Workload Optimization Manager 3.0.7
User Guide
# Contents

What’s New .......................................................................................................................... 7

Introducing Workload Optimization Manager ................................................................. 11
  How Workload Optimization Manager Works .............................................................. 11
    The Desired State ........................................................................................................ 12
    The Market and Virtual Currency .............................................................................. 13
    Risk Index .................................................................................................................. 14
    The Workload Optimization Manager Supply Chain .................................................. 14
  Workload Optimization Manager Targets .................................................................... 15
  Resource Descriptions ................................................................................................. 17

Getting Started .................................................................................................................. 21
  Logging In to Workload Optimization Manager ......................................................... 21
  The Home Page ........................................................................................................... 22
    APPLICATION View ................................................................................................ 22
    ON-PREM View ....................................................................................................... 22
    CLOUD View ........................................................................................................... 23
  Configuring Targets ...................................................................................................... 40
    AWS Billing Families ............................................................................................. 44
    Azure Enterprise Agreements .................................................................................. 46
  Supply Chain of Entities ............................................................................................... 48
  Working With a Scoped View ....................................................................................... 49
    Scoping the Workload Optimization Manager Session ............................................. 50
    Overview Charts ....................................................................................................... 54
    Details View ............................................................................................................. 55
    Scope Policies ........................................................................................................... 57
  List of Entities ............................................................................................................... 61
    Navigating With the Supply Chain ............................................................................ 62
    Viewing Cluster Headroom ...................................................................................... 63
  Workload Optimization Manager Actions ..................................................................... 64
    Actions by Entity Type ............................................................................................. 65
    Action Types ............................................................................................................. 68
    Action Categories ..................................................................................................... 71
    Action Modes .......................................................................................................... 73
    Working With the Generated Actions ...................................................................... 73
    Pending Actions ....................................................................................................... 75
    Actions Tips and Best Practices ................................................................................ 85
  Working With Policies .................................................................................................. 86
    Placement Policies ................................................................................................... 87
## Contents

<table>
<thead>
<tr>
<th>Entity Types - Applications</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automation Policies</td>
<td>.92</td>
</tr>
<tr>
<td>Business Application Policies</td>
<td>117</td>
</tr>
<tr>
<td>Business Transaction Policies</td>
<td>119</td>
</tr>
<tr>
<td>Business Transaction</td>
<td>120</td>
</tr>
<tr>
<td>Service Policies</td>
<td>121</td>
</tr>
<tr>
<td>Application Component Policies</td>
<td>122</td>
</tr>
<tr>
<td>Application Component</td>
<td>123</td>
</tr>
<tr>
<td>Application Topology</td>
<td>124</td>
</tr>
<tr>
<td>Kubernetes CPU Metrics</td>
<td>127</td>
</tr>
<tr>
<td>Container Policies</td>
<td>128</td>
</tr>
<tr>
<td>Container Spec Policies</td>
<td>131</td>
</tr>
<tr>
<td>Workload Controller Policies</td>
<td>132</td>
</tr>
<tr>
<td>Container Pod Policies</td>
<td>137</td>
</tr>
<tr>
<td>Container Cluster</td>
<td>138</td>
</tr>
<tr>
<td>Namespace</td>
<td>139</td>
</tr>
<tr>
<td>Container Cluster</td>
<td>142</td>
</tr>
<tr>
<td>Kubernetes CPU Metrics</td>
<td>143</td>
</tr>
<tr>
<td>Workload Controller</td>
<td>146</td>
</tr>
<tr>
<td>Container Pod Policies</td>
<td>146</td>
</tr>
<tr>
<td>Container Cluster</td>
<td>149</td>
</tr>
<tr>
<td>Kubernetes CPU Metrics</td>
<td>152</td>
</tr>
<tr>
<td>Entity Types - Cloud Infrastructure</td>
<td>155</td>
</tr>
<tr>
<td>Virtual Machine (Cloud)</td>
<td>156</td>
</tr>
<tr>
<td>Cloud VM Policies</td>
<td>161</td>
</tr>
<tr>
<td>Database Server (Cloud)</td>
<td>165</td>
</tr>
<tr>
<td>Volume (Cloud)</td>
<td>166</td>
</tr>
<tr>
<td>Cloud Volume Policies</td>
<td>173</td>
</tr>
<tr>
<td>Database (Cloud)</td>
<td>175</td>
</tr>
<tr>
<td>Cloud Database Policies</td>
<td>178</td>
</tr>
<tr>
<td>Zone</td>
<td>180</td>
</tr>
<tr>
<td>Region</td>
<td>182</td>
</tr>
<tr>
<td>Entity Types - On-prem Infrastructure</td>
<td>185</td>
</tr>
<tr>
<td>Virtual Machine (On-prem)</td>
<td>186</td>
</tr>
<tr>
<td>On-prem VM Policies</td>
<td>192</td>
</tr>
<tr>
<td>Database Server (On-prem)</td>
<td>196</td>
</tr>
<tr>
<td>On-prem Database Server Policies</td>
<td>199</td>
</tr>
<tr>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>Volume (On-prem) ...........................................................................................................</td>
</tr>
<tr>
<td>201</td>
<td>On-prem Volume Policies .................................................................................................</td>
</tr>
<tr>
<td>201</td>
<td>Virtual Datacenter (Private Cloud) .................................................................................</td>
</tr>
<tr>
<td>205</td>
<td>Virtual Datacenter (Container Orchestration) .................................................................</td>
</tr>
<tr>
<td>207</td>
<td>Business User ..................................................................................................................</td>
</tr>
<tr>
<td>209</td>
<td>Business User Policies .....................................................................................................</td>
</tr>
<tr>
<td>210</td>
<td>Desktop Pool ....................................................................................................................</td>
</tr>
<tr>
<td>212</td>
<td>Desktop Pool Policies .......................................................................................................</td>
</tr>
<tr>
<td>214</td>
<td>View Pod ..........................................................................................................................</td>
</tr>
<tr>
<td>215</td>
<td>View Pod Policies .............................................................................................................</td>
</tr>
<tr>
<td>215</td>
<td>Host .................................................................................................................................</td>
</tr>
<tr>
<td>218</td>
<td>Host Policies ....................................................................................................................</td>
</tr>
<tr>
<td>220</td>
<td>Chassis ............................................................................................................................</td>
</tr>
<tr>
<td>221</td>
<td>Datacenter .......................................................................................................................</td>
</tr>
<tr>
<td>222</td>
<td>Storage .............................................................................................................................</td>
</tr>
<tr>
<td>226</td>
<td>Storage Policies ...............................................................................................................</td>
</tr>
<tr>
<td>230</td>
<td>Logical Pool .....................................................................................................................</td>
</tr>
<tr>
<td>232</td>
<td>Logical Pool Policies .......................................................................................................</td>
</tr>
<tr>
<td>233</td>
<td>Disk Array .......................................................................................................................</td>
</tr>
<tr>
<td>234</td>
<td>Disk Array Policies .........................................................................................................</td>
</tr>
<tr>
<td>237</td>
<td>Storage Controller ...........................................................................................................</td>
</tr>
<tr>
<td>237</td>
<td>Storage Controller Policies ............................................................................................</td>
</tr>
<tr>
<td>238</td>
<td>IO Module .......................................................................................................................</td>
</tr>
<tr>
<td>239</td>
<td>Switch .............................................................................................................................</td>
</tr>
<tr>
<td>240</td>
<td>Switch Policies ...............................................................................................................</td>
</tr>
<tr>
<td>242</td>
<td>Plans: Looking to the Future ..........................................................................................</td>
</tr>
<tr>
<td>243</td>
<td>Plan Management .............................................................................................................</td>
</tr>
<tr>
<td>244</td>
<td>Setting Up Plan Scenarios ..............................................................................................</td>
</tr>
<tr>
<td>251</td>
<td>Plan Scenarios and Types ...............................................................................................</td>
</tr>
<tr>
<td>254</td>
<td>Optimize Container Cluster Plan .....................................................................................</td>
</tr>
<tr>
<td>268</td>
<td>Optimize Cloud Plan .......................................................................................................</td>
</tr>
<tr>
<td>276</td>
<td>Migrate to Cloud Plan .....................................................................................................</td>
</tr>
<tr>
<td>286</td>
<td>Alleviate Pressure Plan .................................................................................................</td>
</tr>
<tr>
<td>290</td>
<td>Custom Plan ....................................................................................................................</td>
</tr>
<tr>
<td>301</td>
<td>Configuring Nightly Plans ..............................................................................................</td>
</tr>
<tr>
<td>303</td>
<td>Place: Reserve Workload Resources ..............................................................................</td>
</tr>
<tr>
<td>305</td>
<td>Creating a Reservation .................................................................................................</td>
</tr>
<tr>
<td>307</td>
<td>Managing Reservations .................................................................................................</td>
</tr>
<tr>
<td>309</td>
<td>Dashboards: Focused Views ...........................................................................................</td>
</tr>
</tbody>
</table>
## Contents

Built-in Dashboards .................................................................................................................. 310
  On-Prem Executive Dashboard .............................................................................................. 310
  Cloud Executive Dashboard .................................................................................................. 312
  Container Platform Dashboard .............................................................................................. 313
Creating and Editing Custom Dashboards ............................................................................... 314
Creating and Editing Chart Widgets ....................................................................................... 316
Chart Types ............................................................................................................................. 320
  Actions and Impact Chart Types ......................................................................................... 320
  Status and Details Chart Types ........................................................................................... 327
Cloud Chart Types ................................................................................................................... 336
On-Prem Chart Types .............................................................................................................. 352

Embedded Reporting ............................................................................................................... 355

Creating Groups ....................................................................................................................... 357

Working With Schedules ......................................................................................................... 360
  Creating Schedules ............................................................................................................... 362
Templates: Resource Allocations for New Entities .................................................................. 366
  Creating Templates ............................................................................................................... 367
  VM Template Settings ......................................................................................................... 368
  Host Template Settings ........................................................................................................ 369
  HCI Host Template Settings ................................................................................................. 370
  Storage Template Settings .................................................................................................... 372

Billing and Costs ....................................................................................................................... 374
  RI Purchase Profile ............................................................................................................... 375
  OS Migration Profile ............................................................................................................ 376
  Hardware Costs .................................................................................................................... 377
  Price Adjustments ................................................................................................................ 377
    Creating a Price Adjustment ............................................................................................... 378
    Price Override: Azure ......................................................................................................... 382
    Price Override: AWS .......................................................................................................... 384
  Currency Settings .................................................................................................................. 385

Administrative Tasks ................................................................................................................ 386
  Managing User Accounts ..................................................................................................... 386
    Configuring a Group for SSO Authentication ................................................................... 394
  Maintenance: Logging and Troubleshooting ........................................................................ 396
  The Updates Page .................................................................................................................. 397
  License Configuration ............................................................................................................ 398
  Email Settings ....................................................................................................................... 399
What's New

Workload Optimization Manager is powered by our next-generation architecture, allowing the core platform to scale with large application and infrastructure environments in a single-instance deployment. This eliminates complexity and provides scale-on-demand capabilities, while continuing to assure application performance and health.

Version 3.0.6

- **Azure Government Support**
  Workload Optimization Manager now discovers Azure Government workloads when you add the Azure Government service principal and Enterprise Agreement (EA) accounts as targets. It monitors these workloads and then recommends actions to address performance issues and optimize costs.
  
  For details on the required targets, see "Azure Government Targets" in the Target Configuration Guide.
  
  For details on Azure Government support, see Support for Government Workloads (on page 35).

- **Cumulative Savings and Investments**
  Use the Cumulative Savings and Cumulative Investments charts to track realized/missed savings and investments for your cloud workloads. Information in these charts can help shape 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.

  For details, see Cumulative Savings and Investments Charts (on page 339).

- **Integration with Grafana 8.0**
  Workload Optimization Manager now ships with Grafana 8.0 to add capabilities to Embedded Reporting. With this Grafana version, you can schedule and email tabular reports in CSV format, and use variables when scheduling dashboards.

  When you update to this Workload Optimization Manager version, your existing Grafana version will automatically upgrade as part of the update process.

Version 3.0.5

- **Kubernetes CPU Metrics**
  To meet user requirements and align with Kubernetes specifications, CPU metrics for your Kubernetes platform are now expressed in millicore (mCore) units.

  For details, see Kubernetes CPU Metrics (on page 152).

- **Savings Plans Coverage Chart**
  Use the Savings Plans Coverage chart to track the percentage of VMs covered by AWS Savings Plans. If you have a high percentage of on-demand VMs, you can reduce your monthly costs by increasing Savings Plans coverage.
NOTE:
For cloud environments with both Reserved Instances and Savings Plans, coverage data is included in the **RI and Savings Plans Coverage** chart.

For details, see **Savings Plans Coverage (on page 344)**.

- **Offline Update Script**
To simplify offline updates to Workload Optimization Manager, this release introduces the `isoupdate.sh` script.
For details, see "Offline Update" in the **Installation Guide**.

### Version 3.0.4

- **Support for JBoss Application Servers**
  Workload Optimization Manager now supports JBoss Application Servers as targets. It monitors the utilization of server resources and shows utilization data in charts.
  For details, see "JBoss" in the **Target Configuration Guide**.

- **Enhancement for Kubernetes Node Actions**
  When recommending actions to provision nodes to address growing demand, Workload Optimization Manager can now represent demand from DaemonSets on the new nodes. This ensures that any application workload can be placed with a better representation of available capacity. For actions to suspend nodes, Workload Optimization Manager will continue to recommend suspending the related DaemonSets.

### Version 3.0.3

- **Support for AWS GovCloud (US)**
  Workload Optimization Manager now discovers **AWS GovCloud (US)** workloads when you add GovCloud accounts and the associated billing targets. It monitors these workloads and then recommends actions to address performance issues and optimize costs.
  For details on the required targets, see "AWS GovCloud Targets" in the **Target Configuration Guide**.
  For details on GovCloud support, see **Support for Government Workloads (on page 35)**.

- **Uptime Data for Cloud VMs**
  For cloud VMs, Workload Optimization Manager now 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.
  For details, see **Cloud VM Uptime (on page 38)**.

- **Cost Impact of Kubernetes Node Actions**
  Workload Optimization Manager recommends actions to provision or suspend Kubernetes nodes (VMs) to address workload congestion, meet application demand, and improve infrastructure efficiency. For nodes in the public cloud, charts in the real-time market and in container cluster plans now reflect the investments and savings associated with these actions.
  For details, see **Actions for Kubernetes Nodes (on page 158)**.

### Version 3.0.2

- **Instana Support**
  Workload Optimization Manager now supports the Instana platform as an application target. It discovers the applications that Instana manages and then includes discovered information about these applications in its calculations of environment health.
  For details, see "Instana" in the **Target Configuration Guide**.
Version 3.0.1

- **Container Resource Management**
  - **Optimize Container Cluster Plan**
    This release introduces a new plan type called Optimize Container Cluster. Run this plan to identify performance and efficiency opportunities for a single Kubernetes 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 details, see [Optimize Container Cluster Plan](on page 254).
  - **Labels and Annotations for Container Entities**
    Workload Optimization Manager now discovers Workload Controller labels and annotations as tag properties, and then applies the same properties to the associated Container Specs. You can filter these entities by labels or annotations when you use Search or create groups for your container policies.
  - **Enhancements for Kubernetes Environments**
    - When recommending Move actions for a pod, Workload Optimization Manager now 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.
    - This release introduces the **Number of Replicas** chart. Use the chart to see adjustments to the Application Component replicas over a given time period. For example, if Kubernetes Horizontal Pod Autoscaler (HPA) is enabled for a Service, Workload Optimization Manager tracks adjustments to the number of replicas made by HPA.

- **Cloud Resource Management**
  - **Improved Cloud Analytics**
    We have improved our cloud analytics to increase the accuracy of our recommendations and deliver better results. As you continue to execute the actions that we recommend for your cloud environment, you can expect to see increased RI coverage and additional savings.
    You can see immediate improvements if you have VMs covered by Azure Hybrid Benefit. When calculating costs for these VMs, Workload Optimization Manager now considers workload licensing rates offered through the benefit. This allows Workload Optimization Manager to calculate costs more accurately and recommend correct scaling actions for the VMs.
  - **Azure VM Throughput Scaling**
    Workload Optimization Manager can now recommend scaling uncached IO throughput capacity for Azure VMs. It uses percentile calculations to analyze throughput utilization more accurately and then picks the most suitable tier for scaling.
  - **Support for Azure Site Recovery and Azure Backup Volumes**
    Workload Optimization Manager discovers Azure Site Recovery and Azure Backup volumes when you add Azure targets, and shows capacity and cost for each volume. Because these volumes are critical to business continuity and disaster recovery, Workload Optimization Manager will never recommend deleting them.
    For details, see [Delete Actions](on page 168).
  - **Integration with Azure Monitor Log Analytics**
    Workload Optimization Manager can now collect performance metrics for your Azure VMs via Azure Monitor Log Analytics. It discovers the Log Analytics workspaces that you have created and then retrieves data periodically. With this additional source of data, Workload Optimization Manager can analyze memory utilization for the VMs more accurately.
    For details, see [Performance Metrics for Azure VMs](on page 161).
  - **Currency Settings**
    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 now set a different symbol to match your preferred currency.
    Currency symbols are for display purposes only. Workload Optimization Manager does not convert monetary values when you switch symbols.
What's New

For details, see Currency Settings (on page 385).

- **AWS Savings Plans Utilization**

  Use the Savings Plans Utilization chart to evaluate how well you have utilized your AWS Savings Plans inventory. Workload Optimization Manager calculates utilization based on the billing data that it discovers via AWS Cost Explorer.

  **NOTE:**
  For cloud environments with both Reserved Instances and Savings Plans, utilization data is included in the RI and Savings Plans Utilization chart.

  For details, see Savings Plans Utilization (on page 349).

- **Chart Enhancements**

  For cloud environments, the Necessary Investments and Potential Savings charts have been enhanced for ease of use. In the ring chart or legend, you can now click an action type (for example, Scale Volumes or Buy RI) to display a filtered view of the actions list. In addition, when you click Show All, you will first see the actions that will incur the highest costs or introduce the most savings.

- **On-prem Resource Management**

  - **Support for Oracle WebLogic and IBM WebSphere**

    Workload Optimization Manager now supports Oracle WebLogic and IBM WebSphere as targets, and can recommend resizing heap and thread pool capacity. Information about the resources required to ensure the performance of these targets will also be part of sizing the underlying infrastructure.

    For details, see:
    - "Oracle WebLogic" in the Target Configuration Guide
    - "IBM WebSphere" in the Target Configuration Guide

- **Embedded Reporting**

  If you have enabled Embedded Reporting in your Workload Optimization Manager instance:

  - You can now create custom reports by running SQL queries against the instance’s Embedded Reporting database.
  - Workload Optimization Manager now collects the following data for reports:
    - Metrics: Peak and cluster headroom values
    - Actions: Percentile values for current actions
  - To manage reports, a user must have the Report Editor role. Only one user per instance is allowed to have this role (by default, the local administrator user). You can assign the role to another user.
  - You can now configure an external TimescaleDB server as a data source for your reports.

  For details, see Embedded Reporting (on page 355).
Introducing Workload Optimization Manager

Thank you for choosing Workload Optimization Manager, the premier solution for Application Resource Management (ARM) of cloud and virtual environments.

Application Resource Management is a top-down, 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 (VMs, instances, containers, and services) with a budget to seek the resources that applications need to perform, and sellers to price their available resources (CPU, memory, storage, network) based on utilization in real-time, Workload Optimization Manager keeps your environment within the desired state — operating conditions that achieve the following conflicting goals at the same time:

- Assured application performance
  Prevent bottlenecks, upsize containers/VMs, prioritize workload, and reduce storage latency.
- Efficient use of resources
  Consolidate workloads to reduce infrastructure usage to the minimum, downsize containers, prevent sprawl, and use the most economical cloud offerings.

Workload Optimization Manager is a containerized, microservices architected application running in a Kubernetes environment (or within a VM) on your network or a public cloud VPC. You then assign services running on your network to be Workload Optimization Manager targets. Workload Optimization Manager discovers the entities (physical devices, virtual components and software components) that each target manages, and then performs analysis, anticipates risks to performance or efficiency, and recommends actions you can take 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. On top of that, 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 the _The Supply Chain (on page 48)_ 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

![Desired State Diagram](image_url)

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 upper 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 upper 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 the The Supply Chain [on page 48](#) 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 86).

**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. For a Kubernetes target, it discovers services, namespaces, containers, container pods, and nodes. 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 example, by connecting the discovered Kubernetes nodes with the discovered VMs in vCenter.

For information about specific members of the supply chain, see The Supply Chain (on page 48).

**Supply Chain Terminology**

Cisco 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.

<table>
<thead>
<tr>
<th>Term:</th>
<th>Definition:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commodity</td>
<td>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 <em>commodities</em>, it’s showing the resources a service provides. When the interface shows <em>commodities bought</em>, it’s showing what that service consumes.</td>
</tr>
<tr>
<td>Composed Of</td>
<td>The resources or commodities that make up the given service. For example, in the user interface you might see that a certain VM is <em>composed of</em> commodities such as one or more physical CPUs, an Ethernet interface, and physical memory. Contrast <em>Composed Of</em> with <em>Consumes</em>, where consumption refers to the commodities the VM has bought. Also contrast <em>Composed Of</em> with the commodities a service offers for sale.</td>
</tr>
</tbody>
</table>
Introducing Workload Optimization Manager

<table>
<thead>
<tr>
<th>Term:</th>
<th>Definition:</th>
</tr>
</thead>
<tbody>
<tr>
<td>A host might include four CPUs in its composition, but it offers CPU Cycles as a single commodity.</td>
<td></td>
</tr>
<tr>
<td>Consumes</td>
<td>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.</td>
</tr>
<tr>
<td>Entity</td>
<td>A buyer or seller in the market. For example, a VM or a datastore is an entity.</td>
</tr>
<tr>
<td>Environment</td>
<td>The totality of data center, network, host, storage, VM, and application resources that you are monitoring.</td>
</tr>
<tr>
<td>Inventory</td>
<td>The list of all entities in your environment.</td>
</tr>
<tr>
<td>Risk Index</td>
<td>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. 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.</td>
</tr>
</tbody>
</table>

Workload Optimization Manager Targets

You can assign instances of the following technologies as Workload Optimization Manager targets:

- **Cloud Native**
  - Kubernetes
  - OpenShift 3.3+
- **Fabric and Network**
  - Cisco UCS Manager 3.1+
  - HPE OneView 3.00.04+
- **Guest OS Processes**
  - SNMP
- **Hyperconverged**
  - Hyperflex 2.5, 2.6
  - Nutanix Community Edition
  - VMware vSAN
- **Hypervisors**
  - Citrix XenServer 5.6.x and 6.x
  - Microsoft Hyper-V 2008 R2, Hyper-V 2012/2012 R2, Hyper-v 2016
  - VMware vCenter 5.1, 5.5, 6.0, 6.5, 6.7, and 7.0+
- **Orchestrator**
  - Action Script
  - ServiceNow
- **Private Cloud**
  - Microsoft System Center 2012/2012 R2 Virtual Machine Manager and System Center 2016 Virtual Machine Manager
  - OpenStack Havana – Queens
• Public Cloud
  – Amazon AWS
  – Amazon AWS Billing
  – Microsoft Azure Service Principal
  – Microsoft Enterprise Agreement
• Storage
  – EMC ScaleIO 2.x and 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+
  – NetApp Cmode/7mode using ONTAP 8.0+ (excluding AFF and SolidFire)
  – Pure Storage F-series and M-series arrays
• Virtual Desktop Infrastructure
  – VMware Horizon

The following sections describe these targets. For information about assigning targets to Workload Optimization Manager, see the Target Configuration Guide.

Hypervisors

Workload Optimization Manager can use a range of VM managers as targets. For general discussion, this document refers to the various supported VM managers as hypervisors.

Workload Optimization Manager supports the following hypervisor targets:
  • Microsoft Hyper-V
  • VMware vCenter

Workload Optimization Manager uses hypervisor targets to access information about the managed VMs, hosts, and datastores, and also to execute commands such as provisioning, resizing, or reconfiguring entities in the environment. Through the hypervisor, Workload Optimization Manager can perform system monitoring, report on wasted storage, recommend actions, execute moves for VMs and VM storage, and execute VM reconfiguration (change CPU count, memory, etc.).

The entities Workload Optimization Manager discovers through hypervisor targets include:
  • VMs
  • Physical machines that host VMs
  • Datastores that support the VMs
  • Datacenters

Cloud Managers

Cloud Managers provide a layer of control to deliver virtual infrastructures that can be deployed automatically, or in a self-service offering to customers. They define and manage virtual datacenters (VDCs) – provider VDCs to manage the physical and virtual resources that support the cloud offering, and consumer VDCs that present limited resources to customers.

Workload Optimization Manager supports the following cloud manager targets:
  • Microsoft Virtual Machine Manager (VMM)

Workload Optimization Manager has visibility into the full VDC chain, from the resources provided by the underlying hosts and physical datastores, through the resources consumed by a provider VDC, to the resources consumed by VMs hosted on a consumer VDC.

You can create special Workload Optimization Manager user accounts for consumer VDC customers. Such an account has a limited scope, and the user cannot see any of the resources outside of that scope. In this way, you can offer Workload Optimization Manager to cloud customers without exposing any proprietary infrastructure data to them. For more information, see Managing User Accounts (on page 386).
Introducing Workload Optimization Manager

The entities Workload Optimization Manager discovers through cloud manager targets include:

- **Consumer VDCs**
  Virtual resources that are available to customers.
- **Provider VDCs**
  Physical resources that provide the infrastructure to support Consumer VDCs.

### Storage Managers

Storage managers provide management and distribution of data storage across disk arrays. Storage managers can support thin provisioning, deduplication, and HA architectures. Workload Optimization Manager monitors resource utilization across the storage system to optimize placement and provisioning of volumes and disk arrays, as well as management of storage controller resources.

Workload Optimization Manager supports the following storage manager targets:

- **NetApp Storage Systems running Data ONTAP version 8 or later**
  The actions Workload Optimization Manager can recommend and perform are different for systems running in 7-Mode or Cluster-Mode.
- **EMC VNX Series Storage Systems** — for version details, see the [EMC VNX Support KB article](#).

The entities Workload Optimization Manager discovers through storage manager targets include:

- **Storage Controllers (NetApp controllers/filers, VNX processors)**
- **Disk Arrays (aggregates, clustered aggregates, storage pools, RAID groups)**
- **Datastores (volumes or LUNs)**

### Fabric Managers

Fabric managers provide a point of control for fabrics that unify compute, network, storage, and virtual resources within a single system.

Workload Optimization Manager supports the following fabric manager targets:

- **Cisco UCS Fabric Manager**

The entities Workload Optimization Manager discovers through fabric managers targets include:

- **UCS Domains**
- **Chassis**
- **Fabric Interconnects**
- **IO Modules**

### Resource Descriptions

To perform intelligent workload balancing, Workload Optimization Manager collects raw data from its target servers – hypervisors, cloud management stacks, public cloud accounts, etc. Workload Optimization Manager polls its targets at 10-minute intervals to collect the latest data samples. It then uses these 10-minute data points for analysis and to display data in the GUI.

The way Workload Optimization Manager collects host memory data from vCenter Server illustrates how this works. vCenter Server collects peak metrics from its managed VMs at 20-second intervals. Every ten minutes Workload Optimization Manager polls vCenter Server to collect its last round of data samples (30 samples in 10 minutes). To track a VM’s utilization of host
Introducing Workload Optimization Manager

memory, Workload Optimization Manager requests memory.active data samples from vCenter. From that polling, Workload Optimization Manager can track:

- Peak Memory Utilization - Workload Optimization Manager uses the greatest value in each polling sample. This gives the highest percentage of active memory utilization for the selected VM (or group of VMs), calculated over the selected time period. For a maximum value, Workload Optimization Manager uses the highest observed active memory value in the data sample.
- Average Memory Utilization - Workload Optimization Manager averages all the values in each polling sample.

The following table lists the metrics Workload Optimization Manager collects, and includes details about how they are collected or measured. When the Workload Optimization Manager user interface plots charts of clusters or groups of devices, these charts show the average of the percentage of allocated resources that are used.

<table>
<thead>
<tr>
<th>Resource:</th>
<th>Description:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 2 - 4-CPU Rdy</td>
<td>Wait time in the ready queue on the host, measured in ms. Workload Optimization Manager monitors 1-CPU, 2-CPU, 4-CPU, up to 32-CPU ready queues on hosts. Charts show 1 - 4 CPU values. The charts show the percentage allocated ready queue capacity that is in use on the host. For host charts, this is a measure of the total ready queue wait time for all the VMs running on that host.</td>
</tr>
</tbody>
</table>
| Balloon                       | Ballooning capacity on the PM, measured in KBytes. This capacity is the greater of:  
                                  |  65% of the VMem configured for all powered-on VMs that the PM hosts  
                                  |  The physical memory capacity of the PM  
                                  | Charts show the percentage of the PM’s ballooning capacity that is in use. |
| Buffer                        | For network environments that support buffered switch ports (Arista networks), this resource measures utilization of a port buffer. For example, if a host connects to the network through port 1 on a switch, and that port has enough traffic to cause packet buffering, this resource will show utilization. |
| Connection                    | The connections in use, as a percentage of the maximum connections allowed on the database. Database configuration determines the capacity for this resource. |
| Cooling                       | Allocated cooling indicates the highest acceptable running temperature for a physical device, such as a chassis in a compute fabric. |
| CPU                           | Host CPU capacity, measured in MHz. This shows what percentage of CPU cycles are devoted to processing instructions.  
                                  |  Host charts show the percentage of the host’s CPU capacity that is in use.  
<pre><code>                              |  VM charts show the percentage of the host’s CPU capacity that is consumed by the given VM. |
</code></pre>
<p>| DBMem                         | The memory in use by the database, as a percentage of the allocated capacity. Database configuration determines the capacity for this resource. Note that for databases, Workload Optimization Manager uses this resource to drive actions, instead of the VMem on the hosting VM. This means that actions are driven by the actual memory consumption on the database. |
| Flow0 – InProvider Flow       | For measuring network flow, the flow that is within a single provider – For example, the network flow between VMs that are hosted by the same physical machine. This measures network flow between consumers that are on the same set of closely connected providers. Charts show the percentage of capacity that is utilized. Note that Workload Optimization Manager assumes an unlimited supply of InProvider Flow because this flow does not go across the physical network. |
| Flow1 – InDPOD Flow           | For measuring network flow, the flow that is local to the given DPOD. This measures network flow between consumers that are on the same set of closely connected providers. Charts show the percentage of capacity that is utilized. |
| Flow2 – CrossDPOD Flow        | For measuring network flow, the flow that is between different DPODs. This measures network flow between consumers that are on different sets of closely connected providers. Charts show the percentage of capacity that is utilized. |</p>
<table>
<thead>
<tr>
<th>Resource:</th>
<th>Description:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heap</td>
<td>The heap capacity allocated for an Application Component. Charts show the percentage of capacity that is used by an Application Component.</td>
</tr>
<tr>
<td>HotStorage</td>
<td>For Nutanix platforms, the storage capacity on the server-attached flash.</td>
</tr>
</tbody>
</table>
| IO           | Data rate through the host’s IO adapter, measured in KBytes/sec.  
• Datacenter charts show the average percentage of the host IO capacity that is in use, for all the hosts in the datacenter.  
• Host charts show the percentage of the host’s total IO capacity that is in use.                                                                                                                                                    |
| IOPS         | Storage access operations per second. Charts show the percentage of allocated IOPS capacity that is used on a datastore.                                                                                                                                                                                |
| Latency      | Allocated capacity for latency on a datastore. This measures the latency experienced by all VMs and hosts that access the datastore. Charts show the percentage of allocated latency that is in use on the datastore.                                                                                  |
| Mem          | Host memory, measured in Kbytes.  
• Host charts show the percentage of the host’s memory that is in use.  
• VM charts show the percentage of the host’s memory that is consumed by the given VM.                                                                                                                                                     |
| NET          | Data rate through the host’s Network adapter, measured in Kbytes/sec.  
• Datacenter charts show the average percentage of the host NET capacity that is used for all the hosts in the datacenter.  
• Host charts show the percentage of the host’s total NET capacity that is in use.                                                                                                                                                        |
| nfu (AWS only)| Normalized Factor Unit.  
For RIs in AWS environments, the nfu is a measure of RI capacity that you can use to compare or combine the capacity for different template families. For example, the normalized factors for some template families include:  
• nano: 0.25  
• micro: 0.5  
• small: 1  
• medium: 2  
• large: 4  
Workload Optimization Manager measures RI utilization and coverage in terms of these normalized factors.                                                                                                                                   |
| Power        | A measure of the power that is consumed by a physical device.                                                                                                                                                                                                                                                                                                    |
| RI ratio (Azure only) | For Azure environments, RI ratio is the number of RI units compared to the total number of RI units for a given Workload Optimization Manager scope. Each workload is assigned RI units based on its instance type. For example, here are some instance types with RI units:  
• Standard_DS2_v2: 1  
• Standard_B2ms: 3  
RI ratio information appears in the tooltips of cloud RI charts. Information about the Azure instance types and their RI workloads is provided in the RI Inventory chart.  
Azure RI ratio and AWS NFU are equivalent concepts.                                                                                                                                                |
<p>| Swap         | The rate of memory swapping to disk, in bytes per second. The default capacity is 5,000,000 Byte/sec.                                                                                                                                                                                                   |
| Threads      | Allocated thread capacity. Charts show the percentage of thread capacity that is consumed by an Application Component.                                                                                                                                                                                     |</p>
<table>
<thead>
<tr>
<th>Resource:</th>
<th>Description:</th>
</tr>
</thead>
<tbody>
<tr>
<td>TransactionLog</td>
<td>The disk space devoted to transaction logging for a database.</td>
</tr>
<tr>
<td>Transactions</td>
<td>Transactions per second. Charts show the percentage of allocated transaction capacity that is in use.</td>
</tr>
<tr>
<td>Risk Index</td>
<td>A measure of the impact on 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. For all the resources that impact performance or risk, charts show the Risk Index for the most utilized resource of a given entity. For example, if a host has a Risk Index of 6 for MEM and 12 for CPU, the chart will show the higher value.</td>
</tr>
<tr>
<td>VCPU</td>
<td>The allocated CPU capacity, measured in MHz. Charts show the percentage of VCPU cycles that are devoted to processing instructions.</td>
</tr>
<tr>
<td>VMem</td>
<td>The allocated memory capacity, measured in Kbytes. Charts show the percentage of VMem that is in use.</td>
</tr>
<tr>
<td>VMem</td>
<td>Note that percentages of allocated VMem are measured against whichever is the less of: The VMem limit (if set) or the allocated VMem capacity. This is also true in reports and recommended actions. For example, assume a VM with allocated VMem of 8 GB, but a limit of 4 GB. In this case, the percentage in a chart shows the percentage utilized of 4 GB.</td>
</tr>
<tr>
<td>VStorage</td>
<td>The allocated virtual storage capacity, measured in Kbytes. Charts show the percentage of storage that is in use.</td>
</tr>
</tbody>
</table>
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 22). 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" in the Target Configuration Guide.
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 117).
  - **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 117). 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 120) and Services (on page 122) that make up your Business Applications. You can see finer details and set SLOs at these levels of the application model.

**NOTE:**

**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.
Getting Started

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 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
Getting Started

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 "Private Cloud" in the Target Configuration Guide.

In addition, to view full cost information in AWS, you must have created a Cost and Usage report in your AWS account and you must store it in an S3 bucket.

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 RI resources and put them into active use.

- Charts that show your current Reserved Instance strategy. For details, see RI Charts (on page 34).

- 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

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 charts such as:

- Cloud Estimated Cost
- Cost Breakdown by Cloud Accounts, Component, or Service Provider
- Expenses

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.
For Azure environments, Workload Optimization Manager can track OS costs for existing VMs. For RI Buy actions, Workload Optimization Manager does not include the OS cost. For more information about Azure RIs, see Azure Enterprise Agreements (on page 46).

• IP
For some workloads, you might use IP services that incur a cost. For example, your cloud provider might charge to grant a static IP to a VM. On AWS environments Workload Optimization Manager can include that cost in its calculation and analysis.
Workload Optimization Manager uses this cost information when making VM resize and placement decisions. You can see this information in Expenses charts.

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.
Using scope to minimize this effect. You can create separate dashboards for your DT and Shared Tenancy workloads.
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.

AWS VMs and Azure VMs Without License Costs

Cost Calculation

For these VMs, the calculation for Estimated On-demand Monthly Cost can be expressed as follows:

\[
\text{On-demand Rate} \times \text{Usage Not Covered by RIs} \times \text{Uptime} \times 730 = \text{Estimated On-demand Monthly Cost}
\]

Where:

- **On-demand Rate** is the hourly cost for a VM's instance type *without* RI coverage.
  - For AWS, this rate includes all license costs, but not storage or IP. You can obtain on-demand rates via [Amazon EC2 On-demand Pricing](https://aws.amazon.com/ec2/pricing/on-demand/).
  - For Azure, the rate does *not* include license costs, storage, or IP. You can obtain on-demand rates via [Azure Pricing Calculator](https://azure.microsoft.com/pricing/calculator/).

  **NOTE:**
  Azure VMs covered by Azure Hybrid Benefit do not have license costs.

- **Usage Not Covered by RIs** is the percentage of hourly VM usage not covered by any RI. For example:
  - RI Coverage = 20% (0.2)
  - Usage Not Covered by RIs = 80% (0.8)

- **Uptime** is a percentage value that indicates how long a VM has been running over a period of time (age). Age refers to the number of days that a VM has existed since first discovery. For VMs older than 30 days, 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** represents the number of hours per month that Workload Optimization Manager uses to estimate monthly costs.

The listed items above 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.

**Example**

Assume the following data for a pending scale action for an AWS VM:

<table>
<thead>
<tr>
<th>VIRTUAL MACHINE DETAILS</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAME</td>
</tr>
<tr>
<td>PT_Consistent_S...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VCPU PERCENTILE</th>
<th>VMEM PERCENTILE</th>
<th>NET THROUGHPUT</th>
<th>IO THROUGHPUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 %</td>
<td>0.6 %</td>
<td>94 %</td>
<td>47 %</td>
</tr>
<tr>
<td>800 MHz</td>
<td>1.4 GHz</td>
<td>1 GiB</td>
<td>2 GiB</td>
</tr>
<tr>
<td>468.8 MB/s</td>
<td>468.8 MB/s</td>
<td>260.6 MB/s</td>
<td>260.6 MB/s</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ON-DEMAND RATE</th>
<th>RI COVERAGE</th>
<th>UPTIME</th>
<th>EST. ON-DEMAND MONTHLY COST</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.012/h</td>
<td>$0.021/h</td>
<td>95.3%</td>
<td>$4.1/mo $15/mo</td>
</tr>
</tbody>
</table>

To estimate the monthly on-demand cost for this VM, you would multiply the on-demand rate ($0.012/hour) by the usage not covered by RIs (50%), the uptime (95.3%), and the factor of 730 hours to get an estimated cost of $4.1 per month.
Workload Optimization Manager 3.0.7 User Guide

Current Values

<table>
<thead>
<tr>
<th></th>
<th>Current Values</th>
<th>Values After Action Execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-demand Rate</td>
<td>$0.012/hr</td>
<td>$0.021/hr</td>
</tr>
<tr>
<td>RI Coverage</td>
<td>50% (0.5)</td>
<td>0% (0.0)</td>
</tr>
<tr>
<td>Usage Not Covered by RIs (calculated based on RI coverage)</td>
<td>50% (0.5)</td>
<td>100% (1.0)</td>
</tr>
<tr>
<td>Uptime</td>
<td>95.3% (.953)</td>
<td></td>
</tr>
</tbody>
</table>

Workload Optimization Manager calculates the following:

- **Current Estimated On-demand Monthly Cost:**
  
  \[ \text{Current Cost} = 0.012 \times 0.5 \times 0.953 \times 730 = 4.1 \]

- **Estimated On-demand Monthly Cost after executing the action:**
  
  \[ \text{Estimated Cost} = 0.021 \times 1.0 \times 0.953 \times 730 = 15 \]

**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.
Azure VMs with License Costs

Cost Calculation

For 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.

