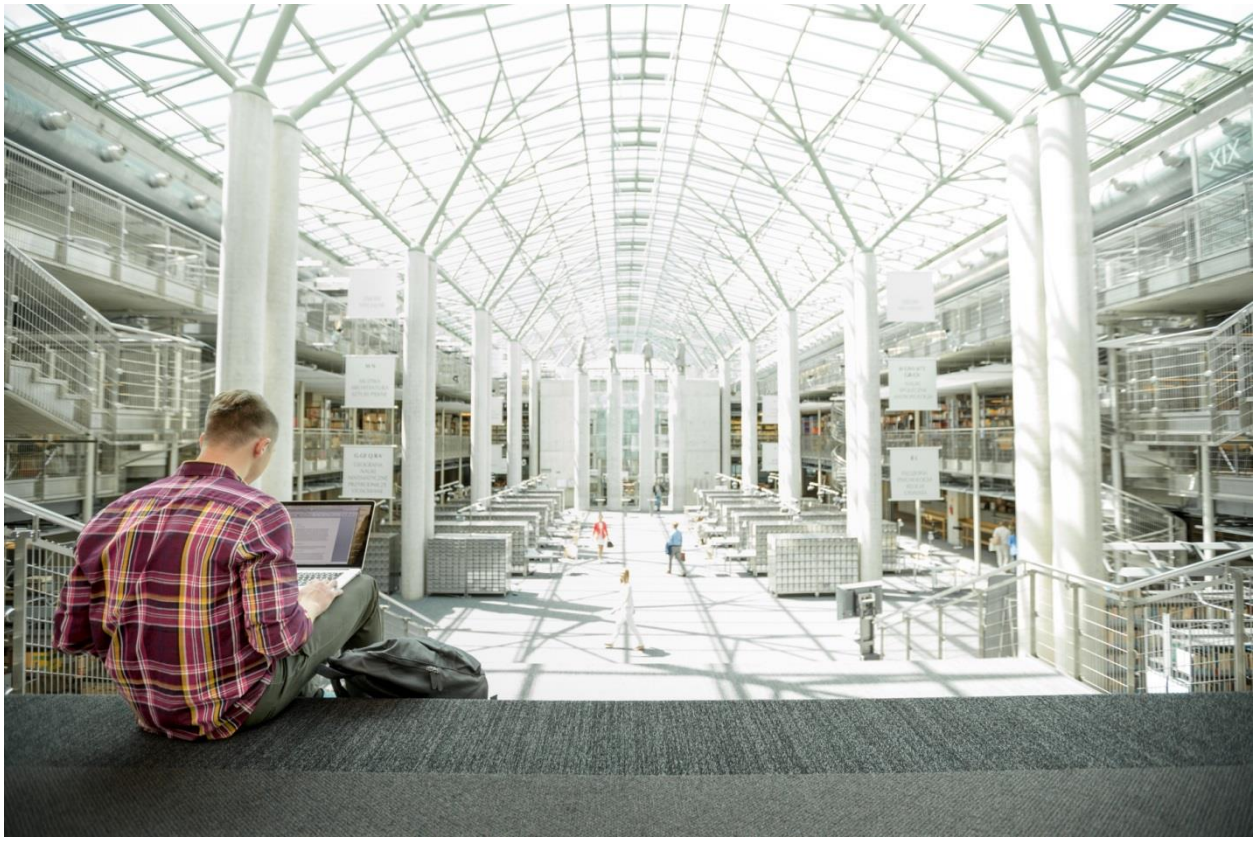


Cisco HyperFlex 3.0 for Virtual Server Infrastructure with VMware ESXi

Deployment Guide for Cisco HyperFlex 3.0 for Virtual Server Infrastructure with VMware ESXi Hypervisor

Last Updated: December 20, 2018

About the Cisco Validated Design Program

The Cisco Validated Design (CVD) program consists of systems and solutions designed, tested, and documented to facilitate faster, more reliable, and more predictable customer deployments. For more information visit:

http://www.cisco.com/go/designzone.

Table of Contents

Executive Summary

With the proliferation of virtualized environments across most IT landscapes, other technology stacks which have traditionally not offered the same levels of simplicity, flexibility, and rapid deployment as virtualized compute platforms have come under increasing scrutiny. In particular, networking devices and storage systems have lacked the agility of hypervisors and virtual servers. With the introduction of Cisco HyperFlex, Cisco has brought the dramatic enhancements of hyperconvergence to the modern datacenter. Cisco HyperFlex systems are based on the Cisco UCS platform, combining Cisco HX-Series x86 servers and integrated networking technologies through the Cisco UCS Fabric Interconnects, into a single management domain, along with industry leading virtualization hypervisor software from VMware, and next-generation software defined storage technology. The combination creates a complete virtualization platform, which provides the network connectivity for the guest virtual machine (VM) connections, and the distributed storage to house the VMs, spread across all of the Cisco UCS x86 servers, versus using specialized storage or networking components. The unique storage features of the HyperFlex log based filesystem enable rapid cloning of VMs, snapshots without the traditional performance penalties, and data deduplication and compression. All configuration, deployment, management, and monitoring of the solution can be done with existing tools for Cisco UCS and VMware, such as Cisco UCS Manager and VMware vCenter, and new integrated HTML based management tools, such as Cisco HyperFlex Connect and Cisco Intersight. This powerful linking of advanced technology stacks into a single, simple, rapidly deployed solution makes Cisco HyperFlex a true second generation hyperconverged platform.

Cisco HyperFlex HXDP 3.0 expands on the existing suite of enterprise class data protection features by introducing additional deployment options to enhance the availability and protection of data stored in the HyperFlex cluster. Stretched clusters provide an installation option which splits the physical location of the HyperFlex cluster nodes across two sites, providing for continued operation even if an entire site fails. Logical availability zones divide the HyperFlex cluster nodes into logical subdivisions, in order to allow data to be spread across the nodes in a more controlled manner. This enhanced distribution of data provides greater resiliency to node failures in larger HyperFlex clusters. Customers can choose to deploy SSD-only All-Flash HyperFlex clusters for improved performance, increased density, and reduced latency, or use HyperFlex hybrid clusters which combine high-performance SSDs and low-cost, high-capacity HDDs to optimize the cost of storing data. Further enhancements include improvements and customization capabilities in the HyperFlex Connect management tool, larger scale 64-node clusters, large form-factor disks for larger storage capacity, and support for Intel Optane NVMe based caching SSDs.

Solution Overview

Introduction

The Cisco HyperFlex System provides an all-purpose virtualized server platform, with hypervisor hosts, networking connectivity, and virtual server storage across a set of Cisco UCS HX-Series x86 rack-mount servers. Legacy datacenter deployments have relied on a disparate set of technologies, each performing a distinct and specialized function, such as network switches connecting endpoints and transferring Ethernet network traffic, and Fibre Channel (FC) storage arrays providing block based storage via a dedicated storage array network (SAN). Each of these systems had unique requirements for hardware, connectivity, management tools, operational knowledge, monitoring, and ongoing support. A legacy virtual server environment was often divided up into areas commonly referred to as silos, within which only a single technology operated, along with their correlated software tools and support staff. Silos could often be divided between the x86 computing hardware, the networking connectivity of those x86 servers, SAN connectivity and storage device presentation, the hypervisors and virtual platform management, and finally the guest VM themselves along with their OS and applications. This model proves to be inflexible, difficult to navigate, and is susceptible to numerous operational inefficiencies.

A more modern datacenter model was developed called a converged infrastructure. Converged infrastructures attempt to collapse the traditional silos by combining these technologies into a more singular environment, which has been designed to operate together in pre-defined, tested, and validated designs. A key component of the converged infrastructure was the revolutionary combination of x86 rack and blade servers, along with converged Ethernet and Fibre Channel networking offered by the Cisco UCS platform. Converged infrastructures leverage Cisco UCS, plus new deployment tools, management software suites, automation processes, and orchestration tools to overcome the difficulties deploying traditional environments, and do so in a much more rapid fashion. These new tools place the ongoing management and operation of the system into the hands of fewer staff, with more rapid deployment of workloads based on business needs, while still remaining at the forefront of flexibility to adapt to workload needs, and offering the highest possible performance. Cisco has had incredible success in these areas with our various partners, developing leading solutions such as Cisco FlexPod, FlashStack, VersaStack, and VxBlock architectures. Despite these advances, because these converged infrastructures contained some legacy technology stacks, particularly in the storage subsystems, there often remained a division of responsibility amongst multiple teams of administrators. Alongside, there is also a recognition that these converged infrastructures can still be a somewhat complex combination of components, where a simpler system would suffice to serve the workloads being requested.

Significant changes in the storage marketplace have given rise to the software defined storage (SDS) system. Legacy FC storage arrays often contained a specialized subset of hardware, such as Fibre Channel Arbitrated Loop (FC-AL) based controllers and disk shelves along with optimized Application Specific Integrated Circuits (ASIC), read/write data caching modules and cards, plus highly customized software to operate the arrays. With the rise of Serial Attached SCSI (SAS) bus technology and its inherent benefits, storage array vendors began to transition their internal hardware architectures to SAS, and with dramatic increases in processing power from recent x86 processor architectures, they also used fewer or no custom ASICs at all. As disk physical sizes shrank, x86 servers began to have the same density of storage per rack unit (RU) as the arrays themselves, and with the proliferation of NAND based flash memory solid state disks (SSD), they also now had access to input/output (IO) devices whose speed rivaled that of dedicated caching devices. If servers themselves now contained storage devices and technology to rival many dedicated arrays on the market, then the major differentiator between them was the software providing allocation, presentation and management of the storage, plus the advanced features many vendors offered. This has led to the rise of software defined storage, where the x86 servers with the storage devices ran software to effectively turn one or more of them, working cooperatively, into a storage array much the same as the

traditional arrays were. In a somewhat unexpected turn of events, some of the major storage array vendors themselves were pioneers in this field, recognizing the technological shifts in the market, and attempting to profit from the software features they offered versus their specialized hardware, as had been done in the past.

Some early uses of SDS systems simply replaced the traditional storage array in the converged architectures as described earlier. That configuration still had a separate storage system from the virtual server hypervisor platform, and depending on the solution provider, still remained separate from the network devices. If the servers that hosted the VMs, and also provided the SDS environment were in fact the same model of server, could they simply do both things at once and collapse the two functions into one? This ultimate combination of resources becomes what the industry has given the moniker of a hyperconverged infrastructure. Hyperconverged infrastructures coalesce the computing, memory, hypervisor, and storage devices of servers into a single platform for virtual servers. There is no longer a separate storage system, as the servers running the hypervisors also provide the software defined storage resources to store the virtual servers, effectively storing the virtual machines on themselves. Now nearly all the silos are gone, and a hyperconverged infrastructure becomes something almost completely self-contained, simpler to use, faster to deploy, easier to consume, yet still flexible and with very high performance. Many hyperconverged systems still rely on standard networking components, such as on-board network cards in the x86 servers, and top-of-rack switches. The Cisco HyperFlex system combines the convergence of computing and networking provided by Cisco UCS, along with next-generation hyperconverged storage software, to uniquely provide the compute resources, network connectivity, storage, and hypervisor platform to run an entire virtual environment, all contained in a single uniform system.

Some key advantages of hyperconverged infrastructures are the simplification of deployment, day to day management operations, as well as increased agility, thereby reducing the amount operational costs. Since hyperconverged storage can be easily managed by an IT generalist, this can also reduce technical debt going forward that is often accrued by implementing complex systems that need dedicated management teams and skillsets.

Audience

The intended audience for this document includes, but is not limited to, sales engineers, field consultants, professional services, IT managers, partner engineering, and customers deploying the Cisco HyperFlex System. External references are provided wherever applicable, but readers are expected to be familiar with VMware specific technologies, infrastructure concepts, networking connectivity, and security policies of the customer installation.

Purpose of this Document

This document describes the steps required to deploy, configure, and manage a Cisco HyperFlex system using the VMware ESXi hypervisor. The document is based on all known best practices using the software, hardware and firmware revisions specified in the document. As such, recommendations and best practices can be amended with later versions. This document showcases the installation, configuration and expansion of Cisco HyperFlex standard and also extended clusters, including both converged nodes and compute-only nodes, in a typical customer datacenter environment. While readers of this document are expected to have sufficient knowledge to install and configure the products used, configuration details that are important to the deployment of this solution are provided in this CVD.

Enhancements for Version 3.0

The Cisco HyperFlex system has several new capabilities and enhancements in version 3.0:

- Multi-hypervisor support allows HyperFlex to be installed with either the VMware ESXi hypervisor, or Microsoft Hyper-V. This document focuses on installation and support of HyperFlex with the VMware ESXi hypervisor.
- Installation of a Cisco HyperFlex cluster can span two physical locations, creating a stretched cluster. A third location is required for running a witness virtual machine to prevent a “split brain” situation.
- HyperFlex clusters can be configured with logical availability zones, which subdivide the nodes into groups and evenly distribute the data across all zones, in order to better tolerate node failures.
- Support for 64 node clusters; up to 32 converged nodes and 32 compute-only nodes can be used per cluster.
- Support for using Intel Optane based NVMe based SSDs as the caching disk in the Cisco HyperFlex all-flash nodes.
- Support for large form factor hard drives in the HyperFlex HX240-M5SL model server for higher storage capacity.
- Enhancements and customizations available for the HyperFlex Connect native HTML5 management GUI.
- Kubernetes support with automated storage and networking deployment via a new FlexVolume driver, creating a fully integrated container platform.

Documentation Roadmap

For the comprehensive documentation suite, refer to the following for the Cisco UCS HX-Series Documentation Roadmap:
https://www.cisco.com/c/en/us/td/docs/hyperconverged_systems/HyperFlex_HX_DataPlatformSoftware/HX_Documentation_Roadmap/HX_Series_Doc_Roadmap.html

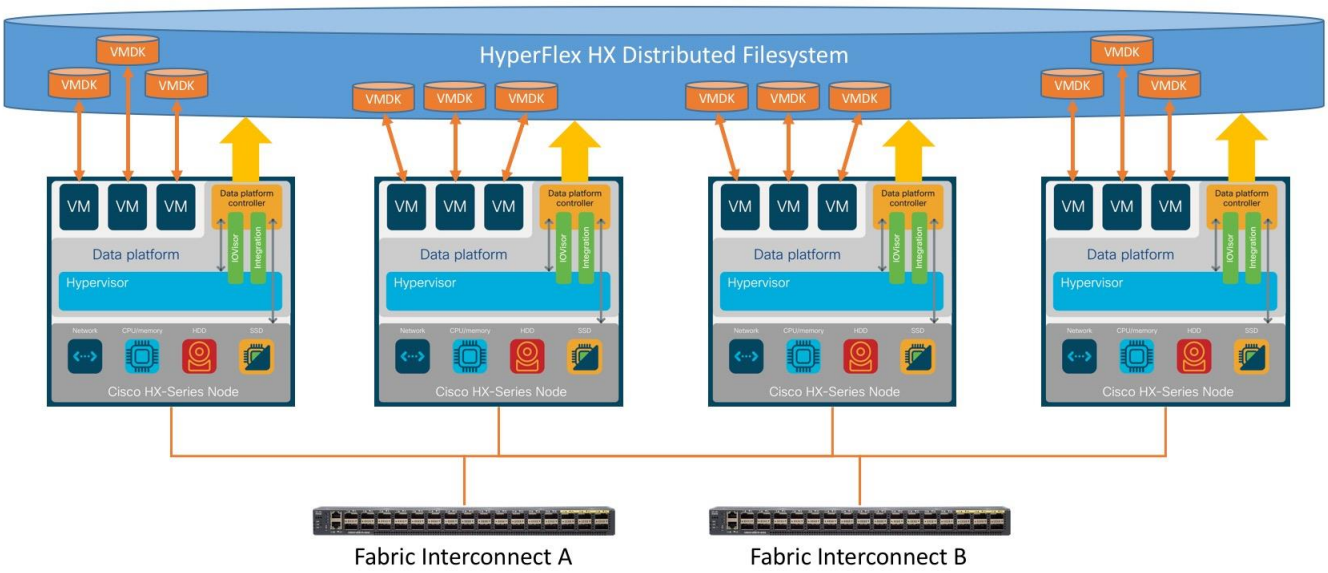
Note: A login is required for the Documentation Roadmap.

Hyperconverged Infrastructure web link: http://hyperflex.io

Solution Summary

The Cisco HyperFlex system provides a fully contained virtual server platform, with compute and memory resources, integrated networking connectivity, a distributed high-performance log based filesystem for VM storage, and the hypervisor software for running the virtualized servers, all within a single Cisco UCS management domain.

Figure 1 HyperFlex System Overview



The following are the components of a Cisco HyperFlex system using the VMware ESXi Hypervisor:

- One pair of Cisco UCS Fabric Interconnects, choose from models:
  - Cisco UCS 6248UP Fabric Interconnect
  - Cisco UCS 6296UP Fabric Interconnect
  - Cisco UCS 6332 Fabric Interconnect
  - Cisco UCS 6332-16UP Fabric Interconnect
- Three to Thirty-Two Cisco HyperFlex HX-Series Rack-Mount Servers, choose from models:
  - Cisco HyperFlex HX220c-M5SX Rack-Mount Servers
  - Cisco HyperFlex HX240c-M5SX Rack-Mount Servers
  - Cisco HyperFlex HX240c-M5L Rack-Mount Servers
  - Cisco HyperFlex HXAF220c-M5SX All-Flash Rack-Mount Servers
  - Cisco HyperFlex HXAF240c-M5SX All-Flash Rack-Mount Servers
  - Cisco HyperFlex HX220c-M4S Rack-Mount Servers
  - Cisco HyperFlex HX240c-M4SX Rack-Mount Servers
  - Cisco HyperFlex HXAF220c-M4S All-Flash Rack-Mount Servers
  - Cisco HyperFlex HXAF240c-M4SX All-Flash Rack-Mount Servers
- Cisco HyperFlex Data Platform Software
- VMware vSphere ESXi Hypervisor
- VMware vCenter Server (end-user supplied)

Optional components for additional compute-only resources are:

- Cisco UCS 5108 Chassis
- Cisco UCS 2204XP, 2208XP or 2304 model Fabric Extenders
- Cisco UCS B200-M3, B200-M4, B200-M5, B260-M4, B420-M4, B460-M4 or B480-M5 blade servers
- Cisco UCS C220-M3, C220-M4, C220-M5, C240-M3, C240-M4, C240-M5, C460-M4 or C480-M5 rack-mount servers

Technology Overview

Cisco Unified Computing System

The Cisco Unified Computing System (Cisco UCS) is a next-generation data center platform that unites compute, network, and storage access. The platform, optimized for virtual environments, is designed using open industry-standard technologies and aims to reduce total cost of ownership (TCO) and increase business agility. The system integrates a low-latency, lossless 10 Gigabit Ethernet or 40 Gigabit Ethernet unified network fabric with enterprise-class, x86-architecture servers. It is an integrated, scalable, multi chassis platform in which all resources participate in a unified management domain.

The main components of Cisco Unified Computing System are:

- Computing: The system is based on an entirely new class of computing system that incorporates rack-mount and blade servers based on Intel Xeon Processors.
- Network: The system is integrated onto a low-latency, lossless, 10-Gbps or 40-Gbps unified network fabric. This network foundation consolidates LANs, SANs, and high-performance computing networks which are often separate networks today. The unified fabric lowers costs by reducing the number of network adapters, switches, and cables, and by decreasing the power and cooling requirements.
- Virtualization: The system unleashes the full potential of virtualization by enhancing the scalability, performance, and operational control of virtual environments. Cisco security, policy enforcement, and diagnostic features are now extended into virtualized environments to better support changing business and IT requirements.
- Storage access: The system provides consolidated access to both SAN storage and Network Attached Storage (NAS) over the unified fabric. By unifying storage access, the Cisco Unified Computing System can access storage over Ethernet, Fibre Channel, Fibre Channel over Ethernet (FCoE), and iSCSI. This provides customers with their choice of storage protocol and physical architecture, and enhanced investment protection. In addition, the server administrators can pre-assign storage-access policies for system connectivity to storage resources, simplifying storage connectivity, and management for increased productivity.
- Management: The system uniquely integrates all system components which enable the entire solution to be managed as a single entity by the Cisco UCS Manager (UCSM). The Cisco UCS Manager has an intuitive graphical user interface (GUI), a command-line interface (CLI), and a robust application programming interface (API) to manage all system configuration and operations.

The Cisco Unified Computing System is designed to deliver:

- A reduced Total Cost of Ownership and increased business agility.
- Increased IT staff productivity through just-in-time provisioning and mobility support.
- A cohesive, integrated system which unifies the technology in the data center. The system is managed, serviced and tested as a whole.
- Scalability through a design for hundreds of discrete servers and thousands of virtual machines and the capability to scale I/O bandwidth to match demand.
- Industry standards supported by a partner ecosystem of industry leaders.

Cisco UCS Fabric Interconnect

The Cisco UCS Fabric Interconnect (FI) is a core part of the Cisco Unified Computing System, providing both network connectivity and management capabilities for the system. Depending on the model chosen, the Cisco UCS Fabric Interconnect offers line-rate, low-latency, lossless 10 Gigabit or 40 Gigabit Ethernet, Fibre Channel over Ethernet (FCoE) and Fibre Channel connectivity. Cisco UCS Fabric Interconnects provide the management and communication backbone for the Cisco UCS C-Series, S-Series and HX-Series Rack-Mount Servers, Cisco UCS B-Series Blade Servers and Cisco UCS 5100 Series Blade Server Chassis. All servers and chassis, and therefore all blades, attached to the Cisco UCS Fabric Interconnects become part of a single, highly available management domain. In addition, by supporting unified fabrics, the Cisco UCS Fabric Interconnects provide both the LAN and SAN connectivity for all servers within its domain.

From a networking perspective, the Cisco UCS 6200 Series uses a cut-through architecture, supporting deterministic, low latency, line rate 10 Gigabit Ethernet on all ports, up to 1.92 Tbps switching capacity and 160 Gbps bandwidth per chassis, independent of packet size and enabled services. The product family supports Cisco low-latency, lossless 10 Gigabit Ethernet unified network fabric capabilities, which increase the reliability, efficiency, and scalability of Ethernet networks. The Fabric Interconnect supports multiple traffic classes over the Ethernet fabric from the servers to the uplinks. Significant TCO savings come from an FCoE-optimized server design in which network interface cards (NICs), host bus adapters (HBAs), cables, and switches can be consolidated.

The Cisco UCS 6300 Series offers the same features while supporting even higher performance, low latency, lossless, line rate 40 Gigabit Ethernet, with up to 2.56 Tbps of switching capacity. Backward compatibility and scalability are assured with the ability to configure 40 Gbps quad SFP (QSFP) ports as breakout ports using 4x10GbE breakout cables. Existing Cisco UCS servers with 10GbE interfaces can be connected in this manner, although Cisco HyperFlex nodes must use a 40GbE VIC adapter in order to connect to a Cisco UCS 6300 Series Fabric Interconnect.

Cisco UCS 6248UP Fabric Interconnect

The Cisco UCS 6248UP Fabric Interconnect is a one-rack-unit (1RU) 10 Gigabit Ethernet, FCoE and Fiber Channel switch offering up to 960 Gbps throughput and up to 48 ports. The switch has 32 1/10-Gbps fixed Ethernet, FCoE, or 1/2/4/8 Gbps FC ports, plus one expansion slot.

Figure 2 Cisco UCS 6248UP Fabric Interconnect

Cisco UCS 6296UP Fabric Interconnect

The Cisco UCS 6296UP Fabric Interconnect is a two-rack-unit (2RU) 10 Gigabit Ethernet, FCoE, and native Fibre Channel switch offering up to 1920 Gbps of throughput and up to 96 ports. The switch has 48 1/10-Gbps fixed Ethernet, FCoE, or 1/2/4/8 Gbps FC ports, plus three expansion slots.

Figure 3 Cisco UCS 6296UP Fabric Interconnect

Cisco UCS 6332 Fabric Interconnect

The Cisco UCS 6332 Fabric Interconnect is a one-rack-unit (1RU) 40 Gigabit Ethernet and FCoE switch offering up to 2560 Gbps of throughput. The switch has 32 40-Gbps fixed Ethernet and FCoE ports. Up to 24 of the ports can be reconfigured as 4x10Gbps breakout ports, providing up to 96 10-Gbps ports.

Figure 4 Cisco UCS 6332 Fabric Interconnect

Cisco UCS 6332-16UP Fabric Interconnect

The Cisco UCS 6332-16UP Fabric Interconnect is a one-rack-unit (1RU) 10/40 Gigabit Ethernet, FCoE, and native Fibre Channel switch offering up to 2430 Gbps of throughput. The switch has 24 40-Gbps fixed Ethernet and FCoE ports, plus 16 1/10-Gbps fixed Ethernet, FCoE, or 4/8/16 Gbps FC ports. Up to 18 of the 40-Gbps ports can be reconfigured as 4x10Gbps breakout ports, providing up to 88 total 10-Gbps ports.

Figure 5 Cisco UCS 6332-16UP Fabric Interconnect

Note: When used for a Cisco HyperFlex deployment, due to mandatory QoS settings in the configuration, the 6332 and 6332-16UP will be limited to a maximum of four 4x10Gbps breakout ports, which can be used for other non-HyperFlex servers.

Cisco HyperFlex HX-Series Nodes

A HyperFlex cluster requires a minimum of three HX-Series “converged” nodes (with disk storage). Data is replicated across at least two of these nodes, and a third node is required for continuous operation in the event of a single-node failure. Each node that has disk storage is equipped with at least one high-performance SSD drive for data caching and rapid acknowledgment of write requests. Each node also is equipped with additional disks, up to the platform’s physical limit, for long term storage and capacity.

Cisco HyperFlex HXAF220c-M5SX All-Flash Node

This small footprint Cisco HyperFlex all-flash model contains a 240 GB M.2 form factor solid-state disk (SSD) that acts as the boot drive, a 240 GB housekeeping SSD drive, either a single 375 GB Optane NVMe

SSD, a 1.6 TB NVMe SSD or 400GB SAS SSD write-log drive, and six to eight 960 GB or 3.8 TB SATA SSD drives for storage capacity. For configurations requiring self-encrypting drives, the caching SSD is replaced with an 800 GB SAS SED SSD, and the capacity disks are also replaced with either 800 GB, 960 GB or 3.8 TB SED SSDs.

Figure 6 HXAF220c-M5SX All-Flash Node

Cisco HyperFlex HXAF240c-M5SX All-Flash Node

This capacity optimized Cisco HyperFlex all-flash model contains a 240 GB M.2 form factor solid-state disk (SSD) that acts as the boot drive, a 240 GB housekeeping SSD drive, either a single 375 GB Optane NVMe SSD, a 1.6 TB NVMe SSD or 400GB SAS SSD write-log drive installed in a rear hot swappable slot, and six to twenty-three 960 GB or 3.8 TB SATA SSD drives for storage capacity. For configurations requiring self-encrypting drives, the caching SSD is replaced with an 800 GB SAS SED SSD, and the capacity disks are also replaced with either 800 GB, 960 GB or 3.8 TB SED SSDs.

Figure 7 HXAF240c-M5SX Node

Note: Either a 375 GB Optane NVMe SSD, a 400 GB SAS SSD or 1.6 TB NVMe SSD caching drive may be chosen. While the NVMe options can provide a higher level of performance, the partitioning of the three disk options is the same, therefore the amount of cache available on the system is the same regardless of the model chosen.

Cisco HyperFlex HX220c-M5SX Hybrid Node

This small footprint Cisco HyperFlex hybrid model contains a minimum of six, and up to eight 1.8 terabyte (TB) or 1.2 TB SAS hard disk drives (HDD) that contribute to cluster storage capacity, a 240 GB SSD housekeeping drive, a 480 GB or 800 GB SSD caching drive, and a 240 GB M.2 form factor SSD that acts as the boot drive. For configurations requiring self-encrypting drives, the caching SSD is replaced with an 800 GB SAS SED SSD, and the capacity disks are replaced with 1.2TB SAS SED HDDs.

Figure 8 HX220c-M5SX Node

Note: Either a 480 GB or 800 GB caching SAS SSD may be chosen. This option is provided to allow flexibility in ordering based on product availability, pricing and lead times. There is no performance, capacity, or scalability benefit in choosing the larger disk.

Cisco HyperFlex HX240c-M5SX Hybrid Node

This capacity optimized Cisco HyperFlex hybrid model contains a minimum of six and up to twenty-three 1.8 TB or 1.2 TB SAS small form factor (SFF) hard disk drives (HDD) that contribute to cluster storage, a 240 GB SSD housekeeping drive, a single 1.6 TB SSD caching drive installed in a rear hot swappable slot, and a 240 GB M.2 form factor SSD that acts as the boot drive. For configurations requiring self-encrypting drives, the caching SSD is replaced with a 1.6 TB SAS SED SSD, and the capacity disks are replaced with 1.2TB SAS SED HDDs.

Figure 9 HX240c-M5SX Node

Cisco HyperFlex HX240c-M5L Hybrid Node

This density optimized Cisco HyperFlex hybrid model contains a minimum of six and up to twelve 6 TB or 8 TB SAS large form factor (LFF) hard disk drives (HDD) that contribute to cluster storage, a 240 GB SSD housekeeping drive and a single 3.2 TB SSD caching drive, both installed in the rear hot swappable slots, and a 240 GB M.2 form factor SSD that acts as the boot drive. Large form factor nodes cannot be configured with self-encrypting disks, and are limited to a maximum of eight nodes in a cluster in the initial release of HyperFlex 3.0.

Figure 10 HX240c-M5L Node

Cisco HyperFlex HXAF220c-M4S All-Flash Node

This small footprint Cisco HyperFlex all-flash model contains two Cisco Flexible Flash (FlexFlash) Secure Digital (SD) cards that act as the boot drives, a single 120 GB or 240 GB solid-state disk (SSD) data-logging drive, a single 400 GB NVMe or a 400GB or 800 GB SAS SSD write-log drive, and six 960 GB or 3.8 terabyte

(TB) SATA SSD drives for storage capacity. For configurations requiring self-encrypting drives, the caching SSD is replaced with an 800 GB SAS SED SSD, and the capacity disks are also replaced with either 800 GB, 960 GB or 3.8 TB SED SSDs.

Figure 11 HXAF220c-M4S All-Flash Node

Cisco HyperFlex HXAF240c-M4SX All-Flash Node

This capacity optimized Cisco HyperFlex all-flash model contains two FlexFlash SD cards that act as boot drives, a single 120 GB or 240 GB solid-state disk (SSD) data-logging drive, a single 400 GB NVMe or a 400GB or 800 GB SAS SSD write-log drive, and six to twenty-three 960 GB or 3.8 terabyte (TB) SATA SSD drives for storage capacity. For configurations requiring self-encrypting drives, the caching SSD is replaced with an 800 GB SAS SED SSD, and the capacity disks are also replaced with either 800 GB, 960 GB or 3.8 TB SED SSDs.

Figure 12 HXAF240c-M4SX Node

Note: In M4 generation server all-flash configurations, either a 400 GB or 800 GB caching SAS SSD may be chosen. This option is provided to allow flexibility in ordering based on product availability, pricing and lead times. There is no performance, capacity, or scalability benefit in choosing the larger disk.

Cisco HyperFlex HX220c-M4S Hybrid Node

This small footprint Cisco HyperFlex hybrid model contains six 1.8 terabyte (TB) or 1.2 TB SAS HDD drives that contribute to cluster storage capacity, a 120 GB or 240 GB SSD housekeeping drive, a 480 GB SAS SSD caching drive, and two Cisco Flexible Flash (FlexFlash) Secure Digital (SD) cards that act as boot drives. For configurations requiring self-encrypting drives, the caching SSD is replaced with an 800 GB SAS SED SSD, and the capacity disks are replaced with 1.2TB SAS SED HDDs.

Figure 13 HX220c-M4S Node

Cisco HyperFlex HX240c-M4SX Hybrid Node

This capacity optimized Cisco HyperFlex hybrid model contains a minimum of six and up to twenty-three 1.8 TB or 1.2 TB SAS HDD drives that contribute to cluster storage, a single 120 GB or 240 GB SSD housekeeping drive, a single 1.6 TB SAS SSD caching drive, and two FlexFlash SD cards that act as the boot drives. For configurations requiring self-encrypting drives, the caching SSD is replaced with a 1.6 TB SAS SED SSD, and the capacity disks are replaced with 1.2TB SAS SED HDDs.

Figure 14 HX240c-M4SX Node

Note: In all M4 generation server configurations, either a 120 GB or 240 GB housekeeping disk may be chosen. This option is provided to allow flexibility in ordering based on product availability, pricing and lead times. There is no performance, capacity or scalability benefit in choosing the larger disk.

Cisco VIC 1227 and 1387 MLOM Interface Cards

The Cisco UCS Virtual Interface Card (VIC) 1227 is a dual-port Enhanced Small Form-Factor Pluggable (SFP+) 10-Gbps Ethernet and Fibre Channel over Ethernet (FCoE)-capable PCI Express (PCIe) modular LAN-on-motherboard (mLOM) adapter installed in the Cisco UCS HX-Series Rack Servers. The VIC 1227 is used in conjunction with the Cisco UCS 6248UP or 6296UP model Fabric Interconnects.

The Cisco UCS VIC 1387 Card is a dual-port Enhanced Quad Small Form-Factor Pluggable (QSFP+) 40-Gbps Ethernet and Fibre Channel over Ethernet (FCoE)-capable PCI Express (PCIe) modular LAN-on-motherboard (mLOM) adapter installed in the Cisco UCS HX-Series Rack Servers. The VIC 1387 is used in conjunction with the Cisco UCS 6332 or 6332-16UP model Fabric Interconnects.

The mLOM slot can be used to install a Cisco VIC without consuming a PCIe slot, which provides greater I/O expandability. It incorporates next-generation converged network adapter (CNA) technology from Cisco, providing investment protection for future feature releases. The card enables a policy-based, stateless, agile server infrastructure that can present up to 256 PCIe standards-compliant interfaces to the host, each dynamically configured as either a network interface card (NICs) or host bus adapter (HBA). The personality of the interfaces is set programmatically using the service profile associated with the server. The number, type (NIC or HBA), identity (MAC address and World Wide Name [WWN]), failover policy, adapter settings, bandwidth, and quality-of-service (QoS) policies of the PCIe interfaces are all specified using the service profile.

Figure 15 Cisco VIC 1227 mLOM Card

Figure 16 Cisco VIC 1387 mLOM Card

Note: Hardware revision V03 or later of the Cisco VIC 1387 card is required for the Cisco HyperFlex HX-series servers.

All-Flash Versus Hybrid

The initial HyperFlex product release featured hybrid converged nodes, which use a combination of solid-state disks (SSDs) for the short-term storage caching layer, and hard disk drives (HDDs) for the long-term storage capacity layer. The hybrid HyperFlex system is an excellent choice for entry-level or midrange storage solutions, and hybrid solutions have been successfully deployed in many non-performance sensitive virtual environments. Meanwhile, there is significant growth in deployment of highly performance sensitive and mission critical applications. The primary challenge to the hybrid HyperFlex system from these highly performance sensitive applications, is their increased sensitivity to high storage latency. Due to the characteristics of the spinning hard disks, it is unavoidable that their higher latency becomes the bottleneck in the hybrid system. Ideally, if all of the storage operations were to occur in the caching SSD layer, the hybrid system's performance will be excellent. But in several scenarios, the amount of data being written and read exceeds the caching layer capacity, placing larger loads on the HDD capacity layer, and the subsequent increases in latency will naturally result in reduced performance.

Cisco All-Flash HyperFlex systems are an excellent option for customers with a requirement to support high performance, latency sensitive workloads. With a purpose built, flash-optimized and high-performance log based filesystem, the Cisco All-Flash HyperFlex system provides:

- Predictable high performance across all the virtual machines on HyperFlex All-Flash and compute-only nodes in the cluster.
- Highly consistent and low latency, which benefits data-intensive applications and databases such as Microsoft SQL and Oracle.
- Support for NVMe caching SSDs, offering an even higher level of performance.

- Future ready architecture that is well suited for flash-memory configuration:
  - Cluster-wide SSD pooling maximizes performance and balances SSD usage so as to spread the wear.
  - A fully distributed log-structured filesystem optimizes the data path to help reduce write amplification.
  - Large sequential writing reduces flash wear and increases component longevity.
  - Inline space optimization, for example; deduplication and compression, minimizes data operations and reduces wear.
- Lower operating cost with the higher density drives for increased capacity of the system.
- Cloud scale solution with easy scale-out and distributed infrastructure and the flexibility of scaling out independent resources separately.

Cisco HyperFlex support for hybrid and all-flash models now allows customers to choose the right platform configuration based on their capacity, applications, performance, and budget requirements. All-flash configurations offer repeatable and sustainable high performance, especially for scenarios with a larger working set of data, in other words, a large amount of data in motion. Hybrid configurations are a good option for customers who want the simplicity of the Cisco HyperFlex solution, but their needs focus on capacity-sensitive solutions, lower budgets, and fewer performance-sensitive applications.

Cisco HyperFlex Compute-Only Nodes

All current model Cisco UCS M4 and M5 generation servers, except the C880 M4 and C880 M5, may be used as compute-only nodes connected to a Cisco HyperFlex cluster, along with a limited number of previous M3 generation servers. Any valid CPU and memory configuration is allowed in the compute-only nodes, and the servers can be configured to boot from SAN, local disks, or internal SD cards. The following servers may be used as compute-only nodes:

- Cisco UCS B200 M3 Blade Server
- Cisco UCS B200 M4 Blade Server
- Cisco UCS B200 M5 Blade Server
- Cisco UCS B260 M4 Blade Server
- Cisco UCS B420 M4 Blade Server
- Cisco UCS B460 M4 Blade Server
- Cisco UCS B480 M5 Blade Server
- Cisco UCS C220 M3 Rack-Mount Servers
- Cisco UCS C220 M4 Rack-Mount Servers
- Cisco UCS C220 M5 Rack-Mount Servers
- Cisco UCS C240 M3 Rack-Mount Servers
- Cisco UCS C240 M4 Rack-Mount Servers

- Cisco UCS C240 M5 Rack-Mount Servers
- Cisco UCS C460 M4 Rack-Mount Servers
- Cisco UCS C480 M5 Rack-Mount Servers

Cisco HyperFlex Data Platform Software

The Cisco HyperFlex HX Data Platform is a purpose-built, high-performance, distributed file system with a wide array of enterprise-class data management services. The data platform's innovations redefine distributed storage technology, exceeding the boundaries of first-generation hyperconverged infrastructures. The data platform has all the features expected in an enterprise shared storage system, eliminating the need to configure and maintain complex Fibre Channel storage networks and devices. The platform simplifies operations and helps ensure data availability. Enterprise-class storage features include the following:

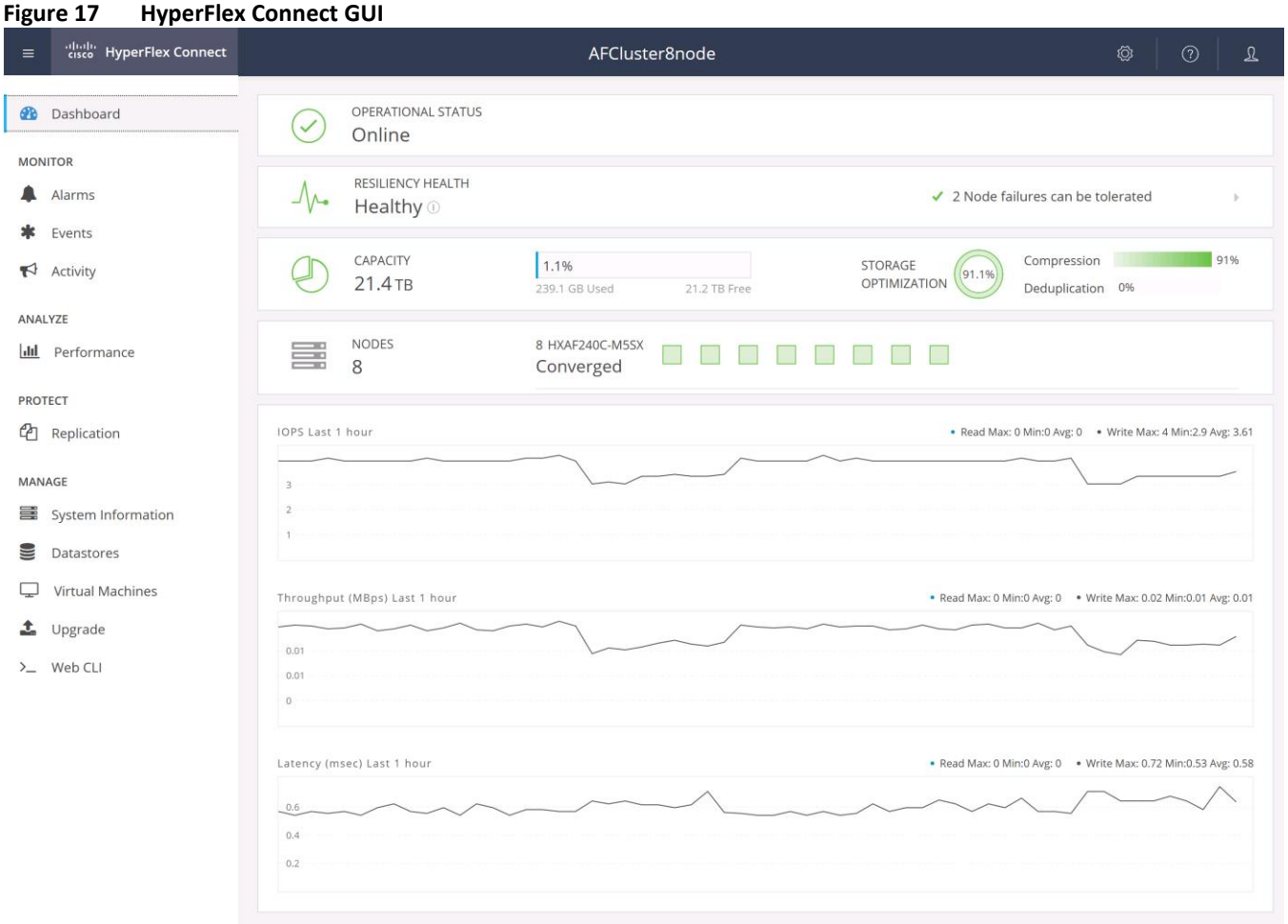
- **Data protection** creates multiple copies of the data across the cluster so that data availability is not affected if single or multiple components fail (depending on the replication factor configured).
- **Stretched clusters** allow nodes to be evenly split between two physical locations, keeping a duplicate copy of all data in both locations, thereby providing protection in case of an entire site failure.
- **Logical availability zones** provide multiple logical grouping of nodes and distributes the data across these groups in such a way that no single group has more than one copy of the data. This enables enhanced protection from node failures, allowing for more nodes to fail while the overall cluster remains online.
- **Deduplication** is always on, helping reduce storage requirements in virtualization clusters in which multiple operating system instances in guest virtual machines result in large amounts of replicated data.
- **Compression** further reduces storage requirements, reducing costs, and the log-structured file system is designed to store variable-sized blocks, reducing internal fragmentation.
- **Replication** copies virtual machine level snapshots from one Cisco HyperFlex cluster to another, to facilitate recovery from a cluster or site failure, via a failover to the secondary site of all VMs.
- **Encryption** stores all data on the caching and capacity disks in an encrypted format, to prevent accidental data loss or data theft. Key management can be done using local Cisco UCS Manager managed keys, or third-party Key Management Systems (KMS) via the Key Management Interoperability Protocol (KMIP).
- **Thin provisioning** allows large volumes to be created without requiring storage to support them until the need arises, simplifying data volume growth and making storage a "pay as you grow" proposition.
- **Fast, space-efficient clones** rapidly duplicate virtual storage volumes so that virtual machines can be cloned simply through metadata operations, with actual data copied only for write operations.
- **Snapshots** help facilitate backup and remote-replication operations, which are needed in enterprises that require always-on data availability.

Cisco HyperFlex Connect HTML5 Management Web Page

An all-new HTML 5 based Web UI is available for use as the primary management tool for Cisco HyperFlex. Through this centralized point of control for the cluster, administrators can create volumes, monitor the data

platform health, and manage resource use. Administrators can also use this data to predict when the cluster will need to be scaled. To use the HyperFlex Connect UI, connect using a web browser to the HyperFlex cluster IP address: http://<hx controller cluster ip>.

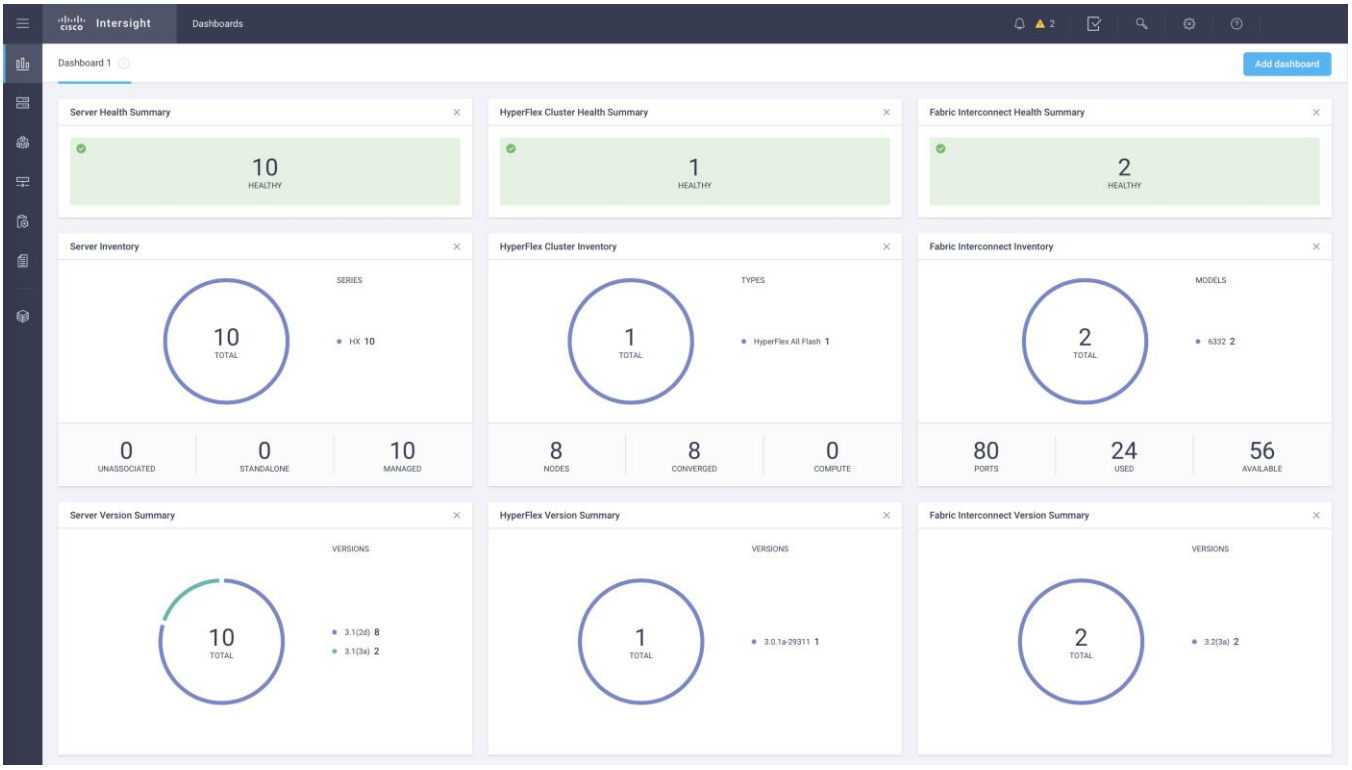
Figure 17 HyperFlex Connect GUI



Cisco Intersight Cloud Based Management

Cisco Intersight (https://intersight.com), previously known as Starship, is the latest visionary cloud-based management tool, designed to provide a centralized off-site management, monitoring and reporting tool for all of your Cisco UCS based solutions. In the initial release of Cisco Intersight, monitoring and reporting is enabled against Cisco HyperFlex clusters. The Cisco Intersight website and framework can be upgraded with new and enhanced features independently of the products that are managed, meaning that many new features and capabilities can come with no downtime or upgrades required by the end users. Future releases of Cisco HyperFlex will enable further functionality along with these upgrades to the Cisco Intersight framework. This unique combination of embedded and online technologies will result in a complete cloud-based management solution that can care for Cisco HyperFlex throughout the entire lifecycle, from deployment through retirement.

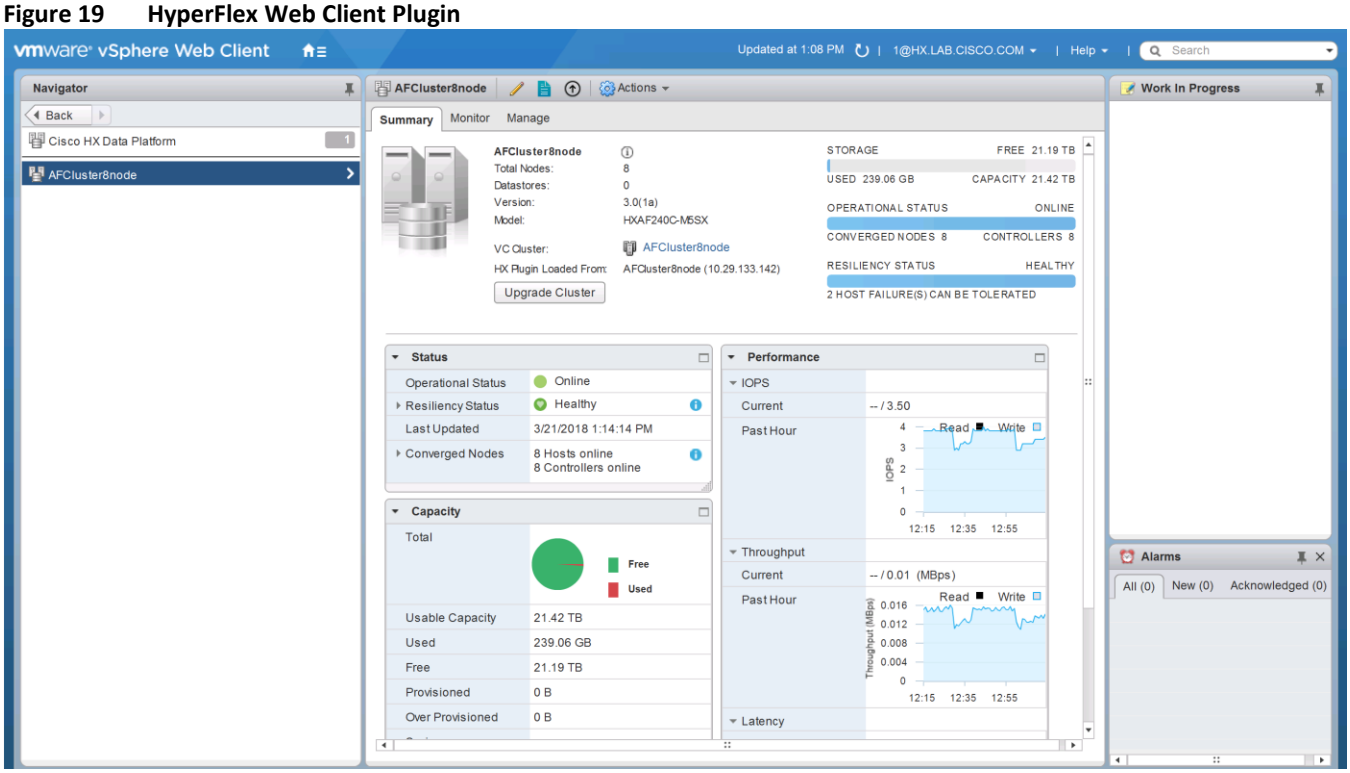
Figure 18 Cisco Intersight



Cisco HyperFlex HX Data Platform Administration Plug-in

The Cisco HyperFlex HX Data Platform is also administered secondarily through a VMware vSphere web client plug-in.

Figure 19 HyperFlex Web Client Plugin



Cisco HyperFlex HX Data Platform Controller

A Cisco HyperFlex HX Data Platform controller resides on each node and implements the distributed file system. The controller runs as software in user space within a virtual machine, and intercepts and handles all I/O from the guest virtual machines. The Storage Controller Virtual Machine (SCVM) uses the VMDirectPath I/O feature to provide PCI passthrough control of the physical server's SAS disk controller. This method gives the controller VM full control of the physical disk resources, utilizing the SSD drives as a read/write caching layer, and the HDDs or SDDs as a capacity layer for distributed storage. The controller integrates the data platform into the VMware vSphere cluster through the use of three preinstalled VMware ESXi vSphere Installation Bundles (VIBs) on each node:

- **IO Visor:** This VIB provides a network file system (NFS) mount point so that the ESXi hypervisor can access the virtual disks that are attached to individual virtual machines. From the hypervisor's perspective, it is simply attached to a network file system. The IO Visor intercepts guest VM IO traffic, and intelligently redirects it to the HyperFlex SCVMs.
- **VMware API for Array Integration (VAAI):** This storage offload API allows vSphere to request advanced file system operations such as snapshots and cloning. The controller implements these operations via manipulation of the filesystem metadata rather than actual data copying, providing rapid response, and thus rapid deployment of new environments.
- **stHypervisorSvc:** This VIB adds enhancements and features needed for HyperFlex data protection and VM replication.

Data Operations and Distribution

The Cisco HyperFlex HX Data Platform controllers handle all read and write operation requests from the guest VMs to their virtual disks (VMDK) stored in the distributed datastores in the cluster. The data platform distributes the data across multiple nodes of the cluster, and also across multiple capacity disks of each

node, according to the replication level policy selected during the cluster setup. This method avoids storage hotspots on specific nodes, and on specific disks of the nodes, and thereby also avoids networking hotspots or congestion from accessing more data on some nodes versus others.

Replication Factor

The policy for the number of duplicate copies of each storage block is chosen during cluster setup, and is referred to as the replication factor (RF).

- **Replication Factor 3:** For every I/O write committed to the storage layer, 2 additional copies of the blocks written will be created and stored in separate locations, for a total of 3 copies of the blocks. Blocks are distributed in such a way as to ensure multiple copies of the blocks are not stored on the same disks, nor on the same nodes of the cluster. This setting can tolerate simultaneous failures of 2 entire nodes in a cluster of 5 nodes or greater, without losing data and resorting to restore from backup or other recovery processes. RF3 is recommended for all production systems.
- **Replication Factor 2:** For every I/O write committed to the storage layer, 1 additional copy of the blocks written will be created and stored in separate locations, for a total of 2 copies of the blocks. Blocks are distributed in such a way as to ensure multiple copies of the blocks are not stored on the same disks, nor on the same nodes of the cluster. This setting can tolerate a failure of 1 entire node without losing data and resorting to restore from backup or other recovery processes. RF2 is suitable for non-production systems, or environments where the extra data protection is not needed. HyperFlex stretched clusters use the RF2 setting, however there are 2 copies of the data kept in both halves of the cluster, so effectively there are four copies stored.

Data Write and Compression Operations

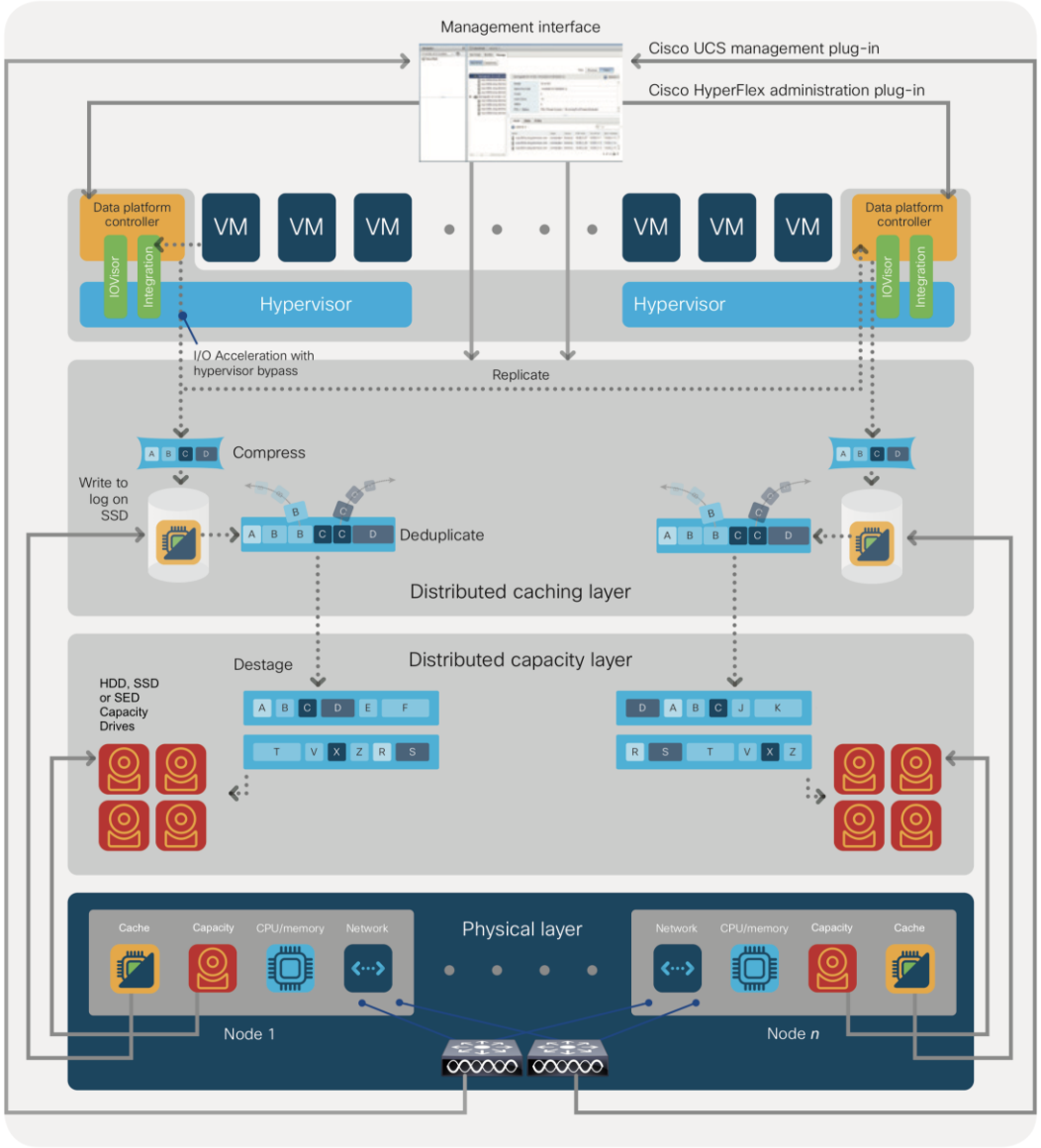
Internally, the contents of each virtual disk are subdivided and spread across multiple servers by the HXDP software. For each write operation, the data is intercepted by the IO Visor module on the node where the VM is running, a primary node is determined for that particular operation via a hashing algorithm, and then sent to the primary node via the network. The primary node compresses the data in real time, writes the compressed data to its caching SSD, and replica copies of that compressed data are written to the caching SSD of the remote nodes in the cluster, according to the replication factor setting. For example, at RF=3 a write will be written to the primary node for that virtual disk address, and two additional writes will be committed in parallel on two other nodes. Because the virtual disk contents have been divided and spread out via the hashing algorithm, this method results in all writes being spread across all nodes, avoiding the problems with data locality and “noisy” VMs consuming all the IO capacity of a single node. The write operation will not be acknowledged until all three copies are written to the caching layer SSDs. Written data is also cached in a write log area resident in memory in the controller VM, along with the write log on the caching SSDs. This process speeds up read requests when reads are requested of data that has recently been written.

Data Destaging and Deduplication

The Cisco HyperFlex HX Data Platform constructs multiple write caching segments on the caching SSDs of each node in the distributed cluster. As write cache segments become full, and based on policies accounting for I/O load and access patterns, those write cache segments are locked and new writes roll over to a new write cache segment. The data in the now locked cache segment is destaged to the HDD capacity layer of the nodes for the Hybrid system or to the SDD capacity layer of the nodes for the All-Flash system. During the destaging process, data is deduplicated before being written to the capacity storage layer, and the resulting data can now be written to the HDDs or SDDs of the server. On hybrid systems, the now deduplicated and compressed data is also written to the dedicated read cache area of the caching SSD, which speeds up read requests of data that has recently been written. When the data is destaged to a HDD, it is written in a single sequential operation, avoiding disk head seek thrashing on the spinning disks

and accomplishing the task in the minimal amount of time. Since the data is already deduplicated and compressed before being written, the platform avoids additional I/O overhead often seen on competing systems, which must later do a read/dedupe/compress/write cycle. Deduplication, compression and destaging take place with no delays or I/O penalties to the guest VMs making requests to read or write data, which benefits both the HDD and SDD configurations.

Figure 20 HyperFlex HX Data Platform Data Movement



Data Read Operations

For data read operations, data may be read from multiple locations. For data that was very recently written, the data is likely to still exist in the write log of the local platform controller memory, or the write log of the local caching layer disk. If local write logs do not contain the data, the distributed filesystem metadata will be queried to see if the data is cached elsewhere, either in write logs of remote nodes, or in the dedicated read cache area of the local and remote caching SSDs of hybrid nodes. Finally, if the data has not been accessed in a significant amount of time, the filesystem will retrieve the requested data from the distributed capacity layer. As requests for reads are made to the distributed filesystem and the data is retrieved from the

capacity layer, the caching SSDs of hybrid nodes populate their dedicated read cache area to speed up subsequent requests for the same data. This multi-tiered distributed system with several layers of caching techniques, ensures that data is served at the highest possible speed, leveraging the caching SSDs of the nodes fully and equally. All-flash configurations, however, do not employ a dedicated read cache because such caching does not provide any performance benefit; the persistent data copy already resides on high-performance SSDs.

In summary, the Cisco HyperFlex HX Data Platform implements a distributed, log-structured file system that performs data operations via two configurations:

- In a Hybrid configuration, the data platform provides a caching layer using SSDs to accelerate read requests and write responses, and it implements a storage capacity layer using HDDs.
- In an All-Flash configuration, the data platform provides a dedicated caching layer using high endurance SSDs to accelerate write responses, and it implements a storage capacity layer also using SSDs. Read requests are fulfilled directly from the capacity SSDs, as a dedicated read cache is not needed to accelerate read operations.

Solution Design

Requirements

The following sections detail the physical hardware, software revisions, and firmware versions required to install a single cluster of the Cisco HyperFlex system. Maximum cluster size of 64 nodes can be obtained by combining 32 converged nodes and 32 compute-only nodes.

Physical Components

Table 1 HyperFlex System Components

Component	Hardware Required
Fabric Interconnects	Two Cisco UCS 6248UP Fabric Interconnects, or Two Cisco UCS 6296UP Fabric Interconnects, or Two Cisco UCS 6332 Fabric Interconnects, or Two Cisco UCS 6332-16UP Fabric Interconnects
Servers	Three to Thirty-Two Cisco HyperFlex HXAF220c-M5SX All-Flash rack servers, or Three to Thirty-Two Cisco HyperFlex HXAF240c-M5SX All-Flash rack servers, or Three to Thirty-Two Cisco HyperFlex HX220c-M5SX Hybrid rack servers, or Three to Thirty-Two Cisco HyperFlex HX240c-M5SX Hybrid rack servers, or Three to Eight Cisco HyperFlex HX240c-M5L Hybrid rack servers, or Three to Thirty-Two Cisco HyperFlex HXAF220c-M4S All-Flash rack servers, or Three to Thirty-Two Cisco HyperFlex HXAF240c-M4SX All-Flash rack servers, or Three to Thirty-Two Cisco HyperFlex HX220c-M4S Hybrid rack servers, or Three to Thirty-Two Cisco HyperFlex HX240c-M4SX Hybrid rack servers

Note: The HX240c-M5L server is limited to a maximum of eight nodes per cluster.

For complete server specifications and more information, please refer to the links below:

Compare models:

http://www.cisco.com/c/en/us/products/hyperconverged-infrastructure/hyperflex-hx-series/index.html#compare-models

HXAF220c-M5SX Spec Sheet:

https://www.cisco.com/c/dam/en/us/products/collateral/hyperconverged-infrastructure/hyperflex-hx-series/hxaf-220c-m5-specsheet.pdf

HXAF240c-M5SX Spec Sheet:

https://www.cisco.com/c/dam/en/us/products/collateral/hyperconverged-infrastructure/hyperflex-hx-series/hxaf-240c-m5-specsheet.pdf

HX220c-M5SX Spec Sheet:

https://www.cisco.com/c/dam/en/us/products/collateral/hyperconverged-infrastructure/hyperflex-hx-series/hx-220c-m5-specsheet.pdf

HX240c-M5SX Spec Sheet:

https://www.cisco.com/c/dam/en/us/products/collateral/hyperconverged-infrastructure/hyperflex-hx-series/hx-240c-m5-specsheet.pdf

HX240c-M5L Spec Sheet:

https://www.cisco.com/c/dam/en/us/products/collateral/hyperconverged-infrastructure/hyperflex-hx-series/hx240c-m5-specsheet.pdf

HXAF220c-M4S Spec Sheet:

http://www.cisco.com/c/dam/en/us/products/collateral/hyperconverged-infrastructure/hyperflex-hx-series/HXAF220c_M4_SpecSheet.pdf

HXAF240c-M4S Spec Sheet:

http://www.cisco.com/c/dam/en/us/products/collateral/hyperconverged-infrastructure/hyperflex-hx-series/HXAF240c_M4_SpecSheet.pdf

HX220c-M4S Spec Sheet:

http://www.cisco.com/c/dam/en/us/products/collateral/hyperconverged-infrastructure/hyperflex-hx-series/datasheet-c78-736817.pdf

HX240c-M4SX Spec Sheet:

http://www.cisco.com/c/dam/en/us/products/collateral/hyperconverged-infrastructure/hyperflex-hx-series/datasheet-c78-736818.pdf

Table 2 lists the hardware component options for the HXAF220c-M5SX server model:

Table 2 HXAF220c-M5SX Server Options

HXAF220c-M5SX options		Hardware Required
Processors		Chose a matching pair of Intel Xeon Processor Scalable Family CPUs
Memory		192 GB to 3 TB of total memory using 16 GB, 32 GB, 64 GB, or 128 GB DDR4 2666 MHz 1.2v modules
Disk Controller		Cisco 12Gbps Modular SAS HBA
SSDs	Standard	• One 240 GB 2.5 Inch Enterprise Value 6G SATA SSD • One 400 GB 2.5 Inch Enterprise Performance 12G SAS SSD, or one 1.6 TB 2.5 Inch Enterprise Performance NVMe SSD, or one 375 GB 2.5 Inch Optane Extreme Performance SSD • Six to eight 3.8 TB 2.5 Inch Enterprise Value 6G SATA SSDs, or six to eight 960 GB 2.5 Inch Enterprise Value 6G SATA SSDs
	SED	• One 240 GB 2.5 Inch Enterprise Value 6G SATA SSD • One 800 GB 2.5 Inch Enterprise Performance 12G SAS SED SSD • Six to eight 3.8 TB 2.5 Inch Enterprise Value 6G SATA SED SSDs, or six to eight 960 GB 2.5 Inch Enterprise Value 6G SATA SED SSDs, or six to eight 800 GB 2.5 Inch Enterprise Performance 12G SAS SED SSDs
Network		Cisco UCS VIC1387 VIC MLOM
Boot Device		One 240 GB M.2 form factor SATA SSD
microSD Card		One 32GB microSD card for local host utilities storage
Optional		Cisco QSA module to convert 40 GbE QSFP+ to 10 GbE SFP+

Table 3 lists the hardware component options for the HXAF240c-M5SX server model:

Table 3 HXAF240c-M5SX Server Options

HXAF240c-M5SX Options	Hardware Required
Processors	Chose a matching pair of Intel Xeon Processor Scalable Family CPUs

HXAF240c-M5SX Options		Hardware Required
Memory		192 GB to 3 TB of total memory using 16 GB, 32 GB, 64 GB, or 128 GB DDR4 2666 MHz 1.2v modules
Disk Controller		Cisco 12Gbps Modular SAS HBA
SSD	Standard	• One 240 GB 2.5 Inch Enterprise Value 6G SATA SSD • One 400 GB 2.5 Inch Enterprise Performance 12G SAS SSD, or one 1.6 TB 2.5 Inch Enterprise Performance NVMe SSD, or one 375 GB 2.5 Inch Optane Extreme Performance SSD • Six to twenty-three 3.8 TB 2.5 Inch Enterprise Value 6G SATA SSDs, or six to twenty-three 960 GB 2.5 Inch Enterprise Value 6G SATA SSDs
	SED	• One 240 GB 2.5 Inch Enterprise Value 6G SATA SSD • One 800 GB 2.5 Inch Enterprise Performance 12G SAS SED SSD • Six to twenty-three 3.8 TB 2.5 Inch Enterprise Value 6G SATA SED SSDs, or six to twenty-three 960 GB 2.5 Inch Enterprise Value 6G SATA SED SSDs, or six to twenty-three 800 GB 2.5 Inch Enterprise Performance 12G SAS SED SSDs
Network		Cisco UCS VIC1387 VIC MLOM
Boot Device		One 240 GB M.2 form factor SATA SSD
microSD Card		One 32GB microSD card for local host utilities storage
Optional		Cisco QSA module to convert 40 GbE QSFP+ to 10 GbE SFP+

Table 4 lists the hardware component options for the HX220c-M5SX server model:

Table 4 HX220c-M5SX Server Options

HX220c-M5SX Options	Hardware Required
Processors	Chose a matching pair of Intel Xeon Processor Scalable Family CPUs
Memory	192 GB to 3 TB of total memory using 16 GB, 32 GB, 64 GB, or 128 GB DDR4 2666 MHz 1.2v modules

HX220c-M5SX Options		Hardware Required
Disk Controller		Cisco 12Gbps Modular SAS HBA
SSDs	Standard	• One 240 GB 2.5 Inch Enterprise Value 6G SATA SSD • One 480 GB 2.5 Inch Enterprise Performance 6G SATA SSD, or one 800 GB 2.5 Inch Enterprise Performance 12G SAS SSD
	SED	• One 240 GB 2.5 Inch Enterprise Value 6G SATA SSD • One 800 GB 2.5 Inch Enterprise Performance 12G SAS SED SSD
HDDs	Standard	Six to eight 1.8 TB or 1.2 TB SAS 12Gbps 10K rpm SFF HDD
	SED	Six to eight 1.2 TB SAS 12Gbps 10K rpm SFF SED HDD
Network		Cisco UCS VIC1387 VIC MLOM
Boot Device		One 240 GB M.2 form factor SATA SSD
microSD Card		One 32GB microSD card for local host utilities storage
Optional		Cisco QSA module to convert 40 GbE QSFP+ to 10 GbE SFP+

Table 5 lists the hardware component options for the HX240c-M5SX server model:

Table 5 HX240c-M5SX Server Options

HX240c-M5SX Options		Hardware Required
Processors		Chose a matching pair of Intel Xeon Processor Scalable Family CPUs
Memory		192 GB to 3 TB of total memory using 16 GB, 32 GB, 64 GB, or 128 GB DDR4 2666 MHz 1.2v modules
Disk Controller		Cisco 12Gbps Modular SAS HBA
SSDs	Standard	• One 240 GB 2.5 Inch Enterprise Value 6G SATA SSD • One 1.6 TB 2.5 Inch Enterprise Performance 12G SAS SSD

HX240c-M5SX Options		Hardware Required
	SED	• One 240 GB 2.5 Inch Enterprise Value 6G SATA SSD • One 1.6 TB 2.5 Inch Enterprise Performance 12G SAS SED SSD
HDDs	Standard	Six to twenty-three 1.8 TB or 1.2 TB SAS 12Gbps 10K rpm SFF HDD
	SED	Six to twenty-three 1.2 TB SAS 12Gbps 10K rpm SFF SED HDD
Network		Cisco UCS VIC1387 VIC MLOM
Boot Device		One 240 GB M.2 form factor SATA SSD
microSD Card		One 32GB microSD card for local host utilities storage
Optional		Cisco QSA module to convert 40 GbE QSFP+ to 10 GbE SFP+

Table 6 lists the hardware component options for the HX240c-M5L server model:

Table 6 HX240c-M5L Server Options

HX240c-M5L Options	Hardware Required
Processors	Chose a matching pair of Intel Xeon Processor Scalable Family CPUs
Memory	192 GB to 3 TB of total memory using 16 GB, 32 GB, 64 GB, or 128 GB DDR4 2666 MHz 1.2v modules
Disk Controller	Cisco 12Gbps Modular SAS HBA
SSDs	• One 240 GB 2.5 Inch Enterprise Value 6G SATA SSD • One 3.2 TB 2.5 Inch Enterprise Performance 12G SAS SSD
HDDs	Six to twelve 6 TB or 8 TB SAS 7.2K rpm LFF HDD
Network	Cisco UCS VIC1387 VIC MLOM
Boot Device	One 240 GB M.2 form factor SATA SSD

HX240c-M5L Options	Hardware Required
microSD Card	One 32GB microSD card for local host utilities storage
Optional	Cisco QSA module to convert 40 GbE QSFP+ to 10 GbE SFP+

Table 7 lists the hardware component options for the HXAF220c-M4S server model:

Table 7 HXAF220c-M4S Server Options

HXAF220c-M4S options		Hardware Required
Processors		Chose a matching pair of Intel E5-2600 v4 CPUs
Memory		128 GB to 1.5 TB of total memory using 16 GB, 32 GB, or 64 GB DDR4 2400 MHz 1.2v modules, or 64 GB DDR4 2133 MHz 1.2v modules
Disk Controller		Cisco 12Gbps Modular SAS HBA
SSDs	Standard	• One 120 GB 2.5 Inch Enterprise Value 6G SATA SSD, or one 240 GB 2.5 Inch Enterprise Value 6G SATA SSD • One 400 GB 2.5 Inch Enterprise Performance 12G SAS SSD, or one 800 GB 2.5 Inch Enterprise Performance 12G SAS SSD, or one 400 GB 2.5 Inch Enterprise Performance NVMe SSD • Six 3.8 TB 2.5 Inch Enterprise Value 6G SATA SSDs, or six 960 GB 2.5 Inch Enterprise Value 6G SATA SSDs
	SED	• One 120 GB 2.5 Inch Enterprise Value 6G SATA SSD, or one 240 GB 2.5 Inch Enterprise Value 6G SATA SSD • One 800 GB 2.5 Inch Enterprise Performance 12G SAS SED SSDs • Six 3.8 TB 2.5 Inch Enterprise Value 6G SATA SED SSDs, or six 960 GB 2.5 Inch Enterprise Value 6G SATA SED SSDs, or six 800 GB 2.5 Inch Enterprise Performance 12G SAS SED SSDs
Network		Cisco UCS VIC1227 VIC MLOM, or Cisco UCS VIC1387 VIC MLOM
Boot Devices		Two 64GB Cisco FlexFlash SD Cards for UCS Servers

Table 8 lists the hardware component options for the HXAF240c-M4SX server model:

Table 8 HXAF240c-M4S Server Options

HXAF240c-M4SX Options		Hardware Required
Processors		Chose a matching pair of Intel E5-2600 v4 CPUs
Memory		128 GB to 1.5 TB of total memory using 16 GB, 32 GB, or 64 GB DDR4 2400 MHz 1.2v modules, or 64 GB DDR4 2133 MHz 1.2v modules
Disk Controller		Cisco 12Gbps Modular SAS HBA
SSD	Standard	• One 120 GB 2.5 Inch Enterprise Value 6G SATA SSD, or one 240 GB 2.5 Inch Enterprise Value 6G SATA SSD (in the rear disk enclosure) • One 400 GB 2.5 Inch Enterprise Performance 12G SAS SSD, or one 800 GB 2.5 Inch Enterprise Performance 12G SAS SSD, or one 400 GB 2.5 Inch Enterprise Performance NVMe SSD • Six to twenty-three 3.8 TB 2.5 Inch Enterprise Value 6G SATA SSDs, or six to twenty-three 960 GB 2.5 Inch Enterprise Value 6G SATA SSDs
	SED	• One 120 GB 2.5 Inch Enterprise Value 6G SATA SSD, or one 240 GB 2.5 Inch Enterprise Value 6G SATA SSD (in a front disk bay) • One 800 GB 2.5 Inch Enterprise Performance 12G SAS SED SSDs • Six to twenty-two 3.8 TB 2.5 Inch Enterprise Value 6G SATA SED SSDs, or six to twenty-two 960 GB 2.5 Inch Enterprise Value 6G SATA SED SSDs, or six to twenty-two 800 GB 2.5 Inch Enterprise Performance 12G SAS SED SSDs
Network		Cisco UCS VIC1227 VIC MLOM, or Cisco UCS VIC1387 VIC MLOM
Boot Devices		Two 64GB Cisco FlexFlash SD Cards for UCS Servers

Table 9 lists the hardware component options for the HX220c-M4S server model:

Table 9 HX220c-M4S Server Options

HX220c-M4S Options	Hardware Required
Processors	Chose a matching pair of Intel E5-2600 v4 CPUs

HX220c-M4S Options		Hardware Required
Memory		128 GB to 1.5 TB of total memory using 16 GB, 32 GB, or 64 GB DDR4 2400 MHz 1.2v modules, or 64 GB DDR4 2133 MHz 1.2v modules
Disk Controller		Cisco 12Gbps Modular SAS HBA
SSDs	Standard	• One 120 GB 2.5 Inch Enterprise Value 6G SATA SSD, or one 240 GB 2.5 Inch Enterprise Value 6G SATA SSD • One 480 GB 2.5 Inch Enterprise Performance 6G SATA SSD
	SED	• One 120 GB 2.5 Inch Enterprise Value 6G SATA SSD, or one 240 GB 2.5 Inch Enterprise Value 6G SATA SSD • One 800 GB 2.5 Inch Enterprise Performance 12G SAS SED SSDs
HDDs	Standard	Six 1.8 TB or 1.2 TB SAS 12Gbps 10K rpm SFF HDD
	SED	Six 1.2 TB SAS 12Gbps 10K rpm SFF SED HDD
Network		Cisco UCS VIC1227 VIC MLOM, or Cisco UCS VIC1387 VIC MLOM
Boot Devices		Two 64GB Cisco FlexFlash SD Cards for Cisco UCS Servers

Table 10 lists the hardware component options for the HX240c-M4SX server model:

Table 10 HX240c-M4SX Server Options

HX240c-M4SX Options		Hardware Required
Processors		Chose a matching pair of Intel E5-2600 v4 CPUs
Memory		128 GB to 1.5 TB of total memory using 16 GB, 32 GB, or 64 GB DDR4 2400 MHz 1.2v modules, or 64 GB DDR4 2133 MHz 1.2v modules
Disk Controller		Cisco 12Gbps Modular SAS HBA
SSDs	Standard	• One 120 GB 2.5 Inch Enterprise Value 6G SATA SSD, or one 240 GB 2.5 Inch Enterprise Value 6G SATA SSD (in the rear disk enclosure) • One 1.6 TB 2.5 Inch Enterprise Performance 6G SATA SSD

HX240c-M4SX Options		Hardware Required
	SED	• One 120 GB 2.5 Inch Enterprise Value 6G SATA SSD, or one 240 GB 2.5 Inch Enterprise Value 6G SATA SSD (in a front disk bay) • One 1.6 TB 2.5 Inch Enterprise Performance 12G SAS SED SSD
HDDs	Standard	Six to twenty-three 1.8 TB or 1.2 TB SAS 12Gbps 10K rpm SFF HDD
	SED	Six to twenty-two 1.2 TB SAS 12Gbps 10K rpm SFF SED HDD
Network		Cisco UCS VIC1227 VIC MLOM, or Cisco UCS VIC1387 VIC MLOM
Boot Devices		Two 64GB Cisco FlexFlash SD Cards for Cisco UCS Servers

Software Components

Table 11 lists the software components and the versions required for the Cisco HyperFlex system:

Table 11 Software Components

Component	Software Required
Hypervisor	VMware ESXi 6.0 Update 3 or 6.5 Update 1 ESXi 6.5 U1 is recommended (CISCO Custom Image for ESXi 6.5 Update 1g: HX-Vmware-ESXi-6.5U1-7967591-Cisco-Custom-6.5.1.0.iso) **Note:** Use of a published Cisco custom ESXi ISO installer file is required when installing/reinstalling ESXi, or upgrading to a newer version prior to installing HyperFlex. An offline bundle file is also provided to upgrade ESXi on running clusters. **Note:** VMware vSphere Standard, Essentials Plus, ROBO, Enterprise or Enterprise Plus licensing is required from VMware.
Management Server	VMware vCenter Server for Windows or vCenter Server Appliance 6.0 U3c or later.

	Refer to http://www.vmware.com/resources/compatibility/sim/interop_matrix.php for interoperability of your ESXi version and vCenter Server. **Note:** Using ESXi 6.5 on the HyperFlex nodes also requires using vCenter Server 6.5.
Cisco HyperFlex Data Platform	Cisco HyperFlex HX Data Platform Software 3.0(1a)
Cisco HyperFlex Witness VM	HyperFlex-Witness-1.0.1.ova
Cisco UCS Firmware	Cisco UCS Infrastructure software, B-Series and C-Series bundles, revision 3.2(3a) or later. **Note:** Cisco UCS Firmware 3.2(3a) is the minimum version recommended for any cluster containing Cisco HX-series M5 generation servers. For UCS domains that have only Cisco HX-series M4 generation servers, Cisco UCS Firmware 3.1(3f) or later is required.

Licensing

Cisco HyperFlex systems must be properly licensed using Cisco Smart Licensing, which is a cloud-based software licensing management solution used to automate many manual, time consuming and error prone licensing tasks. Cisco HyperFlex 2.5 and later communicate with the Cisco Smart Software Manager (CSSM) online service via a Cisco Smart Account, to check out or assign available licenses from the account to the Cisco HyperFlex cluster resources. Communications can be direct via the internet, they can be configured to communicate via a proxy server, or they can communicate with an internal Cisco Smart Software Manager satellite server, which caches and periodically synchronizes licensing data. In a small number of highly secure environments, systems can be provisioned with a Permanent License Reservation (PLR) which does not need to communicate with CSSM. Contact your Cisco sales representative or partner to discuss if your security requirements will necessitate use of these permanent licenses. New HyperFlex cluster installations will operate for 90 days without licensing as an evaluation period, thereafter the system will generate alarms and operate in a non-compliant mode. Systems without compliant licensing will not be entitled to technical support.

For more information on the Cisco Smart Software Manager satellite server, visit this website: https://www.cisco.com/c/en/us/buy/smart-accounts/software-manager-satellite.html

Beginning with Cisco HyperFlex 3.0, licensing of the system requires one license per node from one of three different licensing editions; Edge licenses, Standard licenses, or Enterprise licenses. Depending on the type of cluster being installed, and the features you desire to activate and use in the system, you need to purchase licenses from the appropriate licensing tier. Additional features in the future will be added to the different licensing editions as they are released, the features listed below are current only as of the publication of this document.

Table 13 lists the licensing editions, and the features available with each type of license:

Table 12 HyperFlex System License Editions

HyperFlex Licensing Edition	Edge	Standard (in addition to Edge)	Enterprise (in addition to Standard)
Features Available	3 Node Edge deployments without Fabric Interconnects HX220c model servers only Hybrid or All-Flash Replication Factor 2 only 1 GbE Ethernet only Compression Deduplication HyperFlex native snapshots Rapid Clones Management via vCenter plugin, HyperFlex Connect, or Cisco Intersight	64 Node standard clusters with Fabric Interconnects (32 converged + 32 compute-only) All server models Replication Factor 3 Compute-only nodes 10 GbE or 40 GbE Ethernet HyperFlex native replication Data-at-rest encryption using self-encrypting disks Logical Availability Zones	16 node stretched clusters

Considerations

Version Control

The software revisions listed in Table 11 are the only valid and supported configuration at the time of the publishing of this validated design. Special care must be taken not to alter the revision of the hypervisor, vCenter server, Cisco HX platform software, or the Cisco UCS firmware without first consulting the appropriate release notes and compatibility matrixes to ensure that the system is not being modified into an unsupported configuration.

vCenter Server

The vCenter Server 6.0 Update 3c or later is required due to the requirement for TLS 1.2 with Cisco HyperFlex 3.0. The vCenter server must be installed and operational prior to the installation of the Cisco HyperFlex HX Data Platform software. The following best practice guidance applies to installations of HyperFlex 3.0:

- Do not modify the default TCP port settings of the vCenter installation. Using non-standard ports can lead to failures during the installation.
- It is recommended to build the vCenter server on a physical server or in a virtual environment outside of the HyperFlex cluster. Building the vCenter server as a virtual machine inside the HyperFlex cluster environment is highly discouraged. There is a tech note for multiple methods of deployment if no external vCenter server is already available: http://www.cisco.com/c/en/us/td/docs/hyperconverged_systems/HyperFlex_HX_DataPlatformSoftware/TechNotes/Nested_vcenter_on_hyperflex.html

Note: This document does not cover the installation and configuration of VMware vCenter Server for Windows, or the vCenter Server Appliance.

Scale

Cisco HyperFlex standard clusters currently scale from a minimum of 3 to a maximum of 32 Cisco HX-series converged nodes with small form factor (SFF) disks per cluster. A converged node is a member of the cluster which provides storage resources to the HX Distributed Filesystem. For the compute intensive "extended" cluster design, a configuration with 3 to 32 Cisco HX-series converged nodes can be combined with up to 32 compute nodes. It is required that the number of compute-only nodes should always be less than or equal to number of converged nodes. Once the maximum size of a single cluster has been reached, the environment can be "scaled out" by adding additional HX model servers to the Cisco UCS domain, installing an additional HyperFlex cluster on them, and controlling them via the same vCenter server. A maximum of 8 clusters can be created in a single UCS domain, and up to 100 HyperFlex clusters can be managed by a single vCenter server. When using Cisco Intersight for management and monitoring of Cisco HyperFlex clusters, there are no limits to the number of clusters being managed.

Cisco HyperFlex HX240c-M5L model servers with large form factor (LFF) disks are limited to a maximum of eight nodes per cluster, and cannot be mixed within the same cluster as models with small form factor (SFF) disks.

Cisco HyperFlex systems deployed in a stretched cluster configuration require a minimum of two Cisco HX-series converged nodes per physical site, and support a maximum of eight converged nodes per physical site. Each site requires a pair of Cisco UCS Fabric Interconnects, to form an individual UCS domain in both sites.

Note: At the time of the publication of this document, it is not supported to add compute-only nodes to a stretched cluster, or to expand an existing stretched cluster by adding more converged nodes.

Table 13 lists the minimum and maximum scale for various installations of the Cisco HyperFlex system:

Table 13 HyperFlex Cluster Scale

Cluster Type	Minimum Converged Nodes Required	Maximum Converged Nodes Allowed	Maximum Compute-only Nodes Allowed
Standard with SFF disks	3	32	32
Standard with LFF disks	3	8	8
Stretched	2 per site	8 per site	0

Capacity

Overall usable cluster capacity is based on a number of factors. The number of nodes in the cluster, the number and size of the capacity layer disks, and the replication factor of the HyperFlex HX Data Platform, all affect the cluster capacity. In addition, configuring a cluster as a stretched cluster across two sites modifies the data distribution method, which reduces capacity in favor of data availability. Caching disk sizes are not calculated as part of the cluster capacity.

Disk drive manufacturers have adopted a size reporting methodology using calculation by powers of 10, also known as decimal prefix. As an example, a 120 GB disk is listed with a minimum of 120 x 10^9 bytes of usable addressable capacity, or 120 billion bytes. However, many operating systems and filesystems report their space based on standard computer binary exponentiation, or calculation by powers of 2, also called binary prefix. In this example, 2^10 or 1024 bytes make up a kilobyte, 2^10 kilobytes make up a megabyte, 2^10 megabytes make up a gigabyte, and 2^10 gigabytes make up a terabyte. As the values increase, the disparity between the two systems of measurement and notation get worse, at the terabyte level, the deviation between a decimal prefix value and a binary prefix value is nearly 10 percent.

The International System of Units (SI) defines values and decimal prefix by powers of 10 as follows:

Table 14 SI Unit Values (Decimal Prefix)

Value	Symbol	Name
1000 bytes	kB	Kilobyte
1000 kB	MB	Megabyte
1000 MB	GB	Gigabyte
1000 GB	TB	Terabyte

The International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) defines values and binary prefix by powers of 2 in ISO/IEC 80000-13:2008 Clause 4 as follows:

Table 15 IEC unit values (binary prefix)

Value	Symbol	Name
1024 bytes	KiB	Kibibyte
1024 KiB	MiB	Mebibyte
1024 MiB	GiB	Gibibyte
1024 GiB	TiB	Tebibyte

For the purpose of this document, the decimal prefix numbers are used only for raw disk capacity as listed by the respective manufacturers. For all calculations where raw or usable capacities are shown from the perspective of the HyperFlex software, filesystems or operating systems, the binary prefix numbers are used. This is done primarily to show a consistent set of values as seen by the end user from within the HyperFlex vCenter Web Plugin and HyperFlex Connect GUI when viewing cluster capacity, allocation and consumption, and also within most operating systems.

Table 16 lists a set of HyperFlex HX Data Platform cluster usable capacity values, using binary prefix, for an array of cluster configurations. These values provide an example of the capacity calculations, for determining the appropriate size of HX cluster to initially purchase, and how much capacity can be gained by adding capacity disks. The calculations for these values are listed in Appendix A: Cluster Capacity Calculations. The HyperFlex tool to help with sizing is listed in Appendix B: HyperFlex Sizer.

Table 16 Cluster Usable Capacities

HX-Series Server Model	Node Quantity	Capacity Disk Size (each)	Capacity Disk Quantity (per node)	Cluster Usable Capacity at RF=2	Cluster Usable Capacity at RF=3
HXAF220c-M5SX	8	3.8 TB	8	102.8 TiB	68.6 TiB
		960 GB	8	25.7 TiB	17.1 TiB
		800 GB	8	21.4 TiB	14.3 TiB
HXAF240c-M5SX	8	3.8 TB	6	77.1 TiB	51.4 TiB
			15	192.8 TiB	128.5 TiB
			23	295.7 TiB	197.1 TiB
		960 GB	6	19.3 TiB	12.9 TiB

HX-Series Server Model	Node Quantity	Capacity Disk Size (each)	Capacity Disk Quantity (per node)	Cluster Usable Capacity at RF=2	Cluster Usable Capacity at RF=3
			15	48.2 TiB	32.1 TiB
			23	73.9 TiB	49.3 TiB
		800 GB	6	16.1 TiB	10.7 TiB
			15	40.2 TiB	26.8 TiB
			22	58.9 TiB	39.3 TiB
HX220c-M4S	8	1.2 TB	6	24.1 TiB	16.1 TiB
		1.8 TB	6	36.2 TiB	24.1 TiB
HX240c-M4SX	8	1.2 TB	6	24.1 TiB	16.1 TiB
			15	60.2 TiB	40.2 TiB
			23	92.4 TiB	61.6 TiB
		1.8 TB	6	36.2 TiB	24.1 TiB
			15	90.4 TiB	60.2 TiB
			23	138.6 TiB	92.4 TiB
HX240c-M5L	8	6 TB	6	120.5 TiB	80.3 TiB
			12	241.0 TiB	160.7 TiB
		8 TB	6	160.7 TiB	107.1 TiB
			12	321.3 TiB	214.2 TiB

Note: Capacity calculations methods for both M4 generation and M5 generation servers are identical, except for the large form factor HX240c-M5L, which is only available as an M5 generation model. Calculations will be based upon the number of nodes, the number of capacity disks per node, and the size of the capacity disks. The above table is not a comprehensive list of all capacities and models available.

Note: Stretched clusters keep two identical copies of all data in both sites, therefore the cluster capacities listed above for RF2 must be divided in half if the cluster is installed in a stretched configuration.

Physical Topology

Topology Overview

The Cisco HyperFlex system is composed of a pair of Cisco UCS Fabric Interconnects along with up to thirty-two HX-Series rack-mount servers per cluster. Up to thirty-two compute-only servers can also be added per HyperFlex cluster. Adding Cisco UCS rack-mount servers and/or Cisco UCS 5108 Blade chassis, which house Cisco UCS blade servers, allows for additional compute resources in an extended cluster design. Up to eight separate HX clusters can be installed under a single pair of Fabric Interconnects. The two Fabric Interconnects both connect to every HX-Series rack-mount server, and both connect to every Cisco UCS 5108 blade chassis, and Cisco UCS rack-mount server. Upstream network connections, also referred to as "northbound" network connections are made from the Fabric Interconnects to the customer datacenter network at the time of installation.

Figure 21 HyperFlex Standard Cluster Topology

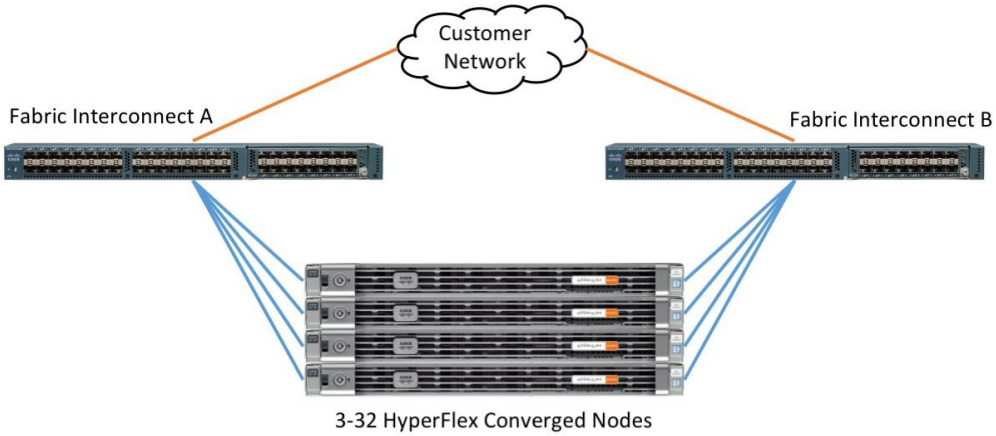
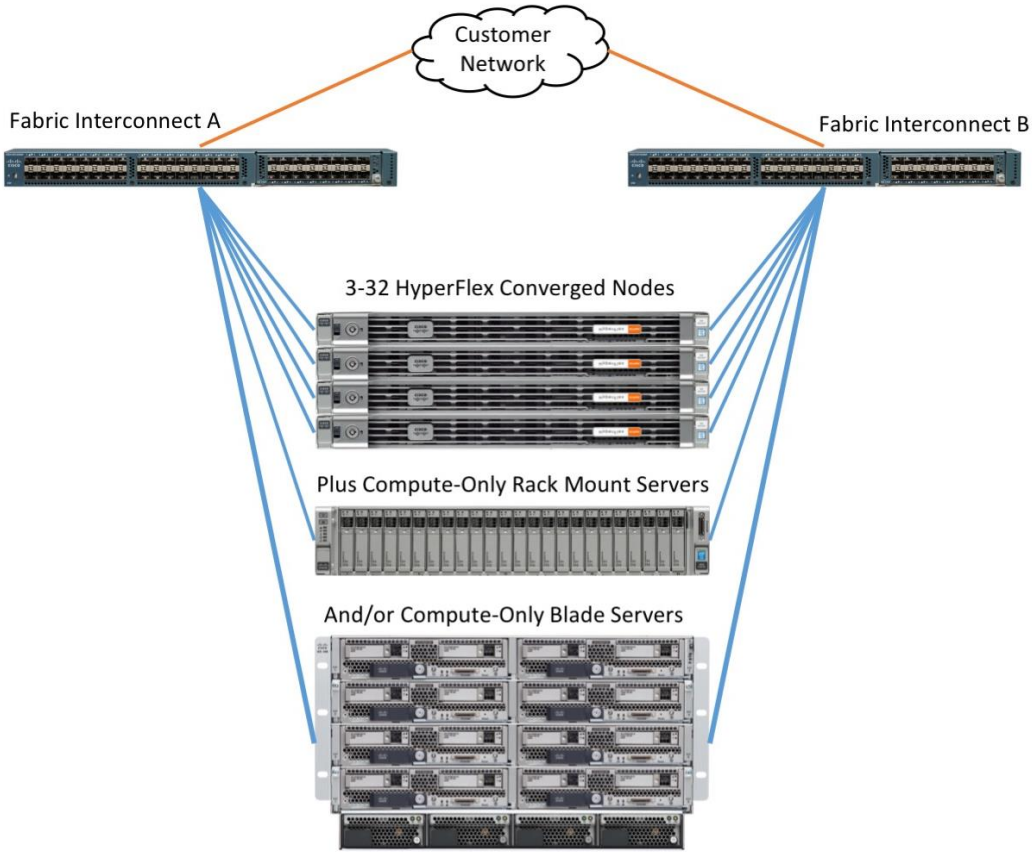


Figure 22 HyperFlex Extended Cluster Topology



Fabric Interconnects

Fabric Interconnects (FI) are deployed in pairs, wherein the two units operate as a management cluster, while forming two separate network fabrics, referred to as the A side and B side fabrics. Therefore, many design elements will refer to FI A or FI B, alternatively called fabric A or fabric B. Both Fabric Interconnects are active at all times, passing data on both network fabrics for a redundant and highly available configuration. Management services, including Cisco UCS Manager, are also provided by the two FIs but in a clustered manner, where one FI is the primary, and one is secondary, with a roaming clustered IP address. This primary/secondary relationship is only for the management cluster, and has no effect on data transmission.

Fabric Interconnects have the following ports, which must be connected for proper management of the Cisco UCS domain:

- **Mgmt:** A 10/100/1000 Mbps port for managing the Fabric Interconnect and the Cisco UCS domain via GUI and CLI tools. This port is also used by remote KVM, IPMI and SoL sessions to the managed servers within the domain. This is typically connected to the customer management network.
- **L1:** A cross connect port for forming the Cisco UCS management cluster. This port is connected directly to the L1 port of the paired Fabric Interconnect using a standard CAT5 or CAT6 Ethernet cable with RJ45 plugs. It is not necessary to connect this to a switch or hub.

- **L2:** A cross connect port for forming the Cisco UCS management cluster. This port is connected directly to the L2 port of the paired Fabric Interconnect using a standard CAT5 or CAT6 Ethernet cable with RJ45 plugs. It is not necessary to connect this to a switch or hub.
- **Console:** An RJ45 serial port for direct console access to the Fabric Interconnect. This port is typically used during the initial FI setup process with the included serial to RJ45 adapter cable. This can also be plugged into a terminal aggregator or remote console server device.

HX-Series Rack-Mount Servers

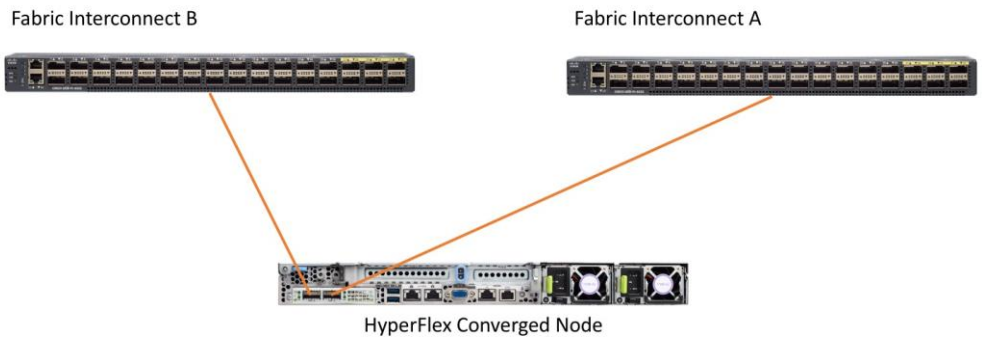
The HX-Series converged servers are connected directly to the Cisco UCS Fabric Interconnects in Direct Connect mode. This option enables Cisco UCS Manager to manage the HX-Series Rack-Mount Servers using a single cable for both management traffic and data traffic. All the Cisco HyperFlex M4 generation servers are configured with the Cisco VIC 1227 or Cisco VIC 1387 network interface card (NIC) installed in a modular LAN on motherboard (MLOM) slot, which has dual 10 Gigabit Ethernet (GbE) or 40 Gigabit Ethernet (GbE) ports. Cisco HyperFlex M5 generation servers can be configured only with the Cisco VIC 1387 card. The standard and redundant connection practice is to connect port 1 of the VIC card (the right-hand port) to a port on FI A, and port 2 of the VIC card (the left-hand port) to a port on FI B (0). The HyperFlex installer checks for this configuration, and that all servers' cabling matches. Failure to follow this cabling practice can lead to errors, discovery failures, and loss of redundant connectivity.

Various combinations of physical connectivity between the Cisco HX-series servers and the Fabric Interconnects are possible, but only specific combinations are supported. For example, use of the Cisco QSA module to convert a 40 GbE QSFP+ port into a 10 GbE SFP+ port is not allowed for use in M4 generation servers, but is allowed for M5 generation servers in order to configure a mixed cluster of M4 and M5 generation servers along with model 6248 or 6296 Fabric Interconnects. Table 17 lists the possible connections, and which of these methods is supported.

Table 17 Supported Physical Connectivity

Fabric Interconnect Model	6248	6296	6332		6332-16UP		
Port Type	**10GbE**	**10GbE**	**40GbE**	**10GbE Breakout**	**40GbE**	**10GbE Breakout**	**10GbE onboard**
M4 with VIC 1227	✓	✓	✕	✕	✕	✕	✕
M4 with VIC 1387	✕	✕	✓	✕	✓	✕	✕
M4 with VIC 1387 + QSA	✕	✕	✕	✕	✕	✕	✕
M5 with VIC 1387	✕	✕	✓	✕	✓	✕	✕
M5 with VIC 1387 + QSA	✓	✓	✕	✕	✕	✕	✕

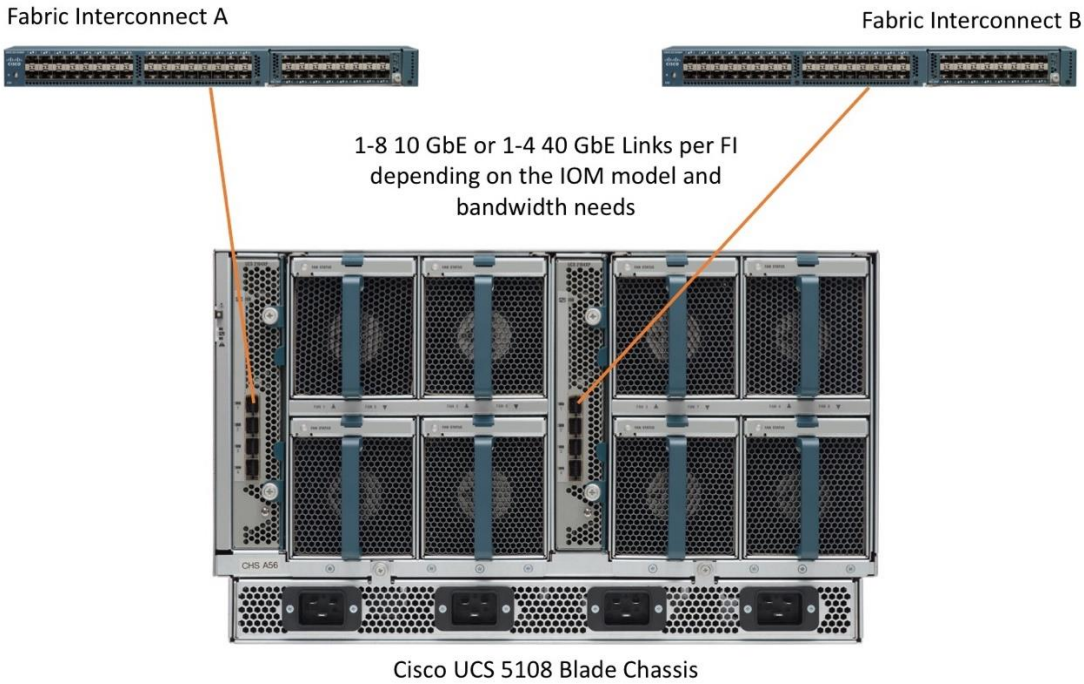
Figure 23 HX-Series Server Connectivity



Cisco UCS B-Series Blade Servers

HyperFlex extended clusters also incorporate 1-16 Cisco UCS blade servers for additional compute capacity. The blade chassis comes populated with 1-4 power supplies, and 8 modular cooling fans. In the rear of the chassis are two bays for installation of Cisco Fabric Extenders. The Fabric Extenders (also commonly called IO Modules, or IOMs) connect the chassis to the Fabric Interconnects. Internally, the Fabric Extenders connect to the Cisco VIC card installed in each blade server across the chassis backplane. The standard connection practice is to connect 1-8 10 GbE links, or 1-4 40 GbE links (depending on the IOMs and FIs purchased) from the left-side IOM, or IOM 1, to FI A, and to connect the same number of 10 GbE or 40 GbE links from the right-side IOM, or IOM 2, to FI B (Figure 24). All other cabling configurations are invalid, and can lead to errors, discovery failures, and loss of redundant connectivity.

