

Cisco HyperFlex 4.0 for Citrix VDI with VMware ESXi for up to 500 Users

Deployment Guide for Cisco HyperFlex 4.0 for Citrix VDI Using Stretch Cluster Converged Nodes for High Availability

Published: July 2021



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, go to:

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

ALL DESIGNS, SPECIFICATIONS, STATEMENTS, INFORMATION, AND RECOMMENDATIONS (COLLECTIVELY, "DESIGNS") IN THIS MANUAL ARE PRESENTED "AS IS," WITH ALL FAULTS. CISCO AND ITS SUPPLIERS DISCLAIM ALL WARRANTIES, INCLUDING, WITHOUT LIMITATION, THE WARRANTY OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OR ARISING FROM A COURSE OF DEALING, USAGE, OR TRADE PRACTICE. IN NO EVENT SHALL CISCO OR ITS SUPPLIERS BE LIABLE FOR ANY INDIRECT, SPECIAL, CONSEQUENTIAL, OR INCIDENTAL DAMAGES, INCLUDING, WITHOUT LIMITATION, LOST PROFITS OR LOSS OR DAMAGE TO DATA ARISING OUT OF THE USE OR INABILITY TO USE THE DESIGNS, EVEN IF CISCO OR ITS SUPPLIERS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

THE DESIGNS ARE SUBJECT TO CHANGE WITHOUT NOTICE. USERS ARE SOLELY RESPONSIBLE FOR THEIR APPLICATION OF THE DESIGNS. THE DESIGNS DO NOT CONSTITUTE THE TECHNICAL OR OTHER PROFESSIONAL ADVICE OF CISCO, ITS SUPPLIERS OR PARTNERS. USERS SHOULD CONSULT THEIR OWN TECHNICAL ADVISORS BEFORE IMPLEMENTING THE DESIGNS. RESULTS MAY VARY DEPENDING ON FACTORS NOT TESTED BY CISCO.

CCDE, CCENT, Cisco Eos, Cisco Lumin, Cisco Nexus, Cisco StadiumVision, Cisco TelePresence, Cisco WebEx, the Cisco logo, DCE, and Welcome to the Human Network are trademarks; Changing the Way We Work, Live, Play, and Learn and Cisco Store are service marks; and Access Registrar, Aironet, AsyncOS, Bringing the Meeting To You, Catalyst, CCDA, CCDP, CCIE, CCIP, CCNA, CCNP, CCSP, CCVP, Cisco, the Cisco Certified Inter-network Expert logo, Cisco IOS, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Cisco Unified Computing System (Cisco UCS), Cisco UCS B-Series Blade Servers, Cisco UCS C-Series Rack Servers, Cisco UCS S-Series Storage Servers, Cisco UCS Manager, Cisco UCS Management Software, Cisco Unified Fabric, Cisco Application Centric Infrastructure, Cisco Nexus 9000 Series, Cisco Nexus 7000 Series, Cisco Prime Data Center Network Manager, Cisco NX-OS Software, Cisco MDS Series, Cisco Unity, Collaboration Without Limitation, EtherFast, EtherSwitch, Event Center, Fast Step, Follow Me Browsing, FormShare, Giga-Drive, HomeLink, Internet Quotient, IOS, iPhone, iQuick Study, LightStream, Linksys, MediaTone, MeetingPlace, MeetingPlace Chime Sound, MGX, Networkers, Networking Academy, Network Registrar, PCNow, PIX, PowerPanels, ProConnect, ScriptShare, SenderBase, SMARTnet, Spectrum Expert, StackWise, The Fastest Way to Increase Your Internet Quotient, TransPath, WebEx, and the WebEx logo are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the United States and certain other countries. (LDW_P1).

All other trademarks mentioned in this document or website are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (0809R)

© 2021 Cisco Systems, Inc. All rights reserved.

Contents

Executive Summary	4
Solution Overview.....	5
Technology Overview	7
Solution Design	34
Design Elements	43
Installation	48
Management	92
Validation.....	96
Build the Virtual Machines and Environment for Workload Testing	98
Test Setup and Configurations.....	191
Test Methodology and Success Criteria.....	193
Test Results.....	207
Summary	219
About the Author	220
Feedback.....	221

Executive Summary

To keep pace with the market, you need systems that support rapid, agile development processes. Cisco HyperFlex™ Systems let you unlock the full potential of hyper-convergence and adapt IT to the needs of your workloads. The systems use an end-to-end software-defined infrastructure approach, combining software-defined computing in the form of Cisco HyperFlex HX-Series Nodes, software-defined storage with the powerful Cisco HyperFlex HX Data Platform, and software-defined networking with the Cisco UCS fabric that integrates smoothly with Cisco® Application Centric Infrastructure (Cisco ACI™).

Together with a single point of connectivity and management, these technologies deliver a pre-integrated and adaptable cluster with a unified pool of resources that you can quickly deploy, adapt, scale, and manage to efficiently power your applications and your business

This document provides an architectural reference and design guide for up to 500 VDI session workload on an 8-node Cisco HyperFlex system Stretch Cluster. We provide deployment guidance and performance data for Citrix Virtual Desktops 1912 LTSR virtual desktops running Microsoft Windows 10 with Office 2016 and Windows Server 2019 for HSD. The solution is a pre-integrated, best-practice data center architecture built on the Cisco Unified Computing System (Cisco UCS), the Cisco Nexus® 9000 family of switches and Cisco HyperFlex Data Platform software version 4.0.2b.

The solution payload is 100 percent virtualized on Cisco HyperFlex HXAF220C-M5SX hyperconverged nodes booting through on-board M.2 SATA SSD drive running VMware hypervisor and the Cisco HyperFlex Data Platform storage controller virtual machine. The virtual desktops are configured with Virtual Desktops 1912 LTSR, which incorporates both traditional persistent and non-persistent virtual Windows 10 desktops, hosted applications, and remote desktop service (RDS) Microsoft Server 2019 based desktops. The solution provides unparalleled scale and management simplicity. Citrix Virtual Desktops Provisioning Services or Machine Creation Services Windows 10 desktops, full clone desktops or Virtual Apps server-based desktops can be provisioned on an eight node Cisco HyperFlex cluster. Where applicable, this document provides best practice recommendations and sizing guidelines for customer deployment of this solution.

Solution Overview

Introduction

The current industry trend in data center design is towards small, granularly expandable hyperconverged infrastructures. By using virtualization along with pre-validated IT platforms, customers of all sizes have embarked on the journey to “just-in-time capacity” using this new technology. The Cisco HyperFlex hyperconverged solution can be quickly deployed, thereby increasing agility, and reducing costs. Cisco HyperFlex uses best of breed storage, server, and network components to serve as the foundation for desktop virtualization workloads, enabling efficient architectural designs that can be quickly and confidently deployed and scaled-out.

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, Citrix and Microsoft specific technologies, infrastructure concepts, networking connectivity, and security policies of the customer installation.

Purpose of this Document

This document provides a step-by-step design, configuration, and implementation guide for the Cisco Validated Design for a Cisco HyperFlex Stretch Cluster system running three different Citrix Virtual Desktops/Virtual Apps workloads with Cisco UCS 6400 series Fabric Interconnects and Cisco Nexus 9000 series switches.

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



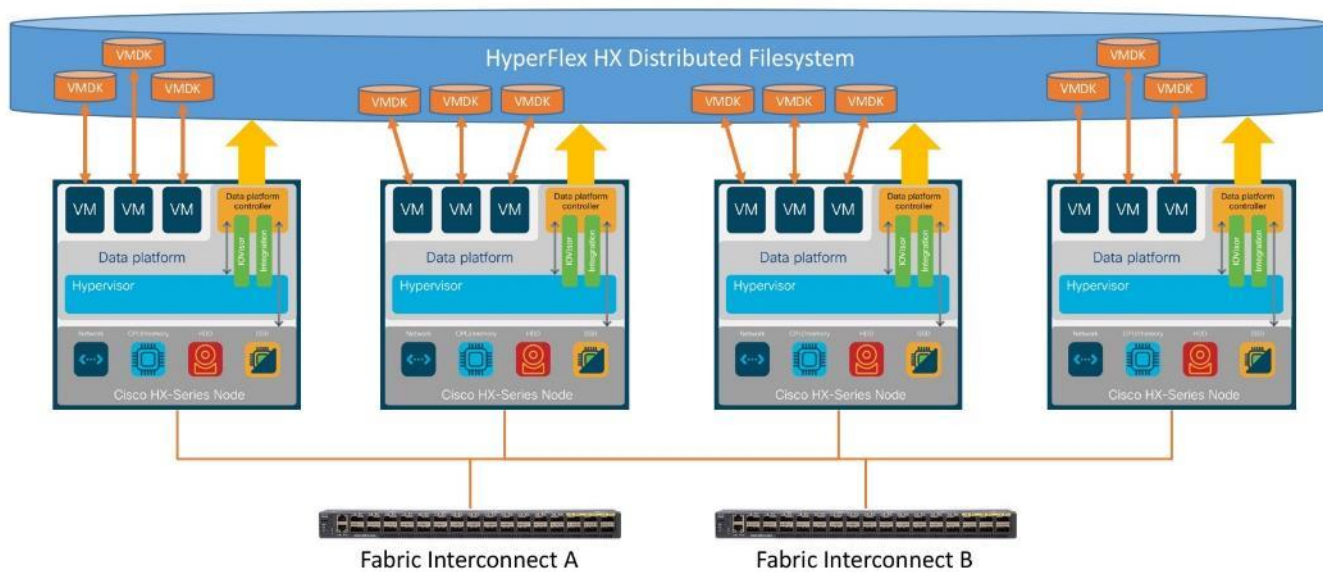
A login is required for the Documentation Roadmap.

For more information about Hyperconverged Infrastructure, go to: <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 6454 Fabric Interconnect
- Eight Cisco HyperFlex HX-Series Rack-Mount Servers, choose from models:
 - Cisco HyperFlex HXAF220c-M5SX All-Flash Rack-Mount Servers
- Cisco HyperFlex Data Platform Software
- VMware vSphere ESXi Hypervisor
- VMware vCenter Server (end-user supplied)

Technology Overview

Cisco Unified Computing System

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, 25 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, 25-Gbps or 40-Gbps unified network fabric, with an option for 100-Gbps uplinks. 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. Cisco UCS can also be managed by Cisco Intersight, a cloud-based management and monitoring platform which offers a single pane of glass portal for multiple Cisco UCS systems across multiple locations.

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 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. The product family supports Cisco low-latency, lossless 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.

Cisco UCS 6454 Fabric Interconnect

The Cisco UCS 6454 54-Port Fabric Interconnect is a One-Rack-Unit (1RU) 10/25/40/100 Gigabit Ethernet, FCoE and Fibre Channel switch offering up to 3.82 Tbps throughput and up to 54 ports. The switch has 28 10/25-Gbps Ethernet ports, 4 1/10/25-Gbps Ethernet ports, 6 40/100-Gbps Ethernet uplink ports and 16 unified ports that can support 10/25-Gbps Ethernet ports or 8/16/32-Gbps Fibre Channel ports. All Ethernet ports are capable of supporting FCoE. Cisco HyperFlex nodes can connect at 10-Gbps or 25-Gbps speeds depending on the model of Cisco VIC card in the nodes and the SFP optics or cables chosen.

Figure 2. Cisco UCS 6454 Fabric Interconnect



Cisco HyperFlex HX-Series Nodes

A standard HyperFlex cluster requires a minimum of three HX-Series “converged” nodes (such as nodes with shared 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.

Figure 3. HXAF220c-M5SX All-Flash Node



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 1.6 TB 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 960 GB or 3.8 TB SED SSDs.

Figure 4. HXAF220c-M5SX All-Flash Node



In HX-series all-flash nodes either a 375 GB Optane NVMe SSD, a 1.6 TB SAS SSD or 1.6 TB NVMe SSD caching drive may be chosen. While the Optane and 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. Caching amounts are not factored in as part of the overall cluster capacity, only the capacity disks contribute to total cluster capacity.

Cisco HyperFlex HX220c-M5SX Hybrid Node

This small footprint Cisco HyperFlex hybrid model contains a minimum of six, and up to eight 2.4 terabyte (TB), 1.8 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 5. HX220c-M5SX Node



Either a 480 GB SATA or 800 GB SAS caching SSD may be chosen. This option is provided to allow flexibility in ordering based on product availability, pricing, and lead times. While the SAS option may provide a slightly higher level of performance, the partitioning of the two disk options is the same, therefore the amount of cache available on the system is the same regardless of the model chosen. Caching amounts are not factored in as part of the overall cluster capacity, only the capacity disks contribute to total cluster capacity.

Cisco VIC 1457 MLOM Interface Cards

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 Cisco UCS VIC 1387 is used in conjunction with the Cisco UCS 6332 or 6332-16UP model Fabric Interconnects.

The Cisco UCS VIC 1457 is a quad-port Small Form-Factor Pluggable (SFP28) mLOM card designed for the M5 generation of Cisco UCS C-Series Rack Servers. The card supports 10-Gbps or 25-Gbps Ethernet and FCoE, where the speed of the link is determined by the model of SFP optics or cables used. The card can be configured to use a pair of single links, or optionally to use all four links as a pair of bonded links. The Cisco UCS VIC 1457 is used in conjunction with the Cisco UCS 6454 model Fabric Interconnect.

The mLOM is 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 6. Cisco VIC 1457 mLOM Card



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:

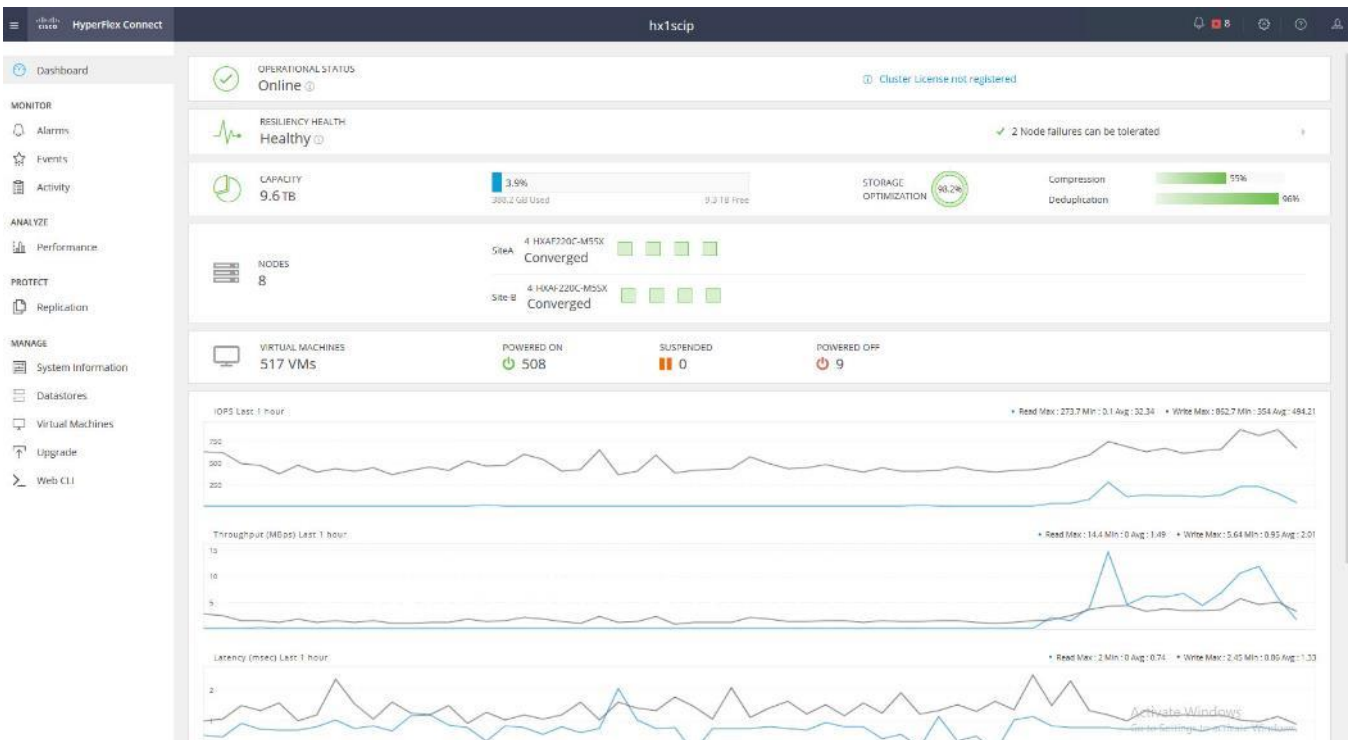
- **Data protection** creates multiple copies of the data across the cluster so that data availability is not affected if single or multiple components fail (depending on the replication factor configured).

-
- **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 HTML 5 based Web UI named HyperFlex Connect 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>.

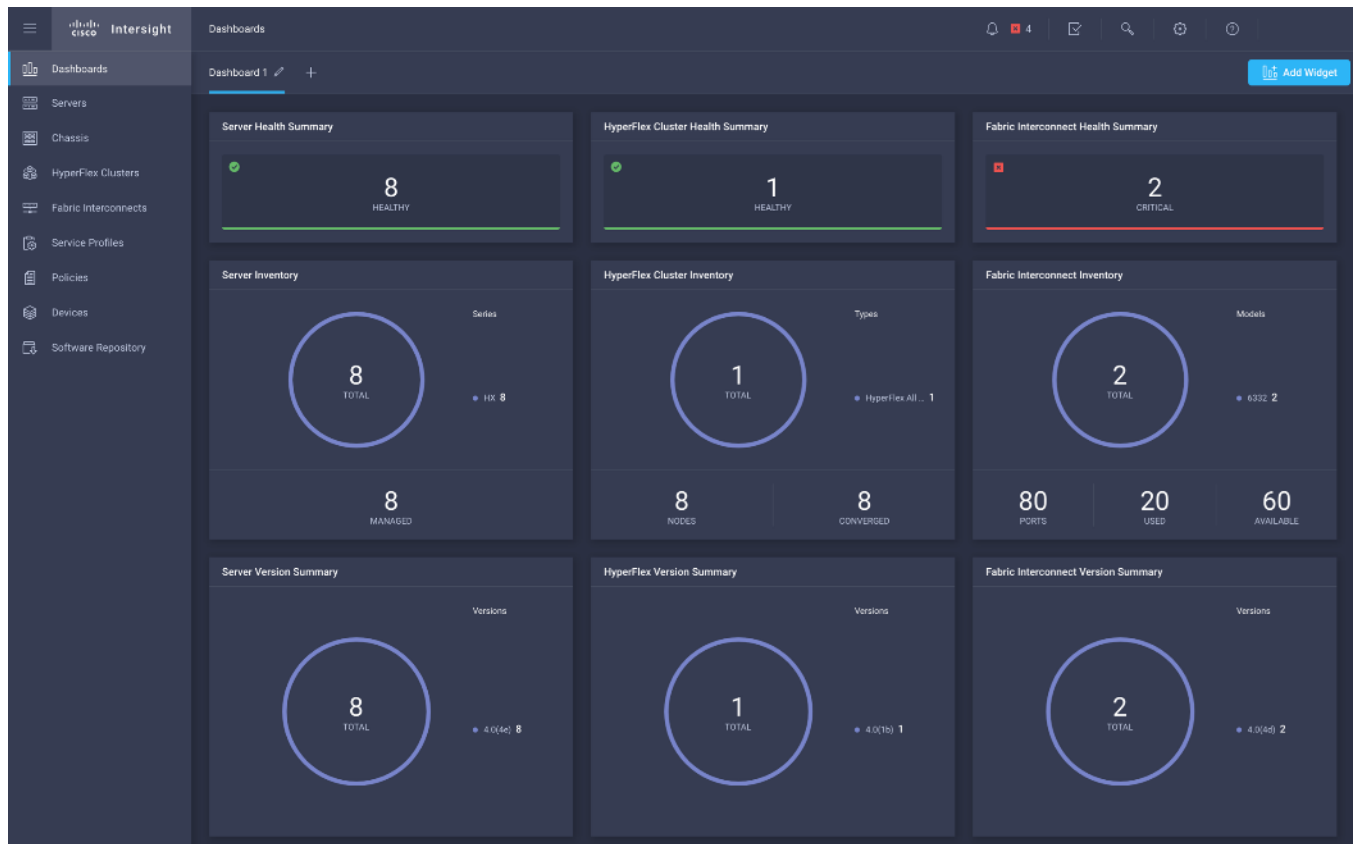
Figure 7. HyperFlex Connect GUI



Cisco Intersight Cloud Based Management

Cisco Intersight (<https://intersight.com>) 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, and can be used to deploy and manage Cisco HyperFlex clusters. Cisco Intersight offers direct links to Cisco UCS Manager and Cisco HyperFlex Connect for systems it is managing and monitoring. The Cisco Intersight website and framework is being constantly upgraded and extended 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. This unique combination of embedded and online technologies results in a complete cloud-based management solution that can care for Cisco HyperFlex throughout the entire lifecycle, from deployment through retirement.

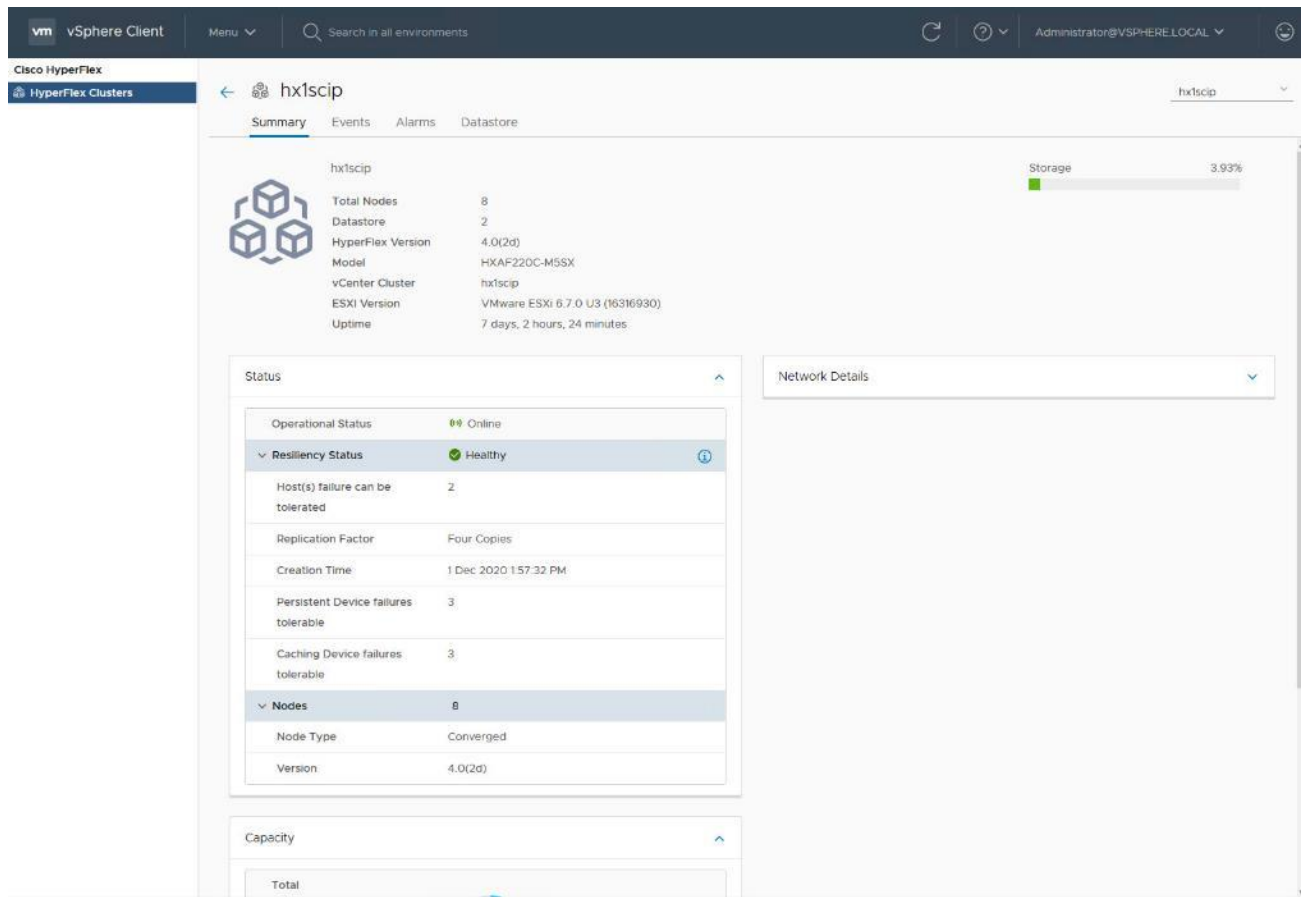
Figure 8. Cisco Intersight



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, which is either deployed automatically by the Cisco HyperFlex installer or the HTML 5 plug-in can be downloaded separately from CCO.

Figure 9. 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 direct PCI passthrough control of the physical server's SAS disk controller, or direct control of the PCI attached NVMe based SSDs. 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.

Cisco HyperFlex HX Stretch Clusters

This section provides an overview of Cisco HyperFlex stretch clusters. It details some of the business reasons for deploying such a cluster. It also discusses some of the physical limitations of such a cluster.

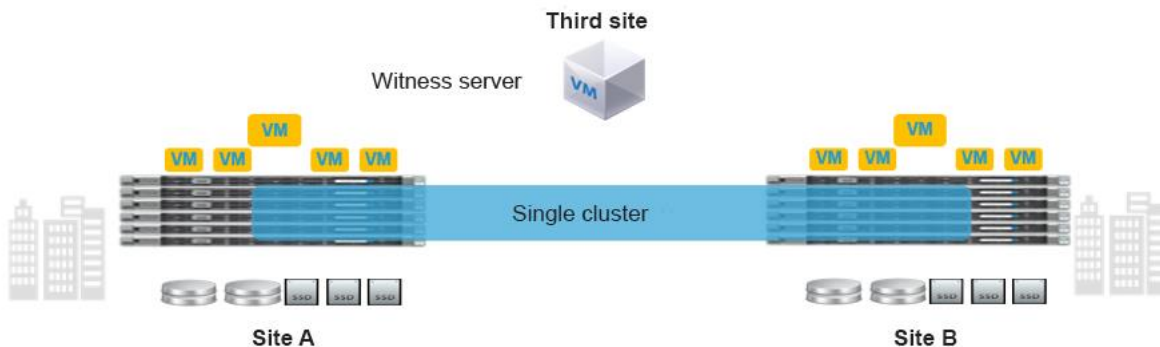
What is a Stretch Cluster?

A stretch cluster is distinct from a non-stretch, or normal, cluster, in that it is designed to offer business continuance in the event of a significant disaster at a data center location. A stretch cluster is geographically redundant, meaning that part of the cluster resides in one physical location and another part resides in a second location. The cluster also requires a “tie breaker” or “witness” component, which should reside in a third, separate location. The goal of this design is to help ensure that the virtual infrastructure remains available even in the event of the complete loss of one site. Of course, many lesser types of failures also can occur, and the system is highly available in the event of these as well. All of these scenarios are discussed later in this document.

People often mistakenly think that a stretch cluster is a set of multiple single clusters. This is not the case. A stretch cluster is, in fact, a single distributed entity and behaves as such in most circumstances. There are a few differences between a normal cluster and a stretch cluster, however. These arise solely from the fact that a stretch cluster must meet some special requirements to provide geographical redundancy for deployments that require it. Georedundancy introduces a few new requirements for the cluster so that certain conditions, such as split brain and node quorum, are handled properly. These are discussed in the following sections.

[Figure 10](#) shows the main features of a stretch cluster.

Figure 10. Three Main Components of a Stretch Cluster Deployment



The following are the characteristics of a stretch cluster:

- A stretch cluster is a single cluster with nodes geographically distributed at different locations.
- Storage is mirrored locally and across each site (but not to the tie-breaker witness).
- Sites need to be connected over a low-latency network to meet the write requirements for applications and for a good end-user experience.
- Geographic failover (virtual machine) is like failover in a regular cluster.
- Node failure in a site is like node failure in a regular cluster.

- Split brain is a condition in which nodes at either site cannot see each other. This condition can lead to problems if a node quorum cannot be determined (so that virtual machines know where to run). Split brain is caused by:
 - Network failure
 - Site failure
- Stretch clusters have a witness: an entity hosted on a third site that is responsible for deciding which site becomes primary after a split-brain condition.

Businesses Need a Stretch Cluster

Businesses require planning and preparation to help ensure business continuity after serious incidents or disasters and to resume normal operations within a reasonably short period. Business continuity is the capability of an organization to maintain essential functions during, as well as after, a disaster. It includes three main elements:

- **Resilience:** Critical business functions and the supporting infrastructure must be designed so that they are materially unaffected by relevant disruptions: for example, through the use of redundancy and spare capacity.
- **Recovery:** Organizations must have in place arrangements to recover or restore critical and less critical business functions that fail for some reason.
- **Contingency:** An organization must establish a generalized capability and readiness to allow it cope effectively with whatever major incidents and disasters may occur, including those that were not, and perhaps could not have been, foreseen. Contingency preparations constitute a last-resort response if resilience and recovery arrangements should prove inadequate in practice.

Stretch Cluster Physical Limitations

Some applications, specifically databases, require write latency of less than 20 milliseconds (ms). Many other applications require latency of less than 10 ms to avoid problems with the application. To meet these requirements, the round-trip time (RTT) network latency on the stretch link between sites in a stretch cluster should be less than 5 ms. The speed of light (3e8 m/s) at the maximum recommended stretch cluster site distance of 100 km (approximately 62 miles) introduces about 1 ms of latency by itself. In addition, time is needed for code path and link hops (from node to fabric interconnect to switch), which also plays a role in determining the maximum site-to-site recommended distance.

Solution Components

A traditional Cisco HyperFlex single-cluster deployment consists of HX-Series nodes in Cisco UCS connected to each other and the upstream switch through a pair of fabric interconnects. A fabric interconnect pair may include one or more clusters. A stretch cluster requires two independent Cisco UCS domains: one for each site. Therefore, a total of four fabric interconnects (two pairs) are required for a stretch cluster. Other clusters can share the same fabric interconnects.

[Figure 11](#) and [Figure 12](#) show typical physical layouts for this kind of deployment. [Figure 11](#) shows a single site with its cabling and independent Cisco UCS domain. [Figure 12](#) shows the racks for site A and site B in a stretch cluster with their respective fabric interconnects and upstream switches. This is an 8-node (4+4) stretch cluster with Cisco HyperFlex HX220c nodes at each location.

Figure 11. Site a for a Stretch Cluster Deployment Showing a Single-Site Rack: the Site Contains 4 HX220c M5 Nodes and 2 Fabric Interconnects with a Single Uplink Switch for the Stretch Layer 2 Network Connecting to Site B

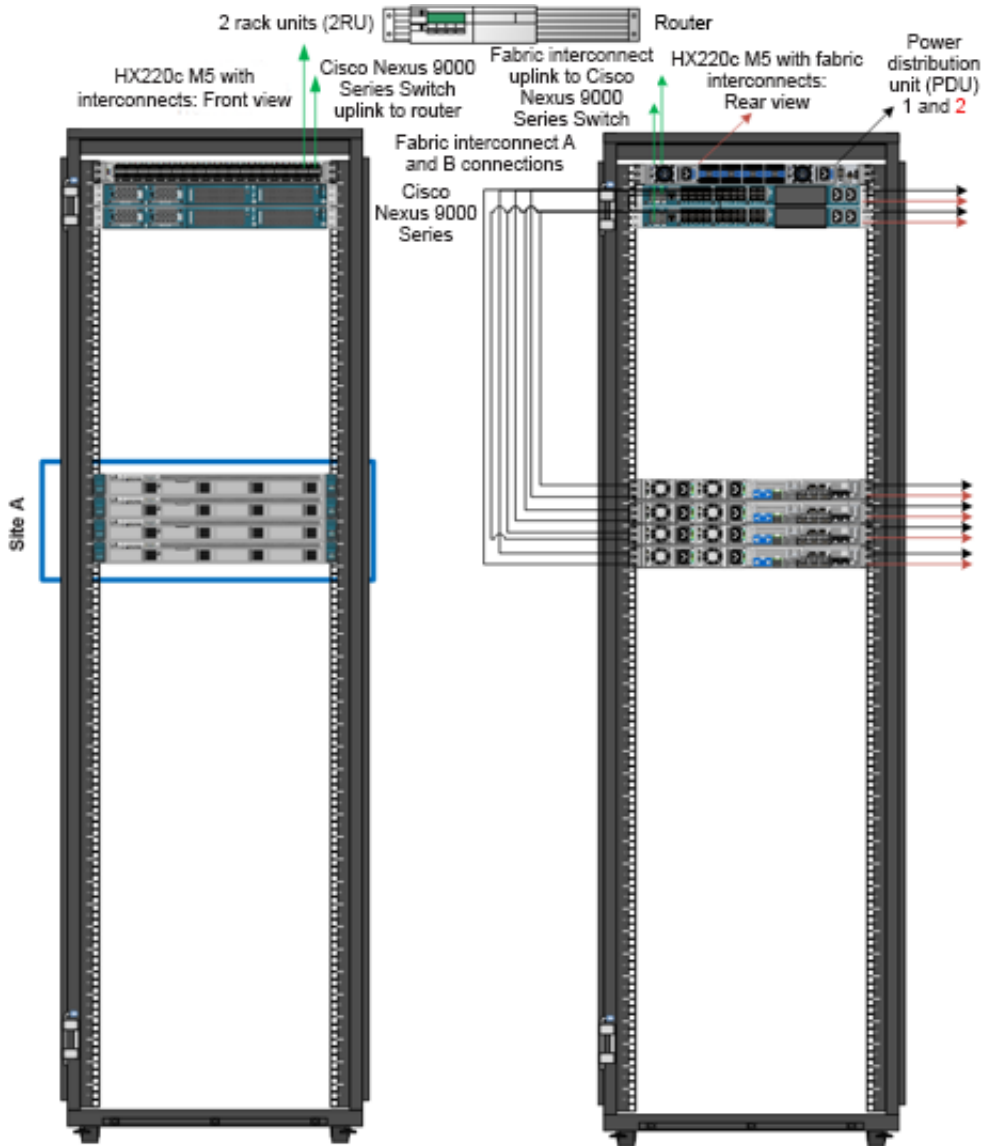
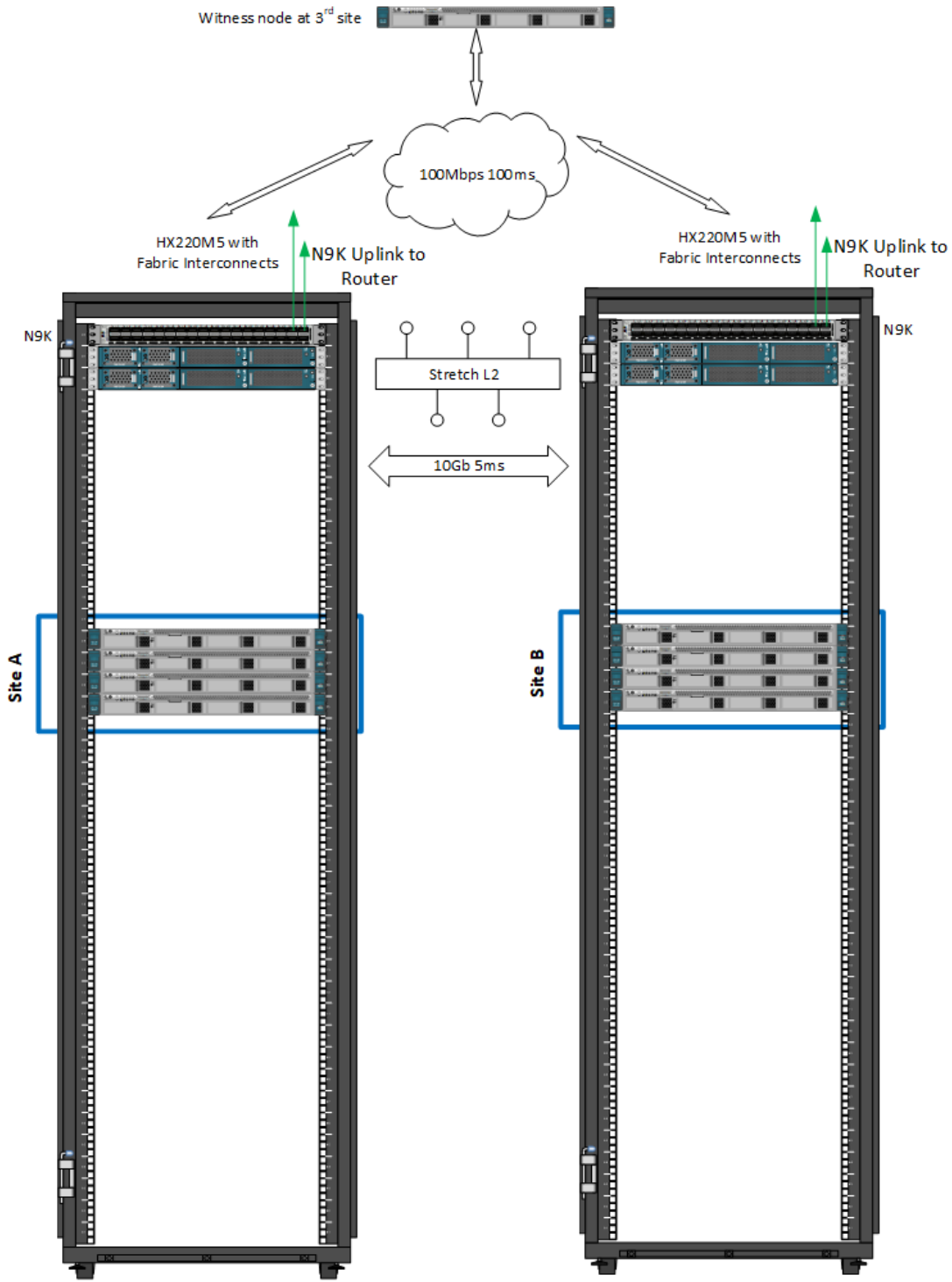


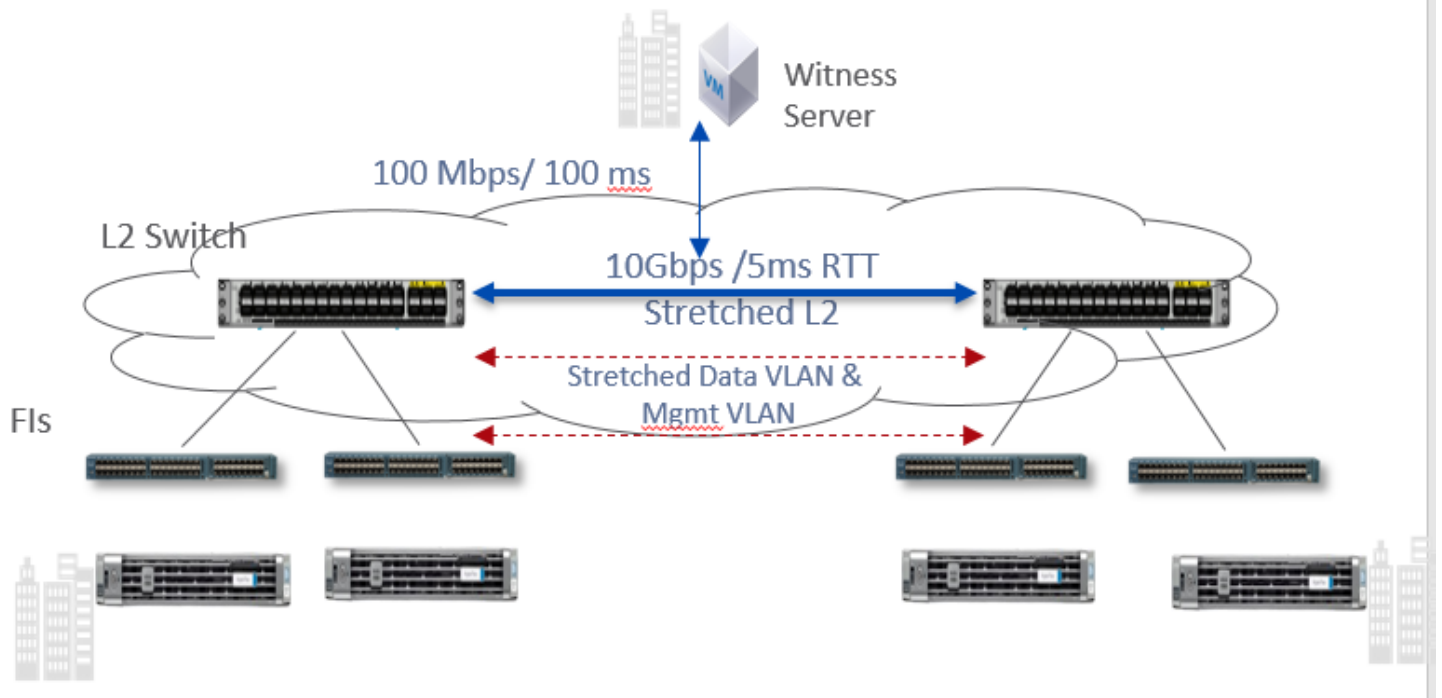
Figure 12. Rack Diagram Showing Site A and Site B with their respective Fabric Interconnects and a Logical Third Site at Another Location for the Stretch Cluster Witness



Stretch Cluster Architecture

This section discusses the specific deployment needs for a stretch cluster, including hardware, networking configuration, VMware requirements (ESXi and vCenter), failure sizing, and characteristics of the witness (Figure 13). VMware vSphere Enterprise Plus is required because Cisco HyperFlex stretch clusters rely on advanced DRS capabilities available only in that premium edition. The requirements are the same across all stacks (even for non-hyperconverged infrastructure [HCI] or traditional storage) that implement stretch or metropolitan clusters on VMware.

