

Accelerate Storage Initiatives with Cisco Intersight Workload Optimizer and FlashStack

Contents

Executive summary	3
Introduction	3
Solution components	4
FlashStack with Cisco Intersight Workload Optimizer	7
Optimizing FlashStack	12
Recommendations	21
Working with policies	31
Conclusion	34
For more information	34

This document describes how Cisco Intersight™ Workload Optimizer with FlashStack enables enterprises to ensure that applications get the guaranteed full-stack quality of service, while efficiently using compute, network, and storage infrastructure and preventing unnecessary overprovisioning.

Executive summary

This document describes how to use the Cisco Intersight™ Workload Optimizer with FlashStack to enhance an organization's application modernization and achieve better infrastructure performance. It provides workload sizing and scaling guidance and describes how to ensure that applications get the compute, network, and storage performance they require to operate virtual infrastructure reliably, while efficiently using infrastructure and preventing unnecessary overprovisioning.

Introduction

Flash storage addresses many of the performance issues for I/O-intensive applications deployed in virtualized data centers. However, problems or bottlenecks can also exist in the compute and network layers, affecting overall application performance. Eliminating performance bottlenecks in just one layer will not expose performance problems in other data center layers. You need to address performance issues across the entire stack, including the compute, network, and storage layers.

With Cisco Intersight Workload Optimizer and FlashStack, enterprises can solve their most important virtualization challenges, improve infrastructure performance, and make full use of their investment in high-performance, high-efficiency Infrastructure.

Main benefits

The benefits of Cisco Intersight Workload Optimizer with FlashStack include the following:

- Ensure the performance for I/O-intensive applications.
- Get the most value from compute, network, and storage resources.
- Gain complete visibility into the entire stack, including application, virtualization, and infrastructure layers.
- Automate the placement of I/O-intensive workloads.
- Prevent bottlenecks at every layer of the data center.
- Simplify management and scaling.

Solution components

Cisco Intersight platform



Figure 1.
Cisco Intersight - Connecting people and technology in a hybrid world

The Cisco Intersight platform is a cloud-based modular platform of integrated SaaS offerings that address your unique requirements and use cases, from managing Cisco data center and 3rd party infrastructure to executing complex workflows for workloads and applications in public clouds. (Figure 1). It is an IT operations management (ITOM) platform through wherein which you can deploy, manage, and support your distributed environment via in the cloud. All that is required is the latest release of the Cisco® software, and you can begin achieving realizing better operations automation in a matter of minutes.

Figure 2 shows a high-level view of the multiple management stages of the Cisco Intersight platform.

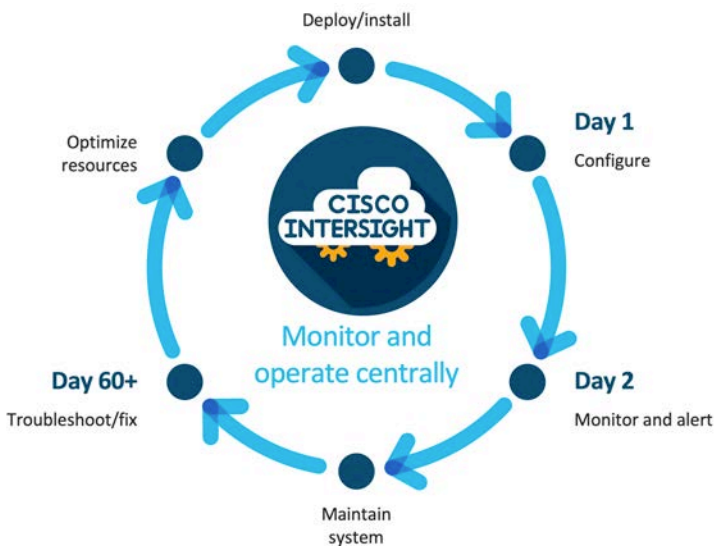


Figure 2.
Cisco Intersight Infrastructure lifecycle

The Cisco Intersight platform provides benefits for the entire infrastructure lifecycle. Our existing Cisco tools—, Cisco UCS® Manager, Cisco UCS Director, and Cisco Integrated Management Controller (IMC)— focus on day-0 and day-1 activities related to around deployment and configuration. Cisco Intersight provides benefits for day 2 and beyond as well. These benefits includes more than just monitoring and alerting. The integration with the Cisco Technical Assistance Center (TAC), predictive analytics, and resource optimization addresses ongoing operations and systems upgrades.

Cisco Intersight Workload Optimizer

Cisco Intersight Workload Optimizer (IWO) is an optional service that can be purchased with or without Cisco Intersight. It extends the optimization capability in Cisco Intersight. It bridges the gap between applications and infrastructure to ensure assure the performance of critical workloads while increasing maximizing efficiency and reducing costs across on-premises and in hybrid cloud environments.

For IT teams challenged with balancing performance, efficiency, and cost in today’s complex, hybrid cloud environments, Cisco Intersight Workload Optimizer is a real-time decision engine that visualizes complex interdependencies across the stack. It simplifies and automates application resource management at scale by leveraging using telemetry data from over 50 targets across a range of hypervisors, compute platforms, container platforms, public clouds, and more with a single tool. It relieves IT teams from the day-to-day management of infrastructure and resource availability, giving them time to promote drive innovation for the business (Figure 3).

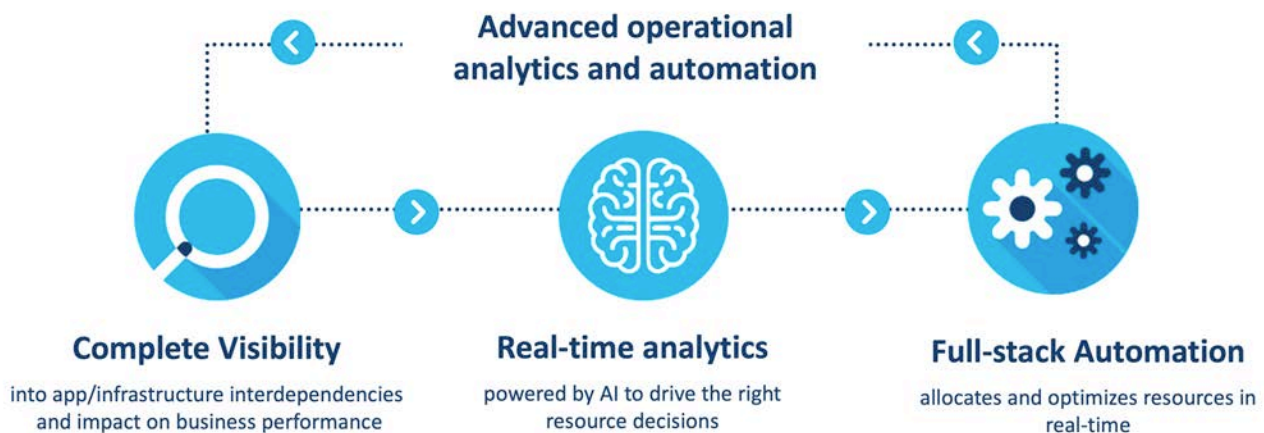


Figure 3. Cisco Intersight Workload Optimizer: - Visibility -> Insight -> Action -> Visibility

Cisco Intersight Workload Optimizer makes it easy to see what is happening in your environment and understand how that affects your applications. You can do the following:

- Gain visibility into the health, use, and performance of your infrastructure stack. Your operations and application teams can discover application and infrastructure interdependencies to make more informed decisions about how to apply and use IT resources.
- Get insight into infrastructure bottlenecks and factors that increase costs. Intelligent analytics stitch together each layer of the application and infrastructure stack, allowing resourcing decisions to be tied to application demand and relevant policies and constraints while factoring in available capacity.
- Trust actions that continuously optimize your infrastructure to deliver application performance. Specific real-time actions ensure that your workloads get the resources they need when they need them for

placement, scaling, and capacity. You can automate the software's decisions to match your level of comfort: recommend (view only), manual (select and apply), or automated (implemented in real time by software).

FlashStack

FlashStack is an exceptional, high-performance converged infrastructure solution that combines compute, network, and storage to provide a modern infrastructure platform for business-critical applications, DevOps, and analytics. It integrates Cisco UCS and Pure Storage All-Flash storage, Cisco Nexus® Family switching, and cloud-based management (Figure 4).

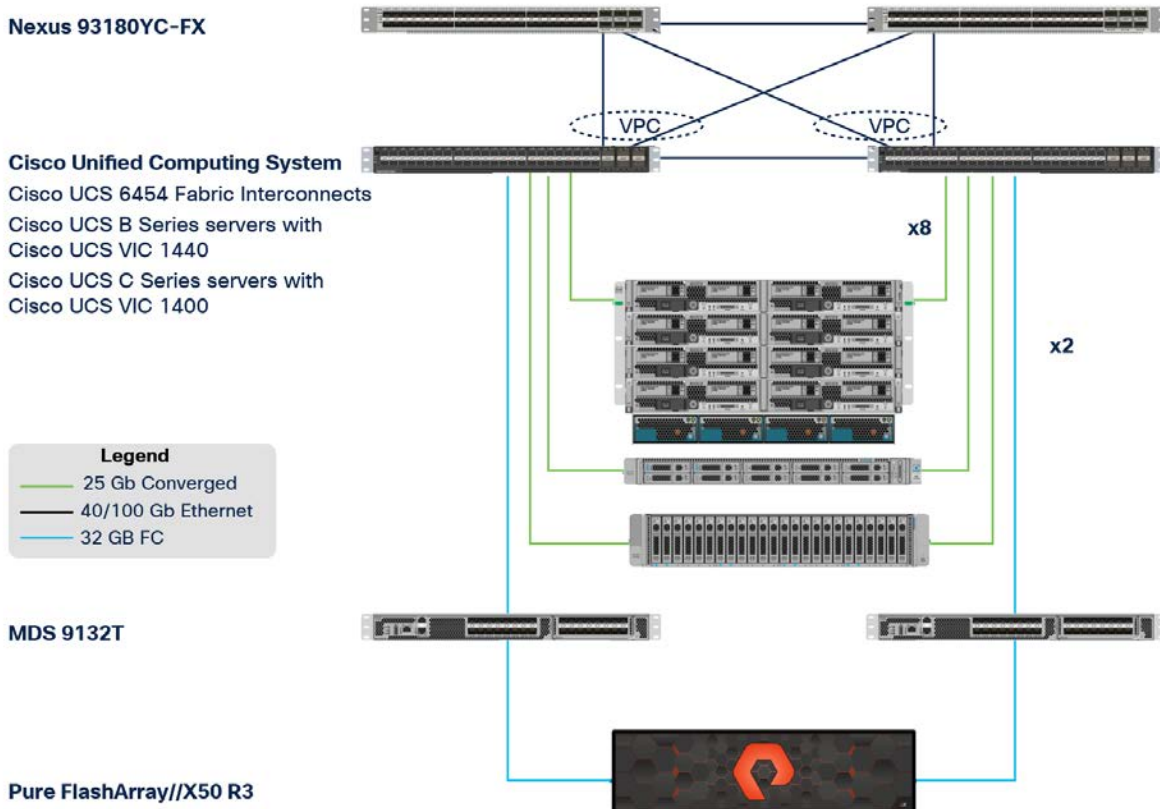


Figure 4.
FlashStack overview

The modern data experience decouples storage software from its hardware, giving you greater scalability, flexibility, and control over your data storage infrastructure. By using FlashArray for data storage, FlashStack delivers a continuous-integration solution augmented with the following benefits:

- Simple architecture that eliminates disparate hardware silos without any trade-offs
- Validated interoperability and confident application deployment
- Infrastructure for both traditional and converged operating models so you can consolidate operations at your own pace
- Continuous integration for multi-hypervisor, bare-metal, and container deployments
- Built for the cloud, including full integration with cloud platforms from Cisco, VMware, OpenStack, and others

FlashStack with Cisco Intersight Workload Optimizer

Prerequisites

1. Cisco Intersight Workload Optimizer must be enabled.
2. Communication must be established between FlashArray targets in the data center and Cisco Intersight Workload Optimizer. Follow these steps:
 - Install a Cisco Intersight Assist appliance in the on-premises data center. The target service must be accessible to the Cisco Intersight Assist appliance.
 - Connect the Cisco Intersight Assist instance with Cisco Intersight.
 - Log in to Cisco Intersight and claim the Cisco Intersight Assist instance as a target.

Cisco Intersight Assist provides a secure way for connected targets to send information and receive control instructions from Cisco Intersight Workload Optimizer, using a secure internet connection. For more information, see the Cisco Intersight Assist Getting Started Guide.

3. You need a service account that Cisco Intersight Workload Optimizer can use to connect to the FlashArray

This account needs privileges to run commands through the Pure Storage API: typically the default pureuser administrative account.

Enabling Cisco Intersight Workload Optimizer

You can enable Cisco Intersight Workload Optimizer either by using a trial license or by purchasing and registering a Cisco Intersight Workload Optimizer license tier.

To enable a trial license, follow these steps:

1. Log in to Cisco Intersight with account administrator privileges.
2. Click the Settings icon and choose Settings > License.
3. Click Start Trial.
4. Choose Workload Optimizer.

You can purchase any of the Cisco Intersight Workload Optimizer license tiers using the Cisco Ordering Tool. For details, see Ordering Information in the Cisco Intersight data sheet.

