Cisco HX Data Platform Overview

Cisco HyperFlex Data Platform (HX Data Platform) is a hyperconverged software appliance that transforms Cisco servers into a single pool of compute and storage resources. It eliminates the need for network storage and enables seamless interoperability between computing and storage in virtual environments. The Cisco HX Data Platform provides a highly fault-tolerant distributed storage system that preserves data integrity and optimizes performance for virtual machine (VM) storage workloads. In addition, native compression and deduplication, reduce storage space occupied by the VMs and VM workloads.

Cisco HX Data Platform has many integrated components. These include: Cisco Fabric Interconnects (FIs), Cisco UCS Manager, Cisco HX specific servers, and Cisco compute only servers; VMware vSphere, ESX servers, and vCenter; and the Cisco HX Data Platform installer, controller VMs, vSphere plug-in, and `stcli` commands.

Cisco HX Data Platform is installed on a virtualized platform such as VMware vSphere. During installation, you specify the Cisco HyperFlex HX Cluster name, and the HX Data Platform creates a hyperconverged storage cluster on each of the nodes. As your storage needs increase and you add nodes in the HX cluster, the HX Data Platform balances the storage across the additional resources. Compute only nodes can be added to increase compute only resources to the storage cluster.

Storage Cluster Physical Components Overview

Cisco HX Data Platform storage clusters contain the following objects. These objects are monitored by the HX Data Platform for the storage cluster. They can be added and removed from the storage cluster.
• **Converged nodes.** Converged nodes are the physical hardware on which the VM runs. They provide computing and storage resources such as disk space, memory, processing, power, and network I/O.

When a converged node is added to the storage cluster, a storage controller VM is installed. The HX Data Platform services are handled through the storage controller VM. Converged nodes add storage resources to your storage cluster through their associated drives.

Adding converged nodes to your storage cluster is performed using the Expand Cluster feature of the HX Data Platform Installer. Removing converged nodes is a command line task.

• **Compute nodes.** Compute nodes add compute resource but not storage capacity to the storage cluster. They are used as means to add compute resources, including CPU and memory. They do not need to have any caching (SSD) or storage (HDD) drives. Compute nodes are optional to the storage cluster.

When a compute node is added to the storage cluster, an agent controller VM is installed. The HX Data Platform services are handled through the agent controller VM. Compute nodes do not add storage resources to your storage cluster.

Adding compute nodes to your storage cluster is performed using the Expand Cluster feature of the HX Data Platform Installer. Removing compute nodes is a command line task.

• **Drives.** There are two types of drives required for any node in the storage cluster: Solid State Drive (SSD) and Hard Disk Driver (HDD). HDD typically provide the physical storage units associated with converged nodes. SSD typically support management.

Adding HDD to existing converged nodes, also adds storage capacity to the storage cluster. When storage is added to any node in the storage cluster, an equal amount of storage must be added to every node in the storage cluster.

When disks are added or removed, the HX Data Platform rebalances the storage cluster to adjust for the change in storage resource.

Adding or removing disks on your converged nodes is not performed through the HX Data Platform. Prior to adding or removing disks, review the best practices. See the server hardware guides for specific instructions to add or remove disks in nodes.

• **Datastores.** Storage capacity and datastore capacity. This is the combined consumable physical storage available to the storage cluster through datastores, and managed by the HX Data Platform.

Datastore are logical containers used by the HX Data Platform to manage your storage use and storage resources.

Datastores are where the host places virtual disk files and other VM files. Datastores hide the specifics of physical storage devices and provide a uniform model for storing VM files.

Additional objects that can be monitored from a storage cluster point of view and are treated as field replaceable units (FRUs), include the following. See the server hardware guides for information about replacing these objects.

• Power Supply Units (PSUs)

• Network Interface Cards (NICs)
HX Data Platform Capacity Overview

In the HX Data Platform the concept of capacity is applied to both datastores and storage clusters. Values are measured in base-2 (GiB/TiB), but for simplicity and consistency are labeled as GB or TB.

- **Cleaner**—A process run on all the storage cluster datastores. After it completes, all the storage cluster datastores total capacity should be in a similar range to the total storage cluster capacity, excluding the metadata. Datastore capacity listed typically will not match the HX storage cluster capacity. See the Cisco HX Data Platform Command Line Interface Reference guide for information on the `cleaner` command.

- **Cluster capacity**—All the storage from all the disks on all the nodes in the storage cluster. This includes uncleaned data and the metadata overhead for each disk.

  The total/used/free capacity of cluster is based on overall storage capacity and how much storage is used.