Figure 13. General Stretch Cluster Network



The first consideration in deploying a stretch cluster is building the proper site-to-site network. A stretch cluster requires a minimum of 10 Gigabit Ethernet connectivity and 5-ms RTT latency on the link. The link needs to be stretch Layer 2 to help ensure network space adjacency for the data storage VLAN network that is used for storage communication. The network between sites requires the following characteristics:

- 10 Gbps (dedicated) for the storage data VLAN.
- 5-ms RTT latency between the two active sites.
- Data VLAN and management VLAN on a stretch Layer 2 VLAN.
- Stretch Layer 2 VLAN between the two sites:
 - Dark fiber and dense wavelength-division multiplexing (DWDM) Layer 2 and 3 technologies are supported.
 - The solution is not currently qualified for Virtual Extensible LAN (VXLAN) unless used with ACI.
 - Stretch Layer 2 characteristics.

- The stretch data VLAN should use jumbo maximum transmission units (MTUs) for best performance. The installer allows for deployment using an MTU of 1500, however.
- The Cisco Nexus® 5000 Series Switches are slightly different than the Cisco Nexus 7000 and 9000 Series Switches. The default network-QoS policy does not accept jumbo MTUs, but you can set up jumbo switch policy across the switches.
- Test the RTT ping using **VMkping -l VMk1 -d -s 8972 x.x.x.x** from any ESXi host in your cluster. This check is also performed by the installer, and if it fails, the installation process will not proceed.
- 100 Mbps and 100-ms RTT latency between the active sites and the witness site.
- Different drives types are supported with different nodes limits. See the release notes for your running or target version to determine which drives and nodes you can use. For example, there are LFF drive restrictions and NVME drives began support in 4.0.2x and onward for the HX220 node type.

Deployment Prerequisites

Some deployment prerequisites exist for stretch clusters related to the qualified hardware. Most of these prerequisites are not based on technical factors but simply reflect test bandwidth and the release cycle. After these items have been qualified, they will be removed from the unsupported-features list, and these capabilities will be available for general deployment.



Check the minor version release notes periodically for changes in the support listings.

Minimum and maximum configuration limitations are as follows:

- Minimum
 - Two fabric interconnects per site
 - Two nodes per site
 - One witness
 - One vCenter instance
 - Replication factor: 2+2
- Maximum
 - Two fabric interconnects per site
 - 2:1 maximum ratio for compute to converged nodes
 - Compute nodes can be added asymmetrically with no restriction
 - 16 small-form-factor (SFF) converged nodes per site (32 total, max cluster 64 with compute)
 - 8 large-form-factor (LFF) converged nodes per site (16 total, max cluster 48 with compute)
 - One witness
 - One vCenter or vCenter with HA instance if there is no database update lag
 - Replication factor: 2+2

Stretch cluster support prerequisites are as follows:

- Self-encrypting drives (SEDs) are not supported.

- Compute-only nodes are supported in HyperFlex 3.5 or higher with a 2:1 ratio to converged nodes. Verify the ratio in the Release Notes for your version.
- ESXi is the only supported hypervisor at this time. Check the release notes for your HX version to see the recommended ESXi version.
- Cisco HyperFlex native replication is supported in HyperFlex 3.5 and greater.
- Expansion of an existing cluster to a stretch cluster is not supported.
- Stretch clusters are supported only in fresh installations. Upgrade from a standalone cluster to a stretch cluster configuration is not supported.
- Stretch Clusters must be symmetric (converged nodes). For production environments, this includes Fabric Interconnects.
- Stretch Clusters must be expanded symmetrically (converged nodes). See the admin guide for your version of HX for workflow details.
- Stretch Clusters can be built and/or expanded asymmetrically with compute nodes.
- Online rolling upgrades are supported only for the HX Data Platform. Cisco UCS Manager upgrades must be performed manually one node at a time.
- Stretch clusters are supported on Cisco M5 nodes only. M4 nodes are not supported.
- Logical availability zones are not currently supported in stretch clusters.
- The witness requires ESXi at the third site (cloud deployment is not currently supported).
- Disk reshuffling is not supported (for example, adding empty nodes and “leveling” the disks out)
- Hardware offload (acceleration) cards are supported starting in HXDP version 4.0.2b and greater
- Node removal is not supported
- Single Socket nodes may or may not be supported, depending on your version of HX. Please see the Release Notes.

About Zones

While logical availability zones are not currently supported in stretch cluster deployments, you may notice that zone information is available when running the `stcli cluster get-zone` command as show below:

```
root@SpringpathControllerOHCWUK9X3N:~# stcli cluster get-zone
zones:
-----
pNodes:
-----
state: ready
name: 192.168.53.136
-----
state: ready
```

```

name: 192.168.53.135
-----
zoneId: 51733a6b98df9784:4f8fc27070894bf4
numNodes: 2
-----
pNodes:
-----
state: ready
name: 192.168.53.138
-----
state: ready
name: 192.168.53.137
-----
zoneId: 7b04a6600e3e3ee5:54c7224773a14a9a
numNodes: 2
-----
isClusterZoneCompliant: True
zoneType: physical
isZoneEnabled: True
numZones: 2

```

LAZ and stretch cluster both are implemented using a basic feature called "zones" and that's why you see 'zone' in some of the output. You will not see "logical zones" which is what would appear under LAZ.

note the "zoneType" on the get-zone output.

On stretch cluster: "zoneType: physical"

On Cluster with LAZ : "zoneType: logical"

Hardware Matching

Stretch Clusters require identical hardware at both sites. This includes node count, type, and drives per node as well. This also applies to expansion. You must expand in converged node pairs.

There are some exceptions to the hardware symmetry requirement. Compute resources are not required to be symmetric between sites. You can have more compute-only nodes on one site than the other. However, care should be taken since a failure scenario from one site with large compute resources to another site with reduced resources may not be sized properly to run the VMs that are started on the surviving site.

Mixing CPU generations is supported within the same family as well. For example, it is ok to mix 8180 Skylake CPUs with 6258R Cascade Lake CPUs. You must, however, size for the less powerful CPU.

A Stretch Cluster will work, such as, deploy properly and functional as expected, if the FIs are different between sites, but identical within the site. This can be useful for lab and testing environments but is not supported by Cisco for production. FIs must be identical within a site and between sites for production.

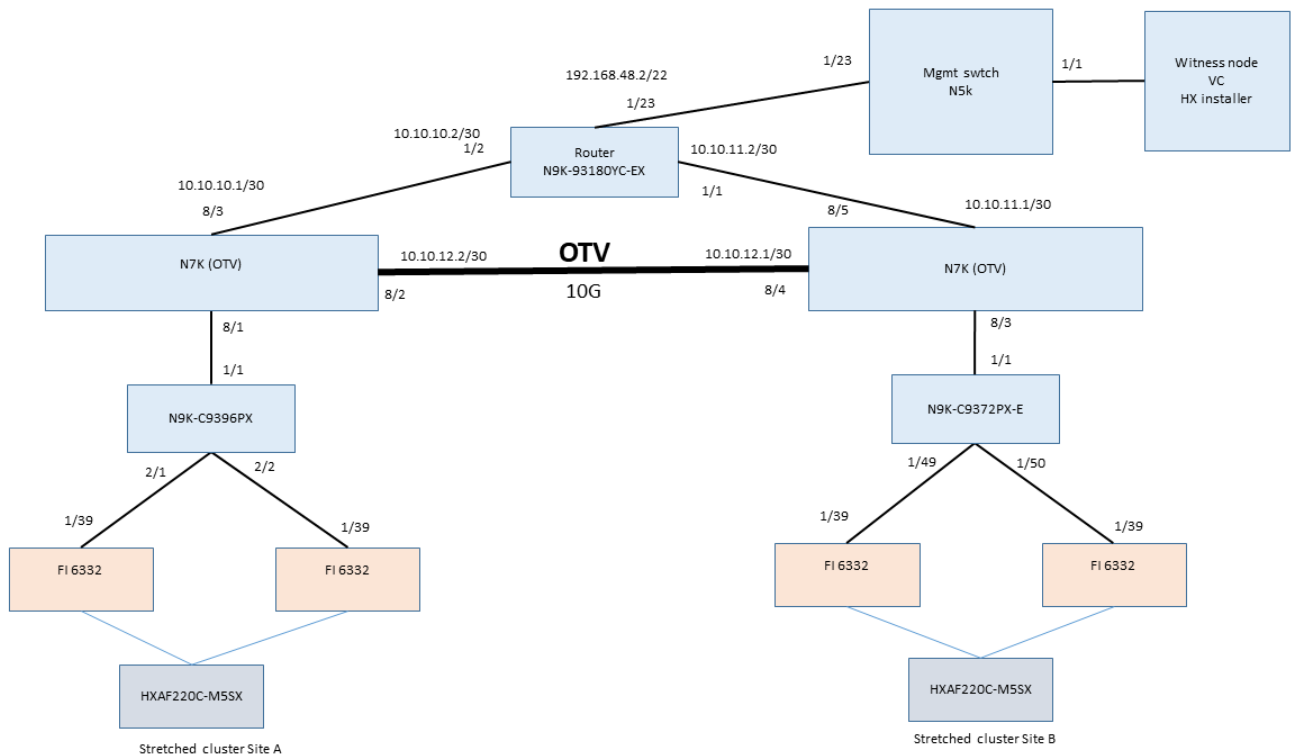
Overlay Networks



Only certain overlay networks are currently qualified for use with Stretch Clusters.

OTV is qualified for use with Stretch Cluster. This means however, that VXLAN and NSX are not supported as stand-alone overlay protocols. VXLAN is supported only with ACI. See the “More Information” section for the CVD describing this deployment.

Cisco Overlay Transport Virtualization (OTV), supported on Nexus Switches, is a networking technology that allows relaying layer 2 (L2) networks over layer 3 (L3) network segments. OTV is important for Stretch Clusters that require stretched L2 storage and management networks when a dedicated dark fiber type site-to-site connection is not available. The tested and validated OTV design is shown below.



This OTV design was tested for the various failure modes discussed later. It was configured to meet the bandwidth and latency requirements necessary for the proper operation of the Stretch Cluster. It is important to note that layering over L3 can introduce latency since the routed network will necessarily have additional device to device hops. When designing and deploying this type of architecture you must ensure that you are still within the site-to-site communication specification for bandwidth and latency.

The following references are for OTV on Cisco Nexus:

<https://www.cisco.com/c/en/us/solutions/data-center-virtualization/overlay-transport-virtualization-otv/index.html>

<https://community.cisco.com/t5/data-center-documents/understanding-overlay-transport-virtualization-otv/tag/3151502>

Fabric Interconnects

Stretch clusters have a specific set of fabric interconnect requirements. Each site is built using its own pair of fabric interconnects in an independent Cisco UCS domain. Therefore, a total of four fabric interconnects are required. The stretch cluster requires a symmetric deployment, meaning that each site must have the same number and type of fabric interconnects and converged nodes. If site A has 4 hybrid nodes, then site B must also have 4 hybrid nodes. As of Cisco HyperFlex 3.0, the maximum cluster size is 8 nodes per site, for a total of 16 (8 + 8). This has increased in 3.5 and above to 16 converged nodes per site (SFF) with up to a 2:1 compute node ratio for a maximum mixed count of 32 per site. Limits for LFF drives are different. See the release notes for your version of HX to get the latest information on the number and type of supported nodes.

Fabric interconnect and node configuration details are as follows:

- A total of four fabric interconnects are required, one pair at each site) in unique Cisco UCS domains.
- Do not mix fabric interconnect models within a domain.
- For the fabric interconnects, Cisco UCS Manager Release 3.2(3e) is required.
- Existing fabric interconnects are supported as long as they work with Cisco M5 nodes.
- Node requirements are as follows:
 - You must have the same number and type of nodes per site: All flash or all hybrid.
 - The maximum cluster size is 16 converged nodes per site starting in 3.5 with a 2:1 maximum compute ratio (max 32 mixed nodes per site).
 - These requirements and maximums change frequently, consult the Release Notes for your version.

Fabric Interconnects Uplink Best Practices

Care should be taken with all deployments of HX when uplinking the Fabric Interconnects to your TOR/edge switches. The best practice surrounding this is designed to make sure that Spanning Tree Protocol (STP) loops are avoided. In a normal cluster these loops will cause FI takeover problems. Due to the multi-domain nature of a stretch cluster, STP storms can bring the system down. When uplinking the FIs to your redundant switches, the virtual port channel (VPC) ports should be set to edge trunk mode so that they do not participate in STP.

This behavior is called out in several location within Cisco documentation but is reiterated here for convenience. For example, the following document call out using spanning-tree port type edge trunk or the need to disable spanning tree on ports connecting to the FIs from upstream switches:

- https://www.cisco.com/c/en/us/td/docs/hyperconverged_systems/HyperFlex_HX_DataPlatformSoftware/Net-work_External_Storage_Management_Guide/b_HyperFlex_Systems_Network_and_External_Storage_Manage-ment_Guide_3_0/b_HyperFlex_Systems_Network_and_External_Storage_Management_Guide_3_0_chapter_01.html

Cisco FIs appear on the network as a collection of endpoints versus another network switch. Internally, the FIs do not participate in spanning-tree protocol (STP) domains, and the FIs cannot form a network loop, as they are

not connected to each other with a layer 2 Ethernet link. The upstream root bridges make all link up/down decisions through STP.

Uplinks need to be connected and active from both FIs. For redundancy, you can use multiple uplinks 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, make uplinks 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. This also 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 unavailable.

When setting the uplinks from the FI as VPC port channels you also need to set the downlink ports, for example, on the Cisco Nexus 9k, to “spanning tree edge” instead of “spanning tree normal”, since the FIs don’t participate in STP. In the absence of this configuration, a spanning tree storm in the N9k will cause a traffic blackhole for HX storage traffic. This in turn will affect all HX traffic in a stretch cluster. In standard clusters, the problem happens only when there is an FI failover.

In clusters without the ability to use vPC or LACP based link aggregation for redundancy, you should use disjoint layer 2.

VMware vCenter

VMware vCenter is a critical component for normal clusters and is vital for a stretch cluster. vCenter, with HA and DRS configured automatically manages virtual machine movement in the event of a site failure. The use of virtual machine host groups in the preferred mode, in which virtual machines are pinned to a site for the purpose of local computing and read I/O, is required for optimal performance in a stretch deployment. Site host groups and the corresponding affinities are created automatically at build time by the Cisco HyperFlex installer.

Data stores also maintain site affinity using host groups as the mechanism to locate the primary copy of virtual machine data. This approach is used to facilitate the asymmetric I/O mechanism that a stretch cluster uses to increase the cluster response time by localizing read I/O while distributing write I/O (two local-site copies and two remote-site copies). Because both sites in a stretch cluster are active, virtual machines at one site or the other do not suffer any “second-class citizen” type scenarios, in which one site has preferential performance relative to another.

In a stretch cluster deployment, a single instance of vCenter is used for both sites. The best approach is to locate this instance at a third location so that it is not affected by site loss. Co-residency with the witness is often the preferred choice because the witness site is required anyway. Nested vCenter (such as, running the cluster’s vCenter instance on the cluster itself) is not supported. vCenter HA (VCHA) is supported with Stretch Cluster. Be aware the VCHA is a high availability deployment of vCenter itself and does not refer to the enabling HA on vCenter (which is a separate requirement for proper cluster failover behavior).

In the vCenter instance, the stretch cluster corresponds to a single ESXi cluster. Be sure to verify that HA and DRS are set up for the stretch cluster.

If the need arises to move the cluster from one vCenter to a new vCenter deployment or a different existing vCenter instance, it will be necessary to perform a cluster re-register. Be sure to see the admin guide for detailed notes, but the general workflow is as follows: Create the cluster object in the new vCenter instance and add the cluster ESXi hosts manually. Be sure the HA/DRS is enabled. The re-register is conducted using STCLI from any node or the CIP-M address.

```
admin@ControllerE2L5LYS7JZ:~$ stcli cluster reregister
```

usage: stcli cluster reregister [-h] --vcenter-datacenter NEWDATACENTER

--vcenter-cluster NEWVCENTERCLUSTER

--vcenter-url NEWVCENTERURL

[--vcenter-sso-url NEWVCENTERSSOURL]

--vcenter-user NEWVCENTERUSER

stcli cluster reregister: error: argument --vcenter-datacenter is required

In a non-stretched cluster this is all that is required to remove the cluster from one vCenter instance and move it to a new one. A stretch cluster, however, requires a few manual steps to complete the process. This is because Host Groups and Affinity Rules are not transferred in the re-registration process. Please note that ICPM needs to be accessible between hosts and vCenter for re-registration to function properly.

A stretch cluster relies on a specific naming convention when interfacing with vCenter for implementation of the affinity rules. This is set up automatically, in advance, with the HX Installer when the cluster sites are built. The host group and affinity group naming must follow this convention: <site name>_{HostGroup, VmGroup, SiteAffinityRule} when rebuilding the groups and rules on the new vCenter host. See the screens below for an example. Here, site 1 is called fi47 and site 2 is fi48. Note the naming convention.

The screenshot shows the vSphere Web Client interface for a cluster named 'sed1-cl'. The 'Configure' tab is active, and the 'VM/Host Groups' section is selected in the left-hand navigation pane. The main content area displays a table of VM/Host Groups with columns for Name and Type. Below the table, there is a section for 'fi47_VmGroup Group Members' listing several VMs.

Name	Type
fi47_VmGroup	VM Group
fi47_HostGroup	Host Group
fi48_VmGroup	VM Group
fi48_HostGroup	Host Group

fi47_VmGroup Group Members	
ubu-vdbench-852-1-clone4	
ubu2-fio-844-1-clone1	
ubu2-fio-850-1-clone3	

sed1-cl | ACTIONS ▾

Summary Monitor **Configure** Permissions Hosts VMs Datastores Networks Updates

- Services
 - vSphere DRS
 - vSphere Availability
- Configuration
 - Quickstart
 - General
 - Licensing
 - VMware EVC
 - VM/Host Groups
 - VM/Host Rules**
 - VM Overrides
 - Host Options
 - Host Profile
 - I/O Filters
- More

VM/Host Rules

+ Add... ✎ Edit... ✖ Delete

Name	Type	Enabled	Conflicts	Defined By
fi47_AffinityRule	Run VMs on Hosts	Yes	0	System
fi48_AffinityRule	Run VMs on Hosts	Yes	0	System

VM/Host Rule Details

Virtual Machines that are members of the VM Group should run on hosts that are members of the Host Group.

+ Add... ✖ Remove

VMware vCenter HA Settings

The settings below are recommended for use in HX Stretch Clusters. This table details the most common settings in vSphere HA that are typically asked about during custom configuration. The screenshots are representative of vCenter 6.5. The cluster will work as designed using the default installation values. If you do not see a value listed below, keep it at the default.

vSphere HA Settings

vSphere HA Turn on HA. Keep Proactive HA disabled.

HXSC-Purple - Edit Cluster Settings

- vSphere DRS
- vSphere Availability**
- Failures and Responses
- Proactive HA Failures and Responses
- Admission Control
- Heartbeat Datastores
- Advanced Options

vSphere Availability

vSphere Availability is comprised of vSphere HA and Proactive HA. To enable Proactive HA you must also enable DRS on the cluster.

Turn ON vSphere HA

Turn on Proactive HA ⓘ

Failure	Response	Details
Host failure	✔ Restart VMs	Restart VMs using VM restart priority ordering.
Proactive HA	✖ Disabled	Proactive HA is not enabled.
Host Isolation	✖ Disabled	VMs on isolated hosts will remain powered on.
Datastore with Permanent Device Loss	✖ Disabled	Datastore protection for All Paths Down and Permanent Device Loss is disabled.
Datastore with All Paths Down	✖ Disabled	Datastore protection for All Paths Down and Permanent Device Loss is disabled.
Guest not heartbeating	✖ Disabled	VM and application monitoring disabled.

Host Monitoring

Enabled

HXSC-Purple - Edit Cluster Settings

Failures and Responses

Failure conditions and responses

You can configure how vSphere HA responds to the failure conditions on this cluster. The following failure conditions are supported: host, host isolation, VM component protection (datastore with PDL and APD), VM and application.

- Enable Host Monitoring**
- Host Failure Response: Restart VMs
- Response for Host Isolation: Disabled
- Datastore with PDL: Disabled
- Datastore with APD: Disabled
- VM Monitoring: Disabled

Virtual Machine Monitoring

Customer Preference – Disabled by default

HXSC-Purple - Edit Cluster Settings

VM Monitoring

Response delay: 3 minutes

Enable heartbeat monitoring

VM monitoring resets individual VMs if their VMware tools heartbeats are not received within a set time. Application monitoring resets individual VMs if their in-guest heartbeats are not received within a set time.

- VM Monitoring**
Turns on VMware tools heartbeats. When heartbeats are not received within a set time, the guest OS is restarted.
- Application Monitoring**
Turns on application heartbeats. When heartbeats are not received within a set time, the guest OS is restarted.

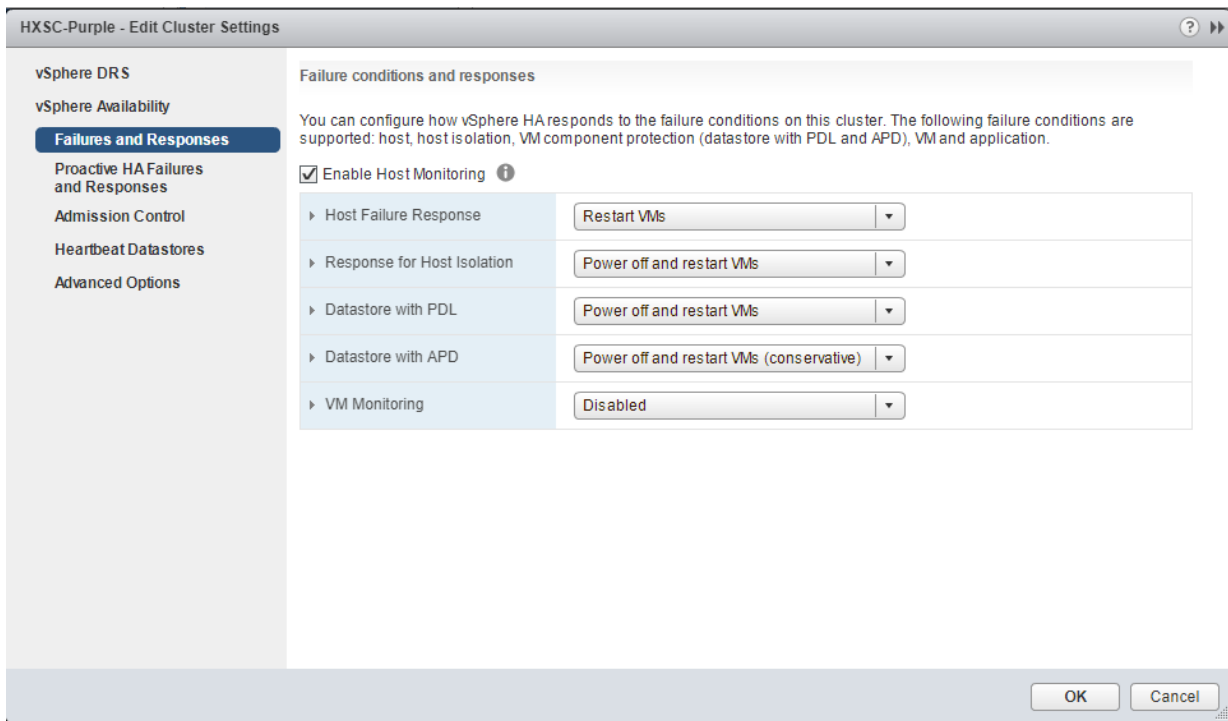
Heartbeat monitoring sensitivity

- Preset**
Low — High
- Custom**
 - Failure interval: 30 seconds
 - Minimum uptime: 120 seconds
 - Maximum per-VM resets: 3
 - Maximum resets time window: No window Within 1 hrs

OK Cancel

Failure conditions and VM Response

Host monitoring is enabled, Response for Host Isolation is set to Power off and Restart VMs. For PDL and APD, select Power off and Restart from the drop-down lists.



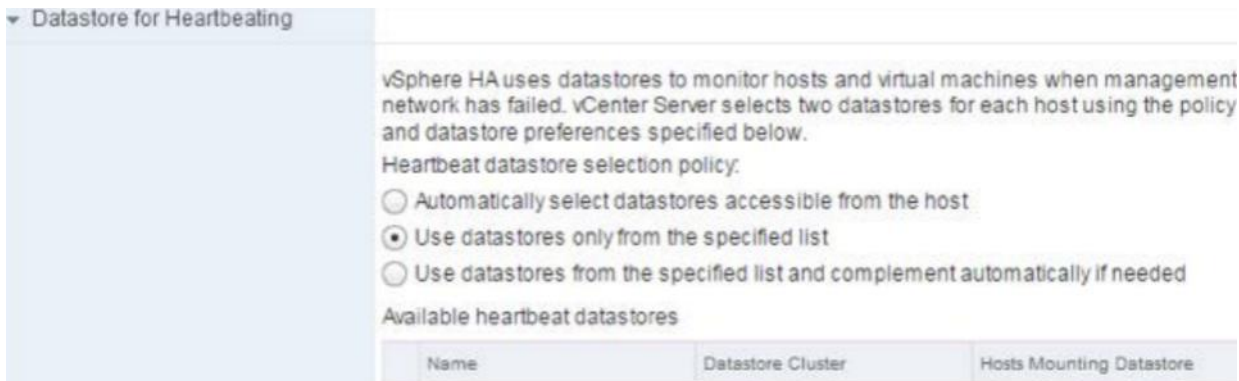
Admission Control

Set to disable

Datastore Heartbeats

“Use datastores only from the specified list” and select HX datastores.

<https://kb.vmware.com/s/article/2004739>



Advanced Settings

das.usedefaultisolationaddress

False

das.isolationaddress0	IP address for Management Network Gateway
das.isolationaddress1	Existing IP address that is outside cluster. Do not use FI VIPs, Cluster IP (CIP), or cluster host IP

Witness Configuration

A quorum is the minimum number of votes that a distributed transaction must obtain to be allowed to perform an operation in a distributed system. A quorum-based technique is implemented to enforce consistent operation in a distributed system. The witness node serves this function. In the event of a split-brain condition, in which both sites are still available but unable to communicate with each other, a virtual machine site leader must be established so that two instances of the same virtual machine are not brought online by HA.

The witness is deployed at a third site and is delivered as an open virtual appliance (OVA) file for use in an infrastructure ESXi deployment at that location. The witness runs an instance of ZooKeeper and becomes a cluster member and contributes its vote when needed to break a tie.

The witness node must have the following characteristics:

- A third independent site is needed to host the witness virtual machine.
- IP address and connectivity for the witness virtual machine is needed to each stretch cluster site.
- The witness must be on a routable Layer 3 network.
- The minimum requirements for the witness node are as follows:
 - Virtual CPUs (vCPUs): 4
 - Memory: 8 GB
 - Storage: 40 GB
 - HA: Optional for the witness node
- Latency of at most 100-ms RTT to each site is required.
- Bandwidth of at least 100 Mbps to each site is required.
- For fastest site-to-site failover times, an RTT latency to the witness of less than 10ms is optimal.
- The node must be deployed separately before the Cisco HyperFlex installer stretch cluster workflow is run.
- The witness behaves as a quorum node, if you are reinstalling the cluster the witness must be reinstalled as well.



There is one witness per cluster. Multiple clusters cannot use the same witness.

While no user data is being sent between the sites and the witness, some storage-cluster metadata traffic is transmitted to the witness site. This traffic is the reason for the 100-Mbps requirement and is in line with competitive products. The witness connection to each site requires 100 Mbps bandwidth with a 100 ms RTT in order to function properly. It is recommended to use a connection with a 100 ms latency for proper system failover behavior. For large clusters and for the best site-to-site failover performance, Cisco recommends witness-to-site latency on the order of 10 ms.

The witness is currently not supported in cloud deployments because of testing limitations. The OVA file has been tested and is supported for the ESXi platform.

If you need to patch the witness virtual machine for any reason, you can take the witness offline temporarily, implement the update, and bring the witness back online. Cisco recommends that you stage this process and practice it on a test witness to help ensure timely reintroduction of the production system when you implement the actual update. The cluster must be in a healthy condition to conduct this operation. If you need assistance, please contact the Cisco Technical Assistance Center (TAC).

I/O Path in a Stretch Cluster

A stretch cluster is in active-active mode at each site: that is, primary copies and read traffic occur for each virtual machine at each site. There is no concept of an active-standby configuration in a stretch cluster. IO Visor, the Cisco HyperFlex file system proxy manager, dictates which nodes service which read and write requests. In general, a stretch cluster behaves the same way as a normal cluster with modifications for host affinity and certain failure scenarios (see section [Stretch Cluster Failure Modes](#)). With virtual machine affinity and a replication factor of 2 + 2, the read and write dynamics are as described in the following sections.

Read Path

Taking advantage of the host group affinity, all read operations for virtual machine data are served locally, meaning that they come from the nodes at the site to which the data store for the virtual machine is assigned. Read operations are first serviced by the node cache if they are available there. If they are not available, they are read from persistent disk space (in a hybrid node) and served to the end user. The read cache in a stretch cluster behaves the same way as in a normal hybrid or all-flash cluster with the exception of local service based on host affinity.

Write Path

Write operations in a stretch cluster are a little more complicated than read operations. This is the case because to achieve data integrity, a write operation is not acknowledged as committed to the virtual machine guest operating system until all copies, local and remote, are internally committed to disk. This means that a virtual machine with affinity to site A will write its two local copies to site A while synchronously writing its two remote copies to site B. Again, IO Visor determines which nodes are used to complete each write operation.

The Cisco HyperFlex file system waits indefinitely for write operations to be acknowledged from all active copies. Thus, if certain nodes or disks that host a copy of data for which a write operation is being implemented are removed, write operations will stall until a failure is detected (based on a timeout value of 10 seconds) or the failure heals automatically without detection. There will be no inconsistency in either case.

I/O operations from virtual machines on site A will be intercepted by IO Visor on site A. IO Visor on site B is not be involved. The write I/O operations are replicated to site B at the data platform level. In the event of virtual machine migration from one site to another—for example, through VMware Storage vMotion from site A to another data store with affinity to site B—IO Visor will conduct a hand-off. When a virtual machine migrates to site B, IO Visor on site B will intercept the I/O operations. This procedure is also part of the virtual machine failover process internally. After the virtual machines have migrated from site A to site B, virtual machine I/O operations will not be intercepted by the site A IO Visor, but rather by the site B IO Visor.

Sizing

Typically, you start sizing exercises by profiling the workload or already knowing the requirements for the virtual machines that you need to run. However, you come by this information, the next step is to use a sizing tool (un-

less you want to do the math yourself). Cisco provides a sizing tool that can run workload estimates for a stretch cluster with a typical VSI profile:

Cisco HyperFlex sizer tool: <https://HyperFlexsizer.cloudapps.cisco.com/ui/index.html#/scenario>

Sizing a stretch cluster requires an understanding of the replication factor used for data protection. Each site runs a replication factor of 2: that is, each site has a primary copy and a replica. Each site also runs a replication factor of 2 for the complementary site, so that for each virtual machine, across both sites, there is a primary copy and three replicas: equivalent to a replication factor of 4. This configuration is required so that any individual site can tolerate the loss of its complementary site and still be able to run. Note that the loss of a site does not guarantee the ability of the surviving site to tolerate a disk or node loss because the affected node might be a zookeeper node. When the cluster is created, a zookeeper leader is elected at a given site. The leader is used to make updates to the ensemble. In the event of a site or zookeeper leader failure, a new leader is elected. This is not configurable.

Survivability while maintaining online status requires a majority zookeeper quorum and more than 50% of nodes (the witness counts as both an active zookeeper node). It is possible that the surviving site could tolerate a node or disk loss (in a cluster greater than 2+2) if that node is not a zookeeper node, but it is not guaranteed.

The data protection and workload profile (I/O requirement) considerations allow you to determine the number and type of disks required to meet your capacity needs. You then need to determine the node count needed to meet your vCPU and virtual machine memory needs.

Here are some sizing guidelines:

- For VSI an option is available in the sizer for selecting the stretch cluster. Use this option for your sizing exercises.
- In general, a stretch cluster uses a replication factor of 4: that is, replication factor 2 + replication factor 2 (a replication factor of 2 at each site with full replication to the complementary site, also at a replication factor of 2). This configuration effectively results in a replication factor of 4.
- You can use a replication factor of 2 for one site and then apply the same factor to the second site. If you want to be able to run all workloads from either site, then you must be sure that you have enough capacity at each site by accounting for the overall workloads and thresholds. The sizer automatically performs this verification for you.
- Consider the virtual machine and vCPU capacity: everything must be able to run comfortably at one site.
- The total virtual machine vCPU capacity is required.
- The total virtual machine memory capacity is required.

Failure Sizing

It is not enough to size your deployment for normal operations. Ideally, you should size your deployment for a scenario in which you have lost a site and the surviving site has lost a non-zookeeper node. This is the worst-case continuous-operation scenario for resource distribution to your overall virtual machine workload. Everything must be able to run comfortably on one site for a stretch cluster deployment to offer true business continuance in the event of a disaster.

If it is sufficient to run only certain virtual machines at the surviving site, you may be able to undersize the system, but you need to be aware of this and take it into consideration when planning disaster-recovery runbooks. Keep in mind that the automated recovery mechanism of the stretch cluster will launch virtual machines from failed sites without user intervention. You may find yourself in a situation in which you need to turn off failover virtual machines if they exceed the capacity of the surviving site.

Bandwidth Considerations for the Inter-Site Link Based on Workloads

Read bandwidth is normally local only, so there is no dependence or impact on the site-to-site link. Non-local VMs, such as, VMs running on nodes that do not have the assigned datastore affinity, will incur link read traffic. This is not the typical situation but should be considered in corner-case scenarios.

Write bandwidth is necessarily relevant to the link: Replicas traverse the link (2 copies). There is also meta data overhead for the filesystem that traverses the link making the write bandwidth some multiplier greater than 2. A typical good estimate is 2.2.

Workloads are almost never 100% read or 100% write. Typical benchmarks use a 70% Read/ 30% Write workload distribution. This means that for a 100,000 IOPS workload, 70,000 would be reads and 30,000 would be writes with a typical block size of 4k in the application. While the cluster writes do not map one-to-one with application writes (they are concatenated and written in chunks), the overall size of the write(s) match.

Link Bandwidth = $WIOPS(2 \text{ replicas})(0.2 \text{ metadata overhead})(4\text{kB}) + RIOPS(4\text{kB}) + ResynchIOPS(2 \text{ replicas})(4\text{kB}) + vMotionBW$

Where WIOPS are Write IOPS, RIOPS are Read IOPS, ResynchIOPS are resynchronization operations from any potential failure recoveries, and vMotionBW is the bandwidth taken up by a VM move (both compute and storage to account for datastore affinity when moved between sites). Resynchronizations only happen on failure recovery and are transitory operations so we will ignore them here. Storage vMotion is also typically not undertaken, but we will consider it in the example below.

Example: 20,000 IOPS total cluster workload, one affinity-displaced VMs contributing 1000 IOPS in 4kB Reads, no resynchronization, and 1 full SVMotion running at 500Mb/s. Assume 70/30 breakdown for the read and writes.

Link BW = $0.7(20000)(2)(0.2)(4\text{kB}) + 1000(4\text{kB}) + 0 + 500\text{Mb/s}$

Link BW = $123,200 \text{ kB/s} + 4000 \text{ kB/s} + 500\text{Mb/s} = 127,200 \text{ kB/s} + 500 \text{ Mb/s} = (127,200) * 8/1024 \text{ Mb/s} + 500 \text{ Mb/s}$

Link BW = $993.4 \text{ Mb/s} + 500 \text{ Mb/s}$

Link BW $\approx 1500 \text{ Mbps}$

Since you will not often do resync or vMotion, this can be considered a peak link value for the 20000 IOPS workload examined. There are times, for example, during large, frequent deletes, where the file system cleaner can incur larger metadata traffic on the link. To estimate those, you can use a multiplier of 1 to 1.5 instead of 0.2 for the (temporary peak) metadata value.

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. This solution's stretch cluster will have a four nodes on each site.