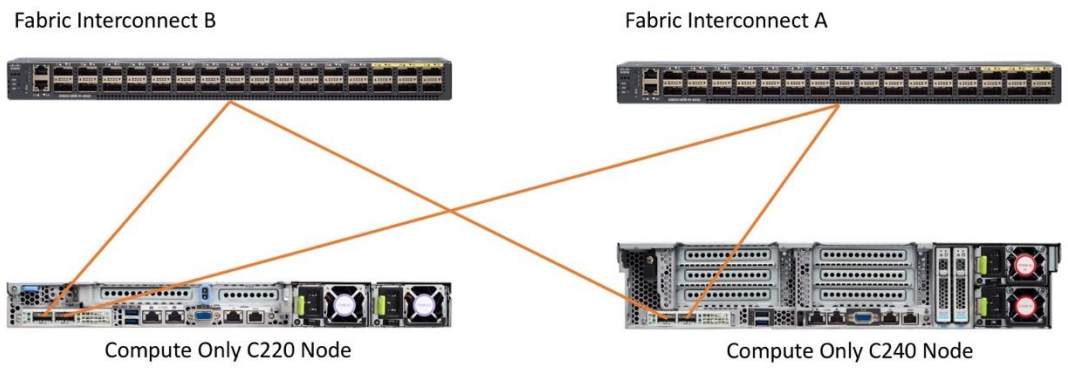
Figure 24 Cisco UCS 5108 Chassis Connectivity



Cisco UCS C-Series Rack-Mount Servers

HyperFlex extended clusters also incorporate 1-32 Cisco UCS Rack-Mount Servers for additional compute capacity. The Cisco UCS C-Series Rack-Mount Servers are connected directly to the Cisco UCS Fabric Interconnects in Direct Connect mode. Internally the Cisco UCS C-Series servers are configured with the Cisco VIC 1227 or Cisco VIC 1387 network interface card (NIC) installed in a modular LAN on motherboard (MLOM) slot, which has dual 10 Gigabit Ethernet (GbE) ports or 40 Gigabit Ethernet (GbE) ports. The standard and redundant connection practice is to connect port 1 of the VIC card to a port on FI A, and port 2 of the VIC card to a port on FI B. Failure to follow this cabling practice can lead to errors, discovery failures, and loss of redundant connectivity.

Figure 25 Cisco UCS C-Series Server Connectivity



Stretched Clusters

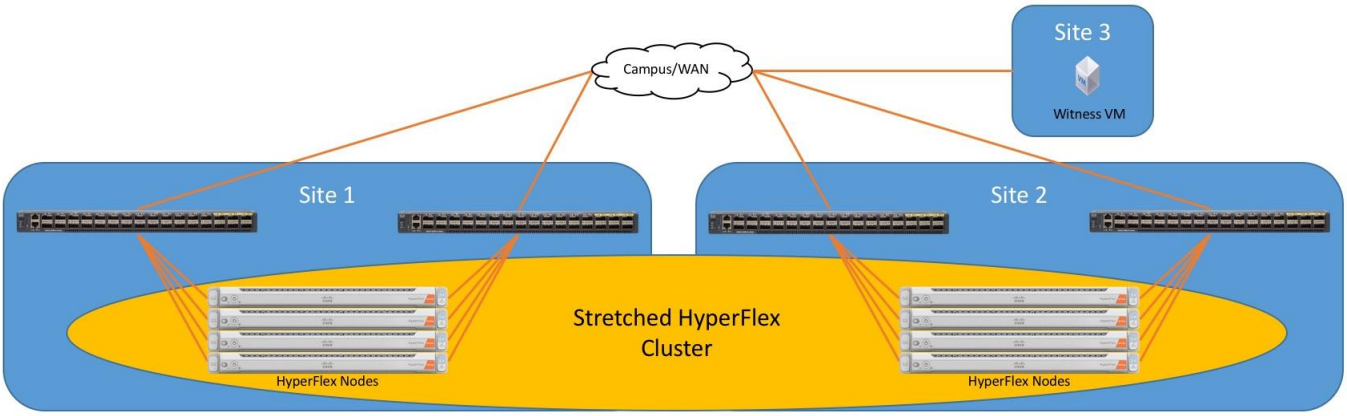
Stretched clusters are a new installation option in Cisco HyperFlex 3.0. Stretched clusters physically locate half of the cluster nodes in one location, while the remaining half are located in a distant secondary location. The data written to the stretched HyperFlex cluster is stored concurrently in both physical locations, therefore this system design provides for additional data protection and availability because the data in the cluster remains available and the cluster remains online, even if an entire site experiences a failure. Since all data written to the cluster is written simultaneously to the two sites, this design can be considered an active/active disaster recovery design. The recovery point objective (RPO, or the maximum amount of data that can potentially be lost prior to a failure) is essentially zero, due to all data being immediately and continuously available in both sites. The recovery time objective (RTO, or amount of time needed to return services to their functional state after a failure) is also near zero from an architectural standpoint, since all data and services are immediately available in the site which remains online, however actual recovery involves restarting the guest virtual machines and that does take some time.

For maximum safety and availability, it is also required to run a “witness” virtual machine as part of a HyperFlex stretched cluster. In any type of clustering or cooperative system design with two halves, there is a risk of a “split brain” scenario, wherein the two sides of the system lose communication with one another, but the system has not actually failed. In this circumstance, if either side has no secondary method besides network communication for determining if the other side has actually failed, in the interest of maximizing system availability it will decide that it is now the authoritative or active half. In truth, both sides would make the same determination, therefore both sides would think they are active and in charge of the overall system, hence the term “split brain”. An example of the issue would be both halves of the HyperFlex stretched cluster trying to run the same VMs in both sites. Recovery from a “split brain” condition can be difficult and time consuming. In order to avoid this scenario, a third system is required to break this tie, or provide additional information and decision-making logic to prevent simultaneous takeover by the two sides. Ideally,

this witness VM is physically separated from the two halves of the stretched cluster, with a reliable network connection to both of the other two sites.

Figure 26 illustrates the stretched cluster design.

Figure 26 Stretched Cluster Design



Stretched Cluster Requirements and Limitations

Installation of HyperFlex stretched clusters are subject to the following requirements and limitations:

- Stretched clusters can only be created during a new cluster installation, existing clusters cannot be converted into stretched clusters by process of expansion or moving nodes.
- Stretched clusters can only be created using M5 generation Cisco HyperFlex HX-series servers with small form factor disks. The HX240c-M5L model server cannot be used in a stretched cluster.
- Stretched clusters can only be installed using the VMware ESXi hypervisor.
- Server models in both sites must match.
- Stretched clusters can make use of hybrid nodes, or all-flash nodes.
- Stretched cluster nodes cannot make use of self-encrypting disks (SEDs).
- A minimum of 2 nodes per site is required, up to a maximum of 8 nodes per site. The number of servers in each site must match.
- vSphere Enterprise Plus licensing is required for the stretched cluster nodes, due to the use of advanced DRS rules necessary for the proper operation of the cluster.
- Stretched clusters cannot be expanded once built, either with additional converged nodes, or with compute-only nodes.
- Stretched clusters cannot participate in HyperFlex native replication, either as a source or as a target.
- A pair of Cisco Fabric Interconnects is required in each site, and the Ethernet speeds of all four models must match. For example, one site cannot use 40 GbE connections to the Cisco HyperFlex converged nodes, while the other site uses 10 GbE connections.

- The only allowed data replication factor setting is RF2, however two copies of all data are stored in both halves of the cluster, effectively creating a cluster running the equivalent of RF4.

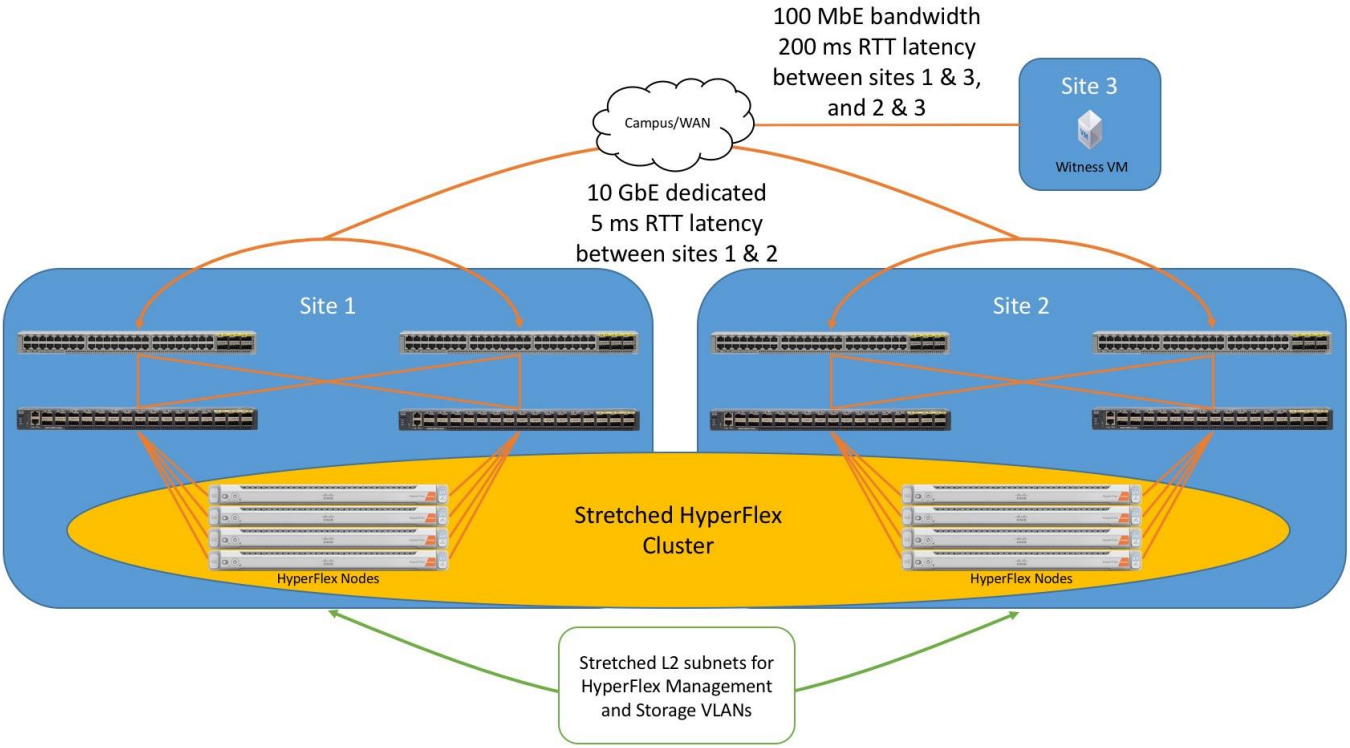
Stretched Cluster Networking

In order to implement stretched clusters, specific network bandwidth and latency requirements must be met. In addition, the configuration of the WAN must present multiple IP layer 2 (L2) subnets, with the same VLAN IDs in both sites, to the Cisco UCS Fabric Interconnects.

- A dedicated 10 GbE full-duplex link is required between the two sites, the link must carry no other traffic. The link bandwidth requirement is as measured from the uplink network switches that are connected to the pair of Cisco Fabric Interconnects in each of the two sites.
- The network latency, as measured from the uplink network switches that are connected to the pair of Cisco Fabric Interconnects in each of the two sites, must not exceed 5 milliseconds (ms) round trip time (RTT).
- The network must provide a minimum of 100 Mbps of bandwidth as measured from the uplink network switches that are connected to the pair of Cisco Fabric Interconnects in each of the two sites, to the witness virtual machine in the third site.
- The network latency, as measured from the uplink network switches that are connected to the pair of Cisco Fabric Interconnects in each of the two sites, must not exceed 200 milliseconds (ms) round trip time (RTT) to the witness virtual machine in the third site.
- The IP layer 2 subnet and VLAN ID used for HyperFlex management network traffic must be present in both sites with the active HyperFlex nodes, by use of various methods of stretched L2 technologies. Layer 3 routing is not allowed.
- The IP layer 2 subnet and VLAN ID used for HyperFlex storage network traffic must be present in both sites with the active HyperFlex nodes, by use of various methods of stretched L2 technologies. Layer 3 routing is not allowed. This is a fundamental change from standard cluster deployments, where the HyperFlex storage network was not required to traverse the network beyond the switches immediately upstream of the Cisco UCS Fabric Interconnects.
- The HyperFlex management network must be able to reach the IP address of the witness server VM, and this can be accessible via a layer 3 routable network connection.
- The stretched L2 subnet for HyperFlex storage traffic must be configured to allow jumbo frame transmission in order to achieve the best performance of the cluster. If jumbo frames cannot be supported, the cluster must be installed without jumbo frames enabled during the cluster configuration step.

0 illustrates the stretched cluster networking requirements.

Figure 27 Stretched Cluster Networking Requirements



Logical Topology

Logical Network Design

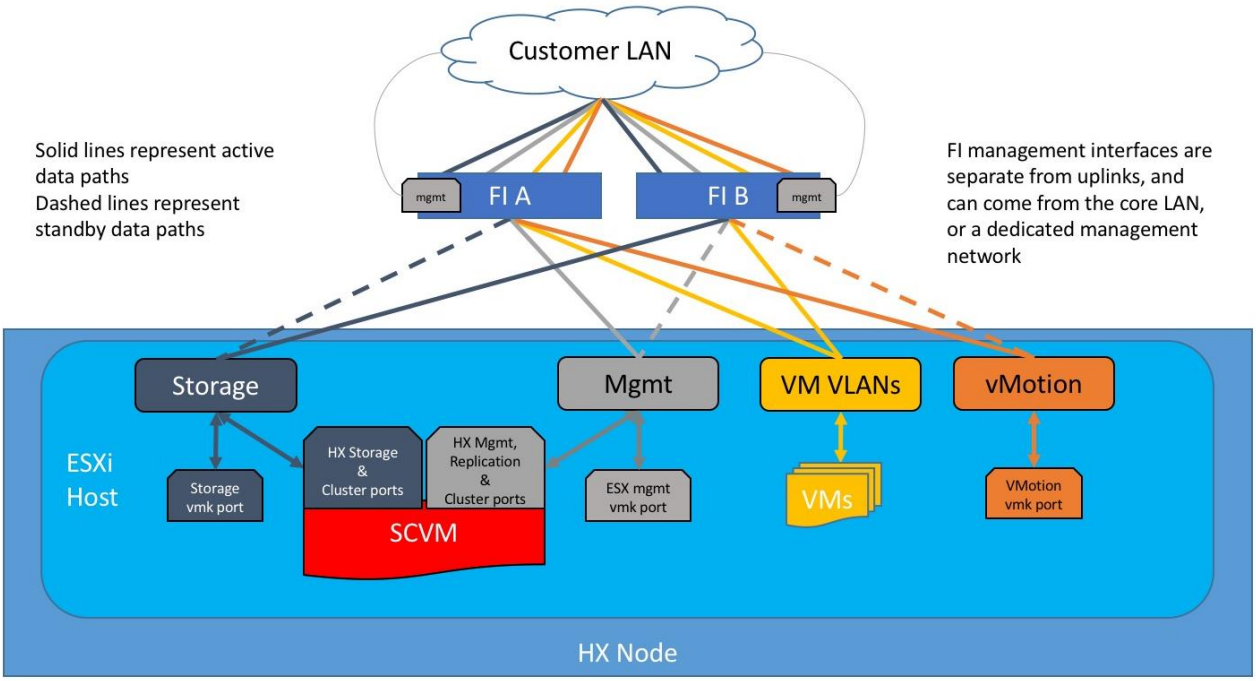
The Cisco HyperFlex system has communication pathways that fall into four defined zones (Figure 28):

- **Management Zone:** This zone comprises the connections needed to manage the physical hardware, the hypervisor hosts, and the storage platform controller virtual machines (SCVM). These interfaces and IP addresses need to be available to all staff who will administer the HX system, throughout the LAN/WAN. This zone must provide access to Domain Name System (DNS) and Network Time Protocol (NTP) services, and allow Secure Shell (SSH) communication. In this zone are multiple physical and virtual components:
  - Fabric Interconnect management ports.
  - Cisco UCS external management interfaces used by the servers and blades, which answer via the FI management ports.
  - ESXi host management interfaces.
  - Storage Controller VM management interfaces.
  - A roaming HX cluster management interface.
  - Storage Controller VM replication interfaces.
  - A roaming HX cluster replication interface.

- **VM Zone:** This zone comprises the connections needed to service network IO to the guest VMs that will run inside the HyperFlex hyperconverged system. This zone typically contains multiple VLANs, that are trunked to the Cisco UCS Fabric Interconnects via the network uplinks, and tagged with 802.1Q VLAN IDs. These interfaces and IP addresses need to be available to all staff and other computer endpoints which need to communicate with the guest VMs in the HX system, throughout the LAN/WAN.
- **Storage Zone:** This zone comprises the connections used by the Cisco HX Data Platform software, ESXi hosts, and the storage controller VMs to service the HX Distributed Data Filesystem. These interfaces and IP addresses need to be able to communicate with each other at all times for proper operation. During normal operation, this traffic all occurs within the Cisco UCS domain, however there are hardware failure scenarios where this traffic would need to traverse the network northbound of the Cisco UCS domain. For that reason, the VLAN used for HX storage traffic must be able to traverse the network uplinks from the Cisco UCS domain, reaching FI A from FI B, and vice-versa. This zone is primarily jumbo frame traffic therefore jumbo frames must be enabled on the Cisco UCS uplinks. In this zone are multiple components:
  - A VMkernel interface used for storage traffic on each ESXi host in the HX cluster.
  - Storage Controller VM storage interfaces.
  - A roaming HX cluster storage interface.
- **VMotion Zone:** This zone comprises the connections used by the ESXi hosts to enable vMotion of the guest VMs from host to host. During normal operation, this traffic all occurs within the Cisco UCS domain, however there are hardware failure scenarios where this traffic would need to traverse the network northbound of the Cisco UCS domain. For that reason, the VLAN used for HX vMotion traffic must be able to traverse the network uplinks from the Cisco UCS domain, reaching FI A from FI B, and vice-versa.

0 illustrates the logical network design.

Figure 28 Logical Network Design



Logical Availability Zones

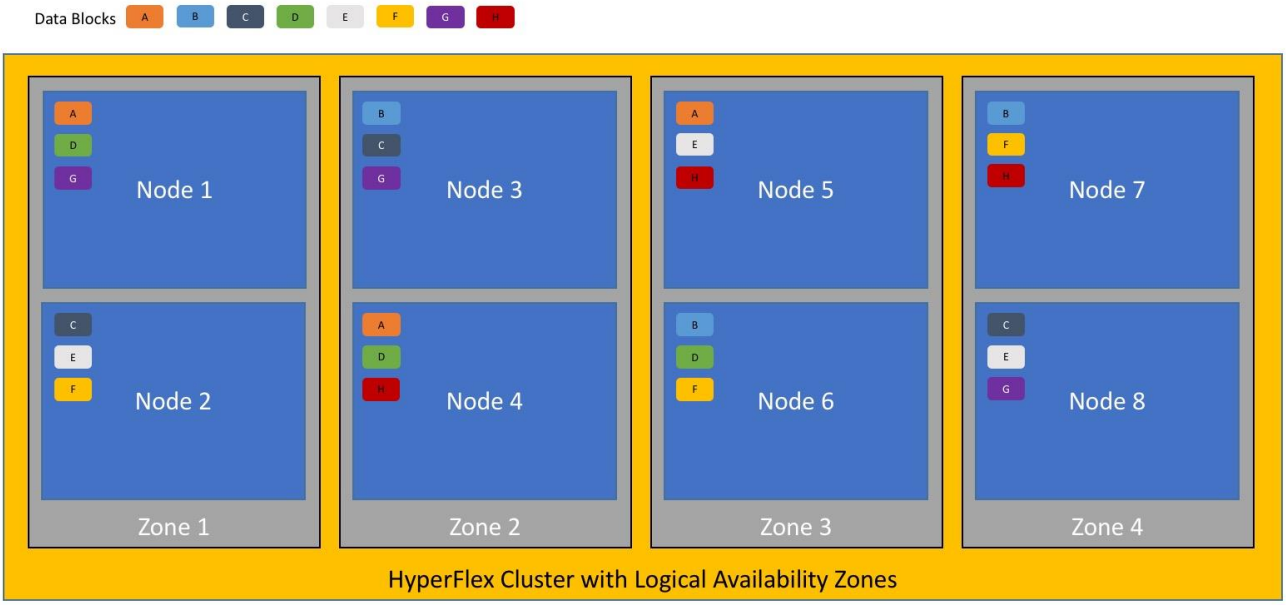
Larger scale HyperFlex clusters are subject to higher failure risks, simply due to the number of nodes in the cluster. While any individual node's risk of failure is the same no matter how many nodes there are, with clusters up to 64 nodes in size, there is a logically higher probability that a single node could fail, when compared to a cluster with fewer nodes. To mitigate these risks in larger scale clusters, a HyperFlex cluster of eight nodes or more, can be configured with a feature called Logical Availability Zones (LAZ). The Logical Availability Zones feature groups 2 or more HyperFlex nodes together into a logically defined zone, a minimum of 3 zones are created, and the data in the cluster is distributed in such a way that no blocks are written to the nodes within a single zone more than once. Due to this enhanced distribution pattern of data across zones, wherein each zone has multiple servers, clusters with LAZ enabled can typically withstand more failures than clusters that operate without this feature enabled. The number of failures that can tolerated varies depending on the number of zones in the cluster, and the number of servers in each of the zones. Generally speaking, multiple node failures across one or two zones will be tolerated better, and with less risk than multiple nodes failing across three or more zones. Note that the failure tolerance shown in the HyperFlex Connect dashboard will always present a "worst case scenario" view, meaning that even though the dashboard may state that two failures can be tolerated, in fact two servers could fail and the cluster can remain online, and the failure tolerance may still remain at two.

Logical availability zones should not be confused with the concept of fault domains. An example of a fault domain would be a subset of the nodes in a single HyperFlex cluster being powered by one uninterruptable power supply (UPS) or connected to one power distribution unit (PDU), meanwhile the remaining nodes would be connected to another UPS or PDU. If one of the UPS' or PDUs were to fail, then there would be a simultaneous failure of multiple nodes. While LAZ may actually prevent the cluster from failing in this scenario, to guarantee it would require that the zone membership be manually controlled, so that a failure of all of the servers protected by a single UPS or PDU, would be distributed in such a way that it would not cause an outage. The LAZ feature is not designed to be manually configured in this way, instead the zone membership is determined automatically by the system. If a HyperFlex cluster needs to be physically split in

half due to a physical limitation, such as the UPS example above, or a distance requirement for fault tolerance, then the cluster should be built as a stretched cluster instead of using LAZ.

Figure 29 illustrates an example of the data distribution method for clusters with Logical Availability Zones enabled, set to replication factor 3, where each zone only contains one of the three copies of the data in the cluster. This cluster consists of eight nodes, which the system configures into four zones.

Figure 29 Logical Availability Zone Data Distribution



Logical availability zones are subject to the following requirements and limitations:

- Only HyperFlex clusters with 8 nodes or more can be configured with logical availability zones during the installation process.
- Logical Availability Zones can be enabled during the HyperFlex cluster installation, or it can be enabled via the command line at a later time. It is recommended to enable this feature during installation, in order to avoid a large migration and reorganization of data across the cluster, which would be necessary to comply with the data distribution rules if LAZ is turned on in a cluster already containing data.
- The number of zones can be manually specified as 3, 4, 5, or you can allow the installer to automatically choose, which is the recommended setting.
- The HyperFlex cluster determines which nodes participate in each zone, and this configuration cannot be modified.
- To maintain the most balanced consumption of space and data distribution, it is recommended that the number of nodes in a cluster are whole multiples of 3, 4, 5, or 7. For example, 8 nodes would evenly divide into 4 zones of 2 servers each, and 9 nodes would divide evenly into 3 zones of 3 servers each. Eleven nodes would create an unbalanced number of nodes across the zones, leading to unbalanced space consumption on the nodes.
- In addition to the previous point, expansion of a cluster should be done in multiples of the number of zones, when the cluster is operating with LAZ enabled. Expanding in such a way preserves a matched number of nodes in each zone, and prevents any unbalance of space consumption. For example, a

cluster with 3 zones should be expanded by adding 3 more nodes, because adding only 1 or 2 nodes would lead to an imbalance, as would adding 4 nodes.

Design Elements

Installing the HyperFlex system is primarily done through a deployable HyperFlex installer virtual machine, available for download at cisco.com as an OVA file. The installer VM performs most of the Cisco UCS configuration work, it can be leveraged to simplify the installation of ESXi on the HyperFlex hosts, and also performs significant portions of the ESXi configuration. Finally, the installer VM is used to install the HyperFlex HX Data Platform software and create the HyperFlex cluster. Because this simplified installation method has been developed by Cisco, this CVD will not give detailed manual steps for the configuration of all the elements that are handled by the installer. Instead, the elements configured will be described and documented in this section, and the subsequent sections will guide you through the manual steps needed for installation, and how to utilize the HyperFlex Installer for the remaining configuration steps.

Network Design

Cisco UCS Uplink Connectivity

Cisco UCS network uplinks connect "northbound" from the pair of Cisco UCS Fabric Interconnects to the LAN in the customer datacenter. All Cisco UCS uplinks operate as trunks, carrying multiple 802.1Q VLAN IDs across the uplinks. The default Cisco UCS behavior is to assume that all VLAN IDs defined in the Cisco UCS configuration are eligible to be trunked across all available uplinks.

Cisco UCS Fabric Interconnects appear on the network as a collection of endpoints versus another network switch. Internally, the Fabric Interconnects do not participate in spanning-tree protocol (STP) domains, and the Fabric Interconnects cannot form a network loop, as they are not connected to each other with a layer 2 Ethernet link. All link up/down decisions via STP will be made by the upstream root bridges.

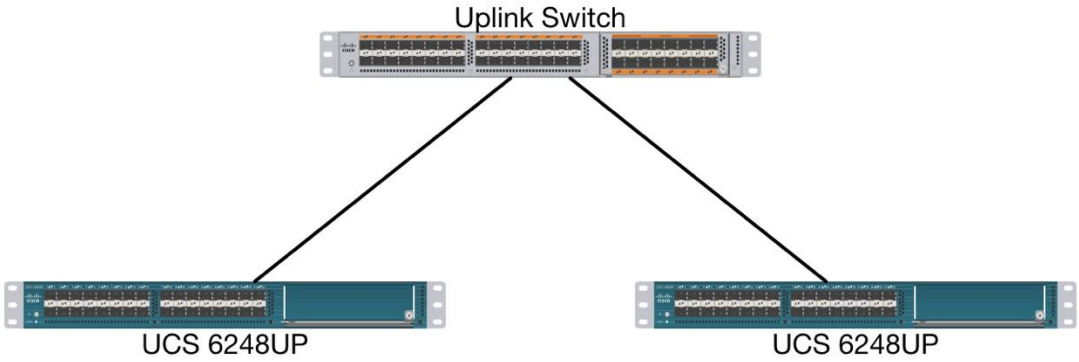
Uplinks need to be connected and active from both Fabric Interconnects. For redundancy, multiple uplinks can be used on each FI, either as 802.3ad Link Aggregation Control Protocol (LACP) port-channels, or using individual links. For the best level of performance and redundancy, uplinks can be made as LACP port-channels to multiple upstream Cisco switches using the virtual port channel (vPC) feature. Using vPC uplinks allows all uplinks to be active passing data, plus protects against any individual link failure, and the failure of an upstream switch. Other uplink configurations can be redundant, but spanning-tree protocol loop avoidance may disable links if vPC is not available.

All uplink connectivity methods must allow for traffic to pass from one Fabric Interconnect to the other, or from fabric A to fabric B. There are scenarios where cable, port or link failures would require traffic that normally does not leave the Cisco UCS domain, to now be forced over the Cisco UCS uplinks. Additionally, this traffic flow pattern can be seen briefly during maintenance procedures, such as updating firmware on the Fabric Interconnects, which requires them to be rebooted. The following sections and figures detail several uplink connectivity options.

Single Uplinks to Single Switch

This connection design is susceptible to failures at several points; single uplink failures on either Fabric Interconnect can lead to connectivity losses or functional failures, and the failure of the single uplink switch will cause a complete connectivity outage.

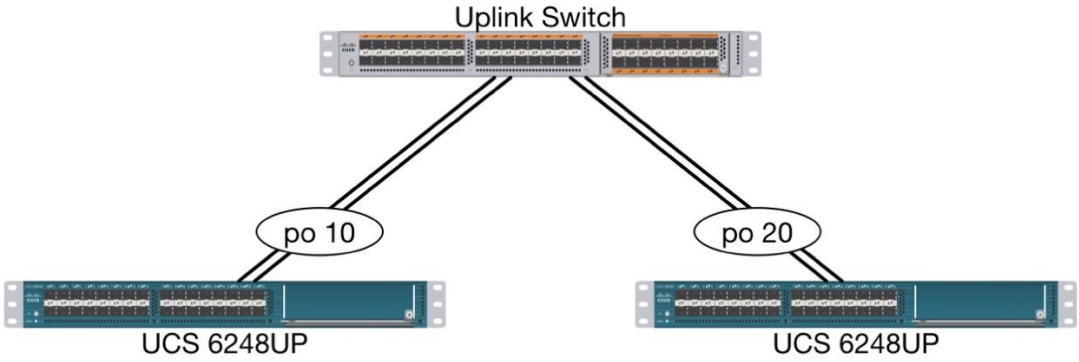
Figure 30 Connectivity with Single Uplink to Single Switch



Port Channels to Single Switch

This connection design is now redundant against the loss of a single link, but remains susceptible to the failure of the single switch.

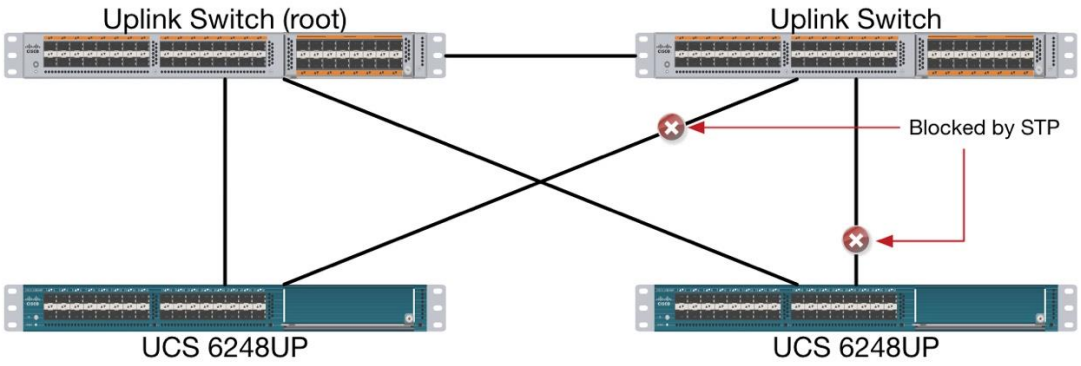
Figure 31 Connectivity with Port-Channels to Single Switch



Single Uplinks or Port Channels to Multiple Switches

This connection design is redundant against the failure of an upstream switch, and redundant against a single link failure. In normal operation, STP is likely to block half of the links to avoid a loop across the two upstream switches. The side effect of this is to reduce bandwidth between the Cisco UCS domain and the LAN. If any of the active links were to fail, STP would bring the previously blocked link online to provide access to that Fabric Interconnect via the other switch. It is not recommended to connect both links from a single FI to a single switch, as that configuration is susceptible to a single switch failure breaking connectivity from fabric A to fabric B. For enhanced redundancy, the single links in the figure below could also be port-channels.

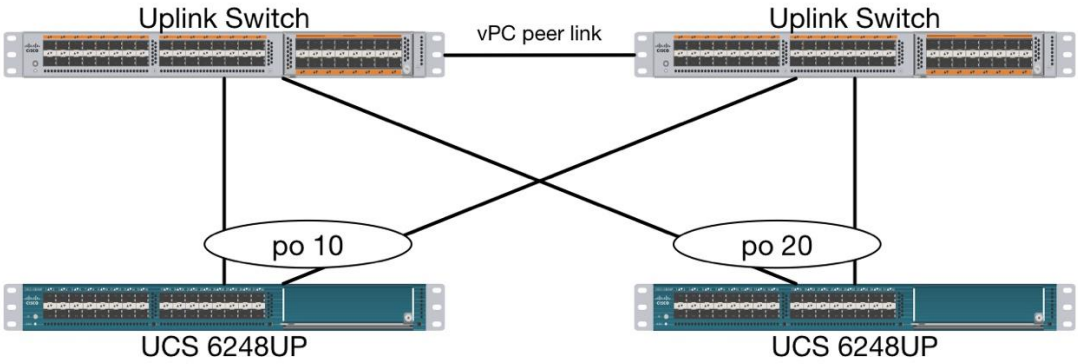
Figure 32 Connectivity with Multiple Uplink Switches



vPC to Multiple Switches

This recommended connection design relies on using Cisco switches that have the virtual port channel feature, such as Catalyst 6000 series switches running VSS, Cisco Nexus 5000 series, and Cisco Nexus 9000 series switches. Logically the two vPC enabled switches appear as one, and therefore spanning-tree protocol will not block any links. This configuration allows for all links to be active, achieving maximum bandwidth potential, and multiple redundancy at each level.

Figure 33 Connectivity with vPC



VLANs and Subnets

For the base HyperFlex system configuration, multiple VLANs need to be carried to the Cisco UCS domain from the upstream LAN, and these VLANs are also defined in the Cisco UCS configuration. The hx-storage-data VLAN must be a separate VLAN ID from the remaining VLANs. Table 18 lists the VLANs created by the HyperFlex installer in Cisco UCS, and their functions:

Table 18 VLANs

VLAN Name	VLAN ID	Purpose
hx-inband-mgmt	Customer supplied	ESXi host management interfaces HX Storage Controller VM management interfaces HX Storage Cluster roaming management interface
hx-inband-repl	Customer supplied	HX Storage Controller VM Replication interfaces HX Storage Cluster roaming replication interface
hx-storage-data	Customer supplied	ESXi host storage VMkernel interfaces HX Storage Controller storage network interfaces HX Storage Cluster roaming storage interface
vm-network	Customer supplied	Guest VM network interfaces
hx-vmotion	Customer supplied	ESXi host vMotion VMkernel interfaces

Note: A dedicated network or subnet for physical device management is often used in datacenters. In this scenario, the mgmto interfaces of the two Fabric Interconnects would be connected to that dedicated network or subnet. This is a valid configuration for HyperFlex installations with the following caveat; wherever the HyperFlex installer is deployed it must have IP connectivity to the subnet of the mgmto interfaces of the Fabric Interconnects, and also have IP connectivity to the subnets used by the hx-inband-mgmt VLANs listed above.

Jumbo Frames

All HyperFlex storage traffic traversing the hx-storage-data VLAN and subnet is configured by default to use jumbo frames, or to be precise, all communication is configured to send IP packets with a Maximum Transmission Unit (MTU) size of 9000 bytes. In addition, the default MTU for the hx-vmotion VLAN is also set to use jumbo frames. Using a larger MTU value means that each IP packet sent carries a larger payload, therefore transmitting more data per packet, and consequently sending and receiving data faster. This configuration also means that the Cisco UCS uplinks must be configured to pass jumbo frames. Failure to configure the Cisco UCS uplink switches to allow jumbo frames can lead to service interruptions during some failure scenarios, including Cisco UCS firmware upgrades, or when a cable or port failure would cause storage traffic to traverse the northbound Cisco UCS uplink switches.

HyperFlex clusters can be configured to use standard size frames of 1500 bytes, however Cisco recommends that this configuration only be used in environments where the Cisco UCS uplink switches are not capable of passing jumbo frames, and that jumbo frames be enabled in all other situations.

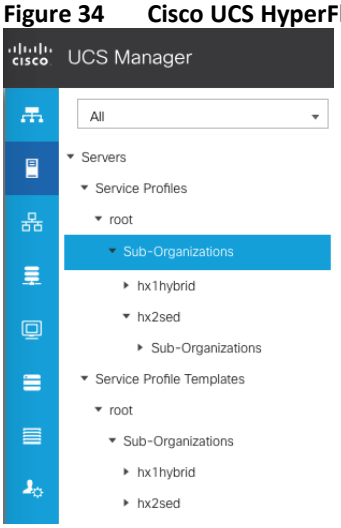
Cisco UCS Design

This section describes the elements within Cisco UCS Manager that are configured by the Cisco HyperFlex installer. Many of the configuration elements are fixed in nature, meanwhile the HyperFlex installer does allow for some items to be specified at the time of creation, for example VLAN names and IDs, external management IP pools and more. Where the elements can be manually set during the installation, those items will be noted in << >> brackets.

Cisco UCS Organization

During the HyperFlex installation a new Cisco UCS Sub-Organization is created. You must specify a unique Sub-Organization name for each cluster during the installation, for example “hx1hybrid”, or “hx2sed”. The sub-organization is created underneath the root level of the Cisco UCS hierarchy, and is used to contain all policies, pools, templates and service profiles used by HyperFlex, which prevents problems from overlapping settings across policies and pools. This arrangement also allows for organizational control using Role-Based Access Control (RBAC) and administrative locales within UCS Manager at a later time if desired. In this way, control can be granted to administrators of only the HyperFlex specific elements of the Cisco UCS domain, separate from control of root level elements or elements in other sub-organizations.

Figure 34 Cisco UCS HyperFlex Sub-Organization



Cisco UCS LAN Policies

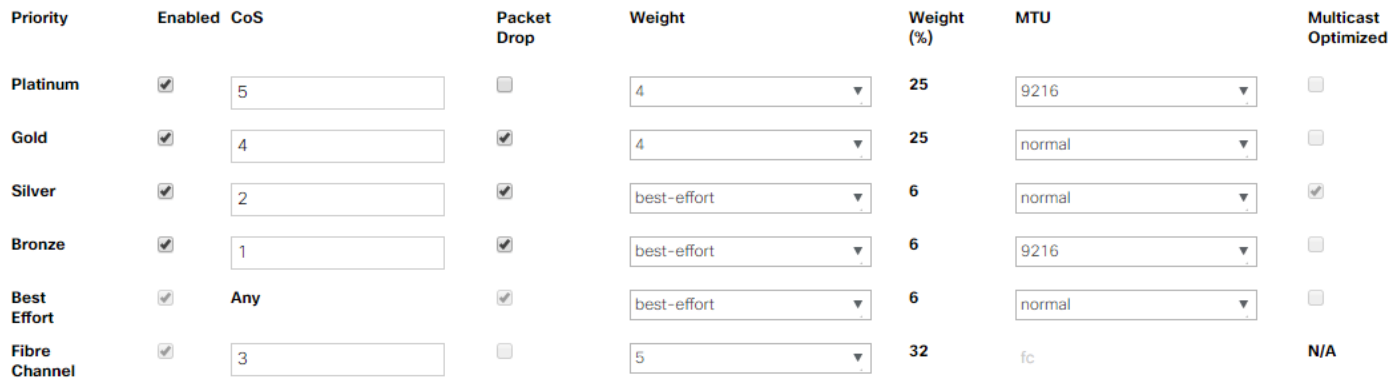
QoS System Classes

Specific Cisco UCS Quality of Service (QoS) system classes are defined for a Cisco HyperFlex system. These classes define Class of Service (CoS) values that can be used by the uplink switches north of the Cisco UCS domain, plus which classes are active, along with whether packet drop is allowed, the relative weight of the different classes when there is contention, the maximum transmission unit (MTU) size, and if there is multicast optimization applied. QoS system classes are defined for the entire Cisco UCS domain, the classes that are enabled can later be used in QoS policies, which are then assigned to Cisco UCS vNICs. The following table and figure details the QoS System Class settings configured for HyperFlex:

Table 19 QoS System Classes

Priority	Enabled	CoS	Packet Drop	Weight	MTU	Multicast Optimized
Platinum	Yes	5	No	4	9216	No
Gold	Yes	4	Yes	4	Normal	No
Silver	Yes	2	Yes	Best-effort	Normal	Yes
Bronze	Yes	1	Yes	Best-effort	9216	No
Best Effort	Yes	Any	Yes	Best-effort	Normal	No
Fibre Channel	Yes	3	No	5	FC	N/A

Figure 35 QoS System Classes



Note: Changing the QoS system classes on a Cisco UCS 6332 or 6332-16UP model Fabric Interconnect requires both FIs to reboot in order to take effect.

QoS Policies

In order to apply the settings defined in the Cisco UCS QoS System Classes, specific QoS Policies must be created, and then assigned to the vNICs, or vNIC templates used in Cisco UCS Service Profiles. Table 20 details the QoS Policies configured for HyperFlex, and their default assignment to the vNIC templates created:

Table 20 HyperFlex QoS Policies

Policy	Priority	Burst	Rate	Host Control	Used by vNIC Template
Platinum	Platinum	10240	Line-rate	None	storage-data-a storage-data-b

Policy	Priority	Burst	Rate	Host Control	Used by vNIC Template
Gold	Gold	10240	Line-rate	None	vm-network-a vm-network-b
Silver	Silver	10240	Line-rate	None	hv-mgmt-a hv-mgmt-b
Bronze	Bronze	10240	Line-rate	None	hv-vmotion-a hv-vmotion-b
Best Effort	Best Effort	10240	Line-rate	None	N/A

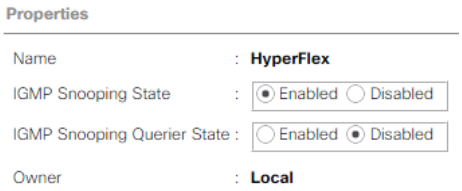
Multicast Policy

A Cisco UCS Multicast Policy is configured by the HyperFlex installer, which is referenced by the VLANs that are created. The policy allows for future flexibility if a specific multicast policy needs to be created and applied to other VLANs, that may be used by non-HyperFlex workloads in the Cisco UCS domain. Table 21 and Figure 36 details the Multicast Policy configured for HyperFlex:

Table 21 Multicast Policy

Name	IGMP Snooping State	IGMP Snooping Querier State
HyperFlex	Enabled	Disabled

Figure 36 Multicast Policy



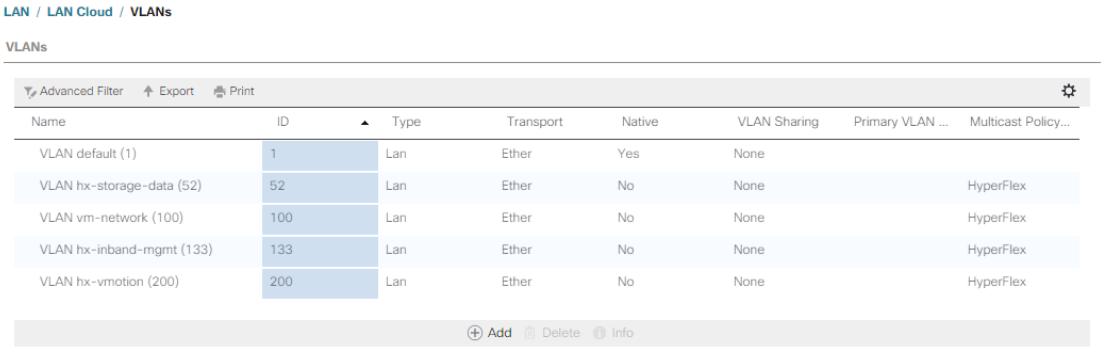
VLANs

VLANs are created by the HyperFlex installer to support a base HyperFlex system, with a VLAN for vMotion, and a single or multiple VLANs defined for guest VM traffic. Names and IDs for the VLANs are defined in the Cisco UCS configuration page of the HyperFlex installer web interface. The VLANs listed in Cisco UCS must already be present on the upstream network, and the Cisco UCS FIs do not participate in VLAN Trunk Protocol (VTP). Table 22 and Figure 37 details the VLANs configured for HyperFlex:

Table 22 Cisco UCS VLANs

Name	ID	Type	Transport	Native	VLAN Sharing	Multicast Policy
<<hx-inband-mgmt>>	<user_defined>	LAN	Ether	No	None	HyperFlex
<<hx-inband-repl>>	<user_defined>	LAN	Ether	No	None	HyperFlex
<<hx-storage-data>>	<user_defined>	LAN	Ether	No	None	HyperFlex
<<vm-network>>	<user_defined>	LAN	Ether	No	None	HyperFlex
<<hx-vmotion>>	<user_defined>	LAN	Ether	No	None	HyperFlex

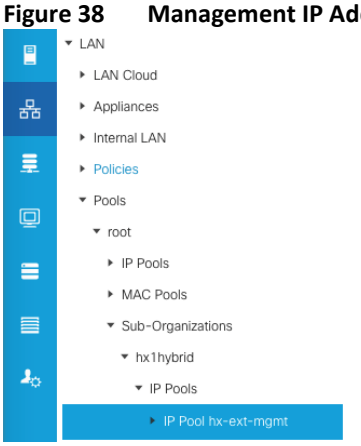
Figure 37 Cisco UCS VLANs



Management IP Address Pool

A Cisco UCS Management IP Address Pool must be populated with a block of IP addresses. These IP addresses are assigned to the Cisco Integrated Management Controller (CIMC) interface of the rack mount and blade servers that are managed in the Cisco UCS domain. The IP addresses are the communication endpoints for various functions, such as remote KVM, virtual media, Serial over LAN (SoL), and Intelligent Platform Management Interface (IPMI) for each rack mount or blade server. Therefore, a minimum of one IP address per physical server in the domain must be provided. The IP addresses are considered to be an "out-of-band" address, meaning that the communication pathway uses the Fabric Interconnects' mgmt0 ports, which answer ARP requests for the management addresses. Because of this arrangement, the IP addresses in this pool must be in the same IP subnet as the IP addresses assigned to the Fabric Interconnects' mgmt0 ports. A new IP pool, named "hx-ext-mgmt" is created in the HyperFlex sub-organization, and populated with a block of IP addresses, a subnet mask, and a default gateway by the HyperFlex installer. The default IP pool named "ext-mgmt", in the root organization is no longer used as of HyperFlex 2.5 for new installations.

Figure 38 Management IP Address Pool



MAC Address Pools

One of the core benefits of the Cisco UCS and Virtual Interface Card (VIC) technology is the assignment of the personality of the card via Cisco UCS Service Profiles. The number of virtual NIC (vNIC) interfaces, their VLAN associations, MAC addresses, QoS policies and more are all applied dynamically as part of the service profile association process. Media Access Control (MAC) addresses use 6 bytes of data as a unique address to identify the interface on the layer 2 network. All devices are assigned a unique MAC address, which is ultimately used for all data transmission and reception. The Cisco UCS and VIC technology picks a MAC address from a pool of addresses, and assigns it to each vNIC defined in the service profile when that service profile is created.

Best practices mandate that MAC addresses used for Cisco UCS domains use 00:25:B5 as the first three bytes, which is one of the Organizationally Unique Identifiers (OUI) registered to Cisco Systems, Inc. The fourth byte (e.g. 00:25:B5:xx) is specified during the HyperFlex installation. The fifth byte is set automatically by the HyperFlex installer, to correlate to the Cisco UCS fabric and the vNIC placement order. Finally, the last byte is incremented according to the number of MAC addresses created in the pool. To avoid overlaps, when you define the values in the HyperFlex installer you must ensure that the first four bytes of the MAC address pools are unique for each HyperFlex cluster installed in the same layer 2 network, and also different from MAC address pools in other Cisco UCS domains which may exist.

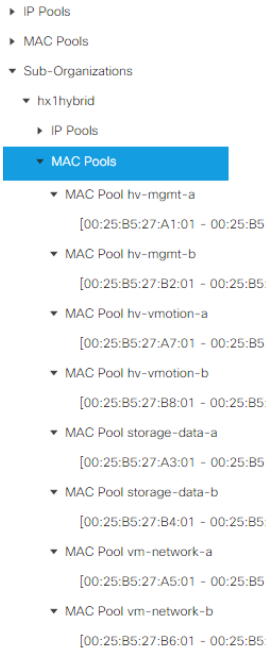
Table 23 details the MAC Address Pools configured for HyperFlex and their default assignment to the vNIC templates created:

Table 23 MAC Address Pools

Name	Block Start	Size	Assignment Order	Used by vNIC Template
hv-mgmt-a	00:25:B5:<xx>:A1:01	100	Sequential	hv-mgmt-a
hv-mgmt-b	00:25:B5:<xx>:B2:01	100	Sequential	hv-mgmt-b
hv-vmotion-a	00:25:B5:<xx>:A7:01	100	Sequential	hv-vmotion-a
hv-vmotion-b	00:25:B5:<xx>:B8:01	100	Sequential	hv-vmotion-b

Name	Block Start	Size	Assignment Order	Used by vNIC Template
storage-data-a	00:25:B5:<xx>:A3:01	100	Sequential	storage-data-a
storage-data-b	00:25:B5:<xx>:B4:01	100	Sequential	storage-data-b
vm-network-a	00:25:B5:<xx>:A5:01	100	Sequential	vm-network-a
vm-network-b	00:25:B5:<xx>:B6:01	100	Sequential	vm-network-b

Figure 39 MAC Address Pools



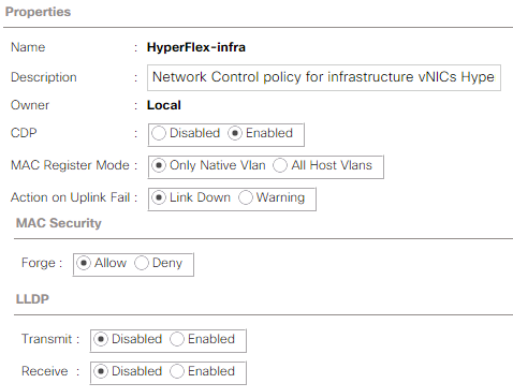
Network Control Policies

Cisco UCS Network Control Policies control various aspects of the behavior of vNICs defined in the Cisco UCS Service Profiles. Settings controlled include enablement of Cisco Discovery Protocol (CDP), MAC address registration, MAC address forging, and the action taken on the vNIC status if the Cisco UCS network uplinks are failed. Two policies are configured by the HyperFlex Installer, HyperFlex-infra is applied to the “infrastructure” vNIC interfaces of the HyperFlex system, and HyperFlex-vm, which is only applied to the vNIC interfaces carrying guest VM traffic. This allows for more flexibility, even though the policies are currently configured with the same settings. Table 24 details the Network Control Policies configured for HyperFlex, and their default assignment to the vNIC templates created:

Table 24 Network Control Policy

Name	CDP	MAC Register Mode	Action on Uplink Fail	MAC Security	Used by vNIC Template
HyperFlex-infra	Enabled	Only Native VLAN	Link-down	Forged: Allow	hv-mgmt-a hv-mgmt-b hv-vmotion-a hv-vmotion-b storage-data-a storage-data-b
HyperFlex-vm	Enabled	Only Native VLAN	Link-down	Forged: Allow	vm-network-a vm-network-b

Figure 40 Network Control Policy



vNIC Templates

Cisco UCS Manager has a feature to configure vNIC templates, which can be used to simplify and speed up configuration efforts. VNIC templates are referenced in service profiles and LAN connectivity policies, versus configuring the same vNICs individually in each service profile, or service profile template. VNIC templates contain all the configuration elements that make up a vNIC, including VLAN assignment, MAC address pool selection, fabric A or B assignment, fabric failover, MTU, QoS policy, Network Control Policy, and more. Templates are created as either initial templates, or updating templates. Updating templates retain a link between the parent template and the child object, therefore when changes are made to the template, the changes are propagated to all remaining linked child objects. An additional feature named “vNIC Redundancy” allows vNICs to be configured in pairs, so that the settings of one vNIC template, designated as a primary template, will automatically be applied to a configured secondary template. For all HyperFlex vNIC templates, the “A” side vNIC template is configured as a primary template, and the related “B” side vNIC template is a secondary. In each case, the only configuration difference between the two templates is which fabric they are configured to connect through. The following tables detail the initial settings in each of the vNIC templates created by the HyperFlex installer:

Table 25 vNIC Template hv-mgmt-a

vNIC Template Name:	hv-mgmt-a	
Setting	**Value**	
Fabric ID	A	
Fabric Failover	Disabled	
Target	Adapter	
Type	Updating Template	
MTU	1500	
MAC Pool	hv-mgmt-a	
QoS Policy	silver	
Network Control Policy	HyperFlex-infra	
VLANs	<<hx-inband-mgmt>>	Native: No

Table 26 vNIC Template hv-mgmt-b

vNIC Template Name:	hv-mgmt-b
Setting	**Value**
Fabric ID	B
Fabric Failover	Disabled
Target	Adapter
Type	Updating Template
MTU	1500
MAC Pool	hv-mgmt-b

QoS Policy	silver	
Network Control Policy	HyperFlex-infra	
VLANs	<<hx-inband-mgmt>>	Native: No

Table 27 vNIC Template hv-vmotion-a

vNIC Template Name:	**hv-vmotion-a**	
Setting	**Value**	
Fabric ID	A	
Fabric Failover	Disabled	
Target	Adapter	
Type	Updating Template	
MTU	9000	
MAC Pool	hv-vmotion-a	
QoS Policy	bronze	
Network Control Policy	HyperFlex-infra	
VLANs	<<hx-vmotion>>	Native: No

Table 28 vNIC Template hx-vmotion-b

vNIC Template Name:	**hv-vmotion-b**
Setting	**Value**
Fabric ID	B
Fabric Failover	Disabled

Target	Adapter	
Type	Updating Template	
MTU	9000	
MAC Pool	hv-vmotion-b	
QoS Policy	bronze	
Network Control Policy	HyperFlex-infra	
VLANs	<<hx-vmotion>>	Native: No

Table 29 vNIC Template storage-data-a

vNIC Template Name:	**storage-data-a**	
Setting	**Value**	
Fabric ID	A	
Fabric Failover	Disabled	
Target	Adapter	
Type	Updating Template	
MTU	9000	
MAC Pool	storage-data-a	
QoS Policy	platinum	
Network Control Policy	HyperFlex-infra	
VLANs	<<hx-storage-data>>	Native: No

Table 30 vNIC Template storage-data-b

vNIC Template Name:	storage-data-b	
Setting	**Value**	
Fabric ID	B	
Fabric Failover	Disabled	
Target	Adapter	
Type	Updating Template	
MTU	9000	
MAC Pool	storage-data-b	
QoS Policy	platinum	
Network Control Policy	HyperFlex-infra	
VLANs	<<hx-storage-data>>	Native: No

Table 31 vNIC Template vm-network-a

vNIC Template Name:	vm-network-a
Setting	**Value**
Fabric ID	A
Fabric Failover	Disabled
Target	Adapter
Type	Updating Template
MTU	1500
MAC Pool	vm-network-a

QoS Policy	gold	
Network Control Policy	HyperFlex-vm	
VLANs	<<vm-network>>	Native: no

Table 32 vNIC Template vm-network-b

vNIC Template Name:	**vm-network-b**	
Setting	**Value**	
Fabric ID	B	
Fabric Failover	Disabled	
Target	Adapter	
Type	Updating Template	
MTU	1500	
MAC Pool	vm-network-b	
QoS Policy	gold	
Network Control Policy	HyperFlex-vm	
VLANs	<<vm-network>>	Native: no

LAN Connectivity Policies

Cisco UCS Manager has a feature for LAN Connectivity Policies, which aggregates all of the vNICs or vNIC templates desired for a service profile configuration into a single policy definition. This simplifies configuration efforts by defining a collection of vNICs or vNIC templates once, and using that policy in the service profiles or service profile templates. The HyperFlex installer configures a LAN Connectivity Policy named HyperFlex, which contains all of the vNIC templates defined in the previous section, along with an Adapter Policy named HyperFlex, also configured by the HyperFlex installer. Table 33 details the LAN Connectivity Policy configured for HyperFlex:

Table 33 LAN Connectivity Policy

Policy Name	Use vNIC Template	vNIC Name	vNIC Template Used	Adapter Policy
HyperFlex	Yes	hv-mgmt-a	hv-mgmt-a	HyperFlex
		hv-mgmt-b	hv-mgmt-b	
		hv-vmotion-a	hv-vmotion-a	
		hv-vmotion-b	hv-vmotion-b	
		storage-data-a	storage-data-a	
		storage-data-b	storage-data-b	
		vm-network-a	vm-network-a	
		vm-network-b	vm-network-b	

Cisco UCS Servers Policies

Adapter Policies

Cisco UCS Adapter Policies are used to configure various settings of the Converged Network Adapter (CNA) installed in the Cisco UCS blade or rack-mount servers. Various advanced hardware features can be enabled or disabled depending on the software or operating system being used. The following figures detail the Adapter Policy named “HyperFlex”, configured for HyperFlex:

Figure 41 Cisco UCS Adapter Policy Resources

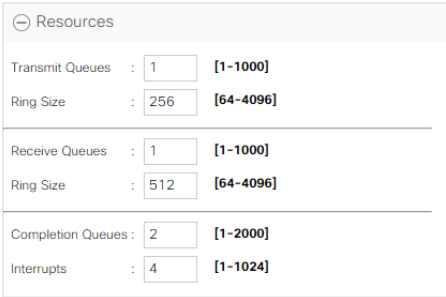
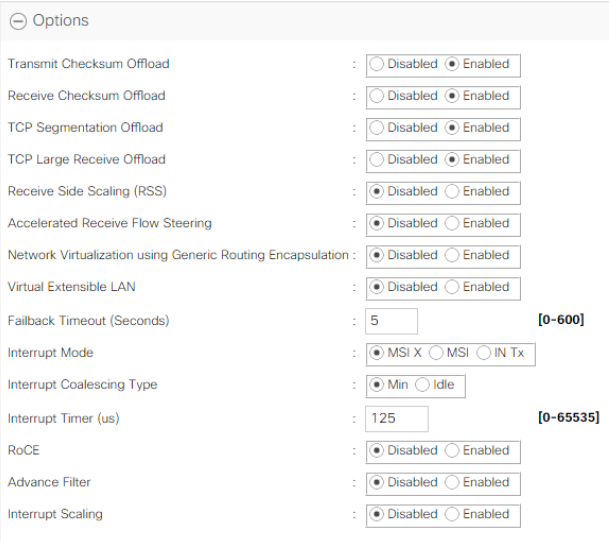


Figure 42 Cisco UCS Adapter Policy Options



BIOS Policies

Cisco HX-Series M4 generation servers have a set of pre-defined BIOS setting defaults defined in Cisco UCS Manager. These settings have been optimized for the Cisco HX-Series servers running HyperFlex. The HyperFlex installer creates a BIOS policy named “HyperFlex”, with all settings set to the defaults, except for enabling the Serial Port A for Serial over LAN (SoL) functionality. This policy allows for future flexibility in case situations arise where the settings need to be modified from the default configuration.

Cisco HX-Series M5 generation servers no longer use pre-defined BIOS setting defaults derived from Cisco UCS Manager, instead the servers have default BIOS tokens set from the factory. The current default token settings can be viewed at the following website: https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/ucs-manager/Reference-Docs/Server-BIOS-Tokens/3-2/b_UCS_BIOS_Tokens.html

A second BIOS policy, named “HyperFlex-m5” is created by the HyperFlex installer to modify the setting of M5 generation servers. The settings modified are as follows:

- System altitude is set to “Auto”
- CPU performance is set to “HPC”
- Processor C1E state is set to “Disabled”
- Power Technology is set to “Performance”

- Energy Performance is set to “Performance”
- Serial Port A is enabled
- Console Redirection is set to Serial Port A

Boot Policies

Cisco UCS Boot Policies define the boot devices used by blade and rack-mount servers, and the order that they are attempted to boot from. Cisco HX-Series M4 generation rack-mount servers have their VMware ESXi hypervisors installed to an internal pair of mirrored Cisco FlexFlash SD cards, therefore they require a boot policy defining that the servers should boot from that location. The HyperFlex installer configures a boot policy named “HyperFlex” specifying boot from the SD cards, which is used by the HyperFlex converged nodes, and should not be modified. The compute-only Cisco UCS blade servers and Cisco UCS Rack-Mount Servers can also boot from SD cards, or they can be configured to boot from local disks, boot from SAN, or via the network using PXE or iSCSI. The HyperFlex installer configures a boot policy named “hx-compute”, which can be modified as needed for the boot method used by the compute-only nodes.

Cisco HX-Series M5 generation rack-mount servers have their VMware ESXi hypervisors installed to an internal M.2 SSD boot drive, therefore they require a unique boot policy defining that the servers should boot from that location. The HyperFlex installer configures a boot policy named “HyperFlex-m5” specifying boot from the M.2 SSDs, referred to as “Embedded Disk”, which is used by the HyperFlex M5 converged nodes, and should not be modified. The HyperFlex installer configures a boot policy named “hx-compute-m5”, which can be modified as needed for the boot method used by the M5 generation compute-only nodes.

The following figures detail the HyperFlex Boot Policies:

Figure 43 Cisco UCS M4 Boot Policy

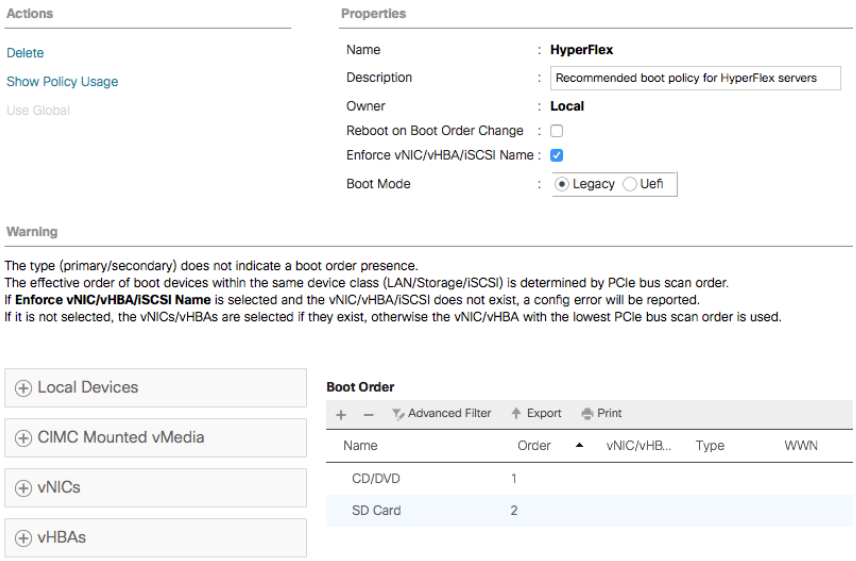
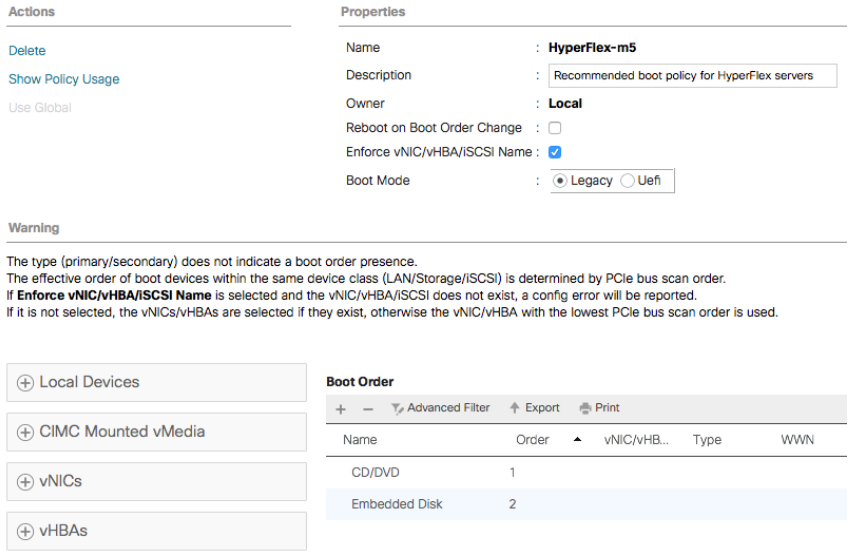


Figure 44 Cisco UCS M5 Boot Policy



Host Firmware Packages

Cisco UCS Host Firmware Packages represent one of the most powerful features of the Cisco UCS platform; the ability to control the firmware revision of all the managed blades and rack-mount servers via a policy specified in the service profile. Host Firmware Packages are defined and referenced in the service profiles. Once a service profile is associated to a server, the firmware of all the components defined in the Host Firmware Package are automatically upgraded or downgraded to match the package. The HyperFlex installer creates two Host Firmware Packages named “HyperFlex” and “HyperFlex-m5” which use the simple package definition method, applying firmware revisions to all components that matches a specific Cisco UCS firmware bundle, versus defining the firmware revisions part by part. The two policies are distinct to allow different generations of HX-Series rack-mount servers to run different firmware packages as needed. The following figures detail the Host Firmware Packages configured by the HyperFlex installer:

Figure 45 Cisco UCS M4 Host Firmware Package

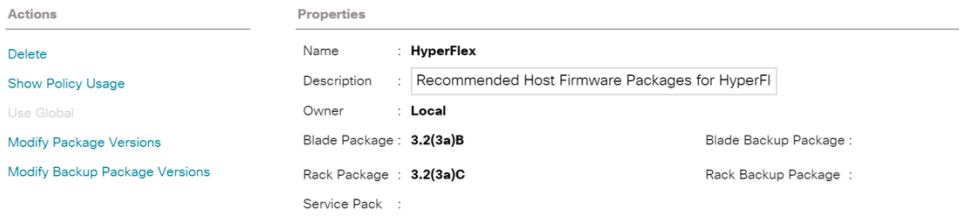
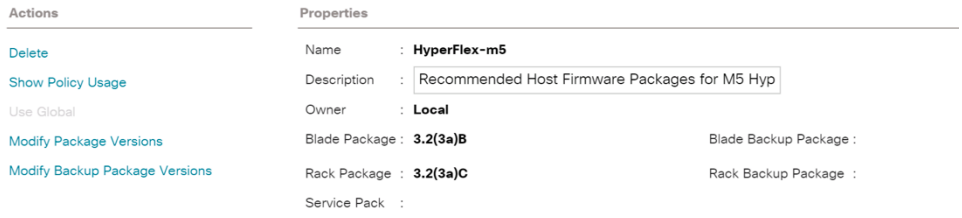


Figure 46 Cisco UCS M5 Host Firmware Package



Local Disk Configuration Policies

Cisco UCS Local Disk Configuration Policies are used to define the configuration of disks installed locally within each blade or rack-mount server, most often to configure Redundant Array of

Independent/Inexpensive Disks (RAID levels) when multiple disks are present for data protection. Since HX-Series converged nodes providing storage resources do not require RAID, the HyperFlex installer creates four Local Disk Configuration Policies, named “HyperFlex”, “HyperFlex-m5”, “hx-compute” and “hx-compute-m5”, all of which allows any local disk configuration. The policy named “HyperFlex” is used by the service profile template named “hx-nodes”, which is for the HyperFlex M4 generation converged servers, and should not be modified. The policy named “HyperFlex-m5” is used by the service profile template named “hx-nodes-m5”, which is for the HyperFlex M5 generation converged servers, and should not be modified. The difference between the two policies is the state of the FlexFlash controller, which is enabled in M4 generation servers to support the SD cards, but is disabled in the M5 generation servers.

Meanwhile, the policies named “hx-compute” and “hx-compute-m5” are used by the service profile templates named “compute-nodes” and “compute-nodes-m5”, which are used by compute-only nodes. The two compute-only node policies can be modified as needed to suit the local disk configuration that will be used in compute-only nodes.

The following figures detail the Local Disk Configuration Policies configured by the HyperFlex installer:

Figure 47 Cisco UCS M4 Local Disk Configuration Policy

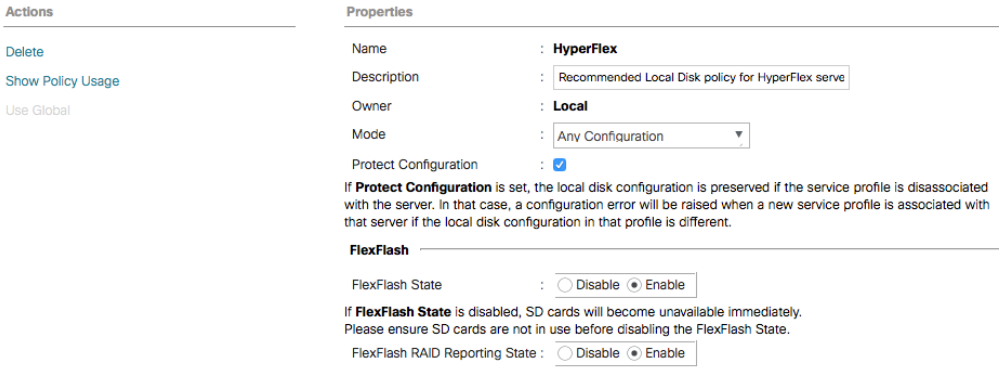
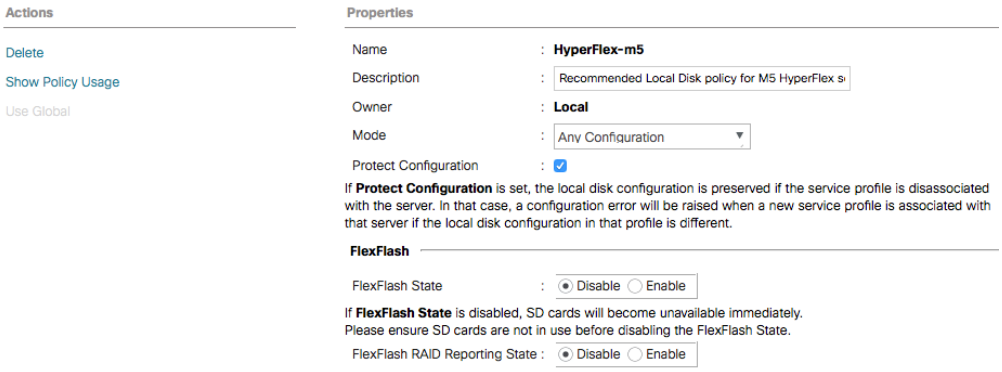


Figure 48 Cisco UCS M5 Local Disk Configuration Policy

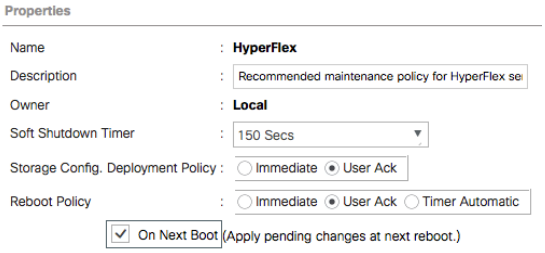


Maintenance Policies

Cisco UCS Maintenance Policies define the behavior of the attached blades and rack-mount servers when changes are made to the associated service profiles. The default Cisco UCS Maintenance Policy setting is “Immediate” meaning that any change to a service profile that requires a reboot of the physical server will result in an immediate reboot of that server. The Cisco best practice is to use a Maintenance Policy set to “user-ack”, which requires a secondary acknowledgement by a user with the appropriate rights within Cisco UCS Manager, before the server is rebooted to apply the changes. The HyperFlex installer creates a

Maintenance Policy named “HyperFlex” with the setting changed to “user-ack”. In addition, the On Next Boot setting is enabled, which will automatically apply changes the next time the server is rebooted, without any secondary acknowledgement. Figure 49 details the Maintenance Policy configured by the HyperFlex installer:

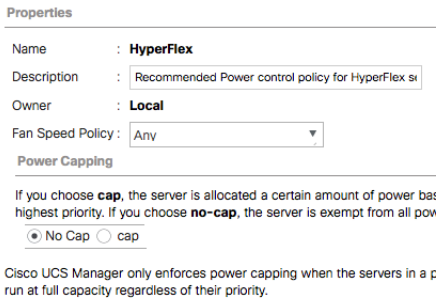
Figure 49 Cisco UCS Maintenance Policy



Power Control Policies

Cisco UCS Power Control Policies allow administrators to set priority values for power application to servers in environments where power supply may be limited, during times when the servers demand more power than is available. The HyperFlex installer creates a Power Control Policy named “HyperFlex” with all power capping disabled, and fans allowed to run at full speed when necessary. Figure 50 details the Power Control Policy configured by the HyperFlex installer:

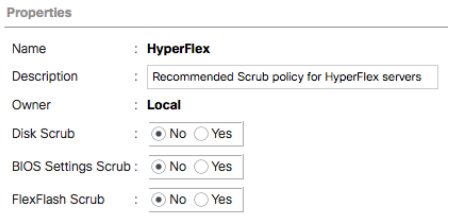
Figure 50 Cisco UCS Power Control Policy



Scrub Policies

Cisco UCS Scrub Policies are used to scrub, or erase data from local disks, BIOS settings and FlexFlash SD cards. If the policy settings are enabled, the information is wiped when the service profile using the policy is disassociated from the server. The HyperFlex installer creates a Scrub Policy named “HyperFlex” which has all settings disabled, therefore all data on local disks, SD cards and BIOS settings will be preserved if a service profile is disassociated. Figure 51 details the Scrub Policy configured by the HyperFlex installer:

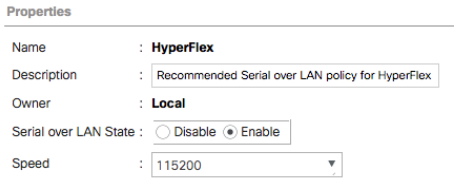
Figure 51 Cisco UCS Scrub Policy



Serial over LAN Policies

Cisco UCS Serial over LAN (SoL) Policies enable console output which is sent to the serial port of the server, to be accessible via the LAN. For many Linux based operating systems, such as VMware ESXi, the local serial port can be configured as a local console, where users can watch the system boot, and communicate with the system command prompt interactively. Since many blade servers do not have physical serial ports, and often administrators are working remotely, the ability to send and receive that traffic via the LAN is very helpful. Connections to a SoL session can be initiated from Cisco UCS Manager. The HyperFlex installer creates a SoL named "HyperFlex" to enable SoL sessions, and uses this feature to configure the ESXi hosts' management networking configuration. Figure 52 details the SoL Policy configured by the HyperFlex installer:

Figure 52 Cisco UCS Serial over LAN Policy



vMedia Policies

Cisco UCS Virtual Media (vMedia) Policies automate the connection of virtual media files to the remote KVM session of the Cisco UCS blades and rack-mount servers. Using a vMedia policy can speed up installation time by automatically attaching an installation ISO file to the server, without having to manually launch the remote KVM console and connect them one-by-one. The HyperFlex installer creates a vMedia Policy named "HyperFlex" for future use, with no media locations defined.

Cisco UCS Service Profile Templates

Cisco UCS Manager has a feature to configure service profile templates, which can be used to simplify and speed up configuration efforts when the same configuration needs to be applied to multiple servers. Service profile templates are used to spawn multiple service profile copies to associate with a group of servers, versus configuring the same service profile manually each time it is needed. Service profile templates contain all the configuration elements that make up a service profile, including vNICs, vHBAs, local disk configurations, boot policies, host firmware packages, BIOS policies and more. Templates are created as either initial templates, or updating templates. Updating templates retain a link between the parent template and the child object, therefore when changes are made to the template, the changes are propagated to all remaining linked child objects. The HyperFlex installer creates four service profile templates, named "hx-nodes", "hx-nodes-m5", "compute-nodes" and "compute-nodes-m5", each with nearly the same configuration, except for the BIOS, firmware, local disk configuration and boot policies. This simplifies future efforts if the configuration of the compute only nodes needs to differ from the configuration of the HyperFlex converged storage nodes. The following tables detail the service profile templates configured by the HyperFlex installer:

Table 34 Cisco UCS Service Profile Template Settings and Values

Service Profile Template Name:	**hx-nodes**
Setting	**Value**
UUID Pool	Hardware Default
Associated Server Pool	None
Maintenance Policy	HyperFlex
Management IP Address Policy	hx-ext-mgmt
Local Disk Configuration Policy	HyperFlex
LAN Connectivity Policy	HyperFlex
Boot Policy	HyperFlex
BIOS Policy	HyperFlex
Firmware Policy	HyperFlex
Power Control Policy	HyperFlex
Scrub Policy	HyperFlex
Serial over LAN Policy	HyperFlex
vMedia Policy	Not defined

Service Profile Template Name:	**hx-nodes-m5**
Setting	**Value**
UUID Pool	Hardware Default
Associated Server Pool	None

Service Profile Template Name:	**hx-nodes-m5**
Setting	**Value**
Maintenance Policy	HyperFlex
Management IP Address Policy	hx-ext-mgmt
Local Disk Configuration Policy	HyperFlex-m5
LAN Connectivity Policy	HyperFlex
Boot Policy	HyperFlex-m5
BIOS Policy	HyperFlex-m5
Firmware Policy	HyperFlex-m5
Power Control Policy	HyperFlex
Scrub Policy	HyperFlex
Serial over LAN Policy	HyperFlex
vMedia Policy	Not defined

Service Profile Template Name:	**compute-nodes**
Setting	**Value**
UUID Pool	Hardware Default
Associated Server Pool	None
Maintenance Policy	HyperFlex
Management IP Address Policy	hx-ext-mgmt
Local Disk Configuration Policy	hx-compute

LAN Connectivity Policy	HyperFlex
Boot Policy	hx-compute
BIOS Policy	HyperFlex
Firmware Policy	HyperFlex
Power Control Policy	HyperFlex
Scrub Policy	HyperFlex
Serial over LAN Policy	HyperFlex
vMedia Policy	Not defined

Service Profile Template Name:	**compute-nodes-m5**
Setting	**Value**
UUID Pool	Hardware Default
Associated Server Pool	None
Maintenance Policy	HyperFlex
Management IP Address Policy	hx-ext-mgmt
Local Disk Configuration Policy	hx-compute-m5
LAN Connectivity Policy	HyperFlex
Boot Policy	hx-compute-m5
BIOS Policy	HyperFlex-m5
Firmware Policy	HyperFlex-m5
Power Control Policy	HyperFlex

Scrub Policy	HyperFlex
Serial over LAN Policy	HyperFlex
vMedia Policy	Not defined

vNIC/vHBA Placement

In order to control the order of detection of the vNICs and vHBAs defined in service profiles, Cisco UCS allows for the definition of the placement of the vNICs and vHBAs across the cards in a blade or rack-mount server, and the order they are seen. In certain hardware configurations, the physical mapping of the installed cards and port extenders to their logical order is not linear, therefore each card is referred to as a virtual connection, or vCon. Because of this, the placement and detection order of the defined vNICs and vHBAs does not refer to physical cards, but instead refers to vCons. Since HX-series servers are configured with a single Cisco UCS VIC mLOM card, the only valid placement is on card number 1, or vCon 1. Therefore, all the vNICs defined in the service profile templates for HX-series servers places them on vCon 1, then their detection order is set.

Through the combination of the vNIC templates created (vNIC Templates), the LAN Connectivity Policy (LAN Connectivity Policies), and the vNIC placement, every VMware ESXi server will detect the same network interfaces in a known and identical order, and they will always be connected to the same VLANs via the same network fabrics. The following table outlines the vNICs, their placement, their order, the fabric they are connected to, their default VLAN, and how they are enumerated by the ESXi hypervisor:

Table 35 vNIC Placement

vNIC	Placement	Order	Fabric	VLAN	ESXi interface enumeration
hv-mgmt-a	1	1	A	<<hx-inband-mgmt>>	vmnic0
hv-mgmt-b	1	2	B	<<hx-inband-mgmt>>	vmnic1
storage-data-a	1	3	A	<<hx-storage-data>>	vmnic2
storage-data-b	1	4	B	<<hx-storage-data>>	vmnic3
vm-network-a	1	5	A	<<vm-network>>	vmnic4
vm-network-b	1	6	B	<<vm-network>>	vmnic5
hv-vmotion-a	1	7	A	<<hx-vmotion>>	vmnic6
hv-vmotion-b	1	8	B	<<hx-vmotion>>	vmnic7

Figure 53 vNIC Placement

vNICs

Filter Export Print

Name	MAC Address	Desired Order	Actual Order	Fabric ID	Desired Placement	Actual Placement	Admin Host Port	Actual Host Port
vNIC hv-mgmt-a	Derived	1	Unspecified	A	1	Any	1	NONE
vNIC hv-mgmt-b	Derived	2	Unspecified	B	1	Any	1	NONE
vNIC hv-vmotion-a	Derived	7	Unspecified	A	1	Any	2	NONE
vNIC hv-vmotion-b	Derived	8	Unspecified	B	1	Any	2	NONE
vNIC storage-data-a	Derived	3	Unspecified	A	1	Any	1	NONE
vNIC storage-data-b	Derived	4	Unspecified	B	1	Any	1	NONE
vNIC vm-network-a	Derived	5	Unspecified	A	1	Any	2	NONE
vNIC vm-network-b	Derived	6	Unspecified	B	1	Any	2	NONE

Note: ESXi VMDirectPath relies on a fixed PCI address for the passthrough devices. If the configuration is changed by adding or removing vNICs or vHBAs, then the order of the devices seen in the PCI tree will change. The ESXi hosts will subsequently need to reboot one additional time in order to repair the configuration, which they will do automatically.

ESXi Host Design

The following sections detail the design of the elements within the VMware ESXi hypervisors, system requirements, virtual networking and the configuration of ESXi for the Cisco HyperFlex HX Distributed Data Platform.

Virtual Networking Design

The Cisco HyperFlex system has a pre-defined virtual network design at the ESXi hypervisor level. Four different virtual switches are created by the HyperFlex installer, each using two uplinks, which are each serviced by a vNIC defined in the Cisco UCS service profile. The vSwitches created are:

- **vswitch-hx-inband-mgmt**: This is the default vSwitch0 which is renamed by the ESXi kickstart file as part of the automated installation. The default VMkernel port, vmk0, is configured in the standard Management Network port group. The switch has two uplinks, active on fabric A and standby on fabric B, without jumbo frames. A second port group is created for the Storage Platform Controller VMs to connect to with their individual management interfaces. A third port group is created for cluster to cluster VM snapshot replication traffic. The VLANs are not Native VLANs as assigned to the vNIC templates, and therefore they are defined in ESXi/vSphere.

- **vswitch-hx-storage-data**: This vSwitch is created as part of the automated installation. A VMkernel port, vmk1, is configured in the Storage Hypervisor Data Network port group, which is the interface used for connectivity to the HX Datastores via NFS. The switch has two uplinks, active on fabric B and standby on fabric A, with jumbo frames highly recommended. A second port group is created for the Storage Platform Controller VMs to connect to with their individual storage interfaces. The VLAN is not a Native VLAN as assigned to the vNIC templates, and therefore they are defined in ESXi/vSphere.

- **vswitch-hx-vm-network**: This vSwitch is created as part of the automated installation. The switch has two uplinks, active on both fabrics A and B, and without jumbo frames. The VLANs are not Native VLANs as assigned to the vNIC templates, and therefore they are defined in ESXi/vSphere.

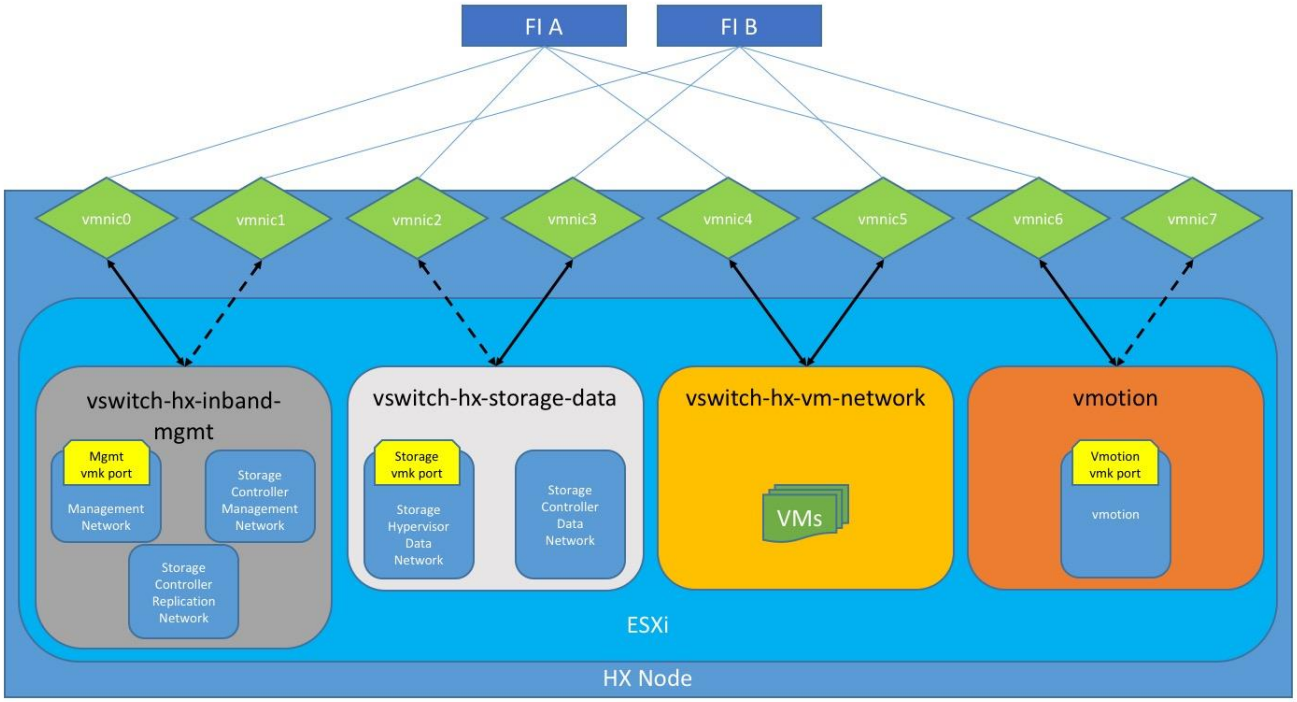
- **vmotion**: This vSwitch is created as part of the automated installation. The switch has two uplinks, active on fabric A and standby on fabric B, with jumbo frames highly recommended. The IP addresses of the VMkernel ports (vmk2) are configured during the post_install script execution. The VLAN is not a Native VLAN as assigned to the vNIC templates, and therefore they are defined in ESXi/vSphere.

The following table and figures help give more details into the ESXi virtual networking design as built by the HyperFlex installer by default:

Table 36 Virtual Switches

Virtual Switch	Port Groups	Active vmnic(s)	Passive vmnic(s)	VLAN IDs	Jumbo
vswitch-hx-in-band-mgmt	Management Network Storage Controller Management Network	vmnic0	vmnic1	<<hx-inband-mgmt>>	no
	Storage Controller Replication Network	vmnic0	vmnic1	<<hx-inband-repl>>	no
vswitch-hx-storage-data	Storage Controller Data Network Storage Hypervisor Data Network	vmnic3	vmnic2	<<hx-storage-data>>	yes
vswitch-hx-vm-network	vm-network-<<VLAN ID>>	vmnic4 vmnic5		<<vm-network>>	no
vmotion	vmotion-<<VLAN ID>>	vmnic6	vmnic7	<<hx-vmotion>>	yes

Figure 54 ESXi Network Design



VMDirectPath I/O Passthrough

VMDirectPath I/O allows a guest VM to directly access PCI and PCIe devices in an ESXi host as though they were physical devices belonging to the VM itself, also referred to as PCI passthrough. With the appropriate driver for the hardware device, the guest VM sends all I/O requests directly to the physical device, bypassing the hypervisor. In the Cisco HyperFlex system, the Storage Platform Controller VMs use this feature to gain full control of the Cisco 12Gbps SAS HBA cards in the Cisco HX-series rack-mount servers. This gives the controller VMs direct hardware level access to the physical disks installed in the servers, which they consume to construct the Cisco HX Distributed Filesystem. Only the disks connected directly to the Cisco SAS HBA are controlled by the controller VMs. Other disks, connected to different controllers, such as the SD cards, remain under the control of the ESXi hypervisor. The configuration of the VMDirectPath I/O feature is done by the Cisco HyperFlex installer and requires no manual steps.

Storage Platform Controller VMs

A key component of the Cisco HyperFlex system is the Storage Platform Controller Virtual Machine running on each of the nodes in the HyperFlex cluster. The controller VMs cooperate to form and coordinate the Cisco HX Distributed Filesystem, and service all the guest VM IO requests. The controller VMs are deployed as a vSphere ESXi agent, which is similar in concept to that of a Linux or Windows service. ESXi agents are tied to a specific host, they start and stop along with the ESXi hypervisor, and the system is not considered to be online and ready until both the hypervisor and the agents have started. Each ESXi hypervisor host has a single ESXi agent deployed, which is the controller VM for that node, and it cannot be moved or migrated to another host, nor should its settings be manually modified in any way. The collective ESXi agents are managed via an ESXi agency in the vSphere cluster.

The storage controller VM runs custom software and services that manage and maintain the Cisco HX Distributed Filesystem. The services and processes that run within the controller VMs are not exposed as part of the ESXi agents to the agency, therefore the ESXi hypervisors nor vCenter server have any direct knowledge of the storage services provided by the controller VMs. Management and visibility into the

function of the controller VMs, and the Cisco HX Distributed Filesystem is done via the HyperFlex Connect HTML management webpage, or a plugin installed to the vCenter server or appliance managing the vSphere cluster. The plugin communicates directly with the controller VMs to display the information requested, or make the configuration changes directed, all while operating within the same web-based interface of the vSphere Web Client. The deployment of the controller VMs, agents, agency, and vCenter plugin are all done by the Cisco HyperFlex installer, and requires no manual steps.

Controller VM Locations

The physical storage location of the controller VMs differs among the Cisco HX-Series rack servers, due to differences with the physical disk location and connections on those server models. The storage controller VM is operationally no different from any other typical virtual machines in an ESXi environment. The VM must have a virtual disk with the bootable root filesystem available in a location separate from the SAS HBA that the VM is controlling via VMDirectPath I/O. The configuration details of the models are as follows:

- **HX220c M5, HXAF220c M5, HX240c M5 and HXAF240c M5**: The server boots the ESXi hypervisor from the internal M.2 form factor SSD. The M.2 SSD is partitioned by the ESXi installer, and the remaining 216 GB of space is used as a VMFS datastore. The controller VM's root filesystem is stored on a 2.5 GB virtual disk, /dev/sda, which is placed on this VMFS datastore. The controller VM has full control of all the front and rear facing hot-swappable disks via PCI passthrough control of the SAS HBA. The controller VM operating system sees the 240 GB SSD, also commonly called the "housekeeping" disk as /dev/sdb, and places HyperFlex binaries and logs on this disk. The remaining disks seen by the controller VM OS are used by the HX Distributed filesystem for caching and capacity layers.

- **HX220c M4 and HXAF220c M4**: The server boots the ESXi hypervisor from the internal mirrored SD cards. The SD card is partitioned by the ESXi installer, and the remaining space is used as a VMFS datastore. The controller VM's root filesystem is stored on a 2.5 GB virtual disk, /dev/sda, which is placed on a this VMFS datastore. The controller VM has full control of all the front facing hot-swappable disks via PCI passthrough control of the SAS HBA. The controller VM operating system sees the 120 GB or 240 GB SSD, also commonly called the "housekeeping" disk as /dev/sdb, and places HyperFlex binaries and logs on this disk. The remaining disks seen by the controller VM OS are used by the HX Distributed filesystem for caching and capacity layers.

- **HX240c M4 and HXAF240c M4**: The server boots the ESXi hypervisor from the internal mirrored SD cards. The HX240c-M4SX and HXAF240c-M4SX server has a built-in SATA controller provided by the Intel Wellsburg Platform Controller Hub (PCH) chip, and the 120 GB or 240 GB housekeeping disk is connected to it, placed in an internal drive carrier. Since this model does not connect the housekeeping disk to the SAS HBA, the ESXi hypervisor remains in control of this disk, and a VMFS datastore is provisioned there, using the entire disk. On this VMFS datastore, a 2.2 GB virtual disk is created and used by the controller VM as /dev/sda for the root filesystem, and an 87 GB virtual disk is created and used by the controller VM as /dev/sdb, placing the HyperFlex binaries and logs on this disk. The front-facing hot swappable disks, seen by the controller VM OS via PCI passthrough control of the SAS HBA, are used by the HX Distributed filesystem for caching and capacity layers.

Note: On the HX240c M4 and HXAF240c M4 model servers, when configured with SEDs, the housekeeping disk is moved to a front disk slot. Since this disk is physically controlled by the SAS HBA in PCI passthrough mode, the configuration of the SCVM virtual disks changes to be the same as that of the HX220c and HXAF220c servers.

The following figures detail the Storage Platform Controller VM placement on the ESXi hypervisor hosts:

Figure 55 HX220c M5 Controller VM Placement

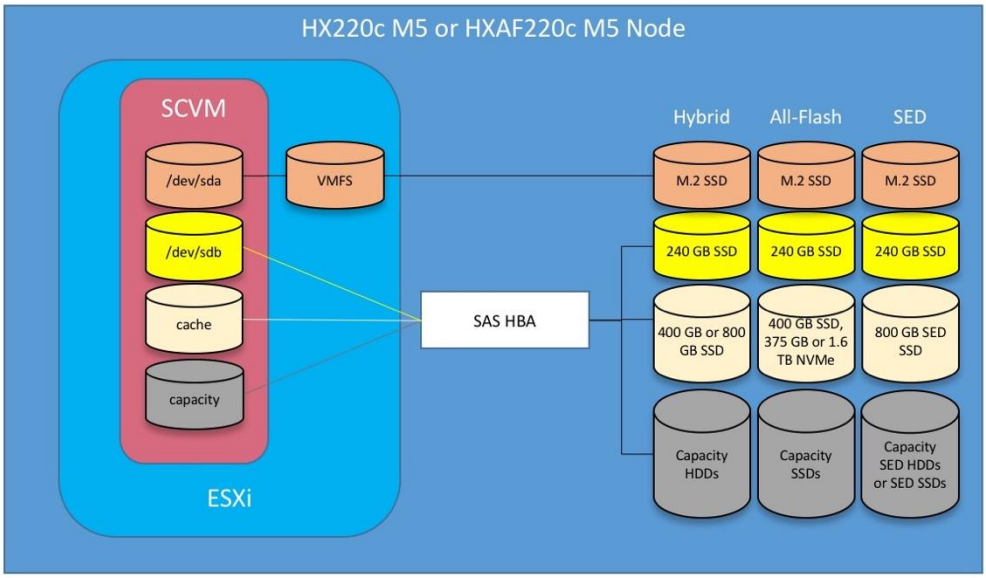

HX220c M5 or HXAF220c M5 Node
SCVM
/dev/sda
/dev/sdb
cache
capacity
VMFS
ESXi
SAS HBA
Hybrid
All-Flash
SED
M.2 SSD
M.2 SSD
M.2 SSD
240 GB SSD
240 GB SSD
240 GB SSD
400 GB or 800 GB SSD
400 GB SSD, 375 GB or 1.6 TB NVMe
800 GB SED SSD
Capacity HDDs
Capacity SSDs
Capacity SED HDDs or SED SSDs

Figure 56 HX240c M5 Controller VM Placement

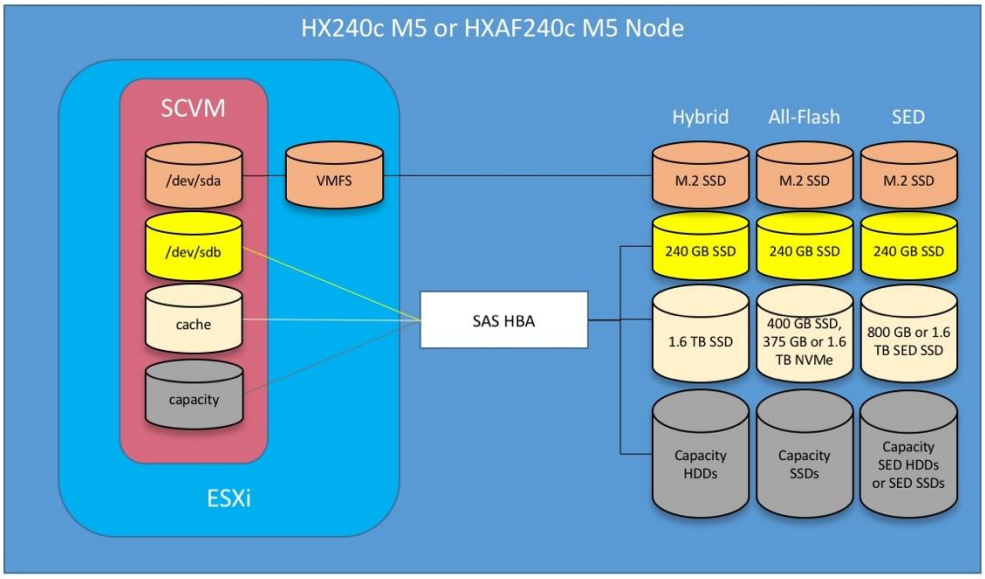


Figure 57 HX220c M4 Controller VM Placement

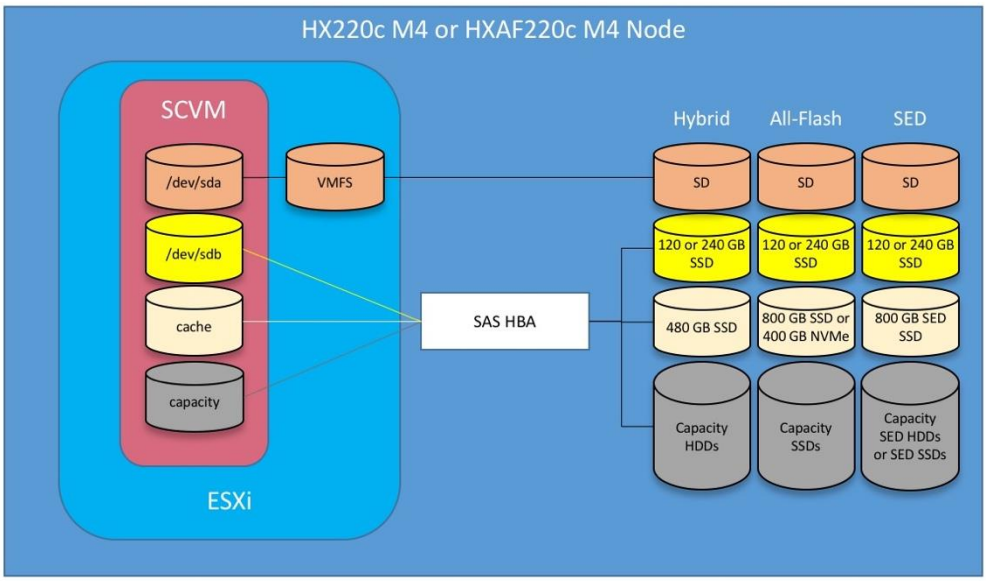
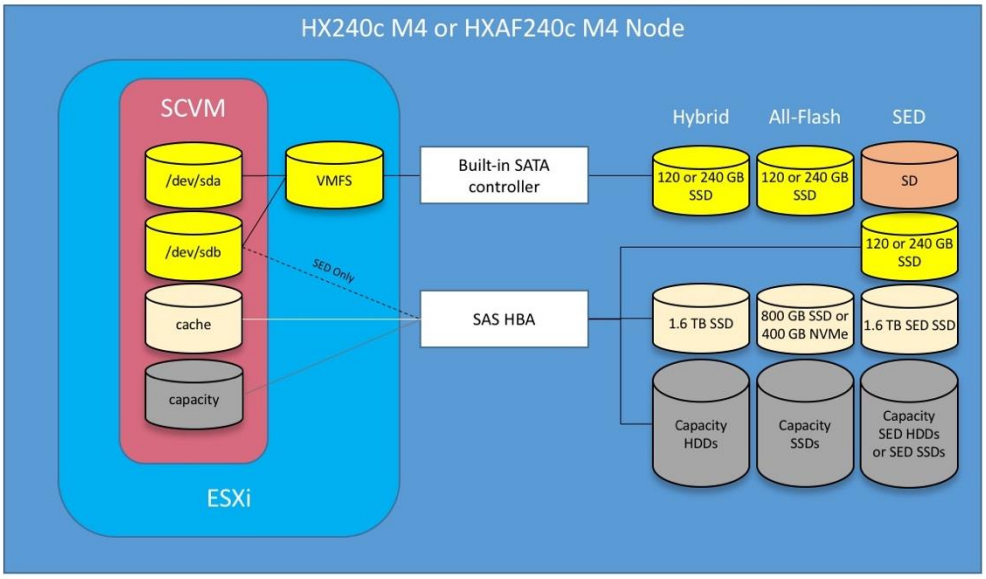


Figure 58 HX240c M4 Controller VM Placement

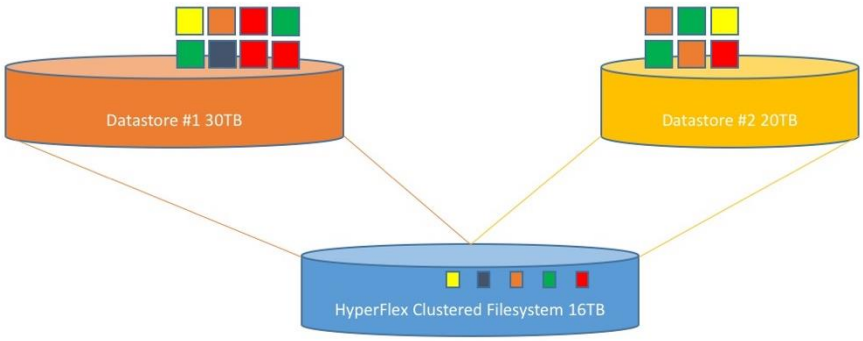


Note: The HyperFlex compute-only Cisco UCS server blades or rack-mount servers also place a lightweight storage controller VM on a 3.5 GB VMFS datastore, which can be provisioned from the SD cards, or placed on a VMFS partition alongside the boot volume if booting from SAN or local disk.

HyperFlex Datastores

A new HyperFlex cluster has no default datastores configured for virtual machine storage, therefore the datastores must be created using the vCenter Web Client plugin or the HyperFlex Connect GUI. It is important to recognize that all HyperFlex datastores are thinly provisioned, meaning that their configured size can far exceed the actual space available in the HyperFlex cluster. Alerts will be raised by the HyperFlex system in HyperFlex Connect or the vCenter plugin when actual space consumption results in low amounts of free space, and alerts will be sent via auto support email alerts. Overall space consumption in the HyperFlex clustered filesystem is optimized by the default deduplication and compression features.

Figure 59 Datastore Example



CPU Resource Reservations

Since the storage controller VMs provide critical functionality of the Cisco HX Distributed Data Platform, the HyperFlex installer will configure CPU resource reservations for the controller VMs. This reservation guarantees that the controller VMs will have CPU resources at a minimum level, in situations where the physical CPU resources of the ESXi hypervisor host are being heavily consumed by the guest VMs. This is a soft guarantee, meaning in most situations the SCVMs are not using all of the CPU resources reserved, therefore allowing the guest VMs to use them. The following table details the CPU resource reservation of the storage controller VMs:

Table 37 Controller VM CPU Reservations

Number of vCPU	Shares	Reservation	Limit
8	Low	10800 MHz	unlimited

Memory Resource Reservations

Since the storage controller VMs provide critical functionality of the Cisco HX Distributed Data Platform, the HyperFlex installer will configure memory resource reservations for the controller VMs. This reservation guarantees that the controller VMs will have memory resources at a minimum level, in situations where the physical memory resources of the ESXi hypervisor host are being heavily consumed by the guest VMs. The following table details the memory resource reservation of the storage controller VMs:

Table 38 Controller VM Memory Reservations

Server Models	Amount of Guest Memory	Reserve All Guest Memory
HX220c-M5SX HXAF220c-M5SX HX220c-M4S HXAF220c-M4S	48 GB	Yes
HX240c-M5SX HXAF240c-M5SX HX240c-M4SX HXAF240c-M4SX	72 GB	Yes
HX240c-M5L	78 GB	Yes

Note: The compute-only nodes have a lightweight storage controller VM, it is configured with only 1 vCPU of 1024MHz and 512 MB of memory reservation.

Installation

Cisco HyperFlex systems are ordered with a factory pre-installation process having been done prior to the hardware delivery. This factory integration work will deliver the HyperFlex servers with the proper firmware revisions pre-set, a copy of the VMware ESXi hypervisor software pre-installed, and some components of the Cisco HyperFlex software already installed. Once on site, the final steps to be performed are reduced and simplified due to the previous factory work. For the purpose of this document, the setup process is described as though this factory pre-installation work was done, thereby leveraging the tools and processes developed by Cisco to simplify the process and dramatically reduce the deployment time.

Installation of the Cisco HyperFlex system is primarily done via a deployable HyperFlex installer virtual machine, available for download at cisco.com as an OVA file. The installer VM performs the Cisco UCS configuration work, the configuration of ESXi on the HyperFlex hosts, the installation of the HyperFlex HX Data Platform software and creation of the HyperFlex cluster. Because this simplified installation method has been developed by Cisco, this CVD will not give detailed manual steps for the configuration of all the elements that are handled by the installer. The following sections will guide you through the prerequisites and manual steps needed prior to using the HyperFlex installer, how to utilize the HyperFlex Installer, and finally how to perform the remaining post-installation tasks.

Prerequisites

Prior to beginning the installation activities, it is important to gather the following information:

IP Addressing

To install the HX Data Platform, an OVF installer appliance must be deployed on a separate virtualization host, which is not a member of the HyperFlex cluster. The HyperFlex installer requires one IP address on the management network and the HX installer appliance IP address must be able to communicate with Cisco UCS Manager, ESXi management IP addresses on the HX hosts, and the vCenter IP addresses where the HyperFlex cluster will be managed.

Additional IP addresses for the Cisco HyperFlex system need to be allocated from the appropriate subnets and VLANs to be used. IP addresses that are used by the system fall into the following groups:

- **Cisco UCS Manager**: These addresses are used and assigned by Cisco UCS manager. Three IP addresses are used by Cisco UCS Manager; one address is assigned to each Cisco UCS Fabric Interconnect, and the third IP address is a roaming address for management of the Cisco UCS cluster. In addition, at least one IP address per Cisco UCS blade or HX-series rack-mount server is required for the hx-ext-mgmt IP address pool, which are assigned to the CIMC interface of the physical servers. Since these management addresses are assigned from a pool, they need to be provided in a contiguous block of addresses. These addresses must all be in the same subnet.
- **HyperFlex and ESXi Management**: These addresses are used to manage the ESXi hypervisor hosts, and the HyperFlex Storage Platform Controller VMs. Two IP addresses per node in the HyperFlex cluster are required from the same subnet, and a single additional IP address is needed as the roaming HyperFlex cluster management interface. These addresses can be assigned from the same subnet at the Cisco UCS Manager addresses, or they may be separate.
- **HyperFlex Replication:** These addresses are used by the HyperFlex Storage Platform Controller VMs for clusters that are configured to replicate VMs to one another. One IP address per HX node is required, plus one additional IP address as a roaming clustered replication interface. These addresses are assigned to a pool as part of a post-installation activity described later in this document, and are

not needed to complete the initial installation of a HyperFlex cluster. These addresses can be from the same subnet as the HyperFlex and ESXi management addresses, but it is recommended that the VLAN ID and subnet be unique.

