *(on page 1044)*.

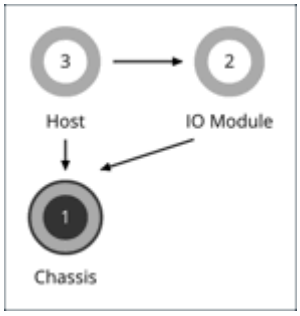
NOTE:
For vCenter, Cisco automatically imports any vSphere Host DRS rules when DRS is enabled, and displays them on the **Settings > Policies** page under **Imported Placement Policies**.

For more information, see Importing Workload Placement Policies *(on page 1044)*.

Chassis

A chassis houses the servers that are part of a computing fabric. It provides compute, memory, storage, and bandwidth resources.

Synopsis



Synopsis	
Provides:	Chassis resources (physical space, cooling, etc.)
Consumes:	N/A
Discovered through:	Fabric Manager targets

NOTE:
When Workload Optimization Manager discovers that blade servers housed in a particular chassis have been designated as vCenter hosts, the supply chain stitches the blade servers and chassis to the corresponding vCenter data center to establish their relationship. When you set the scope to that data center and view the Health chart, you will see the blade servers in the list of hosts. In addition, when the data center is included in a merge policy (a policy that merges data centers for the purpose of VM placement), the VMs in the blade servers apply the policy, allowing them to move between data centers as necessary.

Monitored resources

Workload Optimization Manager monitors the following resources:

- Power

  Power is the measurement of electricity consumed by a given entity, expressed in watts.

- Cooling

  Cooling is the percentage of the acceptable temperature range that is utilized by the entity. As the temperature nears the high or low running temperature limits, this percentage increases.

Actions

None

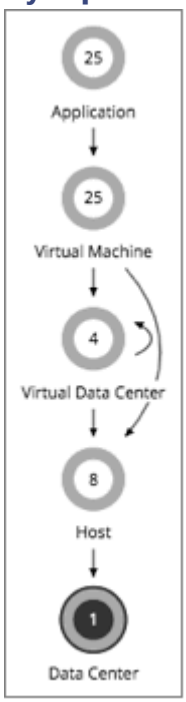
Workload Optimization Manager does not recommend actions for a chassis.

Data center

A data center is the sum of VMs, hosts, datastores, and network devices that are managed by a given hypervisor target. A data center provides compute, memory, storage, and bandwidth resources.

NOTE:
For public cloud environments, a data center is the cloud region. The hosts that get resources from the data center are availability zones within that region. For details, see Region *(on page 738)* and Zone *(on page 736)*.

Synopsis



Synopsis	
Provides:	Compute, memory, storage, and bandwidth resources
Consumes:	N/A
Discovered through:	Hypervisor targets

NOTE:
When Workload Optimization Manager discovers that blade servers housed in a particular chassis have been designated as vCenter hosts, the supply chain stitches the blade servers and chassis to the corresponding vCenter data center to establish their relationship. When you set the scope to that data center and view the Health chart, you will see the blade servers in the list of hosts. In addition, when the data center is included in a merge policy (a policy that merges data centers for the purpose of VM placement), the VMs in the blade servers apply the policy, allowing them to move between data centers as necessary.

Monitored resources

Workload Optimization Manager does not monitor resources directly from the data center, but it does monitor the following resources, aggregated for the hosts in a data center:

- Memory (Mem)

  Memory is the measurement of memory that is reserved or in use.
- CPU

  CPU is the measurement of CPU that is reserved or in use.
- IO

  IO is the utilization of a host's IO adapters.
- Net

  Net is the utilization of data through the host's network adapters.
- Swap

  Swap is the measurement of a host's swap space that is in use.
- Balloon

  Balloon is the measurement of memory that is shared by VMs running on a host.
- CPU ready

  CPU ready is the measurement of a host's ready queue capacity that is in use.

For vCenter targets, the following resources are also monitored:

- Energy

  Energy is the measurement of electricity used by a given entity over a period of time, expressed in watt-hours (Wh).
- Carbon footprint

  Carbon footprint is the measurement of carbon dioxide equivalent (CO2e) emissions for a given entity. Workload Optimization Manager measures carbon footprint in grams.

Actions

None

Workload Optimization Manager does not recommend actions for a datacenter. Instead, it recommends actions for the entities running in the datacenter.

Data center policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Operational constraints

Workload Optimization Manager calculates *(on page 978)* carbon footprint using industry standards that take into account energy consumption, datacenter efficiency, and carbon intensity data. You can create Data Center policies to adjust the calculations according to the requirements of your data centers. For example, a data center in a particular location might have different requirements than data centers in other locations. After you adjust the calculations via policies, Workload Optimization Manager can accurately report your organization's carbon footprint.

Attribute	Default value
Power Usage Effectiveness (PUE)	1.5

Attribute	Default value
Carbon Intensity (CI)	.25 (g/Wh)

- Carbon Intensity (CI) is a measurement of how 'clean' electricity is. It refers to how many grams of carbon dioxide (CO2) are released to produce 1 watt-hour (Wh) of electricity. Electricity that is generated using fossil fuels is more carbon intensive, as the process by which it is generated creates CO2 emissions. Renewable energy sources, such as wind, hydro, or solar power produce next to no CO2 emissions, so their carbon intensity value is much lower and often zero.
- Power Usage Effectiveness (PUE) is a ratio that describes how efficiently a computer data center uses energy; specifically, how much energy is used by the computing equipment. PUE is the ratio of the total amount of energy used by a computer data center facility to the energy delivered to computing equipment. The closer PUE is to 1, the more efficient the computer data center.

Placement policies

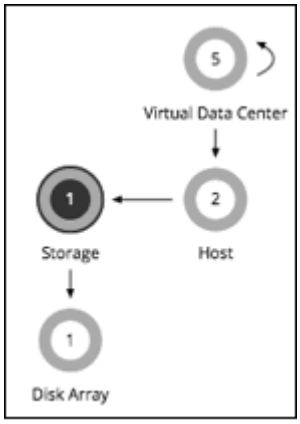
For vCenter environments, you can create placement policies that merge data centers to support cross-vCenter moves. In this case, where a data center corresponds to a vCenter target, the merged clusters can be in different data centers. In this case, you must create two merge policies; one to merge the affected data centers, and another to merge the specific clusters.

For more information, see Creating Placement Policies *(on page 1044)*.

Storage

Workload Optimization Manager represents storage as Datastores. A Datastore is a logical grouping of one or more physical storage devices that serve workload storage requirements.

Synopsis



A Datastore gains its budget by selling resources to the VMs it serves. If utilization of a Datastore is high enough, Workload Optimization Manager can recommend that you provision a new one.

Synopsis	
Provides:	Host resources for VMs to use, including: ▪ Storage Amount ▪ IOPS ▪ Latency
Consumes:	Disk arrays (or aggregates)
Discovered through:	Hypervisor and Storage Controller targets

Monitored resources

Workload Optimization Manager monitors the following resources:

- Storage amount

  Storage amount is the measurement of storage capacity that is in use.
- Storage provisioned

  Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
- Storage access (IOPS)

  Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

  NOTE:
  When it generates actions, Workload Optimization Manager does not consider IOPS throttling that it discovers on storage entities. Analysis uses the IOPS it discovers on Logical Pool or Disk Array entities.

- Latency

  Latency is the measurement of storage latency.

Storage actions

Workload Optimization Manager supports the following actions:

- Move

  For high utilization of physical storage, move datastore to a different disk array (aggregate).
- Provision

  For high utilization of storage resources, provision a new datastore.
- Resize

  Increase or decrease the datastore capacity.
- Start

  For high utilization of storage resources, start a suspended datastore.
- Suspend

  For low utilization of storage resources, move served VMs to other datastores and suspend this one.
- Delete (datastore or volume)

  Delete a datastore or volume that has been suspended for a period of time.
- Delete (unattached files)

  Delete a file on a datastore that has not been accessed or modified for a period of time.

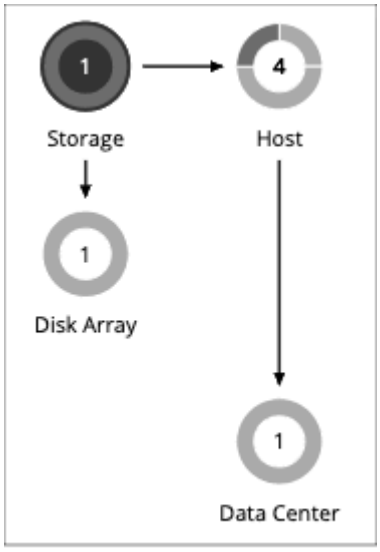
Storage resize actions use Workload Optimization Manager tuned scaling settings. This gives you increased control over the action acceptance mode Workload Optimization Manager will use for the affected actions. For an overview of tuned scaling, see Tuned Scaling for On-prem VMs *(on page 744)*.

You can create placement policies to enforce constraints for storage move actions. For example, you can have a policy that allows storage to only move to certain disk arrays, or a policy that prevents storage from moving to certain disk arrays.

For more information, see Creating Placement Policies *(on page 1044)*.

vSAN storage

Overview



For environments that use hyperconverged infrastructure to provide storage on a vSAN, Workload Optimization Manager can discover the storage provided by a host cluster as a single Storage entity. This Storage entity represents the full storage capacity that is provided by that host cluster.

Workload Optimization Manager supports VMware vSAN, but does not support stretched VSAN clusters. Adding stretched clusters can cause the generation of incorrect storage recommendations and actions.

vSAN storage capacity

When you consider vSAN capacity, you need to compare *Raw Capacity* with *Usable Capacity*.

- Raw Capacity

  Workload Optimization Manager discovers Raw Capacity configured in vCenter and uses it to calculate Usable Capacity. Raw Capacity displays in the Entity Information chart.

- Usable Capacity

  Workload Optimization Manager calculates Usable Capacity and then uses the calculated value to drive scaling actions. Workload Optimization Manager can recommend scaling the Storage Amount, Storage Provisioned, or Storage Access capacity. Usable Capacity displays in the Capacity and Usage chart.

Usable capacity calculation

To calculate Usable Capacity, Workload Optimization Manager considers a variety of attributes, including:

- Raw Capacity and Largest Host Capacity

  Workload Optimization Manager compares the Raw Capacity for all the hosts in the cluster and then uses the largest value as Largest Host Capacity.

- RAID Factor

  Workload Optimization Manager calculates RAID Factor based on the *Failures to Tolerate* (FTT) value and *Redundancy Method* that it discovers. FTT specifies how many failures a given cluster can tolerate, while Redundancy Method specifies the RAID level for the cluster.

FTT	Redundancy method	RAID factor
0	RAID1	1
1	RAID1	1/2
2	RAID1	1/3
1	RAID5/6	3/4
2	RAID5/6	2/3

NOTE:
If discovery fails for some reason, Workload Optimization Manager uses a RAID Factor of 1.

- Host Capacity Reservation, Slack Space Percentage, and Compression Ratio

  You can control the values for these attributes in storage policies. For details about these attributes and their effect on usable capacity calculations, see Hyper-converged Infrastructure Settings *(on page 806)*.

The calculation for Usable Capacity can be expressed as:

```
Usable Capacity = (Raw Capacity - Largest Host Capacity * Host Capacity Reservation) * Slack Space
 Percentage * RAID Factor * Compression Ratio
```

If the result of the calculation is zero or a negative value, Workload Optimization Manager sets the Usable Capacity to 1 MB.

Capacity and usage chart for vSAN storage

The **Capacity and Usage** chart for vSAN storage shows two Storage Amounts – *Consumed* (bought) and *Provided* (sold). This is because vSAN storage can buy and sell commodities to hosts.

For the *Provided* Storage Amount, the *Capacity* value corresponds to *Usable Capacity*, while the *Used* value indicates utilization.

Entity information chart for vSAN storage

The **Entity Information** chart includes the following information:

- HCI Technology Type

  The technology that supports this storage cluster. For this release, Workload Optimization Manager supports VMware vSAN technology.
- Capacity

  Workload Optimization Manager displays rounded values for the following, which might be slightly different from the values it discovers from vCenter:
  - Raw Capacity

    The sum of the Raw Capacity that each storage capacity device provides.
  - Raw Free Space

    How much of the Raw Capacity is not currently in use.
  - Raw Uncommitted Space

    In terms of Raw Capacity, how much space is available according to your thin/thick provisioning.
- Redundancy Method and Failures to Tolerate

Redundancy Method specifies the RAID level employed for the cluster. RAID level impacts how much Usable Capacity you can see for a given Raw Capacity. You can use a RAID calculator to determine how the RAID level impacts your Usable Capacity.

Failures to Tolerate specifies how many capacity device failures a given cluster can tolerate. In practical terms, this means how many hosts can come down at the same time, without affecting storage. This value should match the RAID level.

Actions to add vSAN capacity

To scale up storage amount, you add additional hosts that are configured to include their storage in the vSAN array.

When you scope the session to the vSAN storage, you can see actions to scale:

- Storage Amount
- Storage Provisioned
- Storage Access

The action to scale up the storage indicates the amount of storage you need to add. It appears as a recommended action. In fact, to add storage you must add a new host.

When you scope the session to hosts that provide the capacity devices to the storage, you can see the following actions that are related to scaling up the storage capacity:

- Scale up StorageAmount for Storage [MyVsanStorageCluster]
- Provision Host [VSAN_HostName]

The action to provision a host includes details about the storage cluster. Because you need to manually add hosts to your on-prem environment, this appears as a recommended action.

Planning with vSAN storage

For *Hardware Replace* and *Custom* plans, you can use HCI Host templates to add vSAN capacity. These represent the hosts that add storage capacity to a vSAN cluster. For more information, see HCI Host Template Settings *(on page 1104)*.

Under certain circumstances, *Add Virtual Machines* plans can fail to place workloads, or it can fail to generate actions to increase storage capacity by provisioning new hosts.

- If you scope the plan to a user-created group that only provides vSAN storage, or to a discovered storage cluster group, then the plan can fail to place VMs with multiple volumes. This can occur for VMs that use conventional storage (not vSAN) along with vSAN storage.
- If you scope the plan to a vSAN host group and add VMs, the plan can fail to increase storage capacity by provisioning new hosts. For example, assume you scope the plan to a vSAN host group and add 20 VMs to the environment. In that case, you need hosts to provide compute capacity for the VMs, and you also need hosts to provide storage capacity. The plan can represent the compute provisioning correctly, but it can incorrectly fail to add more storage capacity to the vSAN.
- If the vSAN RAID type is `Raid6/FTT=2`, if you scope the plan to any vSAN groups then the plan will fail to place any of the VMs.

Storage policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

The following are the storage actions and automation support for environments that do not include Disk Array Storage Controllers as targets. For details about these actions, see Storage Actions *(on page 799)*.

Action	Default mode	vCenter	Hyper-V
Delete (Datastore)	Manual	Recommend	Recommend
Delete (Volume)	Manual	Recommend	Recommend
Delete Unattached Files	Manual	Automatable	Recommend
Move	Recommend	Recommend	Recommend
Provision	Recommend	Recommend	Recommend
Resize (Up, Down, Above Max, or Below Min) using tuned scaling	Recommend	Recommend	Recommend
Start	Recommend	Recommend	Recommend
Suspend	Disabled	Recommend	Recommend

For datastores on disk arrays:

Action	Default mode	Dell Compellent	HP 3Par	NetApp ONTAP	Dell EMC VNX	VMAX	Nutanix	Pure Storage
Delete (Volume)	Recommend	Recommend	N/A	Recommend	Recommend	Recommend	Recommend	Recommend
Suspend	Manual	Recommend	Recommend	Recommend	Recommend	Recommend	Recommend	Recommend
Delete (Datastore)	Disabled	Recommend	N/A	Recommend	Recommend	Recommend	Recommend	Recommend
Move	Recommend	Recommend	N/A	Recommend	Recommend	Recommend	Recommend	Recommend
Provision	Recommend	Recommend	Recommend	Recommend	Recommend	Recommend	Recommend	Recommend
Start	Recommend	Recommend	Recommend	Recommend	Recommend	Recommend	Recommend	Recommend
Resize (Up, Down, Above Max, or Below Min) using tuned scaling	Recommend	Recommend	Recommend	Recommend	Recommend	Recommend	Recommend	Recommend

You can use Action Scripts for action orchestration.

For ServiceNow:

- Storage suspend and vSAN storage resize actions will not generate a CR.
- Currently Workload Optimization Manager can only execute a CR for storage provision actions on Pure and Dell Compellent storage.

Utilization constraints

Utilization constraints affect the actions Workload Optimization Manager recommends as it manages your environment. Workload Optimization Manager recommends actions that avoid using these resources beyond the given settings. The values you set here specify what percentage of the existing capacity that Workload Optimization Manager will consider to be 100% of capacity.

Attribute	Default value
Storage Provisioned Utilization	100
IOPS Utilization	100
Storage Amount Utilization	90
Latency Utilization	100

For example, setting 90 for Storage Amount Utilization means that Workload Optimization Manager considers 90% utilization of the physical storage to be 100% of capacity.

Storage settings

Attribute	Default setting/value
Storage Latency Capacity	277 ms
IOPS Capacity	50000
Storage Overprovisioned Percentage	2000
Minimum Unattached Files Size	1000 KB
Directories to Ignore	\.dvsData.*|\.snapshot.*|\.vSphere-HA.*|\.naa.*|\.etc.*|lost\+found.*
Files to Ignore	config\.db|stats\.db.*
Generate Delete Action after	15 Day(s)
Delete file after	30 Day(s)

- Storage Overprovisioned Percentage

  Storage Overprovisioned Percentage sets how much overprovisioning Workload Optimization Manager assumes when recommending actions for VM datastores. For example, if a datastore has a 30 GB capacity, and Storage Overprovisioned Percentage is set to 2000, Workload Optimization Manager will treat the datastore as though it has a capacity of 60 GB, or 200% of the actual datastore capacity.

- IOPS Capacity

  IOPS Capacity is the IOPS setting for individual datastores. To set a specific capacity for one group of datastores, select that group as the property scope and override the global setting for that scope.

  Note that IOPS capacity for a disk array takes precedence — Datastores that are members of a disk array always have the IOPS capacity that is set to the disk array.

  Workload Optimization Manager considers these settings when calculating utilization percentage. For example, assume IOPS Capacity of 500 for datastores. If utilization on a datastore is 250 IOPS, then the datastore is at 50% of capacity for that metric.

- Storage Latency Capacity

Storage Latency Capacity sets the maximum storage latency to tolerate on a datastore, in ms. The default setting is 100 ms.

Workload Optimization Manager measures the latency experienced by all VMs and hosts that access the datastore. Assume a default setting of 100 ms. If a datastore exhibits latency of 50 ms, then the Workload Optimization Manager will show latency utilization of 50%.

For VMAX environments, Workload Optimization Manager discovers SLO for storage latency that you set in VMAX and uses it in analysis. However, if you set a higher storage latency value in a Workload Optimization Manager policy, analysis will use that value instead.

- Minimum Unattached Files Size

  You can make settings to control how Workload Optimization Manager tracks and reports on unattached storage in your environment. Unattached storage is any disk space devoted to files that are not required for operations of the devices or applications in your environment. Unattached storage may indicate opportunities for you to free up disk space, and provide more storage capacity to running VMs and applications.

  If there are groups of datastores you do not want to track for unattached storage, set the given scope and disable datastore browsing there. If you prefer not to use Workload Optimization Manager resources to track unattached storage, leave the global setting checked.

  The settings for **Directories to Ignore** and **Files to Ignore** specify directories and files that Workload Optimization Manager will not consider when looking for unattached data storage space. Separate items in these lists with the OR bar ("|").

- Generate Delete Action after

  Generate Delete Action after specifies the duration of days a file is inactive before the **Delete Unattached Files** action is generated.

- Delete file after

  Delete file after specifies the duration of days a file can be inactive before the delete action can be automatically executed. This setting must be larger than the duration of days the file is inactive.

  Enable automatic deletion of unattached on-prem files by changing the Automation and Orchestration settings for **Delete Unattached Files** actions.

  NOTE:
  When the **Delete Unattached Files** action type is set to automated, Workload Optimization Manager deletes the unattached on-prem files after the specified number of days of inactivity. The action is not executed if the action type is left at the default (manual).

Scaling constraints

Attribute	Default value
Increment Constant for Storage Amount	100 GB
Rate of Resize	High (3)

- Increment Constant for Storage Amount

  This setting controls how many GB to add or subtract when resizing the allocation for a datastore.

- Rate of Resize

  Workload Optimization Manager uses the Rate of Resize setting to determine how to make storage resize changes in a single action.

  - Low

    Change the value by one increment only.

  - Medium

    Change the value by an increment that is 1/4 of the difference between the current value and the optimal value.

  - High

Change the value to be the optimal value.

This default value ensures that resizing to the desired state can be achieved in a single action. This is more efficient than smaller, incremental resizes.

Hyperconverged infrastructure settings

Workload Optimization Manager considers these settings when calculating capacity and utilization for hyperconverged environments.

Attribute	Default setting/value
Compression Ratio	1
Host IOPS Capacity	50000
Slack Space Percentage	25
Host Capacity Reservation	1
Usable Space Includes Compression	Off

NOTE:
Workload Optimization Manager uses Host Capacity Reservation, Slack Space Percentage, and Compression Ratio to calculate vSAN usable capacity and drive scaling actions. For more information about usable capacity and how it is calculated, see vSAN Storage *(on page 800)*.

- Host Capacity Reservation

  When a host must be taken out of service for maintenance, vSphere will evacuate the data from that host and move it to other hosts in the cluster to maintain the integrity of the replication demanded by the storage policy. For this to happen, there must be enough free raw capacity available to accept the data being evacuated.

  Workload Optimization Manager uses this setting to determine how many hosts worth of capacity it should subtract from the raw capacity amount before calculating usable capacity. This is not the same as redundancy. It does not specify how the array distributes data to maintain integrity.

- Host IOPS Capacity

  In addition to calculating usable capacity, Workload Optimization Manager needs an estimate of datastore IOPS capacity (storage access). Workload Optimization Manager uses the value that you set to provide an estimate of effective IOPS capacity for each host in the cluster. Total IOPS capacity is the number of hosts in the cluster multiplied by Host IOPS Capacity.

- Slack Space Percentage

  It is recommended that a vSAN datastore never be filled to prevent vSphere from moving objects/files around the cluster to balance the datastore across all the hosts.

  Workload Optimization Manager reduces usable capacity by the percentage that you set.

- Compression Ratio

  vSAN supports both deduplication and compression, which may increase the amount of usable capacity on the datastore. Workload Optimization Manager does not try to predict the deduplication or compression ratio, but you can choose to include a compression ratio into the usable capacity calculation. This captures the ratio achieved both by compression and deduplication.

  The compression ratio that you set acts as a multiplier on the raw capacity to calculate usable capacity. For example, a compression ratio of 2 would double the amount of usable capacity. The default value of 1 means no compression.

- Usable Space Includes Compression

  Turn this on if you want Workload Optimization Manager to consider the compression ratio when calculating storage utilization and capacity. Whether this is on or off, Workload Optimization Manager always considers compression when calculating utilization of StorageProvisioned.

Placement policies

Workload Optimization Manager supports placement policies for storage and storage clusters.

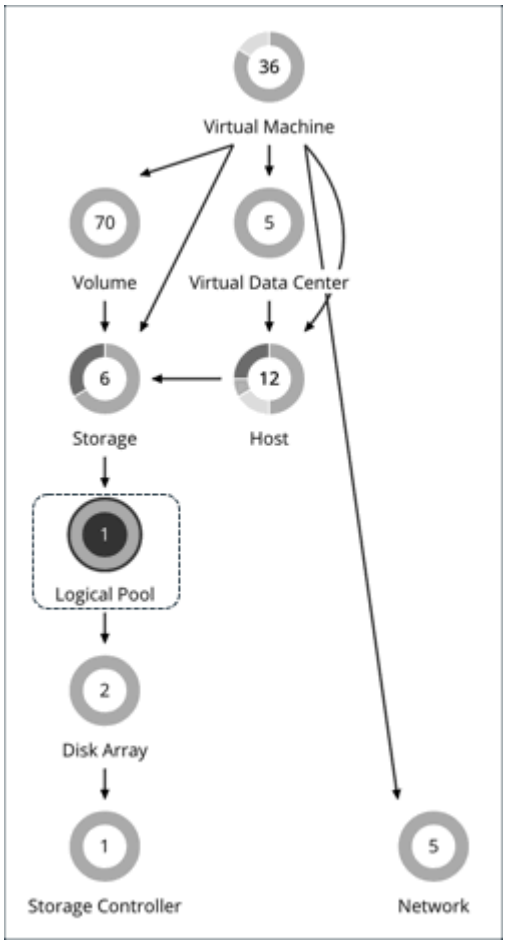
- You can create placement policies to enforce constraints for storage move actions. For example, you can have a policy that allows storage to only move to certain disk arrays, or a policy that prevents storage from moving to certain disk arrays.
- You can create placement policies that merge multiple clusters into a single logical group for workload placement.

For more information, see Creating Placement Policies *(on page 1044)*.

Logical pool

A logical pool represents storage resources that are managed together and presented as a single storage system. Workload Optimization Manager analysis identifies performance and efficiency opportunities for a logical pool. For example, it can recommend moving resources into or out of a logical pool, or aggregating resource capacity within the pool.

Synopsis



Synopsis	
Provides:	Storage resources
Consumes:	Disk array resources
Discovered through:	Storage targets

Monitored resources

Workload Optimization Manager monitors the following resources:

- Storage amount

  Storage amount is the measurement of storage capacity that is in use.
- Storage provisioned

  Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
- Storage access (IOPS)

  Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
- Latency

  Latency is the measurement of storage latency.

Logical pool actions

Workload Optimization Manager supports the following actions:

- Move

  For high utilization of physical storage, move the logical pool to a different disk array (aggregate).
- Provision

  For high utilization of storage resources, provision a new logical pool.
- Resize

  Increase or decrease the logical pool capacity.
- Start

  For high utilization of storage resources, start a suspended logical pool.
- Suspend

  For low utilization of storage resources, move served VMs to other logical pools and suspend this one.

Logical pool policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

Action	Default mode
Move	Disabled
Provi-sion	Disabled
Resize	Recommend
Start	Disabled
Suspend	Disabled

Storage settings

Attribute	Default value
IOPS Capacity	50000

Attribute	Default value
Storage Latency Capacity	100 ms
Storage Overprovisioned Percentage	200

- Storage Latency Capacity

  Storage Latency Capacity sets the maximum storage latency to tolerate on a logical pool, in ms. The default setting is 100 ms.

- Storage Overprovisioned Percentage

  Storage Overprovisioned Percentage sets how much overprovisioning Workload Optimization Manager assumes when recommending actions for logical pools.

- IOPS Capacity

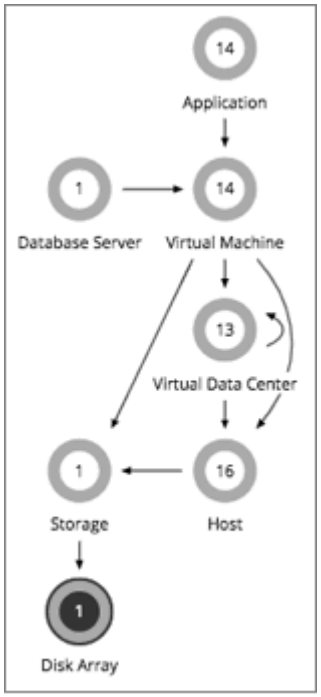
  IOPS Capacity is the IOPS setting for individual logical pools.

NOTE:
Workload Optimization Manager discovers storage latency and IOPS capacities that you set in your environment (for example VMAX) and uses them in its analysis. These capacities will be overridden by values that you set in Workload Optimization Manager policies.

Disk array

A disk array (an aggregate) is a data storage system made up of multiple disk drives. For example, a RAID is an aggregate that implements redundancy and other data management features. A disk array provides storage volumes to serve the storage requirements of physical machines. It uses the resources of one storage controller, which manages the disk array operation.

Synopsis



Synopsis	
Provides:	Storage resources for datastores to use, including: ▪ Storage Amount ▪ Storage Provisioned ▪ IOPS (storage access operations per second) ▪ Latency (capacity for disk latency in ms)
Consumes from:	Storage Controllers
Discovered through:	Storage Controller targets

Monitored resources

Workload Optimization Manager monitors the following resources:

NOTE:
Not all targets of the same type provide all possible commodities. For example, some storage controllers do not expose CPU activity. When a metric is not collected, the corresponding chart in the user interface will not display data.

- Storage amount

  Storage amount is the measurement of storage capacity that is in use.
- Storage provisioned

  Storage provisioned is the utilization of the entity's capacity, including overprovisioning.
- Storage access (IOPS)

  Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.
- Latency

  Latency is the measurement of storage latency.

Disk array actions

Workload Optimization Manager supports the following actions:

- Provision

  For high utilization of the disk array's storage, provision a new disk array. This action can only be executed outside Workload Optimization Manager.
- Start

  For high utilization of disk array, start a suspended disk array. This action can only be executed outside Workload Optimization Manager.
- Suspend

  For low utilization of the disk array's storage, move VMs to other datastores and suspend volumes on the disk array. This action can only be executed outside Workload Optimization Manager.
- Move

  (Only for NetApp Cluster-Mode) For high utilization of Storage Controller resources, Workload Optimization Manager can move an aggregate to another storage controller. The storage controllers must be running.

  For high IOPS or latency, a move is always off of the current disk array. All the volumes on a given disk array show the same IOPS and Latency, so moving to a volume on the same array would not fix these issues.
- Move VM

  For high utilization of Storage on a volume, Workload Optimization Manager can move a VM to another volume. The new volume can be on the current disk array, on some other disk array, or on any other datastore.

For high IOPS or latency, a move is always off of the current disk array. All the volumes on a given disk array show the same IOPS and Latency, so moving to a volume on the same array would not fix these issues.

- Move datastore

  To balance utilization of disk array resources, Workload Optimization Manager can move a datastore to another array.

Disk array policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

The following table describes the default action acceptance modes for disk array actions and automation support for environments that have disk array storage controllers as targets.

Action	Default mode	Dell Compellent	HP 3Par	NetApp ONTAP	VMAX	Dell EMC VNX	Nutanix	Pure Storage	XTremIO
Move	Disabled	N/A	N/A	Recommend	N/A	N/A	N/A	N/A	N/A
Provision	Recommend	Recommend	Recommend	Recommend	N/A	N/A	Recommend	Recommend	N/A
Resize (up)	Recommend	Recommend	Recommend	Recommend	Recommend	Recommend	Recommend	Recommend	N/A
Start	Recommend	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Suspend	Disabled	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Action automation for NetApp storage systems

For NetApp storage systems, the actions Workload Optimization Manager can automatically perform depend on the NetApp version you are running, and whether the system is running in cluster mode:

Automated action	Cluster mode
Move VM between datastores, on the same disk array	Yes
Move VM between datastores on different disk arrays	Yes
Move Datastore between disk arrays on the same storage controller	Yes
Move Datastore between disk arrays on different storage controllers	Yes
Resize Storage	Yes
Resize Disk Array	No – Resize up, only

In addition, for a system running in Cluster-Mode, Workload Optimization Manager can recommend moving an aggregate to another storage controller.

Utilization constraints

Utilization constraints affect the actions Workload Optimization Manager recommends as it manages your environment. Workload Optimization Manager recommends actions that avoid using these resources beyond the given settings. The values

you set here specify what percentage of the existing capacity that Workload Optimization Manager will consider to be 100% of capacity.

Attribute	Default value
Storage Amount Utilization	90

Storage settings

Set capacity for specific storage resources.

Attribute	Default value
SSD Disk IOPS Capacity	50000
15k Disk IOPS Capacity	1600
VSeries LUN IOPS Capacity	5000
Storage Latency Capacity	100 ms
7.2k Disk IOPS Capacity	800
Storage Overprovisioned Percentage	200
IOPS Capacity – a generic setting for disk array IOPS capacity (see **Disk Array IOPS Capacity**).	5000
10k Disk IOPS Capacity	1200
Disk Array IOPS Capacity	10000

NOTE:
Workload Optimization Manager discovers storage latency and IOPS capacities that you set in your environment (for example VMAX) and uses them in its analysis. These capacities will be overridden by values that you set in Workload Optimization Manager policies.

- IOPS capacity

  The capacity of IOPS (IO operations per second) that your storage devices can support. Workload Optimization Manager considers these settings when calculating utilization percentage. For example, assume IOPS Capacity of 5000 for a disk array. If utilization on the array is 2500 IOPS, then the disk array is at 50% of capacity for that metric.

  Note that the IOPS setting for an array will determine IOPS calculations for all the storage on that array. If you made different IOPS settings for individual datastores hosted by the array, Workload Optimization Manager ignores the datastore settings and uses the disk array settings.

  - Various disk IOPS capacity settings, such as SSD Disk IOPS and 7.2k Disk IOPS

    IOPS capacity settings for the different types of physical drives that are discovered on a disk array. If the storage controller exposes the types of disks in the array, Workload Optimization Manager uses multiples of these values to calculate the IOPS capacity of the disk array.

  - Disk array IOPS capacity

    Some disk arrays do not expose data for their individual disks — This is typical for flash arrays, or arrays that aggregate storage utilization across multiple tiers. Workload Optimization Manager uses this setting for the IOPS capacity of such disk arrays. Set it to the global scope to specify IOPS capacity for all disk arrays. To override this setting, set a disk array or group of disk arrays as the property scope, and then set the value you want for **IOPS Capacity**.

NOTE:
The user interface shows a disk array entity for any array that is discovered through a valid disk array or storage controller target. It also shows *placeholder* disk arrays for disk arrays that are not discovered through a configured target. For example, you might have disk arrays that Workload Optimization Manager does not natively support. Or you might have storage that is not hosted by any disk array. Such *placeholder* disk array entities appear with the string "DiskArray-" prefixed to their names. The user interface allows you to set IOPS Capacity to these placeholders, but those settings have no effect. To set IOPS Capacity for that storage, you must set it to the individual datastores.

- Storage overprovisioned percentage

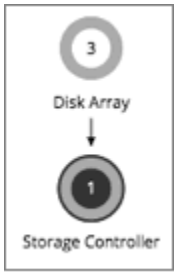
  This setting indicates how much overprovisioning Workload Optimization Manager assumes when recommending actions for disk arrays. For example, if a disk array has a 30 TB capacity, and Storage Overprovisioned Percentage is set to 200, Workload Optimization Manager will treat the datastore as though it has a capacity of 60 TB, or 200% of the actual disk array capacity.

  If the disk array is returning good Data Reduction Rates (DRR) in excess of the Storage Overprovisioned Percentage, actions to provision disk array may be generated. Consider increasing the percentage upwards closer to the real DRR. For example, if DRR is 12 to 1, the percentage could be closer to 1200.

Storage controller

A Storage Controller is a device that manages one or more disk arrays. The storage controller provides CPU cycles to perform storage management tasks for each disk array it manages.

Synopsis



A storage controller gains its budget by selling resources to the disk arrays it manages. If utilization of the storage controller's CPU resources is high enough, Workload Optimization Manager can recommend that you provision a new one and move disk arrays (aggregates) to it.

Synopsis	
Provides:	CPU resources to manage disk arrays.
Consumes:	NA
Discovered through:	Workload Optimization Manager directly accesses storage controller targets.

Monitored resources

Workload Optimization Manager monitors the following resources:

- CPU

  CPU is the measurement of CPU that is reserved or in use.

- Storage amount

  Storage amount is the measurement of storage capacity that is in use.

NOTE:
In NetApp environments, the storage controller shows 100% utilization when there are no more disks in a SPARE state that the storage controller can utilize in an aggregate. This does not indicate that the storage controller has no capacity.

Actions

Workload Optimization Manager supports the following actions:

Provision

For high utilization of the storage controller's CPU, provision a new storage controller, and then move disk arrays to it.

Storage controller policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

Actions for individual Disk Array Storage Controllers:

Action	Default mode	Dell Compellent	HP 3Par	NetApp ONTAP	Dell EMC VNX	VMAX	Nutanix	Pure Storage	XTremIO
Provision	Disabled	Recommend	Recommend	Recommend	Recommend	Recommend	Recommend	Recommend	Recommend

Utilization constraints

Utilization constraints affect the actions Workload Optimization Manager recommends as it manages your environment. Workload Optimization Manager recommends actions that avoid using these resources beyond the given settings. The values you set here specify what percentage of the existing capacity that Workload Optimization Manager will consider to be 100% of capacity.

Attribute	Default value
CPU Utilization	100 Maximum allowed utilization of Storage Controller CPU (from 20 to 100).
Storage Amount Utilization	90 Maximum allowed utilization of storage that is managed by the Storage Controller.

Storage settings

Set capacity for specific storage resources.

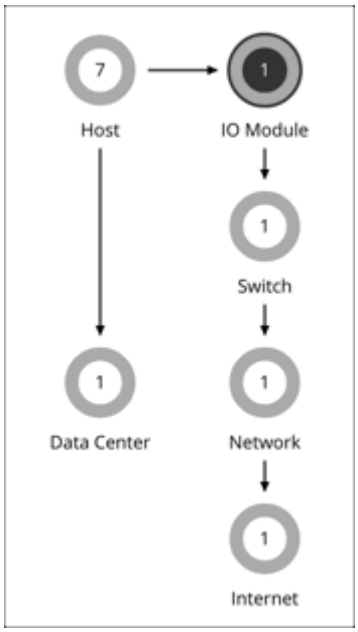
Attribute	Default value
IOPS Capacity	5000
Storage Latency Capacity	100 ms

IO module

An IO Module connects the compute resources on a chassis to the fabric domain via the Fabric Interconnect. It provides the servers on the chassis with Net resources. Typical installations provide two IO Modules per chassis.

Workload Optimization Manager supports IO Modules when you have installed the Fabric Control Module license.

Synopsis



Synopsis	
Provides:	Net resources
Consumes from:	Chassis and Fabric Interconnect
Discovered through:	Fabric Manager targets

Monitored resources

Workload Optimization Manager monitors the following resources:

- Net throughput

  Net throughput is the rate of message delivery over a port.

Actions

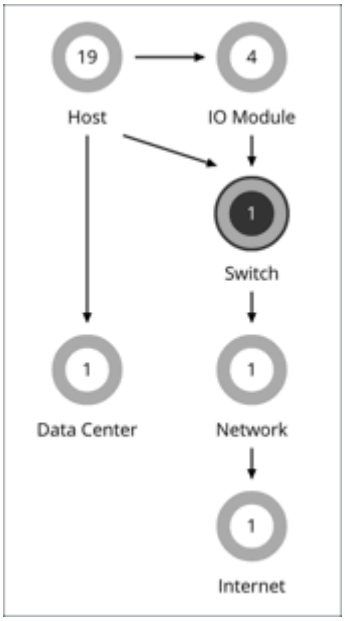
- Suspend

  For low utilization of net resources, move VMs to another host that provides compatible network connectivity and suspend this IO Module.

Switch

A switch connects servers in a computing fabric to the fabric's network and storage resources. It provides network bandwidth to the servers in the platform.

Synopsis



Synopsis	
Provides:	Net resources
Consumes:	N/A
Discovered through:	Fabric Manager targets (such as UCS)

Monitored resources

Workload Optimization Manager monitors the following resources:

- Net throughput

  Net throughput is the rate of message delivery over a port.
- PortChannel

  `PortChannel` is the amalgamation of ports with a shared net throughput and utilization.

Actions

Workload Optimization Manager supports the following actions:

Resize

Resize PortChannel for a switch to increase bandwidth.

Switch policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

Automation workflow

For environments that have Fabric Managers as targets:

Action	Default mode	Cisco UCS
Resize	Recommend	Recommend
Start	Recommend	Recommend
Provision	Recommend	Recommend
Suspend	Disabled	Recommend
Move	Disabled	Recommend

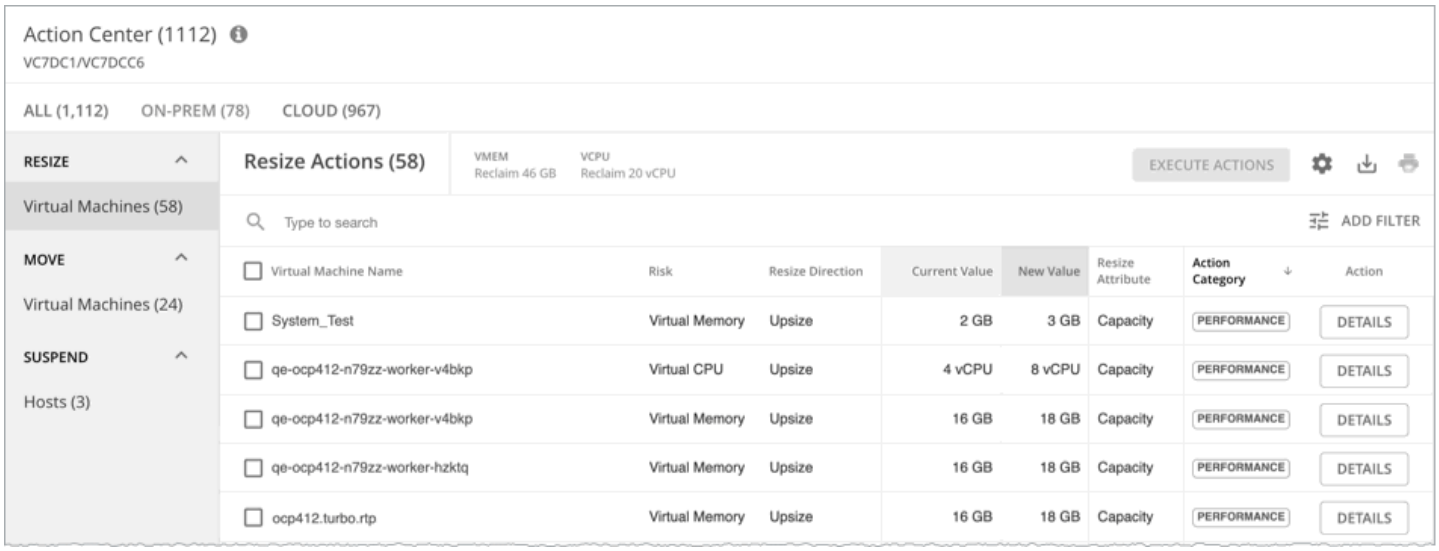
Utilization constraints

Utilization constraints affect the actions Workload Optimization Manager recommends as it manages your environment. Workload Optimization Manager recommends actions that avoid using these resources beyond the given settings. The values you set here specify what percentage of the existing capacity that Workload Optimization Manager will consider to be 100% of capacity.

Attribute	Default value
Switch Net Throughput	70

Workload Optimization Manager actions

After you deploy your targets, Workload Optimization Manager starts its analysis of your environment. This holistic analysis identifies problems and the actions you can take to *resolve* and *avoid* these problems. Workload Optimization Manager then generates a set of actions for that particular analysis and displays it in Action Center for you to investigate.



To get the best results from Workload Optimization Manager, execute these actions promptly and consider automating as many of them as possible. You can execute these actions from the user interface or outside Workload Optimization Manager. To automate these actions, create an automation policy *(on page 1053)* or change action acceptance to *Automatic* in the default policies *(on page 1050)*.

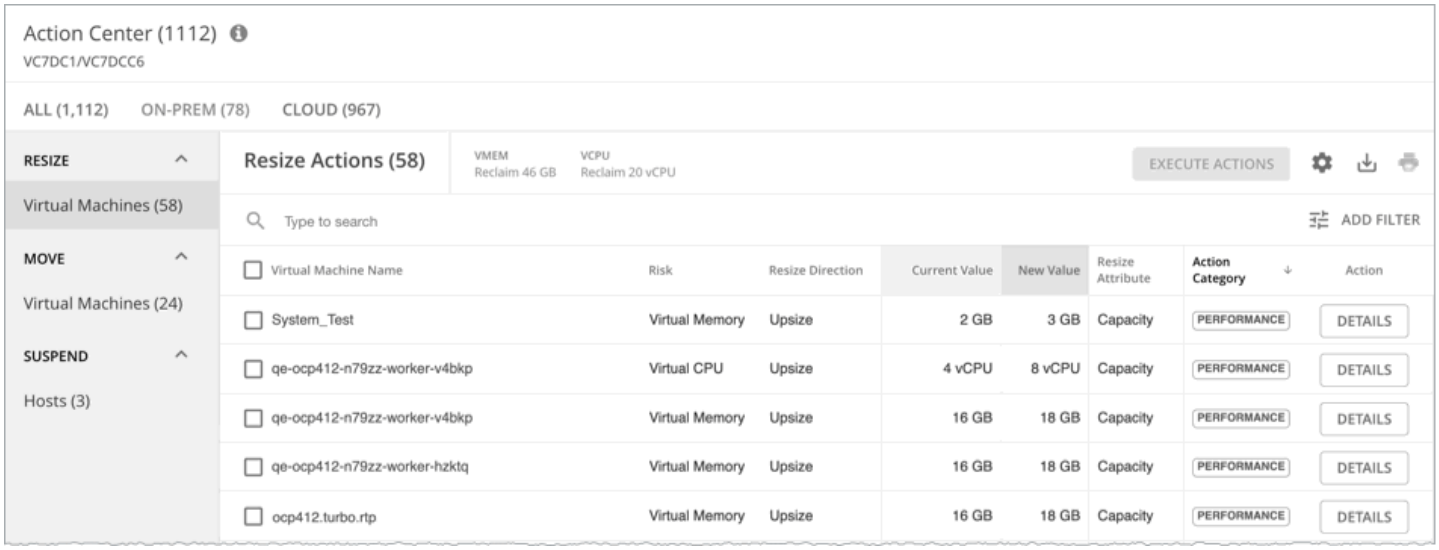
At first glance, individual actions might appear trivial and it is instinctively convenient to ignore them. It is important to keep in mind that executing a single action can impact other workloads in a meaningful way, helping move these other workloads closer

to their desired state. However, if you find that a recommended action is not acceptable (for example, if it violates existing business rules), you can set up a policy with your preferred action.

In some cases, actions can introduce disruptions that you want to avoid at all costs. For example, during critical hours, Workload Optimization Manager might execute a resize action on a mission critical resource, which then requires that resource to restart. It is important to anticipate these disruptions and plan accordingly. For example, you can create a group for all critical resources and then schedule the execution of actions to off-peak hours or weekends.

Working with Action Center

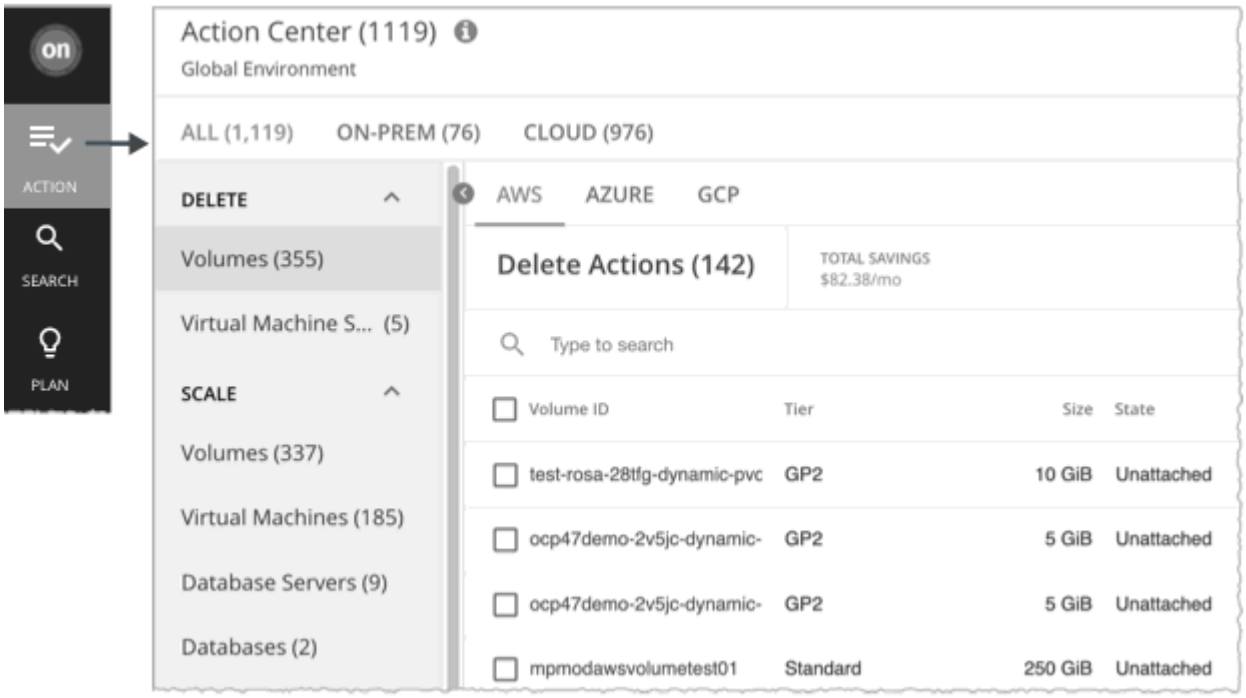
When you start using Workload Optimization Manager, all the actions that the product generates appear as pending. Use Action Center to review and execute pending actions.



Viewing Action Center

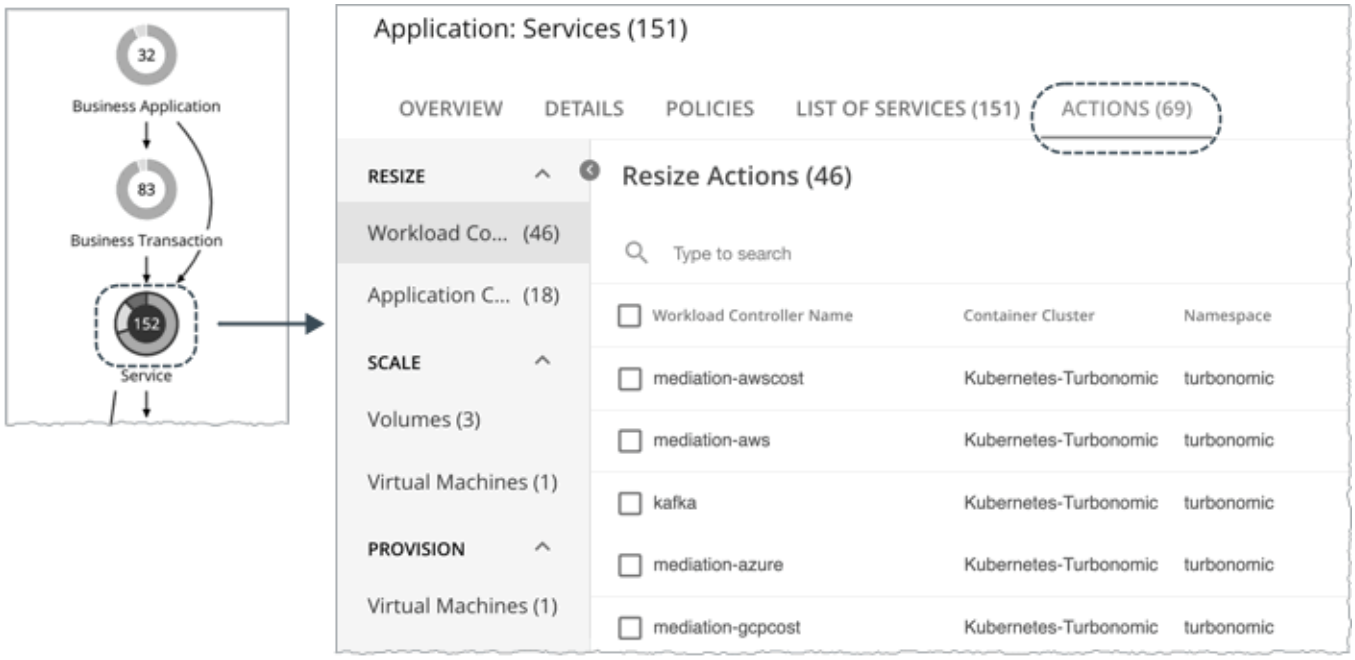
Action Center displays in areas of the product that reference actions, providing a consistent view and experience. Here are some ways you can view Action Center.

- The Action button on the main menu provides easy access to all the current pending actions.



Action Center in this view is its own dedicated page. Bookmark the page or share it with interested parties.

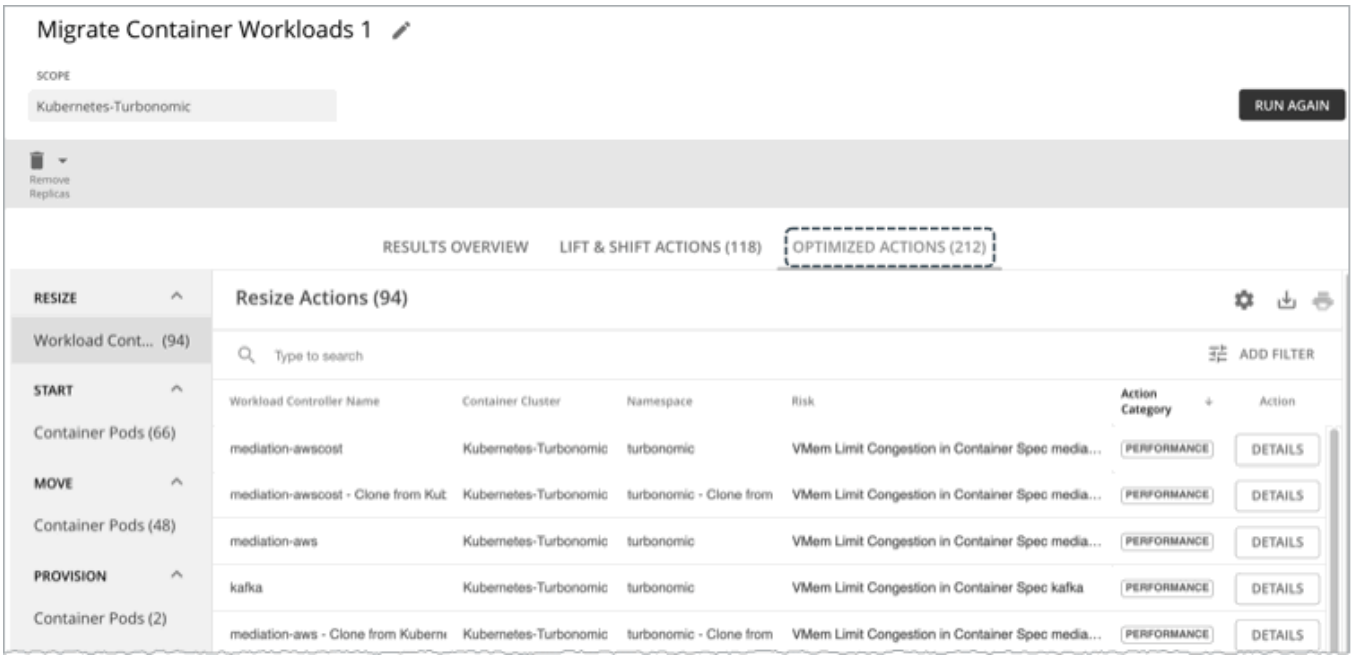
- To focus on actions for a specific entity type, click the entity type in the supply chain, and then click the **Actions** tab.



- You can launch Action Center from action-focused charts, such as the Pending Actions, Top Utilized, and Potential Savings charts.

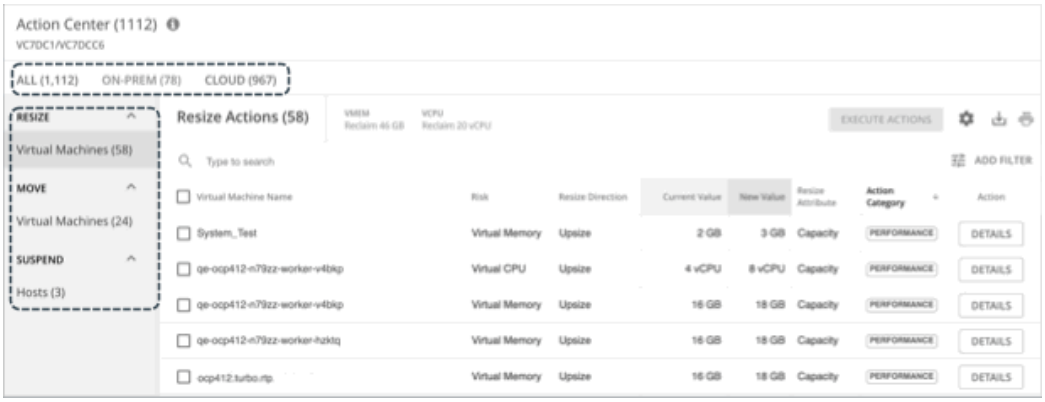

Pending Actions
Global Environment
662
Scaling Actions
76
Start/Buy Actions
3
Stop Actions
360
Delete Actions
5
Discount Optimization Actions
22
Placement Actions
SHOW ALL
Action Center (1119)
Global Environment
ALL (1,119)
ON-PREM (76)
CLOUD (976)
DELETE
Volumes (355)
Virtual Machine S... (5)
SCALE
Volumes (337)
Virtual Machines (185)
Database Servers (9)
Databases (2)
AWS
AZURE
GCP
Delete Actions (142)
Type to search
Top Host Clusters, By Headroom
Global Environment
SORT BY: CPU HEADROOM
Name
Virtual Machin...
Headroom
Time To Exhaustion
Actions
VCSDC1/VCSDC1C1
2
CPU
3%
More than ...
2 ACTIONS
Mem...
50%
42 days
Stor...
2%
More than ...
Total...
50%
VC7DC1/VC7DC1C4-QE/
1
CPU
3%
More than ...
1 ACTION
Mem...
1%
More than ...
Stor...
12%
More than ...
Total...
1%
VC7DC1/VC7DC1C1-QE/
11
CPU
3%
More than ...
4 ACTIONS
Mem...
41%
More than ...
Stor...
12%
More than ...
SHOW ALL
Action Center (4)
MOVE
Virtual Machines (4)
Move Actions (4)
Type to search
Virtual Machine Name
Type
PS DRS - Affinity - Keep On esx41
Host
PS DRS - KeepTogether - Keep On esx
Host
PS DRS - KeepSeparate 1 - Keep On e
Host
PS DRS - AntiAffinity - Keep On esx41
Host
Potential Savings
Global Environment
$14.79 k
Monthly
$7,436.13/mo, 149 Scale Virtual Machines
$3,061.86/mo, 74 Buy RI
$1,780.93/mo, 333 Scale Volumes
$1,119.61/mo, 355 Delete Volumes
$769.25/mo, 14 Other
$621.23/mo, 11 Scale Database Servers
SHOW ALL
Potential Savings (934)
SCALE
Virtual Machines (149)
Volumes (333)
Database Servers (9)
Virtual Machine Specs (1)
Databases (2)
BUY
Reserved Instance (74)
AWS
AZURE
GCP
Scale Actions (57)
TOTAL SAVINGS
$2,734.21/mo
Type to search
Virtual Machine Name
Instance Type
eks-cluster-eks-cluster-ng
m5a.4xlarge
i-09b3a3ebd5f1d713e
t2.2xlarge
eks-cluster-eks-cluster-ng
m5a.4xlarge
ocp47demo-2v5jc-worker
m5a.4xlarge

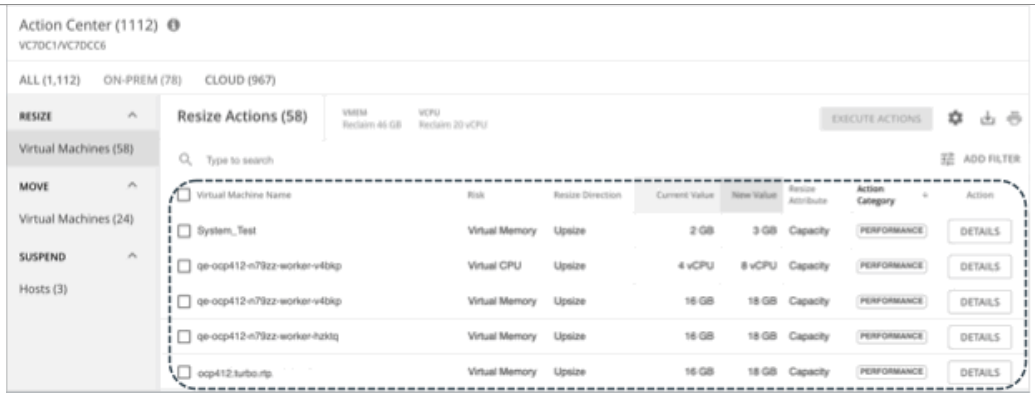
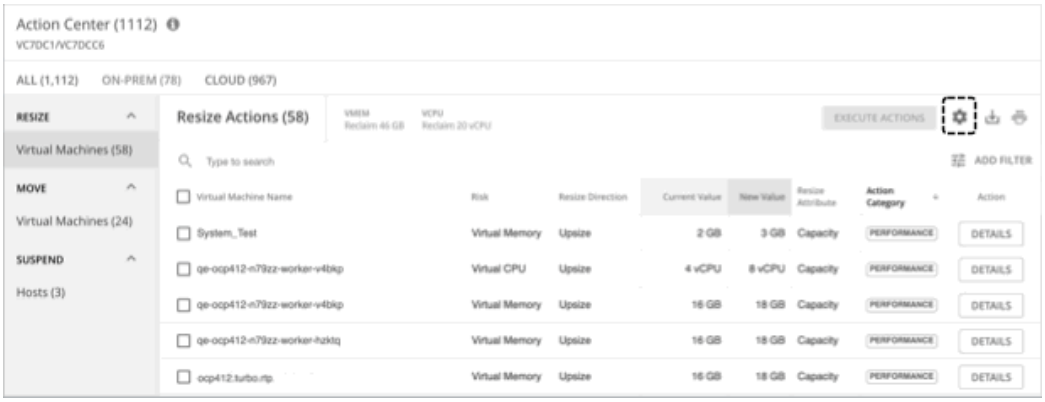
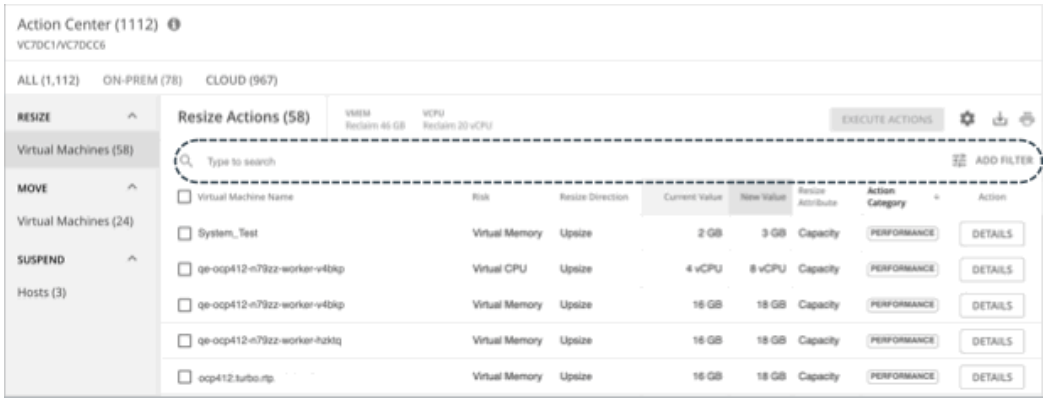
- After you run a plan *(on page 850)*, click an Actions tab to open Action Center.

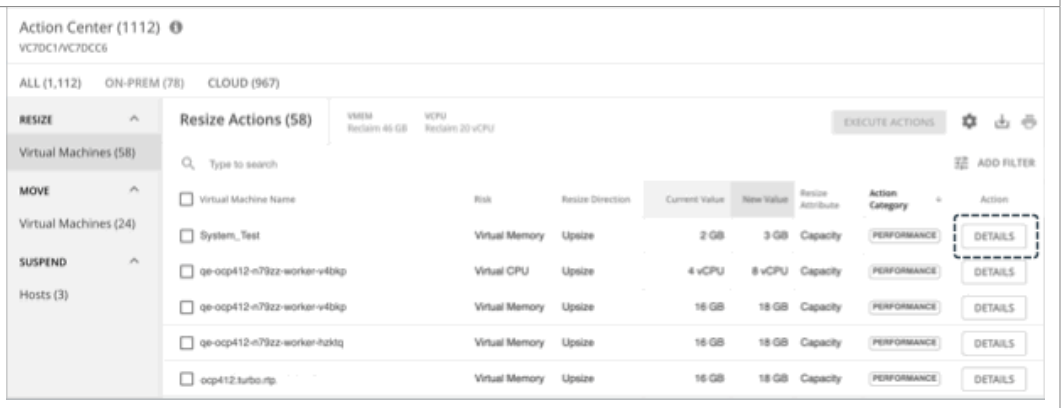


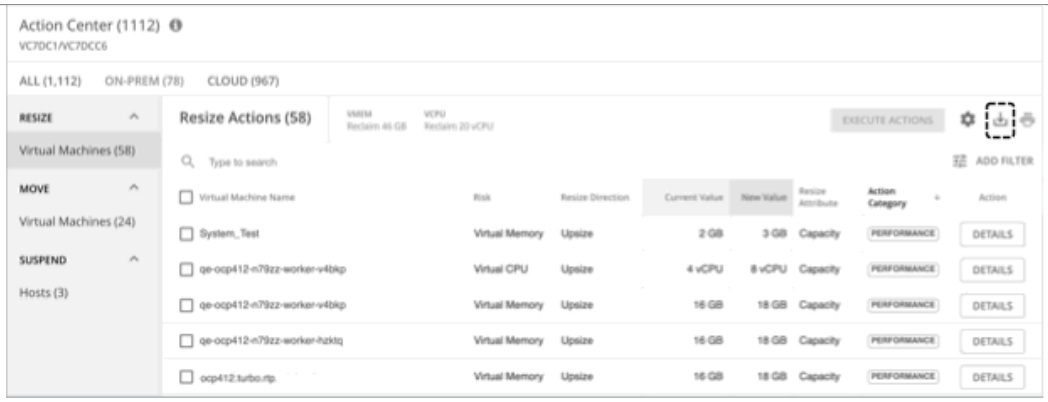
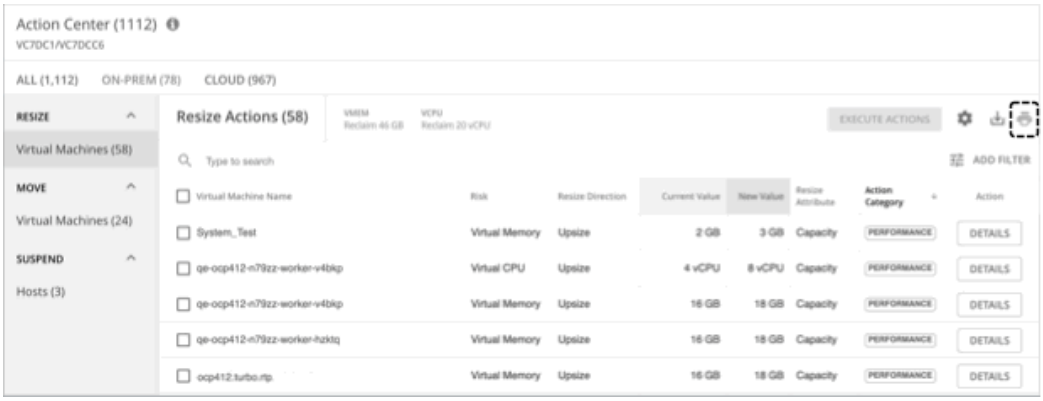
Action Center features

Action Center provides a comprehensive view of actions. The following features are available to help you work with actions with ease.

Feature	Description
Action groups	Action Center provides meaningful groups of actions.  ▪ Actions are grouped by environment type (on-prem or cloud). For a cloud environment, Action Center also groups actions by cloud provider. ▪ Actions are grouped by action types *(on page 845)*. These actions may apply to one or several entities.
Table columns	Actions are presented in tabular format. Columns in the table provide information that is relevant to the action type and entity that you are currently viewing.

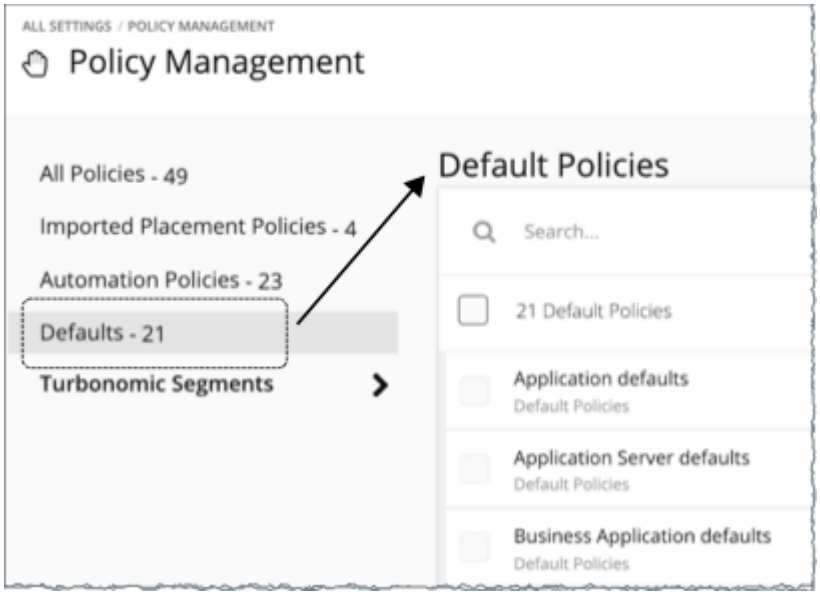
Feature	Description
	
Configure columns	The Configure columns button allows you to customize the display of columns. You can show or hide certain columns, or rearrange them as needed. 
Search and Add Filter button	For a long list of actions, use Search or filters to narrow the results. The filter options you see are those that are relevant to the action type and entity that you are currently viewing. 
Details button	The Details button for each action opens the Action Details *(on page 825)* page, which provides additional information to help you understand why Workload Optimization Manager recommends an action and what you would gain if you execute it.

Feature	Description
	
Action execution	If an action is executable, you can execute it from Action Center. Select one or several actions that you want to execute, and then click **Execute Actions**.  ▪ Some actions may not be executable. For example, the action acceptance *(on page 848)* setting in a policy may have been set to *Recommend*, or the underlying technology for the entity does not support automation. The check box for a non-executable action is grayed out. Review the action and then execute it outside Workload Optimization Manager. ▪ Some actions can only be executed after certain prerequisites are met. For example, in order to suspend Host A, VM_01 in the host must first move to Host B. However, Host B only has capacity for one VM and is currently hosting VM_02. In this case, Host A suspension is blocked by two prerequisite actions – VM_02 moving to another host and VM_01 moving to Host B. The check box for an action with prerequisites is grayed out and includes the prohibition symbol (⊘). Click the **Details** button for the action to view prerequisite information. When all the prerequisites have been satisfied, the action becomes executable.
Download button	Download the actions you are currently viewing, or all the actions that are currently pending.

Feature	Description
	
Print Action Details button	Select one or several actions (maximum 75) that you want to print, and then click **Print Action Details** to open a printable preview of the selected Action Details *(on page 825)* pages. You can then print the action details for all of the selected actions at once. 

Action handling

Workload Optimization Manager will never execute actions automatically, unless you tell it to. If you examine the default policies that ship with the product, you will notice that these policies do not enable automation on any action. Workload Optimization Manager gives you full control over all automation decisions.



At first, your goal may be to evaluate all the actions in Action Center and then execute those that result in immediate improvements in performance and resource utilization. Over time, you develop and fine-tune your action-handling process to meet productivity goals and respond to changing business needs. This process could lead to the following key decisions:

- Disabling actions that should never execute, such as those that violate business rules

  Workload Optimization Manager will not consider recommending disabled actions when it performs its analysis.
- Allowing certain actions to execute automatically, such as those that assure performance on mission-critical resources

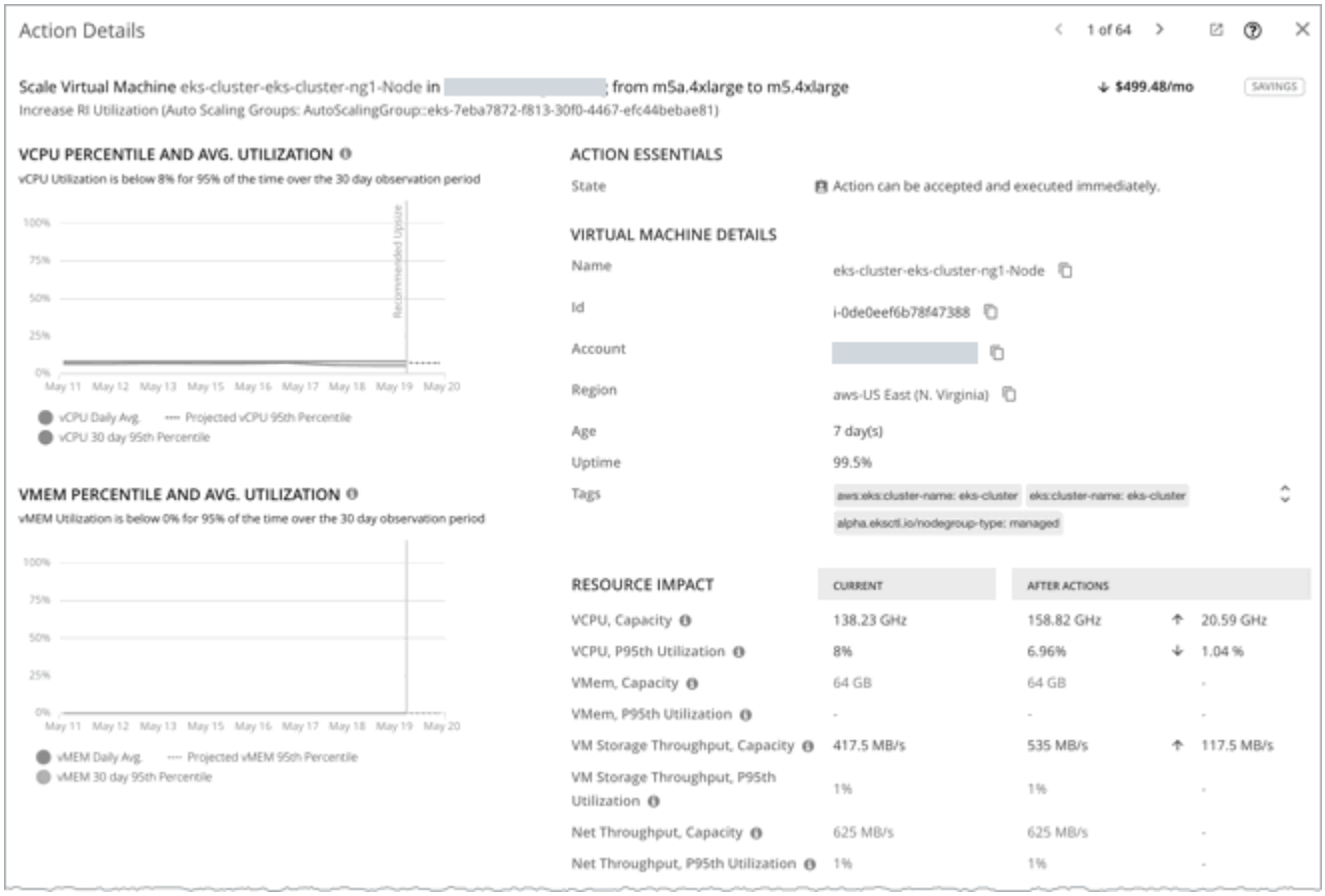
  Automation simplifies your task, while ensuring that workloads continue to have adequate resources to perform optimally. As such, it is important that you set the goal of automating as many actions as possible. This requires evaluating which actions are safe to automate, and on which entities.
- Continuing to let Workload Optimization Manager post certain actions so you can execute them on a case-by-case basis

  For example, certain actions might require the approval of specific individuals. In this case, you would want Workload Optimization Manager to post those actions for review and only execute the actions that receive an approval.

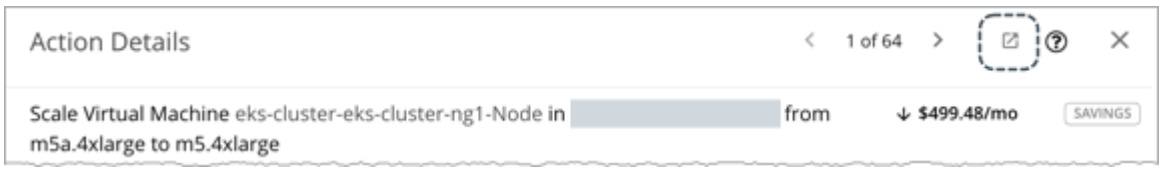
  These are the actions that you would look for in Action Center. These actions stop showing after you execute them, if you disable or automate them, or if the environment changes in the next market analysis such that the actions are no longer needed.

Action details

The Action Details page provides a comprehensive set of information to help you understand why Workload Optimization Manager recommends an action and what you would gain if you execute it.

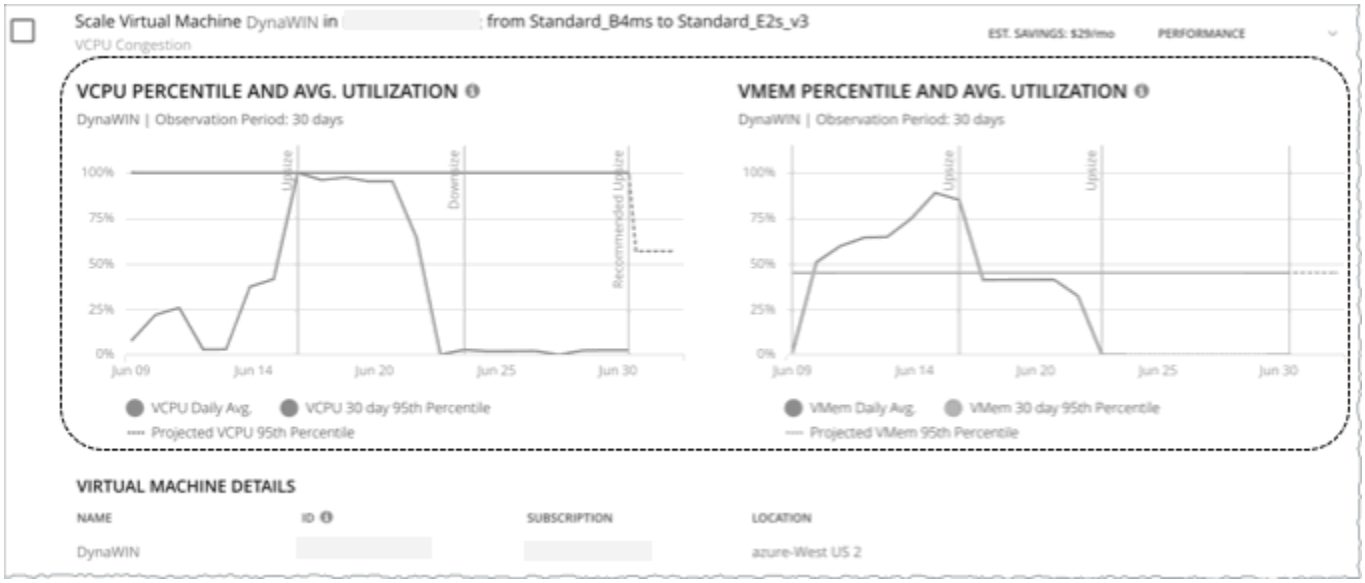


The Action Details page for a specific action has a unique URL (or deep link) that you can share with interested parties. To get the URL, click the share icon and then copy the URL displayed in the address bar of the new tab.



Utilization charts

Workload Optimization Manager uses percentile calculations to measure resource utilization more accurately, and drive actions that improve overall utilization and reduce costs for cloud workloads. When you examine the details for an entity or pending action, you will see charts that highlight resource *utilization percentiles* for a given observation period, and the projected percentiles after you execute the action.



The charts also plot *daily average utilization* for your reference. If you have previously executed scaling actions on the entity, you can see the resulting improvements in daily average utilization. Put together, these charts allow you to easily recognize utilization trends that drive Workload Optimization Manager's recommendations.

Notes:

- You can set constraints in policies to refine the percentile calculations.
- After you execute an action, it might take some time for the charts to reflect the resulting improvements.

Entities with utilization charts

Utilization charts display for actions on the following entity types:

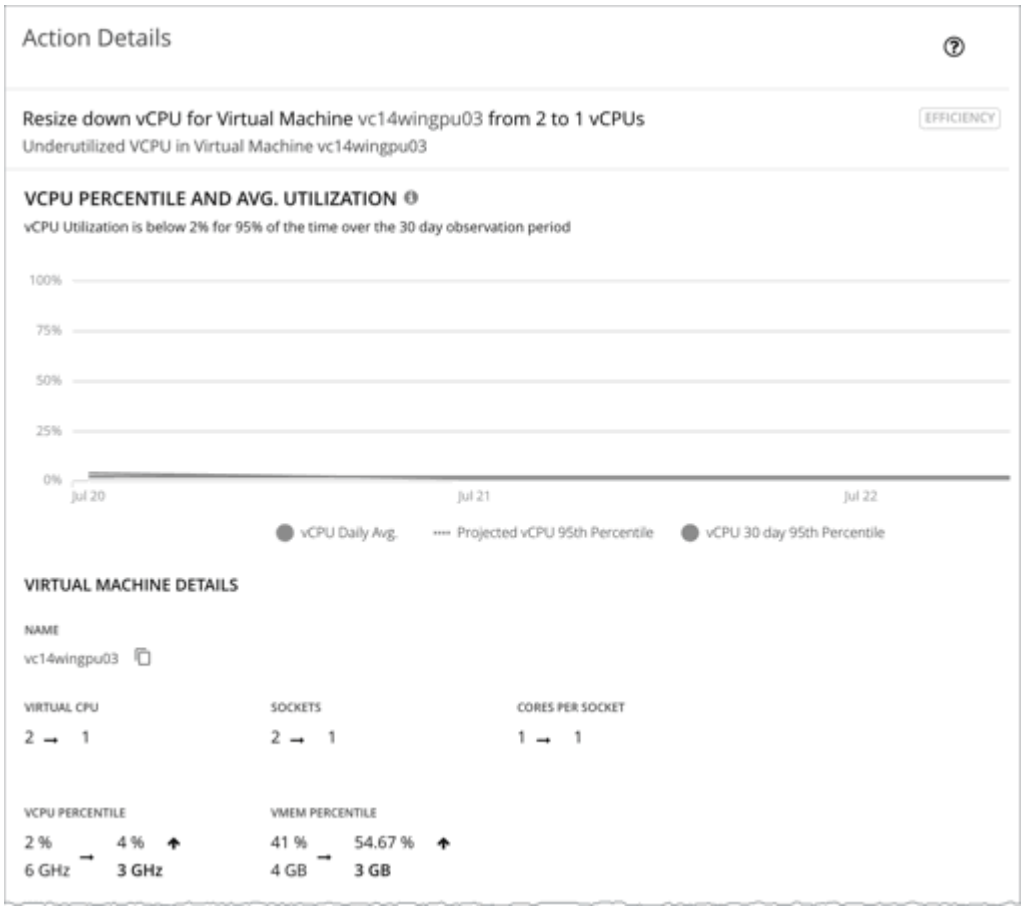
Entity type	Monitored resources		Notes
	Percentile utilization	Average utilization	
Application Component	▪ Heap ▪ Garbage Collection	▪ Heap ▪ Garbage Collection	For Application Components, you will see either a Heap or Garbage Collection chart for Heap resize actions. See Application Component Actions *(on page 553)*.
Virtual Machine	▪ vCPU ▪ vMem	▪ vCPU ▪ vMem	For on-prem VMs, you will see either a VCPU or VMem chart, depending on the commodity that needs to scale. For cloud VMs and VMs in Migrate to Cloud plans, both charts display. These charts also appear when you scope to a given VM (on-prem or cloud) and view the Details page.

Entity type	Monitored resources		Notes
	Percentile utilization	Average utilization	
			They also appear in Migrate to Cloud plan results.
Virtual Machine Spec	▪ vCPU ▪ vMem	▪ vCPU ▪ vMem ▪ Storage ▪ Number of replicas	See Virtual Machine Spec Actions *(on page 674)*.
Database (cloud)	▪ DTU Pricing Model – DTU ▪ vCore Pricing Model – vCPU – vMem – IOPS – Throughput ▪ RU Pricing Model – RU	▪ DTU Pricing Model – DTU – Storage ▪ vCore Pricing Model – vCPU – vMem – IOPS – Throughput – Storage ▪ RU Pricing Model – RU	See Cloud Database *(on page 703)*.
Database Server (AWS RDS)	▪ vCPU ▪ vMem ▪ IOPS	▪ vCPU ▪ vMem ▪ IOPS	See AWS RDS Actions *(on page 685)*.
Database Server (On-prem)	▪ DB Memory ▪ DB Cache Hit Rate	▪ DB Memory ▪ DB Cache Hit Rate	For On-prem Database Servers, you will see either a DB Memory or DB Cache Hit Rate chart for DB Memory resize actions. See Database Server (On-prem) Actions *(on page 769)*.
Document Collection	RU	RU	See Document Collection Actions *(on page 723)*.
Volume (cloud)	▪ IOPS ▪ Throughput	▪ IOPS ▪ Throughput	These charts also appear when you scope to a given volume and view the Details page. See Cloud Volume Actions *(on page 728)*.
Workload Controller	▪ vCPU limits and requests ▪ vMem limits and requests	▪ vCPU limits, throttling, and requests ▪ vMem limits and requests	See Container Actions *(on page 579)*.

Action Details - vCPU resizes for on-prem VMs

The Action Details page provides a comprehensive set of information to help you understand why Workload Optimization Manager recommends a vCPU resize down action and the impact of this action on the VM and its host.

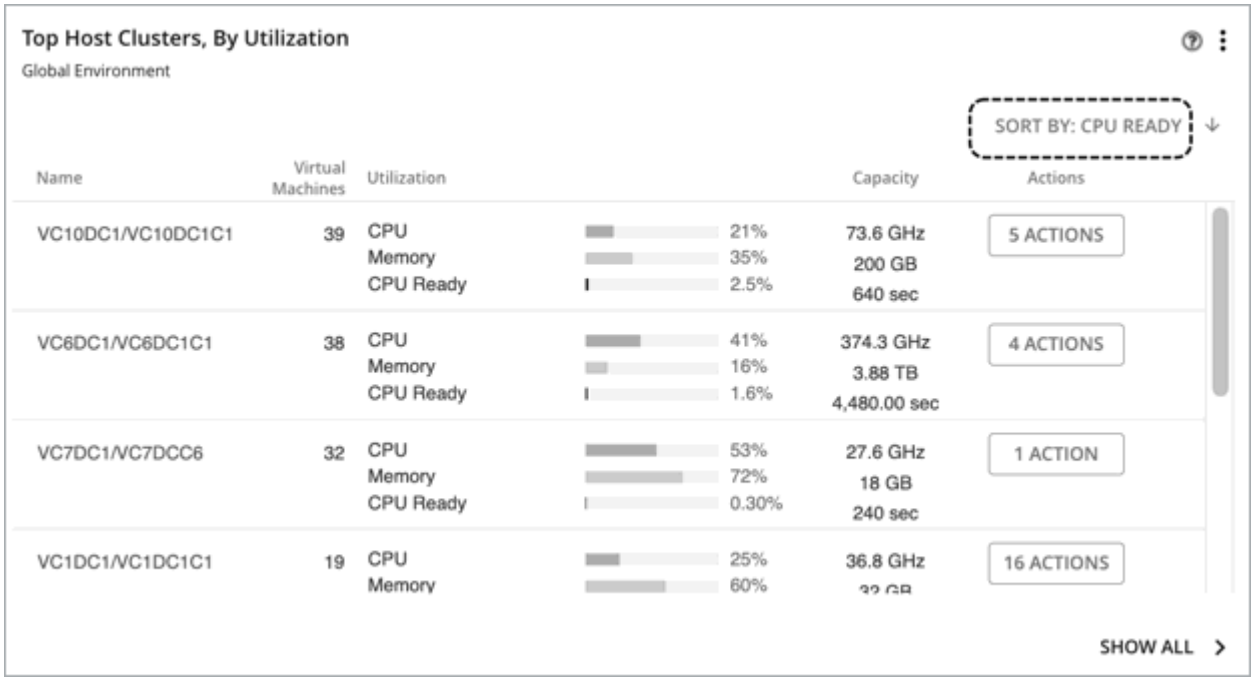
Actions to resize down vCPU can result in significant performance improvements in environments where CPU Ready on virtualization hosts is high.



Addressing CPU ready issues by resizing down vCPU

CPU Ready is a measurement of how long a VM is waiting for access to the host's physical CPU. In VMware environments, a CPU Ready value of 5% (or 1000 milliseconds) is considered a performance risk for general purpose VMs. As the VM and its application increase in priority/performance requirements, the acceptable CPU Ready value decreases.

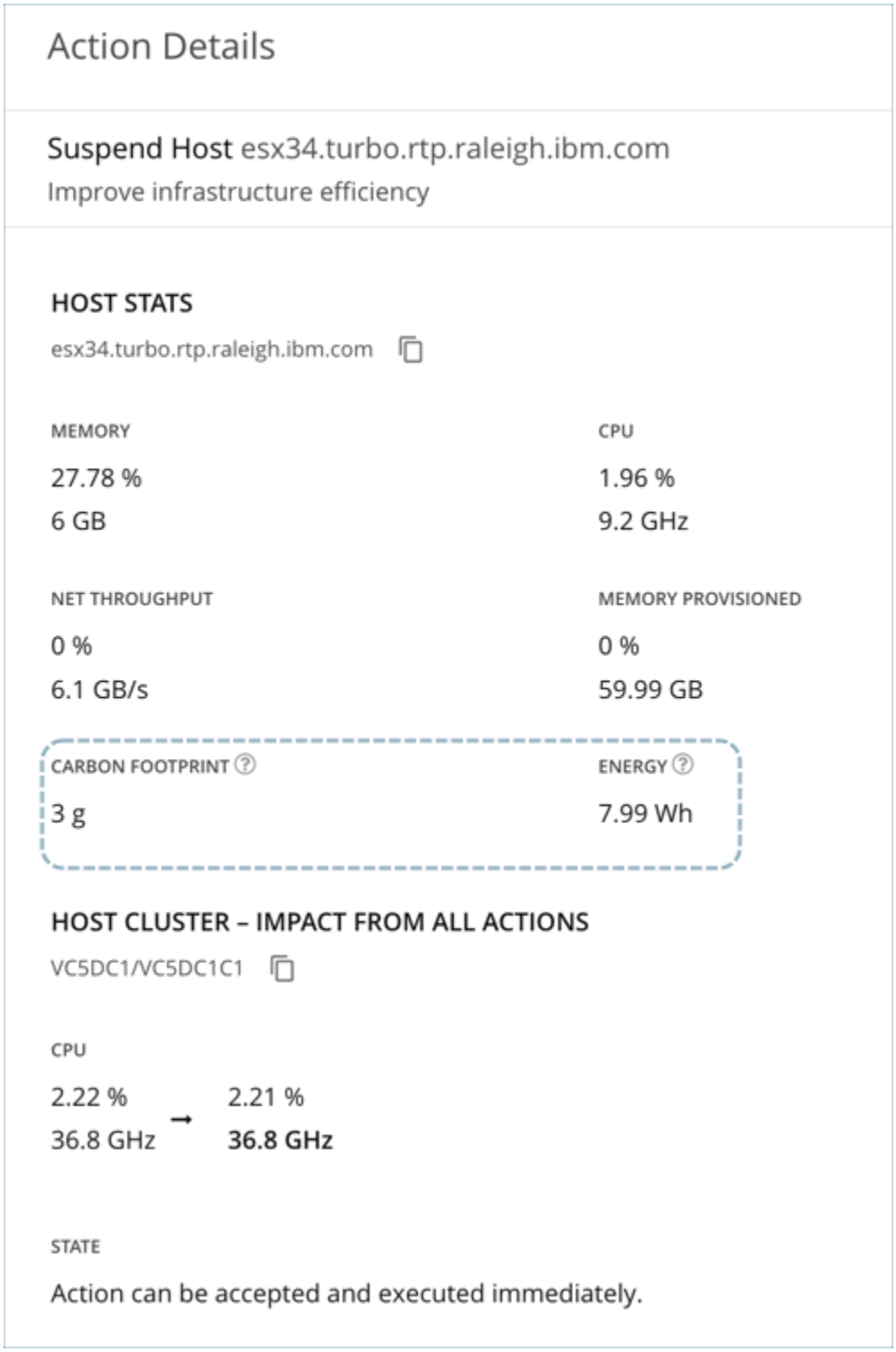
To identify risks related to CPU Ready in your environment, add the Top Host Clusters by Utilization chart to your dashboard and then sort the chart data by CPU Ready.



As you examine the pending actions for these clusters, check for actions to resize down VM vCPU and then prioritize them for execution. These actions offer the greatest opportunity to resolve CPU Ready issues.

Action details- host suspend impact on energy usage and carbon footprint

Workload Optimization Manager shows the projected change in energy usage and carbon footprint in the action details for host suspend actions. This information can help you identify which hosts with executable suspend actions are energy inefficient and therefore the best candidates for suspension.



In Action Center, when you click the **Details** button for a host suspend action, the **Action Details** page includes a **Host Stats** section that shows the current energy usage and carbon footprint of the host and any VMs running on that host. If no VMs are running on the host and you execute the host suspend action, these values become 0. Therefore, these values also reflect the projected reductions in energy usage and carbon footprint that the host suspend action creates.

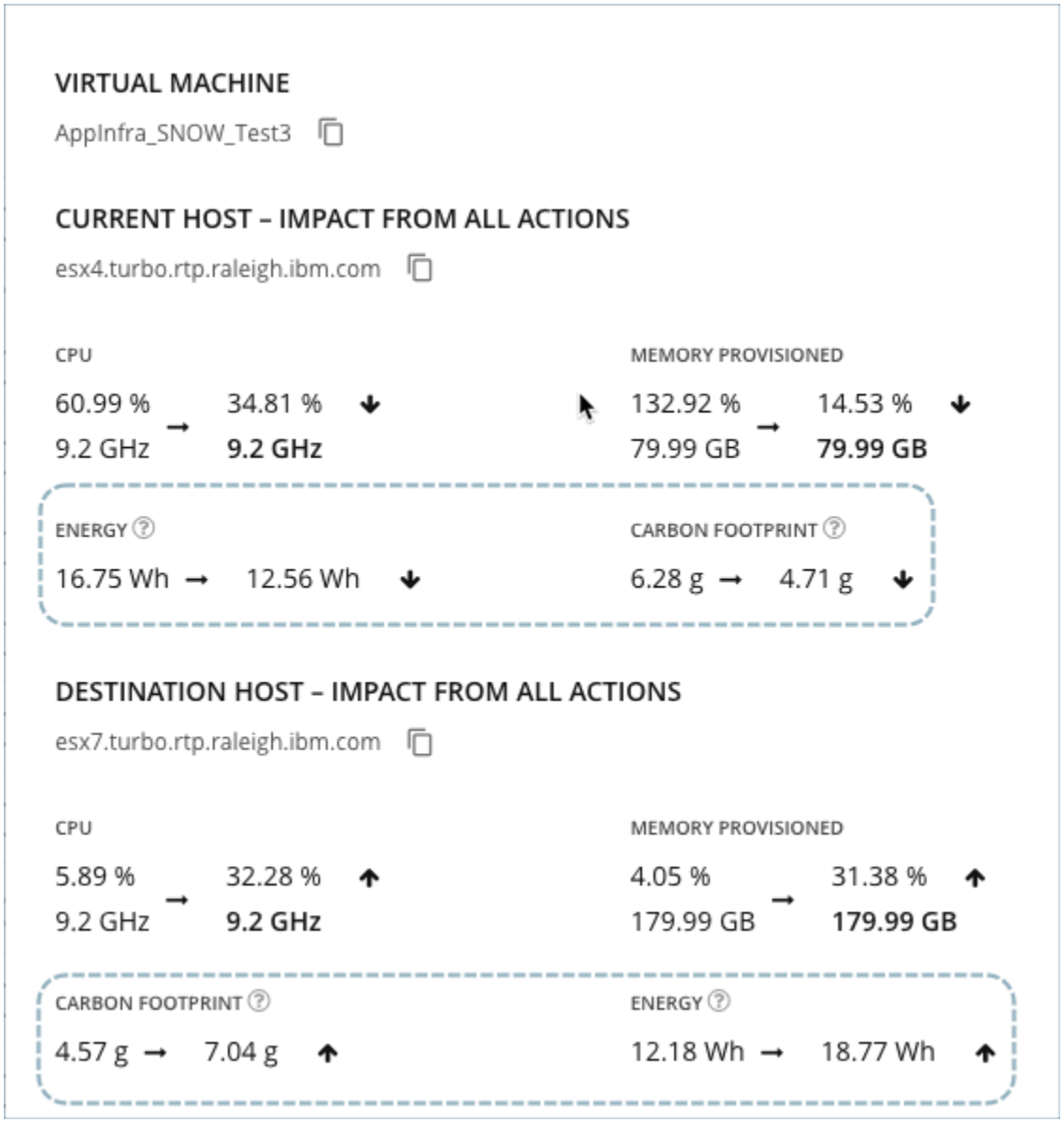
However, if the host still has running VMs when Workload Optimization Manager generates the host suspend action, Workload Optimization Manager also generates actions to move those VMs to other hosts. The host suspend action is blocked until all VM move actions for the host are complete. In this case, the values reflect the combined carbon footprint and combined energy usage of the host and all VMs that are running on that host. To understand the projected reductions, you must first move all the

running VMs off of the host. The updated values then reflect the potential reductions in energy usage and carbon footprint from suspending the host.

For more information about these values, see Carbon Footprint Calculation *(on page 978)* and Calculation of Energy Consumed by VMs *(on page 987)*.

Action details- VM move impact on energy usage and carbon footprint

Workload Optimization Manager shows the projected change in energy usage and carbon footprint in the action details for VM move actions. You can use these metrics to understand projected impact on energy usage and carbon footprint that is associated with moving VMs off of a host.



For both the source and destination hosts, Workload Optimization Manager calculates the projected energy usage and carbon footprint based on all the actions that are currently pending for the host. Workload Optimization Manager takes all each pending action's impact on the host's energy usage and carbon footprint into account to calculate the projected values after all actions are executed.

In Action Center, when you click the **Details** button for a VM move action, the **Action Details** page includes the following sections that show energy usage and carbon footprint data.

- **Current Host - Impact from All Actions**

  This section includes four sustainability values:

  - The current energy usage and carbon footprint of the VM's host
  - The projected energy usage and carbon footprint of that host after the VM moves to another host and all pending actions for the host are executed

- **Destination Host - Impact from All Actions**

  This section includes four sustainability values:

  - The current energy usage and carbon footprint of the host to which the VM moves
  - The projected energy usage and carbon footprint of that host after the VM moves and all pending actions for the host are executed

For more information about these values, see Carbon Footprint Calculation *(on page 978)* and Calculation of Energy Consumed by VMs *(on page 987)*.

Calculating projected energy usage and carbon footprint for VM move Actions

By observing the total CPU utilization and the reported power (electricity consumed) for a host, Workload Optimization Manager generates a power model that is used to compute the power at any level of CPU utilization on that host. From the computed power, Workload Optimization Manager computes the energy usage and carbon footprint metrics for the host. The power model also enables Workload Optimization Manager to project the impact of move actions on the energy usage and carbon footprint of the host.

Workload Optimization Manager derives the projected energy usage on the host following a move action from the projected power, which in turn is derived from the projected CPU that Workload Optimization Manager generates from market analysis.

Similarly, the projected carbon footprint of the host following a move action is computed by applying the carbon-energy relationship on the projected energy computed in the previous step:

```
Projected carbon footprint = projected energy * PUE * CI
```

For more information about the carbon-energy relationship, Power Usage Effectiveness (PUE), and Carbon Intensity (CI), see Carbon Footprint Calculation *(on page 978)*.

Action details- VM resize impact on energy usage and carbon footprint

Workload Optimization Manager shows the projected change in energy usage and carbon footprint in the action details for on-prem VM resize actions. You can use these metrics to understand projected impact on energy usage and carbon footprint that is associated with resizing VMs.



In Action Center, when you click the **Details** button for a VM resize action, the **Resource Impact** table on the **Action Details** page includes rows for **Energy, Usage** and **Carbon Footprint, Usage**.

- The **Energy, Usage** row shows the aggregated energy usage in watt-hours over a 10-minute interval. The **Current** value is the latest measurement for energy usage. The **After Actions** value is the projected 10-minute energy usage after all pending actions are executed.
- The **Carbon footprint, Usage** row shows the aggregated carbon footprint in grams over a 10-minute interval. The **Current** value is the latest carbon footprint measurement. The **After Actions** value is the projected carbon footprint for a 10-minute interval after all pending actions are executed.

For more information about these values, see Carbon Footprint Calculation *(on page 978)* and Calculation of Energy Consumed by VMs *(on page 987)*. For more information about how Workload Optimization Manager calculates the change in these values due to resize actions, see the next section.

Calculating projected energy usage and carbon footprint for on-prem VM resize Actions

Workload Optimization Manager calculates energy usage for each VM on a host, based on two types of host energy:

- **Host overhead energy**: Energy that is consumed by non-VM resources, such as hypervisors. This value includes host idle energy, the energy consumed by a host while it is in an idle state, when no processes are actively executing.

  Host overhead energy is attributed to all member VMs based on each VM's allocated capacity.
- **Host active energy**: The energy consumed by the VMs' operating systems and active processes.

  Host active energy is attributed to all member VMs based on each VM's usage of host CPU.

The calculation can be expressed as follows:

```
(VM Size / Total Size for all VMs) * Host Overhead Energy for VMs +
(VM CPU Utilization / Total CPU Utilization for all VMs) * Host Active Energy for VMs
= VM Energy usage (in Wh)
```

> **NOTE:**
> Workload Optimization Manager rounds the calculated values.

The impact of resize actions on VM energy usage can be understood as a change in the values of `VM Size` and `VM CPU Utilization` in the previous equation. As a VM resizes down, its share of the total size and CPU utilization for all VMs on the host also decreases, which in turn decreases the calculated `VM Energy usage (in Wh)` value.

Similarly, the projected carbon footprint of the VM following a resize action is computed by applying the carbon-energy relationship to the projected energy usage from previous step:

```
Projected carbon footprint = projected energy * PUE * CI
```

For more information about the carbon-energy relationship, Power Usage Effectiveness (PUE), and Carbon Intensity (CI), see Carbon Footprint Calculation *(on page 978)*.

However, the **After Actions** values for energy usage and carbon footprint that appear on the **Action Details** page are the projected values after *all* pending actions are executed, not just resize actions. In some cases, the **After Actions** values might be greater than the **Current** values. This comparison does not indicate that overall energy usage increases when the action are executed. Instead, as a result of pending move actions on the same host, this particular VM might represent a larger share of the overall energy usage and carbon footprint after other VMs are moved off the host.

Actions by entity type

Workload Optimization Manager generates actions based on how entity types use or provide resources, and what each entity type supports.

The following tables show the actions that each entity type supports:

Application entity types

Entity type	Supported actions
Business application	None Workload Optimization Manager does not recommend actions for a Business Application, but it does recommend actions for the underlying Application Components and infrastructure. The Pending Actions chart for a Business Application lists these actions, thus providing visibility into the risks that have a direct impact on the Business Application's performance.
Business transaction	None Workload Optimization Manager does not recommend actions for a Business Transaction, but it does recommend actions for the underlying Application Components and infrastructure. The Pending Actions chart for a Business Transaction lists these actions, thus providing visibility into the risks that have a direct impact on the Business Transaction's performance.
Service	▪ For APM services: None Workload Optimization Manager does not recommend actions for services discovered through APM targets, but it does recommend actions for the underlying Application Components and nodes. The Pending Actions chart for services list these actions, thus providing visibility into the risks that have a direct impact on their performance. ▪ For container platform services: Scale (through workload controllers) Workload Optimization Manager recommends actions to scale the replicas that back the container platform services. Set the scope to the associated Workload Controller to view and execute the actions. For details, see Workload Controller Scale Actions *(on page 598)*.
Application component	Resize

Entity type	Supported actions
	Resize the following resources to maintain performance: ▪ Thread pool Workload Optimization Manager generates thread pool resize actions. These actions are recommend-only and can only be executed outside Workload Optimization Manager. ▪ Connections Workload Optimization Manager uses connection data to generate memory resize actions for on-prem Database Servers. ▪ Heap Workload Optimization Manager generates Heap resize actions if an Application Component provides Heap and Remaining GC Capacity, and the underlying VM or container provides VMem. These actions are recommend-only and can only be executed outside Workload Optimization Manager. **NOTE:** Remaining GC capacity is the measurement of Application Component uptime that is *not* spent on garbage collection (GC). The resources that Workload Optimization Manager can resize depend on the processes that it discovers from your Applications and Databases targets. Refer to the topic for a specific target to see a list of resources that can be resized.

Container platform entity types

Entity type	Supported actions
Service	Scale (through workload controllers) Workload Optimization Manager recommends actions to scale the replicas that back the container platform services. Set the scope to the associated Workload Controller to view and execute the actions. For details, see Workload Controller Scale Actions *(on page 598)*.
Container	Resize Workload Optimization Manager generates container resize actions for Bare Pods where the workload has no controller.
Container spec	Resize (through workload controllers) A container spec retains the historical utilization data of ephemeral containers. Workload Optimization Manager uses this data to make resize decisions that assure optimal utilization of resources. By default, all replicas of the same container for the same workload type resize consistently. For details, see Workload Controller Resize Actions *(on page 593)*.
Namespace	Resize quota Workload Optimization Manager treats quotas defined in a namespace as constraints when making resize decisions. If existing actions would exceed the namespace quotas, Workload Optimization Manager recommends actions to resize up the affected namespace quota.

Entity type	Supported actions
	Note that Workload Optimization Manager does not recommend actions to resize *down* a namespace quota. Such an action reduces the capacity that is already allocated to an application. The decision to resize down a namespace quota should include the application owner. For details, see Namespace Actions *(on page 616)*.
Workload controller	▪ Resize Actions associated with a workload controller resize container specs vertically to assure optimal utilization of resources. This is a natural representation of these actions because the parent controller's container specs for vCPU limits/requests and vMem limits/requests are modified. The workload controller then rolls out the changes in the running environment. For details, see Workload Controller Resize Actions *(on page 593)*. Resize actions for certain types of workloads may be blocked or set to recommend-only, which means that they are currently not executable in Workload Optimization Manager. These actions usually require a dependent action or an external configuration to be completed first. For example, a workload controller resize up action may be temporarily blocked if there is insufficient resources in the corresponding node or namespace. In this case, a node provision or namespace quota resize action must be completed first. For more information, see this topic *(on page 291)*. ▪ Scale Actions associated with a workload controller scale replicas horizontally to maintain Service Level Objectives (SLOs) for your applications. This is a natural representation of these actions because the parent controller's number of replicas are modified. The workload controller then rolls out the changes in the running environment. For details, see Workload Controller Scale Actions *(on page 598)*.
Container pod	▪ Move Move a pod that is controlled by a ReplicationController, ReplicaSet, or Deployment (through a ReplicaSet) to another node (VM) to address performance issues or improve infrastructure efficiency. For example, if a particular node is congested for CPU, you can move pods to a node with sufficient capacity. If a node is underutilized and is a candidate for suspension, you must first move the pods before you can safely suspend the node. For details, see Container Pod Move Actions *(on page 606)*. ▪ Provision or suspend When recommending node provision or suspend actions, Workload Optimization Manager will also recommend provisioning pods (based on demand from DaemonSets) or suspending the related pods. For details, see Container Pod Provision and Suspension Actions *(on page 610)*.
Container platform cluster	None Workload Optimization Manager does not recommend actions for a container platform cluster. Instead, it recommends actions for the containers, pods, nodes (VMs), and volumes in the cluster. Workload Optimization Manager shows all of these actions when you scope to a container platform cluster and view the Pending Actions chart.

Entity type	Supported actions
Container platform node (VM)	A node (cloud or on-prem) is represented as a Virtual Machine entity in the supply chain. **NOTE:** For nodes in the public cloud, Workload Optimization Manager reports the cost savings or investments attached to these actions. For details, see Node Actions *(on page 626)*.

Cloud infrastructure entity types

Entity Type	Supported Actions
Virtual machine (cloud)	For details about cloud VM actions, see the following topics: ▪ Actions for AWS VMs *(on page 637)* ▪ Actions for Azure VMs *(on page 641)* ▪ Actions for Google Cloud VMs *(on page 646)*
Virtual machine spec	▪ For AWS Auto Scaling groups: Stop and start (also known as 'parking' actions) Stop an Auto Scaling group for a given period of time to reduce your cloud expenses, and then start it at a later time. For details, see Parking: Stop or Start Cloud Resources *(on page 942)*. ▪ For Azure Virtual Machine Scale Sets: Stop and start (also known as 'parking' actions) Stop a Virtual Machine Scale Set for a given period of time to reduce your cloud expenses, and then start it at a later time. For details, see Parking: Stop or Start Cloud Resources *(on page 942)*. ▪ For Azure App Service plans: – Scale Scale Azure App Service plans to optimize app performance or reduce costs, while complying with business policies. – Delete Delete empty Azure App Service plans as a cost-saving measure. A plan is considered empty if it is not hosting any running apps. For details, see Virtual Machine Spec Actions *(on page 674)*.
App component spec	None Workload Optimization Manager does not recommend actions for App Component Specs, but it does recommend actions for the underlying Virtual Machine Specs. For details, see Virtual Machine Spec Actions *(on page 674)*.
Database server (cloud)	▪ For AWS RDS: – Stop and start (also known as 'parking' actions) Stop a Database Server for a given period of time to reduce your cloud expenses, and then start it at a later time.

Entity Type	Supported Actions
	– Scale Scale compute and storage resources to optimize performance and costs. For details, see Cloud Database Server Actions *(on page 685)* and Parking: Stop or Start Cloud Resources *(on page 942)*. ▪ For Azure SQL Managed Instances: Scale Change the Managed Instance to use a different instance type or tier to optimize performance and costs. For details, see Actions for SQL Managed Instances *(on page 693)*. ▪ For Azure Cosmos DB Accounts: None Workload Optimization Manager does not recommend actions for a Cosmos DB account but it does recommend actions for the databases *(on page 713)* and document collections *(on page 722)* in the account.
Database (cloud)	▪ Scale SQL database – DTU pricing model Scale DTU and storage resources to optimize performance and costs. – vCore pricing model Scale vCPU, vMem, IOPS, throughput and storage resources to optimize performance and costs. For details, see Scale Actions for SQL Databases *(on page 705)*. ▪ Scale Cosmos DB database Scale request units (RUs) to optimize performance and costs. For details, see Scale Actions for Cosmos DB Databases *(on page 714)*. ▪ Reconfigure Cosmos DB database Remove unused provisioned throughput to reduce costs. For details, see Reconfigure Actions for Cosmos DB Databases *(on page 715)*. ▪ Delete Cosmos DB database Delete a database with provisioned throughput but without any underlying document collection (container) to reduce costs. For details, see Delete Actions for Cosmos DB Databases *(on page 715)*. ▪ Suspend or stop dedicated SQL pool Suspend or stop a dedicated SQL pool (used in Azure Synapse Analytics) to reduce compute costs. – Workload Optimization Manager analysis generates suspend actions for *idle* pools. **NOTE:** Currently, Workload Optimization Manager analysis does not generate actions to start a suspended pool. You can start a suspended pool from the Parking page (see the next item) or from Azure.

Entity Type	Supported Actions
	For details, see Suspend Actions for Dedicated SQL Pools *(on page 712)*. – You can use the Parking page to stop pools (running or idle), either on-demand or according to a schedule. Use the same page to start a stopped pool. For details, see Parking: Stop or Start Cloud Resources *(on page 942)*.
Data warehouse	Suspend Suspend a Redshift provisioned cluster to reduce compute costs. For details, see Suspend actions for Redshift provisioned clusters *(on page 720)*.
Document collection	Scale Scale Request Units (RUs) to optimize performance and costs. For details, see Document Collection Actions *(on page 723)*.
Volume (cloud)	▪ Scale Scale attached volumes to optimize performance and costs. ▪ Delete Delete unattached volumes as a cost-saving measure. Workload Optimization Manager generates an action immediately after discovering an unattached volume. For details, see Cloud Volume Actions *(on page 728)*.
Zone	None Workload Optimization Manager does not recommend actions for a cloud zone.
Region	None Workload Optimization Manager does not recommend actions for a cloud region.

On-prem infrastructure entity types

Entity type	Supported actions
Virtual machine (on-prem)	▪ Resize Resize resource capacity, reservation, or limit to improve performance. ▪ Resize PU or VP (IBM PowerVM) Resize PU (processing units) or VP (virtual processor) to optimize LPAR processing unit allocation and virtual processor capacity based on historical demand collected from the HMC. ▪ Move Move a VM due to high resource utilization on VM or host, excess IOPS or latency in VStorage, workload placement violation, underutilized host (move VM before suspending host). ▪ Reconfigure

Entity type	Supported actions
	Change a VM's configuration to comply with a policy. For hypervisor targets, Workload Optimization Manager can reconfigure VMs that violate vCPU scaling policies. For details, see VCPU Scaling Controls *(on page 754)*. For details, see On-prem VM Actions *(on page 743)*.
Volume (on-prem)	▪ Move Move a VM's volume (virtual storage) due to excess utilization of the current datastore, or for more efficient utilization of datastores in the environment. Points to consider: – The default global policy includes a setting that directs Workload Optimization Manager to use relevant metrics when analyzing and recommending actions for volumes. For details, see Enable Analysis of On-prem Volumes *(on page 1052)*. – Workload Optimization Manager will not recommend moving a volume to a datastore that is currently in maintenance mode. Any volume in that datastore should move to an active datastore (for example, via vMotion). ▪ Reconfigure Reconfigure a VM's volume (virtual storage) to comply with placement policies.
Database server (on-prem)	Resize Resize the following resources: ▪ Connections Workload Optimization Manager uses connection data to generate memory resize actions for on-prem Database Servers. ▪ Database memory (DBMem) Actions to resize database memory are driven by data on the Database Server, which is more accurate than data on the hosting VM. Workload Optimization Manager uses database memory and cache hit rate data to decide whether resize actions are necessary. A high cache hit rate value indicates efficiency. The optimal value is 100% for on-prem (self-hosted) Database Servers, and 90% for cloud Database Servers. When the cache hit rate reaches the optimal value, no action generates even if database memory utilization is high. If utilization is low, a resize down action generates. When the cache hit rate is less than the optimal value but database memory utilization remains low, no action generates. If utilization is high, a resize up action generates. ▪ Transaction log Resize actions based on the transaction log resource depend on support for virtual storage in the underlying hypervisor technology. Currently, Workload Optimization Manager does not support resize actions for Oracle and Database Servers on the Hyper-V platform (due to the lack of API support for virtual storage).
Virtual datacenter	None

Entity type	Supported actions
	Workload Optimization Manager does not recommend actions for a virtual datacenter. Instead, it recommends actions for the entities that provide resources to the virtual datacenter.
Business user	Move Move a business user between desktop pools to address: ▪ Resource congestion on the image When utilization is consistently near capacity for image resources, Workload Optimization Manager can recommend moving a business user to a desktop pool that serves larger images. ▪ Resource congestion on the desktop pool When utilization is consistently near capacity for the desktop pool, Workload Optimization Manager can recommend moving a business user to a desktop pool that has more available resources. **NOTE:** To support moves, you must configure placement policies that merge *similarly configured* desktop pools. For details, see Desktop Pool Placement Policies *(on page 784)*.
Desktop pool	None Workload Optimization Manager does not recommend actions for a desktop pool. It recommends actions for the business users running active sessions in the pool.
View pod	None Workload Optimization Manager does not recommend actions for a view pod. Instead, it recommends actions for the business users that are running active sessions.
Host	▪ Start Start a suspended host when there is increased demand for physical resources. ▪ Provision Provision a new host in the environment when there is increased demand for physical resources. Workload Optimization Manager can then move workloads to that host. ▪ Suspend When physical resources are underutilized on a host, move existing workloads to other hosts and then suspend the host. ▪ Reconfigure Workload Optimization Manager generates this action in response to changing demand for software licenses. For details, see License policy *(on page 1048)*. For details, see Host Actions *(on page 788)*.
Chassis	None Workload Optimization Manager does not recommend actions for a chassis.
Datacenter	None

Entity type	Supported actions
	Workload Optimization Manager does not recommend actions for a datacenter. Instead, it recommends actions for the entities running in the datacenter.
Storage	▪ Move For high utilization of physical storage, move datastore to a different disk array (aggregate). ▪ Provision For high utilization of storage resources, provision a new datastore. ▪ Resize Increase or decrease the datastore capacity. ▪ Start For high utilization of storage resources, start a suspended datastore. ▪ Suspend For low utilization of storage resources, move served VMs to other datastores and suspend this one. ▪ Delete (datastore or volume) Delete a datastore or volume that has been suspended for a period of time. ▪ Delete (unattached files) Delete a file on a datastore that has not been accessed or modified for a period of time. For details, see Storage Actions *(on page 799)*.
Logical pool	▪ Move For high utilization of physical storage, move the logical pool to a different disk array (aggregate). ▪ Provision For high utilization of storage resources, provision a new logical pool. ▪ Resize Increase or decrease the logical pool capacity. ▪ Start For high utilization of storage resources, start a suspended logical pool. ▪ Suspend For low utilization of storage resources, move served VMs to other logical pools and suspend this one.
Disk array	▪ Provision For high utilization of the disk array's storage, provision a new disk array. This action can only be executed outside Workload Optimization Manager. ▪ Start For high utilization of disk array, start a suspended disk array. This action can only be executed outside Workload Optimization Manager. ▪ Suspend For low utilization of the disk array's storage, move VMs to other datastores and suspend volumes on the disk array. This action can only be executed outside Workload Optimization Manager.

Entity type	Supported actions
	▪ Move (Only for NetApp Cluster-Mode) For high utilization of Storage Controller resources, Workload Optimization Manager can move an aggregate to another storage controller. The storage controllers must be running. For high IOPS or latency, a move is always off of the current disk array. All the volumes on a given disk array show the same IOPS and Latency, so moving to a volume on the same array would not fix these issues. ▪ Move VM For high utilization of Storage on a volume, Workload Optimization Manager can move a VM to another volume. The new volume can be on the current disk array, on some other disk array, or on any other datastore. For high IOPS or latency, a move is always off of the current disk array. All the volumes on a given disk array show the same IOPS and Latency, so moving to a volume on the same array would not fix these issues. ▪ Move datastore To balance utilization of disk array resources, Workload Optimization Manager can move a datastore to another array.
Storage controller	Provision For high utilization of the storage controller's CPU, provision a new storage controller, and then move disk arrays to it.
IO module	Suspend For low utilization of net resources, move VMs to another host that provides compatible network connectivity and suspend this IO Module.
Switch	Resize Resize PortChannel for a switch to increase bandwidth.

Action categories

Workload Optimization Manager groups entries in the Actions List by different categories. These categories do not strictly define the severity of an issue, but they indicate the nature of the issue.

Compliance

A virtual environment can include policies that limit availability of resources. It's possible that the environment configuration violates these defined policies. In such cases, Workload Optimization Manager identifies the violation and recommends actions that bring the entity back into compliance.

Efficiency

Efficient utilization of resources is an important part of running in the desired state. Running efficiently maximizes your investment and reduces cost. When Workload Optimization Manager discovers underutilized resources, it recommends actions to consolidate your operations. For example, it can recommend that you move certain VMs onto a different host, and then suspend the original host.

Performance

Ultimately, the reason to manage workloads in your environment is to assure performance and meet productivity goals. Workload Optimization Manager can recommends actions when it detects conditions that directly put performance at risk. You can consider these critical conditions, and you should execute the recommended actions as soon as possible.

Prevention

Workload Optimization Manager constantly monitors conditions, and works to keep your environment running in a desired state. As it finds issues that risk moving the environment out of this state, it recommends preventive actions. You should attend to these issues, and perform the associated actions. If you do not, the environment may drift away from the desired state, and performance may be put at risk.

Action types

Workload Optimization Manager performs the following general types of actions:

- Delete – Remove unused entities (for example, datastores on disk arrays or unattached volumes)
- Move – Move an entity from one provider to another to address performance issues or improve infrastructure efficiency
- Optimization – Increase discount coverage for cloud workloads to reduce on-demand costs
- Provision – Add resource providers to the environment to increase capacity
- Reconfigure – Correct a misconfiguration to bring an entity into compliance
- Scale/Resize – Change the resources allocated to workloads to meet demand
- Start – Start a stopped/suspended entity
- Stop/Suspend – Pause an entity for a period of time to increase efficient use of resources

For information on the actions that Workload Optimization Manager generates for specific entities, see Actions by Entity Type *(on page 835)*.

Placement

Placement actions determine the best provider for a consumer. These include initial placement for a new entity, and move actions that change a consumer to use a different provider. For example, moving a VM assigns it to a different host. Moving a VM's storage means the VM will use a different datastore.

Placement constraints

When making placement decisions, Workload Optimization Manager checks for placement constraints to limit the set of providers for a given consumer. It respects automatic placement constraints, including cluster boundaries and DRS rules. It also considers user-configured constraints defined in a placement policy to ensure compliance to specific business requirements.

Reviewing the constraints on an entity helps you understand the actions that Workload Optimization Manager recommends. If an action seems questionable to you, then you should look at the constraints on the affected entities. It's possible that some policy or constraint is in effect, and it keeps Workload Optimization Manager from recommending a more obvious action. For details, see Entity Placement Constraints *(on page 168)*.

You can run plans to see what happens if you turn off constraints, or disable or enable certain placement policies.

Effective CPU capacity

CPU processor speed is not necessarily an effective indicator of CPU capacity. For example, processor architecture can make a slower CPU have a greater effective capacity. Newer models of machines can often have fewer cores or less clock speed, but still have a higher effective capacity.

When placing VMs on hosts in the on-prem environment, Workload Optimization Manager discovers the effective CPU capacity of your hosts. This increases the accuracy of placement calculations so that newer, more efficient hosts will show a greater effective capacity than less efficient hosts that might have larger or faster processors.

To discover the effective capacity, Workload Optimization Manager uses benchmark data from spec.org. This benchmark data maps to effective capacity settings that Workload Optimization Manager uses to make placement calculations.

You can see a catalog of these benchmark data and choose from listed processors when you edit Host templates. For more information, see Selecting CPUs from the Catalog *(on page 1103)*.

Shared-nothing migration actions

If you have enabled both storage and VM moves, Workload Optimization Manager can perform shared-nothing migrations, which move the VM and the stored VM files simultaneously. For details, see Shared-Nothing Migration *(on page 748)*.

Cross-vCenter vMotion

VMware vSphere 6.0 introduces functionality that enables migration of virtual machines between different vCenter Server instances. Workload Optimization Manager supports this capability through *Merge* placement policies (see Creating Placement Policies *(on page 1044)*). It considers cross-vCenter locations when calculating placement, and can recommend or execute moves to different vCenter servers.

Moves on the public cloud

On the public cloud you do not place workloads on physical hosts. In the Workload Optimization Manager Supply Chain, the Host nodes represent availability zones. Workload Optimization Manager can recommend moving a workload to a different zone, if such a move can reduce your cloud cost. These moves recognize constraints, such as availability of instances types and discounts *(on page 157)* in the given zones.

In AWS environments, a VM can use Elastic Block Stores (EBS) or Instance Storage. If the VM's root storage is EBS, then Workload Optimization Manager can recommend a VM move. However, because Instance Storage is ephemeral and a move would lose the stored data, Workload Optimization Manager does not recommend moving a VM that has Instance Storage as its root storage.

If a VM is running within a billing family, then Workload Optimization Manager only recommends moving that VM to other regions within that billing family.

In AWS environments that use RIs, Workload Optimization Manager recognizes Availability Zones that you have specified for your RI purchases. For move and resize actions, Workload Optimization Manager gives precedence to these RIs in the given zone. All else being equal for a given zone, if you have Zone RIs with reserved capacity and RIs that do not reserve capacity, Workload Optimization Manager will use the Zone RI first.

Scaling

Scaling actions update capacity in your environment. For vertical scaling, Workload Optimization Manager increases or decreases the capacity of resources on existing entities. For horizontal scaling it provisions new providers. For example, provisioning a host adds more compute capacity that is available to run VMs. Provisioning a VM adds capacity to run applications.

Workload Optimization Manager can provision the following:

- Containers
- VMs
- Hosts
- Storage
- Storage Controllers (only for planning scenarios)
- Disk Arrays

Under certain circumstances, Workload Optimization Manager can also recommend that you provision a virtual datacenter.

Discount optimization

To reduce your cloud costs, Workload Optimization Manager can recommend scaling VMs to instance types that are charged discounted rates.

NOTE:
Under most circumstances, when a cloud provider offers a new instance type that is meant to replace an older type, the provider offers it at a lower cost. However, a provider may provide a new instance type with identical costs as the older instance types. If this occurs, and capacity and cost are equal, Workload Optimization Manager cannot ensure that it chooses the newer instance type. To work around this issue, you can create an Action Automation policy that excludes the older instance type.

- Discount Utilization *(on page 1032)*

  This chart shows how well you have utilized your current discount inventory *(on page 1028)*. The desired goal is to maximize the utilization of your inventory and thus take full advantage of the discounted pricing offered by your cloud provider.
- Discount Coverage *(on page 1026)*

  This chart shows the percentage of cloud workloads (VMs and RDS database servers) covered by discounts. For VMs covered by discounts, you can reduce your costs by increasing coverage. To increase coverage, you scale VMs to instance types that have existing capacity.
- Discount Inventory *(on page 1028)*

  This chart lists the cloud provider discounts that you have purchased for your workloads.

Discount optimization actions are not executed by Workload Optimization Manager users. They reflect capacity reassignments performed by your cloud provider.

Configuration

These are reconfigure and resize actions. Reconfigure actions can add necessary network access, or reconfigure storage. Resize actions allocate more or less resource capacity on an entity, which can include adding or reducing VCPUs or VMem on a VM, adding or reducing capacity on a datastore, and adding or reducing volumes in a disk array.

Workload Optimization Manager can reconfigure the following:

- VMs
- Containers
- Storage
- Disk Arrays
- Virtual Datacenters

Start

Workload Optimization Manager can recommend that you start a suspended entity to add capacity to the environment.

Stop

Stop actions suspend entities without removing them from the environment. Suspended capacity is still available to be brought back online, but is currently not available for use. Suspended resources are candidates for termination.

Workload Optimization Manager can suspend the following:

- Application Components
- Container Pods
- Disk Arrays

- Hosts
- Storage (on-prem)
- Virtual datacenter

Delete

Workload Optimization Manager can recommend deleting unnecessary resources to improve infrastructure efficiency. In public cloud environments, delete actions also introduce cost savings. For example, Workload Optimization Manager might recommend that you delete wasted files to free up storage space, or delete unattached storage in your cloud environment.

In on-prem environments, you can sort Delete file actions based on file size or filter based on the file directory or the number of days a file has been unattached to understand the benefits of deleting unattached files. You are able to configure automation policies to automatically delete unattached files after a defined period of non-use, and can manually execute delete actions if automation is not desired.

Action acceptance modes

Action acceptance modes specify the degree of automation for the generated actions. For example, in some environments you might not want to automate resize down of VMs because that is a disruptive action. You would use action acceptance modes in a policy to set that business rule.

Workload Optimization Manager supports the following action acceptance modes:

- Recommend

  Recommend the action so that a user can execute it outside Workload Optimization Manager.

- Manual

  Recommend the action, and provide the option to execute that action through the Workload Optimization Manager user interface.

- Automated

  Execute the action automatically.

  For automated resize or move actions on the same entity, Workload Optimization Manager waits five minutes between each action to avoid failures associated with trying to execute all actions at once. Any action awaiting execution stays in queue. For example, if a VM has both vCPU and vMem resize actions, Workload Optimization Manager could resize vCPU first. After this resize completes, it waits five minutes before resizing vMem.

- Automated when approved

  Execute the action automatically only after they are externally approved. This option is available only after adding a Service Now target.

The Pending Actions charts only count actions in *Recommend* or *Manual* mode.

Automated actions appear in the following charts:

- **All Actions** chart on the **Home Page** and the On-prem Executive Dashboard
- **Accepted Actions** chart on the **Home Page**

Setting action acceptance modes

There are two ways to configure action acceptance modes:

- Change the action acceptance mode in a default policy *(on page 1050)*.
- Create an automation policy *(on page 1053)*, scope the policy to specific entities or groups, and then select the action acceptance mode for each action.

  Workload Optimization Manager allows you to create dynamic groups to ensure that entities discovered in the future automatically add to a group and apply the policy of that group. If a conflict arises as a result of an entity belonging to several groups, the entity applies the policy with the most conservative action.

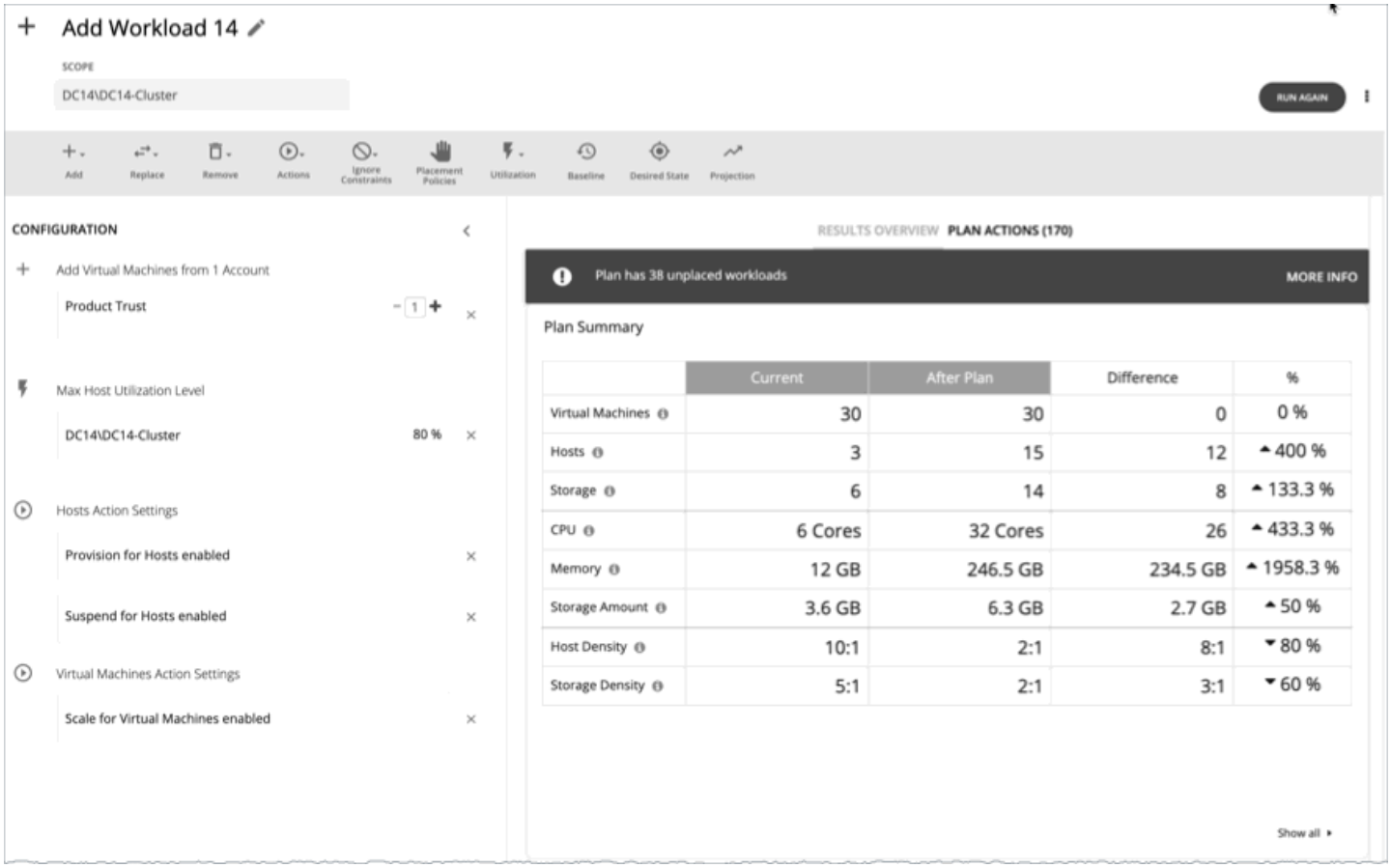
Action acceptance overrides

Under some conditions, Workload Optimization Manager changes the action acceptance mode of an action from *Manual* to *Recommend*.

Workload Optimization Manager makes this change as a safeguard against executing actions that the underlying infrastructure cannot support. For example, assume you have VM move actions set to *Manual*. Then assume Workload Optimization Manager analysis wants to move a VM onto a host that is already utilized fully. In this case, there would be other actions to move workloads *off* of the given host to make room for this new VM. However, because moves are *Manual*, the host might not be properly cleared off yet. In that case, Workload Optimization Manager changes actions to move workloads *to* the host from *Manual* to *Recommend*.

For cloud environments, some instances require workloads to be configured in specific ways before they can move to those instance types. If Workload Optimization Manager recommends moving a workload that is not suitably configured onto one of these instances, then it changes the action acceptance mode from *Manual* to *Recommend*, and then describes the reason.

Plans: looking to the future



Use the Plan Page to run simulations for what-if scenarios that explore possibilities such as:

- Reducing cost while assuring performance for your workloads
- Impact of scaling resources
- Changing hardware supply
- Projected infrastructure requirements

- Optimal workload distribution to meet historical peaks demands
- Optimal workload distribution across existing resources

How Plans work

To run a plan scenario, Workload Optimization Manager creates a snapshot copy of your real-time market and modifies that snapshot according to the scenario. It then uses the Economic Scheduling Engine to perform analysis on that plan market. A scenario can modify the snapshot market by changing the workload, adding or removing hardware resources, or eliminating constraints such as cluster boundaries or placement policies.

As it runs a plan, Workload Optimization Manager continuously analyzes the plan market until it arrives at the optimal conditions that market can achieve. When it reaches that point, the Economic Scheduling Engine cannot find better prices for any of the resources demanded by the workload — the plan stops running, and it displays the results as the plan's desired state. The display includes the resulting workload distribution across hosts and datastores, as well as a list of actions the plan executed to achieve the desired result.

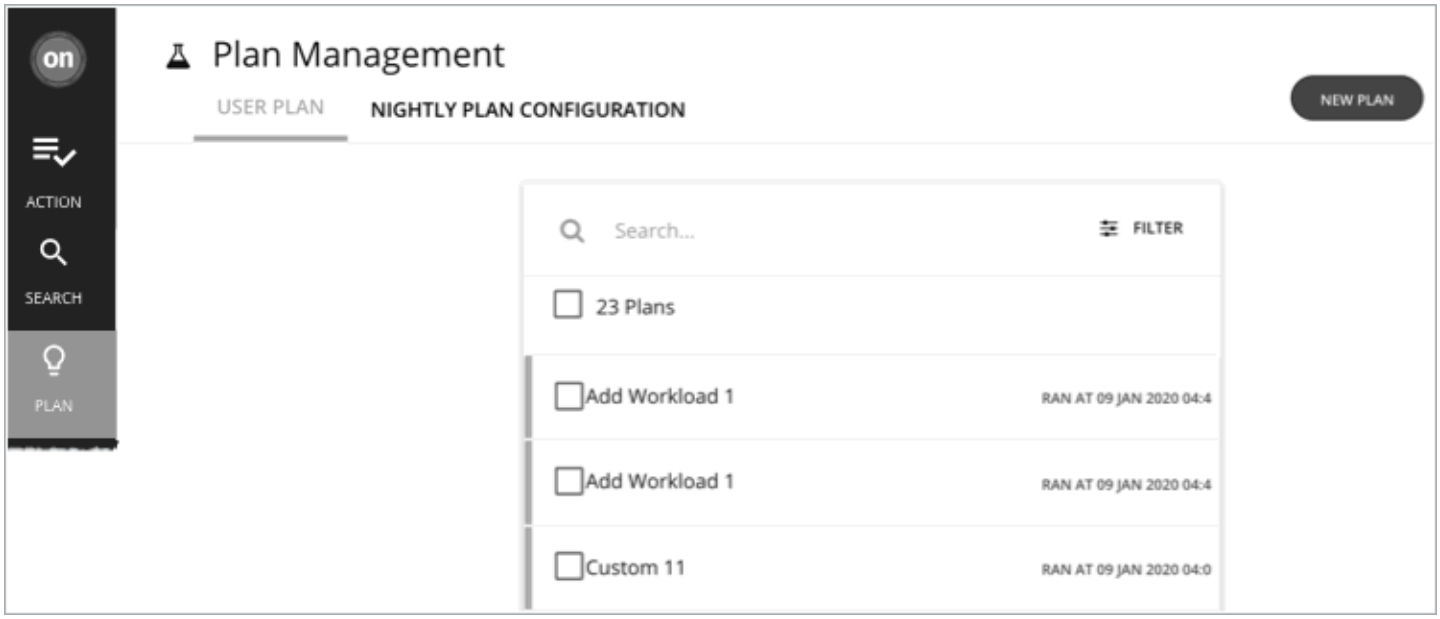
For example, assume a scenario that adds virtual machines to a cluster. To run the plan, Workload Optimization Manager takes a snapshot of the current market, and adds the VMs to the specified cluster. Workload Optimization Manager then runs analysis on the plan market, where each entity in the supply chain shops for the resources it needs, always looking for a better price — looking for those resources from less-utilized suppliers. This analysis continues until all the resources are provided at the best possible price.

The results might show that you can add more workload to your environment, even if you reduce compute resources by suspending physical machines. The recommended actions would then indicate which hosts you can take offline, and how to distribute your virtual machines among the remaining hosts.

Idle workloads

Plans calculate optimal placement and optimal resource allocation for the given workload. However, plans do not include *idle* workloads. This is because an idle VM shows no utilization, so the plan cannot determine optimal placement or what percentage of allocated resources that workload will require when it restarts.

Plan management



The Plan Management Page is your starting point for creating new plans, viewing saved plans, and deleting saved plans that you don't need anymore. To display this page, click **Plan** in the Workload Optimization Manager navigation bar.

- Create new plans

  To create a new plan, click the **NEW PLAN** button. See Setting Up Plan Scenarios *(on page 851)*.

- View saved plans

  After you create and run a plan, Workload Optimization Manager saves it and then shows it in the Plan Management Page. You can open the saved plan to review the results, or you can change its configuration and run it again.

  NOTE:
  You can also view saved plans from the Search page, under the **Plans** category.

- Delete saved plans

  To delete a saved plan, turn on the plan's checkbox and then click the **Delete** button.

- Configure nightly plans

  Workload Optimization Manager runs nightly plans to calculate headroom for the clusters in your on-prem environment. For each cluster plan, you can set which VM template to use in these calculations. See Configuring Nightly Plans *(on page 941)*.

Setting Up plan scenarios

A plan scenario specifies the overall configuration of a plan. Creating the plan scenario is how you set up a what-if scenario to see the results you get if you changed your environment in some way.

Plan entry points

You can begin creating a plan scenario from different places in the user interface:

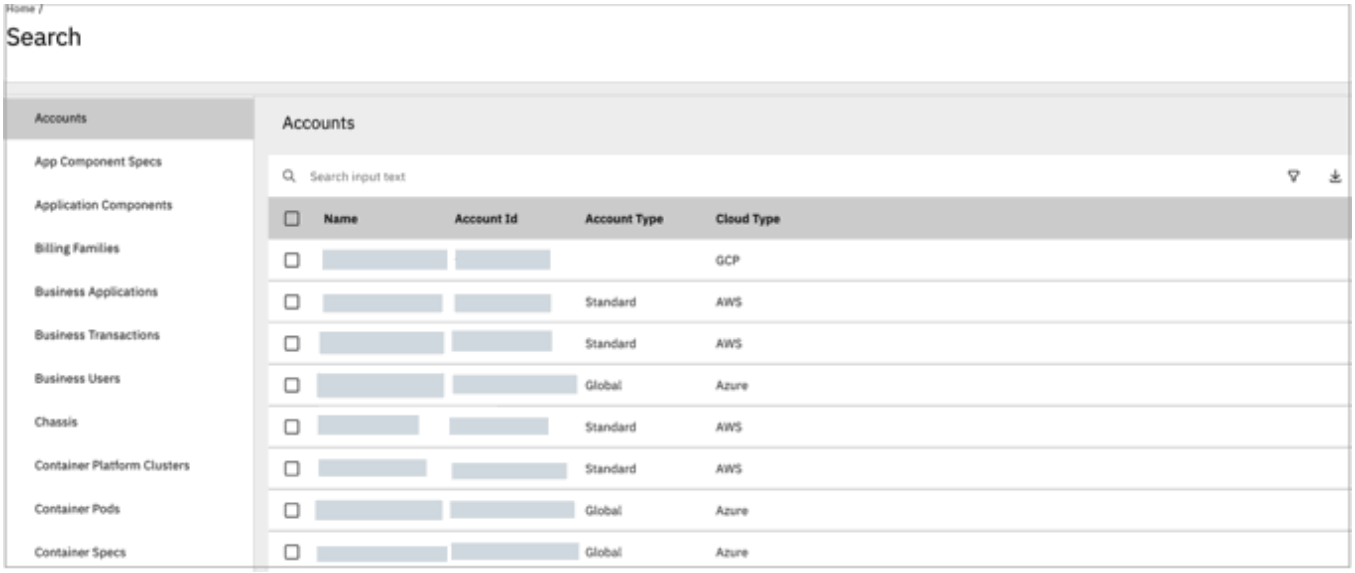
- From the Plan Page

  Navigate to the Plan Page and click **New plan**. This plan has no scope. You will specify the scope after selecting the plan type.



- From the **Home Page**

  To start a plan scenario from the **Home Page**, you must first go to the **Search** page to set the scope.



Set the scope to *a specific* Account, Billing Family, VM Group, or Region to start an Optimize Cloud plan.

- Cloud scope

  If you set the scope to *a specific* Account, Billing Family, VM Group, or Region, you can start an Optimize Cloud or Buy VM Reservations plan.

- On-prem scope

  If you set the scope to *a specific* Host Cluster, Datacenter, Group, Storage Cluster, or Virtual Datacenter, you can start any plan. You may need to go through additional steps, depending on your chosen plan type. For example, if you scope to a host cluster and choose the Add Virtual Machines plan type, the plan wizard prompts you to select the most suitable templates for the VMs you plan to add to the cluster.

- Container platform cluster scope

  If you set the scope to *a specific* container platform cluster, you can start an Optimize Container Cluster or Migrate Container Workloads plan.

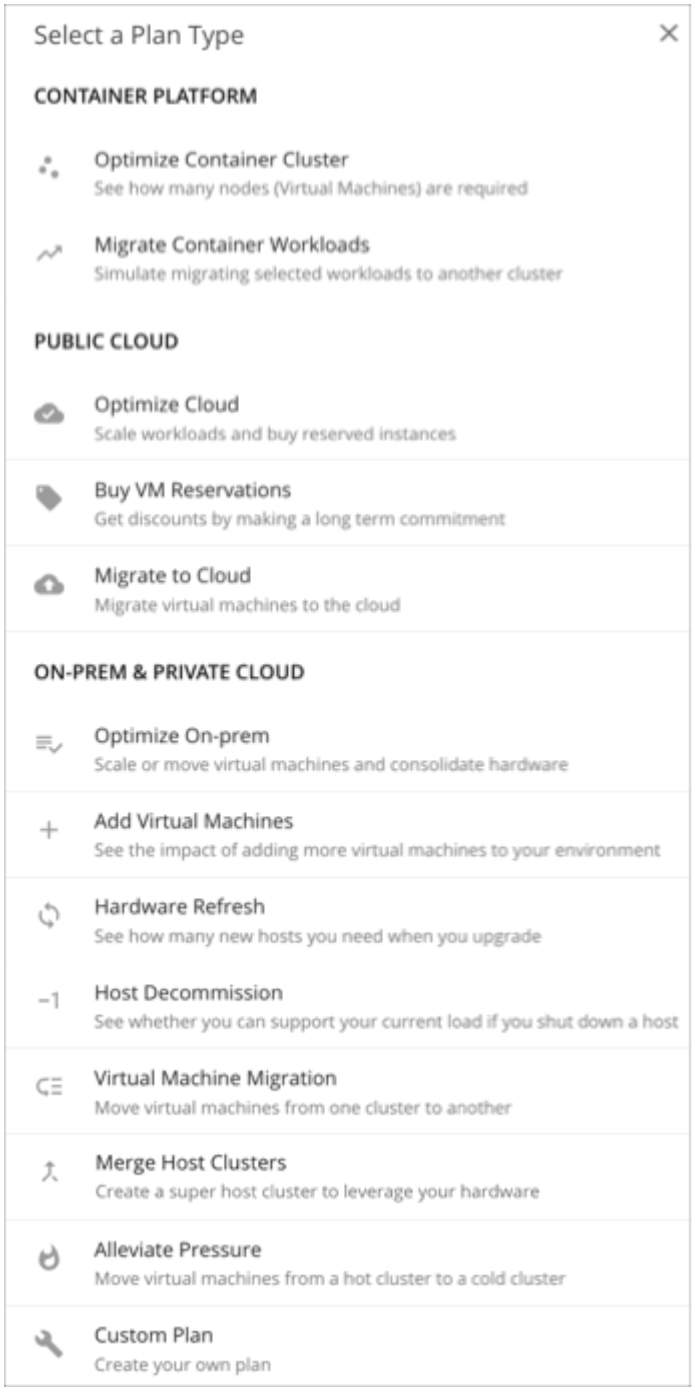
For details, see Scoping the Workload Optimization Manager Session *(on page 161)*.

After setting the scope, the **Plan** button appears in the **Home Page**.



Plan types

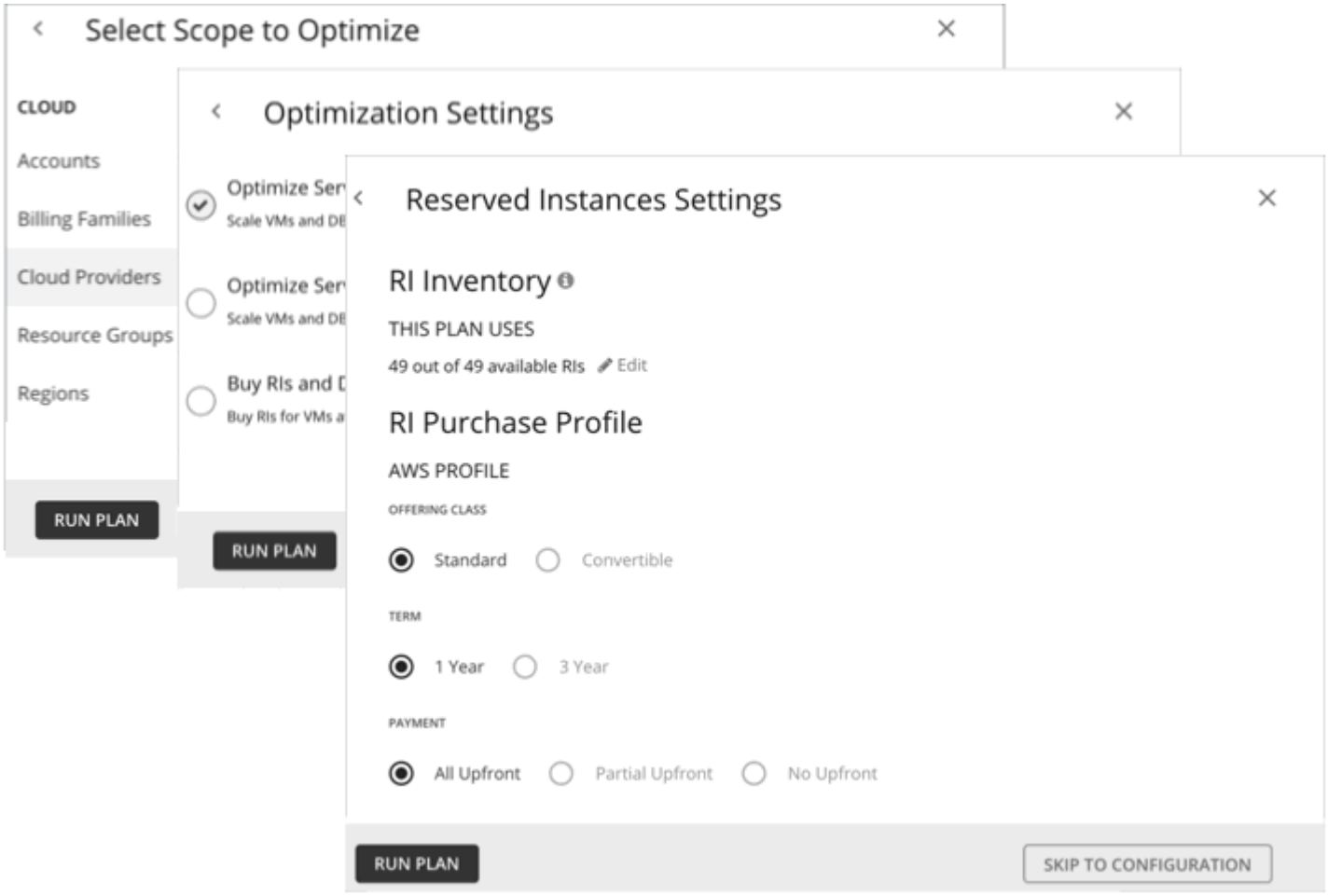
Select from the list of plan types. For more information, see Plan scenarios and types *(on page 857)*.



Workload Optimization Manager opens the appropriate plan wizard.

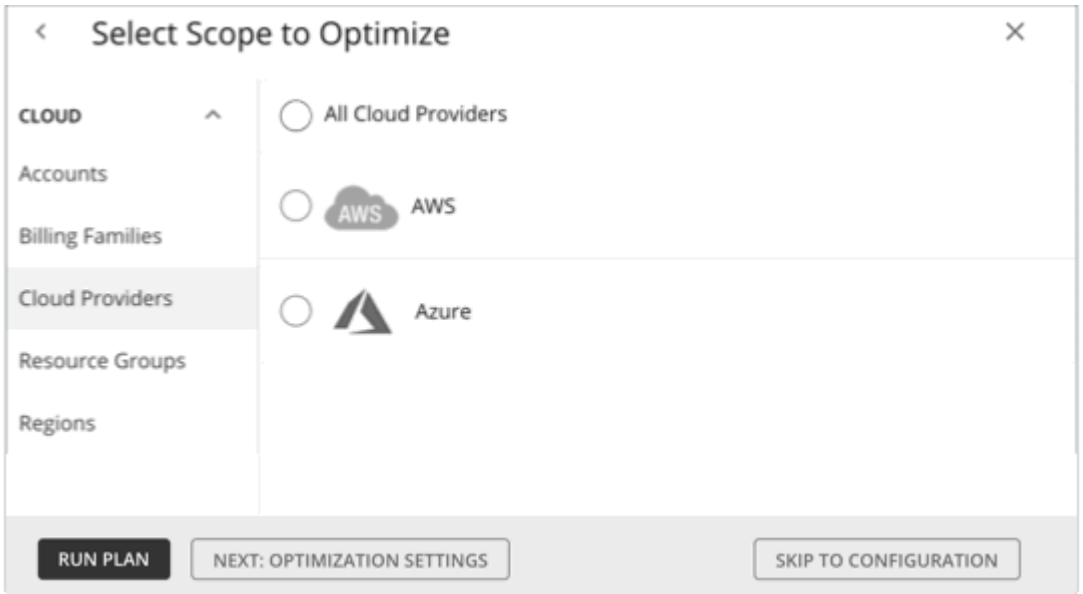
Plan wizards

Each plan type includes a wizard to guide you through creating the scenario. The wizard leads you through the configuration steps to create a plan that answers a specific question. After you make the required settings, you can skip ahead and run the plan, or continue through all the optional steps.



Plan scope

All plans require a scope. For example, to configure an Optimize Cloud plan, you set the scope to all or specific cloud providers or accounts.



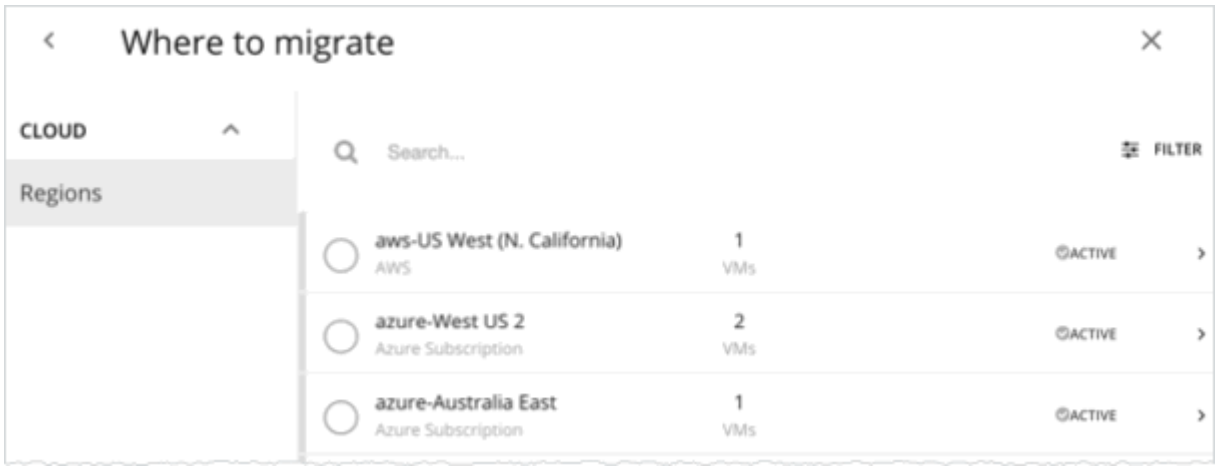
It usually helps to focus on a subset of your environment. For a large environment, plans that are scoped run faster.

To narrow the scope, select a group from the list. The page then refreshes to include only the entities that belong to that group.

Use **Search** or **Filter** to sort through a long list.

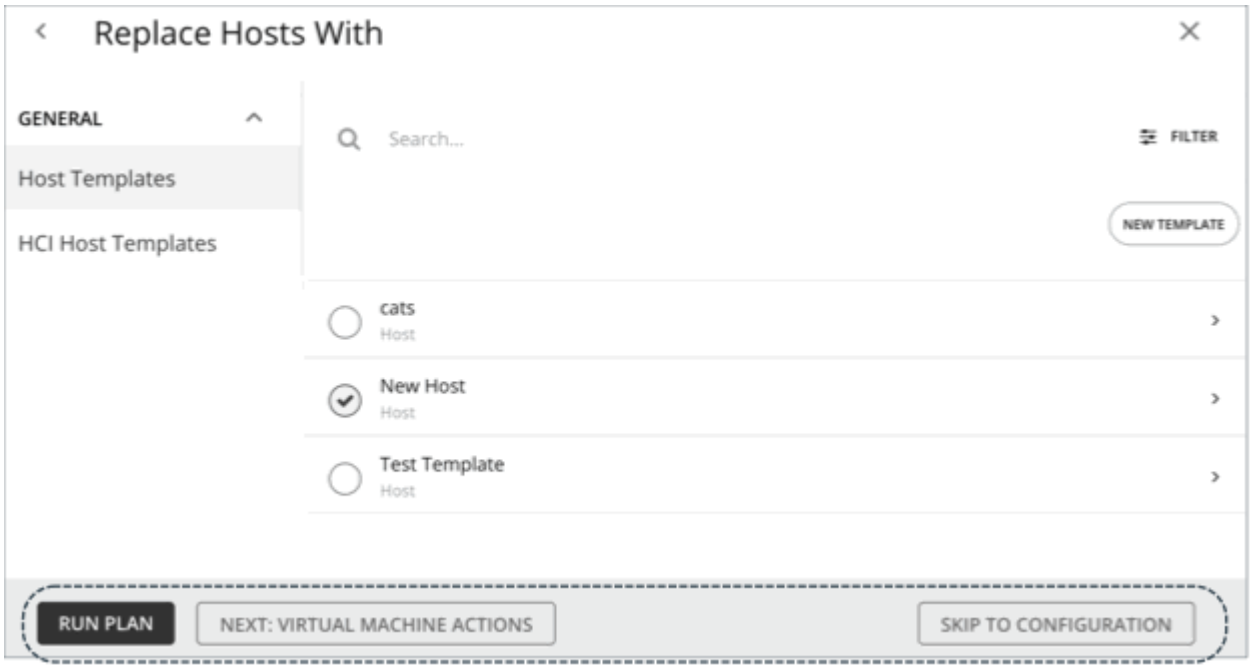
Other plan information

The wizard prompts you for any additional information that is required to run the plan. For example, for a Hardware Refresh plan, you need to identify the hosts that replace the scoped hosts. For a Migrate to Cloud plan, you need to identify the cloud service provider, region, or group you want the scoped workloads to migrate to.



Run the plan

After you provide the minimum required information for running a plan, the wizard shows you the following options:



- **Run Plan**: Immediately run the plan.
- **Next: [Step]**: Continue with the rest of the wizard and then run the plan.

- **Skip to Configuration**: Skip the rest of the wizard and go to the Plan Page to:
  - Customize the plan settings.
  - See a preview of the plan scenario.
  - Run the plan.

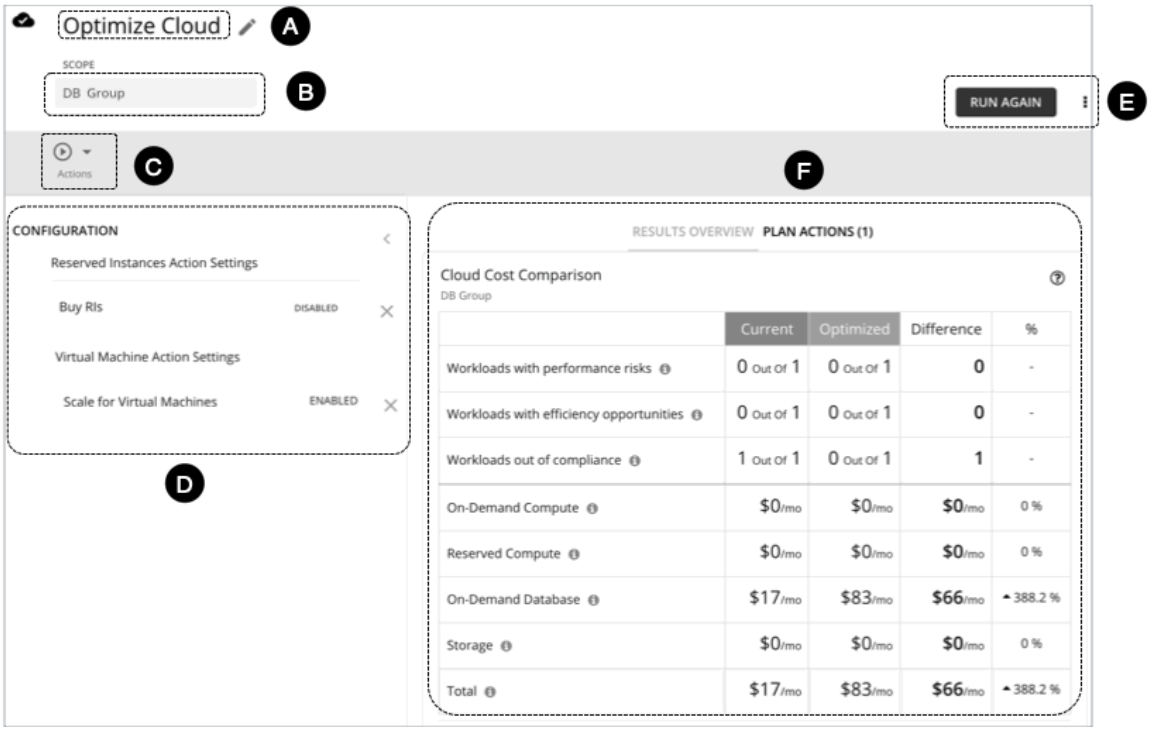
NOTE:
For a custom plan, the only option available is **Configure Plan**. Click this button to open the Plan Page, configure the plan settings, and then run the plan.

The Plan page

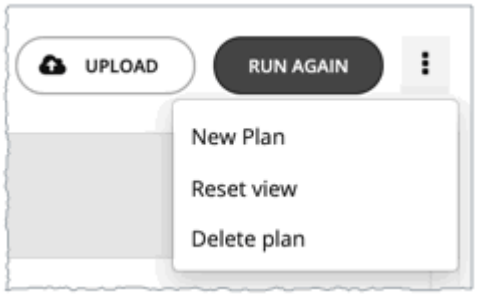
The Plan Page first displays if you skip the wizard or when you run a plan.

For a plan with a large scope, it might take some time before you see the results. You can navigate away from the Plan Page and check the status in the Plan Management Page. You can also cancel a plan that is in progress.

The Plan Page shows the following sections:



Plan page sections	Description
A. Plan name	Workload Optimization Manager automatically generates a name when you create a new plan. Change the name to something that helps you recognize the purpose of this plan.
B. Plan scope	Review the scope that you set in a previous step. **NOTE:** It is not possible to change the scope of the plan in the Plan Page. You must start over if you want a different scope. To start over, click the More options icon (⋮), and then select **New Plan**.

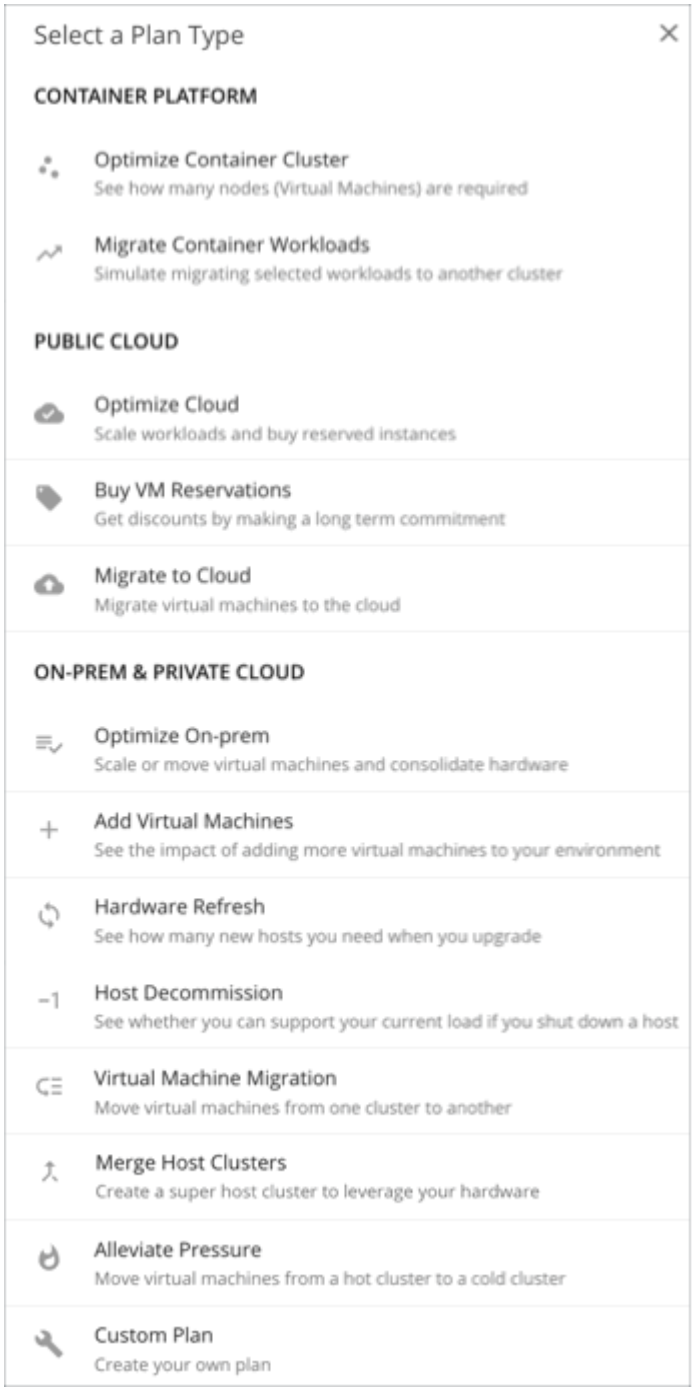
Plan page sections	Description
C. Configuration toolbar	Configure other settings for the plan. You can name the plan, change workload demand and the supply of resources, and specify other changes to the plan market. The toolbar items that display depend on the plan you are creating.
D. Configuration summary	Review the plan's configuration settings. You can remove any setting by clicking the **x** mark on the right. Use the toolbar to change the settings. As you edit the plan scenario, those changes immediately appear in the Configuration summary.
E. Other options	See what else you can do with the plan. ▪ **Upload**: (For Azure only) Upload the results of a Migrate to Cloud plan to the Azure Migrate portal. For details, see Uploading the Results to Azure *(on page 912)*. ▪ **Run** / **Run Again**: – If a plan has not run, click **Run** and then check the plan results. – If the plan ran and you want to run it again with a different set of configuration settings, click **Run Again**. This runs the plan scenario against the market in its current state. ▪ ⋮ : Click to see more options.  – **New Plan**: Configure a new plan. You can choose this option if you want to change the scope of the current plan, which requires that you start over and configure a new plan. – **Reset view**: Restore charts to their default views. For example, if you changed the commodities that are displayed in the Optimized Improvements or Comparison charts, you can discard those changes by choosing this option. – **Delete plan**: Choose if you no longer need the plan.
F. Plan results	Review the results in the charts provided. For a plan that has not run, you see a **Scope Preview** chart and a one-time message that instructs you to run the plan.

Plan Management

Created plans are displayed in the Plan Management Page *(on page 850)*.

Plan scenarios and types

To simulate different plan scenarios, Workload Optimization Manager provides the following general types of plans:



Optimize container cluster

Run an Optimize Container Cluster plan to identify performance and efficiency opportunities for a single cluster. The results show the optimal number of nodes you need to assure performance for your existing workloads, and the impact of actions on the health of your container workloads and infrastructure.

Migrate container workloads

Run a Migrate Container Workloads plan to simulate the migration of container workloads from one cluster to another. The plan compares results from a 'lift-and-shift only' scenario against a Workload Optimization Manager optimized plan. The results further highlight the actions you need to take to maintain and optimize workload performance in the new cluster.

Optimize cloud

For the scope of your public cloud environment that you want to examine, run a plan to see all the opportunities you have to reduce cost while assuring performance for your workloads. This includes comparisons of template and storage usage, and a comparison of current to optimized cost.

Buy VM reservations

Run the Buy VM Reservations plan to see the most cost-effective discount *(on page 157)* purchases that will continue to assure performance for your cloud VMs.

Migrate to cloud

A Migrate to Cloud plan simulates migration of on-prem VMs to the cloud, or migration of VMs from one cloud provider to another.

> **NOTE:**
> For migrations within your on-prem environment, use the *Virtual Machine Migration* plan type.

Optimize On-prem

See the effects of executing certain actions, such as scaling virtual machines, suspending hosts, or provisioning storage, to your on-prem environment.

Add virtual machines

Adding virtual machines increases the demand that you place on your environment's infrastructure. You can set up a plan to add individual VMs or groups of VMs in your environment, or based on templates.

Hardware refresh

Choose hosts that you want to replace with different hardware. For example, assume you are planning to upgrade the hosts in a cluster. How many do you need to deploy, and still assure performance of your applications? Create templates to represent the upgraded hosts and let the plan figure out how many hosts you really need.

To increase the accuracy of the plan results, Workload Optimization Manager analysis considers a cluster's overall resource utilization over the last ten days. The platform identifies the day within those ten days when percentile utilization for the cluster reached 90%, and then uses each VM's actual utilization data *on that day* to perform its analysis.

> **NOTE:**
> If you configure a Hardware Refresh plan to use a baseline snapshot, the plan will use that snapshot's data instead of the cluster's percentile data.

Host decommission

If your environment includes underutilized hardware, you can use a plan to see whether you can decommission hosts without affecting the workloads that depend on them.

Virtual machine migration

Use this plan type to simulate workload migrations within your on-prem environment.

You can see whether you have enough resources to move your workload from its current provider group to another. For example, assume you want to decommission one datacenter and move all its workload to a different datacenter. Does the target

datacenter have enough physical resources to support the workload move? Where should that workload be placed? How can you calculate the effect such a change would have on your overall infrastructure?

To calculate this information, create a plan that:

- Limits the plan scope to two data centers (or clusters) — the one you will decommission, and the one that will take on the extra workload
- Removes all the hardware from the decommissioned datacenter
- Calculates workload placement across datacenter (or cluster) boundaries
- Does not provision new hardware to support the workload

Merge host clusters

See the effects of merging two or more host clusters. For example, you can see if merging the host clusters would require provisioning additional storage to support current demand, or if ignoring cluster boundaries would improve performance and efficiency.

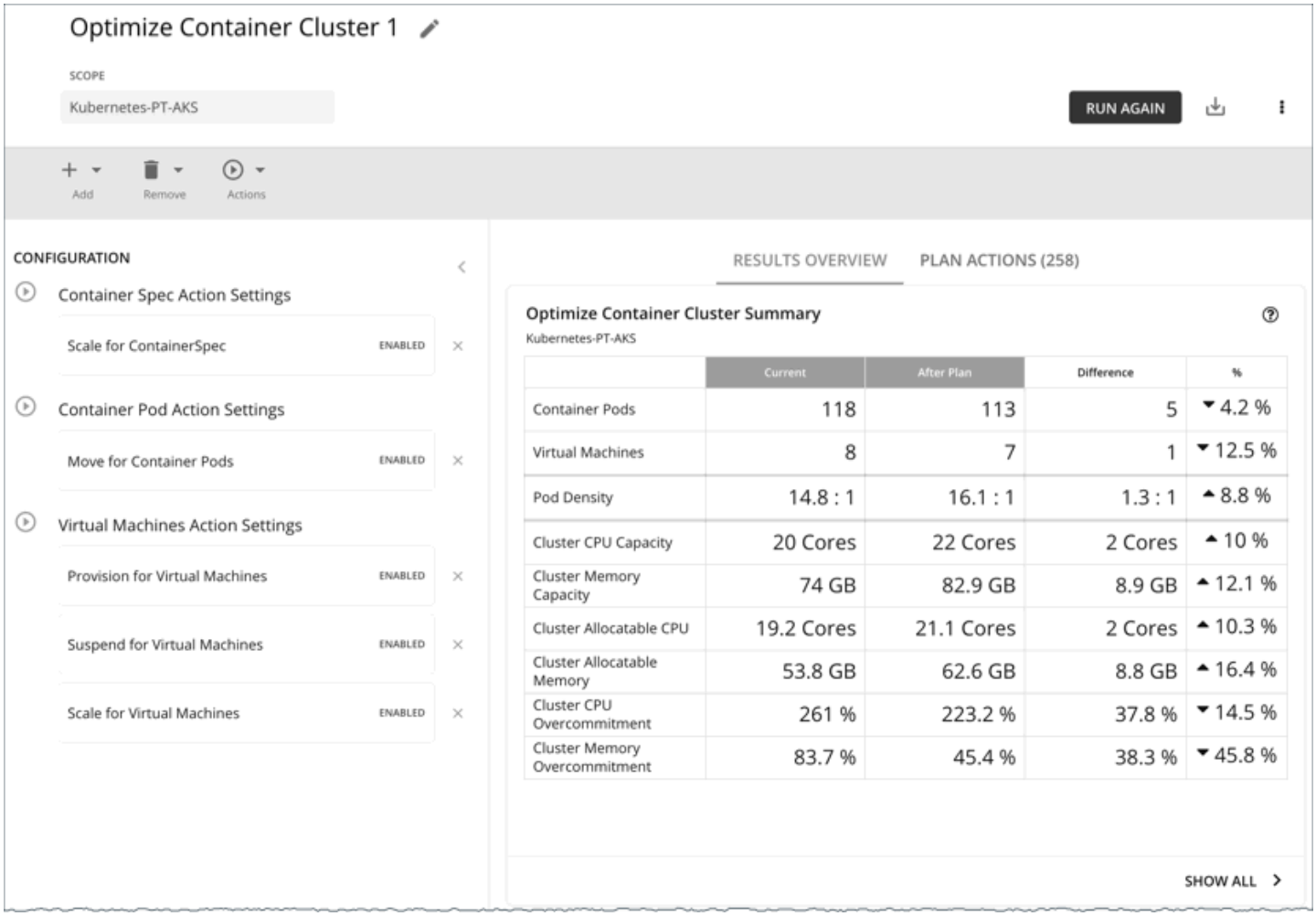
Alleviate pressure

Choose a cluster that shows bottlenecks or other risks to performance, and check to see the minimal changes you can make by migrating some workloads to another cluster. The cluster that is showing risks is a *hot* cluster, and the cluster you will migrate to is a *cold* cluster.

Custom plan

With a custom plan, you skip directly to the plan configuration after specifying the plan scope, and set up whatever type of scenario you want.

Optimize container cluster plan



Run an Optimize Container Cluster plan to identify performance and efficiency opportunities for a single container platform cluster. The results show the optimal number of nodes you need to assure performance for your existing workloads, and the impact of actions on the health of your container workloads and infrastructure. For example, you can see how container resize actions change the limits and requests allocated per namespace, or how node provision/suspend actions impact allocatable capacity for the cluster. For a cluster in the public cloud, the results also include the cost impact of node actions.

You can scope the plan to a:

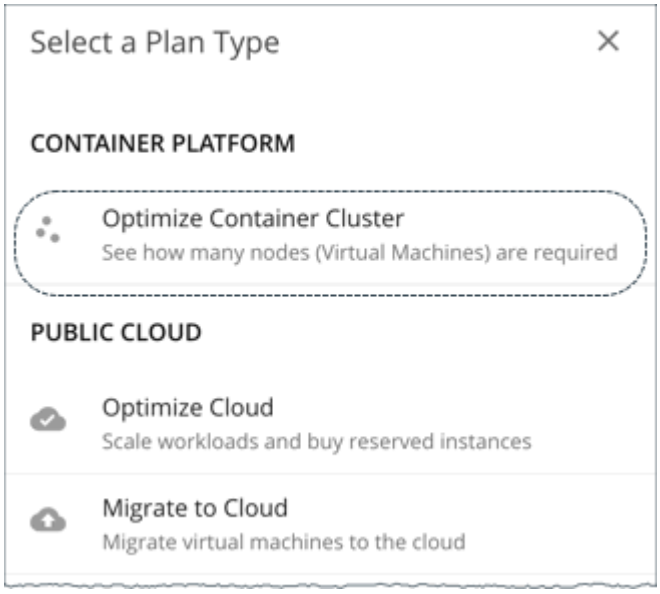
- Standalone cluster
- Cluster in an on-prem or public cloud environment
- Cluster stitched to applications via Data Ingestion Framework *(on page 522)* (DIF)

Scoping to a group within a cluster (such as a group of nodes) is currently not supported.

Configuring an optimize container cluster plan

You can start an Optimize Container Cluster plan when you open the Plan page or set the scope to a container platform cluster.

For an overview of setting up plan scenarios, see Setting Up Plan Scenarios *(on page 851)*.



Scope

Select a container platform cluster to optimize.

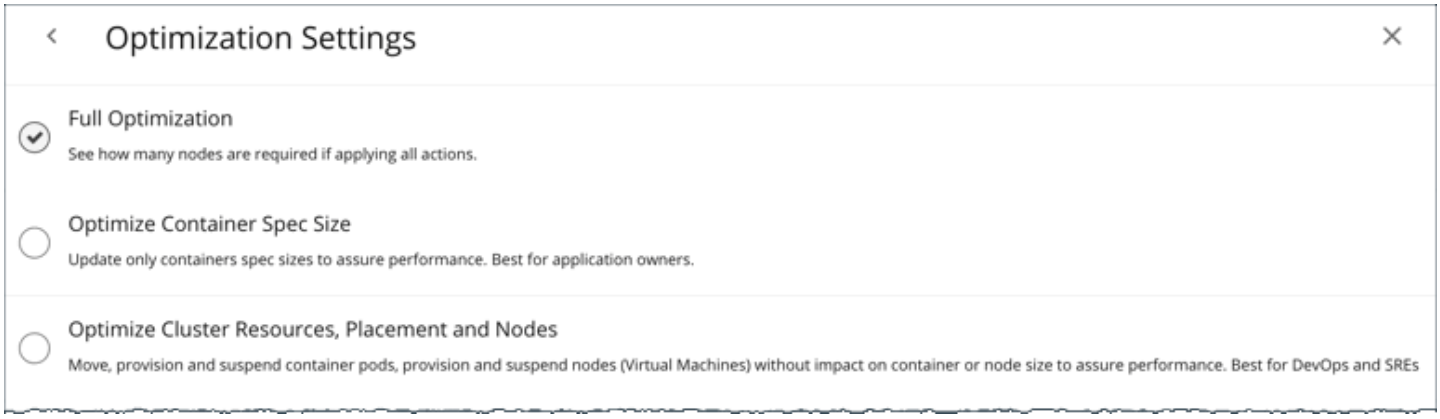


Scoping to a group within a cluster (such as a group of nodes) is currently not supported.

> **NOTE:**
> After selecting a cluster, you can skip the next step (**Optimization Settings**) and run the plan. Workload Optimization Manager runs the **Full Optimization** scenario in this case.

Optimization settings

Choose from the given optimization scenarios.



- Full optimization

  Workload Optimization Manager will recommend all relevant actions to optimize the cluster. For example, it can recommend provisioning nodes or resizing containers to meet application demand, or moving pods from one node to another to reduce congestion.

  Workload Optimization Manager can recommend the following actions:

  - Resize namespace compute resource quotas
  - Resize container limits and requests
  - Move pods
  - Provision or suspend nodes
  - Scale volumes

  NOTE:
  For a cluster in the public cloud, Workload Optimization Manager shows the cost impact of actions on nodes and volumes, to help you track your cloud spend. Workload Optimization Manager only reports the costs attached to these actions, and does *not* perform cost analysis on the cluster.

  For a cluster in an on-prem environment, Workload Optimization Manager can also recommend the following actions:

  - Move VMs
  - Provision or suspend hosts
  - Provision or suspend storage

- Optimize container spec size

  Workload Optimization Manager will only recommend resizing container limits and requests. This is ideal for application owners who manage the containers that their applications run on, but not the underlying container infrastructure.

- Optimize cluster resources, placement, and nodes

  Workload Optimization Manager will recommend *all* relevant actions, *except* resizing container limits and requests. This is ideal for teams who oversee the health of your container infrastructure, and want to evaluate the impact of *not* rightsizing workloads.

After selecting an optimization scenario, you can:

- Run the plan.

  Or

- Choose **Skip to Configuration** to configure additional settings. See the next section for details.

(Optional) Other plan settings

You can fine tune your selected optimization scenario or include additional scenarios before you run the plan.

- Enable or disable actions

  Fine tune your optimization scenario by enabling or disabling actions for containers, pods, or nodes. For example, you may have selected **Full Optimization**, but only for containers, nodes, and pods that are allowed to move. In this case, you would disable move actions for the pods that should never move.

  For clusters in on-prem environments, you can also enable or disable actions for hosts and storage.

  IMPORTANT:
  To avoid seeing inaccurate plan results, do *not* disable all actions.

- Add pods

  See resource changes if you add more pods to the cluster. For example, you might need to provision nodes to accommodate the new pods.

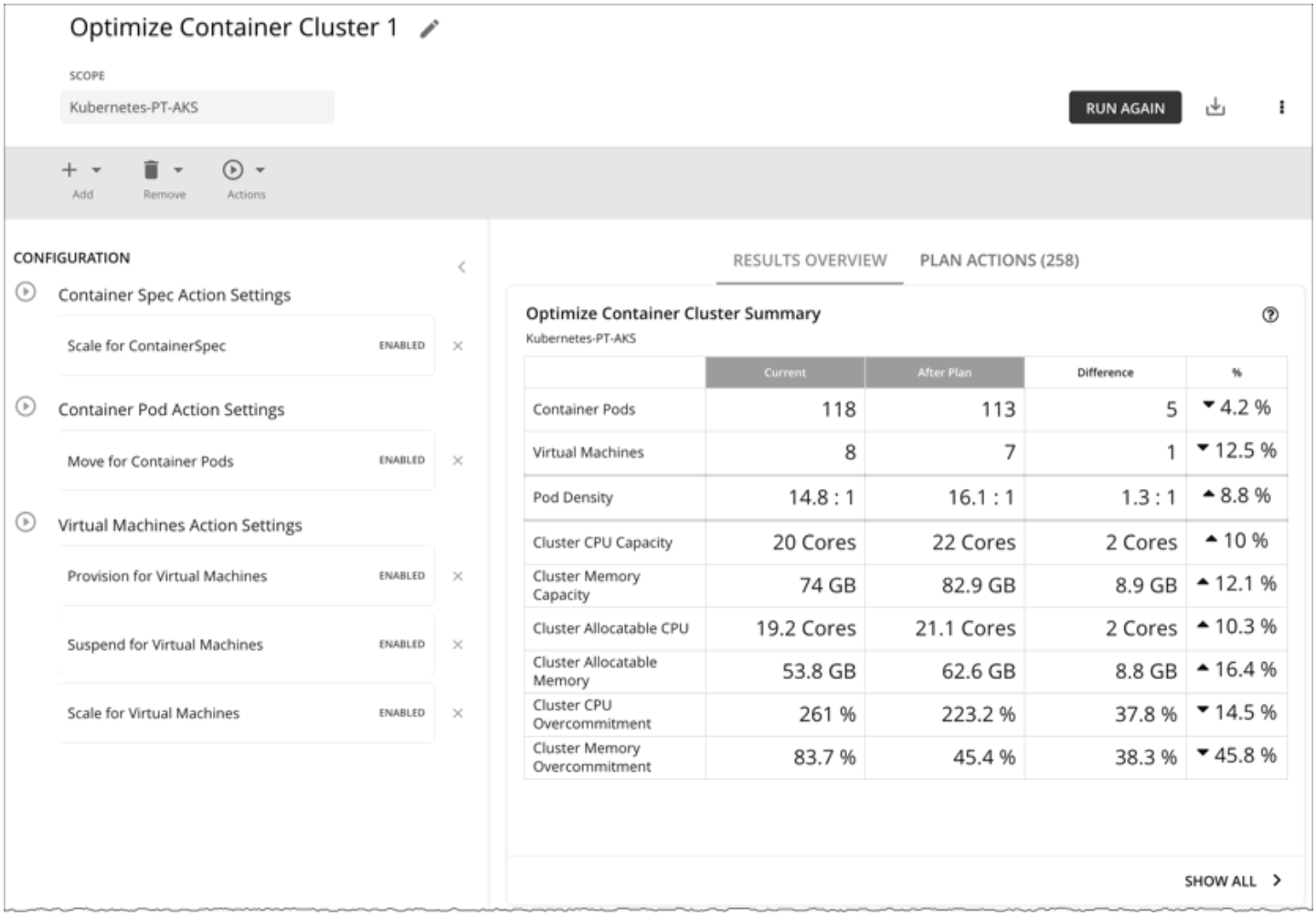
  Select an existing pod within or outside the selected cluster, and then specify how many copies to add. The plan simulates adding pods with the same resources as the selected pod.

- Remove pods or nodes

  See the effect of removing pods or nodes from the cluster. For example, pod density could improve significantly if you remove pods that you no longer need, or certain pods might become unplaced if you remove nodes.

Working with optimize container cluster plan results

After the plan runs, you can view the results to see how the plan settings you configured affect your environment.

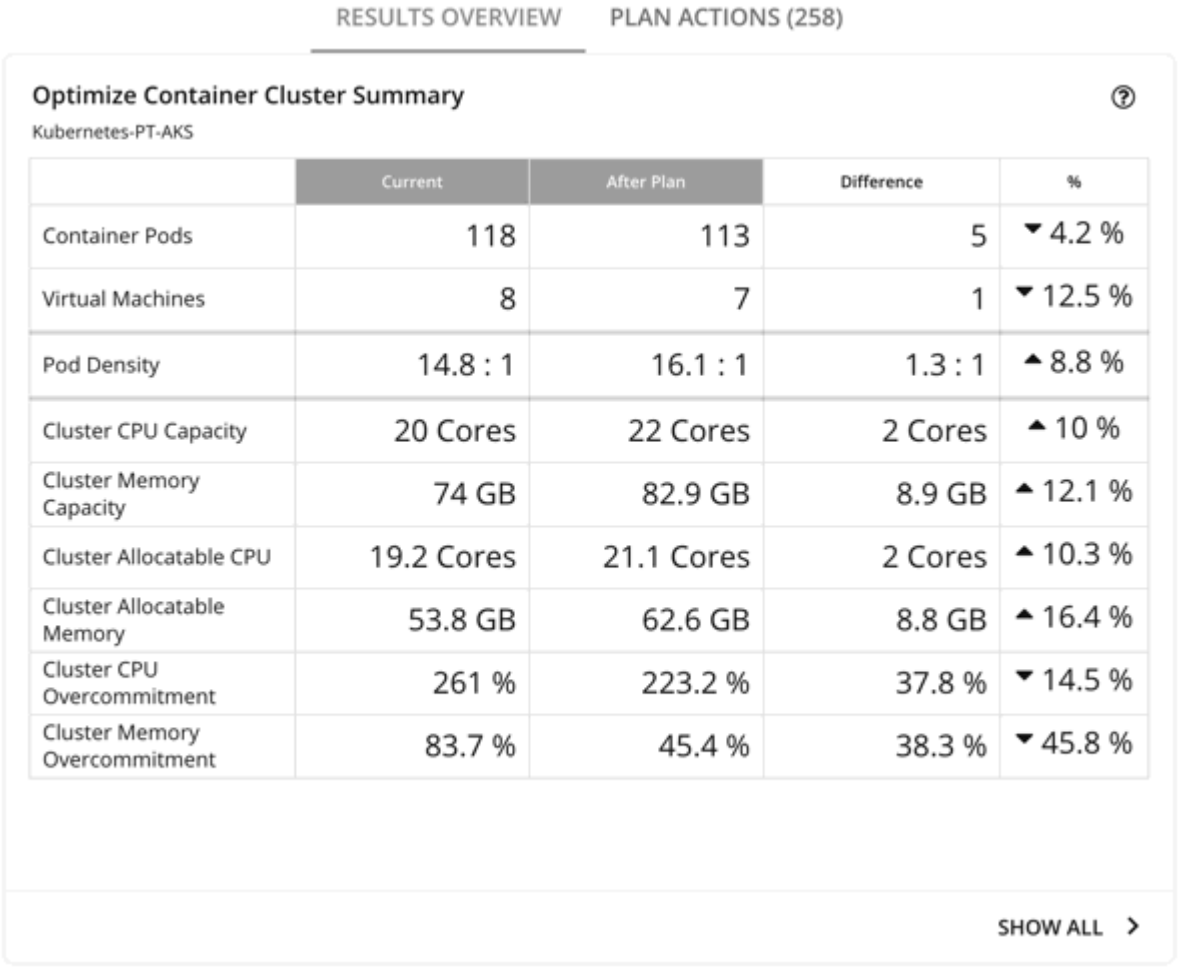


General guidelines

Familiarize yourself with these common terms that appear in the plan results:

- A container pod represents the compute demand from a running pod.
- A node (virtualized or bare metal) is represented as a VM.
- *Used* (or *Usage*) values represent actual resource consumption. For example, a node that consumes 100 MB of memory has a used value of 100 MB.
- *Utilization* values represent used/usage values against capacity. For example, a node that consumes 100 MB of memory against a total capacity of 500 MB has a utilization value of 20%.

Optimize container cluster summary



	Current	After Plan	Difference	%
Container Pods	118	113	5	▼4.2 %
Virtual Machines	8	7	1	▼12.5 %
Pod Density	14.8 : 1	16.1 : 1	1.3 : 1	▲8.8 %
Cluster CPU Capacity	20 Cores	22 Cores	2 Cores	▲10 %
Cluster Memory Capacity	74 GB	82.9 GB	8.9 GB	▲12.1 %
Cluster Allocatable CPU	19.2 Cores	21.1 Cores	2 Cores	▲10.3 %
Cluster Allocatable Memory	53.8 GB	62.6 GB	8.8 GB	▲16.4 %
Cluster CPU Overcommitment	261 %	223.2 %	37.8 %	▼14.5 %
Cluster Memory Overcommitment	83.7 %	45.4 %	38.3 %	▼45.8 %

This chart shows how your container environment and the underlying resources will change after you execute the actions that the plan recommends. The chart shows the following information:

- Container pods

  Count of active container pods in the plan.

- Virtual machines

  Count of active nodes in the plan. This chart does not count "non-participating" entities in the real-time market, such as suspended nodes.

- Pod density

  Average number of pods per node.

  For the total number of pods against the node capacity (maximum pods per node), see the **Number of Consumers** data in the following charts:

  - Nodes (VMs) Optimized Improvements
  - Nodes (VMs) Comparison
  - Container Cluster Optimized Improvements
  - Container Cluster Comparison

- Cluster CPU capacity

  Total CPU capacity for the cluster. The 'After Plan' result indicates how much CPU capacity will result in the optimal number of nodes required to run workloads.

- Cluster memory capacity

  Total memory capacity for the cluster. The 'After Plan' result indicates how much memory capacity will result in the optimal number of nodes required to run workloads.

- Cluster allocatable CPU

  Total amount of cluster CPU available for pod requests. The 'After Plan' result indicates how much of the allocatable CPU capacity will change if you provision or suspend nodes.

- Cluster allocatable memory

  Total amount of cluster memory available for pod requests. The 'After Plan' result indicates how much of the allocatable memory capacity will change if you provision or suspend nodes.

- Cluster CPU overcommitment

  (Only for containers with CPU limits) This indicates whether the CPU limits exceed the capacity of the underlying nodes. A value greater than 100% indicates overcommitment. Workload Optimization Manager manages cluster resources by actual utilization and limit rightsizing so that you can run more workloads with less risk.

  Workload Optimization Manager only calculates overcommitment in plans. The calculation can be expressed as:

  ```
  Overcommitment = Sum of CPU limits for all containers / Sum of CPU capacity for all nodes
  ```

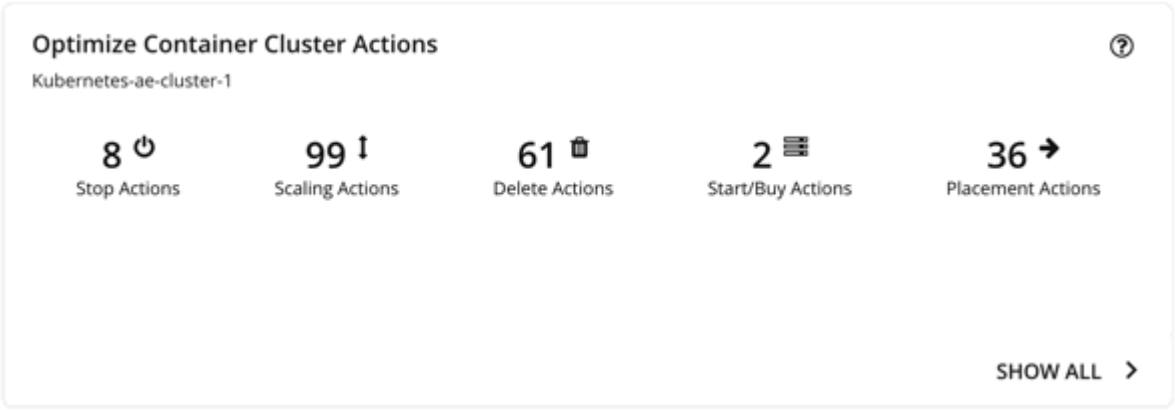
- Cluster memory overcommitment

  (Only for containers with memory limits) This indicates whether the memory limits exceed the capacity of the underlying nodes. A value greater than 100% indicates overcommitment. Workload Optimization Manager manages cluster resources by actual utilization and limit rightsizing so that you can run more workloads with less risk.

  Workload Optimization Manager only calculates overcommitment in plans. The calculation can be expressed as:

  ```
  Overcommitment = Sum of memory limits for all containers / Sum of memory capacity for all nodes
  ```

Optimize container cluster actions



This chart summarizes the actions that you need to execute to achieve the plan results. For example, you might need to resize limits and requests for containers (via the associated Workload Controllers) to address performance issues. Or, you might need to move pods from one node to another to reduce congestion.

Smarter redistribution and workload rightsizing also drive cluster optimization, resulting in the need to provision node(s) based on application demand, or to defragment node resources to enable node suspension.

Workload Optimization Manager can recommend the following actions:

- Resize namespace compute resource quotas
- Resize container limits and requests

NOTE:
Executing several container resize actions can be very disruptive since pods need to restart with each resize. For replicas of the container scale group(s) related to a single Workload Controller, Workload Optimization Manager consolidates resize actions into one *merged action* to minimize disruptions. When a merged action has been executed (via the associated Workload Controller), all resizes for all related container specifications will be changed at the same time, and pods will restart once.

- Move pods
- Provision or suspend nodes
- Scale volumes

NOTE:
For a cluster in the public cloud, Workload Optimization Manager shows the cost impact of actions on nodes and volumes, to help you track your cloud spend. Workload Optimization Manager only reports the costs attached to these actions, and does *not* perform cost analysis on the cluster. See the Optimized Savings and Optimized Investments charts for more information.

For an on-prem cluster, Workload Optimization Manager can also recommend the following actions:

- Move VMs
- Provision or suspend hosts
- Provision or suspend storage

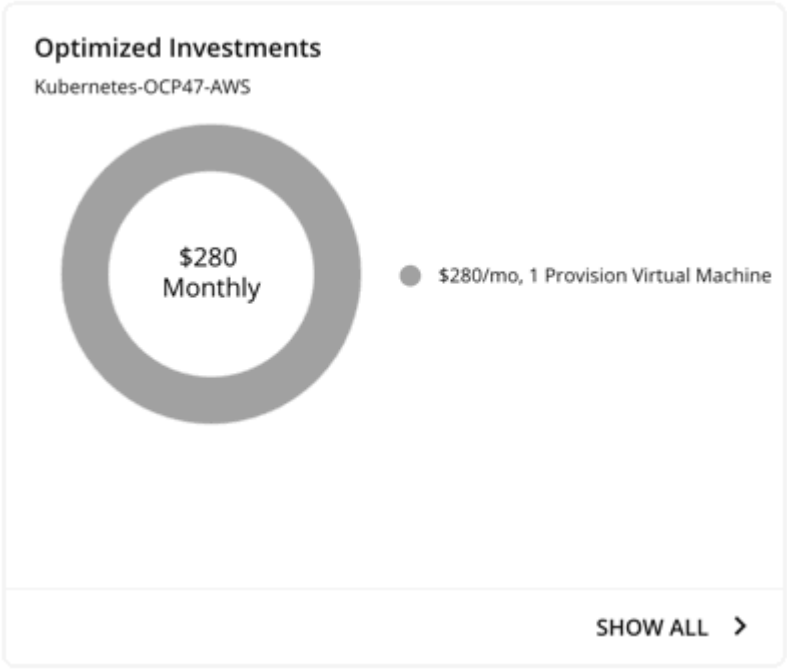
Optimized savings



For a cluster in the public cloud, Workload Optimization Manager shows the savings you would realize if you execute the actions (such as node suspension) that the plan recommends to increase infrastructure efficiency. Note that efficiency is the driver of this action, *not* cost. Cost information is included to help you track your cloud spend.

The chart shows total monthly savings. Click **Show All** to view the actions with cost savings.

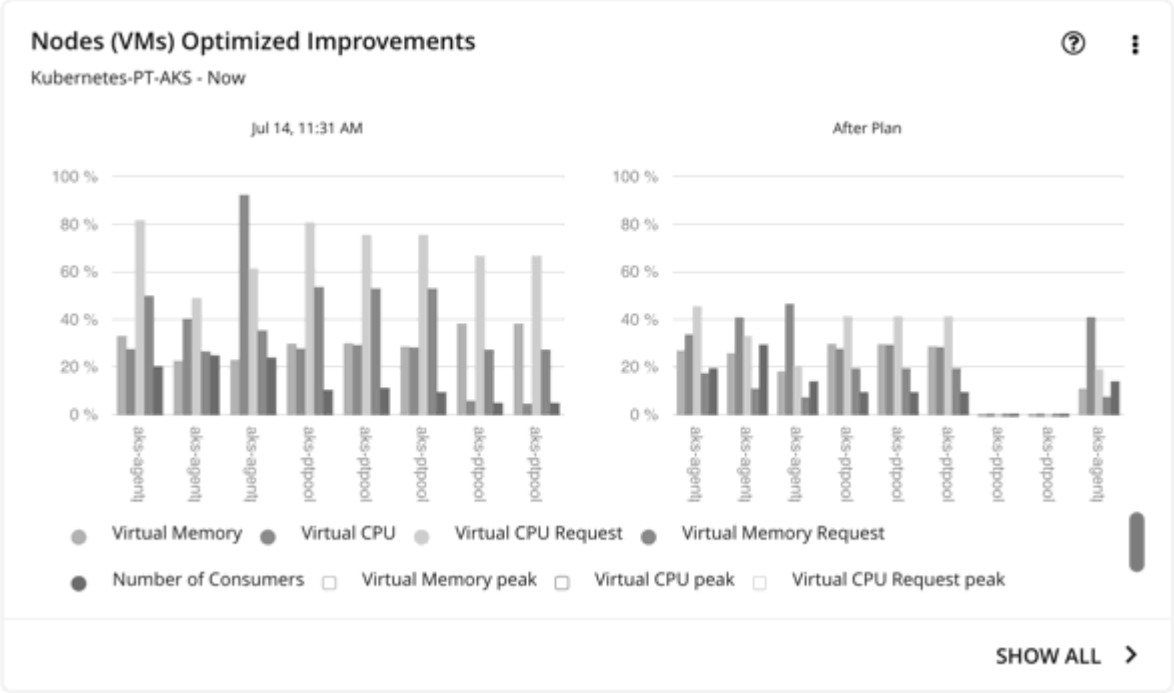
Optimized investments



For a cluster in the public cloud, Workload Optimization Manager shows the costs you would incur if you execute the node and volume scaling actions that the plan recommends to address performance issues. For example, if some applications risk losing performance, Workload Optimization Manager can recommend provisioning nodes to increase capacity. This chart shows how these actions translate to an increase in expenditure. Note that performance and efficiency are the drivers of these actions, *not* cost. Cost information is included to help you plan for the increase in capacity.

The chart shows total monthly investments. Click **Show All** to view the actions that require investments.

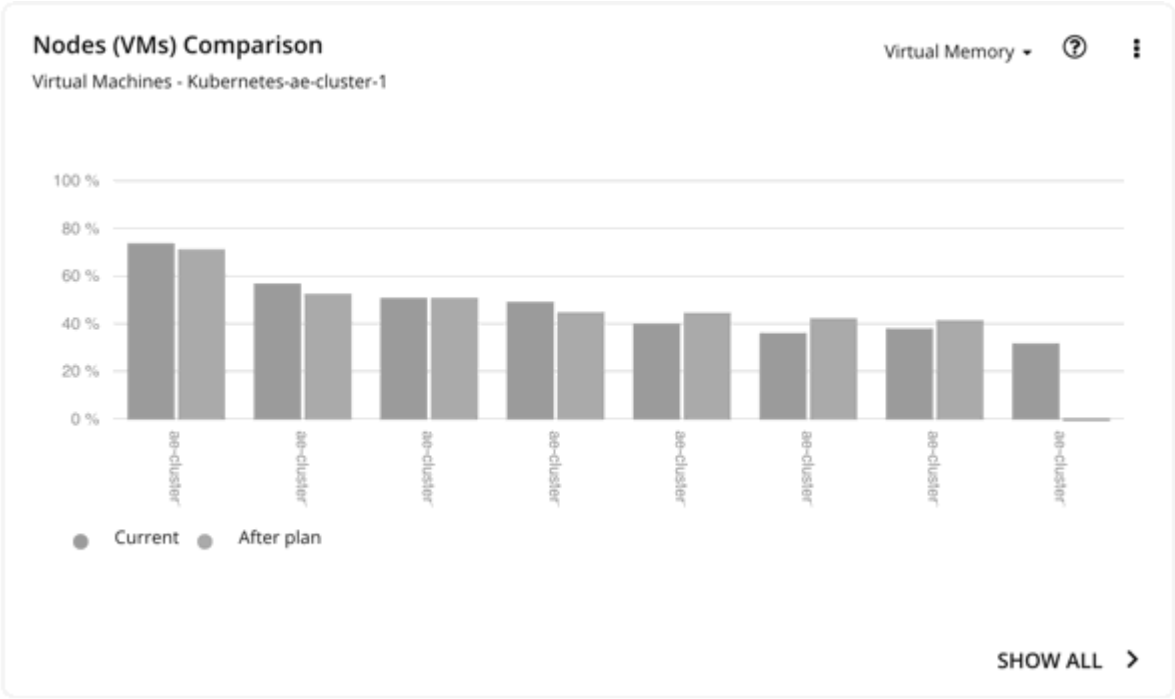
Nodes (VMs) optimized improvements



This chart compares the following before and after the plan:

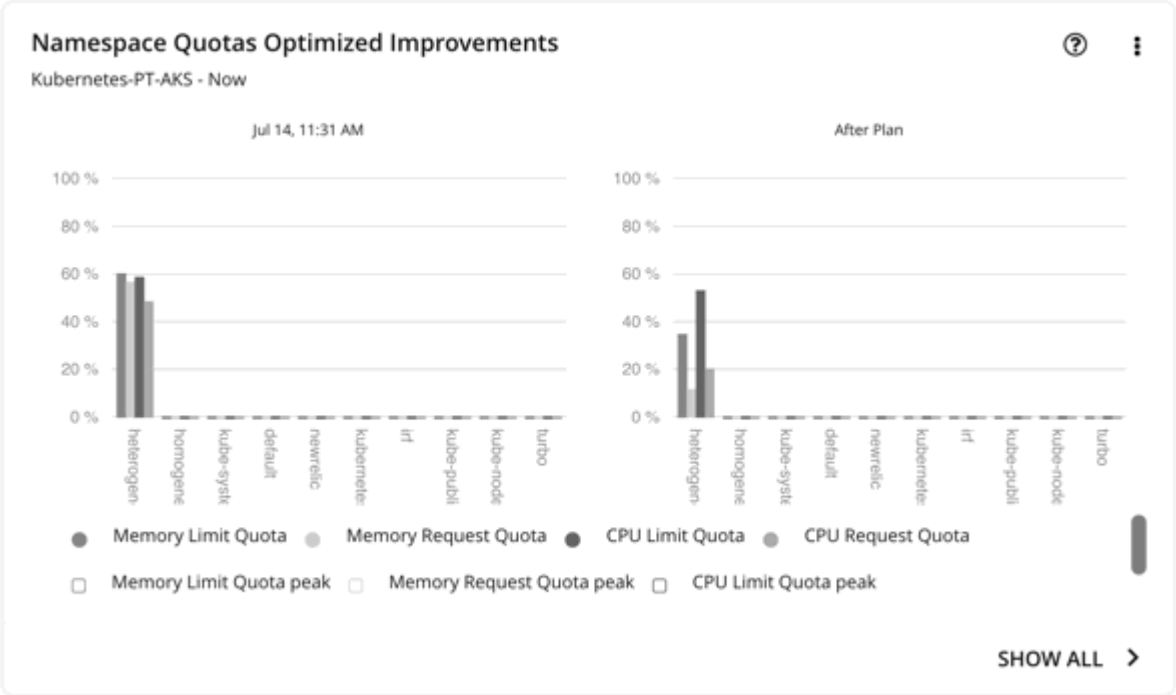
- Utilization of the following for all nodes:
  - vMem
  - vCPU
  - vMem Request
  - vCPU Request
- Number of pods consuming resources against the maximum pod capacity for all the nodes

Nodes (VMs) comparison



This chart compares node resource utilization (one metric at a time) before and after the plan.

Namespace quotas optimized improvements



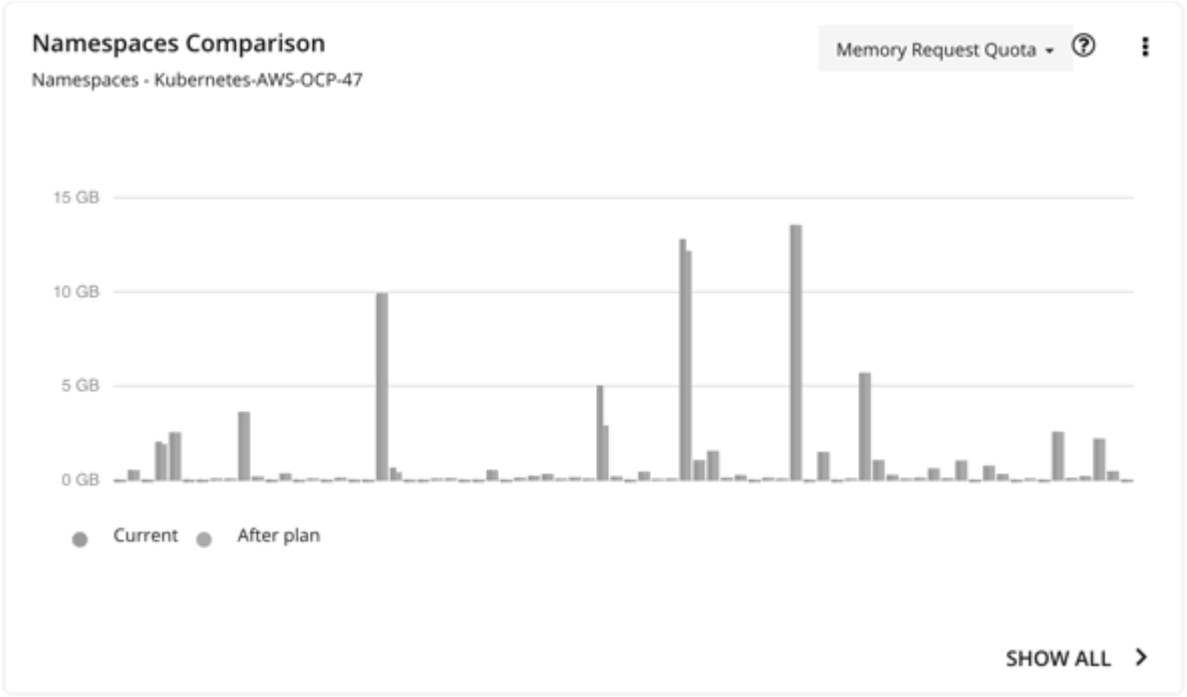
This chart shows pod utilization of resource quotas defined in namespaces. Resource quotas include:

- CPU Limit Quota
- Memory Limit Quota
- CPU Request Quota
- Memory Request Quota

For namespaces without defined quotas, utilization is 0 (zero).

With or without quotas, you can see the sum of pod limits and requests per namespace. Go to the Plan Results page, click the download button, and select **Namespace**. Utilization data in the downloaded file shows these limits and requests. You can also compare usage values in the Namespaces Comparison chart.

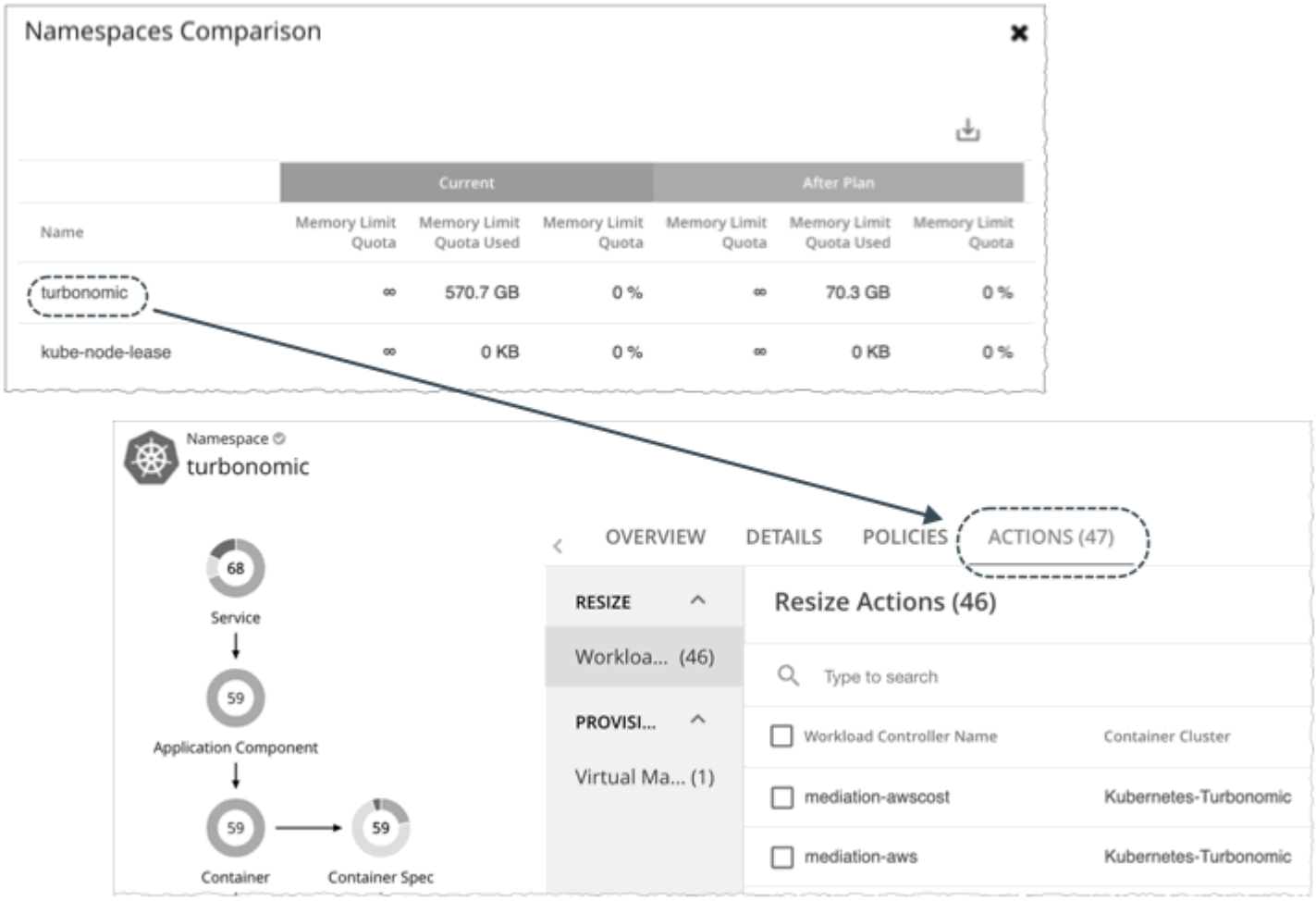
Namespaces comparison



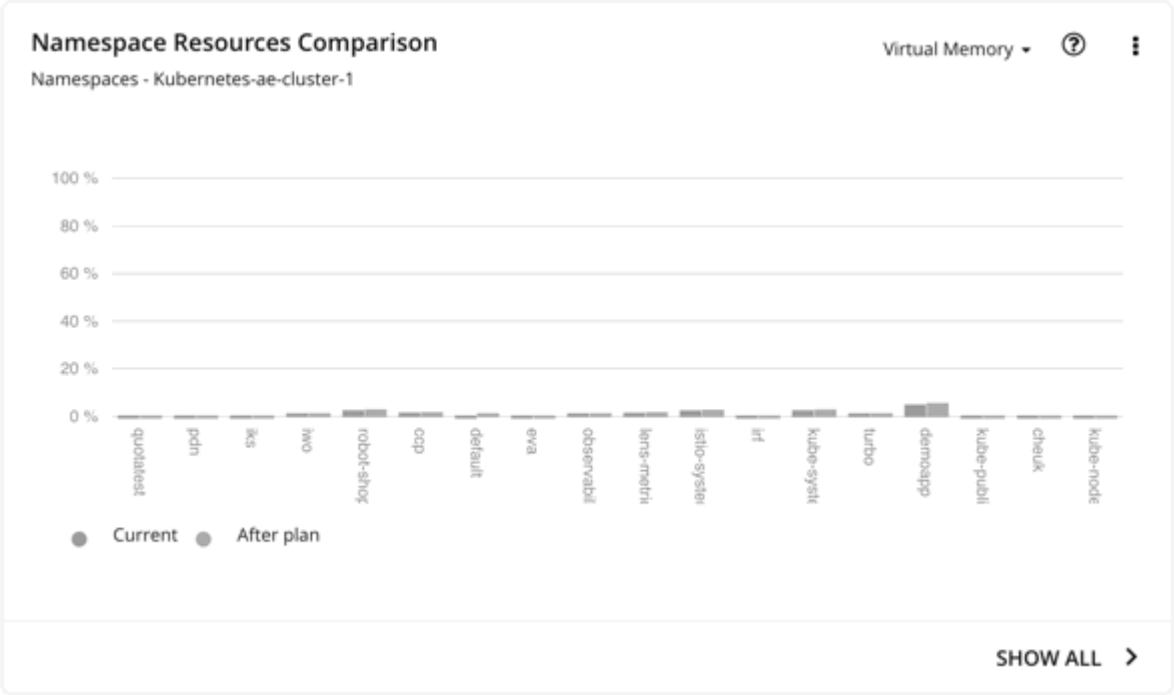
This chart compares namespace quota usage (one metric at a time) before and after the plan.

Use this chart to see how container resizing changes the limits and requests allocated per namespace, whether you leverage quotas or not.

To achieve the 'After Plan' results, click **Show All**. In the Details page that opens, go to the Name column and then click the namespace link. This opens another page with a list of pending actions for the namespace.


Namespaces Comparison
Current
After Plan
Name
Memory Limit Quota
Memory Limit Quota Used
Memory Limit Quota
Memory Limit Quota
Memory Limit Quota Used
Memory Limit Quota
turbonomic
∞
570.7 GB
0 %
∞
70.3 GB
0 %
kube-node-lease
∞
0 KB
0 %
∞
0 KB
0 %
Namespace
turbonomic
OVERVIEW
DETAILS
POLICIES
ACTIONS (47)
68
Service
59
Application Component
59
Container
59
Container Spec
RESIZE
Workloa... (46)
PROVISI...
Virtual Ma... (1)
Resize Actions (46)
Type to search
Workload Controller Name
Container Cluster
mediation-awscost
Kubernetes-Turbonomic
mediation-aws
Kubernetes-Turbonomic

Namespace resources comparison



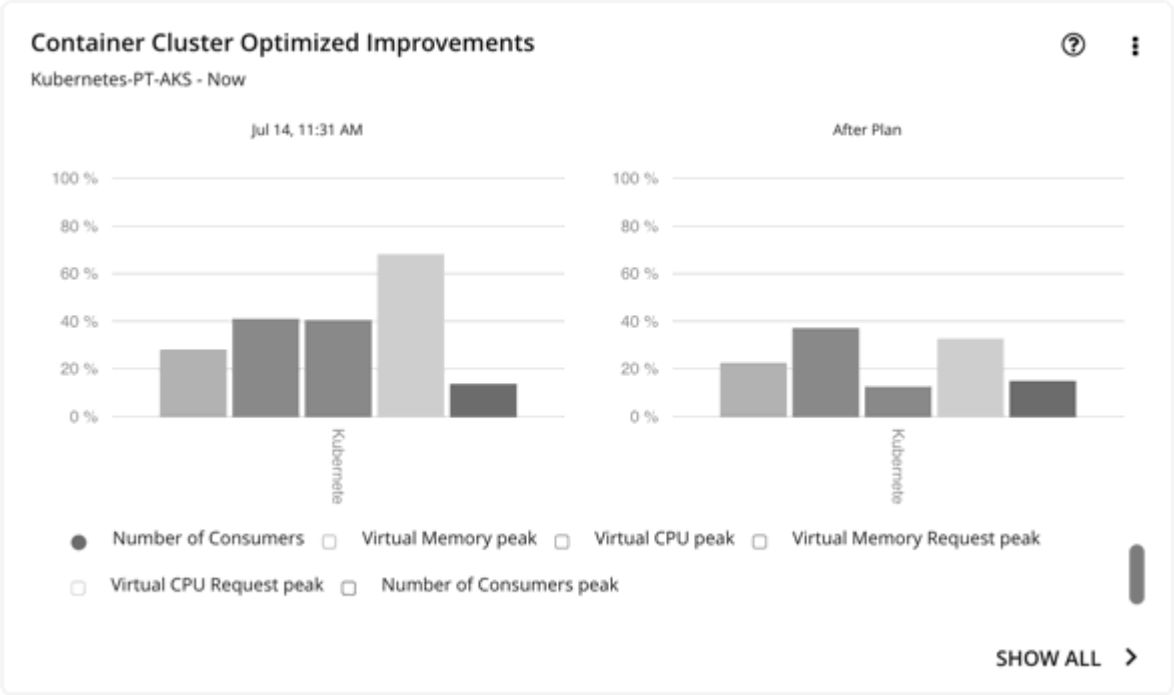
This chart shows how much cluster resources per namespace are utilized by pods. Utilization can be expressed as follows:

```
Utilization = Sum of actual vMem/vCPU used by pods / vMem/vCPU capacity for the cluster
```

This information helps you understand which namespaces use the most cluster resources. You can also use it for showback analysis. vMem and vCPU utilized by pods in the namespaces would change when the number of nodes changes as a result of executing the plan actions.

This chart is especially useful if you do not have resource quotas defined in your namespaces.

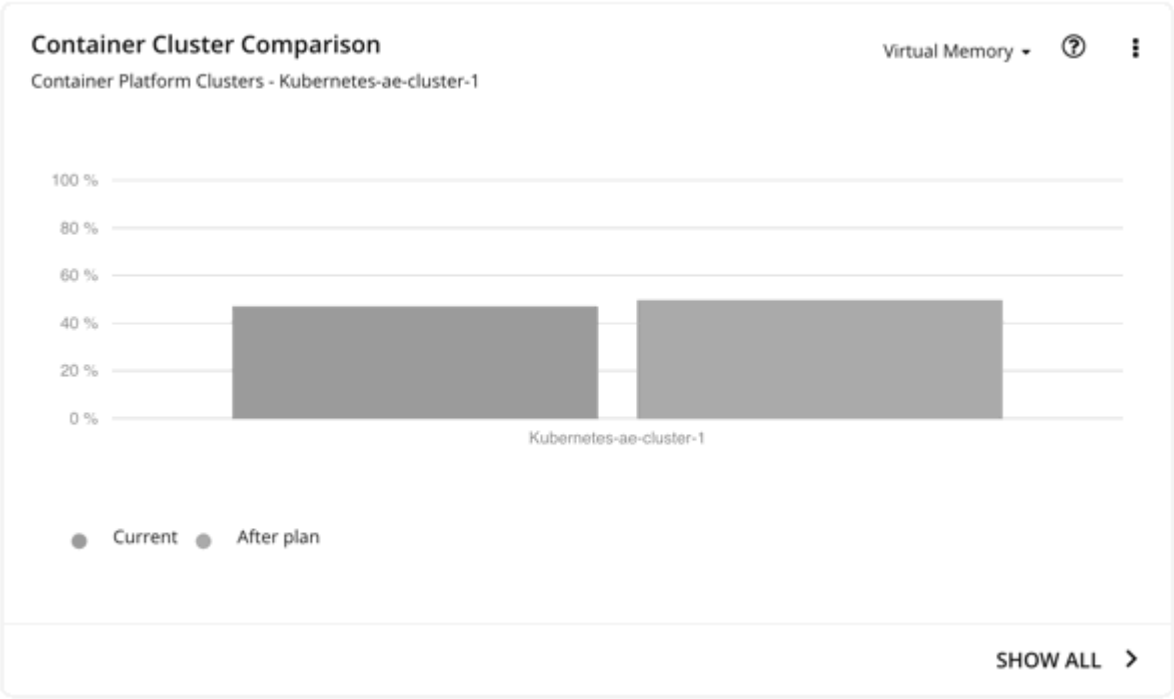
Container cluster optimized improvements



This chart shows the following, assuming you execute all actions in the plan:

- Changes to the utilization of cluster resources
- Overcommitment values

Container cluster comparison



This chart compares the following before and after the plan:

- Utilization of cluster resources (one metric at a time)
- Overcommitment values

Optimized improvements for hosts, storage, and virtual machines

Use these charts if you ran the plan in an on-prem cluster. These charts show how the utilization of resources would change if you accept all of the actions listed in the Plan Actions chart.

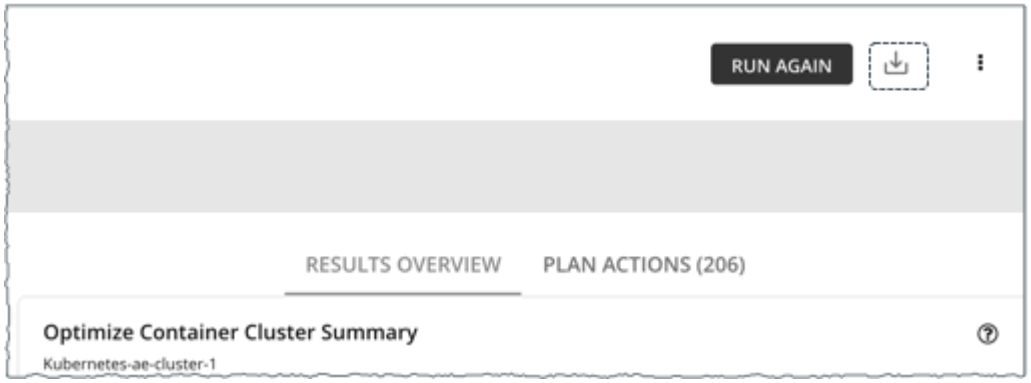
Hosts, storage devices, and virtual machines comparison

Use these charts if you ran the plan in an on-prem cluster. These charts show how the utilization of a particular commodity (such as memory or CPU) for each entity in the plan would change if you execute the recommended actions.

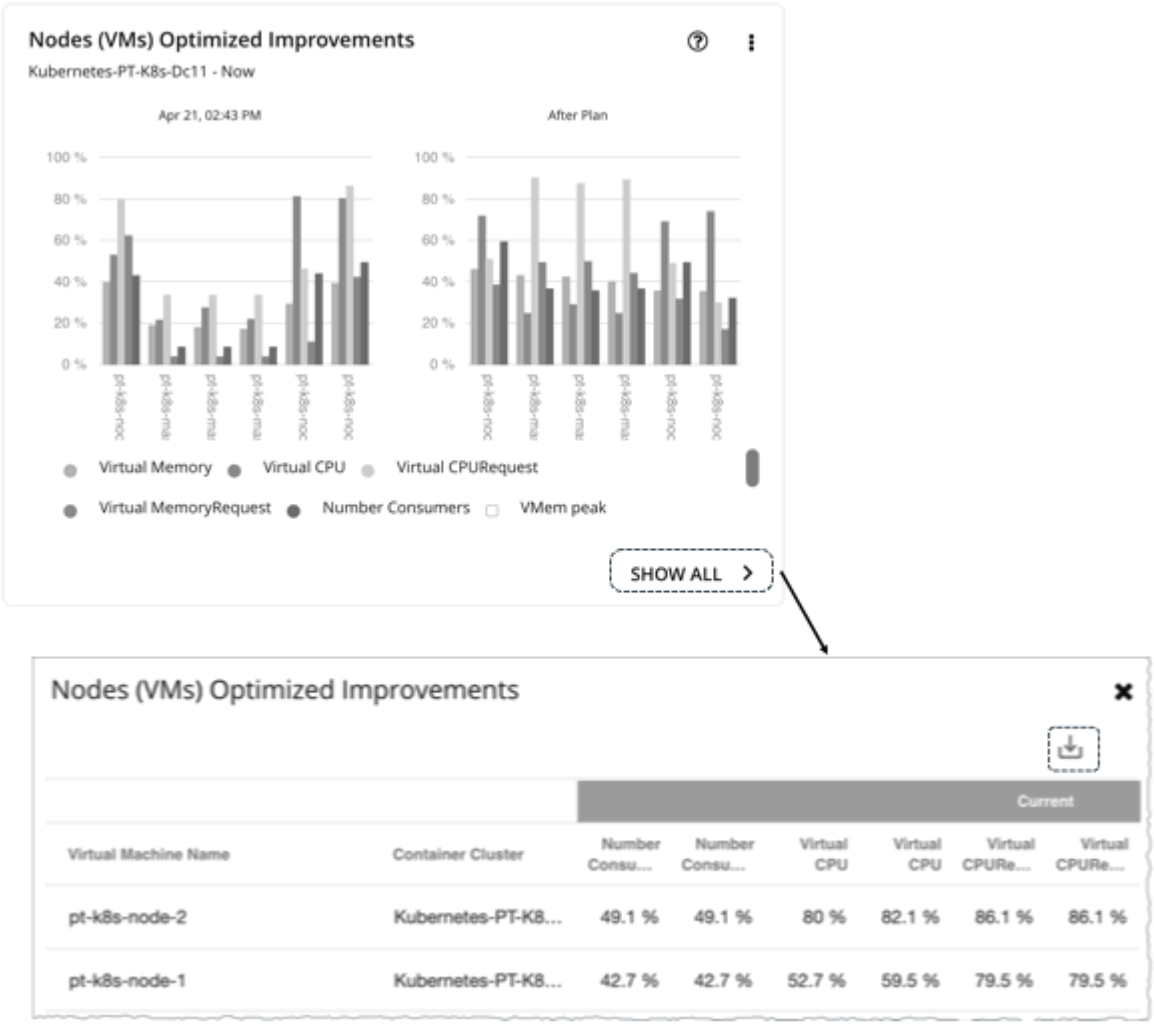
> **NOTE:**
> For the Storage Devices Comparison chart, if you set the view to **VM Per Storage** and click **Show all**, the total number of VMs sometimes does not match the number in the Summary chart. This happens if there are VMs in the plan that use multiple storage devices. The Storage Devices Comparison chart counts those VMs multiple times, depending on the number of storage devices they use, while the Plan Summary chart shows the actual number of VMs.

Downloading plan results

To download results for nodes, namespaces, or the cluster, click the download button on the Plan Results page.



You can also download the plan results shown in individual charts. Click the **Show All** button for a chart, and then the download button on the Details page.



For charts that display infinite capacities (for example, the Namespaces Comparison chart), the downloaded file shows an unusually high value, such as 1,000,000,000 cores, instead of the ∞ symbol.

Re-running the plan

You can run the plan again with the same or a different set of configuration settings. This runs the plan scenario against the market in its current state, so the results you see might be different, even if you did not change the configuration settings.

Use the toolbar for the Configuration section to change the configuration settings.

> **NOTE:**
> It is not possible to change the scope of the plan in the Plan Page. You must start over if you want a different scope. To start over, click the More options icon (⋮), and then select **New Plan**.

When you are ready to re-run the plan, click **Run Again**.

Migrate container workloads plan

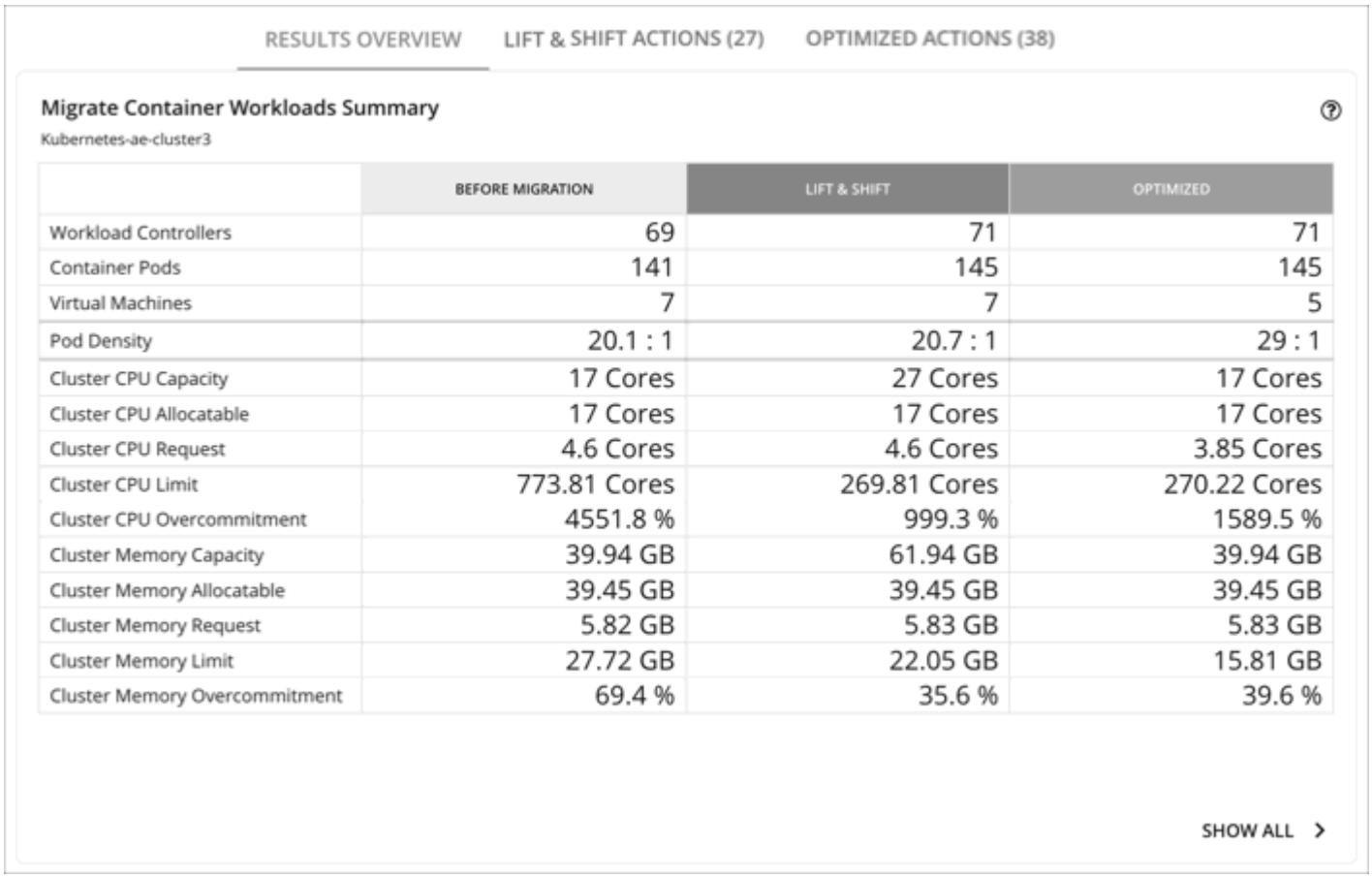
Run a Migrate Container Workloads plan to simulate the migration of workloads from one container platform cluster to another. For example, when you are ready to deploy workloads in your test cluster to production, you can run this plan to evaluate resource requirements in the production cluster. You might also have a need to decommission an existing cluster or consolidate

workloads in different clusters into a single cluster, both of which require moving existing workloads to a new cluster. Run this plan to see if all workloads can be placed in the new cluster.

The destination cluster (i.e., the cluster that you choose for migration) can be a:

- Standalone cluster
- Cluster in an on-prem or public cloud environment. For cloud environments, cost information is included in the plan results.
- Cluster stitched to applications via Data Ingestion Framework *(on page 522)* (DIF)

The plan results show the impact of workload migration on the destination cluster, based on two migration scenarios:



	BEFORE MIGRATION	LIFT & SHIFT	OPTIMIZED
Workload Controllers	69	71	71
Container Pods	141	145	145
Virtual Machines	7	7	5
Pod Density	20.1 : 1	20.7 : 1	29 : 1
Cluster CPU Capacity	17 Cores	27 Cores	17 Cores
Cluster CPU Allocatable	17 Cores	17 Cores	17 Cores
Cluster CPU Request	4.6 Cores	4.6 Cores	3.85 Cores
Cluster CPU Limit	773.81 Cores	269.81 Cores	270.22 Cores
Cluster CPU Overcommitment	4551.8 %	999.3 %	1589.5 %
Cluster Memory Capacity	39.94 GB	61.94 GB	39.94 GB
Cluster Memory Allocatable	39.45 GB	39.45 GB	39.45 GB
Cluster Memory Request	5.82 GB	5.83 GB	5.83 GB
Cluster Memory Limit	27.72 GB	22.05 GB	15.81 GB
Cluster Memory Overcommitment	69.4 %	35.6 %	39.6 %

- Lift & Shift

  The Lift & Shift scenario migrates your container workloads based on the resources currently available in the destination cluster.

- Optimized

  The Optimized scenario identifies opportunities to optimize performance in the destination cluster. For example, after analyzing historical resource utilization, Workload Optimization Manager might recommend resizing limits and requests for containers (via the associated workload controllers) to maintain performance. If you were to migrate workloads based on the resources currently available in the destination cluster, then your applications could risk losing performance.

NOTE:
The plan results do not show data for the source cluster.

The results further highlight the actions you need to take to maintain and optimize workload performance in the destination cluster.

Configuring a migrate container workloads plan

You can begin creating a plan scenario from two places in the user interface:

- Plan Page (Recommended)

  On the menu, click **Plan** to open the Plan Page. Click **New Plan**, and then select **Migrate Container Workloads**.

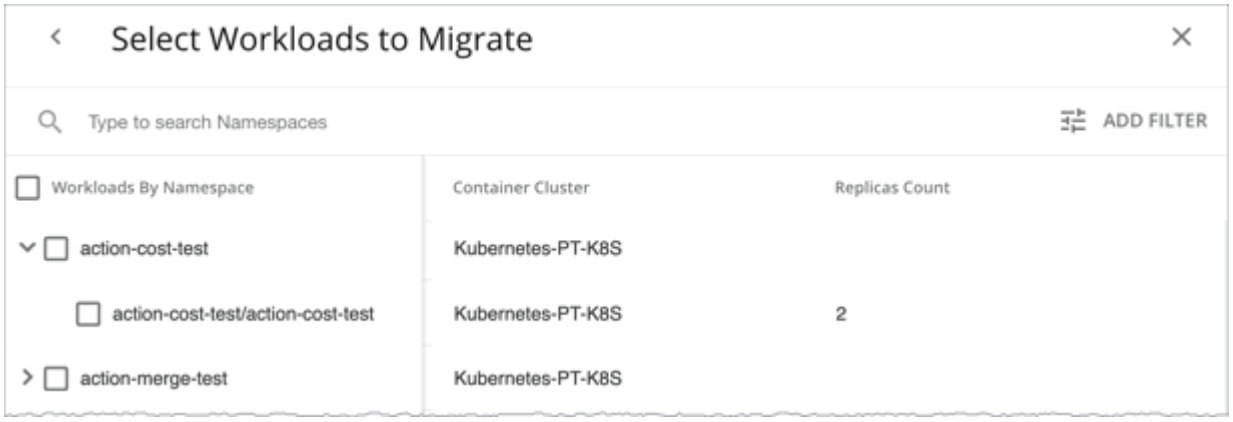
- Supply chain

  To start a plan scenario from the supply chain, set the scope to a container platform cluster and then click **Plan** at the upper-right corner of the page.

For an overview of setting up plan scenarios, see Setting Up Plan Scenarios *(on page 851)*.

Scope

Select the container workloads that you want to migrate.



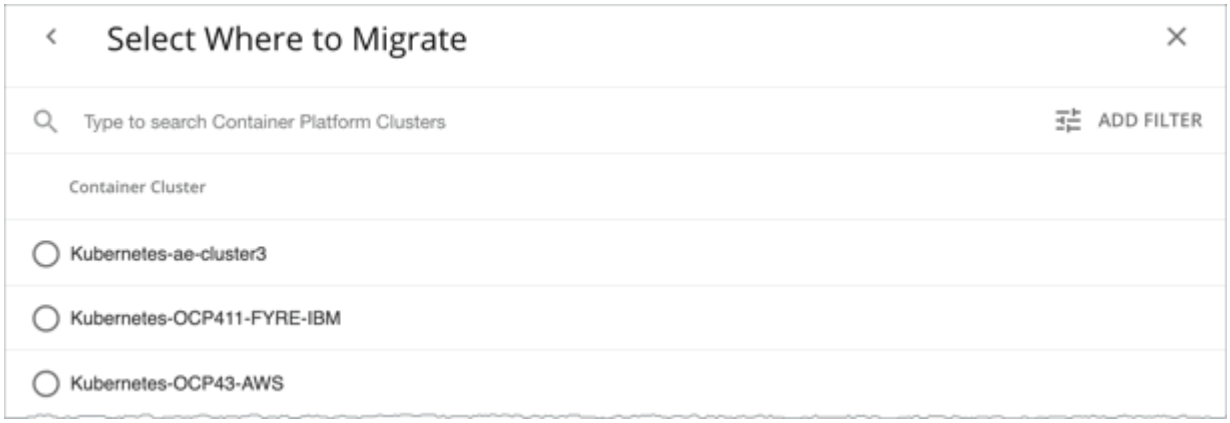
Container workloads are grouped by namespace. Expand a namespace to see individual Workload Controllers that manage workloads, and the number of container replicas for each Workload Controller. You can adjust the number of replicas in a later step. For example, you can add replicas to meet workload demand.

To define your scope, select individual namespaces, Workload Controllers, or both from one or several clusters. For example, if you are migrating workloads in two clusters, select the namespaces in those clusters.

To find specific workloads, type a keyword in the Search bar or click **Add Filter**. You can filter workloads by container platform cluster, namespace, tag, or Workload Controller.

Where to migrate

Choose the destination cluster for the container workloads that you selected.



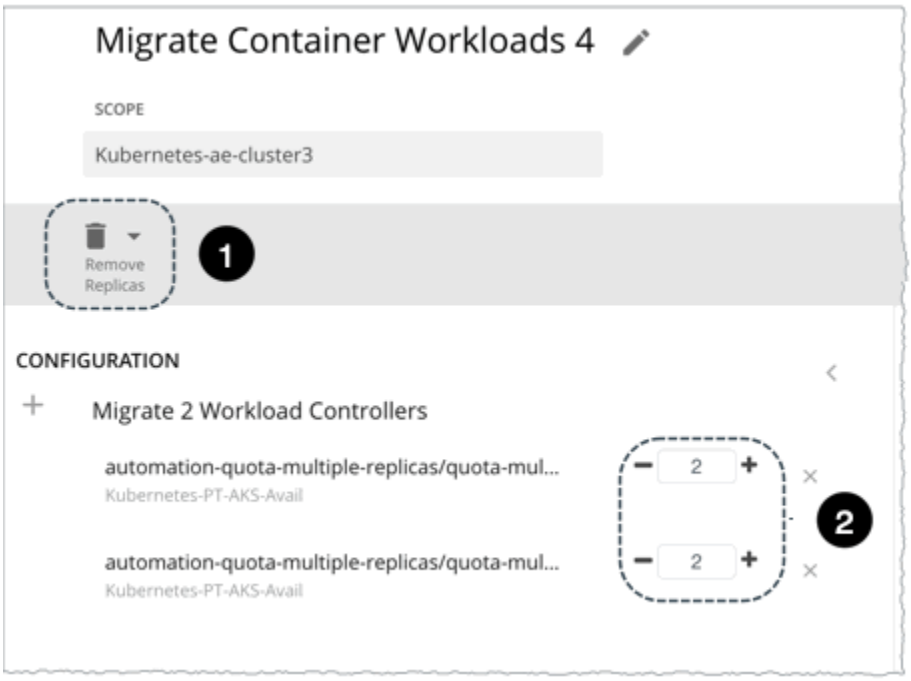
You can migrate to a:

- Stand-alone container platform cluster
- Container platform cluster in an on-prem or public cloud environment
- Container platform cluster that is stitched to applications though the Data Ingestion Framework *(on page 522)* (DIF)

After selecting a cluster, you can:

- Run the plan.

  Or
- Choose **Skip to Configuration** to configure other settings, including:



1. Removing existing replicas from the destination cluster. You may need to do this to make room for the workloads you are migrating.
2. Adjusting the number of replicas for the Workload Controllers in scope. For example, you may need to add replicas to meet workload demand.

Working with migrate container workloads plan results

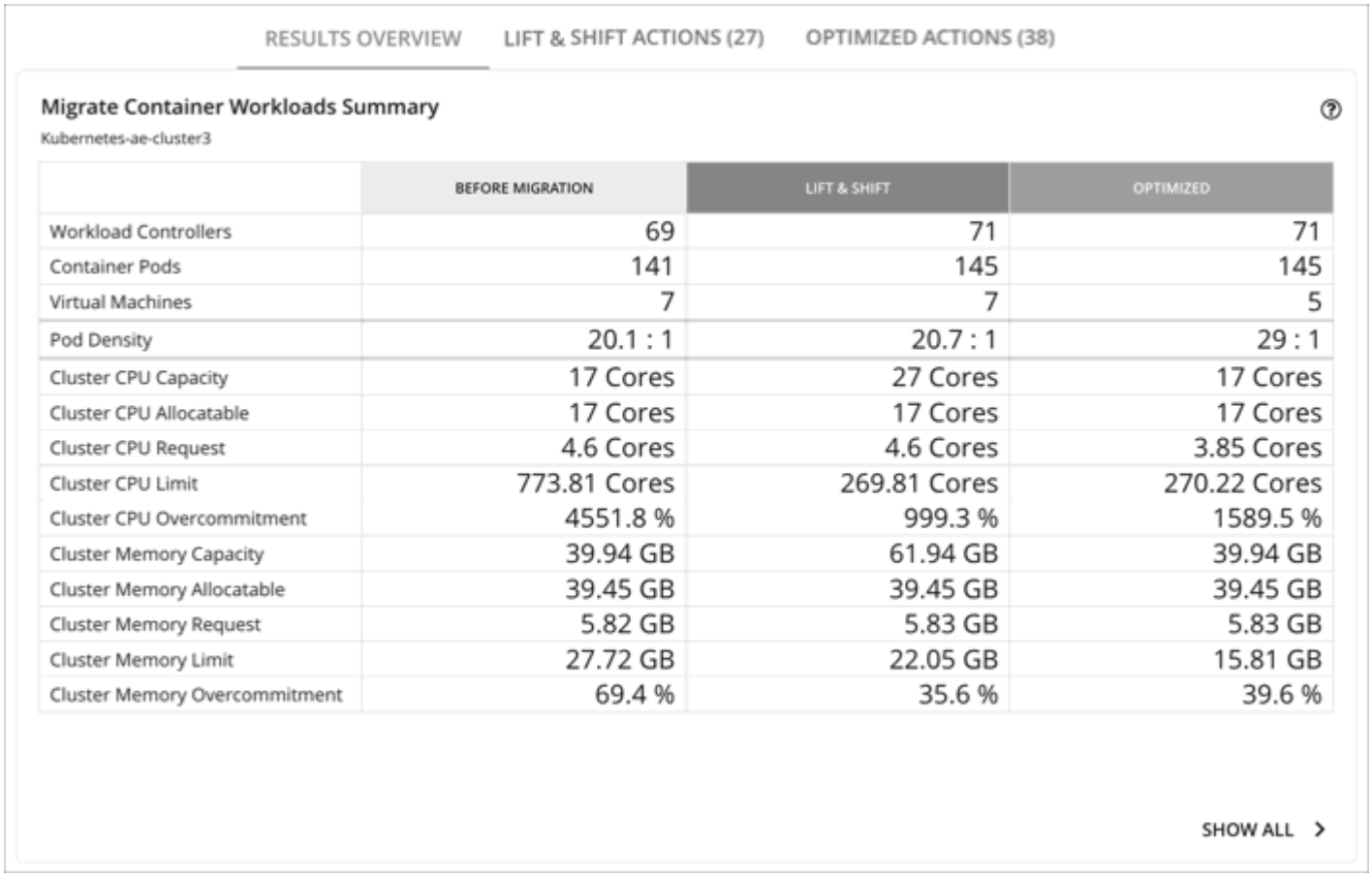
After the plan runs, you can view the results to see the impact of workload migration on the *destination* cluster. The results do *not* show data for the *source* cluster (for example, the cluster that currently hosts the workloads).

General guidelines

Familiarize yourself with these common terms that appear in the plan results:

- A container pod represents the compute demand from a running pod.
- A node (virtualized or bare metal) is represented as a VM.
- *Used* (or *Usage*) values represent actual resource consumption. For example, a node that consumes 100 MB of memory has a used value of 100 MB.
- *Utilization* values represent used/usage values against capacity. For example, a node that consumes 100 MB of memory against a total capacity of 500 MB has a utilization value of 20%.

Migrate container workloads summary



	BEFORE MIGRATION	LIFT & SHIFT	OPTIMIZED
Workload Controllers	69	71	71
Container Pods	141	145	145
Virtual Machines	7	7	5
Pod Density	20.1 : 1	20.7 : 1	29 : 1
Cluster CPU Capacity	17 Cores	27 Cores	17 Cores
Cluster CPU Allocatable	17 Cores	17 Cores	17 Cores
Cluster CPU Request	4.6 Cores	4.6 Cores	3.85 Cores
Cluster CPU Limit	773.81 Cores	269.81 Cores	270.22 Cores
Cluster CPU Overcommitment	4551.8 %	999.3 %	1589.5 %
Cluster Memory Capacity	39.94 GB	61.94 GB	39.94 GB
Cluster Memory Allocatable	39.45 GB	39.45 GB	39.45 GB
Cluster Memory Request	5.82 GB	5.83 GB	5.83 GB
Cluster Memory Limit	27.72 GB	22.05 GB	15.81 GB
Cluster Memory Overcommitment	69.4 %	35.6 %	39.6 %

Table columns

Table columns shows the following information for the destination cluster:

- Before migration

  This column shows the state of the destination cluster before workload migration.
- Lift & Shift

  The Lift & Shift scenario migrates your container workloads based on the resources currently available in the destination cluster.
- Optimized

  The Optimized scenario identifies opportunities to optimize performance in the destination cluster. For example, after analyzing historical resource utilization, Workload Optimization Manager might recommend resizing limits and requests for containers (via the associated workload controllers) to maintain performance. If you were to migrate workloads based on the resources currently available in the destination cluster, then your applications could risk losing performance.

Table rows

Table rows show the following information for the destination cluster:

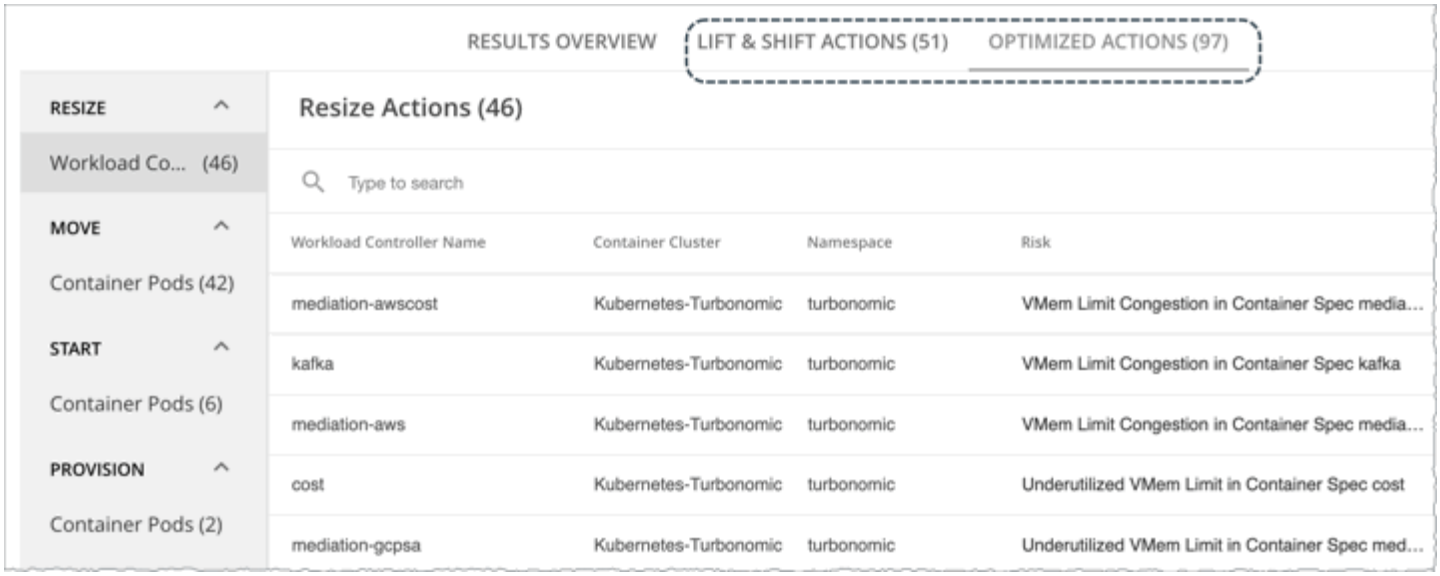
Row	Description
Workload con-trollers	Count of active Workload Controllers. This chart does not count "non-participating" entities in the real-time market, such as inactive Workload Controllers.

Row	Description
Container pods	Count of active (running) container pods.
Virtual machines	Count of active nodes (virtualized or bare metal). This chart does not count "non-participating" nodes in the real-time market, such as suspended nodes.
Pod density	Average number of pods per node. The 'Optimized' result shows if you can improve density by increasing the number of pods per node. For the total number of pods against the node capacity (maximum pods per node), see the **Number of Consumers** data in the following charts: ▪ Nodes (VMs) Optimized Improvements ▪ Container Platform Cluster Optimized Improvements
Cluster CPU capacity	Total CPU capacity for the cluster. The 'Optimized' result indicates how much CPU capacity will result in the optimal number of nodes required to run workloads.
Cluster CPU allocatable	Total amount of cluster CPU available for pod requests. The 'Optimized' result indicates how much of the allocatable CPU capacity will change if you provision or suspend nodes.
Cluster CPU request	Total CPU request capacity for the cluster.
Cluster CPU limit	Total CPU limit capacity for the cluster.
Cluster CPU overcommitment	(Only for containers with CPU limits) This indicates whether the CPU limits exceed the capacity of the underlying nodes. A value greater than 100% indicates overcommitment. Workload Optimization Manager manages cluster resources by actual utilization and limit rightsizing so that you can run more workloads with less risk. Workload Optimization Manager only calculates overcommitment in plans. The calculation can be expressed as: `Overcommitment = Sum of CPU limits for all containers / Sum of CPU capacity for all nodes`
Cluster memory capacity	Total memory capacity for the cluster. The 'Optimized' result indicates how much memory capacity will result in the optimal number of nodes required to run workloads.
Cluster memory allocatable	Total amount of cluster memory available for pod requests. The 'Optimized' result indicates how much of the allocatable memory capacity will change if you provision or suspend nodes.
Cluster memory request	Total memory request capacity for the cluster.
Cluster memory limit	Total memory limit capacity for the cluster.
Cluster memory overcommitment	(Only for containers with memory limits) This indicates whether the memory limits exceed the capacity of the underlying nodes. A value greater than 100% indicates overcommitment. Workload Optimization Manager manages cluster resources by actual utilization and limit rightsizing so that you can run more workloads with less risk. Workload Optimization Manager only calculates overcommitment in plans. The calculation can be expressed as:

Row	Description
	`Overcommitment = Sum of memory limits for all containers / Sum of memory capacity for all nodes`

Plan actions

Workload Optimization Manager shows separate tabs for **Lift & Shift** and **Optimized** migration actions. You can download the list of actions as a CSV file.



These tabs show the actions that you need to execute to achieve the plan results. For example, you might need to resize limits and requests for containers (via the associated Workload Controllers) to address performance issues. Or, you might need to move pods from one node to another to reduce congestion.

Smarter redistribution and workload rightsizing also drive actions, resulting in the need to provision node(s) based on application demand, or to defragment node resources to enable node suspension.