- **Condition**—When the HX Storage Cluster enters a space event state, the Free Space Status fields are displayed. The Condition field lists the space event state. The options are: Warning, Critical, and Alert.

- **Datastore capacity**—The amount of storage available for provisioning to datastores without over-provisioning. Generally, this is similar to the cleaned storage cluster capacity, but it is not an exact match. It does not include metadata or uncleaned data.

  The provisioned/used/free capacity of each datastore is based on datastore (thin) provisioned capacity. Because the datastore is thin provisioned, the provisioned capacity (specified by the administrator when creating the datastore) can be well above the actual storage.

- **Free Capacity, storage cluster**—Same as available capacity. For the storage cluster, this is the difference between the amount available to the storage cluster and the amount used in the storage cluster.

- **Free capacity, datastore**—Same as available capacity. For all the storage cluster datastores, this is the difference between the amount provisioned to all the storage cluster datastores and the amount used on all the storage cluster datastores.

  The amount used on the whole storage cluster is not included in this datastore calculation. Because datastores are frequently over provisioned, the free capacity can indicate a large availability on all the storage cluster datastores, while the storage cluster capacity can indicate a much lower availability.

- **Multiple users**—Can have different datastores with different provisioned capacities. At any point in time, users do not fully utilize their allocated datastore capacity. When allocating datastore capacity to multiple users, it is up to the administrator to ensure that each user’s provisioned capacity is honored at all time.

- **Over-provisioning**—Occurs when the amount of storage capacity allocated to all the datastores exceeds the amount available to the storage cluster.

  It is a common practice to initially over-provision. It allows administrators to allocate the capacity now and backfill the actual storage later.

  The value is the difference between the usable capacity and provisioned capacity.

  It displays zero (0) value, unless more space has been allocated than the maximum physical amount possible.

  Review the over provisioned capacity and ensure that your system does not reach an out-of-space condition.
• **Provisioned**—Amount of capacity allowed to be used by, allocated to, the storage cluster datastores. The provisioned amount is not set aside for the sole use of the storage cluster datastores. Multiple datastores can be provisioned storage from the same storage capacity.

• **Space Needed**—When the HX Storage Cluster enters a space event state, the **Free Space Status** fields are displayed. **Space Needed** indicates the amount of storage that needs to be made available to clear the listed **Condition**.

• **Used**—Amount of storage capacity consumed by the listed storage cluster or datastore.

  HX Data Platform internal meta-data uses 0.5% to 1% space. This might cause the HX Data Platform Plug-in or HX Connect to display a **Used Storage** value even if you have no data in your datastore.

  Storage Used shows how much datastore space is occupied by virtual machine files, including configuration and log files, snapshots, and clones. When the virtual machine is running, the used storage space also includes swap files.

• **Usable Capacity**—Amount of storage in the storage cluster available for use to store data.

**Understanding Capacity Savings**

The Capacity portlet on the Summary tab displays the deduplication and compression savings provided by the storage cluster. For example, with 50% overall savings, a 6TB capacity storage cluster can actually store 9 TB of data.

The total storage capacity saved by the HX Data Platform system is a calculation of two elements:

• **Compression** - How much of the data is compressed.

• **Deduplication** - How much data is deduplicated. Deduplication is a method of reducing storage space by eliminating redundant data. It stores only one unique instance of the data.

Deduplication savings and compression savings are not simply added together. They are not independent operations. They are correlated using the following elements where essentially the amount of unique bytes used for storage is reduced through deduplication. Then the deduplicated storage consumption is compressed to make even more storage available to the storage cluster.

Deduplication and compression savings are particularly useful when working with VM clones.

If the savings is showing 0%, this indicates the storage cluster is new. The total ingested data to the storage cluster is insufficient to determine meaningful storage savings. Wait until sufficient data is written to the storage cluster.

**For example:**

1. **Initial values**

   Given a VM of 100 GB that is cloned 2 times.

   - Total Unique Used Space (TUUS) = 100GB
   - Total Addressable Space (TAS) = 100x2 = 200 GB

   Given, for this example:

   - Total Unique Bytes (TUB) = 25 GB

2. **Deduplication savings**
3. Compression Savings
   \[- \frac{(1 - TUB/TUUS) \times 100}{1 - (25GB / 100GB) \times 100} = 75\%\]

4. Total savings calculated
   \[- \frac{(1 - TUB/TAS) \times 100}{1 - (25GB / 200GB) \times 100} = 87.5\%\]

**Storage Capacity Event Messages**

Cluster storage capacity includes all the storage from all the disks on all the nodes in the storage cluster. This available capacity is used to manage your data.