- **HyperFlex Storage**: These addresses are used by the HyperFlex Storage Platform Controller VMs, and as VMkernel interfaces on the ESXi hypervisor hosts, for sending and receiving data to/from the HX Distributed Data Platform Filesystem. Two IP addresses per node in the HyperFlex cluster are required from the same subnet, and a single additional IP address is needed as the roaming HyperFlex cluster storage interface. It is recommended to provision a subnet that is not used in the network for other purposes, and it is also possible to use non-routable IP address ranges for these interfaces. Finally, if the Cisco UCS domain is going to contain multiple HyperFlex clusters, it is recommended to use a different subnet and VLAN ID for the HyperFlex storage traffic for each cluster. This is a safer method, guaranteeing that storage traffic from multiple clusters cannot intermix.
- **VMotion**: These IP addresses are used by the ESXi hypervisor hosts as VMkernel interfaces to enable vMotion capabilities. One or more IP addresses per node in the HyperFlex cluster are required from the same subnet. Multiple addresses and VMkernel interfaces can be used if you wish to enable multi-nic vMotion, although this configuration would require additional manual steps.

The following tables will assist with gathering the required IP addresses for the installation of an 8 node standard HyperFlex cluster, an 8 node stretched HyperFlex cluster, or a 4+4 extended cluster, by listing the addresses required, plus an example IP configuration:

Note: Table cells shaded in black do not require an IP address.

Table 39 HyperFlex Standard Cluster IP Addressing

Address Group:	UCS	HyperFlex and ESXi Management			HyperFlex Storage		VMotion
VLAN ID:							
Subnet:							
Subnet Mask:							
Gateway:							
Device	UCS Management Addresses	ESXi Management Interfaces	Storage Controller VM Management Interfaces	Storage Controller VM Replication Interfaces	ESXi Hypervisor Storage VMkernel Interfaces	Storage Controller VM Storage Interfaces	VMotion VMkernel Interfaces
Fabric Interconnect							
Fabric Interconnect							
UCS Manager							
HyperFlex Cluster							
HyperFlex Node #1							
HyperFlex Node #2							
HyperFlex Node #3							
HyperFlex Node #4							
HyperFlex Node #5							
HyperFlex Node #6							
HyperFlex Node #7							
HyperFlex Node #8							

Stretched clusters follow the same addressing scheme as a standard HyperFlex cluster, however they require an additional pair of Cisco UCS Fabric Interconnects for the second physical site, therefore they require additional IP addresses, as listed below:

Table 40 HyperFlex Stretched Cluster IP Addressing

Address Group:	UCS	HyperFlex and ESXi Management			HyperFlex Storage		VMotion
VLAN ID:							
Subnet:							
Subnet Mask:							
Gateway:							
Device	UCS Management Addresses	ESXi Management Interfaces	Storage Controller VM Management Interfaces	Storage Controller VM Replication Interfaces	ESXi Hypervisor Storage VMkernel Interfaces	Storage Controller VM Storage Interfaces	VMotion VMkernel Interfaces
FI A (site 1)							
FI B (site 1)							
UCS Manager (site							
FI A (site 2)							
FI B (site 2)							
UCS Manager (site							
HyperFlex Cluster							
HyperFlex Node #1							
HyperFlex Node #2							
HyperFlex Node #3							
HyperFlex Node #4							
HyperFlex Node #5							
HyperFlex Node #6							
HyperFlex Node #7							
HyperFlex Node #8							

An additional IP address is required for a stretched cluster, assigned the witness VM. The IP address can be from a different routable subnet than what is used for the cluster nodes' ESXi management and Storage Controller VM management interfaces.

Table 41 Stretched Cluster Witness VM IP Addressing

VLAN ID:	
Subnet:	
Subnet Mask:	
Gateway:	
Device	IP Addresses
Witness VM	

HyperFlex extended clusters are also addressed similarly to a standard cluster, however the compute-only nodes do not require any IP addresses for the Storage Controller VMs, as shown below:

Table 42 HyperFlex Extended Cluster IP Addressing

Address Group:	UCS	HyperFlex and ESXi Management		HyperFlex Storage	VMotion
VLAN ID:					
Subnet:					
Subnet Mask:					

Gateway:							
Device	**UCS Management Addresses**	**ESXi Management Interfaces**	**Storage Controller VM Management Interfaces**	**Storage Controller VM Replication Interfaces**	**ESXi Hypervisor Storage VMkernel Interfaces**	**Storage Controller VM Storage Interfaces**	**VMotion VMkernel Interfaces**
Fabric Interconnect							
Fabric Interconnect							
UCS Manager							
HyperFlex Cluster							
HyperFlex Node #1							
HyperFlex Node #2							
HyperFlex Node #3							
HyperFlex Node #4							
Compute Node #1							
Compute Node #2							
Compute Node #3							
Compute Node #4							

Table 43 HyperFlex Standard Cluster Example IP Addressing

Address Group:	**UCS**	**HyperFlex and ESXi Management**			**HyperFlex Storage**		**VMotion**
VLAN ID:	133	133		150	51		200
Subnet:	10.29.133.0	10.29.133.0		192.168.150.0	192.168.51.0		192.168.200.0
Subnet Mask:	255.255.255.0	255.255.255.0		255.255.255.0	255.255.255.0		255.255.255.0
Gateway:	10.29.133.1	10.29.133.1		192.168.150.1			
Device	**UCS Management Addresses**	**ESXi Management Interfaces**	**Storage Controller VM Management Interfaces**	**Storage Controller VM Replication Interfaces**	**ESXi Hypervisor Storage VMkernel Interfaces**	**Storage Controller VM Storage Interfaces**	**VMotion VMkernel Interfaces**
Fabric Interconnect	10.29.133.104						
Fabric Interconnect	10.29.133.105						
UCS Manager	10.29.133.106						
HyperFlex Cluster			10.29.133.151	192.168.150.10		192.168.51.20	
HyperFlex Node #1	10.29.133.133	10.29.133.143	10.29.133.152	192.168.150.11	192.168.51.11	192.168.51.21	192.168.200.11
HyperFlex Node #2	10.29.133.134	10.29.133.144	10.29.133.153	192.168.150.12	192.168.51.12	192.168.51.22	192.168.200.12
HyperFlex Node #3	10.29.133.135	10.29.133.145	10.29.133.154	192.168.150.13	192.168.51.13	192.168.51.23	192.168.200.13
HyperFlex Node #4	10.29.133.136	10.29.133.146	10.29.133.155	192.168.150.14	192.168.51.14	192.168.51.24	192.168.200.14
HyperFlex Node #5	10.29.133.137	10.29.133.147	10.29.133.156	192.168.150.15	192.168.51.15	192.168.51.25	192.168.200.15
HyperFlex Node #6	10.29.133.138	10.29.133.148	10.29.133.157	192.168.150.16	192.168.51.16	192.168.51.26	192.168.200.16
HyperFlex Node #7	10.29.133.139	10.29.133.149	10.29.133.158	192.168.150.17	192.168.51.17	192.168.51.27	192.168.200.17
HyperFlex Node #8	10.29.133.140	10.29.133.150	10.29.133.159	192.168.150.18	192.168.51.18	192.168.51.28	192.168.200.18

Note: IP addresses for Cisco UCS Management, plus HyperFlex and ESXi Management can come from the same subnet, or can be separate subnets, as long as the HyperFlex installer can reach them both.

DHCP versus Static IP

By default, the HX installation will assign a static IP address to the management interface of the ESXi servers. Using Dynamic Host Configuration Protocol (DHCP) for automatic IP address assignment in not recommended.

DNS

DNS servers are highly recommended to be configured for querying Fully Qualified Domain Names (FQDN) in the HyperFlex and ESXi Management group. DNS records need to be created prior to beginning the installation. At a minimum, it is highly recommended to create A records and reverse PTR records for the ESXi hypervisor hosts' management interfaces. Additional A records can be created for the Storage Controller Management interfaces, ESXi Hypervisor Storage interfaces, and the Storage Controller Storage interfaces if desired.

The following tables will assist with gathering the required DNS information for the installation, by listing the information required, and an example configuration:

Table 44 DNS Server Information

Item	Value
DNS Server #1	
DNS Server #2	
DNS Domain	
vCenter Server Name	
SMTP Server Name	
UCS Domain Name	
HX Server #1 Name	
HX Server #2 Name	
HX Server #3 Name	
HX Server #4 Name	
HX Server #5 Name	
HX Server #6 Name	

Item	Value
HX Server #7 Name	
HX Server #8 Name	

Table 45 DNS Server Example Information

Item	Value
DNS Server #1	10.29.133.110
DNS Server #2	
DNS Domain	hx.lab.cisco.com
vCenter Server Name	vcenter.hx.lab.cisco.com
SMTP Server Name	outbound.cisco.com
UCS Domain Name	HX1-FI
HX Server #1 Name	hx220-01.hx.lab.cisco.com
HX Server #2 Name	hx220-02.hx.lab.cisco.com
HX Server #3 Name	hx220-03.hx.lab.cisco.com
HX Server #4 Name	hx220-04.hx.lab.cisco.com
HX Server #5 Name	hx220-05.hx.lab.cisco.com
HX Server #6 Name	hx220-06.hx.lab.cisco.com
HX Server #7 Name	hx220-07.hx.lab.cisco.com
HX Server #8 Name	hx220-08.hx.lab.cisco.com

NTP

Consistent time clock synchronization is required across the components of the HyperFlex system, provided by reliable NTP servers, accessible for querying in the Cisco UCS Management network group, and the

HyperFlex and ESXi Management group. NTP is used by Cisco UCS Manager, vCenter, the ESXi hypervisor hosts, and the HyperFlex Storage Platform Controller VMs. The use of public NTP servers is highly discouraged, instead a reliable internal NTP server should be used.

The following tables will assist with gathering the required NTP information for the installation by listing the information required, and an example configuration:

Table 46 NTP Server Information

Item	Value
NTP Server #1	
NTP Server #2	
Timezone	

Table 47 NTP Server Example Information

Item	Value
NTP Server #1	171.68.38.65
NTP Server #2	171.68.38.66
Timezone	(UTC-8:00) Pacific Time

VLANs

Prior to the installation, the required VLAN IDs need to be documented, and created in the upstream network if necessary. At a minimum, there are 4 VLANs that need to be trunked to the Cisco UCS Fabric Interconnects that comprise the HyperFlex system; a VLAN for the HyperFlex and ESXi Management group, a VLAN for the HyperFlex Storage group, a VLAN for the VMotion group, and at least one VLAN for the guest VM traffic. If HyperFlex Replication is to be used, another VLAN must be created and trunked for the replication traffic. The VLAN IDs must be supplied during the HyperFlex Cisco UCS configuration step, and the VLAN names can optionally be customized.

The following tables will assist with gathering the required VLAN information for the installation by listing the information required, and an example configuration:

Table 48 VLAN Information

Name	ID
<<hx-inband-mgmt>>	
<<hx-inband-repl>>	

Name	ID
<<hx-storage-data>>	
<<hx-vm-data>>	
<<hx-vmotion>>	

Table 49 **VLAN Example Information**

Name	ID
hx-inband-mgmt	133
hx-inband-repl	150
hx-storage-data	51
vm-network	100
hx-vmotion	200

Network Uplinks

The Cisco UCS uplink connectivity design needs to be finalized prior to beginning the installation. One of the early manual tasks to be completed is to configure the Cisco UCS network uplinks and verify their operation, prior to beginning the HyperFlex installation steps. Refer to the network uplink design possibilities in the Network Design section.

The following tables will assist with gathering the required network uplink information for the installation by listing the information required, and an example configuration:

Table 50 **Network Uplink Configuration**

Fabric Interconnect Port		Port Channel	Port Channel Type	Port Channel ID	Port Channel Name
A		☐ Yes ☐ No	☐ LACP ☐ vPC		
		☐ Yes ☐ No			
		☐ Yes ☐ No			

Fabric Interconnect Port		Port Channel	Port Channel Type	Port Channel ID	Port Channel Name
		☐ Yes ☐ No			
B		☐ Yes ☐ No	☐ LACP ☐ vPC		
		☐ Yes ☐ No			
		☐ Yes ☐ No			
		☐ Yes ☐ No			

Table 51 **Network Uplink Example Configuration**

Fabric Interconnect Port		Port Channel	Port Channel Type	Port Channel ID	Port Channel Name
A	1/25	☒ Yes ☐ No	☐ LACP ☒ vPC	10	vpc-10
	1/26	☒ Yes ☐ No			
		☐ Yes ☐ No			
		☐ Yes ☐ No			
B	1/25	☒ Yes ☐ No	☐ LACP ☒ vPC	20	vpc-20
	1/26	☒ Yes ☐ No			
		☐ Yes ☐ No			
		☐ Yes ☐ No			

Usernames and Passwords

Several usernames and passwords need to be defined or known as part of the HyperFlex installation process. The following tables will assist with gathering the required username and password information by listing the information required and an example configuration:

Table 52 Usernames and Passwords

Account	Username	Password
HX Installer Administrator	root	<<hx_install_root_pw>>
UCS Administrator	admin	<<ucs_admin_pw>>
ESXi Administrator	root	<<esxi_root_pw>>
HyperFlex Administrator	root	<<hx_admin_pw>>
vCenter Administrator	<<vcenter_administrator>>	<<vcenter_admin_pw>>

Table 53 Example Usernames and Passwords

Account	Username	Password
HX Installer Administrator	root	Cisco123
UCS Administrator	admin	Cisco123
ESXi Administrator	root	Cisco123
HyperFlex Administrator	root	Cisco123!!
vCenter Administrator	administrator@vsphere.local	!QAZ2wsx

Physical Installation

Install the Fabric Interconnects, the HX-Series rack-mount servers, standard C-series rack-mount servers, the Cisco UCS 5108 chassis, the Cisco UCS Fabric Extenders, and the Cisco UCS blades according to their corresponding hardware installation guides listed below. For a stretched cluster deployment, the physical installation is identical to a standard cluster, only it is duplicated in two different physical locations.

Cisco UCS 6200 Series Fabric Interconnect:

http://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/hw/6200-install-guide/6200_HIG.pdf

Cisco UCS 6300 Series Fabric Interconnect:

https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/hw/6300-install-guide/6300_Series_HIG.html

HX220c M5 Server:

https://www.cisco.com/c/en/us/td/docs/hyperconverged_systems/HX_series/HX220c_M5/HX220c_M5.html

HX240c M5 Server:

https://www.cisco.com/c/en/us/td/docs/hyperconverged_systems/HX_series/HX240c_M5/HX240c_M5.html

HX240c M5L Server:

https://www.cisco.com/c/en/us/td/docs/hyperconverged_systems/HX_series/HX240c_M5/HX240c_M5.html

HX220c M4 Server:

http://www.cisco.com/c/en/us/td/docs/hyperconverged_systems/HX_series/HX220c_M4/HX220c/overview.html

HX240c M4 Server:

http://www.cisco.com/c/en/us/td/docs/hyperconverged_systems/HX_series/HX240c_M4/HX240c/overview.html

Cisco UCS 5108 Chassis, Servers and Fabric Extenders:

http://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/hw/chassis-install-guide/ucs5108_install.pdf

Cabling

The physical layout of the HyperFlex system was previously described in section Physical Topology. The Fabric Interconnects, HX-series rack-mount servers, Cisco UCS chassis and blades need to be cabled properly before beginning the installation activities. For a stretched cluster deployment, the physical cabling is identical to a standard cluster, only it is duplicated in two different physical locations.

Table 54 provides an example cabling map for installation of a Cisco HyperFlex system, with eight HyperFlex converged servers, and one Cisco UCS 5108 chassis.

Table 54 Example Cabling Map

Device	Port	Connected To	Port	Type	Length	Note
UCS6248-A	L1	UCS6248-B	L1	CAT5	1FT	
UCS6248-A	L2	UCS6248-B	L2	CAT5	1FT	
UCS6248-A	mgmt0	Customer LAN				
UCS6248-A	1/1	HX Server #1	mLOM port 1	Twinax	3M	Server 1
UCS6248-A	1/2	HX Server #2	mLOM port 1	Twinax	3M	Server 2
UCS6248-A	1/3	HX Server #3	mLOM port 1	Twinax	3M	Server 3
UCS6248-A	1/4	HX Server #4	mLOM port 1	Twinax	3M	Server 4
UCS6248-A	1/5	HX Server #5	mLOM port 1	Twinax	3M	Server 5
UCS6248-A	1/6	HX Server #6	mLOM port 1	Twinax	3M	Server 6

Device	Port	Connected To	Port	Type	Length	Note
UCS6248-A	1/7	HX Server #7	mLOM port 1	Twinax	3M	Server 7
UCS6248-A	1/8	HX Server #8	mLOM port 1	Twinax	3M	Server 8
UCS6248-A	1/9	2204XP #1	IOM1 port 1	Twinax	3M	Chassis 1
UCS6248-A	1/10	2204XP #1	IOM1 port 2	Twinax	3M	Chassis 1
UCS6248-A	1/11	2204XP #1	IOM1 port 3	Twinax	3M	Chassis 1
UCS6248-A	1/12	2204XP #1	IOM1 port 4	Twinax	3M	Chassis 1
UCS6248-A	1/13					
UCS6248-A	1/14					
UCS6248-A	1/15					
UCS6248-A	1/16					
UCS6248-A	1/17					
UCS6248-A	1/18					
UCS6248-A	1/19					
UCS6248-A	1/20					
UCS6248-A	1/21					
UCS6248-A	1/22					
UCS6248-A	1/23					
UCS6248-A	1/24					
UCS6248-A	1/25	Customer LAN				uplink
UCS6248-A	1/26	Customer LAN				uplink
UCS6248-A	1/27					
UCS6248-A	1/28					
UCS6248-A	1/29					
UCS6248-A	1/30					
UCS6248-A	1/31					
UCS6248-A	1/32					

UCS6248-B	L1	UCS6248-A	L1	CAT5	1FT	
UCS6248-B	L2	UCS6248-A	L2	CAT5	1FT	
UCS6248-B	mgmt0	Customer LAN				

Device	Port	Connected To	Port	Type	Length	Note
UCS6248-B	1/1	HX Server #1	mLOM port 2	Twinax	3M	Server 1
UCS6248-B	1/2	HX Server #2	mLOM port 2	Twinax	3M	Server 2
UCS6248-B	1/3	HX Server #3	mLOM port 2	Twinax	3M	Server 3
UCS6248-B	1/4	HX Server #4	mLOM port 2	Twinax	3M	Server 4
UCS6248-B	1/5	HX Server #5	mLOM port 2	Twinax	3M	Server 5
UCS6248-B	1/6	HX Server #6	mLOM port 2	Twinax	3M	Server 6
UCS6248-B	1/7	HX Server #7	mLOM port 2	Twinax	3M	Server 7
UCS6248-B	1/8	HX Server #8	mLOM port 2	Twinax	3M	Server 8
UCS6248-B	1/9	2204XP #1	IOM2 port 1	Twinax	3M	Chassis 1
UCS6248-B	1/10	2204XP #1	IOM2 port 2	Twinax	3M	Chassis 1
UCS6248-B	1/11	2204XP #1	IOM2 port 3	Twinax	3M	Chassis 1
UCS6248-B	1/12	2204XP #1	IOM2 port 4	Twinax	3M	Chassis 1
UCS6248-B	1/13					
UCS6248-B	1/14					
UCS6248-B	1/15					
UCS6248-B	1/16					
UCS6248-B	1/17					
UCS6248-B	1/18					
UCS6248-B	1/19					
UCS6248-B	1/20					
UCS6248-B	1/21					
UCS6248-B	1/22					
UCS6248-B	1/23					
UCS6248-B	1/24					
UCS6248-B	1/25	Customer LAN				uplink
UCS6248-B	1/26	Customer LAN				uplink
UCS6248-B	1/27					
UCS6248-B	1/28					
UCS6248-B	1/29					
UCS6248-B	1/30					

Device	Port	Connected To	Port	Type	Length	Note
UCS6248-B	1/31					
UCS6248-B	1/32					

Cisco UCS Installation

This section describes the steps to initialize and configure the Cisco UCS Fabric Interconnects, to prepare them for the HyperFlex installation.

Cisco UCS Fabric Interconnect A

To configure Fabric Interconnect A, complete the following steps:

1. Make sure the Fabric Interconnect cabling is properly connected, including the L1 and L2 cluster links, and power the Fabric Interconnects on by inserting the power cords.

2. Connect to the console port on the first Fabric Interconnect, which will be designated as the A fabric device. Use the supplied Cisco console cable (CAB-CONSOLE-RJ45=), and connect it to a built-in DB9 serial port, or use a USB to DB9 serial port adapter.

3. Start your terminal emulator software.

4. Create a connection to the COM port of the computer's DB9 port, or the USB to serial adapter. Set the terminal emulation to VT100, and the settings to 9600 baud, 8 data bits, no parity, and 1 stop bit.

5. Open the connection just created. You may have to press ENTER to see the first prompt.

6. Configure the first Fabric Interconnect, using the following example as a guideline:

```
        ---- Basic System Configuration Dialog ----

This setup utility will guide you through the basic configuration of
the system. Only minimal configuration including IP connectivity to
the Fabric interconnect and its clustering mode is performed through these steps.

Type Ctrl-C at any time to abort configuration and reboot system.
To back track or make modifications to already entered values,
complete input till end of section and answer no when prompted
to apply configuration.


Enter the configuration method. (console/gui) ? console

Enter the setup mode; setup newly or restore from backup. (setup/restore) ? setup

You have chosen to setup a new Fabric interconnect. Continue? (y/n): y

Enforce strong password? (y/n) [y]: y

Enter the password for "admin":
Confirm the password for "admin":

Is this Fabric interconnect part of a cluster(select 'no' for standalone)? (yes/no) [n]: yes

Enter the switch fabric (A/B) []: A
```

```
Enter the system name:  HX1-FI

Physical Switch Mgmt0 IP address : 10.29.133.104

Physical Switch Mgmt0 IPv4 netmask : 255.255.255.0

IPv4 address of the default gateway : 10.29.133.1

Cluster IPv4 address : 10.29.133.106

Configure the DNS Server IP address? (yes/no) [n]: yes

  DNS IP address : 10.29.133.110

Configure the default domain name? (yes/no) [n]: yes

  Default domain name : hx.lab.cisco.com

Join centralized management environment (UCS Central)? (yes/no) [n]: no

Following configurations will be applied:

  Switch Fabric=A
  System Name=HX1-FI
  Enforced Strong Password=no
  Physical Switch Mgmt0 IP Address=10.29.133.104
  Physical Switch Mgmt0 IP Netmask=255.255.255.0
  Default Gateway=10.29.133.1
  Ipv6 value=0
  DNS Server=10.29.133.110
  Domain Name=hx.lab.cisco.com

  Cluster Enabled=yes
  Cluster IP Address=10.29.133.106
  NOTE: Cluster IP will be configured only after both Fabric Interconnects are initialized

Apply and save the configuration (select 'no' if you want to re-enter)? (yes/no): yes
Applying configuration. Please wait.

Configuration file - Ok
```

Cisco UCS Fabric Interconnect B

To configure Fabric Interconnect B, complete the following steps:

1. Connect to the console port on the first Fabric Interconnect, which will be designated as the B fabric device. Use the supplied Cisco console cable (CAB-CONSOLE-RJ45=), and connect it to a built-in DB9 serial port, or use a USB to DB9 serial port adapter.

2. Start your terminal emulator software.

3. Create a connection to the COM port of the computer's DB9 port, or the USB to serial adapter. Set the terminal emulation to VT100, and the settings to 9600 baud, 8 data bits, no parity, and 1 stop bit.

4. Open the connection just created. You may have to press ENTER to see the first prompt.

5. Configure the second Fabric Interconnect, using the following example as a guideline:

```
---- Basic System Configuration Dialog ----
```

```
  This setup utility will guide you through the basic configuration of
  the system. Only minimal configuration including IP connectivity to
  the Fabric interconnect and its clustering mode is performed through these steps.

  Type Ctrl-C at any time to abort configuration and reboot system.
  To back track or make modifications to already entered values,
  complete input till end of section and answer no when prompted
  to apply configuration.


  Enter the configuration method. (console/gui) ? console

  Installer has detected the presence of a peer Fabric interconnect. This Fabric interconnect
will be added to the cluster. Continue (y/n) ? y

  Enter the admin password of the peer Fabric interconnect:
    Connecting to peer Fabric interconnect... done
    Retrieving config from peer Fabric interconnect... done
    Peer Fabric interconnect Mgmt0 IPv4 Address: 10.29.133.104
    Peer Fabric interconnect Mgmt0 IPv4 Netmask: 255.255.255.0
    Cluster IPv4 address          : 10.29.133.106

    Peer FI is IPv4 Cluster enabled. Please Provide Local Fabric Interconnect Mgmt0 IPv4 Address

  Physical Switch Mgmt0 IP address : 10.29.133.105


  Apply and save the configuration (select 'no' if you want to re-enter)? (yes/no): yes
  Applying configuration. Please wait.

Configuration file - Ok
```

Cisco UCS Manager

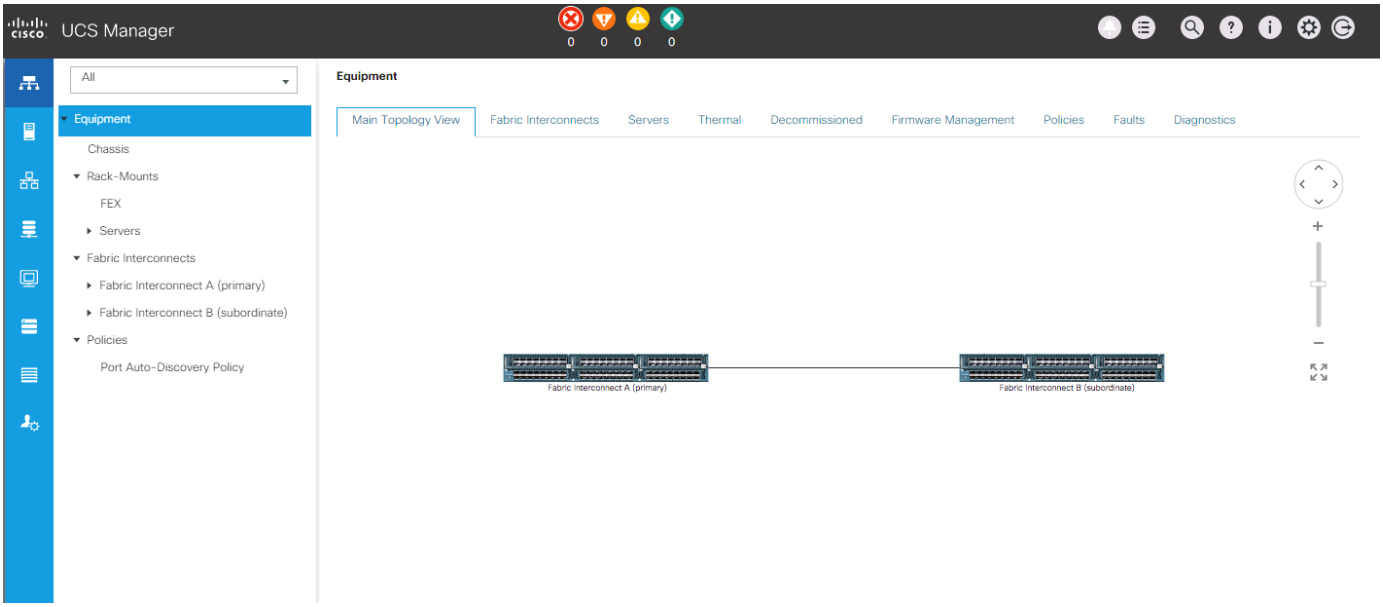
Log in to the Cisco UCS Manager environment by completing the following steps:

1. Open a web browser and navigate to the Cisco UCS Manager Cluster IP address, for example https://10.29.133.106

2. Click the “Launch UCS Manager” HTML link to open the Cisco UCS Manager web client.

3. At the login prompt, enter “admin” as the username, and enter the administrative password that was set during the initial console configuration.

4. Click No when prompted to enable Cisco Smart Call Home, this feature can be enabled at a later time.



Cisco UCS Configuration

Configure the following ports, settings, and policies in the Cisco UCS Manager interface prior to beginning the HyperFlex installation.

Cisco UCS Firmware

Your Cisco UCS firmware version should be correct as shipped from the factory, as documented in the Software Components section. This document is based on Cisco UCS infrastructure, B-series bundle, and C-Series bundle software versions 3.2(3a). If the firmware version of the Fabric Interconnects is older than this version, the firmware must be upgraded to match the requirements prior to completing any further steps. To upgrade the Cisco UCS Manager version, the Fabric Interconnect firmware, and the server bundles, refer to these instructions:

https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/ucs-manager/GUI-User-Guides/Firmware-Mgmt/3-2/b_UCSM_GUI_Firmware_Management_Guide_3_2.html

NTP

To synchronize the Cisco UCS environment time to the NTP server, complete the following steps:

1. In Cisco UCS Manager, click the Admin button on the left-hand side.

2. In the navigation pane, select All > Time Zone Management, and click the carat next to Time Zone Management to expand it.

3. Click Timezone.

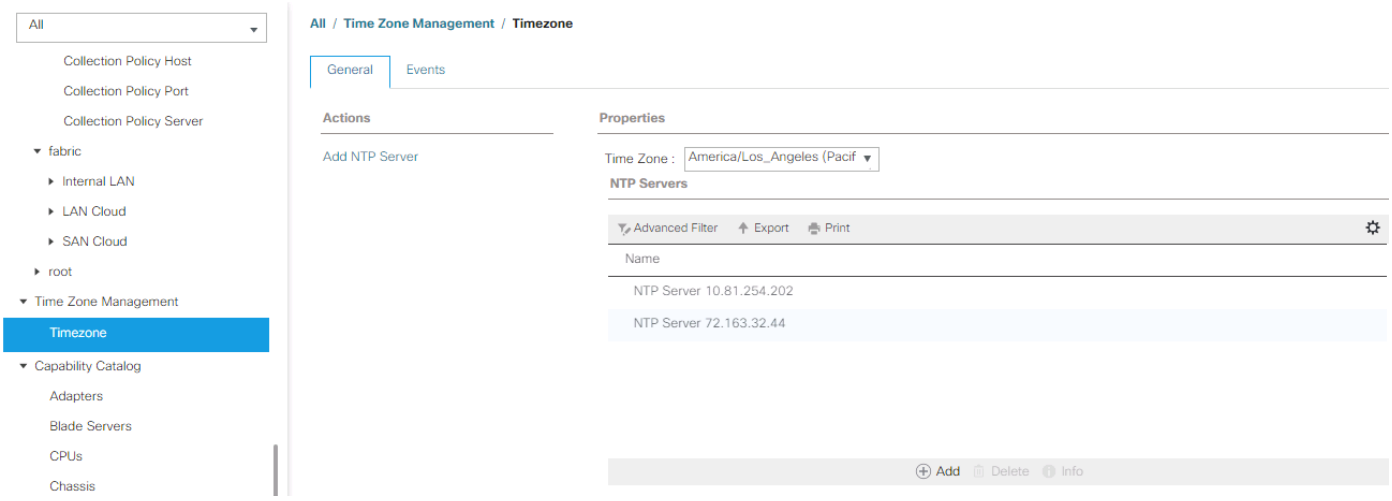
4. In the Properties pane, select the appropriate time zone in the Time Zone menu.

5. Click Add NTP Server.

6. Enter the NTP server IP address and click OK.

7. Click OK.

8. Click Save Changes and then click OK.



Uplink Ports

The Ethernet ports of a Cisco UCS Fabric Interconnect are all capable of performing several functions, such as network uplinks or server ports, and more. By default, all ports are unconfigured, and their function must be defined by the administrator. To define the specified ports to be used as network uplinks to the upstream network, complete the following steps:

1. In Cisco UCS Manager, click the Equipment button on the left-hand side.

2. Select Fabric Interconnects > Fabric Interconnect A > Fixed Module or Expansion Module as appropriate > Ethernet Ports.

3. Select the ports that are to be uplink ports, right click them, and click Configure as Uplink Port.

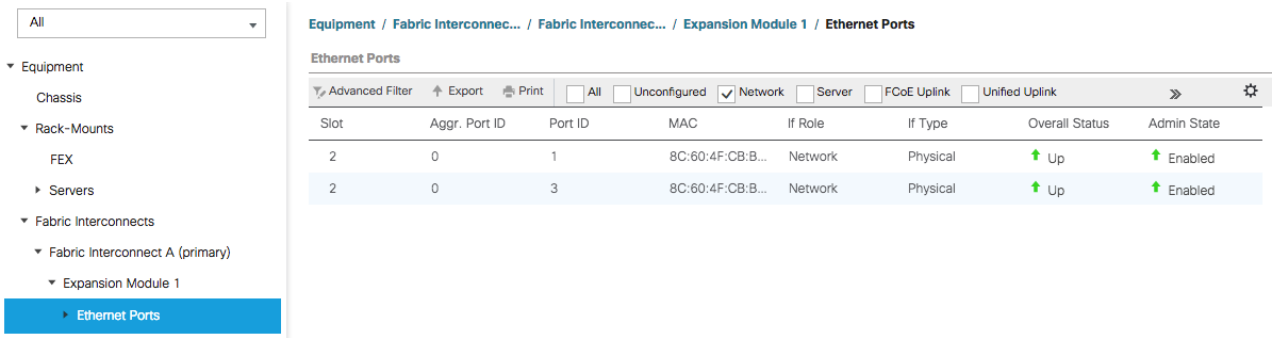
4. Click Yes to confirm the configuration, and click OK.

5. Select Fabric Interconnects > Fabric Interconnect B > Fixed Module or Expansion Module as appropriate > Ethernet Ports.

6. Select the ports that are to be uplink ports, right-click them, and click Configure as Uplink Port.

7. Click Yes to confirm the configuration and click OK.

8. Verify all the necessary ports are now configured as uplink ports, where their role is listed as “Network”.



Uplink Port Channels

If the Cisco UCS uplinks from one Fabric Interconnect are to be combined into a port channel or vPC, you must separately configure the port channels, which will use the previously configured uplink ports. To configure the necessary port channels in the Cisco UCS environment, complete the following steps:

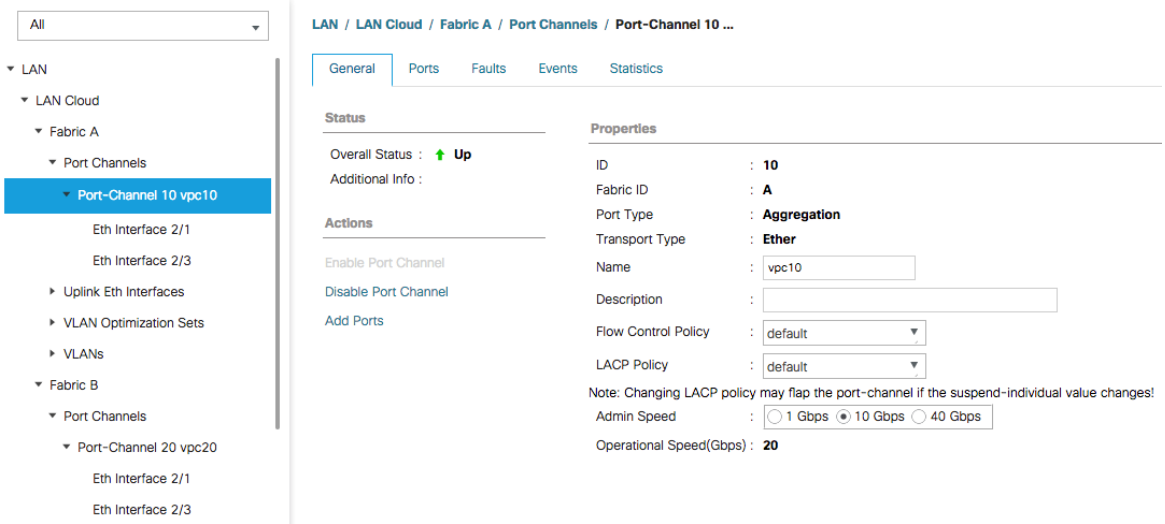
1. In Cisco UCS Manager, click the LAN button on the left-hand side.
2. Under LAN > LAN Cloud, click the carat to expand the Fabric A tree.
3. Right-click Port Channels underneath Fabric A, then click Create Port Channel.
4. Enter the port channel ID number as the unique ID of the port channel (this does not have to match the port-channel ID on the upstream switch).
5. Enter the name of the port channel.
6. Click Next.
7. Click each port from Fabric Interconnect A that will participate in the port channel, and click the >> button to add them to the port channel.
8. Click Finish.
9. Click OK.
10. Under LAN > LAN Cloud, click the carat to expand the Fabric B tree.
11. Right-click Port Channels underneath Fabric B, then click Create Port Channel.
12. Enter the port channel ID number as the unique ID of the port channel (this does not have to match the port-channel ID on the upstream switch).
13. Enter the name of the port channel.
14. Click Next.

15. Click each port from Fabric Interconnect B that will participate in the port channel, and click the >> button to add them to the port channel.

16. Click Finish.

17. Click OK.

18. Verify the necessary port channels have been created. It can take a few minutes for the newly formed port channels to converge and come online.



Chassis Discovery Policy

If the Cisco HyperFlex system will use blades as compute-only nodes in an extended cluster design, additional settings must be configured for connecting the Cisco UCS 5108 blade chassis. The Chassis Discovery policy defines the number of links between the Fabric Interconnect and the Cisco UCS Fabric Extenders which must be connected and active, before the chassis will be discovered. This also effectively defines how many of those connected links will be used for communication. The Link Grouping Preference setting specifies if the links will operate independently, or if Cisco UCS Manager will automatically combine them into port-channels. Cisco best practices recommends using link grouping, and the number of links per side is dependent on the hardware used in Cisco UCS 5108 chassis, and the model of Fabric Interconnects. For 10 GbE connections Cisco recommends 4 links per side, and for 40 GbE connections Cisco recommends 2 links per side.

To configure the necessary policy and setting, complete the following steps:

1. In Cisco UCS Manager, click the Equipment button on the left-hand side, and click Equipment in the top of the navigation tree on the left.

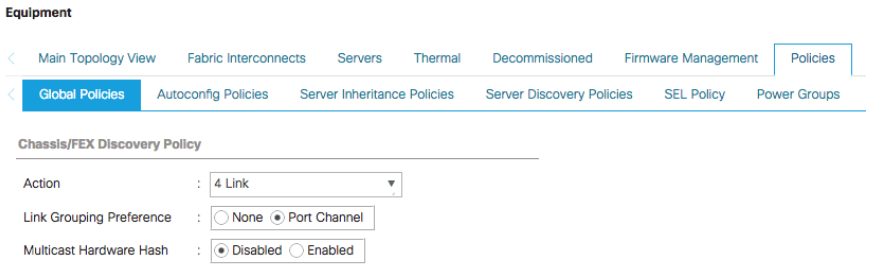
2. In the properties pane, click the Policies tab.

3. Under the Global Policies sub-tab, set the Chassis/FEX Discovery Policy to match the number of uplink ports that are cabled per side, between the chassis and the Fabric Interconnects.

4. Set the Link Grouping Preference option to Port Channel.

5. Click Save Changes.

6. Click OK.



Server Ports

The Ethernet ports of a Cisco UCS Fabric Interconnect connected to the rack-mount servers, or to the blade chassis must be defined as server ports. When a server port is activated, the connected server or chassis will begin the discovery process shortly afterwards. Rack-mount servers and blade chassis are automatically numbered in Cisco UCS Manager in the order which they are first discovered. For this reason, it is important to configure the server ports sequentially in the order you wish the physical servers and/or chassis to appear within Cisco UCS Manager. For example, if you installed your servers in a cabinet or rack with server #1 on the bottom, counting up as you go higher in the cabinet or rack, then you need to enable the server ports to the bottom-most server first, and enable them one-by-one as you move upward. You must wait until the server appears in the Equipment tab of Cisco UCS Manager before configuring the ports for the next server. The same numbering procedure applies to blade server chassis, although chassis and rack-mount server numbers are separate from each other.

Auto Configuration

A new feature in Cisco UCS Manager 3.1(3a) and later is Server Port Auto-Discovery, which automates the configuration of ports on the Fabric Interconnects as server ports when a Cisco UCS rack-mount server or blade chassis is connected to them. The firmware on the rack-mount servers or blade chassis Fabric Extenders must already be at version 3.1(3a) or later in order for this feature to function properly. Enabling this policy eliminates the manual steps of configuring each server port, however it does configure the servers in a somewhat random order. For example, the rack-mount server at the bottom of the stack, which you may refer to as server #1, and you may have plugged into port 1 of both Fabric Interconnects, could be discovered as server 2, or server 5, etc. In order to have fine control of the rack-mount server or chassis numbering and order, the manual configuration steps listed in the next section must be followed.

To configure automatic server port definition and discovery, complete the following steps:

1. In Cisco UCS Manager, click the Equipment button on the left-hand side.

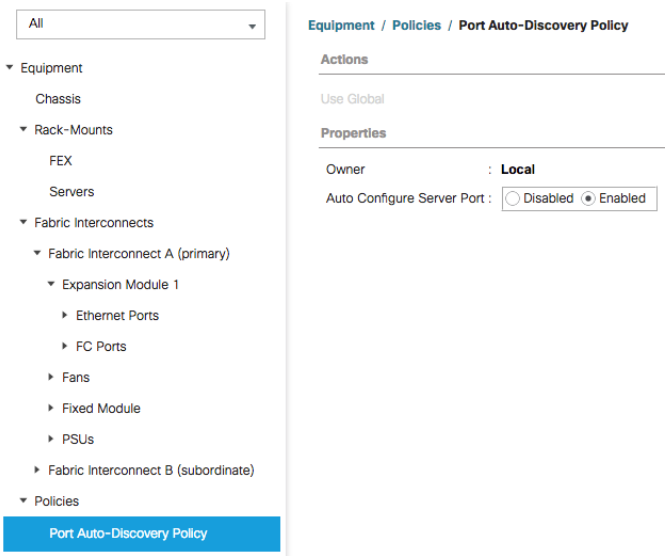
2. In the navigation tree, under Policies, click Port Auto-Discovery Policy

3. In the properties pane, set Auto Configure Server Port option to Enabled.

4. Click Save Changes.

5. Click OK.

6. Wait for a brief period, until the rack-mount servers appear in the Equipment tab underneath Equipment > Rack Mounts > Servers, or the chassis appears underneath Equipment > Chassis.



Manual Configuration

To manually define the specified ports to be used as server ports, and have control over the numbering of the servers, complete the following steps:

1. In Cisco UCS Manager, click the Equipment button on the left-hand side.

2. Select Fabric Interconnects > Fabric Interconnect A > Fixed Module or Expansion Module as appropriate > Ethernet Ports.

3. Select the first port that is to be a server port, right click it, and click Configure as Server Port.

4. Click Yes to confirm the configuration and click OK.

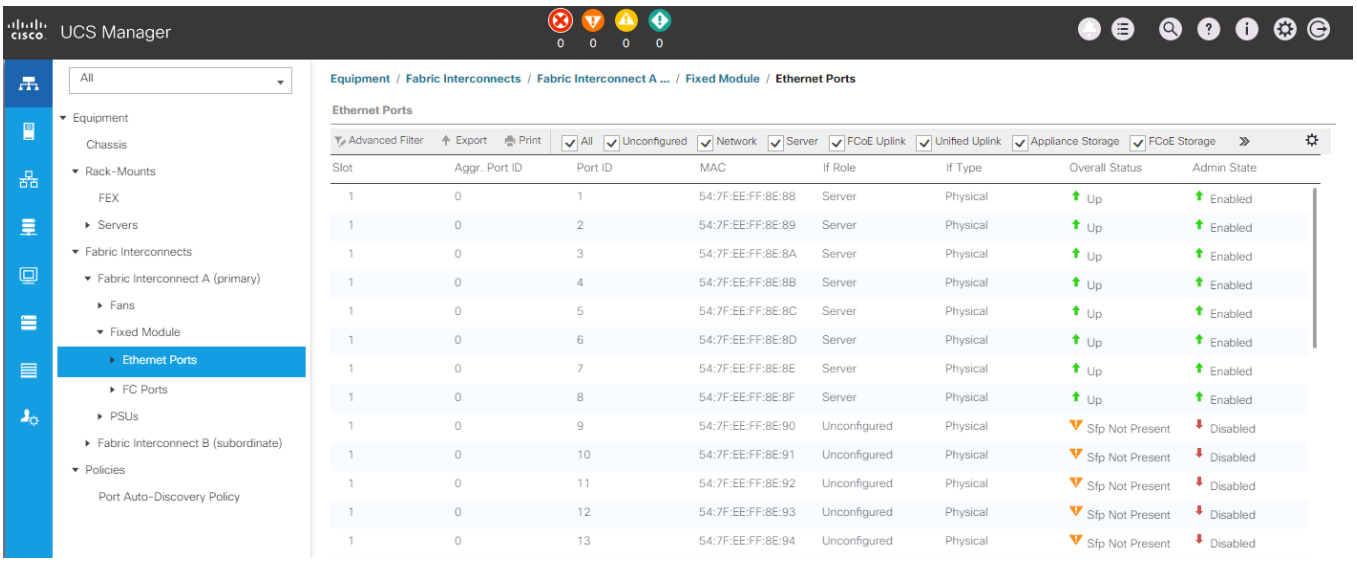
5. Select Fabric Interconnects > Fabric Interconnect B > Fixed Module or Expansion Module as appropriate > Ethernet Ports.

6. Select the matching port as chosen for Fabric Interconnect A that is to be a server port, right-click it, and click Configure as Server Port.

7. Click Yes to confirm the configuration and click OK.

8. Wait for a brief period, until the rack-mount server appears in the Equipment tab underneath Equipment > Rack Mounts > Servers, or the chassis appears underneath Equipment > Chassis.

9. Repeat Steps 1-8 for each pair of server ports, until all rack-mount servers and chassis appear in the order desired in the Equipment tab.



Server Discovery

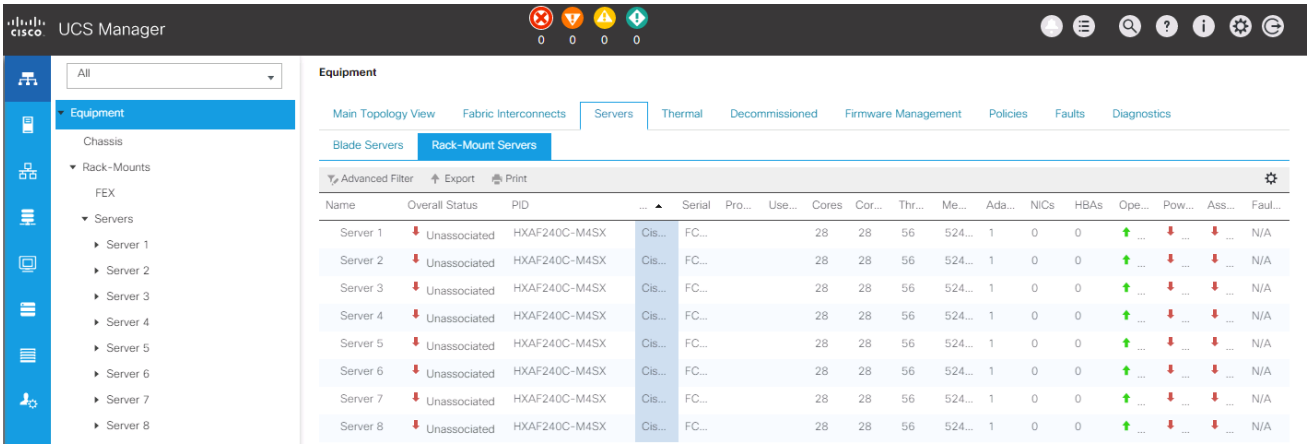
As previously described, when the server ports of the Fabric Interconnects are configured and active, the servers connected to those ports will begin a discovery process. During discovery, the servers' internal hardware inventories are collected, along with their current firmware revisions. Before continuing with the HyperFlex installation processes, which will create the service profiles and associate them with the servers, wait for all of the servers to finish their discovery process and to show as unassociated servers that are powered off, with no errors.

To view the servers' discovery status, complete the following steps:

1. In Cisco UCS Manager, click the Equipment button on the left-hand side, and click Equipment in the top of the navigation tree on the left.

2. In the properties pane, click the Servers tab.

3. Click the Blade Servers or Rack-Mount Servers sub-tab as appropriate, and view the servers' status in the Overall Status column.



HyperFlex Installer Deployment

The Cisco HyperFlex software is distributed as a deployable virtual machine, contained in an Open Virtual Appliance (OVA) file format. The HyperFlex OVA file is available for download at cisco.com:

https://software.cisco.com/download/home/286305544/type/286305994/release/3.0%25281a%2529

This document is based on the Cisco HyperFlex 3.0(1a) release filename: ***Cisco-HX-Data-Platform-Installer-v3.0.1a-29499-esx.ova***

The HyperFlex installer OVA file can be deployed as a virtual machine in an existing VMware vSphere environment, VMware Workstation, VMware Fusion, or other virtualization environment which supports the import of OVA format files. For the purpose of this document, the process described uses an existing ESXi server managed by vCenter to run the HyperFlex installer OVA, and deploying it via the VMware vSphere Web Client.

Installer Connectivity

The Cisco HyperFlex Installer VM must be deployed in a location that has connectivity to the following network locations and services:

- Connectivity to the vCenter Server which will manage the HyperFlex cluster(s) to be installed.
- Connectivity to the management interfaces of the Fabric Interconnects that contain the HyperFlex cluster(s) to be installed.
- Connectivity to the management interface of the ESXi hypervisor hosts which will host the HyperFlex cluster(s) to be installed.
- Connectivity to the DNS server(s) which will resolve host names used by the HyperFlex cluster(s) to be installed.
- Connectivity to the NTP server(s) which will synchronize time for the HyperFlex cluster(s) to be installed.
- Connectivity from the staff operating the installer to the webpage hosted by the installer, and to log in to the installer via SSH.

For complete details of all ports required for the installation of Cisco HyperFlex, refer to Appendix A of the HyperFlex 3.0 Hardening Guide: https://www.cisco.com/c/dam/en/us/support/docs/hyperconverged-infrastructure/hyperflex-hx-data-platform/HX-Hardening_Guide_v3-0.pdf

If the network where the HyperFlex installer VM is deployed has DHCP services available to assign the proper IP address, subnet mask, default gateway, and DNS servers, the HyperFlex installer can be deployed using DHCP. If a static address must be defined, use Table 55 to document the settings to be used for the HyperFlex installer VM:

Table 55 HyperFlex Installer Settings

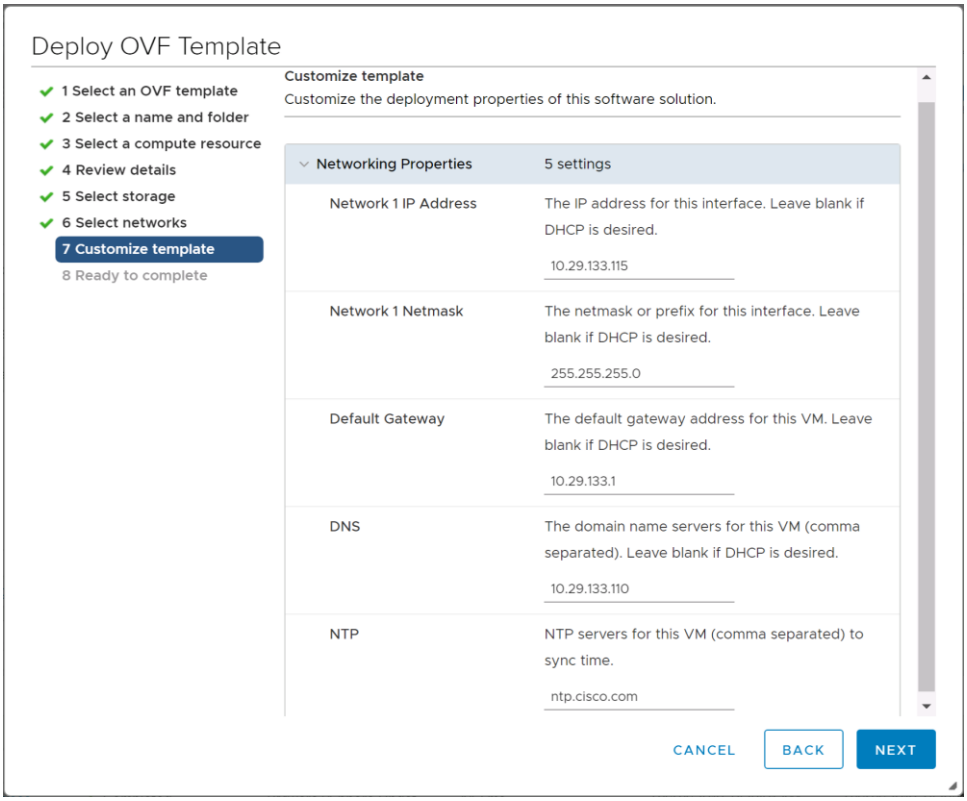
Setting	Value
IP Address	
Subnet Mask	

Setting	Value
Default Gateway	
DNS Server #1	
NTP Servers	

Deploy Installer OVA

To deploy the HyperFlex installer OVA, complete the following steps:

1. Open the vSphere Web Client webpage, or the vSphere HTML5 Web Client webpage of a vCenter server where the installer OVA will be deployed, and log in with admin privileges.
2. In the vSphere Web Client, from the Home view, click Hosts and Clusters.
3. From the Actions menu, click Deploy OVF Template.
4. Click the Local file option, then click Browse and locate the *Cisco-HX-Data-Platform-Installer-v3.0.1a-29499-esx.ova* file, click the file and click Open.
5. Click Next.
6. Modify the name of the virtual machine to be created if desired, and click a folder location to place the virtual machine, then click Next.
7. Click a specific host or cluster to locate the virtual machine and click Next.
8. After the file validation, review the details and click Next.
9. Select a Thin provision virtual disk format, and the datastore to store the new virtual machine, then click Next.
10. Modify the network port group selection from the drop-down list in the Destination Networks column, choosing the network the installer VM will communicate on, and click Next.
11. If DHCP is to be used for the installer VM, leave the fields blank, except for the NTP server value and click Next. If static address settings are to be used, fill in the fields for the DNS server, Default Gateway, NTP Servers, IP address, and subnet mask, then click Next.



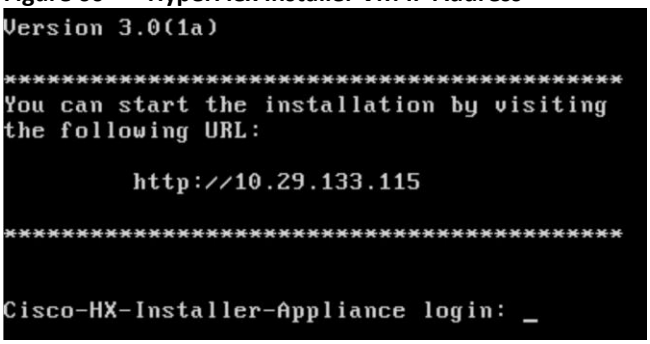
12. Review the final configuration and click Finish.

13. The installer VM will take a few minutes to deploy, once it has deployed, power on the new VM and proceed to the next step.

HyperFlex Installer Web Page

The HyperFlex installer is accessed via a webpage using your local computer and a web browser. If the HyperFlex installer was deployed with a static IP address, then the IP address of the website is already known. If DHCP was used, open the local console of the installer VM. In the console, you will see an interface similar to the example below, showing the IP address that was leased:

Figure 60 HyperFlex Installer VM IP Address



To access the HyperFlex installer webpage, complete the following steps:

1. Open a web browser on the local computer and navigate to the IP address of the installer VM. For example, open http://10.29.133.115

2. Click accept or continue to bypass any SSL certificate errors.

3. At the login screen, enter the username: root

4. At the login screen, enter the default password: Cisco123

5. Verify the version of the installer in the lower right-hand corner of the Welcome page is the correct version.

6. Check the box for “I accept the terms and conditions” and click Login.

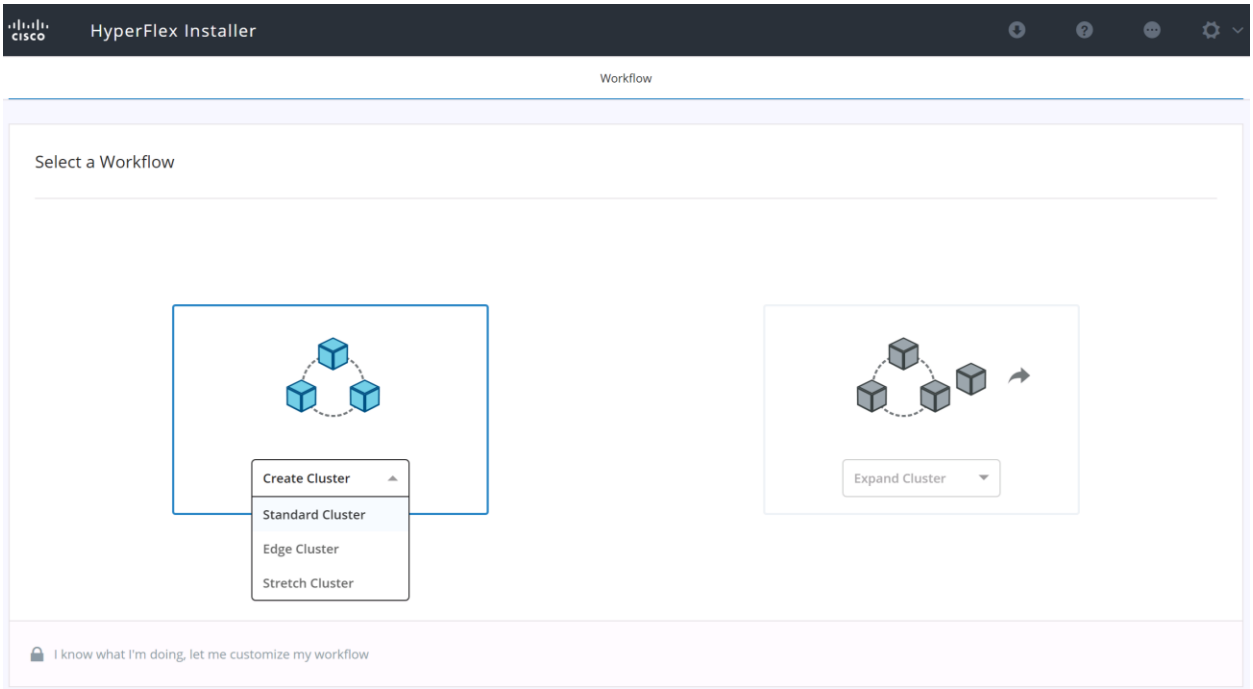
HyperFlex Installation

HyperFlex Standard Cluster Creation

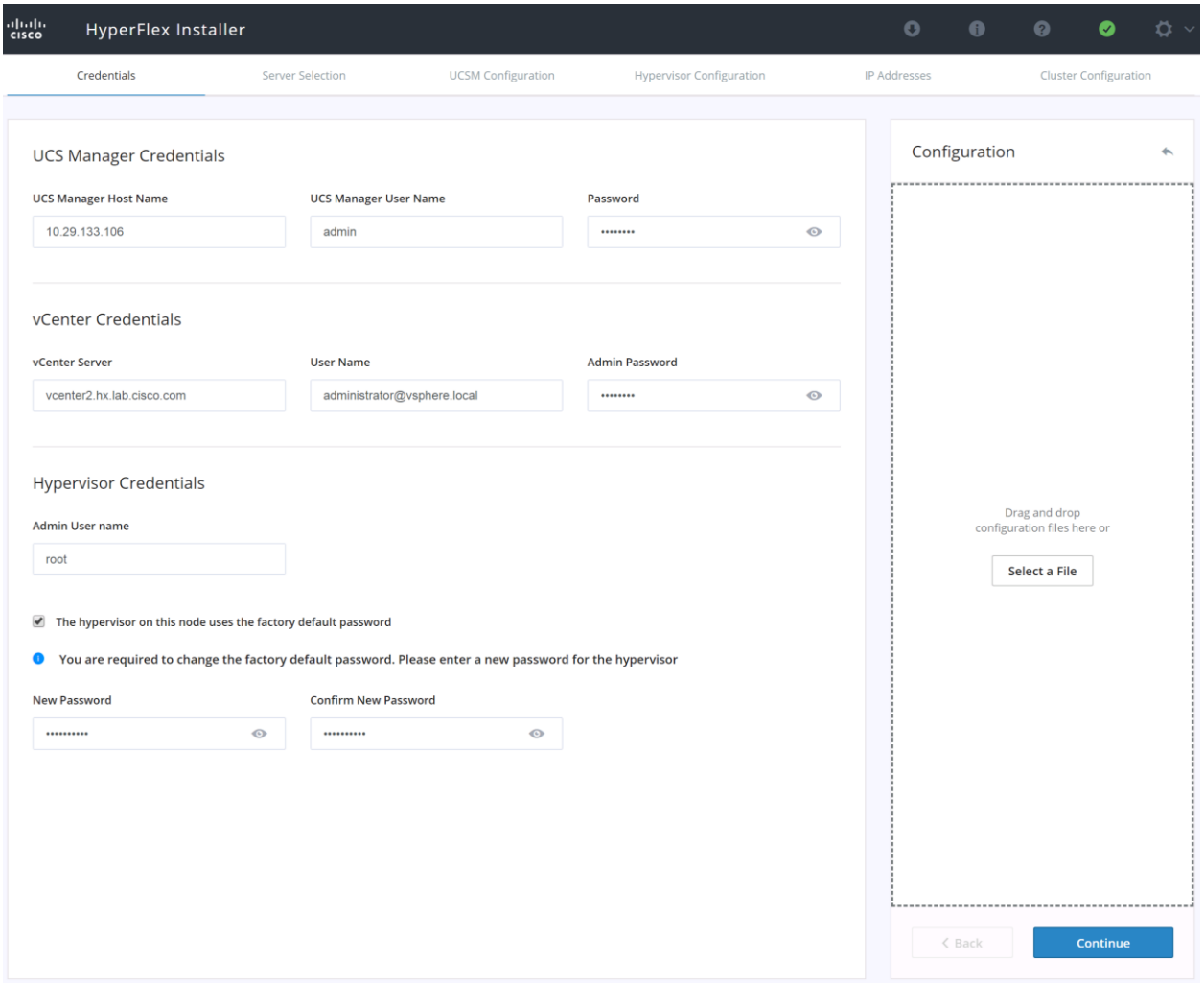
The HyperFlex installer will guide you through the process of setting up your cluster. It will configure Cisco UCS policies, templates, service profiles, and settings, as well as assigning IP addresses to the HX servers that come from the factory with ESXi hypervisor software preinstalled. The installer will deploy the HyperFlex controller VMs and software on the nodes, add the nodes to the vCenter cluster, then finally create the HyperFlex cluster and distributed filesystem. All of these processes can be completed via a single workflow from the HyperFlex Installer webpage.

To install and configure a HyperFlex cluster, complete the following steps:

1. On the HyperFlex installer webpage click the Create Cluster dropdown menu, then click Standard Cluster.



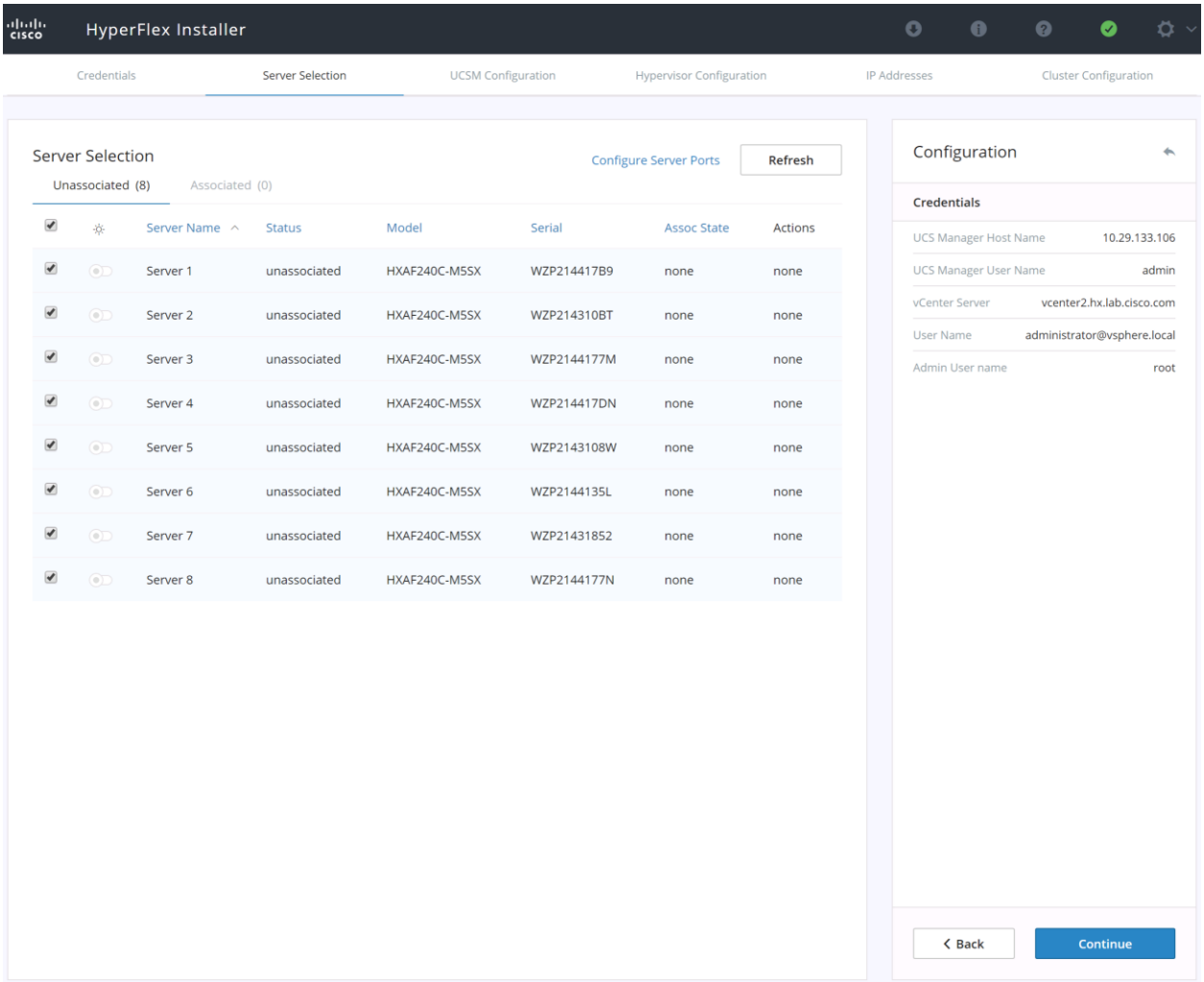
2. Enter the Cisco UCS Manager and vCenter DNS hostname or IP address, the admin usernames, and the passwords. The default Hypervisor credential which comes installed from the factory is username: root with a password of “Cisco123” and you are required to enter a new password for the Hypervisor. Optionally, you can import a JSON file that has the configuration information, except for the appropriate passwords.



3. Click Continue.

4. Select the Unassociated HX server models that are to be used in the new HX cluster and click Continue. If the Fabric Interconnect server ports were not enabled in the earlier step, you have the option to enable them here to begin the discovery process by clicking the Configure Server Ports link.

Note: Using the option to enable the server ports within the HX Installer will not allow you to finely control the server number order, as would be possible when performing this step manually before installing the HyperFlex cluster. To have control of the server number order, perform the steps outlined earlier for manually configuring the server ports. The server discovery can take several minutes to complete, and it will be necessary to periodically click the Refresh button to see the unassociated servers appear once discovery is completed.



5. Enter the VLAN names and VLAN IDs that are to be created in Cisco UCS, multiple comma-separated VLAN IDs for different guest VM networks are allowed here.

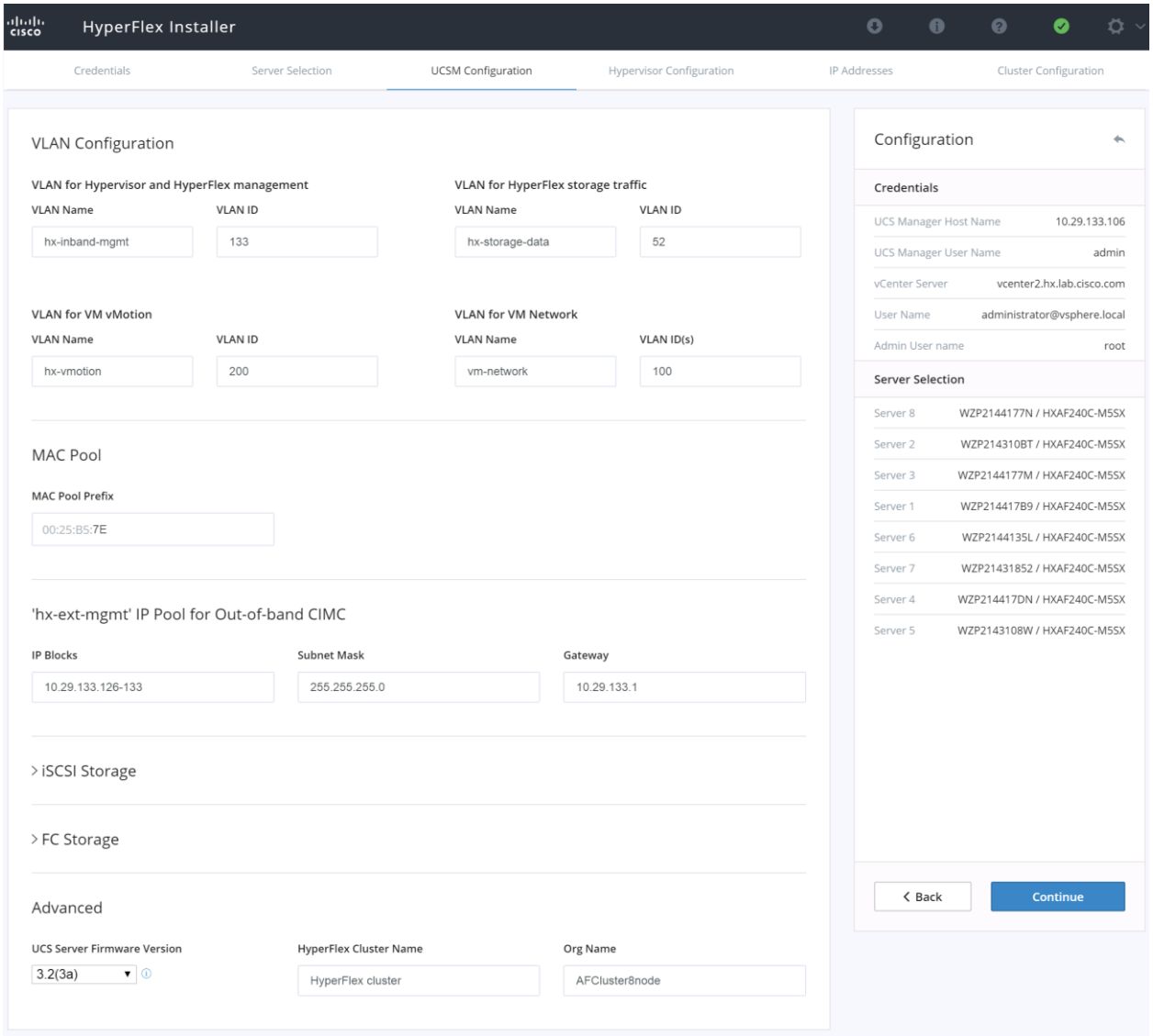
6. Enter the MAC Pool prefix, only enter the 4th byte value, for example: 00:25:B5:ED.

7. Enter the IP address range, subnet mask and gateway to be used by the CIMC interfaces of the servers in this HX cluster.

8. If multiple firmware packages exist on the Fabric Interconnect, choose the version to be installed on the servers that will comprise this cluster. Note that for M5 generation servers, the recommended version is 3.2(3a).

9. Enter a unique Org name for the HyperFlex Cluster.

Important: When deploying a second or any additional clusters, you must put them into a different sub-org, use a different MAC Pool prefix, a unique pool of IP addresses for the CIMC interfaces, and you should also create new VLAN names for the additional clusters. Even if reusing the same VLAN ID, it is prudent to create a new VLAN name to avoid conflicts. For example, for a second cluster change the VLAN names, use a unique MAC Pool prefix, IP address pool, Cluster Name and Org Name so as to not overwrite the original cluster information.

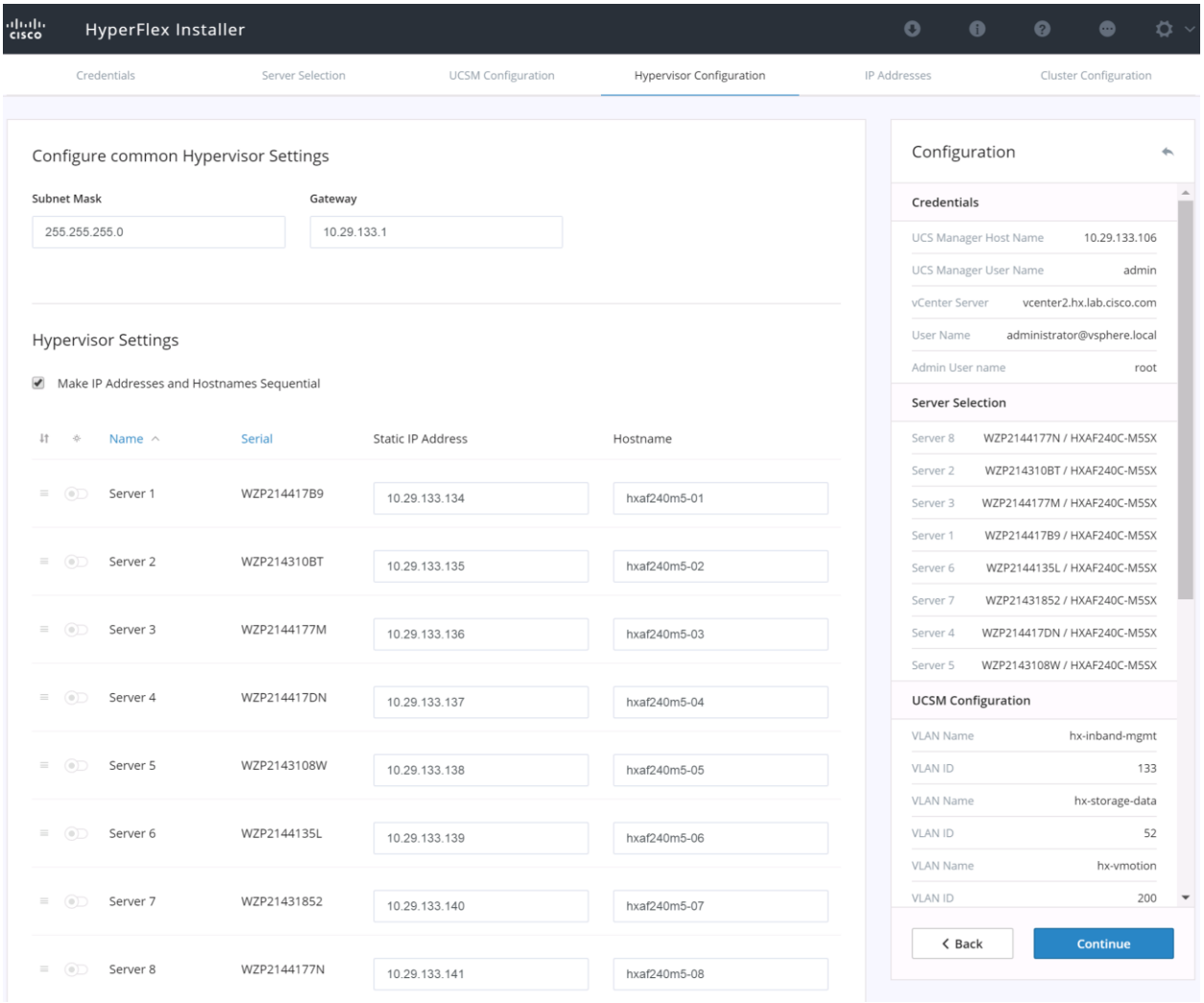


Important: (Optional) If you need to add extra iSCSI vNICs and/or FC vHBAs to connect the HX nodes to an external iSCSI or FC array, enable iSCSI Storage and/or FC Storage here using the procedure described in the following section: [Process for adding additional vHBAs or iSCSI vNICs prior to cluster creation](#).

10. Click Continue.

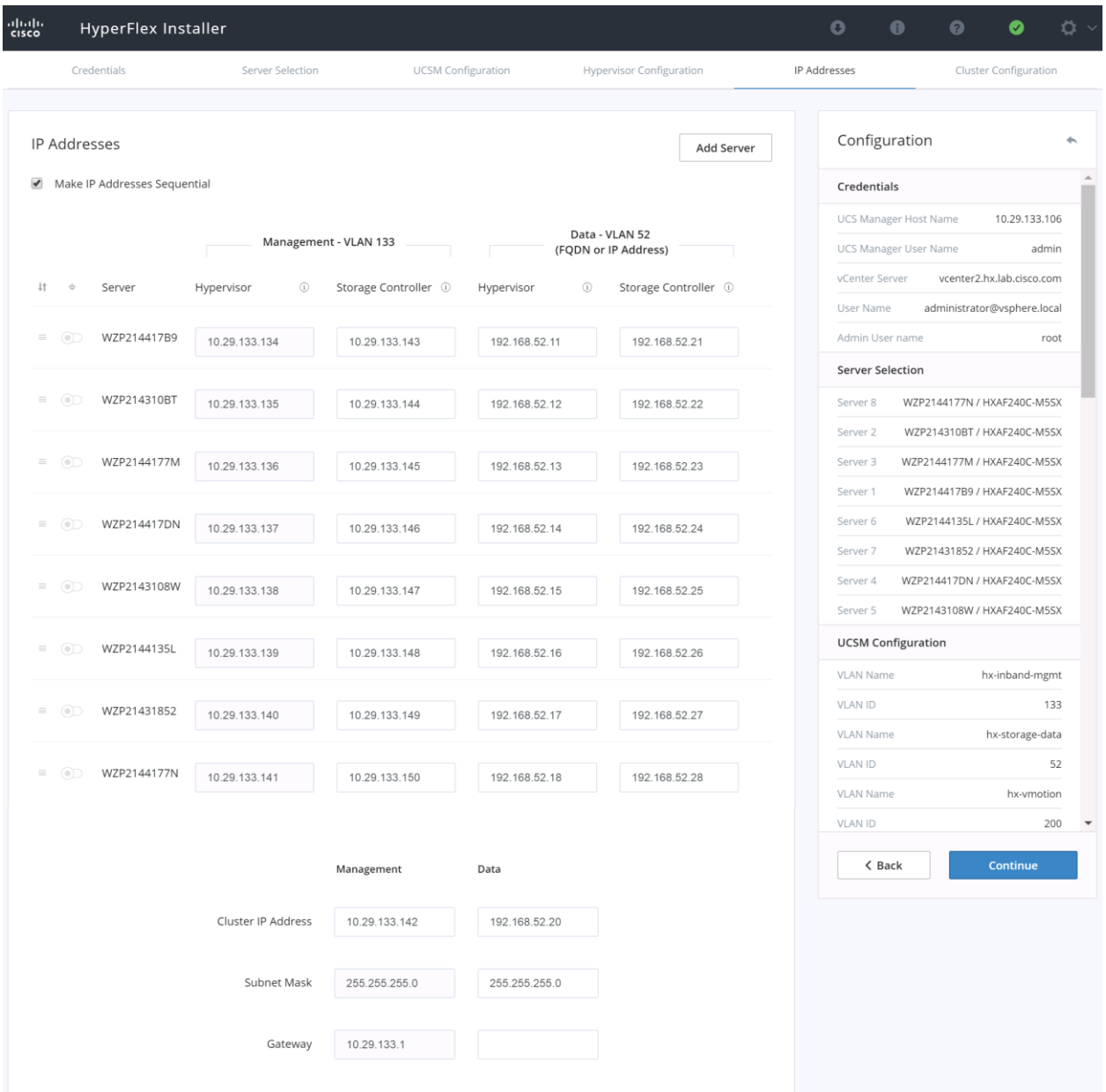
11. Enter the subnet mask, gateway, and IP addresses and hostnames for the Hypervisors. The IP addresses will be assigned via Serial over Lan (SoL) through Cisco UCS Manager to the ESXi host systems as their management IP addresses.

12. Click Continue.



13. Assign the additional IP addresses for the Management and Data networks as well as the cluster IP addresses, then click Continue.

Note: A default gateway is not required for the data network, as those interfaces normally will not communicate with any other hosts or networks, and the subnet can be non-routable.



14. Enter the HX Cluster Name and Replication Factor setting.

15. Enter the Password that will be assigned to the Controller VMs.

16. Enter the Datacenter Name from vCenter, and vCenter Cluster Name. If the Datacenter Name or Cluster Name does not exist, the installer will create them. The Cluster Name must not already exist in the vCenter inventory.

17. Enter the System Services information for DNS, NTP, and Time Zone.

18. Enable Connected Services in order to enable management via Cisco Intersight, and enter the email address to receive service ticket alerts, then scroll down.

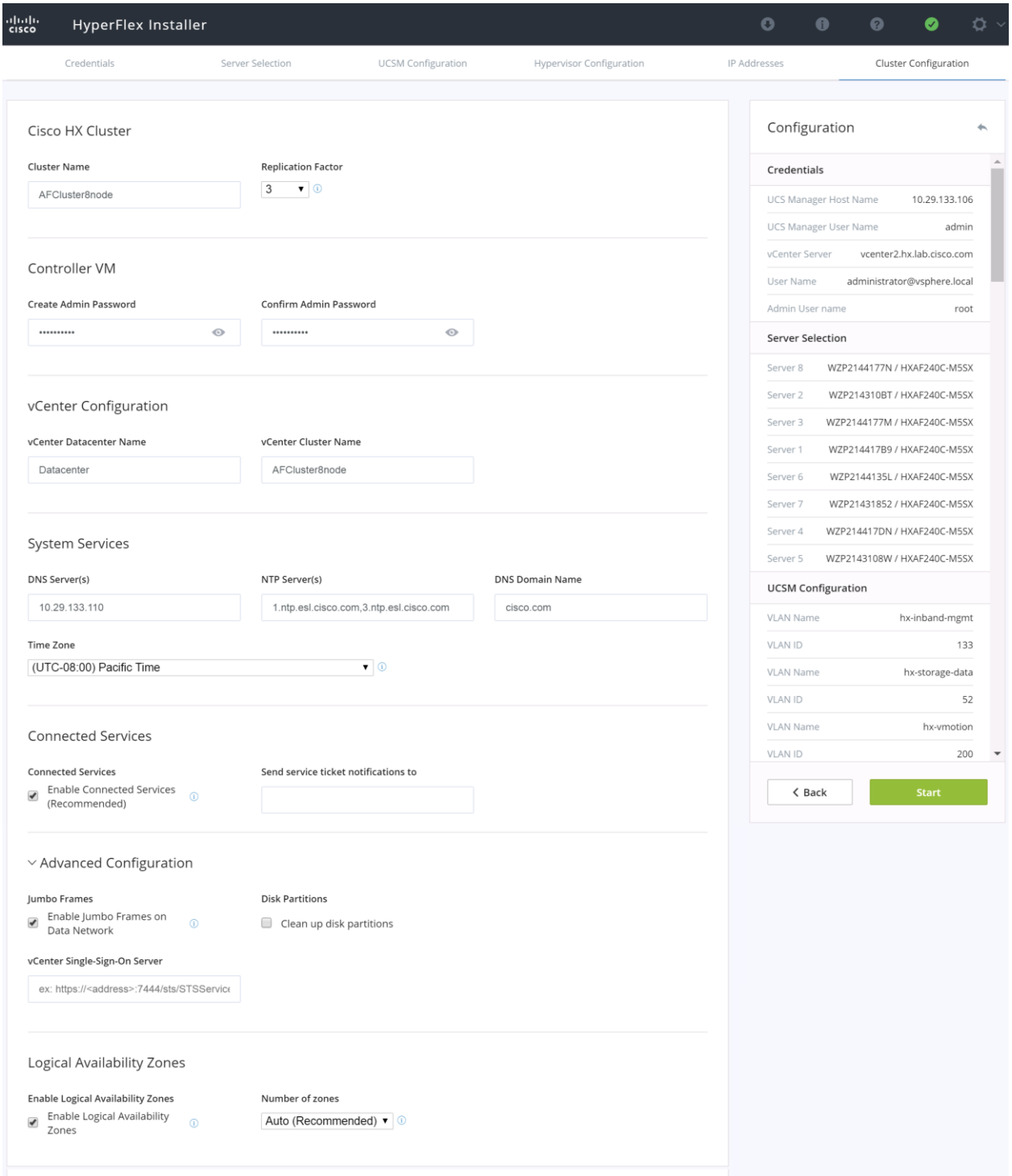
19. Under Advanced Configuration, validate that VDI is not checked (hybrid nodes only). Jumbo Frames should be enabled to ensure the best performance, unless the upstream network is not capable of being configured to transmit jumbo frames. It is not necessary to select Clean up disk partitions for a new cluster installation, but an installation using previously used converged nodes should have the option checked.

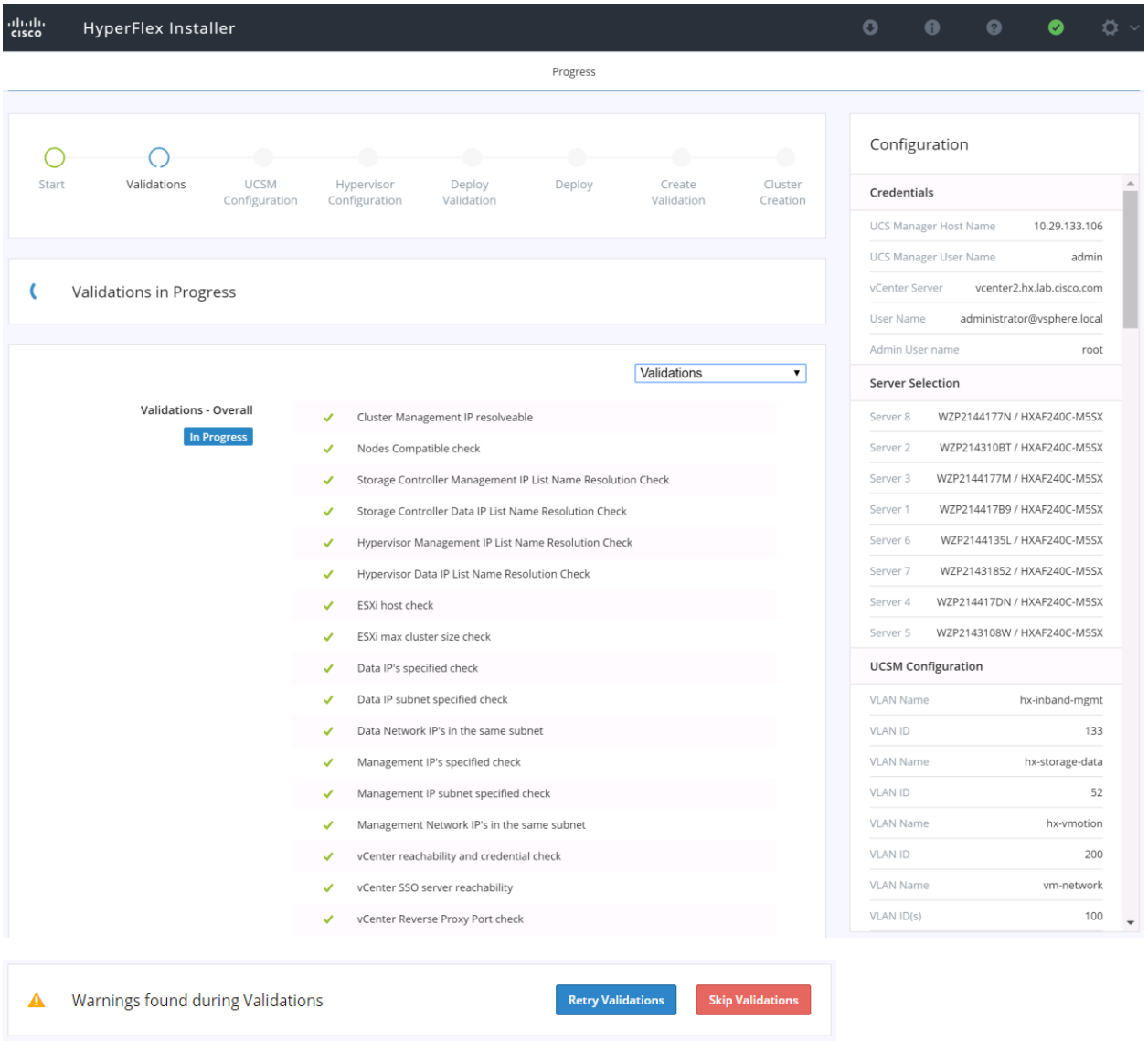
20. For clusters with 8 nodes or more, check the box to Enable Logical Availability Zones. Leave the number of zones set to the recommended Auto setting.

21. Click Start.

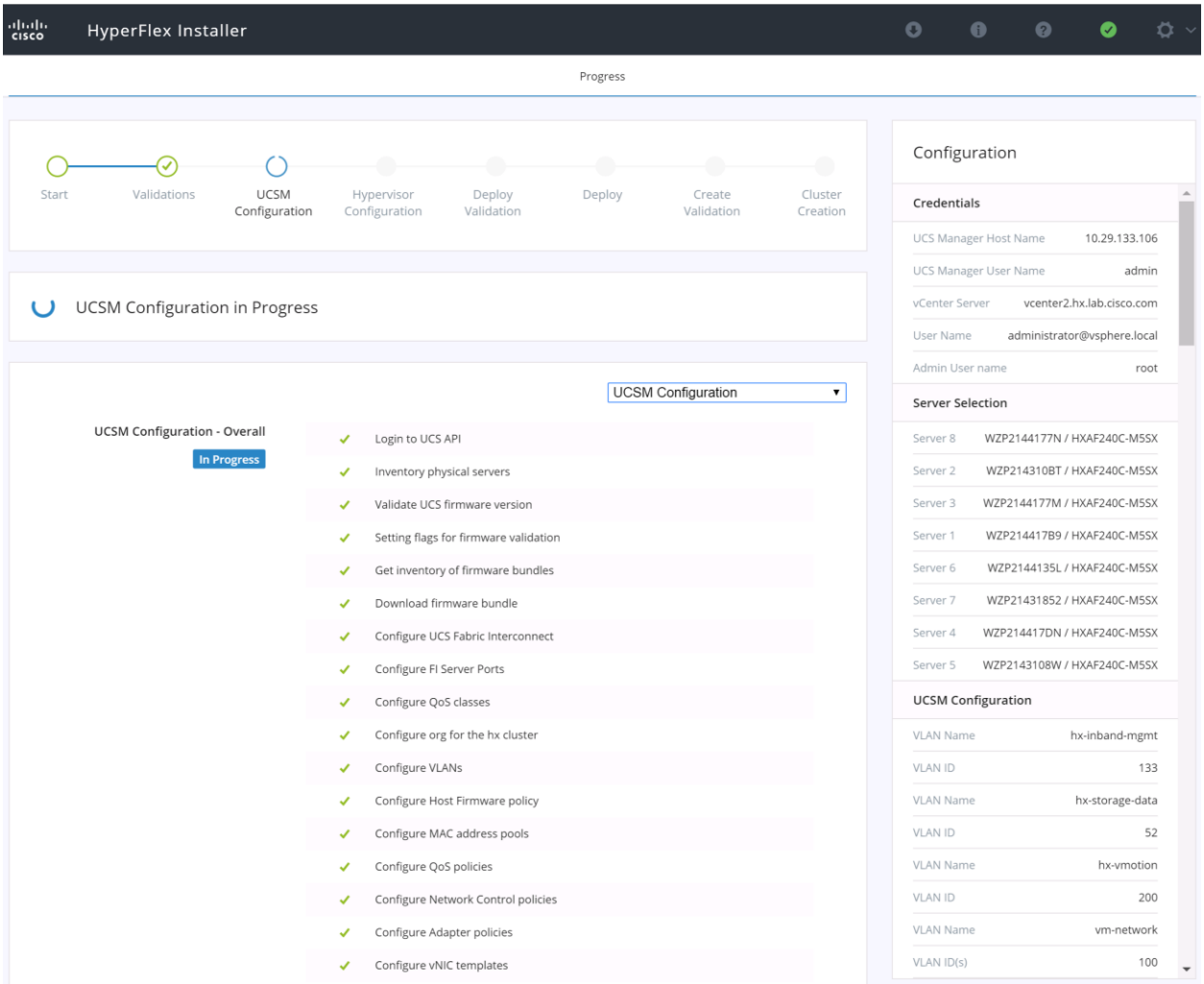
22. Validation of the configuration will now start. If there are warnings, you can review them and click "Skip Validation" if the warnings are acceptable. If there are no warnings, the installer will automatically continue on to the configuration process.

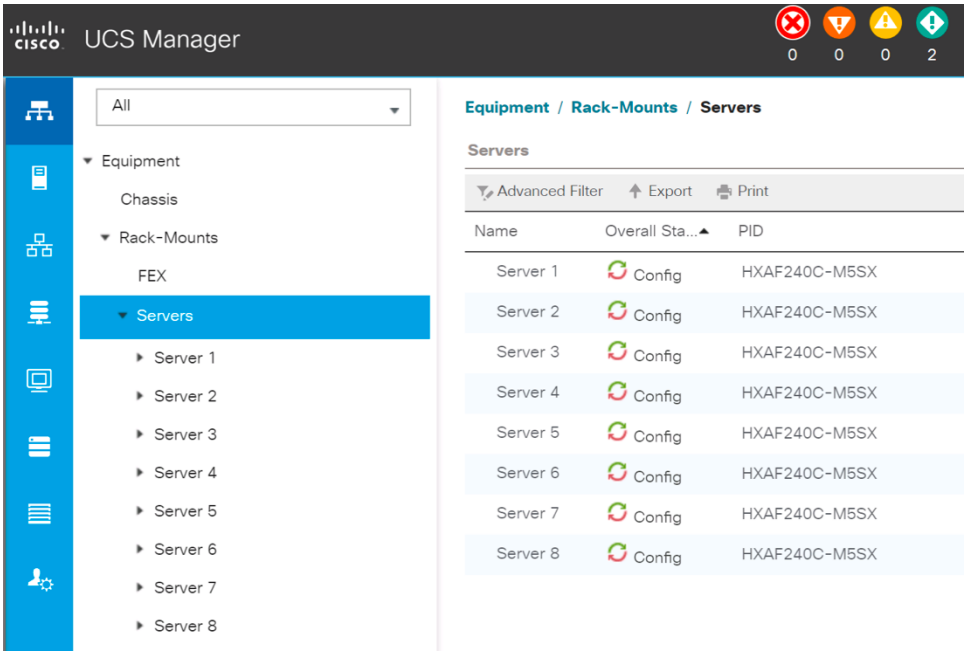
Note: The initial validation will always fail when using new Cisco UCS 6332 or 6332-16UP model Fabric Interconnects. This is due to the fact that changes to the QoS system classes require these models to reboot. If the validation is skipped, the HyperFlex installer will continue the installation and automatically reboot both Fabric Interconnects sequentially. If this is an initial setup of these Fabric Interconnects, and no other systems are running on them yet, then it is safe to proceed. However, if these Fabric Interconnects are already in use for other workloads, then caution must be taken to ensure that the sequential reboots of both Fabric Interconnects will not interrupt those workloads, and that the QoS changes will not cause traffic drops. Contact Cisco TAC for assistance if this situation applies.

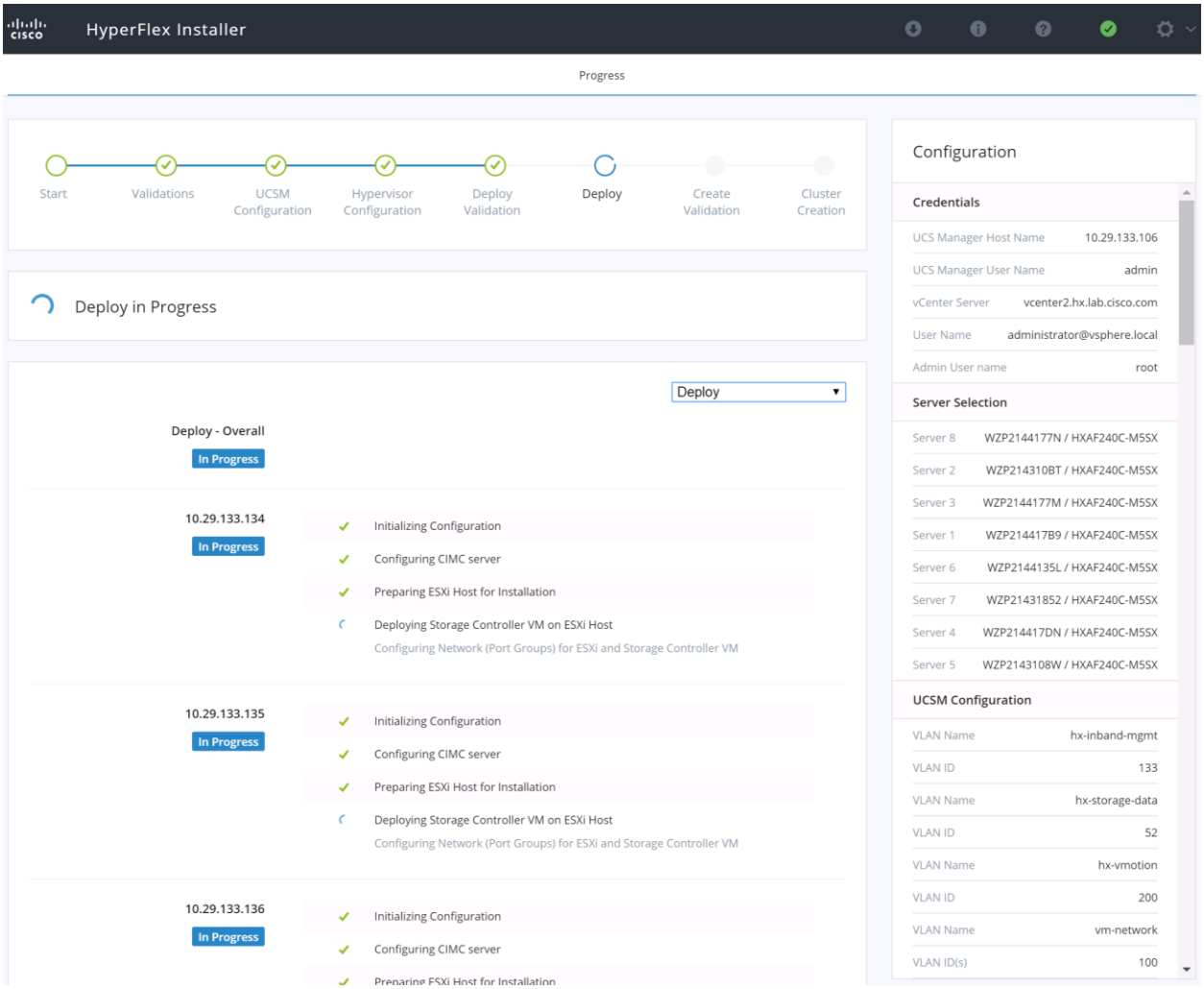

HyperFlex Installer
Credentials
Server Selection
UCSM Configuration
Hypervisor Configuration
IP Addresses
Cluster Configuration
Cisco HX Cluster
Cluster Name
AFCluster8node
Replication Factor
3
Controller VM
Create Admin Password
Confirm Admin Password
vCenter Configuration
vCenter Datacenter Name
Datacenter
vCenter Cluster Name
AFCluster8node
System Services
DNS Server(s)
10.29.133.110
NTP Server(s)
1.ntp.esl.cisco.com,3.ntp.esl.cisco.com
DNS Domain Name
cisco.com
Time Zone
(UTC-08:00) Pacific Time
Connected Services
Connected Services
Enable Connected Services (Recommended)
Send service ticket notifications to
Advanced Configuration
Jumbo Frames
Enable Jumbo Frames on Data Network
Disk Partitions
Clean up disk partitions
vCenter Single-Sign-On Server
ex: https://<address>:7444/sts/STSServic
Logical Availability Zones
Enable Logical Availability Zones
Enable Logical Availability Zones
Number of zones
Auto (Recommended)
Configuration
Credentials
UCS Manager Host Name 10.29.133.106
UCS Manager User Name admin
vCenter Server vcenter2.hx.lab.cisco.com
User Name administrator@vsphere.local
Admin User name root
Server Selection
Server 8 WZP2144177N / HXAF240C-M5SX
Server 2 WZP214310BT / HXAF240C-M5SX
Server 3 WZP2144177M / HXAF240C-M5SX
Server 1 WZP214417B9 / HXAF240C-M5SX
Server 6 WZP2144135L / HXAF240C-M5SX
Server 7 WZP21431852 / HXAF240C-M5SX
Server 4 WZP214417DN / HXAF240C-M5SX
Server 5 WZP2143108W / HXAF240C-M5SX
UCSM Configuration
VLAN Name hx-inband-mgmt
VLAN ID 133
VLAN Name hx-storage-data
VLAN ID 52
VLAN Name hx-vmotion
VLAN ID 200
Back
Start



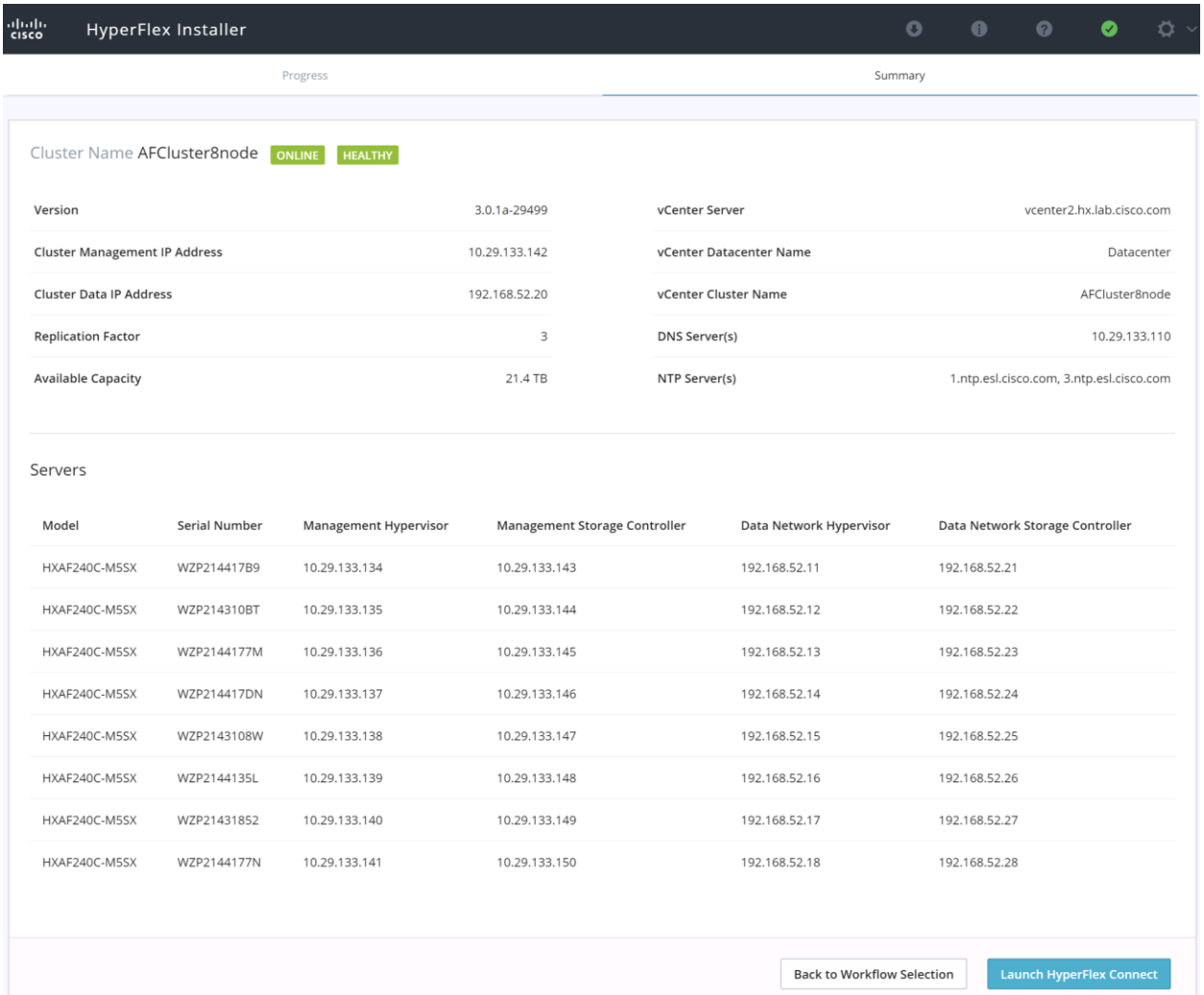
23. After the pre-installation validations, the HX installer will proceed to complete the deployment and perform all the steps listed at the top of the screen along with their status. The process can also be monitored in Cisco UCS Manager and vCenter while the profiles and cluster are created.