1. On-demand Compute Rate Calculation

The calculation for On-demand Compute Rate can be expressed as follows:

\[
\text{On-demand Rate} - (\text{Reserved License Cost} + \text{On-demand License Cost}) = \text{On-demand Compute Rate}
\]

Where:

- **On-demand Rate** is the hourly cost for a VM’s instance type without RI coverage. This does not include license costs, storage, or IP. You can obtain on-demand rates via Azure Pricing Calculator.
- **Reserved License Cost** and **On-demand License Cost** are the hourly costs for the VM’s licenses. You can obtain license costs via Azure Pricing Calculator or the Workload Optimization Manager user interface.
From the user interface, set the scope to the Azure VM and then see the Workload Cost Breakdown chart. In the chart, set the time frame to Last 2 Hours, and then:

- Hover over the second to the last bar in the chart to obtain the current On-demand License Cost and Reserved License Cost.

- Hover over the last bar (after the vertical line) in the chart to obtain the On-demand License Cost and Reserved License Cost after you execute actions.

The On-demand Compute Rate and License Cost (On-demand and Reserved) are then used to calculate Estimated On-demand Monthly Costs.

2. Estimated On-demand Monthly Cost Calculation

The calculation can be expressed as follows:

\[(\text{On-demand Compute Rate} \times \text{Usage Not Covered by RIs}) + \text{License Cost} \times \text{Uptime} \times 730 = \text{Estimated On-demand Monthly Cost}\]
Where:

- **Usage Not Covered by RIs** is the percentage of hourly VM usage not covered by any RI. For example:
  - RI Coverage = 20% (0.2)
  - Usage Not Covered by RIs = 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 over a period of time (age). Age refers to the number of days that a VM has existed since first discovery. For VMs older than 30 days, 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** represents the number of hours per month that Workload Optimization Manager uses to estimate monthly costs.

The listed items above 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.

**Example**

Assume the following data for a pending scale action for an Azure VM with license costs:

<table>
<thead>
<tr>
<th>VIRTUAL MACHINE DETAILS</th>
</tr>
</thead>
<tbody>
<tr>
<td>NAME</td>
</tr>
<tr>
<td>ptaub807disable...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VCPU PERCENTILE</th>
<th>VMEM PERCENTILE</th>
<th>VM OPS PERCENTILE</th>
<th>VM STORAGE THROUGHPUT PERCENTILE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 % → 2.9 % ↑</td>
<td>9 % → 18 % ↑</td>
<td>0 % → 0 %</td>
<td>0 % → 0 %</td>
</tr>
<tr>
<td>33.2 GHz → 22.7 GHz</td>
<td>32 GiB → 16 GiB</td>
<td>6,400 IOPS → 3,200 IOPS</td>
<td>96 MB/s → 48 MB/s</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ON-DEMAND RATE</th>
<th>RI COVERAGE</th>
<th>UPTIME</th>
<th>EST. ON-DEMAND MONTHLY COST</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.45/h → $0.218/h</td>
<td>100% → 0%</td>
<td>96.1%</td>
<td>$129/mo → $153/mo</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Current Values</th>
<th>Values After Action Execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>On-demand Rate</td>
<td>$0.45/hr</td>
</tr>
</tbody>
</table>
1. Workload Optimization Manager first calculates the following:
   - **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 can now calculate Estimated On-demand Monthly Cost based on:
   - **On-demand Compute Rate**

<table>
<thead>
<tr>
<th>On-demand Compute Rate</th>
<th>Current Values</th>
<th>Values After Action Execution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$0.266/hr</td>
<td>$0.13/hr</td>
</tr>
</tbody>
</table>

- **Usage Not Covered by RIs and Uptime**

<table>
<thead>
<tr>
<th>VIRTUAL MACHINE DETAILS</th>
<th>Current Values</th>
<th>Values After Action Execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name: ptaub807disable...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID: 25b90ce-0e08-4...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age: 30+ days</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location: azure-EU2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VCPU PERCENTILE: 2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VMEM PERCENTILE: 9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VM IOPS PERCENTILE: 18%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VM STORAGE THROUGHPUT PERCENTILE: 0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33.2 GHz -&gt; 22.7 GHz</td>
<td>32 GiB -&gt; 16 GiB</td>
<td>6,400 IOPS -&gt; 3,200 IOPS</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Current Values</th>
<th>Values After Action Execution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.45/h</td>
<td>$0.218/h</td>
</tr>
<tr>
<td>100%</td>
<td>9%</td>
</tr>
<tr>
<td>96.1%</td>
<td></td>
</tr>
<tr>
<td>$129/mo</td>
<td>$153/mo</td>
</tr>
</tbody>
</table>

Workload Optimization Manager calculates the following:

- **RI Coverage**
  - Current: 100% (1.0)
  - Values after Action Execution: 0% (0.0)

- **Usage Not Covered by RIs (calculated based on RI coverage)**
  - Current: 0% (0.0)
  - Values after Action Execution: 100% (1.0)

- **Uptime**
  - 96.1% (0.961)

**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.

Resizing Cloud Workloads

To resize a workload (for example, a VM or an RDS instance) on the cloud, Workload Optimization Manager chooses the cloud tier that best matches the workload requirements. This can be to reduce cost by choosing a smaller tier, or it can be to assure performance by choosing a larger tier. To accomplish the resize, Workload Optimization Manager actually moves the workload to the new tier. This can include moving to a new availability zone.

Note that resize decisions also take into account the discount you can realize by using RI purchases. Workload Optimization Manager can recommend to purchase more RI resources. When considering workload resize actions, Workload Optimization Manager can recommend resizing to a larger RI tier because the overall cost will be less.

As it considers a resize, Workload Optimization Manager also considers the storage and network requirements. Even if the compute resources are underutilized on a workload, if the available tiers cannot support the workload’s storage or network requirements then Workload Optimization Manager will not recommend the change.

NOTE:
In AWS environments, under certain circumstances VM resizing can fail. If the restart of the VM initially fails, Workload Optimization Manager waits 30 seconds and tries to restart again. Workload Optimization Manager will try to restart up to four times. If the restart still fails, Workload Optimization Manager assumes the VM cannot start up on the new tier, and it restarts the VM on the old tier.
Scaling on the Public Cloud

On the cloud, scaling actions change the VM to a different instance type. These can include:

- Changing a VM to an instance type with different capacity
- Changing on-demand to RI

For these actions, the action list shows the current cost for the source workload, and also the projected cost given the change. To show the current cost, Workload Optimization Manager uses the actual costs for that workload. However, to show the projected cost it uses an estimate based on average utilization for the VM, for the costs of the given tier.

Note that scaling to an RI can result in running the VM on a larger instance when the cost is lower. This might occur even though the VM does not need that capacity and there are other smaller tiers available.

In Azure environments, there are circumstances where a VM resize can be especially disruptive. In a given region, the infrastructure can be made up of different clusters that have different sets of underlying hardware. Further, some tiers that are available in the given region are only available on different clusters. If Workload Optimization Manager recommends resizing from a tier on one cluster, to a tier on another cluster, then the resize action can take longer to complete than usual.

In both Azure and AWS environments, Workload Optimization Manager conforms to specific instance requirements as it generates resize actions. For more information, see:

- **Azure Instance Requirements** (on page 160)
- **AWS Instance Requirements** (on page 159)

Reserved Instances (RIs)

Workload Optimization Manager analysis takes advantage of AWS and Azure Reserved Instances (RIs) to calculate optimal workload placement and to arrive at the best possible costs for your deployments on the cloud.

**NOTE:**

For AWS environments, 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.

**RI Charts**

The Cloud View in the Homepage includes the following charts that show RI data:

- **Pending Actions** (on page 320)
  If Workload Optimization Manager has found actions you can take to improve performance or to reduce cost, then you can see an overview of them in the Pending Actions chart. To see a listing of the specific actions, click **Show All** at the bottom of the chart. For more about actions, see **Workload Optimization Manager Actions** (on page 64).

- **RI Utilization** (on page 348)
  This chart shows how well you have utilized the Reserved Instance inventory. The chart compares the capacity for all Reserved Instances versus the RI consumption by virtual machines.

- **RI Coverage** (on page 342)
  This chart compares the capacity of your current VM workload to the capacity of workload that is covered by RIs. If you have a high percentage of on-demand workload, then you should be able to reduce your monthly costs by increasing RI coverage. To increase coverage, you resize workloads to instance types that have existing RI capacity. If you need more RI capacity, then Workload Optimization Manager will recommend the RIs that you should buy.

- **RI Inventory** (on page 344)
  This chart lists the RI instance types that are active in your inventory. To see more information, click **Show All** at the bottom of the chart.

- **Recommended RI Purchases** (on page 341)
  This chart shows the projected inventory of pending RI purchases as generated by Workload Optimization Manager. To see more information, click **Show All** at the bottom of the chart.
**RI Purchases**

Workload Optimization Manager can recommend that you purchase RI capacity to reduce costs for your current workload. The analysis looks at workload history for template families to identify RI candidates. This considers the count of workloads in a family, their hours of active-state condition, and RI costs. If a workload shows stable utilization over time, then Workload Optimization Manager identifies it as an RI candidate, and it recommends purchasing RI capacity for that workload. To ensure enough historical data for the analysis, Workload Optimization Manager generates RI Buy actions on a two-week cycle. It also generates a new set of RI Buy actions if you change the RI inventory, or if you restart Workload Optimization Manager.

Be aware of the following:

- Different types of RIs have different costs, so the choice between using on-demand or RI pricing can vary depending on the RI Pricing configuration in your Budgets and Costs settings. For more information, see RI Purchase Profile (on page 375).

- Workload Optimization Manager can only estimate the cost that would result if you execute pending RI Buy actions. This is because the full data is only available after you actually purchase the RIs. Estimates reflect costs you would see after scaling workloads to the newly purchased RI capacity. For scaling to already-purchased RIs, the chart reflects the actual costs.

- As Workload Optimization Manager calculates actions to purchase RI capacity, it assumes that any other pending actions for the workload will also be executed. For example, assume a workload running on an r4.xlarge template. If Workload Optimization Manager recommends changing that instance type to an m5.medium, it can recommend that you purchase an m5 RI to cover the workload and reduce costs. This purchase could be on a region that currently doesn’t have any m5 workloads — The purchase recommendation assumes you will move the workload to that other region.

- Workload Optimization Manager uses a weighted history of workload activity and that suspended VMs are also considered. The longer ago that the workload was suspended, the less weight it has in the RI Buy calculation.

- For AWS RIs:
  - For environments that use the Instance Size Flexible rules, Workload Optimization Manager can recommend that you buy multiple RIs of smaller instance types to cover the resource requirements of larger instance types. For example, rather than buying one t2.small RI, Workload Optimization Manager can recommend that you buy four t2.nano RIs to offer an equivalent discount.
  - For environments that consolidate billing into Billing Families, Workload Optimization Manager recommends purchases for RIs that are within the given billing family. For more information, see AWS Billing Families (on page 44).

**Support for Government Workloads**

AWS GovCloud (US) and Azure Government provide dedicated regions for US government customers and their partners to architect secure cloud solutions and meet regulatory and compliance requirements.

Workload Optimization Manager discovers workloads in these regions when you add the required accounts as targets. For details on the required accounts, see "AWS GovCloud Targets" in the Target Configuration Guide and "Azure Government Targets" in the Target Configuration Guide.

Discovered workloads include:

- AWS VMs (including auto-scaling groups), volumes, database servers, and spot instances
- Azure VMs (including availability/scale sets), volumes, and SQL databases

Workload Optimization Manager recommends actions on VMs, volumes, and SQL databases to address performance issues and optimize costs.

**NOTE:**

Workload Optimization Manager currently does not support Azure Government integration with Application Insights. You can add accounts for Azure Government and Application Insights as targets, but Application Insights will only return performance data for non-government workloads.
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. This chart shows the following:
  - Azure Government subscriptions discovered via the service principal and EA accounts that you have added as targets
  - AWS GovCloud master and member accounts that you have added as targets. Accounts with a star symbol are master accounts.

  For an AWS GovCloud account, the projected billed 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 projected 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.
• **RI and Savings Plans Inventory** chart

The government accounts that you added as targets enable Workload Optimization Manager to gain full insight into the AWS/Azure RIs and AWS Savings Plans that you have purchased for your government workloads. Even as you selectively add secondary targets, Workload Optimization Manager remains aware of all RIs and Savings Plans, 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.
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.

**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.

**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 26)](#).

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.

### Configuring Targets

A target is a service that performs management in your virtual environment. Workload Optimization Manager uses targets to monitor workload and to execute actions in your environment. When you configure a target, you specify the address of the service, and the credentials to connect as a client to it.

For each target, Workload Optimization Manager communicates with the service via the management protocol that it exposes — The REST API, SMI-S, XML, or some other management transport. Workload Optimization Manager uses this communication to discover the managed entities, monitor resource utilization, and execute actions.
To configure a target, you will choose the target type, specify the target’s address, and then provide credentials to access the target.

After you configure a target and add it to your installation, Workload Optimization Manager validates the connection, and then discovers the entities that target manages.

NOTE:
Workload Optimization Manager regularly checks that your targets are valid. If it discovers that a target is invalid it then posts that status to the user interface. Under some circumstances, the target can become valid again, but the status does not update. If you see an Invalid message for a given target, try to manually validate the target again (click **VALIDATE**).

**Configuring a Target**

1. Navigate to the Settings Page.

   ![Settings Page](image)

   Click to navigate to the Settings Page. From there, you can perform a variety of Workload Optimization Manager configuration tasks.

2. Choose Target Configuration.

   ![Target Configuration](image)

   Click to navigate to the Target Configuration Page.

3. Review the list of targets.

   ![Target Configurations](image)

   This page lists all the targets that you currently have configured for Workload Optimization Manager. You can inspect these targets, you can edit them (change address and credentials), and you can add a new target to Workload Optimization Manager.

4. Filter the list of targets.
To work with a long list of targets, you can filter by the target type. You can also type a string in the Search field to filter the list, and you can sort the list by target status or target name.

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.
- Validate – Direct Workload Optimization Manager to validate its connection with the target. For example, if you create a new user account on the target, you can edit the target connection to use that account, and then revalidate.
- Delete – When you delete a target, Workload Optimization Manager removes all the associated entities from its model of the inventory.

6. Expand an entry to see details, or click the entry to edit the target's configuration.

For example, if you entered the wrong username or password, you can change those credentials and validate the target again.
7. Create a new target and add it to Workload Optimization Manager.

- Click the **NEW TARGET** button.
- Choose the target category (Application Server, Hypervisor, Storage, Load Balancer, Network).
- Select the target type (Hyper-V, RHEV, XenServer, vCenter, IBM PowerVM).
First, select the type of target to add. Then for the type you choose, select the specific target technology. For example, select Hypervisor/vCenter to add a VMware vCenter Server target. Then provide the address and credentials for that target.

For more details, including a list of supported targets and configuration requirements, see the Workload Optimization Manager Target Configuration Guide.

AWS Billing Families

As you configure AWS targets, Workload Optimization Manager discovers AWS accounts that are consolidated into billing families. A billing family has one master 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.

In the Targets user interface, master accounts appear in bold, with a star next to them. You can expand the account entry to see the related member accounts. If you expand the entry for a member account, then the related accounts includes the family master, indicated by a star.

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 master and member accounts in the Targets 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 master 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 have enabled Billing Family Recognition, you should 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 Cloud Discounts (on page 377) and Discount Override: AWS (on page 384).

- **You might see master 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 master account, but it has no member accounts.
Getting Started

Azure Enterprise Agreements

You can configure Workload Optimization Manager to manage Azure subscriptions within the context of an Enterprise Agreement (EA). An EA defines specific pricing, including the pricing for Reserved Instances (RIs). When you configure an EA target, set the EA key to your Azure targets, Workload Optimization Manager uses that richer pricing information to calculate workload placement and RI coverage for your Azure environment.

To enable Workload Optimization Manager management of Azure EA environments, you must configure:

- One Microsoft Enterprise Agreement target
- At least one Service Principal target that can discover the underlying Azure subscriptions

For information about Azure targets, see "Microsoft Azure" in the Target Configuration Guide.

In the Targets View, you can identify the targets related to Azure EA as follows:

- EA Targets
  - The target that discovers the EA to track pricing and RI information. You can have one EA target per Workload Optimization Manager deployment.
- Azure Subscription Targets
The targets that manage the workloads in your Azure environment. These are discovered by Service Principal targets. Note that not all subscription targets necessarily participate in the EA. Expand these entries to see the related Service Principal target. For members of the EA, you can see the related EA target as well.

Subscriptions that do not participate in the EA appear as Standalone targets.

**NOTE:**
In rare circumstances, you can have a subscription that is not in use – The subscription has no workloads associated with it. In this case, Workload Optimization Manager identifies the subscription as Standalone. This is because the target cannot discover any cost or usage information that would relate the subscription to its EA.

- **Service Principal Targets**
  The Azure target that you configure to discover Azure subscription targets. Expand the entry to see the discovered targets. If you have configured an EA target, the entry lists that as well, along with the EA enrollment number.

---

**Reserved Instances and Azure EA**

For Azure environments, Workload Optimization Manager can only discover and use RIs if you have configured a Microsoft Enterprise Account target, and if one or more subscriptions participate in that EA.

To discover and manage RIs in Azure environments, Workload Optimization Manager uses both the EA target and the associated subscription targets. On its own, a subscription target exposes costs for pay-as-you-go pricing. The EA target discovers pricing for the available RI instance types. Workload Optimization Manager combines this information to track:

- RI utilization
- RI coverage
- Virtual machine costs (accounting for RIs)

**NOTE:**
This release of Workload Optimization Manager does not support RI discovery and management for Classic VMs and Classic Cloud Services. Also, it does not support RI discovery and management for Suppressed Core virtual machines.

---

**Cost Calculations for Azure Environments**

To understand the reported costs in your Azure environment, consider these points:

- For targets that participate in the EA, Workload Optimization Manager uses the terms of the given EA, and bases costs on the Offer ID that is effective for the given subscription.
- For VMs in Azure, RI pricing does not include the cost of the OS license. However pricing for on-demand VMs does include the license cost.

**NOTE:**
For Microsoft Azure EA environments, the projected cost for RI Purchase actions might not match associated costs you find in the Microsoft Pricing Calculator.

Workload Optimization Manager actions can recommend RI purchases. For these recommendations, the action assumes a free Linux OS, so the cost estimate does not include the OS cost. However, The Microsoft Pricing Calculator does include costs for OS licenses. As a result, when you compare the Workload Optimization Manager cost estimates to the values in the Pricing Calculator, it's likely that the two estimates will not match. This difference also affects the Break Even Point that appears in the Recommended RI Purchases chart. Because the recommended purchases do not include Azure costs for OS licenses, the listed Break Even Point can be optimistic.

- For workloads you migrated from on-prem to the Azure cloud, Workload Optimization Manager recognizes Azure Hybrid Benefit (AHUB) savings for RIs and on-demand workloads. The costs you see in Workload Optimization Manager charts include this benefit. However, remember that recommended actions do not include any license cost, so the actions will not reflect any proposed AHUB savings (see above).
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.
Getting Started

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 117)
- 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.

![List of Virtual Machines](image)

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 50)*
- Navigating With the Supply Chain *(on page 62)*
- Viewing Cluster Headroom *(on page 63)*

**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.
   
   ![Search Icon]
   
   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 type of entities that you want to search through. Find the list of entity types on the left. Select **All** to search the complete environment. Or 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.

3. Use **Search** to filter the listing.
   
   For example, if you’re showing **All** and you search for ”Development”, then you will see all clusters, groups, and entities with ”Development” in their names.
4. Expand an entry to see details.
   For example, expand a group or an entity to see 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.
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.

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 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.

**For different types of groups, click to set a single group as your scope**

**Choose an entity type, and set the scope to one or more of those entities**

**Click to set the scope you have selected**
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 50)
• Create new charts: See Creating and Editing Chart Widgets (on page 316)

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 62)
• Set Scope
  Use Search to set the scope of the viewing session - see Scoping the Workload Optimization Manager Session (on page 50)

Chart Time Frame
You can set a time frame from recent hours to the past year, and set that to the charts in the view. Use the Time Slider to set specific start and end times within that range. 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 50)
- Create new charts: See Creating and Editing Chart Widgets (on page 316)

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 62)
- Set Scope
  Use Search to set the scope of the viewing session - see Scoping the Workload Optimization Manager Session (on page 50)

Chart Time Frame

Choose the time scale you want
You can set a time frame from recent hours to the past year, and set that to the charts in the view. Use the Time Slider to set specific start and end times within that range. 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 has been enabled or disabled. In addition, you can create new policies and apply them to that scope.

To edit a policy, click the policy name. You can then change the policy settings, or enable/disable the policy.

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 has been applied to this scope.

When you create a new policy, it automatically includes the current scope. You can add other groups to the policy scope if you like. Note that 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.

For more information, see Automation Policies (on page 92).
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 for this VM.
### Host Constraints For "Oracle11g-Win-172.32"

When you add constraints, you limit the placement decisions Turbonomic can make for your VM. Remove unnecessary constraints so Turbonomic can discover more placement options.

<table>
<thead>
<tr>
<th>CONSTRAINT TYPE</th>
<th>SCOPE NAME</th>
<th>SOURCE</th>
<th>POTENTIAL HOSTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster boundaries</td>
<td>ACM/ACM Cluster</td>
<td>vCenter</td>
<td>4 Hosts</td>
</tr>
<tr>
<td>Datacenter boundaries</td>
<td>ACM</td>
<td>vCenter</td>
<td>4 Hosts</td>
</tr>
<tr>
<td>Datastore Commodity</td>
<td>Q54/ACM</td>
<td>vCenter</td>
<td>4 Hosts</td>
</tr>
<tr>
<td>Network boundaries</td>
<td>NetworkCommodity/Oracle11g</td>
<td>Turbonomic</td>
<td>12 Hosts</td>
</tr>
<tr>
<td></td>
<td>Win-172.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Segmentation Commodity</td>
<td>My Placement Policy</td>
<td>Turbonomic</td>
<td>16 Hosts</td>
</tr>
<tr>
<td>License/Access Commodity</td>
<td>Linux</td>
<td>Turbonomic</td>
<td>70 Hosts</td>
</tr>
</tbody>
</table>

Current count of potential providers: 4

Click to enable constraint simulations

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 PROVIDERS**
  For the given constraint, how many providers that constraint allows. To see a list of the potential providers, click the POTENTIAL PROVIDERS value.

To dig deeper into how these constraints affect your entity, click **FIND MORE PLACEMENT OPTIONS**. 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. As you do, the POTENTIAL PROVIDERS label updates to show how many providers are available to your entity. To see the resulting list of providers, click the POTENTIAL PROVIDERS label.

By turning off the 4-Host constraints, you have 12 potential hosts for this VM. Click this label to see the resulting list of providers.

Use the toggles to turn off various constraints

Click to see the potential hosts
List of Entities

The list of entities is a quick way to drill down to details about your environment, so 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 62).
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 cluster headroom:

1. Navigate to the Search page.
2. Choose the 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 use the Economic Scheduling Engine to 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 301)
Workload Optimization Manager Actions

After you deploy your targets, Workload Optimization Manager starts to perform market analysis as part of its Application Resource Management process. This holistic analysis identifies problems in your environment 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 the Pending Actions charts for you to investigate.

Workload Optimization Manager can generate the following actions:

<table>
<thead>
<tr>
<th>Action</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provision</td>
<td>Introduce new resource providers to update the environment’s capacity. For example:</td>
</tr>
<tr>
<td></td>
<td>• Provisioning a host adds more compute capacity that is available to VMs.</td>
</tr>
<tr>
<td></td>
<td>• Provisioning a VM adds capacity to run applications.</td>
</tr>
<tr>
<td>Start</td>
<td>Start a suspended entity to add capacity to the environment.</td>
</tr>
<tr>
<td>Resize</td>
<td>Re-allocate resource capacity on an entity. For example, reduce vCPUs or vMem on a VM, or add volumes to a disk array.</td>
</tr>
<tr>
<td>Buy RI</td>
<td>For workloads that are good RI candidates, purchase RI capacity to move your environment toward the RI coverage that you desire.</td>
</tr>
<tr>
<td>Increase RI coverage</td>
<td>Increase RI coverage to reduce costs.</td>
</tr>
<tr>
<td>Reconfigure</td>
<td>• Add necessary network access or reconfigure storage. For example, if a VM is configured to access a network that is not available on the host, the VM must reconfigure to use an available network.</td>
</tr>
<tr>
<td></td>
<td>• Reconfigure container pods.</td>
</tr>
<tr>
<td>Move</td>
<td>Change a consumer to use a different provider, such as moving a VM to a different host. Moving a VM to a different storage means relocating any file-based component that belongs to a virtual machine.</td>
</tr>
<tr>
<td>Suspend</td>
<td>Stop and set resources aside without removing them from the environment.</td>
</tr>
<tr>
<td>Action</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>For example, you might suspend an underutilized host to save it for some time when you really need it, or suspend a virtual machine to save money.</td>
<td></td>
</tr>
<tr>
<td>Delete</td>
<td>Remove storage (for example, datastores on disk arrays or unattached volumes).</td>
</tr>
</tbody>
</table>

### 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:

#### Applications

<table>
<thead>
<tr>
<th>Entity Type</th>
<th>Supported Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business Application</td>
<td>Workload Optimization Manager does not recommend actions for a Business Application, but it does recommend actions for the underlying Application Components and nodes. 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.</td>
</tr>
<tr>
<td>Business Transaction</td>
<td>Workload Optimization Manager does not recommend actions for a Business Transaction, but it does recommend actions for the underlying Application Components and nodes. 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.</td>
</tr>
<tr>
<td>Service</td>
<td>Workload Optimization Manager does not recommend actions for a Service, but it does recommend actions for the underlying Application Components and nodes. The Pending Actions chart for a Service lists these actions, thus providing visibility into the risks that have a direct impact on the Service’s performance.</td>
</tr>
<tr>
<td>Application Component</td>
<td>The default Scaling Policy is <strong>Resize</strong>. With this policy, Workload Optimization Manager will only recommend resize actions. If you change the Scaling Policy to <strong>Horizontal Scale</strong>, it will only recommend provision and suspend actions. For details, see Application Component Actions <em>(on page 126)</em>.</td>
</tr>
</tbody>
</table>

#### Container Platform

<table>
<thead>
<tr>
<th>Entity Type</th>
<th>Supported Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Container, Container Spec, Workload Controller</td>
<td>Workload Optimization Manager supports resize actions for containers. By default, containers resize consistently, which allows all replicas of the same container for the same workload type to resize any commodity consistently. Workload Optimization Manager uses namespace or organization/space quotas as constraints when making resize decisions. Workload Optimization Manager shows container resize actions via Container Specs, and executes them via Workload Controllers. It does not recommend actions for these two entity types. For details, see Container Actions <em>(on page 133)</em>.</td>
</tr>
<tr>
<td>Container Pod</td>
<td>Workload Optimization Manager can recommend the following actions for a container pod:</td>
</tr>
<tr>
<td></td>
<td>• Move</td>
</tr>
<tr>
<td></td>
<td>• Suspend</td>
</tr>
</tbody>
</table>
## Getting Started

<table>
<thead>
<tr>
<th>Entity Type</th>
<th>Supported Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Provision</td>
</tr>
<tr>
<td></td>
<td>• Reconfigure</td>
</tr>
<tr>
<td></td>
<td>For details, see Container Pod Actions (on page 145).</td>
</tr>
<tr>
<td>Namespace</td>
<td>Workload Optimization Manager does not recommend actions for a namespace. Instead, it recommends actions for containers and container pods.</td>
</tr>
<tr>
<td>Container Cluster</td>
<td>Workload Optimization Manager does not recommend actions for a Container 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 Cluster and view the Pending Actions chart.</td>
</tr>
</tbody>
</table>

### Cloud Infrastructure

<table>
<thead>
<tr>
<th>Entity Type</th>
<th>Supported Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual Machine</td>
<td>Workload Optimization Manager can recommend the following actions for a cloud VM:</td>
</tr>
<tr>
<td></td>
<td>• Scale</td>
</tr>
<tr>
<td></td>
<td>• Increase RI Coverage</td>
</tr>
<tr>
<td></td>
<td>• Buy RI</td>
</tr>
<tr>
<td></td>
<td>For details, see Cloud VM Actions (on page 157).</td>
</tr>
<tr>
<td></td>
<td>For a VM used as a Kubernetes node, Workload Optimization Manager can recommend the following actions:</td>
</tr>
<tr>
<td></td>
<td>• Provision</td>
</tr>
<tr>
<td></td>
<td>• Suspend</td>
</tr>
<tr>
<td></td>
<td>For details, see Actions for Kubernetes Nodes (VMs) (on page 158).</td>
</tr>
<tr>
<td>Database</td>
<td>Workload Optimization Manager can recommend scale actions for a cloud database.</td>
</tr>
<tr>
<td></td>
<td>For details, see Cloud Database Actions (on page 177).</td>
</tr>
<tr>
<td>Database Server</td>
<td>Actions for cloud Database Servers will be introduced in a future product release.</td>
</tr>
<tr>
<td>Volume</td>
<td>Workload Optimization Manager can recommend the following actions for a cloud volume:</td>
</tr>
<tr>
<td></td>
<td>• Delete</td>
</tr>
<tr>
<td></td>
<td>• Scale</td>
</tr>
<tr>
<td></td>
<td>For details, see Cloud Volume Actions (on page 168).</td>
</tr>
<tr>
<td>Zone</td>
<td>Workload Optimization Manager does not recommend actions for a cloud zone.</td>
</tr>
<tr>
<td>Region</td>
<td>Workload Optimization Manager does not recommend actions for a cloud region.</td>
</tr>
</tbody>
</table>

### On-prem Infrastructure

<table>
<thead>
<tr>
<th>Entity Type</th>
<th>Supported Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual Machine</td>
<td>Workload Optimization Manager can recommend the following actions for an on-prem VM:</td>
</tr>
<tr>
<td></td>
<td>• Resize</td>
</tr>
<tr>
<td></td>
<td>• Move</td>
</tr>
<tr>
<td></td>
<td>• Move Storage</td>
</tr>
<tr>
<td></td>
<td>• Reconfigure</td>
</tr>
<tr>
<td></td>
<td>For details, see On-prem VM Actions (on page 189).</td>
</tr>
</tbody>
</table>
## Entity Type

<table>
<thead>
<tr>
<th>Entity Type</th>
<th>Supported Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>For a VM used as a Kubernetes node, Workload Optimization Manager can recommend the following actions:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ Provision</td>
</tr>
<tr>
<td></td>
<td>▪ Suspend</td>
</tr>
<tr>
<td>Database Server</td>
<td>Workload Optimization Manager can recommend resize actions for an on-prem Database Server.</td>
</tr>
<tr>
<td></td>
<td>For details, see [Database Server Actions (on page 198)](on page 198).</td>
</tr>
<tr>
<td>Volume</td>
<td>Workload Optimization Manager can recommend move actions for an on-prem volume. To evaluate and execute these actions, set the scope to the VM to which a volume is attached.</td>
</tr>
<tr>
<td>Virtual Datacenter</td>
<td>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.</td>
</tr>
<tr>
<td>Business User</td>
<td>Workload Optimization Manager can recommend move actions for a Business User.</td>
</tr>
<tr>
<td></td>
<td>For details, see [Business User Actions (on page 209)](on page 209).</td>
</tr>
<tr>
<td>Desktop Pool</td>
<td>Workload Optimization Manager does not recommend actions for a desktop pool. It does recommend actions for the business users running active sessions in the pool.</td>
</tr>
<tr>
<td>View Pod</td>
<td>Workload Optimization Manager does not recommend actions for a view pod. Instead, it recommends actions for the business users that are running active sessions.</td>
</tr>
<tr>
<td>Host</td>
<td>Workload Optimization Manager can recommend the following actions for a host:</td>
</tr>
<tr>
<td></td>
<td>▪ Start</td>
</tr>
<tr>
<td></td>
<td>▪ Provision</td>
</tr>
<tr>
<td></td>
<td>▪ Suspend</td>
</tr>
<tr>
<td></td>
<td>▪ Reconfigure</td>
</tr>
<tr>
<td></td>
<td>For details, see [Host Actions (on page 217)](on page 217).</td>
</tr>
<tr>
<td>Chassis</td>
<td>Workload Optimization Manager does not recommend actions for a chassis.</td>
</tr>
<tr>
<td>Datacenter</td>
<td>Workload Optimization Manager does not recommend actions for a datacenter. Instead, it recommends actions for the devices running in the datacenter.</td>
</tr>
<tr>
<td>Storage</td>
<td>Workload Optimization Manager can recommend the following actions for storage:</td>
</tr>
<tr>
<td></td>
<td>▪ Move</td>
</tr>
<tr>
<td></td>
<td>▪ Provision</td>
</tr>
<tr>
<td></td>
<td>▪ Resize</td>
</tr>
<tr>
<td></td>
<td>▪ Start</td>
</tr>
<tr>
<td></td>
<td>▪ Suspend</td>
</tr>
<tr>
<td></td>
<td>▪ Delete Datastore or Volume</td>
</tr>
<tr>
<td></td>
<td>For details, see [Storage Actions (on page 223)](on page 223).</td>
</tr>
<tr>
<td>Logical Pool</td>
<td>Workload Optimization Manager can recommend the following actions for a logical pool:</td>
</tr>
<tr>
<td></td>
<td>▪ Resize</td>
</tr>
<tr>
<td></td>
<td>▪ Provision</td>
</tr>
<tr>
<td></td>
<td>▪ Move</td>
</tr>
</tbody>
</table>
### Entity Type

<table>
<thead>
<tr>
<th>Entity Type</th>
<th>Supported Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disk Array</td>
<td>Workload Optimization Manager can recommend the following actions for a disk array:</td>
</tr>
<tr>
<td></td>
<td>▪ Provision</td>
</tr>
<tr>
<td></td>
<td>▪ Start</td>
</tr>
<tr>
<td></td>
<td>▪ Suspend</td>
</tr>
<tr>
<td></td>
<td>▪ Move (for NetApp Cluster-Mode, only)</td>
</tr>
<tr>
<td></td>
<td>▪ Move VM</td>
</tr>
<tr>
<td></td>
<td>▪ Move Datastore</td>
</tr>
<tr>
<td>Storage Controller</td>
<td>Workload Optimization Manager can recommend provision actions for a Storage Controller.</td>
</tr>
<tr>
<td>IO Module</td>
<td>Workload Optimization Manager does not recommend actions for an IO Module.</td>
</tr>
<tr>
<td>Switch</td>
<td>Workload Optimization Manager can recommend resizing port for a switch to increase size.</td>
</tr>
</tbody>
</table>

### Action Types

Workload Optimization Manager performs the following general types of actions:

- Placement — Place a consumer on a specific provider (place a VM on a Host)
- Scaling — Resize allocation of resources, based on profitability
  - Resize up, shown as a required investment
  - Resize down, shown as savings
- RI Optimization — Purchase RIs for specific workloads or move to RI tiers that are more appropriate for your applications' requirements
- Configuration — Correct a misconfiguration
- Start/Buy — Start a new instance to add capacity to the environment, shown as a required investment. For workloads that are good RI candidates, purchase RI capacity to move your environment toward the RI Coverage that you desire.
- Stop — Suspend an instance to increase efficient use of resources, shown as savings
- Delete — Remove storage (for example, datastores on disk arrays or unattached volumes)

### 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.

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 369).

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 193).

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 87)). 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 RIs 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.

**Storage Resize Actions**

Any storage resize action impacts both the storage entities and the entities managed by the given hypervisor. However, not all hypervisors recognize changes to the storage capacity. After executing a storage resize, Workload Optimization Manager indicates that the resize action has succeeded but a hypervisor might not show the corresponding change in storage capacity. If this occurs, then you must refresh the hypervisor target so Workload Optimization Manager can discover the storage changes.

To avoid this situation, you can set the action mode to *Manual or Recommend* for storage resize actions. In that way, you can perform the resizes yourself, and then manually refresh your hypervisors.

**Scaling on the Public Cloud**

On the cloud, scaling actions change the VM to a different instance type. These can include:

- Changing a VM to an instance type with different capacity
- Changing on-demand to RI

For these actions, the action list shows the current cost for the source workload, and also the projected cost given the change. To show the current cost, Workload Optimization Manager uses the actual costs for that workload. However, to show the projected cost it uses an estimate based on average utilization for the VM, for the costs of the given tier.

Note that scaling to an RI can result in running the VM on a larger instance when the cost is lower. This might occur even though the VM does not need that capacity and there are other smaller tiers available.

In Azure environments, there are circumstances where a VM resize can be especially disruptive. In a given region, the infrastructure can be made up of different clusters that have different sets of underlying hardware. Further, some tiers that are available in the given region are only available on different clusters. If Workload Optimization Manager recommends resizing from a tier on one cluster, to a tier on another cluster, then the resize action can take longer to complete than usual.

In both Azure and AWS environments, Workload Optimization Manager conforms to specific instance requirements as it generates resize actions. For more information, see:

- [Azure Instance Requirements](on page 160)
- [AWS Instance Requirements](on page 159)

**RI Optimization**

Workload Optimization Manager can recommend purchasing RIs for specific workloads or moving to RI tiers that are more appropriate for your applications’ requirements. See the following charts in the dashboard for relevant information:

- **RI Inventory** chart: Shows a list of RI workloads that Workload Optimization Manager discovered, listed by tiers
- **RI Utilization** chart: Shows how well you have utilized your RI inventory
- **RI Coverage** chart: Shows the capacity of your current VM workload compared to the capacity of workload that is covered by RIs

RI optimization actions are not executed by Workload Optimization Manager users. They reflect RI reassignment, which the cloud provider will take care of.

**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/Buy
Workload Optimization Manager can recommend that you start a suspended entity to add capacity to the environment. It can also recommend purchasing RI capacity to reduce costs for your current workload. For details, see RI Purchases (on page 35).

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
- VMs

Delete
Delete actions affect storage. For example, Workload Optimization Manager might recommend that you delete wasted files to free up storage space, or delete unused storage in your cloud environment to reduce storage costs.

Wasted Storage in Azure Environments
In Azure environments, Workload Optimization Manager can identify unmanaged storage as unattached volumes, recommend that you remove this unused storage, and then show estimated savings after you remove this storage and no longer pay for it. The savings that Workload Optimization Manager shows are estimates based on the overall cost for that storage, since Azure does not provide specific values for the cost per volume or cost for the amount of storage that is in use for a given volume. If the estimated savings appear unusually high, then you should identify which storage the actions will remove, and review your billing to calculate the costs with more precision.