Workload Optimization Manager can recommend the following actions:

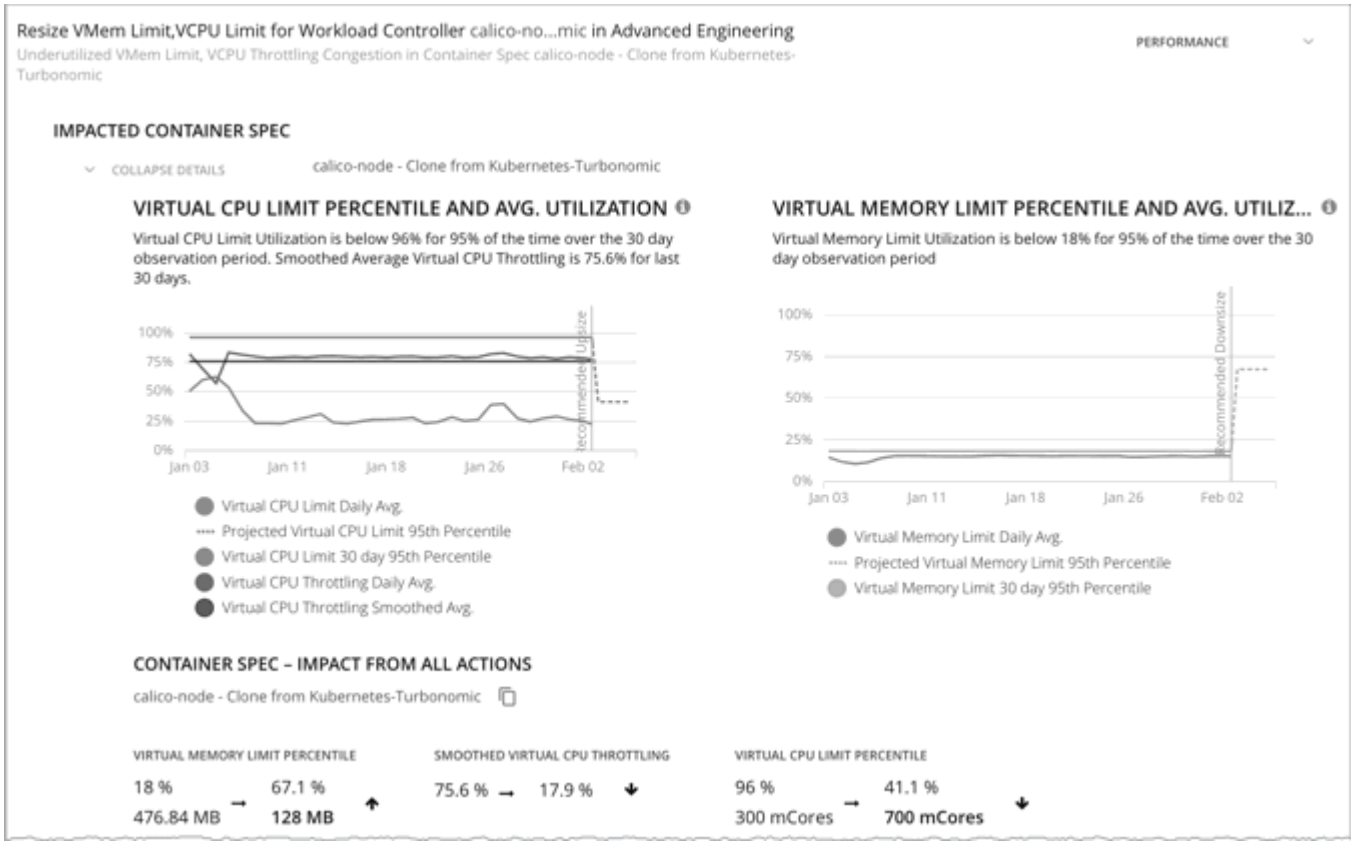
Entity	Lift & shift actions	Optimized actions
Workload controller	None	Resize container limits and requests See Workload Controller Resize Actions *(on page 884)* for additional details.
Namespace	None	Resize quotas (if container resize actions would exceed the namespace quotas)
Pod	▪ Move ▪ Provision (if node provision is required) ▪ Reconfigure	▪ Move ▪ Provision (if node provision is required) ▪ Suspend (if node suspension is required) ▪ Reconfigure
Node (VM)	▪ Provision ▪ Move on-prem nodes	▪ Provision ▪ Suspend ▪ Move on-prem nodes

Entity	Lift & shift actions	Optimized actions
Cloud volume	Scale	Scale
On-prem storage	▪ Provision ▪ Suspend	▪ Provision ▪ Suspend
On-prem host	▪ Provision ▪ Suspend	▪ Provision ▪ Suspend

Additional information:

- Workload controller resize actions

  When you expand an action on a Workload Controller, you will see charts that track VCPU and VMem utilization for the impacted container spec. With these charts, you can easily recognize the utilization trends that Workload Optimization Manager analyzed to make accurate resize decisions.



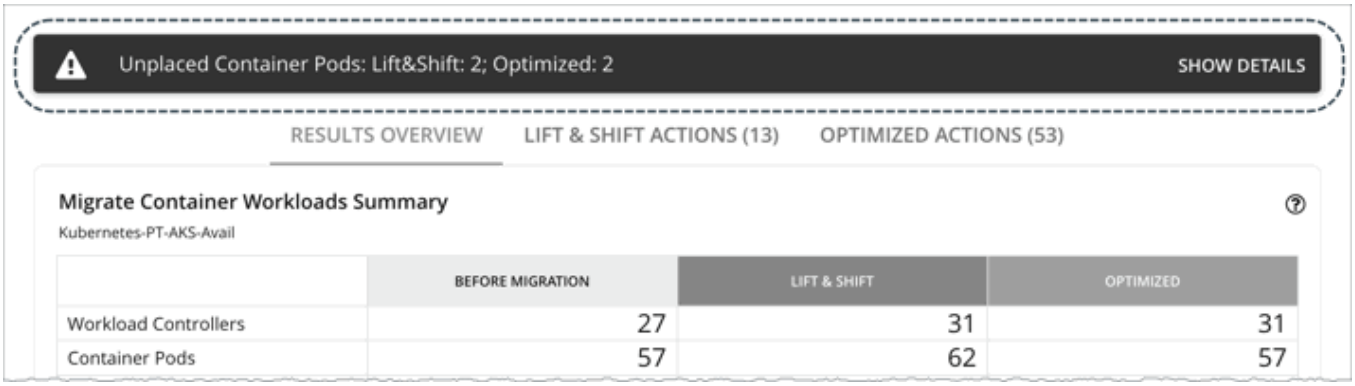
  For more information about these charts, see Utilization Charts *(on page 826)*.

  Executing several container resize actions can be very disruptive since pods need to restart with each resize. For replicas of the container scale groups related to a single Workload Controller, Workload Optimization Manager consolidates resize actions into one *merged action* to minimize disruptions. When a merged action has been executed (via the associated Workload Controller), all resizes for all related container specifications will be changed at the same time, and pods will restart once.

- Pod constraints

  A Migrate Container Workloads plan evaluates pod constraints on the source cluster when making placement decisions for pods. The plan enforces taints, tolerations, and `nodeSelector` specifications if there are matching constraints in the destination cluster. If these constraints cannot be achieved in the destination cluster, the plan ignores them to guarantee placement.

The plan always enforces affinity and anti-affinity constraints, which could result in unplaced pods in the destination cluster. If this happens, the plan generates reconfigure actions for the unplaced pods and shows a notification in the plan results.

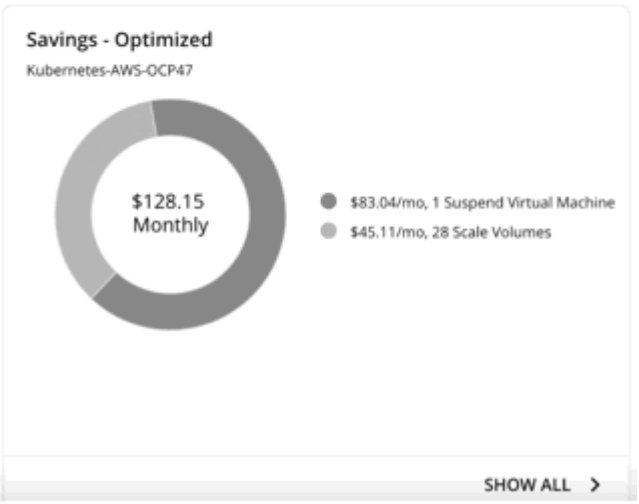


Click **Show Details** to see the list of pods and the reasons for their non-placement. The charts in the plan results do not count these pods.

Savings

For a destination cluster in the public cloud, Workload Optimization Manager shows the savings you would realize if you execute the actions that the plan recommends. The results show separate charts for the **Lift and Shift** and **Optimized** scenarios, so you can compare the impact of actions on your cloud expenses.





For example, in both scenarios, the plan might recommend scaling volumes to new tiers to address performance issues. If these new tiers happen to be more cost-effective than the current tiers, then the actions are treated as cost-saving measures. For the optimized scenario, the plan might also recommend suspending certain nodes to increase infrastructure efficiency, which could introduce additional savings.

Note that application performance and infrastructure efficiency are the drivers of these actions, *not* cost optimization. Cost information is included to help you track your cloud expenses.

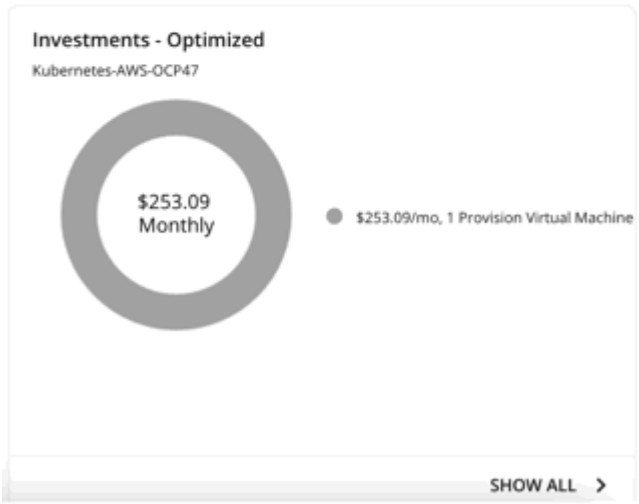
The charts show total monthly savings. Click **Show All** to view the actions with cost savings.

> **NOTE:**
> An empty chart indicates that no savings will be realized after you execute the recommended actions.

Investments

For a destination cluster in the public cloud, Workload Optimization Manager shows the costs you would incur if you execute the actions that the plan recommends. The results show separate charts for the **Optimized** and **Lift and Shift** scenarios, so you can compare the impact of actions on your cloud expenses.





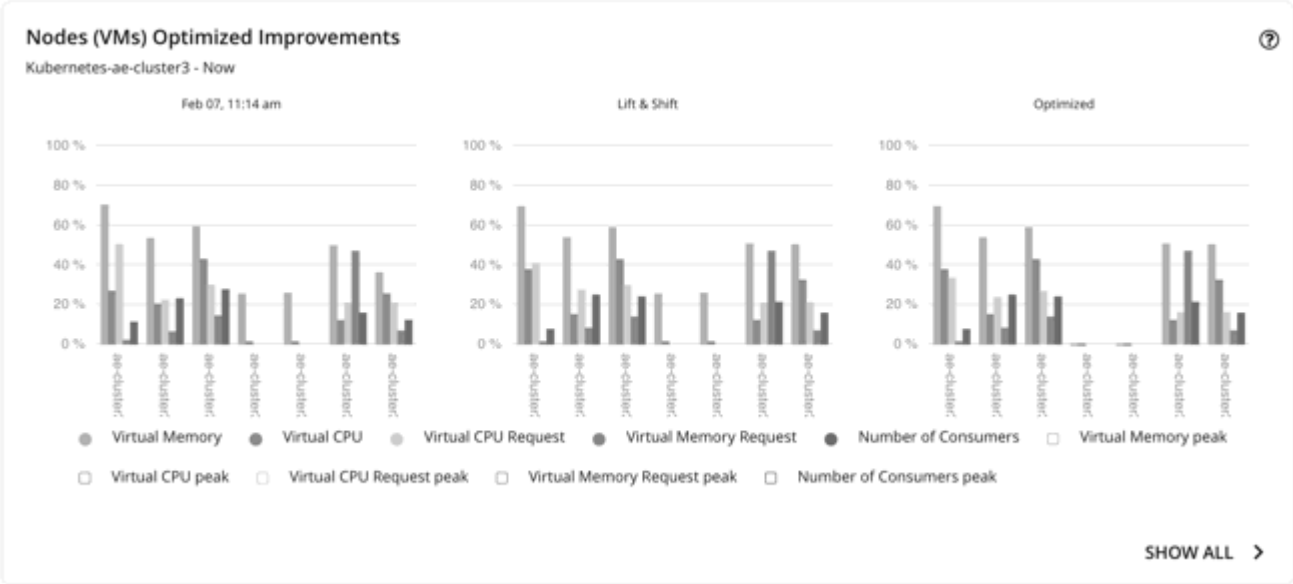
For example, if some applications risk losing performance, the plan might recommend provisioning nodes to increase capacity. The charts show how node provision actions translate to an increase in expenditure. You can use cost information in the charts as you seek approval to execute these actions.

The charts show total monthly investments. Click **Show All** to view the actions that require investments.

> **NOTE:**
> An empty chart indicates that no investments are required to execute the recommended actions.

Nodes (VMs) optimized improvements

This chart compares the following before and after the plan:

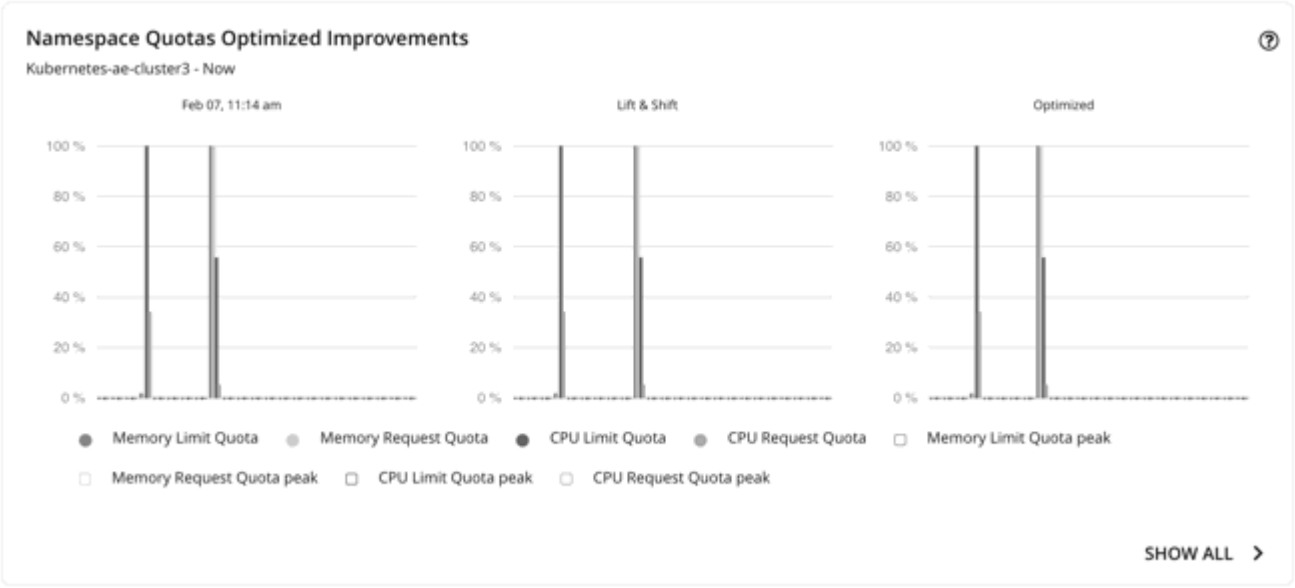


- Utilization of the following resources for all nodes:
  - vMem
  - vCPU
  - vMem Request
  - vCPU Request
- Number of pods consuming resources against the maximum pod capacity for all the nodes

Use this chart to identify nodes that are candidates for suspension, or resources with unusual utilization. You can also use this chart to drill down to a specific resource. Click **Show All** for more details.

Namespace quotas optimized improvements

This chart shows pod utilization of resource quotas defined in namespaces. Resource quotas include:



- CPU Limit Quota
- Memory Limit Quota

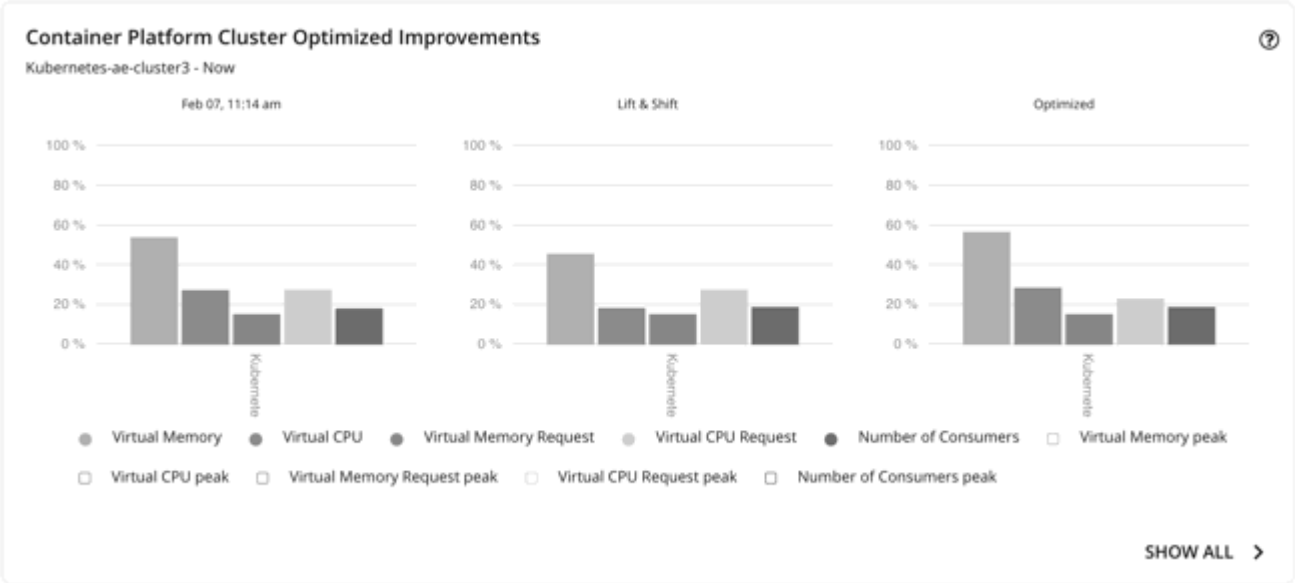
- CPU Request Quota
- Memory Request Quota

The chart highlights utilization improvements as a result of namespace resizing.

For namespaces without defined quotas, utilization is 0 (zero).

Container platform cluster optimized improvements

This chart shows the following, assuming you execute all actions in the plan:



- Changes to the utilization of cluster resources
- Overcommitment values

Optimized improvements for hosts and storage

Use these charts if you ran the plan on an on-prem cluster. These charts show how the utilization of resources would change assuming you accept all of the actions that the plan recommends.

Downloading plan results

You can download the plan results shown in individual charts. Click the **Show All** button for a chart, and then the download button at the Details page.

Re-running the plan

You can run the plan again with the same or a different set of configuration settings. This runs the plan scenario against the market in its current state, so the results you see might be different, even if you did not change the configuration settings.

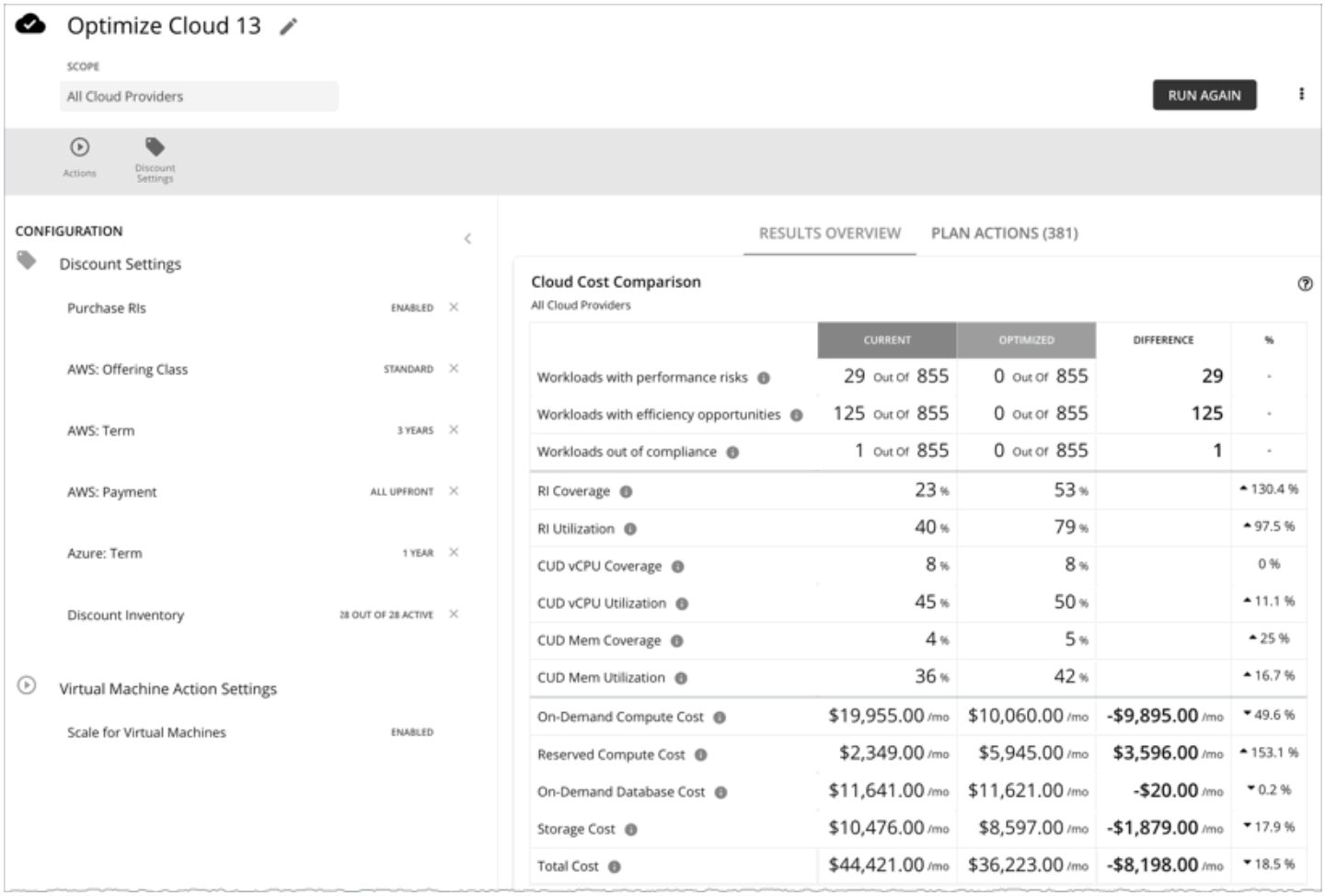
Use the toolbar for the Configuration section to change the configuration settings.

> **NOTE:**
> It is not possible to change the scope of the plan in the Plan Page. You must start over if you want a different scope. To start over, click the More options icon (⋮), and then select **New Plan**.

When you are ready to re-run the plan, click **Run Again**.

Optimize cloud plan

Run the Optimize Cloud plan to see how you can maximize savings while still assuring performance for your applications and workloads. This plan identifies ways to optimize your costs by choosing the best templates (most adequate compute resources), regions, accounts, or resource groups to host your workloads. The plan also identifies workloads that can change over to discounted pricing, and it compares your current costs to the costs you would get after executing the plan recommendations.

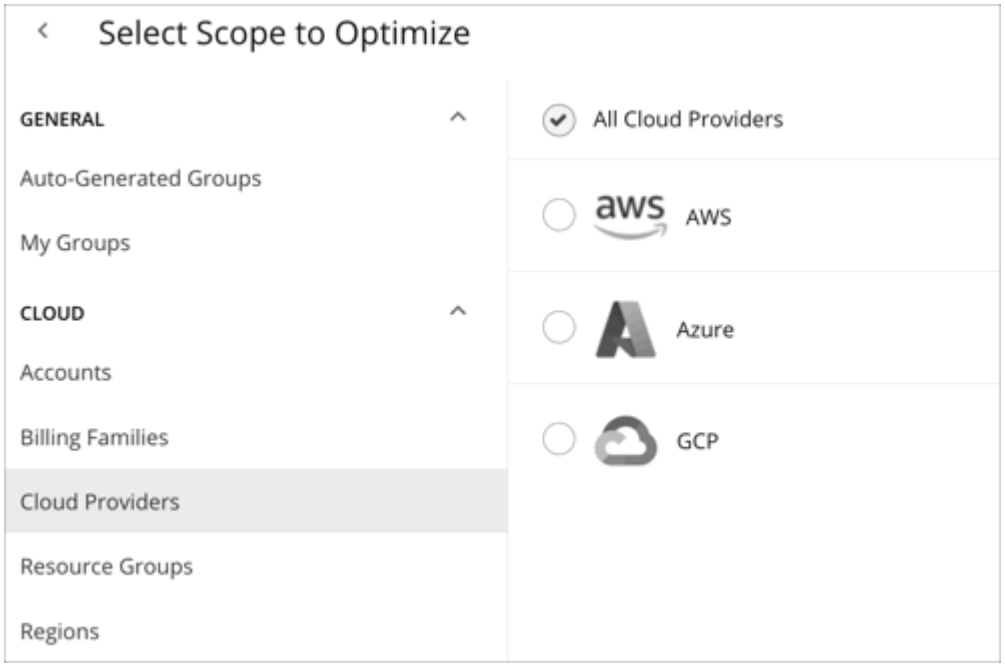


Configuring an optimize cloud plan

For an overview of setting up plan scenarios, see Setting Up Plan Scenarios *(on page 851)*.

Scope

You can scope by:



- Accounts

  Choose AWS accounts, Azure subscriptions, or Google Cloud projects as the plan's scope. The plan does not recommend discount purchases if you scope to this level. To optimize discount purchases for a limited scope, choose a Billing Family.
- Billing Families

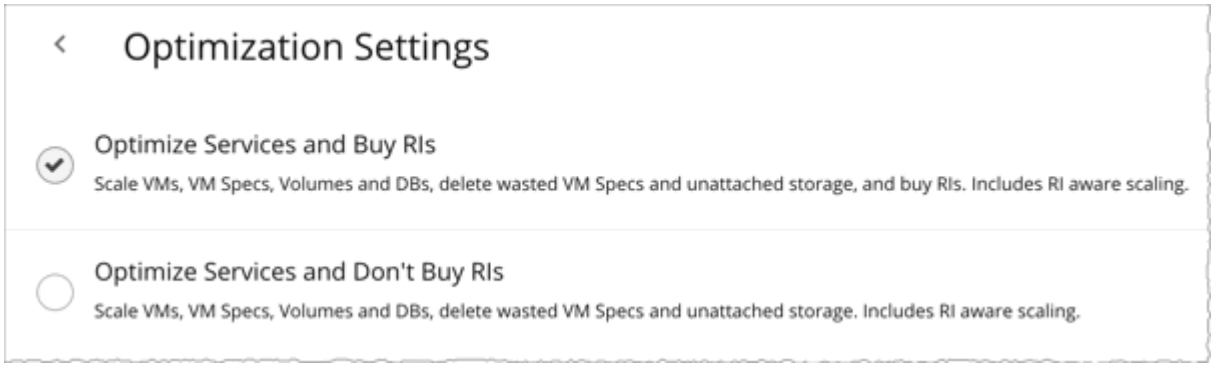
  Include discount purchases in the planning for a scope that is limited to a single billing family. The plan calculates discount purchases through the billing family's management account. Discount purchases are supported for AWS and Azure, but not Google Cloud.
- Cloud Providers

  See how you can optimize all your AWS, Azure, and Google Cloud workloads.
- Resource Groups

  Workload Optimization Manager discovers Azure resource groups. You can select one or more resource groups for the plan scope.
- Regions

  Focus the plan on a provider's region.

Optimization settings



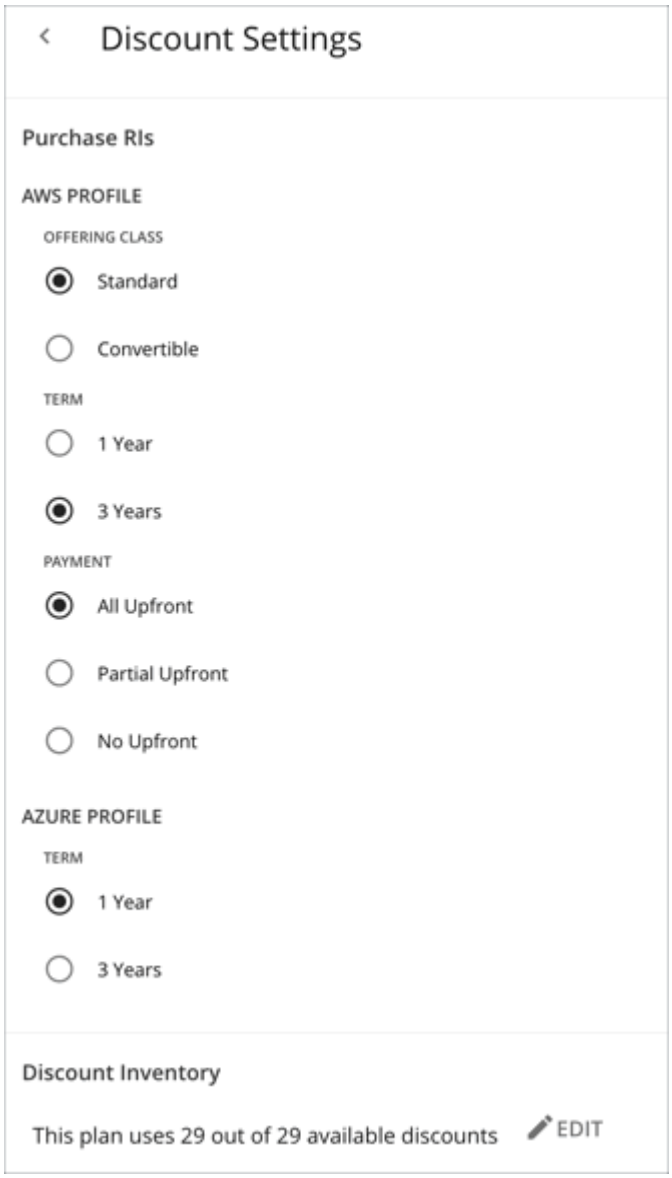
Choose from the optimization options. If you set a plan's scope to a resource group, Workload Optimization Manager optimizes services without recommending discount purchases.

If your goal is to purchase discounts for VMs at their current sizes, use the Buy VM Reservations plan type. For details, see Buy VM Reservations Plan *(on page 897)*.

NOTE:
If you turn on the **Disable All Actions** setting in the global default policy and then run an Optimize Cloud plan with VM scaling and discount purchases enabled, the plan results show inaccurate discount recommendations.

Turn off **Disable All Actions** to resolve this issue. After you turn off this setting, it will take Workload Optimization Manager a week to reflect accurate results in Optimize Cloud plans.

Discount settings



- Currently, **Purchase RIs** applies to only AWS and Azure. For **AWS/Azure Profile**, the settings that you have set up for real-time analysis are selected by default. You can change the settings to see how they affect costs.
  - Offering class

    Choose the offering class that corresponds to the RI types that you typically use in your environment.
  - Term

Choose the payment terms for your discounts. Typically, longer term payment plans cost less per year.

- Payment

  Choose your payment option.

  - All upfront – You make full payment at the start of the term.
  - Partial upfront – You make a portion of the payment at the start of the term, with the remain cost paid at an hourly rate.
  - No upfront – You pay at an hourly rate for the duration of the term.

- For **Discount Inventory**, the discounts for the current scope are selected by default. Click **Edit** to make changes.

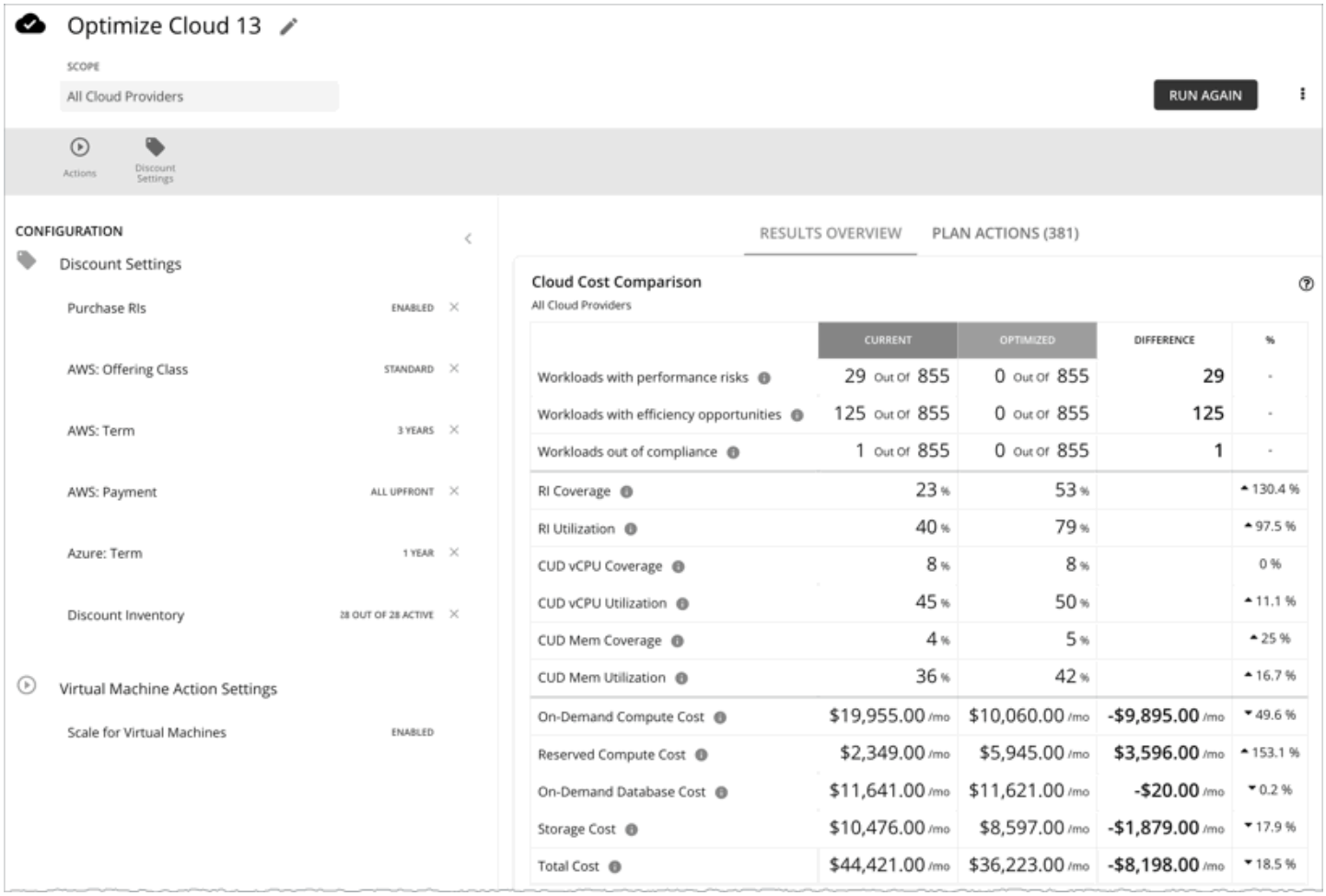
Working with optimize cloud plan results

After the Optimize Cloud plan runs, you can view the results to see how you can maximize savings or make other improvements to your cloud environment.

The plan results:

- Compare current to optimized costs, including on-demand compute, discounted compute, on-demand database, and storage costs
- Compare current and optimized breakdowns of workload tiers
- Compare breakdowns of storage tiers
- Project the discount coverage (how many workloads are covered by discounted pricing) and utilization (percentage of discounts that are active)
- Show the cost benefit of moving workloads from on-demand to discounted pricing

Viewing the results



The plan results include the following charts.

Cloud cost comparison

This chart highlights any difference in cost as a result of optimization. For example, undersized VMs risk losing performance and should therefore scale up. This could contribute to an increase in cost. On the other hand, oversized VMs can scale down to less expensive instances, so cost should go down. The values under the **%** column indicate the percentage of VMs that are affected by optimization cost calculations.

Workload Optimization Manager can also recommend that you purchase discounts to reduce costs. The analysis looks at workload history to identify workloads that can move from on-demand to discounted pricing. This considers the count of workloads in a family, plus their hours of active-state condition, to arrive at the discounts you should purchase. Since discounted costs are incurred at the account level, the Cloud Cost Comparison chart will present discounted costs or charges when you scope to an account or group of accounts (including a billing family).

For AWS clouds, Workload Optimization Manager can get the information it needs to display license costs for database instances. For Azure clouds, Workload Optimization Manager does not display database license costs because Azure does not make that information available.

Workload mapping

This chart shows the types of tiers you currently use, compared to the tiers the plan recommends, including how many of each type, plus the costs for each.

To see a detailed breakdown of the template costs, click **Show changes**.

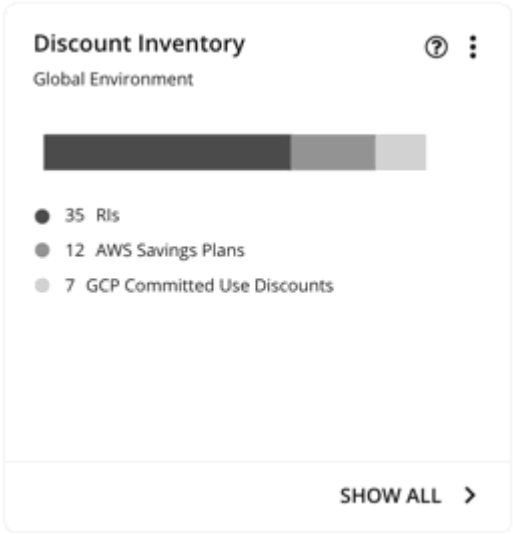
Volume tier summary

This chart shows the current distribution of volumes that support your workloads, and the optimized distribution if you execute the actions that the plan recommends.

The difference in the result reflects the number of unattached volumes. To see a list of unattached volumes, click **Show changes** at the bottom of the chart.

Discount inventory

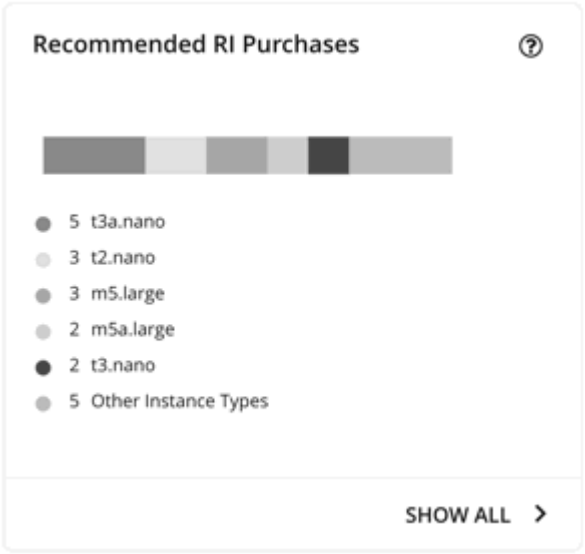
This chart lists the cloud provider discounts discovered in your environment. For a tabular listing, click **Show All** at the bottom of the chart. In the tabular listing, you can see if a discount expired before the specified purchase date.



Recommended RI purchases

Workload Optimization Manager can recommend purchasing instance types at a discounted rate to help you increase the percentage of VMs covered by discounted pricing and reduce on-demand costs. This chart shows your pending purchases. Download the list of purchases and then send it your cloud provider or representative to initiate the purchase process.

> **NOTE:**
> Currently, Workload Optimization Manager can recommend purchasing AWS EC2 RIs and Azure reservations, but only in plans. Purchase actions are not generated in the real-time market and therefore do not show in Action Center.



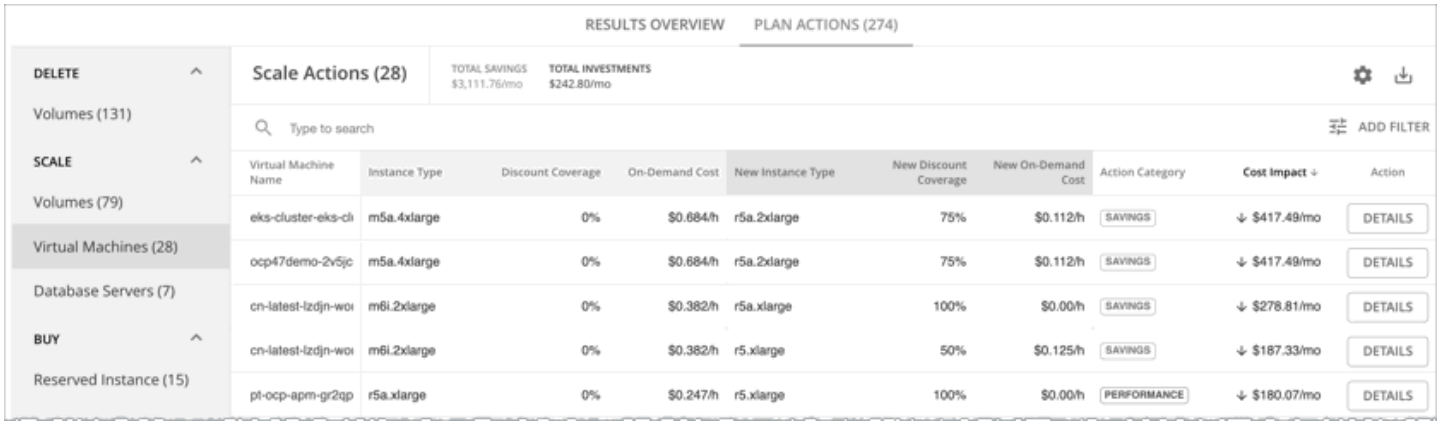
Click **Show All** to see a table with details for each discount.

The table shows the properties, up-front cost, and break-even period for each discount. The break-even period is the time at which savings will exceed the up-front cost, rounded to the month. The Cost Impact column indicates the monthly savings you would realize when you buy a specific discount.

When you choose one or more check boxes, the total count, up-front cost, and savings appear first.

Viewing plan actions

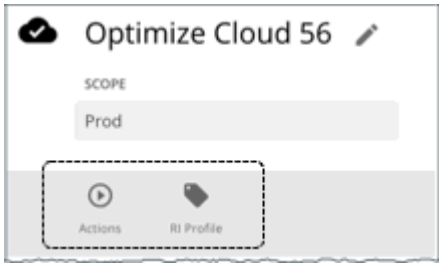
Click the **Plan Actions** tab to view a list of actions that you need to execute to achieve the plan results.



Re-running the plan

You can run the plan again with the same or a different set of configuration settings. This runs the plan scenario against the market in its current state, so the results you see might be different, even if you did not change the configuration settings.

Use the toolbar for the Configuration section to change the configuration settings.



- Actions

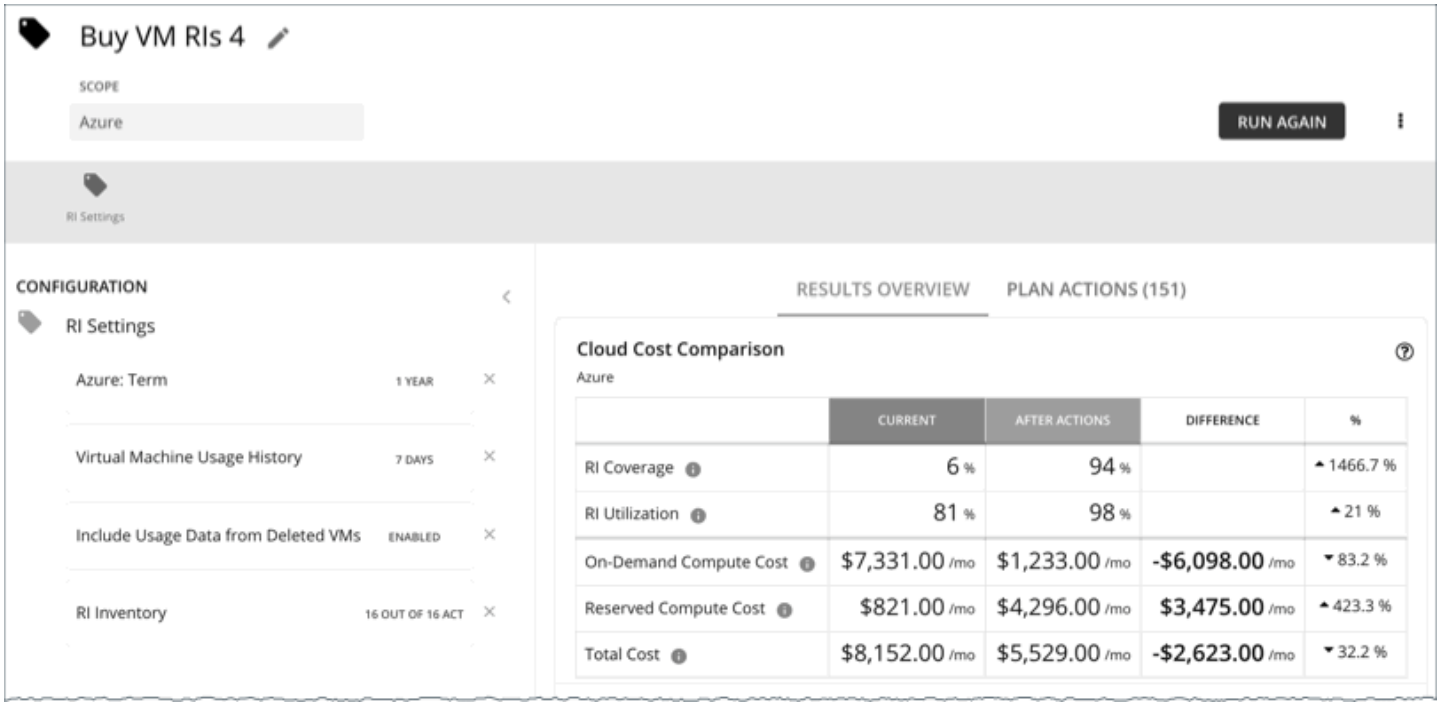
  Use this to enable or disable automatic Scale actions for the virtual machines in the plan.
- Discount settings

  See Discount Settings *(on page 892)*.

NOTE:
It is not possible to change the scope of the plan in the Plan Page. You must start over if you want a different scope. To start over, click the More options icon (⋮), and then select **New Plan**.

When you are ready to re-run the plan, click **Run Again**.

Buy VM reservations plan



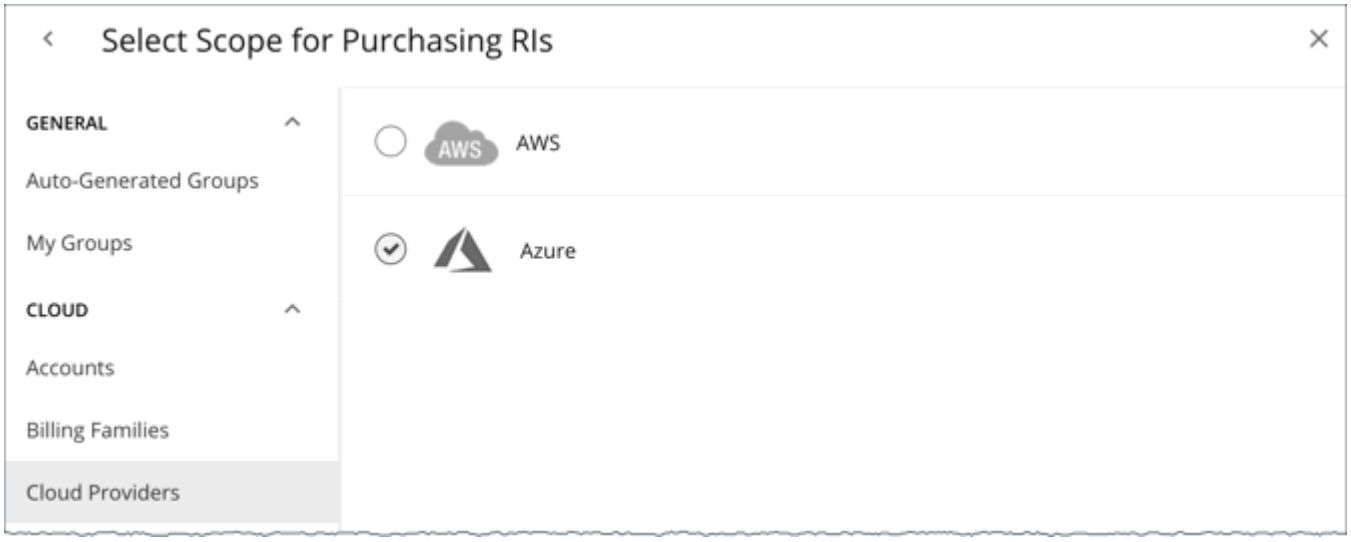
Run the buy VM reservations plan to see discount purchase opportunities that can significantly reduce on-demand costs for your cloud VMs. When calculating purchases, Workload Optimization Manager evaluates all purchasing options for your selected scope and usage data for the VMs in that scope. It then compares your current costs to the costs you would get after executing the plan recommendations.

Currently, Workload Optimization Manager can recommend purchasing AWS EC2 RIs and Azure reservations, but only in plans. Purchase actions are not generated in the real-time market and therefore do not show in Action Center.

Configuring a buy VM reservations plan

For an overview of setting up plan scenarios, see Setting Up Plan Scenarios *(on page 851)*.

Scope



You can scope by:

- Accounts

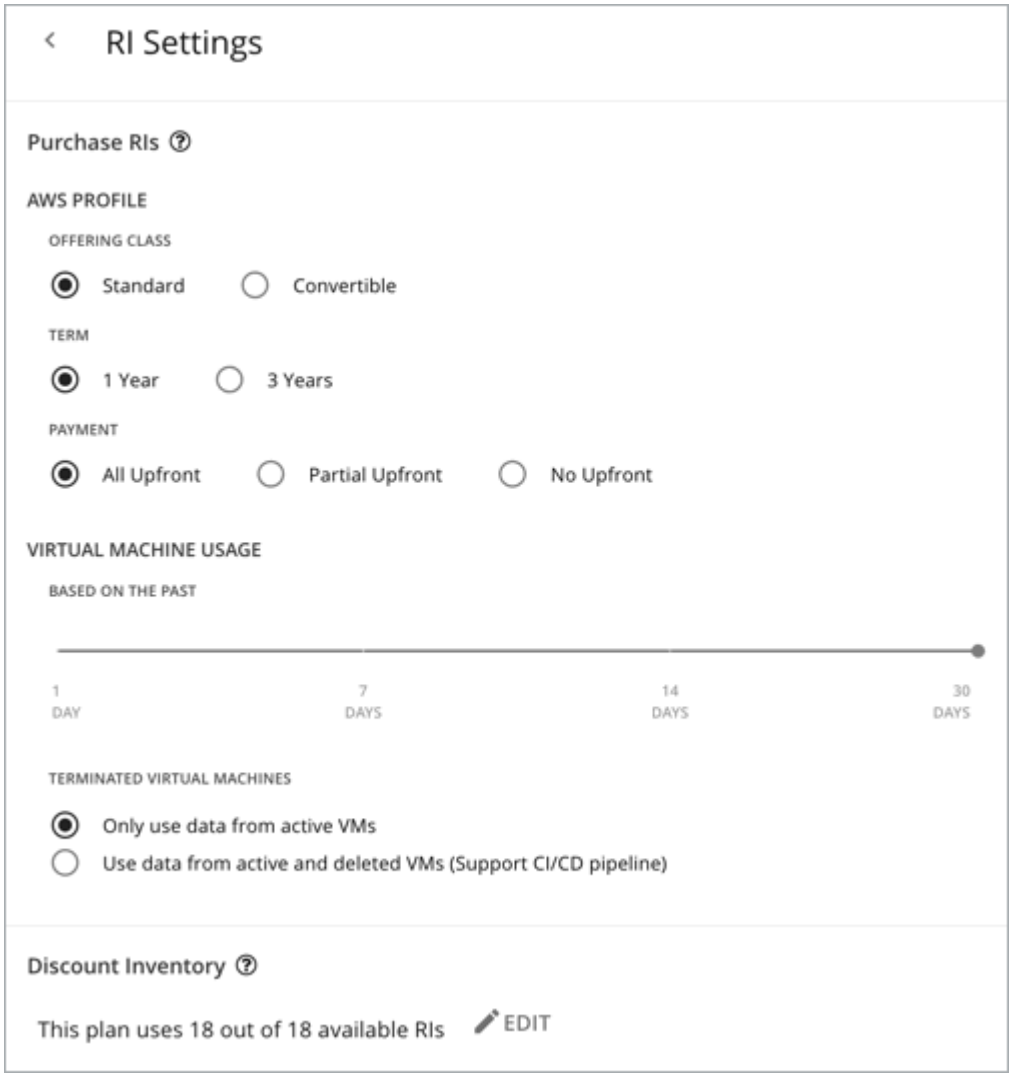
  Choose AWS accounts or Azure subscriptions for the plan's scope.
- Billing families

  Include discount purchases for a billing family. The plan calculates discount purchases through the billing family's primary account.
- Cloud providers

  See purchase opportunities for your AWS or Azure environment.
- Regions

  Focus the plan on a cloud provider's region.

RI settings



Purchase RIs

Allow the plan to buy discounts based on the following configurations:

- Profile

  The settings that you set up for real-time analysis are selected by default. You can change the settings to see how they affect costs.

  - Offering class

    Choose the offering class that corresponds to the RI types that you typically use in your environment.

  - Term

    Choose the payment terms for your discounts. Typically, longer term payment plans cost less per year.

  - Payment

    Choose your payment option.

    - All upfront – You make full payment at the start of the term.
    - Partial upfront – You make a portion of the payment at the start of the term, with the remain cost paid at an hourly rate.
    - No upfront – You pay at an hourly rate for the duration of the term.

- Virtual machine usage

Specify the time frame that you want the plan to use when it calculates your discount purchases.

- Terminated virtual machines
  - Only use data from active VMs

    Select this option if you stop your VMs permanently.
  - Use data from active and deleted VMs (Support CI/CD pipeline)

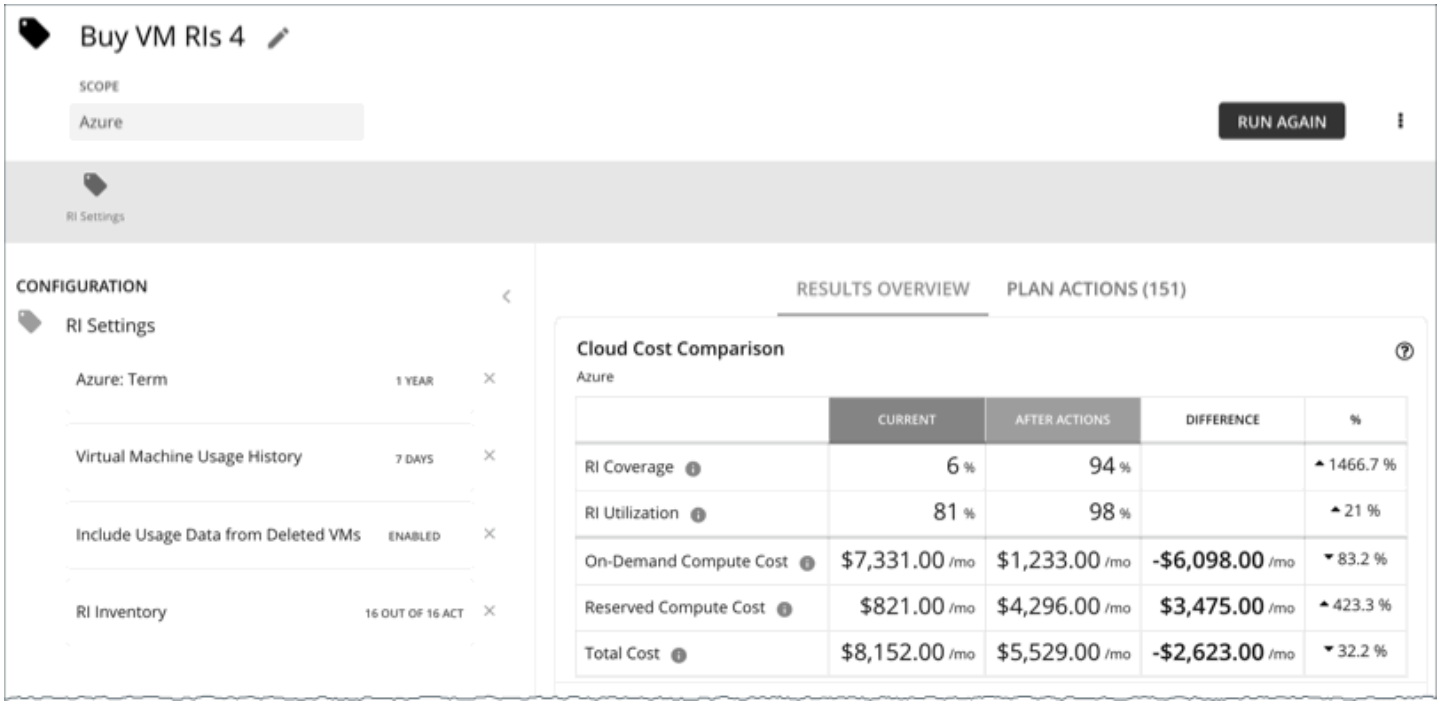
    Select this option if you want to use data from a CI/CD pipeline that regularly deploys and stops VMs.

Discount inventory

Select your discount inventory for the plan. You can use the default selection or any of the available discounts for your scope.

Working with buy VM reservations plan results

After the buy VM reservations runs, you can view the results to see discount and optimization opportunities for your cloud environment.



Viewing the results

The plan results include the following charts.

Cloud cost comparison

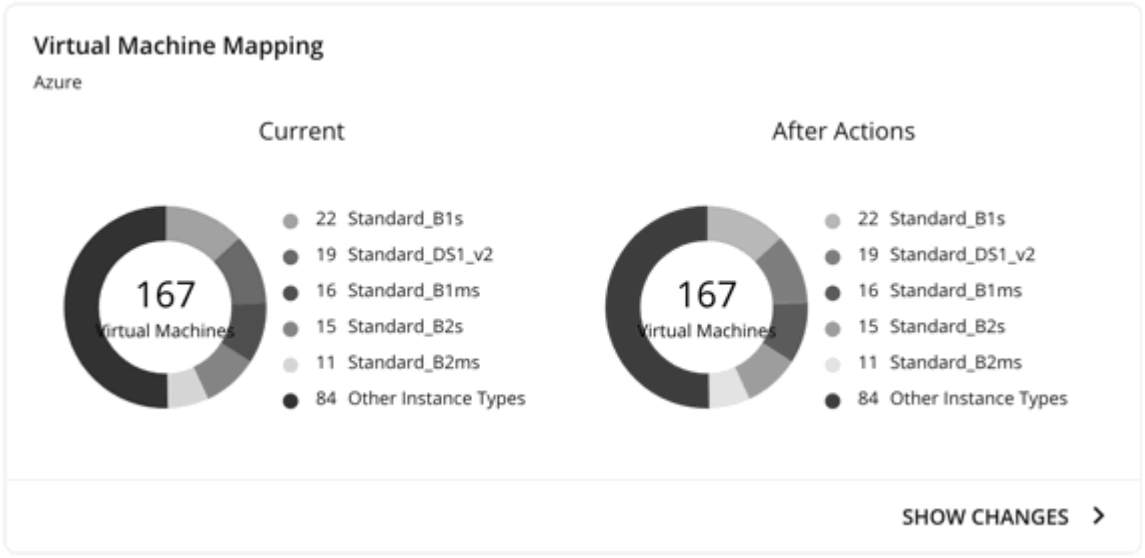
This chart highlights changes to your existing discount coverage and utilization if you execute all the actions that the plan recommends. Actions include increasing coverage or purchasing additional instance types at a discounted rate. Your cloud provider will adjust discount allocations when the actions have completed.

- Analysis evaluates ways to increase your current discount coverage so you can take full advantage of discounted pricing.
- The plan can recommend purchase actions to reduce your costs further. The analysis looks at historical VM usage and uptime to arrive at the number of instance types you should purchase.

You can compare current and after-action costs, including on-demand compute, discounted compute, and total costs. Purchase actions increase your discounted compute cost, but can lower your on-demand compute cost significantly as discount coverage increases. The end result is a reduction to your total cost.

Virtual machine mapping

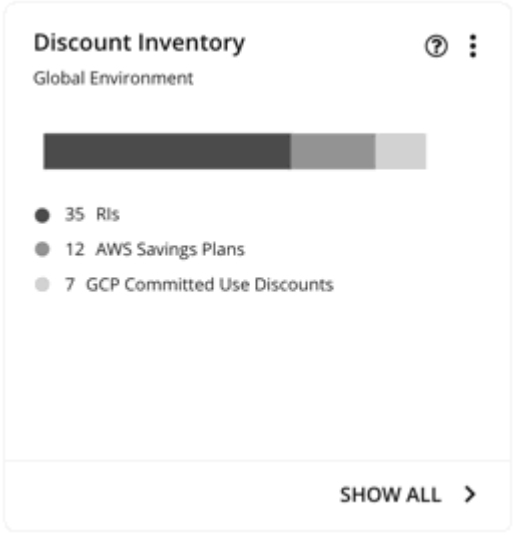
This chart shows the instance types for the VMs included in the plan.



Click **Show Changes** to see details for each VM with discount coverage changes. The table maps VMs to instance types, and shows how changes in discount coverage can reduce on-demand cost.

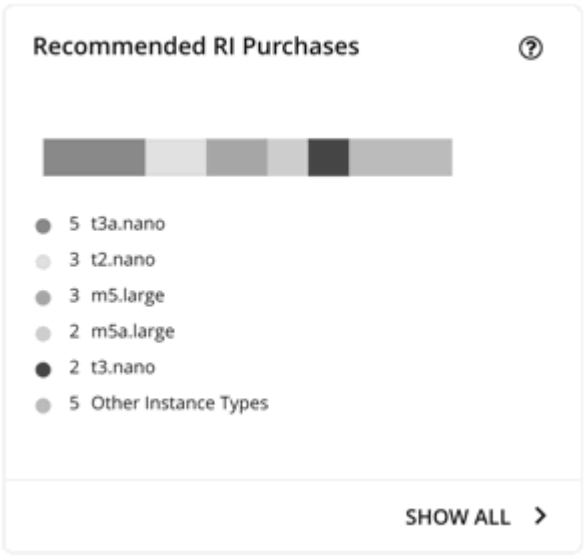
Discount inventory

This chart lists the cloud provider discounts discovered in your environment. For a tabular listing, click **Show All** at the bottom of the chart. In the tabular listing, you can see if a discount expired before the specified purchase date.



Recommended RI purchases

Workload Optimization Manager can recommend purchasing instance types at a discounted rate to help you increase the percentage of VMs covered by discounted pricing and reduce on-demand costs. This chart shows your pending purchases. Download the list of purchases and then send it your cloud provider or representative to initiate the purchase process.



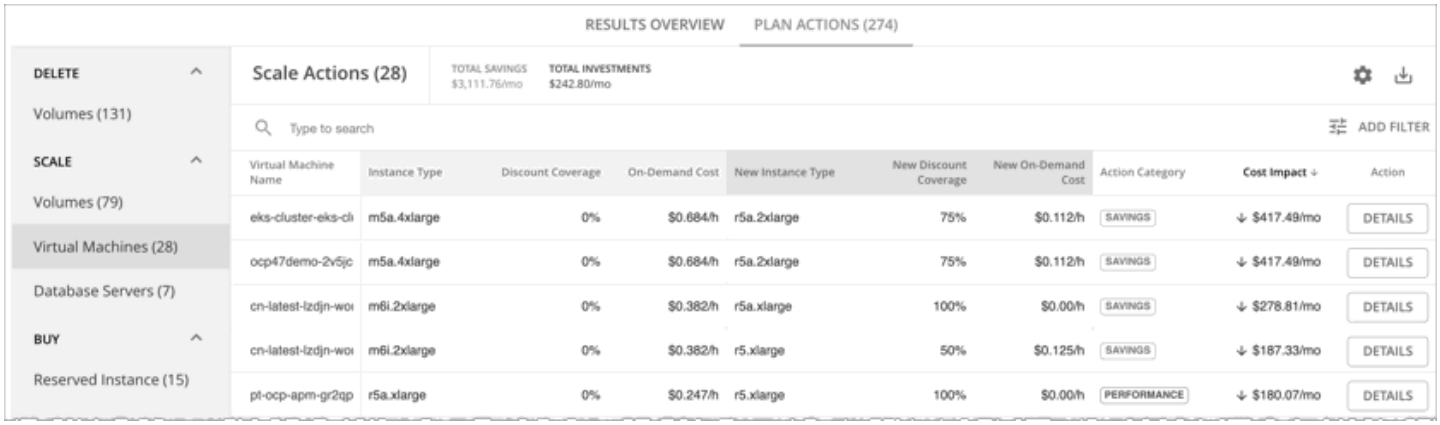
Click **Show All** to see a table with details for each discount.

The table shows the properties, up-front cost, and break-even period for each discount. The break-even period is the time at which savings will exceed the up-front cost, rounded to the month. The Cost Impact column indicates the monthly savings you would realize when you buy a specific discount.

When you choose one or more check boxes, the total count, up-front cost, and savings appear first.

Viewing plan actions

Click the **Plan Actions** tab to view a list of actions that you need to execute to achieve the plan results.



Re-running the plan

You can run the plan again with the same or a different set of configuration settings. This runs the plan scenario against the market in its current state, so the results you see might be different, even if you did not change the configuration settings.

Use the toolbar for the Configuration section to change the configuration settings.

- RI Settings

  Update your purchase settings to see how they impact results. For example, you can configure a longer timeframe so that the plan can include additional VM usage data in its analysis. For details, see Purchase RIs *(on page 899)*.

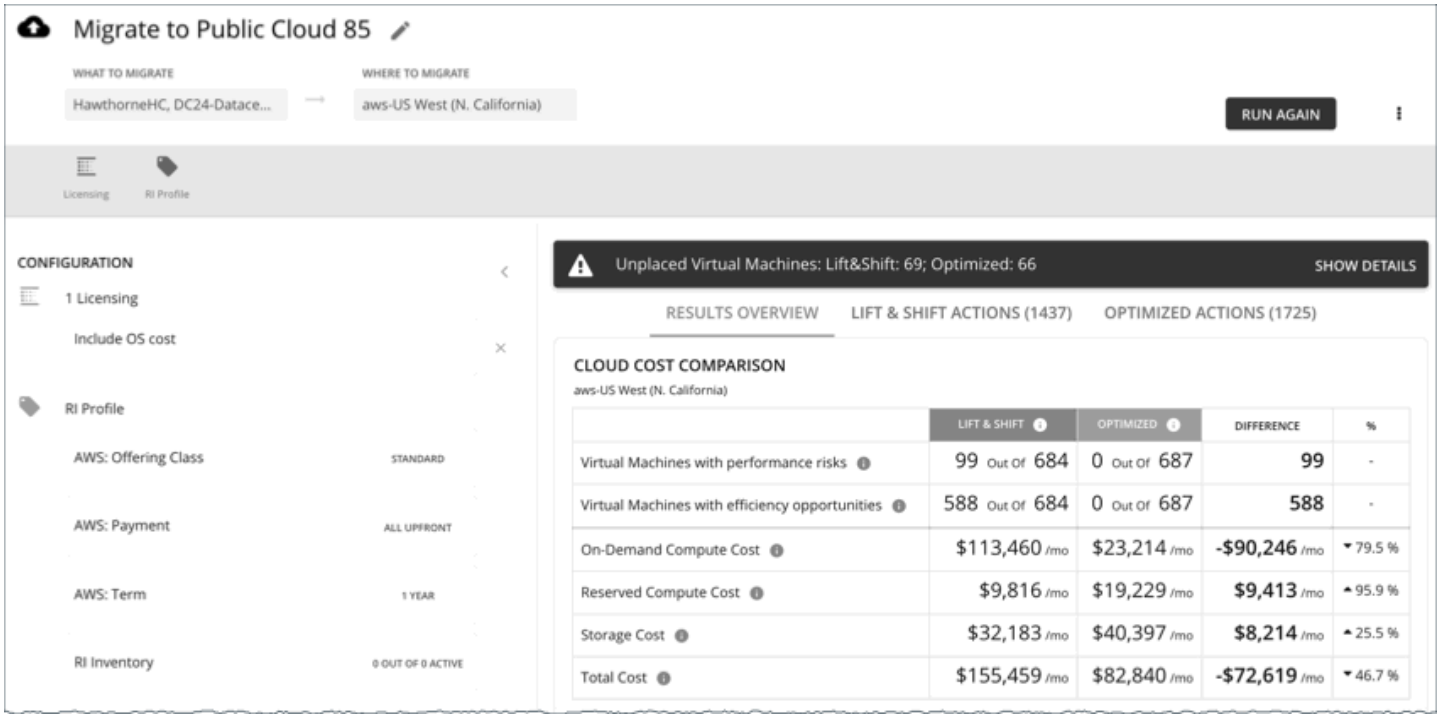
- Discount Inventory

  Use the default selection or any of the available discounts for your scope.

NOTE:
It is not possible to change the scope of the plan in the Plan Page. You must start over if you want a different scope. To start over, click the More options icon ( ), and then select **New Plan**.

When you are ready to re-run the plan, click **Run Again**.

Migrate to cloud plan



A Migrate to Cloud plan simulates migration of on-prem VMs to the cloud, or migration of VMs from one cloud provider to another. This plan focuses on optimizing performance and costs by choosing the most suitable cloud resources for your VMs and the volumes they use. To further optimize your costs, the plan can recommend moving workloads from on-demand to discounted pricing, and purchasing more discounts.

The plan calculates costs according to the billing and price adjustments that you have negotiated with your cloud provider. Costs include compute, service (such as IP services), and license costs. The plan also calculates discount purchases for VMs that can benefit from discounted pricing.

NOTE:
If your instance of Workload Optimization Manager is inoperative for a period of time, that can affect the cost calculations. To calculate costs for a VM that it will migrate to the cloud, Workload Optimization Manager considers the VM's history. For example, if the VM has been stable for 16 of the last 21 days, then Workload Optimization Manager will plan for that VM to use a discount. In this way, the plan calculates the best cost for the migration. However, if Workload Optimization Manager is inoperative for any time, that can impact the historical data such that the plan will *not* recognize a VM as stable, even though it is.

Points to consider:

- AWS includes EC2 Spot Instances that offer steep discounts. Plans that migrate from AWS to Azure do not migrate VMs that run on Spot Instances.
- Currently, Workload Optimization Manager does not support plans that migrate a VM running an AWS EC2 GPU *(on page 638)* or Azure GPU *(on page 643)* instance type to another cloud provider. For example, a plan does not migrate an Azure GPU VM to an AWS region.
- Do not use this plan type to migrate within the same cloud provider (for example, moving VMs from one Azure subscription to another) as a way to test the effect on pricing. The results from such a plan would not be reliable.
- For migrations within your on-prem environment, use the *Virtual Machine Migration* plan type.
- Before migrating, consider turning on a setting *(on page 1052)* in the default global policy that enables metrics collection for on-prem volumes attached to VMs. This allows Workload Optimization Manager to make more accurate placement decisions for the VMs and volumes you are migrating.

The plan results show:

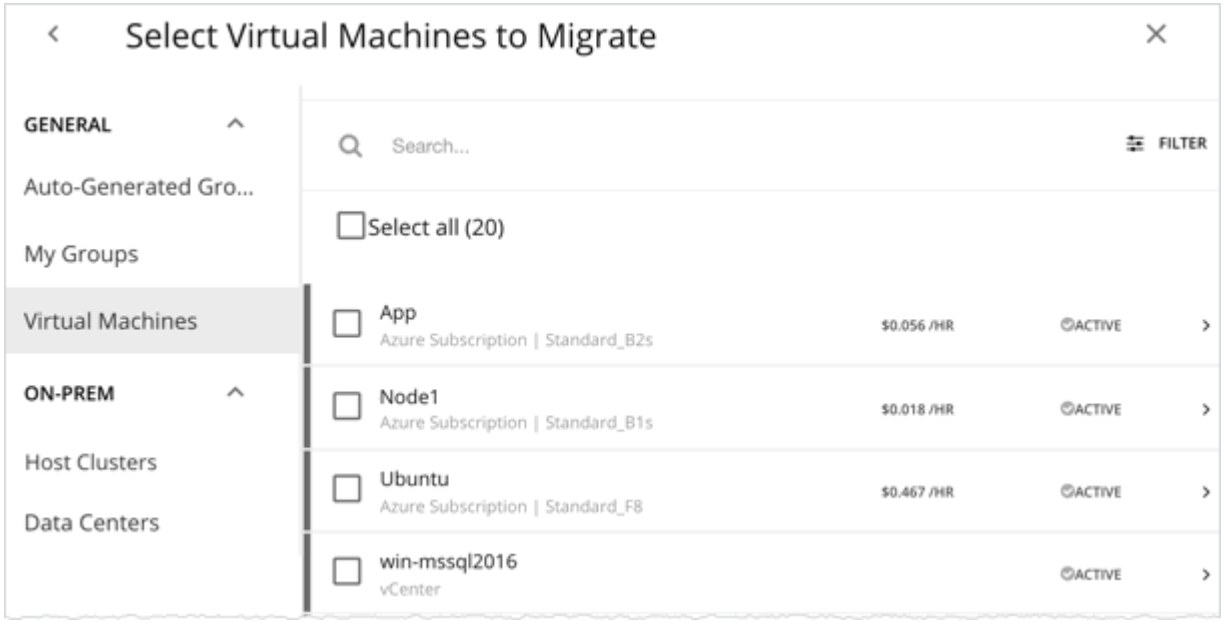
- Projected costs
- Actions to execute your migration and optimize costs and performance
- Optimal cloud instances to use, combining efficient purchase of resources with assured application performance
- The cost benefit of moving workloads from on-demand to discounted pricing
- Discounts you should purchase

Configuring a migrate to cloud plan

For an overview of setting up plan scenarios, see Setting Up Plan Scenarios *(on page 851)*.

Scope

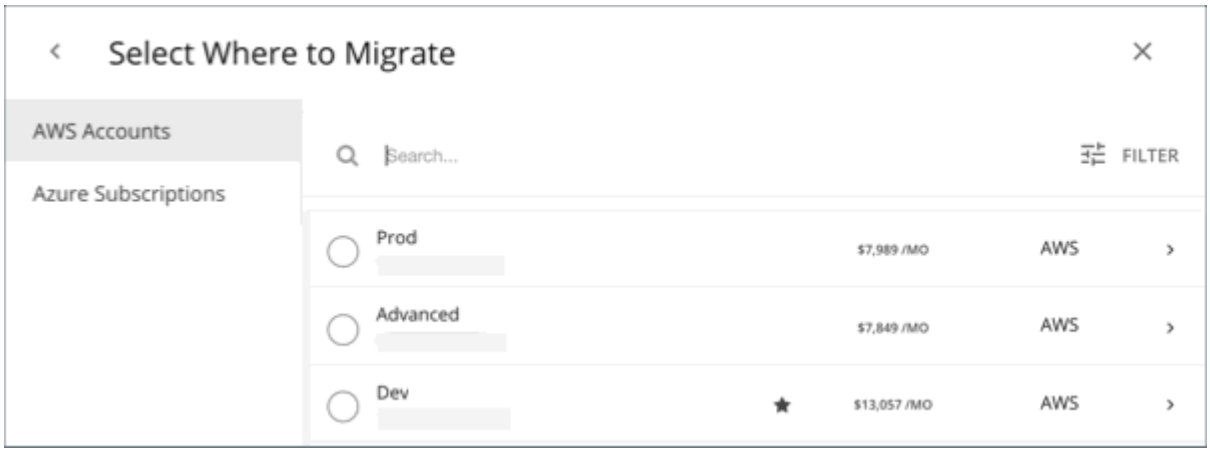
Select the VMs that you want to migrate. You can select VM groups, individual VMs, or both.



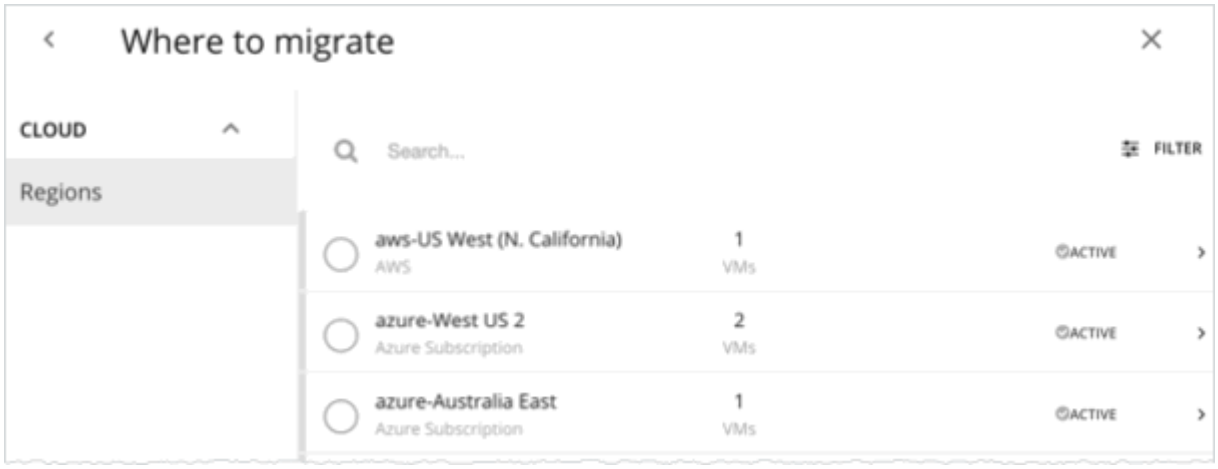
If you select an Auto Scaling Group, Workload Optimization Manager simulates migrating the VMs individually, and not as a group.

Where to migrate

Choose a billing account (AWS account or Azure subscription).



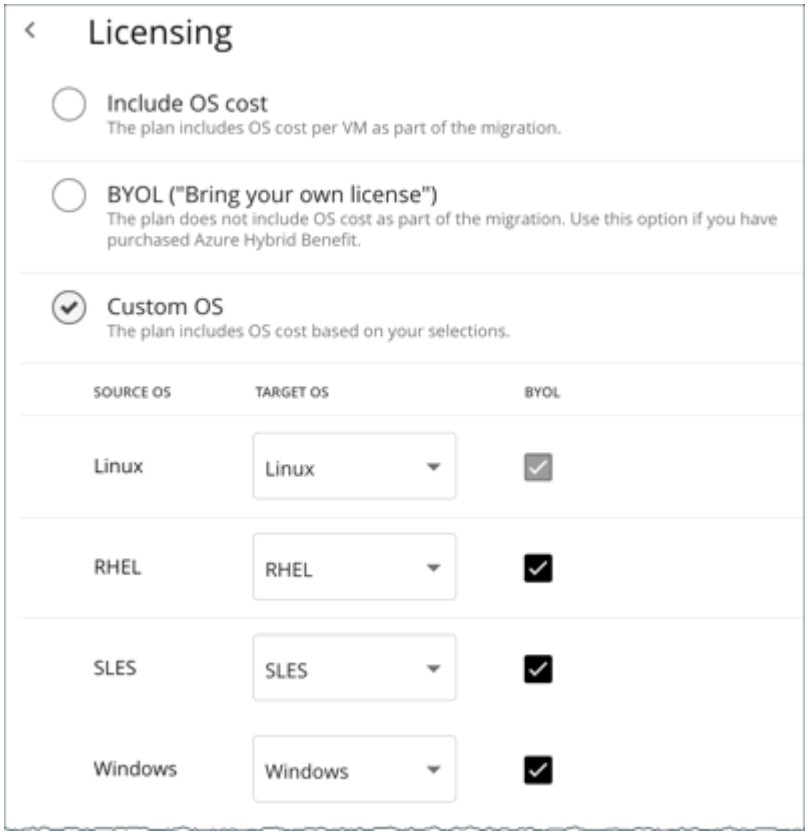
Choose a region. Workload Optimization Manager shows all the regions that you can access from your target cloud accounts.



By default, Workload Optimization Manager considers all instance types in the selected region when making placement decisions for the scoped VMs and the volumes they use. However, you may have set up constraints in policies that limit migration to certain instance types. If there are VMs and volumes in your scope that are affected by those policies, Workload Optimization Manager will only consider the instance types that are defined in the policies.

Licensing (OS migration profile)

Select an OS Profile for this migration.



On the cloud, instances usually include an OS platform to run processes on the VM. As you migrate VMs to the cloud, you can specify the OS you prefer to run. You can keep the same OS that the original VM has, or map it to a different OS.

- Include OS cost

  As Workload Optimization Manager calculates placement for the migrated workloads, it will include costs for instances that provide the same OS that the VM already has.

- BYOL (Bring your own license)

  This is the same as the **Include OS cost** option, except the plan does not include OS licensing costs in any of the cost calculations for on-cloud placement.

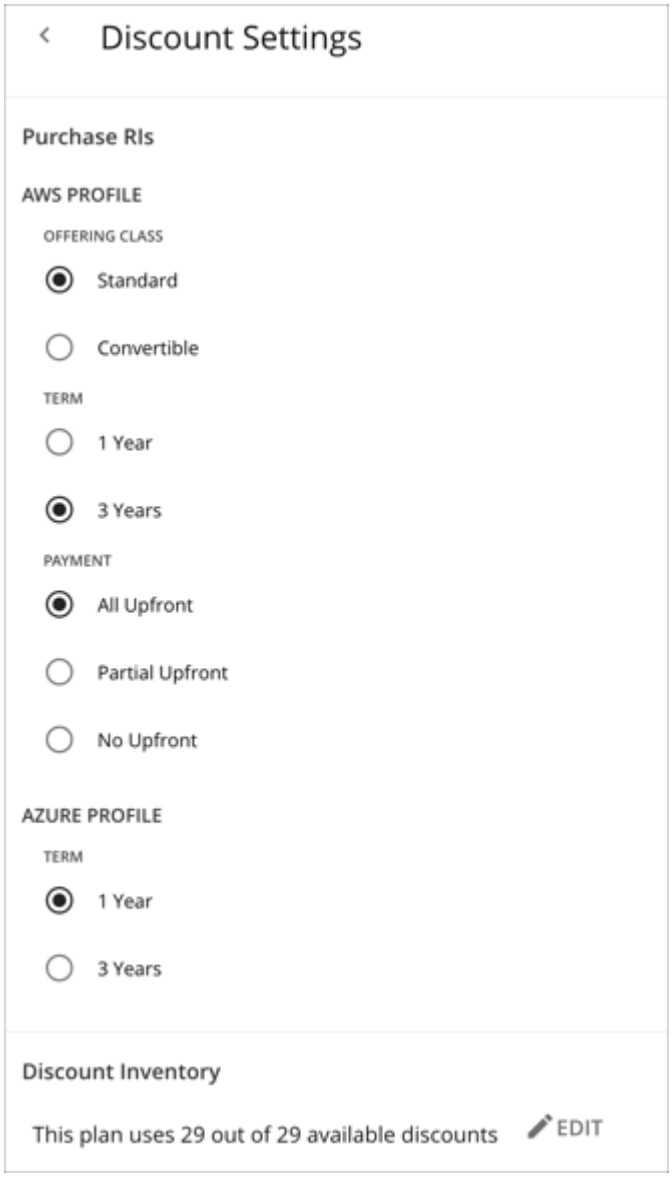
- Custom OS

  For each of the listed OS types, map the migrated VM to the OS you choose. The OS types are:

  - Linux – Any open source distribution of Linux. For the migration, Workload Optimization Manager will choose instances that provide the Linux platform that the cloud service provider delivers as a free platform. Note that this is always BYOL, because it assumes a free OS license.
  - RHEL (Red Hat Enterprise Linux)
  - SLES (SUSE Linux Enterprise Server)
  - Windows

  If you enable **BYOL** for RHEL, SLES, or Windows, Workload Optimization Manager assumes that you are paying for the OS license, and will not include the license cost in the plan results. If you do not enable **BYOL**, Workload Optimization Manager gets the license cost from the service provider and includes that cost in the plan results.

Reserved instances settings



- Currently, **Purchase RIs** applies to only AWS and Azure. For **AWS/Azure Profile**, the settings that you have set up for real-time analysis are selected by default. You can change the settings to see how they affect costs.
  - Offering class

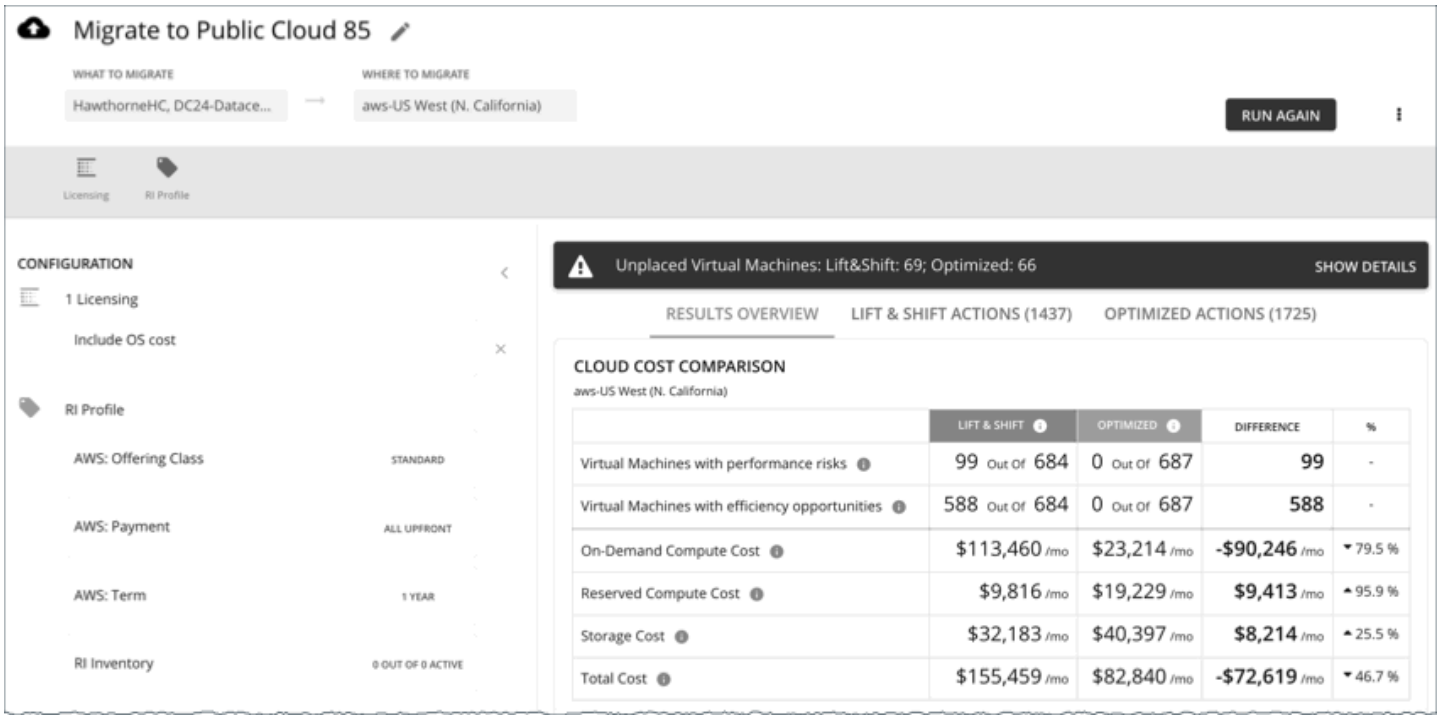
    Choose the offering class that corresponds to the RI types that you typically use in your environment.
  - Term

    Choose the payment terms for your discounts. Typically, longer term payment plans cost less per year.
  - Payment

    Choose your payment option.
    - All upfront – You make full payment at the start of the term.
    - Partial upfront – You make a portion of the payment at the start of the term, with the remain cost paid at an hourly rate.
    - No upfront – You pay at an hourly rate for the duration of the term.
- For **Discount Inventory**, the discounts for the current scope are selected by default. Click **Edit** to make changes.

Working with migrate to cloud plan results

The migrate to cloud plan results show the cloud resources and costs for the VMs you plan to migrate, and the actions required for migration.



Workload Optimization Manager shows results for two migration scenarios:

- Lift & shift

  Lift & Shift migrates your VMs to cloud instances that match their current resource allocations.

- Optimized

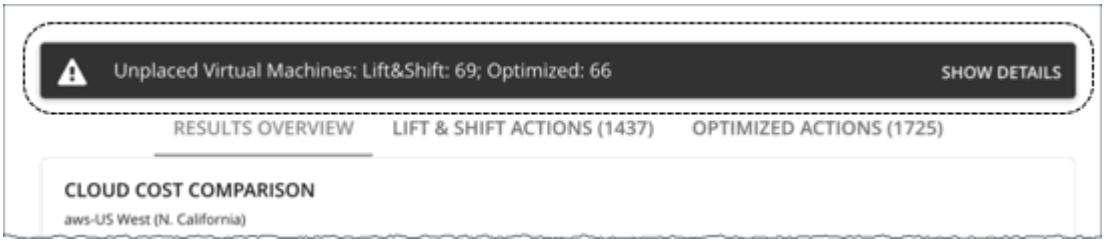
  As Workload Optimization Manager runs the plan, it looks for opportunities to optimize cost and performance. For example, it might discover overprovisioned VMs after analyzing the historical utilization of VM resources. If you were to migrate such VMs to instances that match their current allocations, then you would spend more than necessary. For an optimized migration, Workload Optimization Manager can recommend migrating to less expensive instances while still assuring performance, and then show the resulting savings. In addition, when you examine the actions for an optimized migration, you will see charts that plot the historical utilization data used in the analysis.

Results overview

The Results Overview section shows the following:

- Unplaced VMs

  If the plan's scope includes VMs that cannot be migrated, the results include a notification indicating the number of VMs. Click **Show Details** to see the list of VMs and the reasons for their non-placement.

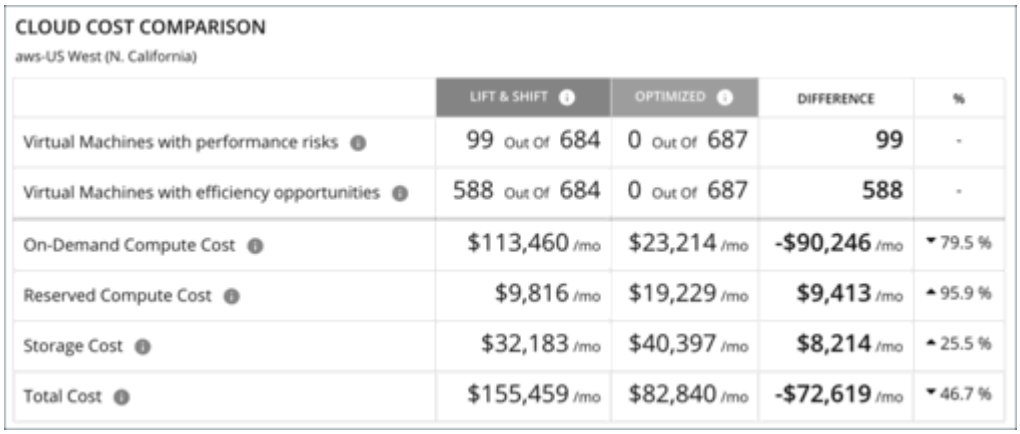
  

  The charts in the plan results do not count these VMs.

Workload Optimization Manager displays adjusted CPU values for unplaced VMs. These values are the actual metrics used in analysis and are calculated using benchmark data. CPU values shown in other places (such as the Capacity and Usage chart) are unadjusted values obtained from targets.

- Cloud cost comparison chart

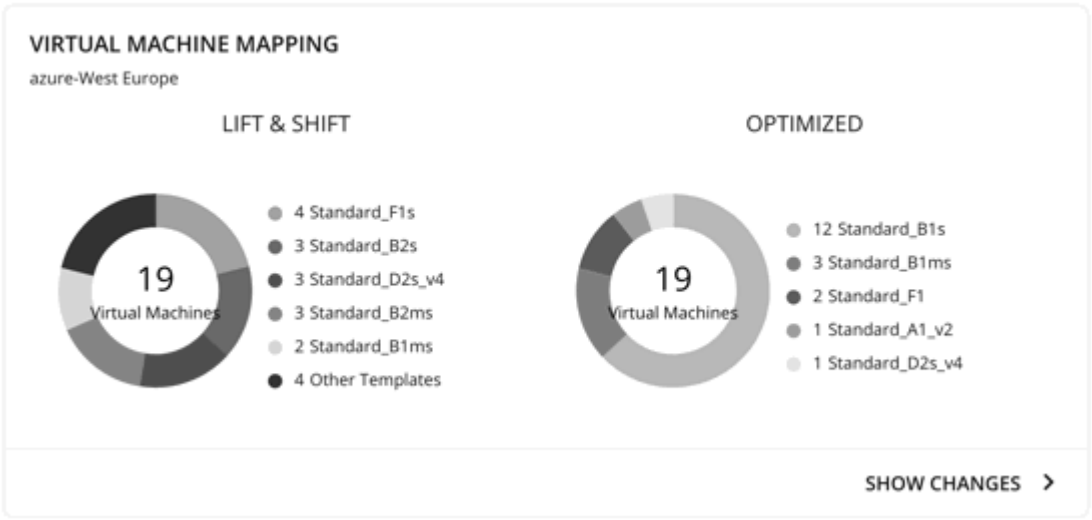
  This chart highlights any difference in cost as a result of optimization. For example, undersized VMs risk losing performance and should therefore scale up. This could contribute to an increase in cost. On the other hand, oversized VMs can scale down to less expensive instances, so cost should go down. The values under the **%** column indicate the percentage of VMs that are affected by optimization cost calculations.



NOTE:
For Azure, the results do not include the license cost for the migrated VMs.

- Virtual machine mapping chart

  This chart gives a breakdown of the instance types that the plan recommends for the migration, including how many of each is needed.



  Click **Show Changes** to see a table with details for each VM in the plan. The table maps VMs to instance types. It also shows the properties and monthly cost for each instance type. Under the **Actions** column, click **Details** to compare Lift & Shift and Optimized actions.

- Volume tier summary chart

  This chart gives a breakdown of the volume types that the plan recommends for the migration, including how many of each is needed.



Click **Show Changes** to see a table with details for each volume in the plan. The table maps the volumes you plan to migrate to the volume types that Workload Optimization Manager recommends. It also shows the properties and monthly cost for each volume type. Under the **Actions** column, click **Details** to compare Lift & Shift and Optimized actions.

- Recommended RI purchases charts

  Workload Optimization Manager can recommend purchasing instance types at a discounted rate to help you increase the percentage of VMs covered by discounted pricing and reduce on-demand costs. This chart shows your pending purchases. Download the list of purchases and then send it your cloud provider or representative to initiate the purchase process.

  NOTE:
  Currently, Workload Optimization Manager can recommend purchasing AWS EC2 RIs and Azure reservations, but only in plans. Purchase actions are not generated in the real-time market and therefore do not show in Action Center.



To identify VMs that are good candidates for discounted pricing, Workload Optimization Manager analysis considers the history of a VM (by default, the last 21 days), and it looks for:

  - Activity

    If the VM's VCPU utilization percentile is 20% or higher, then Workload Optimization Manager considers it an active VM.

- Stability

  If there have been no start, stop, or resize actions for the VM for 16 of the last 21 days, then Workload Optimization Manager considers it stable.

If the current discount inventory cannot support the VM, or if supporting it would exceed your desired coverage, then Workload Optimization Manager can recommend purchasing additional discounts.

Click **Show All** to see a table with details for each discount.

The table shows the properties, up-front cost, and break-even period for each discount. The break-even period is the time at which savings will exceed the up-front cost, rounded to the month. The Cost Impact column indicates the monthly savings you would realize when you buy a specific discount.

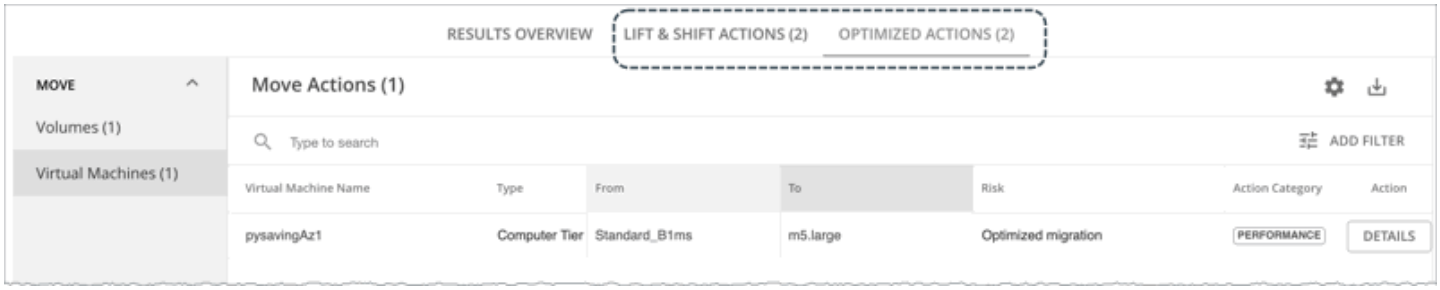
When you choose one or more check boxes, the total count, up-front cost, and savings appear first.

Click **Details** under the **Actions** column to compare Lift & Shift and Optimized actions.

NOTE:
The plan assumes that a discount will always be less expensive than its on-demand counterpart. However, this is not always the case. There might be billing details from service providers that could lead to recommendations to move to a discounted instance type that is more expensive than running on demand.

Plan actions

Workload Optimization Manager shows separate tabs for **Lift & Shift** and **Optimized** migration actions. You can download the list of actions as a CSV file.



For *Optimized* migrations, when you expand an action on a VM, you will see charts that track VCPU and VMem utilization for that VM. With these charts, you can easily recognize the utilization trends that Workload Optimization Manager analyzed to determine the most efficient instance for the VM.



For more information about these charts, see Utilization Charts *(on page 826)*.

Uploading plan results to Azure Migrate

Workload Optimization Manager can upload the plan results and additional plan information to the Azure Migrate portal as part of your migration process. This feature is only available for plans that simulate on-prem VM migration to an Azure region.

Uploaded information includes:

- Basic information for the on-prem VMs, including OS Name and Machine Name
- Target Azure region, VM size, and storage type

  NOTE:
  Azure Migrate does not support automatic selection of OS Disk or manual selection of Ultra Storage disk tiers as part of a migration plan.

- Discount recommendations
- OS license recommendations (based on the licensing option that you selected for the plan)

NOTE:
The Azure Migrate portal displays standardized information provided by third-party migration assessment solutions, including Workload Optimization Manager. Microsoft might not support displaying some information unique to Workload Optimization Manager.

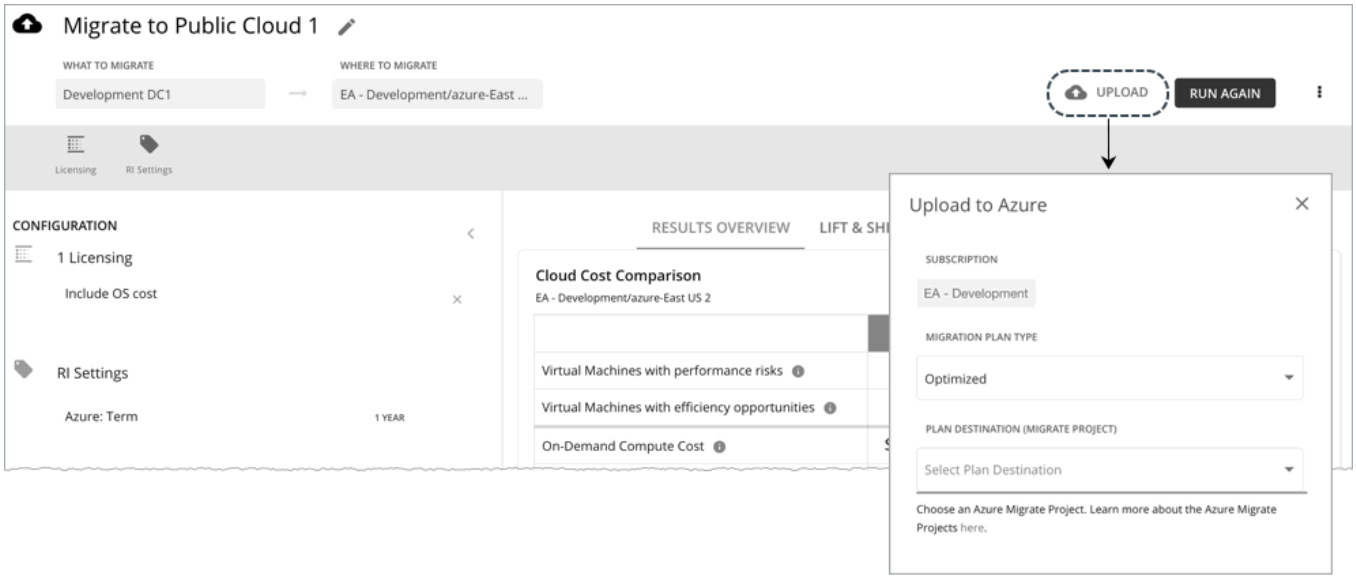
Before uploading the results, be sure to complete the following tasks:

1. Create a project in the Azure Migrate portal.
2. Add Workload Optimization Manager as a migration assessment solution to the project.
3. Set the necessary permissions in the Azure Migrate portal. For details about permissions, see this topic *(on page 462)*.

Consult the Azure documentation for information on completing these tasks.

When you are ready to upload:

1. Click **Upload** on the Plan Page.



2. Specify the following:
   - Migration plan type

     Choose to migrate either the **Lift & Shift** or **Optimized** results.
   - Plan destination (Migrate project)

     Select from the list of Azure Migrate projects. These are the projects belonging to the Azure subscription that you selected for the plan. If you have not created a project for the subscription, go to the Azure Migrate portal and create one.

     WARNING:
     Uploading to a project with existing plan results overwrites those results.

     The upload will fail if another upload targeted at the same destination is already in progress.

3. Click **Submit**.

   The Plan Page updates to display the upload status. Refresh the page periodically to check:
   - If the upload task completed without problems
   - Any upload issues for individual entities
4. When the upload is complete, log in to the Azure Migrate portal and go to the project you selected as the plan destination.

   The project should now display the uploaded information. Use the migration tools identified for the project to start the actual migration.

NOTE:
Repeat the upload procedure if you re-ran the plan and want to upload the new results.

Re-running the plan

You can run the plan again with the same or a different set of configuration settings. This runs the plan scenario against the market in its current state, so the results you see might be different, even if you did not change the configuration settings.

Use the toolbar for the Configuration section to change the configuration settings.

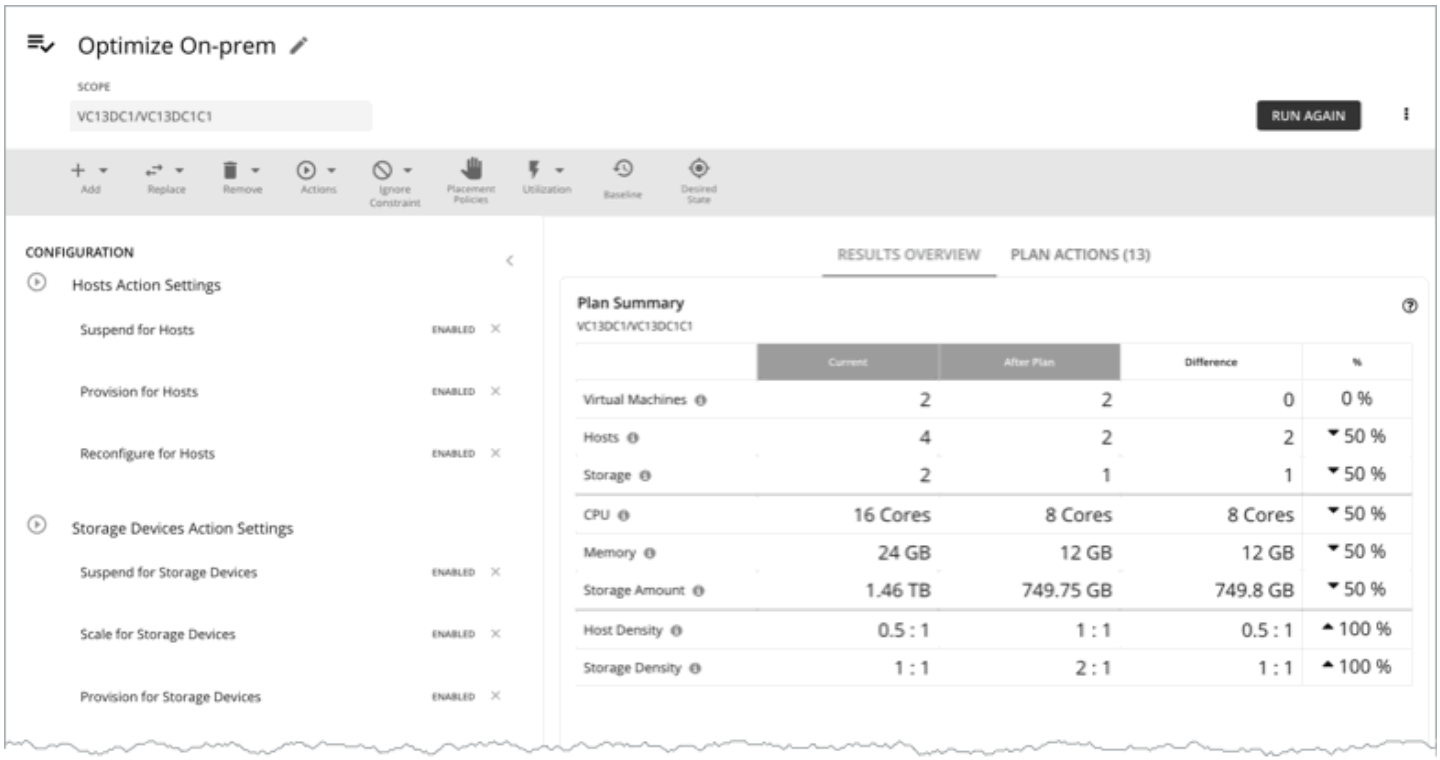
NOTE:
It is not possible to change the scope of the plan in the Plan Page. You must start over if you want a different scope. To start over, click the More options icon ( ), and then select **New Plan**.

When you are ready to re-run the plan, click **Run Again**.

Optimize on-prem plan

Run the Optimize On-prem plan to see the effects of executing certain actions, such as scaling virtual machines, suspending hosts, or provisioning storage, to your on-prem environment.

For an overview of setting up plan scenarios, see Setting Up Plan Scenarios *(on page 851)*.

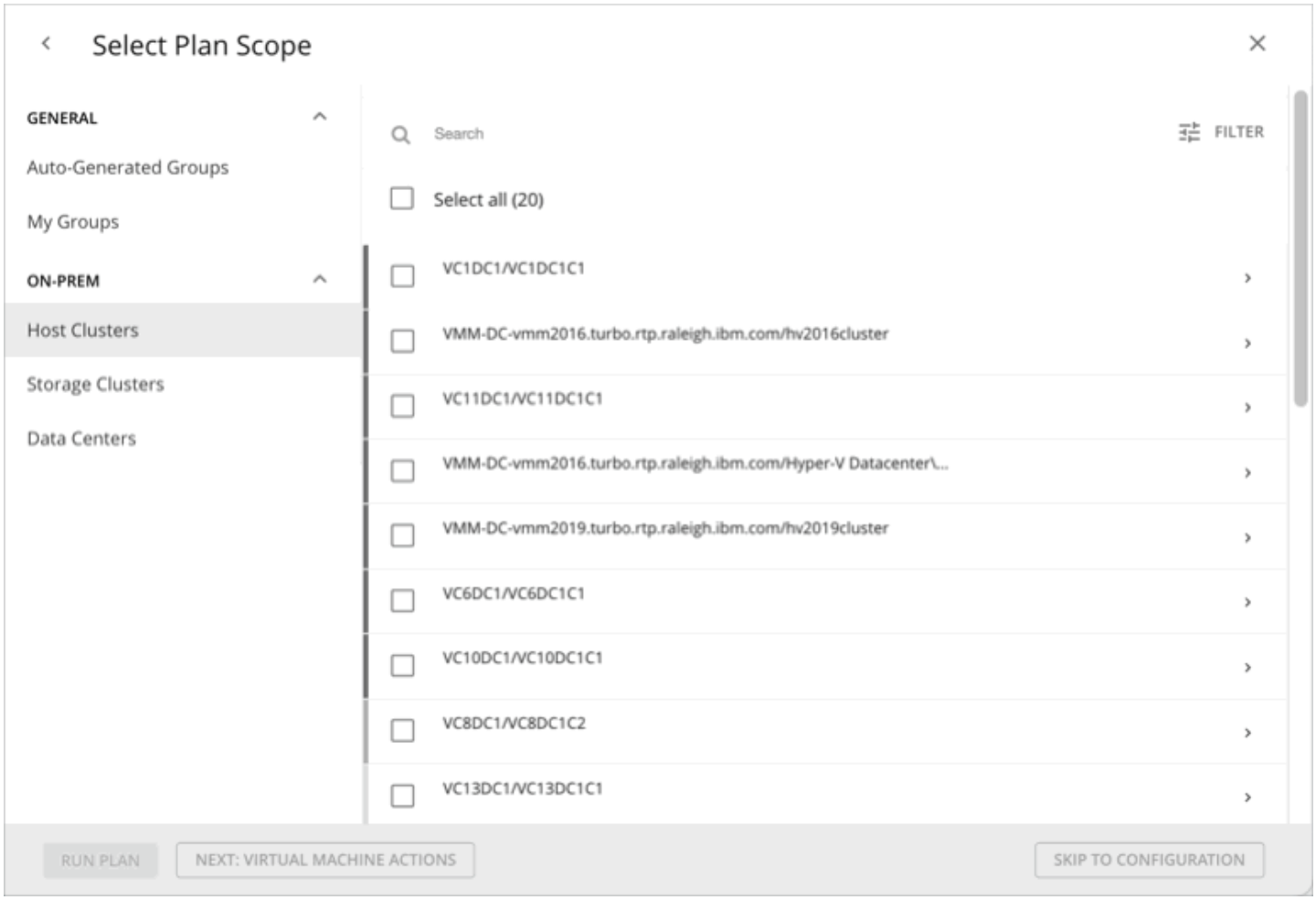


Configuring an optimize on-prem plan

For an overview of setting up plan scenarios, see Setting Up Plan Scenarios *(on page 851)*.

Scope

If you set the scope to a specific Host Cluster, Storage Cluster, Data Center, or group, you can start any plan. You may need to go through additional steps, depending on your chosen plan type. For example, if you **Scope** to a cluster and choose the **Add Virtual Machines plan** type, the plan wizard prompts you to select the most suitable templates for the VMs you plan to add to the cluster.

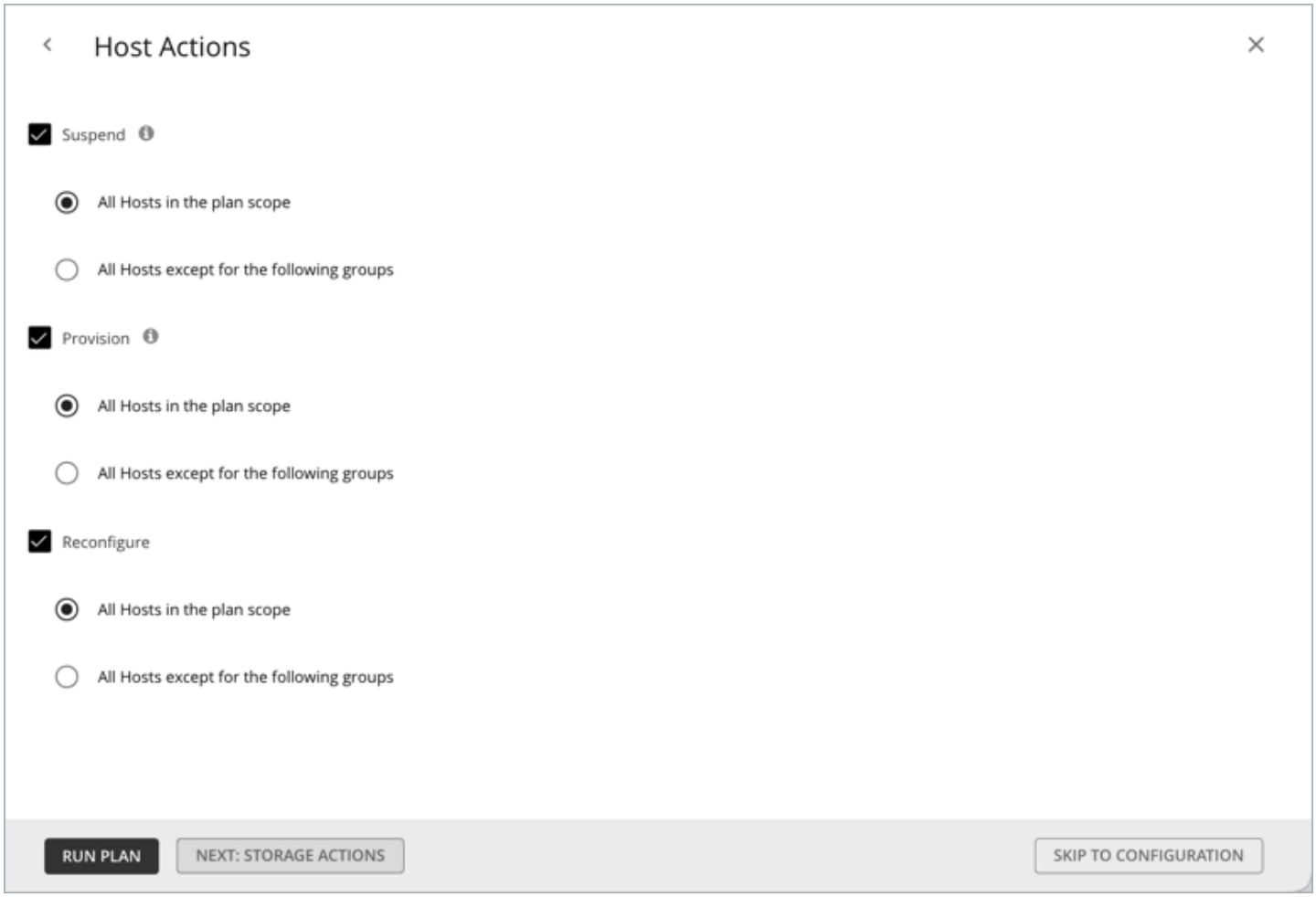


Virtual machine actions

See the effect of enabling or disabling **Scale** actions on the entity included in the plan.

Host actions

See the effect of enabling or disabling **Suspend, Provision**, or **Reconfigure** actions on the entity included in the plan.



- Suspend

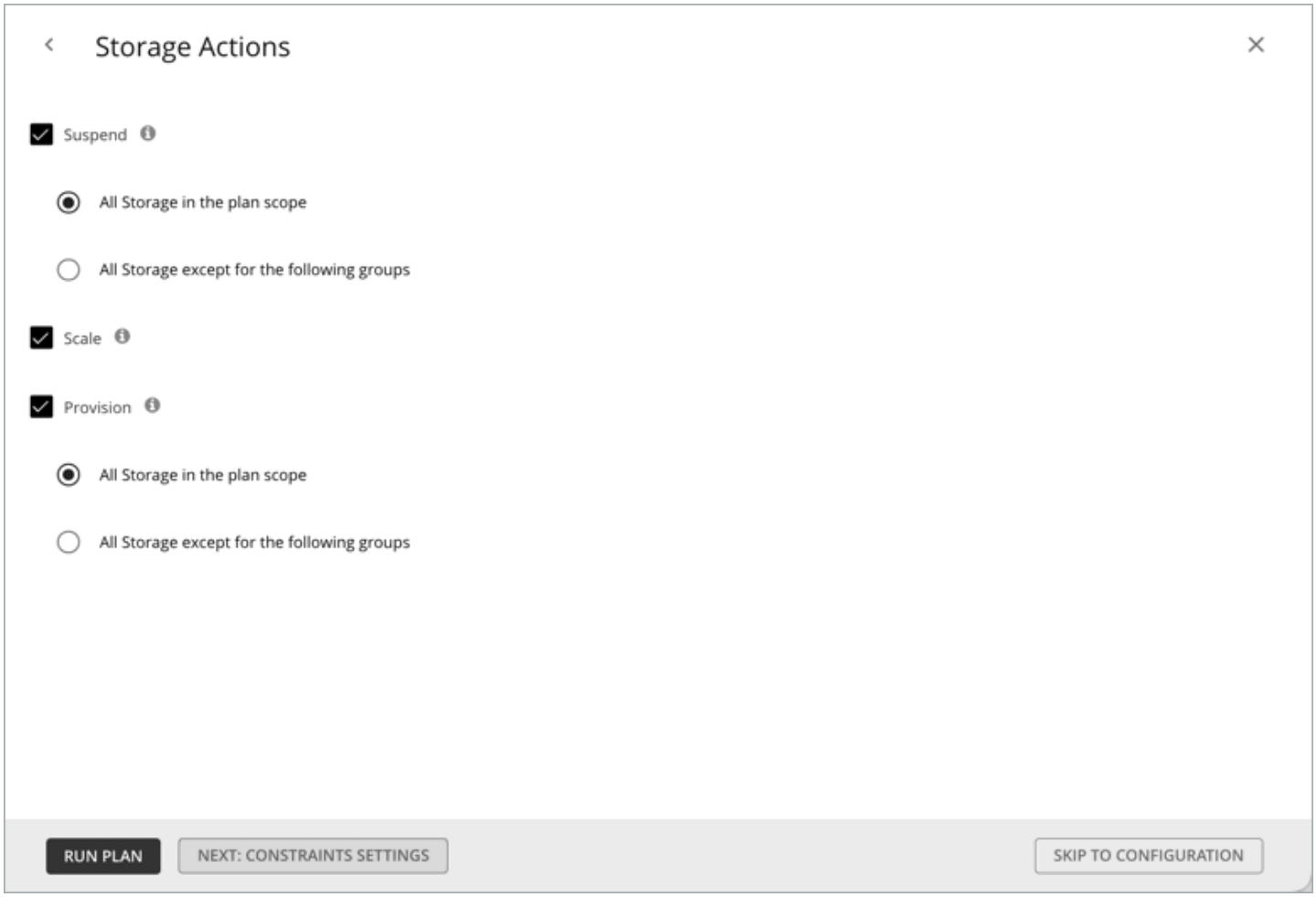
  Workload Optimization Manager removes hosts from your environment if the capacity is not needed to support the workloads in scope.
- Provision

  Workload Optimization Manager adds hosts to your environment if the capacity is not enough to support the workloads in scope. You may need to purchase or provision hardware to meet the plan results.
- Reconfigure

  Workload Optimization Manager generates actions to guide you in changing your hosts. For example, you may need to license a host for a specific application or operating system.

Storage actions

See the effect of enabling or disabling **Suspend, Scale**, or **Provision** actions on the entity included in the plan.



- Suspend

  Workload Optimization Manager removes data stores from your environment if the capacity is not needed to support the workloads in scope.
- Scale

  Workload Optimization Manager modifies your storage device to provide more IOPS or capacity.
- Provision

  Workload Optimization Manager adds storage devices to your environment if the capacity is not enough to support the workloads in scope. You may need to purchase or provision hardware to meet the plan results.

Constraints settings

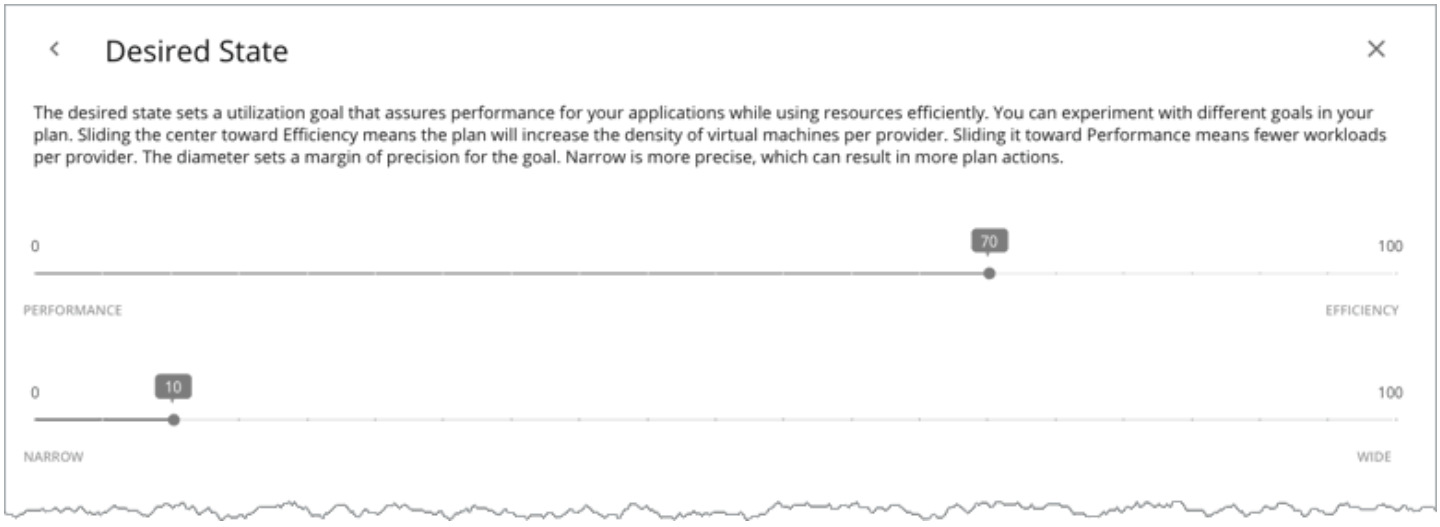
Choose whether to ignore constraints (such as placement policies) for VMs in your environment. By default, VMs are constrained to the cluster, network group, data center, or storage group that their hosts belong to. You can choose to ignore these boundaries.

For example, by default a plan does not consider moving VMs to physical hosts outside of the current cluster. If you disable the Cluster constraint for a VM in your plan, then the plan can evaluate the results of hosting those VMs on any other physical machine within the scope of your plan. If the best results come from moving that VM to a different cluster, then the plan will show that result.

> **NOTE:**
> If you are adding hosts to a plan, and use host templates, then you must turn on **Ignore Constraints**.

Desired state

The desired state sets a utilization goal that assures performance for your applications while using resources efficiently. You can experiment with different goals in your plan. Sliding the center toward Efficiency means the plan will increase the density of virtual machines per provider. Sliding it toward Performance means fewer workloads per provider. The diameter sets a margin of precision for the goal. Narrow is more precise, which can result in more plan actions.



The desired state is a condition in your environment that assures performance for your workloads, while it utilizes your resources as efficiently as possible and you do not overprovision your infrastructure. Workload Optimization Manager uses default Desired State settings to drive its analysis. You should never change the settings for real-time analysis unless you are working directly with Technical support. However, you can change the settings in a plan to see what effect a more or less aggressive configuration would have in your environment.

You can think of the desired state as an n-dimensional sphere that encompasses the fittest conditions your environment can achieve. The multiple dimensions of this sphere are defined by the resource metrics in your environment. Metric dimensions include VMem, storage, CPU, etc. While the metrics on the entities in your environment can be any value, the desired state, this n-dimensional sphere, is the subset of metric values that assures the best performance while achieving the most efficient utilization of resources that is possible.

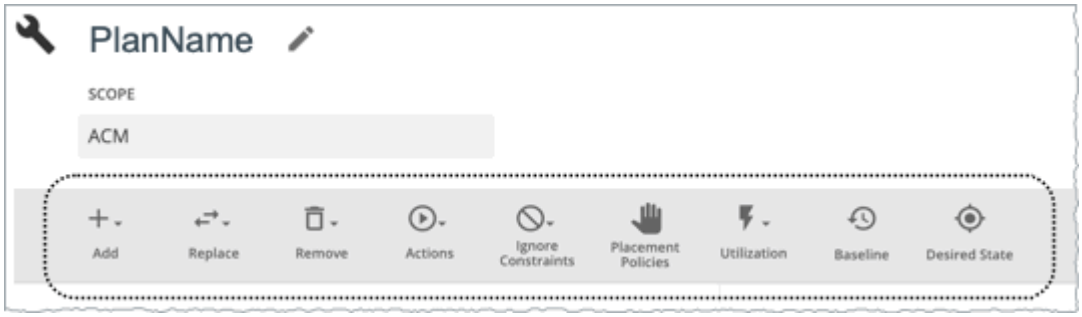
The Desired State settings center this sphere on Performance (more infrastructure to supply the workload demand), or on Efficiency (less investment in infrastructure to supply the workload demand). The settings also adjust the diameter of the sphere to determine the range of deviation from the center that can encompass the desired state. If you specify a large diameter, Workload Optimization Manager will have more variation in the way it distributes workload across hosting devices.

For more information, see The Desired State *(on page 144)*

Click **RUN PLAN** to view the results.

Plan configuration

Use the Plan Configuration toolbar to fine-tune your plan settings. You can change workload demand and the supply of resources, and specify other changes to the plan market.



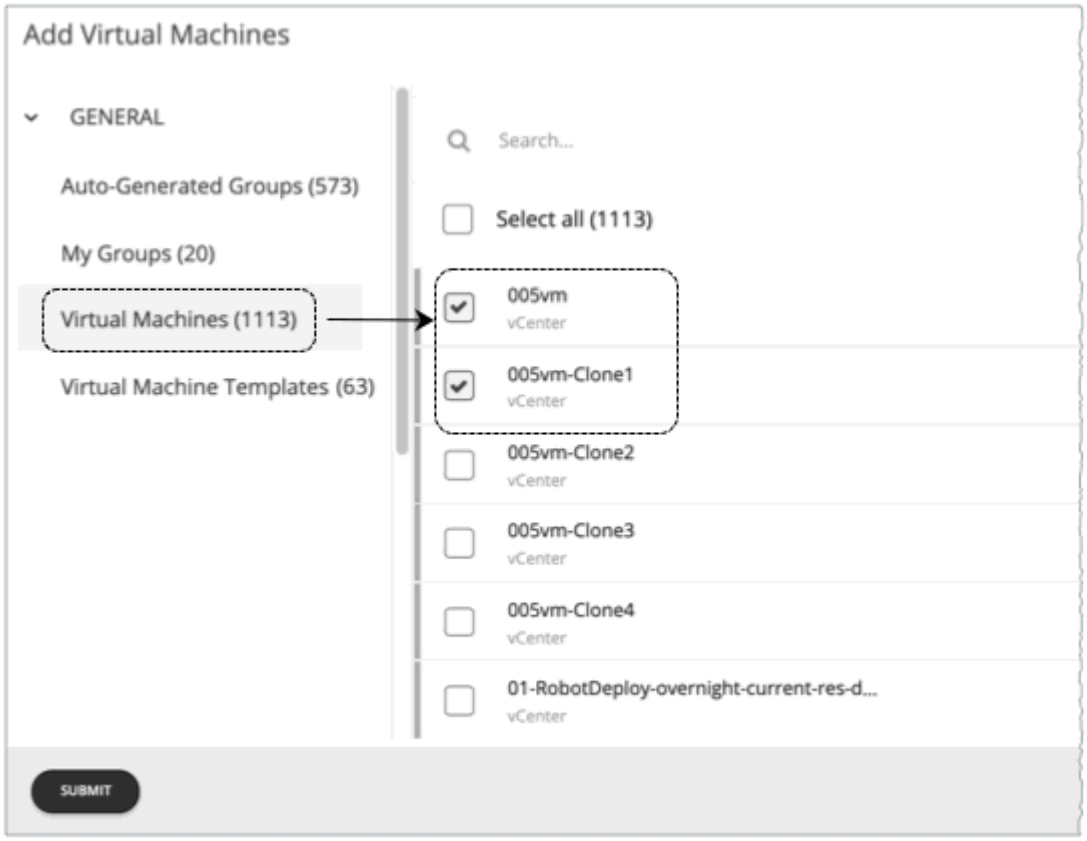
Add

Add virtual machines, hosts, or storage to your plan. For example, when you add hosts, you increase the compute resources for the plan.

- Copy from an entity or template

  Choose an entity or template to copy. This describes the new entities that Workload Optimization Manager will add to the plan. For example, you can run a plan that adds new VMs to a cluster. If you copy from a template, then the plan adds a new VM that matches the resource allocation you have specified for the given template.

  – Option 1: Copy from an entity



  – Option 2: Copy from a template

    If no existing template is satisfactory, create one by clicking **New Template**.



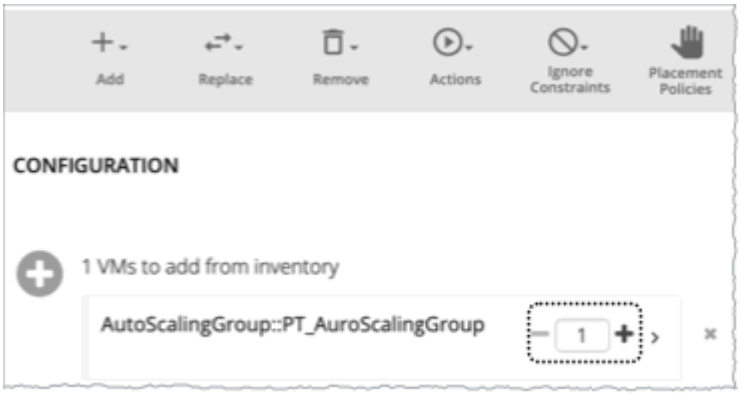
NOTE:
Workload Optimization Manager automatically adds any new template you create to the Template Catalog page (**Settings > Templates**).

It is not possible to use templates for containers or container pods.

Use the **Filter** option to show entities or templates with certain properties (name, number of CPUs, and so on). This makes it easier to sort through a long list.

- Number of copies to add

  After choosing an entity or template, it appears as an entry in the Configuration summary. Then you can set how many copies to add.



Replace

Replacing virtual machine is a way to change the properties of VMs in your plan market. When you replace workload, you select one or more VMs that you want to change, and then you select a template to use in their place. The list of changed VMs displays in the Configuration Summary. You can delete individual entries from the this summary if necessary.

Replacing hosts or storage is a way to plan for a hardware upgrade. For example, if you replace your hosts or data stores with a more powerful template, the plan might show that you can use fewer hosts or data stores, and it will show the best placement for workloads on those entities. You begin by selecting the entities you want to replace, and when you click **REPLACE** you can then choose a template that will replace them. Note that you can only choose a single template for each set of entities you want to have replaced. You can configure different replacements in the same plan, if you want to use more than one template.

Remove

Removing virtual machines frees up resources for other workloads to use. Removing hosts or storage means you have fewer compute or storage resources for your workloads.

If you think you have overprovisioned your environment, you can run a plan to see whether fewer hosts or less storage can still support the same workload.

Actions

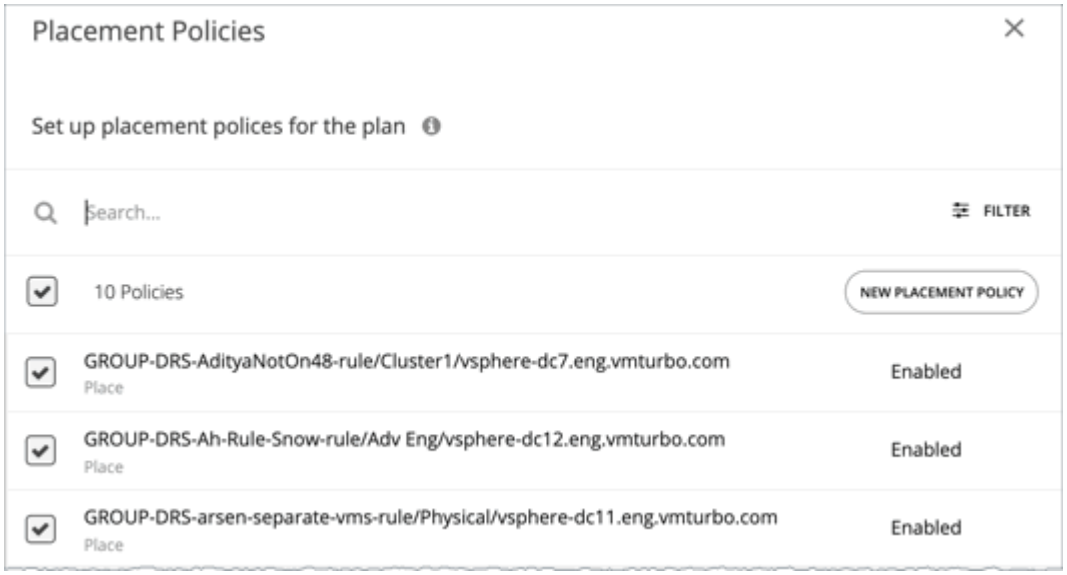
See the effect of enabling or disabling actions on the entities included in the plan. See Virtual Machine Actions *(on page 915)*, Host Actions *(on page 915)*, and Storage Actions *(on page 916)*.

Ignore constraints

Choose to ignore constraints (such as placement policies) for VMs in your environment. See Constraints Settings *(on page 917)*.

Placement policies

By default, the plan includes all the placement policies that apply to the plan scope. Also, these policies are in their real-time state (enabled or disabled).



You can use these settings to enable or disable existing policies, or you can create new policies to apply only to this plan scenario. For information about creating placement policies, see Placement Policies *(on page 1044)*.

Utilization

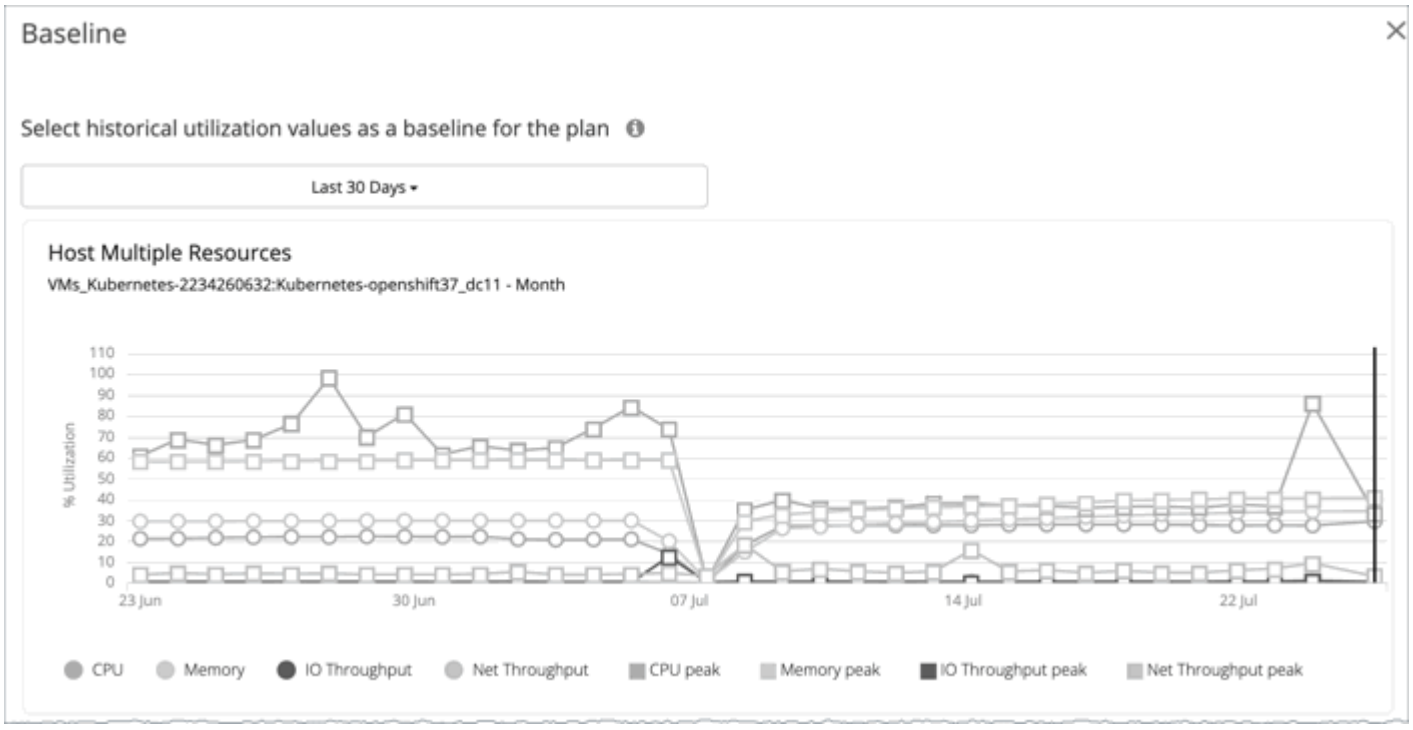
Setting utilization by a certain percentage is a way to increase or decrease the workload for the scope of your plan and any entity added to the plan, or for specific groups. Workload Optimization Manager uses the resulting utilization values as the baseline for the plan.

Max Host Utilization levels specify the percentage of the physical resource that you want to make available in the given plan. By default, hosts have utilization set to 100%. For a given plan, you can set the utilization to a lower value. For example, assume you want to simulate High Availability of 25% for some hosts in the plan. In that case, you can select these hosts and set their utilization levels to 75%.

Max Storage utilization levels specify the percentage of the physical resource that you want to make available in the given plan. By default, storage has utilization set to 100%. For a given plan, you can set the utilization to a lower value. For example, assume you have one data store that you want to share evenly for two clusters of VMs. Also assume that you are creating a plan for one of those clusters. In that case, you can set the data stores to 50% utilization. This saves storage resources for the other cluster that will use this storage.

Baseline

Use these settings to set up the baseline of utilization metrics for your plan.



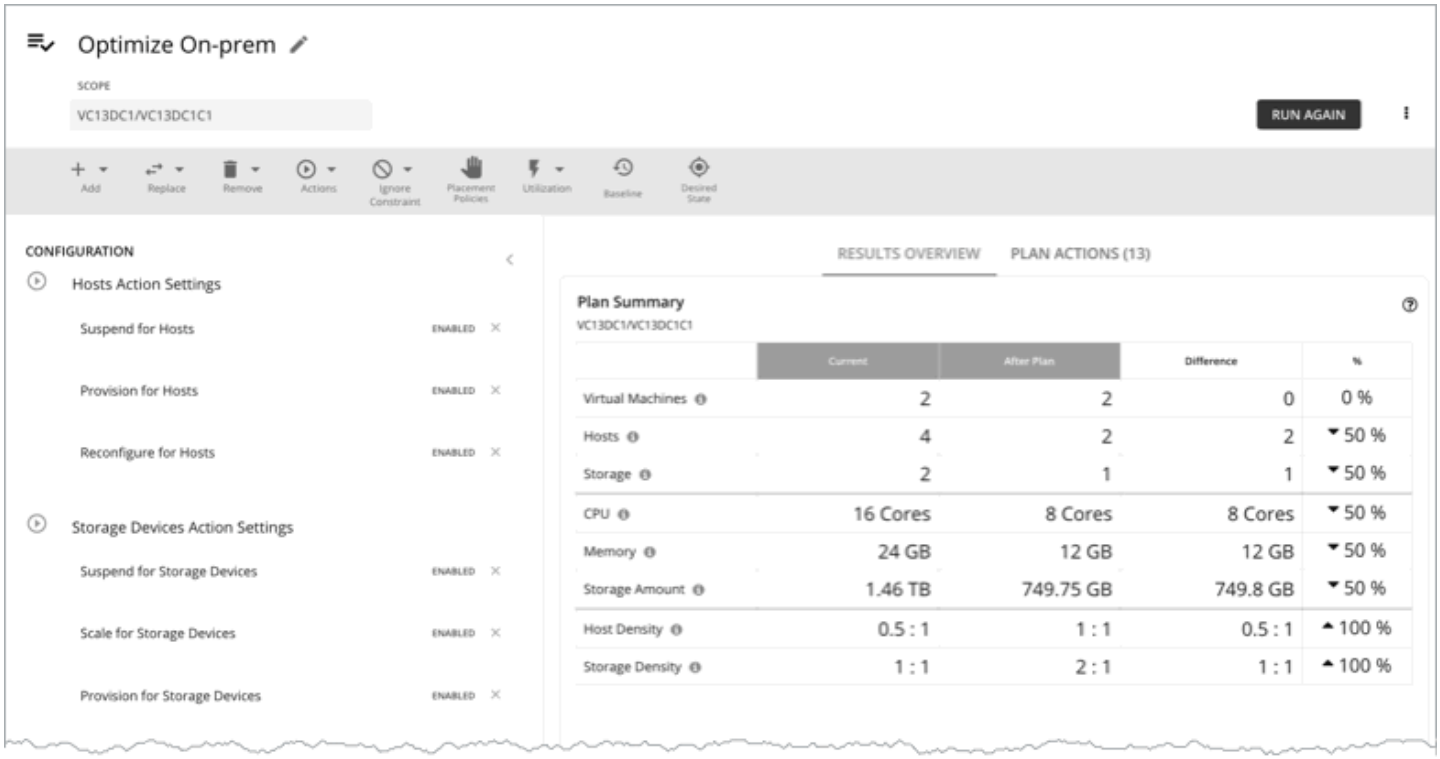
By default, the plan runs against the current state of your environment. You can set up the plan to add or remove entities, or otherwise affect the plan calculations. But the utilization metrics will be based on the current state of the plan. If you run the same plan multiple times, each run begins with a fresh view of your inventory.

You can select from the list of snapshots to load the utilization statistics from a previous time period into the plan. Use this to run the plan against utilization that you experienced in the past. For example, assume a peak utilization period for the month before the winter holidays. During the holidays you want to plan to add new capacity that can better handle that peak. You would set the baseline to the utilization you saw during that peak before the holidays.

Working with optimize on-prem plan results

After the plan runs, you can view the results to see how the plan settings you configured affect your environment.

Viewing the results



The plan results include the following charts:

Plan summary chart

This chart compares your current resources to the resources you would get after executing the plan.

> **NOTE:**
> Under some circumstances, this chart might not count "non-participating" entities in the real-time market, such as suspended VMs or hosts in a failover state. The following charts, on the other hand, count all entities in the real-time market, regardless of state:
>
> - Scope Preview chart (displays before you run the plan)
> - Optimized Improvements and Comparison charts

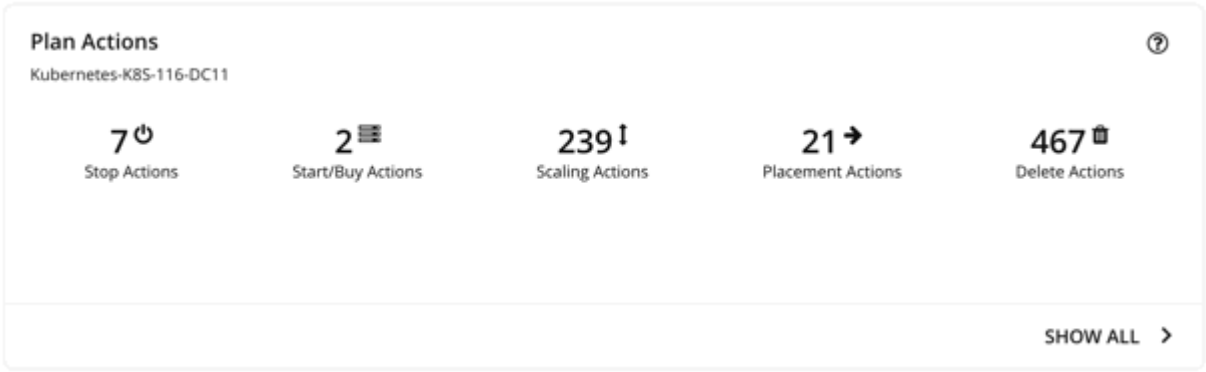
If the plan's scope includes VMs that cannot be placed, the results include a notification indicating the number of VMs. Click **Show Details** to see the list of VMs and the reasons for their non-placement.

Click **Show all** to see savings or investment costs, or to download the chart as a CSV file.

Plan actions chart

This chart summarizes the actions that you need to execute to achieve the plan results. For example, if you run an Alleviate Pressure plan, you can see actions to move workloads from the hot cluster over to the cold cluster. If some VMs are overprovisioned, you might see actions to reduce the capacity for those workloads.

The text chart groups actions by action type *(on page 845)*. The list chart shows a partial list of actions *(on page 817)*.

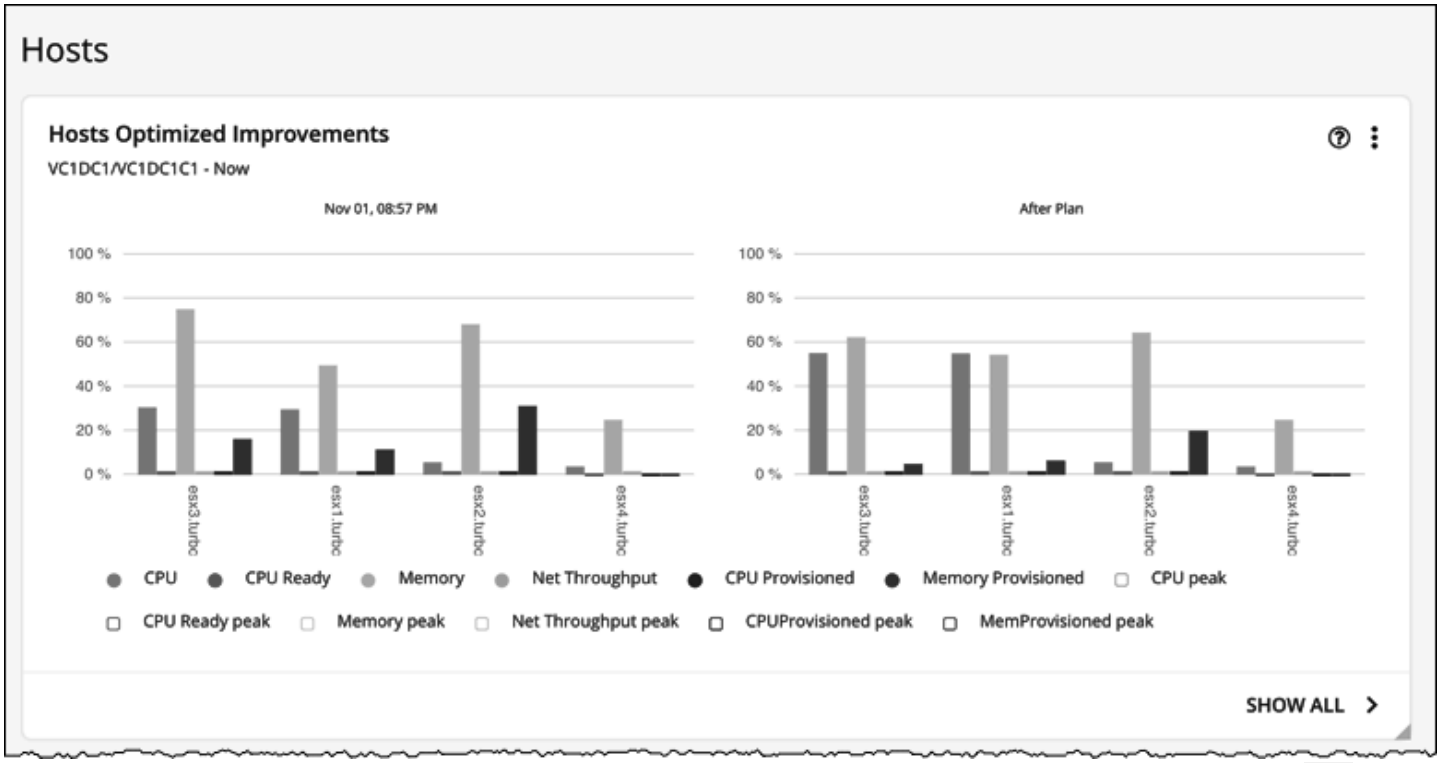


To view action details or download the list of actions as a CSV file:

- Click an action type in the text chart or an individual action in the list chart.
- Click **Show All** at the bottom of the chart.

Optimized improvements charts for hosts, storage, and virtual machines

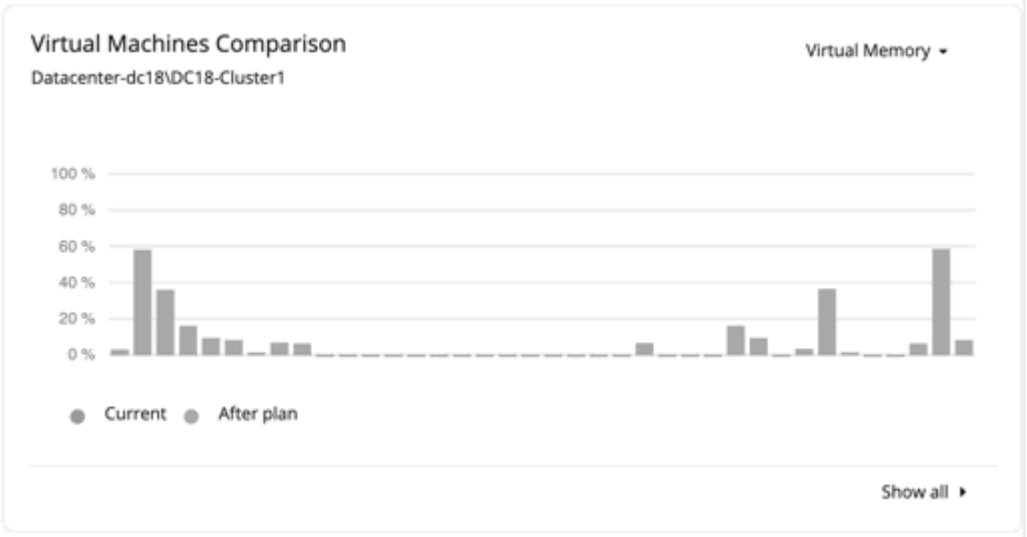
The Optimized Improvements chart shows how the utilization of resources would change assuming you accept all of the actions listed in the Plan Actions chart.



- In many of these charts, you can change the commodities on display. To do this, click the More options icon (⋮) and then select **Edit**. In the new screen that displays, go to the **Commodity** section and then add or remove commodities.
- Click **Show all** to see a breakdown of the current chart data by entity (for example, show CPU, Memory, and IO Throughput utilization for each host), or to download chart data as a CSV file.

Comparison charts for hosts, storage devices, and virtual machines

A Comparison chart shows how the utilization of a particular commodity (such as memory or CPU) for each entity in the plan would change if you execute the actions listed in the Plan Actions chart.



- To change the commodity displayed in the chart, select from the list of commodities in the chart.

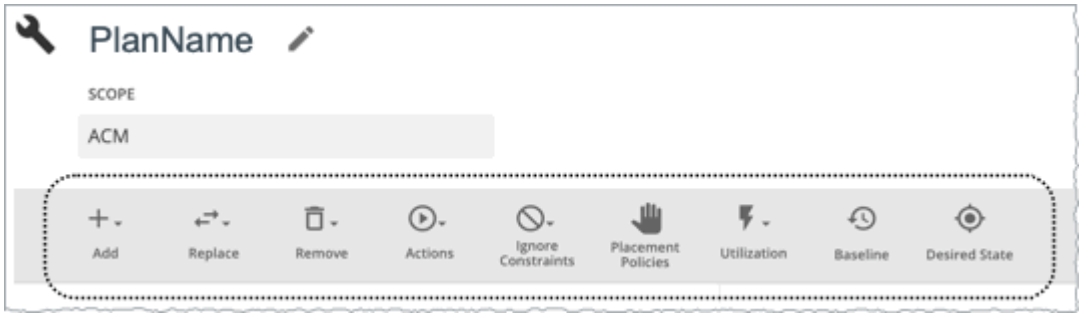
  To restore the default commodity, e, click the More options icon ( ), and then select **Reset view**.
- Click **Show all** at the bottom of the chart to show a breakdown of the current chart data by entity (for example, show Virtual Memory utilization for each virtual machine), or to download the chart as a CSV file.

NOTE:
For the Storage Devices Comparison chart, if you set the view to **VM Per Storage** and click **Show all**, the total number of VMs sometimes does not match the number in the Plan Summary chart. This happens if there are VMs in the plan that use multiple storage devices. The Storage Devices Comparison chart counts those VMs multiple times, depending on the number of storage devices they use, while the Plan Summary chart shows the actual number of VMs.

Re-running the plan

You can run the plan again with the same or a different set of configuration settings. This runs the plan scenario against the market in its current state, so the results you see might be different, even if you did not change the configuration settings.

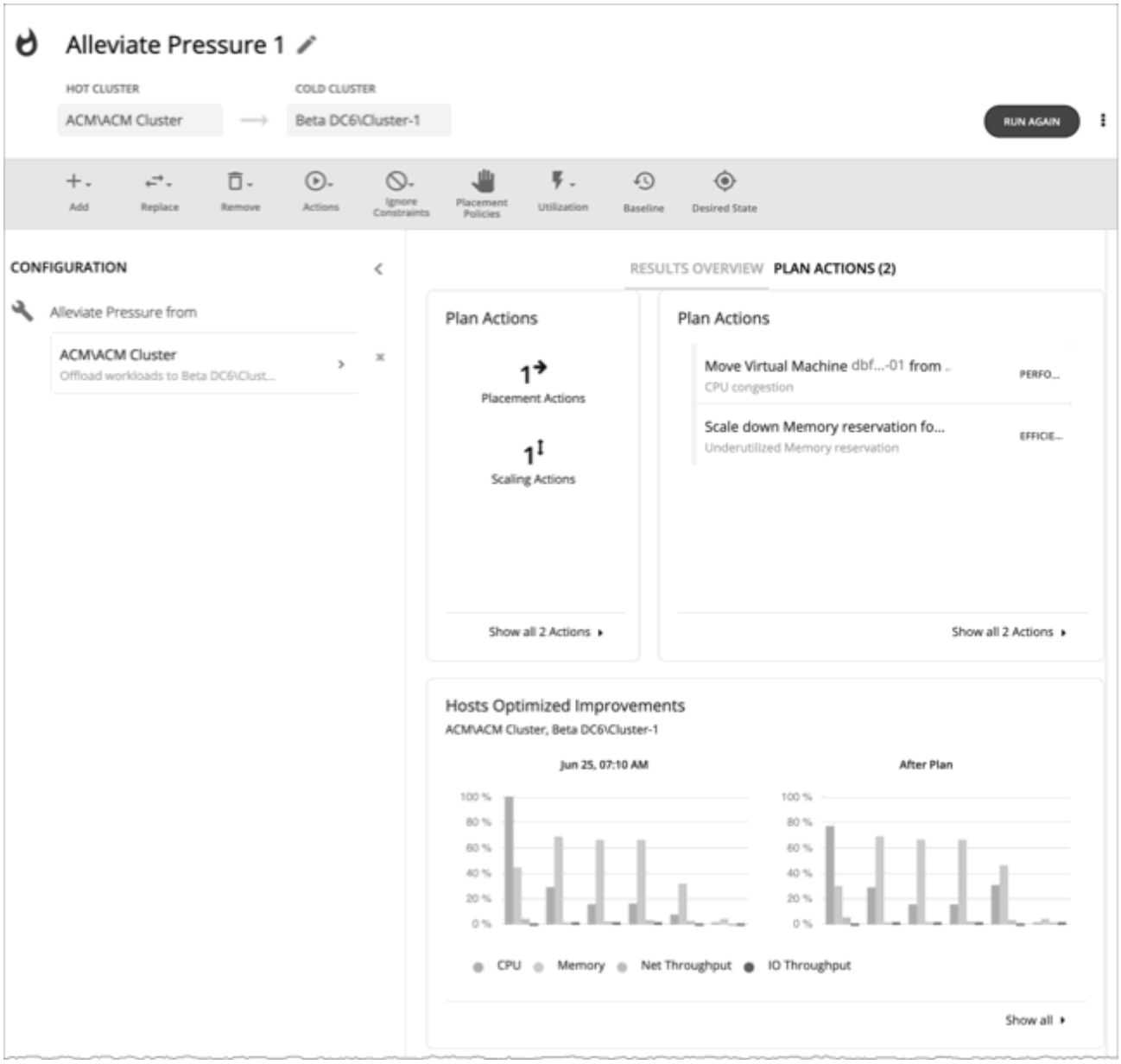
Use the toolbar for the Configuration section to change the configuration settings.



NOTE:
It is not possible to change the scope of the plan in the Plan Page. You must start over if you want a different scope. To start over, click the More options icon ( ), and then select **New Plan**.

When you are ready to re-run the plan, click **Run Again**.

Alleviate pressure plan



Use the Alleviate Pressure plan to find out how to migrate workloads from a stressed or *hot* cluster over to a cluster with more headroom. This plan shows the minimal changes you need to make to reduce risks on the hot cluster.

The plan results:

- Show the actions to migrate workloads from the hot cluster to the cold one
- Compare the current state of your clusters to the optimized state
- Show resulting headroom for both the hot and the cold clusters
- Show trends of workload-to-inventory over time for both clusters

Alleviate Pressure plans make use of the headroom in your clusters. Headroom is the number of VMs the cluster can support, for CPU, Memory and Storage.

To calculate cluster capacity and headroom, Workload Optimization Manager runs nightly plans that take into account the conditions in your current environment. The plans identify the optimal workload distribution for your clusters. This can include moving your current VMs to other hosts within the given cluster, if such moves would result in a more desirable workload distribution. The result of the plan is a calculation of how many more VMs the cluster can support.

To calculate VM headroom, the plan simulates adding VMs to your cluster. The plan assumes a certain capacity for these VMs, based on a specific VM template. For this reason, the count of VMs given for the headroom is an approximation based on that VM template.

To specify the templates these plans use, you can configure the nightly plans for each cluster. For more information, see Configuring Nightly Plans *(on page 941)*

NOTE:
To execute, this plan must ignore certain constraints. The plan ignores cluster constraints to allow migrating workloads from the hot cluster to the cold one. It also ignores network constraints, imported DRS policies, and any Workload Optimization Manager that would ordinarily be in effect.

Configuring an alleviate pressure plan

For an overview of setting up plan scenarios, see Setting Up Plan Scenarios *(on page 851)*.

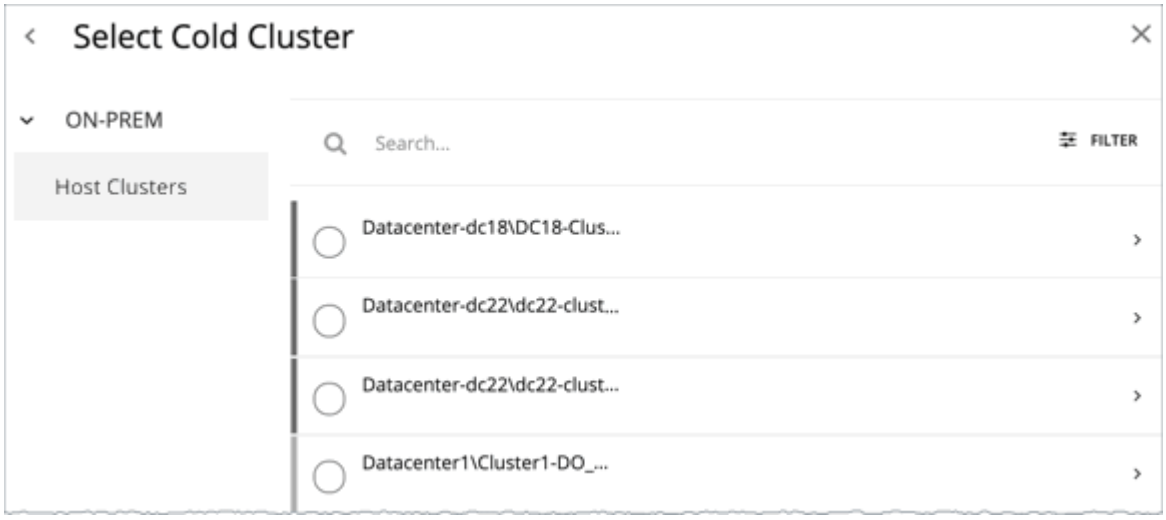
Scope

The wizard first gives you a list for you to choose the hot cluster. This is the cluster that shows risks to performance. The list sorts with the most critical clusters first, and it includes the calculated headroom for CPU, Memory, and Storage in each cluster.



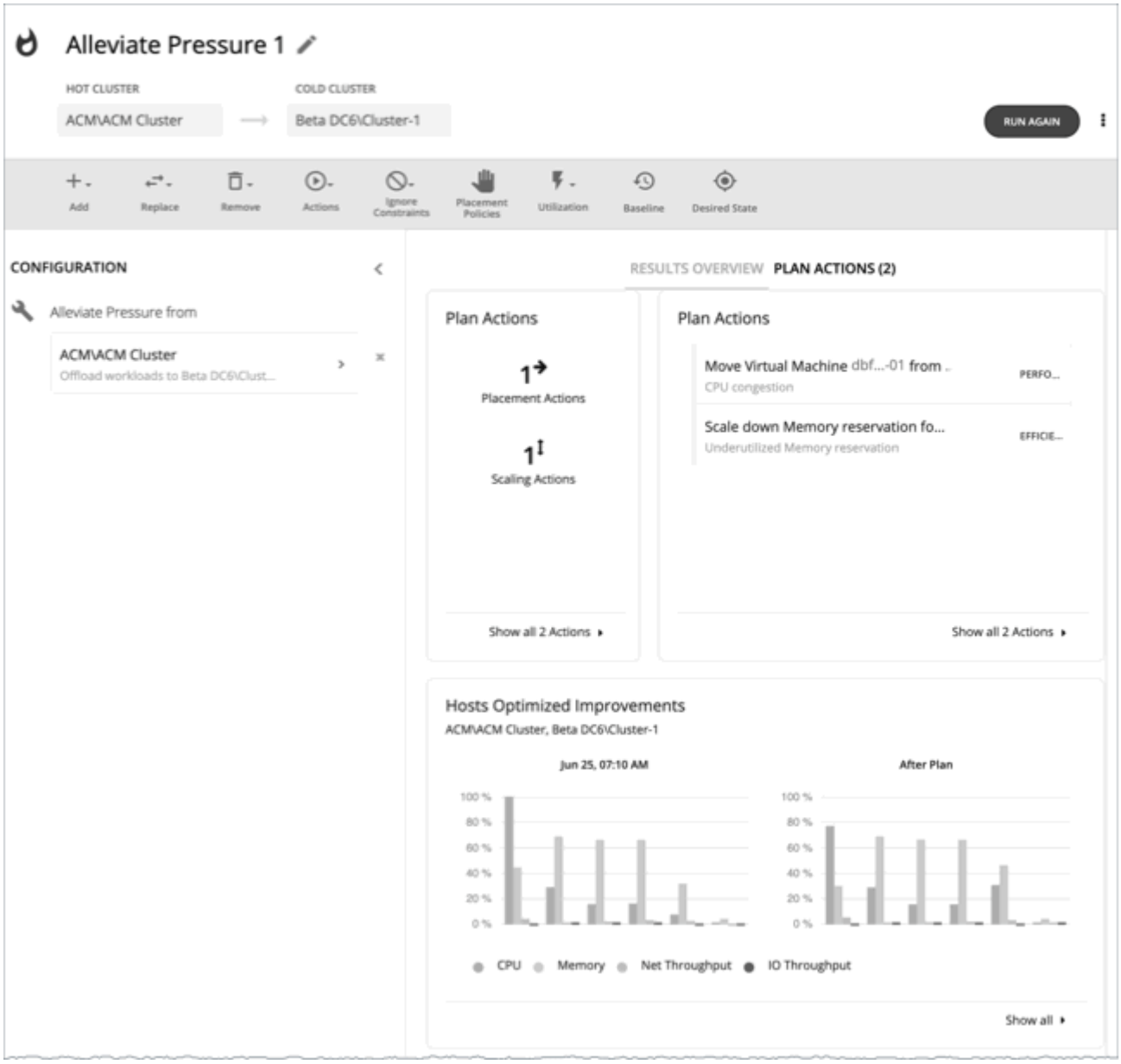
Cold cluster

After you select the hot cluster, choose the cold cluster.



Working with alleviate pressure plan results

After the plan runs, you can view the results to see how the migration of workloads off of your hot cluster affects your environment.



Viewing the results

The results include the following charts:

- Plan actions

  You can see a list of actions to reduce the pressure on the hot cluster. It's typical to see actions to move workloads from the hot cluster over to the cold cluster. If some VMs are overprovisioned, you might see actions to reduce the capacity for those workloads.

- Hosts optimized improvements

  This chart compares the current state of the hot cluster to its state after executing the plan actions. It displays the resource utilization of the cluster's hosts both before and after the plan.

- Headroom

  With these charts, you can compare the headroom between the hot and cold clusters.

- Virtual machines versus hosts and storage

  This chart shows the total number of virtual machines, hosts, and storage in your on-prem environment, and tracks the data over time. Chart information helps you understand and make decisions around capacity and utilization, based on historical and projected demand.

Re-running the plan

You can run the plan again with the same or a different set of configuration settings. This runs the plan scenario against the market in its current state, so the results you see might be different, even if you did not change the configuration settings.

Use the toolbar for the Configuration section to change the configuration settings.



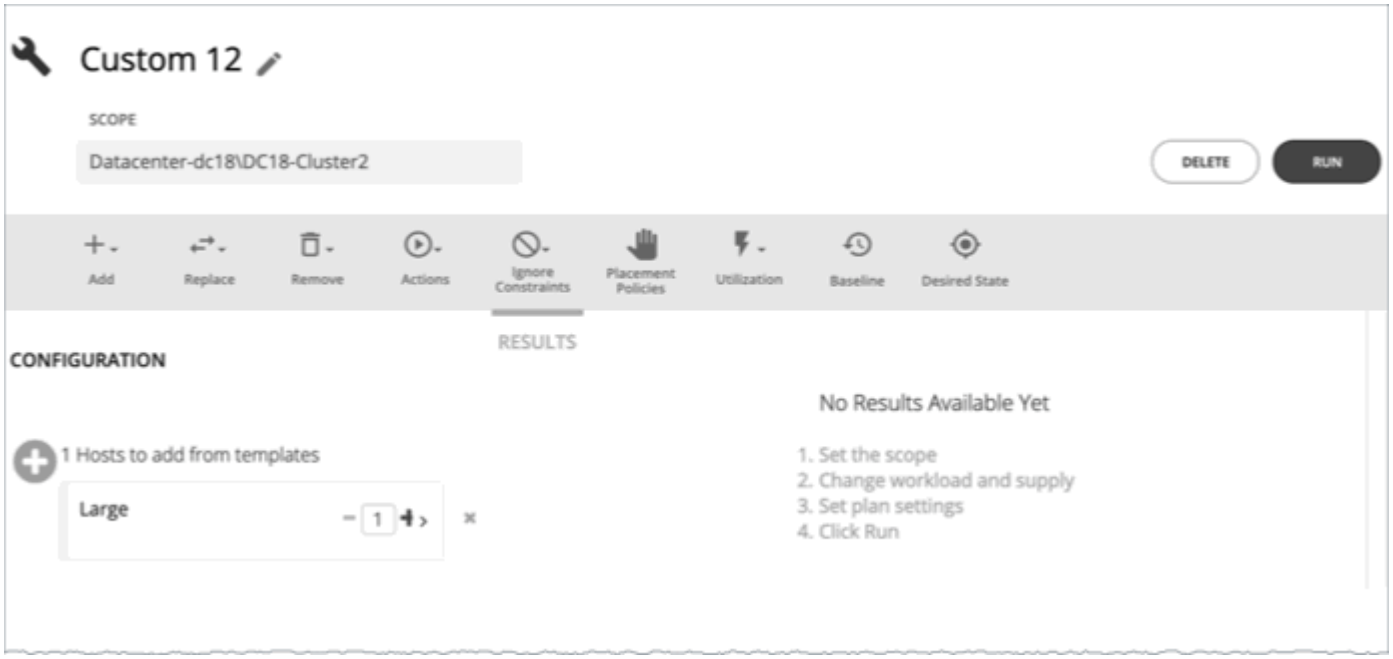
The toolbar items that display are similar to the toolbar items for a custom plan. For details, see Configuring a Custom Plan *(on page 931)*.

> **NOTE:**
> It is not possible to change the scope of the plan in the Plan Page. You must start over if you want a different scope. To start over, click the More options icon ( ), and then select **New Plan**.

When you are ready to re-run the plan, click **Run Again**.

Custom plan

For an overview of setting up plan scenarios, see Settings Up User Plan Scenarios *(on page 851)*.



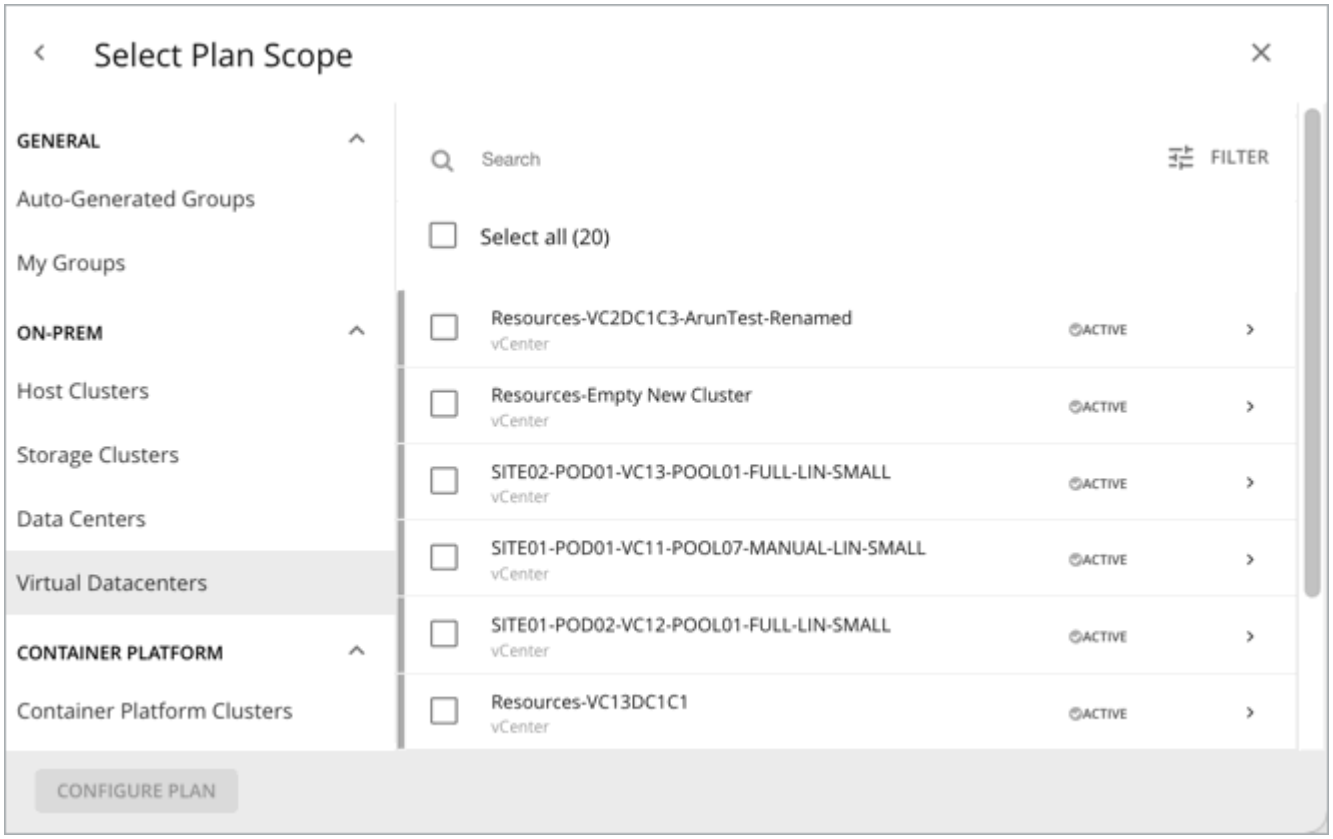
When you create a custom scenario, you specify the plan scope as an initial step, and then skip the plan wizards and jump straight into setting up the plan parameters. You can name the plan, change workload demand and the supply of resources, and specify other changes to the plan market.

Configuring a custom plan

For an overview of setting up plan scenarios, see Setting Up Plan Scenarios *(on page 851)*.

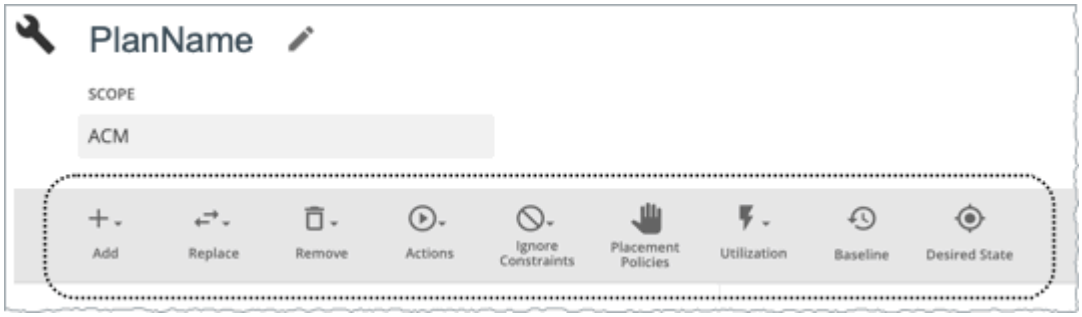
Scope

Specify the plan scope and then click **Configure Plan**.



Plan configuration

Use the Plan Configuration toolbar to fine-tune your plan settings. You can change workload demand and the supply of resources, and specify other changes to the plan market.



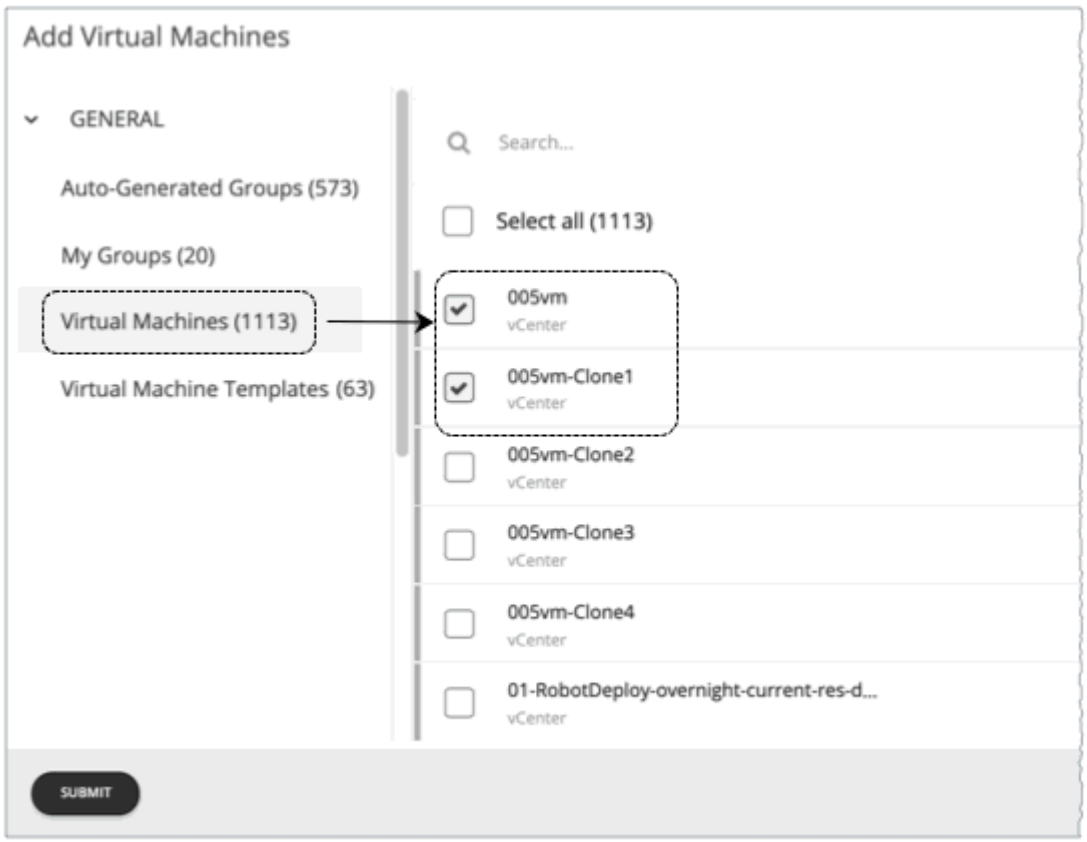
Add

Add virtual machines, hosts, or storage to your plan. For example, when you add hosts, you increase the compute resources for the plan.

- Copy from an entity or template

  Choose an entity or template to copy. This describes the new entities that Workload Optimization Manager will add to the plan. For example, you can run a plan that adds new VMs to a cluster. If you copy from a template, then the plan adds a new VM that matches the resource allocation you have specified for the given template.

  – Option 1: Copy from an entity



  – Option 2: Copy from a template

    If no existing template is satisfactory, create one by clicking **New Template**.



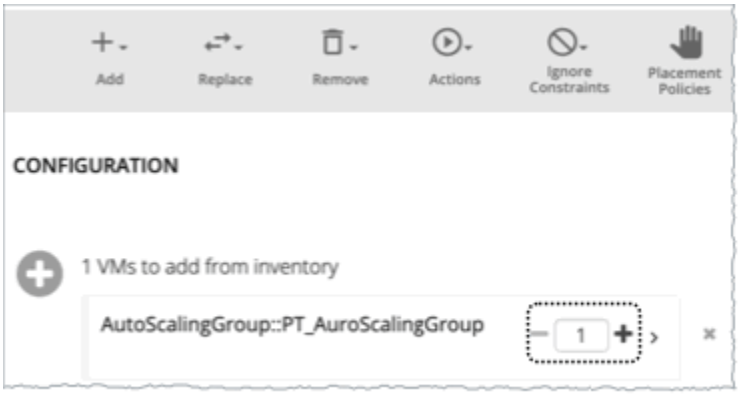
NOTE:
Workload Optimization Manager automatically adds any new template you create to the Template Catalog page (**Settings > Templates**).

It is not possible to use templates for containers or container pods.

Use the **Filter** option to show entities or templates with certain properties (name, number of CPUs, etc.). This makes it easier to sort through a long list.

- Number of copies to add

  After choosing an entity or template, it appears as an entry in the Configuration summary. Then you can set how many copies to add.



Replace

Replacing virtual machine is a way to change the properties of VMs in your plan market. When you replace workload, you select one or more VMs that you want to change, and then you select a template to use in their place. The list of changed VMs displays in the Configuration Summary. You can delete individual entries from the this summary if necessary.

Replacing hosts or storage is a way to plan for a hardware upgrade. For example, if you replace your hosts or data stores with a more powerful template, the plan might show that you can use fewer hosts or data stores, and it will show the best placement for workloads on those entities. You begin by selecting the entities you want to replace, and when you click **REPLACE** you can then choose a template that will replace them. Note that you can only choose a single template for each set of entities you want to have replaced. You can configure different replacements in the same plan, if you want to use more than one template.

Remove

Removing virtual machines frees up resources for other workloads to use.

Removing hosts or storage means you have fewer compute or storage resources for your workloads. If you think you have over-provisioned your environment, you can run a plan to see whether fewer hosts or less storage can still support the same workload.

Actions

See the effect of enabling or disabling actions on the entities included in the plan. For example, you might plan for more workload but know that you don't want to add more hardware, so you disable Provision of hosts for your plan. The results would then indicate if the environment can support the additional workload.

Ignore constraints

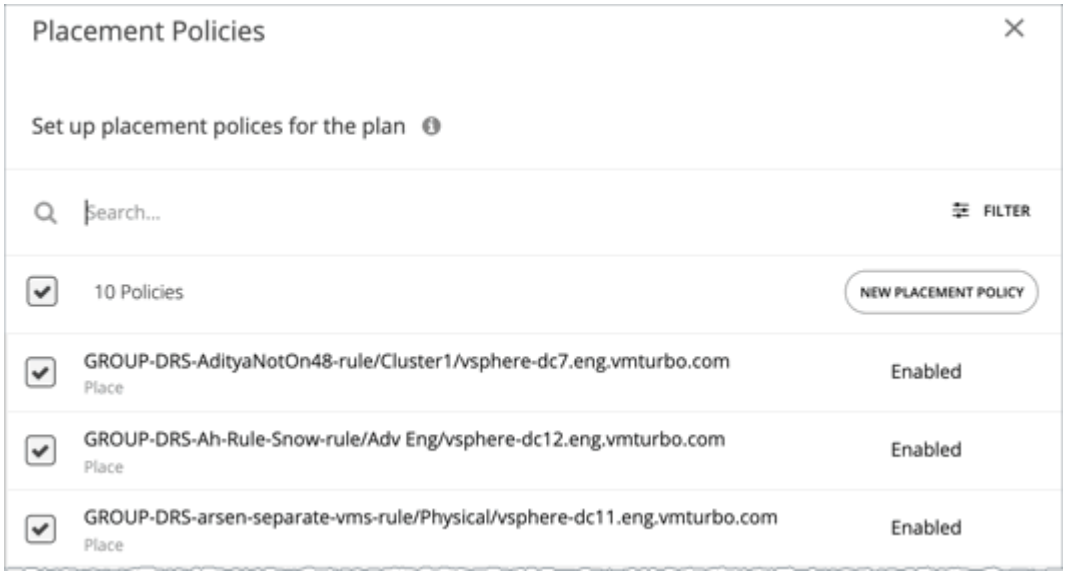
Choose to ignore constraints (such as placement policies) for VMs in your environment. By default, VMs are constrained to the cluster, network group, datacenter, or storage group that their hosts belong to. You can choose to ignore these boundaries.

For example, by default a plan does not consider moving VMs to physical hosts outside of the current cluster. If you disable the Cluster constraint for a VM in your plan, then the plan can evaluate the results of hosting those VMs on any other physical machine within the scope of your plan. If the best results come from moving that VM to a different cluster, then the plan will show that result.

> **NOTE:**
> If you are adding hosts to a plan, and use host templates, then you must turn on **Ignore Constraints**.

Placement policies

By default, the plan includes all the placement policies that apply to the plan scope. Also, these policies are in their real-time state (enabled or disabled).



You can use these settings to enable or disable existing policies, or you can create new policies to apply only to this plan scenario. For information about creating placement policies, see Placement Policies *(on page 1044)*.

Utilization

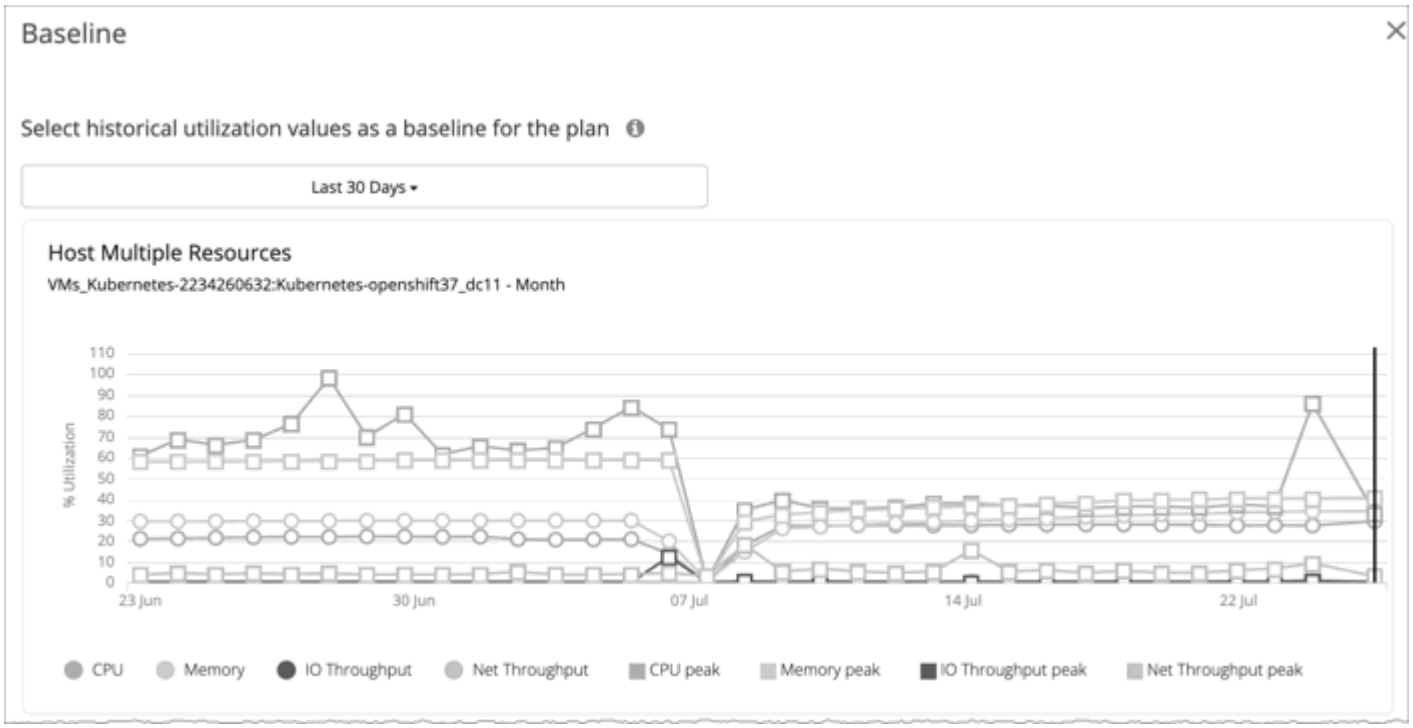
Setting utilization by a certain percentage is a way to increase or decrease the workload for the scope of your plan and any entity added to the plan, or for specific groups. Workload Optimization Manager uses the resulting utilization values as the baseline for the plan.

Max Host Utilization levels specify the percentage of the physical resource that you want to make available in the given plan. By default, hosts have utilization set to 100%. For a given plan, you can set the utilization to a lower value. For example, assume you want to simulate High Availability of 25% for some hosts in the plan. In that case, you can select these hosts and set their utilization levels to 75%.

Max Storage utilization levels specify the percentage of the physical resource that you want to make available in the given plan. By default, storage has utilization set to 100%. For a given plan, you can set the utilization to a lower value. For example, assume you have one data store that you want to share evenly for two clusters of VMs. Also assume that you are creating a plan for one of those clusters. In that case, you can set the data stores to 50% utilization. This saves storage resources for the other cluster that will use this storage.

Baseline

Use these settings to set up the baseline of utilization metrics for your plan.

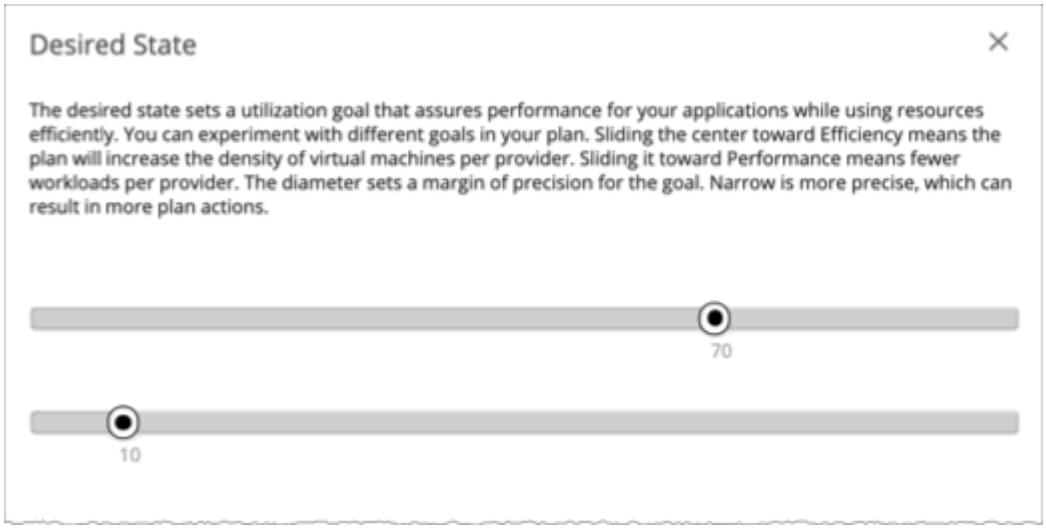


By default, the plan runs against the current state of your environment. You can set up the plan to add or remove entities, or otherwise affect the plan calculations. But the utilization metrics will be based on the current state of the plan. If you run the same plan multiple times, each run begins with a fresh view of your inventory.

You can select from the list of snapshots to load the utilization statistics from a previous time period into the plan. Use this to run the plan against utilization that you experienced in the past. For example, assume a peak utilization period for the month before the winter holidays. During the holidays you want to plan to add new capacity that can better handle that peak. You would set the baseline to the utilization you saw during that pre-holiday peak.

Desired state

The desired state is a condition in your environment that assures performance for your workloads, while it utilizes your resources as efficiently as possible and you do not overprovision your infrastructure. Workload Optimization Manager uses default Desired State settings to drive its analysis. You should never change the settings for real-time analysis unless you are working directly with Technical support. However, you can change the settings in a plan to see what effect a more or less aggressive configuration would have in your environment.



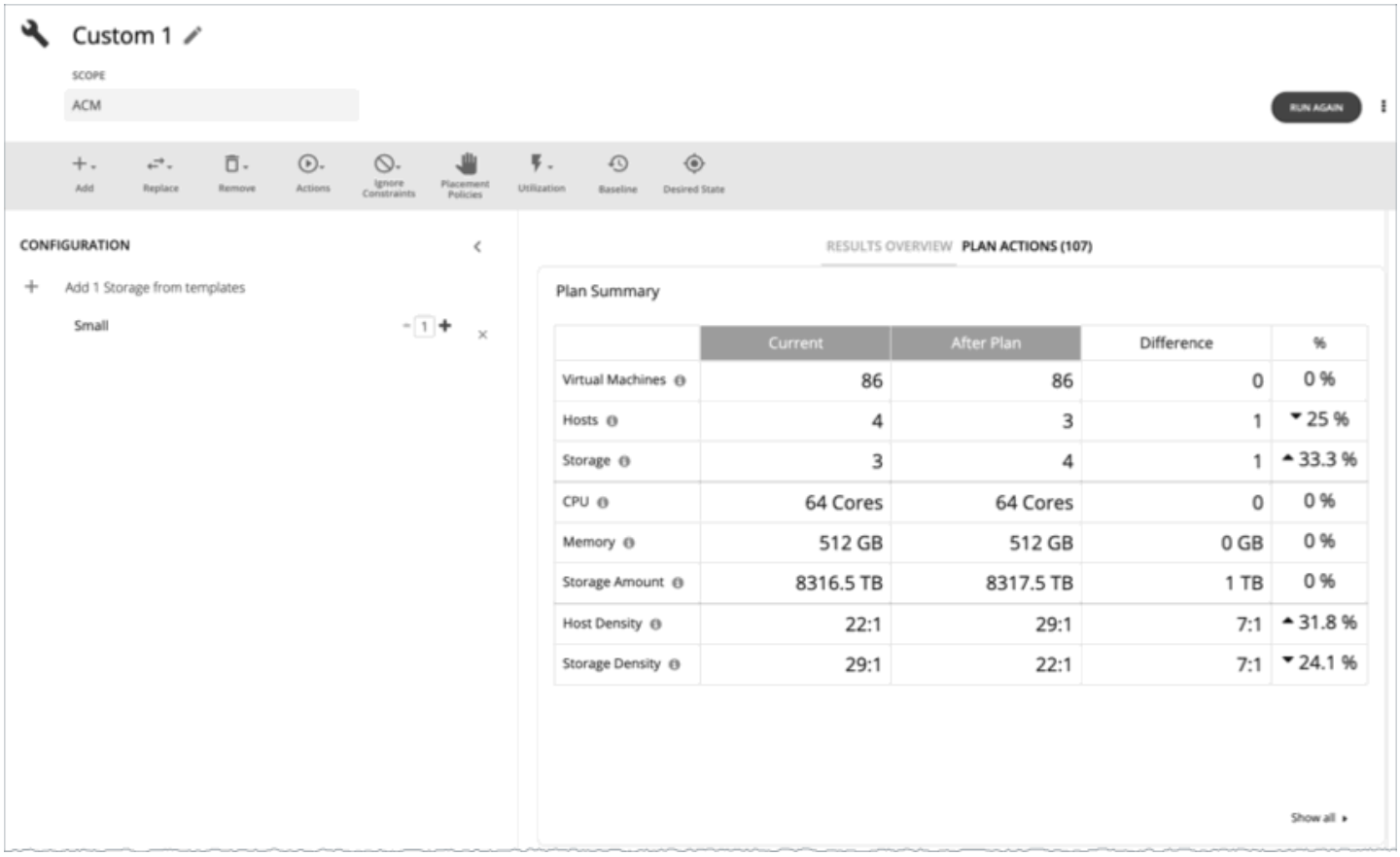
You can think of the desired state as an n-dimensional sphere that encompasses the fittest conditions your environment can achieve. The multiple dimensions of this sphere are defined by the resource metrics in your environment. Metric dimensions include VMem, storage, CPU, etc. While the metrics on the entities in your environment can be any value, the desired state, this n-dimensional sphere, is the subset of metric values that assures the best performance while achieving the most efficient utilization of resources that is possible.

The Desired State settings center this sphere on Performance (more infrastructure to supply the workload demand), or on Efficiency (less investment in infrastructure to supply the workload demand). The settings also adjust the diameter of the sphere to determine the range of deviation from the center that can encompass the desired state. If you specify a large diameter, Workload Optimization Manager will have more variation in the way it distributes workload across hosting devices.

For more information, see The Desired State *(on page 144)*.

Working with custom plan results

After the plan runs, you can view the results to see how the plan settings you configured affect your environment.



Viewing the results

The results include the following charts:

Plan summary chart

This chart compares your current resources to the resources you would get after executing the plan.

> **NOTE:**
> Under some circumstances, this chart might not count "non-participating" entities in the real-time market, such as suspended VMs or hosts in a failover state. The following charts, on the other hand, count all entities in the real-time market, regardless of state:
>
> - Scope Preview chart (displays before you run the plan)
> - Optimized Improvements and Comparison charts

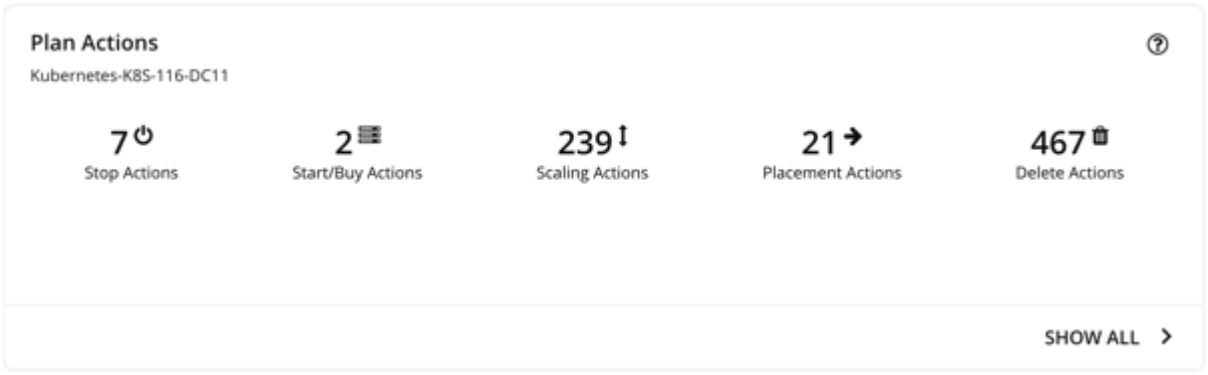
If the plan's scope includes VMs that cannot be placed, the results include a notification indicating the number of VMs. Click **Show Details** to see the list of VMs and the reasons for their non-placement.

Click **Show all** at the bottom of the chart to see savings or investment costs, or to download the chart as a CSV file.

Plan actions chart

This chart summarizes the actions that you need to execute to achieve the plan results. For example, if you run an Alleviate Pressure plan, you can see actions to move workloads from the hot cluster over to the cold cluster. If some VMs are overprovisioned, you might see actions to reduce the capacity for those workloads.

The text chart groups actions by action type *(on page 845)*. The list chart shows a partial list of actions *(on page 817)*.

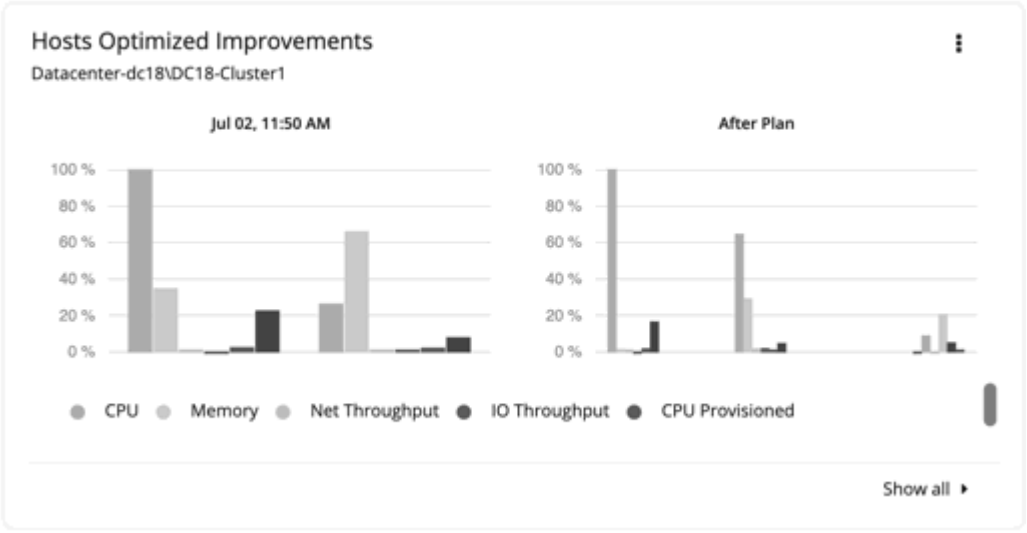


To view action details or download the list of actions as a CSV file:

- Click an action type in the text chart or an individual action in the list chart.
- Click **Show All**.

Optimized improvements charts for hosts, storage, and virtual machines

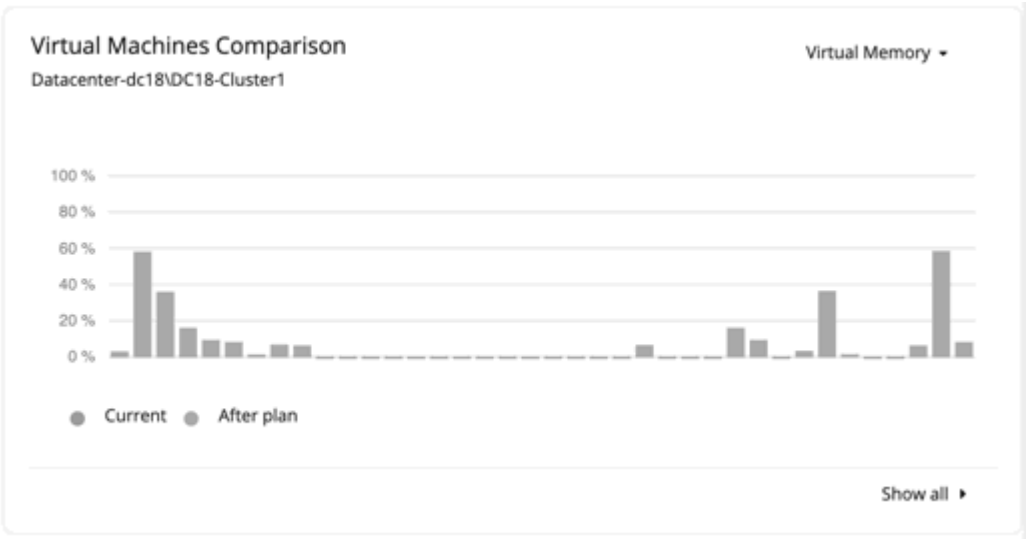
The Optimized Improvements chart shows how the utilization of resources would change assuming you accept all of the actions listed in the Plan Actions chart.



- In many of these charts, you can change the commodities on display. To do this, click the More options icon (⋮) and then select **Edit**. In the new screen that displays, go to the **Commodity** section and then add or remove commodities.
- Click **Show all** to see a breakdown of the current chart data by entity (for example, show CPU, Memory, and IO Throughput utilization for each host), or to download chart data as a CSV file.

Comparison charts for hosts, storage devices, and virtual machines

A Comparison chart shows how the utilization of a particular commodity (such as memory or CPU) for each entity in the plan would change if you execute the actions listed in the Plan Actions chart.



- To change the commodity displayed in the chart, go to the chart and then select from the list of commodities.

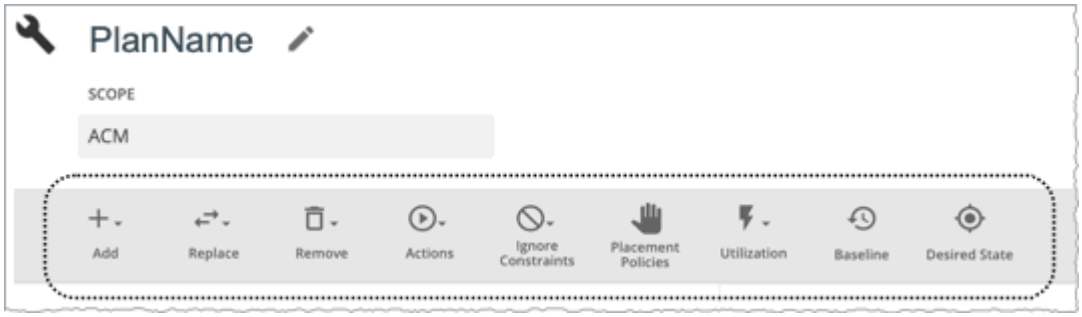
  To restore the default commodity, click the More options icon ( ), and then select **Reset view**.
- Click **Show all** at the bottom of the chart to show a breakdown of the current chart data by entity (for example, show Virtual Memory utilization for each virtual machine), or to download the chart as a CSV file.

NOTE:
For the Storage Devices Comparison chart, if you set the view to **VM Per Storage** and click **Show all**, the total number of VMs sometimes does not match the number in the Plan Summary chart. This happens if there are VMs in the plan that use multiple storage devices. The Storage Devices Comparison chart counts those VMs multiple times, depending on the number of storage devices they use, while the Plan Summary chart shows the actual number of VMs.

Re-running the plan

You can run the plan again with the same or a different set of configuration settings. This runs the plan scenario against the market in its current state, so the results you see might be different, even if you did not change the configuration settings.

Use the toolbar for the Configuration section to change the configuration settings.



For details about these settings, see Configuring a Custom Plan *(on page 931)*.

NOTE:
It is not possible to change the scope of the plan in the Plan Page. You must start over if you want a different scope. To start over, click the More options icon ( ), and then select **New Plan**.

When you are ready to re-run the plan, click **Run Again**.

Configuring nightly plans



Workload Optimization Manager runs nightly plans to calculate headroom for the clusters in your on-prem environment. For each cluster plan, you can:

1. View the plan and its template.
2. Expand to see template details.
3. Set which VM template to use in these calculations.

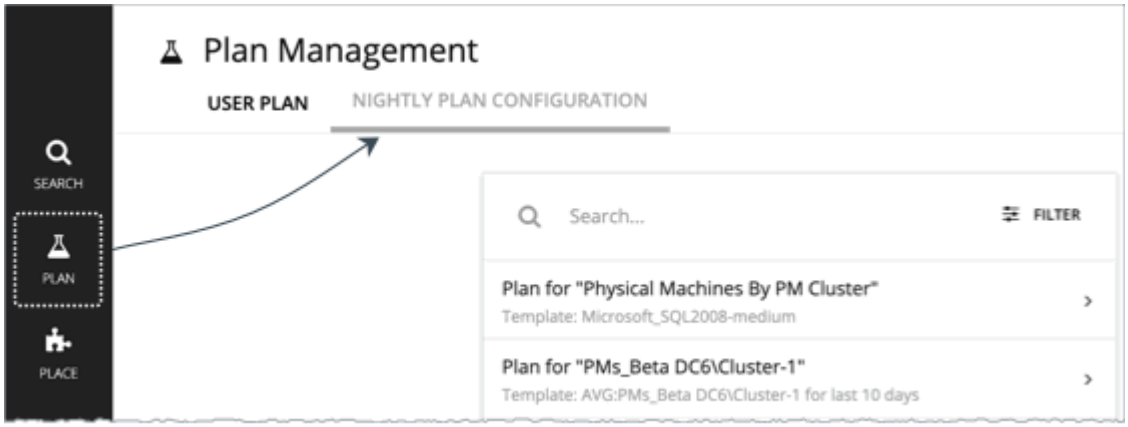
For information about viewing cluster headroom, see Viewing Cluster Headroom *(on page 172)*.

To calculate cluster capacity and headroom, Workload Optimization Manager runs nightly plans that take into account the conditions in your current environment. The plans identify the optimal workload distribution for your clusters. This can include moving your current VMs to other hosts within the given cluster, if such moves would result in a more desirable workload distribution. The result of the plan is a calculation of how many more VMs the cluster can support.

To calculate VM headroom, the plan simulates adding VMs to your cluster. The plan assumes a certain capacity for these VMs, based on a specific VM template. For this reason, the count of VMs given for the headroom is an approximation based on that VM template.

To set templates to use for the nightly plans:

1. Navigate to the Plan Page and click **NIGHTLY PLAN CONFIGURATION**.



   This displays a list of all the nightly plans. Workload Optimization Manager creates a nightly plan for each cluster.
2. Click the plan that you want to configure.

   A fly-out appears that lists all the available templates.
3. Select the template you want for this plan.

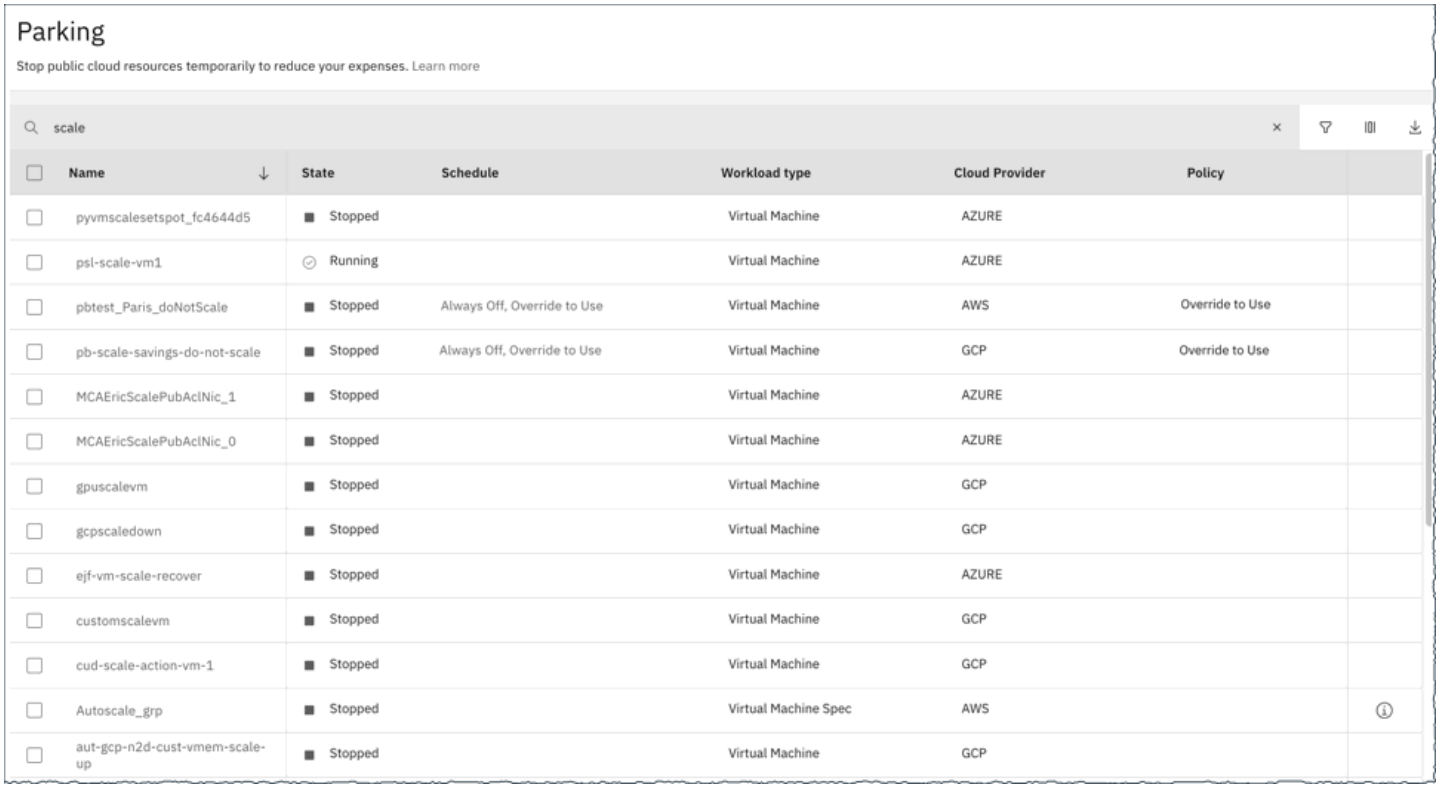
   Choose the template and click **Select**.

Park: stop or start cloud resources

Overview

Workload Optimization Manager supports 'parking' actions for cloud resources. These user-initiated actions *stop* your cloud resources for a given period of time to help you reduce your cloud expenses, and then *start* these resources later when you need them. You can enforce parking actions on demand, or automatically through parking schedules and policies.

When you click **Park** on the main menu, the user interface displays a list of parkable entities.



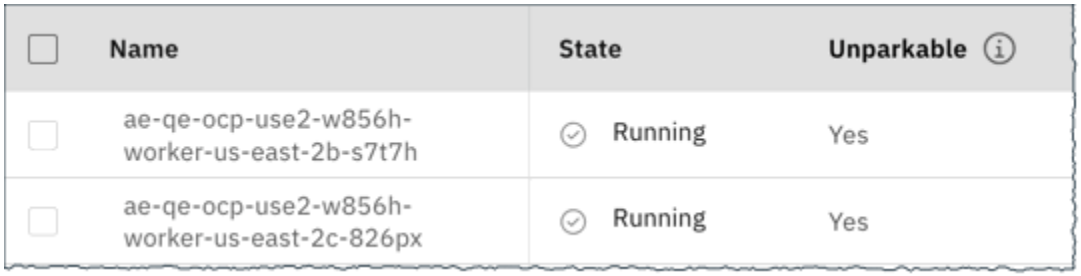
Parkable entities

Currently, Workload Optimization Manager supports parking actions for the following entities:

Supported entity	Parkable resources
Virtual Machine	Virtual machines in AWS, Azure, and Google Cloud
Virtual Machine Spec	AWS Auto Scaling groups and Azure Virtual Machine Scale Sets Points to consider: ▪ Auto Scaling groups and Azure Virtual Machine Scale Sets in the Parking page are only available in the "new design" in the *User Guide* of the user interface. ▪ Currently, the following entities are not discovered as Virtual Machine Specs and are not parkable: – Auto Scaling groups in Amazon Elastic Kubernetes Service (EKS) – Virtual Machine Scale Sets in Azure Kubernetes Service (AKS) – Virtual Machine Scale Sets that apply multiple autoscale profiles

Supported entity	Parkable resources
Database Server	AWS RDS instances ▪ AWS provides guidelines and caveats for parking actions. For example, your accounts will continue to incur certain charges while instances are stopped. For details, visit this AWS page. ▪ AWS RDS clusters and cluster members are not parkable.
Database	Azure dedicated SQL pools (used in Azure Synapse Analytics) Workload Optimization Manager analysis can also generate suspend actions for idle SQL pools. Suspend actions have the same effect as stop actions. For details, see Suspend Actions for Dedicated SQL Pools *(on page 712)*

Some entities that display in the Parking page may be unparkable when you view them because the entities apply a parking policy *(on page 1087)* that disables parking actions. The **Unparkable** column indicates if such a policy currently applies. These entities cannot be parked on demand or automatically.



For example, you may have virtual machines in your production environment that must be running all the time. You can prevent the enforcement of parking actions by creating a parking policy that disables parking actions.

How parking works

Cloud workloads that become idle for a period of time (for example, after business hours) and those that undergo regular maintenance are ideal for parking. When parking actions are enforced, Workload Optimization Manager reverses any conflicting actions on these entities. For example, when a VM has been started via Workload Optimization Manager and a user stops the VM via the cloud provider, Workload Optimization Manager will start the VM after a subsequent discovery cycle.

When a scaling action for a VM or Database Server is in progress, Workload Optimization Manager postpones a stop or start action until the scaling action completes.

Enforcing parking actions

There are several ways to enforce parking actions.

- Execute actions on demand.

  Use the Parking page to execute parking actions on individual entities. For details, see Executing Parking Actions *(on page 947)*.

- Attach a parking schedule to one or several entities.

  Use the Parking page to attach a parking schedule to one or several entities. You must create parking schedules as a prerequisite. For details, see Managing Parking Schedules *(on page 1097)* and Scheduling Parking Actions *(on page 948)*.

NOTE:
When you execute a parking action on demand and the affected entity has an existing schedule, a system prompt notifies you to pause the schedule before executing the action. This prevents potential conflicts between on-demand and scheduled parking actions.

- Create parking policies.

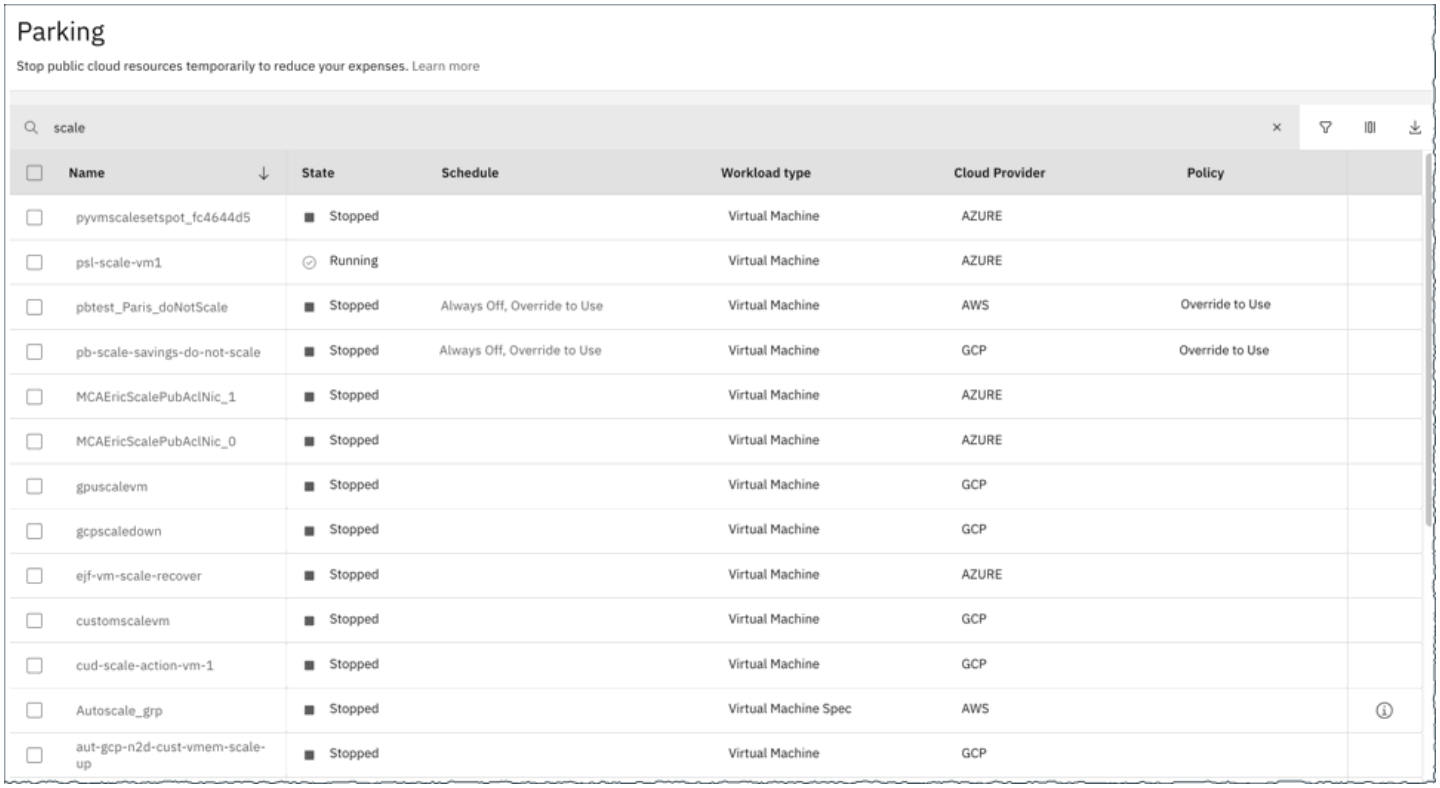
  Use parking policies to enforce actions dynamically and at scale. For example, you can create a parking policy at the cloud provider level so that all existing and newly discovered parkable entities from a specific cloud provider are parked uniformly.

  Use the Policy Management page to create parking policies. For details, see Parking Policies *(on page 1087)*.

Parking policies take precedence over the other enforcement methods. For example, if a parking policy dictates that a VM must be running everyday from 9:00 AM to 6:00 PM, and a Workload Optimization Manager user manually stopped the VM from the Parking page at 3:00 PM, the VM will be started to comply with the parking policy.

Working with the Parking page

After discovering parkable entities, Workload Optimization Manager adds them to the Parking page and shows their state (stopped or running) at the time of discovery. Use the Parking page to execute parking actions or attach parking schedules to these entities.



Users with the Site Administrator, Administrator, Automator, or Parker role can manage the parking actions and settings in the Parking page. Users with other roles have read-only access to the page. For details about roles, see Managing User Accounts *(on page 1116)*.

General tasks:

- For a long list of entities, you can type a search keyword or add filters to narrow the results.

  For filters, valid values display in a dropdown list. These filters support regular expressions and partial keyword matching to help you narrow down the values further.

- To set an entity as your scope, click the entity name.
- To download the data in the Parking page, click the download button.

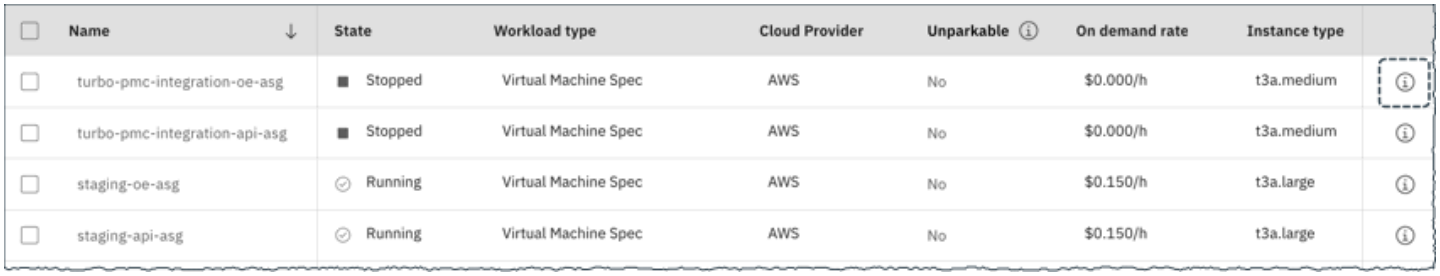
Be aware of the following as you manage parking actions.

- To see and download a list of parking (and non-parking) actions that have been executed successfully, add the All Actions chart to your dashboard and use the tabular view.
- Actual savings associated with parking actions are reflected in the Savings and Cumulative Savings charts, but not in the Parking page.

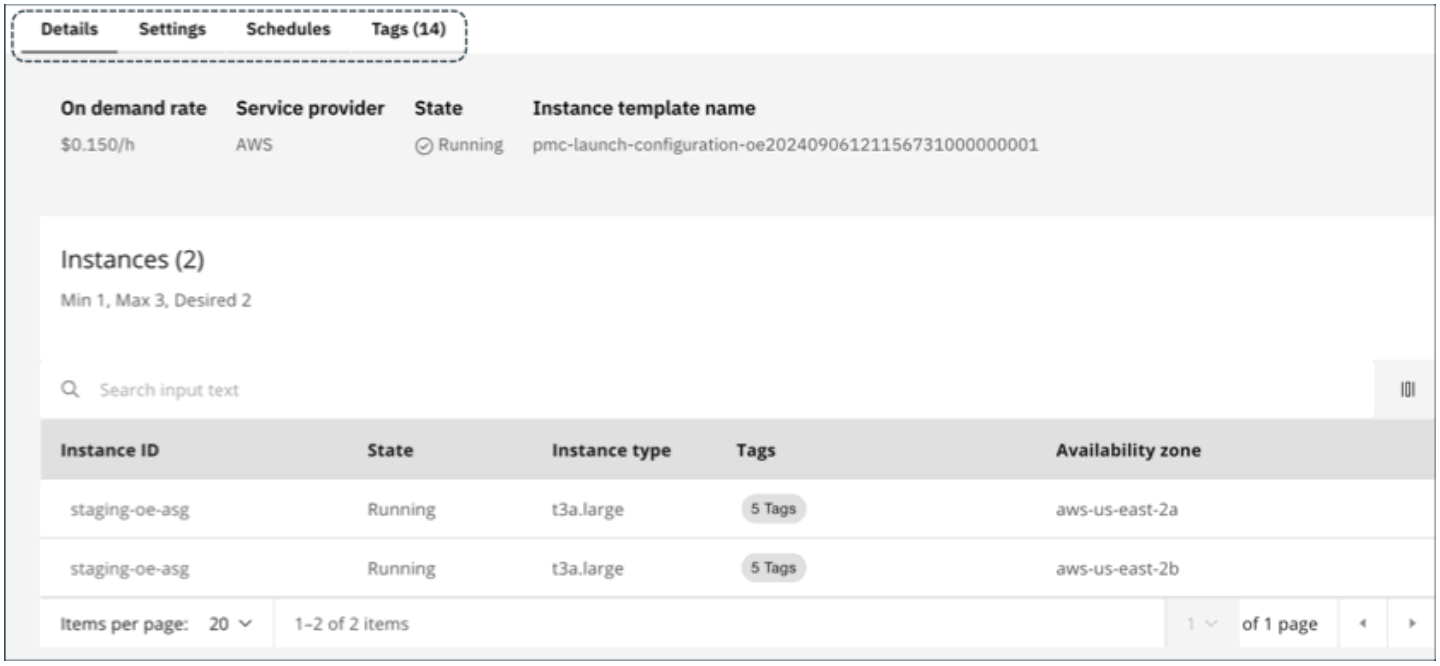
  For an entity that is currently running, the **Cost** column in the Parking page shows the hourly on-demand rate for that entity, for your reference.

Details for AWS Auto Scaling Groups and Azure Virtual Machine Scale Sets

The table in the Parking page shows individual AWS Auto Scaling Groups and Azure Virtual Machine Scale Sets, but not the member instances. To view member instances, click the information icon in the table.



A new page opens and includes the following tabs:



Tab name	Details
Details	This tab shows the instances discovered from an AWS Auto Scaling group or Azure Virtual Machine Scale Set that is currently running.

Tab name	Details
Settings	This tab shows the following scaling limits: ▪ **Discovered** – The current scaling limits as discovered from AWS or Azure ▪ **After start action** – After a start action is executed, the Auto Scaling group or Virtual Machine Scale Set applies the scaling limits assigned to it before it was stopped. ▪ **After stop action** – After a stop action is executed, scaling limits become zero. Workload Optimization Manager persists the scaling limits before the stop action was executed and then assigns the same limits to the Auto Scaling group or Virtual Machine Scale Set when it is started later. The scaling limits that display vary by cloud provider. ▪ For AWS, the min, max, and desired values shown in Workload Optimization Manager are the scaling limits that you configured in AWS. ▪ For Azure: – For Virtual Machine Scale Sets that apply an autoscale profile that scales instances dynamically based on metrics, the default, minimum, and maximum values are the scaling limits that you configured in the profile. – For Virtual Machine Scale Sets that apply a manual autoscale profile or a custom autoscale profile that scales instances to a specific count, the default value is the fixed instance count that you configured in the profile. Minimum and maximum values do not apply and are not shown.
Schedules	This tab shows the schedule that is currently assigned to the Auto Scaling group or Virtual Machine Scale Set.
Tags	This tab shows the tags that are currently assigned to the Auto Scaling group or Virtual Machine Scale Set, and the individual instances.

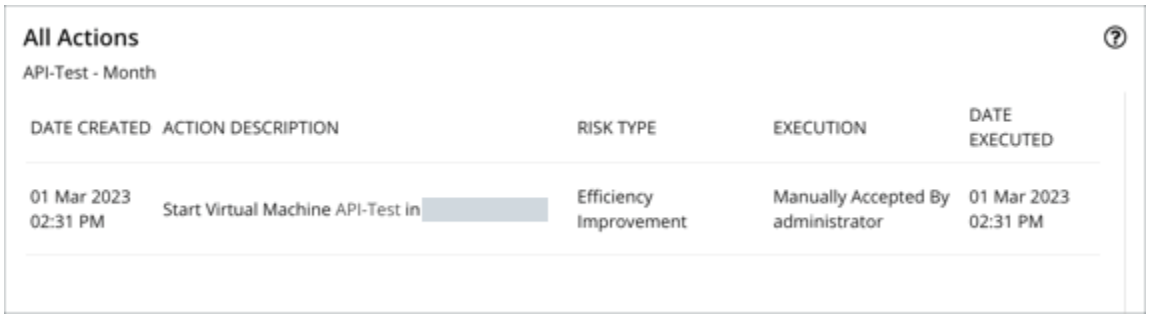
Executing parking actions on demand

Parking actions can only be executed from the Parking page. The Pending Actions, Potential Savings, and Top Accounts charts in the Home Page only show the non-parking actions that you can execute.

To execute a parking action, select one or several entities with the same state, and then click **Stop** or **Start**. In the following example, the selected entities are currently stopped and will be started.



In the All Actions chart, the tabular view shows the following information for the parking actions that you executed:



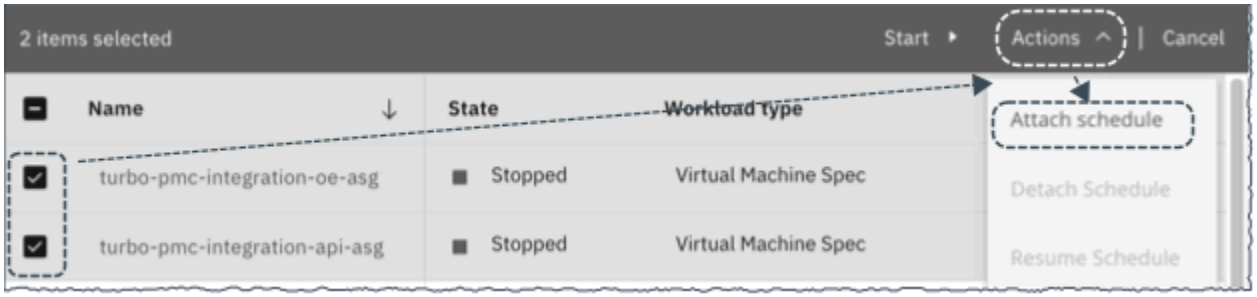
- Execution: Manually Accepted
- Risk Type: Savings (stop actions) or Efficiency Improvement (start actions)

Scheduling parking actions

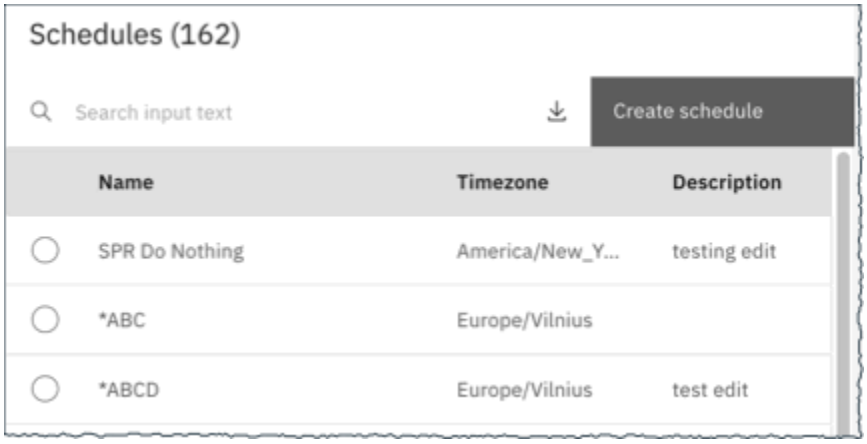
Attach a parking schedule to one or several entities to ensure that parking actions occur automatically at the appropriate times. You must create parking schedules as a prerequisite. For instructions on creating parking schedules, see Managing Parking Schedules *(on page 1097)*.

> **NOTE:**
> Use parking policies *(on page 1087)* to schedule parking actions at scale. The parking schedules that you set up can be reused in these policies.
>
> Parking actions cannot be scheduled using automation policies *(on page 663)*. Automation policies (and the schedules configured in these policies) only apply to scale actions.

To attach a schedule, select one or several entities without a schedule and then select **Attach** from the dropdown menu.

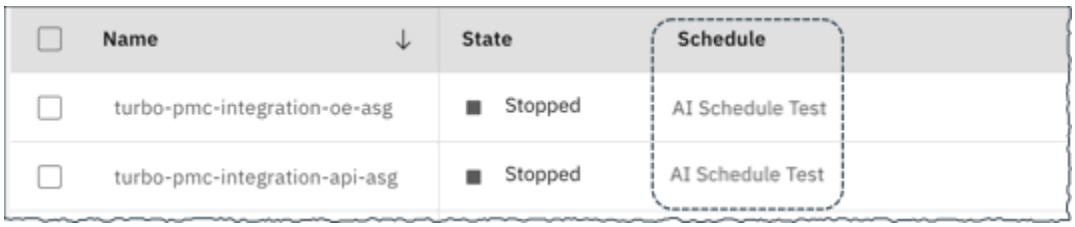


In the Parking Schedule page, choose from a list of previously configured schedules.



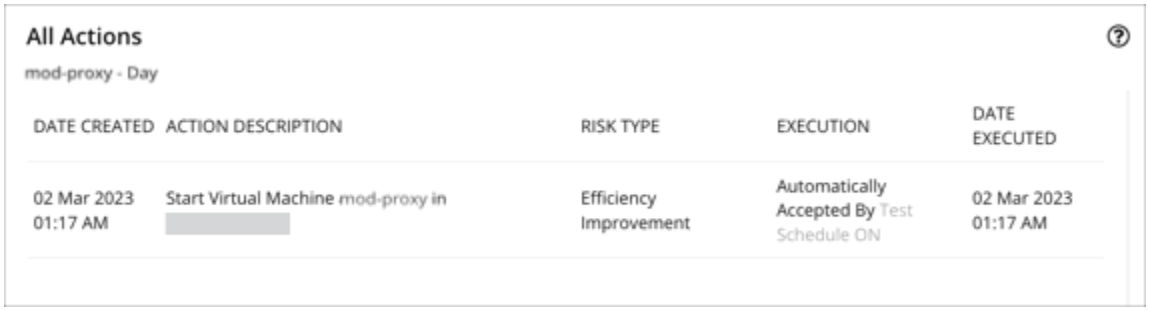
If no schedule is suitable or available, create one by clicking **Create schedule**. Be aware that a schedule created this way will immediately apply to the entities in scope. If this is not your preference, create a new schedule by navigating to **Settings > Schedule** and then clicking **New Schedule > Parking Schedule**.

After attaching a schedule, the name of the schedule appears in the **Schedule** column.



Click the schedule name to view the settings for that schedule. To edit the settings, navigate to **Settings > Schedule** and then open the schedule for editing.

In the All Actions chart, the tabular view shows the following information for scheduled parking actions that have been executed:

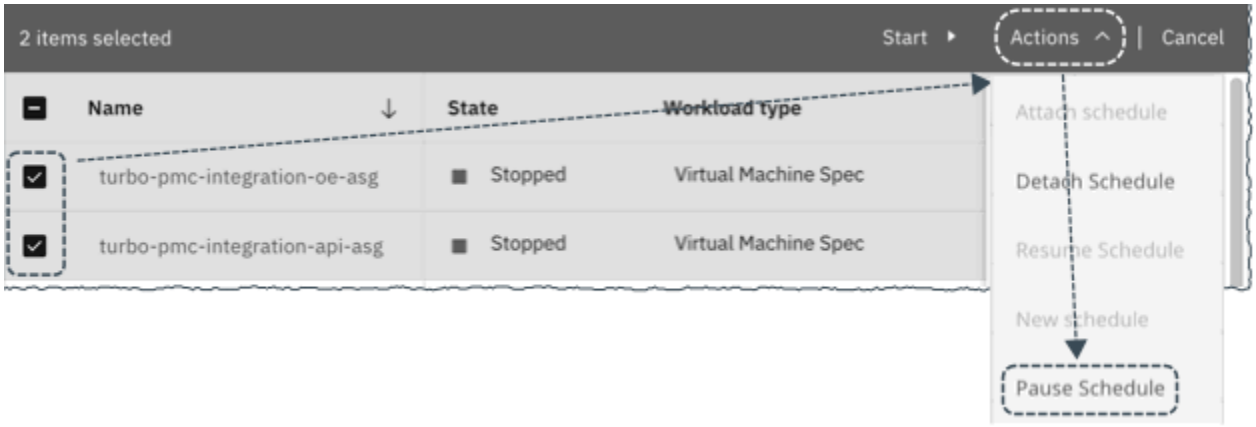


- Execution: Automatically Accepted
- Risk Type: Savings (stop actions) or Efficiency Improvement (start actions)

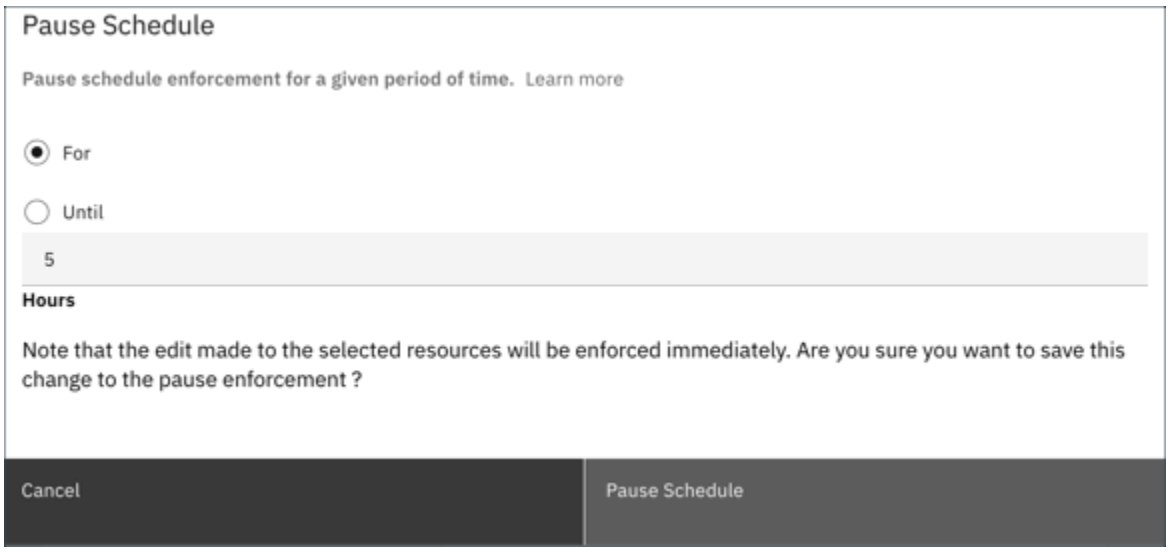
Pausing a parking schedule

You can pause schedule enforcement for a given period of time. For example, you can pause during non-typical work hours or when an unanticipated maintenance window requires automation outside Workload Optimization Manager to start and stop workloads. If the entity was running, it keeps running. If it was stopped, it stays stopped.

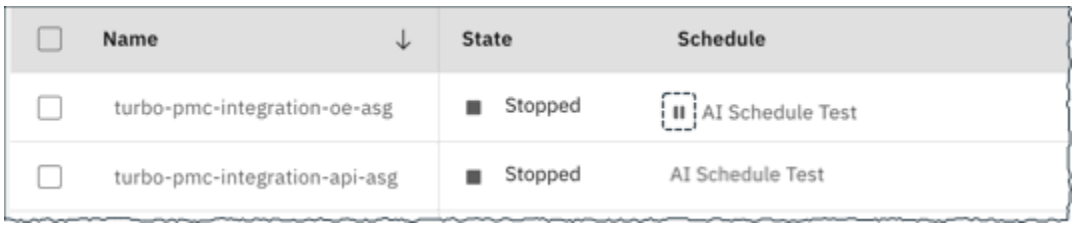
To pause a parking schedule, select one or several entities with a schedule, and then select **Pause Schedule** from the dropdown menu.



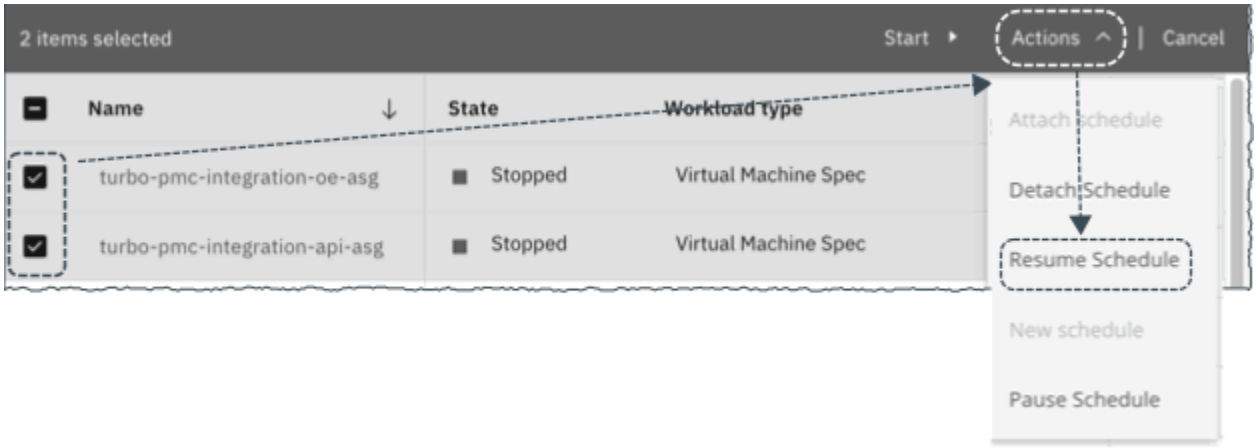
In the Pause Schedule page, specify the duration of the pause.



In the Parking page, an icon appears before the name of the schedule to indicate that it has been paused.



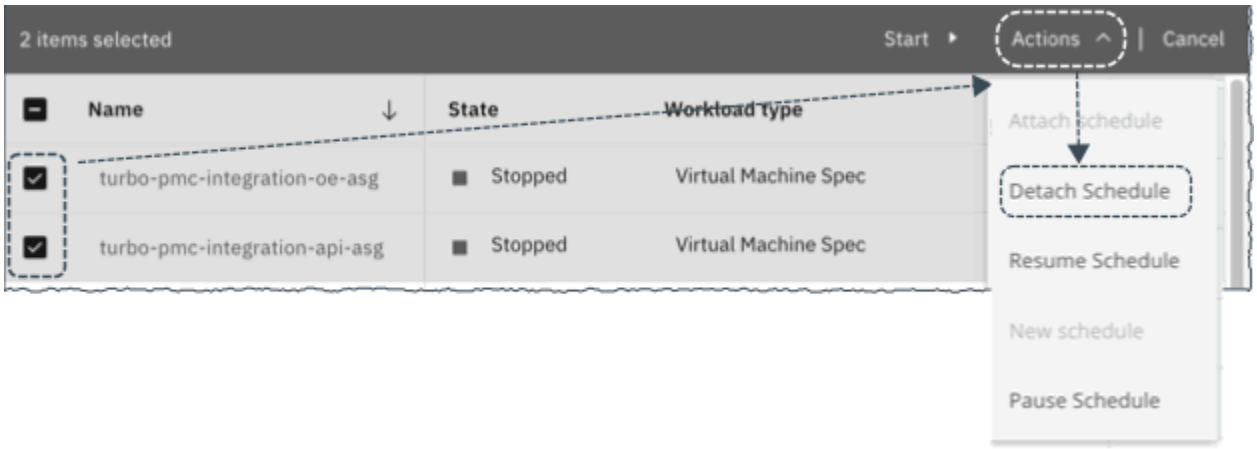
The parking schedule will resume automatically when the pause period is over. To manually resume the schedule, select the affected entities and then select **Resume Schedule** from the dropdown menu.



NOTE:
When you execute a parking action on demand and the affected entity has an existing schedule, a system prompt notifies you to pause the schedule before executing the action. This prevents potential conflicts between on-demand and scheduled parking actions.

Detaching a parking schedule

You can detach parking schedules anytime. Select one or several entities with schedules, and then select **Detach Schedule** from the dropdown menu.



Notice for updated Workload Optimization Manager instances

If you updated your Workload Optimization Manager instance from version 3.5.0 (or older) and noticed that the parking feature is missing in the user interface, perform the following steps in your updated instance:

1. Open `charts_v1alpha1_xl_cr.yaml` for editing.

   ```
   vi /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

2. In the `suspend` section, change the value for `enabled` to `true`.

   ```
   # Suspend service selector
   suspend:
       enabled: true
   ```

3. Apply the change.

```
kubectl apply -f /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
```

4. Restart the UI and API pods.