HyperFlex Installer
Progress
Start
Validations
UCSM Configuration
Hypervisor Configuration
Deploy Validation
Deploy
Create Validation
Cluster Creation
UCSM Configuration in Progress
UCSM Configuration
UCSM Configuration - Overall
In Progress
Login to UCS API
Inventory physical servers
Validate UCS firmware version
Setting flags for firmware validation
Get inventory of firmware bundles
Download firmware bundle
Configure UCS Fabric Interconnect
Configure FI Server Ports
Configure QoS classes
Configure org for the hx cluster
Configure VLANs
Configure Host Firmware policy
Configure MAC address pools
Configure QoS policies
Configure Network Control policies
Configure Adapter policies
Configure vNIC templates
Configuration
Credentials
UCS Manager Host Name 10.29.133.106
UCS Manager User Name admin
vCenter Server vcenter2.hx.lab.cisco.com
User Name administrator@vsphere.local
Admin User name root
Server Selection
Server 8 WZP2144177N / HXAF240C-M5SX
Server 2 WZP214310BT / HXAF240C-M5SX
Server 3 WZP2144177M / HXAF240C-M5SX
Server 1 WZP214417B9 / HXAF240C-M5SX
Server 6 WZP2144135L / HXAF240C-M5SX
Server 7 WZP21431852 / HXAF240C-M5SX
Server 4 WZP214417DN / HXAF240C-M5SX
Server 5 WZP2143108W / HXAF240C-M5SX
UCSM Configuration
VLAN Name hx-inband-mgmt
VLAN ID 133
VLAN Name hx-storage-data
VLAN ID 52
VLAN Name hx-vmotion
VLAN ID 200
VLAN Name vm-network
VLAN ID(s) 100

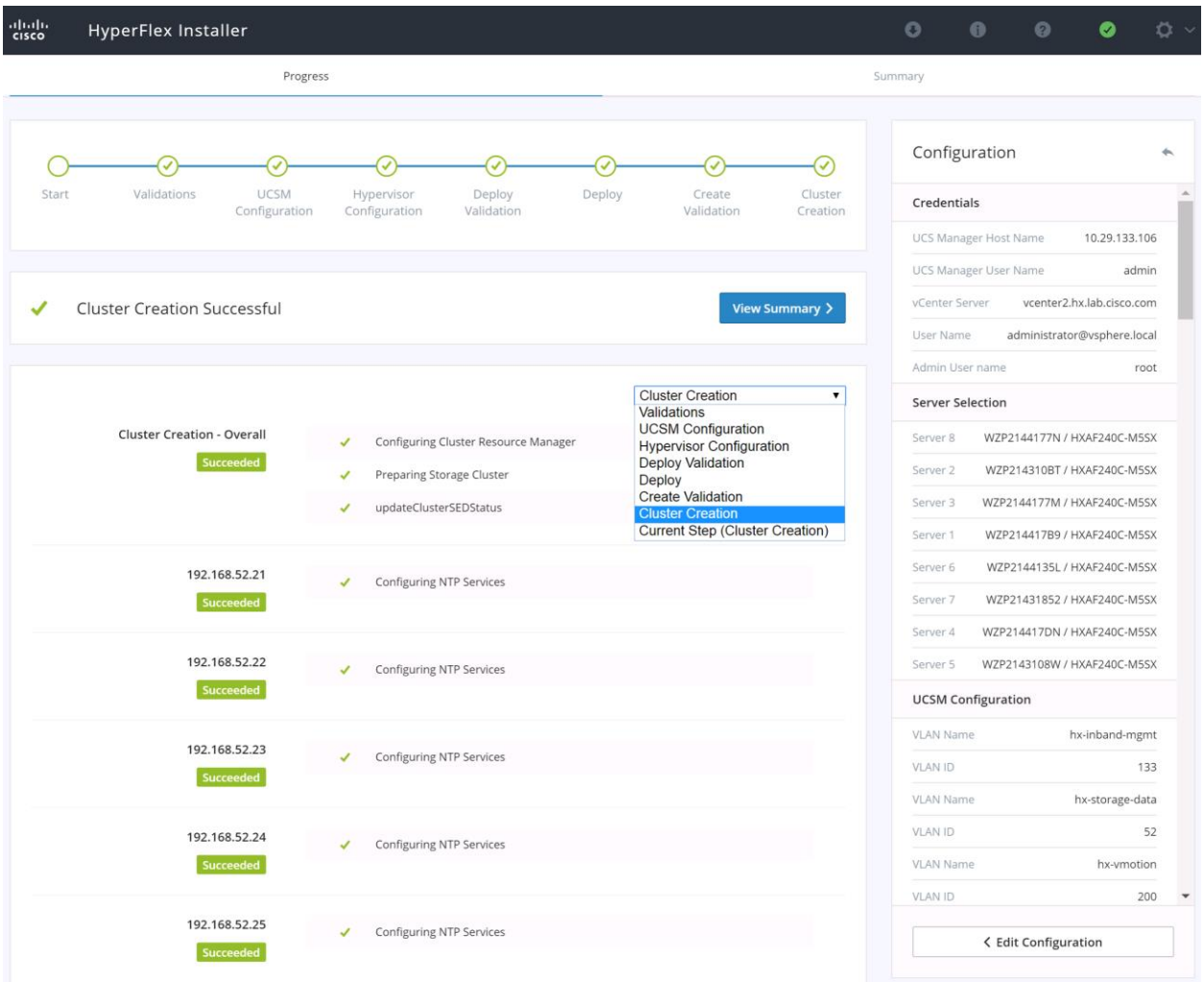

UCS Manager
0 0 0 2
All
Equipment
Chassis
Rack-Mounts
FEX
Servers
Server 1
Server 2
Server 3
Server 4
Server 5
Server 6
Server 7
Server 8
Equipment / Rack-Mounts / Servers
Servers
Advanced Filter
Export
Print
Name
Overall Sta...
PID
Server 1 Config HXAF240C-M5SX
Server 2 Config HXAF240C-M5SX
Server 3 Config HXAF240C-M5SX
Server 4 Config HXAF240C-M5SX
Server 5 Config HXAF240C-M5SX
Server 6 Config HXAF240C-M5SX
Server 7 Config HXAF240C-M5SX
Server 8 Config HXAF240C-M5SX



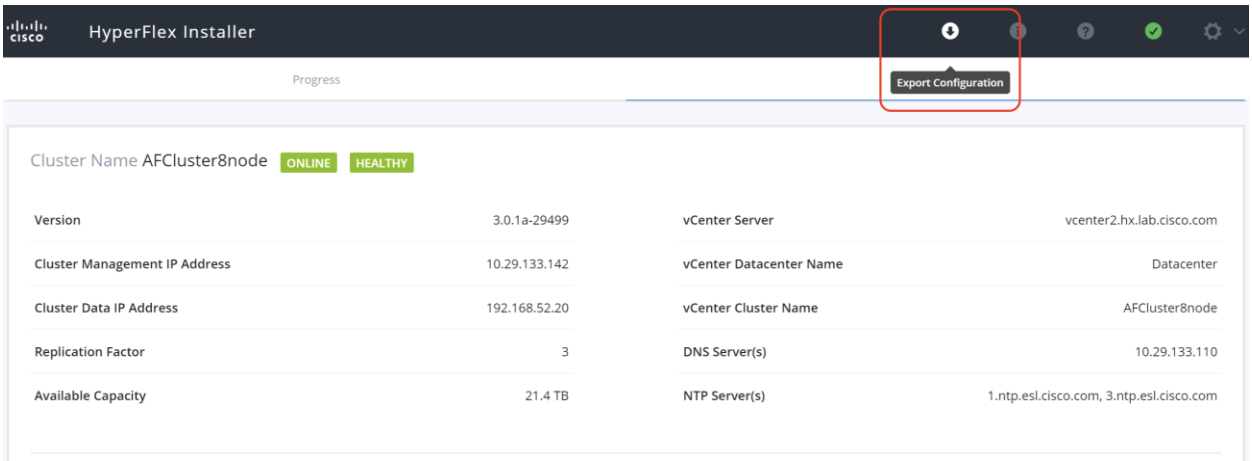
24. Review the Summary screen after the install completes by selecting Summary on the top right of the window.



25. You can also review the details of the installation process after the install completes by selecting Progress on the top left of the window.



26. After the install completes, you may export the cluster configuration by clicking on the downward arrow icon in the top right of the screen. Click OK to save the configuration to a JSON file. This file can be imported to save time if you need to rebuild the same cluster in the future, and be kept as a record of the configuration options and settings used during the installation.



27. Continue with the Post Installation tasks in the next section. It is particularly important to run the post_install script in order to create the vMotion interfaces, the guest VM port groups, and to enable HA and DRS in the cluster.

28. After the installation completes, you can click the Launch HyperFlex Connect button to immediately log in to the HTML5 management GUI.

HyperFlex Stretched Cluster Creation

The HyperFlex installer will guide you through the process of setting up your stretched cluster. It will configure Cisco UCS policies, templates, service profiles, and settings for two separate UCS domains, assign IP addresses to the HX servers that come from the factory with ESXi hypervisor software preinstalled, and deploy the HyperFlex cluster. As with installation of a standard HyperFlex cluster, pre-installation of the physical components, networking configuration, and assembling the necessary IP addressing and configuration information is required.

As an overview of the differences from a standard cluster installation, the following list is provided to assist with the proper order of configuration for a stretched cluster:

1. Physically install and cable the Cisco UCS Fabric Interconnects and the Cisco HX-series servers in site # 1.

2. Physically install and cable the Cisco UCS Fabric Interconnects and the Cisco HX-series servers in site # 2.

3. Configure the upstream networking, ensuring the HyperFlex management and HyperFlex storage VLANs are available in both sites and presented to the Cisco UCS Fabric Interconnects, with Layer 2 adjacency, along with the required bandwidth and latency limits.

4. Gather the required IP addressing and configuration information for the two sites. Stretched clusters require some additional IP addresses for the witness server VM, and the additional pair of Cisco UCS Fabric Interconnects.

5. For stretched clusters it is important to have redundant network services, such as DNS and NTP. You must list at least two DNS and two NTP servers, and at least one of the two servers must be accessible from each site. You must ensure that at least one of the listed DNS and NTP servers will remain accessible by the remaining online HyperFlex stretched cluster nodes, should a site go offline.

6. Perform the initial Cisco UCS Manager configuration in both sites, including uplink configuration, NTP, and server discovery.

7. Deploy the witness server VM to an ESXi server or cluster in the third site.

8. Perform the HyperFlex UCS Configuration in site # 1.

9. Perform the HyperFlex UCS Configuration in site # 2.

10. Perform the HyperFlex Cluster installation.

11. Run the post_install script on the HyperFlex installer VM, in order to enable HA and DRS, configure the guest VM port groups, and configure the vMotion interfaces.

Witness Server Installation

Prior to deploying the witness server VM, complete steps 1-6 above to prepare the two sites for the stretched cluster installation. To deploy the HyperFlex stretched cluster witness VM OVA, complete the following steps:

1. Open the vSphere Web Client webpage, or the vSphere HTML5 Web Client webpage of a vCenter server where the witness OVA will be deployed, and log in with admin privileges.

2. In the vSphere Web Client, from the Home view, click Hosts and Clusters.

3. From the Actions menu, click Deploy OVF Template.

4. Click the Local file option, then click Browse and locate the *HyperFlex-Witness-1.0.1.ova* file, click the file and click Open.

5. Click Next.

6. Modify the name of the virtual machine to be created if desired, and click a folder location to place the virtual machine, then click Next.

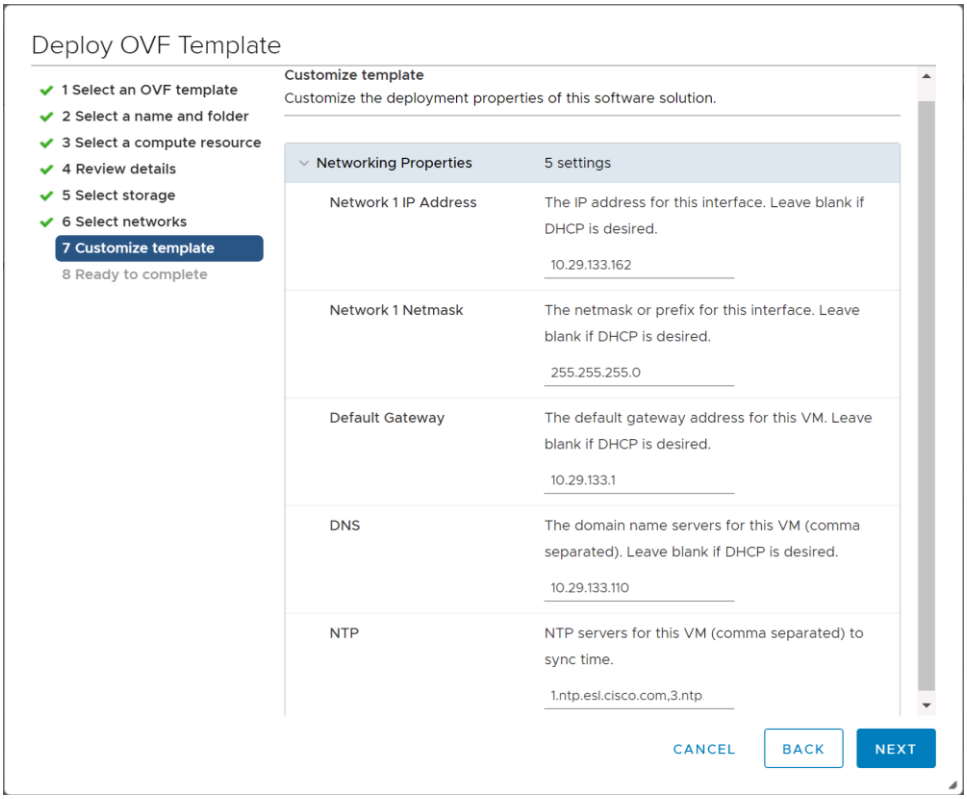
7. Click a specific host or cluster to locate the virtual machine and click Next.

8. After the file validation, review the details and click Next.

9. Select a Thin provision virtual disk format, and the datastore to store the new virtual machine, then click Next.

10. Modify the network port group selection from the drop-down list in the Destination Networks column, choosing the network the witness VM will communicate on, and click Next.

11. Enter the static IP address settings to be used, fill in the fields for the DNS server, Default Gateway, NTP Servers, IP address, and subnet mask, then click Next.



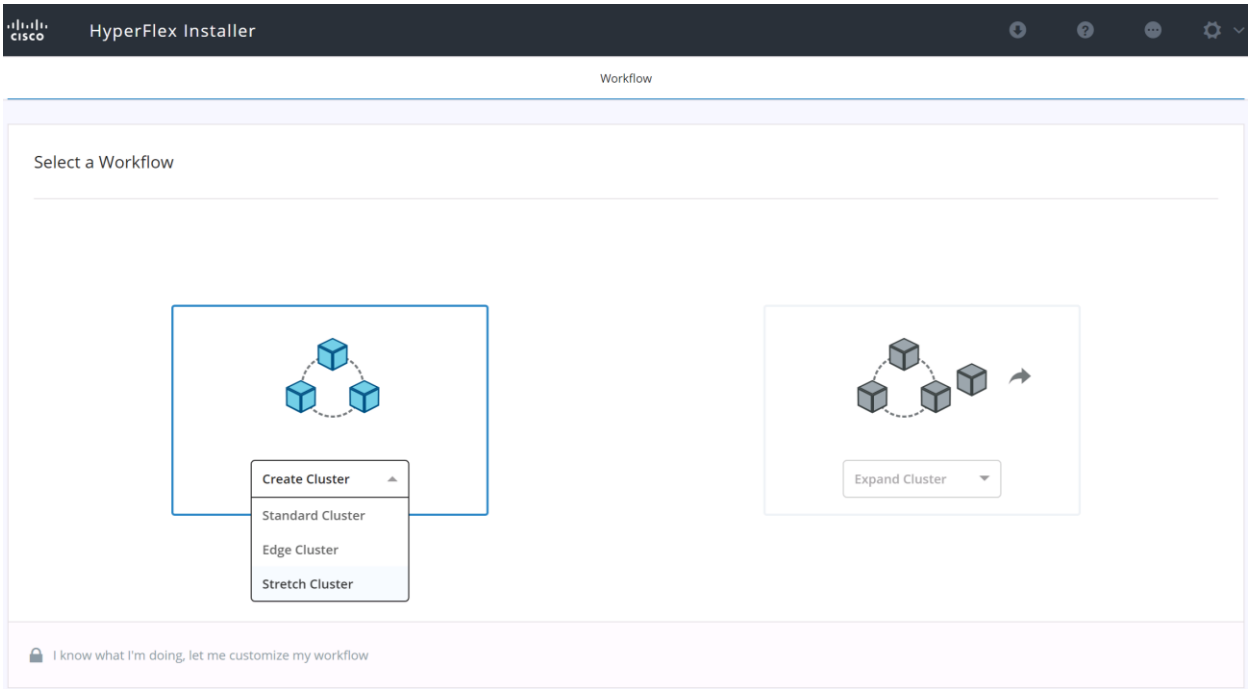
12. Review the final configuration and click Finish.

13. The witness VM will take a few minutes to deploy, once it has deployed, power on the new VM and proceed to the next step.

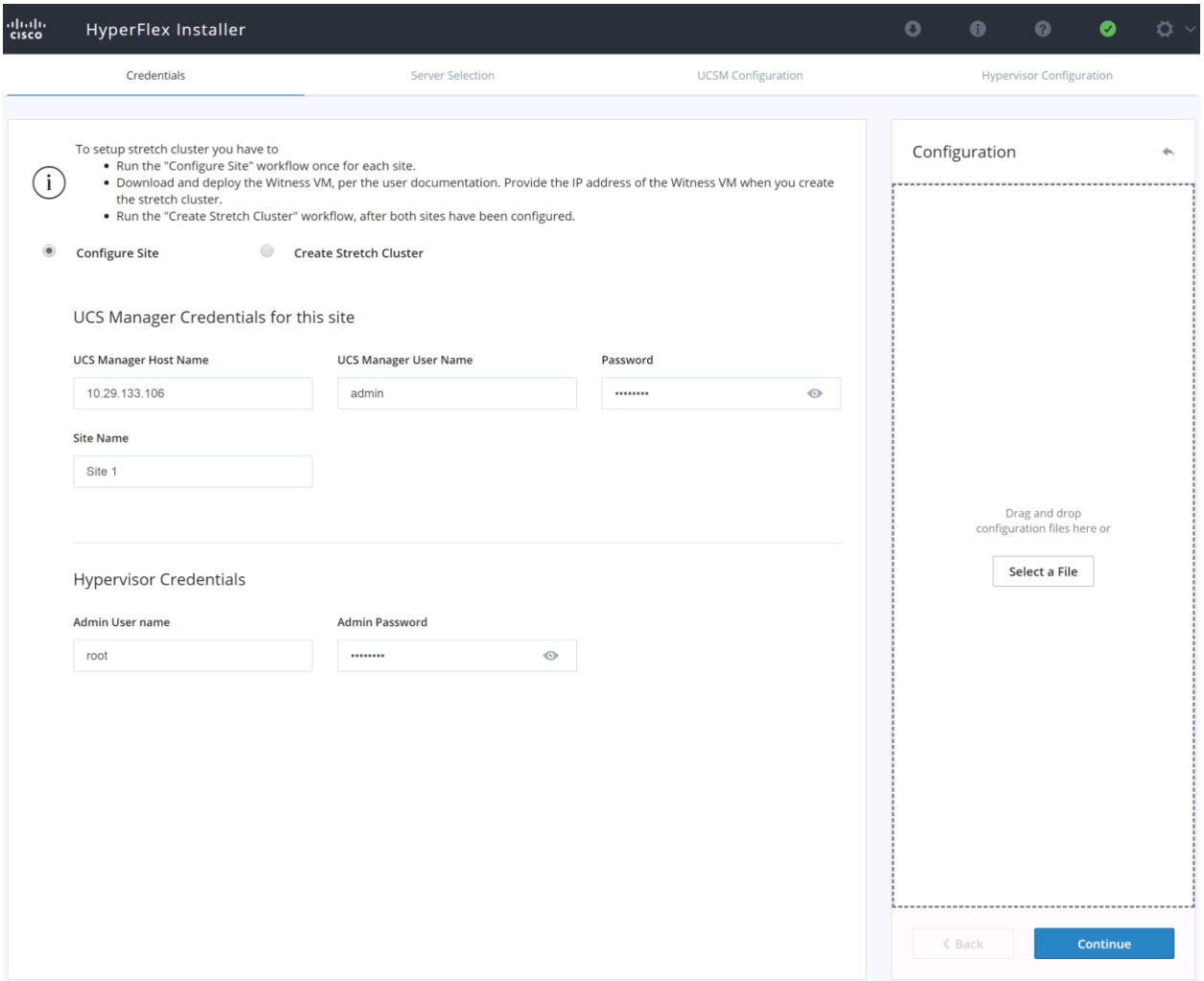
Stretched Cluster Installation

To install and configure a HyperFlex stretched cluster, complete the following steps:

1. On the HyperFlex installer webpage click the Create Cluster dropdown menu, then click Stretch Cluster.



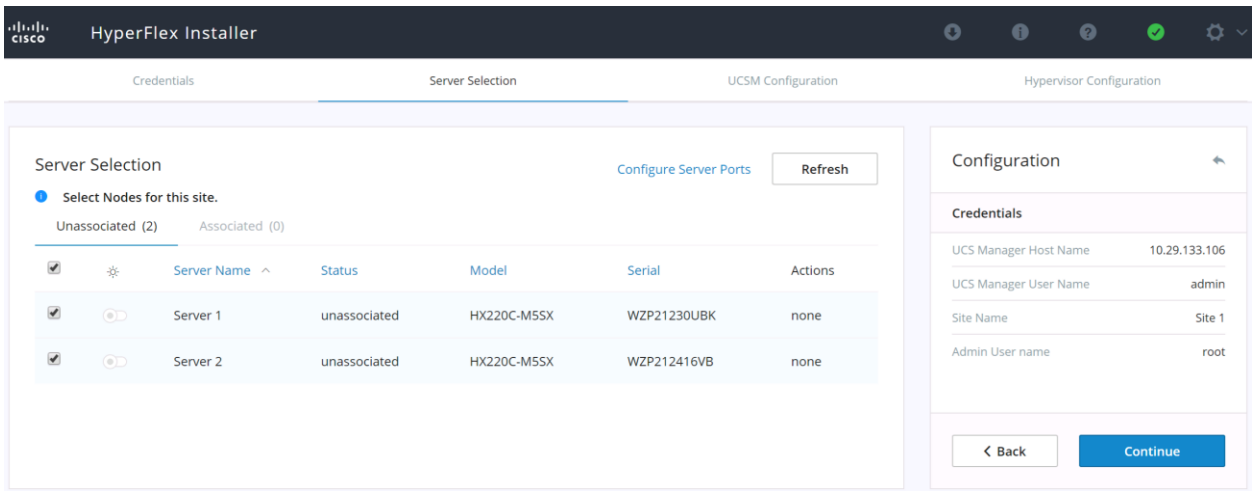
2. Click the radio button for Configure Site, in order to configure the first physical site for the new cluster. Enter the Cisco UCS Manager DNS hostname or IP address, the admin username and the password. Provide a name for the site being configured. The site name will be the name of the physical site for this cluster as seen in the Cisco HyperFlex Connect management webpage. The default Hypervisor credential which comes installed from the factory is username: root with a password of “Cisco123”. Optionally, you can import a JSON file that has the configuration information, except for the appropriate passwords.



3. Click Continue.

4. Select the Unassociated HX server models that are to be used in the new HX cluster for this site and click Continue. If the Fabric Interconnect server ports were not enabled in the earlier step, you have the option to enable them here to begin the discovery process by clicking the Configure Server Ports link.

Note: Using the option to enable the server ports within the HX Installer will not allow you to finely control the server number order, as would be possible when performing this step manually before installing the HyperFlex cluster. To have control of the server number order, perform the steps outlined earlier for manually configuring the server ports. The server discovery can take several minutes to complete, and it will be necessary to periodically click the Refresh button to see the unassociated servers appear once discovery is completed.



5. Enter the VLAN names and VLAN IDs that are to be created in Cisco UCS, multiple comma-separated VLAN IDs for different guest VM networks are allowed here.

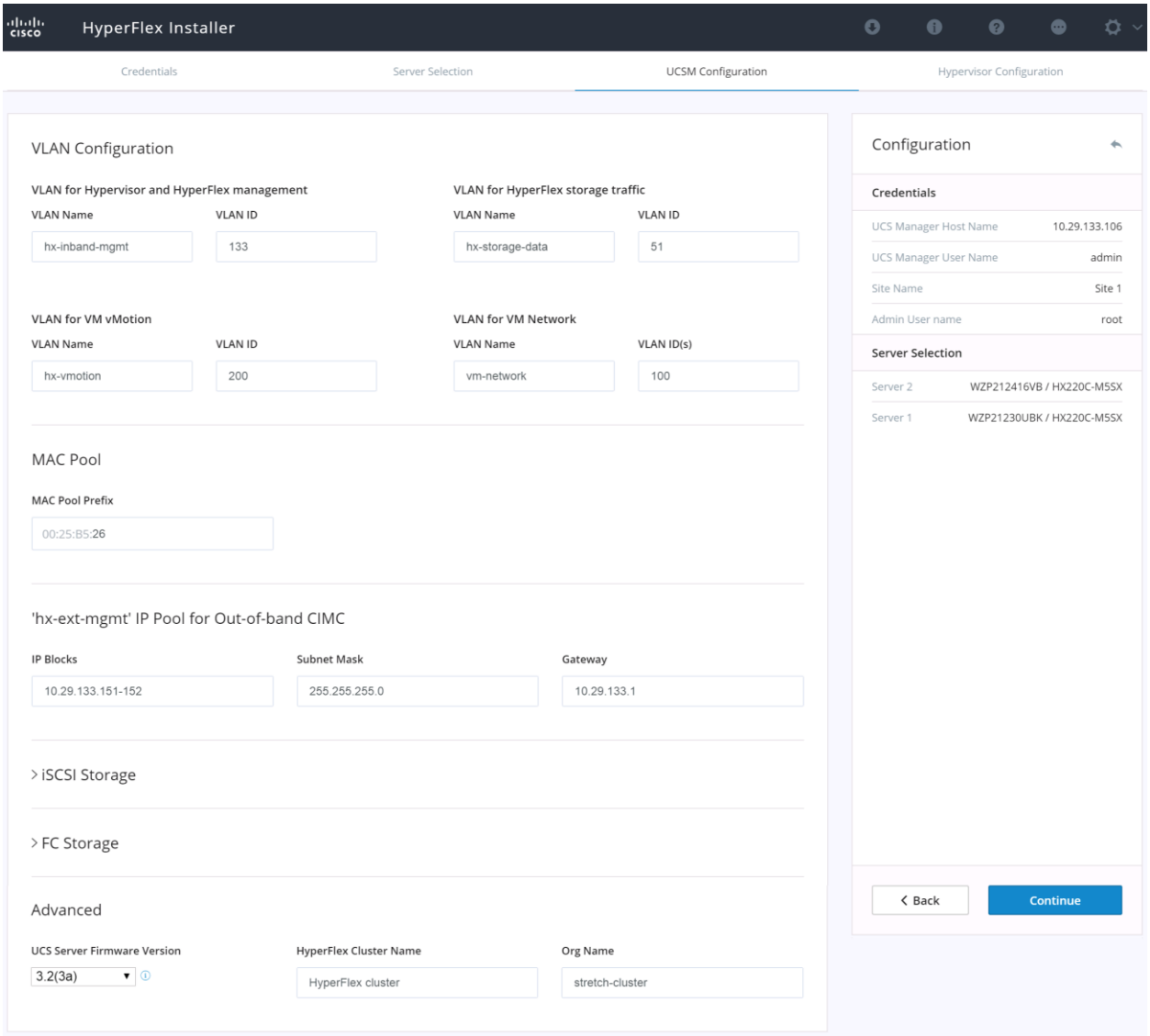
6. Enter the MAC Pool prefix, only enter the 4th byte value, for example: 00:25:B5:ED. This prefix must be unique from any other clusters in the same Cisco UCS domain, and also unique from the prefix used by the second site in this stretched cluster.

7. Enter the IP address range, subnet mask and gateway to be used by the CIMC interfaces of the servers in this HX cluster. This range must be unique from any other clusters in the same Cisco UCS domain, and also unique from the range used by the second site in this stretched cluster.

8. If multiple firmware packages exist on the Fabric Interconnect, choose the version to be installed on the servers that will comprise this cluster. Note that for M5 generation servers, the recommended version is 3.2(3a).

9. Enter a unique Org name for the HyperFlex Cluster. The org name can be the same in both sites of the stretched cluster, because they are in different UCS domains, but must be unique from any other clusters deployed in the same UCS domain.

Important: When deploying a second or any additional clusters, you must put them into a different sub-org, use a different MAC Pool prefix, and you should also create new VLAN names for the additional clusters. Even if reusing the same VLAN ID, it is prudent to create a new VLAN name to avoid conflicts. For example, for a second cluster change the VLAN names, MAC Pool prefix, Cluster Name and Org Name so as to not overwrite the original cluster information.

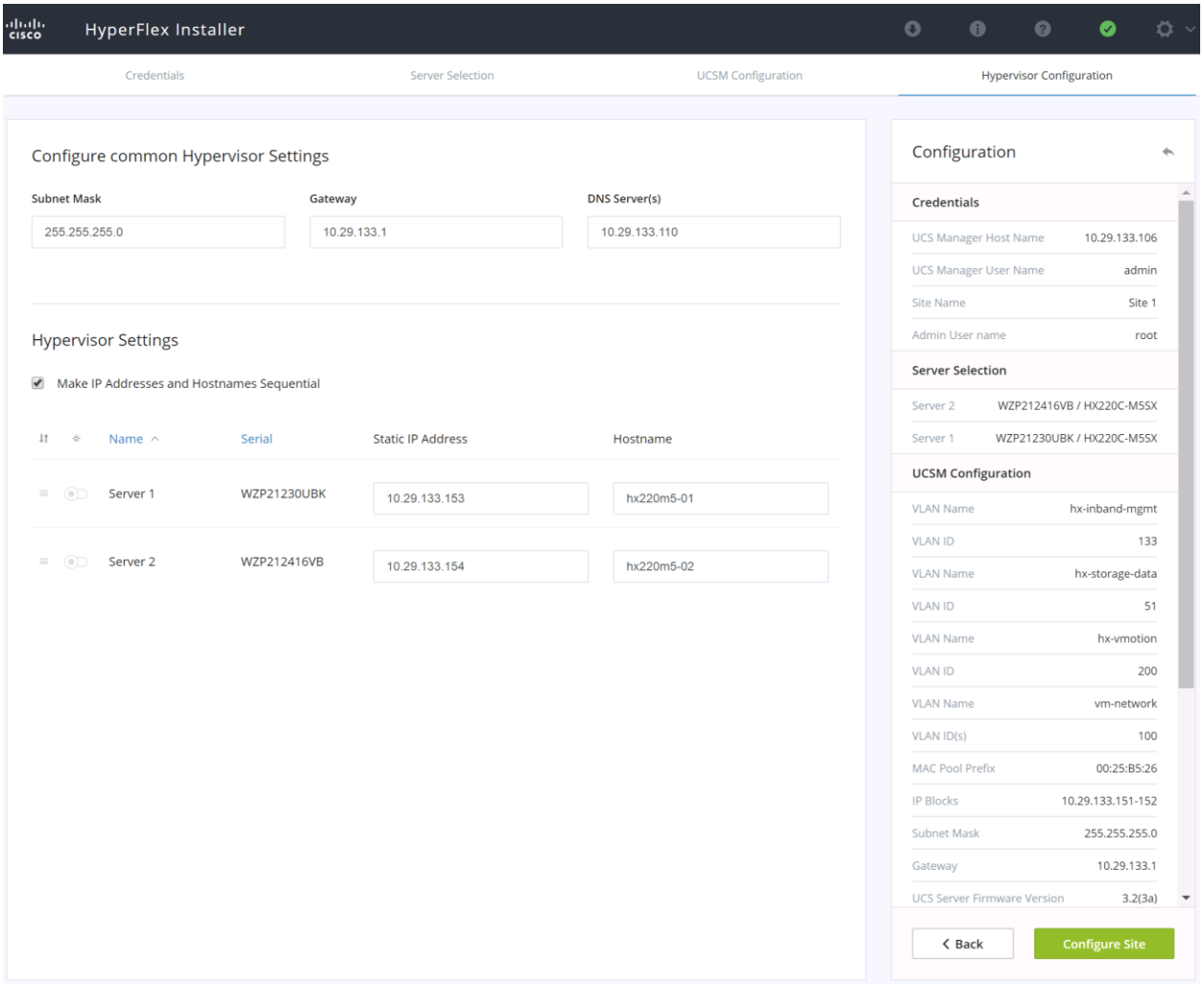


Important: (Optional) If you need to add extra iSCSI vNICs and/or FC vHBAs to connect the HX nodes to an external iSCSI or FC array, enable iSCSI Storage and/or FC Storage here using the procedure described in the following section: Process for adding additional vHBAs or iSCSI vNICs prior to cluster creation.

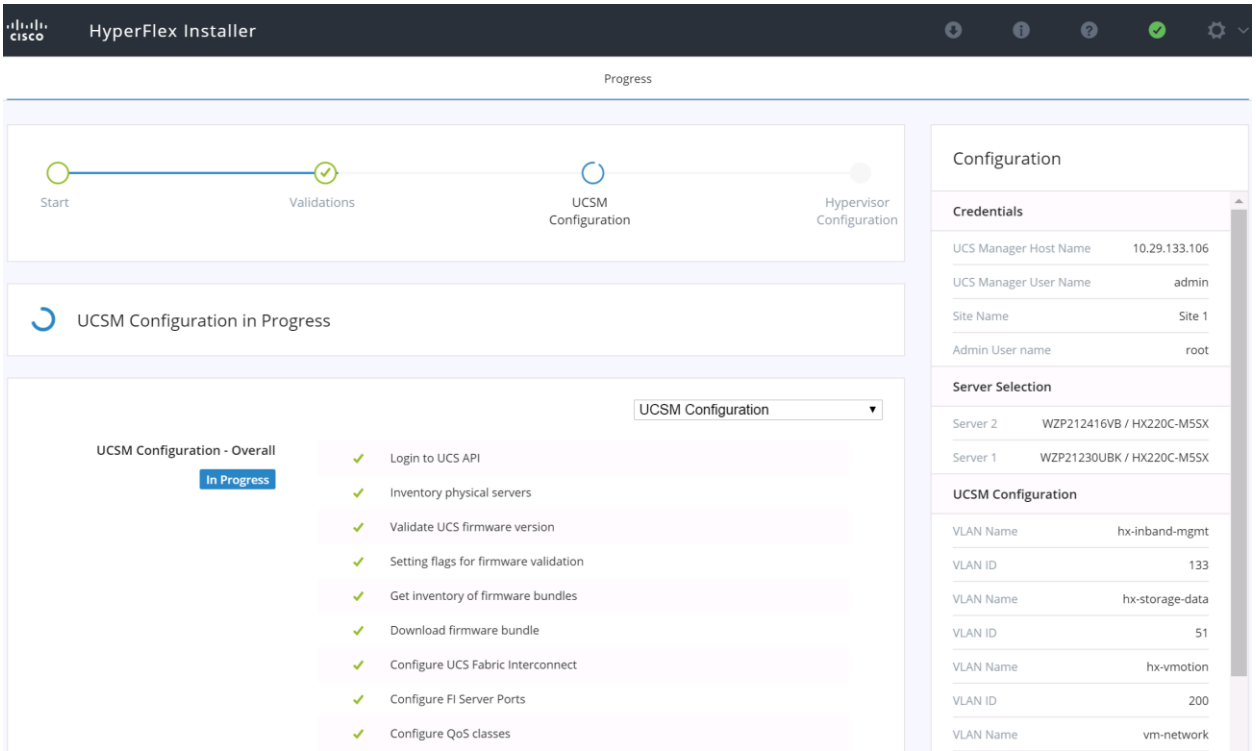
10. Click Continue.

11. Enter the subnet mask, gateway, and IP addresses and hostnames for the Hypervisors. The IP addresses will be assigned via Serial over Lan (SoL) through Cisco UCS Manager to the ESXi host systems as their management IP addresses.

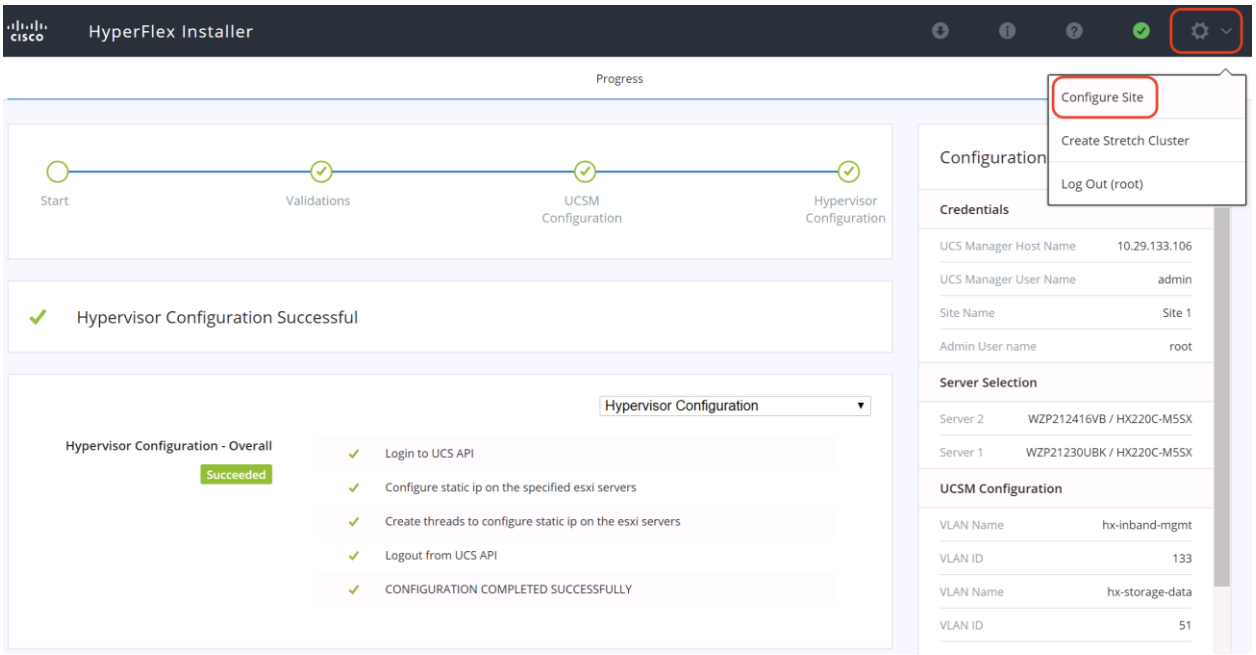
12. Click Configure Site.



13. The installer will validate the configuration options chosen, then perform the UCSM and Hypervisor Configuration steps for this first site of the stretched cluster.

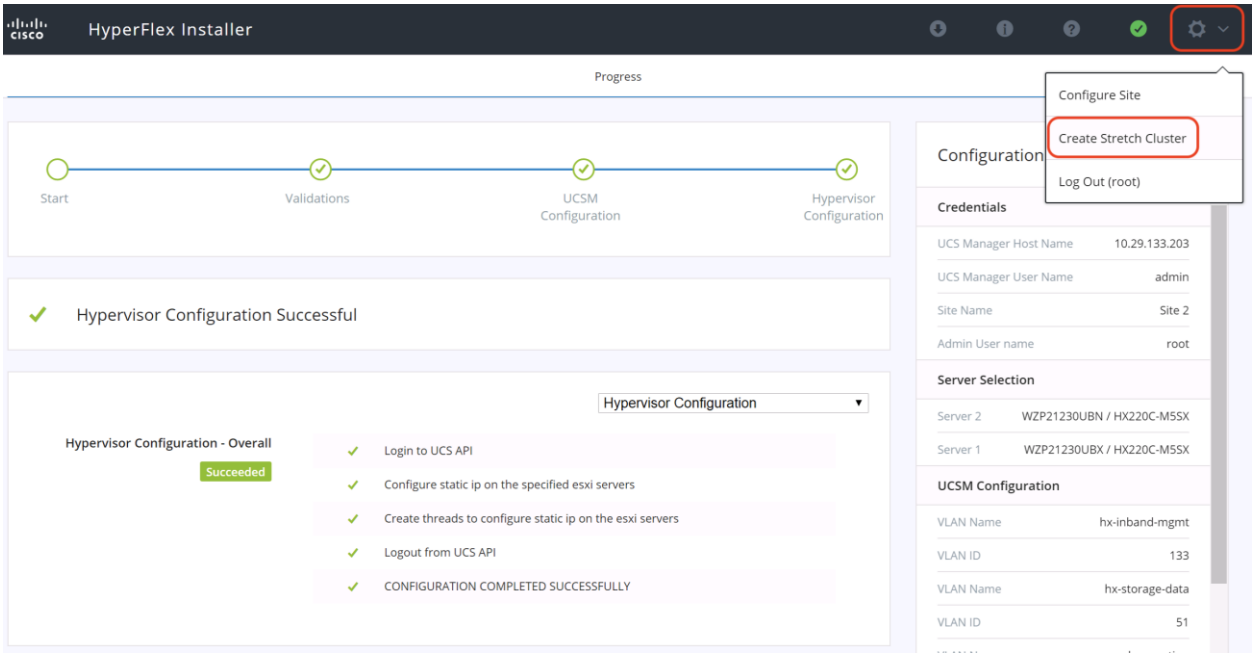


14. After the Hypervisor Configuration step finishes, click the gear shaped icon in the upper-right corner, then click Configure Site to move ahead with configuring the second physical site for this stretched cluster.

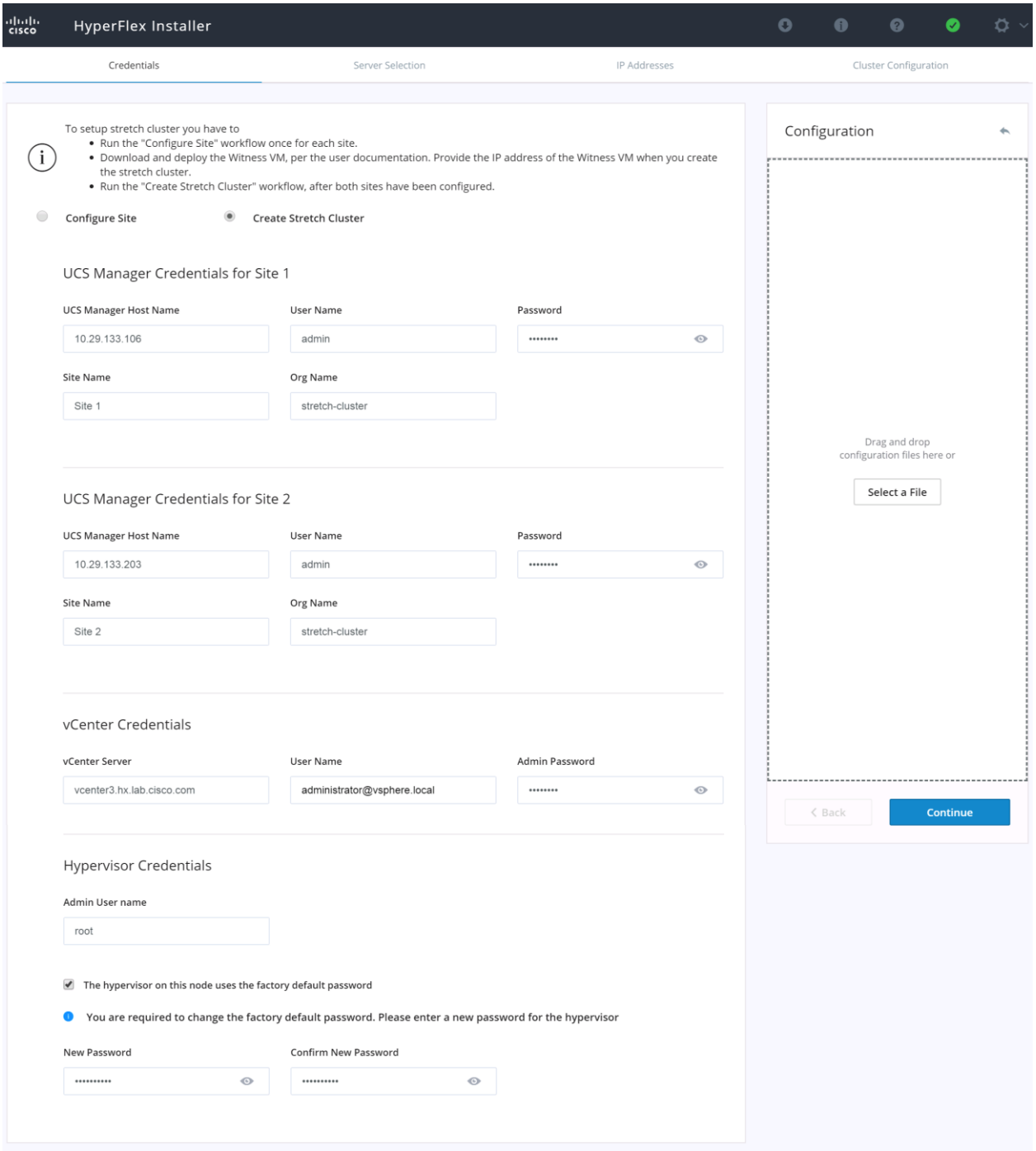


15. Repeat steps 2 through 12 to configure the second site, entering the information and settings that are unique to the second site, such as the second site's UCS Manager address, Site name, MAC address pool prefix, CIMC addresses, Hypervisor IP addresses and hostnames. The VLAN IDs must match.

16. After the Hypervisor Configuration step finishes for the second site, click the gear shaped icon in the upper-right corner, then click Create Stretch Cluster.

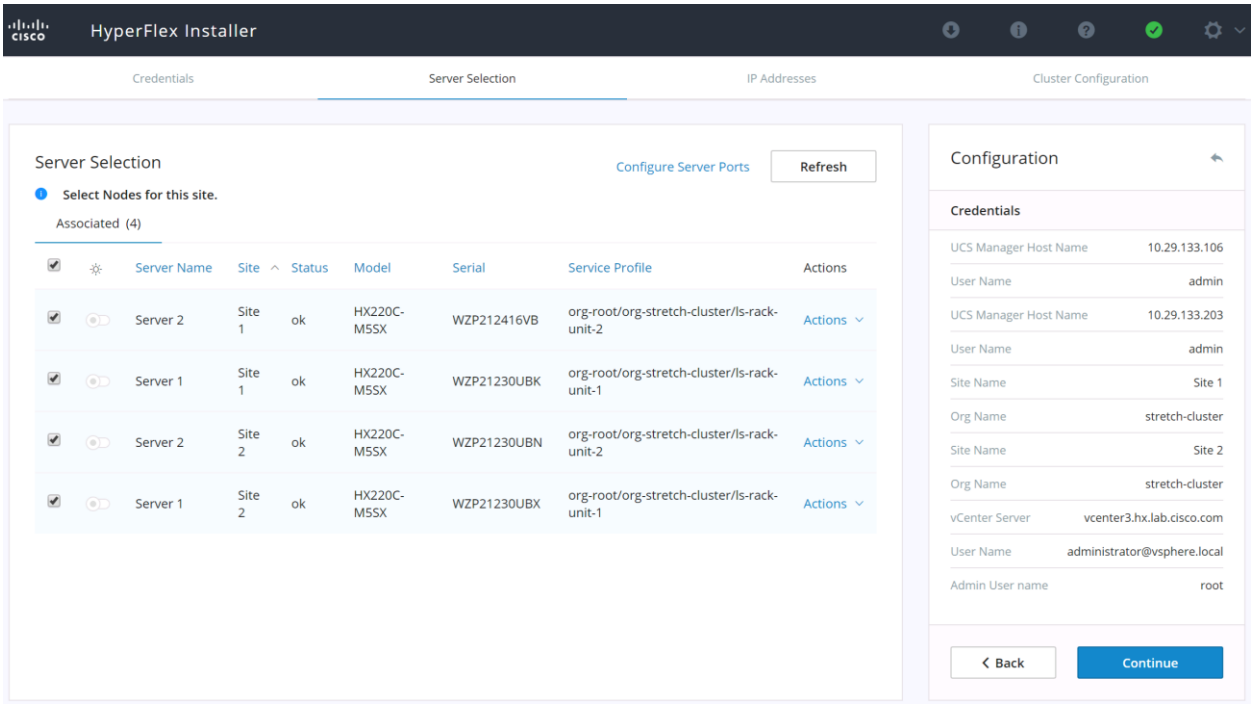


17. Click the radio button for Create Stretch Cluster, in order to complete the installation of the new cluster. Enter the Cisco UCS Manager DNS hostnames or IP addresses for the two previously configured sites, the admin usernames, the passwords, the site names and UCSM org names. Provide the vCenter server DNS hostname or IP address, and an administrative username and password. The default Hypervisor credential which comes installed from the factory is username: root with a password of “Cisco123” and you are required to enter a new password for the Hypervisor. Optionally, you can import a JSON file that has the configuration information, except for the appropriate passwords.



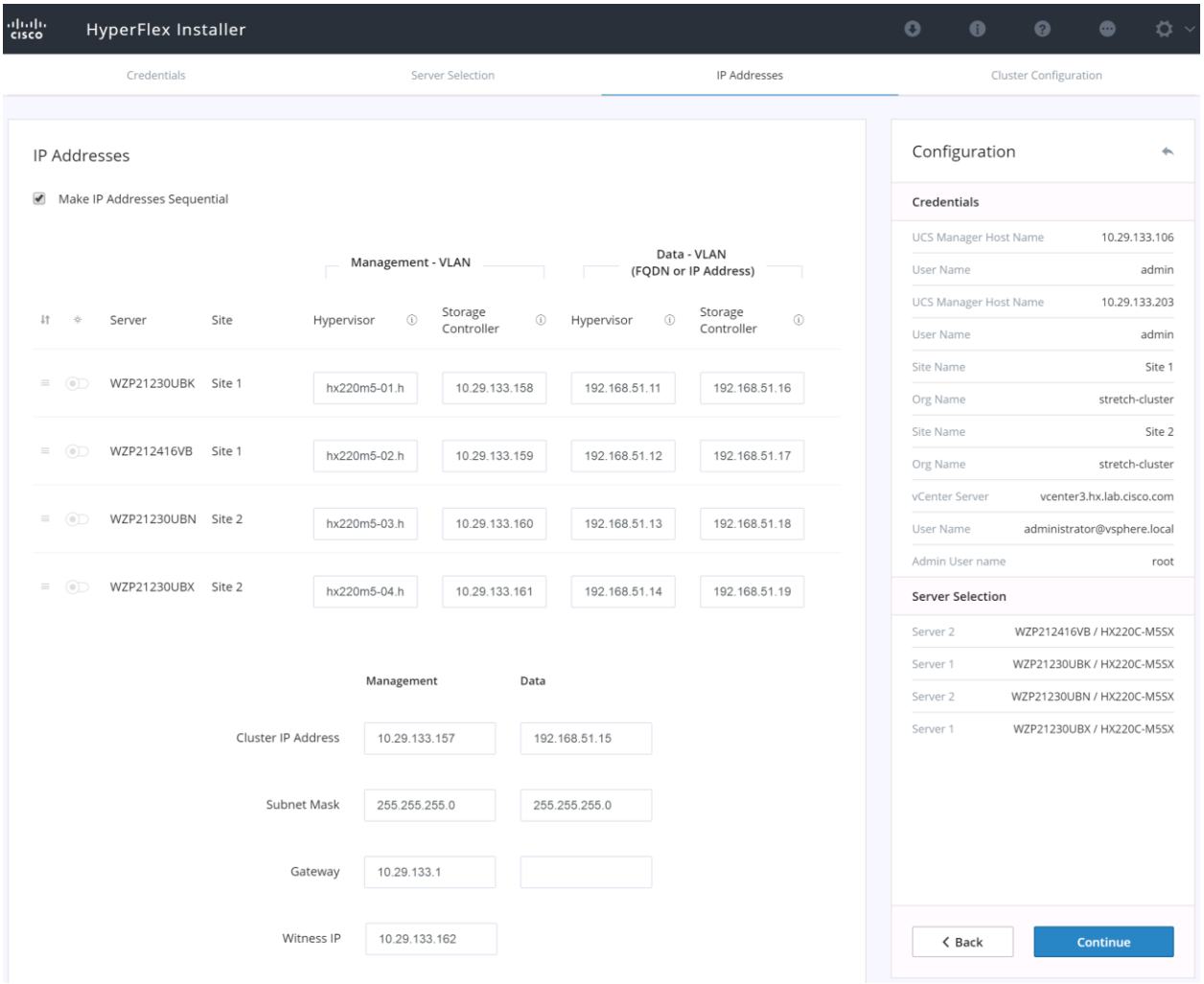
18. Click Continue.

19. Select the Associated HX server models that were previously configured, and are to be used in the new HX cluster for this site and click Continue.



20. Assign the additional IP addresses for the Management and Data networks, the cluster IP addresses, and the Witness VM IP address, then click Continue.

Note: A default gateway is not required for the data network, as those interfaces normally will not communicate with any other hosts or networks, and the subnet can be non-routable.



21. Enter the HX Cluster Name and Replication Factor setting.

22. Enter the Password that will be assigned to the Controller VMs.

23. Enter the Datacenter Name from vCenter, and vCenter Cluster Name. If the Datacenter Name or Cluster Name does not exist, the installer will create them. The Cluster Name must not already exist in the vCenter inventory.

24. Enter the System Services information for DNS, NTP, and Time Zone. You must ensure that at least one of the listed DNS and NTP servers are accessible from each site, and at least one DNS and NTP server will be available in the event of a site going offline.

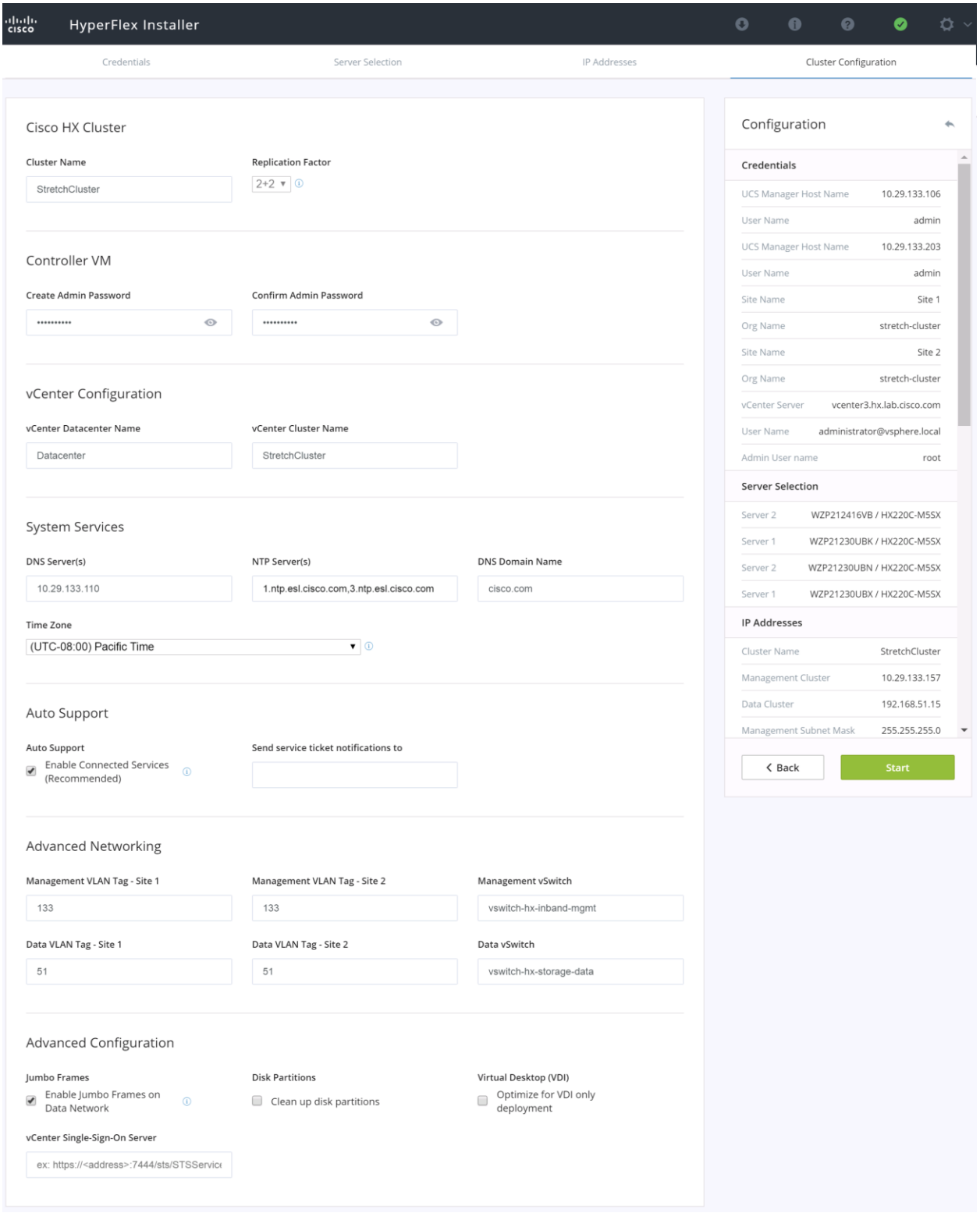
25. Enable Connected Services in order to enable management via Cisco Intersight, and enter the email address to receive service ticket alerts, then scroll down.

26. Under Advanced Networking, enter the matching VLAN IDs for the HyperFlex Management and Storage networks in the two sites.

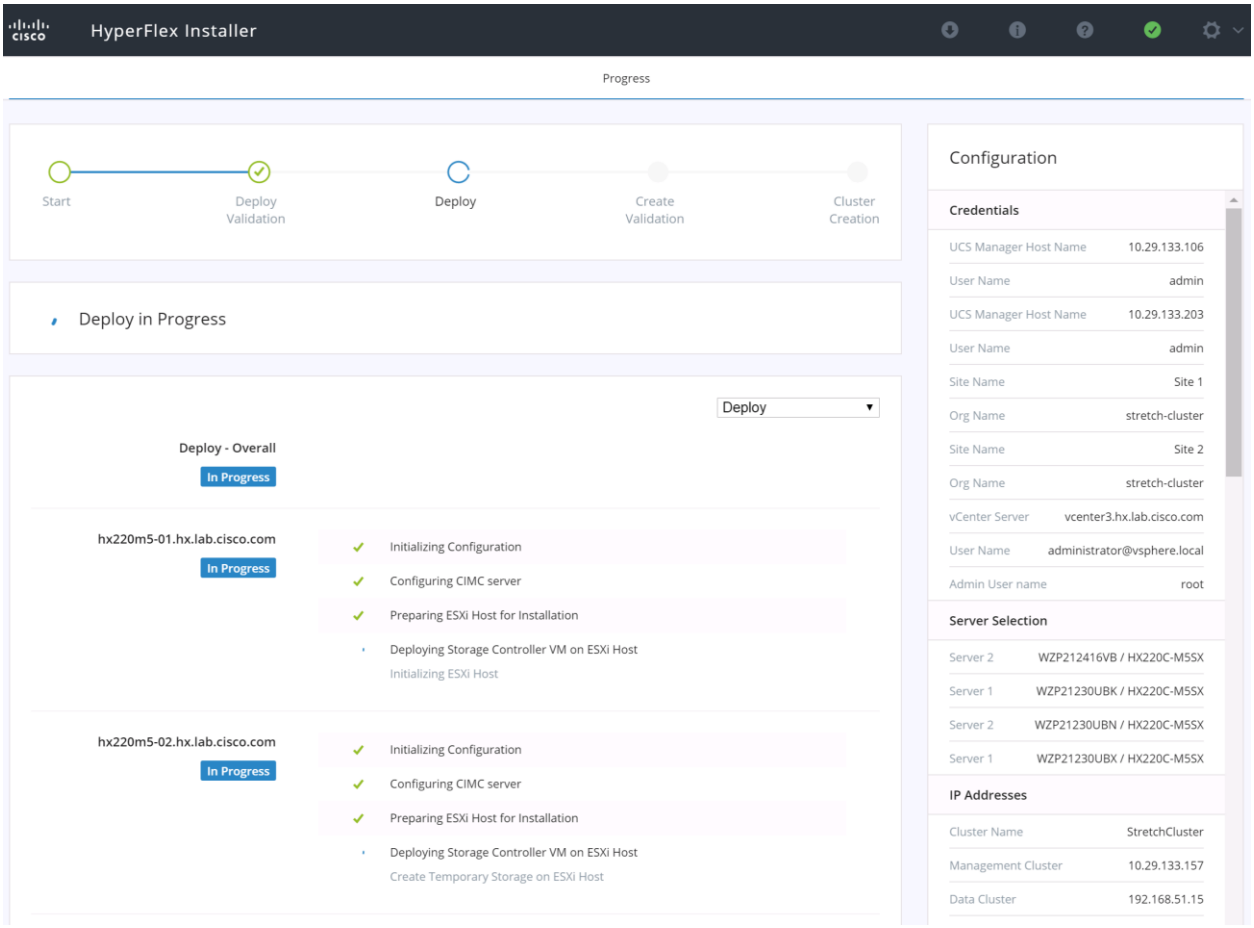
27. Under Advanced Configuration, validate that VDI is not checked (hybrid nodes only). Jumbo Frames should be enabled to ensure the best performance, unless the upstream network is not capable of being

configured to transmit jumbo frames. It is not necessary to select Clean up disk partitions for a new cluster installation, but an installation using previously used converged nodes should have the option checked.

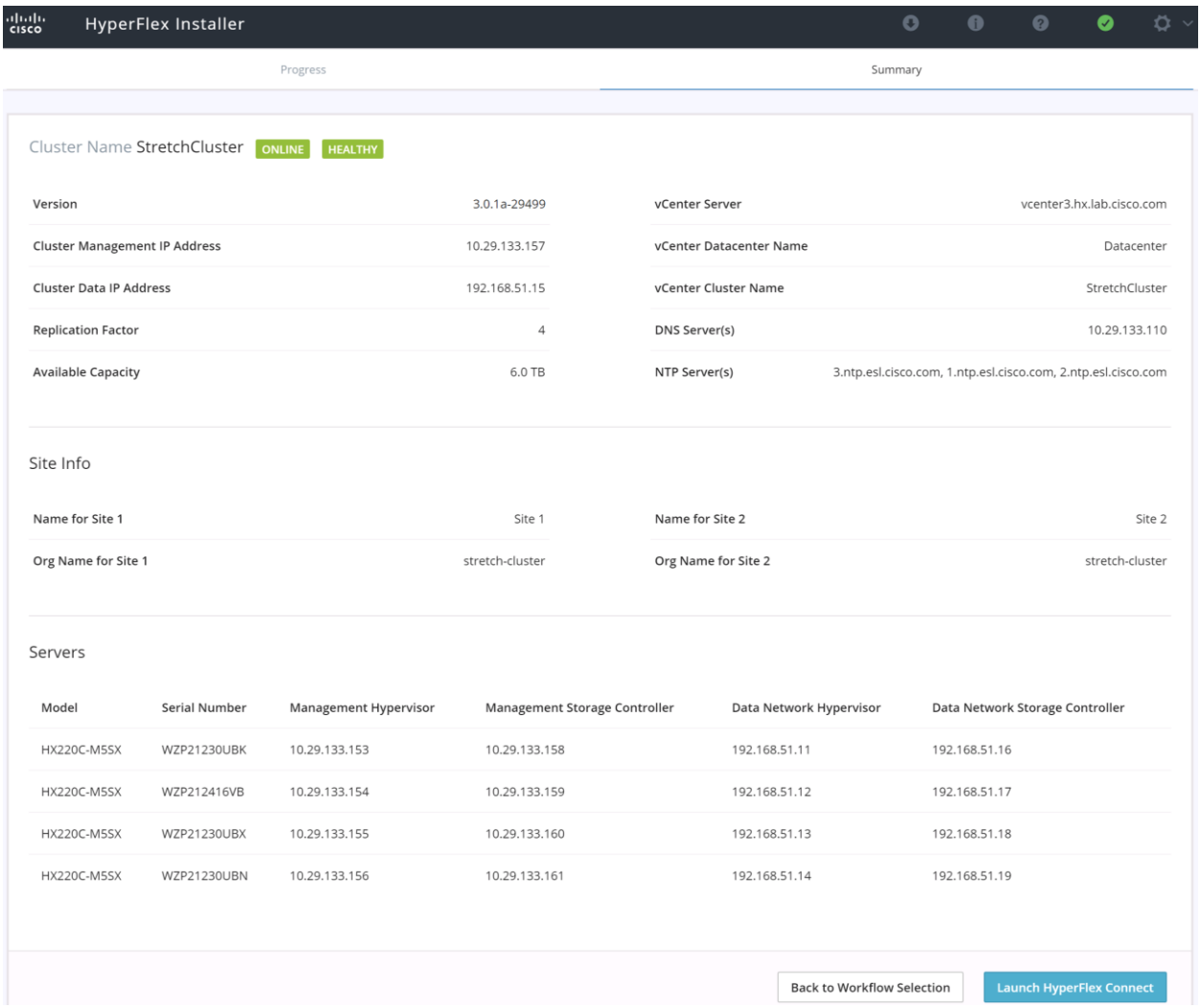
28. Click Start.



29. Validation of the configuration will now start. If there are warnings, you can review them and click “Skip Validation” if the warnings are acceptable. If there are no warnings, the installer will automatically continue on to the configuration process.



30. Review the Summary screen after the install completes.

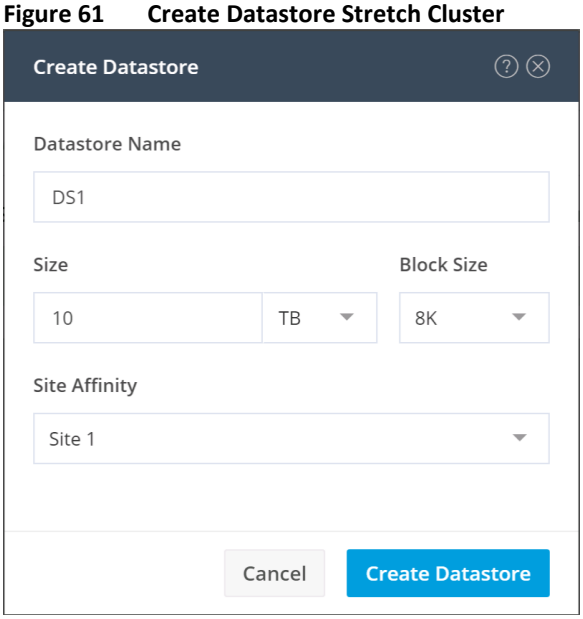


31. Continue with the Post Installation tasks in the next section. It is particularly important to run the post_install script in order to create the vMotion interfaces, the guest VM port groups, and to enable HA and DRS in the cluster. After DRS is enabled, you must also review the DRS site affinity rules which are automatically created as listed below, to ensure the configuration is correct.

Stretched Cluster Datastores

Datastores created in stretched clusters have one additional setting compared to datastores in standard clusters, which is a setting for the datastore's site affinity. The site affinity setting is chosen when creating the datastore in the HyperFlex Connect webpage user interface. Specifying a site association for the datastore ensures that during normal operation, all requests to read data from that datastore will be serviced by the nodes in that specific site, rather than by nodes in the remote site. Maintaining the placement of the VMs using DRS site affinity rules also optimizes performance in a stretched cluster, and these rules should be automatically maintained by the system. In addition to considering the proximity to the users which consume the services provided by the VM, VMs should always be configured to store their virtual disk files in a datastore which is associated with the site where the VM will run. This positive association between the VM's running location, and the site affinity of the datastore which stores the VM's virtual disk files, provides the best possible performance. Changing the configuration to one where there is a negative association, for example a VM running in site #2, but the virtual disk files are stored in a datastore associated with site #1, can lead to reduced storage performance.

Figure 61 Create Datastore Stretch Cluster



vSphere DRS Configuration

VMware vSphere Dynamic Resource Scheduler (DRS) must be configured with site affinity rules in order for the stretched cluster to operate in an optimal manner. VM placement across a stretched cluster uses these site affinity rules, in order to constrain VMs to only run on the nodes in their primary site during normal operation, which is also associated with the site affinity of the datastore which stores the VM's virtual disk files. Site affinity rules and groups are automatically created during the installation, and the rules are created in such a manner that the VMs are allowed to restart and run in the other site in case of a site failure. When VMs are created, they are automatically placed into the VM group associated with the site where they are running. This method helps to balance workloads across all of the nodes in both sites, while retaining the enhanced failover capability of a stretched cluster, in case an entire site was to go offline or otherwise fail.

Figure 62 DRS Host Group

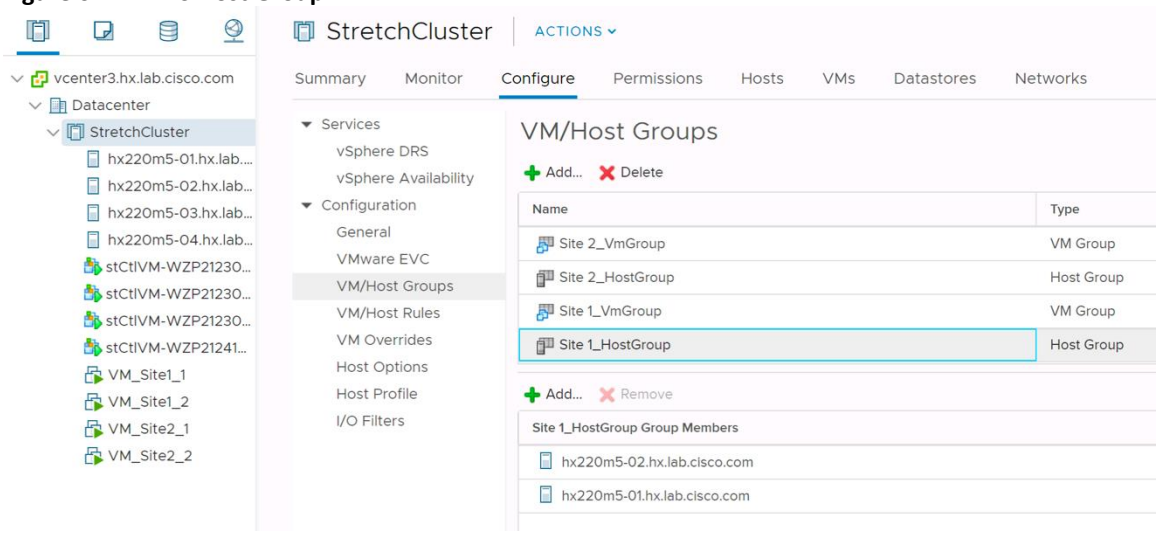
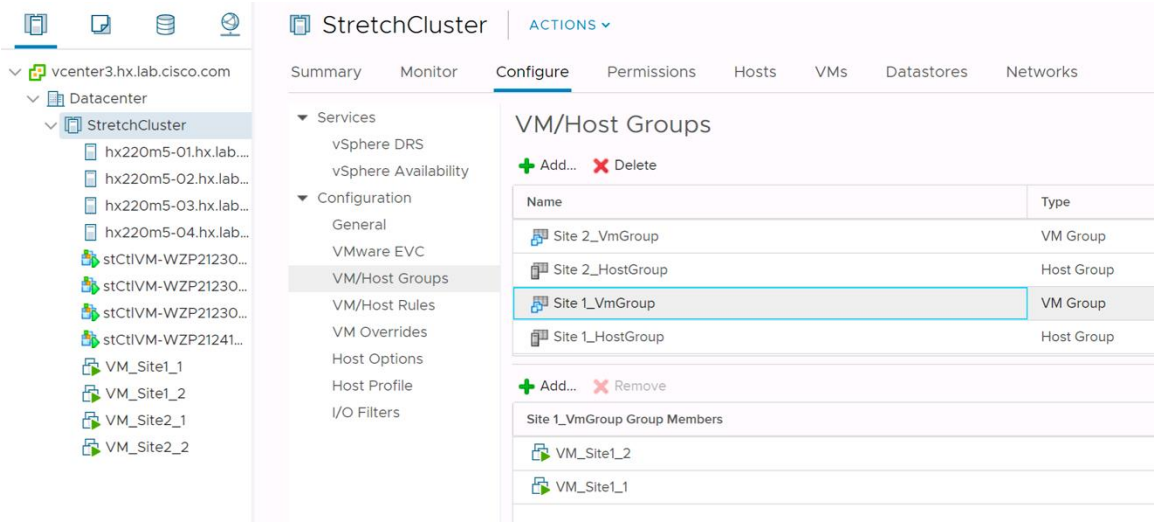


Figure 63 DRS VM Group



Post Installation

After the HyperFlex Installer has completed the installation of the HyperFlex cluster, additional steps may be needed to finalize the configuration of the system before placing virtual machine workloads on the cluster. The following scripts and procedures can be used to complete the installation.

Post Installation Script

To automate many of the post installation procedures and verify the HyperFlex Installer has properly configured Cisco UCS Manager, a script has been provided on the HyperFlex Installer OVA. These steps can also be performed manually in vCenter if preferred. To run this script, complete the following steps:

1. SSH to the installer OVA IP as root with password Cisco123,

   ```
   # ssh root@10.29.133.115
   ```

2. From the CLI of the installer VM, run the script named post_install.

   ```
   # post_install
   ```

3. The installer will already have the information from the just completed HX installation and it will be used by the script. Enter the HX Storage Controller VM root password for the HX cluster (use the one entered during the HX Cluster installation), as well as the vCenter user name and password. You can also enter the vSphere license or complete this task later.

4. Enter “y” to enable HA/DRS if you have the appropriate licenses.

5. Enter “y” to disable the ESXi hosts’ SSH warning.

```
root@Cisco-HX-Installer-Appliance:~# post_install
Script succesfully updated
Logging in to controller 10.29.133.142
HX CVM root password:
Getting ESX hosts from HX cluster...
vCenter URL: 10.29.133.121
Enter vCenter username (user@domain): administrator@vsphere.local
vCenter Password:
Found datacenter Datacenter
Found cluster AFCluster8node

Enter vSphere license key?  (y/n) n

Enable HA/DRS on cluster? (y/n) y

Disable SSH warning? (y/n) y
```

6. Add the vMotion VMkernel interfaces to each node by entering “y”. Input the netmask, the vMotion VLAN ID, and the vMotion IP addresses for each of the hosts as prompted.

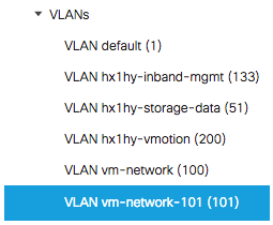
```
Add vmotion interfaces? (y/n) y
 Netmask for vMotion: 255.255.255.0
 VLAN ID: (0-4096) 200
 vMotion IP for hxaf240m5-01.hx.lab.cisco.com: 192.168.200.11
 Adding vmotion-200 to hxaf240m5-01.hx.lab.cisco.com
 Adding vmkernel to hxaf240m5-01.hx.lab.cisco.com
 vMotion IP for hxaf240m5-02.hx.lab.cisco.com: 192.168.200.12
 Adding vmotion-200 to hxaf240m5-02.hx.lab.cisco.com
 Adding vmkernel to hxaf240m5-02.hx.lab.cisco.com
 vMotion IP for hxaf240m5-03.hx.lab.cisco.com: 192.168.200.13
 Adding vmotion-200 to hxaf240m5-03.hx.lab.cisco.com
 Adding vmkernel to hxaf240m5-03.hx.lab.cisco.com
 vMotion IP for hxaf240m5-04.hx.lab.cisco.com: 192.168.200.14
 Adding vmotion-200 to hxaf240m5-04.hx.lab.cisco.com
 Adding vmkernel to hxaf240m5-04.hx.lab.cisco.com
 vMotion IP for hxaf240m5-05.hx.lab.cisco.com: 192.168.200.15
 Adding vmotion-200 to hxaf240m5-05.hx.lab.cisco.com
 Adding vmkernel to hxaf240m5-05.hx.lab.cisco.com
 vMotion IP for hxaf240m5-06.hx.lab.cisco.com: 192.168.200.16
 Adding vmotion-200 to hxaf240m5-06.hx.lab.cisco.com
 Adding vmkernel to hxaf240m5-06.hx.lab.cisco.com
 vMotion IP for hxaf240m5-07.hx.lab.cisco.com: 192.168.200.17
 Adding vmotion-200 to hxaf240m5-07.hx.lab.cisco.com
 Adding vmkernel to hxaf240m5-07.hx.lab.cisco.com
 vMotion IP for hxaf240m5-08.hx.lab.cisco.com: 192.168.200.18
 Adding vmotion-200 to hxaf240m5-08.hx.lab.cisco.com
 Adding vmkernel to hxaf240m5-08.hx.lab.cisco.com
```

7. For stretched clusters, the guest VM port groups must be created. Enter “y” to create the port groups on the ESXi hosts. This option will also create the corresponding VLANs in Cisco UCS Manager, and assign the VLANs to the vm-network vNIC-Template.

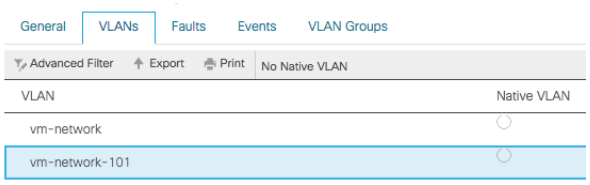
8. For standard clusters, the HyperFlex installer will have already created at least one vm-network port group and assigned the default VM network VLAN input from the cluster installation. Enter “n” to skip this step and use the group(s) that were already created. If desired, additional VM network port groups can be created and the additional VLANs will be added to the vm-networks vSwitch. This option will also create the corresponding VLANs in Cisco UCS Manager, and assign the VLANs to the vm-network vNIC-Template. This script can be rerun at later time as well to create additional VM networks and Cisco UCS VLANs by issuing the command “post_install --vlan”.

```
Add VM network VLANs? (y/n) y
 Attempting to find UCSM IP
 Found UCSM 10.29.133.106, logging with username admin.  Org is AFCluster8node
 UCSM Password:
 Port Group Name to add (VLAN ID will be appended to the name): vm-network
 VLAN ID: (0-4096) 101
 Adding VLAN 101 to FI
 Adding VLAN 101 to vm-network-a VNIC template
 Adding vm-network-101 to hxaf240m5-01.hx.lab.cisco.com
 Adding vm-network-101 to hxaf240m5-02.hx.lab.cisco.com
 Adding vm-network-101 to hxaf240m5-03.hx.lab.cisco.com
 Adding vm-network-101 to hxaf240m5-04.hx.lab.cisco.com
 Adding vm-network-101 to hxaf240m5-05.hx.lab.cisco.com
 Adding vm-network-101 to hxaf240m5-06.hx.lab.cisco.com
 Adding vm-network-101 to hxaf240m5-07.hx.lab.cisco.com
 Adding vm-network-101 to hxaf240m5-08.hx.lab.cisco.com
Add additional VM network VLANs? (y/n) n
```

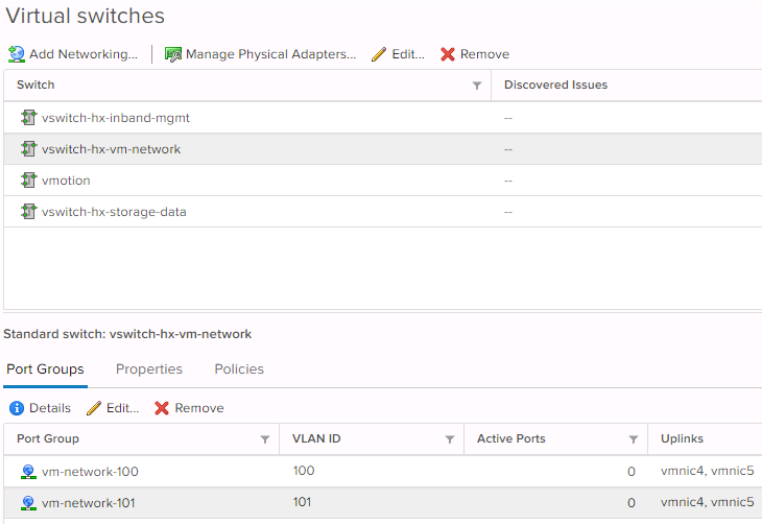
VLANs are created in Cisco UCS:



VLANs are assigned to vNICs:



Port groups are created:



9. Enter “y” to run the health check. Enter the ESX hosts’ root password.

10. The script will complete and provide a summary screen. Validate there are no errors and the cluster is healthy.

```
Host: hxaf240m5-01.hx.lab.cisco.com
    No errors found

Host: hxaf240m5-02.hx.lab.cisco.com
    No errors found

Host: hxaf240m5-03.hx.lab.cisco.com
    No errors found

Host: hxaf240m5-04.hx.lab.cisco.com
    No errors found

Host: hxaf240m5-05.hx.lab.cisco.com
    No errors found

Host: hxaf240m5-06.hx.lab.cisco.com
    No errors found

Host: hxaf240m5-07.hx.lab.cisco.com
    No errors found

Host: hxaf240m5-08.hx.lab.cisco.com
    No errors found
```

Syslog

M4 generation servers will generate an ESXi fault about the host logs not being stored in a permanent location. To clear this fault, and also to store diagnostic logs in a central location in case they are needed after a host failure, it is recommended to enable a syslog destination for permanent storage of the ESXi host logs for all Cisco HyperFlex hosts. It is possible to use the vCenter server as the log destination in this case, or another syslog receiver of your choice.

To configure syslog, complete the following steps:

1. Log on to the ESXi host via SSH as the root user.

2. Enter the following commands, replacing the IP address in the first command with the IP address of the vCenter server that will receive the syslog logs:

```
[root@hx220-01:~] esxcli system syslog config set --loghost='udp://10.29.133.120'
[root@hx220-01:~] esxcli system syslog reload
[root@hx220-01:~] esxcli network firewall ruleset set -r syslog -e true
[root@hx220-01:~] esxcli network firewall refresh
```

3. Repeat for each ESXi host.

Datastores

Create a datastore for storing the virtual machines. This task can be completed by using the vSphere Web Client HX plugin, or by using the HyperFlex Connect HTML management webpage. To configure a new datastore via the HyperFlex Connect webpage, complete the following steps:

1. Use a web browser to open the HX cluster IP management URL.

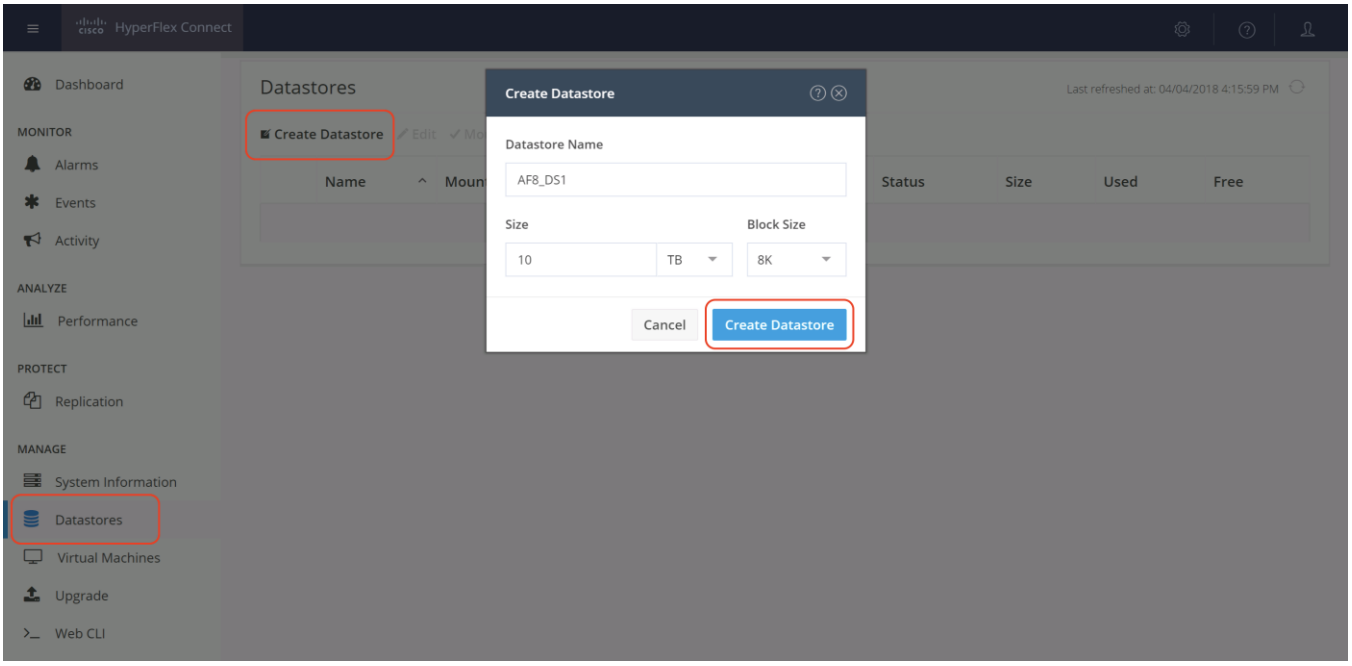
2. Enter a local credential, or a vCenter RBAC credential with administrative rights for the username, and the corresponding password.

3. Click Login.

4. Click Datastores in the left pane and click Create Datastore.

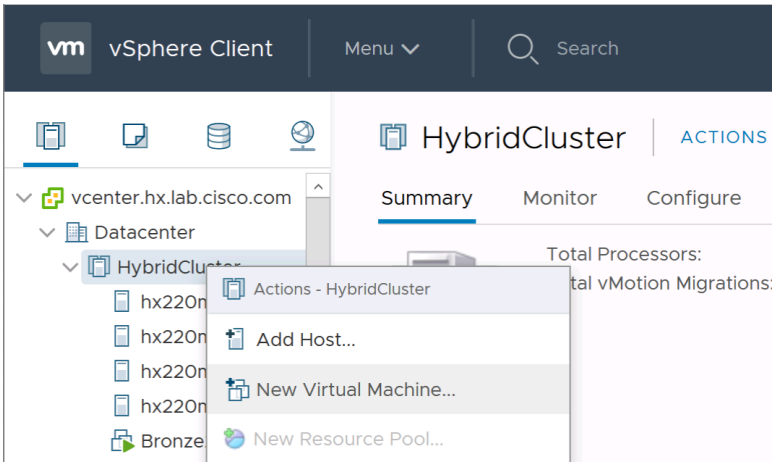
5. In the popup, enter the Datastore Name and size. For most applications, leave the Block Size at the default of 8K. Only dedicated Virtual Desktop Infrastructure (VDI) environments should choose the 4K Block Size option.

6. Click Create Datastore.



Initial Tasks and Testing

In order to perform initial testing and learn about the features in the HyperFlex cluster, create a test virtual machine stored on your new HX datastore in order to take a snapshot and perform a cloning operation.

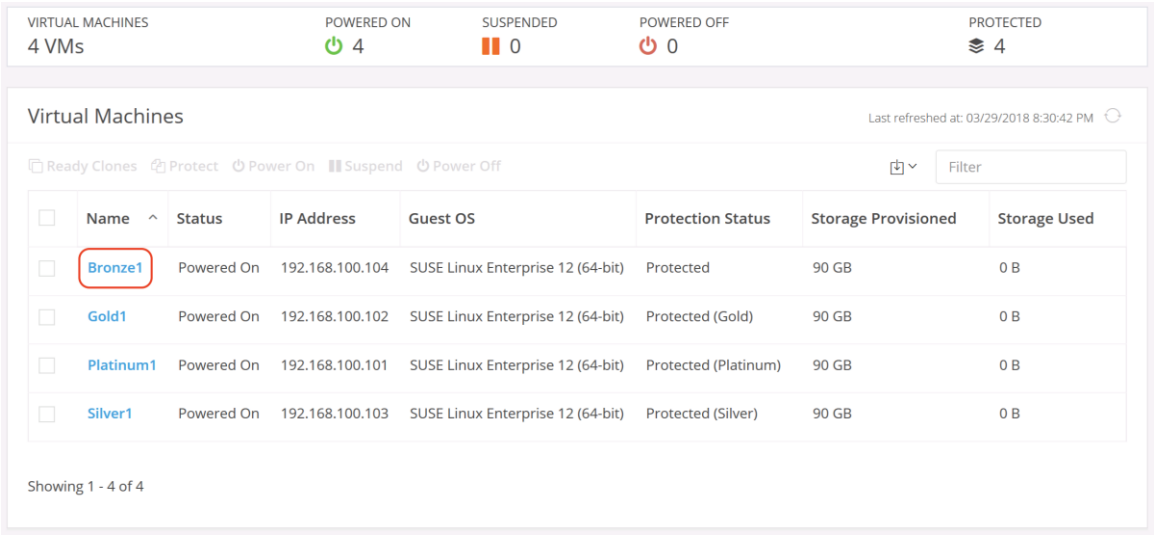


Snapshots

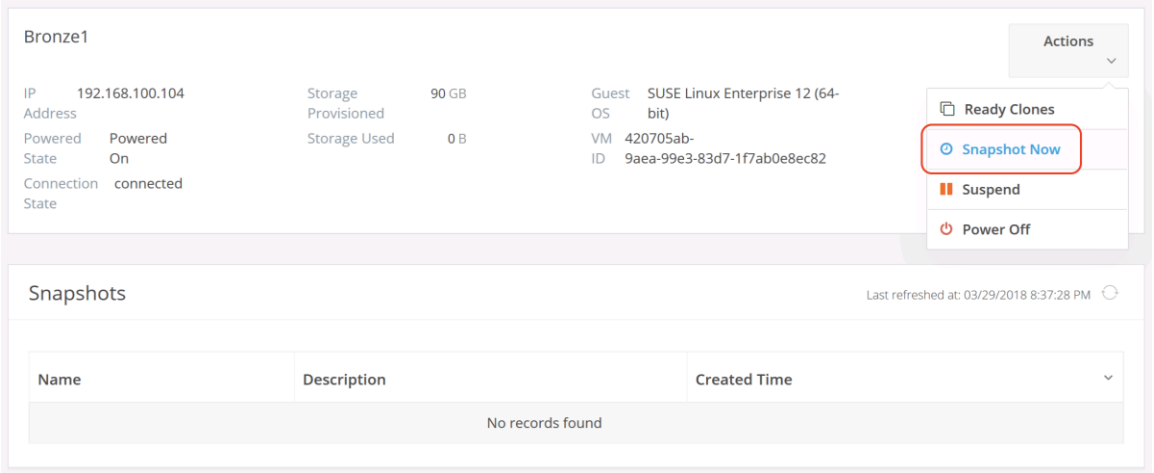
Take a snapshot of the new virtual machine prior to powering it on.

To take an instant snapshot of a VM, complete the following steps:

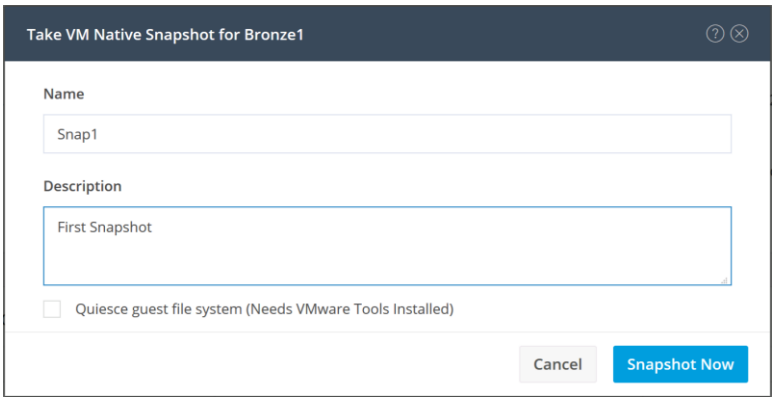
1. In the HyperFlex Connect webpage, click on the Virtual Machines menu, then click on the name of the VM to snapshot.



2. Click the Actions drop-down list, then select Snapshot Now.



3. Input the snapshot name, a description if desired, and choose whether to quiesce the VM, then click Snapshot Now.

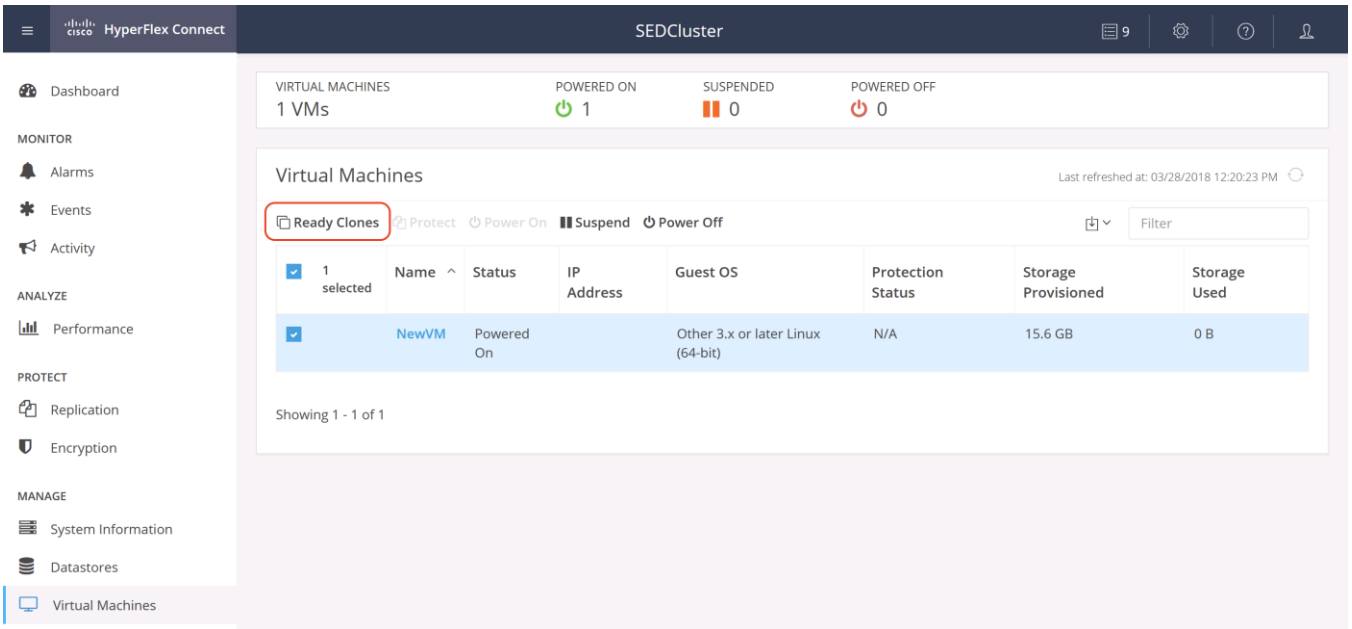


Ready Clones

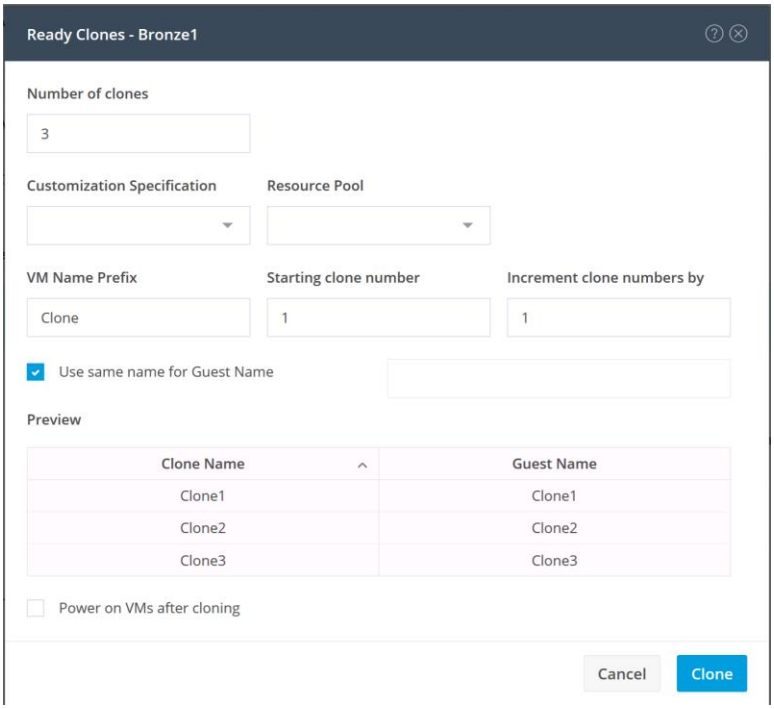
Create a few clones of our test virtual machine.

To create the Ready Clones, complete the following steps:

1. In the HyperFlex Connect webpage, click on the Virtual Machines menu, click the checkbox to select the VM to clone, then click Ready Clones.



2. Input the Number of clones to create, a customization specification if needed, a resource pool if needed, and a naming prefix, then click Clone to start the operation. The clones will be created in seconds.



Auto-Support and Notifications

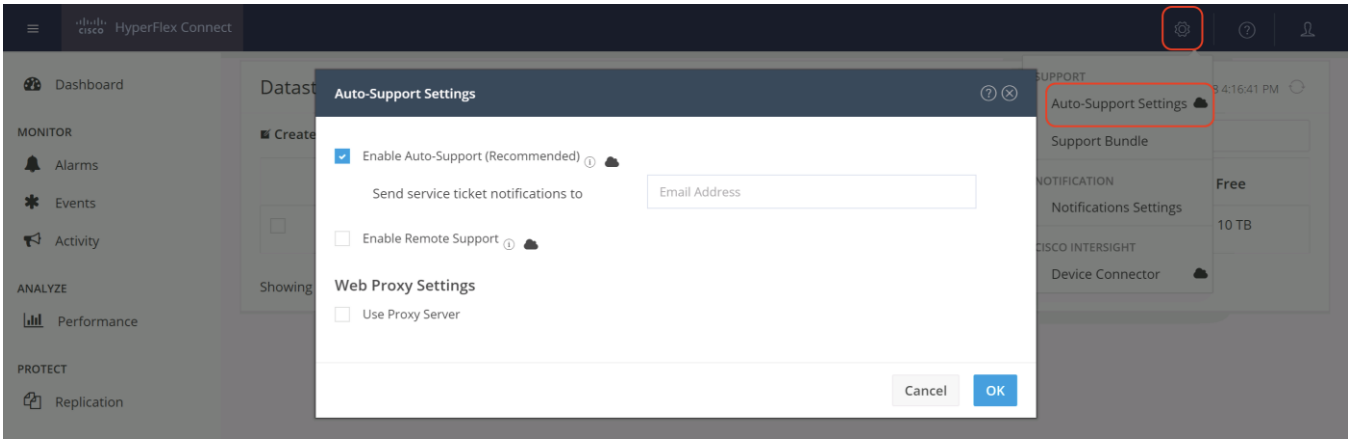
Auto-Support should be enabled for all clusters during the initial HyperFlex installation. Auto-Support enables Call Home to automatically send support information to Cisco TAC, and notifications of tickets to the email address specified. If the settings need to be modified, they can be changed in the HyperFlex Connect HTML management webpage.

A list of events that will automatically open a support ticket with Cisco TAC is as follows:

- Cluster Capacity Changed
- Cluster Unhealthy
- Cluster Health Critical
- Cluster Read Only
- Cluster Shutdown
- Space Warning
- Space Alert
- Space Critical
- Disk Blacklisted
- Infrastructure Component Critical
- Storage Timeout

To change Auto-Support settings, complete the following steps:

1. From the HyperFlex Connect webpage, click the gear shaped icon in the upper right-hand corner, and click Auto-Support Settings.
2. Enable or disable Auto-Support as needed.
3. Enter the email address to receive alerts when Auto-Support events are generated.
4. Enable or disable Remote Support as needed. Remote support allows Cisco TAC to connect to the HX cluster and accelerate troubleshooting efforts.
5. Enter in the information for a web proxy if needed.
6. Click OK.



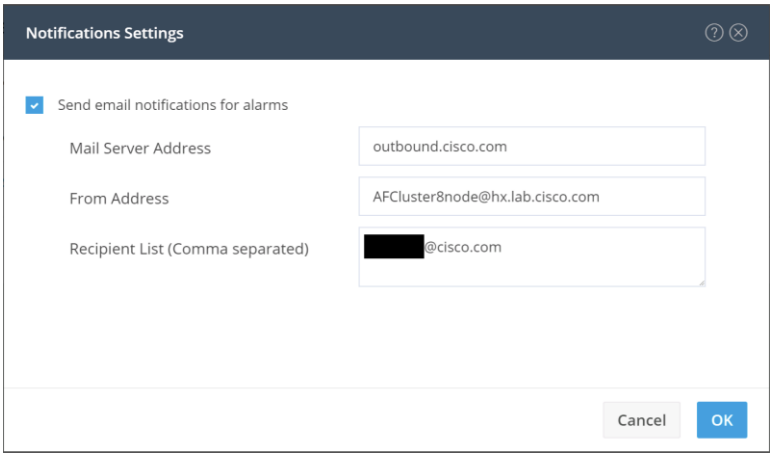
Email notifications that come directly from the HyperFlex cluster can also be enabled.

To enable direct email notifications, complete the following steps:

1. From the HyperFlex Connect webpage, click the gear shaped icon in the upper right-hand corner, and click Notifications Settings.

2. Enter the DNS name or IP address of the outgoing email server or relay, the email address the notifications will come from, and the recipients.

3. Click OK.



Smart Licensing

HyperFlex 2.5 and later utilizes Cisco Smart Licensing, which communicates with a Cisco Smart Account to validate and check out HyperFlex licenses to the nodes, from the pool of available licenses in the account. At the beginning, Smart Licensing is enabled but the HX storage cluster is unregistered and in a 90-day evaluation period or EVAL MODE. For the HX storage cluster to start reporting license consumption, it must be registered with the Cisco Smart Software Manager (SSM) through a valid Cisco Smart Account. Before beginning, verify that you have a Cisco Smart account, and valid HyperFlex licenses are available to be checked out by your HX cluster.

To create a Smart Account, see **Cisco Software Central** > **Request a Smart Account**

https://webapps.cisco.com/software/company/smartaccounts/home?route=module/accountcreation .

To activate and configure smart licensing, complete the following steps:

1. Log into a controller VM. Confirm that your HX storage cluster is in Smart Licensing mode.

```
# stcli license show status

Smart Licensing is ENABLED

Registration:
  Status: UNREGISTERED
  Export-Controlled Functionality: Not Allowed

License Authorization:
  Status: EVAL MODE
  Evaluation Period Remaining: 88 days, 1 hr, 33 min, 41 sec
  Last Communication Attempt: NONE

License Conversion:
 Automatic Conversion Enabled: true
 Status: NOT STARTED

Utility:
  Status: DISABLED

Transport:
  Type: TransportCallHome
```

   Feedback will show Smart Licensing is ENABLED, Status: UNREGISTERED, and the amount of time left during the 90-day evaluation period (in days, hours, minutes, and seconds).

2. Navigate to Cisco Software Central (https://software.cisco.com/) and log in to your Smart Account.

3. From Cisco Smart Software Manager, generate a registration token.

4. In the License pane, click Smart Software Licensing to open Cisco Smart Software Manager.

5. Click Inventory.

6. From the virtual account where you want to register your HX storage cluster, click General, and then click New Token.

7. In the Create Registration Token dialog box, add a short Description for the token, enter the number of days you want the token to be active and available to use on other products, and check Allow export-controlled functionality on the products registered with this token.

8. Click Create Token.

9. From the New ID Token row, click the Actions drop-down list, and click Copy.

10. Log into a controller VM.

11. Register your HX storage cluster, where *idtoken-string* is the New ID Token from Cisco Smart Software Manager.

```
# stcli license register --idtoken idtoken-string
```

12. Confirm that your HX storage cluster is registered.

```
# stcli license show summary
```

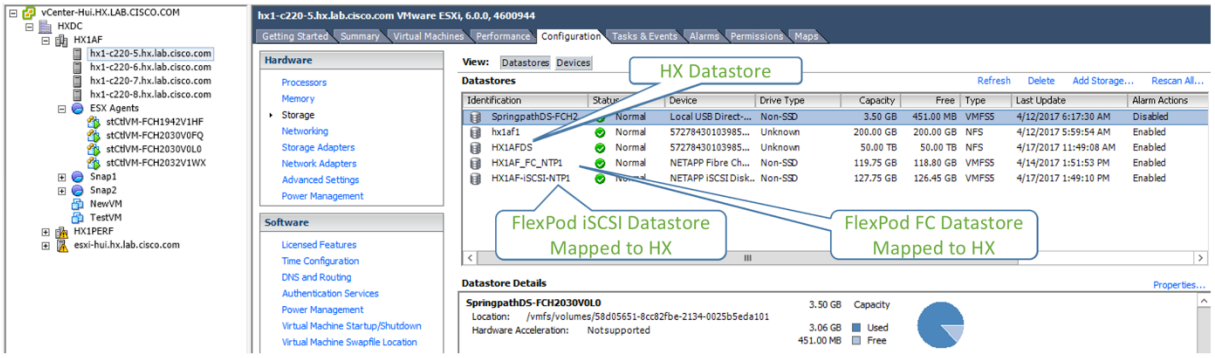
The cluster is now ready. You may run any other preproduction tests that you wish to run at this point.

Additional vHBAs or vNICs

Overview

From HXDP version 1.8 onward, customers have the flexibility to leverage third-party storage infrastructure by connecting external storage arrays to HX systems. As an example, one can map and connect Fibre Channel LUNs from an IBM VersaStack or NFS volumes from a NetApp FlexPod system, and then easily perform a Storage vMotion of virtual machines into the HyperFlex system.

Figure 64 External Storage in HX



In order to connect to other storage systems such as FlexPod via iSCSI or NFS, or an FC SAN, it is recommended that the additional vHBAs or vNICs be added during the creation of the HX cluster. If these are added post cluster creation, the PCI enumeration can change causing PCI passthrough device configuration errors. With HXDP 2.5 and onward, the system can repair these changes automatically via an additional reboot of the ESXi hosts. It is recommended that you do not make such hardware changes after the HX cluster is created. A better option is to add vHBAs or vNICs as necessary while the cluster is created. Both of these processes are documented below.