Physical Components

Table 1. HyperFlex System Components

Component	Hardware Required
Fabric Interconnects	Four Cisco UCS 6454 Fabric Interconnects (Two at each Site)
Servers	Eight Cisco HyperFlex HXAF220c-M5SX All-Flash rack servers (Four at each Site)

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

For the HXAF220c-M5SX Spec sheet, go to:

<https://www.cisco.com/c/dam/en/us/products/collateral/hyperconverged-infrastructure/hyperflex-hx-series/hxaf-220c-m5-specsheet.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 2 nd Generation Intel Xeon 6230 Processor Scalable Family CPUs
Memory		786 GB total memory using 64 GB DDR4 2933 MHz 1.2v modules depending on CPU type
Disk Controller		Cisco 12Gbps Modular SAS HBA
SSDs	Standard	One 240 GB 2.5 Inch Enterprise Value 6G SATA SSD 1.6 TB 2.5 Inch Extreme Performance SAS 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
Network		Cisco UCS VIC1387 VIC MLOM, or

HXAF220c-M5SX options	Hardware Required
	Cisco UCS VIC1457 VIC MLOM
Boot Device	One 240 GB M.2 form factor SATA SSD
microSD Card	One 32GB microSD card for local host utilities storage (Not used in this study)
Optional	




Software Components


The software components of the Cisco HyperFlex system must meet minimum requirements for the Cisco UCS firmware, hypervisor version, and the Cisco HyperFlex Data Platform software in order to interoperate properly.

For additional hardware and software combinations, refer to the public Cisco UCS Hardware Compatibility here: <https://ucshcltool.cloudapps.cisco.com/public/>

[Table 3](#) lists the software components and the versions required for the Cisco HyperFlex 4.0 system.

Table 3. Software Components

Component	Software Required
Hypervisor	<p>VMware ESXi 6.7 Update 3</p> <p>Cisco Custom Image for ESXi 6.7 Update 3 for HyperFlex: HX-ESXi-6.7U3-16316930-Cisco-Custom-6.7.3.3-install-only.iso</p> <hr/> <p> Using 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.</p> <p> ESXi 6.0 is not supported on servers equipped with the Cisco VIC1457 card, or the HXAF220c-M5N model servers. Each of these requires ESXi 6.5 Update 3 or higher.</p> <p> VMware vSphere Standard, Essentials Plus, ROBO, Enterprise or Enterprise Plus licensing is required from VMware.</p> <hr/>
Management Server	<p>VMware vCenter Server for Windows or vCenter Server Appliance 6.0 U3c or later.</p> <p>Refer to http://www.vmware.com/resources/compatibility/sim/interop_matrix.php for interoperability of your ESXi version and vCenter Server.</p>

Component	Software Required
	 Using ESXi 6.5 on the HyperFlex nodes also requires using vCenter Server 6.5. Accordingly, using ESXi 6.7 hosts requires using vCenter Server 6.7.
Cisco HyperFlex Data Platform	Cisco HyperFlex HX Data Platform Software 4.0(2b)
Cisco UCS Firmware	Cisco UCS Infrastructure software, B-Series and C-Series bundles, revision 4.0(4g) or later.

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, go to:
<https://www.cisco.com/c/en/us/buy/smart-accounts/software-manager-satellite.html>

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 desired features to be activated and used in the system, licenses must be purchased 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 4](#) lists an overview of the licensing editions, and the features available with each type of license.

Table 4. HyperFlex System License Editions

HyperFlex Licensing Edition	Edge	Advantage (in addition to Edge)	Premier (in addition to Standard)
Features Available	HyperFlex Edge clusters without Fabric Interconnects 220 SFF model servers only Hybrid or All-Flash ESXi Hypervisor only	HyperFlex standard clusters with Fabric Interconnects 220 and 240 SFF server models and 240 LFF server models	Stretched clusters 220 all-NVMe server models Cisco HyperFlex Acceleration Engine

HyperFlex Licensing Edition	Edge	Advantage (in addition to Edge)	Premier (in addition to Standard)
	Replication Factor 2 only 1 Gb or 10 Gb Ethernet only Compression Deduplication HyperFlex native snapshots Rapid Clones HyperFlex native replication Management via vCenter plugin, HyperFlex Connect, or Cisco Intersight	Replication Factor 3 Hyper-V and Kubernetes platforms Cluster expansions Compute-only nodes up to 1:1 ratio 10 Gb, 25 Gb or 40 Gb Ethernet Data-at-rest encryption using self-encrypting disks Logical Availability Zones	cards Compute-only nodes up to 2:1 ratio

For a comprehensive guide to licensing and all the features in each edition, consult the Cisco HyperFlex Licensing Guide here:

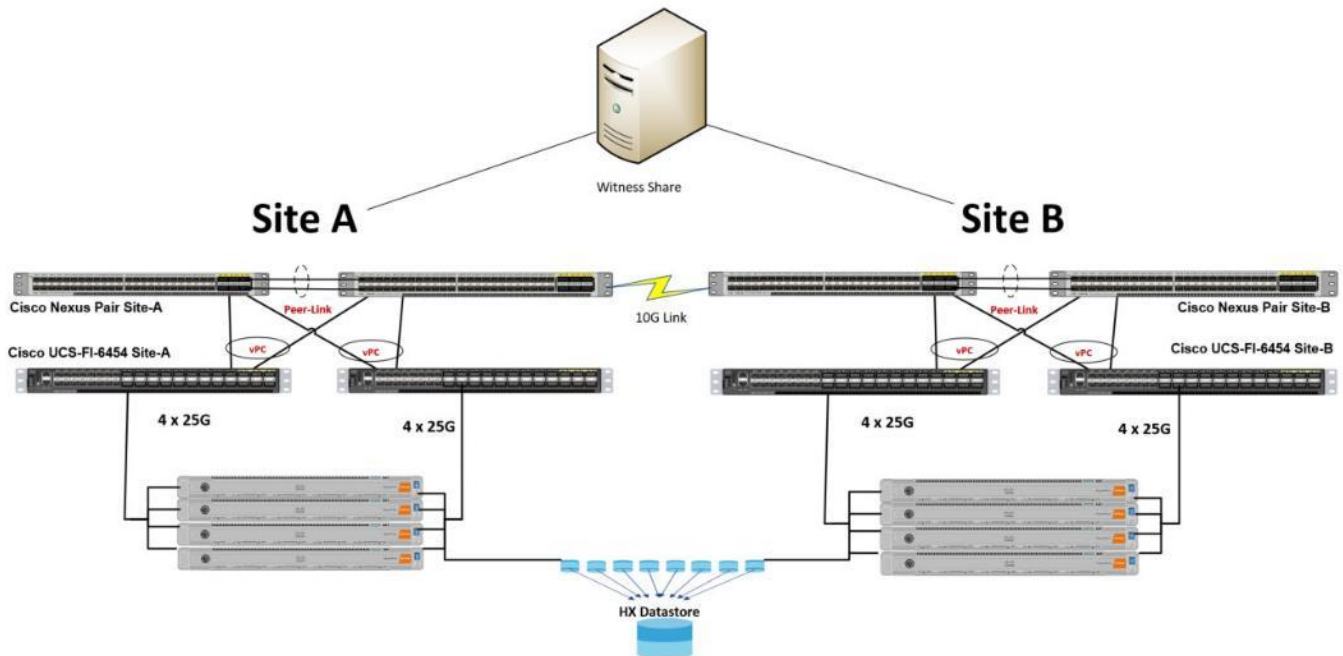
https://www.cisco.com/c/en/us/td/docs/hyperconverged_systems/HyperFlex_HX_DataPlatformSoftware/b_Cisco_HyperFlex_Systems_Ordering_and_Licensing_Guide/b_Cisco_HyperFlex_Systems_Ordering_and_Licensing_Guide_chapter_01001.html

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. 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 “north-bound” network connections are made from the Fabric Interconnects to the customer datacenter network at the time of installation.

Figure 14. HyperFlex Stretch 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. In this Stretch Cluster, there are a total of 4 fabric interconnects, a pair at each site.

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. Cisco HyperFlex M5 generation servers can be configured with the Cisco UCS VIC 1387 or Cisco UCS VIC 1457 cards. The standard and redundant connection practice for the Cisco UCS VIC 1387 is to connect port 1 of the Cisco UCS 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 ([Figure 15](#)). For the Cisco UCS VIC 1457 card, the standard and redundant practice is to connect port 1 of the VIC card (the left-hand most port) to a port on FI A and connect port 3 (the right-center port) to a port on FI B ([Figure 16](#)). An optional configuration method for servers containing the Cisco VIC 1457 card is to cable the servers with 2 links to each FI, using ports 1 and 2 to FI A, and ports 3 and 4 to FI B. The HyperFlex installer checks for these configurations, and that all servers' cabling matches. Failure to follow this cabling best practice can lead to errors, discovery failures, and loss of redundant connectivity.

All nodes within a Cisco HyperFlex cluster must be connected at the same communication speed, for example, mixing 10 Gb with 25 Gb interfaces is not allowed. In addition, for clusters that contain only M5 generation nodes, all of the nodes within a cluster must contain the same model of Cisco VIC cards.

Various combinations of physical connectivity between the Cisco HX-series servers and the Fabric Interconnects are possible, but only specific combinations are supported. [Table 5](#) lists the possible connections, and which of these methods is supported.

Table 5. Supported Physical Connectivity

Fabric Interconnect Model	6248		6296		6332		6332-16UP			6454	
	10GbE	10GbE	40GbE	10GbE Breakout	40GbE	10GbE Breakout	10GbE onboard	10GbE	25GbE		
M4 with VIC 1227	✓	✓	✗	✗	✗	✗	✗	✓	✗		
M4 with VIC 1387	✗	✗	✓	✗	✓	✗	✗	✗	✗		
M4 with VIC 1387 + QSA	✓	✓	✗	✗	✗	✗	✗	✓	✗		
M5 with VIC 1387	✗	✗	✓	✗	✓	✗	✗	✗	✗		
M5 with VIC 1387 + QSA	✓	✓	✗	✗	✗	✗	✗	✓	✗		
M5 with VIC 1457	✓	✓	✗	✗	✗	✗	✗	✓	✓		

Figure 15. HX-Series Server with Cisco UCS VIC 1387 Connectivity

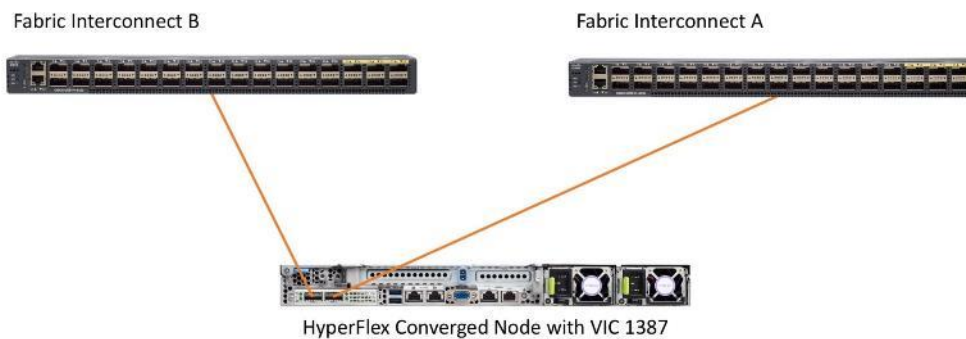
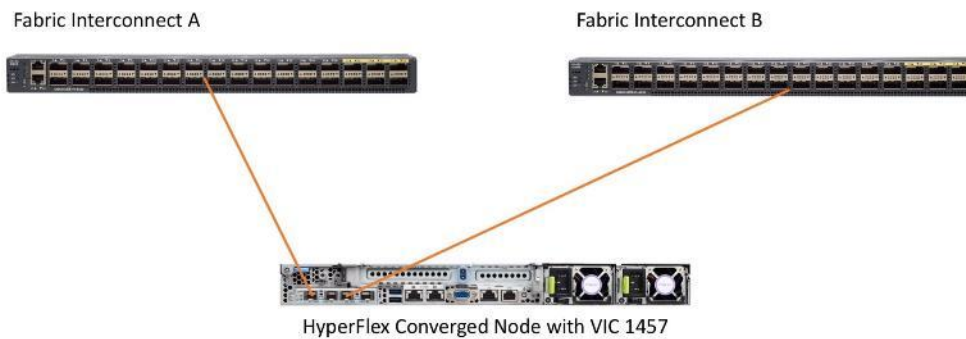


Figure 16. HX-Series Server with Cisco UCS VIC 1457 Connectivity



Cisco UCS C-Series Rack-Mount Servers

HyperFlex extended clusters can 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, 1387 or 1457 network interface card (NIC) installed in a modular LAN on motherboard (MLOM) slot, which have dual 10 Gigabit Ethernet (GbE), quad 10/25 Gigabit Ethernet (GbE) ports or dual 40 Gigabit Ethernet (GbE) ports. The standard and redundant connection practice for connecting standard Cisco UCS C-Series servers to the Fabric Interconnects is identical to the method described earlier for the HX-Series servers.


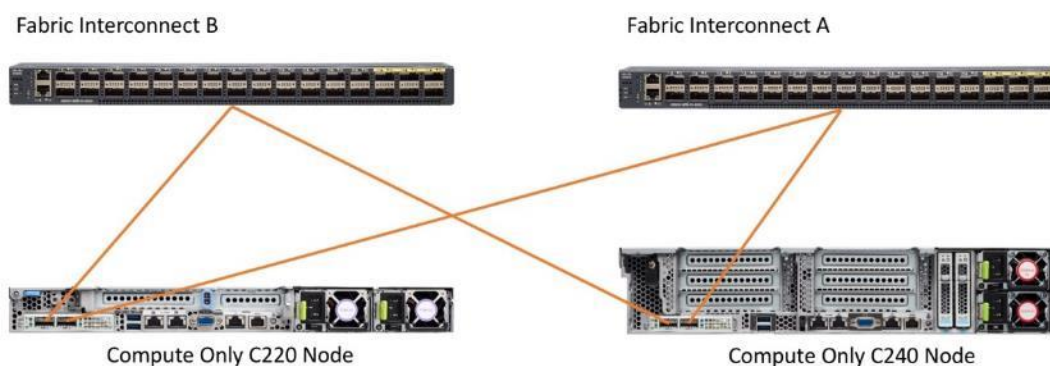
 Failure to follow this cabling practice can lead to errors, discovery failures, and loss of redundant connectivity.

Figure 17. Cisco UCS C-Series Server Connectivity



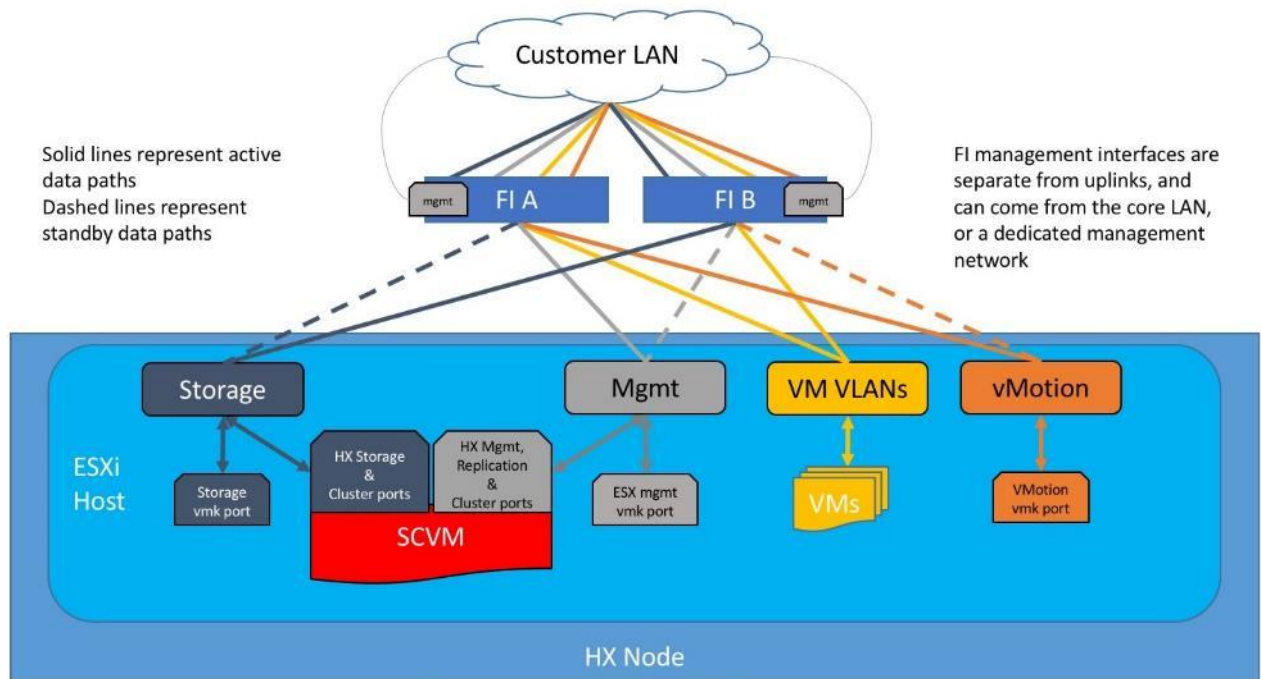
Logical Topology

Logical Network Design

The Cisco HyperFlex system has communication pathways that fall into four defined zones ([Figure 18](#)):

-
- **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 also 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, which 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.

Figure 18. Logical Network Design



Design Elements

Installing the HyperFlex system is done via the Cisco Intersight online management portal, or through a deployable HyperFlex installer virtual machine, available for download at [cisco.com](https://www.cisco.com) as an OVA file. The installer performs most of the Cisco UCS configuration work, and also performs significant portions of the ESXi configuration. Finally, the installer will 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 prerequisite steps needed for installation, and how to then utilize the HyperFlex Installer for the remaining configuration steps. This document focuses on the use of Cisco Intersight for the initial deployment of a Cisco HyperFlex cluster.

Network Design for a Stretch Cluster

For detailed guidance on Network Design for stretch clusters, please refer to the HyperFlex Stretch Cluster for Infrastructure CVD here:

https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/UCS_CVDs/hx_40_vsi_aci_multipod_design.html

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.

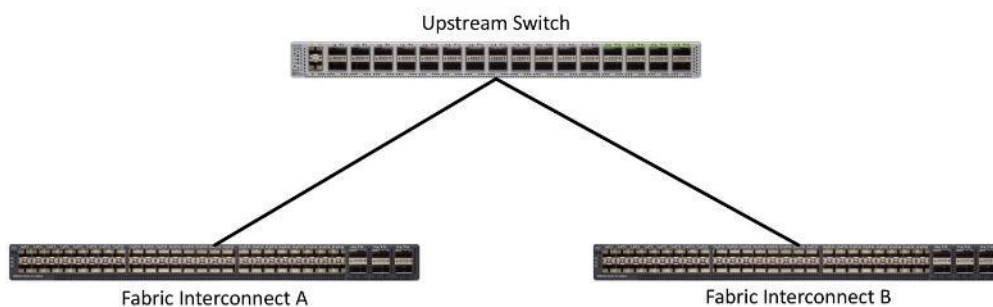
Uplinks need to be connected and active from both Fabric Interconnects. For redundancy, multiple uplinks can be used on each FI, either as 802.3ad Link Aggregation Control Protocol (LACP) port-channels or using individual links. For the best level of performance and redundancy, uplinks can be made as 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, however 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 instead be directed over the Cisco UCS uplinks because that traffic must travel from fabric A to fabric B, or vice-versa. 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. Cisco recommends that the uplink bandwidth configured is greater than or equal to double the bandwidth available to each Hyperflex converged node. For example, if the nodes are connected at 10 Gigabit speeds, then each Fabric Interconnect should have at least 20 Gigabit of uplink bandwidth available. 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.

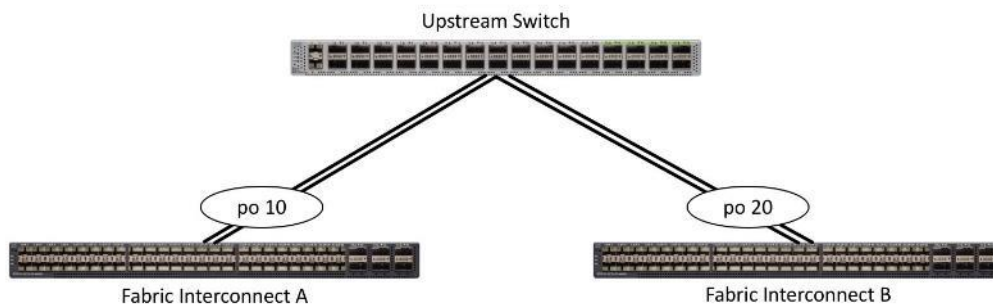
Figure 19. Connectivity with Single Uplink to Single Switch



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.

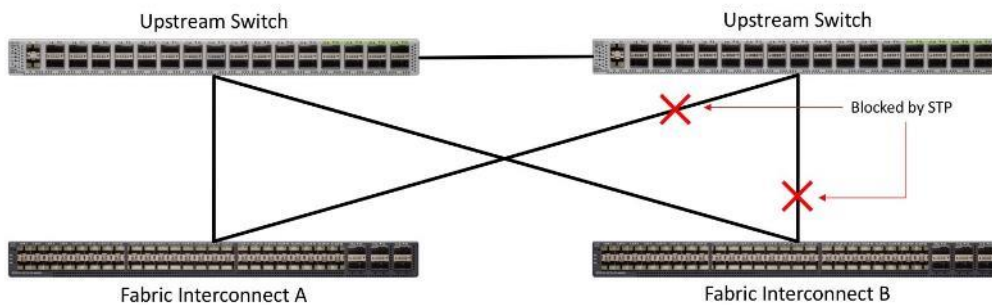
Figure 20. Connectivity with Port-Channels to Single Switch



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.

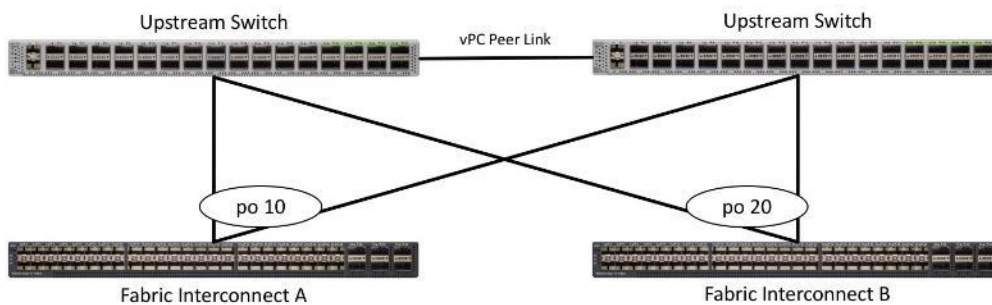
Figure 21. Connectivity with Multiple Uplink Switches



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 22. 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 6](#) lists the VLANs created by the HyperFlex installer in Cisco UCS, and their functions:

Table 6. 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

VLAN Name	VLAN ID	Purpose
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



A dedicated network or subnet for physical device management is often used in datacenters. In this scenario, the mgmt0 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 mgmt0 interfaces of the Fabric Interconnects, and also have IP connectivity to the subnets used by the hx-inband-mgmt VLANs listed above.

Jumbo Frames

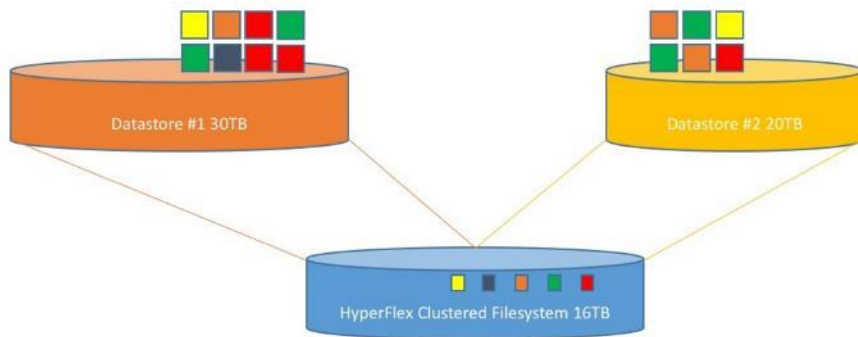
All HyperFlex storage traffic traversing the hx-storage-data VLAN, and subnet is configured by default to use jumbo frames, or to be precise, all communication is configured to send IP packets with a Maximum Transmission Unit (MTU) size of 9000 bytes. In addition, the default MTU for the hx-vmotion VLAN 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.

HyperFlex Datastores for Stretch Clusters

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. When creating a datastore on Site-A the site affinity must be configured for site-a and the same goes for Site-B. Both sites will write copies to its stretch cluster partner site, so the data is available in the event of a site failure. 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 23. Datastore Example



Installation

Cisco UCS Installation for Stretch Clusters

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, follow these 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 : 10.29.132.104

Physical Switch Mgmt0 IPv4 netmask : 255.255.255.0

IPv4 address of the default gateway : 10.29.132.1

Cluster IPv4 address : 10.29.132.106

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

DNS IP address : 10.29.132.110

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

Default domain name : hxdom.local

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=10.29.132.104

Physical Switch Mgmt0 IP Netmask=255.255.255.0

Default Gateway=10.29.132.1

Ipv6 value=0

DNS Server=10.29.132.110

Domain Name=hxdom.local

Cluster Enabled=yes

Cluster IP Address=10.29.132.106

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, follow these 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: 10.29.132.104
Peer Fabric interconnect Mgmt0 IPv4 Netmask: 255.255.255.0
Cluster IPv4 address           : 10.29.132.106
```

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

```
Physical Switch Mgmt0 IP address : 10.29.132.105
```

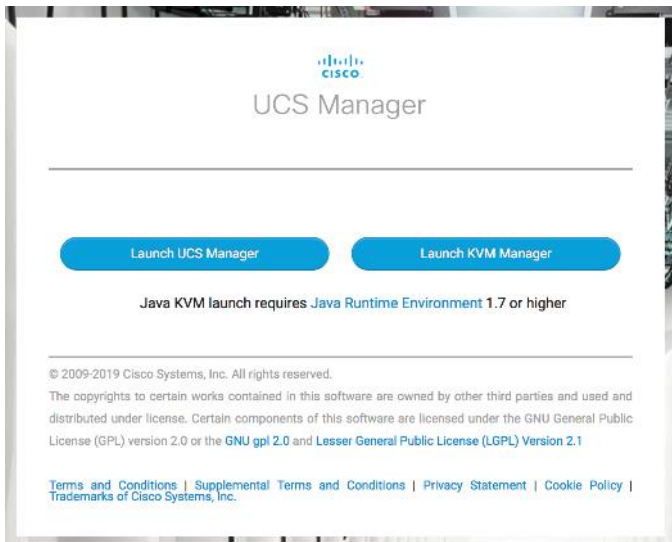
```
Apply and save the configuration (select 'no' if you want to re-enter)? (yes/no): yes
Applying configuration. Please wait.
```

```
Configuration file - Ok
```

Cisco UCS Manager

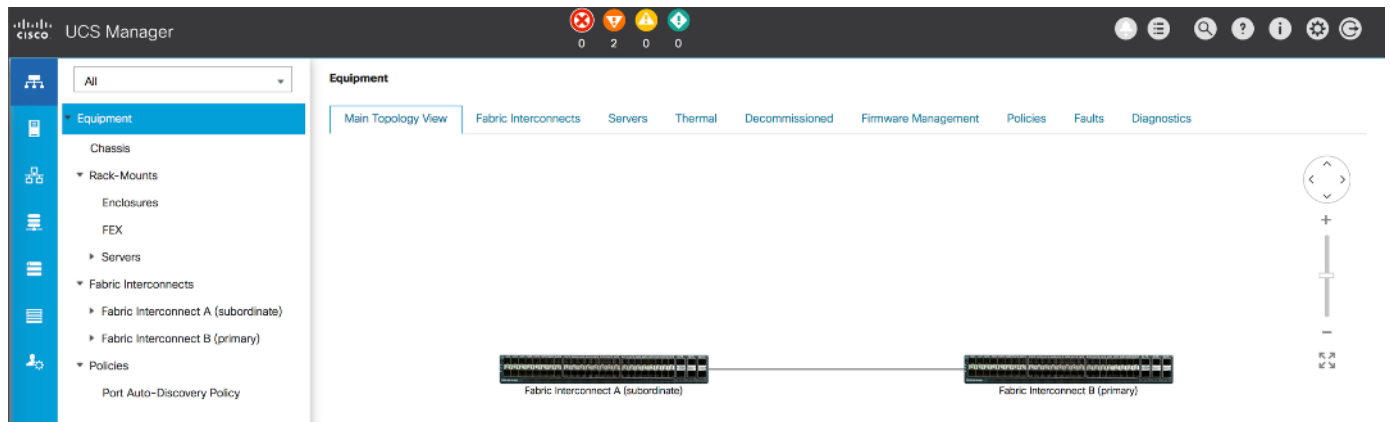
To log into the Cisco UCS Manager environment, follow these steps:

1. Open a web browser and navigate to the Cisco UCS Manager Cluster IP address, for example: <https://10.29.132.106>



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.

- 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 4.0(4d). 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:

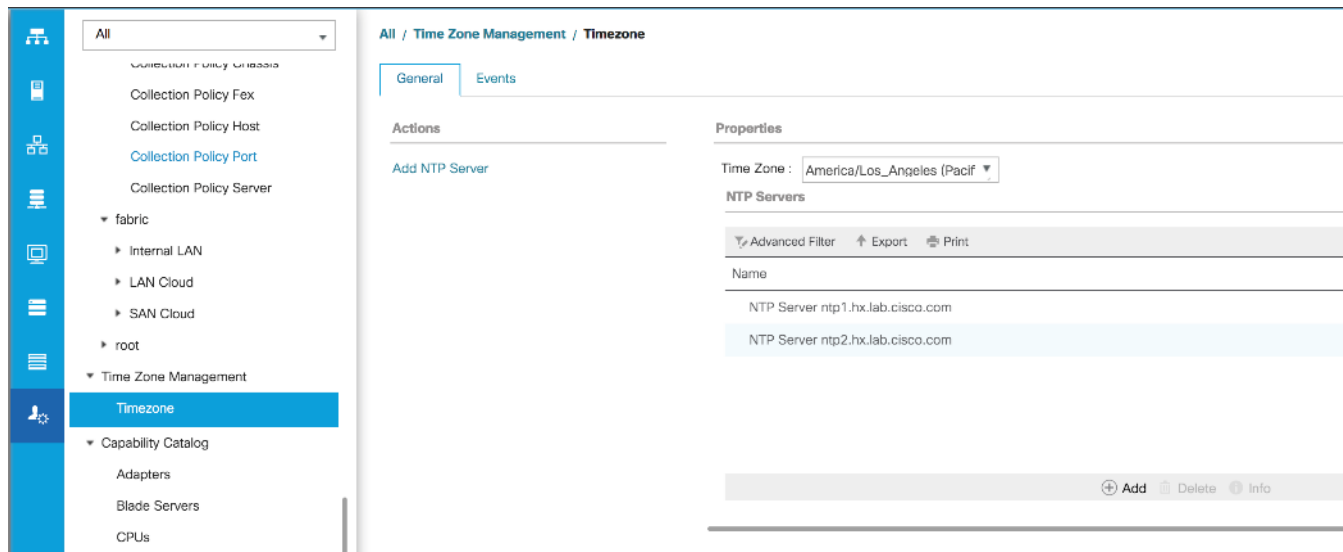
https://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/ucs-manager/GUI-User-Guides/Firmware-Mgmt/4-0/b_UCSM_GUI_Firmware_Management_Guide_4-0.html

NTP

To synchronize the Cisco UCS environment time to the NTP server, follow these steps:

- In Cisco UCS Manager, click Admin.
- In the navigation pane, select All > Time Zone Management, and click the carat next to Time Zone Management to expand it.
- Click Timezone.
- In the Properties pane, select the appropriate time zone in the Time Zone menu.
- Click Add NTP Server.
- Enter the NTP server IP address and click OK.
- 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, follow these steps:

1. In Cisco UCS Manager, click Equipment.
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, then 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.”

Slot	Aggr. Port ID	Port ID	MAC	If Role	If Type	Overall Status	Admin State	Peer
1	0	39	00:DE:FB:DF:B7:A0	Network	Physical	↑ Up	↑ Enabled	
1	0	40	00:DE:FB:DF:B7:A1	Network	Physical	↑ Up	↑ Enabled	

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, follow these steps:

1. In Cisco UCS Manager, click LAN.
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, then 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, then 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, follow these steps:

1. In Cisco UCS Manager, click Equipment, and click Equipment.
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. Set the backplane speed preference to 4x10 Gigabit or 40 Gigabit.

6. Click Save Changes.

7. Click OK.

The screenshot shows the Cisco UCS Manager interface. At the top, there is a navigation bar with tabs: Main Topology View, Fabric Interconnects, Servers, Thermal, Decommissioned, Firmware Management, and Policies. Below this is a sub-navigation bar with tabs: Global Policies, Autoconfig Policies, Server Inheritance Policies, Server Discovery Policies, SEL Policy, and Power Groups. The main content area is titled "Chassis/FEX Discovery Policy" and contains three configuration rows:

- Action : 1 Link (dropdown menu)
- Link Grouping Preference : None Port Channel
- Backplane Speed Preference : 40G 4x10G

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

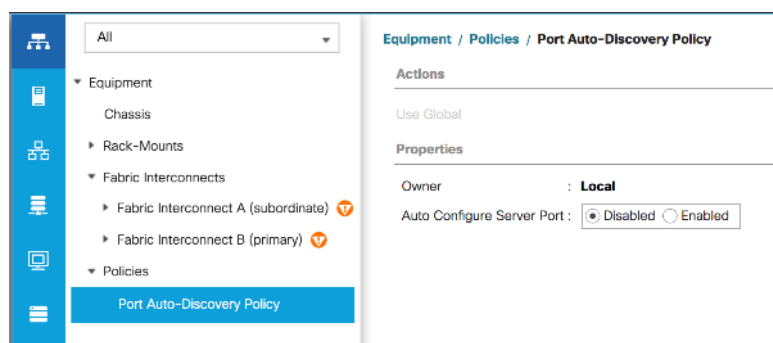
Server Port Auto-Discovery, which automates the configuration of ports on the Fabric Interconnects as server ports when a Cisco UCS rack-mount server or blade chassis is connected to them. The firmware on the rack-mount servers or blade chassis Fabric Extenders must already be at version 3.1(3a) or later in order for this feature to function properly. Enabling this policy eliminates the manual steps of configuring each server port, however it can configure the servers in a somewhat random order depending upon the circumstances. An example of how to use this feature in an orderly manner would be to have the policy already set, then to mount, cable and apply power to each new server one-by-one. In this scenario the servers should be automatically discovered in the order you racked them and applied power.

An example of how the policy can result in unexpected ordering would be when the policy has not been enabled, then all of the new servers are racked, cabled, and have power applied to them. If the policy is enabled afterwards, it will likely not discover the servers in a logical 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 in this scenario, the manual configuration steps listed in the next section must be followed.

To configure automatic server port definition and discovery, follow these steps:

1. In Cisco UCS Manager, click Equipment.
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, follow these steps:

1. In Cisco UCS Manager, click Equipment.
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.

Equipment / Fabric Interconnects / Fabric Interconnect A (subordinate) / Fixed Module / Ethernet Ports

Ethernet Ports

Advanced Filter Export Print All Unconfigured Network Server FCoE Uplink Unified Uplink Appliance Storage FCoE Storage Unified Storage

Slot	Aggr. Port ID	Port ID	MAC	IF Role	IF Type	Overall Status	Admin State	Peer
1	0	17	00:DE:FB:DF:B7:54	Server	Physical	Up	Enabled	sys/rack-unit-1/ad...
1	0	18	00:DE:FB:DF:B7:58	Server	Physical	Up	Enabled	sys/rack-unit-2/ad...
1	0	19	00:DE:FB:DF:B7:5C	Server	Physical	Up	Enabled	sys/rack-unit-3/ad...
1	0	20	00:DE:FB:DF:B7:60	Server	Physical	Up	Enabled	sys/rack-unit-4/ad...
1	0	21	00:DE:FB:DF:B7:64	Server	Physical	Up	Enabled	sys/rack-unit-5/ad...
1	0	22	00:DE:FB:DF:B7:68	Server	Physical	Up	Enabled	sys/rack-unit-6/ad...
1	0	23	00:DE:FB:DF:B7:6C	Server	Physical	Up	Enabled	sys/rack-unit-7/ad...
1	0	24	00:DE:FB:DF:B7:70	Server	Physical	Up	Enabled	sys/rack-unit-8/ad...

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, follow these steps:

1. In Cisco UCS Manager, click Equipment, 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, then view the servers' status in the Overall Status column.

UCS Manager

Equipment

Main Topology View Fabric Interconnects Servers Thermal Decommissioned Firmware Management Policies Faults Diagnostics

Blade Servers Rack-Mount Servers

Advanced Filter Export Print

Name	Overall Status	PID	Model	Serial	Profile	User ...	Cores	Core...	Threa...	Mem...	Adap...	NICs	HBA...	Oper...	Powe...	Asso...	Fault ...
Enclosures																	
Servers																	
Server 1	Unassociated	HXAF240C-MSSX	Cisc...	WZP...			36	36	72	3932...	1	0	0	Off	Off	N...	N/A
Server 2	Unassociated	HXAF240C-MSSX	Cisc...	WZP...			36	36	72	3932...	1	0	0	Off	Off	N...	N/A
Server 3	Unassociated	HXAF240C-MSSX	Cisc...	WZP...			36	36	72	3932...	1	0	0	Off	Off	N...	N/A
Server 4	Unassociated	HXAF240C-MSSX	Cisc...	WZP...			36	36	72	3932...	1	0	0	Off	Off	N...	N/A
Server 5	Unassociated	HXAF240C-MSSX	Cisc...	WZP...			36	36	72	3932...	1	0	0	Off	Off	N...	N/A
Server 6	Unassociated	HXAF240C-MSSX	Cisc...	WZP...			36	36	72	3932...	1	0	0	Off	Off	N...	N/A
Server 7	Unassociated	HXAF240C-MSSX	Cisc...	WZP...			36	36	72	3932...	1	0	0	Off	Off	N...	N/A
Server 8	Unassociated	HXAF240C-MSSX	Cisc...	WZP...			36	36	72	3932...	1	0	0	Off	Off	N...	N/A

HyperFlex Installer VM 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/4.0\(1b\)](https://software.cisco.com/download/home/286305544/type/286305994/release/4.0(1b))

This document is based on the Cisco HyperFlex 4.0(2b) release filename: Cisco-HX-Data-Platform-Installer-v4.0.2b-33133-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 importing 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 4.0 Hardening Guide: https://www.cisco.com/c/dam/en/us/support/docs/hyperconverged-infrastructure/hyperflex-hx-data-platform/HX-Hardening_Guide.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 7](#) to document the settings to be used for the HyperFlex installer VM.

Table 7. HyperFlex Installer Settings

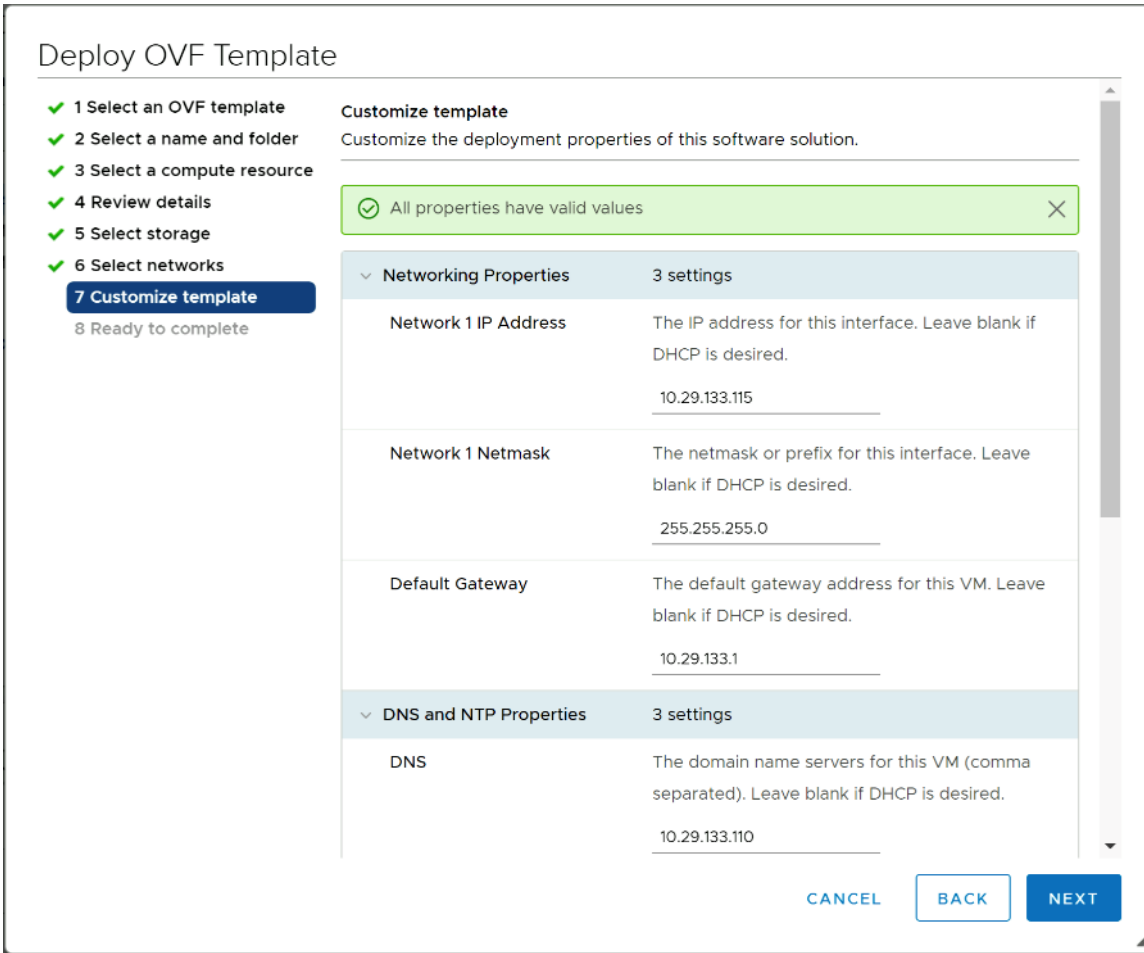
Setting	Value
IP Address	
Subnet Mask	
Default Gateway	

Setting	Value
DNS Server	
NTP Server(s)	
Root Password	

Deploy Installer OVA

To deploy the HyperFlex installer OVA, follow these steps:

1. Open 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. Select the Local file option, then click Choose Files and locate the Cisco-HX-Data-Platform-Installer-v4.0.2b-33133-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.
12. Enter and confirm a new password used to log in to the installer VM after it is deployed, then click Next.



13. Review the final configuration and click Finish.

14. 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 shown in [Figure 24](#), showing the IP address that was leased.

Figure 24. HyperFlex Installer VM IP Address