```
kubectl delete pod <ui-podname>
```

```
kubectl delete pod <api-podname>
```

Place: reserve workload resources

From the Workload Placement Page, you can set up reservations to save the resources you will need to deploy VMs at a future date. Workload Optimization Manager calculates optimal placement for these VMs and then reserves the host and storage resources that they need.

To reserve VMs, you will need to choose a VM template, specify any placement constraints, set how many instances to reserve, and then indicate whether to reserve now or in the future. Because reserved VMs do not yet exist, they do not participate in the real-time market.

About VM templates for reservations

VM templates specify the resource requirements for each reserved VM, including:

- Compute and storage resources that are allocated to each VM
- Consumed factor. This is the percentage of allocated CPU, memory, or storage that the reserved VM will utilize.

For more information about these templates, see VM Template Settings *(on page 1102)*.

About placement of reserved VMs

To determine the best placement for the VMs you want to reserve, Workload Optimization Manager runs a plan that uses the last-generated data in nightly-run headroom plans.

> **NOTE:**
> If you change your environment by adding targets or changing policies, wait until the next run of headroom plans for the affected scope before you create reservations.

When making placement decisions, Workload Optimization Manager considers the following:

- Placement constraints set in the reservation
- Demand capacity

  Workload Optimization Manager calculates demand based on the *resource allocation* and *consumed factor* set in VM templates. For example, to create a reserved VM from a template that assigns 3 GB of virtual memory and a consumed factor of 50%, Workload Optimization Manager calculates 1.5 GB of demand capacity for the reservation.
- Overprovisioned capacity

  For reserved VMs, this corresponds to the resource allocation set in VM templates. Continuing from the previous example, Workload Optimization Manager assumes 3 GB of overprovisioned capacity for a reserved VM created from a template that assigns 3 GB of virtual memory.

  For providers (hosts and storage), Workload Optimization Manager calculates overprovisioned capacity. The default overprovisioned capacity is 1000% for host Mem and CPU, and 200% for storage. A host with 512 GB of memory has an overprovisioned capacity of 5 TB (5120 GB).

Providers must have sufficient *demand* and *overprovisioned* capacity to place a reservation. Workload Optimization Manager analyzes the current and historical utilization of cluster, host, and storage resources to identify viable providers for the VMs

when they are deployed to your on-prem environment. In this way, Workload Optimization Manager can prevent congestion issues after you deploy the VMs.

> **NOTE:**
> Workload Optimization Manager persists historical utilization data in its database so it can continue to calculate placements accurately when market analysis restarts.

The initial placement attempt either succeeds or fails.

- Successful Initial Placements

  If the initial placement attempt is successful, Workload Optimization Manager adds the reserved VM to your inventory.

  In the previous example, a reserved VM that requires 1.5 GB of demand capacity and 3 GB of overprovisioned capacity can be placed on a host with 512 GB of memory (5 TB of overprovisioned capacity), assuming no constraints prevent the placement.

  Note that *actual* and *reserved* VMs share the same resources on providers. This means that provider capacity changes as demand from the actual VMs changes. Workload Optimization Manager polls your environment once per day to identify changes in provider capacity. It then evaluates if it can continue to place the reserved VMs *within the same cluster*, and then shows the latest placement status.

  For example, if the host for a reserved VM is congested at the time of polling, Workload Optimization Manager might decide to move the VM to another host in the cluster that has sufficient capacity. In this case, the placement status stays the same (**Reserved**). Should you decide to deploy the VM at that point, you need to deploy it to the new host. If, on the other hand, there is no longer a suitable host in the cluster, the placement fails and the status changes to **Placement Failed**. Deploying the VM at that point results in congestion. Workload Optimization Manager does *not* retry fulfilling the reservation.

  Reserved VMs are listed on the Workload Placement page. You can also get a list of reserved VMs, with information about each reservation, by making the following API request to the `/reservations` endpoint:

  ```
  GET https://10.10.10.10/api/v3/reservations?status=RESERVED
  ```

  However, reserved VMs are not visible in your application map in the Workload Optimization Manager user interface.

- Failed Initial Placements

  If the initial placement attempt is unsuccessful (for example, if all providers have seen historical congestion), the Workload Placement page shows that the placement has failed and Workload Optimization Manager does *not* retry fulfilling the reservation. You can get a list of VMs for which the placement failed by making the following API request to the `/reservations` endpoint:

  ```
  GET https://10.10.10.10/api/v3/reservations?status=PLACEMENT_FAILED
  ```

Current and future reservations

You can create a current or future reservation from the Workload Placement Page.

- Current Reservation

  Workload Optimization Manager calculates placement immediately and then adds the reserved VMs to your inventory if placement is successful.

  This reservation stays in effect for 24 hours, or until you delete it.

- Future Reservation

  Set the reservation for some time in the future.

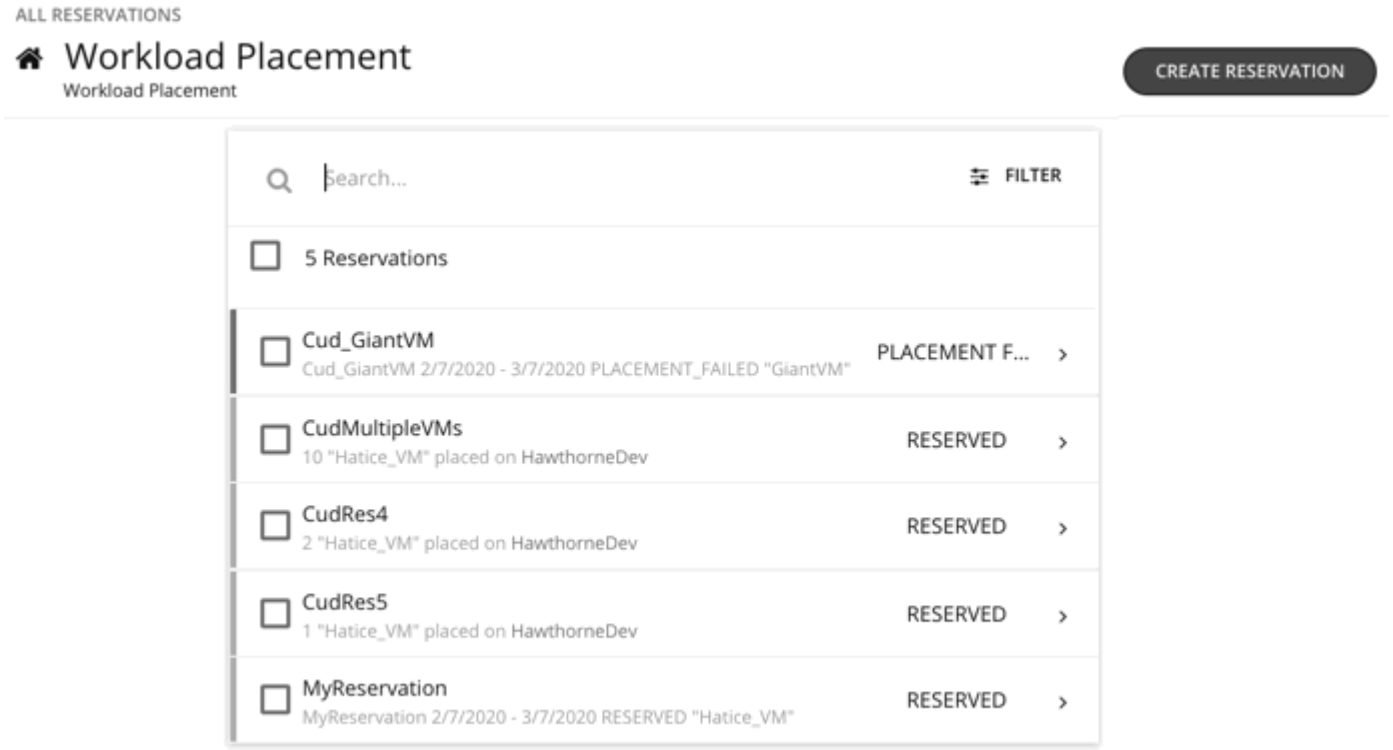
  Workload Optimization Manager does not calculate placement at this time — the future reservation saves the definition, and Workload Optimization Manager will calculate placement at the time of the reservation start date.

  This reservation stays in effect for the duration that you set, or until you delete it.

Displaying the Workload Placement page



To see the reservations that are in effect and to create new reservations, click the **PLACE** button in the Navigation Menu.

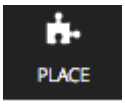


Creating a reservation

Reservations set aside resources for anticipated workload. While a reservation is in the RESERVED state, Workload Optimization Manager continually calculates placement for the reserved VMs.

To create a reservation:

1. Navigate to the Workload Placement page.

2. Create a new reservation.

   In the Workload Placement page, click **CREATE RESERVATION**.

   Workload Optimization Manager displays a list of templates. Choose the template you want, and click **NEXT: CONSTRAINTS**.

3. Optionally, specify placement constraints.

   In the **Constraints** section and choose which constraints to apply to this reservation.

Constraints are optional, but note that these constraints are how you ensure that the template you have chosen is viable in the given locations that Workload Optimization Manager will choose.

The constraints you can choose include:

- Scope

  Choose the datacenter or host cluster that you will limit the reservation to.

- Placement Policy

  This list shows all the placement policies have been created as **Workload Optimization Manager Segments**. Choose which placement policies the reservation will respect.

- Networks

  Workload Optimization Manager discovers the different networks in your environment. Use this constraint to limit workload placement to the networks you choose.

When you are done setting constraints, click **NEXT: RESERVATION SETTINGS**.

4. Make the reservation settings, and create the reservation.

   To finalize the reservation, make these settings:

   - RESERVATION NAME

     The name for the reservation. You should use unique names for all your current reservations. This name also determines the names of the reservation VMs that Workload Optimization Manager creates to reserve resources in your environment. For example, assume the name *MyReservation*. If you reserve three VMs, then Workload Optimization Manager creates three reservation VMs named *MyReservation_0*, *MyReservation_1*, and *MyReservation_2*.

   - VIRTUL MACHINES COUNT

     How many VMs to reserve.

     NOTE:
     You can include up to 100 VMs in a single reservation.

   - RESERVATION DATE

     The time period that you want the reservation to be active. Can be one of:

     – Reserve Now

       Use this to calculate the ideal placement for a workload that you want to deploy today. Workload Optimization Manager begins planning the reservation immediately when you click **CREATE RESERVATION**. The reservation stays in effect for 24 hours – At that time Workload Optimization Manager deletes the reservation.

     – Future Reservation

       This executes the reservation for the date range you specify. Workload Optimization Manager begins planning the reservation on the day you set for START DATE. The END DATE determines when the reservation is no longer valid. At that time, Workload Optimization Manager deletes the reservation.

   When you are finished with the reservation settings, click **CREATE RESERVATION**. Workload Optimization Manager displays the new reservation in the Workload Placement page. Depending on the reservation settings and your environment, the reservation can be in one of the one of the following states:

   - UNFULFILLED

     The reservation request is in the queue, waiting for an ongoing reservation request to complete.

   - INPROGRESS

     Workload Optimization Manager is planning the placement of the reservation workloads.

   - FUTURE

     Workload Optimization Manager is waiting for the START DATE before it will start to plan the reservation.

   - RESERVED

Workload Optimization Manager has planned the reservation, and it found providers for all the VMs in the reservation. As your environment changes, Workload Optimization Manager continues to calculate the placement for the reservation VMs. If at any time it finds that it cannot place all the VMs, it changes the reservation to PLACEMENT FAILED.

- PLACEMENT FAILED

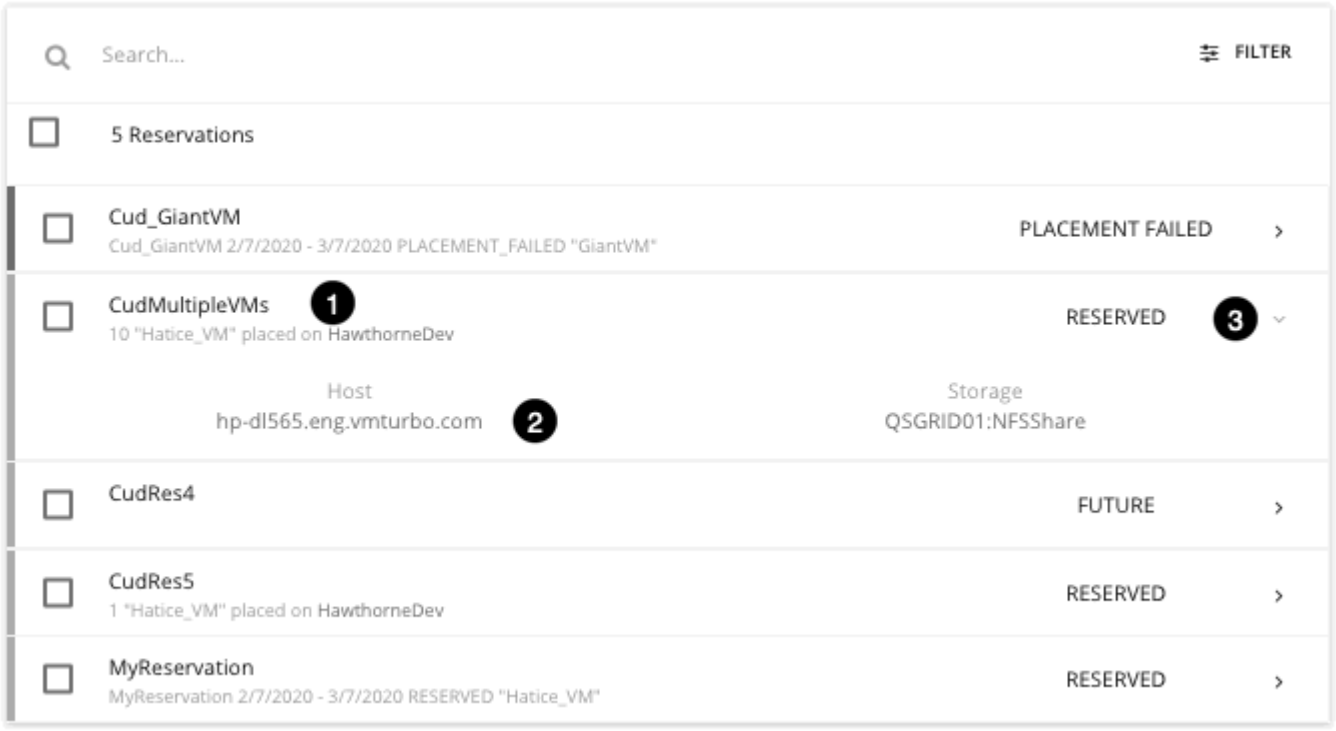
  Workload Optimization Manager cannot place all the reservation VMs. As your environment changes, Workload Optimization Manager continues to calculate placement for the VMs. If at any time it finds that it can place all the VMs, it changes the reservation to RESERVED.

- INVALID

  An error occurred while planning the placement of the reservation VMs.

NOTE:
The list of reservations refreshes whenever you open the Workload Placement page. To see changes in reservation state, navigate away from the page, and navigate back to it again.

Managing reservations



The PLACE page displays the current list of reservations.

1. For an entry in the RESERVED state, you can click the entry name to open the Reservation Settings flyout.
2. To see details about the provider entities or the data center that is hosting the reserved VMs, click the entity name.
3. You can expand items in the list to see some details, or you can click to view the full details.
4. To delete a reservation, select it in the list and click the DELETE icon. This cancels the reservation or deployment.

Deploying workloads to the reserved resources

When you reserve resources, you know that they are available for you to deploy actual VMs in your environment.

1. Note the placement that your reservation calculated.

   Expand the reservation entry in the Workload Placement page and note the hosts and storage that provide resources for your VMs.

2. Delete the reservation.

   Before you deploy the reserved VMs, delete the reservation to free up the Workload Optimization Manager market to manage the placement of the VMs you are about to deploy.

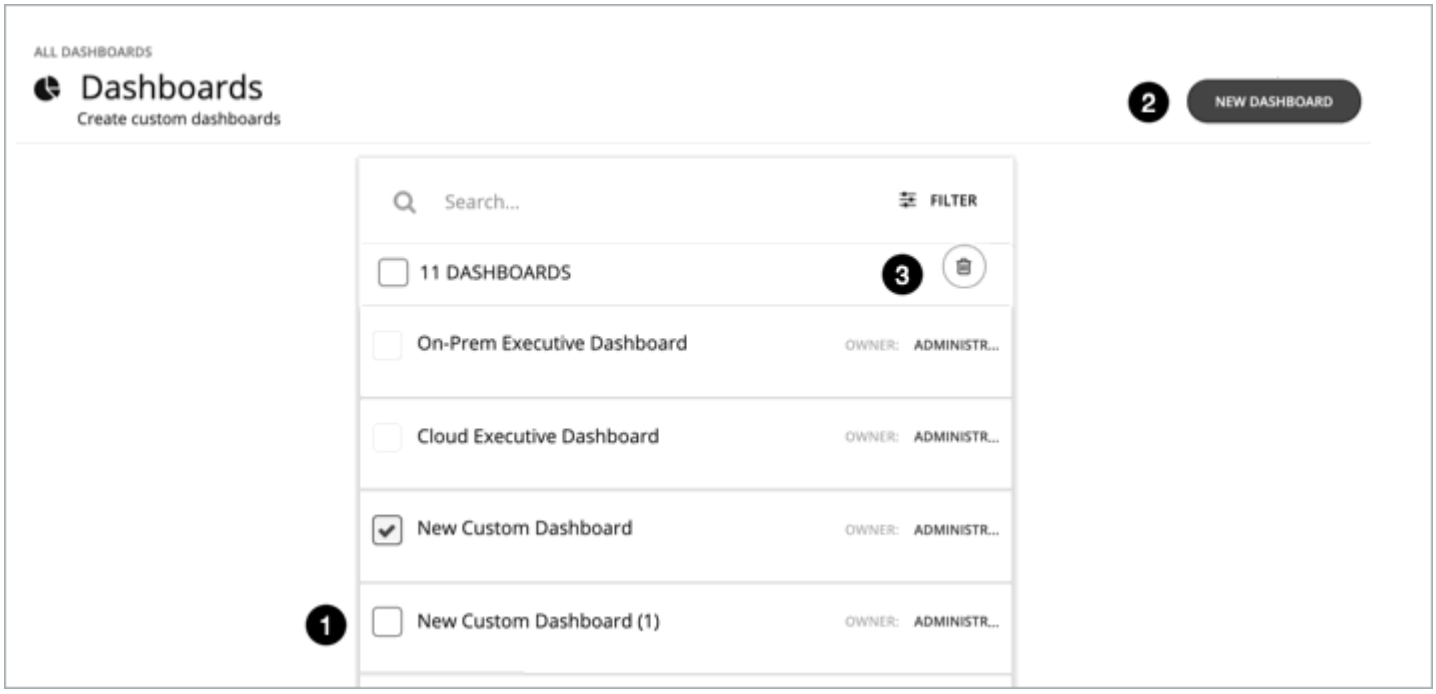
   NOTE:
   When you delete a reservation from the user interface or API, Workload Optimization Manager only marks the reservation for deletion and waits 48 hours before permanently deleting it. You can permanently delete a reservation by using the API's `reservation_force_delete` parameter along with a DELETE call to a specific reservation. When `reservation_force_delete = true`, the system removes the reservation permanently, no matter what state it is in.

3. Deploy the actual VMs.

   In your Hypervisor user interface, deploy the VMs to the hosts and storage that you noted. When you are done, Workload Optimization Manager manages their placement the same as it manages the rest of your environment.

Dashboards: focused views

Dashboards give you views of your environment that focus on different aspects of the environment's health. At a glance, you can gain insights into service performance health, workload improvements over time, actions performed and risks avoided, and savings in cost. For cloud environments, you can see utilization of discounts, potential savings, required investments, and the cost/performance of specific cloud accounts.



The Dashboards page lists all the dashboards that are available to you, including built-in and custom dashboards that your account can access.

1. To view a dashboard, click its name in the list.
2. To create a custom dashboard, click NEW DASHBOARD.
3. To delete a dashboard, select the dashboard and click the delete button

Built-in dashboards give you overviews of your on-prem, cloud, and container environments, showing how you have improved your environment over time.

NOTE:
In charts that show tables, if the table contains more than 500 cells, then the User Interface disables the option to export the chart as PDF. You can still export the chart as a CSV file to load in a spreadsheet.

Built-in dashboards

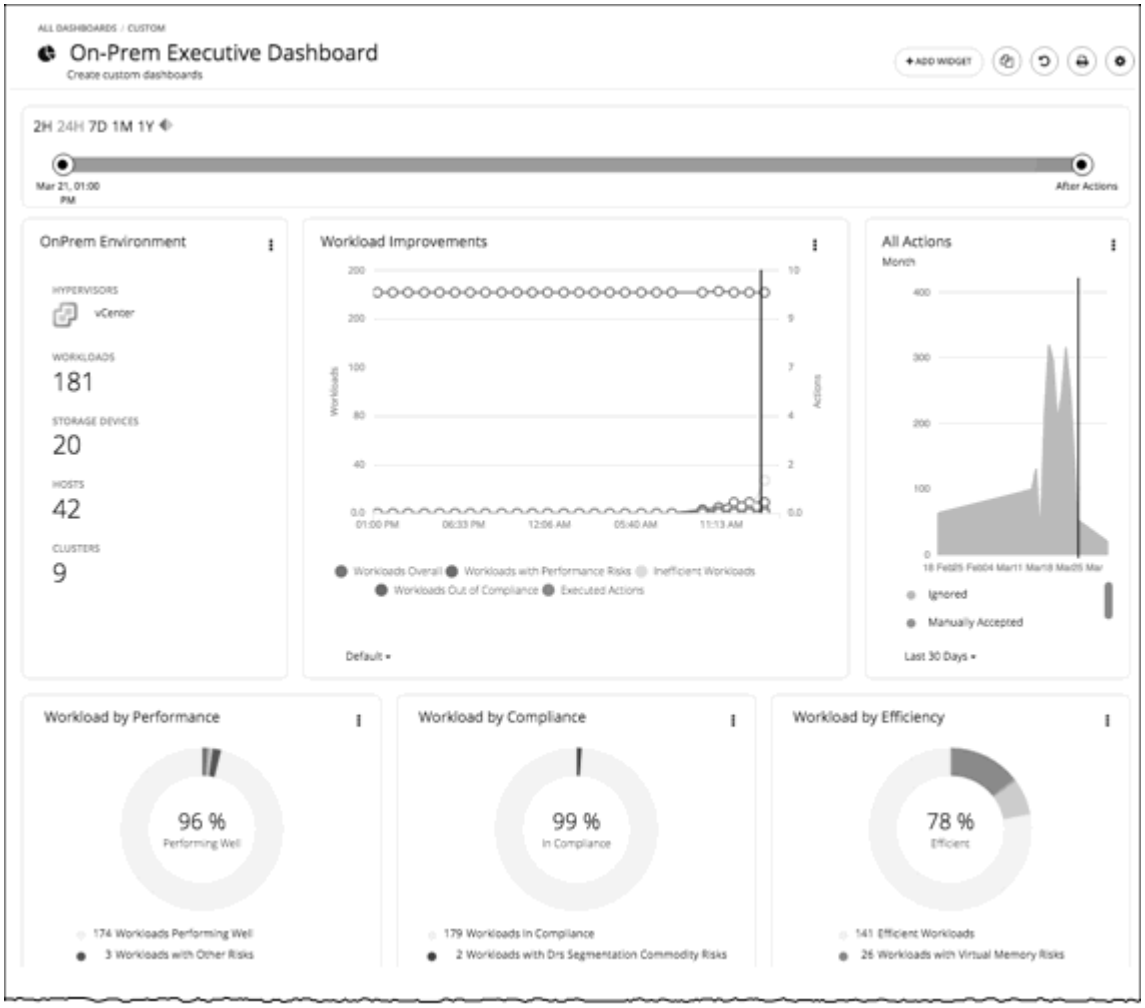
Built-in Dashboards are scorecards of your environment. They demonstrate how well you are improving performance, cost, and compliance, as well as opportunities for further improvements that are available.

Workload Optimization Manager ships with these dashboards:

- On-Prem Executive Dashboard
- On-Prem CPU Ready Dashboard
- Cloud Executive Dashboard
- Container Platform Dashboard

NOTE:
Workload Optimization Manager ships these dashboards with default configurations. To edit a dashboard, you must log in with the administrator user account. Users logged in with that account can add or remove chart widgets, and change widget scopes. For information about editing dashboards, see Creating and Editing Custom Dashboards *(on page 963)*.

On-prem executive dashboard

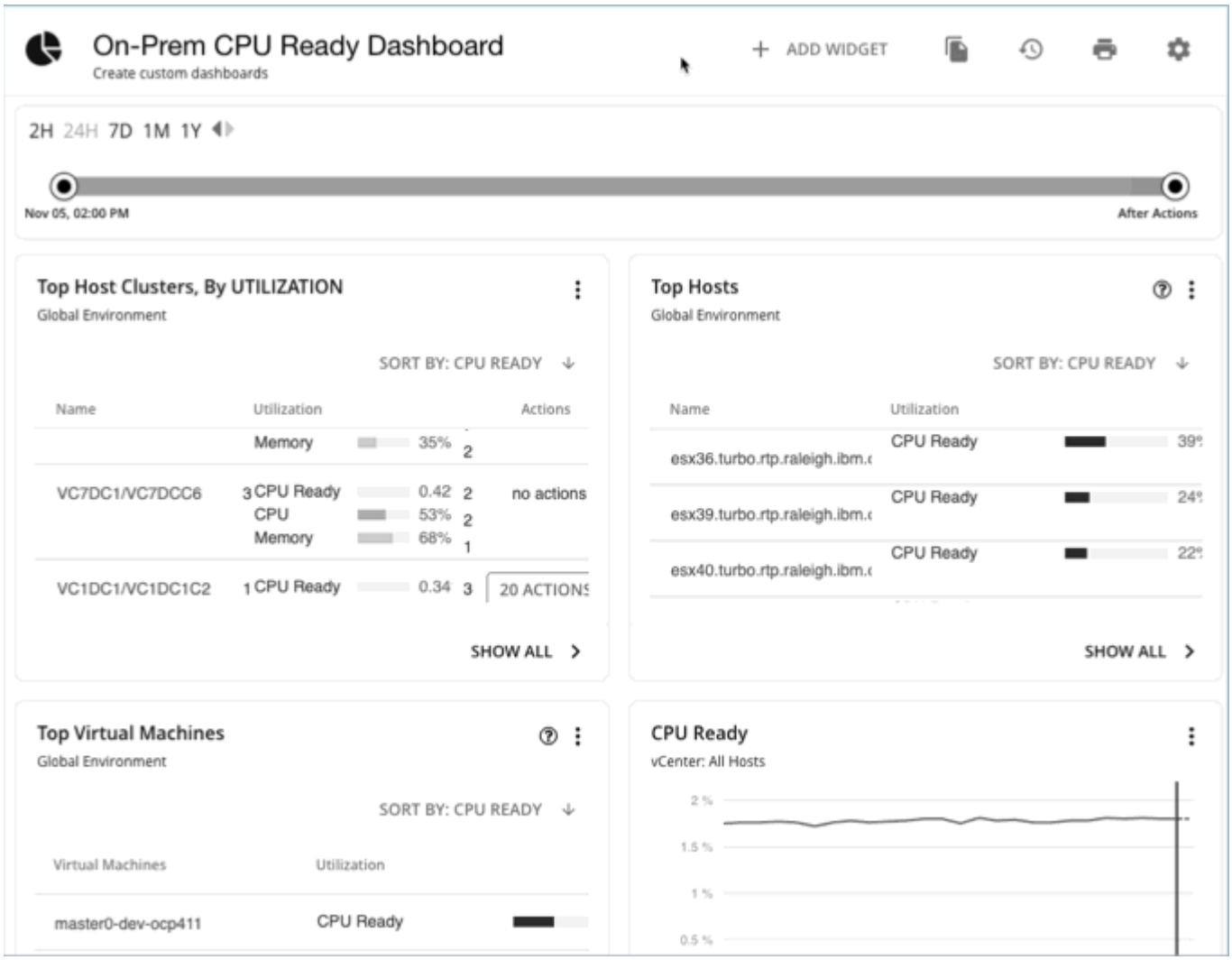


The On-Prem Executive Dashboard shows the overall performance, capacity, and compliance in your on-prem infrastructure. This includes insights into:

- Actions History
  - The **On-Prem Environment** chart widget shows you an overview of your on-prem environment that Workload Optimization Manager is managing and controlling. The chart displays the workloads and the infrastructure that Workload Optimization Manager discovered.
  - The **Workload Improvements** chart widget shows how the efficiency, performance, and policy risks associated with your workloads have disappeared as you have increased your adoption of Workload Optimization Manager Workload Automation. The chart tracks how your workloads have grown as your execution of actions have increased or decreased as your environment achieves and maintains its desired states over time.
  - The **All Actions** chart widget shows the number of actions that Workload Optimization Manager has generated versus the ones executed. This gives you an understanding of where there were more opportunities for improvement that were not taken in the past versus those that are available today.
- Opportunities
  - The Workload by Performance, Workload by Compliance, and Workload by Efficiency chart widgets indicate workload health by showing the risks that are currently in your environment and each classification of those risks. You can click **Show Action** on the chart to reveal all of the outstanding actions that need to be taken to resolve those risks on your workloads.
  - The **Necessary Investments** and **Potential Savings** chart widgets together project how the current actions to improve performance, efficiency, and compliance will impact your costs.

- Current State
  - This chart shows the top host clusters in your on-prem environment by CPU, memory, and storage capacity or utilization. In the default view, the chart shows the top clusters by CPU headroom (available capacity). The chart also shows the time to exhaustion of cluster resources, which is useful for resource planning (for example, you might need to buy more hardware).
  - The **Virtual Machines vs Hosts and Storage** and the **Virtual Machines vs Hosts and Storage -Density** chart widgets show how your overall density has improved in your on-prem environment. A high count of VMs per host or storage means that your workloads are densely packed.

On-Prem CPU ready dashboard



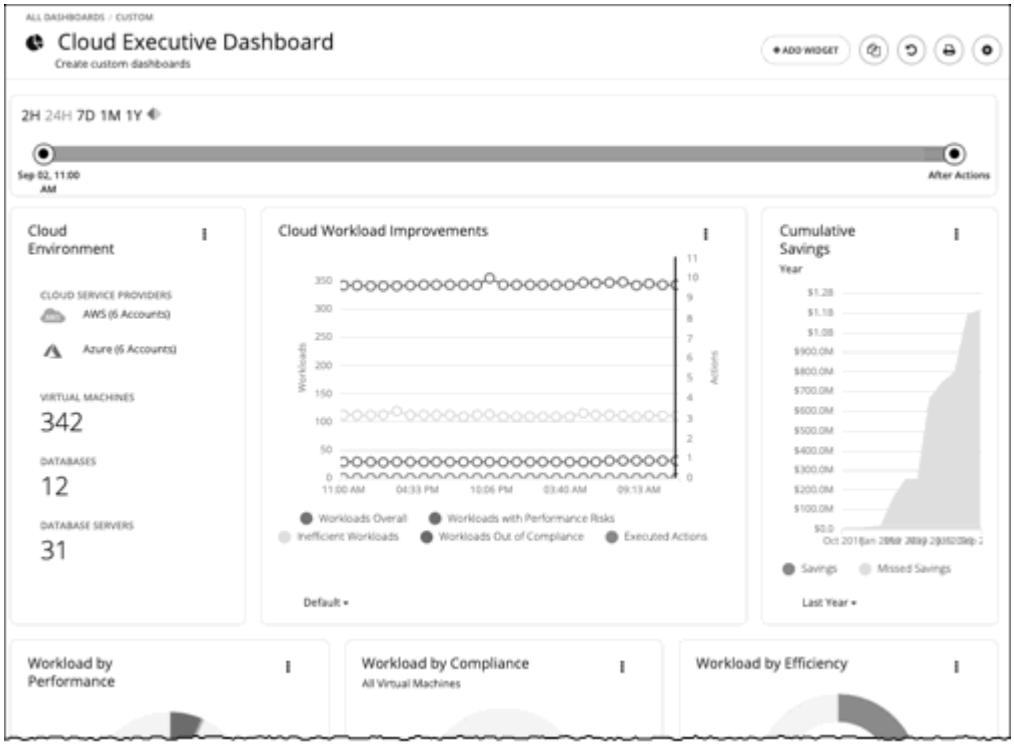
The On-Prem CPU Ready Dashboard shows the Host Clusters, Hosts, and VMs with the highest CPU Ready values. Reviewing the charts in this dashboard can help you tune your CPU Ready settings for the specific workloads that are running in your environment. It includes insights into the following details:

- Host CPU Ready Utilization
  - The **CPU Ready** chart widget shows the overall average utilization of CPU Ready for all hosts that are discovered by vCenter targets in your environment.
  - The **Top Hosts** chart widget shows the hosts with the highest CPU Ready in your environment.
  - The **Top Host Clusters, BY UTILIZATION** chart widget shows the host clusters with the highest CPU Ready in your environment.

- Virtual Machine CPU Ready Utilization
  - The **Top Virtual Machines** chart widget shows the virtual machines with the highest CPU Ready in your environment.
- Effects of Actions on CPU Ready
  - The **Hosts Optimized Improvements** chart widget shows a comparison of the CPU Ready peaks for hosts in your environment before and after actions are executed.

For more information about CPU Ready, see CPU ready chart *(on page 981)*.

Cloud executive dashboard



The Cloud Executive Dashboard shows your overall cloud expenditures and how you can improve performance and reduce cost. This includes insights into:

- Actions History
  - The **Cloud Environment** chart widget shows you an overview of your cloud environment that Workload Optimization Manager is managing and controlling. The chart displays the workloads, cloud service providers, and cloud accounts that you currently have set up as Workload Optimization Manager targets.
  - The **Workload Improvements** chart widget shows how the efficiency, performance, and policy risks associated with your workloads have disappeared as you have increased your adoption of Workload Optimization Manager Workload Automation. The chart tracks how your workloads have grown as your execution of actions have increased or decreased as your environment achieves and maintains its desired states over time.
  - The **Cumulative Savings** chart widget shows you the cost savings for executed cloud actions compared to the cloud actions that you have not executed (missed savings).
- Opportunities
  - The Workload by Performance, Workload by Compliance, and Workload by Efficiency chart widgets indicate workload health by showing the risks that are currently in your environment and each classification of those risks. You can click **Show Action** on the chart to reveal all of the outstanding actions that need to be taken to resolve those risks on your workloads.

  - The **Necessary Investments** and **Potential Savings** chart widgets together project how the current actions to improve performance, efficiency, and compliance will impact your costs.
  - **Cloud Estimated Cost** chart widget shows estimated monthly costs and investments for the cloud. Monthly cost amounts are summarized as amounts with and without actions.
- Current State
  - The **Top Accounts** chart widget shows all of the cloud accounts in your cloud environment and what the utilization is for each account. You can see the number of workloads, estimated monthly costs, saved by actions, and actions taken. In the default view, the chart shows the top cloud accounts and you can click **Show All** button to see all of the accounts. In the Show All list, you can also download the account cost data as a CSV file or PDF.
  - The **Cost Breakdown by Tag** chart widget shows the tags you have assigned to your cloud resources and the costs associated with each of these tagged categories. The **Cost Breakdown by Cloud Service Provider** chart widget is an Expenses chart widget that shows your expenses for each cloud service provider.
  - Usage of Discounts

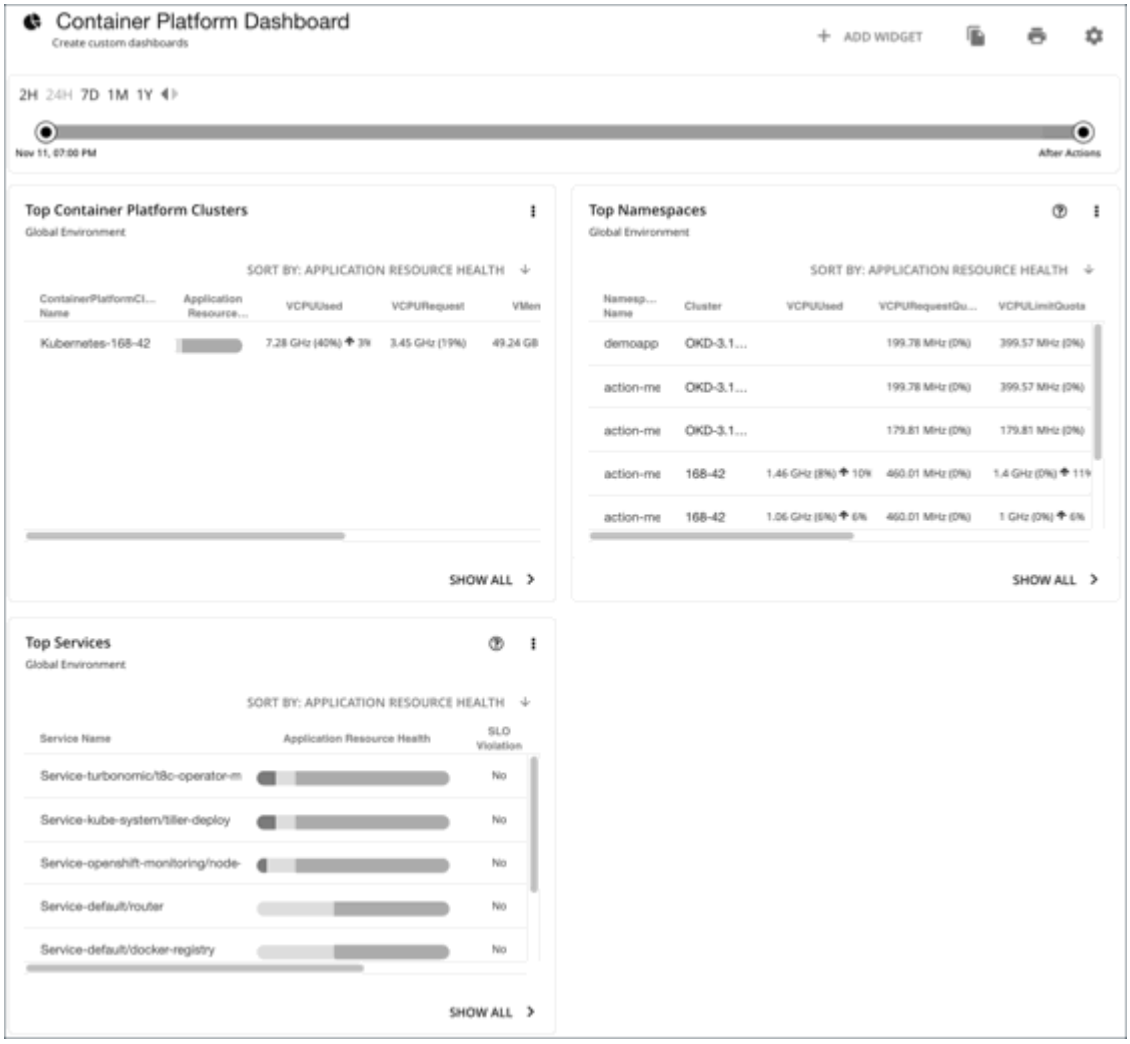
    Discounts reduce cost by offering a subscription-based payment plan. Workload Optimization Manager discovers these discounts and tracks usage patterns to identify workloads that can take advantage of discounted pricing. The Cloud Executive Dashboard shows whether you are getting the most out of your current discounts.

    - Discount Coverage *(on page 1026)*

      This chart shows the percentage of cloud workloads (VMs and RDS database servers) covered by discounts. For VMs covered by discounts, you can reduce your costs by increasing coverage. To increase coverage, you scale VMs to instance types that have existing capacity.
    - Discount Inventory *(on page 1028)*

      This chart lists the cloud provider discounts that you have purchased for your workloads.

Container platform dashboard



The Container Platform Dashboard shows the overall performance, capacity, and health of your container infrastructure. This includes insights into:

- Top Container Platform Clusters

  Assess the health of your clusters and sort them by risk level.

- Top Namespaces

  Identify namespaces that are running out of quota, and how much resources each namespace is using in both quotas and actual utilization.

- Top Services

  Assess the impact of Services on the performance of your applications.

Creating and editing custom dashboards

A custom dashboard is a view that you create to focus on specific aspects of your environment. You can create dashboards that are private to your user account, or dashboards that are visible to any user who logs into your Workload Optimization Manager deployment.

Two common approaches exist for creating custom dashboards:

- Scope First

  You can create a dashboard in which all of the chart widgets focus on the same scope of your environment. For example, you might want to create a dashboard that focuses on costs for a single public cloud account. In that case, as you add chart widgets to the dashboard, you give them all the same scope.

- Data First

  You might be interested in a single type of data for all groups of entities in your environment. For example, each chart widget in the dashboard can focus on Cost Breakdown by Cloud Service, but you set the scope of each chart widget to a different cloud region or zone.

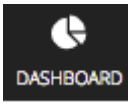
Of course, you can mix and match, according to your needs. You can set any scopes or data sources to the chart widgets in a dashboard to set up whatever organization and focus that you want.

NOTE:
If you set a scope to your Workload Optimization Manager session, the specified scope does not affect your custom dashboards. For information about scoped views, see Working With a Scoped View *(on page 160)*.

Creating a dashboard

To create a custom dashboard:

1. Navigate to the Dashboards Page.

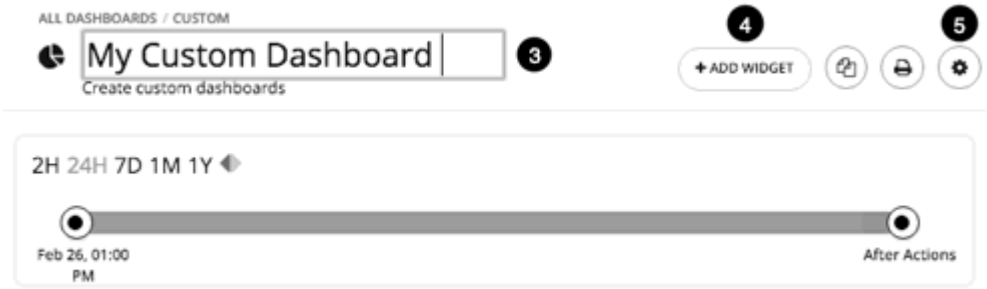
   Click to navigate to the Dashboard Page.

   This page lists all dashboards that are available to you. To view a dashboard, click its name in the list.

2. Create a new dashboard.

   Click **NEW DASHBOARD** to add a new dashboard to your Workload Optimization Manager session. The dashboard appears with a default name and without chart widgets. The time range in the Time Slider is set to 24 hours by default.

3. Name the dashboard.

   Give a name that describes the dashboard. If you will share the dashboard with all Workload Optimization Manager users, the name will help them decide whether to view it.

4. Add chart widgets to the dashboard.

   Add as many chart widgets to the dashboard as you want. See Creating and Editing Chart Widgets *(on page 966)*.

5. Optionally, set the dashboard access.

Click **Gear** to change the setting.

Dashboard access can be:

- **Only Me** – (default) The dashboard is only available to your Workload Optimization Manager user account.
- **All Users** – Every Workload Optimization Manager user can see this dashboard.

As soon as you create a new dashboard, it appears in the list on the Dashboard Page. Users with access to it can click the dashboard name in the list to view it, create a copy of the dashboard, or print the dashboard.

At any time, if you are an administrator or the dashboard owner, you can view and make the following changes to the dashboard:

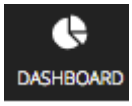
- Add, edit, or delete widgets
- Change the dashboard name
- Change the dashboard access setting

For executive dashboards, only an administrator (username=administrator) can edit an executive dashboard.

Editing a dashboard

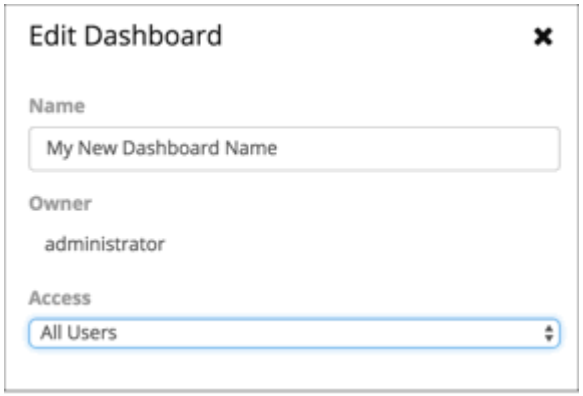
If you have created a dashboard, you can change the name of the dashboard, its access settings, and its chart widgets. To change the chart widgets, see Creating and Editing Chart Widgets *(on page 966)*.

1. Navigate to the Dashboards Page.

   Click to navigate to the Dashboard Page.
2. Click the name of the dashboard that you want to edit.
3. Click **Gear** in the dashboard.

   In the Edit Dashboard flyout, change the dashboard name or set the dashboard access.

   For the dashboard's access, you can set:

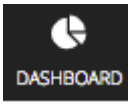
   - **Only Me** – The dashboard is only available to your Workload Optimization Manager user account.
   - **All Users** – Every Workload Optimization Manager user can see this dashboard.
4. When you are done, close the panel.

   Your changes take effect when you close the panel.

Deleting a dashboard

If you are an administrator or the dashboard owner, you can delete a custom dashboard. You cannot delete executive dashboards.

1. Navigate to the Dashboards Page.

   Click to navigate to the Dashboard Page.

   This page lists all dashboards that are available to you.

2. Delete one or more dashboards.

   In the list, choose the check box for each dashboard to delete and click **Trash can**.

Creating and editing chart widgets

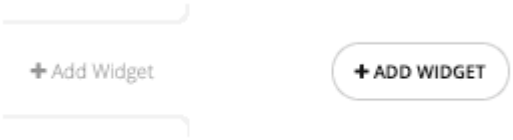
Workload Optimization Manager displays information about your environment in various chart widgets. To focus on the information you need, you can add new chart widgets to scoped views and dashboards, and you can edit existing chart widgets. You can also pull the corners of chart widgets to resize them and change the display order of chart widgets in dashboards.

When you create or edit a chart widget, you can choose various settings. For example, in the Top Utilized chart widget, if you choose Host Clusters as the Entity Type, you can then choose Utilization as the Data Type and Storage Provisioned as the Commodity.

Creating a chart widget

To create a new chart widget:

1. Click **Add Widget** to open the Widget Gallery.

   On a dashboard, click **Add Widget** at the upper-right corner. In a scoped view, click **Add Widget**.

2. Choose a chart widget in the Widget Gallery.

   The Widget Gallery is a list of thumbnail previews of chart widgets.

   You can scroll through the gallery or search it. For example, if you type "Health" in the **Search** field, the results are two chart widgets, Health and Workload Health. You can choose chart widgets from these categories:

   - Actions and Impact
   - Status and Details
   - Cloud
   - On-Prem

   To see the possible displays of a specific chart widget, use the horizontal scroll bar at the bottom of the thumbnail to scroll through the display choices.

   To choose a chart widget to add it to your dashboard, click the thumbnail preview.

   The Widget Preview window with the Edit flyout opens.

3. Configure the settings for your chart widget.

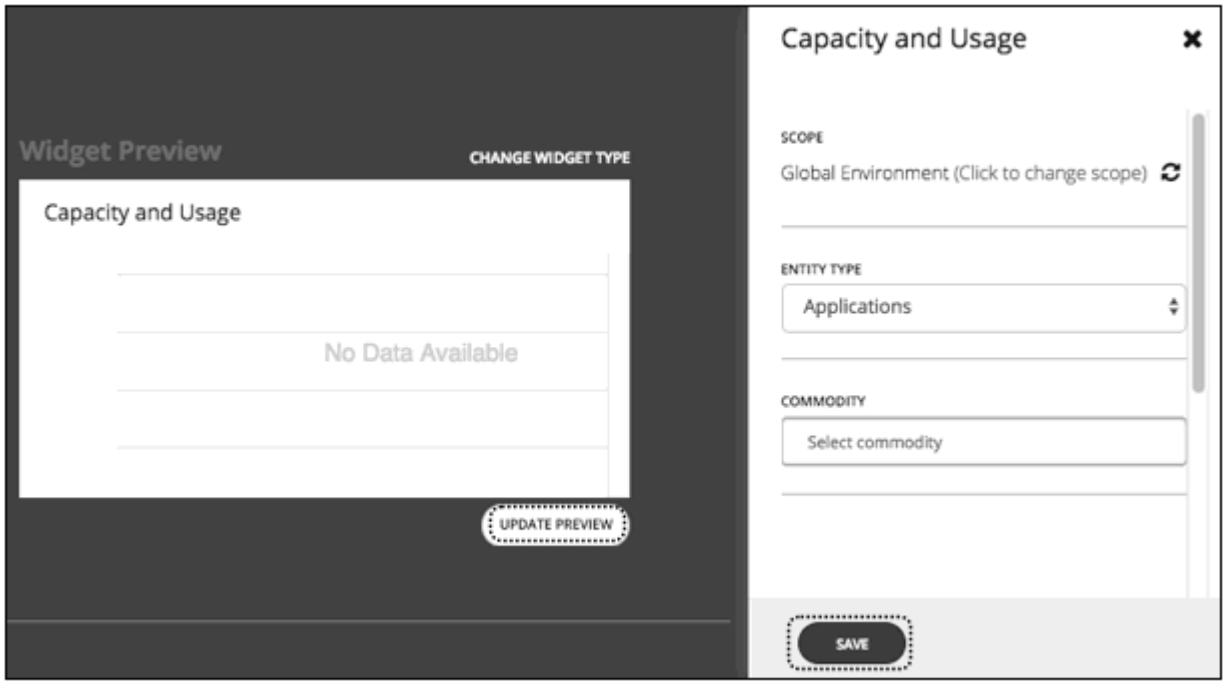
   Chart widget settings determine the data that the chart widget shows.

   In the Edit flyout, choose the settings and click **Update Preview** to display the result in the Widget Preview window.

When you are satisfied with your settings, click **Save**. The chart widget is added to your dashboard.

For information about settings, see Chart Widget Settings *(on page 968)*.

For example:



To delete a chart widget from your dashboard, choose **Delete** in the More options menu at the upper-right corner of the chart widget.

Methods to access chart widget settings

Two methods exist for accessing the chart widget settings in the Edit flyout:

- You can access the settings in the Edit flyout when you add a chart widget to your dashboard after you click a thumbnail preview.
- For an existing chart widget in a dashboard, you can choose **Edit** in the More options menu.

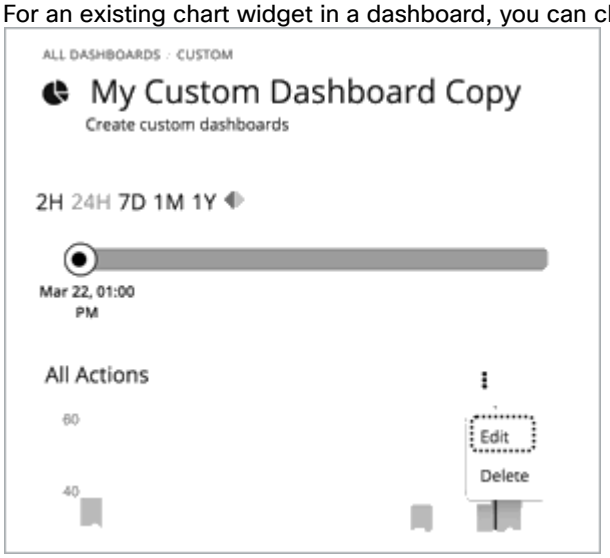


Chart widget settings

Chart widget settings vary according to the type of chart widget. Also, depending on the value that you choose for a setting, more settings appear. The following is a list of frequently used chart widget settings:

- Scope

  The set of entities in your environment that this chart widget represents. By default, the chart widget scope is set to **Global Environment**.

  For every type of chart widget, you have the option to set the chart's scope. To do so:

  1. Click **Click to change scope** to open the Select Scope flyout.
  2. In the Select Scope flyout you can:
     a. Select the scope for the chart: entities, groups, or accounts.

        The ACCOUNTS tab is available depending on the type of chart widget.

        Your choice appears in the **Scope** field.

     b. Search or refine results with a filter.

     c. Display details for the list of entities, groups or accounts.



- Timeframe

  The timeframe for historical data or projections in the chart. Choices for the chart's timeframe are: Default, Last 2 Hours, Last 24 Hours, Last 7 Days, Last 30 Days, and Last Year.

  If you set the timeframe to **Default**, the dashboard Time Slider controls the timeframe setting. For example, if your dashboard Time Slider is set to one month (1M), then all chart widgets with the Default timeframe in that dashboard are set to one month and show information for one month. Note that the dashboard Time Slider does not override the other specific timeframe settings.

- Chart Type

  The chart widget display type. Most chart widgets can display horizontal bar or ring charts. Other display choices can include tabular data, band chart, stacked bar, line, or area charts.

  NOTE:
  For summary charts like horizontal bar and ring charts, when the legend has more than four categories, the remaining categories are represented as a fifth category named "Other."

- Entity Type

  The type of entities or their data that you want to display in this chart widget. Choices vary (for example, Applications, Hosts, Virtual Data Centers, Storage Devices, and so on).

- Commodity

  The resources that you want this chart widget to monitor. Some charts can monitor multiple commodities. Choices vary (for example, CPU, Memory, Virtual Storage, and so on).

Chart types

Workload Optimization Manager provides many different types of charts in the Widget Gallery. To design dashboards, you should be familiar with the data each chart presents. These charts provide information on actions, impact, status of your environment, and details about specific entities, cloud, and on-prem environments.

Actions and impact chart types

These chart widgets provide information on actions, pending actions, risks that you avoided, improvements, and potential savings or investments.

Pending actions charts

Pending Actions charts show the following types of actions that Workload Optimization Manager recommends to improve the current state of your environment.

- Delete – Remove unused entities (for example, datastores on disk arrays or unattached volumes)
- Move – Move an entity from one provider to another to address performance issues or improve infrastructure efficiency
- Optimization – Increase discount coverage for cloud workloads to reduce on-demand costs
- Provision – Add resource providers to the environment to increase capacity
- Reconfigure – Correct a misconfiguration to bring an entity into compliance
- Scale/Resize – Change the resources allocated to workloads to meet demand
- Start – Start a stopped/suspended entity
- Stop/Suspend – Pause an entity for a period of time to increase efficient use of resources

For information on the actions that Workload Optimization Manager generates for specific entities, see Actions by Entity Type *(on page 835)*.

Click **Show all** to see a full list of pending actions in Action Center *(on page 818)*, along with action details and controls to execute actions.

Actions charts

Actions charts keep a history of pending (not executed) and executed actions. These charts use historical data from the Workload Optimization Manager database. You can set the chart to show hourly, daily, or monthly data points.

Filter

You can filter the chart to show **All Actions** (pending and executed actions) or only **Executed Actions**.

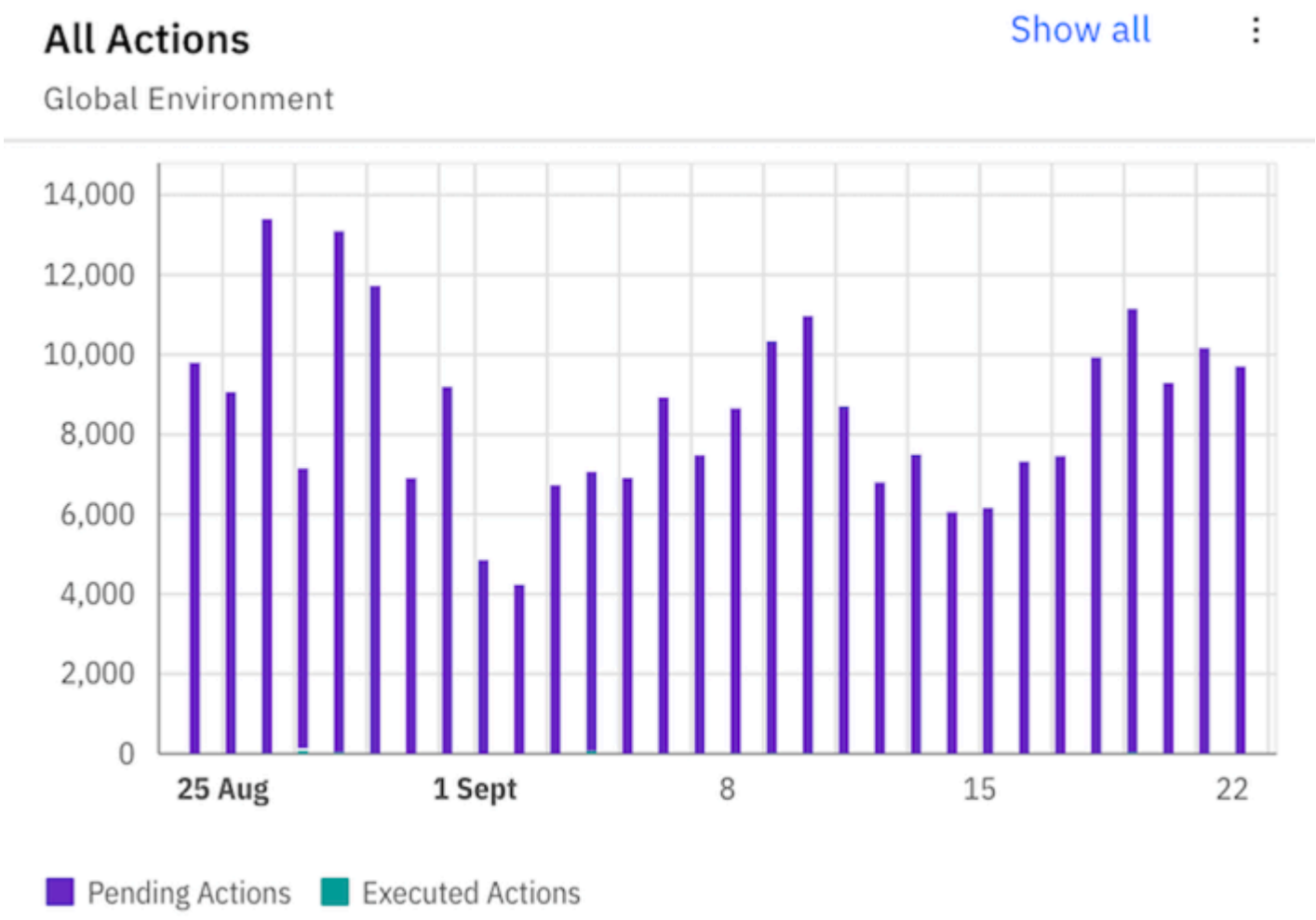
Chart type

You can set the display to:

- Tabular
- Area Chart
- Text
- Stacked Bar Chart

To see the full list of actions, click **Show all**.

Area chart and stacked bar chart



For these charts, each peak represents a time period. Hover over the peak to see the number of unique actions for that time period. In the default view, the bars represent actions per hour in the last 24 hours. For example, the 2:00 PM bar shows actions between 2:00 PM and 2:59 PM.

For these charts, each bar represents a time period. Hover over the bar to see the number of unique actions for that time period. In the default view, the bars represent actions per hour in the last 24 hours. For example, the 2:00 PM bar shows actions between 2:00 PM and 2:59 PM.

A pending action that remains valid for an extended period of time is counted once for each hour, day, and month it remains pending. This also applies to pending actions that go away as conditions in the market change, but are generated again in the

future. Once a pending action is executed, it is counted once (this time, as an executed action) on the hour, day, and month of execution.

Consider the following scenarios.

- An action was generated at 1:25 PM and then executed two hours later at 3:25 PM.
  - For per-hour views (Last 24 Hours or Default), the action will be counted three times – as a *pending* action in the 1:00 PM and 2:00 PM bars, and as an *executed* action in the 3:00 PM bar.
  - For per-day (Last 7 or 30 Days) or per-month (Last Year) views, the action will be counted once (as an executed action) on the day or month of execution.
- An action was generated at 6:20 PM but went away (without being executed) in the next hour. The same action was generated again the next day at 9:10 AM and was executed immediately.
  - For per-hour views, the action will be counted twice – as a pending action in the 6:00 PM bar and as an executed action in next day's 9:00 AM bar.
  - For per-day views, the action will also be counted twice – as a pending action on Day 1 and an executed action on Day 2.
  - For per-month views, the action will be counted once (as an executed action) on the month of execution.

Use these charts to evaluate the rate of action execution, which underscores the importance of executing actions in a timely manner. As pending actions persist, they become more challenging to track and your environment stays in a risky state longer. To reduce potential delays in executing actions, consider action automation.

Timeframe

The chart supports the following timeframes:

- Last 2 hours
- Last 24 hours
- Last 3 days
- Last 7 days
- Last 30 days
- Last 60 days
- Last year
- Custom

Risks avoided charts

As you execute the actions Workload Optimization Manager recommended, you improve your environment's health and avoid performance risks. For public cloud environments, you can avoid unnecessary expenses by executing cost optimization actions. These charts show how many risks you have avoided over time.

Chart Type

You can set the display to:

- Tabular
- Ring Chart
- Horizontal Bar

Optimized improvements charts

Use Optimized Improvements charts in dashboards and plan results to see how the utilization of resources would change if you execute the actions that Workload Optimization Manager recommends. These charts show two bar charts for the entities that

are in scope – one for current consumption and the other for the consumption you would expect to see if you execute all the actions.

- In many of these charts, you can change the commodities on display. To do this, click the More options icon (⋮) and then select **Edit**. In the new screen that displays, go to the **Commodity** section and then add or remove commodities.
- Click **Show all** to see a breakdown of the current chart data by entity (for example, show CPU, Memory, and IO Throughput utilization for each host), or to download chart data as a CSV file.

Potential savings chart

This chart shows potential savings in your cloud expenditure, assuming you execute all the pending actions that Workload Optimization Manager identifies as a result of its analysis. For example, if there are pending actions to scale underutilized virtual machines to less expensive instance types, the Potential Savings chart shows the reduction in monthly costs.

This chart also track discount optimization actions. VM scaling actions may result in freed up capacity on a discounted instance type, which can now be applied to a different VM. Discount optimization actions reflect the potential savings resulting from reassigning the capacity to a different VM. These actions are not executed by Workload Optimization Manager users. They reflect capacity reassignment performed by your cloud provider.

To display a filtered view of the actions, click an action type (for example, **Scale Volumes**) in the ring chart or legend.

Click **Show all** to see a full list of actions in Action Center *(on page 818)*, along with action details and controls to execute actions. By default, the actions are shown in order of largest amounts so you can easily identify the actions that introduce the most savings.

You can also compare instance types, costs, and discount coverage before and after executing the actions, allowing you to easily identify actions with the most savings.

Necessary investments chart

This chart shows the cost of executing all the pending actions that Workload Optimization Manager identifies as a result of its analysis. For example, if some workloads risk losing performance, Workload Optimization Manager recommends scaling actions for the virtual machine to increase capacity. The chart shows how these actions translate to an increase in expenditure.

To display a filtered view of the actions, click an action type (for example, **Scale Volumes**) in the ring chart or legend.

Click **Show all** to see a full list of actions in Action Center *(on page 818)*, along with action details and controls to execute actions. By default, the actions are shown in order of largest amounts so you can easily identify the actions that incur the highest charges.

Status and details chart types

These chart widgets provide information on the status of your environment and details about specific entities.

Health charts

Health charts show the current status of the entity types *(on page 835)* discovered in your environment. For example, you can choose to show the health of all hosts in your on-prem environment or the health of all workloads in a public cloud region.

Chart Type

You can set the display to:

- Text
- Ring Chart
- Horizontal Bar

Basic info charts

The Basic Info charts provide an overview of a single entity or individual Azure resource group that you have chosen as your scope.

Type

You can choose:

- Entity Information

  This chart shows basic information (ID, Name, Type, State, Severity, Target Name, and so on) for the scoped entity or Azure resource group.

- Related Tag Information

  This chart lists any available tag information for the scoped entity or Azure resource group. For example, in a cloud environment, if a virtual machine has tags applied to it, the chart shows those tags for the virtual machine.

Display

The chart shows the information as Tabular.

Capacity and usage charts

These charts list the commodities for the selected entity type, showing the source of the target metric data, their allocated capacity, and how much of that capacity is in use.

Depending on the entity type, you can add different resource commodities that you want to measure. Which commodities you see in the capacity and usage chart depend on the type of entity that you scope to. For example, for a chart of Virtual Machines, you can measure the capacity, usage, and utilization of commodities such as virtual CPU, memory, and storage.

> **NOTE:**
> For a cloud database server, the chart might show incorrect *used* values for vMem and Storage Amount after an action executes. Allow up to 40 minutes for the correct values to display.

Capacity and used sources

When you scope to a specific entity, you can view the source of every commodity collected by a target in Workload Optimization Manager. This chart shows the source of both the available capacity and the utilization of these commodities across all entities in the Workload Optimization Manager environment in the **Capacity Source** and **Used Source** columns. The available sources for the commodity components include:

- Target or Probe name – The source is its parent target. You will see the parent target name (such as vCenter) or "Target" if the data source is not stitched and it originates from one target source.
- Calculation – The source is calculated. See Calculations and derivations *(on page 973)* for more details.
- Policy – The source is a policy setting. See Policy *(on page 976)* for more details.

Calculations and derivations

Review the descriptions in the following table to learn more about the algorithms used to derive the calculated value for the commodity or the source that the values for the commodity are derived from.

Table that describes the calculations of different commodities

Commodity	Metric	Description
Carbon Footprint	Capacity	Multiply the calculated energy capacity by the product of the power usage effectiveness (PUE) and carbon intensity (CI). The following formula expresses the calculation: `(Energy) * [PUE * CI] = Carbon Footprint` For more information, see Carbon Footprint Calculation *(on page 978)*.
	Used	Multiply the calculated energy usage by the product of the power usage effectiveness (PUE) and carbon intensity (CI). The following formula expresses the calculation: `(Energy) * [PUE * CI] = Carbon Footprint`
Connection	Used	Derived from the sold commodity.
CPU	Capacity	Multiply the CPU core MHz by the number of cores for a CPU commodity, if not set. The following formula expresses the calculation: `CPU core MHz * Number of cores for a CPU commodity`
CPU Allocation	Capacity	Multiply the overprovisioned percentage of a commodity by the source commodity capacity. The following formula expresses the calculation: `Commodity Overprovisioned Percentage * Source Commodity Capacity`
	Capacity	Sum of all capacity values of all of the providers of the commodity; if the entity has exactly one provider, it uses the capacity value of a single provider.
CPU Provisioned	Capacity	Multiply the overprovisioned percentage of a commodity by the source commodity capacity. The following formula expresses the calculation: `Commodity Overprovisioned Percentage * Source Commodity Capacity`
	Used	Sum of all used values from the consumers of the commodities.
Energy	Capacity	Energy is calculated by converting the power at 100% CPU utilization from watts to watt-hours. This number is then split among the VMs that are running on a host, based on the number of vCPUs that are allocated to each VM, to calculate the energy capacity for each VM.

	Used	Add the product of the ratio of the VM size to the size of all VMs on the host and the host overhead energy for all VMs on the host to the product of the ratio of the VM CPU utilization to the total CPU utilization for all VMs on the host and the host active energy for all VMs on the host. The following formula expresses the calculation:
		`(VM Size / Total Size for all VMs) * Host overhead energy for VMs +`
		`(VM CPU Utilization / Total CPU Utilization for all VMs) * Host active energy for VMs`
		`= VM energy consumption (in Wh)`
		For more information, see Calculation of Energy Consumed by VMs *(on page 987)*.

Commodity	Metric	Description
Guest Load Application Component	Adjusting the value for the commodities	Subtract the total used values from the consumers of the commodities from the used value sold by the virtual machine. The following formula expresses the calculation: `Used value sold by the VM - Total used values from the consumers of the commodities`
GPU Memory	Capacity	For cloud VMs, the amount of GPU memory provided by an individual GPU card (not the sum of all cards).
	Percentile	For cloud VMs, at each point in time, all GPUs are surveyed and the utilization is taken from the card with the highest memory consumption.
	Used	For cloud VMs, the average memory in use across all GPUs.
Memory Allocation	Capacity	Multiply the overprovisioned percentage of a commodity by the source commodity capacity. The following formula expresses the calculation: `Commodity Overprovisioned Percentage * Source Commodity Capacity`
Memory Provisioned	Capacity	Multiply the overprovisioned percentage of a commodity by the source commodity capacity. The following formula expresses the calculation: `Commodity Overprovisioned Percentage * Source Commodity Capacity`
	Used	Sum of all used values from the consumers of the commodities.
Response Time	Capacity	Derived from the entity disk counts.
Service SLO	Capacity	Derived from the database; if the data is not present in the database, the commodity's capacity value is set to the commodity's used value.
	Used	Average of all used values of the number of replicas of a service.
Storage Access	Capacity	Derived from the provider of the commodity.
	Used	Sum of all used values from the consumers of the commodity.
Storage Access (without a storage target)	Used	Sum of all used values from the consumers of the commodity without a storage target.
Storage Latency	Capacity	Derived from the provider of the commodity.
	Used	Average of all used IOPS-weighted values from the consumers of the commodity.
Storage Latency (without a storage target)	Used	Average of all used IOPS-weighted values from the consumers of the commodity without a storage target.
Storage Provisioned	Capacity	Multiply the overprovisioned percentage of a commodity by the source commodity capacity. The following formula expresses the calculation:

Commodity	Metric	Description
		`Commodity Overprovisioned Percentage * Source Commodity Capacity`
	Used	Sets the used value of the commodity to the capacity value of the Storage Amount commodity if the entity buys a Storage Provisioned commodity and sells the Storage Amount commodity.
Storage Provisioned, Storage Latency, and Storage Access	Used	Merges the identical entities.
Transaction	Capacity	Derived from the database; if the data is not present in the database, the commodity's capacity value is set to the commodity's used value.
VCPU cores	Used	Average of all used values from the consumers of the commodity. The peak value is derived from the maximum values from the consumers of the commodity.
virtual cores	Used value for the number of used virtual cores	Sum of all used values from the consumers of the commodity.
Virtual CPU	Conversion	Converts the CPU units from MHz to millicore.
Virtual CPU Limit Quota	Conversion	Converts the CPU units from MHz to millicore.
Virtual CPU Request	Conversion	Converts the CPU units from MHz to millicore.
Virtual CPU Request Quota	Conversion	Converts the CPU units from MHz to millicore.
Virtual Memory	Used	Verifies that the sum of all used values of consumers of the commodities is lower than the VMEM provided by the virtual machine.

Policy

Review the descriptions in the following table to learn more about the source used to set the policy value for the commodity.

Policy setting	Description
Capacity of the Service SLO commodities	Derived from the SLO values set in the policies.
Capacity of the Response Time commodity	Derived from the value set in the policy.
Capacity of the Response Time commodity for storage targets	Derived from the value set in the policy.
Capacity of the Response Time commodity for Azure cloud target	Derived from the value set in the policy
Capacity of the Storage Access and Storage Latency commodities	Derived from the value set in the policy.

Multiple resources charts

Multiple Resources charts show the historical utilization of commodities for the scoped entity or a group of entities. The vertical bar shows the current moment – plots that extend to the right project utilization into the future.

Commodity

Depending on the entity type, you can add different resource commodities that you want to measure. For example, for a chart of volumes, you can measure commodities such as IO throughput, storage access, and storage amount.

Show peaks

Edit the chart and choose the **Show Peaks** checkbox to include peak information in the chart.

Display

The chart shows the historical utilization and, if chosen, the peak information as a Line chart.

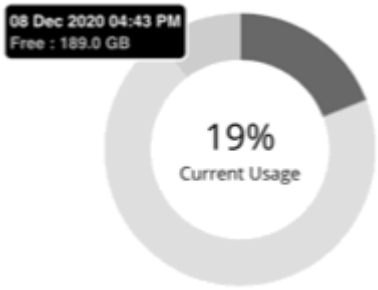
Resources charts

Resources charts show the utilization of a resource over time, for the entities in the chart's scope. The chart title shows the resource that you are plotting, as well as the chart's current scope.

To see finer details about your environment, you can set up charts that show utilization of specific commodities. For example, you can set up a dashboard with a number of Resources charts with their scopes set to the same cluster. Such a dashboard gives you a detailed look at the health of that cluster. Or you could make a dashboard with each chart scoped to a different cluster, but have all the charts show the same resource utilization.

Ring chart

For certain entity types (such as hosts, storage, and disk arrays), you will see a ring chart displayed on the left side of the window that indicates the current overall utilization of a particular resource. Hover over the ring chart to see the following information:



- Free: Available capacity
- Used: Utilized capacity
- Reserved: Unavailable capacity due to utilization constraints

The sum of *Free* and *Used* capacity equals the total allocated capacity.

In addition to showing the currently discovered free and used capacity, Workload Optimization Manager also calculates *Reserved* capacity based on utilization constraints set in policies.

For example, for a cluster with 100 GB of allocated storage, Workload Optimization Manager might discover 80 GB of free capacity, and 20 GB of used capacity. If the cluster is currently applying a storage policy that has a utilization constraint of 90%, then Workload Optimization Manager will show 10 GB of reserved capacity.

Options

Choose **Show Utilization** to see averages and peaks/lows, or **Show Capacity** to see averages and peaks/lows versus capacity.

The **Show Summary** option adds a ring chart to the view, showing the current utilization of the selected commodity.

Chart Type

You can set the following types of display:

- Line Chart

  A line plot showing resource utilization over time. The vertical green bar shows the current moment – Plots that extend to the right project utilization into the future.

- Band Chart

  Lines plot average capacity and average used. The chart shows a band where its thickness indicates peaks and lows.

Carbon footprint chart

Carbon footprint is the measurement of carbon dioxide equivalent (CO2e) emissions for a given entity. Workload Optimization Manager measures carbon footprint in grams.

> **NOTE:**
> An empty chart could be the result of delayed discovery, target validation failure, unavailable data for the given time frame, or an unsupported entity (see the next section for a list of supported entities).

Supported entities

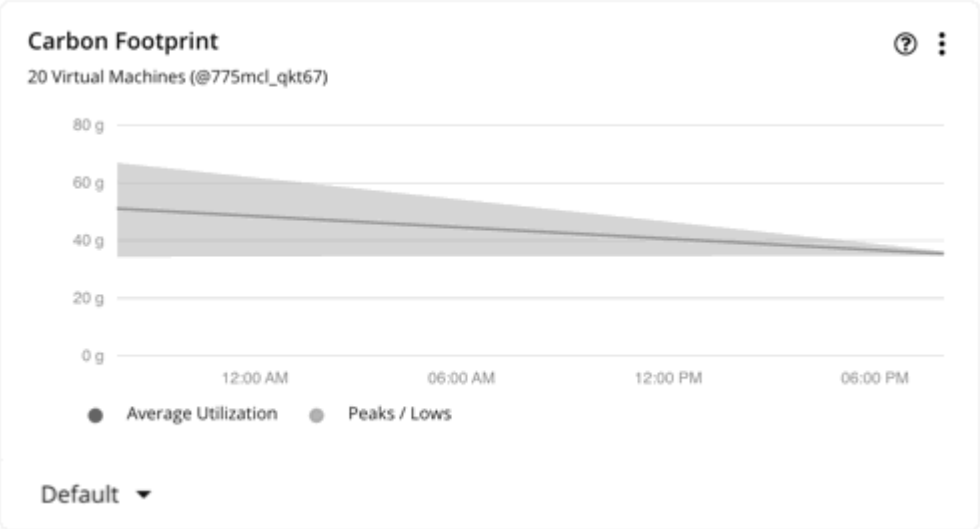
Data is available in the Carbon Footprint chart for on-prem hosts and VMs and cloud VMs.

For cloud VMs, Workload Optimization Manager supports the following providers and services:

- Amazon Web Services (AWS)
  - Amazon Elastic Compute Cloud (EC2)
- Google Cloud Platform (GCP)
  - Google Compute
- Microsoft Azure
  - Azure Compute

Carbon footprint calculation for on-prem hosts and VMs

Workload Optimization Manager collects energy-related data from hosts and VMs at 10-minute intervals, and then uses that data to calculate carbon footprint. When you set the scope to one or several hosts or VM discovered from supported targets, the data that Workload Optimization Manager calculated displays in the Carbon Footprint chart.



The chart shows average and peak/low values over time. Use the selector on the chart to change the time frame.

Workload Optimization Manager also relies on host power data for the supported targets. This data is mandatory for Energy and Carbon Footprint calculations.

The calculation for carbon footprint can be expressed as follows:

```
(Energy Consumption) * [(Power Usage Effectiveness) * (Carbon Intensity)] =
Carbon Footprint
```

Where:

- Energy Consumption is the consumption data collected from the entity.
- Power Usage Effectiveness (PUE) is a ratio that describes how efficiently a computer data center uses energy; specifically, how much energy is used by the computing equipment. PUE is the ratio of the total amount of energy used by a computer data center facility to the energy delivered to computing equipment. The closer PUE is to 1, the more efficient the computer data center.
- Carbon Intensity (CI) is a measurement of how 'clean' electricity is. It refers to how many grams of carbon dioxide (CO2) are released to produce 1 watt-hour (Wh) of electricity. Electricity that is generated using fossil fuels is more carbon intensive, as the process by which it is generated creates CO2 emissions. Renewable energy sources, such as wind, hydro, or solar power produce next to no CO2 emissions, so their carbon intensity value is much lower and often zero.

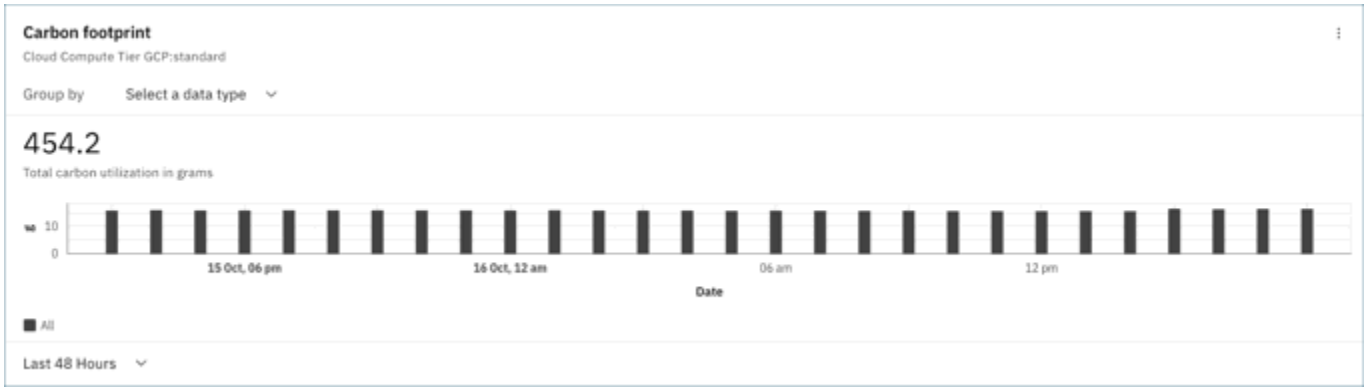
By default, PUE is set to 1.5, while CI is set to 0.25 g/Wh. These values appear in the default policy *(on page 797)* for Data Center entities. You can modify these values directly for a global effect, or set specific values in custom automation policies for Data Centers.

For example, if energy consumption for a host is 50 Wh, carbon footprint will be calculated as follows:

```
(50) * [(1.5) * (.25)] = 18.75 grams of CO2e
```

Calculation of carbon footprint for public cloud VMs

Workload Optimization Manager provides the capability to measure and visualize the current and historical carbon footprint of cloud VMs or groups of VMs (account, region, zone, and more) at an hourly granularity.



The chart shows a series of interval values over time. Use the selector on the chart to change the time frame.

The methodology to calculate the carbon footprint for cloud VMs is based on the Cloud Carbon Footprint methodology, with some precision enhancements, such as using the uptime of a VM and the actual CPU utilization that Workload Optimization Manager collects.

The calculation can be expressed as follows:

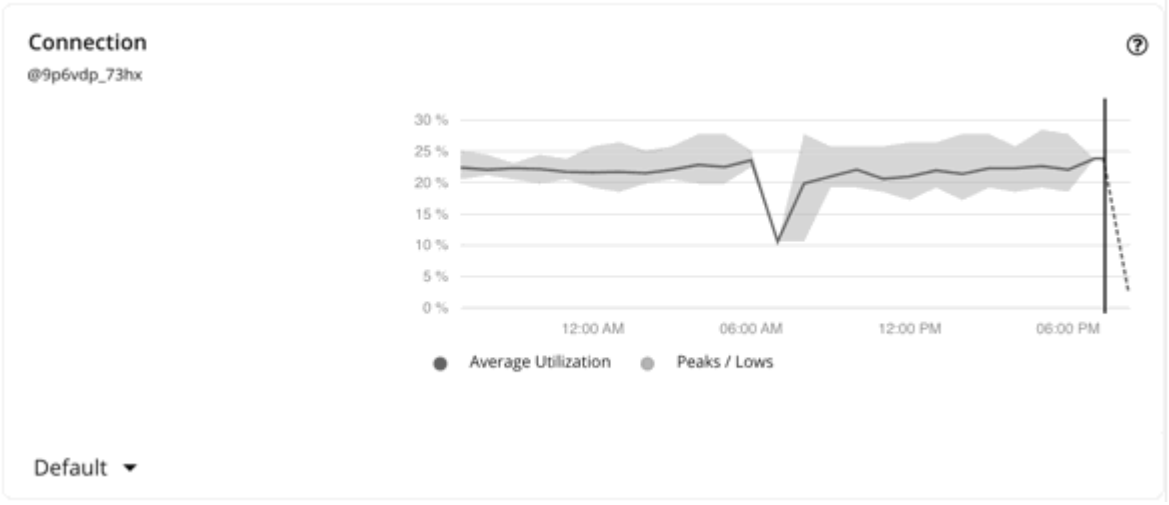
```
Carbon footprint in grams = calculated energy consumption in watt-hours x grid emmission x 1000
```

- **Calculated energy consumption:** The product of vCPU hours, average watts, and power usage effectiveness (PUE). For more information, see Calculation of energy consumed by public cloud VMs *(on page 989)*
- **Grid emission (mt/kWh):** A grid emission factor in metric tons per kilowatt-hour based on cloud service provider and region. For more information, see Grid emissions factors.

Connection chart

Connection is the measurement of database connections utilized by applications.

Workload Optimization Manager collects connection data from Database Servers discovered by Databases, APM, and Cloud targets. When you set the scope to one or several Database Servers, the data that Workload Optimization Manager collected displays in the Connection chart.



The chart shows average and peak/low values over time. Use the selector on the chart to change the time frame.

NOTE:
An empty chart could be the result of delayed discovery, target validation failure, unavailable data for the given time frame, or an unsupported Database Server (see the next section for a list of supported Database Servers).

Supported database servers

Data is available in the Connection chart for Database Servers discovered via the following targets:

Target	Supported database servers
AWS	RDS
Azure	SQL
AppDynamics	MongoDB
Instana	Oracle
MySQL	MySQL
Oracle	Oracle
JBoss	All Database Servers discovered from the target
SQL	SQL
Tomcat	All Database Servers discovered from the target
WebLogic	All Database Servers discovered from the target
WebSphere	All Database Servers discovered from the target

Effect on memory resize/scale actions

Workload Optimization Manager uses connection data to generate memory resize actions for on-prem Database Servers.

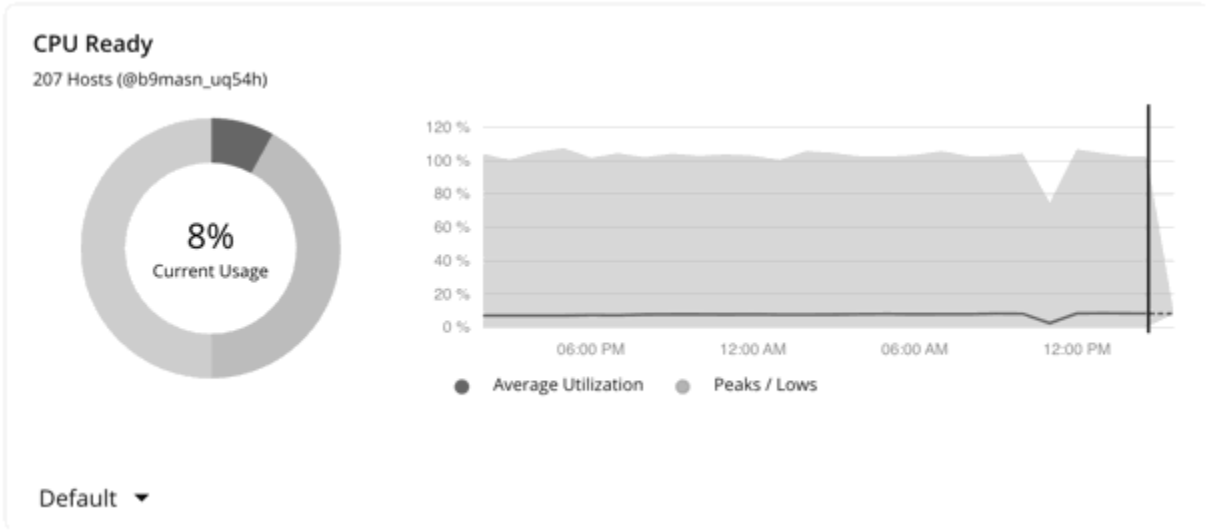
For cloud Database Servers, Workload Optimization Manager uses connection data as a constraint when generating scale actions. For details about scale actions, see Database Server Actions *(on page 685)*.

CPU ready chart

CPU Ready is the measurement of time that a VM is ready to use CPU, but is unable to schedule physical CPU time because host CPU resources are busy.

Workload Optimization Manager collects CPU Ready data from hosts and VMs that are discovered by vCenter targets. It then calculates CPU Ready capacity and utilization to make accurate VM move recommendations.

When you set the scope to one or several VMs or hosts, the CPU Ready chart displays CPU Ready data.



Host CPU ready capacity

Workload Optimization Manager calculates host CPU Ready capacity by using following formula:

```
Host Logical Processors * 20 = Host CPU Ready Capacity
```

The following values are represented in the formula:

- `Host Logical Processors` is the number of logical CPU cores on a host.
- `20` is the standard Ready Queue interval (in seconds) at which the hypervisor measures metrics.

For example, if a host has 10 cores and applies the default host policy, Workload Optimization Manager calculates capacity according to the following formula:

```
10 * 20 = 200
```

Host CPU ready utilization

CPU Ready utilization is the measurement of capacity that is in use. Workload Optimization Manager calculates host CPU Ready utilization by using following formula:

```
Host CPU Ready Average / Host CPU Ready Capacity = Host CPU Ready Utilization
```

The following values are represented in the formula:

- `Host CPU Ready Average` is the average of the 20-second CPU Ready summation values that are collected from vCenter every 10 minutes. Each value is expressed in number of milliseconds.
- `Host CPU Ready Capacity` is the capacity that is calculated by Workload Optimization Manager. For more information, see the previous section.

For example, if raw utilization on a host is 40 and CPU Ready capacity is 100, Workload Optimization Manager uses the following calculation for host CPU Ready utilization.

First, to get the `Host CPU Ready Average` value, Workload Optimization Manager collects 20-second CPU Ready summation values from vCenter during the poll period, for example:

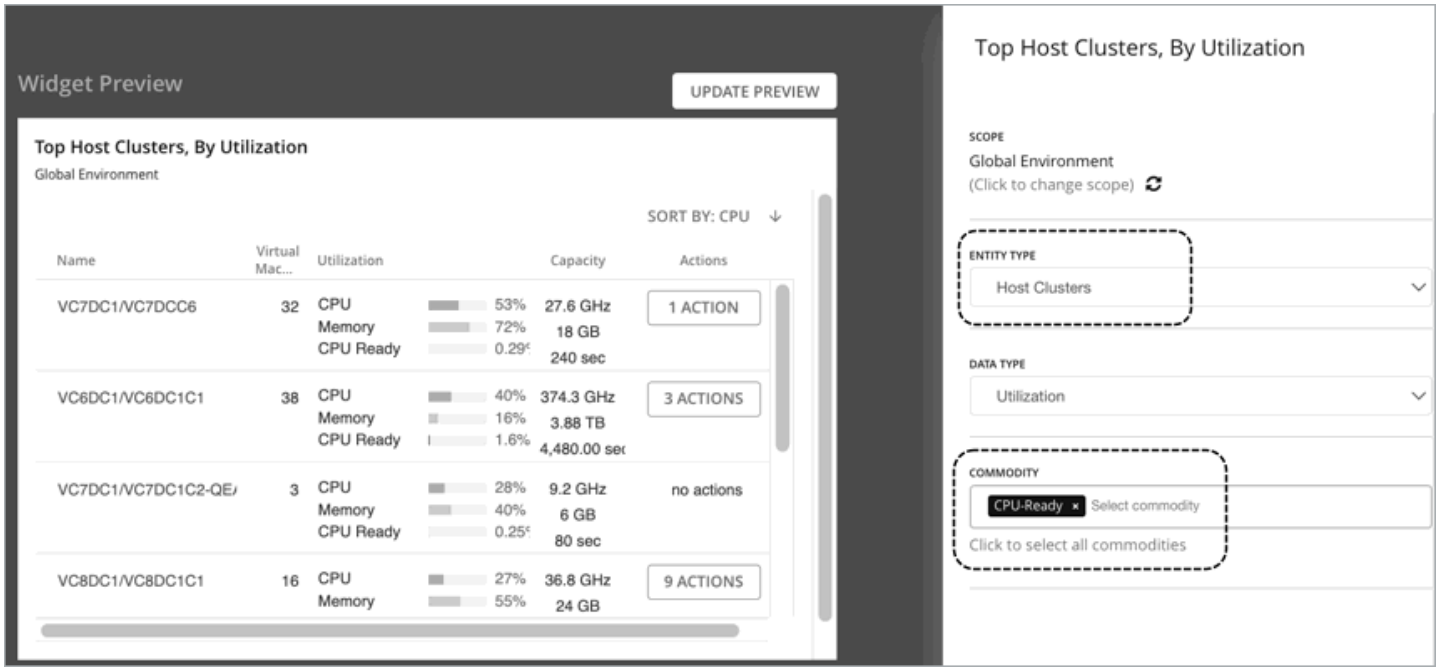
```
AVERAGE(500, 100, 1000, 1500, 1500, 1500, 500, 100, 1000, 1500, 1500, 1500, 500, 100,
1000, 1500, 1500, 1500, 500, 100, 1000, 1500, 1500, 1500, 500, 100, 1000, 1500, 1500,
1500) = 920 ms = 0.92 s
```

Workload Optimization Manager then divides the Host CPU Ready Average value by Host CPU Ready Capacity, which in this example is 100, to get a host CPU Ready utilization value of 0.92, or about 9%

```
0.92/100 = 0.0092 (0.9%)
```

To view the hosts with the highest CPU Ready, add the **Top Utilized** chart to your dashboard. When you configure the chart, select **Hosts** as the entity type, and **CPUReady** as the commodity.

To view the host clusters with the highest CPU Ready, add the **Top Utilized** chart to your dashboard. When you configure the chart, select **Host Clusters** as the entity type and **CPUReady** as the commodity.



This chart shows any pending action to move VMs out of a host cluster due to CPU Ready congestion.

VM CPU ready capacity

Workload Optimization Manager calculates VM CPU Ready capacity by using the following formula:

```
VM Logical Processors * 20 = VM CPU Ready Capacity
```

The following values are represented in the formula:

- `VM Logical Processors` is the number of vCPUs on a VM.
- `20` is the standard Ready Queue interval in seconds at which the hypervisor measures metrics.

For example, for a VM with 2 vCPUs, Workload Optimization Manager uses the following calculation for VM CPU Ready capacity:

```
2 * 20 = 40
```

VM CPU ready utilization

CPU Ready utilization is the measurement of capacity that is in use. Workload Optimization Manager calculates VM CPU Ready utilization by using following formula:

```
Raw Utilization / VM CPU Ready Capacity = VM CPU Ready Utilization
```

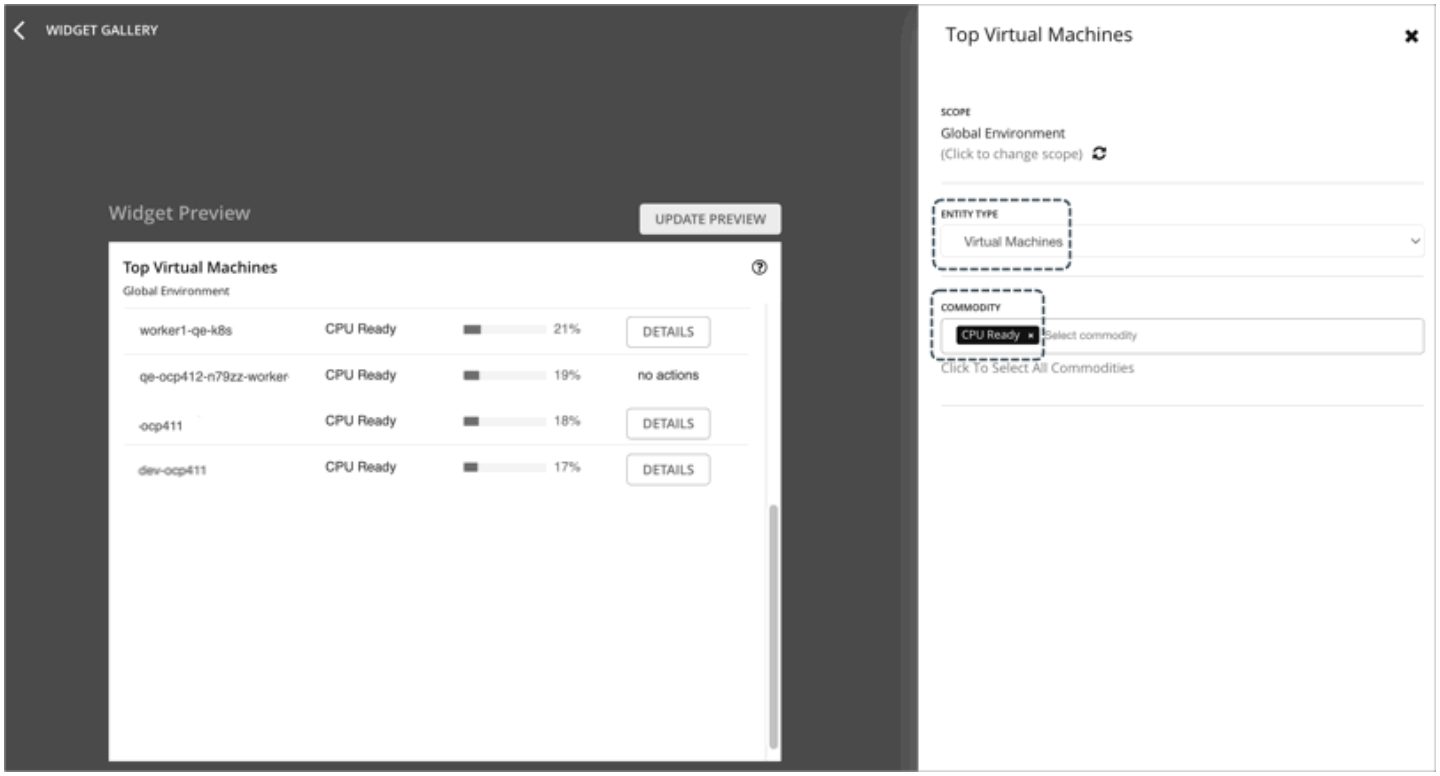
The following values are represented in the formula:

- `Raw Utilization` is the utilization value (in seconds) that is collected from vCenter.
- `VM CPU Ready Capacity` is the capacity that is calculated by Workload Optimization Manager. For more information, see the previous section.

For example, if raw utilization on a VM is 40 and CPU Ready capacity is 100, Workload Optimization Manager uses the following calculation for VM CPU Ready utilization:

```
40/40 = 1 (100%)
```

To view the most utilized VMs, add the **Top Utilized** chart to your dashboard. When you configure the chart, select **virtual machines** as the entity type, and **CPUReady** as the commodity.



This chart shows any pending action to move a VM to a different host due to CPU Ready congestion.

Effect on VM move sctions

Workload Optimization Manager considers host CPU Ready utilization when it makes placement decisions for VMs. `Ready Queue Utilization` is a host policy setting for the percentage of utilization that Workload Optimization Manager considers as full utilization. For example, if utilization reaches 50%, Workload Optimization Manager considers Ready Queue to be fully utilized and the market might generate a move action to remedy the high Ready Queue utilization condition. The default value for this policy setting is 50%. This percentage is suitable for most environments. For environments where applications are sensitive to latency, you can reduce the percentage so that Workload Optimization Manager is more sensitive to CPU Ready when it places VMs. However, CPU Ready utilization is only one among many factors that go into move decisions. In some environments, other factors might offset CPU Ready utilization concerns.

Logical Processors are considered when Workload Optimization Manager places VMs on hosts. For example: If a VM has 64 vCPUs, it must run on a host with at least 64 logical cores. Workload Optimization Manager does not attempt to move a VM to a host with fewer Logical Processors. For VMs, Logical Processors Capacity and Used values equal the number of vCPUs that are configured on the VM. For hosts, Logical Processor Capacity equals the number of host logical cores and the Used value equals the sum of all VM Logical Processors (vCPUs). In this way, the Host Logical Processor Utilization represents the ratio of VM vCPUs to the host logical cores. Overprovisioned hosts are known to contribute to increased risk of CPU Ready latency. Be sure to take any resize down actions on VMs where CPU Ready is observed to reduce CPU overprovision.

NOTE: In VMware environments, the best practice is to keep CPU Ready values as low as possible. A CPU Ready value of 3% indicates a potential performance risk for most applications. If the CPU Ready value is 5% or greater, expect a significant performance impact.

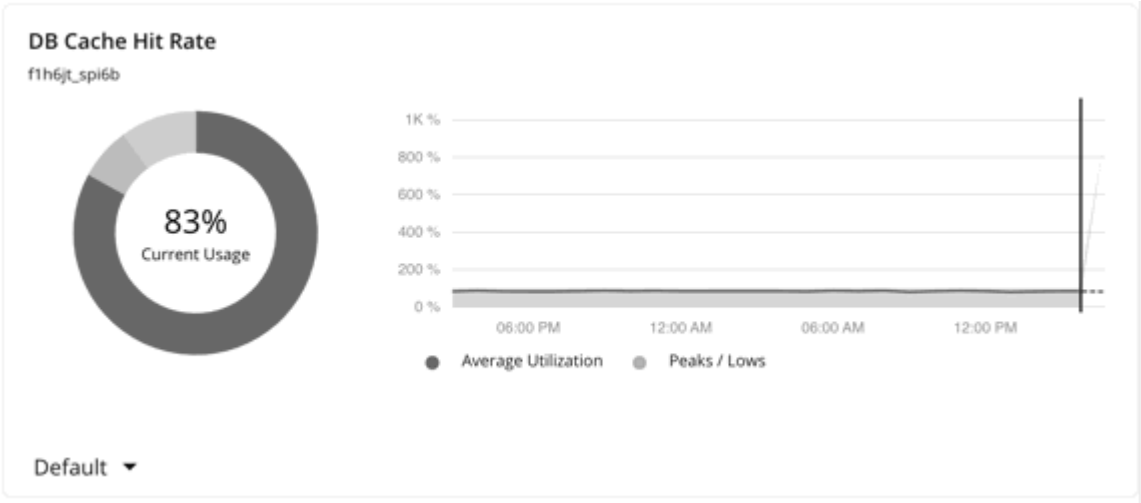
CPU Ready Dashboard: identifying vPU ready risks

A default dashboard *(on page 960)* is available for reviewing CPU Ready in your environment. This dashboard shows the Host Clusters, Hosts, and VMs with the highest CPU Ready values. Reviewing the charts in this dashboard can help you tune your CPU Ready settings for the specific workloads that are running in your environment. To view this dashboard, select **DASHBOARD** from the main navigation menu and click **On-Prem CPU Ready Dashboard**.

DB cache hit rate chart

DB cache hit rate is the measurement of Database Server accesses that result in cache hits, measured as a percentage of hits versus total attempts. A high cache hit rate indicates efficiency.

Workload Optimization Manager collects cache hit rate data from Database Servers discovered by Databases, APM, and Cloud targets. When you set the scope to one or several Database Servers, the data that Workload Optimization Manager collected displays in the DB Cache Hit Rate chart.



The chart shows average and peak/low values over time. Use the selector on the chart to change the time frame.

NOTE:
An empty chart could be the result of delayed discovery, target validation failure, unavailable data for the given time frame, or an unsupported Database Server (see the next section for a list of supported Database Servers).

Supported database servers

Data is available in the DB Cache Hit Rate chart for Database Servers discovered via the following targets:

Target	Supported database servers
AWS	RDS
Azure	SQL Managed Instance
AppDynamics	SQL, Oracle
Dynatrace	SQL
Instana	MySQL, SQL, Oracle

Target	Supported database servers
MySQL	MySQL
New Relic	SQL, MySQL
Oracle	Oracle
SQL	SQL

Effect on memory resize actions

Actions to resize database memory are driven by data on the Database Server, which is more accurate than data on the hosting VM. Workload Optimization Manager uses database memory and cache hit rate data to decide whether resize actions are necessary.

A high cache hit rate value indicates efficiency. The optimal value is 100% for on-prem (self-hosted) Database Servers, and 90% for cloud Database Servers. When the cache hit rate reaches the optimal value, no action generates even if database memory utilization is high. If utilization is low, a resize down action generates.

When the cache hit rate is less than the optimal value but database memory utilization remains low, no action generates. If utilization is high, a resize up action generates.

DB memory chart

Database memory (or DBMem) is the measurement of memory that is utilized by a Database Server.

Workload Optimization Manager collects memory data from Database Servers discovered by Databases and APM targets. When you set the scope to one or several Database Servers, the data that Workload Optimization Manager collected displays in the DB Memory chart.



The chart shows average and peak/low values over time. Use the selector on the chart to change the time frame.

NOTE:
An empty chart could be the result of delayed discovery, target validation failure, unavailable data for the given time frame, or an unsupported Database Server (see the next section for a list of supported Database Servers).

Supported database servers

Data is available in the DB Memory chart for Database Servers discovered via the following targets:

Target	Supported database servers
AppDynamics	Oracle, MongoDB

Target	Supported database servers
Dynatrace	SQL, MySQL
Instana	MySQL, SQL
MySQL	MySQL
Oracle	Oracle
SQL	SQL

Memory resize actions

Actions to resize database memory are driven by data on the Database Server, which is more accurate than data on the hosting VM. Workload Optimization Manager uses database memory and cache hit rate data to decide whether resize actions are necessary.

A high cache hit rate value indicates efficiency. The optimal value is 100% for on-prem (self-hosted) Database Servers, and 90% for cloud Database Servers. When the cache hit rate reaches the optimal value, no action generates even if database memory utilization is high. If utilization is low, a resize down action generates.

When the cache hit rate is less than the optimal value but database memory utilization remains low, no action generates. If utilization is high, a resize up action generates.

Energy chart

Energy is the measurement of electricity used by a given entity over a period of time, expressed in watt-hours (Wh).

> **NOTE:**
> An empty chart could be the result of delayed discovery, target validation failure, unavailable data for the given time frame, or an unsupported entity (see the next section for a list of supported entities).

Supported entities

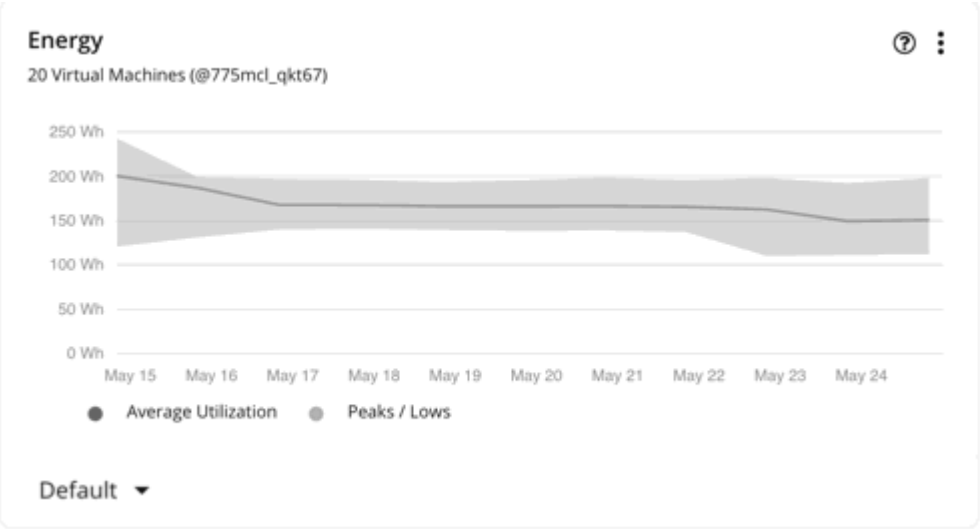
Data is available in the Energy chart for on-prem hosts and VMs and cloud VMs

For cloud VMs, Workload Optimization Manager supports the following providers and services:

- Amazon Web Services (AWS)
  - Amazon Elastic Compute Cloud (EC2)
- Google Cloud Platform (GCP)
  - Google Compute
- Microsoft Azure
  - Azure Compute

Calculation of energy consumed by on-prem VMs

Workload Optimization Manager collects energy-related data from hosts and VMs at 10-minute intervals. For VMs, the data collected from hosts is used to calculate VM energy consumption. When you set the scope to one or several hosts or VM, the data that Workload Optimization Manager collected or calculated displays in the Energy chart.



The chart shows average and peak/low values over time. Use the selector on the chart to change the time frame.

Workload Optimization Manager also relies on host power data for the supported targets. This data is mandatory for Energy and Carbon Footprint calculations.

Workload Optimization Manager calculates energy usage for each VM on a host, based on two types of host energy:

- **Host overhead energy**: Energy that is consumed by non-VM resources, such as hypervisors. This value includes host idle energy, the energy consumed by a host while it is in an idle state, when no processes are actively executing.

  Host overhead energy is attributed to all member VMs based on each VM's allocated capacity.
- **Host active energy**: The energy consumed by the VMs' operating systems and active processes.

  Host active energy is attributed to all member VMs based on each VM's usage of host CPU.

The calculation can be expressed as follows:

```
(VM Size / Total Size for all VMs) * Host Overhead Energy for VMs +
(VM CPU Utilization / Total CPU Utilization for all VMs) * Host Active Energy for VMs
= VM Energy usage (in Wh)
```

> **NOTE:**
> Workload Optimization Manager rounds the calculated values.

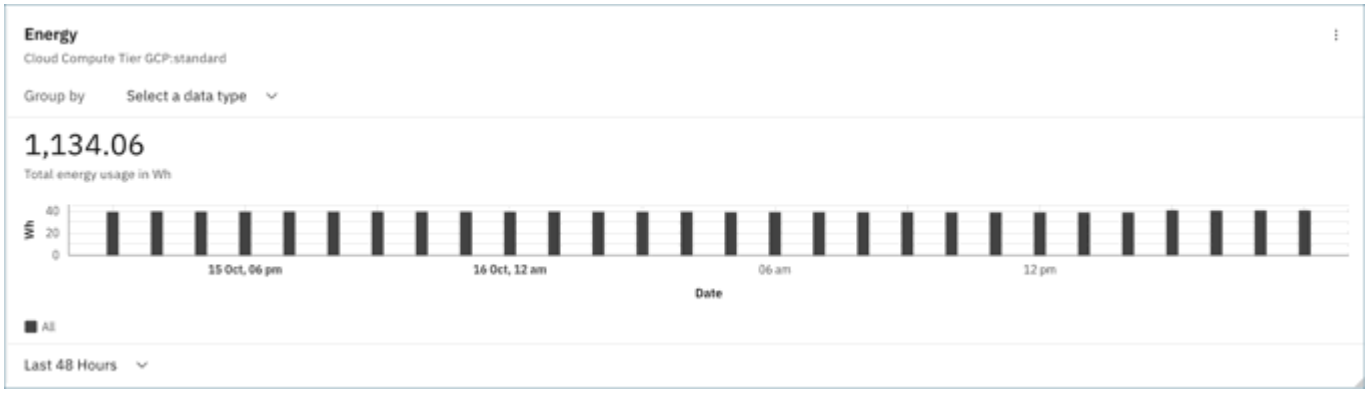
For example, consider a host with three VMs. Currently, host *idle* energy for VMs is 50 Wh, while host *active* energy for VMs is 20 Wh.

Workload Optimization Manager calculates VM energy usage as follows:

VM	Size (cores)	CPU utilization (%)	VM energy usage (Wh)
VM_01	2	50 (x 2 cores) = 100	(2 / 4) * 50 + (100 / 180) * 20 = **36.1**
VM_02	1	80 (x 1 core) = 80	(1 / 4) * 50 + (80 / 180) * 20 = **21.4**
VM_03	1	0 (idle)	(1 / 4) * 50 + (0) * 20 = **12.5**
Total	**4**	**180**	**70**

Calculation of energy consumed by public cloud VMs

Workload Optimization Manager provides the capability to measure and visualize the current and historical energy usage of cloud VMs or groups of VMs (account, region, zone, and more) at an hourly granularity.



The chart shows a series of interval values over time. Use the selector on the chart to change the time frame.

The methodology to calculate energy usage by cloud VMs is based on Cloud Carbon Footprint, with some precision enhancements, such as using the uptime of a VM and the actual CPU utilization that Workload Optimization Manager collects.

The calculation can be expressed as follows:

```
Energy in watt-hours = vCPU hours x average watts x PUE
```

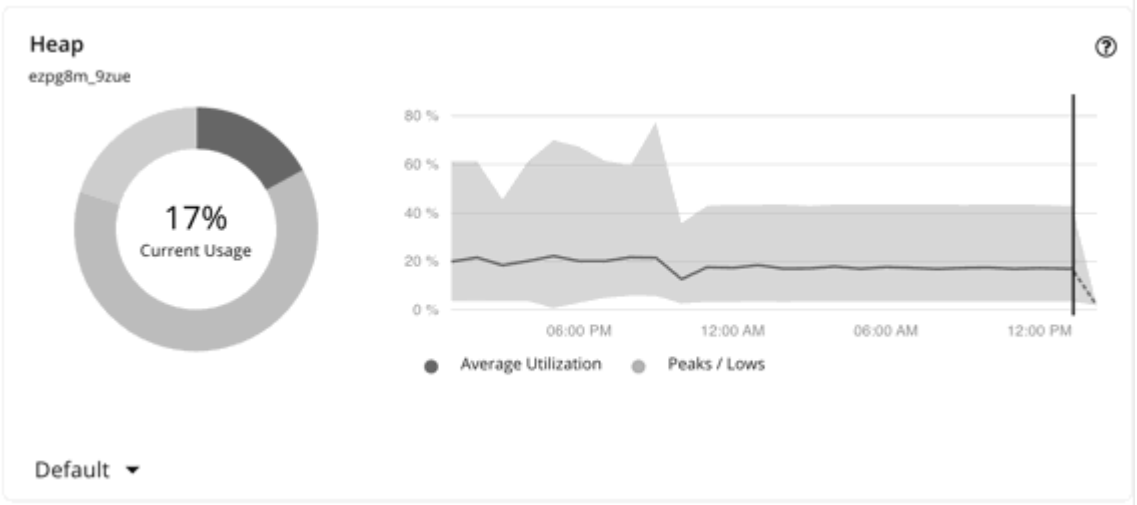
- **vCPU hours:** The total hours that CPUs are in use
  - Calculation: `vCPU hours = number of vCPUs x usage time in hours`
- **Average watts:** The average compute energy at a moment in time, based on processor type & CPU Utilization
  - Calculation: `Average watts = minimum watts + real-time CPU utilization x (maximum watts - minimum watts)`
    - **Minimum watts:** The minimum power usage by a processor in an idle state
    - **Maximum watts:** The maximum power usage by a processor at 100% CPU utilization
    - **Real-time CPU utilization:** The actual CPU utilization retrieved during the time window that energy is computed
- **PUE:** Power usage effectiveness
  Power Usage Effectiveness (PUE) is a ratio that describes how efficiently a computer data center uses energy; specifically, how much energy is used by the computing equipment. PUE is the ratio of the total amount of energy used by a computer data center facility to the energy delivered to computing equipment. The closer PUE is to 1, the more efficient the computer data center.

  For more information, see Power usage effectiveness .

Heap chart

Heap is the portion of a VM or container's memory allocated to individual applications.

Workload Optimization Manager collects heap data from Application Components discovered by Applications and APM targets. When you set the scope to one or several Application Components, the data that Workload Optimization Manager collected displays in the Heap chart.



The chart shows average and peak/low values over time. Use the selector on the chart to change the time frame.

NOTE:
An empty chart could be the result of delayed discovery, target validation failure, unavailable data for the given time frame, or an unsupported Application Component (see the next section for a list of supported Application Components).

Supported application components

Data is available in the Heap chart for Application Components discovered via the following targets:

Target	Supported application components
AppDynamics	Java applications, .NET, Node.js
Dynatrace	Java applications
Instana	Java applications
JBoss	Java applications
JVM	Java applications
New Relic	Java applications, Node.js
Tomcat	Java applications
WebLogic	Java applications
WebSphere	Java applications

Heap resize actions

Workload Optimization Manager generates Heap resize actions if an Application Component provides Heap and Remaining GC Capacity, and the underlying VM or container provides VMem. These actions are recommend-only and can only be executed outside Workload Optimization Manager.

NOTE:
Remaining GC capacity is the measurement of Application Component uptime that is *not* spent on garbage collection (GC).

Number of replicas chart

This chart shows the replicas of Application Components running over a given time period.

Use this chart if:

- SLO-driven scaling is enabled for a Service, and *provision* or *suspend* actions are executed by Workload Optimization Manager. These actions adjust the number of replicas to help you meet your SLO goals.

  Or

- Horizontal Pod Autoscaler (HPA) is enabled for a *Deployment*, *ReplicaSet*, or *StatefulSet* that is exposed as a Service. Workload Optimization Manager discovers adjustments to the number of replicas made by HPA.

The chart shows following information:

- **Capacity** values

  The number of desired pod replicas configured in the workload controller that backs the Service. This can be configured in *Deployment*, *ReplicaSet*, *StatefulSet*, *ReplicationController* or *DeploymentConfig*.

  The chart plots the *maximum* observed capacity within the given time period.

- **Used** values

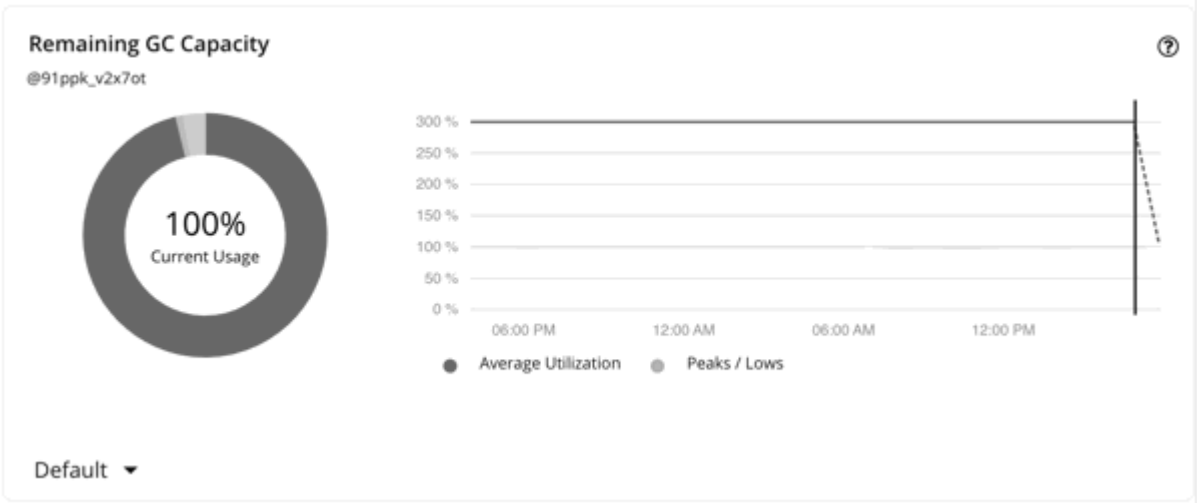
  The number of *ready* pods owned by the workload controller. Pods in other states (for example, pending pods) are not counted.

  The chart plots the *average* used values within the given time period. Hover over the chart to see the minimum and maximum used values.

Remaining GC capacity chart

Remaining GC capacity is the measurement of Application Component uptime that is *not* spent on garbage collection (GC).

Workload Optimization Manager collects GC data from Application Components discovered by Applications and APM targets, and then uses that data to calculate remaining GC capacity. When you set the scope to one or several Application Components, the capacity that Workload Optimization Manager calculated displays in the Remaining GC Capacity chart.



The chart shows average and peak/low values over time. Use the selector on the chart to change the time frame.

> **NOTE:**
> An empty chart could be the result of delayed discovery, target validation failure, unavailable data for the given time frame, or an unsupported Application Component (see the next section for a list of supported Application Components).

Supported application components

Data is available in the Remaining GC Capacity chart for Application Components discovered via the following targets:

Target	Supported application components
AppDynamics	Java applications, .NET
Dynatrace	Java applications
Instana	Java applications
JBoss	Java applications
JVM	Java applications
New Relic	Java applications, Node.js
Tomcat	Java applications
WebLogic	Java applications
WebSphere	Java applications

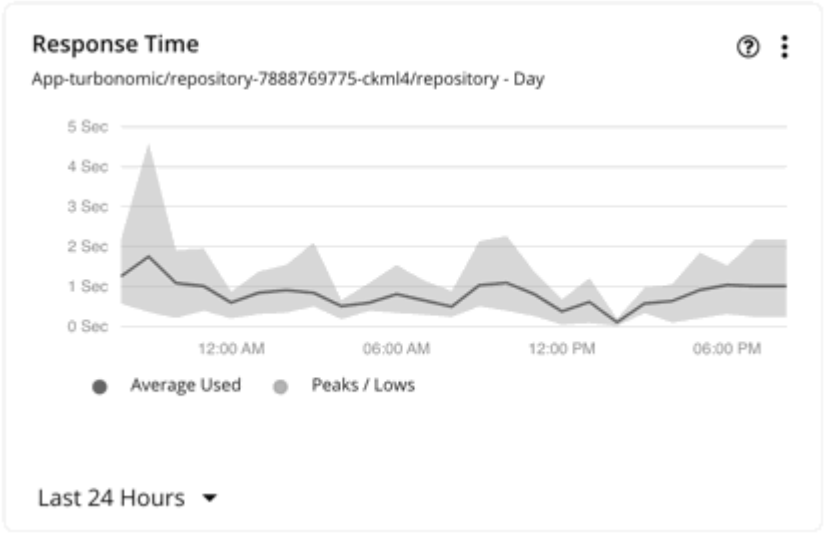
Effect on heap resize actions

Workload Optimization Manager generates Heap resize actions if an Application Component provides Heap and Remaining GC Capacity, and the underlying VM or container provides VMem. These actions are recommend-only and can only be executed outside Workload Optimization Manager.

Response time chart

Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

Workload Optimization Manager collects response time data from entities discovered by Applications, Databases, and APM targets. Entities include Business Applications, Business Transactions, Services, Application Components, and self-hosted Database Servers. When you set the scope to any of these entities, the data that Workload Optimization Manager collected displays in the Response Time chart.



The chart shows average and peak/low values over time. Use the selector on the chart to change the time frame.

NOTE:
An empty chart could be the result of delayed discovery, target validation failure, unavailable data for the given time frame, or an unsupported entity (see the next section for a list of supported entities).

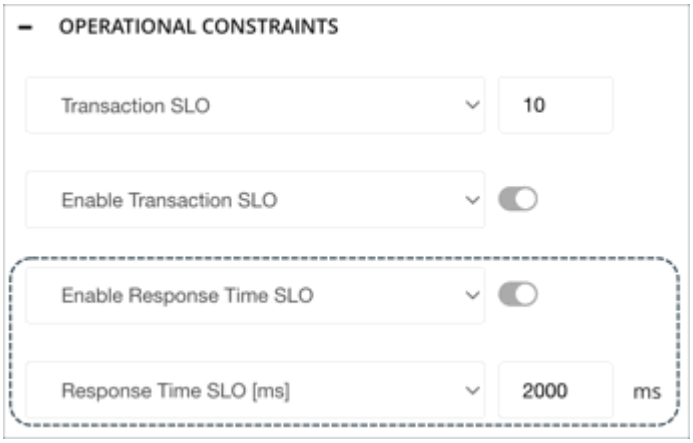
Supported entities

Data is available in the Response Time chart for the following entities:

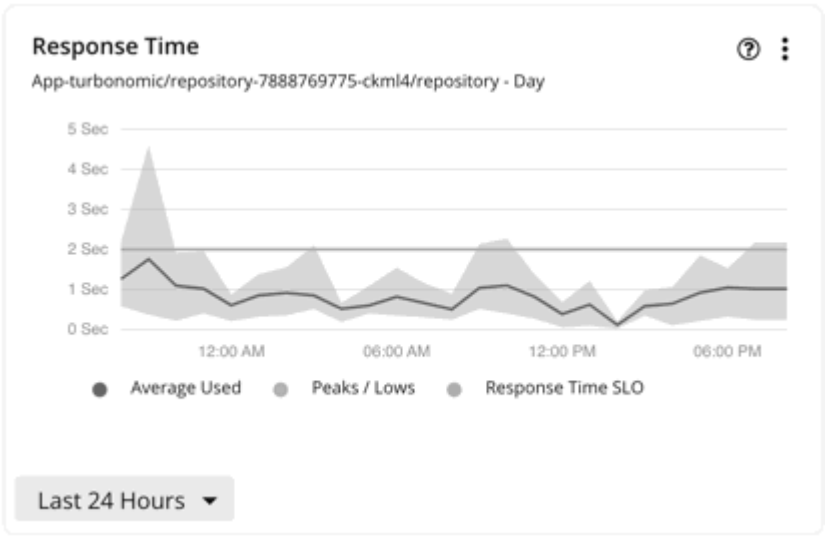
Target	Supported entities
AppDynamics	Business Application, Business Transaction, Service, Application Component
Datadog	Business Application, Business Transaction, Service, Application Component
Dynatrace	Business Application, Service, Application Component
Instana	Business Application, Business Transaction, Service, Application Component, Database Server
JBoss	Application Component
JVM	Application Component
MySQL	Database Server
New Relic	Business Transaction, Service, Application Component, Database Server
Oracle	Database Server
SQL	Database Server
Tomcat	Application Component
WebLogic	Application Component
WebSphere	Application Component

Response time SLOs

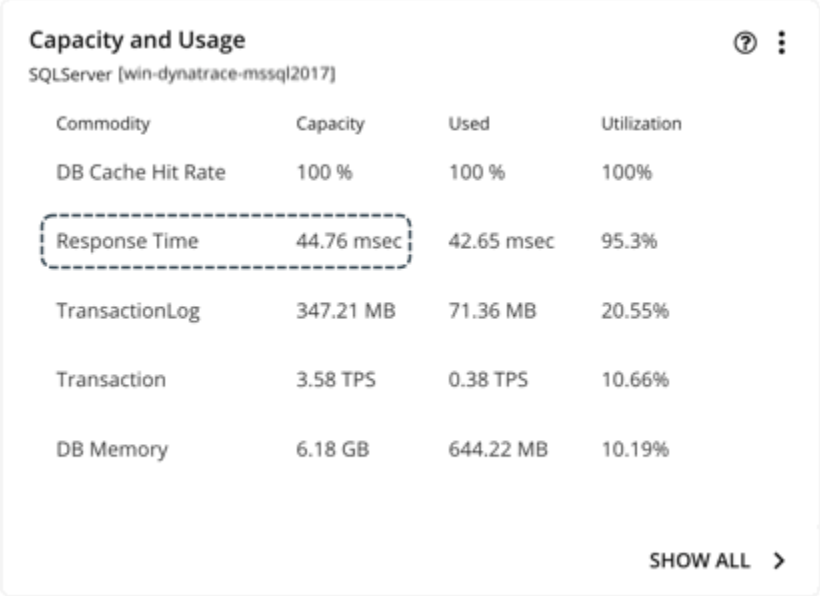
To evaluate the performance of your applications and Database Servers, set Response Time SLOs (Service Level Objectives) as an operational constraint in policies. For applications, you can set the SLO at the Business Application, Business Transaction, Service, or Application Component level.



After you create a policy, the SLO value appears as a solid straight line in the Response Time chart. You can then gauge performance against the given SLO.



If you do not set an SLO, Workload Optimization Manager estimates SLO based on historical Response Time data collected from the target, and then displays the estimated value in the Capacity and Usage chart, as Response Time capacity. This estimated value is *not* reflected in the Response Time chart.

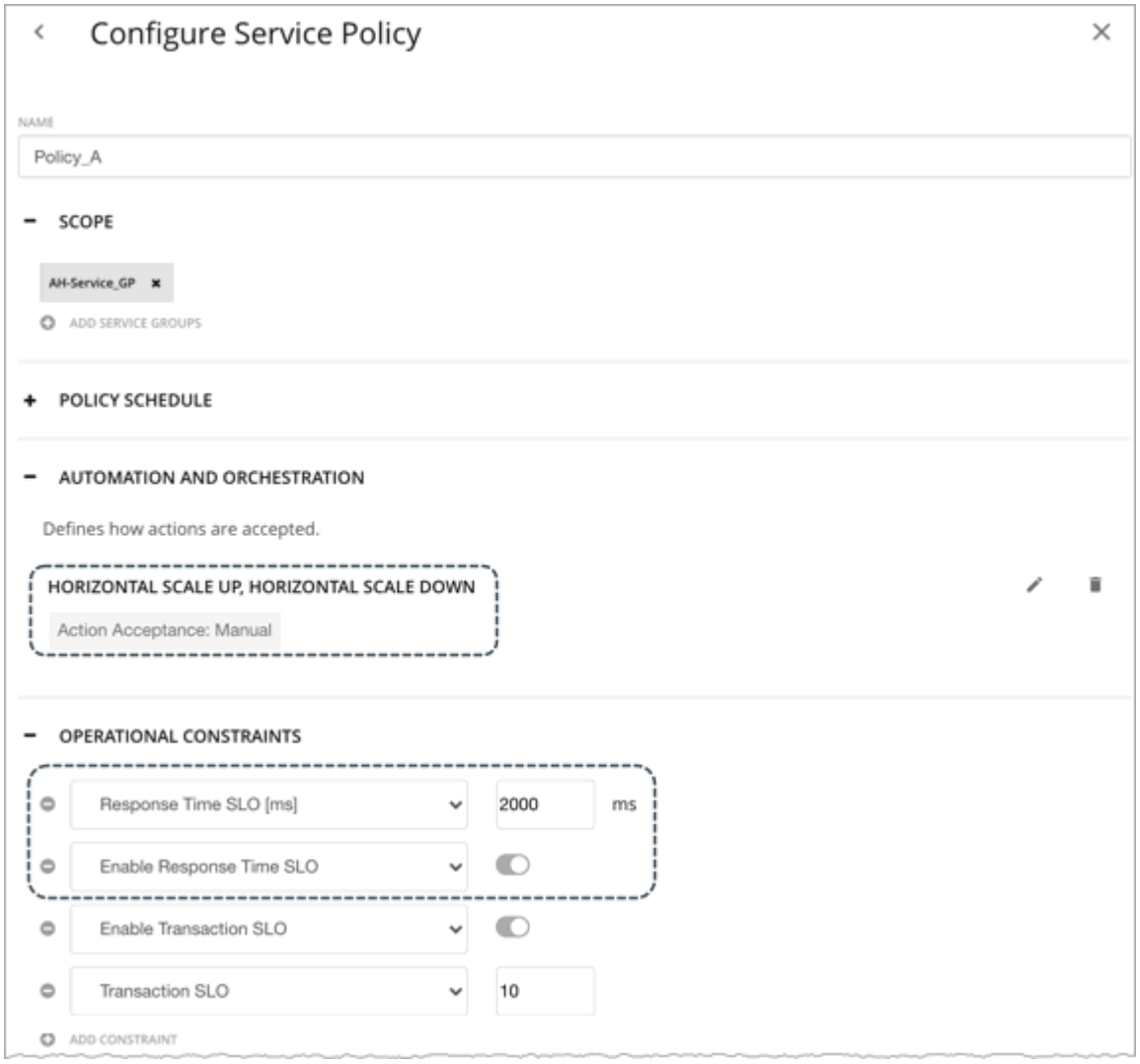


NOTE:
When you set an SLO value, Response Time capacity in the Capacity and Usage chart shows as N/A.

Response time SLOs for container platform services

When you add a container platform target, Workload Optimization Manager discovers services monitored by AppDynamics, Instana, Dynatrace, and New Relic.

To generate actions that adjust pod replicas, services must be discovered by the Kubeturbo agent that you deployed to your environment, as well as collect performance metrics through Instana or DIF (Data Ingestion Framework). In addition, Workload Optimization Manager requires that you turn on horizontal scaling and specify Response Time SLOs in policies for the affected services.



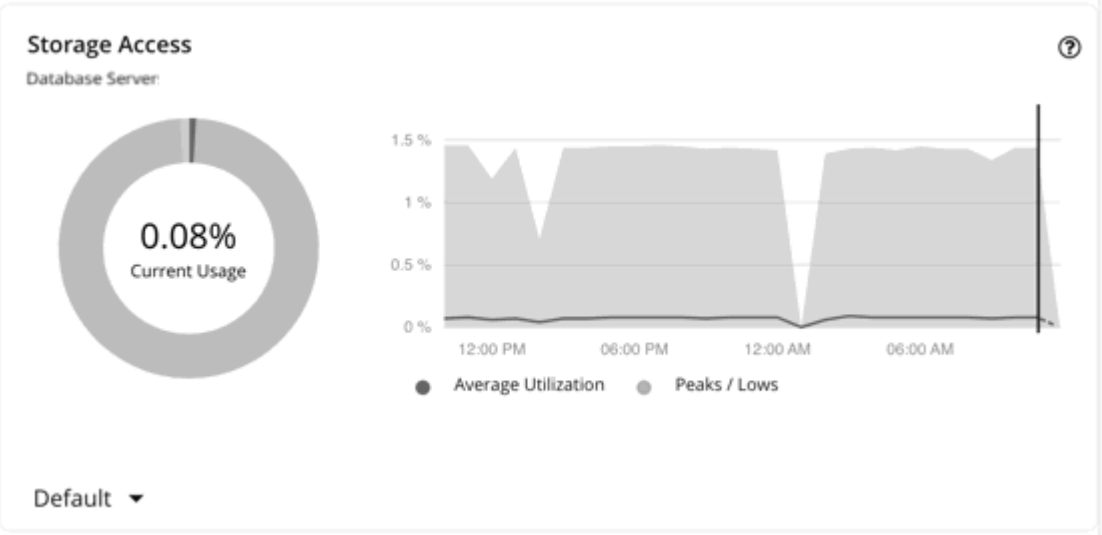
Response time SLO is the desired *weighted average* response time (in milliseconds) of all Application Component replicas associated with a Service.

> **NOTE:**
> If you specified SLOs but turned off horizontal scaling in policies, no actions generate but SLO values will continue to display in the Response Time chart for Services, for your reference. This allows you to gauge performance against those SLOs.

Storage access chart

Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

Workload Optimization Manager collects storage access data from VMs, Database Servers, and storage entities discovered by cloud, on-prem, and storage targets. When you set the scope to these entities, the data that Workload Optimization Manager collected displays in the Storage Access chart.



The chart shows average and peak/low values over time. Use the selector on the chart to change the time frame.

NOTE:
An empty chart could be the result of delayed discovery, target validation failure, unavailable data for the given time frame, or an unsupported entity (see the next section for a list of supported entities).

Supported entities

Data is available in the Storage Access chart for entities discovered via the following targets:

Target type	Target	Supported entities
Cloud	AWS	Virtual Machine, Database Server, Volume
	Azure	Virtual Machine, Database Server, Volume
	Google Cloud	Virtual Machine
Fabric	HPE OneView	Virtual Machine, Storage
Hyperconverged	HyperFlex	Storage, Disk Array
	Nutanix	Storage, Disk Array
Hypervisor	Hyper-V	Virtual Machine, Storage
	vCenter	Virtual Machine, Storage
Storage	EMC ScaleIO	Storage, Disk Array
	EMC VMAX	Storage, Disk Array, Logical Pool
	EMC XtremIO	Storage, Disk Array
	HPE 3PAR	Storage, Disk Array, Logical Pool
	IBM FlashSystem	Storage, Disk Array, Logical Pool
	NetApp	Storage, Disk Array
	Pure Storage	Storage, Disk Array

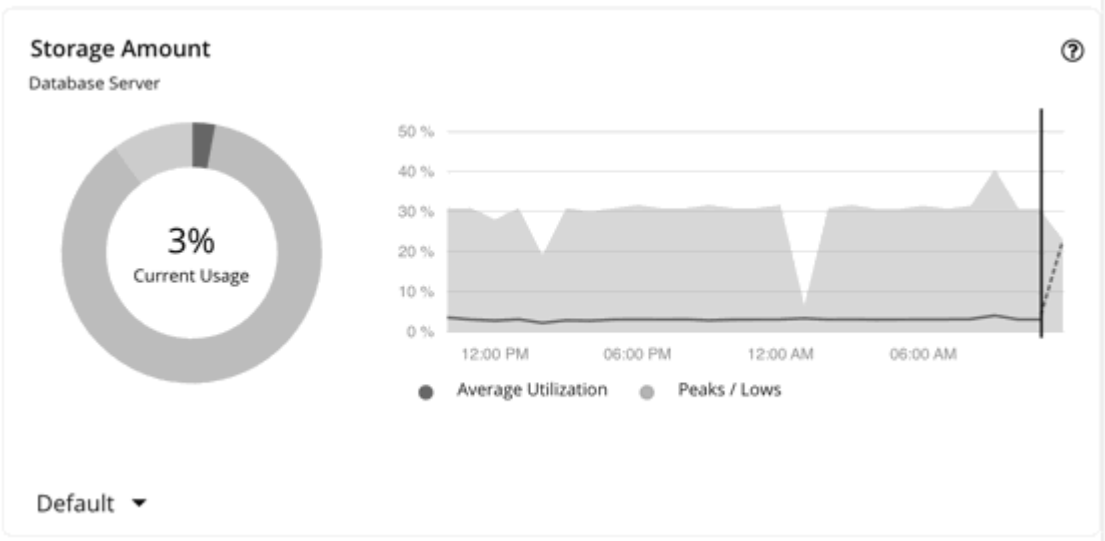
Scale actions

Workload Optimization Manager considers storage access (IOPS) utilization when recommending scaling actions for cloud VMs *(on page 637)*, cloud Database Servers *(on page 685)*, and volumes *(on page 728)*.

Storage amount chart

Storage amount is the measurement of storage capacity that is in use.

Workload Optimization Manager collects storage amount data from VMs, Database Servers, and storage entities discovered by cloud, on-prem, and storage targets. When you set the scope to these entities, the data that Workload Optimization Manager collected displays in the Storage Amount chart.



The chart shows average and peak/low values over time. Use the selector on the chart to change the time frame.

NOTE:
An empty chart could be the result of delayed discovery, target validation failure, unavailable data for the given time frame, or an unsupported entity (see the next section for a list of supported entities).

Supported entities

Data is available in the Storage Amount chart for entities discovered via the following targets:

Target type	Target	Supported entities
Cloud	AWS	Virtual Machine, Database Server (RDS)
	Azure	Virtual Machine, Database (SQL), Database Server (SQL Managed Instance)
	Google Cloud	Virtual Machine
Fabric	HPE OneView	Storage
Hyperconverged	HyperFlex	Storage, Disk Array
	Nutanix	Storage, Disk Array, Storage Controller
Hypervisor	Hyper-V	Storage
	vCenter	Storage

Target type	Target	Supported entities
Storage	EMC ScaleIO	Storage, Disk Array, Storage Controller
	EMC VMAX	Storage, Disk Array, Logical Pool, Storage Controller
	EMC XtremIO	Storage, Disk Array, Storage Controller
	HPE 3PAR	Storage, Disk Array, Logical Pool, Storage Controller
	IBM FlashSystem	Storage, Disk Array, Logical Pool, Storage Controller
	NetApp	Storage, Disk Array, Storage Controller
	Pure Storage	Storage, Disk Array, Storage Controller

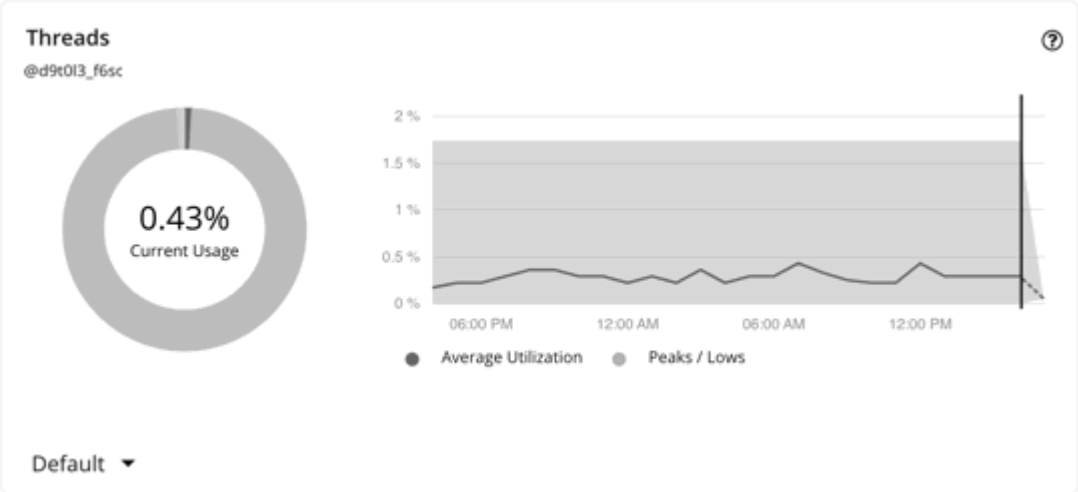
Scale actions

Workload Optimization Manager can recommend scaling cloud VMs *(on page 637)* and cloud Database Servers *(on page 685)* to instance types that can adequately meet storage demand at the lowest possible cost. For cloud Database Servers, Workload Optimization Manager can also recommend scaling up storage amount within the same instance type. Note that scaling up storage amount is non-disruptive but irreversible.

Threads chart

Threads is the measurement of thread capacity utilized by applications.

Workload Optimization Manager collects thread data from Application Components discovered by Applications and APM targets. When you set the scope to one or several Application Components, the data that Workload Optimization Manager collected displays in the Threads chart.



The chart shows average and peak/low values over time. Use the selector on the chart to change the time frame.

NOTE:
An empty chart could be the result of delayed discovery, target validation failure, unavailable data for the given time frame, or an unsupported Application Component (see the next section for a list of supported Application Components).

Supported application components

Data is available in the Threads chart for Application Components discovered via the following targets:

Target	Supported application components
AppDynamics	Java applications, .NET
JBoss	Java applications
JVM	Java applications
New Relic	Java applications
Tomcat	Java applications
WebLogic	Java applications
WebSphere	Java applications

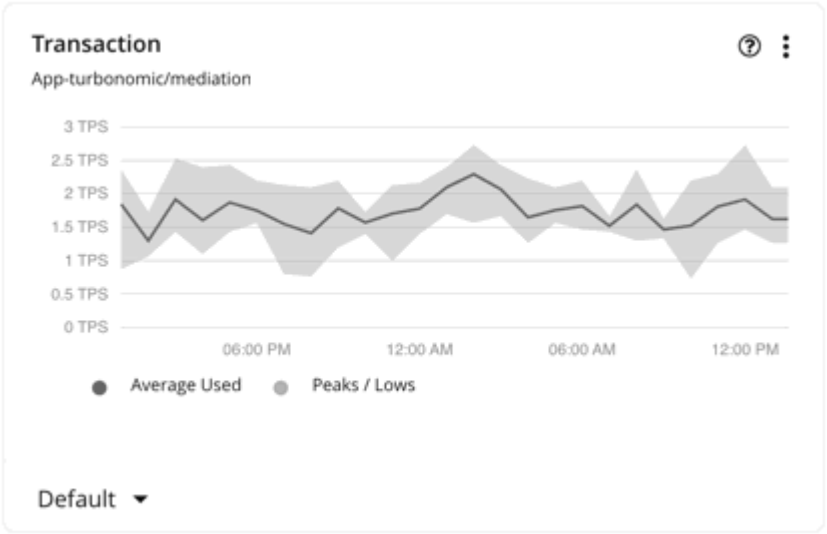
Thread pool resize actions

Workload Optimization Manager generates thread pool resize actions. These actions are recommend-only and can only be executed outside Workload Optimization Manager.

Transaction chart

Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

Workload Optimization Manager collects transaction data from entities discovered by Applications, Databases, and APM targets. Entities include Business Applications, Business Transactions, Services, Application Components, and self-hosted Database Servers. When you set the scope to any of these entities, the data that Workload Optimization Manager collected displays in the Transaction chart.



The chart shows average and peak/low values over time. Use the selector on the chart to change the time frame.

> **NOTE:**
> An empty chart could be the result of delayed discovery, target validation failure, unavailable data for the given time frame, or an unsupported entity (see the next section for a list of supported entities).

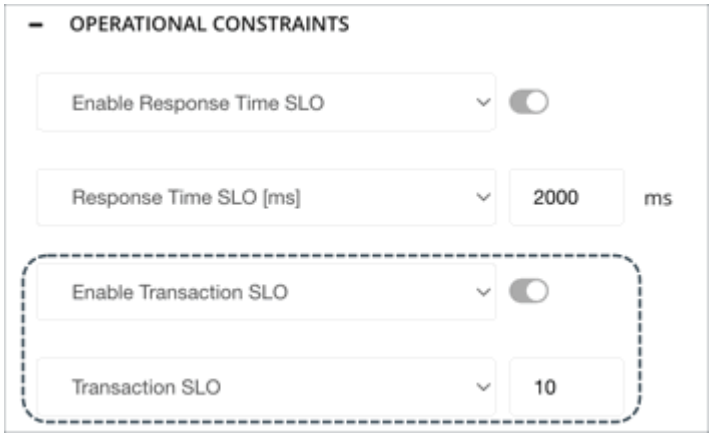
Supported entities

Data is available in the Transaction chart for the following entities:

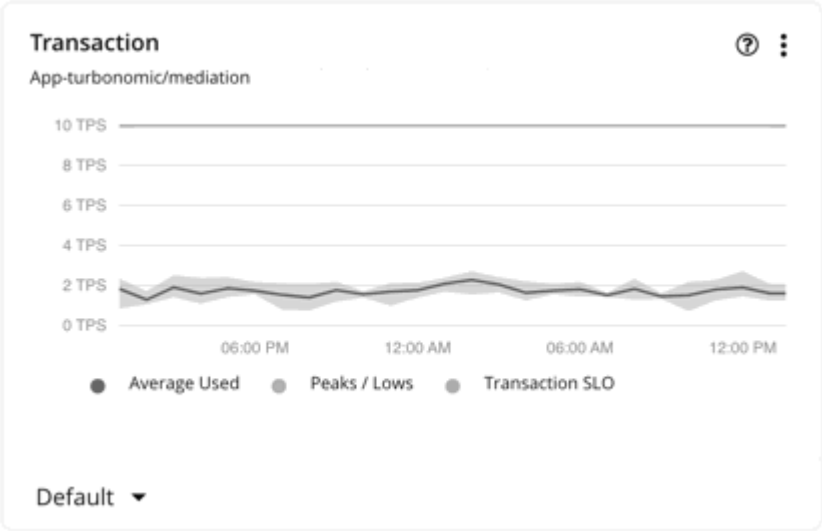
Target	Supported entities
AppDynamics	Business Application, Business Transaction, Service, Application Component, Database Server
Dynatrace	Business Application, Service, Database Server, Application Component
Instana	Business Application, Business Transaction, Service, Database Server
JBoss	Application Component
MySQL	Database Server
New Relic	Business Transaction, Service, Application Component, Database Server
Oracle	Database Server
SQL	Database Server
Tomcat	Application Component
WebLogic	Application Component
WebSphere	Application Component

Transaction SLOs

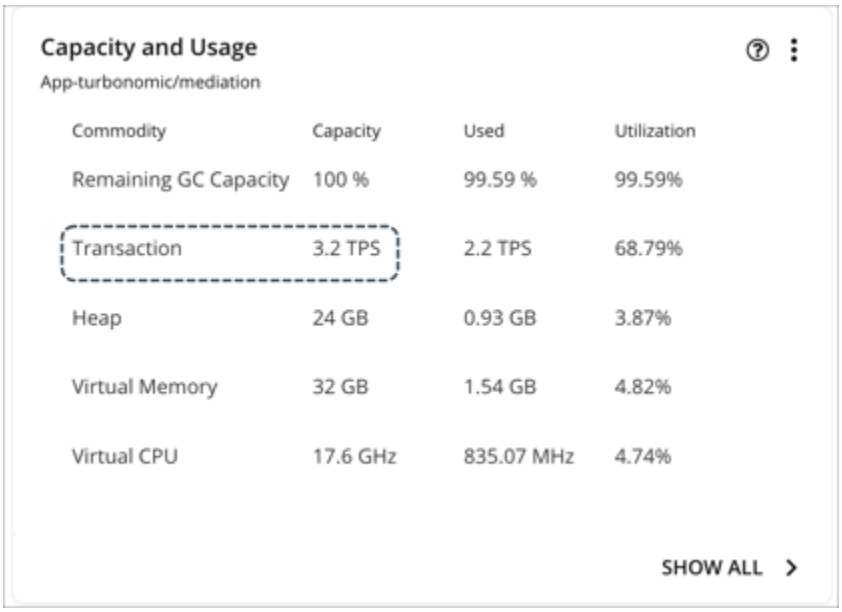
To evaluate the performance of your applications and Database Servers, set Transaction SLOs as an operational constraint in policies. For applications, you can set the SLO at the Business Application, Business Transaction, Service, or Application Component level.



After you create a policy, the SLO value appears as a solid straight line in the Transaction chart. You can then gauge performance against the given SLO.



If you do not set an SLO, Workload Optimization Manager estimates SLO based on historical Transaction data collected from the target, and then displays the estimated value in the Capacity and Usage chart, as Transaction capacity. This estimated value is *not* reflected in the Transaction chart.

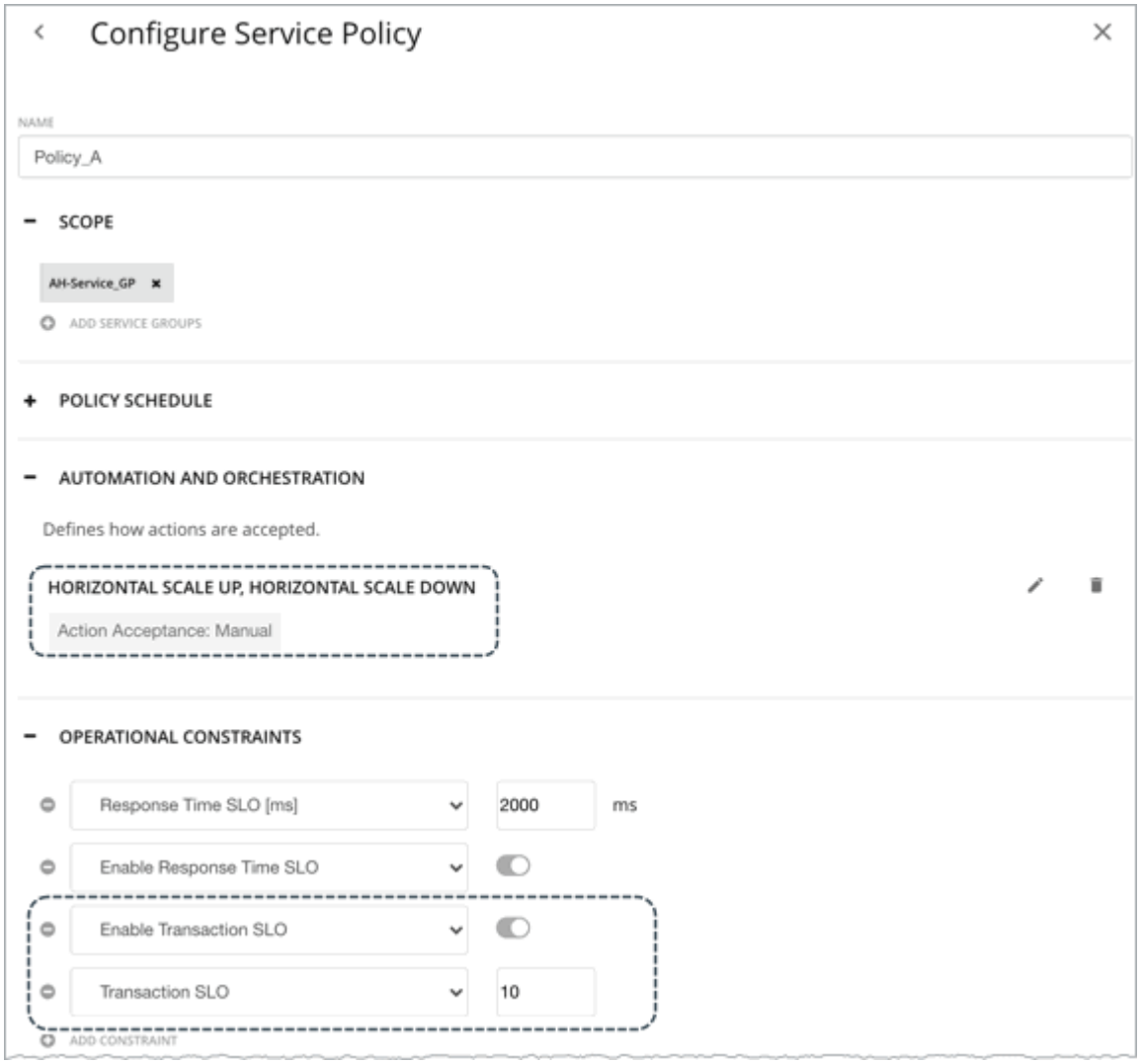


NOTE:
When you set an SLO value, Transaction capacity in the Capacity and Usage chart shows as N/A.

Transaction SLOs for container platform services

When you add a container platform target, Workload Optimization Manager discovers container platform services managed by AppDynamics, Instana, Dynatrace, and New Relic.

To generate actions that adjust pod replicas, container platform services must be discovered by the Kubeturbo pod that you deployed to your environment, as well as collect performance metrics through Instana or DIF (Data Ingestion Framework). In addition, Workload Optimization Manager requires that you turn on horizontal scaling and specify Transaction SLOs in policies for the affected services.



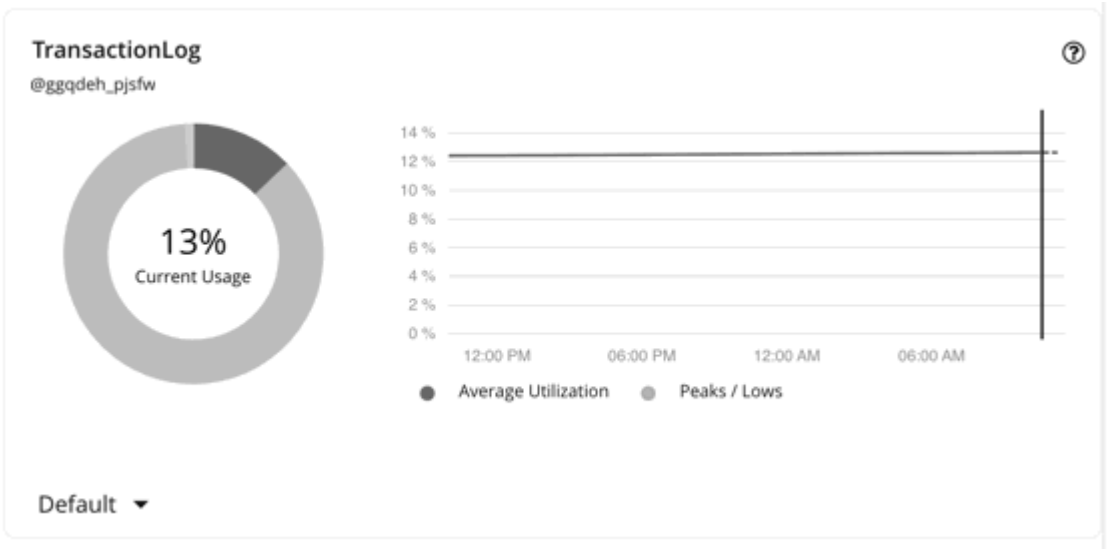
Transaction SLO is the maximum number of transactions per second that each Application Component replica can handle.

> **NOTE:**
> If you specified SLOs but turned off horizontal scaling in policies, no actions generate but SLO values will continue to display in the Transaction chart for services, for your reference. This allows you to gauge performance against those SLOs.

Transaction log chart

Transaction log is the measurement of storage capacity utilized by a Database Server for transaction logging.

Workload Optimization Manager collects transaction log data from Database Servers discovered by Databases and APM targets. When you set the scope to one or several Database Servers, the data that Workload Optimization Manager collected displays in the Transaction Log chart.



The chart shows average and peak/low values over time. Use the selector on the chart to change the time frame.

NOTE:
An empty chart could be the result of delayed discovery, target validation failure, unavailable data for the given time frame, or an unsupported Database Server (see the next section for a list of supported Database Servers).

Supported database servers

Data is available in the Transaction Log chart for Database Servers discovered via the following targets:

Target	Supported database servers
AppDynamics	SQL
Oracle	Oracle
SQL	SQL

Transaction log resize actions

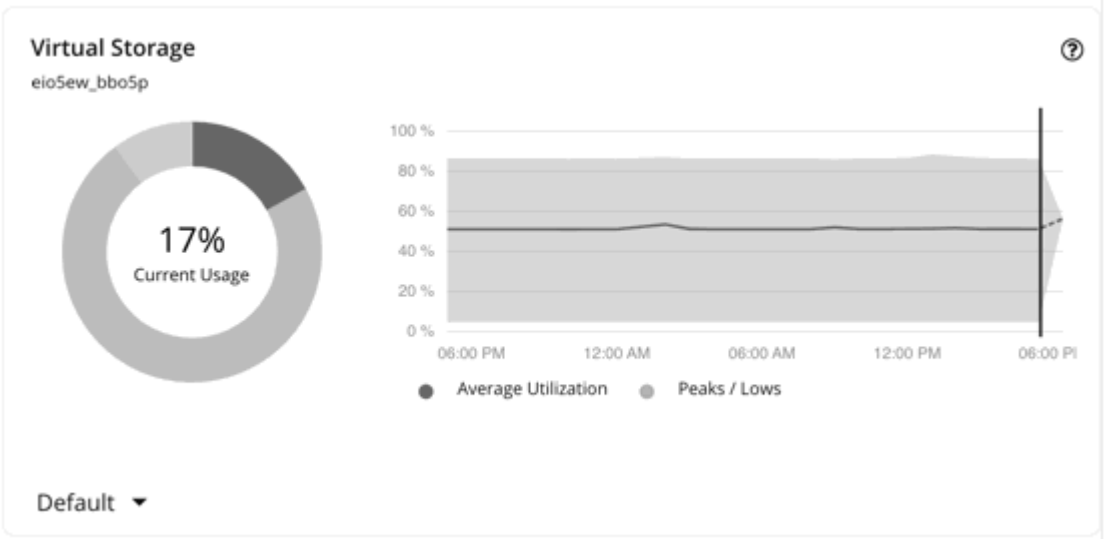
Resize actions based on the transaction log resource depend on support for virtual storage in the underlying hypervisor technology.

Currently, Workload Optimization Manager does not support resize actions for Oracle and Database Servers on the Hyper-V platform (due to the lack of API support for virtual storage).

Virtual storage chart

Virtual ssorage is the measurement of virtual storage capacity that is in use.

Workload Optimization Manager collects virtual storage data from VMs discovered by on-prem and APM targets. When you set the scope to one or several VMs, the data that Workload Optimization Manager collected displays in the Virtual Storage chart.



The chart shows average and peak/low values over time. Use the selector on the chart to change the time frame.

NOTE:
An empty chart could be the result of delayed discovery, target validation failure, unavailable data for the given time frame, or an unsupported entity (see the next section for a list of supported entities).

Supported entities

Data is available in the Virtual Storage chart for entities discovered via the following targets:

Target type	Target	Supported entities
Database	MySQL	Virtual machine
	Oracle	Virtual machine
	SQL	Virtual machine
Fabric	HPE OneView	Virtual machine
Guest OS Process	SNMP	Virtual machine
	WMI	Virtual machine
Hyperconverged	HyperFlex	Virtual machine
	Nutanix	Virtual machine
Hypervisor	Hyper-V	Virtual machine
	vCenter	Virtual machine

Virtual storage actions

Workload Optimization Manager can recommend the following virtual storage actions:

- Move a VM's volume (virtual storage) due to excess utilization of the current datastore, or for more efficient utilization of datastores in the environment.
- Reconfigure a VM's volume (virtual storage) to comply with placement policies.

Top utilized charts

Top Utilized charts show the entities or groups with the most utilization.

Entity Type

Entity types you can choose include:

- Accounts *(on page 1006)* (public cloud)
- Application Components
- Business Applications
- Business Transactions
- Business Users
- Chassis
- Host Clusters *(on page 1006)*
- Containers
- Container Pods
- Container specs
- Data Centers
- Database Servers
- Desktop Pools
- Disk Arrays
- IO Modules
- Internet
- Namespaces
- Networks
- Node Pools
- Hosts
- Resource Groups *(on page 1007)*
- Services
- Storage Devices
- Storage Controllers
- Switches
- View Pods
- Virtual Data Centers
- Virtual Machines
- Volumes
- Wasted Files
- Workload controllers
- Regions
- Zones

Data Type

Depending on the entity type (for example, Host Clusters), you can choose **Headroom** or **Utilization** information in the chart.

Commodity

Depending on the entity type, you can add one or more different resource commodities that you want to measure.

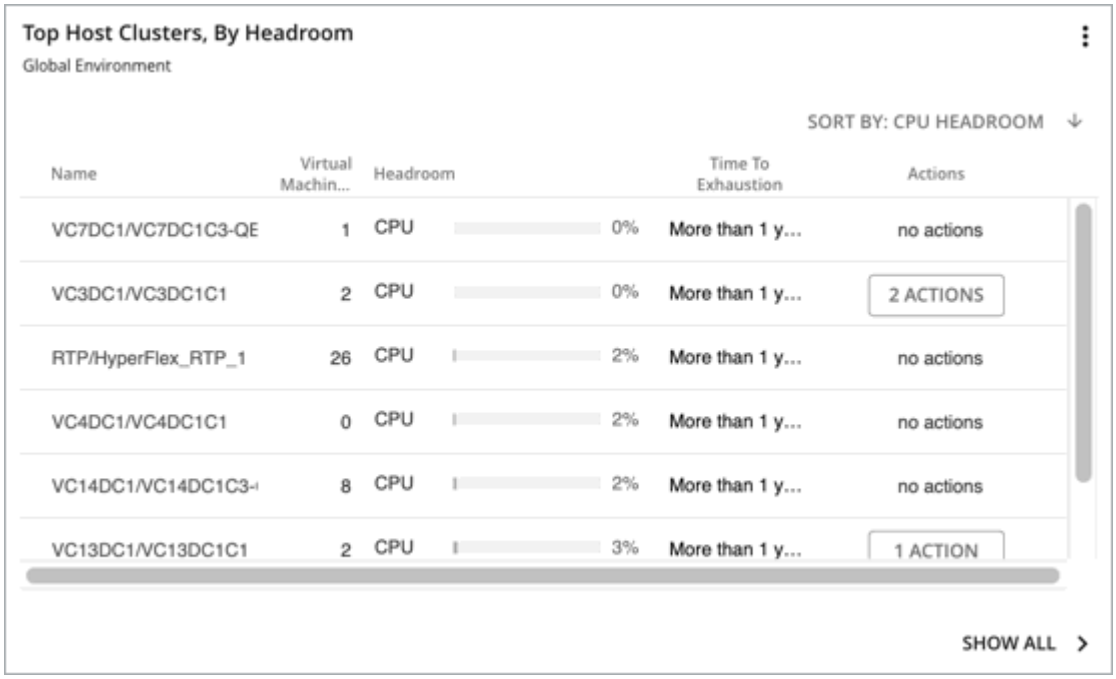
Display

The chart shows data from the last successful discovery and does not track historical data. The chart lists the top entities by utilization of the commodities that you selected or that the system set. Depending on the entity type and scope, you can sort the information. To view the utilization details, hover over the entity to display the tooltip.

To drill down to an entity, click the entity name in the chart. This sets the scope to that entity.

Click the **ACTIONS** button for an entity to examine the actions that are pending for it, and then decide which ones are safe to execute.

Example: A top host clusters chart which can be sorted by CPU headroom or CPU exhaustion.



Top host clusters chart

This chart shows the top host clusters in your on-prem environment by CPU, memory, and storage capacity or utilization. In the default view, the chart shows the top clusters by CPU headroom (available capacity). The chart also shows the time to exhaustion of cluster resources, which is useful for resource planning (for example, you might need to buy more hardware).

To calculate cluster capacity and headroom, Workload Optimization Manager runs nightly plans that take into account the conditions in your current environment. The plans identify the optimal workload distribution for your clusters. This can include moving your current VMs to other hosts within the given cluster, if such moves would result in a more desirable workload distribution. The result of the plan is a calculation of how many more VMs the cluster can support.

Click the value of the **Actions** column for a host cluster to see the actions that Workload Optimization Manager recommends for the resources in that cluster. Then, decide which actions are safe to run.

Click **Show All** to see the list of all host clusters. From this list, you can download capacity data as a CSV file. You can also click a cluster name to set the scope to that cluster and view more details about the current capacity and health of the cluster.

Top accounts chart

This chart lists the cloud accounts with the largest potential savings. These are the savings you would realize if you execute pending actions for your cloud workloads. The **Actions** column shows the number of pending actions for each account. Click a

number to examine these actions and decide which ones are safe to execute. The chart also shows billed costs in the last 30 days.

To set the scope to a particular account, click the account name.

Cloud provider accounts

The chart shows the following accounts discovered from your cloud provider.

Cloud provider	Account
AWS	The chart shows the AWS management and member accounts that you added as targets, including GovCloud *(on page 397)* accounts. For GovCloud accounts, the billed cost is shown as **N/A** since invoicing for these accounts is managed through an associated AWS standard account. Adding this standard account as a target is optional. When added, it appears in the chart and shows the total billed cost for the GovCloud accounts and the standard account itself. **NOTE:** Specific RIs can provide savings for multiple accounts. However, individual accounts show the full RI savings, which can result in exaggerated savings for that account.
Azure	The chart shows the subscriptions discovered through the Azure service principal accounts that you added as targets, including Azure Government *(on page 427)* subscriptions.
Google Cloud	The chart shows the projects discovered through the Google Cloud service accounts that you added as targets.

Show all

Click **Show all** to view additional information, including the number of actions that were executed for individual accounts or workloads, along with the resulting savings.

Use the Search bar to search for non-numerical data in the Name, ID, Cloud provider, and Account type columns. It is not possible to search data in the Workloads, Billed cost, and Actions columns because these columns only show numerical data.

You can download the accounts list as a CSV file.

Top resource groups chart

This chart highlights the estimated monthly cost for the top resource groups in your cloud environment and the savings you would realize if you execute the pending actions. Click the **ACTIONS** button to examine these actions and decide which ones are safe to execute. Click a resource group to set the scope to that group.

The chart also counts actions that have been executed for individual groups, and then shows the resulting savings.

Workload health charts

Workload Health charts show the health of workloads from the compliance, efficiency improvement, and performance assurance perspectives. These charts use current (real-time) data for the workloads chosen for the chart widget scope.

Chart Type

You can set the display to:

- Text
- Ring Chart
- Horizontal Bar

Breakdown

You can choose:

- Workload by Compliance

  A virtual environment can include policies that limit availability of resources. It's possible that the environment configuration violates these defined policies. In such cases, Workload Optimization Manager identifies the violation and recommends actions that bring the entity back into compliance.

- Workload by Efficiency Improvement

  Efficient utilization of resources is an important part of running in the desired state. Running efficiently maximizes your investment and reduces cost. When Workload Optimization Manager discovers underutilized workloads, it recommends actions to optimize operations and save money.

- Workload by Performance Assurance

  Ultimately, the reason to manage workloads in your environment is to assure performance and meet QoS goals. When Workload Optimization Manager detects conditions that directly put QoS at risk, it recommends associated actions to assure performance. You can consider these critical conditions, and you should execute the recommended actions as soon as possible.

Workload Health charts indicate actions that you should consider to improve the health of workloads. To see a list of actions, click **Show Actions** at the end of the chart.

Environment chart

The Environment chart shows the cloud and on-prem targets *(on page 173)* that you added to Workload Optimization Manager and the number of entities for those targets.

You can choose from the following environment types:

- Hybrid (both on-prem and cloud environments)
- Cloud
- On-prem

Workload improvements charts

Workload Improvements charts track the health of workloads in your environment over time, and map the health to the number of actions Workload Optimization Manager has executed in that time period.

In the chart, you can see the significance and value of executed actions:

- Workloads Overall

  This is the total number of workloads over time.

- Workloads with Performance Risks

  These are the workloads that are not performing well.

- Inefficient Workloads

  These are the workloads that are running on under-utilized hosts or are not being utilized.

- Workloads Out of Compliance

These are the workloads that are violating a placement policy. Workloads that are not in compliance might be running on a host or placed on storage, for example, that violate a placement policy.

- Executed actions

  Actions that Workload Optimization Manager executed successfully.

The vertical line shows when the last data point was polled in your environment.

Environment Type

You can choose one of the following views:

- Hybrid (both on-prem and cloud)
- Cloud
- On-Prem

Display

The chart shows the information as a Line chart.

Optimizations and risks charts

Use these charts to view a summary of actions that you can take to optimize resources, address performance risks, and enforce compliance to business rules. Click **Show all** to view relevant actions in Action Center *(on page 818)*.

Chart type

You can set the display to:

- Ring Chart
- Horizontal Bar
- Text

Breakdown by

- Efficiency

  Efficient utilization of resources is an important part of running in the desired state. Running efficiently maximizes your investment and reduces cost. When Workload Optimization Manager discovers underutilized resources, it recommends actions to consolidate your operations. For example, it can recommend that you move certain VMs onto a different host, and then suspend the original host.

- Performance

  Ultimately, the reason to manage workloads in your environment is to assure performance and meet productivity goals. Workload Optimization Manager can recommends actions when it detects conditions that directly put performance at risk. You can consider these critical conditions, and you should execute the recommended actions as soon as possible.

- Compliance

  A virtual environment can include policies that limit availability of resources. It's possible that the environment configuration violates these defined policies. In such cases, Workload Optimization Manager identifies the violation and recommends actions that bring the entity back into compliance.

Cloud chart types

These chart widgets provide information on the status of your cloud environment.

For many cloud chart widgets that display costs and savings, Workload Optimization Manager uses the billing reports from your cloud service providers to build a picture of your overall costs. The data includes all costs that the service provider includes in the billing report. Workload Optimization Manager parses these reports into the formats that it uses for the cloud chart widgets.

Cost breakdown charts

Workload Optimization Manager displays the unbilled billing information from your cloud provider. Use this chart to track *(on page 155)* your expenses and see historical trends.

Categories

The chart can break down costs by the following categories.

- Cost Breakdown by Service Provider

  See costs over time for each cloud service provider in your environment.

  You can open more than one account from a single service provider. If you are running workloads on different service providers, this chart shows the distribution of costs across these providers.

- Cost Breakdown by Account

  The chart shows the cloud accounts *(on page 1006)* with the largest costs. The chart's legend displays up to 20 actual accounts and, if needed, an additional item labeled 'Others' that represents all accounts that are not in the top 20. Hover on a data point to see costs for individual accounts.

- Cost Breakdown by Service

  The chart shows the services with the largest costs. The chart's legend displays up to 20 actual services and, if needed, an additional item labeled 'Others' that represents all services that are not in the top 20. Hover on a data point to see costs for individual services.

- Cost Breakdown By Region

  The chart shows the regions with the largest costs. The chart's legend displays up to 20 actual regions and, if needed, an additional item labeled 'Others' that represents all regions that are not in the top 20. Hover on a data point to see costs for individual regions.

- Cost Breakdown by Tag

  Workload Optimization Manager discovers the tags that you assigned to resources in your environment and then displays the tag keys when you click **Tag** on the chart. Select a tag key to see the tag values with the largest costs.

  The chart's legend displays up to 20 actual tag values and, if needed, an additional item labeled 'Others' that represents all tag values that are not in the top 20. Hover on a data point to see costs for individual tag values.

  > **NOTE:**
  > Workload Optimization Manager stores and displays tag keys in lowercase. If you use these tags in Workload Optimization Manager API requests, specify them in lowercase. Otherwise, the API returns incorrect results.

- Cost Breakdown by Workload

  This chart breaks down costs by the following categories:

  - On-Demand Compute
  - Spot Compute
  - On-Demand Compute License Bundle – only applies to Azure VMs and AWS VMs running Windows OS
  - On-demand License
  - Reserved License
  - Reserved Compute
  - Spot Compute License Bundle

  - On-Demand Storage
  - Other – includes workload costs that are not categorized by Workload Optimization Manager, such as network charges

Points to consider:

  - For VMs fully covered by AWS Reserved Instances, Azure reservations, or Google Cloud committed use discounts, on-demand cost is 0 (zero).
  - For Azure VMs, the chart shows cost data from the day the VM was discovered by Workload Optimization Manager. Cost data before discovery is not available and therefore not reflected in the chart.

Chart Type

You can set the display to:

- Line Chart
- Stacked Bar Chart
- Area Chart

Chart time frame

The chart can display data from the last 7, 30, or 60 days, or the last year. As you change the time frame, Workload Optimization Manager divides the reported information into the appropriate time units to match that time frame. However, the source remains the same. Changing the time frame does not affect the source data or increase data polling.

After installing Workload Optimization Manager for the first time and adding a new cloud target, the amount of cost data that Workload Optimization Manager retrieves varies by cloud provider. For Google Cloud, Workload Optimization Manager retrieves 60 days of data. For AWS and Azure, Workload Optimization Manager retrieves data for the current and previous months. This means that on the day you installed the product and added a new AWS or Azure target, the chart might not show 60 days of data if you select 60 days as the time frame. For example, if the current date is July 15, you will see data from June 1 to July 14 (or 44 days of data). You will see 60 days of data on August 1.

Entity breakdown charts

Entity breakdown charts show the entities that Workload Optimization Manager discovers according to the scope and entity type that you selected. For each entity type, you can use a **Group By** option to break down the discovered entities by state or tier.

For example, if you select a region as your scope, **Virtual Machine** as the entity type, and **State** as the **Group By** option, the chart shows the virtual machines discovered in the region and the number of virtual machines that are active and idle.

Scopes and entity types

The chart supports the following scopes and entity types:

Scope	Entity type
Database	Database
Database server	Database server
Region	All available entity types
Resource group (Azure only)	Virtual machine, virtual machine specs
Virtual machine	Virtual machine, virtual machine specs
Auto-generated groups	All available entity types

Chart type

You can set the display to:

- Ring Chart
- Horizontal Bar
- Tabular

Group by

The following options are supported:

- Tier – Break down entities by cloud tier.
- State – Break down entities by operational state.

Location chart

NOTE:
This chart is deprecated and will be removed from the product in a future release. Delete this chart from any dashboard.

Location charts show cloud provider regions in a world map for which there are discovered workloads. Click on any region to examine more detailed information in a scoped view.

Display

The chart shows the regions in countries in a Map chart.

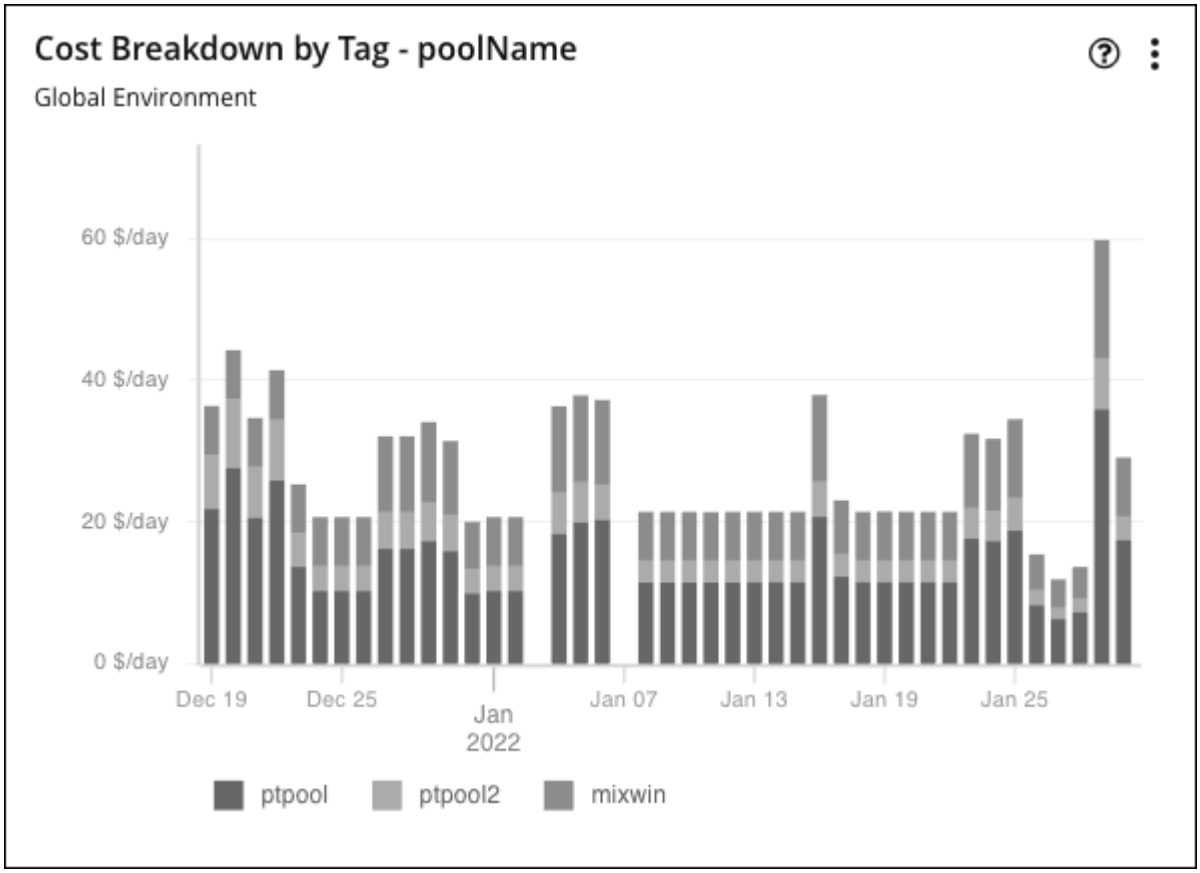
Cost breakdown by tag chart

NOTE:
This chart is deprecated and will be removed from the product in a future release. Delete this chart from any dashboard and use the replacement chart, Cost breakdown.

Cost Breakdown By Tag charts show the costs for tagged cloud entities that Workload Optimization Manager discovered in your AWS, Azure, or Google Cloud environment. For the tagged entities in scope, the chart shows how daily costs change over time.

You choose a tag key to track, and then choose which tag values to include in the chart. Each data point aggregates the costs for all the entities with a given tag/value pair. You can display the cost breakdown in a stacked bar chart or an area chart.

For example, in this stacked bar chart, the tag **poolName** is *workload-type* and the tag **Values** are *ptpool*, *ptpool2*, and *mixwin*.



NOTE:
If you set up an AWS standard data export (CUR 2.0) for use with the AWS Billing target, be aware that AWS replaces tags with special characters and camel cases with underscores (_) in the data export. As a result, tags shown in AWS Cost Explorer and in the Cost Breakdown by Tag chart do not match.

Scope

To display these charts, add them to the default views in the Home Page or to your custom dashboards. By default, these charts are scoped to your global environment. You can change the scope to view granular data.

Timeframe

Currently, the chart can display data from the last 7 or 30 days.

Chart Type

You can set the display to:

- Area Chart
- Stacked Bar Chart

For more detail, hover over a data point. A tooltip appears to show specific values for that date. Click the legend items to show/hide data for specific values.

Tag settings

Choose the Tag/Value pairs you want to display in the chart.

- **Key**: The tag name that you want to chart. Workload Optimization Manager discovers the tags you have configured in your environment.

  You can choose one Key for the chart.

- **Values**: The values that you have configured in your environment for the given Key.

  You can choose multiple values. To shorten the list of values, type a filter string in the Values field.

NOTE:
Tag keys and values are case insensitive, but tag keys are normalized to lowercase. If there are duplicate lowercase keys for an entity, only the smallest tag value is retained.

Cumulative savings and investments charts

Actions for your cloud workloads usually have cost savings or investments attached to them. For example, deleting unattached volumes can lower your costs significantly (savings), while scaling a VM to a different tier to improve performance could incur additional costs (investments). These charts highlight total savings and investments as a result of executing actions directly in Workload Optimization Manager and externally (for example, through your cloud provider). Workload Optimization Manager uses billing data from your cloud provider to calculate savings and investments.

These charts display in the built-in Cloud Executive Dashboard. You can add these charts to the default views in the Home Page or to your custom dashboards.

Points to consider:

- The method for tracking savings and investments has been enhanced to accurately report the impact of actions on your cloud expenses. This enhanced tracking began in version 3.7.3. For cumulative savings and investments before this release, see the **Legacy Cumulative Savings** and **Legacy Cumulative Investments** charts.
- Billing data from your cloud provider generally takes 1 to 2 days to update fully. Savings and investment figures may change until the data is fully updated.
- In the Cloud Executive Dashboard, scoped users can see data for the entire cloud environment.

Scale actions

Scale actions involve moving an entity to a different compute tier or adjusting the entity's allocated capacity. Currently, Workload Optimization Manager calculates savings and investments when scaling the following entities:

- VMs (AWS, Azure, and Google Cloud VMs)
- Volumes (AWS and Azure volumes)
- Databases (Azure SQL vCore/DTU and Cosmos DB)
- Document collections (for Cosmos DB)
- Virtual Machine Specs (Azure App Service plans)
- Database Servers (AWS RDS and Azure SQL Managed Instances)

Workload Optimization Manager calculates savings and investment *per entity*. Calculation is based on the *before-action* and *after-action* costs.

- Before-action cost is the cost of an entity before an action was executed. This is based on the cost snapshot on the day of action execution.
- After-action cost is the cost of an entity as reflected in the daily billing report from the cloud provider.

Savings and investments are the total of all past scaling actions, with the exception that a scale in one direction (such as a scale up) reduces the amounts of previous actions in the opposite direction (such as a scale down), until one or more previous actions have no more effect.

To illustrate, consider three consecutive scale actions for a newly discovered VM and the effect of these actions on the calculated costs. Note that these actions are assumed to have been executed at midnight. Costs are calculated proportionately when actions are executed at other times of the day.

Day	Executed action	Before-action cost	After-action cost	Investments	Savings
01	Scale up	$2	$5	$5 - $2 = **$3**	**$0**
02	Scale down	$5	$4	$4 - $2 = **$2**	$5 - $4 = **$1**
03	Scale down	$4	$2	$2 - $2 = **$0**	$5 - $2 = **$3**

Investment is the difference between the after-action cost on a particular day and the lowest ever before-action cost up to that day ($2 for all three days in the example). Savings is the difference between the largest ever before-action cost up to a particular day ($5 on Days 02 and 03) and the after-action cost on that day.

> **NOTE:**
> When the result of the calculation is a negative amount, Workload Optimization Manager considers savings or investments to be $0.

Workload Optimization Manager reads the daily billing report from the cloud provider to determine the current daily cost of an entity. Calculation also takes into account workload uptime (for VMs) and the effect of consecutive scale actions on the same workload.

Points to consider:

- Calculation can adjust to varying VM uptime and discount coverage over time.
- Calculation stops for terminated entities or entities that Workload Optimization Manager no longer discovers.
- Calculation stops if there is a cost change in the entity that is *not* the result of a Workload Optimization Manager action.

Delete actions

Delete actions always result in savings. Currently, Workload Optimization Manager calculates savings when deleting the following entities:

- Unattached volumes (AWS, Azure, and Google Cloud volumes)
- Empty Virtual Machine Specs (Azure App Service plans)
- Cosmos databases with provisioned throughput but without any underlying document collection

Workload Optimization Manager checks the daily cost of an entity on the day the entity was deleted, and uses that cost to calculate savings from that day onwards. The daily bill from the cloud provider is not used to calculate savings.

Stop and suspend actions

Stop and suspend actions always result in savings. Currently, Workload Optimization Manager calculates savings when:

- Stopping parkable entities *(on page 943)*
- Suspending idle dedicated SQL pools (used in Azure Synapse Analytics)

  > **NOTE:**
  > Dedicated SQL pools are represented as database entities in the supply chain.

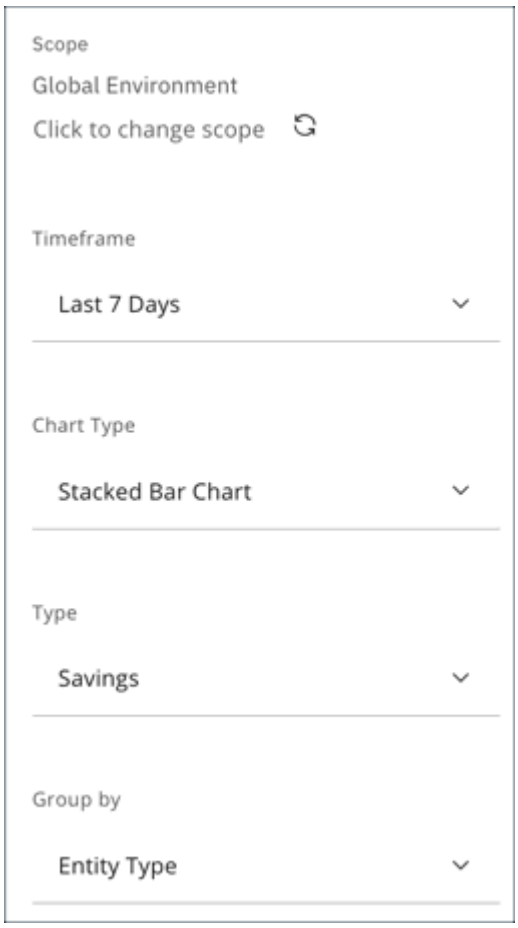
- Suspending idle Amazon Redshift provisioned clusters

  > **NOTE:**
  > Amazon Redshift provisioned clusters are represented as data warehouse entities in the supply chain.

Workload Optimization Manager checks the hourly cost of an entity's instance type at the time the action was executed, and then uses that cost to calculate savings for the duration of the action. Calculation stops as soon as the entity is powered on. The daily bill from the cloud provider is not used to calculate savings.

Chart settings

Click the More options icon ( ), and then select **Edit** to modify the following settings:



- Scope

  By default, the charts are scoped to your global environment. You can change the scope to one or several accounts, billing families, groups, or entities.

- Timeframe
  - Last 7 Days or Last 30 Days – Each data point in the chart shows total savings or investments from previous days until the given day.
  - Last Year – Each data point in the chart shows total savings or investments from previous months until the given month.
- Chart Type

  You can set the display to:
  - Stacked Bar Chart
  - Area Chart
- Type

  Switch between the **Cumulative Savings** and **Cumulative Investments** views. You can also change the displayed data to just **Savings** or **Investments** if you do not wish to see how the savings or investment costs accumulate over time.
- Group By

This setting breaks down data points by the selected group. By default, no group is selected. Each data point only shows the total cumulative values for a given day or month, depending on the selected timeframe.

The following groups are supported:

- Account – Cloud accounts discovered in your environment
- Action type – Scale and delete actions
- Cloud service provider – AWS, Azure and Google Cloud
- Entity type – Virtual Machine, Volume, Database (Azure only), Database Server (AWS and Azure only), Document Collection (Azure only), Virtual Machine Spec (Azure only)
- Region – Regions for the entities
- Resource group (Azure only) – Available resource groups in your environment
- Tag – Tag values for a given tag key

  Select a tag key from the list. For AWS, the tag keys are limited to cost allocation tags.

  NOTE:
  Workload Optimization Manager stores and displays tag keys in lowercase. If you use these tags in Workload Optimization Manager API requests, specify them in lowercase. Otherwise, the API returns incorrect results.

For example, if you selected **Account**:

- The chart represents each account as a distinct color.
- When you hover on a data point, the chart shows a list of accounts, ordered from largest to smallest cumulative values.
- The chart legend shows the individual accounts. Click an account to show/hide values for that account.
- The chart and legend show up to 20 accounts and, if needed, an additional item labeled 'Others' that represents all accounts that are not in the top 20.

Show All

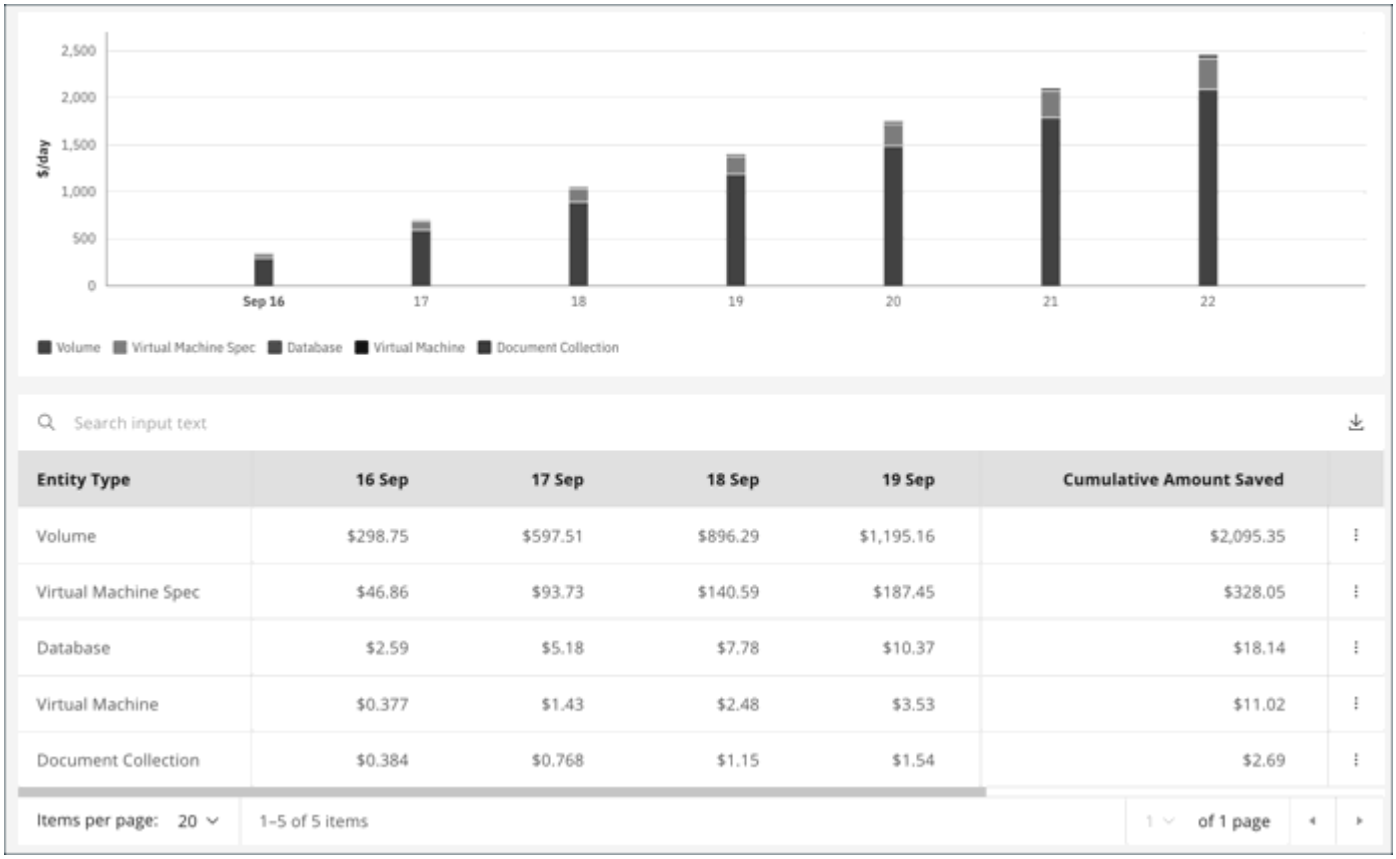
Click **Show All** to view data as a stacked bar or area chart (depending on your selection) and as a table.

The table shows detailed information and provides the following functionality:

- If you filtered by a specific group, the table provides a Search functionality to help you narrow down the data further. For example, if you filtered by account, you can search for account names by specifying an account name as your search keyword. Partial keyword matching is supported.

  The Search functionality is not available if you did not set a filter.

- You can break down data by resource names. To do this, go to the last column, click the icon for a particular row, and then select **Show more details**.
- Data in the table is downloadable.



Entity Type	16 Sep	17 Sep	18 Sep	19 Sep	Cumulative Amount Saved
Volume	$298.75	$597.51	$896.29	$1,195.16	$2,095.35
Virtual Machine Spec	$46.86	$93.73	$140.59	$187.45	$328.05
Database	$2.59	$5.18	$7.78	$10.37	$18.14
Virtual Machine	$0.377	$1.43	$2.48	$3.53	$11.02
Document Collection	$0.384	$0.768	$1.15	$1.54	$2.69

Legacy cumulative savings or investments chart

NOTE:
This chart is deprecated and will be removed from the product in a future release. Delete this chart from any dashboard.

These charts show cumulative savings and investments before version 3.7.3. The method for tracking savings and investments was enhanced in version 3.7.3, and data from this enhanced tracking is available in the **Cumulative Savings** and **Cumulative Investments** charts.

Actions for your cloud workloads usually have cost savings or investments attached to them. For example, deleting unattached volumes can lower your costs significantly (savings), while scaling a VM to a different tier to improve performance could incur additional costs (investments).

These charts highlight:

- Total *realized* savings and investments as a result of executing actions
- Total *missed* savings and investments when actions are not executed

Information in these charts can help determine your action handling policies. For example, you can start automating actions so you don't miss opportunities to assure performance at the lowest possible cost.

Scope

These charts display in the built-in Cloud Executive Dashboard and are scoped to your global environment. You can change the scope to view granular data. You can also add these charts to the default views in the Home Page or to your custom dashboards.

Another way to view granular data is to set the scope (in the supply chain or by using Search) to one or several accounts, billing families, groups, or workloads.

Scale actions

For actions to scale workloads (VMs, Database Servers, databases, or volumes), Workload Optimization Manager calculates savings and investment *per workload* based on the hourly cost of the workload price difference, taking into account workload uptime *(on page 647)* and the effect of consecutive scale actions on the same workload.

- Calculated investments and savings are the total of all past scaling actions, with the exception that a scale in one direction reduces the amounts of previous actions in the opposite direction, until one or more previous actions have no more effect.

  To illustrate:

  Consider three consecutive scale actions for a VM and their effect on the calculation.

  1. A cost increase of $1.00 counts as an investment of $1.00.
  2. A subsequent cost decrease of $0.25 is factored in as:
     - Savings of $0.25 to the total amount in the Cumulative Savings chart
     - An investment of $0.75 to the total amount in the Cumulative Investments chart
  3. Another cost decrease of $1.00 is factored in as:
     - Savings of $1.25 to the total amount in the Cumulative Savings chart
     - An investment of $0.00 to the total amount in the Cumulative Investments chart

  By the time the third action was executed, the initial $1.00 investment has been "undone" (investment amount is $0.00) and is no longer considered when calculating savings and investments for the VM.
- Calculation can adjust to varying VM uptime and discount coverage over time.
- Calculation stops for terminated entities or entities that Workload Optimization Manager no longer discovers.
- Calculation stops if there is a cost change in the entity that is *not* the result of a Workload Optimization Manager action.

Volume delete actions

For actions to delete volumes, Workload Optimization Manager calculates savings accumulated over one year since volume deletion, based on the hourly cost of the deleted volume. It also estimates missed savings based on the hourly cost of the workload price difference and the number of hours that pending actions remain in the system.

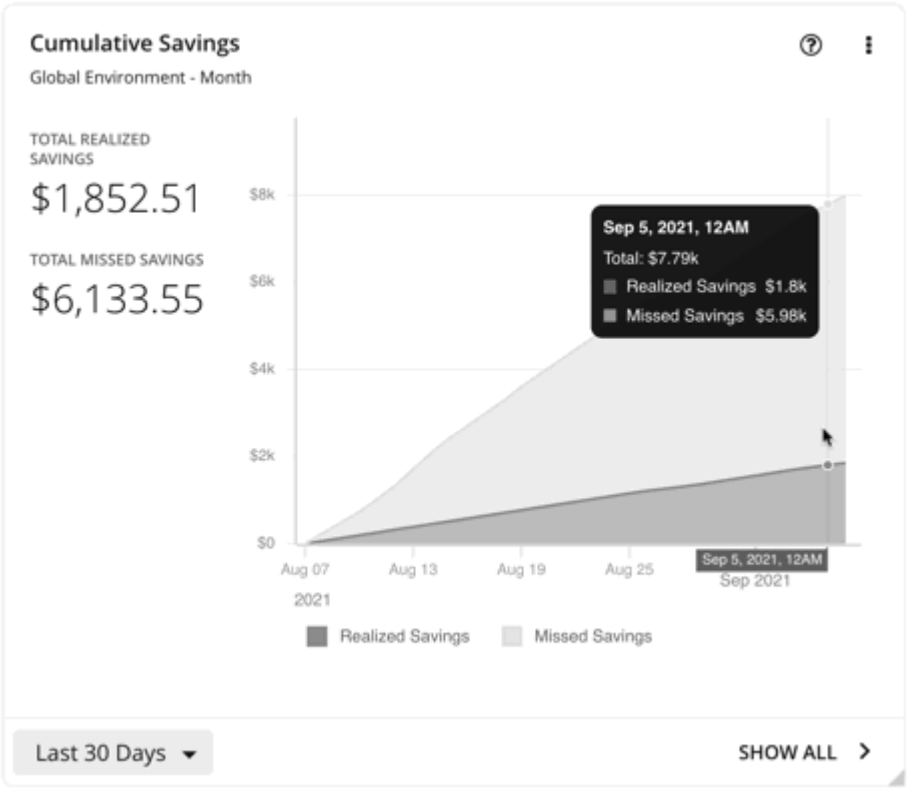
Chart Type

You can set the display to:

- Stacked Bar Chart
- Area Chart

You can edit the chart to switch between the **Cumulative Savings** and **Cumulative Investments** views. You can also change the displayed data to just **Savings** or **Investments** if you do not wish to see how the savings or investment costs accumulate over time.

In this example, Workload Optimization Manager shows monthly realized and missed savings.



In the chart legend, you can click **Realized Savings** or **Missed Savings** to display a filtered view. Click the item again to reset the chart.

Click **Show All** at the end of the chart to view and download data in tabular format.

Savings and investments charts

Actions for your cloud workloads usually have cost savings or investments attached to them. For example, deleting unattached volumes can lower your costs significantly (savings), while scaling a VM to a different tier to improve performance could incur additional costs (investments). These charts highlight total savings and investments as a result of executing actions directly in Workload Optimization Manager and externally (for example, through your cloud provider). Workload Optimization Manager uses billing data from your cloud provider to calculate savings and investments.

You can add these charts to the default views in the Home Page or to your custom dashboards.

Points to consider:

- The method for tracking savings and investments has been enhanced to accurately report the impact of actions on your cloud expenses. This enhanced tracking began in version 3.7.3. For savings and investments before this release, see the **Legacy Cumulative Savings** and **Legacy Cumulative Investments** charts.
- Billing data from your cloud provider generally takes 1 to 2 days to update fully. Savings and investment figures may change until the data is fully updated.

Scale actions

Scale actions involve moving an entity to a different compute tier or adjusting the entity's allocated capacity. Currently, Workload Optimization Manager calculates savings and investments when scaling the following entities:

- VMs (AWS, Azure, and Google Cloud VMs)
- Volumes (AWS and Azure volumes)
- Databases (Azure SQL vCore/DTU and Cosmos DB)
- Document collections (for Cosmos DB)

- Virtual Machine Specs (Azure App Service plans)
- Database Servers (AWS RDS and Azure SQL Managed Instances)

Workload Optimization Manager calculates savings and investment *per entity*. Calculation is based on the *before-action* and *after-action* costs.

- Before-action cost is the cost of an entity before an action was executed. This is based on the cost snapshot on the day of action execution.
- After-action cost is the cost of an entity as reflected in the daily billing report from the cloud provider.

NOTE:
When the result of the calculation is a negative amount, Workload Optimization Manager considers savings or investments to be $0.

Workload Optimization Manager reads the daily billing report from the cloud provider to determine the current daily cost of an entity. Calculation also takes into account workload uptime (for VMs).

Points to consider:

- Calculation can adjust to varying VM uptime and discount coverage over time.
- Calculation stops for terminated entities or entities that Workload Optimization Manager no longer discovers.
- Calculation stops if there is a cost change in the entity that is *not* the result of a Workload Optimization Manager action.

Delete actions

Delete actions always result in savings. Currently, Workload Optimization Manager calculates savings when deleting the following entities:

- Unattached volumes (AWS, Azure, and Google Cloud volumes)
- Empty Virtual Machine Specs (Azure App Service plans)
- Cosmos databases with provisioned throughput but without any underlying document collection

Workload Optimization Manager checks the daily cost of an entity on the day the entity was deleted. The daily bill from the cloud provider is not used to calculate savings.

Stop and suspend actions

Stop and suspend actions always result in savings. Currently, Workload Optimization Manager calculates savings when:

- Stopping parkable entities *(on page 943)*
- Suspending idle dedicated SQL pools (used in Azure Synapse Analytics)

  NOTE:
  Dedicated SQL pools are represented as database entities in the supply chain.

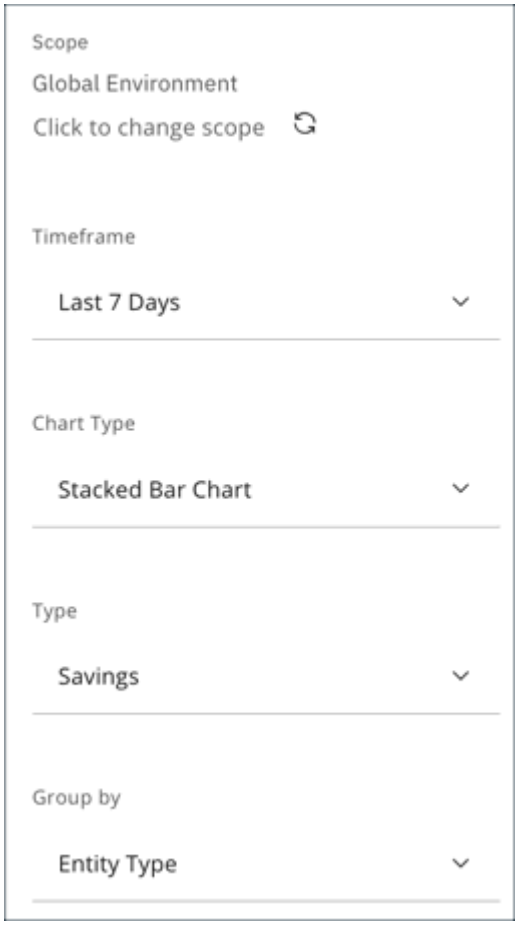
- Suspending idle Amazon Redshift provisioned clusters

  NOTE:
  Amazon Redshift provisioned clusters are represented as data warehouse entities in the supply chain.

Workload Optimization Manager checks the hourly cost of an entity's instance type at the time the action was executed, and then uses that cost to calculate savings for the duration of the action. Calculation stops as soon as the entity is powered on. The daily bill from the cloud provider is not used to calculate savings.

Chart Settings

Click the More options icon (⋮), and then select **Edit** to modify the following settings:



- Scope

  By default, the charts are scoped to your global environment. You can change the scope to one or several accounts, billing families, groups, or entities.
- Timeframe
  - Last 7 Days or Last 30 Days – Each data point in the chart shows total savings or investments for the given day.
  - Last Year – Each data point in the chart shows total savings or investments for the given month.
- Chart Type

  You can set the display to:
  - Stacked Bar Chart
  - Area Chart
- Type

  Switch between the **Savings** and **Investments** views. You can also change the displayed data to **Cumulative Savings** or **Cumulative Investments** to see how the savings or investment costs accumulate over time.
- Group By

  This setting breaks down data points by the selected group. By default, no group is selected. Each data point only shows the total values for a given day or month, depending on the selected timeframe.

  The following groups are supported:
  - Account – Cloud accounts discovered in your environment
  - Action type – Scale and delete actions
  - Cloud service provider – AWS, Azure and Google Cloud

- Entity type – Virtual Machine, Volume, Database (Azure only), Database Server (AWS and Azure only), Document Collection (Azure only), Virtual Machine Spec (Azure only)
- Region – Regions for the entities
- Resource group (Azure only) – Available resource groups in your environment
- Tag – Tag values for a given tag key

  Select a tag key from the list. For AWS, the tag keys are limited to cost allocation tags.

  NOTE:
  Workload Optimization Manager stores and displays tag keys in lowercase. If you use these tags in Workload Optimization Manager API requests, specify them in lowercase. Otherwise, the API returns incorrect results.

For example, if you selected **Account**:

- The chart represents each account as a distinct color.
- When you hover on a data point, the chart shows a list of accounts, ordered from largest to smallest values.
- The chart legend shows the individual accounts. You can click an account to show/hide values for that account.
- The chart and legend show up to 20 accounts and, if needed, an additional item labeled 'Others' that represents all accounts that are not in the top 20.

Show All

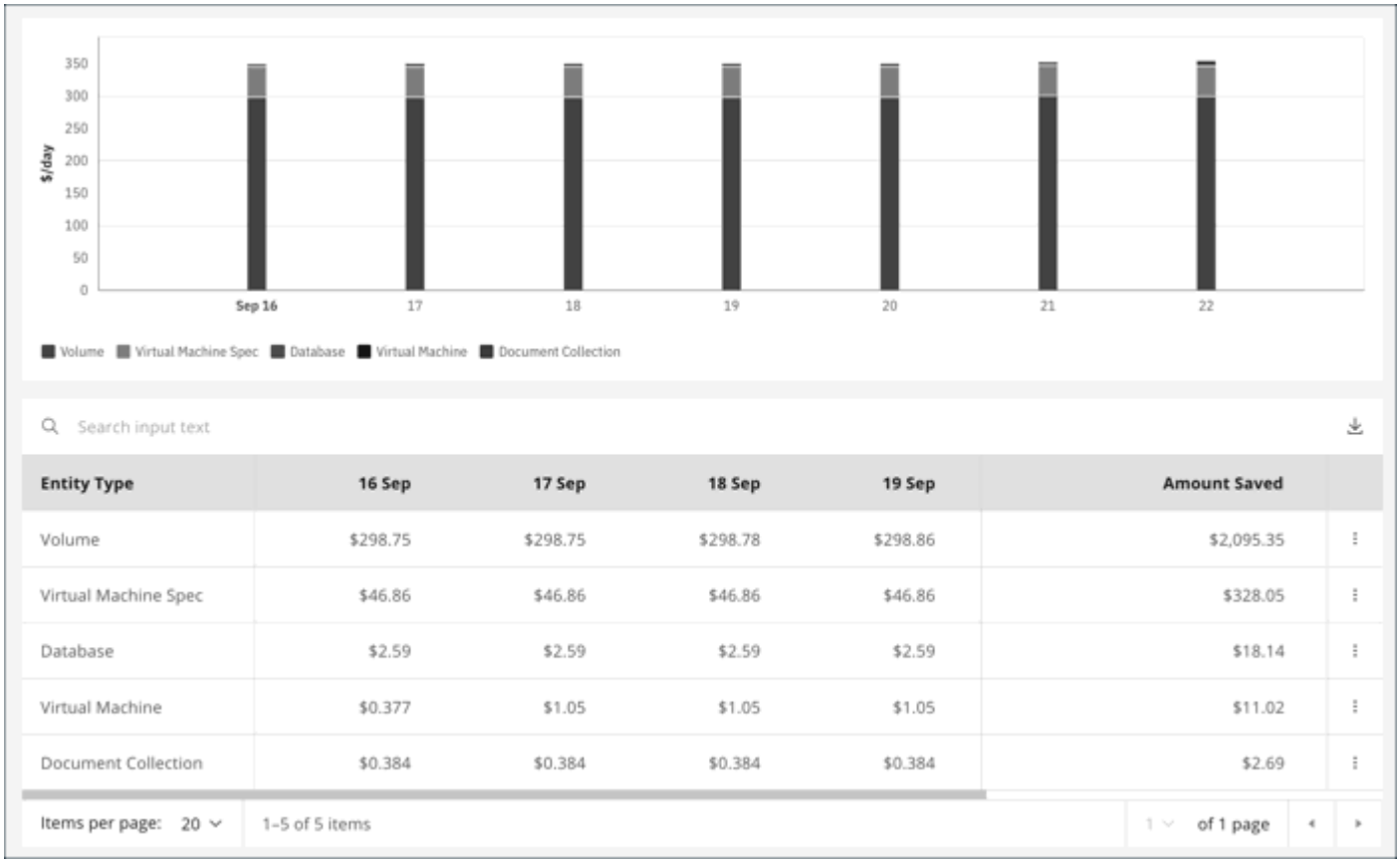
Click **Show All** to view data as a stacked bar or area chart (depending on your selection) and as a table.

The table shows detailed information and provides the following functionality:

- If you filtered by a specific group, the table provides a Search functionality to help you narrow down the data further. For example, if you filtered by account, you can search for account names by specifying an account name as your search keyword. Partial keyword matching is supported.

  The Search functionality is not available if you did not set a filter.

- You can break down data by resource names. To do this, go to the last column, click the icon for a particular row, and then select **Show more details**.
- Data in the table is downloadable.



Legacy savings or investments chart

NOTE:

NOTE:
This chart is deprecated and will be removed from the product in a future release. Delete this chart from any dashboard.

These charts show savings and investments before version 3.7.3. The method for tracking savings and investments was enhanced in version 3.7.3, and data from this enhanced tracking is available in the **Savings** and **Investments** charts.

Actions for your cloud workloads usually have cost savings or investments attached to them. For example, deleting unattached volumes can lower your costs significantly (savings), while scaling a VM to a different tier to improve performance could incur additional costs (investments).

These charts highlight:

- Total *realized* savings and investments as a result of executing actions
- Total *missed* savings and investments when actions are not executed

Information in these charts can help determine your action handling policies. For example, you can start automating actions so you don't miss opportunities to assure performance at the lowest possible cost.

Scope

To display these charts, add them to the default views in the Home Page or to your custom dashboards. By default, these charts are scoped to your global environment. You can change the scope to view granular data.

Another way to view granular data is to set the scope (in the supply chain or by using Search) to one or several accounts, billing families, groups, or workloads.

Scale actions

For actions to scale workloads (VMs, Database Servers, databases, or volumes), Workload Optimization Manager calculates savings and investment *per workload* based on the hourly cost of the workload price difference, taking into account workload uptime *(on page 647)*.

- Calculation can adjust to varying VM uptime and discount coverage over time.
- Calculation stops for terminated entities or entities that Workload Optimization Manager no longer discovers.
- Calculation stops if there is a cost change in the entity that is *not* the result of a Workload Optimization Manager action.

Volume delete actions

For actions to delete volumes, Workload Optimization Manager calculates savings since volume deletion, based on the hourly cost of the deleted volume. It also estimates missed savings based on the hourly cost of the workload price difference and the number of hours that pending actions remain in the system.

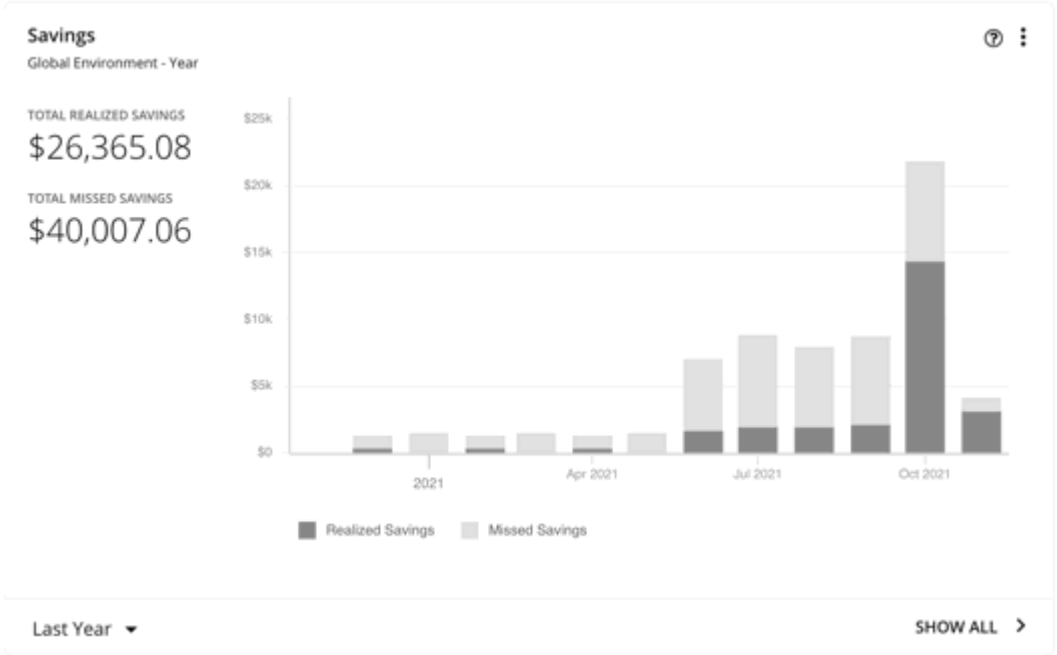
Chart Type

You can set the display to:

- Stacked Bar Chart
- Area Chart

You can edit the chart to switch between the **Savings** and **Investments** views. You can also change the displayed data to **Cumulative Savings** or **Cumulative Investments** to see how the savings or investment costs accumulate over time.

In this example, the chart shows realized and missed savings per month over the last year. It indicates higher rates of realized savings in the last two months as more actions are executed rather than kept pending.



In the chart legend, you can click **Realized Savings** or **Missed Savings** to display a filtered view. Click the item again to reset the chart.

Click **Show All** at the end of the chart to view and download data in tabular format.

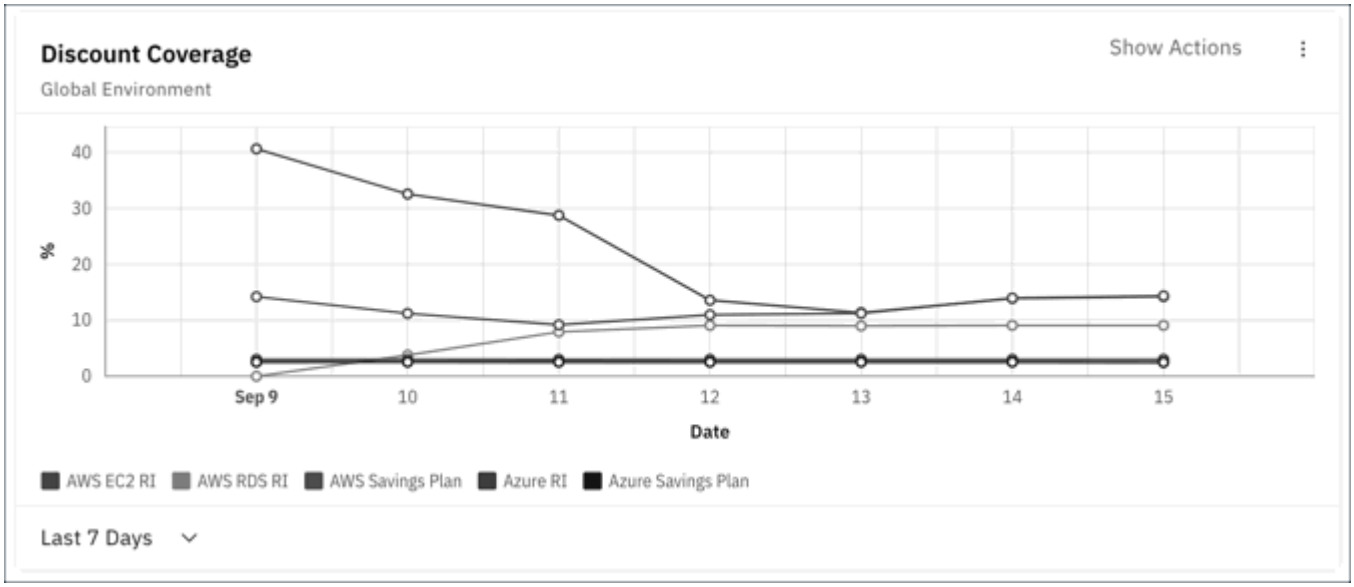
Recommended RI purchases chart

This chart has been removed from the cloud dashboard. Delete it from any dashboard.

This chart remains available in plan results.

Discount coverage chart

This chart shows the percentage of cloud workloads (VMs and RDS database servers) covered by discounts. For VMs covered by discounts, you can reduce your costs by increasing coverage. To increase coverage, you scale VMs to instance types that have existing capacity.



To identify VMs that are good candidates for discounted pricing, Workload Optimization Manager analysis considers the history of a VM (by default, the last 21 days), and it looks for:

- Activity

  If the VM's VCPU utilization percentile is 20% or higher, then Workload Optimization Manager considers it an active VM.
- Stability

  If there have been no start, stop, or resize actions for the VM for 16 of the last 21 days, then Workload Optimization Manager considers it stable.

If the current discount inventory cannot support the VM, or if supporting it would exceed your desired coverage, then Workload Optimization Manager can recommend purchasing additional discounts.

Timeframe

The chart shows daily data points and supports the following timeframes:

- Last 7 Days
- Last 30 Days

AWS EC2/RDS reserved instances and compute savings plans

Workload Optimization Manager uses data exports from AWS to calculate Reserved Instance (RI) and Compute Savings Plan coverage for both standard and GovCloud workloads. Workload Optimization Manager supports a data export created at the management account, but not member accounts. For Workload Optimization Manager to use these data exports, you must add a billing target in the Target Configuration page.

The chart shows historical data. Data for the current day may not be available until the cloud provider has exported it fully.

Hover on a data point in the chart to see the following information:

- The date and time for the data point
- The percentage of coverage. For RIs, coverage is based on normalization factors.

  Normalization factor is a measure of RI capacity that you can use to compare or combine the capacity for different instance families.

  Workload Optimization Manager measures RI coverage in terms of normalization factors. It compares the number of RIs calculated as normalization factors that cover workload capacity with the total number of normalization factors for a given Workload Optimization Manager scope. Each workload is assigned normalized units depending on its instance type.

Points to consider:

- Scope
  - If you set the scope to a specific AWS account, the chart shows the RI coverage for the workloads for the account and any RIs for the billing family.
  - For AWS GovCloud, you can set the scope to the global environment or the standard account/billing family associated with GovCloud.
- RIs

  In AWS, you can turn off RI discount sharing for specific accounts. These accounts will not share any discounts with other accounts. Workload Optimization Manager does not recognize RI coverage or utilization for these accounts. For example, the RI Coverage and RI Utilization charts will show zero values.

Azure reservations and savings plans

Workload Optimization Manager uses billed cost data from Azure to calculate reservation and savings plan coverage for workloads. For Workload Optimization Manager to use billed cost data, you must add an Azure Billing target in the Target Configuration page.

> **NOTE:**
> The use of billed cost data to calculate reservation or savings plan coverage for Azure Government workloads is not supported.

The chart shows historical data. Data for the current day may not be available until the cloud provider has exported it fully.

Hover on a data point in the chart to see the following information:

- The date and time for the data point
- The percentage of coverage
  - For reservations, the percentage of coverage is based on ratios.

    Ratio refers to the number of Azure reservation units that cover workload capacity compared to the total number of reservation units for a given Workload Optimization Manager scope. Each workload is assigned reservation units based on its instance type.
  - For savings plans, the percentage of coverage is based on billed cost data.

If you set the scope to a specific Azure subscription, this chart shows the following:

- Reservation or savings plan coverage for the workloads for the subscription
- Shared reservations or savings plans
- Single-scope reservations or savings plans owned by the subscription

Google Cloud committed use discounts

Workload Optimization Manager uses billing data from Google Cloud to calculate CUD coverage for VM vCPU (cores) and VM memory. For Workload Optimization Manager to use billing data, you must add a Google Cloud Billing target in the Target Configuration page.

The chart shows historical data. Data for the current day may not be available until the cloud provider has exported it fully.

Hover on a data point in the chart to see the following information:

- The date and time for the data point
- The percentage of coverage

Points to consider:

- CUD coverage in this chart and in the Google Cloud console may not match for the following reasons:
  - Workload Optimization Manager uses UTC, while the Google Cloud console uses local time.
  - CUD coverage for VMs running custom machine types is reported by Google Cloud but not by Workload Optimization Manager. If the chart's scope includes these VMs, aggregated data may not match the data shown in Google Cloud. If the chart is scoped solely to these VMs, the chart will not display data.
  - Google Cloud reports CUD coverage for resource reservations, which Workload Optimization Manager does not monitor. When CUDs cover both VMs and resource reservations, Google Cloud reports coverage for both, while Workload Optimization Manager only reports coverage for VMs.
- When you configure a billing export in Google Cloud, you can choose to export standard or detailed usage cost data. Detailed usage cost data is recommended because it includes granular data that Workload Optimization Manager can display in charts, such as discount coverage for individual VMs.

  If you exported standard usage cost data and configured your Google Cloud Billing target to use this data:
  - Coverage data in the chart is only available for scopes that are larger than individual VMs, such as projects or folders. This is because standard usage cost data does not include CUD coverage for individual VMs.
  - Workload Optimization Manager uses retail pricing to calculate coverage. This may result in an underestimation of the coverage if the negotiated price for a machine type is lower than the retail price.

  NOTE:
  The Google Cloud Billing target uses standard usage cost data if you disabled the **Enable Resource Level Detail From Cost Export Table** option in the Target Configuration page for the target.

- When coverage reaches 100%, a rounding issue might result in a value that is slightly higher than 100% (such as 100.01%).

Discount inventory chart

This chart lists the cloud provider discounts that you have purchased for your workloads.

The following table describes the current level of support for the different cloud provider discounts. Discounts that are not listed are not supported.

Discount	Discovered	Considered in analysis
AWS EC2 Reserved Instances (RIs)	Yes	Yes
AWS Compute Savings Plans	Yes	Yes

Discount	Discovered	Considered in analysis
AWS RDS Reserved DB Instances	Yes	No
Azure reservations	Yes	Yes
Azure savings plans	Yes	No
Google Cloud committed use discounts	Yes	Yes

Timeframe

The chart shows daily data points and supports the following timeframes:

- Last 7 Days
- Last 30 Days

Scope

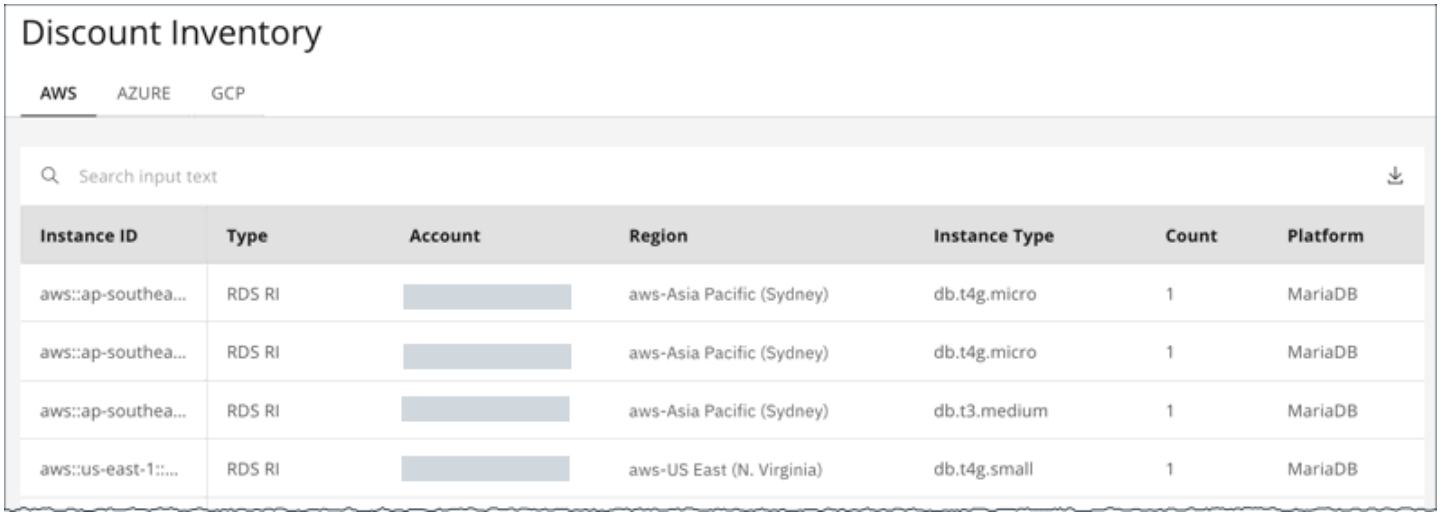
You can set the scope to your global cloud environment or to individual AWS accounts, Azure subscriptions, Google Cloud folders, billing families, or regions.

For AWS GovCloud Savings Plans, you can set the scope to the global environment or the standard account/billing family associated with GovCloud.

For AWS GovCloud RIs, you can set the scope to the global environment, or a GovCloud account or billing family.

Show All

Click **Show All** on the chart to see detailed information for the discounts in scope. If your scope includes multiple cloud providers, each provider will have its own tab.



Each row in the table corresponds to a discount. Note that there can be several discounts for an Azure subscription or AWS/Google Cloud account, and each discount displays as its own row. Table columns show basic information obtained from the cloud provider, such as the name/ID of the discount, the subscription/account that uses that discount, instance type and location, term, and expiration dates. Click a subscription/account to narrow the scope.

The table supports the following general functionality:

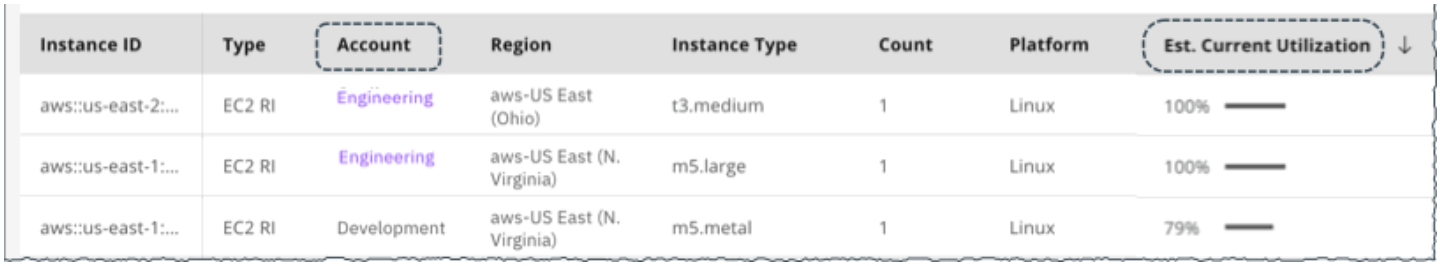
- **Column Sorting**: Click any column heading to sort the list.
- **Download**: Click the Download icon to download the table as a CSV file.

AWS EC2 RIs

When you add an AWS management account as your primary cloud target, Workload Optimization Manager gains full insight into the RIs for your billing families. Even as you selectively add member accounts as secondary targets, Workload Optimization Manager remains aware of all RIs and how they are utilized across the board, and can thus recommend more accurate RI optimization and purchase actions.

> **NOTE:**
> If you added some member accounts as targets, but not a management account, Workload Optimization Manager will not reflect RIs for member accounts that you have not added as targets.

Set the scope to your global environment to view the full inventory. When you click **Show All** in the chart, pay attention to the following columns:



Instance ID	Type	Account	Region	Instance Type	Count	Platform	Est. Current Utilization
aws::us-east-2:...	EC2 RI	Engineering	aws-US East (Ohio)	t3.medium	1	Linux	100%
aws::us-east-1:...	EC2 RI	Engineering	aws-US East (N. Virginia)	m5.large	1	Linux	100%
aws::us-east-1:...	EC2 RI	Development	aws-US East (N. Virginia)	m5.metal	1	Linux	79%

RI	Account column	Est. Current Utilization column
RIs in AWS member accounts that you added to Workload Optimization Manager as targets	This column shows the account name for the RI. Click the name to set the scope to that account. Note that there can be several RIs for an account, and each RI displays as its own row. **NOTE:** If there is a failure to re-validate the account for some reason, Workload Optimization Manager shows it as a *non-added* account in the Discount Inventory page.	This column shows the percentage of RI capacity currently used by VMs in all accounts. Workload Optimization Manager estimates the percentage if there are VMs in non-added accounts that use the RI (since the exact number of VMs is unknown).
RIs in AWS member accounts that you did not add to Workload Optimization Manager	This column shows a grayed-out, non-clickable name to indicate that you have not added the account as a target. Workload Optimization Manager is aware of this account and if the given RI is shared with other accounts because you have added a management account.	This column shows the percentage of RI capacity currently used by VMs in all accounts. Workload Optimization Manager estimates the percentage if there are VMs in non-added accounts that use the RI (since the exact number of VMs is unknown).

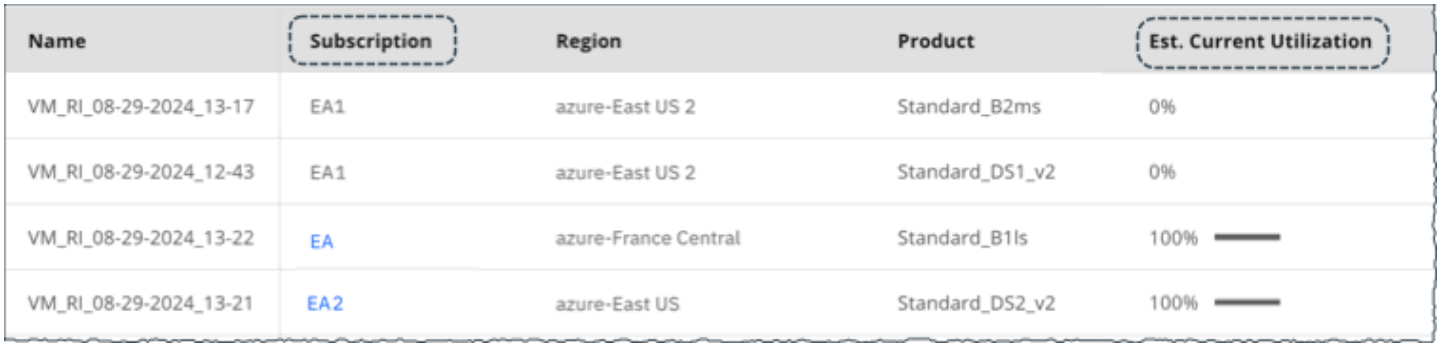
Azure reservations and savings plans

When you add an Azure EA or MCA account as your primary cloud target, Workload Optimization Manager gains full insight into the reservations and savings plans for your billing families. Even as you selectively add subscriptions as secondary targets, Workload Optimization Manager remains aware of all reservations and savings plans, and how they are utilized across the board. For reservations, Workload Optimization Manager can recommend optimization and purchase actions based on your current inventory.

Points to consider:

- It could take Workload Optimization Manager up to a day to discover newly purchased Azure reservations and savings plans.
- There can be delays in billing information updates that Azure makes available to Workload Optimization Manager. If this happens, analysis might use partial billing data in its calculations and show incomplete costs for reservations and savings plans from non-added Azure subscriptions.

Set the scope to your global environment to view the full inventory. When you click **Show All** in the chart, pay attention to the following columns:



Name	Subscription	Region	Product	Est. Current Utilization
VM_RI_08-29-2024_13-17	EA1	azure-East US 2	Standard_B2ms	0%
VM_RI_08-29-2024_12-43	EA1	azure-East US 2	Standard_DS1_v2	0%
VM_RI_08-29-2024_13-22	EA	azure-France Central	Standard_B1ls	100%
VM_RI_08-29-2024_13-21	EA2	azure-East US	Standard_DS2_v2	100%

Discount type	Subscription column	Est. Current Utilization column
Reservations and savings plans in Azure subscriptions that you added to Workload Optimization Manager as targets	This column shows the account name for the reservation or savings plan. Click the name to set the scope to that account. Note that there can be several reservations or savings plans for an account, and each reservation or savings plan displays as its own row. **NOTE:** If there is a failure to re-validate the account for some reason, Workload Optimization Manager shows it as a *non-added* account in the Discount Inventory page.	▪ For reservations, this column shows the percentage of capacity currently used by VMs in all accounts. Workload Optimization Manager estimates the percentage if there are VMs in non-added accounts that use the reservation (since the exact number of VMs is unknown). ▪ For savings plans, a dash symbol displays to indicate that utilization is currently not tracked.
Reservations and savings plans in Azure subscriptions that you did not add to Workload Optimization Manager	This column shows a grayed-out, non-clickable name to indicate that you have not added the account as a target. Workload Optimization Manager is aware of this account and if the given reservation or savings plan is shared with other accounts because you have added an EA or MCA account.	▪ For reservations, this column shows the percentage of reservation capacity currently used by VMs in all accounts. Workload Optimization Manager estimates the percentage if there are VMs in non-added accounts that use the reservation (since the exact number of VMs is unknown). ▪ For savings plans, a dash symbol displays to indicate that utilization is currently not tracked.

Workload Optimization Manager discovers reservations scoped to a *resource group*, but treats them as shared. In the Discount Inventory chart, the scope for these reservations is shown as UNKNOWN.

Name	Subscription	Region	Product	Quantity	Scope
VM_RI_08-29...	EA1	azure-East US 2	Standard_B2ms	1	SHARED
VM_RI_08-29...	EA1	azure-East US 2	Standard_DS1_v2	1	SINGLE
VM_RI_08-29...	EA	azure-France Central	Standard_B1ls	1	UNKNOWN
VM_RI_08-29...	EA2	azure-East US	Standard_DS2_v2	1	SHARED

This could result in unreliable actions, such as scaling VMs within the resource group to other reservations, which could potentially increase costs. If you have existing reservations scoped to a resource group, be sure to change their scope in Azure before executing VM scale actions. For best results, change their scope to *shared*.

AWS savings plans

If you added targets that are AWS accounts with read-only access to the AWS Savings Plans API, Workload Optimization Manager uses this chart to present the Savings Plans that it discovered in your cloud environment (including GovCloud) and the instance types they use.

Savings Plan ID	Account	Type	Payment	Instance Family	Location	Commitment	Term	Start Date	Exp. Date
55555 ...	Prod	Compute	All Upfront	All	All	$0.001/hr	1 Year	Dec 28, 2020	Dec 28, 2021
45555 ...	Prod	EC2	No Upfront	t3	aws-...	$0.001/hr	3 Years	Dec 29, 2020	Dec 29, 2023

Google Cloud committed use discounts

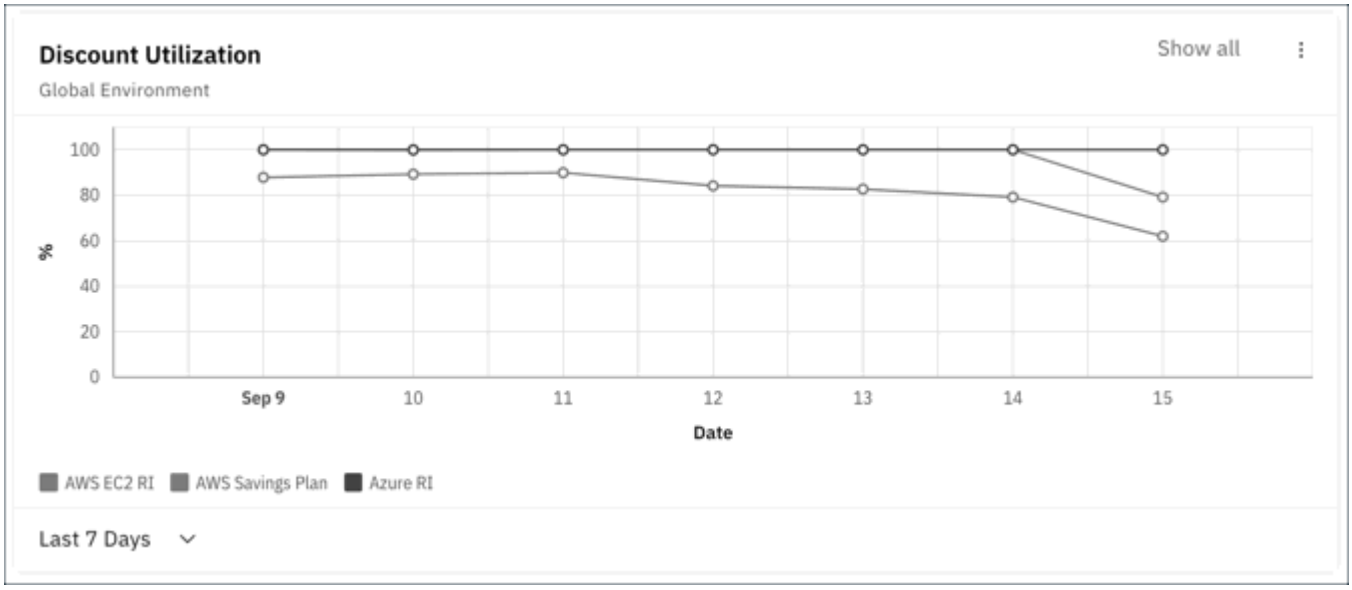
Workload Optimization Manager discovers committed use discounts (CUDs) for your workloads when you add a Google Cloud target in the Target Configuration page.

Name	Account	Status	Region	Type	Payment	Instance Family	Cores	Memory	Term	Start Date	End Date
commitment-1		Active	us-west4	Family Scoped	All upfront	N2	1	4 GB	3 Years	Oct 15, 2021	Oct 15, 2024
commitment-1		Active	europe-north1	Family Scoped	All upfront	N2	1	2 GB	1 Year	Oct 21, 2021	Oct 21, 2022
commitment-1		Active	northamerica-northea	Family Scoped	All upfront	N2D	N/A	2 GB	1 Year	Jan 27, 2022	Jan 27, 2023
commitment-2		Active	us-central1	Family Scoped	All upfront	E2	N/A	4 GB	1 Year	Oct 21, 2021	Oct 21, 2022

This chart also shows the current overall utilization of your CUD inventory for VM vCPU (cores) and VM memory, as estimated by Workload Optimization Manager analysis.

Discount utilization chart

This chart shows how well you have utilized your current discount inventory *(on page 1028)*. The desired goal is to maximize the utilization of your inventory and thus take full advantage of the discounted pricing offered by your cloud provider.



Timeframe

The chart shows daily data points and supports the following timeframes:

- Last 7 Days
- Last 30 Days

Show all

Click **Show all** to see all the discovered discounts and their current utilization.

AWS EC2/RDS reserved instances and compute savings plans

Workload Optimization Manager uses data exports from AWS to calculate Reserved Instance (RI) and Compute Savings Plan utilization for both standard and GovCloud workloads. Workload Optimization Manager supports a data export created at the management account, but not member accounts. In order for Workload Optimization Manager to use these data exports, you must add a billing target in the Target Configuration page.

The chart shows historical data. Data for the current day may not be available until the cloud provider has exported it fully.

Hover on a data point in the chart to see the following information:

- The date and time for the data point
- The percentage of utilization

  For Compute Savings Plans, utilization is based on the total utilized and committed costs per day.

  For RIs, utilization is based on normalization factors.

  Normalization factor is a measure of RI capacity that you can use to compare or combine the capacity for different instance families.

  Workload Optimization Manager measures RI coverage in terms of normalization factors. It compares the number of RIs calculated as normalization factors that cover workload capacity with the total number of normalization factors for a given Workload Optimization Manager scope. Each workload is assigned normalized units depending on its instance type.

Points to consider:

- Scope
  - You can set the scope to your global cloud environment or to individual accounts, billing families, or regions. Scoping to an account shows the RI utilization for the workloads for the entire billing family.
  - For AWS GovCloud RIs, you can set the scope to the global environment, or a GovCloud account or billing family.
  - For AWS GovCloud Savings Plans, you can set the scope to the global environment or the standard account/billing family associated with GovCloud.
- RIs
  - In AWS, you can turn off RI discount sharing for specific accounts. These accounts will not share any discounts with other accounts. Workload Optimization Manager does not recognize RI coverage or utilization for these accounts. For example, the RI Coverage and RI Utilization charts will show zero values.
  - Under very rare circumstances, you can have RIs on payment plans that do not resolve to 1-year or 3-year terms. In this case, AWS does not return pricing data for those RIs. Workload Optimization Manager does not include such RIs in its calculations of RI utilization or RI cost.

Azure reservations

Workload Optimization Manager uses billed cost data from Azure to calculate reservation utilization for workloads. In order for Workload Optimization Manager to use billed cost data, you must add an Azure Billing target in the Target Configuration page.

> **NOTE:**
> The use of billed cost data to calculate reservation utilization for Azure Government workloads is not supported.
>
> Currently, Workload Optimization Manager does not track utilization of Azure savings plans.

The chart shows historical data. Data for the current day may not be available until the cloud provider has exported it fully.

Hover on a data point in the chart to see the following information:

- The date and time for the data point
- The percentage of utilization, based on ratios.

  Ratio refers to the number of Azure reservation units that cover workload capacity compared to the total number of reservation units for a given Workload Optimization Manager scope. Each workload is assigned reservation units based on its instance type.

You can set the scope to your global cloud environment or to individual subscriptions, billing families, or regions. Scoping to a subscription shows utilization for workloads for the entire billing family or for single and shared subscriptions.

Google Cloud committed use discounts

This chart does not show *historical* utilization data for committed use discounts (CUDs) because Google Cloud does not track this data. If Workload Optimization Manager only manages Google Cloud targets, this chart will be empty.

> **NOTE:**
> Workload Optimization Manager uses billing data from Google Cloud to calculate the *current* utilization of your CUD inventory. To view current utilization data, see the Discount Inventory chart.

Cloud estimated cost charts

Cloud Estimated Cost charts show estimated monthly costs and investments for the cloud. Monthly cost amounts are summarized as amounts with and without actions.

Display

The chart shows the information as a Text chart.

Volume summary chart

> **NOTE:**
> This chart is deprecated and will be removed from the product in a future release. Delete this chart from any dashboard and use the replacement chart, Entity breakdown.

To help you manage your costs on the public cloud, these charts show the distribution and costs of volumes for the given scope. In this way, you can see how volume utilization affects your costs. For these charts, Workload Optimization Manager calculates the costs based on the cost information from the cloud targets.

These charts show the following data:

- Tiers

  The chart breaks down volumes by tier (disk type) and shows the volume count and monthly cost for each tier.
- Volume State

  The chart breaks down volumes by attachment state (attached or unattached) and shows the volume count and monthly cost for each state. For unattached volumes, you can reduce your cloud cost by the given amount if you delete these volumes. Click **Show All** and then click the **Details** button for an unattached volume to execute a delete action.

  > **NOTE:**
  > For an Optimize Cloud plan, the Volume Tier Summary chart shows 'Current' and 'Optimized' results. The 'Current' result includes currently unattached volumes that you can delete to reduce costs, while the 'Optimized' result assumes that unattached volumes have been deleted. To see a list of unattached volumes, click **Show Changes** in the chart. For details about Optimize Cloud plans, see Optimize Cloud Plan Results *(on page 893)*.

For a detailed breakdown, click **Show All** on the chart. If you have multiple cloud providers, each provider will have its own tab. Click any column heading to sort the list. The total is displayed when you choose one or more check boxes in the window.

> **NOTE:**
> For Azure environments with VMs in Scale Sets, for any VMs that are powered off, the associated volume shows a utilization of zero GB. This is an accurate presentation of the data that the Azure environment returns for such a powered-off VM. However, it is likely that some of the volume capacity is currently utilized.

Chart unit

If you are scoped to a particular cloud provider, you can sort tiers and volumes by clicking the Edit option on the chart, and then choosing one of the following units:

- **Count** – Sort by volume count, from largest to smallest.
- **Cost** – Sort by monthly cost, from highest to lowest.

On-prem chart types

These chart widgets provide information on the status of your on-prem environment.

Density charts

Density charts show the number of resource consumers (virtual machines or containers) per provider (host or storage). If available, choose the **Show Density** checkbox to see the ratio of consumers to providers.

These charts also show the desired count of virtual machines, assuming you want to fill the headroom completely. Note that the Desired Workloads values are the results of running plans. These plans can calculate workload moves within a cluster to gain more efficiency, but they always respect the cluster boundaries – the plans never move VMs to hosts on different clusters.

To display relevant data, you must set the scope to your global environment or a cluster group. Other scopes are not supported.

You can choose from the following types of charts:

- Virtual Machines vs Hosts

  This chart shows the total number of virtual machines and hosts in your on-prem environment, and tracks the data over time.

  Hover on a number in the chart to see a breakdown by resource status. For example, you can see how many virtual machines are currently active, idle, or suspended. To see the number of virtual machines per host, edit the chart and select the **Show Density** option.

- Virtual Machines vs Hosts - Density

  This chart shows the number of virtual machines per host in your on-prem environment, and tracks the data over time. A high number means that your virtual machines are densely packed, and you can take steps to alleviate pressure on the resource providers.

  If the scope of the chart is a discovered group that is based on host clusters, both historical and current density data appears. If the discovered group is not based on host clusters, only current density data appears.

  Hover on the number in the chart to see a breakdown by host status. For example, you can see how many hosts are currently active or under maintenance.

- Virtual Machines vs Storage

  This chart shows the total number of virtual machines and storage in your on-prem environment, and tracks the data over time.

  Hover on a number in the chart to see a breakdown by resource status. For example, you can see how many virtual machines are currently active, idle, or suspended. To see the number of virtual machines per host, edit the chart and select the **Show Density** option.

- Virtual Machines vs Storage - Density

  This chart shows the number of virtual machines per storage in your on-prem environment, and tracks the data over time. A high number means that your virtual machines are densely packed, and you can take steps to alleviate pressure on the resource providers.

  Hover on the number in the chart to see a breakdown by storage status. For example, you can see how many storage devices are currently active.

- Containers vs Hosts

  This chart shows the total number of containers and hosts in your on-prem environment, and tracks the data over time.

  Hover on a number in the chart to see a breakdown by resource status. For example, you can see how many containers are currently active or idle. To see the number of containers per host, edit the chart and select the **Show Density** option.

- Containers vs Hosts - Density

  This chart shows the number of containers per host in your on-prem environment, and tracks the data over time. A high number means that your containers are densely packed, and you can take steps to alleviate pressure on the resource providers.

  Hover on the number in the chart to see a breakdown by host status. For example, you can see how many hosts are currently active or under maintenance.

- Containers vs Storage

  This chart shows the total number of containers and storage in your on-prem environment, and tracks the data over time.

Hover on a number in the chart to see a breakdown by resource status. For example, you can see how many containers are currently active or idle. To see the number of containers per storage, edit the chart and select the **Show Density** option.

- Containers vs Storage - Density

  This chart shows the number of containers per storage in your on-prem environment, and tracks the data over time. A high number means that your containers are densely packed, and you can take steps to alleviate pressure on the resource providers.

  Hover on the number in the chart to see a breakdown by storage status. For example, you can see how many storage devices are currently active.

- Virtual Machines vs Hosts and Storage

  This chart shows the total number of virtual machines, hosts, and storage in your on-prem environment, and tracks the data over time. Chart information helps you understand and make decisions around capacity and utilization, based on historical and projected demand.

  This chart appears in both dashboards and plan results. If you are viewing this chart in a dashboard, you can hover on a number in the chart to see a breakdown by resource status. For example, you can see how many virtual machines are currently active, idle, or suspended. To see the number of virtual machines per host and storage, edit the chart and select the **Show Density** option.

- Virtual Machines vs Hosts and Storage - Density

  This chart shows the number of virtual machines per host and storage in your on-prem environment, and tracks the data over time. A high number means that your virtually machines are densely packed, and you can take steps to alleviate pressure on the resource providers.

  Hover on a number in the chart to see a breakdown by host or storage status. For example, you can see how many hosts are currently active or under maintenance.

- Containers vs Hosts and Storage

  This chart shows the total number of containers, hosts, and storage in your on-prem environment, and tracks the data over time. Chart information helps you understand and make decisions around capacity and utilization, based on historical and projected demand.

  Hover on a number in the chart to see a breakdown by resource status. For example, you can see how many containers are currently active or idle. To see the number of containers per host and storage, edit the chart and select the **Show Density** option.

- Containers vs Hosts and Storage - Density

  This chart shows the number of containers per host and storage in your on-prem environment, and tracks the data over time. A high number means that your containers are densely packed, and you can take steps to alleviate pressure on the resource providers.

  Hover on a number in the chart to see a breakdown by host or storage status. For example, you can see how many hosts are currently active or under maintenance.

- Headroom

  This chart shows the total number of virtual machines running in the chart's scope, plus the total headroom (capacity) for that scope. With this chart, you can see if there is extra capacity in your clusters to host workloads.

- Host and Storage Units

  This chart shows the number of active hosts and storage devices, and tracks the data over time.

- Virtual Machines and Containers vs Hosts and Storage

  This chart shows the total number of virtual machines, containers, hosts, and storage in your on-prem environment, and tracks the data over time. Chart information helps you understand and make decisions around capacity and utilization, based on historical and projected demand.

  Hover on a number in the chart to see a breakdown by resource status. For example, you can see how many virtual machines are currently active, idle, or suspended.

Chart type

You can set the display to:

- Stacked Bar Chart
- Line Chart

Timeframe

The chart supports the following timeframes:

- Default
- Last 2 hours
- Last 24 hours
- Last 3 days
- Last 7 days
- Last 30 days
- Last 60 days
- Last year
- Custom

Ports charts

Ports charts show the most utilized northbound or southbound ports in your on-prem environment over a given time period. These charts are useful in Fabric environments where you license port channels.

Display

The chart shows the information as Tabular.

Headroom charts

Headroom charts show how much extra capacity your clusters have to host workloads.

To calculate cluster capacity and headroom, Workload Optimization Manager runs nightly plans that take into account the conditions in your current environment. The plans identify the optimal workload distribution for your clusters. This can include moving your current VMs to other hosts within the given cluster, if such moves would result in a more desirable workload distribution. The result of the plan is a calculation of how many more VMs the cluster can support.

To calculate VM headroom, the plan simulates adding VMs to your cluster. The plan assumes a certain capacity for these VMs, based on a specific VM template. For this reason, the count of VMs given for the headroom is an approximation based on that VM template.

Timeframe

The chart supports the following timeframes:

- Default
- Last 2 hours
- Last 24 hours
- Last 7 days
- Last 30 days

Commodity

You can choose from the following types of Headroom charts:

- CPU Headroom
- Memory Headroom
- Storage Headroom
- Total Headroom

Exhaustion time chart

This chart shows the current growth of workloads and projects when workloads will exceed the capacity of your current infrastructure. This is useful for future planning (for example, if you might need to buy more hardware).

You can track CPU, memory, and storage as well as the average monthly Virtual Machine growth and the average VM template. The amount of time is presented as days. For example, storage will be used up in 41 days.

Creating groups

Groups assemble collections of resources for Workload Optimization Manager to monitor and manage. When setting scope for your Workload Optimization Manager session, you can select groups to focus on those specific resources. For example, if you have a number of VMs devoted to a single customer, you can create a group of just those VMs. When running a planning scenario you can set the scope to work with just that group.

Workload Optimization Manager discovers groups that exist in your environment and allows you to create custom groups. Workload Optimization Manager supports two custom-grouping methods:

- Dynamic – You define these groups by specific criteria. You can group services according to naming conventions (all VM names that start with **ny**), resource characteristics (all hosts with four CPUs), or other criteria such as time zone or number of CPUs.

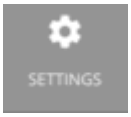
  These groups are dynamic because Workload Optimization Manager updates the group as conditions change.
- Static – You create these groups by selecting the specific group members.

NOTE:
Do not use the Workload Optimization Manager user interface to delete discovered groups. If you do, later analysis cycles will discover them again, and add them to your environment. In the meantime, any analysis that relies on those groups can give unexpected results.

You *can* delete any custom group you have created. Before you do, verify that you do not have any charts, plans, or policies that use the group you want to delete. After you delete the group, such charts, plans, or policies will lose their scope. For example, a policy with no scope has no effect.

To create a group:

1. Navigate to the Settings Page.

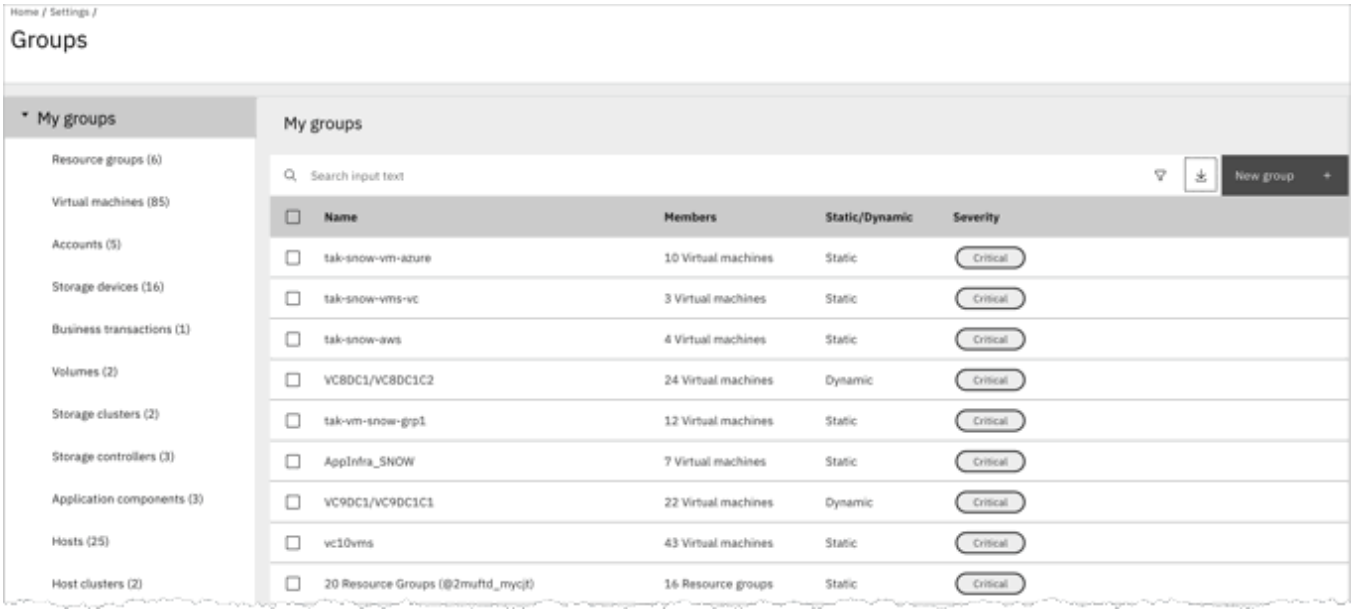
   Click to navigate to the Settings Page. From there, you can perform a variety of Workload Optimization Manager configuration tasks.
2. Choose **Groups**.



Click to navigate to the Group Management Page.

This page lists all the custom groups that you currently have configured for Workload Optimization Manager.

3. Perform tasks in the Group Management page.



You can perform any of the following tasks:

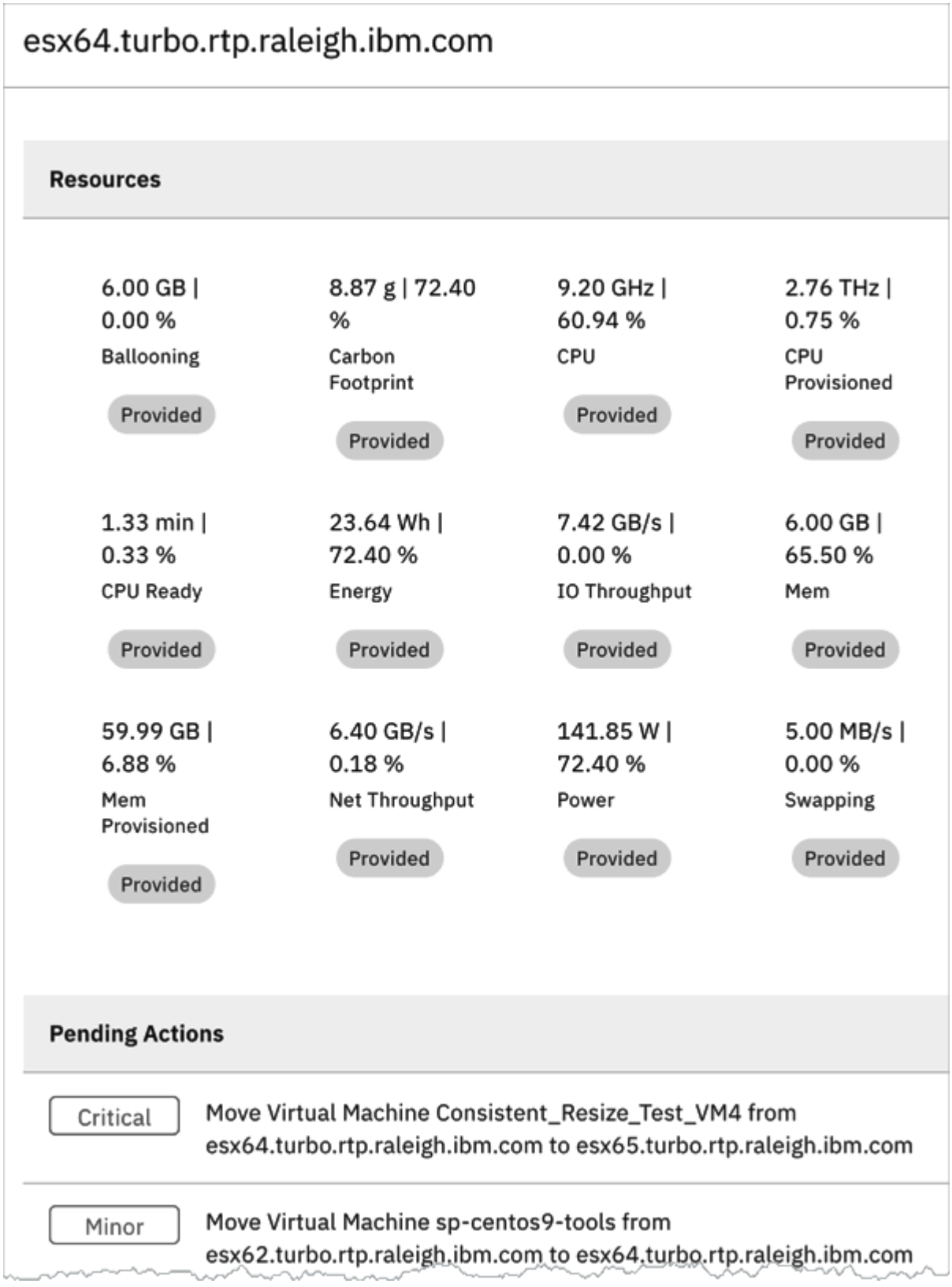
- Sort by name or severity.
- Search for groups.
- Edit a group by clicking the pencil icon.

  For a dynamic group, edit the criteria to select group members. For a static group, you can add or remove specific members.
- Filter by name using regular expressions.
- Download a list of groups.
- Expand an entry to see group details.

  Click the **Show more details** button  to see group details.
- Select an entry to delete the groups
- Create new groups.

To work with a long list of groups, you can filter by group type. For example, only show groups of VMs, or groups of host machines. You can also type a string in the **Search** field to filter the list.

4. Expand an entry to see group details.



The details show you information about related entities such as how many hosts provide resources for a group of VMs. If there are any pending actions for the group, the details list those actions as well.

5. Create a new group.

Click **NEW GROUP**, choose a group type, and then specify the group settings.



- Give the group a unique name. To prevent issues, you should never use duplicate names for groups of the same entity type.
- Set whether the group will be static or dynamic.

  To create a static group, select the member entities from the list. To filter the list, set group criteria.

  To create a dynamic group, set group criteria. The list updates to show the resulting group members.
- When you are finished, save the group.

Working with policies

Policies set business rules for resource allocation analysis, resource status display, and action execution.

Workload Optimization Manager includes the following fundamental types of policies:

- Placement policy

  A placement policy is a set of rules that Workload Optimization Manager must satisfy when generating placement recommendations for on-prem and containerized workloads.

  For details, see Placement Policies *(on page 1043)*.
- Automation policy

  An automation policy is a set of rules that Workload Optimization Manager must satisfy when executing non-parking actions *(on page 835)* on public cloud or on-prem workloads, or changing settings that affect analysis and action generation.

  For details, see Automation Policies *(on page 1049)*.
- Parking policy

A parking policy is a set of rules that Workload Optimization Manager must satisfy when enforcing parking actions (stop and start) on public cloud entities.

For details, see Parking Policies *(on page 1087)*.

Policy Management

Use the Policy Management page to view and manage policies. The page shows the following categories of policies:

- All Policies – All the currently defined policies
- Placement Policies
  - Imported Placement Policies – Placement policies discovered *(on page 1044)* from your targets
  - Workload Optimization Manager Segments – Placement policies created *(on page 1044)* from the Policy Management page
- Automation Policies
  - Imported Automation Policies – Automation policies discovered *(on page 1053)* from your targets
  - User Defined Automation Policies – Automation policies created *(on page 1053)* from the Policy Management page
  - Defaults – Default *(on page 1050)* automation policies for the different entity types

    NOTE:
    To see the placement or automation policies that are applied by a particular entity, set the scope to that entity (from Search or the supply chain), and then click the Policies tab. For more information, see Scope Policies *(on page 167)*.

- Parking Policies

    NOTE:
    All parking policies are manually created from the Policy Management page. There are no default or imported parking policies.

  - Global
  - Cloud Provider
  - Account
  - Resource Groups

Placement policies

A placement policy is a set of rules that Workload Optimization Manager must satisfy when generating placement recommendations for on-prem and containerized workloads.

With these policies, Workload Optimization Manager can recommend placement actions that comply with your business rules. For example, a placement policy can constrain VMs to specific hosts. If a VM needs to move to a different host due to resource congestion, Workload Optimization Manager will generate an action to move the VM to one of the hosts defined in the policy.

Workload Optimization Manager discovers placement policies from your targets, and allows you to create your own policies.

NOTE:
You can enable or disable any placement policy to affect placement calculations in the real-time environment or in plans.

When calculating workload placement, Workload Optimization Manager respects cluster boundaries, networks, and provisioned data stores. In addition, the configuration of your environment can specify logical boundaries, and you can create even

more boundaries within Workload Optimization Manager. These boundaries impose segments on the market that Workload Optimization Manager uses to model your application infrastructure.

In finance, a market segment divides the market according to the criteria different groups of people use when they buy or sell goods and services. Likewise in the Workload Optimization Manager market, a workload placement segment uses criteria to focus the buying and selling of resources within specific groups of entities. This gives you finer control over how Workload Optimization Manager calculates placements.

Imported placement policies

Your on-prem targets can include placement policies of their own. Workload Optimization Manager imports these placement policies, and considers them to be constraints on placement. You cannot disable these imported policies for real-time analysis, but you can disable them for plans.

Workload Optimization Manager imports the following:

- vCenter Server DRS Rules

  See Other Information Imported from vCenter *(on page 353)*
- Virtual Machine Manager Availability Sets

  See Virtual Machine Manager *(on page 368)*
- Flexera One License Specifications

  See Flexera *(on page 354)*

To view imported placement policies, navigate to **Settings > Policies**, and then click **Imported Placement Policies**.

Creating placement policies

A placement policy is a set of rules that Workload Optimization Manager must satisfy when generating placement recommendations for on-prem and containerized workloads.

With these policies, Workload Optimization Manager can recommend placement actions that comply with your business rules. For example, a placement policy can constrain VMs to specific hosts. If a VM needs to move to a different host due to resource congestion, Workload Optimization Manager will generate an action to move the VM to one of the hosts defined in the policy.

Workload Optimization Manager discovers placement policies from your targets, and allows you to create your own policies.

You can create the following placement policies:

- **Place** - Determine which entities use specific providers.

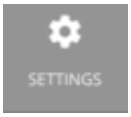
  For more information on Place policies, click here *(on page 1046)*.
- **Don't Place** - Consumers must never run on specific providers.

  For more information on Don't Place policies, click here *(on page 1047)*.
- **Merge** - Merge clusters into a single provider group.

  For more information on Merge policies, click here *(on page 1047)*.
- **License** - Set up hosts to provide licenses for VMs.

  For more information on License policies, click here *(on page 1048)*.

1. Open the Settings Page.

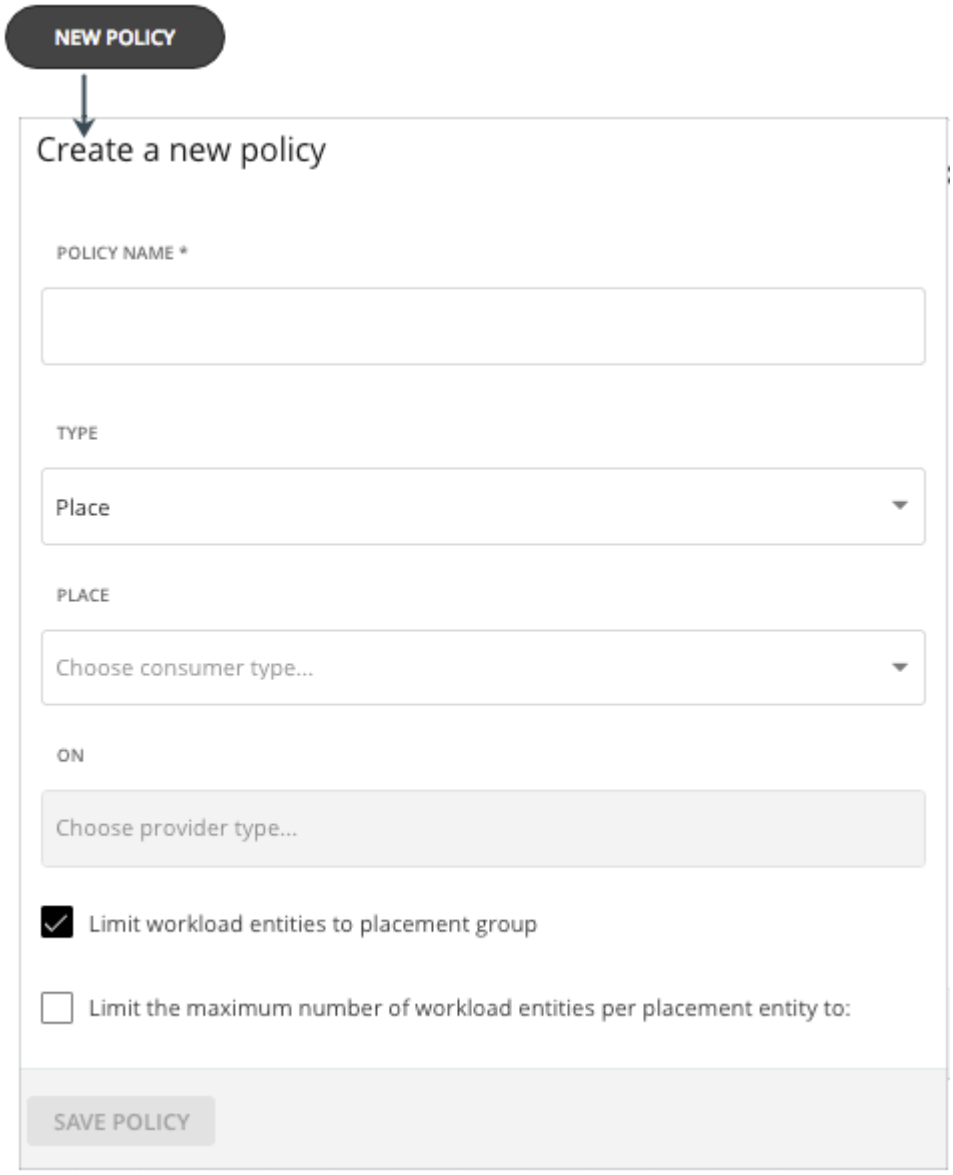
   Click to open the Settings Page. From there, you can perform various Workload Optimization Manager configuration tasks.
2. Choose Policies.



Click to open the Policy Management Page.

This page lists all the policies that you currently configured for Workload Optimization Manager.

3. Create a new policy.



First, select the type of Placement policy to create, then specify the settings:

- Give the policy a name.
- Choose the policy type.
- Specify the settings.

4. Save the policy.

Placement policy types

A placement policy is a set of rules that Workload Optimization Manager must satisfy when generating placement recommendations for on-prem and containerized workloads.

You can create the following placement policies:

- **Place**: Determine which entities use specific providers.

  For example, the VMs in a consumer group can only run on a host that is in the provider group. You can limit the number of consumers that can run on a single provider – for hosts in the provider group, only 2 instances of VMs in the consumer group can run on the same host. Or no more than the specified number of VMs can use the same storage device.
- **Don't place**: Consumers must never run on specific providers.

  For example, the VMs in a consumer group can never run on a host that is in the provider group. You can use such a segment to reserve specialized hardware for certain workloads.
- **Merge**: Merge clusters into a single provider group.

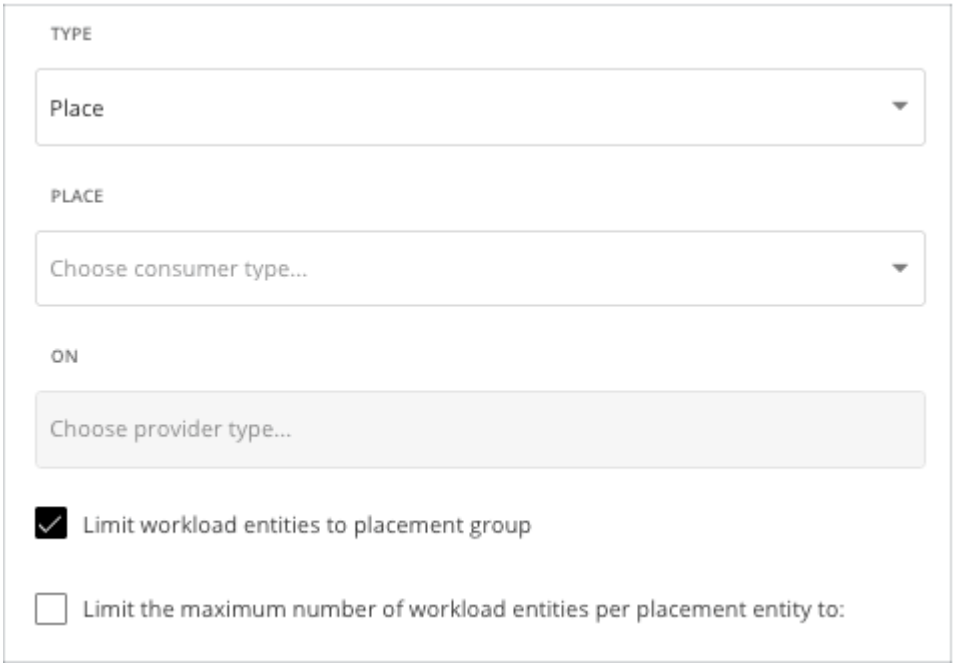
  For example, you can merge three host clusters in a single provider group. This enables Workload Optimization Manager to move workload from a host in one of the clusters to a host in any of the merged clusters to increase efficiency in your environment.
- **License**: Set up hosts to provide licenses for VMs.

  For VMs that require paid licenses, you can create placement policies that set up certain hosts to be the VMs' preferred license providers. Workload Optimization Manager can then recommend consolidating VMs or reconfiguring hosts in response to changing demand for licenses.

Place policy

These policies control where the workload can be placed. For example, you can specify that a VM will be placed only on a host that is a member of a specific cluster. Alternatively, you might specify that any applications in a specific group can be placed only on a data store that is a member of a specific group.

1. Select **Settings > Policies**, then select **New Policy**.
2. Select **Place** from the **Type** drop-down.



3. Specify the consumer type.

The consumer type is a group or cluster of entities that will be placed on the identified providers.

4. Specify the provider type.

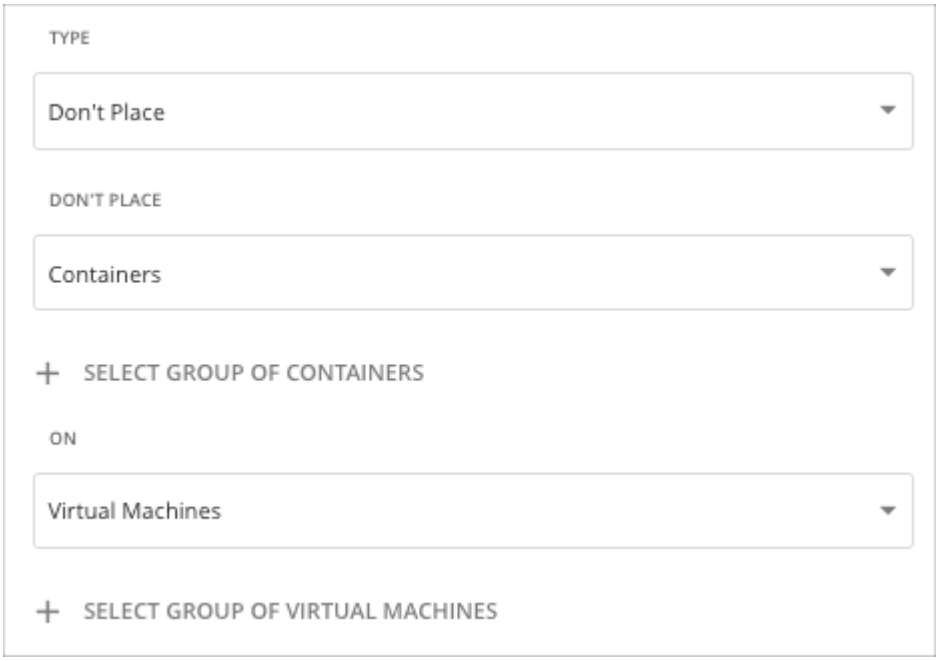
   The provider type is a group or cluster of entities that provide resources to the consumers.

5. The Limit workload entities to placement group is selected by default to place consumer entities only on members of the provider group.
6. Specify whether to limit how many instances of the consumer entities can be placed on a single provider.
7. Save the policy.

Don't place policy

These policies identify groups or clusters that will never host the consumer entities. For example, you can specify that a VM will never be placed on a host that is a member of a specific cluster. Alternatively, you can specify that a set of noncritical applications will never be placed on specialized hardware, as a way to ensure availability for critical applications.

1. Select **Settings > Policies**, then select **New Policy**.
2. Select **Don't Place** from the **Type** drop-down.



3. Specify the consumer type.

   The consumer type is a group or cluster of entities that will be placed on the identified providers.

4. Specify the provider type.

   The provider type is a group or cluster of entities that provide resources to the consumers.

5. Save the policy.

Merge policy

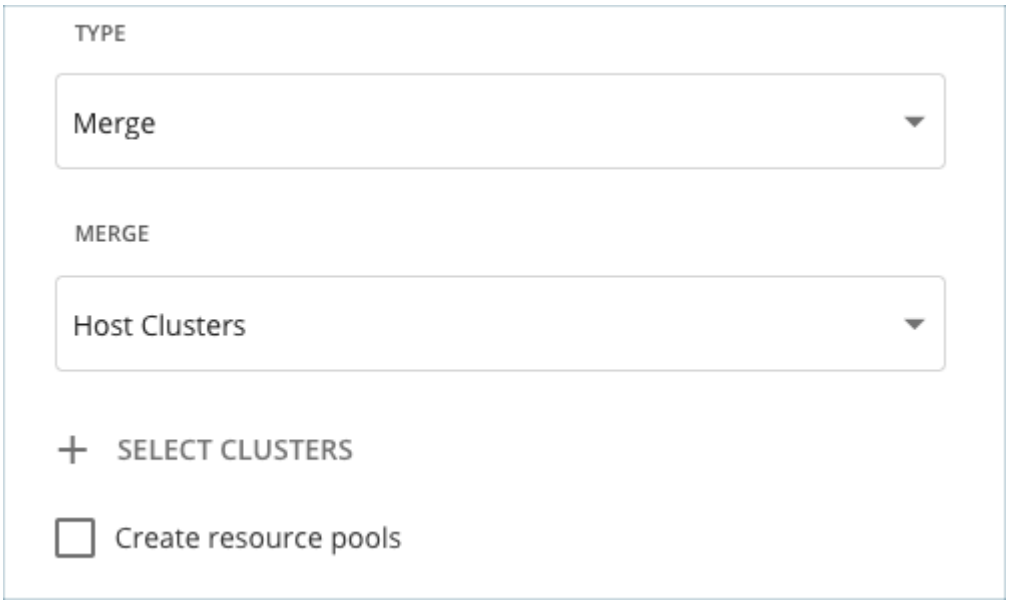
You can create placement policies that merge multiple clusters into a single logical group for workload placement.

For example, your environment might divide hosts into clusters according to hardware vendor, or by some other criteria. Workload placement typically does not cross such cluster boundaries. However, there might be no technical reason to apply these boundaries to workload placement. By creating a larger pool of provider resources, Workload Optimization Manager has even more opportunities to increase efficiency in your environment.

Combining merge and placement policies can provide significant operational advantages during hardware refreshes or data center migrations.

WARNING:
If you have reservations defined, network merge policies do not currently respect these reservations.

1. Select **Settings > Policies**, then select **New Policy**.
2. Select **Merge** from the **Type** drop-down.



3. Choose the type of entity to merge.
4. Select the groups to merge.
5. Create resource pools.

   Select this option to enable the creation of resource pools in the destination cluster during cross-cluster VM move actions. For any resource pool in the source cluster, Workload Optimization Manager creates a resource pool with the same name and default quotas in the destination cluster if one does not already exist. This feature is applicable only in vCenter environments.
6. Save the policy.

For merge policies, keep the following considerations in mind:

- For most policies that merge host and storage clusters, the clusters you place in the Merge segment must be members of the same data center.
- For vCenter environments, you can create placement policies that merge data centers to support cross-vCenter moves. In this case, where a data center corresponds to a vCenter target, the merged clusters can be in different data centers. In this case, you must create two merge policies; one to merge the affected data centers, and another to merge the specific clusters.

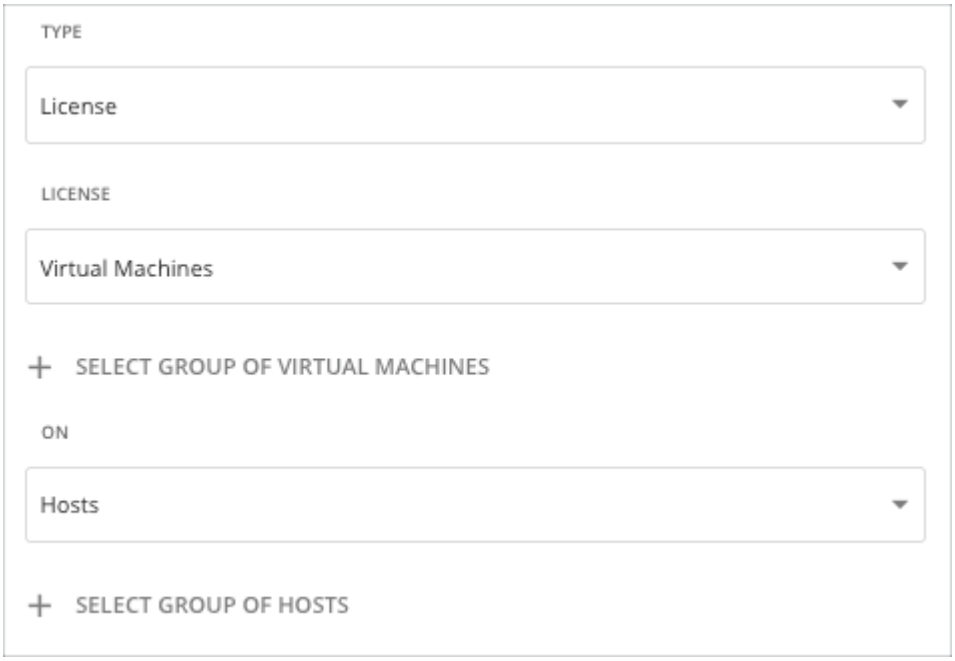
  If the clusters you merge do not use the same network names on their respective data center, configure a network merge policy to define the compatible networks.
- For network merges, if you select to merge "Network A" and "Network B," all networks across data centers with name "Network A" and "Network B" are merged.

License policy

Assume that you have purchased several licenses for a database – you pay for the right to run that database on some hosts. You can create a license policy to identify the hosts that provide the license, and the VMs that can consume that license.

In addition to creating a license policy, you must also create host automation policies to allow Workload Optimization Manager to recommend reconfigure actions on hosts. In the automation policies, add the license-providing hosts and then enable the *Reconfigure* action.

1. Select **Settings > Policies**, then select **New Policy**.
2. Select **License** from the **Type** drop-down.



3. Specify the license consumers (VMs).
4. Specify the license providers (Hosts).
5. Save the policy.

After you create the policy, Workload Optimization Manager can recommend the following actions in response to changing demand for licenses:

- When demand is low, Workload Optimization Manager recommends consolidating VMs on as few license-providing hosts as possible to reduce your license costs. To consolidate, you move VMs to another host and then reconfigure the original hosts to remove their licenses. Note that Workload Optimization Manager will *not* recommend suspending these hosts. Since they remain active, they can be reconfigured to become providers when demand starts to exceed capacity.

  For example, if you have `Host_01` providing a license to `VM_01` and `Host_02` providing a license to `VM_02`, you see two recommendations – move `VM_02` to `Host_01` and then remove the license in `Host_02`. You will not see a recommendation to suspend `Host_02`.
- When demand exceeds capacity, and there are hosts in the policy that currently do not provide licenses, Workload Optimization Manager recommends reconfiguring those hosts to become providers and then moving VMs to those hosts. If all hosts are providing licenses, Workload Optimization Manager recommends adding licenses to the hosts to meet demand.

  These actions are more efficient than provisioning new hosts.

Automation policies

An automation policy is a set of rules that Workload Optimization Manager must satisfy when executing non-parking actions *(on page 835)* on public cloud or on-prem workloads, or changing settings that affect analysis and action generation.

Automation policies include the following settings:

- Scope

  The scope determines which entities this policy affects.

- Policy Schedule

  This setting enables the overall policy at the scheduled time.

- Automation Workflow

  These settings specify whether Workload Optimization Manager or a workflow orchestrator will execute the action. For more information, see Automation Workflow *(on page 1063)*.

- Constraints and other settings

  These settings affect the Workload Optimization Manager analysis of the state of your environment. These include operational, utilization, and scaling constraints.

  The settings you can make are different according to the type of entity this policy will affect. Each setting you add to the policy takes precedence over the default value for that setting.

Default and user-defined automation policies

Workload Optimization Manager ships with default automation policies that are believed to give you the best results based on our analysis. For certain entities in your environment, you can create automation policies as a way to override the defaults.

For example, **Enforce Non Disruptive Mode** is turned off in the default automation policy for on-prem VMs. In most cases, you might want to turn on the setting, and only turn it off for select VMs. In that case, you would turn it on in the default automation policy for VMs, and then create policies for those groups of VMs for which you want to turn it off.

The default and user-defined automation policies take effect in relation to each other. A default policy has a global effect, while a user-defined policy overrides the default policy for the entities within the indicated scope. You should keep the following points in mind:

- User-defined policies override a subset of settings.

  A user-defined policy can override a subset of settings for the entity type. For the remainder, Workload Optimization Manager will use the default policy settings on the indicated scope.

- When an entity applies conflicting user-defined policies, Workload Optimization Manager applies the following tie breakers:
  - A scheduled policy always takes precedence over a non-scheduled policy, even if the non-scheduled policy is more conservative.
  - Among scheduled policies with *identical* schedules, the most conservative setting wins.
  - Among non-scheduled policies, the most conservative setting wins.

  For example, a VM currently belongs to four groups with different policy settings.
  - Group A policy: Resize VM in *Manual* mode every Saturday.
  - Group B policy: Resize VM in *Automatic* mode every Saturday.
  - Group C policy: Resize VM in *Manual* mode (no schedule).
  - Group D policy: Resize VM in *Recommend* mode (no schedule).

  Results:
  - On a Saturday, Groups A and B policies take precedence over Groups C and D policies. The VM ultimately applies the Group A setting because it is more conservative.
  - On all the other days, only Groups C and D policies are active. The VM applies the Group D setting because it is more conservative.

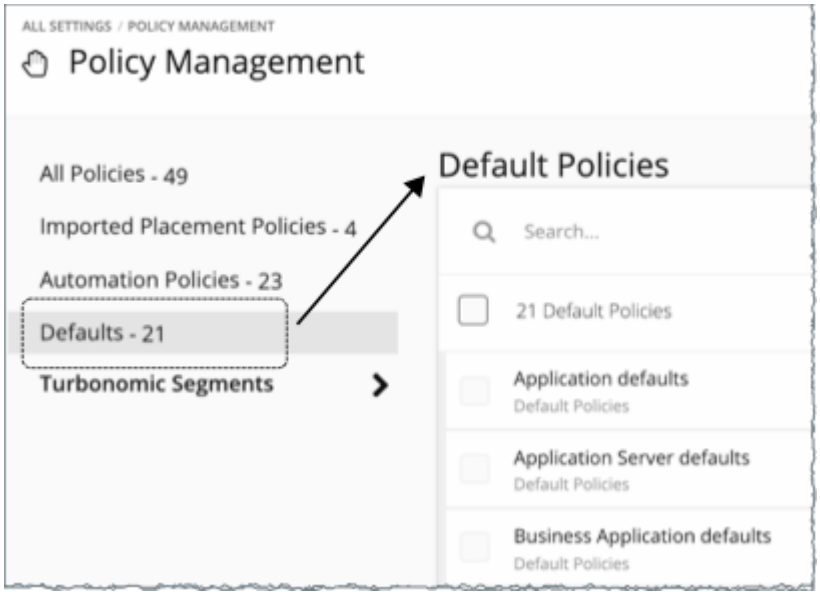
- User-defined policies always take precedence over default policies.

  Even if the default policy has a more conservative setting, the setting in the user-defined policy wins for entities in that scope.

- For a global effect, *always* use default policies.

Default automation policies

To view the default automation policies, navigate to **Settings > Policies**, and then click **Defaults**.



The page displays a list of all the default policies, by entity type. You can click the entity type to view or change the default settings.

Global default policy

Use these settings to modify Workload Optimization Manager analysis globally for any scope of your environment. These defaults affect both default and user-defined automation policies.

Disable all actions

Attribute	Default setting
Disable All Actions	Off

When this is ON, Workload Optimization Manager dos not generate any actions for your environment. For example, assume you have configured a number of polices that automate actions, but you want to stop making changes to the entire environment for a period of time. Turn this ON to stop all execution with a single setting.

VM growth observation period

Attribute	Default value
VM Growth Observation Period	1 month

Use this setting to specify how much historical data the Workload Optimization Manager analysis will use to calculate time to exhaustion of your cluster resources.

Workload Optimization Manager runs nightly plans to calculate headroom for the clusters in your on-prem environment. To review your cluster headroom in dashboards, set the view scope to a cluster. With that scope, the view includes charts to show headroom for that cluster, as well as time to exhaustion of the cluster resources.

To calculate cluster growth trends, analysis uses historical data for the given clusters. With **VM Growth Observation Period**, you can specify how much historical data the headroom analysis will use to calculate time to exhaustion of your cluster resources. For example, if cluster usage is growing slowly, then you can set the observation to a period that is long enough to capture that rate of growth.

If the historical database does not include at least two entries in the monthly data for the cluster, then analysis uses daily historical data.

Allow unlimited host provisioning

Attribute	Default setting
Allow Unlimited Host Provisioning	Off

By default, Workload Optimization Manager allows overprovisioning hosts up to 10 times their memory capacity, and up to 30 times their CPU capacity. When this setting is ON, Workload Optimization Manager removes these overprovisioning limits to allow VM placements on already overprovisioned hosts.

This setting does not stop Workload Optimization Manager from recommending actions to provision new hosts in clusters.

Enable analysis of on-prem volumes

Attribute	Default setting
Enable Analysis of On-prem Volumes	On

On-prem volumes *(on page 772)* represent VM Disks discovered by hypervisor targets. A VM will have one volume for each configured disk and another volume (representing the configuration) that always moves with Disk 1.

- On (default)

  Workload Optimization Manager analyzes resources on each volume independently. In the real-time market and on-prem plans, any action to move VM storage migrates volumes to the most optimal datastore. A Migrate to Cloud plan will recommend storage for each volume.

  For example, assume a VM with three disks. Disks 1 and 3 are on Datastore A, while Disk 2 is on Datastore B.

  - During a storage migration, VM Disk volumes 1, 2, and 3 can migrate to different datastores.
  - A Migrate to Cloud plan will recommend three separate storage disks for VM Disk volumes 1, 2, and 3.

  NOTE:
  When this setting is enabled, your Workload Optimization Manager instance uses more memory and storage to perform its analysis. If you experience performance issues when this setting is enabled, review your VM automation policies *(on page 746)* and verify that Storage Move actions are in *Recommend* or *Manual* mode. In addition, review your storage placement policies *(on page 1043)* to ensure that individual VM volumes can be placed on the expected storage.

- Off

  Workload Optimization Manager analyzes volume resources as part of VM analysis. In the real-time market and on-prem plans, any action to move VM storage ensures that volumes stay together on the underlying datastore. A Migrate to Cloud plan *(on page 903)* will recommend storage per datastore to hold all the VM Disks currently on the datastore.

  For example, assume a VM with three disks. Disks 1 and 3 are on Datastore A, while Disk 2 is on Datastore B.

  - During a storage migration, VM Disk volumes 1 and 3 will stay on the same datastore.
  - A Migrate to Cloud plan will recommend a storage disk for VM Disk volumes 1 and 3, and another storage disk for VM Disk volume 2.

Imported automation policies

As Workload Optimization Manager discovers your environment, it can find configurations that set up scopes that need specific policies. For example:

- HA Configurations

  For vCenter Server environments, Workload Optimization Manager discovers HA cluster settings and translates them into CPU and memory utilization constraints. The discovery creates a group of type *folder* for each HA cluster, and creates a policy that sets the appropriate CPU and memory constraints to that policy.

- Availability Sets

  In public cloud environments, Workload Optimization Manager discovers groups of VMs that should keep all their VMs on the same template. In the Automation Policies list, these appear with the prefix `AvailabilitySet::` on the policy names. You can enable Consistent Resizing for the VMs in each group so Workload Optimization Manager can resize them to the same size.

To view imported automation policies, navigate to **Settings > Policies**, and then click **Imported Automation Policies**.

Creating automation policies

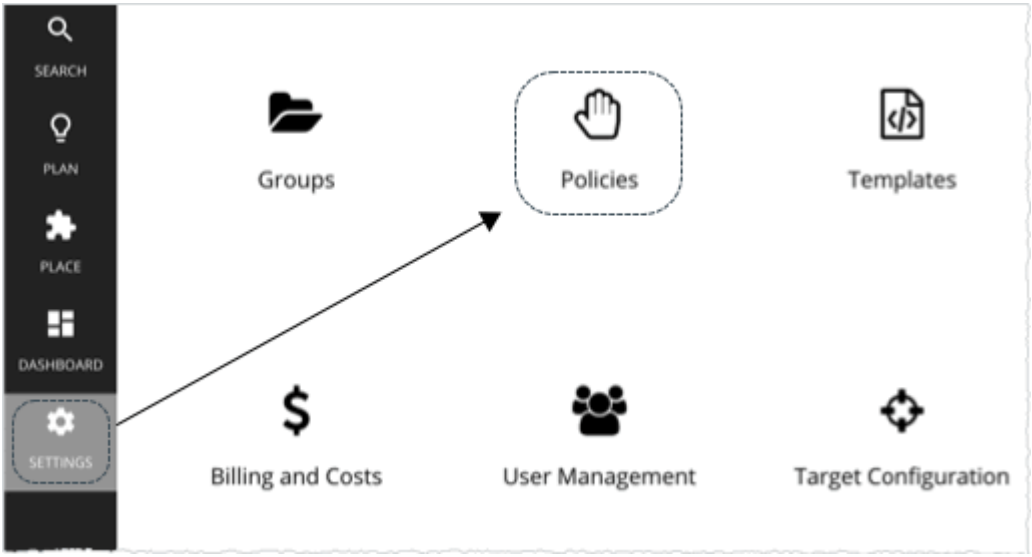
To override the default automation policies, you can create your own policies. These policies specify the settings that you want to change for certain entities in your environment. You can assign a schedule to your policy to set up maintenance windows or other scheduled actions in your environment.

The following list notes use cases for creating automation policies.

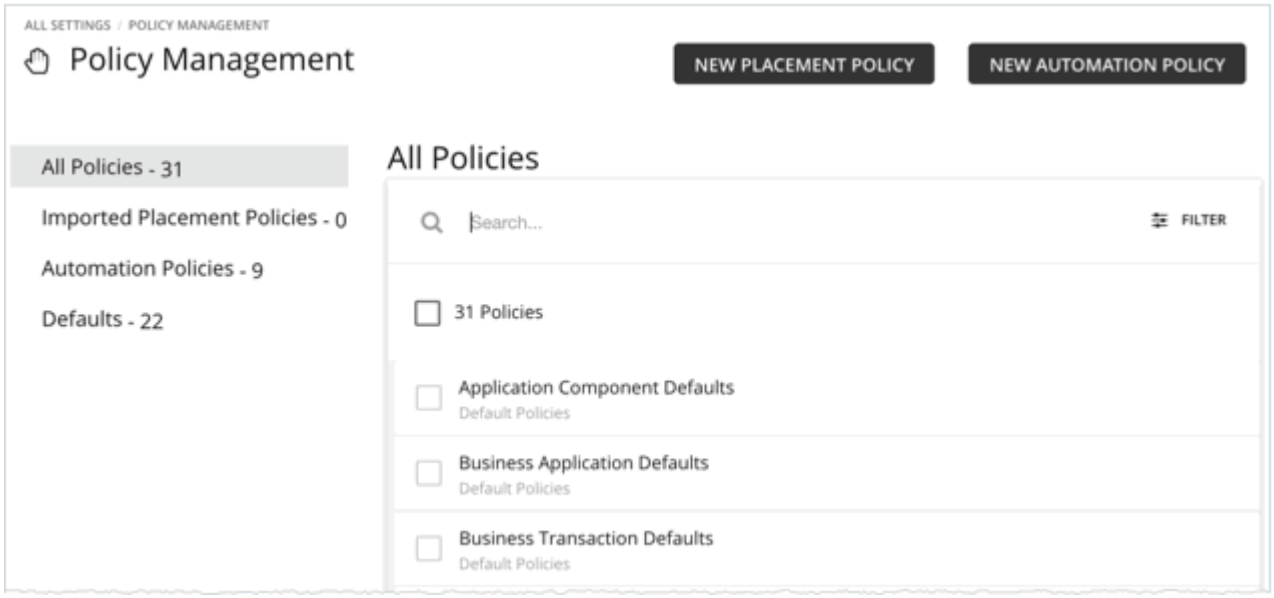
- Workload Optimization Manager uses several settings to guide its analysis of the entities in your environment. The default settings might be fine in most cases, but you might want different analysis for some groups of entities.
- Assume that you want to automate scaling and placement actions for the VMs in your environment. It is common to take a cautious approach, and start by automating clusters that are not critical or in production. You can scope the policy to those clusters, and set the action acceptance mode to Automatic for different actions on those VMs (see Action Acceptance Modes *(on page 848)*).
- If actions require change approval, or integration into DevOps pipelines to complete tasks before, instead of, or after action execution, you should scope those entities to a group and apply an automation policy.

Entry point

Navigate to the Settings Page and then choose **Policies**.



The Policy Management page lists all the currently available policies.

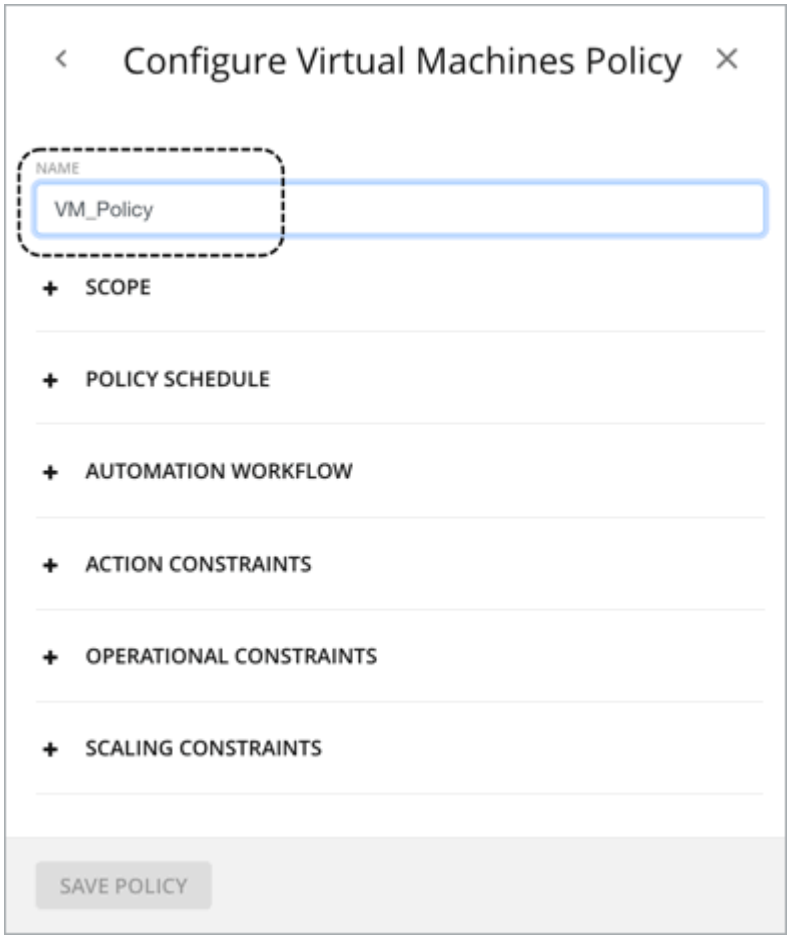


Click **New Policy > Automation Policy**and then select the entity type (such as Virtual Machine).

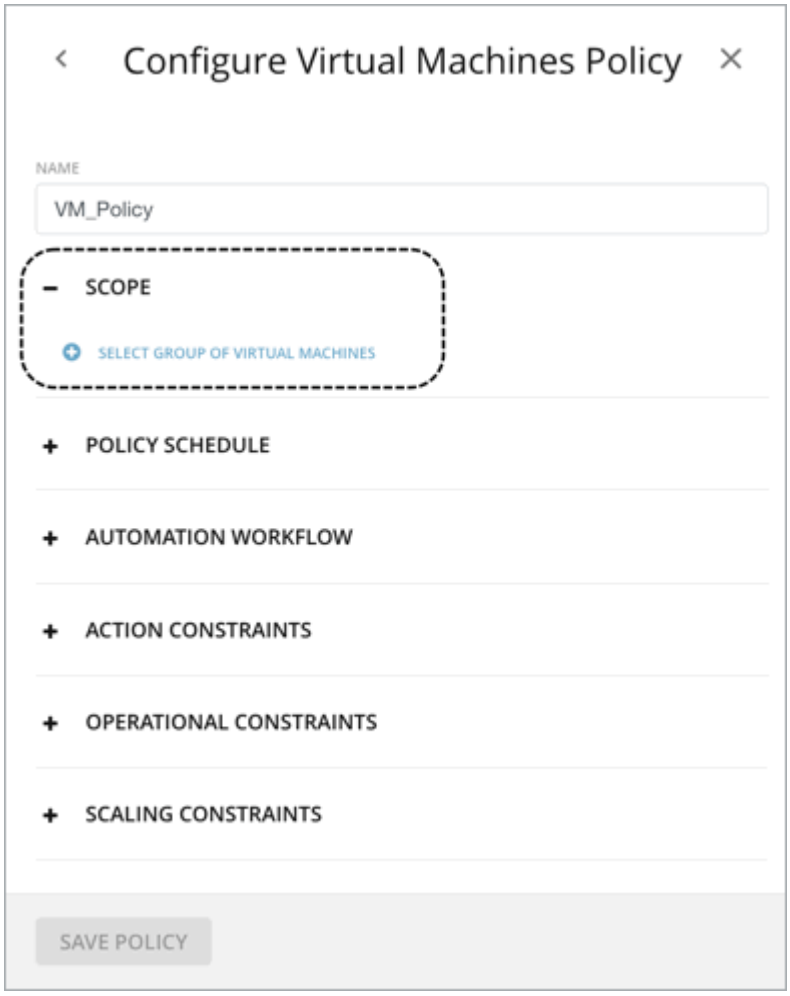
This sets the type of entity that your policy affects. Note that Workload Optimization Manager supports different actions for different types of entities. For example, you cannot add VMem to a storage device. Setting a policy type is the first step that you take to focus on which actions you want to map to your workflows.

Policy name

Name the policy.


Configure Virtual Machines Policy
NAME
VM_Policy
SCOPE
POLICY SCHEDULE
AUTOMATION WORKFLOW
ACTION CONSTRAINTS
OPERATIONAL CONSTRAINTS
SCALING CONSTRAINTS
SAVE POLICY

Scope

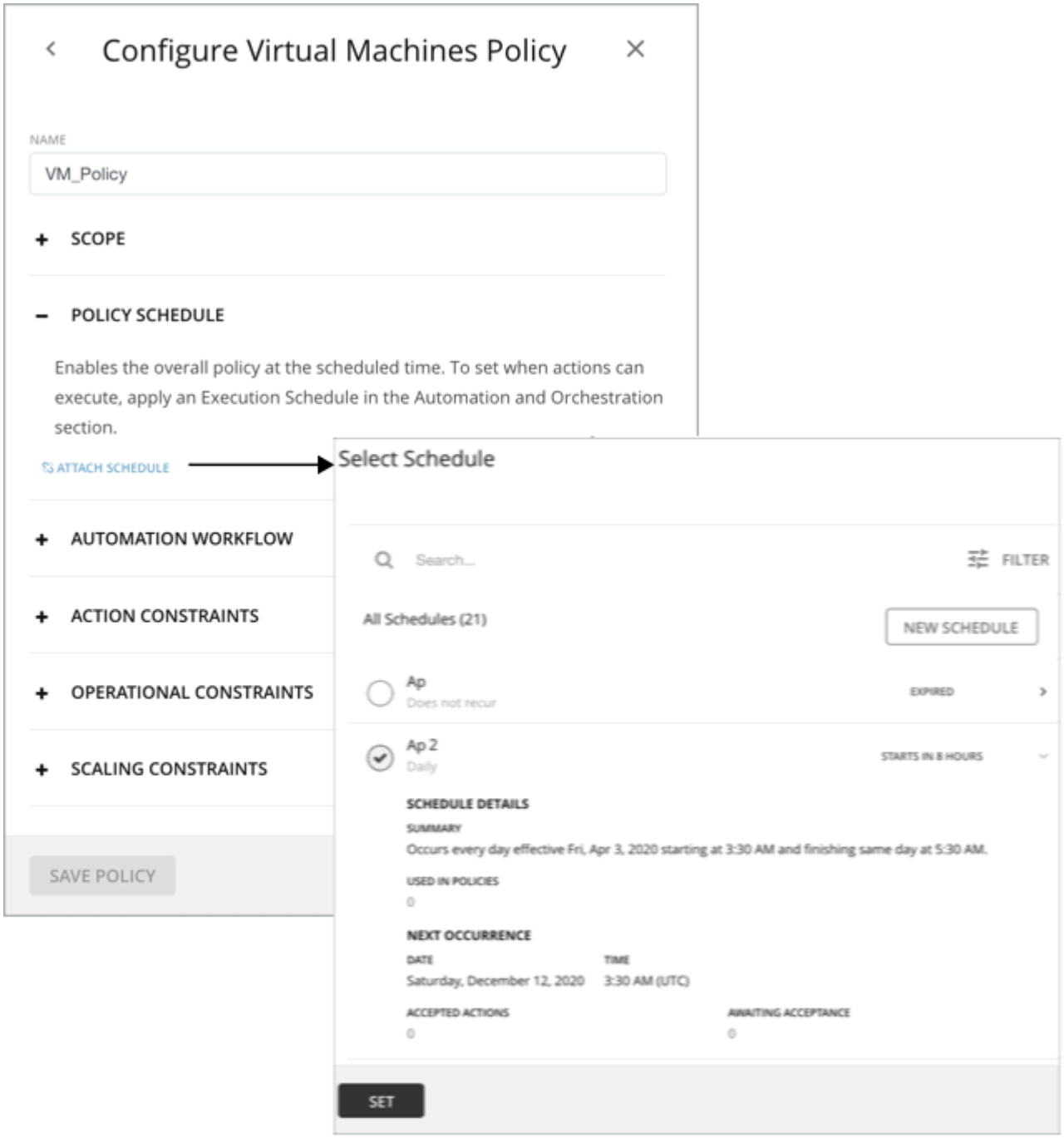


The scope determines which entities this policy affects. Choose one or more groups, or create new groups and add them to the policy scope. These groups match the type of entity that you set for the policy.

In Workload Optimization Manager you can find nested groups (groups of groups). For example, the "By PM Cluster" group contains host clusters, and each host cluster is a group. Do not set the policy scope to a parent of nested groups. When setting up policies, be sure that you set them to individual groups. If necessary, create a custom group for the settings you want to apply.

> **NOTE:**
> A single entity can be a member of multiple groups. This can result in a conflict of settings, where the same entity can have different policy settings. For conflicts among user-defined policy settings, the most conservative setting will take effect. For details, see Default and User-defined Policies *(on page 1050)*.

Policy schedule



For use cases and information about how schedules affect policies, see Policy Schedules *(on page 1061)*.

The **Select Schedule** fly-out lists all the schedules that are currently defined for your instance of Workload Optimization Manager.

Expand a schedule entry to see its details. The details include a summary of the schedule definition, as well as:

- Used in policies

  The number of policies that use this schedule. Click the number to review the policies.

- Next occurrence

  When the schedule will next come into effect.

- Accepted actions

  How many scheduled actions have been accepted to be executed in the next schedule occurrence. Click the number for a list of these actions.

- Awaiting acceptance

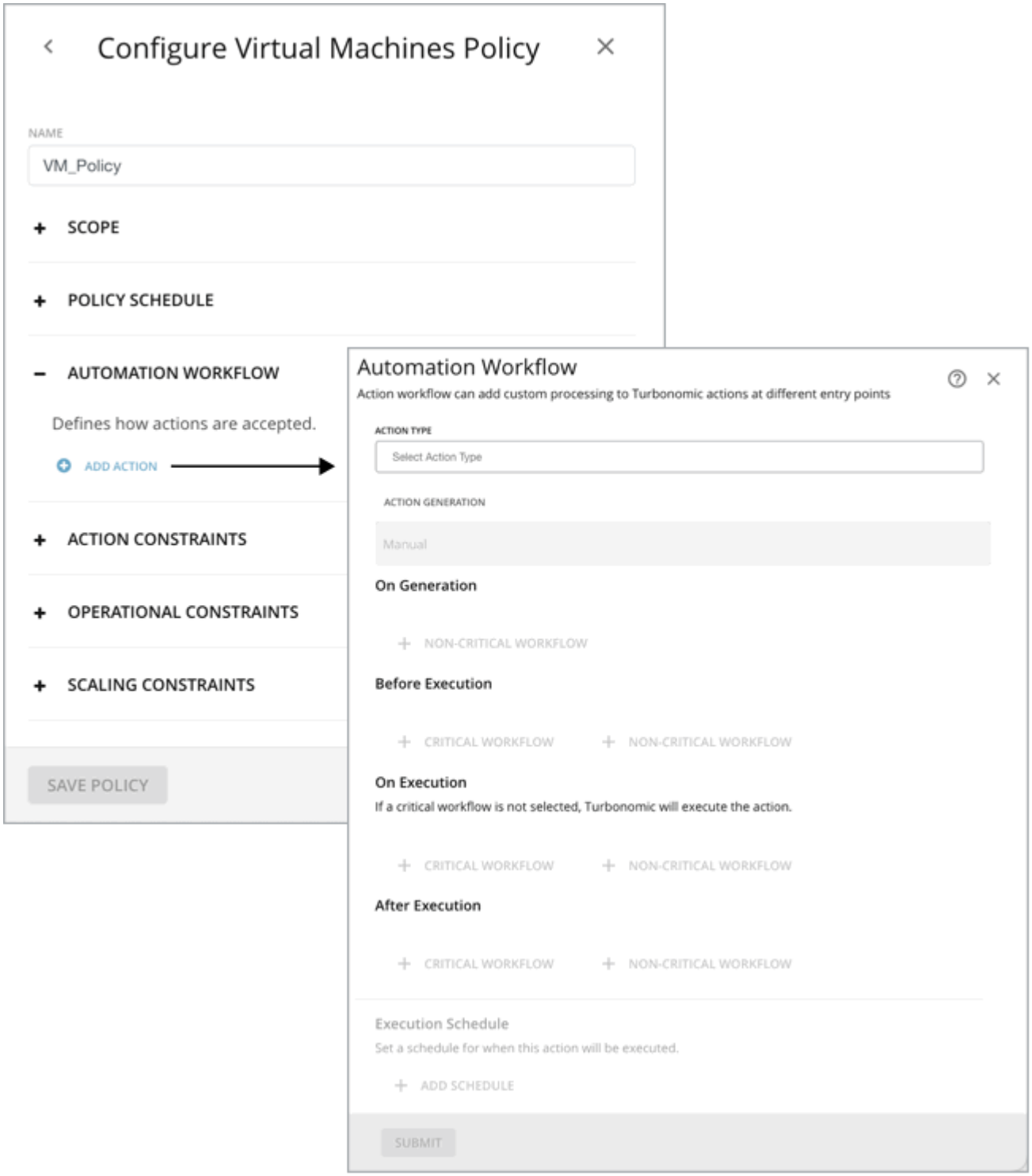
  The number of Manual actions affected by this schedule that are in the Pending Actions list, and have not been accepted. Click the number for a list of these actions.

If none of the listed schedules is suitable for your policy (or if none exists), click **New Schedule**. For details, see Managing Calendar Schedules *(on page 1092)*.

> **NOTE:**
> When you configure a schedule window for a VM resize action, to ensure Workload Optimization Manager will execute the action during the scheduled time, you must turn off the **Enforce Non Disruptive Mode** setting for that scheduled policy. Even if you turn the setting off for the global policy, you still must turn the setting off for your scheduled policy. Otherwise Workload Optimization Manager will not execute the resize action.

Automation workflow

You can define automation workflow settings for different action types within the same policy. For example, for a group of VMs in a policy, you can automate all resize actions, but require suspend actions to go through an approval process through an orchestrator.



- Action type

  See a list of actions that are viable for the policy, and then make your selections.
- Action generation
  - Do not generate actions

    Workload Optimization Manager never considers your selected actions in its calculations. For example, if you do not want to generate Resize actions for VMs in the policy, analysis will still drive toward the desired state, but will do so without considering resizes.

- Generate actions

  Workload Optimization Manager generates your selected actions to address or prevent problems. Choose how you would like the actions to run on generation, before execution, on execution, or after execution.

  - Recommend

    Recommend the action so that a user can execute it outside Workload Optimization Manager.

  - Manual

    Recommend the action, and provide the option to execute that action through the Workload Optimization Manager user interface.

  - Automated

    Execute the action automatically.

    For automated resize or move actions on the same entity, Workload Optimization Manager waits five minutes between each action to avoid failures associated with trying to execute all actions at once. Any action awaiting execution stays in queue. For example, if a VM has both vCPU and vMem resize actions, Workload Optimization Manager could resize vCPU first. After this resize completes, it waits five minutes before resizing vMem.

  - Automated when approved

    Execute the action automatically only after they are externally approved. This option is available only after adding a Service Now target.

  If you have an orchestrator target (such as ServiceNow), and that target includes an installation of the *Workload Optimization Manager Actions* application, you can send the action to the orchestrator.

  For more information, see Automation Workflow *(on page 1063)*.

- On generation, before execution, on execution, and after execution

  By default, generated actions execute without the need for orchestration. Workload Optimization Manager gives you the ability to set up orchestration to affect the execution of actions.

  For more information, see Automation Workflow *(on page 1063)*.

- Execution schedule

  You can defer the execution of generated actions to a non-critical time window. For example, if a workload experiences memory bottlenecks during the week, you can defer the necessary resize to the weekend. Even if the workload has minimal utilization over the weekend, Workload Optimization Manager can recognize the need to resize, and will execute the action.

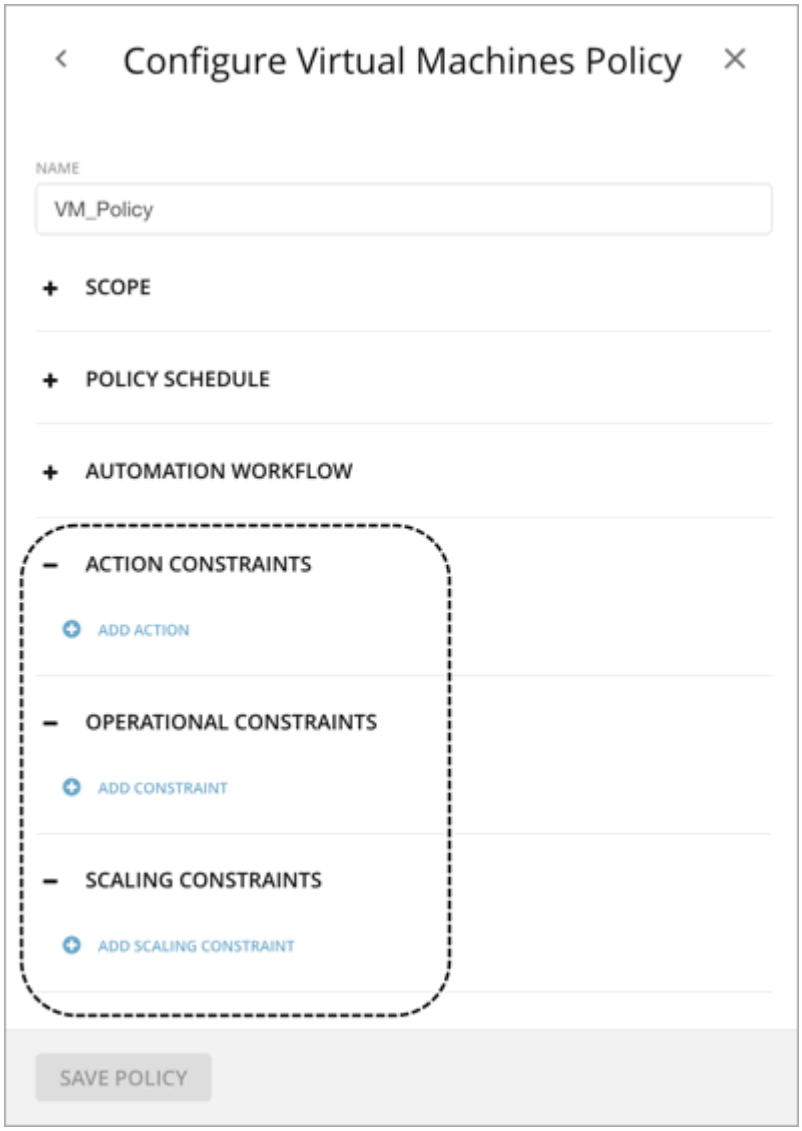
  For more information, see Action Execution Schedules *(on page 1062)*.

Constraints and other settings

Workload Optimization Manager collects metrics to drive the analysis that it uses when it calculates actions for your environment. It compares current utilization and demand against allocated capacities for resources, so it can recommend actions that keep your environment in optimal running condition.

Automation policies include constraints and other settings that you can make to adjust the analysis that Workload Optimization Manager performs. For example, you can set different levels of overprovisioning for host or VM resources, and Workload Optimization Manager will consider that as a factor when deciding on actions.

The settings you can make are different according to the type of entity this policy will affect. Each setting you add to the policy takes precedence over the default value for that setting.



Policy schedules

You can set a schedule for an automation policy, which sets a window of time when the policy takes effect. For example, you can modify the operational or scaling Constraints for a given period of time. These settings affect Workload Optimization Manager analysis, and the actions that the platform generates. You can set up scheduled times when you want to change those settings.

Remember that for user-defined automation policies, it is possible that one entity can be in two different scopes. This means that the entity can be under the effect of two different policies. For this reason, user-defined policies keep the rule, *the most conservative setting wins*. However, a more aggressive user-defined policy takes precedence over the corresponding default automation policy.

You must consider these rules when you add schedules to policies. If the more conservative settings are in a default automation policy, then the scheduled change takes effect. However, if the more conservative settings are in another user-defined policy, then the conservative settings *win*, and the scheduled changes do not take effect.

For details, see Default and User-defined Policies *(on page 1050)*.

Policy schedule and action execution schedule

A scheduled policy can include *actions*. When the policy is in effect, Workload Optimization Manager recommends or automatically executes those actions as they are generated. Some of those actions could be disruptive so you may want to defer their execution to a non-critical time window. In this case, you will need to set an *action execution schedule* within the scheduled policy. For example, you can set a policy that automatically resizes or starts VMs for your customer-facing apps for the entire month of December, in anticipation of an increase in demand. Within this same policy, you can set the resize execution schedule to Monday, from midnight to 7:00 AM, when demand is expected to be minimal.

For more information, see Action Execution Schedules *(on page 1062)*.

Action execution schedules

You can defer the execution of generated actions to a non-critical time window. For example, if mission-critical VMs experience memory bottlenecks during the week, you can defer the necessary memory resizes to the weekend. Even if the VMs have minimal utilization over the weekend, Workload Optimization Manager can recognize the need to resize, and will execute resize actions. For this particular example, you will need to:

1. Create a policy for the VMs.
2. Select *VMem Resize Up* from the list of actions and then set the action mode to either *Automatic* or *Manual*.

   NOTE:
   Execution schedules have no effect on recommended actions. It is therefore not necessary to set up an execution schedule if all the actions in your policy will be in *Recommend* mode.

3. Set an Execution Schedule that starts on Saturday at 8:00 AM and lasts 48 hours.

Execution of scheduled actions

Workload Optimization Manager posts an action at the time that the conditions warrant it, which means that you might see the action in the Pending Actions list even before the execution schedule takes effect. The action details show what schedule affects the given action, and shows the next occurrence of that schedule.

- Automatic

  When the schedule takes effect, Workload Optimization Manager executes any pending automated actions.

- Manual

  Before the execution schedule, the action details for manually executable actions show the action state as `PENDING ACCEPT`. If you accept the action (select it and click **Apply Selected**), then Workload Optimization Manager adds it to the queue of actions to be executed during the maintenance window. The action details show the action state as `AWAITING EXECUTION`. Workload Optimization Manager executes the actions when the schedule takes effect.

Keeping actions valid until the scheduled time

If you have scheduled action execution for a later time, then conditions could change enough that the action is no longer valid. If this happens, and the action remains invalid for 24 hours, then Workload Optimization Manager removes it from the list of pending actions. This action will not be executed.

Workload Optimization Manager includes scaling constraints that work to stabilize action decisions for VMs. The resulting actions are more likely to remain valid up until their scheduled window for execution. You can make these settings in the default or user-defined automation policies.

NOTE:
When you configure an execution schedule for a resize action, to ensure Workload Optimization Manager will execute the action during the scheduled time, you must turn off the **Enforce Non Disruptive Mode** setting for the policy. Even if you turn the setting off for the global policy, you still must turn the setting off for your policy. Otherwise Workload Optimization Manager will not execute the resize action. For information about non disruptive mode, see Non-disruptive Mode *(on page 747)*.

Automation workflow

Automation Workflow specifies whether Workload Optimization Manager will execute an action, or whether Workload Optimization Manager will automate external workflows to effect the change in your environment. In this way, you can integrate supported orchestrators to execute of actions for specific scopes of entities in your environment. For information about Orchestration target requirements, see Orchestrator Targets *(on page 360)*.

Apart from controlling the execution, automation workflows also allow actions to seek external approval from external systems such as ServiceNow before executing the action.

NOTE:
For some orchestration workflows, it is necessary to schedule an action to execute only during a specific maintenance window. Workload Optimization Manager policies can include schedules to enable this use case. However, you must be sure that you do not set the schedule to the policy that declares the orchestration you want. Instead, you should use two policies for the same scope – one to set up the orchestration, and another to schedule the time window during which the action acceptance mode will be `Automated` (to set up the maintenance window). For more information, see Setting Policy Schedules *(on page 1061)*.

About automation workflows

Automation workflows can add custom processing to Workload Optimization Manager actions at different entry points.

You can create an automation workflow that triggers a webhook whenever Workload Optimization Manager recommends moving a VM, or you can create an automation workflow that runs as a replacement for the action that Workload Optimization Manager would execute.

NOTE:
You can specify one or more critical or non-critical workflows. The action is considered failed if any one of the critical workflows returns an error. If a non-critical workflow returns an error, the action will continue and is not marked as failed.

For example:

- Webhook workflow

  To implement a webhook, you create a workflow that specifies parameters such as the HTTP URL, HTTP method, and request body template. You can then use one or more webhook workflow at each action stage to perform activities such as posting to a collaboration channel, sending an email, triggering a notification, or performing the action itself instead of Workload Optimization Manager. For more information about webhooks, see Webhooks *(on page 1075)*.

- Action Script workflow

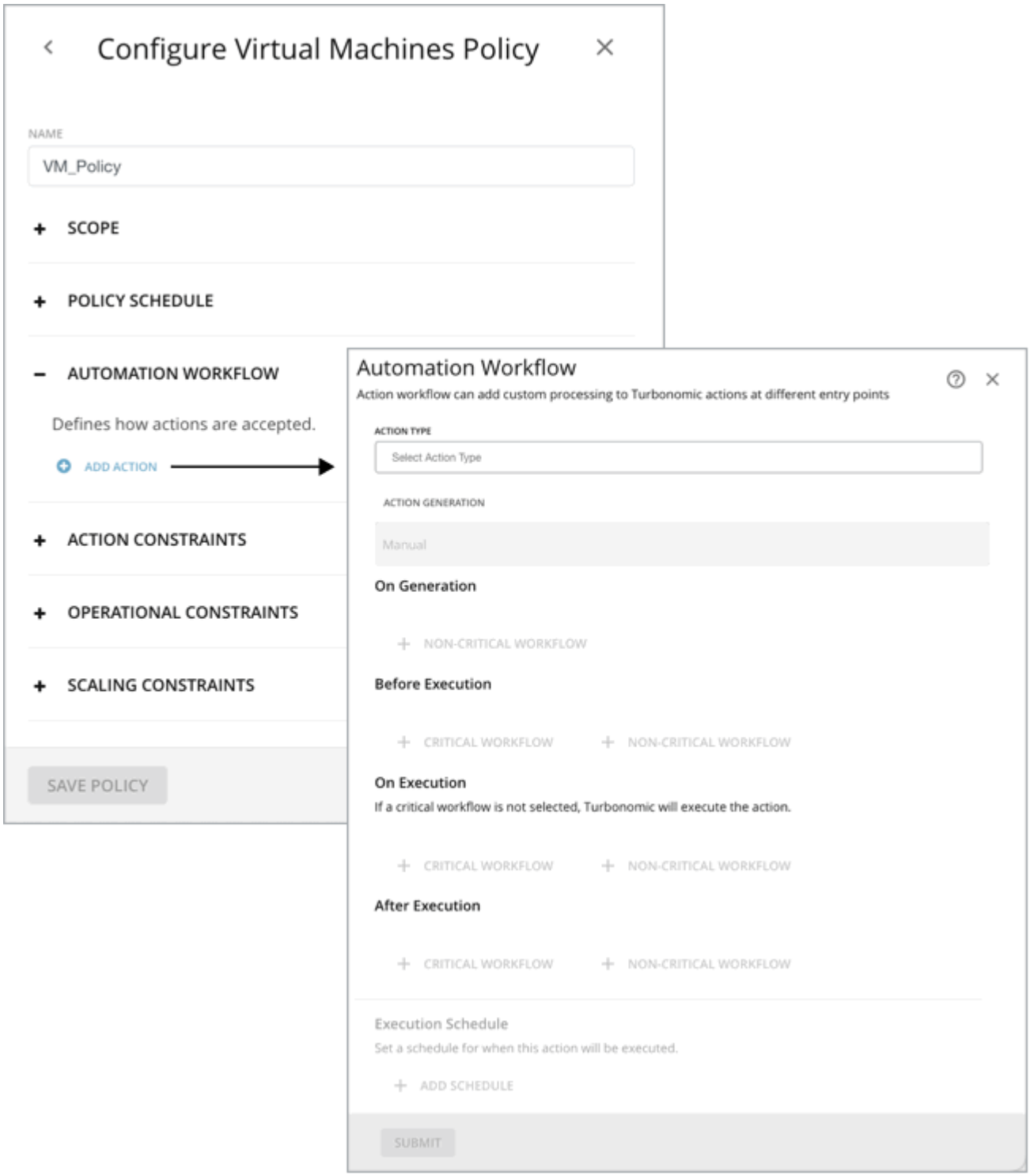
  You can deploy action scripts on a remote machine and configure an Action Script target that communicates with this server. Workload Optimization Manager discovers the exposed scripts and displays them as options you can choose when you specify a workflow in your automation policy. For more information about Action Scripts, see Deploying Action Scripts *(on page 1070)*.

Specifying automation workflows

As you create a policy, you specify the entity type and the scope of entities the policy affects. You can also set modes for specific actions. For example, you can set a mode of Manual for the Resize action for a given scope of VMs.

This section shows how to link orchestration workflows to automation policies. It assumes you have already configured an appropriate Orchestration target. It also assumes that you have configured workflows on that target in such a way that Workload Optimization Manager can discover the workflows and map them to automation policies.

1. On the Policy page for the selected entity, expand **Automation Workflow** and click **Add Action**.



2. Select one or more **Action Types** you want to orchestrate.
3. Select the **Action Generation** mode.

By default, accepted actions execute without the need for orchestration. Workload Optimization Manager gives you the ability to set up orchestration to affect the approval and execution of actions. Actions are staged in the following sequence:

- **On Generation** – Workflows are called when Workload Optimization Manager makes a decision that an action is required.
- **Before Execution** – Workflows are called just before an accepted action is executed.
- **On Execution** – By default, Workload Optimization Manager executes the action using the target that is managing the entity. If a workflow is not specified, Workload Optimization Manager executes the action with its default (Native) action processing.

  If specified, critical workflows are called to execute the action instead of the default Workload Optimization Manager action execution. Non-critical workflows are called on executing the action.
- **After Execution** – Workflows are called just after an accepted action is executed. To finalize an action, specify a critical workflow.

Workflows within each stage are not sequenced and are triggered in the order they were added. Workflows operate in isolation, and there is no data passing between workflows. Depending on the selected Action Generation mode, one or more additional steps may be required for each Action Stage. Refer to the following table for next steps.

Action generation mode	Description	Next steps
Do not generate actions	Actions are not generated. Workload Optimization Manager never considers your selected actions in its calculations. For example, if you do not want to generate Resize actions for VMs in the policy, analysis will still drive toward the desired state, but will do so without considering resizes.	No further steps are required.
Recommend only	Actions are generated but cannot be taken within Workload Optimization Manager.	You can, optionally, add a non-critical workflow on action generation, such as to notify that the action was

Action generation mode	Description	Next steps
		generated so that someone can review it.
Manual	Actions are generated and can be manually accepted for execution by a user. By default, on execution, Workload Optimization Manager executes the action using the target that is managing the entity when a user accepts the action.	You can, optionally, specify critical and non-critical workflows to be triggered On generation, Before, On execution, or After execution.
Automated	Actions are generated and automatically accepted for execution. By default, on execution, Workload Optimization Manager executes the action using the target that is managing the entity. For automated resize or move actions on the same entity, Workload Optimization Manager waits five minutes between each action to avoid failures associated with trying to execute all actions at once.	You can, optionally, specify critical and non-critical workflows to be triggered On generation, Before, On execution, or After execution.

Action generation mode	Description	Next steps
	Any action awaiting execution stays in queue. For example, if a VM has both vCPU and vMem resize actions, Workload Optimization Manager could resize vCPU first. After this resize completes, it waits five minutes before resizing vMem.	
Automated when approved Available only after adding a ServiceNow target	Actions are generated and automatically accepted for execution only after they are externally approved. The ServiceNow target is automatically set as the critical on generation workflow and you can optionally specify additional non-critical workflows. After approval, on execution, Workload Optimization Manager executes the action using the target that is managing the entity.	You can, optionally, specify critical and non-critical workflows to be triggered Before, On execution, or After execution.

4. (Optional) Specify an execution schedule.

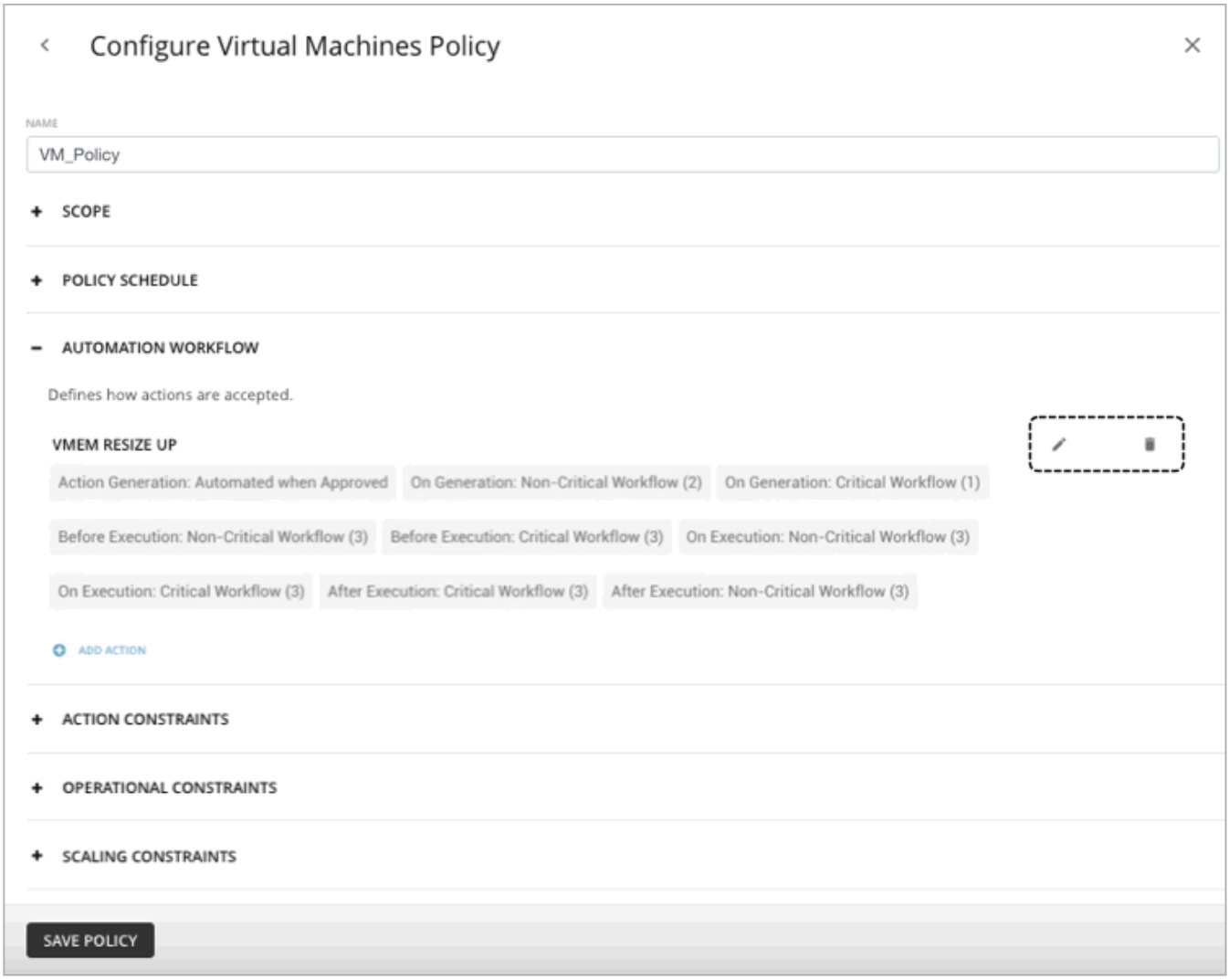
There is no execution schedule by default; Workload Optimization Manager executes the action immediately. If the policy includes a schedule, Workload Optimization Manager executes the action at the scheduled time.

You can defer the execution of generated actions to a non-critical time window. For example, if a workload experiences memory bottlenecks during the week, you can defer the necessary resize to the weekend. Even if the workload has minimal utilization over the weekend, Workload Optimization Manager can recognize the need to resize, and will execute the action.

> **NOTE:**
> Workload Optimization Manager discovers and enforces execution schedules defined in automation workflows. To avoid potential issues with schedules, set the execution schedule either in the orchestrator (such as ServiceNow) or Workload Optimization Manager.

5. When you have made all your settings, save the automation policy.

   The Automation Workflow section of the policy page updates to show the workflows that were specified for the selected policy.



To edit the policy's workflows, click the pencil icon to open the Automation Workflow page or delete the workflow by clicking the delete icon.

Deploying action scripts

Action scripts provide an interface that can add custom processing to Workload Optimization Manager actions.

Action scripts execute on a remote server (a VM or a container) that you have configured as a Workload Optimization Manager target. That server includes a manifest file that identifies the scripts you have deployed, as well the entities and actions they can respond to. Workload Optimization Manager discovers these scripts via the manifest and presents them as orchestration options for actions in automation policies.

For example, assume you have defined a script with:

- `name: MyVmMoveAction`
- `entityType: VIRTUAL_MACHINE`
- `actionType: MOVE`

Following this example, you can use the API to add orchestration to a policy for move actions on VMs. Because you have defined a script for that action, you can specify Action Script as the orchestration type, and you can choose the `MyVmMoveAction` script as the orchestration workflow to perform.

To deploy your action scripts, you will:

- Set up the remote Action Script Server (see Setting Up the Action Script Server *(on page 1070)*)
- Create the action script executables on the remote server (see Creating Action Scripts *(on page 1071)*)
- Deploy the Action Script Manifest on the remote server (see Deploying the Action Script Manifest *(on page 1073)*)

Setting up the action script server

Workload Optimization Manager uses remote servers to run action scripts. Managing the processes remotely means that you do not install custom code on the Workload Optimization Manager server, which eliminates associated security risks there. However, you are responsible for maintaining the security of your Action Script Server to ensure the integrity of your custom code. To accomplish this, the configuration of the remote server must meet certain requirements.

Resource requirements for the server

The remote server can be a VM or a container. The capacity you configure for the server depends entirely on the processes you intend to run on it. Workload Optimization Manager does not impose any special resource requirements on the server.

Configuring command execution

To support execution of your scripts, you must install any software that is necessary to run the scripts, including libraries, language processors, or other processes that your scripts will call.

Workload Optimization Manager calls the scripts as commands on the server. The server must run an SSH service that you configured to support command execution and SFTP operations. Currently, Cisco tested action scripts with the OpenSSH sshd daemon.

The standard port for SSH is 22. You can configure a different port, and provide that for admins who configure the server as an Action Script target.

> **NOTE:**
> An action script can invoke any process you deployed on the remote server. You do not have to run scripts; however, you must be able to invoke the processes from the command line. The script manifest gives Workload Optimization Manager the details it needs to build the command line invocation of each script.

Configuring the action script user account

To run the scripts on your server, Workload Optimization Manager logs on using a user account that is authorized to run the scripts from the command line. You provide the user credentials when you configure the Action Script target. To support this interaction, the user account must meet the following requirements:

- Public Key

  The user must have an RSA public key in the `.ssh/authorized_keys` file. When you configure the Action Script target, you provide this as the Private Token for the target.

  NOTE:
  For the Action Script target to validate, the permissions on the `authorized_keys` file must be set to `600`.

- Security for the `.ssh` Directory

  The Action Script User should be the only user with authorized access. Set file permissions to `700`.

- Supported Shells

  The Action Script User shell can be either the Bourne shell (usually at `/bin/sh`) or the Bourne-Again shell (usually at `/bin/bash`). Workload Optimization Manager passes parameters as it invokes your scripts. At this time it only supports script execution through these shells.

Action script manifest

The Action Script Manifest is a file that declares an array of Script Objects for each script you want to expose to Workload Optimization Manager. You can create the manifest as either a JSON or a YAML file. You can save the Manifest file to any location on your server, so long as the Scripts User has access to that location, and has read and execute privileges. The location of the manifest is provided as part of the Action Script target configuration.

After Workload Optimization Manager validates the target, it then discovers these scripts and presents them in the Orchestration Policy user interface.

For details, see Deploying the Action Script Manifest *(on page 1073)*.

Handling action script timeouts

Workload Optimization Manager limits script execution to 30 minutes. If a script exceeds this limit, Workload Optimization Manager sends a `SIGTERM` to stop the execution of the process.

Note that Workload Optimization Manager does not make any other attempt to end a process. For example, you might implement the script so it traps the `SIGTERM` and continues to run. The process should end at the soonest safe opportunity. However, if the process does not end, then you must implement some way to stop it outside of Workload Optimization Manager. A runaway process continues to use its execution thread, which can block other processes (action scripts or primary processes) if there are no more threads in the pool.

Creating action scripts

An action script can be any executable that a user can invoke from a command line. You can save these executable files anywhere on the server – The Manifest indicates the path to the file (see Deploying the Action Script Manifest *(on page 1073)*). The Action Script user that you have configured for the script server must have access to your script files, with read and execution privileges.

To execute a script, Workload Optimization Manager builds the appropriate SSH command from the manifest information it has discovered. It grants a timeout limit of 30 minutes by default, or the manifest entry can declare a different limit. If the execution exceeds the limit, Workload Optimization Manager sends a `SIGTERM` to terminate the process.

Passing information to the action script

Workload Optimization Manager uses two techniques to pass information about an action to the associated action script:

- Pass general information via environment variables
- Pass full action data via `stdin`

To pass general information into the script, Workload Optimization Manager sets environment variables on the Action Script Server. You can reference these environment variables in your scripts. For example, assume you want to send an email that includes the name of the VM that is an action target. You can get that name via the VMT_TARGET_NAME environment variable.

The following list shows the environment variables that Workload Optimization Manager can set when it executes a script. Note that not all of these variables apply for every action. For example, an action to scale VMEM does not include providers, so the action does not include values for the VMT_CURRENT_INTERNAL, VMT_CURRENT_NAME, VMT_NEW_INTERNAL, or VMT_NEW_NAME variables. If a given variable does not apply, Workload Optimization Manager sets it to an empty string.

- `VMT_ACTION_INTERNAL`

  The UUID for the proposed action. You can use this to access the action via the REST API. For example, your script could accept or cancel the action according to its own criteria.
- `VMT_ACTION_NAME`

  The name of the action.
- `VMT_CURRENT_INTERNAL`

  The internal name for the current provider.
- `VMT_CURRENT_NAME`

  The display name for the current provider.
- `VMT_NEW_INTERNAL`

  The internal name for the new provider.
- `VMT_NEW_NAME`

  The display name for the new provider.
- `VMT_TARGET_INTERNAL`

  The internal name of the entity this action will affect. You can use this to access the target entity via the REST API. For example, you can get historical statistics or you can change settings for the entity.
- `VMT_TARGET_NAME`

  The display name of the entity this action will affect.
- `VMT_TARGET_UUID`

  The UUID of the entity this action will affect.

For some scripts, you might need a complete description of the associated action. For example, assume you want to analyze the utilization metrics for a given resource. The environment variables for passing general information do not include this information.

When it invokes an action script, Workload Optimization Manager passes the complete data for the associated action via `stdin`. Your script can load this into a variable to access the specific data it needs. For example, the following loads `stdin` into `myActionData`:

```
myActionData=$(cat -)
```

`stdin` contains a JSON string that represents of the full data associated with this action. For example, the `myActionData` variable could contain a string similar to:

```
{"actionType":"RIGHT_SIZE","actionItem":[{"actionType":"RIGHT_SIZE","uuid":"143688943343760","targetS
E":{"entityType":"VIRTUAL_MACHINE","id":"4200fcdb-eafe-2a4a-abf5-a7ad2b00555c"...
```

Deploying the action script manifest

The Action Script Manifest is a file that declares an array of Script Objects for each script you want to expose to Workload Optimization Manager. The location of the manifest is provided as part of the Action Script target configuration.

After Workload Optimization Manager validates the target, it then discovers these scripts and presents them in the Orchestration Policy user interface.

Creating the scripts manifest file

You can create the manifest as either a JSON or a YAML file. For example, following are two examples of the same manifest: one in YAML and the other in JSON. Notice that in either case, the manifest is an array of two Script objects.

- YAML Manifest example

```
scripts:
    - name: MyVmMovePrep
      description: Execute this script in preperation to a VM Move
      scriptPath: vmScripts/movePrep.sh
      entityType: VIRTUAL_MACHINE
      actionType: MOVE
      actionPhase: PRE
    - name: MyVmSuspendReplace
      description: Execute this instead of a VM Suspend action
      scriptPath: vmScripts/suspendReplace.sh
      entityType: VIRTUAL_MACHINE
      actionType: SUSPEND
      actionPhase: REPLACE
```

- JSON Manifest example

```
{
    "scripts": [
     {
      "name": "MyVmMovePrep",
      "description": "Execute this script in preperation to a VM Move",
      "scriptPath": "vmScripts/movePrep.sh",
      "entityType": "VIRTUAL_MACHINE",
      "actionType": "MOVE",
      "actionPhase": "PRE"
     },
     {
      "name": "MyVmSuspendReplace",
      "description": "Execute this instead of a VM Suspend action",
      "scriptPath": "vmScripts/suspendReplace.sh",
      "entityType": "VIRTUAL_MACHINE",
      "actionType": "SUSPEND",
      "actionPhase": "REPLACE"
     }
    ]
}
```

You can save the Scripts Manifest file to any location on your server, so long as the Scripts User has access to that location, and has read and execute privileges. You will provide this location as the **Script Path**, which the Workload Optimization Manager administrator will give as part of the Action Script target configuration.

Note that the filename extension for the manifest must match the file format (either YAML or JSON). For example, you should name the file either `MyManifest.yaml` or `MyManifest.json`, respectively.

Declaring script objects

Each script object in the manifest can contain the following fields:

- **`name`**

  Required – The name for this action script. After Workload Optimization Manager discovers your scripts, it displays this name as a Orchestration Workflow choice in the user interface for creating orchestration policies.

- **`description`**

  Optional – A description of the script. The Workload Optimization Manager user interface does not display this description.

- **`scriptPath`**

  Required – The path to the executable for this entry. You can give an absolute path, or a path that is relative to the location of the Scripts Manifest. The Action Script User that you set up for the Action Script server must have read and execute privileges for the executable file.

- **`entityType`**

  Required – The type of entity this script responds to. Can be one of:

  - `Switch`
  - `VIRTUAL_DATACENTER`
  - `STORAGE`
  - `DATABASE_SERVER`
  - `WEB_SERVER`
  - `VIRTUAL_MACHINE`
  - `DISK_ARRAY`
  - `DATA_CENTER`
  - `PHYSICAL_MACHINE`
  - `CHASSIS`
  - `BUSINESS_USER`
  - `STORAGE_CONTROLLER`
  - `IO_MODULE`
  - `APPLICATION_SERVER`
  - `APPLICATION`
  - `CONTAINeR`
  - `CONTAINER_POD`
  - `LOGICAL_POOL`
  - `STORAGE_VOLUME`
  - `DATABASE`
  - `VIEW_POD`
  - `DESKTOP_POOL`

  To configure the same script to respond to actions on different entity types, declare separate entries for that script, one for each entity type.

- **`actionType`**

  Required – The type of action this script responds to. Note that different entity types can support different actions. Can be one of:

  - `START`
  - `MOVE`
  - `SCALE`

Resize on cloud – move workload from one cloud template or tier to another.

  - SUSPEND
  - PROVISION
  - RECONFIGURE
  - RESIZE
  - DELETE
  - RIGHT_SIZE
  - ACTIVATE
  - DEACTIVATE
  - BUY_RI

- **actionPhase**

  Required – Where in the life cycle of an action that you want your script to execute.

  Can be one of:

  - PRE

    For an action that has been accepted, or an AUTOMATED action before it executes, this state is a preparation phase where your script can execute just before the action itself executes.

    Run your script to set up conditions just before the action executes.

  - REPLACE

    For action execution, your script executes instead of the execution that Workload Optimization Manager would perform.

    Run your script as a replacement for the Workload Optimization Manager action.

  - POST

    The action has completed execution, either in a SUCCEEDED, FAILING, or FAILED state.

    FAILING means that the status was checked after the action execution fails, but before the POST script has finished execution.

    Run your script after the action has completed execution.

- **timeLimitSeconds**

  Optional – How long to run the action before assuming a timeout. When execution exceeds this limit, Workload Optimization Manager sends a SIGTERM to terminate the execution of the process.

  If you do not provide a value, Workload Optimization Manager assumes a limit of 30 minutes (1800 seconds).

Webhooks

You can configure automation policies in Workload Optimization Manager to send data through webhooks to external API servers. A webhook is an automated message that Workload Optimization Manager can use to send data to external applications.

You can use webhooks for many different tasks, including the following examples:

- Send notifications to collaboration platforms such as Slack
- Integrate Workload Optimization Manager with workflow management systems
- Override Workload Optimization Manager actions with your own logic

The webhook implementation supports HTTP/S messaging.

To set up a webhook, complete the following steps:

1. Identify the application to receive the webhook.

   Possible applications include collaboration platforms such as Slack, orchestration platforms such as ServiceNow, cloud provider APIs, or a custom application that responds to HTTP/S methods.

2. Create a webhook workflow in your Workload Optimization Manager instance.

   To create a webhook, you can either use the **Settings > Workflows** UI wizard, or the Workload Optimization Manager Rest API. A webhook definition can include the following elements:

   - The URL of the webhook endpoint for the application
   - An HTTP/S method
   - A template for the webhook request body
   - Authentication credentials to access the application
   - Optionally, HTTP/S proxy server details

   For information about creating a webhook workflow with the Workload Optimization Manager API, see Creating Webhook Workflows with the API *(on page 1076)*.

   For information about creating a webhook workflow with the Workflow wizard in the Workload Optimization Manager UI, see Creating webhook workflows with the workflow settings wizard *(on page 1078)*.

3. Create an Automation Policy that uses the webhook workflow.

   Automation Policies include automation workflow settings where you can choose to trigger one or more webhook workflows at each action stage. Workload Optimization Manager can execute a webhook on action generation, before it executes the action, instead of executing the action, and after it executes the action.

   For information about creating policies that use webhook workflows for orchestration, see Automation Workflow *(on page 1063)*.

When Workload Optimization Manager generates or executes the action that you identified in the policy, it sends a message to the URL that you specified in the webhook.

Creating webhook workflows with the API

To implement a webhook, you create a workflow that specifies parameters such as the HTTP URL, HTTP method, and payload template. You can then use this workflow in your automation policy to orchestrate how actions execute.

For more information, see Workflows Endpoint *(on page 1434)*.

To create a workflow, you can use the API to POST a Workflow object to Workload Optimization Manager instance. For example, the following `curl` commands get authorization to access a Workload Optimization Manager server, and then add a simple webhook workflow to that server:

- Authenticate on the server

  This command requests authentication credentials and stores them in a variable you can set to a cookie in subsequent curl headers, where:

  - <T8c_IP_ADDRESS> is the address of the Workload Optimization Manager server
  - <ADMIN_ACCOUNT_NAME> is the name of an account with admin privileges
  - <ADMIN_PWD> is the admin account password

```
JSESSIONID=$(curl \
    --silent \
    --cookie-jar - \
    --insecure \
    https://<T8c_IP_ADDRESS>/vmturbo/rest/login \
    --data "username=<ADMIN_ACCOUNT_NAME>&password=<ADMIN_PWD>" \
    | awk '/JSESSIONID/{print $7}')
```

- Create the workflow

  This command creates the workflow on the server, where:

  - <T8c_IP_ADDRESS> is the address of the Workload Optimization Manager server
  - <WEBHOOK_ADDRESS> is the address of the webhook server

```
curl \
    "https://<T8c_IP_ADDRESS>/api/v3/workflows" \
    --insecure \
```

```
   --compressed \
   --header 'Accept: application/json' \
   --header 'Content-Type: application/json' \
   --header "cookie: JSESSIONID=$JSESSIONID" \
   --request POST \
   --data '
   {
       "displayName": "My_WebHook",
       "className": "Workflow",
       "description": "First webhook attempt.",
       "discoveredBy":
       {
           "readonly": false
       },
      "type": "WEBHOOK",
      "typeSpecificDetails": {
      "url": "http://<WEBHOOK_ADDRESS>",
         "method": "POST",
         "template": "{ \"text\":\"My Webhook Template -- DATA: Action Details:
$action.details\" }",
         "type": "WebhookApiDTO"
      }
   }
   '
```

This is a simple webhook that sends its template to the indicated URL.

The template payload is the string `My Webhook Template -- DATA: Action Details:`, plus the action details that are included in the action's data object. The variable `$action.details` is a reference to a field in the `ActionApiDTO` object that represents the current action. Your template can reference any of the fields in this DTO, starting with `action` as the object name. For example, `$action.createTime` gives you the time the action was created. For a full listing of the `ActionApiDTO` object, see ActionApiDTO *(on page 1485)* or the API Swagger UI.

For pre-defined request payload examples, see the Workload Optimization Manager Orchestration GitHub repository.

Sample webhook application

A webhook workflow sends a message to an application via HTTP. You express the message as a template that can include values from the action data in its payload. This template can express text, JSON, or any other payload that your application can accept.

You can use webhooks to send messages to a number of existing applications, including Slack, Amazon Web Services, and others.

To deploy a simple example, and to test your webhook templates, you can implement a node.js server that receives the webhook message and prints out the template data. If you install this server on a machine in your network, then you can give its URL in the webhook workflow, and test your response to specific actions.

Following is a listing for a node.js web server that you can use.

```
let port = 9090;
const http = require("http");
console.log(`Starting server on port ${port}`);
```

```
http.createServer((request, response) => {
    request.setEncoding('utf8');
    console.log('REQUEST METHOD: ', request.method);
```

```
    let datStr = '';
    request.on('data', chunk => {datStr = datStr + chunk});
    request.on('end', () => {console.log('End of DATA: ', datStr)})
}).listen(port);
```

When you run this program, it prints a message to the console to say that it is running, and to identify the port it listens on.

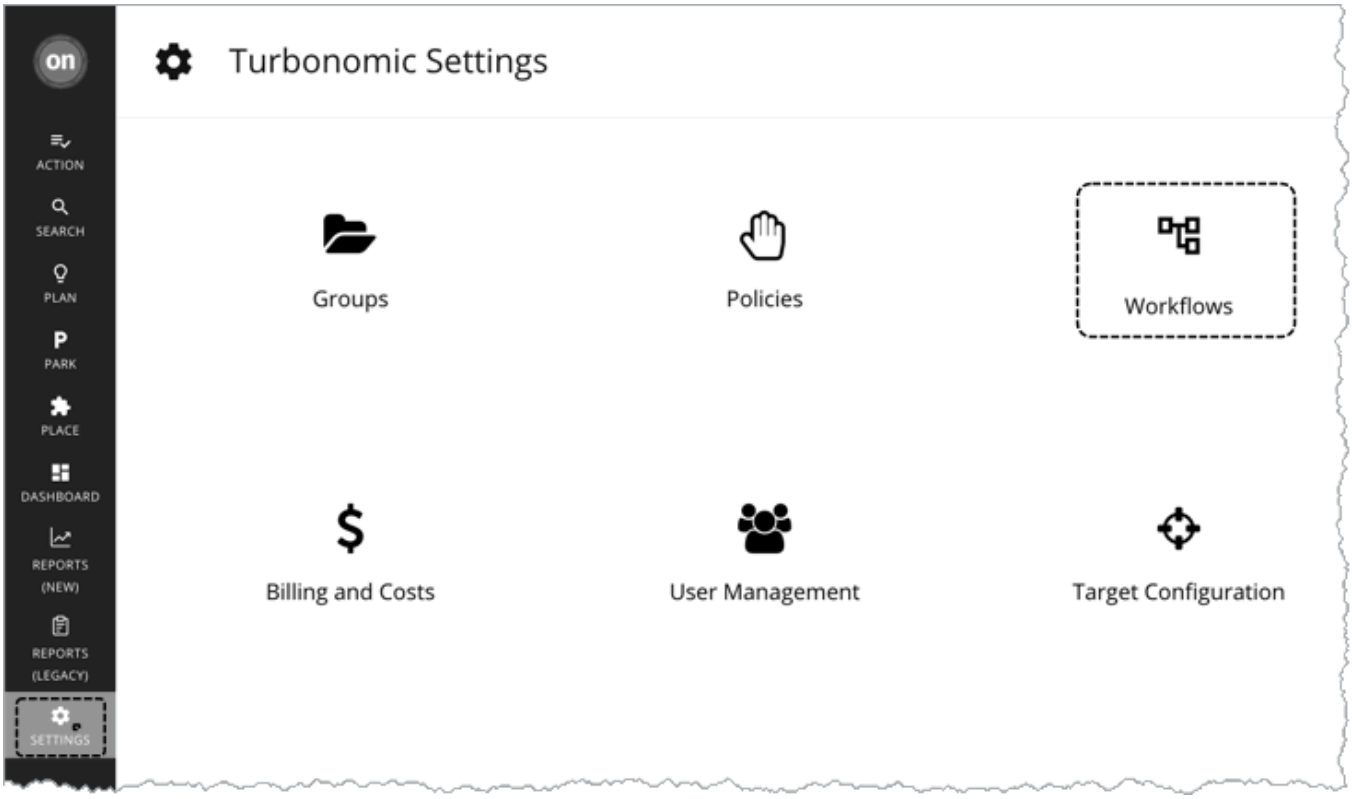
When the server receives a message, it prints out the request method, and then prints out the message payload, as specified in the workflow's `template` field.

If you have configured an Automation Policy to use this workflow, then this server will log a message for each action that Workload Optimization Manager executes on an entity within the policy's scope.

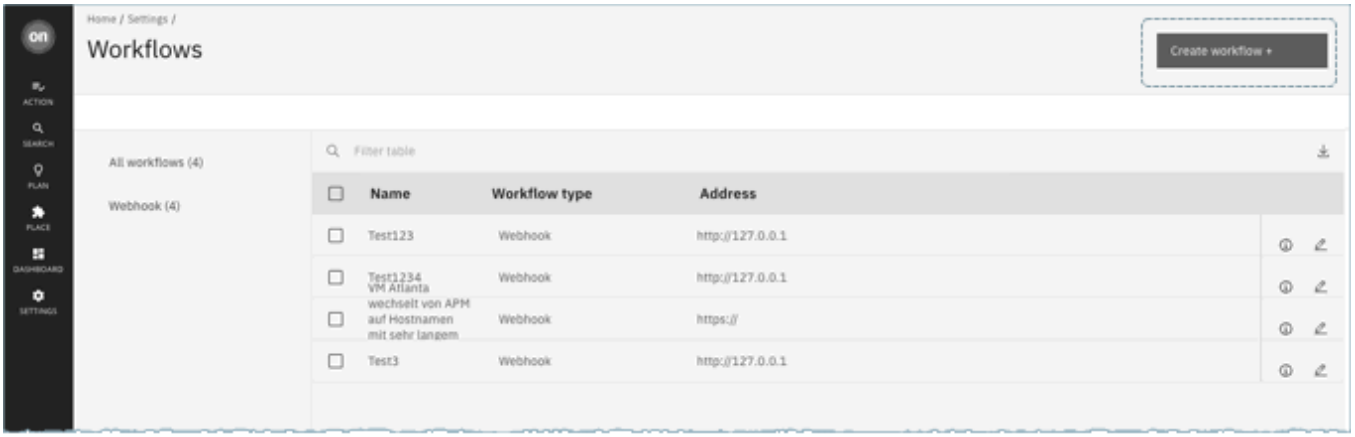
Creating webhook workflows with the workflow settings wizard

To implement a webhook, you create a workflow that specifies parameters such as the HTTP/S URL, HTTP/S method, and request body template. You can then use this webhook in an automation workflow policy.

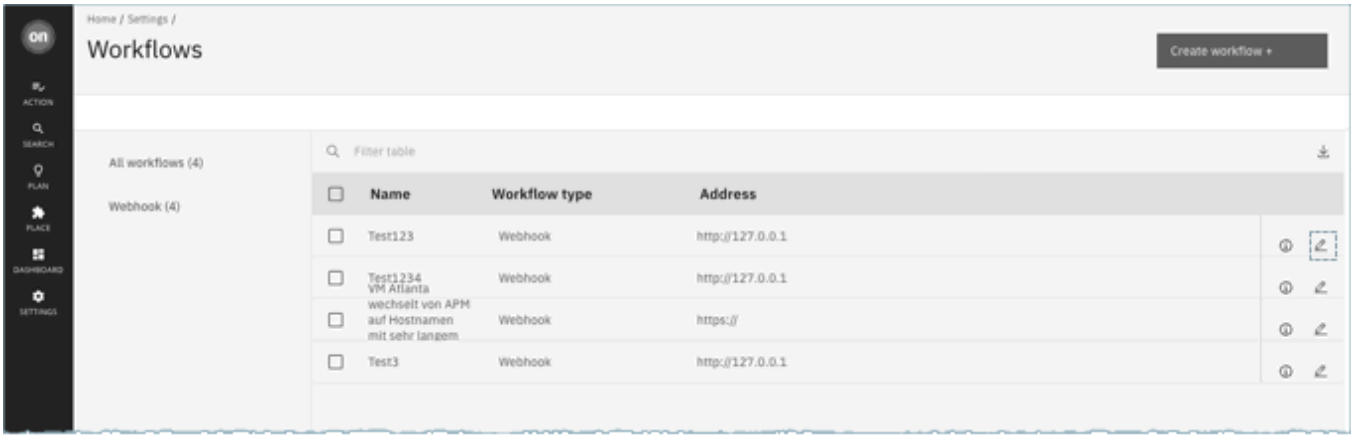
1. In the Workload Optimization Manager UI, select **Settings** from the navigation menu and click **Workflows**.



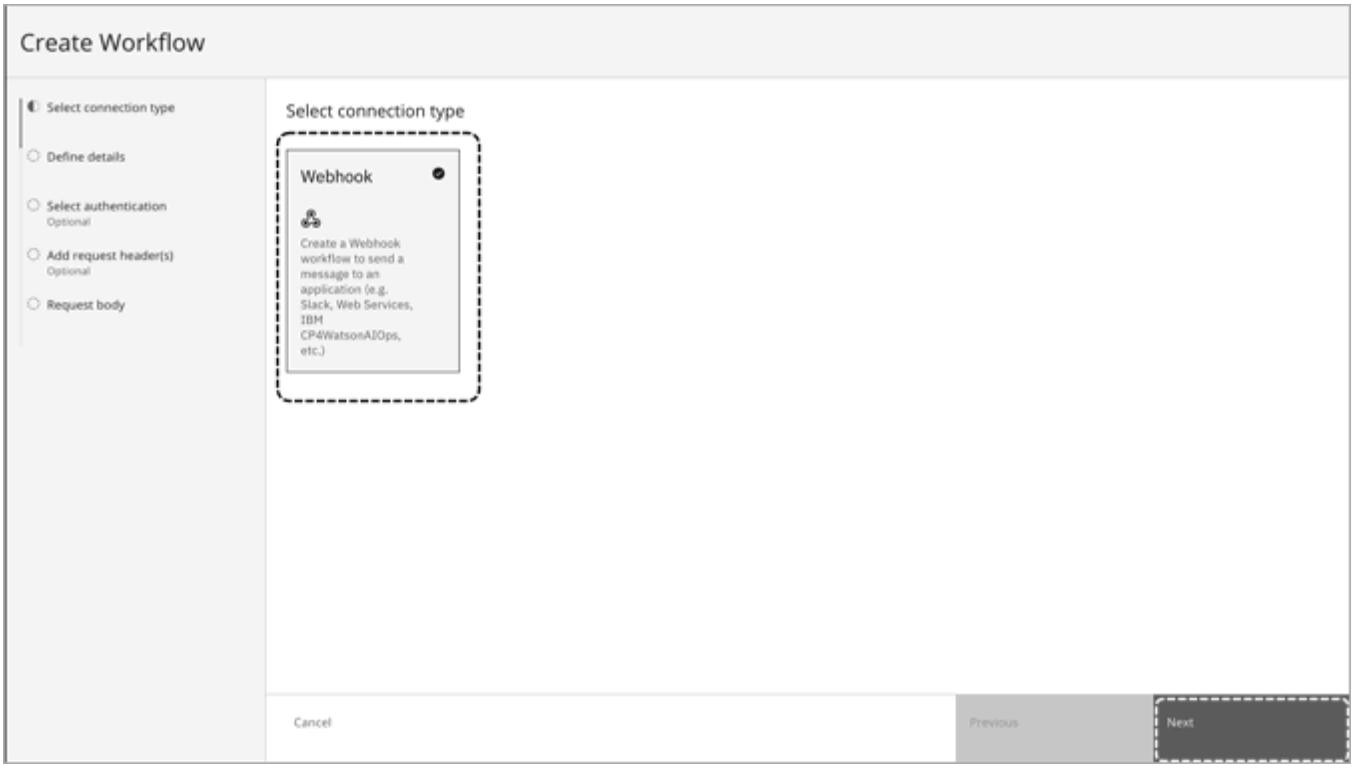
2. Click **Create workflow**.



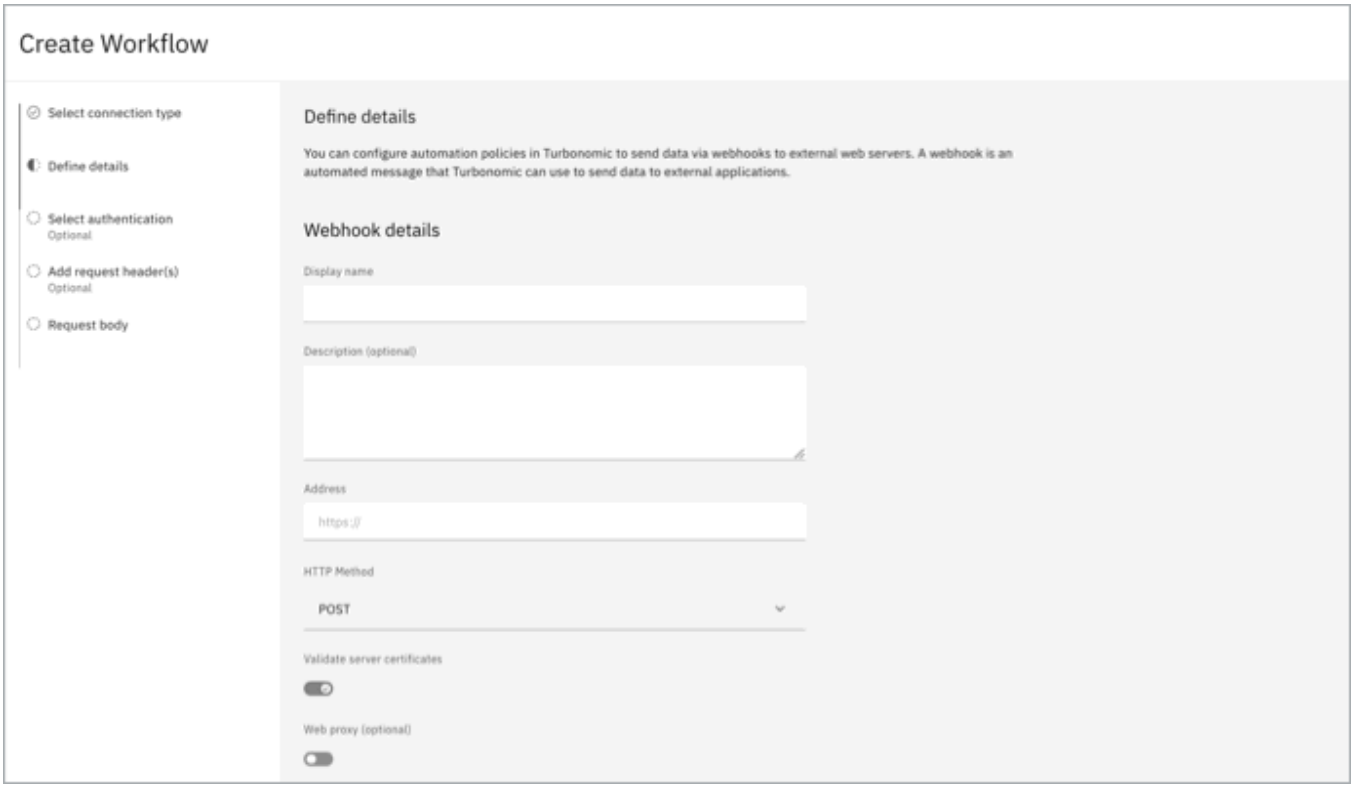
To edit an existing webhook instead of creating a new one, click the pencil icon for the webhook that you want to edit.



3. In the **Select connection type** panel, click **Webhook** and then click **Next**.



4. Define the details for your webhook.



a. **Display name**: Give your webhook a short descriptive name.

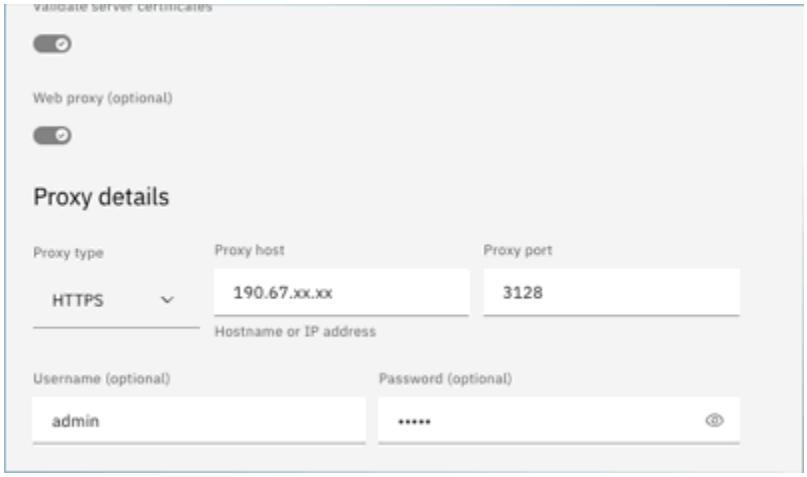
   Description: Optionally, provide a longer description.

   Address: Specify the HTTP/S endpoint for your webhook server.

   HTTP Method: Select the HTTP/S method.

   Validate server certificates: Optionally, deselect to skip server certificate validation (for example, to trust self-signed server certificates).

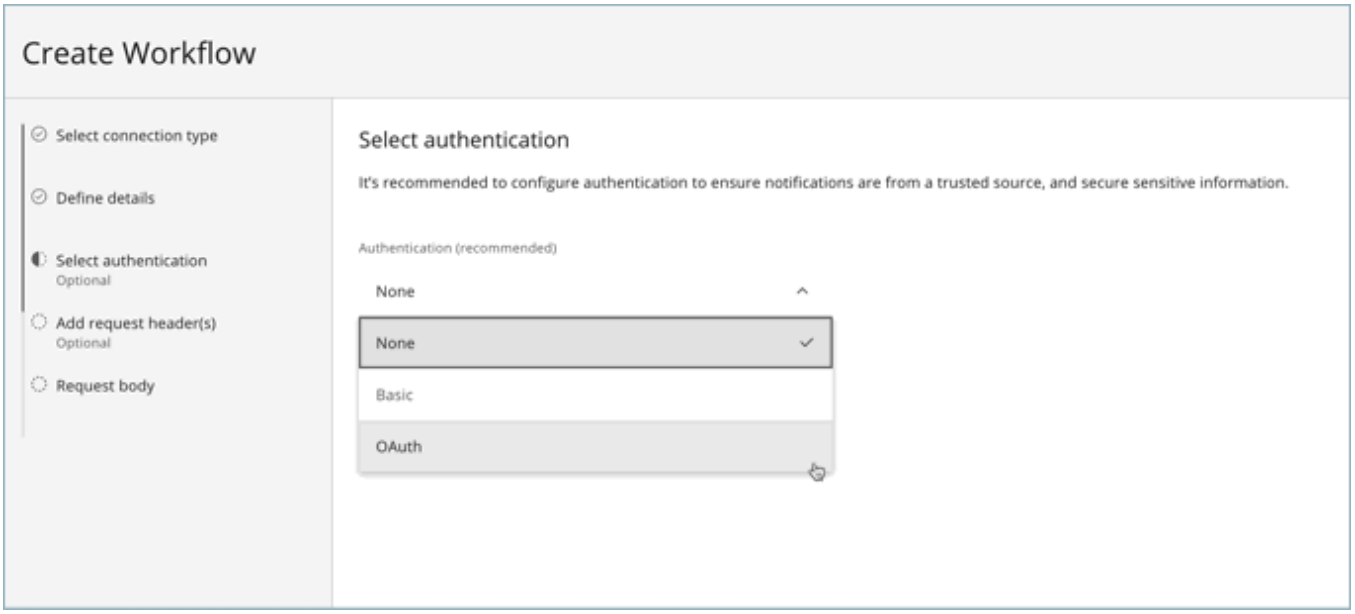
   Web proxy: Optionally, select to enable a proxy host for your webhook.



   Specify the proxy type, host, and port. Optionally, specify a username and password to authenticate to the proxy host.

b. Click **Next**.

5. Configure authentication.

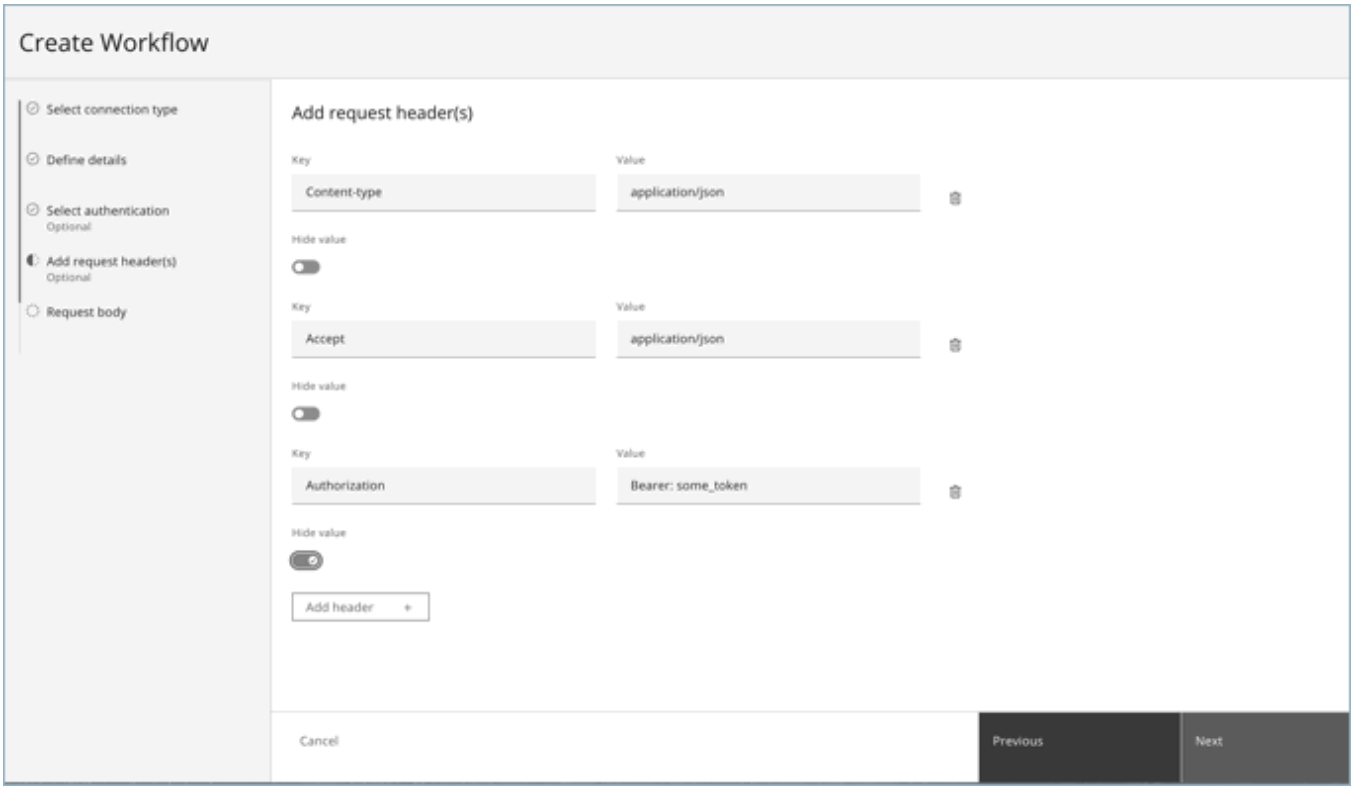


a. You can configure authentication to meet the requirements of your webhook API server. Workload Optimization Manager supports **None**, **Basic**, and **OAuth** authentication.

   - **None**: Bypass authentication.
   - **Basic**: To configure basic authentication, specify a username and password.
   - **OAuth**: To configure OAuth authentication, specify a client ID, client secret, and the URL of your authorization server.

b. Click **Next**.

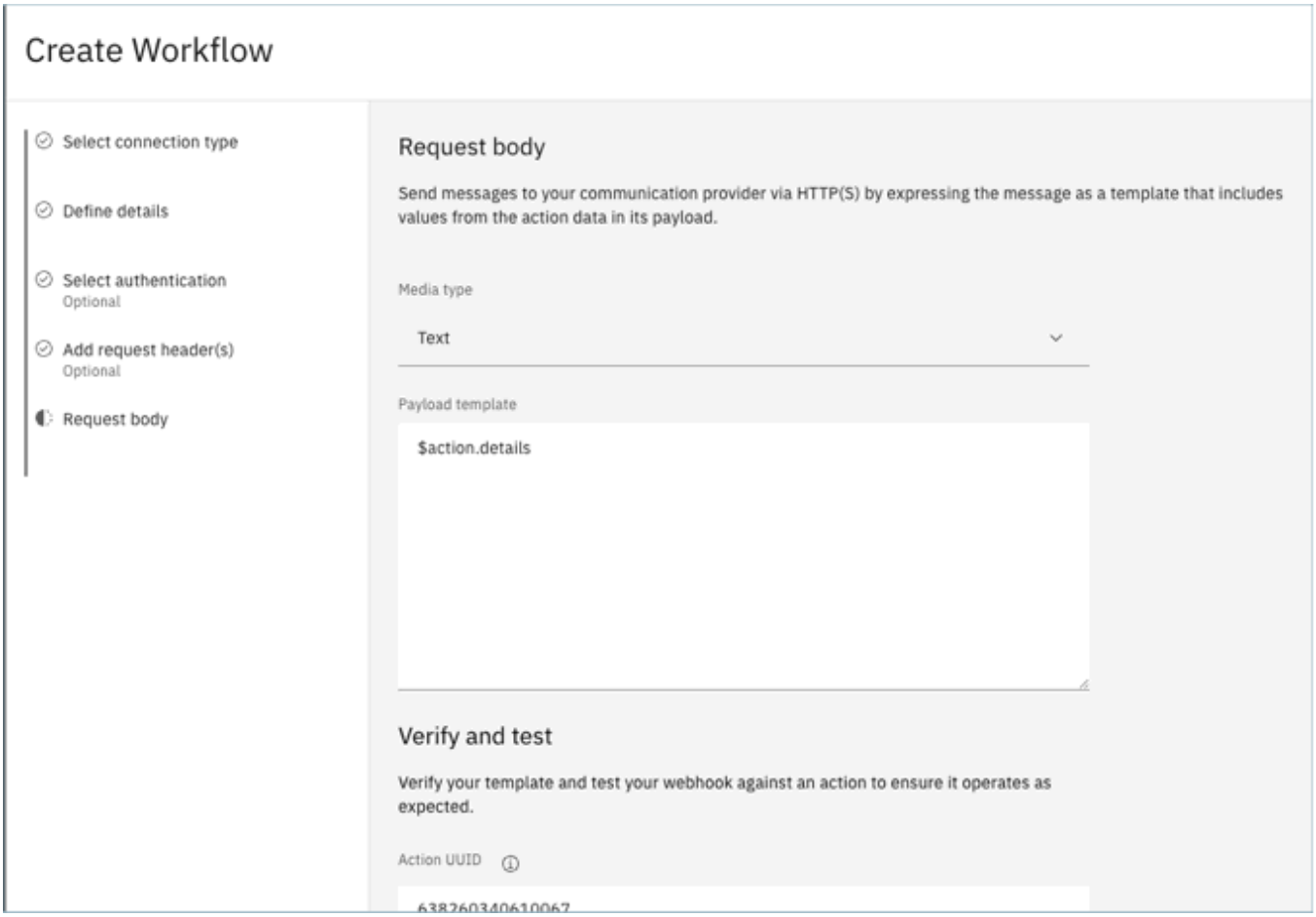
6. Add request headers.



a. To add request headers, click **Add header**. You can specify multiple headers to include in your webhook. If your header includes a secret, you can click the **Hide value** switch so that it does not appear again after it is entered.

b. Click **Next**.

7. Specify the request body.



The request body template uses keywords to substitute action details when the webhook message request is sent. To verify that the keyword substitution is working as wanted, you can provide an existing action UUID to use an as example.

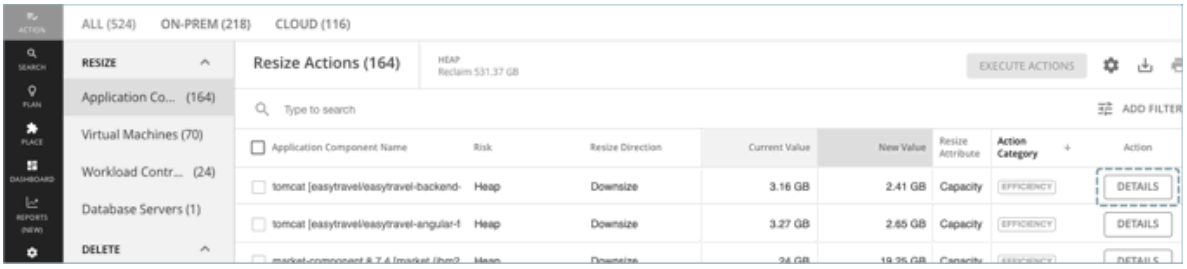
a. **Media type**: Specify a media type for your webhook request body. Choose between `text`, `JSON`, or `XML`.

b. **Payload template**: Specify the action DTO variables to include in the webhook payload.

   In the example, the payload template value is `$action.details`. This variable is a reference to a field in the `ActionApiDTO` object that represents the current action. Your template can reference any of the fields in this DTO, starting with `action` as the object name. For example, `$action.createTime` gives you the time the action was created. For a full listing of the `ActionApiDTO` object, see API Reference *(on page 1133)* or the API Swagger UI.

   For pre-defined request payload examples, see the Workload Optimization Manager Orchestration GitHub repository.

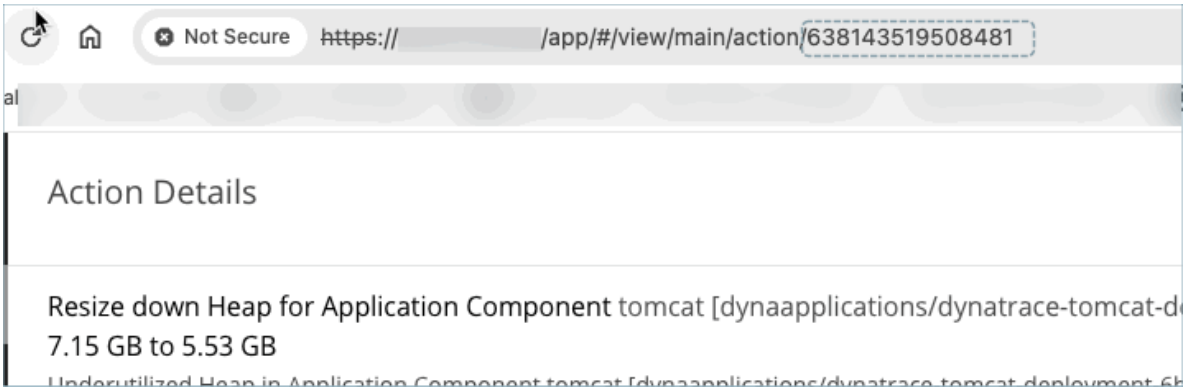
c. **Action UUID:** To test your webhook, specify the UUID value for a Workload Optimization Manager action. To find a suitable action UUID, complete the following steps:
   i. Select **Action** from the main navigation menu.
   ii. Choose an action from the list and click the corresponding **Details** button.



iii. Click the arrow icon ( ) to open the action in a new tab.

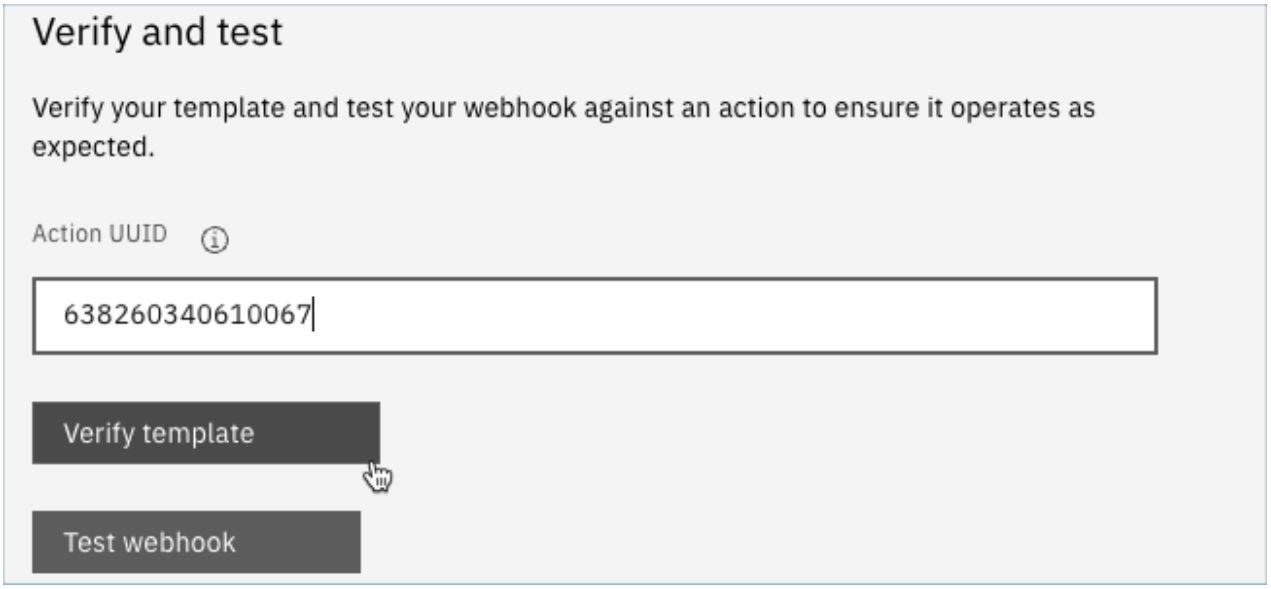


iv. Copy the action UUID from the end of the Workload Optimization Manager URL.



Specify this UUID in the **Action UUID** field for the request body.

d. Optionally, verify the payload template and test the webhook.



- To verify the syntax of your payload template, click **Verify template**. If the template is valid, a checkmark appears and the request payload is displayed. If the template is not valid, an error message appears.
- To test the end-to-end configuration of your webhook, click **Test webhook**. If the configuration is valid and the response that is received from the webhook server is successful, a checkmark appears. If the configuration is not valid, or the webhook server response was not successful, both a plus sign (+) and an error message appear.

e. When you are satisfied with your webhook configuration, click **Create**.

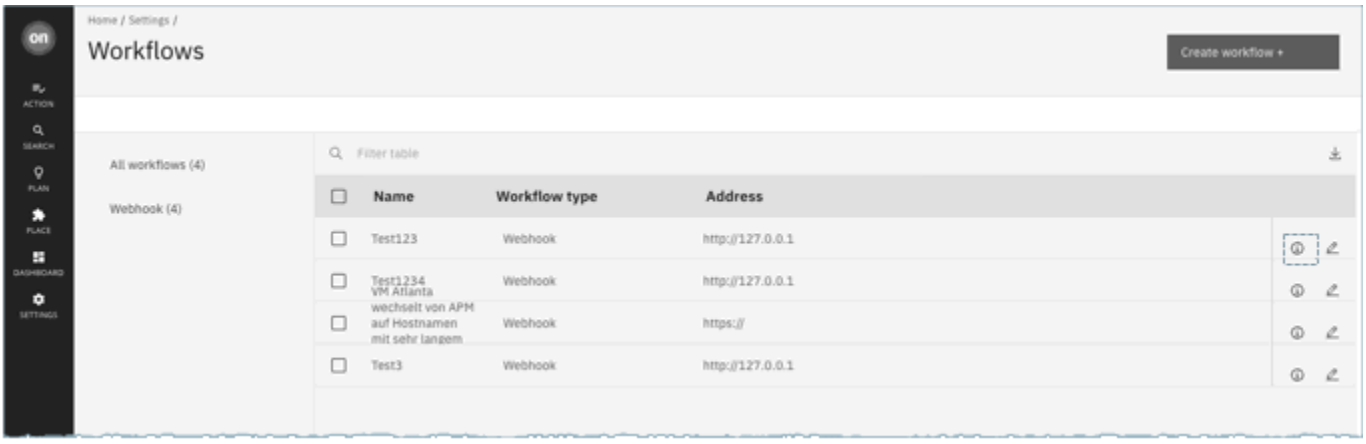
8. Create an Automation Policy that uses the webhook.

   For information about creating policies that use webhook workflows for orchestration, see Automation Workflow *(on page 1063)*.

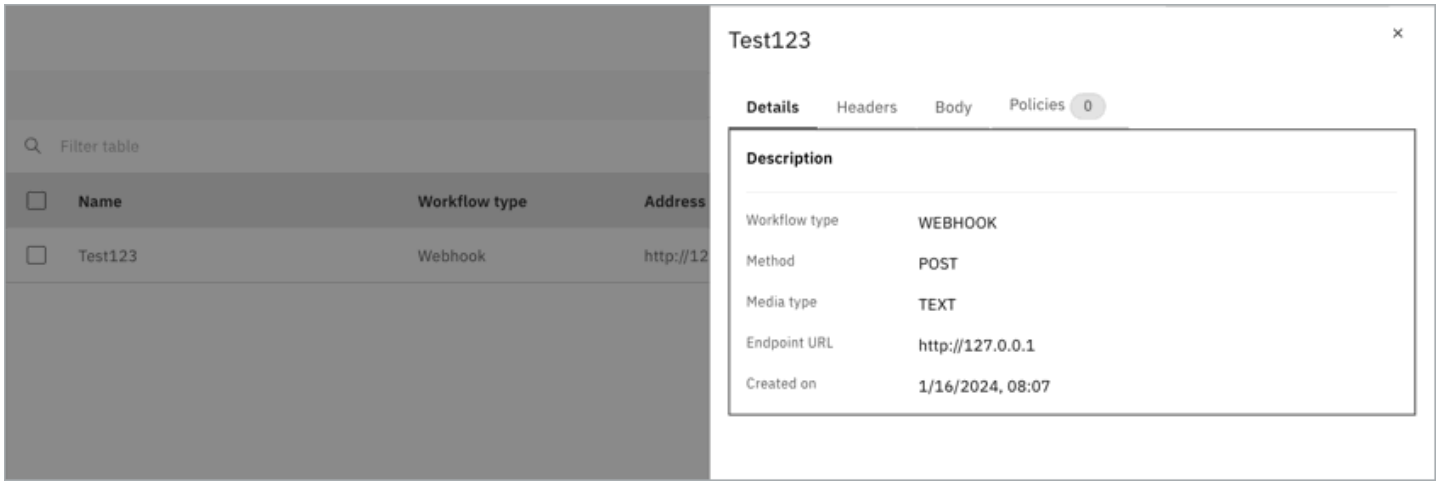
Reviewing webhook details and validation

After you create a webhook, you can quickly review the configuration details from the main **Workflows** page.

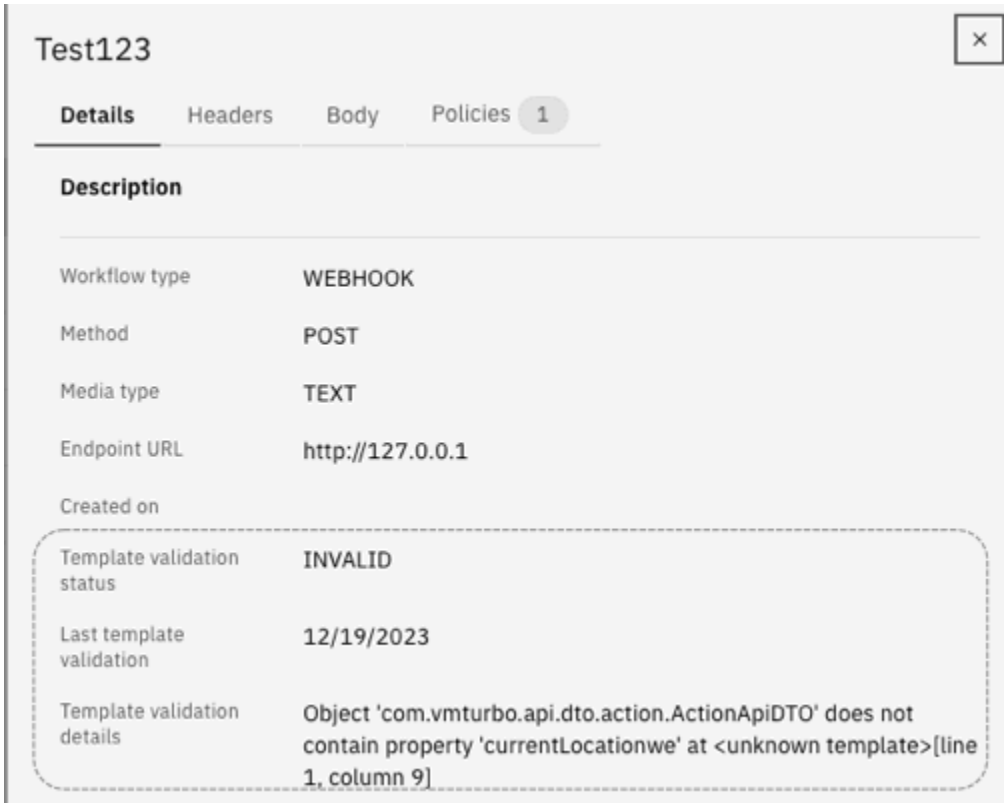
Click the information icon (i) for the webhook that you want to review.



A side panel opens that displays details about the webhook. The panel includes tabs for the headers, request body, and any policies that are associated with the webhook.



If the webhook is invalid for any reason, the panel shows an `INVALID` template validation status and provides details about the validation failure.



Sample webhook application

A webhook workflow sends a message to an application through HTTP/S. You express the message as a template that can include values from the action data in its request body. This template can express text, JSON, or any other format that your application can accept.

You can use webhooks to send messages to various existing applications, including Slack, Amazon Web Services, and others.

To deploy a simple example, and to test your webhook templates, implement a node.js server that receives the webhook message and prints the template data. If you install this server on a machine in your network, then you can give its URL in the webhook workflow, and test your response to specific actions.

The following example is a listing for a node.js web server.

```
let port = 9090;
const http = require("http");
console.log(`Starting server on port ${port}`);

http.createServer((request, response) => {
    request.setEncoding('utf8');
    console.log('REQUEST METHOD: ', request.method);

    let datStr = '';
    request.on('data', chunk => {datStr = datStr + chunk});
    request.on('end', () => {console.log('End of DATA: ', datStr)})
}).listen(port);
```

When you run this program, it prints a message to the console to say that it is running, and to identify the port it listens on.

When the server receives a message, it prints the request method, and then prints the message request body, as specified in the workflow's `template` field.

If you configured an Automation Policy to use this workflow, this server logs a message for each action that Workload Optimization Manager runs on an entity within the policy's scope.

Parking policies

A parking policy is a set of rules that Workload Optimization Manager must satisfy when enforcing parking actions (stop and start) on public cloud entities.

Supported entities include VMs, Virtual Machine Specs (AWS Auto Scaling groups and Azure Virtual Machine Scale Sets), Database Servers (AWS RDS), and Databases (Azure dedicated SQL pools). Click **Park** on the main menu to see a list of parkable entities.

Parking policies take precedence over:

- Parking actions that were manually executed from the Parking page *(on page 945)*.
- Parking schedules that were attached to parkable entities from the Parking page.

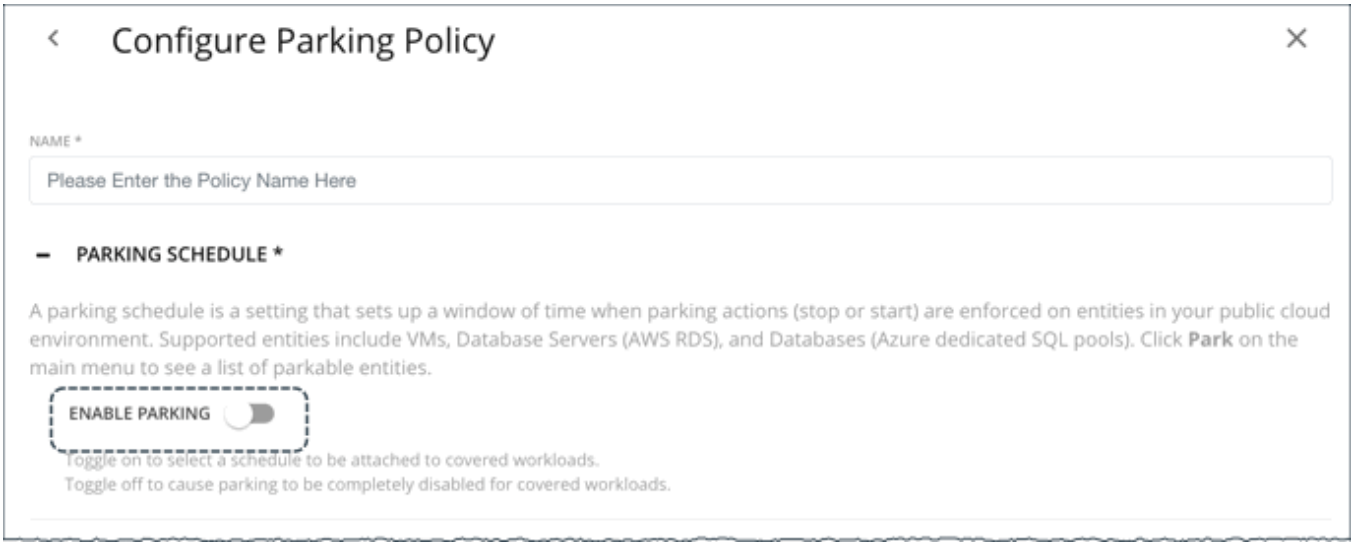
For example, if a parking policy dictates that a VM must be started everyday from 9:00 AM to 6:00 PM, and a Workload Optimization Manager user manually stopped the VM from the Parking page at 3:00 PM, the VM will be started to comply with the parking policy.

There is no default policy for parkable entities. You must create policies to enforce parking actions. In addition, parking policies do not require you to set the scope to specific entities. Entities are matched to policies based on their membership to predefined groups – Global, Cloud Providers, Accounts, and Resource Groups (Azure only). For example, if a policy applies to a specific account, all parkable entities in that account apply the policy, including new entities added in the future.

Prioritization

When an entity matches multiple policies, any resulting conflicts are resolved using prioritization. Parking policies are prioritized in the following ways.

1. Priority by action enforcement



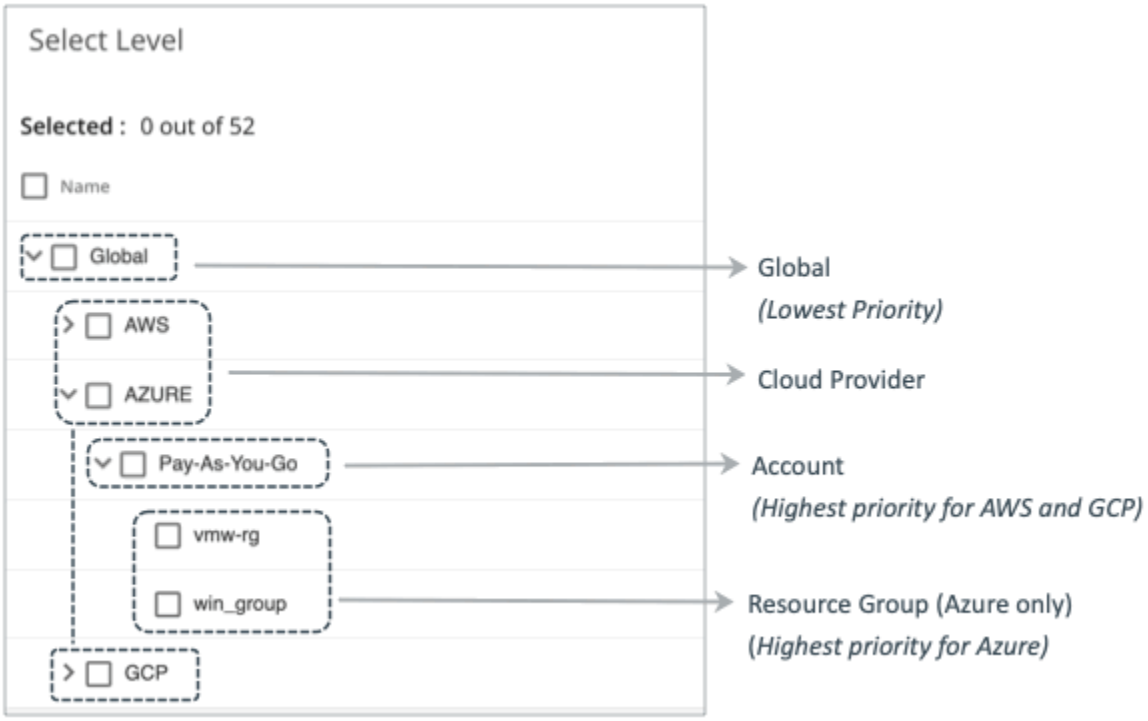
   You can enable or disable the enforcement of parking actions in a policy. For example, you may have virtual machines in your production environment that must be running all the time. You can prevent the enforcement of parking actions by creating a parking policy that disables parking actions.

   A parking policy that disables parking actions has a higher priority than a policy that enables parking actions. When parking actions are disabled in a policy, any entity that applies the policy cannot be parked on demand (in the Parking page) or automatically using schedules. In the Parking page, the **Unparkable** column indicates if parking actions are currently disabled on an entity.

2. Priority by level

   Level represents the first of two tiers of prioritization for policy matching. A policy associated with a lower level (such as Account) has a greater priority than a policy associated with a higher level (such as Global). This mitigates the complexity of prioritizing a large set of policies in a single list.

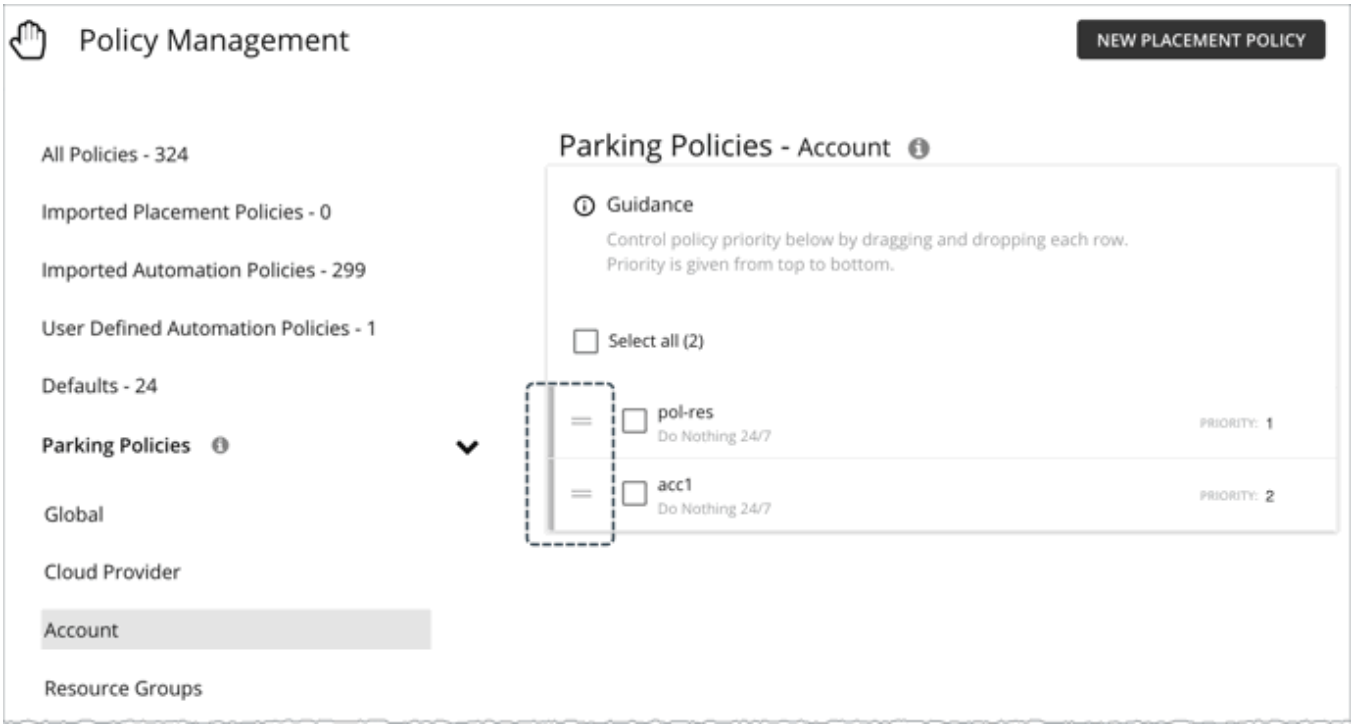
   Level is a required setting when creating parking policies.



Level	Notes
Global	Set a policy at the global level if you have multiple cloud providers and all parkable entities discovered from these providers must apply the same policy. This level has the lowest priority.
Cloud Provider ▪ AWS ▪ Azure ▪ Google Cloud	Set a policy at the cloud provider level if all parkable entities discovered from a specific provider must apply the same policy. This level has a higher priority than Global, but a lower priority than Account or Resource Group.
Account ▪ AWS accounts ▪ Azure subscriptions ▪ Google Cloud projects	Set a policy at the account level if all parkable entities discovered from specific accounts must apply the same policy. For AWS and Google Cloud, this level has the highest priority. For Azure, this level has a lower priority than Resource Group.
Resource Group – Azure only	Set a policy at the Resource Group level if all parkable entities discovered from specific Resource Groups must apply the same policy. This level has the highest priority for Azure. If you do not have Azure targets, this level will not display in the user interface, and you can start at the Account level.

3. Priority within each level

If you have created multiple parking policies for a given level, you can control the priority by rearranging policies via the drag-and-drop functionality in the user interface (highlighted in the following image). Priority is given from top to bottom. The current priority for each policy is indicated on the right.



Points to consider:

- A newly created policy appears last in the list. If you want the policy to have a higher priority, move it up the list.
- Changes to the order of policies take effect on the next policy evaluation cycle. Workload Optimization Manager runs policy evaluations every 10 minutes.
- If there are no entities applying the policy at the top of the list, the next policy will be applied to matching entities.

Example

Assume you created three policies, two at the cloud provider level and one at the global level.

Policy	Level	Policy Settings
A	Cloud Provider (first in the list)	Stop parkable entities Saturday, 1:00 PM - 5:00 PM (America_New_York)
B	Cloud Provider (second in the list)	Stop parkable entities Saturday, 2:00 PM - 3:00 PM (America_New_York)
C	Global	Start parkable entities Monday - Friday, 9:00 AM - 6:00 PM (America_New_York)

A parkable entity that matches all three policies will apply policy A instead of policy B, because policy A has a higher priority. The entity will also ignore Policy C in favor of Policy A. This means that the entity will be stopped every Saturday, between 1:00 PM and 5:00 PM. No parking policies are enforced outside this window.

Creating parking policies

1. Navigate to **Settings > Policies**.

   This opens the Policy Management Page, which lists all the currently available policies.
2. Click **New Policy > Parking Policy**.
3. Provide a meaningful name for the policy.
4. Enable or disable the enforcement of parking actions.
5. If a policy enables parking actions, you can attach a schedule to the policy so that actions can be enforced automatically according to the specified schedule.

   For instructions on creating parking schedules, see Managing Parking Schedules *(on page 1097)*.
6. Configure the level.

   For details about this setting, see Prioritization *(on page 1088)*.
   - Global – Select **Global**.
   - Cloud Provider – Select one or several cloud providers.
   - Account – Expand one or several cloud providers and then select the accounts that will apply the policy.
   - Resource Group (Azure only) – Under **Azure**, expand one or several accounts and then select the Resource Groups that will apply the policy.
7. Select one or several filters.

   Each policy can have filters applied on top of the chosen level. This gives you more fine-grained control over which schedule should apply to each entity. Filters do not the change the priority of a policy.

Filter	Expressions	Input
Name	equals (regex), not equals (regex)	Text
Tag	equals, not equals	Multi-text (tag name and tag value)
Guest OS	equals (regex), not equals (regex)	Text
Instance Type	equals, not equals	Multi-select
Region	equals, not equals	Multi-select

8. Click **Save Policy**.

Managing parking policies

You can perform any of the following tasks after creating policies.

- Drag and drop policies within the same level, as needed. Remember that priority is given from top to bottom.
- Delete a policy by selecting the policy and then clicking the trash bin icon.

Working with schedules

Schedules specify a period of time during which certain events can occur. Workload Optimization Manager uses the following types of schedules:

- Calendar schedule

  A calendar schedule is a setting that sets up a window of time when an automation policy *(on page 1053)* takes effect. This policy can execute non-parking actions *(on page 835)* on public cloud or on-prem entities, or change settings that affect analysis and action generation.
- Parking schedule

A parking schedule is a setting that sets up a window of time when parking actions (stop or start) are enforced on entities in your public cloud environment. Supported entities include VMs, Virtual Machine Specs (AWS Auto Scaling groups and Azure Virtual Machine Scale Sets), Database Servers (AWS RDS), and Databases (Azure dedicated SQL pools). Click **Park** on the main menu to see a list of parkable entities.

Managing calendar schedules

A calendar schedule is a setting that sets up a window of time when an automation policy *(on page 1053)* takes effect. This policy can execute non-parking actions *(on page 835)* on public cloud or on-prem entities, or change settings that affect analysis and action generation.

Creating calendar schedules

1. Navigate to **Settings > Schedule**.
2. In the **Calender** tab, click **New Schedule > Calendar Schedule**.

3. Type a name for the schedule.



4. Set the recurrence for the schedule.

   Choose whether the scheduled period occurs just once, or whether it repeats over time. The settings vary according to the recurrence you choose:

   - Does Not Recur

     This is a one-time schedule window. A non-recurring window has a start date, and no end date. The window starts on the day and time you specify, and remains open for the given duration.

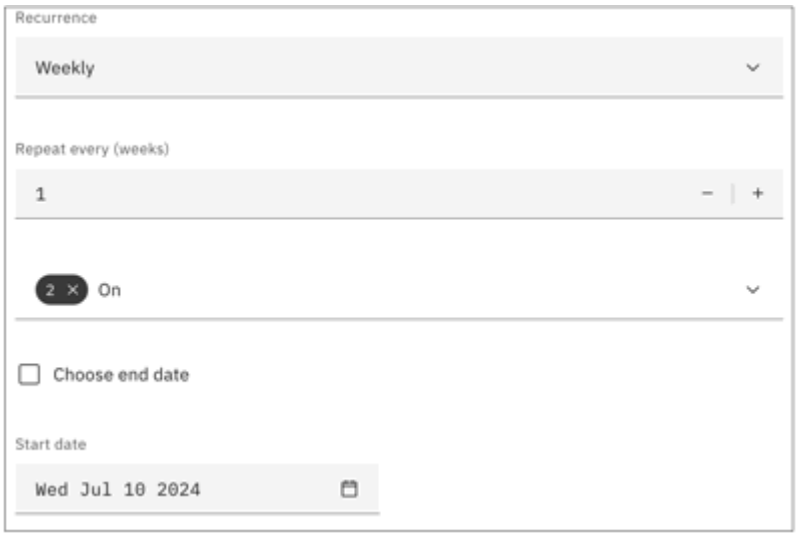
   - Daily

     Repeat this schedule every given number of days. For example, repeating 30 days is similar to repeating monthly, except it repeats by the count of days, not by the calendar month.

     The schedule begins on the **Start Date**, and continues repeating until the **End Date**. If **End Date** is "None", the schedule repeats perpetually.

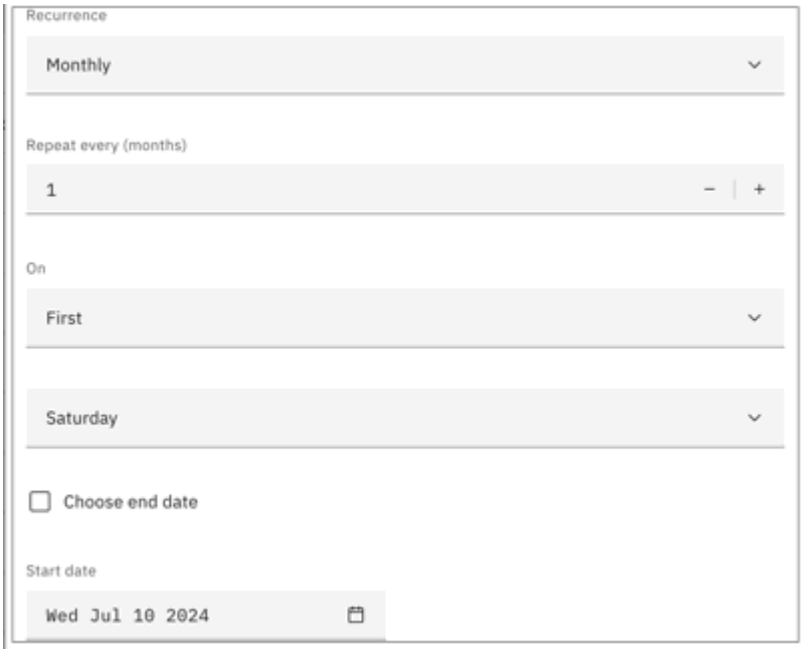
   - Weekly

     Repeat this schedule every given number of weeks, on the week days you specify. For example, to repeat every weekend, set it to repeat every one week on Saturday and Sunday.

     The schedule begins on the **Start Date**, and continues repeating until the **End Date**. If **End Date** is "None", the schedule repeats perpetually.

   - Monthly



Repeat this schedule every given number of months, to begin on a given day in the month. For example, you can schedule a maintenance window to begin on the first Saturday of each month.

The schedule begins on the **Start Date**, and continues repeating until the **End Date**. If **End Date** is "None", the schedule repeats perpetually.

5. Set the start time and duration.

   These settings specify how long the scheduled window remains open. You set the duration in terms of hours and minutes. Using a duration instead of an end time removes ambiguities such as starting before midnight and ending after. However, you should make sure the duration is not longer than the recurrence.

6. Set the time zone.

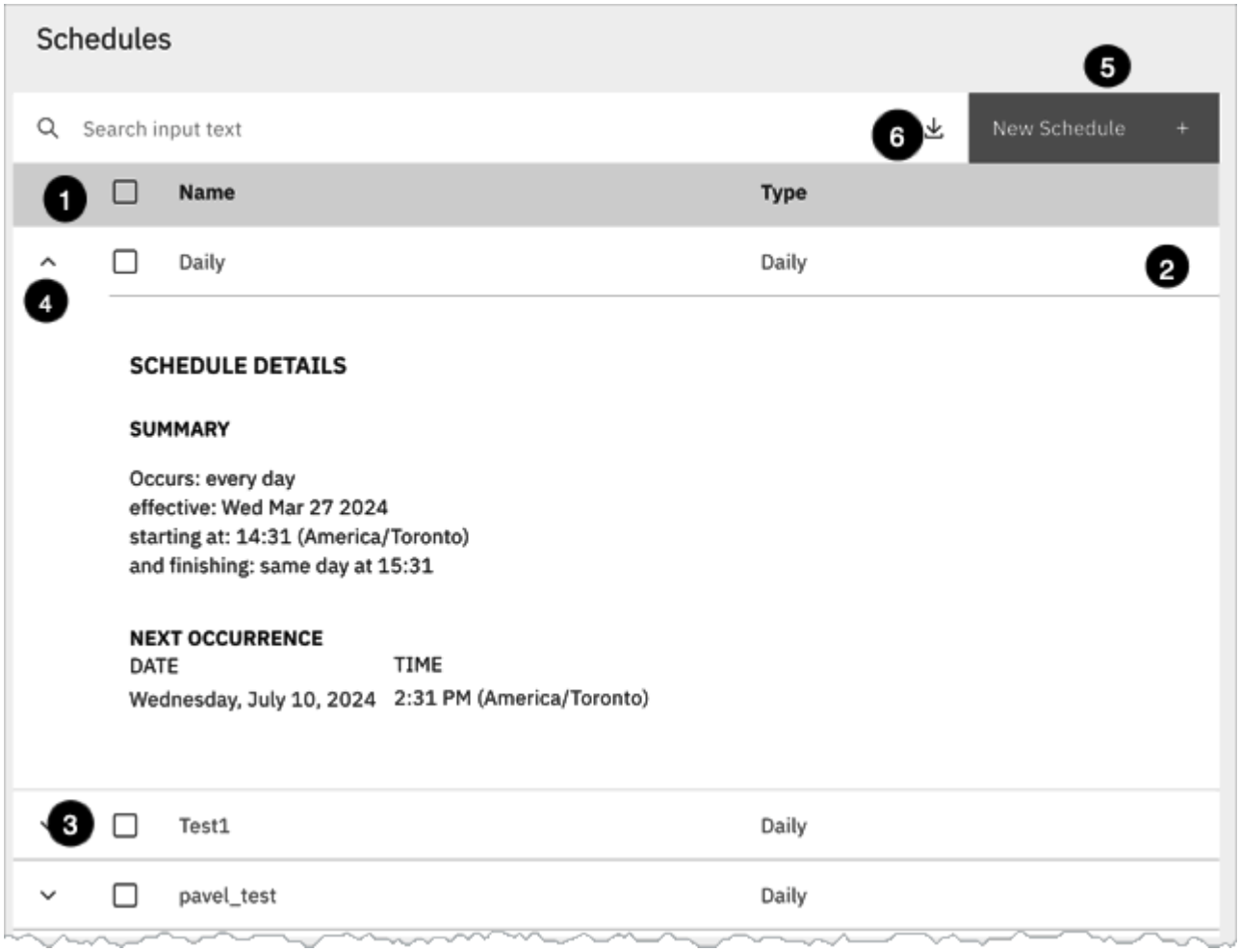
   This gives a reference for the schedule's start time. Workload Optimization Manager uses that reference when it opens and closes the schedule window.

   You see the same time zone setting no matter where you are located. Convert the schedule time to your local time to track when the schedule opens in your working day.

7. Save the schedule.

Viewing calendar schedules

To view calender schedules, navigate to **Settings > Schedule**.



The Schedules page lists all the currently defined schedules. From this page you can:

1. View list of schedules.
2. Click name to edit schedule.

   Click  to edit schedule.
3. Select an entry to delete the schedule.
4. Expand an entry to see schedule details.
5. Create a schedule.
6. Download a list of schedules.

Deleting calendar schedules

Before you delete a schedule, you should view its details to make sure no policies use it. If you delete a schedule that is in use by any policies, Workload Optimization Manager disables the affected policies until you edit them to either:

- Apply a different schedule to the policy and save the change
- Save the policy with no schedule

  Saving with no schedule confirms that you intend for this policy to apply at all times. Because scheduled policies are for special cases, this is usually not what you intend. For example, a scheduled maintenance window can have aggressive action acceptance modes that you do not want to enable during peak hours. If you save the policy with no schedule, then the aggressive settings will take effect at all times.

Workload Optimization Manager posts a confirmation dialog before deleting a schedule that is currently in use.

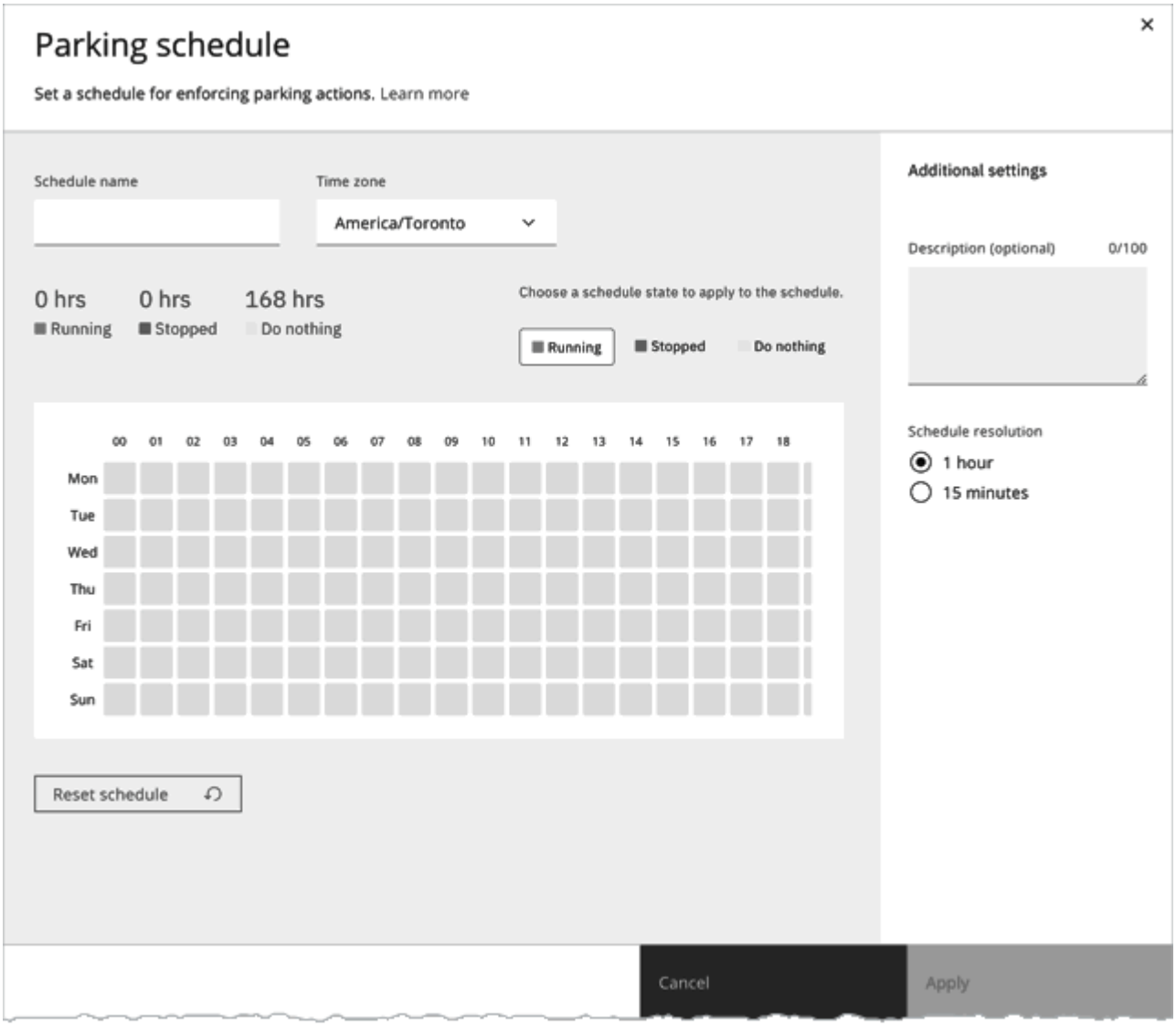
Managing parking schedules

A parking schedule is a setting that sets up a window of time when parking actions (stop or start) are enforced on entities in your public cloud environment. Supported entities include VMs, Virtual Machine Specs (AWS Auto Scaling groups and Azure Virtual Machine Scale Sets), Database Servers (AWS RDS), and Databases (Azure dedicated SQL pools). Click **Park** on the main menu to see a list of parkable entities.

Creating parking schedules

1. Navigate to **Settings > Schedule**.
2. In the **Parking** tab, click **New Schedule > Parking Schedule**.

   The Create Schedule page displays.



3. Provide a name and description.

   When attaching a parking schedule to one or more entities, you will see the *name* of the schedule. It is therefore important to use a name that will help you recognize each schedule and choose the appropriate one.
4. Set the time zone.

This gives a reference for the schedule's start time. Workload Optimization Manager uses that reference when it opens and closes the schedule window.

You see the same time zone setting no matter where you are located. Convert the schedule time to your local time to track when the schedule opens in your working day.

5. Set the schedule resolution.

   By default, each block in the schedule configuration section corresponds to an hour of the day. You can change this to 15-minute increments to give you more flexibility with scheduling. For example, with 15-minute increments, you can enforce parking actions between 8:15 and 20:45. This is not possible with the default schedule resolution.

6. Set a parking schedule.

   Parking schedules recur every week, on the days and times that you specify. These schedules include the following states that define the parking policies that you can enforce.

State	Policy
Running (green)	A parkable entity must be running on the selected days and times. Workload Optimization Manager will start the entity immediately if: ▪ The entity is not running at the beginning of the schedule. ▪ A user stops the entity via the cloud provider.
Stopped (red)	A parkable entity must not be running on the selected days and times. Workload Optimization Manager will stop the entity immediately if: ▪ The entity is running at the beginning of the schedule. ▪ A user starts the entity via the cloud provider.
Do Nothing (gray)	Do not enforce parking actions on the selected days and times. This is the default state for all the blocks.

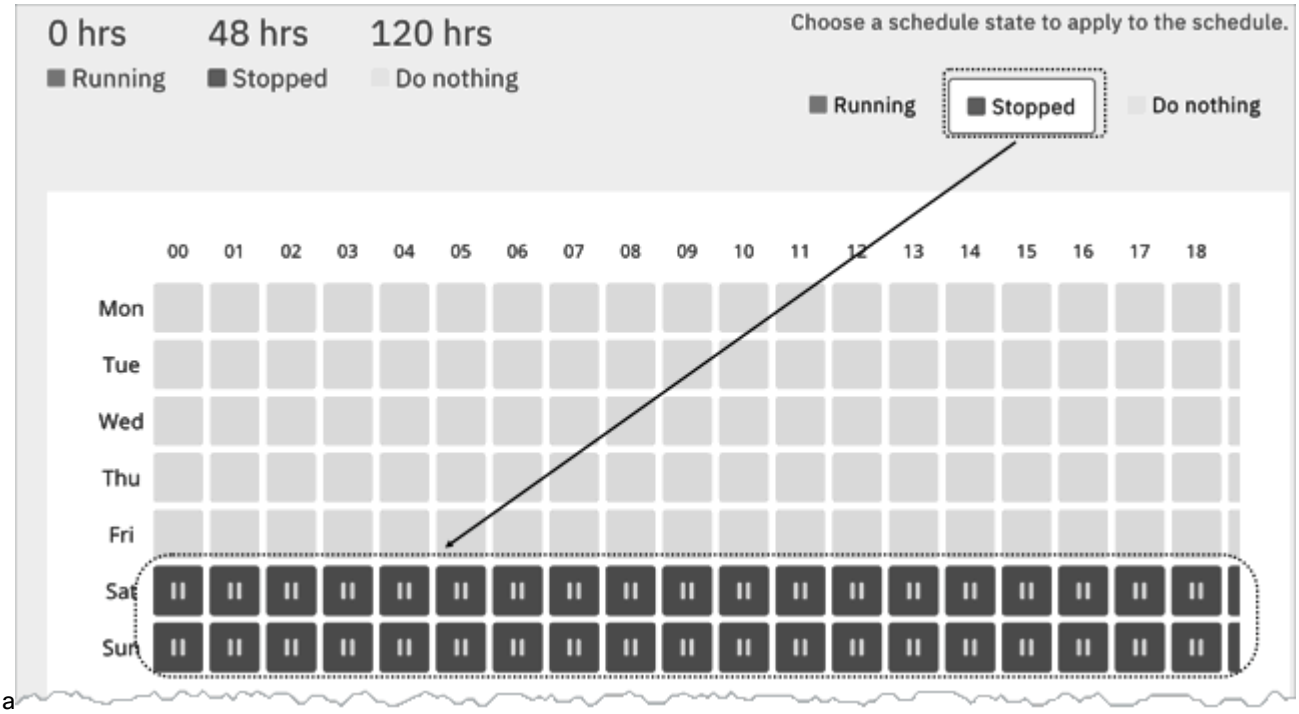
   Basic steps:

   a. Click either the **Running** (green) or **Stopped** (red) button.

   b. Click the blocks for the days and times the action will be enforced. To select multiple contiguous blocks, click and hold, and then drag in any direction.

   c. To undo your selections, click the **Do Nothing** (gray) button and then click the affected blocks.

In the following example, stop actions are enforced every Saturday and Sunday. No parking actions are enforced on all the other days of the week. The selected blocks are shaded red.



a

7. Click **Create**.

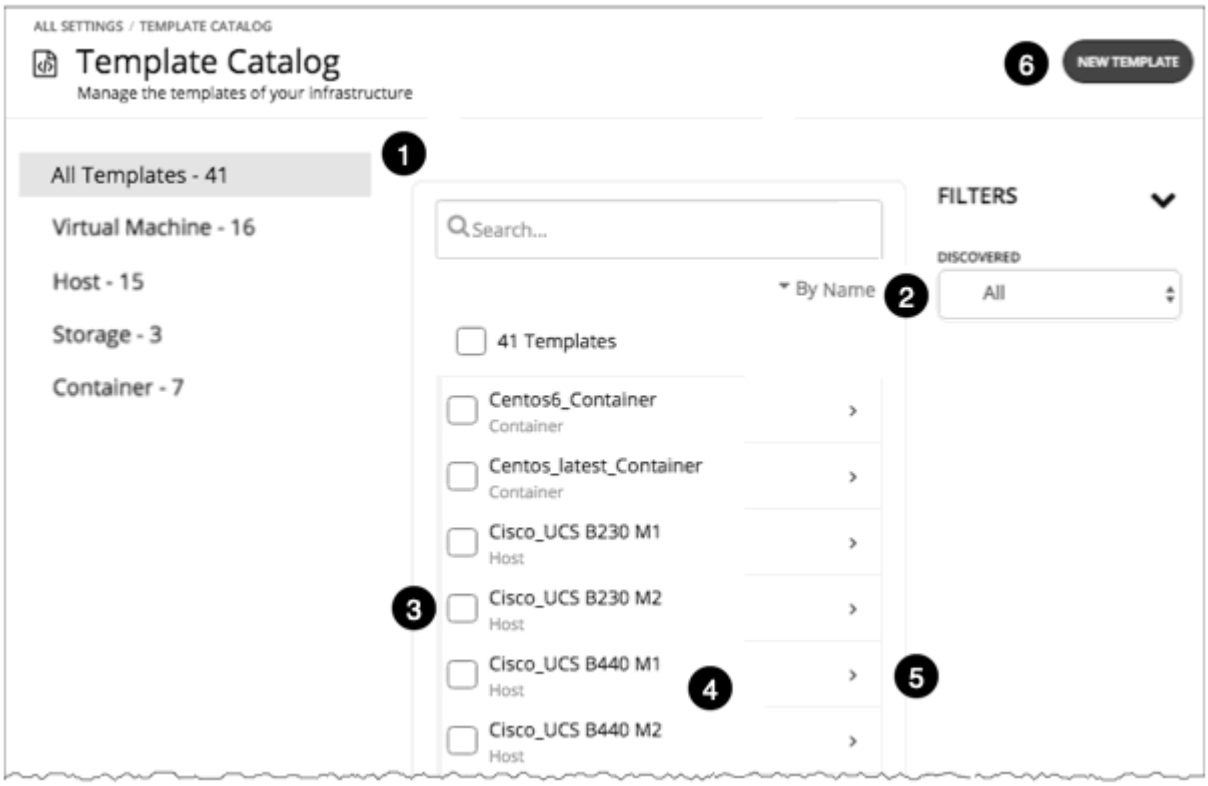
The schedule can now be attached to the parkable entities listed in the Parking page. For details about the Parking page, see Park: Stop or Start Cloud Resources *(on page 942)*.

Deleting parking schedules

Only unattached parking schedules can be deleted.

1. Click the three dots icon for the schedule.
2. Click **Delete**.

Templates: resource allocations for new entities



Workload Optimization Manager uses templates to describe new entities that it deploys in your environment or in plans. The templates specify resource allocations for these entities. For example, you can run a plan that adds new VMs to a cluster. If you add ten copies of a template, then the plan places ten new VMs that match the resource allocation you have specified for the template. For your cloud environment, you can see templates to match the instance types in your cloud accounts and subscriptions.

A VM template definition can include one or more images that Workload Optimization Manager uses to deploy the VM in your environment. The image identifies the actual deployment package, including a path to the physical files (for example an OVA).

As you deploy an instance of a VM template, Workload Optimization Manager chooses the best image for that instance.

The Template Catalog shows the templates that are specified or discovered for your installation of Workload Optimization Manager. From this page, you can:

1. Filter by type.
2. Filter to see only discovered templates.
3. Select templates to delete.
4. Click to edit the template.
5. Expand for details.
6. Create a template.

Creating templates

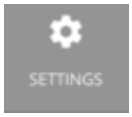
Templates specify the resources for entities that Workload Optimization Manager can deploy in your environment, or in plans.

A VM template definition can include one or more images that Workload Optimization Manager uses to deploy the VM in your environment. The image identifies the actual deployment package, including a path to the physical files (for example an OVA).

The Templates page shows all of the templates that have been specified or discovered for your installation of Workload Optimization Manager. From this page, you an create new templates, edit existing templates, filter the list of templates, and download the list of templates (New UI view only).

Creating templates

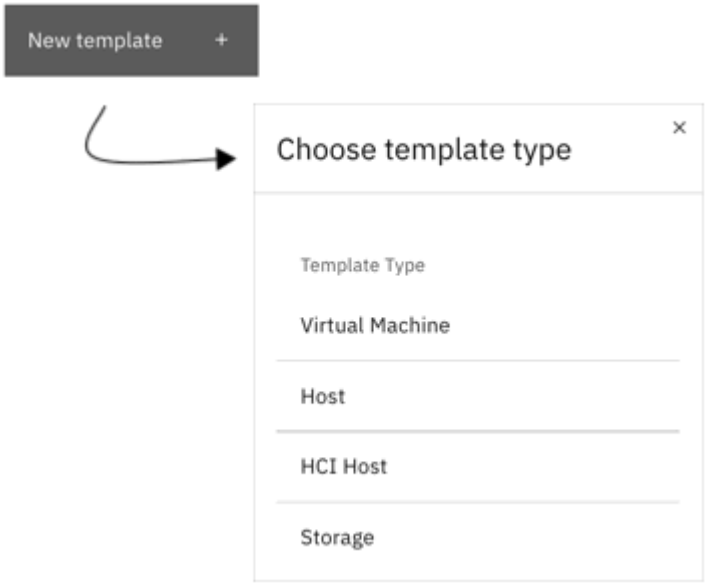
1. Navigate to the Settings page.



2. Choose Templates.



3. Click **New template**.
4. Choose the template type.



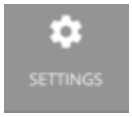
5. Specify the settings for your template.

   When the template window opens, it displays the most common resource settings. For each type of template, you set allocations for different resources.

6. Save your changes.

Editing templates

1. Navigate to the Settings page.



2. Choose Templates.



3. Select the template to edit.

   Hover over the template's name and click the pencil icon.
4. Edit any settings for your template.
5. Save your changes.

VM template settings

A VM template describes the resource allocation that you want to provide for a type of VMs. When Workload Optimization Manager deploys the associated VM to your environment or in a plan, it uses these values to determine the size of the VM. Workload Optimization Manager uses the Size settings to calculate the best placement for a VM of this type.

A VM template can optionally include an image description. When Workload Optimization Manager uses the template to deploy a VM to your environment, it uses the image to access the actual bits that install as the VM instance.

> **NOTE:**
> Workload Optimization Manager generates a special template called *headroomVM*, which it uses to calculate cluster headroom. The Template Catalog shows the template as editable, but you should not edit it because Workload Optimization Manager will overwrite your changes the next time it generates the template.

VM size

- CPU

  The virtual CPUs assigned to the VM. Specify the number of **Cores** and the **VCPU** clock speed – Workload Optimization Manager multiplies these values to calculate the host CPU resources it will allocate when placing the VM.

  The **Utilization** value sets the percentage of allocated CPU that the placed VM will consume. To ensure the host has left over resources for infrastructure tasks, you should assign less than 100%.
- Memory

  The amount of memory to allocate for the VM, in MB.

  The **Utilization** value sets the percentage of allocated memory that the placed VM will consume. To ensure the host has left over resources for infrastructure tasks, you should assign less than 100%.

  Note that you should never allocate less memory than is required for the VM's guest OS.
- Storage

  The storage resources to allocate for this VM.
  - **disk/rdm** – If you choose **rdm**, then the VM can use VMware Raw Device Mapping for its storage.
  - **IOPS** – The capacity for IO operations you give the VM for this datastore.
  - **Size** – The amount of storage capacity, in GB.

  The **Utilization** value sets the percentage of allocated memory that the placed VM will consume. To ensure the storage has left over resources for infrastructure tasks, you should assign less than 100%.

  Note that you can allocate multiple datastores to the VM.
- Network

  The amount of the host's network throughput to assign to the VM, in Mb/s.
- IO

  The amount of throughput on the host's IO bus to assign to the VM, in Mb/s

Host template settings

Host templates describe models of physical hosts that you can deploy in the on-prem data center. As part of capacity planning, you might want to see how to replace your current hosts with different models. To do that, you create templates to represent the hosts you want, and then use those templates when you run hardware replacement plans.

The host template is a collection of these settings:

- CPU

  The processor for this host model. CPU size and speed are not the only factors to determine processing power. Specify the host CPU in the following ways:

  - Select from Catalog

    When you enable **Select from Catalog**, you can open up a catalog of CPU models that Workload Optimization Manager uses to map the model to an effective capacity for the CPU.

  - Cores and CPU Speed

    When you disable **Select from Catalog**, you can specify the number of **Cores** and the **CPU** clock speed – Workload Optimization Manager multiplies these values to calculate the host CPU resources.

- Memory

  The amount of memory to allocate for the VM, in GB.

- Network

  The host's network throughput, in MB/s.

- IO

  The host's IO bus throughput, in MB/s

- Price

  If you know the price of the host model that you're specifying for the template, you can enter it here. When a plan runs, Workload Optimization Manager uses the price to calculate costs or savings when it adds or removes host machines in an on-prem data center.

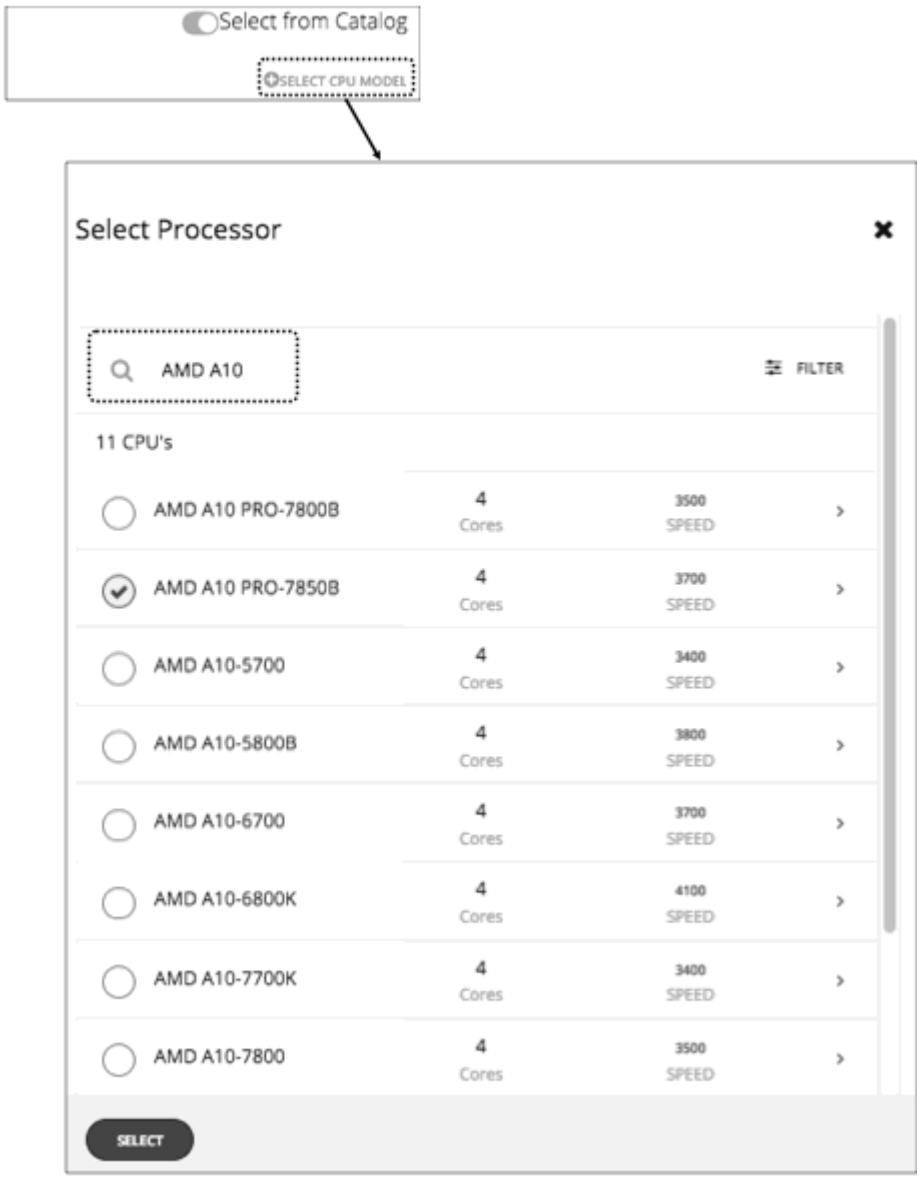
Selecting CPUs from the catalog

CPU processor speed is not necessarily an effective indicator of CPU capacity. For example, processor architecture can make a slower CPU have a greater effective capacity. Newer models of machines can often have fewer cores or less clock speed, but still have a higher effective capacity. Planning is affected in two ways:

- When planning hardware replacement, the plan knows the template's effective capacity. This means that the plan knows how to best place workloads on the new hardware.
- For deployed hosts, Workload Optimization Manager discovers the effective capacity and uses that information when it calculates workload placement.

To build the catalog of CPU capacity, Workload Optimization Manager uses benchmark data from spec.org. When you set up the CPU for a host template, you can search this catalog for the processor you want, and set it to the template.

> **NOTE:**
> Workload Optimization Manager also uses the effective processor capacity when it calculates workload placement in real-time. For more information, see Effective CPU Capacity *(on page 845)*.



HCI host template settings

HCI host templates describe models of physical hosts that support participation in a vSAN. Along with the host compute specifications, you also include specifications for storage capacity and redundancy (RAID level and failover). You can use these templates to plan for changes to your vSAN capacity.

NOTE:
For Hyper-V environments, if you run a Hardware Replace plan that replaces hosts with HCI Host templates, the results can be inconsistent or the plan can fail to place all the VMs in the plan scope. This typically occurs when Workload Optimization Manager detects a configuration issue with VMM or Hyper-V. As a result,Workload Optimization Manager treats the VMs as not controllable and does not attempt to place them.

The HCI Host template is a collection of these settings:

- CPU

  The processor for this host model. CPU size and speed are not the only factors to determine processing power. Specify the host CPU in the following ways:

  - Select from Catalog

    When you enable **Select from Catalog**, you can open up a catalog of CPU models that Workload Optimization Manager uses to map the model to an effective capacity for the CPU.

  - Cores and CPU Speed

    When you disable **Select from Catalog**, you can specify the number of **Cores** and the **CPU** clock speed – Workload Optimization Manager multiplies these values to calculate the host CPU resources.

- Memory

  The amount of memory to allocate for the VM, in GB.

- Network

  The host's network throughput, in MB/s.

- IO

  The host's IO bus throughput, in MB/s

- Storage

  The capacity for this storage.

  - **IOPS** – The effective IOPS capacity.
  - **Size** – Raw storage capacity, in GB. A plan that uses this template computes the effective storage capacity.

- Redundancy

  The redundancy method for this storage is on the virtualized SAN. This combines the RAID level and the number of host failures to tolerate.

- Price

  If you know the price of the host model that you're specifying for the template, you can enter it here. When a plan runs, Workload Optimization Manager uses the price to calculate costs or savings when adding or removing host machines in an on-prem data center.

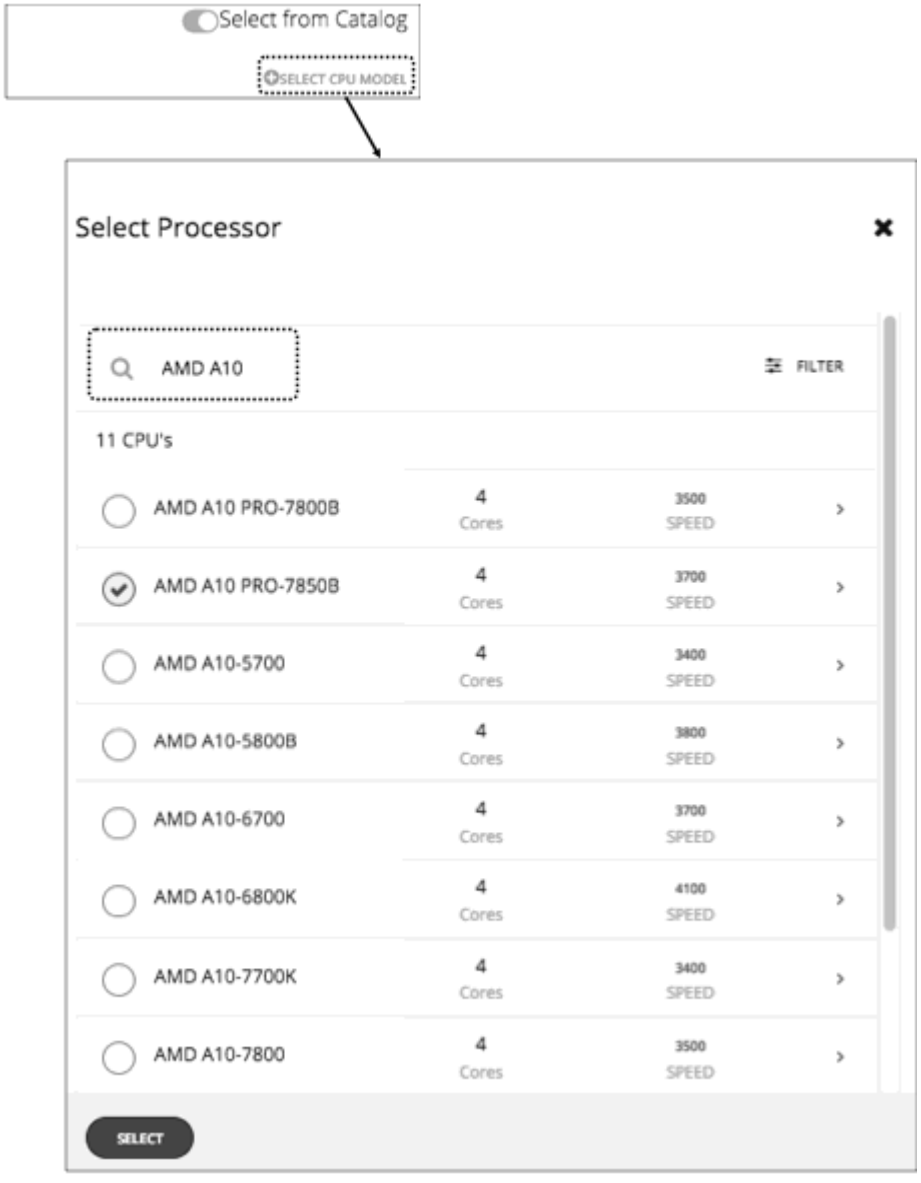
Selecting CPUs from the catalog

CPU processor speed is not necessarily an effective indicator of CPU capacity. For example, processor architecture can make a slower CPU have a greater effective capacity. Newer models of machines can often have fewer cores or less clock speed, but still have a higher effective capacity. Planning is affected in two ways:

- When planning hardware replacement, the plan knows the template's effective capacity. The plan knows how to best place workloads on the new hardware.
- For already deployed hosts, Workload Optimization Manager discovers the effective capacity and uses that information when it calculates workload placement.

To build the catalog of CPU capacity, Workload Optimization Manager uses benchmark data from spec.org. When you set up the CPU for a host template, you can search this catalog for the processor you want, and set it to the template.

NOTE:
Workload Optimization Manager also uses the effective processor capacity when it calculates workload placement in real-time. For more information, see Effective CPU Capacity *(on page 845)*.



Storage template settings

Storage templates describe models of storage that you can deploy in the on-prem datacenter. As part of capacity planning, you might want to see how to replace your current storage with different models. To do that, you create templates to represent the storage you want, and then use those templates when running hardware replacement plans.

The storage template is a collection of these settings:

- Storage

  The capacity for this storage.

  - **IOPS** – The capacity for IO operations on this storage.
  - **Size** – The amount of storage capacity, in GB.

- Price

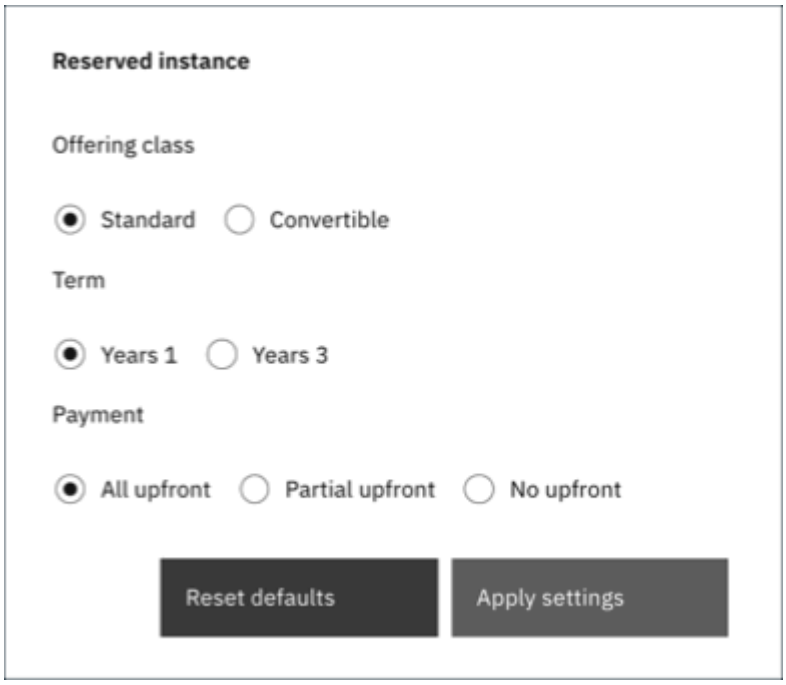
  If you know the price of the storage model that you're specifying for the template, you can enter it here. When running a plan, Workload Optimization Manager can use the price to calculate costs or savings when adding or removing storage in an on-prem datacenter.

Billing and costs

As you work with Workload Optimization Manager, you can set up costs that Workload Optimization Manager uses in its cost analysis.

AWS discounts – reserved instances

Configure a purchase profile. A purchase profile determines the costs that Workload Optimization Manager will use for all discount purchase decisions in your environment. As it sees opportunities to move workloads to another term, Workload Optimization Manager determines the costs based on the profile, and includes cost information in action descriptions. Workload Optimization Manager also uses this information to calculate projected changes in costs.



The following settings for your purchase profile apply to your global AWS environment.

- Offering class

  Choose the offering class that corresponds to the RI types that you typically use in your environment.

- Term

  Choose the payment terms for your discounts. Typically, longer term payment plans cost less per year.

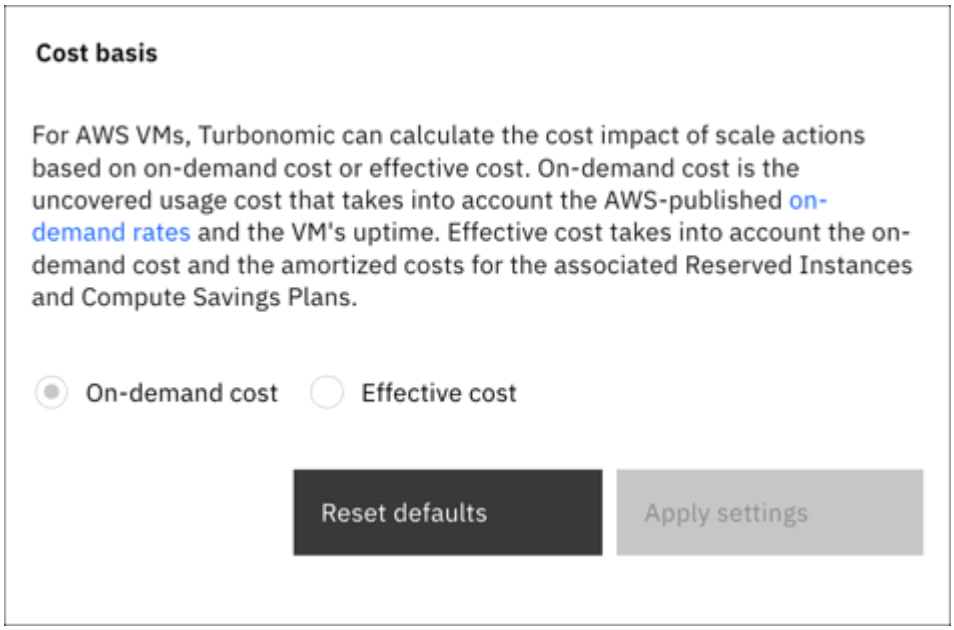
- Payment

Choose your payment option.

- All upfront – You make full payment at the start of the term.
- Partial upfront – You make a portion of the payment at the start of the term, with the remain cost paid at an hourly rate.
- No upfront – You pay at an hourly rate for the duration of the term.

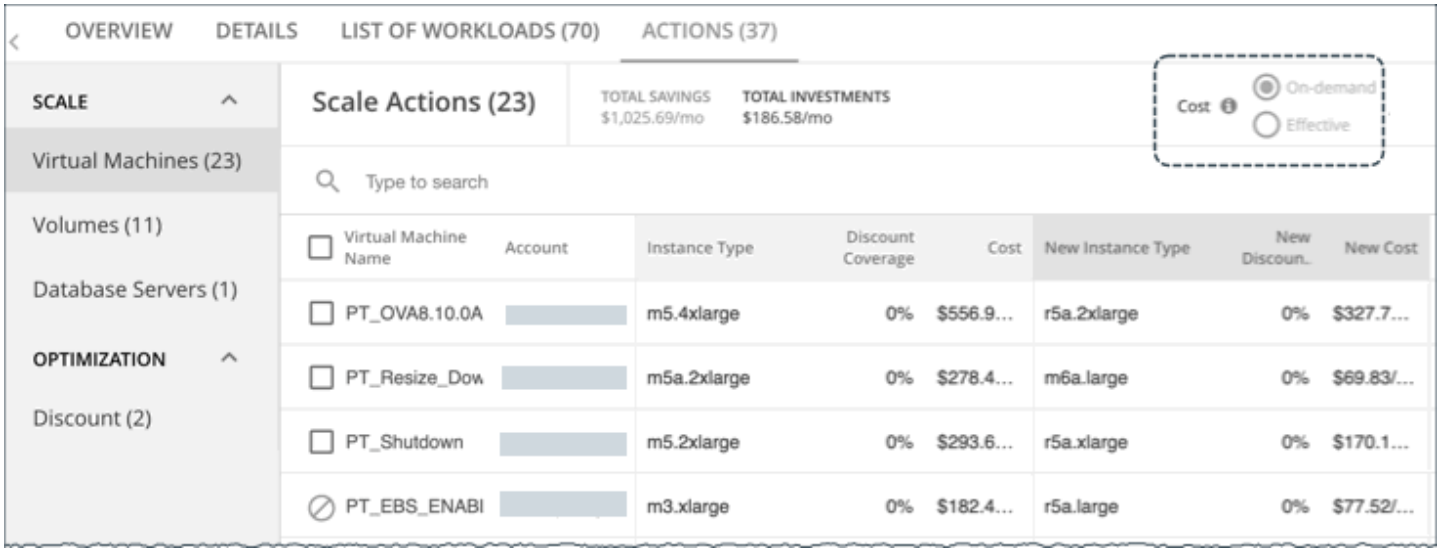
Cost basis

NOTE:
This setting is only available in the "new design" in the *User Guide* of the user interface. In addition, you must set up an AWS Billing target *(on page 389)* to configure this setting.



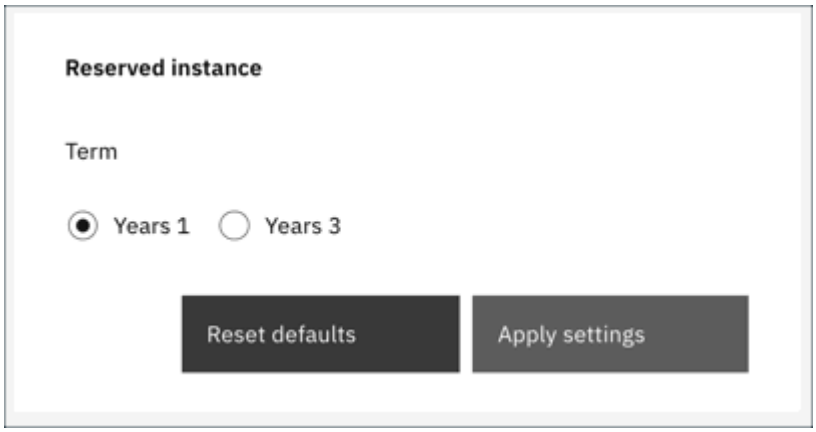
For AWS VMs, Workload Optimization Manager can calculate the cost impact of scale actions based on on-demand cost or effective cost. On-demand cost is the uncovered usage cost that takes into account the AWS-published on-demand rates and the VM's uptime. Effective cost takes into account the on-demand cost and the amortized costs for the associated Reserved Instances and Compute Savings Plans.

Your selection is automatically selected when you view scale actions in Action Center. You can switch selections to see how the calculations change.



Azure discounts

Workload Optimization Manager supports Azure reservations.

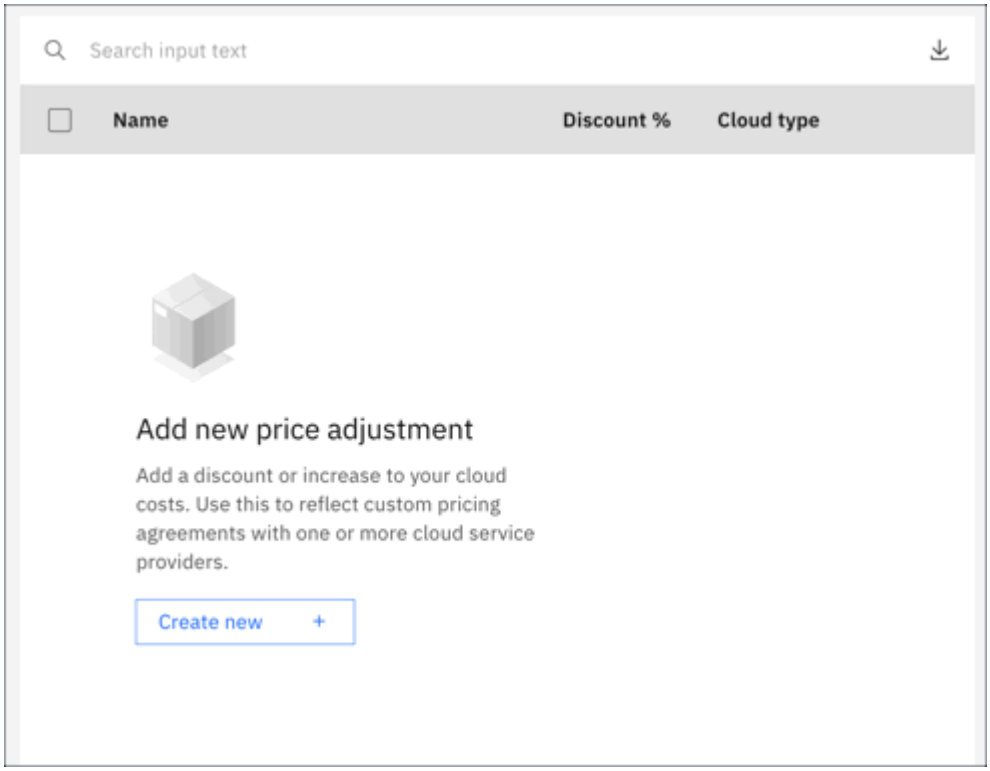


Choose the payment terms for your discounts. Typically, longer term payment plans cost less per year.

This setting applies to your global Azure environment and determines the costs that Workload Optimization Manager will use for all discount purchase decisions in your environment. As it sees opportunities to move workloads to another term, Workload Optimization Manager determines the costs based on the term, and includes cost information in action descriptions. Workload Optimization Manager also uses this information to calculate projected changes in costs.

Price adjustment

Cloud providers can offer their own price lists, including special costs for services or discounts for workloads. However, Workload Optimization Manager does not discover these adjustments. For example, to reflect any AWS discounted prices in Workload Optimization Manager, you must manually configure price adjustments. To configure price adjustments for a specific billing family, you must add the corresponding management account as an AWS target in the Target Configuration page.

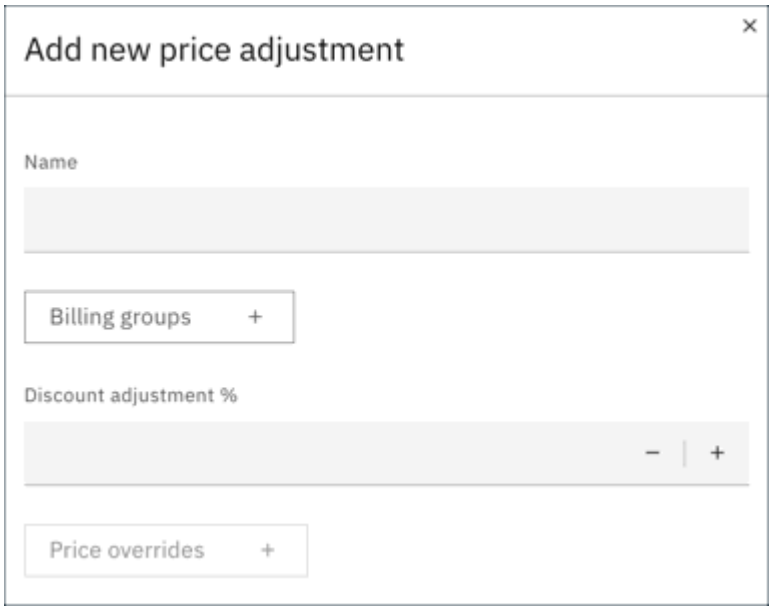


Workload Optimization Manager applies these price adjustments to:

- Costs for workload template families
- Costs for services, including bandwidth, VM licenses, AWS CloudWatch, and AWS DynamoDB

If price adjustments are not set, Workload Optimization Manager will show on-demand pricing, which could result in incorrect cost information in Workload Optimization Manager.

Configure the following settings:



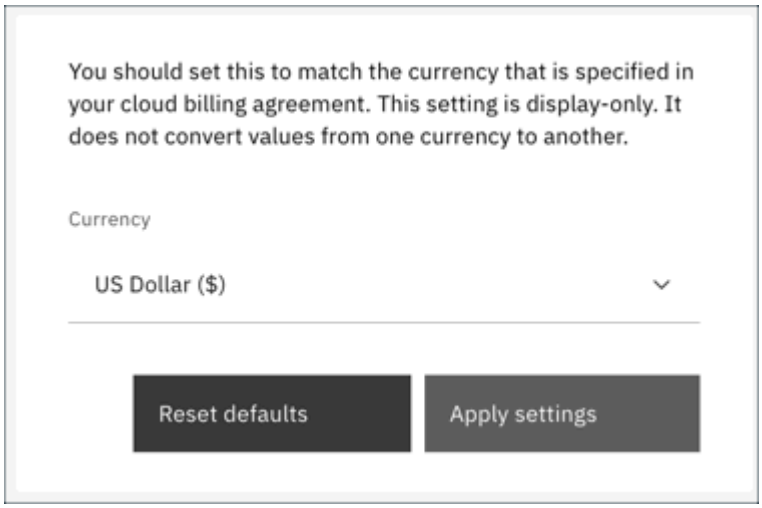
Setting	Instructions
Name	Specify a name.
Billing groups	Click **Billing groups** and then choose the billing group for the scope of this adjustment. A billing group is a set of AWS accounts that are consolidated into a single billing schedule. To consolidate billing, AWS supports billing families of AWS accounts, where there is a management account and member accounts. Workload Optimization Manager lists each billing family as a billing group. **NOTE:** If a particular billing family is missing in the list of billing groups, check if you have added the corresponding management account as an AWS target in the Target Configuration page. This account must be added in order for the billing family to display in the list.
Discount adjustment	Specify a percentage of adjustment between 0 and 99.99%. For any costs within the adjustment scope, Workload Optimization Manager will apply this percentage as it calculates the optimal workload capacity and placement. After you configure an adjustment, Workload Optimization Manager applies it to pricing in the selected scope.
Price overrides	The discount adjustment percentage you just specified applies as a default in the adjustment scope. However, you might have negotiated different prices for specific services or template families in your cloud environment. To configure these special prices, click **Price overrides**. For details, see AWS Price Override *(on page 1113)*.

Points to consider:

- Workload Optimization Manager uses the adjustments that you configure to display costs in the user interface. However, the values for hourly cost per entity, total hourly cost, total monthly cost, or total yearly cost can show inaccuracies on the order of a fraction of a percent. This is due to rounding when calculating the adjusted cost per entity.
- In AWS environments, Workload Optimization Manager does not apply any discounts or other price adjustments to spot compute costs.
- To use a price adjustment with a given billing group, you must increase the memory allocated to the VM that hosts your Workload Optimization Manager instance. Workload Optimization Manager requires that you provide a minimum amount of memory when you install the product. To use price adjustments, Cisco recommends that you increase the allocated memory as follows:
  - For the first price adjustment assigned to one or more billing groups, increase by 4 GB.
  - For each subsequent price adjustment assigned to one or more billing groups, increase by an additional 1 GB.
- Whenever you add, edit, or remove a price adjustment that is in use, you must allow sufficient time for Workload Optimization Manager to fully discover all of the affected environment, and to propagate the changes throughout that environment. In an average environment, this can take up to 30 minutes. As an alternative, you can manually execute rediscovery for the affected cloud subscription or account.

Currency

By default, Workload Optimization Manager uses the dollar symbol ($) when displaying the costs and savings that it discovers or calculates for your cloud workloads. You can set a different symbol to match your preferred currency. For example, if your cloud provider bills you in euros, change the currency symbol to €.

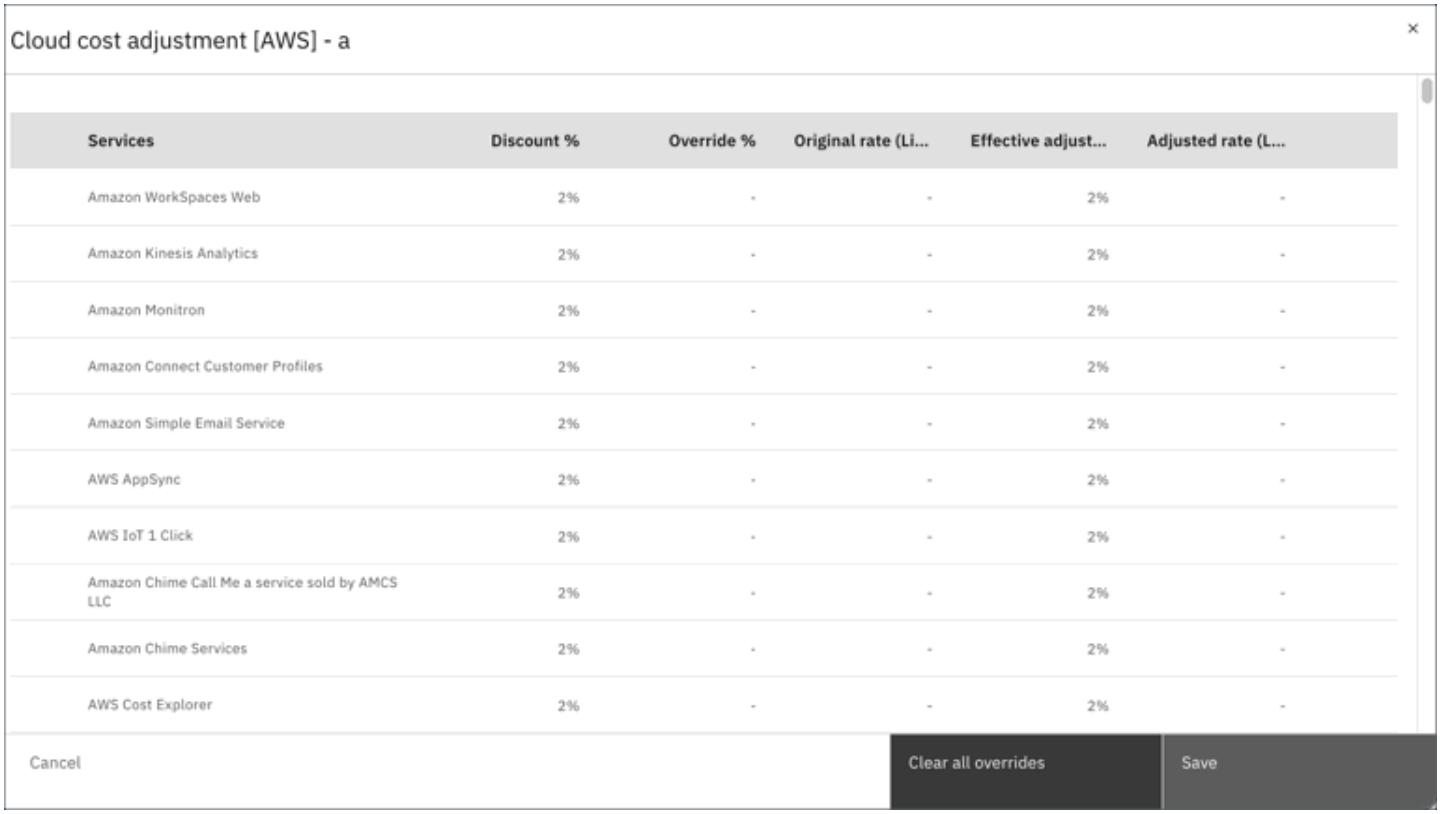


Workload Optimization Manager saves your preference in the local storage of the browser that you used to access the user interface. It reverts to the default symbol if you use another browser or view the user interface in incognito/private mode.

Currency symbols are for display purposes only. Workload Optimization Manager does not convert monetary amounts when you switch symbols.

> **NOTE:**
> For Azure reservations, Workload Optimization Manager can discover costs in non-US dollar currencies. For more information, see this topic *(on page 478)*.

AWS price override



To override the price adjustment for AWS Billing groups, Workload Optimization Manager analysis can use settings for different services that AWS provides to your accounts.

In AWS, you can set up a billing family that includes a management account and a given set of member accounts. Workload Optimization Manager treats the AWS Billing family as a Billing Group. For more information about billing families and accounts, see AWS Billing Families *(on page 1114)*.

Assume you have configured a price adjustment with a discount of 10% for a billing family, to match the overall discount that AWS offers you for that scope. But then assume the account includes extra discounts for some of the services your billing families provide. Then you can create overrides to add the extra discounts to those services.

Workload Optimization Manager uses the adjusted costs in its analysis as it calculates actions. For example, assume a price adjustment of 10% for a billing group, and a discount of 20% for the M4.Large family of templates. As Workload Optimization Manager places a workload, it will consider both the template capacity and the template cost. Even if an M4 template is larger than the workload actually needs, the M4 template could be less expensive because of the added discount. In that case, Workload Optimization Manager will place the workload on the less expensive template.

NOTE:

The Cloud Cost Adjustment table lists the services that are available to you for the AWS Billing family that you have set up as the discount scope. The services this table displays depend on whether the billing family uses the given service, and whether there is any recorded cost at the time that you display the table. For this reason, under some circumstances you might see different services listed in the table.

Under all circumstances, the table lists the services, AWS EC2 Compute, AWS EC2 Reserved Instance, and AWS RDS.

Also, for the Cloud Cost Adjustment table to display CSP Cost and Effective Cost, you must have created a data export in AWS, and you must store it in an S3 bucket.

In the Cloud Cost Adjustment table, you can perform the following:

- Override the price adjustment for a service or template family.

  To add an override, choose the line item for a service, or expand the row for a template family and:

  - Set the Type. Double-click and then choose **Discount** or **Increase**. Press **Enter** to confirm your setting.
  - Specify the percentage for this override, and then press **Enter** to confirm your override. The value you enter here is an absolute value for the discount or increase Workload Optimization Manager will apply for this line item.

  When you're done setting these overrides, click **Save**.
- To remove all overrides and revert back to the PRICE ADJUSTMENT Discount, click **Clear all overrides**.
- To download a report of the discounts for each service, click **Download** and choose CSV or PDF.

The table lists the following information about your discounts:

- Services

  The different cloud services to which you can set an override discount. To see individual workload templates:

  - For Azure, expand **Virtual Machines**
  - For AWS, expand **AWS EC2 Compute** or **EC2 Reserved Instance**
- Type

  Whether this price adjustment will be an increase or a discount. By default, this field shows the setting that you have made for the Price Adjustment. However, you can change it as an override for an individual entry.
- Price adjustment

  The percentage that you have specified for the Price Adjustment setting. This is the general adjustment that Workload Optimization Manager applies by default to the given service.
- Override

  If you have entered a value, this is the price adjustment Workload Optimization Manager applies to the given service.
- Original rate (Linux)

  The Cloud Service Provider's cost for VM templates, per hour. To see these costs, expand the workload services to show specific templates. The cost assumes no charge for the OS license, as though the VM runs Linux.
- Effective adjustment

  The actual adjustment for the given service.
- Adjusted rate (Linux)

  The discounted cost for VM templates, per hour. To see these costs, expand **Virtual Machines** to show specific templates. The cost assumes no charge for the OS license, as though the VM runs Linux.

AWS billing family recognition

As you configure AWS targets, Workload Optimization Manager discovers AWS accounts that are consolidated into billing families. A billing family has one management account, and zero or more member accounts. By recognizing billing families, Workload Optimization Manager more accurately calculates cloud investments and savings, and makes more accurate recommendations for RI coverage.

For RI purchases, different accounts in a billing family can share the same RI resources. At the same time, accounts in other billing families cannot use those RIs. This adds flexibility to your RI coverage, while maintaining order over the billing.

In Workload Optimization Manager, if you enable billing family recognition, then you can see the billing family management and member accounts in the user interface, and Workload Optimization Manager can recommend proper RI purchases within the correct billing families.

To enable billing family recognition, ensure the following as you configure your AWS targets:

- Use the proper role for each AWS target

  To properly discover billing family information for a target, you must give Workload Optimization Manager credentials for an AWS role that includes the permission, `organizations:DescribeOrganization`. With that permission, Workload Optimization Manager can:

  - Discover management accounts and member accounts in different billing families
  - Display the account names in the user interface
  - Discover billing information for each family and account
  - Recommend RI actions that respect billing family boundaries

- Configure targets for the complete billing family

  One billing family can consolidate a number of AWS accounts. For Workload Optimization Manager to include these accounts in its analysis, you must configure each one as a separate target. If you do not configure all the accounts in a billing family, then Workload Optimization Manager cannot discover complete billing information for that family, and its analysis will be based on incomplete information.

  Workload Optimization Manager displays member accounts that have been configured as targets in regular text. For members that Workload Optimization Manager discovers but have not been configured as targets, Workload Optimization Manager displays their names in grayed text.

If you enabled billing family recognition, keep the following points in mind:

- Billing families can grow.

  Workload Optimization Manager regularly checks the membership of your billing families. If it discovers a new member account, it adds that account to the list of members. If you have already configured the account as a target, then Workload Optimization Manager includes the new member in its analysis of billing families. If the new member is not already a target, then Workload Optimization Manager lists the new member in grayed text.

- You can configure discounts per billing family.

  Workload Optimization Manager includes a feature to set a discount for a billing group, and to override that discount for specific template families within that scope. For more information, see AWS Discounts *(on page 1107)* and AWS Price Override *(on page 1113)*.

- You might see management accounts that have no member accounts

  AWS treats every account you create as a part of a billing family. Assume you created an account, but you had no reason to consolidate its billing with any other accounts. In that case, the account appears in the Workload Optimization Manager user interface as a management account, but it has no member accounts.

Administrative tasks

To perform Workload Optimization Manager administrative tasks, you will navigate to different pages from **Settings**.The different tasks you can perform for Workload Optimization Manager include:

- Managing User Accounts *(on page 1116)*

  Create and manage user accounts for Workload Optimization Manager.

- Maintenance: Logging, and Troubleshooting *(on page 1127)*

  Perform general tasks such as setting log levels or sending troubleshooting data to Technical Support.

- License Configuration *(on page 1129)*

  Review the status of your current license, and apply any license upgrades.

Managing user accounts

As an administrator, you specify accounts that grant users access to Workload Optimization Manager. Credentials for user accounts can either be saved locally on the Workload Optimization Manager server or managed externally through Microsoft Active Directory (AD) or another single sign-on (SSO) provider.

The user account configuration determines the following details:

- User authentication

  To configure an account, you set the type of authentication the account uses:

  - Local user: Configure the username and password and save those credentials on the Workload Optimization Manager server.
  - External user: Single user accounts can authenticate through SSO or through AD.
  - External group: Groups of user accounts can authenticate through SSO or AD.
- User authorization

  You can also specify properties that determine the range of access and features for a user:

  - Role: The user role determines access to specific Workload Optimization Manager features.
  - Scope: The user scope determines how much of the environment this user can manage.

As you configure user accounts, you can set up access to specific clusters in your environment. You can even set up accounts for tenant customers, and only show them the virtual workloads they own in their specific virtual datacenters.

IMPORTANT:
For self-hosted Workload Optimization Manager instances, you can configure Workload Optimization Manager to use SSO authentication. When SSO is enabled, Workload Optimization Manager permits logins only through the SSO IdP. Whenever you navigate to your Workload Optimization Manager installation, it redirects you to the SSO Identity Provider (IdP) for authentication before it displays the Workload Optimization Manager user interface.

Before you enable SSO for your Workload Optimization Manager installation, *you must configure at least one SSO user with Workload Optimization Manager administrator privileges*. If you do not, then after you enable SSO, you cannot configure any SSO users in Workload Optimization Manager. To authorize an SSO user as an administrator, use **External authentication** to do one of the following options:

- Configure a single SSO user with administrator authorization.

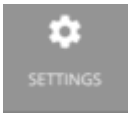
  Add an external user. The username must match an account that the IdP manages.
- Configure an SSO user group with administrator authorization.

  Add an external group. The group name must match a user group on the IdP, and that group must have at least one member.

For information about configuring SSO user groups in SAML, see Configuring a Group for SSO Authentication *(on page 1126)*. For information about configuring SSO authentication for Workload Optimization Manager, see Single Sign-On Authentication *(on page 100)*.

To work with Workload Optimization Manager accounts:

1. Navigate to the Settings page.

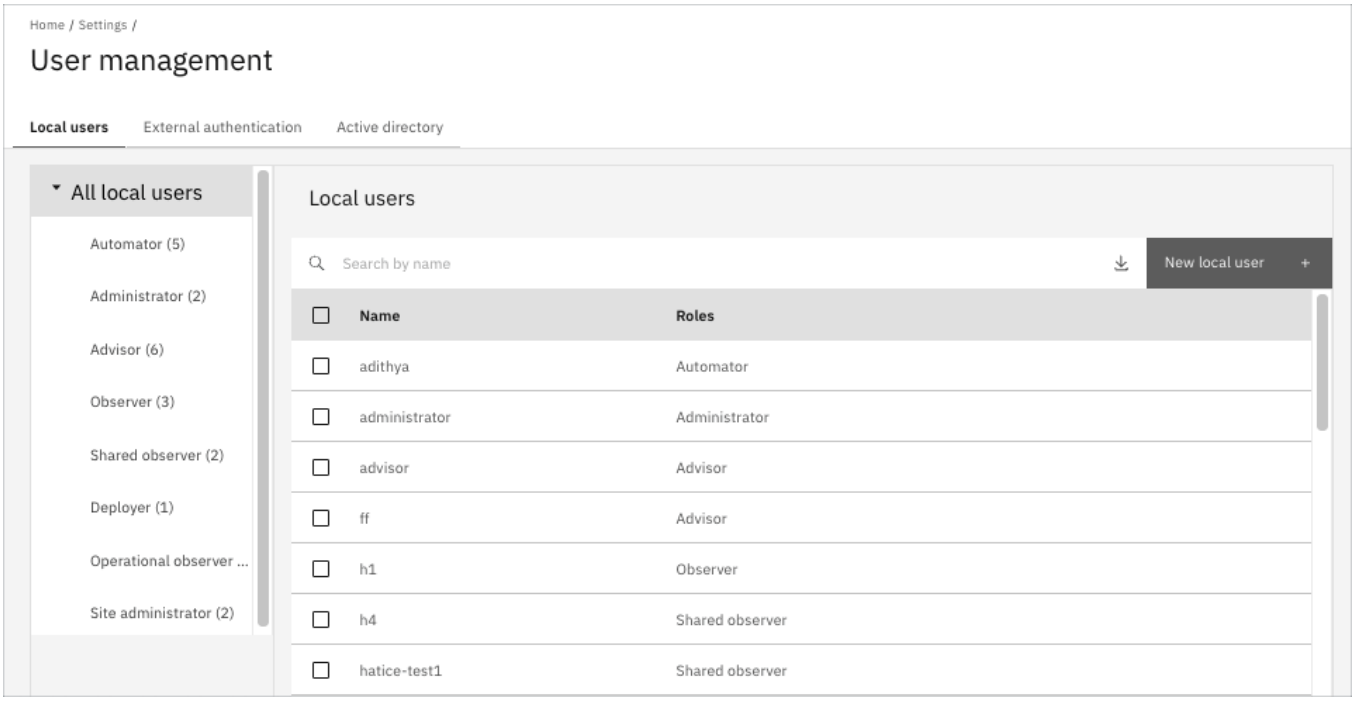
   

   Click to navigate to the Settings page. From there, you can perform a variety of Workload Optimization Manager configuration tasks.
2. Choose User Management.



User Management

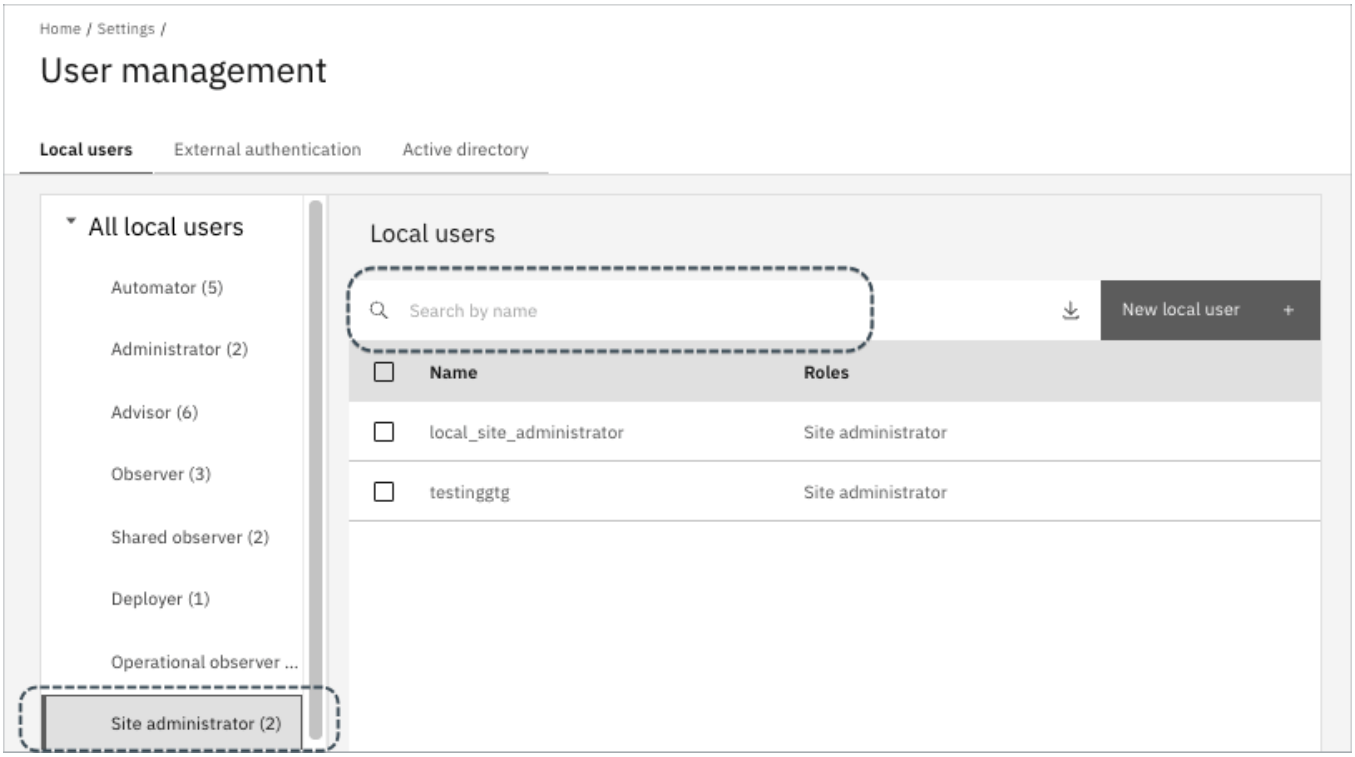
Click to navigate to the User management page.



This page lists all the user accounts that you currently have configured for Workload Optimization Manager. You can perform the following tasks:

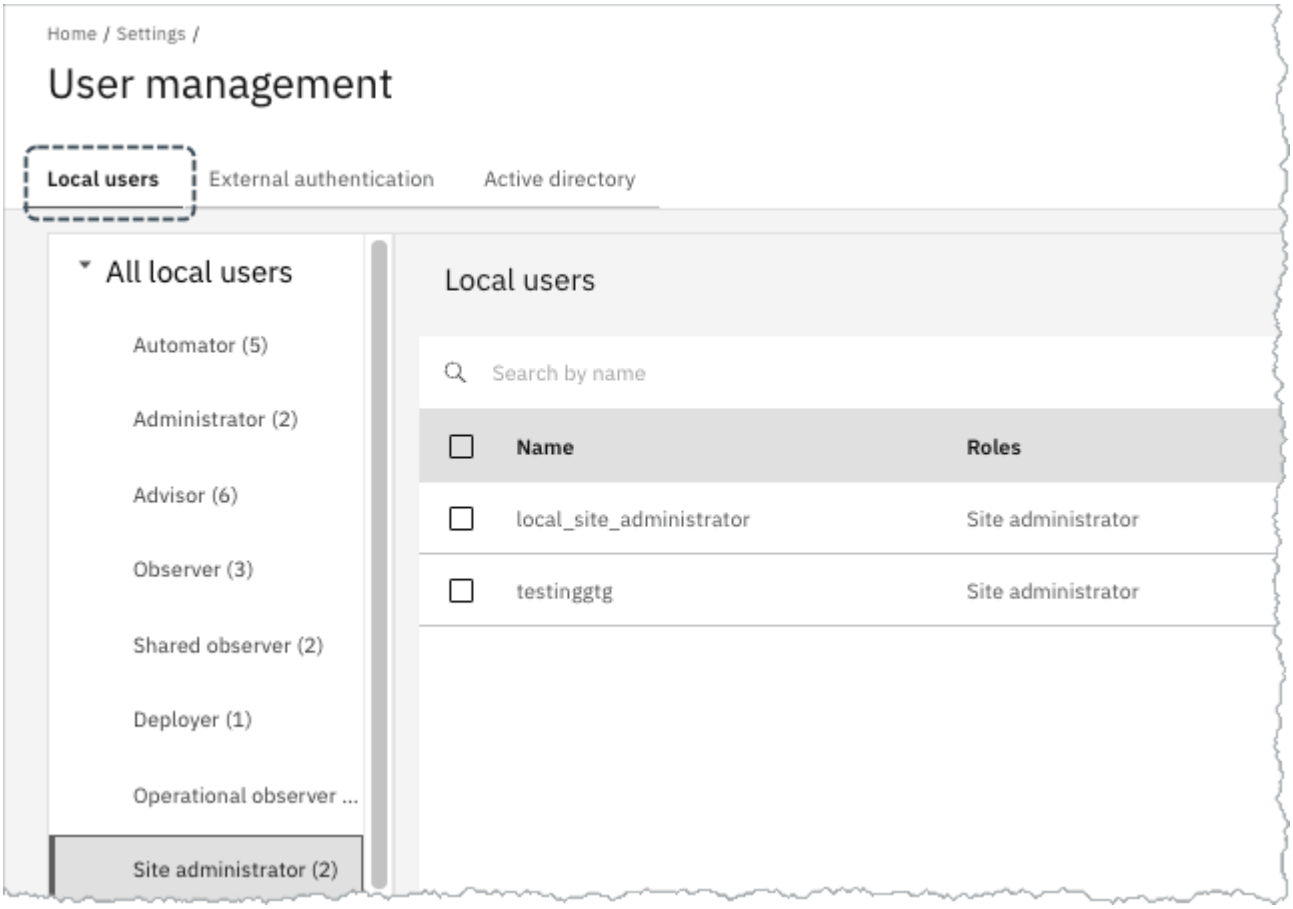
- Manage local or external authentication.
- Delete an account.
- Edit an account by clicking the pencil icon, which appears while hovering over a name.
- Create new user or group account.

3. Filter the list of users.



To work with a long list of users, you can filter by role (for example, only show administrator or only show observer users), or you can type a string in the **Search** field.

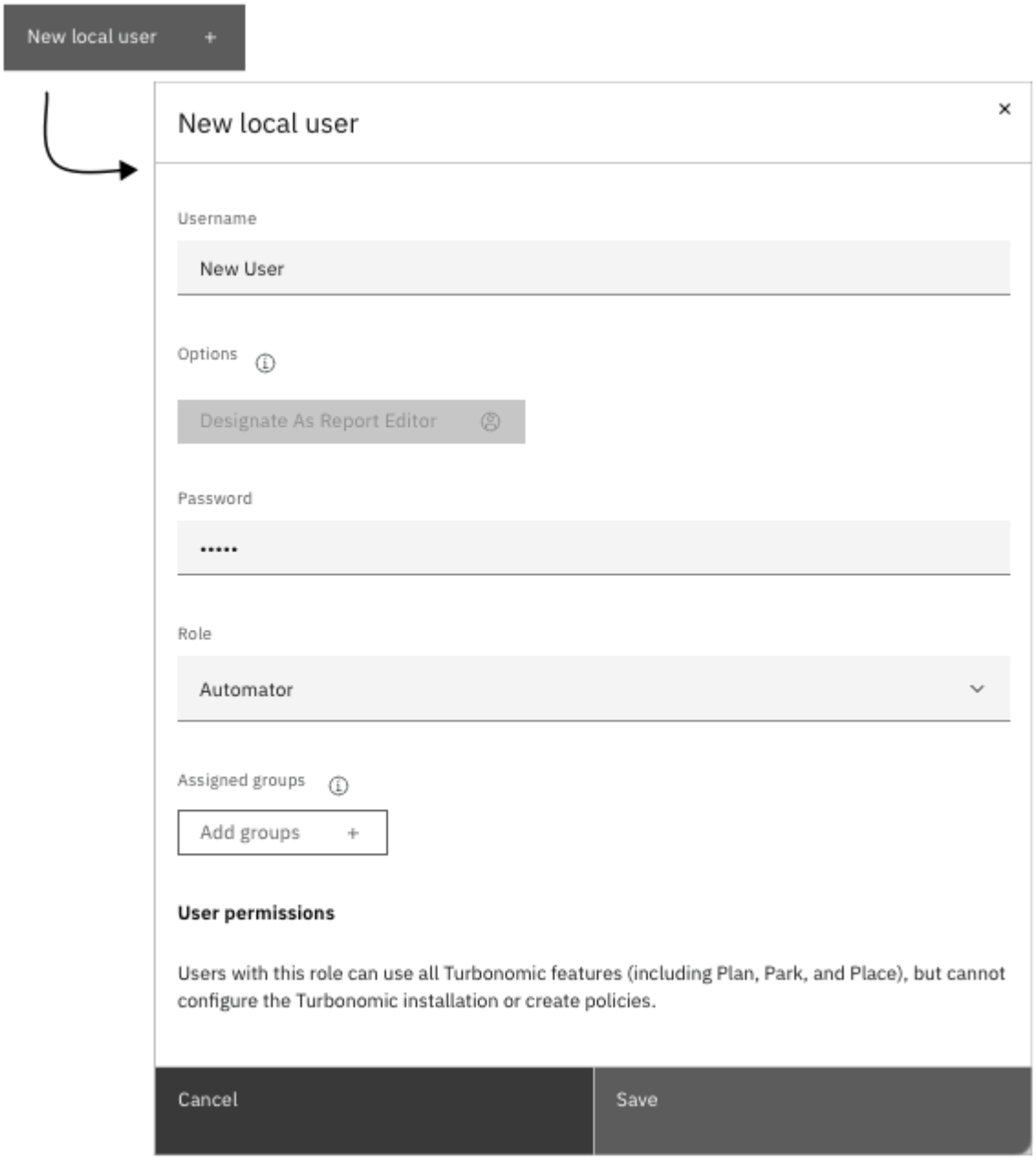
4. Configure local user accounts.



Workload Optimization Manager stores local accounts and their credentials on the Workload Optimization Manager platform. Local authentication is for individual users, only.

When you choose **Local users**, Workload Optimization Manager displays a list of all the local user accounts you have configured for this installation.

5. Create or edit a local user account.


New local user +
New local user
Username
New User
Options
Designate As Report Editor
Password
•••••
Role
Automator
Assigned groups
Add groups +
User permissions
Users with this role can use all Turbonomic features (including Plan, Park, and Place), but cannot configure the Turbonomic installation or create policies.
Cancel
Save

To add a new local user, click **New local user**. To edit an existing account, click the account name in the list. To configure a local account, specify:

- Authentication

  Provide the username and password. Workload Optimization Manager stores these credentials on the local server.

- Authorization – User role
  - Administrator

    Users with this role can use all Workload Optimization Manager features and modify settings to configure the Workload Optimization Manager installation. For Workload Optimization Manager instances hosted in the public cloud, this role is limited to the Workload Optimization Manager representative that manages the instances.

  - Site administrator

    Users with this role can use all Workload Optimization Manager features and modify site-specific settings to configure the Workload Optimization Manager installation. Users can also administer Groups, Policies, Workflows, Templates, Billing/Costs, and Target Configuration, but not Email, Licenses, Updates, and Maintenance. Users can create other user accounts, except accounts with the Administrator role.

  - Automator

    Users with this role can use all Workload Optimization Manager features (including Plan, Park, and Place), but cannot configure the Workload Optimization Manager installation or create policies.

  - Parker

    Users with this role can manually park (stop/start) workloads, manage parking schedules, and view all Workload Optimization Manager charts and data. However, users cannot place workloads, create policies and templates, run plans, or execute any recommended actions.

  - Deployer

    Users with this role can view all Workload Optimization Manager charts and data, use Place to reserve workloads, and create placement policies and templates. However, users cannot run plans or execute any recommended actions.

  - Advisor

    Users with this role can view all Workload Optimization Manager charts and data, and run plans. However, users cannot use Place to reserve workloads, create policies, or execute any recommended actions.

  - Observer

    Users with this role can view the environment, including the Home Page and Dashboards. Users can also use Search to set a scope to the session. For scope, only VM groups and Resource Groups are supported.

  - Operational observer

    Users with this role can view the environment, including the Home Page, Dashboards, Groups, and Policies. Users can also use Search to set a scope to the session.

  - Shared advisor

    Users with this role are scoped users. They can view the Home Page and Dashboards, but only see VMs and Applications. Users cannot execute Workload Optimization Manager actions.

  - Shared observer

    Users with this role are scoped users. They can view the Home Page, custom dashboards, and application maps, but can see only VMs and Applications. They cannot view Executive Dashboards or execute Workload Optimization Manager actions. Shared Observer is the most restricted user role.

  - Report editor

    Users with this role can create, edit, and delete reports. Due to limits to the reporting license, only one user per instance is allowed to have this role (by default, the local **administrator** user). To assign this role to another user, you must first remove it from the current user. Be sure that the new user is *not* a scoped user.

- Authorization – scope or assigned groups (optional)

  You can limit what the user can monitor in your environment. For example, you can scope to a group that contains only the physical machines that support this user's VMs or applications. Click **Add scope** or **Add groups** and choose which groups or clusters this user can see.

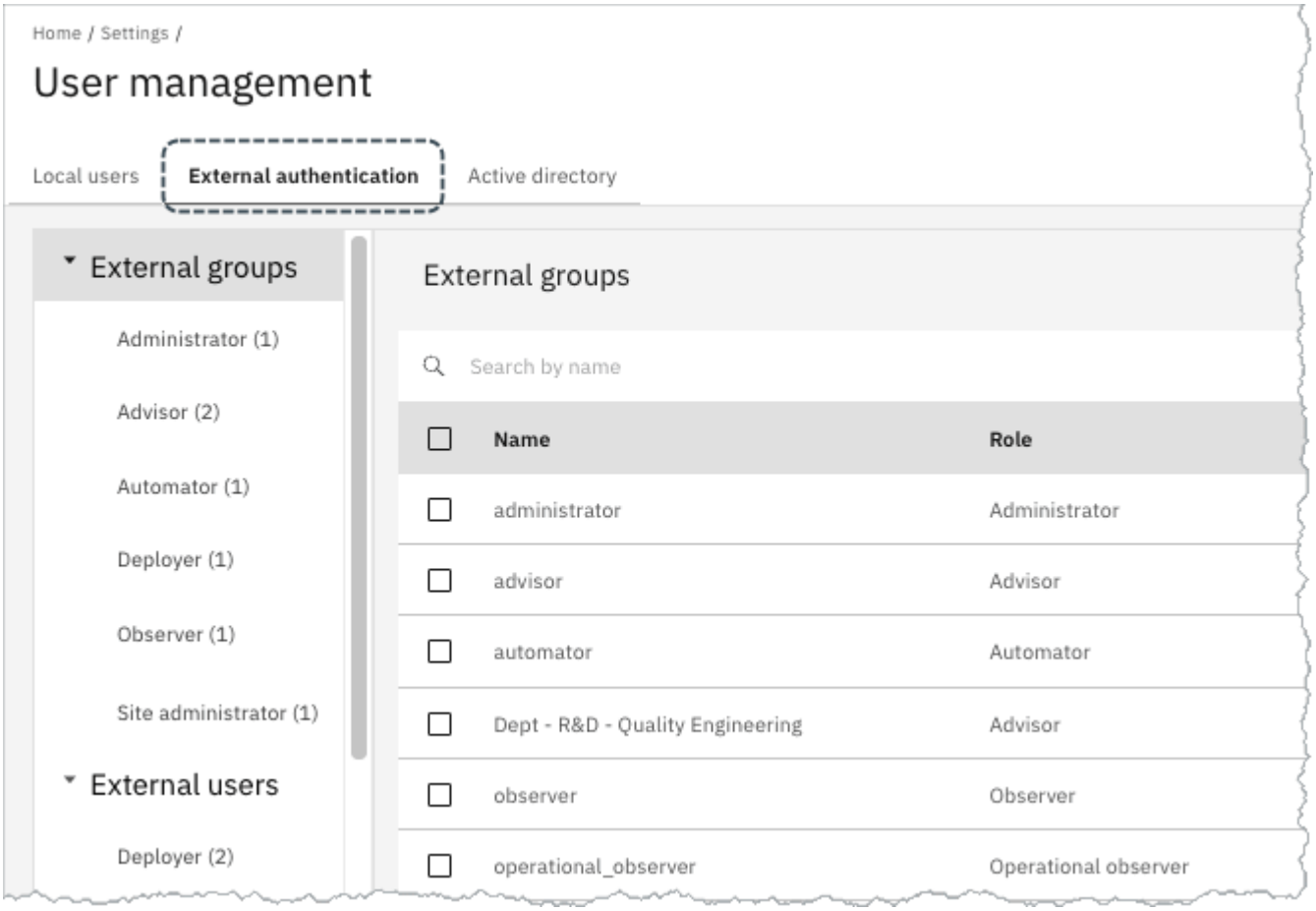
  NOTE:
  Under most circumstances, a scoped user cannot see actions for entities that are outside of the configured scope. However, when zooming in to Host entities, the user can see actions for storage that is outside of the user's scope if the hosts use that storage.

- Permission

  The user permissions are different for each role.

- When you are done, save the account.

6. Configure external authentication.



For external authentication, you configure Workload Optimization Manager to use Single Sign-on (SSO) or Active Directory (AD) services to manage the credentials and authentication of users. You can create external accounts to authorize user groups or individual users.

Points to consider:

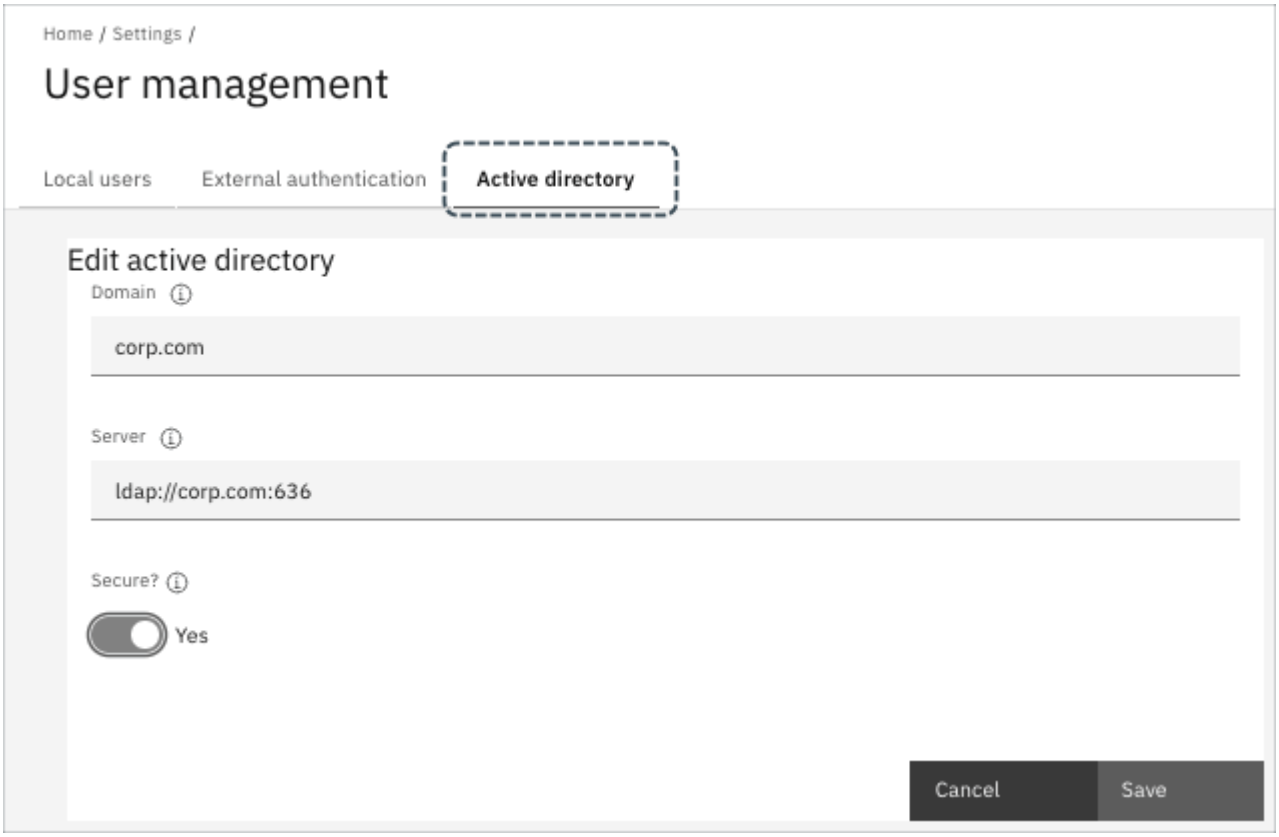
- If an external user is a named external user, Workload Optimization Manager authenticates the user with the permissions defined for that user, independent of membership in any of the configured external groups.
- If an external user is not a named external user, Workload Optimization Manager considers the membership in a defined external group.

- If a user is a member of multiple groups, Workload Optimization Manager authenticates the user through the first group that was returned during the SSO or AD authentication.
- A user always maps to a single external user or external group created in Workload Optimization Manager, and then applies the corresponding role and scope.
- Workload Optimization Manager does not support nested AD groups. AD logins must be for users in a top-level group.

To enable SSO, you must configure access to the given IdP. For information about configuring SSO, see Single Sign-On Authentication *(on page 100)*.

To enable AD you must specify either an AD domain, an AD server, or both. Workload Optimization Manager uses this connection for all AD users.

7. Enable AD authentication.



Click **Active directory** to manage the AD configuration.

- Active Directory domain

  To authenticate AD groups, specify a domain so that AD can find a user with the User Principal Name (UPN). If you specify a domain, but not a server, authentication uses any AD server from that domain. If you want to support only AD users and not groups, the domain is optional.

- Active Directory server

  To disable AD groups, specify a server but do not specify a domain. If you specify a domain and a server, authentication uses that server and also supports groups.

  When you configure an AD server, by default Workload Optimization Manager assumes the AD server port to be 389 or 636. To specify a custom port for the AD server, add the port number to the AD server IP address. For example, `10.10.10.123:444` sets port `444`.

- Secure

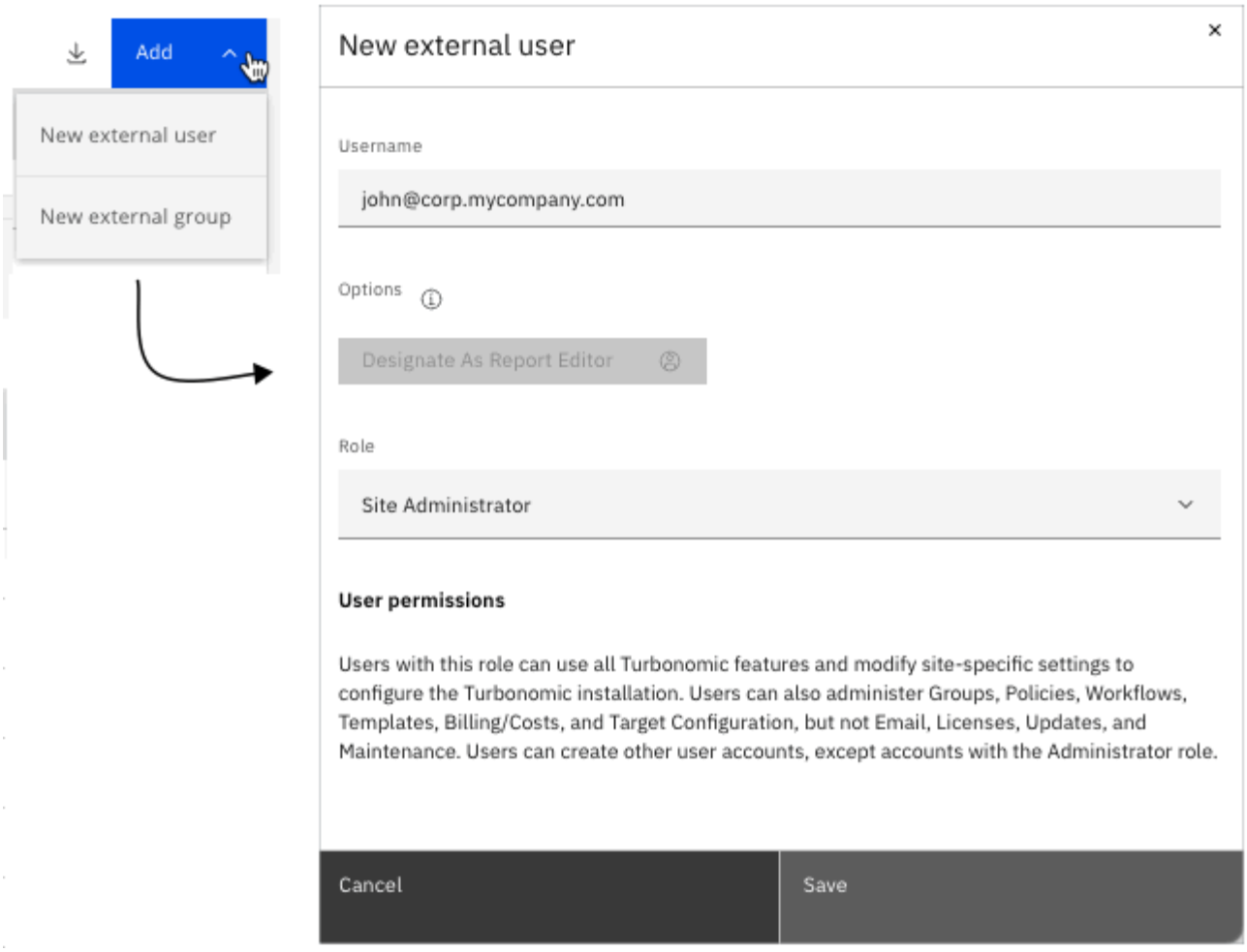
  Use a secure connection to communicate with AD servers. The AD domain must be configured to use LDAPS, and you must import a certificate into the Workload Optimization Manager server. Workload Optimization

Manager can support LDAP channel binding and LDAP signing. To support these Active Directory features, you must configure secure access.

For more information, see Enforcing Secure Access *(on page 77)*.

- When you are done, save your changes.

8. Create or edit an SSO or AD account.



This account can be for a user group or for a single user. To add a new account, click **Add > New external user** or **New external group**. To edit an existing account, click the account name. To configure an external account, specify:

- Authentication

  Provide the group or user name for this account. The name you provide must meet certain requirements, depending on the type of account you are creating:

  - External group - SSO

    Provide a name that matches a group the IdP manages.

  - External group - AD

    The group name must match a group that is accessible from the domain and servers that you configured in **EDIT AD**.

  - External user - SSO

    Provide a user name that matches a user managed by the IdP.

  - External user - AD

    The username must be a valid User Principal Name (UPN). For example, john@corp.mycompany.com.

- Authorization – user role

  - Administrator

    Users with this role can use all Workload Optimization Manager features and modify settings to configure the Workload Optimization Manager installation. For Workload Optimization Manager instances hosted in the public cloud, this role is limited to the Workload Optimization Manager representative that manages the instances.

  - Site administrator

    Users with this role can use all Workload Optimization Manager features and modify site-specific settings to configure the Workload Optimization Manager installation. Users can also administer Groups, Policies, Templates, Billing/Costs, and Target Configuration, but not Email, Licenses, Updates, and Maintenance. Users can create other user accounts, except accounts with the Administrator role.

  - Automator

    Users with this role can use all Workload Optimization Manager features (including Plan, Park, and Place), but cannot configure the Workload Optimization Manager installation or create policies.

  - Parker

    Users with this role can manually park (stop/start) workloads, manage parking schedules, and view all Workload Optimization Manager charts and data. However, users cannot place workloads, create policies and templates, run plans, or execute any recommended actions.

  - Deployer

    Users with this role can view all Workload Optimization Manager charts and data, use Place to reserve workloads, and create placement policies and templates. However, users cannot run plans or execute any recommended actions.

  - Advisor

    Users with this role can view all Workload Optimization Manager charts and data, and run plans. However, users cannot use Place to reserve workloads, create policies, or execute any recommended actions.

  - Observer

    Users with this role can view the environment, including the Home Page and Dashboards. Users can also use Search to set a scope to the session. For scope, only VM groups and Resource Groups are supported.

  - Operational observer

    Users with this role can view the environment, including the Home Page, Dashboards, Groups, and Policies. Users can also use Search to set a scope to the session.

  - Shared advisor

    Users with this role are scoped users. They can view the Home Page and Dashboards, but only see VMs and Applications. Users cannot execute Workload Optimization Manager actions.

  - Shared observer

    Users with this role are scoped users. They can view the Home Page and custom Dashboards, but only see VMs and Applications. Users cannot see Executive Dashboards or execute Workload Optimization Manager actions. This is the most restricted user.

  - Report editor

    Users with this role can create, edit, and delete reports. Due to limits to the reporting license, only one user per instance is allowed to have this role (by default, the local **administrator** user). To assign this role to another user, you must first remove it from the current user. Be sure that the new user is *not* a scoped user.

- Authorization – scope or assigned groups (optional)

  The scope limits what members of this group can monitor. For example, you can scope for access to only the hosts that support this group's VMs or applications. Click **Define scope** or **Add groups** and choose which entities this members of this group can see.

- Permission

  The user permissions are different for each role.

- When you are done, save the account.

Configuring a group for SSO authentication

To use SSO authentication in Workload Optimization Manager, you should configure user groups on the IdP. The IdP can authenticate the group members, and then Workload Optimization Manager can assign the user role and scope according to that group's authentication. To manage personnel changes, you only need to manage the membership in the IdP group. For example, if a user leaves your organization, you only need to remove the member from the group on the IdP. Because authorization on Workload Optimization Manager is by group, that user will not have any authorization settings stored on the Workload Optimization Manager server.

IMPORTANT:
Before you enable SSO for your Workload Optimization Manager installation, *you must configure at least one SSO user with Workload Optimization Manager administrator privileges*. If you do not, then once you enable SSO you will not be able to configure any SSO users in Workload Optimization Manager. To authorize an SSO user as an administrator, use **EXTERNAL AUTHENTICATION** to do one of the following:

- Configure a single SSO user with administrator authorization.

  Add an external user. The username must match an account that is managed by the IdP.

- Configure an SSO user group with administrator authorization.

  Add an external group. The group name must match a user group on the IdP, and that group must have at least one member.

For more information about configuring SSO authentication, see Single Sign-On Authentication *(on page 100)*.

Specifying a group in the SAML response

To support SSO, Workload Optimization Manager recognizes IdP responses that comply with SAML 2.0. To create user groups, for each user response you include an attribute named group, and give the group name as the attribute value. For example, assuming the following users, setting the group attribute for each user assigns that user to the appropriate group.

Users:	Group attribute:
▪ George ▪ Paul	Attribute Name=group, AttributeValue=Beatles

Users:	Group attribute:
▪ John ▪ Ringo	
▪ Smokey ▪ Pete ▪ Ronnie ▪ Claudette ▪ Bobby ▪ Marv	Attribute Name=group, AttributeValue=Miracles

As you specify the user response, to add the user to a group you include a group attribute. For example, to add a user to a group named turbo_admin_group, you would include the following attribute in that user's SAML response:

```
<saml2:Attribute
  Name="group"
  NameFormat="urn:oasis:names:tc:SAML:2.0:attrname-format:unspecified">
  <saml2:AttributeValue
    xmlns:xs="http://www.w3.org/2001/XMLSchema"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:type="xs:string">
        turbo_admin_group
  </saml2:AttributeValue>
</saml2:Attribute>
```

Setting group authorization in Workload Optimization Manager

To set an account role and scope to a user group, you must use the group name that you specify as the value in the given SAML group attribute. In the preceding example, the group value is turbo_admin_group. To set authorization for that group:

1. Open the User Management page to EXTERNAL AUTHENTICATION.

   Navigate to **Settings > User Management**, and display the **EXTERNAL AUTHENTICATION** view.

2. Create a new External Group

   Click **NEW EXTERNAL GROUP**.

3. Provide the group name.

   Be sure to use the name that you specify in the group attribute of the SAML response. For the previous example, use the name turbo_admin_group.

4. Specify the group's authorization

   For the previous example, since this is turbo_admin_group, you should set the **ADMINISTRATOR** role, and you should not set any scope (grant full access to the environment).

   After you configure this group in Workload Optimization Manager, then any member of turbo_admin_group that the IdP returns will have full administrator privileges on your Workload Optimization Manager installation.

The Maintenance page

The Maintenance page provides options for advanced users, such setting logging levels and uploading diagnostics data for Workload Optimization Manager technical support. Contact technical support for assistance with these options.

To view the Maintenance page, navigate to **Settings > Maintenance**. The page includes the following tabs:

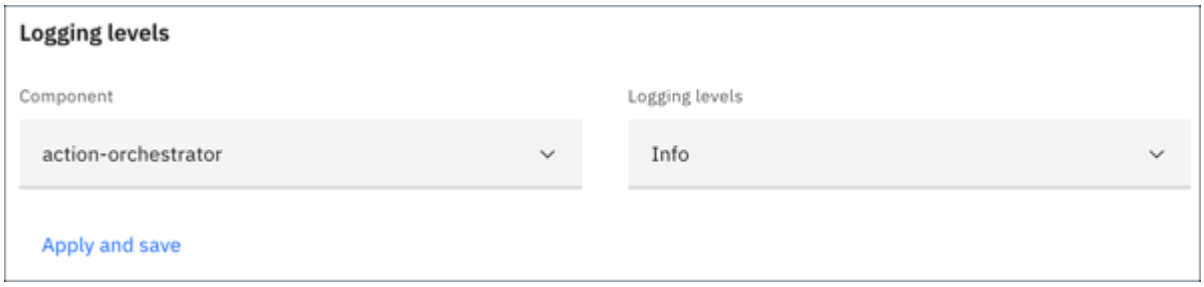
- Maintenance settings

  This tab includes settings for logging levels, HTTP proxy, and data retention.

- Telemetry and diagnostics

  This tab includes settings for diagnostics.

Logging levels

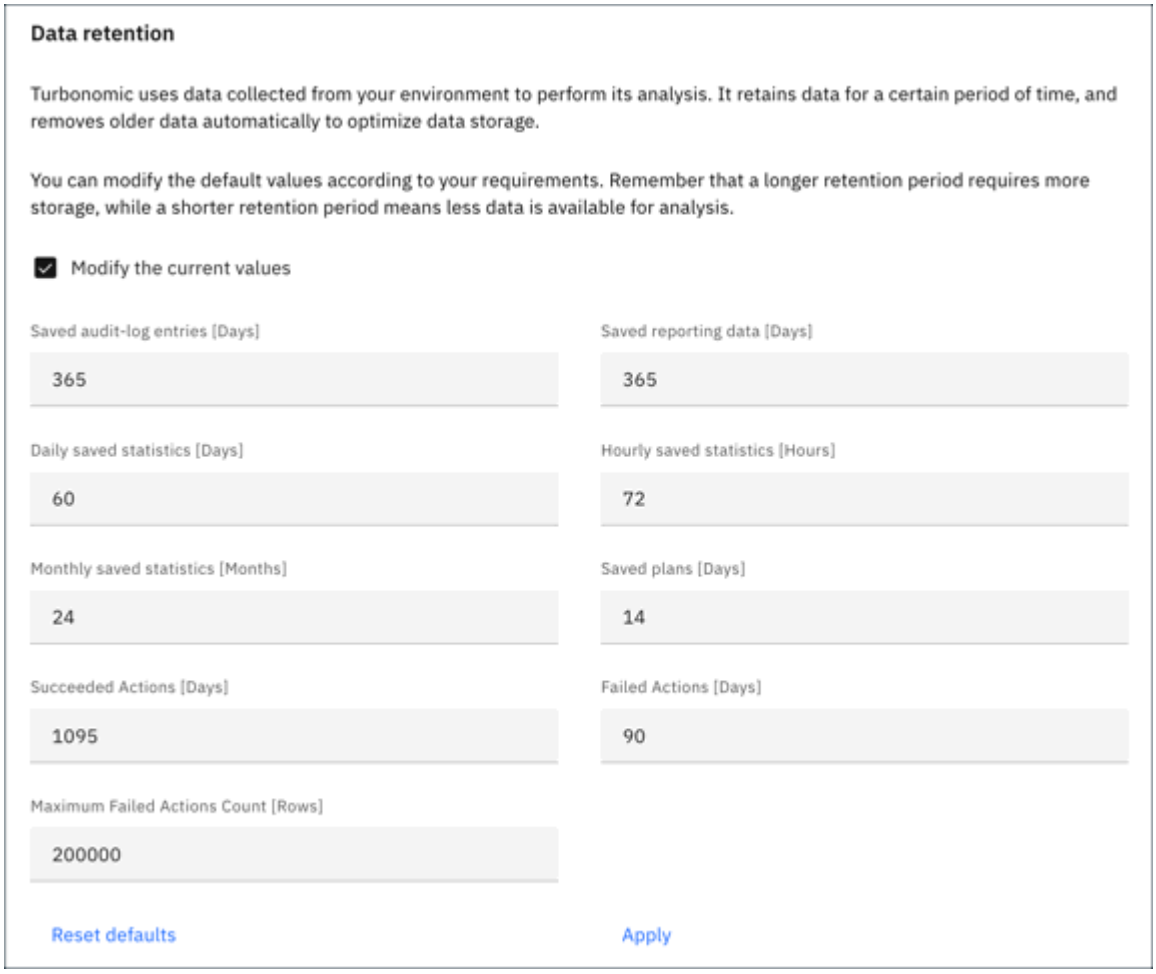


You can set the level of logging for different components of the Workload Optimization Manager platform. Be aware that setting more verbose logging levels increases the disk space that is required to store the log files. You normally change these settings only while you are working with technical support.

HTTP proxy

If your environment requires an HTTP proxy for Workload Optimization Manager to access the web, provide the credentials.

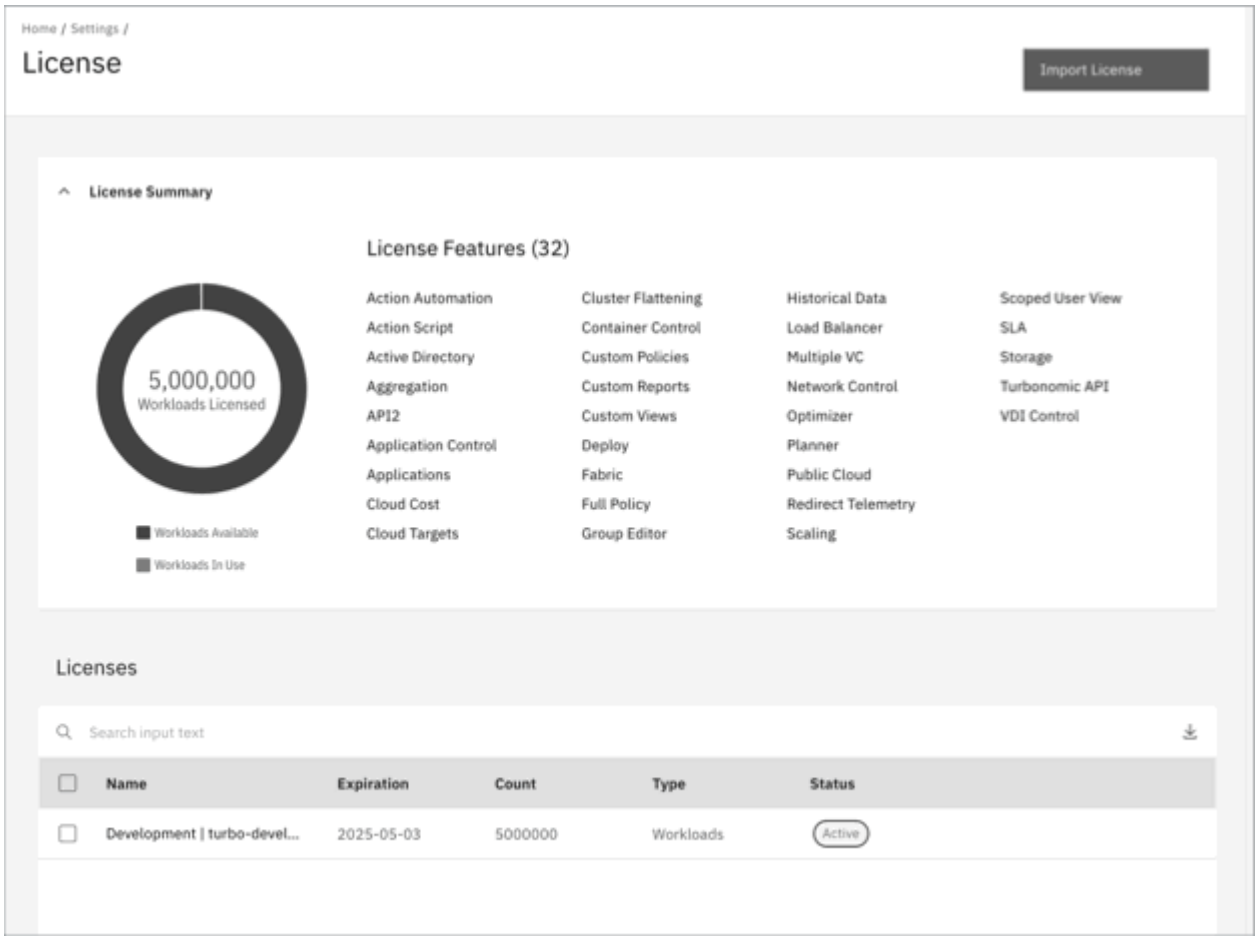
Data retention



Workload Optimization Manager uses data collected from your environment to perform its analysis. Data includes audit log entries, reports, statistics, plans, and actions. The platform retains data for a certain period of time and automatically removes older data to optimize data storage.

You can always modify the default retention values to meet your requirements. A longer retention period requires more storage, while a shorter period means less data is available for analysis.

License configuration



To activate the full range of Workload Optimization Manager features, you must purchase the appropriate license. When you purchase the license, Cisco sends an e-mail message with instructions on how to obtain the license key.

A product license enables specific features as well as a specific number of workloads that you can manage. You can add additional licenses to Workload Optimization Manager as a way to increase the number of workloads your installation can manage. As you add more licenses, they must all support the same feature set.

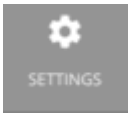
The License page shows the following information:

- Number of workloads that you can manage under the license
- Number of workloads that are currently in use
- Features that the license enables
- A list of licenses and the status for each license

You can search for a license or download a list of licenses from the page.

To navigate to the License page:

1. Navigate to the Settings Page.



2. Choose License.



Activate a license or update your current license

1. Obtain your license.

   Cisco sends an e-mail message with instructions on how to obtain the license key. Save the license file on your local machine so you can upload it to your Workload Optimization Manager installation.

2. Apply the license to your Workload Optimization Manager installation.

   a. Click **Import License** then browse to the license file that you saved and open it.

      Optionally, you can drag the file into the **Enter License** fly-out.

   b. After you upload the file, click **Save**.

   c. Clear your browser cache and reload the Workload Optimization Manager user interface.

After you activate your license, you can add more licenses to increase your workload coverage, or you can license a higher feature set.

> **NOTE:**
> This only applies only to legacy Workload Optimization Manager customers. As you apply new licenses to Workload Optimization Manager, you must be sure that they are for the same edition or feature set. If you try to apply an incompatible license file, Workload Optimization Manager displays an *Invalid Feature Set* error. To apply the new license you must either delete your current license so you can install the new feature set, or you must obtain a different license file that matches your current feature set.

Increase your licensed workload coverage

1. Obtain your additional license.

   Your additional licenses must match the feature set of your current license.

2. Apply the license to your Workload Optimization Manager installation.

Upgrade your license to a higher feature set

1. Obtain your new license for the new features.

   You should obtain a license that supports at least the same number of workloads as your current license.

2. Delete your current license from Workload Optimization Manager.

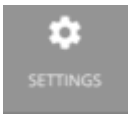
   On the License page, select all the licenses that you currently have installed, then click **Delete**.

3. Apply the license to your Workload Optimization Manager installation.

Email settings

If you enabled the Embedded Reporting component, you can configure email settings to send reports from Workload Optimization Manager.

1. Navigate to the Settings Page.

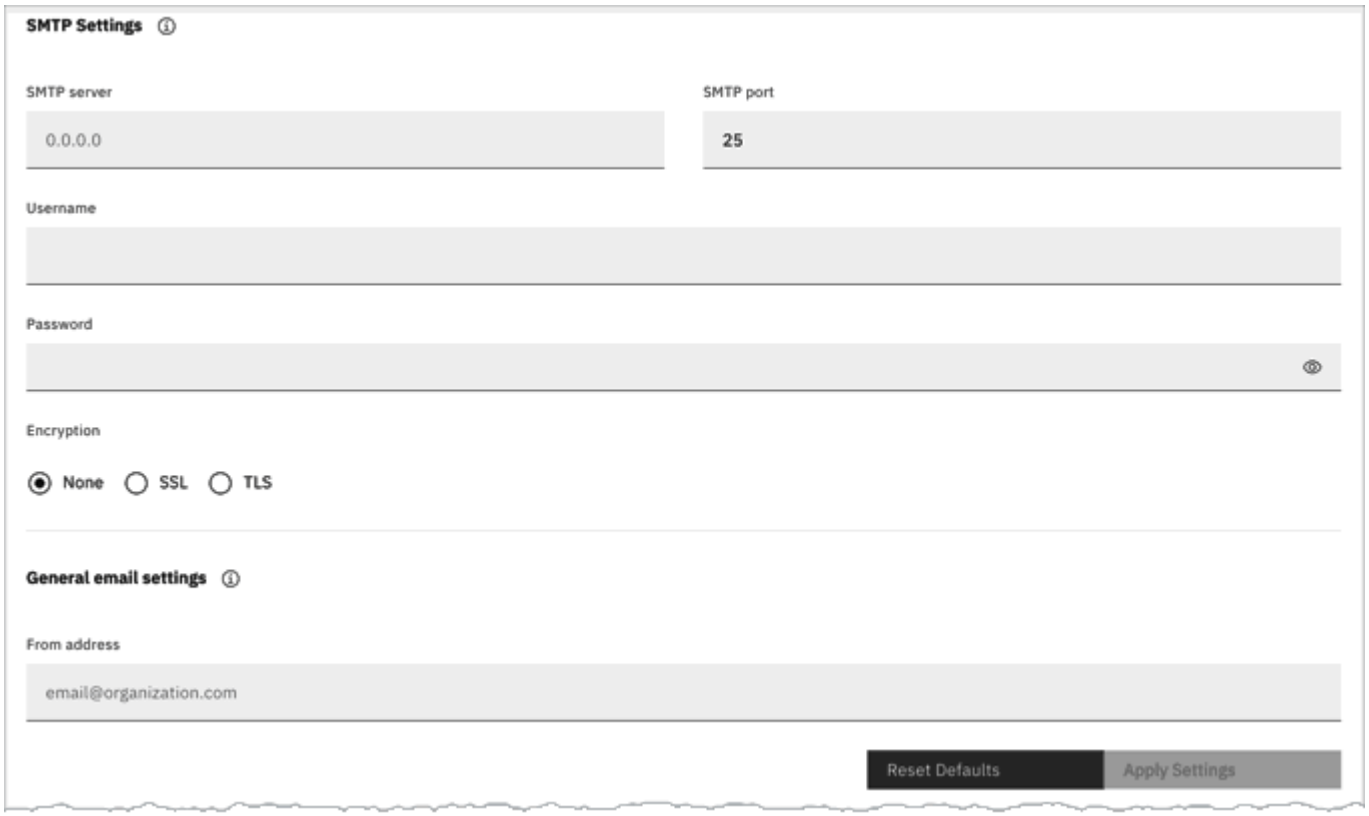
   

   Click to navigate to the Settings Page. From there, you can perform a variety of Workload Optimization Manager configuration tasks.
2. Navigate to the Email Settings Page.

   From here, you can configure:

   - SMTP settings
   - General email settings

SMTP settings



The SMTP Settings fields identify the mail relay server you use on your network to enable email communication from Workload Optimization Manager.

If the server requires authentication, provide the username and password here. You can also choose the following encryption options for notifications:

- None
- Ssl
- Tls

General email settings



Use this setting to specify the return address (the FROM address) for emails that Workload Optimization Manager generates and sends for report subscriptions.

API reference

This guide will help you to use the Workload Optimization Manager REST API as you script interactions with the Workload Optimization Manager software and develop integrations between Workload Optimization Manager and other software applications.

The REST API exposes Workload Optimization Manager data and processing to remote access via HTTP GET, POST, PUSH, and DELETE methods.

The Workload Optimization Manager user interface is a client to this API, and every feature exposed in the user interface is the product of API calls. To understand the API, you should have a working knowledge of the Workload Optimization Manager user interface, the features it exposes, the data it uses, and the use cases that it supports. In most cases, if you want to script some Workload Optimization Manager behavior, you should be able to execute most of that behavior via the user interface. Stepping through these use cases in that context will help you plan and implement your scripts.

To get the most out of the REST API, you should understand how Workload Optimization Manager organizes its underlying data, and how the various REST resources map to that organization. This introductory section includes discussions of the data that underlies:

- Markets and how they represent your environment (see Markets *(on page 1143)*)
- Plans and an overview of how to work with them (see Plans *(on page 1145)*)
- Policies and how they are represented (see Policies) *(on page 1148)*
- Reservations for deployment, and how to set them up (see Reservations *(on page 1157)*)

NOTE:
This guide is not a comprehensive reference to every possible setting and data payload in the REST API. Many endpoint descriptions include samples of API data objects, but they cannot document every possible combination of inputs. For a complete reference to the API, use this guide in conjunction with the Workload Optimization Manager REST Swagger-UI documentation. For more information, see Workload Optimization Manager REST API Swagger Documentation *(on page 1142)*.

API guide revision history

A revision of the API Guide is available for each major release, if the API undergoes significant change between major releases, or to correct significant errors in the API Guide. Each revision of the API Guide includes the following changes:

- Quarterly (major) release – the new revision for the major release, includes fixed issues and improvements to the API
- Bi-weekly (point) release – deprecated features are removed from the API or fully implemented as outlined in the API Deprecation History

For a list of upcoming deprecations, see API Deprecation Contract *(on page 48)*.

Version Family	Releases
3.13	3.13.0 *(on page 1135)*
3.12	3.12.1 *(on page 42)* | 3.12.0 *(on page 1135)*
3.11	3.11.1 *(on page 42)*3.11.0 *(on page 1135)*
3.10	3.10.1 *(on page 42)*|3.10.0 *(on page 1135)*
3.9	3.9.1 *(on page 43)*|3.9.0 *(on page 1136)*
3.8	3.8.1 *(on page 43)*| 3.8.0 *(on page 1136)*
3.7	3.7.1 *(on page 43)*| 3.7.0 *(on page 1136)*
3.6	3.6.1 *(on page 44)* | 3.6.0 *(on page 1137)*
3.5	3.5.1 *(on page 46)* | 3.5.0 *(on page 1137)*
3.4	3.4.1 *(on page 46)* | 3.4.0 *(on page 1137)*

API guide revision history

A revision of the API Guide is available for each major release, if the API undergoes significant change between major releases, or to correct significant errors in the API Guide. Each revision of the API Guide includes the following changes:

- Quarterly (major) release – the new revision for the major release, includes fixed issues and improvements to the API
- Bi-weekly (point) release – deprecated features are removed from the API or fully implemented as outlined in the API Deprecation History

For a list of upcoming deprecations, see API Deprecation Contract *(on page 48)*.

Version Family	Releases
3.13	3.13.0 *(on page 1135)*
3.12	3.12.1 *(on page 42)* | 3.12.0 *(on page 1135)*
3.11	3.11.1 *(on page 42)*3.11.0 *(on page 1135)*
3.10	3.10.1 *(on page 42)*|3.10.0 *(on page 1135)*
3.9	3.9.1 *(on page 43)*|3.9.0 *(on page 1136)*
3.8	3.8.1 *(on page 43)*| 3.8.0 *(on page 1136)*
3.7	3.7.1 *(on page 43)*| 3.7.0 *(on page 1136)*
3.6	3.6.1 *(on page 44)* | 3.6.0 *(on page 1137)*
3.5	3.5.1 *(on page 46)* | 3.5.0 *(on page 1137)*

Version Family	Releases
3.4	3.4.1 *(on page 46)* | 3.4.0 *(on page 1137)*

API change history

Each major release of Workload Optimization Manager includes fixes issues and improvements to the API. The following list includes the changes for each major release.

> **NOTE:**
> Deprecations are implemented in point releases of Workload Optimization Manager. For example, if API features were deprecated in 8.6.3, those deprecations are linked here in the entry for 8.7.0, along with the list of all other fixes and improvements from the 8.6.x point releases. For the latest deprecation history, see API Deprecation History *(on page 41)*.

3.13.0 API Fixes and Improvements

- "Vendor link in the API response for vCenter based hosts"

  For vCenter based hosts, when you execute `GET /entities/{entityUuid}` in the API, the response now shows a vendor link to the appropriate section in vCenter. This link corresponds to the **Vendor link** field in the Entity information chart in the UI.

3.12.0 API Fixes and Improvements

- "Security improvement: Hide exception details in API responses"

  API error responses smust not include `exception` and `class` information, as this could reveal internal implementation details to an attacker.

  Workload Optimization Manager now disables the output details for security exceptions by default. The API caller can still understand the nature of an error by inspecting the `type` and `message` fields.

- "Removal of API parameters for the `/action/availablemodes` API call."

  API Parameters `*entity_type` and `*action_type` are removed from the `/action/availablemodes` API call.

3.11.0 API Fixes and Improvements

- **Improvement:**

  "Added regex validation for `/search` API calls."

  For API developers, regex validation is added for `/search` API calls. If you specify a `*` as the first character of the search when you search for a group in the UI, you no longer receive a 500 error that indicates an internal system error. Instead, you receive a 400 error that correctly indicates invalid user input.

3.10.0 API Fixes and Improvements

For this release (version 3.13.0), there are no new API changes to report. API changes from the previous version family are incorporated in the API Guide.

3.9.0 API Fixes and Improvements

- **Improvement:**

  > "The `GroupType` parameter now accepts `ResourceGroup` as a valid enum."

  For API developers, the `GroupType` parameter in the API now accepts `ResourceGroup` as a valid enum for entities that belong to resource groups, such as VMs. For backward compatibility, the API continues to accept the existing `Resource` enum. However, API responses include only `ResourceGroup`.

3.8.0 API Fixes and Improvements

- **Improvement:**

  > "The `Parking` APIs are now available."

  Workload Optimization Manager supports `parking` actions for cloud resources. These actions *stop* your cloud resources for a period to help you reduce your cloud expenses, and then *start* these resources later when you need them. Use the `Parking` endpoint to get a list of parkable entities and execute parking actions on demand.

  You can also enforce parking actions according to a schedule. In the `Schedules` endpoint, use `/schedules/timespans/` requests to configure parking schedules and attach the schedules to parkable entities. When parking schedules are attached to entities, use the `Parking` endpoint to pause and resume schedules as needed.

- **Improvement:**

  > "The API now allows an `operator` field in the `instanceType` and `newInstanceType` action filters."

  The API now allows an `operator` field in the `instanceType` and `newInstanceType` action filters. The `operator` can be EQ (equals) or NEQ (not equal to). If no value is set, the action filter uses EQ.

- **Improvement:**

  > "The API no longer includes historical price index data in various API methods that return entity statistics."

  The API no longer includes historical price index data in various API methods that return entity statistics. Price index is still available for the most recent real-time projected data, and in responses for plan-related requests.

  This specifically refers to the `StatSnapshotApiDTO` structure when `epoch` is set to `CURRENT` or `HISTORICAL`.

3.7.0 API Fixes and Improvements

- **Improvement:**

  > "API users can now search for entities by their tag as follows: `GET /tags/entities?key=<tagKey>`. "

  API users can now search for entities by their tag by calling `GET /tags/entities?key=<tagKey>`, which includes entities with a slash (/) within a tag string.

  NOTE:
  The existing API call `GET /tags/<tagKey>/entites` will be permanently deleted in a future release. Until its removal, you can use both of these methods.

- **Improvement:**

  > "API users can now filter and sort targets."

  API users can now filter and sort the list of targets to restrict their view to only their relevant targets.

  The following parameters were added to the list of targets request:
  - target type (`AWS, AWS_BILLING, ACTIONSCRIPT, ACTIONSTREAM_KAFKA, APPDYNAMICS, APPINSIGHTS, AZURE_BILLING, AZURE_EA, AZURE_SERVICE_PRINCIPAL, CISCO_HYPERFLEX, CISCO_UCS_MANAGER, DATADOG, DYNATRACE, EMC_SCALEIO, EMC_VMAX, FLEXERA,`

GCP_BILLING, GCP_SERVICE_ACCOUNT, HP_ONEVIEW, HPE_3PAR, HYPER_V, INSTANA, JBOSS, JVM, KUBERNETES, MYSQL, NETAPP, NEWRELIC, NUTANIX, ORACLE, PURE, SNMP, SQLSERVER, SERVICENOW, TOMCAT, VMM, VMWARE_HORIZON, WMI, WEBLOGIC, WEBSPHERE, VCENTER)

  - validation status (FAILED, IN_PROGRESS, SUCCESS)
  - health state (CRITICAL, MAJOR, MINOR, NORMAL)

The following parameters for sorting the list of targets were added: `target type` and `health state`.

- **Fixed Issue:**

  > "Kubeturbo is using a deprecated API for the `APIRequestCount` query."

  Kubeturbo is using a deprecated API for the `APIRequestCount` query.

3.6.0 API Fixes and Improvements

- **Fixed Issue:**

  > "For API developers, when executing requests on the Search endpoint, the API does not paginate the returned result."

  For API developers, when executing requests on the Search endpoint, the API does not paginate the returned result.

3.5.0 API Fixes and Improvements

- **Improvement:**

  > "The API now supports reservations with cluster affinity."

  When creating reservations, you can now specify affinity rules for the reserved VMs. You can ensure all the VMs in a reservation deploy to the same cluster. If you do not specify affinity, the VMs deploy to any clusters in the scope. This is the same as earlier behavior for reservations.

  For more information, see Affinity Rules *(on page 1293)*.

- **Improvement:**

  > "The `/supplychains/leafEntities` endpoint is now available."

  This release introduces the `/supplychains/leafEntities` endpoint that you can use to define your own supply chain topology. This endpoint returns a list of 'leaf entities' related to a 'seed entity'.

  For more information, see Getting Leaf Entities from a Supply Chain Root *(on page 1391)*.

- **Fixed Issue:**

  > "The maximum entries returned for GET and POST methods is 500."

  Starting with Workload Optimization Manager version 3.4.1, GET and POST methods return a maximum of 500 entries when run on ServiceEntities, Groups, or BusinessAccount endpoints. You should always check for pagination in your results.

3.4.0 API Fixes and Improvements

- **Improvement:**

  > "The API now includes the `Plan Destinations` endpoint."

  A plan destination is an external resource where plan results can be uploaded. Currently, you can upload the results of a Migrate to Cloud plan to the Azure Migrate portal. Use the `Plan Destinations` endpoint to get plan destinations and run uploads.

- **Improvement:**

> "When deleting a reservation, you can now specify whether to permanently delete that reservation."

The API now includes the new `reservation_force_delete` parameter that you can pass along with a DELETE call to a specific reservation. When `reservation_force_delete = true`, the system removes the reservation permanently, no matter what state it is in. Before this release, a reservation that you deleted via the API or user interface continued to display for deletion, and was only permanently deleted after 48 hours.

- **Improvement:**

> "The API includes a new LABEL commodity type."

The `commodityType` enum includes a new LABEL commodity type.

- **Improvement:**

> "For API developers, when a reservation you create fails, the returned DTO includes more data about the failure."

When you create a reservation, and the reservation fails, the API returns a DTO to explain the cause of the failure. This DTO includes new fields:

  - `timespan`: Indicates whether the reservation calculation was based on realtime or historical resource utilization data. Can be one of:
    - `Current`: The calculations were based on realtime utilization data
    - `Historical`: The calculations were based on historical utilization
  - `cluster`: The name of the cluster that came the closest to being able to provide the resources that the reservation needed.
  - `hrefCluster`: A link to stats for the closest-seller cluster.
  - `hrefEntity`: A link to stats for the closest-seller entity that is a member of the closest-seller cluster.

For example, the DTO can include data similar to the following for a placement failure:

```
"placements": {
    "failureInfos": [
        {
            "resource": "StorageAmount",
            "timespan": "Historical",
            "closestSeller": {
                "uuid": "74167780552264",
                "displayName": "QSGRID01:DC6DS2",
                "className": "Storage",
                "cluster": "StorageCluster-A",
                "hrefEntity": "/reservations/stats/74167780552264",
                "hrefCluster": "/reservations/stats/74167780552265"
            },
            "maxQuantityAvailable": 248294.59024047852,
            "quantityRequested": 273184.5625,
            "units": "MB"
        }
    ]
}
```

- **Improvement:**

> "Handling of non-executable actions has been improved."

This release improves the handling of actions that are currently not executable due to prerequisite actions.

For example, in order to suspend Host A, VM_01 in the host must first move to Host B. However, Host B only has capacity for one VM and is currently hosting VM_02. In this case, Host A suspension is blocked by two prerequisite actions – VM_02 moving to another host and VM_01 moving to Host B.

The Action Details page for the main action (Host A suspension in the example) now shows the following information:

`"Blocked by actions on target or destination that need to be executed first."`

For earlier releases, the Action Details page indicated that the action is blocked by a policy.

For API developers, the `ActionApiDTO` now shows this state. Workload Optimization Manager sets the `BLOCKED_BY` key of the `relatedActionsCountByType` parameter to `0`, which indicates that an action is currently not executable due to prerequisite actions.

When all the prerequisite actions have been executed, the main action becomes executable.

- **Improvement:**

  > "When you examine a failed reservation via the API, the data now returns a reason for the failure."

  After you create a reservation, you can GET the `reservations` endpoint for data about the current reservations in your environment. The resulting data includes a `status`. If the status is `PLACEMENT_FAILED`, Workload Optimization Manager could not place all the workloads in the reservation for some reason. Starting with version 3.3.4, the `placements` field in the data includes `invalidInfo` to give a reason for the placement failure. The reason can be:

  - `MARKET_CONNECTIVITY_ERROR`: The reservation process could not connect to the Market component for some reason.
  - `MARKET_NOT_READY`: The market does not have data from the Headroom plans for the given reservation scope.
  - `INVALID_CONSTRAINTS`: The reservation policies include conflicting or invalid constraints.

  For example, the data for a single reservation might include the following data:

```
"demandEntities": [
    {
        "template": {
            "uuid": "215167980658768",
            "displayName": "MyVmTemplate",
            "className": "VirtualMachineProfile"
        },
        "placements": {
            "invalidInfo": {
                "invalidReason": "MARKET_CONNECTIVITY_ERROR"
            }
        }
    }
]
```

Getting started with the Workload Optimization Manager REST API

The easiest way to see results with the API is to try calls in the Swagger UI. To use the Swagger UI, navigate to:

`https://<Your_Workload Optimization Manager_URL>/apidoc`

You can use the Swagger UI to navigate to different endpoints and try out their methods. For example, if you navigate to the Users endpoint you can get a listing of all the users that are specified for your Workload Optimization Manager installation. For more information, see Workload Optimization Manager REST API Swagger Documentation *(on page 1142)*.

Ultimately, you will want to implement scripts that use the API to get data, execute actions, or integrate Workload Optimization Manager with other processes. As you work with the API, you need to know about:

- Authentication
- URI Structure

- Response Format
- Time in the Workload Optimization Manager API

Authentication

To use the API, you must have a valid user account on the Workload Optimization Manager instance. Also note that accounts can have different roles. The API will only execute commands that are valid for your user role. For example, to execute Workload Optimization Manager recommended actions, your account must have a role of either `administrator`, `deployer`, or `automator`.

To make API calls, you request an authentication token and pass it with each call to the Workload Optimization Manager API. The token request returns a cookie for your authentication. A common way to use this token is to store the cookie locally, and pass it with your API calls.

Example: `curl -s -k -c /tmp/cookies -H 'accept: application/json' 'https://localhost/api/v3/login?hateoas=true' -d 'username=administrator&password=password'`

Then, each request must use the `-b cookie-filename` parameter to use the session cookie delivered by the login request.

Another approach is to get the authentication header and parse out the authentication cookie. Then you can create a header for each API request that includes the cookie. For example, assume you store the value in a variable named `token`. You could use it like this:

```
headers = {'cookie': token}
r = requests.get('https://10.10.123.456/api/v3/targets/specs', headers=headers, verify=False, stream=True)
```

NOTE: For a more complete explanation, see the Authentication Recipe *(on page 1439)* in the API Cookbook.

URI structure

To use the Workload Optimization Manager REST API, your client will make HTTP requests to specific REST resources. The Workload Optimization Manager REST API supports the standard HTTP methods:

- GET

  Get lists of entities or data objects, get individual items.
- POST

  Create new objects in the Workload Optimization Manager environment, or specify filters for certain queries.
- PUT

  Incrementally modify existing entities or objects.
- DELETE

  Delete entities or objects.

The base URI structure for a Workload Optimization Manager REST API resource is:

```
https://<Your_Workload Optimization Manager_URL>/api/<API_version>/<resource_name>
```

For example, to list the users in your installation:

```
https://111.222.33.44/api/v3/users
```

Response format

The Workload Optimization Manager REST API returns data as JSON objects. Workload Optimization Manager refers to these objects as Data Transfer Objetcs or DTOs. A DTO is an array of key-value pairs that describe the data you have requested, or the data that is the result of executing a POST or PUT. For example, if you GET the users defined for an installation of Workload Optimization Manager, the API returns a DTO similar to this:

```
[
  {
  "links": [
    {
      "rel": "self",
      "href": "https://10.10.10.10/api/v3/users/_4T_7kwY-Ed-WUKbEYSVIDw"
    }
  ],
  "uuid": "_4T_7kwY-Ed-WUKbEYSVIDw",
  "displayName": "Administrator User",
  "username": "administrator",
  "roleUuid": "_4UAioQY-Ed-WUKbEYSVIDw",
  "roleName": "administrator",
  "loginProvider": "Local",
  "type": "DedicatedCustomer",
  "showSharedUserSC": false
  }
]
```


In this case, the DTO is an array of one object. In other words, there is only one user accout defined for this installation of Workload Optimization Manager. The user object begins with a `links` array that gives the URL to this user account. It then follows with properties to describe the given user account.

Note that in most cases to execute a PUT or POST, you will pass the parameters to create the object via a DTO. These DTOs are similar to the associated response DTOs, but they are not identical. For this example of a user account, the response DTO does not include the user account password, but the DTO to create an account must include the password.

You can try out different REST methods in the Swagger UI to see typical response DTOs.

Time in the Workload Optimization Manager API

The Workload Optimization Manager REST API contains requests that require a start time and/or an end time to create or filter information. Time may be entered in three different formats:

- ISO 8601 Date and Time Format

  `YYYY-MM-DDTHH:MM:SS`. For example, `2018-10-07T12:38:17`

- Epoch Time

  Epoch Time is represented as the number of seconds that have elapsed since midnight UTC, January 1, 1970. For example, `1514764800` corresponds to January 1, 2018 12:00:00 AM UTC.

- Relative time

  Relative time is represented as the time relative to when the call is executed. For example, a start time of `-1w` and an end time of `-1d` indicates that the results should include entries from a week before the call is executed, to the day before the call is executed. Relative units are case-sensitive. Workload Optimization Manager supports the following relative units:
  - `m` – minutes
  - `h` – hours
  - `d` – days
  - `w` – weeks
  - `M` – months
  - `y` – years

Pagination in the Workload Optimization Manager API

Some API calls return very large datasets. Workload Optimization Manager recommends using the pagination features at all times. This is particularly important for larger environments managed by Workload Optimization Manager. Via scopes and filters

for individual requests, data may be preliminarily filtered. Using the `limit` and `x-next-cursor` parameters, the data can be returned in manageable chunks.

For example:

```
https://10.10.10.10/api/v3/markets/214075923753936/entities/stats?limit=5&ascending=true
```

The previous request is to obtain statistics for all entities in a specific market. `limit=5` indicates that each page of returned data should have 5 results.

The headers in the returned data will now include `x-next-cursor: 5`. In order to retrieve the next page of results, use the cursor in the next request:

```
https://10.10.10.10/api/v3/markets/214075923753936/entities/stats?
cursor=5&limit=5&ascending=true
```

When the final page is reached, the `x-next-cursor` will be empty.

If a request features the `orderBy, limit,` or `cursor` parameters, using one of those parameters will return paginated results.

Workload Optimization Manager REST API Swagger documentation

You can try out different REST methods in the Swagger documentation to see typical response DTOs.

To access the Workload Optimization Manager Swagger documentation, navigate to:

```
https://<Your_Workload Optimization Manager_URL>/apidoc
```

This documentation uses a standard Swagger-UI application that delivers documentation for all the methods on the different endpoints in the API. This documentation includes a description of the call, a list of parameters to pass, and descriptions of the JSON payloads you POST or PUSH, as well as the payloads you get in the response. If you view the MODEL of a JSON payload, you can navigate the JSON structure to see descriptions of the different object fields.

To access the API, the Swagger-UI requires authentication. If you have already logged into Workload Optimization Manager, your browser session should have a current authentication cookie. The Swagger-UI uses this same cookie for its API authentication.

For each method, the Swagger doc includes a **Try it out** feature where you can set up a call through the Swagger-UI, and then execute it. This generates the following information as a response:

- A curl request for your call

  The curl commands generated by Swagger may need some modification for use in your environment (for example, escaping quotes or adding the -k modifier in order to use HTTP URLs).
- A standard HTTP request for your call
- The server response code
- The response body – The JSON data that the API call returns
- Response headers

NOTE: If you use the SwaggerUI to try out API calls, and the API call returns an error (for example, your call uses incorrect syntax), under some circumstances SwaggerUI does not display the error, and the field for the returned data is empty.

Workload Optimization Manager markets

Workload Optimization Manager uses market-based analysis to perform workload management. To do this, it constructs a model of your environment, representing each entity as a buyer and seller in a market. You can access this model via a named market resource.

At any time, your Workload Optimization Manager appliance can have a number of markets in memory. To get a list of the current markets in your appliance, execute the following URL:

GET: /rest/markets

This listing returns the main real-time market, plus any plan markets that are in memory at the time of the call. Note that you can pass group UUIDs to set a scope for the call. In that case, the call returns only markets for plans that include the passed groups in their scope.

The real-time market

The real-time market performs analysis and workload management on your environment. You can use the real-time market to access entities and get current or historical data about them.

The real-time market displayName is Market. This market should always be in the RUNNING state. Under no circumstances should you use the API to stop this market.

Workload Optimization Manager performs discovery to populate the topology it manages (the collection of entities in the real-time market). For this reason, you should be careful not to delete entities from or add entities to the real-time market.

For the real-time market, you can safely execute POST, PUT, and DELETE calls to:

- Post a query filter to get filtered lists of actions, notifications, and stats for the real-time market
- Post to create placement policies in the real-time market
- Delete to remove placement policies from the real-time market
- Put to edit placement policies on the real-time market
- Post a scenario to the real-time market to run a plan

  When you run a plan scenario against the real-time market, Workload Optimization Manager creates a copy of the real-time market, and modifies the copy according to the plan scenario.

NOTE:
It is possible to execute POST or DELETE methods to modify the topology of the real-time market. However, it is highly unlikely that you would have a reason to do so. You should be aware that changes to the real-time market will affect the analysis Workload Optimization Manager performs. For example, if you remove a VM from the real-time market, Workload Optimization Manager can no longer manage its placement. The VM will still be present in your environment, but it will no longer be managed by Workload Optimization Manager. However, you can't reliably use this technique to make specific entities unavailable to Workload Optimization Manager. For the next discovery pass, Workload Optimization Manager will rediscover the deleted entity, and it will appear in the real-time market again.

Plan markets

Workload Optimization Manager can have markets other than the real-time market in memory. These other markets represent plans.

A plan market begins as a snapshot of the real-time market. You POST a scenario to the real-time market, and Workload Optimization Manager does two things:

- It makes a copy of the real-time market, to create a plan market. This plan market is just like the real-time market, except for any changes that were declared in the scenario. For example, if the scenario declares a scope for the plan, then the plan market only contains entities that are within that scope.
- It runs economic cycles (buy/sell cycles) against this plan market until there are no more meaningful improvements to be made. At this point the plan run is completed.

Note that once Workload Optimization Manager creates the plan market, that market stays in memory until you delete it. Also note that you can later apply a scenario to the plan market. This is how you run a plan on a plan. The logic flow is the same, and Workload Optimization Manager creates yet another plan market to run the analysis.

Internal-use plans

Workload Optimization Manager regularly runs plans to generate data that it displays in the GUI — The Cluster Capacity and Project Cluster Resources dashboards both display data generated by regularly-run plans.

A plan market that is for internal use includes the substring `_BasePlan` in the displayName. You should not modify these plan markets for any reason.

User-created plans

Users can create plans to run what-if scenarios in the environment, and the results are saved as a plan matket. At any time, each user account can have multiple plan markets loaded in the Workload Optimization Manager instance.

You identify planner markets by their names. A plan name is specified as `<PlanType><userName>_<arbitraryID>`.

For example, a valid plan name is `CLOUD_MIGRATION_cud_1493140514716`. To find out which user owns this plan, you can parse out the user's name and query the API for that user's account information.

When you get a plan market, the response includes information such as:

- `uuid`: The market's identifier.
- state: Whether the plan succeeded or not. The state can be one of:
  - `CREATED`
  - `READY_TO_START`
  - `RUNNING`
  - `COPYING`
  - `SUCCEEDED`
  - `STOPPING`
  - `STOPPED`
  - `USER_STOPPED`
  - `DELETING`

As long as a plan market is running, Workload Optimization Manager is still calculating the plan results. If the market is stopped, then the plan has been run and you can access data from this market to see the plan results. You can make a PUT call to the market to stop a plan that is running.

For a plan market, you can execute the same POST, PUT, and DELETE methods that you would execute on the real-time market. These include:

- POST a query filter to get filtered lists of actions, notifications, and stats for the plan market
- POST to create placement policies in the plan market
- DELETE to remove placement policies from the plan market

- PUT to edit placement policies on the plan market
- POST a scenario to the plan market to run a plan on plan

You can also safely execute the following methods on a plan market:

- DELETE a plan market
- PUT to save or stop a plan market

NOTE:
It's possible to stop a plan market through the GUI or the API. Also, a plan may have been stopped before it finished its calculations. In this case, the plan results will be incomplete.

Running plans

A plan is a simulation or what-if scenario that explores the results of possible changes to your environment. To run a plan, Workload Optimization Manager creates a snapshot copy of your real-time market and applies changes to it. It then uses the Economic Scheduling Engine to perform analysis on that plan market.

Before working with plans in the API, you should be familiar with plans via the GUI. You should know what plans can accomplish, and how to specify the plan settings such as:

- Plan scope
- Changes to worklaod in the environment (adding or removing VMs)
- Changes to supply (adding, removing, or reconfiguring hardware)
- Changes to placement (migrating to the cloud or a different cluster)
- Changes to constraints

To create a plan, you first specify a plan scenario. This is an object that contains all the plan settings. The scenario object contains an array of changes, and each change declares a setting that you want to make.

After you create a scenario, you POST it to a market. In most cases you will post it to the real-time market. When you post the scenario to the market, Workload Optimization Manager does two things:

- It makes a copy of the market, to create a plan market. This plan market is just like the original market, except for any changes that were declared in the scenario. For example, if the scenario declares a scope for the plan, then the plan market only contains entities that are within that scope.
- It runs economic cycles (buy/sell cycles) against this plan market until there are no more meaningful improvements to be made. At this point the plan run is completed.

Saving and deleting plan markets

As you run a plan, Workload Optimization Manager performs analysis and shows the results. It also serializes and saves the plan results so users can load the results into the GUI at a later time.

NOTE:
Updates to Workload Optimization Manager can change the data format for the saved plan results. In that event, then the results will be inconsistent. After each update to Workload Optimization Manager, you should regenerate all the plans you want to save, and save them again.

To delete a plan via the API, just execute a DELETE method on the plan market.

Creating scenarios

The first step for a plan is to create the scenario. Note that a scenario is entirely separate from the plan. You create a scenario and then POST it to a market to create the plan. But the scenario still exists in memory, and you can apply it to a different market if you wish. For example, you can apply the same scenario to the real-time market at different times.

To create a scenario, use the `POST: /rest/scenarios/{name}` method. This method takes three parameters:

- `scope`

  An array of group UUIDs. For a scope of multiple groups, you must specify groups of the same type.
- `projection_days`

  For a projection scenario, the days to mark each projection period
- `input`

  A ScenarioApiInputDTO that can specify all the settings for a scenario.

The `input` parameter, a ScenarioApiInputDTO, is an object that you can POST to the API to specify scenario settings. The API includes methods to create and delete scenarios. In most cases, you will create a scenario by POSTing or PUTing a ScenarioApiInputDTO with changes to declare the scenario settings you want. Note that the API includes methods you can use to PUT many of these settings into a specific scenario by passing properties directly in the URL. However, you can also make these settings with the ScenarioApiInputDTO. Learning to use this DTO is the most consistent way to create and modify a scenario.

Note that as you create a scenario you can POST all of its settings in the ScenarioApiInputDTO, or you can create an incomplete scenario, and then PUT other changes into the scenario at a different time.

Creating a projection scenario

Projection plans calculate infrastructure requirements into the future, so your environment can accommodate changes to workload requirements over time. In a single plan you can specify the scope of the plan, how far into the future to project, and by what increments of time. For example, you can project the requirements of a specific datacenter one year into the future, showing how requirements change at one-month intervals.

When you create a projection scenario via the API, you provide the projection periods as an array of the days from today that you want to project. For example, `0` for today, `30` for 30 days from today, `60` for 60 days from today, and so on.

To create a scenario for a projection plan:

- Scope the scenario

  A projection plan requires a scope to the scenario.
- Create a change in the plan of type `PROJECTION_PERIODS`

  This change identifes the plan periods that you want, as an array of days.