To register a Cisco Intersight Workload Optimizer license in Cisco Intersight, do the following:

1. Log in to Cisco Intersight with account administrator privileges.
2. From Settings icon > Settings > License, click Register.
3. On the Set Token page, enter the product instance registration token. Click Cisco Smart Software Manager to obtain your Cisco Intersight registration token. If you do not have a Cisco Smart Account, create one here. You can purchase the subscription and select the required Cisco UCS server volume tier for the selected subscription duration from the same Smart Account. Click Next.
4. On the Set Product page, Toggle on Workload Optimizer and select the required license tier (Essentials, Advantage, or Premier). For more information about the Cisco Intersight Workload Optimizer licensing tiers, see the Cisco Intersight Workload Optimizer Licensing section at https://intersight.com/help/resources/cisco_intersight_workload_optimizer_getting_started_-_intersight_workload_optimizer_licensing

Viewing the supply chain

In Cisco Intersight Workload Optimizer, if you navigate to Workload Optimization > Overview, you can see the consolidated supply chain from Cisco Intersight Workload Optimizer (Figure 5). Market abstraction is fundamental to Workload Optimizer, which models all the elements of applications and infrastructure into a supply chain of buyers and sellers and show the relationships among the various elements. This supply chain represents the flow of resources from the data center, 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 Optimizer provides closed-loop management of resources, from the data center through to the application.

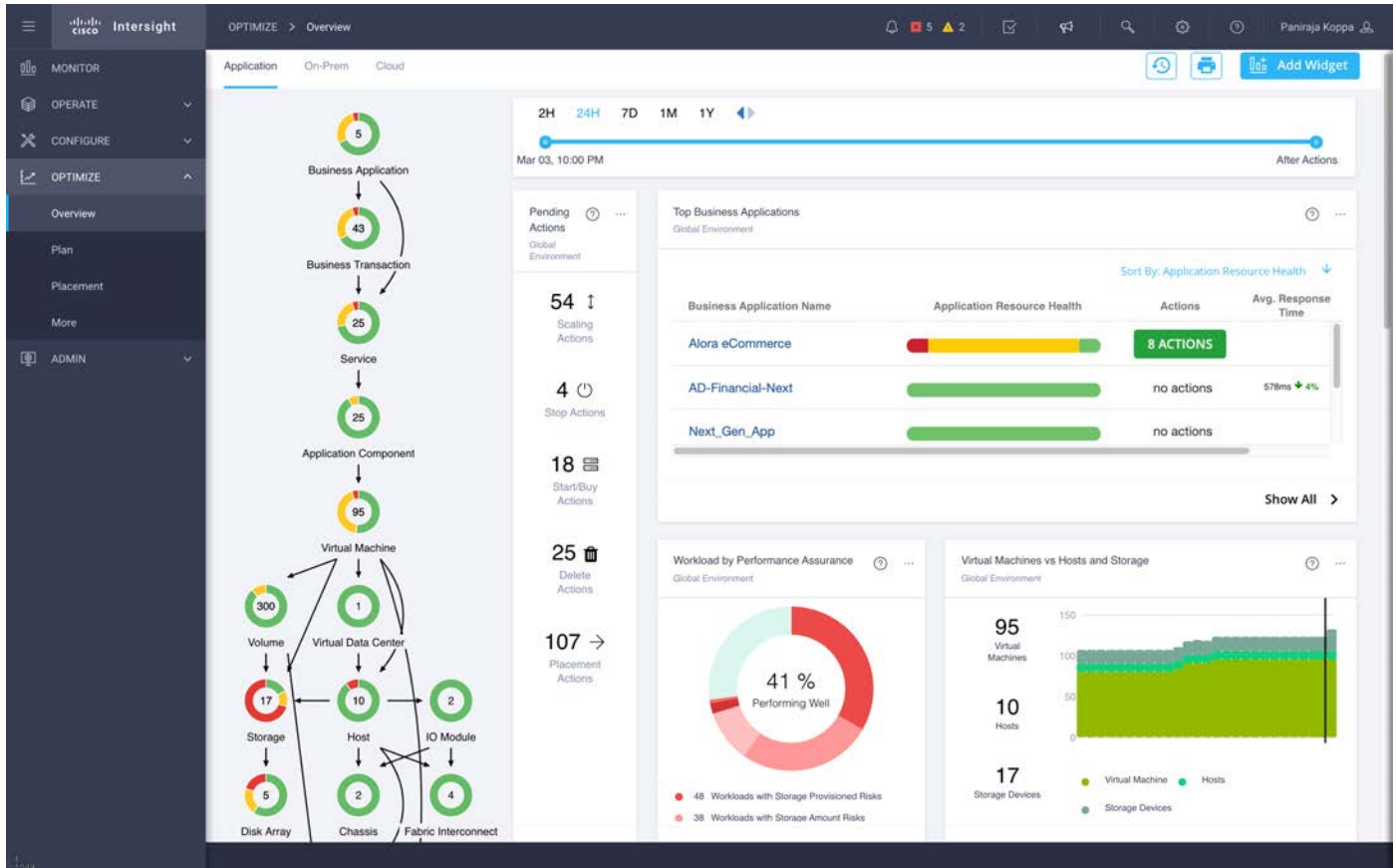


Figure 5.
Overview page

In Figure 5, observe the Pending Actions widget. Workload Optimizer is making multiple recommendations for workloads to perform at the best possible cost while maintaining compliance. It has actions you can take on various elements: scale up or down, delete, start or buy, make a placement, make configuration changes, and stop action.

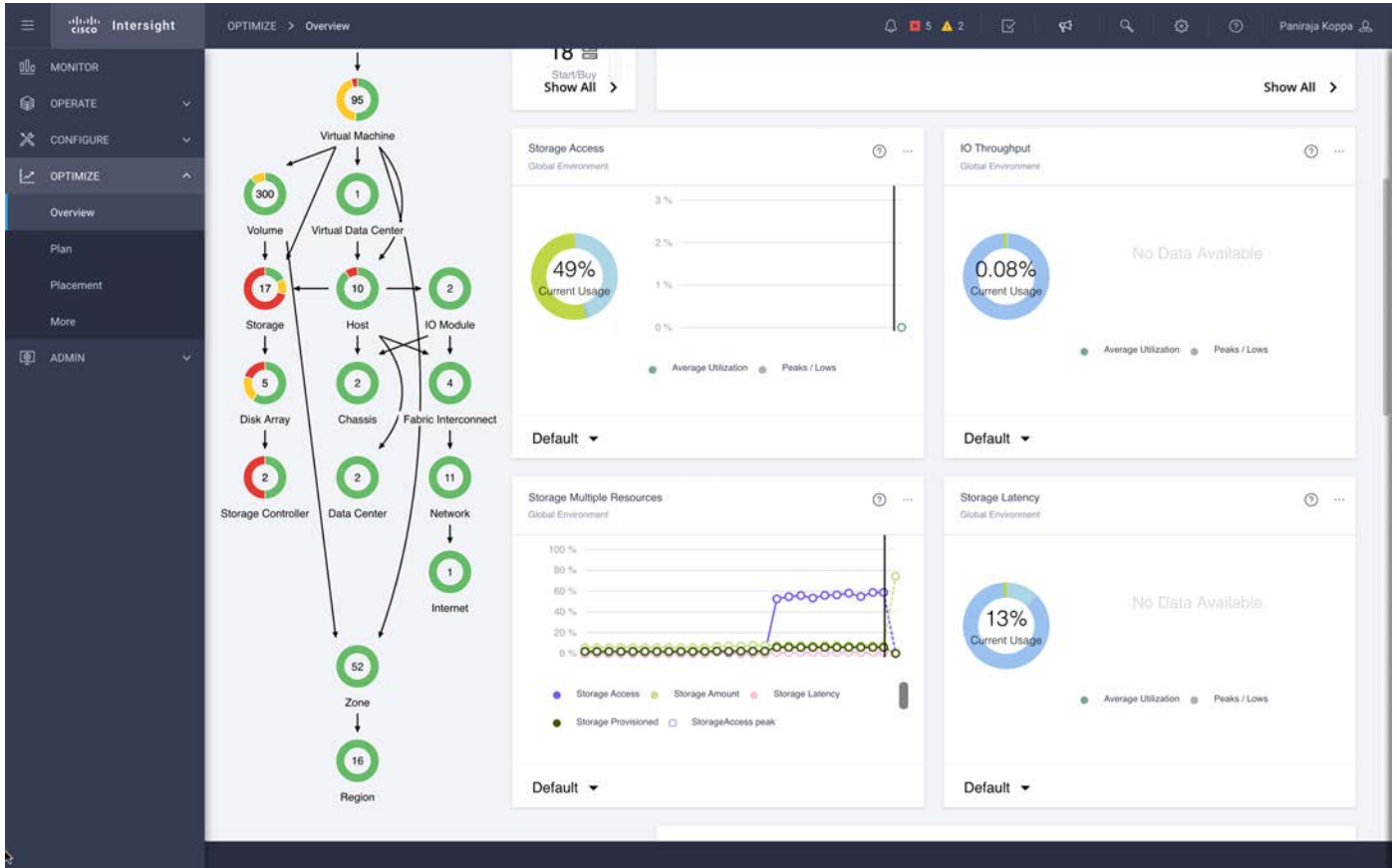


Figure 6.
Overview page with a few storage-related widgets

The view in Figure 6 enables you to really understand what you have from perspective of the infrastructure. You can then look at issues associated with workloads.

Claiming targets

The supply chain view is automatically created using the information gathered from the endpoints you have added. No manual input is required to create this relationship map. Cisco Intersight Workload Optimizer uses targets to monitor workloads and to implement actions. For each target, Workload Optimizer communicates with the service through the management protocol that it exposes: the representational state transfer (REST) API, Storage Management Initiative Specification (SMI-S), XML, or some other management transport. Workload Optimizer uses this communication to discover the managed entities, monitor resource utilization, and implement actions.

In the FlashStack data center, you can claim the targets listed in Table 1. A broad variety of other targets are supported as well.

Table 1. FlashStack-related targets and licensing information

Category	Target	Minimum licensing tier required	Cisco Intersight Assist required
Compute and fabric	Cisco UCS server (standalone) Cisco UCS domain (Cisco UCS Manager managed)	Cisco Intersight Workload Optimizer Essentials	No
Storage	Pure Storage FlashArray	Cisco Intersight Workload Optimizer Essentials	Yes
Hypervisor	VMware vCenter	Cisco Intersight Workload Optimizer Essentials	Yes
Application performance management	Cisco AppDynamics	Cisco Intersight Workload Optimizer Advantage	Yes
Cloud native	Kubernetes	Cisco Intersight Workload Optimizer Advantage	No

To add a target service, follow these steps:

1. Navigate to Admin > Targets > Claim a New Target.
2. Provide the requested information and click Apply to validate those targets and start a new discovery.

Typical information you provide includes the following:

- Target type: Choose among the supported technologies. After you choose the technology, then choose the specific target type for that technology. For example, for Cloud Management, you can choose AWS.
- Host name or IP address: Specify the address of the target service you want to add.
- Username: Enter a valid account username for the target service.
- Password: Enter a password for the target service account.

Claiming Cisco UCS

If the installation of Cisco Intersight has already claimed a Cisco UCS device, then Cisco Intersight Workload Optimizer discovers the Cisco UCS environment automatically.

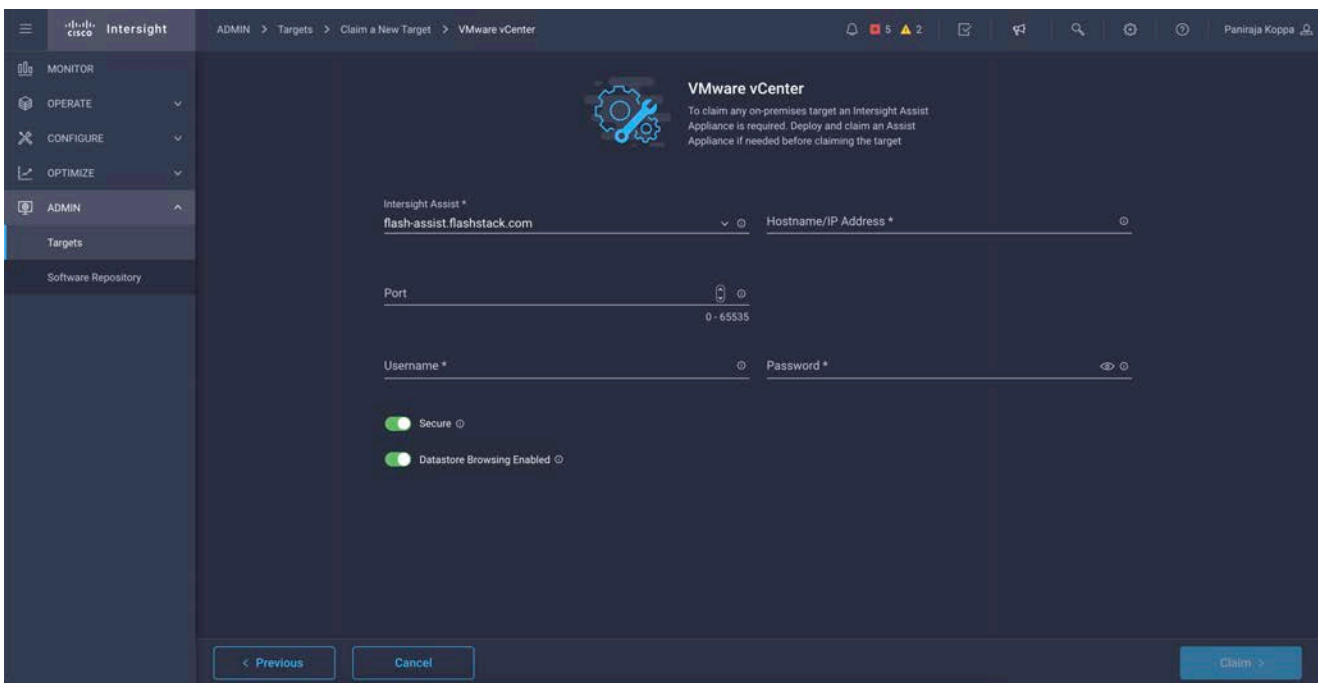
To claim a new Cisco UCS device, follow these steps:

1. Select the Compute / Fabric category and choose the type of device you want for a target.
2. Provide the following information:
 - Device ID: Enter the applicable device ID. Endpoint devices connect to the Cisco Intersight portal through a device connector that is embedded in the management controller (management virtual machine for Cisco UCS Director) of each system. The device connector provides a secure way for connected devices to send information and receive control instructions from the Cisco Intersight portal, using a secure Internet connection.
 - Claim code: The claim code authorizes your access. You can find this code in the device connector.
3. After you provide the information, click Claim. You can see the status of your claimed target on the Targets tab.