```
Version 4.0(1b)

*****
You can start the installation by visiting
the following URL:

      http://10.29.133.115

*****

HyperFlex-Installer login: _
```

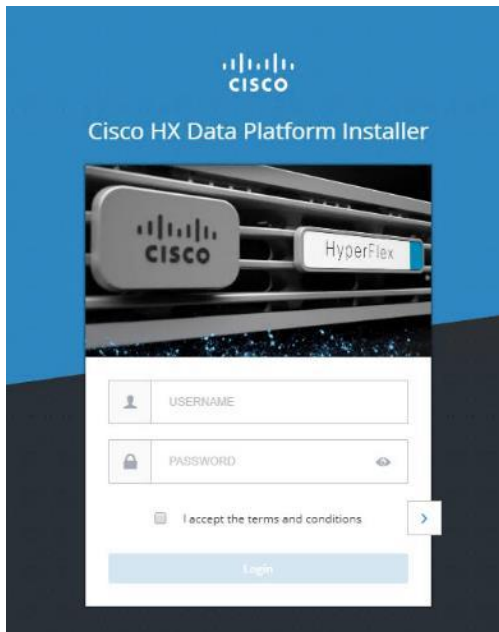
To access the HyperFlex installer webpage, follow these steps:

1. Open a web browser on the local computer and navigate to the IP address of the installer VM. For example, open <http://10.29.132.115>
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 password which was set during the OVA deployment.
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.

Cisco HyperFlex Stretch Cluster Configuration

To configuring the Cisco HyperFlex Cluster, follow this step:

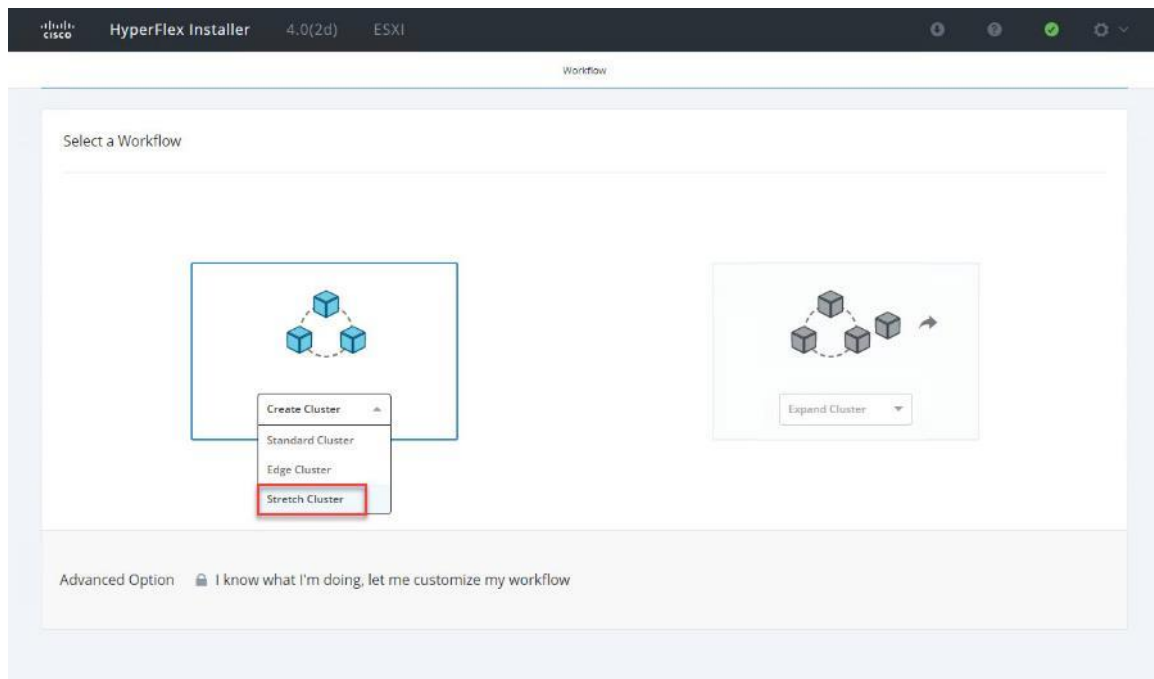
1. Log into the HX Installer virtual machine through a web browser: http://<Installer_VM_IP_Address>.



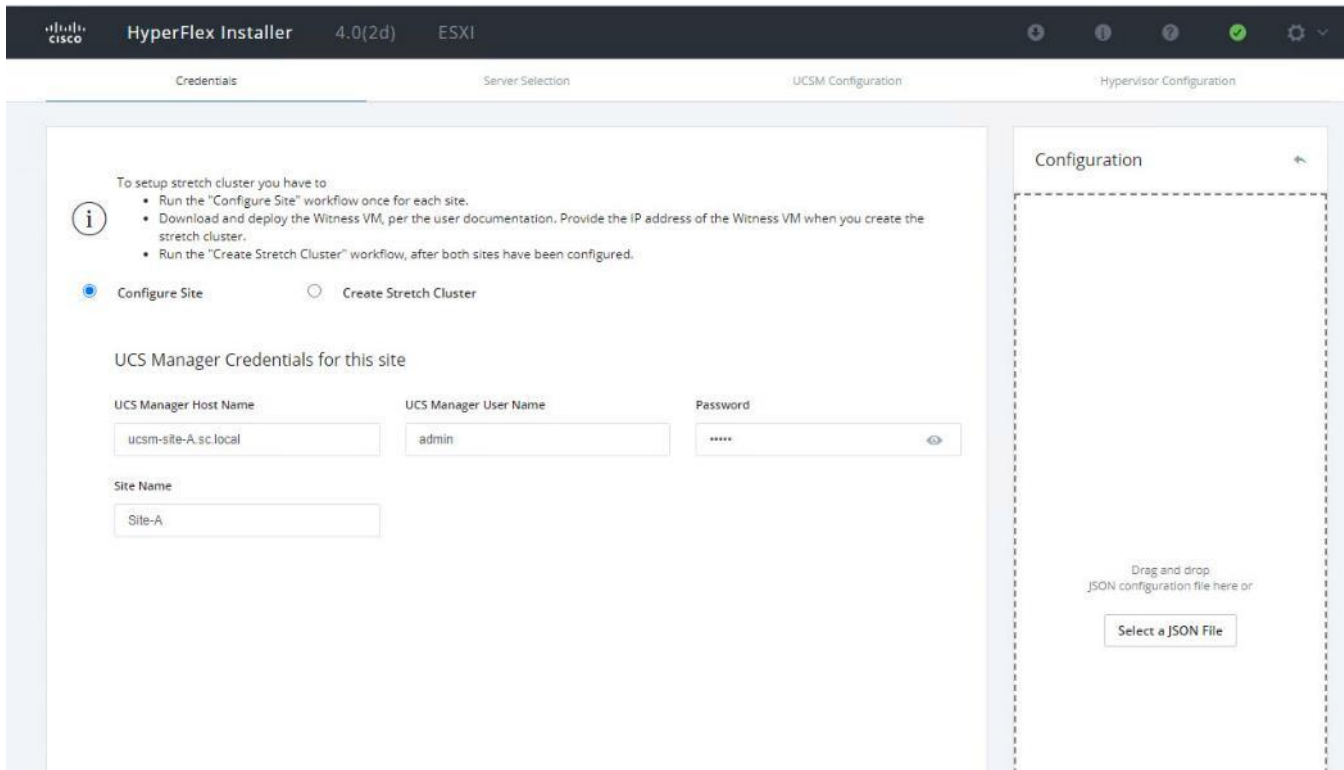
Create a HyperFlex Stretch Cluster

To create a HyperFlex cluster, follow these steps:

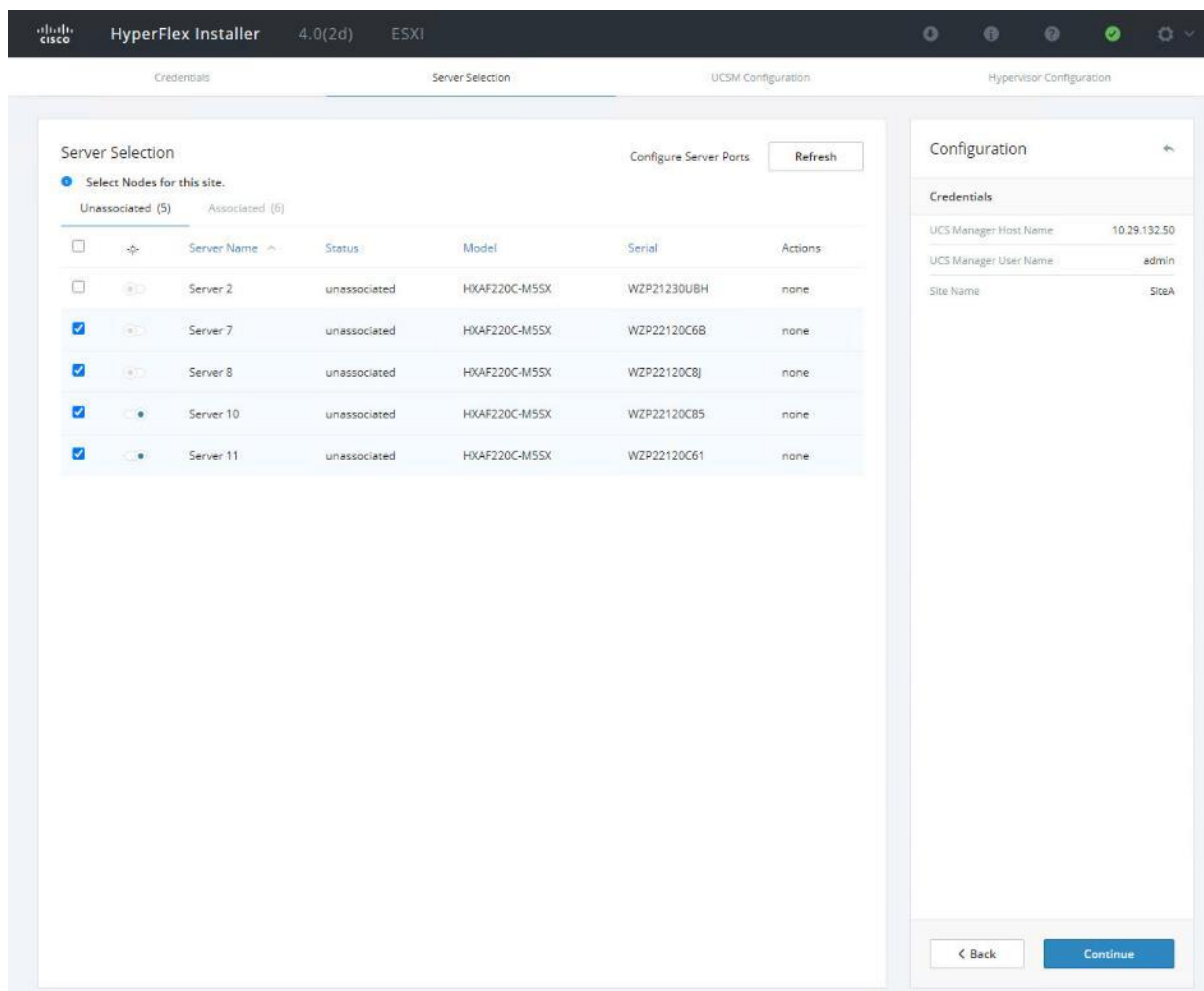
1. Select the workflow for cluster creation to deploy a new HyperFlex cluster on eight Cisco HXAF220c-M5S nodes. From the drop-down list select Stretch Cluster.



2. On the credentials page, enter the access details for Cisco UCS Manager and the Site information. (The Configure Site option must be run on both sites before the Stretch Cluster workflow can be run.) Click Continue.



3. Select the unassociated servers for the site in the HyperFlex installer. To configure a subset of available of the HyperFlex servers, manually click the checkbox for individual servers.
4. Click Continue after completing server selection.



The required server ports can be configured from Installer workflow but it will extend the time to complete server discovery. Therefore, we recommend configuring the server ports and complete HX node discovery in Cisco UCS Manager as described in the [Prerequisites](#) section prior to starting the workflow for HyperFlex installer.

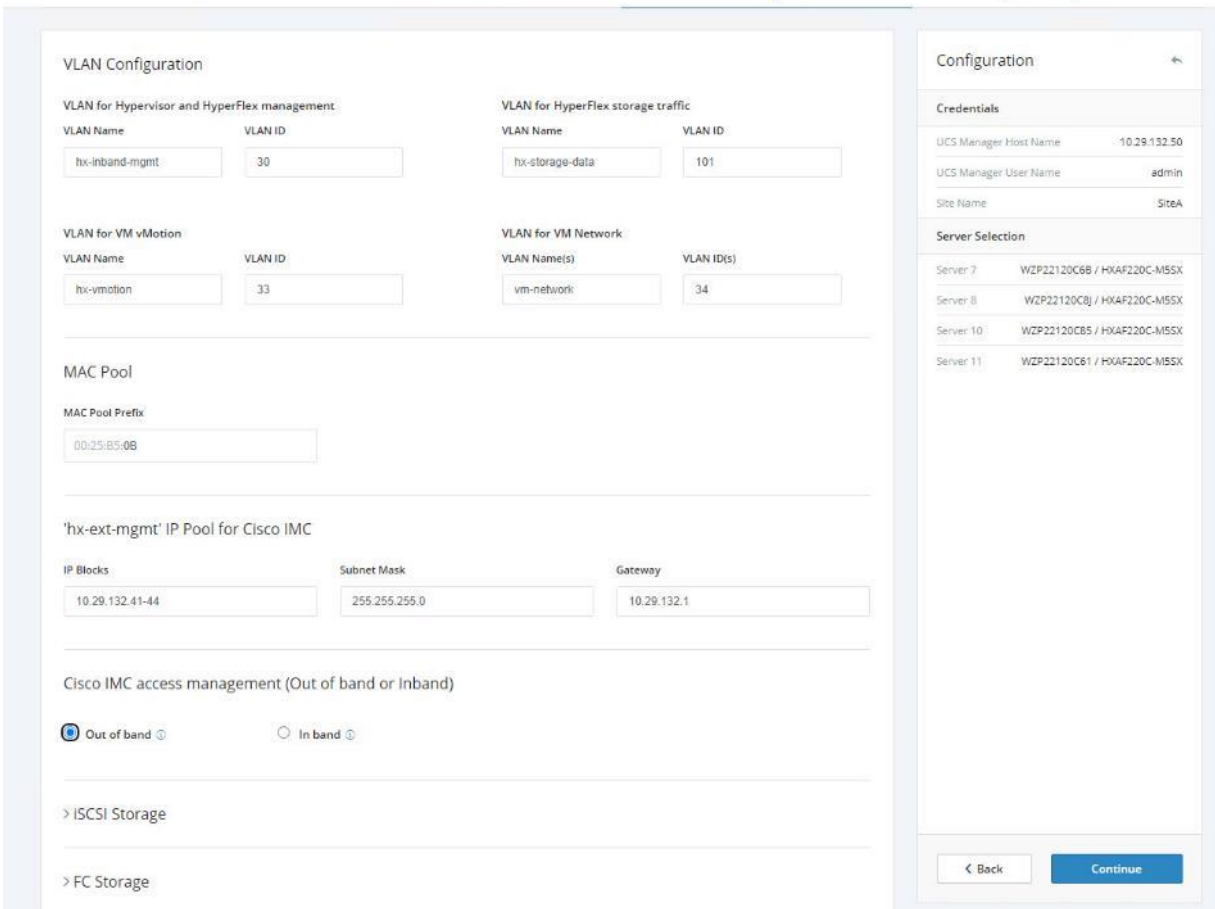
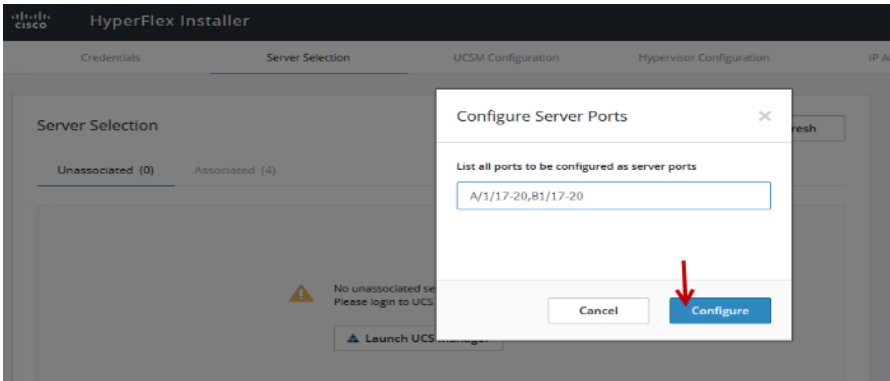
Configure Server Ports (Optional)

If you choose to allow the installer to configure the server ports, follow these steps:

1. Click Configure Server Ports at the top right corner of the Server Selection window.
2. Provide the port numbers for each Fabric Interconnect in the form:

A1/x-y,B1/x-y where A1 and B1 designate Fabric Interconnect A and B and where x=starting port number and y=ending port number on each Fabric Interconnect.

3. Click Configure.



4. Enter the Details for the Cisco UCS Manager Configuration:

- a. Enter VLAN ID for hx-inband-mgmt, hx-storage-data, hx-vmotion, vm-network.
- b. MAC Pool Prefix: The prefix to use for each HX MAC address pool. Please select a prefix that does not conflict with any other MAC address pool across all Cisco UCS domains.
- c. The blocks in the MAC address pool will have the following format:
 - $\{\text{prefix}\}:\{\text{fabric_id}\}\{\text{vnic_id}\}:\{\text{service_profile_id}\}$
 - The first three bytes should always be "00:25:B5".



The first three bytes should always be "00:25:B5."

5. Enter range of IP address to create a block of IP addresses for external management and access to CIMC/KVM.
6. Cisco UCS firmware version is set to 4.0 (4g) which is the required Cisco UCS Manager release for HyperFlex v4.0(2d) installation.
7. Enter HyperFlex cluster name.
8. Enter Org name to be created in Cisco UCS Manager.
9. Click Continue.

10. Repeat steps 1 - 9 for Site-B to configure both sites.

Configure Hypervisor Settings

To configure the Hypervisor settings, follow these steps:

1. In the Configure common Hypervisor Settings section, enter:
 - a. Subnet Mask
 - b. Gateway
 - c. DNS server(s)
2. In the Hypervisor Settings section:
 - a. Select check box Make IP Address and Hostnames Sequential if they are following in sequence.
 - b. Provide the starting IP Address.
 - c. Provide the starting Host Name or enter Static IP address and Host Names manually for each node
3. Click Configure Site.
4. Repeat steps 1 - 3 on Site B.

Deploy Cluster Witness Machine

Download the HyperFlex-Witness-1.x.x.ova appliance at

[https://software.cisco.com/download/home/286305544/type/286305994/release/4.0\(2b\)](https://software.cisco.com/download/home/286305544/type/286305994/release/4.0(2b))

Complete the Deploy OVF Template wizard for the appliance to an ESX host that is NOT in the Stretch Cluster that will be used for VDI.

Be prepared to provide the following to customize the appliance:

- Appliance IP Address
- Subnet Mask
- Default Gateway
- DNS and NTP servers
- Search domains (i.e. cisco.com)
- Root password for the appliance

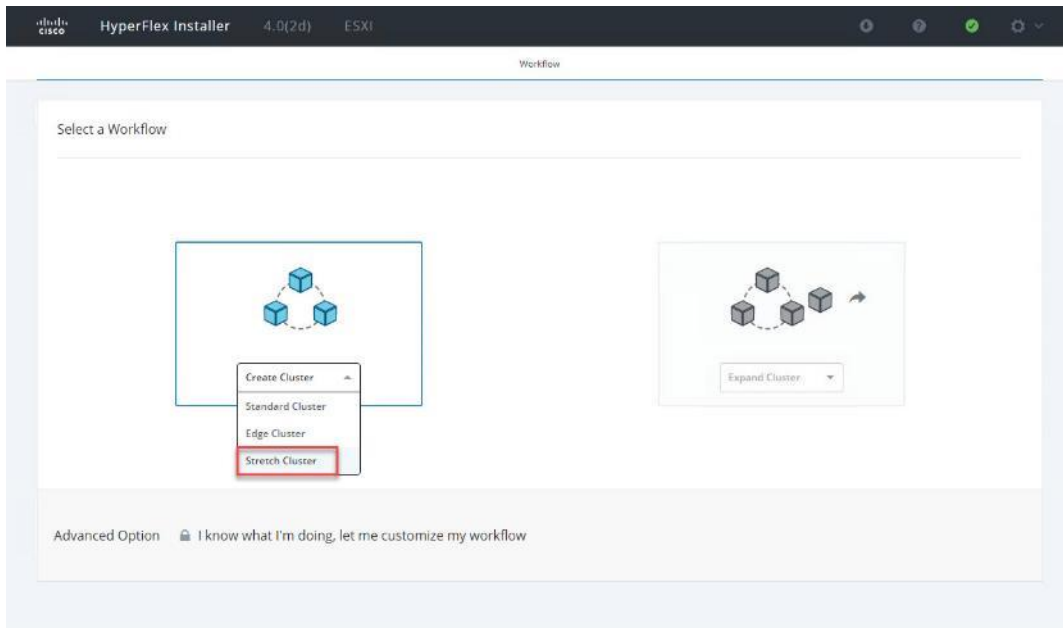


Notate the Witness appliance IP address to use when configuring the stretch cluster in the next steps.

Create Stretch Cluster

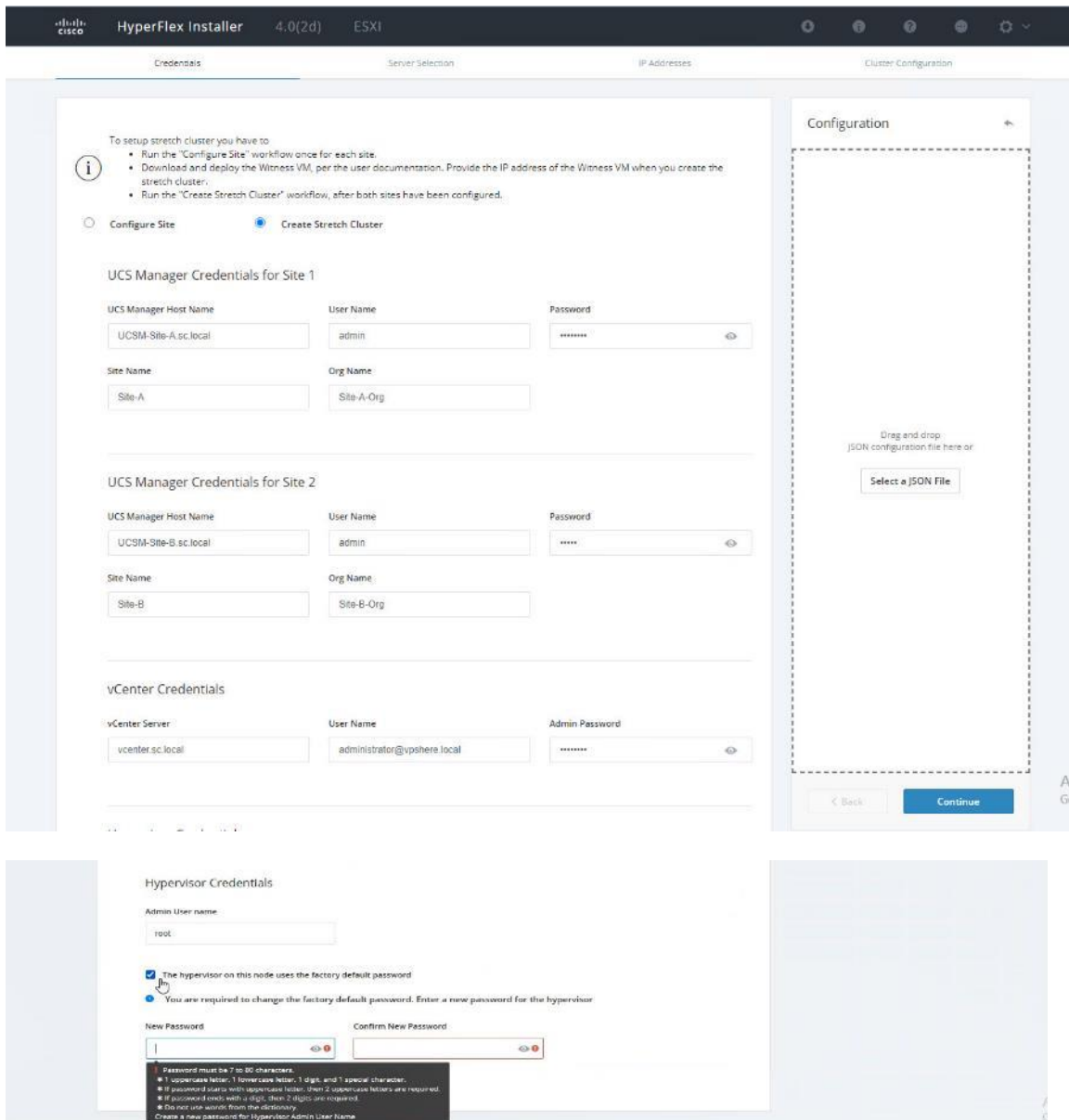
To create the Stretch Cluster, follow these steps:

1. On the installer screen, from the drop-down list select Stretch Cluster.

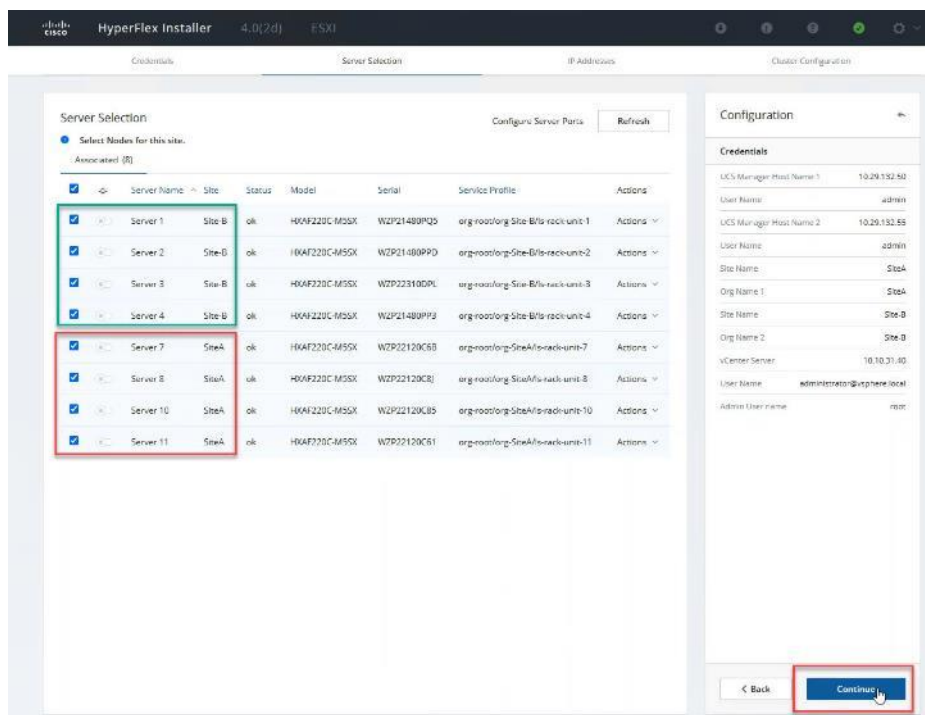


2. Enter the Cluster information for both sites:

- a. UCSM hostname or IP
- b. UCSM username and password
- c. Site Name
- d. Org Name
- e. Vcenter IP/Hostname, username and password.
- f. Hypervisor username and password (the default password from the factory is Cisco123 and must be changed)



3. Select the servers from each site and click Continue.



4. Provide the IP addresses for the management and data networks on both sites.

IP Addresses

To add the IP addresses, follow these steps:

When the IP Addresses page appears, the hypervisor IP address for each node that was configured in the Hypervisor Configuration tab, appears under the Management Hypervisor column.

Three additional columns appear on this page:

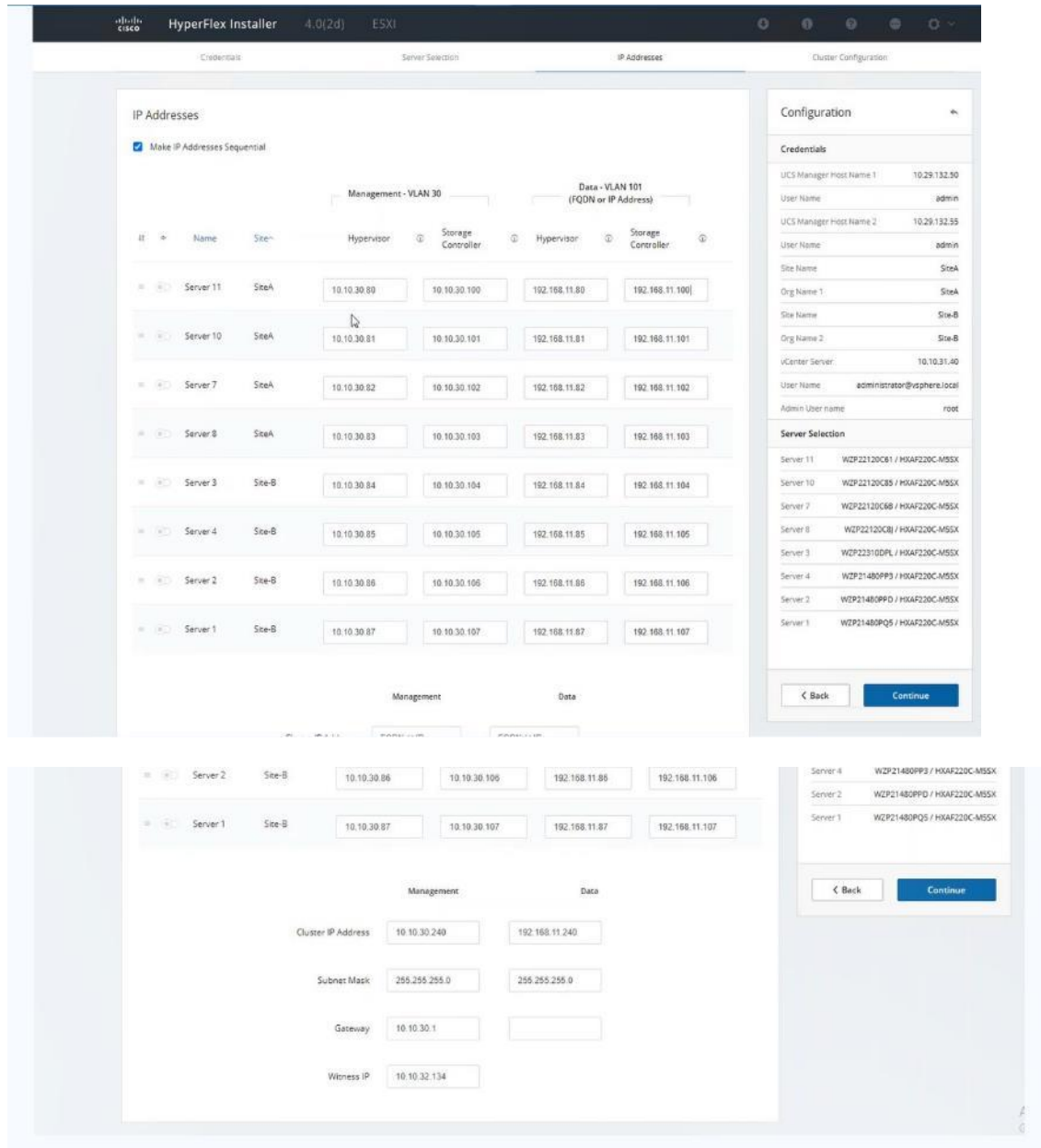
- Storage Controller/Management
- Hypervisor/Data
- Storage Controller/Data



The Data network IP addresses are for vmkernel addresses for storage access by the hypervisor and storage controller virtual machine.

- On the IP Addresses page, check the box Make IP Addresses Sequential or enter the IP address manually for each node for the following requested values:
 - Storage Controller/Management
 - Hypervisor/Data
 - Storage Controller/Data
- Enter subnet and gateway details for the Management and Data subnets configured.

3. Click Continue.



4. On the Cluster Configuration page, enter the following:

- Cluster Name
- Cluster management IP address
- Cluster data IP Address
- Replication Factor for Stretch Cluster is : 2 + 2
- Controller virtual machine password

-
- f. vCenter configuration
 - vCenter Datacenter name
 - vCenter Cluster name
 - g. System Services
 - DNS Server(s)
 - NTP Server(s)
 - Time Zone
 - h. Auto Support
 - Click the check box for Enable Auto Support
 - Mail Server
 - Mail Sender
 - ASUP Recipient(s)
 - i. Advanced Networking
 - Management vSwitch
 - Data vSwitch
 - j. Advanced Configuration
 - Click the check box to Optimize for VDI only deployment
 - Enable jumbo Frames on Data Network
 - Clean up disk partitions (optional)
 - vCenter Single-Sign-On server

Cisco HX Cluster

Cluster Name: Replication Factor:

Controller VM

Create Admin Password: Confirm Admin Password:

vCenter Configuration

vCenter Datacenter Name: vCenter Cluster Name:

System Services

DNS Server(s): NTP Server(s): DNS Domain Name:

Time Zone:

Auto Support

Auto Support: Enable Connected Services (Recommended) Send service ticket notifications to:

When Connected Services are enabled, Cisco periodically collects information about the cluster and its deployment environment for the

Configuration

Credentials

UCS Manager Host Name 1: 10.29.132.50
 User Name: admin
 UCS Manager Host Name 2: 10.29.132.55
 User Name: admin
 Site Name: SiteA
 Org Name 1: SiteA
 Site Name: Site-B
 Org Name 2: Site-B
 vCenter Server: 10.10.31.40
 User Name: administrator@vsphere.local
 Admin User name: root

Server Selection

Server 11	WZP22120C61 / HXAF220C-M55X
Server 10	WZP22120C85 / HXAF220C-M55X
Server 7	WZP22120C68 / HXAF220C-M55X
Server 8	WZP22120C6J / HXAF220C-M55X
Server 3	WZP22310DPL / HXAF220C-M55X
Server 4	WZP21480PP3 / HXAF220C-M55X
Server 2	WZP21480PPD / HXAF220C-M55X
Server 1	WZP21480PQ5 / HXAF220C-M55X

IP Addresses

When Connected Services are enabled, Cisco periodically collects information about the cluster and its deployment environment for the purpose of delivering a better product and support experience.

Web Proxy Settings for Connected Services

Use Proxy Server

Web Proxy Server: Port:

Username: Password:

Advanced Networking

Management VLAN Tag - Site 1: <input type="text" value="30"/>	Management VLAN Tag - Site 2: <input type="text" value="30"/>	Management vSwitch: <input type="text" value="vswitch-hx-inband-mgmt"/>
Data VLAN Tag - Site 1: <input type="text" value="101"/>	Data VLAN Tag - Site 2: <input type="text" value="101"/>	Data vSwitch: <input type="text" value="vswitch-hx-storage-data"/>

Advanced Configuration

Jumbo Frames: Enable Jumbo Frames on Data Network
 Disk Partitions: Clean up disk partitions

vCenter Single-Sign-On Server:



If the QoS system class is not defined as per the requirement HyperFlex installer will go ahead and make required changes. A warning is generated accordingly in HyperFlex Installer workflow. Post-install Configuration.

Prior to putting a new HyperFlex cluster into production, a few post-install tasks must be completed. To automate the post installation procedures and verify the HyperFlex cluster configuration, a `post_install` script has been provided on the HyperFlex Controller VMs. To run this script, follow these steps:

1. SSH to the cluster management IP address and login using `<root>` username and the controller VM password provided during installation. Verify the cluster is online and healthy using “`stcli cluster info`” or “`stcli cluster storage-summary`.”