```
{
    "changes":[
    {
      "type”:"PROJECTION_PERIODS",
      "projectionDays":[0,30,60,90]
    }
  ]
}
```

- For any plan changes that you want to repeat on specific projection periods, specify a `projectionDays` array for that change. Note that the items in this array must match the `projectionDays` that you specified for the `PROJECTION_PERIODS` change.

  For example, assume you specified `"projectionDays":[0,30,60,90]` for the initial projection:
  - Valid: `"projectionDays":[0,30,60,90]`
  - Valid: `"projectionDays":[30,90]`
  - Not valid: `"projectionDays":[0,35,70,90]`

This listing shows a plan scenario that scopes the plan market, sets up a projection for four periods, and adds one VM to the inventory for two out of four projection periods:

```
{
   "displayName":"My Projection",
   "changes": [
     {
       "type": "SCOPE",
       "scope": [
         {
           "uuid": "5678a46e9716657be88b5e1217df91436e13e4ff",
           "className": "Group",
           "entitiesCount": 2,
           "groupType": "VirtualMachine",
         }
       ]
     },
     {
       "type": "PROJECTION_PERIODS",
       "projectionDays": [
         0,
         30,
         60,
         90
       ]
     },
     {
       "type": "ADDED",
       "targets": [
         {
           "uuid": "564d89f4-190b-de1b-cb28-3e2f2126ab48"
         }
       ],
       "value": "1",
       "projectionDays": [
         30
       ]
     }
   ]
}
```

Placement policies in plans

You can enable or disable placement policies in a plan. This is useful to see how the environment would change if you had a different set of constraints. It's important to note that a placement policy exists in the context of a market. To get a placement policy, query the market that contains it. To create a placement policy for a plan, you will create the poolicy and add it to the plan market.

If you want to enable or disable a placement policy in a plan, you specify that as a change in the scenario. Because the placement policy exists in a market, you have to know which market you will run the plan against, and get the policy from that market.

For example, if you will run the plan against the real-time market, you would:

- Query the real-time market for the policy you want
- Give that policy's UUID in the scenario change object
- In the scenario change object — Specify to enable or disable the policy
- When the scenario is complete, POST the scenario to the real-time market

Remember that when you run a plan, Workload Optimization Manager creates a new plan market. This plan market will contain a copy of the placement policy, and will enable or disable it, depending on your setting.

Working with policies in the API

Policies set business rules for resource allocation analysis, resource status display, and action execution.

You can use the Workload Optimization Manager API to manage the following two types of policies:

- Placement policy

  A placement policy is a set of rules that Workload Optimization Manager must satisfy when generating placement recommendations for on-prem and containerized workloads.

  For more information, see Placement Policies Requests *(on page 1148)*.

- Automation policy

  An automation policy is a set of rules that Workload Optimization Manager must satisfy when executing non-parking actions *(on page 835)* on public cloud or on-prem workloads, or changing settings that affect analysis and action generation.

  For more information, see Automation Policies Requests *(on page 1151)*.

In addition to these policy types, Workload Optimization Manager also supports parking policies.

A parking policy is a set of rules that Workload Optimization Manager must satisfy when enforcing parking actions (stop and start) on public cloud entities.

The parking policy API endpoints are currently under development. To create parking policies, use the Workload Optimization Manager UI. For more information, see Parking policies *(on page 1087)*. For information about using the API to list parkable entities or enforce parking actions on demand, see Parking endpoint *(on page 1282)*.

Placement policies requests

A placement policy is a set of rules that Workload Optimization Manager must satisfy when generating placement recommendations for on-prem and containerized workloads.

With these policies, Workload Optimization Manager can recommend placement actions that comply with your business rules. For example, a placement policy can constrain VMs to specific hosts. If a VM needs to move to a different host due to resource congestion, Workload Optimization Manager will generate an action to move the VM to one of the hosts defined in the policy.

Workload Optimization Manager discovers placement policies from your targets, and allows you to create your own policies.

A placement policy exists in the context of a market, either the real-time market or a plan market. Each market has its own set of placement policies. Even if a plan market is an exact copy of the real-time market, the plan market has its own placement policies, each with its own UUID.

Before you work with placement policies via the API, understand how they work in the user interface. You must understand the types of placement policies you can create, and the effect of each on market analysis. For more information, see Placement policies *(on page 1043)*.

The API supports the following types of placement policies:

- `AT_MOST_N`: Only the given number of consumers can run on a single member of the providers group. This is set in the `capacity` property of the policy object.
- `BIND_TO_GROUP`: The consumers can only run on members of the provider group.
- `BIND_TO_COMPLEMENTARY_GROUP`: The consumers cannot run on any members of the provider group.
- `MUST_RUN_TOGETHER`: These consumers must run on the same provider entity.
- `AT_MOST_N_BOUND`: Only the given number of consumers can run on a single member of the providers group, AND The consumers can only run on members of the provider group.
- `MERGE`: Remove cluster boundaries for the specified clusters.
- `BIND_TO_GROUP_AND_LICENSE`: Create a license group.

You can use the following requests to manage your placement policies with the Workload Optimization Manager API.

Creating a placement policy

When you get a policy, the returned object describes the policy type, as well as the consumer and provider groups. When you create a policy, you do not have to provide the full data. You provide:

- `buyerUuid`: The group of consumers for this policy.
- `sellerUuid`: The provider group for this policy.
- `type`: The policy type.
- `policyName`: A display name for this policy.
- `enabled`: Whether to enable the policy in the market. Can be `true` or `false`.
- `capacity`: For an AT_MOST_N or AT_MOST_N_BOUND policy, the number of consumers to allow on a provider entity.
- `mergeType`: For a MERGE policy, the type of clusters to merge. Can be one of `Cluster`, `StorageCluster`, DesktopPools, or `DataCenter`.
- `mergeUuids`: The groups that you want to merge. The group type must match the ``mergeType`.

Because placement policies exist in the context of a market, make a POST request to the `/markets` endpoint and specify the UUID of the market where you want to create a policy.

Example: `POST https://10.10.10.10/api/v3/markets/{market_UUID}/policies`

For example, to create a DON'T PLACE policy, include the following DTO in the request body:

```
{
  "buyerUuid": "f82dbbc2b3366052f3bc1ac8a68c9c06b0eb182a",
  "enabled": false,
  "policyName": "PolicyFromApi",
  "sellerUuid": "4a2f5f132ae690af147ccfd6ea9839e79da3db79",
  "type": "BIND_TO_COMPLEMENTARY_GROUP"
}
```

Getting the details of a specific policy

The single policy request gets the details of a specified policy, given that policy's UUID. After you get the UUIDs for policies that affect a certain scope, you can use this request to view details of those policies.

Example: `GET https://10.10.10.10/api/v3/policies/104:domain-c7:10.10.111.111`

Response: A `PolicyApiDTO` describing details about the policy, including the group information for the provider of the policy.

```
{
  "links":[
    {
      "rel":"self",
      "href":"https://10.10.10.10/api/v3/policies/104:domain-c7:10.10.111.111"
    }
  ],
  "uuid":"104:domain-c7:10.10.111.111",
  "displayName":"GROUP-DRS-LN-rule/Cluster1/10.10.111.111",
  "type":"BIND_TO_GROUP",
  "name":"GROUP-DRS-LN-rule/Cluster1/10.10.111.111",
  "enabled":true,
  "capacity":1000000000,
  "commodityType":"DrsSegmentationCommodity",
  "consumerGroup":{
    "uuid":"7843e134bab4fb841b9cb7dd98ef04acf1f36f7b",
    "displayName":"Buyers-Group-GROUP-DRS-vmGroup1/Cluster1/10.10.111.111",
    "className":"DiscoveredGroup",
    "groupType":"VirtualMachine",
    "isStatic":true,
    "logicalOperator":"AND"
```

```
  },
  "providerGroup":{
    "uuid":"597a35089bfe405291911f3191cf24234d4713d1",
    "displayName":"Sellers-Group-GROUP-DRS-hostGrp1/Cluster1/10.10.111.111",
    "className":"DiscoveredGroup",
    "groupType":"PhysicalMachine",
    "isStatic":true,
    "logicalOperator":"AND"
  }
}
```

Getting policies for a market

Referenced Endpoint: Policies (on page 1285)

To get a list of all current policies affecting a specified market, make a GET request to the `/markets/{market_uuid}/policies` endpoint. For a plan market, this is a list of policies that you created, enabled, disabled, or changed in the plan configuration. For the real-time market, this is a list of policies currently available on your appliance.

To get a single policy, include the UUID of the policy in the request.

Examples:

- All policies: `GET https://10.10.10.10/api/v3/markets/2481646239616/policies`
- One policy:

  `GET https://10.10.10.10/api/v3/markets/2481646239616/policies/_z3LKSTk0EemeycL7hWgJNw`

Response: A list of `PolicyApiDTO`s representing the policies requested.

```
[
   {
      "links":[
         {
            "rel":"self",
            "href":"https://10.10.10.10/api/v3/policies/_z3LxVzk0EemeycL7hWgJNw"
         }
      ],
      "uuid":"_z3LxVzk0EemeycL7hWgJNw",
      "displayName":"DalTestPolMerge",
      "type":"MERGE",
      "name":"DalTestPolMerge",
      "enabled":true,
      "commodityType":"ClusterCommodity",
      "mergeType":"Cluster",
      "mergeGroups":[
         {
            "uuid":"91ff08749fd278925d73b61b0f0f69d79c99fc3a",
            "displayName":"PMs_Beta DC6\\Cluster-1",
            "className":"Group",
            "groupType":"PhysicalMachine",
            "isStatic":true,
            "logicalOperator":"AND"
         },
         {
            "uuid":"8c17d8e79c3db85ed5d3868de622cb89187849ce",
            "displayName":"PMs_Beta DC6\\Cluster-2",
            "className":"Group",
            "groupType":"PhysicalMachine",
            "isStatic":true,
            "logicalOperator":"AND"
         }
      ]
```

```
   },
   {
      "links":[
         {
            "rel":"self",
            "href":"https://10.10.10.10/api/v3/policies/_z3LKSTk0EemeycL7hWgJNw"
         }
      ],
      "uuid":"_z3LKSTk0EemeycL7hWgJNw",
      "displayName":"CommSegmentation/SegmentManager__4T_7kwY-Ed-WUKbEYSVIDw_1551123454966",
      "type":"MUST_RUN_TOGETHER",
      "name":"CommSegmentation/SegmentManager__4T_7kwY-Ed-WUKbEYSVIDw_1551123454966",
      "enabled":false,
      "capacity":1000000000,
      "commodityType":"DrsSegmentationCommodity",
      "consumerGroup":{
         "uuid":"_z2M57Dk0EemeycL7hWgJNw",
         "displayName":"Buyers-Group-GROUP-DRS-KTG-rule/DC13-Cluster/vsphere-dc13.dev.mycorp.com",
         "className":"DiscoveredGroup",
         "groupType":"VirtualMachine",
         "isStatic":true,
         "logicalOperator":"AND"
      },
      "providerGroup":{
         "uuid":"_z2M57zk0EemeycL7hWgJNw",
         "displayName":"Sellers-Group-GROUP-DRS-KTG-rule/DC13-Cluster/vsphere-dc13.dev.mycorp.com",
         "className":"DiscoveredGroup",
         "groupType":"PhysicalMachine",
         "isStatic":true,
         "logicalOperator":"AND"
      }
   },
   ...
]
```


Deleting a placement policy for a market

Referenced Endpoint: Policies (on page 1285)

To delete a policy from a market, include the UUID of the policy and the market. Note that if you delete an imported policy, it reappears after the next discovery cycle of the target it was imported from. To delete imported policies, delete them on the target directly.

Example: `DELETE https://10.10.10.10/api/v3/markets/2481646239616/policies/_z3LxVzk0EemeycL7hWgJNw`

Response: A successful deletion request returns a Response Code of `200`.

Automation policies requests

An automation policy is a set of rules that Workload Optimization Manager must satisfy when executing non-parking actions *(on page 835)* on public cloud or on-prem workloads, or changing settings that affect analysis and action generation.

For more information about automation policies, see Automation policies *(on page 1049)*.

You can create and manage your automation policies with the Workload Optimization Manager API by making requests to the Settings policies endpoint *(on page 1366)* endpoint (`/settingspolicies`).

Getting settings policies

This call returns an array of SettingsPolicyApiDTOs for all the settings policies in your environment. You can pass entity types to limit the listing to policies for those types, and you can limit the list to only default policies.

This request takes the following parameters:

- only_defaults

  If this is `true`, the call only gets the default settings policies
- entity_type

  One or more entity types to limit the list of settings policies

Examples:

- Get all the settings policies: `GET https://10.10.10.10/api/v3/settingspolicies`
- Get all the default settings policies: `GET https://10.10.10.10/api/v3/settingspolicies?only_defaults=true`
- Get settings policies for VMs and Hosts: `GET https://10.10.10.10/api/v3/settingspolicies?entity_type=VirtualMachine&entity_type=PhysicalMachine`
- Get a single settings policy (pass the policy Uuid): `GET https://10.10.10.10/api/v3/settingspolicies/_aVDEwGegEemDFcIzuV5wcQ`

Response: An array of `SettingsPolicyApiDTO`s. This listing shows the settings policies for Container entities.


```
[
  {
    "uuid": "_vyMHEDnSEemHXrxYkw9y2Q",
    "displayName": "Container defaults",
    "entityType": "Container",
    "settingsManagers": [
      {
        "uuid": "marketsettingsmanager",
        "displayName": "Operational Constraints",
        "category": "Analysis",
        "settings": [
          {
            "uuid": "usedIncrement_Container_VMEM",
            "displayName": "Increment constant for Container VMem [MB]",
            "value": "64.0",
            "defaultValue": "64.0",
            "categories": [
              "resizeRecommendationsConstants"
            ],
            "valueType": "NUMERIC",
            "min": 0,
            "max": 10000000,
            "entityType": "Container",
            "sourceGroupName": "Global"
          },
          {
            "uuid": "usedIncrement_Container_VCPU",
            "displayName": "Increment constant for Container VCPU [MHz]",
            "value": "100.0",
            "defaultValue": "100.0",
            "categories": [
              "resizeRecommendationsConstants"
            ],
            "valueType": "NUMERIC",
            "min": 0,
            "max": 1000000,
            "entityType": "Container",
            "sourceGroupName": "Global"
```

```
        },
        {
          "uuid": "enforceConsistentResizingContainer",
          "displayName": "Consistent Resizing",
          "value": "false",
          "defaultValue": "false",
          "categories": [
            "resizeRecommendationsConstants"
          ],
          "valueType": "BOOLEAN",
          "entityType": "Container",
          "sourceGroupName": "Global"
        },
        {
          "uuid": "resizeTargetUtilization_Container",
          "displayName": "Resize Target Utilization",
          "value": "0.0",
          "defaultValue": "0.0",
          "valueType": "NUMERIC",
          "min": 0,
          "max": 100,
          "entityType": "Container",
          "sourceGroupName": "Global"
        }
      ]
    },
    ...
  }
]
```

Creating a settings policy

Referenced Endpoint: Groups (on page 1222)

To create a settings policy, you must do the following:

- Specify a policy name,
- Specify whether to enable or disable the policy in your market.
- Set the policy scope – Specify one or more groups of a given entity type. This determines the entity type for the policy.
- For the of settings you want to make, specify the appropriate settings manager, and then specify an array of setting/value pairs.
- Optionally, specify a schedule window and/or action execution schedule for the policy.

To set the policy scope, you provide an array of objects that contain group Uuids. These objects are effectively abbreviated GroupApiDTOs. Note that the groups must all be for the same type of entity. This specifies the entity type for the settings policy. Use the Groups endpoint to get the Uuid values for the groups you want in the scope. For example, to set the scope to a single group of VMs, use the following scopes object:

```
"scopes": [
    {
      "uuid": "_lSlr0GddEemDFcIzuV5wcQ"
    }
]
```

To specify the actual settings, you provide an array of SettingsManagerApiDTOs. Each one contains an array of settings that you want to make. To see all the settings for a given entity type, refer to the user interface. The setting managers you can choose from are:

- `actionscriptmanager`

  Contains settings related to the use of action scripts.
- `automationmanager`

  Contains settings related to action modes for entities in your environment.
- `cloudsavingsmanager`

  Contains settings related to execution retention for cloud entities.
- `clusterHeadroomConfig`

  Contains settings related to the headroom plans generated by the system.
- `controlmanager`

  Contains settings related to workflows that replace native action handling for generated actions in your environment.
- `emailmanager`

  Contains settings related to SMTP email notifications.
- `hcisettingsmanager`

  Contains settings related to managing hyperconverged infrastructure.
- `marketsettingsmanager`

  Contains settings related to market analysis.
- `osmigrationmanager`

  Contains settings realted to operating system migration for cloud entities.
- `persistencemanager`

  Contains settings related to statistic retention data and reporting.
- `reservedinstancemanager`

  Contains settings related to reserved instances.
- `storagesettingsmanager`

  Contains settings related to storage entity statistics and analysis.
- `telemetrymanager`

  Contains settings related to telemetry opt-in and terms of service.
-

For example, to set the VM start action to AUTOMATED and to set the VCPU maximum to 16, use the following `settingsManagers` array:

```
"settingsManagers": [
    {
      "uuid": "automationmanager",
      "settings": [
        {
          "uuid": "startVM",
          "value": "AUTOMATIC"
        }
      ]
    },
    {
      "uuid": "marketsettingsmanager",
      "settings": [
        {
          "uuid": "resizeVcpuMaximum_VM",
          "value": 16
```

```
        }
      ]
    }
]
```

You can set up a schedule window that determines when the settings policy takes effect. For that you provide a ScheduleApiDTO for the schedule details. It includes: the start and end dates, start and end times,

- Start and end dates.
- Start and end times.
- An optional RecurrenceApiDTO for schedule recurrence. If you don't provide this you will specify a one-time schedule window.

For example, use this to specify a "perpetual" schedule window (does not expire until 2099) that starts at 12:30 am and ends at 7:30 am every Saturday and Sunday.

```
"schedule": {
  "startDate": "2019-04-28T00:30:00-04:00",
  "endDate": "2099-12-31T07:30:00-05:00",
  "startTime": "2019-04-28T00:30:00-04:00",
  "endTime": "2019-04-28T07:30:00-04:00",
  "recurrence": {
    "type": "WEEKLY",
    "daysOfWeek": [
      "Sat",
      "Sun"
    ]
}
```

An example complete data payload for a policy with a policy-level schedule, action execution instruction, and action execution schedule is as follows:

```
{
   "disabled":false,
   "entityType":"VirtualMachine",
   "displayName":"TestPol",
   "scopes":[
      {
         "uuid":"285041619454560"
      }
   ],
   "settingsManagers":[
      {
         "uuid":"automationmanager",
         "displayName":"Action Mode Settings",
         "category":"Automation",
         "settings":[
            {
               "uuid":"storageMove",
               "value":"RECOMMEND",
               "valueType":"STRING",
               "valueObjectType":"String",
               "defaultValue":"RECOMMEND",
               "entityType":"VirtualMachine",
               "displayName":"Storage Move",
               "options":[
                  {
                     "label":"Disabled",
                     "value":"DISABLED"
                  },
                  {
```

```
                     "label":"Recommend",
                     "value":"RECOMMEND"
                  },
                  {
                     "label":"External Approval",
                     "value":"EXTERNAL_APPROVAL"
                  },
                  {
                     "label":"Manual",
                     "value":"MANUAL"
                  },
                  {
                     "label":"Automatic",
                     "value":"AUTOMATIC"
                  }
               ]
            },
            {
               "uuid":"storageMoveExecutionSchedule",
               "value":"285088800975856",
               "valueType":"LIST",
               "valueObjectType":"String",
               "entityType":"VirtualMachine",
               "displayName":"Execution window for Storage Move action"
            }
         ]
      }
   ],
   "schedule":{
      "uuid":"285088797821136",
      "displayName":"TestSched",
      "endTime":"2021-05-24T16:00",
      "startTime":"2021-05-24T14:30",
      "timeZone":"America/Chicago",
      "recurrence":{
         "type":"MONTHLY",
         "interval":1,
         "daysOfMonth":[
            1
         ]
      }
   }
}
```

Example: Using the previous payload, `POST https://10.10.10.10/api/v3/settingspolicies`

Response: The settings policy that you created. Use this response to get the policy Uuid.

Editing a settings policy

To edit a settings policy, you push a new set of data to the Settings Policies end point. The best way to do this is to copy the SettingsPolicyApiDTO that you want to edit, make changes to that DTO, and then push the DTO out to the API. To remove the policy from the scope, pass the SettingsPolicyApiDTO, and also pass `reset_defaults=true` as a parameter.

This request takes the following parameters:

- settingsPolicy_Uuid

  The Uuid of the policy you want to edit

- reset_defaults

  A boolean you can use to reset the policy to default. When `reset_defaults=true`, you remove the policy and reset that scope to default. When you use this setting, pass an empty object `{}` as the data payload.

Examples:

- Assuming you copied a SettingsPolicyApiDTO and then made some changes to it, and pass that at the payload: `PUSH https://10.10.10.10/api/v3/settingspolicies/_2-CbQGoNEemyOoIvFZyv1A?reset_defaults=false`
- Assuming you want to remove the policy (and so you pass an empty object as the payload): PUSH `https://10.10.10.10/api/v3/settingspolicies/_2-CbQGoNEemyOoIvFZyv1A?reset_defaults=true`

Response: The modified `SettingsPolicyApiDTO`.

Deleting a Settings Policy

Deletes the given settings policy.

Example: `DELETE https://10.10.10.10/api/v3/settingspolicies/_aV5ucGoTEemyOoIvFZyv1A`

Response: 200 for success.

Parking policies requests

A parking policy is a set of rules that Workload Optimization Manager must satisfy when enforcing parking actions (stop and start) on public cloud entities.

The parking policy API endpoints are currently under development. To create parking policies, use the Workload Optimization Manager UI. For more information, see Parking policies *(on page 1087)*. For information about using the API to list parkable entities or enforce parking actions on demand, see Parking endpoint *(on page 1282)*.

Calculating reservations and workload placement

Workload Optimization Manager includes the capability to reserve resources for VMs you plan to deploy in the future. Workload Optimization Manager calculates optimal placement for these VMs and then reserves the host and storage resources that they need.

To reserve VMs, you will choose a VM template, specify any placement constraints, set how many instances to reserve, and then indicate whether to reserve now or in the future. Because reserved VMs do not yet exist, they do not participate in the real-time market. VM templates specify the resource requirements for each reserved VM, including:

- Compute and storage resources allocated to each VM
- Consumed factor. This is the percentage of allocated CPU, memory, or storage that the reserved VM will utilize.

Note that before using the API to work with reservations, you should understand how reservations work from the user interface.

With the API, you can perform the following:

- Get list of reservations
- See the current placement for the reservation workloads
- Create reservations
- Delete a current reservation

Getting reservation information

To get a list of current reservations, execute `/rest/reservations`. This returns a list of all active reservations. If you know the UUID of the reservation you want, you can pass it to get data for just that reservation.

Each reservation object includes:

- Display name
- Status – can be:
  - DEPLOYING – Workload Optimization Manager is deploying the workload
  - DEPLOY_SUCCEEDED – The workload was successfully deployed
  - IN_PROGRESS – Placement calculation is in progress
  - PLACEMENT_SUCCEEDED – For a new reservation, the environment has sufficient resources to place the workload; if you specified a reserve date, this will be an active reservation
  - PLACEMENT_FAILED – For a newly created reservation, the environment doesn't have resources to place the workload; if you specified a reserve date, this will be an unfulfilled reservation
  - RETRYING – Workload Optimization Manager is trying to place the workload of an unfulfilled reservation
- Time the reservation was created, time to deploy, and time it will expire
- A description of the reserved workload
- The deployment profile that identifies the physical files that will be copied to deploy the workload, as well as optional placement limitations.
- Statistics for the compute and storage resources the reservation sets aside

For example, this listing shows a reservation for one VM:


```
{
  "uuid": "_kWZHIDDmEeePgeXuo0RRbw",
  "displayName": "MyReservation",
  "count": 1,
  "status": "PLACEMENT_SUCCEEDED",
  "reserveDateTime": "Thu May 04 16:27:29 UTC 2017",
  "expireDateTime": "Thu Aug 31 16:27:29 UTC 2017",
  "deployDateTime": "Thu Aug 31 16:27:29 UTC 2017",
  "reserveCount": 1,
  "demandEntities": [
    {
      "uuid": "_kWgb7TDmEeePgeXuo0RRbw",
      "displayName": "MyReservation_C0",
      "className": "VirtualMachine",
      "template": {
        "uuid": "T423f548d-cadc-e525-6df4-1f90724cf696",
        "displayName": "vsphere-dc3.dev.mycorp.com::TMP-SUSE64",
        "className": "VirtualMachineProfile"
      },
      "deploymentProfile": {
        "uuid": "_gHJ0ICXxEeePgeXuo0RRbw",
        "displayName": "DEP-SUSE64",
        "className": "ServiceCatalogItem"
      },
      "placements": {
        "computeResources": [
          {
            "stats": [
              {
                "name": "numOfCpu",
                "value": 1
              },
              {
                "name": "cpuSpeed",
                "value": 2603
              },
              {
                "name": "cpuConsumedFactor",
                "value": 0.5
```

```
              },
              {
                "name": "memorySize",
                "value": 2097152
              },
              {
                "name": "memoryConsumedFactor",
                "value": 0.75
              },
              {
                "name": "ioThroughput",
                "value": 0
              },
              {
                "name": "networkThroughput",
                "value": 0
              }
            ],
            "provider": {
              "uuid": "Virtual_ESX_42381da5-12fa-1e82-2f1c-887419380d43",
              "displayName": "hp-esx21.corp.vmturbo.com",
              "className": "PhysicalMachine"
            }
          }
        ],
        "storageResources": [
          {
            "stats": [
              {
                "name": "diskSize",
                "value": 18432.363
              },
              {
                "name": "diskIops",
                "value": 0
              }
            ],
            "provider": {
              "uuid": "10545c15-7687ef26",
              "displayName": "QS2:ESXDC3DS1",
              "className": "Storage"
            }
          }
        ]
      }
    }
  ]
}
```


Creating a reservation

To create a reservation, you POST an input DTO that defines the reservation's:

- `action`: The action type: RESERVATION
- `demandName`: The display name of the reservation – If you do not specify names for the added workloads, this will be the root name for new VMs
- `placementParameters`:
  - `constraintIDs`: An array of UUIDs for placement policies that will affect the calculated placement
  - `count`: The number of workloads to place
  - `entityNames`: An array of names for the placed VMs – The array length should equal `count`

  - `geographicRedundancy`: If `true` place the workloads on unique hosts, otherwise Workload Optimization Manager can place multiple workloads on the same host
  - `templateID`: The UUID of the template that you will use to place this workload — Note that the template must include a reference to the deployment profile that you specify in `deploymantParameters`
- `expireDateTime`: When to cancel a reservation if Workload Optimization Manager cannot place all the workloads by that date — for a RESERVATION action, Workload Optimization Manager automatically sets the deploy time to equal this time
- `reserveDateTime`: The time to calculate the workload placement and create the reservation — This time cannot be earlier than the time that you POST the reservation to the API

The following listing shows an input DTO that creates a reservation. It will place four workloads, and it gives specific names ot each one.

```
{
  "action": "RESERVATION",
  "demandName": "MyReservation",
  "expireDateTime": "2017-10-10T12:38:17+00:00",
  "parameters": [
    {
      "deploymentParameters": {
        "deploymentProfileID": "_c9CJMDDAEeePgeXuo0RRbw"
      },
      "placementParameters": {
        "geographicRedundancy": false,
        "count": 4,
        "entityNames": [
          "foo","bar","baz","bonk"
        ],
        "templateID": "_UKsnkJkSEeCHcOXEhzJExA"
      }
    }
  ],
  "reserveDateTime": "2017-05-04T18:22:12+00:00"
}
```

Async placement calculation

When you POST a reservation, Workload Optimization Manager runs a plan to calculate the optimal placement of the workloads. Depending on the size of the reservation, this can take a significant amount of time. The API calculates the placement asynchronously. If your script relies on the reservation, you should verify that the reservation has completed before moving on.

When you execute a reservation, the response to your POST shows that the calculation is in progress:

```
{
    "uuid": "_0dPrYTDlEeePgeXuo0RRbw",
    "displayName": "MyReservation",
    "count": 4,
    "status": "IN_PROGRESS",
    "reserveDateTime": "Thu May 04 18:22:12 UTC 2017",
    "expireDateTime": "Tue Oct 10 12:38:17 UTC 2017"
}
```

To examine the placement results, GET the reservation using the UUID that the API gives in the response.

Using templates

To create a reservation you specify a template that contains workload requirements. VM Templates specify the resources that will be available to the VM, including:

- VCPUs
- Virtual Memory
- Storage
- Network Throughput
- IOPS
- IO Throughput

Note that templates can be created by users, and Workload Optimization Manager also discovers templates that are created by the management services in your environment. For example, a hypervisor or a cloud service provider typically manages a number of its own templates – Workload Optimization Manager discovers these. You should never edit a discovered template.

To see whether a template is discovered, GET the template you're interested in and look for the `discovered` property. This will be `true` or `false`.

Workload Optimization Manager REST API endpoints

The REST API resources give you full access to the Workload Optimization Manager software. This is a complete API that exposes the full set of Workload Optimization Manager capabilities. In fact, the product user interface is implemented as a Workload Optimization Manager client that uses this API.

While the REST API resources are documented in the SwaggerUI, this section provides extra details for the more expressive resources in the API.

To access the Workload Optimization Manager REST SwaggerUI, open a web browser to:

`https://<Your_Workload Optimization Manager_URL>/apidoc`

Utility requests

In the Workload Optimization Manager API Guide, utility requests are the requests within an endpoint that do not return information about your environment, but provide benefits to other calls you may make. For example, `https://10.10.10.10/api/v3/settings/` is a utility request that returns a list of all possible settings managers. This information can be used to find settings on your instance, but the `/settings/` request itself does not provide any instance-specific information.

Actions endpoint

Instead of responding to thresholds, Workload Optimization Manager analyzes the operating conditions and constantly recommends actions to keep the entire environment within the desired state. If you execute these actions (or let Workload Optimization Manager execute them for you), the environment will maintain operating conditions that assure performance while also ensuring the lowest possible cost thanks to efficient utilization of your resources.

Using the actions endpoint, you can:

- Get a complete or filtered list of actions in your environment
- Accept or reject actions

- Get statistics related to a specified action
- Get notifications related to a specified action

Actions utility requests

Utility requests provide metadata that you can use to make related requests, assemble inputDTOs, or see the potential values of a particular DTO.

Getting the valid `actionMode` parameters for a specified entity and action

The actions endpoint contains a utility request that shows the possible values for the `actionMode` parameter for a specified entity and action type. These can be used to set or understand the possibilities for various action settings.

To specify an entity/action type, this request takes the following parameters:

- action_type
  - `START`:

    Start an entity, and add it to the Workload Optimization Manager market.
  - `MOVE`:

    Move an entity from one provider to another. For example, moving a VM between hosts, or a datastore between disk arrays.
  - `SUSPEND`:

    Suspend the entity to exclude its capacity from your environment. Suspend does not remove the entity from your environment, but the market does not include the entity in its analysis. Based on market conditions, analysis can recommend restarting suspended entities, which includes them in market analysis once again.
  - `ADD PROVIDER`:

    This is equivalent to the START action, but for storage entities.
  - `CHANGE`:

    Move a virtual machine's storage between datastores.
  - `PROVISION`:

    Add capacity to your environment.
  - `RECONFIGURE`:

    Analysis recommends this when a VM is not able to correct a misconfiguration via another type of action. For example, analysis can recommend a RECONFIGURE action when a placement policy requires a VM to move to a host on a particular cluster, but cross-cluster moves are not permitted.
  - `DELETE`:

    This is equivalent to the SUSPEND action, but for storage entities.
  - `RESERVE_ON_PM`:

    When using the reservation functionality, this is the action to reserve the resources of a VM on a particular host.
  - `RESERVE_ON_DS`:

    When using the reservation functionality, this is the action to reserve the resources of a VM on a particular datastore.
  - `RESIZE_FOR_EFFICIENCY`:

    A resize that reduces the capacity of a commodity, such as vCPU, or vMEM.
  - `RESIZE_FOR_PERFORMANCE`:

    A resize that increases the capacity of a commodity, such as vCPU, or vMEM.

- entity_type
  - APPLICATION
  - APPLICATION_SERVER
  - BUSINESS_ACCOUNT
  - BUSINESS_UNIT
  - CHASSIS
  - CLOUD_SERVICE
  - CONTAINER
  - CONTAINER_POD
  - DATABASE
  - DATABASE_SERVER
  - DATACENTER
  - DISK_ARRAY
  - IO_MODULE
  - LOAD_BALANCER
  - LOGICAL_POOL
  - NETWORK
  - PHYSICAL_MACHINE
  - STORAGE
  - STORAGE_CONTROLLER
  - SWITCH
  - VIRTUAL_APPLICATION
  - VIRTUAL_DATACENTER
  - VIRTUAL_MACHINE

Example: `GET https://10.10.10.10/api/v3/actions/availablemodes?action_type=PROVISION&entity_type=PhysicalMachine`

Response: An array of strings, where each string is a valid actionMode for the combination provided. In this example, a `PROVISION` action and the `PhysicalMachine` entity. An empty response body indicates that the requested entity does not have an action of that type, usually for practical reasons. For example, the Network entity cannot have a `MOVE` action.


```
[
  "RECOMMEND",
  "DISABLED",
  "MANUAL",
  "AUTOMATIC"
]
```


Getting a list of entry points for actions

The actions endpoint contains a utility request that shows the entry points for other action-related requests. These can be used to find the UUID and details relating to a specific action.

Example: `GET https://10.10.10.10/api/v3/actions`

Response: An array of `LinkApiDTO` objects, where each object contains a URL to get all actions for a specified market, entity, or group.

```
{
  "links":[
    {
      "rel":"Market actions",
      "href":"https://10.10.10.10/api/v3/markets/{uuid}/actions?ascending=true"
    },
    {
      "rel":"Entity actions",
      "href":"https://10.10.10.10/api/v3/entities/{uuid}/actions?ascending=false"
    },
    {
      "rel":"Group actions",
      "href":"https://10.10.10.10/api/v3/groups/{uuid}/actions?ascending=true"
    }
  ]
}
```

Actions requests

Action requests handle acceptance, rejection, filtering, and exploration of actions on your Workload Optimization Manager appliance.

If one or more of the entities in the scope of your request no longer exists in the topology, the ActionApiDTO for such entities will contain minimal information about the entities.

Getting a specified action

Gets the action details, including the creation time and target entity information, for a specified action UUID.

Example: `GET https://10.10.10.10/api/v3/actions/_6QsKkaNWEeiYWYj--U8W5g/details`

Response: The full `ActionApiDTO` for the requested action:

```
{
  "links":[
    {
      "rel":"self",
      "href":"https://10.10.10.10/api/v3/actions/_6QsKkaNWEeiYWYj--U8W5g"
    },
    {
      "rel":"self",
      "href":"https://10.10.10.10/api/v3/actions/2218532354592"
    }
  ],
  "uuid":"_6QsKkaNWEeiYWYj--U8W5g",
  "createTime":"2018-08-19T07:40:57-04:00",
  "actionType":"MOVE",
  "actionState":"PENDING_ACCEPT",
  "actionMode":"MANUAL",
  "details":"Move VirtualMachine Jboss-EAP-10.203 from dc17-host-01.mycorp.com to
 dc17-host-02.mycorp.com",
  "importance":7.4150376,
  "target":{
    "uuid":"421d8db4-3fc6-440b-0575-11383c020c6c",
    "displayName":"Jboss-EAP-10.203",
    "className":"VirtualMachine",
    "aspects":{
      "virtualMachineAspect":{
        "os":"CentOS 4/5/6/7 (64-bit)",
```

```
        "connectedNetworks":[
          {
            "uuid":"34f9f5364cd843aa1fe99900621ee3b51a7dc6e5",
            "displayName":"VM Network"
          }
        ],
        "numVCPUs":2,
        "ebsOptimized":false
      }
    },
    "environmentType":"ONPREM"
  },
  "currentEntity":{
    "uuid":"34313836-3333-5553-4537-33364e385146",
    "displayName":"dc17-host-01.mycorp.com",
    "className":"PhysicalMachine",
    "environmentType":"ONPREM"
  },
  "newEntity":{
    "uuid":"34313836-3333-5553-4537-33394e43424e",
    "displayName":"dc17-host-02.mycorp.com",
    "className":"PhysicalMachine",
    "environmentType":"ONPREM"
  },
  "risk":{
    "uuid":"_CizZoKN2EeiYWYj--U8W5g",
    "subCategory":"Efficiency Improvement",
    "description":"dc17-host-01.mycorp.com can be suspended for efficiency",
    "severity":"MINOR",
    "importance":0
  },
  "actionID":2218532354592
}
```


Getting a filtered list of actions

Gets a list of actions, given a scope UUID and filtered by an ActionApiInputDTO using one or more of the following criteria:

- actionModeList
  - `DISABLED`:

    Do not recommend or perform the action. When you disable an action, the user interface will no longer display actions of that type.
  - `RECOMMENDED`:

    Recommend the action so a user can execute it via the given hypervisor or by other external means
  - `MANUAL`:

    Recommend the action, and provide the option to execute that action through the Workload Optimization Manager user interface or an /actions API request.
  - `AUTOMATED`:

    You have directed Workload Optimization Manager to execute the action automatically.
  - `COLLECTION` **(UNUSED)**:

    Legacy action mode, no longer used.
- actionStateList
  - `DISABLED`:

    An action whose policy has been disabled.
  - `RECOMMEND`:

An action that cannot be automated due to policy settings or must be executed outside the system, such as the purchase of hardware.

  - `PENDING_ACCEPT`:

    When the action mode is set to manual, a recommended action that has not been accepted or cleared.

  - `ACCEPTED`:

    An action that has been accepted, but is not yet in progress.

  - `IN_PROGRESS`:

    An action that is being executed.

  - `SUCCEEDED`:

    An action that was accepted and completed successfully.

  - `FAILED`:

    An action that was accepted and did not complete successfully.

  - `REJECTED`:

    When the action mode is set to manual, an action that has been rejected by the user.

  - `CLEARED`:

    When the action mode is set to manual, an action that is no longer recommended by the market.

  - `ACCOUNTING`:

    For cloud entities, an action to resize in order to use a different instance template.

  - `QUEUED`:

    When more than 10 actions are to be executed for a single target, Workload Optimization Manager will place the 11th and all subsequent actions into a QUEUED state, and they will execute as the previous actions complete.

- actionTypeList
  - `START`:

    Start an entity, and add it to the Workload Optimization Manager market.

  - `MOVE`:

    Move an entity from one provider to another. For example, moving a VM between hosts, or a datastore between disk arrays.

  - `SUSPEND`:

    Suspend the entity to exclude its capacity from your environment. Suspend does not remove the entity from your environment, but the market does not include the entity in its analysis. Based on market conditions, analysis can recommend restarting suspended entities, which includes them in market analysis once again.

  - `ADD PROVIDER`:

    This is equivalent to the START action, but for storage entities.

  - `CHANGE`:

    Move a virtual machine's storage between datastores.

  - `PROVISION`:

    Add capacity to your environment.

  - `RECONFIGURE`:

    Analysis recommends this when a VM is not able to correct a misconfiguration via another type of action. For example, analysis can recommend a RECONFIGURE action when a placement policy requires a VM to move to a host on a particular cluster, but cross-cluster moves are not permitted.

  - `DELETE`:

    This is equivalent to the SUSPEND action, but for storage entities.

  - `RESERVE_ON_PM`:

When using the reservation functionality, this is the action to reserve the resources of a VM on a particular host.

  - `RESERVE_ON_DS`:

    When using the reservation functionality, this is the action to reserve the resources of a VM on a particular datastore.

  - `RESIZE_FOR_EFFICIENCY`:

    A resize that reduces the capacity of a commodity, such as vCPU, or vMEM.

  - `RESIZE_FOR_PERFORMANCE`:

    A resize that increases the capacity of a commodity, such as vCPU, or vMEM.

- cleared

  Default: `false`.

  When `true`, actions that were removed from the action list before being executed will be returned.

- costType
  - `Saving`

    An action that will decrease cloud spending.

  - `Investment`

    An action that will increase cloud spending.

- cumulative

  Default: `false`.

  When `true`, shows the savings for each action in a scope as a cumulative total, rather than showing the savings for individual actions.

- environmentType
  - `ONPREM`:

    Returns only actions related to entities that are part of your on-prem environment.

  - `CLOUD`:

    Returns only actions related to entities that are part of your cloud environment.

- groupBy

  You may group the actions returned by the following criteria: `[actionModes, actionStates, actionTypes, risk, riskSeverity, riskSubCategory]`

- hasReservedInstance

  Default: `false`.

  When `true`, only actions that are related to reserved instances will be returned.

- relatedEntityTypes

  Returns only actions related to entities of these types within the scope. For example, if your scope is a virtual datacenter and you specify VirtualMachine as a related entity type, the request will include only those actions related to VMs in the specified virtual datacenter.

  Choose from the following entity types: `[ Datacenter, PhysicalMachine, VirtualMachine, Storage, Application, Chassis, DiskArray, IOModule, StorageControl, Switch, VirtualDataCenter, VPod, DPod, Container, Database, DatabaseServerContainer, LogicalPool ]`

- riskSeverityList

  Returns only actions of the given severities: `[ Unknown, Normal, Minor, Major, Critical ]`

- riskSubCategoryList
- returns only actions of the given subcategories: `[ Performance Assurance, Efficiency Improvement, Prevention, Compliance ]`

The following commodities may be returned as the related commodity for an action:

- Access
- ActionPermit
- ActiveSessions
- ApplicationCommodity
- Ballooning
- Biclique
- BufferCommodity
- BurstBalance
- CPU
- CPUAllocation
- CPUProvisioned
- ClusterCommodity
- CollectionTime
- Connection
- Cooling
- Coupon
- CrossCloudMoveSvc
- CrossClusterMoveSvc
- DBCacheHitRate
- DBMem
- DISK_ARRAY_ACCESS
- DSPMAccessCommodity
- DataCenterCommodity
- DatastoreCommodity
- DrsSegmentationCommodity
- Extent
- Flow
- FlowAllocation
- HACommodity
- HOST_LUN_ACCESS
- Heap
- HotStorage
- IOThroughput
- ImageCPU
- ImageMem
- ImageStorage
- InstanceDiskSize
- InstanceDiskType
- LICENSE_ACCESS
- LicenseCommodity
- Mem
- MemAllocation
- MemProvisioned
- Move
- NetThroughput
- NetworkCommodity
- NetworkInterfaceCount
- NetworkPolicy
- NumDisk

- NumberConsumers
- NumberConsumersPM
- NumberConsumersStorage
- PORT_CHANNEL
- PoolCPU
- PoolMem
- PoolStorage
- Power
- ProcessingUnits
- Q1VCPU
- Q2VCPU
- Q3VCPU
- Q4VCPU
- Q5VCPU
- Q6VCPU
- Q7VCPU
- Q8VCPU
- Q16VCPU
- Q32VCPU
- Q64VCPU
- QNVCPU
- ResponseTime
- RightSizeDown
- RightSizeSVC
- RightSizeUp
- SLACommodity
- SameClusterMoveSvc
- SegmentationCommodity
- ServiceLevelCluster
- SoftwareLicenseCommodity
- Space
- Storage
- StorageAccess
- StorageAllocation
- StorageAmount
- StorageClusterCommodity
- StorageLatency
- StorageProvisioned
- Swapping
- TemplateAccess
- TenancyAccess
- Threads
- Transaction
- TransactionLog
- Unknown
- VAppAccess
- VCPU
- VCPULimitQuota
- VCPURequest
- VCPURequestQuota

- VDCCommodity
- VMPMAccessCommodity
- VMem
- VMemLimitQuota
- VMemRequest
- VMemRequestQuota
- VStorage
- Zone

Example: `POST https://10.10.10.10/api/v3/actions`

Example input: In this example, a list of actions that meet the following criteria will be returned and grouped by the `actionMode` of the action:

- Is in state `PENDING_ACCEPT`
- Is part of an on-prem environment
- Is an action for a virtual machine
- Is an action generated by the Workload Optimization Manager real-time market

```
{
  "actionInput":{
    "actionStateList":[
      "PENDING_ACCEPT"
    ],
    "environmentType":"ONPREM",
    "groupBy":[
      "actionModes"
    ]
  },
  "relatedType":"VirtualMachine",
  "scopes":[
    "Market"
  ]
}
```

Response: A list of `ActionApiDTO`s that meet the specified criteria.

For information about filtering actions within a specified time period, see *Filtering Actions by Start and End Time* in Actions endpoint tips *(on page 1171)*.

Accepting or rejecting a specified action

This request accepts or rejects an action with the specified UUID, and takes the following parameters:

- action_UUID (Required)

  The UUID of the action.
- accept (Required)

  When `true`, accepts the action. When `false`, rejects the action.

Example: `POST https://10.10.10.10/api/v3/actions/_QJCzNd3JEeij-LI4LMEROw?accept=false`

Response: `true`, if the action was accepted.

Actions endpoint tips

GroupBy field criteria

When using the POST requests to query actions and action statistics, there is a `groupBy` parameter that you may use to group statistics. The valid values for this parameter differ by entity type. Here is the complete list of valid options per entity:

> **NOTE:**
> When sending an inputDTO using this field, only the value is needed.
>
> **Example:**`"groupBy": "vmsByPMName"`

```
{
  "VirtualMachine":{
    "criteria":[
      {
        "filterType":"vmsByName"
      },
      {
        "filterType":"vmsByPMName"
      },
      {
        "filterType":"vmsByStorage"
      },
      {
        "filterType":"vmsByNetwork"
      },
      {
        "filterType":"vmsByApplication"
      },
      {
        "filterType":"vmsByDatabaseServer"
      },
      {
        "filterType":"vmsByDatabaseServerVersion"
      },
      {
        "filterType":"vmsByDC"
      },
      {
        "filterType":"vmsByVDC"
      },
      {
        "filterType":"vmsByDCnested"
      },
      {
        "filterType":"vmsByNumCPUs"
      },
      {
        "filterType":"vmsByMem"
      },
      {
        "filterType":"vmsByGuestName"
      },
      {
        "filterType":"vmsByAltName"
      },
      {
        "filterType":"vmsByClusterName"
      },
```

```
      {
        "filterType":"vmsByDiskArrayName"
      },
      {
        "filterType":"vmsByLogicalPoolName"
      },
      {
        "filterType":"vmsByProfileName"
      },
      {
        "filterType":"vmsByTag"
      },
      {
        "filterType":"vmsByState"
      },
      {
        "filterType":"vmsByBusinessAccountUuid"
      },
      {
        "filterType":"vmsByResourceGroupUuid"
      }
    ]
  },
  "VirtualDataCenter":{
    "criteria":[
      {
        "filterType":"vdcsByName"
      },
      {
        "filterType":"vdcsByVDCName"
      },
      {
        "filterType":"vdcsByTag"
      },
      {
        "filterType":"vdcsByState"
      }
    ]
  },
  "PhysicalMachine":{
    "criteria":[
      {
        "filterType":"pmsByName"
      },
      {
        "filterType":"pmsByStorage"
      },
      {
        "filterType":"pmsByNetwork"
      },
      {
        "filterType":"pmsBySwitch"
      },
      {
        "filterType":"pmsByNumVms"
      },
      {
        "filterType":"pmsByDC"
      },
      {
        "filterType":"pmsByMem"
      },
      {
        "filterType":"pmsByNumCPUs"
      },
```

```
    {
      "filterType":"pmsByVendorName"
    },
    {
      "filterType":"pmsByCPUModel"
    },
    {
      "filterType":"pmsByModel"
    },
    {
      "filterType":"pmsByTimezone"
    },
    {
      "filterType":"pmsByClusterName"
    },
    {
      "filterType":"pmsByTag"
    },
    {
      "filterType":"pmsByState"
    }
  ]
},
"Storage":{
  "criteria":[
    {
      "filterType":"storageByName"
    },
    {
      "filterType":"storageByTag"
    },
    {
      "filterType":"storageByVMs"
    },
    {
      "filterType":"storageByDC"
    },
    {
      "filterType":"storageByPMCluster"
    },
    {
      "filterType":"storageByState"
    }
  ]
},
"Application":{
  "criteria":[
    {
      "filterType":"appsByName"
    },
    {
      "filterType":"appsByTag"
    }
  ]
},
"ApplicationComponent":{
  "criteria":[
    {
      "filterType":"appSrvsByName"
    }
  ]
},
"WebServer":{
  "criteria":[
    {
```

```
        "filterType":"wbSrvsByName"
      }
    ]
  },
  "BusinessApplication":{
    "criteria":[
      {
        "filterType":"busAppsByName"
      }
    ]
  },
  "Database":{
    "criteria":[
      {
        "filterType":"databaseByName"
      },
      {
        "filterType":"databaseByTag"
      },
      {
        "filterType":"databaseByBusinessAccountUuid"
      },
      {
        "filterType":"databaseByResourceGroupUuid"
      }
    ]
  },
  "DatabaseServer":{
    "criteria":[
      {
        "filterType":"databaseServerByName"
      },
      {
        "filterType":"databaseServerByTag"
      },
      {
        "filterType":"databaseServerByBusinessAccountUuid"
      },
      {
        "filterType":"databaseServerByEngine"
      },
      {
        "filterType":"databaseServerByEdition"
      },
      {
        "filterType":"databaseServerByVersion"
      }
    ]
  },
  "VirtualApplication":{
    "criteria":[
      {
        "filterType":"vappsByName"
      }
    ]
  },
  "Cluster":{
    "criteria":[
      {
        "filterType":"clustersByName"
      },
      {
        "filterType":"clustersByTag"
      }
    ]
```

```
  },
  "DataCenter":{
    "criteria":[
      {
        "filterType":"datacentersByName"
      },
      {
        "filterType":"datacentersByTag"
      }
    ]
  },
  "Group":{
    "criteria":[
      {
        "filterType":"groupsByName"
      },
      {
        "filterType":"groupsByTargetName"
      }
    ]
  },
  "StorageCluster":{
    "criteria":[
      {
        "filterType":"storageClustersByName"
      }
    ]
  },
  "DiskArray":{
    "criteria":[
      {
        "filterType":"diskarrayByName"
      }
    ]
  },
  "Zone":{
    "criteria":[
      {
        "filterType":"zonsByName"
      }
    ]
  },
  "Region":{
    "criteria":[
      {
        "filterType":"regsByName"
      }
    ]
  },
  "Network":{
    "criteria":[
      {
        "filterType":"netsByName"
      }
    ]
  },
  "LoadBalancer":{
    "criteria":[
      {
        "filterType":"lbsByName"
      }
    ]
  },
  "Chassis":{
    "criteria":[
```

```
      {
        "filterType":"chasByName"
      }
    ]
  },
  "IOModule":{
    "criteria":[
      {
        "filterType":"ioModuleByName"
      }
    ]
  },
  "StorageController":{
    "criteria":[
      {
        "filterType":"storagecontrollerByName"
      }
    ]
  },
  "DPod":{
    "criteria":[
      {
        "filterType":"dpodByName"
      }
    ]
  },
  "VPod":{
    "criteria":[
      {
        "filterType":"vpodByName"
      }
    ]
  },
  "LogicalPool":{
    "criteria":[
      {
        "filterType":"logicalPoolByName"
      }
    ]
  },
  "Switch":{
    "criteria":[
      {
        "filterType":"switchByName"
      }
    ]
  },
  "Container":{
    "criteria":[
      {
        "filterType":"containersByName"
      },
      {
        "filterType":"containersByVMName"
      }
    ]
  },
  "ContainerPod":{
    "criteria":[
      {
        "filterType":"containerpodsByName"
      },
      {
        "filterType":"containerpodsByVMName"
      }
```

```
    ]
  },
  "ServiceEntity":{
    "criteria":[
      {
        "filterType":"workloadByBusinessAccountUuid"
      },
      {
        "filterType":"workloadByResourceGroupUuid"
      }
    ]
  },
  "Workload":{
    "criteria":[
      {
        "filterType":"workloadByBusinessAccountUuid"
      },
      {
        "filterType":"workloadByResourceGroupUuid"
      }
    ]
  },
  "ResourceGroup":{
    "criteria":[
      {
        "filterType":"resourceGroupByName"
      },
      {
        "filterType":"resourceGroupByUuid"
      },
      {
        "filterType":"resourceGroupByBusinessAccountUuid"
      },
      {
        "filterType":"resourceGroupByTag"
      }
    ]
  },
  "BusinessAccount":{
    "criteria":[
      {
        "filterType":"businessAccountByName"
      },
      {
        "filterType":"businessAccountByUuid"
      },
      {
        "filterType":"subBusinessAccountOfUuid"
      },
      {
        "filterType":"businessAccountValidationStatus"
      },
      {
        "filterType":"businessAccountTargetName"
      },
      {
        "filterType":"businessAccountCloudProvider"
      }
    ]
  },
  "BillingFamily":{
    "criteria":[
      {
        "filterType":"billingFamilyByName"
      }
```

```
    ]
  },
  "ViewPod":{
    "criteria":[
      {
        "filterType":"viewPodByName"
      }
    ]
  },
  "DesktopPool":{
    "criteria":[
      {
        "filterType":"desktopPoolByName"
      },
      {
        "filterType":"desktopPoolByViewPod"
      }
    ]
  },
  "BusinessUser":{
    "criteria":[
      {
        "filterType":"businessUserByName"
      },
      {
        "filterType":"businessUserByDesktopPool"
      }
    ]
  }
}
```

Filtering actions by time

When you make a GET request to the `/actions` endpoint, you can filter the list of returned actions based on a specified start time, end time, or both.

- The `startTime` parameter excludes any actions that were executed (succeeded or failed) before a specified time.
- The `endTime` parameter excludes any actions with a `createTime` value that is after a specified time.

When you define both these parameters, the response includes all current and historical actions that were pending or executed during the specified time period.

For example, the following GET request to the `markets/Market/actions` endpoint returns all actions that were created, pending, or executed between 6:00 and 11:00 on May 8, 2024.

Example: `GET https://10.10.10.10/api/v3/markets/Market/actions`

Example input:

```
{ "startTime": "2024-05-08T06:00:19Z",
  "endTime": "2024-05-08T011:00:19Z",
  ...
}
```

Although the response does not include any actions that were created after the `endTime`, it might include actions that were created before the `startTime`, if they are still pending as of the `startTime`. Therefore, this filtering includes all actions that are either pending or executed during the specified time period, regardless of when those actions were created.

Admin endpoint

The admin endpoint is a collection of requests that enable you to perform various administrative functions. To use the admin endpoint, you must be logged into Workload Optimization Manager as an administrator user.

Using the admin endpoint, you can:

- Export diagnostics locally, and send them to Support
- Get health check results to see if there are issues related to the monitoring and analysis of your environment
- Get or set proxy settings
- Get or set logging levels for various components of your environment
- Get the current product version information and check for updates
- Load existing configuration files for use in your environment

Admin requests

Exporting diagnostics

If you are experiencing problems with Workload Optimization Manager, your support representative might request that you export diagnostic data. You can export the data and then send it to the support representative as requested.

Example: `POST https://10.10.10.10/api/v3/admin/exports`

Response: `true` if successful, `false` if an error occurs while Workload Optimization Manager creates the export. The export file is saved in the `/tmp` directory on your instance with a file name that begins with `bkp-`.

Getting health check results

Get health check results regularly to see whether issues exist that are related to the monitoring and analysis of your environment.

Example: `GET https://10.10.10.10/api/v3/admin/health`

Response: `HealthCategoryResponseDTO` *(on page 1568)* with the health check results.


```
[
  {
    "healthCategory": "TARGET",
    "categoryDisplayName": "Targets",
    "categoryHealthState": "CRITICAL",
    "responseItems": [
      {
        "subcategory": "VALIDATION",
        "healthState": "MAJOR",
        "numberOfItems": 2,
        "recommendations": []
      },
      {
        "subcategory": "DISCOVERY",
        "healthState": "CRITICAL",
        "numberOfItems": 6,
        "recommendations": [
          {
            "errorType": "INTERNAL_PROBE_ERROR",
            "description": "{Description}"
          }
        ]
      },
      {
```

```
        "subcategory": "DISCOVERY",
        "healthState": "NORMAL",
        "numberOfItems": 23,
        "recommendations": []
      }
    ]
  },
  {
    "healthCategory": "ACTION",
    "categoryDisplayName": "Actions",
    "categoryHealthState": "NORMAL",
    "responseItems": [
      {
        "subcategory": "ANALYSIS",
        "healthState": "NORMAL",
        "numberOfItems": 0,
        "recommendations": []
      }
    ]
  },
  {
    "healthCategory": "COMPONENT",
    "categoryDisplayName": "Components",
    "categoryHealthState": "NORMAL",
    "responseItems": [
      {
        "subcategory": "CONTAINER PODS",
        "healthState": "NORMAL",
        "numberOfItems": 45,
        "recommendations": [
          {
            "errorType": "INFO",
            "description": "Pods that are in a running state."
          }
        ]
      }
    ]
  }
]
```


The response includes the following health check categories (`healthCategory`):

- `TARGET` – see whether issues exist with the targets you added, such as validation or discovery failures.
- `ACTION` – see whether issues exist with analyzing your environment, such as analysis timeouts.
- `COMPONENT` – see whether issues exist with Workload Optimization Manager components, such as container pods that are not running.

Ideally, each `healthCategory` shows `"categoryHealthState": "NORMAL"`. If you see a different value, review the issues and recommendations in the `responseItems` key.

Use the following parameters to filter the results:

- `category`

  Get results for a specific `healthCategory`.
- `health_state`

  Get results with a specific `categoryHealthState`. For example, choose `CRITICAL` to see issues that require immediate attention.

Getting and setting a proxy

Gets the current state of the proxy and proxy details (if enabled).

Example: `GET https://10.10.10.10/api/v3/admin/httpproxy`

Response: A `HttpProxyDTO` describing details of the proxy. If no proxy is configured, this request returns `"isProxyEnabled": false`.

```
{
  "isProxyEnabled":true,
  "proxyHost":"10.10.111.1",
  "userName":"UserName",
  "password":"PW"
}
```

To set the proxy, pass an `HttpProxyDTO`, making sure to include `"isProxyEnabled": true` if you want to activate the proxy immediately.

Example: `POST https://10.10.10.10/api/v3/admin/httpproxy`

Example httpProxyDTO:

```
{
  "isProxyEnabled":true,
  "proxyHost":"10.10.111.1",
  "portNumber":"33128",
  "userName":"ProxyUser",
  "password":"Password"
}
```

Response: `true` if successful, `false` if an error occurs while Workload Optimization Manager is setting the proxy.

Getting and setting logging levels by component

You can set the level of logging for different components of the Workload Optimization Manager platform. The default logging level for all components is `INFO`. Setting more verbose logging levels can greatly increase the disk space that is required to store the log files. You normally change these settings only while you're working with a Workload Optimization Manager support representative.

The available logging levels are (in order of least to most verbose): `INFO, WARN, DEBUG,` and `TRACE`.

Example: `GET https://10.10.10.10/api/v3/admin/logginglevels`

Response: A `LoggingApiDTO` with details about your current logging level settings:

```
 {
  "componentLoggingLevel":{
    "Extension":"INFO",
    "Analysis":"INFO",
    "Discovery":"INFO",
    "API":"INFO",
    "Presentation":"INFO",
    "Abstraction":"INFO",
    "Monitoring":"INFO"
  }
}
```

To set the logging level for a specific component, pass a partial LoggingApiDTO containing the component and logging level.

Example: `POST https://10.10.10.10/api/v3/admin/logginglevels`

Example LoggingApiDTO:

```
{
 "componentLoggingLevel":{
   "Extension":"WARN"
 }
}
```

Response: A successful response returns your input in the response body.

Getting the current version

Gets the version of each Workload Optimization Manager package that your instance is running. Set the `check_for_updates` parameter to `true` to also check for updates that might be available.

This request takes the following parameters:

```
check_for_updates
```

Check for available updates before returning version information. Default: `false`.

Example: `GET https://10.10.10.10/api/v3/admin/versions?check_for_updates=true`

Response: A `ProductVersionDTO` describing details about the current version and available updates. If you are running the latest version of Workload Optimization Manager, the updates value reads `"No Updates Available\n"`

```
{
 "versionInfo": "Turbonomic Operations Manager 6.1.7 (Build
 20180813223950000)\nvmt-platform-6.1.7-20180813223950000.i586
 ...,
 "updates": "Available Packages\nvmt-bundle.i586  6.2.0-20180817180112000 vmturbo\nvmt-config.x86_64
 ...,
 "marketVersion": 2
}
```

Loading external configuration files

To help with diagnosis of some issues, a support representative might want you to load configuration files into your Workload Optimization Manager appliance. Use this request only when you are working with a support representative, who can guide you through the process. The support representative specifies what kind of configuration file you are loading, and the contents of that file, by specifying the required parameters.

Example: `POST https://10.10.10.10/api/v3/admin/configfiles?config_type=TOPOLOGY&topology=abc`

Response: `true` if successful, `false` if an error occurs while you are uploading the configuration.

Authorization endpoint

You can use the authorization endpoint to create and manage OAuth 2.0 clients. You can then use these clients to authenticate applications to Workload Optimization Manager APIs. To use the authorization endpoint, you must be logged in to Workload Optimization Manager as an administrator user.

Using the authorization endpoint, you can:

- Create OAuth 2.0 clients
- List OAuth 2.0 clients
- Delete OAuth 2.0 clients

For more information about using OAuth 2.0 clients to authenticate with Workload Optimization Manager APIs, see Authenticating OAuth 2.0 clients with the API *(on page 1441)*.

Authorization requests

Creating an OAuth 2 client

You can create, manage, and authenticate OAuth 2.0 clients by using the Workload Optimization Manager API. For more information about prerequisites, audience configuration, and exchanging client credentials for an access token, see Authenticating OAuth 2.0 clients with the API *(on page 1441)*.

Example: `POST https://10.10.10.10/api/v3/authorization/oauth2/clients`

Request body:

```
{
  "clientName": "my-client",
  "grantTypes": [
    "client_credentials"
  ],
  "clientAuthenticationMethods": [
    "client_secret_post"
  ],
  "scopes": [
    "role:SITE_ADMIN"
  ],
  "tokenSettings": {
    "accessToken": {
      "ttlSeconds": 700
    }
  }
}
```

Specify the following values:

- `clientName`: Choose a name for your client.
- `scopes`: Specify exactly one value in the format `role:<ROLE>`, where `<ROLE>` is a Workload Optimization Manager user role, for example, `role:SITE_ADMIN`. Any access tokens that you generate for this client must have the same role. For more information about the available roles, see step 5 of Managing user accounts *(on page 1116)*.

The following parameters are optional:

- `clientAuthenticationMethods`: This value determines whether the client credentials must be supplied in a basic auth header or in the request body when you request an access token for the client. Specify one of the following values:
  - `client_secret_basic`: the default, which requires the credentials to be supplied as part of a basic auth header
  - `client_secret_post`: the credentials must be supplied in the request body
- `tokenSettings.accessToken.ttlSeconds`: This value determines the lifespan of access tokens that the OAuth 2.0 authorization server issues. The default value is `600` seconds.

Example response:

```
{
  "clientName": "my-client",
```

```
  "clientId": "7613c2ca-cee1-4445-aa99-55c9312852c8",
  "grantTypes": [
    "client_secret"
  ],
  "clientAuthenticationMethods": [
    "client_secret_post"
  ],
  "scopes": [
    "role:SITE_ADMIN"
  ],
  "audience": [
    "turbonomic"
  ],
  "createdAt": 1702042743,
  "tokenSettings": {
    "accessToken": {
      "ttlSeconds": 700
    }
  },
  "clientSecret": "b6YzO0zaRU6EPYNMZPBIbYhmWL1Jb6PTAKyBUTw7VIG8FQDh2ryCaMspoyh2cKzM"
}
```

Listing OAuth 2.0 clients

List the OAuth 2.0 clients in your Workload Optimization Manager instance.

Example: `GET https://10.10.10.10/api/v3/authorization/oauth2/clients`

Response: returns a list of OAuth2ClientApiDTOs for your instance, similar to the following example.

```
{
  "clientName": "my-client",
  "clientId": "7613c2ca-cee1-4445-aa99-55c9312852c8",
  "grantTypes": [
    "client_secret"
  ],
  "clientAuthenticationMethods": [
    "client_secret_post"
  ],
  "scopes": [
    "role:SITE_ADMIN"
  ],
  "audience": [
    "turbonomic"
  ],
  "createdAt": 1702042743,
  "tokenSettings": {
    "accessToken": {
      "ttlSeconds": 700
    }
}
```

Deleting an OAuth 2.0 client

Delete an OAuth 2.0 client from your Workload Optimization Manager instance.

Example: `DELETE https://10.10.10.10/api/v3/authorization/oauth2/clients/[clientId]`

A successful DELETE request does not return a response body. You can confirm that the request succeeded by running the same request again. The response to that request is `404 NOT_FOUND` because the target client was already deleted.

Any tokens that were issued for the deleted client are immediately invalidated and can no longer be used to access Workload Optimization Manager APIs.

Businessunits endpoint

Business units are container objects used for modeling accounts and subscriptions in Public Cloud targets, and for cloud budgets and price adjustements.

Using the businessunits endpoint of the Workload Optimization Manager API, you can:

- For a specified business unit, get a list of:
  - All actions relating to the entities in the business unit
  - Cloud services and pricing models
  - All entities belonging to the business unit
  - Price adjustments
  - Any related (parent, sibling, or child) business units
  - Statistics relating to entities in the business unit
  - Templates available
- Get the supply chain for a specified business unit
- Create a business unit
- Edit a business unit
- Edit a price adjustment
- Delete a business unit

Business units may be one of three types:

- BUDGET

  A business unit that has a budget, and may contain children.
- DISCOUNT

  A business unit that has a discount, and may contain children.
- DISCOVERED

  A business unit that has been discovered by the Workload Optimization Manager instance. These business units should not be edited or deleted, except on the discovered target.

Businessunits requests

Getting business units

Gets a list of all business units created on or discovered by your Workload Optimization Manager instance. This request takes the following parameters:

- type

  The type of business unit to return. `[ BUDGET, DISCOUNT, DISCOVERED ]`
- cloud_type

  The cloud provider whose business units to return. `[ AWS, AZURE, GCP ]`
- has_parent

  Whether to return only business units that have parent business units. Default: `False`. When `true`, all relevant business units will be returned irrespective of parent business units.

Example: `GET https://10.10.10.10/api/v3/businessunits?type=DISCOVERED&cloud_type=AWS`

Response: A list of `BusinessUnitApiDTO`s that meet the selected criteria. In this case, discovered business units that are part of an AWS environment.

```
[
  {
    "uuid": "323871187550",
    "displayName": "Product Trust",
    "className": "BusinessAccount",
    "environmentType": "CLOUD",
    "master": false,
    "hasRelatedTarget": true,
    "budget": {
      "value": 0
    },
    "businessUnitType": "DISCOVERED",
    "cloudType": "AWS",
    "targets": [
      {
        "uuid": "_6glgQEEhEemQ645gSx8RYw",
        "displayName": "PT-AWS",
        "type": "AWS"
      }
    ],
    "costPrice": 0,
    "severity": "Critical",
    "membersCount": 80,
    "memberType": "Workload",
    "related": {
      "rateCard": null,
      "priceAdjustment": null
    },
    "resourceGroupsCount": 0
  },
  {
    "uuid": "001844731978",
    "displayName": "Advanced Engineering",
    "className": "BusinessAccount",
    "environmentType": "CLOUD",
    "master": false,
    "hasRelatedTarget": false,
    "budget": {
      "value": 0
    },
    "businessUnitType": "DISCOVERED",
    "cloudType": "AWS",
    "costPrice": 0,
    "severity": "Normal",
    "membersCount": 0,
    "memberType": "Workload",
    "related": {
      "rateCard": null,
      "priceAdjustment": null
    },
    "resourceGroupsCount": 0
  },
  {
    "uuid": "192821421245",
    "displayName": "Development",
    "className": "BusinessAccount",
    "environmentType": "CLOUD",
    "master": true,
    "hasRelatedTarget": true,
    "budget": {
      "value": 0
    },
    "businessUnitType": "DISCOVERED",
```

```
    "cloudType": "AWS",
    "childrenBusinessUnits": [
      "631949720430",
      "001844731978",
      "323871187550"
    ],
    "targets": [
      {
        "uuid": "_cBVj0MvWEeiFJej-UwMSGg",
        "displayName": "engineering.aws.amazon.com",
        "type": "AWS"
      }
    ],
    "costPrice": 0,
    "severity": "Critical",
    "membersCount": 33,
    "memberType": "Workload",
    "related": {
      "rateCard": null,
      "priceAdjustment": null
    },
    "resourceGroupsCount": 0
  },
  {
    "uuid": "631949720430",
    "displayName": "Turbonomic SaaS",
    "className": "BusinessAccount",
    "environmentType": "CLOUD",
    "master": false,
    "hasRelatedTarget": false,
    "budget": {
      "value": 0
    },
    "businessUnitType": "DISCOVERED",
    "cloudType": "AWS",
    "costPrice": 0,
    "severity": "Normal",
    "membersCount": 0,
    "memberType": "Workload",
    "related": {
      "rateCard": null,
      "priceAdjustment": null
    },
    "resourceGroupsCount": 0
  }
]
```


Deleting a business unit

To delete a business unit, pass the UUID of the business unit in the request. Note, you cannot delete `DISCOVERED` business units.

Example: `DELETE https://10.10.10.10/api/v3/businessunits/49d50ac3f8adbfe649d53e72c42dd9428d1d3835`

Response: A response of `200` indicates successful deletion

Getting cloud service and pricing model pairs for a specified business unit

Gets the cloud service name and pricing model for each cloud service.

Example: `GET https://10.10.10.10/api/v3/businessunits/192821421245/cloudservices`

Response: A list of `CloudServicePricingModelApiDTO`s representing each cloud service available to the business unit:


```
[
  {
    "uuid": "aws::631949720430::CS::AWSCloudTrail",
    "displayName": "AWS CloudTrail",
    "pricingModel": "ON_DEMAND"
  },
  {
    "uuid": "aws::631949720430::CS::AmazonCloudWatch",
    "displayName": "AWS CloudWatch",
    "pricingModel": "ON_DEMAND"
  },
  {
    "uuid": "aws::192821421245::CS::AWSDeveloperSupport",
    "displayName": "AWS Developer Support",
    "pricingModel": "ON_DEMAND"
  },
  {
    "uuid": "aws::001844731978::CS::AmazonDynamoDB",
    "displayName": "AWS DynamoDB",
    "pricingModel": "ON_DEMAND"
  },
  {
    "uuid": "aws::631949720430::CS::AmazonEC2",
    "displayName": "AWS EC2",
    "pricingModel": "ON_DEMAND"
  },
  {
    "uuid": "aws::001844731978::CS::AmazonEKS",
    "displayName": "AWS EKS",
    "pricingModel": "ON_DEMAND"
  },
  ...
]
```


Getting price adjustments related to a specified business unit

Gets any price adjustments affecting the specified business unit.

Example: `GET https://10.10.10.10/api/v3/businessunits/192821421245/priceadjustments`

Response: A list of `BusinessUnitPriceAdjustmentApiDTO`s showing the corrected prices after adjustment:


```
[
  {
    "servicePriceAdjustments": [
      {
        "uuid": "aws::192821421245::CS::AmazonCloudWatch",
        "displayName": "AWS CloudWatch",
        "pricingModel": "ON_DEMAND"
      },
      {
        "uuid": "aws::192821421245::CS::AWSDeveloperSupport",
        "displayName": "AWS Developer Support",
        "pricingModel": "ON_DEMAND"
      },
      {
        "uuid": "aws::192821421245::CS::AmazonDynamoDB",
        "displayName": "AWS DynamoDB",
        "pricingModel": "ON_DEMAND"
```

```
    },
    {
      "uuid": "aws::192821421245::CS::AmazonEC2",
      "displayName": "AWS EC2",
      "pricingModel": "ON_DEMAND",
      "templateDiscounts": [
        {
          "uuid": "aws::VMPROFILE::g3.16xlarge",
          "displayName": "g3.16xlarge",
          "family": "g3",
          "pricesPerDatacenter": [
            {
              "uuid": "aws::ap-southeast-1::DC::ap-southeast-1",
              "displayName": "aws-Asia Pacific (Singapore)",
              "price": 6.68
            },
            {
              "uuid": "aws::us-west-2::DC::us-west-2",
              "displayName": "aws-US West (Oregon)",
              "price": 4.56
            },
            {
              "uuid": "aws::ca-central-1::DC::ca-central-1",
              "displayName": "aws-Canada (Central)",
              "price": 5.664
            },
            {
              "uuid": "aws::eu-west-1::DC::eu-west-1",
              "displayName": "aws-EU (Ireland)",
              "price": 4.84
            },
            {
              "uuid": "aws::eu-west-3::DC::eu-west-3",
              "displayName": "aws-EU (Paris)",
              "price": 0
            },
            {
              "uuid": "aws::ap-northeast-1::DC::ap-northeast-1",
              "displayName": "aws-Asia Pacific (Tokyo)",
              "price": 6.32
            },
            {
              "uuid": "aws::eu-central-1::DC::eu-central-1",
              "displayName": "aws-EU (Frankfurt)",
              "price": 5.7
            },
            {
              "uuid": "aws::ap-south-1::DC::ap-south-1",
              "displayName": "aws-Asia Pacific (Mumbai)",
              "price": 0
            },
            {
              "uuid": "aws::ap-northeast-2::DC::ap-northeast-2",
              "displayName": "aws-Asia Pacific (Seoul)",
              "price": 0
            },
            {
              "uuid": "aws::us-east-1::DC::us-east-1",
              "displayName": "aws-US East (N. Virginia)",
              "price": 4.56
            },
            {
              "uuid": "aws::us-west-1::DC::us-west-1",
              "displayName": "aws-US West (N. California)",
              "price": 6.136
```

```
            },
            {
              "uuid": "aws::sa-east-1::DC::sa-east-1",
              "displayName": "aws-South America (Sao Paulo)",
              "price": 0
            },
            {
              "uuid": "aws::ap-southeast-2::DC::ap-southeast-2",
              "displayName": "aws-Asia Pacific (Sydney)",
              "price": 7.016
            },
            {
              "uuid": "aws::us-east-2::DC::us-east-2",
              "displayName": "aws-US East (Ohio)",
              "price": 4.56
            },
            {
              "uuid": "aws::eu-north-1::DC::eu-north-1",
              "displayName": "aws-EU (Stockholm)",
              "price": 0
            },
            {
              "uuid": "aws::eu-west-2::DC::eu-west-2",
              "displayName": "aws-EU (London)",
              "price": 0
            }
          ]
        },
        {
          "uuid": "aws::VMPROFILE::p3.8xlarge",
          "displayName": "p3.8xlarge",
          "family": "p3",
          "pricesPerDatacenter": [
            {
              "uuid": "aws::ap-southeast-1::DC::ap-southeast-1",
              "displayName": "aws-Asia Pacific (Singapore)",
              "price": 16.936
            },
            {
              "uuid": "aws::us-west-2::DC::us-west-2",
              "displayName": "aws-US West (Oregon)",
              "price": 12.24
            },
            {
              "uuid": "aws::ca-central-1::DC::ca-central-1",
              "displayName": "aws-Canada (Central)",
              "price": 13.464
            },
            {
              "uuid": "aws::eu-west-1::DC::eu-west-1",
              "displayName": "aws-EU (Ireland)",
              "price": 13.22
            },
            {
              "uuid": "aws::eu-west-3::DC::eu-west-3",
              "displayName": "aws-EU (Paris)",
              "price": 0
            },
            {
              "uuid": "aws::ap-northeast-1::DC::ap-northeast-1",
              "displayName": "aws-Asia Pacific (Tokyo)",
              "price": 16.776
            },
            {
              "uuid": "aws::eu-central-1::DC::eu-central-1",
```

```
                "displayName": "aws-EU (Frankfurt)",
                "price": 15.292
              },
              {
                "uuid": "aws::ap-south-1::DC::ap-south-1",
                "displayName": "aws-Asia Pacific (Mumbai)",
                "price": 0
              },
              {
                "uuid": "aws::ap-northeast-2::DC::ap-northeast-2",
                "displayName": "aws-Asia Pacific (Seoul)",
                "price": 16.936
              },
              ...
            ]
          }
        ]
      }
    ]
  }
]
```

Editing price adjustments of a specified business unit

To edit a price adjustment for the specified business unit, send a BusinessUnitPriceAdjustmentApiDTO with the corrected information (in this case, changing the percent discount from 25 to 50.

Example: `PUT https://10.10.10.10/api/v3/businessunits/c0acb428d7b447a5d51f9888d2b6660b5fa5225f`

Example Input:

```
{
  "name": "Daltest1",
  "priceAdjustment": {
    "type": "DISCOUNT",
    "value": 50
  },
  "childrenBusinessUnits": [
    "192821421245"
  ],
  "businessUnitType": "DISCOUNT",
  "uuid": "c0acb428d7b447a5d51f9888d2b6660b5fa5225f"
}
```

Response: The full `BusinessUnitPriceAdjustmentApiDTO` with the corrected values:

```
{
  "uuid": "c0acb428d7b447a5d51f9888d2b6660b5fa5225f",
  "displayName": "Daltest1",
  "className": "BusinessAccount",
  "environmentType": "CLOUD",
  "discount": 50,
  "priceAdjustment": {
    "type": "DISCOUNT",
    "value": 50
  },
  "master": false,
  "hasRelatedTarget": false,
```

```
  "budget": {
    "value": 0
  },
  "businessUnitType": "DISCOUNT",
  "cloudType": "AWS",
  "childrenBusinessUnits": [
    "192821421245"
  ],
  "costPrice": 0,
  "severity": "Normal",
  "membersCount": 0,
  "memberType": "Workload",
  "related": {
    "rateCard": null,
    "priceAdjustment": {
      "uuid": "c0acb428d7b447a5d51f9888d2b6660b5fa5225f",
      "displayName": "Daltest1",
      "className": "BusinessAccount"
    }
  },
  "resourceGroupsCount": 0
}
```


Getting templates available to a business unit

Referenced Endpoint: Templates (on page 1406)

Gets details about each template available to the specified business unit:

Example: `GET https://10.10.10.10/api/v3/businessunits/192821421245/templates`

Response: A list of `TemplateApiDTO`s representing the available templates for the specified business unit:


```
[
  {
    "links": [
      {
        "rel": "self",
        "href": "https://10.10.10.10/api/v3/templates/B4166E9F-589B-357D-AC74-0A534B53CE85"
      }
    ],
    "uuid": "B4166E9F-589B-357D-AC74-0A534B53CE85",
    "displayName": "db.x1e.32xlarge",
    "className": "DatabaseServerProfile",
    "price": 0,
    "model": "db.x1e.32xlarge",
    "dbEngine": "Oracle",
    "dbEdition": "Enterprise",
    "computeResources": [
      {
        "stats": [
          {
            "name": "numOfCpu",
            "value": 128
          },
          {
            "name": "cpuSpeed",
            "units": "MHz",
            "value": 384000
          },
          {
            "name": "ioThroughputSize",
```

```
            "units": "MB/s",
            "value": 546.875
          },
          {
            "name": "memorySize",
            "units": "MB",
            "value": 3997696
          }
        ]
      }
    ],
    "storageResources": [
      {
        "stats": [
          {
            "name": "diskSize",
            "units": "GB",
            "value": 3840
          },
          {
            "name": "diskIops",
            "value": 35000
          }
        ]
      }
    ],
    "deploymentProfile": {
      "uuid": "5C57DCA5-72AB-34BF-A22D-06719D13182E",
      "displayName": "DEP-RDB-'ap-northeast-1'",
      "className": "ServiceCatalogItem",
      "deployParameters": [
        {
          "targetType": "AWS",
          "providers": [
            {
              "provider": {
                "discoveredBy": {
                  "uuid": "_RRn00PMSEeisB7K_chUWVw",
                  "displayName": "vmturbodev.aws.amazon.com",
                  "type": "AWS"
                }
              }
            }
          ]
        }
      ]
    },
    "discovered": true
  },
  {
    "links": [
      {
        "rel": "self",
        "href": "https://10.10.10.10/api/v3/templates/aws::VMPROFILE::g3.16xlarge"
      }
    ],
    "uuid": "aws::VMPROFILE::g3.16xlarge",
    "displayName": "g3.16xlarge",
    "className": "VirtualMachineProfile",
    "description": "g3.16xlarge",
    "computeResources": [
      {
        "stats": [
          {
            "name": "ioThroughputSize",
```

```
            "units": "MB/s",
            "value": 1750
          },
          {
            "name": "networkThroughputSize",
            "units": "MB/s",
            "value": 2500
          },
          {
            "name": "numOfCpu",
            "value": 64
          },
          {
            "name": "cpuSpeed",
            "units": "MHz",
            "value": 6664.6587
          },
          {
            "name": "cpuConsumedFactor",
            "units": "%",
            "value": 50
          },
          {
            "name": "memorySize",
            "units": "MB",
            "value": 499712
          },
          {
            "name": "memoryConsumedFactor",
            "units": "%",
            "value": 75
          },
          {
            "name": "ioThroughput",
            "units": "MB/s",
            "value": 0
          },
          {
            "name": "networkThroughput",
            "units": "MB/s",
            "value": 0
          }
        ]
      }
    ],
    "storageResources": [
      {
        "stats": [
          {
            "name": "diskSize",
            "units": "GB",
            "value": 976562.5
          },
          {
            "name": "diskIops",
            "value": 20000
          },
          {
            "name": "diskConsumedFactor",
            "units": "%",
            "value": 100
          }
        ]
      }
    ],
```

```
    "deploymentProfile": {
      "uuid": "aws::us-east-1::DP::DEP-fake-us-east-1",
      "displayName": "DEP-fake-us-east-1",
      "className": "ServiceCatalogItem",
      "deployParameters": [
        {
          "targetType": "AWS",
          "providers": [
            {
              "provider": {
                "discoveredBy": {
                  "uuid": "_RRn00PMSEeisB7K_chUWVw",
                  "displayName": "vmturbodev.aws.amazon.com",
                  "type": "AWS"
                }
              }
            }
          ]
        }
      ]
    },
    "discovered": true,
    "family": "g3"
  },
  ...
]
```


Entities endpoint

Entities are the objects in your topology that Workload Optimization Manager manages. For example, VMs, hosts, datastores, and applications are all entities. Workload Optimization Manager discovers entities via targets, it monitors and analyzes their utilization of resources, and it recommends actions to change the placement, configuration, or state of entities.

Each market manages a set of entities. Using a market's UUID, you can get a list of entities managed by that market. For each entity you can get a full range of data, including the resources it buys and sells, the providers it buys resources from, actions for the entity, and other associated information.

Remember that more than one market can be resident in memory at a given time – Workload Optimization Manager maintains a real-time market, and there can also be plan markets in memory. As you access market entities, you should keep the following in mind:

- Entities in the real-time market reflect the current state in your physical environment
- You should never add or remove entities in the real-time market
- Entities in a plan market reflect a snapshot of the environment from when the market was created
- Plan markets can be scoped to a subset of your physical environment
- You *can* add or remove entities in a plan market

An interesting point about entities and markets – A plan market can contain entities that are also managed by the real-time market. These entities have the same stats in both markets. Workload Optimization Manager independently performs analysis on these entities in both markets.

Using the entities endpoint of the Workload Optimization Manager API, you can get a list of entities (from the market or group that contains the entities). For an individual entity you can get:

- Actions and related data for an entity
- Constraints assigned to the entity
- Through the entity constraints, get the available providers and consumers for the entity
- Groups the entity belongs to

- Notifications and notification data
- Workload placement policies that affect this entity
- Automation policies that affect this entity
- Stats for the entity
- The supply chain of entities that stitch directly with this entity
- Any tags that are defined for the entity

This endpoint contains references to start and end times for certain operations. For more information, including valid time formats, see Time in the Workload Optimization Manager API *(on page 1141)*

Entities utility requests

Utility requests provide metadata that you can use to make related requests, assemble inputDTOs, or see the potential values of a particular DTO.

Getting a list of entry points for entities

The entities endpoint contains a utility request that shows the entry points you can use to get lists of entities. These can be used to find the UUID and details relating to specific entities.

Example: `GET https://10.10.10.10/api/v3/entities`

Response: An array of Link objects, where each object contains a URL to get all entities for a specified market or group.

```
{
  "links":[
    {
      "rel":"Market actions",
      "href":"https://10.10.10.10/api/v3/markets/{uuid}/entities"
    },
    {
      "rel":"Group actions",
      "href":"https://10.10.10.10/api/v3/groups/{uuid}/entities"
    }
  ]
}
```

Getting a list of entry points for a single entity

If you pass an entity UUID to the entities endpoint, then you get a description of the entity, plus a list of entry points for entity data such as stats, constraints, actions, etc.

Example: `GET https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073`

Response: An array of Link objects, where each object contains a URL to get information about the entity.

```
{
  "links": [
    {
     "rel": "self",
     "href":
 "https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073?include_aspects=false"
    },
    {
     "rel": "To Stats",
```

```
     "href": "https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073/stats"
    },
    {
     "rel": "Entity notifications",
     "href": "https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073/notifications"
    },
    {
      "rel": "Entity actions",
      "href":
 "https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073/actions?ascending=false"
    },
    {
      "rel": "To Constraints",
      "href": "https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073/policies"
    },
    {
      "rel": "To Settings",
      "href":
 "https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073/settings?include_settingspolici
es=false"
    },
    {
     "rel": "To Entity List",
     "href": "https://10.10.10.10/api/v3/entities"
    }
  ],
  ...
}
```

Entities requests

Getting entities

Referenced Endpoint: Groups (on page 1222)

To get a list of entities, start with the object that contains those entities. This is usually a market or a group. You work with individual entities via their UUIDs — A list of entities gives you a list of UUIDs that you can work with.

> **NOTE:**
> When working with groups, you should remember that a list of entities in the group is not the same as a list of group members. Assume a group of host clusters. In that case, the list of entities would be all the hosts in those clusters, while the group members would be a list of clusters, plus all the hosts in each cluster. For more information, see Groups Endpoint *(on page 1222)*.

Example: `GET https://10.10.10.10/api/v3/groups/_BZbFsWXUEemoItN1K8jw3Q/entities`

Response: A list of `EntityApiDTO`s.

```
[
  {
    "links": [
      ...
    ],
    "uuid": "azure::VM::2ffc038a-7af3-4d23-86a9-211da2e4d433",
    "displayName": "e47f9c97-b44d-4461-be02-0304b97e201e",
    "className": "VirtualMachine",
    "priceIndex": 1.0067123,
    "state": "ACTIVE",
```

```
    "severity": "Critical",
    "costPrice": 0.64534247,
    "discoveredBy": {
     "uuid": "_-S8GgFYoEemHGIeet0BOnQ",
     "displayName": "758ad253-cbf5-4b18-8863-3eed0825bf07",
     "type": "Azure"
    },
   "remoteId": "azure::VM::2ffc038a-7af3-4d23-86a9-211da2e4d433",
   "providers": [
      {
       "uuid": "azure::ST::eastus2-managed_premium",
       "displayName": "azure-eastus2-Managed Premium",
       "className": "Storage"
      },
      {
       "uuid": "azure::eastus2::PM::eastus2",
       "displayName": "azure-eastus2",
       "className": "PhysicalMachine"
      }
  ],
  "consumers": [
    {
     "uuid": "azure::APP::2ffc038a-7af3-4d23-86a9-211da2e4d433",
     "displayName": "GuestLoad [e47f9c97-b44d-4461-be02-0304b97e201e]",
     "className": "Application"
    }
  ],
  "template": {
    "uuid": "azure::VMPROFILE::Standard_DS13_v2",
    "displayName": "Standard_DS13_v2",
    "discovered": false,
    "enableMatch": false
  },
  "environmentType": "CLOUD",
  "tags": {
    "instance_group": [
      "diego_cell"
    ]
  }
 },
 {
    ...
  }
]
```


Getting actions related to an entity

Referenced Endpoint: Actions (on page 1161)

If you provide the entity UUID but do not provide an action UUID, this returns a list of ActionApiDTOs for all the actions that are current for the entity. If you provide both the entity UUID and the action UUID, it returns the ActionApiDTO for the specified action.

This request takes the following parameters:

- entity_Uuid

  The UUID of the entity whose actions will be returned.
- action_Uuid

  The UUID of the action you want to get – Required to get a single action.
- cursor

  When making subsequent calls to retrieve additional results, this is the cursor returned by the last call to this method.

- limit

  The maximum number of items to return. Must be a positive integer. If not set, the server will provide a default (20).

- order_by

  The field to order the results by. `[ creation_date, name, risk_category, savings, severity ]`.

- ascending

  Default: true. When false, results will be in descending order.

Examples:

- Get a list of actions for the passed entity: `GET https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073/actions?order_by=severity&ascending=true`
- Get a single action for the entity: `GET https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073/actions/9223304219064809841`

Response: A list of `ActionApiDTO`s representing the actions that are current for the given entity, or a single ActionApiDTO for the specified action. The following listing shows the list of ActionApiDTOs wrapped in an array.


```
[
  {
    "links": [
      {
        "rel": "self",
        "href": "https://10.10.10.10/api/v3/actions/9223304219106592017"
      },
      {
        "rel": "self",
        "href": "https://10.10.10.10/api/v3/actions/2560158594896"
      }
    ],
    "uuid": "9223304219106592017",
    "createTime": "2019-04-23T10:41:37-04:00",
    "actionType": "RIGHT_SIZE",
    "actionState": "PENDING_ACCEPT",
    "actionMode": "MANUAL",
    "details": "Scale down VMem for VirtualMachine vEOS-4 from 2 GB to 623 MB",
    "importance": 0,
    "target": {
      "uuid": "420b1eeb-6190-a9d2-6436-4327ad3ac073",
      "displayName": "vEOS-4",
      "className": "VirtualMachine",
      "discoveredBy": {
        "uuid": "_FJwg0EzHEemaXZFqO1sC0g",
        "displayName": "vsphere-dc11.dev.mycorp.com",
        "type": "vCenter"
      },
      "aspects": {
        "virtualMachineAspect": {
          "os": "Other 2.6.x Linux (32-bit)",
          "connectedNetworks": [
            {
              "uuid": "d2da5316adfff0bf779652bbe7eb76c3307c2ab5",
              "displayName": "Infrastructure"
            },
            {
              "uuid": "bbef68e6ee0e76ba2c82960012b4ba563b78fdbb",
              "displayName": "vmnet4"
            },
            {
              "uuid": "d59c5e7e738ad9d2e2cb45aac5fe0073d36b49fb",
              "displayName": "vmnet6"
            },
```

```
          {
            "uuid": "96a42a4d438d1336e8d2c6962c976d889792af8a",
            "displayName": "vmnet8"
          },
          {
            "uuid": "067bc9d882e9681c55d964295e43134711520145",
            "displayName": "vmnet57"
          },
          {
            "uuid": "05b0ce5abc5cecc4104b26514b11618d77818c6d",
            "displayName": "vmnet59"
          }
        ],
        "numVCPUs": 1,
        "type": "VMEntityAspectApiDTO"
      }
    },
    "environmentType": "ONPREM"
  },
  "currentEntity": {
    "uuid": "_3vIfwmXVEemoItN1K8jw3Q",
    "className": "VMem"
  },
  "newEntity": {
    "uuid": "_5fnl82XVEemoItN1K8jw3Q",
    "className": "VMem",
    "aspects": {
      "virtualMachineAspect": {
        "os": "Linux (Free)",
        "type": "VMEntityAspectApiDTO"
      }
    }
  },
  "currentValue": "2097152.0",
  "newValue": "637952.0",
  "resizeToValue": "637952.0",
  "risk": {
    "uuid": "_5fr3YWXVEemoItN1K8jw3Q",
    "subCategory": "Efficiency Improvement",
    "description": "Underutilized VMem in Virtual Machine 'vEOS-4'",
    "severity": "MINOR",
    "importance": 0,
    "reasonCommodity": "VMem"
  },
  "stats": [
    {
      "name": "costPrice",
      "filters": [
        {
          "type": "savingsType",
          "value": "savings"
        }
      ],
      "units": "$",
      "value": 69.58008
    }
  ],
  "currentLocation": {
    "uuid": "98e12595449d993db8861c1f94c8f86ba277f17b",
    "displayName": "Cloud",
    "className": "DataCenter"
  },
  "newLocation": {
    "uuid": "98e12595449d993db8861c1f94c8f86ba277f17b",
    "displayName": "Cloud",
```

```
      "className": "DataCenter"
    },
    "actionID": 2560158594896
  },
  ...
]
```

Getting a filtered list of actions related to an entity

Referenced Endpoint: *Actions (on page 1161)*

Gets a list of actions, given a market UUID and filtered by an ActionApiInputDTO using one or more of the following criteria:

- actionModeList
  - `DISABLED`:

    Do not recommend or perform the action. When you disable an action, the user interface will no longer display actions of that type.
  - `RECOMMENDED`:

    Recommend the action so a user can execute it via the given hypervisor or by other external means
  - `MANUAL`:

    Recommend the action, and provide the option to execute that action through the Workload Optimization Manager user interface or an /actions API request.
  - `AUTOMATED`:

    You have directed Workload Optimization Manager to execute the action automatically.
  - `COLLECTION` **(UNUSED)**:

    Legacy action mode, no longer used.
- actionStateList
  - `DISABLED`:

    An action whose policy has been disabled.
  - `RECOMMEND`:

    An action that cannot be automated due to policy settings or must be executed outside the system, such as the purchase of hardware.
  - `PENDING_ACCEPT`:

    When the action mode is set to manual, a recommended action that has not been accepted or cleared.
  - `ACCEPTED`:

    An action that has been accepted, but is not yet in progress.
  - `IN_PROGRESS`:

    An action that is being executed.
  - `SUCCEEDED`:

    An action that was accepted and completed successfully.
  - `FAILED`:

    An action that was accepted and did not complete successfully.
  - `REJECTED`:

    When the action mode is set to manual, an action that has been rejected by the user.
  - `CLEARED`:

    When the action mode is set to manual, an action that is no longer recommended by the market.
  - `ACCOUNTING`:

For cloud entities, an action to resize in order to use a different instance template.

  - `QUEUED`:

    When more than 10 actions are to be executed for a single target, Workload Optimization Manager will place the 11th and all subsequent actions into a QUEUED state, and they will execute as the previous actions complete.

- actionTypeList
  - `START`:

    Start an entity, and add it to the Workload Optimization Manager market.
  - `MOVE`:

    Move an entity from one provider to another. For example, moving a VM between hosts, or a datastore between disk arrays.
  - `SUSPEND`:

    Suspend the entity to exclude its capacity from your environment. Suspend does not remove the entity from your environment, but the market does not include the entity in its analysis. Based on market conditions, analysis can recommend restarting suspended entities, which includes them in market analysis once again.
  - `ADD PROVIDER`:

    This is equivalent to the START action, but for storage entities.
  - `CHANGE`:

    Move a virtual machine's storage between datastores.
  - `PROVISION`:

    Add capacity to your environment.
  - `RECONFIGURE`:

    Analysis recommends this when a VM is not able to correct a misconfiguration via another type of action. For example, analysis can recommend a RECONFIGURE action when a placement policy requires a VM to move to a host on a particular cluster, but cross-cluster moves are not permitted.
  - `DELETE`:

    This is equivalent to the SUSPEND action, but for storage entities.
  - `RESERVE_ON_PM`:

    When using the reservation functionality, this is the action to reserve the resources of a VM on a particular host.
  - `RESERVE_ON_DS`:

    When using the reservation functionality, this is the action to reserve the resources of a VM on a particular datastore.
  - `RESIZE_FOR_EFFICIENCY`:

    A resize that reduces the capacity of a commodity, such as vCPU, or vMEM.
  - `RESIZE_FOR_PERFORMANCE`:

    A resize that increases the capacity of a commodity, such as vCPU, or vMEM.
- cleared

  Default: `false`.

  When `true`, actions that were removed from the action list before being executed will be returned.
- costType
  - `Saving`

    An action that will decrease cloud spending.
  - `Investment`

    An action that will increase cloud spending.
- cumulative

Default: `false`.

When `true`, shows the savings for each action in a scope as a cumulative total, rather than showing the savings for individual actions.

- environmentType
  - `ONPREM`:

    Returns only actions related to entities that are part of your on-prem environment.
  - `CLOUD`:

    Returns only actions related to entities that are part of your cloud environment.
- groupBy

  You may group the actions returned by the following criteria: `[actionModes, actionStates, actionTypes, risk, riskSeverity, riskSubCategory]`
- hasReservedInstance

  Default: `false`.

  When `true`, only actions that are related to reserved instances will be returned.
- relatedEntityTypes

  Returns only actions related to entities of these types within the scope. For example, if your scope is a virtual datacenter and you specify VirtualMachine as a related entity type, the request will include only those actions related to VMs in the specified virtual datacenter.

  Choose from the following entity types: `[ Datacenter, PhysicalMachine, VirtualMachine, Storage, Application, Chassis, DiskArray, IOModule, StorageControl, Switch, VirtualDataCenter, VPod, DPod, Container, Database, DatabaseServerContainer, LogicalPool ]`
- riskSeverityList

  Returns only actions of the given severities: `[ Unknown, Normal, Minor, Major, Critical ]`
- riskSubCategoryList
- returns only actions of the given subcategories: `[ Performance Assurance, Efficiency Improvement, Prevention, Compliance ]`

This request takes the following parameters:

- entity_Uuid

  The UUID of the market whose actions will be returned.
- cursor

  When making subsequent calls to retrieve additional results, this is the cursor returned by the last call to this method.
- limit

  The maximum number of items to return. Must be a positive integer. If not set, the server will provide a default (20).
- order_by

  The field to order the results by: `[ creation_date, name, risk_category, savings, severity ]`.
- ascending

  Default: `true`. When `false`, results will be in descending order.

Example: `POST https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073/actions?order_by=severity&ascending=true`

Example input: In this example, a list of actions that meet the following criteria will be returned and grouped by the `actionMode` of the action, in order of severity:

- Is in state `PENDING_ACCEPT`
- Is part of an on-prem environment
- Is an action for a virtual machine

```
{
  "actionInput":{
    "actionStateList":[
      "PENDING_ACCEPT"
    ],
    "environmentType":"ONPREM",
    "groupBy":[
      "actionModes"
    ]
  },
  "relatedType":"VirtualMachine",
  "scopes":[
    "Market"
  ]
}
```

Response: A list of `ActionApiDTO`s that meet the specified criteria.

Getting statistics for a filtered list of actions of an entity

Referenced Endpoint: Actions (on page 1161)

Referenced Endpoint: Stats (on page 1372)

Gets a list of statistics for a filtered list of actions that are current for the passed entity. Statistics returned include the number of actions, total savings/investment, and the number of entities affected by the actions. In addition to the Entity UUID, provide an ActionApiInputDTO using one or more of the following criteria:

- actionModeList
  - `DISABLED`:

    Do not recommend or perform the action. When you disable an action, the user interface will no longer display actions of that type.
  - `RECOMMENDED`:

    Recommend the action so a user can execute it via the given hypervisor or by other external means
  - `MANUAL`:

    Recommend the action, and provide the option to execute that action through the Workload Optimization Manager user interface or an /actions API request.
  - `AUTOMATED`:

    You have directed Workload Optimization Manager to execute the action automatically.
  - `COLLECTION` **(UNUSED)**:

    Legacy action mode, no longer used.
- actionStateList
  - `DISABLED`:

    An action whose policy has been disabled.
  - `RECOMMEND`:

An action that cannot be automated due to policy settings or must be executed outside the system, such as the purchase of hardware.

  - `PENDING_ACCEPT`:

    When the action mode is set to manual, a recommended action that has not been accepted or cleared.

  - `ACCEPTED`:

    An action that has been accepted, but is not yet in progress.

  - `IN_PROGRESS`:

    An action that is being executed.

  - `SUCCEEDED`:

    An action that was accepted and completed successfully.

  - `FAILED`:

    An action that was accepted and did not complete successfully.

  - `REJECTED`:

    When the action mode is set to manual, an action that has been rejected by the user.

  - `CLEARED`:

    When the action mode is set to manual, an action that is no longer recommended by the market.

  - `ACCOUNTING`:

    For cloud entities, an action to resize in order to use a different instance template.

  - `QUEUED`:

    When more than 10 actions are to be executed for a single target, Workload Optimization Manager will place the 11th and all subsequent actions into a QUEUED state, and they will execute as the previous actions complete.

- actionTypeList
  - `START`:

    Start an entity, and add it to the Workload Optimization Manager market.

  - `MOVE`:

    Move an entity from one provider to another. For example, moving a VM between hosts, or a datastore between disk arrays.

  - `SUSPEND`:

    Suspend the entity to exclude its capacity from your environment. Suspend does not remove the entity from your environment, but the market does not include the entity in its analysis. Based on market conditions, analysis can recommend restarting suspended entities, which includes them in market analysis once again.

  - `ADD PROVIDER`:

    This is equivalent to the START action, but for storage entities.

  - `CHANGE`:

    Move a virtual machine's storage between datastores.

  - `PROVISION`:

    Add capacity to your environment.

  - `RECONFIGURE`:

    Analysis recommends this when a VM is not able to correct a misconfiguration via another type of action. For example, analysis can recommend a RECONFIGURE action when a placement policy requires a VM to move to a host on a particular cluster, but cross-cluster moves are not permitted.

  - `DELETE`:

    This is equivalent to the SUSPEND action, but for storage entities.

  - `RESERVE_ON_PM`:

When using the reservation functionality, this is the action to reserve the resources of a VM on a particular host.

  - `RESERVE_ON_DS`:

    When using the reservation functionality, this is the action to reserve the resources of a VM on a particular datastore.

  - `RESIZE_FOR_EFFICIENCY`:

    A resize that reduces the capacity of a commodity, such as vCPU, or vMEM.

  - `RESIZE_FOR_PERFORMANCE`:

    A resize that increases the capacity of a commodity, such as vCPU, or vMEM.

- cleared

  Default: `false`.

  When `true`, actions that were removed from the action list before being executed will be returned.

- costType
  - `Saving`

    An action that will decrease cloud spending.

  - `Investment`

    An action that will increase cloud spending.

- cumulative

  Default: `false`.

  When `true`, shows the savings for each action in a scope as a cumulative total, rather than showing the savings for individual actions.

- environmentType
  - `ONPREM`:

    Returns only actions related to entities that are part of your on-prem environment.

  - `CLOUD`:

    Returns only actions related to entities that are part of your cloud environment.

- groupBy

  You may group the actions returned by the following criteria: `[actionModes, actionStates, actionTypes, risk, riskSeverity, riskSubCategory]`

- hasReservedInstance

  Default: `false`.

  When `true`, only actions that are related to reserved instances will be returned.

- relatedEntityTypes

  Returns only actions related to entities of these types within the scope. For example, if your scope is a virtual datacenter and you specify VirtualMachine as a related entity type, the request will include only those actions related to VMs in the specified virtual datacenter.

  Choose from the following entity types: `[ Datacenter, PhysicalMachine, VirtualMachine, Storage, Application, Chassis, DiskArray, IOModule, StorageControl, Switch, VirtualDataCenter, VPod, DPod, Container, Database, DatabaseServerContainer, LogicalPool ]`

- riskSeverityList

  Returns only actions of the given severities: `[ Unknown, Normal, Minor, Major, Critical ]`

- riskSubCategoryList
- returns only actions of the given subcategories: `[ Performance Assurance, Efficiency Improvement, Prevention, Compliance ]`

Example: `POST https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073/actions/stats`

Example input: Get a list of stats for actions that meet the following criteria, grouped by the `actionMode` of the action, in order of severity:

- Is in state `PENDING_ACCEPT`
- Is part of an on-prem environment
- Is an action for a virtual machine
- Is an action generated by the Workload Optimization Manager real-time market

```
{
  "actionInput":{
    "actionStateList":[
      "PENDING_ACCEPT"
    ],
    "environmentType":"ONPREM",
    "groupBy":[
      "actionModes"
    ]
  },
  "relatedType":"VirtualMachine"
}
```

Response: A list of `StatSnapshotApiDTO`s that aggregate statistics about the filtered actions.

```
[
  {
    "date": "2019-04-24T03:56:34-04:00",
    "statistics": [
      {
        "name": "numActions",
        "values": {
          "max": 1,
          "min": 1,
          "avg": 1,
          "total": 1
        },
        "value": 1
      },
      {
        "name": "costPrice",
        "filters": [
          {
            "type": "property",
            "value": "savings"
          }
        ],
        "units": "$",
        "values": {
          "max": 69.77539,
          "min": 69.77539,
          "avg": 69.77539,
          "total": 69.77539
        },
        "value": 69.77539
      },
      {
        "name": "numEntities",
        "values": {
          "max": 1,
          "min": 1,
```

```
          "avg": 1,
          "total": 1
        },
        "value": 1
      }
    ]
  }
]
```

Getting the Constraints for placement of an entity

When it calculates where to place an entity, Workload Optimization Manager respects whatever placement constraints there might be in your environment. For example, unless you have a Merge Cluster policy in effect, VM placement is constrained to hosts in the same cluster.

This call gets a ConstraintApiDTO object, which describes these constraints for the specified entity. The data in this object is equivalent to the Placement Constraints view that you can see in the user interface. To see this view, you drill down to the policy view for a single entity. In this view you can see the entity's related providers and consumers. You can also open a view for each related entity to show the constraints that limit potential placement for the entity. Also in this view, you can test out different placement options. (For information about testing different placement options, see Inspecting placement options within constraints for an element *(on page 1195)*.)

In the same way, the ConstraintApiDTO contains data objects to describe the current placement relationships, as well as potential changes you could make. The DTO delivers an array of objects for each consumer or provider. Within each object:

- `relation` indicates whether the entity is a provider (`bought`) or consumer (`sold`)
- `relatedEntities` describes the current consumer or provider
- `placementOptions` show other placement you could exploit for this relationship, if you were to remove the current constraints (via merge policies or other means)

Example: `GET https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073/constraints`

Response: This listing shows a subset of the related entities and placement options:

```
[
  {
    "relation": "bought",
    "entityType": "PhysicalMachine",
    "relatedEntities": [
      {
        "links": [
          {
            "rel": "self",
            "href":
 "https://10.10.10.10/api/v3/entities/30333436-3638-5355-4532-313159335631?include_aspects=false"
          }
        ],
        "uuid": "30333436-3638-5355-4532-313159335631",
        "displayName": "hp-dl571.dev.mycorp.com",
        "className": "PhysicalMachine",
        "priceIndex": 6.577457,
        "state": "ACTIVE",
        "severity": "Normal",
        "discoveredBy": {
          "uuid": "_FJwg0EzHEemaXZFqO1sC0g",
          "displayName": "vsphere-dc11.dev.mycorp.com",
          "type": "vCenter"
        },
        "remoteId": "host-30",
        "environmentType": "ONPREM"
      }
```

```
    ],
    "numPotentialEntities": 1,
    "placementOptions": [
      {
        "constraintType": "ClusterCommodity",
        "scope": {
          "uuid": "17fe23bc346750d62dce8ba4274fa99b4739975f",
          "displayName": "Cloud\\Physical"
        },
        "target": {
          "uuid": "_FJwg0EzHEemaXZFqO1sC0g",
          "displayName": "vCenter:vsphere-dc11.dev.mycorp.com",
          "type": "vCenter"
        },
        "numPotentialEntities": 3,
        "key": "Cluster::17fe23bc346750d62dce8ba4274fa99b4739975f"
      },
      {
        "constraintType": "NetworkCommodity",
        "scope": {
          "displayName": "NetworkCommodity/vEOS-4"
        },
        "numPotentialEntities": 2,
        "key": "DistributedVirtualPortgroup::vmnet4"
      },
      {
        "constraintType": "DatastoreCommodity",
        "scope": {
          "uuid": "5a822e24-584bce98-ecb2-441ea15e2534",
          "displayName": "SCIO:ESXDC11DS2"
        },
        "target": {
          "uuid": "_FJwg0EzHEemaXZFqO1sC0g",
          "displayName": "vCenter:vsphere-dc11.dev.mycorp.com",
          "type": "vCenter"
        },
        "numPotentialEntities": 2,
        "key": "Storage::5a822e24-584bce98-ecb2-441ea15e2534"
      },
      {
        "constraintType": "DataCenterCommodity",
        "scope": {
          "uuid": "98e12595449d993db8861c1f94c8f86ba277f17b",
          "displayName": "Cloud"
        },
        "target": {
          "uuid": "_FJwg0EzHEemaXZFqO1sC0g",
          "displayName": "vCenter:vsphere-dc11.dev.mycorp.com",
          "type": "vCenter"
        },
        "numPotentialEntities": 7,
        "key": "DataCenter::98e12595449d993db8861c1f94c8f86ba277f17b"
      }
    ]
  },
  {
    "relation": "sold",
    "entityType": "Application",
      ...
  }
]
```


Inspecting Placement Options Within Constraints for an Element

Within the given constraints that limit placement of an entity, there are usually a number of options for placement. The current placement is just one of many.

For information about getting entity constraints, see Getting the Constraints for Placement of an Entity *(on page 1195)*.

When you get the constraints for an entity, the ConstraintApiDTO includes other placement options for that entity on its providers, and options to place other consumers on the entity. The call in this section passes these placement options and returns the respective providers or consumers that satisfy the resulting constraint changes.

Making this call is equivalent to clicking **Find More Placement Options** in the user interface. To see this in the user interface, set the scope to a single entity, display the POLICIES tab, and scroll down to the Placement Constraints section. Then click **Constraints** next to the provider you want to explore, and click **Find More Placement Options**.

Example: `POST https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073/constraints/entities`

Example input: In this example, two placement options that you copied from a ConstraintApiDTO for the entity you are inspecting was passed. Note that each placement option in the ConstraintApiDTO includes a `constraintType` and a `key`. To pass a placement option, these are the only fields you need. For this example pass a `ClusterCommodity` constraint and a `DatastoreCommodity` constraint.

```
{
  "relation": "bought",
  "placementOptions": [
    {
      "constraintType": "ClusterCommodity",
      "key": "Cluster::17fe23bc346750d62dce8ba4274fa99b4739975f"
    },
    {
      "constraintType": "DatastoreCommodity",
      "key": "Storage::5a822e24-584bce98-ecb2-441ea15e2534"
    }
  ]
}
```

Response: In the data posted in the previous example, the cluster commodity has three potential entities, and the datastore commodity has two potential entities. The response brings back the intersection of these three potential provider entities – Two hosts.

```
[
  {
    "links": [
      {
        "rel": "self",
        "href":
 "https://10.10.10.10/api/v3/entities/30343835-3638-584d-5131-323530394e39?include_aspects=false"
      }
    ],
    "uuid": "30343835-3638-584d-5131-323530394e39",
    "displayName": "hp-dl569.dev.mycorp.com",
    "className": "PhysicalMachine",
    "priceIndex": 1.031998,
    "state": "ACTIVE",
    "severity": "Normal",
    "discoveredBy": {
      "uuid": "_FJwg0EzHEemaXZFqO1sC0g",
      "displayName": "vsphere-dc11.dev.mycorp.com",
      "type": "vCenter"
    },
    "remoteId": "host-3323",
```

```
    "environmentType": "ONPREM"
  },
  {
    "links": [
      {
        "rel": "self",
        "href":
 "https://10.10.10.10/api/v3/entities/30333436-3638-5355-4532-313159335631?include_aspects=false"
      }
    ],
    "uuid": "30333436-3638-5355-4532-313159335631",
    "displayName": "hp-dl571.dev.mycorp.com",
    "className": "PhysicalMachine",
    "priceIndex": 6.916899,
    "state": "ACTIVE",
    "severity": "Normal",
    "discoveredBy": {
      "uuid": "_FJwg0EzHEemaXZFqO1sC0g",
      "displayName": "vsphere-dc11.dev.mycorp.com",
      "type": "vCenter"
    },
    "remoteId": "host-30",
    "environmentType": "ONPREM"
  }
]
```


Getting group membership of an entity

Referenced Endpoint: *Groups (on page 1222)*

The Groups endpoint tells you which entities are in a given group. The `/entities/{entity_Uuid}/groups` call tells you which groups a given entity belongs to. You can make this call in two modes:

- Include relationship paths

  In this mode the results include only those groups that are in a buy or sell relationship with the entity. For example, a VM entity might have relationships with a datacenter and a host cluster. The call would return entries for the datacenter, the host cluster, and the VM itself.

- Do not include relationship paths

  In this mode the call returns every group that includes the passed entity as a member entity. This can be a large number of groups, including all the discovered groups as well as any custom groups you created. The results do not show buy/sell relationships.

This request takes the following parameters:

- entity_Uuid

  The UUID of the entity whose actions will be returned.

- path

  Default: false. When true, results include only related groups, and they show the providers and consumers for each group when applicable.

Examples:

- All groups (`path = false`): `GET https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073/groups?path=false`
- Buy/Sell related groups, only (`path = true`): `GET https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073/groups?path=true`

Response: A list of related groups. For the default (`path = false`), this returns an array of group objects.

```
[
  {
    "links": [
      {
        "rel": "self",
        "href": "https://10.10.10.10/api/v3/groups/_lSlr0GddEemDFcIzuV5wcQ?include_aspects=false"
      }
    ],
    "uuid": "_lSlr0GddEemDFcIzuV5wcQ",
    "displayName": "CUD_Vm_Group_vEOS",
    "className": "Group",
    "environmentType": "ONPREM",
    "entitiesCount": 5,
    "membersCount": 5,
    "groupType": "VirtualMachine",
    "severity": "Normal",
    "isStatic": true,
    "logicalOperator": "OR",
    "activeEntitiesCount": 3
  },
  ...
]
```

Getting settings policies affecting an entity

Referenced Endpoint: *Settings Policies (on page 1366)*

This call returns an array of SettingsPolicyApiDTOs for all the Automation Policies that are effective for the given entity. Note that one entity can be in multiple groups, and so it can have policies that set different values for the same setting. In that case, the most conservative setting takes effect. The data returned by this call does not reflect which setting takes precedence in this case. The call merely reflects what you would see in the user interface.

Example: `GET https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073/settingspolicies`

Response: An array of `SettingsPolicyApiDTO`s that are effective for the entity.

```
[
  {
    "uuid": "_i-QuUGXUEemoItN1K8jw3Q",
    "displayName": "CUD_VM_AUTOMATION_POLICY",
    "entityType": "VirtualMachine",
    "scopes": [
      {
        "uuid": "_BZbFsWXUEemoItN1K8jw3Q",
        "displayName": "CUD_VM_Group",
        "groupType": "VirtualMachine",
        "isStatic": true,
        "logicalOperator": "AND"
      }
    ],
    "settingsManagers": [
      {
        "uuid": "automationmanager",
        "displayName": "Action Mode Settings",
        "category": "Automation",
        "settings": [
          {
            "uuid": "startVM",
            "displayName": "Start",
            "value": "MANUAL",
            "defaultValue": "RECOMMEND",
```

```
            "valueType": "STRING",
            "options": [
              {
                "label": "Disabled",
                "value": "DISABLED"
              },
              {
                "label": "Recommend",
                "value": "RECOMMEND"
              },
              {
                "label": "Manual",
                "value": "MANUAL"
              },
              {
                "label": "Automated",
                "value": "AUTOMATIC"
              }
            ],
            "entityType": "VirtualMachine",
            "sourceGroupName": "Global"
          }
        ]
      }
    ],
    "disabled": false,
    "readOnly": false,
    "default": false
  }
]
```


Getting statistics of an entity

Referenced Endpoint: Stats (on page 1372)

Gets all statistics for the specified entity. This request takes the following parameters:

- entity_uuid

  The entity whose statistics will be returned.
- q

  Used internally by Workload Optimization Manager.

Example: `GET https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073/stats`

Response: A list of `StatSnapshotApiDTO`s representing the current values of the statistics.


```
[
  {
    "displayName": "AdityaReplica",
    "date": "2019-04-24T20:09:46-04:00",
    "statistics": [
      {
        "name": "priceIndex",
        "relatedEntityType": "VirtualMachine",
        "values": {
          "max": 1.01,
          "min": 1.01,
          "avg": 1.01,
          "total": 1.01
        },
```

```
      "value": 1.01
    },
    {
      "displayName": "CPUProvisioned/AdityaReplica",
      "name": "CPUProvisioned",
      "capacity": {
        "max": 3120000,
        "min": 3120000,
        "avg": 3120000,
        "total": 3120000
      },
      "relatedEntityType": "VirtualMachine",
      "filters": [
        {
          "type": "relation",
          "value": "bought"
        }
      ],
      "units": "MHz",
      "values": {
        "max": 5200,
        "min": 5200,
        "avg": 5200,
        "total": 5200
      },
      "value": 5200
    },
    {
      "displayName": "Swapping/AdityaReplica",
      "name": "Swapping",
      "capacity": {
        "max": 40000000,
        "min": 40000000,
        "avg": 40000000,
        "total": 40000000
      },
      "relatedEntityType": "VirtualMachine",
      "filters": [
        {
          "type": "relation",
          "value": "bought"
        }
      ],
      "units": "bit/sec",
      "values": {
        "max": 0,
        "min": 0,
        "avg": 0,
        "total": 0
      },
      "value": 0
    },
    {
      "displayName": "NumDisk/AdityaReplica",
      "name": "NumDisk",
      "capacity": {
        "max": 2147483650,
        "min": 2147483650,
        "avg": 2147483650,
        "total": 2147483650
      },
      "relatedEntityType": "VirtualMachine",
      "filters": [
        {
          "type": "relation",
```

```
            "value": "bought"
          }
        ],
        "values": {
          "max": 1,
          "min": 1,
          "avg": 1,
          "total": 1
        },
        "value": 1
      },
    ...
    ]
  }
]
```

Getting a filtered list of statistics for an entity

Referenced Endpoint: Stats (on page 1372)

Gets a list of statistics filtered by a user-created StatPeriodApiInputDTO. For more information about the criteria that make up the input for thi request, see the Swagger documentation for this request, located at `https://<Your_Workload Optimization Manager_URL/apidpoc`.

Example: `POST https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073/stats`

Example Input: This input will return the values for the VMem, VCPU, and VStorage statistics of the specified entity, ordered first by key, then relatedEntity, then virtualDisk.

```
{
  "statistics": [
    {
      "name": "VMem",
      "groupBy": [
        "key",
        "relatedEntity",
        "virtualDisk"
      ]
    },
    {
      "name": "VCPU",
      "groupBy": [
        "key",
        "relatedEntity",
        "virtualDisk"
      ]
    },
    {
      "name": "VStorage",
      "groupBy": [
        "key",
        "relatedEntity",
        "virtualDisk"
      ]
    }
  ]
}
```

Response: An array of `StatSnapshotApiDTO`s representing the filtered statistics for the entity:


```
[
  {
    "displayName": "vEOS-4",
    "date": "2019-04-25T17:57:02-04:00",
    "statistics": [
      {
        "displayName": "vEOS-4",
        "name": "VCPU",
        "capacity": {
          "max": 2663,
          "min": 2663,
          "avg": 2663,
          "total": 2663
        },
        "relatedEntityType": "VirtualMachine",
        "filters": [
          {
            "type": "virtualDisk",
            "value": ""
          },
          {
            "type": "key",
            "value": null
          },
          {
            "type": "relation",
            "value": "sold"
          }
        ],
        "relatedEntity": {
          "uuid": "c0849b50a2cce61f57ea0e49e2199503e82f3f54",
          "displayName": "GuestLoad[vEOS-4]",
          "className": "Application"
        },
        "units": "MHz",
        "values": {
          "max": 143,
          "min": 91.47,
          "avg": 91.47,
          "total": 91.47
        },
        "value": 91.47
      },
      {
        "displayName": "vEOS-4",
        "name": "VMem",
        "capacity": {
          "max": 2097152,
          "min": 2097152,
          "avg": 2097152,
          "total": 2097152
        },
        "relatedEntityType": "VirtualMachine",
        "filters": [
          {
            "type": "virtualDisk",
            "value": ""
          },
          {
            "type": "key",
            "value": null
          },
```

```
          {
            "type": "relation",
            "value": "sold"
          }
        ],
        "relatedEntity": {
          "uuid": "c0849b50a2cce61f57ea0e49e2199503e82f3f54",
          "displayName": "GuestLoad[vEOS-4]",
          "className": "Application"
        },
        "units": "KB",
        "values": {
          "max": 188740,
          "min": 130022,
          "avg": 130022,
          "total": 130022
        },
        "value": 130022
      }
    ]
  }
]
```


Getting the supply chain of an entity

Referenced Endpoint: Supply Chains (on page 1389)

Workload Optimization Manager models your environment as a market of buyers and sellers. It maps the entities it discovers to a supply chain that represents the buy/sell relationships between these entities.

In the user interface, the default view on the Home Page shows the global supply chain. You can change scope to drill down to a supply chain for the relationships to a single entity. This is the same as the result when you get a supply chain for an entity via the API. The returned data is a SupplyChainApiDTO that expresses the scope for the passed entity.

The supply chain shows tiers of entity types, where depth indicates where in the supply chain levels a specific tier is. When you get a supply chain for an entity, the entity is at `depth = 0`. The depths for tiers above the entity are negative. For example, if a VM is at `depth = 0`, then the application that consumes from the VM would be at `depth = -1`. Likewise, providers to the VM would have positive depths. For example, storage that provides resources to the VM would have `depth = 1`, and a disk array that hosts the storage would have `depth = 2`.

The supply chain request takes the following parameters:

- entity_uuid

  The entity that sets the scope of the supply chain.

- types

  An array of strings for the entity types you want to include in the supply chain. If you leave this blank, the call returns the complete supply chain for this entity. If you specify any types, the returned supply chain contains only those types. Note that this rule also applies to the entity that sets the scope (`entity_uuid`) – If you do not include that entity's type, then the supply chain will not include that entity, even though it sets the scope of the request.

- entity_states

  An array of strings. Limit the supply chain to include only entities in the given state. Entity states are:

  - ACTIVE: The entity is actively consuming resources
  - EVACUATED: All consumers have been moved off of the entity, and it can be suspended or terminated
  - FAILOVER: The entity is reserved for failover and will not participate in the market
  - IDLE: The entity is powered off and not actively consuming market resources
  - LAUNCH: The entity is starting up
  - MAINTENANCE: The entity is in Maintenance mode

  - NOT_MONITORED: Workload Optimization Manager does not currently monitor the entity
  - QUEUED: More than ten actions are to be executed for a single target, and this entity has pending actions in the queue
  - RESOURCE_ALLOCATION: For internal use
  - RESOURCE_RELEASE: For internal use
  - SUSPEND: The entity has been placed in Suspend mode
  - SUSPEND_PENDING: The entity is making preparations to suspend, or is the next entity of its type to suspend
  - TERMINATE_PENDING: The entity is making preparations to terminate, or is the next entity of its type to terminate
  - UNKNOWN: Workload Optimization Manager is unable to obtain the current state of the entity
- environment_type

  Limit the supply chain to a specific environment type. If you do not specify environment type, the supply chain includes all environments, which is the same as HYBRID. The environment types are:
  - HYBRID
  - CLOUD
  - ONPREM
- detail_type

  The level of detail you want in your supply chain. Can be one of:
  - compact: Minimal detail about the supply chain items
  - entity: The supply chain items include full entity descriptions
  - aspects: The supply chain items also include aspects (mata data stored with the entity)
- health

  Whether to include a health summary with the entities in the supply chain. `false` by default.

Example: `GET https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073/supplychains?types=VirtualMachine&types=Storage&types=DiskArray&detail_type=compact&health=true`

This example gets a supply chain for a VM entity, with the following settings: including the VM, storage, and disk arrays:

- Only include the VM, its storage, and the disk arrays
- Give compact details
- Show the entity health

Response: The `SupplyChainApiDTO` for the filtered request:


```
{
  "seMap": {
    "VirtualMachine": {
      "depth": 0,
      "entitiesCount": 1,
      "healthSummary": {
        "Normal": 1
      },
      "stateSummary": {
        "ACTIVE": 1
      },
      "connectedProviderTypes": [
        "PhysicalMachine",
        "Storage",
        "VirtualDataCenter"
      ],
      "connectedConsumerTypes": [
        "Application"
```

```
    ],
    "instances": {
      "420b1eeb-6190-a9d2-6436-4327ad3ac073": {
        "links": [
          {
            "rel": "self",
            "href":
"https://10.10.10.10/api/v3/entities/420b1eeb-6190-a9d2-6436-4327ad3ac073?include_aspects=false"
          }
        ],
        "uuid": "420b1eeb-6190-a9d2-6436-4327ad3ac073",
        "displayName": "vEOS-4",
        "className": "VirtualMachine",
        "priceIndex": 1.1463195,
        "state": "ACTIVE",
        "severity": "Normal",
        "discoveredBy": {
          "uuid": "_FJwg0EzHEemaXZFqO1sC0g",
          "displayName": "vsphere-dc11.dev.mycorp.com",
          "type": "vCenter"
        },
        "remoteId": "vm-100",
        "environmentType": "ONPREM"
      }
    }
  },
  "Storage": {
    "depth": 1,
    "entitiesCount": 2,
    "healthSummary": {
      "Normal": 2
    },
    "stateSummary": {
      "ACTIVE": 2
    },
    "connectedProviderTypes": [
      "DiskArray"
    ],
    "connectedConsumerTypes": [
      "PhysicalMachine",
      "VirtualMachine"
    ],
    "instances": {
      "5a822e24-584bce98-ecb2-441ea15e2534": {
        "links": [
          {
            "rel": "self",
            "href":
"https://10.10.10.10/api/v3/entities/5a822e24-584bce98-ecb2-441ea15e2534?include_aspects=false"
          }
        ],
        "uuid": "5a822e24-584bce98-ecb2-441ea15e2534",
        "displayName": "SCIO:ESXDC11DS2",
        "className": "Storage",
        "priceIndex": 1.2797012,
        "state": "ACTIVE",
        "severity": "Normal",
        "discoveredBy": {
          "uuid": "_FJwg0EzHEemaXZFqO1sC0g",
          "displayName": "vsphere-dc11.dev.mycorp.com",
          "type": "vCenter"
        },
        "remoteId": "datastore-2357",
        "environmentType": "ONPREM",
        "tags": {
```

```
            "Datastore_DC11": [
              "True"
            ]
          }
        },
        "9bd4ee88-99c64661": {
          "links": [
            {
              "rel": "self",
              "href": "https://10.10.10.10/api/v3/entities/9bd4ee88-99c64661?include_aspects=false"
            }
          ],
          "uuid": "9bd4ee88-99c64661",
          "displayName": "QS1:NFSShare",
          "className": "Storage",
          "priceIndex": 20000,
          "state": "ACTIVE",
          "severity": "Normal",
          "discoveredBy": {
            "uuid": "_gOQEYEaREemB8r4uTzazvA",
            "displayName": "vsphere-dc5.dev.mycorp.com",
            "type": "vCenter"
          },
          "remoteId": "datastore-1411",
          "environmentType": "ONPREM",
          "tags": {
            "DimStor": [
              "DimStorTag"
            ],
            "Datastore_DC11": [
              "True"
            ],
            "Test_Tag": [
              "MyStorageTag"
            ]
          }
        }
      }
    },
    "DiskArray": {
      "depth": 2,
      "entitiesCount": 2,
      "healthSummary": {
        "Normal": 2
      },
      "stateSummary": {
        "ACTIVE": 2
      },
      "connectedConsumerTypes": [
        "Storage"
      ],
      "instances": {
        "DiskArray-9bd4ee88-99c64661": {
          "links": [
            {
              "rel": "self",
              "href":
"https://10.10.10.10/api/v3/entities/DiskArray-9bd4ee88-99c64661?include_aspects=false"
            }
          ],
          "uuid": "DiskArray-9bd4ee88-99c64661",
          "displayName": "DiskArray-QS1:NFSShare",
          "className": "DiskArray",
          "priceIndex": 0,
          "state": "ACTIVE",
```

```
          "severity": "Normal",
          "discoveredBy": {
            "uuid": "_oQ2aIEG1EemgjaEqcE-BJw",
            "displayName": "vsphere-dc13.dev.mycorp.com",
            "type": "vCenter"
          },
          "remoteId": "datastore-982",
          "environmentType": "ONPREM"
        },
        "DiskArray-5a822e24-584bce98-ecb2-441ea15e2534": {
          "links": [
            {
              "rel": "self",
              "href":
 "https://10.10.10.10/api/v3/entities/DiskArray-5a822e24-584bce98-ecb2-441ea15e2534?include_aspects=false"
            }
          ],
          "uuid": "DiskArray-5a822e24-584bce98-ecb2-441ea15e2534",
          "displayName": "DiskArray-SCIO:ESXDC11DS2",
          "className": "DiskArray",
          "priceIndex": 0,
          "state": "ACTIVE",
          "severity": "Normal",
          "discoveredBy": {
            "uuid": "_FJwg0EzHEemaXZFqO1sC0g",
            "displayName": "vsphere-dc11.dev.mycorp.com",
            "type": "vCenter"
          },
          "remoteId": "datastore-2357",
          "environmentType": "ONPREM"
        }
      }
    }
  }
}
```


Getting the tags on an entity

Referenced Endpoint: Tags (on page 1392)

Workload Optimization Manager discovers tags and tag values that are assigned to different entities in your environment. This call returns the tag and value sets for the entity you pass.

Example: `GET https://10.10.10.10/api/v3/entities/9bd4ee88-99c64661/tags`

Response: A `TagApiDTO` for the passed entity. This gives an array of key/value pairs, where the value is an array of strings.


```
[
  {
    "key": "Datastore_DC11",
    "values": [
      "True"
    ]
  },
  {
    "key": "Test_Tag",
    "values": [
      "MyStorageTag"
    ]
  },
  {
    "key": "DimStor",
```

```
    "values": [
      "DimStorTag"
    ]
  }
]
```

Groups endpoint

A group is a collection of entities (such as physical machines, containers, clusters, or even other groups) that Workload Optimization Manager can work with as a unit. The most common use of groups is to set scope for display in charts or for processing in plans, and to retrieve statistics for a subset of your environment.

This endpoint contains references to start and end times for certain operations. For more information, including valid time formats, see Time in the Workload Optimization Manager API *(on page 1141)*

Using the groups endpoint of the Workload Optimization Manager API, you can:

- Get a list of all groups on your Workload Optimization Manager appliance
- Create a group
- Edit a group
- Delete a group
- Get a list of actions for a specified groups
- Get a list of entities in a specified group
- Get parent groups of a specified group
- Get a list of members of a specified group
- Get a list of notifications for a specified group
- Get a list of settings for a specified group
- Get a list of policies for a specified group
- Get statistics for a specified group
- Get the supply chain related to a specified group
- Get the tags belonging to a specified group

Group types and class names

Workload Optimization Manager subdivides the entities on your appliance by the following group class name, to aid in both searching and understanding the purpose of a group:

- Folder

  These appear in the user interface as folders, and are for visual organization. Discovered folders represent the folder structure in the target – for example, the vCenter folder structure. In addition, Workload Optimization Manager discovers vCenter Server resource pools, and groups them into folders.

- Group

  Workload Optimization Manager places discovered entities into standard groups, and users can create groups of their own with static or dynamic membership.

- RefGroup

  A RefGroup can be a group of groups- for example, a group of PM cluster groups that make up a hypervisor target- or a group defined by a relationship. For example, the built-in group `PMsByDatacenter` is a RefGroup reflecting the relationship between the datacenters and hosts.

- Cluster / StorageCluster

  Groups that correspond to discovered clusters.

- DiscoveredGroup

  Groups that are defined by a target service. For example, Workload Optimization Manager can discover DRS domains that were defined in vCenter Server.

- MarketGroup

  A group that is based on the infrastructure cost of that entity type. These groups are based on the Infrastructure Cost settings in the Workload Optimization Manager policies.

- StaticMetaGroup / RefMetaGroup / MetaGroup

  Groups used internally by Workload Optimization Manager. These should not be modified in any way.

In addition, each group has a `groupType` that identifies the type of members that group contains. Remember that when you create a group in the user interface or via the API, you cannot add more than one type of member. The different group types include:

- Cluster

  A group of clusters. For example, you could create a group of all PM clusters that include the name "Development".

- ServiceEntity

  An internal group type reserved for groups that Workload Optimization Manager discovers or creates. Note that the group will still contain members of only one type.

- Entity Type
  - `APPLICATION`
  - `APPLICATION_SERVER`
  - `BUSINESS_ACCOUNT`
  - `BUSINESS_UNIT`
  - `CHASSIS`
  - `CLOUD_SERVICE`
  - `CONTAINER`
  - `CONTAINER_POD`
  - `DATABASE`
  - `DATABASE_SERVER`
  - `DATACENTER`
  - `DISK_ARRAY`
  - `IO_MODULE`
  - `LOAD_BALANCER`
  - `LOGICAL_POOL`
  - `NETWORK`
  - `PHYSICAL_MACHINE`
  - `STORAGE`
  - `STORAGE_CONTROLLER`
  - `SWITCH`
  - `VIRTUAL_APPLICATION`
  - `VIRTUAL_DATACENTER`
  - `VIRTUAL_MACHINE`

Understanding the different group types helps you find the group you want. If you know the name of a specific group, then you can filter the returned list for a group of that name. Or if you want to step through all the clusters Workload Optimization Manager has discovered, you can filter all the entries with a `className` of `Cluster` or `StorageCluster`.

Group filter types

Workload Optimization Manager utilizes regex and a `filterType` to build a group. The `filterType` is the criterion to which the regex will match. For example, a `filterType` of `pmsByName` will apply the regex to the display names of physical machines.

The available options for this parameter depend on the entity making up your group:

- ApplicationComponentSpec
  - appComponentSpecsByName
  - appComponentSpecsByBusinessAccountUuid
  - appComponentSpecsByHybridConnectionCount
  - appComponentSpecsByDeploymentSlotCount
- ApplicationComponent
  - appCompsByName
  - applicationComponentByBusinessApplication
  - applicationComponentByBusinessTransaction
  - applicationComponentByService
- AvailabilityZone
  - zonsByName
  - zoneByCloudProvider
- BillingFamily
  - billingFamilyByName
- BusinessAccountFolder
  - businessAccountFolderByName
- BusinessAccount
  - businessAccountByName
  - businessAccountByUuid
  - businessAccountByVendorId
  - businessAccountCloudProvider
  - businessAccountProbeType
  - businessAccountTargetName
  - businessAccountValidationStatus
  - businessAccountByResourceGroup
  - businessAccountByTag
  - subBusinessAccountOfUuid
- BusinessApplication
  - busAppsByName
  - businessApplicationByUserDefinedEntity
- BusinessTransaction
  - busTransByName
  - businessTransactionByBusinessApplication
  - businessTransactionByUserDefinedEntity

- BusinessUser
  - businessUserByName
  - businessUserByDesktopPool
  - businessUserByViewPod
- Chassis
  - chasByName
- CloudService
- Cluster
  - clustersByName
  - clustersByTag
- ComputeTier
  - computeTiersByName
  - computeTiersByMem
  - computeTiersByNumCPUs
  - computeTiersByCloudProvider
- ContainerPlatformCluster
  - containerPlatformClustersByName
- Container
  - containersByName
  - containersByPodName
  - containersByNamespace
  - containersByApplicationComponent
  - containersByBusinessApplication
  - containersByBusinessTransaction
  - containersByService
  - containersByControllerKind
  - containersByTag
  - containersByContainerPlatformCluster
- ContainerPod
  - containerPodsByName
  - containerPodsByVMName
  - containerPodsByNamespace
  - containerPodsByTag
  - containerPodsByControllerKind
  - containerPodsByWorkloadControllerName
  - containerPodsByContainerPlatformCluster
  - containerPodsByState
- ContainerSpec
  - containerSpecsByName
  - containerSpecsByNamespace
  - containerSpecsByControllerKind
  - containerSpecsByContainerPlatformCluster
  - containerSpecsByTag

- DataCenter
  - datacentersByName
  - datacentersByTag
- Database
  - databaseByName
  - databaseByTag
  - databaseByBusinessAccountUuid
  - databaseByResourceGroupName
  - databaseByCloudProvider
  - databaseByApplicationComponent
  - databaseByBusinessApplication
  - databaseByBusinessTransaction
  - databaseByService
  - dbByAvailabilityZoneName
  - dbByRegionName
  - dbByDatabaseServerName
  - databaseByReplicationRole
  - databaseByPricingModel
  - databaseByServiceTier
  - databaseByHoursIdle
  - databaseByAge
- DatabaseServer
  - databaseServerByName
  - databaseServerByTag
  - databaseServerByBusinessAccountUuid
  - databaseServerByResourceGroupName
  - databaseServerByCloudProvider
  - databaseServerByApplicationComponent
  - databaseServerByBusinessApplication
  - databaseServerByBusinessTransaction
  - databaseServerByService
  - databaseServerByEngine
  - databaseServerByEdition
  - databaseServerByVersion
  - databaseServerByState
  - databaseServerByServerConfiguration
  - databaseServerByStorageAutoscaling
  - databaseServerByPerformanceInsights
  - databaseServerByStorageEncryption
  - databaseServerByClusterRole
  - databaseServerByStorageTier
  - databaseServerByAvailabilityZoneName

  - databaseServerByRegionName
  - databaseServerByAge
- DatabaseServerTier
  - databaseServerTiersByName
  - databaseServerTiersByVMem
  - databaseServerTiersByCloudProvider
- DatabaseTier
  - databaseTiersByName
  - databaseTiersByCloudProvider
- DesktopPool
  - desktopPoolByName
  - desktopPoolByViewPod
- DiskArray
  - diskarrayByName
- DocumentCollection
  - dcByName
  - dcByTag
  - dcByBusinessAccount
  - dcByResourceGroup
  - dcByCloudProvider
  - dcByRegion
  - dcByDatabaseServer
- Group
  - groupsByName
- HCIPhysicalMachine
- IOModule
  - ioModuleByName
- Internet
  - internetByName
- LoadBalancer
  - lbsByName
- LogicalPool
  - logicalPoolByName
- Namespace
  - namespacesByName
  - namespacesByContainerPlatformCluster
  - namespacesByTag
- Network
  - netsByName
- NodePool

- PhysicalMachine
  - pmsByName
  - pmsByTargetType
  - pmsByClusterName
  - pmsByDC
  - pmsByState
  - pmsByTag
  - pmsByStorage
  - pmsBySwitch
  - pmsByMem
  - pmsByNumVms
  - pmsByNumCPUs
  - pmsByVendorName
  - pmsByCPUModel
  - pmsByModel
  - pmsByTimezone
  - pmsBySupportedProcessorCompatibilityModes
  - pmsByClusterTag
- ProcessorPool
- Region
  - regsByName
  - regsByCloudProvider
- ReservedInstance
- ResourceGroup
  - resourceGroupByName
  - resourceGroupByTag
  - resourceGroupByBusinessAccountUuid
- Service
  - servicesByName
  - serviceByNamespace
  - serviceByContainerPlatformCluster
  - serviceByKubernetesServiceType
  - serviceByBusinessApplication
  - serviceByTag
  - serviceByBusinessTransaction
  - serviceByUserDefinedEntity
- Storage
  - storageByName
  - storageByVMs
  - storageByPMCluster
  - storageByStorageCluster
  - storageByState
  - storageByTag

  - storageByDC
  - storageWithLocalSupport
- StorageCluster
  - storageClustersByName
- StorageController
  - storagecontrollerByName
- StorageTier
  - storageTierByName
  - storageTiersByCloudProvider
- Switch
  - switchByName
- ViewPod
  - viewPodByName
- VirtualDataCenter
  - vdcsByName
  - vdcsByTag
- VirtualMachine
  - vmsByName
  - vmsByPMName
  - vmsByVDC
  - vmsByDCnested
  - vmsByDiskArrayName
  - vmsByStorage
  - vmsByApplicationComponent
  - vmsByBusinessApplication
  - vmsByBusinessTransaction
  - vmsByService
  - vmsByClusterName
  - vmsByStorageClusterName
  - vmsByDC
  - vmsByState
  - vmsByTag
  - vmsConnectedToEphemeralStorage
  - vmsHotAddMemory
  - vmsHotAddCPU
  - vmsHotRemoveCPU
  - vmsConnectedToEncryptedVolume
  - vmsByGuestName
  - vmsBySockets
  - vmsByCoresPerSocket
  - vmsByNumCPUs
  - vmsByAge

  - vendorToolsInstalled
  - vendorToolsVersion
  - vmsByProcessorCompatibilityMode
  - vmsBySharingMode
  - vmsByHasDedicatedProcessors
  - vmByTargetCategory
  - vmByTargetName
  - vmByTargetType
  - vmsByBusinessAccountUuid
  - vmsByMem
  - vmsByStorageAmount
  - vmsByLogicalPoolName
  - vmsByNetwork
  - vmsByResourceGroupName
  - vmsByAvailabilityZoneName
  - vmsByRegionName
  - vmsByCloudProvider
  - vmsByContainerPlatformCluster
  - vmsByActiveSessions
  - vmsByVendorId
  - vmByComputeTier
- VirtualMachineCluster
  - virtualMachineClustersByName
- VirtualMachineSpec
  - virtualMachineSpecsByName
  - virtualMachineSpecsByTier
  - virtualMachineSpecsByAppCount
  - virtualMachineSpecsByBusinessAccountUuid
  - virtualMachineSpecsByRegionName
  - virtualMachineSpecsByResourceGroup
  - virtualMachineSpecsByComputeTier
  - virtualMachineSpecsByTag
  - virtualMachineSpecsByNumDaysEmpty
- VirtualVolume
  - virtualVolumeByName
  - virtualVolumeWithLocalSupport
  - volumeByTag
  - volumeByAccount
  - volumeByConnectedVm
  - volumeByAttachmentState
  - volumeByNumDaysUnattached
  - volumeByTier

  - volumeByRegion
  - volumeByCloudProvider
  - volumeById
  - volumeByResourceGroupName
  - volumeByDeletable
  - volumeByStorage
  - volumeByStorageGroup
  - volumeByAge
- Workload
  - workloadByName
- WorkloadController
  - workloadControllersByName
  - workloadControllersByNamespace
  - workloadControllersByKind
  - workloadControllersByContainerPlatformCluster
  - workloadControllersByTag

Note that in the case of a discovered resource group, group members can be discovered by multiple targets. In that case, the Workload Optimization Manager API returns a single target for the group.

Groups requests

Pagination for group requests

When you request group data, the API uses pagination to return subsets of the data. You should always use pagination when available to make sure you get your data in manageable chunks. To paginate returned data, the `limit`, `orderBy`, and `cursor` parameters control how the APIs paginate their data. If you do not specify a `limit` for the following requests, the API will paginate the data to a default limit of records:

- `GET /groups`
- `GET /groups/<groupId>/members`
- `GET /groups/<groupId>/entities`

If your scripts make these requests without setting a limit, you should be sure they check for a pagination cursor in the returned data.

For general information about pagination, see Pagination in the Workload Optimization Manager API *(on page 1141)*.

Getting a list of groups

Gets a list of all groups on your Workload Optimization Manager appliance. This will include Workload Optimization Manager internal and default groups, as well as any groups discovered from your targets, in addition to your custom groups.

Example: `GET https://10.10.10.10/api/v3/groups`

Response: A list of `GroupApiDTO`s representing the groups on your appliance.

```
[
  {
    "links": [
      {
```

```
        "rel": "self",
        "href": "https://10.10.10.10/api/v3/groups/_S4k98SFQEd-S4dn3K4_SSA?include_aspects=false"
      }
    ],
    "uuid": "_S4k98SFQEd-S4dn3K4_SSA",
    "displayName": "Datacenters",
    "className": "Group",
    "environmentType": "ONPREM",
    "entitiesCount": 12,
    "membersCount": 12,
    "groupType": "DataCenter",
    "severity": "Critical",
    "isStatic": false,
    "logicalOperator": "AND",
    "activeEntitiesCount": 44
  },
  {
    "links": [
      {
        "rel": "self",
        "href": "https://10.10.10.10/api/v3/groups/_S4k94SFQEd-S4dn3K4_SSA?include_aspects=false"
      }
    ],
    "uuid": "_S4k94SFQEd-S4dn3K4_SSA",
    "displayName": "Chassis",
    "className": "Group",
    "environmentType": "ONPREM",
    "entitiesCount": 2,
    "membersCount": 1,
    "groupType": "Chassis",
    "severity": "Normal",
    "isStatic": false,
    "logicalOperator": "AND",
    "activeEntitiesCount": 2
  },
  ...
]
```


Creating a group

In order to create a group in Workload Optimization Manager, provide a GroupApiDTO with a `groupType` and `isStatic`.

For a dynamic group (`isStatis = false`), you also include the `criteriaList`information:

- caseSensitive

  If `true`, the regex will be evaluated as case-sensitive.
- expType
- For strings, the allowable values are EQ and NEQ (Equals, Does Not Equal). For integers, you may additionally use `GT`, `LT`, `GTE`, `LTE` (Greater than, less than, greater than or equal to, and less than or equal to, respectively).
- expVal

  The value of the regular expression.
- filterType

  See **Group Filter Types** for a full list of available options

For a static group:

- Provide UUIDs for each entity that is a member of the group in the `memberUuidList`
- DO NOT provide any `criteriaList` parameters

Example: `POST https://10.10.10.10/api/v3/groups/`

Example input, showing the creation of a dynamic group of VMs whose names start with "AJ":

```
{
  "isStatic": false,
  "displayName": "Dallas-Dynamic",
  "memberUuidList": [],
  "criteriaList": [
    {
      "expType": "EQ",
      "expVal": "AJ.*",
      "filterType": "vmsByName",
      "caseSensitive": false
    }
  ],
  "groupType": "VirtualMachine"
}
```

Response:

```
{
  "uuid": "_4L0PcE9SEemD69d9a0uLJQ",
  "displayName": "Dallas-Dynamic",
  "className": "Group",
  "environmentType": "ONPREM",
  "entitiesCount": 1,
  "membersCount": 1,
  "groupType": "VirtualMachine",
  "severity": "Major",
  "isStatic": false,
  "logicalOperator": "AND",
  "criteriaList": [
    {
      "expVal": "AJ.*",
      "expType": "EQ",
      "filterType": "vmsByName",
      "caseSensitive": false
    }
  ],
  "activeEntitiesCount": 1
}
```

Example input, showing the creation of a static group of Hosts:

```
{
  "isStatic": true,
  "displayName": "Dallas-Static",
  "memberUuidList": [
    "4C4C4544-0042-5A10-804A-B1C04F384331",
    "4C4C4544-004A-3610-8057-C8C04F464331",
    "Virtual_ESX_42381e13-4256-64bb-f554-d17798ddb2ba"
  ],
  "criteriaList": [],
  "groupType": "PhysicalMachine"
}
```

Response:

```
{
  "links": [
    {
      "rel": "self",
```

```
      "href": "https://10.10.10.10/api/v3/groups/_vn_u0FJ0EemD69d9a0uLJQ?include_aspects=false"
    }
  ],
  "uuid": "_vn_u0FJ0EemD69d9a0uLJQ",
  "displayName": "Dallas-Static",
  "className": "Group",
  "environmentType": "ONPREM",
  "entitiesCount": 3,
  "membersCount": 3,
  "groupType": "PhysicalMachine",
  "severity": "Critical",
  "isStatic": true,
  "logicalOperator": "OR",
  "activeEntitiesCount": 3
}
```

Editing a group

In order to edit a user-created group, send the complete `GroupApiDTO` with your changes as the input to this request.

Example: `PUT https://10.10.10.10/api/v3/groups/_Y6BL4FDQEemD69d9a0uLJQ`

Example Input:A modified `GroupApiDTO`. In this case, the `expVal` parameter has been modified.

```
{
  "isStatic": false,
  "displayName": "Dallas-Host",
  "memberUuidList": [],
  "criteriaList": [
    {
      "expType": "EQ",
      "expVal": "hp*change",
      "filterType": "pmsByName",
      "caseSensitive": false
    }
  ],
  "groupType": "PhysicalMachine",
  "logicalOperator": "AND"
}
```

Response: The full `GroupApiDTO` reflecting your changes:

```
{
  "links": [
    {
      "rel": "self",
      "href": "https://10.10.10.10/api/v3/groups/_Y6BL4FDQEemD69d9a0uLJQ?include_aspects=false"
    }
  ],
  "uuid": "_Y6BL4FDQEemD69d9a0uLJQ",
  "displayName": "Dallas-Host",
  "className": "Group",
  "environmentType": "ONPREM",
  "entitiesCount": 0,
  "membersCount": 0,
  "groupType": "PhysicalMachine",
  "severity": "Normal",
  "isStatic": false,
  "logicalOperator": "AND",
  "criteriaList": [
    {
      "expVal": "hp*change",
```

```
      "expType": "EQ",
      "filterType": "pmsByName",
      "caseSensitive": false
    }
  ],
  "activeEntitiesCount": 0
}
```

Deleting a group

To delete a user-created group, pass the UUID of the group in the URL of the request.

Example: `DELETE https://10.10.10.10/api/v3/groups/_vn_u0FJ0EemD69d9a0uLJQ`

Response: A response of `200` indicates successful deletion.

Getting actions related to a group

Referenced Endpoint: Actions (on page 1161)

Gets a list of actions relating to the entities in the specified group. To get a single action, include the UUID of the action in the request. This request takes the following parameters:

- group_Uuid

  The UUID of the group whose actions will be returned.
- cursor

  When making subsequent calls to retrieve additional results, this is the cursor returned by the last call to this method.
- limit

  The maximum number of items to return. Must be a positive integer. If not set, the server will provide a default (20).
- order_by

  The field to order the results by. `[ creation_date, name, risk_category, savings, severity ]`.
- ascending

  Default: true. When false, results will be in descending order.

Examples:

All actions: `GET https://10.10.10.10/api/v3/groups/421dd4ce-a4c1-231a-71e7-64fbf598bc65/actions?limit=50&order_by=creation_date&ascending=true`

Single action: `GET https://10.10.10.10/api/v3/groups/421dd4ce-a4c1-231a-71e7-64fbf598bc65/actions/_pJg98EQaEemD69d9a0uLJQ`

Response: A list of `ActionApiDTO`s related to the specified group:

```
[
  {
    "links": [
      {
        "rel": "self",
        "href": "https://10.10.10.10/api/v3/actions/_pJg98EQaEemD69d9a0uLJQ"
      },
      {
        "rel": "self",
        "href": "https://10.10.10.10/api/v3/actions/2500817580000"
      }
    ],
    "uuid": "_pJg98EQaEemD69d9a0uLJQ",
    "createTime": "2019-03-11T12:28:03-04:00",
```

```
    "actionType": "RIGHT_SIZE",
    "actionState": "PENDING_ACCEPT",
    "actionMode": "MANUAL",
    "details": "Scale down Memory reservation for Virtual Machine 'AJ_Test' from 2 GB to 1 GB",
    "importance": 0,
    "target": {
      "uuid": "421dd4ce-a4c1-231a-71e7-64fbf598bc65",
      "displayName": "AJ_Test",
      "className": "VirtualMachine",
      "aspects": {
        "virtualMachineAspect": {
          "os": "CentOS 4/5/6/7 (64-bit)",
          "ip": [
            "10.10.10.107"
          ],
          "connectedNetworks": [
            {
              "uuid": "34f9f5364cd843aa1fe99900621ee3b51a7dc6e5",
              "displayName": "VM Network"
            }
          ],
          "numVCPUs": 1,
          "ebsOptimized": false
        }
      },
      "environmentType": "ONPREM"
    },
    "currentEntity": {
      "uuid": "_pJlPYUQaEemD69d9a0uLJQ",
      "className": "Mem"
    },
    "newEntity": {
      "uuid": "_pLCA40QaEemD69d9a0uLJQ",
      "className": "Mem",
      "aspects": {
        "virtualMachineAspect": {
          "os": "Linux (Free)",
          "ebsOptimized": false
        }
      }
    },
    "currentValue": "2097152.0",
    "newValue": "1048576.0",
    "resizeToValue": "1048576.0",
    "risk": {
      "uuid": "_pL_qMEQaEemD69d9a0uLJQ",
      "subCategory": "Efficiency Improvement",
      "description": "Underutilized Memory reservation",
      "severity": "MINOR",
      "importance": 0
    },
    "currentLocation": {
      "uuid": "cdcd967d1e08d20936c3aa9a0879c635214c79ee",
      "displayName": "ACM",
      "className": "DataCenter"
    },
    "newLocation": {
      "uuid": "cdcd967d1e08d20936c3aa9a0879c635214c79ee",
      "displayName": "ACM",
      "className": "DataCenter"
    },
    "actionID": 2500817580000
  },
  {
    "links": [
```

```
    {
      "rel": "self",
      "href": "https://10.10.10.10/api/v3/actions/_XdP4EEQ8EemD69d9a0uLJQ"
    },
    {
      "rel": "self",
      "href": "https://10.10.10.10/api/v3/actions/2501049335840"
    }
  ],
  "uuid": "_XdP4EEQ8EemD69d9a0uLJQ",
  "createTime": "2019-03-11T16:29:28-04:00",
  "actionType": "RIGHT_SIZE",
  "actionState": "PENDING_ACCEPT",
  "actionMode": "MANUAL",
  "details": "Remove VMem limit for VirtualMachine AJ_Test to maximize efficiency",
  "importance": 0,
  "target": {
    "uuid": "421dd4ce-a4c1-231a-71e7-64fbf598bc65",
    "displayName": "AJ_Test",
    "className": "VirtualMachine",
    "aspects": {
      "virtualMachineAspect": {
        "os": "CentOS 4/5/6/7 (64-bit)",
        "ip": [
          "10.10.10.107"
        ],
        "connectedNetworks": [
          {
            "uuid": "34f9f5364cd843aa1fe99900621ee3b51a7dc6e5",
            "displayName": "VM Network"
          }
        ],
        "numVCPUs": 1,
        "ebsOptimized": false
      }
    },
    "environmentType": "ONPREM"
  },
  "currentEntity": {
    "uuid": "_XdP4EkQ8EemD69d9a0uLJQ",
    "className": "VMem"
  },
  "newEntity": {
    "uuid": "_Xnq_YUQ8EemD69d9a0uLJQ",
    "className": "VMem",
    "aspects": {
      "virtualMachineAspect": {
        "os": "Linux (Free)",
        "ebsOptimized": false
      }
    }
  },
  "currentValue": "2097152.0",
  "newValue": "0.0",
  "resizeToValue": "0.0",
  "risk": {
    "uuid": "_XnuCsUQ8EemD69d9a0uLJQ",
    "subCategory": "Efficiency Improvement",
    "description": "VMem congestion in Virtual Machine 'AJ_Test'",
    "severity": "MAJOR",
    "reasonCommodity": "VMem",
    "importance": 0
  },
  "currentLocation": {
    "uuid": "cdcd967d1e08d20936c3aa9a0879c635214c79ee",
```

```
      "displayName": "ACM",
      "className": "DataCenter"
    },
    "newLocation": {
      "uuid": "cdcd967d1e08d20936c3aa9a0879c635214c79ee",
      "displayName": "ACM",
      "className": "DataCenter"
    },
    "actionID": 2501049335840
  }
]
```

Getting the entities in a group

Referenced Endpoint: Entities (on page 1195)

Gets all entities contained within a group.

> **NOTE:** This request will not return the contents of nested groups. In order to see groups of groups, use the `groups/members` request.

Example: `GET https://10.10.10.10/api/v3/groups/_4L0PcE9SEemD69d9a0uLJQ/entities`

Response: A list of `EntityApiDTO`s representing the entities in the specified group.

```
[
  {
    "links": [
      {
        "rel": "self",
        "href": "https://10.10.10.10/api/v3/actions/_pJg98EQaEemD69d9a0uLJQ"
      },
      {
        "rel": "self",
        "href": "https://10.10.10.10/api/v3/actions/2500817580000"
      }
    ],
    "uuid": "_pJg98EQaEemD69d9a0uLJQ",
    "createTime": "2019-03-11T12:28:03-04:00",
    "actionType": "RIGHT_SIZE",
    "actionState": "PENDING_ACCEPT",
    "actionMode": "MANUAL",
    "details": "Scale down Memory reservation for Virtual Machine 'AJ_Test' from 2 GB to 1 GB",
    "importance": 0,
    "target": {
      "uuid": "421dd4ce-a4c1-231a-71e7-64fbf598bc65",
      "displayName": "AJ_Test",
      "className": "VirtualMachine",
      "aspects": {
        "virtualMachineAspect": {
          "os": "CentOS 4/5/6/7 (64-bit)",
          "ip": [
            "10.10.10.107"
          ],
          "connectedNetworks": [
            {
              "uuid": "34f9f5364cd843aa1fe99900621ee3b51a7dc6e5",
              "displayName": "VM Network"
            }
          ],
          "numVCPUs": 1,
          "ebsOptimized": false
```

```
        }
      },
      "environmentType": "ONPREM"
    },
    "currentEntity": {
      "uuid": "_pJlPYUQaEemD69d9a0uLJQ",
      "className": "Mem"
    },
    "newEntity": {
      "uuid": "_pLCA40QaEemD69d9a0uLJQ",
      "className": "Mem",
      "aspects": {
        "virtualMachineAspect": {
          "os": "Linux (Free)",
          "ebsOptimized": false
        }
      }
    },
    "currentValue": "2097152.0",
    "newValue": "1048576.0",
    "resizeToValue": "1048576.0",
    "risk": {
      "uuid": "_pL_qMEQaEemD69d9a0uLJQ",
      "subCategory": "Efficiency Improvement",
      "description": "Underutilized Memory reservation",
      "severity": "MINOR",
      "importance": 0
    },
    "currentLocation": {
      "uuid": "cdcd967d1e08d20936c3aa9a0879c635214c79ee",
      "displayName": "ACM",
      "className": "DataCenter"
    },
    "newLocation": {
      "uuid": "cdcd967d1e08d20936c3aa9a0879c635214c79ee",
      "displayName": "ACM",
      "className": "DataCenter"
    },
    "actionID": 2500817580000
  }
]
```


Getting the parent groups of a group

For a specified group, gets any group that the specified group belongs to.

> **NOTE:** Workload Optimization Manager contains a default parent group for user-created groups called "My Groups". This group should not be modified or deleted.

This request takes the following parameters:

- group_Uuid

  The UUID of the group whose actions will be returned.

- path

  Default: `false`. When `true`, this request returns the parent group, along with any groups traversed to obtain the parent group. For groups that are not nested, this returns the group whose ID you entered and its parent group.

Example: `GET https://10.10.10.10/api/v3/groups/_4L0PcE9SEemD69d9a0uLJQ/groups?path=true`

Response: A list of `GroupApiDTO`s representing the group and parent group.

```
[
  {
    "links": [
      {
        "rel": "self",
        "href": "https://10.10.10.10/api/v3/groups/_S4k68yFvEd-S4dn3T4_SSA?include_aspects=false"
      }
    ],
    "uuid": "_S4k68yFvEd-S4dn3T4_SSA",
    "displayName": "My Groups",
    "className": "Group",
    "environmentType": "ONPREM",
    "entitiesCount": 1,
    "membersCount": 3,
    "groupType": "ServiceEntity",
    "severity": "Major",
    "isStatic": false,
    "logicalOperator": "AND",
    "activeEntitiesCount": 1
  },
  {
    "links": [
      {
        "rel": "self",
        "href": "https://10.10.10.10/api/v3/groups/_4L0PcE9SEemD69d9a0uLJQ?include_aspects=false"
      }
    ],
    "uuid": "_4L0PcE9SEemD69d9a0uLJQ",
    "displayName": "Dallas-Dynamic",
    "className": "Group",
    "environmentType": "ONPREM",
    "entitiesCount": 1,
    "membersCount": 1,
    "groupType": "VirtualMachine",
    "severity": "Major",
    "isStatic": false,
    "logicalOperator": "AND",
    "criteriaList": [
      {
        "expVal": "AJ.*",
        "expType": "EQ",
        "filterType": "vmsByName",
        "caseSensitive": false
      }
    ],
    "activeEntitiesCount": 1
  }
]
```


Getting the members of a group

Referenced Endpoint: Entities (on page 1195)

Gets all members of a group.

> **NOTE:** This request returns nested groups and the member entities of each. To get only the *entities* in the specified group, use the `groups/entities` request.

Example: `GET https://10.10.10.10/api/v3/groups/_4L0PcE9SEemD69d9a0uLJQ/members`

Response: A list of `serviceEntityApiDTO`s representing the members of the group:

```
[
  {
    "links": [
      {
        "rel": "self",
        "href":
"https://10.10.10.10/api/v3/entities/183b37c68211668aecd36eadcdce9110ebc58f85?include_aspects=false"
      }
    ],
    "uuid": "183b37c68211668aecd36eadcdce9110ebc58f85",
    "displayName": "Olga",
    "className": "VirtualDataCenter",
    "priceIndex": 1,
    "state": "ACTIVE",
    "severity": "Normal",
    "discoveredBy": {
      "uuid": "_h3Km4Bs3EemMFrqlXNi1yQ",
      "displayName": "vsphere-dc23.corp.vmturbo.com",
      "type": "vCenter"
    },
    "remoteId": "resgroup-96",
    "environmentType": "ONPREM"
  },
  {
    "links": [
      {
        "rel": "self",
        "href":
"https://10.10.10.10/api/v3/entities/2332ba0f9dc14c07f11adb38e12df43c4bc8ca56?include_aspects=false"
      }
    ],
    "uuid": "2332ba0f9dc14c07f11adb38e12df43c4bc8ca56",
    "displayName": "olga_do_not_touch_1",
    "className": "VirtualDataCenter",
    "priceIndex": 1,
    "state": "ACTIVE",
    "severity": "Normal",
    "discoveredBy": {
      "uuid": "_5kMqwNLnEeiJs9KWCzOdUQ",
      "displayName": "vsphere-dc1.dev.mycorp.com",
      "type": "vCenter"
    },
    "remoteId": "resgroup-1088",
    "environmentType": "ONPREM"
  },
  {
    "links": [
      {
        "rel": "self",
        "href":
"https://10.10.10.10/api/v3/entities/6668eb0940bf970d24eb13279b31a52f3c44a91f?include_aspects=false"
      }
    ],
    "uuid": "6668eb0940bf970d24eb13279b31a52f3c44a91f",
    "displayName": "Gilad",
    "className": "VirtualDataCenter",
    "priceIndex": 1,
    "state": "ACTIVE",
    "severity": "Normal",
    "discoveredBy": {
      "uuid": "_lj_tkH39EeiKy53b3GBglA",
      "displayName": "vsphere-dc5.dev.mycorp.com",
      "type": "vCenter"
    },
    "remoteId": "resgroup-1953",
```

```
    "environmentType": "ONPREM"
  },
  {
    "links": [
      {
        "rel": "self",
        "href":
"https://10.10.10.10/api/v3/entities/20b9e0acf4c6f7d5b00f744d1dc954616095f1f2?include_aspects=false"
      }
    ],
    "uuid": "20b9e0acf4c6f7d5b00f744d1dc954616095f1f2",
    "displayName": "ACM",
    "className": "VirtualDataCenter",
    "priceIndex": 1,
    "state": "ACTIVE",
    "severity": "Normal",
    "discoveredBy": {
      "uuid": "_lj_tkH39EeiKy53b3GBglA",
      "displayName": "vsphere-dc5.dev.mycorp.com",
      "type": "vCenter"
    },
    "remoteId": "resgroup-v1946",
    "environmentType": "ONPREM"
  },
  {
    "links": [
      {
        "rel": "self",
        "href":
"https://10.10.10.10/api/v3/entities/56e6cfd2bc88ef5675df9dc3e06796ceb8b6e301?include_aspects=false"
      }
    ],
    "uuid": "56e6cfd2bc88ef5675df9dc3e06796ceb8b6e301",
    "displayName": "olga_do_not_touch_7",
    "className": "VirtualDataCenter",
    "priceIndex": 1,
    "state": "ACTIVE",
    "severity": "Normal",
    "discoveredBy": {
      "uuid": "_GRNtoNu0Eeilt9UDtI29QA",
      "displayName": "vsphere-dc7.dev.mycorp.com",
      "type": "vCenter"
    },
    "remoteId": "resgroup-1066",
    "environmentType": "ONPREM"
  }
]
```


Getting settings of a group

Referenced Endpoint: Settings (on page 1363)

Gets both global and user-modified settings for the specified group. To get a single setting, pass the setting UUID as part of the request URL. This request takes the following parameters:

- group_Uuid

  The Uuid of the group whose settings will be returned

- include_settingspolices

  Default: `false`. When `true`, returns the settings policies affecting the specified group, in addition to the settings.

Example: `GET https://10.10.10.10/api/v3/groups/_WvaGkOBjEeierZ9b1bxmgg/settings?include_settingspolicies=false`

Response: A list of `SettingsManagerApiDTO`s representing the settings, or settings and settings policies for the specified group:

```
[
  {
    "uuid": "appsettingsmanager",
    "displayName": "Application Discovery",
    "category": "AppDiscovery",
    "settings": [
      {
        "uuid": "Username",
        "displayName": "Username",
        "value": "",
        "defaultValue": "",
        "valueType": "STRING",
        "entityType": "VirtualMachine",
        "sourceGroupName": "Global"
      },
      {
        "uuid": "Password",
        "displayName": "Password",
        "value": "",
        "defaultValue": "",
        "valueType": "STRING",
        "entityType": "VirtualMachine",
        "sourceGroupName": "Global"
      },
      ...
    },
    {
    "uuid": "controlmanager",
    "displayName": "Action Workflow",
    "category": "Automation",
    "settings": [
      {
        "uuid": "moveVMWorkflow",
        "displayName": "Move Workflow",
        "value": "",
        "defaultValue": "",
        "valueType": "STRING",
        "entityType": "VirtualMachine",
        "sourceGroupName": "Global"
      }
    ]
  }
  ...
}
```

Editing a specified setting of a group

Referenced Endpoint: Settings (on page 1363)

Edits a specified setting for the specified group. This request takes the following parameters:

- group_Uuid

  The Uuid of the group whose setting will be edited
- settingsManager_Uuid

  The Uuid of the settings manager whose setting will be edited
- setting_Uuid

  The Uuid of the setting to change

Example: `PUT https://10.10.10.10/api/v3/groups/_WvaGkOBjEeierZ9b1bxmgg/settings/appsettingsmanager/Retry_Interval`

Example Input:

The new value for the setting:

```
{
  "value": "61"
}
```

Response: A `ServiceEntityApiDTO` representing the specified setting, with your changed value(s):

```
{
  "uuid": "Retry_Interval",
  "displayName": "Retry Interval [mins]",
  "value": "61",
  "defaultValue": "60.0",
  "valueType": "NUMERIC",
  "min": 0,
  "max": 50000,
  "entityType": "VirtualMachine",
  "sourceGroupName": "A_Few_ACM_VMs"
}
```

Getting settings policies of a group

Referenced Endpoint: Settings Policies (on page 1366)

Gets the settings policies of the specified group.

Example:

Response:

Getting statistics of a group

Referenced Endpoint: Stats (on page 1372)

Gets all statistics for the specified group:

> **NOTE:**
> When you retrieve statistics for a cluster, using the `/groups` endpoint will return the aggregated statistics for all applicable cluster entities. For individual stats per entity, use the `/stats` endpoint.
>
> Some statistics are only valid for certain scopes. When you request given statistics for a scope that doesn't include them, the API does not return any data. For example, the `numVMs` statistic is valid in only the Market and Cluster scopes – the API omits `numVMs` data from the response when you query other scopes.

Example: `https://10.10.10.10/api/v3/groups/_WvaGkOBjEeierZ9b1bxmgg/stats`

Response: A list of `StatSnapshotApiDTO`s representing the statistics for the specified group:

```
[
  {
    "displayName": "A_Few_ACM_VMs",
    "date": "2019-03-31T17:11:35-04:00",
    "statistics": [
      {
```

```
      "displayName": "(multiple)",
      "name": "priceIndex",
      "values": {
        "max": 1.44,
        "min": 1.03,
        "avg": 1.19,
        "total": 4.77
      },
      "value": 1.19
    },
    {
      "displayName": "(multiple)",
      "name": "Q1VCPU",
      "capacity": {
        "max": 20000,
        "min": 20000,
        "avg": 20000,
        "total": 60000
      },
      "reserved": {
        "max": 10000,
        "min": 10000,
        "avg": 10000,
        "total": 30000
      },
      "filters": [
        {
          "type": "relation",
          "value": "bought"
        }
      ],
      ...
    }
    ...
    ]
  }
```

Getting a filtered list of statistics of a group

Referenced Endpoint: Stats (on page 1372)

Gets a filtered list of statistics for the specified group. For more information about creating a filtered list of statistics, including example POSTs and responses, see *Filtered Statistics* in the Statistics Endpoint *(on page 1372)*

Example: `POST https://10.10.10.10/api/v3/groups/_WvaGkOBjEeierZ9b1bxmgg/stats`

Example input:

```
{
  "endDate": "-1d",
  "startDate": "-1w",
  "statistics": [
    {
      "filters": [
        {
          "type": "relation",
          "value": "bought"
        }
      ]
    }
  ]
```

```
}
```

Response: A list of `StatSnapshotApiDTO`s representing the statistics for the specified group:


```
[
  {
    "displayName": "A_Few_ACM_VMs",
    "date": "2019-03-24T00:00:00-04:00",
    "statistics": [
      {
        "name": "Mem",
        "capacity": {
          "max": 134214448,
          "min": 134214448,
          "avg": 134214448,
          "total": 1073715580
        },
        "filters": [
          {
            "type": "relation",
            "value": "bought"
          }
        ],
        "units": "KB",
        "values": {
          "max": 2073084,
          "min": 679782.1,
          "avg": 1124840.4,
          "total": 4499361.5
        },
        "value": 1124840.4
      },
      {
        "name": "MemAllocation",
        "capacity": {
          "max": 504346624,
          "min": 504346624,
          "avg": 504346624,
          "total": 4034772990
        },
        "filters": [
          {
            "type": "relation",
            "value": "bought"
          }
        ],
        "units": "KB",
        "values": {
          "max": 2073084,
          "min": 679782.1,
          "avg": 1124840.4,
          "total": 4499361.5
        },
        "value": 1124840.4
      },
      {
        "name": "Q1VCPU",
        "capacity": {
          "max": 20000,
          "min": 20000,
          "avg": 20000,
          "total": 120000
        },
        "filters": [
```

```
          {
            "type": "relation",
            "value": "bought"
          }
        ],
        "values": {
          "max": 90,
          "min": 14,
          "avg": 41.08,
          "total": 123.23
        },
        "value": 41.08
      }
      ...
    ]
  }
]
```

Getting the supply chain of a group

Referenced Endpoint: Supply Chains (on page 1389)

Using one or more of the following parameters, you can filter the supply chain that is returned by this request:

- market_Uuid

  The UUID of the market whose supply chain will be returned

- types

  The entity types whose supply chain will be returned. Choose from:

  - APPLICATION
  - APPLICATION_SERVER
  - BUSINESS_ACCOUNT
  - BUSINESS_UNIT
  - CHASSIS
  - CLOUD_SERVICE
  - CONTAINER
  - CONTAINER_POD
  - DATABASE
  - DATABASE_SERVER
  - DATACENTER
  - DISK_ARRAY
  - IO_MODULE
  - LOAD_BALANCER
  - LOGICAL_POOL
  - NETWORK
  - PHYSICAL_MACHINE
  - STORAGE
  - STORAGE_CONTROLLER
  - SWITCH
  - VIRTUAL_APPLICATION

  - VIRTUAL_DATACENTER
  - VIRTUAL_MACHINE
- entity_states

  Only return entities in the specified states. Choose from:
  - ACTIVE:

    Entity is actively consuming resources.
  - EVACUATED:

    Entity is ready to suspend and will not participate in the market, but will not generate a SUSPEND action.

    NOTE: This state replaces the SUSPEND state, and requires additional configuration to enable. Workload Optimization Manager does not recommend enabling this state.
  - FAILOVER:

    This entity is reserved for failover and will not participate in the market.
  - IDLE:

    Entity is powered off and not actively consuming market resources.
  - LAUNCH:

    Entity is starting.
  - MAINTENANCE:

    Entity is in maintenance mode.
  - NOT_MONITORED:

    Entity is not currently monitored by Workload Optimization Manager.
  - QUEUED:

    Entity is in the process of changing from one state to another.

    NOTE: This state is internal to Workload Optimization Manager and will not appear in the UI, but may appear in an API response if the call is executed at the precise moment the entity is changing state.
  - RESOURCE_ALLOCATION:

    Cloud entity is being created.
  - RESOURCE_RELEASE:

    Cloud Entity is being deleted.
  - SUSPEND:

    Entity is suspended or the guest OS is in sleep, standby, or suspended state. The entity is not removed from the environment, but will not be used in market decisions. Based on market conditions, suspended entities can be started, re-adding them to the market.
  - SUSPEND_PENDING:

    Entity is making preparations to suspend, or is the next entity of its type to suspend.
  - TERMINATE_PENDING:

    Entity is in the process of being deleted.
  - UNKNOWN:

    Workload Optimization Manager is unable to obtain the current state of the entity.

    This occurs when the entity is reported to the hypervisor as having a state of DISCONNECTED, ORPHANED, or INVALID.
- environment_type

- Only return actions that affect entities in the specified environment type. `[ CLOUD, HYBRID, ONPREM ]`.
- detail_type

  The level of detail returned for each item in the returned supply chain. `[ compact, entity, aspects ]`.
- health

  Return a small summary of the health of the returned supply chain.

Example: `GET https://10.10.10.10/api/v3/groups/_WvaGkOBjEeierZ9b1bxmgg/supplychains?types=PhysicalMachine&entity_states=ACTIVE&entity_states=SUSPEND&entity_states=MAINTENANCE&envir`

Response: A `SupplychainApiDTO` representing the supply chain of the specified group:


```
 "seMap": {
   "PhysicalMachine": {
     "depth": 1,
     "entitiesCount": 3,
     "healthSummary": {
       "Minor": 1,
       "Normal": 2
     },
     "stateSummary": {
       "ACTIVE": 3
     },
     "connectedProviderTypes": [
       "DataCenter",
       "Storage"
     ],
     "instances": {
       "34313836-3333-5553-4537-33364e385146": {
         "links": [
           {
             "rel": "self",
             "href":
"https://10.10.10.10/api/v3/entities/34313836-3333-5553-4537-33364e385146?include_aspects=false"
           }
         ],
         "uuid": "34313836-3333-5553-4537-33364e385146",
         "displayName": "dc17-host-01.dev.mycorp.com",
         "className": "PhysicalMachine",
         "priceIndex": 3.2925103,
         "state": "ACTIVE",
         "severity": "Normal",
         "discoveredBy": {
           "uuid": "_yXCEIF6hEeiNCKdg4ZVvVQ",
           "displayName": "vsphere-dc17.dev.mycorp.com",
           "type": "vCenter"
         },
         "remoteId": "host-621",
         "environmentType": "ONPREM"
       },
       ...
```


```
     }
   }
 }
}
```

Groups endpoint tips

GroupBy field criteria

When using the POST requests to query actions and action statistics, you can use the `groupBy` parameter to organize the results. The valid values for this parameter differ by entity type:

> **NOTE:**
> When sending an inputDTO using this field, only the value is needed.
>
> **Example:**`"groupBy": "vmsByPMName"`

```
{
  "VirtualMachine":{
    "criteria":[
      {
        "filterType":"vmsByName"
      },
      {
        "filterType":"vmsByPMName"
      },
      {
        "filterType":"vmsByStorage"
      },
      {
        "filterType":"vmsByNetwork"
      },
      {
        "filterType":"vmsByApplication"
      },
      {
        "filterType":"vmsByDatabaseServer"
      },
      {
        "filterType":"vmsByDatabaseServerVersion"
      },
      {
        "filterType":"vmsByDC"
      },
      {
        "filterType":"vmsByVDC"
      },
      {
        "filterType":"vmsByDCnested"
      },
      {
        "filterType":"vmsByNumCPUs"
      },
      {
        "filterType":"vmsByMem"
      },
      {
        "filterType":"vmsByGuestName"
      },
      {
        "filterType":"vmsByAltName"
      },
      {
        "filterType":"vmsByClusterName"
      },
```

```
      {
        "filterType":"vmsByDiskArrayName"
      },
      {
        "filterType":"vmsByLogicalPoolName"
      },
      {
        "filterType":"vmsByProfileName"
      },
      {
        "filterType":"vmsByTag"
      },
      {
        "filterType":"vmsByState"
      },
      {
        "filterType":"vmsByBusinessAccountUuid"
      },
      {
        "filterType":"vmsByResourceGroupUuid"
      }
    ]
  },
  "VirtualDataCenter":{
    "criteria":[
      {
        "filterType":"vdcsByName"
      },
      {
        "filterType":"vdcsByVDCName"
      },
      {
        "filterType":"vdcsByTag"
      },
      {
        "filterType":"vdcsByState"
      }
    ]
  },
  "PhysicalMachine":{
    "criteria":[
      {
        "filterType":"pmsByName"
      },
      {
        "filterType":"pmsByStorage"
      },
      {
        "filterType":"pmsByNetwork"
      },
      {
        "filterType":"pmsBySwitch"
      },
      {
        "filterType":"pmsByNumVms"
      },
      {
        "filterType":"pmsByDC"
      },
      {
        "filterType":"pmsByMem"
      },
      {
        "filterType":"pmsByNumCPUs"
      },
```

```
      {
        "filterType":"pmsByVendorName"
      },
      {
        "filterType":"pmsByCPUModel"
      },
      {
        "filterType":"pmsByModel"
      },
      {
        "filterType":"pmsByTimezone"
      },
      {
        "filterType":"pmsByClusterName"
      },
      {
        "filterType":"pmsByTag"
      },
      {
        "filterType":"pmsByState"
      }
    ]
  },
  "Storage":{
    "criteria":[
      {
        "filterType":"storageByName"
      },
      {
        "filterType":"storageByTag"
      },
      {
        "filterType":"storageByVMs"
      },
      {
        "filterType":"storageByDC"
      },
      {
        "filterType":"storageByPMCluster"
      },
      {
        "filterType":"storageByState"
      }
    ]
  },
  "Application":{
    "criteria":[
      {
        "filterType":"appsByName"
      },
      {
        "filterType":"appsByTag"
      }
    ]
  },
  "ApplicationComponent":{
    "criteria":[
      {
        "filterType":"appSrvsByName"
      }
    ]
  },
  "WebServer":{
    "criteria":[
      {
```

```
        "filterType":"wbSrvsByName"
      }
    ]
  },
  "BusinessApplication":{
    "criteria":[
      {
        "filterType":"busAppsByName"
      }
    ]
  },
  "Database":{
    "criteria":[
      {
        "filterType":"databaseByName"
      },
      {
        "filterType":"databaseByTag"
      },
      {
        "filterType":"databaseByBusinessAccountUuid"
      },
      {
        "filterType":"databaseByResourceGroupUuid"
      }
    ]
  },
  "DatabaseServer":{
    "criteria":[
      {
        "filterType":"databaseServerByName"
      },
      {
        "filterType":"databaseServerByTag"
      },
      {
        "filterType":"databaseServerByBusinessAccountUuid"
      },
      {
        "filterType":"databaseServerByEngine"
      },
      {
        "filterType":"databaseServerByEdition"
      },
      {
        "filterType":"databaseServerByVersion"
      }
    ]
  },
  "VirtualApplication":{
    "criteria":[
      {
        "filterType":"vappsByName"
      }
    ]
  },
  "Cluster":{
    "criteria":[
      {
        "filterType":"clustersByName"
      },
      {
        "filterType":"clustersByTag"
      }
    ]
```

```
  },
  "DataCenter":{
    "criteria":[
      {
        "filterType":"datacentersByName"
      },
      {
        "filterType":"datacentersByTag"
      }
    ]
  },
  "Group":{
    "criteria":[
      {
        "filterType":"groupsByName"
      },
      {
        "filterType":"groupsByTargetName"
      }
    ]
  },
  "StorageCluster":{
    "criteria":[
      {
        "filterType":"storageClustersByName"
      }
    ]
  },
  "DiskArray":{
    "criteria":[
      {
        "filterType":"diskarrayByName"
      }
    ]
  },
  "Zone":{
    "criteria":[
      {
        "filterType":"zonsByName"
      }
    ]
  },
  "Region":{
    "criteria":[
      {
        "filterType":"regsByName"
      }
    ]
  },
  "Network":{
    "criteria":[
      {
        "filterType":"netsByName"
      }
    ]
  },
  "LoadBalancer":{
    "criteria":[
      {
        "filterType":"lbsByName"
      }
    ]
  },
  "Chassis":{
    "criteria":[
```

```
      {
        "filterType":"chasByName"
      }
    ]
  },
  "IOModule":{
    "criteria":[
      {
        "filterType":"ioModuleByName"
      }
    ]
  },
  "StorageController":{
    "criteria":[
      {
        "filterType":"storagecontrollerByName"
      }
    ]
  },
  "DPod":{
    "criteria":[
      {
        "filterType":"dpodByName"
      }
    ]
  },
  "VPod":{
    "criteria":[
      {
        "filterType":"vpodByName"
      }
    ]
  },
  "LogicalPool":{
    "criteria":[
      {
        "filterType":"logicalPoolByName"
      }
    ]
  },
  "Switch":{
    "criteria":[
      {
        "filterType":"switchByName"
      }
    ]
  },
  "Container":{
    "criteria":[
      {
        "filterType":"containersByName"
      },
      {
        "filterType":"containersByVMName"
      }
    ]
  },
  "ContainerPod":{
    "criteria":[
      {
        "filterType":"containerpodsByName"
      },
      {
        "filterType":"containerpodsByVMName"
      }
```

```
    ]
  },
  "ServiceEntity":{
    "criteria":[
      {
        "filterType":"workloadByBusinessAccountUuid"
      },
      {
        "filterType":"workloadByResourceGroupUuid"
      }
    ]
  },
  "Workload":{
    "criteria":[
      {
        "filterType":"workloadByBusinessAccountUuid"
      },
      {
        "filterType":"workloadByResourceGroupUuid"
      }
    ]
  },
  "ResourceGroup":{
    "criteria":[
      {
        "filterType":"resourceGroupByName"
      },
      {
        "filterType":"resourceGroupByUuid"
      },
      {
        "filterType":"resourceGroupByBusinessAccountUuid"
      },
      {
        "filterType":"resourceGroupByTag"
      }
    ]
  },
  "BusinessAccount":{
    "criteria":[
      {
        "filterType":"businessAccountByName"
      },
      {
        "filterType":"businessAccountByUuid"
      },
      {
        "filterType":"subBusinessAccountOfUuid"
      },
      {
        "filterType":"businessAccountValidationStatus"
      },
      {
        "filterType":"businessAccountTargetName"
      },
      {
        "filterType":"businessAccountCloudProvider"
      }
    ]
  },
  "BillingFamily":{
    "criteria":[
      {
        "filterType":"billingFamilyByName"
      }
```

```
    ]
  },
  "ViewPod":{
    "criteria":[
      {
        "filterType":"viewPodByName"
      }
    ]
  },
  "DesktopPool":{
    "criteria":[
      {
        "filterType":"desktopPoolByName"
      },
      {
        "filterType":"desktopPoolByViewPod"
      }
    ]
  },
  "BusinessUser":{
    "criteria":[
      {
        "filterType":"businessUserByName"
      },
      {
        "filterType":"businessUserByDesktopPool"
      }
    ]
  }
}
```

Licenses endpoint

A product license determines the product features you can use, and the number of workloads that you can manage with Workload Optimization Manager. To increase the number of workloads you can manage, you can add additional licenses to your Workload Optimization Manager installation. Note that if you use multiple licenses, they must all support the same feature set.

In all circumstances, you should contact your sales representative to make sure that you get the correct license, and that you know how to install it properly.

Using the licenses endpoint, you can:

- Get a license summary
- Get a list of licenses currently added to your Workload Optimization Manager appliance
- Add a license
- Delete a license

Licenses requests

Getting licenses

Gets a list of licenses. For instances with multiple licenses, the `licenses/summary` request will return a summed total of licensed workload. For example, if you have two licenses for 10,000 workloads each, the `/licenses` request will return both LicenseApiDTOs in a list. The `/licenses/summary` call will return a single summed LicenseApiDTO, which will state that you have capacity for 20,0000 workloads.

Examples:

- **List of Licenses:** `GET https://10.10.10.10/api/v3/licenses`
- **License Summary:** `GET https://10.10.10.10/api/v3/licenses/summary`

Adding a license

Adds a license to your Workload Optimization Manager installation. You may optionally specify the `dryRun` parameter which, when set to true, will validate the license file but not apply it to the appliance.

> **NOTE:**
> This request requires a multipart file as input. Due to this requirement, this request cannot be executed directly from the Swagger.

Example: `POST https://10.10.10.10/api/v3/licenses?file=%5Bobject%20File%5D&dryRun=true`

Deleting a license

Removes the specified license from your Workload Optimization Manager installation. Deleting a license may reduce your licensed workload capacity below your current usage. In this case, you will be unable to add more targets until more workload capacity is available.

Example: `DELETE https://10.10.10.10/api/v3/licenses/_I_WVcJtWEeiGEs_doVlDOw`

Logs requests

The logs endpoint is not implemented at this time.

Markets endpoint

Workload Optimization Manager uses market-based analysis to perform workload management. To do this, it constructs a model of your environment, representing each entity as a buyer and seller in a market. You can access this model via a named market resource. There are two types of markets:

- **Real-Time Market**: The real-time market performs analysis and workload management on your environment. You can use the real-time market to access entities and get current or historical data about them.

  The `displayName` of the real-time market is `Market`. This market should always be in the `RUNNING` state. You should not attempt to use the API to stop this market.

  Workload Optimization Manager performs discovery to populate the topology it manages (the collection of entities in the real-time market). For this reason, you should never use the API add or delete entities in the real-time market.

- **Plan Market**: A plan market begins as a snapshot of the real-time market. You POST a scenario to the real-time market, and Workload Optimization Manager makes a copy of the real-time market to create a plan market. This plan market is just like the real-time market, except for any changes that were declared in the scenario. For example, if the scenario declares a scope for the plan, then the plan market only contains entities that are within that scope.

  Workload Optimization Manager runs economic cycles (buy/sell cycles) against this plan market until there are no more meaningful improvements to be made. At this point the plan run is completed.

  Note that once Workload Optimization Manager creates the plan market, that market stays in memory until you delete it. Also note that you can later apply a scenario to the plan market. This is how you run a plan on a plan. The logic flow is the same, and Workload Optimization Manager creates yet another plan market to run the analysis.

This endpoint contains references to start and end times for certain operations. For more information, including valid time formats, see Time in the Workload Optimization Manager API *(on page 1141)*

Using the markets endpoint of the Workload Optimization Manager API, you can:

- Get a list of markets
- Save or Stop a plan market
- Delete a plan market
- See unplaced entities in a plan market
- Get filtered and unfiltered lists of actions, entities, notifications, policies, reservations, and statistics for a particular market
- Get, edit, and delete placement policies in the real-time market

For more information about managing placement policies with the API, see Working with policies in the API *(on page 1148)*.

Determining the status of the main market

After restarting the Workload Optimization Manager instance, the UI and API may be available before the real-time market has finished initializing and loading the topology of your environment. To check market status, execute the `GET /markets/market_uuid` request with a market UUID of `Market`.

If information about the market is returned, it has initialized. If the real-time market is not initialized, the following response will be returned:

```
{
  "type":400,
  "exception":"com.vmturbo.api.exceptions.OperationFailedException: Infrastructure is not controlled yet.
 Could not get market.",
  "message":"Infrastructure is not controlled yet. Could not get market."
}
```

Markets requests

Getting a list of markets

Returns all the Markets currently stored on your Workload Optimization Manager appliance. To get the details of a single market, include the UUID of the market in the request.

> **NOTE:**
> One market will have the `displayName` of *Market*. This market is the real-time market. You should not add nor delete entities directly to this market using the API, nor should you attempt to execute a STOP action on this market.

Examples:

- All Markets: `GET https://10.10.10.10/api/v3/markets`
- Single Market: `GET https://10.10.10.10/api/v3/markets/_0x3OYUglEd-gHc4L513yOA`

Response: A list of `MarketApiDTO`s representing the markets on your appliance.

```
[
  {
    "links":[
      {
        "rel":"self",
        "href":"https://10.10.10.10/api/v3/markets/_0x3OYUglEd-gHc4L513yOA"
      }
    ],
    "uuid":"_0x3OYUglEd-gHc4L513yOA",
```

```
    "displayName":"Market",
    "className":"Market",
    "state":"RUNNING",
    "unplacedEntities":false,
    "environmentType":"HYBRID"
  },
  ...
]
```

Stopping or saving plan markets

For troubleshooting, you may need to stop or save a plan market that is taking a long time to run, or does not seem to complete. These operations should never be used on the real-time market. Note that a manually-stopped plan market may have incomplete data.

This request takes the following parameters:

- market_Uuid

  The UUID of the market that will be stopped or saved.
- operation

  The operation to perform on the market. `[ stop, save ]`

Example: `POST https://10.10.10.10/api/v3/markets/_NO8_sCi_EemDHtuQzfdzxw?operation=stop`

Response: The `MarketApiDTO` of the specified market, showing its state as `STOPPING`.

```
{
  "links": [
    {
      "rel": "self",
      "href": "https://10.10.10.10/api/v3/markets/_NO8_sCi_EemDHtuQzfdzxw"
    },
    {
      "rel": "Market notifications",
      "href": "https://10.10.10.10/api/v3/markets/_NO8_sCi_EemDHtuQzfdzxw/notifications"
    },
    {
      "rel": "Market actions",
      "href": "https://10.10.10.10/api/v3/markets/_NO8_sCi_EemDHtuQzfdzxw/actions?ascending=true"
    },
    {
      "rel": "To Entities",
      "href": "https://10.10.10.10/api/v3/markets/_NO8_sCi_EemDHtuQzfdzxw/entities"
    },
    {
      "rel": "To Constraints",
      "href": "https://10.10.10.10/api/v3/markets/_NO8_sCi_EemDHtuQzfdzxw/policies"
    },
    {
      "rel": "To Reservations",
      "href": "https://10.10.10.10/api/v3/markets/_NO8_sCi_EemDHtuQzfdzxw/reservations"
    },
    {
      "rel": "To Statistics",
      "href": "https://10.10.10.10/api/v3/markets/_NO8_sCi_EemDHtuQzfdzxw/stats"
    },
    {
      "rel": "To Unplaced Entities",
      "href": "https://10.10.10.10/api/v3/markets/_NO8_sCi_EemDHtuQzfdzxw/unplacedentities"
    },
    {
      "rel": "To Market List",
```

```
      "href": "https://10.10.10.10/api/v3/markets"
    }
  ],
  "uuid": "_NO8_sCi_EemDHtuQzfdzxw",
  "displayName": "ADD_WORKLOAD_administrator_1549313730114",
  "className": "Market",
  "state": "STOPPING",
  "scenario": {
    "uuid": "2452690315520",
    "displayName": "Add Workload 1",
    "owners": [],
    "type": "ADD_WORKLOAD"
  },
  "unplacedEntities": false,
  "runDate": "2019-02-04T15:55:30-05:00",
  "environmentType": "ONPREM"
}
```


Deleting a plan market

Deleting a plan market removes the market entity along with all the topological links, such as actions, notifications, entities, policies and reservations. This operation is permanent. To delete a plan market, pass the UUID of the market in the request.

This request takes the following parameters:

market_Uuid

The UUID of the market that will be deleted.

Example: `DELETE https://10.10.10.10/api/v3/markets/_NO8_sCi_EemDHtuQzfdzxw`

Response: The `MarketApiDTO` of the specified market, showing its state as `DELETING`.


```
{
  "links":[
    {
      "rel":"self",
      "href":"https://10.10.10.10/api/v3/markets/2452692215536"
    }
  ],
  "uuid":"2452692215536",
  "displayName":"ADD_WORKLOAD_administrator_1549313730114",
  "className":"Market",
  "state":"DELETING",
  "scenario":{
    "uuid":"2452692204928",
    "displayName":"Add Workload 1",
    "owners":[
      {
        "links":[
          {
            "rel":"self",
            "href":"https://10.10.10.10/api/v3/users/_4T_7kwY-Ed-WUKbEYSVIDw"
          }
        ],
        "uuid":"_4T_7kwY-Ed-WUKbEYSVIDw",
        "displayName":"Administrator User",
        "username":"administrator",
        "roleUuid":"_4UAioQY-Ed-WUKbEYSVIDw",
        "roleName":"administrator",
        "loginProvider":"Local",
        "type":"DedicatedCustomer",
        "showSharedUserSC":false
```

```
      }
    ],
    "type":"ADD_WORKLOAD"
  },
  "saved":true,
  "unplacedEntities":false,
  "runDate":"2019-02-04T15:55:30-05:00",
  "runCompleteDate":"2019-02-04T15:57:26-05:00",
  "environmentType":"ONPREM"
}
```


Getting a filtered list of actions for a market

Referenced Endpoint: Actions (on page 1161)

Gets a list of actions, given a market UUID and filtered by an ActionApiInputDTO using one or more of the following criteria:

- actionModeList
  - `DISABLED`:

    Do not recommend or perform the action. When you disable an action, the user interface will no longer display actions of that type.
  - `RECOMMENDED`:

    Recommend the action so a user can execute it via the given hypervisor or by other external means
  - `MANUAL`:

    Recommend the action, and provide the option to execute that action through the Workload Optimization Manager user interface or an /actions API request.
  - `AUTOMATED`:

    You have directed Workload Optimization Manager to execute the action automatically.
  - `COLLECTION` **(UNUSED)**:

    Legacy action mode, no longer used.
- actionStateList
  - `DISABLED`:

    An action whose policy has been disabled.
  - `RECOMMEND`:

    An action that cannot be automated due to policy settings or must be executed outside the system, such as the purchase of hardware.
  - `PENDING_ACCEPT`:

    When the action mode is set to manual, a recommended action that has not been accepted or cleared.
  - `ACCEPTED`:

    An action that has been accepted, but is not yet in progress.
  - `IN_PROGRESS`:

    An action that is being executed.
  - `SUCCEEDED`:

    An action that was accepted and completed successfully.
  - `FAILED`:

    An action that was accepted and did not complete successfully.
  - `REJECTED`:

    When the action mode is set to manual, an action that has been rejected by the user.

  - `CLEARED`:

    When the action mode is set to manual, an action that is no longer recommended by the market.
  - `ACCOUNTING`:

    For cloud entities, an action to resize in order to use a different instance template.
  - `QUEUED`:

    When more than 10 actions are to be executed for a single target, Workload Optimization Manager will place the 11th and all subsequent actions into a QUEUED state, and they will execute as the previous actions complete.
- actionTypeList
  - `START`:

    Start an entity, and add it to the Workload Optimization Manager market.
  - `MOVE`:

    Move an entity from one provider to another. For example, moving a VM between hosts, or a datastore between disk arrays.
  - `SUSPEND`:

    Suspend the entity to exclude its capacity from your environment. Suspend does not remove the entity from your environment, but the market does not include the entity in its analysis. Based on market conditions, analysis can recommend restarting suspended entities, which includes them in market analysis once again.
  - `ADD PROVIDER`:

    This is equivalent to the START action, but for storage entities.
  - `CHANGE`:

    Move a virtual machine's storage between datastores.
  - `PROVISION`:

    Add capacity to your environment.
  - `RECONFIGURE`:

    Analysis recommends this when a VM is not able to correct a misconfiguration via another type of action. For example, analysis can recommend a RECONFIGURE action when a placement policy requires a VM to move to a host on a particular cluster, but cross-cluster moves are not permitted.
  - `DELETE`:

    This is equivalent to the SUSPEND action, but for storage entities.
  - `RESERVE_ON_PM`:

    When using the reservation functionality, this is the action to reserve the resources of a VM on a particular host.
  - `RESERVE_ON_DS`:

    When using the reservation functionality, this is the action to reserve the resources of a VM on a particular datastore.
  - `RESIZE_FOR_EFFICIENCY`:

    A resize that reduces the capacity of a commodity, such as vCPU, or vMEM.
  - `RESIZE_FOR_PERFORMANCE`:

    A resize that increases the capacity of a commodity, such as vCPU, or vMEM.
- cleared

  Default: `false`.

  When `true`, actions that were removed from the action list before being executed will be returned.
- costType
  - `Saving`

    An action that will decrease cloud spending.

  - `Investment`

    An action that will increase cloud spending.

- cumulative

  Default: `false`.

  When `true`, shows the savings for each action in a scope as a cumulative total, rather than showing the savings for individual actions.

- environmentType
  - `ONPREM`:

    Returns only actions related to entities that are part of your on-prem environment.
  - `CLOUD`:

    Returns only actions related to entities that are part of your cloud environment.

- groupBy

  You may group the actions returned by the following criteria: `[actionModes, actionStates, actionTypes, risk, riskSeverity, riskSubCategory]`

- hasReservedInstance

  Default: `false`.

  When `true`, only actions that are related to reserved instances will be returned.

- relatedEntityTypes

  Returns only actions related to entities of these types within the scope. For example, if your scope is a virtual datacenter and you specify VirtualMachine as a related entity type, the request will include only those actions related to VMs in the specified virtual datacenter.

  Choose from the following entity types: `[ Datacenter, PhysicalMachine, VirtualMachine, Storage, Application, Chassis, DiskArray, IOModule, StorageControl, Switch, VirtualDataCenter, VPod, DPod, Container, Database, DatabaseServerContainer, LogicalPool ]`

- riskSeverityList

  Returns only actions of the given severities: `[ Unknown, Normal, Minor, Major, Critical ]`

- riskSubCategoryList
- returns only actions of the given subcategories: `[ Performance Assurance, Efficiency Improvement, Prevention, Compliance ]`

This request takes the following parameters:

- market_Uuid

  The UUID of the market whose actions will be returned.

- cursor

  When making subsequent calls to retrieve additional results, this is the cursor returned by the last call to this method.

- limit

  The maximum number of items to return. Must be a positive integer. If not set, the server will provide a default (20).

- order_by

  The field to order the results by. `[ creation_date, disruptiveness, file_size, idleness, name, reversibility, risk_category, savings, severity ]`.

- ascending

  Default: true. When false, results will be in descending order.

Example: `POST https://10.10.10.10/api/v3/markets/Market/actions?order_by=severity&ascending=true`

Example input: In this example, a list of actions that meet the following criteria will be returned and grouped by the `actionMode` of the action, in order of severity:

- Is in state `PENDING_ACCEPT`
- Is part of an on-prem environment
- Is an action for a virtual machine
- Is an action generated by the Workload Optimization Manager real-time market

```
{
  "actionInput":{
    "actionStateList":[
      "PENDING_ACCEPT"
    ],
    "environmentType":"ONPREM",
    "groupBy":[
      "actionModes"
    ]
  },
  "relatedType":"VirtualMachine",
  "scopes":[
    "Market"
  ]
}
```

Response: `ActionApiDTO`s meeting the specified criteria.

```
{
..
  "uuid": "_qFuM8CiOEemDHtuQzfdzxw",
  "createTime": "2019-02-04T10:07:58-05:00",
  "actionType": "RIGHT_SIZE",
  "actionState": "PENDING_ACCEPT",
  "actionMode": "MANUAL",
  "details": "Scale VirtualMachine 8f0e3136-b259-43e5-41f8-0d2c64c62acc from Standard_B2ms to
 Standard_E2s_v3",
  "importance": 0,
  ...
}
```

For information about filtering actions within a specified time period, see *Filtering Actions by Time* in Markets endpoint tips *(on page 1278)*.

Getting statistics for a filtered list of actions for a market

Referenced Endpoint: Actions (on page 1161)

Referenced Endpoint: Stats (on page 1372)

Gets a list of statistics for a filtered list of actions. Statistics returned include the number of actions, total savings/investment, and the number of entities affected by the actions. In addition to the Market UUID, provide an ActionApiInputDTO using one or more of the following criteria:

- actionModeList
  - `DISABLED`:

    Do not recommend or perform the action. When you disable an action, the user interface will no longer display actions of that type.
  - `RECOMMENDED`:

    Recommend the action so a user can execute it via the given hypervisor or by other external means

  - MANUAL:

    Recommend the action, and provide the option to execute that action through the Workload Optimization Manager user interface or an /actions API request.
  - AUTOMATED:

    You have directed Workload Optimization Manager to execute the action automatically.
  - COLLECTION **(UNUSED)**:

    Legacy action mode, no longer used.
- actionStateList
  - DISABLED:

    An action whose policy has been disabled.
  - RECOMMEND:

    An action that cannot be automated due to policy settings or must be executed outside the system, such as the purchase of hardware.
  - PENDING_ACCEPT:

    When the action mode is set to manual, a recommended action that has not been accepted or cleared.
  - ACCEPTED:

    An action that has been accepted, but is not yet in progress.
  - IN_PROGRESS:

    An action that is being executed.
  - SUCCEEDED:

    An action that was accepted and completed successfully.
  - FAILED:

    An action that was accepted and did not complete successfully.
  - REJECTED:

    When the action mode is set to manual, an action that has been rejected by the user.
  - CLEARED:

    When the action mode is set to manual, an action that is no longer recommended by the market.
  - ACCOUNTING:

    For cloud entities, an action to resize in order to use a different instance template.
  - QUEUED:

    When more than 10 actions are to be executed for a single target, Workload Optimization Manager will place the 11th and all subsequent actions into a QUEUED state, and they will execute as the previous actions complete.
- actionTypeList
  - START:

    Start an entity, and add it to the Workload Optimization Manager market.
  - MOVE:

    Move an entity from one provider to another. For example, moving a VM between hosts, or a datastore between disk arrays.
  - SUSPEND:

    Suspend the entity to exclude its capacity from your environment. Suspend does not remove the entity from your environment, but the market does not include the entity in its analysis. Based on market conditions, analysis can recommend restarting suspended entities, which includes them in market analysis once again.
  - ADD PROVIDER:

    This is equivalent to the START action, but for storage entities.

  - `CHANGE`:

    Move a virtual machine's storage between datastores.
  - `PROVISION`:

    Add capacity to your environment.
  - `RECONFIGURE`:

    Analysis recommends this when a VM is not able to correct a misconfiguration via another type of action. For example, analysis can recommend a RECONFIGURE action when a placement policy requires a VM to move to a host on a particular cluster, but cross-cluster moves are not permitted.
  - `DELETE`:

    This is equivalent to the SUSPEND action, but for storage entities.
  - `RESERVE_ON_PM`:

    When using the reservation functionality, this is the action to reserve the resources of a VM on a particular host.
  - `RESERVE_ON_DS`:

    When using the reservation functionality, this is the action to reserve the resources of a VM on a particular datastore.
  - `RESIZE_FOR_EFFICIENCY`:

    A resize that reduces the capacity of a commodity, such as vCPU, or vMEM.
  - `RESIZE_FOR_PERFORMANCE`:

    A resize that increases the capacity of a commodity, such as vCPU, or vMEM.
- cleared

  Default: `false`.

  When `true`, actions that were removed from the action list before being executed will be returned.
- costType
  - `Saving`

    An action that will decrease cloud spending.
  - `Investment`

    An action that will increase cloud spending.
- cumulative

  Default: `false`.

  When `true`, shows the savings for each action in a scope as a cumulative total, rather than showing the savings for individual actions.
- environmentType
  - `ONPREM`:

    Returns only actions related to entities that are part of your on-prem environment.
  - `CLOUD`:

    Returns only actions related to entities that are part of your cloud environment.
- groupBy

  You may group the actions returned by the following criteria: `[actionModes, actionStates, actionTypes, risk, riskSeverity, riskSubCategory]`
- hasReservedInstance

  Default: `false`.

  When `true`, only actions that are related to reserved instances will be returned.
- relatedEntityTypes

Returns only actions related to entities of these types within the scope. For example, if your scope is a virtual datacenter and you specify VirtualMachine as a related entity type, the request will include only those actions related to VMs in the specified virtual datacenter.

Choose from the following entity types: `[ Datacenter, PhysicalMachine, VirtualMachine, Storage, Application, Chassis, DiskArray, IOModule, StorageControl, Switch, VirtualDataCenter, VPod, DPod, Container, Database, DatabaseServerContainer, LogicalPool ]`

- riskSeverityList

  Returns only actions of the given severities: `[ Unknown, Normal, Minor, Major, Critical ]`

- riskSubCategoryList
- returns only actions of the given subcategories: `[ Performance Assurance, Efficiency Improvement, Prevention, Compliance ]`

Example: `POST https://10.10.10.10/api/v3/markets/Market/actions/stats`

Example input: In this example, a list of actions that meet the following criteria will be returned and grouped by the `actionMode` of the action, in order of severity:

- Is in state `PENDING_ACCEPT`
- Is part of an on-prem environment
- Is an action for a virtual machine
- Is an action generated by the Workload Optimization Manager real-time market

```
{
  "actionInput":{
    "actionStateList":[
      "PENDING_ACCEPT"
    ],
    "environmentType":"ONPREM",
    "groupBy":[
      "actionModes"
    ]
  },
  "relatedType":"VirtualMachine",
  "scopes":[
    "Market"
  ]
}
```

Response: A list of `StatSnapshotApiDTO` containing statistics about the filtered actions.

```
[
  {
    "date":"2019-02-05T14:40:08-05:00",
    "statistics":[
      {
        "name":"numActions",
        "values":{
          "max":268,
          "min":268,
          "avg":268,
          "total":268
        },
        "value":268
      },
      {
        "name":"costPrice",
        "filters":[
          {
```

```
            "type":"property",
            "value":"savings"
          }
        ],
        "units":"$",
        "values":{
          "max":24100,
          "min":24100,
          "avg":24100,
          "total":24100
        },
        "value":24100
      },
      ...
    ]
  }
]
```

Getting all entities within a market

Referenced Endpoint: Entities (on page 1195)

Gets a list of all entities in the specified market. This request may be used to retrieve a list of entities used in a plan market.

Example: `GET https://10.10.10.10/api/v3/markets/2452692215536/entities`

Response: A list of `ServiceEntityApiDTO`s, representing the entities in the market. In this case, note that there is an attribute entitled `realtimeMarketRefrence` because these are entities from a plan market.

```
[
  {
    "links":[
      {
        "rel":"self",
        "href":"https://10.10.10.10/api/v3/entities/_QQVtEyi_EemDHtuQzfdzxw?include_aspects=false"
      }
    ],
    "uuid":"_QQVtEyi_EemDHtuQzfdzxw",
    "displayName":"osp-master-2",
    "className":"VirtualMachine",
    "priceIndex":3.6063216,
    "state":"ACTIVE",
    "severity":"Normal",
    "realtimeMarketReference":{
      "uuid":"42000c5d-9ae5-44f0-252c-daabbb423a86"
    },
    "environmentType":"ONPREM"
  },
  ...
]
```

Getting statistics of entities within a market

Referenced Endpoint: Stats (on page 1372)

Referenced Endpoint: Entities (on page 1195)

Gets statistics related to entities in a specified market. For example, this request may be used to see the final calculated statistics for entities in a plan market. For more information about creating a filtered list of statistics, including example POSTs and responses, see *Filtered Statistics* in the Statistics Endpoint *(on page 1372)*.

Getting statistics of a specified group of entities within a market

Referenced Endpoint: Stats (on page 1372)

Referenced Endpoint: Groups (on page 1222)

Referenced Endpoint: Entities (on page 1195)

Gets statistics related to a specific group of entities in a specified market. For example, this request may be used to see the final calculated statistics for a group of hosts in a plan market. For more information about creating a filtered list of statistics, including example POSTs and responses, see *Filtered Statistics* in the Statistics Endpoint *(on page 1372)*.

Getting policies for a market

Referenced Endpoint: Policies (on page 1285)

To get a list of all current policies affecting a specified market, make a GET request to the `/markets/{market_uuid}/policies` endpoint. For a plan market, this is a list of policies that you created, enabled, disabled, or changed in the plan configuration. For the real-time market, this is a list of policies currently available on your appliance.

To get a single policy, include the UUID of the policy in the request.

Examples:

- All policies: `GET https://10.10.10.10/api/v3/markets/2481646239616/policies`
- One policy:

  `GET https://10.10.10.10/api/v3/markets/2481646239616/policies/_z3LKSTk0EemeycL7hWgJNw`

Response: A list of `PolicyApiDTO`s representing the policies requested.


```
[
   {
      "links":[
         {
            "rel":"self",
            "href":"https://10.10.10.10/api/v3/policies/_z3LxVzk0EemeycL7hWgJNw"
         }
      ],
      "uuid":"_z3LxVzk0EemeycL7hWgJNw",
      "displayName":"DalTestPolMerge",
      "type":"MERGE",
      "name":"DalTestPolMerge",
      "enabled":true,
      "commodityType":"ClusterCommodity",
      "mergeType":"Cluster",
      "mergeGroups":[
         {
            "uuid":"91ff08749fd278925d73b61b0f0f69d79c99fc3a",
            "displayName":"PMs_Beta DC6\\Cluster-1",
            "className":"Group",
            "groupType":"PhysicalMachine",
            "isStatic":true,
            "logicalOperator":"AND"
         },
         {
            "uuid":"8c17d8e79c3db85ed5d3868de622cb89187849ce",
            "displayName":"PMs_Beta DC6\\Cluster-2",
            "className":"Group",
            "groupType":"PhysicalMachine",
            "isStatic":true,
            "logicalOperator":"AND"
         }
      ]
```

```
   },
   {
      "links":[
         {
            "rel":"self",
            "href":"https://10.10.10.10/api/v3/policies/_z3LKSTk0EemeycL7hWgJNw"
         }
      ],
      "uuid":"_z3LKSTk0EemeycL7hWgJNw",
      "displayName":"CommSegmentation/SegmentManager__4T_7kwY-Ed-WUKbEYSVIDw_1551123454966",
      "type":"MUST_RUN_TOGETHER",
      "name":"CommSegmentation/SegmentManager__4T_7kwY-Ed-WUKbEYSVIDw_1551123454966",
      "enabled":false,
      "capacity":1000000000,
      "commodityType":"DrsSegmentationCommodity",
      "consumerGroup":{
         "uuid":"_z2M57Dk0EemeycL7hWgJNw",
         "displayName":"Buyers-Group-GROUP-DRS-KTG-rule/DC13-Cluster/vsphere-dc13.dev.mycorp.com",
         "className":"DiscoveredGroup",
         "groupType":"VirtualMachine",
         "isStatic":true,
         "logicalOperator":"AND"
      },
      "providerGroup":{
         "uuid":"_z2M57zk0EemeycL7hWgJNw",
         "displayName":"Sellers-Group-GROUP-DRS-KTG-rule/DC13-Cluster/vsphere-dc13.dev.mycorp.com",
         "className":"DiscoveredGroup",
         "groupType":"PhysicalMachine",
         "isStatic":true,
         "logicalOperator":"AND"
      }
   },
   ...
]
```


Deleting a placement policy for a market

Referenced Endpoint: Policies (on page 1285)

To delete a policy from a market, include the UUID of the policy and the market. Note that if you delete an imported policy, it reappears after the next discovery cycle of the target it was imported from. To delete imported policies, delete them on the target directly.

Example: `DELETE https://10.10.10.10/api/v3/markets/2481646239616/policies/_z3LxVzk0EemeycL7hWgJNw`

Response: A successful deletion request returns a Response Code of `200`.

Getting reservations for a market

Referenced Endpoint: Reservations (on page 1288)

Returns a list of all current reservations for a specified market. For the real-time market, this is a list of reservations currently active on your appliance.

Example: `GET https://10.10.10.10/api/v3/markets/Market/reservations`

Response: A list of `DemandReservationApiDTO`s representing any current reservations.


```
[
   {
      "uuid":"_I6z38Tk8EemeycL7hWgJNw",
```

```
      "displayName":"DalRes2",
      "count":1,
      "status":"RESERVED",
      "reserveDateTime":"Mon Feb 25 15:30:07 EST 2019",
      "expireDateTime":"Mon Mar 25 08:29:54 EDT 2019",
      "deployDateTime":"Tue Mar 26 08:29:54 EDT 2019",
      "reserveCount":1,
      "deployCount":0,
      "demandEntities":[
         {
            "uuid":"_JEZRwTk8EemeycL7hWgJNw",
            "displayName":"DalRes2_C0",
            "className":"VirtualMachine",
            "template":{
               "uuid":"T-ef47f3dbdde39f888052332cbe7d0cf2",
               "displayName":"AVG:PMs_ACM\\ACM Cluster for last 10 days",
               "className":"VirtualMachineProfile"
            },
            "placements":{
               "computeResources":[
                  {
                     "stats":[
                        {
                           "name":"numOfCpu",
                           "value":2
                        },
                        {
                           "name":"cpuSpeed",
                           "units":"MHz",
                           "value":5320
                        },
                        {
                           "name":"cpuConsumedFactor",
                           "units":"%",
                           "value":4.2516437
                        },
                        {
                           "name":"memorySize",
                           "units":"MB",
                           "value":6631
                        },
                        {
                           "name":"memoryConsumedFactor",
                           "units":"%",
                           "value":46.328953
                        },
                        {
                           "name":"ioThroughput",
                           "units":"MB/s",
                           "value":0.05301181
                        },
                        {
                           "name":"networkThroughput",
                           "units":"MB/s",
                           "value":0.103116125
                        }
                     ],
                     "provider":{
                        "uuid":"30333436-3638-5355-4532-313159335631",
                        "displayName":"hp-dl571.dev.mycorp.com",
                        "className":"PhysicalMachine"
                     }
                  }
               ],
               "storageResources":[
```

```
                {
                   "stats":[
                      {
                         "name":"diskSize",
                         "units":"GB",
                         "value":92.468285
                      },
                      {
                         "name":"diskIops",
                         "value":1.6827557
                      }
                   ],
                   "provider":{
                      "uuid":"5a822e24-584bce98-ecb2-441ea15e2534",
                      "displayName":"SCIO:ESXDC11DS2",
                      "className":"Storage"
                   }
                }
             ]
          }
       }
    ]
 }
]
```


Running a scenario on a market

Referenced Endpoint: Scenarios (on page 1298)

When you run a scenario on the real-time market, it creates a plan market showing the results of the scenario. If you run a second scenario a plan market, that is called running a *plan over plan*. This runs a plan new configuration on a market that is already the result of a plan run.

Example: `POST https://10.10.10.10/api/v3/markets/Market/scenarios/2452690315520?ignore_constraints=false`

Response: A `MarketApiDTO` showing the new state of the market- if successful, `COPYING`.


```
{
   "links":[
      {
         "rel":"self",
         "href":"https://10.10.10.10/api/v3/markets/_ps_iUDlBEemeycL7hWgJNw"
      },
      {
         "rel":"Market notifications",
         "href":"https://10.10.10.10/api/v3/markets/_ps_iUDlBEemeycL7hWgJNw/notifications"
      },
      {
         "rel":"Market actions",
         "href":"https://10.10.10.10/api/v3/markets/_ps_iUDlBEemeycL7hWgJNw/actions?ascending=true"
      },
      {
         "rel":"To Entities",
         "href":"https://10.10.10.10/api/v3/markets/_ps_iUDlBEemeycL7hWgJNw/entities"
      },
      {
         "rel":"To Constraints",
         "href":"https://10.10.10.10/api/v3/markets/_ps_iUDlBEemeycL7hWgJNw/policies"
      },
      {
         "rel":"To Reservations",
         "href":"https://10.10.10.10/api/v3/markets/_ps_iUDlBEemeycL7hWgJNw/reservations"
```

```
      },
      {
         "rel":"To Statistics",
         "href":"https://10.10.10.10/api/v3/markets/_ps_iUDlBEemeycL7hWgJNw/stats"
      },
      {
         "rel":"To Unplaced Entities",
         "href":"https://10.10.10.10/api/v3/markets/_ps_iUDlBEemeycL7hWgJNw/unplacedentities"
      },
      {
         "rel":"To Market List",
         "href":"https://10.10.10.10/api/v3/markets"
      }
   ],
   "uuid":"_ps_iUDlBEemeycL7hWgJNw",
   "displayName":"API_Plan_Market",
   "className":"Market",
   "state":"COPYING",
   "scenario":{
      "uuid":"2452690315520",
      "displayName":"Add Workload 1",
      "owners":[
      ],
      "type":"ADD_WORKLOAD"
   },
   "unplacedEntities":false,
   "runDate":"2019-02-25T16:09:34-05:00",
   "environmentType":"ONPREM"
}
```

Getting statistics for a market

Referenced Endpoint: Stats (on page 1372)

Gets statistics about the specified market. This call can be used to see statistics from the real-time market, or to see the resultant statistics in a plan market after the scenario has been applied.

Example: `GET https://10.10.10.10/api/v3/markets/Market/stats`

Response: A list of `StatSnapshotApiDTO`s representing the statistics of the real-time market.

```
[
   {
      "displayName":"PMs",
      "date":"2019-02-23T00:00:00-05:00",
      "statistics":[
         {
            "name":"CPUHeadroom",
            "capacity":{
               "max":8610,
               "min":0,
               "avg":1602.22,
               "total":14420
            },
            "units":"VM",
            "values":{
               "max":8595,
               "min":0,
               "avg":1571.44,
               "total":14143
            },
            "value":1571.44
```

```
        },
        {
           "name":"CPUExhaustion",
           "units":"Day",
           "values":{
              "max":2147483650,
              "min":0,
              "avg":2147483650,
              "total":2147483650
           },
           "value":2147483650
        },
        ...
    ...
]
```

Getting a filtered list of statistics for a market

Referenced Endpoint: Stats (on page 1372)

Gets a list of statistics about the specified market, filtered by the criteria in the StatPeriodApiInputDTO you must pass as an inputDTO to this request. For more information about creating a filtered list of statistics, including example POSTs and responses, see *Filtered Statistics* in the Statistics Endpoint *(on page 1372)*.

Example:

Example Input:

Response: An array of `StatSnapshotApiDTO`s representing the filtered statistics for the market.

Getting the supply chain for a market

Referenced Endpoint: Supply Chains (on page 1389)

Using one or more of the following parameters, you can filter the supply chain that is returned by this request:

- market_Uuid

  The UUID of the market whose supply chain will be returned
- types

  The entity types whose supply chain will be returned. Choose from:
  - `APPLICATION`
  - `APPLICATION_SERVER`
  - `BUSINESS_ACCOUNT`
  - `BUSINESS_UNIT`
  - `CHASSIS`
  - `CLOUD_SERVICE`
  - `CONTAINER`
  - `CONTAINER_POD`
  - `DATABASE`
  - `DATABASE_SERVER`

  - DATACENTER
  - DISK_ARRAY
  - IO_MODULE
  - LOAD_BALANCER
  - LOGICAL_POOL
  - NETWORK
  - PHYSICAL_MACHINE
  - STORAGE
  - STORAGE_CONTROLLER
  - SWITCH
  - VIRTUAL_APPLICATION
  - VIRTUAL_DATACENTER
  - VIRTUAL_MACHINE
- entity_states

  Only return entities in the specified states. Choose from:
  - ACTIVE:

    Entity is actively consuming resources.
  - EVACUATED:

    Entity is ready to suspend and will not participate in the market, but will not generate a SUSPEND action.

    NOTE: This state replaces the SUSPEND state, and requires additional configuration to enable. Workload Optimization Manager does not recommend enabling this state.
  - FAILOVER:

    This entity is reserved for failover and will not participate in the market.
  - IDLE:

    Entity is powered off and not actively consuming market resources.
  - LAUNCH:

    Entity is starting.
  - MAINTENANCE:

    Entity is in maintenance mode.
  - NOT_MONITORED:

    Entity is not currently monitored by Workload Optimization Manager.
  - QUEUED:

    Entity is in the process of changing from one state to another.

    NOTE: This state is internal to Workload Optimization Manager and will not appear in the UI, but may appear in an API response if the call is executed at the precise moment the entity is changing state.
  - RESOURCE_ALLOCATION:

    Cloud entity is being created.
  - RESOURCE_RELEASE:

    Cloud Entity is being deleted.
  - SUSPEND:

Entity is suspended or the guest OS is in sleep, standby, or suspended state. The entity is not removed from the environment, but will not be used in market decisions. Based on market conditions, suspended entities can be started, re-adding them to the market.

  - `SUSPEND_PENDING`:

    Entity is making preparations to suspend, or is the next entity of its type to suspend.

  - `TERMINATE_PENDING`:

    Entity is in the process of being deleted.

  - `UNKNOWN`:

    Workload Optimization Manager is unable to obtain the current state of the entity.

    This occurs when the entity is reported to the hypervisor as having a state of `DISCONNECTED, ORPHANED,` or `INVALID`.

- environment_type
- Only return actions that affect entities in the specified environment type. `[ CLOUD, HYBRID, ONPREM ]`.
- detail_type

  The level of detail returned for each item in the returned supply chain. `[ compact, entity, aspects ]`.
- health

  Return a small summary of the health of the returned supply chain.

Example: `GET https://10.10.10.10/api/v3/markets/Market/supplychains?types=VirtualMachine&entity_states=RESOURCE_ALLOCATION&entity_states=ACTIVE&environment_type=ONPREM&detail_type=aspec`

Response: A `SupplychainApiDTO` representing the entities meeting your criteria. In the previous example, a supply chain including Virtual Machines in the ACTIVE or RESOURCE_ALLOCATION states, in an ONPREM environment, and a health check will be included at the end of the request.