Claiming VMware vCenter targets

To claim VMware vCenter targets, follow these steps:

1. Choose the Hypervisors > VMware vCenter on the Target Configuration page
2. Provide the following information:
 - Host name or IP address: Provide the name or IP address of the vCenter server.
 - Username and password: Provide credentials for the user account that Cisco Intersight Workload Optimizer can use to connect to the vCenter server. Include the domain if required (<domain>\<username>).
3. Enable datastore browsing so that Cisco Intersight Workload Optimizer can discover wasted storage (Figure 7).



The screenshot shows the 'VMware vCenter' configuration page in the Cisco Intersight interface. The breadcrumb navigation is 'ADMIN > Targets > Claim a New Target > VMware vCenter'. The page title is 'VMware vCenter' with a sub-header 'To claim any on-premises target an Intersight Assist Appliance is required. Deploy and claim an Assist Appliance if needed before claiming the target.' The form contains the following fields and options:

- Intersight Assist ***: A dropdown menu with the value 'flash-assist.flashstack.com'.
- Hostname/IP Address ***: A text input field.
- Port**: A text input field with a value of '0 - 65535'.
- Username ***: A text input field.
- Password ***: A text input field with a toggle for visibility.
- Secure**: A checked toggle switch.
- Datastore Browsing Enabled**: A checked toggle switch.

At the bottom of the form, there are three buttons: '< Previous', 'Cancel', and 'Claim >'.

Figure 7.

Claiming VMware vCenter targets

Claiming Pure Storage FlashArray targets

To claim a FlashArray target, follow these steps:

1. Choose Admin > Targets > Claim New Target > Storage > Pure Storage FlashArray on the Target Configuration page (Figure 8).

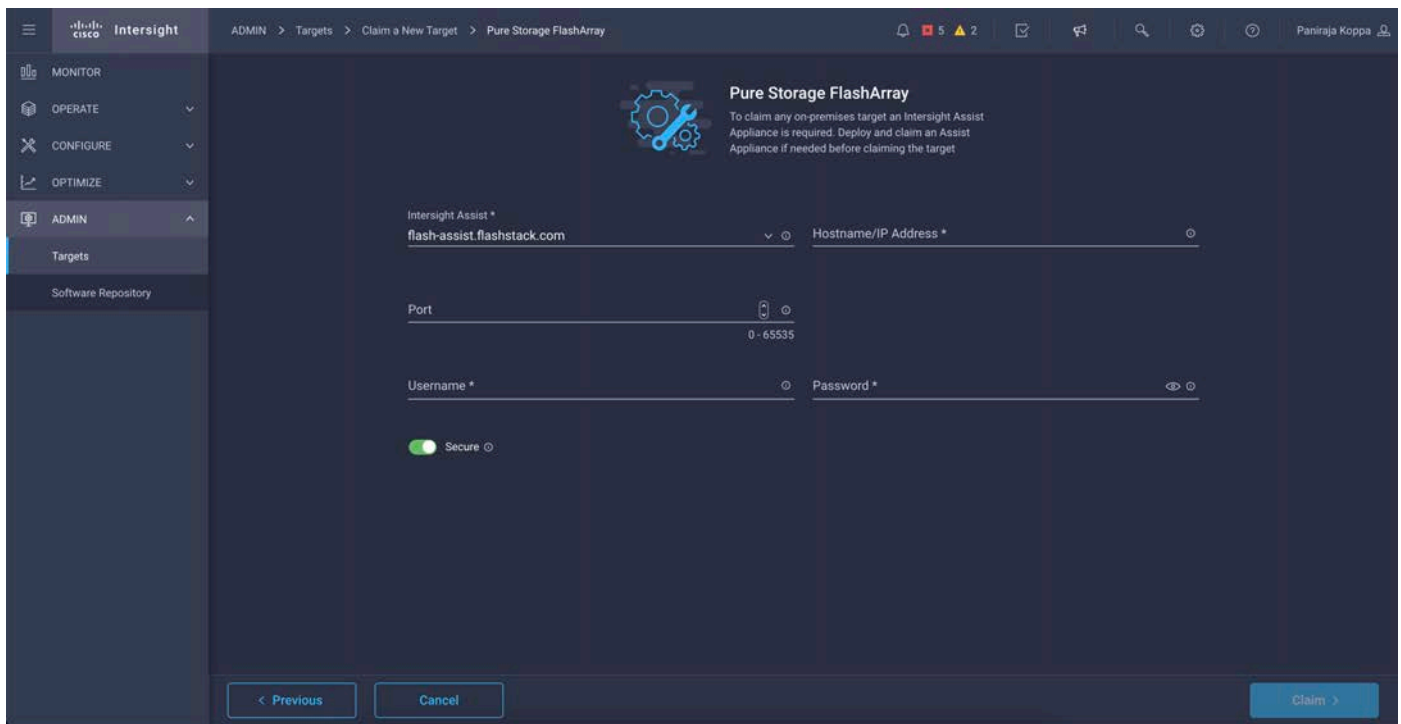


Figure 8.
Claiming a Pure Storage FlashArray target

2. Provide the following information:
 - Address: Provide the name or IP address of the Pure Storage FlashArray.
 - Username and password: Provide the credentials for the service account that Cisco Intersight Workload Optimizer can use to connect to the FlashArray. The username must not contain the domain. For example, Username=jjsmith is correct, but Username=myDomain\jjsmith will result in a failure to validate.
3. Enable Secure option for a secure connection. When this option is enabled, SSL is used to connect to the Pure target. Most Pure installations do not accept insecure connections. If you receive an error message when adding a target with secure connections disabled, try adding the target again with this option enabled.

Optimizing FlashStack

To perform intelligent workload balancing, Cisco Intersight Workload Optimizer collects raw data from various elements of FlashStack such as FlashArray, VMware vCenter, Cisco UCS blade and rack servers, and Kubernetes clusters. It polls these targets at 10-minute intervals to collect the latest data samples. It then uses these 10-minute data points to perform analysis and to display data in the GUI.

Storage entities

After validating the new target, Cisco Intersight Workload Optimizer discovers the connected storage entities. Table 2 compares the terms used in Pure Storage FlashArray with those used in Cisco Intersight Workload Optimizer.

Table 2. Storage entities and FlashArray names

Pure Storage name	Cisco Intersight Workload Optimizer entity
Volume	Storage
Direct Flash Modules (DFM) - SSDs Direct Flash Shelves - Array Shelf	Disk array
Controller	Storage controller

Cisco Intersight Workload Optimizer represents the hypervisor volume (datastore in VMware vSphere) as storage. A datastore is a logical grouping of one or more physical storage devices that serves workload storage requirements.

Monitored resources

Cisco Intersight Workload Optimizer monitors storage resources for Cisco UCS and virtual machines as shown in Table 3.

Table 3. Compute and virtual machine monitored resources

Entity type	Resource
Virtual machine	<p>Virtual memory (vMem)</p> <ul style="list-style-type: none"> The utilization of the VMem allocated to the hosting virtual machine Measured in kilobytes (KB) <p>Virtual CPU (vCPU)</p> <ul style="list-style-type: none"> The utilization of the vCPU allocated to the hosting virtual machine Measured in megahertz (MHz) <p>Virtual storage (vStorage)</p> <ul style="list-style-type: none"> The utilization of the virtual storage capacity allocated for the virtual machine Measured in KB <p>Storage access operations per second</p> <ul style="list-style-type: none"> The utilization of IOPS allocated for the vStorage on the virtual machine Measured in I/O operations per second (IOPS) <p>Latency</p> <ul style="list-style-type: none"> The utilization of latency allocated for the vStorage on the virtual machine Measured in milliseconds (ms)
Physical machine	<p>Memory (Mem)</p> <ul style="list-style-type: none"> The utilization of the physical machine's memory reserved or in use Measured in KB <p>CPU</p> <ul style="list-style-type: none"> The utilization of the physical machine's CPU reserved or in use Measured in MHz <p>I/O</p> <ul style="list-style-type: none"> The utilization of the physical machine's I/O adapters Measured in kilobytes per second (KBps)

Entity type	Resource
	<p>Net</p> <ul style="list-style-type: none"> • The utilization of data through the physical machine’s network adapters • Measured in KBps <p>Swap</p> <ul style="list-style-type: none"> • The utilization of the physical machine’s swap space • Measured in KB <p>Balloon</p> <ul style="list-style-type: none"> • The utilization of shared memory among virtual machines running on the host • VMware ESX only • Measured in KB <p>CPU ready</p> <ul style="list-style-type: none"> • The utilization of the physical machine’s allocated ready queue capacity (measured in KB) that is in use for 1, 2, and 4 CPU ready queues • VMware ESX only • Measured in MHz
Chassis	<p>Power</p> <ul style="list-style-type: none"> • Electricity being consumed by the chassis • Measured in watts (W) <p>Temperature</p> <ul style="list-style-type: none"> • Temperature of the internal components of the chassis • Measured in degrees Celsius (°C)
I/O module	<p>Net throughput</p> <ul style="list-style-type: none"> • Rate of message delivery over a port • Measured in megabits per second (Mbps)
Fabric interconnect	<p>Net throughput</p> <ul style="list-style-type: none"> • Rate of message delivery over a port • Measured in Mbps <p>Port channel</p> <ul style="list-style-type: none"> • Amalgamation of ports with shared net throughput and utilization • Measured in Mbps

Table 4 summarizes the monitored resources for FlashArray.

Table 4. Storage monitored resources

Entity type	Resource
Volume (datastore)	<p>Storage amount</p> <ul style="list-style-type: none"> The utilization of the datastore's capacity Measured in megabytes (MB) <p>Storage provisioned</p> <ul style="list-style-type: none"> The utilization of the datastore's capacity, including overprovisioning Measured in MB <p>Storage access operations per second</p> <ul style="list-style-type: none"> The summation of the read and write access operations per second on the datastore Measured in IOPS <p>Latency</p> <ul style="list-style-type: none"> The utilization of latency on the datastore Measured in ms
Pure Storage shelf array	<p>Disk array</p> <ul style="list-style-type: none"> Storage amount The utilization of the disk array's capacity Measured in MB <p>Storage provisioned</p> <ul style="list-style-type: none"> The utilization of the disk array's capacity, including overprovisioning Measured in MB <p>Storage access operations per second</p> <ul style="list-style-type: none"> The summation of the read and write access operations per second on the disk array Measured in IOPS <p>Latency</p> <ul style="list-style-type: none"> The utilization of latency computed from the latency of each device in the disk array Measured in ms
Controller	<p>CPU</p> <ul style="list-style-type: none"> The utilization of the storage controller's CPU in use Measured in MHz <p>Storage amount</p> <ul style="list-style-type: none"> 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 MB

Actions

Cisco Intersight Workload Optimizer makes multiple recommendations about how to make workloads perform best at the best possible cost while maintaining compliance. It has recommendations for actions you can take on various elements: scale up or down, move, delete, start or buy, make a placement, make configuration changes, and stop action.

Table 5 summarizes the actions for various FlashStack entities.

Table 5. Summary of actions

Entity type	Action
Virtual machine	<ul style="list-style-type: none"> • Provision additional resources (vMem and vCPU) • Move virtual machine • Move virtual machine storage • Reconfigure storage • Reconfigure virtual machine
Physical machines	<ul style="list-style-type: none"> • Start physical machine • Provision physical machine • Suspend physical machine
Chassis	<ul style="list-style-type: none"> • Provision new chassis
Fabric interconnect	<ul style="list-style-type: none"> • Add port to port channel • Remove port from port channel • Add port
Storage controller	<ul style="list-style-type: none"> • Recommendations to provision storage controller
Shelf array	Recommendations to: <ul style="list-style-type: none"> • Provision disk array • Start disk array • Suspend disk array • Move virtual machine • Move datastore
Storage (datastore)	Recommendations to: <ul style="list-style-type: none"> • Suspend • Delete • Move • Provision • Start • Resize (up, down, above maximum, or below minimum, using tuned scaling)

If you plan to automate certain actions, be aware that Cisco Intersight Workload Optimizer doesn't automate the same actions equally for all technologies, because the underlying technologies do not provide the same degree of automation. For example, if you set storage move actions to be automated for all virtual machines, Workload Optimizer will automate storage moves for virtual machines managed by vCenter and Red Hat Enterprise Version (RHEV). However, but it will not automatically perform storage moves for virtual machines managed by Microsoft Hyper-V or Citrix XenServer, because Hyper-V and XenServer do not provide programmatic access to the storage move operation. In this case, Workload Optimizer will recommend that you perform the storage move using the Hyper-V or XenServer console.

Refer to the Cisco Intersight Workload Optimizer User Guide for details.

Cisco Intersight Workload Optimizer user interface overview

Select the On-Prem tab to see details for the on-premises environment (Figure 9). You will see how applications are correlated with virtual machines, clusters, Cisco UCS hosts and chassis, fabric interconnects, storage, etc. You will also see various widgets specific to the on-premises data center.