```
root@SpringpathControllerT7DB8MDX0A:~# stcli cluster storage-summary
address: 169.254.37.1
name: All-NVMe
state: online
uptime: 0 days 2 hours 27 minutes 43 seconds
activeNodes: 8 of 8
compressionSavings: 76.99%
deduplicationSavings: 0.0%
freeCapacity: 13.2T
healingInfo:
  inProgress: False
resiliencyInfo:
  messages:
    Storage cluster is healthy.
  state: 1
  nodeFailuresTolerable: 2
  cachingDeviceFailuresTolerable: 2
  persistentDeviceFailuresTolerable: 2
  zoneResInfoList: None
spaceStatus: normal
totalCapacity: 13.4T
totalSavings: 76.99%
usedCapacity: 148.6G
zkHealth: online
clusterAccessPolicy: lenient
dataReplicationCompliance: compliant
dataReplicationFactor: 3
```

2. Run the following command in the shell, and press enter:

```
/usr/share/springpath/storfs-misc/hx-scripts/post_install.py
```

3. Select the first `post_install` workflow type – New/Existing Cluster.
4. Enter the HX Storage Controller VM root password for the HX cluster (use the one entered during the HX Cluster installation).
5. Enter the vCenter server username and password.

```
root@SpringpathControllerT7DB8MDX0A:~# /usr/share/springpath/storfs-misc/hx-scripts/post_install.py
Select post_install workflow-
1. New/Existing Cluster
2. Expanded Cluster
3. Generate Certificate

Note: Workflow No.3 is mandatory to have unique SSL certificate in the cluster.
      By Generating this certificate, it will replace your current certificate.
      If you're performing cluster expansion, then this option is not required.

Selection: 1
Logging in to controller localhost
HX CVM admin password:
Getting ESX hosts from HX cluster...
vCenter URL: 10.29.133.120
Enter vCenter username (user@domain): administrator@vsphere.local
vCenter Password:
Found datacenter Datacenter
Found cluster All-NVMe

post_install to be run for the following hosts:
hxaf220m5n-01.hx.lab.cisco.com
hxaf220m5n-02.hx.lab.cisco.com
hxaf220m5n-03.hx.lab.cisco.com
hxaf220m5n-04.hx.lab.cisco.com
hxaf220m5n-05.hx.lab.cisco.com
hxaf220m5n-06.hx.lab.cisco.com
hxaf220m5n-07.hx.lab.cisco.com
hxaf220m5n-08.hx.lab.cisco.com
```

6. Enter ESXi host root password (use the one entered during the HX Cluster installation).
7. You must license the vSphere hosts through the script or complete this task in vCenter before continuing. Failure to apply a license will result in an error when enabling HA or DRS in subsequent steps. Enter “n” if you have already registered the license information in vCenter.
8. Enter “y” to enable HA/DRS.
9. Enter “y” to disable the ESXi hosts’ SSH warning. SSH running in ESXi is required in HXDP 2.6.
10. 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.

```

Enter ESX root password:
Enter vSphere license key? (y/n) n
Enable HA/DRS on cluster? (y/n) y
Successfully completed configuring cluster HA.
Successfully completed configuring cluster DRS.

