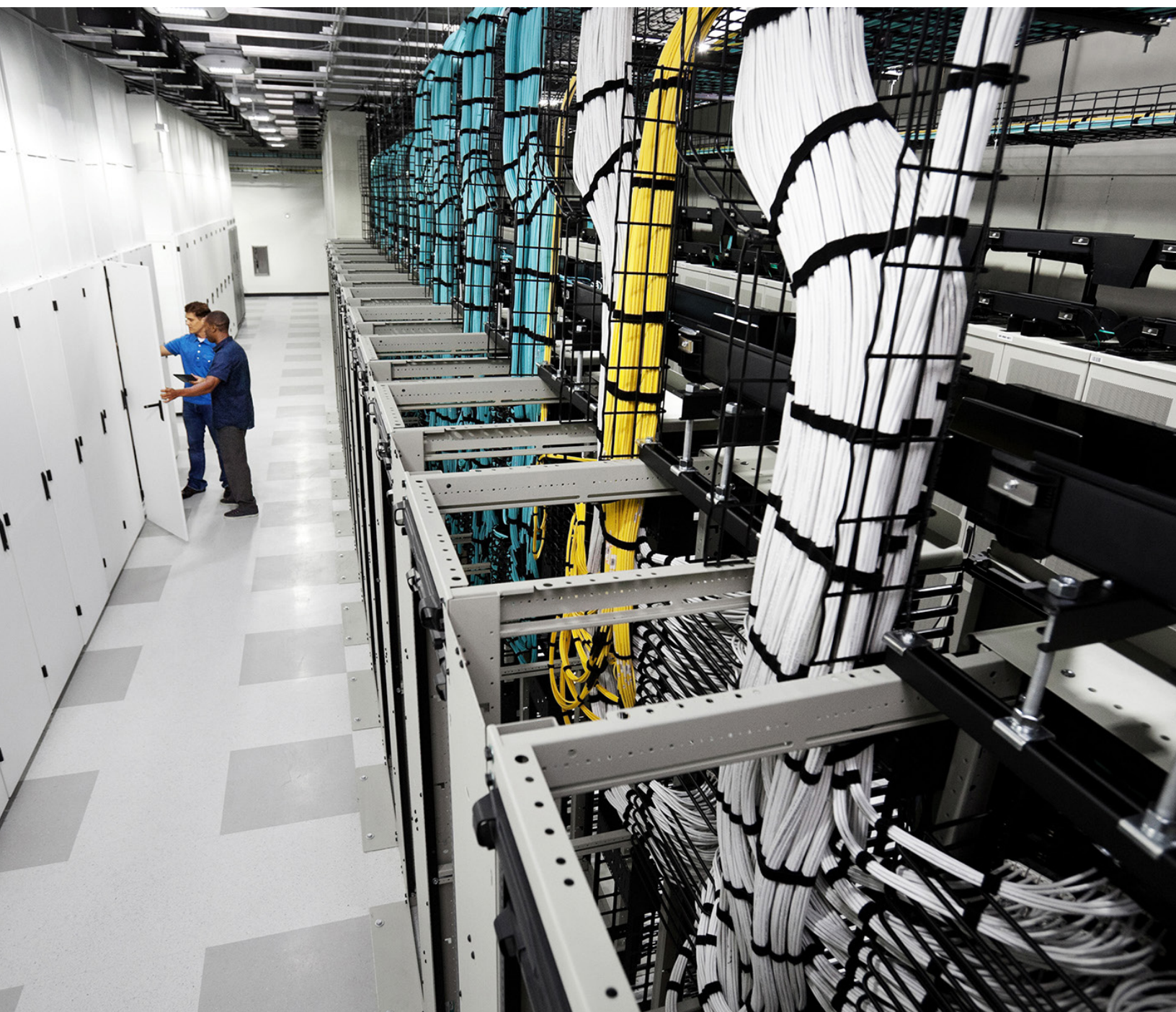


Cisco Cloud and Managed Services for HyperFlex

Addressing the challenges of concurrent use of
overprovisioning



Summary

As part of the management of HyperFlex and hyperconverged infrastructure, Cisco® Cloud and Managed Services (CMS) proactively monitors the overprovisioning levels and keeps track of the allocation and utilization of resources at different virtualization layers, providing comprehensive visibility for capacity management.

Virtualization has become ubiquitous in data centers today as a result of the benefits it delivers such as efficiencies through server consolidation, business agility, and business continuity. When deployed in combination with hyperconverged infrastructure, virtualization delivers features and benefits to shared infrastructure with enhanced flexibility and agility that can grow with the business needs.