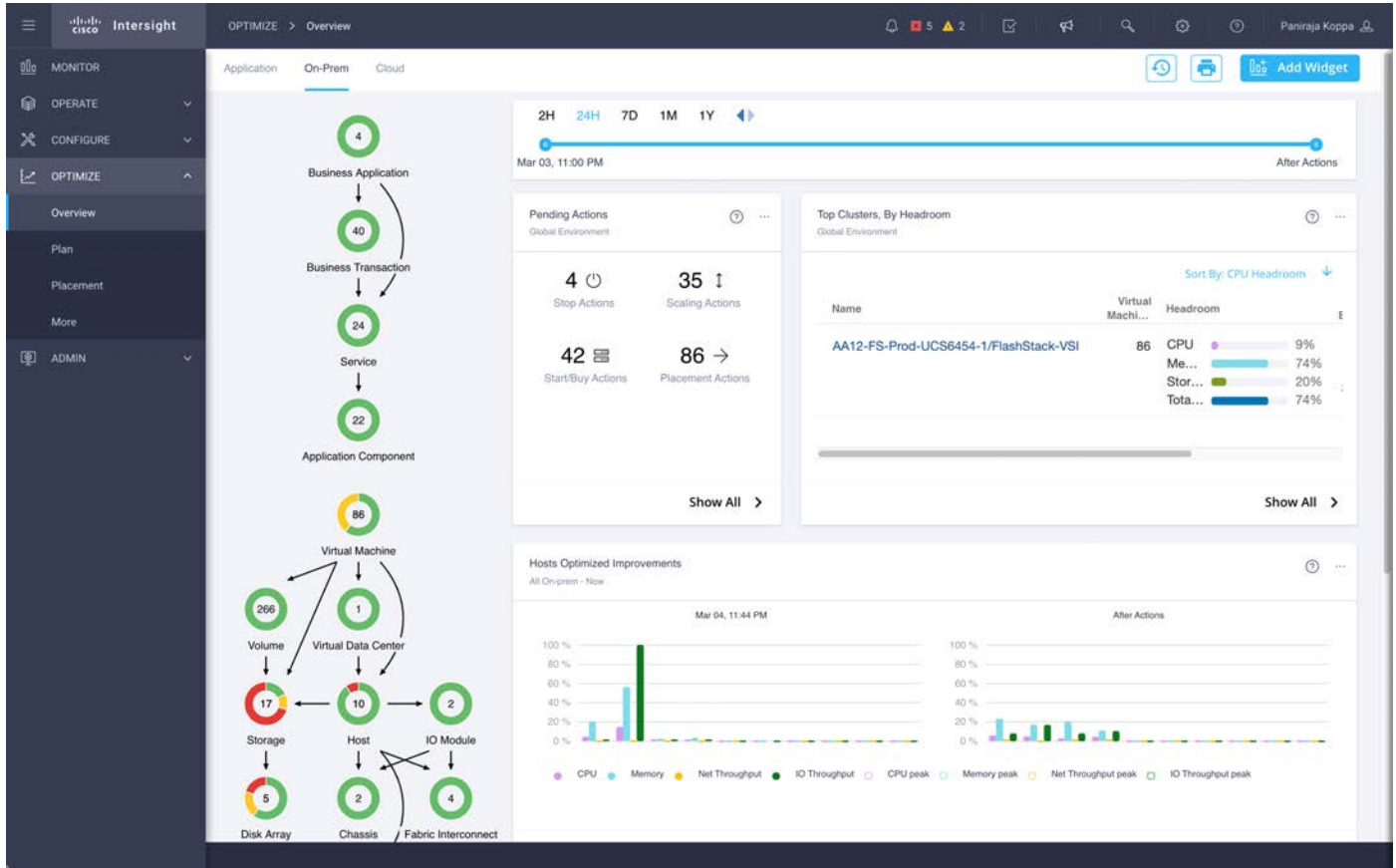


Figure 9.
On-premises view

The supply chain shows all the on-premises entities in your environment. The charts show details about your environment, including the following:

- **Overviews of pending actions:** Workload Optimizer makes multiple recommendations (such as scale up or down, delete, start or buy, placement, configuration change, and stop actions) for how to make workloads perform best with the best resource use. A list will be displayed.
- **Top cluster utilization:** A list of the most-used clusters is displayed. 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 implement the specific actions, click the Actions button for that cluster. To see all the clusters in your environment, click Show All.
- **Optimized improvements:** Compare the current resource utilization with the utilization you would see if you implemented all the pending actions.

Action history

You can see a history of all actions that have been recommended and implemented, or of just the actions that have been accepted and implemented.

By looking at the supply chain, you can see the following:

- The number of entities you have on each tier: Each entry in the supply chain gives a count of the entities of 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 data center. Ring colors indicate how critical the actions are.
 - Green indicates the percentage of entities that have no actions pending.
 - Red indicates the percentage of entities that have performance-related recommendations.
 - Yellow indicates efficiency and compliance recommendations.
- 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 virtual machines are running in a virtual data center, the entry will have another arrow to that as well. This means that your virtual machines consume resources from hosts, storage, and possibly virtual data centers.

To see actual counts of pending actions, hover on a ring to view more details. In Figure 10, you can see that you have 23 storage devices, of which 17 percent have performance-related recommendations.

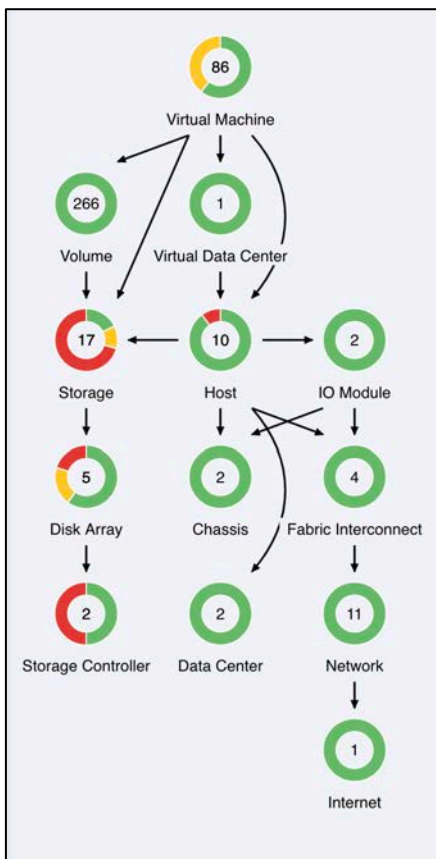


Figure 10.
Supply chain focused on storage entities

You can click an entity tier in the supply chain to see a list of those entities with specific details pertaining to that entity. For example, click Storage to see the details of the storage devices in your environment, such as pending actions, health, top storage devices, capacity, and utilization (Figure 11).

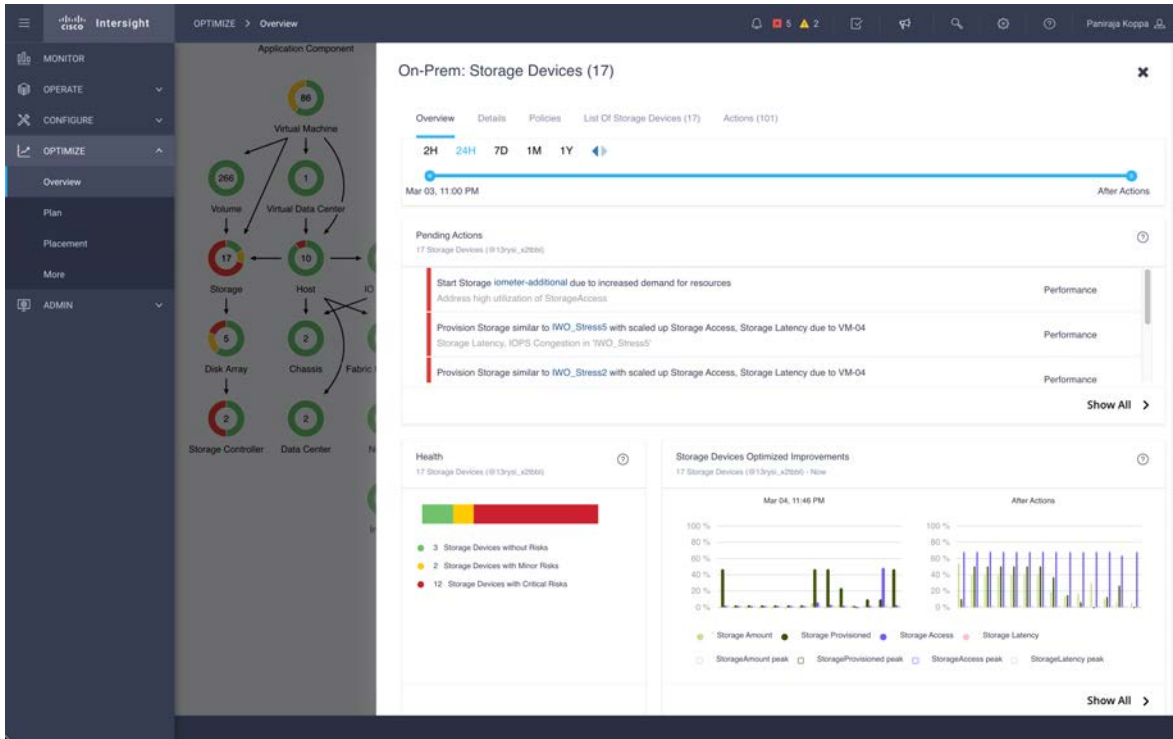


Figure 11.
Details of an on-premises storage device

You can navigate between various tabs to see an overview, details (such as storage amount and storage provisioned), policies, and a list of storage devices and all the actions that you can perform on them (Figure 12 and Figure 13).

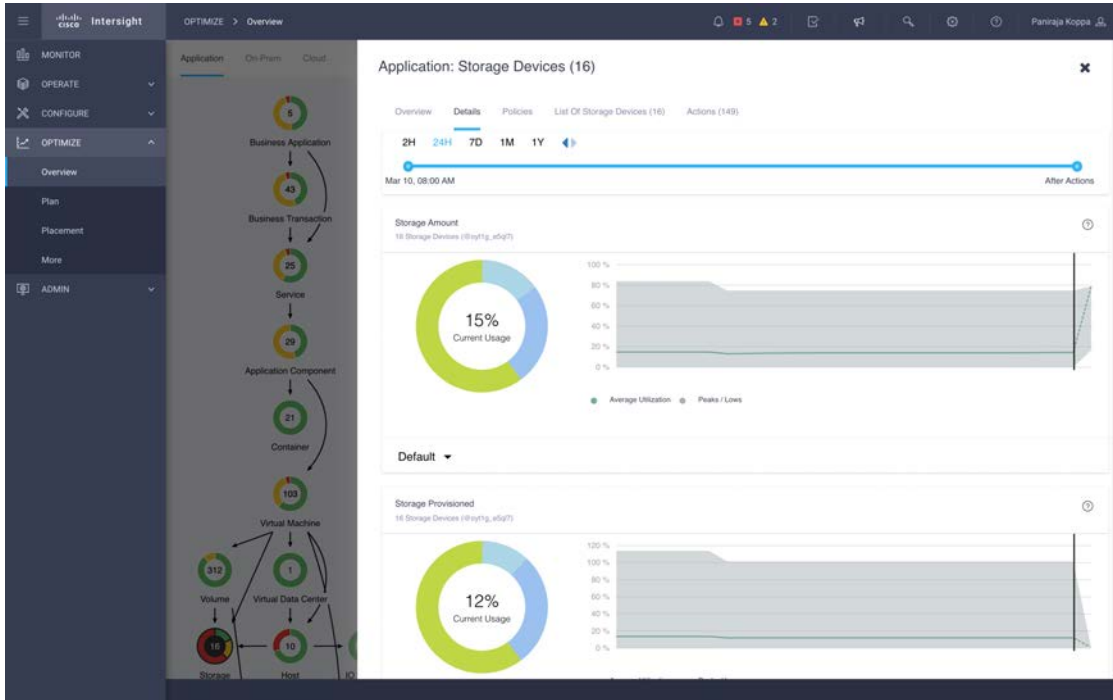


Figure 12.
Details of a storage device

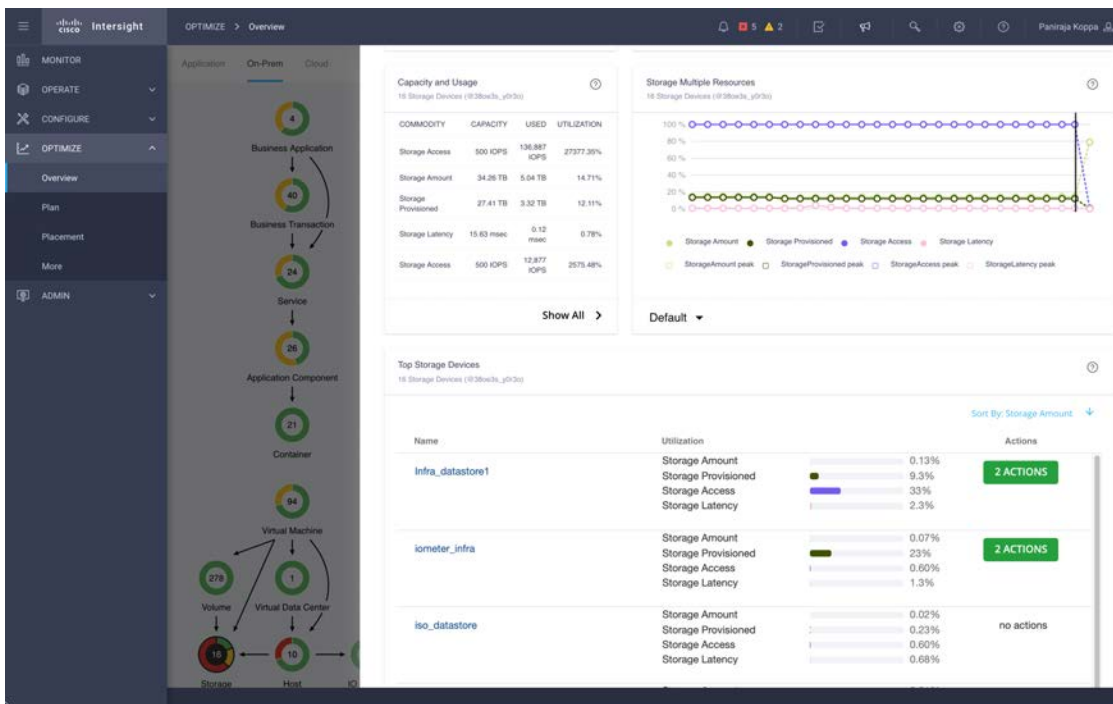


Figure 13.
Details showing storage widgets

Similar operations are possible for other FlashStack entities such as virtual machines, volumes, storage controllers, and other compute elements.