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.

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 in the Performance category. You can consider these critical conditions, and you should execute the recommended actions as soon as possible.

<table>
<thead>
<tr>
<th>Actions</th>
<th>Risks/Opportunities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provision a new VM, Host, Datastore</td>
<td>&lt;Resource&gt; Congestion</td>
</tr>
<tr>
<td>Increase or decrease the number of VCPUs</td>
<td>High utilization of application resources. High utilization of resources on workload, host, or datastore.</td>
</tr>
<tr>
<td>Provision a new container or container pod</td>
<td></td>
</tr>
<tr>
<td>Resize heap for an Application Component</td>
<td></td>
</tr>
<tr>
<td>Scale the resource capacity on an entity</td>
<td></td>
</tr>
</tbody>
</table>
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 resources, it recommends actions to consolidate your operations. For example, it can recommend that you move certain VMs onto a different host. This can free a physical machine to be shut down.

<table>
<thead>
<tr>
<th>Actions</th>
<th>Risks/Opportunities</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Move VM</td>
<td>• Overprovisioning</td>
</tr>
<tr>
<td>• Start or suspend VM</td>
<td>Excess resource capacity in a provider.</td>
</tr>
<tr>
<td>• Buy RI</td>
<td></td>
</tr>
<tr>
<td>• Scale down resource allocation</td>
<td></td>
</tr>
</tbody>
</table>

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 associated actions in the Prevention category. 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 the QoS for some services may be put at risk.

<table>
<thead>
<tr>
<th>Actions</th>
<th>Risks/Opportunities</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Resize vCPU and vMem</td>
<td>• &lt;Resource&gt; Congestion</td>
</tr>
<tr>
<td>• Move VM or storage</td>
<td>High resource utilization on the named VM, host, or datastore. For example, CPU congestion or memory congestion can occur on a VM or physical machine, or an IOPS bottleneck can occur on a datastore.</td>
</tr>
<tr>
<td>• Start VM or host</td>
<td>• Workload Balancing</td>
</tr>
<tr>
<td></td>
<td>Excess workload on a given physical machine that can be addressed by moving a VM to another host.</td>
</tr>
</tbody>
</table>

Compliance

A virtual environment can include policies that limit workload placement or availability of resources. It’s possible that the environment configuration violates these defined policies. It’s also possible that an entity is mis-configured in some way. For example, a VM might be configured to access a network that is not available in its current host cluster. In such cases, Workload Optimization Manager identifies the violation and recommends actions that bring the entity back into compliance.

<table>
<thead>
<tr>
<th>Actions</th>
<th>Risks/Opportunities</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Move VM</td>
<td>• Misconfiguration</td>
</tr>
<tr>
<td>• Move container</td>
<td>Container configuration is in violation of a policy.</td>
</tr>
<tr>
<td>• Provision VM, Host, Datastore</td>
<td>• Placement Violation</td>
</tr>
<tr>
<td></td>
<td>• Misconfiguration</td>
</tr>
<tr>
<td></td>
<td>The configuration violates discovered requirements. For example, a VM is configured to access a network that is not available from the current cluster.</td>
</tr>
</tbody>
</table>
Action Modes

Action 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 modes in a policy to set that business rule.

Workload Optimization Manager supports the following action modes:

- **Recommend** – Recommend the action so a user can execute it via the given hypervisor or by other means
- **Manual** – Recommend the action, and provide the option to execute that action through the Workload Optimization Manager user interface
- **Automatic** – Execute the action automatically

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 Modes

To set action modes for specific entities, you can edit the Workload Optimization Manager automation policies. This is how you specify the default action modes, or set special action modes for a given group or cluster. For more information, see Automation Policies (on page 92).

Action Mode Overrides

Under some conditions, Workload Optimization Manager changes the action 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 mode from Manual to Recommend, and then describes the reason.

Working With the Generated Actions

When you start using Workload Optimization Manager, all the actions that the product generates appear as pending. You can view them in the Pending Actions charts and then decide whether to execute and/or automate them. You can also disable them.
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.

When you first see the pending actions, you execute many of them to see immediate improvements in performance and 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 QoS 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 the Pending Actions charts. They no longer show 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.

**What You Can Do:**
- View and execute pending actions: See [Pending Actions (on page 75)](#).
- See the different display views for the pending actions charts: See [Pending Actions Charts (on page 320)](#).
- Scope pending actions in the **Home Page**: See [Pending Actions Scope (on page 77)](#).
- See a running history of generated and executed actions: See [Actions Charts (on page 322)](#).
- Review the default policies that drive the actions the product generates.
- Create and run plans to simulate different conditions, and see what actions will keep things healthy under those conditions: See [Plan Management (on page 243)](#).

**Pending Actions**

Workload Optimization Manager treats all the non-automated actions that it generates as pending and shows them in the Pending Actions charts.

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 96)](#) or change the action mode to **Automatic** in the [default policies (on page 93)](#).

Workload Optimization Manager can execute up to five actions at a time, and queues any new incoming actions for later execution.
Default Pending Actions Charts

Each time you log in to the user interface, Workload Optimization Manager immediately shows the Pending Actions charts on the Home Page's HYBRID view. These charts provide a summary of the actions that require your attention, and entry points to the Pending Actions List (on page 79).

NOTE:
You can also add these charts to any of your dashboards (on page 309).

By default, a text chart and a list chart display in the Home Page, with the scope set to Global Environment. You can change the chart type by clicking the icon on the upper-right corner of the chart. For details about the available chart types, see Pending Actions Charts (on page 320).

Pending Actions - Text Chart

The text chart shows the estimated costs or savings associated with the pending actions, and the number of actions for each action type (on page 68).
Pending Actions - List Chart

The list chart shows a partial list of pending actions, ordered by the severity of the associated problems.

Pending Actions Scope

To perform Application Resource Management, Workload Optimization Manager identifies actions you can take to avoid problems before they occur. You can perform these actions manually, direct Workload Optimization Manager to perform the actions on command, or direct Workload Optimization Manager to perform actions automatically as they arise.

There are several ways to scope pending actions in the Home Page.
To view all pending actions, click **Show all Actions** in the Pending Actions chart.

Click one of the following to narrow the scope of pending actions:

- An entity type in the supply chain.
  
  Workload Optimization Manager generates actions based on how entity types use or provide resources, and what each entity type supports. For details on the actions that each entity type supports, see [Actions by Entity Type](on page 65).
  
  Only entity types with risks (critical, major, or minor) have pending actions. Hover on the entity type to see a breakdown of risks.

- An action type in the text chart

- An entity name in the list chart

**NOTE:**

If you are in the **ON-PREM** or **CLOUD** view, the text chart displays by default. Switch to the list chart to see the entity names.

If you clicked **Show all Actions** or an action type, the **Pending Actions List** displays immediately.

If you clicked an entity type or an entity name, an Overview page displays first. In that page, click the **Actions** tab to view the Pending Actions List.
The Pending Actions List includes additional features to narrow the scope further. You can search for specific actions using meaningful keywords or use filters. For details, see  Pending Actions List (on page 79).

Pending Actions List

The Pending Actions List includes all the actions that Workload Optimization Manager currently recommends for the given scope (for details, see Pending Actions Scope (on page 77)).

You can select actions to execute, and you can expand action items to see more details.
A. Actions List

Each row in the actions list shows:

- The specific action that Workload Optimization Manager recommends.
- If applicable, the estimated investment needed to successfully execute the action or the resulting savings after performing the action.
- The action category (on page 71).

By default, actions display by the severity of the associated problems, indicated by the thin colored line before the checkbox. Use the Filter functionality to change the order by other categories.

Select one or several actions to execute and click **Apply Selected**.

If you see an action with:

- A grayed-out checkbox ( )
  The action is recommended-only, which means you have to perform the action outside Workload Optimization Manager. This occurs when the action mode is **Recommend** or if the underlying technology for the entity does not support automation.

- A grayed-out checkbox and a prohibition symbol ( )
  You need to perform some prerequisite steps outside Workload Optimization Manager before you can execute the action. Hover on the checkbox to see the prerequisite steps.

B. Action Details

Click the arrow icon to expand the entry and view action details.
Action details include:
- A description of the recommended action, such as **Scale Virtual Machine**.

**NOTE:**
The action item gives the names of the affected entities. You can click on these entity names to drill down and set the Home Page scope to that specific entity. To return after drilling down to an entity in the action details, use the browser’s Back button.
- Immediately below the description, a summary of requirements, risks, opportunities, or reasons for the recommended action.
- The impact of executing the action.

For more information, see *Action Details (on page 82)*.

### C. Search

For a long list of pending actions, use search to narrow the results.

![Search Example]

### D. Filter and Sort

**Filter pending actions**

![Filter Example]

**Sort pending actions**

![Sort Example]
When you click **Filter**, you can:

- Filter the list by [action type](on page 68), [action mode](on page 73), [action category](on page 71), action prerequisite, or any combination of these items.
- Sort the actions in ascending or descending order by severity, name of the action target, risk category, or savings amount.

Workload Optimization Manager determines action severity by the amount of improvement the affected entities will gain by executing the action. Action severities are:

- **Minor** – Issues that affect cost or workload distribution, but not impact the QoS your users will experience
- **Major** – Issues that can affect QoS and should be addressed
- **Critical** – Issues that affect the QoS that your environment can deliver, and you are strongly advised to address them

For example:

- To see only the actions that you can execute through the Workload Optimization Manager user interface, filter the list by action mode and select **Manually executable**.

![FILTER](image)

- To see only resize actions that are manually executable and that give efficiency improvements, set the filter as follows:

![FILTER](image)

**E. Download**

Download the pending actions list as a CSV file.

**Action Details**

Each action in the Pending Actions list comes with a description and additional details to help you understand why Workload Optimization Manager recommends it and what you would gain if you execute it.

At first glance, some 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.
Example

In the image shown above, the action details indicate that scaling the virtual machine to a different instance type impacts RI coverage in a meaningful way. By increasing RI coverage from 0% to 100%, the projected hourly on-demand cost drops to $0, bringing estimated savings of $502 per month.

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 a 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:

<table>
<thead>
<tr>
<th>Entity Type</th>
<th>Monitored Resources</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual Machine</td>
<td>• vCPU</td>
<td>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. They also appear in Migrate to Cloud plan results.</td>
</tr>
<tr>
<td></td>
<td>• vMEM</td>
<td></td>
</tr>
<tr>
<td>Database (cloud)</td>
<td>• DTU</td>
<td>See <a href="#">Cloud Database Actions (on page 177)</a>.</td>
</tr>
<tr>
<td></td>
<td>• Storage</td>
<td></td>
</tr>
<tr>
<td>Volume (cloud)</td>
<td>• IOPS</td>
<td>These charts also appear when you scope to a given volume and view the Details page. See <a href="#">Cloud Volume Actions (on page 168)</a>.</td>
</tr>
<tr>
<td></td>
<td>• Throughput</td>
<td></td>
</tr>
<tr>
<td>Container Spec</td>
<td>• vCPU limits and requests</td>
<td>See <a href="#">Container Actions (on page 133)</a>.</td>
</tr>
<tr>
<td></td>
<td>• vMEM limits and requests</td>
<td></td>
</tr>
</tbody>
</table>

**Uptime Information for Cloud VMs**
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.
Getting Started

For more information, see Cloud VM Uptime (on page 38).

Action Details for AWS Workloads

In AWS environments, Workload Optimization Manager considers a VM's used and reserved memory to calculate virtual memory utilization, and drives actions based on the calculated value. This may not always match the values seen in CloudWatch or at the OS level of the VM.

According to the AWS FAQ, "In C5, portions of the total memory for an instance are reserved from use by the Operating System including areas used by the virtual BIOS for things like ACPI tables and for devices like the virtual video RAM." When Workload Optimization Manager recommends moving to one of these instances, the action details use the capacity that is reported by the instance template. However, subsequent reporting of the Mem capacity for the given instance uses the values that Workload Optimization Manager discovers in the environment.

Actions Tips and Best Practices

To get the best results from Workload Optimization Manager’s Application Resource Management, you should set as many actions as possible to Automated. If you want to approve any changes, set the actions to Manual.

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, scope the group in an automation policy, set the action mode to Automatic, and then set the schedule to off-peak hours or weekends. For details on setting schedules, see Setting Policy Schedules (on page 105).

Resize Actions

Allow VMs that have hot-add enabled to automatically resize up.

Use Tuned Scaling to automatically resize VM and storage resources when the resize amount falls within an acceptable range, and for Workload Optimization Manager to notify you when the amount falls outside the range so you can take the most appropriate action. For details, see Tuned Scaling for On-prem VMs (on page 189).

After executing a storage resize, Workload Optimization Manager indicates that the resize action has succeeded but the hypervisor might not show the corresponding change in storage capacity. If this occurs, perform a manual refresh of the hypervisor so it can discover the storage changes.

Move Actions

Workload Optimization Manager recommends automating host and storage migration.

Use placement constraints if you have placement requirements for specific workloads in your environment (for example, all production virtual machines moving only to specific clusters). Workload Optimization Manager can automatically import placement policies when you add a target, or you can create new placement policies. For more information, see Placement Policies (on page 87).

Working With Policies

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 two 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 ships with default settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.

The Policy Management page shows all the currently defined policies. From this page you can:

- Create new policies.
- Delete a user-created policy.
- Edit a default or user-created policy.
- Enable or disable discovered placement policies. For a Workload Optimization Manager segment (a placement policy that was created in Workload Optimization Manager), you can edit the policy definition as well as enable/disable it.

To see the policies that are applied to a scope, go to the Search page and set the Workload Optimization Manager session to that scope. Then show the Policy view. For more information, see Scope Policies (on page 57).
Things You Can Do

- Manage Imported Placement Policies – Importing Workload Placement Policies (on page 87)
- Create a Placement Policy – Creating Placement Policies (on page 87)
- Create a Scoped Automation Policy – Creating Scoped Automation Policies (on page 96)

Placement Policies

For planning and optimization, Workload Optimization Manager recommends actions to place workloads such as applications, or VMs on their providers (hosts, datastores, disk arrays, networks, etc.). Workload Optimization Manager can recommend these actions, or execute them automatically.

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 within Workload Optimization Manager you can create even more boundaries. 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 moves. When managing segments you can:

- Review the placement policies that Workload Optimization Manager has discovered. These are policies that have been defined in your environment, outside of Workload Optimization Manager. See Importing Workload Placement Policies (on page 87).
- Create placement segments that restrict workload placement according to specific rules. See Creating Placement Policies (on page 87).

NOTE:
You can enable or disable any imported policy or created workload placement segment to affect placement calculations in the real-time environment or in plans.

Importing Workload Placement Policies

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.

Workload Optimization Manager imports:

- vCenter Server DRS Rules
  See "Other Information Imported from vCenter" in the Target Configuration Guide
- Virtual Machine Manager Availability Sets
  See "Virtual Machine Manager" in the Target Configuration Guide

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.

Creating Placement Policies

Placement Policies set up constraints to affect how Workload Optimization Manager calculates the placement of workloads in your environment. In this way, you can direct Workload Optimization Manager to recommend actions that satisfy business rules for your enterprise.

Workload Optimization Manager discovers Placement policies that have been defined in your environment, and you can also create Placement policies through the Workload Optimization Manager user interface. Note that you can enable or disable any Placement policy, both for real-time analysis and for planning scenarios.
Workload Optimization Manager supports 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.

1. **Navigate to the Settings Page.**

   ![Settings](image)

   Click to navigate to the Settings Page. From there, you can perform a variety of Workload Optimization Manager configuration tasks.

2. **Choose Policies.**

   ![Policies](image)

   Click to navigate to the Policy Management Page.

   This page lists all the policies that you currently have configured for Workload Optimization Manager.

3. **Create a new Placement policy.**
First, select the type of Placement policy to create, then specify the settings:

- Give the policy a name
- Choose the policy type and make the settings
- Save the policy when you’re done

4. Create a **Place** policy.
These policies control where workload can be placed. For example, you can specify that a VM will only be placed on a host that is a member of a specific cluster. Or you could specify that any applications in a specific group can only be placed on a datastore that is a member of a specific group.

- **Specify the consumer group** – The group or cluster of entities that will be placed on the identified providers
- **Specify the provider group** – The group or cluster of entities that will provide resources to the consumers
- **Limit workload entities to placement group** – Set the policy to only place consumer entities on members of the provider group
- **Limit the maximum number of workload entities per placement entity to** – Limit how many instances of the consumer entities can be placed on a single provider

5. Create a **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. Or you can specify that a set of non-critical applications will never be placed on specialized hardware, as a way to ensure availability for critical applications.

- **Specify the consumer group** — The group or cluster of entities that will be excluded from the identified providers
- **Specify the provider group** — The group or cluster of entities that will not provide resources to the consumers

6. Create a **Merge** policy.

<table>
<thead>
<tr>
<th>POLICY NAME</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PLACE DON'T PLACE MERGE LICENSE</td>
</tr>
</tbody>
</table>

**MERGE**

Choose consumer type.. SELECT CLUSTERS

You can create placement policies that merge multiple clusters into a single logical group for the purpose of 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.

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 datacenter.
- For vCenter environments, you can create placement policies that merge datacenters to support cross-vCenter moves. In this case, where a datacenter corresponds to a given vCenter target, the merged clusters can be in different datacenters. In this case you must create two merge policies; one to merge the affected datacenters, and another to merge the specific clusters.

Also note that the clusters you merge must use the same network names on their respective datacenters.
- For cloud environments, you can create policies to merge datacenters. Use these merge policies to support VM moves that find better costs on other zones.

To create a Merge policy, choose the type of entity to merge, and then select the groups you will merge.

7. Create a **License** policy.

<table>
<thead>
<tr>
<th>POLICY NAME</th>
<th>TYPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PLACE DON'T PLACE MERGE LICENSE</td>
</tr>
</tbody>
</table>

**LICENSE**

Choose consumer type.. SELECT GROUP

ON

Choose provider type... SELECT GROUP
Assume you have purchased a number of licenses for a database – you pay for the right to run that database on a certain number of hosts. You can create a license policy to identify the hosts that provide the license, and the VMs that can consume that license.

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 will see two recommendations – move VM_02 to Host_01 and then remove the license in Host_01. You will not see a recommendation to suspend Host_01.

- 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 currently providing licenses, Workload Optimization Manager recommends adding licenses to the hosts to meet demand.

  These actions are more efficient than provisioning new hosts.

To create a License policy:

- Specify the license consumers (VMs).
- Specify the license providers (hosts).

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.

8. When you have made all your settings, be sure to save the Policy.

### Automation Policies

As Workload Optimization Manager gathers metrics, it compares the metric values against specified constraint and capacity settings to determine whether a metric exhibits a problem, and what actions to recommend or execute to avoid a problem. Workload Optimization Manager uses Automation Policies to guide its analysis and resulting actions. These policies can specify:

- Action Automation
  Whether to execute automatically or manually, or whether to just recommend the action. For more information, see Action Automation (on page 106).

- Constraints and Other Settings
  Settings that affect the Workload Optimization Manager analysis of the state of your environment. These include operational, utilization, and scaling constraints.

For more information, see Constraints and Other Settings (on page 116).

### Default and Scoped Automation Policies

Workload Optimization Manager ships with default Automation Policy setting for the different types of entities it can discover in your environment. The settings for these default policies should be adequate to meet your initial business requirements. These policies apply to the global scope – Unless you override them, they affect all the entities in your environment. For more information, see Working With Default Automation Policies (on page 93).

Workload Optimization Manager can include scoped Action Policies, which override the default settings for certain entities. With these policies you specify one or more groups of entities as the policy scope. You can also set a schedule to the policy to specify maintenance windows, or to support orchestration workflows that require approval before executing the given action. For more information, see Working With Scoped Automation Policies (on page 95) and Setting Policy Schedules (on page 105).
**Working With Default Automation Policies**

Workload Optimization Manager ships with default Automation Policy settings for the different types of entities it can discover in your environment. The settings for these default policies should be adequate to meet your initial business requirements. These policies apply to the global scope – Unless you override their settings, they affect all the entities in your environment.

Over time you might learn that you want to make global changes to certain policy settings. For example, **Enforce Non Disruptive Mode** is turned off by default. You might learn that in most cases you want to turn it on, and only turn it off for select scopes. In that case, you would turn it on in the default Automation Policy for VMs, and then set scoped policies for those groups of VMs for which you want to turn it off.

**Relationships Between Default and Scoped Policies**

Your default Automation Policies and scoped Automation Policies take effect in relation to each other. A default policy has a global effect, while a scoped policy overrides the default policy for the entities within the indicated scope. You should keep the following points in mind:

- Scoped policies set overrides to specific settings
  A scoped policy can override a subset of settings for the entity type, and for the remainder Workload Optimization Manager will use the default policy settings on the indicated scope.

- Among scoped policies, the most conservative setting wins
  It’s possible to set up policies with conflicts on individual entities. Assume two groups, Group_A and Group_B. Now imagine that one entity is a member of both groups. What happens if you create two different Automation Policies, one for Group_A and another for Group_B? In that case, the entity that is in both groups can have different policy settings.

  For example, the Group_A policy could set the Suspend action to **Manual**, while the setting for Group_B is **Recommend**. Workload Optimization Manager always uses the most conservative setting. For this case, the **Recommend** setting is most conservative, so it wins.

- Scoped policies always take precedence over default policies
  Even if the default policy has a more conservative setting, the setting in the scoped policy wins for entities in that scope.

- For a global effect, always use default policies
  Because of the conservative setting wins rule for scoped policies, you should never use a scoped policy to set a global effect. For example, you can create a scoped policy for the **All VMs** group. If you then specify a conservative setting for that policy, no other scoped policy can specify a more aggressive setting – the conservative setting will always win.

  For this reason, you should always use default Automation Policies whenever you want to achieve a global effect.

**Viewing and Editing Default Automation Policies**

To view or edit your default policies:

1. Navigate to the Settings Page.
   
   ![Settings](image)

   Click to navigate to the Settings Page. From there, you can perform a variety of Workload Optimization Manager configuration tasks.

2. Choose Policies.
   
   ![Policies](image)

   Click to navigate to the Policy Management Page.

   This page lists all the policies that you currently have configured for Workload Optimization Manager.

3. On the Policy Management page, click **Defaults**.
Getting Started

The page displays a list of all the default policies, by entity type.

4. Click the entity type whose default settings you wish to view or change.

   A fly-out appears with all the settings for that default policy. You can navigate to view different settings

5. Optionally, edit settings for this default policy.

   Navigate to the settings you want to change, and enter a different value for each.

6. When you’re done, click **Save and Apply**.

**Global Default Policy**

Use these settings to modify Workload Optimization Manager analysis globally for any scope of your environment. These defaults affect both scoped automation policies and default automation policies.

**ACTION AUTOMATION**

**Disable All Actions**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disable All Actions</td>
<td>OFF</td>
</tr>
</tbody>
</table>

When this is ON, Workload Optimization Manager does not generate any actions for your environment. For example, assume you have configured a number of policies 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.

**OPERATIONAL CONSTRAINTS**

**VM Growth Observation Period**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>VM Growth Observation Period</td>
<td>1 month</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allow Unlimited Host Provisioning</td>
<td>OFF</td>
</tr>
</tbody>
</table>

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.

### Working With Scoped Automation Policies

To override the current default Automation Policies, you can create scoped policies. These specify settings you want to change for certain entities in your environment. For these policies, you assign the policy to one or more groups of entities. In addition, you can assign a schedule to a scoped policy to set up maintenance windows or other scheduled actions in your environment.

Reasons to create scoped Automation Policies include:

- **Change the Analysis Settings for Certain Entities**
  
  Workload Optimization Manager uses a number of 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. You can configure scoped policies to modify Operational Constraints or Scaling Constraints. For more information, see [Constraints and Other Settings](on page 116).

- **Phase In Action Automation**
  
  Assume 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 mode to Automatic for different actions on those VMs (see [Action Modes](on page 73)).

For the steps to create a scoped policy, see [Creating Scoped Automation Policies](on page 96).

### Discovered Scoped 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.
Creating Scoped Automation Policies
Create a new scoped Automation Policy from the Policy Management Page.

1. Entry Point
Navigate to the Settings Page and then choose Policies.

This opens the Policy Management Page, which lists all the currently available policies.

Click NEW AUTOMATION POLICY and then select the policy type.
This sets the type of entity that your policy will affect. Note that Workload Optimization Manager supports different actions for different types of entities. For example, you cannot add VMem to a storage device. Setting policy type is the first step you take to focus on which actions you want to map to your workflows.

2. Policy Name

Name the policy.
3. Scope

The scope determines which entities this policy will affect. Choose one or more groups, or create new groups and add them to the policy scope. These groups match the type of entity you have 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 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 Action Policy settings. For conflicts among scoped policy settings, the most conservative setting will take effect. For more details, see Policy Scope (on page 104).
4. Policy Schedule

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.

For use cases and information about how schedules affect policies, see Policy Schedules (on page 105).
• **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**. See [Creating Schedules](page 362).

**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.

### 5. Automation and Orchestration

You can define automation and orchestration 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 via an Orchestrator (such as ServiceNow).
5.1. Action Type

See a list of actions that are viable for the policy, and then make your selections.
5.2. Action Generation and Acceptance

- 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 from the following Action Acceptance modes to indicate how you would like the actions to execute:
  - Recommend – Recommend the action so a user can execute it via the given hypervisor or by other means
  - Manual – Recommend the action, and provide the option to execute that action through the Workload Optimization Manager user interface
  - Automatic – Execute the action automatically

If you have a ServiceNow target, and that target includes an installation of the Workload Optimization Manager Actions application, you can send the action to ServiceNow. Choose from the following options:

- Generate Action then Send Record to ServiceNow
- Generate Action then Request Approval from ServiceNow

For more information, see Action Orchestration (on page 106).

5.3. Before Execution, Action 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 Action Orchestration (on page 106).

5.4. 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 105).

6. Constraints and Other Settings

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. For information about the settings you can make, see Constraints and Other Settings (on page 116).
Policy Scope

You must declare a scope whenever you make a scoped Automation Policy. The scope determines which entities will be affected by the policy settings. To set scope, you assign one or more groups to the policy. You can use discovered groups, or you can create your own groups. For information about creating groups, see Creating Groups (on page 357).

Relationships Between Default and Scoped Policies

Your default Automation Policies and scoped Automation Policies take effect in relation to each other. A default policy has a global effect, while a scoped policy overrides the default policy for the entities within the indicated scope. You should keep the following points in mind:

- Scoped policies set overrides to specific settings
  A scoped policy can override a subset of settings for the entity type, and for the remainder Workload Optimization Manager will use the default policy settings on the indicated scope.
- Among scoped policies, the most conservative setting wins
  It’s possible to set up policies with conflicts on individual entities. Assume two groups, Group_A and Group_B. Now imagine that one entity is a member of both groups. What happens if you create two different Automation Policies, one for Group_A and another for Group_B? In that case, the entity that is in both groups can have different policy settings.
  For example, the Group_A policy could set the Suspend action to Manual, while the setting for Group_B is Recommend. Workload Optimization Manager always uses the most conservative setting. For this case, the Recommend setting is most conservative, so it wins.
• Scoped policies always take precedence over default policies
  Even if the default policy has a more conservative setting, the setting in the scoped policy wins for entities in that scope.
• For a global effect, always use default policies
  Because of the conservative setting wins rule for scoped policies, you should never use a scoped policy to set a global effect. For example, you can create a scoped policy for the All VMs group. If you then specify a conservative setting for that policy, no other scoped policy can specify a more aggressive setting – the conservative setting will always win.
  For this reason, you should always use default Automation Policies whenever you want to achieve a global effect.

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 it generates. You can set up scheduled times when you want to change those settings.

Remember that for scoped automation policies, it is possible that one entity can be in two different scopes – This means the entity can be under the effect of two different policies. For this reason, scoped policies keep the rule, the most conservative setting wins. However, a more aggressive scoped policy takes precedence over the corresponding default automation policy. For more details, see Policy Scope (on page 104).

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 scoped policy, then the conservative settings win, and the scheduled changes do not take effect.

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 105).

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 scoped 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 default or scoped 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 192)*.

**Action Automation**

To avoid problems in your environment, Workload Optimization Manager analysis identifies actions that you can execute to keep things in optimal running order. You can specify the **degree of automation** you want for these given 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 modes** in a policy to set that business rule.

Action 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 modes in a policy to set that business rule.

Workload Optimization Manager supports the following action modes:

- **Recommend** – Recommend the action so a user can execute it via the given hypervisor or by other means
- **Manual** – Recommend the action, and provide the option to execute that action through the Workload Optimization Manager user interface
- **Automatic** – Execute the action automatically

**Action Mode Configuration**

There are two ways to configure action modes:

- Change the action mode in a default policy. For details, see *Working With Default Automation Policies (on page 93)*.
- Create an automation policy, scope the policy to specific entities or groups, and then select the action 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.

For details, see *Creating Scoped Automation Policies (on page 96)*.

**Action Orchestration**

Action Orchestration specifies whether Workload Optimization Manager will execute an action, or whether Workload Optimization Manager will pass the action request to an orchestrator or action script that will execute its own workflow 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.
About Orchestrators

Action 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 mode is Manual or Automatic, then when Workload Optimization Manager recommends the action, it will direct the orchestrator to use the mapped workflow to execute it.

Workload Optimization Manager supports integration with ServiceNow. You can configure policies that log Workload Optimization Manager actions in your ServiceNow instance, and that submit actions for approval in ServiceNow workflows.

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. For information about Orchestration target requirements, see the Target Configuration Guide.

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 mode will be Automatic (to set up the maintenance window). For more information, see Setting Policy Schedules (on page 105).

About Action Scripts

Action Scripts provide a script interface that can add custom processing to Workload Optimization Manager actions at different entry points. For example, you can create a script that sends an email whenever Workload Optimization Manager recommends moving a VM, or you can create a script that runs as a replacement for the action that Workload Optimization Manager would execute.

You deploy action scripts on a remote machine, and configure an Action Script target that communicates with this Action Script server. Workload Optimization Manager discovers the exposed scripts and displays them as options you can choose when you specify an Action Script in your orchestration policy.

For more information about Action Scripts, see Deploying Action Scripts (on page 111).
Specifying Action Orchestration

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.

1. Expand **Automation and Orchestration** and click **Add Action**. Then select the action type you want to orchestrate.
There is no orchestration for this action by default. The following table describes the supported orchestration workflows.
<table>
<thead>
<tr>
<th>Workflow 1: Generate Actions</th>
<th>Workflow 2: Generate Action then Send Record to ServiceNow</th>
<th>Workflow 3: Generate Action then Request Approval from ServiceNow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Send a record of the generated actions to ServiceNow.</td>
<td>Defer the generated actions to your ServiceNow workflow for approval. Workload Optimization Manager passes control for this action to your ServiceNow workflow as a Change Request (CR).</td>
</tr>
</tbody>
</table>
| Prerequisites                | • A ServiceNow target that includes an installation of the Workload Optimization Manager Actions application  
• An appropriate workflow set up for Records as part of the Workload Optimization Manager Actions installation  
• An Action Script target (if using an Action Script to control action execution) | • A ServiceNow target that includes an installation of the Workload Optimization Manager Actions application  
• An appropriate workflow set up for Approvals as part of the Workload Optimization Manager Actions installation  
• An Action Script target (if using an Action Script to control action execution) |
| Action Acceptance           | Choose from the following:  
• Recommend – Recommend the action so a user can execute it via the given hypervisor or by other means  
• Manual – Recommend the action, and provide the option to execute that action through the Workload Optimization Manager user interface  
• Automatic – Execute the action automatically | Action acceptance automatically changes to "External Approval". If the action is approved, the action executes using the default Action Acceptance mode. |
| Before Execution            | The default is Do Nothing. Select Use Action Script to run an action script that is set up for the PRE entry point. The action script name must match the entity type and action type. For example, you can post an email notification to your team that an action has been generated. | |
| Action Execution            | The default is Native. Workload Optimization Manager executes the action with its default action processing. Select Use Action Script to let Workload Optimization Manager execute a matching action script in place of its default action processing. You must have created and deployed an action script that matches the given entry point (for action execution, REPLACE), the given action, and the given entity type. | |
| 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. | 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. |
Workflow 1: 
Generate Actions

Workflow 2: 
Generate Action then Send Record to ServiceNow

Workflow 3: 
Generate Action then Request Approval from ServiceNow

Note: Workload Optimization Manager discovers and enforces execution schedules defined in ServiceNow approval workflows. To avoid potential issues with schedules, set the execution schedule either in ServiceNow or Workload Optimization Manager.

After Execution

The default is Do Nothing. Select Use Action Script to run an action script that is set up for the POST entry point. The action script name must match the entity type and action type.

The default is Do Nothing. Other options include:

- Create Record in ServiceNow
  Workload Optimization Manager registers the action in the ServiceNow log, showing that the given action has been executed.
- Use Action Script
  Run an action script that is set up for the POST entry point. The action script name must match the entity type and action type.

2. When you have made all your settings, be sure to save the action policy.

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 111))
- Create the action script executables on the remote server (see Creating Action Scripts (on page 112))
- Deploy the Action Script Manifest on the remote server (see Deploying the Action Script Manifest (on page 113))

Setting Up the Action Script Server

Workload Optimization Manager uses remote servers to execute 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.
Getting Started

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. This includes libraries, language processors, or other processes that your scripts will invoke.

Workload Optimization Manager invokes the scripts as commands on the server. The server must run an SSH service that you have configured to support command execution and SFTP operations. At this time, Cisco has 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 that an action script can invoke any process you have deployed on the remote server. You do not have to run scripts per se. 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 execute the scripts on your server, Workload Optimization Manager logs on via a user account that is authorized to execute 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 a 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.
- **Security for the `.ssh` Directory**
  - The Action Script User should be the only user with authorized access. You should set file permissions to 600.
- **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.

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 terminate the execution of the process.

Note that Workload Optimization Manager does not make any other attempt to terminate a process. For example you could implement the script so it traps the SIGTERM and continues to run. The process should terminate at the soonest safe opportunity. However, if the process does not terminate, then you must implement some way to terminate it outside of Workload Optimization Manager. Note that a runaway process continues to use its execution thread. This 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 113). 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:

```bash
myActionData=$(cat -)
```

stdin contains a JSON string that represents the full data associated with this action. For example, the myActionData variable could contain a string similar to:

```json
{"actionType":"RIGHT_SIZE","actionItem":[]}
```

Deploying the Action Script Manifest

The Action Script Manifest identifies the scripts that you want to expose to Workload Optimization Manager. You provide the location of the manifest 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

The Scripts Manifest is a file that declares an array of Script Objects for each script you want to expose. 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:**

```yaml
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:**

```json
{
  "scripts": [
    {
      "name": "MyVmMovePrep",
      "description": "Execute this script in preparation 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 an 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
  - STORAGE_CONTROLLER
  - IO_MODULE
  - APPLICATION_SERVER
  - APPLICATION
  - CONTAINeR
  - CONTAINER_POD
  - LOGICAL_POOL
  - STORAGE_VOLUME
  - DPOD
  - VPOD
  - DATABASE
  - LOAD_BALANCER

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
  - SUSPEND
  - TERMINATE
  - SPAWN
  - ADD_PROVIDER
  - CHANGE
  - REMOVE_PROVIDER
  - PROVISION
  - RECONFIGURE
  - RIGHT_SIZE
  - RESIZE_CAPACITY
  - WARN
  - RECONFIGURE_THRESHOD
• 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 in stead 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).

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.

Action 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.

Workload Optimization Manager ships with default policies that show all the constraints and settings you can make for each policy. These take effect until you create and apply a policy with different values. For the steps in creating a new policy, see Creating Scoped Automation Policies (on page 96). You can edit the defaults if you want to change analysis globally.
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.

Business Application

A Business Application is a logical grouping of Business Transactions (on page 120), Services (on page 122), Application Components (on page 124), 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**

<table>
<thead>
<tr>
<th>Synopsis</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Budget</td>
<td>Business Applications have unlimited budget.</td>
</tr>
<tr>
<td>Provides</td>
<td>The complete application to end users</td>
</tr>
<tr>
<td>Consumes from</td>
<td>Business Transactions, Services, Application Components, Database Servers, and the underlying nodes</td>
</tr>
<tr>
<td>Discovery</td>
<td>Workload Optimization Manager discovers the following:</td>
</tr>
<tr>
<td></td>
<td>▪ AppDynamics Business Applications</td>
</tr>
<tr>
<td></td>
<td>▪ Dynatrace Applications</td>
</tr>
<tr>
<td></td>
<td>If you do not have these targets, you can create your own Business Applications using the Application Topology feature. For details, see Application Topology (on page 128).</td>
</tr>
</tbody>
</table>

**Monitored Resources**

Workload Optimization Manager monitors the following:

- **Response Time**
  - The utilization of the server’s allocated response time
  - Measured in Milliseconds (ms)
- **Transactions**
  - The utilization of the allocated transactions per second for the given entity
  - Measured in transactions per second

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**

Workload Optimization Manager does not recommend actions for a Business Application, but it does recommend actions for the underlying Application Components and nodes. 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 settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.

**Action Automation and Orchestration**

Workload Optimization Manager does not recommend actions for a Business Application, but it does recommend actions for the underlying Application Components and nodes. 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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting/Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enable Transaction SLO</td>
<td>Off</td>
</tr>
<tr>
<td></td>
<td>Workload Optimization Manager estimates SLO based on monitored values.</td>
</tr>
<tr>
<td>Transaction SLO</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>If you enable SLO, Workload Optimization Manager uses the default value of 10. You can change this to a different value.</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting/Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enable Response Time SLO</td>
<td>Off</td>
</tr>
<tr>
<td></td>
<td>Workload Optimization Manager estimates SLO based on monitored values.</td>
</tr>
<tr>
<td>Response Time SLO [ms]</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>If you enable SLO, Workload Optimization Manager uses the default value of 2000. You can change this to a different value.</td>
</tr>
</tbody>
</table>

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. For more information, see Business Application (on page 117).

Synopsis

<table>
<thead>
<tr>
<th>Synopsis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Budget:</strong></td>
</tr>
<tr>
<td><strong>Provides:</strong></td>
</tr>
<tr>
<td><strong>Consumes from:</strong></td>
</tr>
</tbody>
</table>
Synopsis

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 Topology feature. For details, see Application Topology (on page 128).

Monitored Resources

Workload Optimization Manager monitors the following:

- Response Time
  - The utilization of the server’s allocated response time
    - Measured in Milliseconds (ms)
- Transactions
  - The utilization of the allocated transactions per second for the given business transaction
  - Measured in transactions per second

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

Workload Optimization Manager does not recommend actions for a Business Transaction, but it does recommend actions for the underlying Application Components and nodes. 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 settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.

Action Automation and Orchestration

Workload Optimization Manager does not recommend actions for a Business Transaction, but it does recommend actions for the underlying Application Components and nodes. 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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting/Value</th>
</tr>
</thead>
</table>
| 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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting/Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enable Response Time SLO</td>
<td>Off</td>
</tr>
<tr>
<td></td>
<td>Workload Optimization Manager estimates SLO based on monitored values.</td>
</tr>
<tr>
<td>Response Time SLO [ms]</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>If you enable SLO, Workload Optimization Manager uses the default value of 2000. You can change this to a different value.</td>
</tr>
</tbody>
</table>

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.