Disable SSH warning? (y/n) y

Add vmotion interfaces? (y/n) y
Netmask for vMotion: 255.255.255.0
VLAN ID: (0-4096) 200
vMotion MTU is set to use jumbo frames (9000 bytes). Do you want to change to 1500 bytes? (y/n) n
vMotion IP for hxaf220m5n-01.hx.lab.cisco.com: 192.168.200.61
Adding vmotion-200 to hxaf220m5n-01.hx.lab.cisco.com
Adding vmkernel to hxaf220m5n-01.hx.lab.cisco.com
vMotion IP for hxaf220m5n-02.hx.lab.cisco.com: 192.168.200.62
Adding vmotion-200 to hxaf220m5n-02.hx.lab.cisco.com
Adding vmkernel to hxaf220m5n-02.hx.lab.cisco.com
vMotion IP for hxaf220m5n-03.hx.lab.cisco.com: 192.168.200.63
Adding vmotion-200 to hxaf220m5n-03.hx.lab.cisco.com
Adding vmkernel to hxaf220m5n-03.hx.lab.cisco.com
vMotion IP for hxaf220m5n-04.hx.lab.cisco.com: 192.168.200.64
Adding vmotion-200 to hxaf220m5n-04.hx.lab.cisco.com
Adding vmkernel to hxaf220m5n-04.hx.lab.cisco.com
vMotion IP for hxaf220m5n-05.hx.lab.cisco.com: 192.168.200.65
Adding vmotion-200 to hxaf220m5n-05.hx.lab.cisco.com
Adding vmkernel to hxaf220m5n-05.hx.lab.cisco.com
vMotion IP for hxaf220m5n-06.hx.lab.cisco.com: 192.168.200.66
Adding vmotion-200 to hxaf220m5n-06.hx.lab.cisco.com
Adding vmkernel to hxaf220m5n-06.hx.lab.cisco.com
vMotion IP for hxaf220m5n-07.hx.lab.cisco.com: 192.168.200.67
Adding vmotion-200 to hxaf220m5n-07.hx.lab.cisco.com
Adding vmkernel to hxaf220m5n-07.hx.lab.cisco.com
vMotion IP for hxaf220m5n-08.hx.lab.cisco.com: 192.168.200.68
Adding vmotion-200 to hxaf220m5n-08.hx.lab.cisco.com
Adding vmkernel to hxaf220m5n-08.hx.lab.cisco.com

```

11. You may add VM network portgroups for guest VM traffic. Enter “n” to skip this step and create the portgroups manually in vCenter. Or if desired, VM network portgroups can be created and added to the vm-network vSwitch. This step will add identical network configuration to all nodes in the cluster.
12. Enter “y” to run the health check on the cluster.
13. A summary of the cluster will be displayed upon completion of the script. Make sure the cluster is healthy.

Initial Tasks and Testing

Datstores for a Stretch Cluster

When creating datstores for a stretch cluster, we recommend creating two datstores, one on each site with 50% of the total workload residing on each site datstore. For example, for 500 total VDI users, 250 desktops would be on datstore-site-A and 250 desktops would be on datstore-site-B.

Create a datstore for storing the virtual machines. This task can be completed by using the HyperFlex Connect HTML management webpage. To configure a new datstore via the HyperFlex Connect webpage, follow these 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. Select Site Affinity for each site (datastore-site-A would have 'Site-A' for Site Affinity, datastore-site-B would have 'Site-B' for Site Affinity.)
7. Click Create Datastore.

HX Connect

Create Datastore
ⓘ ⌵

Datastore Name

Size

 GB

Block Size

Site Affinity

Cancel
Create Datastore

Datastores

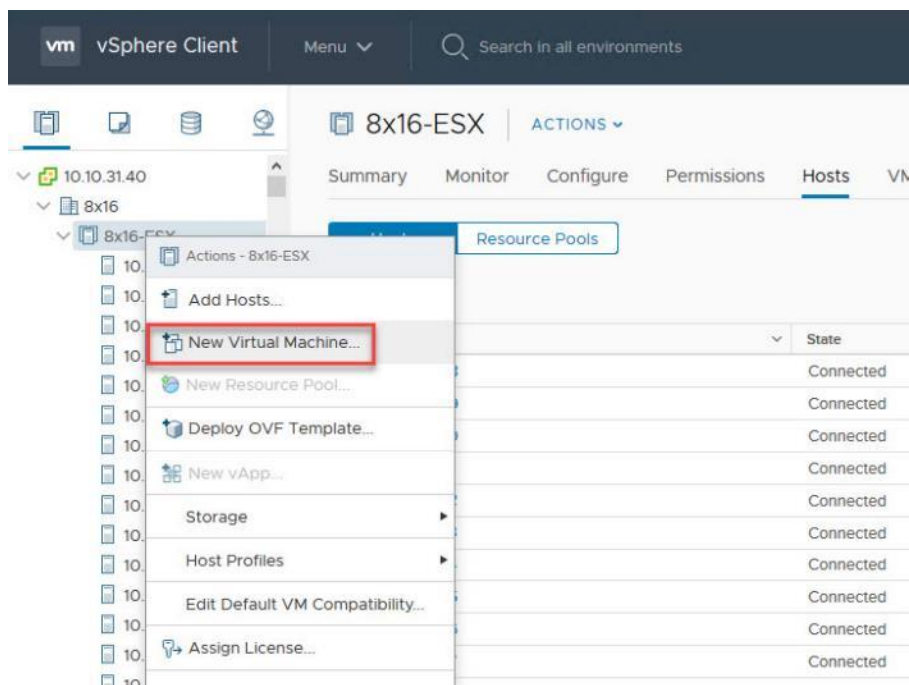
Last refreshed at: 10/31/2017 4:30:16 PM

	Name	Mount Summary	Site Affinity	Status	Size	Used	Free
<input type="checkbox"/>	DS1	MOUNTED	Site A	Normal	100 TB	245 GB	100 TB
<input type="checkbox"/>	DS2	MOUNTED	Site B	Normal	100 GB	0 B	100 GB

Showing 1 - 2 of 2

Create VM

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.



Audit Logging

By default, the HyperFlex controller VMs store logs locally for many functions, including the filesystem logs, security auditing, CLI commands and shell access, single sign-on logs, and more. These logs are rotated periodically and could be lost if there were a total failure of a controller VM. In order to store these logs externally from the HyperFlex cluster, audit logging can be enabled in HX Connect to send copies of these logs to an external syslog server. From this external location, logs can be monitored, generate alerts, and stored long term. HX Connect will not monitor the available disk space on the syslog destination, nor will it generate an alarm if the destination server is full. To enable audit logging, follow these steps:

1. Use a web browser to open the HX cluster IP management URL.
2. From the HyperFlex Connect webpage, click the gear shaped icon in the upper right-hand corner, and click Audit Log Export Settings.
3. Click to check the box to Enable audit log export to an external syslog server.
4. Enter the syslog server IP address and TCP port.
5. Choose TCP or TLS as the connection type. If using TLS, client certificate and private key pair files must be provided. Alternatively, a self-signed certificate can be used. Click browse to select the appropriate files.
6. Click OK.

7. Audit log exports can be temporarily disabled or completely deleted at a later time from the same location.

To store ESXi diagnostic logs in a central location in case they are needed to help diagnose 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 for ESXi, follow these 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://10.29.132.120'
[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 these steps for each ESXi host.

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.

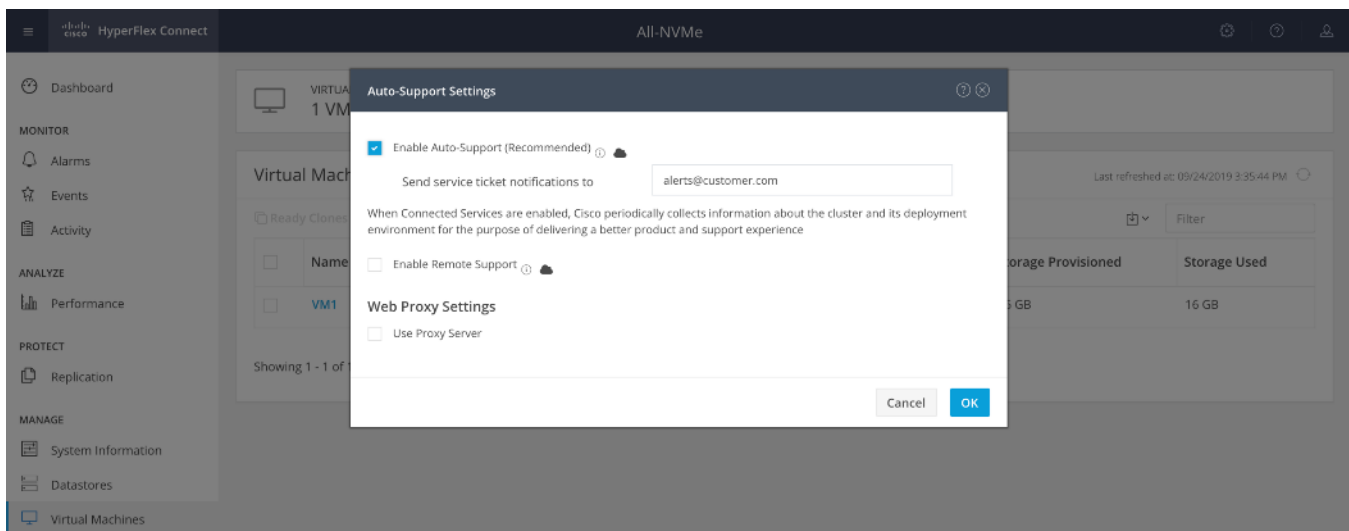
A list of events that will automatically open a support ticket with Cisco TAC is as follows:

- 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, follow these 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.

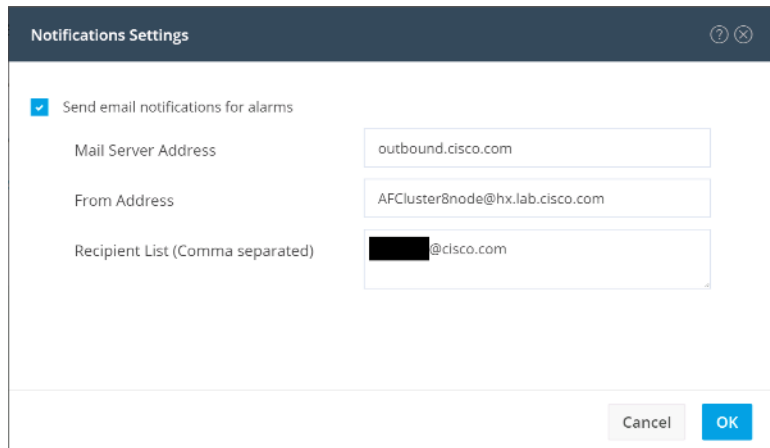


Email notifications that come directly from the HyperFlex cluster can also be enabled.

To enable direct email notifications, follow these 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.



Notifications Settings

Send email notifications for alarms

Mail Server Address: outbound.cisco.com

From Address: AFCluster8node@hx.lab.cisco.com

Recipient List (Comma separated): [REDACTED]@cisco.com

Cancel OK

Smart Licensing

HyperFlex 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, follow these 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

2. 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).
3. Navigate to Cisco Software Central (<https://software.cisco.com/>) and log in to your Smart Account.
4. From Cisco Smart Software Manager, generate a registration token.
5. In the License pane, click Smart Software Licensing to open Cisco Smart Software Manager.
6. Click Inventory.
7. From the virtual account where you want to register your HX storage cluster, click General, and then click New Token.
8. 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.
9. Click Create Token.
10. From the New ID Token row, click the Actions drop-down list, and click Copy.
11. Log into a controller VM.
12. Register your HX storage cluster, where *idtoken-string* is the New ID Token from Cisco Smart Software Manager.

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

13. 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.

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 6.7 Update 3 ISO release with the filename: HX-ESXi-6.7U3-15160138-Cisco-Custom-6.7.3.3-install-only.iso and is available on the Cisco web site:

[https://software.cisco.com/download/home/286305544/type/286305994/release/4.0\(2a\)](https://software.cisco.com/download/home/286305544/type/286305994/release/4.0(2a))

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 HyperFlex Cluster

If a Cisco HyperFlex cluster needs to be reinstalled, contact your local Cisco account or support team in order to be provided with a cluster cleanup guide.

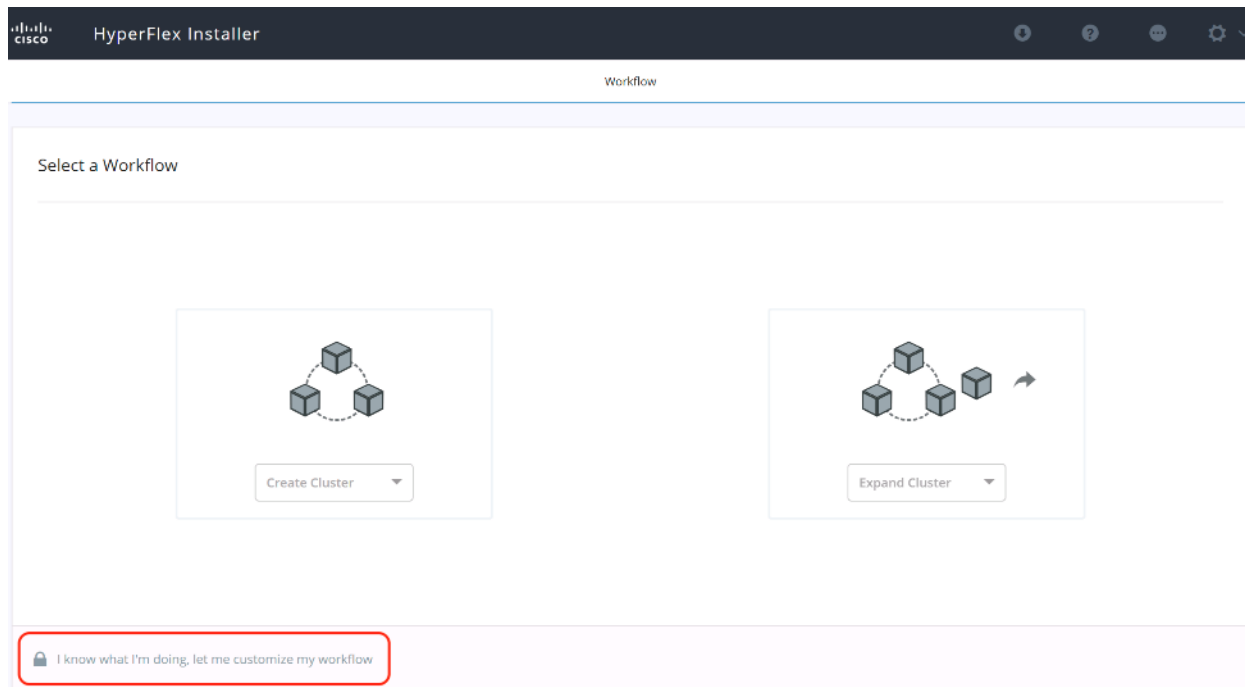


This process will be destructive and result in the loss of all the VMs and all the data stored in the HyperFlex distributed filesystem.

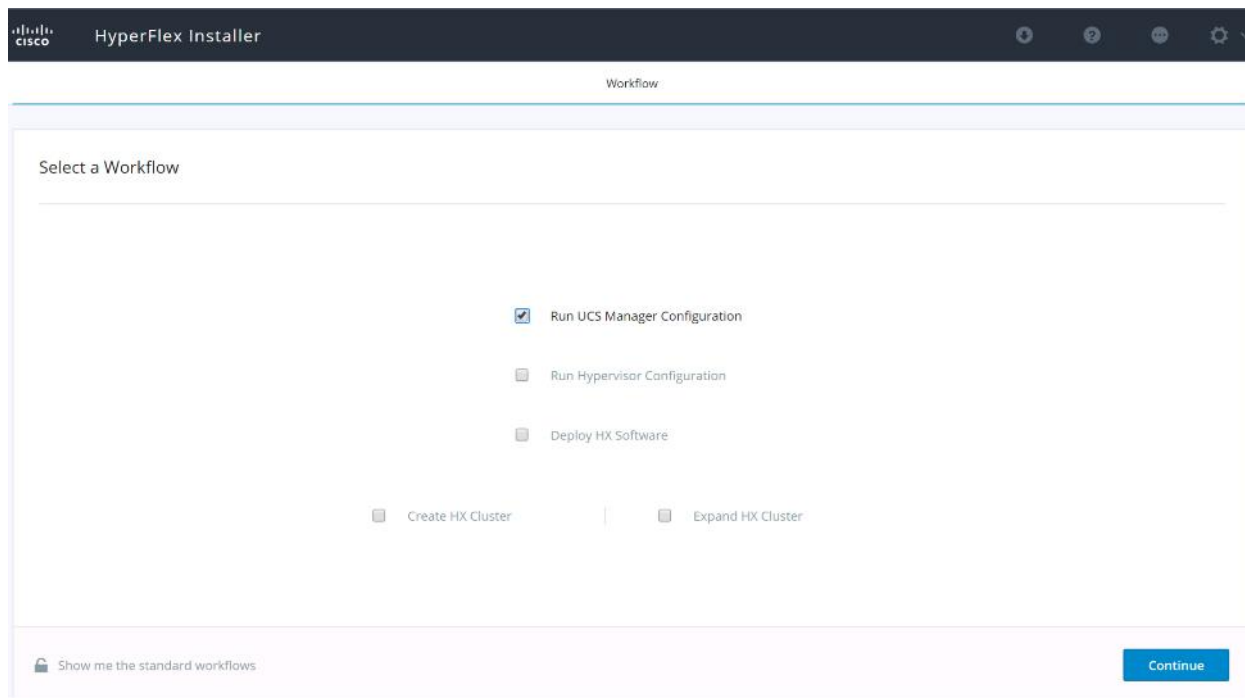
To reinstall a HyperFlex cluster, follow these steps:

1. Clean up the existing environment by:
 - a. Delete the existing HX virtual machines and HX datastores.
 - b. Destroy the HX cluster.
 - c. Remove the HX cluster from vCenter.
 - d. Remove the vCenter MOB entries for the HX extension.
 - e. Delete the HX sub-organization and HX VLANs in Cisco UCS Manager.

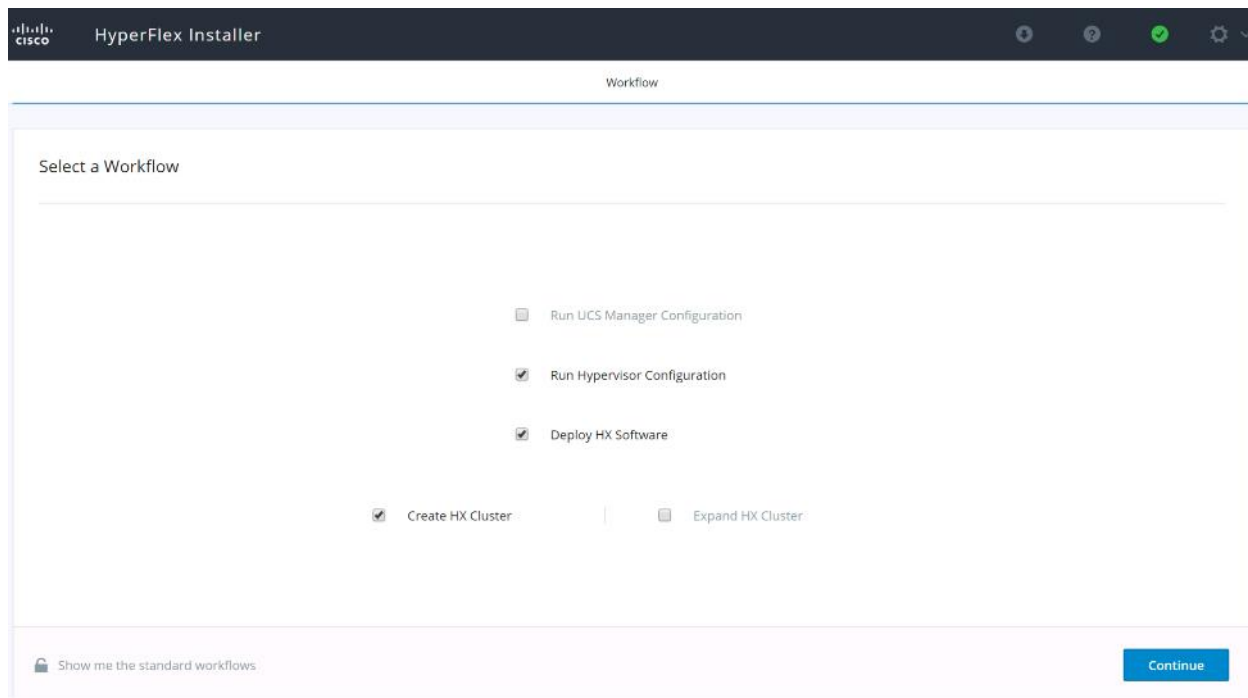
- Using the HX OVA-based installer VM, use the customized version of the installation workflow by selecting the “I know what I am doing” link.



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



- 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](#).
- 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.



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 applied. 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, follow these steps:

1. Copy the *HX-ESXi-6.7U3-15160138-Cisco-Custom-6.7.3.3-install-only.iso* file to an available web server folder, NFS share or CIFS share. In this example, an open internal web server folder is used.
2. In Cisco UCS Manager, click Servers.
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 CIFS as the protocol.
8. Enter the IP address of the CIFS server where the file was copied, for example: 10.29.132.120
9. Select None as the Image Variable Name.
10. Enter HX-ESXi-6.7U3-15160138-Cisco-Custom-6.7.3.3-install-only.iso as the Remote File.
11. Enter the Remote Path to the installation file.

Create vMedia Mount



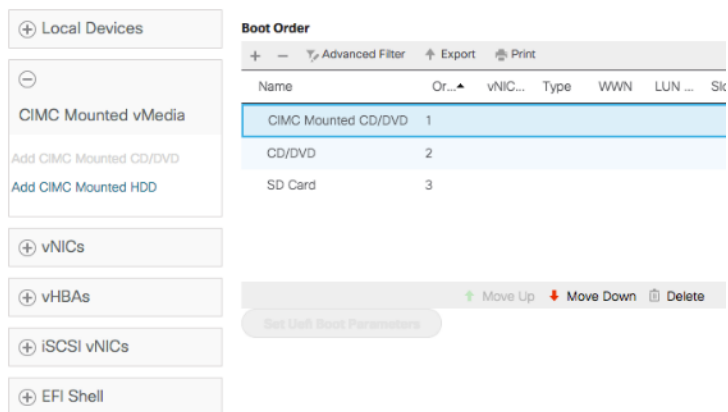
Name	:	<input type="text" value="ESXi"/>
Description	:	<input type="text"/>
Device Type	:	<input checked="" type="radio"/> CDD <input type="radio"/> HDD
Protocol	:	<input type="radio"/> NFS <input checked="" type="radio"/> CIFS <input type="radio"/> HTTP <input type="radio"/> HTTPS
Authentication Protocol	:	<input type="text" value="Ntlm"/>
Hostname/IP Address	:	<input type="text" value="10.29.132.145"/>
Image Name Variable	:	<input checked="" type="radio"/> None <input type="radio"/> Service Profile Name
Remote File	:	<input type="text" value="HX-ESXi-6.7U3-15160138-Cisco-Custom-6.7.3.3-"/>
Remote Path	:	<input type="text" value="iso"/>
Username	:	<input type="text"/>
Password	:	<input type="password"/>
Remain on Fieect	:	<input type="checkbox"/>

OK

Cancel

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.