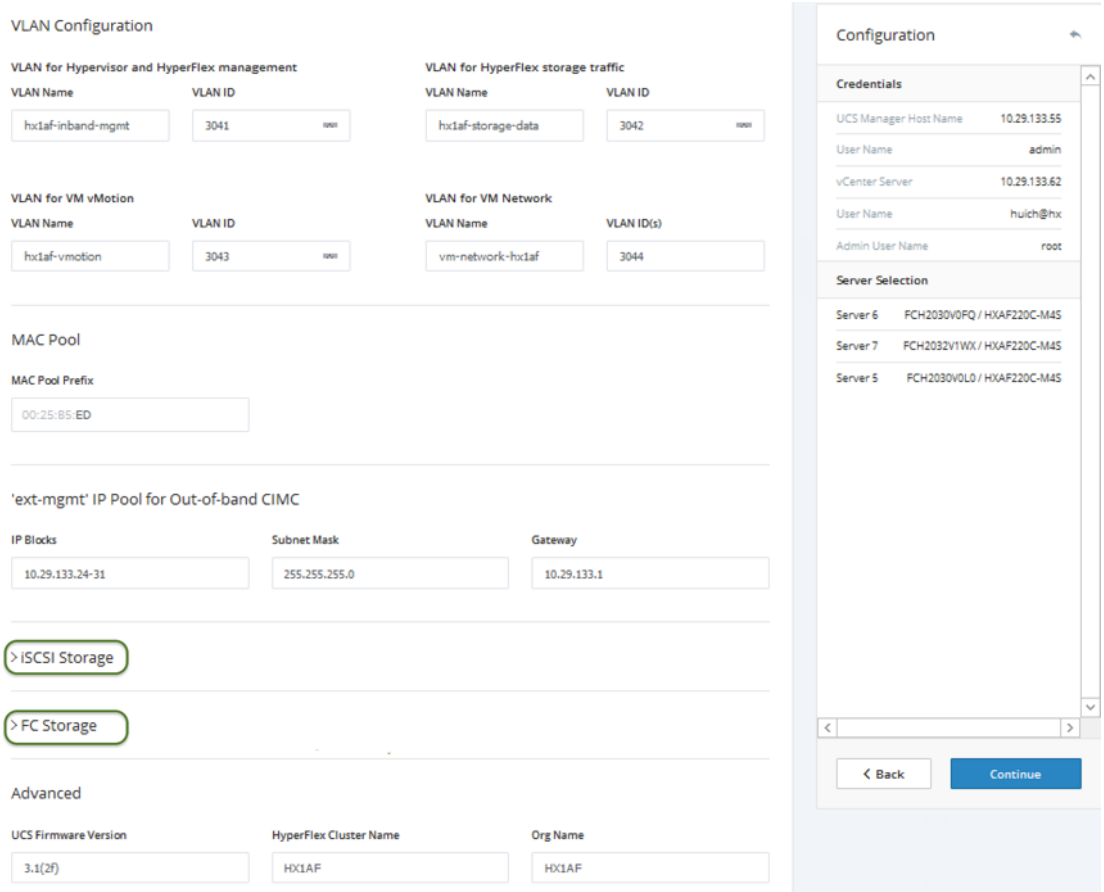
In this section, only the addition of FC vHBAs or iSCSI vNICs to HX hosts is documented. A more detailed procedure about connecting other iSCSI or NFS storage to HX cluster is in the Appendix.

Note: Although in this CVD we use iSCSI as example to connect HX to external IP storage devices, the vNICs added by this procedure could be used for connecting to NFS storage devices, or other IP connectivity.

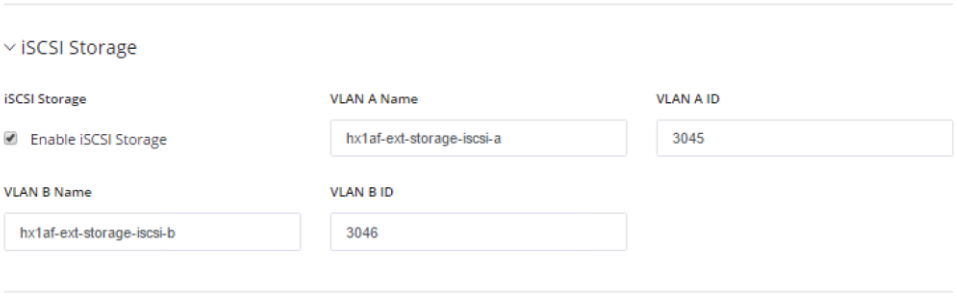
Adding vHBAs or iSCSI vNICs During HX Cluster Creation

From HXDP 2.0 onward, the HX installer supports adding supplemental vHBAs or vNICs as a part of the cluster creation. An overview of this procedure is as follows:

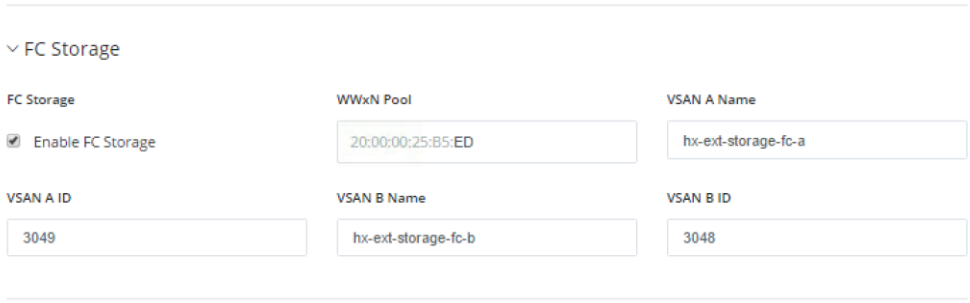
1. Open the HyperFlex Installer from a web browser, login as root user.
2. On the HyperFlex Installer webpage select a Workflow of Cluster Creation to start a fresh cluster installation.
3. Continue with appropriate inputs until you get to the page for Cisco UCS Manager configuration.



4. Click the > carat to expand iSCSI Storage configuration. Check the box Enable iSCSI Storage if you want to create additional vNICs to connect to the external iSCSI storage systems. Enter a VLAN name and ID for Fabric A and B dual connections.



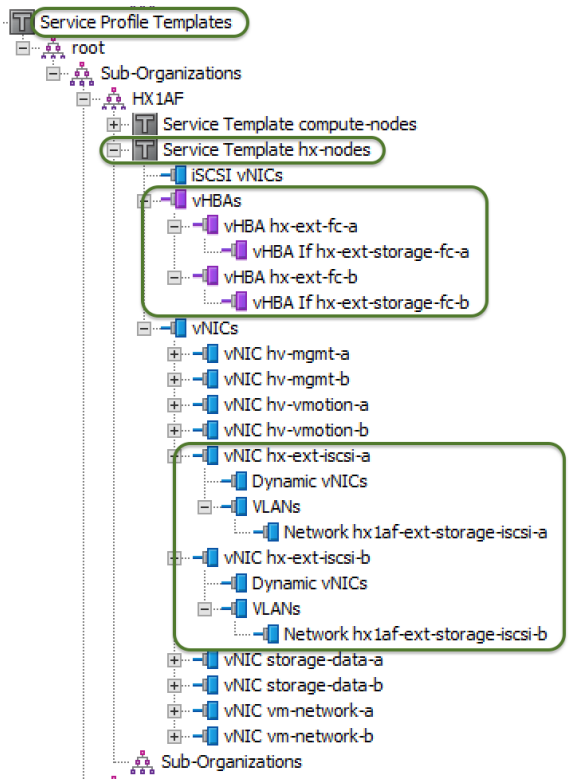
5. Click the > carat to expand FC Storage configuration. Check the box Enable FC Storage if you want to create Fibre Channel vHBAs to connect to the external FC or FCoE storage systems. Enter WWxN Pool prefix (For example: 20:00:00:25:B5:ED, only enter the last byte value), VSAN names and IDs for Fabric A and B dual connections.



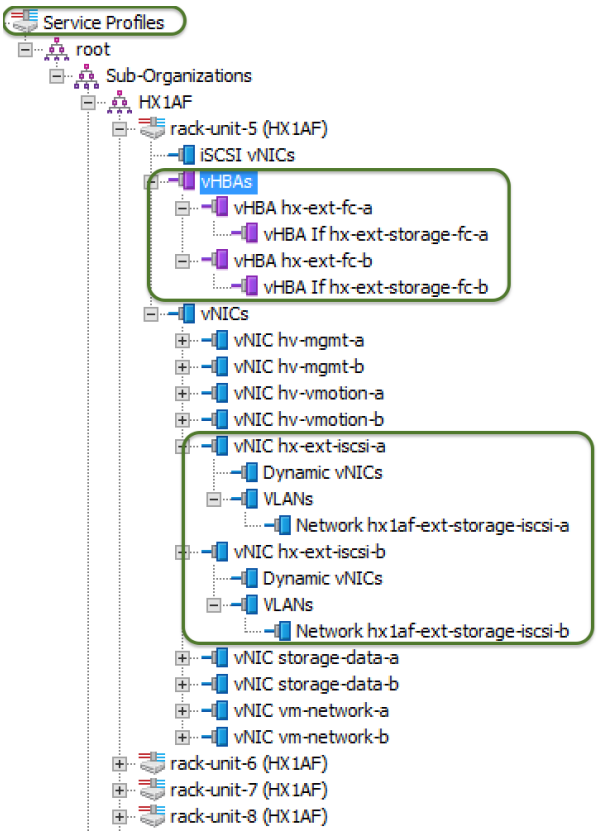
6. Continue and complete the inputs for all the remaining cluster configuration tasks, start the cluster creation and wait for the completion.

Note: You can choose to enable either only iSCSI, only FC, or both according to your needs.

7. After the install is completed, the additional dual vHBAs and/or dual vNICs are created for the Service Profile Templates named “hx-nodes” and “compute-nodes”.

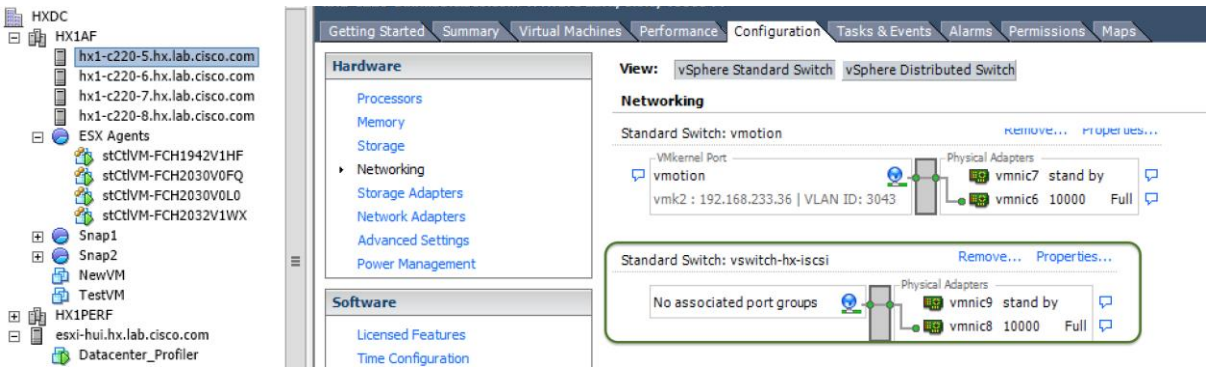


8. For each HX node, dual vHBAs and/or dual iSCSI vNICs are created as well.



Note: In Cisco UCS Manager, the additional vNICs are configured as standard vNICs, not as iSCSI vNICs, as iSCSI vNICs are specifically used for iSCSI boot adapters.

9. In vCenter, a standard vSwitch vswitch-hx-iscsi is created on each HX ESXi host. Further configuration to create iSCSI VMkernel ports needs to be done manually for storage connections (see Appendix D).



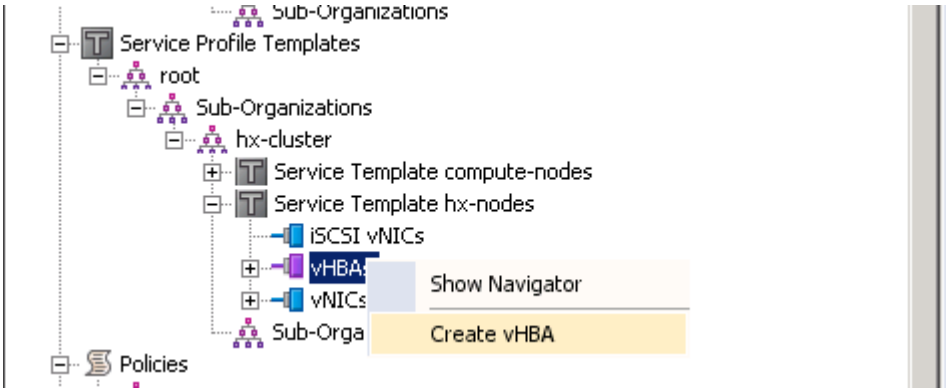
Adding vHBAs or vNICs to an Existing HX Cluster

Should you decide to add additional storage such as a FlexPod after you have already installed your cluster, the following procedure can be used for adding vHBAs or vNICs that could cause PCI re-enumeration upon an ESXi host reboot. Beginning with HXDP 2.5, the DirectPath I/O configuration will repair itself automatically via an additional reboot of the node. Therefore, it is recommended you do not reboot multiple nodes at once after making these hardware changes, as it could lead to a cluster failure. Validate the health state of each host, and the HX cluster before rebooting or performing the procedure on subsequent nodes. In this

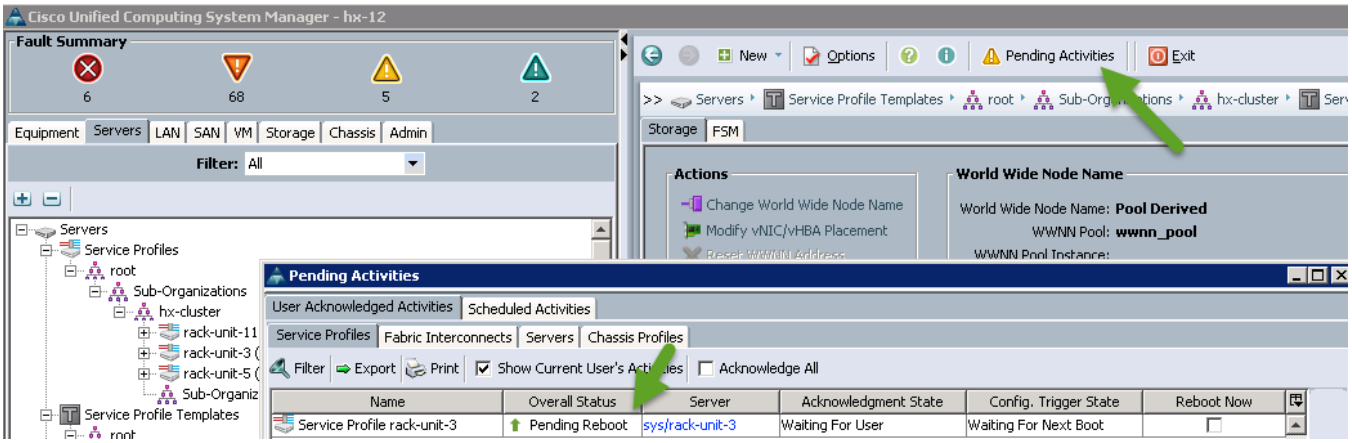
example, we will be adding vHBAs after an HX cluster is created via the Cisco UCS service profile template. We will reboot one ESXi node at a time in a rolling upgrade fashion so there will be no outage.

To add vHBAs or vNICs, complete the following steps:

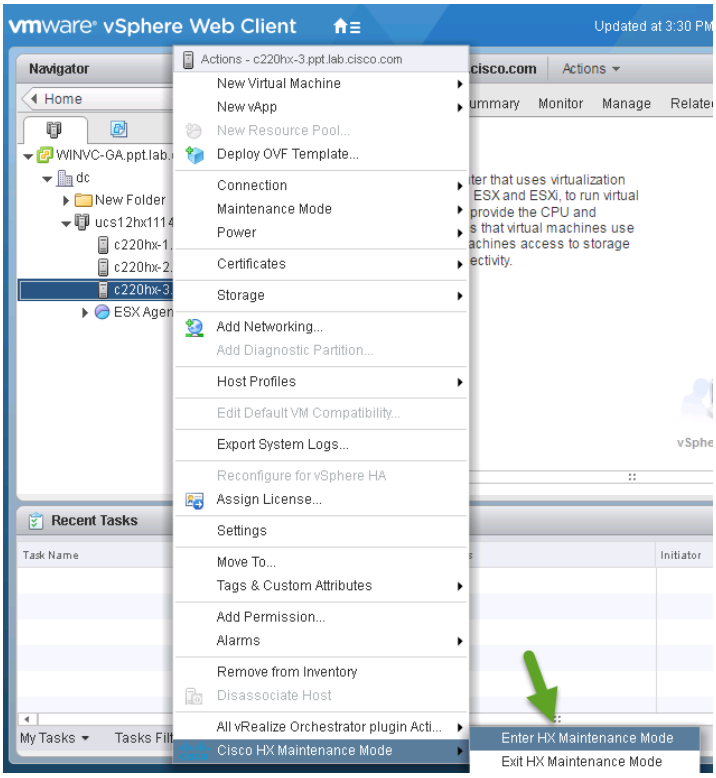
1. Example of hardware change: Add vHBAs to the Service Profile Templates for HX (refer to Cisco UCS documentation for your storage device such as a FlexPod CVD for configuring the vHBAs).



2. After adding the vHBAs to the templates, the servers will be in a Pending Reboot state and require a reboot to add the new interface. **Do NOT reboot the HX servers at this time.**



3. Using HyperFlex Connect, or the vSphere Web Client, place one of the HX ESXi hosts in HX-Maintenance Mode.

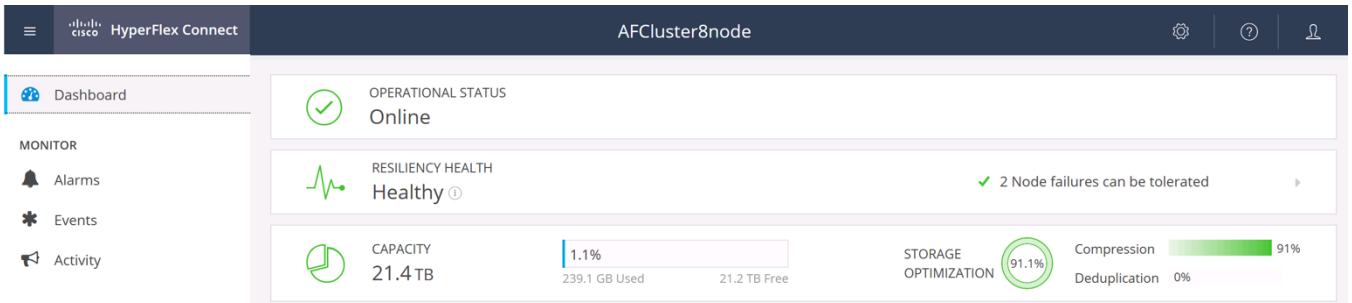


4. After the host has entered Maintenance Mode, reboot the associated node to complete the addition of the new hardware.

5. After the node has rebooted, the HXDP software will detect that the DirectPath I/O configuration has changed, and must be reconfigured. This will result in one additional automatic reboot of the node.

6. After the second reboot, exit the ESXi host from maintenance mode, the SCVM should start automatically without errors.

7. Check the health status of the cluster, validating that the cluster is healthy before proceeding to reboot the next node. The cluster health status can be viewed from HyperFlex Connect.



8. Repeat steps 3 through 7 for each node in the cluster as necessary, until all of the nodes have been re-booted and the new vNICs or vHBAs are present.

ESXi Hypervisor Installation

HX nodes come from the factory with a copy of the ESXi hypervisor pre-installed, however there are scenarios where it may be necessary to redeploy or reinstall ESXi on an HX node. In addition, this process can be used to deploy ESXi on rack mount or blade servers that will function as HX compute-only nodes. The HyperFlex system requires a Cisco custom ESXi ISO file to be used, which has Cisco hardware specific drivers pre-installed, and customized settings configured to ease the installation process. The Cisco custom ESXi ISO file is available to download at cisco.com.

ESXi Kickstart ISO

The HX custom ISO is based on the Cisco custom ESXi 6.5 Update 1 ISO release with the filename: ***HX-Vmware-ESXi-6.5U1-7967591-Cisco-Custom-6.5.1.0.iso*** and is available on the Cisco web site:

https://software.cisco.com/download/home/286305544/type/286305994/release/3.0%25281a%2529

The custom Cisco HyperFlex ESXi ISO will automatically perform the following tasks with no user interaction required:

- Accept the End User License Agreement
- Configure the root password to: Cisco123
- Install ESXi to the internal mirrored Cisco FlexFlash SD cards, or the internal M.2 SSD
- Set the default management network to use vmnic0, and obtain an IP address via DHCP
- Enable SSH access to the ESXi host
- Enable the ESXi shell
- Enable serial port com1 console access to facilitate Serial over LAN access to the host
- Configure the ESXi configuration to always use the current hardware MAC address of the network interfaces, even if they change
- Rename the default vSwitch to vswitch-hx-inband-mgmt

Reinstall HX Cluster

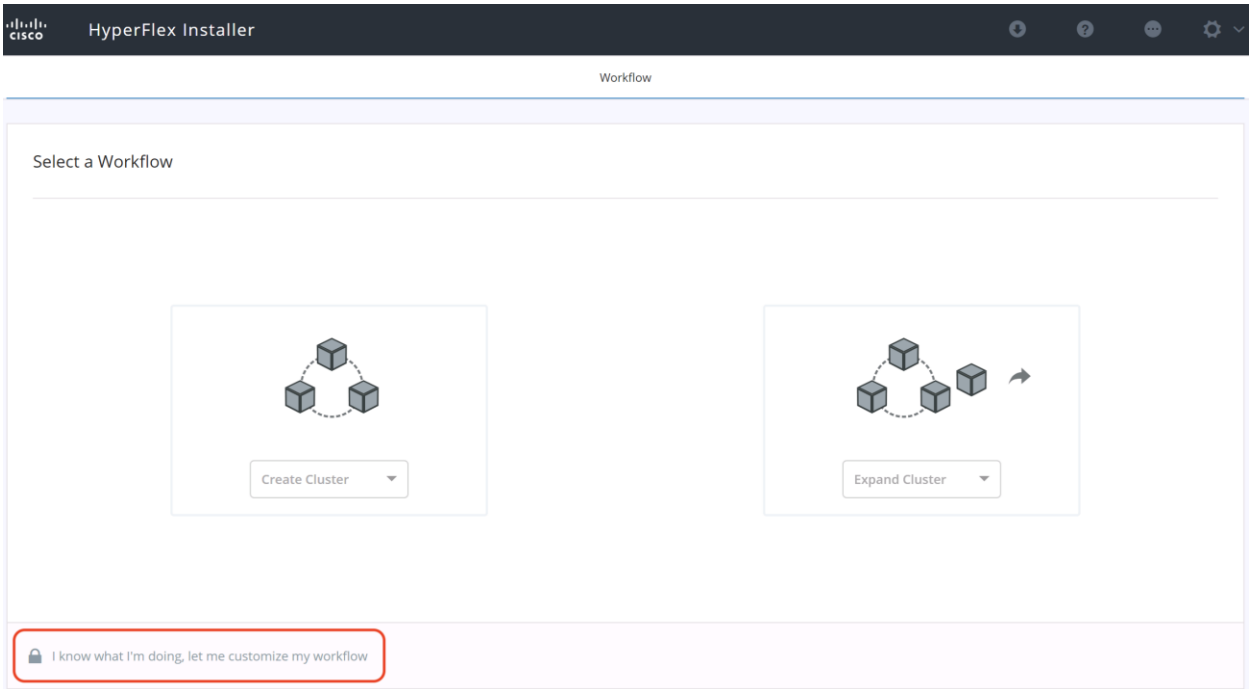
If a Cisco HyperFlex cluster needs to be reinstalled, reference the HyperFlex Cleanup Guide posted here: https://cisco.app.box.com/v/hx-cleanup

Note that the process will be destructive, and result in the loss of all the VMs and all the data stored in the HyperFlex distributed filesystem.

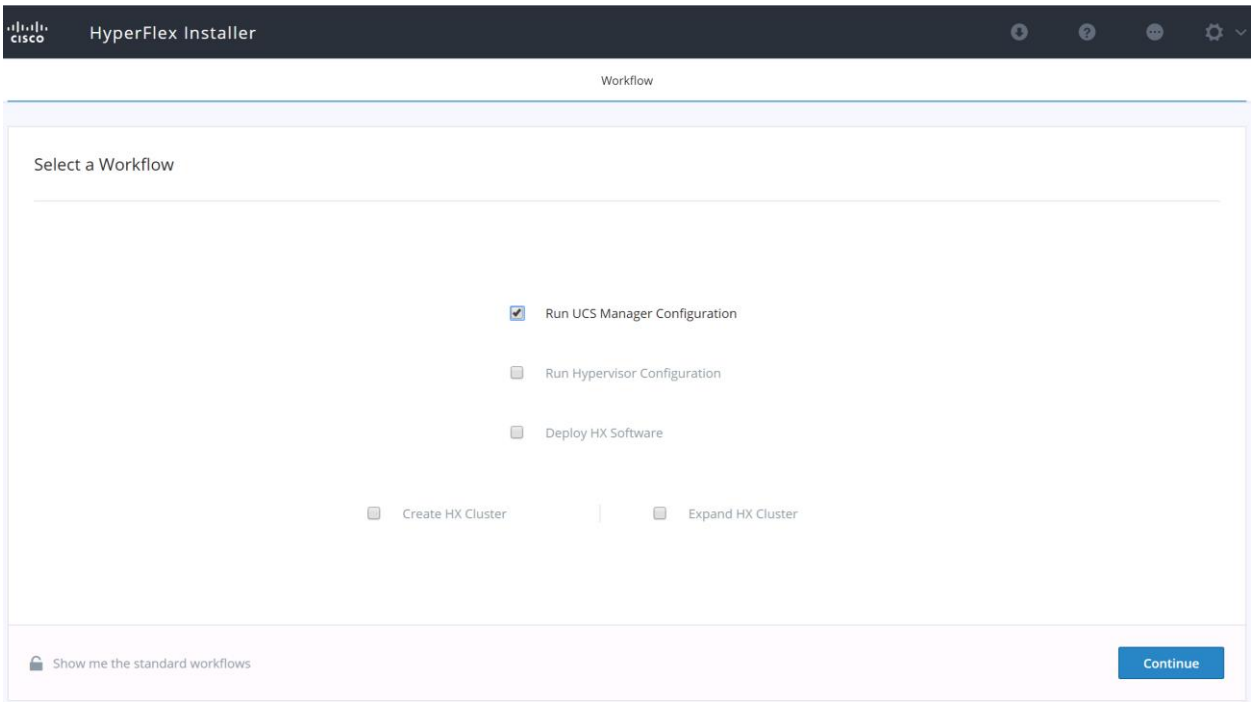
A high-level example of a HX rebuild procedure would be:

1. Clean up the existing environment by:
   - Deleting existing HX virtual machines and HX datastores.
   - Removing the HX cluster in vCenter.
   - Removing vCenter MOB entries for the HX extension.
   - Deleting HX sub-organization and HX VLANs in Cisco UCS Manager.

2. Run HX installer, use the customized version of the installation workflow by selecting the “I know what I am doing” link.

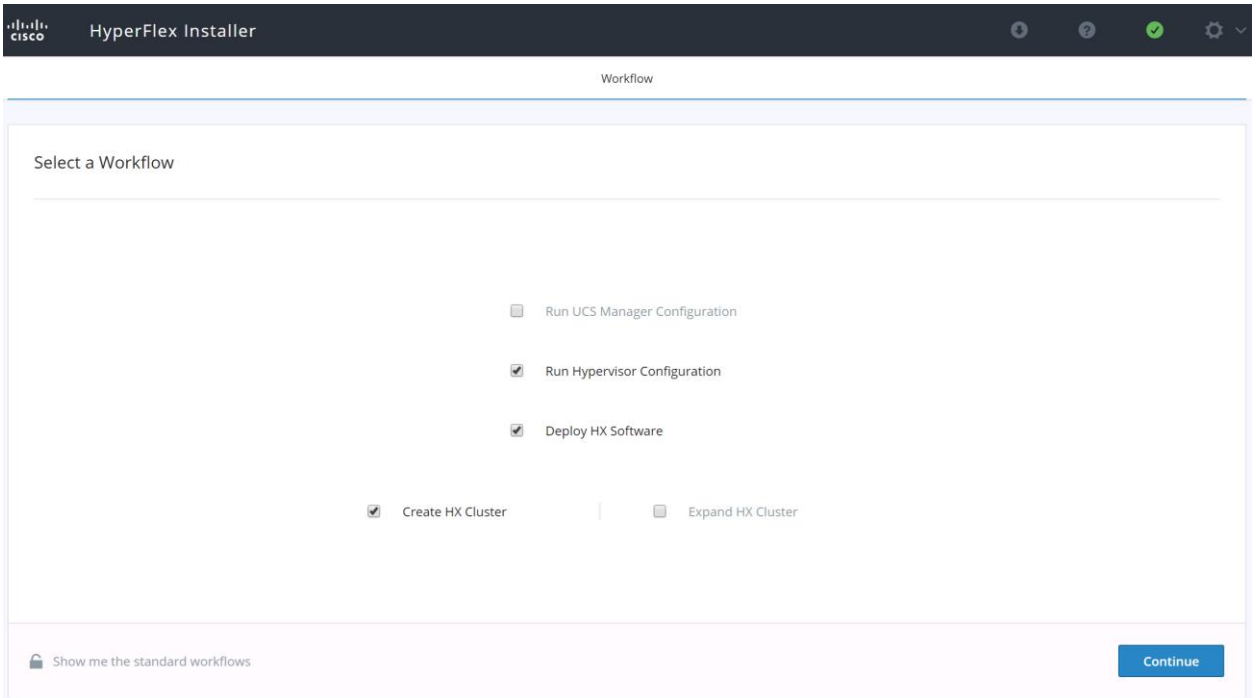


3. Use customized workflow and only choose the “Run UCS Manager Configuration” option, click Continue.



4. When the Cisco UCS Manager configuration is complete, HX hosts are associated with HX service profiles and powered on. Now perform a fresh ESXi installation using the custom ISO image and following the steps in section Cisco UCS vMedia and Boot Policies.

5. When the ESXi fresh installations are all finished, use the customized workflow and select the remaining 3 options; ESXi Configuration, Deploy HX Software, and Create HX Cluster, to continue and complete the HyperFlex cluster installation.



More information on the various installation methods can be found in the Getting Started Guide.

Cisco UCS vMedia and Boot Policies

By using a Cisco UCS vMedia policy, the custom Cisco HyperFlex ESXi installation ISO file can be mounted to all of the HX servers automatically. The existing vMedia policy, named “HyperFlex” must be modified to mount this file, and the boot policy must be modified temporarily to boot from the remotely mounted vMedia file. Once these two tasks are completed, the servers can be rebooted, and they will automatically boot from the remotely mounted vMedia file, installing and configuring ESXi on the servers.

WARNING! While vMedia policies are very efficient for installing multiple servers, using vMedia policies as described could lead to an accidental reinstall of ESXi on any existing server that is rebooted with this policy. Please be certain that the servers being rebooted while the policy is in effect are the servers you wish to reinstall. Even though the custom ISO will not continue without a secondary confirmation, extreme caution is recommended. This procedure needs to be carefully monitored and the boot policy should be changed back to original settings immediately after the intended servers are rebooted, and the ESXi installation begins. Using this policy is only recommended for new installs or rebuilds. Alternatively, you can manually select the boot device using the KVM console during boot, and pressing F6, instead of making the vMedia device the default boot selection.

To configure the Cisco UCS vMedia and Boot Policies, complete the following steps:

1. Copy the ***HX-Vmware-ESXi-6.5U1-7967591-Cisco-Custom-6.5.1.0.iso*** file to the HX Installer VM via SCP or SFTP, placing it in the folder /var/www/localhost/images/.

2. In Cisco UCS Manager, click the Servers button on the left-hand side of the screen.

3. Expand Servers > Policies > root > Sub-Organizations > <<HX_ORG>> > vMedia Policies, and click vMedia Policy HyperFlex.

4. In the configuration pane, click Create vMedia Mount.

5. Enter a name for the mount, for example: ESXi.

6. Select the CDD option.

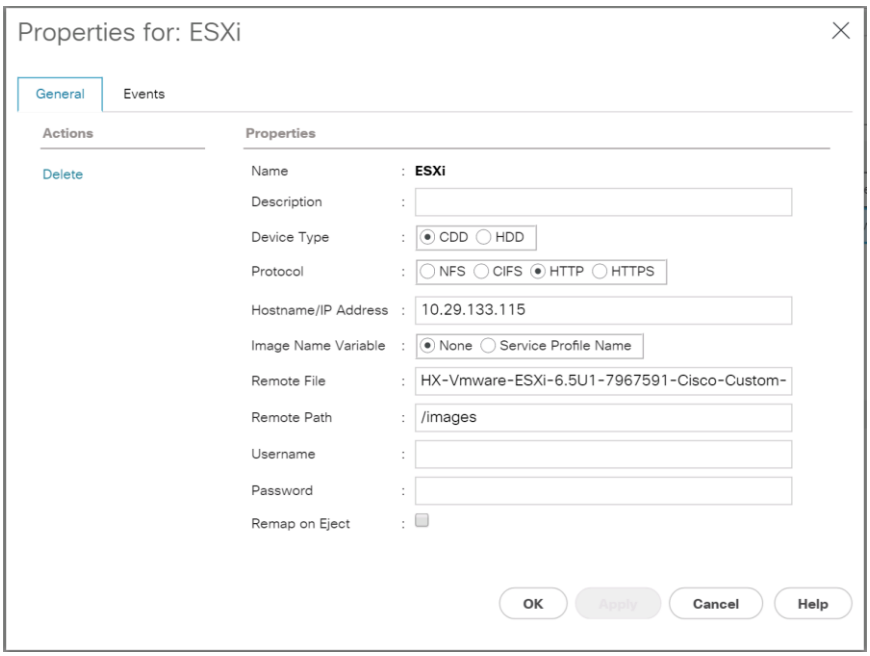
7. Select HTTP as the protocol.

8. Enter the IP address of the HyperFlex installer VM, for example: 10.29.133.115

9. Select None as the Image Variable Name.

10. Enter HX-Vmware-ESXi-6.5U1-7967591-Cisco-Custom-6.5.1.0.iso as the Remote File.

11. Enter /images as the Remote Path.



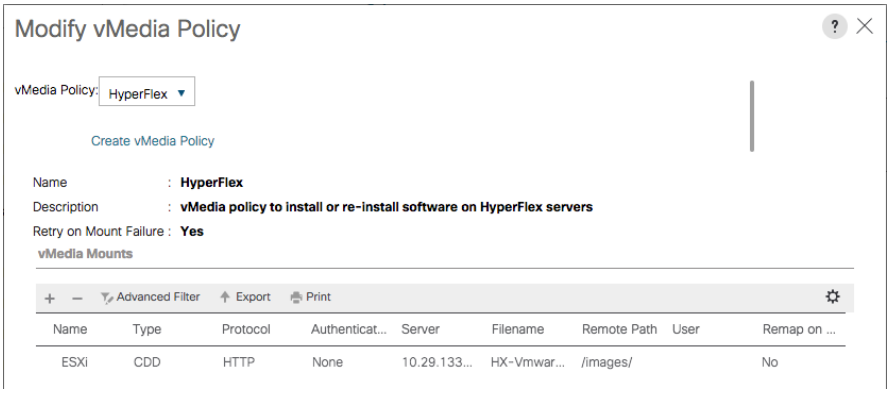
12. Click OK.

13. Select Servers > Service Profile Templates > root > Sub-Organizations > <<HX_ORG>> > Service Template hx-nodes.

14. In the configuration pane, click the vMedia Policy tab.

15. Click Modify vMedia Policy.

16. Chose the HyperFlex vMedia Policy from the drop-down selection and click OK twice.



17. For Compute-Only nodes (if necessary), select Servers > Service Profile Templates > root > Sub-Organizations > <<HX_ORG>> > Service Template compute-nodes.

18. In the configuration pane, click the vMedia Policy tab.

19. Click Modify vMedia Policy.

20. Chose the HyperFlex vMedia Policy from the drop-down selection and click OK twice.

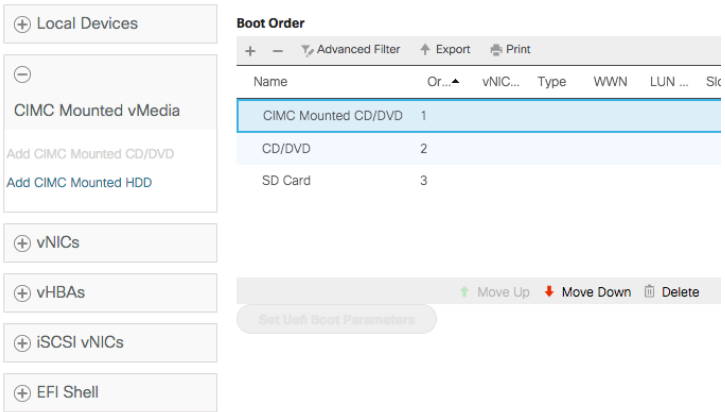
21. Select Servers > Policies > root > Sub-Organizations > <<HX_ORG>> > Boot Policy HyperFlex.

22. In the navigation pane, expand the section titled CIMC Mounted vMedia.

23. Click the entry labeled Add CIMC Mounted CD/DVD.

24. Select the CIMC Mounted CD/DVD entry in the Boot Order list, and click the Move Up button until the CIMC Mounted CD/DVD entry is listed first.

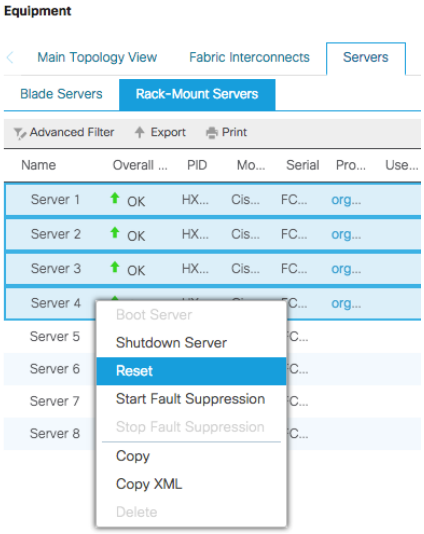
25. Click Save Changes and click OK.



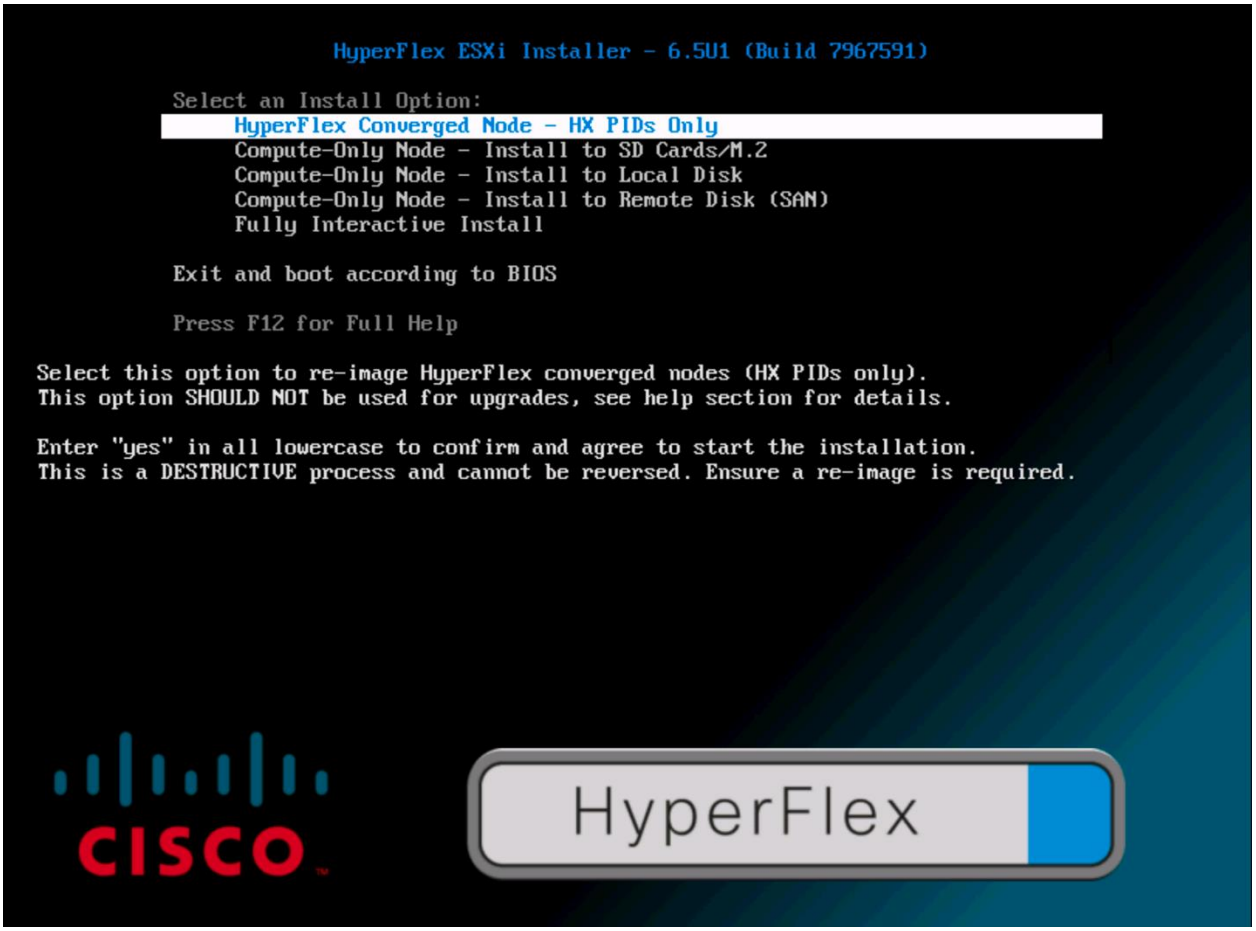
Install ESXi

To begin the installation after modifying the vMedia policy, Boot policy and service profile template, the servers need to be rebooted. To complete the reinstallation, it is necessary to open a remote KVM console session to each server being worked on. To open the KVM console and reboot the servers, complete the following steps:

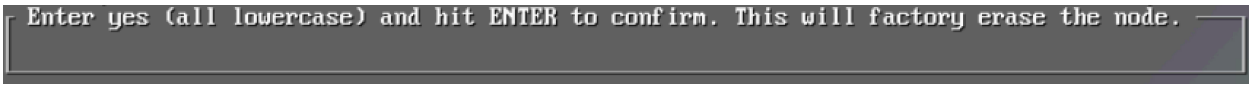
1. In Cisco UCS Manager, click the Equipment button on the left-hand side.
2. Expand Equipment > Rack mounts > Servers > Server 1.
3. In the configuration pane, click KVM Console.
4. The remote KVM Console window will open in a new browser tab. Click Continue to any security alerts that appear, and click the hyperlink to start the remote KVM session.
5. Repeat Steps 2-4 for all additional servers whose console you need to monitor during the installation.
6. In Cisco UCS Manager, click the Equipment button on the left-hand side.
7. Expand Equipment > Rack-Mount Servers > Servers.
8. In the configuration pane, click the first server to be rebooted, then shift+click the last server to be re-booted, selecting all of the servers.
9. Right-click the mouse and click Reset.



10. Click OK.
11. Select Power Cycle and click OK.
12. Click OK. The servers you are monitoring in the KVM console windows will now immediately reboot, and boot from the remote vMedia mount. Alternatively, the individual KVM consoles can be used to perform a power cycle one-by-one.
13. When the server boots from the installation ISO file, you will see a customized Cisco boot menu. In the Cisco customized installation boot menu, select “HyperFlex Converged Node – HX PIDs Only” and press enter.



14. Enter “yes” in all lowercase letters, and press Enter to confirm and install ESXi.



15. (Optional) When installing Compute-Only nodes, the appropriate Compute-Only Node option for the boot location to be used should be selected. The “Fully Interactive Install” option should only be used for debugging purposes.

16. The ESXi installer will continue the installation process automatically, there may be error messages seen on screen, but they can be safely ignored. When the process is complete, the standard ESXi console screen will be seen as below:

Undo vMedia and Boot Policy Changes

When all the servers have booted from the remote vMedia file and begun their installation process, the changes to the boot policy need to be quickly undone, to prevent the servers from going into a boot loop, constantly booting from the installation ISO file. To revert the boot policy settings, complete the following steps:

1. Select Servers > Policies > root > Sub-Organizations > <<HX_ORG>> > Boot Policy HyperFlex.
2. Select the CIMC Mounted CD/DVD entry in the Boot Order list, and click Delete.
3. Click Save Changes and click OK.

The changes made to the vMedia policy and service profile template may also be undone once the ESXi installations have all completed fully, or they may be left in place for future installation work.

HyperFlex Cluster Expansion

The process to expand a HyperFlex cluster can be used to grow an existing HyperFlex cluster with additional converged storage nodes, or to expand an existing cluster with additional compute-only nodes to create an extended cluster.

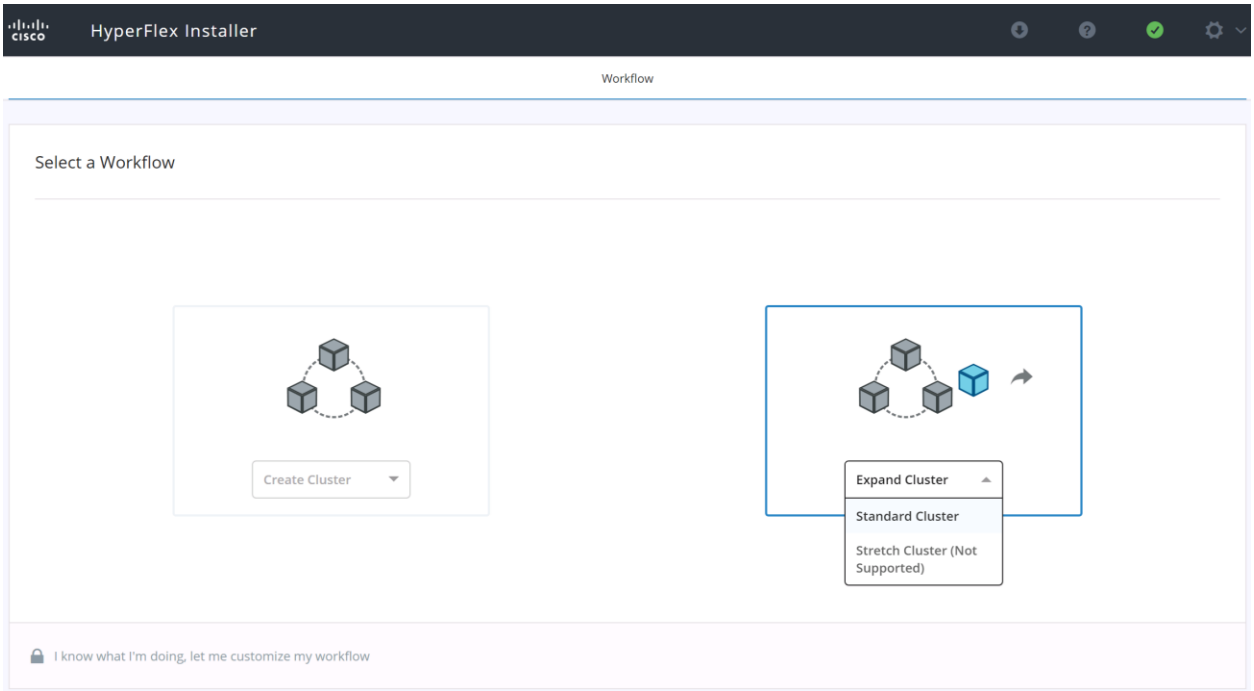
Expansion with Compute-Only Nodes

The following technical guidelines must be followed when adding compute-only nodes to a Cisco HyperFlex cluster:

- The number of compute-only nodes cannot exceed the number of HyperFlex converged nodes within a single HyperFlex cluster.
- Note, that at the time of the publication of this document, it is not supported to add compute-only nodes to a stretched cluster.
- The Cisco UCS infrastructure firmware revision, which provides the firmware for Cisco UCS Manager and the Fabric Interconnects, must be maintained at the minimum version required for the HyperFlex converged nodes, or higher, at all times.
- The version of VMware ESXi installed on the compute-only nodes must be compatible with the Cisco HyperFlex version in use, and it must match the version installed on the HyperFlex converged nodes.
- While the CPU models and memory capacities between the compute-only nodes and the HyperFlex converged nodes do not have to match, configuring the nodes to have similar capacities is recommended.
- Care must be taken that the addition of the compute-only nodes will not significantly impact the HyperFlex cluster by creating additional load, or by consuming too much space. Pay close attention to the space consumption and performance requirements of any net-new VMs that will run on the additional compute-only nodes, and note the current cluster performance and space utilization. If no new VMs will be created, then the current cluster performance will not be impacted.
- Mixing different models of compute-only nodes is allowed within the same cluster. Example: using Cisco UCS C220 M3 and C240 M4 servers as compute-only nodes is allowed.
- Connectivity between compute-only nodes and the HyperFlex cluster must be within the same Cisco UCS domain, and must be 10 GbE or 40 GbE. Connecting compute-only nodes from a different Cisco UCS domain is not allowed, and connecting standalone rack-mount servers from outside of the Cisco UCS domain is not allowed.
- Blade servers installed in the Cisco UCS 5108 Blade Chassis can connect through 10 GbE or 40 GbE chassis links, using the Cisco UCS 2204XP, 2208XP, or 2304 model Fabric Extenders. The Fabric Extenders, also called I/O Modules (IOMs), are typically installed in pairs, and connect the 5108 chassis to the Fabric Interconnects, which provide all the networking and management for the blades. Care must be taken not to oversubscribe and saturate the chassis links.
- Mixing CPU generations will require configuring VMware Enhanced vMotion Compatibility (EVC) in order to allow vMotion to work between the compute-only nodes and the converged nodes. Enabling EVC typically requires all VMs to be powered off. If it is known ahead of time that EVC will be needed, then it is easier to enable EVC on the vCenter cluster prior to installing HyperFlex.
- Compute-only nodes can be configured to boot from SAN, local disks, or internal SD cards. No other internal storage should be present in a compute-only node. Manual configuration of the boot policy will be necessary if booting from any device other than SD cards.
- Compute-only nodes can be configured with additional vNICs or vHBAs in order to connect to supported external storage arrays via NFS, iSCSI or Fibre Channel, in the same way as HyperFlex converged nodes are allowed to do.

The HX installer has a wizard for Cluster Expansion with converged nodes and compute-only nodes, however the compute-only node process requires some additional manual steps to install the ESXi hypervisor on the nodes. To expand an existing HyperFlex cluster with compute-only nodes, creating an extended HyperFlex cluster, complete the following steps:

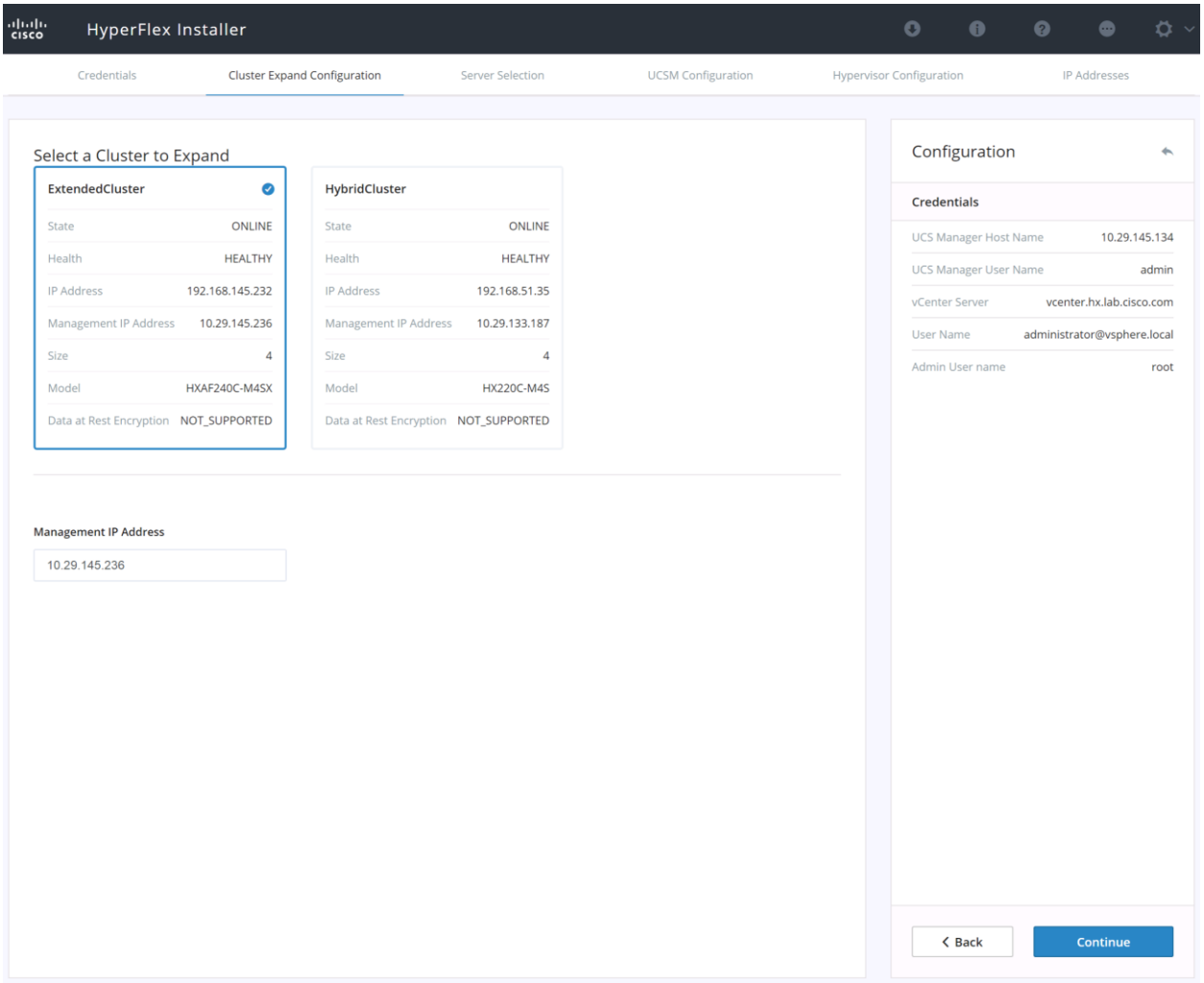
1. On the HyperFlex installer webpage click the dropdown menu for Expand Cluster, then click Standard Cluster.



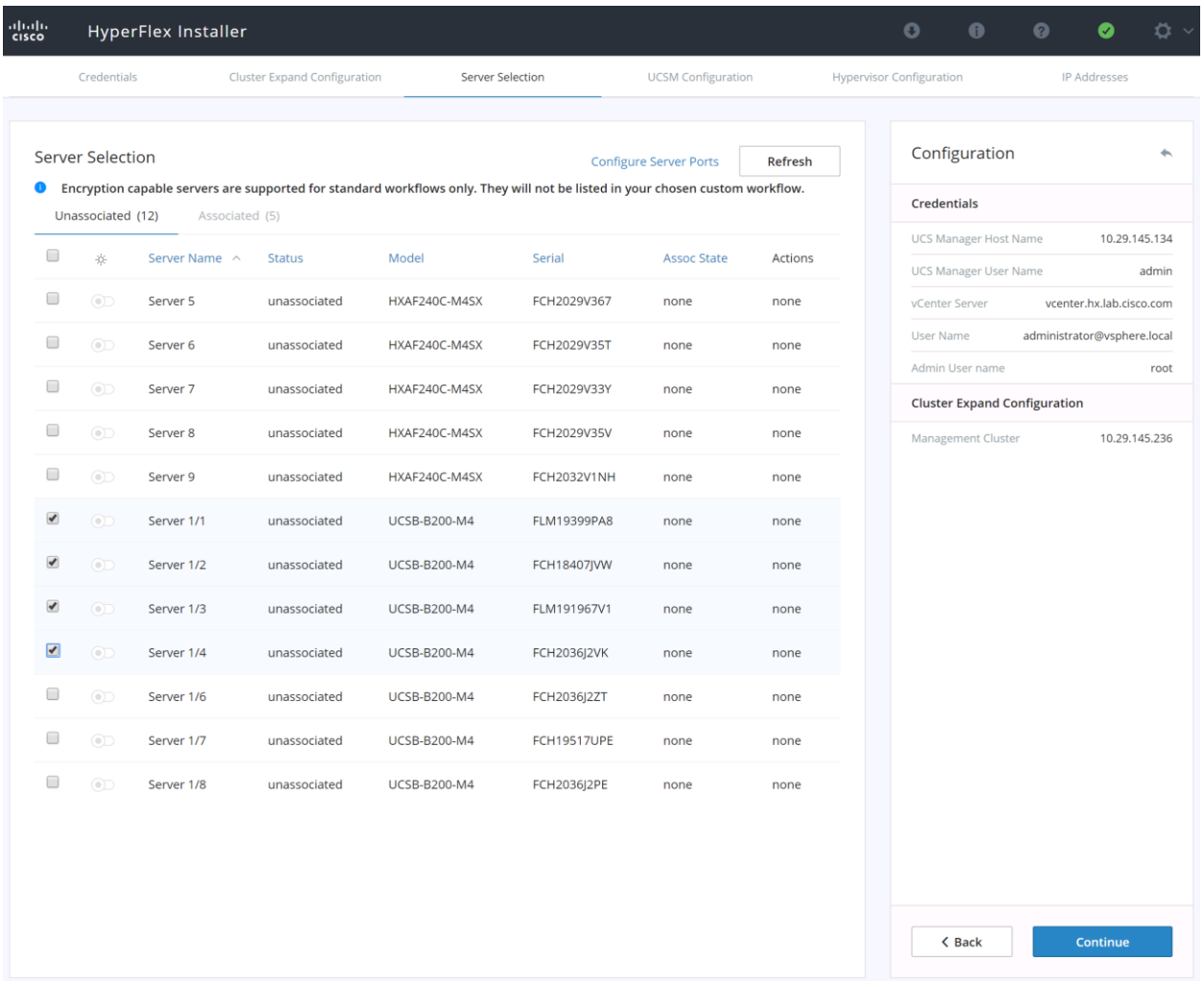
2. Enter the Cisco UCS Manager and vCenter DNS hostname or IP address, the admin usernames, and the passwords. The default Hypervisor credential which comes installed from the factory is username: root with a password of “Cisco123” and are already entered in the installer. You can select the option to see the passwords in clear text. Optionally, you can import a JSON file that has the configuration information, except for the appropriate passwords.

3. Click Continue.

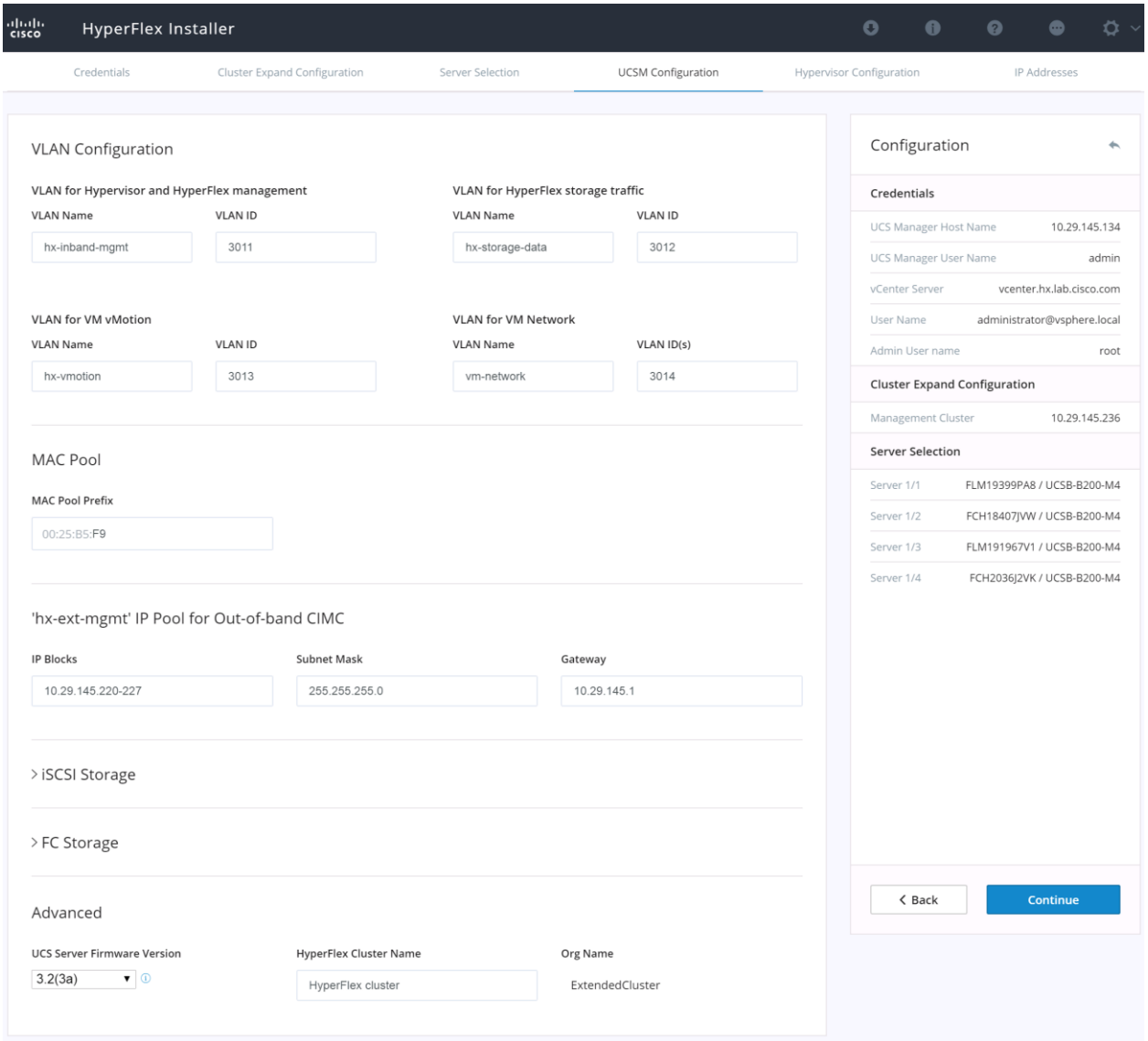
4. Select the HX cluster to expand and click Continue. If the installer has been reset, or otherwise does not detect a cluster to expand, enter the HX cluster management IP address instead.



5. From the list of unassociated servers, select the blade or rack-mount servers you wish to add to the cluster as compute-only nodes, then click Continue.

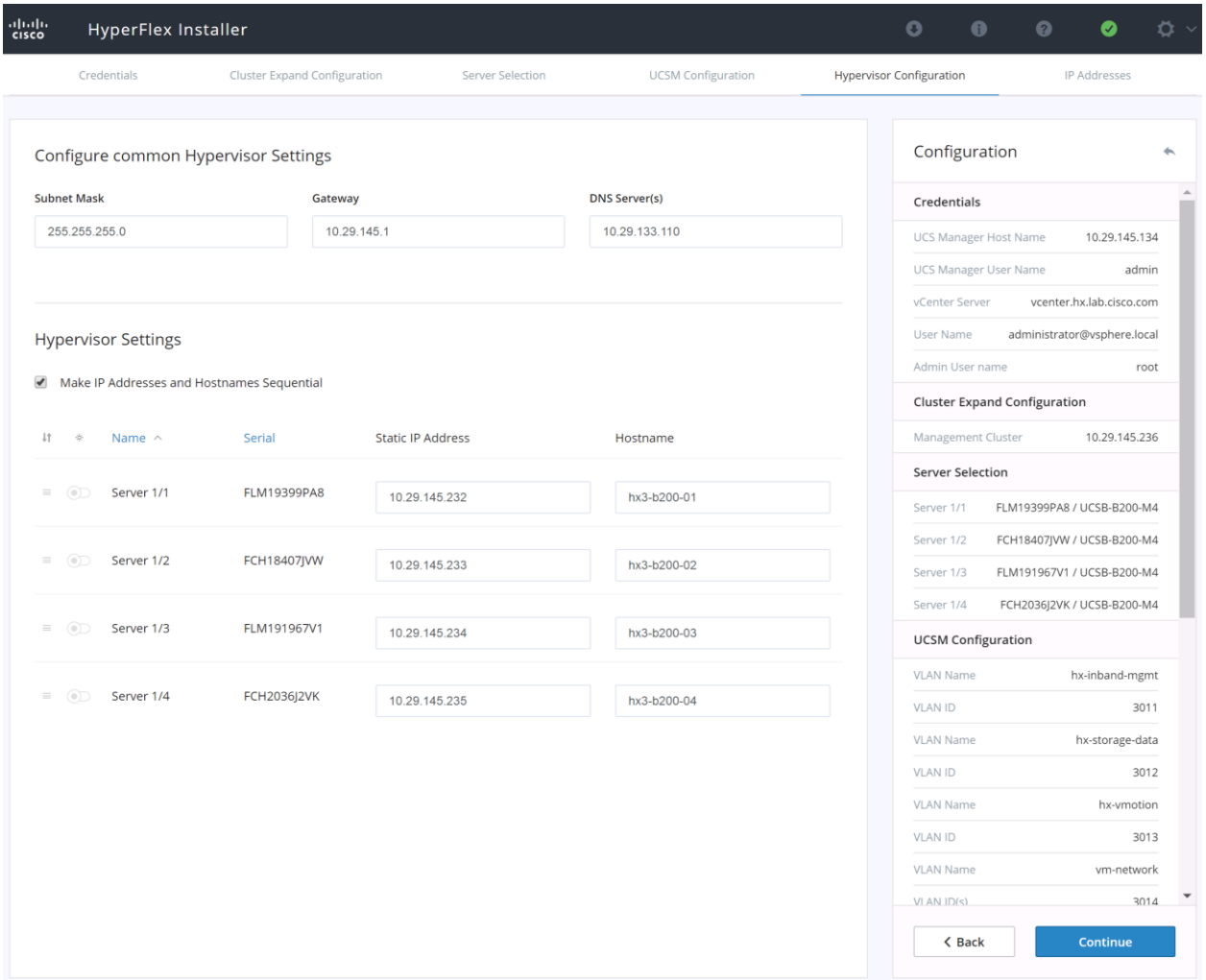


6. On the Cisco UCS Manager Configuration page, enter the VLAN settings, Mac Pool Prefix, UCS hx-ext-mgmt IP Pool for CIMC, iSCSI Storage setting, FC Storage setting, and UCS Firmware version, making sure that all the values match the existing settings for the cluster being expanded.



7. Click Continue.

8. Enter the subnet mask, gateway, DNS, and IP addresses for the Hypervisors (ESXi hosts) as well as host names. The IPs will be assigned through Cisco UCS Manager to the new ESXi hosts.



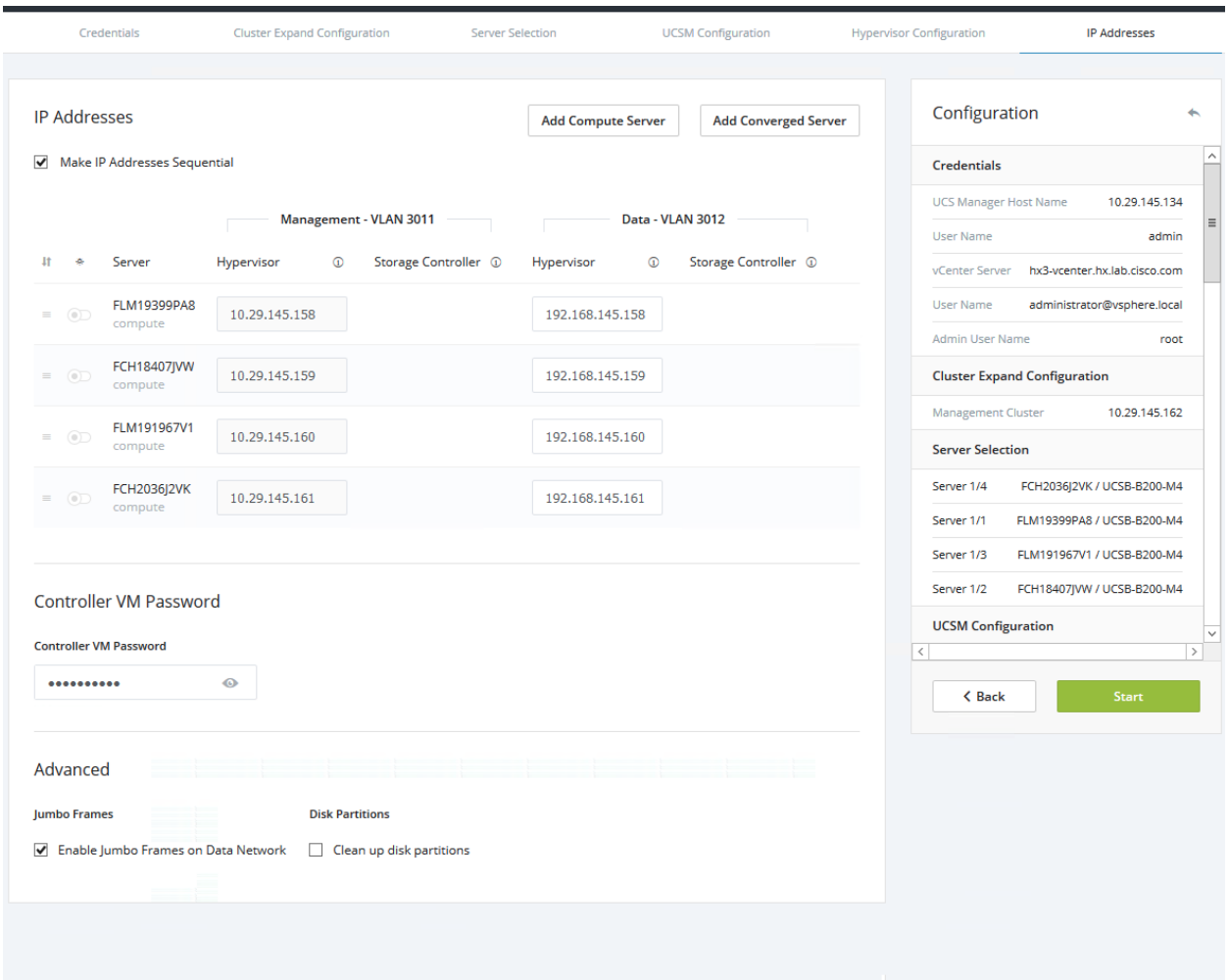
9. Click Continue.

10. Enter the additional IP addresses for the Hypervisor Data network of the new ESXi hosts.

11. Enter the current password that is set on the Controller VMs.

12. Since compute-only nodes have no local storage disks, you do not need to select Clean up disk partitions.

13. (Optional) At this step you can manually add more servers for expansion if these servers already have service profiles associated and the hypervisor is ready, by clicking on Add Compute Server or Add Converged Server and then entering the IP addresses for the storage controller management and data networks.

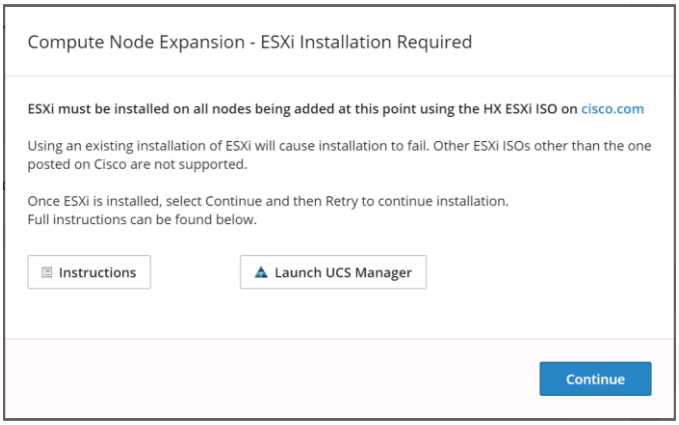


14. Click Start.

15. Validation of the configuration will now start. After validation, the installer will create the compute-only node service profiles and associate them with the selected servers. Once the service profiles are associated, the installer will move on to the Hypervisor Configuration step. If the hypervisor is already installed, the move ahead to step 39. If the ESXi hypervisor has not been previously installed on the compute-only nodes, the installer will stop. The message shown alerts you to the need to install the ESXi hypervisor onto the compute-only nodes. Continue to step 16 and do not click Continue until the hypervisor has been installed. The following steps show how to install ESXi onto the new compute-only nodes.

16. Click the Instructions button to see the steps in a PDF document. If necessary, click the Launch UCS Manager button to log in to Cisco UCS Manager in another browser tab.

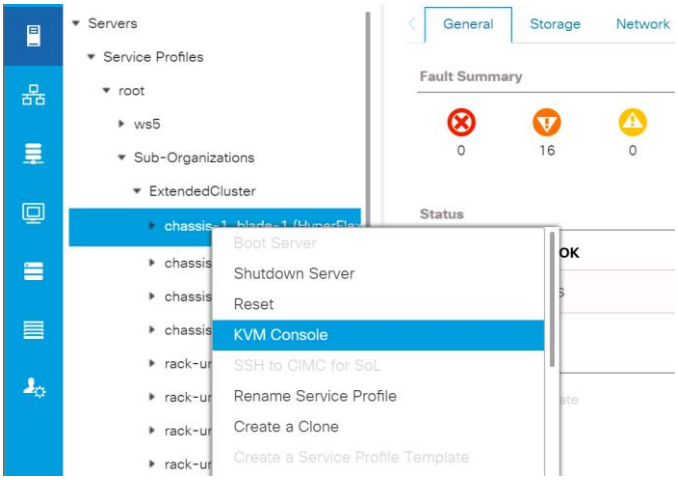
Note: Do not click Continue at this time.



17. In Cisco UCS Manager, click the Servers button on the left-hand side.

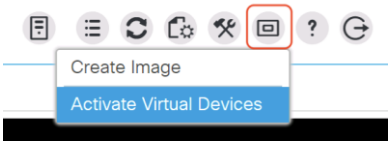
18. Expand Servers > Service Profiles > root > Sub-Organizations > <<HX_ORG>>.

19. Each new compute-only node will have a new service profile, for example: chassis-1_blade-1. Right-click the new service profile and click KVM Console. The remote KVM console will open in a new browser tab. Accept any SSL errors or alerts, then click the link to launch the KVM console.

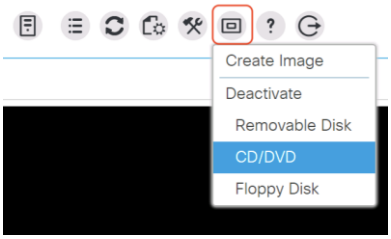


20. Repeat step 19 for each new service profile, that is associated with the new compute-only nodes.

21. In the remote KVM tab, click the Virtual Media button in the top right-hand corner of the screen, then click Activate Virtual Devices.

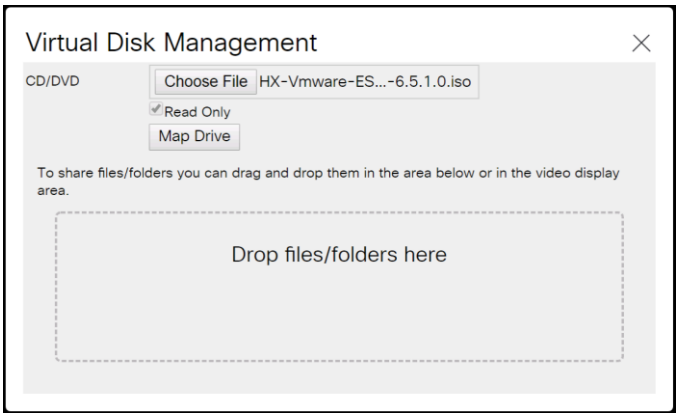


22. In the remote KVM tab, click the Virtual Media button in the top right-hand corner of the screen, then click the CD/DVD option.



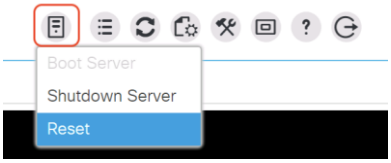
23. Click Choose File, browse for the Cisco custom ESXi ISO installer file, and click Open.

24. Click Map Drive.



25. Repeat steps 21-24 for all the new compute-only nodes.

26. In the remote KVM tab, click the Server Actions button in the top right-hand corner of the screen, the click Reset.



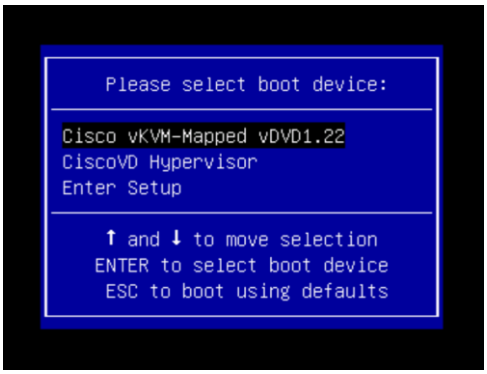
27. Click OK.

28. Choose the Power Cycle option, then click OK.

29. Click OK.

30. Observe the server going through the POST process until the following screen is seen. When it appears, press the F6 key to enter into the boot device selection menu.

31. Select Cisco vKVM-mapped vDVD1.22, then press Enter.

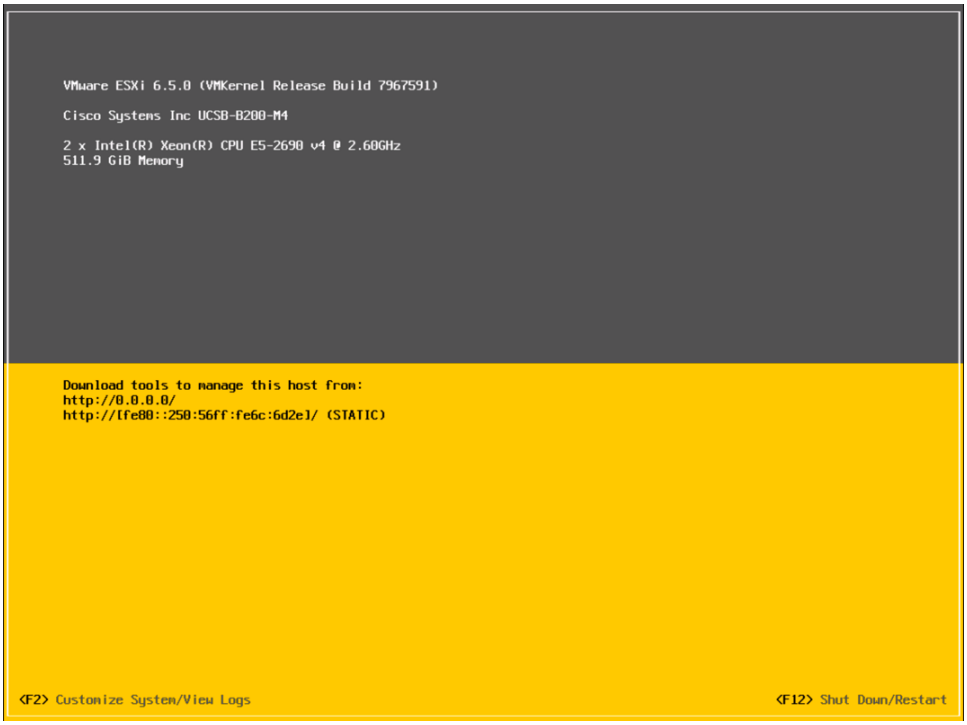


32. The server will boot from the remote KVM mapped ESXi ISO installer and display the following screen:

33. Select the appropriate installation option for the compute-only node you are installing, either installing to SD cards, local disks, or booting from SAN, then press Enter.

34. Type “yes” and press Enter to accept the warning and continue the installation.

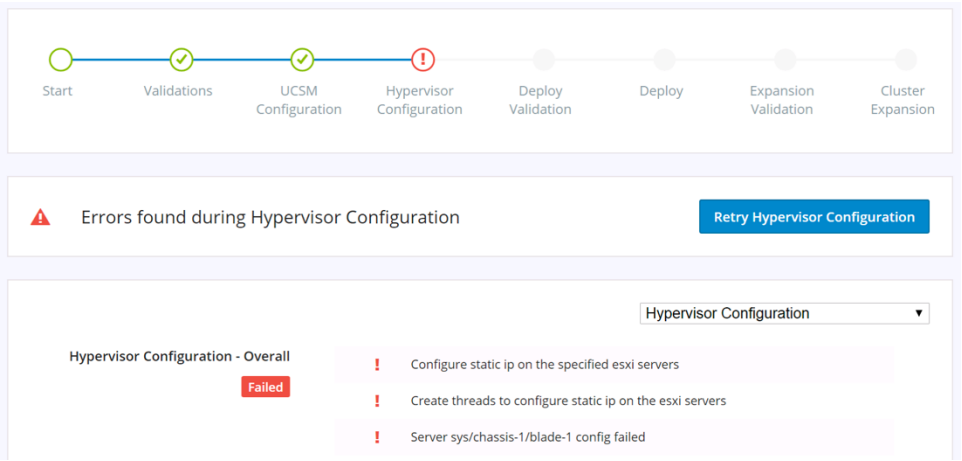
35. The ESXi installer will now automatically perform the installation to the boot media. As you watch the process, some errors may be seen, but they can be ignored. Once the new server has completed the ESXi installation, it will be waiting at the console status screen seen below.



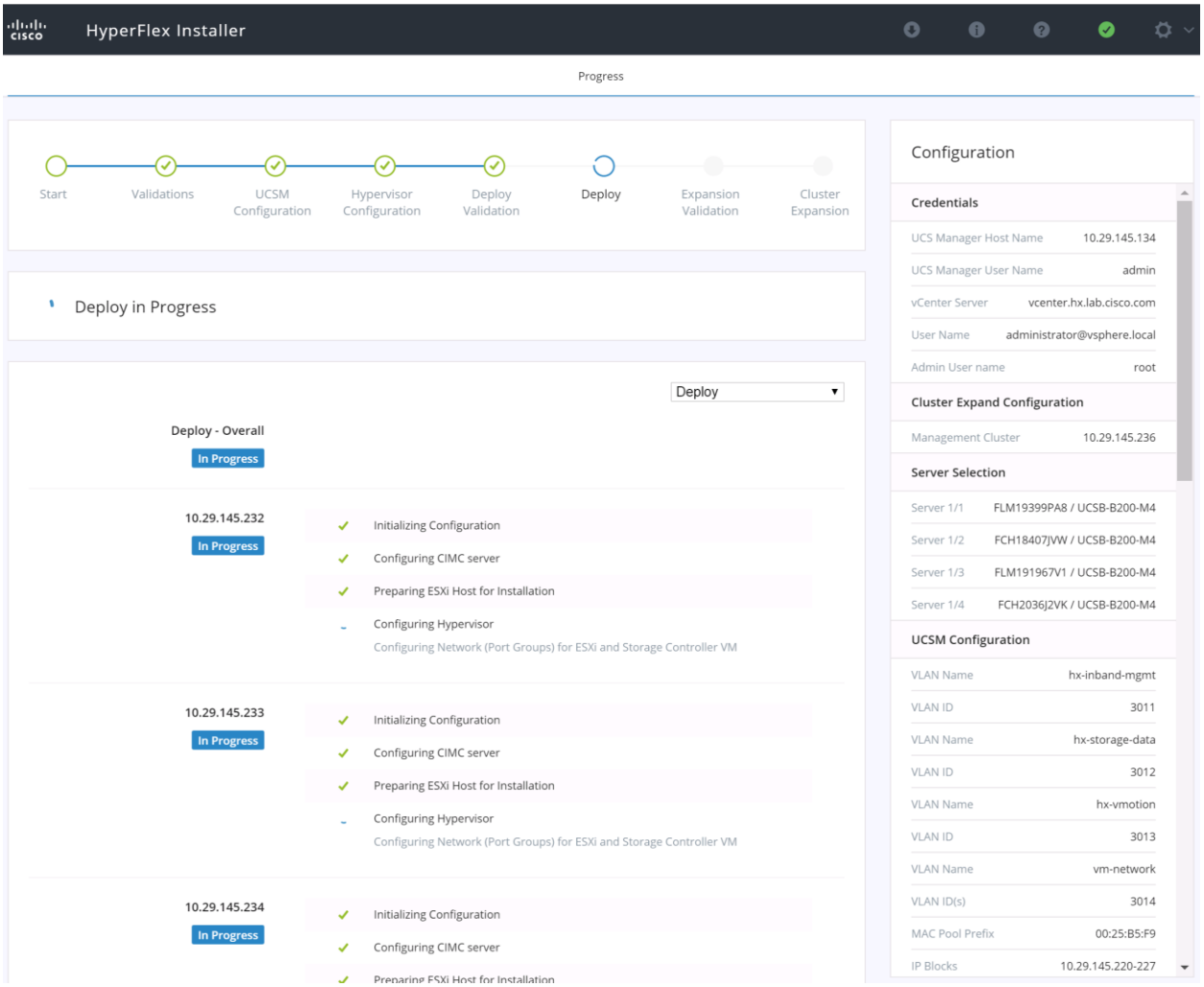
36. Repeat steps 26-35 for all the additional new compute-only nodes being added to the HX cluster.

37. When all the new nodes have finished their fresh ESXi installations, return to the HX installer, where the error in step 15 was seen. Click Continue.

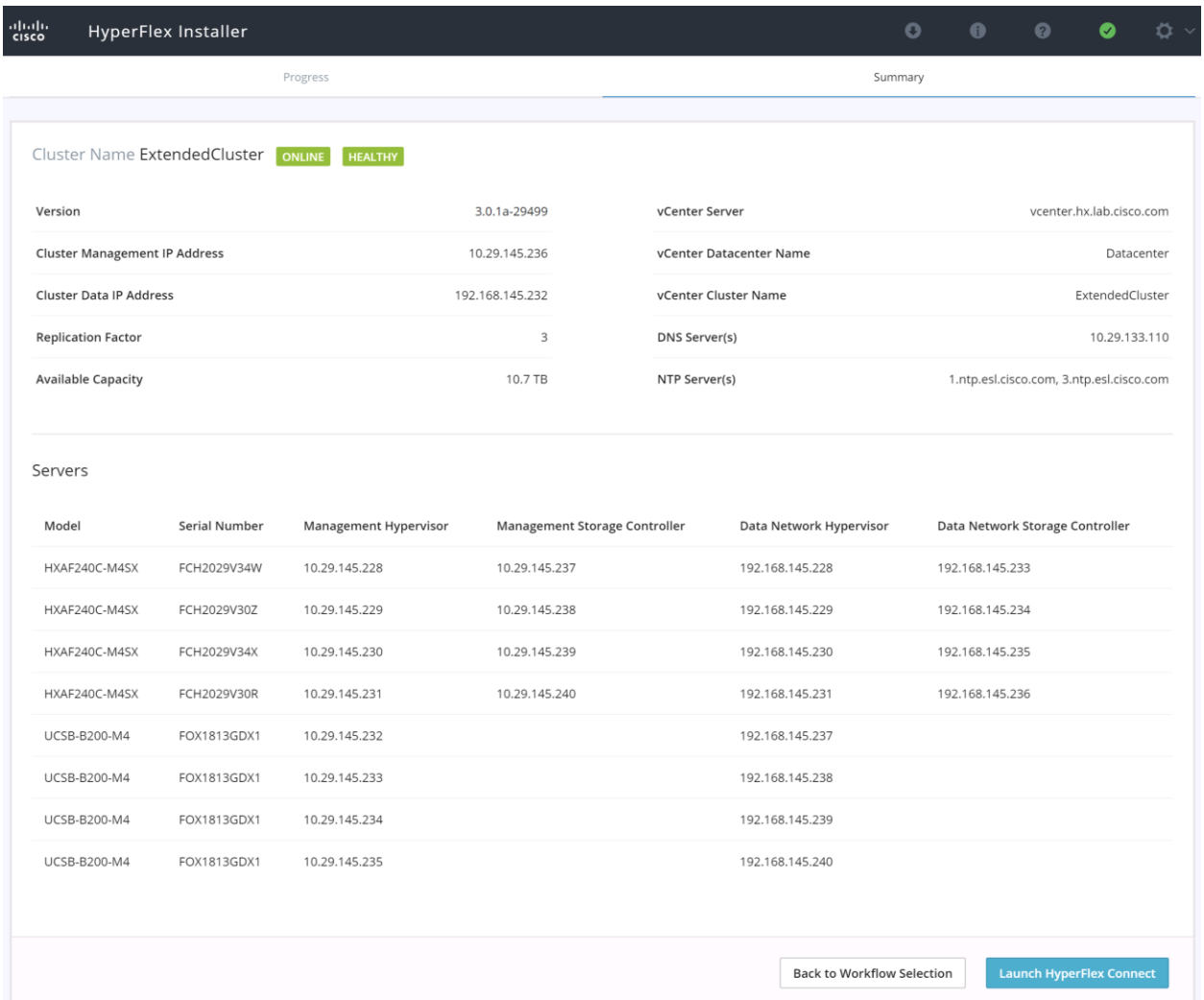
38. Click Retry Hypervisor Configuration.



39. The HX installer will now proceed to complete the deployment and perform all the steps listed at the top of the screen along with their status.



40. When the expansion is completed, a summary screen showing the status of the expanded cluster and the expansion operation is shown.



After the install has completed, the compute-only nodes are added to the cluster and now have access to the existing HX datastores, but some manual post installation steps are required. Most steps can be done by running the post_install script from the HX Installer VM, similar to when performing a new installation, or via a custom script. A list of additional configuration steps necessary includes:

- Disable SSH warning
- Creation of the guest VM port groups
- Creation of the vMotion vmkernel port
- Syslog Server Configuration

Note: If at a later time the post_install script needs to be run against a specific HX cluster, the cluster can be specified by using the --cluster-ip switch, and entering the cluster's management IP address.

Example: PowerCLI script to complete tasks on the ESXi host.

```
# Configure_ESXi_post_install.ps1
# Description: Configures ESXi options and settings after HyperFlex installation.
# Usage: Modify the variables to specify the ESXi root password, the servers to be
# configured, the guest VLAN ID, and the IP addresses used for the vMotion VMkernel
```

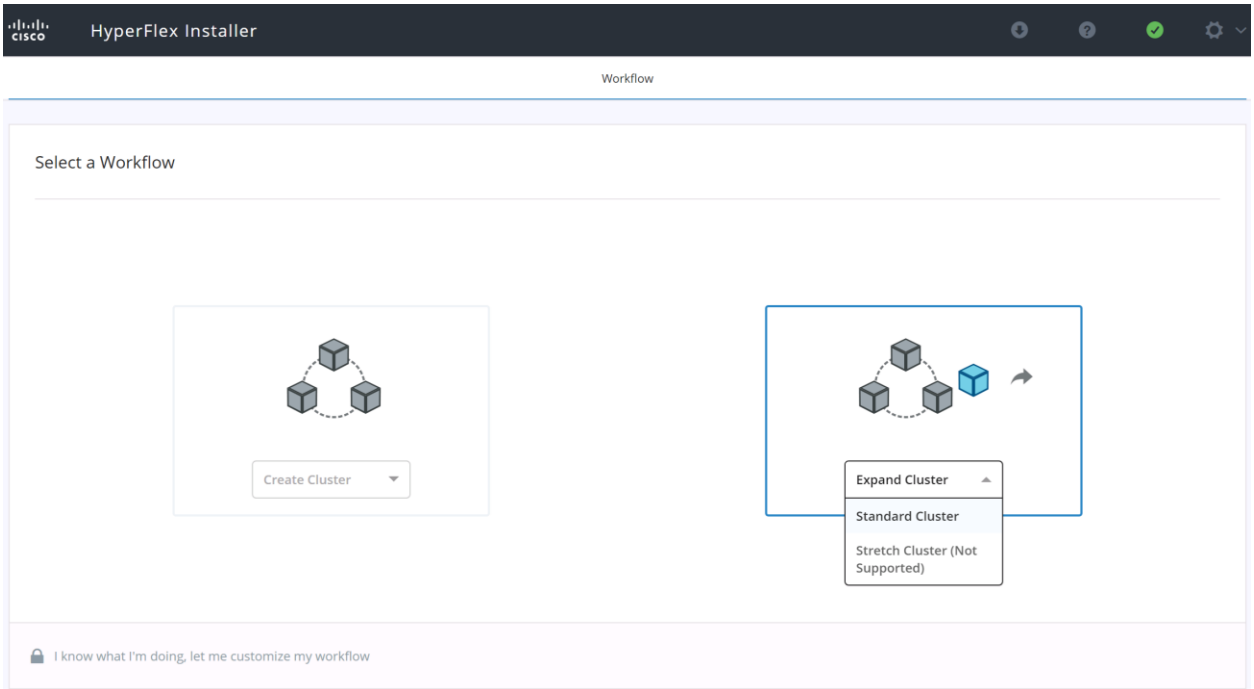
```
# interfaces.
#
Set-PowerCLIConfiguration -InvalidCertificateAction Ignore -Confirm:$false | Out-Null
$domainname ="hx.lab.cisco.com"
$rootpw = "Cisco123"
$servers="hx220-01.hx.lab.cisco.com","hx220-02.hx.lab.cisco.com","hx220-
03.hx.lab.cisco.com","hx220-04.hx.lab.cisco.com","hx220-05.hx.lab.cisco.com","hx220-
06.hx.lab.cisco.com","hx220-07.hx.lab.cisco.com","hx220-08.hx.lab.cisco.com"
$ip=11

Foreach ($server in $servers) {

# connect to the ESXi host server
Connect-VIServer -server $server -user root -password $rootpw
$vmhost = Get-VMHost -Name $server

#disable shell warning
$vmhost | Set-VMHostAdvancedConfiguration UserVars.SuppressShellWarning 1

#configure syslog traffic to send to vCenter or syslog server
Set-VMHostSysLogServer -SysLogServer '10.29.133.63:514' -VMHost $vmhost

# retrieve the virtual switch configurations
$vswitch2 =  Get-VirtualSwitch -VMHost $vmhost -Name vswitch-hx-vm-network
$vswitch3 =  Get-VirtualSwitch -VMHost $vmhost -Name vmotion

# create a port group for the guest VMs
New-VirtualPortGroup -VirtualSwitch $vswitch2 -Name "VM-Network" -VLanID 100

# create the vmotion port group and VMkernel interface
$vmip="192.168.233."+$ip
New-VMHostNetworkAdapter -VMHost $vmhost -VirtualSwitch $vswitch3 -PortGroup "vmotion" -Mtu 9000
-VMotionEnabled $true -IP $vmip -SubnetMask 255.255.255.0 -Confirm:$false
$ip=$ip+1

Disconnect-VIServer -server $server -Confirm:$false
}
```

To validate our configuration, vMotion a VM to the new compute-only node. You can validate that your VM is now running on the compute only node through the Summary tab of the VM.

Expansion with Converged Nodes

The HX installer has a wizard for Cluster Expansion with Converged Nodes. This procedure is very similar to the initial HyperFlex cluster setup. The following process assumes a new Cisco HX node has been ordered, therefore it is pre-configured from the factory with the proper hardware, firmware, and ESXi hypervisor installed. Note, that at the time of the publication of this document, it is not supported to add converged nodes to a stretched cluster. To add converged storage nodes to an existing HyperFlex cluster, complete the following steps:

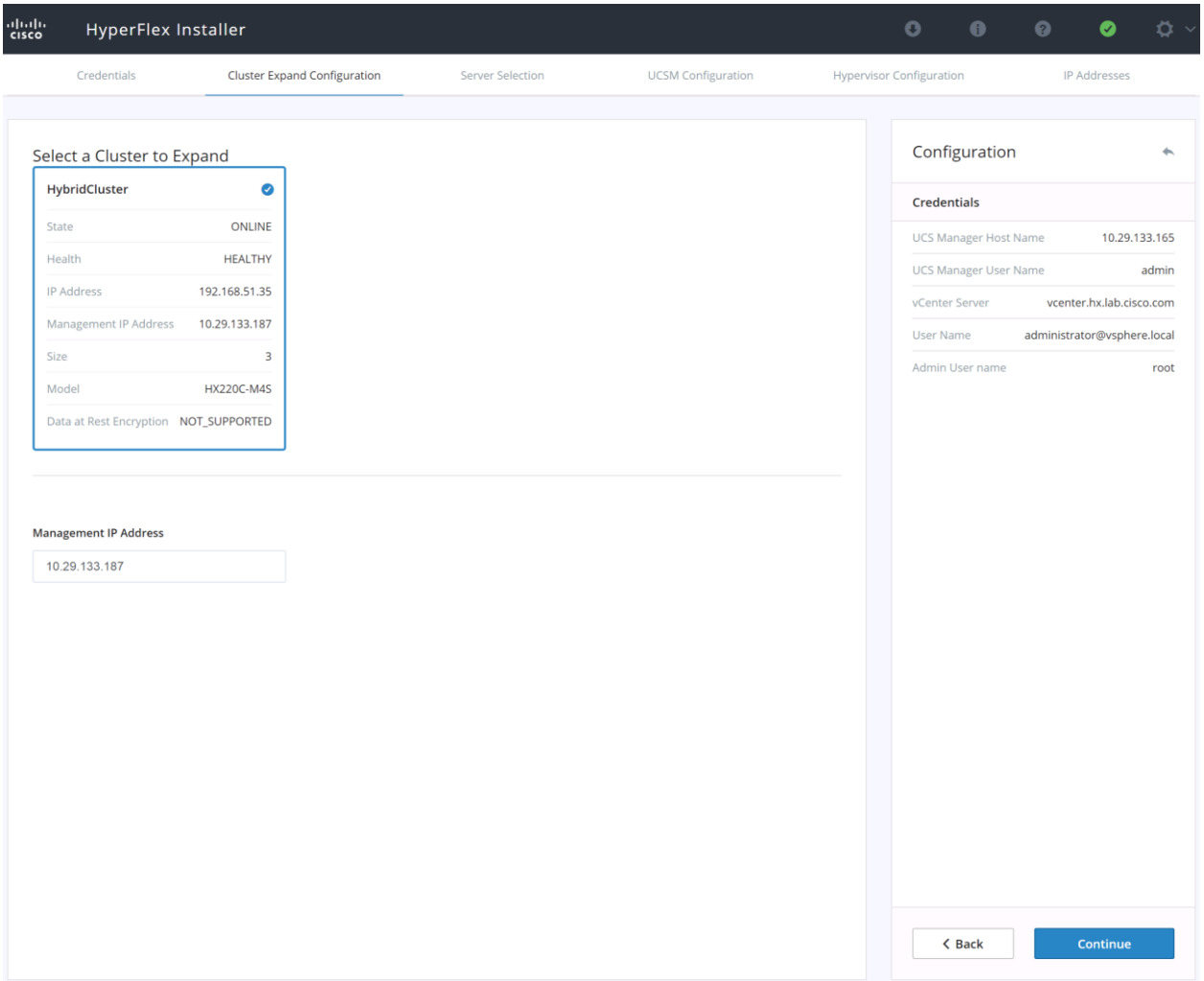
1. On the HyperFlex installer webpage click the dropdown menu for Expand Cluster, then click Standard Cluster.



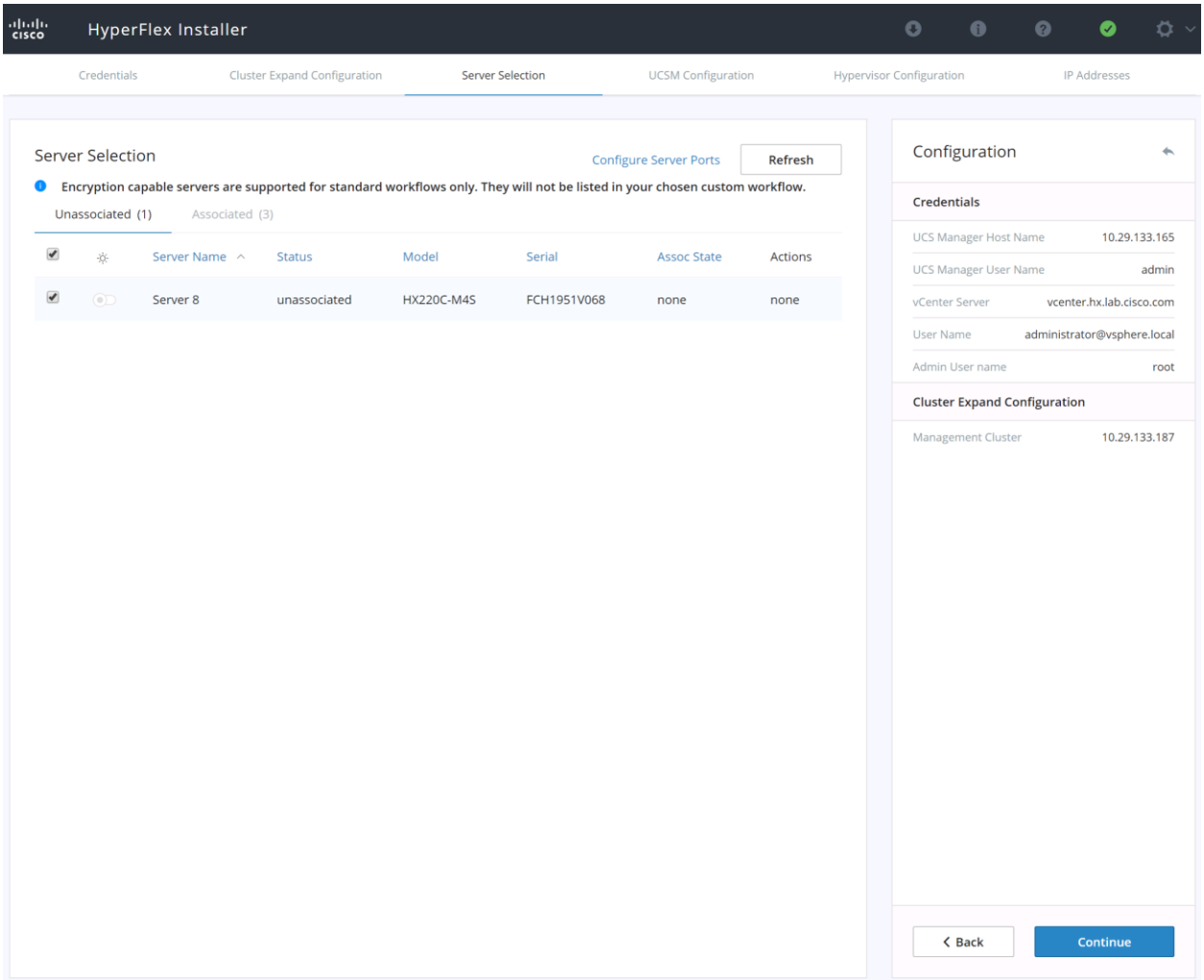
2. Enter the Cisco UCS Manager and vCenter DNS hostname or IP address, the admin usernames, and the passwords. The default Hypervisor credential which comes installed from the factory is username: root with a password of “Cisco123” and are already entered in the installer. You can select the option to see the passwords in clear text. Optionally, you can import a JSON file that has the configuration information, except for the appropriate passwords.

3. Click Continue.

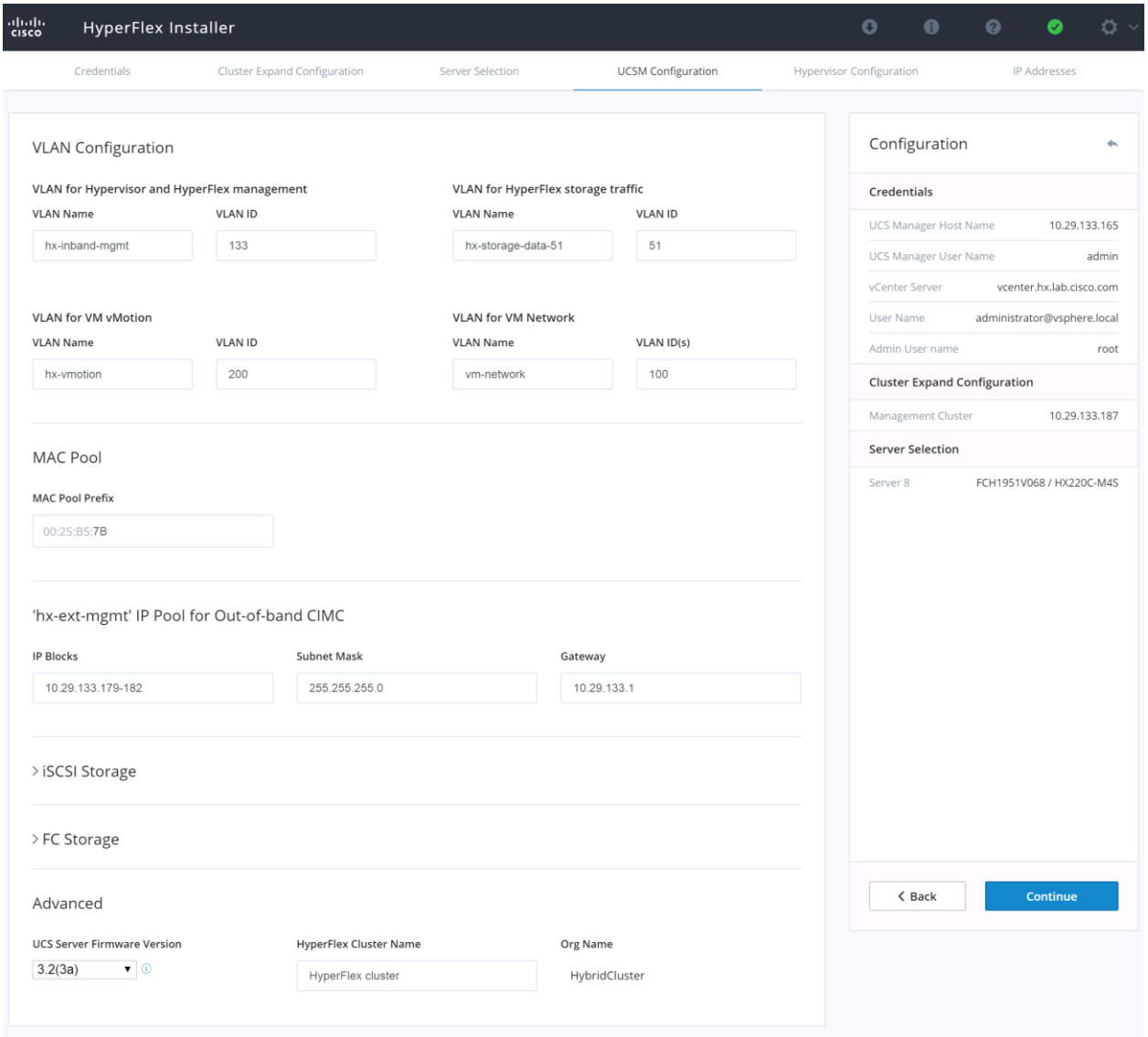
4. Select the HX cluster to expand and click Continue. If the installer has been reset, or otherwise does not detect a cluster to expand, enter the HX cluster management IP address instead.