```
{
   "seMap":{
      "VirtualMachine":{
         "depth":-1,
         "entitiesCount":335,
         "healthSummary":{
            "Major":6,
            "Minor":148,
            "Normal":181
         },
         "stateSummary":{
            "ACTIVE":335
         },
         "connectedProviderTypes":[
            "PhysicalMachine",
            "Storage",
            "VirtualDataCenter"
         ],
         "connectedConsumerTypes":[
            "Application"
         ],
         "instances":{
            "421edc7a-761d-9a79-9378-7140caecd50f":{
               "links":[
                  {
                     "rel":"self",

"href":"https://10.10.10.10/api/v3/entities/421edc7a-761d-9a79-9378-7140caecd50f?include_aspects=false"
                  }
               ],
               "uuid":"421edc7a-761d-9a79-9378-7140caecd50f",
               "displayName":"vm-max-test-netapp",
```

```
               "className":"VirtualMachine",
               "priceIndex":1.0070138,
               "state":"ACTIVE",
               "severity":"Normal",
               "discoveredBy":{
                  "uuid":"_8c98wBwnEem2J-ksdl1CgA",
                  "displayName":"vsphere-dc7.dev.mycorp.com",
                  "type":"vCenter"
               },
               "remoteId":"vm-281",
               "aspects":{
                  "virtualMachineAspect":{
                     "os":"SUSE Linux Enterprise 11 (64-bit)",
                     "connectedNetworks":[
                        {
                           "uuid":"7d6d67a50cfe1979797de159220841a37a55a50c",
                           "displayName":"VM network"
                        }
                     ],
                     "numVCPUs":1,
                     "ebsOptimized":false
                  }
               },
               "environmentType":"ONPREM"
            }           ...
         ]
      ...
   ...
...
}
```

Markets endpoint tips

Unplaced entities

When running a plan in Workload Optimization Manager, you can add workload or remove infrastructure – if the remaining resources are insufficient for the workload, entities may be "unplaced". Workload Optimization Manager provides an array of the unplaced workloads.

NOTE: An example will be forthcoming.

Finding a previously run plan market

While the Workload Optimization Manager UI makes a plan look like one item, it is not. A plan comprises two parts:

- A scenario, which represents the configuration options for the plan For more information about plan configuration, see Scenarios *(on page 1298)*.
- The plan market that is created when a scenario is run – this contains the actions, notifications, etc. that make up the state of the market after the plan.

The simplest way to discover a plan market is to execute the `GET https://10.10.10.10/api/v3/markets` and disregard the main market. The main market DTO entry will be similar to the following:

```
{
  "uuid":"777777",
  "displayName":"Market",
  "className":"Market",
  "state":"RUNNING",
  "stateProgress":100,
```

```
  "unplacedEntities":false,
  "environmentType":"HYBRID"
}
```

The other markets that the request returns are plan markets. Plan market DTOs are similar to:

```
{
  "uuid":"214172431889696",
  "displayName":"Add Workload 2",
  "className":"Market",
  "state":"SUCCEEDED",
  "stateProgress":100,
  "scenario":{
    "uuid":"214172431885856",
    "displayName":"Add Workload 2",
    "owners":[
      {
        "uuid":"2993567556912",
        "showSharedUserSC":false
      }
    ]
    ...
}
```

The outer-most `uuid` and `displayName` parameters are the UUID and display name of plan market. The `state` field shows the results of the plan run, where SUCCEEDED means the plan finished successfully, and `FAILED` means the plan did not finish running- usually because it was canceled by the user.

Within the plan itself, the `scenario` parameter includes information about the scenario used on the market. Note the inner `uuid` and `displayName` are the UUID and display name of the scenario, which you can use to see all markets created using this scenario.

Finally, iterating over the return of the previous `/markets` call, and looking at each market and comparing the UUID of the scenario (in this case, `214172431889696`), you can see how many times the scenario was run, and also each market containing the results of that run.

Filtering actions by time

When you make a GET request to the `/actions` endpoint, you can filter the list of returned actions based on a specified start time, end time, or both.

- The `startTime` parameter excludes any actions that were executed (succeeded or failed) before a specified time.
- The `endTime` parameter excludes any actions with a `createTime` value that is after a specified time.

When you define both these parameters, the response includes all current and historical actions that were pending or executed during the specified time period.

For example, the following GET request to the `markets/Market/actions` endpoint returns all actions that were created, pending, or executed between 6:00 and 11:00 on May 8, 2024.

Example: `GET https://10.10.10.10/api/v3/markets/Market/actions`

Example input:

```
{ "startTime": "2024-05-08T06:00:19Z",
  "endTime": "2024-05-08T011:00:19Z",
  ...
}
```

Although the response does not include any actions that were created after the `endTime`, it might include actions that were created before the `startTime`, if they are still pending as of the `startTime`. Therefore, this filtering includes all actions that are either pending or executed during the specified time period, regardless of when those actions were created.

Notifications endpoint

Notifications are system and market messages from your Workload Optimization Manager instance.

This endpoint contains references to start and end times for certain operations. For more information, including valid time formats, see Time in the Workload Optimization Manager API *(on page 1141)*

Using the notifications endpoint, you can:

- Get all notifications
- Get a filtered list of notifications
- Get related actions for a specified notification

Notification categories

Each notification from your Workload Optimization Manager instance belongs to a particular category of the following:

- Discovery
- Monitoring
- Control
- Mediation
- Healthcheck
- InterAppliance
- MarketAction
- MarketProblem
- Presentation
- Deploy
- Security

Notifications requests

Getting notifications

Gets a list of system and market level notifications. For a list of notifications, the API returns an array of LogEntryApiDTOs. Each LogEntryApiDTO will contain details about the notification, including the time, state, severity, and importance of the notification. To get a single notification, include the UUID of the notification in the request.

This request takes the following parameters:

- start_time

  The start time for a historic range of notifications.
- end_time

  The end time for a historic range of notifications.
- category

  The category of notifications to return. `[Discovery, Monitoring, Control, Mediation, Healthcheck, InterAppliance, MarketAction, MarketProblem, Presentation, Deploy, Security]`
- cursor

When making subsequent calls to retrieve additional results, this is the cursor returned by the last call to this method.

- limit

  The maximum number of items to return. Must be a positive integer. If not set, the server will provide a default (20).

- order_by

  The field to order the results by. This value can be any field found in the response.

- ascending

  Default: true. When false, results will be in descending order.

Examples:

- **All Notifications (no parameters):** `GET https://10.10.10.10/api/v3/notifications`
- **All Notifications (parameters):** `GET https://10.10.10.10/api/v3/notifications?start_time=1237056032&end_time=1238056032&limit=100&order_by=name&ascending=true`
- **Single Notification:** `GET https://10.10.10.10/api/v3/notifications/_5wcf4J5WEeinZ-6Uu4soUw`

Response:


```
[
  {
    "links":[
      {
        "rel":"self",
        "href":"https://10.10.10.10/api/v3/notifications/_5wcf4J5WEeinZ-6Uu4soUw"
      }
    ],
    "uuid":"_5wcf4J5WEeinZ-6Uu4soUw",
    "logActionTime":1534095671518,
    "targetSE":"Hyper-V:hp-cluster.corp.mycorp.com",
    "category":"Discovery",
    "subCategory":"Validation",
    "description":"Validation of target hp-cluster.corp.mycorp.com failed",
    "shortDescription":"Validation Failed",
    "state":"NOTIFY",
    "severity":"MAJOR",
    "importance":0
  },
  {
    "links":[
      {
        "rel":"self",
        "href":"https://10.10.10.10/api/v3/notifications/_sQ7xUJ5XEeinZ-6Uu4soUw"
      }
    ],
    "uuid":"_sQ7xUJ5XEeinZ-6Uu4soUw",
    "logActionTime":1534096010470,
    "targetSE":"Hyper-V:hp-cluster.corp.mycorp.com",
    "category":"Discovery",
    "subCategory":"Discovery",
    "description":"Could not get domain for the 'hp-cluster.corp.mycorp.com' target",
    "shortDescription":"Discovery failed",
    "state":"NOTIFY",
    "severity":"MAJOR",
    "importance":0
  },
  ...
]
```

Parking endpoint

Workload Optimization Manager supports 'parking' actions for cloud resources. These user-initiated actions *stop* your cloud resources for a given period of time to help you reduce your cloud expenses, and then *start* these resources later when you need them. You can enforce parking actions on demand, or automatically through parking schedules and policies.

Using the Parking endpoint of the Workload Optimization Manager API, you can:

- Get a list of parkable entities
- Enforce parking actions on demand

To enforce parking actions according to a schedule, use `/schedules/timespans/` requests in the `Schedules` endpoint. With these requests, you can configure parking schedules and then attach the schedules to parkable entities. For details, see Parking Schedules Requests *(on page 1345)*.

When parking schedules have been attached to entities, you can use the Parking endpoint to pause and resume schedules as needed.

Parking requests

Pagination for parking requests

When you request parking entities, the API uses pagination to return subsets of the data. You should always use pagination when available to make sure you get your data in manageable chunks. To paginate returned data, the `limit` and `cursor` parameters control how the APIs paginate their data.

Example: `POST https://10.10.10.10/api/v3/parking/entities?cursor=10&limit=5&order_by=DISPLAY_NAME`

This example returns the 11th through 16th entities ordered by display name.

For general information about pagination, see Pagination in the Workload Optimization Manager API *(on page 1141)*.

Getting parkable entities

Provides a list of entities that can be stopped or started.

Example: `POST https://10.10.10.10/api/v3/parking/entities`

Results may be sorted by one of the following: `DISPLAY_NAME, TYPE, STATE, ACCOUNT_NAME, PROVIDER, REGION_NAME, INSTANCE_TYPE` or `COST`. The default sort order is `ascending` but it may be reversed by including `ascending=false` in the query string.

Example: `POST https://10.10.10.10/api/v3/parking/entities?order_by=DISPLAY_NAME&ascending=true`

Getting a filtered list of parkable entities

The simplest form of filtering is to provide one of the following main-level keys:

Field	Possible Values	Example
`providers`	`AWS, AZURE, GCP`	`{"providers": ["AWS", "GCP"]}`
`status`	`STOPPED, RUNNING, STARTING, STOPPING`	`{"status": ["RUNNING", "STARTING"]}`

Field	Possible Values	Example
entityTypes	VirtualMachine, DatabaseServer **NOTE:** DatabaseServer only applies to AWS RDS instances.	{"entityTypes": ["VirtualMachine"]}

Example: POST https://10.10.10.10/api/v3/parking/entities

Example input that fetches all running VMs in all of your AWS targets:

```
{"entityTypes": ["VirtualMachine"], "status":["RUNNING"], "providers": ["AWS"]}
```

Example: POST https://10.10.10.10/api/v3/parking/entities

Example input, showing the same query with the addition of the field logicalOperator. The logicalOperator value can either be AND or OR, and will control whether the supplied filters are all required or are optional. For example, this request now fetches all entities that are a virtual machine, or are running, or are provided by an AWS target:

```
{"entityTypes": ["VirtualMachine"], "status": "RUNNING"], "providers": ["AWS"], "logicalOperator": "OR"}
```

Advanced Filtering: More powerful filtering is available by using criteriaList. This object supports additional fields and provides more control over your filter expressions, including regular expression filtering. Each element in a criteriaList contains the following keys: expVal, expType, and filterType.

filterType	expType	Constraints
displayName	EQ, NEQ, RXEQ, RXNEQ	
accountName	EQ, NEQ, RXEQ, RXNEQ	
regionName	EQ, NEQ, RXEQ, RXNEQ	
instanceType	EQ, NEQ, RXEQ, RXNEQ	
scheduleName	EQ, NEQ, RXEQ, RXNEQ	
tags	EQ, NEQ, RXEQ, RXNEQ	Discovered tags containing pipes | may produce unexpected results when using an expType of RXEQ or RXNEQ
cost	EQ, NEQ, GT, LT, GTE, LTE	

Example: POST https://10.10.10.10/api/v3/parking/entities

Example input that fetches all entities within the "Engineering" account:

```
{
  "criteriaList": [
    {
      "expVal": "Engineering",
      "expType": "EQ",
      "filterType": "accountName"
    }
  ]
}
```

Advanced filtering may also take advantage of logicalOperator, if desired.

Example: `POST https://10.10.10.10/api/v3/parking/entities`

Example input that returns all entities in either the "Engineering" or "Test" accounts:

```
{
  "criteriaList": [
    {
      "expVal": "Engineering",
      "expType": "EQ",
      "filterType": "accountName"
    },
    {
      "expVal": "Test",
      "expType": "EQ",
      "filterType": "accountName"
    }
  ]
  "logicalOperator": "OR"
}
```

Executing parking actions

You may start or stop parkable entities on demand by submitting a manual action request. Manual actions are executed immediately.

> **NOTE:**
> Manual actions will not be accepted if the entities are controlled by a schedule. For these entities, you must first pause schedule enforcement.

Example: `POST https://10.10.10.10/api/v3/parking/bulkActionRequest`

Example input that generates actions to start the provided entities:

```
{"action":"START","entityUuids":["74913973295604"]}
```

Constraints: `action` can either be START or STOP.

Response: This API responds with one element for each `entityUuid` provided and indicates the resulting `state`.

```
[{"state":"STARTING","entityUUID":"74913973295604"}]
```

If an action could not be generated for a particular `entityUuid`, the response will contain an error message:

```
[{"error":"unable to toggle the entity","entityUUID":"74913973295605"}]
```

If an action is requested for multiple entities, a response may contain a partial success. Check the response body to understand whether the action succeeded for a particular entity.

```
[{"state":"STARTING","entityUUID":"74913973295604"}, {"error":"unable to toggle the
 entity","entityUUID":"74913973295605"}]
```

Pausing and resuming parking schedules

Parking schedules must first be configured and attached to entities using `/schedules/timespans` requests. For details, see Parking Schedules Requests *(on page 1345)*.

After attaching schedules to entities, you can pause or resume schedules using the Parking endpoint.

Pausing Parking Schedules: There are times when pausing schedule enforcement is useful. The most common use is when you need to pause an environment during non-typical work hours or when an unanticipated maintenance window requires automation outside Workload Optimization Manager to start and stop workloads. Like attaching and detaching schedules *(on page 1350)*, you may pause or resume 1-N entities from schedule enforcement.

Example: `POST https://10.10.10.10/api/v3/parking/enforcementPauses`

Example input that pauses schedule enforcement for 1-N entities using a duration in minutes (only supported unit):

```
{"entityUuids":["74913973295604"],"durationMinutes":15}
```

Example input that pauses schedule enforcement for 1-N entities until 12:26 PM Central (UTC-5) in ISO 8601 date and time format:

```
{"entityUuids":["74913973295604"],"until":"2023-05-16T14:26:00-05:00"}
```

Resuming Parking Schedules Example: `DELETE https://10.10.10.10/api/v3/parking/enforcementPauses`

Example input that resumes schedule enforcement for 1-N entities:

```
{"entityUuids":["74913973295604"]}
```

Policies endpoint

Policies set business rules to control how Workload Optimization Manager analyzes resource allocation, how it displays resource status, and how it recommends or executes actions. Workload Optimization Manager includes the following fundamental types of policies:

- Placement Policies

  To modify workload placement decisions, Workload Optimization Manager divides its market into segments that constrain the valid placement of workloads. Workload Optimization Manager discovers placement rules that are defined by the targets in your environment, and you can create your own segments.

- Automation Policies

  Workload Optimization Manager uses Automation Policies to guide its analysis and resulting actions. These policies can specify Automation Workflows and Constraints (operational, utilization, and scaling constraints).

- Parking Policies

  A parking policy is a set of rules that Workload Optimization Manager must satisfy when enforcing parking actions (stop and start) on public cloud entities.

Workload Optimization Manager ships with default settings that we believe will give you the best results from our analysis and control. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation for that scope, or change the utilization constraints.

The hypervisors that you set as targets can include placement policies of their own. Workload Optimization Manager imports these placement policies, and you can choose to enable or disable them as you wish. By default, Workload Optimization Manager enables imported placement policies.

Using the policies endpoint, you can:

- Get API entry points for placement policies
- See details about a specific placement policy

The policies endpoint does not handle automation policies or parking policies.

For more information about managing policies with the API, see Working with policies in the API *(on page 1148)*.

Policies utility requests

Utility requests provide metadata that you can use to make related requests, assemble inputDTOs, or see the potential values of a particular DTO.

The policies endpoint contains a utility request that shows the entry points for policy-related requests. These can be used to find the UUID and details relating to a specific policy.

Example: `GET https://10.10.10.10/api/v3/policies`

Response: An array of Link objects, where each object contains a URL to get all policies for a specified market, entity, or group.

```
{
  "links":[
    {
      "rel":"Market policies",
      "href":"https://10.10.10.10/api/v3/markets/{uuid}/policies"
    },
    {
      "rel":"Entity policies",
      "href":"https://10.10.10.10/api/v3/entities/{uuid}/policies"
    },
    {
      "rel":"Group policies",
      "href":"https://10.10.10.10/api/v3/groups/{uuid}/policies"
    }
  ],
  "enabled":false
}
```

Policies requests

Getting a list of urls for policy-related requests

The policies endpoint contains a utility request that shows the entry points for policy-related requests. These can be used to find the UUID and details relating to a specific policy.

Utility requests provide metadata that you can use to make related requests, assemble inputDTOs, or see the potential values of a particular DTO.

Example: `GET https://10.10.10.10/api/v3/policies`

Response: An array of Link objects, where each object contains a URL to get all policies for a specified market, entity, or group.

```
{
  "links":[
    {
      "rel":"Market policies",
      "href":"https://10.10.10.10/api/v3/markets/{uuid}/policies"
    },
    {
      "rel":"Entity policies",
      "href":"https://10.10.10.10/api/v3/entities/{uuid}/policies"
    },
    {
      "rel":"Group policies",
      "href":"https://10.10.10.10/api/v3/groups/{uuid}/policies"
    }
  ],
  "enabled":false
```

```
}
```

Getting the details of a specific policy

The single policy request gets the details of a specified policy, given that policy's UUID. After you get the UUIDs for policies that affect a certain scope, you can use this request to view details of those policies.

Example: `GET https://10.10.10.10/api/v3/policies/104:domain-c7:10.10.111.111`

Response: A `PolicyApiDTO` describing details about the policy, including the group information for the provider of the policy.

```
{
  "links":[
    {
      "rel":"self",
      "href":"https://10.10.10.10/api/v3/policies/104:domain-c7:10.10.111.111"
    }
  ],
  "uuid":"104:domain-c7:10.10.111.111",
  "displayName":"GROUP-DRS-LN-rule/Cluster1/10.10.111.111",
  "type":"BIND_TO_GROUP",
  "name":"GROUP-DRS-LN-rule/Cluster1/10.10.111.111",
  "enabled":true,
  "capacity":1000000000,
  "commodityType":"DrsSegmentationCommodity",
  "consumerGroup":{
    "uuid":"7843e134bab4fb841b9cb7dd98ef04acf1f36f7b",
    "displayName":"Buyers-Group-GROUP-DRS-vmGroup1/Cluster1/10.10.111.111",
    "className":"DiscoveredGroup",
    "groupType":"VirtualMachine",
    "isStatic":true,
    "logicalOperator":"AND"
  },
  "providerGroup":{
    "uuid":"597a35089bfe405291911f3191cf24234d4713d1",
    "displayName":"Sellers-Group-GROUP-DRS-hostGrp1/Cluster1/10.10.111.111",
    "className":"DiscoveredGroup",
    "groupType":"PhysicalMachine",
    "isStatic":true,
    "logicalOperator":"AND"
  }
}
```

Creating a policy

Policies are created for a specific market- either a plan market, or the real-time market. For more information about policy creation, see the Markets *(on page 1258)* endpoint.

Price lists utility requests

Utility requests provide metadata that you can use to make related requests, assemble inputDTOs, or see the potential values of a particular DTO.

The price list endpoint contains a sample rate card file and RateCardApiDTO. To access this sample:

1. Execute the `GET https://10.10.10.10/api/v3/pricelists` request, and locate the price list whose `displayName` begins with `Sample Azure Rate Card`.
2. Record the UUID of this rate card.
3. Execute the `https://10.10.10.10/api/v3/pricelists/{uuid_from_step_2}/export` request, and download the file that makes up the response body.

Price lists requests

The price lists endpoint is not implemented at this time.

Reservations endpoint

Workload Optimization Manager uses reservations to save the resources you will need if you want to deploy workloads at a future date. Workload Optimization Manager calculates optimal placement for these workloads, and then it reserves whatever resources the different hosts and storage entities will need to support those workloads. For example, you can instruct Workload Optimization Manager on 01 January to reserve the resources required for 100 workloads on 07 January.

Using the reservations endpoint of the Workload Optimization Manager API, you can:

- Get a list of reservations
- Create a reservation
- Get information about a specific reservation
- Delete a reservation

Reservation states

Reservations will be in one of the following states:

- `FUTURE`

  The reservation is set to occur after the current date.
- `INITIAL`

  The reservation is being created; occurs at the start time of the reservation.
- `INPROGRESS`

  The reservation request has been sent to the market provider(s), and is awaiting results.
- `INVALID`

  This error state occurs when a reservation is placed on an unready market (for example, when the market component is restarting) or one of the constraints or templates used in the reservation has been deleted between the creation of the reservation, and the start date.
- `PLACEMENT_FAILED`

  The reservation has completed, but there were insufficient resources to place all entites.
- `RESERVED`

  The reservation has completed and was able to obtain all needed resources.
- `UNFULFILLED`

  The reservation has been added to the database, but has not yet been sent to the market.

Reservations in the product

From the Workload Placement Page, you can set up reservations to save the resources you will need to deploy workloads at a future date. Workload Optimization Manager uses its intelligent workload management to calculate optimal placement for these workloads, and then it reserves whatever resources the different hosts and storage entities will need to support those workloads.

When creating a reservation in the product, you will select the VM template, any constraints (scope, policy, network), then give your reservation a name, virtual machine count, and start and end dates.

Reservations requests

Getting reservations

For a list of reservations, the API returns a list of DemandReservationApiDTOs. Each DemandReservationApiDTO will contain the entity information for the reserved instances, the count to be deployed, and current status. To get the DTO of a particular reservation, include the UUID in the request.

This request takes the following parameters:

Status

Returns only reservations in the specified status. `[PLACEMENT_FAILED, UNFULFILLED, RESERVED, INPROGRESS, INITIAL, FUTURE, INVALID]`

Examples:

- All reservations with a status of RESERVED:

  ```
  GET https://10.10.10.10/api/v3/reservations?status=RESERVED
  ```

- A single reservation:

  ```
  GET https://10.10.10.10/api/v3/reservations/_TszrkcwHEei7A9DOcc4f1g
  ```

Response:


```
[
  {
    "uuid": "215446856141616",
    "displayName": "MyReservation",
    "count": 1,
    "status": "RESERVED",
    "reserveDateTime": "2022-11-01T13:28:00Z",
    "expireDateTime": "2022-12-01T14:28:00Z",
    "constraintInfos": [
      {
        "uuid": "74635127126226",
        "constraintType": "DATA_CENTER"
      }
    ],
    "reservationDeployed": false,
    "mode": "NO_GROUPING",
    "grouping": "NONE",
    "demandEntities": [
      {
        "template": {
          "uuid": "215364307544240",
          "displayName": "headroomVM",
          "className": "VirtualMachineProfile"
        },
        "placements": {
          "computeResources": [
```

```
            {
              "stats": [ ...
               Stats for CPU, CPUProvisioned, Mem, MemProvisioned, IOThroughput, NetThroughput
              ... ],
              "provider": {
                "uuid": "74635127126242",
                "displayName": "hp-dl3111.eng.vmturbo.com",
                "className": "PhysicalMachine"
              },
              "linkedResources": [
                {
                  "uuid": "285741371272387",
                  "displayName": "DC13/Physical Hosts_Cluster2 - Intel"
                }
              ]
            }
          ],
          "storageResources": [
            {
              "stats": [ ...
               Stats for StorageProvisioned, StorageAmount, StorageAccess
              ... ],
              "provider": {
                "uuid": "74635124839222",
                "displayName": "QSGRID01:DC1314Shared",
                "className": "Storage"
              }
            }
          ]
        }
      }
    ]
  }
]
```

Creating a placement reservation

To create a placement reservation, POST a completed DemandReservationApiInputDTO as input to this request. The DTO specifies reservation type, the template to use, the UUIDs of constraints such as placement or network policies, and other attributes of the reservation.

Example: `POST https://10.10.10.10/api/v3/reservations`

Example inputDTO, to execute a reservation for two instances of a template:

```
{
  "demandName": "MyReservation1",
  "action": "RESERVATION",
  "mode": "NO_GROUPING",
  "grouping": "NONE",
  "reserveDateTime": "2022-11-01T21:03:22Z",
  "expireDateTime": "2022-12-01T21:03:22Z",
  "parameters": [
    {
      "placementParameters": {
        "count": 2,
        "templateID": "215364307544240",
        "entityNames": [
          "MyRes1_VM1",
          "MyRes1_VM2"
        ],
        "constraintIDs": [
          "285815619396128"
```

```
        ]
      }
    }
  ]
}
```

Response: A `DemandReservationApiDTO` that will include the status of the placement attempt. Note that the status of this reservation is FUTURE. When Workload Optimization Manager executes the reservation, if it succeeds in placing the reserved VMs then the `placements` field will contain descriptions of the host and storage providers for the two VMs.

```
{
  "uuid": "215447101807136",
  "displayName": "MyReservation1",
  "count": 2,
  "status": "FUTURE",
  "reserveDateTime": "2022-11-01T21:03:22Z",
  "expireDateTime": "2022-12-01T21:03:22Z",
  "constraintInfos": [
    {
      "uuid": "285815619396128",
      "constraintType": "POLICY"
    }
  ],
  "reservationDeployed": false,
  "mode": "NO_GROUPING",
  "grouping": "NONE",
  "demandEntities": [
    {
      "template": {
        "uuid": "215364307544240",
        "displayName": "headroomVM",
        "className": "VirtualMachineProfile"
      },
      "placements": {}
    },
    {
      "template": {
        "uuid": "215364307544240",
        "displayName": "headroomVM",
        "className": "VirtualMachineProfile"
      },
      "placements": {}
    }
  ]
}
```

Deleting a reservation

Deletes the specified reservation and removes all reserved resources associated with that reservation. This will also remove any pending deployment related to the deleted reservation.

Example: `DELETE https://10.10.10.10/api/v3/reservations/_i-FLwdLiEeiuY7KDiLoLWg`

Response: A successful deletion request returns a Response Code of `200`.

Reservations endpoint tips

Customizing a reservation

When creating a reservation, you should be familiar with placement policies *(on page 1285)*, templates *(on page 1406)*, scope constraints, and network constraints.

> **NOTE:**
> Use network constraint to limit workload placement to the networks that Workload Optimization Manager discovered in your environment.

This request takes the following parameters of the `DemandReservationApiInputDTO`:

demandName

The display name of the reservation – If you do not specify names for the reserved workloads, they will all get this as the root name, with a numeric suffix. For example, MyReservation1, MyReservation2, etc.

action

The action type: This is always `RESERVATION`.

reserveDateTime

The time to calculate the workload placement and create the reservation – This time cannot be earlier than the time that you POST the reservation to the API.

expireDateTime

When to cancel the reservation. At this time, Workload Optimization Manager will delete the reservation. To take advantage of the reserved resources, you should deploy the actual VMs to their reserved locations before this time.

parameters

A `DemandReservationParametersDTO`, which is itself composed of a `PlacementParametersDTO` and `DeploymentParametersDTO`.

In order to pass a completed `DemandReservationParametersDTO`, you will create a `PlacementParametersDTO` with the following parameters:

count

The number of workloads in this reservation

templateID

The UUID of the template used for creating the workload required for the placement

entityNames

Name of the placed workload, which will be affixed with a numeric suffix for reservations of more than one workload

constraintIDs

The UUIDs of the scope, group and any policies or network constraints that the reservation will adhere to

For example, the following POST creates a reservation of 10 VMs, scoped to a specific virtual data center with a UUID of 73365042986774 and network with a UUID of 73365048169751. Because it does not specify `entityNames` (an array of names for the VMs), Workload Optimization Manager will generate the VM names.

```
{
  "action":"RESERVATION",
  "demandName":"DalReservation",
  "reserveDateTime":"2020-03-11T15:44:41Z",
  "expireDateTime":"2020-04-11T15:44:41Z",
  "parameters":[
    {
      "placementParameters":{
```

```
        "count":"10",
        "templateID":"214103358655713",
        "constraintIDs":[
          "73365042986774",
          "73365048169751"
        ]
      }
    }
  ]
}
```

Affinity rules

When creating reservations, you can specify affinity rules for the reserved VMs, to ensure all the VMs in a reservation deploy to the same cluster. If you do not specify affinity, the VMs deploy to any clusters in the reservation's scope. These settings are optional, but if you choose to specify affinity, you must specify compatible values for both settings.

These setting respect other constraints you might have set for the reservation. For example, if you scope the reservation to a group of clusters, with a cluster affinity rule Workload Optimization Manager will choose one cluster from that scope and try to place all the VMs on it.

To specify affinity for a reservation, you declare:

- `mode` (optional):

  Can be one of:

  - `AFFINITY` - Group reserved VMs in the same cluster. The `grouping` parameter must be set to `CLUSTER`. If there is no cluster that can host the requested VMs, the reservation will be `UNFULLFILLED`.
  - `NO_GROUPING` - Do not group reserved VMs in the same cluster. The `grouping` parameter must be set to `NONE`.
- `grouping` (optional):

  Can be one of:

  - `CLUSTER` - Group reserved VMs. The `mode` parameter must be set to `CLUSTER`.
  - `NONE` - Do not group reserved VMs. The `mode` parameter must be set to `NO_GROUPING`.

```
{
  "demandName": "MyReservation1",
  "action": "RESERVATION",
  "mode": "AFFINITY",
  "grouping": "CLUSTER",
  ...
}
```

Getting reservation information

To get a list of current reservations, execute the `GET https://10.10.10.10/api/v3/reservations` request. This returns a list of all active reservations. If you know the UUID of the reservation you want to retrieve, you can pass it to get data for just that reservation.

Each reservation object describes:

- Display name
- Status – can be:
  - `IN_PROGRESS` – Placement calculation is in progress
  - `PLACEMENT_SUCCEEDED` – For a new reservation, the environment has sufficient resources to place the workload; if you specified a reserve date, this will be an active reservation

  - PLACEMENT_FAILED – For a newly created reservation, the environment doesn't have resources to place the workload; if you specified a reserve date, this will be an unfulfilled reservation
  - RETRYING – Workload Optimization Manager is trying to place the workload of an unfulfilled reservation
- Time the reservation was created, and time it will expire
- A description of the reserved workload
- Statistics for the compute and storage resources the reservation sets aside

For example, this listing shows a successful reservation for one VM:


```
{
  "uuid":"_kWZHIDDmEeePgeXuo0RRbw",
  "displayName":"MyReservation",
  "count":1,
  "status":"PLACEMENT_SUCCEEDED",
  "reserveDateTime":"Thu May 04 16:27:29 UTC 2017",
  "expireDateTime":"Thu Aug 31 16:27:29 UTC 2017",
  "deployDateTime":"Thu Aug 31 16:27:29 UTC 2017",
  "reserveCount":1,
  "demandEntities":[
    {
      "uuid":"_kWgb7TDmEeePgeXuo0RRbw",
      "displayName":"MyReservation_C0",
      "className":"VirtualMachine",
      "template":{
        "uuid":"T423f548d-cadc-e525-6df4-1f90724cf696",
        "displayName":"vsphere-dc3.dev.mycorp.com::TMP-SUSE64",
        "className":"VirtualMachineProfile"
      },
      "placements":{
        "computeResources":[
          {
            "stats":[
              {
                "name":"numOfCpu",
                "value":1
              },
              {
                "name":"cpuSpeed",
                "value":2603
              },
              {
                "name":"cpuConsumedFactor",
                "value":0.5
              },
              {
                "name":"memorySize",
                "value":2097152
              },
              {
                "name":"memoryConsumedFactor",
                "value":0.75
              },
              {
                "name":"ioThroughput",
                "value":0
              },
              {
                "name":"networkThroughput",
                "value":0
              }
            ],
            "provider":{
```

```
              "uuid":"Virtual_ESX_42381da5-12fa-1e82-2f1c-887419380d43",
              "displayName":"hp-esx21.corp.vmturbo.com",
              "className":"PhysicalMachine"
            }
          }
        ],
        "storageResources":[
          {
            "stats":[
              {
                "name":"diskSize",
                "value":18432.363
              },
              {
                "name":"diskIops",
                "value":0
              }
            ],
            "provider":{
              "uuid":"10545c15-7687ef26",
              "displayName":"QS2:ESXDC3DS1",
              "className":"Storage"
            }
          }
        ]
      }
    }
  ]
}
```


Creating a reservation

To create a reservation, you POST an input DTO that defines the reservation's:

- `action`: The action type: RESERVATION
- `demandName`: The display name of the reservation – If you do not specify names for the added workloads, this will be the root name for new VMs
- `placementParameters`:
  - `constraintIDs`: An array of UUIDs for placement policies that will affect the calculated placement
  - `count`: The number of workloads to place
  - `entityNames`: An array of names for the placed VMs – The array length should equal `count`
  - `geographicRedundancy`: If `true` place the workloads on unique hosts, otherwise Workload Optimization Manager can place multiple workloads on the same host
  - `templateID`: The UUID of the template that you will use to place this workload – Note that the template must include a reference to the deployment profile that you specify in `deploymantParameters`
- `expireDateTime`: When to cancel a reservation if Workload Optimization Manager cannot place all the workloads by that date – for a RESERVATION action, Workload Optimization Manager automatically sets the deploy time to equal this time
- `reserveDateTime`: The time to calculate the workload placement and create the reservation – This time cannot be earlier than the time that you POST the reservation to the API

The following listing shows an input DTO that creates a reservation. It will place four workloads, and it gives specific names ot each one.

```
{
  "action": "RESERVATION",
  "demandName": "MyReservation",
  "expireDateTime": "2017-10-10T12:38:17+00:00",
  "parameters": [
```

```
    {
      "deploymentParameters": {
        "deploymentProfileID": "_c9CJMDDAEeePgeXuo0RRbw"
      },
      "placementParameters": {
        "geographicRedundancy": false,
        "count": 4,
        "entityNames": [
          "foo","bar","baz","bonk"
        ],
        "templateID": "_UKsnkJkSEeCHcOXEhzJExA"
      }
    }
  ],
  "reserveDateTime": "2017-05-04T18:22:12+00:00"
}
```

Reserved instances endpoint

Workload Optimization Manager analysis takes advantage of AWS Reserved Instances and Azure Reserved VM Instances to calculate optimal workload placement and to arrive at the best possible costs for your deployments on the cloud.

Using the reserved instances endpoint of the Workload Optimization Manager API, you can:

- Get a single reserved instance or list of reserved instances present in your environment
- Get statistics related to the reserved instances

Reserved instances requests

Getting reserved instances

For a list of reserved instances, the API returns an array of ReservedInstanceApiDTOs. Each ReservedInstanceApiDTO will contain the validation status, target type and category, and input field names and values. To get a single reserved instance, include the UUID of the reserved instance in the request.

This request takes the following parameters:

scope

Returns only reserved instances in the specified scope. This can be the scope of a group of entities or target. For example, to see reserved instances for a single AWS target, use that target's UUID here.

Examples:

- All Reserved instances: `GET https://10.10.10.10/api/v3/reservedinstances`
- Reserved Instances for a Specified Scope: `GET https://10.10.10.10//api/v3/reservedinstances?scope=_zeCtIMAVEei2-fP3xzx1Fw`

Response:

```
[
  {
    "uuid": "aws::eu-west-1::RI::a0643140-85d7-4c5e-a7ba-4ec2fc47bb76",
    "displayName": "i3.large",
    "className": "ReservedInstance",
    "template": {
      "uuid": "aws::VMPROFILE::i3.large",
      "displayName": "i3.large",
      "discovered": false,
      "family": "i3"
```

```
  },
  "platform": "LINUX",
  "type": "STANDARD",
  "location": {
    "uuid": "aws::eu-west-1::DC::eu-west-1",
    "displayName": "aws-EU (Ireland)"
  },
  "accountId": "19265476245",
  "targetId": "_zeCtIMAVEei2-fP3xzx1Fw",
  "masterAccountId": "194356375",
  "term": {
    "units": "Year",
    "values": {
      "max": 1,
      "min": 1,
      "avg": 1,
      "total": 1
    },
    "value": 1
  },
  "payment": "PARTIAL_UPFRONT",
  "onDemandPrice": {
    "units": "$/h",
    "values": {
      "max": 0.172,
      "min": 0.172,
      "avg": 0.172,
      "total": 0.172
    },
    "value": 0.172
  },
  "costPrice": {
    "units": "$/h",
    "values": {
      "max": 0.11193608,
      "min": 0.11193608,
      "avg": 0.11193608,
      "total": 0.11193608
    },
    "value": 0.11193608
  },
  "coupons": {
    "capacity": {
      "max": 16,
      "min": 16,
      "avg": 16
    },
    "units": "RICoupon",
    "values": {
      "max": 0,
      "min": 0,
      "avg": 0,
      "total": 0
    },
    "value": 0
  },
  "totalCoupons": 0,
  "expDate": "Sat Apr 06 19:11:19 EDT 2019",
  "instanceCount": 1,
  "tenancy": "DEFAULT",
  "upFrontCost": 490,
  "actualHourlyCost": 0.0560000017285347,
  "effectiveHourlyCost": 0.11193607478789543
},
...
```

```
]
```

Getting entities covered by reserved instances

Given the UUID of a reserved instance, the API returns a list of entities covered by the reserved instance.

This request takes the following parameters:

reservedInstance_Uuid

The UUID of the reserved instance.

Example: `GET https://10.10.10.10/api/v3/reservedinstances/1234/coveredEntities`

Response: An array of `BaseApiDTO`s describing the entities covered by the specified RI.

```
[
   {
      "uuid":"73943736161360",
      "displayName":"pyaus",
      "className":"VirtualMachine"
   },
   {
      "uuid":"73812178348464",
      "displayName":"ptaub807disabled",
      "className":"VirtualMachine"
   },
   {
      "uuid":"73746051159552",
      "displayName":"pymm",
      "className":"VirtualMachine"
   }
]
```

Roles requests

The roles endpoint is not implemented at this time.

Scenarios endpoint

To run a plan scenario, Workload Optimization Manager creates a snapshot copy of your real-time market and modifies that snapshot according to the scenario. It then uses the Economic Scheduling Engine to perform analysis on that plan market. A scenario can modify the snapshot market by changing the workload, adding or removing hardware resources, or changing or eliminating constraints such as cluster boundaries or placement policies.

As it runs a plan, Workload Optimization Manager continuously analyzes the plan market until it arrives at the optimal conditions that market can achieve. When it reaches that point, the Economic Scheduling Engine cannot find better prices for any of the resources demanded by the workload – the plan stops running, and it displays the results as the plan's desired state.

The display includes the resulting workload distribution across hosts and datastores, as well as a list of actions the plan executed to achieve the desired result.

For example, assume a scenario that adds virtual machines to a cluster. To run the plan, Workload Optimization Manager takes a snapshot of the current market, and adds the VMs to the specified cluster. Workload Optimization Manager then runs analysis on the plan market, where each entity in the supply chain shops for the resources it needs, always looking for a better price –

looking for those resources from less-utilized suppliers. This analysis continues until all the resources are provided at the best possible price.

The results might show that you can add more workload to your environment, even if you reduce compute resources by suspending physical machines. The recommended actions would then indicate which hosts you can take offline, and how to distribute your virtual machines among the remaining hosts.

The settings you make in a scenario correspond to the plan settings you can make in the user interface. These include:

- Plan scope
- Changes to workload (adding, removing, or replacing VMs or containers)
- Changes to supply (adding, removing, replacing PMs or storage)
- Enable/disable placement policies and other constraints
- Changes to action modes
- Enable/disable provisioning of supply
- Enable/disable resizing of workloads

Using the scenarios endpoint of the Workload Optimization Manager API, you can:

- Get a list of scenarios on your instance
- Create a scenario
- Get a list of markets that are connected to a particular scenario

Scenarios requests

Getting scenarios

The API returns a ScenarioApiDTO including the following fields when you execute a `GET /scenarios` request:

- uuid

  The unique identifier for this scenario.

- displayName

  By default, Workload Optimization Manager creates a display name that matches the scenario type. You may rename the scenario after creation using a PUT request.

- owners

  An array of user accounts that can use the scenario. Currently, this is always the user that created the scenario.

- type

  The type of scenario. When a user creates a scenario in the GUI, this matches the given type that user chooses from the Plan Wizard.

- scope

  An array of scope objects to specify the scope of the plan.

- projectionDays

  An array of days to specify how far into the future to project the plan. For example, a value of `[30, 60, 120]` would return projections after one, two, and four months, respectively.

- topologyChanges

  An array of additions or deletions to the set of entities in the plan scope.

- loadChanges

  An array of changes to the utilization percentage for different groups of entities within the plan scope.

- configChanges

  An array of configuration changes for different groups of entities within the plan scope.

- `changes - DEPRECATED`

  An array of scenario settings that make changes to the market before running the plan.

Example: `GET https://10.10.10.10/api/v3/scenarios?show_all=false`

Response:


```
{
    "uuid": "214704587455872",
    "displayName": "Migrate to Public Cloud 1_CLOUD_MIGRATION__CONSUMPTION",
    "type": "CLOUD_MIGRATION",
    "scope": [
      {
        "uuid": "73935388327649",
        "displayName": "aws-EU (Frankfurt)",
        "className": "Region"
      }
    ],
    "projectionDays": [
      0
    ],
    "topologyChanges": {
      "migrateList": [
        {
          "projectionDay": 0,
          "source": {
            "uuid": "285041726554576",
            "displayName": "Unknown"
          },
          "destination": {
            "uuid": "73935388327649",
            "displayName": "aws-EU (Frankfurt)",
            "className": "Region",
            "state": "ACTIVE",
            "severity": "Minor",
            "discoveredBy": {
              "uuid": "73935444966640",
              "displayName": "testmachine",
              "category": "Public Cloud",
              "type": "AWS",
              "readonly": false
            },
            "severityBreakdown": {
              "MINOR": 1
            },
            "numRelatedVMs": 2,
            "environmentType": "CLOUD",
            "vendorIds": {
              "vmturboamp": "aws::eu-central-1::DC::eu-central-1"
            }
          },
          "sources": [
            {
              "uuid": "285041726554576",
              "displayName": "Unknown"
            }
          ],
          "destinations": [
            {
              "uuid": "73935388327649",
              "displayName": "aws-EU (Frankfurt)",
              "className": "Region",
              "state": "ACTIVE",
              "severity": "Minor",
              "discoveredBy": {
```

```
              "uuid": "73935444966640",
              "displayName": "testmachine",
              "category": "Public Cloud",
              "type": "AWS",
              "readonly": false
            },
            "severityBreakdown": {
              "MINOR": 1
            },
            "numRelatedVMs": 2,
            "environmentType": "CLOUD",
            "vendorIds": {
              "vmturboamp": "aws::eu-central-1::DC::eu-central-1"
            }
          }
        ],
        "destinationEntityType": "VirtualMachine",
        "removeNonMigratingWorkloads": true
      }
    ]
  },
  "loadChanges": {
    "utilizationList": [],
    "maxUtilizationList": []
  },
  "configChanges": {
    "addPolicyList": [],
    "removePolicyList": [],
    "automationSettingList": [],
    "removeConstraintList": [],
    "riSettingList": [
      {
        "uuid": "ri.aws.preferredOfferingClass",
        "displayName": "Type",
        "value": "STANDARD",
        "valueObjectType": "String"
      },
      {
        "uuid": "ri.aws.preferredPaymentOption",
        "displayName": "Payment",
        "value": "ALL_UPFRONT",
        "valueObjectType": "String"
      },
      {
        "uuid": "ri.aws.preferredTerm",
        "displayName": "Term",
        "value": "YEARS_1",
        "valueObjectType": "String"
      }
    ],
    "osMigrationSettingList": [
      {
        "uuid": "linuxTargetOs",
        "displayName": "linuxTargetOs",
        "value": "LINUX",
        "valueObjectType": "String"
      },
      {
        "uuid": "linuxByol",
        "displayName": "linuxByol",
        "value": "false",
        "valueObjectType": "String"
      },
      {
        "uuid": "rhelTargetOs",
```

```
          "displayName": "rhelTargetOs",
          "value": "RHEL",
          "valueObjectType": "String"
        },
        {
          "uuid": "rhelByol",
          "displayName": "rhelByol",
          "value": "false",
          "valueObjectType": "String"
        },
        {
          "uuid": "slesTargetOs",
          "displayName": "slesTargetOs",
          "value": "SLES",
          "valueObjectType": "String"
        },
        {
          "uuid": "slesByol",
          "displayName": "slesByol",
          "value": "false",
          "valueObjectType": "String"
        },
        {
          "uuid": "windowsTargetOs",
          "displayName": "windowsTargetOs",
          "value": "WINDOWS",
          "valueObjectType": "String"
        },
        {
          "uuid": "windowsByol",
          "displayName": "windowsByol",
          "value": "false",
          "valueObjectType": "String"
        },
        {
          "uuid": "selectedMigrationProfileOption",
          "displayName": "selectedMigrationProfileOption",
          "value": "MATCH_SOURCE_TO_TARGET_OS",
          "valueObjectType": "String"
        },
        {
          "uuid": "matchToSource",
          "displayName": "matchToSource",
          "value": "true",
          "valueObjectType": "String"
        }
      ],
      "includedCoupons": {
        "includedCouponOidsList": [],
        "iswhiteList": true
      },
      "subscription": {
        "uuid": "73935382668544",
        "displayName": "EngineeringSample",
        "className": "BusinessAccount",
        "accountId": "323871187550",
        "master": false,
        "cloudType": "AWS",
        "riSupported": false,
        "hasRelatedTarget": false
      }
    }
  }
```

Creating a scenario

In order to create a scenario, see the Scenario DTO breakdown *(on page 1303)*. For examples of complete ScenarioApiDTOs, see Scenario configuration examples *(on page 1314)*.

Deleting a scenario

Deletes the specified scenario. Note that this does not delete plan markets you create using that scenario.

Example: `DELETE https://10.10.10.10/api/v3/scenarios/__xwWYAE3EeiNKPbzHdETIA`

Response: A successful deletion request returns a Response Code of `200`.

Scenario DTO breakdown

Scenario DTO overview

A scenario DTO assembles configuration settings that you use to set up and run a plan. To run a plan, you will POST the scenario to a market. For most plans, this will be the real-time market. In the case of a plan-over-plan, this will be the market created from the first scenario.

Workload Optimization Manager will then run a plan based on the combination of the scenario and the market state.

These configuration settings fall into one of several groups:

- Configuration Changes (ConfigChangesApiDTO)

  Adding and removing placement policies, removing constraints, and providing automation settings, OS migration settings, reserved instance settings, and associated business units.
- Load Changes (LoadChangesApiDTO)

  Modifications to the current and max utilization values, and peak baseline values.
- Time-Based Topology Changes (TimeBasedTopologyChangesApiDTO)

  Settings to include reserved VMs and periodically add historical workload to the environment.
- Topology Changes (TopologyChangesApiDTO)

  Settings to add, remove, replace, and migrate workload, or to specify the clusters for an Alleviate Pressure plan.

DTO configuration changes

The scenario configuration changes, located in the `configChanges` parameter of the ScenarioApiDTO, enable you to perform the following tasks:

- addPolicyList

  Add a placement policy to the market generated for this plan.
- automationSettingList

  Change one of the available automation plan settings: VM Resize, Host Provision/Suspend, and Storage Provision/Suspend.
- osMigrationSettingList

  For a Migrate to Cloud plan, this parameter will contain the source and destination OS platforms, and their BYOL (Bring Your Own License) status.
- removeConstraintList

  Remove a constraint from the market generated for this plan.
- removePolicyList

  Remove a placement policy from the market generated for this plan.

- riSettingList

  Add information about your reserved cloud instances.
- subscription

  Associate a business unit to the plan.

Adding and Removing Placement Policies

Referenced Endpoint: Policies (on page 1285)

When running plans, you might want to see what your environment would look like with or without certain policies enabled, or with a new policy present. You can see the effects of these changes by passing the UUIDs of current policies, or a PolicyApiDTO for a new policy.

Example input (Current Policies):

```
"addPolicyList": [
     {
       "uuid": "_VnNS0CsqEemvEL3szcXU1Q",
       "uuid": "522a3317-39f5-1a97-3b53-d725927a6ae8",
       "uuid": "52672058-a93b-d0a9-bbd9-5acd33f27882"
     }
   ]
```

Example input (New Policy):

```
"addPolicyList": [
     {
       "capacity": null,
       "consumerGroup": {
         "uuid": "8c17d8e79c3db85ed5d3868de622cb89187849ce",
         "displayName": "PMs_Beta DC6\\Cluster-2",
         "className": "Group",
         "groupType": "PhysicalMachine",
         "isStatic": true,
         "logicalOperator": "AND"
       },
       "enabled": true,
       "mergeGroups": null,
       "mergeType": null,
       "name": "DalTestPolMerge",
       "displayName": "DalTestPolMerge",
       "providerGroup": {
         "uuid": "91ff08749fd278925d73b61b0f0f69d79c99fc3a",
         "displayName": "PMs_Beta DC6\\Cluster-1",
         "className": "Group",
         "groupType": "PhysicalMachine",
         "isStatic": true,
         "logicalOperator": "AND"
       },
       "type": "MERGE",
       "uuid": ""
     }
   ]
```

Setting Plan Automation Settings

Use the plan automation settings to see the effect of your scenario if VM resizes, Host Provisioning / Suspension, and Storage Provisioning / Suspension are enabled or disabled.

Example Input:

```
"automationSettingList":[
  {
    "uuid":"resize",
    "displayName":"Resize for VMS disabled",
    "value":"false",
    "entityType":"VirtualMachine"
  },
  {
    "uuid":"provisionPM",
    "displayName":"Provision for PHYSICALMACHINES disabled",
    "value":"false",
    "entityType":"PhysicalMachine"
  },
  {
    "uuid":"suspendPM",
    "displayName":"Suspend for PHYSICALMACHINES disabled",
    "value":"false",
    "entityType":"PhysicalMachine"
  },
  {
    "uuid":"provisionDS",
    "displayName":"Provision for STS disabled",
    "value":"false",
    "entityType":"Storage"
  },
  {
    "uuid":"suspendDS",
    "displayName":"Suspend for STS enabled",
    "value":"true",
    "entityType":"Storage"
  }
]
```

Additionally, you can use these settings to adjust operational constraints, such as provisioned values for CPU and MEM:

Example Input:

```
"automationSettingList":[
  {
    "uuid":"capacity_MemProvisioned",
    "displayName":"Mem Overprovisioned Percentage",
    "value":"9876.0",
    "defaultValue":"1000.0",
    "categories":[
      "utilizationThresholds"
    ],
    "valueType":"NUMERIC",
    "min":1,
    "max":1000000,
    "entityType":"PhysicalMachine",
    "sourceGroupName":"Global"
  }
```

Providing Cloud Reserved Instance Information

Workload Optimization Manager enables you to upload RI pricing information specific to the scenario.

Example Input:

```
"riSettingList": [
      {
        "uuid": "preferredTerm",
```

```
        "displayName": "Term",
        "value": "YEARS_1",
        "entityType": "YEARS_1"
      },
      {
        "uuid": "preferredPaymentOption",
        "displayName": "Payment",
        "value": "PARTIAL_UPFRONT",
        "entityType": "PARTIAL_UPFRONT"
      },
      {
        "uuid": "preferredOfferingClass",
        "displayName": "Type",
        "value": "STANDARD",
        "entityType": "STANDARD"
      },
      {
        "uuid": "preferredCoverage",
        "displayName": "Coverage",
        "value": "80",
        "entityType": "80"
      },
      {
        "uuid": "riCoverageOverride",
        "displayName": "RI Coverage Override",
        "value": "false",
        "entityType": "false"
      }
    ]
```

Removing Constraints

Referenced Endpoint: Entities (on page 1195)

When running plans, you might want to see what your environment would look like with or without certain constraints. For example, if a group of virtual machines was allowed to move to any cluster of your environment.

In order to remove a constraint, you must know its UUID. To get a list of constraints for this entity, execute the `/entities/ENTITY_UUID/constraints` request for the entity whose constraint you wish to remove.

```
"removeConstraintList":[
  {
    "constraintType":"ClusterCommodity",
    "projectionDay":0,
    "target":{
      "uuid":"a6d365f4fbf03fb9a2e8cdb29e38045e31cee41b",
      "displayName":"Beta DC6_vm",
      "className":"VirtualMachine"
    }
  }
]
```

Setting OS Migration Settings for a Migrate to Cloud Plan

When performing a Migrate to Cloud plan, you might want to use custom OS migration settings. You can specify both the OS mapping, and the BYOL (bring your own license) status of each OS in the mapping.

Example Input:

```
"osMigrationSettingList":[
  {
    "uuid":"linuxTargetOs",
```

```
    "displayName":"linuxTargetOs",
    "value":"LINUX"
  },
  {
    "uuid":"linuxByol",
    "displayName":"linuxByol",
    "value":"true"
  },
  {
    "uuid":"rhelTargetOs",
    "displayName":"rhelTargetOs",
    "value":"RHEL"
  },
  {
    "uuid":"rhelByol",
    "displayName":"rhelByol",
    "value":"false"
  },
  {
    "uuid":"slesTargetOs",
    "displayName":"slesTargetOs",
    "value":"WINDOWS"
  },
  {
    "uuid":"slesByol",
    "displayName":"slesByol",
    "value":"true"
  },
  {
    "uuid":"windowsTargetOs",
    "displayName":"windowsTargetOs",
    "value":"WINDOWS"
  },
  {
    "uuid":"windowsByol",
    "displayName":"windowsByol",
    "value":"true"
  },
  {
    "uuid":"showMatchSourceToTargetOsOption",
    "displayName":"showMatchSourceToTargetOsOption",
    "value":"true"
  },
  {
    "uuid":"showByolOption",
    "displayName":"showByolOption",
    "value":"false"
  },
  {
    "uuid":"showCustomOsOption",
    "displayName":"showCustomOsOption",
    "value":"true"
  },
  {
    "uuid":"matchToSource",
    "displayName":"matchToSource",
    "value":"false"
  },
  {
    "uuid":"selectedMigrationProfileOption",
    "displayName":"selectedMigrationProfileOption",
    "value":"CUSTOM_OS"
  }
]
```

Scenario DTO load changes

The scenario load changes, located in the `loadChanges` parameter of the ScenarioApiDTO, make workload utilization changes that enable you to change the following values:

- baselineDate

  If your plan should be run using historical data, the date of the data to be used.
- maxUtilizationList

  Changes the max utilization values for a group of entities in your plan.
- peakBaselineList

  Changes the peak baseline values for a group of entities in your plan.
- utilizationList

  Changes the utilization values for a group of entities in your plan.

Changing the Baseline Date

Example input:

```
"baselineDate": 1552622400000
```

Changing the Utilization For a Group of Entities

Example input:

```
  "utilizationList": [
    {
      "percentage": -12,
      "projectionDay": 0,
      "target": {
        "displayName": "Global Environment",
        "uuid": "Market"
      }
    }
  ]
```

Changing the Maximum Utilization For a Group of Entities

Example input:

```
  "maxUtilizationList": [
    {
      "maxPercentage": 82,
      "projectionDay": 0,
      "target": {
        "displayName": "All On-prem Hosts",
        "className": "Group",
        "uuid": "_PE0v-YEUEee_hYfzgV9uYg"
      }
    },
    {
      "maxPercentage": 85,
      "projectionDay": 0,
      "target": {
        "displayName": "ACM_datastore",
        "className": "Group",
        "uuid": "c31a6f99f4409700115fc61fca71e5449ad68700"
      }
    },
    {
```

```
      "maxPercentage": 77,
      "projectionDay": 0,
      "target": {
        "displayName": "Beta DC6_datastore",
        "className": "Group",
        "uuid": "5544c35ed3d7fc0d33879d3457c219d7e40fc176"
      }
    }
  ]
}
```

Scenario DTO time-based topology changes

The scenario time-based topology changes, located in the `timebasedTopologyChanges` parameter of the ScenarioApiDTO, make time-dependent topology changes to your scenario, such as:

- addHistoryVMs

  If `true`, Workload Optimization Manager will add additional workload to your scenario each month equal to the workload added to the real-time market in the previous month.

- includeReservation

  If `true`, Workload Optimization Manager will take current reservations and reflect their resource utilization when the scenario runs.

Adding Historical VM Usage and/or Including VM Reservations

Example input:

```
"timebasedTopologyChanges": [
  {
    "addHistoryVMs": "true",
    "includeReservation": "true"
  }
]
```

Scenario DTO topology changes

The scenario topology changes, located in the `topologyChanges` parameter of the ScenarioApiDTO, make topology changes to your scenario, such as:

- addList

  A list of entities to add to your environment as part of plan configuration.

- migrateList

  A list of entities to move from one set of providers to another as part of plan configuration.

- relievePressureList

  In a Relieve Pressure plan, the source and target clusters.

- removeList

  A list of entities to be removed from your environment as part of plan configuration.

- replaceList

  A list of entities and the templates to replace them in your environment as part of plan configuration.

Adding Workload

Adding workload to a scenario simulates the effect of increased VM or container density in your environment. You can add multiple copies of VM groups or individual VMs.

Example input:

```
"addList": [
  {
    "count": 3,
    "projectionDays": [
      0
    ],
    "target": {
      "uuid": "421dd4ce-a4c1-231a-71e7-64fbf598bc65",
      "displayName": "AJ_Test",
      "className": "VirtualMachine"
    }
  },
  {
    "count": 10,
    "projectionDays": [
      0
    ],
    "target": {
      "uuid": "421e503d-d0c0-75f9-8cce-79421924a90d",
      "displayName": "olg_test",
      "className": "VirtualMachine"
    }
  },
  {
    "count": 5,
    "projectionDays": [
      0
    ],
    "target": {
      "uuid": "420f1840-297f-5a39-9543-b72046b9126f",
      "displayName": "opensuse_with_stress_template",
      "className": "VirtualMachine"
    }
  },
  {
    "count": 1,
    "projectionDays": [
      0
    ],
    "target": {
      "uuid": "53523c9a4c5d67962a96f90f59456b0c3e72aca0",
      "displayName": "Development67_vm",
      "className": "VirtualMachine"
    }
  }
]
```

Migrating Workload to the Cloud

Workload Optimization Manager includes a special plan to simulate migration of on-prem workload to the cloud, or migration of workloads from one cloud to another. This plan focuses on optimizing your costs on the cloud by choosing the best templates (most adequate compute resources) and regions to host your workloads. If analysis discovers on-prem workloads that are good candidates for Reserved Instances (RIs), then it recommends migrating to those templates, and can even recommend purchasing more RI capacity. For each group of virtual machines to be migrated (source), you must provide a destination.

Example input:

```
"migrateList": [
  {
    "projectionDay": 0,
    "source": {
```

```
        "uuid": "4656c0ba6b976d05d1a9470ee673672022e8e926",
        "displayName": "ACM_vm",
        "className": "VirtualMachine",
        "severity": "Major"
      },
      "destination": {
        "uuid": "GROUP-PMsByTargetType_Azure",
        "displayName": "PMs_Azure",
        "className": "PhysicalMachine"
      }
    },
    {
      "projectionDay": 0,
      "source": {
        "uuid": "fe23c8642efc5ad2ed4fa1cc502dfb179b8dc9d2",
        "displayName": "Development DC1_vm",
        "className": "VirtualMachine",
        "severity": "Major"
      },
      "destination": {
        "uuid": "GROUP-PMsByTargetType_Azure",
        "displayName": "PMs_Azure",
        "className": "PhysicalMachine"
      }
    },
    {
      "projectionDay": 0,
      "source": {
        "uuid": "f0e26cb417d8143173270b7f01c7213e027cd9e5",
        "displayName": "Development DC7_vm",
        "className": "VirtualMachine",
        "severity": "Major"
      },
      "destination": {
        "uuid": "GROUP-PMsByTargetType_Azure",
        "displayName": "PMs_Azure",
        "className": "PhysicalMachine"
      }
    }
  ]
```


Relieve Pressure

Including this parameter creates an Alleviate Pressure plan to find out how to migrate workloads from a stressed or hot cluster over to a cluster with more headroom.

Example input:


```
  "relievePressureList": [
    {
      "projectionDay": 0,
      "sources": [
        {
          "uuid": "11c63ebd77647e7b9c9fb85dd2d4c77464cbeac2",
          "displayName": "DC13\\DC13-Cluster",
          "className": "Cluster"
        }
      ],
      "destinations": [
        {
          "uuid": "57a4026a69ba4b46ebc430cb62d5c55eeea00695",
          "displayName": "DC13\\Physical Hosts_Cluster2 - Intel",
          "className": "Cluster"
        }
```

```
        ]
      }
    ]
```

Removing Workload

Removing workload in a scenario configuration simulates the effect of decreased VM or container density in your environment. You can remove VM groups or individual VMs.

Example input:

```
  "removeList": [
    {
      "projectionDay": 0,
      "target": {
        "uuid": "5E4F6056-8783-45BC-9070-35CF09457E98",
        "displayName": "AH-TestHyperV-VM001",
        "className": "VirtualMachine"
      }
    },
    {
      "projectionDay": 0,
      "target": {
        "uuid": "421d62c2-964a-0f4f-32cb-ec4cec9e26db",
        "displayName": "Anshuman_MidServer",
        "className": "VirtualMachine"
      }
    },
    {
      "projectionDay": 0,
      "target": {
        "uuid": "421e048b-f66b-539c-f743-7dcc7b40943b",
        "displayName": "Gilad_XL_Storage_Testing - Pure-1",
        "className": "VirtualMachine"
      }
    },
    {
      "projectionDay": 0,
      "target": {
        "uuid": "421e9199-7159-04f2-29fc-6a501f44d1be",
        "displayName": "Gilad_XL_Storage_Testing - Pure-2",
        "className": "VirtualMachine"
      }
    },
    {
      "projectionDay": 0,
      "target": {
        "uuid": "421d9ec6-ab97-84d0-4b12-2db88ae43e2a",
        "displayName": "MS-SQLServer-3",
        "className": "VirtualMachine"
      }
    }
  ]
```

Replacing Workload

When you replace workload, you select one or more VMs or containers that you want to change (target), and then you select a template to use in their place.

Example input:

```
  "replaceList": [
    {
      "projectionDay": 0,
      "target": {
        "uuid": "421ef003-e152-5a25-6e03-19698340cd01",
        "displayName": "centos_imp_7",
        "className": "VirtualMachine",
        "severity": "Minor"
      },
      "template": {
        "uuid": "_wIBz8Jj4EeC6nYMiQT1jqA",
        "displayName": "Microsoft_SQL2008-small",
        "className": "VirtualMachineProfile"
      }
    },
    {
      "projectionDay": 0,
      "target": {
        "uuid": "42001807-bc8f-b144-8ff6-eb0d0d32cc7a",
        "displayName": "cheuk-OpsMgr-172.113",
        "className": "VirtualMachine",
        "severity": "Minor"
      },
      "template": {
        "uuid": "_wIBz8Jj4EeC6nYMiQT1jqA",
        "displayName": "Microsoft_SQL2008-small",
        "className": "VirtualMachineProfile"
      }
    },
    {
      "projectionDay": 0,
      "target": {
        "uuid": "4211806a-65f2-8be1-1015-a4abe1a2ab6f",
        "displayName": "dim-dc13amd-6.2.7_01",
        "className": "VirtualMachine",
        "severity": "Minor"
      },
      "template": {
        "uuid": "_wIBz8Jj4EeC6nYMiQT1jqA",
        "displayName": "Microsoft_SQL2008-small",
        "className": "VirtualMachineProfile"
      }
    },
    {
      "projectionDay": 0,
      "target": {
        "uuid": "42008beb-264e-b0f6-2295-f3443e4ce7d6",
        "displayName": "enOpenShiftInfra1",
        "className": "VirtualMachine",
        "severity": "Minor"
      },
      "template": {
        "uuid": "_wIBz8Jj4EeC6nYMiQT1jqA",
        "displayName": "Microsoft_SQL2008-small",
        "className": "VirtualMachineProfile"
      }
    },
    {
      "projectionDay": 0,
      "target": {
        "uuid": "420036f6-27c3-0252-439d-e23bcff63d02",
        "displayName": "enOpenShiftInfra2",
        "className": "VirtualMachine",
        "severity": "Minor"
      },
```

```
      "template": {
        "uuid": "_wIBz8Jj4EeC6nYMiQT1jqA",
        "displayName": "Microsoft_SQL2008-small",
        "className": "VirtualMachineProfile"
      }
    },
    {
      "projectionDay": 0,
      "target": {
        "uuid": "4200ba09-71eb-3084-026e-1770c23abaa4",
        "displayName": "enOpenshiftDNS",
        "className": "VirtualMachine",
        "severity": "Minor"
      },
      "template": {
        "uuid": "_wIBz8Jj4EeC6nYMiQT1jqA",
        "displayName": "Microsoft_SQL2008-small",
        "className": "VirtualMachineProfile"
      }
    }
  ]
```


Scenario configuration examples

Basic scenario POSTs and responses

Most scenarios consist of multiple configuration changes. However, the UI provides several starting points for custom configuration based on plan type.

Workload Optimization Manager provides the same starting point in the API via the `/scenarios` endpoint and the following list of POST calls and server responses:

The following examples are simplified in order to accomplish a single goal. There are many other options you might choose in conjunction with those in the following examples.

Add workload

POST:


```
{
  "configChanges": {
    "addPolicyList": [],
    "automationSettingList": [],
    "removeConstraintList": [],
    "removePolicyList": [],
    "riSettingList": [],
    "osMigrationSettingList": [],
    "subscription": {}
  },
  "displayName": "Add Workload- Basic",
  "loadChanges": {
    "utilizationList": [],
    "maxUtilizationList": []
  },
  "projectionDays": [
    0
  ],
  "scope": [
    {
      "uuid": "Market",
```

```
      "displayName": "Global Environment",
      "className": "Market"
    }
  ],
  "topologyChanges": {
    "addList": [
      {
        "count": 50,
        "projectionDays": [
          0
        ],
        "target": {
          "uuid": "421dd4ce-a4c1-231a-71e7-64fbf598bc65",
          "displayName": "AJ_Test",
          "className": "VirtualMachine"
        }
      }
    ],
    "migrateList": [],
    "removeList": [],
    "replaceList": [],
    "relievePressureList": []
  },
  "type": "ADD_WORKLOAD"
}
```

Response:

```
{
  "uuid": "2502057162928",
  "displayName": "Add Workload- Basic",
  "owners": [
    {
      "uuid": "_4T_7kwY-Ed-WUKbEYSVIDw",
      "username": "administrator",
      "showSharedUserSC": false
    }
  ],
  "type": "ADD_WORKLOAD",
  "scope": [
    {
      "uuid": "_0x3OYUglEd-gHc4L513yOA",
      "displayName": "Market",
      "className": "Market"
    }
  ],
  "projectionDays": [
    0
  ],
  "topologyChanges": {
    "addList": [
      {
        "projectionDays": [
          0
        ],
        "target": {
          "uuid": "421dd4ce-a4c1-231a-71e7-64fbf598bc65",
          "displayName": "AJ_Test",
          "className": "VirtualMachine"
        },
        "count": 50
      }
    ]
  },
```

```
  "loadChanges": {},
  "configChanges": {},
  "changes": [
    {
      "index": 1,
      "type": "SCOPE",
      "scope": [
        {
          "uuid": "_0x3OYUglEd-gHc4L513yOA",
          "displayName": "Market",
          "className": "Market"
        }
      ]
    },
    {
      "index": 2,
      "type": "PROJECTION_PERIODS",
      "projectionDays": [
        0
      ]
    },
    {
      "index": 3,
      "type": "ADDED",
      "description": "50 AJ_Test",
      "targets": [
        {
          "uuid": "421dd4ce-a4c1-231a-71e7-64fbf598bc65",
          "displayName": "AJ_Test",
          "className": "VirtualMachine",
          "severity": "Major"
        }
      ],
      "value": "50",
      "projectionDays": [
        0
      ]
    }
  ],
  "timebasedTopologyChanges": {}
}
```


Decommission hosts

POST:


```
{
  "configChanges": {
    "addPolicyList": [],
    "automationSettingList": [
      {
        "uuid": "provisionPM",
        "displayName": "PROVISION for PHYSICALMACHINES disabled",
        "value": "false",
        "entityType": "PhysicalMachine"
      }
    ],
    "removeConstraintList": [],
    "removePolicyList": [],
    "riSettingList": [],
    "osMigrationSettingList": [],
    "subscription": {}
  },
```

```
  "displayName": "Decommission Hosts- Basic",
  "loadChanges": {
    "utilizationList": [],
    "maxUtilizationList": []
  },
  "projectionDays": [
    0
  ],
  "scope": [
    {
      "uuid": "Market",
      "displayName": "Global Environment",
      "className": "Market"
    }
  ],
  "topologyChanges": {
    "addList": [],
    "migrateList": [],
    "removeList": [
      {
        "projectionDay": 0,
        "target": {
          "uuid": "4C4C4544-0042-5A10-804A-B1C04F384331",
          "displayName": "dell-pe1944.corp.vmturbo.com",
          "className": "PhysicalMachine"
        }
      },
      {
        "projectionDay": 0,
        "target": {
          "uuid": "4C4C4544-004A-3610-8057-C8C04F464331",
          "displayName": "dell-pe1953.corp.vmturbo.com",
          "className": "PhysicalMachine"
        }
      },
      {
        "projectionDay": 0,
        "target": {
          "uuid": "4C4C4544-0059-4D10-8030-B3C04F464331",
          "displayName": "dell-pe1954.corp.vmturbo.com",
          "className": "PhysicalMachine"
        }
      }
    ],
    "replaceList": [],
    "relievePressureList": []
  },
  "type": "DECOMMISSION_HOST"
}
```


Response:


```
{
  "uuid": "2502061312640",
  "displayName": "Decommission Hosts- Basic",
  "owners": [
    {
      "uuid": "_4T_7kwY-Ed-WUKbEYSVIDw",
      "username": "administrator",
      "showSharedUserSC": false
    }
  ],
  "type": "DECOMMISSION_HOST",
  "scope": [
```

```
    {
      "uuid": "_0x3OYUglEd-gHc4L513yOA",
      "displayName": "Market",
      "className": "Market"
    }
  ],
  "projectionDays": [
    0
  ],
  "topologyChanges": {
    "removeList": [
      {
        "projectionDay": 0,
        "target": {
          "uuid": "4C4C4544-0042-5A10-804A-B1C04F384331",
          "displayName": "dell-pe1944.corp.vmturbo.com",
          "className": "PhysicalMachine"
        }
      },
      {
        "projectionDay": 0,
        "target": {
          "uuid": "4C4C4544-004A-3610-8057-C8C04F464331",
          "displayName": "dell-pe1953.corp.vmturbo.com",
          "className": "PhysicalMachine"
        }
      },
      {
        "projectionDay": 0,
        "target": {
          "uuid": "4C4C4544-0059-4D10-8030-B3C04F464331",
          "displayName": "dell-pe1954.corp.vmturbo.com",
          "className": "PhysicalMachine"
        }
      }
    ]
  },
  "loadChanges": {},
  "configChanges": {
    "automationSettingList": [
      {
        "uuid": "provisionPM",
        "displayName": "Provision",
        "value": "false",
        "defaultValue": "RECOMMEND",
        "valueType": "STRING",
        "options": [
          {
            "label": "Disabled",
            "value": "DISABLED"
          },
          {
            "label": "Recommend",
            "value": "RECOMMEND"
          },
          {
            "label": "Manual",
            "value": "MANUAL"
          },
          {
            "label": "Automated",
            "value": "AUTOMATIC"
          }
        ],
        "entityType": "PhysicalMachine",
```

```
      "sourceGroupName": "Global"
    }
  ]
},
"changes": [
  {
    "index": 1,
    "type": "SCOPE",
    "scope": [
      {
        "uuid": "_0x3OYUglEd-gHc4L513yOA",
        "displayName": "Market",
        "className": "Market"
      }
    ]
  },
  {
    "index": 2,
    "type": "PROJECTION_PERIODS",
    "projectionDays": [
      0
    ]
  },
  {
    "index": 3,
    "type": "REMOVED",
    "description": "dell-pe1944.corp.vmturbo.com",
    "targets": [
      {
        "uuid": "4C4C4544-0042-5A10-804A-B1C04F384331",
        "displayName": "dell-pe1944.corp.vmturbo.com",
        "className": "PhysicalMachine",
        "severity": "Critical"
      }
    ],
    "value": "1",
    "projectionDays": [
      0
    ]
  },
  {
    "index": 4,
    "type": "REMOVED",
    "description": "dell-pe1953.corp.vmturbo.com",
    "targets": [
      {
        "uuid": "4C4C4544-004A-3610-8057-C8C04F464331",
        "displayName": "dell-pe1953.corp.vmturbo.com",
        "className": "PhysicalMachine",
        "severity": "Critical"
      }
    ],
    "value": "1",
    "projectionDays": [
      0
    ]
  },
  {
    "index": 5,
    "type": "REMOVED",
    "description": "dell-pe1954.corp.vmturbo.com",
    "targets": [
      {
        "uuid": "4C4C4544-0059-4D10-8030-B3C04F464331",
        "displayName": "dell-pe1954.corp.vmturbo.com",
```

```
          "className": "PhysicalMachine",
          "severity": "Critical"
        }
      ],
      "value": "1",
      "projectionDays": [
        0
      ]
    },
    {
      "index": 6,
      "type": "SET_ACTION_SETTING",
      "description": "Provision for PhysicalMachine disabled",
      "targets": [
        {
          "uuid": "PhysicalMachine"
        }
      ],
      "value": "PhysicalMachine",
      "enable": false,
      "name": "provision",
      "projectionDays": [
        0
      ]
    }
  ],
  "timebasedTopologyChanges": {}
}
```


Migrate to cloud

POST:


```
{
  "configChanges": {
    "addPolicyList": [],
    "automationSettingList": [],
    "removeConstraintList": [],
    "removePolicyList": [],
    "riSettingList": [],
    "osMigrationSettingList": [],
    "subscription": {}
  },
  "displayName": "Migrate to Public Cloud- Basic",
  "loadChanges": {
    "utilizationList": [],
    "maxUtilizationList": []
  },
  "projectionDays": [
    0
  ],
  "scope": [
    {
      "uuid": "4656c0ba6b976d05d1a9470ee673672022e8e926",
      "displayName": "ACM_vm",
      "groupType": "VirtualMachine",
      "className": "Group"
    },
    {
      "uuid": "_lRrTYB--EeewItqBJctLGw",
      "displayName": "All Cloud Zones",
      "groupType": "PhysicalMachine",
      "className": "Group"
```

```
    }
  ],
  "topologyChanges": {
    "addList": [],
    "migrateList": [
      {
        "projectionDay": 0,
        "source": {
          "uuid": "4656c0ba6b976d05d1a9470ee673672022e8e926",
          "displayName": "ACM_vm",
          "className": "VirtualMachine"
        },
        "destination": {
          "uuid": "_lRrTYB--EeewItqBJctLGw",
          "displayName": "All Cloud Zones",
          "className": "PhysicalMachine"
        }
      }
    ],
    "removeList": [
      {
        "projectionDay": 0,
        "target": {
          "uuid": "_nuoxwETSEemi-4wSMHOOWA",
          "displayName": "All VMs In All Cloud Zones",
          "className": "VirtualMachine"
        }
      }
    ],
    "replaceList": [],
    "relievePressureList": []
  },
  "type": "CLOUD_MIGRATION"
}
```

Response:

```
{
  "uuid": "2502081867536",
  "displayName": "Migrate to Public Cloud- Basic",
  "owners": [
    {
      "uuid": "_4T_7kwY-Ed-WUKbEYSVIDw",
      "username": "administrator",
      "showSharedUserSC": false
    }
  ],
  "type": "CLOUD_MIGRATION",
  "scope": [
    {
      "uuid": "4656c0ba6b976d05d1a9470ee673672022e8e926",
      "displayName": "ACM_vm",
      "className": "Group",
      "groupType": "VirtualMachine",
      "isStatic": true,
      "logicalOperator": "AND"
    },
    {
      "uuid": "_lRrTYB--EeewItqBJctLGw",
      "displayName": "All Cloud Zones",
      "className": "Group",
      "groupType": "PhysicalMachine",
      "isStatic": true,
      "logicalOperator": "AND"
```

```
    }
  ],
  "projectionDays": [
    0
  ],
  "topologyChanges": {
    "removeList": [
      {
        "projectionDay": 0,
        "target": {
          "uuid": "_nuoxwETSEemi-4wSMHOOWA",
          "displayName": "All VMs In All Cloud Zones",
          "className": "Group"
        }
      }
    ],
    "migrateList": [
      {
        "projectionDay": 0,
        "source": {
          "uuid": "4656c0ba6b976d05d1a9470ee673672022e8e926",
          "displayName": "ACM_vm",
          "className": "Group"
        },
        "destination": {
          "uuid": "_lRrTYB--EeewItqBJctLGw",
          "displayName": "All Cloud Zones",
          "className": "Group"
        }
      }
    ]
  },
  "loadChanges": {},
  "configChanges": {
    "osMigrationSettingList": [
      {
        "uuid": "matchToSource",
        "value": "true"
      }
    ]
  },
  "changes": [
    {
      "index": 1,
      "type": "SCOPE",
      "scope": [
        {
          "uuid": "4656c0ba6b976d05d1a9470ee673672022e8e926",
          "displayName": "ACM_vm",
          "className": "Group",
          "entitiesCount": 83,
          "groupType": "VirtualMachine",
          "severity": "Major",
          "logicalOperator": "AND"
        },
        {
          "uuid": "_lRrTYB--EeewItqBJctLGw",
          "displayName": "All Cloud Zones",
          "className": "Group",
          "entitiesCount": 78,
          "groupType": "PhysicalMachine",
          "severity": "Normal",
          "logicalOperator": "AND"
        }
      ]
```

```
  },
  {
    "index": 2,
    "type": "PROJECTION_PERIODS",
    "projectionDays": [
      0
    ]
  },
  {
    "index": 3,
    "type": "REMOVED",
    "description": "All VMs In All Cloud Zones",
    "targets": [
      {
        "uuid": "_nuoxwETSEemi-4wSMHOOWA",
        "displayName": "All VMs In All Cloud Zones",
        "className": "Group",
        "entitiesCount": 154,
        "groupType": "VirtualMachine",
        "severity": "Major",
        "logicalOperator": "AND"
      }
    ],
    "value": "1",
    "projectionDays": [
      0
    ]
  },
  {
    "index": 4,
    "type": "MIGRATION",
    "description": "Migrate [ACM_vm] to All Cloud Zones",
    "targets": [
      {
        "uuid": "4656c0ba6b976d05d1a9470ee673672022e8e926",
        "displayName": "ACM_vm",
        "className": "Group",
        "entitiesCount": 83,
        "groupType": "VirtualMachine",
        "severity": "Major",
        "logicalOperator": "AND"
      },
      {
        "uuid": "_lRrTYB--EeewItqBJctLGw",
        "displayName": "All Cloud Zones",
        "className": "Group",
        "entitiesCount": 78,
        "groupType": "PhysicalMachine",
        "severity": "Normal",
        "logicalOperator": "AND"
      }
    ],
    "projectionDays": [
      0
    ]
  },
  {
    "index": 5,
    "type": "SET_TARGET_OS",
    "targets": [
      {
        "uuid": "4656c0ba6b976d05d1a9470ee673672022e8e926",
        "displayName": "ACM_vm",
        "className": "Group",
        "entitiesCount": 83,
```

```
          "groupType": "VirtualMachine",
          "severity": "Major",
          "logicalOperator": "AND"
        }
      ],
      "projectionDays": [
        0
      ]
    }
  ],
  "timebasedTopologyChanges": {}
}
```

Optimize cloud

POST:

```
{
  "configChanges": {
    "addPolicyList": [],
    "automationSettingList": [
      {
        "uuid": "resize",
        "displayName": "resize for VMS enabled",
        "value": "true",
        "entityType": "VirtualMachine"
      }
    ],
    "removeConstraintList": [],
    "removePolicyList": [],
    "riSettingList": [
      {
        "uuid": "preferredOfferingClass",
        "displayName": "Type",
        "value": "STANDARD",
        "entityType": "STANDARD"
      },
      {
        "uuid": "preferredTerm",
        "displayName": "Term",
        "value": "YEARS_3",
        "entityType": "YEARS_3"
      },
      {
        "uuid": "preferredPaymentOption",
        "displayName": "Payment",
        "value": "ALL_UPFRONT",
        "entityType": "ALL_UPFRONT"
      },
      {
        "uuid": "preferredCoverage",
        "displayName": "Coverage",
        "value": "80",
        "entityType": "80"
      },
      {
        "uuid": "riCoverageOverride",
        "displayName": "RI Coverage Override",
        "value": "false",
        "entityType": "false"
      }
    ],
    "osMigrationSettingList": [],
```

```
    "subscription": {}
  },
  "displayName": "Optimize Cloud- Basic",
  "loadChanges": {
    "utilizationList": [],
    "maxUtilizationList": []
  },
  "projectionDays": [
    0
  ],
  "scope": [
    {
      "uuid": "GROUP-PMsByTargetType_AWS",
      "displayName": "PMs_AWS",
      "groupType": "PhysicalMachine",
      "className": "Group"
    }
  ],
  "topologyChanges": {
    "addList": [],
    "migrateList": [],
    "removeList": [],
    "replaceList": [],
    "relievePressureList": []
  },
  "type": "OPTIMIZE_CLOUD"
}
```

Response:

```
{
  "uuid": "2502117573632",
  "displayName": "Optimize Cloud- Basic",
  "owners": [
    {
      "uuid": "_4T_7kwY-Ed-WUKbEYSVIDw",
      "username": "administrator",
      "showSharedUserSC": false
    }
  ],
  "type": "OPTIMIZE_CLOUD",
  "scope": [
    {
      "uuid": "be95c99fc2a2195ac896204766a52d0e141818a6",
      "displayName": "PMs_AWS",
      "className": "Group",
      "groupType": "PhysicalMachine",
      "isStatic": true,
      "logicalOperator": "AND"
    }
  ],
  "projectionDays": [
    0
  ],
  "topologyChanges": {},
  "loadChanges": {},
  "configChanges": {
    "automationSettingList": [
      {
        "uuid": "resize",
        "displayName": "resize",
        "value": "true"
      }
    ],
```

```
  "riSettingList": [
    {
      "uuid": "preferredPaymentOption",
      "displayName": "Payment",
      "value": "ALL_UPFRONT",
      "defaultValue": "ALL_UPFRONT",
      "valueType": "STRING",
      "options": [
        {
          "label": "ALL_UPFRONT",
          "value": "ALL_UPFRONT"
        },
        {
          "label": "PARTIAL_UPFRONT",
          "value": "PARTIAL_UPFRONT"
        },
        {
          "label": "NO_UPFRONT",
          "value": "NO_UPFRONT"
        }
      ],
      "sourceGroupName": "Global"
    },
    {
      "uuid": "preferredTerm",
      "displayName": "Term",
      "value": "YEARS_3",
      "defaultValue": "YEARS_1",
      "valueType": "STRING",
      "options": [
        {
          "label": "YEARS_1",
          "value": "YEARS_1"
        },
        {
          "label": "YEARS_3",
          "value": "YEARS_3"
        }
      ],
      "sourceGroupName": "Global"
    },
    {
      "uuid": "preferredCoverage",
      "displayName": "Coverage",
      "value": "80",
      "defaultValue": "80",
      "valueType": "NUMERIC",
      "min": 0,
      "max": 100,
      "sourceGroupName": "Global"
    },
    {
      "uuid": "riCoverageOverride",
      "displayName": "RI Coverage Override",
      "value": "false",
      "defaultValue": "false",
      "valueType": "BOOLEAN",
      "sourceGroupName": "Global"
    },
    {
      "uuid": "preferredOfferingClass",
      "displayName": "Type",
      "value": "STANDARD",
      "defaultValue": "STANDARD",
      "valueType": "STRING",
```

```
      "options": [
        {
          "label": "STANDARD",
          "value": "STANDARD"
        },
        {
          "label": "CONVERTIBLE",
          "value": "CONVERTIBLE"
        }
      ],
      "sourceGroupName": "Global"
    }
  ]
},
"changes": [
  {
    "index": 1,
    "type": "SCOPE",
    "scope": [
      {
        "uuid": "be95c99fc2a2195ac896204766a52d0e141818a6",
        "displayName": "PMs_AWS",
        "className": "Group",
        "entitiesCount": 50,
        "groupType": "PhysicalMachine",
        "severity": "Normal",
        "logicalOperator": "AND"
      }
    ]
  },
  {
    "index": 2,
    "type": "PROJECTION_PERIODS",
    "projectionDays": [
      0
    ]
  },
  {
    "index": 3,
    "type": "ENABLED",
    "description": "Resize enabled",
    "targets": [],
    "enable": true,
    "projectionDays": [
      0
    ]
  },
  {
    "index": 4,
    "type": "RI_SETTINGS",
    "targets": [
      {
        "uuid": "preferredPaymentOption|ALL_UPFRONT"
      },
      {
        "uuid": "preferredTerm|YEARS_3"
      },
      {
        "uuid": "preferredCoverage|80"
      },
      {
        "uuid": "riCoverageOverride|false"
      },
      {
        "uuid": "preferredOfferingClass|STANDARD"
```

```
        }
      ],
      "projectionDays": [
        0
      ]
    }
  ],
  "timebasedTopologyChanges": {}
}
```

On-prem workload migration (cluster to cluster)

POST:

```
{
  "configChanges": {
    "addPolicyList": [],
    "automationSettingList": [],
    "removeConstraintList": [],
    "removePolicyList": [],
    "riSettingList": [],
    "osMigrationSettingList": [],
    "subscription": {}
  },
  "displayName": "On-Prem Workload Migration 1",
  "loadChanges": {
    "utilizationList": [],
    "maxUtilizationList": []
  },
  "projectionDays": [
    0
  ],
  "scope": [
    {
      "uuid": "3e47ff3351f3247367c00c08fdc5bb974695c443",
      "displayName": "DC23-Datacenter\\DC23-cluster1",
      "groupType": "PhysicalMachine",
      "className": "Cluster"
    }
  ],
  "topologyChanges": {
    "addList": [
      {
        "count": 1,
        "projectionDays": [
          0
        ],
        "target": {
          "uuid": "67d24a6f265151876e87f39475fb4ee4b4170098",
          "displayName": "AvailabilitySet::ARSEN-AVS-NCUS",
          "className": "VirtualMachine"
        }
      }
    ],
    "migrateList": [],
    "removeList": [],
    "replaceList": [],
    "relievePressureList": []
  },
  "type": "WORKLOAD_MIGRATION"
}
```

Response:


```
{
  "uuid": "2502063893936",
  "displayName": "On-Prem Workload Migration- Basic",
  "owners": [
    {
      "uuid": "_4T_7kwY-Ed-WUKbEYSVIDw",
      "username": "administrator",
      "showSharedUserSC": false
    }
  ],
  "type": "WORKLOAD_MIGRATION",
  "scope": [
    {
      "uuid": "_0x3OYUglEd-gHc4L513yOA",
      "displayName": "Market",
      "className": "Market"
    }
  ],
  "projectionDays": [
    0
  ],
  "topologyChanges": {
    "addList": [
      {
        "projectionDays": [
          0
        ],
        "target": {
          "uuid": "a6d365f4fbf03fb9a2e8cdb29e38045e31cee41b",
          "displayName": "Beta DC6_vm",
          "className": "Group"
        },
        "count": 1
      }
    ]
  },
  "loadChanges": {},
  "configChanges": {},
  "changes": [
    {
      "index": 1,
      "type": "SCOPE",
      "scope": [
        {
          "uuid": "_0x3OYUglEd-gHc4L513yOA",
          "displayName": "Market",
          "className": "Market"
        }
      ]
    },
    {
      "index": 2,
      "type": "PROJECTION_PERIODS",
      "projectionDays": [
        0
      ]
    },
    {
      "index": 3,
      "type": "ADDED",
      "description": "1 Beta DC6_vm",
      "targets": [
        {
```

```
          "uuid": "a6d365f4fbf03fb9a2e8cdb29e38045e31cee41b",
          "displayName": "Beta DC6_vm",
          "className": "Group",
          "entitiesCount": 56,
          "groupType": "VirtualMachine",
          "severity": "Major",
          "logicalOperator": "AND"
        }
      ],
      "value": "1",
      "projectionDays": [
        0
      ]
    }
  ],
  "timebasedTopologyChanges": {}
}
```


Hardware refresh

POST:


```
{
  "configChanges": {
    "addPolicyList": [],
    "automationSettingList": [],
    "removeConstraintList": [],
    "removePolicyList": [],
    "riSettingList": [],
    "osMigrationSettingList": [],
    "subscription": {}
  },
  "displayName": "Hardware Refresh- Basic",
  "loadChanges": {
    "utilizationList": [],
    "maxUtilizationList": []
  },
  "projectionDays": [
    0
  ],
  "scope": [
    {
      "uuid": "Market",
      "displayName": "Global Environment",
      "className": "Market"
    }
  ],
  "topologyChanges": {
    "addList": [],
    "migrateList": [],
    "removeList": [],
    "replaceList": [
      {
        "projectionDay": 0,
        "target": {
          "uuid": "4C4C4544-0042-5A10-804A-B1C04F384331",
          "displayName": "dell-pe1944.corp.vmturbo.com",
          "className": "PhysicalMachine",
          "severity": "Critical"
        },
        "template": {
          "uuid": "_juCtwJkfEeCX4rLBy_KD2g",
          "displayName": "HP_DL580 G7",
```

```
          "className": "PhysicalMachineProfile"
        }
      },
      {
        "projectionDay": 0,
        "target": {
          "uuid": "4C4C4544-004A-3610-8057-C8C04F464331",
          "displayName": "dell-pe1953.corp.vmturbo.com",
          "className": "PhysicalMachine",
          "severity": "Critical"
        },
        "template": {
          "uuid": "_juCtwJkfEeCX4rLBy_KD2g",
          "displayName": "HP_DL580 G7",
          "className": "PhysicalMachineProfile"
        }
      },
      {
        "projectionDay": 0,
        "target": {
          "uuid": "4C4C4544-0059-4D10-8030-B3C04F464331",
          "displayName": "dell-pe1954.corp.vmturbo.com",
          "className": "PhysicalMachine",
          "severity": "Critical"
        },
        "template": {
          "uuid": "_juCtwJkfEeCX4rLBy_KD2g",
          "displayName": "HP_DL580 G7",
          "className": "PhysicalMachineProfile"
        }
      },
      {
        "projectionDay": 0,
        "target": {
          "uuid": "Virtual_ESX_4238e933-3923-8334-37fe-3aec8fbd046f",
          "displayName": "hp-esx22.dev.mycorp.com",
          "className": "PhysicalMachine",
          "severity": "Critical"
        },
        "template": {
          "uuid": "_juCtwJkfEeCX4rLBy_KD2g",
          "displayName": "HP_DL580 G7",
          "className": "PhysicalMachineProfile"
        }
      }
    ],
    "relievePressureList": []
  },
  "type": "RECONFIGURE_HARDWARE"
}
```

Response:

```
{
  "uuid": "2502070918448",
  "displayName": "Hardware Refresh- Basic",
  "owners": [
    {
      "uuid": "_4T_7kwY-Ed-WUKbEYSVIDw",
      "username": "administrator",
      "showSharedUserSC": false
    }
  ],
  "type": "RECONFIGURE_HARDWARE",
```

```
"scope": [
  {
    "uuid": "_0x3OYUglEd-gHc4L513yOA",
    "displayName": "Market",
    "className": "Market"
  }
],
"projectionDays": [
  0
],
"topologyChanges": {
  "replaceList": [
    {
      "projectionDay": 0,
      "target": {
        "uuid": "4C4C4544-0042-5A10-804A-B1C04F384331",
        "displayName": "dell-pe1944.corp.vmturbo.com",
        "className": "PhysicalMachine"
      },
      "template": {
        "uuid": "_juCtwJkfEeCX4rLBy_KD2g",
        "displayName": "HP_DL580 G7",
        "className": "PhysicalMachineProfile"
      }
    },
    {
      "projectionDay": 0,
      "target": {
        "uuid": "4C4C4544-004A-3610-8057-C8C04F464331",
        "displayName": "dell-pe1953.corp.vmturbo.com",
        "className": "PhysicalMachine"
      },
      "template": {
        "uuid": "_juCtwJkfEeCX4rLBy_KD2g",
        "displayName": "HP_DL580 G7",
        "className": "PhysicalMachineProfile"
      }
    },
    {
      "projectionDay": 0,
      "target": {
        "uuid": "4C4C4544-0059-4D10-8030-B3C04F464331",
        "displayName": "dell-pe1954.corp.vmturbo.com",
        "className": "PhysicalMachine"
      },
      "template": {
        "uuid": "_juCtwJkfEeCX4rLBy_KD2g",
        "displayName": "HP_DL580 G7",
        "className": "PhysicalMachineProfile"
      }
    },
    {
      "projectionDay": 0,
      "target": {
        "uuid": "Virtual_ESX_4238e933-3923-8334-37fe-3aec8fbd046f",
        "displayName": "hp-esx22.dev.mycorp.com",
        "className": "PhysicalMachine"
      },
      "template": {
        "uuid": "_juCtwJkfEeCX4rLBy_KD2g",
        "displayName": "HP_DL580 G7",
        "className": "PhysicalMachineProfile"
      }
    }
  ]
```

```
 },
 "loadChanges": {},
 "configChanges": {},
 "changes": [
   {
     "index": 1,
     "type": "SCOPE",
     "scope": [
       {
         "uuid": "_0x3OYUglEd-gHc4L513yOA",
         "displayName": "Market",
         "className": "Market"
       }
     ]
   },
   {
     "index": 2,
     "type": "PROJECTION_PERIODS",
     "projectionDays": [
       0
     ]
   },
   {
     "index": 3,
     "type": "REPLACED",
     "description": "Replace [dell-pe1944.corp.vmturbo.com, dell-pe1953.corp.vmturbo.com,
dell-pe1954.corp.vmturbo.com, hp-esx... with HP_DL580 G7",
     "targets": [
       {
         "uuid": "4C4C4544-0042-5A10-804A-B1C04F384331",
         "displayName": "dell-pe1944.corp.vmturbo.com",
         "className": "PhysicalMachine",
         "severity": "Critical"
       },
       {
         "uuid": "4C4C4544-004A-3610-8057-C8C04F464331",
         "displayName": "dell-pe1953.corp.vmturbo.com",
         "className": "PhysicalMachine",
         "severity": "Critical"
       },
       {
         "uuid": "4C4C4544-0059-4D10-8030-B3C04F464331",
         "displayName": "dell-pe1954.corp.vmturbo.com",
         "className": "PhysicalMachine",
         "severity": "Critical"
       },
       {
         "uuid": "Virtual_ESX_4238e933-3923-8334-37fe-3aec8fbd046f",
         "displayName": "hp-esx22.dev.mycorp.com",
         "className": "PhysicalMachine",
         "severity": "Critical"
       },
       {
         "uuid": "_juCtwJkfEeCX4rLBy_KD2g",
         "displayName": "HP_DL580 G7",
         "className": "PhysicalMachineProfile"
       }
     ],
     "value": "0",
     "projectionDays": [
       0
     ]
   }
 ],
 "timebasedTopologyChanges": {}
```

```
}
```

Alleviate pressure

POST:

```
{
  "configChanges": {
    "addPolicyList": [],
    "automationSettingList": [],
    "removeConstraintList": [],
    "removePolicyList": [],
    "riSettingList": [],
    "osMigrationSettingList": [],
    "subscription": {}
  },
  "displayName": "Alleviate Pressure- Basic",
  "loadChanges": {
    "utilizationList": [],
    "maxUtilizationList": []
  },
  "projectionDays": [
    0
  ],
  "scope": [
    {
      "uuid": "11c63ebd77647e7b9c9fb85dd2d4c77464cbeac2",
      "displayName": "DC13\\DC13-Cluster",
      "className": "Cluster"
    },
    {
      "uuid": "57a4026a69ba4b46ebc430cb62d5c55eeea00695",
      "displayName": "DC13\\Physical Hosts_Cluster2 - Intel",
      "className": "Cluster"
    }
  ],
  "topologyChanges": {
    "addList": [],
    "migrateList": [],
    "removeList": [],
    "replaceList": [],
    "relievePressureList": [
      {
        "projectionDay": 0,
        "sources": [
          {
            "uuid": "11c63ebd77647e7b9c9fb85dd2d4c77464cbeac2",
            "displayName": "DC13\\DC13-Cluster",
            "className": "Cluster"
          }
        ],
        "destinations": [
          {
            "uuid": "57a4026a69ba4b46ebc430cb62d5c55eeea00695",
            "displayName": "DC13\\Physical Hosts_Cluster2 - Intel",
            "className": "Cluster"
          }
        ]
      }
    ]
  },
  "type": "ALLEVIATE_PRESSURE"
```

```
}
```

Response:


```
{
  "uuid": "2531404249984",
  "displayName": "On-Prem Workload Migration 1",
  "owners": [
    {
      "uuid": "_4T_7kwY-Ed-WUKbEYSVIDw",
      "username": "administrator",
      "showSharedUserSC": false
    }
  ],
  "type": "WORKLOAD_MIGRATION",
  "scope": [
    {
      "uuid": "3e47ff3351f3247367c00c08fdc5bb974695c443",
      "displayName": "DC23-Datacenter\\DC23-cluster1",
      "className": "Cluster",
      "groupType": "PhysicalMachine",
      "isStatic": true,
      "logicalOperator": "AND"
    }
  ],
  "projectionDays": [
    0
  ],
  "topologyChanges": {
    "addList": [
      {
        "projectionDays": [
          0
        ],
        "target": {
          "uuid": "67d24a6f265151876e87f39475fb4ee4b4170098",
          "displayName": "AvailabilitySet::ARSEN-AVS-NCUS",
          "className": "DiscoveredGroup"
        },
        "count": 1
      }
    ]
  },
  "loadChanges": {},
  "configChanges": {},
  "changes": [
    {
      "index": 1,
      "type": "SCOPE",
      "scope": [
        {
          "uuid": "3e47ff3351f3247367c00c08fdc5bb974695c443",
          "displayName": "DC23-Datacenter\\DC23-cluster1",
          "className": "Cluster",
          "entitiesCount": 3,
          "groupType": "PhysicalMachine",
          "severity": "Critical",
          "logicalOperator": "AND"
        }
      ]
    },
    {
      "index": 2,
      "type": "PROJECTION_PERIODS",
```

```
      "projectionDays": [
        0
      ]
    },
    {
      "index": 3,
      "type": "ADDED",
      "description": "1 AvailabilitySet::ARSEN-AVS-NCUS",
      "targets": [
        {
          "uuid": "67d24a6f265151876e87f39475fb4ee4b4170098",
          "displayName": "AvailabilitySet::ARSEN-AVS-NCUS",
          "className": "DiscoveredGroup",
          "entitiesCount": 3,
          "groupType": "VirtualMachine",
          "severity": "Critical",
          "logicalOperator": "AND"
        }
      ],
      "value": "1",
      "projectionDays": [
        0
      ]
    }
  ],
  "timebasedTopologyChanges": {}
}
```


Optimize current environment

The result of a plan run with no custom configuration is an optimization of your current environment, respecting all current policies and constraints.

POST:


```
{
  "configChanges": {
    "addPolicyList": [],
    "automationSettingList": [],
    "removeConstraintList": [],
    "removePolicyList": [],
    "riSettingList": [],
    "osMigrationSettingList": [],
    "subscription": {}
  },
  "displayName": "Custom- No Config",
  "loadChanges": {
    "utilizationList": [],
    "maxUtilizationList": []
  },
  "projectionDays": [],
  "scope": [
    {
      "uuid": "Market",
      "displayName": "Global Environment",
      "className": "Market"
    }
  ],
  "topologyChanges": {
    "addList": [],
    "migrateList": [],
    "removeList": [],
    "replaceList": [],
```

```
    "relievePressureList": []
  },
  "type": "CUSTOM"
}
```

Response:

```
{
  "uuid": "2502119526800",
  "displayName": "Custom- No Config",
  "owners": [
    {
      "uuid": "_4T_7kwY-Ed-WUKbEYSVIDw",
      "username": "administrator",
      "showSharedUserSC": false
    }
  ],
  "type": "CUSTOM",
  "scope": [
    {
      "uuid": "_0x3OYUglEd-gHc4L513yOA",
      "displayName": "Market",
      "className": "Market"
    }
  ],
  "changes": [
    {
      "index": 1,
      "type": "SCOPE",
      "scope": [
        {
          "uuid": "_0x3OYUglEd-gHc4L513yOA",
          "displayName": "Market",
          "className": "Market"
        }
      ]
    }
  ]
}
```

Scenario future load configuration example

Future load plans

Future load plans are a type of plan in which commodities from active VMs are taken and applied to passive or low-workload VMs, simulating the effect on the environment if these passive VMs became active. For example, you may want to investigate the effect on your environment if your disaster recovery VMs were forced to run at historical peak values. In this case, the production VMs whose commodities will be used are the active VMs, and the currently idle disaster recovery VMs are the passive VMs.

Currently, there are two future load plan types:

- **PEAKOFPEAKS**: The peak-of-peaks future load plan replicates in the passive VMs every commodity of the active VMs reaching its historical peak utilization simultaneously, and the plan result will reflect the effect on your environment.
- **CUSTOM**: The custom future load plan replicates in the passive VMs every commodity of the active VMs reaching its current utilization simultaneously, and the plan result will reflect the effect on your environment.

The setup and configuration for both future load plans differ only in the content of the `type` parameter- all other steps will be identical.

The settings you make in a scenario correspond to the plan settings you can make in the user interface. These include:

- Plan scope
- Changes to workload (adding, removing, or replacing VMs or containers)
- Changes to supply (adding, removing, replacing PMs or storage)
- Enable/disable placement policies and other constraints
- Changes to action modes
- Enable/disable provisioning of supply
- Enable/disable resizing of workloads

Future load prerequisites

In order to run a future load plan, you must retrieve statistics for the entities to be included in the plan. Typically, this is done utilizing the `POST /markets/market_uuid/entities/stats` request, with `startDate` and `scopes` parameters.

Example input for statistic retrieval:

```
{
  "period": {
    "startDate": 1237056032
  },
  "scopes": [
    "4223ae0c-c99b-f6ca-c1ef-3f367bdf5528",
    "4223bb37-0090-f332-0140-cc3cce7b3735",
    "4223dc55-78f2-2cd7-9abf-b09b1be835b4",
    "42231932-83af-10d9-f0cd-26c8a08a9e18",
    "422374a3-2c64-b613-a9a7-1f900a2314b4"
  ]
}
```

Once the call is completed, copy the response in its entirety- or if using Swagger, click the Download button next to the response and copy the contents of the file.

Running the future load plan

After creating the ScenarioApiDTO, the response will include the uuid of the created scenario. Use this uuid along with the uuid of the market to run the plan against (typically the real-time market) in the `POST /markets/market_uuid/scenarios/scenario_uuid` request.

Viewing the peak of peak plan

After running the plan, the response will include the uuid of the created plan market, which you should note. Results can be viewed using the `GET /markets/market_uuid` request.

Schedules endpoint

Schedules specify a period of time during which certain events can occur. Workload Optimization Manager uses the following types of schedules:

- Calendar Schedule

  A calendar schedule is a setting that sets up a window of time when an *automation policy* takes effect. This policy can execute *non-parking* actions on public cloud or on-prem entities, or change settings that affect analysis and action generation.

To configure calendar schedules, you must use `/schedules` requests. With these requests, you can:

- Get calendar schedules
- Create calendar schedules
- Edit or delete calendar schedules
- View all policies using a specific calendar schedule

NOTE: When you configure a schedule window for a resize action, to ensure Workload Optimization Manager will execute the action during the scheduled time, you must turn off the Enforce Non Disruptive Mode setting for that scheduled policy. Even if you turn the setting off for the global policy, you still must turn the setting off for your scheduled policy. Otherwise Workload Optimization Manager will not execute the resize action.

- Parking Schedule

  A parking schedule is a setting that sets up a window of time when *parking* actions (stop or start) are enforced on entities in your public cloud environment.

  To configure parking schedules, you must use `/schedules/timespans/` requests. With these requests, you can:

  - Get parking schedules
  - Create parking schedules
  - Attach parking schedules to parkable entities
  - Detach parking schedules from parkable entities
  - Edit or delete parking schedules

Calendar schedules requests

The requests listed in this topic only apply to calendar schedules used in automation policies. For details about parking schedules that can be attached to parkable entities, see Parking Schedules Requests *(on page 1345)*.

Getting calendar schedules

Gets a list of schedules. For a list of schedules, the API returns an array of ScheduleApiDTOs. Each ScheduleApiDTO will contain details about the schedule, including the start and end time and date and recurrence. To get a single schedule, include the UUID of the notification in the request.

Examples:

- **All Schedules:** `GET https://10.10.10.10/api/v3/schedules`
- **Single Schedule:** `GET https://10.10.10.10/api/v3/schedules/284466929277824`

Response:


```
[
  {
    "uuid": "284466927892512",
    "displayName": "DalSched1",
    "startTime": "2020-02-29T16:30",
    "endTime": "2020-02-29T19:30",
    "recurrence": {
      "type": "WEEKLY",
      "daysOfWeek": [
        "Mon",
        "Fri"
      ],
      "interval": 1
    },
    "timeZone": "America/Denver",
    "nextOccurrence": "2020-03-02T16:30:00",
```

```
    "nextOccurrenceTimestamp": 1583191800000
  },
  {
    "uuid": "284466929277824",
    "displayName": "DalSched2",
    "startTime": "2020-02-29T16:30",
    "endTime": "2020-02-29T17:45",
    "recurrence": {
      "type": "MONTHLY",
      "daysOfMonth": [
        5
      ],
      "interval": 3
    },
    "timeZone": "America/Denver",
    "nextOccurrence": "2020-05-05T16:30:00",
    "nextOccurrenceTimestamp": 1588717800000
  }
]
```

Adding a calendar schedule

In order to create a schedule in Workload Optimization Manager, you must provide a valid ScheduleApiDTO.

Example: `POST https://10.10.10.10/api/v3/schedules`

Example input, showing a schedule to be run weekly on Friday between the hours of 0000 and 0400 that starts on 2019-12-20 and has no end date:

```
{
  "displayName":"DallasSched",
  "endTime":"2019-12-20T04:00",
  "startTime":"2019-12-20T00:00",
  "timeZone":"America/Denver",
  "recurrence":{
    "type":"WEEKLY",
    "interval":1,
    "daysOfWeek":[
      "Fri"
    ]
  }
}
```

Getting policies using a calendar schedule

Given a schedule ID, retrieve any policies associated with that schedule.

Example: `GET https://10.10.10.10/api/v3/schedules/284466929277824/settingsPolicy`

Response: An array of `SettingspolicyApiDTO`s that represent the policies using the specified schedule.

```
[
  {
    "uuid": "284466929430656",
    "displayName": "DalStorage",
    "entityType": "Storage",
    "scopes": [
      {
```

```
        "uuid": "284460958528132",
        "displayName": "HawthorneHC_datastore",
        "isStatic": true,
        "logicalOperator": "AND"
      }
    ],
    "settingsManagers": [
      {
        "uuid": "automationmanager",
        "displayName": "Action Mode Settings",
        "category": "Automation",
        "settings": [
          {
            "uuid": "suspend",
            "displayName": "Suspend",
            "value": "MANUAL",
            "defaultValue": "MANUAL",
            "valueType": "STRING",
            "valueObjectType": "String",
            "options": [
              {
                "label": "Disabled",
                "value": "DISABLED"
              },
              {
                "label": "Recommend",
                "value": "RECOMMEND"
              },
              {
                "label": "Manual",
                "value": "MANUAL"
              },
              {
                "label": "Automatic",
                "value": "AUTOMATIC"
              }
            ],
            "entityType": "Storage"
          }
        ]
      }
    ],
    "schedule": {
      "uuid": "284466929277824",
      "displayName": "DalSched2",
      "startTime": "2020-02-29T16:30",
      "endTime": "2020-02-29T17:45",
      "recurrence": {
        "type": "MONTHLY",
        "daysOfMonth": [
          5
        ],
        "interval": 3
      },
      "timeZone": "America/Denver",
      "nextOccurrence": "2020-05-05T16:30:00",
      "nextOccurrenceTimestamp": 1588717800000
    },
    "disabled": false,
    "readOnly": false,
    "default": false
  }
]
```

Calendar schedules tips

When creating schedules, you can set multiple timing criteria. The following examples show schedules with advanced timing. For basic schedules, see the Schedules Endpoint *(on page 1338)*.

Bi-weekly calendar schedule with specific days

This schedule is set to run biweekly, indicated by the `"type": "WEEKLY"` and `"interval": 2` fields. As indicated by the `daysOfWeek` array, this schedule is active on Monday, Wednesday, and Friday.

InputDTO:

```
{
  "displayName":"Schedule1",
  "endDate":"2020-07-10",
  "endTime":"2020-05-06T19:30",
  "startTime":"2020-05-06T15:30",
  "timeZone":"America/Denver",
  "recurrence":{
    "type":"WEEKLY",
    "interval":2,
    "daysOfWeek":[
      "Mon",
      "Wed",
      "Fri"
    ]
  }
}
```

The API response for this schedule:

```
{
  "uuid":"284559431391120",
  "displayName":"DallasTest",
  "endDate":"2020-07-10",
  "startTime":"2020-05-06T15:30",
  "endTime":"2020-05-06T19:30",
  "recurrence":{
    "type":"WEEKLY",
    "daysOfWeek":[
      "Mon",
      "Wed",
      "Fri"
    ],
    "interval":2
  },
  "timeZone":"America/Denver",
  "nextOccurrence":"2020-05-06T15:30:00",
  "nextOccurrenceTimestamp":1588800600000
}
```

Six-month calendar schedule with specific week of the month

This schedule is set to run every six months, indicated by the `"type": "MONTHLY"` and `"interval": 6` fields. As indicated by the `daysOfWeek` array this schedule is active on Wednesday, and the `"weekOfTheMonth": 2` indicates that this schedule will only run on the second week of the month.

This schedule was created on May 5th, 2020. Converting the timestamp in the API Response to a human-readable date, you see that the next run is on Wednesday, May 13, 2020 3:30:00 PM, the second Wednesday in May.

Input DTO:

```
{
  "displayName":"DallasTest2",
  "endDate":"2020-10-19",
  "endTime":"2020-05-06T19:15",
  "startTime":"2020-05-06T15:30",
  "timeZone":"America/Denver",
  "recurrence":{
    "type":"MONTHLY",
    "interval":6,
    "daysOfWeek":[
      "Wed"
    ],
    "weekOfTheMonth":[
      2
    ]
  }
}
```

The API response for this schedule:

```
{
  "uuid":"284559434037456",
  "displayName":"DallasTest2",
  "endDate":"2020-10-19",
  "startTime":"2020-05-06T15:30",
  "endTime":"2020-05-06T19:15",
  "recurrence":{
    "type":"MONTHLY",
    "daysOfWeek":[
      "Wed"
    ],
    "weekOfTheMonth":[
      2
    ],
    "interval":6
  },
  "timeZone":"America/Denver",
  "nextOccurrence":"2020-05-13T15:30:00",
  "nextOccurrenceTimestamp":1589405400000
}
```

Getting policies attached to a specific calendar schedule

In order to see what policies are using a schedule, use the `https://10.10.10.10/api/v3/schedules/{schedule_UUID}/settingsPolicy` request.

Response: An array of `SettingspolicyApiDTO`s that represent the policies using the specified schedule, whose details are also included.

```
[
  {
    "uuid": "284466929430656",
    "displayName": "DalStorage",
```

```
  "entityType": "Storage",
  "scopes": [
    {
      "uuid": "284460958528132",
      "displayName": "HawthorneHC_datastore",
      "isStatic": true,
      "logicalOperator": "AND"
    }
  ],
  "settingsManagers": [
    {
      "uuid": "automationmanager",
      "displayName": "Action Mode Settings",
      "category": "Automation",
      "settings": [
        {
          "uuid": "suspend",
          "displayName": "Suspend",
          "value": "MANUAL",
          "defaultValue": "MANUAL",
          "valueType": "STRING",
          "valueObjectType": "String",
          "options": [
            {
              "label": "Disabled",
              "value": "DISABLED"
            },
            {
              "label": "Recommend",
              "value": "RECOMMEND"
            },
            {
              "label": "Manual",
              "value": "MANUAL"
            },
            {
              "label": "Automatic",
              "value": "AUTOMATIC"
            }
          ],
          "entityType": "Storage"
        }
      ]
    }
  ],
  "schedule": {
    "uuid": "284466929277824",
    "displayName": "DalSched2",
    "startTime": "2020-02-29T16:30",
    "endTime": "2020-02-29T17:45",
    "recurrence": {
      "type": "MONTHLY",
      "daysOfMonth": [
        5
      ],
      "interval": 3
    },
    "timeZone": "America/Denver",
    "nextOccurrence": "2020-05-05T16:30:00",
    "nextOccurrenceTimestamp": 1588717800000
  },
  "disabled": false,
  "readOnly": false,
  "default": false
}
```

```
]
```

Parking schedules requests

The requests listed in this topic currently only apply to parking schedules and parkable entities. For details about calendar schedules used in automation policies, see Calendar Schedules Requests *(on page 1339)*.

Getting parking schedules

Gets a list of parking schedules. For a list of schedules, the API returns an array of parking schedule metadata. To get a single parking schedule, include the parking schedule UUID in the URL.

Examples:

- **All Schedules:** `GET https://10.10.10.10/api/v3/schedules/timespans`
- **Single Schedule:** `GET https://10.10.10.10/api/v3/schedules/timespans/778668020368144`

All Schedules Response:

```
[
  {
    "uuid": "778601653004480",
    "displayName": "testing",
    "timeZone": "America/New_York"
  },
  {
    "uuid": "778603154693344",
    "displayName": "everyday of the week",
    "timeZone": "America/New_York"
  },
  {
    "uuid": "778668020368144",
    "displayName": "Sample Schedule",
    "description": "This is a sample schedule where entities are active during normal business hours but
 may undergo maintenance in the early hours of Sunday.",
    "timeZone": "America/Chicago"
  }
]
```

Single Schedule Response:

```
{
  "uuid": "778668020368144",
  "displayName": "Sample Schedule",
  "description": "This is a sample schedule where entities are active during normal business hours but may
 undergo maintenance in the early hours of Sunday.",
  "timeZone": "America/Chicago",
  "timeSpans": {
    "type": "WeekDayTimeSpansApiDTO",
    "sunday": [
      {
        "begins": "00:00",
        "ends": "01:59",
        "policy": {
          "type": "ParkingItemApiDTO",
          "state": "OFF"
        }
      },
      {
        "begins": "02:00",
        "ends": "03:59",
```

```
        "policy": {
          "type": "ParkingItemApiDTO",
          "state": "IGNORE"
        }
      },
      {
        "begins": "04:00",
        "ends": "23:59",
        "policy": {
          "type": "ParkingItemApiDTO",
          "state": "OFF"
        }
      }
    ],
    "monday": [
      {
        "begins": "00:00",
        "ends": "06:59",
        "policy": {
          "type": "ParkingItemApiDTO",
          "state": "OFF"
        }
      },
      {
        "begins": "07:00",
        "ends": "16:59",
        "policy": {
          "type": "ParkingItemApiDTO",
          "state": "ON"
        }
      },
      {
        "begins": "17:00",
        "ends": "23:59",
        "policy": {
          "type": "ParkingItemApiDTO",
          "state": "OFF"
        }
      }
    ],
    "tuesday": [
      {
        "begins": "00:00",
        "ends": "06:59",
        "policy": {
          "type": "ParkingItemApiDTO",
          "state": "OFF"
        }
      },
      {
        "begins": "07:00",
        "ends": "16:59",
        "policy": {
          "type": "ParkingItemApiDTO",
          "state": "ON"
        }
      },
      {
        "begins": "17:00",
        "ends": "23:59",
        "policy": {
          "type": "ParkingItemApiDTO",
          "state": "OFF"
        }
      }
```

```
  ],
  "wednesday": [
    {
      "begins": "00:00",
      "ends": "06:59",
      "policy": {
        "type": "ParkingItemApiDTO",
        "state": "OFF"
      }
    },
    {
      "begins": "07:00",
      "ends": "16:59",
      "policy": {
        "type": "ParkingItemApiDTO",
        "state": "ON"
      }
    },
    {
      "begins": "17:00",
      "ends": "23:59",
      "policy": {
        "type": "ParkingItemApiDTO",
        "state": "OFF"
      }
    }
  ],
  "thursday": [
    {
      "begins": "00:00",
      "ends": "06:59",
      "policy": {
        "type": "ParkingItemApiDTO",
        "state": "OFF"
      }
    },
    {
      "begins": "07:00",
      "ends": "16:59",
      "policy": {
        "type": "ParkingItemApiDTO",
        "state": "ON"
      }
    },
    {
      "begins": "17:00",
      "ends": "23:59",
      "policy": {
        "type": "ParkingItemApiDTO",
        "state": "OFF"
      }
    }
  ],
  "friday": [
    {
      "begins": "00:00",
      "ends": "06:59",
      "policy": {
        "type": "ParkingItemApiDTO",
        "state": "OFF"
      }
    },
    {
      "begins": "07:00",
      "ends": "16:59",
```

```
        "policy": {
          "type": "ParkingItemApiDTO",
          "state": "ON"
        }
      },
      {
        "begins": "17:00",
        "ends": "23:59",
        "policy": {
          "type": "ParkingItemApiDTO",
          "state": "OFF"
        }
      }
    ],
    "saturday": [
      {
        "begins": "00:00",
        "ends": "23:59",
        "policy": {
          "type": "ParkingItemApiDTO",
          "state": "OFF"
        }
      }
    ]
  }
}
```

Adding a parking schedule

To create a parking schedule, you must provide a valid ParkingItemApiDTO.

Example: `POST https://10.10.10.10/api/v3/schedules/timespans`

Example input, showing a schedule that causes entities to stay powered on Monday through Friday from 7:00 AM to 5:00 PM. Entities are off all other times except for a two-hour window on Sunday morning between 2:00 AM and 4:00 AM where the schedule does not enforce a powered state:

```
{
  "displayName": "Sample Schedule",
  "timeZone": "America/Chicago",
  "description": "This is a sample schedule where entities are active during normal business hours but may
 undergo maintenance in the early hours of Sunday.",
  "timeSpans": {
    "type": "WeekDayTimeSpansApiDTO",
    "monday": [
      {
        "policy": {"type": "ParkingItemApiDTO", "state": "OFF"},
        "begins": "00:00",
        "ends": "06:59"
      },
      {
        "policy": {"type": "ParkingItemApiDTO", "state": "ON"},
        "begins": "07:00",
        "ends": "16:59"
      },
      {
        "policy": {"type": "ParkingItemApiDTO", "state": "OFF"},
        "begins": "17:00",
        "ends": "23:59"
      }
    ],
    "tuesday": [
      {
        "policy": {"type": "ParkingItemApiDTO", "state": "OFF"},
```

```
      "begins": "00:00",
      "ends": "06:59"
    },
    {
      "policy": {"type": "ParkingItemApiDTO", "state": "ON"},
      "begins": "07:00",
      "ends": "16:59"
    },
    {
      "policy": {"type": "ParkingItemApiDTO", "state": "OFF"},
      "begins": "17:00",
      "ends": "23:59"
    }
  ],
  "wednesday": [
    {
      "policy": {"type": "ParkingItemApiDTO", "state": "OFF"},
      "begins": "00:00",
      "ends": "06:59"
    },
    {
      "policy": {"type": "ParkingItemApiDTO", "state": "ON"},
      "begins": "07:00",
      "ends": "16:59"
    },
    {
      "policy": {"type": "ParkingItemApiDTO", "state": "OFF"},
      "begins": "17:00",
      "ends": "23:59"
    }
  ],
  "thursday": [
    {
      "policy": {"type": "ParkingItemApiDTO", "state": "OFF"},
      "begins": "00:00",
      "ends": "06:59"
    },
    {
      "policy": {"type": "ParkingItemApiDTO", "state": "ON"},
      "begins": "07:00",
      "ends": "16:59"
    },
    {
      "policy": {"type": "ParkingItemApiDTO", "state": "OFF"},
      "begins": "17:00",
      "ends": "23:59"
    }
  ],
  "friday": [
    {
      "policy": {"type": "ParkingItemApiDTO", "state": "OFF"},
      "begins": "00:00",
      "ends": "06:59"
    },
    {
      "policy": {"type": "ParkingItemApiDTO", "state": "ON"},
      "begins": "07:00",
      "ends": "16:59"
    },
    {
      "policy": {"type": "ParkingItemApiDTO", "state": "OFF"},
      "begins": "17:00",
      "ends": "23:59"
    }
  ],
```

```
    "saturday": [
      {
        "policy": {"type": "ParkingItemApiDTO", "state": "OFF"},
        "begins": "00:00",
        "ends": "23:59"
      }
    ],
    "sunday": [
      {
        "policy": {"type": "ParkingItemApiDTO", "state": "OFF"},
        "begins": "00:00",
        "ends": "01:59"
      },
      {
        "policy": {"type": "ParkingItemApiDTO", "state": "IGNORE"},
        "begins": "02:00",
        "ends": "03:59"
      },
      {
        "policy": {"type": "ParkingItemApiDTO", "state": "OFF"},
        "begins": "04:00",
        "ends": "23:59"
      }
    ]
  }
}
```

Constraints:

- `displayName` value must be unique.
- `timeZone` value must be consistent with the IANA Timezone Database (for example, UTC or America/New_York).
- `begins` and `ends` values are HH:MM representations of a 24-hour clock and are considered inclusive.
- If time spans (or days of the week) are omitted, the system behaves the same as if a corresponding time span with a policy state of `IGNORE` exists.
- Time spans cannot overlap.

Attaching parking schedules to entities

Once a parking schedule is created and your parkable entities are discovered, you can attach a schedule to one or more entities.

Example: `POST https://10.10.10.10/api/v3/schedules/timespans/778668074428784/entities`

Constraints:

- Parking schedule only.
- Entity must be a `VirtualMachine` discovered from cloud targets (AWS, Azure, or GCP), or a `DatabaseServer` (AWS RDS instance).

Example input that attaches an existing parking schedule to 1-N entities:

```
{"entityUuids":["74913973295604","74913973295605","74913973295606"]}
```

Detaching parking schedules

You can detach a parking schedule from one or more entities.

Example: `DELETE https://10.10.10.10/api/v3/schedules/timespans/778668074428784/entities`

Example input that detaches an existing parking schedule from a 1-N entity.

```
{"entityUuids":["74913973295604","74913973295606"]}
```

Constraints: Parking schedule only.

Updating a parking schedule

Update an existing parking schedule. The following example shows a request to update a schedule to enforce a powered off state from midnight Monday to midnight Tuesday. The schedule does not enforce a state any other time.

Example: `PUT https://10.10.10.10/api/v3/schedules/timespans/778668020368144`

Response: An array of `SettingspolicyApiDTO`s that represent the policies using the specified schedule.

```
{
  "timeZone": "America/Chicago",
  "displayName": "Sample Schedule - Updated",
  "description": "Do nothing all week, except on Monday where the attached VMs should be powered off all
 day",
  "timeSpans": {
    "type": "WeekDayTimeSpansApiDTO",
    "monday": [
      {
        "policy": {
          "state": "OFF",
          "type": "ParkingItemApiDTO"
        },
        "begins": "00:00",
        "ends": "23:59"
      }
    ]
  }
}
```

Constraints:

- `displayName` value must be unique.
- `timeZone` value must be consistent with the IANA Timezone Database (for example, UTC or America/New_York).
- `begins` and `ends` values are HH:MM representations of a 24-hour clock and are considered inclusive.
- If time spans (or days of the week) are omitted, the system behaves the same as if a corresponding time span with a policy state of `IGNORE` exists.
- Time spans may not overlap.

Deleting a parking schedule

Remove an unattached parking schedule.

Example: `DELETE https://10.10.10.10/api/v3/schedules/timespans/778668020368144`

Constraints: Parking schedules to be deleted must not be attached to any entity.

Search endpoint

A search is a query performed on the entities that participate in the Workload Optimization Manager market. The most common use of search is to locate a particular object, service entity, or group of entities that share a characteristic.

`Object` and `Entity` refer to different items in your Workload Optimization Manager topology:

Using the search endpoint of the Workload Optimization Manager API, you can:

- Obtain criteria for use in detailed searching or building groups
- Get detailed information about any object in your Workload Optimization Manager environment

Search utility requests

Utility requests provide metadata that you can use to make related requests, assemble inputDTOs, or see the potential values of a particular DTO.

The search endpoint contains a utility request that shows the various criteria that you may use to search, for each entity type in the Workload Optimization Manager supply chain, and a second that shows the options for a specified criterion. You can then use this information to build your search criteria for more detailed groups.

The available filterTypes depend on the entity making up your search:

- ApplicationComponentSpec
  - appComponentSpecsByName
  - appComponentSpecsByBusinessAccountUuid
  - appComponentSpecsByHybridConnectionCount
  - appComponentSpecsByDeploymentSlotCount
- ApplicationComponent
  - appCompsByName
  - applicationComponentByBusinessApplication
  - applicationComponentByBusinessTransaction
  - applicationComponentByService
- AvailabilityZone
  - zonsByName
  - zoneByCloudProvider
- BillingFamily
  - billingFamilyByName
- BusinessAccountFolder
  - businessAccountFolderByName
- BusinessAccount
  - businessAccountByName
  - businessAccountByUuid
  - businessAccountByVendorId
  - businessAccountCloudProvider
  - businessAccountProbeType
  - businessAccountTargetName
  - businessAccountValidationStatus
  - businessAccountByResourceGroup
  - businessAccountByTag
  - subBusinessAccountOfUuid
- BusinessApplication
  - busAppsByName
  - businessApplicationByUserDefinedEntity

- BusinessTransaction
  - busTransByName
  - businessTransactionByBusinessApplication
  - businessTransactionByUserDefinedEntity
- BusinessUser
  - businessUserByName
  - businessUserByDesktopPool
  - businessUserByViewPod
- Chassis
  - chasByName
- CloudService
- Cluster
  - clustersByName
  - clustersByTag
- ComputeTier
  - computeTiersByName
  - computeTiersByMem
  - computeTiersByNumCPUs
  - computeTiersByCloudProvider
- ContainerPlatformCluster
  - containerPlatformClustersByName
- Container
  - containersByName
  - containersByPodName
  - containersByNamespace
  - containersByApplicationComponent
  - containersByBusinessApplication
  - containersByBusinessTransaction
  - containersByService
  - containersByControllerKind
  - containersByTag
  - containersByContainerPlatformCluster
- ContainerPod
  - containerPodsByName
  - containerPodsByVMName
  - containerPodsByNamespace
  - containerPodsByTag
  - containerPodsByControllerKind
  - containerPodsByWorkloadControllerName
  - containerPodsByContainerPlatformCluster
  - containerPodsByState

- ContainerSpec
  - containerSpecsByName
  - containerSpecsByNamespace
  - containerSpecsByControllerKind
  - containerSpecsByContainerPlatformCluster
  - containerSpecsByTag
- DataCenter
  - datacentersByName
  - datacentersByTag
- Database
  - databaseByName
  - databaseByTag
  - databaseByBusinessAccountUuid
  - databaseByResourceGroupName
  - databaseByCloudProvider
  - databaseByApplicationComponent
  - databaseByBusinessApplication
  - databaseByBusinessTransaction
  - databaseByService
  - dbByAvailabilityZoneName
  - dbByRegionName
  - dbByDatabaseServerName
  - databaseByReplicationRole
  - databaseByPricingModel
  - databaseByServiceTier
  - databaseByHoursIdle
  - databaseByAge
- DatabaseServer
  - databaseServerByName
  - databaseServerByTag
  - databaseServerByBusinessAccountUuid
  - databaseServerByResourceGroupName
  - databaseServerByCloudProvider
  - databaseServerByApplicationComponent
  - databaseServerByBusinessApplication
  - databaseServerByBusinessTransaction
  - databaseServerByService
  - databaseServerByEngine
  - databaseServerByEdition
  - databaseServerByVersion
  - databaseServerByState
  - databaseServerByServerConfiguration
  - databaseServerByStorageAutoscaling

  - databaseServerByPerformanceInsights
  - databaseServerByStorageEncryption
  - databaseServerByClusterRole
  - databaseServerByStorageTier
  - databaseServerByAvailabilityZoneName
  - databaseServerByRegionName
  - databaseServerByAge
- DatabaseServerTier
  - databaseServerTiersByName
  - databaseServerTiersByVMem
  - databaseServerTiersByCloudProvider
- DatabaseTier
  - databaseTiersByName
  - databaseTiersByCloudProvider
- DesktopPool
  - desktopPoolByName
  - desktopPoolByViewPod
- DiskArray
  - diskarrayByName
- DocumentCollection
  - dcByName
  - dcByTag
  - dcByBusinessAccount
  - dcByResourceGroup
  - dcByCloudProvider
  - dcByRegion
  - dcByDatabaseServer
- Group
  - groupsByName
- HCIPhysicalMachine
- IOModule
  - ioModuleByName
- Internet
  - internetByName
- LoadBalancer
  - lbsByName
- LogicalPool
  - logicalPoolByName
- Namespace
  - namespacesByName
  - namespacesByContainerPlatformCluster
  - namespacesByTag

- Network
  - netsByName
- NodePool
- PhysicalMachine
  - pmsByName
  - pmsByTargetType
  - pmsByClusterName
  - pmsByDC
  - pmsByState
  - pmsByTag
  - pmsByStorage
  - pmsBySwitch
  - pmsByMem
  - pmsByNumVms
  - pmsByNumCPUs
  - pmsByVendorName
  - pmsByCPUModel
  - pmsByModel
  - pmsByTimezone
  - pmsBySupportedProcessorCompatibilityModes
  - pmsByClusterTag
- ProcessorPool
- Region
  - regsByName
  - regsByCloudProvider
- ReservedInstance
- ResourceGroup
  - resourceGroupByName
  - resourceGroupByTag
  - resourceGroupByBusinessAccountUuid
- Service
  - servicesByName
  - serviceByNamespace
  - serviceByContainerPlatformCluster
  - serviceByKubernetesServiceType
  - serviceByBusinessApplication
  - serviceByTag
  - serviceByBusinessTransaction
  - serviceByUserDefinedEntity
- Storage
  - storageByName
  - storageByVMs
  - storageByPMCluster

  - storageByStorageCluster
  - storageByState
  - storageByTag
  - storageByDC
  - storageWithLocalSupport
- StorageCluster
  - storageClustersByName
- StorageController
  - storagecontrollerByName
- StorageTier
  - storageTierByName
  - storageTiersByCloudProvider
- Switch
  - switchByName
- ViewPod
  - viewPodByName
- VirtualDataCenter
  - vdcsByName
  - vdcsByTag
- VirtualMachine
  - vmsByName
  - vmsByPMName
  - vmsByVDC
  - vmsByDCnested
  - vmsByDiskArrayName
  - vmsByStorage
  - vmsByApplicationComponent
  - vmsByBusinessApplication
  - vmsByBusinessTransaction
  - vmsByService
  - vmsByClusterName
  - vmsByStorageClusterName
  - vmsByDC
  - vmsByState
  - vmsByTag
  - vmsConnectedToEphemeralStorage
  - vmsHotAddMemory
  - vmsHotAddCPU
  - vmsHotRemoveCPU
  - vmsConnectedToEncryptedVolume
  - vmsByGuestName
  - vmsBySockets

  - vmsByCoresPerSocket
  - vmsByNumCPUs
  - vmsByAge
  - vendorToolsInstalled
  - vendorToolsVersion
  - vmsByProcessorCompatibilityMode
  - vmsBySharingMode
  - vmsByHasDedicatedProcessors
  - vmByTargetCategory
  - vmByTargetName
  - vmByTargetType
  - vmsByBusinessAccountUuid
  - vmsByMem
  - vmsByStorageAmount
  - vmsByLogicalPoolName
  - vmsByNetwork
  - vmsByResourceGroupName
  - vmsByAvailabilityZoneName
  - vmsByRegionName
  - vmsByCloudProvider
  - vmsByContainerPlatformCluster
  - vmsByActiveSessions
  - vmsByVendorId
  - vmByComputeTier
- VirtualMachineCluster
  - virtualMachineClustersByName
- VirtualMachineSpec
  - virtualMachineSpecsByName
  - virtualMachineSpecsByTier
  - virtualMachineSpecsByAppCount
  - virtualMachineSpecsByBusinessAccountUuid
  - virtualMachineSpecsByRegionName
  - virtualMachineSpecsByResourceGroup
  - virtualMachineSpecsByComputeTier
  - virtualMachineSpecsByTag
  - virtualMachineSpecsByNumDaysEmpty
- VirtualVolume
  - virtualVolumeByName
  - virtualVolumeWithLocalSupport
  - volumeByTag
  - volumeByAccount
  - volumeByConnectedVm

    - volumeByAttachmentState
    - volumeByNumDaysUnattached
    - volumeByTier
    - volumeByRegion
    - volumeByCloudProvider
    - volumeById
    - volumeByResourceGroupName
    - volumeByDeletable
    - volumeByStorage
    - volumeByStorageGroup
    - volumeByAge
- Workload
    - workloadByName
- WorkloadController
    - workloadControllersByName
    - workloadControllersByNamespace
    - workloadControllersByKind
    - workloadControllersByContainerPlatformCluster
    - workloadControllersByTag

Example: `GET https://10.10.10.10/api/v3/search/criteria`

Response: An array of filter criteria, separated by entity type. For example, this is the filter criteria for Virtual Machine entities:


```
"VirtualMachine": {
    "criteria": [
      {
        "inputType": "*",
        "elements": "displayName",
        "filterCategory": "property",
        "filterType": "vmsByName"
      },
      {
        "inputType": "*",
        "elements": "HostedBy:displayName",
        "filterCategory": "entity",
        "filterType": "vmsByPMName"
      },
      {
        "inputType": "*",
        "elements": "LayeredOver:Storage:displayName",
        "filterCategory": "entity",
        "filterType": "vmsByStorage"
      },
      {
        "inputType": "*",
        "elements": "LayeredOver:Network:displayName",
        "filterCategory": "entity",
        "filterType": "vmsByNetwork"
      },
      {
        "inputType": "*",
        "elements": "Hosts:Application:displayName",
        "filterCategory": "entity",
        "filterType": "vmsByApplication"
```

```
    },
    {
      "inputType": "*",
      "elements": "HostedBy:PhysicalMachine:HostedBy:DataCenter:displayName",
      "filterCategory": "entity",
      "filterType": "vmsByDC"
    },
    {
      "inputType": "*",
      "elements": "LayeredOver:VirtualDataCenter:displayName",
      "filterCategory": "entity",
      "filterType": "vmsByVDC"
    },
    {
      "inputType": "*",
      "elements": "AllMemberOf:VirtualDataCenter:displayName",
      "filterCategory": "entity",
      "filterType": "vmsByDCnested"
    },
    {
      "inputType": "#",
      "elements": "ExtendedBy:VMTopoExt:numCpus",
      "filterCategory": "property",
      "filterType": "vmsByNumCPUs"
    },
    {
      "inputType": "#",
      "elements": "Commodities:VMem:capacity",
      "filterCategory": "property",
      "filterType": "vmsByMem"
    },
    {
      "inputType": "*",
      "elements": "ExtendedBy:VMTopoExt:guestName",
      "filterCategory": "property",
      "filterType": "vmsByGuestName"
    },
    {
      "inputType": "*",
      "elements": "ExtendedBy:VMTopoExt:altName",
      "filterCategory": "property",
      "filterType": "vmsByAltName"
    },
    {
      "inputType": "*",
      "elements": "HostedBy:PhysicalMachine:MemberOf:Cluster:displayName",
      "filterCategory": "entity",
      "filterType": "vmsByClusterName"
    },
    {
      "inputType": "*",
      "elements": "LayeredOver:Storage:LayeredOver:DiskArray:displayName",
      "filterCategory": "entity",
      "filterType": "vmsByDiskArrayName"
    },
    {
      "inputType": "*",
      "elements": "LayeredOver:Storage:LayeredOver:LogicalPool:displayName",
      "filterCategory": "entity",
      "filterType": "vmsByLogicalPoolName"
    },
    {
      "inputType": "*",
      "elements": "tags",
      "filterCategory": "property",
```

```
      "filterType": "vmsByTag"
    },
    {
      "inputType": "*",
      "elements": "state",
      "filterCategory": "property",
      "filterType": "vmsByState"
    },
    {
      "inputType": "*",
      "elements": "belongsTo:uuid",
      "filterCategory": "entity",
      "filterType": "vmsByBusinessAccountUuid"
    },
    {
      "inputType": "*",
      "elements": "MemberOf:ResourceGroup:uuid",
      "filterCategory": "entity",
      "filterType": "vmsByResourceGroupUuid"
    }
  ]
}
```

Getting the Options for a Criterion

For a specific search criterion and entity type, you can get the possible values. For example, you may want to retrieve all tags for a particular group of hosts, or see the available state options for a particular group of virtual machines. This request takes the following parameters:

- criteria_key

  The criterion to retrieve options for. Currently, Workload Optimization Manager accepts the following criteria keys: `[ state, tags ]`. Any other input that returns results should be considered internal or early access, and should not be relied on for scripting purposes.

- scopes

  The UUID of the scope to retrieve the criterion values for. For example, to retrieve all tags for virtual machines in a particular cluster, put the cluster UUID here.

- entity_type

  The entity whose state or tag options will be returned. For example, if `scopes` is passed with the UUID of a cluster, and this parameter is passed with a value of VirtualMachine, the request will return the options for each virtual machine in that cluster.

- order_by

  The field to order the results by. `[ creation_date, name, risk_category, savings, severity ]`.

- ascending

  Default: true. When false, results will be in descending order.

- environment_type

  Returns only objects in the specified environment type. `[ Cloud, Hybrid, Onprem ]`

Example: `GET https://10.10.10.10/api/v3/search/criteria/state/options?scopes=7c8d50fae46c09b68b0091641146970304226b4&entity_type=VirtualMachine`

Response: A `CriteriaOptionApiDTO` with the available options. For example, the following response indicates that all virtual machines on the cluster whose UUID was passed are either in active or idle states.

```
[
  {
    "value": "ACTIVE"
  },
  {
```

```
    "value": "IDLE"
  }
]
```

Search requests

Getting object information by UUID

Gets information about any object in your Workload Optimization Manager environment.

`Object` and `Entity` refer to different items in your Workload Optimization Manager topology:

> **NOTE:**
> For searches, the API paginates the returned data. If you do not specify a pagination limit, then the API returns pages set to the default limit of 100. You can set a maximum of 500 for the pagination limit. If you make no settings, and your search results include more than 100 entries, the results will be paginated. Your scripts should always check for a pagination cursor when processing search results.
>
> You can change the default and max settings in the Workload Optimization Manager cr.yaml file. Find the entries for:
>
> - `apiPaginationDefaultLimit`
> - `apiPaginationMaxLimit`
>
> If you set the `objectType` parameter has a value of `Group`, the response will include objects related to an entity that perform grouping. For example, resource groups and clusters both perform grouping, and will be included in the response.
>
> To further refine your search, include a value for the `group_type` parameter. This parameter describes the membership of the group you are searching for. For example, a search with an `objectType` of `Group` and `group_type` of `VirtualMachine` will return all groups of virtual machines in the current scope.

Example: `https://10.10.10.10/api/v3/search/Market`

Response: A DTO with information about the retrieved object. In this example, the retrieved item is the real-time Market:

```
{
  "links": [
    {
      "rel": "self",
      "href": "https://10.10.10.10/api/v3/markets/_0x3OYUglEd-gHc4L513yOA"
    }
  ],
  "uuid": "_0x3OYUglEd-gHc4L513yOA",
  "displayName": "Market",
  "className": "Market",
  "state": "RUNNING",
  "unplacedEntities": false,
  "environmentType": "HYBRID"
}
```

> **NOTE:**
> When making this request, you can use the `groupBy` parameter to order the results. The `/search` endpoint does not support the `utilization` value for the `groupBy` parameter. When this value is used, the API will return the results ordered by `displayName`.

Settings endpoint

The settings endpoint can be used to change settings in a variety of settings managers. For example, using the `actionscriptmanager`, you can change action script settings, and using the `automationmanager`, you can change automation settings on your instance.

Using the settings endpoint of the Workload Optimization Manager API, you can:

- Get a list of available setting managers and settings
- Edit a setting

The following settings managers are available to be modified:

- `actionscriptmanager`

  Contains settings related to the use of action scripts.
- `automationmanager`

  Contains settings related to action modes for entities in your environment.
- `cloudsavingsmanager`

  Contains settings related to execution retention for cloud entities.
- `clusterHeadroomConfig`

  Contains settings related to the headroom plans generated by the system.
- `controlmanager`

  Contains settings related to workflows that replace native action handling for generated actions in your environment.
- `emailmanager`

  Contains settings related to SMTP email notifications.
- `hcisettingsmanager`

  Contains settings related to managing hyperconverged infrastructure.
- `marketsettingsmanager`

  Contains settings related to market analysis.
- `osmigrationmanager`

  Contains settings realted to operating system migration for cloud entities.
- `persistencemanager`

  Contains settings related to statistic retention data and reporting.
- `reservedinstancemanager`

  Contains settings related to reserved instances.
- `storagesettingsmanager`

  Contains settings related to storage entity statistics and analysis.
- `telemetrymanager`

  Contains settings related to telemetry opt-in and terms of service.
-

For more information about each settingsmanager, execute the `GET https://10.10.10.10/api/v3/settings` request.

Settings utility requests

Utility requests provide metadata that you can use to make related requests, assemble inputDTOs, or see the potential values of a particular DTO.

The settings endpoint contains a utility request that shows each settings manager, and the settings available.

Example: `GET https://10.10.10.10/api/v3/settings/`

Response: An array of `SettingsManagerApiDTO`s, where each object is a settings manager and contains the settings for that manager.

Settings requests

Getting current settings for a specified settings manager

Gets the current settings for the specified settings manager. This request is often used to see the settings of a particular settings manager before updating.

Example: `GET https://10.10.10.10/api/v3/settings/automationmanager`

Response: A list of `SettingApiDTO`s representing the settings for the specified settings manager:


```
[
  {
    "uuid": "startVM",
    "displayName": "Start",
    "value": "RECOMMEND",
    "defaultValue": "RECOMMEND",
    "valueType": "STRING",
    "options": [
      {
        "label": "Disabled",
        "value": "DISABLED"
      },
      {
        "label": "Recommend",
        "value": "RECOMMEND"
      },
      {
        "label": "Manual",
        "value": "MANUAL"
      },
      {
        "label": "Automated",
        "value": "AUTOMATIC"
      }
    ],
    "entityType": "VirtualMachine",
    "sourceGroupName": "Global"
  },
  {
    "uuid": "startPM",
    "displayName": "Start",
    "value": "RECOMMEND",
    "defaultValue": "RECOMMEND",
    "valueType": "STRING",
    "options": [
      {
        "label": "Disabled",
        "value": "DISABLED"
      },
      {
        "label": "Recommend",
        "value": "RECOMMEND"
      },
      {
```

```
        "label": "Manual",
        "value": "MANUAL"
      },
      {
        "label": "Automated",
        "value": "AUTOMATIC"
      }
    ],
    "entityType": "PhysicalMachine",
    "sourceGroupName": "Global"
  },
  {
    "uuid": "startDS",
    "displayName": "Start",
    "value": "RECOMMEND",
    "defaultValue": "RECOMMEND",
    "valueType": "STRING",
    "options": [
      {
        "label": "Disabled",
        "value": "DISABLED"
      },
      {
        "label": "Recommend",
        "value": "RECOMMEND"
      },
      {
        "label": "Manual",
        "value": "MANUAL"
      },
      {
        "label": "Automated",
        "value": "AUTOMATIC"
      }
    ],
    "entityType": "Storage",
    "sourceGroupName": "Global"
  },
  ...
]
```


Editing settings

To edit a Setting, pass the new value in the body of this request. This request takes the following parameters:

- settingsManager_uuid

  The UUID of the settingsmanager whose setting will be changed.
- setting_uuid

  The setting whose value will be changed.

Example: `PUT https://10.10.10.10/api/v3/settings/automationmanager/moveVM`

Example Input: A SettingApiDTO. In this case, the value is being changed from `MANUAL` to `RECOMMEND`:


```
{
  "uuid": "moveVM",
  "displayName": "Move",
  "value": "RECOMMEND",
  "defaultValue": "MANUAL",
  "valueType": "STRING",
  "options": [
    {
```

```
      "label": "Disabled",
      "value": "DISABLED"
    },
    {
      "label": "Recommend",
      "value": "RECOMMEND"
    },
    {
      "label": "Manual",
      "value": "MANUAL"
    },
    {
      "label": "Automated",
      "value": "AUTOMATIC"
    }
  ],
  "entityType": "VirtualMachine",
  "sourceGroupName": "Global"
}
```

Response: The `SettingApiDTO` reflecting your changes:

```
{
  "uuid": "moveVM",
  "displayName": "Move",
  "value": "RECOMMEND",
  "defaultValue": "MANUAL",
  "valueType": "STRING",
  "options": [
    {
      "label": "Disabled",
      "value": "DISABLED"
    },
    {
      "label": "Recommend",
      "value": "RECOMMEND"
    },
    {
      "label": "Manual",
      "value": "MANUAL"
    },
    {
      "label": "Automated",
      "value": "AUTOMATIC"
    }
  ],
  "entityType": "VirtualMachine",
  "sourceGroupName": "Global"
}
```

Settings policies endpoint

A Settings Policy is the collection of settings that apply to an entity. In Workload Optimization Manager, there is a default policy for each type of entity. These global defaults are the base settings for entities, and any settings policies that you create will specify deltas to the defaults. Note that you can edit the global policies.

When you create a custom settings policy, you assign a scope to it – one or more groups of entities you want to affect with the policy. For these policies, you specify just the settings that will be different from the defaults.

In addition to global default policies and custom policies, Workload Optimization Manager also discovers groups that need special settings, and creates policies for those groups. For example, when Workload Optimization Manager discovers Azure Availability Sets and AWS Autoscaling Groups, it generates settings policies for those groups that turn on Consistent Resizing for the group.

Using the settings policies endpoint, you can:

- Get a listing of the settings policies in your environment
- Get details for a single settings policy
- Create settings policies
- Edit settings policies
- Delete settings policies

For more information about managing policies with the API, see Working with policies in the API *(on page 1148)*.

Settings policies requests

Getting settings policies

This call returns an array of SettingsPolicyApiDTOs for all the settings policies in your environment. You can pass entity types to limit the listing to policies for those types, and you can limit the list to only default policies.

This request takes the following parameters:

- only_defaults

  If this is `true`, the call only gets the default settings policies
- entity_type

  One or more entity types to limit the list of settings policies

Examples:

- Get all the settings policies: `GET https://10.10.10.10/api/v3/settingspolicies`
- Get all the default settings policies: `GET https://10.10.10.10/api/v3/settingspolicies?only_defaults=true`
- Get settings policies for VMs and Hosts: `GET https://10.10.10.10/api/v3/settingspolicies?entity_type=VirtualMachine&entity_type=PhysicalMachine`
- Get a single settings policy (pass the policy Uuid): `GET https://10.10.10.10/api/v3/settingspolicies/_aVDEwGegEemDFcIzuV5wcQ`

Response: An array of `SettingsPolicyApiDTO`s. This listing shows the settings policies for Container entities.


```
[
  {
    "uuid": "_vyMHEDnSEemHXrxYkw9y2Q",
    "displayName": "Container defaults",
    "entityType": "Container",
    "settingsManagers": [
      {
        "uuid": "marketsettingsmanager",
        "displayName": "Operational Constraints",
        "category": "Analysis",
        "settings": [
          {
            "uuid": "usedIncrement_Container_VMEM",
            "displayName": "Increment constant for Container VMem [MB]",
            "value": "64.0",
            "defaultValue": "64.0",
            "categories": [
              "resizeRecommendationsConstants"
```

```
            ],
            "valueType": "NUMERIC",
            "min": 0,
            "max": 10000000,
            "entityType": "Container",
            "sourceGroupName": "Global"
          },
          {
            "uuid": "usedIncrement_Container_VCPU",
            "displayName": "Increment constant for Container VCPU [MHz]",
            "value": "100.0",
            "defaultValue": "100.0",
            "categories": [
              "resizeRecommendationsConstants"
            ],
            "valueType": "NUMERIC",
            "min": 0,
            "max": 1000000,
            "entityType": "Container",
            "sourceGroupName": "Global"
          },
          {
            "uuid": "enforceConsistentResizingContainer",
            "displayName": "Consistent Resizing",
            "value": "false",
            "defaultValue": "false",
            "categories": [
              "resizeRecommendationsConstants"
            ],
            "valueType": "BOOLEAN",
            "entityType": "Container",
            "sourceGroupName": "Global"
          },
          {
            "uuid": "resizeTargetUtilization_Container",
            "displayName": "Resize Target Utilization",
            "value": "0.0",
            "defaultValue": "0.0",
            "valueType": "NUMERIC",
            "min": 0,
            "max": 100,
            "entityType": "Container",
            "sourceGroupName": "Global"
          }
        ]
      },
      ...
  }
]
```


Creating a settings policy

Referenced Endpoint: Groups (on page 1222)

To create a settings policy, you must do the following:

- Specify a policy name,
- Specify whether to enable or disable the policy in your market.
- Set the policy scope – Specify one or more groups of a given entity type. This determines the entity type for the policy.

- For the of settings you want to make, specify the appropriate settings manager, and then specify an array of setting/value pairs.
- Optionally, specify a schedule window and/or action execution schedule for the policy.

To set the policy scope, you provide an array of objects that contain group Uuids. These objects are effectively abbreviated GroupApiDTOs. Note that the groups must all be for the same type of entity. This specifies the entity type for the settings policy. Use the Groups endpoint to get the Uuid values for the groups you want in the scope. For example, to set the scope to a single group of VMs, use the following scopes object:

```
"scopes": [
    {
      "uuid": "_lSlr0GddEemDFcIzuV5wcQ"
    }
]
```

To specify the actual settings, you provide an array of SettingsManagerApiDTOs. Each one contains an array of settings that you want to make. To see all the settings for a given entity type, refer to the user interface. The setting managers you can choose from are:

- `actionscriptmanager`

  Contains settings related to the use of action scripts.
- `automationmanager`

  Contains settings related to action modes for entities in your environment.
- `cloudsavingsmanager`

  Contains settings related to execution retention for cloud entities.
- `clusterHeadroomConfig`

  Contains settings related to the headroom plans generated by the system.
- `controlmanager`

  Contains settings related to workflows that replace native action handling for generated actions in your environment.
- `emailmanager`

  Contains settings related to SMTP email notifications.
- `hcisettingsmanager`

  Contains settings related to managing hyperconverged infrastructure.
- `marketsettingsmanager`

  Contains settings related to market analysis.
- `osmigrationmanager`

  Contains settings realted to operating system migration for cloud entities.
- `persistencemanager`

  Contains settings related to statistic retention data and reporting.
- `reservedinstancemanager`

  Contains settings related to reserved instances.
- `storagesettingsmanager`

  Contains settings related to storage entity statistics and analysis.
- `telemetrymanager`

  Contains settings related to telemetry opt-in and terms of service.
-

For example, to set the VM start action to AUTOMATED and to set the VCPU maximum to 16, use the following `settingsManagers` array:

```
"settingsManagers": [
    {
      "uuid": "automationmanager",
      "settings": [
        {
          "uuid": "startVM",
          "value": "AUTOMATIC"
        }
      ]
    },
    {
      "uuid": "marketsettingsmanager",
      "settings": [
        {
          "uuid": "resizeVcpuMaximum_VM",
          "value": 16
        }
      ]
    }
]
```

You can set up a schedule window that determines when the settings policy takes effect. For that you provide a ScheduleApiDTO for the schedule details. It includes: the start and end dates, start and end times,

- Start and end dates.
- Start and end times.
- An optional RecurrenceApiDTO for schedule recurrence. If you don't provide this you will specify a one-time schedule window.

For example, use this to specify a "perpetual" schedule window (does not expire until 2099) that starts at 12:30 am and ends at 7:30 am every Saturday and Sunday.

```
"schedule": {
  "startDate": "2019-04-28T00:30:00-04:00",
  "endDate": "2099-12-31T07:30:00-05:00",
  "startTime": "2019-04-28T00:30:00-04:00",
  "endTime": "2019-04-28T07:30:00-04:00",
  "recurrence": {
    "type": "WEEKLY",
    "daysOfWeek": [
      "Sat",
      "Sun"
    ]
}
```

An example complete data payload for a policy with a policy-level schedule, action execution instruction, and action execution schedule is as follows:

```
{
   "disabled":false,
   "entityType":"VirtualMachine",
   "displayName":"TestPol",
   "scopes":[
      {
         "uuid":"285041619454560"
      }
   ],
   "settingsManagers":[
      {
         "uuid":"automationmanager",
         "displayName":"Action Mode Settings",
```

```
        "category":"Automation",
        "settings":[
           {
              "uuid":"storageMove",
              "value":"RECOMMEND",
              "valueType":"STRING",
              "valueObjectType":"String",
              "defaultValue":"RECOMMEND",
              "entityType":"VirtualMachine",
              "displayName":"Storage Move",
              "options":[
                 {
                    "label":"Disabled",
                    "value":"DISABLED"
                 },
                 {
                    "label":"Recommend",
                    "value":"RECOMMEND"
                 },
                 {
                    "label":"External Approval",
                    "value":"EXTERNAL_APPROVAL"
                 },
                 {
                    "label":"Manual",
                    "value":"MANUAL"
                 },
                 {
                    "label":"Automatic",
                    "value":"AUTOMATIC"
                 }
              ]
           },
           {
              "uuid":"storageMoveExecutionSchedule",
              "value":"285088800975856",
              "valueType":"LIST",
              "valueObjectType":"String",
              "entityType":"VirtualMachine",
              "displayName":"Execution window for Storage Move action"
           }
        ]
     }
  ],
  "schedule":{
     "uuid":"285088797821136",
     "displayName":"TestSched",
     "endTime":"2021-05-24T16:00",
     "startTime":"2021-05-24T14:30",
     "timeZone":"America/Chicago",
     "recurrence":{
        "type":"MONTHLY",
        "interval":1,
        "daysOfMonth":[
           1
        ]
     }
  }
}
```


Example: Using the previous payload, `POST https://10.10.10.10/api/v3/settingspolicies`

Response: The settings policy that you created. Use this response to get the policy Uuid.

Editing a settings policy

To edit a settings policy, you push a new set of data to the Settings Policies end point. The best way to do this is to copy the SettingsPolicyApiDTO that you want to edit, make changes to that DTO, and then push the DTO out to the API. To remove the policy from the scope, pass the SettingsPolicyApiDTO, and also pass `reset_defaults=true` as a parameter.

This request takes the following parameters:

- settingsPolicy_Uuid

  The Uuid of the policy you want to edit

- reset_defaults

  A boolean you can use to reset the policy to default. When `reset_defaults=true`, you remove the policy and reset that scope to default. When you use this setting, pass an empty object `{}` as the data payload.

Examples:

- Assuming you copied a SettingsPolicyApiDTO and then made some changes to it, and pass that at the payload: `PUSH https://10.10.10.10/api/v3/settingspolicies/_2-CbQGoNEemyOoIvFZyv1A?reset_defaults=false`
- Assuming you want to remove the policy (and so you pass an empty object as the payload): PUSH `https://10.10.10.10/api/v3/settingspolicies/_2-CbQGoNEemyOoIvFZyv1A?reset_defaults=true`

Response: The modified `SettingsPolicyApiDTO`.

Deleting a Settings Policy

Deletes the given settings policy.

Example: `DELETE https://10.10.10.10/api/v3/settingspolicies/_aV5ucGoTEemyOoIvFZyv1A`

Response: 200 for success.

Statistics endpoint

Statistics overview

Statistics are pieces of information that Workload Optimization Manager discovers while continuously analyzing your environment. The most common use of statistics is to look at a particular statistic over a specific time frame.

Using the stats endpoint of the Workload Optimization Manager API, you can:

- Get a list of stats requests in other endpoints
- Get a filtered list of statistics over a specific time frame
- Get a list of statistics for a specified entity
- Get a filtered list of statistics for a specified entity

Statistic timings

Statistics associated to headroom and capacity planning are written to the Workload Optimization Manager database daily, with a timestamp of midnight. Other statistics are collected and stored in the database in ten-minute increments. When retrieving groups of statistics that include both headroom and actively-monitored information, the timestamps returned will reflect the last-written date for each statistic.

Statistics utility requests

Utility requests provide metadata that you can use to make related requests, assemble inputDTOs, or see the potential values of a particular DTO.

The stats endpoint contains a utility request that shows common entry points for statistic-related requests. These can be used to obtain statistics for particular entities or groups.

Example: `GET https://10.10.10.10/api/v3/stats`

Response: An array of Link objects, where each object contains a URL to get all statistics for a specified UUID, entity, or group.

Statistics requests

Getting statistics for an entity

Gets all statistics for the specified entity. This request takes the following parameters:

- entity_uuid

  The entity whose statistics will be returned.

- q

  Used internally by Workload Optimization Manager.

NOTE:
When you retrieve statistics for a cluster, using the `/groups` endpoint will return the aggregated statistics for all applicable cluster entities. For individual stats per entity, use the `/stats` endpoint.

Some statistics are only valid for certain scopes. When you request given statistics for a scope that doesn't include them, the API does not return any data. For example, the `numVMs` statistic is valid in only the Market and Cluster scopes – the API omits `numVMs` data from the response when you query other scopes.

Example: https://10.10.10.10/api/v3/stats/423f26a7-1132-df1c-64e2-8a040480b046

Response: A list of `StatSnapshotApiDTO`s representing the current values of the statistics.


```
[
  {
    "displayName": "AdityaReplica",
    "date": "2019-04-24T20:09:46-04:00",
    "statistics": [
      {
        "name": "priceIndex",
        "relatedEntityType": "VirtualMachine",
        "values": {
          "max": 1.01,
          "min": 1.01,
          "avg": 1.01,
          "total": 1.01
        },
        "value": 1.01
      },
      {
        "displayName": "CPUProvisioned/AdityaReplica",
        "name": "CPUProvisioned",
        "capacity": {
          "max": 3120000,
          "min": 3120000,
          "avg": 3120000,
```

```
        "total": 3120000
      },
      "relatedEntityType": "VirtualMachine",
      "filters": [
        {
          "type": "relation",
          "value": "bought"
        }
      ],
      "units": "MHz",
      "values": {
        "max": 5200,
        "min": 5200,
        "avg": 5200,
        "total": 5200
      },
      "value": 5200
    },
    {
      "displayName": "Swapping/AdityaReplica",
      "name": "Swapping",
      "capacity": {
        "max": 40000000,
        "min": 40000000,
        "avg": 40000000,
        "total": 40000000
      },
      "relatedEntityType": "VirtualMachine",
      "filters": [
        {
          "type": "relation",
          "value": "bought"
        }
      ],
      "units": "bit/sec",
      "values": {
        "max": 0,
        "min": 0,
        "avg": 0,
        "total": 0
      },
      "value": 0
    },
    {
      "displayName": "NumDisk/AdityaReplica",
      "name": "NumDisk",
      "capacity": {
        "max": 2147483650,
        "min": 2147483650,
        "avg": 2147483650,
        "total": 2147483650
      },
      "relatedEntityType": "VirtualMachine",
      "filters": [
        {
          "type": "relation",
          "value": "bought"
        }
      ],
      "values": {
        "max": 1,
        "min": 1,
        "avg": 1,
        "total": 1
      },
```

```
        "value": 1
      },
    ...
    ]
  }
]
```

Getting a filtered list of statistics

Gets a list of statistics filtered by a `StatPeriodApiInputDTO` that you create. For more information about the criteria that make up the input for thi request, see the Swagger documentation for this request, located at `https://<Your_Workload Optimization Manager_URL/apidpoc`.

Example: POST https://10.10.10.10/api/v3/stats/423f26a7-1132-df1c-64e2-8a040480b046

Example Input: This input will return the values for the VMem, VCPU, and VStorage statistics of the specified entity, ordered first by key, then relatedEntity, then virtualDisk.

```
{
  "statistics": [
    {
      "name": "VMem",
      "groupBy": [
        "key",
        "relatedEntity",
        "virtualDisk"
      ]
    },
    {
      "name": "VCPU",
      "groupBy": [
        "key",
        "relatedEntity",
        "virtualDisk"
      ]
    },
    {
      "name": "VStorage",
      "groupBy": [
        "key",
        "relatedEntity",
        "virtualDisk"
      ]
    }
  ]
}
```

Response: A list of `StatSnapshotApiDTO`s representing the filtered statistics for the entity:

```
[
  {
    "displayName": "AdityaReplica",
    "date": "2019-04-24T20:28:06-04:00",
    "statistics": [
      {
        "displayName": "AdityaReplica",
        "name": "VMem",
        "capacity": {
          "max": 1048576,
          "min": 1048576,
          "avg": 1048576,
```

```
        "total": 1048576
      },
      "relatedEntityType": "VirtualMachine",
      "filters": [
        {
          "type": "virtualDisk",
          "value": ""
        },
        {
          "type": "key",
          "value": null
        },
        {
          "type": "relation",
          "value": "sold"
        }
      ],
      "relatedEntity": {
        "uuid": "5d2f7b69eb3bf506089feba37d31b0c376b02e34",
        "displayName": "GuestLoad[AdityaReplica]",
        "className": "Application"
      },
      "units": "KB",
      "values": {
        "max": 0,
        "min": 0,
        "avg": 0,
        "total": 0
      },
      "value": 0
    },
    {
      "displayName": "AdityaReplica",
      "name": "VCPU",
      "capacity": {
        "max": 5200,
        "min": 5200,
        "avg": 5200,
        "total": 5200
      },
      "relatedEntityType": "VirtualMachine",
      "filters": [
        {
          "type": "virtualDisk",
          "value": ""
        },
        {
          "type": "key",
          "value": null
        },
        {
          "type": "relation",
          "value": "sold"
        }
      ],
      "relatedEntity": {
        "uuid": "5d2f7b69eb3bf506089feba37d31b0c376b02e34",
        "displayName": "GuestLoad[AdityaReplica]",
        "className": "Application"
      },
      "units": "MHz",
      "values": {
        "max": 19,
        "min": 14.93,
        "avg": 14.93,
```

```
          "total": 14.93
        },
        "value": 14.93
      }
    ]
  }
]
```

Getting a filtered list of statistics for all objects in a specified scope

For the specified scope, gets a filtered list of statistics for each entity. If you do not provide any statistics in the input, the response contains basic information about each entity. For more information about the criteria that make up the input for thi request, see the Swagger documentation for this request, located at `https://<Your_Workload Optimization Manager_URL/apidpoc`.

Example: POST https://10.10.10.10/api/v3/stats?ascending=true

Example Input: This input will return the VMem values for each entity in the real-time market from a period of five days prior to the request, to one day prior to the request.

```
{
  "scopes": [
    "Market"
  ],
  "period": {
    "startDate": "-5d",
    "endDate": "-1d",
    "statistics": [
      {
        "name": "VMem"
      }
    ]
  },
  "relatedType": "VirtualMachine"
}
```

Response: A list of `StatSnapshotApiDTO`s representing the filtered statistics for the entities in the specified scope:

```
[
  {
    "uuid": "4200829c-1272-c0a4-2d17-55f2cc0ef1b4",
    "displayName": "vm-c67a1cdd-609e-4282-809c-e0def5cd0d7a",
    "className": "VirtualMachine",
    "environmentType": "ONPREM",
    "stats": [
      {
        "displayName": "vm-c67a1cdd-609e-4282-809c-e0def5cd0d7a",
        "date": "2019-04-19T00:00:00-04:00",
        "statistics": [
          {
            "name": "VMem",
            "capacity": {
              "max": 16777216,
              "min": 16777216,
              "avg": 16777216,
              "total": 33554432
            },
            "relatedEntityType": "VirtualMachine",
            "filters": [
              {
                "type": "relation",
```

```
              "value": "sold"
            }
          ],
          "units": "KB",
          "values": {
            "max": 1509949.5,
            "min": 335544.3,
            "avg": 587202.56,
            "total": 1174405.1
          },
          "value": 587202.56
        }
      ]
    },
    {
      "displayName": "vm-c67a1cdd-609e-4282-809c-e0def5cd0d7a",
      "date": "2019-04-20T00:00:00-04:00",
      "statistics": [
        {
          "name": "VMem",
          "capacity": {
            "max": 16777216,
            "min": 16777216,
            "avg": 16777216,
            "total": 33554432
          },
          "relatedEntityType": "VirtualMachine",
          "filters": [
            {
              "type": "relation",
              "value": "sold"
            }
          ],
          "units": "KB",
          "values": {
            "max": 1509949.5,
            "min": 369098.75,
            "avg": 587202.56,
            "total": 1174405.1
          },
          "value": 587202.56
        }
      ]
    },
    {
      "displayName": "vm-c67a1cdd-609e-4282-809c-e0def5cd0d7a",
      "date": "2019-04-22T00:00:00-04:00",
      "statistics": [
        {
          "name": "VMem",
          "capacity": {
            "max": 16777216,
            "min": 16777216,
            "avg": 16777216,
            "total": 33554432
          },
          "relatedEntityType": "VirtualMachine",
          "filters": [
            {
              "type": "relation",
              "value": "sold"
            }
          ],
          "units": "KB",
          "values": {
```

```
            "max": 12582912,
            "min": 0,
            "avg": 452984.84,
            "total": 905969.7
          },
          "value": 452984.84
        }
      ]
    },
    {
      "displayName": "vm-c67a1cdd-609e-4282-809c-e0def5cd0d7a",
      "date": "2019-04-23T00:00:00-04:00",
      "statistics": [
        {
          "name": "VMem",
          "capacity": {
            "max": 16777216,
            "min": 16777216,
            "avg": 16777216,
            "total": 33554432
          },
          "relatedEntityType": "VirtualMachine",
          "filters": [
            {
              "type": "relation",
              "value": "sold"
            }
          ],
          "units": "KB",
          "values": {
            "max": 838860.8,
            "min": 0,
            "avg": 167772.16,
            "total": 335544.3
          },
          "value": 167772.16
        }
      ]
    }
  ]
},
{
  "uuid": "39EBB726-41D9-4BC8-A41D-CA301854309C",
  "displayName": "Gilad_OM-38071_SMB_VM2",
  "className": "VirtualMachine",
  "environmentType": "ONPREM"
},
{
  "uuid": "97A3F420-B7AD-414D-A543-FD308AC774A6",
  "displayName": "Gilad_VMM_Test_VM_2",
  "className": "VirtualMachine",
  "environmentType": "ONPREM"
},
{
  "uuid": "4211f183-69a8-3676-3e53-0151365500ff",
  "displayName": "james-ubuntu-dc13",
  "className": "VirtualMachine",
  "environmentType": "ONPREM",
  "stats": [
    {
      "displayName": "james-ubuntu-dc13",
      "date": "2019-04-19T00:00:00-04:00",
      "statistics": [
        {
          "name": "VMem",
```

```
            "capacity": {
              "max": 1048576,
              "min": 1048576,
              "avg": 1048576,
              "total": 2097152
            },
            "relatedEntityType": "VirtualMachine",
            "filters": [
              {
                "type": "relation",
                "value": "sold"
              }
            ],
            "units": "KB",
            "values": {
              "max": 83886.08,
              "min": 8388.61,
              "avg": 18874.37,
              "total": 37748.73
            },
            "value": 18874.37
          }
        ]
      },
      {
        "displayName": "james-ubuntu-dc13",
        "date": "2019-04-20T00:00:00-04:00",
        "statistics": [
          {
            "name": "VMem",
            "capacity": {
              "max": 1048576,
              "min": 1048576,
              "avg": 1048576,
              "total": 2097152
            },
            "relatedEntityType": "VirtualMachine",
            "filters": [
              {
                "type": "relation",
                "value": "sold"
              }
            ],
            "units": "KB",
            "values": {
              "max": 62914.56,
              "min": 13631.49,
              "avg": 18874.37,
              "total": 37748.73
            },
            "value": 18874.37
          }
        ]
      },
      {
        "displayName": "james-ubuntu-dc13",
        "date": "2019-04-21T00:00:00-04:00",
        "statistics": [
          {
            "name": "VMem",
            "capacity": {
              "max": 1048576,
              "min": 1048576,
              "avg": 1048576,
              "total": 2097152
```

```
          },
          "relatedEntityType": "VirtualMachine",
          "filters": [
            {
              "type": "relation",
              "value": "sold"
            }
          ],
          "units": "KB",
          "values": {
            "max": 73400.32,
            "min": 11534.34,
            "avg": 17825.79,
            "total": 35651.59
          },
          "value": 17825.79
        }
      ]
    },
    {
      "displayName": "james-ubuntu-dc13",
      "date": "2019-04-22T00:00:00-04:00",
      "statistics": [
        {
          "name": "VMem",
          "capacity": {
            "max": 1048576,
            "min": 1048576,
            "avg": 1048576,
            "total": 2097152
          },
          "relatedEntityType": "VirtualMachine",
          "filters": [
            {
              "type": "relation",
              "value": "sold"
            }
          ],
          "units": "KB",
          "values": {
            "max": 83886.08,
            "min": 8388.61,
            "avg": 19922.94,
            "total": 39845.89
          },
          "value": 19922.94
        }
      ]
    },
    {
      "displayName": "james-ubuntu-dc13",
      "date": "2019-04-23T00:00:00-04:00",
      "statistics": [
        {
          "name": "VMem",
          "capacity": {
            "max": 1048576,
            "min": 1048576,
            "avg": 1048576,
            "total": 2097152
          },
          "relatedEntityType": "VirtualMachine",
          "filters": [
            {
              "type": "relation",
```

```
              "value": "sold"
            }
          ],
          "units": "KB",
          "values": {
            "max": 73400.32,
            "min": 3145.73,
            "avg": 18874.37,
            "total": 37748.73
          },
          "value": 18874.37
        }
      ]
    }
  ]
},
{
  "uuid": "421d2165-36cd-7f6a-3a43-08e09f6f4e0f",
  "displayName": "ACM-LexCorp-Control",
  "className": "VirtualMachine",
  "environmentType": "ONPREM",
  "stats": [
    {
      "displayName": "ACM-LexCorp-Control",
      "date": "2019-04-19T00:00:00-04:00",
      "statistics": [
        {
          "name": "VMem",
          "capacity": {
            "max": 2097152,
            "min": 2097152,
            "avg": 2097152,
            "total": 4194304
          },
          "relatedEntityType": "VirtualMachine",
          "filters": [
            {
              "type": "relation",
              "value": "sold"
            }
          ],
          "units": "KB",
          "values": {
            "max": 167772.16,
            "min": 18874.37,
            "avg": 37748.73,
            "total": 75497.47
          },
          "value": 37748.73
        }
      ]
    },
    {
      "displayName": "ACM-LexCorp-Control",
      "date": "2019-04-20T00:00:00-04:00",
      "statistics": [
        {
          "name": "VMem",
          "capacity": {
            "max": 2097152,
            "min": 2097152,
            "avg": 2097152,
            "total": 4194304
          },
          "relatedEntityType": "VirtualMachine",
```

```
          "filters": [
            {
              "type": "relation",
              "value": "sold"
            }
          ],
          "units": "KB",
          "values": {
            "max": 125829.12,
            "min": 18874.37,
            "avg": 37748.73,
            "total": 75497.47
          },
          "value": 37748.73
        }
      ]
    },
    {
      "displayName": "ACM-LexCorp-Control",
      "date": "2019-04-21T00:00:00-04:00",
      "statistics": [
        {
          "name": "VMem",
          "capacity": {
            "max": 2097152,
            "min": 2097152,
            "avg": 2097152,
            "total": 4194304
          },
          "relatedEntityType": "VirtualMachine",
          "filters": [
            {
              "type": "relation",
              "value": "sold"
            }
          ],
          "units": "KB",
          "values": {
            "max": 104857.6,
            "min": 20971.52,
            "avg": 35651.59,
            "total": 71303.17
          },
          "value": 35651.59
        }
      ]
    },
    {
      "displayName": "ACM-LexCorp-Control",
      "date": "2019-04-22T00:00:00-04:00",
      "statistics": [
        {
          "name": "VMem",
          "capacity": {
            "max": 2097152,
            "min": 2097152,
            "avg": 2097152,
            "total": 4194304
          },
          "relatedEntityType": "VirtualMachine",
          "filters": [
            {
              "type": "relation",
              "value": "sold"
            }
```

```
            ],
            "units": "KB",
            "values": {
              "max": 272629.75,
              "min": 20971.52,
              "avg": 73400.32,
              "total": 146800.64
            },
            "value": 73400.32
          }
        ]
      },
      {
        "displayName": "ACM-LexCorp-Control",
        "date": "2019-04-23T00:00:00-04:00",
        "statistics": [
          {
            "name": "VMem",
            "capacity": {
              "max": 2097152,
              "min": 2097152,
              "avg": 2097152,
              "total": 4194304
            },
            "relatedEntityType": "VirtualMachine",
            "filters": [
              {
                "type": "relation",
                "value": "sold"
              }
            ],
            "units": "KB",
            "values": {
              "max": 251658.23,
              "min": 23068.67,
              "avg": 79691.77,
              "total": 159383.55
            },
            "value": 79691.77
          }
        ]
      }
    ]
  },
  {
    "uuid": "420f3462-ca50-fedf-aa8f-223bd39b1852",
    "displayName": "shai-redhat",
    "className": "VirtualMachine",
    "environmentType": "ONPREM",
    "stats": [
      {
        "displayName": "shai-redhat",
        "date": "2019-04-19T00:00:00-04:00",
        "statistics": [
          {
            "name": "VMem",
            "capacity": {
              "max": 1048576,
              "min": 1048576,
              "avg": 1048576,
              "total": 2097152
            },
            "relatedEntityType": "VirtualMachine",
            "filters": [
              {
```

```
                "type": "relation",
                "value": "sold"
              }
            ],
            "units": "KB",
            "values": {
              "max": 20971.52,
              "min": 0,
              "avg": 1048.58,
              "total": 2097.15
            },
            "value": 1048.58
          }
        ]
      },
      ...
    ]
  }
]
```

Complex statistic queries

Using the POST `/stats` request, you can retrieve a variety of statistics. For example, this is a complex query made by the Workload Optimization Manager UI:

```
{
  "statistics":[
    {
      "name":"CPU",
      "relatedEntityType":"PhysicalMachine",
      "groupBy":[
        "key",
        "virtualDisk"
      ],
      "filters":[
        {
          "type":"relation",
          "value":"sold"
        }
      ]
    },
    {
      "name":"Mem",
      "relatedEntityType":"PhysicalMachine",
      "groupBy":[
        "key",
        "virtualDisk"
      ],
      "filters":[
        {
          "type":"relation",
          "value":"sold"
        }
      ]
    },
    {
      "name":"NetThroughput",
      "relatedEntityType":"PhysicalMachine",
      "groupBy":[
        "key",
        "virtualDisk"
      ],
      "filters":[
```

```
      {
        "type":"relation",
        "value":"sold"
      }
    ]
  },
  {
    "name":"IOThroughput",
    "relatedEntityType":"PhysicalMachine",
    "groupBy":[
      "key",
      "virtualDisk"
    ],
    "filters":[
      {
        "type":"relation",
        "value":"sold"
      }
    ]
  },
  {
    "name":"Q1VCPU",
    "relatedEntityType":"PhysicalMachine",
    "groupBy":[
      "key",
      "virtualDisk"
    ],
    "filters":[
      {
        "type":"relation",
        "value":"sold"
      }
    ]
  },
  {
    "name":"Q2VCPU",
    "relatedEntityType":"PhysicalMachine",
    "groupBy":[
      "key",
      "virtualDisk"
    ],
    "filters":[
      {
        "type":"relation",
        "value":"sold"
      }
    ]
  },
  {
    "name":"Q4VCPU",
    "relatedEntityType":"PhysicalMachine",
    "groupBy":[
      "key",
      "virtualDisk"
    ],
    "filters":[
      {
        "type":"relation",
        "value":"sold"
      }
    ]
  },
  {
    "name":"Q8VCPU",
    "relatedEntityType":"PhysicalMachine",
```

```
    "groupBy":[
      "key",
      "virtualDisk"
    ],
    "filters":[
      {
        "type":"relation",
        "value":"sold"
      }
    ]
  },
  {
    "name":"Q16VCPU",
    "relatedEntityType":"PhysicalMachine",
    "groupBy":[
      "key",
      "virtualDisk"
    ],
    "filters":[
      {
        "type":"relation",
        "value":"sold"
      }
    ]
  },
  {
    "name":"Q32VCPU",
    "relatedEntityType":"PhysicalMachine",
    "groupBy":[
      "key",
      "virtualDisk"
    ],
    "filters":[
      {
        "type":"relation",
        "value":"sold"
      }
    ]
  },
  {
    "name":"Ballooning",
    "relatedEntityType":"PhysicalMachine",
    "groupBy":[
      "key",
      "virtualDisk"
    ],
    "filters":[
      {
        "type":"relation",
        "value":"sold"
      }
    ]
  },
  {
    "name":"Swapping",
    "relatedEntityType":"PhysicalMachine",
    "groupBy":[
      "key",
      "virtualDisk"
    ],
    "filters":[
      {
        "type":"relation",
        "value":"sold"
      }
```

```
      ]
    },
    {
      "name":"CPUProvisioned",
      "relatedEntityType":"PhysicalMachine",
      "groupBy":[
        "key",
        "virtualDisk"
      ],
      "filters":[
        {
          "type":"relation",
          "value":"sold"
        }
      ]
    },
    {
      "name":"MemProvisioned",
      "relatedEntityType":"PhysicalMachine",
      "groupBy":[
        "key",
        "virtualDisk"
      ],
      "filters":[
        {
          "type":"relation",
          "value":"sold"
        }
      ]
    },
    {
      "name":"Flow",
      "relatedEntityType":"PhysicalMachine",
      "groupBy":[
        "key",
        "virtualDisk"
      ],
      "filters":[
        {
          "type":"relation",
          "value":"sold"
        }
      ]
    },
    {
      "name":"FlowAllocation",
      "relatedEntityType":"PhysicalMachine",
      "groupBy":[
        "key",
        "virtualDisk"
      ],
      "filters":[
        {
          "type":"relation",
          "value":"sold"
        }
      ]
    },
    {
      "name":"Cooling",
      "relatedEntityType":"PhysicalMachine",
      "groupBy":[
        "key",
        "virtualDisk"
      ],
```

```
      "filters":[
        {
          "type":"relation",
          "value":"sold"
        }
      ]
    },
    {
      "name":"Power",
      "relatedEntityType":"PhysicalMachine",
      "groupBy":[
        "key",
        "virtualDisk"
      ],
      "filters":[
        {
          "type":"relation",
          "value":"sold"
        }
      ]
    },
    {
      "name":"Space",
      "relatedEntityType":"PhysicalMachine",
      "groupBy":[
        "key",
        "virtualDisk"
      ],
      "filters":[
        {
          "type":"relation",
          "value":"sold"
        }
      ]
    }
  ]
}
```

The previous request is made in context of a datacenter scope, so the response includes information for each host within that datacenter. The `virtualDisk` value for the `groupBy` parameter ensures that all virtual disks appear in the response as individual line items, rather than an aggregation.

Additionally, the `"filters":[{"type":"relation","value":"sold"}]` found throughout the call indicates that the response includes those commodities *sold* by these hosts, rather than purchased from a lower level in the supply chain.

Supplychains endpoint

Workload Optimization Manager models your environment as a market of buyers and sellers. It discovers different types of entities in your environment via the targets you have configured for your installation. Discovery maps these entities to the supply chain so Workload Optimization Manager can monitor them and manage the workloads they support. For example, for a hypervisor target Workload Optimization Manager discovers VMs, the hosts and datastores that provide resources to the VMs, and the applications that use VM resources. The entities in your environment form a chain of supply and demand where some entities provide resources while others consume the supplied resources.

Using the supplychains endpoint, you can:

- Get a full or filtered supply chain related to an entity in your environment
- Get the statistics for the supply chain related to an entity in your environment

Supplychains requests

Getting a supply chain for the specified entity

Gets the full supply chain for the specified entity. For example, if the UUID of a virtual machine is used, you will see the applications that are hosted by that virtual machine, the hosts and datastores that it consumes from, the cluster or datacenter that the hosts consume from, the disk arrays and storage controllers that the datastores consume from, and so on.

In a supply chain, depth refers to the distance in the supply chain from the entity called. For example, if the UUID of a host is used, virtual machines consuming from that host will have a depth of `-1`, applications consuming from those virtual machines will have a depth of `-2`, other hosts that are in the supply chain will have a depth of `0`, and the cluster of the host will have a depth of `1`.

Example: `GET https://10.10.10.10/api/v3/supplychains?uuids=Virtual_ESX_4238042f-7465-e92a-5c28-a7a929307247&health=false`

Response: A `SupplychainApiDTO` representing the supply chain of the specified entity– in this case, a host. Note that when there is a single instance of an entity, Workload Optimization Manager will return the details of that entity. In the response, you can see this with the `Storage Controller` entity.


```
{
  "seMap":{
    "Application":{
      "depth":-2,
      "entitiesCount":12,
      "stateSummary":{
        "ACTIVE":5,
        "SUSPEND":1,
        "IDLE":6
      }
    },
    ...
 "StorageController":{
      "depth":3,
      "entitiesCount":1,
      "stateSummary":{
        "ACTIVE":1
      },
      "instances":{
        "be89058c-e786-11e4-82ba-dfc30ffeb3ee":{
          "links":[
            {
              "rel":"self",

 "href":"https://10.10.10.10/api/v3/entities/be89058c-e786-11e4-82ba-dfc30ffeb3ee?include_aspects=false"
            }
          ],
          "uuid":"be89058c-e786-11e4-82ba-dfc30ffeb3ee",
          "displayName":"dataontap-mycorp-cm1",
          "className":"StorageController",
          "priceIndex":20000,
          "state":"ACTIVE",
          "severity":"Normal",
          "discoveredBy":{
            "uuid":"_adGLMJwFEei_xJ5Ai6-rnQ",
            "displayName":"data-cluster.eng.mycorp.com",
            "type":"NetApp"
          },
          "remoteId":"be89058c-e786-11e4-82ba-dfc30ffeb3ee",
          "environmentType":"ONPREM"
        }
      }
    }
```

```
  }
}
```

Getting leaf entities from a supply chain root

Gets an array of entities that are at the top for a given supply chain scope. You provide an array of UUIDs to *seed* the starting scope. The request then traverses up the supply chain from those seed entities, until it finds entities that have no consumers. These are *leaf* entities for that seeded scope.

- uuids

  A list of UUIDs to seed the request scope. For example, you can give uuids for multiple VMs, and the request will find all leaf entities from all VMs in the list. If you give the uuid for a group, the request treats the group members as a list of entities.

- filterEntityTypes

  A list of entity types to filter out of the request. You can provide any valid entity class name. The request stops traversing the supply chain at the filtered entity. For example, assume you have a supply chain that grows from 1 VM > 4 ApplicationComponent > 2 Service > 1 BusinessApplication. A request starting from the VM would return the single Business Application. If you filter out BusinessApplication, it would return the two Service entities. If you filter out Service, it would return 4 ApplicationComponent entities. Note that the Swagger documentation for this call gives a list of valid filterEntityTypes.

- cursor

  When making subsequent calls to retrieve additional results, this is the cursor returned by the last call to this method.

- limit

  The maximum number of items to return. Must be a positive integer. If not set, the server will provide a service-specific maximum limit.

For example, in the Workload Optimization Manager user interface, you can search for the BussinessApplication entity named "App [app]". If you set the scope to that business application, then you can click in the Supply Chain to select the Namespace entity. Find the **Entity Information** chart and copy the ID value. That is the uuid for the app Namespace. If you get the leaf entities for that Namespace, you can see all the Service, ApplicationComponent, and BusinessApplication entities that support the Workload Optimization Manager application. For example, the following Service entry is for the Hyper-V target component:

```
  {
    "uuid": "74626820648480",
    "displayName": "Service-app/mediation-hyperv",
    "className": "Service"
  }
```

Now assume you want to use a VM group as the seeded scope, and you want to exclude any Service entities or higher in the Supply Chain:

Example: GET https://10.10.168.102/api/v3/supplychains/leafEntities?uuids=74626820647952&filterEntityTypes=Service

```
[
  {
    "uuid": "74699295824896",
    "displayName": "App-app/cost-67b7b9bb4d-bl8kt/cost",
    "className": "Application Component"
  },
  {
    "uuid": "74699295824908",
    "displayName": "App-app/mediation-vmax-6cd7f4ff-nzqp6/mediation-vmax",
    "className": "Application Component"
  },
  {
    "uuid": "74699295824897",
    "displayName": "App-app/mediation-azurebilling-c585c846f-mvnl7/mediation-azurebilling",
```

```
    "className": "Application Component"
  },
  ...
]
```

Getting statistics for a supply chain

Gets a list of statistics filtered by a user-created SupplyChainStatsApiInputDTO. For more information about the criteria that make up the input for thi request, see the Swagger documentation for this request, located at `https://<Your_Workload Optimization Manager_URL/apidpoc`.

Example: POST https://10.10.168.102/api/v3/supplychains/stats

Example Input: This input will return the number of applications running on the specified UUID (in this case, a Datacenter):

```
{
  "uuids": [
    "e27e5c2f85d91107ac3f26fd63d4478e808f590b"
  ],
  "types": [
    "Application"
  ],
  "environmentType": "ONPREM"
}
```

Response: A list of `StatSnapshotApiDTO`s representing the filtered statistics for the entity:

```
[
  {
    "date": "2019-05-01T17:45:42-04:00",
    "statistics": [
      {
        "name": "entities",
        "filters": [],
        "value": 25
      }
    ]
  }
]
```

Tags endpoint

Tags overview

Tags are metadata attached to resources that you can use to organize your environment. Tags are most often used to delineate resource groupings for technical, automation, business, or security purposes.

Workload Optimization Manager discovers tags for AWS, Azure, and vCenter resources.

Using the tags endpoint, you can:

- Get a list of available tags
- Get the members belonging to a specified tag

Tags requests

Pagination for tags requests

When you request tag data, the API uses pagination to return subsets of the data. You should always use pagination when available to make sure you get your data in manageable chunks. To paginate returned data, the `limit`, `orderBy`, and `cursor` parameters control how the APIs paginate their data. If you do not specify a `limit` for the following requests, the API will paginate the data to a default limit of records:

- `GET /tags`
- `GET /tags/entities?key=<tagKey>`

If your scripts make these requests without setting a limit, you should be sure they check for a pagination cursor in the returned data.

For general information about pagination, see Pagination in the Workload Optimization Manager API *(on page 1141)*.

Getting tags

For a list of tags, the API returns a list of TagApiDTOs. Each TagApiDTO will contain the key/value pair of the tag. The key can be used to see the entities belonging to the tag.

Example: `https://10.10.10.10/api/v3/tags`

Response:

```
[
  {
    "key":"owner",
    "values":[
      "QA_Azure"
    ]
  },
  {
    "key":"director",
    "values":[
      "p-bosh",
      "bosh-init"
    ]
  },
  ...
]
```

Filtering tags by scope, entity, and environment

Gets a filtered list of tags for the specified scope, entity type, and/or environment. For example, you can use this to find tags only in a particular cluster, tags that affect at least one VM, and/or limit your result to only cloud entities. This request accepts the following parameters:

- scopes

  A list of UUIDs representing the scope of the request. For example, providing `Market` as a scope will search the entire real-time market. If you provide the UUID of a host, only entities on that host would be in the response.
- entity_type
  - `APPLICATION`
  - `APPLICATION_SERVER`
  - `BUSINESS_ACCOUNT`
  - `BUSINESS_UNIT`

  - CHASSIS
  - CLOUD_SERVICE
  - CONTAINER
  - CONTAINER_POD
  - DATABASE
  - DATABASE_SERVER
  - DATACENTER
  - DISK_ARRAY
  - IO_MODULE
  - LOAD_BALANCER
  - LOGICAL_POOL
  - NETWORK
  - PHYSICAL_MACHINE
  - STORAGE
  - STORAGE_CONTROLLER
  - SWITCH
  - VIRTUAL_APPLICATION
  - VIRTUAL_DATACENTER
  - VIRTUAL_MACHINE
- environment_type

  Returns only tags from entities in the specified environment type. `[ Cloud, Hybrid, Onprem ]`

Example: `GET https://10.10.10.10/api/v3/tags?scopes=Market&entity_type=VirtualMachine&environment_type=ONPREM`

Response: a list of VMs running on premises in the real-time market with tags.

Getting Entities For A Specified Tag

Referenced Endpoint: Entities (on page 1195)

Gets a list of entities that have the specified tag. An entity is any device in your environment that Workload Optimization Manager manages, such as applications, physical machines, VMs, storage, and datacenters.. You can use this information to see where entities with the specified tag are located, what other tags are associated to those entities, and detailed information about each entity.

Example: `GET https://10.10.10.10/api/v3/tags/entities?key=owner`

Response: A list of entities that have the tag `owner`:


```
   [
 {
   "links":[
     {
       "rel":"self",

 "href":"https://10.10.10.10/api/v3/entities/azure::VM::375ddfb4-3127-44eb-9088-3c7645212f81?include_aspec
ts=false"
     }
   ],
   "uuid":"azure::VM::375ddfb4-3127-44eb-9088-3c7645212f81",
   "displayName":"oleg-westus2-unmanaged-std",
   "className":"VirtualMachine",
   "priceIndex":3.2109845,
```

```
    "state":"ACTIVE",
    "severity":"Normal",
    "costPrice":0.024,
    "discoveredBy":{
      "uuid":"_GLpj4JwIEei_xJ5Ai6-rnQ",
      "displayName":"mycorp.management.core.windows.net",
      "type":"Azure"
    },
    "remoteId":"azure::VM::375ddfb4-3127-44eb-9088-3c7645212f81",
    "template":{
      "uuid":"azure::VMPROFILE::Basic_A0",
      "displayName":"Basic_A0",
      "discovered":false
    },
    "environmentType":"CLOUD",
    "tags":{
      "owner":[
        "PT_Azure"
      ],
      "AutoSuspend":[
        "false"
      ]
    }
  },
  {
    "links":[
      {
        "rel":"self",

 "href":"https://10.10.10.10/api/v3/entities/azure::VM::c94e96bf-975e-47da-8c52-7ac90161cc87?include_aspec
ts=false"
      }
    ],
    "uuid":"azure::VM::c94e96bf-975e-47da-8c52-7ac90161cc87",
    "displayName":"oleg-westus2-managed-avail-set-1",
    "className":"VirtualMachine",
    "priceIndex":2.446491,
    "state":"ACTIVE",
    "severity":"Normal",
    "costPrice":0.018,
    "discoveredBy":{
      "uuid":"_GLpj4JwIEei_xJ5Ai6-rnQ",
      "displayName":"mycorp.management.core.windows.net",
      "type":"Azure"
    },
    "remoteId":"azure::VM::c94e96bf-975e-47da-8c52-7ac90161cc87",
    "template":{
      "uuid":"azure::VMPROFILE::Standard_B1s",
      "displayName":"Standard_B1s",
      "discovered":false
    },
    "environmentType":"CLOUD",
    "tags":{
      "owner":[
        "PT_Azure"
      ],
      "AutoSuspend":[
        "false"
      ]
    }
  }
]
```

Targets endpoint

A target is a service that performs management functions in your virtual environment. Workload Optimization Manager uses targets to monitor workload and to execute actions in your environment.

To discover entities in your environment, you add different targets to your Workload Optimization Manager installation. Workload Optimization Manager then uses these targets to collect data from your environment.

Workload Optimization Manager uses probes to connect to targets and discover their entities. The Workload Optimization Manager GUI provides forms for users to specify different target instances. Note that each probe may require different settings, and it's possible that different versions of Workload Optimization Manager supports different sets of probes.

For complete target information, review Target Configuration *(on page 173)*. You should not modify targets without a full understanding of how they interact with Workload Optimization Manager.

Using the targets endpoint of the Workload Optimization Manager API, you can:

- Get a single target or list of targets present in your environment
- Get a list of probes available in your Workload Optimization Manager installation
- Get the actions or statistics for a target
- Get the workflows discovered by a target
- Add or edit a target
- Trigger validation or rediscovery on a target
- Delete a target

Targets utility requests

Utility requests provide metadata that you can use to make related requests, assemble inputDTOs, or see the potential values of a particular DTO.

The targets endpoint contains a utility request called `specs` that shows the information required for probe registry for each target available to Workload Optimization Manager, including required input fields and expected target type and category values.

> **NOTE:**
> For a list of probes available to your specific Workload Optimization Manager installation, review your organization's licensing agreement.

Example: `GET https://10.10.10.10/api/v3/targets/specs`

Response: An array of `TargetApiDTO` objects, where each object contains an array of input fields. This example shows one entry for the JVM probe in the Application Server category.

```
[
 {
  [
    ...
    {
      "category":"Application Server",
      "identifyingFields":[
        "nameOrAddress"
      ],
      "inputFields":[
        {
          "displayName":"Scope/Address",
          "name":"nameOrAddress",
          "isMandatory":true,
```

```
          "isSecret":false,
          "valueType":"GROUP_SCOPE",
          "description":"The group of VMs or the address of the JVM server"
        },
        {
          "displayName":"Port Number",
          "name":"targetPort",
          "defaultValue":"1099",
          "isMandatory":true,
          "isSecret":false,
          "valueType":"NUMERIC",
          "description":"The JMX port that connects with the JVM server"
        },
        {
          "displayName":"Username",
          "name":"username",
          "isMandatory":true,
          "isSecret":false,
          "valueType":"STRING",
          "description":"The JVM server JMX password, enter a space if the server has no security"
        },
        {
          "displayName":"Password",
          "name":"password",
          "isMandatory":true,
          "isSecret":true,
          "valueType":"STRING",
          "description":"The JVM server JMX port, enter a space if the server has no security"
        }
      ],
      "type":"JVM"
    },
    ...
  }
]
```

Targets requests

Target requests handle creation, configuration, validation, rediscovery, and deletion of targets on your Workload Optimization Manager appliance.

Pagination for targets requests

When you request target entities, the API uses pagination to return subsets of the data. You should always use pagination when available to make sure you get your data in manageable chunks. To paginate returned data, the `limit`, `orderBy`, and `cursor` parameters control how the APIs paginate their data.

If you do not specify a `limit` for the `GET /targets/<targetId>/entities` request, the API will paginate the data to a default limit of records. If your script makes this request without setting a limit, you should be sure it checks for a pagination cursor in the returned data.

For general information about pagination, see Pagination in the Workload Optimization Manager API *(on page 1141)*.

Getting targets

For a list of targets, the API returns an array of TargetApiDTO objects. Each TargetApiDTO object will contain the validation status, target type and category, and input field names and values. To get a single target, include the UUID of the target in the request.

This request takes the following parameters:

environment_type

Returns only targets in the specified environment type. `[ Cloud, Hybrid, Onprem ]`

Examples:

- All Targets: `GET https://10.10.10.10/api/v3/targets`
- Single Target: `GET https://10.10.10.10/api/v3/targets/__xwWYAE33eku7PbzHdETIA`

Response: A single entry in an array of targets. Notice that the password field is secret, so the response does not contain the password value.


```
[
  {
  "uuid":"__xwWYAE33eku7PbzHdETIA",
  "category":"Hypervisor",
  "type":"vCenter",
  "status":"Validated",
  "lastValidated":"2018-06-26T14:25:08-04:00",
  "inputFields":[
    {
      "displayName":"Address",
      "name":"address",
      "value":"vsphere-demo.mycorp.com",
      "isMandatory":true,
      "isSecret":false,
      "valueType":"STRING",
      "description":"IP or address of vCenter Server",
      "verificationRegex":".*"
    },
    {
      "displayName":"Username",
      "name":"username",
      "value":"corp\\my.user",
      "isMandatory":true,
      "isSecret":false,
      "valueType":"STRING",
      "description":"User name to use to connect to a target",
      "verificationRegex":".*"
    },
    {
      "displayName":"Password",
      "name":"password",
      "isMandatory":true,
      "isSecret":true,
      "valueType":"STRING",
      "description":"Password to use to connect to a target",
      "verificationRegex":".*"
    }
  ]
```


Adding a target

Adds a target to your Workload Optimization Manager installation. You must pass a completed TargetApiDTO object as an inputDTO. For more information about the required fields of the TargetApiDTO for the target you want to add, execute the `/targets/specs` utility request.

Example: `POST https://10.10.10.10/api/v3/targets`

Example inputDTO:

```
{
  "category":"Hypervisor",
  "inputFields":[
    {
      "name":"username",
      "value":"corp\\my.name"
    },
    {
      "name":"password",
      "value":"Octopath"
    },
    {
      "name":"nameOrAddress",
      "value":"10.10.10.10"
    }
  ],
  "type":"vCenter"
}
```

Rediscover or validate all targets

Manually triggers a full rediscovery or validation of all targets in your environment. This is often done to immediately see multiple changes that would normally not be reflected until a daily full discovery is performed, such as displaying updated Orchestrator workflows.

This call takes the following parameters:

- validate

  When true, validates all targets in your environment
- rediscover

  When true, performs a full rediscovery on all targets in your environment

Examples:

- Rediscovery:

  `POST https://10.10.10.10/api/v3/targets?validate=true`
- Validation:

  `POST https://10.10.10.10/api/v3/targets?rediscover=true`

Rediscover or Validate A Single Target

Manually triggers a full rediscovery or validation of a single target in your environment. This is often done to immediately see a change that would normally not be reflected until a daily full discovery is performed, such as displaying updated Orchestrator workflows.

This call takes the following parameters:

- validate

  When true, validates the specified target
- rediscover

  When true, performs a full rediscovery on the specified target

Examples:

- Rediscovery: `https://10.10.10.10/api/v3/targets/__xwWYAE33eku7PbzHdETIA?rediscover=true`
- Validation: `https://10.10.10.10/api/v3/targets/__xwWYAE33eku7PbzHdETIA?validate=true`

Response: The `TargetApiDTO` object of the specified target:

```
{
  "uuid":"__xwWYAE3EeiNKPbzHdETIA",
  "category":"Hypervisor",
  "inputFields":[
    ...
  ],
  "lastValidated":"2018-06-27T21:45:56-04:00",
  "status":"Validated",
  "type":"vCenter"
}
```

Editing target input fields

Edits the input fields of an existing target, and accepts the modified TargetApiDTO as input.

> **NOTE:** You can GET the target first using the GET `/targets/specs` to see the fields required for that target. Note that you must pass all input fields in the modified TargetApiDTO object.

Example: PUT `https://10.10.10.10/api/v3/targets/_HgzQ4CdCEeaUWYEyo3vsyw`

Example input, editing the username for a single target:

```
[
  {
    "uuid":"_HgzQ4CdCEeaUWYEyo3vsyw",
    "category":"Hypervisor",
    "type":"vCenter",
    "status":"Validated",
    "lastValidated":1474575346109,
    "inputFields":[
      {
        "displayName":"Address",
        "name":"address",
        "value":"vsphere.mycorp.com",
        "isMandatory":true,
        "isSecret":false,
        "valueType":"STRING"
      },
      {
        "displayName":"Username",
        "name":"username",
        "value":"corp\\my.name.changed",
        "isMandatory":true,
        "isSecret":false,
        "valueType":"STRING"
      },
      {
        "displayName":"Password",
        "name":"password",
        "value":"Octopath",
        "isMandatory":true,
        "isSecret":true
      }
    ]
  }
]
```

Deleting a target

Deletes the specified target and removes all entities managed by that target. This will also remove any pending actions related to the deleted target.

Example: `DELETE https://10.10.10.10/api/v3/targets/__xwWYAE3EeiNKPbzHdETIA`

Response: A successful deletion request returns a Response Code of `200`.

Actions for a specified target

Referenced Endpoint: Actions (on page 1161)

These requests get full or filtered lists of actions per target for use in reports, or to find a particular action.

> **NOTE:** Once you have the action UUID, you can use requests from the actions endpoint to post changes to that action.

Getting an Unfiltered Array of Actions

Gets a list of current actions in your environment associated to entities discovered by the specified target.

Example: `GET https://10.10.10.10/api/v3/targets/__xwWYAE3EeiNKPbzHdETIA/actions`

Getting a Filtered Array of Actions

Gets a filtered list of actions discovered by a single target. Pass an inputDTO containing one or more of the following criteria:

- actionModeList
  - `DISABLED`:

    Do not recommend or perform the action. When you disable an action, the user interface will no longer display actions of that type.
  - `RECOMMENDED`:

    Recommend the action so a user can execute it via the given hypervisor or by other external means
  - `MANUAL`:

    Recommend the action, and provide the option to execute that action through the Workload Optimization Manager user interface or an /actions API request.
  - `AUTOMATED`:

    You have directed Workload Optimization Manager to execute the action automatically.
  - `COLLECTION` **(UNUSED)**:

    Legacy action mode, no longer used.
- actionStateList
  - `DISABLED`:

    An action whose policy has been disabled.
  - `RECOMMEND`:

    An action that cannot be automated due to policy settings or must be executed outside the system, such as the purchase of hardware.
  - `PENDING_ACCEPT`:

    When the action mode is set to manual, a recommended action that has not been accepted or cleared.
  - `ACCEPTED`:

    An action that has been accepted, but is not yet in progress.
  - `IN_PROGRESS`:

    An action that is being executed.

  - SUCCEEDED:

    An action that was accepted and completed successfully.
  - FAILED:

    An action that was accepted and did not complete successfully.
  - REJECTED:

    When the action mode is set to manual, an action that has been rejected by the user.
  - CLEARED:

    When the action mode is set to manual, an action that is no longer recommended by the market.
  - ACCOUNTING:

    For cloud entities, an action to resize in order to use a different instance template.
  - QUEUED:

    When more than 10 actions are to be executed for a single target, Workload Optimization Manager will place the 11th and all subsequent actions into a QUEUED state, and they will execute as the previous actions complete.
- actionTypeList
  - START:

    Start an entity, and add it to the Workload Optimization Manager market.
  - MOVE:

    Move an entity from one provider to another. For example, moving a VM between hosts, or a datastore between disk arrays.
  - SUSPEND:

    Suspend the entity to exclude its capacity from your environment. Suspend does not remove the entity from your environment, but the market does not include the entity in its analysis. Based on market conditions, analysis can recommend restarting suspended entities, which includes them in market analysis once again.
  - ADD PROVIDER:

    This is equivalent to the START action, but for storage entities.
  - CHANGE:

    Move a virtual machine's storage between datastores.
  - PROVISION:

    Add capacity to your environment.
  - RECONFIGURE:

    Analysis recommends this when a VM is not able to correct a misconfiguration via another type of action. For example, analysis can recommend a RECONFIGURE action when a placement policy requires a VM to move to a host on a particular cluster, but cross-cluster moves are not permitted.
  - DELETE:

    This is equivalent to the SUSPEND action, but for storage entities.
  - RESERVE_ON_PM:

    When using the reservation functionality, this is the action to reserve the resources of a VM on a particular host.
  - RESERVE_ON_DS:

    When using the reservation functionality, this is the action to reserve the resources of a VM on a particular datastore.
  - RESIZE_FOR_EFFICIENCY:

    A resize that reduces the capacity of a commodity, such as vCPU, or vMEM.
  - RESIZE_FOR_PERFORMANCE:

    A resize that increases the capacity of a commodity, such as vCPU, or vMEM.

- cleared

  Default: `false`.

  When `true`, actions that were removed from the action list before being executed will be returned.
- costType
  - `Saving`

    An action that will decrease cloud spending.
  - `Investment`

    An action that will increase cloud spending.
- cumulative

  Default: `false`.

  When `true`, shows the savings for each action in a scope as a cumulative total, rather than showing the savings for individual actions.
- environmentType
  - `ONPREM`:

    Returns only actions related to entities that are part of your on-prem environment.
  - `CLOUD`:

    Returns only actions related to entities that are part of your cloud environment.
- groupBy

  You may group the actions returned by the following criteria: `[actionModes, actionStates, actionTypes, risk, riskSeverity, riskSubCategory]`
- hasReservedInstance

  Default: `false`.

  When `true`, only actions that are related to reserved instances will be returned.
- relatedEntityTypes

  Returns only actions related to entities of these types within the scope. For example, if your scope is a virtual datacenter and you specify VirtualMachine as a related entity type, the request will include only those actions related to VMs in the specified virtual datacenter.

  Choose from the following entity types: `[ Datacenter, PhysicalMachine, VirtualMachine, Storage, Application, Chassis, DiskArray, IOModule, StorageControl, Switch, VirtualDataCenter, VPod, DPod, Container, Database, DatabaseServerContainer, LogicalPool ]`
- riskSeverityList

  Returns only actions of the given severities: `[ Unknown, Normal, Minor, Major, Critical ]`
- riskSubCategoryList
- returns only actions of the given subcategories: `[ Performance Assurance, Efficiency Improvement, Prevention, Compliance ]`

Example: `POST https://10.10.10.10/api/v3/targets/__xwWYAE3EeiNKPbzHdETIA/actions`

Example inputDto, filtering for actions that are pending acceptance:

```
{
  "actionStateList":[
    "PENDING_ACCEPT"
  ]
}
```

Response: An array of actions applicable to the specified target that are in the `PENDING_ACCEPT` state.

```
[
  {
    "links":[
      {
        "rel":"self",
        "href":"https://10.10.10.10/api/v3/actions/_uictIJuHEeics6ASgJ-kdg"
      },
      {
        "rel":"self",
        "href":"https://10.10.10.10/api/v3/actions/2204259249360"
      }
    ],
    "uuid":"_uictIJuHEeics6ASgJ-kdg",
    "createTime":"2018-08-08T23:53:08-04:00",
    "actionType":"RIGHT_SIZE",
    "actionState":"PENDING_ACCEPT",
    "actionMode":"MANUAL",
    "details":"Scale down VMem for VirtualMachine btc-1 from 8 GB to 7 GB",
    "importance":0,
    "target":{
      "uuid":"4223dcd6-0b28-6ae6-1066-689912ced3a1",
      "displayName":"btc-1",
      "className":"VirtualMachine",
      "aspects":{
        "virtualMachineAspect":{
          "os":"Other Linux (64-bit)",
          "connectedNetworks":[
            {
              "uuid":"74c91ba4fa390030f47f91e6e1ee77ff7b4db353",
              "displayName":"VM Network"
            }
          ],
          "numVCPUs":1,
          "ebsOptimized":false
        }
      },
      "environmentType":"ONPREM"
    },
    "currentEntity":{
      "uuid":"_ui4yAZuHEeics6ASgJ-kdg",
      "className":"VMem"
    },
    "newEntity":{
      "uuid":"_ui4yA5uHEeics6ASgJ-kdg",
      "className":"VMem"
    },
    "currentValue":"8388608.0",
    "newValue":"7340032.0",
    "resizeToValue":"7340032.0",
    "risk":{
      "uuid":"_ujB78ZuHEeics6ASgJ-kdg",
      "subCategory":"Efficiency Improvement",
      "description":"Underutilized VMem in Virtual Machine 'btc-1'",
      "severity":"MINOR",
      "reasonCommodity":"VMem",
      "importance":0
    },
    "stats":[
      {
        "name":"costPrice",
        "filters":[
          {
            "type":"savingsType",
            "value":"savings"
          }
```

```
        ],
        "units":"$",
        "value":50
      }
    ],
    "actionID":2204259249360
  },
  ...
  }
]
```

Entities for a specified target

Referenced Endpoint: Entities (on page 1195)

Gets a list of entities discovered by the specified target. This is used most often for reporting, dashboard, or inventory purposes.

Getting An Array of Entities: Gets a list of all entities discovered by the specified target.

An entity is any device in your environment that Workload Optimization Manager manages, such as applications, physical machines, VMs, storage, and datacenters.

Example:`GET https://10.10.10.10/api/v3/targets/__xwWYAE3EeiNKPbzHdETIA/actions`

Response: An array of `EntityApiDTO` objects belonging to a vCenter target:

```
[
  {
    "links":[
      {
        "rel":"self",

 "href":"https://10.10.10.10/api/v3/entities/vsphere-mycorp.com-ResourcePool-resgroup-8?include_aspects=fa
lse"
      }
    ],
    "uuid":"vsphere-mycorp.com-ResourcePool-resgroup-8",
    "displayName":"Resources-Cluster1",
    "className":"VirtualDataCenter",
    "priceIndex":1e-12,
    "state":"ACTIVE",
    "severity":"Normal",
    "discoveredBy":{
      "uuid":"__xwWYAE3EeiNKPbzHdETIA",
      "displayName":"vsphere-mycorp.com",
      "type":"vCenter"
    },
    "remoteId":"resgroup-8",
    "environmentType":"ONPREM"
  },
  {
    "links":[
      {
        "rel":"self",

 "href":"https://10.10.10.10/api/v3/entities/vsphere-mycorp.com-GuestLoad-421ddc94-8d0c-504a-62f5-493802e9
6d0a?include_aspects=false"
      }
    ],
    "uuid":"vsphere-mycorp.com-GuestLoad-421ddc94-8d0c-504a-62f5-493802e96d0a",
    "displayName":"GuestLoad[Nexus1000V]",
    "className":"Application",
    "priceIndex":1,
```

```
    "state":"ACTIVE",
    "severity":"Normal",
    "discoveredBy":{
      "uuid":"__xwWYAE3EeiNKPbzHdETIA",
      "displayName":"vsphere-mycorp.com",
      "type":"vCenter"
    },
    "remoteId":"vm-51",
    "environmentType":"ONPREM"
  },
  ...
  }
]
```

Templates endpoint

Workload Optimization Manager uses templates to reserve resources and deploy workload in your environment, to calculate supply or demand changes in a plan, and to calculate workloads for cloud environments.

Using the templates endpoint, you can:

- Get a list of templates available to your Workload Optimization Manager appliance
- Get information needed to create templates, such as supported OSes and CPUs
- Create a template
- Edit and Delete templates

Templates utility requests

Utility requests provide metadata that you can use to make related requests, assemble inputDTOs, or see the potential values of a particular DTO.

The templates endpoint contains utility requests that show the possible values for several optional fields used to create detailed templates. These can be used to find the supported cloud OSes and CPU models.

Getting a list of CPU models

Gets a list of valid CPU Models. If given a CPU Model, Workload Optimization Manager can more accurately reflect your CPU capacity in market decisions.

Example: `GET https://10.10.10.10/api/v3/templates/physicalmachine/cpumodels`

Result: A list of CpuModelApiDTOs, describing the cores, speed, and internal scaling factor for each CPU. This information will be used to provide more granular CPU data, if passed as part of a TemplateApiInputDTO:

```
[
  {
    "numCores": 4,
    "speed": 3500,
    "scalingFactor": 1.116909571123129,
    "modelName": "AMD A10 PRO-7800B"
  },
  {
    "numCores": 4,
    "speed": 3700,
    "scalingFactor": 0.9745581799084503,
    "modelName": "AMD A10 PRO-7850B"
  },
 ...
```

```
   {
    "numCores": 8,
    "speed": 1415,
    "scalingFactor": 1.1638948014750774,
    "modelName": "UltraSPARC T2 Plus"
  }
]
```

Templates requests

Template requests handle creation, configuration, editing and deletion of templates on your Workload Optimization Manager appliance.

Getting a List of Templates

Gets a list of templates available to your Workload Optimization Manager instance. This list will include both templates discovered from underlying targets and those created through the Workload Optimization Manager UI. To get a specific template, pass the template UUID as part of the request.

Examples:

- All Templates: `GET https://10.10.10.10/api/v3/templates`
- Single Target: `GET https://10.10.10.10/api/v3/templates/_v0Q70MpiEd-hypXfJzX8Wg`

Response: A list of `TemplateApiDTO`s:

```
[
  {
    "links": [
      {
        "rel": "self",
        "href": "https://10.10.10.10/api/v3/templates/_v0Q70MpiEd-hypXfJzX8Wg"
      }
    ],
    "uuid": "_v0Q70MpiEd-hypXfJzX8Wg",
    "displayName": "Small",
    "className": "PhysicalMachineProfile",
    "price": 5000,
    "computeResources": [
      {
        "stats": [
          {
            "name": "numOfCores",
            "value": 4
          },
          {
            "name": "cpuSpeed",
            "units": "MHz",
            "value": 2992
          },
          {
            "name": "ioThroughputSize",
            "units": "MB/s",
            "value": 890
          },
          {
            "name": "memorySize",
            "units": "GB",
            "value": 32759
          },
          {
```

```
            "name": "networkThroughputSize",
            "units": "MB/s",
            "value": 512
          }
        ]
      }
    ],
    "infrastructureResources": [
      {
        "stats": [
          {
            "name": "powerSize",
            "value": 1
          },
          {
            "name": "spaceSize",
            "value": 1
          },
          {
            "name": "coolingSize",
            "value": 1
          }
        ]
      }
    ],
    "discovered": false
  },
  {
    "links": [
      {
        "rel": "self",
        "href": "https://10.10.10.10/api/v3/templates/_v0Q70cpiEd-hypXfJzX8Wg"
      }
    ],
    "uuid": "_v0Q70cpiEd-hypXfJzX8Wg",
    "displayName": "Medium",
    "className": "PhysicalMachineProfile",
    "price": 10000,
    "computeResources": [
      {
        "stats": [
          {
            "name": "numOfCores",
            "value": 4
          },
          {
            "name": "cpuSpeed",
            "units": "MHz",
            "value": 2992
          },
          {
            "name": "ioThroughputSize",
            "units": "MB/s",
            "value": 1780
          },
          {
            "name": "memorySize",
            "units": "GB",
            "value": 65519
          },
          {
            "name": "networkThroughputSize",
            "units": "MB/s",
            "value": 512
          }
```

```
        ]
      }
    ],
    "infrastructureResources": [
      {
        "stats": [
          {
            "name": "powerSize",
            "value": 1
          },
          {
            "name": "spaceSize",
            "value": 1
          },
          {
            "name": "coolingSize",
            "value": 1
          }
        ]
      }
    ],
    "discovered": false
  },
 ...
]
```

Creating a template

Creates a template for use in your environment. You must pass a completed TemplateApiInputDTO as the inputDTO. You can create templates for virtual machines, physical machines, storage, and containers.

Example: `POST https://10.10.10.10/api/v3/templates`

Example inputs:

- Container

```
{
  "computeResources":[
    {
      "stats":[
        {
          "name":"ioThroughput",
          "value":0
        },
        {
          "name":"networkThroughput",
          "value":0
        },
        {
          "name":"memorySize",
          "value":8192
        },
        {
          "name":"cpuSpeed",
          "value":1536
        }
      ]
    }
  ],
  "storageResources":[
    {
```

```
      "stats":[
        {
          "name":"diskSize",
          "value":20
        }
      ]
    }
  ],
  "className":"ContainerProfile",
  "displayName":"ContainerTemplate1"
}
```

- Physical Machine

```
{
  "cpuModel":"AMD A10-5700",
  "computeResources":[
    {
      "stats":[
        {
          "name":"ioThroughputSize",
          "value":1000
        },
        {
          "name":"networkThroughputSize",
          "value":1000
        },
        {
          "name":"memorySize",
          "value":8192
        },
        {
          "name":"numOfCores",
          "value":4
        },
        {
          "name":"cpuSpeed",
          "value":"3400"
        }
      ]
    }
  ],
  "price":0,
  "className":"PhysicalMachineProfile",
  "displayName":"HostTemplate1"
}
```

- Storage

```
{
  "storageResources":[
    {
      "stats":[
        {
          "name":"diskIops",
          "value":250
        },
        {
          "name":"diskSize",
          "value":750
        }
```

```
      ]
    }
  ],
  "price":0,
  "className":"StorageProfile",
  "displayName":"StorageTemplate1"
}
```

- Virtual Machine

```
{
  "computeResources":[
    {
      "stats":[
        {
          "name":"ioThroughput",
          "value":0
        },
        {
          "name":"networkThroughput",
          "value":500
        },
        {
          "name":"memorySize",
          "value":4096
        },
        {
          "name":"memoryConsumedFactor",
          "value":70
        },
        {
          "name":"numOfCpu",
          "value":2
        },
        {
          "name":"cpuSpeed",
          "value":1024
        },
        {
          "name":"cpuConsumedFactor",
          "value":75
        }
      ]
    }
  ],
  "storageResources":[
    {
      "type":"disk",
      "stats":[
        {
          "name":"diskIops",
          "value":50
        },
        {
          "name":"diskSize",
          "value":20
        },
        {
          "name":"diskConsumedFactor",
          "value":70
        }
      ]
    }
```

```
    ]
  }
```

Response: The `TemplateApiInputDTO` for the template created. In this case, a virtual machine template:


```
{
  "uuid":"_lmaQoLUtEei5nORDMStFGA",
  "displayName":"VirtualMachineTemplate1",
  "className":"VirtualMachineProfile",
  "computeResources":[
    {
      "stats":[
        {
          "name":"numOfCpu",
          "value":2.0
        },
        {
          "name":"cpuSpeed",
          "units":"MHz",
          "value":1024.0
        },
        {
          "name":"cpuConsumedFactor",
          "units":"%",
          "value":75.0
        },
        {
          "name":"memorySize",
          "units":"MB",
          "value":4096.0
        },
        {
          "name":"memoryConsumedFactor",
          "units":"%",
          "value":70.0
        },
        {
          "name":"ioThroughput",
          "units":"MB/s",
          "value":0.0
        },
        {
          "name":"networkThroughput",
          "units":"MB/s",
          "value":500.0
        }
      ]
    }
  ],
  "storageResources":[
    {
      "stats":[
        {
          "name":"diskIops",
          "value":50.0
        },
        {
          "name":"diskSize",
          "units":"GB",
          "value":20.0
        },
        {
          "name":"diskConsumedFactor",
          "units":"%",
```

```
          "value":70.0
        }
      ],
      "type":"disk"
    }
  ],
  "deploymentProfile":{
    "uuid":"_llbZMrUtEei5nORDMStFGA",
    "displayName":"VirtualMachineTemplate1_DeploymentProfile",
    "className":"ServiceCatalogItem",
    "deployParameters":[
      {
        "targetType":"vCenter",
        "providers":[
          {
            "provider":{
              "uuid":"dfe657268310d8b128d441b2a34ec0721e72ee7e",
              "displayName":"Development DC7",
              "className":"DataCenter",
              "discoveredBy":{
                "uuid":"_OjF3UJwFEei_xJ5Ai6-rnQ",
                "displayName":"vsphere-dc7.mycorp.com",
                "type":"vCenter"
              }
            },
            "parameters":[
              {
                "parameterType":"image",
                "properties":[
                  {
                    "name":"id",
                    "value":"path/to/image"
                  }
                ]
              }
            ]
          }
        ]
      }
    ]
  },
  "discovered":false
}
```


Editing a template

Edits a user-created template, given the template UUID and the edited DTO. To edit discovered templates, edit them on the target from which they originate. Note that you must pass a complete TemplateApiDTO with the edited fields.

Example: `PUT https://10.10.10.10/api/v3/templates/_lmaQoLUtEei5nORDMStFGA`

Example input: Note that both the `displayName` and `diskConsumedFactor` values were changed:


```
{
  "computeResources":[
    {
      "stats":[
        {
          "name":"ioThroughput",
          "value":0
        },
        {
          "name":"networkThroughput",
```

```
          "value":500
        },
        {
          "name":"memorySize",
          "value":4096
        },
        {
          "name":"memoryConsumedFactor",
          "value":70
        },
        {
          "name":"numOfCpu",
          "value":2
        },
        {
          "name":"cpuSpeed",
          "value":1024
        },
        {
          "name":"cpuConsumedFactor",
          "value":75
        }
      ]
    }
  ],
  "storageResources":[
    {
      "type":"disk",
      "stats":[
        {
          "name":"diskIops",
          "value":50
        },
        {
          "name":"diskSize",
          "value":20
        },
        {
          "name":"diskConsumedFactor",
          "value":75
        }
      ]
    }
  ],
  "className":"VirtualMachineProfile",
  "displayName":"VirtualMachineTemplate1_Modified",
  "deploymentProfileId":"_llbZMrUtEei5nORDMStFGA"
}
```

Response: The modified `TemplateApiDTO` reflecting the submitted changes.

Deleting a template

Deletes the specified template. This will not delete associated deployment profiles, and deployments already scheduled to use this template will not succeed.

Example: `DELETE https://10.10.10.10/api/v3/templates/_lmaQoLUtEei5nORDMStFGA`

Response: A successful deletion will return `true`.

Topology definitions endpoint

Workload Optimization Manager gives you the ability to create your own Business Applications, Business Transactions, and Services without the need to ingest additional application data into the Workload Optimization Manager platform. This is especially useful in cases where there are gaps in the application stack shown in the Workload Optimization Manager supply chain. For example, in the absence of an application monitoring target such as AppDynamics or Dynatrace, you will not see Business Applications in your supply chain. This feature addresses those gaps.

When you create a new application entity, you identify interrelated application entities and nodes in your existing environment for which you want to measure performance, so Workload Optimization Manager can link them in a supply chain and represent them as a unified group. You can monitor overall performance for the group in the context of the new application entity, and drill down to the individual entities and nodes for finer details.

Using the topology definitions endpoint of the Workload Optimization Manager API, you can:

- Get a single topology definition or list of topology definitions present in your environment
- Create a topology definition
- Edit a topology definition
- Delete a topology definition

Topology definition requests

Topology definition requests handle creation, configuration, editing, and deletion of topology definitions on your Workload Optimization Manager appliance.

Getting Topology Definitions

For a list of topology definitions, the API returns an array of TopologyDataDefinitionApiDTO. Each TopologyDataDefinitionApiDTO will contain the UUID, entity type, and connection attributes for the defined topology. To get a single topology definition, include the UUID of the topology definition in the request.

Examples:

- All Topology Definitions: `GET https://10.10.10.10/api/v3/topologydefinitions`
- Single Topology Definition: `GET https://10.10.10.10/api/v3/topologydefinitions/284711321829968`

Response:

```
{
  "uuid":"284711313871856",
  "displayName":"Daltest1__App",
  "entityType":"BusinessApplication",
  "entityDefinitionData":{
    "namePrefix":"Daltest1_",
    "entityType":"VirtualMachine",
    "generationAndConnectionAttribute":{
      "expVal":"App",
      "expType":"EQ",
      "filterType":"virtualMachinesByTag",
      "caseSensitive":true,
      "entityType":null,
      "singleLine":false
    }
  }
}
```

Creating a manual topology definition

To create a manual topology definition, you must provide the `entityType` of the definition, and manual mapping data to the entities to use. Consider the following input DTO to create a manual topology definition:

```
{
  "entityType":"BusinessApplication",
  "entityDefinitionData":{
    "manualConnectionData":{
      "BusinessTransaction":{
        "staticConnections":[
          "73585155548499",
          "73585155548500"
        ]
      },
      "Service":{
        "staticConnections":[
          "73585155548487",
          "73585155608929"
        ]
      },
      "ApplicationComponent":{
        "staticConnections":[
          "73596151905427",
          "73596151905425",
          "73596151905428",
          "73596151905429"
        ]
      },
      "VirtualMachine":{
        "staticConnections":[
          "73586657216608",
          "73585300874097",
          "73586657216576"
        ]
      },
      "DatabaseServer":{
        "staticConnections":[
          "73585314633044",
          "73585314632869",
          "73585314632946"
        ]
      },
      "Container":{
        "staticConnections":[
          "73585323455544",
          "73582124234229",
          "73585385438946"
        ]
      },
      "ContainerPod":{
        "staticConnections":[
          "73585334243044",
          "73585314324535",
          "73585314533256"
        ]
      }
    }
  },
  "displayName":"DalManualtest_"
}
```

In the `manualConnectionData` parameter of the `entityDefinitionData`, the UUIDs of the related entities have been provided for each level of the supply chain for which the topology should be instantiated.

The creation can be as simple or complex as desired. For example, a simple manual topology may look like the following:

```
{
  "entityType":"BusinessTransaction",
  "entityDefinitionData":{
    "manualConnectionData":{
      "Service":{
        "staticConnections":[
          "73585155548487"
        ]
      },
      "Container":{
        "staticConnections":[
          "73596151905298"
        ]
      }
    }
  },
  "uuid":"284711514573216",
  "displayName":"DalManualSimple"
}
```

Creating an automated topology definition

To create an automated topology definition, you must provide the `entityType` of the definition, and the data that makes up the regular expression used to determine dynamic group membership. Consider the following input DTO to create an automated topology definition:

```
{
   "entityType":"BusinessApplication",
   "entityDefinitionData":{
      "entityType":"VirtualMachine",
      "namePrefix":"Daltest1_",
      "generationAndConnectionAttribute":{
         "expType":"EQ",
         "expVal":"App",
         "filterType":"vmsByTag",
         "caseSensitive":false
      }
   }
}
```

In the `generationAndConnectionAttribute` parameter of the `entityDefinitionData`, the information that makes up the regular expression is sent. In this instance, it will match any virtual machine whose tag value equals "App".

Editing a topology definition

Edits the input fields of an existing topology definition, and accepts the modified TopologyDataDefinitionApiDTO as input.

> **NOTE:** You can GET the topology definition first using the `GET /topologydefinition/UUID` request to see the current configuration. Note that you must pass all fields in the modified TopologyDataDefinitionApiDTO, not only the changed fields.

A successful edit will return the modified TopologyDataDefinitionApiDTO.

Example: `PUT https://10.10.10.10/api/v3/topologydefinitions`

Deleting a topology definition

Deletes the specified topology definition and removes all created entities. This will also remove any pending actions related to the deleted entities.

Example: `DELETE https://10.10.10.10/api/v3/topologydefinitions/284711321829968`

Response: A successful deletion request returns a Response Code of `200`.

Users endpoint

As an administrator, you specify accounts that grant users access to Workload Optimization Manager. Credentials for user accounts can either be saved locally on the Workload Optimization Manager server or managed externally through Microsoft Active Directory (AD) or another single sign-on (SSO) provider.

The user account configuration determines the following details:

- User authentication

  To configure an account, you set the type of authentication the account uses:

  - Local user: Configure the username and password and save those credentials on the Workload Optimization Manager server.
  - External user: Single user accounts can authenticate through SSO or through AD.
  - External group: Groups of user accounts can authenticate through SSO or AD.
- User authorization

  You can also specify properties that determine the range of access and features for a user:

  - Role: The user role determines access to specific Workload Optimization Manager features.
  - Scope: The user scope determines how much of the environment this user can manage.

You can access the users endpoint of the Workload Optimization Manager API to create and manage accounts for local users, external users, and external groups. You must have administrator or site administrator role privileges to access the users endpoint.

Managing local users

Administrators can access the users endpoint to create and manage local user accounts. Workload Optimization Manager stores the credentials for these accounts on the local server.

Adding local users

To add a local user, make a POST request to the users endpoint. The following parameters are required:

- `username`: Specify a username for the account.
- `password`: Specify the password for this user's account.
- `roles`: Specify a user role to grant the user access to specific Workload Optimization Manager features. Specify the role in uppercase letters.
- `loginProvider`: For a local user, specify `Local`.
- `type`: Specify `DedicatedCustomer`.

Optionally, you can specify one or more scopes for the user, which limit what the user can monitor in your environment. For example, you can scope to a group that contains only the physical machines that support this user's VMs or applications. This parameter is not valid for users who are in the ADMINISTRATOR role.

- `scope`: Specify one or more scopes by using the numeric UUID value for each scope. To get a list of the groups that are available to scope to in your Workload Optimization Manager instance, along with their display names and UUIDs, make a GET request to the search endpoint, for example:

  ```
  GET https://10.10.10.10/api/v3/search?types=Group
  ```

For more information about user roles and scopes, see Managing user accounts *(on page 1116)*.

Example:

```
POST https://10.10.10.10/api/v3/users
```

Example input

```
{
  "username": "myuser",
  "password": "mypassword",
  "roles": [
    {
      "name": "DEPLOYER"
    }
  ],
  "loginProvider": "Local",
  "type": "DedicatedCustomer",
  "scope": [
  {
    "uuid":"286125332284065"
  },
  {
    "uuid":"266725335284010"
  }
 ]
}
```

Response:

```
{
  "uuid": "4776186363888",
  "displayName": "myuser",
  "username": "mypassword",
  "roles": [
    {
      "name": "DEPLOYER"
    }
  ],
  "loginProvider": "Local",
  "type": "DedicatedCustomer",
  "scope": [
    {
      "uuid": "286125332284065",
      "isStatic": true,
      "logicalOperator": "AND",
      "groupClassName": "GroupApiDTO"
    }
    {
      "uuid": "266725335284010",
      "isStatic": true,
      "logicalOperator": "AND",
```

```
      "groupClassName": "GroupApiDTO"
    }
  ],
  "showSharedUserSC": false
}
```

Maintaining local users

To edit the details of an existing user or to delete a user, you must know the UUID value for the user.

- To get a list of all the currently configured user accounts, make a GET request to the users endpoint:

  ```
  GET https://10.10.10.10/api/v3/users
  ```

  The API returns an array of UserApiDTO *(on page 1708)* objects. Each `UserApiDTO` object includes the UUID, display name, username, roles, scope, and other details about a user account.

  Response:

  ```
  [
    {
      "uuid": "0000000000111",
      "displayName": "extTestUser",
      "username": "extTestUser",
      "roles": [
        {
          "name": "DEPLOYER"
        }
      ],
      "loginProvider": "LDAP",
      "type": "DedicatedCustomer",
      "showSharedUserSC": false
    },
    {
      "uuid": "0000000000112",
      "displayName": "externaluser1",
      "username": "externaluser1",
      "roles": [
        {
          "name": "OBSERVER"
        }
      ],
      "loginProvider": "LDAP",
      "type": "DedicatedCustomer",
      "showSharedUserSC": false
    },
    {
      "uuid": "0000000000113",
      "displayName": "administrator",
      "username": "administrator",
      "roles": [
        {
          "name": "ADMINISTRATOR"
        }
      ],
      "loginProvider": "Local",
      "type": "DedicatedCustomer",
      "showSharedUserSC": false
    },
    {
      "uuid": "0000000000114",
      "displayName": "newTestUser",
      "username": "newTestUser",
      "roles": [
  ```

```
      {
        "name": "DEPLOYER"
      }
    ],
    "loginProvider": "Local",
    "type": "DedicatedCustomer",
    "showSharedUserSC": false
  }
]
```

  - To get details about a specific user, append the UUID of the user to the request:

    ```
    GET https://10.10.10.10/api/v3/users/0000000000111
    ```

  - To get information about the currently logged-in user, append `/me` to the request:

    ```
    GET https://10.10.10.10/api/v3/users/me
    ```

- To edit the details for a user, make a PUT request to the users endpoint and include the user's UUID:

  ```
  PUT https://10.10.10.10/api/v3/users/0000000000114
  ```

  The following parameters are required in the request body, regardless of whether you want to modify them:
  - `username`
  - `roles`
  - `loginProvider`
  - `type`

  Example:

  Include the required parameters in the request body and specify the values that you want to modify. In the following example, the role for the `newTestUser` user is changed from DEPLOYER to OBSERVER. The other parameters remain unchanged:

  ```
  {
    "username": "newTestUser",
    "roles": [
      {
        "name": "OBSERVER"
      }
    ],
    "loginProvider": "Local",
    "type": "DedicatedCustomer"
  }
  ```

  Response:

  ```
  {
    "uuid": "0000000000114",
    "displayName": "newTestUser",
    "username": "newTestUser",
    "roles": [
      {
        "name": "OBSERVER"
      }
    ],
    "loginProvider": "Local",
    "type": "DedicatedCustomer",
    "showSharedUserSC": false
  }
  ```

- To delete a user, make a DELETE request to the `users` endpoint and include the user's UUID:

```
DELETE https://10.10.10.10/api/v3/users/0000000000111
```

Managing single sign-on users

Administrators can access the users endpoint to create and manage user accounts that authenticate by single sign-on (SSO) through an external identity provider (IdP). Workload Optimization Manager recognizes IdP responses that comply with SAML 2.0.

For information about authenticating with Microsoft Active Directory (AD), see Managing Active Directory users *(on page 1427)*.

Prerequisites

To enable SSO, you must configure access to the IdP. For information about configuring SSO, see Single Sign-On Authentication *(on page 100)*.

IMPORTANT:
For self-hosted Workload Optimization Manager instances, you can configure Workload Optimization Manager to use SSO authentication. When SSO is enabled, Workload Optimization Manager permits logins only through the SSO IdP. Whenever you navigate to your Workload Optimization Manager installation, it redirects you to the SSO Identity Provider (IdP) for authentication before it displays the Workload Optimization Manager user interface.

Before you enable SSO for your Workload Optimization Manager installation, *you must configure at least one SSO user with Workload Optimization Manager administrator privileges*. If you do not, then after you enable SSO, you cannot configure any SSO users in Workload Optimization Manager. To authorize an SSO user as an administrator, use **External authentication** to do one of the following options:

- Configure a single SSO user with administrator authorization.

  Add an external user. The username must match an account that the IdP manages.
- Configure an SSO user group with administrator authorization.

  Add an external group. The group name must match a user group on the IdP, and that group must have at least one member.

For information about configuring SSO user groups in SAML, see Configuring a Group for SSO Authentication *(on page 1126)*. For information about configuring SSO authentication for Workload Optimization Manager, see Single Sign-On Authentication *(on page 100)*.

To get information about the currently configured SSO provider, make a GET request to the `users/saml` endpoint:

```
GET https://10.10.10.10/api/v3/users/saml
```

Adding SSO users

To add an SSO user, make a POST request to the users endpoint. The following parameters are required:

- `username`: Specify a username for the account. The username must match a username that is specified on the IdP.
- `roles`: Specify a user role to grant the user access to specific Workload Optimization Manager features. Specify the role in uppercase letters.
- `loginProvider`: For an SSO user, specify LDAP.
- `type`: Specify `DedicatedCustomer`.

Optionally, you can specify one or more scopes for the user, which limit what the user can monitor in your environment. For example, you can scope to a group that contains only the physical machines that support this user's VMs or applications. This parameter is not valid for users who are in the ADMINISTRATOR role.

- `scope`: Specify one or more scopes by using the numeric UUID value for each scope. To get a list of the groups that are available to scope to in your Workload Optimization Manager instance, along with their display names and UUIDs, make a GET request to the search endpoint, for example:

  ```
  GET https://10.10.10.10/api/v3/search?types=Group
  ```

For more information about user roles and scopes, see Managing user accounts *(on page 1116)*.

Example:

```
POST https://10.10.10.10/api/v3/users
```

Example input

```
{
  "username": "mySsoUser",
  "roles": [
    {
      "name": "DEPLOYER"
    }
  ],
  "loginProvider": "LDAP",
  "type": "DedicatedCustomer",
  "scope": [
  {
    "uuid":"286125332284065"
  },
  {
    "uuid":"266725335284010"
  }
 ]
}
}
```

Response:

```
curl -X 'POST' \
  'https://10.10.10.10/api/v3/users' \
  -H 'accept: application/json' \
  -H 'Content-Type: application/json' \
  -d '{
  "username": "mySsoUser",
  "roles": [
    {
      "name": "DEPLOYER"
    }
  ],
  "loginProvider": "LDAP",
  "type": "DedicatedCustomer",
  "scope": [
    {
      "uuid": "286125332284065",
      "isStatic": true,
      "logicalOperator": "AND",
      "groupClassName": "GroupApiDTO"
    },
    {
      "uuid": "266725335284010",
```

```
      "isStatic": true,
      "logicalOperator": "AND",
      "groupClassName": "GroupApiDTO"
    }
  ],
  "showSharedUserSC": false
}'
```

Maintaining SSO Users

To edit the details of an existing user or to delete a user, you must know the UUID value for the user.

- To get a list of all the currently configured user accounts, make a GET request to the `users` endpoint:

  ```
  GET https://10.10.10.10/api/v3/users
  ```

  The API returns an array of UserApiDTO *(on page 1708)* objects. Each `UserApiDTO` object contains the universal unique identifier (UUID), display name, username, encoded password, roles, scope, and other details about a user account.

  Response:

  ```
  [
    {
      "uuid": "0000000000111",
      "displayName": "extTestUser",
      "username": "extTestUser",
      "roles": [
        {
          "name": "DEPLOYER"
        }
      ],
      "loginProvider": "LDAP",
      "type": "DedicatedCustomer",
      "showSharedUserSC": false
    },
    {
      "uuid": "0000000000112",
      "displayName": "externaluser1",
      "username": "externaluser1",
      "roles": [
        {
          "name": "OBSERVER"
        }
      ],
      "loginProvider": "LDAP",
      "type": "DedicatedCustomer",
      "showSharedUserSC": false
    },
    {
      "uuid": "0000000000113",
      "displayName": "administrator",
      "username": "administrator",
      "roles": [
        {
          "name": "ADMINISTRATOR"
        }
      ],
      "loginProvider": "Local",
      "type": "DedicatedCustomer",
      "showSharedUserSC": false
    },
    {
      "uuid": "0000000000114",
  ```

```
    "displayName": "newTestUser",
    "username": "newTestUser",
    "roles": [
      {
        "name": "DEPLOYER"
      }
    ],
    "loginProvider": "Local",
    "type": "DedicatedCustomer",
    "showSharedUserSC": false
  }
]
```

  - To get details about a specific user, append the UUID of the user to the request:

    ```
    GET https://10.10.10.10/api/v3/users/0000000000111
    ```

  - To get information about the currently logged-in user, append `/me` to the request:

    ```
    GET https://10.10.10.10/api/v3/users/me
    ```

- To edit the details for a user, make a PUT request to the `users` endpoint and include the user's UUID:

  ```
  PUT https://10.10.10.10/api/v3/users/0000000000111
  ```

  The following parameters are required, regardless of whether you want to modify them:
  - `username`
  - `roles`
  - `loginProvider`
  - `type`

  Include the required parameters in the request body and specify the values that you want to modify. In the following example, the role for the `extTestUser` user is changed from DEPLOYER to OBSERVER. The other parameters remain unchanged:

  Example:

  ```
  {
    "username": "extTestUser",
    "roles": [
      {
        "name": "OBSERVER"
      }
    ],
    "loginProvider": "LDAP",
    "type": "DedicatedCustomer"
  }
  ```

  Response:

  ```
  {
    "uuid": "0000000000111",
    "displayName": "extTestUser",
    "username": "extTestUser",
    "roles": [
      {
        "name": "OBSERVER"
      }
    ],
    "loginProvider": "LDAP",
    "type": "DedicatedCustomer",
  ```

```
  "showSharedUserSC": false
}
```

- To delete a user, make a DELETE request to the `users` endpoint and include the user's UUID:

```
DELETE https://10.10.10.10/api/v3/users/0000000000111
```

Adding SSO groups

To add an SSO group, make a POST request to the `/users/ldap/groups` endpoint. The following parameters are required:

- `displayName`: Specify the name of the external group that you are authorizing. The group name must be valid for the authentication system that you are relying on.
- `roleName`: Specify a role to grant group members access to specific Workload Optimization Manager features.
- `type`: Specify `DedicatedCustomer`.

Optionally, you can specify one or more scopes for the group, which limit what the group can monitor in your environment. For example, you can scope to a group that contains only the physical machines that support this group's VMs or applications. This parameter is not valid for groups that are in the `ADMINISTRATOR` role.

- `scope`: Specify one or more scopes by using the numeric UUID value for each scope. To get a list of the groups that are available to scope to in your Workload Optimization Manager instance, along with their display names and UUIDs, make a GET request to the search endpoint, for example:

```
GET https://10.10.10.10/api/v3/search?types=Group
```

For more information about user roles and scopes, see Managing user accounts *(on page 1116)*.

Example:

```
POST https://10.10.10.10/api/v3/users/ldap/groups
```

Example input

```
{
  "uuid": "mySsoGroup",
  "displayName: "mySsoGroup",
  "rolename": "DEPLOYER",
  "type": "DedicatedCustomer"
}
```

Response:

```
curl -X 'POST' \
  'https://10.10.10.10/api/v3/users' \
  -H 'accept: application/json' \
  -H 'Content-Type: application/json' \
  -d '{
  "uuid": "mySsoGroup",
  "displayName: "mySsoGroup",
  "rolename": "DEPLOYER",
  "type": "DedicatedCustomer"
}'
```

Maintaining SSO groups

To edit the details of an existing SSO group or to delete an SSO group, you must know the UUID value for the group.

- To get a list of all the currently configured SSO groups, make a GET request to the `users/ldap/groups` endpoint:

  ```
  GET https://10.10.10.10/api/v3/users/ldap/groups
  ```

  The API returns an array of ActiveDirectoryApiDTO *(on page 1502)* objects. Each `ActiveDirectoryGroupApiDTO` object contains the UUID, role, and other details about a group.

- To edit the details for a group, make a PUT request to the `users/ldap/groups` endpoint.

  ```
  PUT https://10.10.10.10/api/v3/users/ldap/groups
  ```

  The following parameters are required in the request body, regardless of whether you want to modify them:

  - `displayName`
  - `roles`
  - `type`

  Include the required parameters in the request body and specify the values that you want to modify. In the following example, the role for the `mySsoGroup` group is changed from DEPLOYER to OBSERVER. The other parameters remain unchanged:

  ```
  {
    "displayName": "mySsoGroup",
    "rolename": "OBSERVER",
    "type": "DedicatedCustomer"
  }
  ```

  Response:

  ```
  {
  "uuid": "mySsoGroup",
  "displayName": "mySsoGroup",
  "type": "DedicatedCustomer",
  "roleName": "DEPLOYER" }
  ```

- To delete a group, make a DELETE request to the `users/ldap/groups` endpoint and include the group's UUID:

  ```
  DELETE https://10.10.10.10/api/v3/users/ldap/groups/mySsoGroup
  ```

Managing Active Directory users

Administrators can access the users endpoint to create and manage user accounts and groups that authenticate by using Microsoft Active Directory (AD).

Prerequisites

To enable AD you must specify either an AD domain, an AD server, or both. Workload Optimization Manager uses this connection for all AD users.

Use a secure connection to communicate with AD servers. The AD domain must be configured to use LDAPS, and you must import a certificate into the Workload Optimization Manager server. Workload Optimization Manager can support LDAP channel binding and LDAP signing. To support these Active Directory features, you must configure secure access.

For more information, see Enforcing Secure Access *(on page 77)*.

To connect to an AD server, make a POST request to the `users/ldap` endpoint. Configure your request with the following parameters:

- `domainName`: Specify the domain for AD authentication.

  To authenticate AD groups, specify a domain so that AD can find a user with the User Principal Name (UPN). If you specify a domain, but not a server, authentication uses any AD server from that domain. If you want to support only AD users and not groups, the domain is optional.
- `loginProviderURI`: Specify the server for AD authentication.

  To disable AD groups, specify a server but do not specify a domain. If you specify a domain and a server, authentication uses that server and also supports groups.

  When you configure an AD server, by default Workload Optimization Manager assumes the AD server port to be 389 or 636. To specify a custom port for the AD server, add the port number to the AD server IP address. For example, `10.10.10.123:444` sets port 444.
- `isSecure`: Specify `true` for this setting to enable a secure connection to the AD server.

  Use a secure connection to communicate with AD servers. The AD domain must be configured to use LDAPS, and you must import a certificate into the Workload Optimization Manager server. Workload Optimization Manager can support LDAP channel binding and LDAP signing. To support these Active Directory features, you must configure secure access.

  For more information, see Enforcing Secure Access *(on page 77)*.

Example:

```
POST https://10.10.10.10/api/v3/users/ldap
```

Example input:

```
{
  "loginProviderURI": "10.10.10.123:444",
  "isSecure": "true"
}
```

To get information about the currently configured AD server and AD groups, make a GET request to the `users/ldap` endpoint.

Example:

```
GET https://10.10.10.10/api/v3/users/ldap
```

Response:

```
[
  {
    "loginProviderURI": "ldaps://10.10.10.123:444",
    "isSecure": true,
    "groups": [
      {
        "uuid": "adGroup",
        "displayName": "adGroup",
        "type": "DedicatedCustomer",
        "roleName": "ADVISOR"
      },
      {
        "uuid": "administrator",
        "displayName": "administrator",
        "type": "DedicatedCustomer",
        "roleName": "ADMINISTRATOR"
      },
      {
        "uuid": "advisor",
        "displayName": "advisor",
        "type": "DedicatedCustomer",
        "roleName": "ADVISOR"
```

```
      },
      {
        "uuid": "automator",
        "displayName": "automator",
        "type": "DedicatedCustomer",
        "roleName": "AUTOMATOR"
      },
      {
        "uuid": "deployer",
        "displayName": "deployer",
        "type": "DedicatedCustomer",
        "roleName": "DEPLOYER"
      },
      {
        "uuid": "guyexternalgrouptest",
        "displayName": "guyexternalgrouptest",
        "type": "DedicatedCustomer",
        "roleName": "OBSERVER"
      },
      {
        "uuid": "observer",
        "displayName": "observer",
        "type": "DedicatedCustomer",
        "roleName": "OBSERVER"
      },
      {
        "uuid": "operational_observer",
        "displayName": "operational_observer",
        "type": "DedicatedCustomer",
        "roleName": "OPERATIONAL_OBSERVER"
      },
      {
        "uuid": "site_admin",
        "displayName": "site_admin",
        "type": "DedicatedCustomer",
        "roleName": "SITE_ADMIN"
      },
      {
        "uuid": "testgroop",
        "displayName": "testgroop",
        "type": "DedicatedCustomer",
        "roleName": "OBSERVER"
      }
    ]
  }
]
```


Adding AD Users

To add an AD user, make a POST request to the users endpoint. The following parameters are required:

- `username`: Specify a username for the account. The username must be a valid User Principal Name (UPN). For example, `john@corp.mycompany.com`.
- `roles`: Specify a user role to grant the user access to specific Workload Optimization Manager features.
- `loginProvider`: For an AD user, specify LDAP.
- `type`: Specify `DedicatedCustomer`.

Optionally, you can specify one or more scopes for the user, which limit what the user can monitor in your environment. For example, you can scope to a group that contains only the physical machines that support this user's VMs or applications. This parameter is not valid for users who are in the ADMINISTRATOR role.

- `scope`: Specify one or more scopes by using the numeric UUID value for each scope. To get a list of the groups that are available to scope to in your Workload Optimization Manager instance, along with their display names and UUIDs, make a GET request to the `/search` endpoint, for example:

  ```
  GET https://10.10.10.10/api/v3/search?types=Group
  ```

For more information about user roles and scopes, see Managing user accounts *(on page 1116)*.

Example:

```
POST https://10.10.10.10/api/v3/users
```

Example input

```
{
  "username": "john@corp.mycompany.com",
  "roles": [
    {
      "name": "DEPLOYER"
    }
  ],
  "loginProvider": "LDAP",
  "type": "DedicatedCustomer"
  "scope": [
  {
    "uuid":"286125332284065"
  },
  {
    "uuid":"266725335284010"
  }
 ]
}
```

Response:

```
{
  "uuid": "4776186363888",
  "displayName": "john@corp.mycompany.com",
  "username": "john@corp.mycompany.com",
  "roles": [
    {
      "name": "DEPLOYER"
    }
  ],
  "loginProvider": "LDAP",
  "type": "DedicatedCustomer",
  "scope": [
    {
      "uuid": "286125332284065",
      "isStatic": true,
      "logicalOperator": "AND",
      "groupClassName": "GroupApiDTO"
    },
    {
      "uuid": "266725335284010",
      "isStatic": true,
      "logicalOperator": "AND",
      "groupClassName": "GroupApiDTO"
```

```
    }
  ],
  "showSharedUserSC": false
}
```

Maintaining AD Users

To edit the details of an existing user or to delete a user, you must know the UUID value for the user.

- To get a list of all the currently configured user accounts, make a GET request to the `users` endpoint:

  ```
  GET https://10.10.10.10/api/v3/users
  ```

  The API returns an array of UserApiDTO *(on page 1708)* objects. Each `UserApiDTO` object contains the universal unique identifier (UUID), display name, username, encoded password, roles, scope, and other details about a user account.

  Response:

  ```
  [
    {
      "uuid": "0000000000111",
      "displayName": "extTestUser",
      "username": "extTestUser",
      "roles": [
        {
          "name": "DEPLOYER"
        }
      ],
      "loginProvider": "LDAP",
      "type": "DedicatedCustomer",
      "showSharedUserSC": false
    },
    {
      "uuid": "0000000000112",
      "displayName": "externaluser1",
      "username": "externaluser1",
      "roles": [
        {
          "name": "OBSERVER"
        }
      ],
      "loginProvider": "LDAP",
      "type": "DedicatedCustomer",
      "showSharedUserSC": false
    },
    {
      "uuid": "0000000000113",
      "displayName": "administrator",
      "username": "administrator",
      "roles": [
        {
          "name": "ADMINISTRATOR"
        }
      ],
      "loginProvider": "Local",
      "type": "DedicatedCustomer",
      "showSharedUserSC": false
    },
    {
      "uuid": "0000000000114",
      "displayName": "newTestUser",
      "username": "newTestUser",
      "roles": [
        {
  ```

```
        "name": "DEPLOYER"
      }
    ],
    "loginProvider": "Local",
    "type": "DedicatedCustomer",
    "showSharedUserSC": false
  }
]
```

  - To get details about a specific user, append the UUID of the user to the request:

    ```
    GET https://10.10.10.10/api/v3/users/0000000000111
    ```

  - To get information about the currently logged-in user, append `/me` to the request:

    ```
    GET https://10.10.10.10/api/v3/users/me
    ```

- To edit the details for a user, make a PUT request to the `users` endpoint and include the user's UUID:

  ```
  PUT https://10.10.10.10/api/v3/users/0000000000111
  ```

  The following parameters are required, regardless of whether you want to modify them:
  - `username`
  - `roles`
  - `loginProvider`
  - `type`

  Include the required parameters in the request body and specify the values that you want to modify. In the following example, the role for the `extTestUser` user is changed from DEPLOYER to OBSERVER. The other parameters remain unchanged.

  Example:

  ```
  {
    "username": "extTestUser",
    "roles": [
      {
        "name": "OBSERVER"
      }
    ],
    "loginProvider": "LDAP",
    "type": "DedicatedCustomer"
  }
  ```

  Response:

  ```
  {
    "uuid": "0000000000111",
    "displayName": "extTestUser",
    "username": "extTestUser",
    "roles": [
      {
        "name": "OBSERVER"
      }
    ],
    "loginProvider": "LDAP",
    "type": "DedicatedCustomer",
    "showSharedUserSC": false
  }
  ```

- To delete a user, make a DELETE request to the `users` endpoint and include the user's UUID:

```
DELETE https://10.10.10.10/api/v3/users/0000000000111
```

Adding AD Groups

To add an AD group, make a POST request to the `/users/ldap/groups` endpoint. The following parameters are required:

- `displayName`: Specify the name of the group. The group name must match a group that is accessible from the AD domain and servers that you configured for your Workload Optimization Manager instance.
- `roleName`: Specify a role to grant group members access to specific Workload Optimization Manager features.
- `type`: Specify `DedicatedCustomer`.

Optionally, you can specify one or more scopes for the group, which limit what the group can monitor in your environment. For example, you can scope to a group that contains only the physical machines that support this group's VMs or applications. This parameter is not valid for groups that are in the ADMINISTRATOR role.

- `scope`: Specify one or more scopes by using the numeric UUID value for each scope. To get a list of the groups that are available to scope to in your Workload Optimization Manager instance, along with their display names and UUIDs, make a GET request to the `/search` endpoint, for example:

```
GET https://10.10.10.10/api/v3/search?types=Group
```

For more information about user roles and scopes, see Managing user accounts *(on page 1116)*.

Example:

```
POST https://10.10.10.10/api/v3/users/ldap/groups
```

Example input:

```
{
  "displayName: "myAdGroup",
  "rolename": "DEPLOYER",
  "type": "DedicatedCustomer"
}
```

Response:

```
curl -X 'POST' \
  'https://10.10.10.10/api/v3/users' \
  -H 'accept: application/json' \
  -H 'Content-Type: application/json' \
  -d '{
  "uuid": "myAdGroup",
  "displayName: "myAdGroup",
  "rolename": "DEPLOYER",
  "type": "DedicatedCustomer"
}'
```

NOTE:
If a user is a member of multiple groups, then Workload Optimization Manager logs the user on via the first SSO or AD that successfully authenticates the user. Also, Workload Optimization Manager does not support nested AD groups. AD logins must be for users in an upper-level group.

Maintaining AD Groups

To edit the details of an existing AD group or to delete an AD group, you must know the UUID value for the group.

- To get a list of all the currently configured AD groups, make a GET request to the `users/ldap/groups` endpoint:

  ```
  GET https://10.10.10.10/api/v3/users/ldap/groups
  ```

  The API returns an array of ActiveDirectoryApiDTO *(on page 1502)* objects. Each `ActiveDirectoryGroupApiDTO` object contains the UUID, role, and other details about a group.

- To edit the details for a group, make a PUT request to the `users/ldap/groups` endpoint.