One of the primary features of virtualization is overprovisioning. Overprovisioning is the practice of allocating more resources to a virtual server than physically available in the underlying hardware. Although overprovisioning enables higher asset utilization, concurrent use of virtualization at different levels requires coordinated management.

Cisco CMS maintains a holistic view of all virtualization elements and integrated components, including network, compute, and storage, keeping track of performance and capacity and proactively managing incidents that can affect the health of the solution.

As part of the management of hyperconverged infrastructure, CMS proactively monitors the overprovisioning levels and keeps track of the allocation and utilization of resources at different virtualization layers, providing comprehensive visibility for capacity management.

Data Center Virtualization and Resource Overallocation

Today's data centers are highly virtualized. Virtualization technologies were developed to solve the problem of underutilized resources while isolating the elements consuming these resources. For example, VMware vSphere virtualizes servers, intermediating the allocation of hardware resources (CPU, memory, and disk) to virtual machines.

A primary feature of virtualization is the efficient utilization of resources. It is a common practice to allocate to virtual machines more resources than the hardware can accommodate, assuming the workloads will not use the resources at the same point in time. Even the oversized workload will not consume all the resources allocated to them. This practice of overprovisioning increases efficiency and decreases the cost of ownership of servers and storage.

It's a common practice to overprovision the storage by rates of up to 1:2, given compression and deduplication techniques are used at some point.

Highly virtualized environments do have challenges. When overlapping different virtualization technologies, IT administrators can lose track of overprovisioned levels of real resources. This challenge can take place because one layer of virtualization can overallocate a virtualized resource that might have already been overprovisioned.

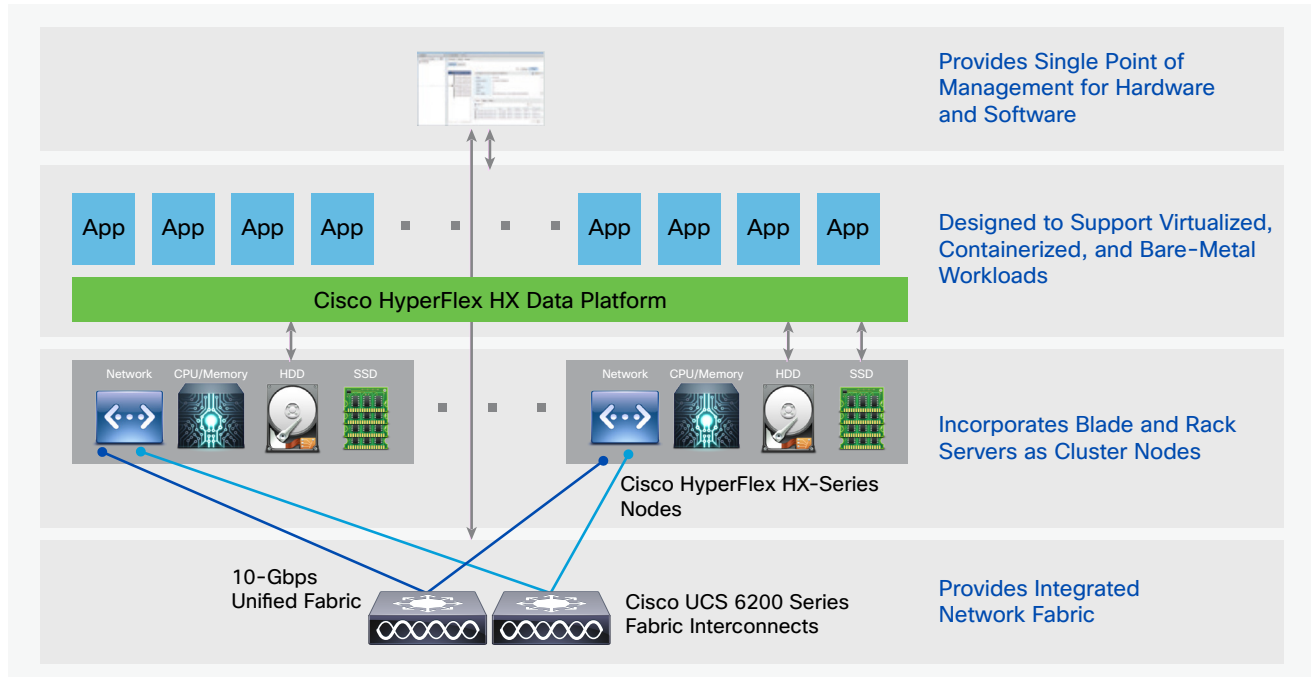
The following are reasons why datastores might be overprovisioned:

- Planned overallocation of resources
- Virtual machines committed space (vDisks)
- Virtual machines thin-provisioned vDisks
- Virtual machines uncommitted space for snapshots (this effectively take up to 2 times the size of all vDisks per snapshot)

Cisco HyperFlex

Cisco HyperFlex is a hyperconverged infrastructure product that uses HX220c and HX240c servers' local disks and abstracts them in the form of a single storage system that can be used to create and mount VMware vSphere NFS datastores on the underlying ESXi clustered hosts. (See Figure 1.)