5. Select the unassociated HX servers you want to add to the existing HX cluster. Click Continue.

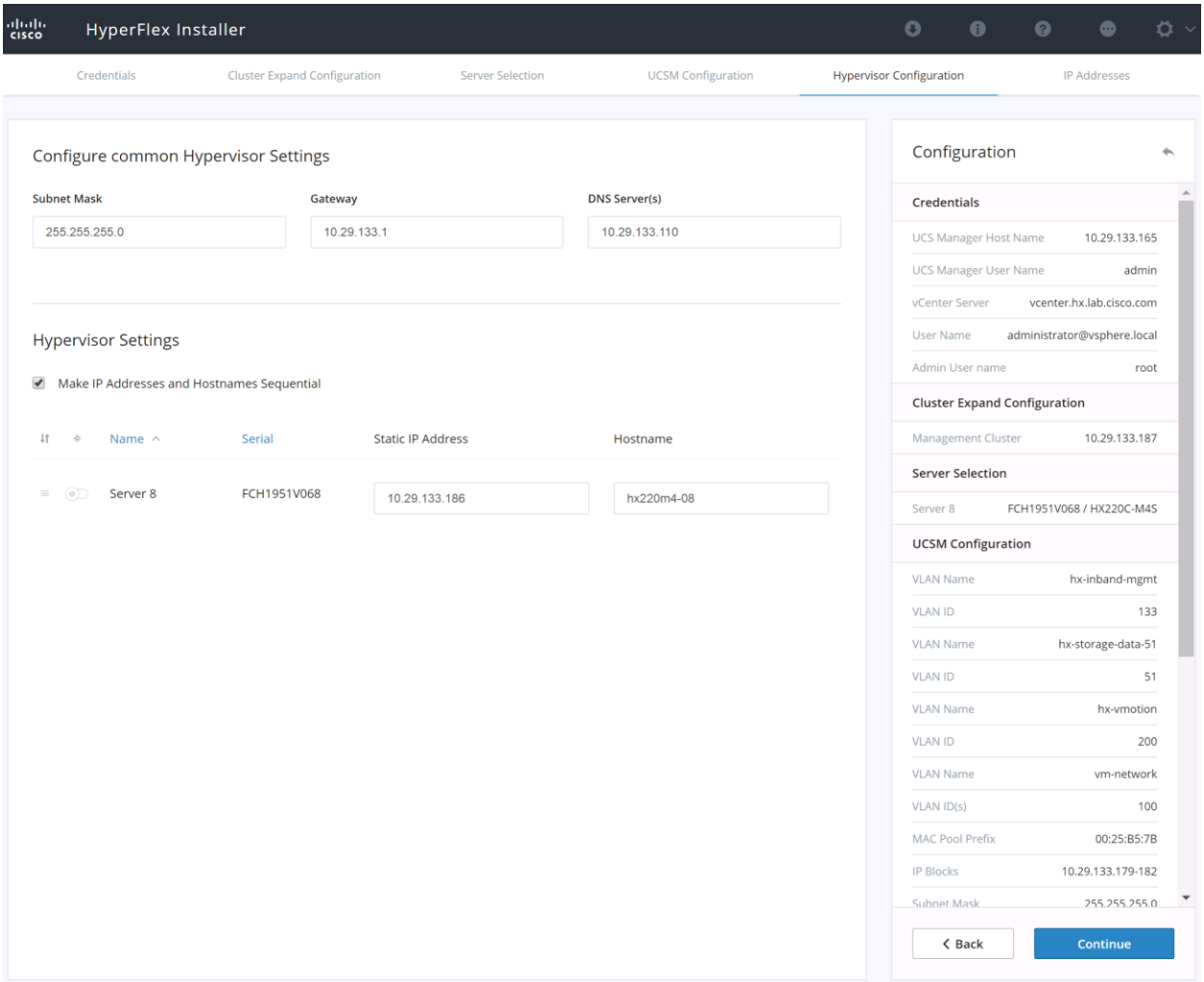


6. On the Cisco UCS Manager Configuration page, enter the VLAN settings, Mac Pool Prefix, UCS hx-ext-mgmt IP Pool for CIMC, iSCSI Storage setting, FC Storage setting, UCS Firmware version, making sure that all the values match the existing settings for the cluster being expanded.



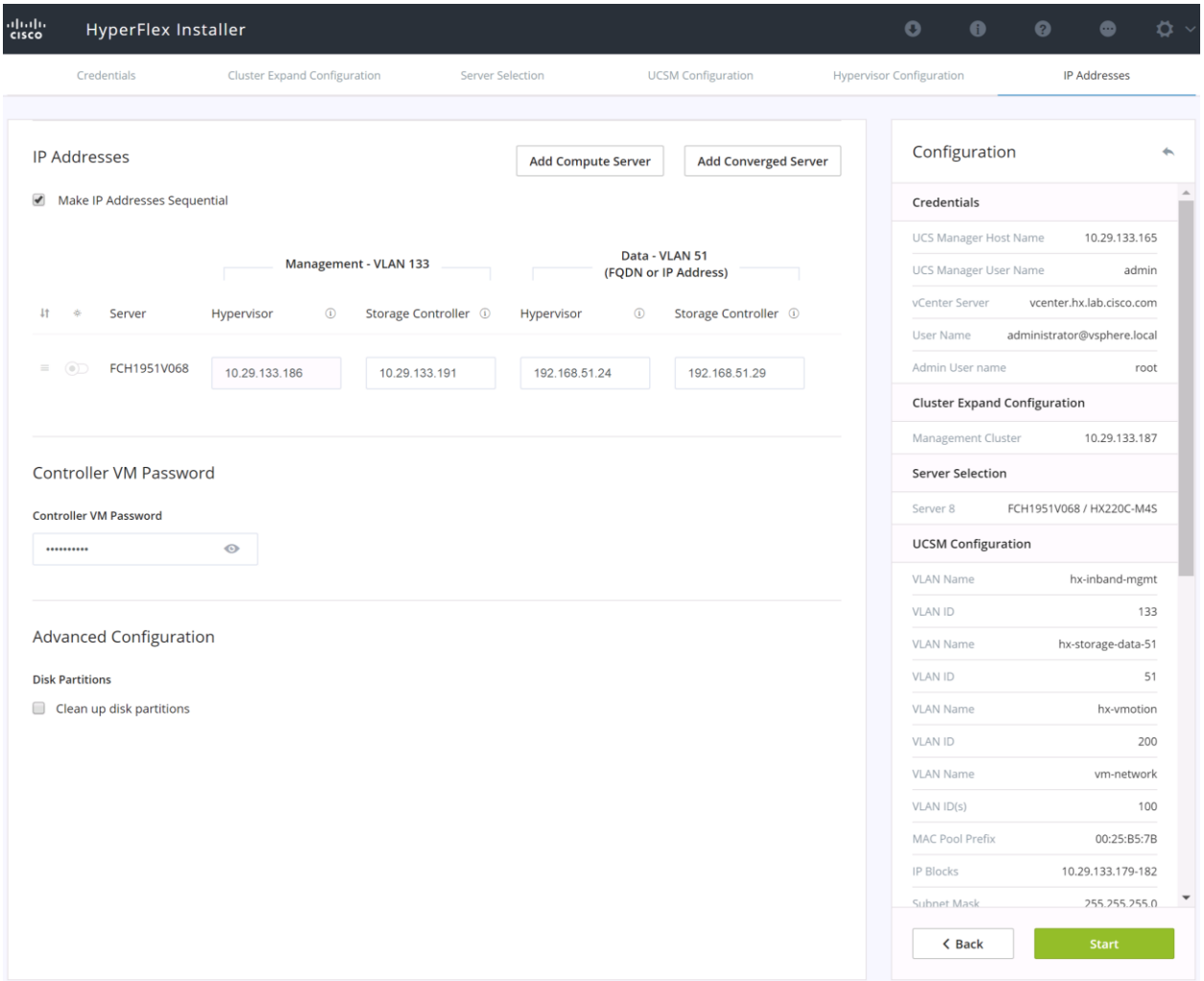
7. Click Continue.

8. Enter the subnet mask, gateway, DNS, and IP addresses for the Hypervisors (ESXi hosts) as well as host names. The IPs will be assigned through Cisco UCS Manager to ESXi systems.



9. Click Continue.

10. Enter the additional IP addresses for the Management and Data networks of the new nodes.



11. Enter the current password that is set on the Controller VMs.

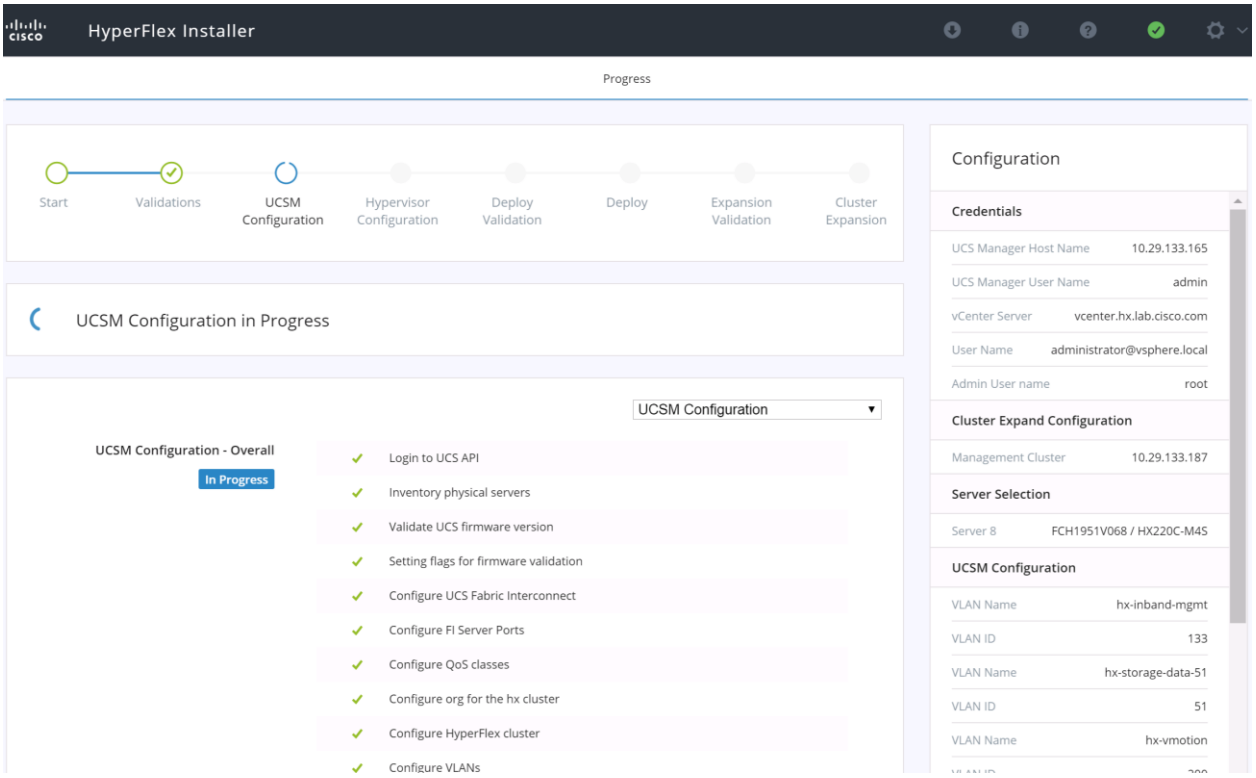
12. Enable Jumbo Frames, and if the server has been used previously, then select Clean up disk partitions.

13. (Optional) At this step you can manually add more servers for expansion if these servers already have service profiles associated and the hypervisor is ready, by clicking on Add Compute Server or Add Converged Server and then entering the IP addresses for the storage controller management and data networks.

14. Click Start.

15. Validation of the configuration will now start. If there are warnings, you can review and click “Skip Validation” if the warnings are acceptable (e.g. you might get the warning from Cisco UCS Manger validation that the guest VLAN is already assigned). If there are no warnings, the validation will automatically continue on to the configuration process.

16. The HX installer will now proceed to complete the deployment and perform all the steps listed at the top of the screen along with their status.



17. You can review the summary screen after the install completes by selecting Summary on the top right of the window.

After the install has completed, the new converged node is added to the cluster, and its storage, CPU, and RAM resources are immediately available, however the new node still requires some post installation steps in order to be consistent with the configuration of the existing nodes. For example, the new converged node will not have a vMotion vmkernel interface, and it may not have all of the guest VM networks configured. The easiest method to make the changes is to use the post_install script, or the configuration can be done manually.

Expansion with M5 Generation Servers for Mixed Clusters

Existing HyperFlex clusters can be expanded by adding new M5 generation servers to a cluster of M4 generation servers, creating a mixed cluster. The creation of mixed HyperFlex clusters is subject to a number of limitations and guidelines, as outlined below:

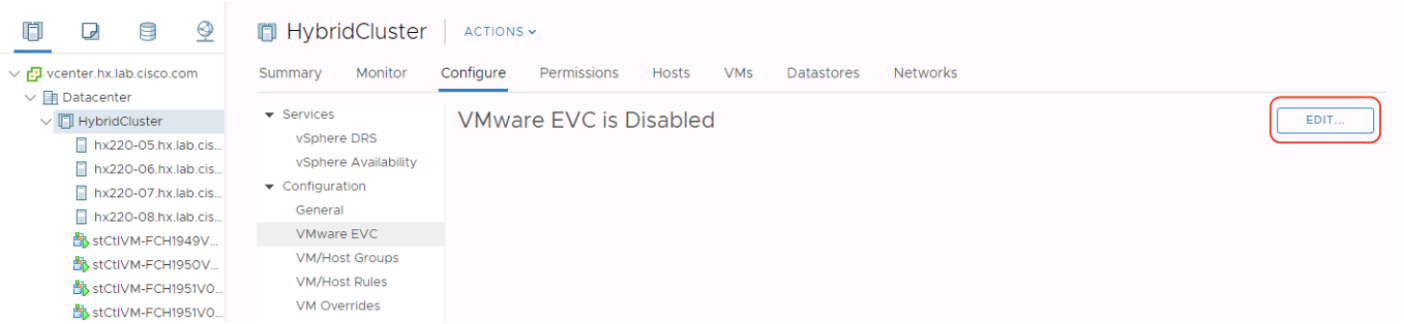
- A mixed cluster cannot be created by performing an initial cluster installation with a mixture of M4 and M5 generation converged nodes. The only supported method for creating a mixed cluster is to create an initial cluster of only M4 converged nodes, and then expanding that cluster with additional M5 converged nodes.
- Creating an initial cluster of only M5 generation converged nodes, and then expanding that cluster with M4 generation converged nodes is not supported.
- Once a cluster is operating as a mixed cluster with both M4 and M5 generation converged nodes, that cluster can later be expanded further with more M4 or M5 generation converged nodes.
- All of the new M5 generation converged nodes must match the form factor, type, encryption, and disk capacity and quantity of the existing M4 generation converged nodes. For example:

  - All nodes in the cluster must be HX220c or HX240c models. It is not allowed to mix HX220c and HX240c within the same cluster under any circumstance.
  - All nodes in the cluster must be hybrid nodes, or all-flash nodes. It is not allowed to mix hybrid nodes and all-flash nodes within the same cluster under any circumstance.
  - All nodes in the cluster must contain exclusively standard disks. Mixed clusters cannot be created using nodes with self-encrypting disks (SEDs).
  - Capacity disk size and quantities must match across all existing and additional nodes being added. As an example, the HX220c-M4S node can only contain 6 capacity disks, therefore the new HX220c-M5SX nodes can only contain 6 capacity disks, even though they can physically contain up to 8 disks.
  - The storage size of all capacity disks must match, but the caching disk sizes do not need to match.
- Existing HyperFlex clusters can be expanded with compute-only nodes of any model or generation in nearly any combination. For example, the following combinations are possible:
  - Expanding an existing mixed cluster with any model of supported compute-only nodes.
  - Expanding an existing M5 generation cluster with any model of supported compute-only nodes, including older M4 and M3 servers.
  - Expanding an existing M4 generation cluster with any model of supported compute-only nodes, including new M5 servers.
- While CPU models in the M4 and M5 generation servers are different, it is recommended to attempt to match CPU core counts and frequencies between the servers as much as possible.
- While RAM speeds and optimal DIMM layouts in the M4 and M5 generation servers are different, it is recommended to attempt to match RAM amounts between the servers.
- Cisco UCS Manager and Fabric Interconnect firmware must be upgraded to version 3.2(2d) or later at minimum to support the new M5 generation servers prior to beginning the expansion.
- The existing M4 generation HyperFlex cluster must be upgraded to HXDP 2.6 before it can be expanded with the additional M5 generation nodes.
- Existing M4 generation servers are recommended to have their firmware revisions upgraded to the included components in version 3.2(2d) or later if their current versions are older than 3.1(3c), after the upgrade to HXDP 2.6.
- The expansion tasks must be completed using the HXDP 2.6 Installer OVA.
- A HyperFlex cluster running an HXDP version older than 1.8 cannot coexist with any M5 generation cluster within the same Cisco UCS domain. This restriction applies to a mixed cluster with M5 and M4 generation nodes, or a non-mixed cluster of only M5 generation nodes.
- The version of VMware ESXi installed on the compute-only nodes and converged nodes of a cluster must all match, regardless of the model of server.
- In order to support the necessary VMware Enhanced vCPU Compatibility (EVC) settings for a mixed cluster, the managing vCenter server must be upgraded to version 6.5. The ESXi hypervisor can remain at version 6.0.

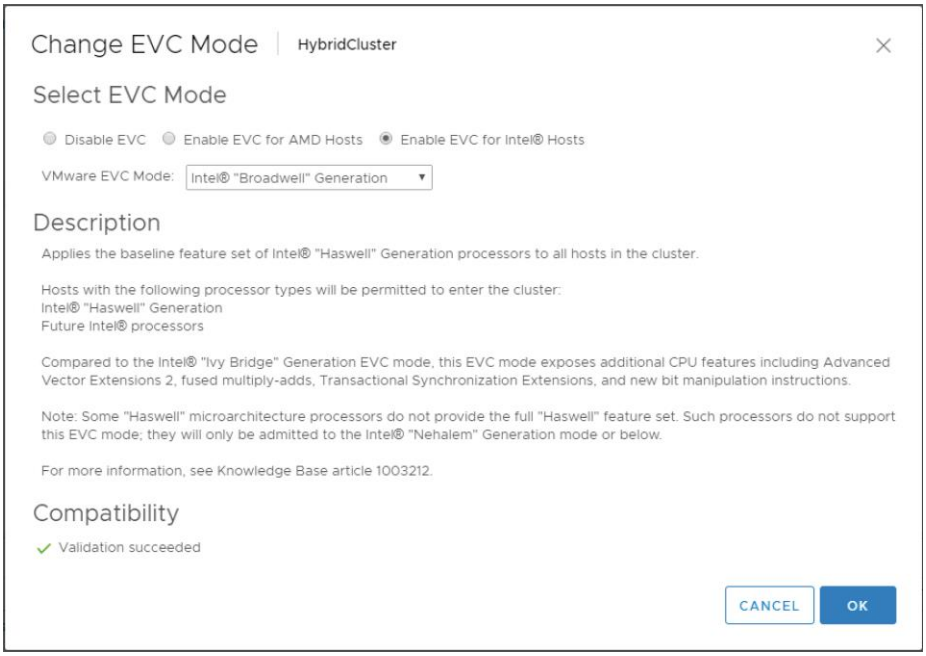
- EVC mode must be enabled in the existing M4 generation cluster, via the vCenter 6.5 server, prior to expanding the cluster to enable Intel Broadwell compatibility mode. Failure to do so will lead to failures of the VMs when attempting to vMotion the VMs between the different generations of servers. The HyperFlex installer will detect this condition and alert you to the requirement.
- EVC mode must also be enabled if an existing M5 generation cluster is being expanded with M4 or M3 generation compute-only nodes. This manual procedure will require a shutdown of the existing cluster to complete.

To expand an existing HyperFlex cluster with M4 generation converged nodes with additional M5 generation converged nodes, follow the steps in the previous section, choosing the new M5 generation servers as the nodes to add to the existing cluster. If EVC mode must be enabled in the cluster, the HyperFlex installer will alert you to the requirement. To enable EVC mode prior to the expansion, creating a mixed cluster, complete the following steps:

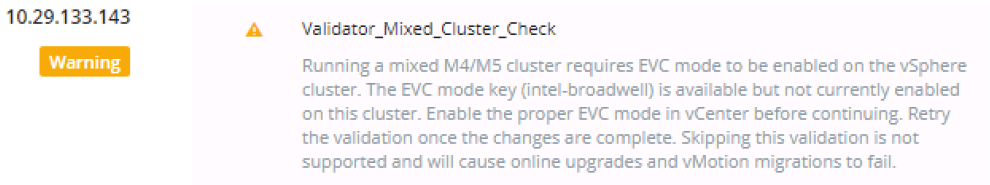
1. Log in to the vCenter 6.5 HTML5 vSphere Client.
2. From the Hosts and Clusters view, click on the cluster which is the existing M4 generation based HyperFlex cluster that needs to have EVC mode enabled.
3. On the right-hand side of the screen, click Configure.
4. Under the Configuration menu, click VMware EVC, then click the Edit button on the right.



5. Click the radio button labeled “Enable EVC for Intel Hosts”
6. Click on the drop-down menu to choose the generation of CPU to use as the baseline for EVC mode. For example, if the M4 generation servers are equipped with Intel E5 v3 processors, you would choose Intel “Haswell” Generation, but if the CPUs are Intel E5 v4 models, you would choose Intel “Broadwell” Generation.
7. Ensure that the Compatibility check at the bottom shows Validation Succeeded, then click OK.



If the cluster expansion is attempted before enabling EVC mode, the HyperFlex installer will generate a validation error during the deployment validation phase, as seen below. When this happens, do not cancel or restart the cluster expansion, instead enable EVC mode in vCenter 6.5. Once EVC mode has been enabled using the previous steps, return to the HyperFlex installer and click on the Retry Deploy Validation button. The cluster expansion should continue without any further errors.



Management

HyperFlex Connect

HyperFlex Connect is the new, easy to use, and powerful primary management tool for HyperFlex clusters. HyperFlex Connect is an HTML5 web-based GUI tool which runs on all of the HX nodes, and is accessible via the cluster management IP address.

Local Access

Logging into HyperFlex Connect can be done using pre-defined local accounts. In order to log in with a local account prepend “local/” to the account name, for example, local/root. The password for the default root account is set during the cluster creation as the cluster password. Using local access is only recommended when vCenter direct or SSO credentials are not available.

Role-Based Access Control

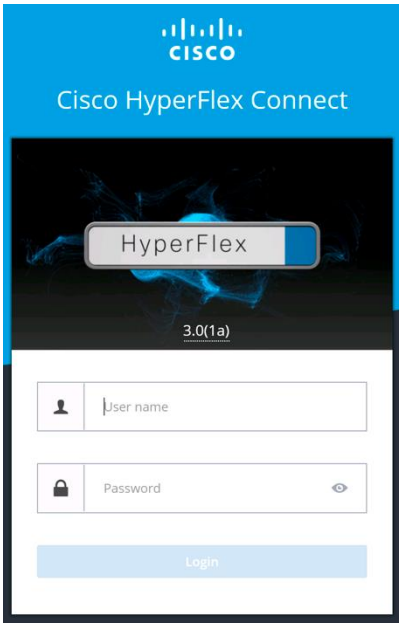
HyperFlex Connect provides Role-Based Access Control (RBAC) via integrated authentication with the vCenter Server managing the HyperFlex cluster. You can have two levels of rights and permissions within the HyperFlex cluster:

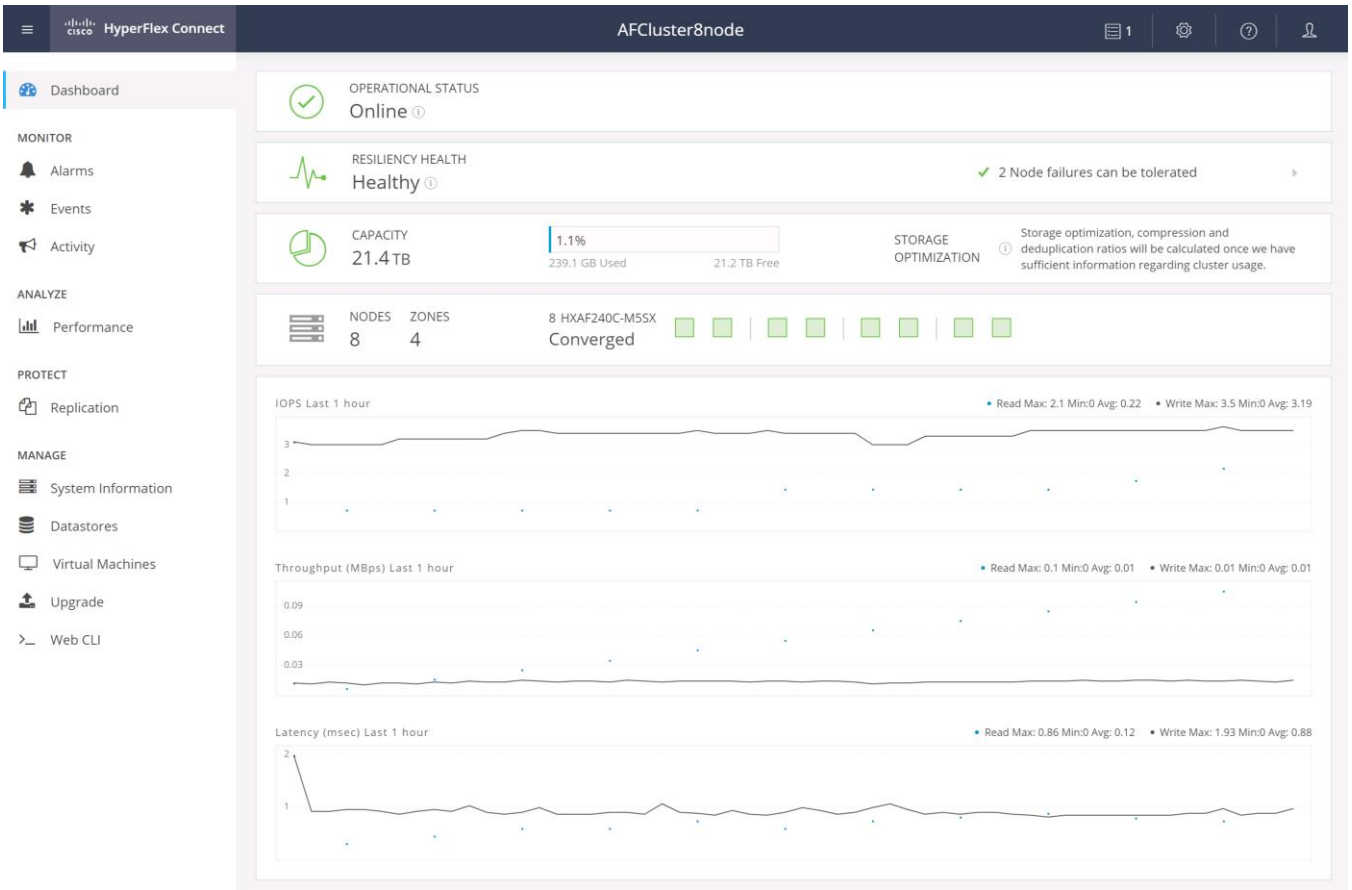
- Administrator: Users with administrator rights in the managing vCenter server will have read and modify rights within HyperFlex Connect. These users can make changes to the cluster settings and configuration.
- Read-Only: Users with read-only rights in the managing vCenter server will have read rights within HyperFlex Connect. These users cannot make changes to the cluster settings and configuration.

Users can log in to HyperFlex Connect using direct vCenter credentials, for example, administrator@vsphere.local, or using vCenter Single Sign-On (SSO) credentials, such as an Active Directory user, for example, domain\user. Creation and management of RBAC users and rights must be done via the vCenter Web Client or vCenter 6.5 HTML5 vSphere Client.

To manage the HyperFlex cluster using HyperFlex Connect, complete the following steps:

1. Using a web browser, open the HyperFlex cluster’s management IP address via HTTPS, for example, https://10.29.133.160.
2. Enter a local credential, such as local/root, or a vCenter RBAC credential for the username, and the corresponding password.
3. Click Login.
4. The Dashboard view will be shown after a successful login.





Dashboard

From the Dashboard view, several elements are presented:

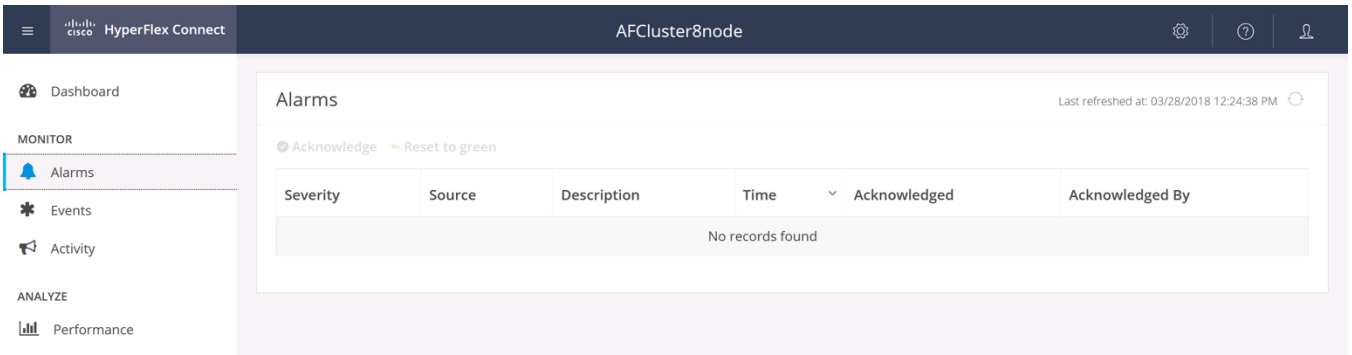
- Cluster operational status, overall cluster health, and the cluster's current node failure tolerance.

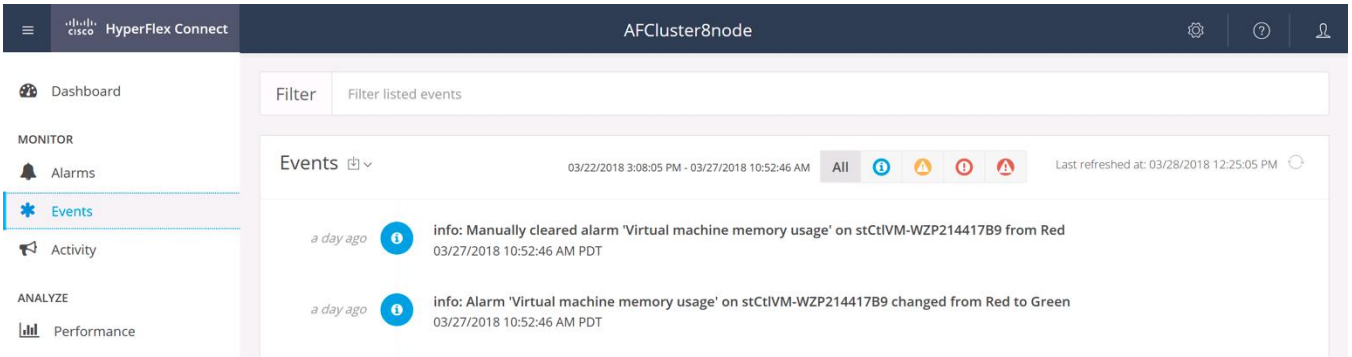
- Cluster storage capacity, used and free space, compression and deduplication savings, and overall cluster storage optimization statistics.
- Cluster size and individual node health.
- Cluster IOPs, storage throughput, and latency for the past 1 hour.

Monitor

HyperFlex Connect provides for additional monitoring capabilities, including:

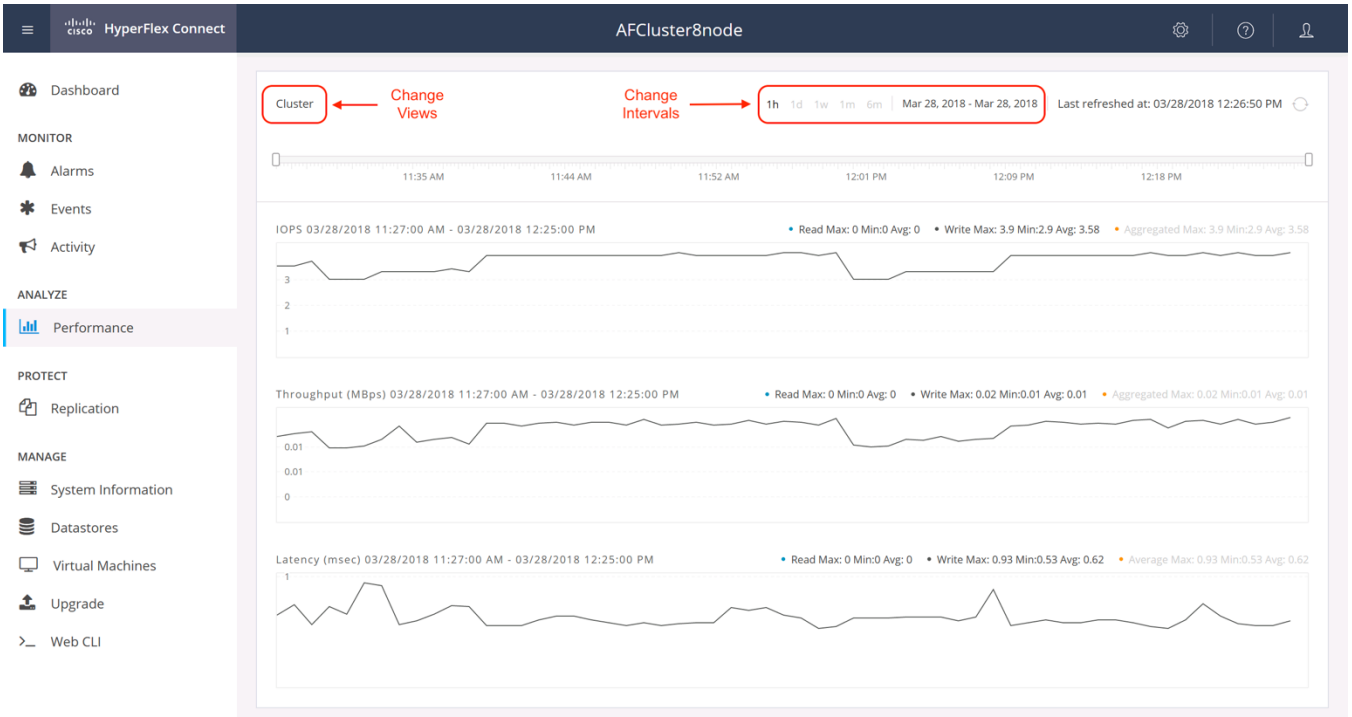
- Alarms: Cluster alarms can be viewed, acknowledged and reset.
- Event Log: The cluster event log can be viewed, specific events can be filtered for, and the log can be exported.
- Activity Log: Recent job activity, such as ReadyClones can be viewed and the status can be monitored.





Analyze

The historical and current performance of the HyperFlex cluster can be analyzed via the built-in performance charts. The default view shows read and write IOPs, bandwidth, and latency over the past 1 hour for the entire cluster. Views can be customized to see individual nodes or datastores, and change the timeframe shown in the charts.



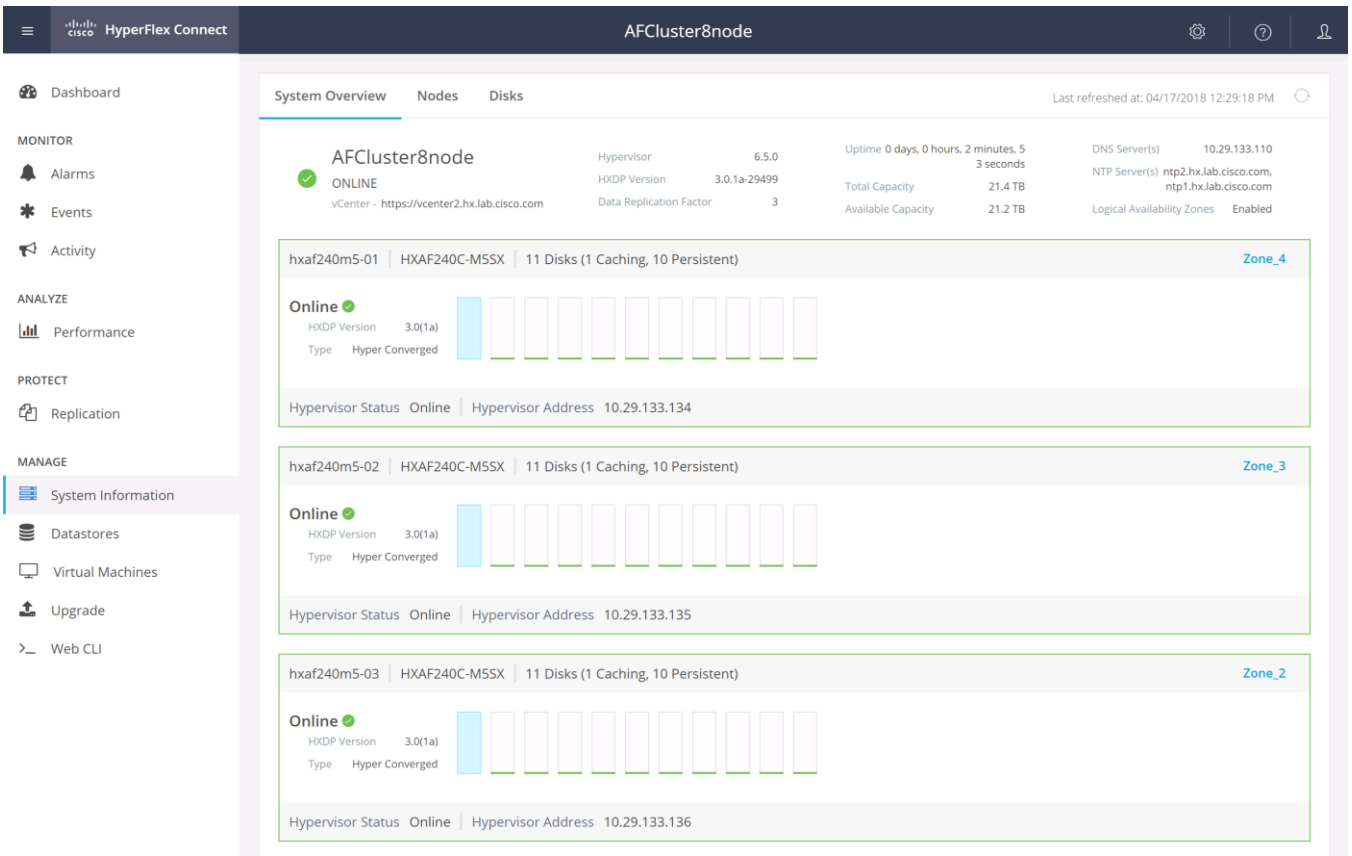
Protect

HyperFlex Connect is used as the management tool for all configuration of HyperFlex Data Protection features, including VM replication and data-at-rest encryption. Configuration of these features is covered in later sections of this document.

Manage

HyperFlex Connect presents several views and elements for managing the HyperFlex cluster:

- System Information: Presents a detailed view of the cluster configuration, software revisions, hosts, disks, and cluster uptime. Support bundles can be generated to be shared with Cisco TAC when technical support is needed. Views of the individual nodes and the individual disks are available. In these views, nodes can be placed into HX Maintenance Mode, and disks can be securely erased, as described later in this document.
- Datastores: Presents the datastores present in the cluster, and allows for datastores to be created, mounted, unmounted, edited or deleted, as described earlier in this document as part of the cluster setup.
- Virtual Machines: Presents the VMs present in the cluster, and allows for the VMs to be powered on or off, cloned via HX ReadyClone, Snapshots taken, and protected via native replication, as described later in this document.
- Upgrade: Upgrades to the HXDP software, and Cisco UCS firmware can be initiated from this view.
- Web CLI: A web-based interface, from which CLI commands can be issued and their output seen, as opposed to directly logging into the SCVMs via SSH.



Cisco Intersight Cloud-Based Management

Cisco Intersight management is enabled via embedded code running on the Cisco UCS Fabric Interconnects, and in the Cisco HyperFlex software, known as device connectors. To enable Intersight management, the device connectors are registered online at the Cisco Intersight website, https://intersight.com when logged into the website with a valid cisco.com account used to manage your environments. Cisco Intersight can be used to manage and monitor HyperFlex clusters and UCS domains with the following software revisions:

- Cisco UCS Manager and Infrastructure Firmware version 3.2 and later
- Cisco HyperFlex software version 2.5(1a) or later

The Cisco UCS Fabric Interconnects, and the Cisco HyperFlex nodes must have DNS lookup capabilities and access to the internet. If direct access to the internet is not available, the systems can be configured to connect via an HTTPS proxy server.

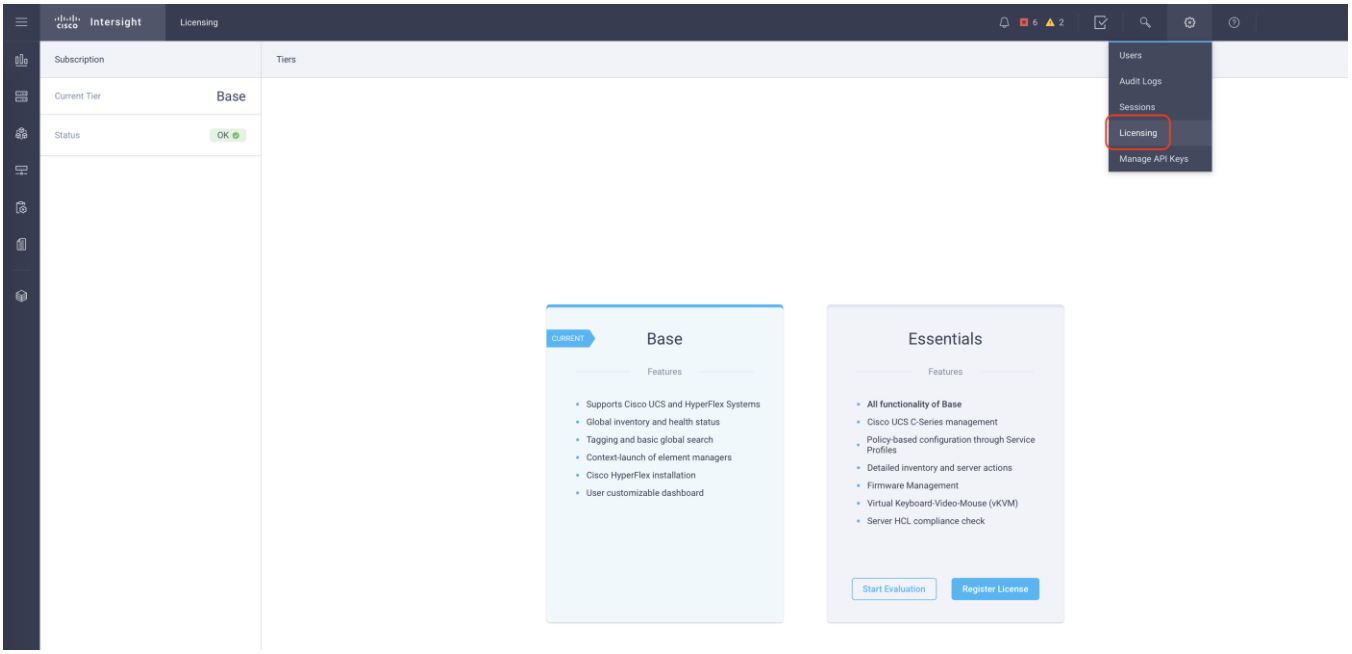
Cisco Intersight Licensing

Cisco Intersight is offered in two editions; a Base license which is free to use, and offers a large variety of monitoring, inventory and reporting features, and an added cost Essentials license, which adds advanced monitoring, server policy and profile configuration, firmware management, virtual KVM features, and more. New features and capabilities will be added to the different licensing tiers over time. A 90-day trial of the Essentials license is available for use as an evaluation period.

To configure Cisco Intersight licensing, complete the following steps:

1. Using a web browser, log on to the Cisco Intersight webpage at https://intersight.com (you must have a valid cisco.com CCO account).

2. In the Dashboards view, click the gear shaped icon in the upper right-hand corner, then click on Licensing.



3. Click Start Evaluation to begin a 90-day Essentials license trial, or click Register License.

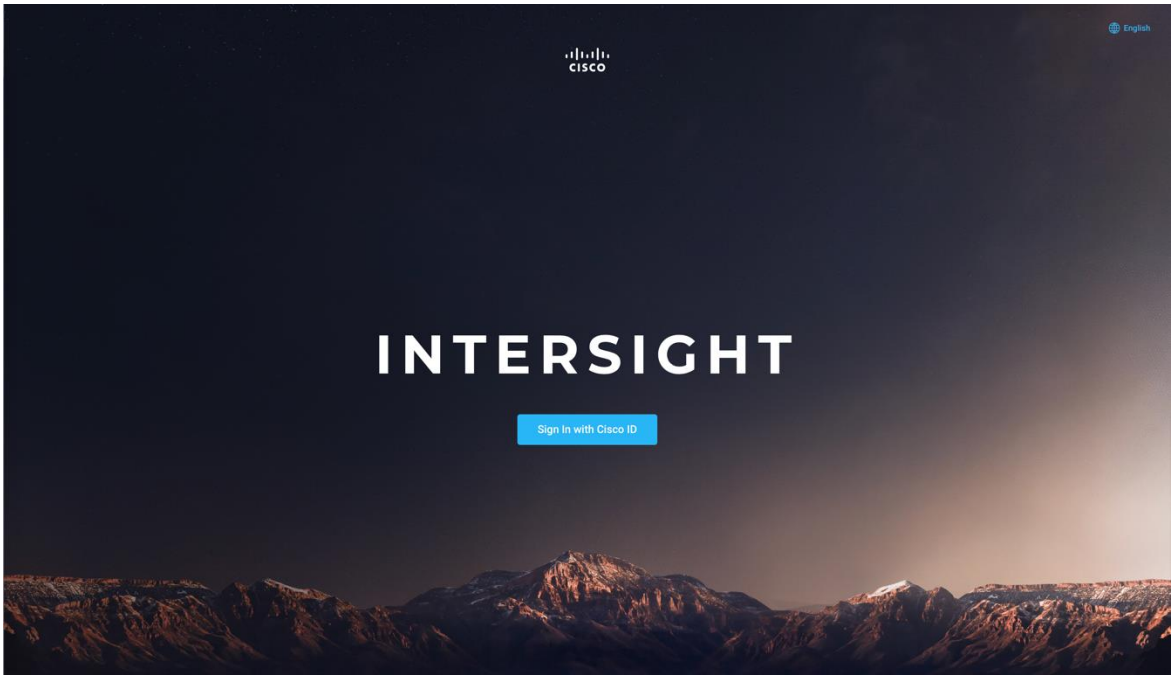
Cisco Intersight HyperFlex Management

To connect Cisco Intersight to the Cisco HyperFlex cluster(s), and the Cisco UCS Domain(s) in your environments, complete the following steps:

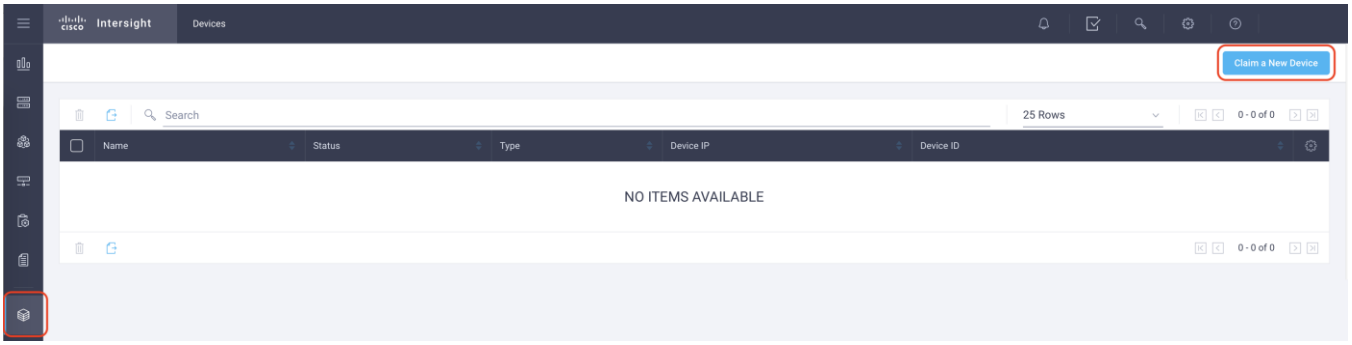
Connecting Cisco UCS Manager

1. Using a web browser, log on to the Cisco UCS Manager webpage.

2. From a second browser window or tab, log on to the Cisco Intersight webpage at https://intersight.com (you must have a valid cisco.com CCO account).



3. In the left-hand navigation buttons, click Devices.

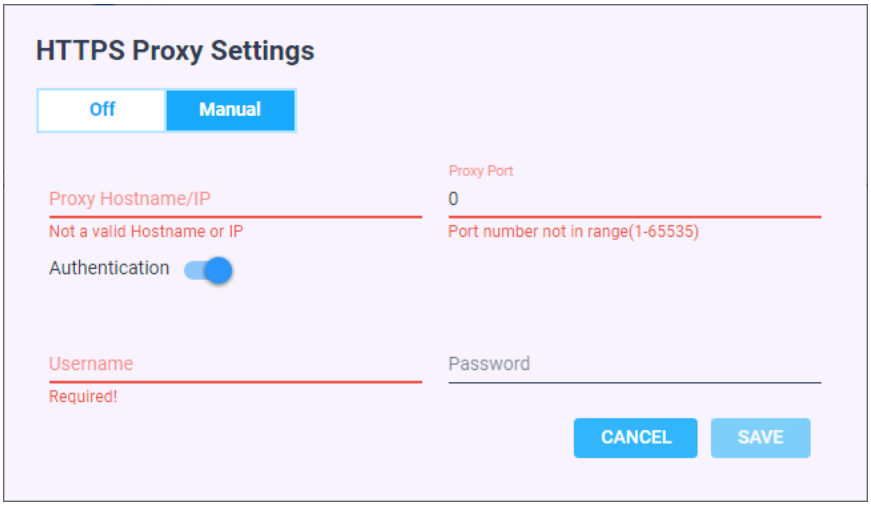


4. Click the “Claim A New Device” button in the top right-hand corner.

5. In Cisco UCS Manager, in the left-hand navigation buttons, click Admin.

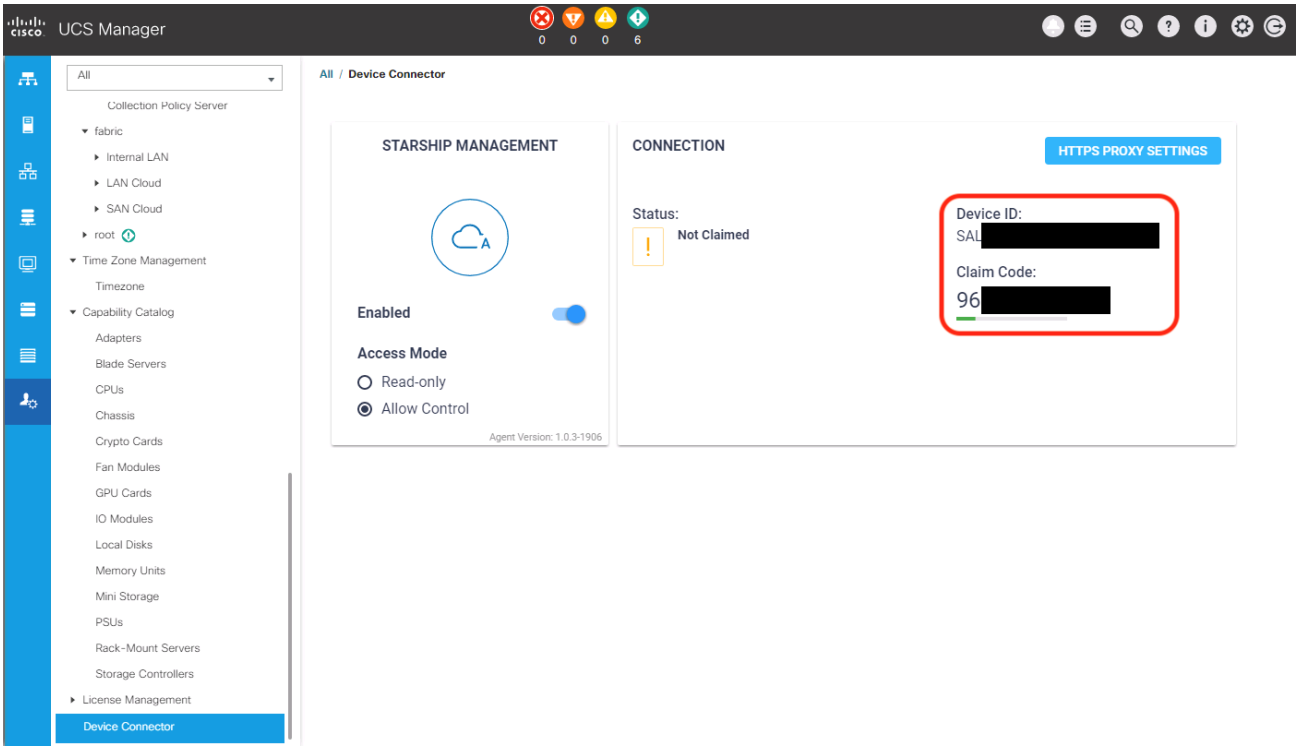
6. In the Admin tree, click Device Connector at the bottom.

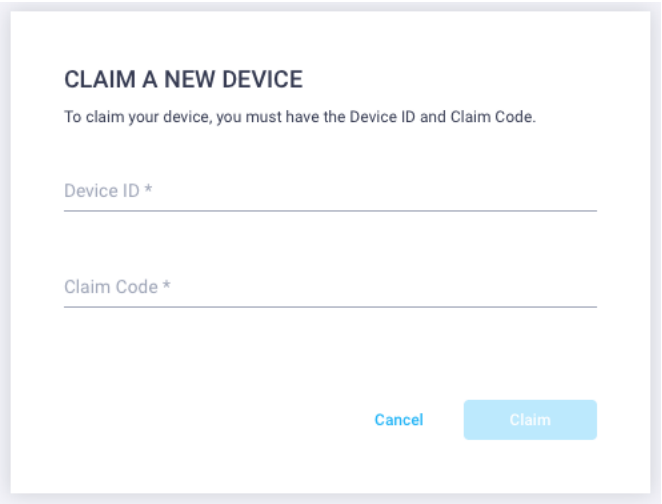
7. If necessary, click the HTTPS Proxy Settings button, and click the Manual button. Enter the Proxy server IP address or DNS hostname, the TCP port, enable authentication then enter a username and password if necessary, then click Save.



8. If desired, the Access Mode for Cisco Intersight can be set to Read-only, or management can be disabled from this screen.

9. In the main UCS Manager screen, you will see a Device ID and a Claim Code for this Cisco UCS Domain. Copy these two codes to the Device ID and Claim Code fields in the Cisco Intersight "Claim A New Device" window, then click Claim.



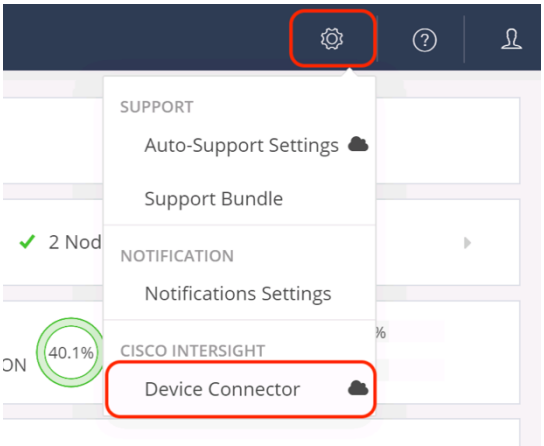


10. The Cisco UCS Domain will now show the system as Claimed in the Device Connector screen.

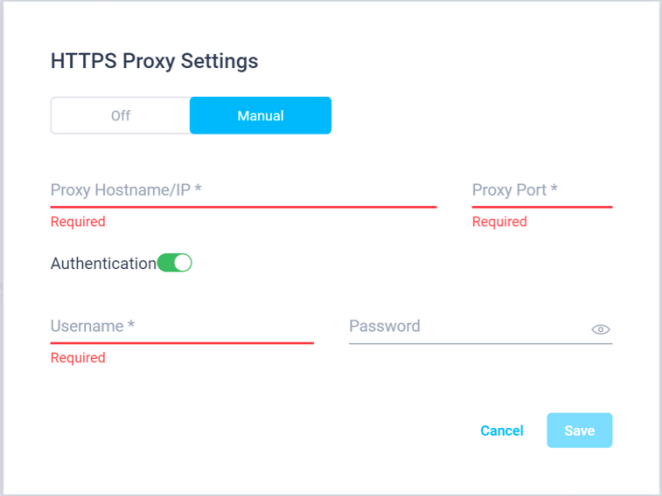
Connecting Cisco HyperFlex Clusters

To connect Cisco HyperFlex Clusters, complete the followings steps:

1. Use a web browser to open the HX Connect webpage at the cluster's management IP address, for example: https://10.29.133.151
2. Enter a local credential or a vCenter RBAC credential for the username and the corresponding password.
3. Click Login.
4. From a second browser window or tab, log on to the Cisco Intersight webpage at https://intersight.com (you must have a valid cisco.com CCO account).
5. In the left-hand navigation buttons, click on Devices.
6. Click the "Claim A New Device" button in the top right-hand corner.
7. In the HyperFlex Connect Dashboard page, click Edit Settings in the top right-hand corner, then click Device Connector.

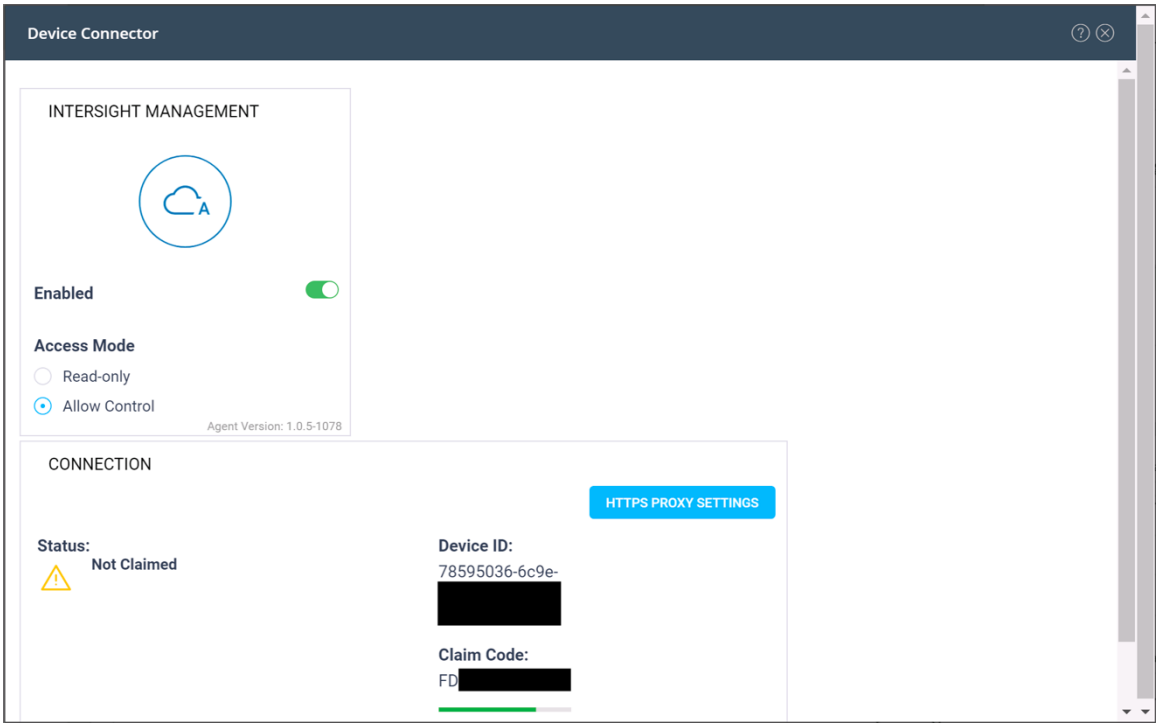


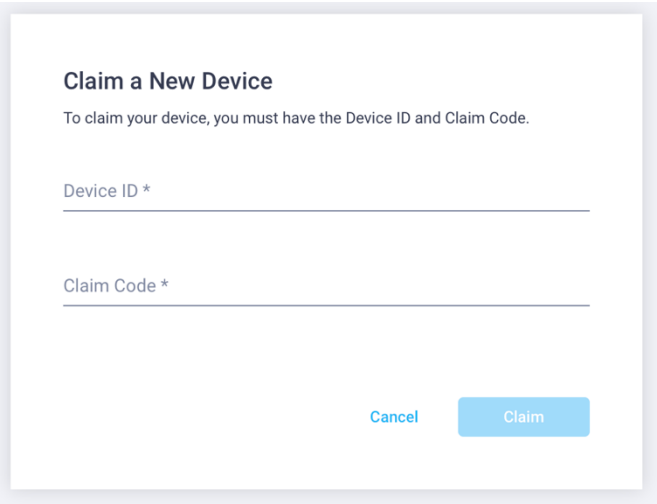
8. If necessary, click the HTTPS Proxy Settings button, and click the Manual button. Enter the Proxy server IP address or DNS hostname, the TCP port, enable authentication then enter a username and password if necessary, then click Save.



9. If desired, the Access Mode for Cisco Intersight can be set to Read-only, or management can be disabled from this screen.

10. In the HyperFlex Connect screen, you will see a Device ID and a Claim Code for this HyperFlex cluster. Copy these two codes to the Device ID and Claim Code fields in the Cisco Intersight "Claim A New Device" window, then click Claim.

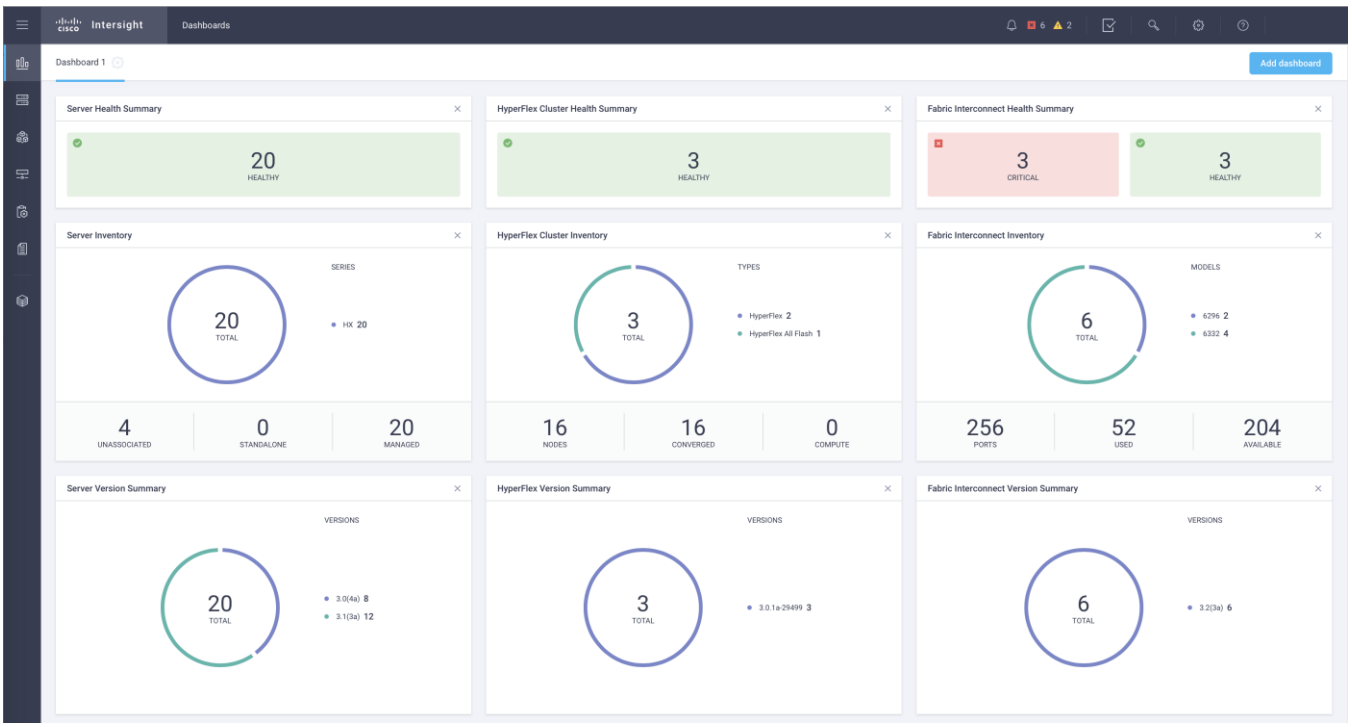




11. The Cisco HyperFlex Cluster will now show the system as Claimed in the Device Connector screen.

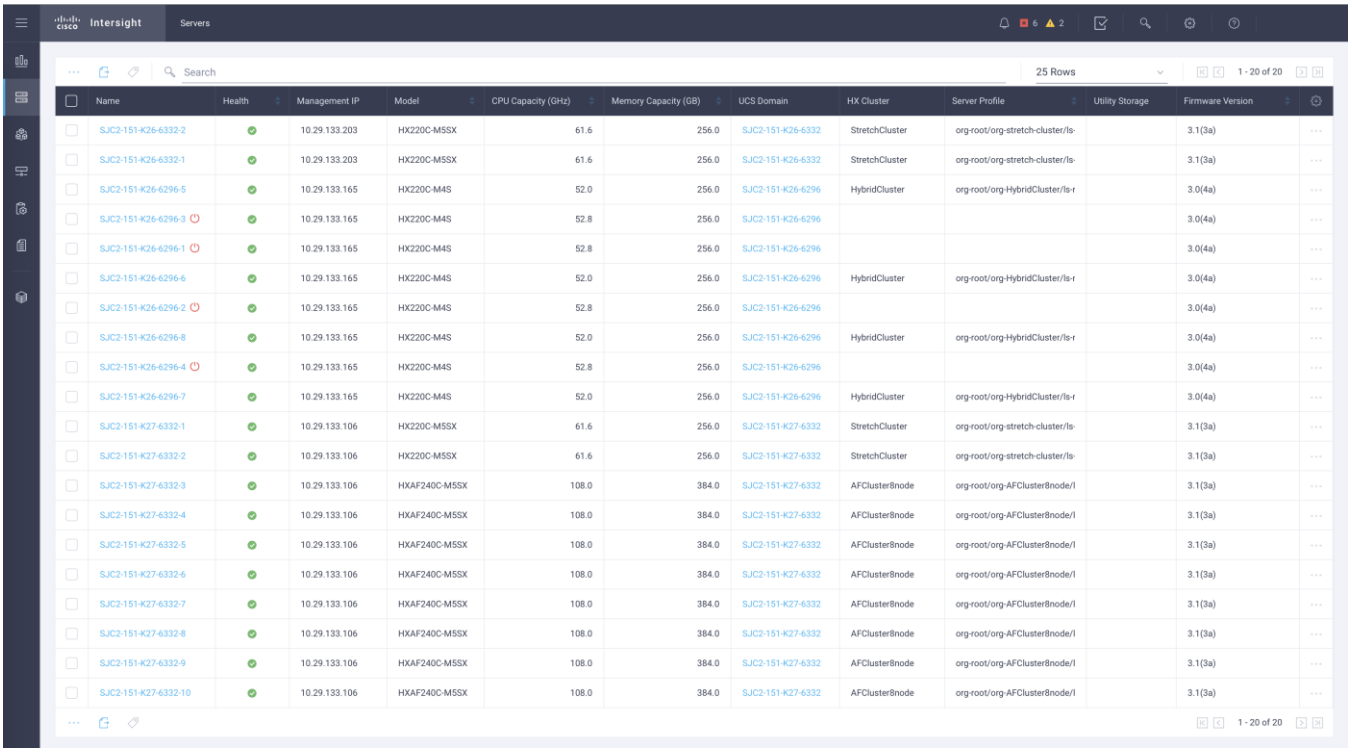
Dashboard

The Cisco Intersight Dashboard provides a single screen overview of all connected Cisco UCS Domains, the servers within those domains, the HyperFlex Clusters running in the domains, along with their health statuses, storage utilization, port counts, and more. Elements on the screen are clickable and will drill down into other sections of the page to view further details.



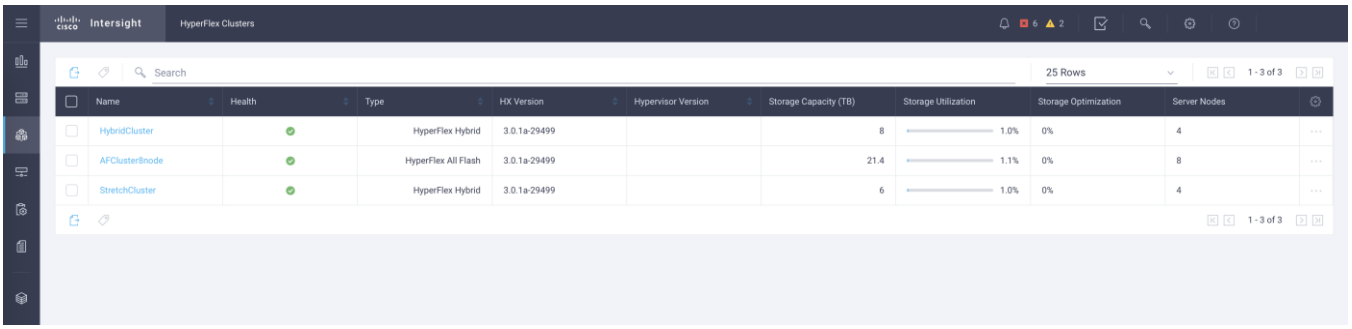
Servers

The Servers screen provides details of all the individual servers within the connected and managed UCS domains.



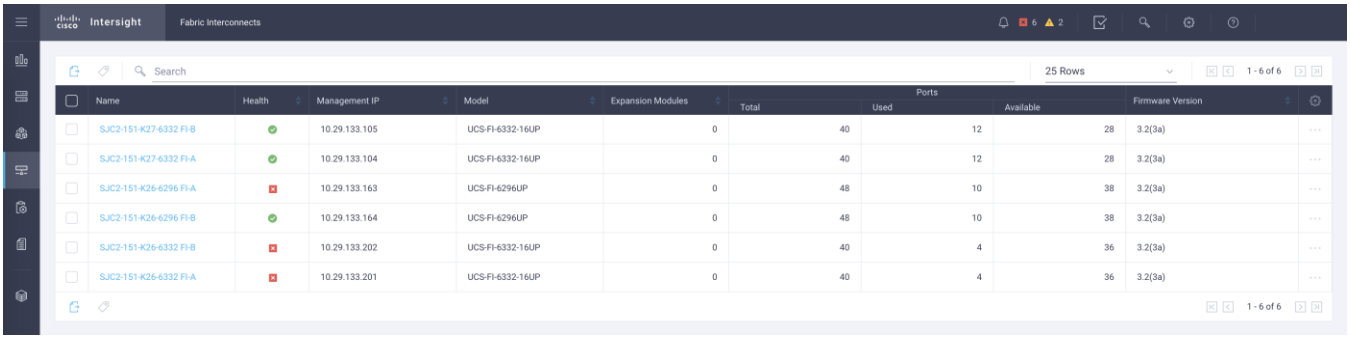
HyperFlex Clusters

The HyperFlex Clusters screen provides details of all the HyperFlex clusters that are connected and managed by Cisco Intersight. By clicking the ellipses (...) the HyperFlex Connect GUI for the clusters can be directly connected to in another browser window or tab.



Fabric Interconnects

The Fabric Interconnects screen provides details of all the UCS domains that are connected and managed by Cisco Intersight. By clicking the ellipses (...) the Cisco UCS Manager webpage for the domain can be directly connected to in another browser window or tab, or a session can be opened to the CLI of the Fabric Interconnect.



Profiles and Policies

Cisco Intersight Service Profiles and Policies pages are only available with Intersight Essentials licensing, except for configuring a Cisco HyperFlex Cluster Profile. At the time of the publishing of this document, using Cisco Intersight to configure a HyperFlex Profile and perform the HyperFlex installation is only supported for HyperFlex Edge systems. As such, the details of the use and configuration for HyperFlex Edge is covered in a separate document. Deployment of standard and stretched HyperFlex 3.0 clusters using Cisco Intersight is a forthcoming feature and will be covered in a future revision to this document.

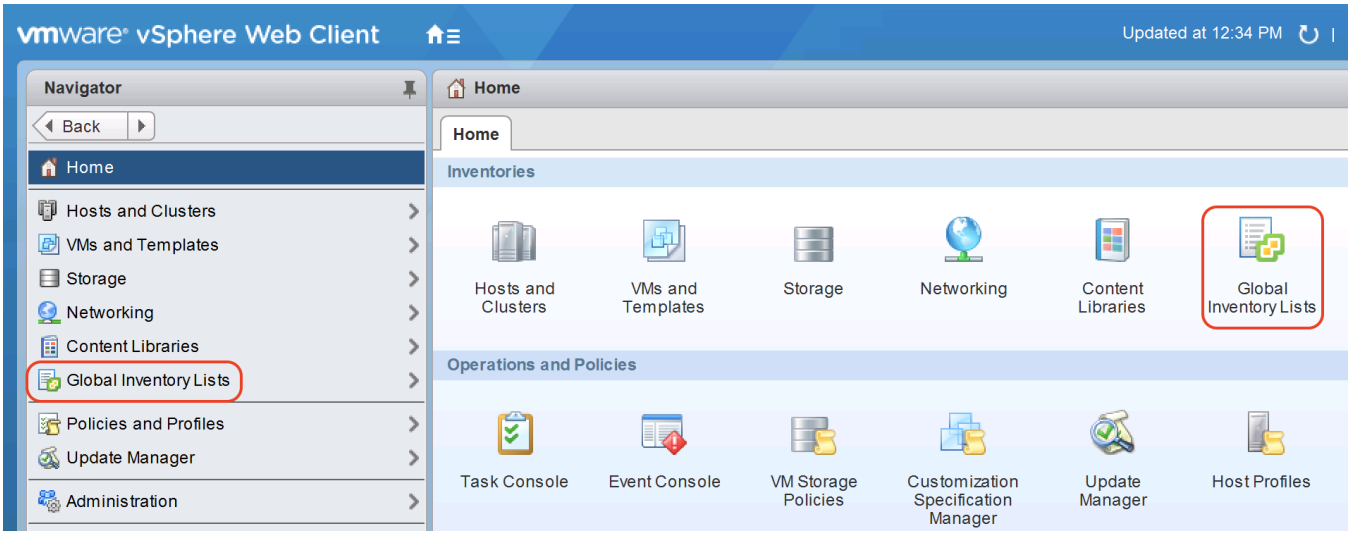
vCenter Web Client Plugin

The Cisco HyperFlex vCenter Web Client Plugin is installed by the HyperFlex installer to the specified vCenter server or vCenter appliance. The plugin is accessed as part of the vCenter Web Client (Flash) interface, and is a secondary tool used to monitor and configure the HyperFlex cluster. This plugin is not integrated into the new vCenter 6.5 HTML5 vSphere Client. In order to manage a HyperFlex cluster via an HTML5 interface, i.e. without the Adobe Flash requirement, use the new HyperFlex Connect management tool.

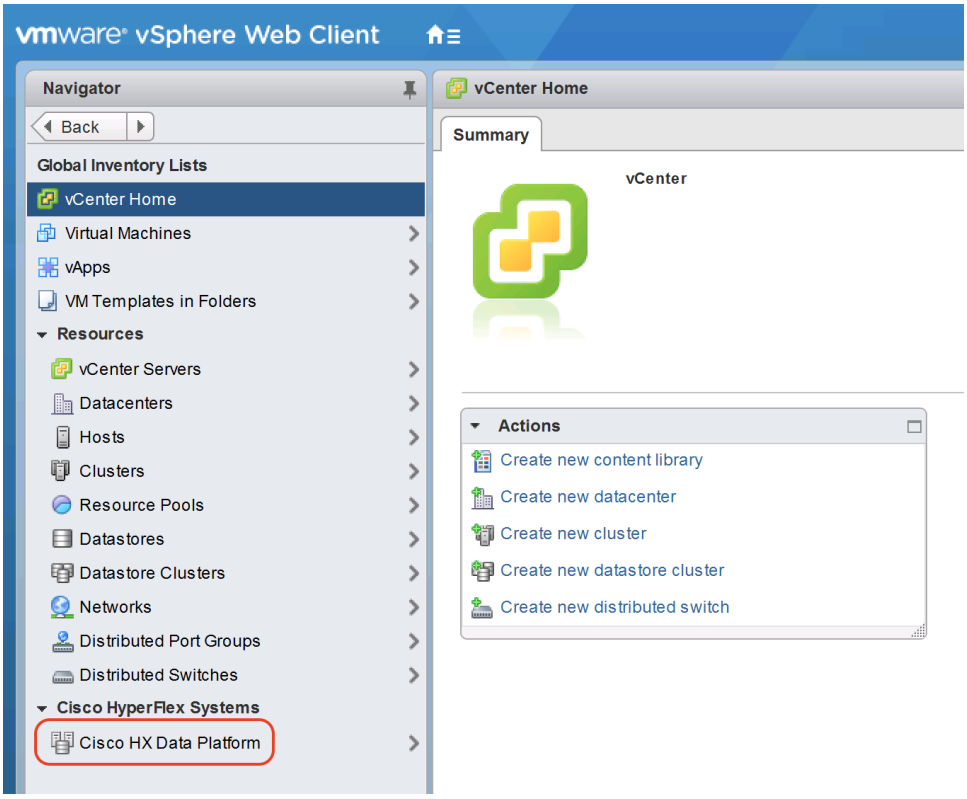
To manage the HyperFlex cluster using the vCenter Web Client Plugin for vCenter 6.5, complete the following steps:

1. Open the vCenter Web Client, and login with admin rights.

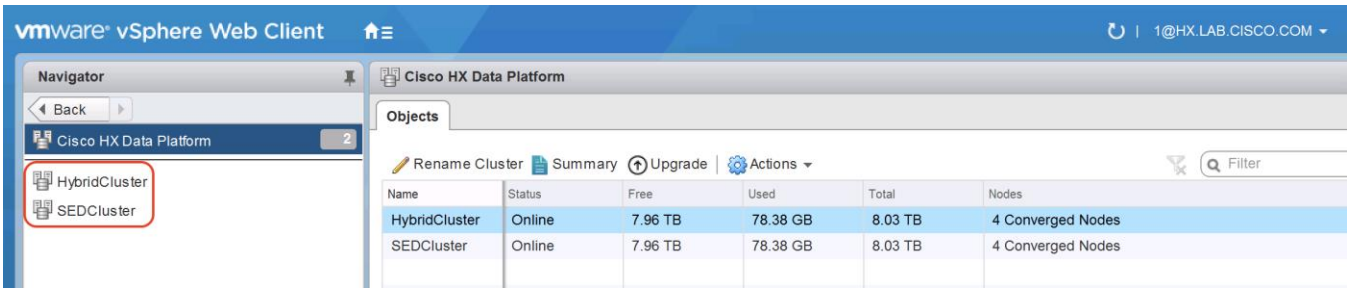
2. In the home pane, from the home screen click Global Inventory Lists.



3. In the Navigator pane, click Cisco HX Data Platform.



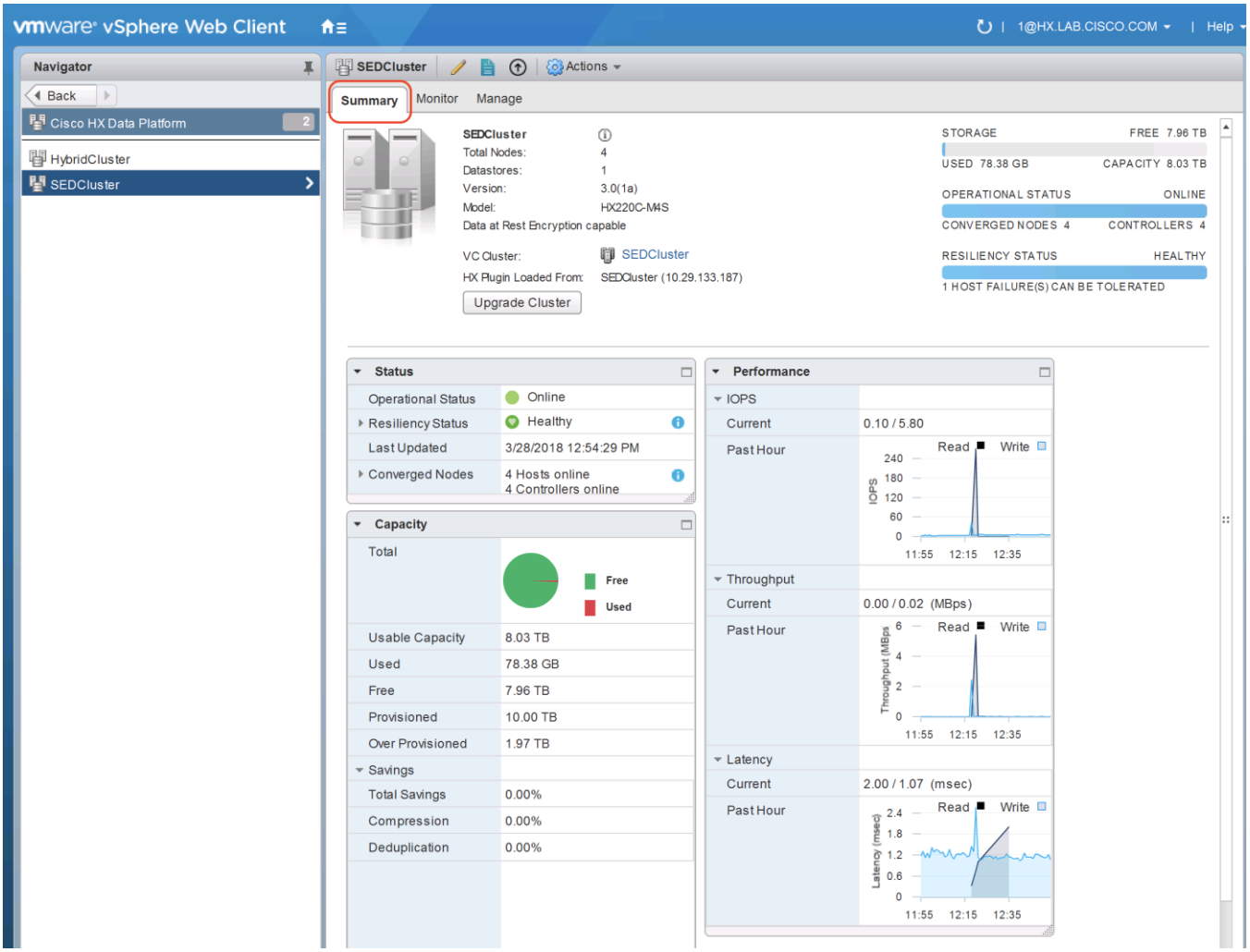
4. In the Navigator pane, choose the HyperFlex cluster you want to manage and click the name.



Summary

From the Web Client Plugin Summary screen, several elements are presented:

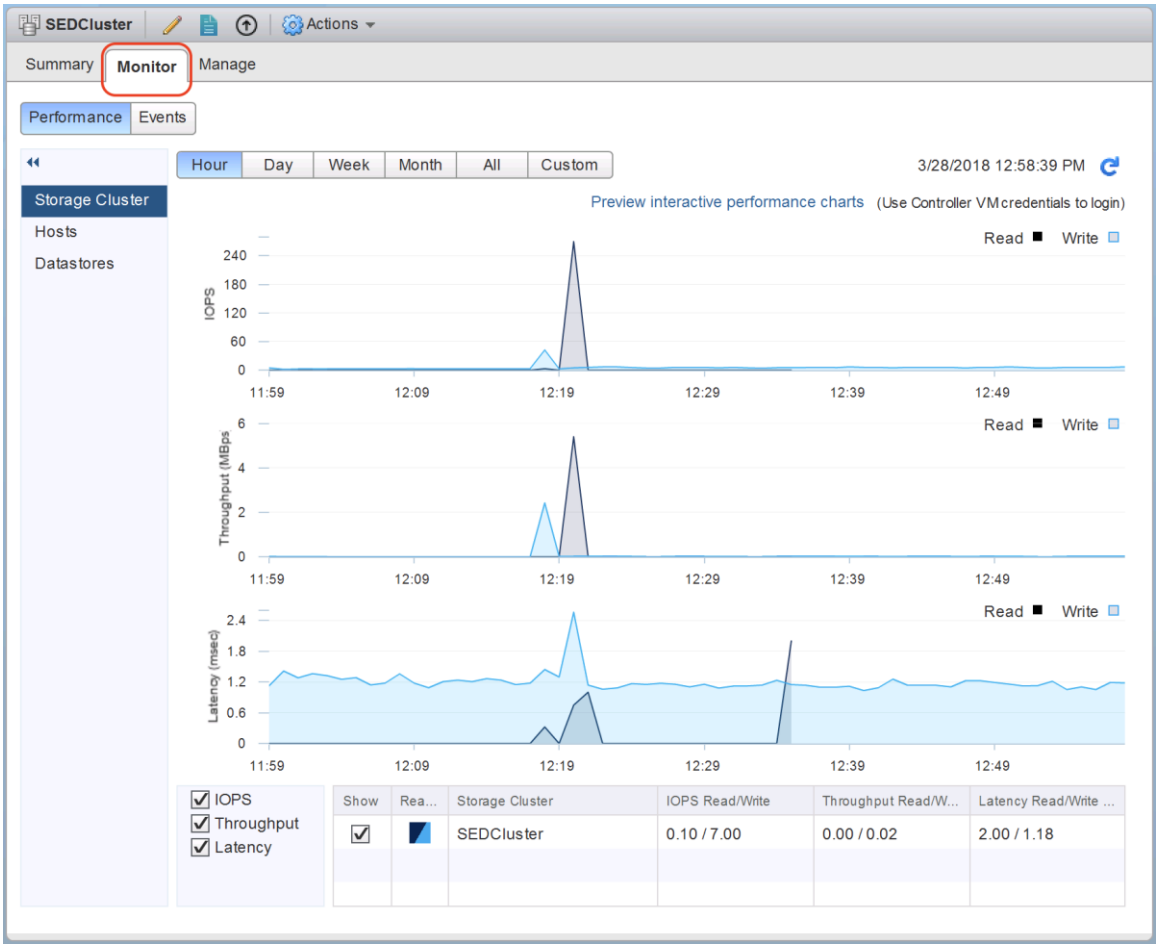
- Overall cluster usable capacity, used capacity, free capacity, datastore capacity provisioned, and the amount of datastore capacity provisioned beyond the actual cluster capacity.
- Deduplication and compression savings percentages calculated against the data stored in the cluster.
- The cluster operational status, the health state, and the number of node failures that can occur before the cluster goes into read-only or offline mode.
- A snapshot of performance over the previous hour, showing IOPS, throughput, and latencies.



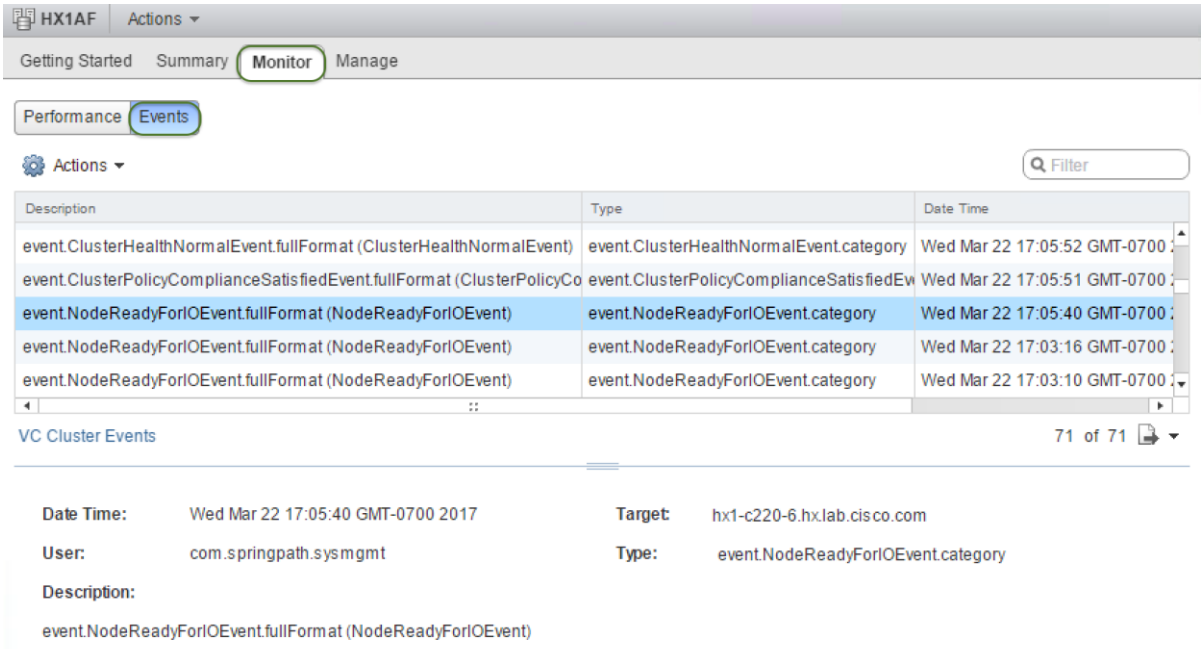
Monitor

From the Web Client Plugin Monitor tab, several elements are presented:

- Clicking the Performance button displays a larger view of the performance charts. If a full webpage screen view is desired, click the Preview Interactive Performance charts hyperlink. Enter the username (root) and the password for the HX controller VM to continue.



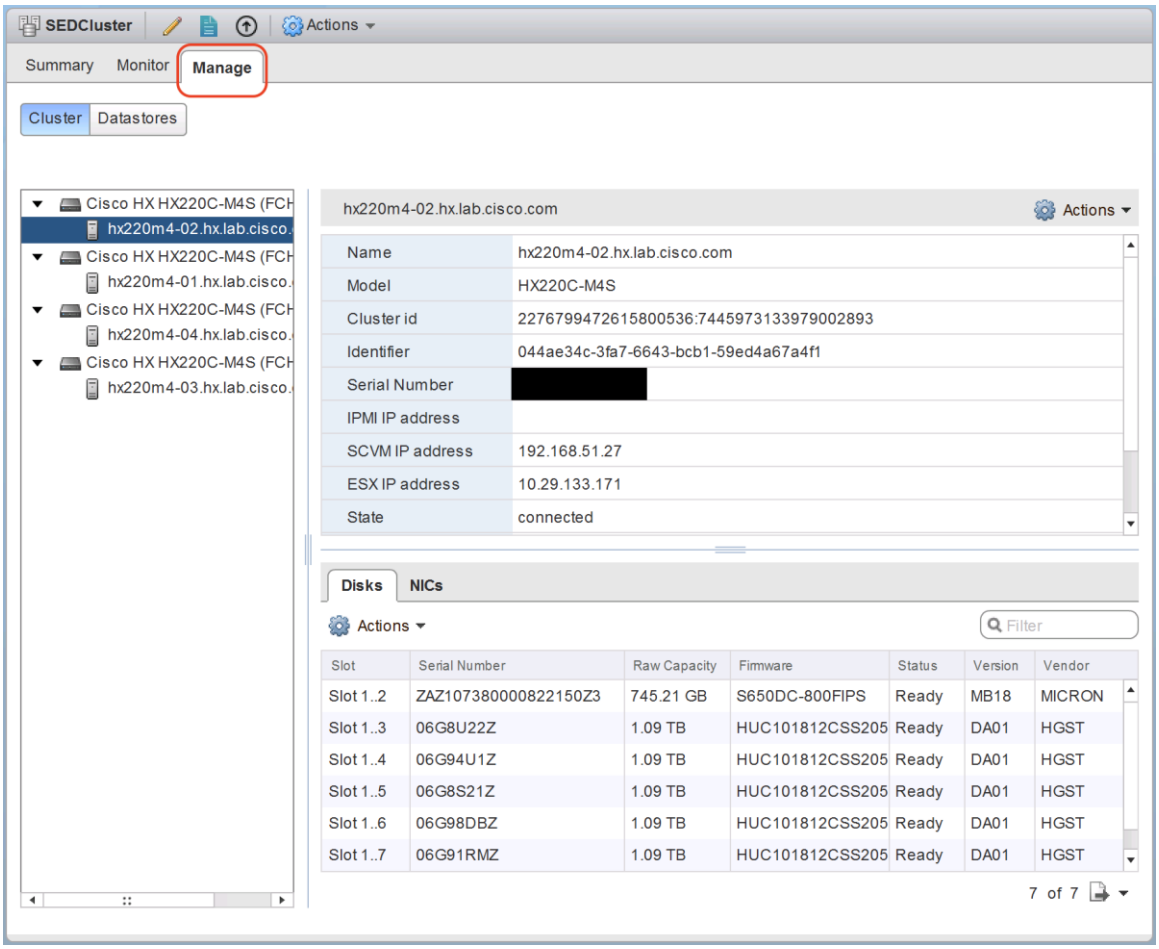
- Clicking the Events button displays a HyperFlex event log, which can be used to diagnose errors and view system activity events.



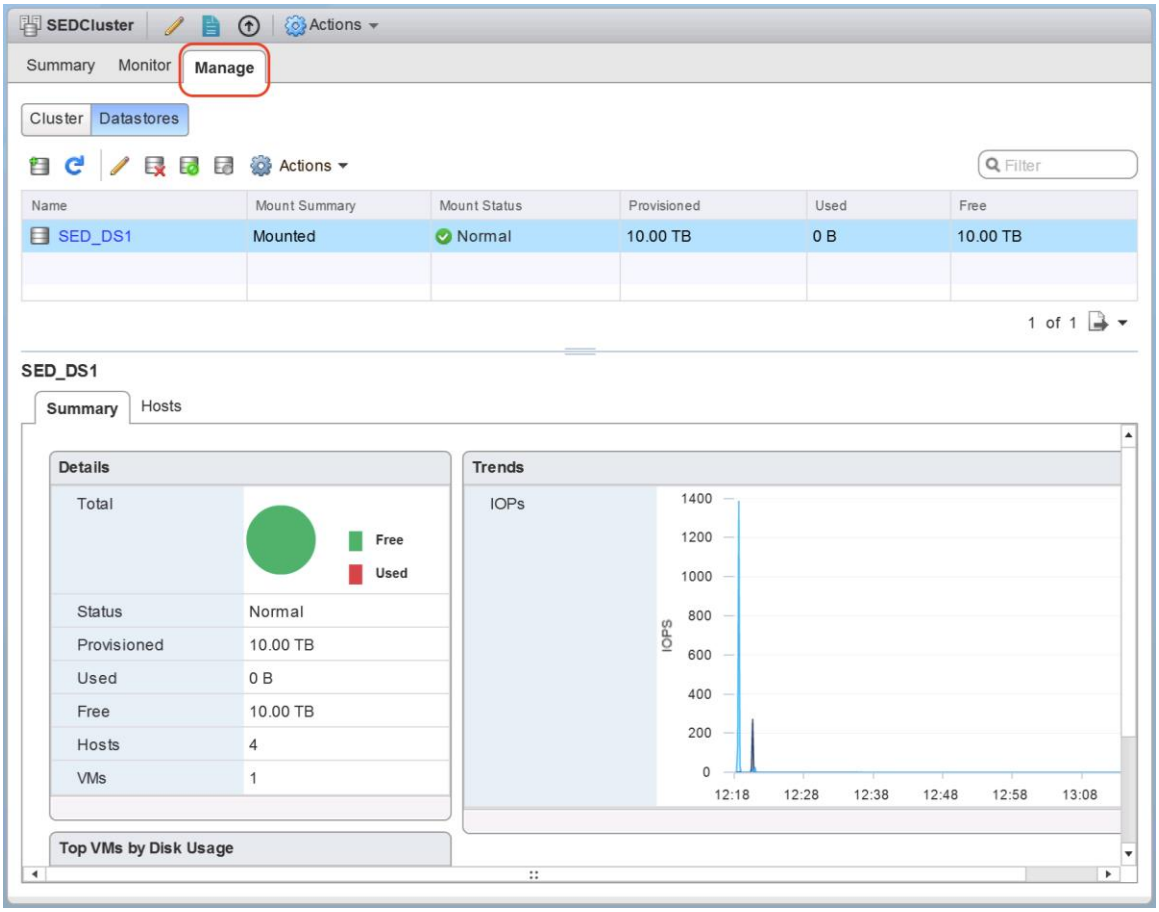
Manage

From the Web Client Plugin Manage tab, several elements are presented:

- Clicking the Cluster button displays an inventory of the HyperFlex cluster and the physical assets of the cluster hardware.



- Clicking the Datastores button allows datastores to be created, edited, deleted, mounted and unmounted, along with space summaries and performance snapshots of that datastore.



Management Best Practices

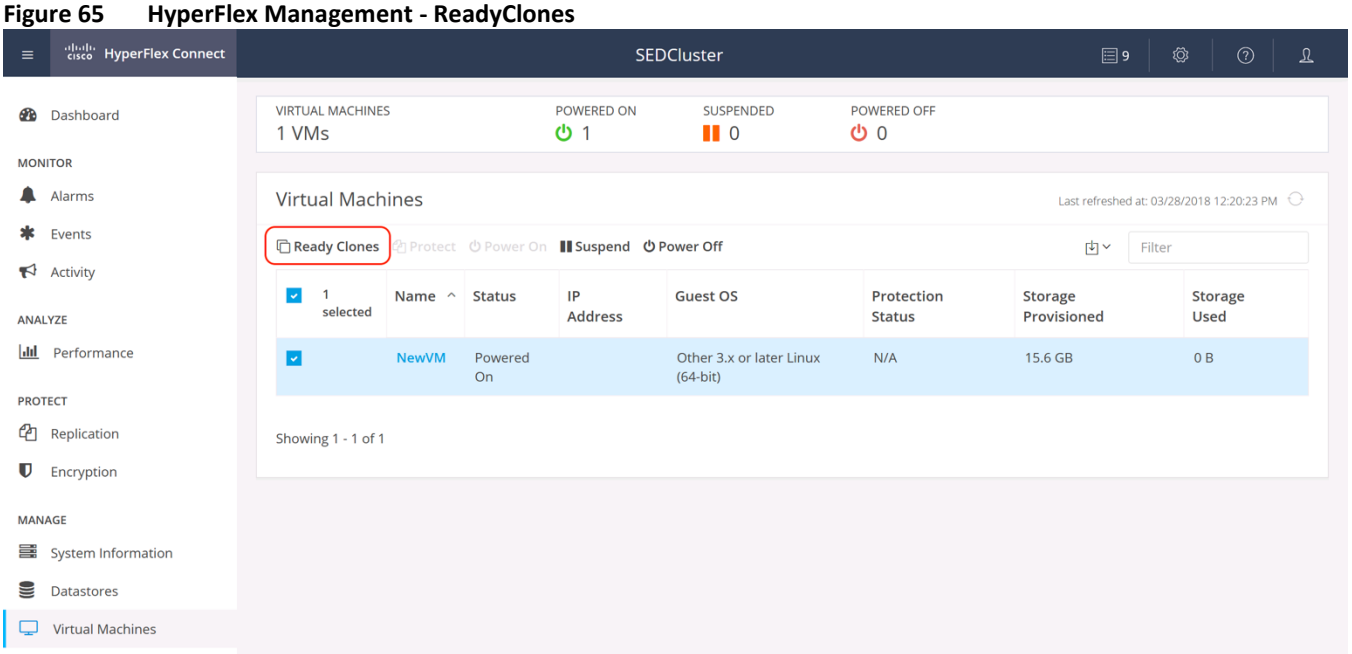
In this section, various best practices and guidelines are given for management and ongoing use of the Cisco HyperFlex system. These guidelines and recommendations apply only to the software versions upon which this document is based, listed in Software Components.

ReadyClones

For the best possible performance and functionality of the virtual machines that will be created using the HyperFlex ReadyClone feature, the following guidelines for preparation of the base VMs to be cloned should be followed:

- Base VMs must be stored in a HyperFlex datastore.
- All virtual disks of the base VM must be stored in the same HyperFlex datastore.
- Base VMs can only have HyperFlex native snapshots, no VMware redo-log based snapshots can be present.
- For very high IO workloads with many clone VMs leveraging the same base image, it might be necessary to use multiple copies of the same base image for groups of clones. Doing so prevents referencing the same blocks across all clones and could yield an increase in performance. This step is typically not required for most uses cases and workload types.