  ```
  PUT https://10.10.10.10/api/v3/users/ldap/groups
  ```

  The following parameters are required in the request body, regardless of whether you want to modify them:

  - `displayName`
  - `roles`
  - `type`

  Include the required parameters in the request body and specify the values that you want to modify. In the following example, the role for the `myAdGroup` group is changed from DEPLOYER to OBSERVER. The other parameters remain unchanged:

  Example input:

  ```
  {
    "displayName": "myAdGroupr",
    "rolename": "OBSERVER",
    "type": "DedicatedCustomer"
  }
  ```

- **Response:**

  ```
  {
  "uuid": "mySsoGroup",
  "displayName": "mySsoGroup",
  "type": "DedicatedCustomer",
  "roleName": "DEPLOYER" }
  ```

- To delete a group, make a DELETE request to the `users/ldap/groups` endpoint and include the group's UUID:

  ```
  DELETE https://10.10.10.10/api/v3/users/ldap/groups/myADgroup
  ```

Workflows endpoint

When you add orchestrator targets to your Workload Optimization Manager installation, they assign workflows that run multiple actions to modify your environment. Workload Optimization Manager discovers workflows that you define on the orchestrator. You can then set up an automation policy that maps workflows to actions. If the action mode is Manual or Automatic, when Workload Optimization Manager recommends the action, it directs the orchestrator to use the mapped workflow to run it.

For more information about specific orchestrator targets, see Orchestrator Targets *(on page 360)*. For more information about execution of workflows, see Automation Workflow *(on page 1063)*.

You can also use the workflows endpoint to create and manage webhooks.

A webhook is an HTTP custom callback that alters the behavior of a web page or web application. You can use webhooks to integrate Workload Optimization Manager with other applications, such as Slack.

You can complete the following tasks by using the workflows endpoint:

- Get a list of all workflows.
- Get detailed information about a specified workflow.
- Get a list of policies that are associated with a specified workflow.
- Create, test, and modify a webhook.

Workflows requests

Getting workflows

For a list of workflows (such as webhooks) that are configured for your Workload Optimization Manager instance, the API returns an array of `WorkflowApiDTO` objects. Each `WorkflowApiDTO` object contains detailed information about that workflow, including description and parameters. To get a single workflow, include the UUID of the workflow in the request.

This call takes the following parameter:

Examples:

- All workflows: `GET https://10.10.10.10/api/v3/workflows`
- Single workflow: `GET https://10.10.10.10/api/v3/workflows/637998092280496`

Response:


```
[
  {
    "links":[
      {
        "rel":"self",
        "href":"https://10.10.10.10/api/v3/workflows/637998092280496"
      }
    ],
    "uuid":"637998092280496",
    "displayName":"APIC Container Delete Firewall Rule Action",
    "className":"Workflow",
    "actionType":"NONE",
    "parameters":[
      {
        "name":"APIC_TENANT_L4L7_EPG_ACE_IDENTITY",
        "isMandatory":true,
        "isSecret":false,
        "specificValueType":"ApicTenantL4L7EPGAceIdentity"
      },
      {
        "name":"L4L7Identity",
        "isMandatory":true,
        "isSecret":false,
        "specificValueType":"gen_text_input"
      },
      {
        "name":"ACLName",
        "isMandatory":true,
        "isSecret":false,
        "specificValueType":"gen_text_input"
      },
      {
        "name":"ACEName",
        "isMandatory":true,
        "isSecret":false,
        "specificValueType":"gen_text_input"
      }
    ],
```

```
    "discoveredBy":{
      "uuid":"__U1MIJv-Eei_xJ5Ai6-rnQ",
      "displayName":"ucsd.eng.mycorp.com",
      "type":"Cisco UCS Director"
    }
  ],
  ...
 }
]
```

Getting workflows of a specific type

By adding the WorkflowApiDTO *(on page 1723)* `type` parameter to the request, you can filter workflows from a specific source.

For example, the response to the following request is an array of `WorkflowApiDTO` objects that are configured as webhooks:

Example: `GET https://10.10.10.10/api/v3/workflows?type=WEBHOOK`

Getting a list of policies that are associated with a workflow

You can use the `/api/v3/workflows/{workflow_Uuid}/settingspolicies` endpoint to get a list of all the policies that are associated with a workflow that you specify by UUID. The API returns the UUID and display name for any policies that are associated with the specified workflow. For more information about policies, see Working with policies *(on page 1042)*.

Example: `GET https://10.10.10.10/api/v3/workflows/637998092280496/settingspolicies`

Response:

```
[
  {
    "uuid": "286242929267568",
    "displayName": "Send message to slack on action generation policy",
    "className": "BaseApiDTO"
  },
  {
    "uuid": "286242929902288",
    "displayName": "Send message to slack before execution policy",
    "className": "BaseApiDTO"
  },
  {
    "uuid": "286242929988464",
    "displayName": "Shutdown VM webhook policy",
    "className": "BaseApiDTO"
  },
  {
    "uuid": "286242930068560",
    "displayName": "Send message to slack after execution policy",
    "className": "BaseApiDTO"
  }
]
```

Creating and testing webhooks

A webhook is an HTTP custom callback that alters the behavior of a web page or web application. You can use webhooks to integrate Workload Optimization Manager with other applications, such as Slack.

To create a webhook, make a POST request to the `/workflows` endpoint:

Example: `POST https://10.10.10.10/api/v3/workflows`

The body of the request must include a WorkflowApiDTO *(on page 1723)* that provides details about your webhook.

```
{
    "displayName": "Example webhook workflow DTO",
    "className": "Workflow",
    "description": "Example of workflow request payload",
    "discoveredBy": {
        "readonly": false
    },
    "type": "WEBHOOK",
    "typeSpecificDetails": {
        "url": "<YOUR_WEBHOOK_URL>",
        "method": "POST",
        "template": "$converter.toJson($action)",
        "authenticationMethod": "NONE",
        "trustSelfSignedCertificates": false,
        "type": "WebhookApiDTO"
    }
}
```

For pre-defined request payload examples, see the Workload Optimization Manager Orchestration GitHub repository.

In the response body, make note of the UUID value, as you need it to test or modify your webhook.

Response:

```
{
  "uuid": "638205616014777",
  "displayName": "Example webhook workflow DTO",
  "className": "Workflow",
  "description": "Example of workflow request payload",
  "discoveredBy": {
    "readonly": false
  },
  "type": "WEBHOOK",
  "typeSpecificDetails": {
    "url": "https://webhook.site/182d298f-6089-4b71-a37d-cc756e8ba204",
    "method": "POST",
    "template": "$converter.toJson($action)",
    "authenticationMethod": "NONE",
    "trustSelfSignedCertificates": false,
    "templateValidationStatus": {
      "validationStatus": "VALID",
      "validationErrors": [],
      "lastTemplateValidationTime": "2023-12-01T17:49:46Z"
    },
    "mediaType": "TEXT",
    "type": "WebhookApiDTO"
  },
  "creationTime": "2023-12-01T17:49:46Z",
  "lastModifiedTime": "2023-12-01T17:49:46Z"
}
```

For more information, see Creating webhook workflows with the API *(on page 1076)*.

To test your webhook, send a POST request to the `/workflows` endpoint and append the UUID value for your webhook:

Example: `POST https://10.10.10.10/api/v3/workflows/{workflow_Uuid}`

In the request body, include the UUID value of a Workload Optimization Manager action.

```
{
  "operation": "TEST",
  "actionId": <YOUR_ACTION_ID>
}
```

If the call is successful, you get the following response:

Response:

```
{
    "succeeded": true,
    "details": "The call to the webhook endpoint was successful."
}
```

A successful POST request for a TEST operation also executes the webhook against the configured external endpoint. For example, if the webhook is integrated with Slack, a successful POST request for a TEST operation triggers output in the integrated Slack channel.

Modifying a webhook

To modify an existing webhook, make a PUT request to the `/workflows` endpoint and append the UUID value for your webhook:

Example: `PUT https://10.10.10.10/api/v3/workflows/{workflow_Uuid}`

In the request body, include the full `WorkflowApiDTO` with any changed values:

```
{
    "displayName": "Example webhook workflow DTO",
    "className": "Workflow",
    "description": "updated description of workflow request payload",
    "discoveredBy": {
        "readonly": false
    },
    "type": "WEBHOOK",
    "typeSpecificDetails": {
        "url": "<NEW_WEBHOOK_URL>",
        "method": "POST",
        "template": "$converter.toJson($action)",
        "authenticationMethod": "NONE",
        "trustSelfSignedCertificates": false,
        "type": "WebhookApiDTO"
    }
}
```

If the call is successful, you get the following response:

Response:

```
{
  "succeeded": true,
  "details": "The call to webhook endpoint was successful."
}
```

Workload Optimization Manager cookbook

Purpose

The Workload Optimization Manager cookbook is intended to assist in performing specific tasks using the API, and answering commonly-asked questions while presenting Workload Optimization Manager API best practices.

Cookbooks

Each cookbook contains one or more recipes. Each recipe will contain all the necessary information to complete the specified task, along with a basic script showing the functionality of the recipe.

Authenticating with the API

Objective

To obtain an authentication token after logging in with the API, and use that token on subsequent API calls.

Authentication in the API

To use the API, you must have a valid user account on the Workload Optimization Manager instance. Also note that accounts can have different roles. The API will only execute commands that are valid for your user role. For example, to execute Workload Optimization Manager recommended actions, your account must have a role of either `administrator`, `deployer`, or `automator`.

To make API calls, you request an authentication token and pass it with each call to the Workload Optimization Manager API. The token request returns a cookie for your authentication. A common way to use this token is to store the cookie locally, and pass it with your API calls.

Example: `curl -s -k -c /tmp/cookies -H 'accept: application/json' 'https://localhost/api/v3/login?hateoas=true' -d 'username=administrator&password=password'`

Then, each request must use the `-b cookie-filename` parameter to use the session cookie delivered by the login request.

Another approach is to get the authentication header and parse out the authentication cookie. Then you can create a header for each API request that includes the cookie. For example, assume you store the value in a variable named `token`. You could use it like this:

```
headers = {'cookie': token}
r = requests.get('https://10.10.123.456/api/v3/targets/specs', headers=headers, verify=False, stream=True)
```

Prerequisites

In order to obtain an authentication token, you must have the following information:

- The IP address of the Workload Optimization Manager instance.
- The username and password of the user who will be logged in.

Procedure

To obtain an authentication token:

1. Construct the API payload.

   The input body must be URL-encoded (`x-www-form-urlencoded`). Use the following format with your credentials in place of `myUsername` and `myPassword`:

   ```
   username=myUsername&password=myPassword
   ```

2. Make the API request.

   Use the `POST https://[INSTANCE_URL]/api/v3/login` request, with the input body constructed in the previous step.

3. Retrieve the authentication token from the response headers.

   Example headers:

```
{
  'Server':'nginx',
  'Date':'Mon, 01 Feb 2021 20:17:02 GMT',
  'Content-Type':'application/json',
  'Transfer-Encoding':'chunked',
  'Connection':'keep-alive',
  'Vary':'Accept-Encoding',
  'Set-Cookie':'JSESSIONID=node0jzerbqte09pe1j29x6ypo92fh17.node0; Path=/; HTTPOnly; Secure',
  'Expires':'Thu, 01 Jan 1970 00:00:00 GMT',
  'X-Content-Type-Options':'nosniff',
  'X-XSS-Protection':'1; mode=block',
  'Strict-Transport-Security':'max-age=63072000; includeSubDomains',
  'X-Frame-Options':'SAMEORIGIN',
  'X-Turbo-Upstream':'API',
  'Content-Encoding':'gzip'
}
```

The authentication token is the first `;`-separated tocen in the value of the `Set-Cookie` header. In this example, it is `JSESSIONID=node0jzerbqte09pe1j29x6ypo92fh17.node0`.

Using the authentication token in subsequent API requests

Once you have obtained an authentication token, you can use that token to make further API requests. To do so, include a header in your request with the name of `cookie` and a value of the token. For example:

```
{'cookie': 'JSESSIONID=node0jzerbqte09pe1j29x6ypo92fh17.node0' }
```

Script example

These listings show a function that gets the authentication token from the cookie, and a script that invokes the function and prints out the token value.

Function to get the token:

To invoke this function, pass the IP address, the username, and password as arguments. The function builds the authentication payload and requests the token from the given Workload Optimization Manager instance. It then parses out the first token in the `Set-Cookie` header and returns that value.

```
import json
import requests
import urllib3
urllib3.disable_warnings()
```

```
def get_cookie(ip, username, password):
 payload = {'username': username, 'password': password}
 r = requests.post(f'https://{ip}/api/v3/login', data=payload, verify=False)
 r.encoding = 'JSON'
 rh = r.headers
 token = rh['Set-Cookie'].split(';')[0]
```

```
 return token
```

Script that calls the function:

Assume the script is saved as `use_login.py`. Also assume the function is in the file `./login.py`. To invoke this script, open a shell at the script file location and enter:

`./use-login.py <My_IP_Address> <My_Username> <My_Pwd>`

where the three arguments are the IP address of your Workload Optimization Manager, your username, and your password, respectively. The script imports the `./login.py` file and checks for four arguments (the script name and the three arguments you passed). It then executes the `login.get_cookie()` function. The function returns a token, which the script then prints out.

```
#!/usr/bin/env python3
```

```
import sys
import login
```

```
if len(sys.argv) != 4:
 print("Incorrect arguments list.")
 sys.exit()
```

```
token = login.get_cookie(sys.argv[1], sys.argv[2], sys.argv[3])
print(f'TOKEN IS: {token}')
```

Script result

The result of this script displays the authentication token:

```
TOKEN IS: JSESSIONID=node0jzerbqte09pe1j29x6ypo92fh17.node0
```

Authenticating OAuth 2.0 clients with the API

Objective

Create, manage, and authenticate with OAuth 2.0 clients by using the Workload Optimization Manager API.

Prerequisites

To use the API to create and manage OAuth 2.0 clients and tokens, you must have the following prerequisites:

- Workload Optimization Manager version 8.12.4 or later
- `t8c-operator` version `42.57` or later
- `ADMINISTRATOR` or `SITE_ADMIN` user role privileges. Only users in the `ADMINISTRATOR` role can create OAuth 2.0 clients that have the `ADMINISTRATOR` role.

To interact with OAuth 2.0 APIs, you must be authenticated with the Workload Optimization Manager instance. For more information, see Authenticating with the API *(on page 1439)*.

Creating and authenticating an OAuth 2.0 client

The following steps describe the process to create an OAuth 2.0 client and exchange the client credentials for an access token that you use to authenticate calls to Workload Optimization Manager APIs. You can also easily try these calls by using the Workload Optimization Manager swagger UI. For more information, see Workload Optimization Manager REST API Swagger Documentation *(on page 1142)*.

Currently, Workload Optimization Manager supports only the OAuth 2.0 client credentials flow.

1. Create an OAuth 2.0 client.

   Construct a `POST https://[INSTANCE_URL]/api/v3/authorization/oauth2/clients` request with the following payload, where `[INSTANCE_URL]` is the URL for your Workload Optimization Manager instance.

```
{
  "clientName": "my-client",
  "grantTypes": [
    "client_credentials"
  ],
  "clientAuthenticationMethods": [
    "client_secret_post"
  ],
  "scopes": [
    "role:SITE_ADMIN"
  ],
  "tokenSettings": {
    "accessToken": {
      "ttlSeconds": 700
    }
  }
}
```

Specify the following values:

- `clientName`: Choose a name for your client.
- `scopes`: Specify exactly one value in the format `role:`<ROLE>, where <ROLE> is a Workload Optimization Manager user role, for example, `role:SITE_ADMIN`. Any access tokens that you generate for this client must have the same role. The following roles are available:
  - `role:ADMINISTRATOR` (only if the logged-in user is an ADMINISTRATOR)
  - `role:SITE_ADMIN`
  - `role:AUTOMATOR`
  - `role:DEPLOYER`
  - `role:ADVISOR`
  - `role:OBSERVER`
  - `role:OPERATIONAL_OBSERVER`
  - `role:SHARED_ADVISOR`
  - `role:SHARED_OBSERVER`
  - `role:REPORT_EDITOR`

  For more information about the permissions that these roles grant, see step 5 of Managing user accounts *(on page 1116)*.

The following parameters are optional:

- `clientAuthenticationMethods`: This value determines whether the client credentials must be supplied in a basic auth header or in the request body when you request an access token for the client. Specify one or both of the following values:
  - `client_secret_basic`: the default, which requires the credentials to be supplied as part of a basic auth header
  - `client_secret_post`: the credentials must be supplied in the request body
- `tokenSettings.accessToken.ttlSeconds`: This value determines the lifespan of access tokens that the OAuth 2.0 authorization server issues. The default value is `600` seconds.

The response is similar to the following example.

```
{
  "clientName": "my-client",
  "clientId": "7613c2ca-cee1-4445-aa99-55c9312852c8",
  "grantTypes": [
    "client_secret"
  ],
  "clientAuthenticationMethods": [
    "client_secret_post"
```

```
  ],
  "scopes": [
    "role:SITE_ADMIN"
  ],
  "audience": [
    "turbonomic"
  ],
  "createdAt": 1702042743,
  "tokenSettings": {
    "accessToken": {
      "ttlseconds": 700
    }
  },
  "clientSecret": "b6YzO0zaRU6EPYNMZPBIbYhmWL1Jb6PTAKyBUTw7VIG8FQDh2ryCaMspoyh2cKzM"
}
```

Copy the values for `clientId` and `clientSecret`. The `clientSecret` cannot be recovered if lost.

2. Exchange the client credentials for an access token.

   To exchange your credentials for an access token, you must call the `https://[INSTANCE_URL]:443/oauth2/token` endpoint. This endpoint is a standard 0Auth 2.0 API and you can use any library that supports OAuth 2.0 to facilitate this transaction programmatically. For more information, see the documentation for your chosen library. The following substeps describe how to compose the call manually.

   - If your client uses `client_secret_basic` as the authentication method, you must include the client credentials in a basic auth header, not in the request body. Construct a `POST https://[INSTANCE_URL]:443/oauth2/token` request with the following payload, where `[INSTANCE_URL]` is the URL for your Workload Optimization Manager instance.

     ```
     grant_type=client_credentials&scope=role:SITE_ADMIN
     ```

     Specify the following values:
     - `grant_type`: Specify `client_credentials`.
     - `scope`: This value must match the configured scope for your client, for example, `role:SITE_ADMIN`.
   - If your client uses `client_secret_post` as the authentication method, you must include the client credentials in the request body. Construct a `POST https://[INSTANCE_URL]:443/oauth2/token` request with the following payload, where `[INSTANCE_URL]` is the URL for your Workload Optimization Manager instance.

     ```
     grant_type=client_credentials&scope=role:SITE_ADMIN&client_id=7613c2ca-cee1-4445-aa99-55c9312
     852c8&client_secret=b6YzO0zaRU6EPYNMZPBIbYhmWL1Jb6PTAKyBUTw7VIG8FQDh2ryCaMspoyh2cKzM
     ```

     Specify the following values:
     - `client_id`: Specify the client ID for your OAuth 2.0 client that you copied from the API response when you created the client.
     - `client_secret`: Specify the client secret for your OAuth 2.0 client that you copied from the API response when you created the client.
     - `grant_type`: Specify `client_credentials`.
     - `scope`: This value must match the configured scope for your client, for example, `role:SITE_ADMIN`.

   The response is similar to the following example.

```
{
  "access_token":
 "GiLUVmth6LH6nSocLYtd4DvUxIy4L9zntTbZX7NQVOlq31-kle_ZEDkcJiiZkx0-KtKjJAR78sQ1qR6LJCIqQ_HSO_i0KBfHz
zZgc-0x8gBy8FpDnWLdUENuRnlBt2_t",
  "scope": "role:SITE_ADMIN",
  "token_type": "Bearer",
  "expires_in": 599
}
```

Copy the value of the `access_token` to authenticate calls to Workload Optimization Manager APIs.

3. Use the access token to access Workload Optimization Manager APIs.

   Include the authorization token in a bearer authorization header of a request to a Workload Optimization Manager API that is available to your role, such as the `GET licenses` API. For example, you can set the token value as a variable in your terminal:

   ```
   token=GiLUVmth6LH6nSocLYtd4DvUxIy4L9zntTbZX7NQVOlq31-kle_ZEDkcJiiZkx0-KtKjJAR78sQ1qR6LJCIqQ_HSO_i0K
   BfHzzZgc-0x8gBy8FpDnWLdUENuRnlBt2_t
   ```

   When you make a request to a Workload Optimization Manager API, include the variable in a bearer authorization header, as shown in the following curl example:

   ```
   curl -H "Authorization: Bearer ${token}" https://$host:443/vmturbo/rest/licenses
   ```

Listing or deleting OAuth 2.0 clients

After you create OAuth 2.0 clients for your Workload Optimization Manager instance, you can list or delete the clients by making API calls.

- List the OAuth 2.0 clients in your Workload Optimization Manager instance.

  Make a `GET https://[INSTANCE_URL]/api/v3/authorization/oauth2/clients` request, where `[INSTANCE_URL]` is the URL for your Workload Optimization Manager instance. The response is similar to the following example.

  ```
  {
    "clientName": "my-client",
    "clientId": "7613c2ca-cee1-4445-aa99-55c9312852c8",
    "grantTypes": [
      "client_secret"
    ],
    "clientAuthenticationMethods": [
      "client_secret_post"
    ],
    "scopes": [
      "role:SITE_ADMIN"
    ],
    "audience": [
      "turbonomic"
    ],
    "createdAt": 1702042743,
    "tokenSettings": {
      "accessToken": {
        "ttlSeconds": 700
      }
  }
  ```

  The `clientSecret` values are not returned.

- Delete an OAuth2 client from your Workload Optimization Manager instance.

  Make a `DELETE https://[INSTANCE_URL]/api/v3/authorization/oauth2/clients/[clientId]` request, where `[clientId]` is the `clientId` value for the client that you want to delete and `[INSTANCE_URL]` is the URL for your Workload Optimization Manager instance.

  A successful DELETE request does not return a response body. You can confirm that the request succeeded by running the same request again. The response to that request is `404 NOT_FOUND` because the target client was already deleted.

  Any tokens that were issued for the deleted client are immediately invalidated and can no longer be used to access Workload Optimization Manager APIs.

Troubleshooting OAuth 2.0 clients

If you cannot access the `authorization/oauth2/clients/` API, make sure that you satisfy all the prerequisites *(on page 1441)* and that your session is not expired.

If the token you get from the OAuth 2.0 API is not working, check the following values:

- Make sure that the token is not expired. The default lifespan of a token is 10 minutes. If the token expires, you must retrieve a new one.
- Ensure that you specified a scope in your request to the OAuth2 API *(on page 1443)*. If you don't include a scope in your request, the token is not valid.

Plan results cookbook

This cookbook explores plan markets in Workload Optimization Manager. Interpreting and parsing this plan market will enable you to better understand how and why Workload Optimization Manager has performed a certain action on a particular entity, how the resources- both used and available- have changed, and more as recipes are added to the cookbook.

> **NOTE:** The Workload Optimization Manager UI has certain plan configuration options, available to present a base set of configuration changes to accomplish a particular goal. For example, the Add Workload plan is preconfigured to expect additional workloads to be provisioned. To see how Workload Optimization Manager creates these configurations, view the DTOs for the UI plan configurations at Scenario Examples *(on page 1314)*.

Related endpoints

When interpreting plan results, the following endpoints are commonly used to retrieve and clarify data:

- Actions *(on page 1161)*: Retrieving the actions for the plan market will enable you to see what entities were directly affected during the run of the plan.
- Entities *(on page 1195)*: Retrieving the entities for the plan market will enable you to see a full listing of the entities involved in the plan.
- Groups *(on page 1222)*: Retrieving the groups for the plan market will enable you to more easily obtain UUID information for the entities in the specified group.
- Statistics *(on page 1372)*: Retrieving the market statistics for the plan market as a whole, a group, or an entity will enable you to easily see a broad overview of the state of the entity utilization after the plan has finished.

Use case index

The Plan Results Cookbook has the following recipes:

- Seeing Changes to a Subset of Entities *(on page 1445)*

Seeing changes to a subset of entities

Objective

After running a plan, you may want to see the effects of your plan on a particular entity or group of entities. This can help you better understand the effect your configuration changes would have on a subset of plan entities, or to explore why Workload Optimization Manager made specific action recommendations.

For example, assume that there is a virtual machine that shares dependencies with a number of other entities. Apart from the overal plan results, you might want to take a closer look at how the plan impacts this VM in particular.

Prerequisites

In order to investigate the changes to the specified entity, you need to know:

- The UUID of a successfully completed plan market

  To obtain this UUID, either get it from the API response sent when you POST a scenario to the market, or execute the `GET /markets` request, and find the market whose name corresponds with the name of the scenario. Alternately, you can find the scenario's UUID in the `scenario` parameter of the plan market. Verify that the `state` parameter has a value of SUCCEEDED. For example, the UUID of the scenario that created the following plan market is `214410933912976`, and the UUID of the plan market itself is `214410933923968`:

  ```
  {
      "uuid":"214410933923968",
      "displayName":"Optimize Cloud 2",
      "className":"Market",
      "state":"SUCCEEDED",
      "stateProgress":100,
      "scenario":{
          "uuid":"214410933912976",
          "displayName":"Optimize Cloud 2",
          ...
      }
  ...
  }
  ```

  For cases where the same scenario has been applied to multiple plans, be sure to check the plan's `runDate` parameter corresponds to the time you expect the plan to have run.

  For more information about creating and running scenarios, see Scenario DTO Breakdown *(on page 1303)*.

- The UUID(s) of the entities you want to inspect

  Entity UUIDs in a plan market are identical to their real-time market counterparts. This means you can get the UUIDs either from searching the plan market for a related field (for example, the `displayName` of the entity) or searching in the Workload Optimization Manager user interface to find the entity UUID.

  You can also use the API's Search endpoint to search for entities by different criteria.

Related endpoints

When interpreting plan results, the following endpoints are commonly used to retrieve and clarify data:

- Actions *(on page 1161)*: Retrieving the actions for the plan market will enable you to see what entities were directly affected during the run of the plan.
- Entities *(on page 1195)*: Retrieving the entities for the plan market will enable you to see a full listing of the entities involved in the plan.
- Groups *(on page 1222)*: Retrieving the groups for the plan market will enable you to more easily obtain UUID information for the entities in the specified group.
- Statistics *(on page 1372)*: Retrieving the market statistics for the plan market as a whole, a group, or an entity will enable you to easily see a broad overview of the state of the entity utilization after the plan has finished.

Procedure

To investigate changes to a particular subset of plan entities:

1. Get the list of UUIDs you want investigate.

   To get all the entities in a market:

   ```
   markets/{market_uuid}/entities
   ```

If the scope of the plan is small, you can use this to get all the entities in the plan, and then iterate through the results for other attributes such as entity type or display name.

```
markets/{market_uuid}/entities
```

For a large scope, you can use the Search endpoint to find VMs by specific criteria. This works because the plan market and the real-time market use the same UUID values for the same entities. For example, the following call gets all VMs that have the string FOO in their names:

```
https://10.10.123.456/api/v3/search?q=FOO&types=VirtualMachine&query_type=CONTAINS
```

2. Get the list of plan actions.

   Run `GET /markets/{market_Uuid}/actions` against the plan market. You can also use `POST /markets/{market_uuid}/actions` to post a filter for advanced action queries, such as only returning actions of a particular type, or particular risk severity.

   Another way to get plan actions is to use the Search endpoint. You can POST search parameters that limit the returned actions by text in the description, cost type, or other data. For example, the following search returns only actions on VMs:

```
searchSpec = {
 "actionRelationTypeFilter": [
  "NONE"
 ],
 "costType": "SAVING",
 "descriptionQuery": {
  "query": "Virtual Machine",
  "type": "CONTAINS",
  "caseSensitive": "true"
 },
 "detailLevel": "STANDARD"
}
```

   With the previous payload, the following POST would return VM actions from the plan:

```
requests.post(f'https://{ip}/api/v3/markets/{marketUuid}/actions',
  headers=headers,
  json = searchSpec,
  verify=False,
  stream=True)
```

3. Search the API response and find the actions relevant to your entities.

   Each action will have a `target` parameter, which will in turn have a `uuid`. This UUID identifies the entity the action executed on. This is the best way to locate your entity in a list of actions.

   If no actions have the specified `target > uuid` value, then no actions were performed on the entity in question.

   > **NOTE:** You can use `GET /markets/{market_Uuid}/entities` to verify that the UUID you are searching for exists in the plan market. If it does not, it means that the entity was not included in the scope of the plan, or the entity is considered an idle workload. An idle workload shows no utilization, so the plan cannot determine optimal placement or what percentage of allocated resources that workload will require when it restarts.

4. Analyze the results.

   Each action contains the information detailed in the ActionApiDTO data, including the details of the action, old/new values for any changed resources, a record of deletion for DELETE actions, etc.

   Search through this DTO and make note of any parameters important to you. (For a DTO reference, see ActionApiDTO *(on page 1485)*.)

Script example

The following script searches through the list of actions for a virtual machine, and returns the requested fields for quick comparison. In this case, the UUID, type, state, reason, risk category, and risk subcategory are returned for the action, and paired with the display name, template UUID, and location of the entity.

This script is easily extensible- for example, if `currentLocation > displayName` is equal to `newLocation > displayName`, you may choose not to show either of those fields, as no change involving the VM location was made. You might also use these results to make further calls- for example, using the template UUID to make a `GET /templates/{template_Uuid}` request to return more information about the template being used for a resize action.

Note that the script uses dot-notation to name the fields you want to inspect. It includes the `getFieldFromDict` function to convert those fields to nested Python dictionary values. Because the API returns JSON data, a function like this could be a generally useful utility.

```
#!/usr/bin/env python3
```

```
#
# Commandline args: platformIP, username, pwd, marketID, VmIDs as comma-sep string (no spaces)
# EXAMPLE: ./entity_changes.py 10.10.123.456 MyUser MyPwd 215426452602368
 74635124762083,74635124762088,74635124762090
#
import json
import requests
import urllib3
import re
```

```
import sys
sys.path.append('../login') #This makes login visible...
import login
```

```
#
# This function takes json-like dot-notation fields and returns python dictionary values.
# That makes it easier to specify the DTO fields you want to view.
#
def getFieldFromDict(field, dict):
 fToks = field.split('.')
 l = len(fToks)
 workingDict = dict
 for item in fToks:
  if item in workingDict:
   ret = workingDict[item]
  else:
   return "ITEM NOT IN OBJECT"
  l -= 1
  if l == 0:
   return ret
  else:
   workingDict = ret
 return ret
```

```
#
# Ths list of fields you want to query, in dot-notation
#
RetrieveFields = [
"uuid",
"actionType",
"actionState",
"details",
"currentEntity.displayName",
"newEntity.displayName",
"template.uuid",
```

```
"risk.subCategory",
"risk.description",
"currentLocation.displayName",
"newLocation.displayName"
]
```

```
#
# A search spec to get actions that are on VMs.
#
searchSpec = {
  "actionRelationTypeFilter": [
    "NONE"
  ],
  "costType": "SAVING",
  "descriptionQuery": {
    "query": "Virtual Machine",
    "type": "CONTAINS",
    "caseSensitive": "true"
  },
  "detailLevel": "STANDARD"
}
```

```
#
# Make sure you have the correct count of arguments.
#
if len(sys.argv) != 6:
 print("Incorrect arguments list.")
 sys.exit()
```

```
#
# Get the args into well named vars, and also authenticate.
#
ip = sys.argv[1]
marketUuid = sys.argv[4]
ElemsToInspect = sys.argv[5].split(',')
```

```
token = login.get_cookie(ip, sys.argv[2], sys.argv[3])
headers = {'cookie': token}
```

```
#
# Search plan for actions on VMs
#
actions_r = requests.post(f'https://{ip}/api/v3/markets/{marketUuid}/actions', headers=headers, json =
 searchSpec, verify=False, stream=True)
```

```
if actions_r.status_code != 200:
    print(f'Market Request Status is {actions_r.status_code}.')
    sys.exit()
```

```
#
# For each action that targets one of the entities you pass in the command line,
# parse the object and display the field values.
#
print("Plan Changes for Entity(ies):")
print(sys.argv[5])
print("-----------------------")
r_json = actions_r.json()
for item in r_json:
 target = item['target']
 if 0 < ElemsToInspect.count(target["uuid"]):
  print(f'VM ACTION: {target["displayName"]}')
  for field in RetrieveFields:
   val = getFieldFromDict(field, item)
   print(f'    {field}: {val}')
```

Script result

The result of this script displays the requested information for each plan entity:

```
Plan Changes for Entity(ies):
74635124762083,74635124762088
----------------------
VM ACTION: test-rosa-28tfg-worker-us-west-2a-5hr8j
    uuid: 145057709553360
    actionType: SCALE
    actionState: SUCCEEDED
    details: Scale Virtual Machine test-rosa-28tfg-worker-us-west-2a-5hr8j from m5.xlarge to r5a.large in
 Advanced Engineering
    currentEntity.displayName: m5.xlarge
    newEntity.displayName: r5a.large
    template.uuid: 74635137532084
    risk.subCategory: Efficiency Improvement
    risk.description: Underutilized VCPU, Net Throughput
    currentLocation.displayName: aws-US West (Oregon)
    newLocation.displayName: aws-US West (Oregon)
VM ACTION: test-rosa-28tfg-infra-us-west-2a-mw867
    uuid: 145057709553376
    actionType: SCALE
    actionState: SUCCEEDED
    details: Scale Virtual Machine test-rosa-28tfg-infra-us-west-2a-mw867 from r5.xlarge to r5a.xlarge in
 Advanced Engineering
    currentEntity.displayName: r5.xlarge
    newEntity.displayName: r5a.xlarge
    template.uuid: 74635137531741
    risk.subCategory: Efficiency Improvement
    risk.description: Increase RI Utilization
    currentLocation.displayName: aws-US West (Oregon)
    newLocation.displayName: aws-US West (Oregon)
```

Seeing the changes for a group of entities

You can easily modify this script to see the changes for a particular group of entities without needing to know each UUID.

To perform this modification, consume the result of a `GET /groups/{group_uuid}/entities` request. Example result:

```
[
  {
    "uuid":"73422419372894",
    "displayName":"PT_RHEL",
    "className":"VirtualMachine",
    "priceIndex":4.33,
    "state":"ACTIVE",
    "severity":"Normal",
    "costPrice":0.24104111,
    "discoveredBy":{
      "uuid":"73422416392368",
      "displayName":"producttrust",
      "category":"Public Cloud",
      "type":"AWS",
      "readonly":false
```

```
    },
    ...
  },
  {
    "uuid":"73422419372988",
    "displayName":"PT_Resize_Up_Target",
    "className":"VirtualMachine",
    "priceIndex":10074.822,
    "state":"ACTIVE",
    "severity":"Normal",
    "costPrice":0.078387,
    "discoveredBy":{
      "uuid":"73422416392368",
      "displayName":"producttrust",
      "category":"Public Cloud",
      "type":"AWS",
      "readonly":false
    },
    ...
  },
  {
    "uuid":"73422419372921",
    "displayName":"PT_Consistent_Scaling_VM_1",
    "className":"VirtualMachine",
    "priceIndex":100.357,
    "state":"ACTIVE",
    "severity":"Normal",
    "costPrice":0.0159137,
    "discoveredBy":{
      "uuid":"73422416392368",
      "displayName":"producttrust",
      "category":"Public Cloud",
      "type":"AWS",
      "readonly":false
    },
    ...
  }
]
```

Using this result, you can either use each UUID found, or filter based on entities. For example, to iterate through a group and get the UUID of all VM entities:

```
foreach (var entity in groupApiResponse)
{
  if (entity["className"] == VirtualMachine)
  {
    Console.Writeline(entity["uuid"])
  }
}
```

Creating a static group

Objective

To create a static group of entities.

A static group contains a fixed set of entities. This example creates a group of VMs, but you can create a group of any entity type.

The script will get the unique IDs of member entities by their display names, and add them to a list that it then uses to display the group members.

The script also sets up the group type and entity type when it creates the payload data. It then uses the API to create the group, and verify that the group contains the entities you specified.

Prerequisites

In order to create a static group, you must have the following information:

- The display names of the members in the group
- The entity type you will use for this group
- The IP address and login credentials for a Workload Optimization Manager instance

Related endpoints

This recipe uses the following endpoints:

- Groups *(on page 1222)*

  It uses the `POST /groups` request to create the group, and the `GET /groups/{group_uuid}` request to verify the group membership.
- Search *(on page 1351)*

  It uses the `GET /search` request to obtain the UUIDs of the group members from the display names.

Script setup

To set up the script, provide values from your environment for the following:

- `ip`

  The IP address of your Workload Optimization Manager instance.
- `username`

  The username for the account Workload Optimization Manager will use to access the API.
- `password`

  The password for the account Workload Optimization Manager will use to access the API.
- `names_list`

  A list of display names for the entities you want to add to the group. Note that the entities must all be of the same type. If entities of more than one type are provided, group creation will fail.
- `group_name`

  The display name for the group. This name will appear in the user interface.
- `entity_type`

  The entity type for the group and its member entities. The script will use this in creating the Group input data, and also when executing a search for the entities' UUID.

> **NOTE:**
> Workload Optimization Manager does not support creation of groups with more than one entity type at this time.

Authentication

The first step of any script is to authenticate accessc to your Workload Optimization Manager instance, retrieve the authentication token and store it for use with subsequent calls. To authenticate, this script uses the values you set for `ip`, `username`, and `password`. For more detail, see the Authentication Recipe *(on page 1439)*.

API Request: `https://{ip}/vmturbo/rest/login`

UUID retrieval

After it logs in, the script iterates through your list of entity names, passing each in turn to the `GET /search` request. It uses the returned EntityApiDTO to get the UUIDs for each matching entity.

The search response is an array of entity objects, and each entity object includes the entity's UUID. The script loops through the display names you provide in `names_list`, and executes a separate search request for each name. Each response should be an array of one item. The script adds the `uuid` of that item to the `group_members_uuids list`.

> **NOTE:**
> Another method to obtain the UUIDs is to execute a filtered `POST /search` request. This script uses repeated `GET /search` requests for simplicity.

Sample Request: `GET https://10.10.10.10/api/v3/search?q=zrlinuxclone106&types=VirtualMachine&entityTypes=VirtualMachine`

Sample Return:


```
[
  {
    "uuid":"74011863734144",
    "displayName":"zrlinuxclone106",
    "className":"VirtualMachine",
    "state":"ACTIVE",
    "severity":"Critical",
    "discoveredBy":{
      "uuid":"73935378522272",
      "displayName":"vsphere-dc23.eng.vmturbo.com",
      "category":"Hypervisor",
      "type":"vCenter",
      "readonly":false
    },
    "severityBreakdown":{
      "CRITICAL":1
    },
    "providers":[
      {
        "uuid":"73935378875792",
        "displayName":"hp-esx108.eng.vmturbo.com",
        "className":"PhysicalMachine"
      },
      {
        "uuid":"73935378875736",
        "displayName":"QS2_GRID01:DC23DS1",
        "className":"Storage"
      },
      {
        "uuid":"73995333587472",
        "displayName":"Resources-hp-esx108.eng.vmturbo.com",
        "className":"VirtualDataCenter"
      }
    ],
    "environmentType":"ONPREM",
    "vendorIds":{
      "vsphere-dc23.eng.vmturbo.com":"vm-2844"
    }
  }
]
```

Constructing the input DTO and creating the group

Once the script has built the list of UUIDs for the group members, it can build a `GroupApiDTO` to describe the group you want. The DTO needs values for the following properties:

- `isStatic`: true to create a static group
- `displayName`: The group name that you set for the group_name variable
- `memberUuidList`: The array of UUID strings that the script has created
- `criteriaList`: For a static group, pass an empty array
- `groupType`: The entity type for this group -- the value you set for the entity_type variable

Using the obtained UUIDs and the previous information, the script will construct the input GroupApiDTO *(on page 1564)*. The script creates a Python dictionary, which it passes as the input body of the `POST /groups` request.

Example Input DTO:

```
{
  "isStatic":"true",
  "displayName":"TestGrp2",
  "memberUuidList":[
    "74011863734144",
    "73997881750336"
  ],
  "criteriaList":[

  ],
  "groupType":"VirtualMachine"
}
```

Sample Request: `POST https://{ip}/api/v3/groups/`

The response to this request will be the `GroupApiDTO` for the created group.

Sample Response:

```
{
  "uuid":"285123335341792",
  "displayName":"TestGrp2",
  "className":"Group",
  "membersCount":2,
  "entitiesCount":2,
  "groupType":"VirtualMachine",
  "severity":"CRITICAL",
  "environmentType":"ONPREM",
  "isStatic":true,
  "logicalOperator":"AND",
  "memberUuidList":[
    "74011863734144",
    "73997881750336"
  ],
  "temporary":false,
  "activeEntitiesCount":1,
  "cloudType":"UNKNOWN",
  "memberTypes":[
    "VirtualMachine"
  ],
  "entityTypes":[
    "VirtualMachine"
  ],
```

```
  "groupClassName":"GroupApiDTO"
}
```

Verification

To verify that the group was created successfully, the script uses the returned UUID to get the group data via the `/groups` endpoint. If the request gets the group, the returned DTO should be identical to the response from the request to create the group.

If the group does not exist for the given UUID, the API returns the request status `404`, with the message, `Group not found: <Passed_UUID>`. This indicates that the script failed to create the group.

Sample Response: `https://{ip}/api/v3/groups/{group_uuid}`

You can choose to only use a subset of the provided response. For example, this script prints the `displayName` and `UUID` for each group member.

Sample Script Response:

```
Group Members:
```

```
Display name: boao.min-spare
OID: 73997881750336
Display name: zrlinuxclone106
OID: 74011863734144
```

Script example

```
import json
import requests
import urllib3
import re
```

```
import sys
sys.path.append('../login') #This makes login visible...
import login
```

```
#
# Make sure you have the correct count of arguments.
#
if len(sys.argv) != 7:
 print("Incorrect arguments list.")
 sys.exit()
```

```
#
# Get the args into well named vars, and also authenticate.
#
ip = sys.argv[1]
groupName = sys.argv[4]
entityType = sys.argv[5]
entsListByName = sys.argv[6].split(',')
```

```
token = login.get_cookie(ip, sys.argv[2], sys.argv[3])
authHeader = {'cookie': token}
```

```
#
```

```
# Get UUIDs for entities in entsListByName
#
memberIds = []
for name in entsListByName:
 response = requests.get(
  f'https://{ip}/api/v3/search?q={name}&types={entityType}&query_type=EXACT',
  headers=authHeader, verify=False)
 if response.status_code == 200:
     response_json = json.loads(response.text)
     memberIds.append(response_json[0]['uuid'])
 if response.status_code != 200:
     print(f'Search Request Status for {name} is {response.status_code}.')
```

```
#
# Assemble the payload object to create the group.
#
group_input_dto = {}
group_input_dto['isStatic'] = 'true'
group_input_dto['displayName'] = groupName
group_input_dto['memberUuidList'] = memberIds
group_input_dto['criteriaList'] = []
group_input_dto['groupType'] = entityType
```

```
#
# Create the group
#
response = requests.post(f'https://{ip}/api/v3/groups', json=group_input_dto, headers=authHeader,
 verify=False, stream=True)
```

```
if response.status_code != 200:
 print(f'Create Group failed. StatusCode: {response.status_code}\n{response.text}')
 sys.exit()
```

```
groupResponse_json = json.loads(response.text)
groupId = groupResponse_json['uuid']
```

```
#
# Inspect the group
#
groupCheck_response = requests.get(f'https://{ip}/api/v3/groups/{groupId}', headers=authHeader,
 verify=False)
groupCheck_json = json.loads(groupCheck_response.text)
```

```
if groupCheck_json['uuid']:
 verifiedId = groupCheck_json['uuid']
 groupName = groupCheck_json['displayName']
 print(f'Group \"{groupName}\" successfully created with a UUID of: {verifiedId}')
else:
 print('Error in group creation, no UUID found.')
 sys.exit()
```

```
# Get a list of Group members
response = requests.get(
 f'https://{ip}/api/v3/groups/{groupId}/entities', headers=authHeader, verify=False)
if response.status_code != 200:
 print(f'Could not get group members. Status is {response.status_code}.')
 sys.exit()
```

```
groupMembers_json = json.loads(response.text)
```

```
# Print a list of group members
print('\nGroup Members:')
for item in groupMembers_json:
```

```
print('Display name: ' + item['displayName'])
print('OID: ' + item['uuid'])
```

Creating a dynamic group

Objective

To create a dynamic group of entities.

A dynamic group contains a set of entities discovered by filter criteria. In this case, the script will create a group of all virtual machines which reside on a specific host. As virtual machines are added and removed from the host, the membership of this group will reflect those changes.

The script will use the `/search` endpoint to get the unique ID of the host using its display name, and then retrieve a list of expected virtual machines to verify against. The script then creates the criteria list and inserts that value into the rest of the payload data before creating the group and verifying that the membership is as expected.

Prerequisites

In order to create a dynamic group, you must have the following information:

- The display name of the host

  The entity type you will use for this group

  The IP address and login credentials for a Workload Optimization Manager instance

Related endpoints

This recipe uses information from the following endpoints:

- Groups *(on page 1222)*

  NOTE:
  All possible criterion values can be found in the Groups Endpoint Tips *(on page 1250)*.

- Search *(on page 1351)*

Script setup

To set up the script, provide values from your environment for the following:

- `ip`

  The IP of your Workload Optimization Manager instance.
- `username`

  The username for the account Workload Optimization Manager will use to access the API.
- `password`

  The password for the account Workload Optimization Manager will use to access the API.
- `host_name`

  The display name of the host whose virtual machines will be added to the group.
- `group_name`

  The display name for the group. This name will appear in the user interface.

Authentication

The first step of any script is to authenticate to your Workload Optimization Manager instance, retrieve the authentication token and store it for use with subsequent calls. To authenticate, this script uses the values you set for `ip`, `username`, and `password`. For more detail, see the Authentication Recipe *(on page 1439)*.

API Request: `https://{ip}/vmturbo/rest/login`

Get expected initial membership

After it logs in, the displays a list of virtual machines that reside on the host. These will be the group members.

The script uses the `/search` endpoint and retrieves the EntityApiDTO for the host. Using the `consumers` parameter, you can see each entity that purchases resources from this host. The script lists the consumers with a `className` that matches the entity type.

Sample Request: `GET https://{ip}/api/v3/search?q={host_name}&types=PhysicalMachine'&detail_type=entity'`

Create the filter criterion

To create a dynamic group, you pass filter criteria to identify matching entities. You pass these criteria as an array. For this script, you pass a single criterion in the array.

Example `criteria_list`:

```
[
  {
    "expType":"RXEQ",
    "expVal":"host_name",
    "filterType":"vmsByPMName",
    "caseSensitive":"false"
  }
]
```

Construct the input DTO and create the group

The script will take the `criterion` parameter and use it to construct the GroupApiDTO *(on page 1564)*, with the following parameters:

- `isStatic`

  When true, the created group is static. The script sets this parameter to `false`.
- `displayName`

  The display name for the group. This name will appear in the user interface. The script sets this parameter
- `memberUuidList`

  For a dynamic group, pass an empty array.
- `criteriaList`

  The filtering criteria used to determine group membership. The script sets this parameter equal to the `criteria_list` variable.
- `groupType`

  The entity type of the group members. The script sets this parameter equal to the `group_entity_type` variable.

API Request: `POST https://{ip}/api/v3/groups/`

The response to this request will be the `GroupApiDTO` for the created group.

Sample Response:


```
{
  "uuid":"285126581114848",
  "displayName":"TestGrp2",
  "className":"Group",
  "membersCount":3,
  "entitiesCount":3,
  "groupType":"VirtualMachine",
  "severity":"CRITICAL",
  "environmentType":"ONPREM",
  "isStatic":false,
  "logicalOperator":"AND",
  "criteriaList":[
    {
      "expVal":"hp-esx119.eng.vmturbo.com",
      "expType":"RXEQ",
      "filterType":"vmsByPMName",
      "caseSensitive":false,
      "entityType":null,
      "singleLine":false
    }
  ],
  "memberUuidList":[
    "73943582577376",
    "73943582577568",
    "73943582577744"
  ],
  "temporary":false,
  "activeEntitiesCount":2,
  "cloudType":"UNKNOWN",
  "memberTypes":[
    "VirtualMachine"
  ],
  "entityTypes":[
    "VirtualMachine"
  ],
  "groupClassName":"GroupApiDTO"
}
```


Verify the group

To verify that the group was created successfully, the script uses the returned UUID to get the group data via the `/groups` endpoint. If the request gets the group, the membership of the group should match what was printed to the console in the *Get Expected Initial Membership* portion of the script, assuming that no virtual machines have been added or removed since that time.

If the group does not exist for the given UUID, the API returns the request status `404`, with the message, `Group not found: <Passed_UUID>`. This indicates that the script failed to create the group.

API Request: `https://{ip}/api/v3/groups/{group_uuid}`

You might choose to only use a subset of the provided response. For example, this script prints the `displayName` and `UUID` for each group member.

Sample Script Response:

```
Group Members:
```

```
Display name: DC26-P02-03
OID: 73943582577376
Display name: DC26-P02-02
OID: 73943582577568
Display name: DC26-P02-01
OID: 73943582577744
```

Script example

```
#
# Commandline args: platformIP, username, pwd, group_name, entity_type, host_name
# EXAMPLE: ./entity_changes.py 10.10.123.456 MyUser MyPwd MyTestGroup VirtualMachine MyHost
#
import json
import requests
import urllib3
import re
```

```
import sys
sys.path.append('../login') #This makes login visible...
import login
```

```
#
# Make sure you have the correct count of arguments.
#
if len(sys.argv) != 7:
 print("Incorrect arguments list.")
 sys.exit()
```

```
#
# Get the args into well named vars, and also authenticate.
#
ip = sys.argv[1]
group_name = sys.argv[4]
entity_type = sys.argv[5]
host_name = sys.argv[6]
```

```
token = login.get_cookie(ip, sys.argv[2], sys.argv[3])
authHeader = {'cookie': token}
```

```
#
# Get Expected Initial Membership
#
response = requests.get(
    f'https://{ip}/api/v3/search?q={host_name}&types=PhysicalMachine'
    f'&detail_type=entity',
    headers=authHeader, verify=False)
get_vms_response_json = json.loads(response.text)
```

```
print(f'CONSUMERS: {get_vms_response_json[0]["consumers"]}')
print(f'The {entity_type} members of this dynamic group will be:')
for entity in get_vms_response_json[0]['consumers']:
 if entity['className'] == entity_type:
  print(entity)
```

```
#
# Create the filter criterion
#
```

```
criterion = {}
criterion['expType'] = "RXEQ"
criterion['expVal'] = host_name
criterion['filterType'] = "vmsByPMName"
criterion['caseSensitive'] = "false"
```

```
#
# Construct the Input DTO and Create the Group
#
group_input_dto = {}
group_input_dto['isStatic'] = 'false'
group_input_dto['displayName'] = group_name
group_input_dto['memberUuidList'] = []
group_input_dto['criteriaList'] = [criterion]
group_input_dto['groupType'] = entity_type
```

```
response = requests.post(
    f'https://{ip}/api/v3/groups', headers=authHeader, json=group_input_dto, verify=False, stream=True)
group_creation_response_json = json.loads(response.text)
group_uuid = group_creation_response_json['uuid']
```

```
#
# Verify the group
#
response = requests.get(
    f'https://{ip}/api/v3/groups/{group_uuid}/entities', headers=authHeader, verify=False)
group_members_response_json = json.loads(response.text)
```

```
# Print a list of group members
print('\nGROUP MEMBERS:')
for item in group_members_response_json:
    print('Display name: ' + item['displayName'])
    print('OID: ' + item['uuid'])
```

Reference: Internal Workload Optimization Manager API requests

Internal Requests are requests that are used internally by Workload Optimization Manager, but are not intended for customer use. The structure data returned by these requests is subject to change without notice.

The following requests have been marked as Internal Only:

- /entities

  `DELETE /entities/loggingEntities`

  `PUT /entities/{entity_Uuids}/loggingEntities`

  `GET /entities/{entity_Uuid}/aspects/{aspect_name}/actions`

  `POST /entities/{entity_Uuid}/aspects/{aspect_name}/actions`

  `GET /entities/{entity_Uuid}/aspects/{aspect_name}/stats`

  `POST /entities/{entity_Uuid}/aspects/{aspect_name}/stats`

  `POST /entities/{entity_Uuid}/tags`

  `DELETE /entities/{entity_Uuid}/tags`

  `DELETE /entities/{entity_Uuid}/tags/{tag_key}`

- /groups

```
GET /groups/{group_Uuid}/aspects/{aspect_name}/actions
POST /groups/{group_Uuid}/aspects/{aspect_name}/actions
GET /groups/{group_Uuid}/aspects/{aspect_name}/stats
POST /groups/{group_Uuid}/aspects/{aspect_name}/stats
POST /groups/{group_Uuid}/notifications/stats
POST /groups/{group_Uuid}/tags
DELETE /groups/{group_Uuid}/tags
DELETE /groups/{group_Uuid}/tags/{tag_key}
```

- /markets

  ```
  PUT /markets/{market_Uuid}/policies/{policy_Uuid}
  ```

- /reports

  ```
  POST /reports/templates
  ```

- /scenarios

  ```
  PUT /scenarios/{scenario_Uuid}
  PUT /scenarios/{scenario_Uuid}/entities/{entity_Uuid}
  DELETE /scenarios/{scenario_Uuid}/entities/{entity_Uuid}
  PUT /scenarios/{scenario_Uuid}/groups/{group_Uuid}
  PUT /scenarios/{scenario_Uuid}/policies
  DELETE /scenarios/{scenario_Uuid}/policies/{policy_Uuid}
  PUT /scenarios/{scenario_Uuid}/templates/{template_name}
  ```

- /widgetsets

  The entire /widgetsets endpoint is Internal Only.

Reference: Scenario peak-of-peak full scenario input DTO

For reference purposes, see the following example for a complete input DTO for a Peak of Peaks plan.

```
    {
"uuid":"XXXXDemo",
"displayName":"XXXXDemo",
"owners":[
  {
    "uuid":"_4T_7kwY-Ed-WUKbEYSVIDw",
    "username":"administrator",
    "showSharedUserSC":false
  }
],
"type":"PEAKOFPEAKS",
"scope":[
  {
    "uuid":"_PE0v-YEUEee_hYfzgV9uYg",
    "displayName":"All On-prem Hosts",
    "className":"All On-prem Hosts"
  }
],
"topologyChanges":{
```

```
},
"loadChanges":{
  "maxUtilizationList":[
    {
      "maxPercentage":70,
      "projectionDay":0,
      "target":{
        "displayName":"All On-prem Hosts",
        "className":"Group",
        "uuid":"_PE0v-YEUEee_hYfzgV9uYg"
      }
    }
  ]
},
"overlayStatsList":[
  {
    "uuid":"4223ae0c-c99b-f6ca-c1ef-3f367bdf5528",
    "displayName":"centos_imp_2",
    "className":"VirtualMachine",
    "environmentType":"ONPREM",
    "stats":[
      {
        "displayName":"centos_imp_2",
        "date":"2019-02-01T00:00:00-05:00",
        "statistics":[
          {
            "displayName":"centos_imp_2",
            "date":"2019-03-14T15:12:36-04:00",
            "statistics":[
              {
                "name":"priceIndex",
                "values":{
                  "max":20000,
                  "min":20000,
                  "avg":20000,
                  "total":20000
                },
                "value":20000
              },
              {
                "name":"CPUProvisioned",
                "capacity":{
                  "max":104000,
                  "min":104000,
                  "avg":104000,
                  "total":104000
                },
                "filters":[
                  {
                    "type":"relation",
                    "value":"bought"
                  }
                ],
                "relatedEntity":{
                  "uuid":"Virtual_ESX_4238255f-86cc-4cb9-0f47-11ff689f0b22",
                  "displayName":"hp-esx4.dev.mycorp.com",
                  "className":"PhysicalMachine"
                },
                "units":"MHz",
                "values":{
                  "max":5200,
                  "min":5200,
                  "avg":5200,
                  "total":5200
                },
```

```
      "value":5200
    },
    {
      "name":"Mem",
      "capacity":{
        "max":16776692,
        "min":16776692,
        "avg":16776692,
        "total":16776692
      },
      "filters":[
        {
          "type":"relation",
          "value":"bought"
        }
      ],
      "relatedEntity":{
        "uuid":"Virtual_ESX_4238255f-86cc-4cb9-0f47-11ff689f0b22",
        "displayName":"hp-esx4.dev.mycorp.com",
        "className":"PhysicalMachine"
      },
      "units":"KB",
      "values":{
        "max":3555836,
        "min":3554799.8,
        "avg":3554799.8,
        "total":3554799.8
      },
      "value":3554799.8
    },
    {
      "name":"MemProvisioned",
      "capacity":{
        "max":167766912,
        "min":167766912,
        "avg":167766912,
        "total":167766912
      },
      "filters":[
        {
          "type":"relation",
          "value":"bought"
        }
      ],
      "relatedEntity":{
        "uuid":"Virtual_ESX_4238255f-86cc-4cb9-0f47-11ff689f0b22",
        "displayName":"hp-esx4.dev.mycorp.com",
        "className":"PhysicalMachine"
      },
      "units":"KB",
      "values":{
        "max":4194304,
        "min":4194304,
        "avg":4194304,
        "total":4194304
      },
      "value":4194304
    },
    {
      "name":"StorageAmount",
      "capacity":{
        "max":3301375,
        "min":3301375,
        "avg":3301375,
        "total":3301375
```

```
      },
      "filters":[
        {
          "type":"relation",
          "value":"bought"
        }
      ],
      "relatedEntity":{
        "uuid":"6f76b47e-e6f0a3fe",
        "displayName":"QSGRID01:ESXDC1DS1",
        "className":"Storage"
      },
      "units":"MB",
      "values":{
        "max":725.32,
        "min":725.32,
        "avg":725.32,
        "total":725.32
      },
      "value":725.32
    },
    {
      "name":"NetThroughput",
      "capacity":{
        "max":3072000,
        "min":3072000,
        "avg":3072000,
        "total":3072000
      },
      "filters":[
        {
          "type":"relation",
          "value":"bought"
        }
      ],
      "relatedEntity":{
        "uuid":"Virtual_ESX_4238255f-86cc-4cb9-0f47-11ff689f0b22",
        "displayName":"hp-esx4.dev.mycorp.com",
        "className":"PhysicalMachine"
      },
      "units":"Kbit/sec",
      "values":{
        "max":0,
        "min":0,
        "avg":0,
        "total":0
      },
      "value":0
    },
    {
      "name":"MemAllocation",
      "capacity":{
        "max":52503552,
        "min":52503552,
        "avg":52503552,
        "total":52503552
      },
      "filters":[
        {
          "type":"key",
          "value":"VirtualDataCenter::2332ba0f9dc14c07f11adb38e12df43c4bc8ca56"
        },
        {
          "type":"relation",
          "value":"bought"
```

```
        }
      ],
      "relatedEntity":{
        "uuid":"2332ba0f9dc14c07f11adb38e12df43c4bc8ca56",
        "displayName":"olga_do_not_touch_1",
        "className":"VirtualDataCenter"
      },
      "units":"KB",
      "values":{
        "max":3555836,
        "min":3554799.8,
        "avg":3554799.8,
        "total":3554799.8
      },
      "value":3554799.8
    },
    {
      "name":"VCPU",
      "capacity":{
        "max":5200,
        "min":5200,
        "avg":5200,
        "total":5200
      },
      "filters":[
        {
          "type":"relation",
          "value":"sold"
        }
      ],
      "units":"MHz",
      "values":{
        "max":5214,
        "min":5201.9,
        "avg":5201.9,
        "total":5201.9
      },
      "value":5201.9
    },
    {
      "name":"Ballooning",
      "capacity":{
        "max":16776692,
        "min":16776692,
        "avg":16776692,
        "total":16776692
      },
      "filters":[
        {
          "type":"relation",
          "value":"bought"
        }
      ],
      "relatedEntity":{
        "uuid":"Virtual_ESX_4238255f-86cc-4cb9-0f47-11ff689f0b22",
        "displayName":"hp-esx4.dev.mycorp.com",
        "className":"PhysicalMachine"
      },
      "units":"KB",
      "values":{
        "max":0,
        "min":0,
        "avg":0,
        "total":0
      },
```

```
      "value":0
    },
    {
      "name":"StorageAccess",
      "capacity":{
        "max":5000,
        "min":5000,
        "avg":5000,
        "total":5000
      },
      "filters":[
        {
          "type":"relation",
          "value":"bought"
        }
      ],
      "relatedEntity":{
        "uuid":"6f76b47e-e6f0a3fe",
        "displayName":"QSGRID01:ESXDC1DS1",
        "className":"Storage"
      },
      "units":"IOPS",
      "values":{
        "max":0,
        "min":0,
        "avg":0,
        "total":0
      },
      "value":0
    },
    {
      "name":"CPUAllocation",
      "capacity":{
        "max":31096,
        "min":31096,
        "avg":31096,
        "total":31096
      },
      "filters":[
        {
          "type":"key",
          "value":"VirtualDataCenter::2332ba0f9dc14c07f11adb38e12df43c4bc8ca56"
        },
        {
          "type":"relation",
          "value":"bought"
        }
      ],
      "relatedEntity":{
        "uuid":"2332ba0f9dc14c07f11adb38e12df43c4bc8ca56",
        "displayName":"olga_do_not_touch_1",
        "className":"VirtualDataCenter"
      },
      "units":"MHz",
      "values":{
        "max":5214,
        "min":5201.9,
        "avg":5201.9,
        "total":5201.9
      },
      "value":5201.9
    },
    {
      "name":"IOThroughput",
      "capacity":{
```

```
                "max":11264000,
                "min":11264000,
                "avg":11264000,
                "total":11264000
              },
              "reserved":{
                "max":9011200,
                "min":9011200,
                "avg":9011200,
                "total":9011200
              },
              "filters":[
                {
                  "type":"relation",
                  "value":"bought"
                }
              ],
              "relatedEntity":{
                "uuid":"Virtual_ESX_4238255f-86cc-4cb9-0f47-11ff689f0b22",
                "displayName":"hp-esx4.dev.mycorp.com",
                "className":"PhysicalMachine"
              },
              "units":"Kbit/sec",
              "values":{
                "max":0,
                "min":0,
                "avg":0,
                "total":0
              },
              "value":0
            },
            {
              "name":"StorageLatency",
              "capacity":{
                "max":100,
                "min":100,
                "avg":100,
                "total":100
              },
              "filters":[
                {
                  "type":"relation",
                  "value":"bought"
                }
              ],
              "relatedEntity":{
                "uuid":"6f76b47e-e6f0a3fe",
                "displayName":"QSGRID01:ESXDC1DS1",
                "className":"Storage"
              },
              "units":"msec",
              "values":{
                "max":0.5,
                "min":0.5,
                "avg":0.5,
                "total":0.5
              },
              "value":0.5
            },
            {
              "name":"VMem",
              "capacity":{
                "max":4194304,
                "min":4194304,
                "avg":4194304,
```

```
        "total":4194304
      },
      "filters":[
        {
          "type":"relation",
          "value":"sold"
        }
      ],
      "units":"KB",
      "values":{
        "max":3397384,
        "min":3233807,
        "avg":3233807,
        "total":3233807
      },
      "value":3233807
    },
    {
      "name":"CPU",
      "capacity":{
        "max":10400,
        "min":10400,
        "avg":10400,
        "total":10400
      },
      "filters":[
        {
          "type":"relation",
          "value":"bought"
        }
      ],
      "relatedEntity":{
        "uuid":"Virtual_ESX_4238255f-86cc-4cb9-0f47-11ff689f0b22",
        "displayName":"hp-esx4.dev.mycorp.com",
        "className":"PhysicalMachine"
      },
      "units":"MHz",
      "values":{
        "max":5214,
        "min":5201.9,
        "avg":5201.9,
        "total":5201.9
      },
      "value":5201.9
    },
    {
      "name":"VStorage",
      "capacity":{
        "max":2646,
        "min":1014,
        "avg":2238,
        "total":8952
      },
      "filters":[
        {
          "type":"key",
          "value":"VirtualMachine::08bb59f519b55c22415be8d623b5ccdf578b85cc"
        },
        {
          "type":"relation",
          "value":"sold"
        }
      ],
      "units":"MB",
      "values":{
```

```
                "max":1016.14,
                "min":142.4,
                "avg":797.71,
                "total":3190.82
              },
              "value":797.71
            }
          ]
        }
      ]
    },
    {
      "uuid":"4223bb37-0090-f332-0140-cc3cce7b3735",
      "displayName":"centos_imp_3",
      "className":"VirtualMachine",
      "environmentType":"ONPREM",
      "stats":[
        {
          "displayName":"centos_imp_3",
          "date":"2019-02-01T00:00:00-05:00",
          "statistics":[
            {
              "displayName":"centos_imp_3",
              "date":"2019-03-14T15:12:36-04:00",
              "statistics":[
                {
                  "name":"priceIndex",
                  "values":{
                    "max":20000,
                    "min":20000,
                    "avg":20000,
                    "total":20000
                  },
                  "value":20000
                },
                {
                  "name":"CPUProvisioned",
                  "capacity":{
                    "max":104000,
                    "min":104000,
                    "avg":104000,
                    "total":104000
                  },
                  "filters":[
                    {
                      "type":"relation",
                      "value":"bought"
                    }
                  ],
                  "relatedEntity":{
                    "uuid":"Virtual_ESX_4238a511-8593-4469-2751-46611e09b6a8",
                    "displayName":"hp-esx1.dev.mycorp.com",
                    "className":"PhysicalMachine"
                  },
                  "units":"MHz",
                  "values":{
                    "max":5200,
                    "min":5200,
                    "avg":5200,
                    "total":5200
                  },
                  "value":5200
                },
                {
                  "name":"Swapping",
```

```
      "capacity":{
        "max":40000000,
        "min":40000000,
        "avg":40000000,
        "total":40000000
      },
      "filters":[
        {
          "type":"relation",
          "value":"bought"
        }
      ],
      "relatedEntity":{
        "uuid":"Virtual_ESX_4238a511-8593-4469-2751-46611e09b6a8",
        "displayName":"hp-esx1.dev.mycorp.com",
        "className":"PhysicalMachine"
      },
      "units":"bit/sec",
      "values":{
        "max":0,
        "min":0,
        "avg":0,
        "total":0
      },
      "value":0
    },
    {
      "name":"NumDisk",
      "capacity":{
        "max":2147483650,
        "min":2147483650,
        "avg":2147483650,
        "total":2147483650
      },
      "filters":[
        {
          "type":"relation",
          "value":"bought"
        }
      ],
      "relatedEntity":{
        "uuid":"Virtual_ESX_4238a511-8593-4469-2751-46611e09b6a8",
        "displayName":"hp-esx1.dev.mycorp.com",
        "className":"PhysicalMachine"
      },
      "values":{
        "max":1,
        "min":1,
        "avg":1,
        "total":1
      },
      "value":1
    },
    {
      "name":"Mem",
      "capacity":{
        "max":16776692,
        "min":16776692,
        "avg":16776692,
        "total":16776692
      },
      "filters":[
        {
          "type":"relation",
          "value":"bought"
```

```
        }
      ],
      "relatedEntity":{
        "uuid":"Virtual_ESX_4238a511-8593-4469-2751-46611e09b6a8",
        "displayName":"hp-esx1.dev.mycorp.com",
        "className":"PhysicalMachine"
      },
      "units":"KB",
      "values":{
        "max":3069332,
        "min":3062497,
        "avg":3062497,
        "total":3062497
      },
      "value":3062497
    },
    {
      "name":"MemProvisioned",
      "capacity":{
        "max":167766912,
        "min":167766912,
        "avg":167766912,
        "total":167766912
      },
      "filters":[
        {
          "type":"relation",
          "value":"bought"
        }
      ],
      "relatedEntity":{
        "uuid":"Virtual_ESX_4238a511-8593-4469-2751-46611e09b6a8",
        "displayName":"hp-esx1.dev.mycorp.com",
        "className":"PhysicalMachine"
      },
      "units":"KB",
      "values":{
        "max":4194304,
        "min":4194304,
        "avg":4194304,
        "total":4194304
      },
      "value":4194304
    },
    {
      "name":"StorageAmount",
      "capacity":{
        "max":3301375,
        "min":3301375,
        "avg":3301375,
        "total":3301375
      },
      "filters":[
        {
          "type":"relation",
          "value":"bought"
        }
      ],
      "relatedEntity":{
        "uuid":"6f76b47e-e6f0a3fe",
        "displayName":"QSGRID01:ESXDC1DS1",
        "className":"Storage"
      },
      "units":"MB",
      "values":{
```

```
        "max":719,
        "min":719,
        "avg":719,
        "total":719
      },
      "value":719
    },
    {
      "name":"NetThroughput",
      "capacity":{
        "max":3072000,
        "min":3072000,
        "avg":3072000,
        "total":3072000
      },
      "filters":[
        {
          "type":"relation",
          "value":"bought"
        }
      ],
      "relatedEntity":{
        "uuid":"Virtual_ESX_4238a511-8593-4469-2751-46611e09b6a8",
        "displayName":"hp-esx1.dev.mycorp.com",
        "className":"PhysicalMachine"
      },
      "units":"Kbit/sec",
      "values":{
        "max":0,
        "min":0,
        "avg":0,
        "total":0
      },
      "value":0
    },
    {
      "name":"MemAllocation",
      "capacity":{
        "max":52503552,
        "min":52503552,
        "avg":52503552,
        "total":52503552
      },
      "filters":[
        {
          "type":"key",
          "value":"VirtualDataCenter::2332ba0f9dc14c07f11adb38e12df43c4bc8ca56"
        },
        {
          "type":"relation",
          "value":"bought"
        }
      ],
      "relatedEntity":{
        "uuid":"2332ba0f9dc14c07f11adb38e12df43c4bc8ca56",
        "displayName":"olga_do_not_touch_1",
        "className":"VirtualDataCenter"
      },
      "units":"KB",
      "values":{
        "max":3069332,
        "min":3062497,
        "avg":3062497,
        "total":3062497
      },
```

```
      "value":3062497
    },
    {
      "name":"VCPU",
      "capacity":{
        "max":5200,
        "min":5200,
        "avg":5200,
        "total":5200
      },
      "filters":[
        {
          "type":"relation",
          "value":"sold"
        }
      ],
      "units":"MHz",
      "values":{
        "max":5203,
        "min":5194.67,
        "avg":5194.67,
        "total":5194.67
      },
      "value":5194.67
    },
    {
      "name":"Ballooning",
      "capacity":{
        "max":16776692,
        "min":16776692,
        "avg":16776692,
        "total":16776692
      },
      "filters":[
        {
          "type":"relation",
          "value":"bought"
        }
      ],
      "relatedEntity":{
        "uuid":"Virtual_ESX_4238a511-8593-4469-2751-46611e09b6a8",
        "displayName":"hp-esx1.dev.mycorp.com",
        "className":"PhysicalMachine"
      },
      "units":"KB",
      "values":{
        "max":0,
        "min":0,
        "avg":0,
        "total":0
      },
      "value":0
    },
    {
      "name":"StorageAccess",
      "capacity":{
        "max":5000,
        "min":5000,
        "avg":5000,
        "total":5000
      },
      "filters":[
        {
          "type":"relation",
          "value":"bought"
```

```
        }
      ],
      "relatedEntity":{
        "uuid":"6f76b47e-e6f0a3fe",
        "displayName":"QSGRID01:ESXDC1DS1",
        "className":"Storage"
      },
      "units":"IOPS",
      "values":{
        "max":0,
        "min":0,
        "avg":0,
        "total":0
      },
      "value":0
    },
    {
      "name":"CPUAllocation",
      "capacity":{
        "max":31096,
        "min":31096,
        "avg":31096,
        "total":31096
      },
      "filters":[
        {
          "type":"key",
          "value":"VirtualDataCenter::2332ba0f9dc14c07f11adb38e12df43c4bc8ca56"
        },
        {
          "type":"relation",
          "value":"bought"
        }
      ],
      "relatedEntity":{
        "uuid":"2332ba0f9dc14c07f11adb38e12df43c4bc8ca56",
        "displayName":"olga_do_not_touch_1",
        "className":"VirtualDataCenter"
      },
      "units":"MHz",
      "values":{
        "max":5203,
        "min":5194.67,
        "avg":5194.67,
        "total":5194.67
      },
      "value":5194.67
    },
    {
      "name":"IOThroughput",
      "capacity":{
        "max":11264000,
        "min":11264000,
        "avg":11264000,
        "total":11264000
      },
      "reserved":{
        "max":9011200,
        "min":9011200,
        "avg":9011200,
        "total":9011200
      },
      "filters":[
        {
          "type":"relation",
```

```
            "value":"bought"
          }
        ],
        "relatedEntity":{
          "uuid":"Virtual_ESX_4238a511-8593-4469-2751-46611e09b6a8",
          "displayName":"hp-esx1.dev.mycorp.com",
          "className":"PhysicalMachine"
        },
        "units":"Kbit/sec",
        "values":{
          "max":0,
          "min":0,
          "avg":0,
          "total":0
        },
        "value":0
      },
      {
        "name":"StorageLatency",
        "capacity":{
          "max":100,
          "min":100,
          "avg":100,
          "total":100
        },
        "filters":[
          {
            "type":"relation",
            "value":"bought"
          }
        ],
        "relatedEntity":{
          "uuid":"6f76b47e-e6f0a3fe",
          "displayName":"QSGRID01:ESXDC1DS1",
          "className":"Storage"
        },
        "units":"msec",
        "values":{
          "max":0.33,
          "min":0.33,
          "avg":0.33,
          "total":0.33
        },
        "value":0.33
      },
      {
        "name":"Q2VCPU",
        "capacity":{
          "max":20000,
          "min":20000,
          "avg":20000,
          "total":20000
        },
        "reserved":{
          "max":10000,
          "min":10000,
          "avg":10000,
          "total":10000
        },
        "filters":[
          {
            "type":"relation",
            "value":"bought"
          }
        ],
```

```
      "relatedEntity":{
        "uuid":"Virtual_ESX_4238a511-8593-4469-2751-46611e09b6a8",
        "displayName":"hp-esx1.dev.mycorp.com",
        "className":"PhysicalMachine"
      },
      "units":"msec",
      "values":{
        "max":11,
        "min":11,
        "avg":11,
        "total":11
      },
      "value":11
    },
    {
      "name":"VMem",
      "capacity":{
        "max":4194304,
        "min":4194304,
        "avg":4194304,
        "total":4194304
      },
      "filters":[
        {
          "type":"relation",
          "value":"sold"
        }
      ],
      "units":"KB",
      "values":{
        "max":2894068,
        "min":2747267,
        "avg":2747267,
        "total":2747267
      },
      "value":2747267
    },
    {
      "name":"CPU",
      "capacity":{
        "max":10400,
        "min":10400,
        "avg":10400,
        "total":10400
      },
      "filters":[
        {
          "type":"relation",
          "value":"bought"
        }
      ],
      "relatedEntity":{
        "uuid":"Virtual_ESX_4238a511-8593-4469-2751-46611e09b6a8",
        "displayName":"hp-esx1.dev.mycorp.com",
        "className":"PhysicalMachine"
      },
      "units":"MHz",
      "values":{
        "max":5203,
        "min":5194.67,
        "avg":5194.67,
        "total":5194.67
      },
      "value":5194.67
    },
```

```
                {
                  "name":"VStorage",
                  "capacity":{
                    "max":2646,
                    "min":1014,
                    "avg":2238,
                    "total":8952
                  },
                  "filters":[
                    {
                      "type":"key",
                      "value":"VirtualMachine::199bb3ec4f8da89f01c126a6b3df4c1321c4b07d"
                    },
                    {
                      "type":"relation",
                      "value":"sold"
                    }
                  ],
                  "units":"MB",
                  "values":{
                    "max":1016.01,
                    "min":142.4,
                    "avg":797.61,
                    "total":3190.42
                  },
                  "value":797.61
                }
              ]
            }
          ]
        }
      ]
    },
    "timebasedTopologyChanges":{
```

```
    }
  }
```


Reference: Tables

Reference tables overview

This section contains reference tables for common API parameters, separated by endpoint. Note that filter tables may contain other reference tables for completeness.

Action Reference Tables

Action Mode:

- `DISABLED`:

  Do not recommend or perform the action. When you disable an action, the user interface will no longer display actions of that type.
- `RECOMMENDED`:

  Recommend the action so a user can execute it via the given hypervisor or by other external means
- `MANUAL`:

  Recommend the action, and provide the option to execute that action through the Workload Optimization Manager user interface or an /actions API request.

- `AUTOMATED`:

  You have directed Workload Optimization Manager to execute the action automatically.
- `COLLECTION` **(UNUSED)**:

  Legacy action mode, no longer used.

Action State:

- `DISABLED`:

  An action whose policy has been disabled.
- `RECOMMEND`:

  An action that cannot be automated due to policy settings or must be executed outside the system, such as the purchase of hardware.
- `PENDING_ACCEPT`:

  When the action mode is set to manual, a recommended action that has not been accepted or cleared.
- `ACCEPTED`:

  An action that has been accepted, but is not yet in progress.
- `IN_PROGRESS`:

  An action that is being executed.
- `SUCCEEDED`:

  An action that was accepted and completed successfully.
- `FAILED`:

  An action that was accepted and did not complete successfully.
- `REJECTED`:

  When the action mode is set to manual, an action that has been rejected by the user.
- `CLEARED`:

  When the action mode is set to manual, an action that is no longer recommended by the market.
- `ACCOUNTING`:

  For cloud entities, an action to resize in order to use a different instance template.
- `QUEUED`:

  When more than 10 actions are to be executed for a single target, Workload Optimization Manager will place the 11th and all subsequent actions into a QUEUED state, and they will execute as the previous actions complete.

Action Type:

- `START`:

  Start an entity, and add it to the Workload Optimization Manager market.
- `MOVE`:

  Move an entity from one provider to another. For example, moving a VM between hosts, or a datastore between disk arrays.
- `SUSPEND`:

  Suspend the entity to exclude its capacity from your environment. Suspend does not remove the entity from your environment, but the market does not include the entity in its analysis. Based on market conditions, analysis can recommend restarting suspended entities, which includes them in market analysis once again.
- `ADD PROVIDER`:

  This is equivalent to the START action, but for storage entities.
- `CHANGE`:

  Move a virtual machine's storage between datastores.
- `PROVISION`:

Add capacity to your environment.

- `RECONFIGURE`:

  Analysis recommends this when a VM is not able to correct a misconfiguration via another type of action. For example, analysis can recommend a RECONFIGURE action when a placement policy requires a VM to move to a host on a particular cluster, but cross-cluster moves are not permitted.

- `DELETE`:

  This is equivalent to the SUSPEND action, but for storage entities.

- `RESERVE_ON_PM`:

  When using the reservation functionality, this is the action to reserve the resources of a VM on a particular host.

- `RESERVE_ON_DS`:

  When using the reservation functionality, this is the action to reserve the resources of a VM on a particular datastore.

- `RESIZE_FOR_EFFICIENCY`:

  A resize that reduces the capacity of a commodity, such as vCPU, or vMEM.

- `RESIZE_FOR_PERFORMANCE`:

  A resize that increases the capacity of a commodity, such as vCPU, or vMEM.

Action Filter Criteria:

- actionModeList
  - `DISABLED`:

    Do not recommend or perform the action. When you disable an action, the user interface will no longer display actions of that type.

  - `RECOMMENDED`:

    Recommend the action so a user can execute it via the given hypervisor or by other external means

  - `MANUAL`:

    Recommend the action, and provide the option to execute that action through the Workload Optimization Manager user interface or an /actions API request.

  - `AUTOMATED`:

    You have directed Workload Optimization Manager to execute the action automatically.

  - `COLLECTION` **(UNUSED)**:

    Legacy action mode, no longer used.

- actionStateList
  - `DISABLED`:

    An action whose policy has been disabled.

  - `RECOMMEND`:

    An action that cannot be automated due to policy settings or must be executed outside the system, such as the purchase of hardware.

  - `PENDING_ACCEPT`:

    When the action mode is set to manual, a recommended action that has not been accepted or cleared.

  - `ACCEPTED`:

    An action that has been accepted, but is not yet in progress.

  - `IN_PROGRESS`:

    An action that is being executed.

  - `SUCCEEDED`:

    An action that was accepted and completed successfully.

  - `FAILED`:

An action that was accepted and did not complete successfully.

  - `REJECTED`:

    When the action mode is set to manual, an action that has been rejected by the user.

  - `CLEARED`:

    When the action mode is set to manual, an action that is no longer recommended by the market.

  - `ACCOUNTING`:

    For cloud entities, an action to resize in order to use a different instance template.

  - `QUEUED`:

    When more than 10 actions are to be executed for a single target, Workload Optimization Manager will place the 11th and all subsequent actions into a QUEUED state, and they will execute as the previous actions complete.

- actionTypeList
  - `START`:

    Start an entity, and add it to the Workload Optimization Manager market.

  - `MOVE`:

    Move an entity from one provider to another. For example, moving a VM between hosts, or a datastore between disk arrays.

  - `SUSPEND`:

    Suspend the entity to exclude its capacity from your environment. Suspend does not remove the entity from your environment, but the market does not include the entity in its analysis. Based on market conditions, analysis can recommend restarting suspended entities, which includes them in market analysis once again.

  - `ADD PROVIDER`:

    This is equivalent to the START action, but for storage entities.

  - `CHANGE`:

    Move a virtual machine's storage between datastores.

  - `PROVISION`:

    Add capacity to your environment.

  - `RECONFIGURE`:

    Analysis recommends this when a VM is not able to correct a misconfiguration via another type of action. For example, analysis can recommend a RECONFIGURE action when a placement policy requires a VM to move to a host on a particular cluster, but cross-cluster moves are not permitted.

  - `DELETE`:

    This is equivalent to the SUSPEND action, but for storage entities.

  - `RESERVE_ON_PM`:

    When using the reservation functionality, this is the action to reserve the resources of a VM on a particular host.

  - `RESERVE_ON_DS`:

    When using the reservation functionality, this is the action to reserve the resources of a VM on a particular datastore.

  - `RESIZE_FOR_EFFICIENCY`:

    A resize that reduces the capacity of a commodity, such as vCPU, or vMEM.

  - `RESIZE_FOR_PERFORMANCE`:

    A resize that increases the capacity of a commodity, such as vCPU, or vMEM.

- cleared

  Default: `false`.

  When `true`, actions that were removed from the action list before being executed will be returned.

- costType
  - `Saving`

    An action that will decrease cloud spending.
  - `Investment`

    An action that will increase cloud spending.
- cumulative

  Default: `false`.

  When `true`, shows the savings for each action in a scope as a cumulative total, rather than showing the savings for individual actions.
- environmentType
  - `ONPREM`:

    Returns only actions related to entities that are part of your on-prem environment.
  - `CLOUD`:

    Returns only actions related to entities that are part of your cloud environment.
- groupBy

  You may group the actions returned by the following criteria: `[actionModes, actionStates, actionTypes, risk, riskSeverity, riskSubCategory]`
- hasReservedInstance

  Default: `false`.

  When `true`, only actions that are related to reserved instances will be returned.
- relatedEntityTypes

  Returns only actions related to entities of these types within the scope. For example, if your scope is a virtual datacenter and you specify VirtualMachine as a related entity type, the request will include only those actions related to VMs in the specified virtual datacenter.

  Choose from the following entity types: `[ Datacenter, PhysicalMachine, VirtualMachine, Storage, Application, Chassis, DiskArray, IOModule, StorageControl, Switch, VirtualDataCenter, VPod, DPod, Container, Database, DatabaseServerContainer, LogicalPool ]`
- riskSeverityList

  Returns only actions of the given severities: `[ Unknown, Normal, Minor, Major, Critical ]`
- riskSubCategoryList
- returns only actions of the given subcategories: `[ Performance Assurance, Efficiency Improvement, Prevention, Compliance ]`

Entity Reference Tables

Entity State:

- ACTIVE:

  Entity is actively consuming resources.
- EVACUATED:

  Entity is ready to suspend and will not participate in the market, but will not generate a SUSPEND action.

  > **NOTE:** This state replaces the SUSPEND state, and requires additional configuration to enable. Workload Optimization Manager does not recommend enabling this state.

- FAILOVER:

  This entity is reserved for failover and will not participate in the market.

- `IDLE`:

  Entity is powered off and not actively consuming market resources.
- `LAUNCH`:

  Entity is starting.
- `MAINTENANCE`:

  Entity is in maintenance mode.
- `NOT_MONITORED`:

  Entity is not currently monitored by Workload Optimization Manager.
- `QUEUED`:

  Entity is in the process of changing from one state to another.

  NOTE: This state is internal to Workload Optimization Manager and will not appear in the UI, but may appear in an API response if the call is executed at the precise moment the entity is changing state.
- `RESOURCE_ALLOCATION`:

  Cloud entity is being created.
- `RESOURCE_RELEASE`:

  Cloud Entity is being deleted.
- `SUSPEND`:

  Entity is suspended or the guest OS is in sleep, standby, or suspended state. The entity is not removed from the environment, but will not be used in market decisions. Based on market conditions, suspended entities can be started, re-adding them to the market.
- `SUSPEND_PENDING`:

  Entity is making preparations to suspend, or is the next entity of its type to suspend.
- `TERMINATE_PENDING`:

  Entity is in the process of being deleted.
- `UNKNOWN`:

  Workload Optimization Manager is unable to obtain the current state of the entity.

  This occurs when the entity is reported to the hypervisor as having a state of `DISCONNECTED, ORPHANED,` or `INVALID`.

Entity Type:

- `APPLICATION`
- `APPLICATION_SERVER`
- `BUSINESS_ACCOUNT`
- `BUSINESS_UNIT`
- `CHASSIS`
- `CLOUD_SERVICE`
- `CONTAINER`
- `CONTAINER_POD`
- `DATABASE`
- `DATABASE_SERVER`
- `DATACENTER`
- `DISK_ARRAY`
- `IO_MODULE`

- LOAD_BALANCER
- LOGICAL_POOL
- NETWORK
- PHYSICAL_MACHINE
- STORAGE
- STORAGE_CONTROLLER
- SWITCH
- VIRTUAL_APPLICATION
- VIRTUAL_DATACENTER
- VIRTUAL_MACHINE

Reference: DTO index

Data transfer objects (DTOs) are used to encapsulate and send data to the Workload Optimization Manager API. The following DTOs are used by Workload Optimization Manager:

> **NOTE:**
> The DTO Index is a work in progress, and included in a draft stage for reference. The information contained in the Index is accurate, but may not be complete.

AcceleratorCardApiDTO

Description

Model to describe info about an Accelerator/GPU Card

Optional Parameters

acceleratorFamily

- **type:** string
- **description:** Accelerator/GPU Micro architecture family E.G NVIDIA_TURING, NVIDIA_AMPERE, AMD_RADEON_PRO

acceleratorModel

- **type:** string
- **description:** Accelerator/GPU Micro model. E.G NVIDIA_T4, NVIDIA_T4G, NVIDIA_A10G, AMD_V520

numCores

- **type:** integer
- **format:** int32
- **description:** Number of Accelerator/GPU cards

memoryPerCore

- **type:** number
- **format:** double
- **description:** Memory in GB per Accelerator/GPU card

partitions

- **type:** integer
- **format:** int32
- **description:** Accelerator partitions if partial GPUs.

ActionApiDTO

Description

Model to describe an Action

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

actionImpactID

- **type:** integer
- **format:** int64
- **description:** The ID for the action, which will persist across restarts

marketID

- **type:** integer
- **format:** int64
- **description:** The ID of the market for which the action was generated

createTime

- **type:** string
- **description:** Creation time

updateTime

- **type:** string
- **description:** Update time

clearTime

- **type:** string
- **description:** Execution time

actionType

- **type:** string
- **description:** Type
- **enum:** ['START', 'MOVE', 'SCALE', 'ALLOCATE', 'SUSPEND', 'PROVISION', 'RECONFIGURE', 'RESIZE', 'DELETE', 'RIGHT_SIZE', 'BUY_RI']

actionState

- **type:** string
- **description:** Action State.
- **enum:** ['ACCEPTED', 'REJECTED', 'PRE_IN_PROGRESS', 'POST_IN_PROGRESS', 'IN_PROGRESS', 'SUCCEEDED', 'FAILED', 'DISABLED', 'QUEUED', 'CLEARED', 'ACCOUNTING', 'READY', 'FAILING', 'BEFORE_EXEC', 'IN_PROGRESS_NON_CRITICAL']

actionMode

- **type:** string
- **description:** Action Mode
- **enum:** ['DISABLED', 'RECOMMEND', 'EXTERNAL_APPROVAL', 'MANUAL', 'AUTOMATIC', 'COLLECTION']

userName

- **type:** string
- **description:** Username of the user that accepted the action [DEPRECATED - Use: userInfo.userName in executionStatus.executorInfo]

details

- **type:** string
- **description:** A user-readable string describing the action

importance

- **type:** number
- **format:** float
- **description:** Numeric value that describes the priority of the action

target

- **description:** Target entity for an action. For example, the VM in a Resize Action, or the host for a VM move
- **reference:** ServiceEntityApiDTO *(on page 1668)*

currentEntity

- **description:** Current entity, such as the current host that a VM resides on for a VM move action
- **reference:** ServiceEntityApiDTO *(on page 1668)*

newEntity

- **description:** Destination entity, such as the host that a VM will move to for a VM move action
- **reference:** ServiceEntityApiDTO *(on page 1668)*

currentValue

- **type:** string
- **description:** Current value of a property, for example vMEM for a VM resize action

newValue

- **type:** string
- **description:** Calculated value to resize to, such as vMEM for a VM resize action

valueUnits

- **type:** string
- **description:** Units of the currentValue and newValue, such as KB for a VM vMEM resize action

resizeAttribute

- **type:** string
- **description:** The commodity attribute to be resized
- **enum:** ['CAPACITY', 'LIMIT', 'RESERVATION']

template

- **description:** Template used for the action, such as in a Cloud entity provision or move action
- **reference:** TemplateApiDTO *(on page 1697)*

risk

- **description:** Cause of the action
- **reference:** LogEntryApiDTO *(on page 1581)*

prerequisites

- **type:** array
- **description:** Description(s) of any prerequisite(s) that must occur before execution is possible. If this field is populated, the action would otherwise be executable but has been blocked by this prerequisite, and is displayed accordingly. If this field is not populated, no prerequisite has caused the action mode to be modified, and the action is displayed as normal.
- – **type:** string

prioritizedAutomationOpportunities

- **type:** array
- **description:** The Action is generated using the 'Prioritze Automation' setting, indicating that it is possible to have more savings if these opportunities are satisfied
- – **type:** string

stats

- **type:** array
- **description:** Statistics, such as Mem, vCPU, costPrice
- – **reference:** StatApiDTO *(on page 1679)*

progress

- **type:** integer
- **format:** int32
- **description:** Moved the 'progress' percentage value inside 'executionStatus', available when action detail level is 'EXECUTION' - [DEPRECATED - Use: executionStatus.progress]

currentLocation

- **description:** The region (DataCenter) where the current service entity is located, for cloud migration actions
- **reference:** DiscoveredEntityApiDTO *(on page 1554)*

newLocation

- **description:** The region, represented as a DataCenter entity, where the target service entity will be located for cloud migration actions
- **reference:** DiscoveredEntityApiDTO *(on page 1554)*

virtualDisks

- **type:** array
- **description:** Virtual Disk used for the action, for example in a storage move
- – **reference:** VirtualDiskApiDTO *(on page 1713)*

policy

- **description:** The policy associated with the action
- **reference:** PolicyApiDTO *(on page 1622)*

relatedSettingsPolicies

- **type:** array
- **description:** The related settings policies associated with the action
- – **reference:** BaseApiDTO *(on page 1509)*

actionSchedule

- **description:** The schedule associated with the action
- **reference:** ActionScheduleApiDTO *(on page 1498)*

compoundActions

- **type:** array
- **description:** Property for compound move actions
- – **reference:** ActionApiDTO *(on page 1485)*

reservedInstance

- **description:** The Reserved Instance associated with the action
- **reference:** ReservedInstanceApiDTO *(on page 1647)*

reservedInstancePlatform

- **type:** string
- **description:** The platform of the recommended RI if the action has a supersaving

externalActionName

- **type:** string
- **description:** The name under which this action appears in third-party software or external approval backend

externalActionUrl

- **type:** string
- **description:** URL to third-party software showing this action. Suitable for external action approval

executionStatus

- **description:** Status details of the execution of the Action. Include percentage progress, error messages and execution times.
- **reference:** ActionExecutionAuditApiDTO *(on page 1493)*

executionCharacteristics

- **description:** Characteristics of the execution of the Action. Include disruptiveness and reversibility.
- **reference:** ActionExecutionCharacteristicApiDTO *(on page 1494)*

entityCharacteristics

- **description:** Characteristics of the entity the action is for.
- **reference:** DBEntityAspectApiDTO *(on page 1540)*

source

- **type:** string
- **description:** Defines what generated this action.
- **enum:** ['MARKET', 'PARKING']

relatedActionsCountByType

- **type:** object
- **description:** [BETA]: Map of related actions count by relationship type that have some relation to this Action
- **additionalProperties:** {'type': 'integer', 'format': 'int32'}

actionID

- **type:** integer
- **format:** int64
- **description:** Id of the action

ActionApiInputDTO

Description

Model to describe a Filter for Actions

Required Parameters

startTime

- **type:** string
- **description:** Start time of the period to filter, based on Actions Creation Time

Optional Parameters

actionStateList

- **type:** array
- **description:** List of States to filter.
- – **type:** string
  – **enum:** ['ACCEPTED', 'REJECTED', 'PRE_IN_PROGRESS', 'POST_IN_PROGRESS', 'IN_PROGRESS', 'SUCCEEDED', 'FAILED', 'DISABLED', 'QUEUED', 'CLEARED', 'ACCOUNTING', 'READY', 'FAILING', 'BEFORE_EXEC', 'IN_PROGRESS_NON_CRITICAL']

actionModeList

- **type:** array
- **description:** List of Modes to filter
- – **type:** string
  – **enum:** ['DISABLED', 'RECOMMEND', 'EXTERNAL_APPROVAL', 'MANUAL', 'AUTOMATIC', 'COLLECTION']

riskSeverityList

- **type:** array
- **description:** List of Risk Severities to filter
- – **type:** string

riskSubCategoryList

- **type:** array
- **description:** List of Risk Severities to filter
- – **type:** string

actionRelationTypeFilter

- **type:** array
- **description:** Filters the main list of actions based on the types of relationships they have to other actions.
- – **type:** string
  – **enum:** ['NONE', 'BLOCKED_BY', 'BLOCKING', 'CAUSED_BY', 'CAUSING']

environmentType

- **type:** string
- **description:** Filter the actions by Environment Type
- **enum:** ['ONPREM', 'CLOUD']

costType

- **type:** string
- **description:** Filter the actions by cost Type
- **enum:** ['SAVING', 'INVESTMENT']

descriptionQuery

- **description:** This query will match the details field of the ActionApiDTO to return actions that match the specified expression.
- **reference:** QueryInputApiDTO *(on page 1634)*

riskQuery

- **description:** This query will match the risk description field of the ActionApiDTO to return actions that match the specified expression.
- **reference:** QueryInputApiDTO *(on page 1634)*

executionCharacteristics

- **description:** Characteristics of the execution of the action, including disruptiveness and reversibility.
- **reference:** ActionExecutionCharacteristicInputApiDTO *(on page 1494)*

savingsAmountRange

- **description:** This field will match the costPrice statistic field of the ActionApiDTO to return actions that have a costPrice amount between minValue and maxValue, inclusive
- **reference:** RangeInputApiDTO *(on page 1637)*

effectiveCostSavingsAmountRange

- **description:** This field will match the effectiveCost statistic field of the ActionApiDTO to return actions that have an effectiveCost amount between minValue and maxValue, inclusive
- **reference:** RangeInputApiDTO *(on page 1637)*

unattachedDays

- **description:** [BETA]: Filter the actions based on an unattached days range and is applicable only to delete wasted file actions.
- **reference:** RangeInputApiDTO *(on page 1637)*

hasSchedule

- **type:** boolean
- **description:** If true, only actions that have an associated execution schedule will be returned.

hasPrerequisites

- **type:** boolean
- **description:** If true, only actions that have a prerequisite preventing execution of the action will be returned.

endTime

- **type:** string
- **description:** End time of the period to filter, based on Actions Creation Time

cleared

- **type:** boolean
- **description:** Filter by Actions Executed

cumulative

- **type:** boolean
- **description:** Show the action savings as cumulative values

hasReservedInstance

- **type:** boolean
- **description:** Filter the actions by whether having related reserved instance

excludeNotifications

- **type:** boolean
- **description:** Exclude the notifications from the result

limitEntities

- **type:** integer
- **format:** int32
- **description:** Limit number of entities in the result

detailLevel

- **type:** string
- **description:** Detail level of the Action, 'standard' is the default value
- **enum:** ['STANDARD', 'EXECUTION']

vendorId

- **type:** string
- **description:** Vendor Id indicating the related vendor to the entity in the actions

actionTypeList

- **type:** array
- **description:** List of Types to filter
- – **type:** string
  – **enum:** ['NONE', 'START', 'MOVE', 'SCALE', 'ALLOCATE', 'SUSPEND', 'TERMINATE', 'SPAWN', 'ADD_PROVIDER', 'CHANGE', 'REMOVE_PROVIDER', 'PROVISION', 'RECONFIGURE', 'RESIZE', 'RESIZE_CAPACITY', 'WARN', 'RECONFIGURE_THRESHOLD', 'DELETE', 'RIGHT_SIZE', 'RESERVE_ON_PM', 'RESERVE_ON_DS', 'RESIZE_FOR_EFFICIENCY', 'RESIZE_FOR_PERFORMANCE', 'CROSS_TARGET_MOVE', 'MOVE_TOGETHER', 'ACTIVATE', 'DEACTIVATE', 'BUY_RI']

relatedEntityTypes

- **type:** array
- **example:** Selecting a VCD, and specifying '"related_entities": ["VirtualMachine"] will include the actions for any VM in the given VDC
- **description:** Related Entity Types to include from the supply chain for this entity
- – **type:** string

targetEntityTypes

- **type:** array
- **example:** Specifying '"target_entities": ["VirtualMachine"] will include the actions where the target entity is a VM.
- **description:** Filter the actions where the target entity of the action matches the provided list of entity types
- – **type:** string

relatedCloudServiceProviderIds

- **type:** array
- **description:** Filter the actions by the related Cloud Service Providers.
- – **type:** string

groupBy

- **type:** array
- **description:** Fields used to Group the Stats
- – **type:** string

instanceTypeFilter

- **description:** [BETA]: Filter the actions based on a list of Instance Type Uuids.
- **reference:** InstanceTypeFilterApiDTO *(on page 1573)*

onDemandCostFilter

- **description:** [BETA]: Filter the actions based on a on-demand cost range.
- **reference:** OnDemandCostFilterApiDTO *(on page 1597)*

effectiveCostFilter

- **description:** Filter the actions based on a effective cost range.
- **reference:** EffectiveCostFilterApiDTO *(on page 1554)*

discountCoverageFilter

- **description:** [BETA]: Filter the actions based on a discount coverage.
- **reference:** DiscountCoverageFilterApiDTO *(on page 1553)*

accountFilter

- **description:** [BETA]: Filter the actions based on a list of account Uuid.
- **reference:** StringListFilterApiDTO *(on page 1686)*

resourceGroupFilter

- **description:** [BETA]: Filter the actions based on a list of resource group Uuid.
- **reference:** StringListFilterApiDTO *(on page 1686)*

tagValueFilter

- **description:** Filter actions based on a list of tag values.
- **reference:** StringListFilterApiDTO *(on page 1686)*

virtualDiskDisplayName

- **description:** This query will match the virtual disk displayName field of the ActionApiDTO to return actions that match the specified expression.
- **reference:** QueryInputApiDTO *(on page 1634)*

ActionDetailsApiDTO

Description

Parent class for action details. Implementing classes include: CloudResizeActionDetailsApiDTO, NoDetailsApiDTO, RIBuyActionDetailsApiDTO, CloudProvisionActionDetailsApiDTO, CloudSuspendActionDetailsApiDTO, OnPremResizeActionDetailsApiDTO and ReconfigureActionDetailsApiDTO.

Required Parameters

actionDetailType

- **type:** string

actionType

- **type:** string

ActionExecutionApiDTO

Description

Model to describe an Action Execution.

Required Parameters

acceptTime

- **type:** string
- **description:** Time when actions were accepted (in ISO8601 format - YYYY-MM-DDTHH:MM:SS)

actionIds

- **type:** array
- **description:** List of action IDs
- – **type:** string
- **maxItems:** 2147483647
- **minItems:** 1

id

- **type:** string
- **description:** Action Execution ID

skippedActions

- **type:** array
- **description:** List of actions that were skipped in execution
- – **reference:** SkippedActionApiDTO *(on page 1679)*

ActionExecutionAuditApiDTO

Description

Model to describe an Action Execution status.

Required Parameters

executionTime

- **type:** string
- **description:** Start of the execution time

executorInfo

- **description:** Information about the executor
- **reference:** ExecutorInfoApiDTO *(on page 1560)*

state

- **type:** string
- **description:** State of the execution
- **enum:** ['ACCEPTED', 'REJECTED', 'PRE_IN_PROGRESS', 'POST_IN_PROGRESS', 'IN_PROGRESS', 'SUCCEEDED', 'FAILED', 'DISABLED', 'QUEUED', 'CLEARED', 'ACCOUNTING', 'READY', 'FAILING', 'BEFORE_EXEC', 'IN_PROGRESS_NON_CRITICAL']

Optional Parameters

completionTime

- **type:** string
- **description:** End of the execution time

message

- **type:** string
- **description:** Status message, if the Action failed, it will contains the Error details

progress

- **type:** integer
- **format:** int32
- **description:** Action execution progress value, from 0 to 100

ActionExecutionCharacteristicApiDTO

Description

Model to describe action execution characteristics

Optional Parameters

disruptiveness

- **type:** string
- **description:** whether action is disruptive.
- **enum:** ['DISRUPTIVE', 'NON_DISRUPTIVE']

reversibility

- **type:** string
- **description:** whether action is reversible.
- **enum:** ['REVERSIBLE', 'IRREVERSIBLE']

ActionExecutionCharacteristicInputApiDTO

Description

Model to describe action execution characteristics

Optional Parameters

disruptiveness

- **type:** string
- **description:** whether action is disruptive.
- **enum:** ['DISRUPTIVE', 'NON_DISRUPTIVE']

reversibility

- **type:** string
- **description:** whether action is reversible.
- **enum:** ['REVERSIBLE', 'IRREVERSIBLE']

ActionExecutionInputApiDto

Description

Model to describe an Action Execution input.

Required Parameters

actionIds

- **type:** array
- **description:** List of action IDs
- – **type:** string
- **maxItems:** 2147483647
- **minItems:** 1

ActionResourceImpactStatApiInputDTO

Description

Model to describe an input for Action Resource Impact Stats

Required Parameters

actionResourceImpactStatList

- **type:** array
- **description:** List of Action Resource Impact Stat Inputs
- – **reference:** ActionResourceImpactStatInput *(on page 1497)*

Optional Parameters

actionStateList

- **type:** array
- **description:** List of States to filter.
- – **type:** string
- – **enum:** ['ACCEPTED', 'REJECTED', 'PRE_IN_PROGRESS', 'POST_IN_PROGRESS', 'IN_PROGRESS', 'SUCCEEDED', 'FAILED', 'DISABLED', 'QUEUED', 'CLEARED', 'ACCOUNTING', 'READY', 'FAILING', 'BEFORE_EXEC', 'IN_PROGRESS_NON_CRITICAL']

actionModeList

- **type:** array
- **description:** List of Modes to filter
- – **type:** string
  – **enum:** ['DISABLED', 'RECOMMEND', 'EXTERNAL_APPROVAL', 'MANUAL', 'AUTOMATIC', 'COLLECTION']

riskSeverityList

- **type:** array
- **description:** List of Risk Severities to filter
- – **type:** string

riskSubCategoryList

- **type:** array
- **description:** List of Risk Severities to filter
- – **type:** string

actionRelationTypeFilter

- **type:** array
- **description:** Filters the main list of actions based on the types of relationships they have to other actions.
- – **type:** string
  – **enum:** ['NONE', 'BLOCKED_BY', 'BLOCKING', 'CAUSED_BY', 'CAUSING']

environmentType

- **type:** string
- **description:** Filter the actions by Environment Type
- **enum:** ['ONPREM', 'CLOUD']

costType

- **type:** string
- **description:** Filter the actions by cost Type
- **enum:** ['SAVING', 'INVESTMENT']

descriptionQuery

- **description:** This query will match the details field of the ActionApiDTO to return actions that match the specified expression.
- **reference:** QueryInputApiDTO *(on page 1634)*

riskQuery

- **description:** This query will match the risk description field of the ActionApiDTO to return actions that match the specified expression.
- **reference:** QueryInputApiDTO *(on page 1634)*

executionCharacteristics

- **description:** Characteristics of the execution of the action, including disruptiveness and reversibility.
- **reference:** ActionExecutionCharacteristicInputApiDTO *(on page 1494)*

savingsAmountRange

- **description:** This field will match the costPrice statistic field of the ActionApiDTO to return actions that have a costPrice amount between minValue and maxValue, inclusive
- **reference:** RangeInputApiDTO *(on page 1637)*

effectiveCostSavingsAmountRange

- **description:** This field will match the effectiveCost statistic field of the ActionApiDTO to return actions that have an effectiveCost amount between minValue and maxValue, inclusive
- **reference:** RangeInputApiDTO *(on page 1637)*

unattachedDays

- **description:** This field will match the unattachedDays resource impacts field of the ActionResourceImpactStatApiInputDTO to return actions that have a unattachedDays count between minValue and maxValue, inclusive
- **reference:** RangeInputApiDTO *(on page 1637)*

hasSchedule

- **type:** boolean
- **description:** If true, only actions that have an associated execution schedule will be returned.

hasPrerequisites

- **type:** boolean
- **description:** If true, only actions that have a prerequisite preventing execution of the action will be returned.

ActionResourceImpactStatInput

Description

Model to describe an Action Resource Impact Stat Input

Required Parameters

actionType

- **type:** string
- **description:** Action Type
- **enum:** ['NONE', 'START', 'MOVE', 'SCALE', 'ALLOCATE', 'SUSPEND', 'TERMINATE', 'SPAWN', 'ADD_PROVIDER', 'CHANGE', 'REMOVE_PROVIDER', 'PROVISION', 'RECONFIGURE', 'RESIZE', 'RESIZE_CAPACITY', 'WARN', 'RECONFIGURE_THRESHOLD', 'DELETE', 'RIGHT_SIZE', 'RESERVE_ON_PM', 'RESERVE_ON_DS', 'RESIZE_FOR_EFFICIENCY', 'RESIZE_FOR_PERFORMANCE', 'CROSS_TARGET_MOVE', 'MOVE_TOGETHER', 'ACTIVATE', 'DEACTIVATE', 'BUY_RI']

commodityType

- **type:** string
- **example:** VCPU
- **description:** Commodity Type that could be impacted by the action. For example, VCPU or VMEM for a Resize VM action.
- **enum:** ['ACTION_PERMIT', 'ACTIVE_SESSIONS', 'APPLICATION', 'BALLOONING', 'BUFFER_COMMODITY', 'CLUSTER', 'COLLECTION_TIME', 'COOLING', 'CONNECTION', 'COUPON', 'CPU', 'CPU_ALLOCATION', 'CPU_PROVISIONED', 'CPU_REQUEST_ALLOCATION', 'CROSS_CLOUD_MOVE_SVC', 'CROSS_CLUSTER_MOVE_SVC', 'DATACENTER', 'DATASTORE', 'DB_CACHE_HIT_RATE', 'DB_MEM', 'DISK_ARRAY_ACCESS', 'DRS_SEGMENTATION', 'DSPM_ACCESS', 'EXTENT', 'FLOW', 'FLOW_ALLOCATION', 'HEAP', 'HOST_LUN_ACCESS', 'IMAGE_CPU', 'IMAGE_MEM', 'IMAGE_STORAGE', 'INSTANCE_DISK_SIZE', 'INSTANCE_DISK_TYPE', 'IO_THROUGHPUT', 'LICENSE_ACCESS', 'MEM', 'MEM_ALLOCATION', 'MEM_REQUEST_ALLOCATION', 'MEM_PROVISIONED', 'NET_THROUGHPUT', 'NETWORK', 'NUM_DISK', 'NUMBER_CONSUMERS', 'NUM_VCORE', 'POOL_CPU', 'POOL_MEM', 'POOL_STORAGE', 'PORT_CHANNEL', 'POWER', 'Q16_VCPU', 'Q1_VCPU', 'Q2_VCPU', 'Q32_VCPU', 'Q4_VCPU', 'Q64_VCPU', 'Q8_VCPU', 'RESPONSE_TIME', 'SAME_CLUSTER_MOVE_SVC', 'SEGMENTATION', 'SLA_COMMODITY', 'SOFTWARE_LICENSE_COMMODITY', 'SPACE', 'STORAGE_ACCESS', 'STORAGE_ALLOCATION', 'STORAGE_AMOUNT', 'STORAGE_CLUSTER', 'STORAGE_LATENCY', 'STORAGE_PROVISIONED', 'SWAPPING', 'THREADS', 'TRANSACTION', 'TRANSACTION_LOG', 'VCPU', 'VCPU_LIMIT_QUOTA', 'VCPU_REQUEST', 'VCPU_REQUEST_QUOTA', 'VDC', 'VMEM', 'VMEM_LIMIT_QUOTA', 'VMEM_REQUEST', 'VMEM_REQUEST_QUOTA', 'VMPM_ACCESS', 'VSTORAGE', 'ZONE', 'LICENSE_COMMODITY', 'Q3_VCPU', 'NUMBER_CONSUMERS_PM', 'Q6_VCPU', 'Q7_VCPU', 'QN_VCPU', 'RIGHT_SIZE_SVC', 'RIGHT_SIZE_DOWN', 'MOVE', 'Q5_VCPU', 'STORAGE', 'NUMBER_CONSUMERS_STORAGE', 'ACCESS', 'RIGHT_SIZE_UP', 'VAPP_ACCESS', 'HOT_STORAGE', 'HA_COMMODITY', 'NETWORK_POLICY', 'SERVICE_LEVEL_CLUSTER', 'PROCESSING_UNITS', 'TENANCY_ACCESS', 'TEMPLATE_ACCESS', 'BURST_BALANCE', 'DESIRED_COUPON', 'NETWORK_INTERFACE_COUNT', 'BICLIQUE', 'TAINT', 'LABEL', 'PEER_TO_PEER_AFFINITY', 'PEER_TO_PEER_ANTI_AFFINITY', 'ENERGY', 'CARBON', 'VIRTUAL_PROCESSORS', 'CPU_READY', 'PROCESSING_UNITS_PROVISIONED', 'HARDWARE_VERSION', 'GPU_MEM_ALLOCATION', 'GPU_MEM', 'GPU_COUNT', 'GPU_TENSOR_ENGINE', 'GPU_FP64_ENGINE', 'GPU_FP32_ENGINE', 'GPU_FP16_ENGINE', 'GPU_MEMORY_BANDWIDTH', 'CORE_SPEED', 'VGPU_MEM', 'CONCURRENT_QUERIES', 'VGPU_COMPUTE', 'GPU', 'GPU_REQUEST', 'GPU_REQUEST_QUOTA', 'QUEUING_TIME', 'SERVICE_TIME', 'RAID_HOSTS', 'UNKNOWN']

targetEntityType

- **type:** string
- **description:** Target entity type for an action. For example, VirtualMachine for a resize virtual machine action
- **enum:** ['ApplicationComponent', 'VirtualMachine', 'Storage', 'DiskArray', 'Database', 'DatabaseServer', 'Container', 'LogicalPool']

ActionScheduleApiDTO

Description

Model to describe an Action Schedule.

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

nextOccurrence

- **type:** string
- **description:** The time of the next occurrence represented in ISO8601 local time (YYYY-MM-DDTHH:MM:SS). For example, 2019-04-04T19:15.

nextOccurrenceTimestamp

- **type:** integer
- **format:** int64
- **description:** The time of the next occurence of this schedule represented in Timestamp in milliseconds. For example, 1562100619000

timeZoneId

- **type:** string
- **description:** The timezone of the schedule

mode

- **type:** string
- **description:** The action mode affected by the schedule
- **enum:** ['DISABLED', 'RECOMMEND', 'EXTERNAL_APPROVAL', 'MANUAL', 'AUTOMATIC', 'COLLECTION']

acceptedByUserForMaintenanceWindow

- **type:** boolean
- **description:** Boolean signifying acceptance or rejection of the action.

userName

- **type:** string
- **description:** The user who accepted the action for the maintenance window.

remaingTimeActiveInMs

- **type:** integer
- **format:** int64
- **description:** The amount of time that the schedule is going to be active at the time of call in milliseconds. This field will not be populated if the schedule is not currently active.

ActionScopesApiInputDTO

Description

Model to describe the request for Actions in multiple Scopes

Required Parameters

actionInput

- **description:** List of requested actions
- **reference:** ActionApiInputDTO *(on page 1489)*

Optional Parameters

scopes

- **type:** array
- **description:** List of uuid to use as scope
- – **type:** string

relatedType

- **type:** string
- **description:** Get the actions for the entity type related to the scopes

ActionScopesResourceImpactStatApiInputDTO

Description

Model to describe the request for Actions Resource Impacts in multiple Scopes

Required Parameters

actionResourceImpactStatApiInput

- **description:** Requested action resource impact stats
- **reference:** ActionResourceImpactStatApiInputDTO *(on page 1495)*

scopes

- **type:** array
- **description:** List of uuid to use as scope
- – **type:** string

ActionStateChangeApiDTO

Description

Model to describe an Action State Change.

Required Parameters

state

- **type:** string
- **description:** Action state
- **enum:** ['ACCEPTED', 'REJECTED', 'PRE_IN_PROGRESS', 'POST_IN_PROGRESS', 'IN_PROGRESS', 'SUCCEEDED', 'FAILED', 'DISABLED', 'QUEUED', 'CLEARED', 'ACCOUNTING', 'READY', 'FAILING', 'BEFORE_EXEC', 'IN_PROGRESS_NON_CRITICAL']

stateChangeTime

- **type:** string
- **description:** Action state change time

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

ActionWorkflowsApiDTO

Description

Model that describes details associated with a action and workflow pair.

Required Parameters

actionId

- **type:** string
- **description:** ActionId associated with the workflow

actionState

- **type:** string
- **description:** Action state
- **enum:** ['ACCEPTED', 'REJECTED', 'PRE_IN_PROGRESS', 'POST_IN_PROGRESS', 'IN_PROGRESS', 'SUCCEEDED', 'FAILED', 'DISABLED', 'QUEUED', 'CLEARED', 'ACCOUNTING', 'READY', 'FAILING', 'BEFORE_EXEC', 'IN_PROGRESS_NON_CRITICAL']

actionTransitionTime

- **type:** string
- **description:** Action state change time

workflowId

- **type:** string
- **description:** WorkflowId associated with the action

workflowState

- **type:** string
- **description:** Workflow state
- **enum:** ['WORKFLOW_RUNNING', 'WORKFLOW_FAILED', 'WORKFLOW_SUCCEEDED']

workflowTransitionTime

- **type:** string
- **description:** Workflow state change time

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

ActiveDirectoryApiDTO

Description

Model to describe an Active Directory. It contains domain name and server name which are global to Turbonomic server.

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

domainName

- **type:** string
- **description:** The domain name of current active directory

loginProviderURI

- **type:** string
- **description:** The server name of current active directory

isSecure

- **type:** boolean
- **description:** The server name of current active directory

groups

- **type:** array
- **description:** Active directory groups defined in Ops Mgr for current active directory
- – **reference:** ActiveDirectoryGroupApiDTO *(on page 1502)*

ActiveDirectoryGroupApiDTO

Description

Model to describe an Active Dierctory Group. Active Directory groups are used to manage user authentication. You can specify role and scope for the group, and any member of the AD group can log into Turbonomic with those privileges.

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

type

- **type:** string
- **description:** The user type for current active directory group

roleName

- **type:** string
- **description:** The user role name of current active directory group

scope

- **type:** array
- **description:** The scope of current active directory group
- – **reference:** GroupApiDTO *(on page 1564)*

AddObjectApiDTO

Description

Model to describe an add change in the scenario

Optional Parameters

projectionDays

- **type:** array
- **description:** List of projection days, indicate when to apply the changes
- – **type:** integer
  – **format:** int32

target

- **description:** Target of the change, entity or group
- **reference:** BaseApiDTO *(on page 1509)*

count

- **type:** integer
- **format:** int32
- **description:** How many times to apply the add change

targetEntityType

- **type:** string
- **description:** Type of entity to add from group target

AggregateCommodityFieldApiDTO

Description

Inherits From

This DTO inherits from FieldApiDTO *(on page 1562)*

Required Parameters

aggregation

- **type:** string
- **description:** The type of aggregation.
- **readOnly:** True
- **enum:** ['TOTAL', 'AVERAGE']

commodityAttribute

- **type:** string
- **description:** The commodity attribute.
- **readOnly:** True
- **enum:** ['USED', 'CAPACITY', 'PEAK', 'CURRENT_UTILIZATION', 'WEIGHTED_HISTORICAL_UTILIZATION', 'PERCENTILE_HISTORICAL_UTILIZATION']

commodityType

- **type:** string
- **description:** The commodity name.
- **readOnly:** True
- **enum:** ['ACTION_PERMIT', 'ACTIVE_SESSIONS', 'APPLICATION', 'BALLOONING', 'BUFFER_COMMODITY', 'CLUSTER', 'COLLECTION_TIME', 'COOLING', 'CONNECTION', 'COUPON', 'CPU', 'CPU_ALLOCATION', 'CPU_PROVISIONED', 'CPU_REQUEST_ALLOCATION', 'CROSS_CLOUD_MOVE_SVC', 'CROSS_CLUSTER_MOVE_SVC', 'DATACENTER', 'DATASTORE', 'DB_CACHE_HIT_RATE', 'DB_MEM', 'DISK_ARRAY_ACCESS', 'DRS_SEGMENTATION', 'DSPM_ACCESS', 'EXTENT', 'FLOW', 'FLOW_ALLOCATION', 'HEAP', 'HOST_LUN_ACCESS', 'IMAGE_CPU', 'IMAGE_MEM', 'IMAGE_STORAGE', 'INSTANCE_DISK_SIZE', 'INSTANCE_DISK_TYPE', 'IO_THROUGHPUT', 'LICENSE_ACCESS', 'MEM', 'MEM_ALLOCATION', 'MEM_REQUEST_ALLOCATION', 'MEM_PROVISIONED', 'NET_THROUGHPUT', 'NETWORK', 'NUM_DISK', 'NUMBER_CONSUMERS', 'NUM_VCORE', 'POOL_CPU', 'POOL_MEM', 'POOL_STORAGE', 'PORT_CHANNEL', 'POWER', 'Q16_VCPU', 'Q1_VCPU', 'Q2_VCPU', 'Q32_VCPU', 'Q4_VCPU', 'Q64_VCPU', 'Q8_VCPU', 'RESPONSE_TIME', 'SAME_CLUSTER_MOVE_SVC', 'SEGMENTATION', 'SLA_COMMODITY', 'SOFTWARE_LICENSE_COMMODITY', 'SPACE', 'STORAGE_ACCESS', 'STORAGE_ALLOCATION', 'STORAGE_AMOUNT', 'STORAGE_CLUSTER', 'STORAGE_LATENCY', 'STORAGE_PROVISIONED', 'SWAPPING', 'THREADS', 'TRANSACTION', 'TRANSACTION_LOG', 'VCPU', 'VCPU_LIMIT_QUOTA', 'VCPU_REQUEST', 'VCPU_REQUEST_QUOTA', 'VDC', 'VMEM', 'VMEM_LIMIT_QUOTA', 'VMEM_REQUEST', 'VMEM_REQUEST_QUOTA', 'VMPM_ACCESS', 'VSTORAGE', 'ZONE', 'LICENSE_COMMODITY', 'Q3_VCPU', 'NUMBER_CONSUMERS_PM', 'Q6_VCPU', 'Q7_VCPU', 'QN_VCPU', 'RIGHT_SIZE_SVC', 'RIGHT_SIZE_DOWN', 'MOVE', 'Q5_VCPU', 'STORAGE', 'NUMBER_CONSUMERS_STORAGE', 'ACCESS', 'RIGHT_SIZE_UP', 'VAPP_ACCESS', 'HOT_STORAGE', 'HA_COMMODITY', 'NETWORK_POLICY', 'SERVICE_LEVEL_CLUSTER', 'PROCESSING_UNITS', 'TENANCY_ACCESS', 'TEMPLATE_ACCESS', 'BURST_BALANCE', 'DESIRED_COUPON', 'NETWORK_INTERFACE_COUNT', 'BICLIQUE', 'TAINT', 'LABEL', 'PEER_TO_PEER_AFFINITY', 'PEER_TO_PEER_ANTI_AFFINITY', 'ENERGY', 'CARBON', 'VIRTUAL_PROCESSORS', 'CPU_READY', 'PROCESSING_UNITS_PROVISIONED', 'HARDWARE_VERSION', 'GPU_MEM_ALLOCATION', 'GPU_MEM', 'GPU_COUNT', 'GPU_TENSOR_ENGINE', 'GPU_FP64_ENGINE', 'GPU_FP32_ENGINE', 'GPU_FP16_ENGINE', 'GPU_MEMORY_BANDWIDTH', 'CORE_SPEED', 'VGPU_MEM', 'CONCURRENT_QUERIES', 'VGPU_COMPUTE', 'GPU', 'GPU_REQUEST', 'GPU_REQUEST_QUOTA', 'QUEUING_TIME', 'SERVICE_TIME', 'RAID_HOSTS', 'UNKNOWN']

AggregatedHealthResponseDTO

Description

A DTO for a single aggregated health response item.

Required Parameters

healthState

- **type:** string
- **description:** The (aggregated) health state for the subcategory.
- **enum:** ['CRITICAL', 'MAJOR', 'MINOR', 'NORMAL']

numberOfItems

- **type:** integer
- **format:** int32
- **description:** Number of entities whose state matches healthState.

subcategory

- **type:** string
- **description:** Health subcategory.

Optional Parameters

recommendations

- **type:** array
- **description:** A list of recommendations specific for each problem (error) type.
- – **reference:** Recommendation *(on page 1637)*

AppliedScopeApiDTOObject

Description

Model to describe the scnope of resources to which a cloud commitment can be applied

Optional Parameters

appliedScopeType

- **type:** object
- **description:** The group or entity type to which the cloud commitment is applicable

appliedScopeIds

- **type:** array
- **description:** The IDs of the group or entity to which the cloud commitment is applicable
- – **type:** string

AppliedScopeApiDTO

Description

Model to describe the scnope of resources to which a cloud commitment can be applied

Optional Parameters

appliedScopeType

- **type:** object
- **description:** The group or entity type to which the cloud commitment is applicable

appliedScopeIds

- **type:** array
- **description:** The IDs of the group or entity to which the cloud commitment is applicable
- – **type:** string

AutomatedEntityDefinitionData

Description

Inherits From

This DTO inherits from IEntityDefinitionData *(on page 1570)*

Required Parameters

entityType

- **type:** string
- **description:** Connected entity type
- **enum:** ['ApplicationComponent', 'BusinessTransaction', 'Container', 'ContainerPod', 'DatabaseServer', 'Service', 'VirtualMachine']

membersCount

- **type:** integer
- **format:** int32
- **description:** Dynamic tag data Members count

namePrefix

- **type:** string
- **description:** Naming prefix for automated entity definition

Optional Parameters

generationAndConnectionAttribute

- **description:** Defined criteria of connected entities
- **reference:** FilterApiDTO *(on page 1563)*

AzureVirtualMachineScaleSetDTO

Description

Model to describe the specific properties of Azure Virtual Machine Scale Set, using which a group of load balanced VMs can be created and managed

Required Parameters

imageReference

- **description:** Image reference properties of Azure Virtual Machine Scale Set
- **reference:** AzureVirtualMachineScaleSetImageReferenceDTO *(on page 1508)*

provisioningState

- **type:** string
- **description:** The provisioning state is the status of a user-initiated, control-plane operation on Azure Virtual Machine Scale Set
- **enum:** ['CREATING', 'UPDATING', 'SUCCEEDED', 'FAILED', 'DELETING', 'MIGRATING']

scalingMode

- **type:** string
- **description:** Scaling mode enables auto scaling to add and remove instances either manually or by custom autoscaling
- **enum:** ['MANUAL', 'AUTOSCALE_METRIC_BASED', 'AUTOSCALE_SPECIFIC_INSTANCE_COUNT']

upgradePolicy

- **description:** Upgrade policy of Azure Virtual Machine Scale Set
- **reference:** AzureVirtualMachineScaleSetUpgradePolicyDTO *(on page 1509)*

Optional Parameters

autoScaleProfileCount

- **type:** integer
- **format:** int32
- **example:** [1]
- **description:** Number of autoscaled profiles in Azure Virtual Machine Scale Set

scaleInPolicy

- **type:** string
- **description:** The order in which the Azure VMs are selected for deletion during Scale-in operation
- **enum:** ['DEFAULT', 'NEWEST_VM', 'OLDEST_VM']

overProvisioningEnabled

- **type:** boolean

singlePlacementGroupEnabled

- **type:** boolean

AzureVirtualMachineScaleSetImageReferenceDTO

Description

Model to represent the property for source OS image of Azure Virtual Machine Scale Set

Required Parameters

publisher

- **type:** string
- **example:** [canonical]
- **description:** Publisher of Azure Virtual Machine Scale Set

sku

- **type:** string
- **example:** [20_04-lts-gen2]
- **description:** SKU of Azure Virtual Machine Scale Set

version

- **type:** string
- **example:** [latest]
- **description:** Version of Azure Virtual Machine Scale Set

Optional Parameters

offer

- **type:** string
- **example:** [0001-com-ubuntu-server-local]
- **description:** Offer of Azure Virtual Machine Scale Set

AzureVirtualMachineScaleSetUpgradePolicyDTO

Description

Model to represent the upgrade policy of Azure Virtual Machine Scale Set determines how virtual machines can be brought up-to-date with the latest scale set model

Required Parameters

upgradePolicyMode

- **type:** string
- **description:** Upgrade policy mode of Azure Virtual Machine Scale Set
- **enum:** ['AUTOMATIC', 'MANUAL', 'ROLLING']

Optional Parameters

autoOSUpgradeEnabled

- **type:** boolean

BaseApiDTO

Description

Basic information of an entity

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

BaseCommodityApiDTO

Description

Describes a commodity

Required Parameters

type

- **type:** string
- **description:** The type of the commodity
- **enum:** ['ACTION_PERMIT', 'ACTIVE_SESSIONS', 'APPLICATION', 'BALLOONING', 'BUFFER_COMMODITY', 'CLUSTER', 'COLLECTION_TIME', 'COOLING', 'CONNECTION', 'COUPON', 'CPU', 'CPU_ALLOCATION', 'CPU_PROVISIONED', 'CPU_REQUEST_ALLOCATION', 'CROSS_CLOUD_MOVE_SVC', 'CROSS_CLUSTER_MOVE_SVC', 'DATACENTER', 'DATASTORE', 'DB_CACHE_HIT_RATE', 'DB_MEM', 'DISK_ARRAY_ACCESS', 'DRS_SEGMENTATION', 'DSPM_ACCESS', 'EXTENT', 'FLOW', 'FLOW_ALLOCATION', 'HEAP', 'HOST_LUN_ACCESS', 'IMAGE_CPU', 'IMAGE_MEM', 'IMAGE_STORAGE', 'INSTANCE_DISK_SIZE', 'INSTANCE_DISK_TYPE', 'IO_THROUGHPUT', 'LICENSE_ACCESS', 'MEM', 'MEM_ALLOCATION', 'MEM_REQUEST_ALLOCATION', 'MEM_PROVISIONED', 'NET_THROUGHPUT', 'NETWORK', 'NUM_DISK', 'NUMBER_CONSUMERS', 'NUM_VCORE', 'POOL_CPU', 'POOL_MEM', 'POOL_STORAGE', 'PORT_CHANNEL', 'POWER', 'Q16_VCPU', 'Q1_VCPU', 'Q2_VCPU', 'Q32_VCPU', 'Q4_VCPU', 'Q64_VCPU', 'Q8_VCPU', 'RESPONSE_TIME', 'SAME_CLUSTER_MOVE_SVC', 'SEGMENTATION', 'SLA_COMMODITY', 'SOFTWARE_LICENSE_COMMODITY', 'SPACE', 'STORAGE_ACCESS', 'STORAGE_ALLOCATION', 'STORAGE_AMOUNT', 'STORAGE_CLUSTER', 'STORAGE_LATENCY', 'STORAGE_PROVISIONED', 'SWAPPING', 'THREADS', 'TRANSACTION', 'TRANSACTION_LOG', 'VCPU', 'VCPU_LIMIT_QUOTA', 'VCPU_REQUEST', 'VCPU_REQUEST_QUOTA', 'VDC', 'VMEM', 'VMEM_LIMIT_QUOTA', 'VMEM_REQUEST', 'VMEM_REQUEST_QUOTA', 'VMPM_ACCESS', 'VSTORAGE', 'ZONE', 'LICENSE_COMMODITY', 'Q3_VCPU', 'NUMBER_CONSUMERS_PM', 'Q6_VCPU', 'Q7_VCPU', 'QN_VCPU', 'RIGHT_SIZE_SVC', 'RIGHT_SIZE_DOWN', 'MOVE', 'Q5_VCPU', 'STORAGE', 'NUMBER_CONSUMERS_STORAGE', 'ACCESS', 'RIGHT_SIZE_UP', 'VAPP_ACCESS', 'HOT_STORAGE', 'HA_COMMODITY', 'NETWORK_POLICY', 'SERVICE_LEVEL_CLUSTER', 'PROCESSING_UNITS', 'TENANCY_ACCESS', 'TEMPLATE_ACCESS', 'BURST_BALANCE', 'DESIRED_COUPON', 'NETWORK_INTERFACE_COUNT', 'BICLIQUE', 'TAINT', 'LABEL', 'PEER_TO_PEER_AFFINITY', 'PEER_TO_PEER_ANTI_AFFINITY', 'ENERGY', 'CARBON', 'VIRTUAL_PROCESSORS', 'CPU_READY', 'PROCESSING_UNITS_PROVISIONED', 'HARDWARE_VERSION', 'GPU_MEM_ALLOCATION', 'GPU_MEM', 'GPU_COUNT', 'GPU_TENSOR_ENGINE', 'GPU_FP64_ENGINE', 'GPU_FP32_ENGINE', 'GPU_FP16_ENGINE', 'GPU_MEMORY_BANDWIDTH', 'CORE_SPEED', 'VGPU_MEM', 'CONCURRENT_QUERIES', 'VGPU_COMPUTE', 'GPU', 'GPU_REQUEST', 'GPU_REQUEST_QUOTA', 'QUEUING_TIME', 'SERVICE_TIME', 'RAID_HOSTS', 'UNKNOWN']

Optional Parameters

key

- **type:** string
- **description:** The specific commodity

units

- **type:** string
- **description:** The units in which the commodity is measured, if any

BaseGroupApiDTO

Description

Base model to describe a Group.

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

membersCount

- **type:** integer
- **format:** int32
- **description:** Number of members of the Group

BasicActionApiDTO

Description

Model to tersely describe an action, providing only a set of basic fields

Optional Parameters

actionType

- **type:** string
- **description:** Type
- **enum:** ['START', 'MOVE', 'SCALE', 'ALLOCATE', 'SUSPEND', 'PROVISION', 'RECONFIGURE', 'RESIZE', 'DELETE', 'RIGHT_SIZE', 'BUY_RI']

details

- **type:** string
- **description:** A user-readable string describing the action

target

- **description:** Target entity for an action. For example, the VM in a Resize Action, or the host for a VM move
- **reference:** DiscoveredEntityApiDTO *(on page 1554)*

currentEntity

- **description:** Current entity, such as the current host that a VM resides on for a VM move action
- **reference:** DiscoveredEntityApiDTO *(on page 1554)*

newEntity

- **description:** Destination entity, such as the host that a VM will move to for a VM move action
- **reference:** DiscoveredEntityApiDTO *(on page 1554)*

actionID

- **type:** integer
- **format:** int64
- **description:** Id of the action

BillingFamilyApiDTO

Description

Inherits From

This DTO inherits from GroupApiDTO *(on page 1564)*

Optional Parameters

masterAccountUuid

- **type:** string
- **description:** The UUID of the master account.

businessUnitApiDTOList

- **type:** array
- **description:** List of business units, for all accounts in this billing family.
- – **reference:** BusinessUnitApiDTO *(on page 1514)*

billingCostDataExists

- **type:** boolean
- **description:** Indicates if the billing family has associated cost data.

BooleanConditionApiDTO

Description

Inherits From

This DTO inherits from ConditionApiDTO *(on page 1535)*

Required Parameters

value

- **type:** boolean
- **description:** The value of the field must be equal to this boolean value
- **readOnly:** True

BooleanFieldValueApiDTO

Description

Inherits From

This DTO inherits from FieldValueApiDTO *(on page 1562)*

Required Parameters

value

- **type:** boolean
- **description:** The value.
- **readOnly:** True

BulkActionRequestApiDTO

Description

BulkActionRequest

Optional Parameters

state

- **type:** string
- **description:** state
- **enum:** ['RUNNING', 'STOPPED', 'SUSPENDED', 'MAINTENANCE', 'FAILOVER', 'UNKNOWN', 'STARTING', 'STOPPING']

error

- **type:** string
- **description:** error

entityUUID

- **type:** string

BulkActionRequestInputDTO

Description

BulkActionInput

Optional Parameters

action

- **type:** string
- **description:** action
- **enum:** ['START', 'STOP']

entityUuids

- **type:** array
- **description:** entityUuids
- – **type:** string

BusinessUnitApiDTO

Description

Model to describe the property of a business unit.

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

environmentType

- **type:** string
- **description:** Environment type
- **enum:** ['CLOUD', 'ONPREM', 'HYBRID']

discount

- **type:** number
- **format:** float
- **description:** Entity discount percentage

priceAdjustment

- **description:** Entity price adjustment percentage
- **reference:** PriceAdjustmentDTO *(on page 1629)*

accountId

- **type:** string
- **description:** Account ID

master

- **type:** boolean
- **description:** Flag which shows whether this account is master or not

associatedTargetId

- **type:** integer
- **format:** int64
- **description:** Target associated with this account, if workloads related to this account are being discovered. If no workloads are discovered for this account, the associatedTargetId will be empty.

budget

- **description:** Budget of the business unit, user only need to set value and units(monthly)
- **reference:** StatApiDTO *(on page 1679)*

businessUnitType

- **type:** string
- **description:** Type of the business unit instance
- **enum:** ['BUDGET', 'DISCOUNT', 'PRICE_ADJUSTMENT', 'DISCOVERED']

cloudType

- **type:** string
- **description:** Type of the cloud target which is related to current business account.
- **enum:** ['AWS', 'AZURE', 'GCP', 'UNKNOWN']

accountType

- **type:** string
- **description:** The category of the current cloud account.
- **enum:** ['STANDARD', 'GOVERNMENT_US']

riSupported

- **type:** boolean
- **description:** Flag which shows whether this account is RI supported

childrenBusinessUnits

- **type:** array
- **description:** UUIDs of the discovered children business units that will share the budget or discounts configuration of this business unit.
- – **type:** string

targets

- **type:** array
- **description:** Targets associated with the business unit
- – **reference:** TargetApiDTO *(on page 1690)*

costPrice

- **type:** number
- **format:** float
- **description:** Cost price per Hour

severity

- **type:** string
- **description:** Calculated using the highest severity of the member entities
- **enum:** ['UNKNOWN', 'NORMAL', 'MINOR', 'MAJOR', 'CRITICAL']

membersCount

- **type:** integer
- **format:** int32
- **description:** Number of members of the business unit

memberType

- **type:** string
- **description:** The type of service entities comprising the business unit

resourceGroupsCount

- **type:** integer
- **format:** int32
- **description:** The number of Resource Groups belonging to the business unit

displayable

- **type:** boolean
- **description:** Whether business unit is displayable or not

pricingIdentifiers

- **type:** object
- **description:** Pricing idenifiers for the business unit
- **additionalProperties:** {'type': 'string'}

BusinessUnitApiInputDTO

Description

Model to describe the parameters for creating a business unit.

Required Parameters

name

- **type:** string
- **description:** Name of the business unit

Optional Parameters

budget

- **description:** Budget of the business unit, user only need to set value
- **reference:** StatApiDTO *(on page 1679)*

childrenBusinessUnits

- **type:** array
- **description:** UUIDs of the discovered children business units, that will share the budget or discounts configuration of this business unit.
  - **type:** string

discount

- **type:** number
- **format:** float
- **description:** Business unit enterprise-level discount percentage. It will come to effect if cloud service or template-level discounts are not configured.

priceAdjustment

- **description:** Business unit enterprise-level price adjustment percentage. It will come to effect if cloud service or template-level discounts are not configured.
- **reference:** PriceAdjustmentDTO *(on page 1629)*

targets

- **type:** array
  - **type:** string

nameWithBusinessAccountType

- **type:** string

BusinessUnitPriceAdjustmentApiDTO

Description

Price adjustment values for cloud services and templates related to a business unit. The price adjustment configured on an user-defined business unit apply tocloud services and templates offered by the children business units.

Optional Parameters

servicePriceAdjustments

- **type:** array
- **description:** Price adjustment information for the related cloud services and templates.
- – **reference:** CloudServicePriceAdjustmentApiDTO *(on page 1530)*

serviceDiscounts

- **type:** array
- **description:** Discount information for the related cloud services and templates
- – **reference:** CloudServicePriceAdjustmentApiDTO *(on page 1530)*

BusinessUserEntityAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

sessions

- **type:** array
- **description:** Information about sessions to entities (VM or other) established by this BusinessUser.
- – **reference:** BusinessUserSessionApiDTO *(on page 1517)*

BusinessUserSessionApiDTO

Description

Session information for a connection established by BusinessUser to an entity (VM or other)

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

businessUserUuid

- **type:** string
- **description:** UUID of the connected BusinessUser

connectedEntityUuid

- **type:** string
- **description:** UUID of the entity that user has a connection to

duration

- **type:** integer
- **format:** int64
- **description:** Time elapsed since the session start in seconds- sessions are always timed

ChangePasswordApiDTO

Description

Optional Parameters

oldPassword

- **type:** string

newPassword

- **type:** string

ClientConfigurationApiDTO

Description

Client configuration information

Optional Parameters

clientFirewallAllowlist

- **type:** array
- **description:** Endpoints that client sites need to be able to access
- – **reference:** ClientConfigurationEndpoint *(on page 1518)*

ClientConfigurationEndpoint

Description

Optional Parameters

name

- **type:** string
- **description:** The endpoint name

type

- **type:** string
- **description:** The endpoint type (LoadBalancer or Route)

host

- **type:** string
- **description:** The host string of the endpoint. May be null

ip

- **type:** string
- **description:** The IP address of the endpoint

port

- **type:** integer
- **format:** int32
- **description:** The port of the endpoint

ClientInputDTO

Description

Required Parameters

name

- **type:** string
- **description:** Name of client

supportedServices

- **type:** array
- **description:** List of supported services
- – **type:** string

ClientNetworkRawTokenApiDTO

Description

Inherits From

This DTO inherits from ClientNetworkTokenApiOutputDTO *(on page 1521)*

Optional Parameters

apiVersion

- **type:** string

kind

- **type:** string

metadata

- **reference:** ClientNetworkRawTokenMetadata *(on page 1520)*

data

- **type:** object
- **additionalProperties:** {'type': 'string'}

ClientNetworkRawTokenMetadata

Description

Optional Parameters

name

- **type:** string

annotations

- **type:** object
- **additionalProperties:** {'type': 'string'}

labels

- **type:** object
- **additionalProperties:** {'type': 'string'}

ClientNetworkSitesMetadataApiDTO

Description

Optional Parameters

id

- **type:** string
- **description:** id

name

- **type:** string
- **description:** name

ClientNetworkTokenApiDTO

Description

Inherits From

This DTO inherits from ClientNetworkTokenApiOutputDTO *(on page 1521)*

Optional Parameters

id

- **type:** string
- **description:** id

tokenData

- **type:** string
- **description:** Token data

ClientNetworkTokenApiOutputDTO

Description

One of {ClientNetworkTokenApiDTO, ClientNetworkRawTokenApiDTO}

This is an unspecified object.

ClientNetworkTokenInputDTO

Description

Optional Parameters

claimLimit

- **type:** integer
- **format:** int32
- **example:** 1
- **description:** Claim Limit: Number of times the token can be reused
- **minimum:** 1
- **maximum:** 100

expiration

- **type:** integer
- **format:** int32
- **example:** 15
- **description:** Expiration time (in minutes) to claim the token
- **minimum:** 1
- **maximum:** 10080

ClientNetworkTokensMetadataApiDTO

Description

Optional Parameters

id

- **type:** string
- **description:** id

created

- **type:** string
- **description:** Created

claimsMade

- **type:** integer
- **format:** int32
- **description:** Claims Made

claimsRemaining

- **type:** integer
- **format:** int32
- **description:** Claims Remaining

claimExpiration

- **type:** string
- **description:** Claim Expiration

ClientServiceApiDTO

Description

Optional Parameters

id

- **type:** string
- **description:** id

supportedServices

- **type:** array
- **description:** Supported services
  - **type:** string

secret

- **type:** string
- **description:** Secret

name

- **type:** string
- **description:** Name of client

CloudAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

zone

- **description:** The zone of the cloud entity
- **reference:** BaseApiDTO *(on page 1509)*

region

- **description:** The region of the cloud entity
- **reference:** BaseApiDTO *(on page 1509)*

businessAccount

- **description:** The BusinessAccount of the cloud entity
- **reference:** BaseApiDTO *(on page 1509)*

resourceGroup

- **description:** The ResourceGroup of the cloud entity
- **reference:** BaseApiDTO *(on page 1509)*

template

- **description:** The template used to generate the entity
- **reference:** BaseApiDTO *(on page 1509)*

billingType

- **type:** string
- **description:** The Cloud Billing type

tier

- **type:** string
- **description:** The Tier of the storage

riCoveragePercentage

- **type:** number
- **format:** float
- **description:** The RI coverage percentage

riCoverage

- **description:** Used value stores VM coupons in use, capacity stores coupon capacity
- **reference:** StatApiDTO *(on page 1679)*

onDemandCost

- **type:** number
- **format:** float
- **description:** On-demand cost paid for workload per Hour (includes Compute and License on-demand costs)

enhancedNetworking

- **type:** string
- **description:** The type of Enhanced Networking supported. If the entity is an AWS VM, this field will be populated by a value describing what kind of enhanced networking, if any, the VM is currently using. If the entity is not an AWS VM, this field is not relevant and will not be populated or displayed.
- **enum:** ['NONE', 'ENA']

enaActive

- **type:** string
- **description:** Whether ENA is active. If the entity is an AWS VM, this field will be populated by a value describing whether ENA is currently in use or not. If the entity is not an AWS VM, this field is not relevant and will not be populated or displayed.
- **enum:** ['Active', 'Inactive']

virtualizationType

- **type:** string
- **description:** The virtualization type. If the entity is an AWS VM, this field will be populated by a value describing what kind of virtualization type, if any, the VM is currently using. If the entity is not an AWS VM, this field is not relevant and will not be populated or displayed.
- **enum:** ['PVM', 'HVM']

architecture

- **type:** string
- **description:** If the entity is an AWS VM, this field will be populated by a value describing what is the architecture of the VM.If the entity is not an AWS VM, this field is not relevant and will not be populated or displayed.
- **enum:** ['32-bit', '64-bit']

nvme

- **type:** string
- **description:** NVMe driver support. If the entity is an AWS VM, this field will be true if NVMe driver installed. If the entity is not an AWS VM, this field is not relevant and will not be populated or displayed.
- **enum:** ['Installed', 'Not installed']

entityUptime

- **description:** The entity uptime
- **reference:** EntityUptimeApiDTO *(on page 1559)*

tenancy

- **type:** string
- **description:** The tenancy type for VMs
- **enum:** ['DEFAULT', 'DEDICATED', 'HOST']

pathSegments

- **type:** array
- **description:** Optional path segments that can represent hierarchical context for this entity such as the names of parents containing this entity.
- – **type:** string

cloudServiceName

- **type:** string
- **description:** Optional parameter that represents the name of the cloud service
- **enum:** ['AMAZON_RDS', 'AMAZON_EC2', 'AMAZON_EC2_AUTO_SCALING', 'AZURE_APP_SERVICE', 'AZURE_COSMOS_DB', 'AZURE_DISK_STORAGE', 'AZURE_SQL_DATABASE', 'AZURE_SQL_MANAGED_INSTANCE', 'AZURE_SYNAPSE_SQL_POOL', 'AZURE_VIRTUAL_MACHINES', 'AZURE_VIRTUAL_MACHINE_SCALE_SETS', 'GCP_DISK_STORAGE', 'GCP_VIRTUAL_MACHINES']

CloudCommitmentAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

startTimeInMilliseconds

- **type:** integer
- **format:** int64
- **description:** The start time of the Cloud Commitment in UTC.

expirationTimeInMilliseconds

- **type:** integer
- **format:** int64
- **description:** The expiration time of the Cloud Commitment in UTC.

termInMilliseconds

- **type:** integer
- **format:** int64
- **description:** The term of the Cloud Commitment.

cloudCommitmentScopeType

- **type:** string
- **description:** The scope of the Cloud Commitment.
- **enum:** ['FamilyScoped', 'FamilyAndOSScoped', 'CloudServiceScoped']

cloudCommitmentScopeDTO

- **description:** The info about the scope of the cloud commitment.
- **reference:** CloudCommitmentScopeDTO *(on page 1527)*

cloudCommitmentScopeInfo

- **description:** The commitment scope details.
- **reference:** CloudCommitmentScopeInfoApiDTO *(on page 1527)*

cloudCommitmentPricing

- **description:** The commitment pricing information
- **reference:** CloudCommitmentPricingApiDTO *(on page 1527)*

cloudCommitmentCapacityApiDTO

- **description:** The capacity specific info of the cloud commitment.
- **reference:** CloudCommitmentCapacityApiDTO *(on page 1526)*

payment

- **type:** string
- **description:** Payment option of the cloud commitment.
- **enum:** ['ALL_UPFRONT', 'PARTIAL_UPFRONT', 'NO_UPFRONT']

providerSpecificType

- **type:** string
- **description:** The provider specific type.
- **enum:** ['SAVINGS_PLAN', 'COMMITTED_USE', 'RESERVED_INSTANCE']

commitmentStatus

- **type:** string
- **description:** The cloud commitment status.
- **enum:** ['Unknown', 'Pending', 'Active', 'Expired', 'Canceled']

instanceCount

- **type:** integer
- **format:** int32
- **description:** Number of commitment instances with the same type, template, platform, etc.

term

- **description:** Term, unit is Year
- **reference:** StatApiDTO *(on page 1679)*

CloudCommitmentCapacityApiDTO

Description

Model to describe cloud commitment capacity

Optional Parameters

instanceCapacity

- **type:** integer
- **format:** int64
- **description:** The instance capacity of the cloud commitment.

spendCapacity

- **type:** number
- **format:** double
- **description:** The spend capacity of the cloud commitment.

commoditiesBought

- **type:** object
- **description:** The capacity for each type commodity bought by the commitment.
- **additionalProperties:** {'type': 'number', 'format': 'double'}

CloudCommitmentPricingApiDTO

Description

Model to describe cloud commitment pricing

Optional Parameters

reservedInstanceType

- **type:** string
- **description:** The commitment offering type, e.g. standard or convertible
- **enum:** ['STANDARD', 'CONVERTIBLE']

paymentOption

- **type:** string
- **description:** The commitment payment option
- **enum:** ['ALL_UPFRONT', 'PARTIAL_UPFRONT', 'NO_UPFRONT']

fixedCost

- **type:** number
- **format:** double
- **description:** The fixed cost of the commitment

recurringCost

- **type:** number
- **format:** double
- **description:** The recurring cost of the commitment

effectiveCost

- **description:** Effective hourly cost of the commitment
- **reference:** StatApiDTO *(on page 1679)*

CloudCommitmentScopeDTO

Description

Model to describe cloud committment scope.

Optional Parameters

cloudFamilyReferenceApiDTO

- **description:** The family scoped type specific info of the cloud commitment. If null, indicates service scoped cloud commitment
- **reference:** CloudFamilyReferenceApiDTO *(on page 1528)*

CloudCommitmentScopeInfoApiDTO

Description

A container of scope related objects that may apply to a cloud commitment

Optional Parameters

topologyScopes

- **type:** array
- **description:** A list of topology related scope objects that may apply to a cloud commitment
- – **reference:** TopologyScopeApiDTO *(on page 1706)*

appliedScope

- **description:** A scope representing the eligible resources to which the discount can be applied
- **reference:** AppliedScopeApiDTOObject *(on page 1506)*

CloudFamilyReferenceApiDTO

Description

Model describe a reference to a cloud family

Optional Parameters

family

- **type:** string
- **description:** The family of the Cloud Commitment.

entityType

- **type:** string
- **description:** The entity type the Cloud Commitment applies to.
- **enum:** ['Application', 'ApplicationComponent', 'ApplicationServer', 'AvailabilityZone', 'BusinessAccount', 'BusinessApplication', 'BusinessTransaction', 'BusinessUser', 'Chassis', 'CloudService', 'ComputeTier', 'Container', 'ContainerPod', 'ContainerSpec', 'Database', 'DatabaseServer', 'DatabaseServerTier', 'DatabaseTier', 'DataCenter', 'DesktopPool', 'DiskArray', 'DPod', 'HCIPhysicalMachine', 'HypervisorServer', 'Internet', 'IOModule', 'LoadBalancer', 'LogicalPool', 'Namespace', 'Network', 'PhysicalMachine', 'ProcessorPool', 'Region', 'ReservedInstance', 'Service', 'ServiceProvider', 'Storage', 'StorageController', 'StorageTier', 'Switch', 'Unknown', 'ViewPod', 'VirtualApplication', 'VirtualDataCenter', 'VirtualMachine', 'VirtualVolume', 'VPod', 'WorkloadController', 'ContainerPlatformCluster', 'CloudCommitment', 'ApplicationComponentSpec', 'VirtualMachineSpec', 'DocumentCollection', 'NodeGroup', 'PowerVMPhysicalMachine', 'DataWarehouse', 'DataWarehouseTier']

CloudProvisionActionDetailsApiDTO

Description

Inherits From

This DTO inherits from ActionDetailsApiDTO *(on page 1492)*

Optional Parameters

onDemandRate

- **type:** number
- **format:** float

onDemandCost

- **type:** number
- **format:** float

subType

- **type:** string

CloudResizeActionDetailsApiDTO

Description

Inherits From

This DTO inherits from ActionDetailsApiDTO *(on page 1492)*

Optional Parameters

onDemandRateBefore

- **type:** number
- **format:** float

onDemandRateAfter

- **type:** number
- **format:** float

onDemandCostBefore

- **type:** number
- **format:** float

onDemandCostAfter

- **type:** number
- **format:** float

riCoverageBefore

- **description:** RI coupons used and capacity of the target entity before action
- **reference:** StatApiDTO *(on page 1679)*

riCoverageAfter

- **description:** RI coupons used and capacity of the target entity after action
- **reference:** StatApiDTO *(on page 1679)*

Reserved instance rate before

- **type:** number
- **format:** float

Reserved instance rate after

- **type:** number
- **format:** float

commitmentRateBefore

- **type:** number
- **format:** float

commitmentRateAfter

- **type:** number
- **format:** float

effectiveCostBefore

- **type:** number
- **format:** float

effectiveCostAfter

- **type:** number
- **format:** float

commitmentCoverageBefore

- **reference:** StatApiDTO *(on page 1679)*

commitmentCoverageAfter

- **reference:** StatApiDTO *(on page 1679)*

totalDiscountCoverageBefore

- **reference:** StatApiDTO *(on page 1679)*

totalDiscountCoverageAfter

- **reference:** StatApiDTO *(on page 1679)*

entityUptime

- **description:** The entity uptime which was taken into account when generating the action
- **reference:** EntityUptimeApiDTO *(on page 1559)*

subType

- **type:** string

CloudServicePriceAdjustmentApiDTO

Description

Price adjustments for cloud service and related templates specified for a certain pricing model.

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

environmentType

- **type:** string
- **description:** Environment type
- **enum:** ['CLOUD', 'ONPREM', 'HYBRID']

discount

- **type:** number
- **format:** float
- **description:** Entity discount percentage

priceAdjustment

- **description:** Entity price adjustment percentage
- **reference:** PriceAdjustmentDTO *(on page 1629)*

pricingModel

- **type:** string
- **description:** Pricing model, 'On demand' for time-based, 'Reserved' for coupon-based, 'Spot' for customer bidding
- **enum:** ['ON_DEMAND', 'RESERVED', 'SPOT']

templateDiscounts

- **type:** array
- **description:** Discount values for the templates offered by the cloud service under certain pricing model.
- – **reference:** TemplatePriceAdjustmentDTO *(on page 1701)*

templatePriceAdjustments

- **type:** array
- **description:** Price adjustment values for the templates offered by the cloud service under certain pricing model.
- – **reference:** TemplatePriceAdjustmentDTO *(on page 1701)*

CloudServicePricingModelApiDTO

Description

Cloud service + pricing model pair that is available for a business unit.

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

environmentType

- **type:** string
- **description:** Environment type
- **enum:** ['CLOUD', 'ONPREM', 'HYBRID']

discount

- **type:** number
- **format:** float
- **description:** Entity discount percentage

priceAdjustment

- **description:** Entity price adjustment percentage
- **reference:** PriceAdjustmentDTO *(on page 1629)*

pricingModel

- **type:** string
- **description:** Pricing model, 'On demand' for time-based, 'Reserved' for coupon-based, 'Spot' for customer bidding
- **enum:** ['ON_DEMAND', 'RESERVED', 'SPOT']

CloudSuspendActionDetailsApiDTO

Description

Inherits From

This DTO inherits from ActionDetailsApiDTO *(on page 1492)*

Optional Parameters

onDemandRate

- **type:** number
- **format:** float

onDemandCost

- **type:** number
- **format:** float

subType

- **type:** string

CommodityDetailsApiDTO

Description

Information on relation type and its related commodities

Required Parameters

commodityType

- **type:** string
- **description:** Relation Type
- **enum:** ['ACTION_PERMIT', 'ACTIVE_SESSIONS', 'APPLICATION', 'BALLOONING', 'BUFFER_COMMODITY', 'CLUSTER', 'COLLECTION_TIME', 'COOLING', 'CONNECTION', 'COUPON', 'CPU', 'CPU_ALLOCATION', 'CPU_PROVISIONED', 'CPU_REQUEST_ALLOCATION', 'CROSS_CLOUD_MOVE_SVC', 'CROSS_CLUSTER_MOVE_SVC', 'DATACENTER', 'DATASTORE', 'DB_CACHE_HIT_RATE', 'DB_MEM', 'DISK_ARRAY_ACCESS', 'DRS_SEGMENTATION', 'DSPM_ACCESS', 'EXTENT', 'FLOW', 'FLOW_ALLOCATION', 'HEAP', 'HOST_LUN_ACCESS', 'IMAGE_CPU', 'IMAGE_MEM', 'IMAGE_STORAGE', 'INSTANCE_DISK_SIZE', 'INSTANCE_DISK_TYPE', 'IO_THROUGHPUT', 'LICENSE_ACCESS', 'MEM', 'MEM_ALLOCATION', 'MEM_REQUEST_ALLOCATION', 'MEM_PROVISIONED', 'NET_THROUGHPUT', 'NETWORK', 'NUM_DISK', 'NUMBER_CONSUMERS', 'NUM_VCORE', 'POOL_CPU', 'POOL_MEM', 'POOL_STORAGE', 'PORT_CHANNEL', 'POWER', 'Q16_VCPU', 'Q1_VCPU', 'Q2_VCPU', 'Q32_VCPU', 'Q4_VCPU', 'Q64_VCPU', 'Q8_VCPU', 'RESPONSE_TIME', 'SAME_CLUSTER_MOVE_SVC', 'SEGMENTATION', 'SLA_COMMODITY', 'SOFTWARE_LICENSE_COMMODITY', 'SPACE', 'STORAGE_ACCESS', 'STORAGE_ALLOCATION', 'STORAGE_AMOUNT', 'STORAGE_CLUSTER', 'STORAGE_LATENCY', 'STORAGE_PROVISIONED', 'SWAPPING', 'THREADS', 'TRANSACTION', 'TRANSACTION_LOG', 'VCPU', 'VCPU_LIMIT_QUOTA', 'VCPU_REQUEST', 'VCPU_REQUEST_QUOTA', 'VDC', 'VMEM', 'VMEM_LIMIT_QUOTA', 'VMEM_REQUEST', 'VMEM_REQUEST_QUOTA', 'VMPM_ACCESS', 'VSTORAGE', 'ZONE', 'LICENSE_COMMODITY', 'Q3_VCPU', 'NUMBER_CONSUMERS_PM', 'Q6_VCPU', 'Q7_VCPU', 'QN_VCPU', 'RIGHT_SIZE_SVC', 'RIGHT_SIZE_DOWN', 'MOVE', 'Q5_VCPU', 'STORAGE', 'NUMBER_CONSUMERS_STORAGE', 'ACCESS', 'RIGHT_SIZE_UP', 'VAPP_ACCESS', 'HOT_STORAGE', 'HA_COMMODITY', 'NETWORK_POLICY', 'SERVICE_LEVEL_CLUSTER', 'PROCESSING_UNITS', 'TENANCY_ACCESS', 'TEMPLATE_ACCESS', 'BURST_BALANCE', 'DESIRED_COUPON', 'NETWORK_INTERFACE_COUNT', 'BICLIQUE', 'TAINT', 'LABEL', 'PEER_TO_PEER_AFFINITY', 'PEER_TO_PEER_ANTI_AFFINITY', 'ENERGY', 'CARBON', 'VIRTUAL_PROCESSORS', 'CPU_READY', 'PROCESSING_UNITS_PROVISIONED', 'HARDWARE_VERSION', 'GPU_MEM_ALLOCATION', 'GPU_MEM', 'GPU_COUNT', 'GPU_TENSOR_ENGINE', 'GPU_FP64_ENGINE', 'GPU_FP32_ENGINE', 'GPU_FP16_ENGINE', 'GPU_MEMORY_BANDWIDTH', 'CORE_SPEED', 'VGPU_MEM', 'CONCURRENT_QUERIES', 'VGPU_COMPUTE', 'GPU', 'GPU_REQUEST', 'GPU_REQUEST_QUOTA', 'QUEUING_TIME', 'SERVICE_TIME', 'RAID_HOSTS', 'UNKNOWN']

filters

- **type:** array
- **description:** Describes the grouping this commodity is included in, for examplewhether the commodity has real time stats available.
- – **reference:** StatFilterApiDTO *(on page 1682)*

CommodityFieldApiDTO

Description

Inherits From

This DTO inherits from FieldApiDTO *(on page 1562)*

Required Parameters

commodityAttribute

- **type:** string
- **description:** The commodity attribute.
- **readOnly:** True
- **enum:** ['USED', 'CAPACITY', 'PEAK', 'CURRENT_UTILIZATION', 'WEIGHTED_HISTORICAL_UTILIZATION', 'PERCENTILE_HISTORICAL_UTILIZATION']

commodityType

- **type:** string
- **description:** The commodity name.
- **readOnly:** True
- **enum:** ['ACTION_PERMIT', 'ACTIVE_SESSIONS', 'APPLICATION', 'BALLOONING', 'BUFFER_COMMODITY', 'CLUSTER', 'COLLECTION_TIME', 'COOLING', 'CONNECTION', 'COUPON', 'CPU', 'CPU_ALLOCATION', 'CPU_PROVISIONED', 'CPU_REQUEST_ALLOCATION', 'CROSS_CLOUD_MOVE_SVC', 'CROSS_CLUSTER_MOVE_SVC', 'DATACENTER', 'DATASTORE', 'DB_CACHE_HIT_RATE', 'DB_MEM', 'DISK_ARRAY_ACCESS', 'DRS_SEGMENTATION', 'DSPM_ACCESS', 'EXTENT', 'FLOW', 'FLOW_ALLOCATION', 'HEAP', 'HOST_LUN_ACCESS', 'IMAGE_CPU', 'IMAGE_MEM', 'IMAGE_STORAGE', 'INSTANCE_DISK_SIZE', 'INSTANCE_DISK_TYPE', 'IO_THROUGHPUT', 'LICENSE_ACCESS', 'MEM', 'MEM_ALLOCATION', 'MEM_REQUEST_ALLOCATION', 'MEM_PROVISIONED', 'NET_THROUGHPUT', 'NETWORK', 'NUM_DISK', 'NUMBER_CONSUMERS', 'NUM_VCORE', 'POOL_CPU', 'POOL_MEM', 'POOL_STORAGE', 'PORT_CHANNEL', 'POWER', 'Q16_VCPU', 'Q1_VCPU', 'Q2_VCPU', 'Q32_VCPU', 'Q4_VCPU', 'Q64_VCPU', 'Q8_VCPU', 'RESPONSE_TIME', 'SAME_CLUSTER_MOVE_SVC', 'SEGMENTATION', 'SLA_COMMODITY', 'SOFTWARE_LICENSE_COMMODITY', 'SPACE', 'STORAGE_ACCESS', 'STORAGE_ALLOCATION', 'STORAGE_AMOUNT', 'STORAGE_CLUSTER', 'STORAGE_LATENCY', 'STORAGE_PROVISIONED', 'SWAPPING', 'THREADS', 'TRANSACTION', 'TRANSACTION_LOG', 'VCPU', 'VCPU_LIMIT_QUOTA', 'VCPU_REQUEST', 'VCPU_REQUEST_QUOTA', 'VDC', 'VMEM', 'VMEM_LIMIT_QUOTA', 'VMEM_REQUEST', 'VMEM_REQUEST_QUOTA', 'VMPM_ACCESS', 'VSTORAGE', 'ZONE', 'LICENSE_COMMODITY', 'Q3_VCPU', 'NUMBER_CONSUMERS_PM', 'Q6_VCPU', 'Q7_VCPU', 'QN_VCPU', 'RIGHT_SIZE_SVC', 'RIGHT_SIZE_DOWN', 'MOVE', 'Q5_VCPU', 'STORAGE', 'NUMBER_CONSUMERS_STORAGE', 'ACCESS', 'RIGHT_SIZE_UP', 'VAPP_ACCESS', 'HOT_STORAGE', 'HA_COMMODITY', 'NETWORK_POLICY', 'SERVICE_LEVEL_CLUSTER', 'PROCESSING_UNITS', 'TENANCY_ACCESS', 'TEMPLATE_ACCESS', 'BURST_BALANCE', 'DESIRED_COUPON', 'NETWORK_INTERFACE_COUNT', 'BICLIQUE', 'TAINT', 'LABEL', 'PEER_TO_PEER_AFFINITY', 'PEER_TO_PEER_ANTI_AFFINITY', 'ENERGY', 'CARBON', 'VIRTUAL_PROCESSORS', 'CPU_READY', 'PROCESSING_UNITS_PROVISIONED', 'HARDWARE_VERSION', 'GPU_MEM_ALLOCATION', 'GPU_MEM', 'GPU_COUNT', 'GPU_TENSOR_ENGINE', 'GPU_FP64_ENGINE', 'GPU_FP32_ENGINE', 'GPU_FP16_ENGINE', 'GPU_MEMORY_BANDWIDTH', 'CORE_SPEED', 'VGPU_MEM', 'CONCURRENT_QUERIES', 'VGPU_COMPUTE', 'GPU', 'GPU_REQUEST', 'GPU_REQUEST_QUOTA', 'QUEUING_TIME', 'SERVICE_TIME', 'RAID_HOSTS', 'UNKNOWN']

ComputeTierAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Required Parameters

instanceStorageSize

- **type:** number
- **format:** float
- **description:** Size per instance storage in GB. This will be 0 if no instance storage's are present for compute tier.

numInstanceStorages

- **type:** number
- **format:** float
- **description:** Number of instance storage disks. Compute tiers can have 0 or more instance storage disks.

tierFamily

- **type:** string
- **description:** Family of compute tier (for example 'M4' for M4.large)

Optional Parameters

customConfigurable

- **type:** boolean
- **example:** True
- **description:** If the compute tier has commodities that can be configured per VM.

acceleratorCards

- **type:** array
- **description:** The Accelerator/GPU cards
- – **reference:** AcceleratorCardApiDTO *(on page 1484)*

ConditionApiDTO

Description

A condition on a field of an entity or group.

Required Parameters

conditionType

- **type:** string
- **description:** Type of compared values.
- **readOnly:** True
- **enum:** ['NUMBER', 'TEXT', 'INCLUSION', 'INTEGER', 'BOOLEAN']

field

- **description:** Field to compare.
- **readOnly:** True
- **reference:** FieldApiDTO *(on page 1562)*

operator

- **type:** string
- **description:** Operator to use.
- **readOnly:** True
- **enum:** ['EQ', 'NEQ', 'GT', 'LT', 'GE', 'LE', 'LIKE', 'IN']

ConfigChangesApiDTO

Description

Policies and Settings changes in the scenario

Optional Parameters

addPolicyList

- **type:** array
- **description:** List of Placement Policies to add
- – **reference:** PolicyApiDTO *(on page 1622)*

removePolicyList

- **type:** array
- **description:** List of Placement Policies to remove
- – **reference:** PolicyApiDTO *(on page 1622)*

automationSettingList

- **type:** array
- **description:** List of automation settings
- – **reference:** SettingApiDTOString *(on page 1673)*

removeConstraintList

- **type:** array
- **description:** List of constraints to remove
- – **reference:** RemoveConstraintApiDTO *(on page 1642)*

riSettingList

- **type:** array
- **description:** List of Reserved Instance settings
- – **reference:** SettingApiDTO *(on page 1675)*

osMigrationSettingList

- **type:** array
- **description:** List of OS migration settings
- – **reference:** SettingApiDTOString *(on page 1673)*

includedCoupons

- **description:** List of Included Coupons (RIs) and related information
- **reference:** IncludedCouponsApiDTO *(on page 1570)*

subscription

- **description:** Subscription
- **reference:** BusinessUnitApiDTO *(on page 1514)*

ConstraintApiDTO

Description

Model to describe the property of constraints, corresponding to entities constraints view in ui.

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

relation

- **type:** string
- **enum:** ['sold', 'bought']

entityType

- **type:** string
- **description:** entity type of providers or consumers

relatedEntities

- **type:** array
- **description:** if relation is sold they are consumers, if bought they are providers
- – **reference:** ServiceEntityApiDTO *(on page 1668)*

numPotentialEntities

- **type:** integer
- **format:** int32
- **description:** how many other providers/consumers can fulfill the constraint

placementOptions

- **type:** array
- **description:** A list of constraints prevent finding better placement for an entity
- – **reference:** PlacementOptionApiDTO *(on page 1619)*

ConstraintApiInputDTO

Description

Model to describe the property of constraints

Optional Parameters

relation

- **type:** string
- **enum:** ['sold', 'bought']

placementOptions

- **type:** array
- **description:** A list of constraints prevent finding better placement for an entity
- – **reference:** PlacementOptionApiInputDTO *(on page 1620)*

entityTypeFilter

- **type:** string
- **description:** potential entity type filter
- **enum:** ['Application', 'ApplicationComponent', 'ApplicationServer', 'AvailabilityZone', 'BusinessAccount', 'BusinessApplication', 'BusinessTransaction', 'BusinessUser', 'Chassis', 'CloudService', 'ComputeTier', 'Container', 'ContainerPod', 'ContainerSpec', 'Database', 'DatabaseServer', 'DatabaseServerTier', 'DatabaseTier', 'DataCenter', 'DesktopPool', 'DiskArray', 'DPod', 'HCIPhysicalMachine', 'HypervisorServer', 'Internet', 'IOModule', 'LoadBalancer', 'LogicalPool', 'Namespace', 'Network', 'PhysicalMachine', 'ProcessorPool', 'Region', 'ReservedInstance', 'Service', 'ServiceProvider', 'Storage', 'StorageController', 'StorageTier', 'Switch', 'Unknown', 'ViewPod', 'VirtualApplication', 'VirtualDataCenter', 'VirtualMachine', 'VirtualVolume', 'VPod', 'WorkloadController', 'ContainerPlatformCluster', 'CloudCommitment', 'ApplicationComponentSpec', 'VirtualMachineSpec', 'DocumentCollection', 'NodeGroup', 'PowerVMPhysicalMachine', 'DataWarehouse', 'DataWarehouseTier']

ContainerPlatformContextAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

namespace

- **type:** string
- **description:** The namespace that a container platform entity belong to – [DEPRECATED – Use: namespaceEntity]

containerPlatformCluster

- **type:** string
- **description:** The container platform cluster name that an entity belongs to – [DEPRECATED – Use: containerClusterEntity]

workloadControllerEntity

- **description:** The workload controller entity that an entity belongs to
- **reference:** BaseApiDTO *(on page 1509)*

namespaceEntity

- **description:** The namespace entity that an entity belongs to
- **reference:** BaseApiDTO *(on page 1509)*

containerClusterEntity

- **description:** The container platform cluster entity that an entity belongs to
- **reference:** BaseApiDTO *(on page 1509)*

serviceEntity

- **description:** The service associated with the entity
- **reference:** BaseApiDTO *(on page 1509)*

CostInputApiDTO

Description

Model to describe the request for Cost Statistics by a Time range

Optional Parameters

tagFilters

- **type:** array
- **description:** List of Tag Filters. NOTE: Required non-empty when used with costGroupBys set to "TAG"
- – **reference:** TagApiDTO *(on page 1690)*

costGroupBys

- **type:** array
- **description:** List of values to group the results by. NOTE: When used non-empty tagFilters required
- – **type:** string
  – **enum:** ['TAG', 'TAG_GROUP', 'TAG_VAL', 'ENTITY', 'ENTITY_TYPE', 'ACCOUNT', 'REGION', 'CLOUD_SERVICE', 'SERVICE_PROVIDER', 'CLOUD_TIER_PROVIDER', 'COST_CATEGORY', 'PRICE_MODEL']

startDate

- **type:** string
- **description:** Start time of the period to filter. If startDate and endDate not provided current stats will be processed

endDate

- **type:** string
- **description:** End time of the period to filter. If startDate and endDate not provided current stats will be processed

CpuModelApiDTO

Description

Model to describe a Template: contains the resources used to Deploy a VM or to Add Workload/Supply in a Plan

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

numCores

- **type:** integer
- **format:** int32
- **description:** Number of Cores

speed

- **type:** integer
- **format:** int32
- **description:** Speed

scalingFactor

- **type:** number
- **format:** double
- **description:** Scaling Factor

modelName

- **type:** string

CriteriaOptionApiDTO

Description

Generic model used to describe a search criteria option

Optional Parameters

value

- **type:** string
- **description:** value

displayName

- **type:** string
- **description:** value to display, more readable

subValues

- **type:** array
- **description:** possible sub-values associated with the main value
  - **type:** string

DBEntityAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

dbEngine

- **type:** string
- **description:** The Database engine

dbEdition

- **type:** string
- **description:** The Database edition

dbVersion

- **type:** string
- **description:** The Database engine version

licenseModel

- **type:** string
- **description:** The license model of the database engine

deploymentType

- **type:** string
- **description:** The deployment type of the database engine

maxConcurrentWorkers

- **type:** integer
- **format:** int32
- **description:** Maximum concurrent workers allowed for the DB

maxConcurrentSessions

- **type:** integer
- **format:** int32
- **description:** Maximum concurrent sessions allowed for the DB

pricingModel

- **type:** string
- **example:** DTU
- **description:** Database Pricing Model.
- **enum:** ['DTU', 'vCore', 'DWU', 'RU']

replicationRole

- **type:** string
- **example:** Primary
- **description:** Replication role if DB is part of geo replicated DB. Currently applies to Azure DB only.
- **enum:** ['Primary', 'Secondary', 'None']

storageTier

- **type:** string
- **example:** GP2
- **description:** Storage Tier

dbServerName

- **type:** string
- **description:** The server name of a database

serviceTier

- **type:** string
- **example:** Hyperscale
- **description:** The service tier of a database, generally defines the storage architecture, space and I/O limits, and business continuity options related to availability and disaster recovery.
- **enum:** ['Free', 'Basic', 'Standard', 'Premium', 'GeneralPurpose', 'BusinessCritical', 'Hyperscale']

computeTier

- **type:** string
- **example:** Provisioned Compute
- **description:** The compute tier of a database. Compute tier options include the provisioned and serverless compute tiers

hardwareGeneration

- **type:** string
- **example:** Gen4
- **description:** Database hardware generation. The generation of hardware that underlies the DB instance

backupPolicy

- **type:** string
- **example:** Enabled
- **description:** The policy to backup and restore a database.

workspace

- **type:** string
- **description:** The resource group where the database is located.

idleTime

- **type:** integer
- **format:** int64
- **description:** Time duration for a database being idle in milliseconds.

poolType

- **type:** string
- **description:** Type of database pool.

resourceId

- **type:** string
- **description:** Resource identifier of the database. For example, the resource id in Azure, the ARN in AWS, or the full resource name in GCP.

serverConfiguration

- **type:** string
- **description:** The Database server configuration

creationTimeStamp

- **type:** integer
- **format:** int64
- **description:** Timestamp in milliseconds for the creation of the database.

DatabaseServerEntityAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

dbEngine

- **type:** string
- **description:** The Database engine

dbEdition

- **type:** string
- **description:** The Database edition

dbVersion

- **type:** string
- **description:** The Database engine version

licenseModel

- **type:** string
- **description:** The license model of the database engine

deploymentType

- **type:** string
- **description:** The deployment type of the database engine

maxConcurrentWorkers

- **type:** integer
- **format:** int32
- **description:** Maximum concurrent workers allowed for the DB

maxConcurrentSessions

- **type:** integer
- **format:** int32
- **description:** Maximum concurrent sessions allowed for the DB

pricingModel

- **type:** string
- **example:** DTU
- **description:** Pricing Model followed by DB.

storageTier

- **type:** string
- **example:** GP2
- **description:** Storage Tier

clusterRole

- **type:** string
- **description:** Cluster Role
- **enum:** ['Writer', 'Reader']

hourlyBilledOps

- **type:** number
- **format:** double
- **description:** Hourly Billed Ops - Count of operations per hour

featureStateMap

- **type:** object
- **description:** Features
- **additionalProperties:** {'type': 'string', 'enum': ['Enabled', 'Disabled', 'Unknown']}

resourceId

- **type:** string
- **description:** Resource identifier of the database server. For example, the resource id in Azure, the ARN in AWS, or the full resource name in GCP.

creationTimeStamp

- **type:** integer
- **format:** int64
- **description:** Timestamp in milliseconds for the creation of the database server.

serverConfiguration

- **type:** string
- **description:** The Database server configuration

replicas

- **type:** array
- **description:** Replicas associated with this database
- – **reference:** ReplicaApiDTO *(on page 1644)*

databases

- **type:** array
- **description:** Databases that run on this instance
- – **type:** string

fullyQualifiedDomainName

- **type:** string
- **description:** Fully qualified domain name

backupStorageRedundancy

- **type:** string
- **description:** Backup Storage Redundancy - refers to how data is duplicated across storage locations

DatabaseServerTierAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Required Parameters

tierFamily

- **type:** string
- **description:** Family of database server tier (for example 'M5' for db.M5.large)

DatabaseServerTierInfoApiDTO

Description

Model to describe the database server tier to which a cloud commitment applies

Optional Parameters

databaseEngine

- **type:** string
- **description:** The database engine to which a cloud commitment applies
- **enum:** ['UNKNOWN', 'MARIADB', 'MYSQL', 'POSTGRESQL', 'AURORAMYSQL', 'AURORAPOSTGRESQL', 'SQLSERVER', 'ORACLE', 'DB2']

deploymentType

- **type:** string
- **description:** The deployment type that the commitment covers
- **enum:** ['SINGLE_AZ', 'MULTI_AZ']

licenseType

- **type:** string
- **description:** The license type associated with the commitment
- **enum:** ['LICENSE_INCLUDED', 'AHUB', 'BRING_YOUR_OWN_LICENSE', 'NO_LICENSE_REQUIRED']

DatabaseTierAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Required Parameters

tierFamily

- **type:** string
- **description:** Family of database tier (for example 'S' for S4)

DemandEntityInfoDTO

Description

Model to describe a reservation entity

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

template

- **reference:** BaseApiDTO *(on page 1509)*

deploymentProfile

- **reference:** BaseApiDTO *(on page 1509)*

operationalCost

- **type:** string

placements

- **reference:** PlacementInfoDTO *(on page 1618)*

DemandReservationApiDTO

Description

Basic information of an entity

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

prefix

- **type:** string

count

- **type:** integer
- **format:** int32

status

- **type:** string

reserveDateTime

- **type:** string

expireDateTime

- **type:** string

deployDateTime

- **type:** string

reserveCount

- **type:** integer
- **format:** int32

deployCount

- **type:** integer
- **format:** int32

placementExpirationDateTime

- **type:** string

placementResultMessage

- **type:** string
- **description:** Details of placement during failure or success

constraintInfos

- **type:** array
- **description:** The constraints user specified when running reservation which limit where templates could be placed.
- – **reference:** ReservationConstraintApiDTO *(on page 1645)*

reservationDeployed

- **type:** boolean
- **description:** When true, the reservation will be included in the system and will keep utilizing the resources from host and storage until the corresponding virtual machine utilization starts to reflect on the historically computed utilization.

mode

- **type:** string
- **description:** Demand related behavior for reservation of NO_GROUPING/AFFINITY. This value interacts with the grouping value which is specifying the type of grouping that the demand should consider the mode for. NO_GROUPING and NONE must be paired together and cannot be paired with other values. Combinations such as: (NO_GROUPING - NONE), (AFFINITY - CLUSTER) are valid. Examples of invalid combinations are (NO_GROUPING - CLUSTER), (AFFINITY - NONE).
- **enum:** ['NO_GROUPING', 'AFFINITY']

grouping

- **type:** string
- **description:** Grouping related behavior for reservation of NONE and CLUSTER. This value interacts with the mode value.
- **enum:** ['NONE', 'CLUSTER']

scope

- **type:** array
- **description:** List of uuids of groups (currently limited to Clusters or Host Groups) whose members are the only possible providers of that type of entity. These are further constrained by the constraintIds passed in placement parameters. If multiple groups of providers are specified then it is possible that entities will be placed within different groups.
- – **type:** string

demandEntities

- **type:** array
- – **reference:** DemandEntityInfoDTO *(on page 1546)*

DemandReservationApiInputDTO

Description

Required Parameters

action

- **type:** string
- **description:** Action requested for the reservation
- **enum:** ['RESERVATION']

demandName

- **type:** string
- **description:** Name of the reservation
- **minLength:** 1
- **maxLength:** 2147483647

parameters

- **type:** array
- **description:** Reservation Parameters
- – **reference:** DemandReservationParametersDTO *(on page 1549)*
- **maxItems:** 1
- **minItems:** 1

Optional Parameters

mode

- **type:** string
- **description:** Demand related behavior for reservation of NO_GROUPING/AFFINITY. This value interacts with the grouping value which is specifying the type of grouping that the demand should consider the mode for. NO_GROUPING and NONE must be paired together and cannot be paired with other values. Combinations such as: (NO_GROUPING - NONE), (AFFINITY - CLUSTER) are valid. Examples of invalid combinations are (NO_GROUPING - CLUSTER), (AFFINITY - NONE).
- **enum:** ['NO_GROUPING', 'AFFINITY']

grouping

- **type:** string
- **description:** Grouping related behavior for reservation of NONE and CLUSTER. This value interacts with the mode value.
- **enum:** ['NONE', 'CLUSTER']

reserveDateTime

- **type:** string
- **description:** Date to specify when to reserve the Demand

expireDateTime

- **type:** string
- **description:** Date to specify when the Demand request will expire, required if action = RESERVATION

deployDateTime

- **type:** string
- **description:** Date to specify when to deploy the Demand

scope

- **type:** array
- **description:** List of uuids of groups (currently limited to Clusters or Host Groups) whose members are the only possible providers of that type of entity. These are further constrained by the constraintIds passed in placement parameters. If multiple groups of providers are specified then it is possible that entities will be placed within different groups.
- – **type:** string

DemandReservationParametersDTO

Description

Required Parameters

placementParameters

- **description:** Placement parameters
- **reference:** PlacementParametersDTO *(on page 1620)*

Optional Parameters

deploymentParameters

- **description:** Deployment parameters, required if action = DEPLOYMENT, RESERVATION
- **reference:** DeploymentParametersDTO *(on page 1550)*

DeploymentInfoDTO

Description

Contains deployment information about the running Turbonomic instance.

Optional Parameters

clientId

- **type:** string
- **example:** turbonomic1
- **description:** The client Id

clientSvcAccount

- **type:** string
- **example:** turbonomic1-sa
- **description:** The client service account

oidcProviderUrl

- **type:** string
- **example:** https://rh-oidc.s3.us-east-1.amazonaws.com/22ejnvnnturfmt6km08idd0nt4hekbn7
- **description:** The OIDC provider URL

DeploymentParametersDTO

Description

Optional Parameters

deploymentProfileID

- **type:** string
- **description:** Deployment Profile uuid

highAvailability

- **type:** boolean
- **description:** High availability

priority

- **type:** string
- **description:** Workload priority
- **enum:** ['LOW', 'NORMAL', 'MISSION_CRITICAL']

DeploymentProfileApiDTO

Description

Model to describe a Deployment Profile, including target and provider entity parameters

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

account

- **description:** Business account related to the Deployment Profile
- **reference:** BaseApiDTO *(on page 1509)*

deployParameters

- **type:** array
- **description:** Target and Provider parameters of the Deployment Profile
- – **reference:** DeploymentProfileTargetApiDTO *(on page 1552)*

DeploymentProfileParamApiDTO

Description

Provider-specific Deployment Profile parameters

Optional Parameters

parameterType

- **type:** string
- **description:** Parameter type. For vCenter, this can be a folder, osConfiguration, or image. For all other technologies, this must be an image.
- **enum:** ['image', 'folder', 'osConfiguration']

properties

- **type:** array
- **description:** A NameValueInputDTO describing the name and value of the parameter.
- – **reference:** NameValueInputDTO *(on page 1589)*

DeploymentProfileProviderApiDTO

Description

Deployment profile Provider properties

Optional Parameters

provider

- **description:** Provider UUID
- **reference:** BaseApiDTO *(on page 1509)*

parameters

- **type:** array
- **description:** A DeploymentProfileParamApiDTO describing the parameters for the provider entities
- – **reference:** DeploymentProfileParamApiDTO *(on page 1551)*

DeploymentProfileTargetApiDTO

Description

Deployment Profile Target properties

Optional Parameters

targetType

- **type:** string
- **description:** Target type for the deployment profile
- **enum:** ['vCenter', 'AWS', 'Azure', 'Softlayer']

providers

- **type:** array
- **description:** DeploymentProfileProviderApiInputDTO describing the provider entity parameters
- – **reference:** DeploymentProfileProviderApiDTO *(on page 1551)*

DesktopPoolEntityAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

cloneType

- **type:** string
- **description:** Type of a clone provided by the DesktopPool
- **enum:** ['FULL', 'LINKED', 'INSTANT']

provisionType

- **type:** string
- **description:** How clones are provisioned in the DesktopPool
- **enum:** ['UPFRONT', 'ON_DEMAND']

assignmentType

- **type:** string
- **description:** Whether the same clone is assigned each time when user logs back in
- **enum:** ['PERSISTENT', 'DYNAMIC']

masterTemplateUuid

- **type:** string
- **description:** UUID of a master template that is used to create clones. Only one of master VM and master template can be set.

masterVirtualMachineUuid

- **type:** string
- **description:** UUID of a master virtual machine that is used to create clones. Only one of master VM and master template can be set.

masterVirtualMachineSnapshot

- **type:** string
- **description:** Name of the snapshot of a master virtual machine that is used to create clones.

getvCenterClusterName

- **type:** string
- **description:** Name of the vCenter cluster that this DesktopPool draws resources from.

vendorId

- **type:** string
- **description:** Vendor Identifier for the DesktopPool.

DetailDataApiDTO

Description

A single entity detail, represented as key-value pair

Required Parameters

key

- **type:** string
- **description:** Detail key

value

- **type:** string
- **description:** Detail value

Optional Parameters

critical

- **type:** boolean

DiscountCoverageFilterApiDTO

Description

Model to describe the discount coverage filter information for the action filtering.

Optional Parameters

currentDiscountCoverageFilter

- **description:** Filter the actions based on a current discount coverage.
- **reference:** RangeInputApiDTO *(on page 1637)*

newDiscountCoverageFilter

- **description:** Filter the actions based on a new discount coverage.
- **reference:** RangeInputApiDTO *(on page 1637)*

DiscoveredEntityApiDTO

Description

Model to describe a discovered entity in the system

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

environmentType

- **type:** string
- **description:** Environment type
- **enum:** ['CLOUD', 'ONPREM', 'HYBRID']

discoveredBy

- **description:** Target that discovered the entity
- **reference:** TargetApiDTO *(on page 1690)*

vendorIds

- **type:** object
- **description:** The mapping of target identifier to vendor-provided identity of this entity on the remote target
- **additionalProperties:** {'type': 'string'}

EffectiveCostFilterApiDTO

Description

Model to describe effective cost range for action filtering.

Optional Parameters

currentEffectiveCostRange

- **description:** Filter the actions by the range for current effective cost.
- **reference:** RangeInputApiDTO *(on page 1637)*

newEffectiveCostRange

- **description:** Filter the actions by the range for destination effective cost. Applicable only for scale actions.
- **reference:** RangeInputApiDTO *(on page 1637)*

EntityActionsApiDTO

Description

Basic informations of an entity + actions

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

realtimeMarketReference

- **description:** In case of a Plan Market Entity, this contains the original entity from the Real Time Market
- **reference:** BaseApiDTO *(on page 1509)*

actions

- **type:** array
- **description:** List of actions for this entity
- – **reference:** ActionApiDTO *(on page 1485)*

EntityAspect

Description

Parent class for entity aspects

Required Parameters

type

- **type:** string

Optional Parameters

sli

- **type:** array
- – **type:** string

EntityCostApiDTO

Description

Model to describe the Costs of an entity

Optional Parameters

value

- **type:** number
- **format:** float
- **description:** The cost value

hairCutRatio

- **type:** number
- **format:** float
- **description:** The hair cut ratio, it represents the portion of cost will be taken into consideration

costType

- **type:** string
- **description:** Type of the Cost

costUnit

- **type:** string
- **description:** Unit of the Cost

EntityCountRequestApiDTO

Description

A query used to fetch entity counts.

Optional Parameters

groupBy

- **type:** array
- **description:** Criteria to group the results by.
- **readOnly:** True
- – **type:** string
- – **enum:** ['ENTITY_TYPE', 'ENVIRONMENT_TYPE']

EntityDetailsApiDTO

Description

Model to describe an entity with all its details

Required Parameters

uuid

- **type:** integer
- **format:** int64
- **description:** The internal UUID of the entity

Optional Parameters

details

- **type:** array
- **description:** A list of details, represented as key-value
- – **reference:** DetailDataApiDTO *(on page 1553)*

EntityMetadataRequestApiDTO

Description

A query used to fetch entity metadata.

Required Parameters

entityType

- **type:** string
- **description:** Entity type to fetch metadata for.
- **readOnly:** True
- **enum:** ['Application', 'ApplicationComponent', 'ApplicationServer', 'AvailabilityZone', 'BusinessAccount', 'BusinessApplication', 'BusinessTransaction', 'BusinessUser', 'Chassis', 'CloudService', 'ComputeTier', 'Container', 'ContainerPod', 'ContainerSpec', 'Database', 'DatabaseServer', 'DatabaseServerTier', 'DatabaseTier', 'DataCenter', 'DesktopPool', 'DiskArray', 'DPod', 'HCIPhysicalMachine', 'HypervisorServer', 'Internet', 'IOModule', 'LoadBalancer', 'LogicalPool', 'Namespace', 'Network', 'PhysicalMachine', 'ProcessorPool', 'Region', 'ReservedInstance', 'Service', 'ServiceProvider', 'Storage', 'StorageController', 'StorageTier', 'Switch', 'Unknown', 'ViewPod', 'VirtualApplication', 'VirtualDataCenter', 'VirtualMachine', 'VirtualVolume', 'VPod', 'WorkloadController', 'ContainerPlatformCluster', 'CloudCommitment', 'ApplicationComponentSpec', 'VirtualMachineSpec', 'DocumentCollection', 'NodeGroup', 'PowerVMPhysicalMachine', 'DataWarehouse', 'DataWarehouseTier']

EntityPriceDTO

Description

Context-specific price. For example the base rate of a template in the context of certain business unit.

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

price

- **type:** number
- **format:** float
- **description:** Price in the context of a related entity

EntityQueryApiDTO

Description

A query used to search topology entities.

Required Parameters

select

- **description:** Information to be fetched for each entity.
- **readOnly:** True
- **reference:** SelectEntityApiDTO *(on page 1666)*

Optional Parameters

where

- **description:** Conditions that each retrieved entity must satisfy.
- **readOnly:** True
- **reference:** WhereApiDTO *(on page 1719)*

pagination

- **description:** Pagination information.
- **readOnly:** True
- **reference:** PaginationApiDTO *(on page 1602)*

EntityStatsApiDTO

Description

Basic informations of an entity + statistics

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

environmentType

- **type:** string
- **description:** Environment type
- **enum:** ['CLOUD', 'ONPREM', 'HYBRID']

realtimeMarketReference

- **description:** In case of a Plan Market Entity, this contains the original entity from the Real Time Market
- **reference:** BaseApiDTO *(on page 1509)*

stats

- **type:** array
- **description:** Collected monitoring statistics
- – **reference:** StatSnapshotApiDTO *(on page 1684)*

EntityUptimeApiDTO

Description

Model to describe the uptime of an entity

Optional Parameters

creationTimestamp

- **type:** integer
- **format:** int64
- **description:** The creation timestamp of the entity

uptimeDurationInMilliseconds

- **type:** integer
- **format:** int64
- **description:** Uptime duration in milliseconds for the total analyzed duration

totalDurationInMilliseconds

- **type:** integer
- **format:** int64
- **description:** Total analyzed duration in milliseconds for which uptime is calculated

uptimePercentage

- **type:** number
- **format:** double
- **description:** The uptime of the entity as a percentage

EnumFieldValueApiDTO

Description

Inherits From

This DTO inherits from FieldValueApiDTO *(on page 1562)*

Required Parameters

value

- **type:** string
- **description:** The string representation of the value.
- **readOnly:** True

ExecutorInfoApiDTO

Description

Model to describe the entity that triggered action execution. such as a user or a schedule.

Required Parameters

type

- **type:** string
- **description:** The type of the executor. such as a user or a schedule.
- **enum:** ['USER', 'SCHEDULE', 'EXECUTED_EXTERNALLY']

Optional Parameters

scheduleInfo

- **description:** Information about the schedule that triggered this action. Omitted if the action was not triggered by a schedule.
- **reference:** ExecutorScheduleApiDTO *(on page 1561)*

userInfo

- **description:** Information about the user who triggered this action. Omitted if the action was not triggered by a user.
- **reference:** ExecutorUserApiDTO *(on page 1561)*

ExecutorScheduleApiDTO

Description

Model to describe the Schedule that executed the action.

Required Parameters

scheduleName

- **type:** string
- **description:** Name of the schedule

ExecutorUserApiDTO

Description

Model to describe the user who triggered this action.

Required Parameters

userName

- **type:** string
- **description:** Name of the user

FailedResourceApiDTO

Description

A description of the resource that caused an entity to be unplaced

Required Parameters

commodity

- **description:** The commodity that caused the entity to be unplaced
- **reference:** BaseCommodityApiDTO *(on page 1509)*

requestedAmount

- **type:** number
- **format:** double
- **description:** The requested amount of the commodity that caused the entity to be unplaced

Optional Parameters

maxAvailable

- **type:** number
- **format:** double
- **description:** The maximum available amount of the requested commodity. Populated only if there are one or more entities selling the requested commodity.

FamilyScopeApiDTO

Description

Model to describe the instance family of a cloud commitment

Optional Parameters

family

- **type:** string
- **description:** The name of the instance family

FieldApiDTO

Description

A field of an entity or group.

Optional Parameters

fieldType

- **type:** string
- **description:** The kind of the field.
- **enum:** ['PRIMITIVE', 'COMMODITY', 'RELATED_ENTITY', 'RELATED_GROUP', 'RELATED_ACTION', 'AGGREGATE_COMMODITY', 'MEMBER']

FieldValueApiDTO

Description

A field of an entity or group with its value.

Required Parameters

field

- **description:** Field to compare.
- **readOnly:** True
- **reference:** FieldApiDTO *(on page 1562)*

valueType

- **type:** string
- **description:** Type of compared values.
- **readOnly:** True
- **enum:** ['NUMBER', 'TEXT', 'ENUM', 'INTEGER', 'BOOLEAN', 'MULTI_TEXT']

FieldValueTypeApiDTO

Description

A field of an entity or group associated with the type of its values.

Required Parameters

field

- **description:** Field for which information is presented.
- **readOnly:** True
- **reference:** FieldApiDTO *(on page 1562)*

valueType

- **type:** string
- **description:** Type of values of this field.
- **readOnly:** True
- **enum:** ['NUMBER', 'TEXT', 'ENUM', 'INTEGER', 'BOOLEAN', 'MULTI_TEXT']

Optional Parameters

allowableValues

- **type:** array
- **description:** Allowable string values for this field.
- **readOnly:** True
- – **type:** string

FilterApiDTO

Description

Model used to filter Groups or Entities, every filter is built using a filterType + expType + expVal

Required Parameters

caseSensitive

- **type:** boolean
- **description:** Whether the regex is case sensitive

expType

- **type:** string
- **description:** Type of comparator used while building a filter, the allowable values are 'EQ', 'NEQ', 'RXEQ', 'RXNEQ' for strings, 'EQ', 'NEQ', 'GT', 'LT', 'GTE', 'LTE' for numbers'EX', 'NEX' for inclusion in a set of options
- **enum:** ['EQ', 'NEQ', 'GT', 'LT', 'GTE', 'LTE', 'RXEQ', 'RXNEQ', 'EX', 'NEX']

expVal

- **type:** string
- **description:** Regex to be searched for

filterType

- **type:** string
- **description:** Internal name for criterion such as pmsByName, pmsByDC, storageByPMCluster

Optional Parameters

entityType

- **type:** string
- **description:** The entity type which this filter should be applied on

singleLine

- **type:** boolean
- **description:** Whether the regex dot matches all characters including line break characters. Defaults to false.

FrontdoorIdpApiDTO

Description

Model to describe a Frontdoor IDP. It contains IDP URL and is FD enabled or not

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

isFrontdoorEnabled

- **type:** boolean

GroupApiDTO

Description

Model to describe a Group: a collection of Entities. Subclasses include BillingFamilyApiDTO and ResourceGroupApiDTO.

Required Parameters

groupClassName

- **type:** string

isStatic

- **type:** boolean
- **description:** True if group is static

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

membersCount

- **type:** integer
- **format:** int32
- **description:** Number of members of the Group

entitiesCount

- **type:** integer
- **format:** int32
- **description:** Number of entities of the Group

costPrice

- **type:** number
- **format:** float
- **description:** Cost of the Group per Hour: sum of the costs of the member entities

groupType

- **type:** string
- **description:** The type of service entities comprising the group

severity

- **type:** string
- **description:** Calculated using the highest severity of the member entities
- **enum:** ['UNKNOWN', 'NORMAL', 'MINOR', 'MAJOR', 'CRITICAL']

state

- **type:** string
- **description:** Calculated using the state of the member entities
- **enum:** ['UNKNOWN', 'ACTIVE']

environmentType

- **type:** string
- **description:** Environment type
- **enum:** ['HYBRID', 'CLOUD', 'ONPREM', 'UNKNOWN']

scope

- **type:** array
- **description:** Scope within which the criteria have to be applied to the list of specific entity type
  - **type:** string

logicalOperator

- **type:** string
- **description:** Logical operator to be applied across all the criteria used to create dynamic group.It can take values 'AND', 'OR', 'XOR'

criteriaList

- **type:** array
- **description:** Criteria list of elements, expression type, expression value, seType – needed to create a dynamic group
- – **reference:** FilterApiDTO *(on page 1563)*

memberUuidList

- **type:** array
- **description:** Uuid list for members of the group – if group is static
- – **type:** string

realtimeMarketReference

- **description:** In case of a Plan Market Group, this contains the original group from the Real Time Market
- **reference:** BaseApiDTO *(on page 1509)*

stats

- **type:** array
- **description:** Collected monitoring statistics
- – **reference:** StatSnapshotApiDTO *(on page 1684)*

temporary

- **type:** boolean
- **description:** The Group is only valid for a limited period of time, it will be removed automatically

activeEntitiesCount

- **type:** integer
- **format:** int32
- **description:** The active entities count of a group

cloudType

- **type:** string
- **description:** Type of the Cloud if environmentType = CLOUD.[DEPRECATED – Use: targetType
- **enum:** ['AWS', 'AZURE', 'GCP', 'HYBRID', 'UNKNOWN']

source

- **description:** The Source of the Group
- **reference:** TargetApiDTO *(on page 1690)*

memberTypes

- **type:** array
- **description:** The types for immediate members of the group.
- – **type:** string

entityTypes

- **type:** array
- **description:** The types of entities contained in the group. This includes types of entities in nested levels of the group if the group is nested.
- – **type:** string

aspects

- **type:** object
- **description:** Additional info about the Group categorized as Aspects
- **additionalProperties:** {'$ref': '#/definitions/EntityAspect'}

vendorIds

- **type:** object
- **description:** The mapping of target identifier to vendor-provided identity of this group, if the group is discovered
- **additionalProperties:** {'type': 'string'}

groupOrigin

- **type:** string
- **description:** Origin of entities and groups
- **enum:** ['DISCOVERED', 'USER', 'SYSTEM']

targetType

- **type:** string
- **description:** Target type of group entities. TargetType UNKNOWN if multiple target types

GroupCountRequestApiDTO

Description

A query used to fetch group counts.

Optional Parameters

groupBy

- **type:** array
- **description:** Criteria to group the results by.
- **readOnly:** True
- **uniqueItems:** True
- – **type:** string
  – **enum:** ['GROUP_TYPE', 'ORIGIN']

GroupMetadataRequestApiDTO

Description

A query used to fetch group metadata.

Required Parameters

groupType

- **type:** string
- **description:** Group type to fetch metadata for.
- **readOnly:** True
- **enum:** ['Group', 'Resource', 'Cluster', 'VMCluster', 'StorageCluster', 'BusinessAccountFolder', 'BillingFamily', 'NodePool']

GroupQueryApiDTO

Description

A query used to search groups.

Required Parameters

select

- **description:** Information to be fetched for each group.
- **readOnly:** True
- **reference:** SelectGroupApiDTO *(on page 1667)*

Optional Parameters

where

- **description:** Conditions that each retrieved group must satisfy.
- **readOnly:** True
- **reference:** WhereApiDTO *(on page 1719)*

pagination

- **description:** Pagination information.
- **readOnly:** True
- **reference:** PaginationApiDTO *(on page 1602)*

HealthCategoryResponseDTO

Description

A DTO to collect the aggregated health data for one health category.

Required Parameters

categoryDisplayName

- **type:** string
- **description:** The reported health category name.

categoryHealthState

- **type:** string
- **description:** The health state for the reported category.
- **enum:** ['CRITICAL', 'MAJOR', 'MINOR', 'NORMAL']

healthCategory

- **type:** string
- **description:** The reported health category.
- **enum:** ['TARGET', 'ACTION', 'COMPONENT']

Optional Parameters

responseItems

- **type:** array
- **description:** A list of individual health response items for this health category.
- – **reference:** AggregatedHealthResponseDTO *(on page 1505)*

HttpProxyDTO

Description

Model to describe http proxy settings, like username, password, port number and host

Required Parameters

isProxyEnabled

- **type:** boolean
- **description:** Whether proxy is enabled or not

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

proxyHost

- **type:** string
- **description:** Proxy host

proxyPortNumber

- **type:** integer
- **format:** int32
- **description:** Proxy port number

userName

- **type:** string
- **description:** Proxy username

password

- **type:** string
- **description:** Proxy password

isProxySecure

- **type:** boolean
- **description:** Whether proxy is secure: SSL is used to connect to the proxy host. Defaults for 'false'.

IEntityDefinitionData

Description

Parent class for topology data definition

Required Parameters

definitionType

- **type:** string
- **enum:** ['MANUAL', 'AUTOMATED']

IManualConnectionsData

Description

Parent class for manual connections.

Required Parameters

connectionType

- **type:** string
- **enum:** ['DYNAMIC', 'STATIC', 'GROUP', 'GROUPS']

membersCount

- **type:** integer
- **format:** int32

IncludedCouponsApiDTO

Description

Included RI/Coupon and related information in the scenario

Optional Parameters

includedCouponOidsList

- **type:** array
- **description:** User selected list of RIs to be included in plans
  - **type:** integer
  - **format:** int64

iswhiteList

- **type:** boolean

InclusionConditionApiDTO

Description

Inherits From

This DTO inherits from ConditionApiDTO *(on page 1535)*

Required Parameters

value

- **type:** array
- **description:** A set of strings. The value must belong to one of those.
- **readOnly:** True
- – **type:** string

InputFieldApiDTO

Description

Basic information of an entity

Required Parameters

name

- **type:** string
- **description:** Name of the field, used for field identification.

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

value

- **type:** string
- **description:** Field value. Used if field holds primitive value (String, number or boolean

defaultValue

- **type:** string
- **description:** Default value of the field

isMandatory

- **type:** boolean
- **description:** Whether the field is mandatory. Valid targets must have all the mandatory fields set.

isSecret

- **type:** boolean
- **description:** Whether the field is secret. This means, that field value is stored in an encrypted value and not shown in any logs.

isMultiline

- **type:** boolean
- **description:** Whether the field can have multiline values. Can be only true when the valueType field is set to STRING.

isTargetDisplayName

- **type:** boolean
- **description:** Whether the field is used as a target display name.

valueType

- **type:** string
- **description:** Type of the value this field holds
- **enum:** ['STRING', 'BOOLEAN', 'NUMERIC', 'INTEGER', 'LIST', 'OBJECT', 'GROUP_SCOPE', 'SECRET']

specificValueType

- **type:** string
- **description:** A specific Type of the value that can't be described in valueType

description

- **type:** string
- **description:** Additional information about what the input to the field should be

verificationRegex

- **type:** string
- **description:** The regex pattern that needs to be satisfied for the input field text

groupProperties

- **type:** array
- **description:** Group scope structure, filled if this field represents group scope value
- – **type:** array
  - **type:** string

allowedValues

- **type:** array
- **description:** Potential values for this field. If nonempty, value must be one of the choices in the list.
- – **type:** string

dependencyKey

- **type:** string
- **description:** A key of a dependency account value field. If this value is not set, field is declared without any dependencies. If an annotated account value field has a dependency it means that it's presence is logically driven by the dependency field. This field is expected to be available ONLY when field specified by {@code dependencyKey} has value {@code dependencyValue}. If this field is not available because of unmet dependency, this field will not be validated and stored in target configuration data. If a dependency for this field is unmet value of this field will be null

dependencyValue

- **type:** string
- **example:** one|two|four
- **description:** Value of a field of account values referenced by {@code dependencyKey} to enable (activate) the annotated field. Value is specified using a regular expression. Specifying just a string will also work correctly. If a referenced account value field is a boolean field, only "true" or "false" are supported as values. If a referenced account value is an enumeration ({@code allowedValues} is specified) this value is only restricted to contain a subset of the allowed values united using a pipe {@code one|two|four}.

InstanceTypeFilterApiDTO

Description

Model to describe instanceType filter information for the action filtering.

Optional Parameters

currentInstanceTypeFilter

- **description:** Filter the actions by matching current instanceType and given the instanceType Uuid list.
- **reference:** InstanceTypeIdsApiDTO *(on page 1573)*

newInstanceTypeFilter

- **description:** Filter the actions by matching new instanceType and given the instanceType Uuid list.
- **reference:** InstanceTypeIdsApiDTO *(on page 1573)*

InstanceTypeIdsApiDTO

Description

Model to describe instanceType Uuid list and operator for the action filtering.

Optional Parameters

instanceTypeIds

- **type:** array
- **description:** InstanceType Uuid list for the action filter.
- – **type:** string

operator

- **type:** string
- **description:** Operator for instanceType Uuid filter.the allowable values are 'EQ' and 'NEQ'
- **enum:** ['EQ', 'NEQ']

IntegerConditionApiDTO

Description

Inherits From

This DTO inherits from ConditionApiDTO *(on page 1535)*

Required Parameters

value

- **type:** integer
- **format:** int64
- **description:** Value to compare against.
- **readOnly:** True

IntegerFieldValueApiDTO

Description

Inherits From

This DTO inherits from FieldValueApiDTO *(on page 1562)*

Required Parameters

value

- **type:** integer
- **format:** int64
- **description:** The value.
- **readOnly:** True

Optional Parameters

units

- **type:** string
- **description:** Units of measurement.
- **readOnly:** True

KubeturboConfigInputDTO

Description

Required Parameters

clusterName

- **type:** string
- **example:** dev-ocp
- **description:** Name of container platform cluster. Will show up as Target Name in the target page.

deployThroughTSC

- **type:** boolean
- **description:** Determines whether to deploy kubeturbo through TSC

targetSubtype

- **type:** string
- **example:** RedHatOpenShift
- **description:** Target subtype

Optional Parameters

namespace

- **type:** string
- **example:** turbo
- **description:** Namespace where kubeturbo will be deployed

permission

- **type:** string
- **description:** Permission for kubeturbo deployed in a cluster
- **enum:** ['OPTIMIZE', 'MONITOR']

privateRegistry

- **type:** string
- **example:** icr.io/cpopen/turbonomic/kubeturbo
- **description:** Private image registry for kubeturbo

privateRegistryUsername

- **type:** string
- **example:** username
- **description:** Private image registry username for kubeturbo

privateRegistryPassword

- **type:** string
- **example:** password
- **description:** Private image registry password for kubeturbo

proxyServerURL

- **type:** string
- **example:** http://proxyserver:proxyport
- **description:** Proxy server address

serverURL

- **type:** string
- **example:** https://Turbo_server_URL
- **description:** Turbo server address

serverVersion

- **type:** string
- **example:** 8.13.5
- **description:** Turbo server version

oauthClientID

- **type:** string
- **description:** OAuth client ID

oauthSecret

- **type:** string
- **description:** OAuth client secret

tscToken

- **type:** string
- **description:** TSC Token

KubeturboDeployOutputDTO

Description

Bash script to deploy kubeturbo

Optional Parameters

script

- **type:** string
- **example:** #!/usr/bin/env bash echo "Sample"
- **description:** Bash script to deploy kubeturbo

LicenseApiDTO

Description

Model to describe properties of a license

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

licenseOwner

- **type:** string
- **description:** The owner of current license

email

- **type:** string
- **description:** The email of current license

customerId

- **type:** string
- **description:** The customer id

expirationDate

- **type:** string
- **description:** The expiration date of current license

features

- **type:** array
- **description:** List of features contained in current license
- **uniqueItems:** True
- – **type:** string

numLicensedEntities

- **type:** integer
- **format:** int32
- **description:** The number of licensed sockets

numInUseEntities

- **type:** integer
- **format:** int32
- **description:** The number of sockets in use

edition

- **type:** string
- **description:** The edition of the current license

filename

- **type:** string
- **description:** The name of the uploaded file

countedEntity

- **type:** string
- **description:** Indicates which entity type is counted with respect to licensing
- **enum:** ['VM', 'SOCKET']

licenseSchemaVersion

- **type:** string
- **description:** License format version

engagementType

- **type:** string
- **description:** Engagement type

companyName

- **type:** string
- **description:** Company name

parentCompanyName

- **type:** string
- **description:** Parent company name

territory

- **type:** string
- **description:** Sales territory

iscId

- **type:** string
- **description:** ISC ID

iscPovId

- **type:** string
- **description:** ISC POV ID

maxReportEditorsCount

- **type:** integer
- **format:** int32
- **description:** Maximum number of report editors

expirationDateTime

- **type:** string
- **format:** date-time

errorReasons

- **type:** array
- **description:** Indicate why the license is invalid
- **uniqueItems:** True
- – **type:** string
  – **enum:** ['DUPLICATE_LICENSE', 'EXPIRED', 'INVALID_EMAIL', 'INVALID_CONTENT_TYPE', 'INVALID_LICENSE_KEY', 'INVALID_LICENSE_TYPE_CWOM_ONLY', 'INVALID_LICENSE_TYPE_FREEMIUM_NOT_ALLOWED', 'INVALID_FEATURE_SET', 'INCOMPATIBLE']

isValid

- **type:** boolean
- **description:** Indicate if current license is valid or not
- **readOnly:** True

isExpired

- **type:** boolean
- **description:** Indicate if current license is expired
- **readOnly:** True

LinkNotificationCriteria

Description

Model to describe criteria information for a notification category.

Optional Parameters

severities

- **type:** array
- **description:** Health state criteria to filter when linking notifications.
- – **type:** string
  – **enum:** ['CRITICAL', 'MAJOR', 'MINOR', 'NORMAL']

LinkNotificationRequestApiDTO

Description

Model to describe criteria for notification link request.

Required Parameters

criteria

- **type:** object
- **description:** Notification criteria for a category
- **additionalProperties:** {'$ref': '#/definitions/LinkNotificationCriteria'}

LinkRelation

Description

This is an unspecified object.

Link

Description

Optional Parameters

rel

- **readOnly:** True
- **reference:** LinkRelation *(on page 1579)*

href

- **type:** string
- **readOnly:** True

hreflang

- **type:** string
- **readOnly:** True

media

- **type:** string
- **readOnly:** True

title

- **type:** string
- **readOnly:** True

type

- **type:** string
- **readOnly:** True

deprecation

- **type:** string
- **readOnly:** True

profile

- **type:** string
- **readOnly:** True

name

- **type:** string
- **readOnly:** True

LoadChangesApiDTO

Description

Workload utilization changes in the scenario

Optional Parameters

baselineDate

- **type:** string
- **description:** Baseline date

peakBaselineList

- **type:** array
- **description:** List of peak baseline changes
- – **reference:** PeakBaselineApiDTO *(on page 1618)*

utilizationList

- **type:** array
- **description:** List of utilization changes
- – **reference:** UtilizationApiDTO *(on page 1709)*

maxUtilizationList

- **type:** array
- **description:** List of max utilization changes
- – **reference:** MaxUtilizationApiDTO *(on page 1586)*

overlayStatsList

- **type:** array
- **description:** List of stat changes to override the plan commodities values with the provided input
- – **reference:** EntityStatsApiDTO *(on page 1558)*

LogEntryApiDTO

Description

Basic information of an entity

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

logActionTime

- **type:** integer
- **format:** int64

targetSE

- **type:** string

targetUuid

- **type:** string

currentSE

- **type:** string

currentSEUuid

- **type:** string

newSE

- **type:** string

newSEUuid

- **type:** string

category

- **type:** string

subCategory

- **type:** string

description

- **type:** string

shortDescription

- **type:** string

state

- **type:** string

severity

- **type:** string

count

- **type:** integer
- **format:** int64

total

- **type:** integer
- **format:** int64

user

- **type:** string

type

- **type:** string

importance

- **type:** number
- **format:** float

reasonCommodities

- **type:** array
- **description:** The distinct set of commodities that were the reason for the action. Not all actions are driven by commodities. So its possible that this can be an empty list.
- **uniqueItems:** True
- – **type:** string

reasonCommodityRelationType

- **type:** string
- **description:** Relation type (bought/sold) of the reason commodity
- **enum:** ['sold', 'bought']

logDBId

- **type:** integer
- **format:** int64

LoggingApiDTO

Description

Model to describe the component logging level

Optional Parameters

componentLoggingLevel

- **type:** object
- **description:** A map from component name to logging level, component could be Discovery, Monitoring,Presentation, Abstraction, Analysis, Extension; logging levels could be INFO, WARN, DEBUG,TRACE
- **additionalProperties:** {'type': 'string', 'enum': ['FATAL', 'ERROR', 'WARN', 'INFO', 'DEBUG', 'TRACE', 'UNKNOWN']}

ManualDynamicConnections

Description

Inherits From

This DTO inherits from IManualConnectionsData *(on page 1570)*

Optional Parameters

dynamicConnectionCriteria

- **type:** array
- **description:** Connection criteria to define dynamically connected entities.
- – **reference:** FilterApiDTO *(on page 1563)*

ManualEntityDefinitionData

Description

Inherits From

This DTO inherits from IEntityDefinitionData *(on page 1570)*

Required Parameters

manualConnectionData

- **type:** object
- **description:** Entity connections set manually
- **additionalProperties:** {'$ref': '#/definitions/IManualConnectionsData'}

ManualGroupConnections

Description

Inherits From

This DTO inherits from IManualConnectionsData *(on page 1570)*

Required Parameters

connectedGroup

- **description:** Group for a connection.
- **reference:** BaseGroupApiDTO *(on page 1510)*

ManualGroupsConnections

Description

Inherits From

This DTO inherits from IManualConnectionsData *(on page 1570)*

Required Parameters

connectedGroups

- **type:** array
- **description:** Groups for a connection.
- – **reference:** BaseGroupApiDTO *(on page 1510)*

ManualStaticConnections

Description

Inherits From

This DTO inherits from IManualConnectionsData *(on page 1570)*

Optional Parameters

staticConnections

- **type:** array
- **description:** Oids to be statically connected to the created entity
- – **type:** string

MarketApiDTO

Description

Model to describe a Market

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

state

- **type:** string
- **description:** State
- **enum:** ['CREATED', 'READY_TO_START', 'RUNNING', 'COPYING', 'SUCCEEDED', 'STOPPING', 'STOPPED', 'DELETING']

stateProgress

- **type:** integer
- **format:** int32
- **description:** State progress indicated by a percentage, used in RUNNING state

scenario

- **description:** Scenario used to create the Plan, only for Plan Market
- **reference:** ScenarioApiDTO *(on page 1659)*

saved

- **type:** boolean
- **description:** Shows if a Plan Market is saved

unplacedEntities

- **type:** boolean
- **description:** Unplaced entities, shows if there are entities that could not be placed in the Plan Market

runDate

- **type:** string
- **description:** Shows when a Plan Market started running

runCompleteDate

- **type:** string
- **description:** Shows when a Plan Market was completed

violations

- **type:** array
- **description:** Violations generated in a Plan Market
- – **reference:** ServiceEntityApiDTO *(on page 1668)*

relatedPlanMarkets

- **type:** array
- **description:** Related Plan Markets generated by particular Scenarios
- – **reference:** MarketApiDTO *(on page 1584)*

environmentType

- **type:** string
- **description:** Environment type
- **enum:** ['HYBRID', 'CLOUD', 'ONPREM']

planDestination

- **description:** Plan destination where the plan results would be stored, only for Plan Market.
- **reference:** BaseApiDTO *(on page 1509)*

MasterImageEntityAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

displayName

- **type:** string
- **description:** DisplayName of the masterImage used to create new VMs

numVcpus

- **type:** integer
- **format:** int32
- **description:** The number of vCPUs allocated to each VM generated from the master image

mem

- **type:** number
- **format:** float
- **description:** The amount of Mem allocated to each VM generated from the master image

storage

- **type:** number
- **format:** float
- **description:** The amount of Storage allocated to each VM generated from the master image

MaxUtilizationApiDTO

Description

Model to describe a max utilization change in the scenario

Optional Parameters

projectionDay

- **type:** integer
- **format:** int32
- **description:** Projection day, indicate when to apply the change

target

- **description:** Target of the change, entity or group. If empty, assumes global scope
- **reference:** BaseApiDTO *(on page 1509)*

maxPercentage

- **type:** integer
- **format:** int32
- **description:** Max utilization threshold percentage change

selectedEntityType

- **type:** string
- **description:** Type of entity to select in global scope configuration

globalScope

- **type:** boolean

MemberFieldApiDTO

Description

Inherits From

This DTO inherits from FieldApiDTO *(on page 1562)*

Required Parameters

property

- **type:** string
- **description:** The property of interest.
- **readOnly:** True
- **enum:** ['COUNT']

Optional Parameters

entityType

- **type:** string
- **description:** Restricts the property to entities of this type.
- **readOnly:** True
- **enum:** ['Application', 'ApplicationComponent', 'ApplicationServer', 'AvailabilityZone', 'BusinessAccount', 'BusinessApplication', 'BusinessTransaction', 'BusinessUser', 'Chassis', 'CloudService', 'ComputeTier', 'Container', 'ContainerPod', 'ContainerSpec', 'Database', 'DatabaseServer', 'DatabaseServerTier', 'DatabaseTier', 'DataCenter', 'DesktopPool', 'DiskArray', 'DPod', 'HCIPhysicalMachine', 'HypervisorServer', 'Internet', 'IOModule', 'LoadBalancer', 'LogicalPool', 'Namespace', 'Network', 'PhysicalMachine', 'ProcessorPool', 'Region', 'ReservedInstance', 'Service', 'ServiceProvider', 'Storage', 'StorageController', 'StorageTier', 'Switch', 'Unknown', 'ViewPod', 'VirtualApplication', 'VirtualDataCenter', 'VirtualMachine', 'VirtualVolume', 'VPod', 'WorkloadController', 'ContainerPlatformCluster', 'CloudCommitment', 'ApplicationComponentSpec', 'VirtualMachineSpec', 'DocumentCollection', 'NodeGroup', 'PowerVMPhysicalMachine', 'DataWarehouse', 'DataWarehouseTier']

direct

- **type:** boolean
- **description:** Whether this field refers to direct or indirect members (default = direct).
- **readOnly:** True

MigrateObjectApiDTO

Description

Model to describe a replace change in the scenario

Optional Parameters

projectionDay

- **type:** integer
- **format:** int32
- **description:** Projection day, indicate when to apply the change

source

- **description:** Source of the migration, entity or group - [DEPRECATED - Use: sources]
- **reference:** BaseApiDTO *(on page 1509)*

destination

- **description:** Destination of the migration - [DEPRECATED - Use: destinations]
- **reference:** BaseApiDTO *(on page 1509)*

sources

- **type:** array
- **description:** Sources of the migration, entities or groups.
- – **reference:** BaseApiDTO *(on page 1509)*

destinations

- **type:** array
- **description:** Destinations of the migration, entities or groups.
- – **reference:** BaseApiDTO *(on page 1509)*

destinationEntityType

- **type:** string
- **description:** Entity type of migration destination, default is VirtualMachine. Case 1, migration source VirtualMachine: destination entity must be VirtualMachine. Case 2, migration source Database Server: destination entity can be VirtualMachine or DatabaseServer.
- **enum:** ['VirtualMachine', 'DatabaseServer']

removeNonMigratingWorkloads

- **type:** boolean
- **description:** If true, remove the existing Virtual Machines, Databases and Database Servers from the scope that are not migrating, so that only the newly migrated entities are included in the plan results.

MultiTextFieldValueApiDTO

Description

Inherits From

This DTO inherits from FieldValueApiDTO *(on page 1562)*

Required Parameters

value

- **type:** array
- **description:** The values associated with the field.
- **readOnly:** True
- – **type:** string

NameValueInputDTO

Description

Deployment Profile provider parameters

Optional Parameters

name

- **type:** string
- **description:** Parameter type. For vCenter, this can be a name or id. For all other target types, this must be an id.
- **enum:** ['name', 'id']

value

- **type:** string

NetworkEntityAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

networkName

- **type:** string
- **description:** Network's name.

clusterName

- **type:** string
- **description:** Network's cluster name. [DEPRECATED - EOL: 8.15.1 - Use: clusterNames]

dataCenterName

- **type:** string
- **description:** Network's data center name.

clusterNames

- **type:** array
- **description:** Clusters that the network is part of.
  - **type:** string

vlanId

- **type:** integer
- **format:** int32

vswitchType

- **type:** string

NoSqlEntityAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

accountThroughputLimit

- **type:** integer
- **format:** int32
- **description:** The throughput capacity limit that is set at account-level.

noSqlApiType

- **type:** string
- **description:** The NoSql engine or api the current entity is using.
- **enum:** ['NOSQL', 'MONGODB', 'CASSANDRA', 'GREMLIN', 'TABLE']

throughputType

- **type:** string
- **description:** Indicates the way of how throughput is allocate to current entity.
- **enum:** ['DEDICATED', 'SHARED', 'NONE']

autoscaling

- **type:** string
- **description:** Indicates whether the entity has autoscale setting enabled.
- **enum:** ['Enabled', 'Disabled', 'Unknown']

maxThroughput

- **type:** integer
- **format:** int32
- **description:** The maximum throughput amount provisioned for current entity.

minThroughput

- **type:** integer
- **format:** int32
- **description:** The minimum throughput amount provisioned for current entity.

effectiveThroughput

- **type:** number
- **format:** double
- **description:** The throughput amount provisioned for current entity, or the amount from host entity or sum of children entities.

databaseName

- **type:** string
- **description:** The associated database name.

NumberConditionApiDTO

Description

Inherits From

This DTO inherits from ConditionApiDTO *(on page 1535)*

Required Parameters

value

- **type:** number
- **format:** double
- **description:** Value to compare against.
- **readOnly:** True

NumberFieldValueApiDTO

Description

Inherits From

This DTO inherits from FieldValueApiDTO *(on page 1562)*

Required Parameters

value

- **type:** number
- **format:** double
- **description:** The value.
- **readOnly:** True

Optional Parameters

units

- **type:** string
- **description:** Units of measurement.
- **readOnly:** True

OAuth2ClientApiDTO

Description

Optional Parameters

clientName

- **type:** string
- **description:** The name of the client

clientId

- **type:** string
- **description:** The ID of the client

grantTypes

- **type:** array
- **description:** The OAuth2 grants that the client can perform in order to acquire a token
- **uniqueItems:** True
- – **type:** string
  – **enum:** ['client_credentials']

clientAuthenticationMethods

- **type:** array
- **description:** The authentication methods that the client supports when acquiring a token. * client_secret_basic: Provide the credentials using the HTTP Basic Authentication Scheme * client_secret_post: Provide the credentials in the HTTP request body as a form parameter
- **uniqueItems:** True
- – **type:** string
  – **enum:** ['client_secret_basic', 'client_secret_post']

scopes

- **type:** array
- **description:** The scopes that the client can request when acquiring a token
- **uniqueItems:** True
- – **type:** string

audience

- **type:** array
- **description:** The audience that the client can request when acquiring a token
- **uniqueItems:** True
- – **type:** string

createdAt

- **type:** integer
- **format:** int64
- **description:** Creation timestamp

tokenSettings

- **reference:** OAuth2ClientOutputTokenSettings *(on page 1594)*

OAuth2ClientInputAccessToken

Description

Optional Parameters

ttlSeconds

- **type:** integer
- **format:** int64
- **example:** 600
- **description:** The lifespan of access tokens acquired by the client in seconds
- **minimum:** 1
- **maximum:** 2592000

OAuth2ClientInputTokenSettings

Description

Optional Parameters

accessToken

- **reference:** OAuth2ClientInputAccessToken *(on page 1593)*

OAuth2ClientOutputAccessToken

Description

Optional Parameters

ttlSeconds

- **type:** integer
- **format:** int64
- **description:** The lifespan of access tokens acquired by the client in seconds

OAuth2ClientOutputTokenSettings

Description

Optional Parameters

accessToken

- **reference:** OAuth2ClientOutputAccessToken *(on page 1594)*

OAuth2CreateClientApiDTO

Description

Optional Parameters

clientName

- **type:** string
- **description:** The name of the client

clientId

- **type:** string
- **description:** The ID of the client

grantTypes

- **type:** array
- **description:** The OAuth2 grants that the client can perform in order to acquire a token
- **uniqueItems:** True
- – **type:** string
  – **enum:** ['client_credentials']

clientAuthenticationMethods

- **type:** array
- **description:** The authentication methods that the client supports when acquiring a token. * client_secret_basic: Provide the credentials using the HTTP Basic Authentication Scheme * client_secret_post: Provide the credentials in the HTTP request body as a form parameter
- **uniqueItems:** True
- – **type:** string
  – **enum:** ['client_secret_basic', 'client_secret_post']

scopes

- **type:** array
- **description:** The scopes that the client can request when acquiring a token
- **uniqueItems:** True
- – **type:** string

audience

- **type:** array
- **description:** The audience that the client can request when acquiring a token
- **uniqueItems:** True
- – **type:** string

createdAt

- **type:** integer
- **format:** int64
- **description:** Creation timestamp

tokenSettings

- **reference:** OAuth2ClientOutputTokenSettings *(on page 1594)*

clientSecret

- **type:** string
- **description:** The client secret

OAuth2CreateClientApiInputDTO

Description

Required Parameters

clientName

- **type:** string
- **description:** The name of the client

grantTypes

- **type:** array
- **description:** The OAuth2 grants that the client can perform in order to acquire a token
- **uniqueItems:** True
- – **type:** string
  – **enum:** ['client_credentials']

scopes

- **type:** array
- **description:** The scopes that the client can request when acquiring a token
- **uniqueItems:** True
- – **type:** string
- **maxItems:** 1
- **minItems:** 1

Optional Parameters

clientAuthenticationMethods

- **type:** array
- **description:** The authentication methods that the client supports when acquiring a token. * client_secret_basic: Provide the credentials using the HTTP Basic Authentication Scheme * client_secret_post: Provide the credentials in the HTTP request body as a form parameter
- **uniqueItems:** True
- – **type:** string
  – **enum:** ['client_secret_basic', 'client_secret_post']

tokenSettings

- **reference:** OAuth2ClientInputTokenSettings *(on page 1593)*

OAuthDataApiDTO

Description

Model to define the oAuth data.

Required Parameters

authorizationServerUrl

- **type:** string
- **description:** The URL of the authorization server.

clientId

- **type:** string
- **description:** The client id used for oAuth authorization.

clientSecret

- **type:** string
- **description:** The client secret used for oAuth authorization.

grantType

- **type:** string
- **description:** The oAuth grant type.
- **enum:** ['CLIENT_CREDENTIALS']

Optional Parameters

scope

- **type:** string
- **description:** The oAuth scope.

OnDemandCostFilterApiDTO

Description

Model to describe on-demand cost range for action filtering.

Optional Parameters

currentOnDemandCostRange

- **description:** Filter the actions by the range for current on-demand cost.
- **reference:** RangeInputApiDTO *(on page 1637)*

newOnDemandCostRange

- **description:** Filter the actions by the range for destination on-demand cost. Applicable only for scale actions.
- **reference:** RangeInputApiDTO *(on page 1637)*

OnPremCostsManagerApiDTO

Description

Model to describe a Category Manager for the on prem cost manager

Required Parameters

category

- **type:** string
- **description:** Category of the Manager

settings

- **type:** array
- **description:** List of Settings contained in the Manager
- – **reference:** SettingApiDTOSerializable *(on page 1671)*

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

OnPremResizeActionDetailsApiDTO

Description

Inherits From

This DTO inherits from ActionDetailsApiDTO *(on page 1492)*

Optional Parameters

vcpuBefore

- **type:** integer
- **format:** int32
- **description:** vCPUs before the change

vcpuAfter

- **type:** integer
- **format:** int32
- **description:** vCPUs after the change

socketsBefore

- **type:** integer
- **format:** int32
- **description:** sockets before the change

socketsAfter

- **type:** integer
- **format:** int32
- **description:** sockets after the change

coresPerSocketBefore

- **type:** integer
- **format:** int32
- **description:** cores per socket before the change

coresPerSocketAfter

- **type:** integer
- **format:** int32
- **description:** cores per socket after the change

nonDisruptiveMaxPU

- **type:** number
- **format:** double
- **description:** Non-disruptive Processing Units max resize constraint

nonDisruptiveMinPU

- **type:** number
- **format:** double
- **description:** Non-disruptive Processing Units min resize constraint

subType

- **type:** string

OrderByApiDTO

Description

A field to order results by.

Required Parameters

field

- **description:** The field to order by.
- **readOnly:** True
- **reference:** FieldApiDTO *(on page 1562)*

Optional Parameters

ascending

- **type:** boolean
- **description:** Whether the order is ascending (defaults to false).
- **readOnly:** True

PMDiskAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

diskCapacity

- **type:** integer
- **format:** int64
- **description:** Disk Capacity

diskRole

- **type:** string
- **description:** Disk Role
- **enum:** ['CAPACITY', 'CACHE']

PMDiskGroupAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

disks

- **type:** array
- **description:** Disks
- – **reference:** PMDiskAspectApiDTO *(on page 1599)*

PMEntityAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

diskGroups

- **type:** array
- **description:** Disk Groups
- – **reference:** PMDiskGroupAspectApiDTO *(on page 1599)*

dedicatedFailoverHost

- **type:** boolean
- **description:** If true, this host is a failover host.

cpuModel

- **type:** string
- **example:** Intel(R) Xeon(R) CPU E5-2670 0 @ 2.60GHz
- **description:** The CPU model of the host

connectedNetworks

- **type:** array
- **example:** Broadcom BCM5708C NetXtreme II GigE
- **description:** Connected Networks
- – **type:** string

automationLevel

- **type:** string
- **example:** Fully Automated
- **description:** Vendor provided automation level
- **enum:** ['FULLY_AUTOMATED', 'PARTIALLY_AUTOMATED', 'NOT_AUTOMATED', 'DISABLED']

migrationLevel

- **type:** string
- **example:** Conservative
- **description:** Vendor provided migration level

machineTypeAndModel

- **type:** string
- **example:** 9040-MR9
- **description:** Machine Type and Model

serialNumber

- **type:** string
- **example:** 130330X
- **description:** Serial Number

totalInstalledPhysicalProcessorUnits

- **type:** integer
- **format:** int32
- **example:** 80
- **description:** Total Installed Physical Processor Units

supportedProcessorCompatibilityModes

- **type:** string
- **example:** POWER7, POWER8, POWER9
- **description:** Supported Processor Compatibility Modes

latestSupportedProcessorGeneration

- **type:** string
- **example:** 9
- **description:** Latest Supported Processor Generation

totalInstalledPhysicalMemory

- **type:** integer
- **format:** int32
- **example:** 4096
- **description:** Total Installed Physical Memory

hypervisorUsedMemory

- **type:** integer
- **format:** int32
- **example:** 4096
- **description:** Hypervisor Used Memory

memoryAssignedToVMs

- **type:** integer
- **format:** int32
- **example:** 4096
- **description:** Memory Assigned To VMs

logicalMemoryBlockSize

- **type:** integer
- **format:** int32
- **example:** 256
- **description:** Logical Memory Block Size

supportedVirtualGpuTypes

- **type:** string
- **example:** 1q, 2q, 4q, 8q, 16q
- **description:** Comma separated list of Supported vGPU Types

installedGpuDevices

- **type:** string
- **example:** T4, A16
- **description:** Comma separated list of Installed GPU Devices

hardwareVersion

- **type:** string
- **example:** 19.0
- **description:** Hardware Compatibility Version

processorPools

- **type:** array
  - **type:** string

PaginationApiDTO

Description

Pagination parameters.

Optional Parameters

orderBy

- **type:** array
- **description:** Order by these fields.
- **readOnly:** True
  - **reference:** OrderByApiDTO *(on page 1599)*

cursor

- **type:** string
- **description:** Cursor for the next batch.
- **readOnly:** True

limit

- **type:** integer
- **format:** int32
- **description:** Maximum number of records to fetch.
- **readOnly:** True

ParkableEntityApiDTO

Description

Parkable Entity

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

environmentType

- **type:** string
- **example:** ONPREM
- **description:** On-Prem environment type
- **enum:** ['HYBRID', 'CLOUD', 'ONPREM', 'UNKNOWN']

accountOID

- **type:** integer
- **format:** int64
- **example:** 74613608411667
- **description:** Unique ID of the account

accountName

- **type:** string
- **example:** dev-account
- **description:** Display name of the account

provider

- **type:** string
- **example:** AWS
- **description:** Name of the cloud provider
- **enum:** ['AWS', 'AZURE', 'GCP', 'HYBRID', 'UNKNOWN']

state

- **type:** string
- **example:** STOPPED
- **description:** Last known state of the entity
- **enum:** ['RUNNING', 'STOPPED', 'SUSPENDED', 'MAINTENANCE', 'FAILOVER', 'UNKNOWN', 'STARTING', 'STOPPING']

parkable

- **type:** boolean
- **example:** True
- **description:** Signifies whether the entity can be parked. [DEPRECATED - EOL: 8.15.1 - Use: restrictions]

entityType

- **type:** string
- **example:** COMPUTE
- **description:** Type of the entity
- **enum:** ['VirtualMachine', 'DatabaseServer', 'Database', 'VirtualMachineSpec']

regionOID

- **type:** integer
- **format:** int64
- **example:** 74613608411668
- **description:** Unique ID of the region, the entity belongs to

regionName

- **type:** string
- **example:** us-east-1
- **description:** Name of the region

instanceType

- **type:** string
- **example:** t3.micro
- **description:** Type of the instance

targetOID

- **type:** integer
- **format:** int64
- **example:** 746136084116679
- **description:** Unique ID of the target

datacenterName

- **type:** string
- **example:** VC1
- **description:** Display name of the Data Center

cost

- **type:** number
- **format:** double
- **example:** 10.24
- **description:** On demand hourly cost of the entity

tags

- **type:** array
- **description:** List of tags attached to the entity
- – **reference:** TagApiDTO *(on page 1690)*

schedules

- **type:** array
- **description:** List of schedules attached to the entity
- – **reference:** ParkableEntityScheduleApiDTO *(on page 1611)*

groups

- **type:** array
- **description:** List of the Groups attached to the entity
- – **reference:** ParkableEntityGroupApiDTO *(on page 1609)*

guestOSName

- **type:** string
- **example:** Linux
- **description:** Guest OS Name of the entity

restrictions

- **type:** array
- **description:** List of the restrictions placed on the entity
- – **reference:** ParkableEntityRestrictionApiDTO *(on page 1611)*

controlledBy

- **type:** integer
- **format:** int64
- **example:** 74613608411668
- **description:** Unique ID of the entity that this entity is controlled by. For example, if this is a Virtual Machine that is part of a scale group, this would be the ID of the scale group

smartParkingRecommendation

- **description:** [BETA]: Smart Parking Schedule Recommendation for the entity
- **reference:** ParkableEntitySmartParkingRecommendationApiDTO *(on page 1611)*

schedulePausedUntil

- **type:** integer
- **format:** int64
- **example:** 1673273639
- **description:** The time until which the schedule is paused only applicable if entity has parking schedule, in epoch time format

cloudServiceName

- **type:** string
- **description:** Cloud service name is the resource bought from the cloud service provider
- **enum:** ['AMAZON_RDS', 'AMAZON_EC2', 'AMAZON_EC2_AUTO_SCALING', 'AZURE_APP_SERVICE', 'AZURE_COSMOS_DB', 'AZURE_DISK_STORAGE', 'AZURE_SQL_DATABASE', 'AZURE_SQL_MANAGED_INSTANCE', 'AZURE_SYNAPSE_SQL_POOL', 'AZURE_VIRTUAL_MACHINES', 'AZURE_VIRTUAL_MACHINE_SCALE_SETS', 'GCP_DISK_STORAGE', 'GCP_VIRTUAL_MACHINES']

customInstanceType

- **type:** string
- **example:** n2d-custom-2-1024
- **description:** The entity's custom instance type. Similar to instanceType but provides customization details about the type of instance the workload is based on

ParkableEntityDetailsApiDTO

Description

Details about a Parkable Entity. This includes the summary information from /parking/entities API along with additional details specific to the entity.

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

environmentType

- **type:** string
- **example:** ONPREM
- **description:** On-Prem environment type
- **enum:** ['HYBRID', 'CLOUD', 'ONPREM', 'UNKNOWN']

accountOID

- **type:** integer
- **format:** int64
- **example:** 74613608411667
- **description:** Unique ID of the account

accountName

- **type:** string
- **example:** dev-account
- **description:** Display name of the account

provider

- **type:** string
- **example:** AWS
- **description:** Name of the cloud provider
- **enum:** ['AWS', 'AZURE', 'GCP', 'HYBRID', 'UNKNOWN']

state

- **type:** string
- **example:** STOPPED
- **description:** Last known state of the entity
- **enum:** ['RUNNING', 'STOPPED', 'SUSPENDED', 'MAINTENANCE', 'FAILOVER', 'UNKNOWN', 'STARTING', 'STOPPING']

parkable

- **type:** boolean
- **example:** True
- **description:** Signifies whether the entity can be parked. [DEPRECATED - EOL: 8.15.1 - Use: restrictions]

entityType

- **type:** string
- **example:** COMPUTE
- **description:** Type of the entity
- **enum:** ['VirtualMachine', 'DatabaseServer', 'Database', 'VirtualMachineSpec']

regionOID

- **type:** integer
- **format:** int64
- **example:** 74613608411668
- **description:** Unique ID of the region, the entity belongs to

regionName

- **type:** string
- **example:** us-east-1
- **description:** Name of the region

instanceType

- **type:** string
- **example:** t3.micro
- **description:** Type of the instance

targetOID

- **type:** integer
- **format:** int64
- **example:** 746136084116679
- **description:** Unique ID of the target

datacenterName

- **type:** string
- **example:** VC1
- **description:** Display name of the Data Center

cost

- **type:** number
- **format:** double
- **example:** 10.24
- **description:** On demand hourly cost of the entity

tags

- **type:** array
- **description:** List of tags attached to the entity
- – **reference:** TagApiDTO *(on page 1690)*

schedules

- **type:** array
- **description:** List of schedules attached to the entity
- – **reference:** ParkableEntityScheduleApiDTO *(on page 1611)*

groups

- **type:** array
- **description:** List of the Groups attached to the entity
- – **reference:** ParkableEntityGroupApiDTO *(on page 1609)*

guestOSName

- **type:** string
- **example:** Linux
- **description:** Guest OS Name of the entity

restrictions

- **type:** array
- **description:** List of the restrictions placed on the entity
- – **reference:** ParkableEntityRestrictionApiDTO *(on page 1611)*

controlledBy

- **type:** integer
- **format:** int64
- **example:** 74613608411668
- **description:** Unique ID of the entity that this entity is controlled by. For example, if this is a Virtual Machine that is part of a scale group, this would be the ID of the scale group

smartParkingRecommendation

- **description:** [BETA]: Smart Parking Schedule Recommendation for the entity
- **reference:** ParkableEntitySmartParkingRecommendationApiDTO *(on page 1611)*

schedulePausedUntil

- **type:** integer
- **format:** int64
- **example:** 1673273639
- **description:** The time until which the schedule is paused only applicable if entity has parking schedule, in epoch time format

cloudServiceName

- **type:** string
- **description:** Cloud service name is the resource bought from the cloud service provider
- **enum:** ['AMAZON_RDS', 'AMAZON_EC2', 'AMAZON_EC2_AUTO_SCALING', 'AZURE_APP_SERVICE', 'AZURE_COSMOS_DB', 'AZURE_DISK_STORAGE', 'AZURE_SQL_DATABASE', 'AZURE_SQL_MANAGED_INSTANCE', 'AZURE_SYNAPSE_SQL_POOL', 'AZURE_VIRTUAL_MACHINES', 'AZURE_VIRTUAL_MACHINE_SCALE_SETS', 'GCP_DISK_STORAGE', 'GCP_VIRTUAL_MACHINES']

customInstanceType

- **type:** string
- **example:** n2d-custom-2-1024
- **description:** The entity's custom instance type. Similar to instanceType but provides customization details about the type of instance the workload is based on

virtualMachineSpecDetails

- **description:** Additional details about the entity if the ParkingEntityType is VirtualMachineSpec
- **reference:** ParkableVirtualMachineSpecApiDTO *(on page 1613)*

ParkableEntityEnforcementPauseApiDTO

Description

represents the partial failure per entity for enforcement pause on list of parkable entities

Optional Parameters

error

- **type:** string
- **example:** entityUuid not found
- **description:** error indicating the reason for failure

entityUUID

- **type:** string

ParkableEntityEnforcementPauseInputDTO

Description

represents the enforcement pause on list of parkable entities either until a given time or in duration of minutes

Optional Parameters

until

- **type:** string
- **example:** 2022-11-29T01:02:03+00:00
- **description:** Time when the enforcement pause on parkable entities ends. Should be in ISO 8601 format

durationMinutes

- **type:** integer
- **format:** int32
- **example:** 60
- **description:** duration in minutes to keep enforcement pause on parkable entities

entityUuids

- **type:** array
- **description:** list of parkable entities to pause enforcement on
  - **type:** string

ParkableEntityGroupApiDTO

Description

represents groups which are attached to the parkable entity

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

groupType

- **type:** string
- **example:** Resource
- **description:** Type of the Group
- **enum:** ['RESOURCE']

ParkableEntityInputDTO

Description

Optional Parameters

entityTypes

- **type:** array
  - **type:** string
  - **enum:** ['VirtualMachine', 'DatabaseServer', 'Database', 'VirtualMachineSpec']

status

- **type:** array
  - **type:** string
  - **enum:** ['RUNNING', 'STOPPED', 'SUSPENDED', 'MAINTENANCE', 'FAILOVER', 'UNKNOWN', 'STARTING', 'STOPPING']

providers

- **type:** array
- **description:** Parking Cloud Providers
  - **type:** string
  - **enum:** ['AWS', 'AZURE', 'GCP', 'HYBRID', 'UNKNOWN']

parkableState

- **type:** string
- **description:** Parkable state
- **enum:** ['UNPARKABLE', 'PARKABLE']

criteriaList

- **type:** array
  - **reference:** FilterApiDTO *(on page 1563)*

logicalOperator

- **type:** string
- **description:** Logical operator to be applied across all the criteria. It can take values 'AND', 'OR'. If omitted, defaults to 'AND'
- **enum:** ['AND', 'OR']

ParkableEntityRestrictionApiDTO

Description

Parent class for restrictions details. Implementing classes include: ParkableEntityUnparkableRestrictionApiDTO.

Required Parameters

restrictionType

- **type:** string
- **enum:** ['UNPARKABLE']

ParkableEntityScheduleApiDTO

Description

Represents time span based schedule display name and uuid of the time span based schedule attached to the parkable entity

Optional Parameters

displayName

- **type:** string
- **example:** Up 7-5 on weekdays
- **description:** Display name of the time span based schedule the parkable entities is attached to

uuid

- **type:** string
- **example:** 74613608411667
- **description:** UUID of the time span based schedule the parkable entities are attached to

pauseEnforcementUntil

- **type:** integer
- **format:** int64
- **example:** 1673273639
- **description:** Pause Enforcement time over the entity, in epoch time format

attachedBy

- **description:** Details of the user/policy responsible for attaching the schedule to entity
- **reference:** ParkingEntitySchedulerApiDTO *(on page 1615)*

source

- **type:** string
- **description:** [BETA]: Time span schedule source
- **enum:** ['USER', 'SMART_PARK_RECOMMENDATION']

ParkableEntitySmartParkingRecommendationApiDTO

Description

[BETA]: Smart parking recommendation details

Optional Parameters

uuid

- **type:** string
- **description:** [BETA]: Smart Parking schedule uuid

displayName

- **type:** string
- **description:** [BETA]: Smart Parking schedule name

ParkableEntityStatsInputDTO

Description

Required Parameters

groupBy

- **type:** array
- **description:** List of groupBys to filter
- – **type:** string
  – **enum:** ['ENTITY_TYPE', 'CSP']

ParkableEntityUUIDSetDTO

Description

list of parkable entity uuids

Required Parameters

entityUuids

- **type:** array
- **description:** list of parkable entity uuids
- – **type:** string

ParkableEntityUnparkableRestrictionApiDTO

Description

Inherits From

This DTO inherits from ParkableEntityRestrictionApiDTO *(on page 1611)*

Optional Parameters

attachedBy

- **description:** Details of the user/policy responsible for adding the unparkable restriction
- **reference:** ParkingEntitySchedulerApiDTO *(on page 1615)*

ParkableVirtualMachineSpecApiDTO

Description

Details about a parkable Virtual Machine Spec entity such as a cloud scale group.A cloud scale group is is a collection of virtual machine (VM) instances that are created from a common source. Ex: AWS Auto Scaling Groups, Azure Virtual Machine Scale Sets.

Optional Parameters

instanceTemplateName

- **type:** string
- **description:** Name of template or configuration used to create the scale group. Ex: Name of Launch Template or Launch Configuration in AWS, Auto Scale Profile in Azure or the Instance Template name in Google."

currentConfiguration

- **description:** The current minimum group size, desired group size and maximum group size for a scale group
- **reference:** ParkableVirtualMachineSpecConfigApiDTO *(on page 1613)*

startConfiguration

- **description:** The minimum group size, desired group size and maximum group size that will be set for the scale group when a START action is executed
- **reference:** ParkableVirtualMachineSpecConfigApiDTO *(on page 1613)*

instances

- **type:** array
- **description:** The virtual machine instances in a cloud scale group
- – **reference:** ParkableVirtualMachineSpecInstanceApiDTO *(on page 1614)*

scalingMode

- **type:** string
- **description:** The scaling mode of a cloud scale group
- **enum:** ['AZURE_AUTOSCALE_METRIC_BASED', 'AZURE_AUTOSCALE_SPECIFIC_INSTANCE_COUNT', 'AZURE_MANUAL_SCALE']

ParkableVirtualMachineSpecConfigApiDTO

Description

The minimum group size, desired group size and maximum group size for a scale group

Optional Parameters

minSize

- **type:** integer
- **format:** int32
- **description:** Minimum group size

maxSize

- **type:** integer
- **format:** int32
- **description:** Maximum group size

desiredSize

- **type:** integer
- **format:** int32
- **description:** Default size of group when no provider-initiated scaling actions are applied

ParkableVirtualMachineSpecInstanceApiDTO

Description

Details about a virtual machine instance within a parkable Virtual Machine Spec entity

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

instanceType

- **type:** string
- **example:** t3.micro
- **description:** Type of the instance

state

- **type:** string
- **example:** STOPPED
- **description:** Last known state of the instance
- **enum:** ['RUNNING', 'STOPPED', 'SUSPENDED', 'MAINTENANCE', 'FAILOVER', 'UNKNOWN', 'STARTING', 'STOPPING']

tags

- **type:** array
- **description:** List of tags attached to the instance
- – **reference:** TagApiDTO *(on page 1690)*

availabilityZoneName

- **type:** string
- **example:** aws-us-west-2a
- **description:** Name of the availability zone the instance belongs to

ParkingEntitySchedulerApiDTO

Description

Represents the details of the user/policy responsible for attaching the restriction/schedule to parkable entity

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

type

- **type:** string
- **example:** POLICY
- **description:** Whether user or policy was responsible for attaching the restriction/schedule
- **enum:** ['USER', 'POLICY']

ParkingItemApiDTO

Description

Inherits From

This DTO inherits from ScheduleItemApiDTO *(on page 1662)*

Required Parameters

state

- **type:** string
- **description:** expected state during the time span
- **enum:** ['IGNORE', 'ON', 'OFF']

ParkingPolicyApiDTO

Description

Model to describe a parking policy, used to assign timespan based schedules to entities based on global, cloud provider, account and resource group level filters

Required Parameters

impacts

- **description:** Impacts to be taken by the policy.
- **reference:** ParkingPolicyImpactsApiDTO *(on page 1616)*

level

- **type:** string
- **description:** Specifies the level at which the policy is applicable.
- **enum:** ['GLOBAL', 'CLOUD_PROVIDER', 'ACCOUNT', 'RESOURCE_GROUP']

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

scopes

- **type:** array
- **description:** List of scope object to which the parking policy is applicable, like details of account for account level policies.
- – **reference:** BaseApiDTO *(on page 1509)*

criteriaList

- **type:** array
- **description:** List of filters to get the desired the entities to which the policy is applicable.
- – **reference:** FilterApiDTO *(on page 1563)*

priority

- **type:** integer
- **format:** int32
- **description:** Priority is used to identify the applicable priority if there are multiple policies at same level

ParkingPolicyImpactsApiDTO

Description

Impacts to be taken by the policy.

Optional Parameters

attachSchedule

- **description:** time span based schedule to be attached by policy.
- **reference:** ParkingPolicyTimeSpansApiDTO *(on page 1617)*

restrictUnparkable

- **type:** boolean
- **description:** to mark unparkable by policy.

ParkingPolicyStatsInputDTO

Description

Required Parameters

groupBy

- **type:** array
- **description:** List of groupBys to filter
- – **type:** string
  – **enum:** ['LEVEL']

ParkingPolicyTimeSpansApiDTO

Description

Parking Policy time span schedule

Optional Parameters

timespanScheduleUuid

- **type:** string
- **description:** timespanScheduleUuid used to uniquely recognize the time span Schedule.

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

PatchedTargetDataApiDTO

Description

Model to describe a patched target data.

Required Parameters

patchedFields

- **type:** array
- **description:** Patched target fields.
- – **reference:** PatchedTargetFieldApiDTO *(on page 1618)*

probeType

- **type:** string
- **description:** Probe type of the patched target.

PatchedTargetFieldApiDTO

Description

Model to describe a field in patched target data.

Optional Parameters

fieldName

- **type:** string
- **description:** Patched field name.

fieldValue

- **type:** string
- **description:** Patched field value.

PeakBaselineApiDTO

Description

Model to describe a peak baseline change in the scenario

Optional Parameters

date

- **type:** string
- **description:** Date of the utilization peak

target

- **description:** Target of the change
- **reference:** BaseApiDTO *(on page 1509)*

PlacementInfoDTO

Description

Optional Parameters

computeResources

- **type:** array
- **description:** Compute resources that the reservation is placed in
- – **reference:** ResourceApiDTO *(on page 1651)*

storageResources

- **type:** array
- **description:** Storage resources that the reservation is placed in
- – **reference:** ResourceApiDTO *(on page 1651)*

networkResources

- **type:** array
- **description:** Network resources that the reservation is placed in
- – **reference:** ResourceApiDTO *(on page 1651)*

failureInfos

- **type:** array
- **description:** Details about why the reservation failed. This will only be populated if the reservation failed.
- – **reference:** ReservationFailureInfoDTO *(on page 1646)*

invalidInfo

- **description:** Details about why the reservation is invalid. This will only be populated if the reservation is invalid.
- **reference:** ReservationInvalidInfoApiDTO *(on page 1647)*

PlacementOptionApiDTO

Description

Model to describe the property of provider options.

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

constraintType

- **type:** string
- **description:** constraint type, if a commodity is AccessCommodity, then the class name of that commodity is a constraint type
- **enum:** ['ApplicationCommodity', 'ClusterCommodity', 'DataCenterCommodity', 'DatastoreCommodity etc']

scope

- **description:** scope or placement policy
- **reference:** BaseApiDTO *(on page 1509)*

target

- **description:** for placement policy, if it's an user placement policy is null.
- **reference:** TargetApiDTO *(on page 1690)*

numPotentialEntities

- **type:** integer
- **format:** int32
- **description:** example: other VMs, which buys this constraint but not on this host

key

- **type:** string
- **description:** commodity key, this information is needed when trying to retrieve the intersaction of providers or consumers

currentEntities

- **type:** array
- **description:** for consumers only. example: VMs, which are currently on this host per constraint they are buying
- – **reference:** ServiceEntityApiDTO *(on page 1668)*

PlacementOptionApiInputDTO

Description

Model to describe the property of provider options.

Optional Parameters

constraintType

- **type:** string
- **description:** constraint type, if a commodity is AccessCommodity, then the class name of that commodity is a constraint type
- **enum:** ['ApplicationCommodity', 'ClusterCommodity', 'DataCenterCommodity', 'DatastoreCommodity etc']

key

- **type:** string
- **description:** commodity key, this information is needed when trying to retrieve the intersection of providers or consumers

PlacementParametersDTO

Description

Required Parameters

count

- **type:** integer
- **format:** int32
- **description:** Number of Virtual Machines to Place

Optional Parameters

templateID

- **type:** string
- **description:** Template Uuid used for the Placement

entityNames

- **type:** array
- **description:** Name of the Placed Virtual Machine
- – **type:** string

constraintIDs

- **type:** array
- **description:** List of uuids that will define the constraints that new VMs should be compliant to, allowed uuids are: Cluster, Datacenter, VirtualDataCenter, Network, Placement Policy. The possible providers are the intersection of all constraints passed.
- **uniqueItems:** True
- – **type:** string

geographicRedundancy

- **type:** boolean
- **description:** Enable Geographic Redundancy

PlanDestinationApiDTO

Description

Model to describe a plan destination. A plan destination is an external destination where plan results can be uploaded.

Required Parameters

businessUnit

- **description:** Business account to plan destination.
- **reference:** BaseApiDTO *(on page 1509)*

exportDescription

- **type:** string
- **description:** Plan export description message on this plan destination.

exportProgressPercentage

- **type:** integer
- **format:** int32
- **description:** Plan export progress percentage on this plan destination.

exportState

- **type:** string
- **description:** Plan export state on this plan destination.
- **enum:** ['NONE', 'REJECTED', 'IN_PROGRESS', 'SUCCEEDED', 'FAILED']

hasExportedData

- **type:** boolean
- **description:** Boolean, true if this plan destination has uploaded data.

market

- **description:** Plan market currently uploaded to plan destination.
- **reference:** BaseApiDTO *(on page 1509)*

numErrors

- **type:** integer
- **format:** int32
- **description:** Number of errors on the plan destination.

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

PolicyApiDTO

Description

Model to describe a Workload policy, used to manage the entities in the Environment. e.g. Restrict the number of VMs that can use a given datastore, Limit VM mobility by restricting a collection of VMs to a specified cluster or group of hosts, Expand VM mobility by merging a number of clusters, Keep workload running on hosts that have the required licenses

Required Parameters

type

- **type:** string
- **description:** Placement Policy Type
- **enum:** ['AT_MOST_N', 'BIND_TO_GROUP', 'BIND_TO_COMPLEMENTARY_GROUP', 'MUST_RUN_TOGETHER', 'AT_MOST_N_BOUND', 'MERGE', 'BIND_TO_GROUP_AND_LICENSE', 'MUST_NOT_RUN_TOGETHER', 'EXCLUSIVE_BIND_TO_GROUP']

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

name

- **type:** string
- **description:** Name of the policy

enabled

- **type:** boolean
- **description:** Shows if the policy is enabled

capacity

- **type:** integer
- **format:** int32
- **description:** Indicate the max workload entities at the same time per placement entites, used when type is AT_MOST_N_BOUND

commodityType

- **type:** string
- **description:** Commodity type used for the policy

mergeType

- **type:** string
- **description:** Type of Group that is merged, used only when type is MERGE
- **enum:** ['Cluster', 'StorageCluster', 'DataCenter', 'DesktopPool']

consumerGroup

- **description:** Consumer Group, used when type is any but MERGE
- **reference:** BaseApiDTO *(on page 1509)*

providerGroup

- **description:** Provider Group, used when type is any but MERGE
- **reference:** BaseApiDTO *(on page 1509)*

mergeGroups

- **type:** array
- **description:** List of Groups to merge, used when type is MERGE
- – **reference:** BaseApiDTO *(on page 1509)*

providerEntityType

- **type:** string
- **description:** The entity type for the provider. This property is only applicable when policy type is set to MUST_RUN_TOGETHER or MUST_NOT_RUN_TOGETHER. When policy type is set to MUST_RUN_TOGETHER or MUST_NOT_RUN_TOGETHER, this field should be set with type of provider where members of consumer group must run or must not run together on it depending on the policy type.
- **enum:** ['Application', 'ApplicationComponent', 'ApplicationServer', 'AvailabilityZone', 'BusinessAccount', 'BusinessApplication', 'BusinessTransaction', 'BusinessUser', 'Chassis', 'CloudService', 'ComputeTier', 'Container', 'ContainerPod', 'ContainerSpec', 'Database', 'DatabaseServer', 'DatabaseServerTier', 'DatabaseTier', 'DataCenter', 'DesktopPool', 'DiskArray', 'DPod', 'HCIPhysicalMachine', 'HypervisorServer', 'Internet', 'IOModule', 'LoadBalancer', 'LogicalPool', 'Namespace', 'Network', 'PhysicalMachine', 'ProcessorPool', 'Region', 'ReservedInstance', 'Service', 'ServiceProvider', 'Storage', 'StorageController', 'StorageTier', 'Switch', 'Unknown', 'ViewPod', 'VirtualApplication', 'VirtualDataCenter', 'VirtualMachine', 'VirtualVolume', 'VPod', 'WorkloadController', 'ContainerPlatformCluster', 'CloudCommitment', 'ApplicationComponentSpec', 'VirtualMachineSpec', 'DocumentCollection', 'NodeGroup', 'PowerVMPhysicalMachine', 'DataWarehouse', 'DataWarehouseTier']

enableCreateResourcePool

- **type:** boolean
- **description:** Enable create resource pool during cross VM moves, used when type is any but MERGE.When enabled resource pool, of which VM is part of at source , is created in destination if it does not exists in destination during a cross cluster or vcenter VM move.By default this is disabled(false)

PolicyApiInputDTO

Description

Model to describe a Placement Policy

Required Parameters

policyName

- **type:** string
- **description:** Name of the policy

type

- **type:** string
- **description:** Placement Policy Type
- **enum:** ['AT_MOST_N', 'BIND_TO_GROUP', 'BIND_TO_COMPLEMENTARY_GROUP', 'MUST_RUN_TOGETHER', 'AT_MOST_N_BOUND', 'MERGE', 'BIND_TO_GROUP_AND_LICENSE', 'MUST_NOT_RUN_TOGETHER', 'EXCLUSIVE_BIND_TO_GROUP']

Optional Parameters

sellerUuid

- **type:** string
- **description:** Uuid of the provider Group in a Placement Policy

buyerUuid

- **type:** string
- **description:** Uuid of the consumer Group in a Placement Policy

mergeUuids

- **type:** array
- **description:** Uuids of the Groups in a Merge Policy
- – **type:** string

mergeType

- **type:** string
- **description:** Group Type in a Merge Policy
- **enum:** ['Cluster', 'StorageCluster', 'DataCenter', 'DesktopPool']

capacity

- **type:** integer
- **format:** int32
- **description:** For an AT_MOST_N or AT_MOST_N_BOUND policy, the number of consumers to allow on a provider entity

enableCreateResourcePool

- **type:** boolean
- **description:** Enable create resource pool during cross VM moves, used when type is any but MERGE.When enabled resource pool, of which VM is part of at source , is created in destination if it does not exists in destination during a cross cluster or vcenter VM move.By default this is disabled(false)

enabled

- **type:** boolean
- **description:** Whether to enable the policy

providerEntityType

- **type:** string
- **description:** The entity type for the provider. This property is only applicable when the policy type is set to MUST_RUN_TOGETHER or MUST_NOT_RUN_TOGETHER. This field is set with the provider type (for example, PhysicalMachine) the policy should respect.
- **enum:** ['Application', 'ApplicationComponent', 'ApplicationServer', 'AvailabilityZone', 'BusinessAccount', 'BusinessApplication', 'BusinessTransaction', 'BusinessUser', 'Chassis', 'CloudService', 'ComputeTier', 'Container', 'ContainerPod', 'ContainerSpec', 'Database', 'DatabaseServer', 'DatabaseServerTier', 'DatabaseTier', 'DataCenter', 'DesktopPool', 'DiskArray', 'DPod', 'HCIPhysicalMachine', 'HypervisorServer', 'Internet', 'IOModule', 'LoadBalancer', 'LogicalPool', 'Namespace', 'Network', 'PhysicalMachine', 'ProcessorPool', 'Region', 'ReservedInstance', 'Service', 'ServiceProvider', 'Storage', 'StorageController', 'StorageTier', 'Switch', 'Unknown', 'ViewPod', 'VirtualApplication', 'VirtualDataCenter', 'VirtualMachine', 'VirtualVolume', 'VPod', 'WorkloadController', 'ContainerPlatformCluster', 'CloudCommitment', 'ApplicationComponentSpec', 'VirtualMachineSpec', 'DocumentCollection', 'NodeGroup', 'PowerVMPhysicalMachine', 'DataWarehouse', 'DataWarehouseTier']

PortChannelApiDTO

Description

Model to describe a Port Channel

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

capacity

- **description:** Capacity values
- **reference:** StatValueApiDTO *(on page 1685)*

values

- **description:** Contains Average, Min, Max, Total
- **reference:** StatValueApiDTO *(on page 1685)*

units

- **type:** string
- **description:** Units, used for Commodities stats. E.G. $/h

ports

- **type:** array
- **description:** List of ports in the channel
- – **reference:** StatApiDTO *(on page 1679)*

filters

- **type:** array
- **description:** describe the grouping options used to generate the output
- – **reference:** StatFilterApiDTO *(on page 1682)*

PortsAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

portChannels

- **type:** array
- **description:** Port Channels
- – **reference:** PortChannelApiDTO *(on page 1625)*

ports

- **type:** array
- **description:** Independent ports
- – **reference:** StatApiDTO *(on page 1679)*

PowerVmPMEntityAspectDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

activeMemoryExpansionCapable

- **type:** boolean
- **example:** True
- **description:** Declares if PM has active memory expansion

lparSuspendCapable

- **type:** boolean
- **example:** True
- **description:** Declares if PM can suspend VMs

ibmiLparSuspendCapable

- **type:** boolean
- **example:** True
- **description:** Declares if PM can suspend Ibmi VMs

trustedPlatformModuleCapable

- **type:** boolean
- **example:** True
- **description:** Declares if PM has trusted boot

remoteRestartCapable

- **type:** boolean
- **example:** True
- **description:** Declares if PM has remote restart

secureBootCapable

- **type:** boolean
- **example:** True
- **description:** Declares if PM has secure boot

hasMspEnabledVios

- **type:** boolean
- **example:** True
- **description:** Declares if PM has any MSP enabled VIOS

PowerVmVMEntityAspectDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

activeMemoryExpansionEnabled

- **type:** boolean
- **example:** True
- **description:** Declares if VM has active memory expansion enabled

lparSuspendCapable

- **type:** boolean
- **example:** True
- **description:** Declares if VM can suspend

ibmiLparSuspendCapable

- **type:** boolean
- **example:** True
- **description:** Declares if Ibmi VM can suspend

trustedPlatformModuleCapable

- **type:** boolean
- **example:** True
- **description:** Declares if VM has trusted boot

remoteRestartCapable

- **type:** boolean
- **example:** True
- **description:** Declares if VM host has remote restart

secureBootCapable

- **type:** boolean
- **example:** True
- **description:** Declares if VM has secure boot

redundantErrorPathReportingEnabled

- **type:** boolean
- **example:** True
- **description:** Declares if VM has redundant error path reporting enabled

hasLastActivatedProfile

- **type:** boolean
- **example:** True
- **description:** Declares if VM has last activated profile

hasPhysicallo

- **type:** boolean
- **example:** True
- **description:** Declares if VM has physical adapters

hasHugePageSizesConfigured

- **type:** boolean
- **example:** True
- **description:** Declares if VM has huge page sizes configured

PriceAdjustmentDTO

Description

Optional Parameters

type

- **type:** string
- **enum:** ['DISCOUNT', 'INCREASE']

value

- **type:** number
- **format:** float

PrimitiveFieldApiDTO

Description

Inherits From

This DTO inherits from FieldApiDTO *(on page 1562)*

Required Parameters

fieldName

- **type:** string
- **description:** The name of the primitive field.
- **readOnly:** True

ProbeApiDTO

Description

Model to describe a probe.

Required Parameters

category

- **type:** string
- **description:** Probe category.

id

- **type:** integer
- **format:** int64
- **description:** Probe id.

type

- **type:** string
- **description:** Probe type.

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

error

- **type:** string
- **description:** Unexpected error while collecting probe information

ProbeDeploymentApiDTO

Description

Optional Parameters

name

- **type:** string
- **description:** Name of the Probe

version

- **type:** string
- **description:** Version of the Probe

image

- **type:** string
- **description:** Probe image path

imageHash

- **type:** string
- **description:** Probe image hash

ProbePropertyApiDTO

Description

Model to describe a probe property

Required Parameters

name

- **type:** string
- **description:** Probe property name.

value

- **type:** string
- **description:** Probe property value.

Optional Parameters

probeId

- **type:** integer
- **format:** int64
- **description:** Probe id.

targetId

- **type:** integer
- **format:** int64
- **description:** Target id.

ProbePropertyNameValuePairApiDTO

Description

Model to describe a name/value pair for probe properties.

Required Parameters

name

- **type:** string
- **description:** Probe property name.

value

- **type:** string
- **description:** Probe property value.

ProductCapabilityDTO

Description

Model to describe Product Capability information.

Optional Parameters

deploymentMode

- **type:** string
- **description:** Deployment mode, affects hiding / showing multiple UI views, options
- **enum:** ['HYBRID_SAAS', 'SERVER', 'SAAS_TRIAL', 'CLOUDABILITY_PREMIUM']

searchApiEnabled

- **type:** boolean
- **description:** Controls use of new search api

parkingEnabled

- **type:** boolean
- **description:** Controls hiding/showing parking related pages in UI

frontEndTelemetryEnabled

- **type:** boolean
- **description:** Controls enabling/disabling front end telemetry in UI

jdbcHttpSessionEnabled

- **type:** boolean
- **description:** Controls use of JDBC persistent HTTP Session

saasReportingEnabled

- **type:** boolean
- **description:** Controls hiding/showing SaaS reports page in UI

trialExpirationTime

- **type:** string
- **description:** [BETA]: When the user's trial expires in ISO 8601 Format. For SAAS_TRIAL deploymentMode

featureFlags

- **type:** object
- **description:** [BETA]: A map of feature flags. DISCLAIMER: Do not rely on these for scripting, can be removed at any time.
- **additionalProperties:** {'type': 'boolean'}

reportingEnabled

- **type:** boolean

ProductVersionDTO

Description

Contains version information about the running Turbonomic instance.

Optional Parameters

versionInfo

- **type:** string
- **example:** "versionInfo": "Turbonomic Operations Manager 8.0.1-SNAPSHOT (Build \"20200922092548000\") \"2020-09-22 20:10:04\"\n\naction-orchestrator: 8.0.1-SNAPSHOT\napi: 8.0.1-SNAPSHOT\nauth: 8.0.1-SNAPSHOT\ncost: 8.0.1-SNAPSHOT\nextractor: 8.0.1-SNAPSHOT \ngroup: 8.0.1-SNAPSHOT\nhistory: 8.0.1-SNAPSHOT\nmarket: 8.0.1-SNAPSHOT\nmediation-actionscript: 8.0.1-SNAPSHOT\nmediation-appdynamics: 8.0.1-SNAPSHOT\nmediation-appinsights: 8.0.1-SNAPSHOT\nmediation-aws: 8.0.1-SNAPSHOT\nmediation-awsbilling: 8.0.1-SNAPSHOT \nmediation-awscost: 8.0.1-SNAPSHOT\nmediation-azure: 8.0.1-SNAPSHOT\nmediation-azurecost: 8.0.1-SNAPSHOT\nmediation-azureea: 8.0.1-SNAPSHOT\nmediation-azuresp: 8.0.1-SNAPSHOT \nmediation-azurevolumes: 8.0.1-SNAPSHOT\nmediation-dynatrace: 8.0.1-SNAPSHOT\nmediation-hpe3par: 8.0.1-SNAPSHOT\nmediation-hyperv: 8.0.1-SNAPSHOT\nmediation-netapp: 8.0.1-SNAPSHOT \nmediation-nutanix: 8.0.1-SNAPSHOT\nmediation-oneview: 8.0.1-SNAPSHOT\nmediation-pure: 8.0.1-SNAPSHOT\nmediation-ucs: 8.0.1-SNAPSHOT\nmediation-vcenter: 8.0.1-SNAPSHOT\nmediation-vcenterbrowsing: 8.0.1-SNAPSHOT\nmediation-vmax: 8.0.1-SNAPSHOT\nmediation-vmm: 8.0.1-SNAPSHOT\nplan-orchestrator: 8.0.1-SNAPSHOT\nrepository: 8.0.1-SNAPSHOT\ntopology-processor: 8.0.1-SNAPSHOT"
- **description:** A multi-line string describing the version information, including that of the various internal components.

version

- **type:** string
- **example:** 8.0.1-SNAPSHOT
- **description:** The Operations Manager product version, including major and minor versions

build

- **type:** string
- **example:** 20200922211058775
- **description:** The build number associated with this deployment.

commit

- **type:** string
- **example:** 1183dd2
- **description:** The commit associated with this deployment.

branch

- **type:** string
- **example:** 8.0.1
- **description:** The branch associated with this deployment.

gitDescription

- **type:** string
- **example:** 1183dd2bb8ebca3dbcaf708e0c323cabc7b8e352
- **description:** Specific source control information associated with this deployment.

updates

- **type:** string
- **description:** Description of available updates--not yet implemented in XL.

marketVersion

- **type:** integer
- **format:** int32
- **example:** 2
- **description:** The version of the market

apiVersion

- **type:** string
- **example:** 3.1
- **description:** The version of the REST API.

PurchaseProfileDTO

Description

Model to describe a set of preferences for purchasing RIs

Required Parameters

cloudType

- **type:** string
- **description:** Cloud type
- **enum:** ['AWS', 'AZURE', 'GCP', 'HYBRID', 'UNKNOWN']

paymentOption

- **type:** string
- **description:** Preferred payment option
- **enum:** ['ALL_UPFRONT', 'PARTIAL_UPFRONT', 'NO_UPFRONT']

reservedInstanceType

- **type:** string
- **description:** Preferred offering class (Standard or Convertible)
- **enum:** ['STANDARD', 'CONVERTIBLE']

termYears

- **type:** integer
- **format:** int32
- **description:** Number of years in preferred term

QueryInputApiDTO

Description

Model to describe an input query

Required Parameters

query

- **type:** string
- **description:** Query string. The format of this depends on the QueryType chosen. For example, for a REGEX QueryType, this should be a regular expression.

type

- **type:** string
- **description:** Query type
- **enum:** ['EXACT', 'REGEX', 'CONTAINS']

Optional Parameters

caseSensitive

- **type:** boolean
- **description:** If true, the query string will match case-sensitivity. This field is only used when QueryType is not REGEX.

RIBuyActionDetailsApiDTO

Description

Inherits From

This DTO inherits from ActionDetailsApiDTO *(on page 1492)*

Optional Parameters

estimatedOnDemandCost

- **type:** number
- **format:** float

estimatedRICoverage

- **type:** number
- **format:** float

historicalDemandDataset

- **type:** array
- – **reference:** StatSnapshotApiDTO *(on page 1684)*

subType

- **type:** string

RIPurchaseProfilesSettingApiDTO

Description

Inherits From

This DTO inherits from SettingApiDTO *(on page 1675)*

Required Parameters

value

- **description:** Value of the RI Purchase Profiles of type ReservedInstancePurchaseSettingsDTO.
- **reference:** ReservedInstancePurchaseSettingsDTO *(on page 1650)*

Optional Parameters

defaultValue

- **description:** Default value before any change
- **readOnly:** True
- **reference:** ReservedInstancePurchaseSettingsDTO *(on page 1650)*

RangeApiDTO

Description

Model to describe the Range of a Value in the Settings

Optional Parameters

step

- **type:** number
- **format:** double
- **description:** The Step scale factor of the slider

labels

- **type:** array
- **description:** Labels to show in the values of the slider
  - **type:** string

customStepValues

- **type:** array
- **description:** Custom step values when the slider step is not constant - [DEPRECATED - Use: stepValues]
  - **type:** integer
  - **format:** int32

stepValues

- **type:** array
- **description:** Custom step values when the slider step is not constant
  - **type:** number
  - **format:** float

RangeInputApiDTO

Description

Model to describe an input range. Either minValue or maxValue needs to be defined.

Optional Parameters

minValue

- **type:** number
- **format:** float
- **description:** Minimum value of the input range

maxValue

- **type:** number
- **format:** float
- **description:** Maximum value of the input range

Recommendation

Description

A DTO for a single recommendation to an error.

Required Parameters

description

- **type:** string
- **description:** The recommended user action to investigate/mitigate the problem.

errorType

- **type:** string
- **description:** The problem this recommendation is produced for.

ReconfigureActionDetailsApiDTO

Description

Inherits From

This DTO inherits from ActionDetailsApiDTO *(on page 1492)*

Optional Parameters

vcpuBefore

- **type:** integer
- **format:** int32
- **description:** vCPUs before the change

vcpuAfter

- **type:** integer
- **format:** int32
- **description:** vCPUs after the change

socketsBefore

- **type:** integer
- **format:** int32
- **description:** sockets before the change

socketsAfter

- **type:** integer
- **format:** int32
- **description:** sockets after the change

coresPerSocketBefore

- **type:** integer
- **format:** int32
- **description:** cores per socket before the change

coresPerSocketAfter

- **type:** integer
- **format:** int32
- **description:** cores per socket after the change

RecurrenceApiDTO

Description

Model to describe the Schedule for a Settings Policy

Required Parameters

type

- **type:** string
- **description:** Type of the recurrence, Enum: DAILY, WEEKLY, MONTHLY
- **enum:** ['DAILY', 'WEEKLY', 'MONTHLY']

Optional Parameters

daysOfWeek

- **type:** array
- **description:** The day of week that the recurrence will occur. Required if the type of recurrence is WEEKLY, or if the type is MONTHLY and a week for the monthly recurrence has been set.
-
  - **type:** string
  - **enum:** ['Sun', 'Mon', 'Tue', 'Wed', 'Thu', 'Fri', 'Sat']

daysOfMonth

- **type:** array
- **description:** When recurrence type is MONTHLY, the date to apply the monthly recurrence
- – **type:** integer
  – **format:** int32

weekOfTheMonth

- **type:** array
- **description:** When recurrence type is MONTHLY, which week of the month to apply the recurrence. A value of -1 indicates last week, 0 indicates this week, and so on.
- – **type:** integer
  – **format:** int32

interval

- **type:** integer
- **format:** int32
- **description:** Frequency of the recurrence based on the recurrence type. For example, an interval of '2' with a recurrence type of MONTHLY will occur every two months.

RegionAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

latitude

- **type:** number
- **format:** double
- **description:** The latitude of the region used along with longitude to determine physical location.

longitude

- **type:** number
- **format:** double
- **description:** The longitude of the region used along with latitude to determine physical location.

numWorkloads

- **type:** integer
- **format:** int32
- **description:** Number of workloads (VMs, databases, etc.) in this region.

RelatedActionApiDTO

Description

Model to describe the relationship between two actions

Optional Parameters

action

- **description:** The basic action DTO for the related action
- **reference:** BasicActionApiDTO *(on page 1511)*

actionRelationType

- **type:** string
- **description:** Action relation type
- **enum:** ['BLOCKED_BY', 'BLOCKING', 'CAUSED_BY', 'CAUSING']

RelatedEntityFieldApiDTO

Description

Inherits From

This DTO inherits from FieldApiDTO *(on page 1562)*

Required Parameters

property

- **type:** string
- **description:** The related entity property of interest.
- **readOnly:** True
- **enum:** ['NAMES', 'COUNT']

relatedEntityType

- **type:** string
- **description:** The type of related entities of interest.
- **readOnly:** True
- **enum:** ['Application', 'ApplicationComponent', 'ApplicationServer', 'AvailabilityZone', 'BusinessAccount', 'BusinessApplication', 'BusinessTransaction', 'BusinessUser', 'Chassis', 'CloudService', 'ComputeTier', 'Container', 'ContainerPod', 'ContainerSpec', 'Database', 'DatabaseServer', 'DatabaseServerTier', 'DatabaseTier', 'DataCenter', 'DesktopPool', 'DiskArray', 'DPod', 'HCIPhysicalMachine', 'HypervisorServer', 'Internet', 'IOModule', 'LoadBalancer', 'LogicalPool', 'Namespace', 'Network', 'PhysicalMachine', 'ProcessorPool', 'Region', 'ReservedInstance', 'Service', 'ServiceProvider', 'Storage', 'StorageController', 'StorageTier', 'Switch', 'Unknown', 'ViewPod', 'VirtualApplication', 'VirtualDataCenter', 'VirtualMachine', 'VirtualVolume', 'VPod', 'WorkloadController', 'ContainerPlatformCluster', 'CloudCommitment', 'ApplicationComponentSpec', 'VirtualMachineSpec', 'DocumentCollection', 'NodeGroup', 'PowerVMPhysicalMachine', 'DataWarehouse', 'DataWarehouseTier']

RelatedGroupFieldApiDTO

Description

Inherits From

This DTO inherits from FieldApiDTO *(on page 1562)*

Required Parameters

property

- **type:** string
- **description:** The related group property of interest.
- **readOnly:** True
- **enum:** ['NAMES', 'COUNT']

relatedGroupType

- **type:** string
- **description:** The type of related groups of interest.
- **readOnly:** True
- **enum:** ['Group', 'Resource', 'Cluster', 'VMCluster', 'StorageCluster', 'BusinessAccountFolder', 'BillingFamily', 'NodePool']

RelatedStatsApiDTO

Description

Information on relation type and its related commodities

Required Parameters

commodityDetailsList

- **type:** array
- **description:** Commodity detail list
- – **reference:** CommodityDetailsApiDTO *(on page 1532)*

relationType

- **type:** string
- **description:** Relation Type
- **enum:** ['sold', 'bought']

RelievePressureObjectApiDTO

Description

Model to describe a relieve pressure change in the scenario: minimum amount of Actions Moves that will relieve the pressure on the over-utilized clusters

Optional Parameters

projectionDay

- **type:** integer
- **format:** int32
- **description:** Projection day, indicate when to apply the change

sources

- **type:** array
- **description:** Sources of the change, List of Clusters that are overloaded
- – **reference:** BaseApiDTO *(on page 1509)*

destinations

- **type:** array
- **description:** Destinations of the change, List of Clusters where VMs can move to
- – **reference:** BaseApiDTO *(on page 1509)*

RemoveConstraintApiDTO

Description

Model to describe a constraint change in the scenario

Optional Parameters

constraintType

- **type:** string
- **description:** Type of the constraint to remove. To ignore all constraints – use AllCommodities. This will ignore all constraints not just Network/Cluster/StorageCluster/DC. GlobalIgnoreConstraint – deprecated – use AllCommodities instead.
- **enum:** ['DataCenterCommodity', 'ClusterCommodity', 'NetworkCommodity', 'DatastoreCommodity', 'StorageClusterCommodity', 'GlobalIgnoreConstraint', 'AllCommodities']

projectionDay

- **type:** integer
- **format:** int32
- **description:** Projection day, indicate when to apply the change

target

- **description:** Target of the change, when not provided change applied to plan scope
- **reference:** BaseApiDTO *(on page 1509)*

targetEntityType

- **type:** string
- **description:** EntityType to target with constraint configuration
- **enum:** ['VirtualMachine', 'Container', 'ContainerPod']

RemoveObjectApiDTO

Description

Model to describe a remove change in the scenario

Optional Parameters

projectionDay

- **type:** integer
- **format:** int32
- **description:** Projection day, indicate when to apply the change

targetEntityType

- **type:** string
- **description:** Type of entity to remove from group target

target

- **description:** Target of the change, entity or group
- **reference:** BaseApiDTO *(on page 1509)*

ReplaceObjectApiDTO

Description

Model to describe a migration change in the scenario

Optional Parameters

projectionDay

- **type:** integer
- **format:** int32
- **description:** Projection day, indicate when to apply the change

target

- **description:** Target of the change, entity or group
- **reference:** BaseApiDTO *(on page 1509)*

template

- **description:** Template used for the replace
- **reference:** BaseApiDTO *(on page 1509)*

targetEntityType

- **type:** string
- **description:** Type of entity to remove from group target

ReplaceWebhookBodyRequestApiDTO

Description

The details of a request to substitute action fields in a Webhook body template.

Required Parameters

webhookBody

- **type:** string
- **description:** The Webhook body being substituted.

ReplaceWebhookBodyResponseApiDTO

Description

Result of substituting action field values in a Webhook body template.

Required Parameters

succeeded

- **type:** boolean
- **description:** The Status of webhook body replacement.

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

error

- **type:** string
- **description:** The error that occurred during replacement, if any.

replacedBody

- **type:** string
- **description:** The replaced body after substitution of action fields.

ReplicaApiDTO

Description

Model to describe Replica attributes

Required Parameters

type

- **type:** string

Optional Parameters

name

- **type:** string
- **description:** Replica name

failoverPriority

- **type:** string
- **description:** Failover priority level of this replica

clusterRole

- **type:** string
- **description:** Role of this replica.
- **enum:** ['Writer', 'Reader']

deploymentType

- **type:** string
- **description:** Deployment type
- **enum:** ['SINGLE_AZ', 'MULTI_AZ']

sli

- **type:** array
  - **type:** string

RequestHeader

Description

Model to describe a request header.

Required Parameters

name

- **type:** string
- **description:** Request header name

Optional Parameters

value

- **type:** string
- **description:** Request header value. On create, required. On update, not provided means re-use existing.

isSecret

- **type:** boolean
- **description:** Whether the request header value should be encrypted

ReservationConstraintApiDTO

Description

The constraints user specified when running reservation which limit where templates could be placed.

Required Parameters

constraintType

- **type:** string
- **description:** Type of constraint
- **enum:** ['CLUSTER', 'DATA_CENTER', 'VIRTUAL_DATA_CENTER', 'POLICY', 'NETWORK']

uuid

- **type:** string
- **description:** UUID of the constraint

Optional Parameters

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

ReservationFailureInfoDTO

Description

Failure information when reservation fails

Required Parameters

closestSeller

- **description:** provider that came closest to meeting the requirements
- **reference:** BaseApiDTO *(on page 1509)*

closestSellerCluster

- **description:** Closest provider's cluster information
- **reference:** BaseApiDTO *(on page 1509)*

maxQuantityAvailable

- **type:** number
- **format:** double
- **description:** Quantity of resource available on closestSeller

quantityRequested

- **type:** number
- **format:** double
- **description:** Quantity of resource requested

resource

- **type:** string
- **description:** Most constrained resource

timespan

- **type:** string
- **description:** Failure is in historical or current
- **enum:** ['HISTORICAL', 'CURRENT', 'PROJECTED', 'PLAN_SOURCE', 'PLAN_PROJECTED']

Optional Parameters

units

- **type:** string
- **description:** Units for the resource

ReservationInvalidInfoApiDTO

Description

The reason the reservation is invalid

Optional Parameters

invalidReason

- **type:** string
- **description:** The reason why the reservation is invalid
- **enum:** ['MARKET_CONNECTIVITY_ERROR', 'MARKET_NOT_READY', 'INVALID_CONSTRAINTS']

ReservedInstanceApiDTO

Description

Model to describe a Reserved Instance

Required Parameters

cloudType

- **type:** string
- **description:** Cloud Type
- **enum:** ['AWS', 'AZURE', 'GCP', 'HYBRID', 'UNKNOWN']

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

template

- **description:** Related template
- **reference:** BaseApiDTO *(on page 1509)*

platform

- **type:** string
- **description:** Platform name
- **enum:** ['UNKNOWN', 'LINUX', 'RHEL', 'SUSE', 'WINDOWS', 'WINDOWS_WITH_SQL_STANDARD', 'WINDOWS_WITH_SQL_WEB', 'WINDOWS_WITH_SQL_SERVER_ENTERPRISE', 'LINUX_WITH_SQL_STANDARD', 'LINUX_WITH_SQL_WEB', 'LINUX_WITH_SQL_SERVER_ENTERPRISE']

type

- **type:** string
- **description:** Reserved instance type
- **enum:** ['STANDARD', 'CONVERTIBLE']

location

- **description:** Location, either a zone or a data center if zone is not available
- **reference:** BaseApiDTO *(on page 1509)*

accountId

- **type:** string
- **description:** Account ID

accountDisplayName

- **type:** string
- **description:** Account display name

targetId

- **type:** string
- **description:** The associated target ID of the account

masterAccountId

- **type:** string
- **description:** Master account ID

term

- **description:** Term, unit is Year
- **reference:** StatApiDTO *(on page 1679)*

payment

- **type:** string
- **description:** Payment option
- **enum:** ['ALL_UPFRONT', 'PARTIAL_UPFRONT', 'NO_UPFRONT']

onDemandPrice

- **description:** On-demand price, saving is RI price minus on-demand price
- **reference:** StatApiDTO *(on page 1679)*

costPrice

- **description:** Total monthly cost
- **reference:** StatApiDTO *(on page 1679)*

coupons

- **description:** Total coupons and used coupons, utilization = numOfCouponsUsed / numOfCoupons
- **reference:** StatApiDTO *(on page 1679)*

totalCoupons

- **type:** number
- **format:** float
- **description:** Total coupons provided by the cbtp the vm matched with

expDate

- **type:** string
- **description:** Expiration date

instanceCount

- **type:** integer
- **format:** int32
- **description:** Number of reserved instance with the same type, template, platform etc..

tenancy

- **type:** string
- **description:** Tenancy
- **enum:** ['DEFAULT', 'DEDICATED', 'HOST']

upFrontCost

- **type:** number
- **format:** double
- **description:** Up front cost

actualHourlyCost

- **type:** number
- **format:** double
- **description:** Actual hourly cost

effectiveHourlyCost

- **type:** number
- **format:** double
- **description:** Effective hourly cost

isExpired

- **type:** boolean
- **description:** Whether the RI be expired by a date provided

expDateEpochTime

- **type:** integer
- **format:** int64
- **description:** Expiration Date in epoch time

scopeType

- **type:** string
- **description:** Whether the RI scope is Single or Shared (Azure)
- **enum:** ['SINGLE', 'SHARED', 'UNKNOWN']

appliedScopes

- **type:** array
- **description:** List of applied scope UUIDs- only provided when scope is Single (Azure)
- – **type:** string

orderID

- **type:** string
- **description:** The Order ID of the RI (Azure)

trueID

- **type:** string
- **description:** The ID of the RI

coveredEntityCount

- **type:** integer
- **format:** int32
- **description:** Count of workload entities covered by the reserved instance

undiscoveredAccountsCoveredCount

- **type:** integer
- **format:** int32
- **description:** Count of undiscovered accounts covered by the reserved instance

toBuy

- **type:** boolean

sizeFlexible

- **type:** boolean

ReservedInstancePurchaseSettingsDTO

Description

Optional Parameters

awsProfile

- **reference:** PurchaseProfileDTO *(on page 1634)*

azureProfile

- **reference:** PurchaseProfileDTO *(on page 1634)*

ResizeActionDetailsApiDTO

Description

Parent class for resize action details. Implementing classes include: CloudResizeActionDetailsApiDTO.

Required Parameters

actionDetailType

- **type:** string

actionType

- **type:** string

Optional Parameters

subType

- **type:** string

ResourceApiDTO

Description

Model to describe the Resources of a Template

Optional Parameters

stats

- **type:** array
- **description:** Statistics, e.g. Number of CPU, Memory size, Disk size, ...
  - **reference:** StatApiDTO *(on page 1679)*

provider

- **description:** Entity provider
- **reference:** BaseApiDTO *(on page 1509)*

type

- **type:** string
- **description:** Type of the resource

template

- **type:** string
- **description:** Template used to fulfill the resources

relatedResources

- **type:** array
- **description:** ID of resources related to this resource
  - **type:** string

linkedResources

- **type:** array
- **description:** resources related to this resource. For now, this will only contain the cluster
  - **reference:** BaseApiDTO *(on page 1509)*

ResourceGroupApiDTO

Description

Inherits From

This DTO inherits from GroupApiDTO *(on page 1564)*

Optional Parameters

parentUuid

- **type:** string
- **description:** Parent Business Account uuid (e.g subscription in Azure) where the Resource Group belongs to.

parentDisplayName

- **type:** string
- **description:** Parent Business Account display name (e.g subscription in Azure) where the Resource Group belongs to.

RoleApiDTO

Description

Model to describe a User Role, it describe the permissions of an user over the Ops. Manager capabilities

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

name

- **type:** string
- **description:** Name of the role

SAMLIdpApiDTO

Description

Model to describe an SAML IDP. It contains IDP URL and is SAML ONLY enabled or not.

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

idpURL

- **type:** string
- **description:** The SAML IDP URL

isSAMLOnly

- **type:** boolean

isSingleLogoutEnabled

- **type:** boolean

STEntityAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

displayName

- **type:** string
- **description:** Display Name

name

- **type:** string
- **description:** Name

productName

- **type:** string
- **description:** Product Name

productType

- **type:** string
- **description:** Product Type

deviceId

- **type:** string
- **description:** Device ID

deviceName

- **type:** string
- **description:** Device Name

externalNames

- **type:** array
- **description:** External Names
  - **type:** string

altName

- **type:** string
- **description:** Alternative Name

ssdDiskCount

- **type:** integer
- **format:** int64
- **description:** SSD Disk Count

rpm7200DiskCount

- **type:** integer
- **format:** int64
- **description:** RPM 7200 Disk Count

rpm10KDiskCount

- **type:** integer
- **format:** int64
- **description:** RPM 10K Disk Count

rpm15KDiskCount

- **type:** integer
- **format:** int64
- **description:** RPM 15K Disk Count

getvSeriesDiskCount

- **type:** integer
- **format:** int64
- **description:** VSeries Disk Count

ssdDiskIOPSCapacity

- **type:** number
- **format:** double
- **description:** SSD Disk IOPS Capacity

rpm7200DiskIOPSCapacity

- **type:** number
- **format:** double
- **description:** RPM 7200 Disk IOPS Capacity

rpm10KDiskIOPSCapacity

- **type:** number
- **format:** double
- **description:** RPM 10K Disk IOPS Capacity

rpm15KDiskIOPSCapacity

- **type:** number
- **format:** double
- **description:** RPM 15K Disk IOPS Capacity

getvSeriesDiskIOPSCapacity

- **type:** number
- **format:** double
- **description:** VSeries Disk IOPS Capacity

technologyType

- **type:** string
- **description:** HCI Technology Type

rawCapacity

- **type:** integer
- **format:** int64
- **description:** HCI Raw Capacity

rawFreespace

- **type:** integer
- **format:** int64
- **description:** HCI Raw Freespace

rawUncommitted

- **type:** integer
- **format:** int64
- **description:** HCI Raw Uncommitted

redundancyMethod

- **type:** string
- **description:** HCI Redundancy Method

failuresToTolerate

- **type:** integer
- **format:** int32
- **description:** HCI Failures To Tolerate

spaceReservationPct

- **type:** integer
- **format:** int32
- **description:** HCI Space Reservation Percentage

minimumHosts

- **type:** integer
- **format:** int32
- **description:** HCI Minimum Hosts Required

isHciStorage

- **type:** boolean

vendorLink

- **type:** string
- **description:** Link to navigate to the entity in vendor UI.

SaaSReportingConfigApiDTO

Description

SaaS Reporting Configuration Information

Required Parameters

configuration

- **type:** string
- **description:** SaaS Reporting Encoded Configuration

enabled

- **type:** boolean
- **description:** SaaS Reporting Configuration Information

SavingsInvestmentsApiDTO

Description

Model to describe Savings/Investments

Required Parameters

savingsInvestmentsTotal

- **description:** The total Saving/Investment value
- **reference:** SavingsInvestmentsSnapshotApiDTO *(on page 1658)*

Optional Parameters

csp

- **description:** The csp associated with this savings/investments value
- **reference:** BaseApiDTO *(on page 1509)*

account

- **description:** The account associated with this savings/investments value
- **reference:** BaseApiDTO *(on page 1509)*

resourceGroup

- **description:** The resourceGroup associated with this savings/investments value
- **reference:** BaseApiDTO *(on page 1509)*

region

- **description:** The region associated with this savings/investments value
- **reference:** BaseApiDTO *(on page 1509)*

entity

- **description:** The entity associated with this savings/investments value
- **reference:** BaseApiDTO *(on page 1509)*

actionChainId

- **type:** string
- **description:** The actionChainId associated with this savings/investments value

entityType

- **type:** string
- **description:** The entityType associated with this savings/investments value
- **enum:** ['Application', 'ApplicationComponent', 'ApplicationServer', 'AvailabilityZone', 'BusinessAccount', 'BusinessApplication', 'BusinessTransaction', 'BusinessUser', 'Chassis', 'CloudService', 'ComputeTier', 'Container', 'ContainerPod', 'ContainerSpec', 'Database', 'DatabaseServer', 'DatabaseServerTier', 'DatabaseTier', 'DataCenter', 'DesktopPool', 'DiskArray', 'DPod', 'HCIPhysicalMachine', 'HypervisorServer', 'Internet', 'IOModule', 'LoadBalancer', 'LogicalPool', 'Namespace', 'Network', 'PhysicalMachine', 'ProcessorPool', 'Region', 'ReservedInstance', 'Service', 'ServiceProvider', 'Storage', 'StorageController', 'StorageTier', 'Switch', 'Unknown', 'ViewPod', 'VirtualApplication', 'VirtualDataCenter', 'VirtualMachine', 'VirtualVolume', 'VPod', 'WorkloadController', 'ContainerPlatformCluster', 'CloudCommitment', 'ApplicationComponentSpec', 'VirtualMachineSpec', 'DocumentCollection', 'NodeGroup', 'PowerVMPhysicalMachine', 'DataWarehouse', 'DataWarehouseTier']

actionType

- **type:** string
- **description:** The actionType associated with this savings/investments value
- **enum:** ['NONE', 'START', 'MOVE', 'SCALE', 'ALLOCATE', 'SUSPEND', 'TERMINATE', 'SPAWN', 'ADD_PROVIDER', 'CHANGE', 'REMOVE_PROVIDER', 'PROVISION', 'RECONFIGURE', 'RESIZE', 'RESIZE_CAPACITY', 'WARN', 'RECONFIGURE_THRESHOLD', 'DELETE', 'RIGHT_SIZE', 'RESERVE_ON_PM', 'RESERVE_ON_DS', 'RESIZE_FOR_EFFICIENCY', 'RESIZE_FOR_PERFORMANCE', 'CROSS_TARGET_MOVE', 'MOVE_TOGETHER', 'ACTIVATE', 'DEACTIVATE', 'BUY_RI']

savingsInvestmentsSnapshot

- **type:** array
- **description:** The list of Savings/Investments snapshots based on the selected time period
  - **reference:** SavingsInvestmentsSnapshotApiDTO *(on page 1658)*

tag

- **description:** The tag key and value associated with this savings/investments value
- **reference:** SavingsInvestmentsTagApiDTO *(on page 1659)*

SavingsInvestmentsApiInputDTO

Description

Model to describe the request of a SavingsInvestment.

Optional Parameters

scopes

- **type:** array
- **description:** List of uuid to use as scope
  - **type:** string

filters

- **type:** array
- **description:** Filters to apply to the requested savings/investment. Supported values: CSP, ACCOUNT, RESOURCE_GROUP, REGION, ENTITY, ACTION_CHAIN, ENTITY_TYPE, ACTION_TYPE
  - **reference:** SavingsInvestmentsFilterApiDTO *(on page 1658)*

groupBy

- **type:** array
- **description:** Execute a groupBy on the values of the savings/investment. Supported values: CSP, ACCOUNT, RESOURCE_GROUP, REGION, ENTITY, ACTION_CHAIN, ENTITY_TYPE, ACTION_TYPE
- – **type:** string
  – **enum:** ['CSP', 'ACCOUNT', 'RESOURCE_GROUP', 'REGION', 'ENTITY', 'ACTION_CHAIN', 'ENTITY_TYPE', 'ACTION_TYPE', 'TAG_KEY', 'TAG_VAL', 'TAG_TYPE']

startDate

- **type:** string
- **description:** Start time of the period to filter, if empty the current savings/investment will be processed

endDate

- **type:** string
- **description:** End time of the period to filter, if empty the current savings/investment will be processed

includeSnapshots

- **type:** boolean
- **description:** Include time period snapshots of savings/investment. Could be daily or monthly

includeOnlySavings

- **type:** boolean
- **description:** Include only the Savings values

includeOnlyInvestments

- **type:** boolean
- **description:** Include only the Investments values

SavingsInvestmentsFilterApiDTO

Description

Model to describe the filter applied to a Savings/Investments response

Required Parameters

type

- **type:** string
- **description:** type of the filter, E.G: ACTION_TYPE, RESOURCE_GROUP ...
- **enum:** ['CSP', 'ACCOUNT', 'RESOURCE_GROUP', 'REGION', 'ENTITY', 'ACTION_CHAIN', 'ENTITY_TYPE', 'ACTION_TYPE', 'TAG_KEY', 'TAG_VAL', 'TAG_TYPE']

value

- **type:** string
- **description:** value of the filter

SavingsInvestmentsSnapshotApiDTO

Description

Model to describe the Saving/Investment of a snapshot, contains a list of Savings/Investments for a moment in time

Required Parameters

currency

- **type:** string
- **description:** Currency

investments

- **type:** number
- **format:** double
- **description:** Investments value

savings

- **type:** number
- **format:** double
- **description:** Savings value

Optional Parameters

date

- **type:** string
- **description:** When the Saving/Investment was recorded

SavingsInvestmentsTagApiDTO

Description

Tag information

Optional Parameters

key

- **type:** string
- **description:** Tag key

value

- **type:** string
- **description:** Tag value

ScenarioApiDTO

Description

Model to describe a What-If Scenario, composed by a list of changes

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

owners

- **type:** array
- **description:** Users that owns the scenario
- – **reference:** UserApiDTO *(on page 1708)*

type

- **type:** string
- **description:** Type of the scenario

scope

- **type:** array
- **description:** List of entities/group that define the Scope
- – **reference:** BaseApiDTO *(on page 1509)*

projectionDays

- **type:** array
- **description:** List of days that define when the scenario should run
- – **type:** integer
  – **format:** int32

topologyChanges

- **description:** Changes to apply to the topology of the plan market
- **reference:** TopologyChangesApiDTO *(on page 1704)*

loadChanges

- **description:** Workload utilization changes in the scenario
- **reference:** LoadChangesApiDTO *(on page 1580)*

configChanges

- **description:** Configuration changes in the scenario
- **reference:** ConfigChangesApiDTO *(on page 1536)*

timebasedTopologyChanges

- **reference:** TimeBasedTopologyChangesApiDTO *(on page 1703)*

ScheduleApiDTO

Description

Model to describe the Schedule for a Settings Policy

Required Parameters

endTime

- **type:** string
- **format:** date-time
- **example:** 2019-04-04T19:15
- **description:** End time in hours and minutes for the scheduled period in ISO8601. For example, 2019-04-04T19:15

startDate

- **type:** string
- **format:** date-time
- **example:** 2019-04-04T19:15
- **description:** Start date of the schedule period, expressed using ISO8601 Format (YYYY-MM-DDTHH:MM:SS). This date is local to the timezone of the instance.

startTime

- **type:** string
- **format:** date-time
- **example:** 2019-04-04T19:15
- **description:** Start time in hours and minutes for the scheduled period in ISO8601. For example, 2019-04-04T17:15

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

endDate

- **type:** string
- **format:** date
- **example:** 2019-04-04
- **description:** End date of the schedule period, expressed using ISO8601 Format (YYYY-MM-DDTHH:MM:SS). This date is local to the timezone of the instance.

recurrence

- **description:** Recurrence setup
- **reference:** RecurrenceApiDTO *(on page 1638)*

timeZone

- **type:** string
- **description:** The time zone name based on tz database

nextOccurrence

- **type:** string
- **description:** The date and time of the next occurence of this schedule represented in ISO8601 local time (YYYY-MM-DDTHH:MM:SS). For example, 2019-04-04T19:15

nextOccurrenceTimestamp

- **type:** integer
- **format:** int64
- **description:** The date and time of the next occurence of this schedule represented in Timestamp in milliseconds. For example, 1562100619000

remaingTimeActiveInMs

- **type:** integer
- **format:** int64
- **description:** The amount of time that the schedule is going to be active at the time of call in milliseconds. This field will not be populated if the schedule is not currently active.

ScheduleEntityResponseApiDTO

Description

represents a response of the attach/detach time span schedule to/from the parkable entity

Optional Parameters

error

- **type:** string
- **example:** entityUuid not found
- **description:** error indicating the reason for failure

entityUUID

- **type:** string

ScheduleItemApiDTO

Description

Base model to describe a ScheduleItem.

Required Parameters

type

- **type:** string
- **example:** ParkingItemApiDTO
- **description:** ScheduleItem type discriminator. Describes which subclass of ScheduleItem is returned.

ScheduleTimeSpansApiDTO

Description

Parking time span schedule

Required Parameters

timeZone

- **type:** string
- **example:** America/New_York
- **description:** Timezone for the time spans in this schedule. Should be chosen from the IANA Timezone Database here: [https://www.iana.org/time-zones]. See the list in wikipedia here: [https://en.wikipedia.org/wiki/List_of_tz_database_time_zones]

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

description

- **type:** string
- **example:** Up 7-5 on weekdays
- **description:** Description of the schedule

timeSpans

- **description:** Weekly schedule time spans
- **reference:** TimeSpanGroupApiDTO *(on page 1704)*

source

- **type:** string
- **description:** [BETA]: Time span schedule source
- **enum:** ['USER', 'SMART_PARK_RECOMMENDATION']

ScheduleTimeSpansInputApiDTO

Description

Model to describe a filter for schedule timespans.

Optional Parameters

criteriaList

- **type:** array
- **description:** The list of filters to apply in the search request.
- – **reference:** FilterApiDTO *(on page 1563)*

logicalOperator

- **type:** string
- **description:** Logical operator to be applied across all the criteria.
- **enum:** ['AND', 'OR']

ScopeUuidsApiInputDTO

Description

Model to describe the request by UUIDs

Required Parameters

uuids

- **type:** array
- **description:** List of uuids
- – **type:** string
- **maxItems:** 2147483647
- **minItems:** 1

Optional Parameters

marketId

- **type:** string
- **description:** Market ID associated with the requested items

SearchAllQueryApiDTO

Description

A query used to search across the entire topology (entities and groups)

Required Parameters

select

- **description:** Defines what subset of types (entities and groups) should be returned.
- **readOnly:** True
- **reference:** SelectAllApiDTO *(on page 1666)*

where

- **description:** Conditions that each retrieved object must satisfy (currently only name is supported).
- **readOnly:** True
- **reference:** WhereApiDTO *(on page 1719)*

Optional Parameters

pagination

- **description:** Pagination information.
- **readOnly:** True
- **reference:** PaginationApiDTO *(on page 1602)*

SearchCountRecordApiDTO

Description

A count of entities.

Required Parameters

count

- **type:** integer
- **format:** int32
- **description:** Count of entities.
- **readOnly:** True

groupBys

- **type:** array
- **description:** List of fields describing the groupBys that this count represents
- **readOnly:** True
- – **reference:** FieldValueApiDTO *(on page 1562)*

SearchQueryRecordApiDTO

Description

A record with information about an entity or a group.

Required Parameters

oid

- **type:** integer
- **format:** int64
- **description:** Unique identifier for the entity or group.
- **readOnly:** True

values

- **type:** array
- **description:** Information about the entity or group.
- **readOnly:** True
- – **reference:** FieldValueApiDTO *(on page 1562)*

SelectAllApiDTO

Description

Specification of what subset of types (entities and groups) should be returned.

Optional Parameters

entityTypes

- **type:** array
- **description:** Entity types to include in the response.
- **readOnly:** True
- – **type:** string
- – **enum:** ['Application', 'ApplicationComponent', 'ApplicationServer', 'AvailabilityZone', 'BusinessAccount', 'BusinessApplication', 'BusinessTransaction', 'BusinessUser', 'Chassis', 'CloudService', 'ComputeTier', 'Container', 'ContainerPod', 'ContainerSpec', 'Database', 'DatabaseServer', 'DatabaseServerTier', 'DatabaseTier', 'DataCenter', 'DesktopPool', 'DiskArray', 'DPod', 'HCIPhysicalMachine', 'HypervisorServer', 'Internet', 'IOModule', 'LoadBalancer', 'LogicalPool', 'Namespace', 'Network', 'PhysicalMachine', 'ProcessorPool', 'Region', 'ReservedInstance', 'Service', 'ServiceProvider', 'Storage', 'StorageController', 'StorageTier', 'Switch', 'Unknown', 'ViewPod', 'VirtualApplication', 'VirtualDataCenter', 'VirtualMachine', 'VirtualVolume', 'VPod', 'WorkloadController', 'ContainerPlatformCluster', 'CloudCommitment', 'ApplicationComponentSpec', 'VirtualMachineSpec', 'DocumentCollection', 'NodeGroup', 'PowerVMPhysicalMachine', 'DataWarehouse', 'DataWarehouseTier']

groupTypes

- **type:** array
- **description:** Group types to include in the response.
- **readOnly:** True
- – **type:** string
- – **enum:** ['Group', 'Resource', 'Cluster', 'VMCluster', 'StorageCluster', 'BusinessAccountFolder', 'BillingFamily', 'NodePool']

SelectEntityApiDTO

Description

Specification of what fields should be fetched in an entity search response.

Required Parameters

entityType

- **type:** string
- **description:** Entity Type
- **readOnly:** True
- **enum:** ['Application', 'ApplicationComponent', 'ApplicationServer', 'AvailabilityZone', 'BusinessAccount', 'BusinessApplication', 'BusinessTransaction', 'BusinessUser', 'Chassis', 'CloudService', 'ComputeTier', 'Container', 'ContainerPod', 'ContainerSpec', 'Database', 'DatabaseServer', 'DatabaseServerTier', 'DatabaseTier', 'DataCenter', 'DesktopPool', 'DiskArray', 'DPod', 'HCIPhysicalMachine', 'HypervisorServer', 'Internet', 'IOModule', 'LoadBalancer', 'LogicalPool', 'Namespace', 'Network', 'PhysicalMachine', 'ProcessorPool', 'Region', 'ReservedInstance', 'Service', 'ServiceProvider', 'Storage', 'StorageController', 'StorageTier', 'Switch', 'Unknown', 'ViewPod', 'VirtualApplication', 'VirtualDataCenter', 'VirtualMachine', 'VirtualVolume', 'VPod', 'WorkloadController', 'ContainerPlatformCluster', 'CloudCommitment', 'ApplicationComponentSpec', 'VirtualMachineSpec', 'DocumentCollection', 'NodeGroup', 'PowerVMPhysicalMachine', 'DataWarehouse', 'DataWarehouseTier']

Optional Parameters

fields

- **type:** array
- **description:** Fields to fetch.
- **readOnly:** True
- – **reference:** FieldApiDTO *(on page 1562)*

SelectGroupApiDTO

Description

Specification of what fields should be fetched in a group search response.

Required Parameters

groupType

- **type:** string
- **description:** Group Type.
- **readOnly:** True
- **enum:** ['Group', 'Resource', 'Cluster', 'VMCluster', 'StorageCluster', 'BusinessAccountFolder', 'BillingFamily', 'NodePool']

Optional Parameters

fields

- **type:** array
- **description:** Fields to fetch.
- **readOnly:** True
- – **reference:** FieldApiDTO *(on page 1562)*

ServiceEntityApiDTO

Description

Model to describe an Entity, e.g. Virtual Machine, Physical Machine, Storage

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

environmentType

- **type:** string
- **description:** Environment type
- **enum:** ['CLOUD', 'ONPREM', 'HYBRID']

discoveredBy

- **description:** Target that discovered the entity
- **reference:** TargetApiDTO *(on page 1690)*

vendorIds

- **type:** object
- **description:** The mapping of target identifier to vendor-provided identity of this entity on the remote target
- **additionalProperties:** {'type': 'string'}

priceIndex

- **type:** number
- **format:** float
- **description:** Price index of the Entity

state

- **type:** string
- **description:** State
- **enum:** ['IDLE', 'RESOURCE_ALLOCATION', 'LAUNCH', 'ACTIVE', 'RESOURCE_RELEASE', 'SUSPEND', 'SUSPEND_PENDING', 'TERMINATE_PENDING', 'NOT_MONITORED', 'MAINTENANCE', 'FAILOVER', 'UNKNOWN', 'EVACUATED', 'QUEUED', 'NOT_FOUND']

severity

- **type:** string
- **description:** Highest severity of the Actions related to this entity
- **enum:** ['UNKNOWN', 'NORMAL', 'MINOR', 'MAJOR', 'CRITICAL']

unplacementDetails

- **description:** Additional information provided for unplaced entities in a plan
- **reference:** UnplacementDetailsApiDTO *(on page 1707)*

placedOn

- **type:** string
- **description:** The comma-separated classNames of entities successfully placed on, used for Unplaced entities in a Plan Market - [DEPRECATED - EOL: 8.15.1 - Use: unplacementDetails.placedOn]

notPlacedOn

- **type:** string
- **description:** The comma-separated classNames of entities could not be placed on, used for Unplaced entities in a Plan Market - [DEPRECATED - EOL: 8.15.1 - Use: unplacementDetails.reasons]

unplacedExplanation

- **type:** string
- **description:** The reason(s) the entity could not be placed - [DEPRECATED - EOL: 8.15.1 - Use: unplacementDetails.reasons]

costPrice

- **type:** number
- **format:** float
- **description:** Cost price per Hour

severityBreakdown

- **type:** object
- **description:** A Map of severity along with their count. E.g. {CRITICAL: 1, NORMAL: 10}
- **additionalProperties:** {'type': 'integer', 'format': 'int64'}

remoteId

- **type:** string
- **description:** The identifier of this instance on the remote target from which it was discovered. [DEPRECATED - EOL: 8.15.1 - Use: vendorIds]

realtimeMarketReference

- **description:** In case of a Plan Market Entity, this contains the original entity from the Real Time Market
- **reference:** BaseApiDTO *(on page 1509)*

providers

- **type:** array
- **description:** List of Providers
- – **reference:** BaseApiDTO *(on page 1509)*

consumers

- **type:** array
- **description:** List of Consumers
- – **reference:** BaseApiDTO *(on page 1509)*

template

- **description:** Related Template
- **reference:** TemplateApiDTO *(on page 1697)*

numRelatedVMs

- **type:** integer
- **format:** int32
- **description:** Number of related VMs

connectedEntities

- **type:** array
- **description:** Limited list of connected entities relevant to the service entity. This will not contain all connected entities.
- – **reference:** BaseApiDTO *(on page 1509)*

aspects

- **type:** object
- **description:** Additional info about the Entity categorized as Aspects
- **additionalProperties:** {'$ref': '#/definitions/EntityAspect'}

tags

- **type:** object
- **description:** Tags are the metadata defined in name/value pairs. Each name can have multiple values.
- **additionalProperties:** {'type': 'array', 'items': {'type': 'string'}}

onDemandRateBefore

- **type:** number
- **format:** float
- **description:** On-demand rate before actions

onDemandRateAfter

- **type:** number
- **format:** float
- **description:** On-demand rate after actions

staleness

- **type:** string
- **description:** [BETA]: Whether the entity is outdated, i.e. discovered too long ago to be considered for analytical purposes
- **enum:** ['CURRENT', 'STALE']

entityCosts

- **type:** array
- **description:** Cost info for the entity
- – **reference:** EntityCostApiDTO *(on page 1556)*

SettingActivePolicyApiDTO

Description

Model to describe a Settings Policy that affect the value of a Setting

Required Parameters

numEntities

- **type:** integer
- **format:** int32
- **description:** Number of entities affected by the Settings Policy value

settingsPolicy

- **description:** Settings Policy that affects the final value of a Setting
- **reference:** BaseApiDTO *(on page 1509)*

value

- **type:** string
- **description:** Setting value declared in the Settings Policy

SettingApiDTOSerializable

Description

Generic model to describe a Setting. Sub types include StringSettingApiDTO and RIPurchaseProfilesSettingApiDTO.

Required Parameters

value

- **description:** Value of the setting. Type indicated by the `valueObjectType` field. The default type is `String`
- **$ref:** #/definitions/object

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

valueDisplayName

- **type:** string
- **description:** Display Name of the Value, in case the value is a Template or another entity
- **readOnly:** True

defaultValue

- **description:** Default value before any change
- **readOnly:** True
- **$ref:** #/definitions/object

categories

- **type:** array
- **description:** Category and sub categories of the setting
- **readOnly:** True
- – **type:** string

valueType

- **type:** string
- **description:** Define the type of the value
- **enum:** ['STRING', 'NUMERIC', 'INTEGER', 'BOOLEAN', 'LIST', 'SECRET']

valueObjectType

- **type:** string
- **description:** The type of complex object values. Defaults to 'String'.
- **enum:** ['String', 'ReservedInstancePurchaseSettingsDTO']

options

- **type:** array
- **description:** List of allowable value for this setting
- **readOnly:** True
- – **reference:** SettingOptionApiDTO *(on page 1676)*

min

- **type:** number
- **format:** double
- **description:** If type is NUMERIC, this define the min limit
- **readOnly:** True

max

- **type:** number
- **format:** double
- **description:** If type is NUMERIC, this define the max limit
- **readOnly:** True

entityType

- **type:** string
- **description:** Entity type for this Setting

range

- **description:** Range options for the value of the setting
- **readOnly:** True
- **reference:** RangeApiDTO *(on page 1636)*

scope

- **type:** string
- **description:** Define the visibility of the setting. E.G. a Local Setting is only visible when associated with an Entity/Group
- **readOnly:** True
- **enum:** ['GLOBAL', 'LOCAL']

activeSettingsPolicies

- **type:** array
- **description:** List of Settings Policies that are affecting the final value of this setting
- – **reference:** SettingActivePolicyApiDTO *(on page 1670)*

sourceGroupName

- **type:** string
- **description:** Source group name for setting value

sourceGroupUuid

- **type:** string
- **description:** Source group uuid for setting value

SettingApiDTOString

Description

Generic model to describe a Setting. Sub types include StringSettingApiDTO and RIPurchaseProfilesSettingApiDTO.

Required Parameters

value

- **description:** Value of the setting. Type indicated by the `valueObjectType` field. The default type is `String`
- **$ref:** #/definitions/object

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

valueDisplayName

- **type:** string
- **description:** Display Name of the Value, in case the value is a Template or another entity
- **readOnly:** True

defaultValue

- **description:** Default value before any change
- **readOnly:** True
- **$ref:** #/definitions/object

categories

- **type:** array
- **description:** Category and sub categories of the setting
- **readOnly:** True
- – **type:** string

valueType

- **type:** string
- **description:** Define the type of the value
- **enum:** ['STRING', 'NUMERIC', 'INTEGER', 'BOOLEAN', 'LIST', 'SECRET']

valueObjectType

- **type:** string
- **description:** The type of complex object values. Defaults to 'String'.
- **enum:** ['String', 'ReservedInstancePurchaseSettingsDTO']

options

- **type:** array
- **description:** List of allowable value for this setting
- **readOnly:** True
- – **reference:** SettingOptionApiDTO *(on page 1676)*

min

- **type:** number
- **format:** double
- **description:** If type is NUMERIC, this define the min limit
- **readOnly:** True

max

- **type:** number
- **format:** double
- **description:** If type is NUMERIC, this define the max limit
- **readOnly:** True

entityType

- **type:** string
- **description:** Entity type for this Setting

range

- **description:** Range options for the value of the setting
- **readOnly:** True
- **reference:** RangeApiDTO *(on page 1636)*

scope

- **type:** string
- **description:** Define the visibility of the setting. E.G. a Local Setting is only visible when associated with an Entity/Group
- **readOnly:** True
- **enum:** ['GLOBAL', 'LOCAL']

activeSettingsPolicies

- **type:** array
- **description:** List of Settings Policies that are affecting the final value of this setting
- – **reference:** SettingActivePolicyApiDTO *(on page 1670)*

sourceGroupName

- **type:** string
- **description:** Source group name for setting value

sourceGroupUuid

- **type:** string
- **description:** Source group uuid for setting value

SettingApiDTO

Description

Generic model to describe a Setting. Sub types include StringSettingApiDTO and RIPurchaseProfilesSettingApiDTO.

Required Parameters

value

- **description:** Value of the setting. Type indicated by the `valueObjectType` field. The default type is `String`
- **$ref:** #/definitions/object

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

valueDisplayName

- **type:** string
- **description:** Display Name of the Value, in case the value is a Template or another entity
- **readOnly:** True

defaultValue

- **description:** Default value before any change
- **readOnly:** True
- **$ref:** #/definitions/object

categories

- **type:** array
- **description:** Category and sub categories of the setting
- **readOnly:** True
- – **type:** string

valueType

- **type:** string
- **description:** Define the type of the value
- **enum:** ['STRING', 'NUMERIC', 'INTEGER', 'BOOLEAN', 'LIST', 'SECRET']

valueObjectType

- **type:** string
- **description:** The type of complex object values. Defaults to 'String'.
- **enum:** ['String', 'ReservedInstancePurchaseSettingsDTO']

options

- **type:** array
- **description:** List of allowable value for this setting
- **readOnly:** True
- – **reference:** SettingOptionApiDTO *(on page 1676)*

min

- **type:** number
- **format:** double
- **description:** If type is NUMERIC, this define the min limit
- **readOnly:** True

max

- **type:** number
- **format:** double
- **description:** If type is NUMERIC, this define the max limit
- **readOnly:** True

entityType

- **type:** string
- **description:** Entity type for this Setting

range

- **description:** Range options for the value of the setting
- **readOnly:** True
- **reference:** RangeApiDTO *(on page 1636)*

scope

- **type:** string
- **description:** Define the visibility of the setting. E.G. a Local Setting is only visible when associated with an Entity/Group
- **readOnly:** True
- **enum:** ['GLOBAL', 'LOCAL']

activeSettingsPolicies

- **type:** array
- **description:** List of Settings Policies that are affecting the final value of this setting
- – **reference:** SettingActivePolicyApiDTO *(on page 1670)*

sourceGroupName

- **type:** string
- **description:** Source group name for setting value

sourceGroupUuid

- **type:** string
- **description:** Source group uuid for setting value

SettingOptionApiDTO

Description

Optional Parameters

label

- **type:** string

value

- **type:** string

SettingsManagerApiDTO

Description

Model to describe a Category Manager for the settings

Required Parameters

category

- **type:** string
- **description:** Category of the Manager

settings

- **type:** array
- **description:** List of Settings contained in the Manager
- – **reference:** SettingApiDTOSerializable *(on page 1671)*

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

SettingsPolicyApiDTO

Description

Model to describe a Setting Policy: a group of Settings that are applied based on Scope and/or Schedule

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

entityType

- **type:** string
- **description:** Type of the Settings Policy. E.G. VirtualMachine, PhysicalMachine, ...

scopes

- **type:** array
- **description:** List of scopes where the settings are applied, on create/edit, only Uuid is required. If null it's considered Global scope
- – **reference:** GroupApiDTO *(on page 1564)*

settingsManagers

- **type:** array
- **description:** List of groups of Settings to activate in this Policy
- – **reference:** SettingsManagerApiDTO *(on page 1677)*

onPremCostsManagers

- **type:** array
- **description:** List of groups of Settings to activate for on prem costs policy
- – **reference:** OnPremCostsManagerApiDTO *(on page 1597)*

schedule

- **description:** Schedule to define when the settings are applied
- **reference:** ScheduleApiDTO *(on page 1660)*

disabled

- **type:** boolean
- **description:** Control if the policy is Enabled or Disabled

note

- **type:** string
- **description:** Note for this Policy

readOnly

- **type:** boolean
- **description:** If true, prevents users from deleting or modifying the SettingPolicy.

default

- **type:** boolean

SiteInformationApiDTO

Description

Hybrid SaaS Network Site information

Optional Parameters

tunnelComponents

- **description:** Skupper components image version
- **reference:** TunnelComponents *(on page 1707)*

siteId

- **type:** string
- **description:** Site Id

clientOperatorVersion

- **type:** string
- **description:** Client Operator Image version

autoUpdateEnabled

- **type:** boolean
- **description:** Is Auto Update Enabled?

lastUpdated

- **type:** integer
- **format:** int64
- **description:** Site Information Last Updated in unix epoch seconds format

probes

- **type:** array
- **description:** List of Probes
- – **reference:** ProbeDeploymentApiDTO *(on page 1630)*

SkippedActionApiDTO

Description

Model to describe an action that was skipped in execution.

Required Parameters

actionId

- **type:** string
- **description:** Action ID

Optional Parameters

reason

- **type:** string
- **description:** Reason why action was skipped

StatApiDTO

Description

Model to describe a Statistic

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

name

- **type:** string
- **description:** Name

capacity

- **description:** Capacity values, used for Commodities stats
- **reference:** StatValueApiDTO *(on page 1685)*

reserved

- **description:** Reserved values, used for Commodities stats
- **reference:** StatValueApiDTO *(on page 1685)*

relatedEntityType

- **type:** string
- **description:** E.G.virtual machine

filters

- **type:** array
- **description:** describe the grouping options used to generate the output
- – **reference:** StatFilterApiDTO *(on page 1682)*

relatedEntity

- **description:** Provider or Consumer of the Statistic, 'relatedEntity' and 'numRelatedEntities' are mutually exclusive
- **reference:** BaseApiDTO *(on page 1509)*

numRelatedEntities

- **type:** integer
- **format:** int32
- **description:** Number of providers or consumers for this statistic, 'relatedEntity' and 'numRelatedEntities' are mutually exclusive

units

- **type:** string
- **description:** Units, used for Commodities stats. E.G. $/h

values

- **description:** Contains Average, Min, Max, Total
- **reference:** StatValueApiDTO *(on page 1685)*

value

- **type:** number
- **format:** float
- **description:** Simple value, equal to values.avg.

commoditySource

- **description:** Source for a commodity's capacity and usage, used for Commodities stats
- **reference:** StatCommoditySourceApiDTO *(on page 1682)*

histUtilizations

- **type:** array
- **description:** Commodity historical utilization values
- – **reference:** StatHistUtilizationApiDTO *(on page 1683)*

StatApiInputDTO

Description

Model to describe the request of a Stat. Please note that the costPrice stat is deprecated and will EOL in 8.12.1. Use the dedicated 'cost' endpoints in /entities, /groups and /markets instead.

Optional Parameters

name

- **type:** string
- **description:** Name of the stat

relatedEntityType

- **type:** string
- **description:** Used to get stats from different entity types in the supply chain of {uuid} E.G.costPrice of the VMs where the scope is a PM

filters

- **type:** array
- **description:** Filters to apply to the requested stats. Commodities supported values: key = {commodity_key}, relation = sold|bought, virtualDisk = {uuid}. Cost Price supported values: cumulative = true|false, cloudService = {uuid}, target = {uuid}, CSP = {name}, category = {name}, businessUnit = {uuid}, costComponent = COMPUTE|IP|LICENSE|STORAGE, costComponentArtifact = {key}. Please note that the costPrice stat is deprecated and will EOL in 8.12.1. Use the dedicated 'cost' endpoints in /entities, /groups and /markets instead.Action stats supported values: actionTypes = {actionType}, actionModes: {action_mode}, actionStates: {action_state}, riskSubCategory: Performance Assurance|Efficiency Improvement|Prevention|Compliance, riskSeverity: UNKNOWN|NORMAL|MINOR|MAJOR|CRITICAL
- – **reference:** StatFilterApiDTO *(on page 1682)*

groupBy

- **type:** array
- **description:** Execute a groupBy on the values of the stats. Commodities supported values: key, relatedEntity, virtualDisk. Cost Price supported values: cloudService, target, CSP, category, businessUnit, costComponent, costComponentArtifact. Please note that the costPrice stat is deprecated and will EOL in 8.12.1. Use the dedicated 'cost' endpoints in /entities, /groups and /markets instead.Action stats supported values: actionTypes, actionModes, actionStates, risk, riskSubCategory, riskSeverity
- – **type:** string

historyType

- **type:** string
- **description:** Query the histUtilization table for these historyTypes
- **enum:** ['Percentile', 'Smoothed']

StatCommoditySourceApiDTO

Description

Model to describe commodity source for a single statistic, e.g. commodity source for VCPU (capacity and usage)

Optional Parameters

capacityCommoditySource

- **type:** string
- **description:** Commodity source for capacity

usedCommoditySource

- **type:** string
- **description:** Commodity source for usage

capacitySourceIndex

- **type:** integer
- **format:** int32
- **description:** Commodity source operation index for capacity

usedSourceIndex

- **type:** integer
- **format:** int32
- **description:** Commodity source operation index for usage

StatFilterApiDTO

Description

Model to describe the filter applied to a Statistic

Required Parameters

type

- **type:** string
- **description:** type of the filter, E.G: actionTypes, category, ...

Optional Parameters

value

- **type:** string
- **description:** value of the filter

displayName

- **type:** string
- **description:** display name of the value, E.G: displayName if 'value' is an oid or an enum

StatHistUtilizationApiDTO

Description

Model to describe a historical utilization statistic, e.g. the P95 vCPU or P95 vMEM of a VM over (up to) NN days observation period.

Optional Parameters

type

- **type:** string
- **description:** Historical utilization type

usage

- **type:** number
- **format:** float
- **description:** Historical usage

capacity

- **type:** number
- **format:** float
- **description:** Historical capacity

resizeMaxScalingObservationPeriod

- **type:** number
- **format:** float
- **description:** Resize max scaling observation period. Affects on amount of data points considered for action generation process. Value in days.

resizeScalingAggressiveness

- **type:** number
- **format:** float
- **description:** Resize scaling aggressiveness. Describes how aggressively Turbonomic will resize in response of resource utilization. Value in percents. 100% least aggressive, 95% most aggressive.

StatPeriodApiInputDTO

Description

Model to describe the request for Statistics by a Time range

Required Parameters

statistics

- **type:** array
- **description:** List of requested statistics, if empty all stats will be processed
- – **reference:** StatApiInputDTO *(on page 1681)*

Optional Parameters

startDate

- **type:** string
- **description:** Start time of the period to filter, if empty the current stats will be processed

endDate

- **type:** string
- **description:** End time of the period to filter, if empty the current stats will be processed

StatScopesApiInputDTO

Description

Model to describe the request for Statistics in multiple Scopes

Optional Parameters

scopes

- **type:** array
- **description:** List of uuid to use as scope
- – **type:** string

period

- **description:** List of requested statistics
- **reference:** StatPeriodApiInputDTO *(on page 1683)*

relatedType

- **type:** string
- **description:** Get the stats for the entity type related to the scopes

StatSnapshotApiDTO

Description

Model to describe the Statistics of a snapshot, contains a list of Statistics for a moment in time

Required Parameters

statistics

- **type:** array
- **description:** List of statistics, e.g. Price Index, Commodities values, Cost price, Number of VMs in a Group, ...
- – **reference:** StatApiDTO *(on page 1679)*

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

date

- **type:** string
- **description:** When the statistics were recorded

epoch

- **type:** string
- **description:** An indicator of whether the snapshot represents historical, projected or plan data. Should be used in conjunction with the date field to understand the context of this stat snapshot.
- **enum:** ['HISTORICAL', 'CURRENT', 'PROJECTED', 'PLAN_SOURCE', 'PLAN_PROJECTED']

StatValueApiDTO

Description

Model to describe a single statistic, e.g. the CPU of a PM Group

Optional Parameters

max

- **type:** number
- **format:** float
- **description:** Max value

min

- **type:** number
- **format:** float
- **description:** Min value

avg

- **type:** number
- **format:** float
- **description:** Average value

total

- **type:** number
- **format:** float
- **description:** Total value - sum of average values in a group

totalMax

- **type:** number
- **format:** float
- **description:** Total max value - sum of max values in a group

totalMin

- **type:** number
- **format:** float
- **description:** Total min value - sum of min values in a group

StringListFilterApiDTO

Description

Model to describe a filter by a string list and a operator.

Required Parameters

values

- **type:** array
- **description:** String list for a comparison.
- – **type:** string

Optional Parameters

operator

- **type:** string
- **description:** Operator for a comparison.
- **enum:** ['EQ', 'NEQ']

caseSensitive

- **type:** boolean
- **description:** If true, the query string will match case-sensitivity.

StringSettingApiDTO

Description

Inherits From

This DTO inherits from SettingApiDTO *(on page 1675)*

Required Parameters

value

- **type:** string
- **description:** Value of the setting. Type indicated by the `valueObjectType` field. The default type is `String`

Optional Parameters

defaultValue

- **type:** string
- **description:** Default value before any change
- **readOnly:** True

SupplyChainStatsApiInputDTO

Description

Model to describe a Filter for Entities

Required Parameters

uuids

- **type:** array
- **description:** List of uuids used to build the supply chain
- – **type:** string
- **maxItems:** 2147483647
- **minItems:** 1

Optional Parameters

types

- **type:** array
- **description:** List of Types to filter
- – **type:** string

states

- **type:** array
- **description:** List of States to filter
- – **type:** string
  – **enum:** ['IDLE', 'RESOURCE_ALLOCATION', 'LAUNCH', 'ACTIVE', 'RESOURCE_RELEASE', 'SUSPEND', 'SUSPEND_PENDING', 'TERMINATE_PENDING', 'NOT_MONITORED', 'MAINTENANCE', 'FAILOVER', 'UNKNOWN', 'EVACUATED', 'QUEUED', 'NOT_FOUND']

groupBy

- **type:** array
- **description:** List of fields used to group the Stats
- – **type:** string
  – **enum:** ['entityType', 'state', 'severity', 'riskSubCategory', 'template', 'target', 'businessUnit', 'resourceGroup', 'nodePool']

environmentType

- **type:** string
- **description:** Filter the actions by Environment Type
- **enum:** ['ONPREM', 'CLOUD']

SupplychainApiDTO

Description

Model to describe the Supply Chain: the Environment is described as a chain between Producers and Consumers

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

seMap

- **type:** object
- **description:** Model to describe the type of Entity associated with a SupplyChainEntry
- **additionalProperties:** {'$ref': '#/definitions/SupplychainEntryDTO'}

SupplychainEntryDTO

Description

Model to describe a Supply Chain Entry

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

depth

- **type:** integer
- **format:** int32
- **description:** Number to indicate the level of depth of the Entity type in the Supply Chain

entitiesCount

- **type:** integer
- **format:** int32
- **description:** Number of entity in this level of depth

healthSummary

- **type:** object
- **description:** Entities count by Severity
- **additionalProperties:** {'type': 'integer', 'format': 'int32'}

stateSummary

- **type:** object
- **description:** Entities count by State
- **additionalProperties:** {'type': 'integer', 'format': 'int32'}

connectedProviderTypes

- **type:** array
- **description:** List of provider types connected to this level of depth
- **uniqueItems:** True
- – **type:** string

connectedConsumerTypes

- **type:** array
- **description:** List of consumer types connected to this level of depth
- **uniqueItems:** True
- – **type:** string

instances

- **type:** object
- **description:** List of entities in this level of depth
- **additionalProperties:** {'$ref': '#/definitions/ServiceEntityApiDTO'}

SustainabilityInputApiDTO

Description

Model to describe the request for Sustainability Metrics by a Time range

Optional Parameters

sustainabilityGroupBys

- **type:** array
- **description:** List of values to group the results by.
- – **type:** string
- – **enum:** ['ACCOUNT', 'AVAILABILITY_ZONE', 'ENTITY', 'ENTITY_TYPE', 'REGION', 'SERVICE_PROVIDER']
- **maxItems:** 1
- **minItems:** 0

startDate

- **type:** string
- **description:** Start time of the period to filter. If startDate and endDate not provided current metrics will be processed

endDate

- **type:** string
- **description:** End time of the period to filter. If startDate and endDate not provided current metrics will be processed

TagApiDTO

Description

Tag information

Optional Parameters

key

- **type:** string
- **description:** Tag key

values

- **type:** array
- **description:** Tag value
- – **type:** string

TagKeyApiDTO

Description

Optional Parameters

key

- **type:** string
- **description:** Tag key

TargetApiDTO

Description

Model to describe a Target. Targets are used to discover the components to create a connected virtual environment

Required Parameters

type

- **type:** string
- **description:** Probe type. Turbonomic ships with the following probe types: AppDynamics, Dynatrace, Datadog, Apptio Cloudability, Arista, AWS, Azure, Cisco APIC, Cisco HyperFlex, Cisco UCS Central, Cisco UCS Director, Cisco UCS Manager, CloudFoundry, CloudStack, Dell Compellent, EMC ScaleIO, EMC VMAX, EMC VNX, EMC VPLEX, EMC XtremIO, Hitachi Vantara, HP OneView, HPE 3PAR, Hyper-V, IBM PowerVM, Instana, JBoss, JVM, MSExchange, MySQL, NetApp, NetFlow, NetScaler, Nutanix, OpenStack, Oracle, Pivotal Ops Manager, Pure, Red Hat Virtualization Manager, sFlow, SNMP, SoftLayer, SQLServer, Tetration, Tomcat, vCenter, vCloudDirector, VMM, WebLogic, WebSphere, WMI, XenServer,

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

category

- **type:** string
- **example:** Hypervisor
- **description:** Probe category.

isProbeRegistered

- **type:** boolean
- **example:** False
- **description:** Indicator that is used to determine whether the associated probe is running and registered with the system

uiCategory

- **type:** string
- **example:** Application Server, Billing, Cloud Management, Applications and Databases, Fabric and Network, Guest OS Processes, HYPERCONVERGED, HYPERVISOR, Load Balancer, Operations Manager Appliance, ORCHESTRATOR, PaaS, Storage, Storage Browsing, WINDOWS APPLICATION. There are also categories Custom and UNKNOWN to be used for new probe development.
- **description:** Probe UI category.

identifyingFields

- **type:** array
- **description:** Array of field names used to differentiate probes of the same category. Most typically either the name/address field, or a combination of address and port.
  - **type:** string

inputFields

- **type:** array
- **description:** List of all InputFieldApiDTO objects for the target, analogous to the target addition form fields in the GUI.
  - **reference:** InputFieldApiDTO *(on page 1571)*

lastValidated

- **type:** string
- **description:** Date of the last validation - [DEPRECATED - Use: healthSummary.timeOfLastSuccessfulDiscovery]

status

- **type:** string
- **example:** Validated, VALIDATION FAILED, UNKNOWN
- **description:** Description of the status of the most recent Validation Operation - [DEPRECATED - Use: healthSummary.healthState]

derivedTargets

- **type:** array
- **description:** Targets derived (created by discovery of) from this target. Populated depending on the TargetRelationship parameter.
  - **reference:** TargetApiDTO *(on page 1690)*

patchedTargets

- **type:** array
- **description:** Targets patched by this target
- – **reference:** PatchedTargetDataApiDTO *(on page 1617)*

parentTargets

- **type:** array
- **description:** Targets that created this target. Populated depending on the TargetRelationship parameter.
- – **reference:** TargetApiDTO *(on page 1690)*

subType

- **type:** string
- **example:** RedHatOpenShift
- **description:** Target subType.

readonly

- **type:** boolean
- **description:** Whether the target cannot be changed through public APIs.

health

- **description:** [BETA]: A summary of the overall health of the target. Only returned if TargetDetailLevel is HEALTH or HEALTH_DETAILS
- **reference:** TargetHealthApiDTO *(on page 1693)*

lastTargetOperationStages

- **type:** array
- **description:** [BETA]: The status at each stage of the most recent validation or discovery, which ever was more recent. Only returned if TargetDetailLevel is HEALTH_DETAILS
- – **reference:** TargetOperationStageApiDTO *(on page 1694)*

lastEditTime

- **type:** string
- **example:** 2021-07-07T20:33:49Z
- **description:** [BETA]: The last time the target was edited in ISO-8601 Format (YYYY-MM-DDTHH:MM:SSZ). For example 2021-07-07T20:33:49Z

lastEditUser

- **type:** string
- **description:** [BETA]: The username for last user who edited the target.

healthSummary

- **description:** The health summary of the target.
- **reference:** TargetHealthSummaryApiDTO *(on page 1694)*

TargetErrorDetailsApiDTO

Description

Additional details related to a target error.

Required Parameters

targetErrorDetailsClass

- **type:** string

TargetHealthApiDTO

Description

A DTO that describes the health of a single target.

Required Parameters

healthCategory

- **type:** string
- **description:** The type of the health checks that are done.
- **enum:** ['TARGET', 'ACTION', 'COMPONENT']

healthClassDiscriminator

- **type:** string
- **example:** TargetHealthApiDTO
- **description:** Health type discriminator. Describes which subclass of BaseHealthApiDTO is returned.

rollupState

- **type:** string
- **example:** NORMAL
- **description:** The health state of the target including its derived targets.
- **enum:** ['CRITICAL', 'MAJOR', 'MINOR', 'NORMAL']

targetName

- **type:** string
- **description:** Target display name.

uuid

- **type:** string
- **description:** UUID of the target.

Optional Parameters

errorText

- **type:** string
- **description:** Error text (if there's any).

timeOfFirstFailure

- **type:** string
- **description:** Time of first failure of validation/discovery (if the target is in failed state; ISO-8601 formatted String).

timeOfLastSuccessfulDiscovery

- **type:** string
- **example:** 2021-07-07T20:33:49Z
- **description:** [BETA]: The last time a successful discovery finished on the target in ISO-8601 Format (YYYY-MM-DDTHH:MM:SSZ). For example 2021-07-07T20:33:49Z

targetErrorDetails

- **type:** array
- **description:** Error details related to a target. Supports multiple errors details.
- – **reference:** TargetErrorDetailsApiDTO *(on page 1692)*

targetStatusSubcategory

- **type:** string
- **enum:** ['DISCOVERY', 'VALIDATION', 'DUPLICATION', 'DELAYED_DATA']

healthState

- **type:** string
- **enum:** ['CRITICAL', 'MAJOR', 'MINOR', 'NORMAL']

TargetHealthSummaryApiDTO

Description

A DTO that describes target health.

Required Parameters

healthState

- **type:** string
- **example:** NORMAL
- **description:** The target health state.
- **enum:** ['CRITICAL', 'MAJOR', 'MINOR', 'NORMAL']

rollupState

- **type:** string
- **example:** NORMAL
- **description:** The health state of the target including its derived targets.
- **enum:** ['CRITICAL', 'MAJOR', 'MINOR', 'NORMAL']

Optional Parameters

timeOfLastSuccessfulDiscovery

- **type:** string
- **example:** 2021-07-07T20:33:49Z
- **description:** The last time a successful discovery finished on the target in ISO-8601 Format (YYYY-MM-DDTHH:MM:SSZ). For example 2021-07-07T20:33:49Z

TargetOperationStageApiDTO

Description

Holds all information related to a stage of discovery or validation.

Optional Parameters

description

- **type:** string
- **description:** The description of what this stage is doing (ideally in less that 70 characters). Example of description are "validating credentials" or "retrieving business applications".

status

- **description:** Holds all the information of this particular instance of the discovery or validation stage.
- **reference:** TargetOperationStageStatusApiDTO *(on page 1695)*

subStages

- **type:** array
- **description:** List with the subStages of he current stage.
- – **reference:** TargetOperationStageApiDTO *(on page 1694)*

optional

- **type:** boolean
- **description:** Is stage marked as optional.

TargetOperationStageStatusApiDTO

Description

Holds all the information of this particular instance of the discovery or validation stage.

Optional Parameters

state

- **type:** string
- **description:** The state of this stage at the end of the discovery or validation.
- **enum:** ['DID_NOT_RUN', 'SUCCESS', 'FAILURE']

summary

- **type:** string
- **description:** The short message (ideally 70 characters or less) explaining the stage status. Example of such messages are "invalid credentials" or "4 business application retrieved".

fullExplanation

- **type:** string
- **description:** The full message explaining the stage status. This explanation may include the instructions or hints on how to address issue if there is one.

TargetPropertiesApiDTO

Description

Model for registered Target properties (categories, types, subtypes)

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

categories

- **type:** array
- **example:** Container Platform
- **description:** Target categories
- **uniqueItems:** True
- – **type:** string

types

- **type:** array
- **example:** Kubernetes
- **description:** Target types
- **uniqueItems:** True
- – **type:** string

subtypes

- **type:** array
- **example:** RedHatOpenShift
- **description:** Target subtypes
- **uniqueItems:** True
- – **type:** string

TargetThirdPartyErrorDetailsApiDTO

Description

Inherits From

This DTO inherits from TargetErrorDetailsApiDTO *(on page 1692)*

Optional Parameters

targetErrorType

- **type:** string
- **description:** Target validation/discovery error type.
- **enum:** ['INTERNAL_PROBE_ERROR', 'CONNECTIVITY_ERROR', 'UNAUTHENTICATED', 'TOKEN_UNAVAILABLE', 'VERSION_NOT_SUPPORTED', 'DATA_ACCESS_ERROR', 'DUPLICATION', 'DELAYED_DATA', 'THIRD_PARTY_FAILURE']

errorCode

- **type:** string
- **example:** 0x423232, 404, 400
- **description:** Error code related to the failure.

endPoint

- **type:** string
- **description:** End point which caused the third party api failure. This can be a complete URLor a method name while using the third party SDK.

TelemetryDTO

Description

Model to describe the status of telemetry

Optional Parameters

telemetryTermsViewed

- **type:** boolean
- **description:** Controls when an administrator should be shown the telemetry opt in

telemetryEnabled

- **type:** boolean
- **description:** Controls whether analytics and telemetry are enabled

tenantId

- **type:** string
- **description:** [BETA]: tenantId for Pendo agent. Only available in SAAS_TRIAL deploymentMode

ibmUniqueId

- **type:** string
- **description:** [BETA]: ibmUniqueId (also called iui) for Pendo agent. Only available in SAAS_TRIAL deploymentMode

TemplateApiDTO

Description

Model to describe a Template: contains the resources used to Deploy a VM or to Add Workload/Supply in a Plan

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

price

- **type:** number
- **format:** float
- **description:** The estimated cost of the template. Used for Physical Machine and Storage templates where the hardware cost is known.

model

- **type:** string
- **example:** Cisco UCS B200 M5 (Intel Xeon Gold 5115, 2.40 GHz), PowerEdge R940 (Intel Xeon Gold 6136, 3.00 GHz), etc.
- **description:** The model of the entire Physical Machine system.

cpuModel

- **type:** string
- **example:** Intel Xeon Gold 6130, AMD Opteron 6166 HE, etc.
- **description:** The model of the CPU in the Physical Machine.

vendor

- **type:** string
- **example:** Cisco, Microsoft, HP, Novell
- **description:** Vendor of the templated entity

description

- **type:** string
- **description:** Description

image

- **type:** string
- **description:** Container Profile image

imageTag

- **type:** string
- **description:** (UNUSED) Container Profile image tag

cmdWithArgs

- **type:** string
- **description:** (UNUSED) Container command with Arguments

dbEngine

- **type:** string
- **example:** PostgreSql, AuroraMySQL, SqlServer
- **description:** Database engine, used for Database templates

dbEdition

- **type:** string
- **example:** Standard One, Enterprise, Premium_P2
- **description:** Database edition, used for Database templates

computeResources

- **type:** array
- **description:** ResourceApiDTO describing the compute resources. Max 1 per template.
- – **reference:** ResourceApiDTO *(on page 1651)*
- **maxItems:** 1
- **minItems:** 0

storageResources

- **type:** array
- **description:** ResourceApiDTO describing the storage resources
- – **reference:** ResourceApiDTO *(on page 1651)*

networkResources

- **type:** array
- **description:** ResourceApiDTO describing the network resources
- – **reference:** ResourceApiDTO *(on page 1651)*

infrastructureResources

- **type:** array
- **description:** ResourceApiDTO describing the infrastructure resources. Max 1 per template
- – **reference:** ResourceApiDTO *(on page 1651)*
- **maxItems:** 1
- **minItems:** 0

deploymentProfile

- **description:** Deployment profile associated with the template
- **reference:** DeploymentProfileApiDTO *(on page 1550)*

discovered

- **type:** boolean
- **example:** False
- **description:** Indicates if the template is discovered or manually created
- **enum:** [True, False]

family

- **type:** string
- **example:** standardDSv2Family, standardNCFamily, r5d, r3
- **description:** For cloud templates, the instance family.

enableMatch

- **type:** boolean
- **example:** False
- **description:** Add to Infrastructure Cost Policy. Infrastructure Cost policies group hardware devices according to their cost

TemplateApiInputDTO

Description

Model to describe a Template: contains the resources used to Deploy a VM or to Add Workload/Supply in a Plan

Optional Parameters

displayName

- **type:** string
- **description:** Name of the template

className

- **type:** string
- **description:** className
- **enum:** ['VirtualMachine', 'PhysicalMachine', 'Storage', 'Container']

price

- **type:** number
- **format:** float
- **description:** Cost price

model

- **type:** string
- **description:** The Model of the entire Physical Machine system. For example: Cisco UCS B200 M5 (Intel Xeon Gold 5115, 2.40 GHz) vs. PowerEdge R940 (Intel Xeon Gold 6136, 3.00 GHz)

cpuModel

- **type:** string
- **description:** The model of the CPU in the Physical Machine. This is separate from the model of the entire Physical Machine. For example: Intel Xeon Gold 6130 vs. AMD Opteron 6166 HE

vendor

- **type:** string
- **description:** Vendor

description

- **type:** string
- **description:** Description

image

- **type:** string
- **description:** Profile image, used for Container templates

imageTag

- **type:** string
- **description:** Profile image tag, used for Container templates

cmdWithArgs

- **type:** string
- **description:** Command with Arguments, used for Container templates

computeResources

- **type:** array
- **description:** Compute resources: Number of CPU, CPU speed, Memory size, ...
- – **reference:** ResourceApiDTO *(on page 1651)*
- **maxItems:** 1
- **minItems:** 0

storageResources

- **type:** array
- **description:** Storage resources: Disk I/0, Disk Size, Percentage of Disk consumed
- – **reference:** ResourceApiDTO *(on page 1651)*

networkResources

- **type:** array
- **description:** Network resources
- – **reference:** ResourceApiDTO *(on page 1651)*

infrastructureResources

- **type:** array
- **description:** Infrastructure resources: Power, Size, Cooling
- – **reference:** ResourceApiDTO *(on page 1651)*
- **maxItems:** 1
- **minItems:** 0

deploymentProfileId

- **type:** string
- **description:** Ids of the Deployment Profiles associated with this Template

enableMatch

- **type:** boolean
- **description:** Add to Infrastructure Cost Policy. Infrastructure Cost policies group hardware devices according to their cost

TemplatePriceAdjustmentDTO

Description

Price adjustment information for a template within a business unit and specified pricing model.

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

environmentType

- **type:** string
- **description:** Environment type
- **enum:** ['CLOUD', 'ONPREM', 'HYBRID']

discount

- **type:** number
- **format:** float
- **description:** Entity discount percentage

priceAdjustment

- **description:** Entity price adjustment percentage
- **reference:** PriceAdjustmentDTO *(on page 1629)*

family

- **type:** string
- **description:** Family of a template, group of similar purpose templates. Available only for virtual machine templates.

pricesPerDatacenter

- **type:** array
- **description:** Template base rates per data center, provided for information purpose.
- – **reference:** EntityPriceDTO *(on page 1557)*

TemplateValidationStatusApiDTO

Description

Model to define the template validation status data.

Required Parameters

validationStatus

- **type:** string
- **description:** The ValidationStatus is the result of template validation.
- **enum:** ['NOTFOUND', 'INVALID', 'VALID']

Optional Parameters

validationErrors

- **type:** array
- **description:** The list of validation errors by action type.
- – **type:** string

lastTemplateValidationTime

- **type:** string
- **description:** The lastTemplateValidationTime is the time in ISO-8601 format when the Webhook template was last validated.

TextConditionApiDTO

Description

Inherits From

This DTO inherits from ConditionApiDTO *(on page 1535)*

Required Parameters

value

- **type:** string
- **description:** A regular expression to be matched against the value of the field.
- **readOnly:** True

TextFieldValueApiDTO

Description

Inherits From

This DTO inherits from FieldValueApiDTO *(on page 1562)*

Required Parameters

value

- **type:** string
- **description:** The value.
- **readOnly:** True

TimeBasedTopologyChangesApiDTO

Description

Topology changes based on a timeframe in the scenario

Optional Parameters

includeReservation

- **type:** boolean
- **description:** Add the workload that is currently reserved

reservationCount

- **type:** integer
- **format:** int32
- **description:** # of currently reserved VMs
- **readOnly:** True

addHistoryVMs

- **type:** boolean
- **description:** Should historical vmNetGrowth be considered

vmNetGrowth

- **type:** integer
- **format:** int32
- **description:** Number of workloads added during the last month
- **readOnly:** True

TimeSpanApiDTO

Description

Schedule Time span information

Required Parameters

begins

- **type:** string
- **description:** Start time of the time span
- **pattern:** ^([0-1]?[0-9]|2[0-3]):[0-5][0-9]$

ends

- **type:** string
- **description:** End time of the time span
- **pattern:** ^([0-1]?[0-9]|2[0-3]):[0-5][0-9]$

policy

- **description:** policy to be applied during this time span
- **reference:** ScheduleItemApiDTO *(on page 1662)*

TimeSpanGroupApiDTO

Description

Base model to describe a TimeSpanGroup.

Required Parameters

type

- **type:** string
- **example:** WeekDayTimeSpansApiDTO
- **description:** TimeSpan group type discriminator. Describes which subclass of TimeSpanGroupApiDTO is returned.

TopologyChangesApiDTO

Description

Topology changes in the scenario

Optional Parameters

addList

- **type:** array
- **description:** List of add changes in the Topology
- – **reference:** AddObjectApiDTO *(on page 1503)*

removeList

- **type:** array
- **description:** List of remove changes in the Topology
- – **reference:** RemoveObjectApiDTO *(on page 1643)*

replaceList

- **type:** array
- **description:** List of replace changes in the Topology
- – **reference:** ReplaceObjectApiDTO *(on page 1643)*

migrateList

- **type:** array
- **description:** List of migrate changes in the Topology
- – **reference:** MigrateObjectApiDTO *(on page 1588)*

relievePressureList

- **type:** array
- **description:** List of relieve pressure changes in the Topology
- – **reference:** RelievePressureObjectApiDTO *(on page 1641)*

TopologyDataDefinitionApiDTO

Description

Model to describe a user-defined topology definition.

Required Parameters

entityDefinitionData

- **description:** Entity definition data
- **reference:** IEntityDefinitionData *(on page 1570)*

entityType

- **type:** string
- **description:** The type of service entity being defined
- **enum:** ['BusinessApplication', 'BusinessTransaction', 'Service']

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

customEntityUuidList

- **type:** array
  - **type:** string

discoveryState

- **type:** string
- **description:** Discovery status for application topology.
- **enum:** ['IN_PROGRESS', 'CREATED', 'DETACHED']

contextBased

- **type:** boolean
- **description:** Defines if this Definition is Context-Based.

TopologyScopeApiDTO

Description

Model to describe cloud commitment's topology scoped attributes

Optional Parameters

cloudServiceId

- **type:** string
- **description:** The cloud service ID of the cloud commitment, e.g. aws::CS::AmazonRDS

cloudTierId

- **type:** string
- **description:** The cloud tier ID of the cloud commitment, e.g. db.t3.micro

tenancy

- **type:** string
- **description:** The tenancy of the cloud commitment, e.g. shared or dedicated
- **enum:** ['DEFAULT', 'DEDICATED', 'HOST']

familyScope

- **description:** The family scope of the cloud commitment if applicable
- **reference:** FamilyScopeApiDTO *(on page 1562)*

databaseServerTierInfo

- **description:** The database server tier info of the cloud commitment if applicable
- **reference:** DatabaseServerTierInfoApiDTO *(on page 1545)*

TunnelComponents

Description

Optional Parameters

router

- **type:** string
- **description:** Skupper Router image information

siteController

- **type:** string
- **description:** Skupper Site Controller image information

serviceController

- **type:** string
- **description:** Skupper service controller image information

configSync

- **type:** string
- **description:** Skupper Config Sync image information

UnplacementDetailsApiDTO

Description

Additional information provided for entities in a plan that are unplaced

Required Parameters

reasons

- **type:** array
- **description:** Detailed information about why the entity could not be placed
- – **reference:** UnplacementReasonApiDTO *(on page 1707)*

Optional Parameters

placedOn

- **type:** array
- **description:** For Unplaced entities in a Plan Market, the providers that this entity was able to place on
- – **reference:** BaseApiDTO *(on page 1509)*

UnplacementReasonApiDTO

Description

Details to describe why an entity could not be placed

Optional Parameters

failedResources

- **type:** array
- **description:** A list of resources that caused the entity not to be placed
- – **reference:** FailedResourceApiDTO *(on page 1561)*

resourceOwner

- **description:** The entity owning the resources that caused unplacement, if different than this entity. For example, a volume that caused a VM to be unplaced.
- **reference:** BaseApiDTO *(on page 1509)*

providerType

- **type:** string
- **description:** The entity type that sells the commodities that caused unplacement.
- **enum:** ['Application', 'ApplicationComponent', 'ApplicationServer', 'AvailabilityZone', 'BusinessAccount', 'BusinessApplication', 'BusinessTransaction', 'BusinessUser', 'Chassis', 'CloudService', 'ComputeTier', 'Container', 'ContainerPod', 'ContainerSpec', 'Database', 'DatabaseServer', 'DatabaseServerTier', 'DatabaseTier', 'DataCenter', 'DesktopPool', 'DiskArray', 'DPod', 'HCIPhysicalMachine', 'HypervisorServer', 'Internet', 'IOModule', 'LoadBalancer', 'LogicalPool', 'Namespace', 'Network', 'PhysicalMachine', 'ProcessorPool', 'Region', 'ReservedInstance', 'Service', 'ServiceProvider', 'Storage', 'StorageController', 'StorageTier', 'Switch', 'Unknown', 'ViewPod', 'VirtualApplication', 'VirtualDataCenter', 'VirtualMachine', 'VirtualVolume', 'VPod', 'WorkloadController', 'ContainerPlatformCluster', 'CloudCommitment', 'ApplicationComponentSpec', 'VirtualMachineSpec', 'DocumentCollection', 'NodeGroup', 'PowerVMPhysicalMachine', 'DataWarehouse', 'DataWarehouseTier']

closestSeller

- **description:** The seller that came closest to satisfying the commodities. This information is not always available.
- **reference:** BaseApiDTO *(on page 1509)*

placementProblem

- **type:** string
- **description:** Categorizes the reason which caused unplacement
- **enum:** ['COSTS_NOT_FOUND', 'NOT_CONTROLLABLE', 'UNSATISFIED_COMMODITIES']

UserApiDTO

Description

Model to describe an User

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

username

- **type:** string
- **description:** Username

password

- **type:** string
- **description:** Encoded password

roles

- **type:** array
- **description:** Roles assigned to the user.
  - **reference:** RoleApiDTO *(on page 1652)*

loginProvider

- **type:** string
- **description:** Provider for the login
- **enum:** ['Local', 'LDAP']

type

- **type:** string
- **description:** type
- **enum:** ['DedicatedCustomer', 'SharedCustomer']

scope

- **type:** array
- **description:** List of scopes that the User is allowed to see
  - **reference:** GroupApiDTO *(on page 1564)*

atdScope

- **type:** array
- **description:** List of Application Topology Definitions that the User is allowed to see
  - **reference:** BaseApiDTO *(on page 1509)*

features

- **type:** array
- **description:** List of features that the user is allowed to see
  - **type:** string

authToken

- **type:** string
- **description:** The secured user token. [DEPRECATED - EOL: 8.15.1]

showSharedUserSC

- **type:** boolean

UtilizationApiDTO

Description

Model to describe a utilization change in the scenario

Optional Parameters

projectionDay

- **type:** integer
- **format:** int32
- **description:** Projection day, indicate when to apply the change

target

- **description:** Target of the change, entity or group
- **reference:** BaseApiDTO *(on page 1509)*

percentage

- **type:** integer
- **format:** int32
- **description:** Utilization percentage change

VMEntityAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

os

- **type:** string
- **description:** The operating system type. E.G. 'Linux', 'Windows' and 'Unknown' if not found.

ip

- **type:** array
- **description:** The IP address. Can be IPv4 or IPv6
  - **type:** string

connectedNetworks

- **type:** array
- **description:** The connected Networks
  - **reference:** BaseApiDTO *(on page 1509)*

numVCPUs

- **type:** integer
- **format:** int32
- **description:** Number of vCPUs

acceleratorCards

- **type:** array
- **description:** The Accelerator/GPU cards
  - **reference:** AcceleratorCardApiDTO *(on page 1484)*

coresPerSocketRatio

- **type:** integer
- **format:** int32
- **description:** Cores per socket ratio

processorPoolName

- **type:** string
- **description:** Processor Pool

hasDedicatedProcessors

- **type:** boolean
- **example:** True
- **description:** Has Dedicated Processors

processorCompatibilityMode

- **type:** string
- **example:** POWER8
- **description:** Processor Compatibility Mode

activeMemoryExpansionEnabled

- **type:** boolean
- **example:** True
- **description:** Active Memory Expansion Enabled

resourceMonitoringControlState

- **type:** string
- **example:** active
- **description:** Resource Monitoring Control (RMC) State

sharingMode

- **type:** string
- **example:** uncapped
- **description:** Sharing Mode

dedicatedSharingMode

- **type:** string
- **example:** share_idle_pros
- **description:** Dedicated Sharing Mode

nonDisruptiveMaxPU

- **type:** number
- **format:** double
- **description:** Non-disruptive Processing Units max resize constraint

nonDisruptiveMinPU

- **type:** number
- **format:** double
- **description:** Non-disruptive Processing Units min resize constraint

ebsOptimized

- **type:** boolean
- **description:** EBS Optimized

sessions

- **type:** array
- **description:** Information about sessions established by a BusinessUser to this VirtualMachine.
- – **reference:** BusinessUserSessionApiDTO *(on page 1517)*

vendorToolsVersion

- **type:** string
- **description:** Vendor Tools Version

resourceId

- **type:** string
- **description:** Resource identifier of the virtual machine. For example, the resource id in Azure, the ARN in AWS, or the full resource name in GCP.

creationTimeStamp

- **type:** integer
- **format:** int64
- **description:** Timestamp in milliseconds for the creation of the virtual machine.

vgpuProfile

- **type:** string
- **description:** vGPU profile used by virtual machine.

hardwareVersion

- **type:** string
- **example:** 19.0
- **description:** Hardware Compatibility Version

passthroughGPUIds

- **type:** array
- **example:** 0000:81:00.0
- **description:** The PCI address ID for passthrough GPUs
- – **type:** string

passthroughGPUModels

- **type:** array
- **example:** NVIDIA A100-PCIE-40GB
- **description:** The Model for passthrough GPUs
- – **type:** string

minimumHosts

- **type:** integer
- **format:** int32
- **description:** Minimum hosts required based on the HCI storage policy.

vendorLink

- **type:** string
- **description:** Link to navigate to the entity in vendor UI.

ahublicense

- **type:** boolean

VirtualDiskApiDTO

Description

Model to describe a Virtual Disk

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

tier

- **type:** string
- **description:** The tier of the Virtual Disk

actions

- **type:** array
- **description:** The related action: Move or Delete
- – **reference:** ActionApiDTO *(on page 1485)*

stats

- **type:** array
- **description:** The related statistics for this virtual disk
- – **reference:** StatApiDTO *(on page 1679)*

attachedVirtualMachine

- **description:** The Virtual Machine if the disk is attached
- **reference:** BaseApiDTO *(on page 1509)*

provider

- **description:** The Storage provider
- **reference:** BaseApiDTO *(on page 1509)*

dataCenter

- **description:** The DataCenter where the VirtualDisk resides
- **reference:** BaseApiDTO *(on page 1509)*

environmentType

- **type:** string
- **description:** Environment type
- **enum:** ['CLOUD', 'ONPREM', 'HYBRID']

lastModified

- **type:** integer
- **format:** int64
- **description:** Last Modified

businessAccount

- **description:** Business Account
- **reference:** BaseApiDTO *(on page 1509)*

resourceGroup

- **description:** Resource Group
- **reference:** BaseApiDTO *(on page 1509)*

creationTime

- **type:** string
- **description:** Creation time

skuName

- **type:** string
- **description:** Sku name

snapshotId

- **type:** string
- **description:** Snapshot id

encryption

- **type:** string
- **description:** Encryption

attachmentState

- **type:** string
- **description:** Attachment state
- **enum:** ['AVAILABLE', 'IN_USE', 'UNKNOWN']

diskRole

- **type:** string
- **description:** Disk Role
- **enum:** ['OS Disk', 'Data Disk']

name

- **type:** string
- **description:** Name

ephemeral

- **type:** string
- **description:** Ephemeral

lastAttachedVm

- **type:** string
- **description:** Name of the VM this volume was last attached to.

numDaysUnattached

- **type:** string
- **description:** Number of days ago this volume was last attached.

hourlyBilledOps

- **type:** number
- **format:** double
- **description:** Average number of billed operations per hour.

creationTimeStamp

- **type:** integer
- **format:** int64
- **description:** Timestamp in milliseconds for the creation of the virtual disk.

attachedVMStorageCompatibility

- **type:** string
- **description:** Storage compatibility for virtualDisks's attached VM(s). The compatibility will affect scale analysis for the volume. If an attached VM is not PREMIUM compatible, the volume will never scale to a premium StorageTier
- **enum:** ['PREMIUM', 'STANDARD']

storageUsageType

- **type:** string
- **description:** Usage type for the volume, i.e., what the volume is used for.
- **enum:** ['SITE_RECOVERY', 'BACKUP', 'STANDARD']

resourceId

- **type:** string
- **description:** Resource identifier of the volume. For example, the resource id in Azure, the ARN in AWS, or the full resource name in GCP.

redundancyType

- **type:** string
- **description:** Specifies the redundancy type for the volume

VirtualDisksAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

virtualDisks

- **type:** array
- **description:** List of Virtual Disks
- – **reference:** VirtualDiskApiDTO *(on page 1713)*

VirtualMachineSpecAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

cloudTierIds

- **type:** string
- **example:** [t3.micro]
- **description:** Type of the instances

minSize

- **type:** integer
- **format:** int32
- **description:** Minimum group size

maxSize

- **type:** integer
- **format:** int32
- **description:** Maximum group size

desiredSize

- **type:** integer
- **format:** int32
- **description:** Default size of group when no provider-initiated scaling actions are applied

instanceTemplateName

- **type:** string
- **description:** Name of template or configuration used to create the scale group. Ex: Name of Launch Template or Launch Configuration in AWS or the Instance Template name in Google

instanceTemplateVersion

- **type:** string
- **description:** Version of template or configuration used to create the scale group. Ex: Version of Launch Template in AWS

availabilityZones

- **type:** string
- **example:** [us-east-1a]
- **description:** Available Zones

terminationPolicies

- **type:** string
- **example:** [Default]
- **description:** Termination policies

resourceId

- **type:** string
- **description:** Resource ID containing region, account ID, AWS Auto Scaling Group ID etc

accountId

- **type:** string
- **description:** Account ID

currentInstanceCount

- **type:** integer
- **format:** int32
- **description:** Current number of instances in this Scale Group

suspendedProcesses

- **type:** string
- **description:** List of suspended processes

creationTimeStamp

- **type:** integer
- **format:** int64
- **description:** Date/time first discovered

azureVirtualMachineScaleSet

- **description:** Azure Virtual Machine Scale Set related properties
- **reference:** AzureVirtualMachineScaleSetDTO *(on page 1507)*

VirtualVolumeEntityAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

files

- **type:** array
- **description:** Files comprised in the volume
- – **type:** string

WebhookApiDTO

Description

Inherits From

This DTO inherits from WorkflowAspect *(on page 1724)*

Required Parameters

method

- **type:** string
- **description:** The http method used to make the request.
- **enum:** ['GET', 'POST', 'PUT', 'DELETE', 'PATCH']

url

- **type:** string
- **description:** The URL that HTTP request is made to.

Optional Parameters

template

- **type:** string
- **description:** The template for the body of request.

authenticationMethod

- **type:** string
- **description:** The authentication method to use for the request.
- **enum:** ['NONE', 'BASIC', 'OAUTH']

username

- **type:** string
- **description:** The username for the authenticated request.

password

- **type:** string
- **description:** The password for the authenticated request.

trustSelfSignedCertificates

- **type:** boolean
- **description:** If true, self-signed certificates will be trusted when using HTTPS connections. Defaults to 'false'.

headers

- **type:** array
- **description:** The request headers.
- – **reference:** RequestHeader *(on page 1645)*

templateValidationStatus

- **description:** The status of Webhook Template Validation.
- **reference:** TemplateValidationStatusApiDTO *(on page 1702)*

policyValidationStatus

- **type:** object
- **description:** The status of Webhook Template Validation by action types in related policy(s). The key is the policyId.
- **additionalProperties:** {'$ref': '#/definitions/TemplateValidationStatusApiDTO'}

mediaType

- **type:** string
- **description:** The media type of the workflow template
- **enum:** ['JSON', 'XML', 'TEXT']

proxy

- **description:** The proxy settings.
- **reference:** HttpProxyDTO *(on page 1569)*

oauthData

- **reference:** OAuthDataApiDTO *(on page 1596)*

WeekDayTimeSpansApiDTO

Description

Inherits From

This DTO inherits from TimeSpanGroupApiDTO *(on page 1704)*

Optional Parameters

sunday

- **type:** array
- **description:** Schedule time spans information for sunday
- – **reference:** TimeSpanApiDTO *(on page 1704)*

monday

- **type:** array
- **description:** Schedule time spans information for monday
- – **reference:** TimeSpanApiDTO *(on page 1704)*

tuesday

- **type:** array
- **description:** Schedule time spans information for tuesday
- – **reference:** TimeSpanApiDTO *(on page 1704)*

wednesday

- **type:** array
- **description:** Schedule time spans information for wednesday
- – **reference:** TimeSpanApiDTO *(on page 1704)*

thursday

- **type:** array
- **description:** Schedule time spans information for thursday
- – **reference:** TimeSpanApiDTO *(on page 1704)*

friday

- **type:** array
- **description:** Schedule time spans information for friday
- – **reference:** TimeSpanApiDTO *(on page 1704)*

saturday

- **type:** array
- **description:** Schedule time spans information for saturday
- – **reference:** TimeSpanApiDTO *(on page 1704)*

WhereApiDTO

Description

Filter for entity or group queries.

Optional Parameters

relatedToScopes

- **type:** array
- **description:** Restrict to entities related to these scopes.
- **readOnly:** True
- – **type:** integer
  – **format:** int64

conditions

- **type:** array
- **description:** Restrict to entities satisfying these conditions.
- **readOnly:** True
- – **reference:** ConditionApiDTO *(on page 1535)*

WidgetApiDTO

Description

Basic information of an entity

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

type

- **type:** string

scope

- **reference:** BaseApiDTO *(on page 1509)*

startPeriod

- **type:** string

endPeriod

- **type:** string

row

- **type:** integer
- **format:** int32

column

- **type:** integer
- **format:** int32

sizeRows

- **type:** integer
- **format:** int32

sizeColumns

- **type:** integer
- **format:** int32

widgetsetId

- **type:** integer
- **format:** int64

widgetElements

- **type:** array
  - **reference:** WidgetElementApiDTO *(on page 1721)*

widgetSource

- **type:** string
- **enum:** ['UI_DEFAULT', 'USER']

WidgetElementApiDTO

Description

Basic information of an entity

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

row

- **type:** integer
- **format:** int32

column

- **type:** integer
- **format:** int32

type

- **type:** string

widgetId

- **type:** integer
- **format:** int64

properties

- **type:** object
- **additionalProperties:** {'type': 'string'}

stats

- **type:** array
- – **reference:** StatApiInputDTO *(on page 1681)*

datasets

- **type:** array
- – **type:** string

WidgetsetApiDTO

Description

Basic information of an entity

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

scope

- **type:** string

startPeriod

- **type:** string

endPeriod

- **type:** string

widgets

- **type:** array
- – **reference:** WidgetApiDTO *(on page 1720)*

category

- **type:** string

scopeType

- **type:** string

username

- **type:** string

isSharedWithAllUsers

- **type:** boolean

WorkflowApiDTO

Description

Model to describe a Workflow

Required Parameters

discoveredBy

- **description:** Target that discovered this workflow
- **reference:** TargetApiDTO *(on page 1690)*

type

- **type:** string
- **description:** The type of this workflow
- **enum:** ['UCSD', 'ACTION_SCRIPT', 'SERVICENOW', 'ACTIONSTREAM_KAFKA', 'WEBHOOK']

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

description

- **type:** string
- **description:** Description of the workflow

entityType

- **type:** string
- **description:** Type of the entity associated

actionType

- **type:** string
- **description:** Type of the action associated

parameters

- **type:** array
- **description:** Input parameters for the workflow
- – **reference:** InputFieldApiDTO *(on page 1571)*

scriptPath

- **type:** string
- **description:** Full path to workflow script (e.g. path to an ActionScript executable

actionPhase

- **type:** string
- **description:** Action Phase to which workflow applies

timeLimitSeconds

- **type:** integer
- **format:** int64
- **description:** Time limit for workflow execution, in seconds

typeSpecificDetails

- **description:** The information about this workflow that is specific to its type.
- **reference:** WorkflowAspect *(on page 1724)*

creationTime

- **type:** string
- **description:** The creationTime is the time in ISO-8601 format when the workflow was originally created.

lastModifiedTime

- **type:** string
- **description:** The lastModifiedTime is the time in ISO-8601 format when the workflow was last modified.

WorkflowAspect

Description

Parent model to describe a type-specific workflow details

Required Parameters

type

- **type:** string

WorkflowOperationRequestApiDTO

Description

The details of the requested operation on the workflow.

Required Parameters

actionId

- **type:** integer
- **format:** int64
- **description:** The ID of the action that we perform the workflow operation on

operation

- **type:** string
- **description:** The operation that is being performed on the workflow
- **enum:** ['TEST']

WorkflowOperationResponseApiDTO

Description

Result of a workflow operation.

Required Parameters

succeeded

- **type:** boolean
- **description:** True if the workflow operation has succeeded.

Optional Parameters

uuid

- **type:** string

displayName

- **type:** string
- **description:** A user readable name of the api object. Cannot be blank. The UTF-8 encoding must be at most 255 bytes.

className

- **type:** string

details

- **type:** string
- **description:** Details of workflow operation result.

WorkloadControllerAspectApiDTO

Description

Inherits From

This DTO inherits from EntityAspect *(on page 1555)*

Optional Parameters

controllerType

- **type:** string
- **description:** Information about workload controller type.
- **enum:** ['CRON_JOB_INFO', 'CUSTOM_CONTROLLER_INFO', 'DAEMON_SET_INFO', 'DEPLOYMENT_INFO', 'JOB_INFO', 'REPLICA_SET_INFO', 'REPLICATION_CONTROLLER_INFO', 'STATEFUL_SET_INFO']

customControllerType

- **type:** string
- **description:** Information about custom controller type. Only set when the controllerType is CUSTOM_CONTROLLER_INFO.

controllerReplicaCount

- **type:** integer
- **format:** int32
- **description:** The number of active replicas controlled by the workload controller.

Reference: Entity statistics

Workload Optimization Manager monitors the following statistics for entities in your environment:

ApplicationComponent statistics

Heap

Heap is the portion of a VM or container's memory allocated to individual applications.

Produces

Not implemented

RemainingGcCapacity

Remaining GC capacity is the measurement of Application Component uptime that is *not* spent on garbage collection (GC).

ResponseTime

Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

Threads

Threads is the measurement of thread capacity utilized by applications.

Transaction

Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

VCPU

Virtual CPU is the measurement of CPU that is in use.

VMem

Virtual memory (vMem) is the measurement of memory that is in use.

BusinessApplication statistics

Produces

Not implemented

ResponseTime

Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

Transaction

Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

BusinessTransaction statistics

Produces

Not implemented

ResponseTime

Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

Transaction

Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

BusinessUser statistics

ActiveSessions

Active sessions is the number of active sessions in a desktop pool, expressed as a percentage of the pool's capacity.

ImageCPU

ImageCPU is the measurement of CPU that is utilized by a user's desktop images, expressed as a percentage of CPU capacity.

ImageMem

ImageMem is the measurement of memory that is utilized by a user's desktop images, expressed as a percentage of memory capacity.

ImageStorage

ImageStorage is the measurement of storage that is utilized by a user's desktop images, expressed as a percentage of storage capacity.

PoolCPU

Pool CPU is the measurement of CPU that is utilized by active sessions in a desktop pool.

PoolMem

Pool memory is the measurement of memory that is utilized by active sessions in a desktop pool.

PoolStorage

Pool storage is the measurement of storage that is utilized by active sessions in a desktop pool.

Produces

Not implemented

TotalSessions

Total sessions is the number of active and disconnected (non-terminated) sessions in a desktop pool, expressed as a percentage of the pool's capacity.

Chassis statistics

Cooling

Cooling is the percentage of the acceptable temperature range that is utilized by the entity. As the temperature nears the high or low running temperature limits, this percentage increases.

Power

Power is the measurement of electricity consumed by a given entity, expressed in watts.

Produces

Not implemented

Space

Not implemented

ContainerPlatformCluster statistics

NumberConsumers

Number of consumers is the total number of pods against the VM (Kubernetes node) capacity.

Produces

Not implemented

VCPU

Virtual CPU is the measurement of CPU that is in use.

VCPURequest

If applicable, vCPU request is the virtual CPU (in mCores) utilized by a container against the CPU request.

VMem

Virtual memory (vMem) is the measurement of memory that is in use.

VMemRequest

If applicable, vMem request is the virtual memory utilized by a container against the memory request.

ContainerPod statistics

NumberConsumers

Number of consumers is the total number of pods against the VM (Kubernetes node) capacity.

Produces

Not implemented

StorageAmount

Storage amount is the measurement of storage capacity that is in use.

VCPU

Virtual CPU is the measurement of CPU that is in use.

VCPULimitQuota

vCPU limit quota is the total amount of virtual CPU limits (in mCores) for all pods allocated to the namespace against the namespace quota.

VCPURequest

If applicable, vCPU request is the virtual CPU (in mCores) utilized by a container against the CPU request.

VCPURequestQuota

If applicable, vCPU request quota is the amount of virtual CPU request (in mCores) a pod has allocated against the namespace quota.

VMem

Virtual memory (vMem) is the measurement of memory that is in use.

VMemLimitQuota

If applicable, vMem limit quota is the amount of virtual memory limit a pod has allocated against the namespace quota.

VMemRequest

If applicable, vMem request is the virtual memory utilized by a container against the memory request.

VMemRequestQuota

If applicable, vMem request quota is the amount of virtual memory request a pod has allocated against the namespace quota.

VStorage

Virtual ssorage is the measurement of virtual storage capacity that is in use.

ContainerSpec statistics

Produces

Not implemented

VCPU

Virtual CPU is the measurement of CPU that is in use.

VCPURequest

If applicable, vCPU request is the virtual CPU (in mCores) utilized by a container against the CPU request.

VCPUThrottling

vCPU throttling is the throttling of container virtual CPU that could impact response time, expressed as the percentage of throttling for all containers associated with a Container Spec. In the Capacity and Usage chart for containers, *used* and *utilization* values reflect the actual throttling percentage, while *capacity* value is always 100%.

VMem

Virtual memory (vMem) is the measurement of memory that is in use.

VMemRequest

If applicable, vMem request is the virtual memory utilized by a container against the memory request.

Container statistics

Produces

Not implemented

VCPU

Virtual CPU is the measurement of CPU that is in use.

VCPULimitQuota

vCPU limit quota is the total amount of virtual CPU limits (in mCores) for all pods allocated to the namespace against the namespace quota.

VCPURequest

If applicable, vCPU request is the virtual CPU (in mCores) utilized by a container against the CPU request.

VCPURequestQuota

If applicable, vCPU request quota is the amount of virtual CPU request (in mCores) a pod has allocated against the namespace quota.

VCPUThrottling

vCPU throttling is the throttling of container virtual CPU that could impact response time, expressed as the percentage of throttling for all containers associated with a Container Spec. In the Capacity and Usage chart for containers, *used* and *utilization* values reflect the actual throttling percentage, while *capacity* value is always 100%.

VMem

Virtual memory (vMem) is the measurement of memory that is in use.

VMemLimitQuota

If applicable, vMem limit quota is the amount of virtual memory limit a pod has allocated against the namespace quota.

VMemRequest

If applicable, vMem request is the virtual memory utilized by a container against the memory request.

VMemRequestQuota

If applicable, vMem request quota is the amount of virtual memory request a pod has allocated against the namespace quota.

DataCenter statistics

Ballooning

Balloon is the measurement of memory that is shared by VMs running on a host.

CarbonFootprint

Carbon footprint is the measurement of carbon dioxide equivalent (CO2e) emissions for a given entity. Workload Optimization Manager measures carbon footprint in grams.

Cooling

Cooling is the percentage of the acceptable temperature range that is utilized by the entity. As the temperature nears the high or low running temperature limits, this percentage increases.

CoreSpeed

CoreSpeed is CPU Clock Speed of the entity.

CPU

CPU is the measurement of CPU that is reserved or in use.

Measurement of ready queue capacity that is in use on an entity with 1 CPU

Measurement of ready queue capacity that is in use on an entity with 2 CPUs

Measurement of ready queue capacity that is in use on an entity with 4 CPUs

Measurement of ready queue capacity that is in use on an entity with 8 CPUs

CPUAllocation

CPU allocation is the aggregate CPU capacity and utilization on the host

CPUProvisioned

CPU provisioned is the aggregate overprovisioned CPU capacity and utilization on the host

CPUReady

CPU ready is the measurement of a host's ready queue capacity that is in use.

Energy

Energy is the measurement of electricity used by a given entity over a period of time, expressed in watt-hours (Wh).

GpuMemAllocation

Not implemented

HardwareVersion

Hardware version specifies hardware versions that are compatible with VMs

HostLunAccess

Not implemented

HotStorage

Hot storage is the storage capacity on the server-attached flash

IOThroughput

I/O throughput is the measurement of an entity's throughput to the underlying storage.

Mem

Memory is the measurement of memory that is reserved or in use.

MemAllocation

Memory allocation is the aggregate memory capacity and utilization on the host

MemProvisioned

Memory provisioned is the aggregate overprovisioned memory capacity and utilization on the host

NetThroughput

Net throughput is the rate of message delivery over a port.

numCPUs

`numCPUs` is the number of CPUs in the datacenter.

numHosts

`NumHosts` is the number of hosts in the datacenter.

numSockets

`numSockets` is the number of sockets in the datacenter.

NumVCore

`NumVCore` is the number of CPU cores on the entity.

Power

Power is the measurement of electricity consumed by a given entity, expressed in watts.

ProcessingUnits

For PowerVM, processing unit is a unit of measurement for shared processing power across one or more virtual processors. One shared processing unit on one virtual processor accomplishes approximately the same work as one dedicated processor.

ProcessingUnitsProvisioned

`ProcessingUnitsProvisioned` is the processing units provisioned for the host.

Produces

Not implemented

RaidHosts

Not implemented

SoftwareLicenseCommodity

Not implemented

Space

Not implemented

StorageAccess

Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

StorageAmount

Storage amount is the measurement of storage capacity that is in use.

StorageLatency

Storage latency is the capacity, reserved amount, and utilized amount of latency for storage transactions, in msecs.

StorageProvisioned

Storage provisioned is the utilization of the entity's capacity, including overprovisioning.

Swapping

Swap is the measurement of a host's swap space that is in use.

VirtualProcessors

For PowerVM, virtual processor (VP) is the guest OS representation of the capacity of the VM. VPs have a one-to-one correlation with PUs. Therefore ,1 VP that is utilized in a VM equates to 1 PU being used. However, VPs do not have to be equal to PUs on a system. VPs are a virtual representation. Therefore, they can be overprovisioned.

DatabaseServer statistics

Connection

Connection is the measurement of database connections utilized by applications.

costPrice

CostPrice is the on-demand monthly cost of an entity.

DBCacheHitRate

DB cache hit rate is the measurement of Database Server accesses that result in cache hits, measured as a percentage of hits versus total attempts. A high cache hit rate indicates efficiency.

DBMem

Database memory (or DBMem) is the measurement of memory that is utilized by a Database Server.

LicenseAccess

Not implemented

Produces

Not implemented

ResponseTime

Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

RU

Request unit (RU) is a performance currency that abstracts CPU, IOPS, and memory that are required to perform the database operations supported by Azure Cosmos DB. Azure Cosmos DB normalizes the cost of all database operations using RUs.

StorageAccess

Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

StorageAmount

Storage amount is the measurement of storage capacity that is in use.

Transaction

Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

TransactionLog

Transaction log is the measurement of storage capacity utilized by a Database Server for transaction logging.

VCPU

Virtual CPU is the measurement of CPU that is in use.

VMem

Virtual memory (vMem) is the measurement of memory that is in use.

VStorage

Virtual ssorage is the measurement of virtual storage capacity that is in use.

Database statistics

Connection

Connection is the measurement of database connections utilized by applications.

costPrice

CostPrice is the on-demand monthly cost of an entity.

DTU

DTU is the measurement of compute capacity for the database. DTU represents CPU, memory, and IOPS/IO Throughput bundled as a single commodity.

DWU

DWU (Data Warehousing Unit) is the measurement of compute capacity for the dedicated SQL pool. DWU represents CPU, memory, and IO Throughput bundled as a single commodity.

IOThroughput

I/O throughput is the measurement of an entity's throughput to the underlying storage.

LicenseAccess

Not implemented

Produces

Not implemented

RU

Request unit (RU) is a performance currency that abstracts CPU, IOPS, and memory that are required to perform the database operations supported by Azure Cosmos DB. Azure Cosmos DB normalizes the cost of all database operations using RUs.

StorageAccess

Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

StorageAmount

Storage amount is the measurement of storage capacity that is in use.

VCPU

Virtual CPU is the measurement of CPU that is in use.

VMem

Virtual memory (vMem) is the measurement of memory that is in use.

DesktopPool statistics

ActiveSessions

Active sessions is the number of active sessions in a desktop pool, expressed as a percentage of the pool's capacity.

CPUAllocation

CPU allocation is the aggregate CPU capacity and utilization on the host

ImageCPU

ImageCPU is the measurement of CPU that is utilized by a user's desktop images, expressed as a percentage of CPU capacity.

ImageMem

ImageMem is the measurement of memory that is utilized by a user's desktop images, expressed as a percentage of memory capacity.

ImageStorage

ImageStorage is the measurement of storage that is utilized by a user's desktop images, expressed as a percentage of storage capacity.

MemAllocation

Memory allocation is the aggregate memory capacity and utilization on the host

PoolCPU

Pool CPU is the measurement of CPU that is utilized by active sessions in a desktop pool.

PoolMem

Pool memory is the measurement of memory that is utilized by active sessions in a desktop pool.

PoolStorage

Pool storage is the measurement of storage that is utilized by active sessions in a desktop pool.

Produces

Not implemented

TotalSessions

Total sessions is the number of active and disconnected (non-terminated) sessions in a desktop pool, expressed as a percentage of the pool's capacity.

DiskArray statistics

DiskArrayAccess

Not implemented

Extent

Not implemented

Produces

Not implemented

StorageAccess

Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

StorageAmount

Storage amount is the measurement of storage capacity that is in use.

StorageLatency

Storage latency is the capacity, reserved amount, and utilized amount of latency for storage transactions, in msecs.

StorageProvisioned

Storage provisioned is the utilization of the entity's capacity, including overprovisioning.

IOModule statistics

Cooling

Cooling is the percentage of the acceptable temperature range that is utilized by the entity. As the temperature nears the high or low running temperature limits, this percentage increases.

NetThroughput

Net throughput is the rate of message delivery over a port.

PortChannel

`PortChannel` is the amalgamation of ports with a shared net throughput and utilization.

Power

Power is the measurement of electricity consumed by a given entity, expressed in watts.

Produces

Not implemented

LogicalPool statistics

DiskArrayAccess

Not implemented

Extent

Not implemented

Produces

Not implemented

StorageAccess

Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

StorageAmount

Storage amount is the measurement of storage capacity that is in use.

StorageLatency

Storage latency is the capacity, reserved amount, and utilized amount of latency for storage transactions, in msecs.

StorageProvisioned

Storage provisioned is the utilization of the entity's capacity, including overprovisioning.

Namespace statistics

Produces

Not implemented

VCPU

Virtual CPU is the measurement of CPU that is in use.

VCPULimitQuota

vCPU limit quota is the total amount of virtual CPU limits (in mCores) for all pods allocated to the namespace against the namespace quota.

VCPURequest

Virtual CPU request is the CPU currently guaranteed by all containers on the node with a CPU request. The capacity for this resource is the node allocatable capacity, which is the amount of resources available for pods and can be less than the physical capacity.

VCPURequestQuota

If applicable, vCPU request quota is the amount of virtual CPU request (in mCores) a pod has allocated against the namespace quota.

VMem

Virtual memory (vMem) is the measurement of memory that is in use.

VMemLimitQuota

If applicable, vMem limit quota is the amount of virtual memory limit a pod has allocated against the namespace quota.

VMemRequest

Virtual memory request is the memory currently guaranteed by all containers on the node with a memory request. The capacity for this resource is the Node Allocatable capacity, which is the amount of resources available for pods and can be less than the physical capacity.

Memory currently guaranteed by all containers on the node with a memory request

VMemRequestQuota

vMem request quota is the total amount of virtual memory request for all pods allocated to the namespace against the namespace quota.

PhysicalMachine statistics

Ballooning

Balloon is the measurement of memory that is shared by VMs running on a host.

CarbonFootprint

Carbon footprint is the measurement of carbon dioxide equivalent (CO2e) emissions for a given entity. Workload Optimization Manager measures carbon footprint in grams.

Cooling

Cooling is the percentage of the acceptable temperature range that is utilized by the entity. As the temperature nears the high or low running temperature limits, this percentage increases.

CoreSpeed

CoreSpeed is CPU Clock Speed of the entity.

CPU

CPU is the measurement of CPU that is reserved or in use.

Measurement of ready queue capacity that is in use on an entity with 16 CPUs

Measurement of ready queue capacity that is in use on an entity with 1 CPU

Measurement of ready queue capacity that is in use on an entity with 2 CPUs

Measurement of ready queue capacity that is in use on an entity with 32 CPUs

Measurement of ready queue capacity that is in use on an entity with 4 CPUs

Measurement of ready queue capacity that is in use on an entity with 8 CPUs

CPUAllocation

CPU allocation is the aggregate CPU capacity and utilization on the host

CPUProvisioned

CPU provisioned is the aggregate overprovisioned CPU capacity and utilization on the host

CPUReady

CPU ready is the measurement of a host's ready queue capacity that is in use.

Energy

Energy is the measurement of electricity used by a given entity over a period of time, expressed in watt-hours (Wh).

GpuMemAllocation

Not implemented

HardwareVersion

Hardware version specifies hardware versions that are compatible with VMs

HostLunAccess

Not implemented

HotStorage

Hot storage is the storage capacity on the server-attached flash

IOThroughput

I/O throughput is the measurement of an entity's throughput to the underlying storage.

Mem

Memory is the measurement of memory that is reserved or in use.

MemAllocation

Memory allocation is the aggregate memory capacity and utilization on the host

MemProvisioned

Memory provisioned is the aggregate overprovisioned memory capacity and utilization on the host

NetThroughput

Net throughput is the rate of message delivery over a port.

numCPUs

`numCPUs` is the number of CPUs on the entity.

numSockets

`numSockets` is the number of sockets in the datacenter.

NumVCore

`NumVCore` is the number of CPU cores on the entity.

PCIPassthroughAccessCommodity

`PCIPassthroughAccessCommodity` specifies whether to prevent move actions on entities with Passthrough PCI devices.

Power

Power is the measurement of electricity consumed by a given entity, expressed in watts.

ProcessingUnits

For PowerVM, processing unit is a unit of measurement for shared processing power across one or more virtual processors. One shared processing unit on one virtual processor accomplishes approximately the same work as one dedicated processor.

ProcessingUnitsProvisioned

`ProcessingUnitsProvisioned` is the processing units provisioned for the host.

Produces

Not implemented

RaidHosts

Not implemented

SoftwareLicenseCommodity

Not implemented

Space

Not implemented

StorageAccess

Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

StorageAmount

Storage amount is the measurement of storage capacity that is in use.

StorageLatency

Storage latency is the capacity, reserved amount, and utilized amount of latency for storage transactions, in msecs.

StorageProvisioned

Storage provisioned is the utilization of the entity's capacity, including overprovisioning.

Swapping

Swap is the measurement of a host's swap space that is in use.

VirtualProcessors

For PowerVM, virtual processor (VP) is the guest OS representation of the capacity of the VM. VPs have a one-to-one correlation with PUs. Therefore ,1 VP that is utilized in a VM equates to 1 PU being used. However, VPs do not have to be equal to PUs on a system. VPs are a virtual representation. Therefore, they can be overprovisioned.

Service statistics

NumberReplicas

Number of replicas is the total number of VM instances underlying an App Service plan.

Produces

Not implemented

ResponseTime

Response time is the elapsed time between a request and the response to that request. Response time is typically measured in seconds (s) or milliseconds (ms).

Transaction

Transaction is a value that represents the per-second utilization of the transactions that are allocated to a given entity.

StorageController statistics

Produces

Not implemented

StorageAccess

Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

StorageAmount

Storage amount is the measurement of storage capacity that is in use.

StorageLatency

Storage latency is the capacity, reserved amount, and utilized amount of latency for storage transactions, in msecs.

Storage statistics

Extent

Not implemented

Produces

Not implemented

RaidHosts

Not implemented

StorageAccess

Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

StorageAmount

Storage amount is the measurement of storage capacity that is in use.

StorageLatency

Storage latency is the capacity, reserved amount, and utilized amount of latency for storage transactions, in msecs.

StorageProvisioned

Storage provisioned is the utilization of the entity's capacity, including overprovisioning.

Switch statistics

NetThroughput

Net throughput is the rate of message delivery over a port.

PortChannel

`PortChannel` is the amalgamation of ports with a shared net throughput and utilization.

Produces

Not implemented

ViewPod statistics

ActiveSessions

Active sessions is the number of active sessions in a desktop pool, expressed as a percentage of the pool's capacity.

Produces

Not implemented

TotalSessions

Total sessions is the number of active and disconnected (non-terminated) sessions in a desktop pool, expressed as a percentage of the pool's capacity.

VirtualDataCenter statistics

CPUAllocation

CPU allocation is the aggregate CPU capacity and utilization on the host

MemAllocation

Memory allocation is the aggregate memory capacity and utilization on the host

Produces

Not implemented

VirtualMachine statistics

Ballooning

Balloon is the measurement of memory that is shared by VMs running on a host.

CarbonFootprint

Carbon footprint is the measurement of carbon dioxide equivalent (CO2e) emissions for a given entity. Workload Optimization Manager measures carbon footprint in grams.

CoreSpeed

CoreSpeed is CPU Clock Speed of the entity.

costPrice

CostPrice is the on-demand monthly cost of an entity.

CPU

CPU is the measurement of CPU that is reserved or in use.

CPUAllocation

CPU allocation is the aggregate CPU capacity and utilization on the host

CPUProvisioned

CPU provisioned is the aggregate overprovisioned CPU capacity and utilization on the host

CPUReady

CPU ready is the measurement of a host's ready queue capacity that is in use.

Energy

Energy is the measurement of electricity used by a given entity over a period of time, expressed in watt-hours (Wh).

GpuMemAllocation

Not implemented

HardwareVersion

Hardware version specifies hardware versions that are compatible with VMs

HostLunAccess

Not implemented

InstanceDiskCount

Not implemented

InstanceDiskSize

Not implemented

InstanceDiskType

Not implemented

IopsSSDRead

IopsSSDRead is the per-second measurement of read operations on a Google Cloud SSD disk.

IopsSSDWrite

IopsSSDWrite is the per-second measurement of write operations on a Google Cloud SSD disk.

IopsStandardRead

IopsStandardRead is the per-second measurement of read operations on a Google Cloud standard disk.

IopsStandardWrite

IopsStandardWrite is the per-second measurement of write operations on a Google Cloud standard disk.

IOThroughput

I/O throughput is the measurement of an entity's throughput to the underlying storage.

IOThroughputRead

IOThroughputRead is the measurement of an entity's read operations from the underlying storage.

IOThroughputWrite

IOThroughputWrite is the measurement of an entity's write operations from the underlying storage.

LicenseAccess

Not implemented

Mem

Memory is the measurement of memory that is reserved or in use.

MemAllocation

Memory allocation is the aggregate memory capacity and utilization on the host

MemProvisioned

Memory provisioned is the aggregate overprovisioned memory capacity and utilization on the host

NetThroughput

Net throughput is the rate of message delivery over a port.

NetThroughputInbound

Net throughput inbound is the rate of inbound message delivery over a port.

NetThroughputOutbound

Net throughput outbound is the rate of outbound message delivery over a port.

NetworkInterfaceCount

Network interface count is the number of Network Interface Cards (NICs).

NumberConsumers

Number of consumers is the total number of pods against the VM (Kubernetes node) capacity.

NumDisk

`NumDisks` is the number of disks on the entity.

NumVCore

`NumVCore` is the number of CPU cores on the entity.

numVCPUs

`NumVCPUs` is the number of CPUs on the entity.

PCIPassthroughAccessCommodity

`PCIPassthroughAccessCommodity` specifies whether to prevent move actions on entities with Passthrough PCI devices.

ProcessingUnits

For PowerVM, processing unit is a unit of measurement for shared processing power across one or more virtual processors. One shared processing unit on one virtual processor accomplishes approximately the same work as one dedicated processor.

ProcessingUnitsProvisioned

`ProcessingUnitsProvisioned` is the processing units provisioned for the host.

Produces

Not implemented

RaidHosts

Not implemented

StorageAccess

Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

StorageAmount

Storage amount is the measurement of storage capacity that is in use.

StorageLatency

Storage latency is the capacity, reserved amount, and utilized amount of latency for storage transactions, in msecs.

StorageProvisioned

Storage provisioned is the utilization of the entity's capacity, including overprovisioning.

Swapping

Swap is the measurement of a host's swap space that is in use.

TenancyAccess

Not implemented

VCPU

Virtual CPU is the measurement of CPU that is in use.

VCPURequest

If applicable, vCPU request is the virtual CPU (in mCores) utilized by a container against the CPU request.

VirtualProcessors

For PowerVM, virtual processor (VP) is the guest OS representation of the capacity of the VM. VPs have a one-to-one correlation with PUs. Therefore ,1 VP that is utilized in a VM equates to 1 PU being used. However, VPs do not have to be equal to PUs on a system. VPs are a virtual representation. Therefore, they can be overprovisioned.

VMem

Virtual memory (vMem) is the measurement of memory that is in use.

VMemRequest

Virtual memory request is the memory currently guaranteed by all containers on the VM (Kubernetes node) with a memory request.

VStorage

Virtual ssorage is the measurement of virtual storage capacity that is in use.

Zone

Zone is the location within a given region that serves as a datacenter for the entity.

VirtualVolume statistics

costPrice

CostPrice is the on-demand monthly cost of an entity.

IOThroughput

I/O throughput is the measurement of an entity's throughput to the underlying storage.

Produces

Not implemented

StorageAccess

Storage access, also known as IOPS, is the per-second measurement of read and write access operations on a storage entity.

StorageAmount

Storage amount is the measurement of storage capacity that is in use.

StorageLatency

Storage latency is the capacity, reserved amount, and utilized amount of latency for storage transactions, in msecs.

StorageProvisioned

Storage provisioned is the utilization of the entity's capacity, including overprovisioning.

WorkloadController statistics

Produces

Not implemented

VCPULimitQuota

vCPU limit quota is the total amount of virtual CPU limits (in mCores) for all pods allocated to the namespace against the namespace quota.

VCPURequestQuota

If applicable, vCPU request quota is the amount of virtual CPU request (in mCores) a pod has allocated against the namespace quota.

VMemLimitQuota

If applicable, vMem limit quota is the amount of virtual memory limit a pod has allocated against the namespace quota.

VMemRequestQuota

If applicable, vMem request quota is the amount of virtual memory request a pod has allocated against the namespace quota.

Integration - Embedded reporting

The Workload Optimization Manager platform includes an embedded reporting component that you can choose to enable when you install the platform. Embedded reporting stores a history of your managed environment and then presents selective snapshots of this history with a set of standard dashboards and reports. You can also create your own dashboards and reports to focus on other areas of interest.

Dashboards and charts in Workload Optimization Manager embedded reporting are powered by the Grafana observability platform. With Grafana, it's easy to navigate the existing dashboards, and to make your own charts and dashboards with no coding required. You can also create custom reports via SQL queries against the embedded reporting database. For more information about creating custom reports, see Custom reports *(on page 1765)*.

The embedded reporting feature uses a PostgreSQL database with the TimescaleDB extension installed to manage the reporting data. The OVA-based installation of Workload Optimization Manager includes both a PostgreSQL database server and TimescaleDB extension. Alternatively, you can deploy a containerized or an external PostgreSQL database with the TimescaleDB extension.

For instructions to enable embedded reporting, see "Enabling embedded reporting" in the *Workload Optimization Manager Installation Guide*.

Enabling embedded reporting

Workload Optimization Manager includes an embedded reporting component that you can choose to enable when you install the platform. Use embedded reporting to understand application resource management trends and to share insights with stakeholders through reports and dashboards.

Embedded reporting runs as its own component, as part of the Workload Optimization Manager platform. This architecture enhances performance and reduces storage requirements. It stores a history of your managed environment and then presents selective snapshots of this history through a set of standard dashboards and reports. You can create your own dashboards and reports to focus on other areas of concern.

Embedded reports use TimescaleDB, which is an extension that is installed on a PostgreSQL database sever . These instructions are for the default configuration of TimescaleDB. As an alternative, you can configure an external deployment of TimescaleDB to manage your embedded reports data. For more information, see Configuring an External TimescaleDB *(on page 1800)*.

The method that you use to enable embedded reports depends on the version status of your Workload Optimization Manager instance. Two options are available:

- Run the enable_reporting.py script *(on page 1752)*

  You can use this option if you installed Workload Optimization Manager as a new VM image (OVA or VHD; see Installing on a Virtual Machine Image *(on page 50)*).

- Edit the `charts_v1alpha1_xl_cr.yaml` file.

  You can use this option if you installed Workload Optimization Manager as a Kubernetes cluster, not using the VM image for your Workload Optimization Manager installation. For more information, see Enabling embedded reporting in Kubernetes and Red Hat OpenShift Container Platform *(on page 1757)*

Using the script interface

If you installed Workload Optimization Manager as a VM image, the script to enable embedded reports is already installed on your installation at `/opt/local/bin/enable_reporting.py`.

1. Open an SSH terminal session to your Workload Optimization Manager instance.

   Log in with the System Administrator that you set up when you installed Workload Optimization Manager:

   - Username: `turbo`
   - Password: `[your_private_password]`

2. Go to the script directory.

   ```
   cd /opt/local/bin
   ```

3. Run the script.

   ```
   ./enable_reporting.py
   ```

   The script prompts you for two passwords:

   - Grafana admin password

     This password enables access to Grafana from external URLs and also from the extractor component that feeds data to Grafana.

     Do not use special characters.

     IMPORTANT:
     This step is the only time that you should change the Grafana Admin password.

   - Grafana database password

     This password enables communication between Grafana and the PostgreSQL database that stores the reporting data.

   After you supply the passwords, the script displays a confirmation message similar to the following example:

   ```
   Succesfully applied new changes
    to /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml.
   Backup written to /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml.bak
   ```

   This indicates that the script successfully updated the Workload Optimization Manager configuration. The script then applies the changed configuration to enable the embedded reports feature. It should display messages similar to the following example:

   ```
   Applying CR file /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   Warning: kubectl apply should be used on resource created by either
   kubectl create --save-config or kubectl apply
   xl.charts.helm.k8s.io/xl-release configured
   Waiting for changes to take effect...
   Restarting api pod to apply configuration changes.
   ```