Figure 1. Cisco HyperFlex Solution



Hyperconverged infrastructure significantly reduces the time and complexity to scale out compute nodes and storage by using local disks and avoiding the complexity of building and configuring legacy storage systems over SAN or Ethernet.

The Cumulative Overprovisioning Challenge

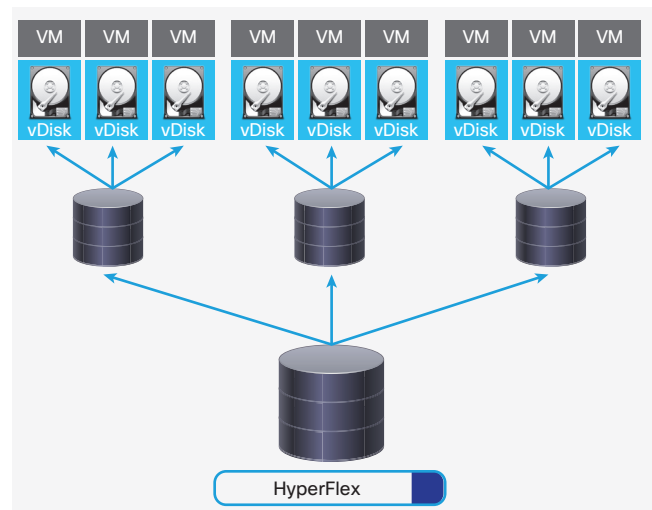
Cisco HyperFlex in conjunction with VMware vSphere uses virtualized compute and storage. The double layer of storage virtualization brings a challenge that is referred to hereafter as “cumulative overprovisioning.” Both technologies allow for overallocation of storage resources for higher efficiency of use. For example, taking a top-down approach, VMware allows for datastore overprovisioning, and HyperFlex also allows for overprovisioning of the real storage cluster capacity, creating a double layer of storage space overallocation.

When VMware vSphere datastores are created on a Cisco HyperFlex storage cluster, the system allows for the sum of datastore capacity to be higher than the actual storage cluster capacity.

When a user creates VMware vSphere virtual machines in vCenter, the system allows virtual machines to take more space on the containing datastore than the actual capacity of the datastore by using thin-provisioned disks and snapshots (accounted as uncommitted space).

The risk goes beyond running out of datastore space based on how much space virtual machine vDisks are taking (committed and uncommitted space). It is important to remember that as virtual machines start filling their vDisks, not only are the datastores filling up with disk information, which increases exponentially if there is uncommitted space utilization in the case there are snapshots of the virtual machines, but the HyperFlex storage cluster (which is the real storage space) is also filling up, at a higher rate. Datastore on VMware vSphere infrastructure is contributing to filling HyperFlex storage space. (See Figure 2.)

Figure 2. Cisco HyperFlex Storage Virtualization Layers



IT administration often overlooks the underlying storage space and takes into account only VMware vSphere datastore capacity planning.

Sample Case

Table 1 lists solution use case details.