For more information, see Application Component (on page 124) and Business Application (on page 117).

**Synopsis**
### Synopsis

**Budget:** Services have unlimited budget.

**Provides:** Response time and transactions to [Business Transactions](#) and Business Applications

**Consumes from:** Application Components, Database Servers, and the underlying nodes

**Discovery:** For APM targets, Workload Optimization Manager discovers the following:
- AppDynamics Tiers
- Dynatrace Services
- NewRelic APM Applications / NewRelic Services (New Relic ONE)

If you do not have an APM target, you can create your own Services using the Application Topology feature. For details, see [Application Topology](#).

For Kubernetes, Workload Optimization Manager discovers Kubernetes and New Relic Services through the Kubeturbo pod that you have deployed in your environment.

### Monitored Resources

Workload Optimization Manager monitors the following:
- **Response Time**
  - The utilization of the server’s allocated response time
  - Measured in Milliseconds (ms)
- **Transactions**
  - The utilization of the allocated transactions per second for the given service
  - Measured in transactions per second

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

Workload Optimization Manager does not recommend actions for a Service, but it does recommend actions for the underlying Application Components and nodes. The Pending Actions chart for a Service lists these actions, thus providing visibility into the risks that have a direct impact on the Service’s performance.

### Service Policies

Workload Optimization Manager ships with default settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change *action automation* or *constraints* for that scope. You can create policies that override the defaults for the scopes you specify.

### Action Automation and Orchestration

Workload Optimization Manager does not recommend actions for a Service, but it does recommend actions for the underlying Application Components and nodes. The Pending Actions chart for a Service lists these actions, thus providing visibility into the risks that have a direct impact on the Service’s performance.

### Transaction SLO

Enable this SLO if you are monitoring performance through your Services.
### Entity Types - Applications

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting/Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enable Transaction SLO</td>
<td>Off</td>
</tr>
<tr>
<td></td>
<td>Workload Optimization Manager estimates SLO based on monitored values.</td>
</tr>
<tr>
<td>Transaction SLO</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>If you enable SLO, Workload Optimization Manager uses the default value of 10. You can change this to a different value.</td>
</tr>
</tbody>
</table>

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 Services.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting/Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enable Response Time SLO</td>
<td>Off</td>
</tr>
<tr>
<td></td>
<td>Workload Optimization Manager estimates SLO based on monitored values.</td>
</tr>
<tr>
<td>Response Time SLO [ms]</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>If you enable SLO, Workload Optimization Manager uses the default value of 2000. You can change this to a different value.</td>
</tr>
</tbody>
</table>

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%.

### Application Component

An Application Component is a software component, application code, or a unit of processing within a Service (on page 122) that consumes resources to enable it to perform its function for the Business Application (on page 117). 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

<table>
<thead>
<tr>
<th>Budget:</th>
<th>Application Components have unlimited budget.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides:</td>
<td>Response Time and Transactions to Services, <a href="#">Business Transactions (on page 120)</a>, and Business Applications</td>
</tr>
<tr>
<td></td>
<td>Response Time, Transactions, Heap, Remaining GC Capacity, and Threads to end users</td>
</tr>
<tr>
<td>Consumes:</td>
<td>Compute resources from nodes</td>
</tr>
<tr>
<td>Discovery:</td>
<td>Workload Optimization Manager discovers the following:</td>
</tr>
<tr>
<td></td>
<td>Apache Tomcat</td>
</tr>
<tr>
<td></td>
<td>AppDynamics Nodes</td>
</tr>
<tr>
<td></td>
<td>Dynatrace Processes</td>
</tr>
<tr>
<td></td>
<td>NewRelic APM Application Instances</td>
</tr>
<tr>
<td></td>
<td>Application Insights Components</td>
</tr>
<tr>
<td></td>
<td>SNMP</td>
</tr>
<tr>
<td></td>
<td>WMI</td>
</tr>
<tr>
<td></td>
<td>Data Ingestion Framework metrics for Kubernetes environments</td>
</tr>
</tbody>
</table>

### Monitored Resources

The exact resources monitored will differ based on application type. This list includes all resources you may see.

- Virtual CPU (VCPU)
  - The utilization of the VCPU allocated to the hosting VM
  - Measured in Megahertz (MHz)
- Virtual Memory (VMem)
The utilization of the VMem allocated to the hosting VM
Measured in Kilobytes (KB)
- Transactions
  The utilization of the allocated transactions per second for the given entity
  Measured in transactions per second
- Heap
  The utilization of the application server’s heap
  Measured in Kilobytes (KB)
- Response Time
  The utilization of the server’s allocated response time
  Measured in Milliseconds (ms)
- Connections
  The utilization of the connection capacity. Only applicable to database servers
  Measured in Connections
- Remaining Garbage Collection Capacity
  The percentage of server uptime spent not performing garbage collection
  Measured in uptime (%)
- Threads
  The utilization of the server’s thread capacity
  Measured in number of Threads

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. To include SLOs in the analysis, set SLO values in policies.

Application Component Actions

The default Scaling Policy is Resize. With this policy, Workload Optimization Manager will only recommend resize actions. If you change the Scaling Policy to Horizontal Scale, it will only recommend provision and suspend actions.

For resize actions, Workload Optimization Manager can recommend resizing the following resources:
- Thread Pool
- Connections
- Heap
  For Heap, Workload Optimization Manager will generate recommendations if all of the following are satisfied:
    - The Application Component sells Heap and Remaining Garbage Collection (GC) Capacity.
    - The underlying VM or container sells VMem.

For provision and suspend actions, Workload Optimization Manager will generate recommendations if all of the following are satisfied:
- An application is horizontally scalable and is hosted on Kubernetes.
- There are Services that consume Response Time or Transaction commodities from the application.
- Workload Optimization Manager can retrieve Transaction or Response Time metrics for the application. Contact your Workload Optimization Manager representative for more information.
- You set Transaction or Response Time SLOs in policies.

NOTE:
Before setting SLOs, check that the supply chain has stitched the Application Components with the underlying container infrastructure. If stitching has not occurred and you set SLOs, the product might generate inaccurate actions.
Application Component Policies

Workload Optimization Manager ships with default settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.

Action Automation and Orchestration

The default Scaling Policy is Resize. With this policy, Workload Optimization Manager will only recommend resize actions. If you change the Scaling Policy to Horizontal Scale, it will only recommend provision and suspend actions.

For details, see Application Component Actions. (on page 126)

<table>
<thead>
<tr>
<th>Action</th>
<th>Default Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provision</td>
<td>Recommend</td>
</tr>
<tr>
<td>Suspend</td>
<td>Recommend</td>
</tr>
<tr>
<td>Resize heap (up or down)</td>
<td>Recommend</td>
</tr>
<tr>
<td>Resize thread pool (up or down)</td>
<td>Recommend</td>
</tr>
<tr>
<td>Resize connections (up or down)</td>
<td>Recommend</td>
</tr>
</tbody>
</table>

You can use Action Scripts (on page 111) for action orchestration. Third-party orchestrators (such as ServiceNow) are not supported.

Transaction SLO

Enable this SLO to monitor the performance of your Application Components.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting/Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enable Transaction SLO</td>
<td>Off</td>
</tr>
<tr>
<td></td>
<td>Workload Optimization Manager estimates SLO based on monitored values.</td>
</tr>
<tr>
<td>Transaction SLO</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>If you enable SLO, Workload Optimization Manager uses the default value of 10. You can change this to a different value.</td>
</tr>
</tbody>
</table>

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 to monitor the performance of your Application Components.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting/Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enable Response Time SLO</td>
<td>Off</td>
</tr>
<tr>
<td></td>
<td>Workload Optimization Manager estimates SLO based on monitored values.</td>
</tr>
<tr>
<td>Response Time SLO [ms]</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>If you enable SLO, Workload Optimization Manager uses the default value of 2000. You can change this to a different value.</td>
</tr>
</tbody>
</table>
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%.

**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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heap Utilization (%)</td>
<td>80</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heap Scaling Increment (MB)</td>
<td>128</td>
</tr>
</tbody>
</table>

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.

**Application Topology**

Workload Optimization Manager gives you the ability to create your own Business Applications (on page 117), Business Transactions (on page 120), and Services (on page 122) without the need to ingest additional application data into the platform. This is especially useful in environments 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. User-created application entities address 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.
Workload Optimization Manager does not perform analysis on any user-created application entity, but it aggregates the underlying risks the same way it does for auto-discovered entities.

After you create an application entity, Workload Optimization Manager counts it in the global supply chain and adds it to the relevant charts (for example, if you created a new Service that has performance risks, you might see it listed in the Top Services chart). Drill down to the newly created entity to monitor its performance. You can also use Search to find the application entity and set it as your scope.

**Creating Application Entities**

1. Navigate to the **Settings** Page.

2. Choose **Application Topology**.
3. Click **New Application Topology** and then choose **Automatic** or **Manual**.
   - **Automatic**
     Create a new application entity composed of tagged entities. For example, create a new Business Application composed of VMs with the "Production" tag.
   - **Manual**
     Create a new application entity composed of a specific set of application entities and nodes.

4. Select the application entity type that you want to create.

5. If you chose **Automatic**:
   - Type an entity name prefix to help you easily identify the application entities that Workload Optimization Manager will create for you, and then specify the tags that will identify the underlying entities.
   - If you chose **Manual**:
     Give the application entity a name and then add the underlying application entities and nodes.

6. Click **Create 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

An application container is a standalone, executable image of software that includes components to host an application.

**Synopsis**

<table>
<thead>
<tr>
<th>Budget:</th>
<th>A container obtains its budget by selling resources to the hosted application.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides:</td>
<td>Resources for the applications to use:</td>
</tr>
<tr>
<td></td>
<td>• Virtual CPU</td>
</tr>
<tr>
<td></td>
<td>• Virtual Memory</td>
</tr>
<tr>
<td>Consumes:</td>
<td>Resources from container pods, virtual machines, and virtual datacenters.</td>
</tr>
<tr>
<td></td>
<td>Note that container pods are not applicable for Pivotal Cloud Foundry.</td>
</tr>
</tbody>
</table>
**Synopsis**

<table>
<thead>
<tr>
<th>Discovered through:</th>
<th>For Kubernetes, Workload Optimization Manager discovers containers through the Kubeturbo pod that you have deployed in your environment.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>For Dynatrace and AppDynamics hosted on containers:</td>
</tr>
<tr>
<td></td>
<td>• Dynatrace: Workload Optimization Manager discovers containers through the metadata of processes.</td>
</tr>
<tr>
<td></td>
<td>• AppDynamics: Workload Optimization Manager discovers containers through container objects.</td>
</tr>
</tbody>
</table>

**Monitored Resources**

Workload Optimization Manager monitors the following resources for a container:

**VMem**

The virtual memory utilized by the container against the memory limit (if no limit is set, then node capacity is used). Measured in Megabytes (MB)

**VCPU**

The virtual CPU utilized by the container against the CPU limit (if no limit is set, then node capacity is used). Measured in millicores (mCores)

**VMem Request**

If applicable, the virtual memory utilized by the container against the memory request. Measured in Megabytes (MB)

**VCPU Request**

If applicable, the virtual CPU utilized by the container against the CPU request. Measured in millicores (mCores)

**Container Actions**

**Resize Actions**

Workload Optimization Manager supports resize actions for containers. By default, containers resize consistently, which allows all replicas of the same container for the same workload type to resize any commodity consistently. Workload Optimization Manager uses namespace or organization/space quotas as constraints when making resize decisions.

**Action Merging**

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.

**Action Visibility**

Workload Optimization Manager shows the container resize actions that it generates when you set the scope to Container Specs, which are persistent entities. You will *not* see actions when you set the scope to Containers.

To view the actions, go to the Pending Actions chart of the Container Spec and then click **Show All** to see the full list.

When you expand an action, the action details show the following:
• Individual and merged resize actions (read-only)

• Utilization charts

Workload Optimization Manager uses percentile calculations to measure container resource utilization more accurately, and drive actions that improve overall utilization. When you examine the details for a pending scaling action on a Container Spec, which persists historical utilization data for containers, 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 resize actions on the associated Workload Controller, the charts also show the resulting improvements in daily average utilization. Put together, these charts allow you to easily recognize utilization 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 141).*

• A link to the associated Workload Controller, under the **Container Spec Details** section

Workload Controllers execute resize actions. Click the link to set the scope to the Workload Controller and execute the action.

**Action Execution**

To execute container resize actions, set the scope to a Workload Controller and then select an action to execute.
Default Action Modes and Their Effect on Action Execution

By default, container resize actions are set in *Manual* mode. This means that Workload Optimization Manager will not execute any action automatically, and you can manually select the actions that you want to execute. The following table summarizes the default settings and behavior.

<table>
<thead>
<tr>
<th>Entity</th>
<th>Action Mode in Default Policy</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Container Spec</td>
<td>Manual</td>
<td>Workload Optimization Manager shows actions in the Container Spec’s Pending Actions chart. Since execution is via the Workload Controller, the checkboxes for the actions are grayed out. For each action, you can click the link to the associated Workload Controller to execute the action.</td>
</tr>
<tr>
<td>Workload Controller</td>
<td>Manual</td>
<td>Workload Optimization Manager shows actions in the Workload Controller’s Pending Actions chart, and enables the checkboxes for the actions so you can execute them.</td>
</tr>
</tbody>
</table>

You can override the default action modes to meet certain requirements.

- **Execute Actions Outside Workload Optimization Manager**
  
  If you prefer to execute certain actions outside Workload Optimization Manager, create a Container Spec policy and set the action mode for Resize actions to *Recommend*. Doing this hides all the actions at the Workload Controller level.

  **NOTE:**
  
  To see read-only actions at the Workload Controller level, keep the default Container Spec action mode (*Manual*), and set the action mode for Workload Controller to *Recommend*.

- **Disable Actions**
If you do not want Workload Optimization Manager to generate actions for certain Container Specs, create a Container Spec policy and set the action generation for Resizes to Do Not Generate Actions. This will disregard the action mode set in Workload Controllers.

- **Automate Actions**
  To automate the execution of generated actions, you will need two policies - one for Container Spec and one for Workload Controller. For the Container Spec policy, set the action mode for Resize actions to Manual. For the Workload Controller, set it to Automatic.

**Tuned Scaling for Container Specs**

Workload Optimization Manager can automate resizes if the resize values fall within a normal range, and then post more conservative actions when resize values fall outside the range. To do this, you would set specific action modes and tuned scaling values in policies.

For example, consider resizing vMem limits. As memory demand increases, Workload Optimization Manager can automatically execute vMem limit resizes that fall within the normal range. If the Container Spec requests memory beyond the normal range, Workload Optimization Manager will either ignore the action or post it for you to review, depending on the tuned scaling settings that you configured.

Assume the following tuned scaling settings in policies:

- The **Operational Constraints** settings in the Container Spec policy specify 100 MB to 500 MB as the normal range.
- With the **Resize** action modes in the Container Spec and Workload Controller policies set to Manual and Automatic, respectively, Workload Optimization Manager will automate resize up actions that are below the max threshold and resize down actions that are above the min threshold.
- In the Container Spec policy, since the action mode for vMem Limit Resize Above Max and vMem Limit Resize Below Min is set to Disabled, Workload Optimization Manager will not generate resize actions that fall outside the normal range.
NOTE:
If the action mode is Recommend, Workload Optimization Manager will post the actions for you to review. You can only execute these actions outside Workload Optimization Manager.

- Since vMEM increment constant is not defined in the Container Spec policy, Workload Optimization Manager uses the default value of 128 MB.

With these two policies in effect:
- If a Container Spec with 200 MB of vMEM limit needs to resize to 328 MB, Workload Optimization Manager automatically resizes to 328 MB.
- If a Container Spec with 200 MB of vMEM limit needs to resize to 72 MB, Workload Optimization Manager does not generate the action. vMEM limit remains at 200 MB.

NOTE:
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, Workload Optimization Manager uses the most conservative settings for the affected entities.

For tuned scaling, the effective action 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 Policy Scope (on page 104).

Container Policies
Workload Optimization Manager ships with default settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.

Action Automation and Orchestration
Workload Optimization Manager supports resize actions for containers. By default, containers resize consistently, which allows all replicas of the same container for the same workload type to resize any commodity consistently. Workload Optimization Manager uses namespace or organization/space quotas as constraints when making resize decisions.

Workload Optimization Manager shows actions via Container Specs, and executes them via Workload Controllers.

<table>
<thead>
<tr>
<th>Action</th>
<th>Default Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Container</td>
<td>Container Spec</td>
</tr>
<tr>
<td>Resize</td>
<td>N/A</td>
</tr>
</tbody>
</table>

For a detailed explanation of the default modes for container entities, see Default Action Modes and Their Effect on Action Execution (on page 135).

Consistent Resizing
- For groups in scoped policies:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consistent Resizing</td>
<td>Off</td>
</tr>
</tbody>
</table>

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 Kubernetes 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. It is a persistent entity that retains the historical utilization data of containers, which Workload Optimization Manager leverages to make container sizing decisions. Utilization data includes:

- vCPU used by all container replicas
- vCPU request capacity (if applicable)
- vMem used by all container replicas
- vMem request capacity (if applicable)

### Synopsis

<table>
<thead>
<tr>
<th></th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Budget:</td>
<td>N/A</td>
</tr>
<tr>
<td>Provides:</td>
<td>N/A</td>
</tr>
<tr>
<td>Consumes:</td>
<td>N/A</td>
</tr>
</tbody>
</table>
**Synopsis**

| Discovered through: | Kubeturbo Mediation Pod |

**Monitored Resources**

When you view the resources for a Container Spec, you will see the historical usage of any instance of a container running for the workload (assuming the workload name stays the same). The chart shows the trend of usage even with restarts or redeployments.

**Actions**

Workload Optimization Manager does not recommend actions to perform directly on Container Specs. Instead, it recommends actions to perform on containers. However, since containers are ephemeral while Container Specs are persistent, Workload Optimization Manager shows these actions at the Container Spec level.

To view the actions, set the scope to the Container Spec for related containers. Go to the Pending Actions chart and click **Show All** to see the full list.

For more information about container actions, including the various settings that control the way they are executed, see *Container Actions (on page 133)*.

**Container Spec Policies**

Workload Optimization Manager ships with default settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change *action automation* or *constraints* for that scope. You can create policies that override the defaults for the scopes you specify.

**Action Automation and Orchestration**

Workload Optimization Manager supports resize actions for containers. By default, containers resize consistently, which allows all replicas of the same container for the same workload type to resize any commodity consistently. Workload Optimization Manager uses namespace or organization/space quotas as constraints when making resize decisions.

Workload Optimization Manager shows actions via Container Specs, and executes them via Workload Controllers.

<table>
<thead>
<tr>
<th>Action</th>
<th>Default Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Container</td>
</tr>
<tr>
<td>Re-size</td>
<td>N/A</td>
</tr>
</tbody>
</table>

For a detailed explanation of the default modes for container entities, see *Default Action Modes and Their Effect on Action Execution (on page 135)*.

The following settings affect tuned scaling:

<table>
<thead>
<tr>
<th>Setting</th>
<th>Default Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>vCPU Request Resize Below Min</td>
<td>Recommend</td>
</tr>
<tr>
<td>vCPU Limit Resize Above Max</td>
<td>Recommend</td>
</tr>
<tr>
<td>vCPU Limit Resize Below Min</td>
<td>Recommend</td>
</tr>
<tr>
<td>vMem Request Resize Below Min</td>
<td>Recommend</td>
</tr>
<tr>
<td>vMem Limit Resize Above Max</td>
<td>Recommend</td>
</tr>
<tr>
<td>vMem Limit Resize Below Min</td>
<td>Recommend</td>
</tr>
</tbody>
</table>
The default mode of \textit{Recommend} means that when resize values in actions fall outside the normal range (as defined in Container Spec policies), Workload Optimization Manager will post the actions for you to review. You can only execute these actions outside Workload Optimization Manager. If you set the action mode to \textit{Disabled}, Workload Optimization Manager will not generate the actions.

For an overview of tuned scaling, see \textit{Tuned Scaling for Container Specs (on page 136)}.

### Resize Thresholds

Workload Optimization Manager uses resize thresholds as operational constraints to set up tuned scaling for Container Specs. For an overview of tuned scaling, see \textit{Tuned Scaling for Container Specs (on page 136)}.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>VCPU Request Resize Min Threshold (mCores)</td>
<td>10</td>
</tr>
<tr>
<td>VCPU Limit Resize Min Threshold (mCores)</td>
<td>500</td>
</tr>
<tr>
<td>VCPU Limit Resize Max Threshold (mCores)</td>
<td>64000</td>
</tr>
<tr>
<td>VMEM Request Resize Min Threshold (MB)</td>
<td>10</td>
</tr>
<tr>
<td>VMEM Resize Min Threshold (MB)</td>
<td>10</td>
</tr>
<tr>
<td>VMEM Resize Max Threshold (MB)</td>
<td>1048576</td>
</tr>
</tbody>
</table>

### Increment Constants

Workload Optimization Manager recommends changes in terms of the specified resize increments.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increment constant for VCPU Limit and VCPU Request (mCores)</td>
<td>100</td>
</tr>
<tr>
<td>Increment constant for VMEM Limit and VMEM Request (MB)</td>
<td>128</td>
</tr>
</tbody>
</table>

For example, assume the vCPU request increment is 100 mCores and you have requested 800 mCores for a container. Workload Optimization Manager could recommend to reduce the request by 100, down to 700 mCores.

For VMEM, you should not set the increment value to be lower than what is necessary for the container to operate. If the VMEM increment is too low, then it’s possible that Workload Optimization Manager would allocate insufficient VMEM. For a container that is underutilized, Workload Optimization Manager will reduce VMEM allocation by the increment amount, 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

(For the default policy only)

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate of Resize</td>
<td>High</td>
</tr>
</tbody>
</table>

When resizing resources for a container, Workload Optimization Manager calculates the optimal values for VMEM and vCPU. 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, as follows:

- **Low**
  Change the value by one increment, only. For example, if the resize action calls for increasing VMEM, and the increment is set at 128, Workload Optimization Manager increases VMEM by 128 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).

### 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**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggressiveness</td>
<td>99th Percentile</td>
</tr>
</tbody>
</table>

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 CPU 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 CPU 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 CPU for that container.

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 99% of the samples CPU never exceeded 50%. If you set **Aggressiveness** to 99th Percentile, then Workload Optimization Manager can see this as an opportunity to reduce CPU 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 – The least aggressive, recommended for critical workloads that need maximum guaranteed performance at all times.
- 99th Percentile (Default) – The recommended setting to achieve maximum performance and savings.
- 90th Percentile – Most aggressive, recommended for non-production workloads that can stand higher resource utilization.

- **Max Observation Period**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max Observation Period</td>
<td>Last 30 Days</td>
</tr>
</tbody>
</table>

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**
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

## Workload Controller

A *Workload Controller* is a Kubernetes controller that watches the state of your pods and then requests changes where needed. You can execute container resize actions when you set the scope to a Workload Controller.

### Synopsis

<table>
<thead>
<tr>
<th>Budget</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides</td>
<td>N/A</td>
</tr>
<tr>
<td>Consumes</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Entity Types - Container Platform

Synopsis

| Discovered through | Kubeturbo Mediation Pod |

Monitored Resources

Workload Optimization Manager does not monitor resources for Workload Controllers.

Actions

A Workload Controller executes container actions, but Workload Optimization Manager does not recommend actions to perform on the Workload Controller itself. When you scope to a Workload Controller and view the actions list, the actions apply to containers.

For more information about container actions, including the various settings that control the way they are executed, see Container Actions (on page 133).

Workload Controller Policies

Workload Optimization Manager ships with default settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.

Action Automation and Orchestration

Workload Optimization Manager supports resize actions for containers. By default, containers resize consistently, which allows all replicas of the same container for the same workload type to resize any commodity consistently. Workload Optimization Manager uses namespace or organization/space quotas as constraints when making resize decisions.

Workload Optimization Manager shows actions via Container Specs, and executes them via Workload Controllers.

<table>
<thead>
<tr>
<th>Action</th>
<th>Default Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Container</td>
</tr>
<tr>
<td>Re-size</td>
<td>N/A</td>
</tr>
</tbody>
</table>

For a detailed explanation of the default modes for container entities, see Default Action Modes and Their Effect on Action Execution (on page 135).

Action orchestration is currently not supported.

Container Pod

A ContainerPod is a Kubernetes pod, which is a group of one or more containers with shared storage or network resources and a specification for how to run the containers together.
## Synopsis

**Budget:**
A container pod obtains its budget by selling resources to containers.

**Provides:**
Resources for containers to use:
- Virtual CPU
- Virtual Memory

**Consumes:**
Resources from virtual machines and namespaces.

**Discovered through:**
Workload Optimization Manager discovers Kubernetes pods through the Kubeturbo pod that you have deployed in your environment.

### Monitored Resources

Workload Optimization Manager monitors the following resources for a container pod:
VMem
The virtual memory utilized by the pod against the node physical capacity. Measured in Megabytes (MB)

VCPU
The virtual CPU utilized by the pod against the node physical capacity. Measured in millicores (mCores)

VMem Request
The virtual memory request allocated by the pod against the node allocatable capacity. Measured in Megabytes (MB)

VCPU Request
The virtual CPU request allocated by the pod against the node allocatable capacity. Measured in millicores (mCores)

VMem Request Quota
If applicable, The amount of virtual memory request the pod has allocated against the namespace quota. Measured in Megabytes (MB)

VCPU Request Quota
If applicable, The amount of virtual CPU request the pod has allocated against the namespace quota. Measured in millicores (mCores)

VMem Limit Quota
If applicable, The amount of virtual memory limit the pod has allocated against the namespace quota. Measured in Megabytes (MB)

VCPU Limit Quota
If applicable, The amount of virtual CPU limit the pod has allocated against the namespace quota. Measured in millicores (mCores)

Container Pod Actions
Workload Optimization Manager can execute or recommend these actions for a container pod in Kubernetes platforms:

• Move
• Suspend
• Provision
• Reconfigure

Move Actions
The following items affect the Move actions that Workload Optimization Manager recommends for pods:

• Constraints
  Workload Optimization Manager respects the following constraints when making placement decisions for pods:
  – Kubernetes taints for nodes and tolerations for pods are treated 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.
  – Workload Optimization Manager imports Kubernetes 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.
  – Workload Optimization Manager recognizes pod affinity and anti-affinity 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 VMs, or a policy that prevents pods from moving to certain VMs.

  For more information, see Creating Placement Policies (on page 87).

• 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, while maintaining that one replica continues to run. You can disable temporary increases in quotas, but be aware that this will result in failure to move pods. To disable increases, set the following in the yaml resource for KubeTurbo deployment:

  \texttt{update-quota-to-allow-moves=false}

## Container Pod Policies

Workload Optimization Manager ships with default settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change \textit{action automation} or \textit{constraints} for that scope. You can create policies that override the defaults for the scopes you specify.

### Action Automation and Orchestration

For details about container pod actions, see Container Pod Actions (on page 145).

<table>
<thead>
<tr>
<th>Action</th>
<th>Default Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Move</td>
<td>Manual</td>
</tr>
<tr>
<td>Suspend</td>
<td>Manual</td>
</tr>
<tr>
<td>Provision</td>
<td>Manual</td>
</tr>
<tr>
<td>Reconfigure</td>
<td>Recommend</td>
</tr>
</tbody>
</table>

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 VMs, or a policy that prevents pods from moving to certain VMs.

For more information, see Creating Placement Policies (on page 87).

## Namespace

A namespace is a logical pool of resources in a Kubernetes environment 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**

<table>
<thead>
<tr>
<th>Budget:</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides:</td>
<td>N/A</td>
</tr>
<tr>
<td>Consumes:</td>
<td>N/A</td>
</tr>
<tr>
<td>Discovered through:</td>
<td>Kubeturbo Mediation Pod</td>
</tr>
</tbody>
</table>

**Resource Quotas**

A namespace can include the following compute resource quotas to describe its capacity:

**VMem Request Quota**

The total amount of virtual memory request for all pods allocated to the namespace against the namespace quota. Measured in Megabytes (MB)

**VCPU Request Quota**

The total amount of virtual CPU request for all pods allocated to the namespace against the namespace quota. Measured in millicores (mCores)

**VMem Limit Quota**

The total amount of virtual memory limit for all pods allocated to the namespace against the namespace quota. Measured in Megabytes (MB)

**VCPU Limit Quota**

The total amount of virtual CPU limit for all pods allocated to the namespace against the namespace quota. Measured in millicores (mCores)

Workload Optimization Manager treats quotas defined in namespaces as constraints when making sizing decisions for containers. When you scope to a namespace in the supply chain, 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.
### Capacity and Usage

#### nsquota

<table>
<thead>
<tr>
<th>Commodity</th>
<th>Capacity</th>
<th>Used</th>
<th>Utilization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory Request Quota</td>
<td>640 MB</td>
<td>640 MB</td>
<td>100%</td>
</tr>
<tr>
<td>CPU Limit Quota</td>
<td>500 mCores</td>
<td>500 mCores</td>
<td>100%</td>
</tr>
<tr>
<td>Memory Limit Quota</td>
<td>1.25 GB</td>
<td>1.25 GB</td>
<td>100%</td>
</tr>
<tr>
<td>CPU Request Quota</td>
<td>250 mCores</td>
<td>100 mCores</td>
<td>40%</td>
</tr>
<tr>
<td>Virtual Memory Request</td>
<td>90.99 GB</td>
<td>640 MB</td>
<td>0.69%</td>
</tr>
</tbody>
</table>

#### openshift-sdn

<table>
<thead>
<tr>
<th>Commodity</th>
<th>Capacity</th>
<th>Used</th>
<th>Utilization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory Request Quota</td>
<td>∞</td>
<td>4.99 GB</td>
<td>0%</td>
</tr>
<tr>
<td>CPU Request Quota</td>
<td>∞</td>
<td>1.03 Cores</td>
<td>0%</td>
</tr>
<tr>
<td>Memory Limit Quota</td>
<td>∞</td>
<td>0 KB</td>
<td>0%</td>
</tr>
<tr>
<td>CPU Limit Quota</td>
<td>∞</td>
<td>0 mCores</td>
<td>0%</td>
</tr>
<tr>
<td>Virtual Memory Request</td>
<td>192.04 GB</td>
<td>4.99 GB</td>
<td>2.6%</td>
</tr>
</tbody>
</table>

**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).

**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.

**Monitored Resources**

Workload Optimization Manager monitors actual utilization of VMem, VCPU, VMem Requests and VCPU Requests against cluster capacity.
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.

To see which namespaces use the most cluster resources, set the scope to a container cluster and see the **Top Namespaces** chart. You can use the data in the chart for showback analysis.

**Actions**

Workload Optimization Manager does not recommend actions for a namespace. Instead, it recommends actions for containers and container pods.

For more information, see [Container Actions](on page 133) and [Container Pod Actions](on page 145).

**Container Cluster**

A Container Cluster is a Kubernetes cluster that Workload Optimization Manager discovers through Kubeturbo. With this entity type, Workload Optimization Manager can fully link the entire container infrastructure with the underlying nodes, and then present all actions on containers and nodes in a single view. This gives you full visibility into the actions that impact the health of your container environment.
Synopsis

<table>
<thead>
<tr>
<th>Entity Types - Container Platform</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Entity Types - Container Platform</th>
</tr>
</thead>
</table>

| Budget: | N/A |
| Provides: | N/A |
| Consumes: | N/A |
| Discovered through: | Kubeturbo Mediation Pod |

**Monitored Resources**

Workload Optimization Manager does not monitor resources for Kubernetes clusters.
**Actions**

Workload Optimization Manager does not recommend actions for a Container 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 Cluster and view the Pending Actions chart.

For actions to suspend or provision nodes in the *public cloud*, Workload Optimization Manager includes cost information (investments or savings) attached to those actions. Note that Workload Optimization Manager generates these actions not to optimize costs, but to assure performance and efficiency for your container infrastructure. Workload Optimization Manager reports costs to help you track your cloud spend.

To view cost information, set the scope to a cluster in the public cloud and view the Necessary Investments or Potential Savings charts. You can also set the scope to the global cloud environment to see total costs, or to an individual node.

**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.

Kubernetes CPU Metrics

To meet user requirements and align with Kubernetes specifications, Workload Optimization Manager uses millicore (mCore) as the base unit for CPU metrics for your Kubernetes platform.
These include metrics for the following CPU-related commodities:

- vCPU
- vCPU Request
- vCPU Limit Quota
- vCPU Request Quota

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 Container Spec or Workload Controller to view pending resize actions (on page 133) for a container, you will see utilization and resize values in mCores.
- When you create Container Spec policies (on page 139), resize thresholds and increment constants for CPU-related commodities are set in mCores.
- For an Optimize Container Cluster plan, the plan results (on page 257) for CPU-related commodities are shown in mCores.
For nodes (VMs) and Application Components:

- For nodes stitched to your Kubernetes platform, the base unit for vCPU Request is also mCore, since this commodity is provided only to Kubernetes.
- For both nodes and Application Components (standalone or stitched to your Kubernetes 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.

<table>
<thead>
<tr>
<th>Entity</th>
<th>CPU Commodity</th>
<th>vCPU Request</th>
<th>vCPU Limit Quota</th>
<th>vCPU Request Quota</th>
</tr>
</thead>
<tbody>
<tr>
<td>Container</td>
<td>mCore</td>
<td>mCore</td>
<td>mCore</td>
<td>mCore</td>
</tr>
<tr>
<td>Container Spec</td>
<td>mCore</td>
<td>mCore</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Workload Controller</td>
<td>N/A</td>
<td>N/A</td>
<td>mCore</td>
<td>mCore</td>
</tr>
<tr>
<td>Container Pod</td>
<td>mCore</td>
<td>mCore</td>
<td>mCore</td>
<td>mCore</td>
</tr>
<tr>
<td>Namespace</td>
<td>mCore</td>
<td>mCore</td>
<td>mCore</td>
<td>mCore</td>
</tr>
<tr>
<td>Container Cluster</td>
<td>mCore</td>
<td>mCore</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Node (VM)</td>
<td>MHz</td>
<td>mCore</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Application Component</td>
<td>MHz</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

This feature is available starting in version 3.0.5. For customers updating to version 3.0.5 or later:

- This feature does not require you to update your Kubituro image after the update.
- For time series charts, metrics generated after the update are actual mCore values, but pre-update metrics are the same (unconverted) values in MHz displayed in mCore units. This results in unexpected data in charts immediately after the update.

For example:

If vCPU Limit for a Container Spec was resized from 1300 MHz to 1200 MHz before you updated Workload Optimization Manager, data points in charts correctly show these values in MHz.

Immediately after the update:

- When you view the Virtual CPU chart for the Container Spec, Workload Optimization Manager will show a capacity value of 1200 mCores (which is 1200 MHz in reality) for the last data point before the update, and the equivalent value of 500 mCores for the first data point after the update. This gives the impression of a resize down action between the data points, even if no such action was executed.
- Assume Workload Optimization Manager recommends an action to resize VCPU Limit for the Container Spec from 500 to 700 mCores. When you view the details for this action, the utilization chart will show unexpected data.

  - For the actual recommended action, the data point shows current capacity as 1200 mCores, instead of 500 mCores. The new value after executing the action correctly shows as 700 mCores.
  - For the last resize action before the update, the data point shows the same MHz values (1300 and 1200), but in mCore units.
  - One day after the update, a new data point displays in the chart, indicating that capacity was resized from 1200 mCores to 500 mCores, even if no actual resize action was executed.

Over time, data points with unexpected values will begin to fall out of range and newer data points will reflect actual mCore values.

- For increment constants in Container Spec policies, the default value of 100 remains unchanged, but the unit changes from MHz to mCores. This means that each resize action will now increase or decrease capacity by 100 mCores, instead of 100 MHz.
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.

Synopsis
## Synopsis

<table>
<thead>
<tr>
<th>Budget:</th>
<th>A VM gains its budget by selling resources to the applications it hosts.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides:</td>
<td>Resources for hosted applications to use:</td>
</tr>
<tr>
<td></td>
<td>▪ VMEM (Kbytes)</td>
</tr>
<tr>
<td></td>
<td>▪ VCPU (MHz)</td>
</tr>
<tr>
<td></td>
<td>▪ VStorage</td>
</tr>
<tr>
<td></td>
<td>▪ IOPS (storage access operations per second)</td>
</tr>
<tr>
<td></td>
<td>▪ Latency (capacity for disk latency in ms)</td>
</tr>
<tr>
<td></td>
<td>▪ Memory and CPU Requests (for Kubernetes environments)</td>
</tr>
<tr>
<td>Consumes:</td>
<td>Resources from cloud zones</td>
</tr>
<tr>
<td>Discovered through:</td>
<td>Cloud targets</td>
</tr>
</tbody>
</table>

## Monitored Resources

Workload Optimization Manager monitors the following resources for a cloud VM:

- **Virtual Memory (VMem)**
  
  The utilization of the VMem allocated to the hosting VM
  
  Measured in Kilobytes (KB)

- **Virtual CPU (VCPU)**
  
  The utilization of the VCPU allocated to the hosting VM
  
  Measured in Megahertz (MHz)

- **Storage Amount**
  
  The utilization of the datastore’s capacity
  
  Measured in Megabytes (MB)

- **Storage Access Operations Per Second (IOPS)**
  
  The utilization of IOPS allocated for the VStorage on the VM
  
  Measured in IOPS

- **Net Throughput**
  
  Rate of message delivery over a port
  
  Measured in KB/s

- **I/O Throughput**
  
  The throughput to the underlying storage for the entity
  
  Measured in KB/s

- **Latency**
  
  The utilization of latency allocated for the VStorage on the VM
  
  Measured in milliseconds (ms)

## Cloud VM Actions

Workload Optimization Manager can recommend the following actions for a cloud VM:

- **Scale**
- **Increase RI Coverage**
- **Buy RI**
Scale Actions for Cloud VMs

For scale actions, you change the VM instance to use a different instance type or tier. You can choose Cloud Scale All, Cloud Scale for Performance, or Cloud Scale for Savings.

- You can direct Workload Optimization Manager to only execute cloud VM scaling actions that improve performance (Cloud Scale for Performance) or reduce costs (Cloud Scale for Savings). The default action 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.
- Cloud Scale All enables all scaling actions, including those that result in efficiency improvements and increased costs.

**NOTE:**
The Move/Compute Scale action available in version 7.22.5 or earlier has been separated into two actions starting in version 7.22.6 – Move (for on-prem VMs) and Cloud Compute Scale (for cloud VMs). In version 8.0.5, Cloud Compute Scale has been renamed Cloud Scale All.

If you disabled Move/Compute Scale and then updated to 7.22.6 or later, only Move actions are disabled. To apply the same action to cloud VMs, create policies for the affected VMs and then disable Cloud Scale All.

- When policy conflicts arise, Cloud Scale All overrides the other two scaling options in most cases. For more information, see Relationship Between Scoped and Default Policies (on page 93).