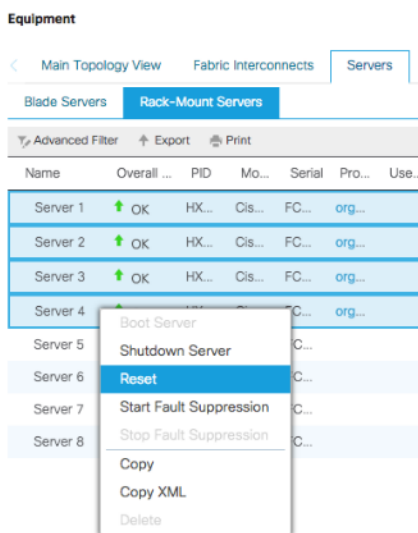


Install ESXi

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, follow these steps:

1. In Cisco UCS Manager, click Equipment.
2. Expand Equipment > Rack mounts > Servers > Server 1.
3. In the configuration pane, click KVM Console.

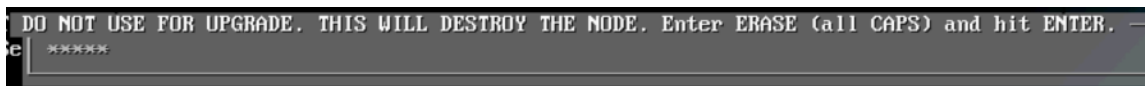
4. The remote KVM Console window will open in a new browser tab. Click Continue to any security alerts that appear and click the hyperlink to start the remote KVM session.
5. Repeat Steps 2-4 for all additional servers whose console you need to monitor during the installation.
6. In Cisco UCS Manager, click Equipment.
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 rebooted, 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 "ERASE" in all uppercase 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 temporarily, but they can be safely ignored. When the process is complete, the standard ESXi console screen will be seen as below:

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. The default predefined administrative account is named “admin”. The password for the default admin 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 into 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, follow these steps:

1. Using a web browser, open the HyperFlex cluster’s management IP address via HTTPS.
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:

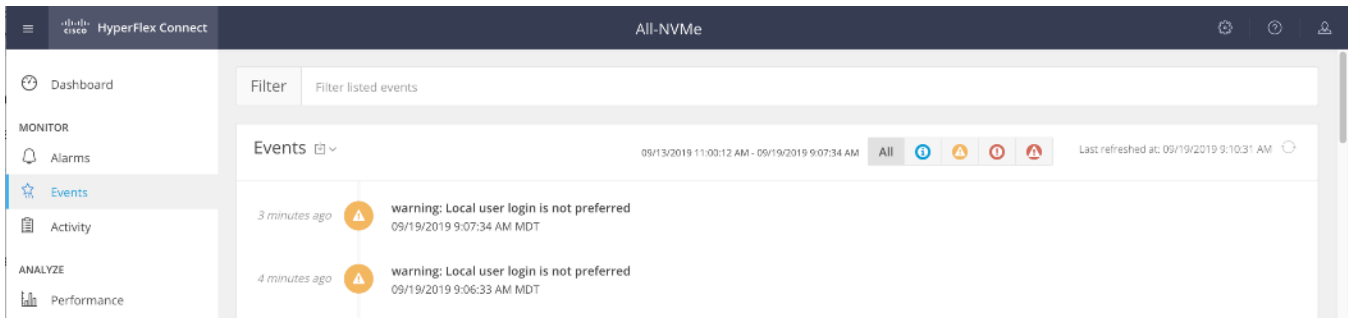
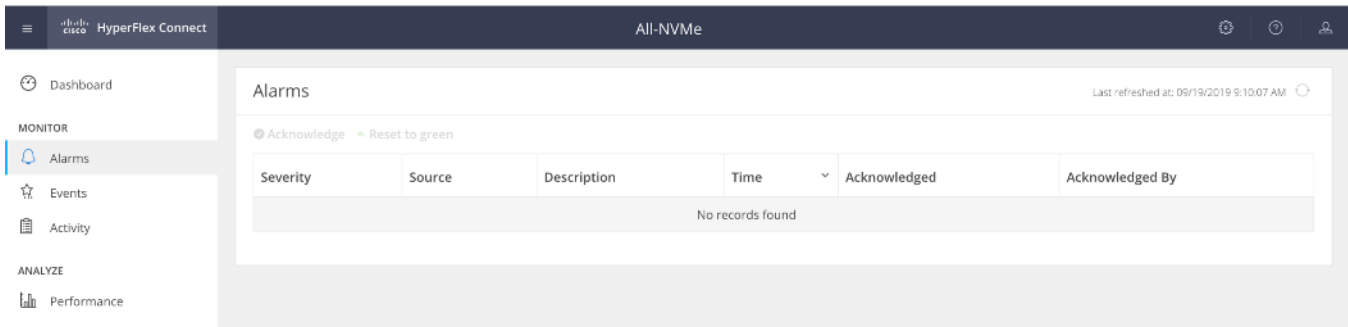
- Cluster operational status, overall cluster health, and the cluster’s current node failure tolerance.
- Cluster storage capacity used and free space, compression and deduplication savings, and overall cluster storage optimization statistics.
- Cluster size and individual node health.

- Cluster IOPs, storage throughput, and latency for the past 1 hour.

Monitor

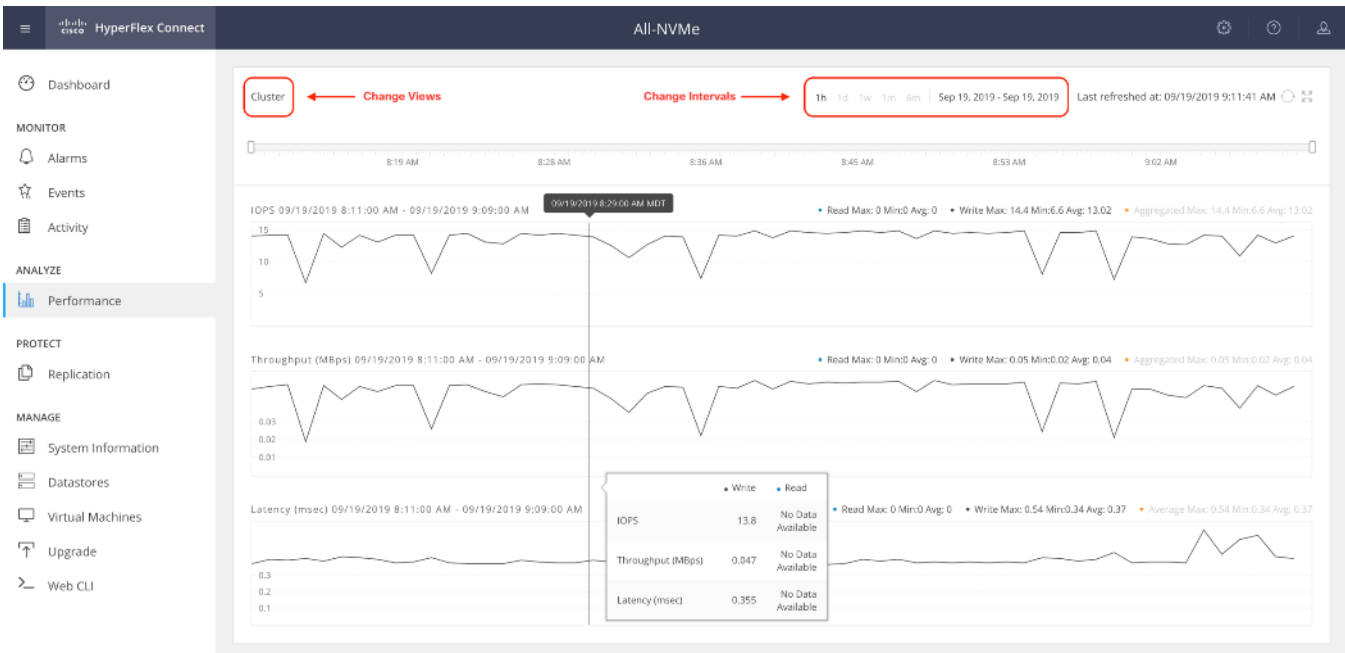
HyperFlex Connect provides for additional monitoring capabilities, including:

- 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.

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 self-encrypting disks can be securely erased.
- **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.
- **Upgrade:** One-click upgrades to the HXDP software, ESXi host 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.

- Dashboard
- MONITOR
 - Alarms
 - Events
 - Activity
- ANALYZE
 - Performance
- PROTECT
 - Replication
- MANAGE
 - System Information
 - Datastores
 - Virtual Machines
 - Upgrade
 - Web CLI

System Overview Nodes Disks Last refreshed at: 09/19/2019 9:18:58 AM

All-NVMe ONLINE Actions

vCenter	https://vcenter.hx.lab.cisco.com	Hypervisor	6.7.0-13473784	Total Capacity	13.39 TB	DNS Server(s)	10.29.133.110
Uptime	5 days, 22 hours, 7 minutes, 38 seconds	HXDP Version	4.0.1b-33133	Available Capacity	13.25 TB	NTP Server(s)	ntp2.hx.lab.cisco.com,ntp1.hx.la b.cisco.com
				Data Replication Factor	3	Controller Access over SSH	Enabled

hxaf220m5n-01 | HXAF220C-M55N | 7 Disks (1 Caching, 6 Persistent)

Online HXDP Version 4.0(1b)

Type Hyper Converged

Hypervisor Status Online | Hypervisor Address 10.29.133.174

hxaf220m5n-02 | HXAF220C-M55N | 7 Disks (1 Caching, 6 Persistent)

Online HXDP Version 4.0(1b)

Type Hyper Converged

Hypervisor Status Online | Hypervisor Address 10.29.133.175

hxaf220m5n-03 | HXAF220C-M55N | 7 Disks (1 Caching, 6 Persistent)

Online HXDP Version 4.0(1b)

Type Hyper Converged

Hypervisor Status Online | Hypervisor Address 10.29.133.176

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 HX Connect Dashboard summary screen.
3. Create a test virtual machine that accesses the HyperFlex datastore and is able to perform read/write operations.
4. Perform a 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 its 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 in the HX Connect Dashboard.
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 in the HX Connect Dashboard after a brief time to restart the services on that node. vSphere DRS should rebalance the VM distribution across the cluster over time.



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.

Build the Virtual Machines and Environment for Workload Testing

Software Infrastructure Configuration

This section details how to configure the software infrastructure components that comprise this solution.

Install and configure the infrastructure virtual machines by following the process provided in [Table 8](#).

Table 8. Test Infrastructure Virtual Machine Configuration

Configuration	Citrix Virtual Desktops Controllers Virtual Machines	Citrix Provisioning Services Servers Virtual Machines
Operating System	Microsoft Windows Server 2019	Microsoft Windows Server 2019
Virtual CPU amount	6	8
Memory amount	8 GB	12 GB
Network	VMNIC	Network
Disk-1 (OS) size and location	40 GB	Disk-1 (OS) size and location
Disk-2 size and location	500GB	Disk-2 (Data) Paravirtual SCSI adapter with ReFS format
Configuration	Microsoft Active Directory DC's Virtual Machines	Citrix Profile Servers Virtual Machines
Operating system	Microsoft Windows Server 2019	Operating system
Virtual CPU amount	4	
Memory amount	4 GB	
Network	VMNIC	
Disk size and location	40 GB	
Configuration	Microsoft SQL Server Virtual Machine	Citrix StoreFront Virtual Machine
Operating system	Microsoft Windows Server 2019	Microsoft Windows Server 2019
Virtual CPU amount	8	4
Memory amount	16 GB	8 GB
Network	VMNIC	Network
Disk-1 (OS) size and location	40 GB	Disk-1 (OS) size and location
Disk-2 size and location	200 GB Infra-DS volume	Disk-2 size and location

Configuration	Citrix License Server Virtual Machine	NetScaler VPX Appliance Virtual Machine
Operating system	Microsoft Windows Server 2019	NS11.1 52.13.nc
Virtual CPU amount	4	2
Memory amount	4 GB	2 GB
Network	VMNIC	Network
Disk size and location	40 GB	20 GB

Prepare the Master Images

This section details how to create the golden (or master) images for the environment. virtual machines for the master images must first be installed with the software components needed to build the golden images. For this CVD, the images contain the basics needed to run the Login VSI workload.

To prepare the master virtual machines for the Hosted Virtual Desktops (HVDs) and Hosted Shared Desktops (HSDs), there are three major steps to complete when the base virtual machine has been created:

- Installing OS
- Installing application software
- Installing the Virtual Delivery Agents (VDAs)

The master image HVD and HSD virtual machines were configured as listed in [Table 9](#).

Table 9. HVD and HSD Configurations

Configuration	HVDI Virtual Machines	HSD Virtual Machines
Operating system	Microsoft Windows 10 64-bit	Microsoft Windows Server 2019
Virtual CPU amount	2	8
Memory amount	4.0 GB (reserved)	24 GB (reserved)
Network	VMNIC vm-network	VMNIC vm-network
Citrix PVS vDisk size and location	24 GB	40 GB
Citrix PVS write cache	6 GB	24 GB
Disk size		
Additional software used for	Microsoft Office 2016	Microsoft Office 2016

Configuration	HVDI Virtual Machines	HSD Virtual Machines
testing	Login VSI 4.1.32 (Knowledge Worker Workload)	Login VSI 4.1.32 (Knowledge Worker Workload)

Install and Configure Citrix Desktop Delivery Controller, Citrix Licensing, and StoreFront

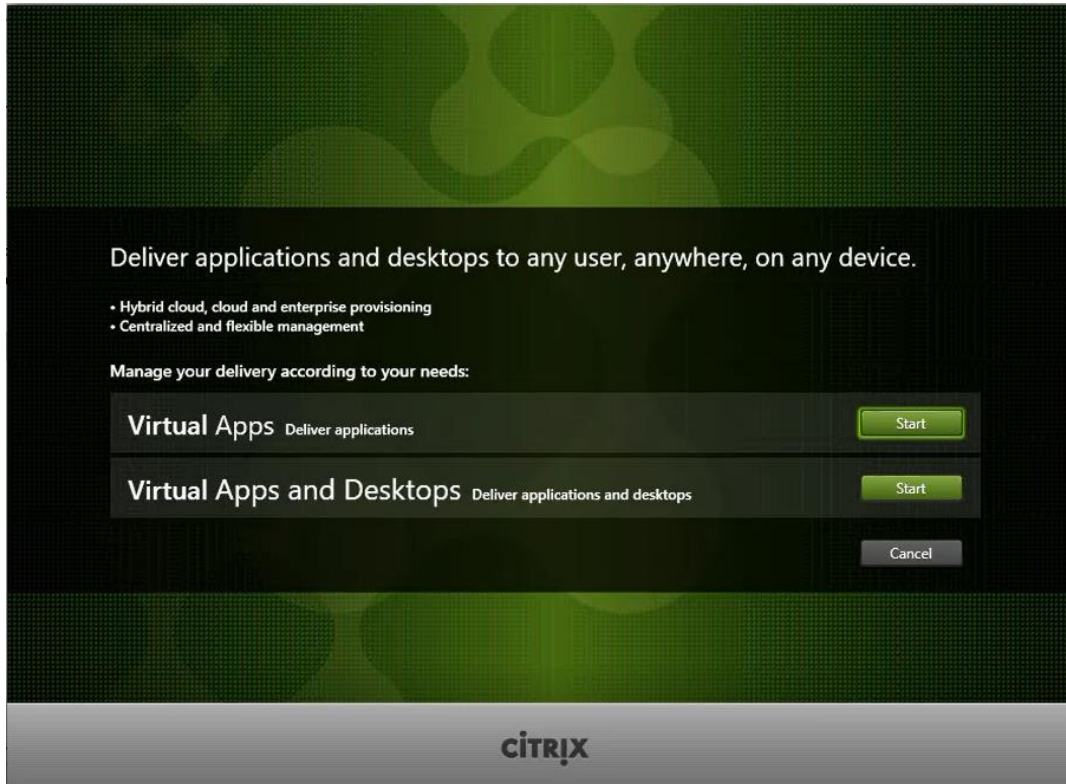
This section details the installation of the core components of the Citrix Virtual Apps and Desktops 1912 LTSR system. This CVD provides the process to install two Desktop Delivery Controllers to support hosted shared desktops (HSD), non-persistent virtual desktops (VDI), and persistent virtual desktops (VDI).

The process of installing the Desktop Delivery Controller also installs other key Citrix Desktop software components, including Studio, which is used to create and manage infrastructure components, and Director, which is used to monitor performance and troubleshoot problems.

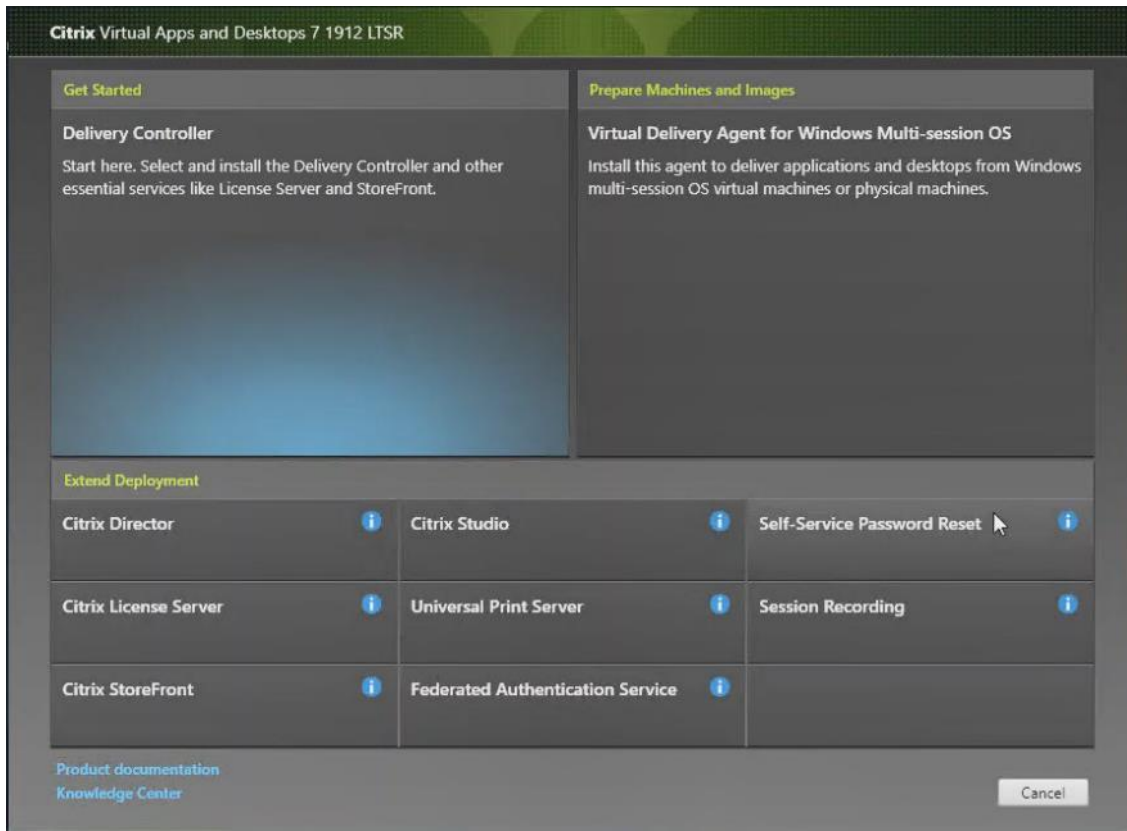
Install Citrix License Server

To install the Citrix License Server, follow these steps:

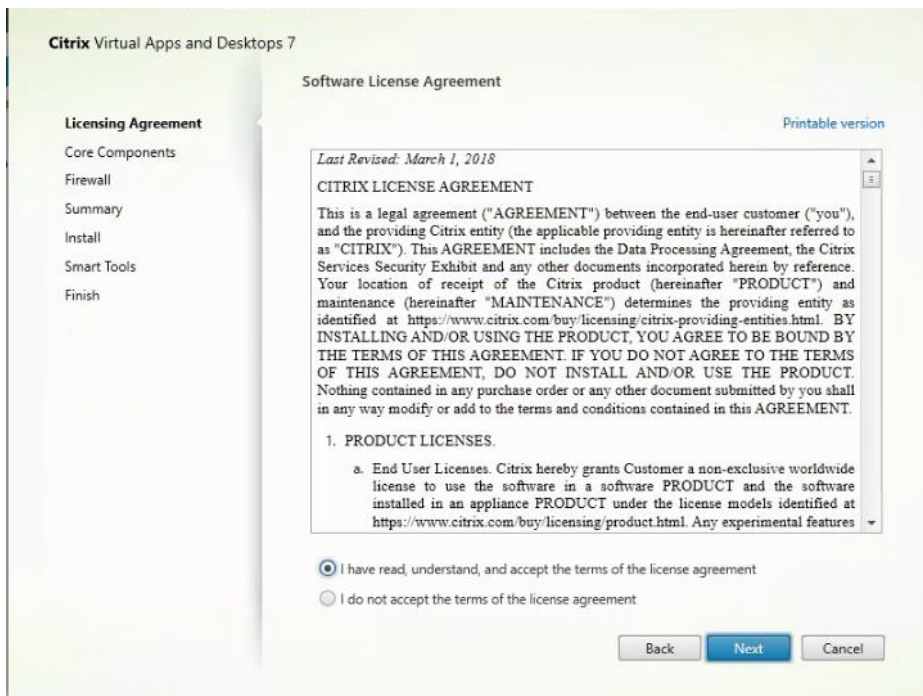
1. To begin the installation, connect to the first Citrix License server and launch the installer from the Citrix Virtual Apps and Desktops 1912 LTSR ISO.
2. Click Start.



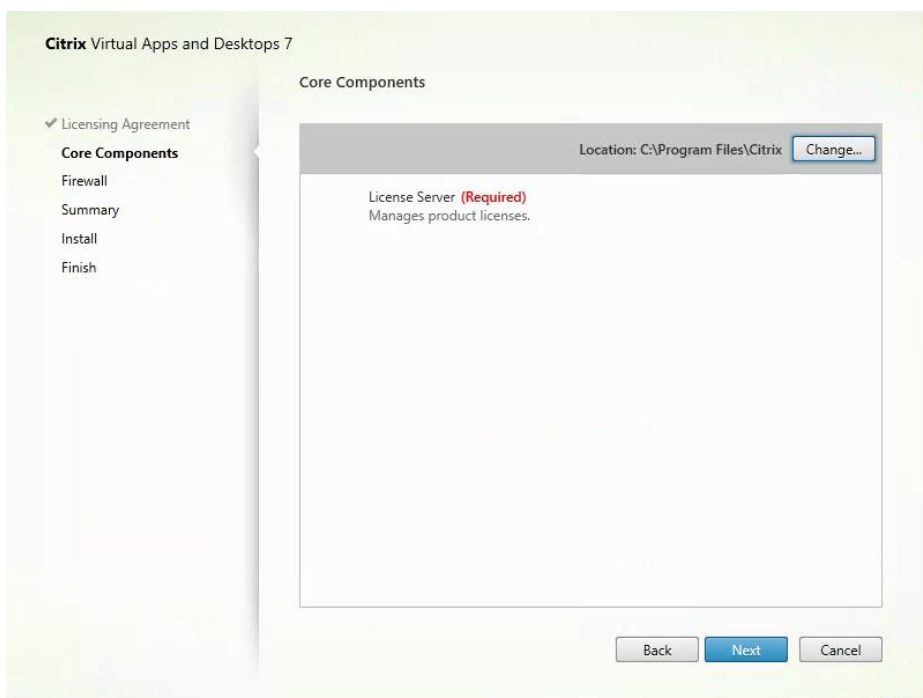
3. Click "Extend Deployment - Citrix License Server."



4. Read the Citrix License Agreement.
5. If acceptable, indicate your acceptance of the license by selecting the “I have read, understand, and accept the terms of the license agreement” radio button.
6. Click Next.

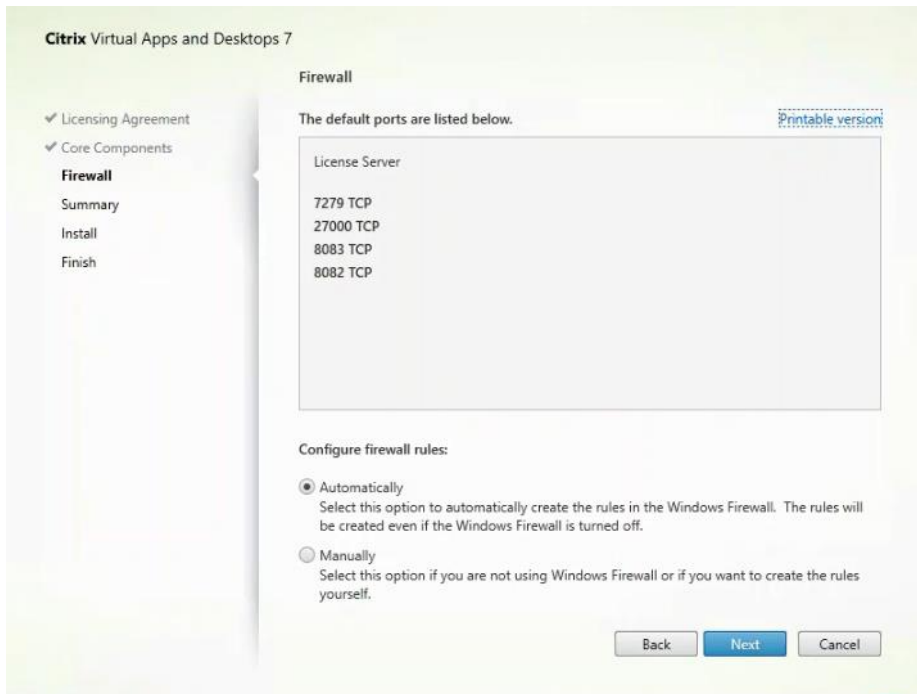


7. Click Next.

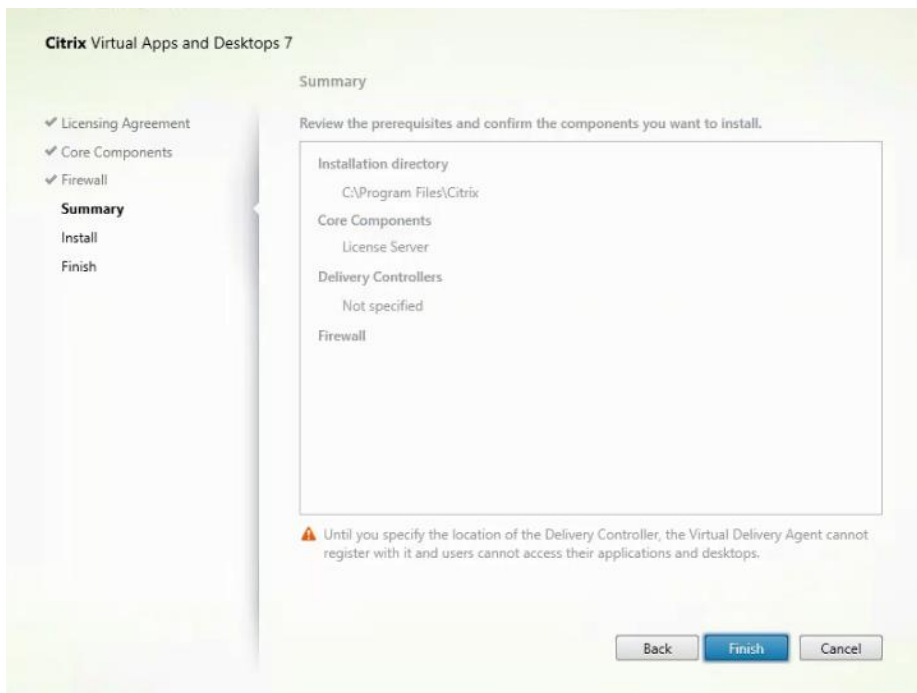


8. Select the default ports and automatically configured firewall rules.

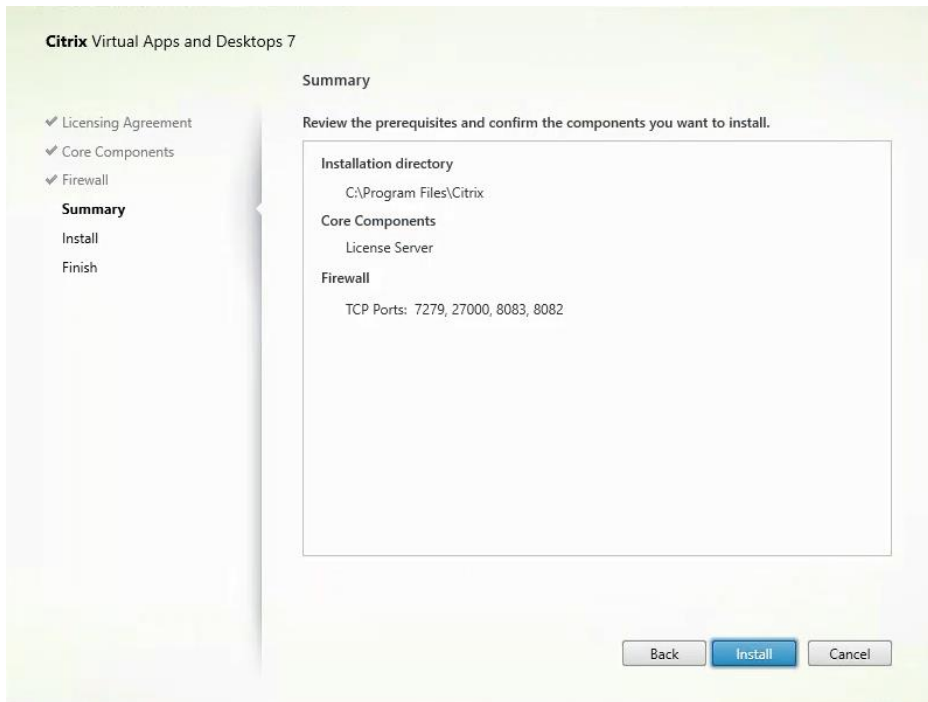
9. Click Next.



10. Click Install.



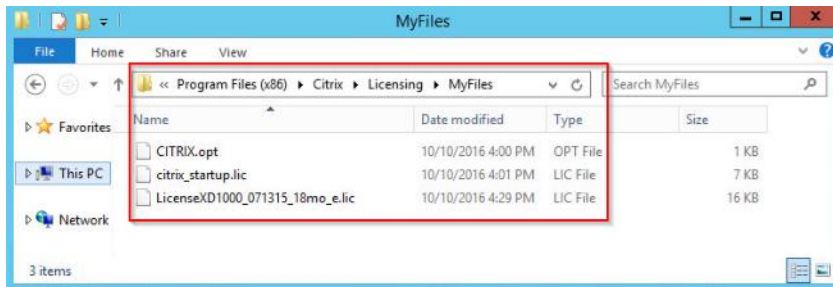
11. Click Finish to complete the installation.



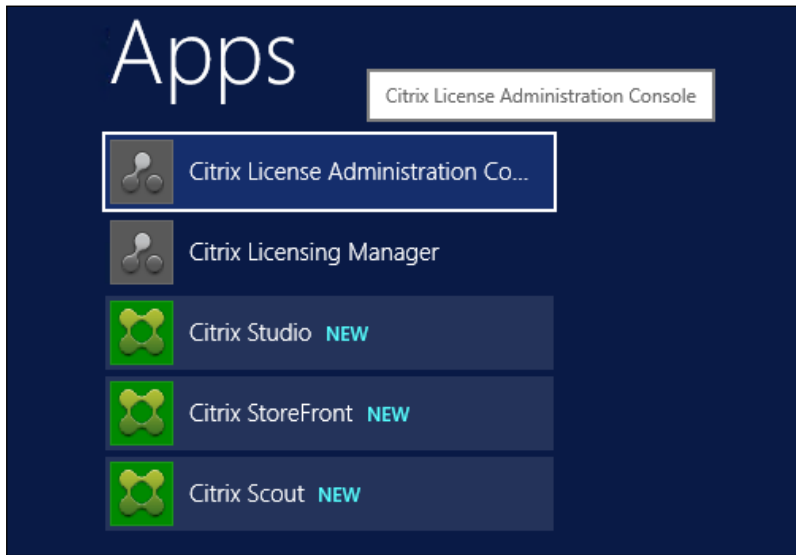
Install Citrix Licenses

To install the Citrix Licenses, follow these steps:

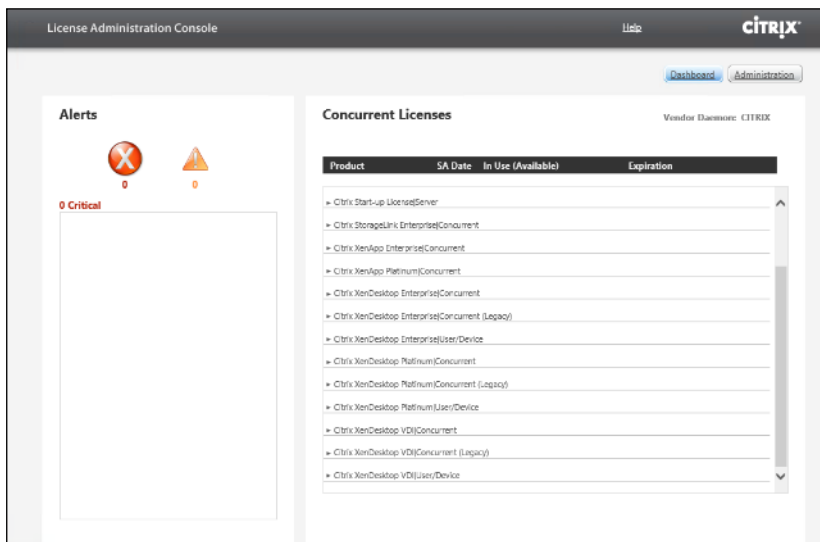
1. Copy the license files to the default location (C:\Program Files (x86)\Citrix\Licensing\ MyFiles) on the license server.



2. Restart the server or Citrix licensing services so that the licenses are activated.
3. Run the application Citrix License Administration Console.



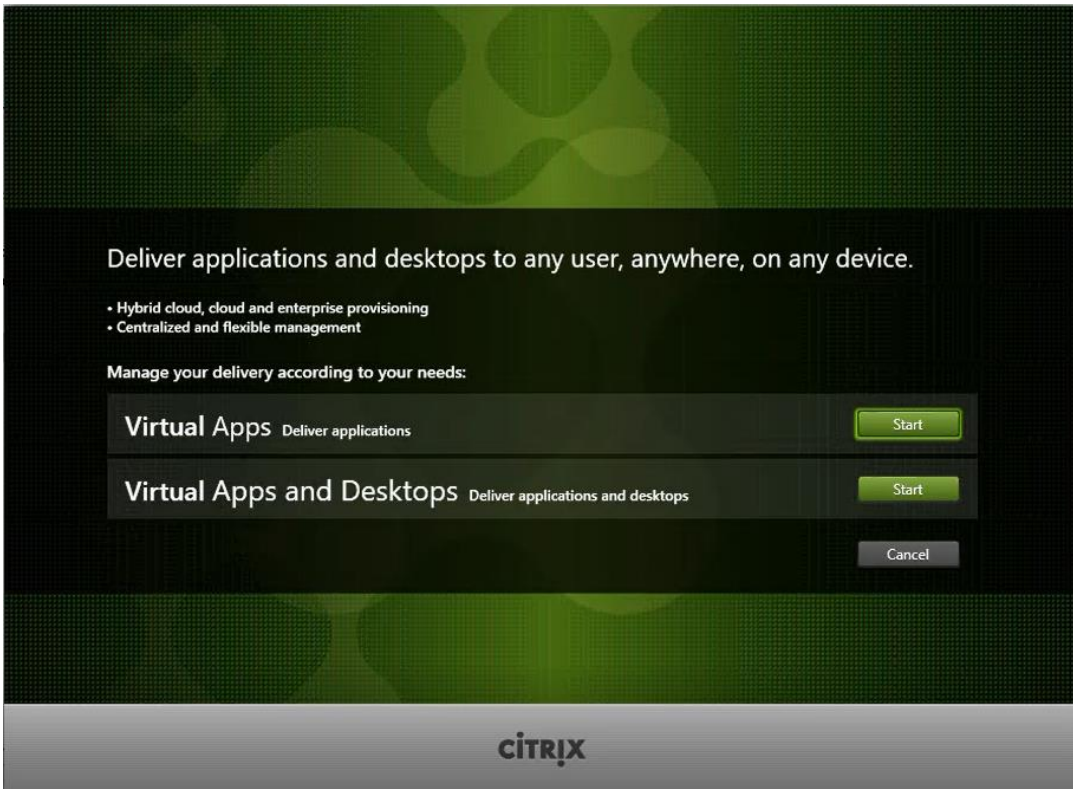
4. Confirm that the license files have been read and enabled correctly.



Install Citrix Desktop Broker/Studio

To install Citrix Desktop, follow these steps:

1. Connect to the first Citrix VDI server and launch the installer from the Citrix Desktop 1912 LTSR ISO.
2. Click Start.

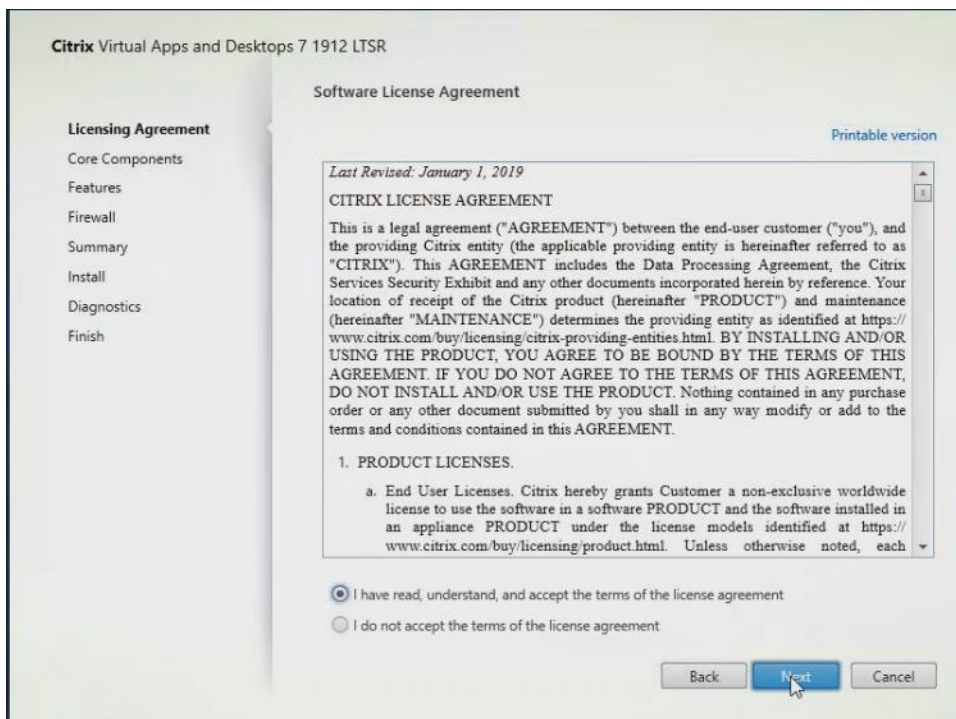


The installation wizard presents a menu with three subsections.

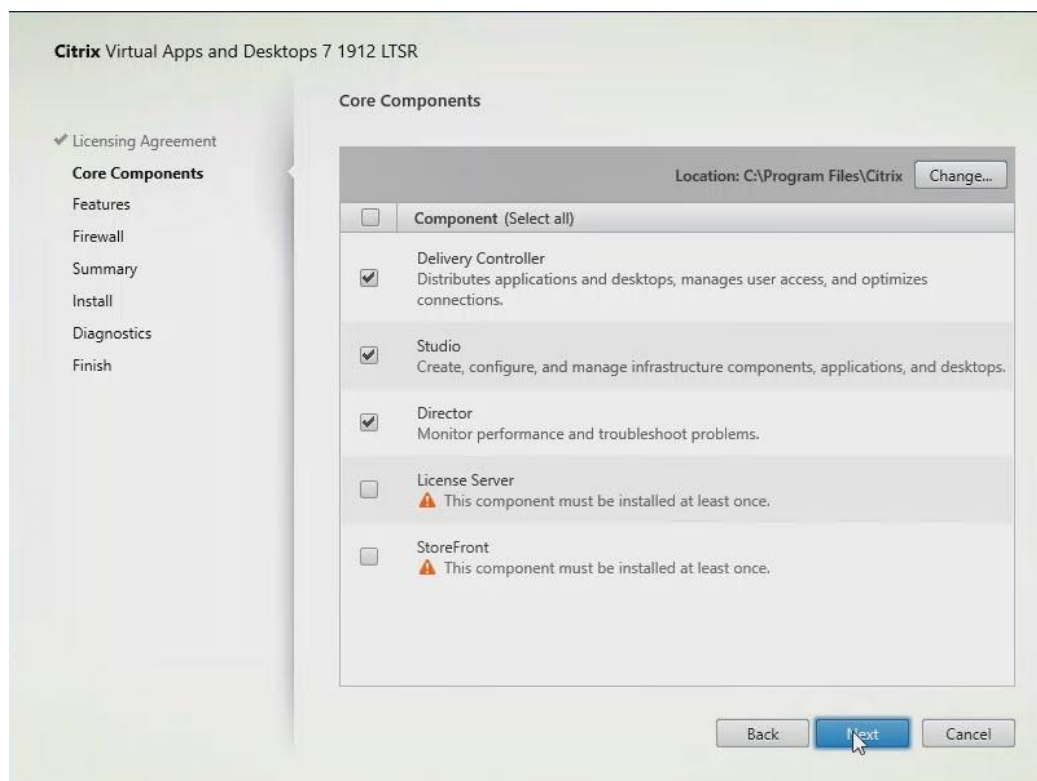
3. Click "Get Started - Delivery Controller."



4. Read the Citrix License Agreement and if acceptable, indicate your acceptance of the license by selecting the “I have read, understand, and accept the terms of the license agreement” radio button.
5. Click Next.

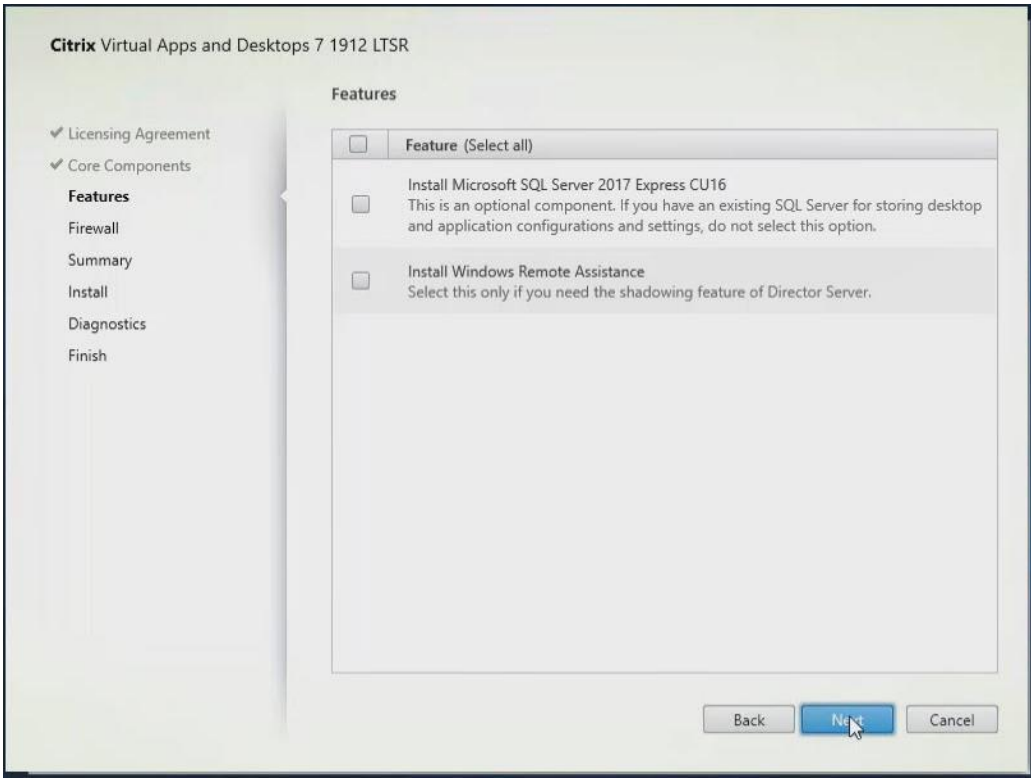