Scoping

You can scope to one particular entity and view the entire supply chain and other details only for that entity. This scoped view allows you to zoom in on an entity in your environment and get a quick picture of the system health, actions, relationship with other entities, and all other details for that scope.

The example in Figure 14 shows scoping to a storage entity, by choosing Storage > List of Storage Devices and clicking one of the storage entities.

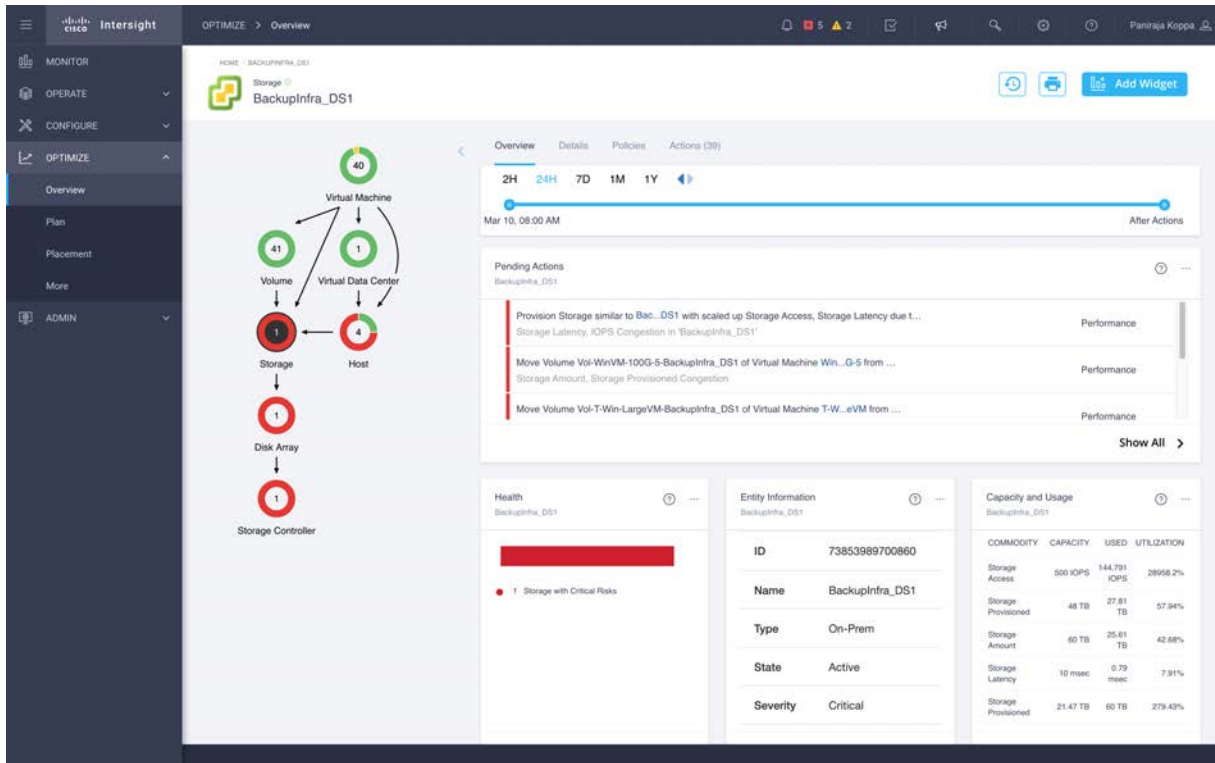


Figure 14.
Scoping to a storage entity in Flash Array

From this supply chain scoped to a storage volume, you can see to which disk array this volume belongs, which filer owns it, which virtual machines are making use of the datastore from this volume, etc.

Pending actions, health, capacity, utilization, etc. are specific to this particular storage volume.

Recommendations

Cisco Intersight Workload Optimizer makes multiple recommendations about how to make workloads perform best at the best possible cost while maintaining compliance. This section presents some examples of recommendations for virtual machines, hosts, and storage.

Recommendations for virtual machines

Cisco Intersight Workload Optimizer discovers virtual machines through added VMware targets. If utilization is high enough, Workload Optimizer can allocate more resources to the virtual machine, provision another instance, or move the virtual machine to a host that has more resources. If utilization falls off, Workload Optimizer can post recommendations to reduce the amount of resources.

The following sections present some examples of recommendations that Cisco Intersight Workload Optimizer provides for virtual machines.

Recommendations to ensure virtual machine compliance

When calculating workload placement, Cisco Intersight Workload Optimizer respects cluster boundaries, networks, and provisioned datastores. These boundaries impose segments on the market view that Workload Optimizer uses to model virtual infrastructure. The example in Figure 15 shows a virtual machine that doesn't comply with a policy set in a node. (A group of virtual machines must run on a storage device.)

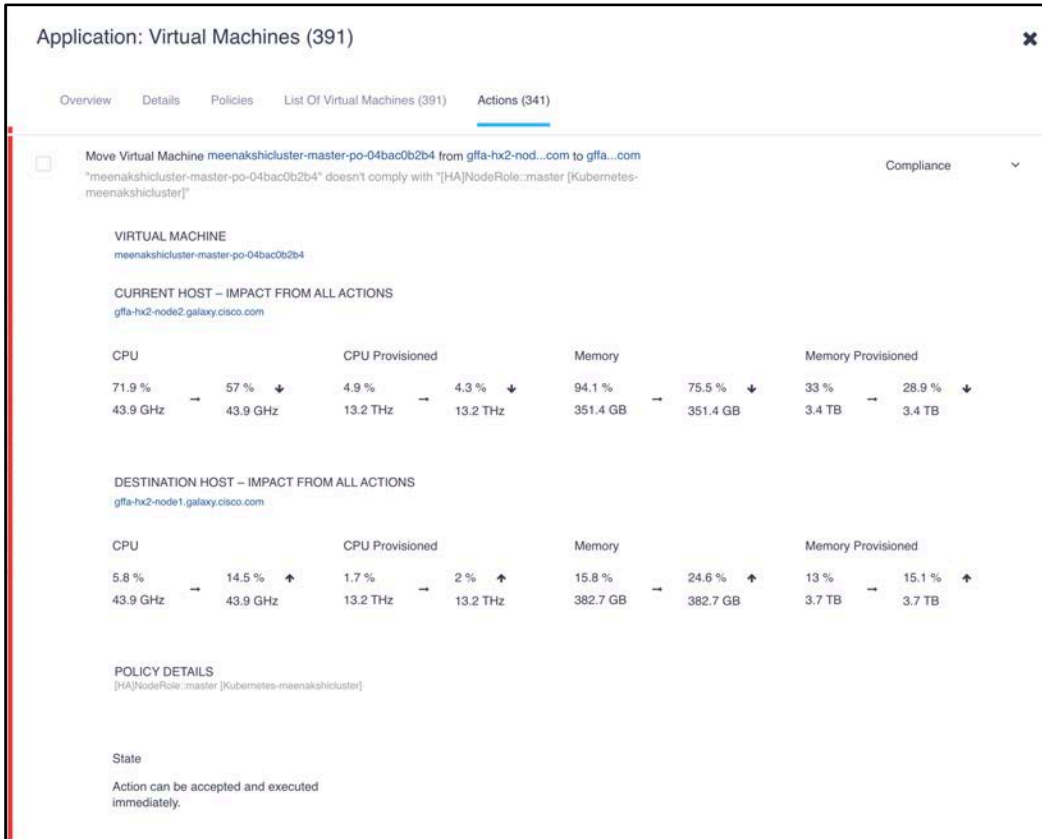


Figure 15. Recommending a compliance action for a virtual machine

Recommendations to improve overall performance of virtual machines

Cisco Intersight Workload Optimizer monitors the utilization of the resources allocated to the hosting virtual machine (such as vCPU and vMem) and recommends provisioning additional resources or reducing the amount of resources.

In the example in Figure 16, CPU utilization is high in a virtual machine, and hence Workload Optimizer recommends resizing the vCPU up.

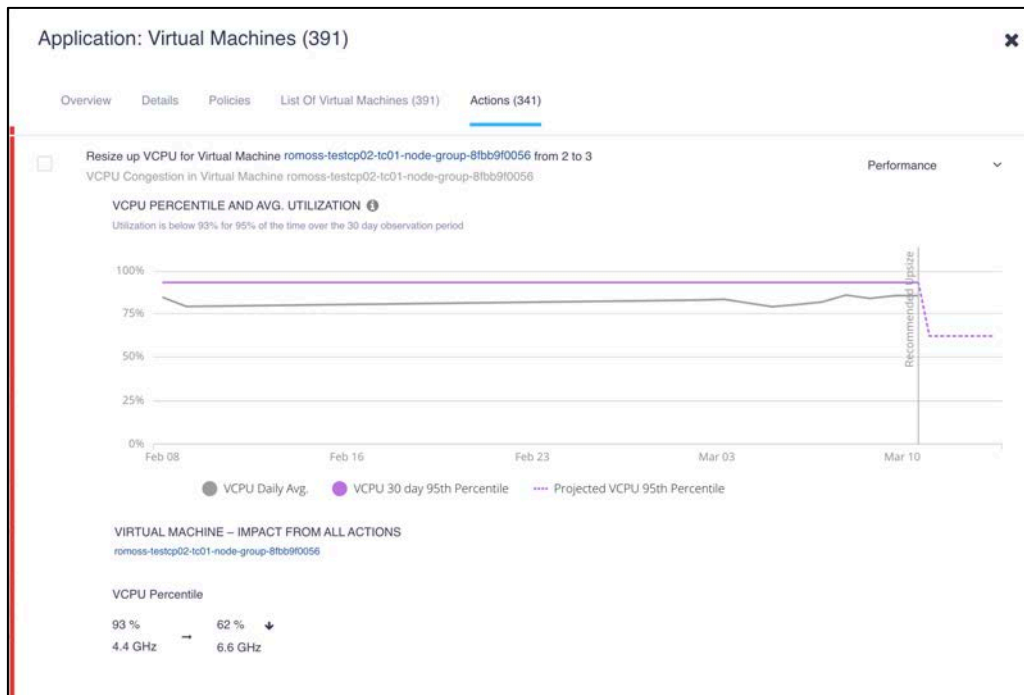


Figure 16.
vCPU resize up recommendation for virtual machine performance

Recommendations to improve infrastructure utilization

The example in Figure 17 shows an underutilized vCPU in a virtual machine. Resize the vCPU down to increase the percentage utilization of vCPU cycles that are devoted to processing instruction.

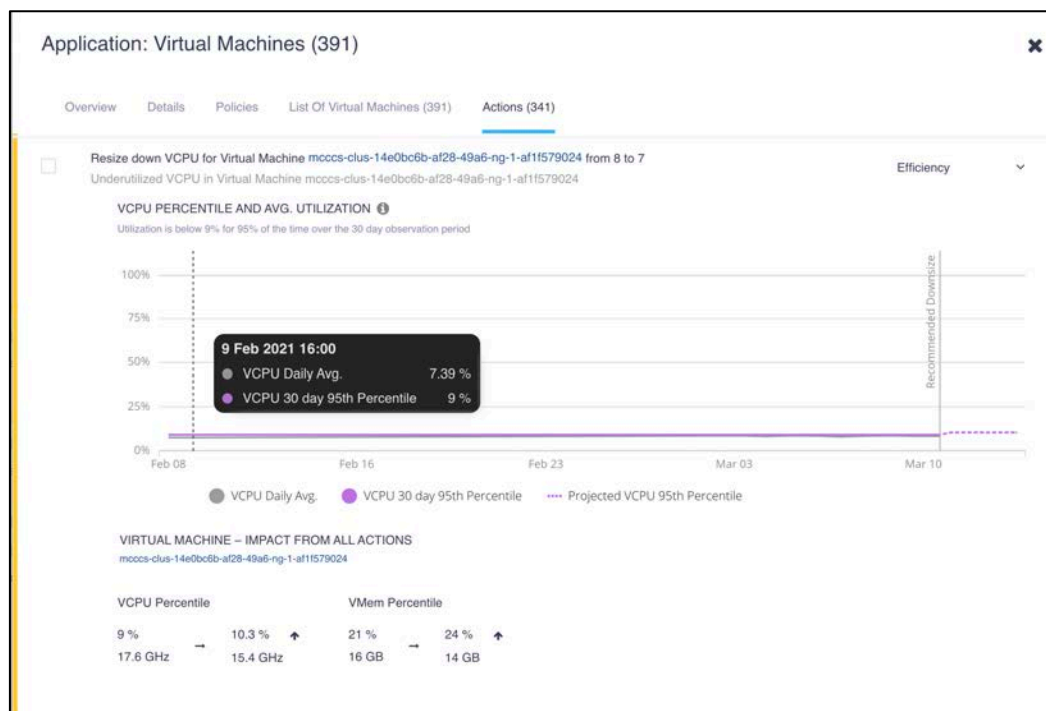


Figure 17.
vCPU resize down recommendation

Cisco Intersight Workload Optimizer monitors the utilization of the vMem allocated to the hosting virtual machine and recommends provisioning additional memory or reduce the amount of memory.

In the example in Figure 18, vMem utilization is below 13 percent for 95 percent of the time over the 30-day observation period, so Workload Optimizer recommends resizing vMem to increase the utilization percentage.