Error messages are issued if your data storage needs consume high amounts of available capacity, the performance and health of your storage cluster are affected. The error messages are displayed in vCenter Alarms panels, HX Connect, and HX Data Platform Plug-in Alarms and Events pages.

---

**Note**

When the warning or critical errors appear:

Add additional drives or nodes to expand capacity. Additionally, consider deleting unused virtual machines and snapshots. Performance is impacted until storage capacity is reduced.

- **SpaceWarningEvent** – Issues an error. This is a first level warning.
  Cluster performance is affected.
  Reduce the amount of storage capacity used to below the warning threshold, of 70% total HX Storage Cluster capacity.

- **SpaceAlertEvent** – Issues and error. Space capacity usage remains at error level.
  This alert is issued after storage capacity has been reduced, but is still above the warning threshold.
  Cluster performance is affected.
  Continue to reduce the amount of storage capacity used, until it is below the warning threshold, of 80% total HX Storage Cluster capacity.

- **SpaceCriticalEvent** – Issues and error. This is a critical level warning.
  Cluster is in a read only state.
  Do not continue the storage cluster operations until you reduce the amount of storage capacity used to below this warning threshold, of 92% total HX Storage Cluster capacity.
**HX Data Platform High Availability Overview**

The HX Data Platform High Availability (HA) feature ensures that the storage cluster maintains at least two copies of all your data during normal operation with three or more fully functional nodes.

If nodes or disks in the storage cluster fail, the cluster's ability to function is affected. If more than one node fails or one node and disk(s) on a different node fail, it is called a *simultaneous failure*.

The number of nodes in the storage cluster, combined with the Data Replication Factor and Access Policy settings, determine the state of the storage cluster that results from node failures.

---

**Note**

Before using the HX Data Platform HA feature, enable DRS and vMotion on the vSphere Web Client.

---

**Storage Cluster Status**

HX Data Platform storage cluster status information is available through the HX Data Platform plug-in and the storage controller VM stcli commands. Storage cluster status is described through resiliency and operational status values.

Storage cluster status is described through two reported status elements.

- **Operational Status** - Describes the ability of the storage cluster to perform the functions storage management and storage cluster management of the cluster. How well can the storage cluster perform operations.

- **Resiliency Status** - Describes the ability of the storage clusters to tolerate node failures within the storage cluster. How well can the storage cluster handle disruptions.

Two settings that work with each other affect when the storage cluster transitions into particular operational and resiliency status states.

- **Data Replication Factor** - Sets the number of redundant data replicas.

- **Cluster Access Policy** - Sets the level of data protection and data loss.

---

**Operational Status Values**

Cluster Operational Status indicates the operational status of the storage cluster and the ability for the applications to perform I/O.

The Operational Status options are:

- **Online**—Cluster is ready for IO.

- **Offline**—Cluster is not ready for IO.
• **Out of space**—Either the entire cluster is out of space or one or more disks are out of space. In both cases, the cluster cannot accept write transactions, but can continue to display static cluster information.

• **Readonly**—Cluster cannot accept write transactions, but can continue to display static cluster information.

• **Unknown**—This is a transitional state while the cluster is coming online.

Other transitional states might be displayed during cluster upgrades and cluster creation.

Color coding and icons are used to indicate various status states. Click icons to display additional information such as reason messages that explain what is contributing to the current state.

### Resiliency Status Values

Resiliency status is the data resiliency health status and ability of the storage cluster to tolerate failures.

Resiliency Status options are:

- **Healthy**—The cluster is healthy with respect to data and availability.

- **Warning**—Either the data or the cluster availability is being adversely affected.

- **Unknown**—This is a transitional state while the cluster is coming online.

Color coding and icons are used to indicate various status states. Click an icon to display additional information, such as reason messages that explain what is contributing to the current state.

### HX Data Platform Cluster Tolerated Failures

If nodes or disks in the storage cluster fail, the cluster's ability to function is affected. If more than one node fails or one node and disk(s) on a different node fail, it is called a *simultaneous failure*.

How the number of node failures affect the storage cluster is dependent upon:

- **Number of nodes in the cluster**. The response by the storage cluster is different for clusters with 3 to 4 nodes and 5 or greater nodes.

- **Data Replication Factor**. This is set during HX Data Platform installation and cannot be changed. The options are 2 or 3 redundant replicas of your data across the storage cluster.

  **Attention** Data Replication Factor 3 is the recommended option.

- **Access Policy**. This can be changed from the default setting after the storage cluster is created. The options are strict, for protecting against data loss, or lenient, to support longer storage cluster availability.