Table 1. Cisco HyperFlex Solution Use Case Details

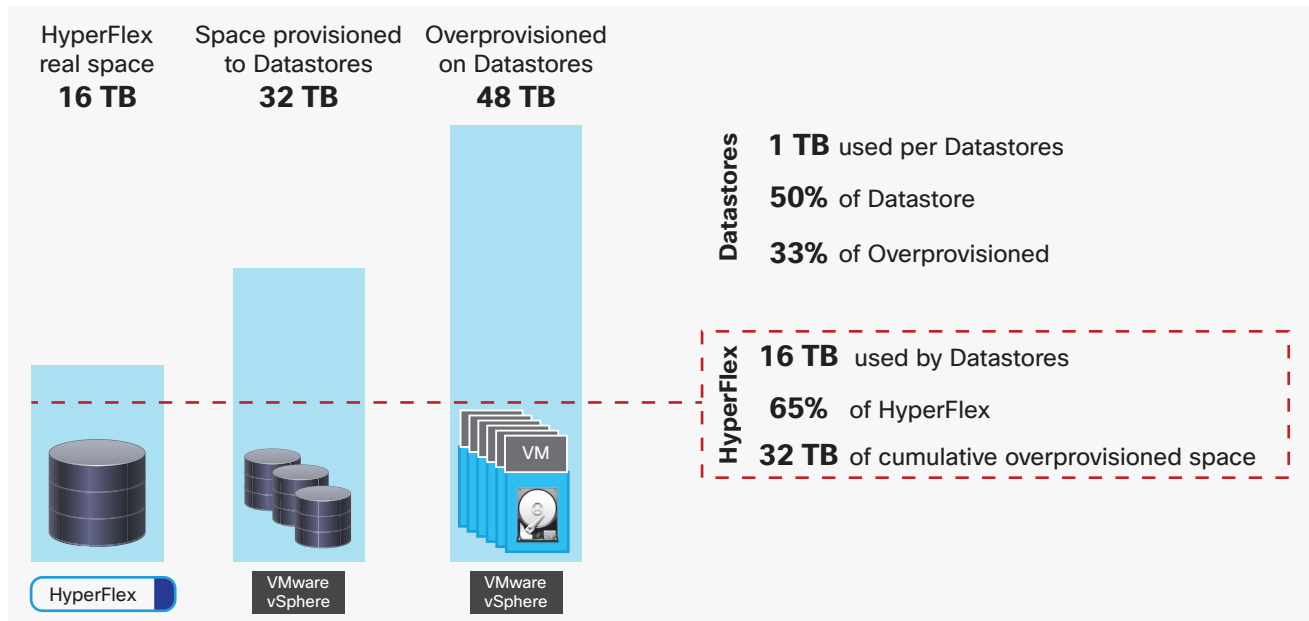
HyperFlex space	16 TB
HyperFlex savings (deduplication/compression)	35%
Number of datastores	16
Capacity of datastores	2 TB
Overprovision per datastore	1 TB
Total provisioned to datastores	32 TB
Overprovisioned (HyperFlex)	16 TB
Overprovisioned (datastores)	16 TB

For example, consider a use case where a HyperFlex storage cluster with an effective capacity of 16 TB (usable size of a traditional 8 converged nodes cluster with HX220c appliance) and total savings (deduplication and compression) of 35% is used to provision 16 Datastores with 2TB each (see Table 1).

Table 2. HyperFlex Usage When Datastores Reach 1 TB of Used Space

Used space per datastore (50%)	1 TB
Used space by datastores	16 TB
Potential space required by datastores	48 TB
Percent used of potential required space	33%
HyperFlex savings (deduplication/compression)	35%
HyperFlex space used	65%
Total cumulative overprovisioned space	32 TB

Figure 3. HyperFlex Usage When Each Datastore Reaches 1 TB of Used Space



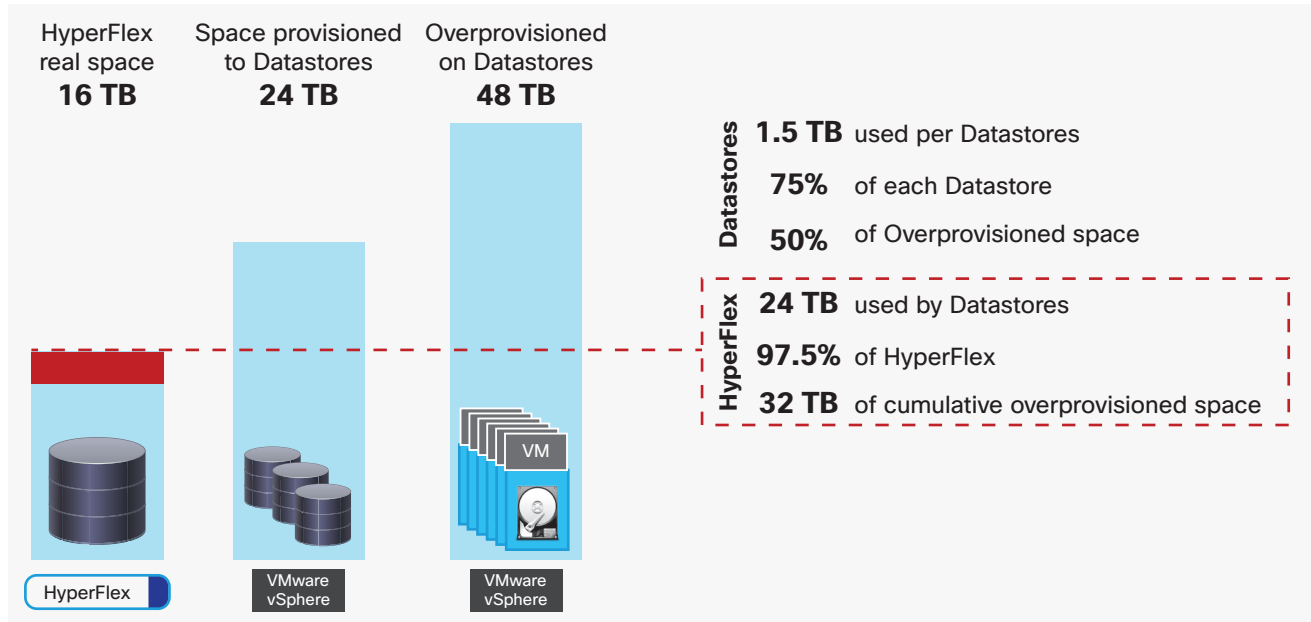
For the same example, consider that each datastore will be overprovisioned by 1 TB given all virtual machines have a snapshot.