Cloud VMs with Failed Sizing

For workload on the public cloud, if Workload Optimization Manager tries to execute a scale action but the action fails, then Workload Optimization Manager places the affected VM in a special group named Cloud VMs with Failed Sizing. Under normal circumstances this group will be empty. But in case some actions have failed, you can review the contents of this group to inspect the individual VMs. As soon as Workload Optimization Manager successfully executes a scale action on a VM in this group, it then removes the VM from the group.

**NOTE:**
When Workload Optimization Manager places a VM in this group, it restarts the VM to ensure that it is running correctly with its original configuration.

By default Workload Optimization Manager does not include any action policies for this group. Whatever action mode is set to the given VMs remains in effect while the VMs are in this group. You can create a policy and scope the policy to this group. For example, assume you see typical failures for actions that Workload Optimization Manager tries to execute during working hours. In that case, you can create a scheduling window that enables scale actions during off hours. That can help to automatically execute the actions and remove the VMs from this group.

Note that the VMs in this group could already be in a scope that is affected by another actions policy. Remember that with competing policies, the most conservative policy wins. When working with the Cloud VMs with Failed Sizing group, this can have unintended consequences. Assume you have VMs with automated scale actions, and you create a policy that sets the action mode to Manual for this group. Assume a failed scale action places a VM into this group. In that case the more conservative action mode takes effect, and the VM will use Manual mode. Because of a failed scale action, the VM does not automate subsequent scale actions.

Actions for Kubernetes Nodes (VMs)

For a VM used as a Kubernetes node, Workload Optimization Manager can recommend the following actions:

- **Provision**
  Workload Optimization Manager can recommend provisioning nodes to address workload congestion or meet application demand. This action could lead to a related action to scale the volumes for the nodes to improve I/O throughput.
- **Suspend**
- **Workload Optimization Manager** can recommend suspending nodes after you have consolidated pods or defragmented node resources to improve infrastructure efficiency.

When recommending provision actions, Workload Optimization Manager considers demand from DaemonSets to ensure 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. Workload Optimization Manager represents number of consumers with DaemonSets and respects the maximum number of pods allowed for that node.
When recommending suspend actions, Workload Optimization Manager will also recommend suspending the related DaemonSets.

For nodes in the public cloud, Workload Optimization Manager reports the cost savings or investments attached to these 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 Reserved Instances 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 cluster (on page 149) or the global cloud environment to view aggregated cost information.

**AWS VMs**

### AWS Instance Requirements

In AWS 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 sets the action to Recommend Only, and describes the reason. Workload Optimization Manager will not automate the move, even if you have set the action mode for that scope to Automatic. You can execute the move manually, after you have properly configured the instance.

Note that if you have workloads 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 workloads, and then create a placement policy for that scope. In the policy, **Excluded Templates** to exclude the instance types that do require ENA support. For information about placement policies, see Automation Policies (on page 92). For information about excluding instance types, see Cloud Instance Types (on page 164).

The instance requirements that Workload Optimization Manager recognizes are:

- **Enhanced Network Adapters**
  
  Some workloads 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 moving a workload that does not support ENA onto an instance that does. To make that move, you must perform the required configuration of the workload before you can execute the move. If you move a non-ENA VM to an instance that requires ENA, then AWS cannot start up the VM after the move. Before executing the move, you must enable ENA on the VM.

  For information about ENA configuration, see “Enabling Enhanced Networking with the Elastic Network Adapter (ENA) on Windows Instances” in the AWS documentation.

- **Linux AMI Virtualization Type**
  
  An Amazon Linux AMI can use ParaVirtual (PV) or Hardware Virtual Machine (HVM) virtualization. Workload Optimization Manager can recommend moving an instance type 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 workload. Workload Optimization Manager can recommend moving a 32-bit workload to an instance that only supports a 64-bit platform.

- **NVMe Block**
  
  Some instances expose EBS volumes as NVMe block devices, but not all workloads are configured with NVMe drivers. Workload Optimization Manager can recommend moving such a workload to an instance that supports NVMe. Before executing the move, you must install the NVMe drivers on the workload.

In addition, Workload Optimization Manager recognizes processor types that you currently use for your workloads. For move or resize actions, Workload Optimization Manager keeps your workloads on instance types with compatible processors:

- **GPU-based instances:**
  
  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.
• ARM-based instances
  If your workload is on an ARM-based instance, then Workload Optimization Manager will only recommend moves and resizes to other compatible ARM-based instance types.

**Resizing Storage Capacity in AWS Environments**

When a VM needs more storage capacity, Workload Optimization Manager recommends actions to move the it 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 storage actions.

If the root storage for your workload is Instance Storage, then Workload Optimization Manager will not recommend a storage action. This is because Instance Storage is ephemeral, and such an action would cause the workload to lose all the stored data.

If the root storage is EBS, then Workload Optimization Manager recommends storage actions. EBS is persistent, and the data will remain after the action. However, if the workload uses Instance Storage for extra storage, then Workload Optimization Manager does not include that storage in its calculations or actions.

**Azure VMs**

**Azure Resource Group Discovery**

To discover Azure Resource Groups, you can set up the following targets:

• Microsoft Azure service principle targets
• Microsoft Azure Enterprise Agreement (EA) targets

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 328)](on page 328).

**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.
Performance Metrics for Azure VMs
To analyze the utilization of Azure VM resources, Workload Optimization Manager collects performance metrics (such as memory and CPU usage) from Azure periodically. It collects metrics in the following ways:
- Azure storage account
- For VMs with basic diagnostics enabled, Workload Optimization Manager collects performance metrics that Azure publishes via this storage account.
- Azure Monitor Log Analytics

Rather than enabling diagnostics on a per-VM basis, you may have created Azure Monitor Log Analytics workspaces to centralize the management of your Azure VM configurations. Workload Optimization Manager discovers these workspaces when you add Azure targets, and then retrieves performance metrics periodically.

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 162).

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.

Cloud VM Policies
Workload Optimization Manager ships with default settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.

Action Automation and Orchestration
For details about cloud VM actions, see Cloud VM Actions (on page 157).

Cloud Scale
Change the VM instance to use a different instance type or tier. You can choose Cloud Scale All, Cloud Scale for Performance, or Cloud Scale for Savings.

<table>
<thead>
<tr>
<th>Action</th>
<th>Default Mode</th>
<th>AWS</th>
<th>Azure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud Scale All</td>
<td>Manual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cloud Scale for Perfor-</td>
<td>Manual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cloud Scale for Savings</td>
<td>Manual</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Other Actions

<table>
<thead>
<tr>
<th>Action</th>
<th>Default Mode</th>
<th>AWS</th>
<th>Azure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buy RI</td>
<td>Recommend</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Entity Types - Cloud Infrastructure

<table>
<thead>
<tr>
<th>Action</th>
<th>Default Mode</th>
<th>AWS</th>
<th>Azure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provision</td>
<td>Recommend</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Start</td>
<td>Recommend</td>
<td></td>
<td>Auto</td>
</tr>
<tr>
<td>Reconfigure (based on custom policies)</td>
<td>Recommend</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suspend</td>
<td>Recommend</td>
<td></td>
<td>Auto</td>
</tr>
</tbody>
</table>

### Scaling Target Utilization

**For VCPU, VMEM, and IO/Net Throughput Utilization:**

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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scaling Target VCPU Utilization</td>
<td>70</td>
</tr>
<tr>
<td>The target utilization as a percentage of VCPU capacity.</td>
<td></td>
</tr>
<tr>
<td>Scaling Target VMEM Utilization</td>
<td>90</td>
</tr>
<tr>
<td>The target utilization as a percentage of memory capacity.</td>
<td></td>
</tr>
<tr>
<td>Scaling Target IO Throughput Utilization</td>
<td>90</td>
</tr>
<tr>
<td>The target utilization as a percentage of IO throughput capacity.</td>
<td></td>
</tr>
<tr>
<td>Scaling Target Net Throughput Utilization</td>
<td>90</td>
</tr>
<tr>
<td>The target utilization as a percentage of network throughput capacity.</td>
<td></td>
</tr>
</tbody>
</table>

**For IOPS Utilization:**

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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scaling Target IOPS Utilization (Azure VMs only)</td>
<td>70</td>
</tr>
<tr>
<td>For Azure environments, the target percentile value Workload Optimization Manager will attempt to match.</td>
<td></td>
</tr>
</tbody>
</table>

For details on how IOPS utilization affects scaling decisions, see [IOPS-aware Scaling for Azure VMs (on page 161)](on page 161).

### Aggressiveness and Observation Periods

Workload Optimization Manager uses these settings to calculate utilization percentiles for vCPU, VMEM, and IOPS (Azure VMs only). It then recommends actions to improve utilization based on the observed values for a given time period.

- **Aggressiveness**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggressiveness</td>
<td>95th Percentile</td>
</tr>
</tbody>
</table>

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 above 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**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max Observation Period</td>
<td>Last 30 Days</td>
</tr>
</tbody>
</table>

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**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min Observation Period</td>
<td>None</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud Instance Types</td>
<td>None</td>
</tr>
</tbody>
</table>

By default, Workload Optimization Manager considers all instance types currently available for scaling when making scaling decisions for VMs. However, you may have set up your cloud VMs to only scale to or avoid certain instance types to reduce complexity and cost, improve RI utilization, or meet application demand. Use this setting to identify the instance types that VMs can scale to.

Click **Edit** to set your preferences. In the new page that displays, expand a **cloud tier** (a family of instance types, such as `a1` for AWS or `B-series` for Azure) to see individual instance types and the resources allocated to them. If you have several cloud providers, each provider will have its own tab.

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.

### Consistent Resizing

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enable Consistent Resizing</td>
<td>Off</td>
</tr>
</tbody>
</table>

**For groups in scoped 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.

**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 creates a policy that identifies compute tiers to exclude from that set.
  
  - For Scale Sets, Workload Optimization Manager automatically enables Consistent Resizing across all the VMs in the 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 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.

Reasons to employ Consistent Resizing for a group include:

- **Load Balancing**
  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.

- **High Availability (HA)**
  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.

### Instance Store Aware Scaling

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instance Store Aware Scaling</td>
<td>Off</td>
</tr>
</tbody>
</table>

For AWS environments:

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.

### Database Server (Cloud)

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.

**NOTE:**

Azure SQL Databases discovered by Workload Optimization Manager appear as *Database* entities in the supply chain. For details, see [Database (Cloud)](on page 175).
**Synopsis**

<table>
<thead>
<tr>
<th>Budget:</th>
<th>A cloud Database Server has unlimited budget.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides:</td>
<td>VCPU, VMem</td>
</tr>
<tr>
<td>Consumes:</td>
<td>Resources in the availability zone</td>
</tr>
<tr>
<td>Discovered through:</td>
<td>AWS targets</td>
</tr>
</tbody>
</table>

**Monitored Resources**

Workload Optimization Manager monitors the following resources for a cloud Database Server:

- **Virtual Memory (VMem)**
  - The utilization of the VMem allocated to the hosting VM
  - Measured in Kilobytes (KB)

- **Virtual CPU (VCPU)**
  - The utilization of the VCPU allocated to the hosting VM
  - Measured in Megahertz (MHz)

**Actions**

Actions for cloud Database Servers will be introduced in a future product release.

**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

| Budget: | A cloud volume gains its budget by selling resources to the VMs that it serves. |
| Provides: | Storage resources for VMs to use: |
|           | - Storage Access |
|           | - Storage Amount |
|           | - IO Throughput |
| Consumes: | Storage services provided by Zones or Regions |

- **Budget:** A cloud volume gains its budget by selling resources to the VMs that it serves.
- **Provides:** Storage resources for VMs to use:
  - Storage Access
  - Storage Amount
  - IO Throughput
- **Consumes:** Storage services provided by Zones or Regions
Entity Types - Cloud Infrastructure

<table>
<thead>
<tr>
<th>Synopsis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovered through: AWS and Azure targets</td>
</tr>
</tbody>
</table>

**Monitored Resources**

Workload Optimization Manager monitors the following resources for a cloud volume:

- **Storage Access**
  The percentage of the volume’s capacity for storage access operations (measured in IOPS) that is in use.

- **IO Throughput**
  The percentage of the volume’s capacity for IO throughput (measured in MB/s) that is in use.

**NOTE:**
Workload Optimization Manager discovers Storage Amount (disk size) for AWS/Azure volumes, but does not monitor utilization. For a Kubeturbo (container) deployment that includes AWS/Azure volumes, Kubeturbo monitors Storage Amount utilization for the volumes. You can view utilization information in the Capacity and Usage chart.

**Cloud Volume Actions**

Workload Optimization Manager can recommend the following actions for a cloud volume:

- **Delete**
- **Scale**

**Delete Actions**

Delete unattached volumes as a cost-saving measure.

**NOTE:**
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.

**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 **RSVaultBackup** 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 the **Delete** action in that policy.

**Scale Actions**

Scale attached volumes to optimize performance, address overprovisioning, or reduce 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 171).
• 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.

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 174).

**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:
This information helps you decide whether to take the action.

Workload Optimization Manager polls your cloud volumes every 6 hours, and then records their state (attached or unattached) at the time of polling. It does not account for changes in state between polls. An unattached period of 1 day means that Workload Optimization Manager has discovered the volume as unattached in the last four polls.

Once Workload Optimization Manager discovers an unattached volume, it immediately recommends that you delete it. If a currently unattached volume is not deleted and is subsequently discovered as attached, Workload Optimization Manager removes the Delete action attached to it, and then resets the unattached period.

NOTE:
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.

To see the last known VM attached to the volume, click DETAILS.

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 173)*.

**Non-disruptive and Reversible Scaling Actions**
Workload Optimization Manager indicates whether a pending action is non-disruptive or reversible.
Non-disruptive
Executing storage scaling actions can sometimes be disruptive if the VM must be rebooted to execute a storage change. For example, Azure Standard and Premium scaling actions are disruptive. When a storage action is disruptive, expect a single reboot (usually 2–3 minutes of downtime).

The following scaling actions are non-disruptive:
- Scaling IOPS and throughput on Azure Ultra storage
- All scaling actions on AWS storage

Reversible
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.
Cloud Volume Policies

Workload Optimization Manager ships with default settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.

Action Automation and Orchestration

For details about cloud volume actions, see Cloud Volume Actions (on page 168).

<table>
<thead>
<tr>
<th>Action</th>
<th>Default Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale</td>
<td>Manual</td>
</tr>
<tr>
<td>Delete</td>
<td>Manual</td>
</tr>
</tbody>
</table>

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.

- **Aggressiveness**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggressiveness</td>
<td>95th Percentile</td>
</tr>
</tbody>
</table>

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 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 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**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max Observation Period</td>
<td>Last 30 Days</td>
</tr>
</tbody>
</table>

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**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min Observation Period</td>
<td>None</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scaling Target IOPS/Throughput Utilization</td>
<td>70</td>
</tr>
</tbody>
</table>

**Cloud Storage Tiers**

By default, Workload Optimization Manager considers all storage tiers currently available for scaling when making scaling decisions for volumes. However, you may have set up your cloud volumes to only scale to or avoid certain tiers to reduce complexity and cost, or meet demand. Use this setting to identify the tiers that volumes can scale to.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud Storage Tiers</td>
<td>None</td>
</tr>
</tbody>
</table>

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.
Database (Cloud)

A Database is a fully managed database engine running in Azure environments. Workload Optimization Manager discovers SQL Databases through your Azure targets. In particular, it discovers the single database resource type that uses the DTU (Database Transaction Unit) purchase model. With this model, Azure bundles CPU, memory, and IOPS as a single DTU commodity. Workload Optimization Manager monitors DTU and storage utilization for these databases, and drives scaling actions to improve utilization and reduce costs.

NOTE:
For more information about DTUs, see the Azure documentation.

AWS RDS databases appear as Database Server entities in the supply chain. For details, see Database Server (Cloud) (on page 165).
## Synopsis

<table>
<thead>
<tr>
<th><strong>Budget:</strong></th>
<th>A database has unlimited budget.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Provides:</strong></td>
<td>Transactions to end users</td>
</tr>
<tr>
<td><strong>Consumes:</strong></td>
<td>DTU and storage resources in the Azure region</td>
</tr>
<tr>
<td><strong>Discovered through:</strong></td>
<td>Azure targets</td>
</tr>
</tbody>
</table>
Monitored Resources

Workload Optimization Manager monitors the following resources for a cloud database:

- **DTU**
  
  DTU capacity for the database. DTU represents CPU, memory, and IOPS bundled as a single commodity.

- **Storage**
  
  Storage capacity for the database.

Workload Optimization Manager drives scaling actions based on DTU and storage utilization, and treats the following limits as constraints when making scaling decisions:

- **Maximum concurrent sessions**
  
  Maximum number of database connections at a time.

- **Maximum concurrent workers**
  
  Maximum number of database processes that can handle queries at a time.

Cloud Database Actions

Scale Actions

Workload Optimization Manager can recommend scale actions for a cloud database.

A single action can scale both DTU 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.

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.

![Necessary Investments](image1)

*13 Databases (@31kqpx, 50mns)*

$15 Monthly

$15/mo, 1 Scale Database

![Potential Savings](image2)

*13 Databases (@31kqpx, 50mns)*

$225 Monthly

$225/mo, 4 Scale Databases

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

Utilization Charts for Scale Actions

Workload Optimization Manager uses percentile calculations to measure DTU and storage utilization 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 database, you will see charts that highlight DTU and storage 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 179).

Cloud Database Policies
Workload Optimization Manager ships with default settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.

Action Automation and Orchestration
For details about cloud database actions, see Cloud Database Actions (on page 177).

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scaling Target DTU Utilization</td>
<td>70</td>
</tr>
</tbody>
</table>
Storage Amount 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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage Amount Utilization (%)</td>
<td>90</td>
</tr>
</tbody>
</table>

For example, a value of 90 means that Workload Optimization Manager considers 90% utilization to be 100% of capacity. Workload Optimization Manager recommends actions that avoid utilization beyond the given value.

Cloud Instance Types

By default, Workload Optimization Manager considers all instance types currently available for scaling when making scaling decisions for databases. However, you may have set up your cloud databases to only scale to or avoid certain instance types to reduce complexity and cost, or meet demand. Use this setting to identify the instance types that databases can scale to.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud Instance Types</td>
<td>None</td>
</tr>
</tbody>
</table>

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 and the resources allocated to them.

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.

NOTE:
This policy setting is not available in plans.

Aggressiveness and Observation Period

Workload Optimization Manager uses these settings to calculate utilization percentiles for DTU and storage. It then recommends actions to improve utilization based on the observed values for a given time period.

- **Aggressiveness**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggressiveness</td>
<td>95th Percentile</td>
</tr>
</tbody>
</table>

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 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 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

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max Observation Period</td>
<td>Last 14 Days</td>
</tr>
</tbody>
</table>

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](#).

## 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**

- **Budget:** Workload Optimization Manager assumes a Zone has infinite resources.
- **Provides:** Compute and storage resources to VMs.
- **Consumes:** Region resources.
- **Discovered through:** Workload Optimization Manager discovers Zones through public cloud targets.

**Monitored Resources**

For public cloud environments, Workload Optimization Manager discovers the resources that an availability zone provides, including:

- **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.
- **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 Reserved Instance billing.

**Workload Optimization Manager** monitors the following resources for a Zone:

- **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**

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**

**Budget:** Workload Optimization Manager assumes a Region has infinite resources.

**Provides:** Hosting and storage resources to Zones.

**Consumes:** NA

**Discovered through:** Cloud service accounts, such as accounts on Amazon AWS, or subscriptions on Microsoft Azure.

---

### 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**

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.
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**

<table>
<thead>
<tr>
<th>Provides:</th>
<th>Resources for hosted applications to use:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• VMEM (Kbytes)</td>
</tr>
<tr>
<td></td>
<td>• VCPU (MHz)</td>
</tr>
<tr>
<td></td>
<td>• VStorage</td>
</tr>
<tr>
<td></td>
<td>• IOPS (storage access operations per second)</td>
</tr>
<tr>
<td></td>
<td>• Latency (capacity for disk latency in ms)</td>
</tr>
<tr>
<td></td>
<td>• Memory and CPU Requests (for Kubernetes environments)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consumes:</th>
<th>Physical host resources, including CPU and Mem</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Storage</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Discovered through:</th>
<th>Hypervisor targets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>When Workload Optimization Manager discovers VMs in a Pivotal Operations Manager environment, it assigns VM names that identify the VM in the context of your Pivotal environment. The name is expressed in the following tokens:</td>
</tr>
<tr>
<td></td>
<td>{PCF Job name}#{index number}#{deployment ID}#{IaaS VM name}</td>
</tr>
</tbody>
</table>

**Monitored Resources**

Workload Optimization Manager monitors the following resources for a VM:

- **Virtual Memory (VMem)**
  - The utilization of the VMem allocated to the hosting VM
  - Measured in Kilobytes (KB)

- **Virtual CPU (VCPU)**
  - The utilization of the VCPU allocated to the hosting VM
  - Measured in Megahertz (MHz)

- **Virtual Storage (VStorage)**
  - The utilization of the virtual storage capacity allocated for the VM
  - Measured in Kilobytes (KB)

- **Storage Access Operations Per Second (IOPS)**
  - The utilization of IOPS allocated for the VStorage on the VM
  - Measured in IOPS

- **Latency**
  - The utilization of latency allocated for the VStorage on the VM
  - Measured in milliseconds (ms)

For VMs that host Kubernetes pods:

- **Memory Request Allocation**
  - The memory available to the VM to support the ResourceQuota request parameter for a given VDC (Kubernetes namespace).

- **CPU Request Allocation**
  - The CPU available to the VM to support the ResourceQuota request parameter for a given VDC (Kubernetes namespace).

- **Virtual Memory Request**
  - The memory currently requested by containers. The capacity for this resource is the Node Allocatable capacity (the amount of resources available for pods).

- **Virtual CPU Request**
The CPU currently requested by containers. The capacity for this resource is the Node Allocatable capacity (the amount of resources available for pods).

- **MemAllocation**
  The memory ResourceQuota limit parameter for a given VDC (Kubernetes namespace).

- **CPUAllocation**
  The CPU ResourceQuota limit parameter for a given VDC (Kubernetes namespace).

### On-prem VM Actions

#### Overview of Actions

For on-prem VMs, actions can change resources in the following ways:

- **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.

- **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.

#### Supported Actions

Workload Optimization Manager can recommend the following actions for a VM:

- **Resize Up**
  - High resource utilization on VM

- **Resize Down**
  - Low resource utilization on VM that must not shut down

- **Move VM due to:**
  - High resource utilization on VM
  - High resource utilization on hosting PM
  - Excess IOPS or Latency in VStorage
  - Workload placement violation
  - Hosting PM is underutilized (move before suspending PM)

- **Move VM Storage**
  For excess utilization of the current datastore, or for more efficient utilization of datastores in the environment.

- **Reconfigure Storage**
  For overutilized storage resources, add VStorage capacity.
  For underutilized storage resources, remove VStorage capacity.

- **Reconfigure**
  Change network and storage configuration. For example, Workload Optimization Manager recommends this action if the VM is configured to use a network that it cannot access.

#### Tuned Scaling for On-prem VMs

For resizing VMs and Storage, Workload Optimization Manager includes tuned scaling action settings. These settings give 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 post more conservative actions (Manual or Recommend) when the issue lies outside of the scaling 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.

To configure tuned scaling, create a VM or Storage policy (see Creating Automation Policies (on page 96)). Under Action Automation, configure the action mode for the various resize actions, which are listed in the table below for your reference. Note that 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. Finally, under Operational Constraint, specify the tuned scaling range.

<table>
<thead>
<tr>
<th>Entity Type</th>
<th>Resize Actions</th>
<th>Operational Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>VM</td>
<td>▪ VCPU Resize Up</td>
<td>▪ VCPU Max Size</td>
</tr>
<tr>
<td></td>
<td>▪ VCPU Resize Down</td>
<td>▪ VCPU Min Size</td>
</tr>
<tr>
<td></td>
<td>▪ VCPU Resize Above Max</td>
<td>▪ VMEM Max Size</td>
</tr>
<tr>
<td></td>
<td>▪ VCPU Resize Below Min</td>
<td>▪ VMEM Min Size</td>
</tr>
<tr>
<td></td>
<td>▪ VMEM Resize Up</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ VMEM Resize Down</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ VMEM Resize Above Max</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ VMEM Resize Below Min</td>
<td></td>
</tr>
<tr>
<td>Storage</td>
<td>▪ Storage Resize Up</td>
<td>▪ Storage Max Size</td>
</tr>
<tr>
<td></td>
<td>▪ Storage Resize Down</td>
<td>▪ Storage Min Size</td>
</tr>
<tr>
<td></td>
<td>▪ Storage Resize Above Max</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ Storage Resize Below Min</td>
<td></td>
</tr>
</tbody>
</table>

For example, assume the following settings:
With this policy in effect, Workload Optimization Manager will post the following actions:

- A VM with 6 VCPUs requests 2 new VCPUs: Automatic
- A VM with 8 VCPUs requests 2 new VCPUs: Manual
- A VM with 2 VCPUs requests to resize down to 1 VCPU: Disabled (Workload Optimization Manager does not post the action)

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 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 Policy Scope (on page 104).

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 105).

**Actions for Kubernetes Nodes (VMs)**

For a VM used as a Kubernetes node, Workload Optimization Manager can recommend the following actions:

- Provision
  
  Workload Optimization Manager can recommend provisioning nodes to address workload congestion or meet application demand. This action could lead to a related action to scale the volumes for the nodes to improve I/O throughput.
- Suspend
- Workload Optimization Manager can recommend suspending nodes after you have consolidated pods or defragmented node resources to improve infrastructure efficiency.

When recommending provision actions, Workload Optimization Manager considers demand from DaemonSets to ensure 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. Workload Optimization Manager represents number of consumers with DaemonSets and respects the maximum number of pods allowed for that node.

When recommending suspend actions, Workload Optimization Manager will also recommend suspending the related DaemonSets.

**On-prem VM Policies**

Workload Optimization Manager ships with default settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.

**Action Automation and Orchestration**

For details about on-prem VM actions, see [On-prem VM Actions](on page 189). You can use [Action Scripts](on page 111) 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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enforce Non-disruptive Mode</td>
<td>Off</td>
</tr>
</tbody>
</table>

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.

**Hypervisor VMEM for Resize**

For on-prem environments, Workload Optimization Manager discovers VMEM utilization and can recommend actions to resize the VMEM capacity on a VM. For environments that do not include Applications and Databases as targets, the data that analysis uses to make these recommendations comes from the underlying hypervisors. However, that data is not always sufficient to result in accurate resize recommendations. Use the **Use Hypervisor VMEM for Resize** setting to determine how to generate VMEM recommendations.
<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypervisor VMEM for Resize</td>
<td>On</td>
</tr>
</tbody>
</table>

- **On**
  When your environment includes Applications and Databases as targets, Workload Optimization Manager uses the VMEM metrics those targets discover. If a scope of VMs does not fall under those targets, then analysis *will* generate VMEM resize actions for that scope. In this case, analysis uses the VMEM metrics it discovers from the underlying hypervisors.

- **Off**
  When your environment includes Applications and Databases as targets, Workload Optimization Manager uses the VMEM metrics those targets discover. If a scope of VMs does not fall under those targets, then analysis *will not* generate VMEM resize actions for that scope.

### 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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shared-nothing Migration</td>
<td>Off</td>
</tr>
</tbody>
</table>

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 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 (Operational Constraints)

Workload Optimization Manager uses these settings to set up tuned scaling actions for the VM. For an overview of tuned scaling, see [Tuned Scaling for On-prem VMs](on page 189).

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>VCPU Resize Max Threshold (in Cores)</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>Tuned Scaling Range Upper Limit</td>
</tr>
<tr>
<td>VCPU Resize Min Threshold (in Cores)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Tuned Scaling Range Lower Limit</td>
</tr>
<tr>
<td>VMEM Resize Max Threshold (MB)</td>
<td>131072</td>
</tr>
<tr>
<td></td>
<td>Tuned Scaling Range Upper Limit</td>
</tr>
<tr>
<td>VMEM Resize Min Threshold (MB)</td>
<td>512</td>
</tr>
<tr>
<td></td>
<td>Tuned Scaling Range Lower Limit</td>
</tr>
</tbody>
</table>

### 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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting/Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resize VStorage</td>
<td>Disabled</td>
</tr>
<tr>
<td>Increment constant for VStorage [GB]</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>If you enable resize, Workload Optimization Manager uses the default value of 1024. You can change this to a different value.</td>
</tr>
</tbody>
</table>

### Resize Increments

These increments specify how many units to add or subtract when resizing the given resource allocation for a VM.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increment constant for VMEM [MB]</td>
<td>1024</td>
</tr>
<tr>
<td>Increment constant for VCPU [MHz]</td>
<td>1800</td>
</tr>
</tbody>
</table>

For resize increments, you should consider the following:

- For VMem, you should not set the increment value to be lower 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.
- For VCPU, the increment affects resize of VCPU limits and reservations in MHz, and it also affects the addition/removal of cores for VCPU capacity on a VM.
For limits and reservations, Workload Optimization Manager recommends changes in terms of the specified resize increment. For example, assume the increment is 1800 MHz and you have reserved 3000 MHz for a VM. Workload Optimization Manager could recommend to reduce the reservation by 1800, down to 1200 MHz.

For VCPUs, Workload Optimization Manager can only resize allocation one core at a time. This means a resize is to the nearest core count that matches or exceeds the resize increment. Assume the cores all have a clock speed of 2000 MHz. If the resize increment is 1800 MHz, then a resize up will recommend to add one more core at 2000 MHz.

**Rate of Resize**

When resizing resources for a VM, Workload Optimization Manager calculates the optimal values for VMem, VCPU and VStorage. 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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate of Resize</td>
<td>Medium (2)</td>
</tr>
</tbody>
</table>

- **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**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enable Consistent Resizing</td>
<td>Off</td>
</tr>
</tbody>
</table>

For groups in scoped 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.

**NOTE:**

For consistent resizing in on-prem environments, 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. If you automate resizes, Workload Optimization Manager executes each resize individually in a way that avoids disruption to your workloads.

**Reasons to employ Consistent Resizing for a group include:**

- **Load Balancing**
  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.
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.

Ignore NVMe Constraints

Workload Optimization Manager recognizes when a VM instance includes an NVMe driver. To respect NVMe constraints, it will not recommend a move or resize to an instance type that does not also include an NVMe driver. If you ignore NVMe constraints, then Workload Optimization Manager is free to resize or move the instance to a type that does not include an NVMe driver.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ignore NVMe Constraints</td>
<td>Off</td>
</tr>
</tbody>
</table>

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 87).

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 87).

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

| Budget: | On-prem Database Servers have unlimited budget. |
| 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 MSSQL processes  
• NewRelic Infrastructure Integration (NRI): MySql, MsSql, MongoDB, OracleDB |
Monitored Resources

Workload Optimization Manager monitors the following resources for an on-prem Database Server:

- **Virtual Memory (VMem)**
  The utilization of the VMem allocated to the hosting VM
  Measured in Kilobytes (KB)
- **Virtual CPU (VCPU)**
  The utilization of the VCPU allocated to the hosting VM
  Measured in Megahertz (MHz)
- **Transactions**
  The utilization of the allocated transactions per second for the given entity
  Measured in transactions per second
- **Heap**
  The utilization of the application server’s heap
  Measured in Kilobytes (KB)
- **DBMem**
  The memory utilized by the database, as a of the memory capacity that is allocated to the database. Note that this resource is more accurate than the VMem resource on the hosting VM. With this resource, Workload Optimization Manager can drive resize and move actions based on the memory consumed by the database, not the memory consumed by the VM.
  The utilization of the allocated transactions per second for the given virtual application
  Measured in transactions per second
- **Connections**
  The utilization of the connection capacity. Only applicable to database servers
  Measured in Connections
- **Remaining Garbage Collection Capacity**
  The percentage of server uptime spent not performing garbage collection
  Measured in uptime (%)
- **Threads**
  The utilization of the server’s thread capacity
  Measured in number of Threads
- **Transaction Log**
  The utilization of the server’s capacity for storage devoted to transaction logs
  Measured in Kilobytes (KB)
- **DB Cache Hit Rate**
  The percentage of accesses that result in cache hits.
  Measured as a percentage of hits vs total attempts (%)
NOTE:
Resize actions based on the Transaction Log resource depend on support for vStorage in the underlying hypervisor technology. Because current versions of Hyper-V do not provide API support for vStorage, Workload Optimization Manager cannot support Transaction Log resize actions for database servers running on the Hyper-V platform.

On-prem Database Server Policies
Workload Optimization Manager ships with default settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.

Action Automation and Orchestration

<table>
<thead>
<tr>
<th>Action</th>
<th>Default Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resize</td>
<td>Manual</td>
</tr>
<tr>
<td>Resize DBMem (Up/Down)</td>
<td>Manual</td>
</tr>
</tbody>
</table>

You can use Action Scripts (on page 111) and third-party orchestrators (such as ServiceNow) for action orchestration.

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%.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting/Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enable Transaction SLO</td>
<td>Off</td>
</tr>
<tr>
<td>Transaction SLO</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>If you enable SLO, Workload Optimization Manager uses the default value of 10. You can change this to a different value.</td>
</tr>
</tbody>
</table>

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%.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting/Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enable Response Time SLO</td>
<td>Off</td>
</tr>
<tr>
<td></td>
<td>Workload Optimization Manager estimates SLO based on monitored values.</td>
</tr>
<tr>
<td>Response Time SLO [ms]</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>If you enable SLO, Workload Optimization Manager uses the default value of 2000. You can change this to a different value.</td>
</tr>
</tbody>
</table>

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.


<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBMem Utilization (%)</td>
<td>100</td>
</tr>
</tbody>
</table>

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.

**DBMem Scaling Increment**

This increment specifies how many units to add or subtract when scaling DBMem.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBMem Scaling Increment (MB)</td>
<td>128</td>
</tr>
</tbody>
</table>

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.

**Volume (On-prem)**

A 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**

| Budget: | An on-prem volume gains its budget by selling resources to the VMs that it serves. |
| 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 can recommend move actions for an on-prem volume. To evaluate and execute these actions, set the scope to the VM to which a volume is attached.

**On-prem Volume Policies**

Workload Optimization Manager ships with default settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.

**Action Automation and Orchestration**

<table>
<thead>
<tr>
<th>Action</th>
<th>Default Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Move</td>
<td>Manual</td>
</tr>
</tbody>
</table>

**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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud Storage Tiers</td>
<td>None</td>
</tr>
</tbody>
</table>

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 Datacenter (Private Cloud)**

A virtual datacenter (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 Datacenters.
Provider Virtual Datacenters

A provider virtual datacenter (vDC) is a collection of physical resources (hosts and datastores) within a cloud stack. The cloud administrator has access to these resources, and defines the datacenter members. A Provider vDC is created to manage resources that will be allocated to external customers through one or more Consumer vDCs.

Synopsis

<table>
<thead>
<tr>
<th>Budget:</th>
<th>A Provider vDC gains its budget by selling resources to the Consumer vDCs that it hosts. If utilization falls off, the datacenter loses budget. Ultimately, if the budget isn’t enough to pay for the services it consumes, Workload Optimization Manager will recommend decommisioning the Provider vDC.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides:</td>
<td>Physical resources such as hosts and datastores to Consumer vDCs.</td>
</tr>
<tr>
<td>Consumes:</td>
<td>Hosts and datastores from the physical infrastructure</td>
</tr>
<tr>
<td>Discovered through:</td>
<td>Workload Optimization Manager discovers vDCs through private cloud stack managers such as vCloud Director.</td>
</tr>
</tbody>
</table>
Monitored Resources

Workload Optimization Manager monitors the following resources for a Provider vDC:

- **Memory (Mem)**
  The utilization of the Datacenter’s memory reserved or in use
  Measured in Kilobytes (KB)

- **CPU**
  The utilization of the Datacenter’s CPU reserved or in use
  Measured in Megahertz (MHz)

- **Storage**
  The utilization of the storage attached to the Provider vDC.
  Measured in Kilobytes (KB)

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.

Consumer Virtual Datacenters

A Consumer Virtual Datacenter (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 Datacenters use the resources supplied by a Provider Datacenter.
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.

Provides: Resources to host virtual systems.

Consumes: Provider vDC

Discovered through: Workload Optimization Manager discovers vDCs through cloud stack managers such as vCloud Director.

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 for a Consumer vDC:

- **Memory (Mem)**
  - The utilization of the Datacenter’s memory reserved or in use
  - Measured in Kilobytes (KB)

- **CPU**
  - The utilization of the Datacenter’s CPU reserved or in use
  - Measured in Megahertz (MHz)

- **Storage**
  - The utilization of the storage attached to the Consumer vDC.
  - Measured in Kilobytes (KB)

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.

Virtual Datacenter (Container Orchestration)

A virtual datacenter (vDC) is a collection or pool of resources that groups the resources around specific requirements or business needs. These vDCs can implement boundaries for the cloud infrastructure, and then can establish tenant groups on that infrastructure.

Container Orchestration platforms like Kubernetes or Cloud Foundry use logical pools of resources to manage scheduling of workload. For example, administrators can pool resources for different organizations within the enterprise, and assign different policies to each pool. Workload Optimization Manager represents these pools as Virtual Datacenters (vDCs).
Synopsis

Workload Optimization Manager creates vDCs for the following platforms:

- **Kubernetes**
  
  If you deployed Kubeturbo 7.22.0 or earlier, each namespace appears in Workload Optimization Manager as a vDC. A namespace includes a ResourceQuota object to determine the capacity of this vDC.
  
  If you deployed Kubeturbo 7.22-ea or later, the supply chain displays Kubernetes namespace *on page 146* as a separate entity. The supply chain also models workload controllers *on page 142* and container specs *on page 138* to account for historical utilization of container resources.

- **Cloud Foundry**
  
  An Org appears in Workload Optimization Manager as a vDC. The Org includes a current Quota Plan, which determines the capacity of this vDC.
  
  For Cloud Foundry and Pivotal Cloud Foundry, Workload Optimization Manager imports the Quota Plan of the Organization and Spaces where Workload Optimization Manager represents MemAllocation as the memory limit (maximum memory allowed) and number of consumers as the total number of containers allowed.
Entity Types - On-prem Infrastructure

Synopsis

| Budget: | The 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 containers or container pods. |
| Provides: | Resources to host containers or container pods. |
| Consumes: | Resources from VMs (nodes or diego cells). |
| Discovered through: | Workload Optimization Manager discovers these vDCs through Kubeturbo pods, Cloud Foundry targets, or Pivotal Operations Manager targets. |

Monitored Resources

Workload Optimization Manager monitors the following resources for a Container Orchestrator vDC:

**VMem Request Quota**

The total amount of virtual memory request for all pods allocated to the namespace against the namespace quota. Measured in Megabytes (MB)

**VCPU Request Quota**

The total amount of virtual CPU request for all pods allocated to the namespace against the namespace quota. Measured in millicores (mCores)

**VMem Limit Quota**

The total amount of virtual memory limit for all pods allocated to the namespace against the namespace quota. Measured in Megabytes (MB)

**VCPU Limit Quota**

The total amount of virtual CPU limit for all pods allocated to the namespace against the namespace quota. Measured in millicores (mCores)

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

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.
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.