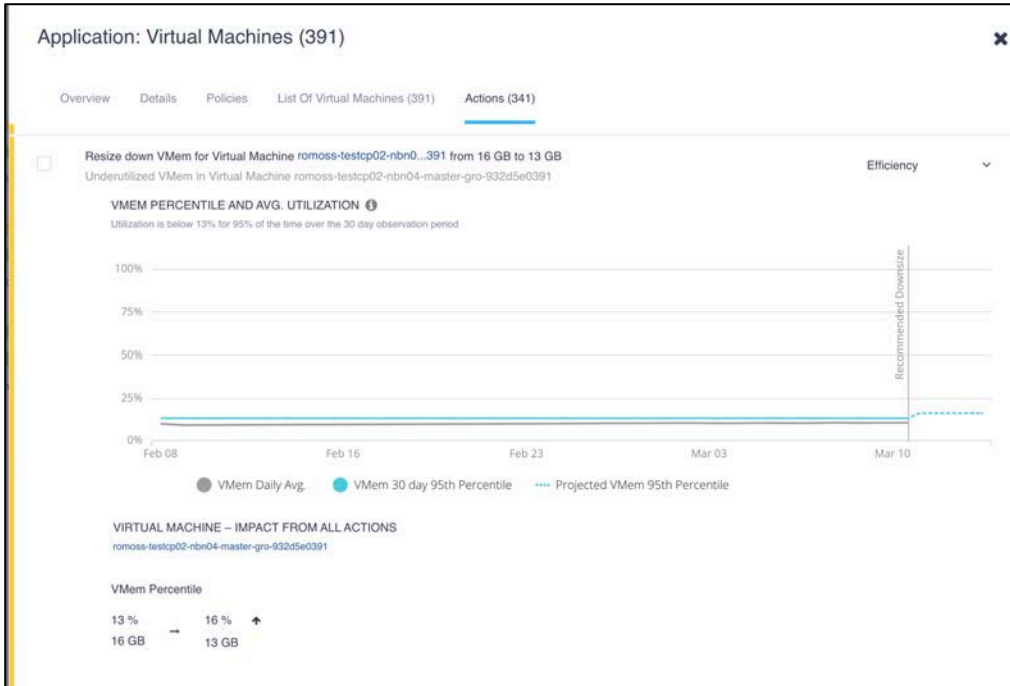


Figure 18.
vMem recommendations for efficiency

Recommendations to move virtual machines from a host for performance

Cisco Intersight Workload Optimizer recommends moving virtual machines from one host to another so that overall performance of virtual machines is increased.

The example in Figure 19 shows that memory congestion is occurring in the source host, and that by moving a virtual machine, it can be avoided.

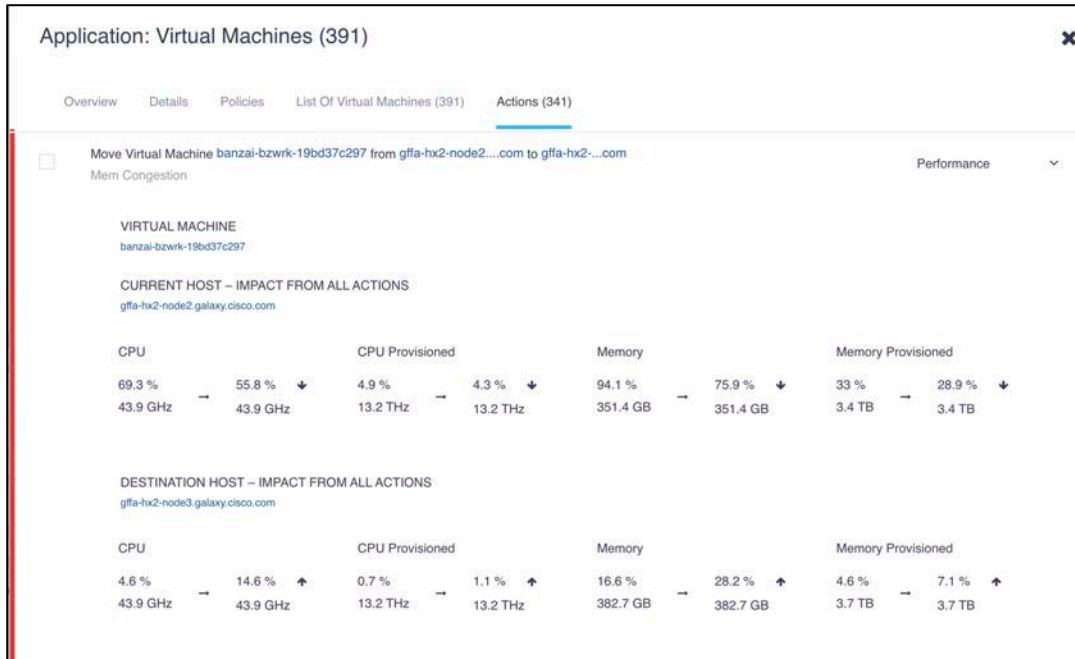


Figure 19.
Move recommendations for virtual machines for performance

Figure 20 shows a move virtual machine recommendation in response to CPU congestions in the host.

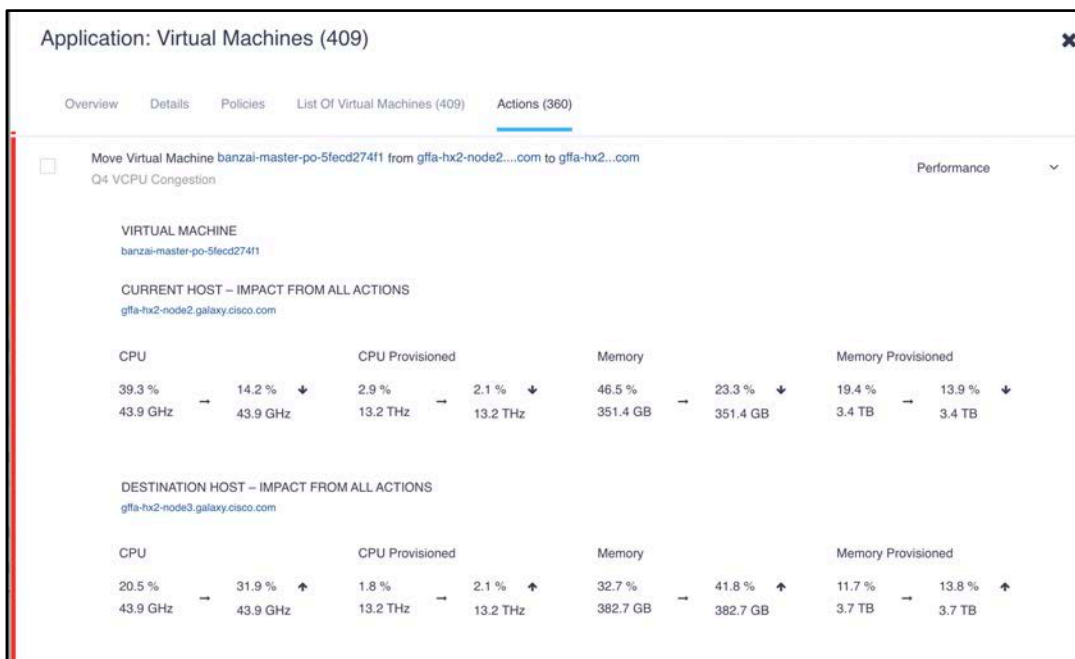


Figure 20.
Move recommendations due to CPU congestion in host

Recommendations to move virtual machines from a host for efficiency

Cisco Intersight Workload Optimizer recommends moving virtual machines from a host so that the host can be suspended without any impact on the performance of the virtual machines.

In the example in Figure 21, Workload Optimizer recommends moving virtual machines from one host to two other hosts to increase infrastructure efficiency. Figure 22 shows the results.

<input type="checkbox"/>	Move Virtual Machine TEST-LNX from gffa-ucsesx-1.galaxy.cisco.com to gffa-ucsesx-4.galaxy.cisco.com gffa-ucsesx-1.galaxy.cisco.com can be suspended to improve efficiency	Efficiency	>
<input type="checkbox"/>	Move Virtual Machine MB-Centos7 from gffa-esx-5.galaxy.cisco.com to gffa-esx-3.galaxy.cisco.com gffa-esx-5.galaxy.cisco.com can be suspended to improve efficiency	Efficiency	>
<input type="checkbox"/>	Move Virtual Machine ucs-ccp-tenant-image-1.16.3-ubuntu18-6.1.0 from gffa-esx-5...com to gffa-...com gffa-esx-5.galaxy.cisco.com can be suspended to improve efficiency	Efficiency	>
<input type="checkbox"/>	Move Virtual Machine GFFA-VPN from gffa-esx-5.galaxy.cisco.com to gffa-esx-3.galaxy.cisco.com gffa-esx-5.galaxy.cisco.com can be suspended to improve efficiency	Efficiency	>
<input type="checkbox"/>	Move Virtual Machine vmware-cloudcenter-centos7-worker from gffa-esx-5.g...com to gffa-es...com gffa-esx-5.galaxy.cisco.com can be suspended to improve efficiency	Efficiency	>
<input type="checkbox"/>	Move Virtual Machine ccs-ucs-n-52556dac-c290...1-0 from gffa-ucs...com to gff...com gffa-ucsesx-1.galaxy.cisco.com can be suspended to improve efficiency	Efficiency	>
<input type="checkbox"/>	Move Virtual Machine ccs5-vmware-remote-5.1.2-new from gffa-ucsesx-1....com to gffa-ucs...com gffa-ucsesx-1.galaxy.cisco.com can be suspended to improve efficiency	Efficiency	>

Figure 21.

Move recommendations for virtual machines from one host to multiple other hosts for efficiency

Application: Virtual Machines (391)										
Overview Details Policies List Of Virtual Machines (391) Actions (341)										
<input type="checkbox"/>	Move Virtual Machine MB-Centos7 from gffa-esx-5.galaxy.cisco.com to gffa-esx-3.galaxy.cisco.com gffa-esx-5.galaxy.cisco.com can be suspended to improve efficiency								Efficiency	▼
VIRTUAL MACHINE MB-Centos7										
CURRENT HOST – IMPACT FROM ALL ACTIONS gffa-esx-5.galaxy.cisco.com										
CPU		CPU Provisioned		Memory		Memory Provisioned				
0.3 %	→ 0.2 % ↓	0.2 %	→ 0 % ↓	8.6 %	→ 2.6 % ↓	2.4 %	→ 0 % ↓			
38.3 GHz	→ 38.3 GHz	11.5 THz	→ 11.5 THz	127.7 GB	→ 127.7 GB	1.2 TB	→ 1.2 TB			
DESTINATION HOST – IMPACT FROM ALL ACTIONS gffa-esx-3.galaxy.cisco.com										
CPU		CPU Provisioned		Memory		Memory Provisioned				
7.6 %	→ 7.5 % ↓	0.6 %	→ 0.7 % ↑	16 %	→ 17.4 % ↑	3.5 %	→ 4.1 % ↑			
72.6 GHz	→ 72.6 GHz	21.8 THz	→ 21.8 THz	511.9 GB	→ 511.9 GB	5 TB	→ 5 TB			

Figure 22.

Impact of virtual machine move

Recommendations for virtual machines in a storage context

Cisco Intersight Workload Optimizer discovers on-premises datastores through VMware targets and storage controllers. A disk array provides storage volumes to serve the storage requirements of physical machines. It uses the resources of one storage controller, which 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.

Recommendations to delete idle files

Figure 23 shows Cisco Intersight Workload Optimizer monitoring the idle and unused files in the datastore and recommending that such files be deleted. These files could be Virtual Machine Disk (VMDK) files for a deleted virtual machine, dump files, and other temporary files.

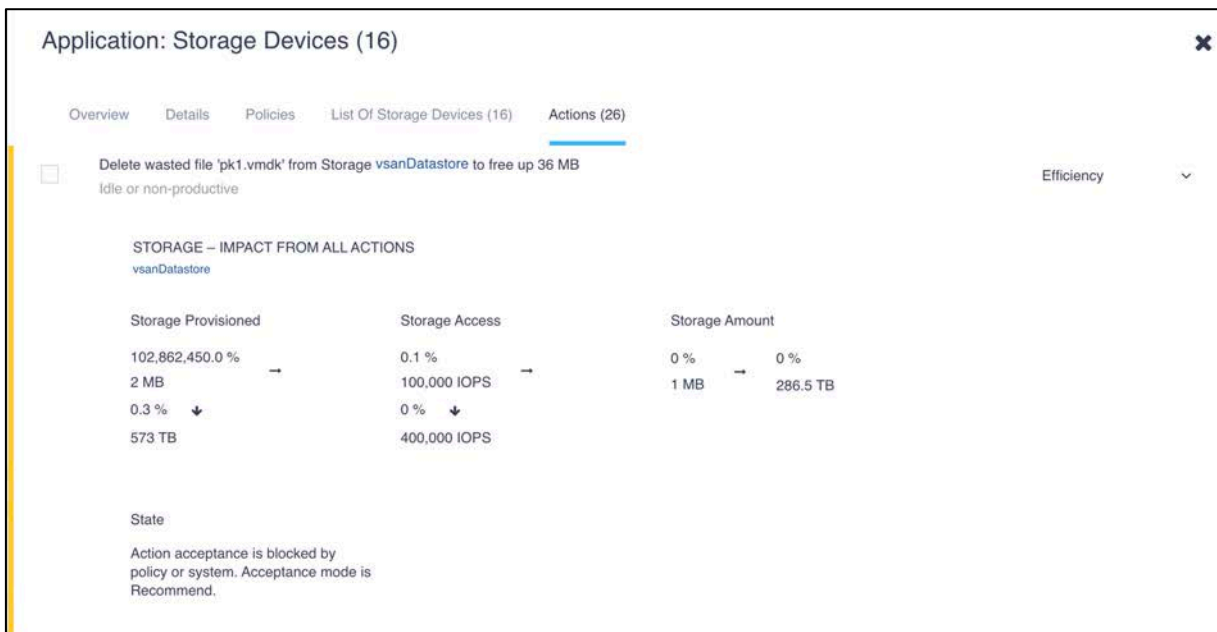


Figure 23.