Considering used space on each datastore is around 1 TB, it is already expected that HyperFlex will be at 65 percent capacity (see Table 2 and Figure 3).

Table 3. HyperFlex Usage When Each Datastore Reaches 1.5 TB of Used Space

Used space per datastore (75%)	1.5 TB
Used space by datastores	24 TB
Potential space required by datastores	48 TB
Percent used of potential required space	50%
HyperFlex savings (deduplication/compression)	35%
HyperFlex space used	97.5%
Total cumulative overprovisioned space	32 TB

Figure 4. HyperFlex Usage When Datastores Reach 1.5 TB of Used Space



When datastore used space reaches 1.5 TB, around 75 percent each, which is not a concerning utilization level, they will sum up to 24 TB, and real storage will be operating at its capacity limit (considering compression and deduplication of data are increasing storage efficiency) and can run out of space at any time (see Table 3 and Figure 4).

If not planned and tracked properly, real storage space can be exhausted before the datastore reaches its provisioned capacity and can severely affect the entire virtual infrastructure.

Some of the problems associated with filling up the storage space include:

- Virtual machines that stop running
- Unresponsive applications
- Database corruption
- Services outage

CMS and Cumulative Overprovisioning Tracking

Cisco Cloud and Managed Services provides a complete portfolio of day 2 operate services that increase the overall availability of IT while customers can focus on driving innovation and delivering on business demands. (See Figure 5.)

Figure 5. Cloud and Managed Services Capabilities



The Cisco CMS platform, through the use of the respective HyperFlex, Cisco UCS® Manager, and vCenter APIs, discovers all components and subcomponents, including HyperFlex converged and compute nodes, datastores, underlying servers, and virtual machines. It can further establish dynamic component mapping and relationships (DCM-R) between relevant solution components. After components and subcomponents are discovered, the system will extract configuration, capacity, and performance metrics for monitoring, threshold, events, and reporting purposes.

In addition to maintaining the health of the HyperFlex, CMS tracks the HyperFlex Storage Cluster and vSphere Datastores capacity, usage, provisioned and overprovisioned levels. CMS tool applies analytics to estimate the cumulative end-to-end overprovisioned space from VMs allocated to Datastores down to HyperFlex real storage capacity and generates events and incidents based on utilization. CMS also provides detailed reports with customer defined thresholds to identify the Datastores with higher levels of overprovisioning, risk status and HyperFlex real storage risk status.

CMS uses collected data to analyze the Datastores' capacity, usage and overprovisioned levels (including uncommitted space), identify and highlight critical levels of overprovisioning.

CMS system will further account for all overprovisioned space across all Datastores and compare to the HyperFlex Storage Cluster capacity, usage and overprovisioning level to calculate the cumulative overprovisioned space over the HyperFlex Storage.

Based on usage vs. overprovisioning ratio, Datastores and HyperFlex Cluster are marked "AT RISK" or "OK".

Customer has access to the analytic reports via the CMS customer service portal. These reports allow customers to closely track resources usage and preventing out of storage space that can cause a collapse of the virtual infrastructure.

Cisco CMS helps our customers simplify the adoption of new technologies, allows them to maximize their return on their investment in the new technologies, and most importantly to effectively and successfully drive their business.

For More Information

Get started today with Cisco Cloud and Managed Services data center solutions by visiting us on Cisco.com or sending email to askcms@external.cisco.com.