<table>
<thead>
<tr>
<th>Synopsis</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Budget:</strong></td>
<td>A Business User has unlimited budget.</td>
</tr>
<tr>
<td><strong>Provides:</strong></td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Consumes:</strong></td>
<td>Resources from the underlying desktop pools:</td>
</tr>
<tr>
<td></td>
<td><em>Sessions</em></td>
</tr>
<tr>
<td></td>
<td><em>Pool Memory</em></td>
</tr>
<tr>
<td></td>
<td><em>Pool Storage</em></td>
</tr>
<tr>
<td></td>
<td><em>Pool CPU</em></td>
</tr>
</tbody>
</table>

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.
**Synopsis**

Discovered through: The LDAP server that manages these users. 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 for a Business User:

- **ImageCPU**
  
  CPU utilization, as a percentage of CPU capacity for the user's desktop image or images.

- **ImageMem**
  
  Memory utilization, as a percentage of Memory capacity for the user's desktop image or images.

- **ImageStorage**
  
  Storage utilization, as a percentage of storage capacity for the user's desktop image or images.

**Business User Actions**

- **Move Business User**

  Workload Optimization Manager can move a business user to address:

  - Resource congestion on the image – When utilization is consistently near capacity for image resources, Workload Optimization Manager can recommend moving the 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 the business user to a desktop pool that has more available resources.

**NOTE:**

To support moves to different desktop pools, you must configure placement policies that merge *similarly configured* desktop pools. See [Desktop Pool Placement Policies](on page 213).

**Business User Policies**

Workload Optimization Manager ships with default settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change *action automation* or *constraints* for that scope. You can create policies that override the defaults for the scopes you specify.

**Action Automation and Orchestration**

For details about Business User actions, see [Business User Actions](on page 209).

<table>
<thead>
<tr>
<th>Action</th>
<th>Default Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Move</td>
<td>Recommend</td>
</tr>
</tbody>
</table>

**Image Target Utilization**

Workload Optimization Manager tracks utilization of desktop image resources for the Business Users in your Virtual Desktop Infrastructure (VDI) environment.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
</table>
| Image CPU Target Utilization | 70  
  The target utilization as a percentage of CPU capacity.                        |
<p>| Image MEM Target Utilization  | 90                                                                                         |</p>
<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image Storage Target Utilization</td>
<td>90</td>
</tr>
<tr>
<td>The target utilization as a percentage of storage capacity.</td>
<td></td>
</tr>
</tbody>
</table>

**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**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggressiveness</td>
<td>95th Percentile</td>
</tr>
</tbody>
</table>

  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**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max Observation Period</td>
<td>Last 7 Days</td>
</tr>
</tbody>
</table>

  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
  - More Elastic – Last 30 Days
  - (Default) 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.

<table>
<thead>
<tr>
<th>Synopsis</th>
<th>A Desktop Pool gets its budget by selling resources to Business Users.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides</td>
<td>Resources for Business Users to use:</td>
</tr>
<tr>
<td></td>
<td>▪ PoolMEM</td>
</tr>
<tr>
<td></td>
<td>▪ PoolCPU</td>
</tr>
<tr>
<td></td>
<td>▪ Sessions</td>
</tr>
<tr>
<td>Consumes</td>
<td>▪ Compute and storage resources from the associated Virtual Datacenters</td>
</tr>
<tr>
<td></td>
<td>▪ Sessions from the underlying View Pod</td>
</tr>
<tr>
<td>Discovered through</td>
<td>The VDI management target.</td>
</tr>
<tr>
<td></td>
<td>For VMware Horizon View, the target is the View Connection Server.</td>
</tr>
</tbody>
</table>
Monitored Resources

Workload Optimization Manager monitors the following resources for a Desktop Pool:

- **Pool CPU**
  The CPU available to the pool that is in use by active sessions.

- **Pool Memory**
  The memory available to the pool that is in use by active sessions.

- **Pool Storage**
  The storage capacity available to the pool that is in use by active sessions.

- **Active Sessions**
  How many active sessions are on the pool as a percentage of the pool’s capacity as defined in the Workload Optimization Manager policy.

- **Total Sessions**
  How many active and disconnected (non-terminated) sessions are on the pool, as a percentage of the pool’s capacity.

Actions

Workload Optimization Manager does not recommend actions for a desktop pool. It does recommend actions for the business users running active sessions in the pool.

Desktop Pool Policies

Workload Optimization Manager ships with default settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.

Action Automation and Orchestration

Workload Optimization Manager does not recommend actions for a desktop pool. It does recommend 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**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily observation windows</td>
<td>3 windows per day</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>Window</th>
<th>Time range</th>
<th>Average utilization</th>
</tr>
</thead>
<tbody>
<tr>
<td>W1</td>
<td>00:00 – 08:00</td>
<td>10%</td>
</tr>
<tr>
<td>W2</td>
<td>08:00 – 16:00</td>
<td>80%</td>
</tr>
<tr>
<td>W3</td>
<td>16:00 – 24:00</td>
<td>40%</td>
</tr>
</tbody>
</table>
Average utilization for this day *without* the benefit of observation windows would be 44%. By using observation windows we 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**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max Observation Period</td>
<td>Last 7 Days</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pool CPU Utilization</td>
<td>95</td>
</tr>
<tr>
<td>Pool Mem Utilization</td>
<td>95</td>
</tr>
<tr>
<td>Pool Storage Utilization</td>
<td>95</td>
</tr>
</tbody>
</table>

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 larger 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 larger 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 87)*.
View Pod

For Virtual Desktop Infrastructure (VDI) environments, a View Pod groups together a given set of Desktop Pools.

**Synopsis**

<table>
<thead>
<tr>
<th>Budget:</th>
<th>A View Pod has unlimited budget.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides:</td>
<td>Active Sessions.</td>
</tr>
<tr>
<td>Consumes:</td>
<td>N/A</td>
</tr>
<tr>
<td>Discovered through:</td>
<td>The VDI management target. For VMware Horizon View, the target is the View Connection Server.</td>
</tr>
</tbody>
</table>

**Monitored Resources**

Workload Optimization Manager monitors the following resources for a Desktop Pool:

- **Active Sessions**
  - How many active sessions are on the pool as a percentage of the pool’s capacity as defined in the Workload Optimization Manager policy.
- **Total Sessions**
  - How many active and disconnected (non-terminated) sessions are on the pool, as a percentage of the pool’s capacity.

**Actions**

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 Scoped Automation Policies (on page 96). For information about view pod policies, see View Pod Policies (on page 215).

1. Create a new scoped 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 357).

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 settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.

Action Automation and Orchestration
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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active Sessions Capacity</td>
<td>8000</td>
</tr>
</tbody>
</table>

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 214).

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’s most usual 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 224).

On the public cloud a host is an availability zone. This is where your cloud workloads run. For details, see Zone (on page 180).

**Synopsis**

| Budget: | 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 isn’t enough to pay for the services it consumes, Workload Optimization Manager will recommend to suspend or power off the host. |
| Provides: | Host resources for VMs to use: |
| | • Mem (Kbytes) |
| | • CPU (MHz) |
| | • IO (throughput on the I/O bus) |
| | • Net (network throughput) |
| | • Swap (swap rate capacity measured in bytes/sec) |
| | • Ballooning (sharing of memory among hosted VMs) |
| | • CPU Ready Queue (wait time on the queue in ms) |
| Consumes: | Datacenter resources (physical space, cooling, etc.) and storage. |
**Synopsis**

**Discovered through:** Workload Optimization Manager discovers hosts 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 on a host:

- **Memory (Mem)**
  - The utilization of the PM’s memory reserved or in use
  - Measured in Kilobytes (KB)

- **CPU**
  - The utilization of the PM’s CPU reserved or in use
  - Measured in Megahertz (MHz)

- **IO**
  - The utilization of the PM’s IO adapters
  - Measured in Kilobytes per second (KB/s)

- **Net**
  - The utilization of data through the PM’s network adapters
  - Measured in Kilobytes per second (KB/s)

- **Swap**
  - The utilization of the PM’s swap space
  - Measured in Kilobytes (KB)

- **Balloon**
  - The utilization of shared memory among VMs running on the host. ESX-only
  - Measured in Kilobytes (KB)

- **CPU Ready**
  - The utilization of the PM’s allocated ready queue capacity (measured in Kbytes) that is in use, for 1, 2, and 4 CPU ready queues. ESX-only
  - Measured in Megahertz (MHz)

---

**Host Actions**

Workload Optimization Manager can recommend the following actions for a host:

- **Start**
  - For increased demand on physical resources, start up a suspended host.

- **Provision**
  - For increased demand of physical resources, install a new host in the environment. Workload Optimization Manager will then move workload to that host.

- **Suspend**
  - For underutilized resources on a host, move existing workload to other hosts and 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 91)](http://example.com)
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).

<table>
<thead>
<tr>
<th>Workload Optimization Manager Vendor Migration Level</th>
<th>vCenter Migration Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Conservative)</td>
<td>5</td>
</tr>
<tr>
<td>2 (Less Conservative)</td>
<td>4</td>
</tr>
<tr>
<td>3 (Moderate)</td>
<td>3</td>
</tr>
<tr>
<td>4 (Less Aggressive)</td>
<td>2</td>
</tr>
<tr>
<td>5 (Aggressive)</td>
<td>1</td>
</tr>
</tbody>
</table>

Host Policies
Workload Optimization Manager ships with default settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change *action automation* or *constraints* for that scope. You can create policies that override the defaults for the scopes you specify.

Action Automation and Orchestration
For details about host actions, see [Host Actions (on page 217)](#).

<table>
<thead>
<tr>
<th>Action</th>
<th>Default Mode</th>
<th>vCenter</th>
<th>XenServer</th>
<th>Hyper-V</th>
<th>RHEV</th>
<th>UCS (blades only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start</td>
<td>Recommend</td>
<td><img src="image" alt="Auto" /></td>
<td><img src="image" alt="Rcmd" /></td>
<td><img src="image" alt="Rcmd" /></td>
<td><img src="image" alt="Auto" /></td>
<td><img src="image" alt="Auto" /></td>
</tr>
<tr>
<td>Suspend</td>
<td>Recommend</td>
<td><img src="image" alt="Auto" /></td>
<td><img src="image" alt="Rcmd" /></td>
<td><img src="image" alt="Rcmd" /></td>
<td><img src="image" alt="Auto" /></td>
<td><img src="image" alt="Auto" /></td>
</tr>
<tr>
<td>Provision</td>
<td>Recommend</td>
<td><img src="image" alt="Rcmd" /></td>
<td><img src="image" alt="Rcmd" /></td>
<td><img src="image" alt="Rcmd" /></td>
<td><img src="image" alt="Rcmd" /></td>
<td><img src="image" alt="Auto" /></td>
</tr>
<tr>
<td>Reconfigure</td>
<td>Recommend</td>
<td><img src="image" alt="Rcmd" /></td>
<td><img src="image" alt="Rcmd" /></td>
<td><img src="image" alt="Rcmd" /></td>
<td><img src="image" alt="Rcmd" /></td>
<td><img src="image" alt="Rcmd" /></td>
</tr>
</tbody>
</table>

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.
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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory Overprovisioned Percentage</td>
<td>1000</td>
</tr>
<tr>
<td>Net Throughput</td>
<td>50</td>
</tr>
<tr>
<td>Ready Queue Utilization</td>
<td>50</td>
</tr>
<tr>
<td>Memory Utilization</td>
<td>100</td>
</tr>
<tr>
<td>IO Throughput</td>
<td>50</td>
</tr>
<tr>
<td>CPU Overprovisioned Percentage</td>
<td>1000</td>
</tr>
<tr>
<td>CPU Utilization</td>
<td>100</td>
</tr>
<tr>
<td>Swapping Utilization</td>
<td>20</td>
</tr>
</tbody>
</table>

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 1000 for Memory Overprovisioned Percentage means that overprovisioning memory by 5 times the physical capacity shows up as 50% utilization of the Mem Overprovisioned capacity in Workload Optimization Manager.
- 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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diameter</td>
<td>10</td>
</tr>
<tr>
<td>Center</td>
<td>70</td>
</tr>
</tbody>
</table>

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.

**Placement Policies**
You can create placement policies that merge multiple clusters into a single logical group for the purpose of 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 87).

**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 87).

**Chassis**
A chassis houses the servers that are part of a computing fabric. It provides compute, memory, storage, and bandwidth resources.

**Synopsis**

<table>
<thead>
<tr>
<th>Synopsis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Budget:</strong></td>
</tr>
<tr>
<td><strong>Provides:</strong></td>
</tr>
<tr>
<td><strong>Consumes:</strong></td>
</tr>
<tr>
<td><strong>Discovered through:</strong></td>
</tr>
</tbody>
</table>

**Monitored Resources**
Workload Optimization Manager monitors the following resources for the servers in a chassis:

- **Power**
  - Electricity being consumed by the Chassis
  - Measured in Watts (W)
• Cooling
  The percentage of the acceptable temperature range that is utilized by this chassis. As the chassis temperature nears the high or low running temperature limits, this percentage increases.

Actions
Workload Optimization Manager does not recommend actions for a chassis.

Datacenter
A datacenter is the sum of VMs, PMs, datastores, and network devices that are managed by a given hypervisor target. A datacenter provides compute, memory, storage, and bandwidth resources.

NOTE:
For public cloud environments, a datacenter is the cloud region. The hosts that get resources from the datacenter are availability zones within that region. For details, see Region (on page 182) and Zone (on page 180).

Synopsis

<table>
<thead>
<tr>
<th>Summary</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Budget:</td>
<td>A Datacenter has unlimited budget.</td>
</tr>
<tr>
<td>Provides:</td>
<td>Compute, memory, storage, and bandwidth resources</td>
</tr>
<tr>
<td>Consumes:</td>
<td>N/A</td>
</tr>
<tr>
<td>Discovered through:</td>
<td>Workload Optimization Manager discovers Datacenters through hypervisor targets.</td>
</tr>
</tbody>
</table>
Monitored Resources

Workload Optimization Manager does not monitor resources directly from the datacenter, but it does monitor the following resources, aggregated for the hosts in a datacenter:

- **Memory (Mem)**
  The utilization of the PM’s memory reserved or in use
  Measured in Kilobytes (KB)

- **CPU**
  The utilization of the PM’s CPU reserved or in use
  Measured in Megahertz (MHz)

- **IO**
  The utilization of the PM’s IO adapters
  Measured in Kilobytes per second (KB/s)

- **Net**
  The utilization of data through the PM’s network adapters
  Measured in Kilobytes per second (KB/s)

- **Swap**
  The utilization of the PM’s swap space
  Measured in Kilobytes (KB)

- **Balloon**
  The utilization of shared of memory among VMs running on the host. ESX-only
  Measured in Kilobytes (KB)

- **CPU Ready**
  The utilization of the PM’s allocated ready queue capacity (measured in Kbytes) that is in use, for 1, 2, and 4 CPU ready queues. ESX-only
  Measured in Kilobytes (KB)

Actions

Workload Optimization Manager does not recommend actions for a datacenter. Instead, it recommends actions for the devices running in the datacenter.

Placement Policies

For vCenter environments, you can create placement policies that merge datacenters to support cross-vCenter moves. In this case, where a datacenter corresponds to a given vCenter target, the merged clusters can be in different datacenters. In this case you must create two merge policies; one to merge the affected datacenters, and another to merge the specific clusters.

For more information, see [Creating Placement Policies (on page 87)](#).

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.

### Provides:
- Host resources for VMs to use:
  - Storage amount
  - IOPS (storage access operations per second)
  - Latency (capacity for disk latency in ms)

### Consumes:
- Disk arrays (or aggregates)

### Discovered through:
- Workload Optimization Manager discovers on-prem Datastores through hypervisor targets and storage controllers.

## Monitored Resources

Workload Optimization Manager monitors the following resources for a datastore:

- **Storage Amount**
  - The utilization of the datastore's capacity
  - Measured in Megabytes (MB)

- **Storage Provisioned**
  - The utilization of the datastore's capacity, including overprovisioning.
  - Measured in Megabytes (MB)

- **Storage Access Operations Per Second (IOPS)**
  - The summation of the read and write access operations per second on the datastore
  - Measured in Operations per second

- **Latency**
  - The utilization of latency on the datastore
  - Measured in Milliseconds (ms)

## Storage Actions

Workload Optimization Manager recommends the following actions for a datastore:

- **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.

Storage resize actions use Workload Optimization Manager tuned scaling settings. This gives you increased control over the action 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 189)].

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 87)].

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.

Workload Optimization Manager supports VMware vSAN.

**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.

<table>
<thead>
<tr>
<th>FTT</th>
<th>Redundancy Method</th>
<th>RAID Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>RAID1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>RAID1</td>
<td>1/2</td>
</tr>
<tr>
<td>2</td>
<td>RAID1</td>
<td>1/3</td>
</tr>
<tr>
<td>1</td>
<td>RAID5/6</td>
<td>3/4</td>
</tr>
<tr>
<td>2</td>
<td>RAID5/6</td>
<td>2/3</td>
</tr>
</tbody>
</table>

**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 229).

The calculation for Usable Capacity can be expressed as:

\[
\text{Usable Capacity} = (\text{Raw Capacity} - \text{Largest Host Capacity} \times \text{Host Capacity Reservation}) \times \text{Slack Space Percentage} \times \text{RAID Factor} \times \text{Compression Ratio}
\]

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.
Entity Types - On-prem Infrastructure

- 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 370)](on page 370).

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 settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.
### Action Automation and Orchestration

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 223).

<table>
<thead>
<tr>
<th>Action</th>
<th>Default Mode</th>
<th>vCenter</th>
<th>XenServer</th>
<th>Hyper-V</th>
<th>RHEV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delete (Volume)</td>
<td>Recommend</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
</tr>
<tr>
<td>Suspend</td>
<td>Manual</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
</tr>
<tr>
<td>Delete (Datastore)</td>
<td>Disabled</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
</tr>
<tr>
<td>Move</td>
<td>Recommend</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
</tr>
<tr>
<td>Provision</td>
<td>Recommend</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
</tr>
<tr>
<td>Start</td>
<td>Recommend</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
</tr>
<tr>
<td>Resize (Up, Down, Above Max, or Below Min - using tuned scaling)</td>
<td>Recommend</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
</tr>
</tbody>
</table>

For datastores on disk arrays:

<table>
<thead>
<tr>
<th>Action</th>
<th>Dell Compellent</th>
<th>HP 3Par</th>
<th>NetApp</th>
<th>ONTAP</th>
<th>VNX</th>
<th>VMAX</th>
<th>Nutanix</th>
<th>Pure Storage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delete (Volume)</td>
<td>Recommend</td>
<td>Nspt</td>
<td>Rcmd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suspend</td>
<td>Manual</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delete (Datastore)</td>
<td>Disabled</td>
<td>Rcmd</td>
<td>Nspt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Move</td>
<td>Recommend</td>
<td>Rcmd</td>
<td>Nspt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provision</td>
<td>Recommend</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Start</td>
<td>Recommend</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resize (Up, Down, Above Max, or Below Min - using tuned scaling)</td>
<td>Recommend</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage Amount Utilization</td>
<td>90</td>
</tr>
<tr>
<td>IOPS Utilization</td>
<td>100</td>
</tr>
<tr>
<td>Latency Utilization</td>
<td>100</td>
</tr>
</tbody>
</table>

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**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting/Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directories to Ignore</td>
<td>.dvsData.*</td>
</tr>
<tr>
<td>Files to Ignore</td>
<td>Empty String</td>
</tr>
<tr>
<td>Storage Latency Capacity [ms]</td>
<td>100</td>
</tr>
<tr>
<td>Storage Overprovisioned Percentage</td>
<td>200</td>
</tr>
<tr>
<td>IOPS Capacity</td>
<td>50000</td>
</tr>
</tbody>
</table>

**Wasted Storage Management**

You can make settings to control how Workload Optimization Manager tracks and reports on wasted storage in your environment. Wasted storage is any disk space devoted to files that are not required for operations of the devices or applications in your environment. Wasted storage may indicate opportunities for you to free up disk space, and provide more storage capacity to running VMs and applications.

**NOTE:**
It’s possible that a single datastore can be managed by more than one instance of vCenter Server. Browsing over such a datastore can result in conflicting values for wasted storage in reports and in the Improve Overall Efficiency dashboard. You should not enable datastore browsing for a scope that includes such a datastore.

If there are groups of datastores you don’t want to track for wasted storage, set the given scope and disable datastore browsing there. If you prefer not to use Workload Optimization Manager resources to track wasted 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 wasted data storage space. Separate items in these lists with the OR bar (“|”).

**Storage Latency Capacity**

This 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.

**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 200, 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.

### Scaling Constraints

This setting controls how many GB to add or subtract when resizing the allocation for a datastore.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increment Constant for Storage Amount [GB]</td>
<td>1 GB</td>
</tr>
</tbody>
</table>

### Hyperconverged Infrastructure Settings

Workload Optimization Manager considers these settings when calculating capacity and utilization for hyperconverged environments.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Setting/Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host Capacity Reservation</td>
<td>1</td>
</tr>
<tr>
<td>Host IOPS Capacity</td>
<td>50000</td>
</tr>
<tr>
<td>Slack Space Percentage</td>
<td>25</td>
</tr>
<tr>
<td>Compression Ratio</td>
<td>1</td>
</tr>
<tr>
<td>Usable Space Includes Compression</td>
<td>Off</td>
</tr>
</tbody>
</table>

**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 224).

- **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 Storage Provisioned.

**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 the purpose of workload placement.

For more information, see [Creating Placement Policies](on page 87).

**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

<table>
<thead>
<tr>
<th>Budget:</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides:</td>
<td>Storage resources</td>
</tr>
<tr>
<td>Consumes:</td>
<td>Disk array resources</td>
</tr>
<tr>
<td>Discovered through:</td>
<td>Storage targets</td>
</tr>
</tbody>
</table>

#### Monitored Resources

Workload Optimization Manager monitors the following resources for a logical pool:

- **Storage Amount**
  - The utilization of the logical pool's capacity.
  - Measured in Megabytes (MB)

- **Storage Provisioned**
  - The utilization of the logical pool's capacity, including overprovisioning.
  - Measured in Megabytes (MB)

- **Storage Access Operations Per Second (IOPS)**
  - The summation of the read and write access operations per second on the logical pool.
  - Measured in Operations per second

- **Latency**
Logical Pool Actions
Workload Optimization Manager can recommend the following actions for a logical pool:

- Resize
- Provision
- Move
- Start
- Suspend

Logical Pool Policies
Workload Optimization Manager ships with default settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.

Action Automation and Orchestration

<table>
<thead>
<tr>
<th>Action</th>
<th>Default Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suspend</td>
<td>Disabled</td>
</tr>
<tr>
<td>Start</td>
<td>Disabled</td>
</tr>
<tr>
<td>Move</td>
<td>Disabled</td>
</tr>
<tr>
<td>Provision</td>
<td>Disabled</td>
</tr>
</tbody>
</table>

Storage Settings

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage Latency Capacity [ms]</td>
<td>100</td>
</tr>
<tr>
<td>Storage Overprovisioned Percentage</td>
<td>200</td>
</tr>
<tr>
<td>IOPS Capacity</td>
<td>50000</td>
</tr>
</tbody>
</table>

- **Storage Latency Capacity**
  This 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

<table>
<thead>
<tr>
<th>Synopsis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Budget:</strong> A disk array gains its budget by selling resources to the datastores it serves. If utilization of a disk array is high enough, Workload Optimization Manager can recommend that you provision a new one.</td>
</tr>
<tr>
<td><strong>Provides:</strong> Storage resources for datastores to use:</td>
</tr>
<tr>
<td>- Storage amount</td>
</tr>
<tr>
<td>- Storage Provisioned</td>
</tr>
<tr>
<td>- IOPS (storage access operations per second)</td>
</tr>
<tr>
<td>- Latency (capacity for disk latency in ms)</td>
</tr>
<tr>
<td><strong>Consumes:</strong> Storage controllers</td>
</tr>
<tr>
<td><strong>Discovered through:</strong> Workload Optimization Manager discovers disk arrays through storage controller targets.</td>
</tr>
</tbody>
</table>

Monitored Resources

Workload Optimization Manager monitors the following resources for a disk array:

**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 widget in the UI will display no data.

- Storage Amount
  The utilization of the Disk Array’s capacity.
Measured in Megabytes (MB)
• Storage Provisioned
  The utilization of the Disk Array’s capacity, including overprovisioning.
  Measured in Megabytes (MB)
• Storage Access Operations Per Second (IOPS)
  The summation of the read and write access operations per second on the disk array
  Measured in Operations per second
• Latency
  The utilization of latency, computed from the latency of each device in the disk array.
  Measured in milliseconds (ms)

**Disk Array Actions**

Workload Optimization Manager recommends the following actions for a disk array:

• Provision
  For high utilization of the disk array’s storage, provision a new disk array (recommendation, only).

• Start
  For high utilization of disk array, start a suspended disk array (recommendation, only).

• Suspend
  For low utilization of the disk array’s storage, move VMs to other datastores and suspend volumes on the disk array (recommendation, only).

• Move (for NetApp Cluster-Mode, only)
  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 settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.

**Action Automation and Orchestration**

The following table describes the default action mode for disk array actions and automation support for environments that have Disk Array Storage Controllers as targets.

<table>
<thead>
<tr>
<th>Action</th>
<th>Default Mode</th>
<th>Dell Compellent</th>
<th>HP 3Par</th>
<th>NetApp ONTAP</th>
<th>VMAX</th>
<th>VNX</th>
<th>Nutanix</th>
<th>Pure Storage</th>
<th>XTremIO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Move</td>
<td>Disabled</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Action</td>
<td>Default Mode</td>
<td>Dell Compellent</td>
<td>HP 3Par</td>
<td>NetApp ONTAP</td>
<td>VMAX</td>
<td>VNX</td>
<td>Nutanix</td>
<td>Pure Storage</td>
<td>XTremIO</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------</td>
<td>-----------------</td>
<td>---------</td>
<td>--------------</td>
<td>------</td>
<td>-----</td>
<td>---------</td>
<td>--------------</td>
<td>---------</td>
</tr>
<tr>
<td>Provision</td>
<td>Rcmand</td>
<td></td>
<td>Rcmand</td>
<td>(C-Mode, only)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resize (up)</td>
<td>Rcmand</td>
<td>Rcmand</td>
<td>Rcmand</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Start</td>
<td>Rcmand</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suspend</td>
<td>Disabled</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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:

<table>
<thead>
<tr>
<th>Automated Action</th>
<th>7-Mode</th>
<th>Cluster-Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Move VM between datastores, on the same disk array</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Move VM between datastores on different disk arrays</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Move Datastore between disk arrays on the same storage controller</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Move Datastore between disk arrays on different storage controllers</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Resize Storage</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Resize Disk Array</td>
<td>No – Resize up, only</td>
<td>No – Resize up, only</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage Amount Utilization</td>
<td>90</td>
</tr>
</tbody>
</table>

**Storage Settings**

Set capacity for specific storage resources.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IOPS Capacity</td>
<td>5000</td>
</tr>
</tbody>
</table>
**Entity Types - On-prem Infrastructure**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A generic setting for disk array IOPS capacity (see <strong>Disk Array IOPS Capacity</strong> below).</td>
<td></td>
</tr>
<tr>
<td>VSeries LUN IOPS Capacity</td>
<td>5000</td>
</tr>
<tr>
<td>7.2k Disk IOPS Capacity</td>
<td>800</td>
</tr>
<tr>
<td>10k Disk IOPS Capacity</td>
<td>1200</td>
</tr>
<tr>
<td>15k Disk IOPS Capacity</td>
<td>1600</td>
</tr>
<tr>
<td>SSD Disk IOPS Capacity</td>
<td>50000</td>
</tr>
<tr>
<td>Disk Array IOPS Capacity</td>
<td>10000</td>
</tr>
<tr>
<td>Storage Overprovisioned Percentage</td>
<td>200</td>
</tr>
<tr>
<td>Storage Latency Capacity [ms]</td>
<td>100</td>
</tr>
</tbody>
</table>

**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 (**SSD Disk IOPS, 7.2k Disk IOPS**, etc)
    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**
  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 DiskArray 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.
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**

<table>
<thead>
<tr>
<th>Budget:</th>
<th>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.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides:</td>
<td>CPU resources to manage disk arrays.</td>
</tr>
<tr>
<td>Consumes:</td>
<td>NA</td>
</tr>
<tr>
<td>Discovered through:</td>
<td>Workload Optimization Manager directly accesses storage controller targets.</td>
</tr>
</tbody>
</table>

**Monitored Resources**

Workload Optimization Manager monitors the following resources for a storage controller:

- **CPU**
  - The utilization of the Storage Controller’s CPU in use
  - Measured in Megahertz (MHz)

- **Storage Amount**
  - The utilization of the storage controller's capacity. The storage allocated to a storage controller is the total of all the physical space available to aggregates managed by that storage controller
  - Measured in Megabytes (MB)

**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 recommends the following actions for a storage controller:

- **Provision (recommendation, only)**
  - 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 settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes
of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.

### Action Automation and Orchestration

Actions for individual Disk Array Storage Controllers:

<table>
<thead>
<tr>
<th>Action</th>
<th>Default Mode</th>
<th>Dell Compellent</th>
<th>HP 3Par</th>
<th>NetApp ONTAP</th>
<th>VNX</th>
<th>VMAX</th>
<th>Nutanix</th>
<th>Pure Storage</th>
<th>XTremIO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provision</td>
<td>Disabled</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
<td>Rcmd</td>
</tr>
</tbody>
</table>

### 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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage Amount Utilization</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Maximum allowed utilization of storage that is managed by the Storage Controller.</td>
</tr>
<tr>
<td>CPU Utilization</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>Maximum allowed utilization of Storage Controller CPU (from 20 to 100).</td>
</tr>
</tbody>
</table>

### Storage Settings

Set capacity for specific storage resources.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>IOPS Capacity</td>
<td>5000</td>
</tr>
<tr>
<td>Storage Latency Capacity [ms]</td>
<td>100</td>
</tr>
</tbody>
</table>

### 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

<table>
<thead>
<tr>
<th>Budget:</th>
<th>An IO Module gains its budget by selling Net resources to a physical machine.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides:</td>
<td>Net resources</td>
</tr>
<tr>
<td>Consumes:</td>
<td>Chassis and Fabric Interconnect</td>
</tr>
<tr>
<td>Discovered through:</td>
<td>Workload Optimization Manager discovers IO Modules through the fabric managers that use them.</td>
</tr>
</tbody>
</table>

Monitored Resources

Workload Optimization Manager monitors the following resources for an IO Module:

- **NetThroughput**
  - Rate of message delivery over a port
  - Measured in Megabits per second (Mb/s)

Actions

Workload Optimization Manager does not recommend actions for an 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

<table>
<thead>
<tr>
<th>Budget:</th>
<th>A switch gains its budget by selling Net resources to the IO Modules.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides:</td>
<td>Net resources</td>
</tr>
<tr>
<td>Consumes:</td>
<td>N/A</td>
</tr>
<tr>
<td>Discovered through:</td>
<td>Workload Optimization Manager discovers switches through managers of fabric platforms (such as UCS) that use them.</td>
</tr>
</tbody>
</table>

Monitored Resources

Workload Optimization Manager monitors the following resources for a switch:

- **NetThroughput**
  - Rate of message delivery over a port
  - Measured in Mb/s
- **PortChannel**
  - Amalgamation of ports with a shared net throughput and utilization
  - Measured in Mb/s

Actions

Workload Optimization Manager can recommend resizing port for a switch to increase size.

Switch Policies

Workload Optimization Manager ships with default settings that we believe will give you the best results from our analysis. These settings are specified in a set of default automation policies for each type of entity in your environment. For some scopes of your environment, you might want to change these settings. For example, you might want to change action automation or constraints for that scope. You can create policies that override the defaults for the scopes you specify.
# Action Automation and Orchestration

For environments that have Fabric Managers as targets:

<table>
<thead>
<tr>
<th>Action</th>
<th>Default Mode</th>
<th>Cisco UCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resize</td>
<td>Recommend</td>
<td>Rcmd</td>
</tr>
<tr>
<td>Start</td>
<td>Recommend</td>
<td>Rcmd</td>
</tr>
<tr>
<td>Provision</td>
<td>Recommend</td>
<td>Rcmd</td>
</tr>
<tr>
<td>Suspend</td>
<td>Disabled</td>
<td>Rcmd</td>
</tr>
<tr>
<td>Move</td>
<td>Disabled</td>
<td>Rcmd</td>
</tr>
</tbody>
</table>

## 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.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Default Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switch Net Throughput</td>
<td>70</td>
</tr>
</tbody>
</table>
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 244).
- 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 301).

  NOTE:
  By default, Workload Optimization Manager saves plans after you run them. However, when you update Workload Optimization Manager to a new major version these saved plans do not carry over to the update.
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 would get if you changed your environment in some way.

This topic walks you through the general process of setting up a plan scenario.

1. 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.

  - Cloud scope
    
    If you set the scope to a **specific** Account, Billing Family, VM Group, or Region, you can start an Optimize Cloud plan.

  - On-prem scope

    If you set the scope to a **specific** 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 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 cluster scope

If you set the scope to a specific Container Platform Cluster, you can start an Optimize Container Cluster plan. For details, see Scoping the Workload Optimization Manager Session (on page 50).

After setting the scope, the Plan button appears in the Home Page.

2. Plan Types

Select from the list of plan types. For more information, see Plan Scenarios and Types (on page 251).
3. Plan Wizards

Each plan type includes a wizard to guide you through creating the scenario. The wizard leads you through the required 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.
4. Plan Scope

All plans require a scope. For example, to configure a Hardware Refresh plan, you set the scope to the hosts that you plan to replace.
It usually helps to focus on a subset of your environment. For a very large environment, scoped plans run faster.

To narrow the scope, select a group from the list on the left side of the page. The page then refreshes to include only the entities belonging to that group.

Use Search or Filter to sort through a long list.

5. Additional Plan Information

The wizard prompts you for any additional information required to run the plan. For example, for a Hardware Refresh plan, you need to identify the hosts that will replace the scoped hosts.

6. Run the Plan

After you provide the minimum required information for running a plan, the wizard shows you the following options:
7. The Plan Page

The Plan Page first displays if you skip the wizard or as soon as 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:
Plans: Looking to the Future

### Plan Page Sections

<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Plan name</td>
<td>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.</td>
</tr>
<tr>
<td>B. Plan scope</td>
<td>Review the scope that you set in a previous step. <strong>NOTE:</strong> It is not possible to change the scope of the plan in the Plan Page. You will need to start over if you want a different scope. To start over, go to the top-right section of the page, click the More options icon ( ), and then select New Plan.</td>
</tr>
<tr>
<td>C. Configuration toolbar</td>
<td>Configure additional 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.</td>
</tr>
<tr>
<td>D. Configuration summary</td>
<td>Review the plan’s configuration settings. You can remove any setting by clicking the x mark on the right. Use the toolbar on top to change the settings. As you make changes to the plan scenario, those changes immediately appear in the Configuration summary.</td>
</tr>
</tbody>
</table>
| E. Additional options | See what else you can do with the plan.  
  - **Run / Run Again:**  
    - If a plan has not run, click Run and then check the plan results.  
    - If the plan has run 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. |
<table>
<thead>
<tr>
<th>Plan Page Sections</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>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.</td>
<td></td>
</tr>
<tr>
<td>Reset view: Restore charts to their default views. For example, if you changed the commodities displayed in the Optimized Improvements or Comparison charts, you can discard those changes by choosing this option.</td>
<td></td>
</tr>
<tr>
<td>Delete plan: Choose if you no longer need the plan.</td>
<td></td>
</tr>
</tbody>
</table>

**F. Plan results**
Review the results in the charts provided. For a plan that has not run, you will see a Scope Preview chart and a one-time message instructing you to run the plan.

**8. Plan Management**
All the plans you have created display in the Plan Management Page (on page 243).

**Plan Scenarios and Types**
To simulate different plan scenarios, Workload Optimization Manager provides the following general types of plans:
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 suggestions to buy RIs, comparisons of template and storage usage, and a comparison of current to optimized cost.

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.

**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 datacenters (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 Clusters**

See the effects of merging two or more clusters. For example, you can see if merging the 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.

You would also choose Custom Plan if you need to run plans that include containers and container pods.
Run an Optimize Container Cluster plan to identify performance and efficiency opportunities for a single Kubernetes 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:
- Standalone container cluster
- Container cluster in an on-prem or public cloud environment
- Container cluster stitched to applications via Data Ingestion Framework (DIF)

Scoping to a group within a Kubernetes 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 Kubernetes cluster.

For an overview of setting up plan scenarios, see Setting Up Plan Scenarios (on page 244).
Select a Kubernetes cluster to optimize.

Scoping to a group within a Kubernetes 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.

**2. Optimization Settings**
Choose from the given optimization scenarios.
Plans: Looking to the Future

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 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) Additional 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 Kubernetes 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.

![Optimize Container Cluster 1](image)

**General Guidelines**
Familiarize yourself with these common terms that appear in many sections of the plan results:

- A container pod represents the compute demand from a running pod.
- A Kubernetes 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

**RESULTS OVERVIEW**

<table>
<thead>
<tr>
<th></th>
<th>Current</th>
<th>After Plan</th>
<th>Difference</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Container Pods</td>
<td>118</td>
<td>113</td>
<td>5</td>
<td>▼4.2 %</td>
</tr>
<tr>
<td>Virtual Machines</td>
<td>8</td>
<td>7</td>
<td>1</td>
<td>▼12.5 %</td>
</tr>
<tr>
<td>Pod Density</td>
<td>14.8 : 1</td>
<td>16.1 : 1</td>
<td>1.3 : 1</td>
<td>▼8.8 %</td>
</tr>
<tr>
<td>Cluster CPU Capacity</td>
<td>20 Cores</td>
<td>22 Cores</td>
<td>2 Cores</td>
<td>▲10 %</td>
</tr>
<tr>
<td>Cluster Memory Capacity</td>
<td>74 GB</td>
<td>82.9 GB</td>
<td>8.9 GB</td>
<td>▲12.1 %</td>
</tr>
<tr>
<td>Cluster Allocatable CPU</td>
<td>19.2 Cores</td>
<td>21.1 Cores</td>
<td>2 Cores</td>
<td>▲10.3 %</td>
</tr>
<tr>
<td>Cluster Allocatable Memory</td>
<td>53.8 GB</td>
<td>62.6 GB</td>
<td>8.8 GB</td>
<td>▲16.4 %</td>
</tr>
<tr>
<td>Cluster CPU Overcommitment</td>
<td>261 %</td>
<td>223.2 %</td>
<td>37.8 %</td>
<td>▲14.5 %</td>
</tr>
<tr>
<td>Cluster Memory Overcommitment</td>
<td>83.7 %</td>
<td>45.4 %</td>
<td>38.3 %</td>
<td>▲45.8 %</td>
</tr>
</tbody>
</table>

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:

\[
\text{Overcommitment} = \frac{\text{Sum of CPU limits for all containers}}{\text{Sum of CPU capacity for all nodes}}
\]

- **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:

\[
\text{Overcommitment} = \frac{\text{Sum of memory limits for all containers}}{\text{Sum of memory capacity for all nodes}}
\]

### 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 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. 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.
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.
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 top-right section of 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.
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.
Namespace Resources Comparison

This chart shows how much cluster resources per namespace are utilized by pods. Utilization can be expressed as follows:

\[
\text{Utilization} = \frac{\text{Sum of actual vMem/vCPU used by pods}}{\text{vMem/vCPU capacity for the cluster}}
\]

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 on an on-prem Kubernetes cluster. These charts show how the utilization of resources would change assuming 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 on an on-prem Kubernetes 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 container cluster, click the download button at the top-right section of 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 at the top-right section of 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 \( \infty \) 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 on top of 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 will need to start over if you want a different scope. To start over, go to the top-right section of the page, click the More options icon (\( \mathbin{\text{\scalebox{1.4}{\text{⋯}}}} \)), and then select **New Plan**.