### Cluster State with Number of Failed Nodes

The tables below list how the storage cluster functionality changes with the listed number of simultaneous node failures.

**Cluster State in 5+ Node Cluster with Number of Failed Nodes**
Cluster State in 3 - 4 Node Clusters with Number of Failed Nodes

<table>
<thead>
<tr>
<th>Replication Factor</th>
<th>Access Policy</th>
<th>Number of Failed Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Read/Write</td>
</tr>
<tr>
<td>3</td>
<td>Lenient or Strict</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Lenient</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Strict</td>
<td>--</td>
</tr>
</tbody>
</table>

Cluster State with Number of Nodes with Failed Disks

The table below lists how the storage cluster functionality changes with the number of nodes that have one or more failed disks. Note that the node itself has not failed but disk(s) within the node have failed. For example: 2 indicates that there are 2 nodes that each have at least one failed disk.

There are two possible types of disks on the servers: SSDs and HDDs. When we talk about multiple disk failures in the table below, it's referring to the disks used for storage capacity. For example: If a cache SSD fails on one node and a capacity SSD or HDD fails on another node the storage cluster remains highly available, even with an Access Policy strict setting.

The table below lists the worst case scenario with the listed number of failed disks. This applies to any storage cluster 3 or more nodes. For example: A 3 node cluster with Replication Factor 3, while self-healing is in progress, only shuts down if there is a total of 3 simultaneous disk failures on 3 separate nodes.

Note

HX storage clusters are capable of sustaining serial disk failures, (separate disk failures over time). The only requirement is that there is sufficient storage capacity available for support self-healing. The worst-case scenarios listed in this table only apply during the small window while HX is completing the automatic self-healing and rebalancing.

3+ Node Cluster with Number of Nodes with Failed Disks

<table>
<thead>
<tr>
<th>Replication Factor</th>
<th>Access Policy</th>
<th>Failed Disks on Number of Different Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Read/Write</td>
</tr>
<tr>
<td>3</td>
<td>Lenient</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Strict</td>
<td>1</td>
</tr>
</tbody>
</table>
Data Replication Factor Settings

Data Replication Factor cannot be changed after the storage cluster is configured.

Data Replication Factor is set when you configure the storage cluster. Data Replication Factor defines the number of redundant replicas of your data across the storage cluster. The options are 2 or 3 redundant replicas of your data.

- If you have hybrid servers (servers that contain both SSD and HDDs), then the default is 3.
- If you have all flash servers (servers that contain only SSDs), then you must explicitly select either 2 or 3 during HX Data Platform installation.

Choose a Data Replication Factor. The choices are:

- **Data Replication Factor 3** — Keep three redundant replicas of the data. This consumes more storage resources, and ensures the maximum protection for your data in the event of node or disk failure. **Attention** Data Replication Factor 3 is the recommended option.

- **Data Replication Factor 2** — Keep two redundant replicas of the data. This consumes fewer storage resources, but reduces your data protection in the event of node or disk failure.

## Cluster Access Policy

The Cluster Access Policy works with the Data Replication Factor to set levels of data protection and data loss prevention. There are two Cluster Access Policy options. The default is lenient. It is not configurable during installation, but can be changed after installation and initial storage cluster configuration.

- **Strict** - Applies policies to protect against data loss.
  
  If nodes or disks in the storage cluster fail, the cluster's ability to function is affected. If more than one node fails or one node and disk(s) on a different node fail, it is called a simultaneous failure. The strict setting helps protect the data in event of simultaneous failures.

- **Lenient** - Applies policies to support longer storage cluster availability. This is the default.
Responses to Storage Cluster Node Failures

A storage cluster healing timeout is the length of time the HX Data Platform plug-in waits before automatically healing the storage cluster. If a disk fails, the healing timeout is 1 minute. If a node fails, the healing timeout is 2 hours. A node failure timeout takes priority if a disk and a node fail at the same time or if a disk fails after node failure, but before the healing is finished.

When the cluster resiliency status is Warning, the HX Data Platform system supports the following storage cluster failures and responses.

Optionally, click the associated Cluster Status or Resiliency Status icons, in the HX Data Platform plug-in, to display reason messages that explain what is contributing to the current state.