Recommendation to delete unused files

Recommendations to move volumes to prevent congestion

Workload Optimizer can allocate more resources to a virtual machine, provision another instance, or move the virtual machine to a host that has more resources if storage utilization is high. If utilization falls off, Workload Optimizer can post recommendations to reduce the amount of resources.

The example in Figure 24 shows several recommendations to move a virtual machine's volume from one storage entity to another. Workload Optimizer makes these recommendations to avoid IOPS congestion, latency congestion, and congestion related to the amount of storage provisioned.

Application: Volumes (312) ✕

Overview Policies List Of Volumes (312) Actions (103)

<input type="checkbox"/>	Move Volume Vol-SNP_C_WinVM-100G-9-BackupInfra_DS2 of Virtual Machine SNP_...G-9 from ... IOPS Congestion	Performance	>
<input type="checkbox"/>	Move Volume Vol-2-Lin-10-BackupInfra_DS2 of Virtual Machine 2-Li...-10 from Backup...DS2 to ... IOPS Congestion	Performance	>
<input type="checkbox"/>	Move Volume Vol-2-Lin-9-BackupInfra_DS2 of Virtual Machine 2-Lin-9 from Backup...DS2 to Veea...S1M IOPS Congestion	Performance	>
<input type="checkbox"/>	Move Volume Vol-2-Lin-6-BackupInfra_DS2 of Virtual Machine 2-Lin-6 from Backup...DS2 to Veea...S1M IOPS Congestion	Performance	>
<input type="checkbox"/>	Move Volume Vol-2-Lin-5-BackupInfra_DS2 of Virtual Machine 2-Lin-5 from Backup...DS2 to Veea...S1M IOPS Congestion	Performance	>
<input type="checkbox"/>	Move Volume Vol-1-Lnx-1-BackupInfra_DS1 of Virtual Machine 1-Lnx-1 from Backup...DS1 to Vee...S1M Storage Amount, Storage Provisioned Congestion	Performance	>
<input type="checkbox"/>	Move Volume Vol-WinVM-100G-5-BackupInfra_DS1 of Virtual Machine Win...G-5 from Bac...DS1 to ... Storage Amount, Storage Provisioned Congestion	Performance	>
<input type="checkbox"/>	Move Volume Vol-WinVM-100G-2-BackupInfra_DS2 of Virtual Machine Win...G-2 from Bac...DS2 to ... IOPS Congestion	Performance	>
<input type="checkbox"/>	Move Volume Vol-Pani-windows-controller-iometer_infra of Virtual Machine Pani...ler from iom...fra to ... IOPS, Storage Latency Congestion	Performance	>
<input type="checkbox"/>	Move Volume Vol-T-WinLarge-FLC1-BackupInfra_DS1 of Virtual Machine T-W...LC1 from Bac...DS1 to ... Storage Amount, Storage Provisioned Congestion	Performance	>
<input type="checkbox"/>	Move Volume Vol-intersight-assist-Infra_datastore1 of Virtual Machine inters...ist from Inf...re1 to ... Storage Provisioned, IOPS Congestion	Performance	>
<input type="checkbox"/>	Move Volume Vol-T1-WinVM18-BackupInfra_DS1 of Virtual Machine T1-...M18 from Bac...DS1 to ... Storage Amount, Storage Provisioned Congestion	Performance	>

Figure 24.
Recommendations to avoid IOPS congestion

The example in Figure 25 shows a recommendation to move a virtual machine from one datastore to another because there is congestion on the storage provisioned.

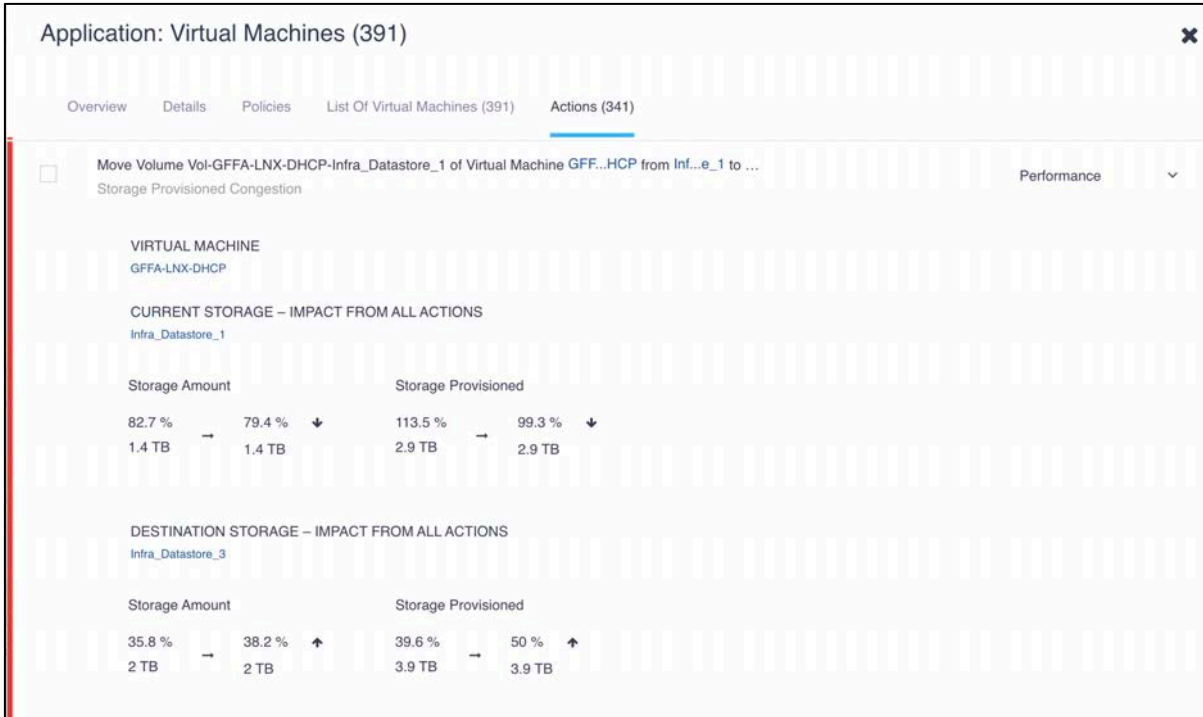


Figure 25.
Recommendations due to congestion on storage provisioned

Recommendations to move volumes to improve infrastructure efficiency

Figure 26 shows recommendations to move a virtual machine's volume to a different storage entity so that a volume can be suspended.

<input type="checkbox"/>	Move Volume Vol-VM-16-iometer-additional of Virtual Machine VM-16 from iometer...nal to Veea...S1M iometer-additional can be suspended to improve efficiency	Efficiency	>
<input type="checkbox"/>	Move Volume Vol-VM-15-iometer-additional of Virtual Machine VM-15 from iometer...nal to Veea...S1M iometer-additional can be suspended to improve efficiency	Efficiency	>
<input type="checkbox"/>	Move Volume Vol-VM-11-iometer-additional of Virtual Machine VM-11 from iometer...nal to Veea...S1M iometer-additional can be suspended to improve efficiency	Efficiency	>
<input type="checkbox"/>	Move Volume Vol-VM-08-iometer-additional of Virtual Machine VM-08 from iometer...nal to Veea...S1M iometer-additional can be suspended to improve efficiency	Efficiency	>
<input type="checkbox"/>	Move Volume Vol-VM-06-iometer-additional of Virtual Machine VM-06 from iometer...nal to Veea...S1M iometer-additional can be suspended to improve efficiency	Efficiency	>
<input type="checkbox"/>	Move Volume Vol-VM-03-iometer-additional of Virtual Machine VM-03 from iometer...nal to Veea...S1M iometer-additional can be suspended to improve efficiency	Efficiency	>
<input type="checkbox"/>	Move Volume Vol-VM-12-iometer-additional of Virtual Machine VM-12 from iometer...nal to Veea...S1M iometer-additional can be suspended to improve efficiency	Efficiency	>

Figure 26.
Recommendations to move a virtual machine's volume to a different datastore

Recommendations for hosts

A host is a server that runs processes, including hypervisor processes to host virtual workloads. Cisco Intersight Workload Optimizer monitors the utilization of memory, CPU, I/O, network, and swap file resources and ballooning and recommends actions such as suspend a host, start a host, and provision a new host.

In the example in Figure 27, Workload Optimizer recommends provisioning a new host to prevent throughput congestion.

The screenshot displays the 'Application: Hosts (10)' interface. The 'Actions (7)' tab is active, showing a list of recommendations. The first recommendation is 'Provision Host similar to vm-infra-esxi-04.flashstack.com' with the reason 'Throughput Congestion'. Below this, a detailed view for the 'CURRENT HOST - IMPACT FROM ALL ACTIONS' for 'vm-infra-esxi-04.flashstack.com' is shown, indicating 'No Impact' and that 'Action acceptance is blocked by policy or system. Acceptance mode is Recommend.' Below this, five more recommendations are listed, all for 'Provision Host similar to vm-infra-esxi-03.flashstack.com' with the reason 'Throughput Congestion'. Each recommendation has a 'Performance' dropdown menu.

Figure 27.
Recommendations to provision a new host to avoid throughput congestion

Figure 28 shows a similar recommendation to provision a new host to prevent memory congestion.

The screenshot displays the 'Application: Hosts (27)' interface. The 'Actions (41)' tab is active, showing a recommendation for 'Provision Host similar to gffa-hx1-node2.galaxy.cisco.com' with the reason 'Mem. Congestion'. Below this, a detailed view for the 'CURRENT HOST - IMPACT FROM ALL ACTIONS' for 'gffa-hx1-node2.galaxy.cisco.com' is shown. It includes a table comparing current host metrics with 'Memory Provisioned' metrics:

Memory	Memory Provisioned	CPU
86.6 %	10.1 %	54.5 %
351.7 GB	3.4 TB	43.9 GHz
84.9 % ↓	9.8 % ↓	53.7 % ↓
351.7 GB	3.4 TB	43.9 GHz

Below the table, the 'State' is indicated as 'Action acceptance is blocked by policy or system. Acceptance mode is Recommend.'

Figure 28.
Recommendations to provision a new host to avoid memory congestion

Working with policies

Policies set business rules to control the way that Cisco Intersight Workload Optimizer analyzes resource allocation, displays resource status, and recommends or implements actions (Figure 29).

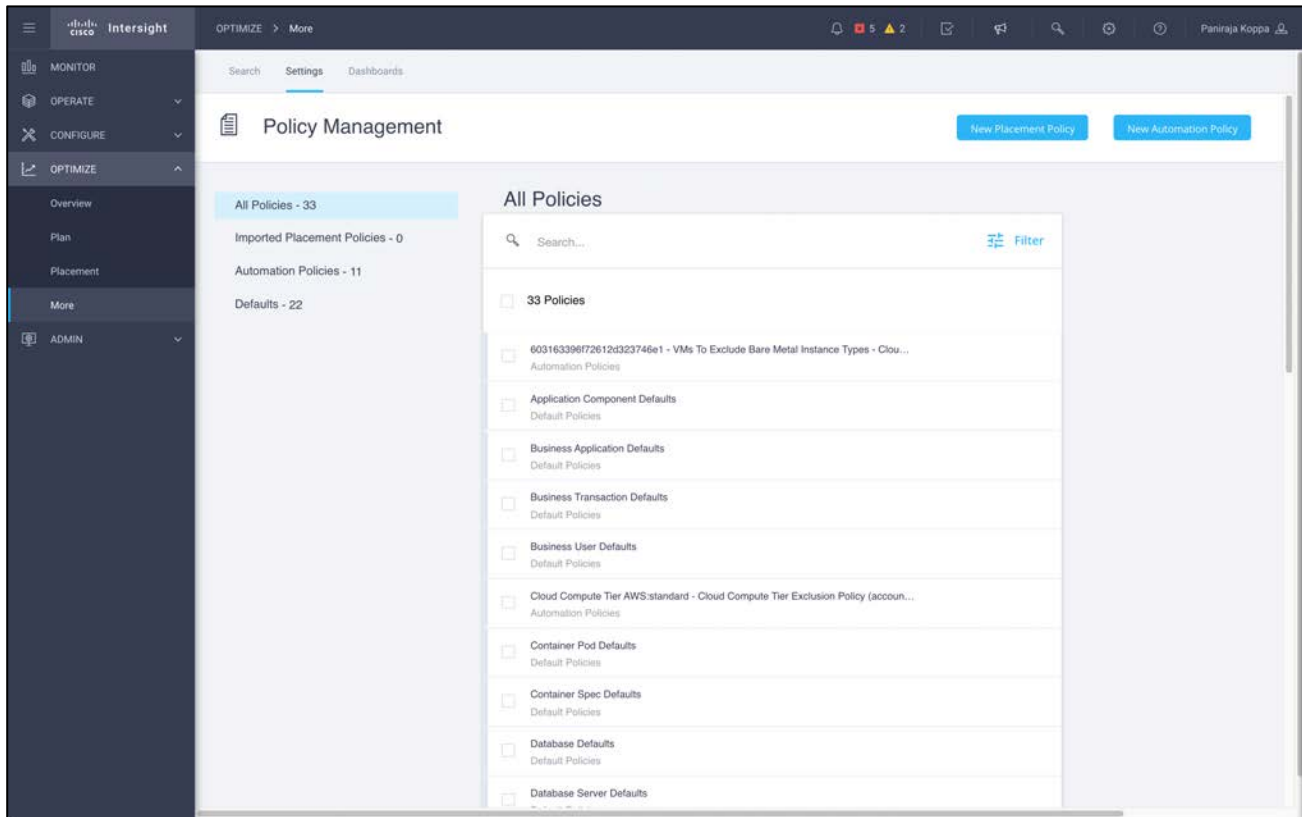


Figure 29.
Policy management

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

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

Cisco Intersight Workload Optimizer includes two fundamental types of policies: placement policies and automation policies

Placement policies

For optimization, Cisco Intersight Workload Optimizer recommends actions to place workloads for certain objects such as virtual machines on a disk array, host, etc. Workload Optimizer can recommend these actions or implement them automatically.