When you are ready to re-run the plan, click **Run Again** on the top-right section of the page.

### 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 RI pricing plans, 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 244).

1. Scope

You can scope by:
Select Scope to Optimize

- **Cloud**
  - Accounts (7)
    - AWS
  - Billing Families (3)
    - AWS
  - Cloud Providers (2)
    - Azure
  - Resource Groups (35)
  - Regions (43)

- **Accounts**
  Choose an AWS account or Azure subscription for the plan’s scope. If you choose an Account for the scope, then the plan will not calculate RI Buy actions. To optimize RI purchases for a limited scope, choose a Billing Family.

- **Billing Families**
  Include RI purchases in the planning for a scope that is limited to a single billing family. The plan calculates RI purchases through the billing family’s master account.

- **Cloud Providers**
  See how you can optimize all your AWS or Azure 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.

2. Optimization Settings

- **Optimize Services and Buy RIs**
  Scale VMs and DBs, delete unattached storage, and buy RIs. Includes RI aware scaling.

- **Optimize Services and Don’t Buy RIs**
  Scale VMs and DBs, delete unattached storage. Includes RI aware scaling.

- **Buy RIs and Don’t Optimize Services**
  Buy RIs for VMs at their current sizes.

Choose from the given optimization options. Note that if you set a plan’s scope to a resource group, Workload Optimization Manager will optimize services without recommending new RI purchases.
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 RI purchases enabled, the plan results show inaccurate RI recommendations.

Turn off **Disable All Actions** to resolve this issue. Be aware that after you turn off this setting, it will take Workload Optimization Manager a week to reflect accurate results in Optimize Cloud plans.

### 3. Reserved Instances Settings

For **Plan RI Inventory**, the RIs for the current scope are selected by default. Click **Edit** to make changes.

For **RI Purchase 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. For more information about RI Purchase, see [RI Purchase Profile](on page 375).

- **OFFERING CLASS**
  
  For AWS environments, choose the offering class that corresponds to the RI types that you typically use in your environment.

- **TERM**
  
  For AWS and Azure environments, choose the payment terms you contract for your RIs. **TERM** can be one of **1 Year** or **3 Year**. Typically, longer term payment plans cost less per year.

- **PAYMENT**
The payment option that you prefer for your AWS RIs:
- All Upfront – You make full payment at the start of the RI 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 for the RIs at an hourly rate, for the duration of the term.

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, reserved compute, on-demand database, and storage costs
- Compare current and optimized breakdowns of templates used
- Compare breakdowns of storage tiers in use
- Project the RI coverage (how many workloads use RI) and utilization (percentage of RIs that are active)
- Identify candidates for Reserved Instance (RI) pricing, and show the cost benefits you can see by running those workloads on templates that are reserved on your public cloud provider.

Viewing the Results

![Optimize Cloud 56 Results](image-url)
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 RI capacity to reduce costs. The analysis looks at workload history to identify RI candidates. This considers the count of workloads in a family, plus their hours of active-state condition, to arrive at the RI capacity you should purchase. Since RI costs are incurred at the account level, the Cloud Cost Comparison chart will present RI 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** at the bottom of the chart.

- **Volume Tier Breakdown**
  This chart shows the distribution of storage that supports your workloads in Current and Optimized graphs.
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.

- **Plan RI Inventory**
  This chart shows the RI workloads that Workload Optimization Manager discovers and lists them by tiers. For a tabular listing, click **Show All** at the bottom of the chart. In the tabular listing, you can see if an RI expired before the specified purchase date.

- **Recommended RI Purchases**
This chart shows the RIs the plan recommends to buy. To see the details, click **SHOW ALL** at the bottom of the chart.

### Viewing Plan Actions

Click the **Plan Actions** tab on top of the page to view a list of actions that you need to execute to achieve the plan results. You can download the list of actions as a CSV file.

### 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 on top of 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.

• **RI Profile**

  Change RI Purchase Profile settings that match the cloud settings you have set up for real-time analysis. For details, see [Reserved Instance Settings (on page 271)](on_page_271).

**NOTE:**

It is not possible to change the scope of the plan in the Plan Page. You will need to start over if you want a different scope. To start over, go to the top-right section of the page, click the More options icon ( ), and then select **New Plan**.

When you are ready to re-run the plan, click **Run Again** on the top-right section of the page.

**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. If analysis discovers VMs that are good candidates for Reserved Instances (RIs), then it recommends migrating to those instances, and can even recommend purchasing more RI capacity.

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 RI purchases for VMs that can benefit from such 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 an RI. 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. A plan that migrates from AWS to Azure will not migrate VMs that run on Spot Instances.
- 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.

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
- VMs that are good candidates for Reserved Instance (RI) pricing, and the cost benefits you would realize by running those VMs as RI instances
- RIs you should purchase (if you need more RI capacity)

Configuring a Migrate to Cloud Plan

For an overview of setting up plan scenarios, see Setting Up Plan Scenarios (on page 244).

1. **Scope**

Select the VMs that you want to migrate. You can select VM groups and/or individual VMs.

If you select an Auto Scaling Group, Workload Optimization Manager simulates migrating the VMs individually, and not as a group.

2. **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 defined in the policies.

3. 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.
4. Reserved Instances Settings

For **Plan RI Inventory**, the RIs for the current scope are selected by default. Click **Edit** to make changes.

For **RI Purchase 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. For more information about RI Purchase, see [RI Purchase Profile](on page 375).

- **OFFERING CLASS**
  For AWS environments, choose the offering class that corresponds to the RI types that you typically use in your environment.

- **TERM**
  For AWS and Azure environments, choose the payment terms you contract for your RIs. TERM can be one of **1 Year** or **3 Year**. Typically, longer term payment plans cost less per year.

- **PAYMENT**
  The payment option that you prefer for your AWS RIs:
  - **All Upfront** – You make full payment at the start of the RI 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 for the RIs at an hourly rate, for the duration of the term.

**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.

• **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.

- **Volume Mapping 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**

  These charts list the RIs the plan recommends for the migration, including how many of each is needed.

  ![Volumes Mapping Diagram]

  To identify RI candidates, Workload Optimization Manager 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.

  - **RI Inventory (AWS only)**
    
    For AWS environments, Workload Optimization Manager compares the RI candidates to your current inventory of RI resources, plus your desired RI coverage. If the inventory can support the VM, then Workload Optimization Manager considers it an AWS RI candidate. If the inventory cannot support the VM, or if supporting it would exceed your desired RI coverage, then Workload Optimization Manager can recommend purchasing more RI capacity.
NOTE:
The plan assumes that an RI 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 migrate to an RI that is more expensive than running on demand.

Click Show All to see a table with details for each RI. The table shows the properties, up-front cost, and break-even period for each RI investment. It also indicates the monthly savings you would realize when you buy a specific RI. Under the Actions column, click Details to compare Lift & Shift and Optimized actions.

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.
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 on top of 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 will need to start over if you want a different scope. To start over, go to the top-right section of the page, click the More options icon (⋮), and then select New Plan.

When you are ready to re-run the plan, click Run Again on the top-right section of the page.
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 use the Economic Scheduling Engine to 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 301)

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 244).

1. 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.

```
  Select Hot Cluster

  ON-PREM

  Cluster

  Datacenter-dc18\DC18-Cluster1
  Datacenter-dc22\dc22-cluster1
  Datacenter-dc22\dc22-cluster2
  DC13\DC13-Cluster

```

2. 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 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.

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 on top of 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 291).

NOTE:
It is not possible to change the scope of the plan in the Plan Page. You will need to start over if you want a different scope. To start over, go to the top-right section of the page, click the More options icon ( ), and then select New Plan.

When you are ready to re-run the plan, click Run Again on the top-right section of the page.

Custom Plan
For an overview of setting up plan scenarios, see Settings Up User Plan Scenarios (on page 244).
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 244).

1. Scope

Specify the plan scope and then click Configure Plan at the bottom of the page.

2. 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.

2.1. 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.

Set how many copies to add
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 datastores with a more powerful template, the plan might show that you can use fewer hosts or datastores, 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.

2.3. 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.

2.4. 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.

2.5. 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.

### 2.6. 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 87).

### 2.7. 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 datastores to 50% utilization. This saves storage resources for the other cluster that will use this storage.

2.8. 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.

2.9. 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 12).

**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 68). The list chart shows a partial list of actions ([on page 64]).
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, go to the top-right section of the chart, 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.
    
    To restore the default commodities, use the Reset view option at the top-right section of the page.

  - Click Show all at the bottom of the chart 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 top-right section of a chart and then select from the list of commodities.

To restore the default commodity, go to the top-right section of the page, 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 on top of the Configuration section to change the configuration settings.

For details about these settings, see Configuring a Custom Plan (on page 291).

NOTE:
It is not possible to change the scope of the plan in the Plan Page. You will need to start over if you want a different scope. To start over, go to the top-right section of the page, click the More options icon ( ), and then select New Plan.

When you are ready to re-run the plan, click Run Again on the top-right section of the page.
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 set which VM template to use in these calculations.

For information about viewing cluster headroom, see Viewing Cluster Headroom (on page 63).

To calculate cluster capacity and headroom, Workload Optimization Manager runs nightly plans that take into account the conditions in your current environment. The plans use the Economic Scheduling Engine to 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**.

   ![Plan Management](image)

   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**.
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 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 368).

About Placement of Reserved VMs

To determine the best placement for the VMs you want to reserve, Workload Optimization Manager runs a plan using the last-generated data in nightly-run headroom plans.

NOTE:
If you have changed 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, if you want 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 will have 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 will prevent the placement.

  Note that actual and reserved VMs share the same resources on providers. This means that provider capacity will change 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 will result in congestion. Workload Optimization Manager will not retry fulfilling the reservation.

- **Failed Initial Placements**
  
  If the initial placement attempt is unsuccessful (for example, if all providers have seen historical congestion), Workload Optimization Manager shows that the placement has failed and will not retry fulfilling the reservation.

### 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 currently 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.

   **NOTE:**

   In OpenStack environments, when you set up reservations to deploy workloads via OpenStack templates you must constrain the deployment to the OpenStack datacenters that support the given template.
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*.

- **VIRTUAL 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

Click the name to open the Reservation Settings fly-out

Expand the list entry for details

The PLACE page displays the current list of reservations. You can expand items in the list to see some details, or you can click to view the full details. You can also select items to delete them, which cancels the reservation or deployment.

For an entry in the RESERVED state, you can click the entry name to open the Reservation Settings fly-out.

To delete a reservation, select it in the list and click the DELETE icon.

To see details about the provider entities, or the datacenter that is hosting the reserved VMs, click that entity name.
Deploying Workloads to the Reserved Resources

When you reserve resources, you know that they will be available for you to deploy actual VMs in your environment. To deploy these VMs, you should:

1. Note the placement that your reservation has calculated.
   Expand the reservation entry in the Workload Placement page and note the hosts and storage that will provide resources for your VMs.

2. Delete the reservation.
   Before you deploy the reserved VMs, you should delete the reservation. This frees up the Workload Optimization Manager market to manage the placement of the VMs you are about to deploy.

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 will manage 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 reserved instances, 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. To view a dashboard, click its name in the list.

From the Dashboard page, you can also create your own custom dashboards.

**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
- 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 314).
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 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). It also shows time to exhaustion of cluster resources, which is useful for future 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.
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.
  - The **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.

**NOTE:**
For the Version 7 family of Workload Optimization Manager, cloud features are Early Access, only. These features appear in the user interface and the documentation, but this version does not enable them for use.
• 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 Reserved Instances

Reserved Instances (RIs) reduce cost by offering a subscription-based payment plan. Workload Optimization Manager discovers these RI plans and tracks usage patterns to identify workloads that are good RI candidates. The Cloud Executive Dashboard shows whether you are getting the most out of your current RI strategy.

The **RI Utilization** chart widget shows how well you have utilized the reservation inventory. The chart compares the capacity for all reservations versus the RI consumption by virtual machines.

The **RI Coverage** chart widget compares the capacity of your current VM workload to the capacity of workload that is covered by Reserved Instances.

---

**Container Platform Dashboard**

![Container Platform Dashboard](image-url)

---

**Top Container Platform Clusters**

Global Environment

- **Kubernetes-168-42**
  - Application Resource: 7.23 GHz (40%) | 3.45 GHz (19%)
  - VM: 49.24 GB

**Top Namespaces**

Global Environment

- **demapp**
  - Cluster: OKI-3.1...
  - VCPUUsed: 199.78 MHz (50%)
  - VCPURequest: 399.57 MHz (50%)

**Top Services**

Global Environment

- **Service-turbomc/3c-operator-m**
  - Application Resource Health: 100

---
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 49).

### Creating a Dashboard

To create a custom dashboard:

1. Navigate to the Dashboards Page.

   ![Dashboard Icon](image)

   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.

   **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 316).
5. Optionally, set the dashboard access.
   Click Gear to change the setting.
   Dashboard access can be:
   - **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.
   By default, access is set to **Only Me**.

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.

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 316).

To edit a dashboard’s name or change its access settings:
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 dashboard’s Edit fly-out, make your changes.

**Change the dashboard name**

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 fly-out panel.
   Your changes take effect when you close the fly-out.

**Deleting a Dashboard**

If you are an administrator or the dashboard owner, you can delete a custom dashboard. You cannot delete executive dashboards.

To delete 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.
2. Delete one or more dashboards.
   In the list, choose the checkbox for each dashboard you want 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 a variety of settings. For example, in the Top Utilized chart widget, if you choose 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.

   - **In a scoped view**
   - **On a dashboard**

   On a dashboard, click **Add Widget** at the top-right corner. In a scoped view, click **Add Widget** on the right above the charts.

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 fly-out opens.

3. Configure the settings for your chart widget.

   Chart widget settings determine the data that the chart widget will show.

   - In the Edit fly-out, choose the settings and click **Update Preview** to display the result in the Widget Preview pane.
   - 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 319)**.

   For example:
To delete a chart widget from your dashboard, choose **Delete** in the More options menu at the top-right corner of the chart widget.

**Methods to Access Chart Widget Settings**

Two methods exist for accessing the chart widget settings in the Edit fly-out:

- You can access the settings in the Edit fly-out 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 at the top-right corner.
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, additional settings may 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 fly-out.
  2. In the Select Scope fly-out, choose the entity, group, or account that you want.

    The ACCOUNTS tab is available depending on the type of chart widget.

    Your choice appears in the **Scope** field.

- **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’s 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 actions that Workload Optimization Manager recommends to improve the current state of your environment.

Chart Type

You can set the display to:

• Text
• Ring Chart
• Horizontal Bar
• List

Examples:

• Text
  The text chart shows the number of actions for each action type. It gives a quick visual indication of the kinds of actions that are pending. For details, see Action Types (on page 68).
Dashboards: Focused Views

- **Ring Chart**
  The ring chart counts the number of actions for each action type. It gives a quick visual indication of the kinds of actions that are pending. For details, see *Action Types (on page 68)*.

- **Horizontal Bar**
  The horizontal bar chart counts the number of actions for each action type. It gives a quick visual indication of the kinds of actions that are pending. For details, see *Action Types (on page 68)*
List

The list chart shows an abbreviated listing of the actions for the chart’s scope. For details about the different actions generated by the product, see Actions (on page 64).

At the bottom of the chart, click Show All Actions to see a full list of pending actions that are in the scope of the chart, along with action details and controls to execute actions. For details, see Pending Actions List (on page 79).

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

Stacked Bar Chart

For the Stacked Bar Chart, 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. The 2:00 PM bar, for example, 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 the chart 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.

Tabular Chart
To see the full list (on page 79) of actions, click Show All at the bottom of the chart.
As you execute the actions Workload Optimization Manager has recommended, you improve your environment’s health and avoid risks to performance or cost. These charts show how many risks you have avoided over time. For example, the charts can show how many over-provisioning and congestion risks you avoided.

**Chart Type**

You can set the display to:
- Text
- Ring Chart
- Horizontal Bar

**Optimized Improvements Charts**

Workload Optimization Manager automatically executes or recommends actions, depending on the policies that you set up. For the recommended actions, you can use Optimized Improvements charts to show how utilization of resources would change assuming you accept all of the pending actions ([on page 320](#)).

**Entity Type**

Entity types you can choose include:
- Business Applications
- Business Transactions
- Services
- Application Components
- Chassis
- Containers
- Container Pods
- Container Specs
- Namespaces
- Workload Controllers
- Data Centers
- Databases
Dashboard: Focused Views

- Database Servers
- Disk Arrays
- IO Modules
- Internet
- Logical Pool
- Networks
- Hosts
- Regions
- Storage Devices
- Storage Controllers
- Switches
- Virtual Data Centers
- Virtual Machines
- Volumes
- Zones

Commodity
Depending on the entity type, you can add different resource commodities that you want to measure. For a chart of Hosts, you can measure commodities such as CPU, Memory, and even network flow between VMs that are on the same host (In-Provider Flow) or on other hosts (In-DPOD or Cross-DPOD Flow).

Display
Optimized Improvements 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.

Example: An Optimized Improvements chart for applications

![Applications Optimized Improvements](image)

Potential Savings or Investments Charts
These charts show potential savings or necessary investments 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 some workloads risk losing performance, Workload Optimization Manager might recommend scaling actions for the virtual machine to increase resources. The Necessary Investments chart shows how these actions translate to an increase in expenditure.

Workload Optimization Manager 3.0.7 User Guide
On the other hand, if there are pending actions to scale a virtual machine, which result in reduced monthly costs, the Potential Savings chart shows the reduced cost that would result from those actions.

This chart also track RI optimization actions. Virtual machine scaling actions may result in a freed up RI, which can now be applied to a different virtual machine. RI optimization actions reflect the potential savings resulting from reassigning the RI to a different virtual machine. These actions are not executed by Workload Optimization Manager users. They reflect RI reassignment, which the cloud provider will take care of.

The projected amounts include on-demand costs for VMs. For information about on-demand cost calculations, see Estimated On-demand Monthly Costs for Cloud VMs (on page 26).

**Type**

You can choose Potential Savings or Necessary Investments.

**Chart Type**

You can set the display to:

- Text
- Ring Chart
- Horizontal Bar

For the ring chart, you can click an action type (for example, Scale Volumes) in the chart or legend to display a filtered view of the actions list.

**Show All**

Click Show all at the bottom of the chart to see a breakdown of savings or investments by action/entity type and entity. By default, the actions are shown in order of largest amounts so you can easily identify which ones will incur the highest costs or introduce the most savings.

For example, you can see the savings you would realize if you execute all Scale actions on the virtual machines included in the chart’s scope.

The table then breaks down the total savings by individual virtual machines, and includes links to the specific actions that you need to perform to realize those savings.

You can also compare instance types, costs, and RI coverage before and after executing the actions, allowing you to easily identify actions with the most savings.
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 your environment, by entity type. For example, you can choose to show the health of all hosts in your environment, or the health of all the workloads running on a public cloud region.

Entity Type

Entity types you can choose include:

- Business Applications
- Business Transactions
- Services
- Application Components
- Chassis
- Containers
- Container Pods
- Container Specs
- Namespaces
- Workload Controllers
- Data Centers
- Databases
- Database Servers
- Disk Arrays
- IO Modules
- Internet
- Logical Pool
- Networks
- Hosts
- Regions
- Storage Devices
- Storage Controllers
- Switches
- Virtual Data Centers
- Virtual Machines
- Volumes
- Zones

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 resources for the selected entity type, showing their allocated capacity and how much of that capacity is in use.

Entity Type
Entity types you can choose include:

- Business Applications
- Business Transactions
- Services
- Application Components
- Containers
- Container Pods
- Container Specs
- Namespaces
- Workload Controllers
- Data Centers
- Database Servers
- Disk Arrays
- Logical Pool
- Networks
- Hosts
- Regions
- Zones
- Storage Devices
- Storage Controllers
- Virtual Machines
- Volumes
Commodity
Depending on the entity type, you can add different resource commodities that you want to measure. For example, for a chart of Virtual Machines, you can measure 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. It could take up to 40 minutes for the correct values to display.

**Display**
The chart shows the information as Tabular.

**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.

**Entity Type**
Entity types you can choose include:
- Business Applications
- Business Transactions
- Services
- Application Components
- Containers
- Container Pods
- Container Specs
- Namespaces
- Workload Controllers
- Data Centers
- Database Servers
- Disk Arrays
- Logical Pool
- Networks
- Hosts
- Regions
- Zones
- Storage Devices
- Storage Controllers
- Virtual Machines
- Volumes

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 on the left 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.

Commodity

You can set a Resources chart to one of the following resources:

- Ballooning
  Percentage of the host’s ballooning capacity that is in use.
- CPU / CPU Allocation / CPU Provisioned
  Percentage of CPU cycles that are devoted to processing instructions.
- Connection
  The connections in use, as a percentage of the maximum connections allowed on the database. Database configuration determines the capacity for this resource.
- DBCacheHitRate
  Percentage utilization of the database server’s allocated cache hit rate, where a greater value indicates fewer disk reads for data.
- DBMem
  The memory in use by the database, as a percentage of the allocated capacity. Database configuration determines the capacity for this resource. Note that for databases, Workload Optimization Manager uses this resource to drive actions, instead of the VMem on the hosting VM. This means that actions are driven by the actual memory consumption on the database.
- Disk Array Access
- Host LUN Access
- **Heap**
  The heap capacity allocated for an application. Charts show the percentage of capacity that is used by an Application Component.
- **IO Throughput**
  Data rate through the host’s IO adapter, measured in KBytes/sec.
- **Memory / Memory Allocation / Memory Provisioned**
  Host memory, measured in Kbytes.
- **Number of Disks**
- **Number of Replicas**
  For details, see [Number of Replicas chart (on page 332)].
- **Net Throughput**
  Data rate through the host’s Network adapter, measured in Kbytes/sec.
- **Port Channel**
- **Q16 / Q1 / Q2 / Q32 / Q4 / Q64 / Q8 VCPU**
- **Remaining GC Capacity**
  Percentage of CPU time not spent on garbage collection (GC).
- **Response Time**
  Percentage utilization of a resource’s allocated response time.
- **SLA Commodity**
- **Storage Access**
  Storage access operations per second.
- **Storage Allocation**
- **Storage Amount**
- **Storage Latency**
  Percentage of allocated latency that is in use on the storage.
- **Storage Provisioned**
  How much the given storage is over-subscribed. Storage Provisioned capacity is the storage capacity multiplied by the Storage Overprovisioned Percentage (200 by default). The higher this value, the greater the risk that storage is over-committed.
- **Swapping**
  The rate of memory swapping to disk, in bytes per second.
- **Tenancy Access**
- **Threads**
  Percentage of thread capacity that is consumed by an Application Component.
- **Transaction**
  Percentage of an application’s allocated transaction capacity that is in use.
- **TransactionLog**
  The disk space devoted to transaction logging for a database.
- **Virtual CPU**
  The allocated CPU capacity, measured in MHz.
- **Virtual Memory**
  The allocated memory capacity, measured in Kbytes.
  Note that percentages of allocated VMem are measured against the VMem limit (if set) or the allocated VMem capacity, whichever is less. This is also true in reports and recommended actions. For example, assume a VM with allocated VMem of 8 GB, but a limit of 4 GB. In this case, the percentage in a chart shows the percentage utilized of 4 GB.
- **Virtual Storage**
  The allocated virtual storage capacity, measured in Kbytes.
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.

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*

- Kubernetes **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.

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 334)** (public cloud)
- Business Applications
- Business Transactions
- Services
- Application Components
- Zones
- Chassis
- **Clusters (on page 334)** (of hosts)
- Containers
- Container Pods
Dashboards: Focused Views

- Container Specs
- Namespaces
- Workload Controllers
- Data Centers
- Databases
- Database Servers
- Disk Arrays
- IO Modules
- Internet
- Logical Pool
- Networks
- Hosts
  - Resource Groups (on page 335)
- Regions
- Storage Devices
- Storage Controllers
- Switches
- Virtual Data Centers
- Virtual Machines
- Volumes
- Wasted Files

**Data Type**
Depending on the entity type (for example, 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 lists the top entities by utilization of the commodities that you or the system has 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. For cloud entities, the estimated cost for those entities also display.

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 clusters chart which can be sorted by CPU headroom or CPU exhaustion.
Top Clusters Chart

This chart shows the top 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). It also shows time to exhaustion of cluster resources, which is useful for future 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 use the Economic Scheduling Engine to 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 ACTIONS button for a given cluster to see the actions that Workload Optimization Manager recommends to keep cluster resources in the desired state, and then decide which ones are safe to execute.

Click Show All to see all of the clusters. In the Show All list, you can also download capacity data as a CSV file. Click a cluster name to set the scope to that cluster and view more details about its current capacity and health.

Top Accounts Chart

This chart lists all the managed accounts in your cloud environment. It highlights the estimated monthly cost for each account 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. You can also click an account name to set the scope to that account.

Click Show all to view additional information, including the total estimated monthly cost for all accounts, and the number of actions that have been executed for individual accounts or workloads, along with the resulting savings. You can also download the accounts list as a CSV file.

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.

For AWS GovCloud environments, you will see all the GovCloud accounts (master and member accounts) and the associated standard accounts that you have added as targets. The standard accounts are not required, but adding them lets you see the projected billed costs for your GovCloud workloads. For Azure Government, you will see the subscriptions discovered via the
service principal and EA targets that you have added. For more information, see Support for Government Workloads (on page 35).

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 bottom of the chart.

Environment Charts
Environment charts provide an overview of your environment. They show the targets that you are managing and count the entities that Workload Optimization Manager has discovered through those targets. For example, you can display the cloud service providers, hypervisors, and the number of workloads.

**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 Text chart type.

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.

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.

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.

NOTE:
In order for Workload Optimization Manager to access AWS monthly reports, you must have created a Cost and Usage report in your AWS account and you must store it in an S3 bucket.

Billing Breakdown Charts
Billing Breakdown charts show your expenditure on cloud services, so you can track overall cost, cost by region, or cost by cloud accounts. Workload Optimization Manager discovers pricing for cloud services through the cloud accounts and Azure subscriptions that you configured as targets.
Chart Type
You can set the display to:

- Text
- Ring Chart
- Horizontal Bar

Expenses Charts
To help you manage costs for your public cloud environment, Cisco tracks compute, storage, license, and IP costs for the workloads in your environment. Are you spending too much on your cloud resources? You can see how your expenses evolve and keep track of these costs over time.

Cisco uses the cost for services and workload expenses to track your cloud spend. See Tracking Cloud Cost (on page 24) for more information about service cost data, compute, storage, license, and IP costs.

Commodity
You can choose:

- **Expenses**
  See your hourly expenses over time, as well as overall monthly and yearly costs.

- **Average Expenses**
  See your average cost per Virtual Machine, as well as overall monthly and yearly costs.

- **Billed Cost by Service Provider**
  See costs over time for each cloud service provider that you use in your cloud environment. For example, you can compare the costs you incur on AWS to costs on Azure.

  You can open more than one account from a single service provider. If you are running workloads on different service providers, then this chart shows the distribution of costs across them.

- **Billed Cost by Account**
  See costs over time for each account that you have set up as a target in Workload Optimization Manager.

  Each public cloud target that you configure for Workload Optimization Manager represents a public cloud account. You can choose one or more specific accounts depending on your configured public cloud targets. This chart gives you a quick read out of your costs. You can see whether one account shows unusually high cost, or perhaps an account is hardly used at all and you can consider closing it down.

- **Billed Cost by Service**
  This chart shows costs over time for each cloud provider service that you use in your cloud environment. For example, you can see costs for your Amazon EC2 instance or AWS Cost Explorer.

  To evaluate your use of different services, you can follow your expenditure for each one. Note that for AWS clouds the service names begin with "Amazon" or "AWS". Other services show the names as they are presented in the service provider’s billing report. You can also set the scope of this chart to an Azure Resource Group or a group of Resource Groups.

- **Workload Cost Breakdown**
  This chart shows costs over time for each component of your cloud utilization. The vertical line indicates when the last data point was polled from your environment. Data points to the right of the vertical line are projections into the future.

  You can see costs for:
  - On-Demand Compute
  - RI Compute
  - Spot Compute
  - On-Demand License
  - Reserved License
  - Storage
  - IP (static IPs for workloads)
This chart reflects on-demand costs for VMs based on uptime and other factors. For details about on-demand cost calculations, see Estimated On-demand Monthly Costs for Cloud VMs (on page 26).

Chart Type
You can set the display to:
- Line Chart
- Stacked Bar Chart
- Area Chart

Reading a Cost Breakdown Chart
The chart tracks cost over time and displays a tooltip with the date for the data point and the given values.

Cost information comes from billing reports. As you change the time scale, Workload Optimization Manager divides the reported information into the appropriate time units to match that scale. However, the source remains the same - Changing the scale does not affect the source data, or increase data polling.

Cloud Tier Breakdown Charts
Cloud Tier charts show the cloud tiers that Workload Optimization Manager discovers for the chart widget scope. For example, if the Chart Widget Scope is set to All Cloud VMs and the Entity Type is set to Virtual Machine, the chart shows all the cloud tiers that the workloads use.

Entity Type
You can choose any entity type in the list.

Chart Type
You can set the display to:
- Text
- Ring Chart
- Horizontal Bar

Location Charts
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 Charts
Cost Breakdown By Tag charts show the costs for tagged cloud entities and commodities that Workload Optimization Manager discovered.

Tag Settings
Choose a key from the list and one or more corresponding values.

Custom X-axis
You can also define an X-axis using one of two commodities, Cost Breakdown by Zone or Cost Breakdown by Region, and choose one or more values for the commodity.

Chart Type
You can set the display to:

- Area Chart
- Stacked Bar Chart

Example: The Tag setting key is workload-type and the values are other and production.

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 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 shape 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.
**Dashboards: Focused Views**

**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, 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 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 temporarily stops for the hours that a workload is offline, and then resumes when the workload is online again and is running with the same configuration.

- Calculation stops for terminated workloads or workloads that Workload Optimization Manager no longer discovers.

**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:
- Text and Area Chart
- Area Chart
- Text

You can also edit the chart to switch between the Cumulative Savings and Cumulative Investments views.

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.

**Recommended RI Purchases Charts**

Recommended RI Purchases charts show the projected inventory of pending Reserved Instance purchases as generated by Workload Optimization Manager. The charts show the Reserved Instance workloads that Workload Optimization Manager discovers, and lists them by the available templates.

To see the RI information for each template, click Show all at the bottom of the chart. If your scope includes both AWS and Azure cloud targets, click AWS or Azure. Click any column heading to sort the list. For example, you can sort the list by the break-even period (The time at which RI savings will exceed the purchase cost of the RI, rounded to the month). When you choose one or more checkboxes, the total count, up-front cost, and savings appear at the top.

**Chart Type**

You can set the display to:

- Text
- Ring Chart
- Horizontal Bar
Examples:

• Horizontal Bar

  Recommended RI Purchases

  - 14 Standard_B1s
  - 13 c3a.nano
  - 10 Standard_B1ms
  - 10 Standard_E2s_v3
  - 46 Other Templates

  Show all

• Viewing the AWS Show all list

<table>
<thead>
<tr>
<th>AWS</th>
<th>AZURE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Buy Actions</strong></td>
<td>20</td>
</tr>
<tr>
<td><strong>Type to search</strong></td>
<td></td>
</tr>
<tr>
<td>□ Amount</td>
<td>□ Instance Type</td>
</tr>
<tr>
<td>□ Advanced</td>
<td>r3a.large</td>
</tr>
<tr>
<td>□ Advanced</td>
<td>c5a.large</td>
</tr>
<tr>
<td>□ Advanced</td>
<td>r3a.large</td>
</tr>
</tbody>
</table>

• Viewing the Azure Show all list

<table>
<thead>
<tr>
<th>AWS</th>
<th>AZURE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Buy Actions</strong></td>
<td>27</td>
</tr>
<tr>
<td><strong>Type to search</strong></td>
<td></td>
</tr>
<tr>
<td>□ Subscription</td>
<td>□ Product Name</td>
</tr>
<tr>
<td>□ Dev</td>
<td>Standard_D2s_v3</td>
</tr>
<tr>
<td>□ Dev</td>
<td>Standard_B1s</td>
</tr>
<tr>
<td>□ Dev</td>
<td>Standard_DS1_v2</td>
</tr>
</tbody>
</table>

**RI and Savings Plans Coverage Chart**

This chart shows the percentage of VMs covered by Reserved Instances (RI) and AWS Savings Plans. If you have a high percentage of on-demand VMs, you should be able to reduce your monthly costs by increasing coverage. To increase coverage, you scale workloads to instance types that have existing capacity.
NOTE:
For cloud environments that only have RI data, the chart’s name is **RI Coverage**. Similarly, for environments with only Savings Plans data, the chart’s name is **Savings Plans Coverage**.

See the following topics for more information:

- [RI Coverage](on page 343)
- [Savings Plans Coverage](on page 344)

## RI Coverage Chart

This chart shows the percentage of VMs covered by Reserved Instances (RI). If you have a high percentage of on-demand VMs, you should be able to reduce your monthly costs by increasing RI coverage. To increase coverage, you scale workloads to instance types that have existing RI capacity. If you need more RI capacity, then Workload Optimization Manager will recommend RI purchase actions.

### Scope

If you set the scope to a specific AWS account, this chart shows the RI coverage for the workloads for the account, plus any RIs for the billing family. For Azure, if you set the scope to a specific Azure subscription, this chart shows the RI coverage for the workloads for the subscription, plus any shared RIs and single-scope RIs owned by this subscription.

### Data Points

The data point on the solid vertical line shows the latest data that was polled from your environment. Data points to the left of the vertical line represent historical data, while data points to the right are projections into the future.

Hover on a data point in the chart to see the following information:

- The date and time for the data point
- The percentage of RI coverage, based on NFUs or ratios.
  - **NFU** (AWS only)
    
    NFU 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 these normalized factors. The chart shows the number of RIs calculated as NFUs that cover workload capacity compared to the total number of NFUs for the workloads in the chart’s scope. Each workload is assigned normalized factor units depending on its instance type.
  - **Ratio** (Azure only)
    
    Ratio refers to the number of RI units that cover workload capacity compared to the total number of RI units for the workloads in the chart’s scope. Each workload is assigned RI units based on its instance type.
Savings Plans Coverage Chart
This chart shows the percentage of VMs covered by AWS Savings Plans. If you have a high percentage of on-demand VMs, you should be able to reduce your monthly costs by increasing Savings Plans coverage. To increase coverage, you scale workloads to instance types that have existing Savings Plans capacity.

Scope
To view data for AWS Savings Plans, you must:

- Set the scope to your global environment.
- Choose a timeframe that shows daily or monthly data points (Last 7 Days, Last 30 Days, or Last Year)

AWS measures your Savings Plans commitment in $/hour, but shows daily and monthly costs in Cost Explorer. Workload Optimization Manager polls Cost Explorer periodically to obtain the latest cost data, and then uses that data to calculate Savings Plans utilization or coverage per day. For this reason, you will not see Savings Plans data if you choose a timeframe that shows hourly data points (Last 2 Hours or Last 24 Hours).

NOTE:
AWS timestamps data in UTC, but the chart presents data in your local time. This difference could result in the appearance of a missing day in the chart, but has no effect on data completeness (the chart always reflects the complete data set).

Data Points
The chart shows historical data, represented by data points to the left of the solid vertical line. Data for the current day is not available since the latest data that AWS provides is always a few days old. In addition, Workload Optimization Manager does not project coverage into the future.

Hover on a data point in the chart to see the following information:

- The date and time for the data point
- The percentage of Savings Plans coverage

RI and Savings Plans Inventory Chart
This chart shows the Reserved Instances and Savings Plans in your cloud environment (including AWS GovCloud and Azure Government (on page 35)) and lists them by the instance types they use.
NOTE: For cloud environments that only have RI data, the chart’s name is RI Inventory. Similarly, for environments with only Savings Plans data, the chart’s name is Savings Plans Inventory.

See the following topics for more information:

- RI Inventory (on page 345)
- Savings Plans Inventory (on page 348)

RI Inventory Chart

This chart shows the Reserved Instances (RIs) in your cloud environment (including AWS GovCloud and Azure Government (on page 35)) and lists them by the instance types they use.

NOTE:
It could take Workload Optimization Manager up to a day to discover newly purchased Azure RIs.

Chart Type

RI Inventory
Global Environment

- 12 t2.micro
- 6 Standard_B1ms
- 6 Standard_B1s
- 5 Standard_DS2_v2
- 5 t3.nano
- 28 Other

SHOW ALL
You can set the display to:

- Text
- Ring Chart
- Horizontal Bar

Show All

Click **Show All** at the bottom of the chart to see detailed information for the RIs in scope. If your scope includes both AWS and Azure accounts, each provider will have its own tab.

Each row in the table corresponds to an RI. Table columns show basic information obtained from the cloud provider, such as reservation/order ID, instance type name and count, location, term, and expiration date. Workload Optimization Manager calculates estimated utilization and savings, and effective cost.

The table also supports the following general functionality:

- **Totals**: At the top of the page, you will see the total RI count, cost, and savings. As you select one or more checkboxes, the information changes to reflect the totals for your selections.
- **Column Sorting**: Click any column heading to sort the list.
- **Download**: Click the Download icon at the upper right section of the page to download the table as a CSV file.

The table presents richer details when you set the scope to the global environment. For details, see the next section, **Full RI Inventory**.

**Full RI Inventory**

When you add an Azure EA account or an AWS master account as your primary cloud target, Workload Optimization Manager gains full insight into the RIs for your billing families. Even as you selectively add Azure subscriptions or AWS 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 actions for the cloud environment that it discovered.

Set the scope to your global environment to view your full RI inventory.

**NOTE:**
If you do not add an AWS master account, Workload Optimization Manager always shows a partial RI inventory, regardless of scope.
When you click **Show All** at the bottom of the chart, pay attention to the following information shown in the table:

- **For RIs in added accounts (Azure subscriptions or AWS member accounts):**

  - **Subscription column (for Azure) or Account column (for AWS)**
    
    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 RI Inventory page.

  - **Linked VMs column**
    
    This column shows how many VMs in the account use the RI. Click the number to view the VMs.

    A star symbol (*) after the number indicates that there are VMs from non-added accounts that also use the RI. Since you have not added those accounts, the exact number of VMs is unknown.

  - **Estimated utilization column**
    
    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).

- **For RIs in non-added accounts:**

  - **Account column (for AWS) or Subscription column (for Azure)**
    
    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 master or EA account.

  - **Linked VMs**
    
    If the number under this column is 1 or more, the number indicates VMs from added accounts that use the RI. Click the number to view the VMs.

    If the number is 0 (zero), and is not followed by a star symbol (*), then the RI is currently not used anywhere.

    A star symbol (*) after the number indicates that there are VMs from other non-added accounts that also use the RI. Since you have not added those accounts, the exact number of VMs is unknown.

  - **Estimated utilization column**
    
    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).