```
pod "api-65cf47986f-jxszd" deleted
Changes have been successfully applied. Embedded reporting is now enabled.
```

4. Verify your installation.

   ```
   ./enable_reporting.py --validate
   ```

   If embedded reports are successfully enabled, the script output should be similar to the following example:

   ```
   No obvious embedded reporting installation errors detected.
   ```

Editing the Workload Optimization Manager YAML file

These instructions describe how to locate and edit the `charts_v1alpha1_xl_cr.yaml` for the VM image installation. If you installed on a Kubernetes node cluster, the Workload Optimization Manager configuration file is likely in a different location.

To enable embedded reports, complete the following tasks:

- Enable the components that implement the embedded reporting.
- Restart the API pod.
- Double-check the installation.
- Optionally enable email subscriptions.

Enable the Grafana Exporter, TimescaleDB, and data extraction processes by editing the `charts_v1alpha1_xl_cr.yaml` file.

1. Open an SSH terminal session to your Workload Optimization Manager instance.

   Log in with the System Administrator that you set up when you installed Workload Optimization Manager:

   - Username: `turbo`
   - Password: `[your_private_password]`

2. Open the following file in a text editor:

   ```
   /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
   ```

3. Specify the IP address of the Workload Optimization Manager instance for external access to the embedded reporting database.

   In the `global:` section of the file, add the following line, where *<Platform_IP>* is the IP address of your instance:

   ```
   global:
     externalTimescaleDBIP: <Platform_IP>
   ```

   If you are using a cloud-based TimescaleDB server, you can alternatively specify your instance by its fully qualified domain name (FQDN):

   ```
   global:
     externalTimescaleDBName: <FQDN>
   ```

   You can also customize port for the TimescaleDB instance by specifying a port value for `externalTimescaleDBPort`.

   ```
   global:
     externalTimescaleDBName: <FQDN>
     externalTimescaleDBPort: 1143
   ```

   NOTE: If you intend to use an instance with a custom DB root user, you must also specify the PostgreSQL default root user credentials, as shown in the following example:

   ```
    properties:
      global:
        dbRootUsername: admin
        dbRootPassword: MyComplexPass
        dbs.postgresDefault:
          rootUserName: postgres
          rootPassword: <The PostgreSQL user's password. The default password in OVA is
   vmturbo>
   ```

4. Enable the Grafana component.

   Find the `grafana:` section in the `crds/charts_v1alpha1_xl_cr.yaml` file, and uncomment the line, `enabled: true`.

5. Enable `postgres` as the database type.

   Enabling `postgres` sets persistent storage of historical data for embedded reports. In the `grafana:` section, find the subsection for `grafana.ini: database:` and uncomment the line, `type: postgres`.

   The changes that you have made so far should be similar to:

   ```
   global:
       externalTimescaleDBIP: <Platform_IP>
       ...
   ```

   ```
   grafana:
       enabled: true
       adminPassword: admin
       grafana.ini:
         database:
           type: postgres
           ...
   ```

6. Change the admin and database passwords.

   It is good practice to change any passwords, and not keep their default values.

IMPORTANT:
Use only alpha-numeric characters for these passwords.

These passwords enable communication between the various embedded reports components. Some of the components accept only alpha-numeric characters. If you use special characters, then the components cannot communicate. Further, the steps to correct these passwords require assistance from your Support representative.

To set the passwords:

- Set the Grafana admin password.

  This password enables access to Grafana from external URLs and also from the extractor component that feeds data to Grafana. In the `grafana:` section, change the value of `adminPassword`.

  Do not use special characters.

  Assume that your password is `MyNewGrafanaPassword`. Then, you would set `adminPassword:` **`MyNewGrafanaPassword`**

  IMPORTANT:
  This step is the only time that you should change the Grafana Admin password.

- Set the Grafana database password.

  This password enables communication between Grafana and the PostgreSQL database that stores the reporting data. In the `grafana:` section, find the subsection for `grafana.ini: database: password:` and change the password value.

```
grafana:
    enabled: true
    adminPassword: MyNewGrafanaPassword
    grafana.ini:
      database:
        type: postgres
        password: MyNewDatabasePassword
```

7. Enable the reporting processes.

   Just after the `properties:` section that you added, and at the same level to it, add the following entries to enable the reporting processes:

```
  reporting:
    enabled: true
  timescaledb:
    enabled: true
  extractor:
    enabled: true
```

It is important that you align these entries with the indentation for the `grafana:` section and the `properties:` section. The changes that you have made should now be similar to:

```
global:
    externalTimescaleDBIP: <Platform_IP>
    ...
```

```
grafana:
    enabled: true
    adminPassword: MyNewGrafanaPassword
    grafana.ini:
      database:
```

```
        type: postgres
        password: MyNewDatabasePassword
```

```
properties:
    extractor:
        grafanaAdminPassword: MyNewGrafanaPassword
```

```
reporting:
    enabled: true
timescaledb:
    enabled: true
extractor:
    enabled: true
```

8. When you are done editing the `charts_v1alpha1_xl_cr.yaml` file, save and apply your changes.
   - Save your changes and quit the text editor.
   - Run the command to apply the changes.

     ```
     kubectl apply -f \
       /opt/turbonomic/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml
     ```

   - Delete the `api` pod. If extractor was previously running, delete the `extractor` pod.

     Deleting these pods triggers them to restart, which loads the changes you made.

     Run the following command to get the full pod names, run the command,

     ```
     kubectl get pods -n turbonomic
     ```

     Find the entries for the pods that begin with `api` and `extractor`. For example, assume that the entries are:

     ```
     ...
     api-7887c66f4b-shndq                          1/1     Running   0
     ...
     extractor-5b86976bc8-vxwz4                    1/1     Running   0
     ...
     ```

     Run the following commands:

     ```
     kubectl delete pod -n turbonomic api-7887c66f4b-shndq
     ```

     ```
     kubectl delete pod -n turbonomic extractor-5b86976bc8-vxwz4
     ```

9. In an SSH terminal session for your Workload Optimization Manager instance, set the password for PostgreSQL and enable access to psql on the appliance's local host. Run the following commands:

   ```
   sudo -iu postgres psql -c "ALTER ROLE postgres WITH PASSWORD 'vmturbo'"
   ```

   ```
   sudo -iu postgres psql -c "CREATE ROLE turbo WITH SUPERUSER CREATEDB CREATEROLE LOGIN REPLICATION
    BYPASSRLS PASSWORD 'vmturbo'"
   ```

10. Verify your installation.
    - Verify that the embedded reports pods are running.

      ```
      kubectl get pods -n turbonomic
      ```

      The output is similar to the following example:

      ```
      NAME                                          READY   STATUS    RESTARTS
      extractor-7759dbcb47-vs6hr                    1/1     Running   0
      grafana-84ccb4bfb-l7sp7                       1/1     Running   0
      ```

- Verify that PostgreSQL is running.

  The PostgreSQL database should be running as a daemon on the Workload Optimization Manager server machine. To check the status, run the command:

  ```
  sudo systemctl status postgresql-15.service
  ```

  In versions prior to Workload Optimization Manager 8.14.2, the supported PostgreSQL version was 12.x. In those versions, run the `sudo systemctl status postgresql-15.service` command.

  The output is similar to the following example:

  ```
  [turbo@upg ~]$ sudo systemctl status postgresql-15.service
  ● postgresql-15.service - PostgreSQL 15 database server
      Loaded: loaded (/usr/lib/systemd/system/postgresql-15.service; enabled; preset:
  disabled)
      Active: active (running) since Tue 2024-11-05 12:12:15 EST; 6 days ago
        Docs: https://www.postgresql.org/docs/15/static/
     Process: 61045 ExecStartPre=/usr/pgsql-15/bin/postgresql-15-check-db-dir ${PGDATA}
  (code=exited, status=0/SUCCESS)
    Main PID: 61050 (postmaster)
       Tasks: 12 (limit: 209714)
      Memory: 1.3G
         CPU: 37min 29.975s
      CGroup: /system.slice/postgresql-15.service
              ├─53565 "postgres: grafana_backend grafana 10.11.12.13(52082) idle"
              ├─61050 /usr/pgsql-15/bin/postmaster -D /var/lib/pgsql/15/data/
              ├─61052 "postgres: logger "
              ├─61053 "postgres: checkpointer "
              ├─61054 "postgres: background writer "
              ├─61056 "postgres: walwriter "
              ├─61057 "postgres: autovacuum launcher "
              ├─61058 "postgres: TimescaleDB Background Worker Launcher "
              ├─61059 "postgres: logical replication launcher "
              ├─61148 "postgres: TimescaleDB Background Worker Scheduler "
              ├─61283 "postgres: TimescaleDB Background Worker Scheduler "
              └─66529 "postgres: grafana_backend grafana 10.11.12.13(50788) idle"
  ```

Enabling embedded reporting in Kubernetes and Red Hat OpenShift Container Platform

To enable embedded reporting in a native Kubernetes cluster, such as Azure Kubernetes Service, Amazon Elastic Kubernetes Service, or Red Hat OpenShift, update your Workload Optimization Manager `charts_v1alpha1_xl_cr.yaml` file to create a containerized `timescaledb` pod that runs the PostgreSQL database.

NOTE:

In Red Hat OpenShift, Grafana does not accept the global security context. Instead, set the Red Hat OpenShift security context (`scc`) at the group level. Assuming `turbonomic` is the project where Workload Optimization Manager is deployed, run the following command from the command line in the Red Hat OpenShift cluster:

```
oc adm policy add-scc-to-group anyuid system:serviceaccounts:turbonomic
```

If Workload Optimization Manager is deployed in a different project or namespace, replace `turbonomic` with your project or namespace. Although setting `scc` at the group level is not considered a best practice, this is the only option to enable Grafana in Red Hat OpenShift.

1. First, enable only the Timescaledb pod to ensure that the reporting database comes up first before you enable any of the other reporting components.

   Add the following lines to your Workload Optimization Manager `charts_v1alpha1_xl_cr.yaml` file:

```
  timescaledb:
    enabled: true
  postgresql:
    persistence:
      size: 500Gi
```

a. Optionally, to set resource requests and limits, add lines similar to the following example to the `postgresql` section:

```
postgresql:
  persistence:
    size: 500Gi
### Optional to set resource requests and limits
    resources:
      requests:
        memory: 1Gi
```

b. **Red Hat OpenShift only:** Add the following lines to enable the security context.

```
    volumePermissions:
      enabled: true
```

2. Apply the updated `charts_v1alpha1_xl_cr.yaml` file and wait for the `timescaledb` pod to be ready.
   a. Confirm that the `timescaledb` pod is ready in the user interface (UI) for your container platform or from the command line.
   b. Confirm that the PostreSQL database is ready by checking that the logs for the `timescaledb` pod include a `database system is ready to accept connections` message.
3. Next, enable the extractor, extractor properties, reporting, and Grafana reporting components.

   Add the following lines to your `charts_v1alpha1_xl_cr.yaml` file:

```
properties:
  extractor:
    grafanaAdminPassword: [your_private_password]
grafana:
  adminPassword: [your_private_password]
  enabled: true
  grafana.ini:
    database:
      password: [your_private_password]
      type: postgres
extractor:
  enabled: true
reporting:
  enabled: true
```

4. Apply the updated `charts_v1alpha1_xl_cr.yaml` file.

   The following example shows a sample configuration for the Workload Optimization Manager `charts_v1alpha1_xl_cr.yaml` with all the reporting components enabled and the Red Hat OpenShift security context requirements configured. The optional lines to set resource requests and limits are specified in the `postgresql` section:

```
#default for OCP apiVersion: charts.helm.k8s.io/v1 kind: Xl metadata:  name: xl-release spec:
#Red Hat OpenShift only
#obtain uid from project, using annotation openshift.io/sa.scc.uid-range or scc.supplemental-groups
#for embedded reporting use this uid as the value for securityContext: in the following sections
securityContext:
  fsGroup: 1000760000
### Red Hat OpenShift only - if needed
nginx:
```

```
  nginxIsPrimaryIngress: false
  httpsRedirect: false
### End of Red Hat OpenShift only section ##
properties:
  extractor:
    grafanaAdminPassword: [your_private_password]
grafana:
  adminPassword: [your_private_password]
  enabled: true
  grafana.ini:
    database:
      password: [your_private_password]
      type: postgres
extractor:
  enabled: true
reporting:
  enabled: true
timescaledb:
  enabled: true
postgresql:
  persistence:
    size: 500Gi
### Optional to set resource requests and limits
    resources:
      requests:
        memory: 1Gi
### End of optional section
### Red Hat OpenShift only
  securityContext:
    runAsUser: 1000760000
  volumePermissions:
    enabled: true
    securityContext:
      runAsUser: 1000760000
```

5. To enable the **Embedded Reporting** link in the Workload Optimization Manager UI, delete the API pod.

   From your container platform UI or the command line, find the pod that begins with `api-`, for example `api-7887c66f4b-shndq`, and delete it.

   Run the following command to get the full pod names, run the command,

   ```
   kubectl get pods -n turbonomic
   ```

   Find the pod that begins with `api`. For example, assume that the entry is:

   ```
   ...
   api-7887c66f4b-shndq                        1/1     Running   0
   ...
   ```

   Run the following command:

   ```
   kubectl delete pod -n turbonomic api-7887c66f4b-shndq
   ```

6. In the Workload Optimization Manager UI, click **Reports** in the main navigation menu.

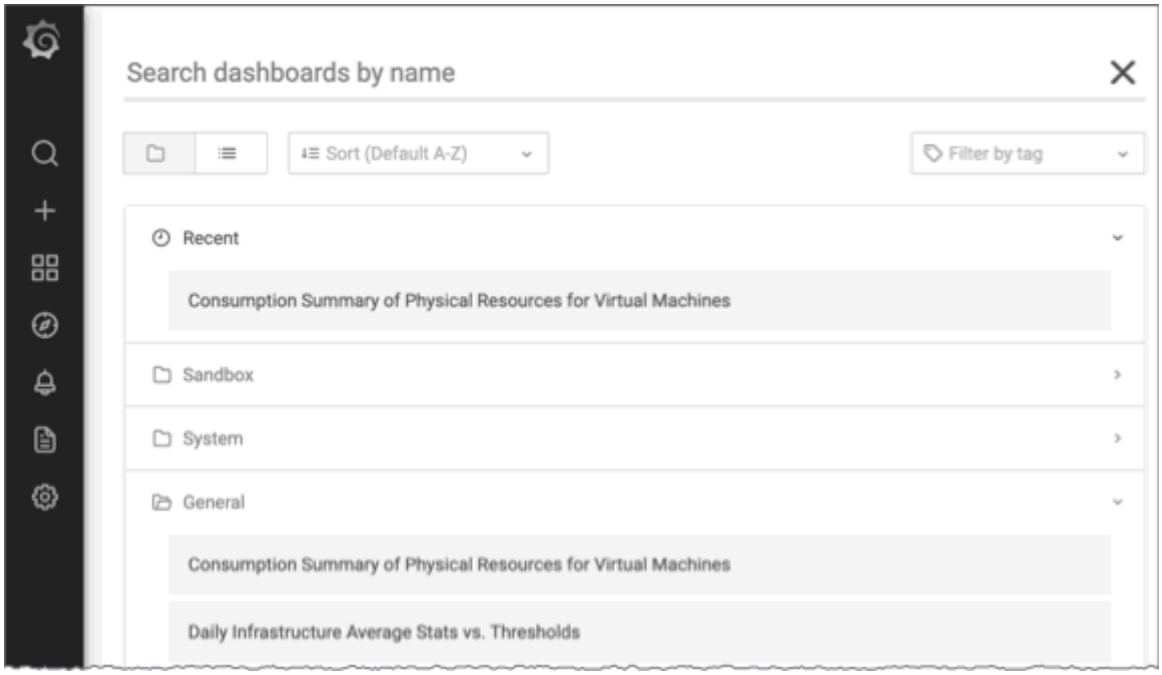
   You are redirected to the Grafana reports UI.

Navigating to the Embedded Reports page

After you enable embedded reports, open the Workload Optimization Manager user interface and click **Reports** in the navigation bar.



Dashboards and charts open in a new browser tab.



You can search for specific dashboards or browse folders to find the dashboards you want. You can also create custom reports.

Dashboards and charts are powered by the Grafana® observability platform. With Grafana, you can navigate the existing dashboards, and to make your own charts and dashboards with no coding required. If you are new to Grafana and need help with getting started, read the documentation available at:

https://grafana.com/docs/grafana/latest/

> **NOTE:**
> Starting with version 3.4.1, Workload Optimization Manager no longer has an agreement to use the Enterprise license of Grafana, and includes the open source Community license instead. After you update an earlier version of Workload Optimization Manager to version 3.4.1 or later, Grafana automatically switches to the Community license and you can use embedded reporting as usual. However, the Community license does not support PDF reports. Contact your Workload Optimization Manager representative for guidance on PDF reports.

Creating custom reports

To create custom reports, you must run SQL queries against the embedded reports database (`Turbo Timescale`).



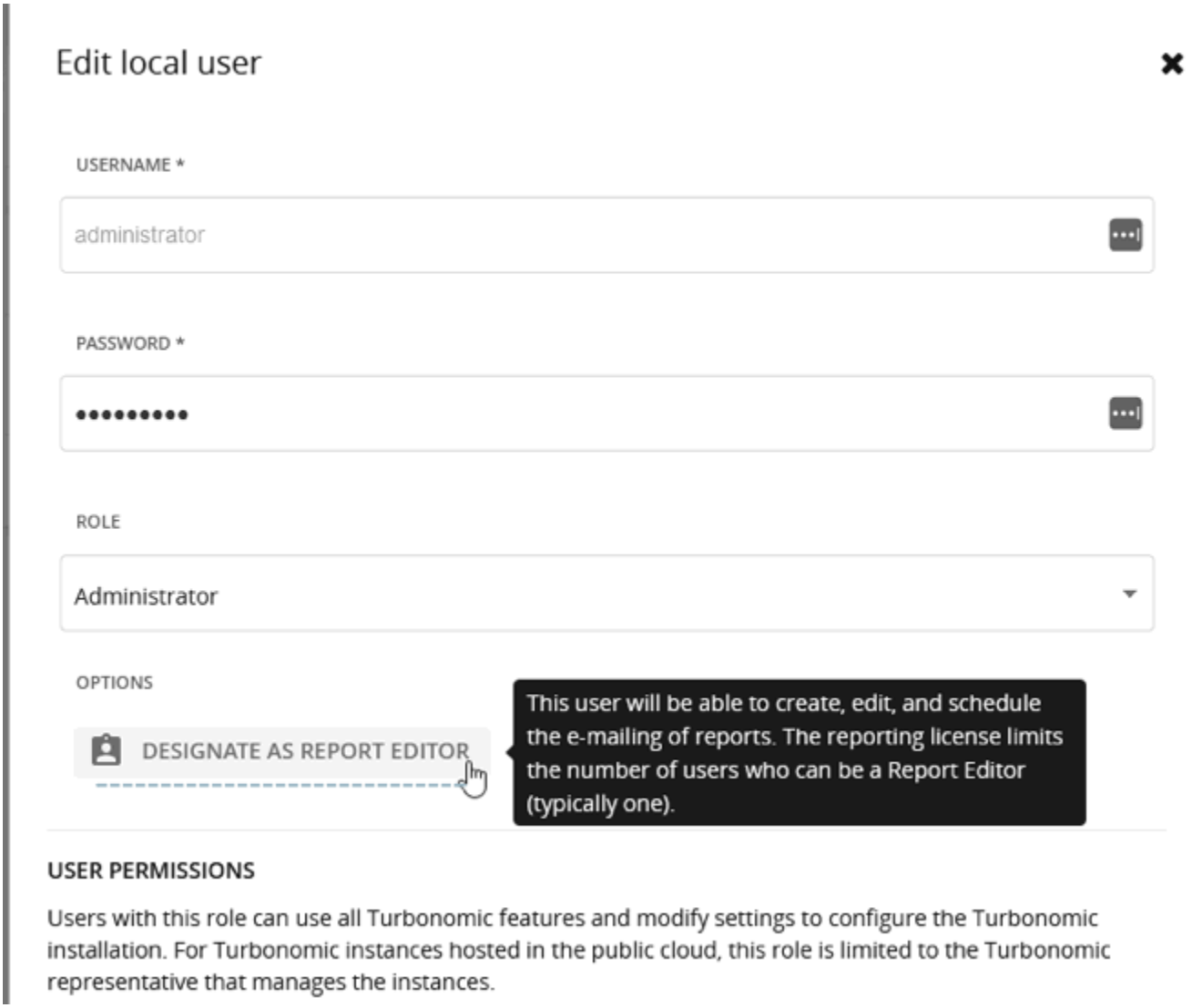
The database schema includes certain tables against which you can run queries. For more information about the schema, see Integration - Embedded reporting *(on page 1751)*.

(Optional) Report editing

To create and edit reports, a user must have Report Editor privileges. This user can have any role, as long as it is not a *shared* or *scoped* role. Only one user per Workload Optimization Manager instance is allowed to have Report Editor privileges. You can grant these privileges to another user.

To create a user account with Report Editor privileges:

1. In the Workload Optimization Manager user interface, navigate to **Settings / User Management**.
2. Choose the user account that you want to configure as a Report Editor. You can either edit an existing account or create a new one.
3. Choose a role for the user. The user can have any role, as long as it is not a *shared* or *scoped* role.
4. In the **Options** section, choose **DESIGNATE AS REPORT EDITOR**.



5. Set any other properties for the user account that you want, and save the user account.

NOTE:
It can take up to 30 minutes before the Reports page shows the Report Editor username. This usually occurs if you have changed the Report Editor multiple times.

To access reports from the user interface, a user must have the Administrator or Site Administrator role, or a non-administrator role without a defined scope. For example, a user with the Observer role but without a scope can access reports.

The default Shared Observer and Shared Adviser roles require scopes, so users with these roles cannot access reports.

Embedded reports storage requirement estimates

The embedded reports feature uses a TimescaleDB server to manage the chart data, which is a PostgreSQL server that runs with the TimescaleDB extension. Configure the datastore for your Workload Optimization Manager instance so it has enough space to support the TimescaleDB requirements.

When you initially enable embedded reports, estimate the storage that you need and configure the platform storage. If you already enabled embedded reports, check your current storage configuration and decide whether it meets your needs now and into the future.

The storage that your TimescaleDB requires depends on:

- Data retention period

  How long to store the TimescaleDB data.
- The size of your environment

  The count of entities Workload Optimization Manager manages in your environment. This count changes over time. Think of it as the average number of entities in your environment over the data retention period.

Also, increased entity count increases the data requirement, as does other activity. Storage requirements can increase over time for reasons such as the following examples:

- You add entities such as workloads, application components, storage, or hosts to your environment.
- You configure new targets.

Storage estimates lookup tables

Cisco investigates the TimescaleDB storage requirements for different topologies and retention periods. The following table lists the estimates that were calculated. Your environment might have different requirements.

	Number of entities						
Retention period	10k	25k	50k	100k	250k	500k	1000k
6 months	36 GB	91 GB	182 GB	364 GB	910 GB	1.8 TB	3.6 TB
1 year	72 GB	181 GB	361 GB	723 GB	1.8 TB	3.5 TB	7 TB
2 years	144 GB	361 GB	721 GB	1.4 TB	3.5 TB	7.2 TB	14 TB

The default installation grants a disk quota of 200 GB to the TimescaleDB. For the default installation, the database can support the following entity counts:

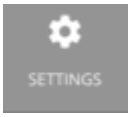
Retention period	Entity count
6 months	55k
1 year	27k
2 years	14k

Setting the data retention period

By default, Workload Optimization Manager sets the retention period for embedded reports to be 365 days. You can see the currently set retention period, and change it in the Workload Optimization Manager user interface.

To execute these actions, navigate to the Maintenance Options page:

1. Navigate to the Settings Page.

   Click to navigate to the Settings Page.
2. Choose Maintenance Options.

   Maintenance Options
3. Set the data retention period for embedded reports.

   In the **Data Retention** group of controls, find the field for **Saved Reporting Data**, which displays the current data retention period for embedded reports, in days. The default is 365 days.

   To change the retention period, enter a different number of days, and then click **Apply Settings**.

Increasing storage capacity for TimescaleDB

If you estimate the storage requirements for embedded reports after you install Workload Optimization Manager, you might learn that you need to increase the storage capacity that is available to the TimescaleDB.

The platform uses Logical Volume Management (LVM) to manage the VM disks. To increase database storage, add a disk to the VM, and then use it to extend the LVM logical volume, `/dev/turbo/var_lib_mysql`. This logical volume serves both the historical database and the embedded reports database. When you do that, you increase the quota for the TimescaleDB.

For more information, see Increasing Available Disk Space *(on page 74)*

Estimating entity count

To get a sense of entity count in your environment, search the Workload Optimization Manager log files for an INFO message that lists the number of entities for each discovery cycle. You can search for the string `INFO [Stages$BroadcastStage]`. The INFO string is similar to the following example:

```
topology-processor-6f6486df64-zf 2021-09-27 20:51:33,724 INFO [Stages$BroadcastStage] : Successfully sent
 1505 entities within topology...
```

This example shows that you have 1505 entities in the topology. Consider how your inventory changes over time. For example, you can check the entity count over time to see whether it increases regularly.

Alternatively, you can search the logs with the following command to see the counts and sizes of the entities in the full supply chain. This example assumes that Workload Optimization Manager is deployed in the `turbonomic` namespace. If your deployment is in a different namespace, edit the command to specify the correct namespace after the `-n` flag.

```
kubectl logs $(kubectl get pods -n turbonomic | grep 'rsyslog' | awk '{print $1}') | grep broadcast | grep
 -i "successfully sent"
```

In Red Hat OpenShift environments, the corresponding command is similar to the following example.

```
oc logs $(oc get pods -n turbonomic | grep 'rsyslog' | awk '{print $1}') | grep broadcast | grep -i
 "successfully sent"
```

Custom reports

Built-in and custom reports are displayed on the Grafana dashboard page. You can import reports from the Workload Optimization Manager open source site, or you can create your own using SQL Server queries into the data sets.

Importing reports

You can find a set of custom reports to import on the Workload Optimization Manager Open Source site at https://github.com/turbonomic/visualization/tree/main/embedded-visualization. This page lists different categories of reports that you can use to track your environment.

To import a custom report:

1. In the embedded reports page, create a folder to store your imported JSON files.

   In embedded reports page, navigate to **Dashboards / Browse**. on that page, click **New Folder** to create a folder for you imported JSON files.

   For example, create a folder named `{{ site.data.vars.Product_Short }}_Github_Reports`.

   After you create the folder, navigate back to the **Dashboards / Browse** page.

2. Navigate to the report you want to import.

   Go to https://github.com/turbonomic/visualization/tree/main/embedded-visualization and browse to find the report you want.

3. Copy the JSON file to the clipboard.

   In the report entry, navigate to the JSON file. In GitHub, display the file as **Raw JSON**, then select the JSON and copy it to the clipboard.

4. Import the JSON file into your Reports.

   In the embedded reports page, click **Import**, and then paste the JSON you copied into the **Import Via Panel JSON** field. Then click **Load**.

   To complete the import:

   - Give the name you want for the report.
   - Choose the folder that you just created to store the report.
   - If you want to share this report with other Grafana installations, make a note of the UID, or change it to a value you can remember.
   - Click **Import**.

The embedded reports page now displays your imported report.

Creating reports

To compose custom reports, your Workload Optimization Manager user account must have **Report Editor** permission. When you navigate from Workload Optimization Manager to the Grafana Reports page, in addition to browsing dashboards, you can create new dashboards. To create a custom report:

1. Create a dashboard.

   Navigate to the Grafana Dashboards page, then click the **+** icon in the menu bar to create a new dashboard.

2. Click to create a new panel.

   The page for the new panel includes a tabbed section with a **Query** tab.

3. In the **Query** tab, set the **Data Source** field to *Turbo Timescale*.

   This value sets up the panel to display data from the {{ site.data.vars.Product_Short }} Timescale database.

4. Edit the panel's query.

The panel displays a default query. You can edit this query to change the data that the panel displays. At the end of the default query, click **Edit SQL**. This displays a field where you can specify the SQL that you want.

For information about experimenting with different queries, see Exploring SQL Queries *(on page 1766)*

To help you build these queries, this document describes the schema tables, and provides some sample queries against them.

Exploring SQL queries

To work with SQL queries, your {{ site.data.vars.Product_Short }} user account must have **Report Editor** permission. To explore queries:

1. Navigate to the Explore page.

   The page displays with a sample query already in place

2. Choose to explore Turbo Timescale.

   Next to the page title, **Explore**, you can choose which database to work with. Choose the database `Turbo Timescale`.

   If you installed an external TimescaleDB, be sure to choose th name that you gave to that database.

3. Edit the **Explore** query.

   The page displays a default query. You can edit this query to change the data that the page displays. At the end of the default query, click **Edit SQL**. This displays a field where you can specify the SQL that you want.

   To test your query, click **Run Query**. If your query includes data points over time, you can choose **Format As: Time Series**. For tabular data, choose **Format As: Table**.

   To help you build these queries, this document describes the schema tables, and provides some sample queries against them.

Schema tables

The embedded reports data set includes the following Schema Tables to represent Workload Optimization Manager data.

action_group table

Data Object	Type	Nullable	Primary	Description	Reference
id	int	False	False		
type	enum	False	False	The specific action type - Provision, Reconfigure, Scale, etc.	action_type *(on page 1791)*
category	enum	False	False	The category of the given action.	action_category *(on page 1791)*
state	enum	False	False	That action type.	action_state *(on page 1791)*
mode	enum	False	False	The action mode - automated, manual, etc.	action_mode *(on page 1791)*
risks	text[]	True	False	An array of strings to describe the	

Data Object	Type	Nullable	Primary	Description	Reference
				risks that drive the action.	

cloud_service_cost table

For a given cloud service that is associated with a cloud target, the cost of that service.

Data Object	Type	Nullable	Primary	Description	Reference
time	timestamp with time zone	False	True		
account_oid	bigint	False	True	The ID of the cloud account or subscription this service is associated with.	
cloud_service_oid	bigint	False	True	The ID of the cloud service.	
cost	real	False	False	The cost of the service.	

completed_action Table

This table lists the current set of completed actions. A completed action is an action that completed its execution, whether the action succeeded or failed.

This table does not contain "in progress" or "queued" actions.

Data Object	Type	Nullable	Primary	Description	Reference
recommendation_time	timestamp with time zone	False	False	When analysis recommended the action.	
acceptance_time	timestamp with time zone	False	False	For manual actions, when the action was accepted for execution.	
completion_time	timestamp with time zone	False	False	When the action completed. Can be for success or failure.	
action_oid	bigint	False	False	The ID of the action. Use this to get action details.	
type	enum	False	False	The action type.	action_type *(on page 1791)*
severity	enum	False	False	The action severity.	severity *(on page 1800)*

Data Object	Type	Nullable	Primary	Description	Reference
category	enum	False	False	The action category.	action_category *(on page 1791)*
target_entity_id	bigint	False	False	The ID of the entity that the action affects. For example, for a VMem resize, the VM that is resized.	
involved_entities	bigint[]	True	False	An array of IDs for entities involved in the action. This includes the target entity.	
description	text	False	False	The text description of the action.	
savings	double precision	False	False	The cost savings this action results in. A negative savings indicates an investment.	
final_state	enum	False	False	Whether the action succeded or failed.	terminal_state *(on page 1800)*
final_message	text	False	False	A message for the final state. If the action failed, this should contain an error message.	
attrs	jsonb	True	False	Additional (often type-specific) info for the action. For example, in resizes this will contain the commodity type and from/to amounts.	action.attrs *(on page 1789)*

Sample use cases

To get the target entities for all the completed actions, join this table with the `entities` table. This example lists the completed action type and time, and the entity name type:

```
select completed_action.type AS action_type, completed_action.completion_time AS time, entity.name,
 entity.type AS ent_type
FROM completed_action
INNER JOIN entity ON completed_action.target_entity_id=entity.oid
```

Entity table

The Entity table contains information about which entities appear in the topologies and when. This table gets updated with each market cycle.

Data Object	Type	Nullable	Primary	Description	Reference
oid	bigint	False	True	The entity ID.	
type	enum	False	False	The entity type.	entity_type *(on page 1793)*
name	text	False	False	The current display name for the entity. While display names may change over time, this is always the most recent display name.	
environment	enum	True	False	Whether the entity is in an on-prem, cloud, or hybrid environment.	environment_type *(on page 1795)*
attrs	jsonb	True	False	Additional data for the entity.	entity.attrs *(on page 1786)* group.attrs *(on page 1788)*
first_seen	timestamp with time zone	False	False	When the entity was first discovered in the topology.	
last_seen	timestamp with time zone	False	False	When the entity last existed in the topology. For an entity that still exists, the value is 9999-12-31T23:59:59.	

Sample use cases

Count entities that meet specific criteria:

For a given time interval, count the number of PHYSICAL_MACHINE entities that have 4 CPUS.

```
select count(*)
from entity e
where e.type = 'PHYSICAL_MACHINE'
    and attrs->'num_cpus'= '4'
    and ('2021-02-01T05:00:00Z','2021-02-18T04:59:59Z') OVERLAPS (e.first_seen, e.last_seen)
```

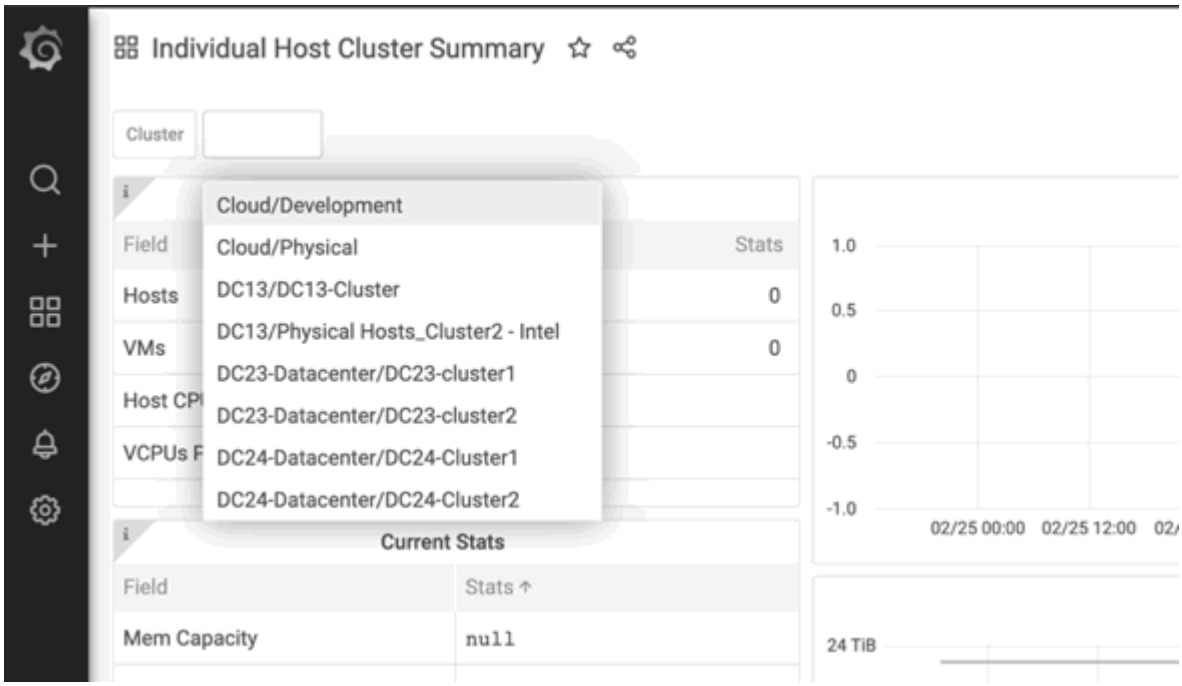
Find clusters that were configured for a given time range:

Queries like this are common to generate values for grafana variables that you can run reports against.

```
SELECT name, oid
FROM entity
WHERE
```

```
  entity.type = 'COMPUTE_CLUSTER'
  AND ('2021-02-01T05:00:00Z','2021-02-18T04:59:59Z') OVERLAPS (first_seen, last_seen)
ORDER BY 1
```

For example, here's a list of variables that a user is choosing:



entity_cost table

For a given cloud entity (a workload such as a VM or a DB service), this table represents a given cost associated with that entity.

Data Object	Type	Nullable	Primary	Description	Reference
time	timestamp with time zone	False	False		
entity_oid	bigint	False	False	The entity ID.	
category	enum	False	False	The category of service cost - IP service, storage service, on-demand or RI compute, etc.	cost_category *(on page 1792)*
source	enum	False	False	The type of billing data that provides the cost data.	cost_source *(on page 1792)*
cost	real	False	False	The calculated cost.	

entity_savings table

Workload Optimization Manager tracks actions that result in savings or investments in your environment. Actions that add resources usually require an investment, and actions that suspend or reduce resources usually give you a saving. This table describes the cost change of an action in either case.

This table can be in an action whether you have already executed the action or not. The `savings_type` field describes:

- Whether the action is a saving or investment
- Whether the action has been executed or not

Data Object	Type	Nullable	Primary	Description	Reference
time	timestamp with time zone	False	True	For a realized saving or investment, when the action was executed. For a missed saving or investment, when the action was last recommended.	
entity_oid	bigint	False	True	The object ID for the entity that has seen a saving or investment.	
savings_type	enum	False	True	Whether the action resulted in an investment or saving, and whether the action executed or not.	savings_type *(on page 1799)*
stats_value	real	False	False	The value of the investment or saving for the related action.	

entity_types_with_cost table

This table lists the types of entities that can have costs associated with them. (For costs associated with entities, see entity_cost Table *(on page 1770)*.)

Data Object	Type	Nullable	Primary	Description	Reference
type	enum	True	False	The entity type.	entity_type *(on page 1793)*

File table

This table stores the paths to files in the environment, as well as properties of the files such as last modification or file size.

Data Object	Type	Nullable	Primary	Description	Reference
volume_oid	bigint	False	True	If present, the ID of the volume that stores this file.	
path	text	False	True	The path to the file.	
type	enum	False	False	The file type - ISO, log file, snapshot, etc.	file_type *(on page 1795)*
file_size_kb	bigint	True	False	The file size in kb.	
modification_time	timestamp with time zone	True	False	When the file was last modified.	
storage_oid	bigint	False	False	If present, the ID of the disk that stores this file.	
is_attached	boolean	False	False	Whether the file is attached or unattached.	
hash	bigint	False	False		

historical_entity_attrs table

This table tracks the historical attributes that are associated with an entity. Each row represents the value of an attribute of an entity at a point in time. New rows are written when the value of a property changes between topologies, and at regular intervals (for example daily) to keep a recent value available within a reasonable time range.

Data Object	Type	Nullable	Primary	Description	Reference
time	timestamp with time zone	False	False	When this attribute was recorded.	
entity_oid	bigint	False	False	The ID of the entity this attribute pertains to.	
type	enum	False	False	The attribute type.	attr_type *(on page 1792)*
bool_value	boolean	True	False	If the attribute is boolean, the boolean value.	
int_value	int	True	False	If the attribute is an integer, the integer value. If the attribute is an enum, this is the OID of the enum in the pg_enum table (to guard against name or order changes). The in-	

Data Object	Type	Nullable	Primary	Description	Reference
				t_value should be castable to the enum via a custom cast (e.g. int_value::entity_state = 'POWERED_ON').	
long_value	bigint	True	False	If the attribute is a long, the long value.	
double_value	double precision	True	False	If the attribute is a double, the double value.	
string_value	text	True	False	If the attribute is a string, the string value.	
int_arr_value	int[]	True	False	If the attribute is a list of ints, the list values. If the attribute is a list of enums, these are the OIDs of the enums in the pg_enum table (to guard against name or order changes).	
long_arr_value	bigint[]	True	False	If the attribute is a list of longs, the long values.	
string_arr_value	text[]	True	False	If the attribute is a list of strings, the string values.	
double_arr_value	float8[]	True	False	If the attribute is a list of doubles, the double values.	
json_value	jsonb	True	False	If the attribute is a JSON object, the given object.	

Sample use cases

Get All Powered On Entities:

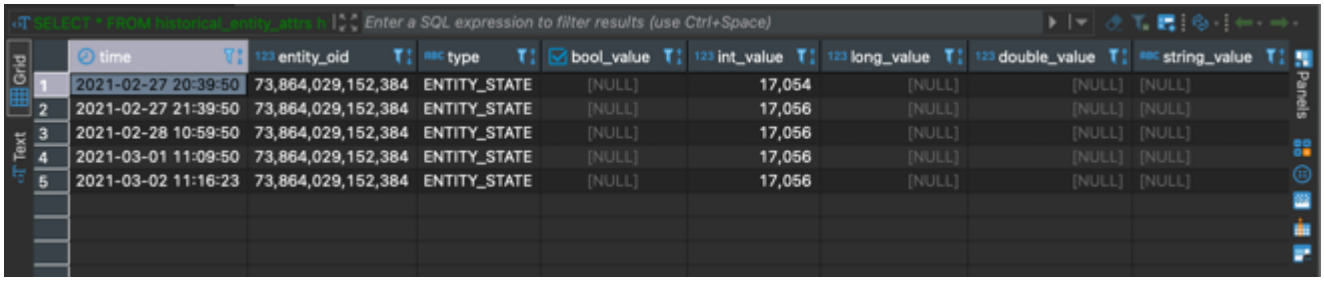
This query returns all the entities in the historical_entity_attrs that do not have a powered_off state in a certain time range.

```
select *
from historical_entity_attrs h1
where h1.type = 'ENTITY_STATE' and h1.int_value::entity_state != 'POWERED_OFF' AND h1.time between
 $__timeFrom() and $__timeTo()
```

Gap-filling data:

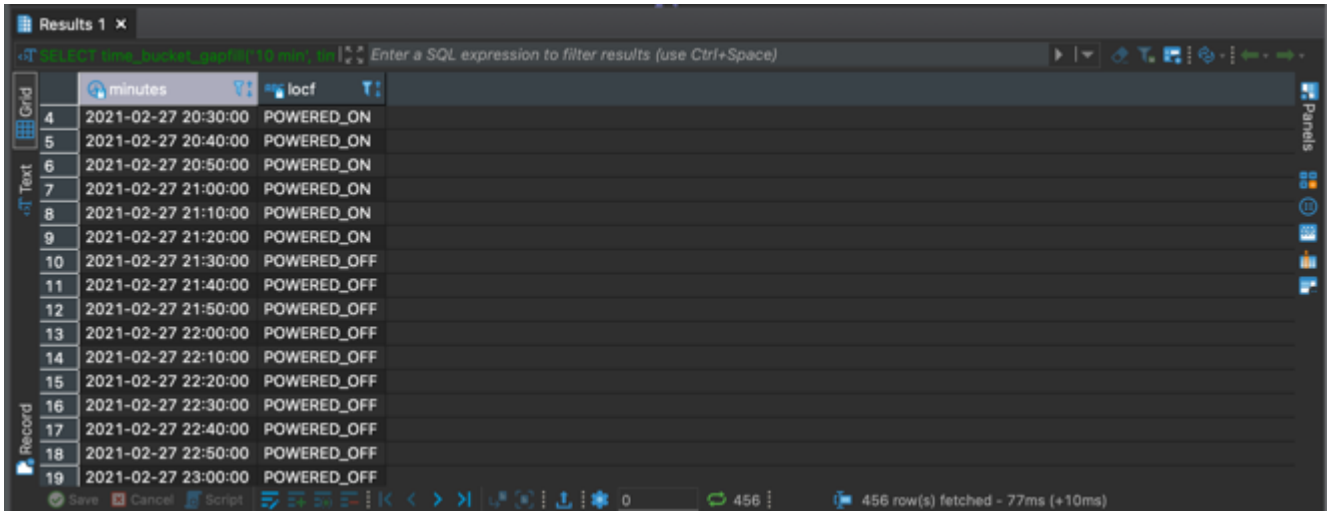
Data in this table is not written at every broadcast. Therefore, a query for those data in a time range might get only a few data points. Consider this query for the values for one entity in a time range:

```
SELECT * FROM historical_entity_attrs h1
where h1.type = 'ENTITY_STATE'
    and entity_oid ='73864029152384'
    and time between (timestamptz '2021-03-01T01:00:00Z' - interval '24 hour') and '2021-03-03T04:59:59Z'
    order by time;
```



To solve this problem, use a timescaledb function that fills the time ranges with data, as if those data were written at a constant interval. To data at every interval `t` when only data at `t1` and `t4` exists, this function can generate data at `t2` and `t3` with the same value of `t1`. The function is called `time_bucket_gapfill`, as shown in the following sample query. Notice how in the result shows data points at a constant interval of 10 minutes:

```
SELECT
  time_bucket_gapfill('10 min', time) AS minutes,
  locf(max(h1.int_value::entity_state))
FROM historical_entity_attrs h1
where h1.type = 'ENTITY_STATE'
--     AND h1.int_value::entity_state = 'POWERED_ON'
    and entity_oid ='73864029152384'
    and time between (timestamptz '2021-03-01T01:00:00Z' - interval '24 hour') and '2021-03-03T04:59:59Z'
GROUP BY minutes
```

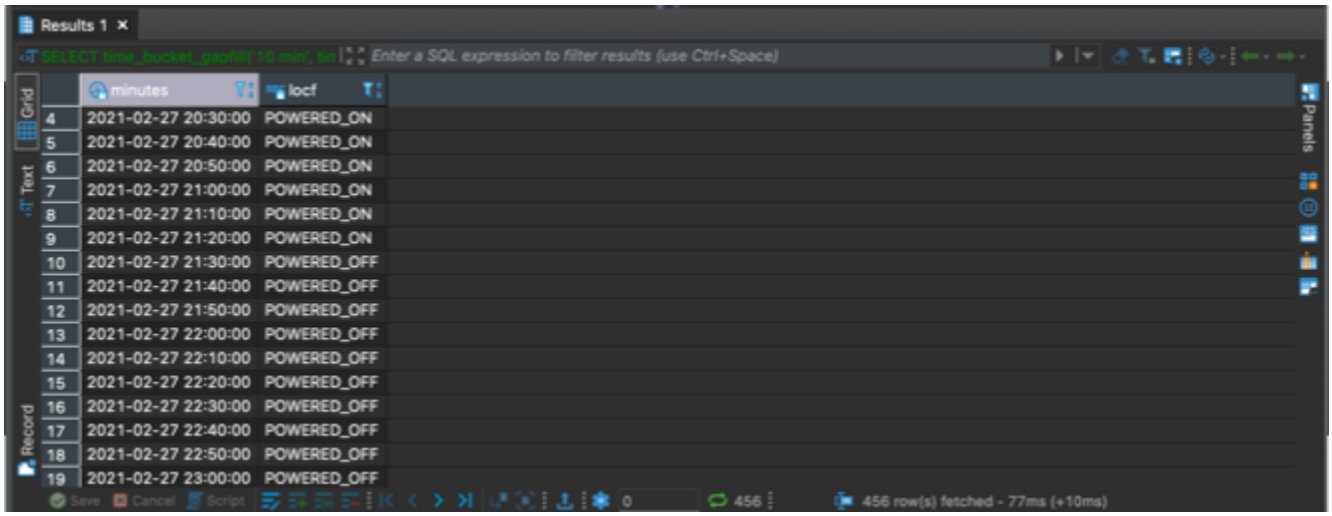


Casting Entity States to Integers:

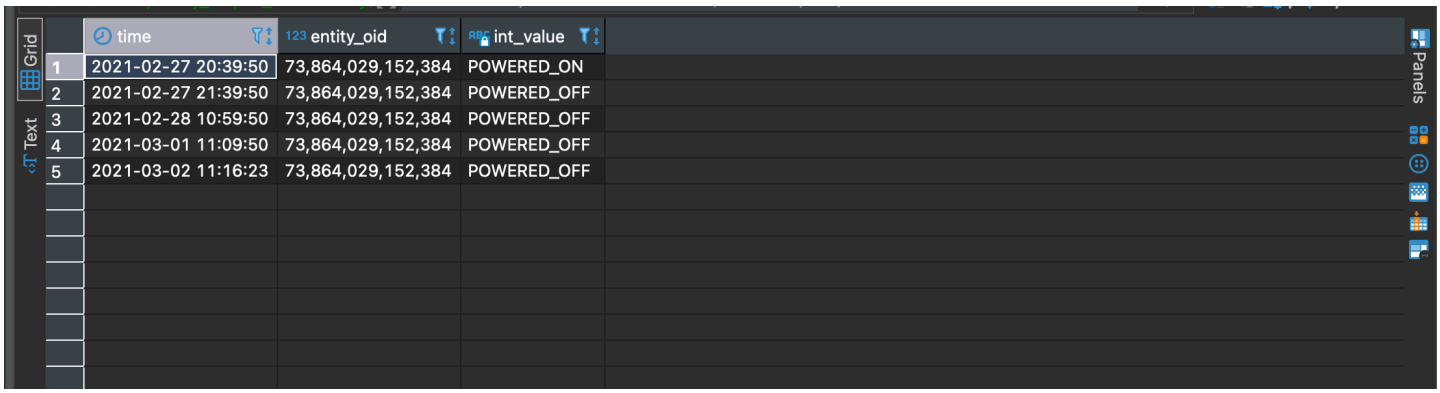
Entity States are internally represented as integers. Therefore, a property such as POWERED_OFF is stored as a number. To work with this number, the queries need to cast the integer to the corresponding entity state. Use the following cast, supported by Postgres: `int_value::entity_state` . So for example, look at the following two queries, the first one without the cast, and the second query with it:

```
SELECT time, entity_oid, entity_state FROM historical_entity_attrs h1
where h1.type = 'ENTITY_STATE'
```

```
    and entity_oid ='73864029152384'
    and time between (timestamptz '2021-03-01T01:00:00Z' - interval '24 hour') and '2021-03-03T04:59:59Z'
    order by time;
```




```
SELECT time, entity_oid, int_value::entity_state FROM historical_entity_attrs h1
where h1.type = 'ENTITY_STATE'
    and entity_oid ='73864029152384'
    and time between (timestamptz '2021-03-01T01:00:00Z' - interval '24 hour') and '2021-03-03T04:59:59Z'
    order by time;
```



Metric table

This table contains metric values for entities that appeared in the topology. Most metrics are values that are associated with commodities bought or sold by the entity. Bought and sold commodities appear in separate records – one for the buyer and one for the seller. In addition to topology metrics, this table contains headroom properties, which are generated daily by headroom plans.

Data Object	Type	Nullable	Primary	Description	Reference
time	timestamp with time zone	False	False	When analysis discovered thIS metric. For head-room metrics, when the head-room plan was created.	

Data Object	Type	Nullable	Primary	Description	Reference
entity_oid	bigint	False	False	The ID of the entity this metric pertains to.	
type	enum	False	False	The metric type.	metric_type *(on page 1795)*
provider_oid	bigint	True	False	The ID of the entity that provides this metric. For example, for VMem, the ID of the host that provides the memory.	
key	text	True	False	The commodity key. This value can be null if no key is present.	
current	double precision	True	False	Current utilization of commodity in provider entity.	
capacity	double precision	True	False	The capacity of the commodity in the provider entity.	
utilization	double precision	True	False	Overall utilization of the commodity in the provider entity.	
consumed	double precision	True	False	Amount of commodity currently used by consumer entity.	
peak_current	double precision	True	False	Peak amount of current used commodity in the provider entity.	
peak_consumed	double precision	True	False	Peak amount of commodity currently used by the consumer entity.	
entity_type	enum	False	False	The type of the consumer entity.	entity_type *(on page 1793)*

Sample use cases

VMem utilization for a Virtual Machine over time:

This query returns all the VM memory metrics that are collected for a specified virtual machine for a time interval. The time interval is truncated to the hour.

```
SELECT m.time, avg(m.utilization) AS avg_util
  FROM  metric m,
    scope_to($__timeFrom()::timestamptz - INTERVAL '24 hour',  date_trunc('hour',
 $__timeTo()::timestamptz),
      'VIRTUAL_MACHINE', ARRAY[0]::bigint[], '0' = '0') s
  WHERE m.entity_oid = s.oid
    and m.time BETWEEN s.from_time AND s.to_time
    AND m.type = 'VMEM'
    AND m.provider_oid IS NULL
    AND m.time BETWEEN $__timeFrom()::timestamptz - interval '24 hour' AND date_trunc('hour',
 $__timeTo()::timestamptz)
  GROUP BY 1
```

Storage Amount of Physical Machines Running On a Cluster:

This query returns the max values of storage amounts for all the physical machines running in a cluster. The only parameter that is given to the query other than the time range is the `cluster_id`. An internal function in Postgres can convert the `cluster_id` to all the OIDs of the physical machines.

```
SELECT
     date_trunc('day', m.time) as time,
     m.entity_oid,
     MAX(current) FILTER (WHERE m.type = 'STORAGE_AMOUNT') as stor_used,
     MAX(capacity) FILTER (WHERE m.type = 'STORAGE_AMOUNT') as stor_cap,
     MAX(current) FILTER (WHERE m.type = 'STORAGE_PROVISIONED') as stor_prov
   FROM
     metric m,
     scope_to($__timeFrom(), $__timeTo(), 'STORAGE', ARRAY[$CluserId+0]::bigint[], $CluserId = '0') s
   WHERE
     m.entity_oid = s.oid
     AND m.time BETWEEN s.from_time AND s.to_time
     AND m.type in ('STORAGE_AMOUNT', 'STORAGE_PROVISIONED')
     AND m.time between $__timeFrom() and $__timeTo()
  GROUP BY 1, 2
```

Physical Machine Average Memory Utilization – Week Over Week By Day:

Returns the average daily memory utilization for virtual machines both for the current week and the past week.

```
WITH metrics AS (
  SELECT m.time, m.utilization
    FROM metric m,
    scope_to($__timeFrom()::timestamptz - INTERVAL '7 day', $__timeTo(), 'PHYSICAL_MACHINE',
 ARRAY[0]::bigint[], '0' = '0') s
    WHERE
      m.entity_oid = s.oid
      AND m.time BETWEEN s.from_time AND s.to_time
      AND m.type = 'MEM'
      AND m.time BETWEEN ($__timeFrom()::TIMESTAMPTZ - interval '7 day') AND date_trunc('day',
 $__timeTo()::TIMESTAMPTZ)
    GROUP BY m.time, m.entity_oid, m.utilization
),
DATA AS (
  SELECT m.time, avg(m.utilization) AS avg_util
  FROM metrics m
  GROUP BY m.time
)
SELECT
  INTERVAL '1 DAY' + CASE
    WHEN time BETWEEN date_trunc('day', $__timeFrom()::TIMESTAMPTZ - interval '7 day')
      AND date_trunc('day', $__timeTo()::TIMESTAMPTZ - interval '7 day')
    THEN date_trunc('day', time + interval '7 day')
    ELSE date_trunc('day', time)
  END as "time",
```

```
  CASE
    WHEN time BETWEEN date_trunc('day', $__timeFrom()::TIMESTAMPTZ - interval '7 day')
      AND date_trunc('day', $__timeTo()::TIMESTAMPTZ - interval '7 day')
    THEN 'Previous Week'
    ELSE 'Week Ending ' || DATE($__timeTo()::TIMESTAMPTZ)::TEXT
  END as metric,
  avg(avg_util)
FROM data
GROUP BY 1, 2
ORDER BY 1, 2
```

Representation of Powered Off Virtual Machines:

Ideally, metrics of a turned off virtual machine are still written on the table, assuming that the probe sends them in the entity dto. Most used values for commodities have NULL values, instead of having 0's. This configuration is needed to not affect potential averages over time, while the capacity has the same value, independent of the state of the machine. Still, some used values for commodities aren't NULL, such as storage, since the machine consumes those commodities even in a powered off state.

pending_action table

This table lists the latest actions that have been recommended by the market. The table repopulates at a configurable interval (not necessarily every 10-minute market cycle). When it repopulates, all actions are replaced with the latest action recommendations.

This table does not include actions that are in progress.

Data Object	Type	Nullable	Primary	Description	Reference
recommendation_time	timestamp with time zone	False	False	When analysis recommended this action.	
action_oid	bigint	False	False	The action ID. Use this to get more action details.	
type	enum	False	False	the action type.	action_type *(on page 1791)*
severity	enum	False	False	The action severity	severity *(on page 1800)*
category	enum	False	False	The action category	action_category *(on page 1791)*
target_entity_id	bigint	False	False	The ID of the entity this action affects.	
involved_entities	bigint[]	True	False	An array of IDs for all the entities involved in the action. Includes the target entity.	
description	text	False	False	The text description of the action.	
savings	double precision	False	False	The cost savings this action results in. A negative savings indi-	

Data Object	Type	Nullable	Primary	Description	Reference
				cates an investment.	
attrs	jsonb	True	False	Additional (often type-specific) info for the action. For example, in resizes this will contain the commodity type and from/to amounts.	action.attrs *(on page 1789)*

Sample use cases

To get the target entities for all the pending actions, join this table with the `entities` table. This example lists the pending action type, entity name, and the entity type:

```
select pending_action.type AS action_type, entity.name, entity.type AS ent_type
FROM pending_action
INNER JOIN entity ON pending_action.target_entity_id=entity.oid
```

pending_action_stats table

Data Object	Type	Nullable	Primary	Description	Reference
time	timestamp with time zone	False	False		
scope_oid	bigint	False	False		
entity_type	enum	False	False		entity_type *(on page 1793)*
environment_type	enum	False	False		environment_type *(on page 1795)*
action_group	int	False	False		
prior_action_count	int	False	False		
cleared_action_count	int	False	False		
new_action_count	int	False	False		
involved_entity_count	int	False	False		
savings	real	False	False		
investments	real	False	False		

run_health_checks table

This table lists the health checks that are current in the {{ site.data.vars.Product_Short }} notifications.

Data Object	Type	Nullable	Primary	Description	Reference
check_name	text	True	False	The display name of the health check.	
passed	boolean	True	False	Whether the health check passed or failed.	
messages	text[]	True	False	The text description of the health check results.	

Scope table

The Scope table tracks the association of two entities as they are related in each other's scope. The `seed_oid` identifies the containing entity (the entity that defines the scope), and the `scoped_oid` identifies the entity that is within the scope. This table updates with each market cycle.

Data object	Type	Nullable	Primary	Description	Reference
seed_oid	bigint	False	True	The ID of the entity whose scope is the subject of the given record. For example, for a cluster-to-host relationship, the seed can be the cluster, and the scoped entity would be the host.	
scoped_oid	bigint	False	True	ID of the entity that is part of the seed scope.	
scoped_type	enum	False	False	Entity type for the scoped entity.	entity_type *(on page 1793)*
start	timestamp with time zone	False	True	When this relationship first appeared in the topology. This covers the initial appearance, or an appearance after a gap (the relationship was temporarily broken).	
finish	timestamp with time zone	False	False	When this relationship last appeared in the topology. A value of 9999-12-31T23:59:59 means the relationship is not yet	

Data object	Type	Nullable	Primary	Description	Reference
				known to have terminated.	

Examples

Example 1:

seed_oid	scoped_oid	scoped_type	start	finish
111	999	PHYSICAL_-MACHINE	2021-01-01T05:00:00Z	9999-12-31T23:59:59
999	111	COMPUTE_CLUSTER	2021-01-01T05:00:00Z	9999-12-31T23:59:59

The first row indicates that cluster_1 (oid: 111) contains Physical Machine_1 (oid: 999). The `start` time indicates the first time that the entity was discovered. This relationship is ongoing and is represented by a `finish` timestamp of 9999-12-31T23:59:59. Row 2 shows that Physical Machine_1 is contained by Cluster_1.

Example 2:

seed_oid	scoped_oid	scoped_type	start	finish
111	999	PHYSICAL_-MACHINE	2021-01-01T05:00:00Z	2021-01-31T05:00:00Z
999	111	COMPUTE_CLUSTER	2021-01-01T05:00:00Z	2021-01-31T05:00:00Z
222	999	PHYSICAL_-MACHINE	2021-01-31T05:00:00Z	9999-12-31T23:59:59
999	222	COMPUTE_CLUSTER	2021-01-31T05:00:00Z	9999-12-31T23:59:59

These records show that Physical_Machine_1 moved to Cluster_2 (oid: 2222) on January 31. The timestamp for the end of the relationship with Cluster_1 is the same as the timestamp for the start of the relationship with Cluster_2.

Sample use cases

Count Physical Machines in Clusters Over A Given Time Interval

```
WITH all_relationships AS (
    select
        e.name cluster_name, s.seed_oid cluster_oid, s.scoped_oid pm_oid
    FROM
        entity e
    JOIN scope s on e.oid = s.seed_oid
        and s.scoped_type ='PHYSICAL_MACHINE'
    WHERE e.type = 'COMPUTE_CLUSTER'
        AND ('2021-02-01T05:00:00Z','2021-02-18T04:59:59Z') OVERLAPS (s.start, s.finish)
), non_duplicate_relationships AS (
--    The same relationships between 2 entities can exist in given time interval if they move in and out
 of each others scopes
--   these duplicates entries at varying time intervals are accounted for
    select *
    from all_relationships
    GROUP BY cluster_name, cluster_oid, pm_oid
)
select cluster_name, count(*)
from non_duplicate_relationships
group by cluster_name
```

scope_to table

Data Object	Type	Nullable	Primary	Description	Reference
oid	bigint	True	False		
from_time	timestamp with time zone	True	False		
to_time	timestamp with time zone	True	False		
seed	bigint	True	False		

scope_to_types table

Data Object	Type	Nullable	Primary	Description	Reference
oid	bigint	True	False		
from_time	timestamp with time zone	True	False		
to_time	timestamp with time zone	True	False		
seed	bigint	True	False		
e_type	enum	True	False		entity_type *(on page 1793)*

topology_stats table

Data Object	Type	Nullable	Primary	Description	Reference
time	timestamp with time zone	False	False		
attrs	jsonb	True	False		

wasted_file table

This table stores the paths to the current wasted files in the environment, as well as properties of the files such as last modification or file size.

Data Object	Type	Nullable	Primary	Description	Reference
path	text	True	False	The file path.	
file_size_kb	bigint	True	False	The file size.	
modification_time	timestamp with time zone	True	False	When the file was last modified.	
storage_oid	bigint	True	False	The ID of the storage that contains the file.	

Data Object	Type	Nullable	Primary	Description	Reference
storage_name	text	True	False	The name of the storage (denormalized for reporting).	

Sample use cases

Get information about storage files for a data store:

The query returns the storage name, the file path, the site of the file and the last modified time for each file of a data store. The query accepts the $storage_oid as a parameter.

```
SELECT storage_name as "Datastore",
       path as "File Path",
       file_size_kb as "File Size",
       modification_time as "Last Modified Time"
  FROM wasted_file
 WHERE file_size_kb >= 0
   AND ARRAY[storage_oid]::text[] && Array['$storage_oid']::text[]
 ORDER BY 3 DESC
```

JSON data objects

The following JSON objects represent data about specific entities in your environment.

ActionCommodity JSON

The capacity and utilization of a commodity that is affected by a given action.

Data Object	Description	Type
capacity	The available capacity of the given commodity.	float
percentileUtilization	For entities and resources that support it, the calculated percentile of utilization.	Double
used	The amount of the resource currently in use.	float
utilization	The percentage of capacity in use.	float

ActionEntity JSON

For a given action, a description of the entity this action changes.

Data Object	Description	Type	Reference	Repeated
oid	The ID of the entity.	Long		
name	The entity name. This corresponds to the	String		

Data Object	Description	Type	Reference	Repeated
	name that displays in the user interface.			
type	The entity type.	/enums/entity_type	entity_type *(on page 1793)*	
attrs		JSON	entity.attrs *(on page 1786)*	
newMetric		JSON		True
newRelated		JSON		True

ActionImpactedCost JSON

For cloud entities, the on-demand costs or RI coverage that is affected by a given action. The action shows one object for the cost before the action, and another for the cost after the action.

Data Object	Description	Type
onDemandCost	For on-demand billing, the entity cost.	Float
onDemandRate	For on-demand billing, the rate of cost.	Float
riCoveragePercentage	For RIs, the percentage of RI coverage.	Float

ActionImpactedCosts JSON

For a given action, descriptions of the Before and After costs.

Data Object	Description	Type	Reference
afterActions	A description of what the cost will be after the action executes.	/shared/ActionImpactedCost	ActionImpactedCost *(on page 1784)*
beforeActions	A description of the current cost, before the action executes.	/shared/ActionImpactedCost	ActionImpactedCost *(on page 1784)*

ActionImpactedEntity JSON

For a given action, a description of an entity that is affected by that action.

Data Object	Description	Type	MapKeyType	Reference
affectedCosts	A description of the before and after costs for this entity.	/shared/ActionImpactedCosts		ActionImpactedCosts *(on page 1784)*
affectedMetrics	A description of the before and after metrics for this entity.	/shared/ImpactedMetric	/enums/metric_type	ImpactedMetric *(on page 1789)*
name	The entity name.	String		
oid	The entity ID.	Long		

Data Object	Description	Type	MapKeyType	Reference
settings	For entities that support percentile calculations, the aggressiveness and observation period for this entity.	/shared/EntitySettings		EntitySettings *(on page 1787)*
type	The entity type.	/enums/entity_type		entity_type *(on page 1793)*
attrs		JSON		entity.attrs *(on page 1786)*
newMetric		JSON		True
newRelated		JSON		True

BuyRiInfo JSON

The information that analysis can use to calculate Buy RI actions.

Data Object	Description	Type	Reference
computeTier	The compute tier, or workload family for the action. The ActionEntity object will be for an entity of type COMPUTE_TIER.	/shared/ActionEntity	ActionEntity *(on page 1783)*
count		Integer	
masterAccount	The account that manages the RI billing.	/shared/ActionEntity	ActionEntity *(on page 1783)*
region	The region that will host the RIs.	/shared/ActionEntity	ActionEntity *(on page 1783)*
target	The cloud account or subscription that will manage the RIs.	/shared/ActionEntity	ActionEntity *(on page 1783)*

CommodityChange JSON

The change (from - to) of an entity's commodity that is proposed for a given action.

Data Object	Description	Type	Reference
attribute		String	
commodityType	The type of commodity the action will change - MEM, IO_THROUGHPUT, etc.	/enums/metric_type	metric_type *(on page 1795)*
from	The commodity value before the action executes.	Float	
percentileChange	For entities and commodities that support percentile calculations, the before and af-	/shared/CommodityPercentileChange	CommodityPercentileChange *(on page 1786)*

Data Object	Description	Type	Reference
	ter values for percentile utilization and settings.		
target	A description of the entity this action will change.	/shared/ActionEntity	ActionEntity *(on page 1783)*
to	The commodity value after the action executes.	Float	
unit	The disp;layed units of measurement.	String	

CommodityPercentileChange JSON

If an action changes a commodity that participates in percentile-based analysis, the percentile settings, and the commodity values before and after the action.

Data Object	Description	Type
after	The percentile utilization after the action executes	Double
aggressiveness	The aggressiveness setting used to calculate the percentile.	Integer
before	The percentile utilization before the action executes.	Double
observationPeriodDays	The observation period used to calculate the percentile.	Integer

DeleteInfo JSON

A description of a file is unattached and should be deleted.

Data Object	Description	Type	Format
filePath	The path to the file.	String	
fileSize	The size of the file.	Double	
lastModifiedTimestamp	When the file was last modified.	String	yyyy-MM-dd'T'HH:mm:ss'Z'
unattachedDays	How many days the file has been unattached.	Integer	
unit	The displayed unit of measure for the file size.	String	

entity.attrs JSON

Additional descriptions for a given entity. The values an instance of this object can have are different for different entity types.

Data Object	Description	Type	SupportedEntity-Types	Repeated	Reference
attachment_state	Whether a given volume is attached or unattached.	/enums/AttachmentState	VIRTUAL_VOLUME	False	AttachmentState *(on page 1790)*
connected_networks	A list of network names that a given VM is connected to.	MULTI_TEXT	VIRTUAL_MACHINE	True	
cpu_model	For a physical host, the model of CPU on that host.	TEXT	PHYSICAL_MACHINE	False	
guest_os_name	The name of the OS running on a given VM.	TEXT	VIRTUAL_MACHINE	False	
guest_os_type	The specific type for the OS running on a given VM.	/enums/OSType	VIRTUAL_MACHINE	False	OSType *(on page 1790)*
is_local	Whether a storage is local to the machine, or networked.	BOOLEAN	STORAGE	False	
model	The specific model of the given physical host.	TEXT	PHYSICAL_MACHINE	False	
num_cpus	The number of CPUs on the given physical host.	NUMBER	PHYSICAL_MACHINE	False	
tags	A list of tags that are assigned to the given entity.	TEXT_MULTIMAP (Embedded Reporting) or List (Data Exporter)	ALL	False	
targets	The target that manages the given entity.	/shared/Target	ALL	True	Target *(on page 1789)*
timezone	The timezone that is set for the clock on the physical machine.	TEXT	PHYSICAL_MACHINE	False	
vendor_id		TEXT	VIRTUAL_VOLUME,BUSINESS_ACCOUNT		

EntitySettings JSON

For an entity that participates in percentile-based analysis, the aggressiveness setting and observation period.

Data Object	Description	Type
percentileAggressiveness	The current aggressiveness setting.	Integer
percentileObservationPeriodDays	The current observation period.	Integer

group.attrs JSON

Additional descriptions for a given group.

Data Object	Description	Type	Repeated	Reference	SupportedEntityTypes	SupportedGroupTypes
dynamic	Whether this group is static or dynamic. true for dynamic.	BOOLEAN	False		STORAGE_CLUSTER,COMPUTE_CLUSTER,GROUP,K8S_CLUSTER,BILLING_FAMILY,RESOURCE_GROUP	RESOURCE_GROUP,COMPUTE_CLUSTER,K8S_CLUSTER,BILLING_FAMILY,GROUP,K8S_CLUSTER,STORAGE_CLUSTER
member_types		MULTI_TEXT	True		STORAGE_CLUSTER,COMPUTE_CLUSTER,GROUP,K8S_CLUSTER,BILLING_FAMILY,RESOURCE_GROUP	RESOURCE_GROUP,COMPUTE_CLUSTER,K8S_CLUSTER,BILLING_FAMILY,GROUP,K8S_CLUSTER,STORAGE_CLUSTER
origin		TEXT	False		STORAGE_CLUSTER,COMPUTE_CLUSTER,GROUP,K8S_CLUSTER,BILLING_FAMILY,RESOURCE_GROUP	RESOURCE_GROUP,COMPUTE_CLUSTER,K8S_CLUSTER,BILLING_FAMILY,GROUP,K8S_CLUSTER,STORAGE_CLUSTER
tags		TEXT_MULTIMAP (Embedded Reporting) or List (Data Exporter)	False		GROUP,RESOURCE_GROUP,COMPUTE_CLUSTER,K8S_CLUSTER,STORAGE_CLUSTER,BILLING_FAMILY	ALL
targets		/shared/Target	False	Target *(on page 1789)*	GROUP,RESOURCE_GROUP,COMPUTE_CLUSTER,K8S_CLUSTER,STORAGE_CLUSTER,BILLING_FAMILY	ALL

ImpactedMetric JSON

For a metric on an entity that is affected by an action, a description of the impacted metric.

Data Object	Description	Type	Reference
afterActions	A description of what the metric utilization will be after the action executes.	/shared/ActionCommodity	
beforeActions	A description of the current utilization, before the action executes.	/shared/ActionCommodity	ActionCommodity *(on page 1783)*

MoveChange JSON

For a vMotion action, the changed placement of the affected entity.

Data Object	Description	Type	Repeated	Reference
from	Where the entity will move from.	/shared/ActionImpactedEntity	False	ActionImpactedEntity *(on page 1784)*
resource		/shared/ActionEntity	True	ActionEntity *(on page 1783)*
to	Where the entity will move to.	/shared/ActionImpactedEntity	False	ActionImpactedEntity *(on page 1784)*

Target JSON

Data Object	Description	Type
category		String
entityVendorId		String
name		String
oid		Long
type		String

action.attrs JSON

Extra data that pertains to a given action.

JSON Object	Description	Type	MapKeyType	Reference
buyRiInfo	For an action to buy RIs, additional information about the RIs.	/shared/BuyRiInfo		BuyRiInfo *(on page 1785)*
deleteInfo	For an action to delete a file, a description of that file.	/shared/DeleteInfo		DeleteInfo *(on page 1786)*

JSON Object	Description	Type	MapKeyType	Reference
moveInfo	For a move action, a description of from and to for the move.	/shared/MoveChange	/enums/entity_type	MoveChange *(on page 1789)*
resizeInfo	For a resize action, a description of the commodity changes.	/shared/Commodity-Change	/enums/metric_type	CommodityChange *(on page 1785)*
scaleInfo	For scaling a cloud VM, a description of the from and to for the VM move.	/shared/MoveChange	/enums/entity_type	MoveChange *(on page 1789)*
target	Extra information about the entity this action impacts.	/shared/ActionIm-pactedEntity		ActionImpactedEntity *(on page 1784)*

Enums

These enums specify values for certain states or types.

AttachmentState Enum

Whether a given is attached or unattached.

- ATTACHED:
- UNATTACHED:

OSType Enum

The OS types that Workload Optimization Manager discovers on hosts and VMs.

- LINUX:
- LINUX_WITH_SQL_ENTERPRISE:
- LINUX_WITH_SQL_STANDARD:
- LINUX_WITH_SQL_WEB:
- RHEL:
- SUSE:
- UBUNTU_PRO:
- UNKNOWN_OS:
- WINDOWS:
- WINDOWS_BYOL:
- WINDOWS_WITH_SQL_ENTERPRISE:
- WINDOWS_WITH_SQL_STANDARD:
- WINDOWS_WITH_SQL_WEB:

action_category Enum

The different categories for a given action.

- COMPLIANCE:
- EFFICIENCY_IMPROVEMENT:
- PERFORMANCE_ASSURANCE:
- PREVENTION:
- UNKNOWN:
- SAVING:

action_mode Enum

The different modes for an action, such as automated or manual.

- AUTOMATIC:
- DISABLED:
- EXTERNAL_APPROVAL:
- MANUAL:
- RECOMMEND:

action_state Enum

The different states for an action, such as completed or pending.

- ACCEPTED:
- CLEARED:
- FAILED:
- FAILING:
- IN_PROGRESS:
- POST_IN_PROGRESS:
- PRE_IN_PROGRESS:
- QUEUED:
- READY:
- REJECTED:
- SUCCEEDED:
- BEFORE_EXEC:
- IN_PROGRESS_NON_CRITICAL:

action_type Enum

The different types of action.

- ACTIVATE:
- ALLOCATE:
- BUY_RI:
- DEACTIVATE:
- DELETE:

- MOVE:
- NONE:
- PROVISION:
- RECONFIGURE:
- RESIZE:
- SCALE:
- START:
- SUSPEND:

attr_type Enum

The supported types of attributes. For each type, you can find different data stored in the attribute.

- ENTITY_STATE:
- NUM_VCPU:
- VOLUME_ATTACHED:
- IDLE_POWER:

cost_category Enum

For different cost data, the service or category that the cost pertains to.

- IP:
- ON_DEMAND_COMPUTE:
- ON_DEMAND_LICENSE:
- RESERVED_LICENSE:
- RI_COMPUTE:
- SPOT:
- STORAGE:
- TOTAL:
- CLOUD_COMMITMENT_USAGE:

cost_source Enum

The type of billing agreement that is the source of the given cost data.

- BUY_RI_DISCOUNT:
- ENTITY_UPTIME_DISCOUNT:
- ON_DEMAND_RATE:
- RI_INVENTORY_DISCOUNT:
- TOTAL:
- UNCLASSIFIED:
- CLOUD_COMMITMENT_DISCOUNT:

entity_state Enum

The running state for an entity, such as powered off, suspended, or running.

- FAILOVER:
- MAINTENANCE:
- POWERED_OFF:
- POWERED_ON:
- SUSPENDED:
- UNKNOWN:

entity_type Enum

The set of supported entities that Workload Optimization Manager manages.

- ACTION_MANAGER:
- APPLICATION:
- APPLICATION_COMPONENT:
- APPLICATION_COMPONENT_SPEC:
- APPLICATION_SERVER:
- AVAILABILITY_ZONE:
- BILLING_FAMILY:
- BUSINESS:
- BUSINESS_ACCOUNT:
- BUSINESS_APPLICATION:
- BUSINESS_ENTITY:
- BUSINESS_TRANSACTION:
- BUSINESS_USER:
- CHASSIS:
- CLOUD_COMMITMENT:
- CLOUD_SERVICE:
- COMPUTE_CLUSTER:
- COMPUTE_RESOURCE:
- COMPUTE_TIER:
- CONTAINER:
- CONTAINER_PLATFORM_CLUSTER:
- CONTAINER_POD:
- CONTAINER_SPEC:
- DATABASE:
- DATABASE_SERVER:
- DATABASE_SERVER_TIER:
- DATABASE_TIER:
- DATACENTER:
- DESIRED_RESERVED_INSTANCE:
- DESKTOP_POOL:

- DISK_ARRAY:
- DISTRIBUTED_VIRTUAL_PORTGROUP:
- DOCUMENT_COLLECTION:
- DPOD:
- GROUP:
- HCI_PHYSICAL_MACHINE:
- HYPERVISOR_SERVER:
- INFRASTRUCTURE:
- INTERNET:
- IO_MODULE:
- IP:
- K8S_CLUSTER:
- LICENSING_SERVICE:
- LOAD_BALANCER:
- LOGICAL_POOL:
- MAC:
- MOVER:
- NAMESPACE:
- NETWORK:
- NETWORKING_ENDPOINT:
- NODE_POOL:
- OPERATOR:
- PHYSICAL_MACHINE:
- PORT:
- PROCESSOR_POOL:
- REGION:
- RESERVED_INSTANCE:
- RESERVED_INSTANCE_SPECIFICATION:
- RESOURCE_GROUP:
- RIGHT_SIZER:
- SAVINGS:
- SERVICE:
- SERVICE_ENTITY_TEMPLATE:
- SERVICE_PROVIDER:
- STORAGE:
- STORAGE_CLUSTER:
- STORAGE_CONTROLLER:
- STORAGE_TIER:
- STORAGE_VOLUME:
- SWITCH:
- TARGET:
- THIS_ENTITY:
- THREE_TIER_APPLICATION:

- VIEW_POD:
- VIRTUAL_APPLICATION:
- VIRTUAL_DATACENTER:
- VIRTUAL_MACHINE:
- VIRTUAL_MACHINE_SPEC:
- VIRTUAL_VOLUME:
- VLAN:
- VM_SPEC:
- VPOD:
- WEB_SERVER:
- WORKLOAD_CONTROLLER:
- _NONE_:
- DATA_WAREHOUSE:
- DATA_WAREHOUSE_TIER:
- NODE_GROUP:
- POWERVM_PHYSICAL_MACHINE:

environment_type Enum

Whether the environment is cloud, on-prem, or hybrid.

- CLOUD:
- HYBRID:
- ON_PREM:
- UNKNOWN_ENV:

file_type Enum

Different types of file, such as configuration file or ISO image.

- CONFIGURATION:
- DISK:
- ESXCONSOLE:
- ISO:
- LOG:
- MEMORY:
- SNAPSHOT:
- SWAP:

metric_type Enum

The different types of metrics that Workload Optimization Manager discovers for entities.

- ACTIVE_SESSIONS:

  Count of the number of actively logged in sessions Count
- BALLOONING:

Number of kb the host is dedicated for ballooning kilobytes

- BUFFER_COMMODITY:

  Utilization of a port buffer (fabric, UCS specific) megabytes
- CARBON:

  Carbon dioxide equivalent (CO2e) emissions for a given entity grams
- CONCURRENT_QUERIES:

  Number of concurrent queries to a workload. When there is no request, this metric is zero.
- CONNECTION:

  Number of connections onto the application Count
- COOLING:

  acceptable temperature range that is utilized by an entity percentage
- CPU:

  CPU used by host (current column) or host CPU consumed by VMs (consumed column). MHz
- CPU_ALLOCATION:

  For VMs (consumed column) - VM buys from VDC or from desktop pool (VDI). For hosts - PMs sells to VDCs. MHz
- CPU_HEADROOM:

  Number of VMs that can fit in all of the available CPU (on a cluster) Count
- CPU_PROVISIONED:

  Capacity is based on the over-provisioned value in policy MHz
- CPU_READY:

  Ready queue ms
- DB_CACHE_HIT_RATE:

  The percentage of accesses that result in cache hits. Percentage
- DB_MEM:

  DB mem utilization or DM capacity kilobytes
- DTU:

  Capacity defined in the cloud Count
- ENERGY:

  Electricity consumed by a given entity over a period of time watt-hours
- EXTENT:

  Internal NA
- FLOW:

  Amount of network traffic per second between two nodes kilobytes/sec
- FLOW_ALLOCATION:

  bytes
- GPU:

  GPU resources
- GPU_MEM:

  Amount of GPU memory provided
- GPU_REQUEST:

  Amount of GPU memory requested
- GPU_REQUEST_QUOTA:

  Quota allocated for GPU requests
- HEAP:

  Amount of heap configured for an application kilobytes
- IMAGE_CPU:

For VDI, usage of the business user MHz

- IMAGE_MEM:

  For VDI, usage of the business user kilobytes
- IMAGE_STORAGE:

  For VDI, usage of the business user megabytes
- IO_THROUGHPUT:

  Amount of traffic through the IO adapter kilobytes/sec
- LLM_CACHE:

  Cache for an LLM service
- MEM:

  Physical machine – usage of memory. VM – consumed memory. kilobytes
- MEM_ALLOCATION:

  Physical machine – what the host provides to VDCs. VM (consumed column) – what the VM buys from VDC or desktop pool. kilobytes
- MEM_HEADROOM:

  Number of VMs that can fit in the available memory on the cluster Count
- MEM_PROVISIONED:

  Capacity, based on the over-provisioned value in policy kilobytes
- NET_THROUGHPUT:

  Amount of traffic through the Network adapter kilobytes/sec
- POOL_CPU:

  For VDI, usage of the desktop pool MHz
- POOL_MEM:

  For VDI, usage of the desktop pool kilobytes
- POOL_STORAGE:

  For VDI, usage of the desktop pool megabytes
- PORT_CHANEL:

  Traffic on ports for UCS Mgr/HPE1 kilobytes/sec
- POWER:

  Electricity consumed by a given entity watts
- PROCESSING_UNITS:

  unit of measure for shared processing power across one or more virtual processors PU
- Q16_VCPU:

  VM – Sum of the Ready Queue usage of all of the CPU on the VM. To normalize the value, divide it by the number of vCPU. Host – Average of the underlying Ready Queue values from all of the corresponding VMs with the same number of vCPUs. ms
- Q1_VCPU:

  VM – Sum of the Ready Queue usage of all of the CPU on the VM. To normalize the value, divide it by the number of vCPU. Host – Average of the underlying Ready Queue values from all of the corresponding VMs with the same number of vCPUs. ms
- Q2_VCPU:

  VM – Sum of the Ready Queue usage of all of the CPU on the VM. To normalize the value, divide it by the number of vCPU. Host – Average of the underlying Ready Queue values from all of the corresponding VMs with the same number of vCPUs. ms
- Q32_VCPU:

  VM – Sum of the Ready Queue usage of all of the CPU on the VM. To normalize the value, divide it by the number of vCPU. Host – Average of the underlying Ready Queue values from all of the corresponding VMs with the same number of vCPUs. ms

- Q3_VCPU:

  VM – Sum of the Ready Queue usage of all of the CPU on the VM. To normalize the value, divide it by the number of vCPU. Host – Average of the underlying Ready Queue values from all of the corresponding VMs with the same number of vCPUs. ms
- Q4_VCPU:

  VM – Sum of the Ready Queue usage of all of the CPU on the VM. To normalize the value, divide it by the number of vCPU. Host – Average of the underlying Ready Queue values from all of the corresponding VMs with the same number of vCPUs. ms
- Q5_VCPU:

  VM – Sum of the Ready Queue usage of all of the CPU on the VM. To normalize the value, divide it by the number of vCPU. Host – Average of the underlying Ready Queue values from all of the corresponding VMs with the same number of vCPUs. ms
- Q64_VCPU:

  VM – Sum of the Ready Queue usage of all of the CPU on the VM. To normalize the value, divide it by the number of vCPU. Host – Average of the underlying Ready Queue values from all of the corresponding VMs with the same number of vCPUs. ms
- Q6_VCPU:

  VM – Sum of the Ready Queue usage of all of the CPU on the VM. To normalize the value, divide it by the number of vCPU. Host – Average of the underlying Ready Queue values from all of the corresponding VMs with the same number of vCPUs. ms
- Q7_VCPU:

  VM – Sum of the Ready Queue usage of all of the CPU on the VM. To normalize the value, divide it by the number of vCPU. Host – Average of the underlying Ready Queue values from all of the corresponding VMs with the same number of vCPUs. ms
- Q8_VCPU:

  VM – Sum of the Ready Queue usage of all of the CPU on the VM. To normalize the value, divide it by the number of vCPU. Host – Average of the underlying Ready Queue values from all of the corresponding VMs with the same number of vCPUs. ms
- QN_VCPU:

  VM – Sum of the Ready Queue usage of all of the CPU on the VM. To normalize the value, divide it by the number of vCPU. Host – Average of the underlying Ready Queue values from all of the corresponding VMs with the same number of vCPUs. ms
- QUEUING_TIME:

  Amount of time that a request spends in a queue before it is processed. When there is no request, this metric is zero.
- REMAINING_GC_CAPACITY:

  Percentage of time not spent on garbage collection Percentage
- RESPONSE_TIME:

  Application response ms
- SERVICE_TIME:

  Amount of processing time needed to generate the next token. This metric is relatively stable for a given model and GPU resource. When there is no request, this metric is unavailable.
- SLA_COMMODITY:

  User-defined value Number
- STORAGE_ACCESS:

  Number of I/O (read + write) within two discovery cycles count
- STORAGE_ALLOCATION:

  Allocated storage megabytes
- STORAGE_AMOUNT:

  Used storage megabytes
- STORAGE_HEADROOM:

Count of VM that can fit for the storage in a cluster in terms of storage Count

- `STORAGE_LATENCY`:

  Acceptable latency. User can set capacity for most storage, is discovered for NetApp. ms
- `STORAGE_PROVISIONED`:

  Capacity, based on the over-provisioned value in policy megabytes
- `SWAPPING`:

  Memory swapping bytes/sec
- `THREADS`:

  Threads in use Count
- `TOTAL_HEADROOM`:

  Min value of CPU, Mem, IO, of the headrooms Count
- `TOTAL_SESSIONS`:

  Total number of possible sessions Count
- `TRANSACTION`:

  Transactions for the application transactions/sec
- `TRANSACTION_LOG`:

  Size of the transaction log megabytes
- `VCPU`:

  CPU in use MHz
- `VCPU_LIMIT_QUOTA`:

  Amount of virtual CPU limit a pod has allocated against the namespace quota millicores
- `VCPU_REQUEST`:

  Virtual CPU request allocated by a pod against the node allocatable capacity millicores
- `VCPU_REQUEST_QUOTA`:

  Amount of virtual CPU request a pod has allocated against the namespace quota millicores
- `VCPU_THROTTLING`:

  Throttling of container virtual CPU that could impact response time percentage
- `VIRTUAL_PROCESSORS`:

  Guest OS representation of the capacity of the VM VP
- `VMEM`:

  Memory sold by a VM. Memory sold and bought for a container kilobytes
- `VMEM_LIMIT_QUOTA`:

  Total amount of virtual memory limit for all pods allocated to the namespace against the namespace quota kilobytes
- `VMEM_REQUEST`:

  Virtual memory request allocated by a pod against the node allocatable capacity kilobytes
- `VMEM_REQUEST_QUOTA`:

  Amount of virtual memory request a pod has allocated against the namespace quota kilobytes
- `VSTORAGE`:

  Storage utilized by an entity megabytes

savings_type Enum

The different types of savings that Workload Optimization Manager actions can give.

- `MISSED_INVESTMENTS`:

  Actions you could have taken that would inscrease your investment to provide more resources to your applications.
- `MISSED_SAVINGS`:

Actions you could have taken that could have reduced your costs.

- `REALIZED_INVESTMENTS`:

  Actions you have taken that inscrease your investment to provide more resources to your applications.
- `REALIZED_SAVINGS`:

  Actions you have taken that have reduced your costs.

severity Enum

For a given action, the severity of the issue that it addresses.

- `CRITICAL`:
- `MAJOR`:
- `MINOR`:
- `NORMAL`:

terminal_state Enum

The ultimate state of an action.

- `FAILED`:
- `SUCCEEDED`:

Configuring an external TimescaleDB

You can configure External Reports to use an external deployment of TimescaleDB. You can use this as an alternative to the default deployment.

To enable an external deployment, you must complete the following tasks:

- Deploy or access an external TimescaleDB instance. This instance is a PostgreSQL server with the TimescaleDB extension installed. The DB server must be deployed in a way that is accessible to your installation of Workload Optimization Manager. It can be deployed on the cloud or in an on-prem VM. As of Workload Optimization Manager 8.14.2, the embedded reports feature currently supports PostgreSQL 15.x and TimescaleDB 2.17.1. Prior to Workload Optimization Manager 8.14.2, the embedded reports feature supported PostgreSQL 12.x and TimescaleDB 2.10.2. For more information, see the following resources:
  - Creating an External TimescaleDB *(on page 1801)*
  - Install TimescaleDB
- Provision users, databases, and schemas on the TimescaleDB instance. To provision these database objects, you can manually provision or enable Workload Optimization Manager to automatically provision. For more information, see:
  - Manual provisioning of users and databases *(on page 1801)*
  - Automatic provisioning of users and databases *(on page 1804)*
- Edit properties in the Workload Optimization Manager `charts_v1alpha1_xl_cr.yaml` file. The location of the `charts_v1alpha1_xl_cr.yaml` file depends on whether you are configuring an OVA installation or a Kubernetes Node installation of Workload Optimization Manager.
- Optionally, enable secure connections between Workload Optimization Manager and the TimescaleDB instance. For information about enabling TLS/SSL for database connections, see the PostgreSQL documentation Secure TCP/IP Connections with SSL.

  This link provides steps to create a self-signed certificate that works with embedded reporting. However, self-signed certificates are not recommended for production instances.

Creating an external TimescaleDB

You can create an external TimescaleDB service that you host in your environment. Alternatively, you can use an existing TimescaleDB service that is deployed on the cloud or in your enterprise by the DB Management team.

The TimescaleDB service that Workload Optimization Manager uses must satisfy these conditions:

- **Deployment**: The TimescaleDB is a PostgreSQL server with the TimescaleDB extension. It must be deployed in a way that is accessible to your installation of Workload Optimization Manager. It can be deployed on the cloud or in an on-prem VM.

  For information about installing Timescale, see Install TimescaleDB.
- **Supported version**: As of Workload Optimization Manager 8.14.2, the embedded reports feature currently supports PostgreSQL 15.x and TimescaleDB 2.17.1. Prior to Workload Optimization Manager 8.14.2, the embedded reports feature supported PostgreSQL 12.x and TimescaleDB 2.10.2.
- An account with global R/W privileges on the TimescaleDB instanceWorkload Optimization ManagerThe embedded reports feature requires a specific set of databases to be created on the TimescaleDB instance. It is typical practice for an administrator to manually create the necessary databases. Alternatively, you can create a global R/W user account for the Workload Optimization Manager platform so it can create the databases automatically.
- **Entry point**: The TimescaleDB must provide an entry point through a DNS Name or IP address that you can access from Workload Optimization Manager. You configure this access in the Workload Optimization Manager `charts_v1alpha1_xl_cr.yaml` file.

To create the TimescaleDB, complete the following tasks:

1. Install the DB package. For information about installing Timescale, see Install TimescaleDB.
2. Create a global R/W user account that can be used to create databases on the DB service.

   For information about automatically creating databases and users through R/W access, see Automatic Provisioning of Users and Databases *(on page 1804)*.
3. Optional: Manually create the users and databases that embedded reports use.

   If you cannot grant Workload Optimization Manager global R/W access to the DB, you must manually create the databases and users for embedded reports.

   For information about manually creating databases and users, see Manual Provisioning of Users and Databases *(on page 1801)*.

Manual Provisioning of users and databases

For Workload Optimization Manager to automatically create the required databases for embedded reports, the Workload Optimization Manager `charts_v1alpha1_xl_cr.yaml` file must specify credentials for a user account that has global R/W access to the DB service. To ensure this access, you can store the credentials in cleartext in the `charts_v1alpha1_xl_cr.yaml` file.

If Automatic Provisioning is appropriate for your environment, then you can go directly to Automatic Provisioning of Users and Databases *(on page 1804)*. That topic is where you specify the global R/W account that embedded reports can use to access your TimescaleDB.

If Automatic Provisioning is not appropriate for your environment, you must access the DB service with an R/W account and manually add the databases and user accounts that embedded reports can use. Then, after you add the databases to your TimescaleDB, you will edit the Workload Optimization Manager `charts_v1alpha1_xl_cr.yaml` file to use these databases.

Manually provisioning databases and users on the TimescaleDB

The following steps provision the database objects that embedded reporting requires. These examples use default names for illustration. You can substitute your own names for databases and users. Remember the names that you use because you must match them as you edit the Workload Optimization Manager `charts_v1alpha1_xl_cr.yaml` file.

To provision the databases and users, open a command session on the TimescaleDB, and run the following commands:

- Create two databases – one for `extractor` data, one for `grafana` data: The extractor database manages the Workload Optimization Manager data stream, and the Grafana database manages data for reporting within Grafana.

  Run the following commands, where you can specify your own database names instead of `extractor` and `grafana`:

  ```
  CREATE DATABASE extractor;
  CREATE DATABASE grafana;
  ```

- Create database users: For the extractor database, you create an R/W user, a read-only group, and a read-only user. Also create an R/Wn user for the Grafana database.

  Run the following commands, where you can specify your own names for instances of USER or ROLE:

  ```
  -\- main read/write user for extractor data
  CREATE USER extractor PASSWORD '<password>';
  -\- group for users with read-only access to extractor data
  CREATE ROLE readers_extractor_extractor;
  -\- read-only user for extractor data as a member of that group
  CREATE USER query PASSWORD '<password>';
  GRANT CONNECT ON DATABASE extractor TO readers_extractor_extractor;
  GRANT readers_extractor_extractor TO query;
  -\- read-write user for grafana data
  CREATE USER grafana_backend PASSWORD '<password>';
  ```

- Create and prepare the schema for extractor data. Connect to the extractor database and run these commands, where you can specify your own name for the SCHEMA, and you grant privileges to the users and roles you previously created:

  ```
  CREATE SCHEMA extractor;
  -\- read/write user has full access
  GRANT ALL PRIVILEGES ON SCHEMA extractor TO extractor;
  -\- all users in readers group have read-only access
  GRANT USAGE on SCHEMA extractor TO readers_extractor_extractor;
  GRANT SELECT ON ALL TABLES IN SCHEMA extractor TO readers_extractor_extractor;
  -\- make sure readers get access to any tables added in the future
  ALTER DEFAULT PRIVILEGES IN SCHEMA extractor GRANT SELECT ON TABLES TO readers_extractor_extractor;
  -\- make the extractor and query users use the extractor schema by default
  ALTER ROLE extractor SET search_path TO `extractor`;
  ALTER ROLE query SET search_path TO `extractor`;
  -\- install the timescaledb plugin into the extractor database using the extractor schema
  CREATE EXTENSION timescaledb SCHEMA extractor;
  ```

- Create and prepare the schema for Grafana data. Connect to the Grafana database and run these commands, where you can specify your own name for the SCHEMA, and you grant privileges to the users and roles you previously created:

  ```
  CREATE SCHEMA grafana;
  -\- read/write user has full access
  GRANT ALL PRIVILEGES ON SCHEMA grafana TO grafana_backend;
  -\- make sure the grafana user uses the grafana schema by default
  ALTER ROLE grafana_backend SET search_path TO 'grafana';
  ```

The previous commands provision the required databases and users. Your provisioning must be complete and correct for embedded reports to properly collect data and display it in Grafana reports and dashboards.

Editing the CR file for manually created databases

After you provision the required database objects, edit the Workload Optimization Manager `charts_v1alpha1_xl_cr.yaml` file to use these databases.

To edit the `charts_v1alpha1_xl_cr.yaml` file:

1. Open the `charts_v1alpha1_xl_cr.yaml` file for editing.

   The location of the file depends on the type of Workload Optimization Manager installation that you are configuring:

   - For an OVA installation of Workload Optimization Manager:

     Open an SSH terminal session on your Workload Optimization Manager instance.

     Log in with the System Administrator that you set up when you installed Workload Optimization Manager:

     - Username: `turbo`
     - Password: `<your_private_password>`

     Then edit the file:

     `/opt/Workload Optimization Manager/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml`

   - Workload Optimization Manager on a Kubernetes node or node cluster:

     Open the following file for editing:

     `deploy/crds/charts_v1alpha1_xl_cr.yaml`

2. Specify the endpoint for connecting to the external database.

   You can use the database service DNS, or you can use an IP address.

   Add the endpoint as the `externalTimescaleDBIP` property in the `spec: global:` section of the `.cr` file:

   ```
   spec:
     global:
       externalTimescaleDBIP: <host-or-IP>
   ```

3. Specify Workload Optimization Manager access to the databases that you provisioned.

   As you specify users and databases, be sure to match the names you used to provision users, databases, and schema. These instructions use the default naming.

   You specify:

   - Global read-only access to the `extractor` database `/spec/properties/global/dbs` This grants the `extractor` and the `api` components read-only access to the `extractor` database that you previously provisioned. Those components use the `query` user account.
   - Read/write access to the `extractor` and `grafana` databases `/spec/properties/extractor/dbs` This grants the `extractor` component read/write access to the `extractor` and `grafana` databases that you previously provisioned.
   - Read/write access for Grafana to access the `grafana` database `/spec/grafana/grafana.ini/database`

   Edit the `charts_v1alpha1_xl_cr.yaml` file to add the following entries:

   ```
   spec:
     properties:
       global:
         dbs:
           extractor:
             databaseName: extractor
             schemaName: extractor
             query:
               userName: query
               password: <defaault-password>
       extractor:
         dbs:
           extractor:
             userName: extractor
             password: <default-password>
           grafana:
             schemaName: grafana
     grafana:
   ```

```
    grafana.ini:
      database:
        name: grafana
        user: grafana_backend
        password: <default-password>
```

4. (Recommended) Block automatic provisioning by Workload Optimization Manager on the TimescaleDB.

   If you manually provisioned the embedded reports objects on the TimescaleDB instance, disable the options for Workload Optimization Manager to automatically run provisioning on the TimescaleDB instance.

   NOTE:
   Workload Optimization Manager cannot provision on the TimescaleDB instance if you do not specify a global user account.

   Set the following properties to `false`:

   - `shouldProvisionDatabase`: `spec/properties/global/dbs/postgresDefault/shouldProvisionDatabase` Determintes whether Workload Optimization Manager attempts to provision any database in scope of the definition.
   - `shouldProvisionUser`: `spec/properties/global/dbs/postgresDefault/shouldProvisionUser` Determines whether Workload Optimization Manager attempts to provision any logins in scope of the definition.
   - `destructiveProvisioningEnabled`: `spec/properties/global/dbs/postgresDefault/destructiveProvisioningEnabled` Determines whether, during its provisioning operations, Workload Optimization Manager can perform destructive operations like dropping databases, schemas, or users that are mis-configured. In the default installation scenario, the `public` schema that is created by PostgreSQL in any new database is dropped to reduce the complexity of the overall model. Because this operation is considered destructive, this option is `true` by default.

   Edit the `charts_v1alpha1_xl_cr.yaml` file to add the following properties:

```
spec:
  properties:
    global:
      dbs:
        postgresDefault:
            shouldProvisionDatabase: false
            shouldProvisionUser: false
            destructiveProvisioningEnabled: false
```

   Placing these settings in the `/spec/properties/global/dbs/postgresDefault` block ensures that they apply to all PostgreSQL databases and users who are used by any Turbonomic component. If your pre-provisioning is more selective, you can move the settings to more narrowly targeted property blocks like `/spec/properties/extractor/dbs/grafana` or `/spec/properties/global/dbs/extractor`.

Automatic provisioning of users and databases

For Workload Optimization Manager to automatically create the required databases for embedded reports, the Workload Optimization Manager `charts_v1alpha1_xl_cr.yaml` file must specify credentials for a user account that has global R/W access to the DB service. To enable this access, store the credentials in cleartext in the `charts_v1alpha1_xl_cr.yaml` file.

If Automatic Provisioning is not appropriate for your environment, you must access the DB service with an R/W account and manually add the databases and user accounts that embedded reports can use. See Manual Provisioning of Users and Databases *(on page 1801)*.

To enable an external TimescaleDB for embedded reports, you must edit the Workload Optimization Manager `charts_v1alpha1_xl_cr.yaml` file.

1. Open the `charts_v1alpha1_xl_cr.yaml` file for editing.

   The location of the file depends on the type of Workload Optimization Manager installation that you are configuring:

   - For an OVA installation of Workload Optimization Manager:

     Open an SSH terminal session on your Workload Optimization Manager instance.

     Log in with the System Administrator that you set up when you installed Workload Optimization Manager:

     - Username: `turbo`
     - Password: `<your_private_password>`

     Then edit the file:

     `/opt/Workload Optimization Manager/kubernetes/operator/deploy/crds/charts_v1alpha1_xl_cr.yaml`

   - Workload Optimization Manager on a Kubernetes node or node cluster:

     Open the following file for editing:

     `deploy/crds/charts_v1alpha1_xl_cr.yaml`

2. Specify the endpoint for connecting to the external database.

   For the connection endpoint, specify either the fully qualified domain name of the external server or its IP address.

   Add the endpoint to the `spec/global/externalTimescaleDBIP` property in the `charts_v1alpha1_xl_cr.yaml` file:

   ```
   spec:
   ...
     global:
       externalTimescaleDBIP: <host-or-IP>
   ```

3. Specify the global R/W account that Workload Optimization Manager uses to log in to the TimescaleDB service.

   This enables Workload Optimization Manager to automatically create the databases, schemas, and users that embedded reports requires.

   NOTE:
   Only complete this step if you want to enable global RW access to the TimescaleDB service in the `charts_v1alpha1_xl_cr.yaml` file. You specify the account credentials in clear text.

   If you choose not to specify the global account here, then you must manually provision the database objects that embedded reports requires. See Manual Provisioning of Users and Databases *(on page 1801)*.

   To grant the access Workload Optimization Manager needs, specify the username and password for an account that has global R/W privileges on the TimescaleDB instance. Add these credentials in the `/spec/properties/global/dbs/postgresDefault` block of the cr.yaml file, as follows:

   ```
   spec:
     properties:
       global:
         dbs:
           postgresDefault:
             rootUserName: <root-user>
             rootPassword: <root-password>
   ```

Reference: Working with YAML files

YAML is the primary file format to create and configure resources on kubernetes, including everything to do with the Workload Optimization Manager platform. The Custom Resource YAML provides a convenient, single place to define the majority of configuration details for Workload Optimization Manager. When editing YAML files, you must be careful to respect the file syntax, including indents in the file. General rules for edits include:

- Always uses spaces, not tabs, for all indentation. Since spacing and indentations matter, and can yield an invalid result or a parameter completely skipped over, you should work with YAML files with an editor that supports using vertical lines associated with indentations to visually spot a misalignment. If your editor of choice makes this difficult, you can use the linux `expand` utility when you're done, to convert tabs to equivalent spaces.
- Indention uses two spaces per level.
- Be careful to keep the same indentation for all properties in a given section.
- Never use the same property name twice in the same section. Doing this will render the YAML file invalid, though you will likely not see any notification of a problem. Rather, all but one of the property definitions will be silently ignored.

Spacing matters

For the Workload Optimization Manager Custom Resource, indentation defines where parameters are applied (globally or to specific components), so ensure you line up the text appropriately. The following example shows examples of specifications that are applied at a global level (`{“spec”:{“global”:[{“tag”:”8.6.4”}]}}`) which sets the container image tag for all instances. Then indented you see properties that are global for the remote database (`{“spec”:{“properties”:{"global":[{“dbPort”:”6033”}]}}}`) describes a property of `dbPort` that would be set for remote DB connections.

Each line is indented two spaces from the higher level.

```
spec:
  global:
    repository: turbonomic
    tag: 3.13.0
  properties:
    global:
      dbPort: 6033
  kubeturbo:
    enabled: true
  aws:
    enabled: true
```

Combine properties correctly

A YAML file is read beginning to end, and if there are different parameters that apply to the same component, they need to be combined. The following example shows a YAML where properties for the `ui` component of image tag *and* memory limit resources.

```
spec:
  global:
    repository: turbonomic
    tag: 3.13.0
  ui:
    image:
      tag: 3.13.0
  properties:
    global:
      dbPort: 6033
  kubeturbo:
    enabled: true
  aws:
    enabled: true
  ui:
    resources:
```

```
      limits:
        memory: 4Gi
```

This YAML will not set both image tag and memory limit resources for the `ui` component because they are set in two different sections. The memory limits would be applied since that is the last section to be read, and it would overwrite the first set of properties with the image tag. The following YAML will set both image tag and memory limit resources for the `ui` component:

```
spec:
  global:
    repository: turbonomic
    tag: 3.13.0
  ui:
    image:
      tag: 3.13.0
    resources:
      limits:
        memory: 4Gi
  properties:
    global:
      dbPort: 6033
  kubeturbo:
    enabled: true
  aws:
    enabled: true
```

To avoid having the same property name twice in the same section, search for an existing instance of the key you want to add. With the complex embedded structure in many k8s resources, this can be quite challenging; however, you can use an online validator that does not accept duplicate keys to check your entire CR file (such as YAMLint – The YAML Validator).

Finding a property

This documentation refers to specific properties in the CR file using a "path" expression. For example, assume the path `/spec/global/repository` to designate a `repository` property in the file. You can find the property in the file as follows:

1. Find a line that says `spec:` with no indentation at all.
2. Between that line and the next unindented line (not counting comments, which start with #), find a line that says `global:` and is at the next level of indentation.
3. Between that line and the next line with the same indentation, find a line that starts with `repository:`. That line is where the addressed property is defined.

In this example, find the `repository` property specified by `/spec/global/repository`:

```
apiVersion: charts.helm.k8s.io/v1alpha1
kind: Xl
metadata:
  name: xl-release
spec:
  properties:
    global:
      repository:        # This is NOT the correct property
        ...
```

```
  # Global settings
  global:
    repository:          # This is the one you want
```

The first `repository` property is not at `/spec/global/repository`, but at `/spec/properties/global/repository`.