Figure 65 HyperFlex Management - ReadyClones

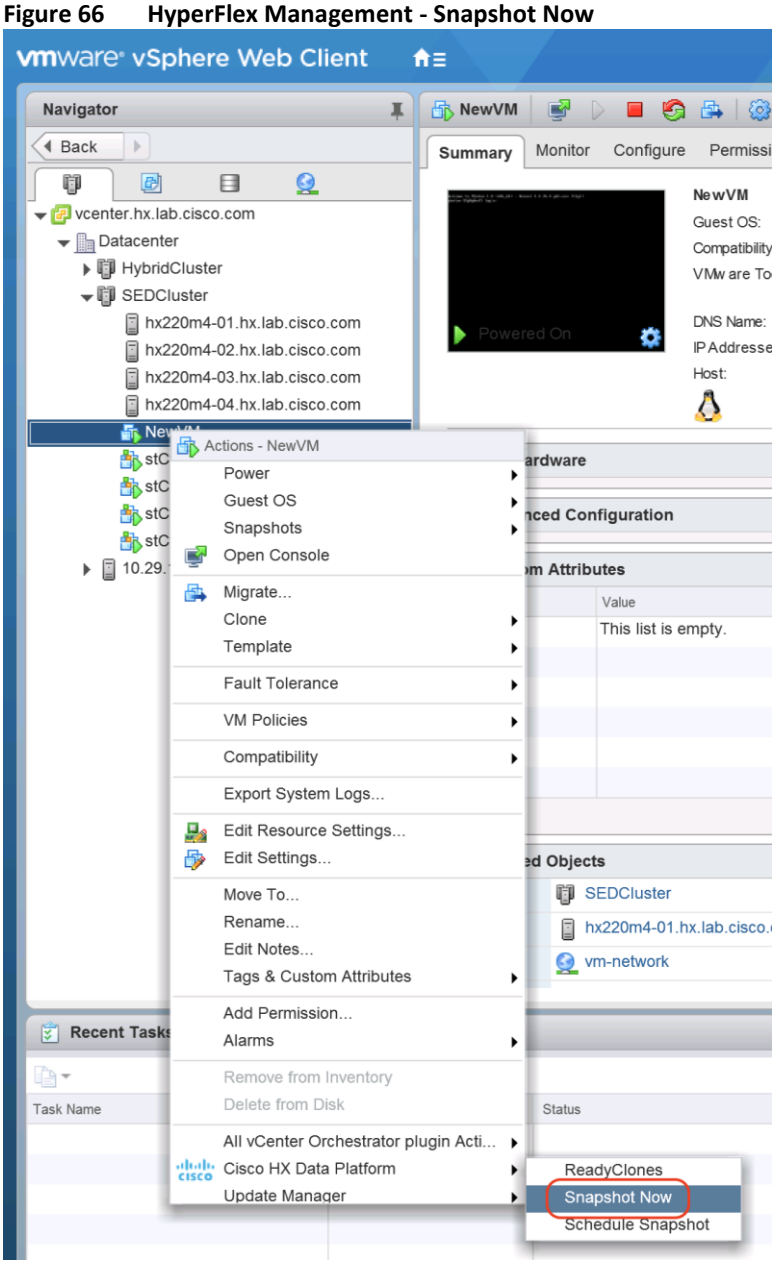


Snapshots

HyperFlex native snapshots are high performance snapshots that are space-efficient, crash-consistent, and application consistent, taken by the HyperFlex Distributed Filesystem, rather than using VMware redo-log based snapshots. For the best possible performance and functionality of HyperFlex native snapshots, the following guidelines should be followed:

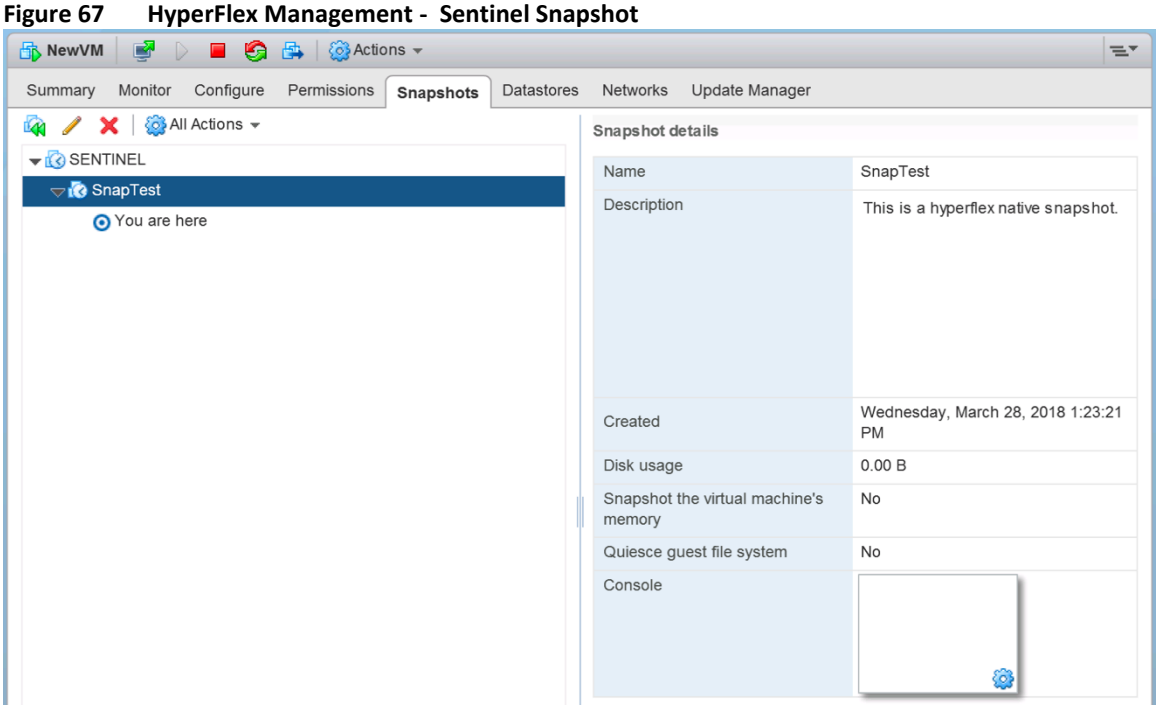
- Make sure that the first snapshot taken of a guest VM is a HyperFlex native snapshot, by using the "Cisco HX Data Platform" menu item in the vSphere Web Client, and choosing Snapshot Now or Schedule Snapshot. Failure to do so reverts to VMware redo-log based snapshots. (Figure 66)

Figure 66 HyperFlex Management - Snapshot Now



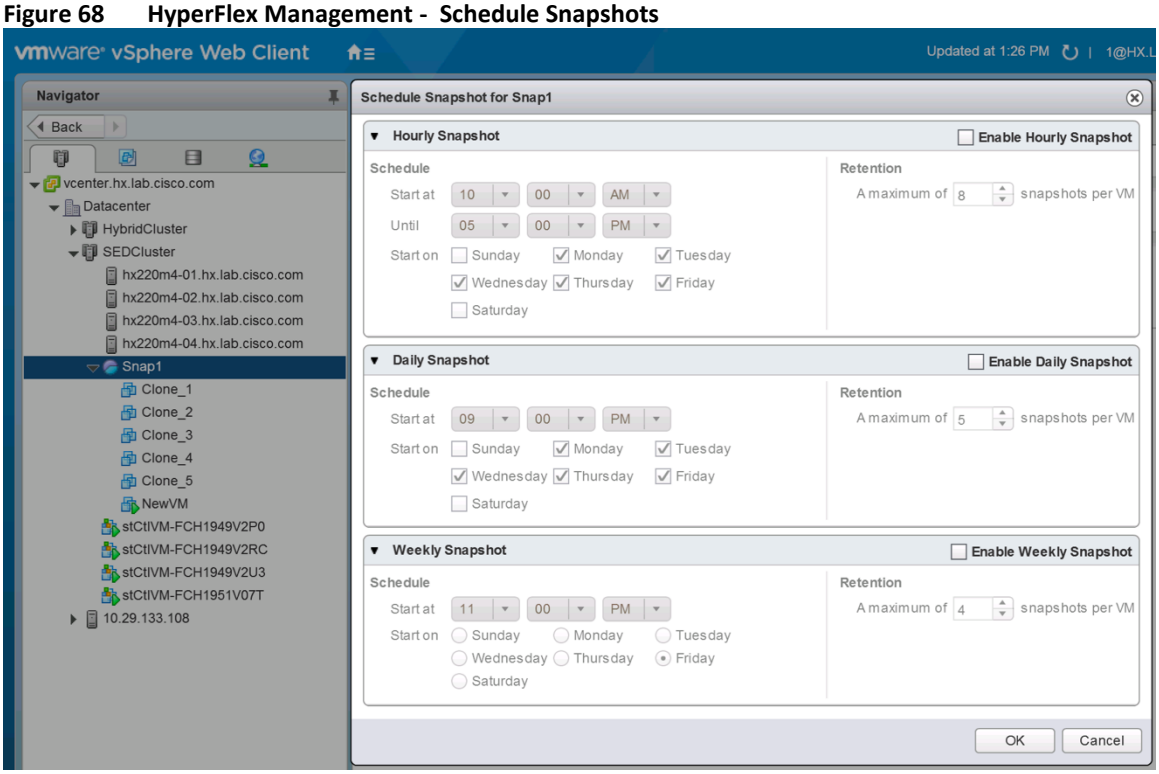
- A Sentinel snapshot becomes a base snapshot that all future snapshots are added to, and prevents the VM from reverting to VMware redo-log based snapshots. Failure to do so can cause performance degradation when taking snapshots later, while the VM is performing large amounts of storage IO.
- Additional snapshots can be taken via the “Cisco HX Data Platform” menu, or the standard vSphere client snapshot menu. As long as the initial snapshot was a HyperFlex native snapshot, each additional snapshot is also considered to be a HyperFlex native snapshot.
- Do not delete the Sentinel snapshot unless you are deleting all the snapshots entirely.
- Do not revert the VM to the Sentinel snapshot. (Figure 67)

Figure 67 HyperFlex Management - Sentinel Snapshot



- If large numbers of scheduled snapshots need to be taken, distribute the time of the snapshots taken by placing the VMs into multiple folders or resource pools. For example, schedule two resource groups, each with several VMs, to take snapshots separated by 15 minute intervals in the scheduler window. Snapshots will be processed in batches of 8 at a time, until the scheduled task is completed. (Figure 68)

Figure 68 HyperFlex Management - Schedule Snapshots



Storage vMotion

The Cisco HyperFlex Distributed Filesystem can create multiple datastores for storage of virtual machines. While there can be multiple datastores for logical separation, all of the files are located within a single distributed filesystem. As such, performing storage vMotions of virtual machine disk files has little value in the HyperFlex system. Furthermore, storage vMotions create additional filesystem consumption and generate additional unnecessary metadata within the filesystem, which must later be cleaned up via the filesystem's internal cleaner process.

Note: It is recommended to not perform a storage vMotion of a guest VM between datastores within the same HyperFlex cluster. Storage vMotion between different HyperFlex clusters, or between HyperFlex and non-HyperFlex datastores are permitted.

Virtual Disk Placement

HyperFlex clusters can create multiple datastores for logical separation of virtual machine storage, yet the files are all stored in the same underlying distributed filesystem. The only difference between one datastore and another are their names and their configured sizes. Due to this, there is no compelling reason for a virtual machine's virtual disk files to be stored on a particular datastore versus another.

Note: All of the virtual disks that make up a single virtual machine must be placed in the same datastore. Spreading the virtual disks across multiple datastores provides no benefit, can cause ReadyClone and Snapshot errors, and lead to degraded performance in stretched clusters.

Maintenance Mode

In HyperFlex Connect, from the System Information screen, in the Nodes view, the individual nodes can be placed into HX Maintenance Mode. Also, within the vCenter Web Client, a specific menu entry for "HX Maintenance Mode" has been installed by the HyperFlex plugin. This option directs the storage platform controller on the node to shutdown gracefully, redistributing storage IO to the other nodes with minimal impact. Using the standard Maintenance Mode menu in the vSphere Web Client, or the vSphere (thick) Client can be used, but graceful failover of storage IO and shutdown of the controller VM is not guaranteed.

Note: In order to minimize the performance impact of placing a HyperFlex converged storage node into maintenance mode, it is recommended to use the HX Maintenance Mode menu selection to enter or exit maintenance mode whenever possible.

Figure 69 HyperFlex Connect - HX Maintenance Mode

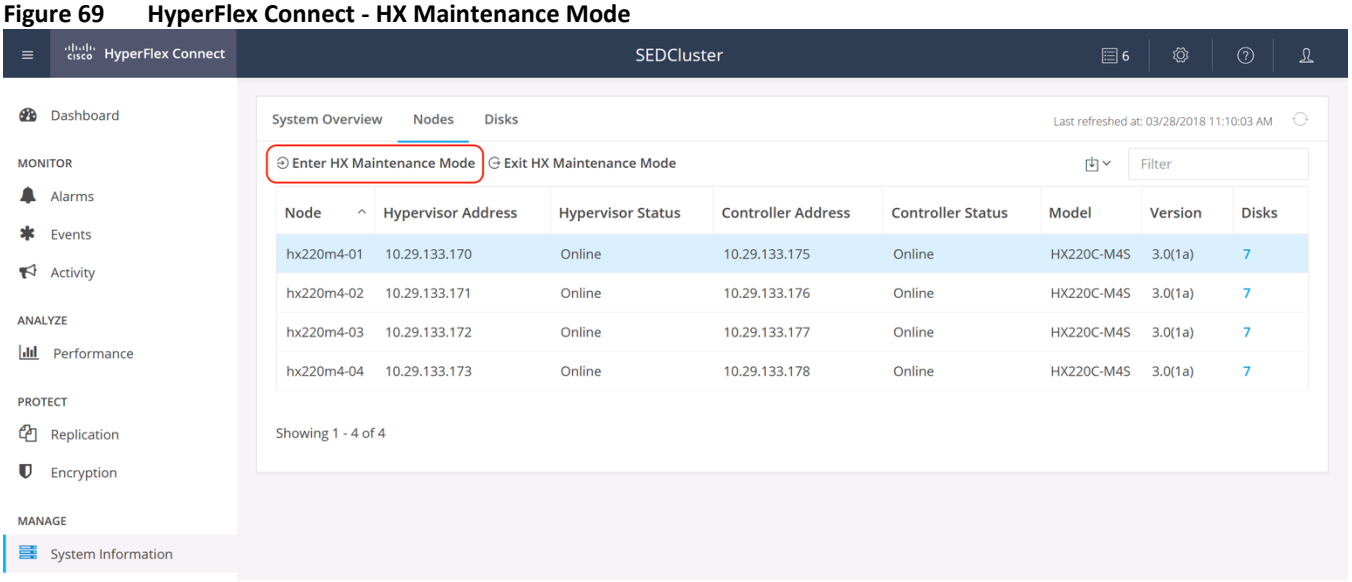
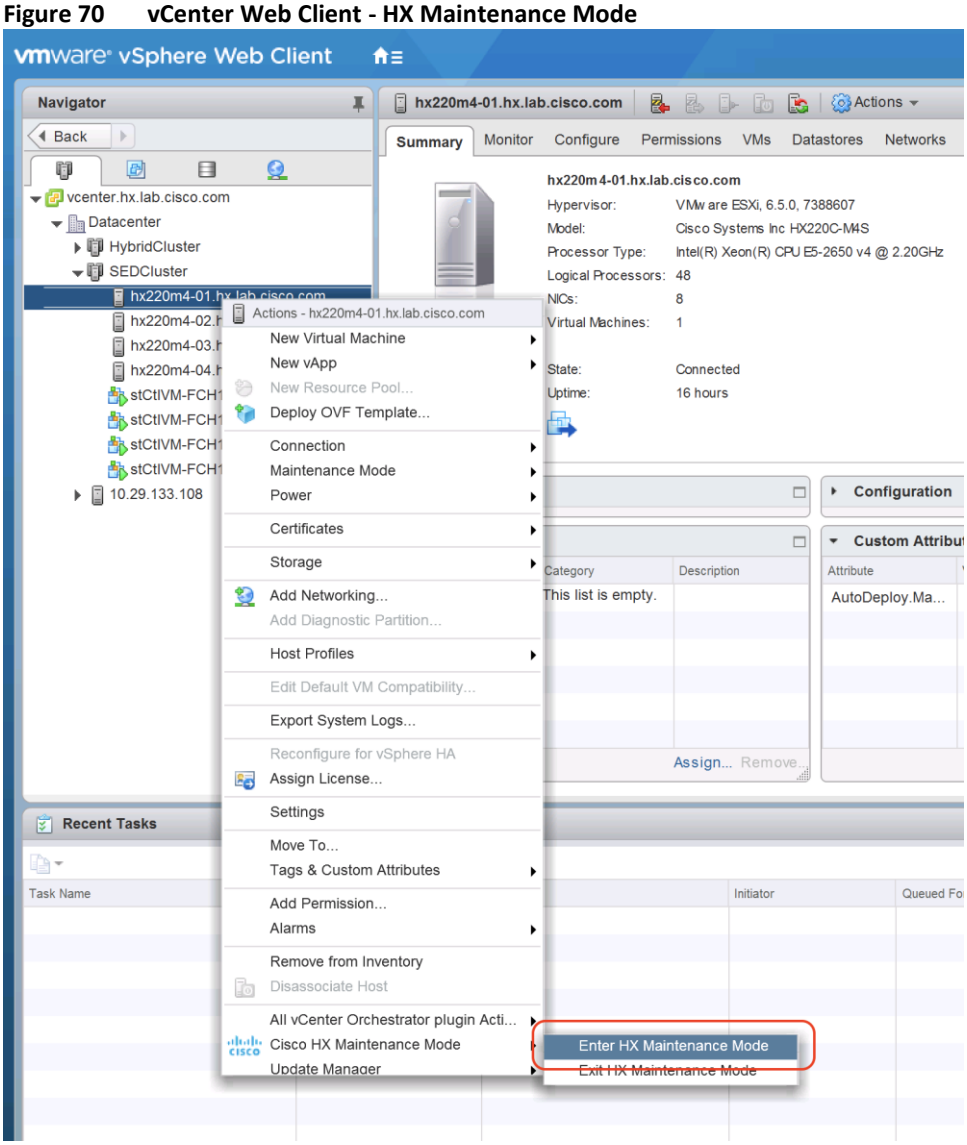


Figure 70 vCenter Web Client - HX Maintenance Mode



Encryption

HyperFlex 2.5 introduced new data protection features, including data-at-rest encryption. HyperFlex clusters can be ordered with self-encrypting disks (SED) which encrypt all of the data stored on them. A cluster using SEDs will store all of its data in an encrypted format, and the disks themselves perform the encryption and decryption functions. Since the hardware handles all the encryption and decryption functions, no additional load is placed on the CPUs of the HyperFlex nodes. Storing the data in an encrypted format prevents data loss and data theft, by making the data on the disk unreadable if it is removed from the system. This protection of the data enables HyperFlex to be used in environments where high security is desired, or compliance with industry security standards is required, such as healthcare providers (HIPAA), financial accounting systems (SOX), credit card transactions (PCI), and more.

Each SED contains a factory generated data encryption key (DEK) which is stored on the drive in a secured manner, and is used by the internal encryption circuitry to perform the encryption of the data. In truth, an SED always encrypts the data, but the default operation mode is known as the unlocked mode, wherein the drive can be placed into any system and the data can be read from it. To provide complete security, the SED needs to be locked, and reconfigured into what is called auto-unlock mode. This is accomplished via software, using another encryption key, called the authentication key (AK). The authentication key is generated externally from the SED and used to encrypt the DEK. When an SED operates in auto-unlock mode its DEK is encrypted, so when the SED is powered on, the AK must be provided by the system, via the disk controller, to decrypt the DEK, which then allows the data to be read. Once unlocked, the SED will continue to operate normally until it loses power, when it will automatically lock itself. If a locked SED is removed from the system, then there is no method for providing the correct AK to unlock the disk, and the data on the disk will remain encrypted and unreadable.

In order to configure a HyperFlex cluster for encryption, all of the disks on all of the nodes of the cluster must be SEDs. The authentication keys which are used to encrypt the data encryption keys on the disks must be supplied by the HyperFlex cluster. The authentication keys can be provided in one of three ways:

- Local keys in Cisco UCS Manager derived from an encryption passphrase. Local keys are simpler to configure, and are intended for use in testing, proof-of-concept builds, or environments where an external Key Management System (KMS) is not available. Local key configurations create a single authentication key (AK) which is used to encrypt all the disks on all the nodes of the cluster.
- Remote keys, where Cisco UCS Manager retrieves the keys via Key Management Interoperability Protocol (KMIP) from a remote KMS. The client/server communications between the HX nodes and the KMIP server are secured using trusted certificate authority (CA) signed keys, created from certificate signing requests (CSR). Remote key configurations create a unique authentication key for each node, and that AK is used for all disks on that node, providing an even higher level of security.
- Remote keys, where Cisco UCS Manager retrieves the keys via Key Management Interoperability Protocol (KMIP) from a remote KMS, but the client/server communications between the HX nodes and the KMIP server are secured using self-signed certificates.

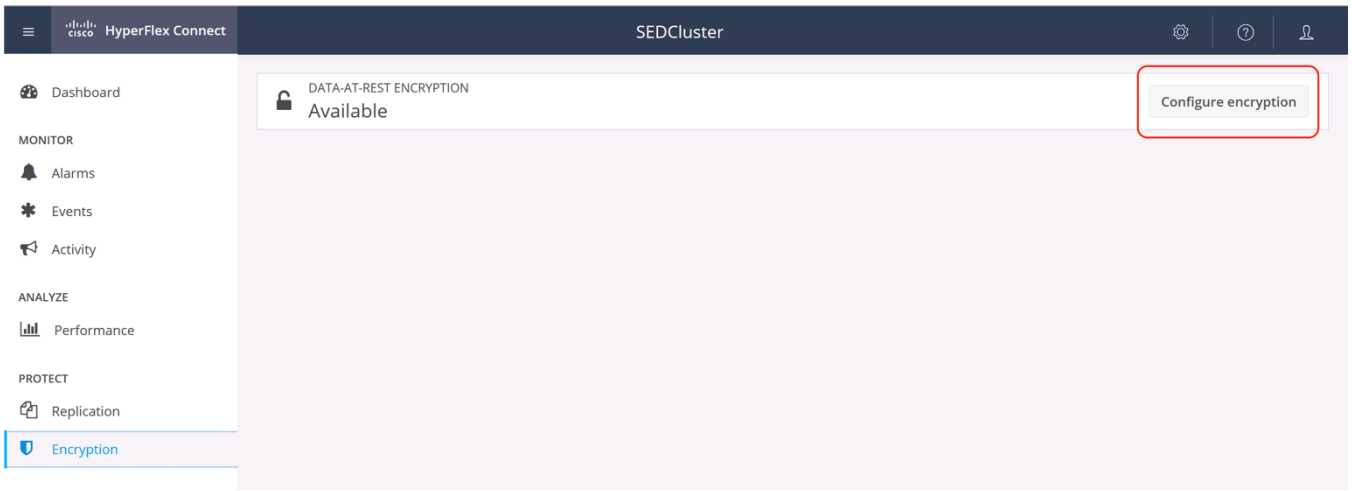
Cisco has tested remote and self-signed keys using KMS systems, including Gemalto SafeNet KeySecure, and Vormetric DSM. A large number of steps are required to perform the configuration of a certificate authority (CA), root certificates, and signing certificates. Additionally, these steps are significantly different depending on the KMS being used. Because of this, the specific steps needed to configure encryption with remote keys is not covered in this design document.

Note: The HyperFlex Connect encryption menu and configuration options are only available when the cluster contains encryption capable hardware on all of the nodes.

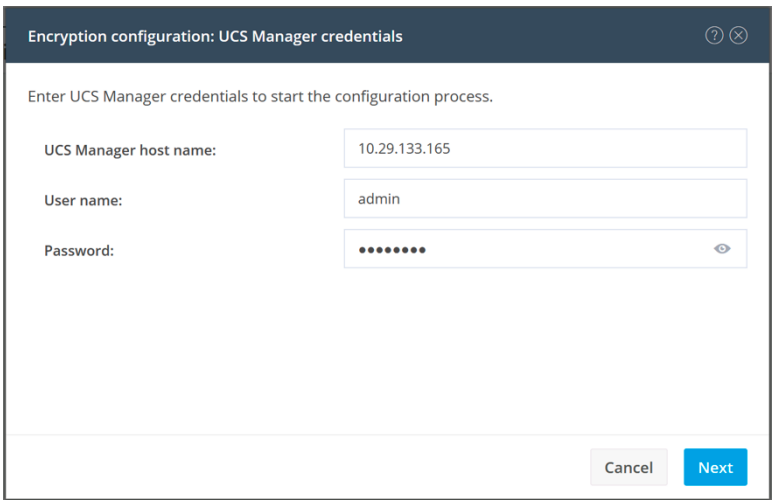
To enable encryption using locally managed keys in Cisco UCS Manager, complete the following steps:

1. Open HyperFlex Connect and log in with admin privileges.

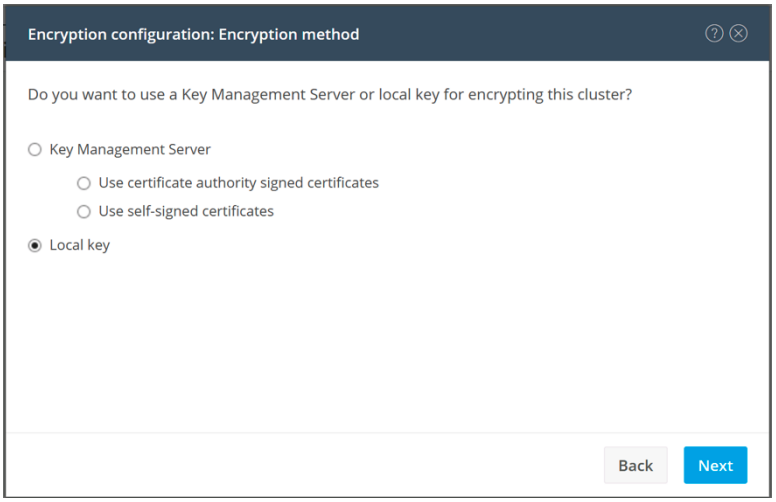
2. Click Encryption in the menu on the left, then click the Configure encryption button.



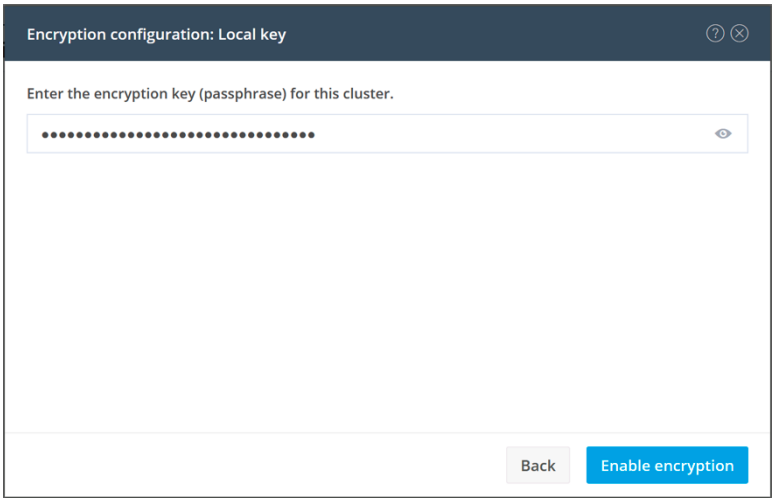
3. Enter the Cisco UCS Manager IP address or hostname, an administrative username, and password, then click Next.



4. Click the option for Local key, then click Next.



5. Enter an encryption key passphrase, which must be exactly 32 characters long, then click Enable Encryption.

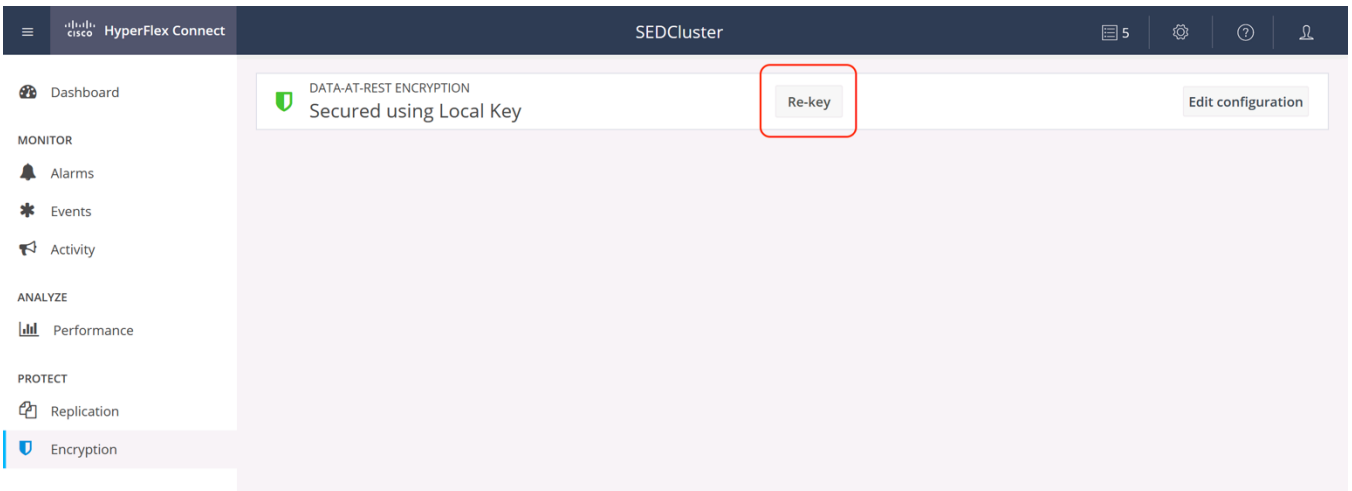


Rekey

At any time, it may be determined for security purposes that it is necessary to regenerate the authentication keys in the cluster, which are used to unlock the encrypted contents of the disks. A rekey operation can be run to regenerate the keys, in case the existing keys may have been compromised, or as part of company policy. A rekey operation is non-destructive to the existing data, and the data remains encrypted at all times. To rekey the drives, complete the following steps:

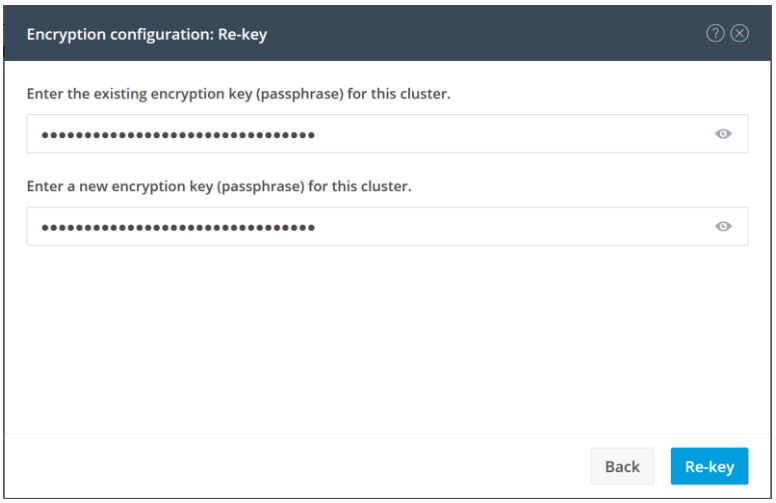
1. Open HyperFlex Connect and log in with admin privileges.

2. Click Encryption in the menu on the left, then click the Re-key button.



3. Enter the Cisco UCS Manager IP address or hostname, an administrative username, and password, then click Next.

4. Enter the existing encryption passphrase, and a new 32 character encryption passphrase, then click Re-key.



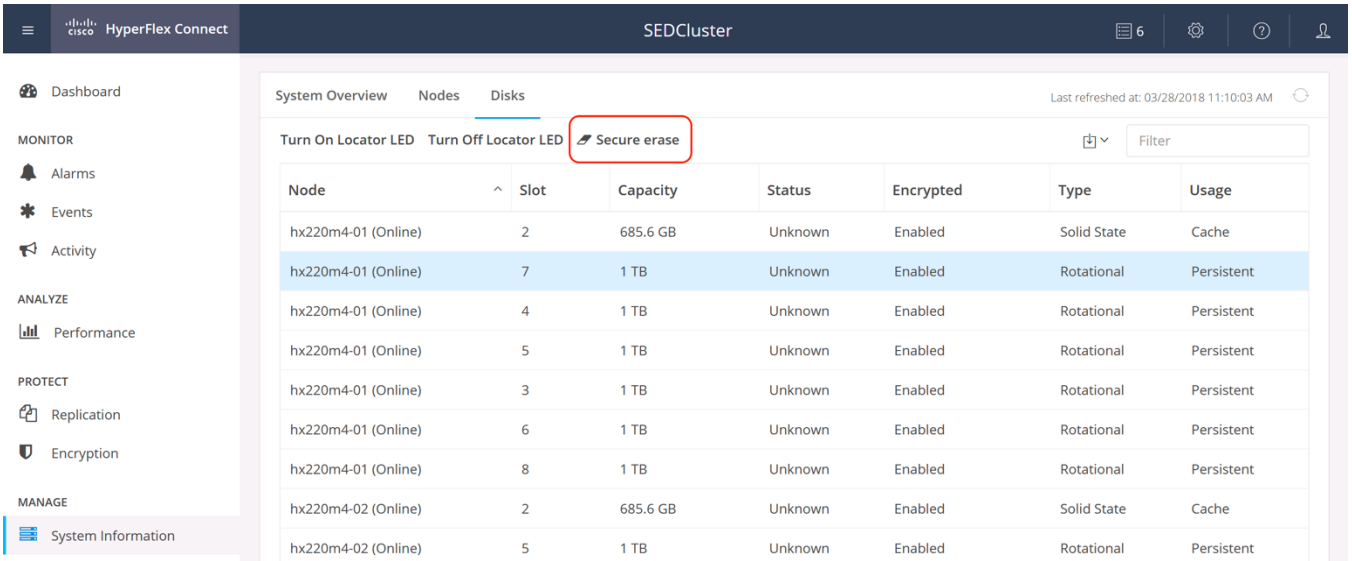
Secure Erase

If an encrypted drive is failed, a predicted failure alarm is triggered, or if a drive is otherwise going to be removed from a node, the drive can be securely erased before its removal. Erasing a drive is a destructive event to the data on that disk, however the data still exists as replicas in other locations across the cluster. A disk secure erase will trigger an event in the cluster similar to a disk failure, and the lost data segments will be recreated in other online locations in the cluster, in order to return the data to its configured replication factor. To securely erase a drive, complete the following steps:

1. Open HyperFlex Connect and log in with admin privileges.

2. Click System Information in the menu on the left, then click Disks.

3. Highlight the disk to be erased, then click Secure Erase.



4. For a cluster using local encryption keys, enter the encryption passphrase, for remote key configurations, no action is necessary.
5. Click Secure Erase.
6. Click “Yes, erase this disk” at the confirmation pop-up.
7. When complete, the disk status will change to “Ok to remove”.
8. Remove the disk from the HX node.

WARNING! If an SED is securely erased, it cannot be put back into service in the same or even a different HX cluster. The only method to reuse an erased SED is to insert the drive into an HX node and install/reinstall that cluster from scratch.

Replication

HyperFlex 2.5 introduced new data protection features, including snapshot-based VM level replication between two HyperFlex clusters. Replication can be used to migrate or recover a single VM in the secondary HX cluster, groups of VMs can be coordinated and recovered, or all VMs can be recovered as part of a disaster recovery scenario. In order to start using replication, two HyperFlex clusters must be installed and have network connectivity between them. The clusters can be extended clusters, and it is possible to replicate between hybrid and all-flash clusters. The clusters are allowed to use self-encrypting disks or standard disks in either location, both of them, or neither of them, there is no restriction in that respect. To avoid complications with duplicate VM IDs, it is recommended that the two replicating HyperFlex clusters be managed by two different VMware vCenter servers.

After a HyperFlex cluster is installed, none of the networking configuration required for replication is in place. In order to use replication, the replication networking must first be configured in HyperFlex Connect, which automates the changes in Cisco UCS Manager, configures the ESXi port groups, and assigns the new replication IP addresses to the SCVMs. Once the networking configuration work is completed for both clusters that will replicate to each other, a partnership, or pairing between the two clusters is established in HyperFlex Connect. After this replication pair is established, VMs can be protected individually, or they can be placed into protection groups, which are created to protect multiple VMs with the same replication settings. VMs can be replicated in intervals as often as once per 5 minutes, up to once per 24 hours, which is analogous to the Recovery Point Objective (RPO). Care must be taken to ensure that the two clusters have enough storage capacity to store the replicated snapshots of the remote cluster's VMs, and also have enough CPU and RAM resources to run those VMs in case they must be recovered. HyperFlex Connect can be used to monitor the status of the protected VMs, and to initiate migrations, test recoveries, or failovers of the VMs. In some cases, individual tasks must be done using the stcli command line interface, but these tasks are made easier when using the WebCLI tool in the HyperFlex Connect webpage due to auto-completion of information.

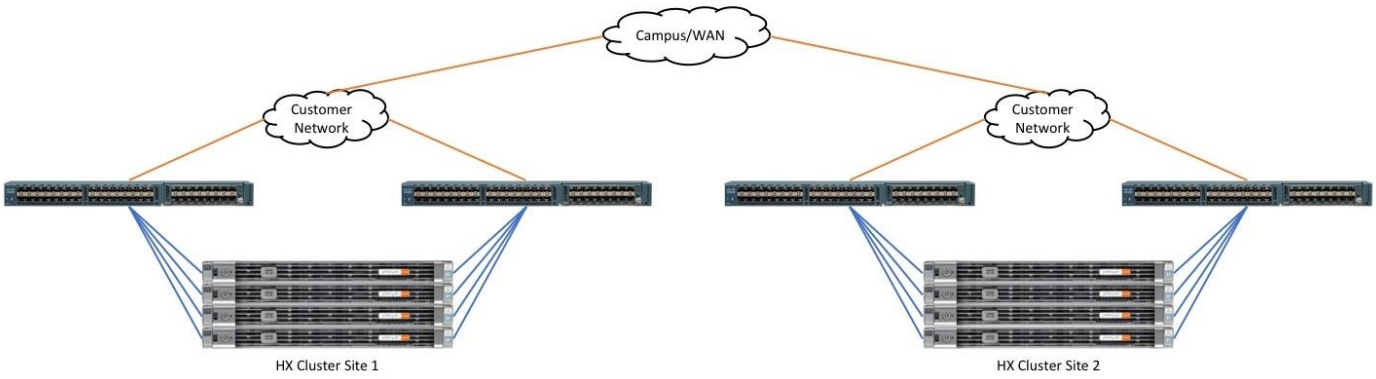
Replication Networking

The two HyperFlex clusters that will replicate must have TCP/IP connectivity between them, and additional IP addresses must be provided to an internal IP address pool that the HX SCVMs will use. The minimum number of IP addresses required is the number of nodes in the cluster, plus 1 additional address. More addresses than are currently needed can be placed into the pool to allow for future growth of the HX cluster. An existing VLAN ID and subnet can be used, although it is more typical to configure a specific VLAN and subnet to carry replication traffic that will traverse the campus or WAN links between the two clusters. The VLANs that will be used for replication traffic must already be trunked to the Cisco UCS Fabric Interconnects from the northbound network by the upstream switches, and this configuration step must be done manually prior to beginning the HyperFlex Connect configuration. The bandwidth usage of the replication traffic can be set to a limit so as not to saturate the interconnecting network links, or it may be left unlimited. The bandwidth consumption will be directly affected by the number of VMs being protected, and the frequency of their replication.

The interconnection between the two clusters at the two sites can be done in several ways. In most cases, the uplinks from the HX clusters will carry all the needed VLAN IDs on the same set of interfaces, including HX management, vMotion, storage traffic, guest VM traffic, and the replication traffic. In some cases, it is desired that the replication traffic will traverse a set of independent uplinks, which is referred to as a split L2 topology. Due to a technical limitation of implementing a split L2 topology, the configuration of replication networking cannot accommodate a split L2 configuration. Specifically, a single Cisco UCS vNIC cannot carry multiple VLANs that traverse multiple uplink groups. Since the default configuration uses vmnic0 and vmnic1 to carry HX management traffic and replication traffic, both of those VLANs must arrive to UCS across a

single set of uplinks. The replication subnets and VLANs used in the two sites can be different routed subnets, or they can be a single subnet if other technologies, such as OTV, are in use by the WAN. Replication traffic originates and terminates on the SCVMs running on each HX host.

Figure 71 Replication Networking

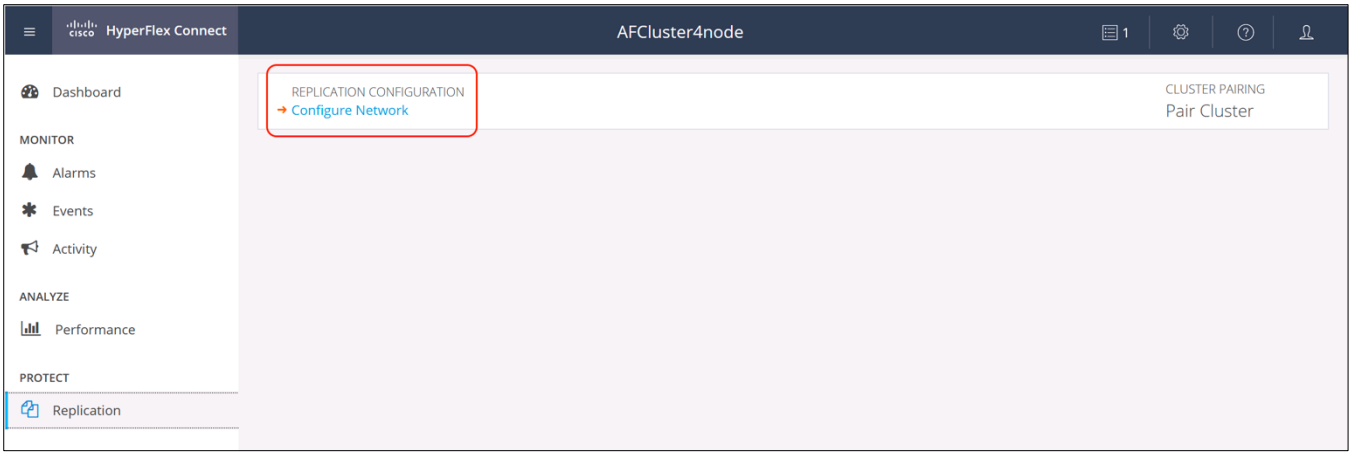


Configuring replication networking in HyperFlex Connect automates the following tasks:

- Creates the replication VLAN in Cisco UCS Manager.
- Adds the new replication VLAN to the VNIC templates named hv-mgmt-a and hv-mgmt-b in the appropriate sub-organization in Cisco UCS Manager.
- Sets the VLAN ID of the Storage Controller Replication Network port group on all ESXi nodes.
- Creates a pool of IP addresses internal to the HyperFlex cluster, from which each SCVM will draw one IP address, plus 1 additional IP will be used as a roaming clustered address.
- Instructs the SCVMs to request an individual IP address, and configures the clustered IP address.

To configure the replication network, complete the following steps:

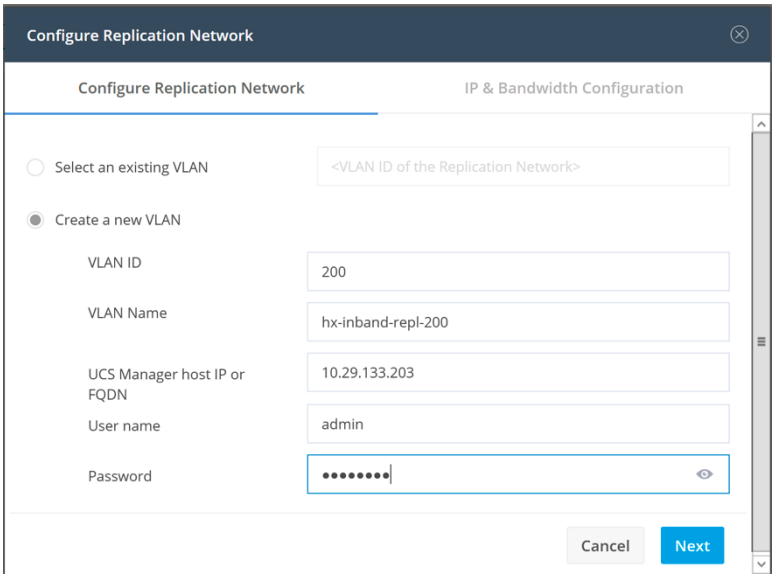
1. Open HyperFlex Connect and log in with admin privileges.
2. Click Replication in the menu on the left, then click the Configure Network link.



3. In the Configure Replication Network section, click the radio button to either use an existing VLAN in UCS Manager for replication, or Create a new VLAN.

4. If you use an existing VLAN, enter the VLAN ID of the existing VLAN.

5. If you are creating a new VLAN, enter the new VLAN ID, the VLAN name, along with the Cisco UCS Manager IP address or hostname, an administrative username, and password, then click Next.

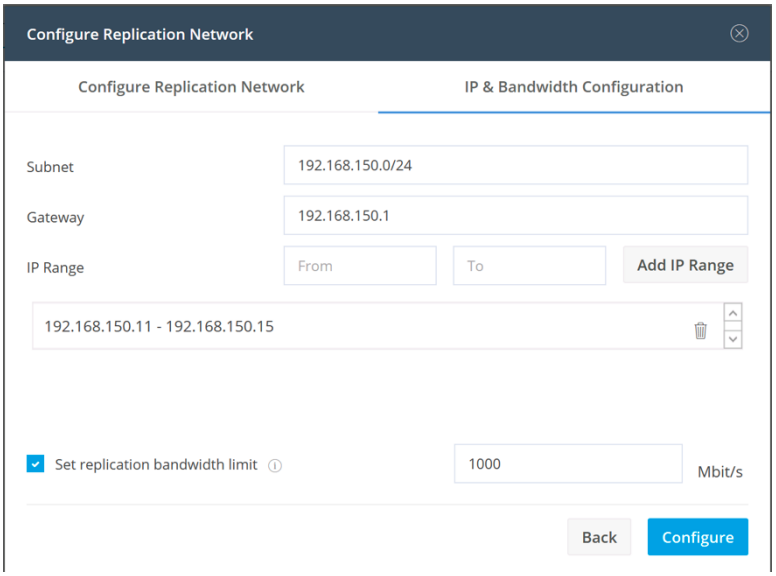


6. In the IP & Bandwidth Configuration section, enter the replication subnet in CIDR notation, i.e. a.b.c.d/n, and the gateway IP address for the subnet.

7. Enter the starting and ending IP addresses for the range that will be added to the pool assigned to the SCVMs, and click the Add button.

8. If outbound bandwidth limits must be set, check the box to enable it and enter a value between 10 and 100,000 Mbps. Cisco recommends limiting the bandwidth to 1000 Mbps or less.

9. Click Configure.



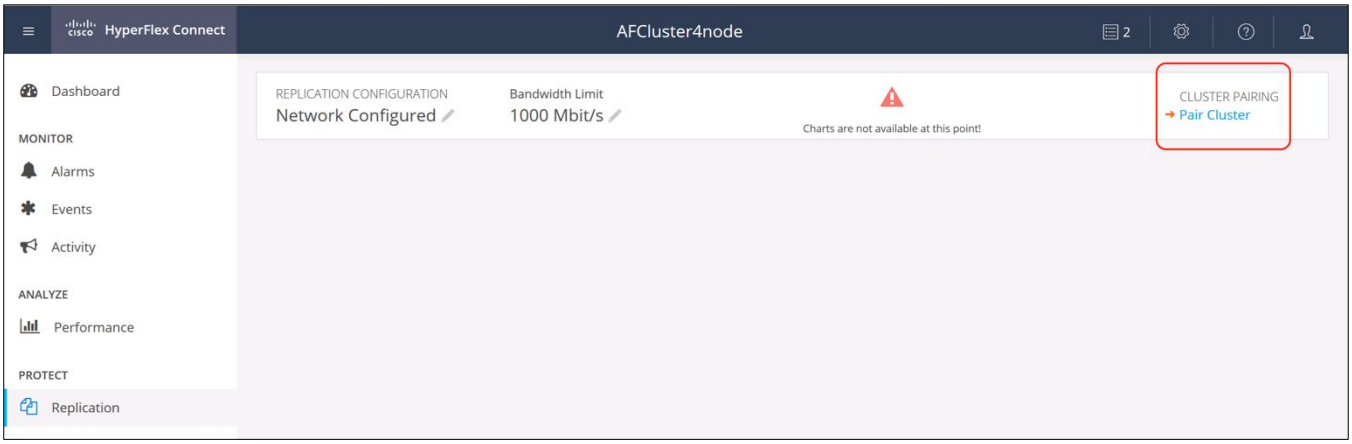
Replication Pairing

The two HyperFlex clusters that will be able to replicate VMs to each other must first be paired before the replication can begin, and a datastore on each cluster is also paired. Prior to pairing, the replication networking on both clusters must be configured and datastores must have been created on both clusters. Each HyperFlex cluster can be paired with only one other cluster, and each datastore can only be paired with one other datastore. The snapshots and VMs on the source cluster, in their source datastore, will be replicated to the paired cluster and stored in the paired datastore. You must know the administrative login credentials of the remote cluster, and the remote cluster's management IP address in order to proceed.

To configure the replication pair, perform the following steps:

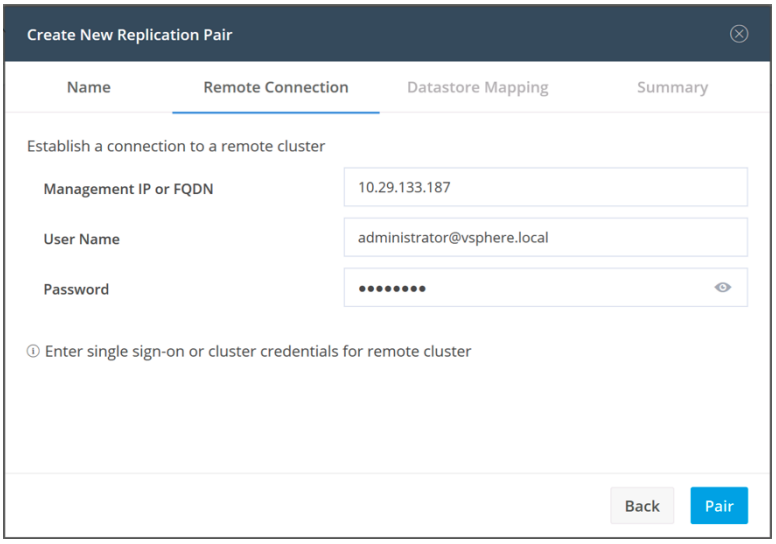
1. Open HyperFlex Connect and log in with admin privileges.

2. Click Replication in the menu on the left, then click the Pair Cluster link.

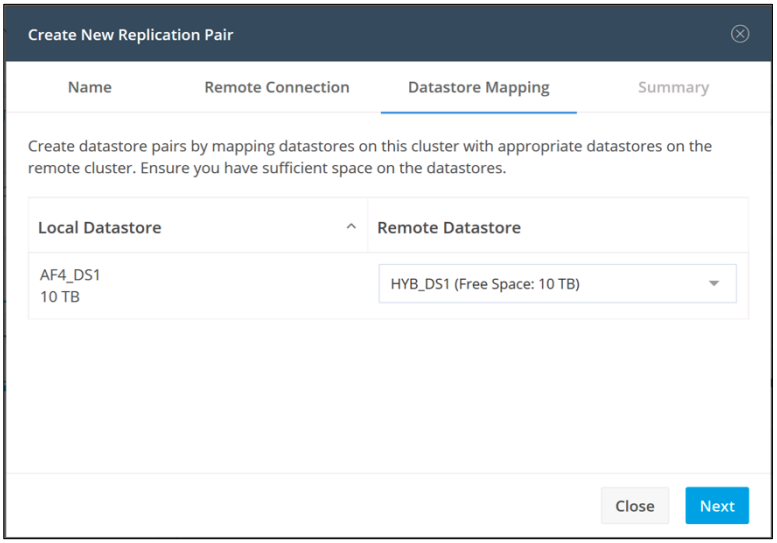


3. Enter a name for the replication pair, then click Next.

4. Enter the cluster management IP address or FQDN for the remote cluster, the username, and the password, then click Pair. The username and password must have admin rights in the vCenter server managing the remote cluster.



5. Pick the remote datastore to pair with the local datastore in the cluster, then click Next.



6. At the summary screen, click Map Datastores.

Protection Groups

Once a replication pair is established, and datastores are mapped to each other across two HX clusters, VM Protection can be configured. VMs can be protected individually, or they can be added to a new or existing Protection Group. Protection Groups can be created to allow for a common configuration of replication parameters to be applied to a collection of VMs, without configuring them individually. A good example would be creating multiple Protection Groups for several classes of protection, each with a different replication schedule, such as a “Platinum” group with a 5-minute schedule, a “Gold” group with a 15-minute schedule, a “Silver” group with a 2 or 4 hour schedule, and a “Bronze” group with a 12 or 24 hour schedule.

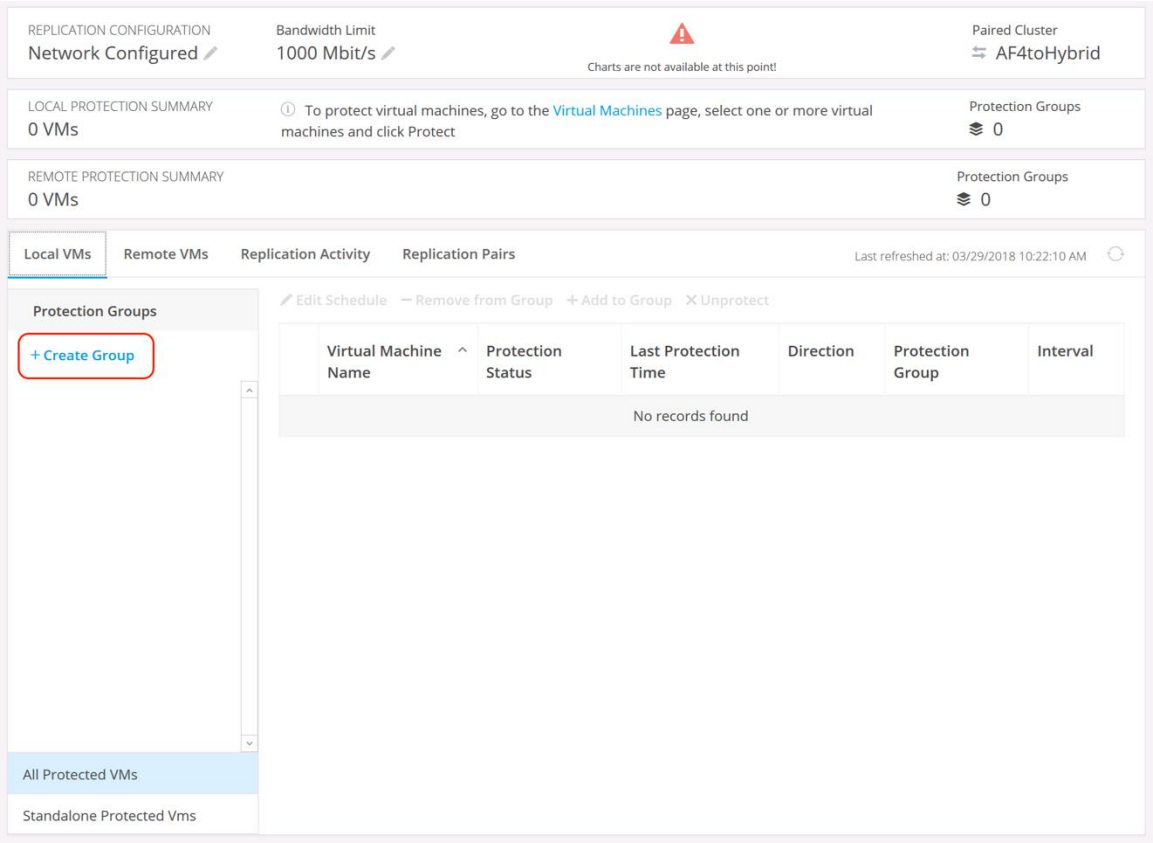
Protection Group names must be unique to both clusters that are paired together, and a maximum of 32 VMs can be added to a protection group.

Migration or recovery operations can be carried out against an entire protection group. If a protection group is halted, marking it for recovery, then all VMs within the group must be recovered on the secondary, or target cluster. If a VM is a member of a protection group, it cannot be individually migrated or recovered. If an individual VM must be migrated or recovered, but it is a member of a protection group, that VM must be removed from the group, thereby unprotecting it, then it must be individually protected again. Care must be taken that the individual protection replicates at least one snapshot before attempting a migration or recovery.

To create a Protection Group, complete the following steps:

1. Open HyperFlex Connect and log in with admin privileges.

2. Click Replication in the menu on the left, in the Local VMs view, under the Protection Groups view, click Create Group.

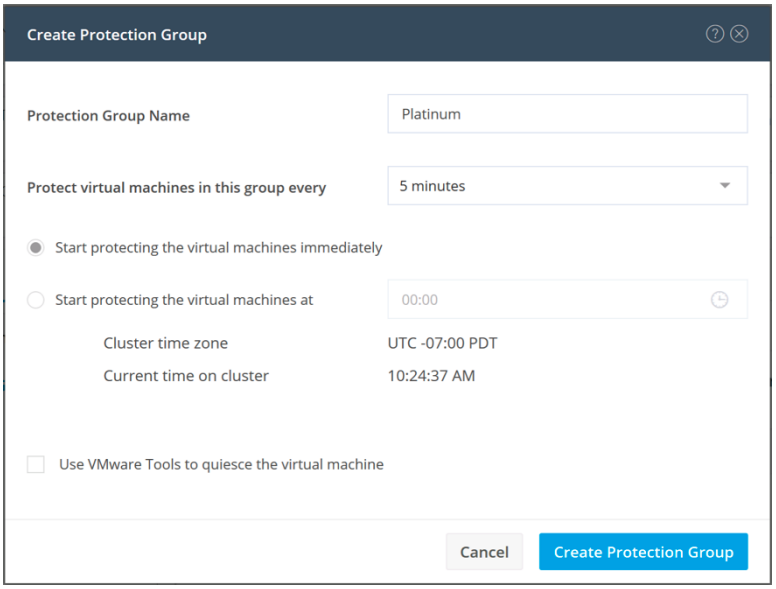


3. Enter a name for the group.

4. Choose the replication interval from the drop-down menu.

5. Choose a time for the replication to start, either immediately or at a future time.

6. Check the box if you wish to quiesce the VM's activity via VMware Tools during the snapshot, then click Create Protection Group.

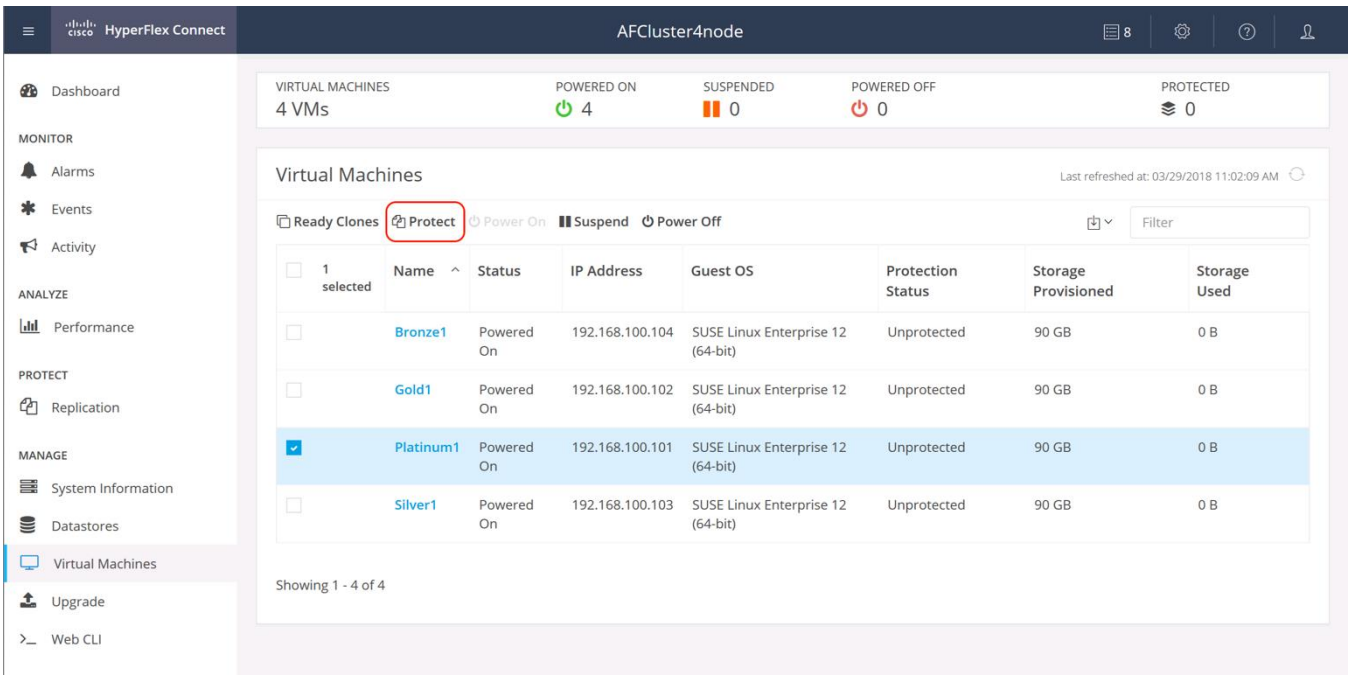


Virtual Machine Protection

Virtual machines can be configured for protection, i.e. replication, individually, or be placed into a Protection Group. The protection settings that can be configured on an individual VM are the same as the settings that are configured for a protection group. In most cases, it is easier to configure multiple Protection Groups, each with the settings that are required, and then add VMs to those groups. This process simplifies operations and helps ensure that replication schedules are not set improperly.

To protect a virtual machine, or group of virtual machines, complete the following steps:

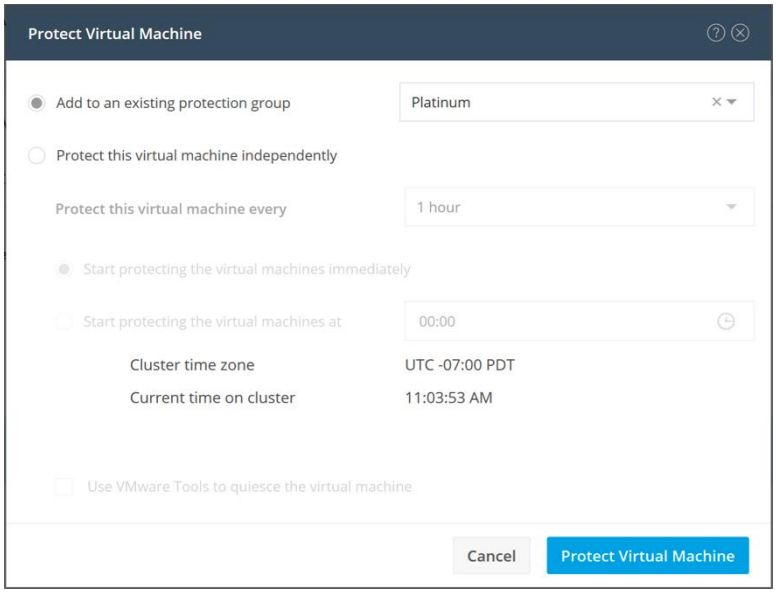
1. Open HyperFlex Connect and log in with admin privileges.
2. Click Virtual Machines in the menu on the left.
3. Check the box next to one or more VMs in the list, then click Protect.



4. Choose the option Add to an existing protection group, and choose the group to add the VM(s) to, then click Protect Virtual Machine.

   Or

5. Choose the option Protect this virtual machine independently, then choose the replication interval, choose a time for the replication to start, either immediately or at a future time, and choose if you would like to use the VMware Tools to quiesce the virtual machines, then click Protect Virtual Machine.



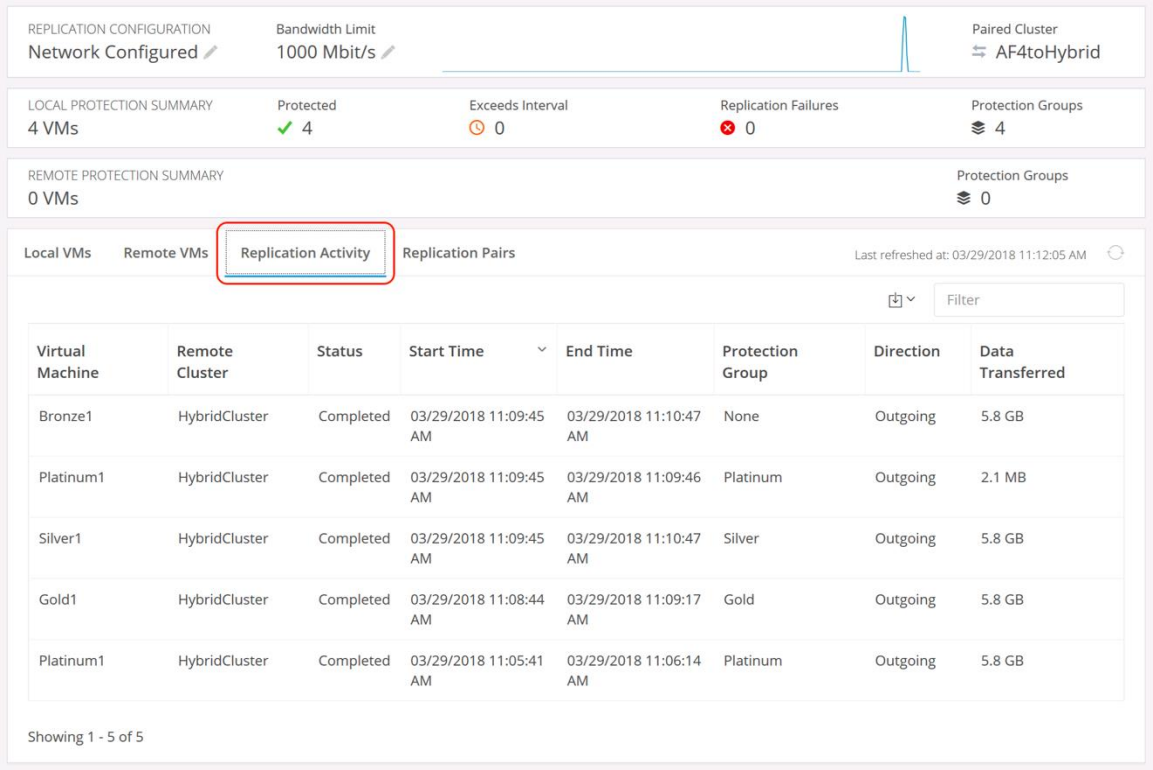
Note: When selecting multiple VMs to protect, the only options available are to place those VMs into a protection group, or create a new protection group. To protect multiple VMs with individual settings, each VM must be configured for protection, one-by-one.

Replication Monitoring

The HyperFlex Connect HTML GUI can be used to monitor the status of ongoing VM protection and replication.

The Replication Activity view shows the status of each individual snapshot replication operation.

Figure 72 Replication Activity View



The Local VMs view shows the protection status of all Protection Groups, Standalone VMs, or all VMs on the local system which have been configured for protection. The green Protected status indicates that the VM or group is being successfully protected according to the configured replication interval, or RPO.

Figure 73 Local VMs View

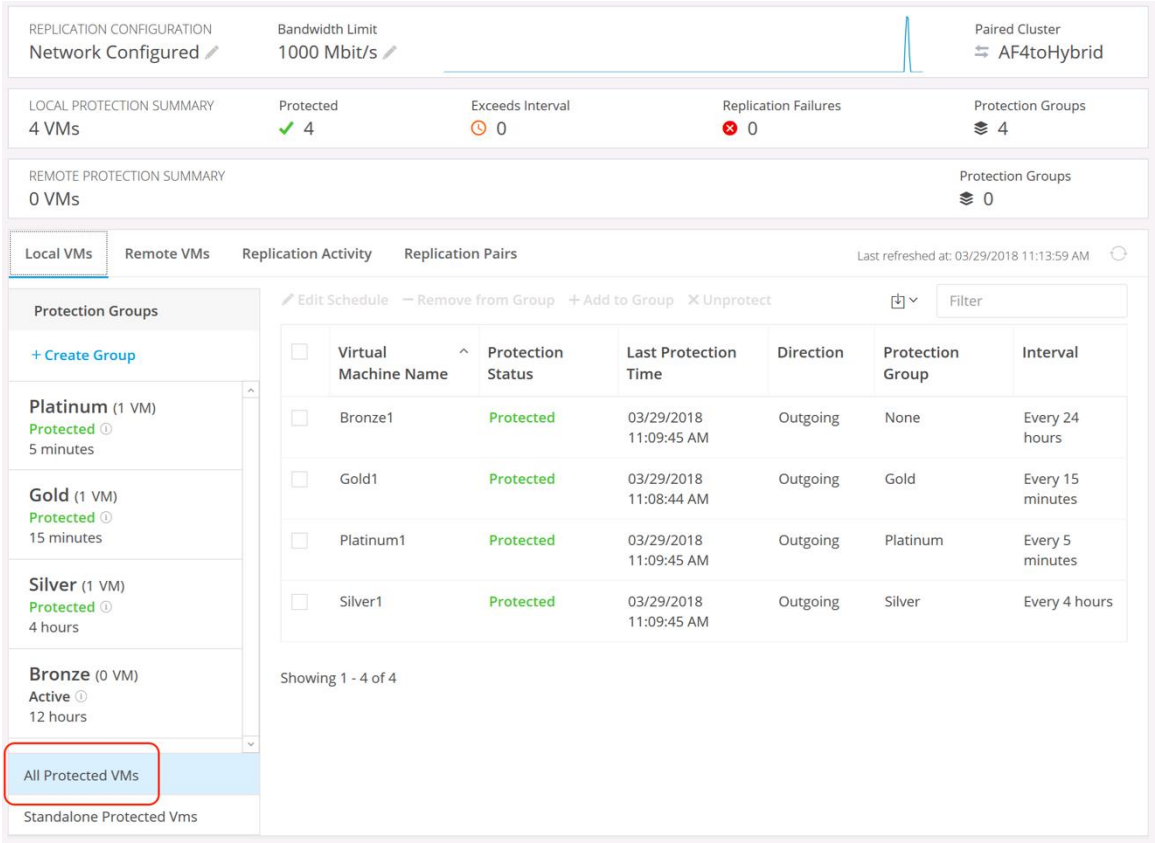
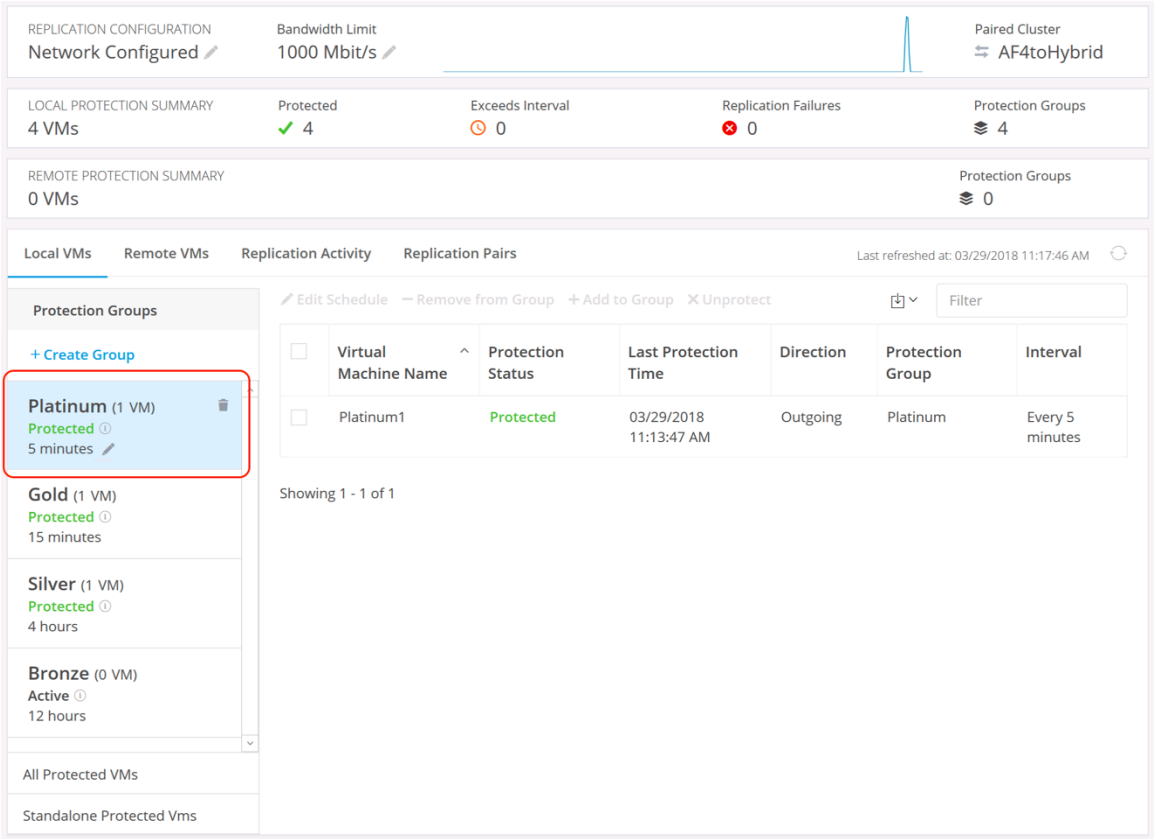
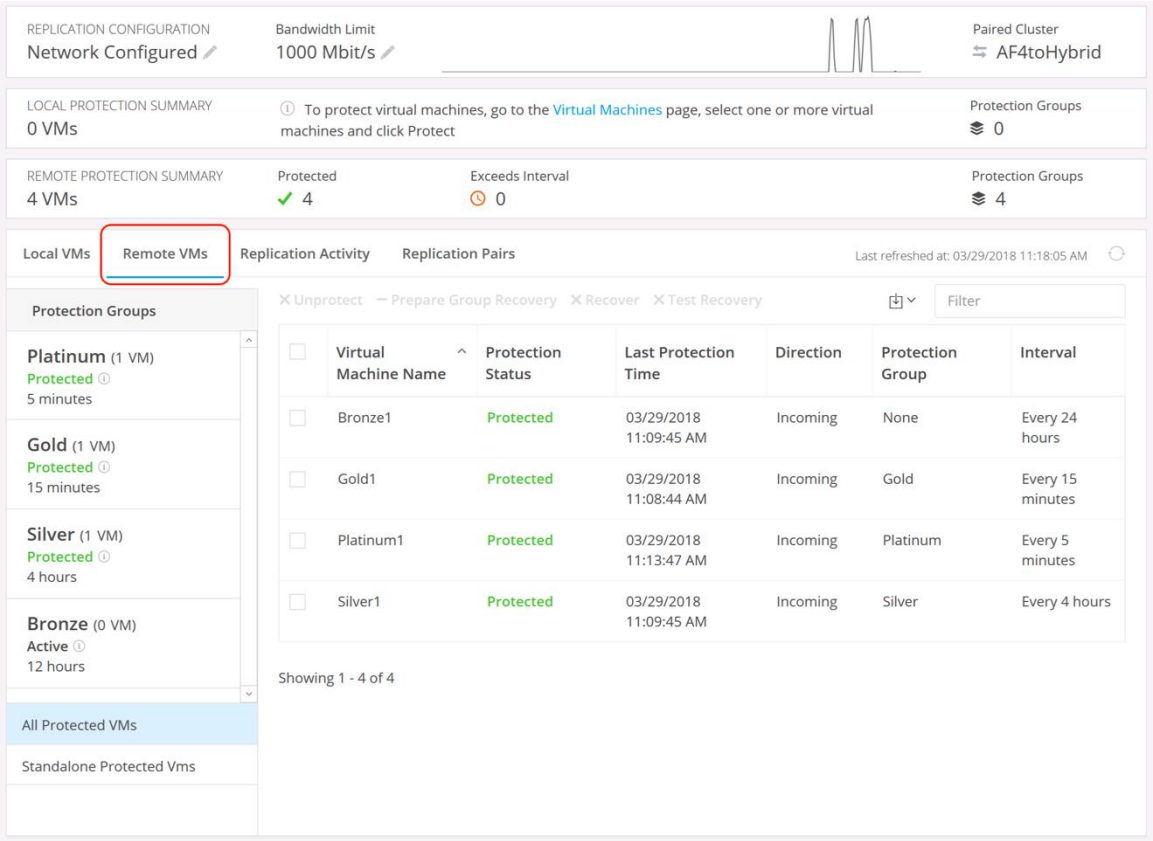


Figure 74 Local Protection Groups



On the receiving, or destination cluster, the same views of the VMs and Protection groups that are incoming can be seen from the Remote VMs view.

Figure 75 Remote VMs View



The Bandwidth Monitor in the upper right-hand corner can be viewed to quickly indicate the replication bandwidth used, or the graph can be clicked on to see an expanded view of the bandwidth consumed by the outgoing or incoming replication traffic.

Figure 76 Bandwidth Monitor

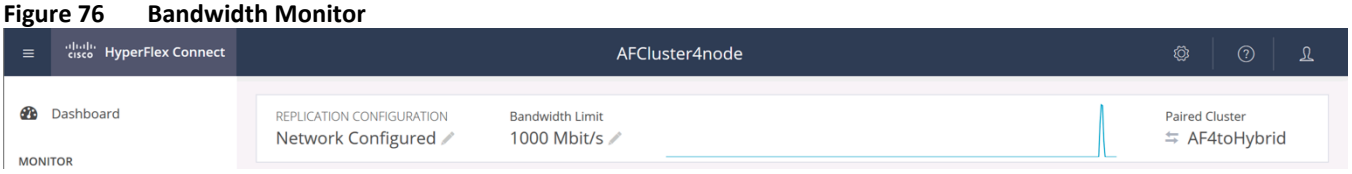
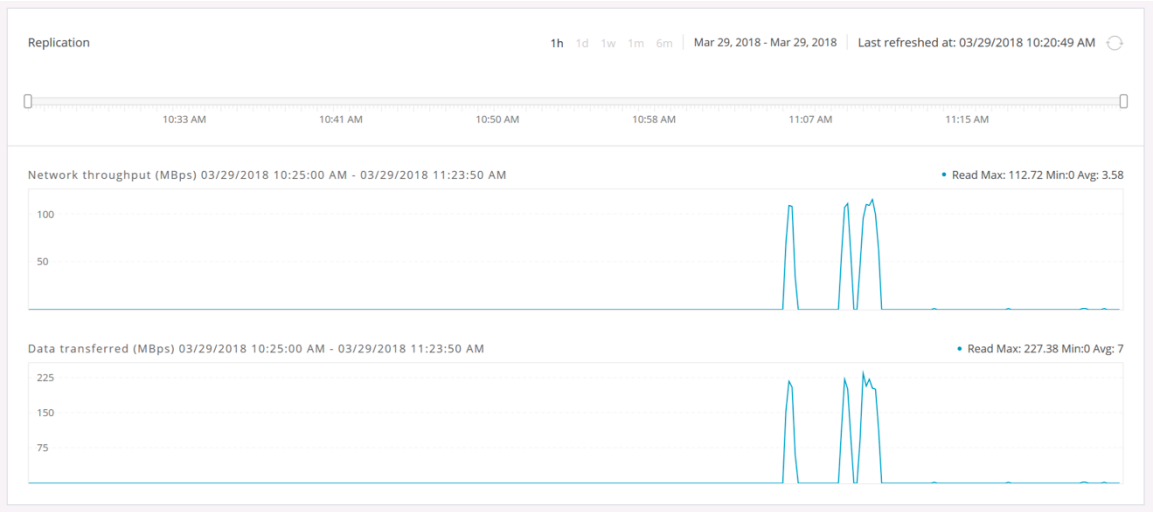


Figure 77 Bandwidth Charts



All of the replication monitoring views can also be accessed via the secondary, or target HX cluster, and the same VMs and protection groups will be presented, only as incoming VMs and groups instead of outgoing. Two paired HX clusters can replicate VMs in both directions, therefore the replication status of all VMs and Protection Groups, incoming and outgoing, are presented in the replication views of both clusters.

Replication Management

When configured, replication will run continuously in the background according to the configured schedules for the VMs and Protection Groups. If it is necessary to pause replication, for example during a maintenance activity such as an upgrade, replication can be paused and resumed via the HyperFlex CLI.

To pause replication, complete the following steps:

1. Log in to the source HyperFlex cluster's management IP address via SSH as root.
2. At the command line, enter the command:

```
stcli dp schedule pause
```

To resume replication, complete the following steps:

3. Log in to the source HyperFlex cluster's management IP address via SSH as root.
4. At the command line, enter the command:

```
stcli dp schedule resume
```

Virtual Machine Recovery Operations

The snapshots taken by the HX Data Protection engine are separate from the HyperFlex native snapshots. Data Protection snapshots are triggered and tracked by the HX Data Platform software internally, and can only be used for the recovery of a virtual machine in the secondary, or target paired HX cluster. These snapshots are not visible in the snapshot manager of the VMware vSphere Web Client, the C# (thick) vSphere Client, or HTML5 vSphere Client, therefore they cannot be used to roll back a VM to an earlier state in the primary cluster location. In order to have the ability to roll back a VM to an earlier snapshot in the

primary, or source location, regular HX native snapshots must be scheduled on the VMs in addition to the Data Protection replication snapshots.

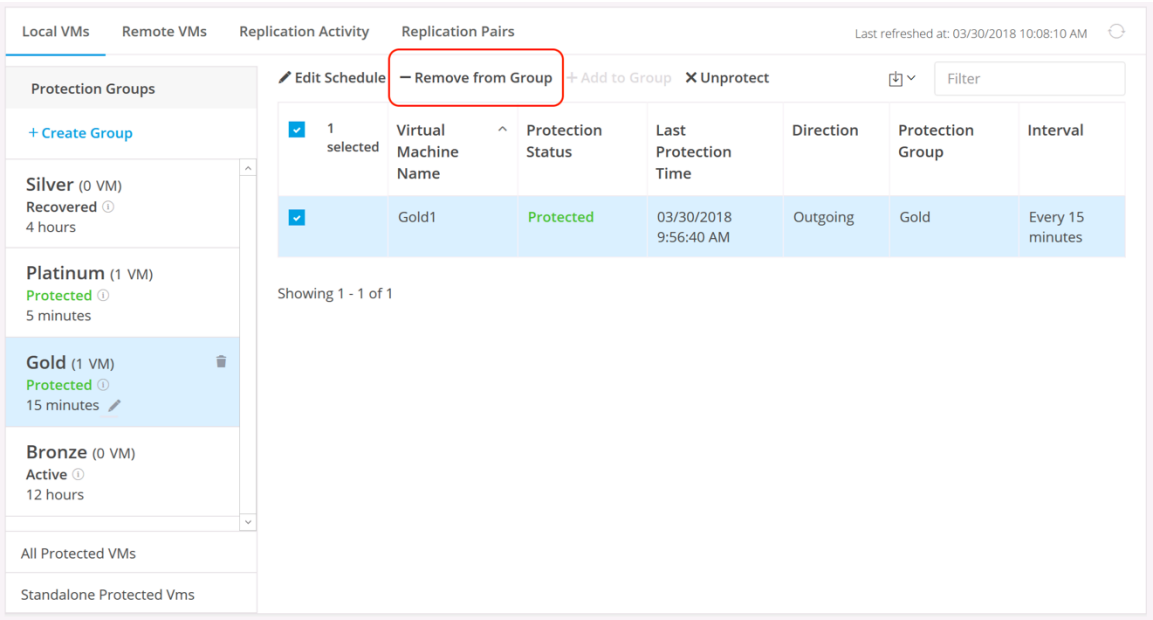
Virtual Machine Migration

When routine scheduled maintenance activities are required, or for other planned management purposes, virtual machines can be migrated from the source cluster to the target cluster. Migration of a virtual machine leaves the replication pairing between the two clusters in place, so that the VM can be protected again in the opposite direction of the original replication. As an overview of the process, a VM migration includes:

- Preparing the VM for migration, which shuts down the source VM and replicates the final changes.
- Removing the VM from a Protection Group if it is a member of one.
- Performing a recovery of the VM on the secondary, or target cluster.
- Power on the migrated VM and test its functionality.
- Preparing the VM for reverse protection, which unregisters the source VM from the original cluster, and transfers ownership of the VM to the cluster where it was migrated.
- Enabling reverse protection of the VM, which transfers a delta copy of changes from the recovered VM back to the source VM files, which still exist.

To perform a virtual machine migration, complete the following steps:

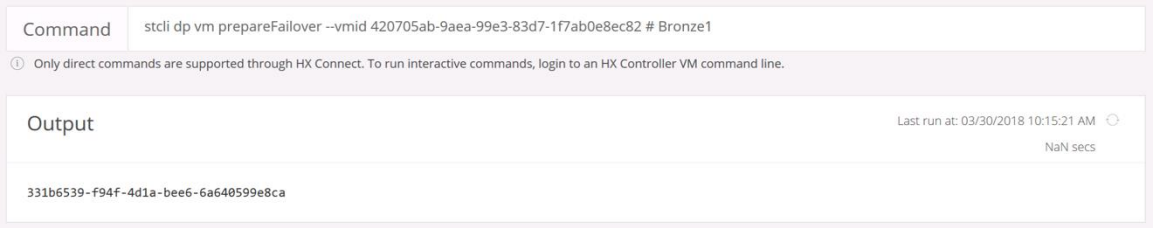
1. Log in to the primary, or source HyperFlex cluster's HyperFlex Connect webpage.
2. Click on Replication in the menu on the left.
3. Click on the Local VMs menu to display the VMs and Protection Groups that are being replicated from this cluster.
4. Locate the VM you wish to migrate, either in the Protection Groups list, All Protected VMs list, or Standalone Protected VMs list.
5. If the VM is a member of a Protection Group, you must first remove the VM from the group prior to performing the migration. Click on the checkbox next to the VM to be migrated, then click Remove from Group. If the VM is not in a group you may continue to step 7.



6. Click the Remove from Protection Group button on the alert window that appears. The VM will be protected according to the same schedule as it was in the group, however now it will be a standalone VM.

7. Click on Web CLI in the menu on the left.

8. Begin entering the command “stcli dp vm prepareFailover” into the Command field, the command should auto-complete with a list of possible VMs to pick from, including their names and VM IDs. Click the command for the VM you wish to prepare for failover.



9. Click Run to confirm the command to be run in the Web CLI. The command output will be shown in the Output section, which will list a job ID.

10. To verify the job has completed, enter the following command in the Web CLI command field:

    ```
    stcli dp vm hxtask --id <<job ID>> --vmid <<VM ID>>
    ```

    for example:

    ```
    stcli dp vm hxtask --vmid 420705ab-9aea-99e3-83d7-1f7ab0e8ec82 --id 331b6539-f94f-4d1a-bee6-6a640599e8ca
    ```

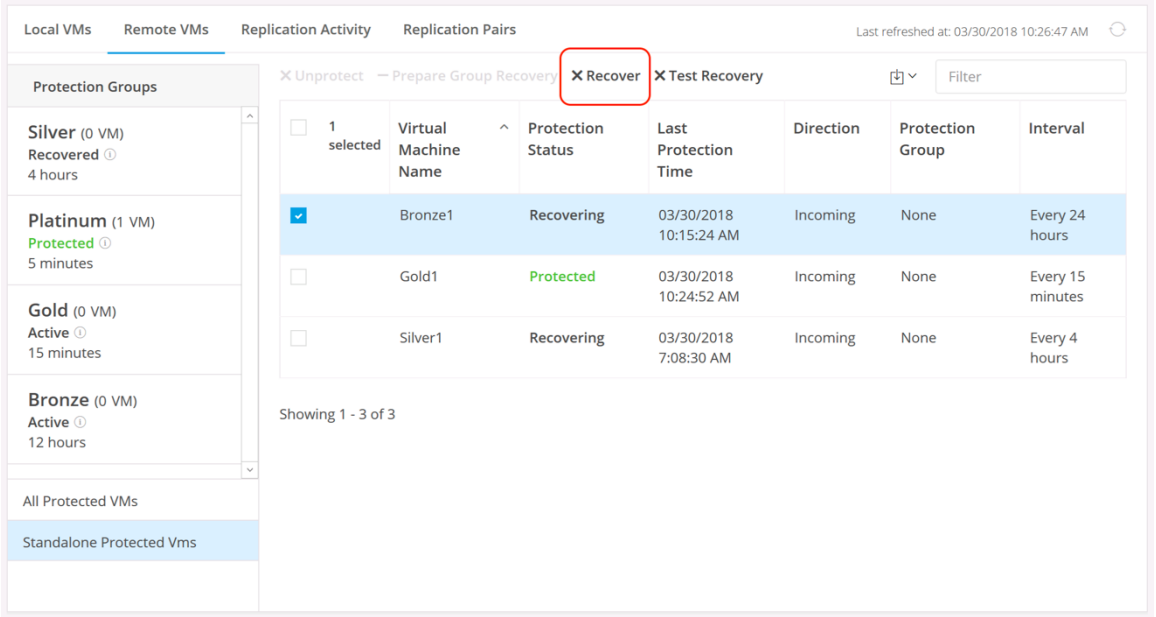
11. Verify in the Output that the job state shows “state: completed”

12. Log in to the secondary, or target HyperFlex cluster’s HyperFlex Connect webpage.

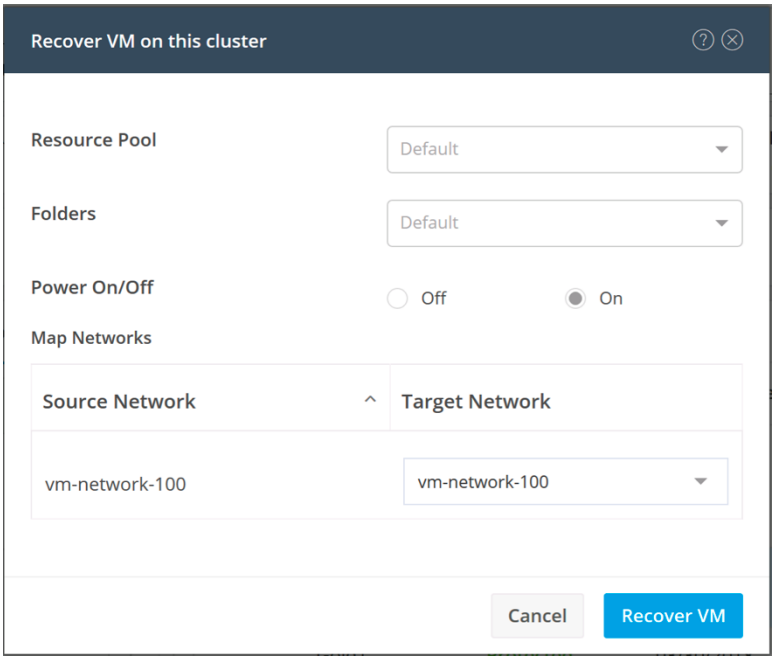
13. Click Replication in the menu on the left.

14. Verify the status of the Remote VM you wish to migrate is listed as “Recovering”.

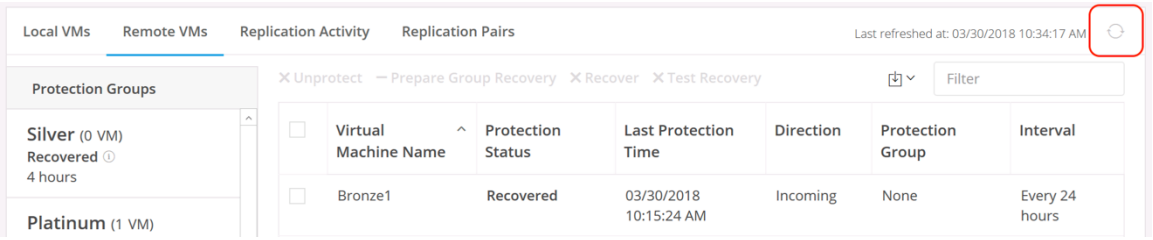
15. Click the checkbox next to the VM to be migrated, then click Recover.



16. Enter the recovery parameters for an existing Resource Pool or VM Folder to place the recovered VM into, if desired. Choose to power the VM on or not, and choose to map the source network to an existing network on the new cluster, then click Recover VM.



17. The recovery process will take a minute or two to complete, during that time the VM status will show “Recovering”. You may refresh the view, until the VM status shows “Recovered”.



18. Log in to the primary, or source HyperFlex cluster's HyperFlex Connect webpage.

19. Click Web CLI in the menu on the left.

20. Begin entering the command "stcli dp vm prepareReverseProtect" into the Command field, the command should auto-complete with a list of possible VMs to pick from, including their name and VM ID. Click the command for the VM you just migrated, to prepare that VM for protection again in the reverse direction of the original replication.

21. Click Run to confirm the command to be run in the Web CLI. The command output will be shown in the Output section, which will list a job ID.

22. To verify the job has completed, enter the following command in the Web CLI command field:

    ```
    stcli dp vm hxtask --id <<job ID>> --vmid <<VM ID>>
    ```

    for example:

    ```
    stcli dp vm hxtask --vmid 420705ab-9aea-99e3-83d7-1f7ab0e8ec82 --id 331b6539-f94f-4d1a-bee6-6a640599e8ca
    ```

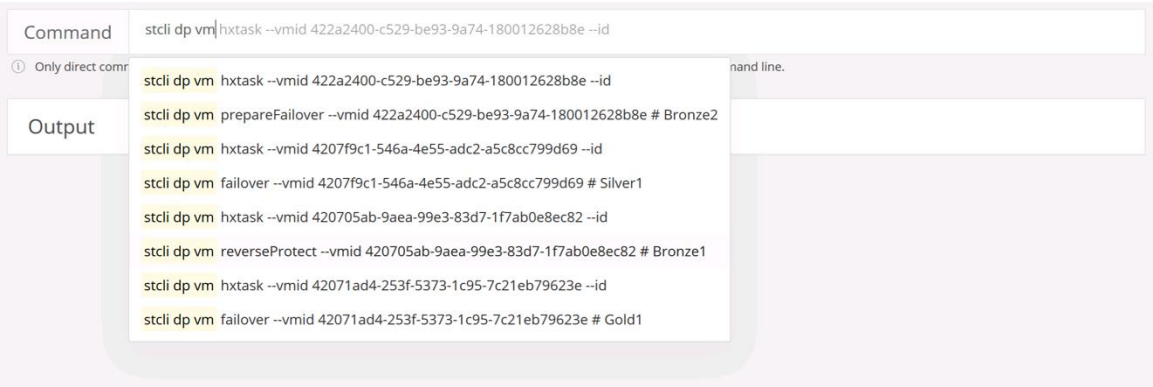
23. Verify in the Output that the job state shows "state: completed"

24. In the Replication menu, confirm that the recovered VM is listed as a Remote VM, and its status is "Protecting".

25. Log in to the secondary, or target HyperFlex cluster's HyperFlex Connect webpage.

26. Click Web CLI in the menu on the left.

27. Begin entering the command "stcli dp vm reverseProtect" into the Command field, the command should auto-complete with a list of possible VMs to pick from, including their name and VM ID. Click the command for the VM you just migrated, and wish to protect again in the reverse direction of the original replication.



28. Click Run to confirm the command to be run in the Web CLI. The command output will be shown in the Output section, which will list a job ID.

29. To verify the job has completed, enter the following command in the Web CLI command field:

    ```
    stcli dp vm hxtask --id <<job ID>> --vmid <<VM ID>>
    ```

    for example:

    ```
    stcli dp vm hxtask --vmid 420705ab-9aea-99e3-83d7-1f7ab0e8ec82 --id 331b6539-f94f-4d1a-bee6-6a640599e8ca
    ```

30. Verify in the Output that the job state shows “state: completed”.

31. In the Replication menu, confirm that the recovered VM is listed as a Local VM, and its status is “Protected”. If the recovery operation has taken longer than the configured replication interval. A warning status of “Exceeds Interval” may be displayed. Once a new replication task completes, the status should return to normal.

32. Perform any necessary post-recovery tasks on the VM, such as changing IP addresses, or updating DNS records, in order to make the VM and its applications available on the network.

33. Repeat steps 5 through 32 for each VM you wish to migrate to the other cluster.

Virtual Machine Recovery Testing

A virtual machine recovery test can be conducted to verify that recovery of a VM can be completed successfully. The recovery test does not cause any interruption to the ongoing replication of the VM, nor does it break the replication pairing between the two clusters. VMs in Protection Groups can be tested at any time and do not have to be removed from the group.

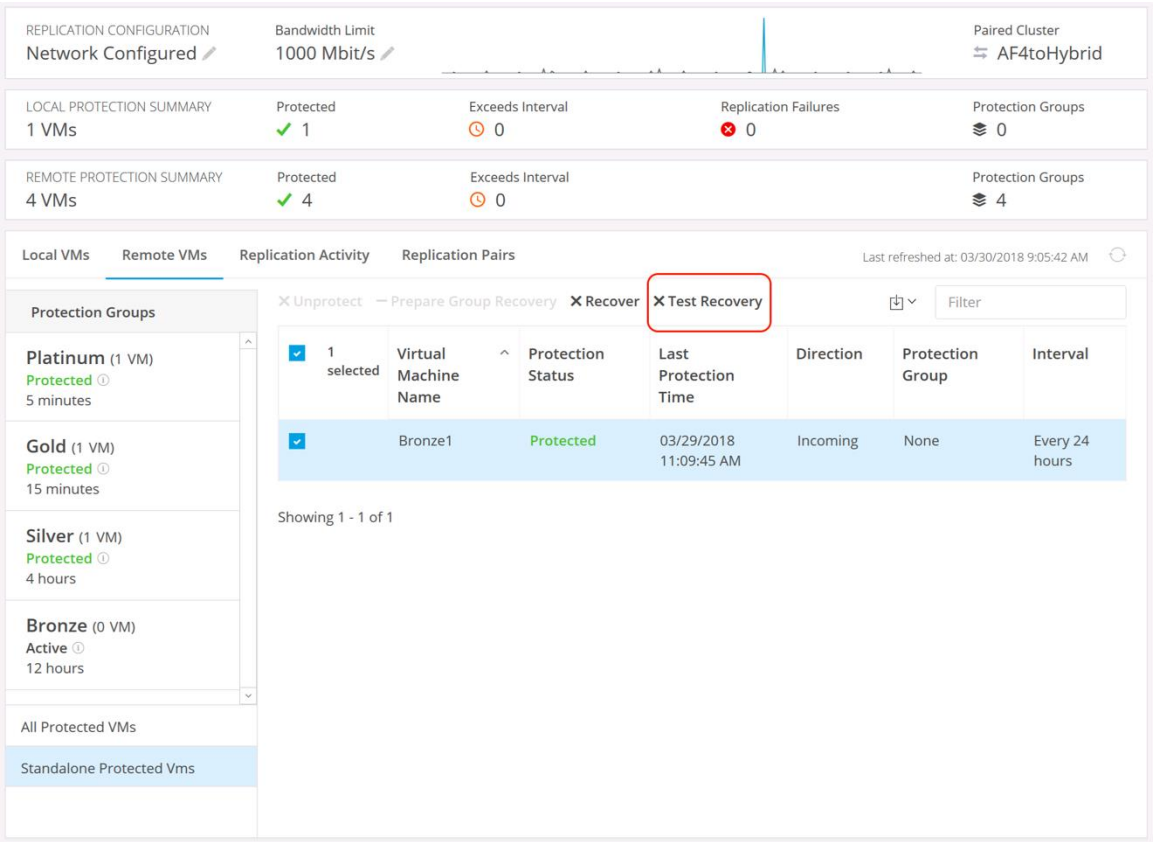
To perform a virtual machine recovery test, complete the following steps:

1. Log in to the secondary, or target HyperFlex cluster’s HyperFlex Connect webpage.

2. Click Replication in the menu on the left.

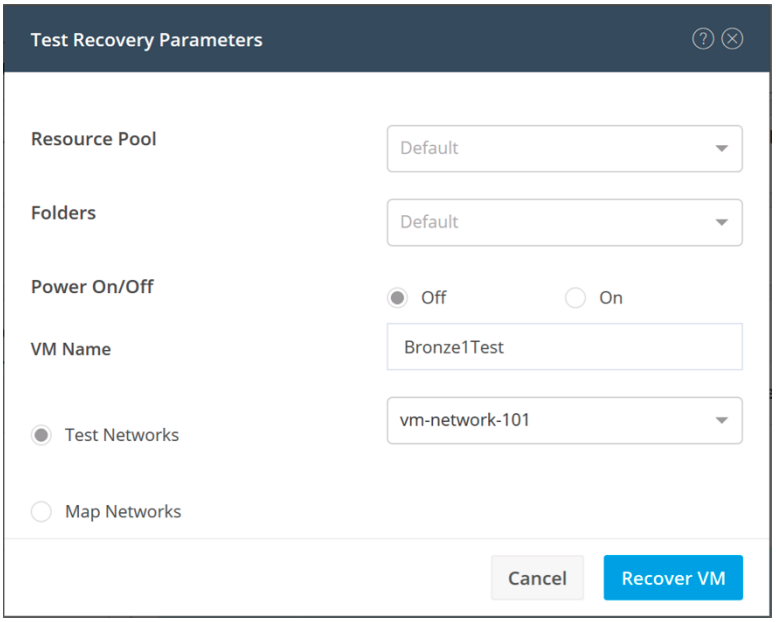
3. Click the Remote VMs menu to display the VMs and Protection Groups that are being replicated to this cluster.

4. Locate the VM you wish to test recovery of, either in the Protection Groups list, All Protected VMs list, or Standalone Protected VMs list. Click the checkbox next to the VM to select it, then click Test Recovery.



5. Enter the recovery parameters for an existing Resource Pool or VM Folder to place the test recovery VM into, if desired. If you do not make a choice the VM will be recovered to a newly created VM folder named HxRecoveryTest, and not placed into a Resource Pool.

6. Choose to power the VM on or not, provide a name for the VM, and choose to either connect the VM to a specific test network or map it to an existing network on the new cluster, then click Recover VM.



7. Once the job completes, verify the VM has been recovered to the HxRecoveryTest folder, or the folder you designated.

8. Power on the recovered VMs via the vSphere Web Client or the HTML5 vSphere Client to test their functionality.

9. Repeat steps 4 through 8 for each VM you wish to test.

10. Once the testing is completed, you may delete the test recovery VMs.

Virtual Machine Disaster Recovery

In the case of a site outage, or the failure of a cluster, VMs can be recovered to their state as of the last successfully transmitted snapshot, running on the secondary, or target cluster as part of a disaster recovery operation. The recovery operation described assumes that the primary, or source site and cluster is either offline, or isolated in such a way that it can no longer communicate with the secondary, or target site, nor can it be managed. A recovery operation stops all replication between the two clusters, so that replication can be reestablished at a later time after the faults or outages have been repaired. As an overview of the process, a VM disaster recovery includes:

- Removing the VMs from a Protection Group if they are a member of one.
- Performing a recovery of the VMs on the secondary, or target cluster.
- Power on the migrated VMs and test their functionality.

To perform a virtual machine migration, complete the following steps:

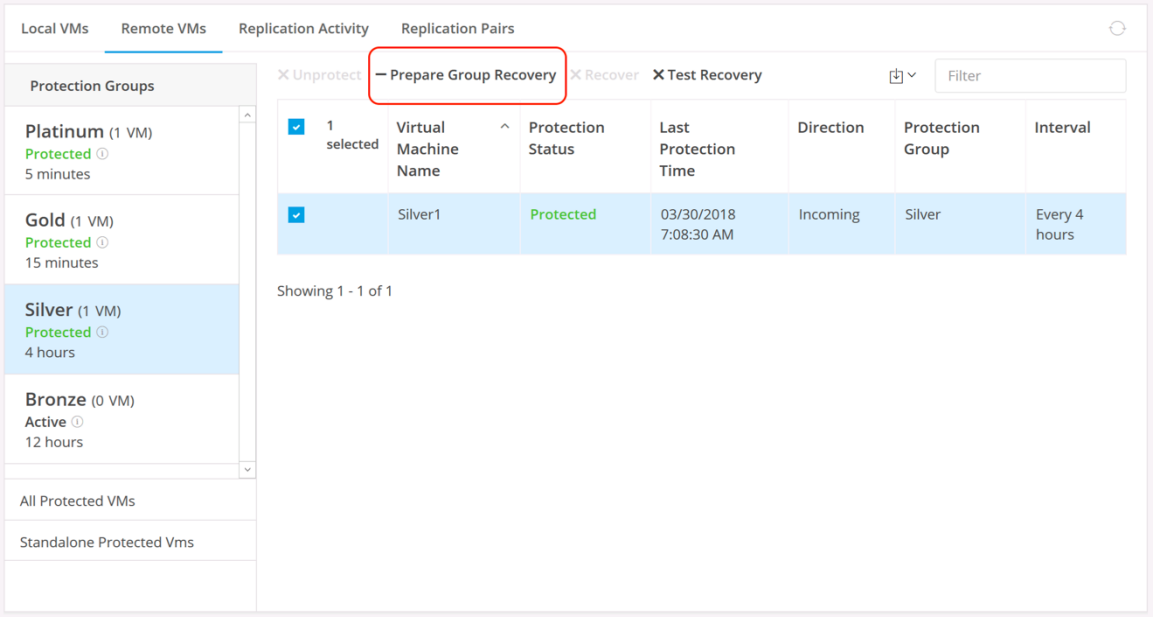
1. Log in to the secondary, or target HyperFlex cluster's HyperFlex Connect webpage.

2. Click Replication in the menu on the left.

3. Click the Remote VMs menu to display the VMs and Protection Groups that are being replicated to this cluster.

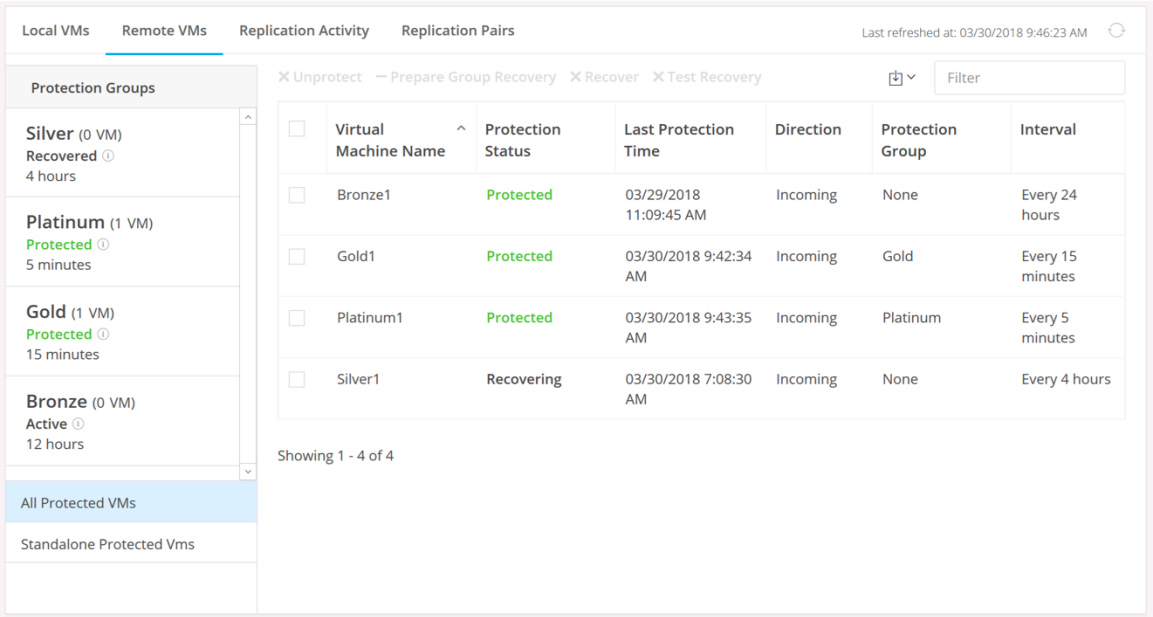
4. Locate the VM you wish to test recovery of, either in the Protection Groups list, All Protected VMs list, or Standalone Protected VMs list.

5. If the VM is a member of a Protection Group, click the checkbox next to the VM to select it, then click Prepare Group Recovery. If the VM is not in a group you may continue to step 7.



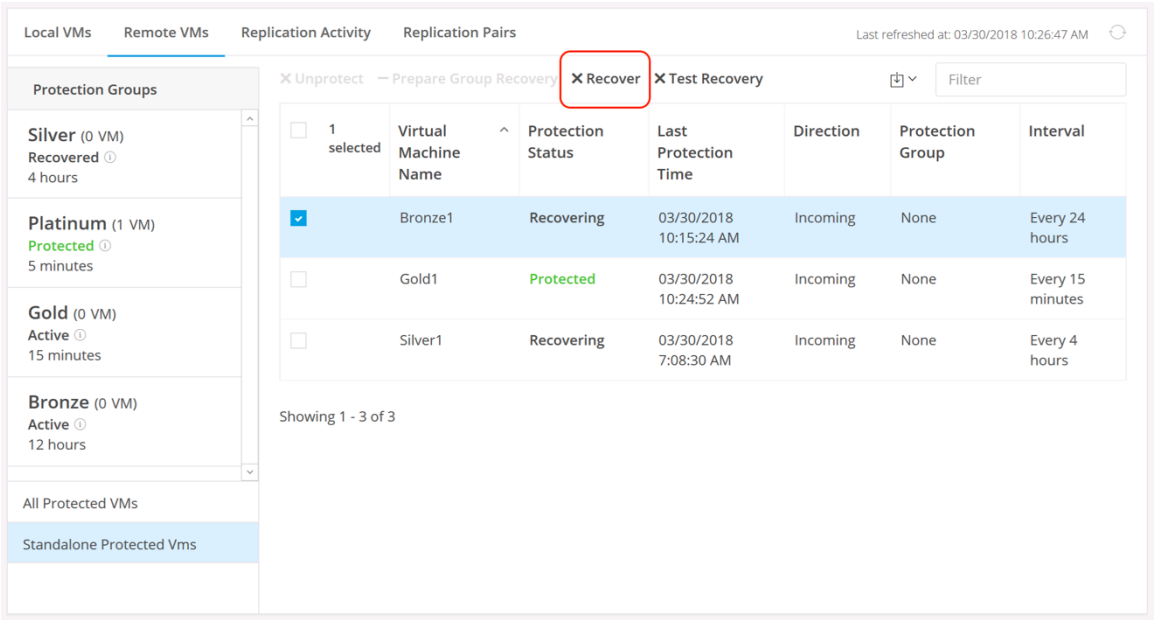
WARNING! If a Protection Group is prepared for recovery, all VMs in that group must be migrated or recovered. There is no way to resume replication of a Protection Group once it has been marked as recovered. If a single VM needs to be migrated, and it is part of a Protection Group, the VM must be removed from the group and protected individually before attempting to migrate or recover the VM.