**NOTE:**

In Azure environments, 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 RIs from non-added Azure subscriptions.
Partial RI Inventory
Workload Optimization Manager shows a partial RI inventory if:

- You added some of your AWS member accounts as targets, but not an AWS master account.
  In this case, Workload Optimization Manager will not reflect RIs for member accounts that you have not added as targets.
- You change the scope (in the RI Inventory chart) from your global environment to individual AWS member accounts or Azure subscriptions.
  In this case, Workload Optimization Manager will only reflect RIs for the selected account or subscription.

Savings Plans Inventory Chart
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 (on page 35)) and the instance types they use.

Chart Type
You can set the display to:

- Text
- Ring Chart
- Horizontal Bar

Show All
Click Show All at the bottom of the chart to see detailed information for the Savings Plans in scope.

Each row in the table corresponds to a Savings Plan. Table columns show basic information obtained from AWS, such as Savings Plan ID/type, instance family, location, term, and inclusive dates.

<table>
<thead>
<tr>
<th>Savings Plan ID</th>
<th>Account</th>
<th>Type</th>
<th>Payment</th>
<th>Instance Family</th>
<th>Location</th>
<th>Commitment</th>
<th>Term</th>
<th>Start Date</th>
<th>Exp. Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>55555</td>
<td>Prod</td>
<td>Compute</td>
<td>All Upfront</td>
<td>All</td>
<td>All</td>
<td>$0.001/hr</td>
<td>1 Year</td>
<td>Dec 28, 2020</td>
<td>Dec 28, 2021</td>
</tr>
<tr>
<td>45555</td>
<td>Prod</td>
<td>EC2</td>
<td>No Upfront</td>
<td>t3</td>
<td>aws...</td>
<td>$0.001/hr</td>
<td>3 Years</td>
<td>Dec 29, 2020</td>
<td>Dec 25, 2023</td>
</tr>
</tbody>
</table>

RI and Savings Plans Utilization Chart
This chart shows how well you have utilized your current RI and AWS Savings Plans inventory (on page 344). The desired goal is to maximize the utilization of your inventory and thus take full advantage of the discounted pricing offered through RIs and Savings Plans.

NOTE:
For cloud environments that only have RI data, the chart’s name is RI Utilization. Similarly, for environments with only Savings Plans data, the chart’s name is Savings Plans Utilization.

See the following topics for more information:
- RI Utilization (on page 348)
- Savings Plans Utilization (on page 349)

RI Utilization Chart
Use this chart to see how well you have utilized your current Reserved Instance inventory. The desired goal is to maximize the utilization of your inventory and thus take full advantage of the discounted pricing offered through Reserved Instances.
Scope
You can set the scope to your global cloud environment or to individual accounts/subscriptions, billing families, or regions. Scoping to an account/subscription shows the RI utilization for the workloads for the entire billing family or for single and shared subscriptions.

Data Points
The data point on the solid vertical line shows the latest data that was polled from your environment. Data points to the left of the vertical line represent historical data, while data points to the right are projections into the future.

Hover on a data point in the chart to see the following information:

- The date and time for the data point
- The percentage of RI utilization, based on NFUs or ratios.
  - NFU (AWS only)
    NFU 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 utilization in terms of these normalized factors. The chart shows the number of RIs calculated as NFUs that are consumed by cloud workloads compared to the total number of RIs in the chart’s scope.
  - Ratio (Azure only)
    Ratio refers to the number of RI units in use compared to the total number of RI units in the chart’s scope. Each workload is assigned RIs based on its instance type.

Savings Plans Utilization Chart
Use this chart to see how well you have utilized your current Savings Plans inventory. The desired goal is to maximize the utilization of your inventory and thus take full advantage of the discounted pricing offered through Savings Plans.

Scope
To view data for AWS Savings Plans, you must:
- Set the scope to your global environment.
- Choose a timeframe that shows daily or monthly data points (Last 7 Days, Last 30 Days, or Last Year)
  AWS measures your Savings Plans commitment in $/hour, but shows daily and monthly costs in Cost Explorer. Workload Optimization Manager polls Cost Explorer periodically to obtain the latest cost data, and then uses that data to calculate Savings Plans utilization or coverage per day. For this reason, you will not see Savings Plans data if you choose a timeframe that shows hourly data points (Last 2 Hours or Last 24 Hours).

NOTE:
AWS timestamps data in UTC, but the chart presents data in your local time. This difference could result in the appearance of a missing day in the chart, but has no effect on data completeness (the chart always reflects the complete data set).
Dashboards: Focused Views

Data Points
The chart shows historical data, represented by data points to the left of the solid vertical line. Data for the current day is not available since the latest data that AWS provides is always a few days old. In addition, Workload Optimization Manager does not project utilization into the future.

Hover on a data point in the chart to see the following information:

- The date and time for the data point
- The percentage of Savings Plans utilization, based on the total utilized and committed costs per day

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.

Storage Summary Charts
To help you manage your costs on the public cloud, these charts show the distribution of storage for the given scope, cost, potential savings, and information about unattached storage. In this way, you can see how storage utilization affects your costs. For these charts, Workload Optimization Manager calculates the costs based on the cost information from the cloud targets.

For a detailed breakdown, click Show all at the bottom of the chart. If your scope includes both AWS and Azure cloud targets, click AWS or Azure to see the details. Click any column heading to sort the list. When you choose one or more checkboxes, the total appears at the top.

NOTE:
For Azure environments with VMs in Scale Sets, for any VMs that are powered off, the associated storage 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 storage capacity is currently utilized.

Chart Unit
Choose one of the following:
- Count to see how many storage tiers or volumes exist by storage type.
- Cost to see the monthly cost by storage type.
Chart Displays

Examples:

- **Costs**
  The chart shows the monthly costs for all storage tiers or volumes. You can also choose **Count** to list how many storage tiers or volumes exist by storage type. This display is available for real-time views and dashboards.

- **Unattached Storage**
  The chart shows how many unattached storage tiers or volumes exist. You can also choose **Cost** to list the monthly costs of the unattached storage. This chart is available for real-time views and dashboards.

Monthly Savings or Investments Totals Charts

Monthly Savings or Investments Totals charts help you examine the monthly savings or investments for executed cloud actions. For example, if an executed action causes an increase in the price, this is an investment. These charts also show the missed monthly savings or missed performance investments that you could have achieved for recommended cloud actions, if you executed them.

For this chart’s scope, you can choose an account or subscription, a group of accounts or subscriptions, or use the default, Global Environment. If you use the default Global Environment, the chart will automatically use all cloud accounts for its scope. Other examples of scope settings are: An AWS billing family, an Azure subscription, the All AWS Accounts predefined group, or the All Azure Accounts predefined group.

For all actions except Suspend, savings and investments are estimated based on the hourly cost of workload price differences and 730 hours per month of workload usage. Savings from Suspend actions are estimated based on the hourly cost of workload price differences and actual suspend times as defined in the suspension policy.

Missed savings and investments are estimated based on the hourly cost of workload price differences and the number of hours that recommended actions exist in the system.

Monthly Savings or Investments Totals charts calculate data on a monthly basis since your update of Workload Optimization Manager to version 2.3.0. Historical data stored in the database prior to version 2.3.0 is not included.

Chart Type

You can set the display to:

- **Stacked Bar Chart**
- **Tabular**

Examples:

- **Stacked Bar**
  This chart shows the monthly totals of savings or investments for each of the last seven days. It also shows the missed monthly savings or performance investments that you could achieve by executing recommended cloud actions.
In the chart legend, you can also choose an item to change the display of the chart. Click the item again to reset the chart. For example, if you want to examine investment information, click Investments in the legend.

- **Tabular**
  This chart shows the monthly totals of savings or investments for each of the last seven days. It also shows the missed monthly savings or performance investments that you could achieve by executing recommended cloud actions.

---

## 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.
Chart Type
You can set the display to:

- Stacked Bar Chart
- Line Chart

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 use the Economic Scheduling Engine to 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.

You can specify the following types of Headroom charts:

- CPU Headroom
- Memory Headroom
- Storage Headroom

Commodity
You can choose:

- CPU Headroom
- Memory Headroom
- Storage Headroom

Display
The chart shows the information as an Area chart.
Example:
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.

Display

The chart shows the information as a Text chart.
Embedded Reporting

The Workload Optimization Manager platform 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 via 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 via a set of standard dashboards and reports. You can create your own dashboards and reports to focus on other areas of concern.

Dashboards and charts 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. If you are new to Grafana and need help getting started, read the documentation available at:

https://grafana.com/docs/grafana/latest/

Enabling Embedded Reporting

To view and create reports, you must enable Embedded Reporting as part of the Workload Optimization Manager installation process. For details, see the Installation Guide.

If you have enabled Embedded Reporting in a Workload Optimization Manager instance that is older than version 8.1.1 (7.22.x to 8.1.0), and then updated that instance to version 8.1.1 or later, be aware of the following:

- Version 8.1.1 or later includes a new Embedded Reporting schema that provides additional reporting capabilities.
- Reports will begin their historical data from the time that you completed the update.
- The Embedded Reports page includes a folder named Old Reports, which is marked with the date of the update. This folder contains all of your reports before the update, using historical data from before the change to the new schema.
- The On-Prem Virtual Machine Rightsizing Recommendations report appears in the Old Reports folder, but it is empty. The table used by this report has been rewritten as part of the new schema, and historical data for the report is not saved after you update.
- If you want to merge the old reports with your new reports, please contact your support representative for assistance.

Report Editor Role

To manage reports, a user must have the Report Editor role and must not be a scoped user. Only one user per instance is allowed to have this role (by default, the local administrator user). You can grant Report Editor privileges to another user, from the User Management page.

**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.
Embedded Reporting Main Page

When you have enabled Embedded Reporting, the Workload Optimization Manager Navigation Bar displays the REPORTS icon.

Click this icon to open the Embedded Reports page in a new browser tab. This tab displays the Search dashboards by name view.

You can search for specific dashboards or browse folders to find the dashboards you want. You can also create custom reports.

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.
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. These groups include PM clusters, and entities grouped by different logical boundaries. For example, Workload Optimization Manager discovers Storage by Disk Array, Physical Machines by Datacenter, and VMs by Network. In addition, Workload Optimization Manager discovers pools such as virtual datacenters, or folders that implement specific HA policies.

You can also 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 physical machines 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.

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.
Creating Groups

This page lists all the custom groups that you currently have configured for Workload Optimization Manager. You can:

- Expand an entry to see group details
- Select an entry to delete the group
- Click a group name to edit it
  - For a dynamic group, you can edit the set of criteria that select the group members. For a static group, you can add or subtract specific members.
- 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.

3. 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.

4. Create a new group.
Click **NEW GROUP**.

Next, choose a group type.

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.

  - Specify group criteria.

  These criteria are entity attributes that determine group membership. You might create a group of all VMs that have 4 VCPUs. You can choose properties of the member entities, and you can choose properties of entities that are related to the members. For example, you can make a group of VMs that are hosted by PMs with the substring "Development" in their names.

  As you set criteria, the list of entities updates to show the member entities. You also can sort the list by severity (per the most critical entity in group) or group name.

  Note that you can use regular expression to express your match strings.

- When you are finished, save the group.

  **Save** adds this group to the **My Groups** collection.
Working With Schedules

Workload Optimization Manager schedules specify a specific time range during which certain events can occur. Workload Optimization Manager currently uses schedules in scoped policies to set up windows of time when the policy can execute certain actions, or when the policy changes settings that affect analysis and action generation.

**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.
The Schedules page lists all the currently defined schedules. From this page you can:

- Select an entry to delete the schedule.
- Select an entry to defer the next occurrence.

Workload Optimization Manager calculates when the next scheduled window will open. If you want cancel the scheduled occurrence one time, you can select the schedule and defer the upcoming occurrence. This defers the schedule wherever it is applied. If the schedule is applied to more than one policy, this will defer all the policies that use this schedule. Before you defer a schedule, you should expand the details and review all the policies that use this schedule.

- Expand an entry to see schedule 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.

- Create new schedules
  See [Creating Schedules](on page 362).

### Deleting 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, or...
- 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 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.

### Creating Schedules

To create a new schedule:

1. Navigate to the Settings Page.

   ![Settings Icon](settings-icon.png)

   Click to navigate to the Settings Page. From there, you can perform a variety of Workload Optimization Manager configuration tasks.

2. Choose Schedules.

   ![Schedule Icon](schedule-icon.png)

   Click to navigate to the Schedule Management Page.

   This page lists all the schedules that you currently have configured for Workload Optimization Manager. You can edit the schedules in the list, or you can create new schedules.
3. Create a new schedule.

Click **New Schedule** to open the new schedule fly-out. Then name 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**

**RECURRENCE**

Weekly

**REPEAT EVERY**

1 weeks

**ON**

Mo Tu Wd Th Fr Sa Su

**START DATE**

Wed Jun 19 2019

**END DATE**

None

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**

**RECURRENCE**

Monthly

**REPEAT EVERY**

1 months

**ON**

First Saturday

**START DATE**

Wed Jun 19 2019

**END DATE**

None

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. The Workload Optimization Manager server uses that reference when it opens and closes the schedule window. Users see the same time zone setting no matter where they are located – They should convert the schedule time to their local time if they want to track when the schedule opens in their working day.

7. **When the settings are complete, save the schedule.**
Templates: Resource Allocations for New Entities

Workload Optimization Manager uses templates to describe new entities that it will deploy 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 given 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).

The Template Catalog shows all of the templates that have been specified or discovered for your installation of Workload Optimization Manager. From this page, you can also create new templates and edit existing ones.

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 Template Catalog shows all of the templates that have been specified or discovered for your installation of Workload Optimization Manager. From this page, you can also create new templates and edit existing ones.

Creating and Editing Templates

To create a new template, navigate to the Template Catalog and click NEW TEMPLATE. To edit a template, click the template’s name. When you create a new template, the first step is to choose the entity type.

1. Navigate to the Settings Page.

2. Choose Templates.

3. Create or edit a template
   To create a new template, navigate to the Template Catalog and click NEW TEMPLATE. To edit a template, click the template’s name.

4. If you’re creating a new template, choose the entity type.

5. Make the settings for your template.
For each type of template, you set allocations for different resources. You can make templates of the following types:

- Virtual Machine
- Host
- Storage
- Container

6. Make the settings for your template, and then save your changes.

When the template window opens, it displays the most common resource settings. You can expand the settings to see the full collection for that template type.

7. Save your changes.

After you have made your settings and named the template, click CREATE or SAVE.

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 datacenter. 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 running hardware replacement plans.

The host template is a collection of these settings:

- **CPU**
  The processor for this host model. Note that CPU size and speed are not the only factors to determine processing power. To address this, you can 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 MB.

- **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 running a plan, Workload Optimization Manager can use the price to calculate costs or savings when adding or removing host machines in an on-prem datacenter.

**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. This can affect planning in two ways:

- When planning hardware replacement, the plan knows the template’s effective capacity. This means 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 calculating 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 calculating workload placement in real-time. For more information, see Effective CPU Capacity (on page 69).
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 will not attempt to place them.
The HCI Host template is a collection of these settings:

- **CPU**
  The processor for this host model. Note that CPU size and speed are not the only factors to determine processing power. To address this, you can 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 MB.

- **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 will compute the effective storage capacity.

- **Redundancy**
  The redundancy method for this storage 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 running a plan, Workload Optimization Manager can use the price to calculate costs or savings when adding or removing host machines in an on-prem datacenter.

### 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. This can affect planning in two ways:

- **When planning hardware replacement**, the plan knows the template’s effective capacity. This means 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 calculating 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 calculating workload placement in real-time. For more information, see [Effective CPU Capacity (on page 69)](#).
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 calculations. This setup includes:

- **Reserved Instance Settings**
  To recommend placing workloads on Reserved Instances (RIs), Workload Optimization Manager uses the real pricing plans that are available to the targets public cloud accounts. Setting up an RI Purchase Profile adds even more detail to the pricing structure that Workload Optimization Manager uses in its calculations.

- **OS Migration Profiles**
  For Migrate to Cloud plans, Workload Optimization Manager calculates the best placement for workloads that you want to move onto the public cloud. The migration includes choosing the OS for each migrated VM. The OS Profile that you configure here configures the default for how to manage the OS choices in migration plans.

- **Hardware Costs**
  As it generates reports and plans, Workload Optimization Manager can show estimated savings and costs for changes in the hosts and storage in your environment. To calculate these values, Workload Optimization Manager uses the prices you set up for Hardware Costs.

- **Price Adjustment**
  Cloud service 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 discounted prices in the Workload Optimization Manager display and in Workload Optimization Manager analysis, you must manually configure those discounts. In Workload Optimization Manager, you configure such discounts via **Price Adjustments** for specific billing groups in your cloud environment.
Billing and Costs

RI Purchase Profile

To recommend placing workloads on Reserved Instances (RIs), Workload Optimization Manager uses the real pricing plans that are available to the targets public cloud accounts. Setting up an RI Purchase Profile adds even more detail to the pricing structure that Workload Optimization Manager uses in its calculations.

The RI Purchase Profile determines the costs that Workload Optimization Manager will use for all RI decisions in your environment. As it sees opportunities to move workloads to an RI term, Workload Optimization Manager determines the costs based on the purchase profile, and includes the cost information in action descriptions. Workload Optimization Manager also uses this information to calculate projected changes in cost, and to calculate costs for plan results.

Note that the settings you make here globally affect all of your public cloud environment.

To set up the RI Purchase Profile, navigate to Settings > Billing and Costs, and display the RESERVED INSTANCE SETTINGS tab. Then make the settings for your purchase profile:

- **OFFERING CLASS**
  For AWS environments, choose the offering class that corresponds to the RI types that you typically use in your environment.

- **TERM**
  For AWS and Azure environments, choose the payment terms you contract for your RIs. TERM can be one of 1 Year or 3 Year. Typically, longer term payment plans cost less per year.

- **PAYMENT**
  The payment option that you prefer for your AWS RIs:
  - All Upfront – You make full payment at the start of the RI 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 for the RIs at an hourly rate, for the duration of the term.

When you are satisfied with your RI Purchase Profile settings, click APPLY SETTINGS. Or to reset the form, click RESET DEFAULTS.
OS Migration Profile

For Migrate to Cloud plans, Workload Optimization Manager calculates the best placement for workloads that you want to move onto the public cloud. The migration includes choosing the OS for each migrated VM. The OS Profile that you configure here configures the default for how to manage the OS choices in migration plans.

To set up the OS Profile that plans will use by default, navigate to Settings > Billing and Costs, and display the OS MIGRATION PROFILE tab. Then make the settings for your OS profile:

<table>
<thead>
<tr>
<th>Licensing</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Include OS cost</td>
<td>The plan includes OS cost per VM as part of the migration.</td>
<td></td>
</tr>
<tr>
<td>BYOL (“Bring your own license”)</td>
<td>The plan does not include OS cost as part of the migration. Use this option if you have purchased Azure Hybrid Benefit.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Custom OS</th>
<th>The plan includes OS cost based on your selections.</th>
<th></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>SOURCE OS</th>
<th>TARGET OS</th>
<th>BYOL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linux</td>
<td>Linux</td>
<td></td>
</tr>
<tr>
<td>RHEL</td>
<td>RHEL</td>
<td></td>
</tr>
<tr>
<td>SLES</td>
<td>SLES</td>
<td></td>
</tr>
<tr>
<td>Windows</td>
<td>Windows</td>
<td></td>
</tr>
</tbody>
</table>
The OS Migration Profile determines how Workload Optimization Manager will map the OS of each workload as it places that workload on the cloud destination. This includes how to choose VM templates that provide the OS you want, and whether to include the license cost in the Migrate to Cloud plan results. To configure an OS Migration Profile, choose from:

- 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.

When you are satisfied with your changes, click APPLY SETTINGS. Or to reset the form, click RESET DEFAULTS.

Hardware Costs

As it generates reports and plans, Workload Optimization Manager can show estimated savings and costs for changes in the hosts and storage in your environment. To calculate these values, Workload Optimization Manager uses the prices you set up for Hardware Costs.

Workload Optimization Manager uses the costs you set up here to estimate investments or savings for changes to hardware in your environment. For plans, if the plan uses templates and you have assigned prices to them, then Workload Optimization Manager uses those values. Otherwise, the plan uses the values that you set up here.

To specify Hardware Costs, navigate to Settings > Billing and Costs, and display the HARDWARE COSTS tab. Then make the settings for different hardware resources:

- HOST HARDWARE COST
  The cost for an average physical host machine in your environment.
- COST OF CPU PER UNIT
  The average cost of each CPU socket in your hosts.
- COST OF MEMORY PER GB
  The average cost of memory for your hosts, per GB.
- COST OF STORAGE PER TB
  The average cost of storage capacity for disk storage, per TB.

When you are satisfied with your changes, click APPLY SETTINGS. Or to reset the form, click RESET DEFAULTS.

Price Adjustments

Cloud service 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 discounted prices in the Workload Optimization Manager display and in Workload Optimization Manager analysis, you must manually configure those
Billing and Costs

discounts. In Workload Optimization Manager, you configure such discounts via **Price Adjustments** for specific billing groups in your cloud environment.

Workload Optimization Manager applies these price adjustments to:

- Costs for workload template families, including:
  - Compute
  - RI Compute
- Costs for services, including:
  - Azure Active Directory
  - Azure Stack
  - Bandwidth
  - VM Licenses
  - AWS CloudWatch
  - AWS DynamoDB
  - And others

Note that in AWS environments, Workload Optimization Manager does not apply any discounts or other price adjustments to Spot Compute costs.

The general steps to configure a price adjustment are:

- Create the price adjustment:
  - Specify the adjustment scope
    To do this, you choose which cloud service provider is giving you the adjustment, and then choose a billing group to set the scope of the adjustment.
  - Choose the Type
    The price adjustment can be a Discount or an Increase. In most cases you will specify discounts for the price adjustment. While this sets the type for the overall adjustment, you can override the type for specific line items.
  - Specify a Price Adjustment setting
    The Price Adjustment is the overall adjustment that your cloud service provider offers for the billing groups in your current scope. For example, AWS might offer you a 10% discount for a given account. For that billing group, you would specify a 10% Discount for the Price Adjustment setting.

- Specify Price Overrides
  While your service provider might offer a general price adjustment for the billing group you chose, it might also offer further discounts for select services or template families. Or it might offer discounts for some template families, but price increases for some other services. You can configure these differences as Price Overrides.

**NOTE:**
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.

**Creating a Price Adjustment**

A price adjustment configures adjusted workload pricing that you have negotiated with your Cloud Provider. After you configure an adjustment, Workload Optimization Manager applies it to pricing in the affected cloud scope.

To create a price adjustment in Workload Optimization Manager, you identify the adjustment’s scope – the subscriptions or billing families the adjustment applies to – and then set the type and percentage for the price adjustment. This specifies an overall adjustment for the workloads that fall within the billing group. You can later drill into the adjustment to specify overrides for specific template families or services.
Billing and Costs

Notes:

- 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. The *Workload Optimization Manager Installation Guide* recommends that you provide a minimum of 16 GB 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.

To create a price adjustment:

1. Navigate to the Settings Page.

   ![Settings](image)

   Click to navigate to the Settings Page. From there, you can perform a variety of configuration tasks.

2. Choose Billing and Costs.

   ![Billing and Costs](image)

   Click to navigate to the Billing and Costs page.

3. Display the PRICE ADJUSTMENT tab.

   Click the **PRICE ADJUSTMENT** tab to see all of the adjustments that have been configured for your environment. In this list you can:
   - Click an entry to see details and edit the adjustment
   - Select an entry to delete the adjustment
   - Create new price adjustments

4. Create the price adjustment.
First click **NEW PRICE ADJUSTMENT**, then specify the following settings to configure a price adjustment:

- Give the adjustment a name.
- To set the scope for this adjustment, choose its Billing Groups.

  **Click in the BILLING GROUPS field to display the Billing Groups fly-out.**

  In the Billing Groups fly-out, choose the cloud service provider you want to work with and then choose the billing group for the scope of this adjustment.
A Billing Group is a set of cloud service provider accounts that are consolidated into a single billing schedule. Billing group details depend on your service provider:

- **Azure**: For Azure environments, Workload Optimization Manager lists each Azure subscription as a billing group.
- **AWS**: To consolidate billing, AWS supports billing families of AWS accounts, where there is a *master* account and other *member* accounts. Workload Optimization Manager lists each billing family as a billing group. You can choose a billing family to set the scope of this adjustment.

After you have chosen your billing group, click **SAVE** to return to the Add New Price Adjustment fly-out.

- Set the **Type** for this price adjustment – Choose either **Discount** or **Increase**.
- Specify a percentage of adjustment as the **Price Adjustment**.

Enter the percentage in the **PRICE ADJUSTMENT** field. The acceptable value depends on the type of adjustment:

- For a discount: 0 - 99.99%
- For an increase: 0 - 999.99%

This is the general percentage of adjustment (increase or discount) for the current scope. For any costs within the adjustment scope, Workload Optimization Manager will apply this percentage as it calculates the optimal workload capacity and placement.

**NOTE:**
If you set an overall adjustment of 0%, then Workload Optimization Manager enforces a Type setting of Discount. The end result is the same, because an increase or a discount of 0% is the same.

5. Specify any price overrides for this price adjustment.
The PRICE 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 to open the Cloud Cost Adjustment fly-out. The overrides you can specify depend on the cloud service provider that manages the discount scope you have set:

- Azure – See Price Override: Azure (on page 382)
- AWS – See Price Override: AWS (on page 384)

6. Save your work.

After you have configured the price adjustment, click SAVE.

**Price Override: Azure**

<table>
<thead>
<tr>
<th>SERVICES</th>
<th>TYPE</th>
<th>PRICE ADJUSTMENT %</th>
<th>OVERRIDE %</th>
<th>ORIGINAL RATE (Linux)</th>
<th>EFFECTIVE ADJUSTMENT %</th>
<th>ADJUSTED RATE (Linux)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth</td>
<td>Discount</td>
<td>10 %</td>
<td>%</td>
<td></td>
<td>10 %</td>
<td></td>
</tr>
<tr>
<td>SQL Database</td>
<td>Discount</td>
<td>40 %</td>
<td>15 %</td>
<td></td>
<td>15 %</td>
<td></td>
</tr>
<tr>
<td>Virtual Machines</td>
<td>Discount</td>
<td>10 %</td>
<td>%</td>
<td></td>
<td>10 %</td>
<td></td>
</tr>
<tr>
<td>standardNIC</td>
<td>Discount</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard Database</td>
<td>Discount</td>
<td>10 %</td>
<td>27%</td>
<td></td>
<td>27%</td>
<td></td>
</tr>
<tr>
<td>Standard Database</td>
<td>Discount</td>
<td>10 %</td>
<td>%</td>
<td></td>
<td>10 %</td>
<td></td>
</tr>
</tbody>
</table>

To override the PRICE ADJUSTMENT setting for Azure billing groups, Workload Optimization Manager analysis can use settings for different services that Azure provides to subscriptions.
Assume your price adjustment specifies a discount of 10% for an Azure subscription. But then assume the subscription includes extra discounts for some of the services the subscription provides. Then you can create overrides to add the extra discounts for those services. For more information about Azure subscriptions and cost calculations, see Azure Enterprise Agreements (on page 46).

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.
Price Override: AWS

To override the PRICE ADJUSTMENT setting 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 master 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 44).

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 Cost and Usage report 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.

**Currency Settings**

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 €. To do this, go to **Settings > Billing and Costs** and then click the **Currency** tab.

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.
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 386)*
  Create and manage user accounts for Workload Optimization Manager.
- **Viewing the Update page** *(on page 397)*
  See information about your current version.
- **License Configuration** *(on page 398)*
  Review the status of your current license, and apply any license upgrades.

Managing User Accounts

As an administrator, you specify accounts that grant users specific access to Workload Optimization Manager. User accounts determine the following for a given user login:

- **User Authentication**
  To configure an account, you set the type of authentication the account will use:
  - Local User – Configure the username and password and save those credentials on the Workload Optimization Manager server.
  - External User – Single user accounts that authenticate through Single Sign-on (SSO) or through Microsoft Active Directory (AD).
  - External Group – A group of user accounts that authenticate through SSO or AD.

- **User Authorization**
  Properties that determine the range of access and features for a given user:
  - Role – Access to specific Workload Optimization Manager features
  - Scope – 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:
You can configure Workload Optimization Manager to use SSO authentication. 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 configuring SSO user groups in SAML, see Configuring a Group for SSO Authentication (on page 394).

For information about configuring SSO authentication for Workload Optimization Manager, see "Single Sign-On Authentication" in the Workload Optimization Manager Installation Guide.

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.
Manage local or external authentication

This page lists all the user accounts that you currently have configured for Workload Optimization Manager. You can:

- Click to manage LOCAL USERS or EXTERNAL AUTHENTICATION
- Select an entry to delete the account
- Click a name to edit the account
- Create new user or group account
- Configure Active Directory settings

3. Filter the list of users.

Filter by role

To work with a long list of users, you can filter by role (for example, only show administrator or only show observer users). You can also type a string in the **Search** field to filter the list, and you can sort the list by name.

4. Work with Local user accounts.
Choose to work with local user accounts

<table>
<thead>
<tr>
<th>LOCAL USERS</th>
<th>EXTERNAL AUTHENTICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Local Users</td>
<td></td>
</tr>
<tr>
<td>Administrator - 1</td>
<td></td>
</tr>
<tr>
<td>administrator - 1</td>
<td></td>
</tr>
<tr>
<td>Advisor - 1</td>
<td></td>
</tr>
<tr>
<td>Automator - 2</td>
<td></td>
</tr>
<tr>
<td>Site Administrator - 2</td>
<td></td>
</tr>
</tbody>
</table>

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.
Administrative Tasks

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**
    
    Can use all Workload Optimization Manager features, and can modify settings to configure the Workload Optimization Manager installation.
  
  - **Site Administrator**
    
    Can use all Workload Optimization Manager features, and can modify site-specific settings to configure the Workload Optimization Manager installation. Can administer Groups, Policies, Templates, Billing/Costs, Users (cannot create users with Administrator role), Target Configuration, but cannot administer Email, Licenses, Updates, and Maintenance.
  
  - **Automator**
    
    Can use all of the Workload Optimization Manager features including Plan and Place, but cannot configure the Workload Optimization Manager installation or create policies.
– **Deployer**
  Can view all Workload Optimization Manager charts and data, can use Place to reserve workloads, and can create policies and templates. However, this role cannot run plans or execute any recommended actions.

– **Advisor**
  Can view all Workload Optimization Manager charts and data, and can run plans, but cannot use Place to reserve workloads, or execute any recommended actions.

– **Observer**
  Can view the environment, including the Home Page, and Dashboards. Can also use Search to set a scope to the session. For scope, only VM groups and Resource Groups are supported.

– **Operational Observer**
  Can view the environment, including the Home Page, Dashboards, Groups, and Policies. Can also use Search to set a scope to the session.

– **Shared Advisor**
  A scoped user. Can only see VMs and Applications, and cannot execute Workload Optimization Manager actions. Can view the Home Page and Dashboards.

– **Shared Observer**
  A scoped user. Can only see VMs and Applications, and cannot execute Workload Optimization Manager actions. Can view the Home Page and custom Dashboards, but not Executive Dashboards. This is the most restricted user.

– **Report Editor**
  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 (optional)**
  The scope limits what the user can monitor. 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** 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.

6. Work with EXTERNAL AUTHENTICATION to set up SSO or AD accounts.

Choose to work with externally authenticated users or groups
For External Authentication, you configure Workload Optimization Manager to use SSO or AD services to manage the credentials and authentication of users. You can create external accounts to authorize user groups or individual users.

**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 note that 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” in the Workload Optimization Manager Installation Guide.

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.

You must specify or edit the AD connection

To enable AD, click **CONNECT TO AD** and configure:

- **Active Directory Domain** – To authenticate AD groups, specify a domain so that AD can find a given user via the User Principal Name (UPN). If you specify a domain, but not a server, authentication uses any AD server from that domain.
- **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 will use that server, and will also support 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 when communicating with AD servers. Note that the AD domain must be configured to use LDAPS, and you must have imported a certificate into the Workload Optimization Manager server. For more information, see “Enforcing Secure Access” in the Workload Optimization Manager Installation Guide.

8. Create or edit an SSO or AD account.
Administrative Tasks

This account can be for a user group or for a single user. To add a new account, click **NEW EXTERNAL GROUP** or **NEW EXTERNAL USER**. To edit an existing account, click the account name. To configure an external account, specify:

- **Authentication:**
  - **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**
    Can use all Workload Optimization Manager features, and can modify settings to configure the Workload Optimization Manager installation.
  - **Site Administrator**
    Can use all Workload Optimization Manager features, and can modify site-specific settings to configure the Workload Optimization Manager installation. Can administer Groups, Policies, Templates, Billing/Costs, Users (cannot create users with Administrator role), Target Configuration, but cannot administer Email, Licenses, Updates, and Maintenance.

This account can be for a user group or for a single user. To add a new account, click **NEW EXTERNAL GROUP** or **NEW EXTERNAL USER**. To edit an existing account, click the account name. To configure an external account, specify:

- **Authentication:**
  - **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**
    Can use all Workload Optimization Manager features, and can modify settings to configure the Workload Optimization Manager installation.
  - **Site Administrator**
    Can use all Workload Optimization Manager features, and can modify site-specific settings to configure the Workload Optimization Manager installation. Can administer Groups, Policies, Templates, Billing/Costs, Users (cannot create users with Administrator role), Target Configuration, but cannot administer Email, Licenses, Updates, and Maintenance.
Administrative Tasks

- **Automator**
  Can use all of the Workload Optimization Manager features including Plan and Place, but cannot configure the Workload Optimization Manager installation or create policies.

- **Deployer**
  Can view all Workload Optimization Manager charts and data, can use Place to reserve workloads, and can create policies and templates. However, this role cannot run plans or execute any recommended actions.

- **Advisor**
  Can view all Workload Optimization Manager charts and data, and can run plans, but cannot use Place to reserve workloads, or execute any recommended actions.

- **Observer**
  Can view the environment, including the Home Page, Dashboards, and Groups (VM Groups and Resource Groups only). Can also use Search to set a scope to the session. For scope, only VM groups and Resource Groups are supported.

- **Operational Observer**
  Can view the environment, including the Home Page, Dashboards, Groups, and Policies. Can also use Search to set a scope to the session.

- **Shared Advisor**
  A scoped user. Can only see VMs and Applications, and cannot execute Workload Optimization Manager actions. Can view the Home Page and Dashboards.

- **Shared Observer**
  A scoped user. Can only see VMs and Applications, and cannot execute Workload Optimization Manager actions. Can view the Home Page and custom Dashboards, but not Executive Dashboards. This is the most restricted user.

- **Report Editor**
  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 (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** and choose which entities this members of this group can see.

### 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” in the Workload Optimization Manager Installation Guide.

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.

<table>
<thead>
<tr>
<th>Users</th>
<th>Group Attribute</th>
</tr>
</thead>
<tbody>
<tr>
<td>George</td>
<td>Attribute Name=group, AttributeValue=Beatles</td>
</tr>
<tr>
<td>Paul</td>
<td></td>
</tr>
<tr>
<td>John</td>
<td></td>
</tr>
<tr>
<td>Ringo</td>
<td></td>
</tr>
<tr>
<td>Smokey</td>
<td>Attribute Name=group, AttributeValue=Miracles</td>
</tr>
<tr>
<td>Pete</td>
<td></td>
</tr>
<tr>
<td>Ronnie</td>
<td></td>
</tr>
<tr>
<td>Claudette</td>
<td></td>
</tr>
<tr>
<td>Bobby</td>
<td></td>
</tr>
<tr>
<td>Marv</td>
<td></td>
</tr>
</tbody>
</table>

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 above example, the group value is turbo_admin_group. To set authorization for that group:

1. Open the User Management page to EXTERNAL AUTHENTICATION.
2. Navigate to Settings > User Management, and display the EXTERNAL AUTHENTICATION view.
3. Create a new External Group
4. 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 above example, use the name `turbo_admin_group`.

4. Specify the group’s authorization
   For the above 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.

### Maintenance: Logging and Troubleshooting

The Maintenance Options Page provides tools to set logging levels and to export data for technical support, and import diagnostic files from Technical Support. Many of these tools are for advanced users. You should contact Cisco technical support before you use them.

To execute these actions, navigate to the Maintenance Options page:

1. Navigate to the Settings Page.
   ![Settings Button]
   Click to navigate to the Settings Page.

2. Choose Maintenance Options.

### Data Retention

Workload Optimization Manager gathers metrics from your environment to provide historical reports. To optimize data storage, it consolidates the data into three groups - Hourly, Daily, and Monthly. Daily statistics consolidate Hourly data, and Monthly statistics consolidate Daily data. Workload Optimization Manager also saves plans, reports, and audit log entries.

You can always modify the default values to meet your requirements. Remember that the longer the retention period, the more storage is required.
HTTP Proxy

If your environment requires an HTTP proxy for Workload Optimization Manager to access the web, provide the credentials here.

Diagnostics

If you are experiencing problems with Workload Optimization Manager, your support engineer might request that you export diagnostic data. You can export the data and then send it to the support engineer as requested.

Logging Levels

You can set the level of logging for different components of the Workload Optimization Manager platform. You should be aware that setting more verbose logging levels increases the disk space required to store the log files. You normally change these settings only while you’re working with a Workload Optimization Manager support engineer.

The Updates Page

Use the Updates page to get information about your Workload Optimization Manager version.

The ABOUT button shows the current version and build of your Workload Optimization Manager installation. It also lists the platform components by name and version.

NOTE:
For complete update instructions, see the Installation Guide.
To navigate to the Updates page:

1. Navigate to the Settings Page.

   ![Settings Page](image)

   Click to navigate to the Settings Page.

2. Choose Updates.

   ![Updates](image)

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 the license file to you in an e-mail message.

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 you installation can manage. Note that as you add more licenses, they must all support the same feature set.

The License Configuration page shows you:

- The number of active workloads you can manage under this license
- How many workloads are currently active
- The set of features this license enables
- A list of current, active licenses
To navigate to the License Configuration page:

1. Navigate to the Settings Page.

2. Choose License.

To activate a license or to update your current license:

1. Obtain your license.
   Cisco sends the license file to you in an e-mail message. 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.
   First click **IMPORT LICENSE**. Then browse to the license file that you saved and open it. Or you can drag the file into the **Enter License** fly-out.
   After you have uploaded the file, click **SAVE**.

After you have activated your license, you can then add more licenses to increase your workload coverage, or you can license a higher feature set.

**NOTE:**
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.

After you install a new license, it is a good idea to clear your browser cache.

To increase your licensed workload coverage:

1. Obtain your additional license.
   Note that your additional licenses must match the feature set of your current license.

2. Apply the license to your Workload Optimization Manager installation.

To 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**

Configure email settings to enable email communication 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**

<table>
<thead>
<tr>
<th>SMTP SERVER</th>
<th>SMTP PORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0.0.0</td>
<td>25</td>
</tr>
</tbody>
</table>

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**

<table>
<thead>
<tr>
<th>'FROM' ADDRESS</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="mailto:email@organization.com">email@organization.com</a></td>
</tr>
</tbody>
</table>

Use this setting to specify the return address (the FROM address) for emails that Workload Optimization Manager generates and sends.