When calculating workload placement, Workload Optimizer respects cluster boundaries, networks, and provisioned datastores. In addition, the configuration of your environment can specify logical boundaries, and within Workload Optimizer you can create even more boundaries. These boundaries impose segments on the market that Workload Optimizer uses to model your application infrastructure. Workload Optimizer discovers placement rules that are defined by the targets in your environment, and you can create your own segments.

Cisco Intersight Workload Optimizer supports the following placement policies:

- Place: Determine which entities use specific providers.
- Don't place: Consumers must never run on specific providers.
- Merge: Merge clusters into a single provider group.
- License: Set up hosts with paid licenses to be the preferred providers for virtual machines or applications that require those licenses.

Figure 30 shows a placement policy that explicitly specifies not to place a group of virtual machines on certain hosts.

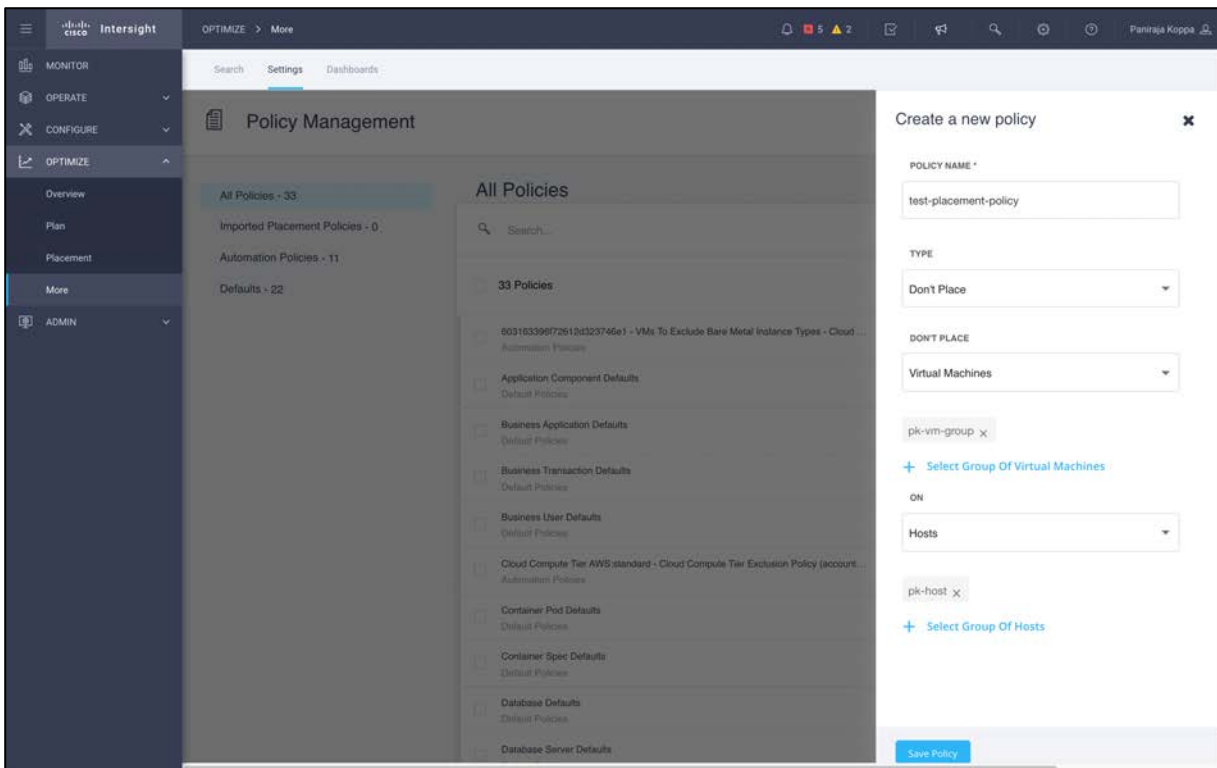


Figure 30.
Placement policy

Automation policies

To help you avoid problems in your environment, Cisco Intersight Workload Optimizer analysis identifies actions that you can implement to keep your system running in an optimal way. You can specify the degree of automation you want for these specific actions. For example, in some environments you may not want to automate the resizing down of virtual machines because that is a disruptive action. You can use action modes in a policy to set that business rule.

Cisco Intersight Workload Optimizer supports the following action modes:

- **Recommend:** Recommend the action so a user can implement it through the given hypervisor or by other means.
- **Manual:** Recommend the action and provide the option to implement that action through the Cisco Intersight Workload Optimizer user interface.
- **Automated:** Implement the action automatically.

Cisco Intersight Workload Optimizer ships with default automation policy settings for the various types of entities that it can discover (Figure 31). The settings for these default policies should be adequate to meet initial business requirements. These policies apply to the global scope. You can edit these settings based on your requirements.

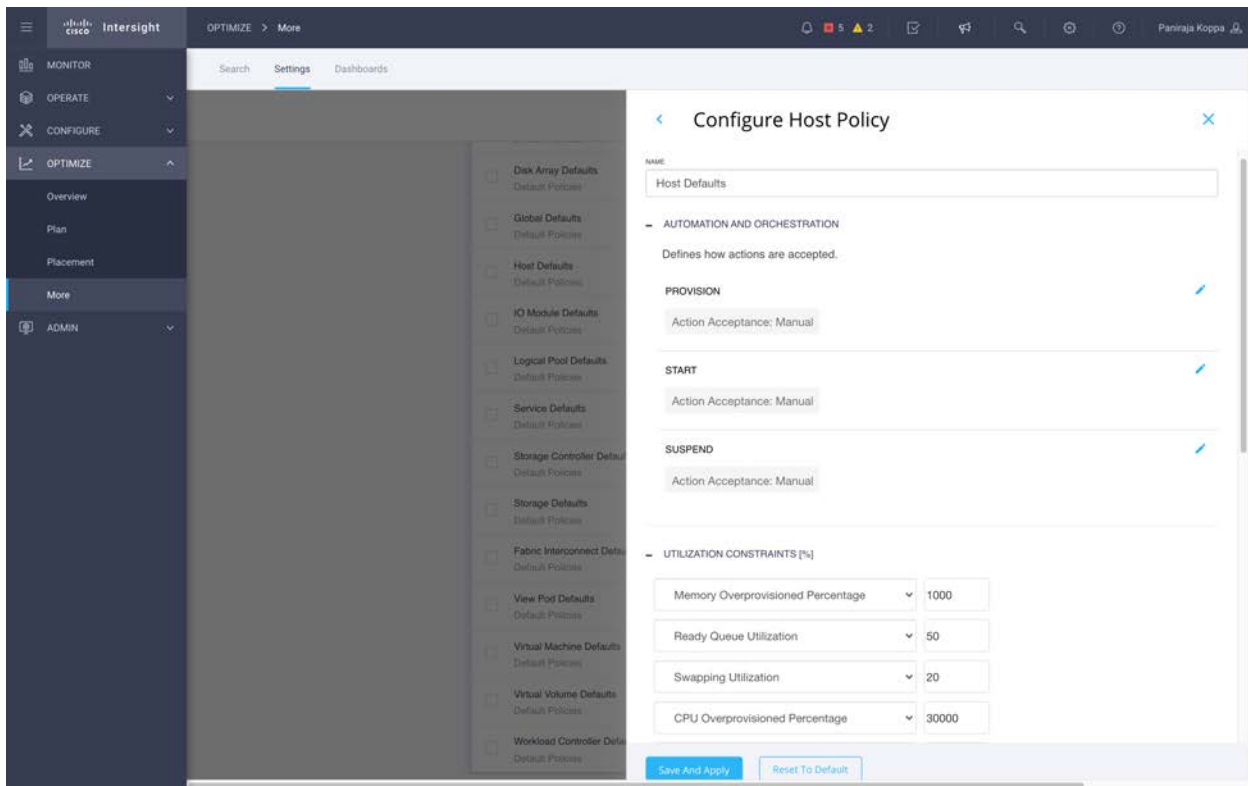


Figure 31.
Default automation policy for a compute host

If you need to change these settings for a subset of entities and not for the global scope (for example, you may want to change action automation for a specific scope or change the utilization constraints), you can override the current default automation policies and create scoped policies. In addition, you can assign a schedule to a scoped policy to set up maintenance windows or other scheduled actions in your environment.

Figure 32 shows an automation policy created for virtual machines and applied on a subset of virtual machines to enable vCPU resize up and down and vMem resize up and down actions to occur automatically. The policy also changes the vCPU utilization constraint from the default setting of 70 to 60.

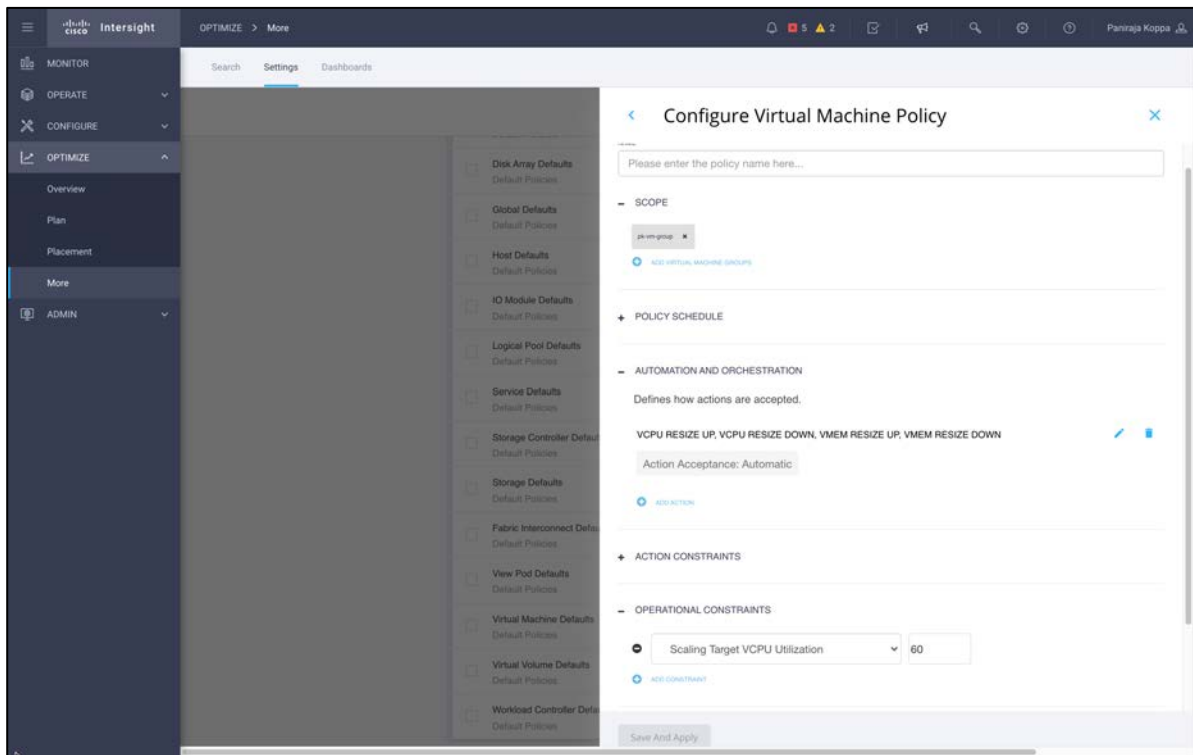


Figure 32.
Sample automation policy for storage

Conclusion

With Cisco Intersight Workload Optimizer and FlashStack, IT operations teams can easily perform and automate application resource management at scale and ensure performance of I/O-intensive applications. This solution provides full-stack visibility of applications and the underlying infrastructure, intelligent insights, and automated actions. The solution also reduces costs and preserves capital by effectively increasing the resource utilization of the FlashStack environment.

For more information

For additional information, see the following resources:

- FlashStack solution: <https://www.flashstack.com/>
- Cisco Intersight Workload Optimizer data sheet: <https://www.cisco.com/c/en/us/products/collateral/cloud-systems-management/intersight-workload-optimizer/datasheet-c78-744509.html>
- Information about Cisco Intersight Workload Optimizer: https://www.cisco.com/c/en_in/products/cloud-systems-management/intersight-workload-optimizer/index.html

-
- Cisco Intersight data sheet and ordering information:
<https://www.cisco.com/c/en/us/products/collateral/servers-unified-computing/intersight/datasheet-c78-739433.html>
 - Cisco Intersight Workload Optimizer online help:
<https://intersight.com/help/resources-cisco-intersight-workload-optimizer>
 - Cisco Intersight Workload Optimizer Getting Started Guide:
<https://intersight.com/help/resources/cisco-intersight-workload-optimizer-getting-started>
 - Cisco Intersight Workload Optimizer User Guide:
https://cdn.intersight.com/components/an-hulk/1.0.9-685/docs/cloud/data/resources/iwo/Cisco_Intersight_Workload_Optimizer_User_Guide.pdf
 - Cisco Intersight Workload Optimizer Target Configuration Guide:
https://cdn.intersight.com/components/an-hulk/1.0.9-685/docs/cloud/data/resources/iwo/Cisco_Intersight_Workload_Optimizer_Target_Configuration_Guide.pdf

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at <https://www.cisco.com/go/offices>.

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: <https://www.cisco.com/go/trademarks>. Third-party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)