6. Click the Prepare button on the alert windows that appears. Preparing the group for recovery removes the VMs from the group, and halts all replication, effectively making all the VMs into standalone replicas. The process may take a couple of minutes to be reflected in the destination cluster. Once the process is completed, the Protection Group will show as “Recovered” with no VMs in it, and the VMs that were in the group will show a status of “Recovering”.

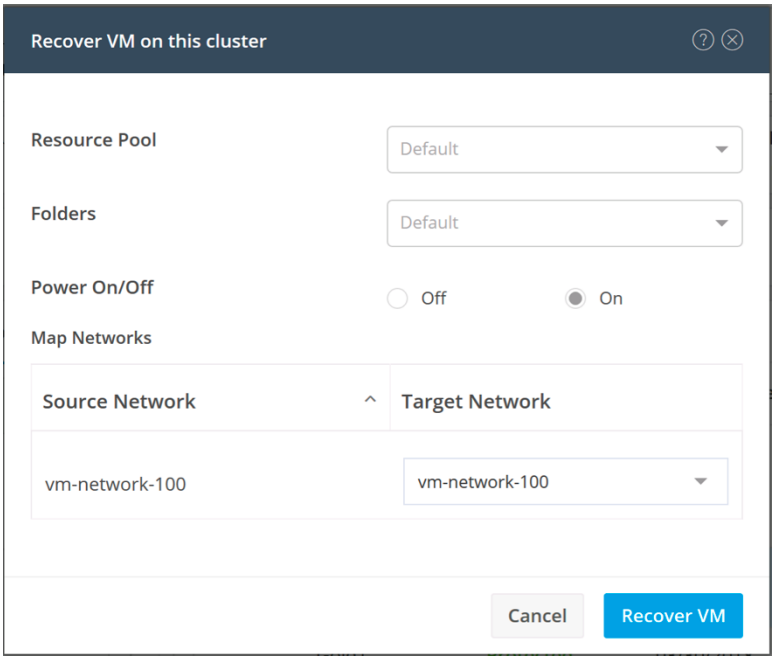


7. Verify the status of the VMs you removed from any Protection Groups are as “Recovering”. Standalone VMs will appear as “Protected”.

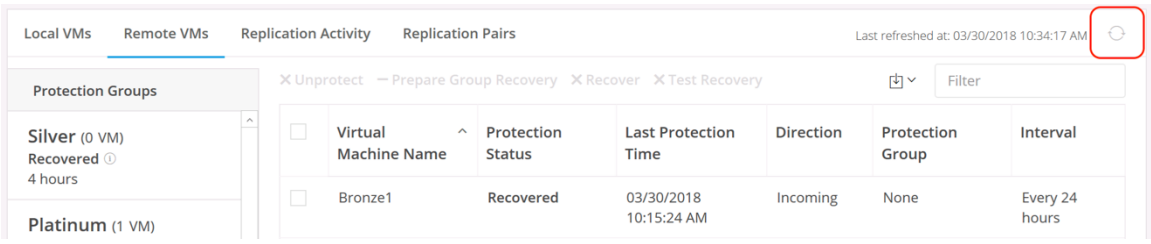
8. Click the checkbox next to the VM to be recovered, then click Recover.



9. Enter the recovery parameters for an existing Resource Pool or VM Folder to place the recovered VM into, if desired. Choose to power the VM on or not, and choose to map the source network to an existing network on the new cluster, then click Recover VM.



10. The recovery process will take a minute or two to complete, during that time the VM status will show “Recovering”. You may refresh the view, until the VM status shows “Recovered”.



Note: Parallel recovery operations of up to 20 VMs can be done as long as each VM is individually protected, or the VM is a member of a separate Protection Group. Parallel recovery of multiple VMs that are in the same Protection Group cannot be done.

11. Perform any post-migration tasks that may be necessary to the VMs, such as IP addressing changes or other operations that are necessary to return the VM and its applications to normal service.

12. Repeat steps 4 through 11 for each VM you need to recover in the secondary cluster.

Disaster Recover Post Operations

After a disaster has been declared, and all VMs have been recovered to the secondary, or target cluster, work can begin to repair the primary, or source cluster. As an overview of the process, post-recovery operations include:

- Repairing or bringing the original cluster back online.
- Preparing the VMs for reverse protection, which unregisters the source VMs from the original cluster, and transfers ownership of the VMs to the cluster where they were recovered.
- Enabling reverse protection of the VM, which transfers a delta copy of changes from the recovered VM back to the source VM files, which still exist.
- If desired, perform VM migrations to move the VMs back to the original cluster where they ran.
- Delete and recreate the Protection Groups, and move the VMs back into the groups.

To complete disaster recovery post operation steps on the original source cluster, complete the following steps:

1. Repair the faults or failures in the original source HX cluster, and bring the cluster online in a healthy state. If for any reason the original cluster had to be reinstalled, these steps can all be skipped, and a new replication pairing, and VM protections must be established.

2. Log in to the original primary, or source HyperFlex cluster's HyperFlex Connect webpage after the repairs are completed.

3. In the Replication menu, confirm that the recovered VMs are listed as a Local VMs, and their status is "Recovered".

4. Click on Web CLI in the menu on the left.

5. Begin entering the command "stcli dp vm prepareReverseProtect" into the Command field, the command should auto-complete with a list of possible VMs to pick from, including their name and VM ID. Click the

command for a VM you recovered, and wish to prepare for protection again in the reverse direction of the original protection.

6. Click Run to confirm the command to be run in the Web CLI. The command output will be shown in the Output section, which will list a job ID.

7. To verify the job has completed, enter the following command in the Web CLI command field:

   ```
   stcli dp vm hxtask --id <<job ID>> --vmid <<VM ID>>
   ```

   for example:

   ```
   stcli dp vm hxtask --vmid 420705ab-9aea-99e3-83d7-1f7ab0e8ec82 --id 331b6539-f94f-4d1a-bee6-6a640599e8ca
   ```

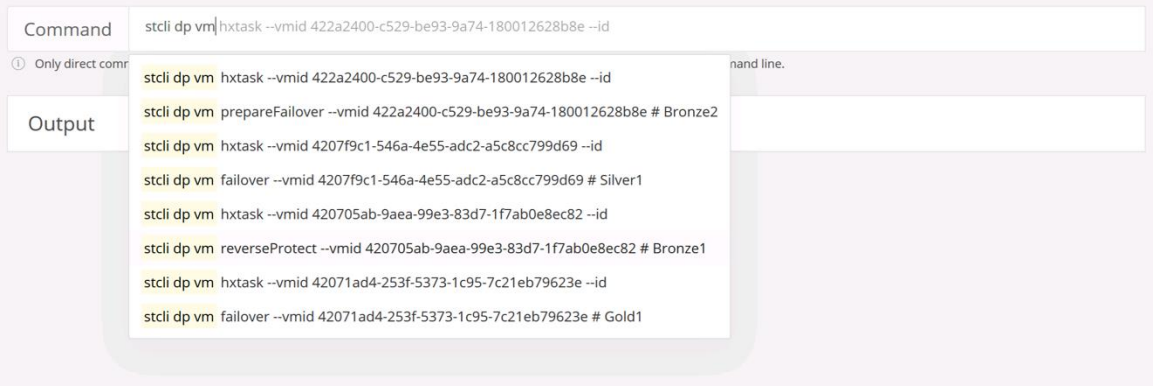
8. Verify in the Output that the job state shows “state: completed”

9. In the Replication menu, confirm that the recovered VM is listed as a Remote VM, and its status is “Protecting”.

10. Log in to the secondary, or target HyperFlex cluster’s HyperFlex Connect webpage.

11. Click Web CLI in the menu on the left.

12. Begin entering the command “stcli dp vm reverseProtect” into the Command field, the command should auto-complete with a list of possible VMs to pick from, including their name and VM ID. Click the command for the VM you just migrated, and wish to protect again in the reverse direction of the original replication.



13. Click Run to confirm the command to be run in the Web CLI. The command output will be shown in the Output section, which will list a job ID.

14. To verify the job has completed, enter the following command in the Web CLI command field:

```
stcli dp vm hxtask --id <<job ID>> --vmid <<VM ID>>
```

for example:

```
stcli dp vm hxtask --vmid 420705ab-9aea-99e3-83d7-1f7ab0e8ec82 --id 331b6539-f94f-4d1a-
bee6-6a640599e8ca
```

15. Verify in the Output that the job state shows “state: completed”.

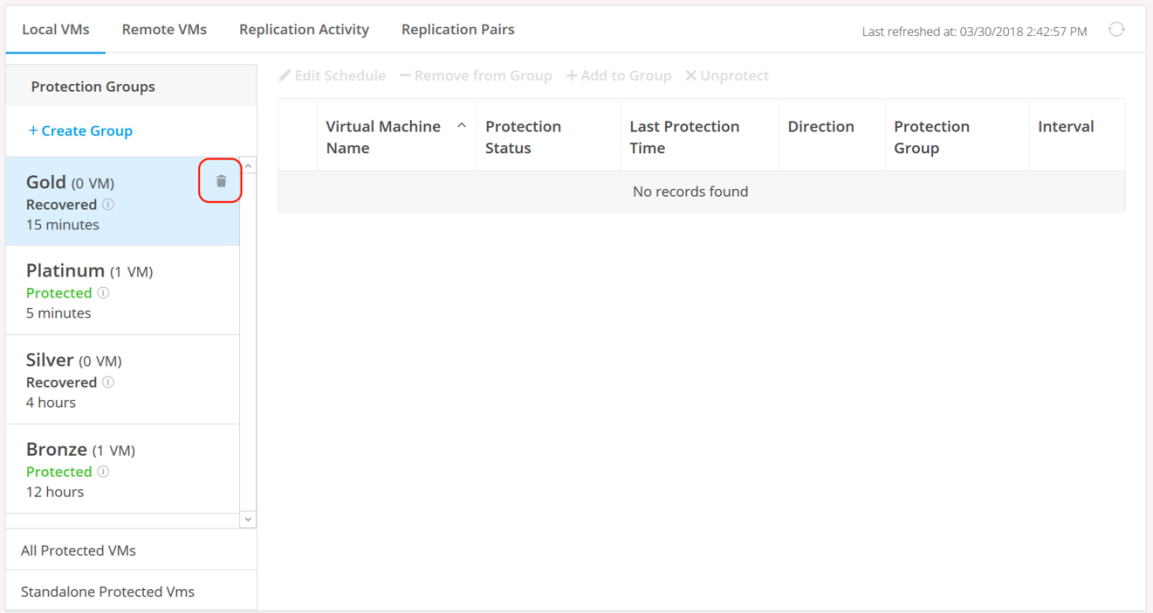
16. In the Replication menu, confirm that the recovered VM is listed as a Local VM, and its status is “Protected”. If the recovery operation has taken longer than the configured replication interval. A warning status of “Exceeds Interval” may be displayed. Once a new replication task completes, the status should return to normal.

17. Perform any necessary post-recovery tasks on the VM, such as changing IP addresses, or updating DNS records, in order to make the VM and its applications available on the network.

18. Repeat steps 5 through 17 for each VM that was recovered as part of the disaster recovery process.

19. Once all recovered VMs are configured for reverse protection, a new migration task can be started to return the VMs back to their original running location.

20. After the VMs are migrated back to their original running cluster, if an entire Protection Group was migrated, the Protection Group status will show as Recovered. The Protection Group must be deleted, as it is no longer possible to add VMs to a recovered group, nor is it possible to make the group active again. From the HyperFlex Connect Replication page of the primary, or source cluster, view the Local VMs list and find the Protection Groups that were recovered. Click the small trash can icon to delete the group, and click the Delete button on the warning prompt that appears.



21. The Protection groups can now be recreated and the VMs added back to them as was done originally.

Validation

This section provides a list of items that should be reviewed after the HyperFlex system has been deployed and configured. The goal of this section is to verify the configuration and functionality of the solution, and ensure that the configuration supports core availability requirements.

Post Install Checklist

The following tests are critical to functionality of the solution, and should be verified before deploying for production:

1. Verify the expected number of converged storage nodes and compute-only nodes are members of the HyperFlex cluster in the vSphere Web Client plugin manage cluster screen.

2. Verify the expected cluster capacity is seen in the vSphere Web Client plugin summary screen. (See Appendix A)

3. Create a test virtual machine that accesses the HyperFlex datastore and is able to perform read/write operations.

4. Perform the virtual machine migration (vMotion) of the test virtual machine to a different host on the cluster.

5. During the vMotion of the virtual machine, make sure the test virtual machine can perform a continuous ping to default gateway and to check if the network connectivity is maintained during and after the migration.

Verify Redundancy

The following redundancy checks can be performed to verify the robustness of the system. Network traffic, such as a continuous ping from VM to VM, or from vCenter to the ESXi hosts should not show significant failures (one or two ping drops might be observed at times). Also, all of the HyperFlex datastores must remain mounted and accessible from all the hosts at all times.

1. Administratively disable one of the server ports on Fabric Interconnect A which is connected to one of the HyperFlex converged storage hosts. The ESXi virtual switch uplinks for fabric A should now show as failed, and the standby uplinks on fabric B will be in use for the management and vMotion virtual switches. Upon administratively re-enabling the port, the uplinks in use should return to normal.

2. Administratively disable one of the server ports on Fabric Interconnect B which is connected to one of the HyperFlex converged storage hosts. The ESXi virtual switch uplinks for fabric B should now show as failed, and the standby uplinks on fabric A will be in use for the storage virtual switch. Upon administratively re-enabling the port, the uplinks in use should return to normal.

3. Place a representative load of guest virtual machines on the system. Put one of the ESXi hosts in maintenance mode, using the HyperFlex HX maintenance mode option. All the VMs running on that host should be migrated via vMotion to other active hosts through vSphere DRS, except for the storage platform controller VM, which will be powered off. No guest VMs should lose any network or storage accessibility during or after the migration. This test assumes that enough RAM is available on the remaining ESXi hosts to accommodate VMs from the host put in maintenance mode. The HyperFlex cluster will show in an unhealthy state.

4. Reboot the host that is in maintenance mode, and exit it from maintenance mode after the reboot. The storage platform controller will automatically start when the host exits maintenance mode. The HyperFlex cluster will show as healthy after a brief time to restart the services on that node. VSphere DRS should rebalance the VM distribution across the cluster over time.

Note: Many vCenter alerts automatically clear when the fault has been resolved. Once the cluster health is verified, some alerts may need to be manually cleared.

5. Reboot one of the two Cisco UCS Fabric Interconnects while traffic is being sent and received on the storage datastores and the network. The reboot should not affect the proper operation of storage access and network traffic generated by the VMs. Numerous faults and errors will be noted in Cisco UCS Manager, but all will be cleared after the FI comes back online.

Appendix

A: Cluster Capacity Calculations

A HyperFlex HX Data Platform cluster capacity is calculated as follows:

(((<capacity disk size in GB> X 10^9) / 1024^3) X <number of capacity disks per node> X <number of HyperFlex nodes> X 0.92) / replication factor

Divide the result by 1024 to get a value in TiB

The replication factor value is 3 if the HX cluster is set to RF=3, and the value is 2 if the HX cluster is set to RF=2.

The 0.92 multiplier accounts for an 8% reservation set aside on each disk by the HX Data Platform software for various internal filesystem functions.

Calculation example:

<capacity disk size in GB> = 1200 for 1.2 TB disks

<number of capacity disks per node> = 15 for an HX240c-M4SX model server

<number of HyperFlex nodes> = 8

replication factor = 3

Result: (((1200*10^9)/1024^3)*15*8*0.92)/3 = 41127.2049

41127.2049 / 1024 = 40.16 TiB

A stretched cluster maintains data identically across both halves of the cluster, therefore it effectively doubles the replication factor. For example, the only allowed replication factor for a stretched cluster is RF2, meaning it will store 2 copies of the data on the nodes in site 1, and also store 2 copies of the data on the nodes in site 2. Because of this, the capacity of a stretched cluster is effectively reduced by 50 percent compared to RF2. The calculation above can use a value of 4 for the replication factor to determine the capacity of a stretched cluster.

B: HyperFlex Sizer

HyperFlex sizer is a cloud based end-to-end tool that can help the customers and partners find out how many Cisco HyperFlex nodes are needed and how the nodes can be configured to meet their needs for the compute resources, storage capacity and performance requirements in the datacenter. The sizing guidance of the HX system is calculated according to the information of workloads collected from the users. This cloud application can be accessed from anywhere at Cisco website (CCO login required):

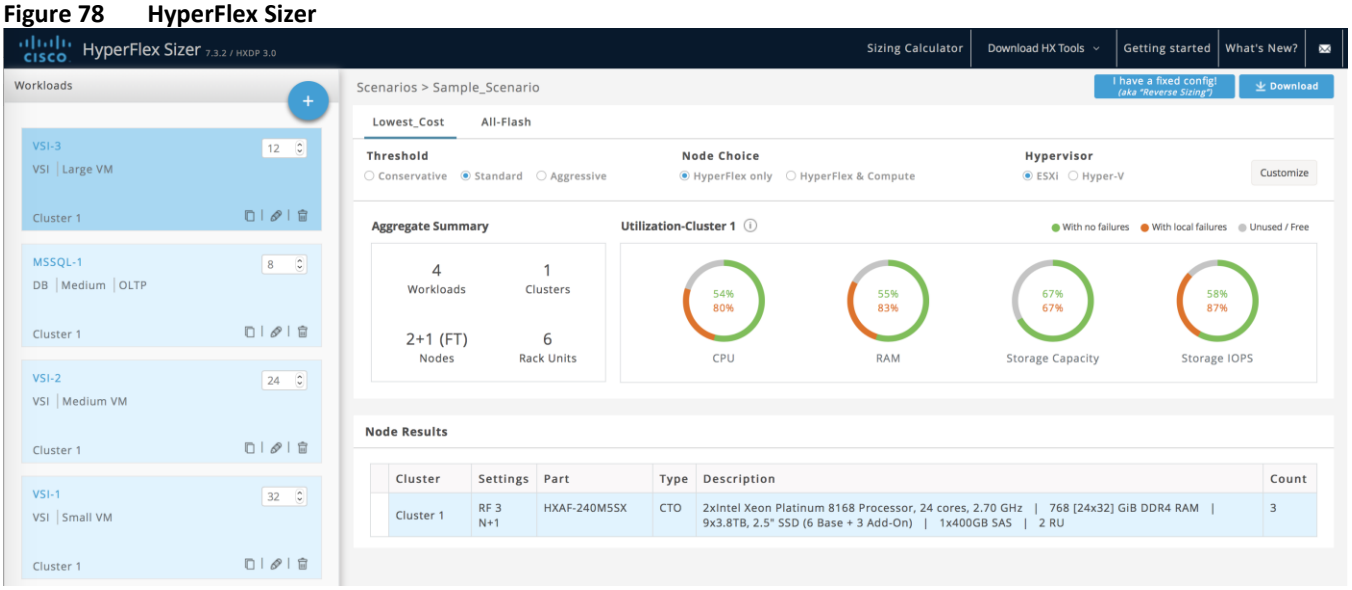
https://hyperflexsizer.cloudapps.cisco.com

Improvements in the sizing tool for HXDP 3.0 release include:

- Support for Hyper-V hypervisors
- Support for stretched clusters
- Support for 64 node clusters

- Support for Large-Form-Factor disks (LFF)
- Support for 1.8 TB drives in hybrid nodes
- Support for Intel Optane NVMe caching drives in M5 all-flash nodes
- Support for Microsoft Exchange server workloads

Figure 78 HyperFlex Sizer



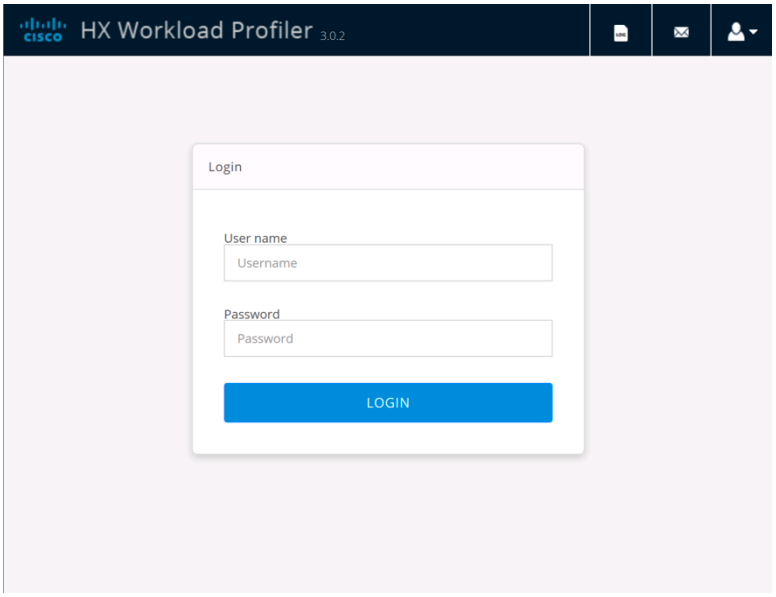
Note: The HyperFlex Sizer tool is designed to provide general guidance in evaluating the optimum solution for using selected Cisco products. The tool is not intended as a substitute for your own judgment or for that of your professional advisors.

C: HyperFlex Workload Profiler

Also available at the https://hyperflexsizer.cloudapps.cisco.com website is an updated HyperFlex Workload Profiler, version 3.0. The HyperFlex Workload Profiler tool is used to capture storage usage and performance statistics from an existing VMware ESX cluster, enabling you to use that data to assist with sizing a HyperFlex cluster which would assume that workload. The workload profiler is distributed as an OVA file, which can be deployed using static or DHCP assigned addressing, on an existing VMware ESXi host. Once deployed, the profiler tool connects to an existing VMware vCenter server to gather storage statistics for the selected ESXi hosts. To capture performance data using the HyperFlex Workload Profiler, complete the following steps:

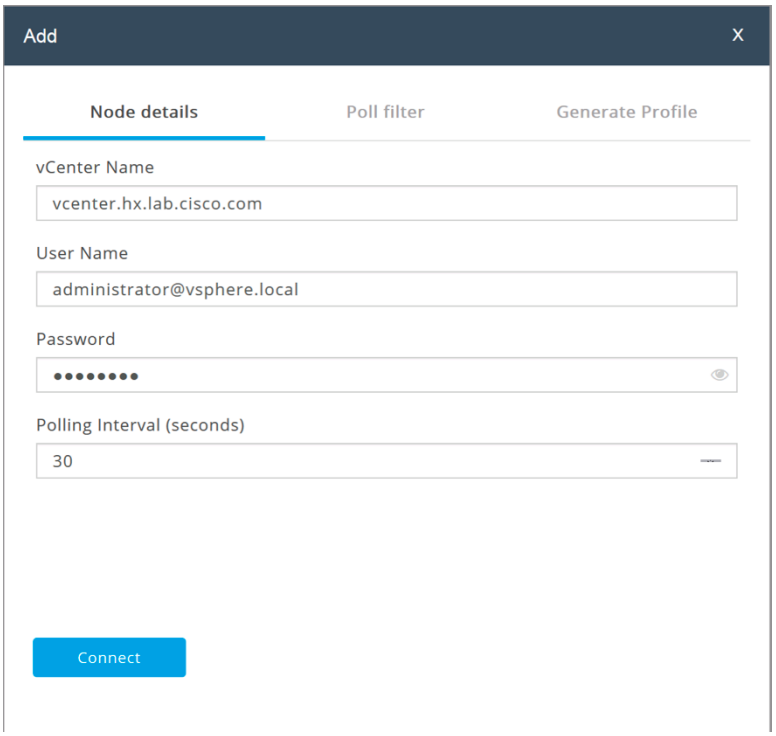
1. Deploy the HyperFlex Workload Profiler VM by using the Deploy OVF Template wizard, on the chosen existing ESXi host. Assign a static IP address or use DHCP addressing as part of the deployment wizard, you may optionally change the default password.

2. Using a web browser, navigate to the IP address assigned or leased by the Workload Profiler VM.



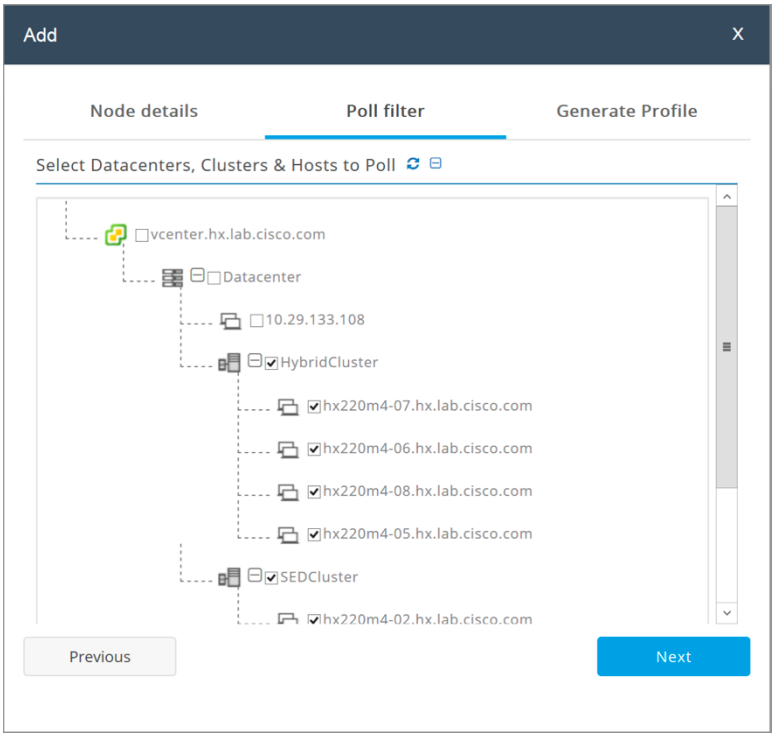
3. Enter the username and password, the default username and password is “monitoring”, or use the password previously entered, then click Login.

4. On first login, a wizard to add a system to be monitored will run. Enter the vCenter server name or IP, a username with administrative rights, and the password, then click Connect.



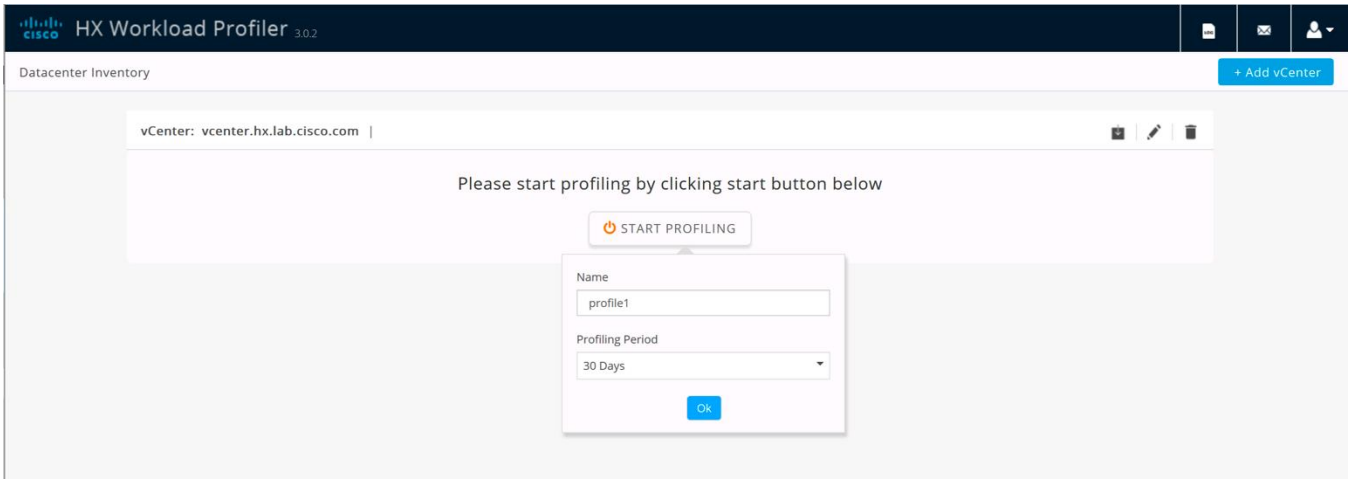
5. When the vCenter server is connected, click Next to select the hosts to monitor.

6. Check the box or boxes next to the hosts to poll for data, then click Next.



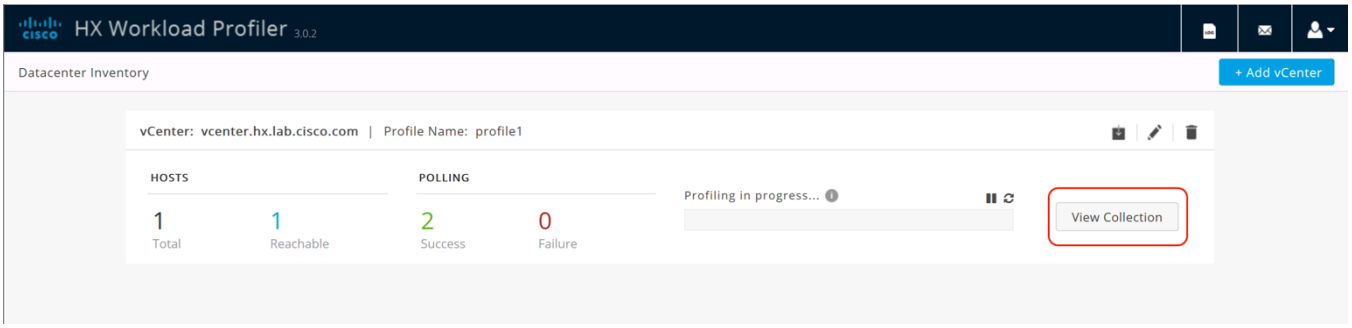
7. Choose to generate a Quick Profile, which will not generate detailed performance data, or a Detailed Profile, then click Save.

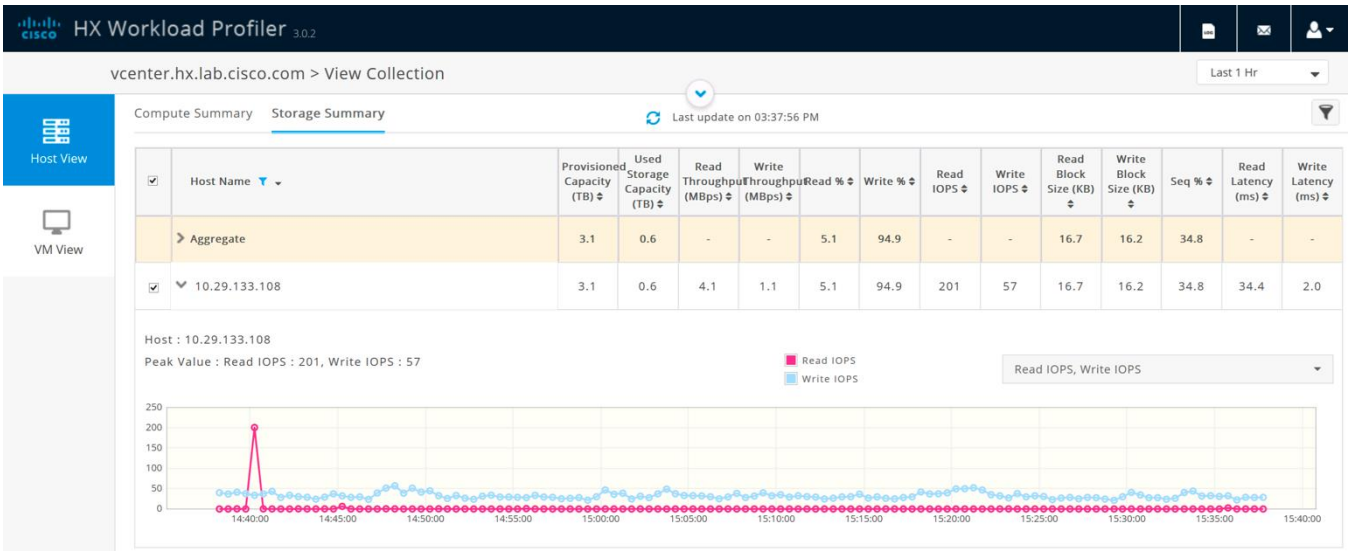
8. In the main screen, the vCenter server being polled will be listed. Click the Start Profiling button.



9. Choose a time interval to collect data on the system, then click OK. A 30-day collection is recommended for accurate sizing activities.

10. At any time during the collection polling, the data can be viewed by clicking on the View Collection button. The data for CPU and memory utilization, and storage statistics can be viewed, as an aggregate of all hosts, one host at a time, or from a per VM perspective.





11. When the collection is complete, the complete dataset can be exported as a comma-separated file, and the data can be automatically imported into the HyperFlex sizer tool to help with computing and storage sizing efforts, or otherwise analyzed to help with sizing decisions.

D: Example Cisco Nexus 9372 Switch Configurations

Switch A

```
hostname HX-9K-A

no feature telnet
no telnet server enable
cfs eth distribute
feature interface-vlan
feature lacp
feature vpc

ip domain-lookup
ip domain-list cisco.com
ip name-server 171.70.168.183 173.36.131.10
logging event link-status default
policy-map type network-qos jumbo
  class type network-qos class-default
    mtu 9216
system qos
  service-policy type network-qos jumbo
clock timezone PST -8 0
```

```
clock summer-time PST
ntp server 171.68.38.65
ntp server 171.68.38.66

vrf context management
vlan 1
vlan 133
  name Management
vlan 51
  name HXCluster1
vlan 100
  name VM-Prod-100
vlan 200
  name VMotion

cdp enable

vpc domain 50
  role priority 10
  peer-keepalive destination 10.29.133.102 source 10.29.133.101
  auto-recovery
  delay restore 150

interface Vlan1

interface port-channel50
  description VPC-Peer
  switchport mode trunk
  switchport trunk allowed vlan 1,51,100,133,200
  spanning-tree port type network
  vpc peer-link

interface port-channel10
  description VPC to 6248-A
  switchport mode trunk
  switchport trunk allowed vlan 51,100,133,200
  spanning-tree port type edge trunk
  spanning-tree bpduguard enable
  mtu 9216
  vpc 10

interface port-channel20
  description VPC to 6248-B
  switchport mode trunk
  switchport trunk allowed vlan 51,100,133,200
  spanning-tree port type edge trunk
  spanning-tree bpduguard enable
  mtu 9216
  vpc 20

interface Ethernet1/1
  description uplink
  switchport mode trunk
  switchport trunk allowed vlan 51,100,133,200
  spanning-tree port type network

interface Ethernet1/2
  description NX9372-A_P1/2--UCS6248-A_2/1
  switchport mode trunk
  switchport trunk allowed vlan 51,100,133,200
  channel-group 10 mode active

interface Ethernet1/4
  description NX9372-A_P1/4--UCS6248-B_2/1
```

```
  switchport mode trunk
  switchport trunk allowed vlan 51,100,133,200
  channel-group 20 mode active

interface Ethernet1/47
  description NX9372-A_P1/47--NX9372-B_P1/47
  switchport mode trunk
  switchport trunk allowed vlan 1,51,100,133,200
  channel-group 50 mode active

interface Ethernet1/48
  description NX9372-A_P1/48--NX9372-B_P1/48
  switchport mode trunk
  switchport trunk allowed vlan 1,51,100,133,200
  channel-group 50 mode active

interface mgmt0
  ip address 10.29.133.101/24

vrf context management
ip route 0.0.0.0/0 10.29.133.1
```

Switch B

```
hostname HX-9K-B

no feature telnet
no telnet server enable
cfs eth distribute
feature interface-vlan
feature lacp
feature vpc

ip domain-lookup
ip domain-list cisco.com
ip name-server 171.70.168.183 173.36.131.10
logging event link-status default
policy-map type network-qos jumbo
  class type network-qos class-default
    mtu 9216
system qos
  service-policy type network-qos jumbo
clock timezone PST -8 0
clock summer-time PST
ntp server 171.68.38.65
ntp server 171.68.38.66

vrf context management
vlan 1
vlan 133
  name Management
vlan 51
  name HXCluster1
vlan 100
  name VM-Prod-100
vlan 200
  name VMotion

cdp enable

vpc domain 50
  role priority 10
  peer-keepalive destination 10.29.133.101 source 10.29.133.102
```

```
  auto-recovery
  delay restore 150

interface Vlan1

interface port-channel50
  description VPC-Peer
  switchport mode trunk
  switchport trunk allowed vlan 1,51,100,133,200
  spanning-tree port type network
  vpc peer-link

interface port-channel10
  description VPC to 6248-A
  switchport mode trunk
  switchport trunk allowed vlan 51,100,133,200
  spanning-tree port type edge trunk
  spanning-tree bpduguard enable
  mtu 9216
  vpc 10

interface port-channel20
  description VPC to 6248-B
  switchport mode trunk
  switchport trunk allowed vlan 51,100,133,200
  spanning-tree port type edge trunk
  spanning-tree bpduguard enable
  mtu 9216
  vpc 20

interface Ethernet1/1
  description uplink
  switchport mode trunk
  switchport trunk allowed vlan 51,100,133,200
  spanning-tree port type network

interface Ethernet1/2
  description NX9372-A_P1/2--UCS6248-A_2/3
  switchport mode trunk
  switchport trunk allowed vlan 51,100,133,200
  channel-group 10 mode active

interface Ethernet1/4
  description NX9372-A_P1/4--UCS6248-B_2/3
  switchport mode trunk
  switchport trunk allowed vlan 51,100,133,200
  channel-group 20 mode active

interface Ethernet1/47
  description NX9372-B_P1/47--NX9372-A_P1/47
  switchport mode trunk
  switchport trunk allowed vlan 1,51,100,133,200
  channel-group 50 mode active

interface Ethernet1/48
  description NX9372-B_P1/48--NX9372-A_P1/48
  switchport mode trunk
  switchport trunk allowed vlan 1,51,100,133,200
  channel-group 50 mode active

interface mgmt0
  ip address 10.29.133.102/24

vrf context management
```

```
ip route 0.0.0.0/0 10.29.133.1
```

E: Example Connecting to External Storage Systems

The following examples demonstrate scenarios where a newly built HX cluster connects to the existing third-party storage devices, using either iSCSI or FC protocols. The new HX system is built with its own Fabric Interconnect switches then connecting to upstream Ethernet switches or Fibre Channel switches where the existing storage devices reside.

Connecting to iSCSI Storage

The HX installer can guide you through the process of setting up your HX cluster allowing you to connect to existing third-party storage systems via the iSCSI protocol. The installer will automatically configure Cisco UCS profiles, and HX cluster nodes with extra vNICs for iSCSI, and proper VLANs in the setup. The procedure is described here in this CVD. It is assumed that the third-party storage system is already configured per a Cisco Validated Design and all networking configuration is completed on the upstream switches. For iSCSI, the VLANs are configured on the A fabric and B fabric separately, as per best practice. In this example topology, the HX hosts connect to the Cisco UCS Fabric Interconnects, that are in turn connected to the upstream Ethernet switches, e.g. Nexus 9000 series. The third-party storage is connected to the Ethernet switches. To configure the HX system with iSCSI external storage for HyperFlex, complete the following steps:

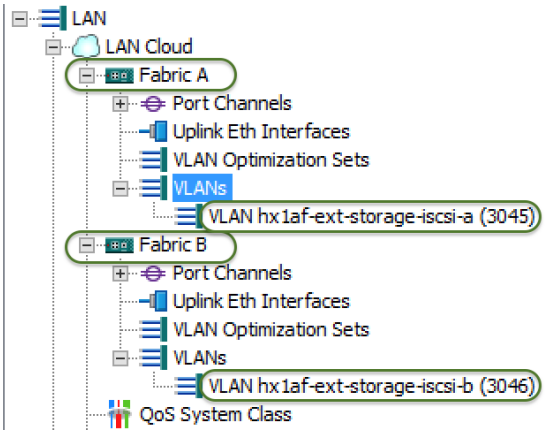
1. Prior to installation of HX, identify the iSCSI settings from the existing environment. Make sure that the third-party storage device has two iSCSI VLANs. Record them in the following table (Table 56). This information will be needed for later use during the HX install. Record the IP addresses of the iSCSI controller interfaces for the A and B path targets, and the iSCSI IQN name of the target device. Depending on how the redundant storage paths are configured in the production, more than two controlling interfaces might be recorded here. For example, in a Cisco validated FlexPod setup, where the NetApp storage array connects to Cisco Nexus 9000 series switches via VPC, normally four iSCSI IP addresses are assigned, two for each path (A and B).

Table 56 iSCSI Storage Settings

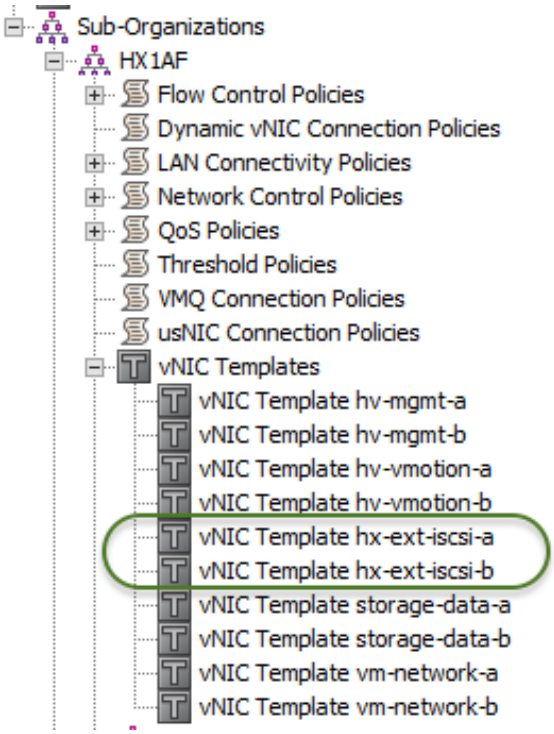
Items	Fabric A	Fabric B	
iSCSI VLAN ID			
iSCSI Target Ports	IP Address-A	IP Address-B	iSCSI IQN Name
iSCSI Storage Controller #1			
iSCSI Storage Controller #2			

2. Follow these steps to create HX cluster with the external storage adapters using the same VLAN ID's obtained from Step 1 for both Fabric A and B. Upon completion of HX install two additional vNICs for iSCSI will be created for each HX host.

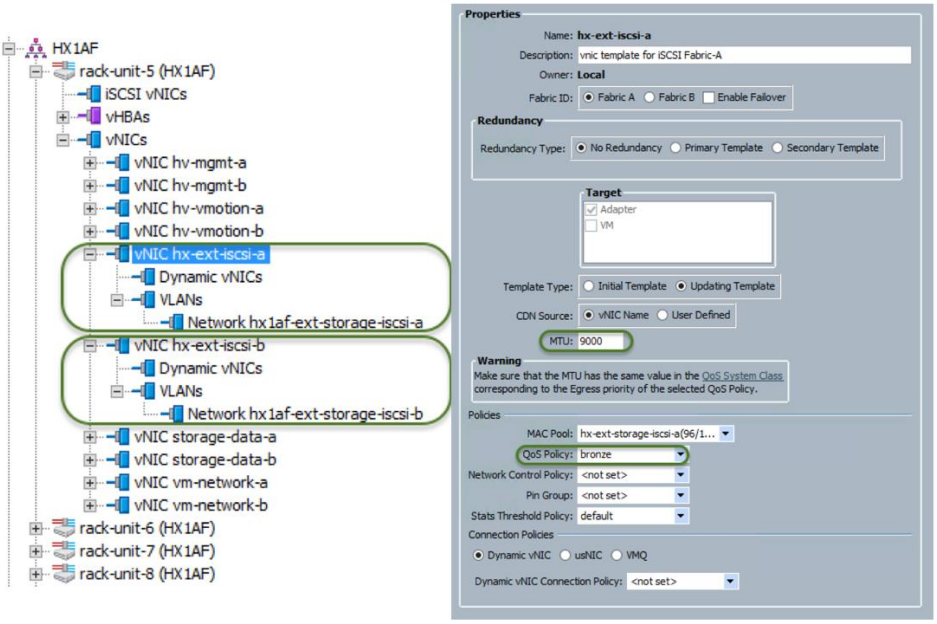
3. Open Cisco UCS Manager, expand LAN > LAN Cloud > Fabric A > VLANs, then Fabric B > VLANs to verify that the iSCSI VLANs are created and assigned to Fabric A and B.



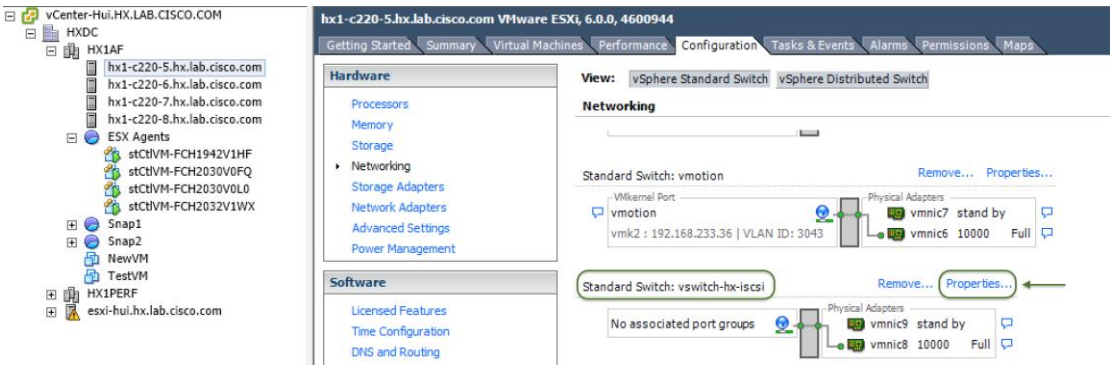
4. On the LAN tab, expand Policies > root > Sub-Organizations, go to the HX sub-organization just created, view the iSCSI templates that were created.



5. In Cisco UCS Manager, Expand Servers > Service Profiles > root > Sub-Organizations, go to the HX sub-organization just created, verify the iSCSI vNICs on all HX servers. Click one vNIC, view the properties of that iSCSI adapter. Make sure Jumbo MTU 9000 is set.



6. Set up the networking for the vSphere iSCSI switch. Login to vCenter and select the first node of the HX cluster in the left screen, then on the right screen select the Configuration tab, select Networking in the hardware pane, then scroll to the iSCSI switch. Click Properties.



7. Click Add.

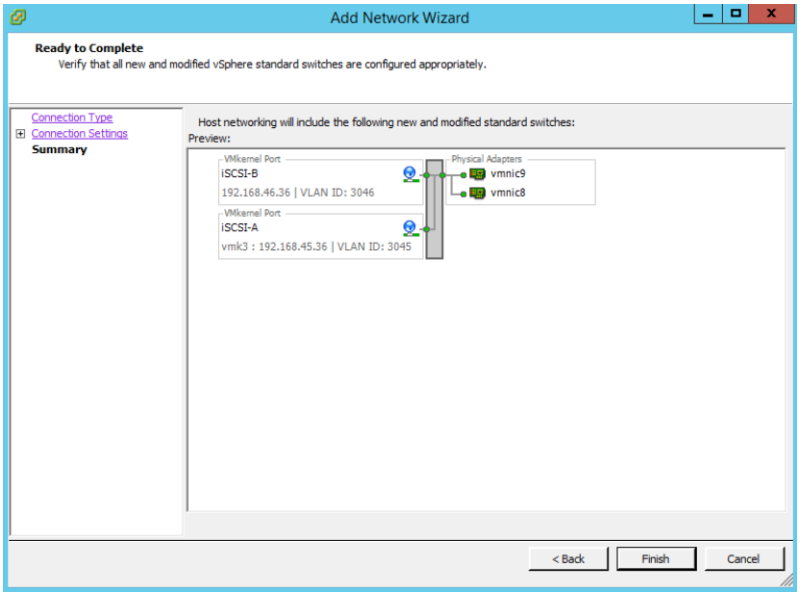
8. Select VMkernel and click Next.

9. Name iSCSI-A for the Network Label and input iSCSI VLAN ID for the A Fabric, then click Next.

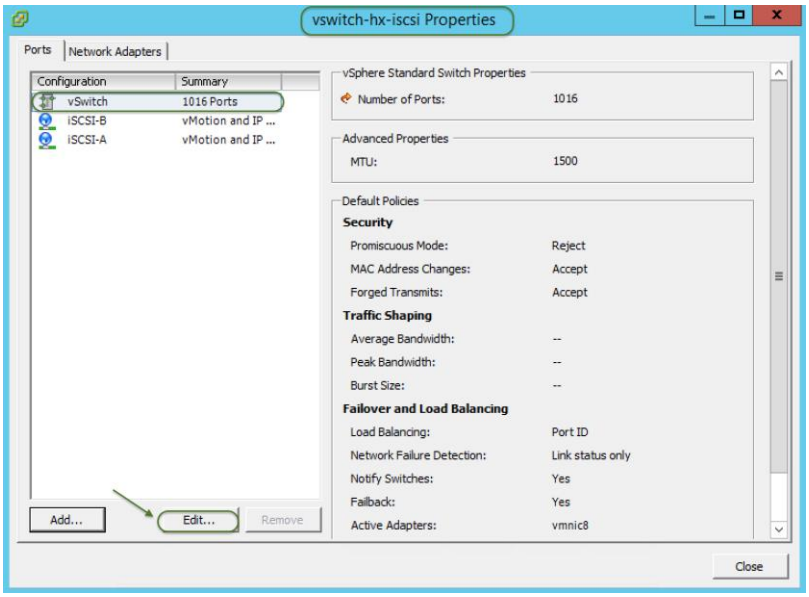
10. Add the IP address for subnet for Fabric-A and click Next.

11. Click Finish to complete addition of iSCSI VMkernel port for A Fabric.

12. Repeat Steps 7-11 to add VMkernel Port for iSCSI-B.



13. Back to the vSwitch Properties page, highlight the vSwitch and click Edit.

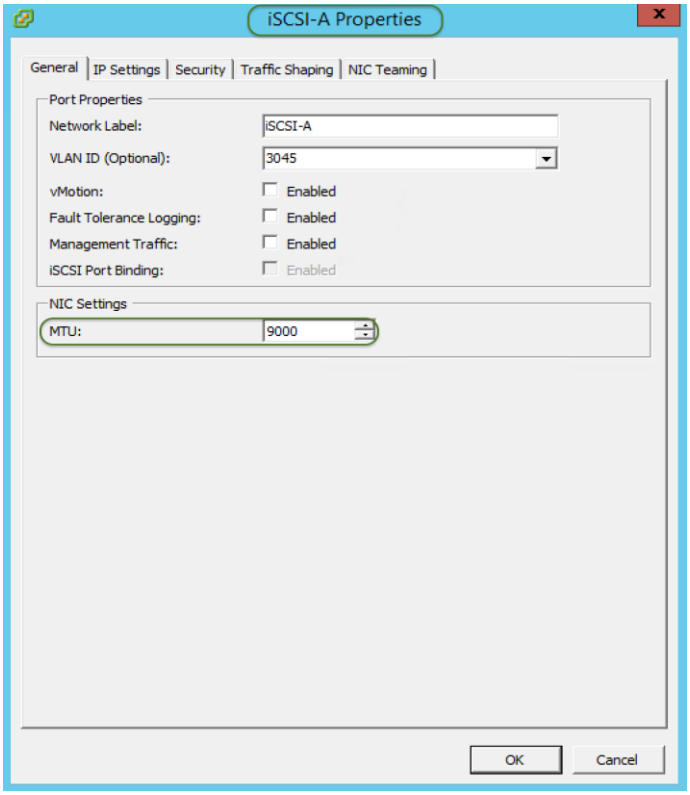


14. Change MTU for vSwitch to 9000.

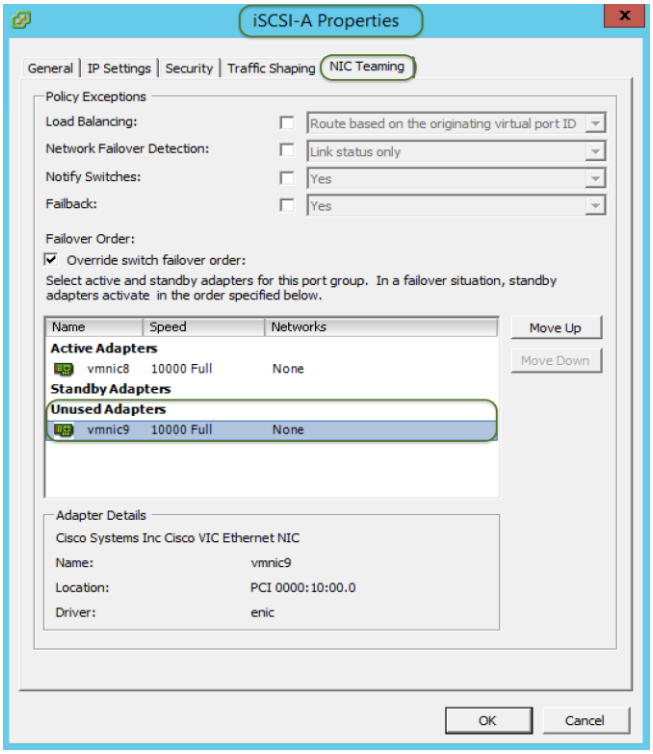
15. Select the NIC Teaming tab and make both adapters active by moving the standby adapter up. Click OK.

16. Highlight the iSCSI-A VMkernel port and click Edit in the vSwitch Properties page.

17. Change the port MTU t0 9000.



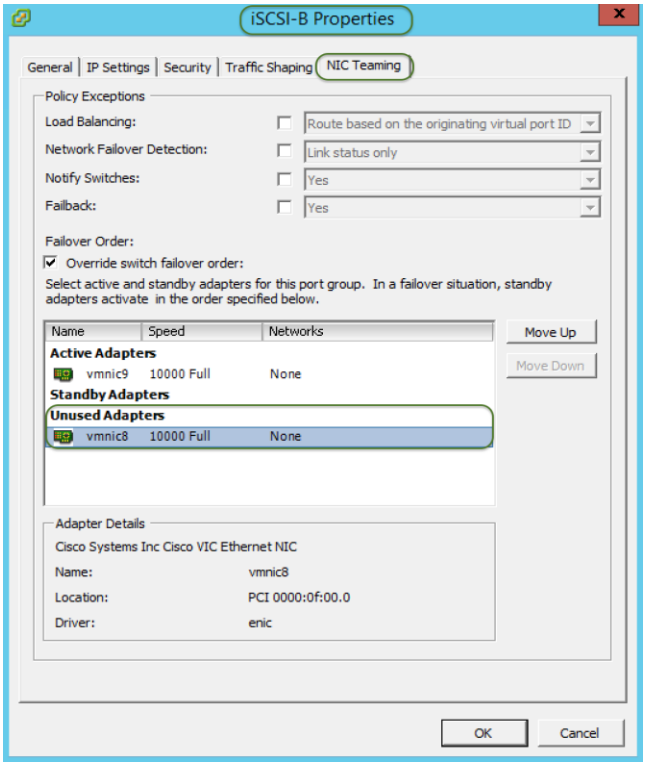
18. Select the NIC Teaming tab. Choose the option of Override switch failover order, highlight vmnic9 and move it to Unused Adapters as this adapter is for the iSCSI-B connection. Click OK.



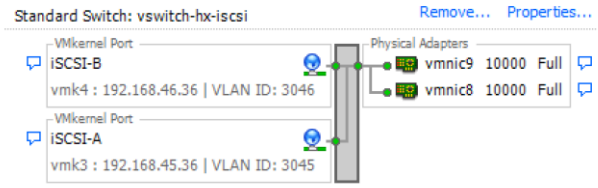
19. Highlight the iSCSI-B VMkernel port and click Edit.

20. Change the port MTU t0 9000.

21. Select the NIC Teaming tab. Select the Override switch failover order, highlight vmnic8 and move it to Unused Adapters as this adapter is for the iSCSI-A connection. Click OK.

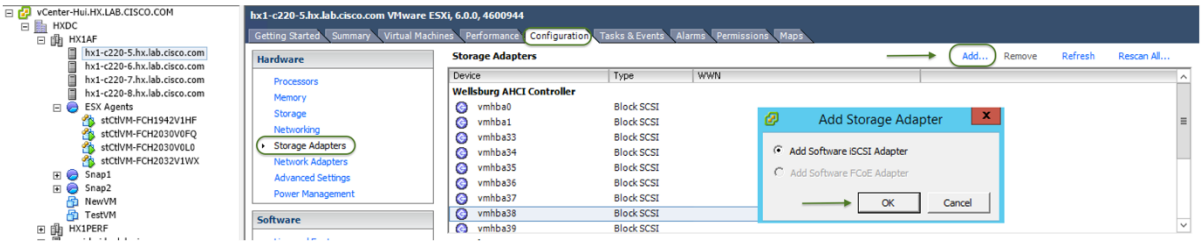


22. Click Close and review the iSCSI vSwitch. Now you should have two IP addresses used in the vSwitch on separate VLANs.



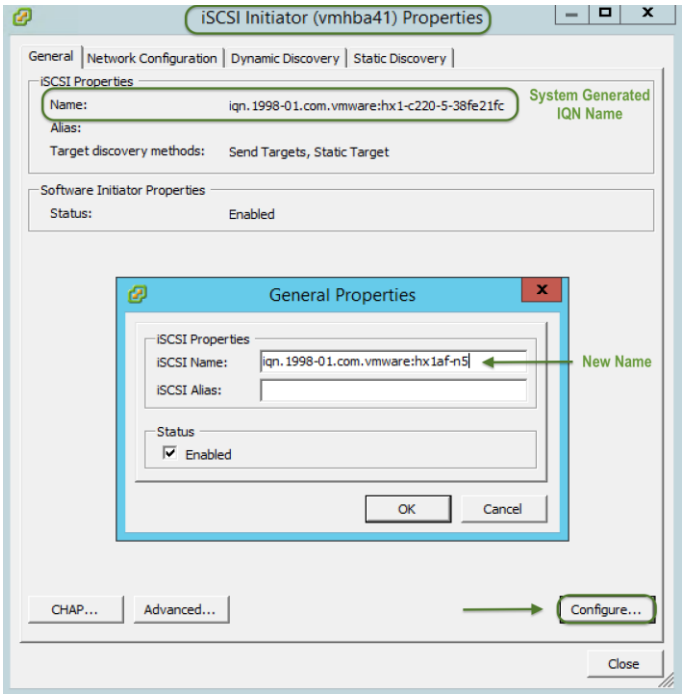
23. Repeat Steps 6-22 to configure the iSCSI vSwitch for the other HX nodes in the cluster.

24. Add the software iSCSI adapters on HX hosts. Select the first node of the HX cluster in the left screen, then on the right screen select the Configuration tab, select Storage Adapters in the hardware pane and click Add, then click OK to Add Software iSCSI Adapters, and then click OK again.

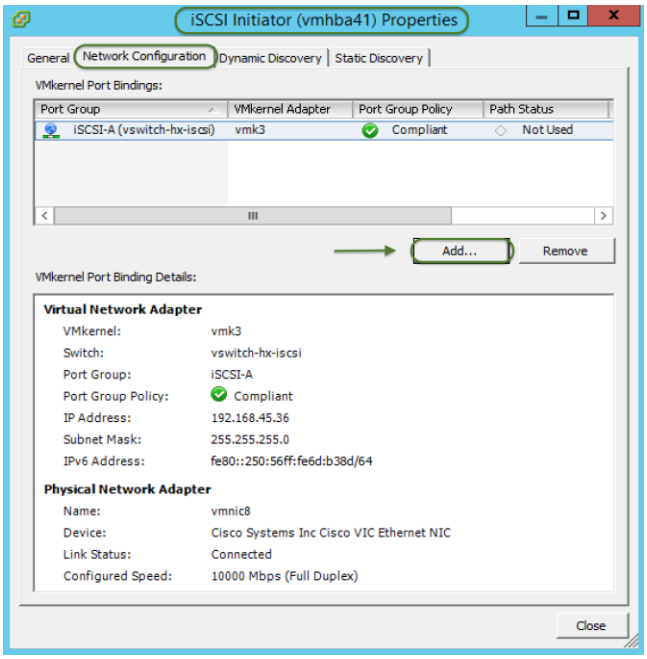


25. Scroll down and right-click the newly created software initiator, right-click and select Properties.

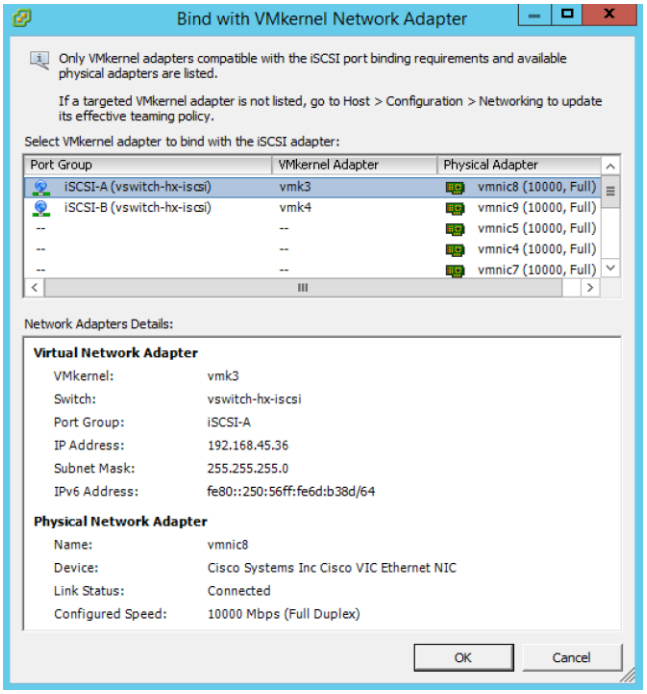
26. Click Configure to change the iSCSI IQN name to a customized name.



27. Click the Network Configuration tab and click Add to bind the VMkernel Adapters to the software iSCSI adapter.



28. Select iSCSI-A and click OK.



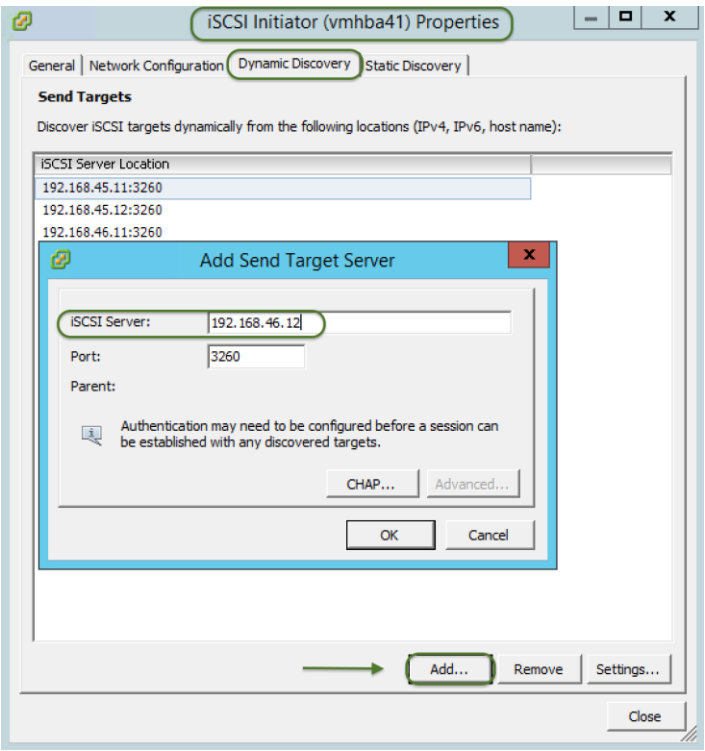
29. Click Add again and select iSCSI-B and click OK.

30. Copy and record the initiator name, IP addresses of iSCSI-A and iSCSI-B VMkernel ports to the following table. Save these values for later use to add to the initiator group created on the storage array.

Table 57 HX iSCSI Initiators

Items	Fabric A	Fabric B	
iSCSI VLAN ID			
HX Hosts	IP Address-A	IP Address-B	iSCSI IQN Name
HX Server #1 iSCSI Initiator			
HX Server #2 iSCSI Initiator			
HX Server #3 iSCSI Initiator			
HX Server #4 iSCSI Initiator			
HX Server #5 iSCSI Initiator			
HX Server #6 iSCSI Initiator			
HX Server #7 iSCSI Initiator			
HX Server #8 iSCSI Initiator			

31. Click the Dynamic Discovery tab and click Add and enter the first IP address that you recorded from your storage device network interface. Click OK. Click Add again until all the interfaces for your storage controllers are entered.

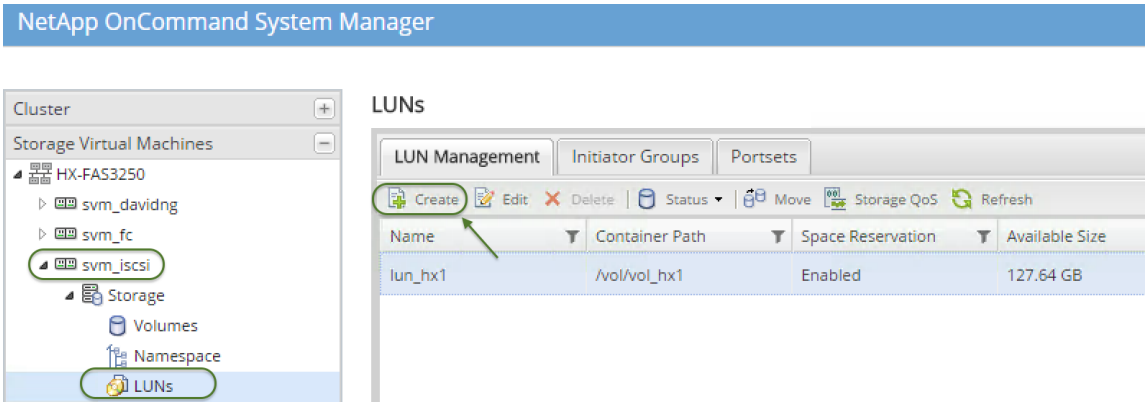


32. Click Close. You do not need to rescan the host bus adapter at this point, so choose No to the scan popup.

33. Repeat Steps 24-32 adding the software iSCSI adapters for the remaining HX nodes.

34. Now create iSCSI initiator groups and then create an iSCSI LUN on the storage system and map it to the HX system. In this example, we are using NetApp OnCommand System Manager GUI to create a LUN on a FAS3250 array. Please consult your storage documentation to accomplish the same tasks. It is assumed you have already configured your iSCSI storage as shown in the CVD.

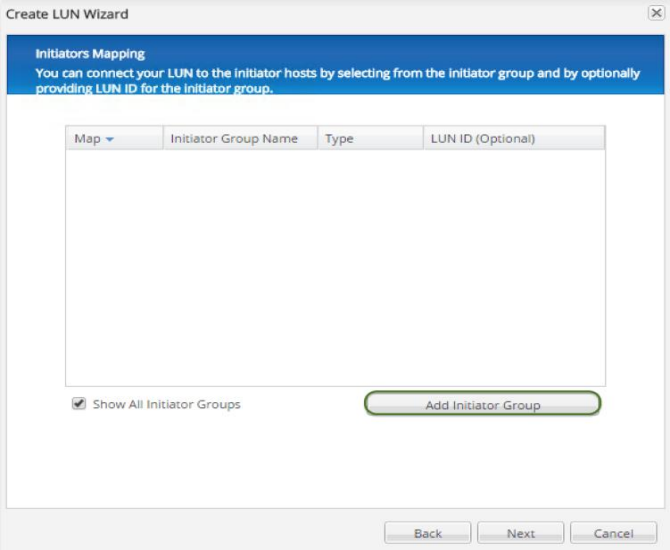
35. Open NetApp OnCommand System Manager GUI from the web browser, select the pre-configured iSCSI Storage Virtual Machine, expand Storage, then LUNs; from the right pane, click Create. This will open Create LUN wizard.



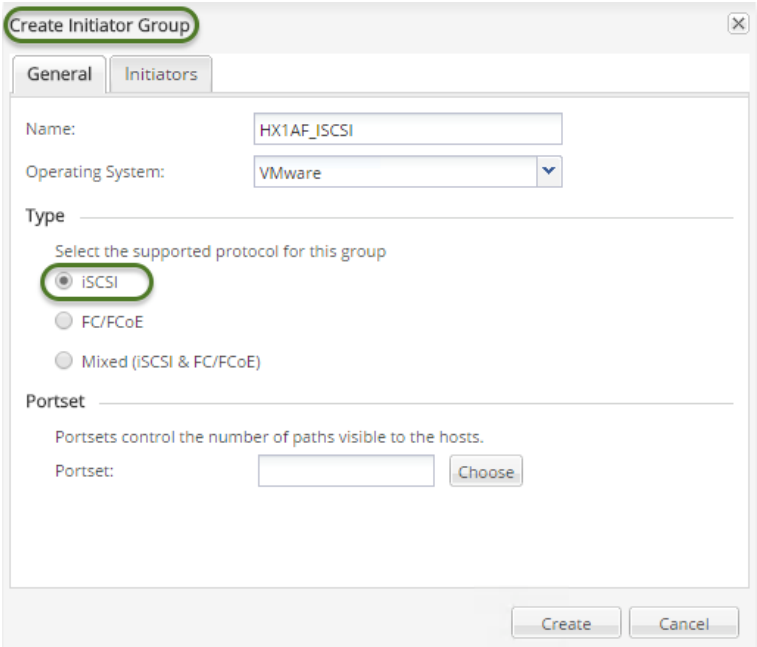
36. Click Next on the General Properties page, enter the LUN Name, Type and Size. Click Next.

37. Check “Select an existing volume or qtree for this LUN”, browse and select an existing volume, then click Next.

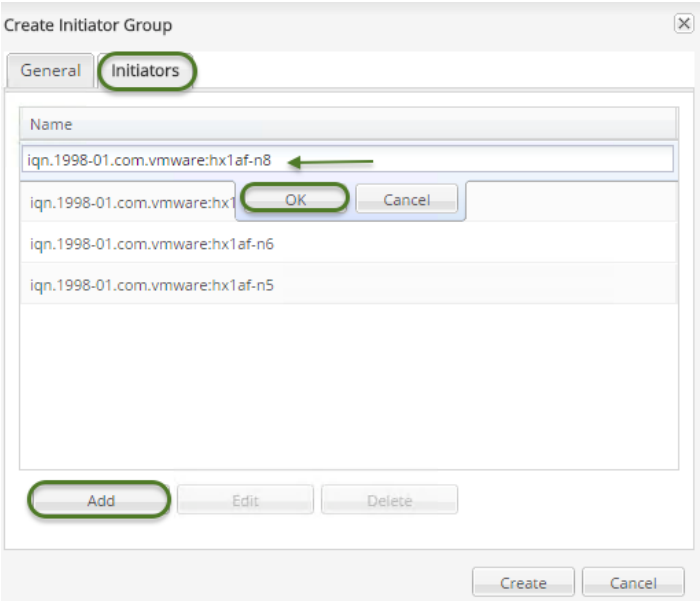
38. On Initiators Mapping page, select Add Initiator Group.



39. In Create Initiator Group wizard, on the General tab, enter Name, Operation System, and select Type of iSCSI for the Initiator Group to be created.

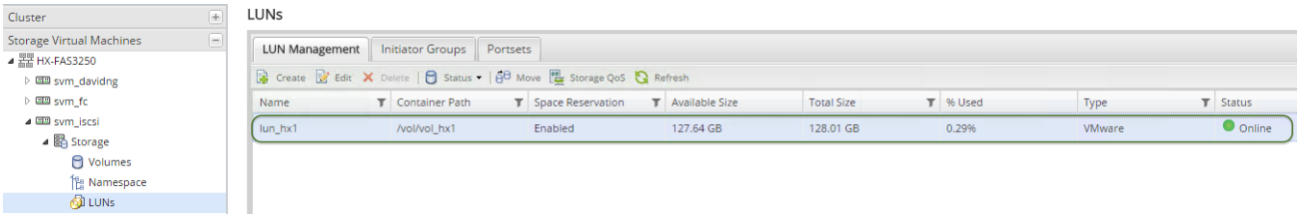


40. On Initiators tab, click Add then enter the iSCSI IQN Name of the first HX host (copy from Table 57), click OK.

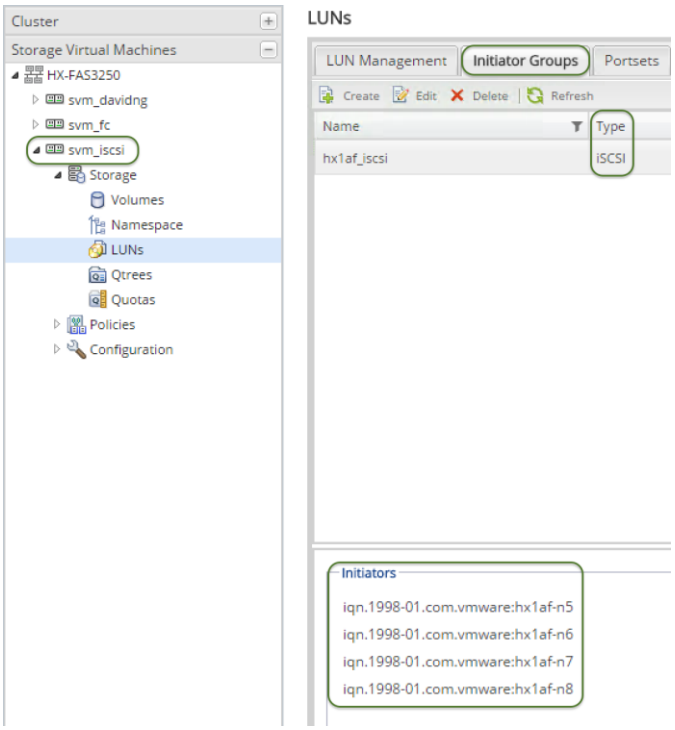


41. Repeat Step 40 until the IQN names of all HX iSCSI adapters are added. Select Create to create the Initiator Group.

42. The Create Initiator Group Wizard closes and reverts to the Initiators Mapping page of the Create LUN wizard. Select the HX initiator group that is just created, click Next three times then click Finish to complete the LUN creation.



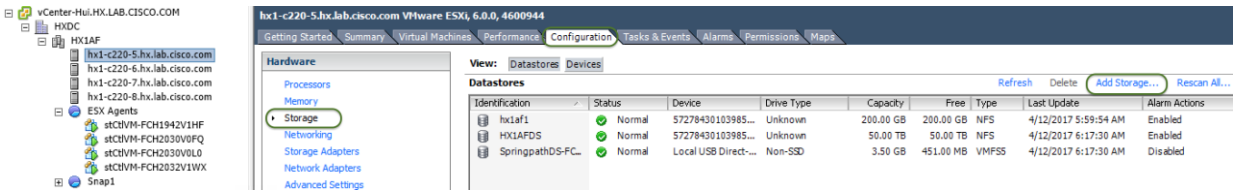
43. Check the iSCSI initiators mapped to this LUN.



44. With a mapped LUN, you can rescan the iSCSI software initiator. Login to the vCenter again, in the configuration tab, right-click the iSCSI software adapter and click Rescan or click Rescan All at the top of the pane (do this for each host).

45. The iSCSI disk will show up in the details pane.

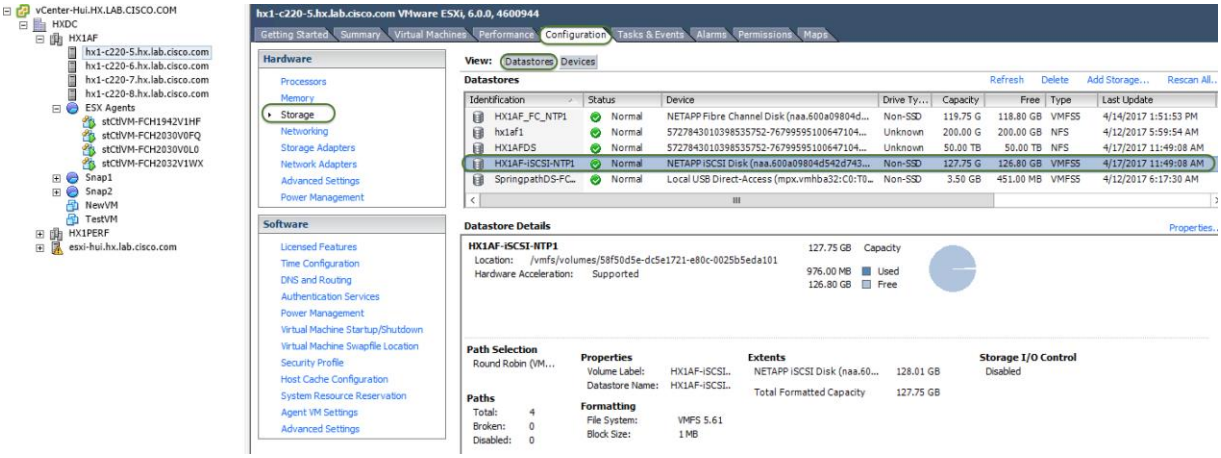
46. Add the disk to the cluster by selecting Storage in the Hardware pane, then Add Storage in the Configuration tab.



47. Leave Disk/LUN selected and click Next.

48. Now the NetApp iSCSI LUN will be detected. Highlight the disk and click Next, and then click Next again.

49. Enter the new Datastore name and click Next then Finish. A new iSCSI datastore for the HX cluster will be created.



50. You can now create VM's on this new datastore and migrate data between HX and the iSCSI datastore.

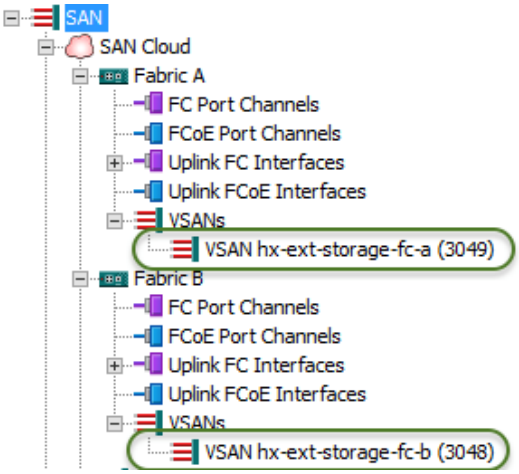
Connecting to Fibre Channel Storage

The HX installer can guide you through the process of setting up your HX cluster allowing you to leverage existing third-party storage via the Fibre Channel protocol. It will automatically configure Cisco UCS profiles, and HX cluster nodes with vHBAs, proper VSAN, and WWPN assignments, simplifying the setup. The procedure is described here in this CVD. It is assumed that the third-party storage system is already configured per a Cisco Validated Design and all networking configuration, including Fibre Channel for connecting via the upstream switches, is completed as well. In this example, you will be using Cisco MDS Fibre Channel switches that are connected to the Cisco UCS Fabric Interconnects, which are configured with Fibre Channel unified ports in End Host mode. Changing the identity of unified ports on a Cisco UCS Fabric Interconnect requires that the FIs are rebooted, so this task should be completed prior to the installation of the HyperFlex cluster(s). The third-party storage is connected to the MDS switches.

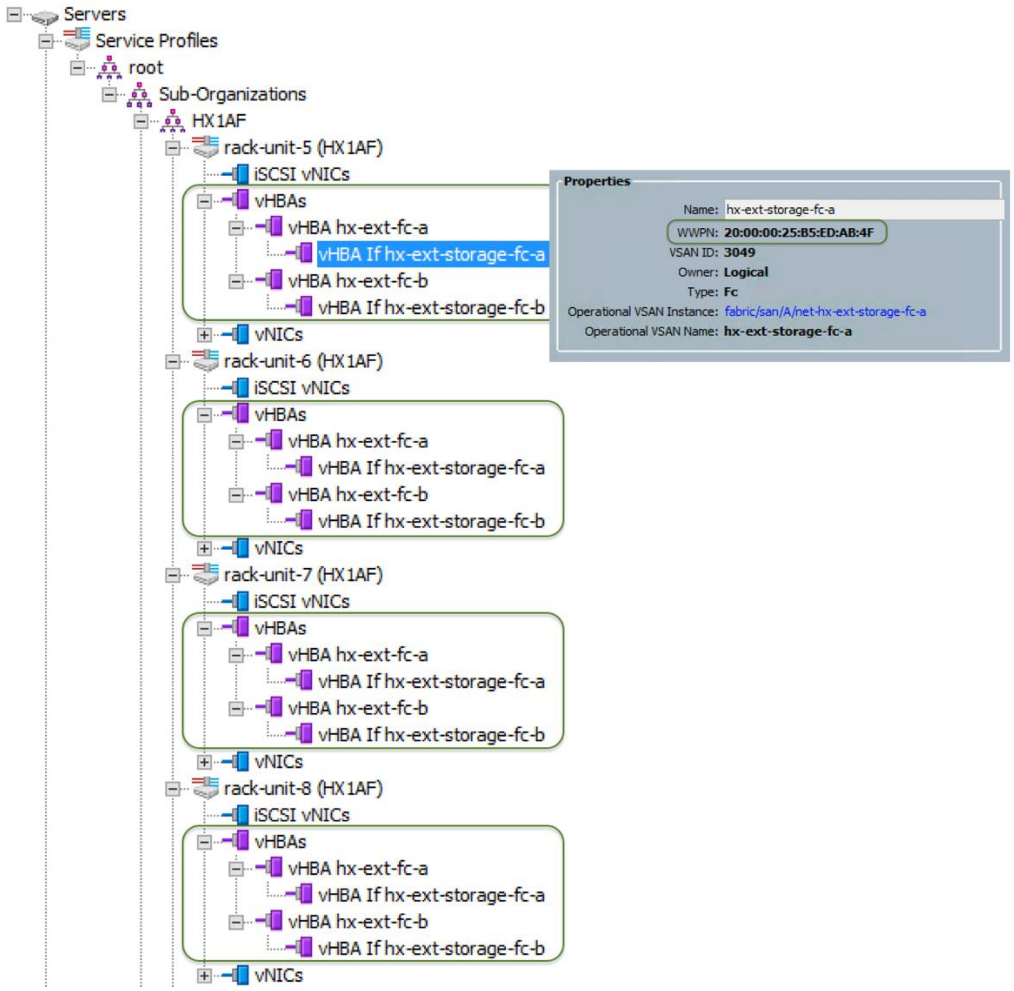
Note: It is required that you obtain the VSAN IDs being used in your current environment for the storage device that is already configured. This can be obtained from the SAN tab in Cisco UCS Manager, or from the upstream Fibre Channel switches.

1. Follow these steps for the HX cluster installation using the same VSAN IDs obtained from Step 1 for both Fabric A and B. Upon completion of HX install, two VSANs and two vHBAs (one for Fabric A and one for Fabric B) for each HX host will be created.

2. Open Cisco UCS Manager, Expand SAN > SAN Cloud > Fabric A > VSANs, then Fabric B > VSANs, verify the right VSANs are generated:



3. In Cisco UCS Manager, Expand Servers > Service Profiles > root > Sub-Organizations, go to the HX sub-organization you just created, verify vHBAs on all HX servers:

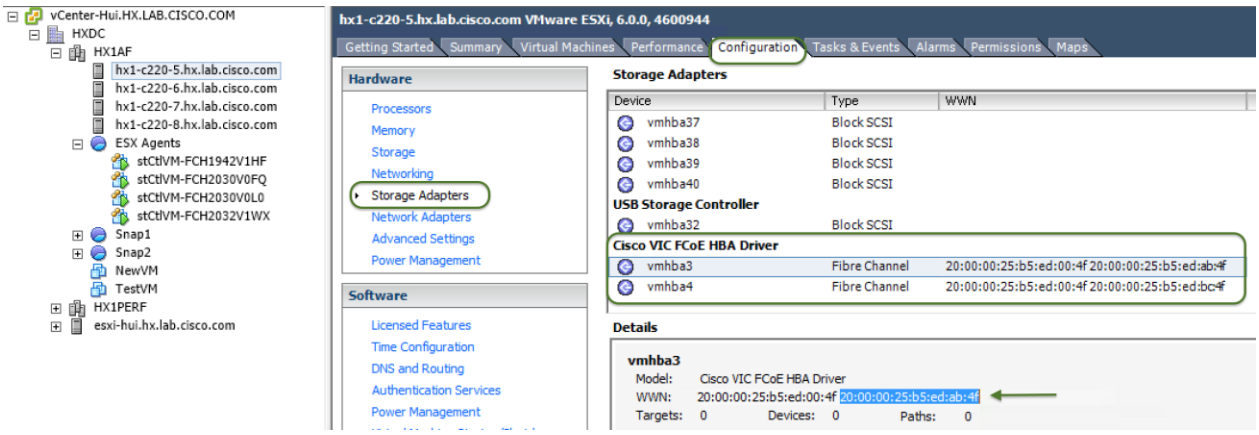


4. Record all the WWPN's for each HX node in the following table. These values are needed later for the zone configuration on the FC switches. You can copy the WWPN value by clicking on the vHBA in Cisco UCS Manager and the in the right pane, right-clicking the WWPN to copy.

Table 58 WWPNs on HX Hosts

Items		Fabric A	Fabric B
HX Server #1	WWPN		
	Alias		
HX Server #2	WWPN		
	Alias		
HX Server #3	WWPN		
	Alias		
HX Server #4	WWPN		
	Alias		
HX Server #5	WWPN		
	Alias		
HX Server #6	WWPN		
	Alias		
HX Server #7	WWPN		
	Alias		
HX Server #8	WWPN		
	Alias		

5. Alternatively, you can copy the WWPN value on the ESXi host in vCenter on the Configuration tab > Storage Adapters > Cisco VIC FCoE HBA Driver > <<vmhba>>.



6. The WWPNs for the storage ports will also be recorded. These values are needed later for zone configuration on the FC switches. You can get that information from your storage device's management tool.

Table 59 Storage WWPNs

Items		Fabric A	Fabric B
Storage Device Port #1	WWPN		
	Alias		
Storage Device Port #2	WWPN		
	Alias		
Storage Device Port #3	WWPN		
	Alias		
Storage Device Port #4	WWPN		
	Alias		

7. Login to the MDS switch for A Fabric (MDS A), verify all HX vHBAs for A fabric have login to the name server and verify they are in the same VSAN as the target storage ports. Example:

```
HX1-C25-MDSA(config-vsan-db)# show flogi database vsan 3049

--------------------------------------------------------------------------------

INTERFACE        VSAN    FCID           PORT NAME               NODE NAME

--------------------------------------------------------------------------------

fc1/1            3049  0xba0000  20:1f:8c:60:4f:8d:dc:c0 2b:e9:8c:60:4f:8d:dc:c1

fc1/1            3049  0xba0001  20:00:00:25:b5:ed:ab:4f 20:00:00:25:b5:ed:00:4f

fc1/1            3049  0xba0002  20:00:00:25:b5:ed:ab:5f 20:00:00:25:b5:ed:00:5f

fc1/1            3049  0xba0003  20:00:00:25:b5:ed:ab:2f 20:00:00:25:b5:ed:00:2f

fc1/1            3049  0xba0004  20:00:00:25:b5:ed:ab:3f 20:00:00:25:b5:ed:00:3f

fc1/49           3049  0xba0020  50:0a:09:85:8d:b2:b9:0c 50:0a:09:80:8d:b2:b9:0c

fc1/49           3049  0xba0021  20:01:00:a0:98:1e:9c:9c 20:00:00:a0:98:1e:9c:9c
```

8. Complete the following steps to create the WWPN aliases using the values form the table. Example:

```
configure terminal

device-alias database

device-alias name HX1AF-N5a pwwn 20:00:00:25:b5:ed:ab:4f

device-alias name HX1AF-N6a pwwn 20:00:00:25:b5:ed:ab:5f

device-alias name HX1AF-N7a pwwn 20:00:00:25:b5:ed:ab:2f

device-alias name HX1AF-N8a pwwn 20:00:00:25:b5:ed:ab:3f
```

```
device-alias name FAS3250-010c pwwn 20:01:00:a0:98:1e:9c:9c

device-alias commit
```

9. Create the zones and add device-alias members (or PWWN members) for the HX servers. Example:

```
zone name HX1AF-N5a vsan 3049

 member device-alias HX1AF-N5a

 member device-alias FAS3250-010c

 exit

zone name HX1AF-N6a vsan 3049

 member device-alias HX1AF-N6a

 member device-alias FAS3250-010c

 exit

zone name HX1AF-N7a vsan 3049

 member device-alias HX1AF-N7a

 member device-alias FAS3250-010c

 exit

zone name HX1AF-N8a vsan 3049

 member device-alias HX1AF-N8a

 member device-alias FAS3250-010c

 exit
```

10. Create a zoneset and add the zones. Example:

```
zoneset name HX1AF-a vsan 3049

member HX1AF-N5a

member HX1AF-N6a

member HX1AF-N7a

member HX1AF-N8a

exit
```

11. Activate the zoneset. Example:

```
zoneset activate name HX1AF-a vsan 3049
```

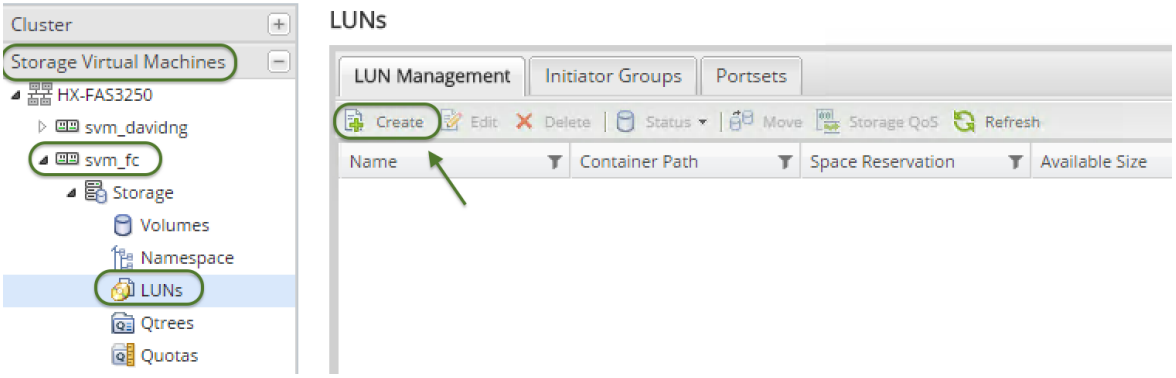
12. Validate the active zoneset and verify that all HX vHBAs and the target storage ports are logged into the switch (indicated by the * next to the devices). Example:

```
HX1-C25-MDSA(config)# show zoneset active vsan 3049

zoneset name HX1AF-a vsan 3049

  zone name HX1AF-N5a vsan 3049

  * fcid 0xba0001 [pwwn 20:00:00:25:b5:ed:ab:4f] [HX1AF-N5a]

  * fcid 0xba0021 [pwwn 20:01:00:a0:98:1e:9c:9c] [FAS3250-010c]


  zone name HX1AF-N6a vsan 3049

  * fcid 0xba0002 [pwwn 20:00:00:25:b5:ed:ab:5f] [HX1AF-N6a]

  * fcid 0xba0021 [pwwn 20:01:00:a0:98:1e:9c:9c] [FAS3250-010c]


  zone name HX1AF-N7a vsan 3049

  * fcid 0xba0003 [pwwn 20:00:00:25:b5:ed:ab:2f] [HX1AF-N7a]

  * fcid 0xba0021 [pwwn 20:01:00:a0:98:1e:9c:9c] [FAS3250-010c]


  zone name HX1AF-N8a vsan 3049

  * fcid 0xba0004 [pwwn 20:00:00:25:b5:ed:ab:3f] [HX1AF-N8a]

  * fcid 0xba0021 [pwwn 20:01:00:a0:98:1e:9c:9c] [FAS3250-010c]
```

13. Login to the MDS switch for the B fabric (MDS B) and complete steps 7-12 to create and activate the FC zones on the B side FC fabric.

14. Create initiator groups and then create a LUN on the storage system and map it to the HX system. In this example, we are using NetApp OnCommand System Manager GUI to create a LUN on a FAS3250 array. Please consult your storage documentation to accomplish the same tasks. It is assumed you have pre-existing FC storage configurations on an array as shown in this CVD.

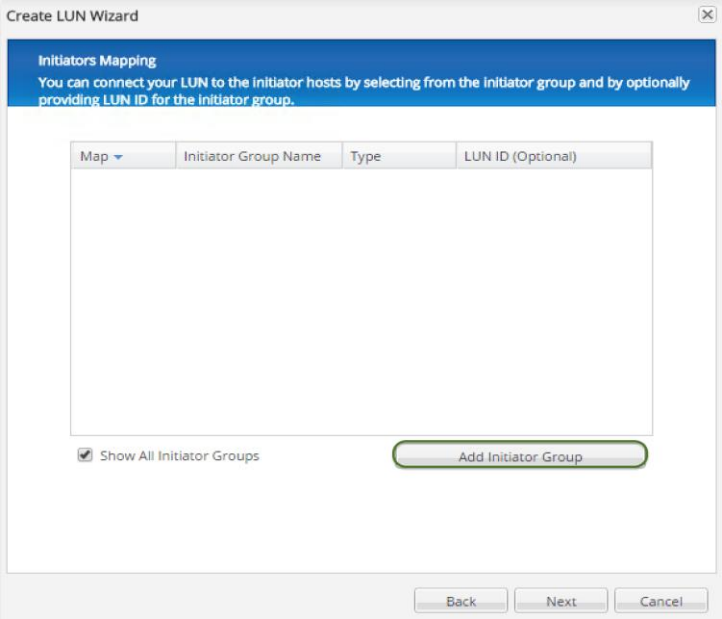
15. Open NetApp OnCommand System Manager GUI from the web browser, select the pre-configured FC Storage Virtual Machine, expand Storage, then LUNs; from the right pane, click Create. This opens the Create LUN wizard.



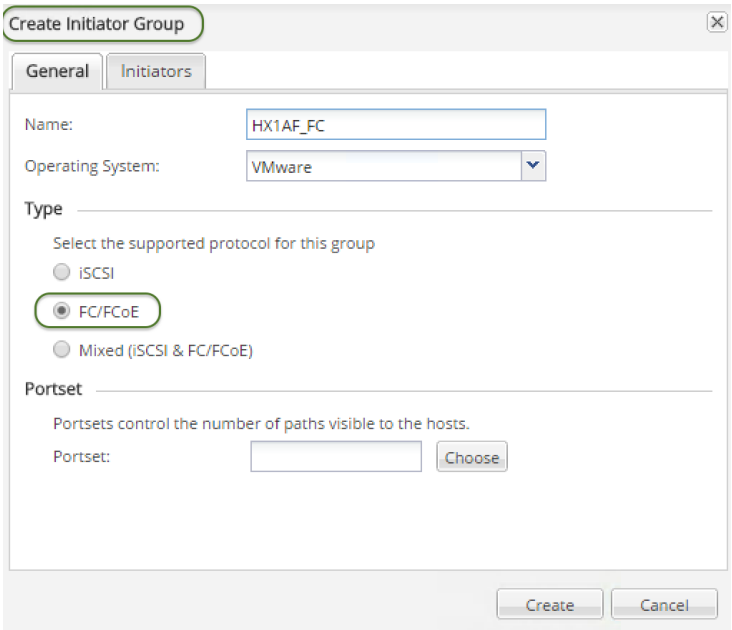
16. Click Next. In General Properties page, enter LUN Name, Type and Size. Click Next.

17. Check “Select an existing volume or qtree for this LUN”, browse and select an existing volume, then click Next.

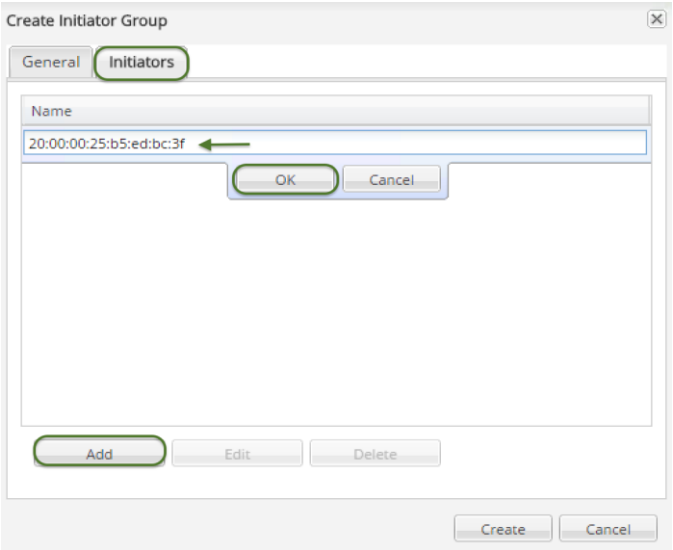
18. On Initiators Mapping page, select Add Initiator Group.



19. In Create Initiator Group wizard, on General tab, enter Name, Operation System, and select Type of FC/FCoE for the Initiator Group to be created.

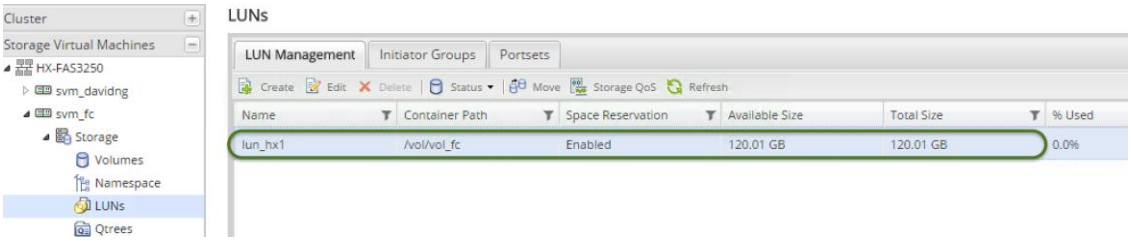


20. On Initiators tab, click Add then enter the WWPN of the first HX vHBA, click OK.

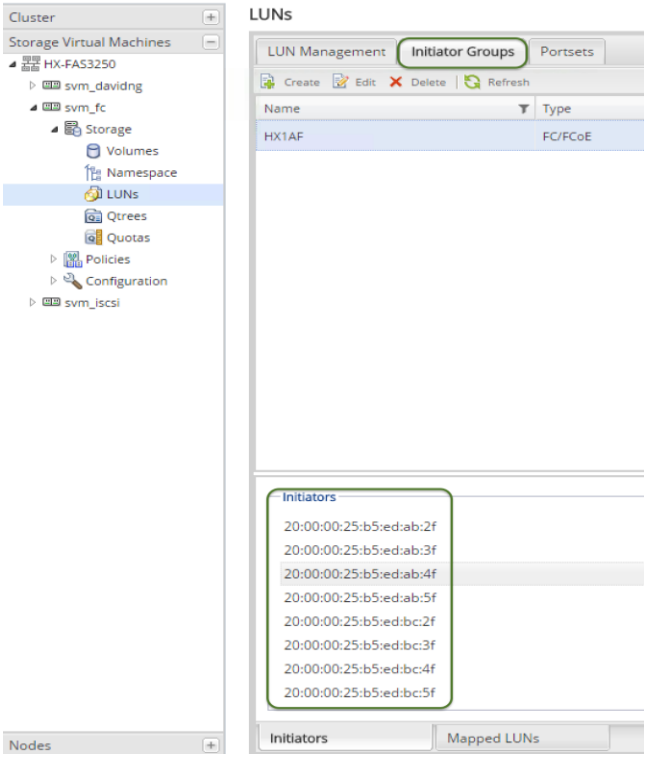


21. Repeat Step 21 until the WWPNs of all HX vHBAs (on both Fabric A and B) are added. Select Create to create the Initiator Group.

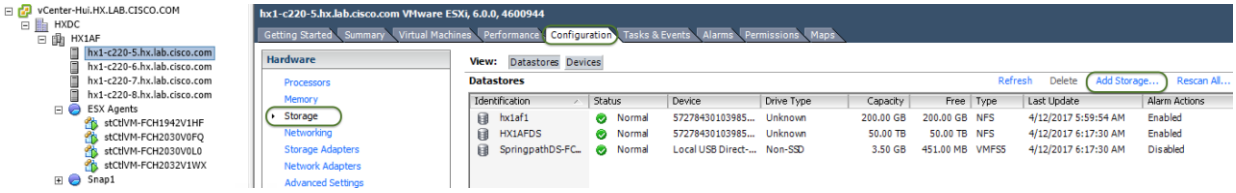
22. The Add Initiator Group Wizard exits back to Initiators Mapping page of the Create LUN wizard. Select the HX initiator group that is just created, click Next three times then click Finish to complete the LUN creation.



23. Check the initiators mapped to this LUN.

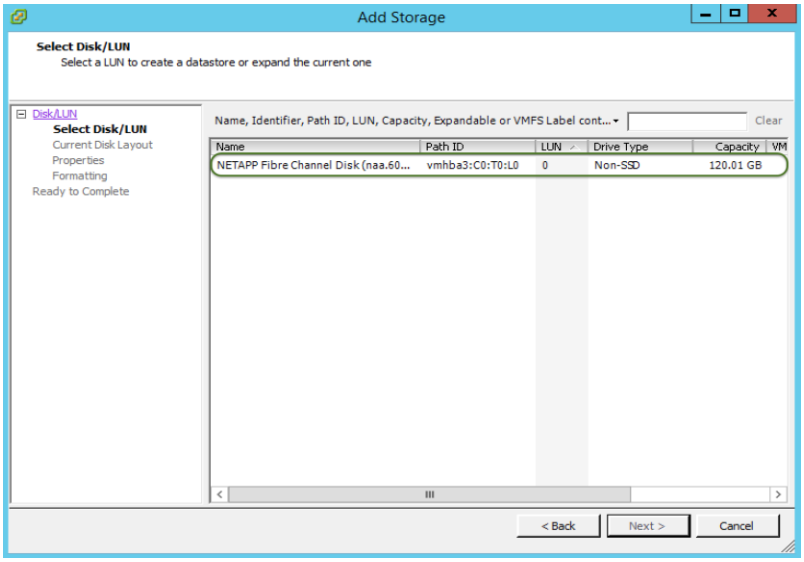


24. Return to vCenter and from the Configuration tab, select Storage, then Add Storage.



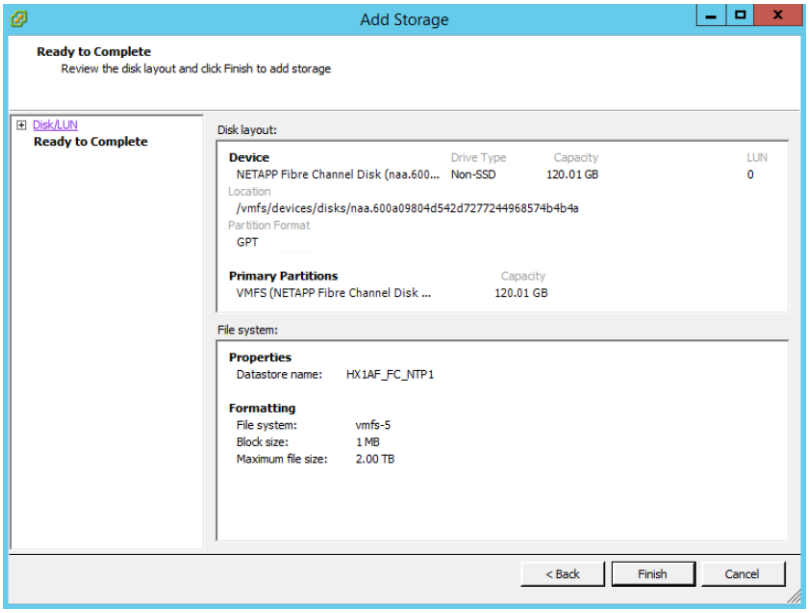
25. Leave Disk/LUN selected and click Next.

26. The NetApp Fibre Channel LUN just created will be detected. Highlight the disk and click Next, then click Next again.

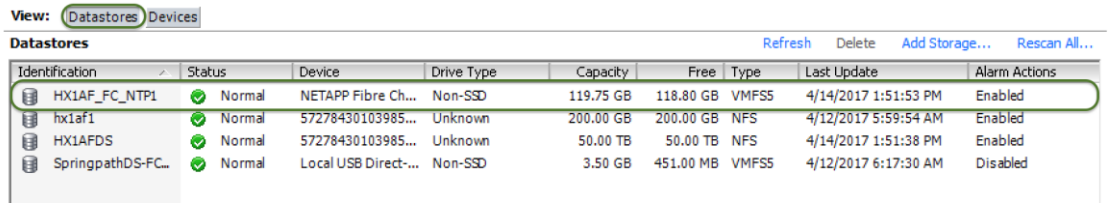


27. Enter the Name of the Datastore and click Next.

28. Click Next for maximum available space as desired, then click Finish.



29. You can now review your datastores in the configuration tab, and perform storage migration of any VM's if necessary.



F: Adding HX to an Existing Cisco UCS Domain

For a scenario where HX nodes are added to an existing Cisco UCS domain, caution is advised. A Cisco UCS firmware upgrade or changes to the configuration on the upstream switches may be required as part of the installation. All of these changes could be disruptive to the existing systems and workloads, and need to be carefully planned and implemented within a maintenance window. It is recommended that you contact Cisco TAC, or your Cisco sales engineer support team for assistance when you need to connect HX nodes to an existing Cisco UCS domain.

About the Authors

Brian Everitt, Technical Marketing Engineer, Cisco UCS Data Center Engineering Group, Cisco Systems, Inc.

Brian is an IT industry veteran with over 20 years of experience deploying server, network, and storage infrastructures for companies around the world. During his tenure at Cisco, he has been a lead Advanced Services Solutions Architect for Microsoft solutions, virtualization, and SAP Hana on Cisco UCS. Currently his focus is on Cisco's portfolio of Software Defined Storage (SDS) and Hyperconverged Infrastructure solutions. Brian has earned multiple certifications from Microsoft, Cisco, and VMware.

Hui Chen, Technical Marketing Engineer, Cisco UCS Data Center Engineering Group, Cisco Systems, Inc.

Hui is a network and storage veteran with over 15 years of experience on Fibre Channel-based storage area networking, the LAN/SAN convergence systems, and how to build end-to-end; from the server to storage, and solutions in the data center. Currently he focuses on Cisco's Software Defined Storage (SDS) and Hyperconverged Infrastructure (HCI) solutions. Hui is also a seasoned CCIE.