<table>
<thead>
<tr>
<th>Cluster Size</th>
<th>Number of Simultaneous Failures</th>
<th>Entity Failed</th>
<th>Maintenance Action to Take</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 nodes</td>
<td>1</td>
<td>One node.</td>
<td>The storage cluster does not automatically heal. Replace the failed node to restore storage cluster health.</td>
</tr>
</tbody>
</table>
|              | 2                             | Two or more disks on two nodes are blacklisted or failed. | 1. If one SSD fails, the storage cluster does not automatically heal.  
- Replace the faulty SSD and restore the system by rebalancing the cluster  
2. If one HDD fails or is removed, the disk is blacklisted immediately. The storage cluster automatically begins healing within a minute.  
3. If more than one HDD fails, the system might not automatically restore storage cluster health.  
- If the system is not restored, replace the faulty disks and restore the system by rebalancing the cluster |
<table>
<thead>
<tr>
<th>Cluster Size</th>
<th>Number of Simultaneous Failures</th>
<th>Entity Failed</th>
<th>Maintenance Action to Take</th>
</tr>
</thead>
</table>
| 4 nodes      | 1                               | One node.     | If the node does not recover in two hours, the storage cluster starts healing by rebalancing data on the remaining nodes. To recover the failed node immediately and fully restore the storage cluster:  
1. Check that the node is powered on and restart it if possible. You might need to replace the node.  
2. Rebalance the cluster |
| 4 nodes      | 2                               | Two or more disks on two nodes. | If two SSDs fail, the storage cluster does not automatically heal. If the disk does not recover in one minute, the storage cluster starts healing by rebalancing data on the remaining nodes. |
## Responses to Storage Cluster Node Failures

<table>
<thead>
<tr>
<th>Cluster Size</th>
<th>Number of Simultaneous Failures</th>
<th>Entity Failed</th>
<th>Maintenance Action to Take</th>
</tr>
</thead>
<tbody>
<tr>
<td>5+ nodes</td>
<td>2</td>
<td>Up to two nodes.</td>
<td>If the node does not recover in two hours, the storage cluster starts healing by rebalancing data on the remaining nodes. To recover the failed node immediately and fully restore the storage cluster: 1. Check that the node is powered on and restart it if possible. You might need to replace the node. 2. Rebalance the cluster. If the storage cluster shuts down, see Troubleshooting, Two Nodes Fail Simultaneously Causes the Storage Cluster to Shutdown section.</td>
</tr>
<tr>
<td>5+ nodes</td>
<td>2</td>
<td>Two nodes with two or more disk failures on each node.</td>
<td>The system automatically triggers a rebalance after a minute to restore storage cluster health.</td>
</tr>
</tbody>
</table>
If the disk does not recover in **one minute**, the storage cluster starts healing by rebalancing data on the remaining nodes.

If the node does not recover in **two hours**, the storage cluster starts healing by rebalancing data on the remaining nodes.

If a node in the storage cluster fails and a disk on a different node also fails, the storage cluster starts healing the failed disk (without touching the data on the failed node) in one minute. If the failed node does not come back up after two hours, the storage cluster starts healing the failed node as well.

To recover the failed node immediately and fully restore the storage cluster:

1. Check that the node is powered on and restart it if possible. You might need to replace the node.

2. Rebalance the cluster

Review the table above and perform the action listed.
HX Data Platform ReadyClones Overview

HX Data Platform ReadyClones is a pioneer storage technology that enables you to rapidly create and customize multiple cloned VMs from a host VM. It is an expansion feature that is used for creating multiple copies of VMs that can then be used as standalone VMs.

A ReadyClone, similar to a standard clone, is a copy of an existing VM. The existing VM is called the host VM. When the cloning operation is complete, the ReadyClone is a separate guest VM.

Changes made to a ReadyClone do not affect the host VM. A ReadyClone's MAC address and UUID are different from those of the host VM.

Installing a guest operating system and applications can be time consuming. With ReadyClone, you can make many copies of a VM from a single installation and configuration process.

Clones are useful when you deploy many identical VMs to a group.

HX Data Platform Native Snapshots Overview

HX Data Platform Native Snapshots are a backup feature that saves versions (states) of working VMs. VMs can be reverted back to native snapshots.

Use the HX Data Platform plug-in to take native snapshots of your VMs. HX Data Platform native snapshot options include: create a native snapshot, revert to any native snapshot, and delete a native snapshot. Timing options include: Hourly, Daily, and Weekly, all in 15 minute increments.

A native snapshot is a reproduction of a VM that includes the state of the data on all VM disks and the VM power state (on, off, or suspended) at the time the native snapshot is taken. Take a native snapshot to save the current state of the VM, so that you have the option to revert to the saved state.

You can take a native snapshot when a VM is powered on, powered off, or suspended. For additional information about VMware snapshots, see the VMware KB, Understanding virtual machine snapshots in VMware ESXi and ESX (1015180) at, http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=1015180