6. Select the components to be installed on the first Delivery Controller Server:
 - a. Delivery Controller
 - b. Studio
 - c. Director
7. Click Next.



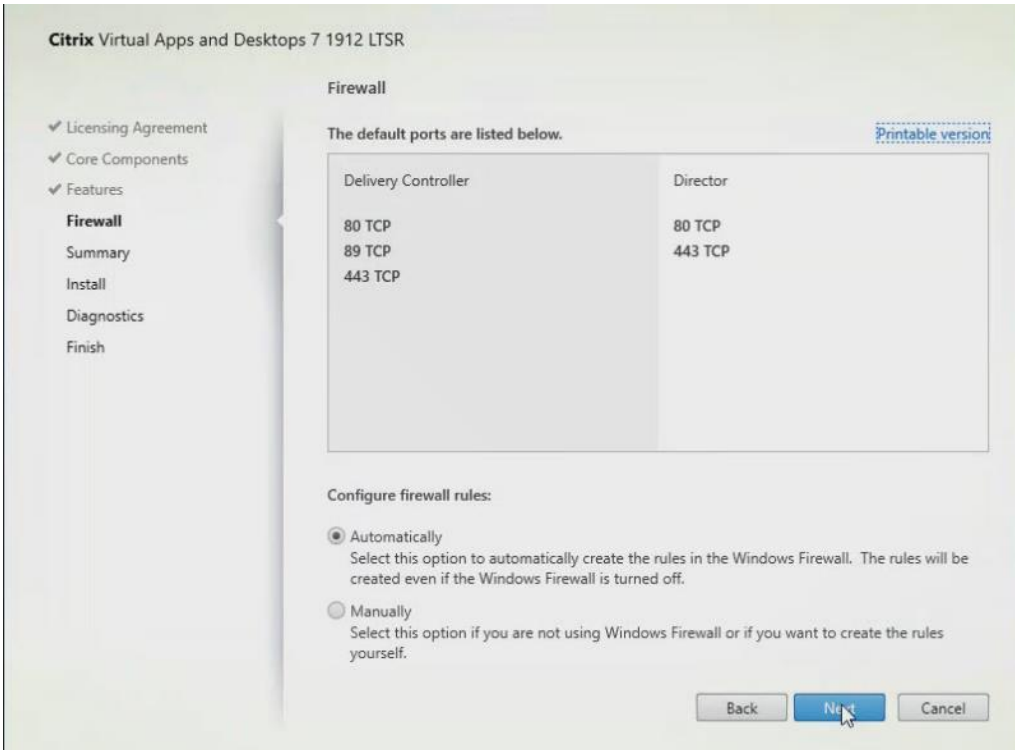
Dedicated StoreFront and License servers should be implemented for large-scale deployments.

8. Since a SQL Server will be used to Store the Database, leave “Install Microsoft SQL Server 2012 SP1 Express” unchecked.
9. Click Next.



10. Select the default ports and automatically configured firewall rules.

11. Click Next.

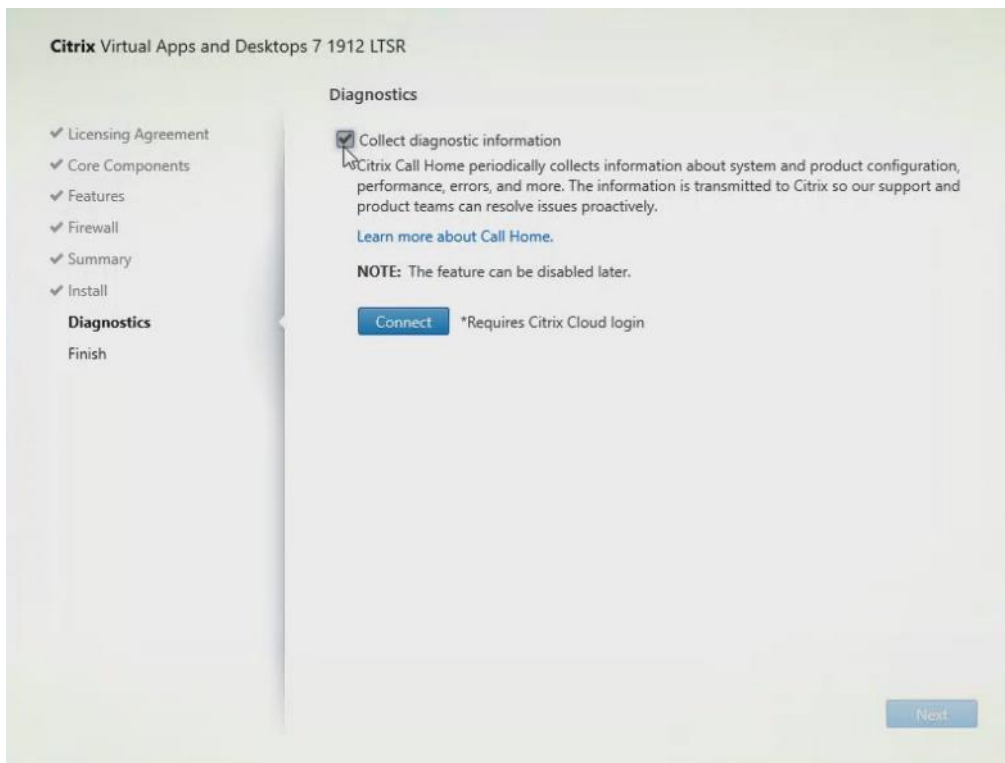


12. Click Install.



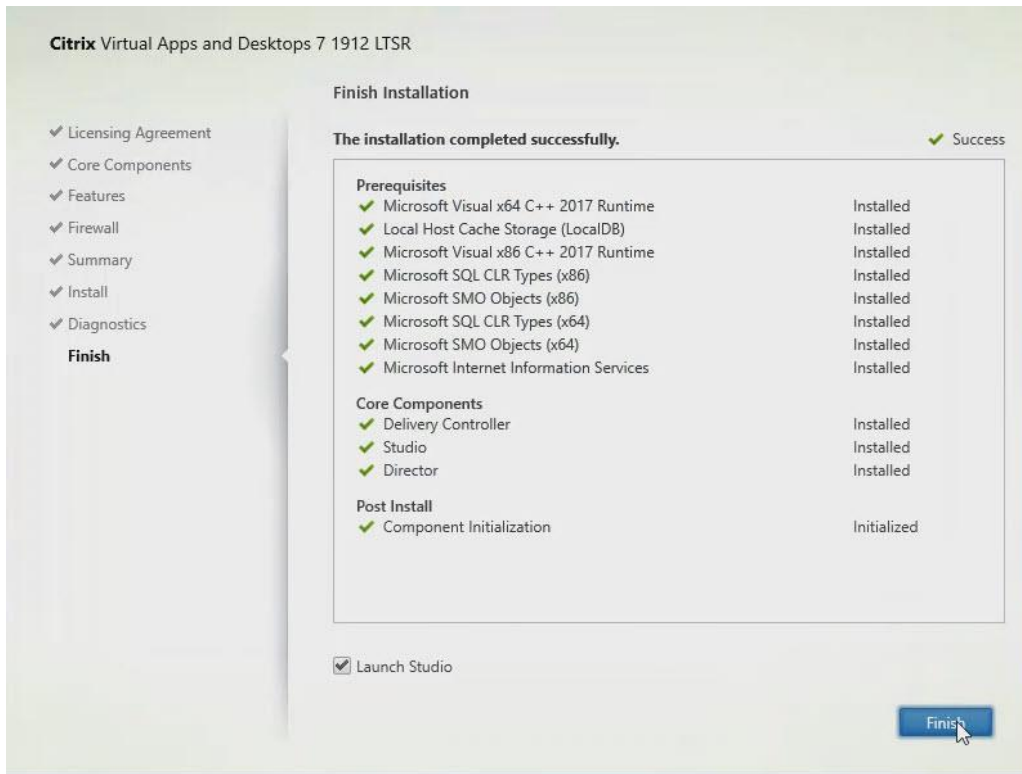
13. (Optional) Click the Call Home participation.

14. Click Next.



15. Click Finish to complete the installation.

16. (Optional) Check Launch Studio to launch Citrix Studio Console.



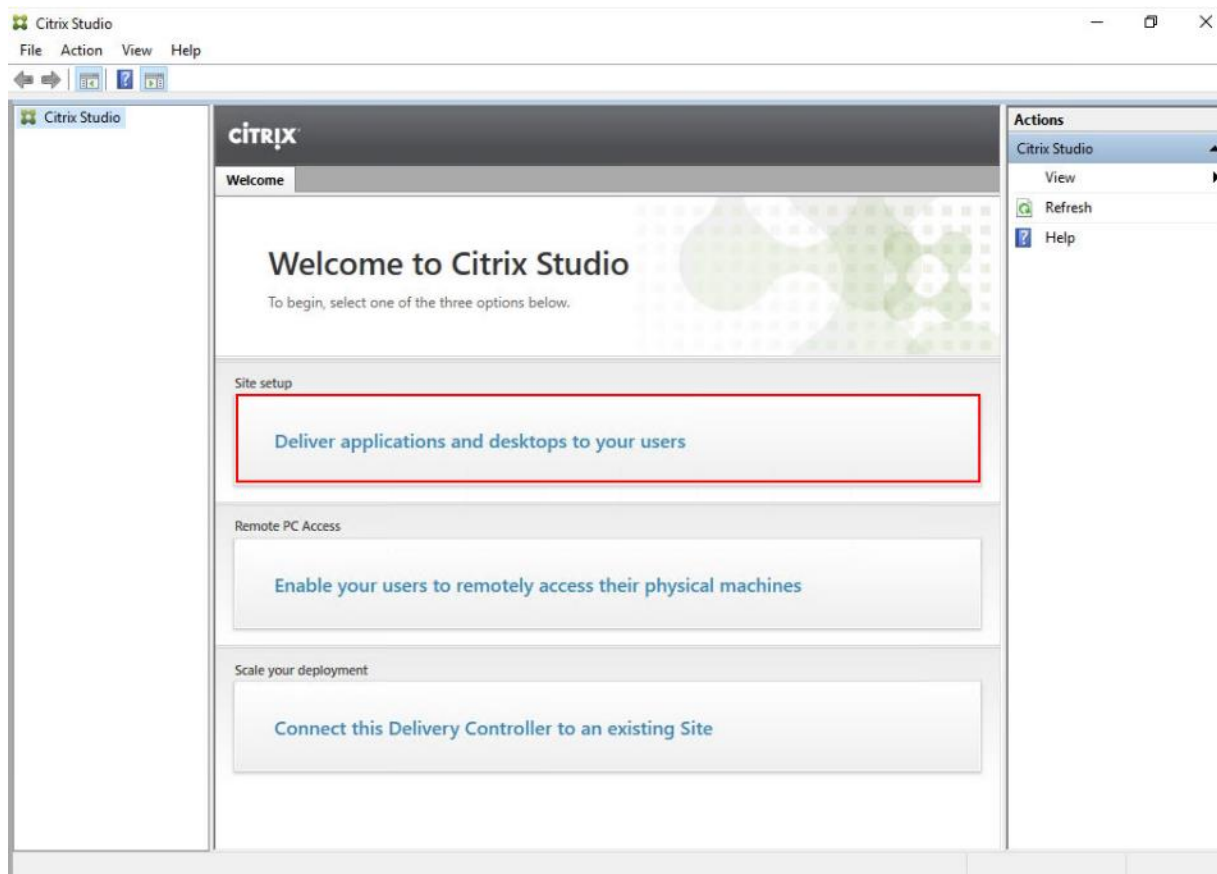
Configure the Citrix VDI Site

Citrix Studio is a management console that allows you to create and manage infrastructure and resources to deliver desktops and applications. Replacing Desktop Studio from earlier releases, it provides wizards to set up your environment, create workloads to host applications and desktops, and assign applications and desktops to users.

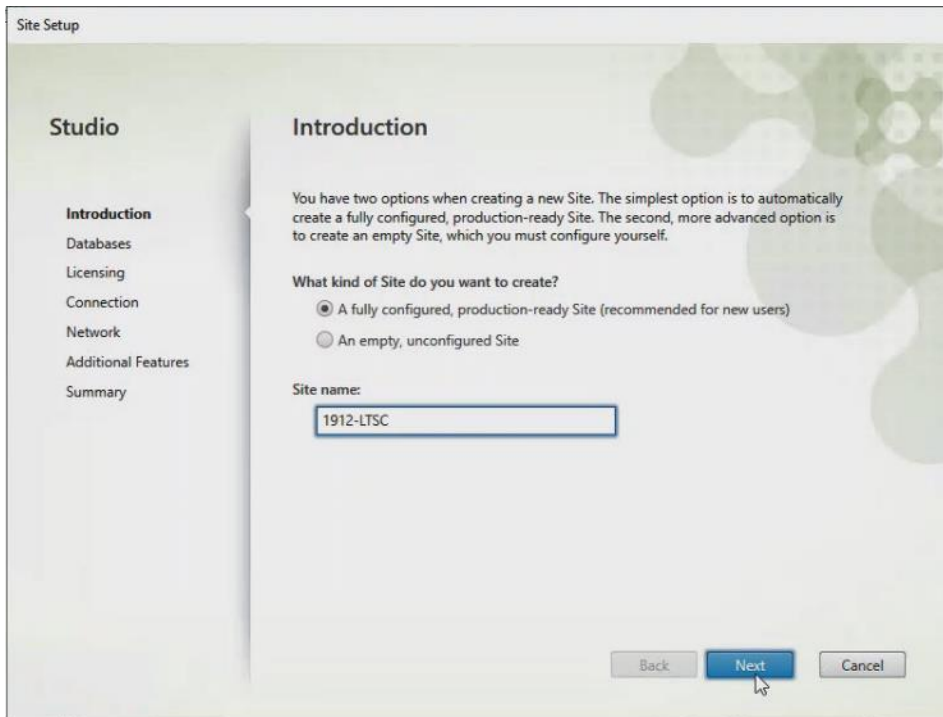
Citrix Studio launches automatically after the Citrix VDI Delivery Controller installation, or if necessary, it can be launched manually. Citrix Studio is used to create a Site, which is the core Citrix VDI environment consisting of the Delivery Controller and the Database.

To configure Citrix VDI, follow these steps:

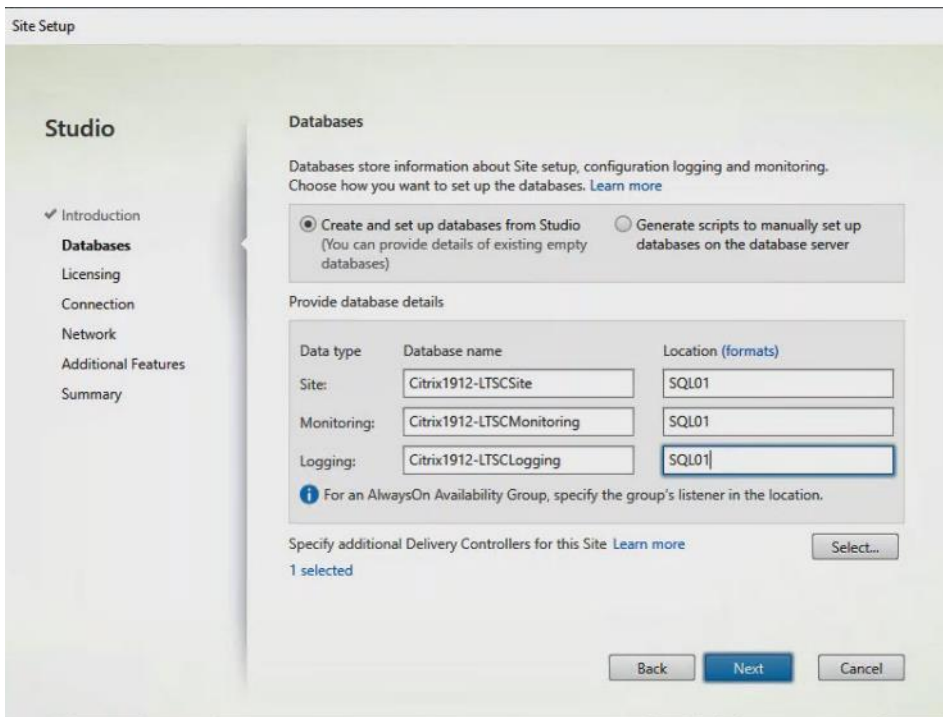
1. From Citrix Studio, click Deliver applications and desktops to your users.



2. Select the “A fully configured, production-ready Site” radio button.
3. Enter a site name.
4. Click Next.



5. Provide the Database Server Locations for each data type and click Next.



6. For an AlwaysOn Availability Group, use the group's listener DNS name.

7. Provide the FQDN of the license server.

8. Click Connect to validate and retrieve any licenses from the server.



If no licenses are available, you can use the 30-day free trial or activate a license file.

9. Select the appropriate product edition using the license radio button.

10. Click Next.

Site Setup

Studio

- Introduction
- Databases
- Licensing**
- Connection
- Network
- Additional Features
- Summary

Licensing

License server address: Connected to trusted server
View certificate

I want to:

- Use the free 30-day trial
You can add a license later.
- Use an existing license
The product list below is generated by the license server.

Product	Model
<input checked="" type="radio"/> Citrix XenDesktop Platinum	User/Device
<input type="radio"/> Citrix XenApp Platinum	Concurrent

11. Select the Connection type of 'Microsoft System Center Virtual Machine Manager'.

12. Enter the Connection Address to the SCVMM Server.

13. Enter the username (in username@domain format) for the vCenter account.

14. Provide the password for the Domain Admin account.

15. Provide a connection name.

16. Select the Studio tools radio button.

Add Connection and Resources

Studio

- Connection
- Storage Management
- Storage Selection
- Network
- Summary

Connection

Use an existing Connection

8x16

Create a new Connection

Connection type: VMware vSphere®

Connection address: *Example: https://vmware.example.com/sdk*

[Learn about user permissions](#)

User name: *Example: domain\username*

Password:

Connection name: *Example: MyConnection*

Create virtual machines using:

Studio tools (Machine Creation Services)
Select this option when using AppDisks, even if you are using Citrix Provisioning.

Other tools

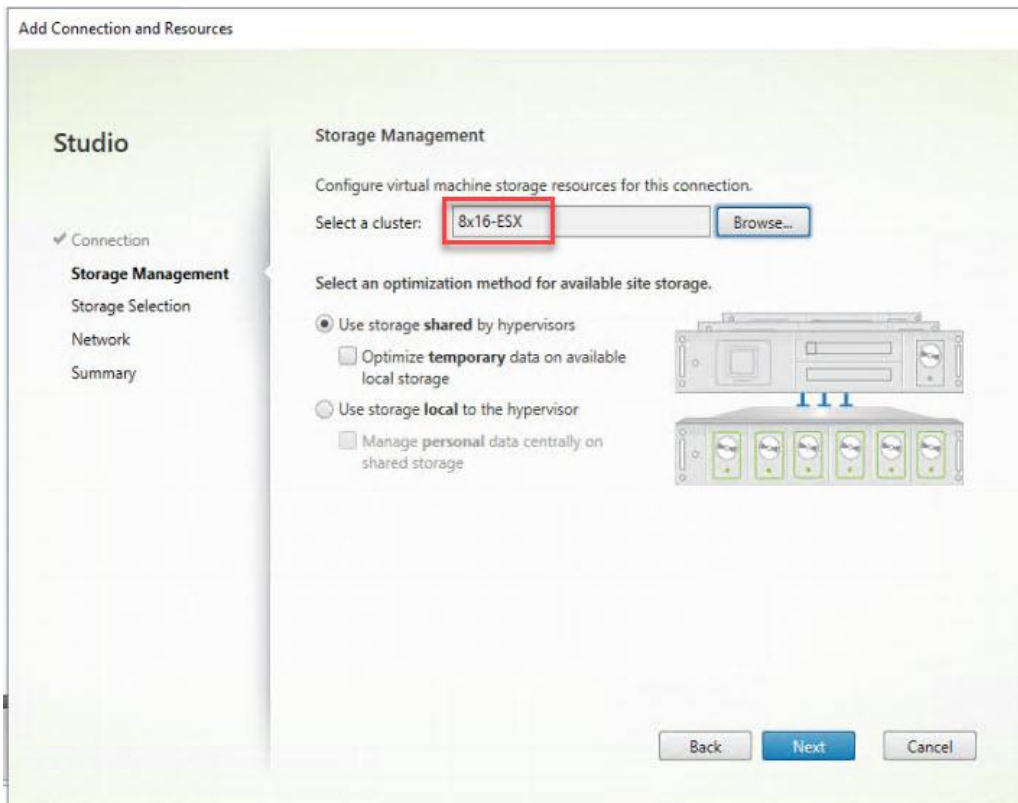
Back Next Cancel

17. Click Next.

18. Select HyperFlex Cluster that will be used by this connection.

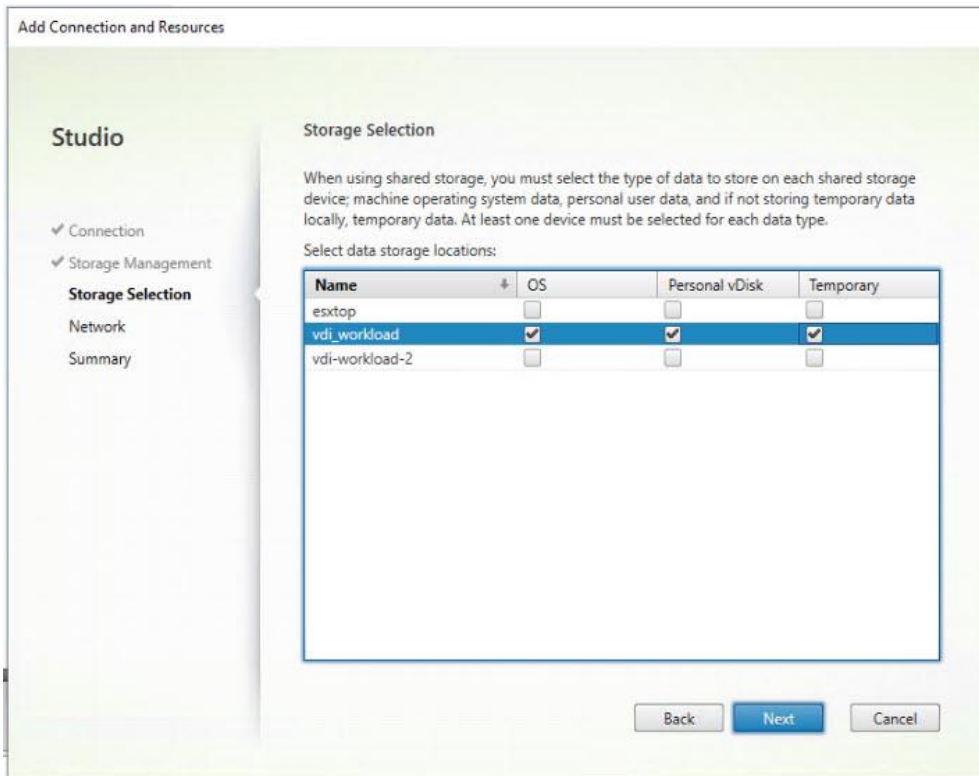
19. Check Studio Tools radio button required to support desktop provisioning task by this connection.

20. Click Next.



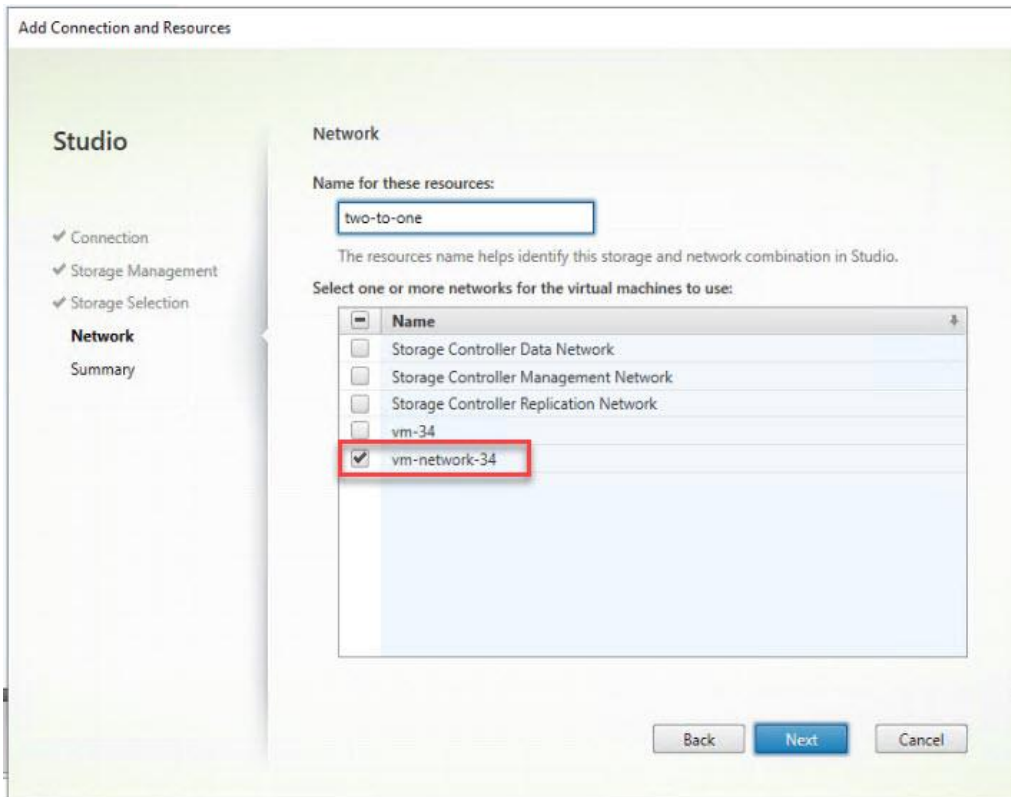
21. Make Storage selection to be used by this connection.

22. Click Next.



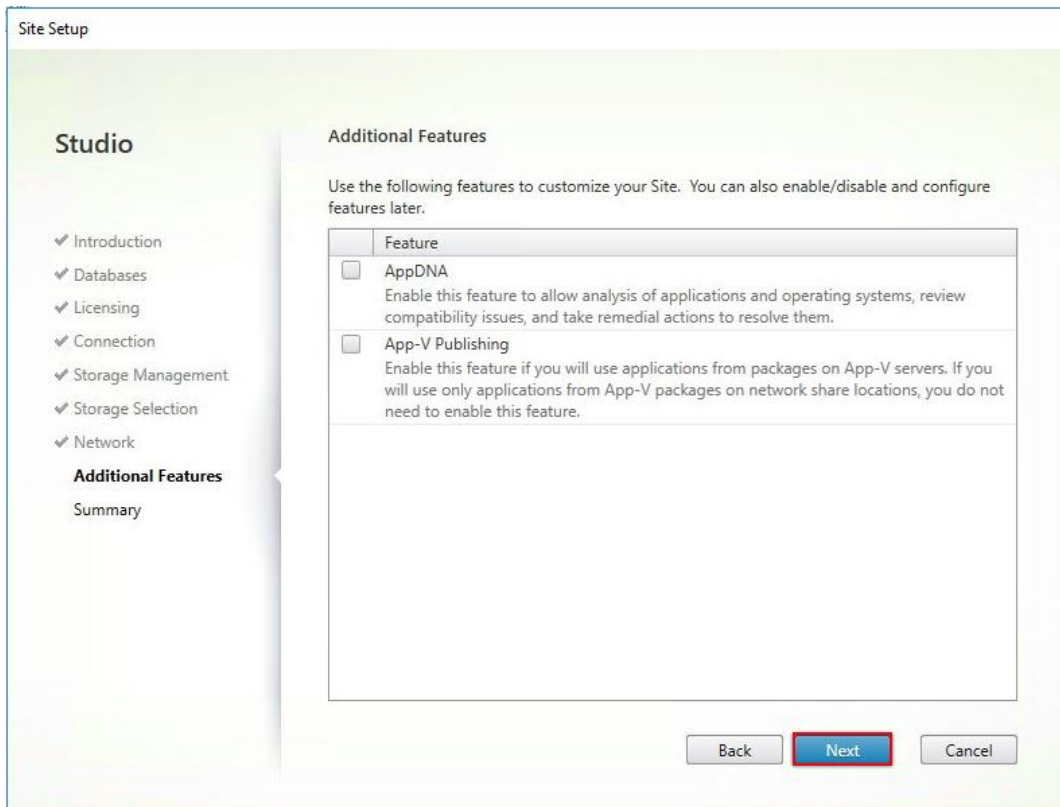
23. Make Network selection to be used by this connection.

24. Click Next.



25. Select Additional features.

26. Click Next.

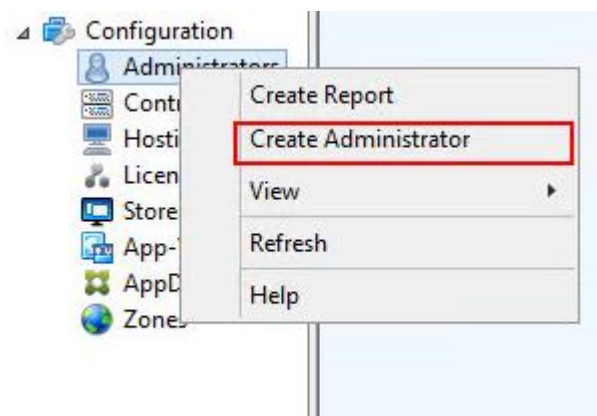


27. Review Site configuration Summary and click Finish.

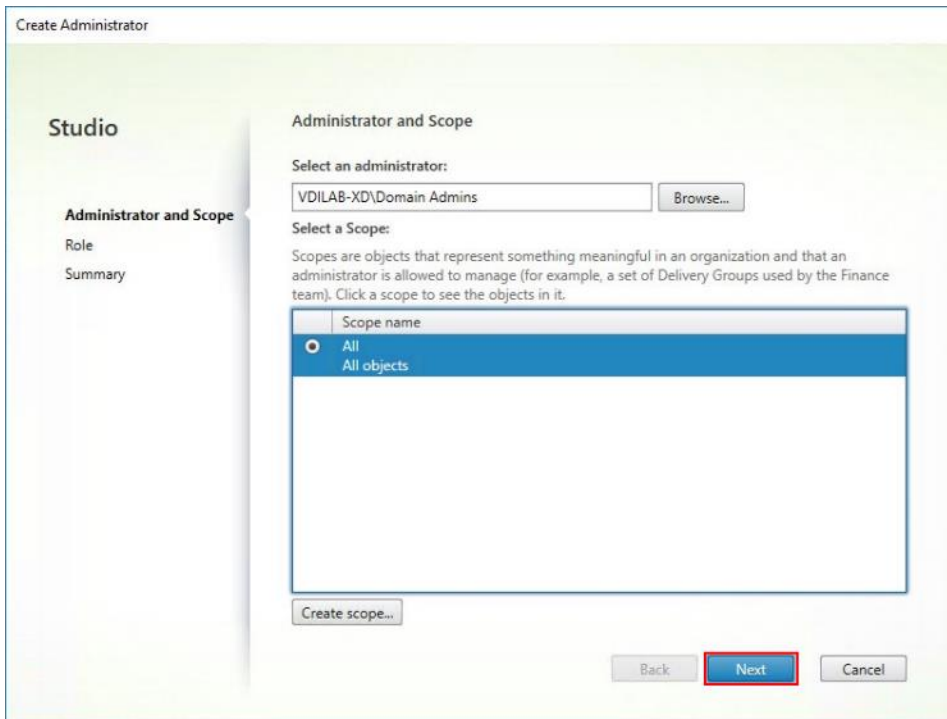
Configure the Citrix VDI Site Administrators

To configure the Citrix VDI site administrators, follow these steps:

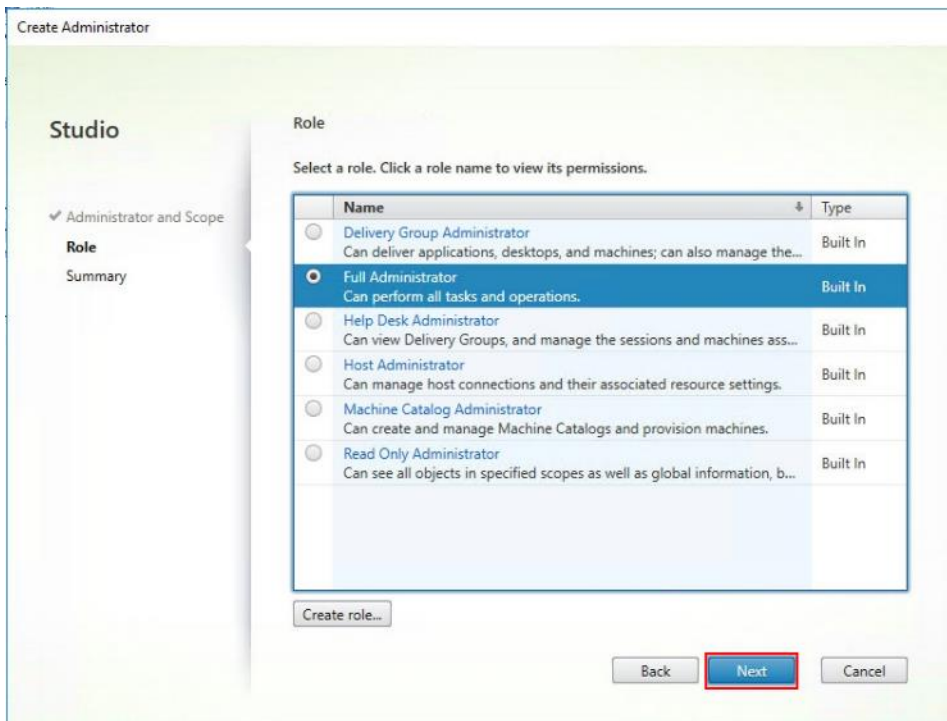
1. Connect to the Citrix VDI server and open Citrix Studio Management console.
2. From the Configuration menu, right-click Administrator and select Create Administrator from the drop-down list.



3. Select/Create appropriate scope and click Next.



4. Choose an appropriate Role.



5. Review the Summary, check Enable administrator, and click Finish.

Create Administrator

Studio

- ✓ Administrator and Scope
- ✓ Role
- Summary**

Summary

Administrator:	VDILAB-XD\Domain Admins
Scope:	All
Role:	Full Administrator

Enable administrator
Clear check box to disable the administrator. No settings will be lost.
[Save full permissions report](#)

Back Finish Cancel

Configure Additional Desktop Controller

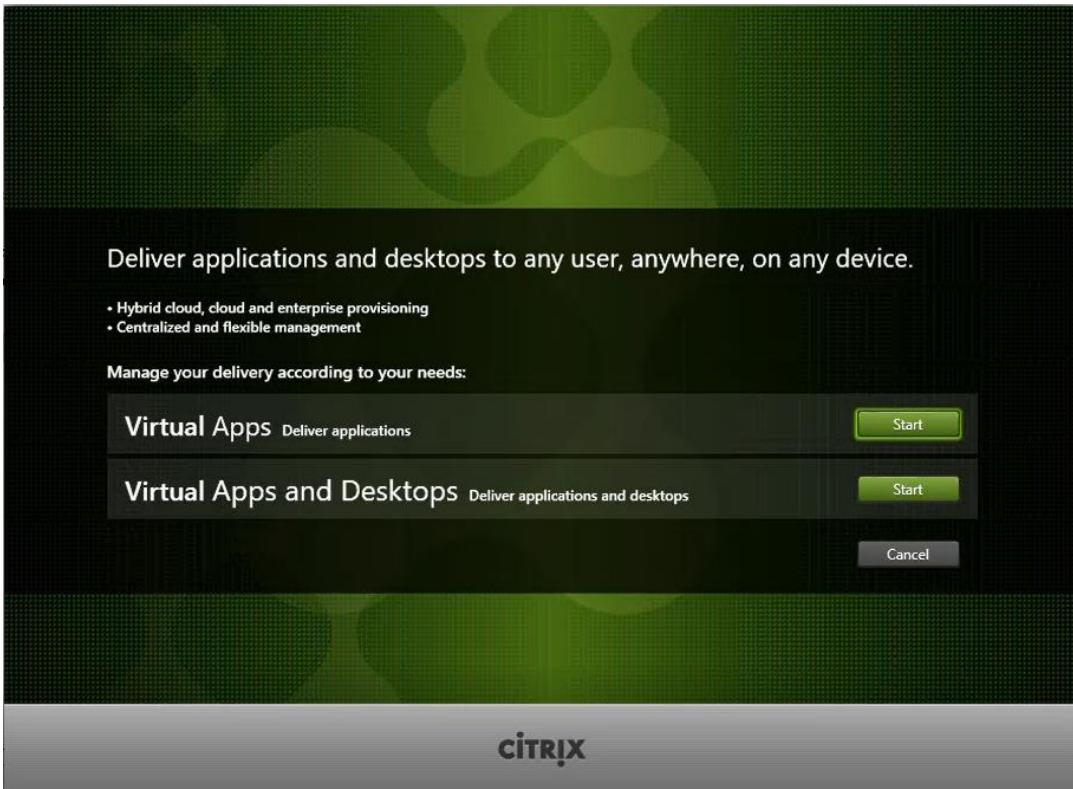
After the first controller is completely configured and the Site is operational, you can add additional controllers.



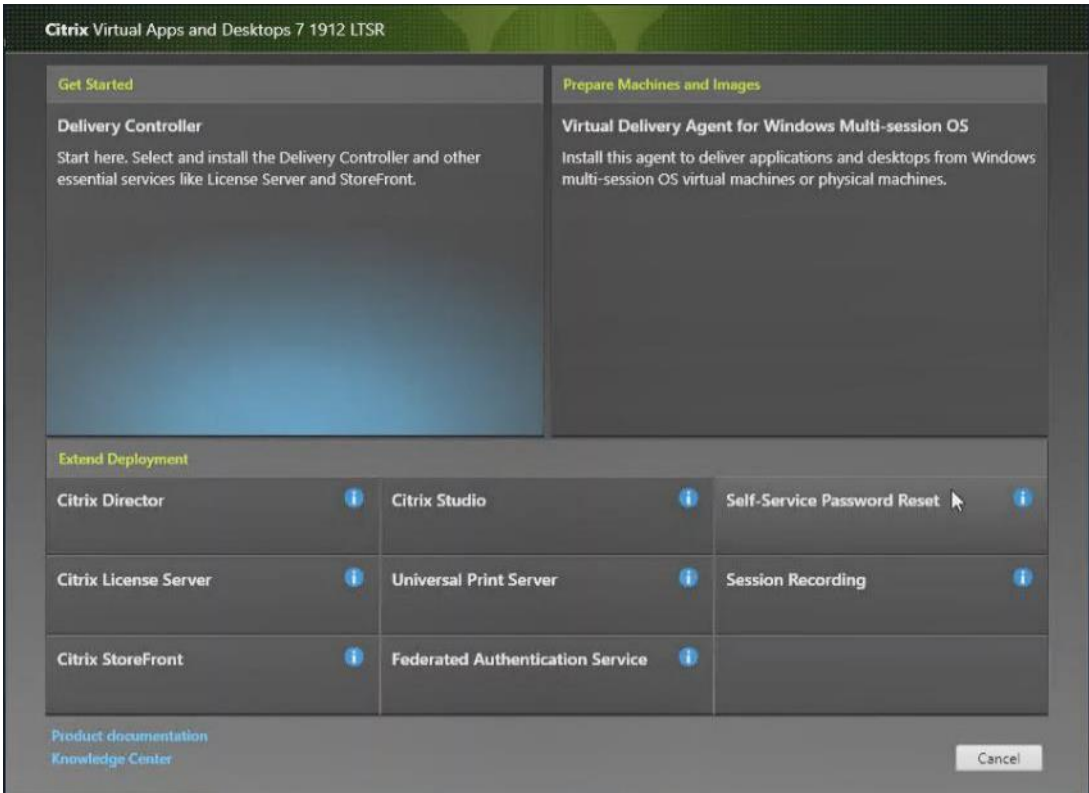
In this CVD, we created two Delivery Controllers.

To configure additional Citrix Desktop controllers, follow these steps:

1. To begin the installation of the second Delivery Controller, connect to the second Citrix VDI server and launch the installer from the Citrix Virtual Apps and Desktops ISO.
2. Click Start.



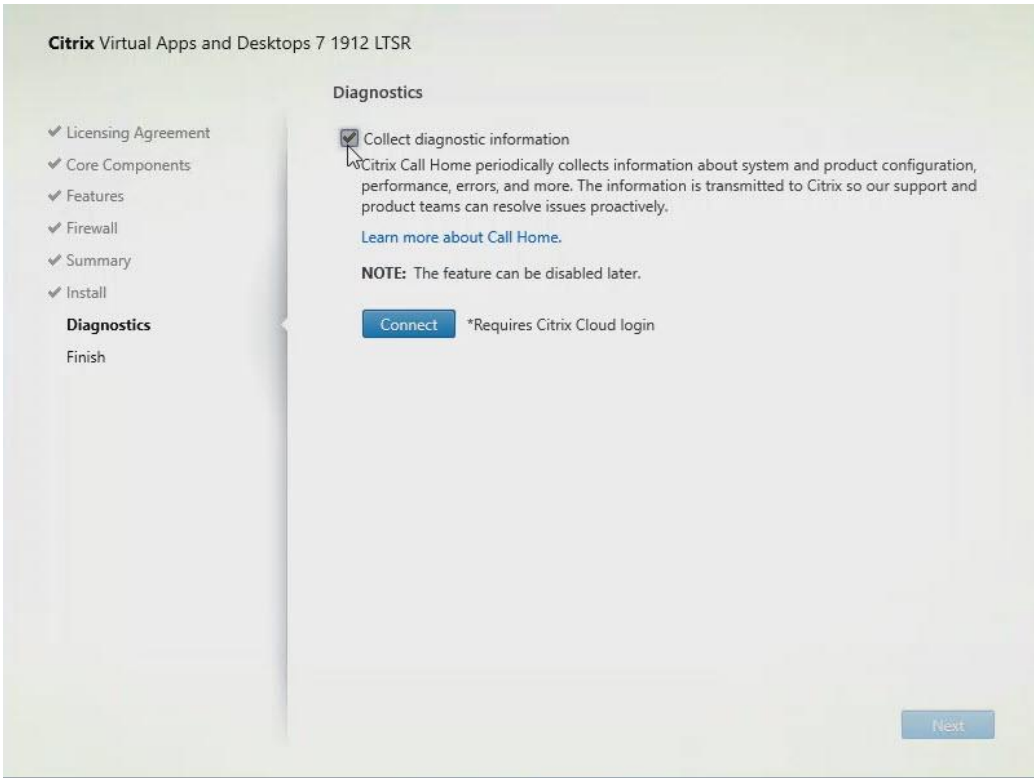
3. Click Delivery Controller.



- Repeat these steps used to install the first Delivery Controller, including the step of importing an SSL certificate for HTTPS between the controller and Hyper-V.
- Review the Summary configuration.
- Click Install.

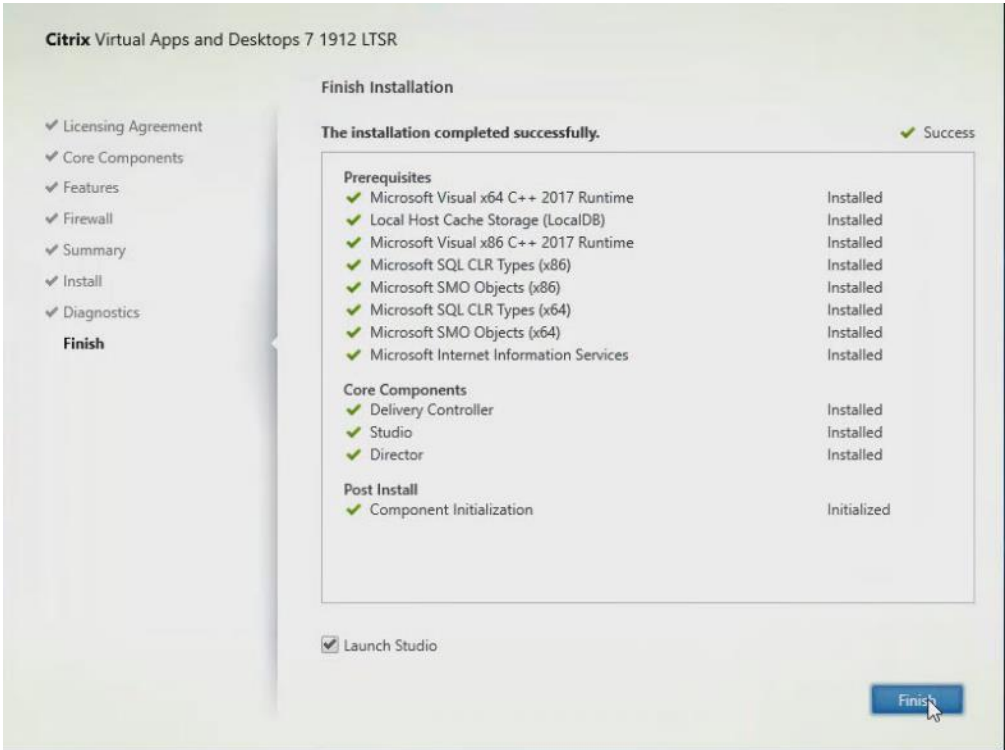


- (Optional) Click "Collect diagnostic information."
- Click Next.



9. Verify the components installed successfully.

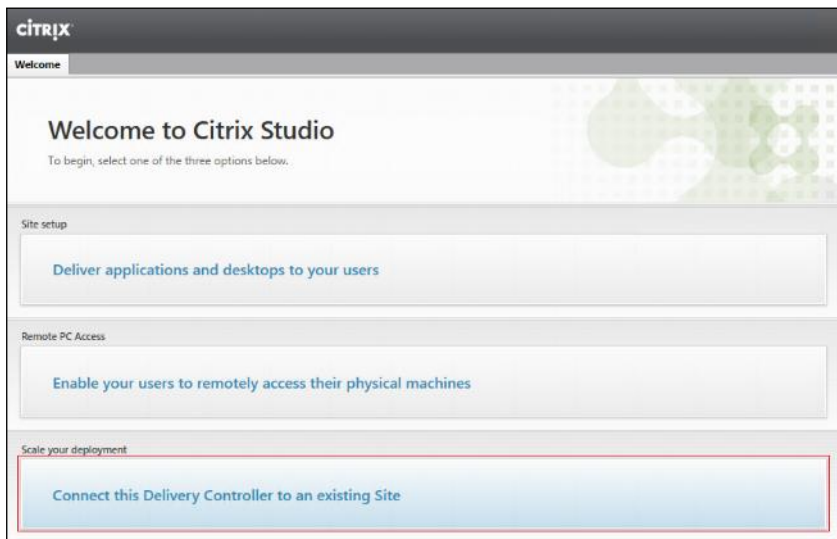
10. Click Finish.



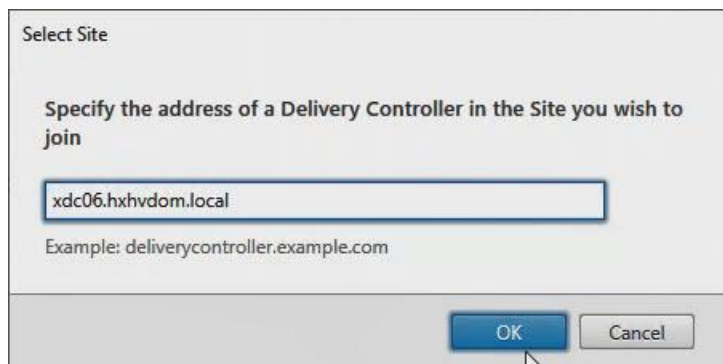
Add the Second Delivery Controller to the Citrix Desktop Site

To add the second Delivery Controller to the Citrix Desktop Site, follow these steps:

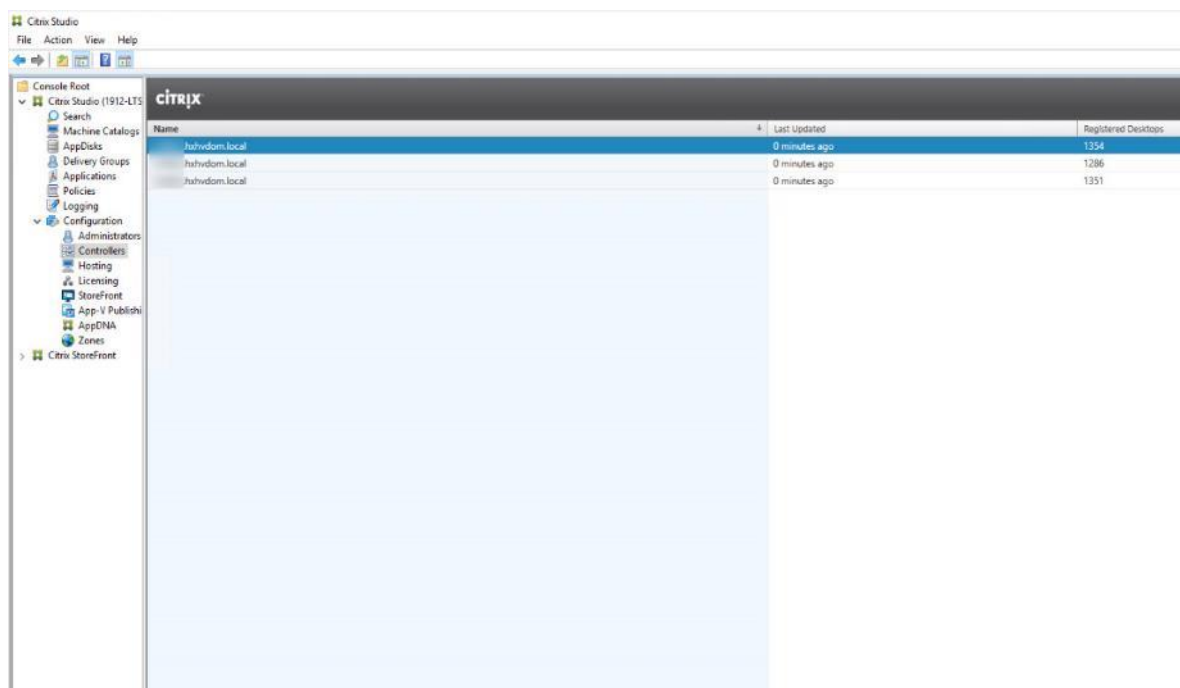
1. In Desktop Studio click Connect this Delivery Controller to an existing Site.



2. Enter the FQDN of the first delivery controller.
3. Click OK.



4. Click Yes to allow the database to be updated with this controller's information automatically.
5. When complete, test the site configuration and verify the Delivery Controller has been added to the list of Controllers.



Install and Configure StoreFront

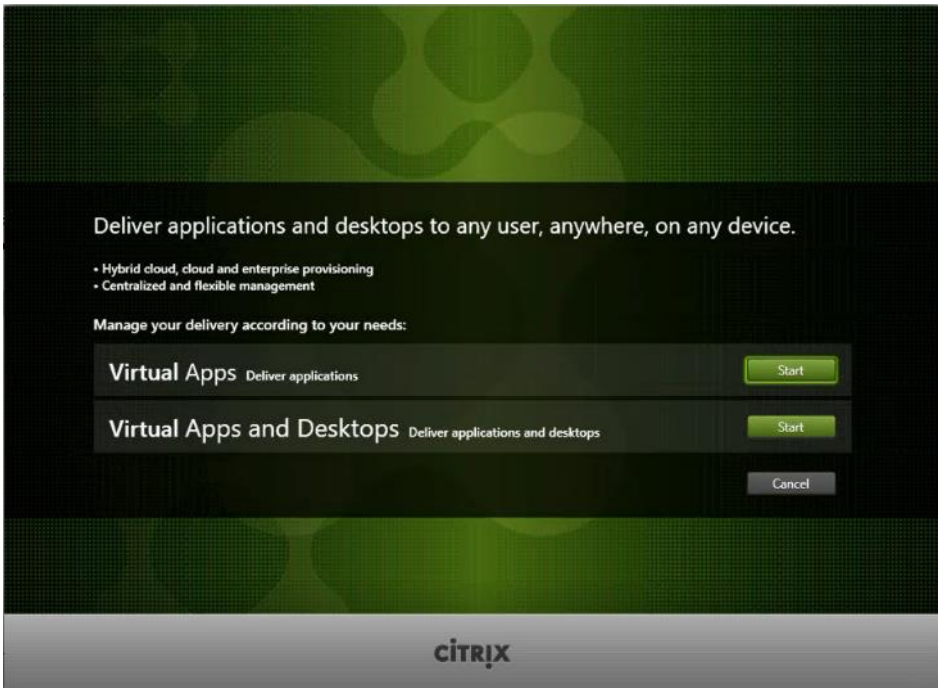
Citrix StoreFront stores aggregate desktops and applications from Citrix VDI sites, making resources readily available to users.



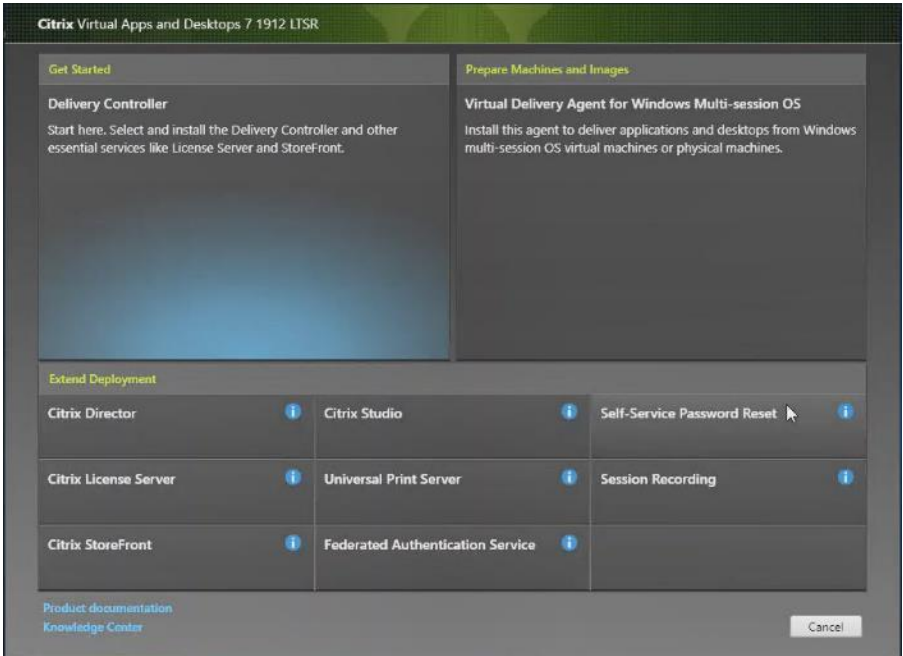
In this CVD, we created two StoreFront servers on dedicated virtual machines.

To install and configure StoreFront, follow these steps:

1. To begin the installation of the StoreFront, connect to the first StoreFront server and launch the installer from the Citrix Desktop 1912 LTSR ISO.
2. Click Start.

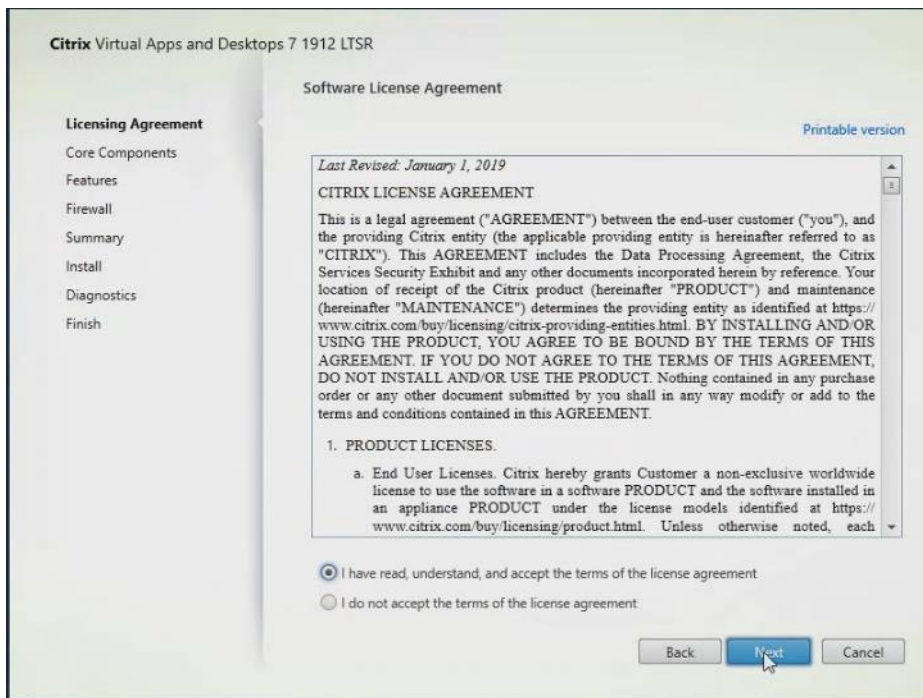


3. Click Extend Deployment Citrix StoreFront.

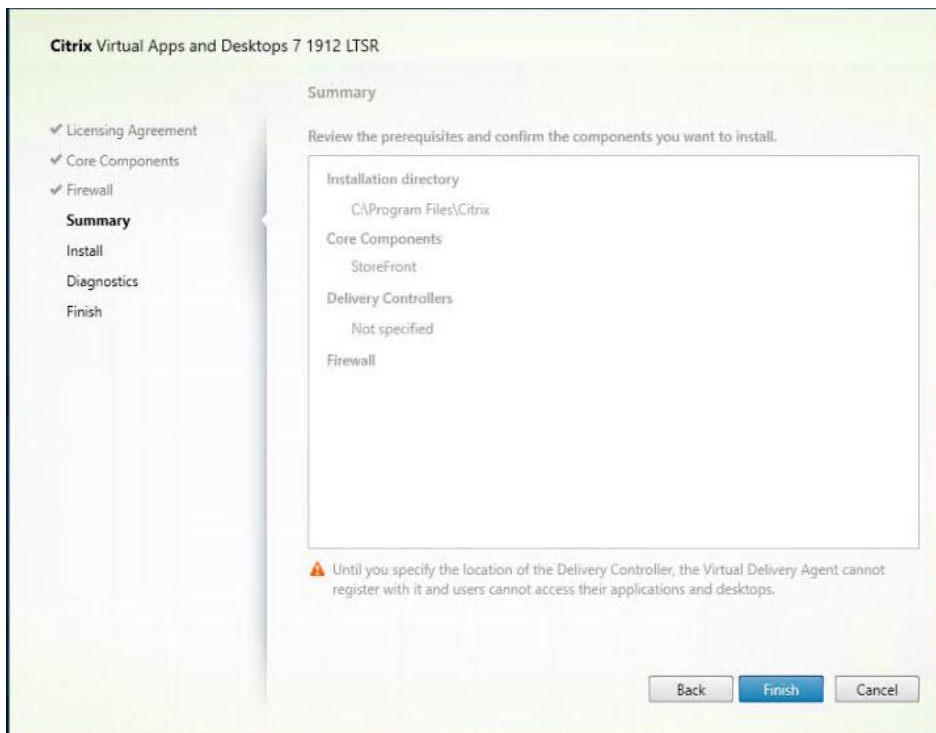


4. If acceptable, indicate your acceptance of the license by selecting the “I have read, understand, and accept the terms of the license agreement” radio button.

5. Click Next.



6. Select Storefront and click Next.



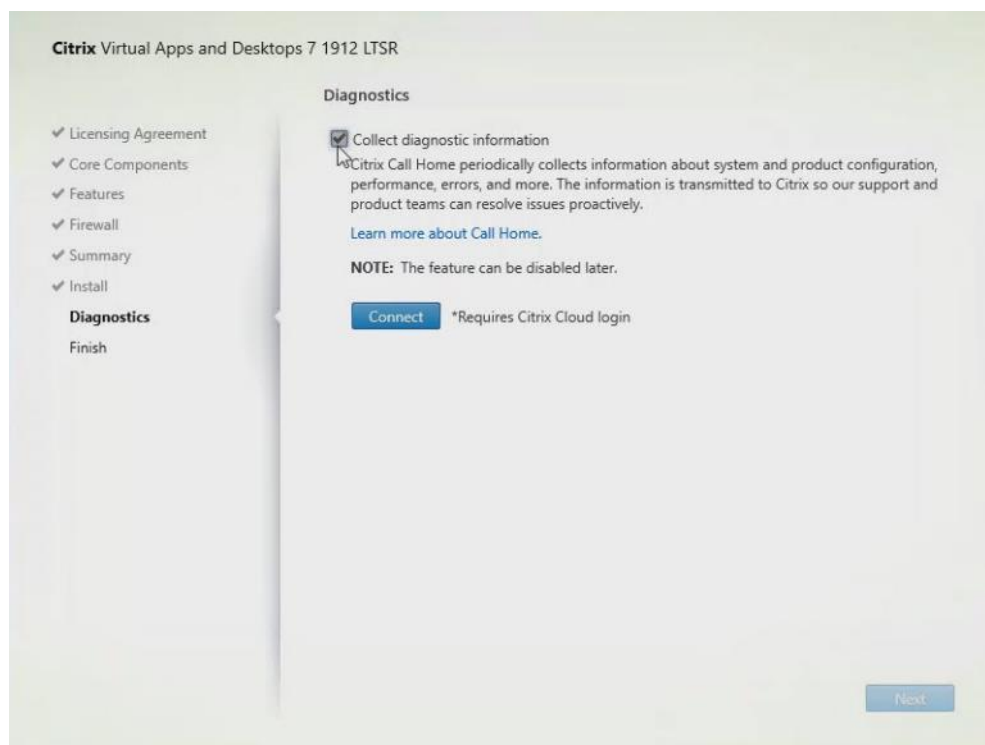
7. Select the default ports and automatically configured firewall rules.

8. Click Next.

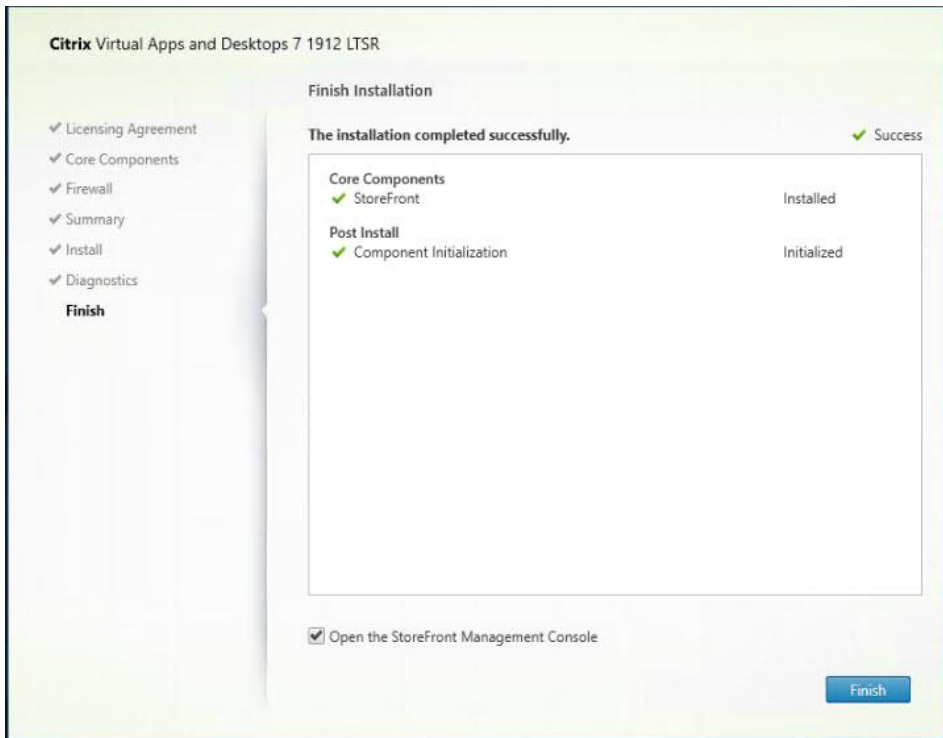
9. Click Install.

10. (Optional) Click Collect diagnostic information.

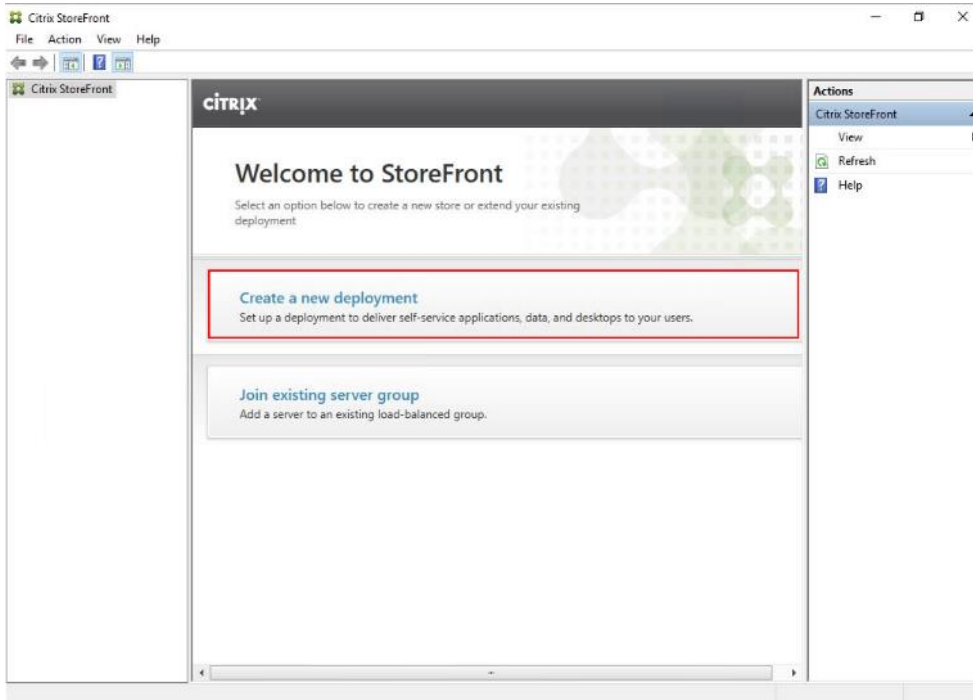
11. Click Next.



12. Click Finish.



13. Click Create a new deployment.



14. Specify the URL of the StoreFront server and click Next.



For a multiple server deployment use the load balancing environment in the Base URL box.

Create New Deployment

StoreFront

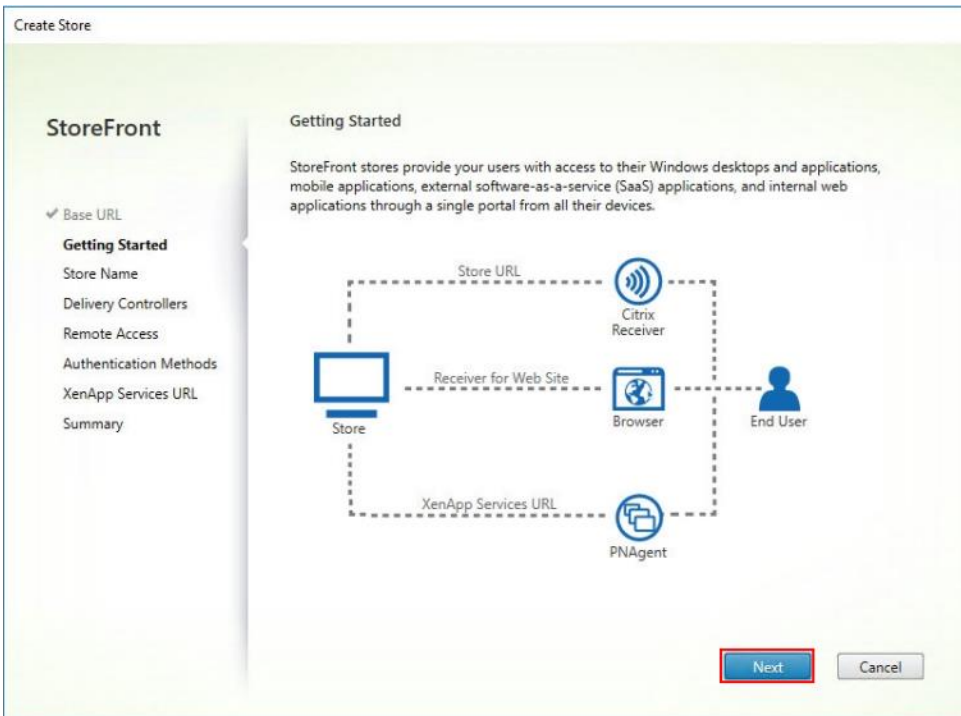
- Base URL
- Getting Started
- Store Name
- Delivery Controllers
- Remote Access
- Authentication Methods
- XenApp Services URL
- Summary

Enter a Base URL

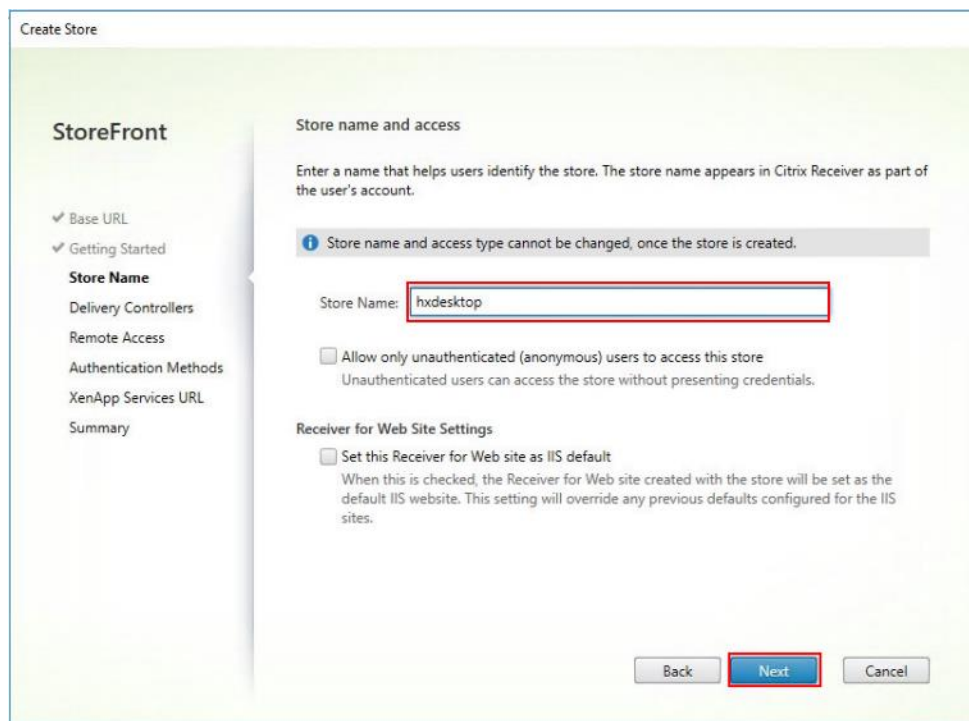
Confirm the base URL for services hosted on this deployment. For multiple server deployments, specify the load-balanced URL for the server group.

Base URL:

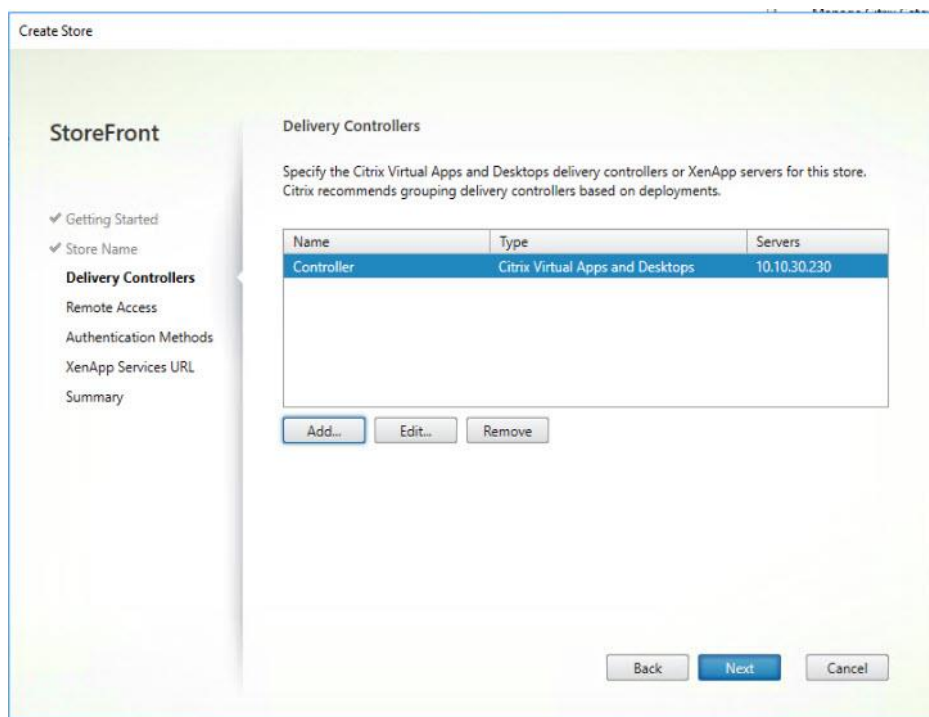
15. Click Next.



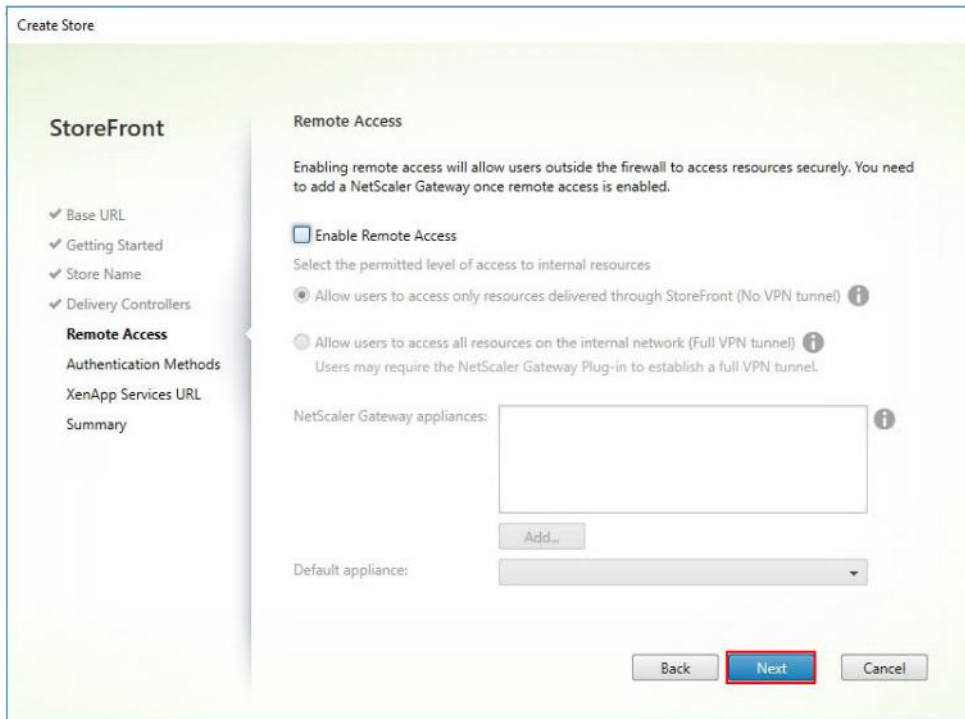
16. Specify a name for your store and click Next.



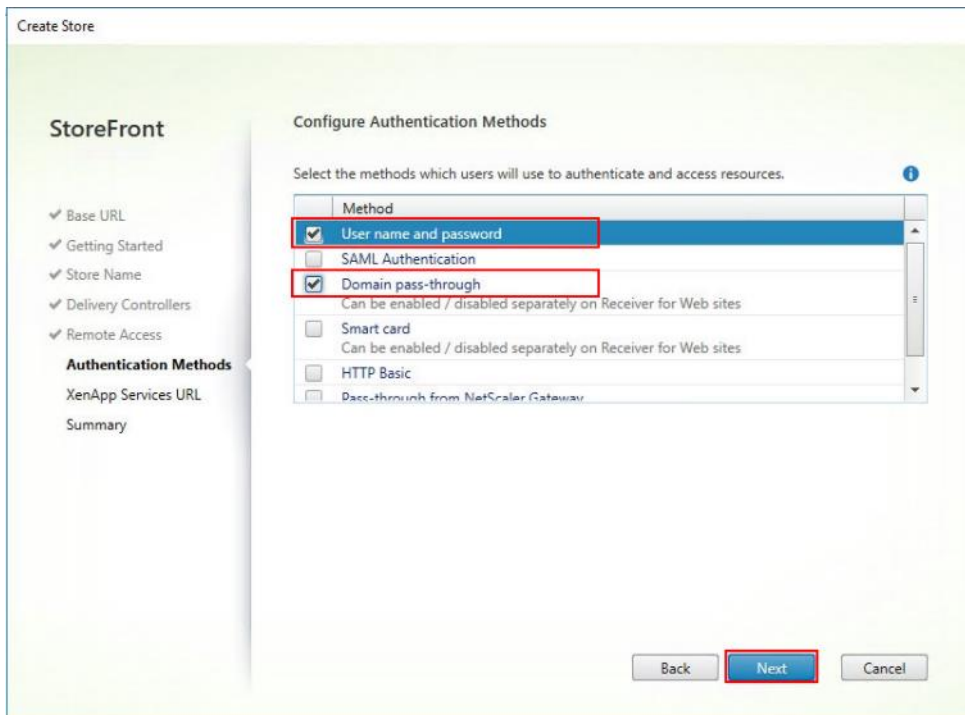
17. Add the required Delivery Controllers to the store and click Next.



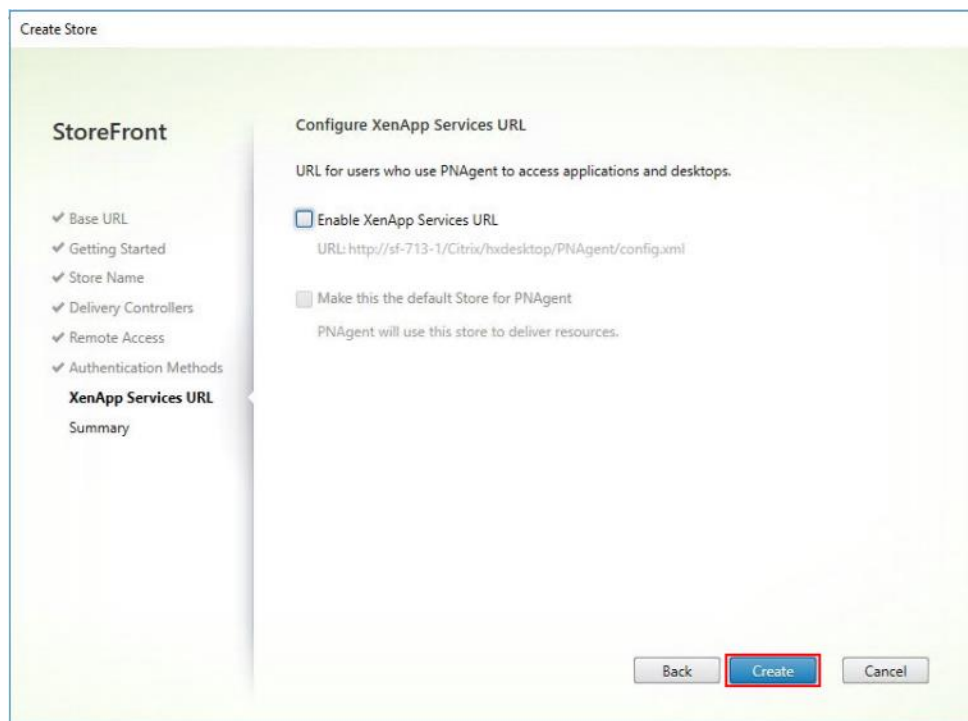
18. Specify how connecting users can access the resources, in this environment only local users on the internal network are able to access the store and click Next.



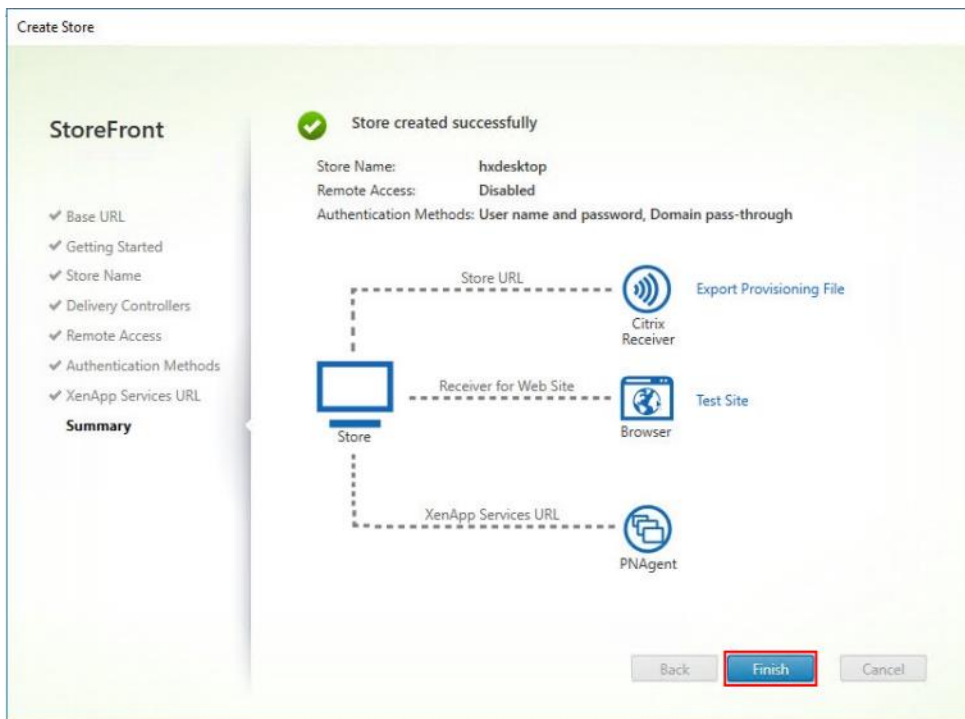
19. On the “Authentication Methods” page, select the methods your users will use to authenticate to the store and click Next. You can select from the following methods as shown below:



20. Username and password: Users enter their credentials and are authenticated when they access their stores.
21. Domain pass-through: Users authenticate to their domain-joined Windows computers and their credentials are used to log them on automatically when they access their stores.
22. Configure the XenApp Service URL for users who use PNAgent to access the applications and desktops and click Create.



23. After creating the store click Finish.

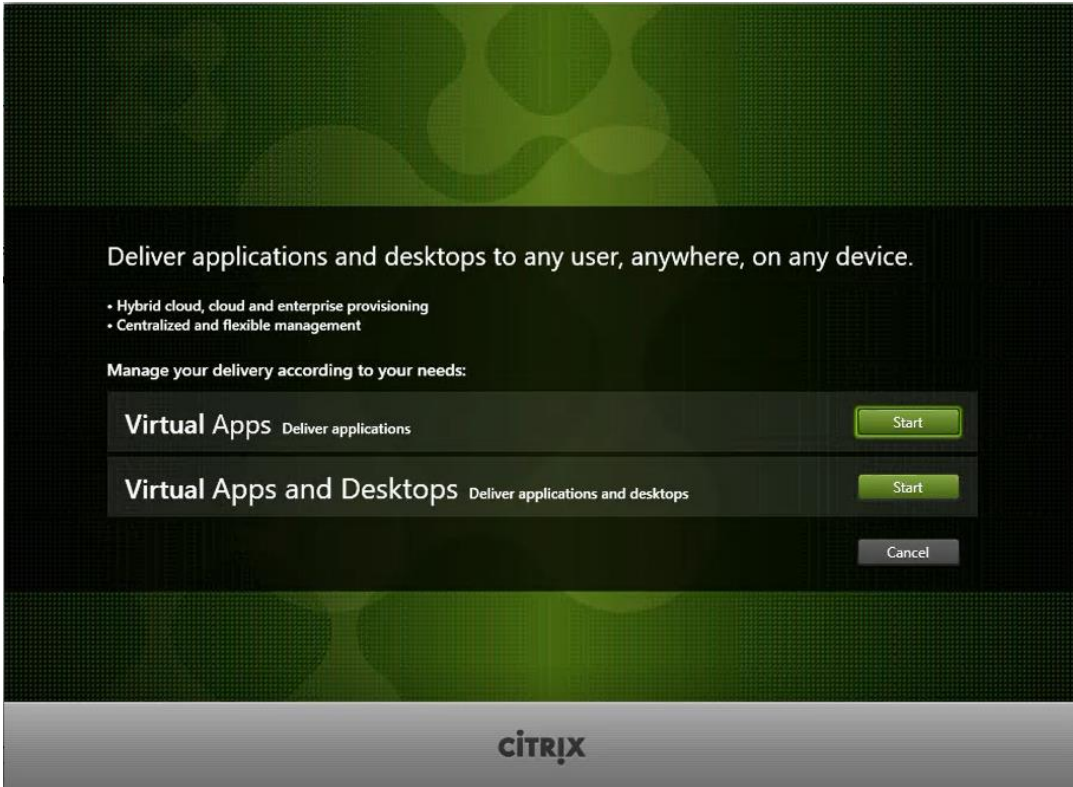


Additional StoreFront Configuration

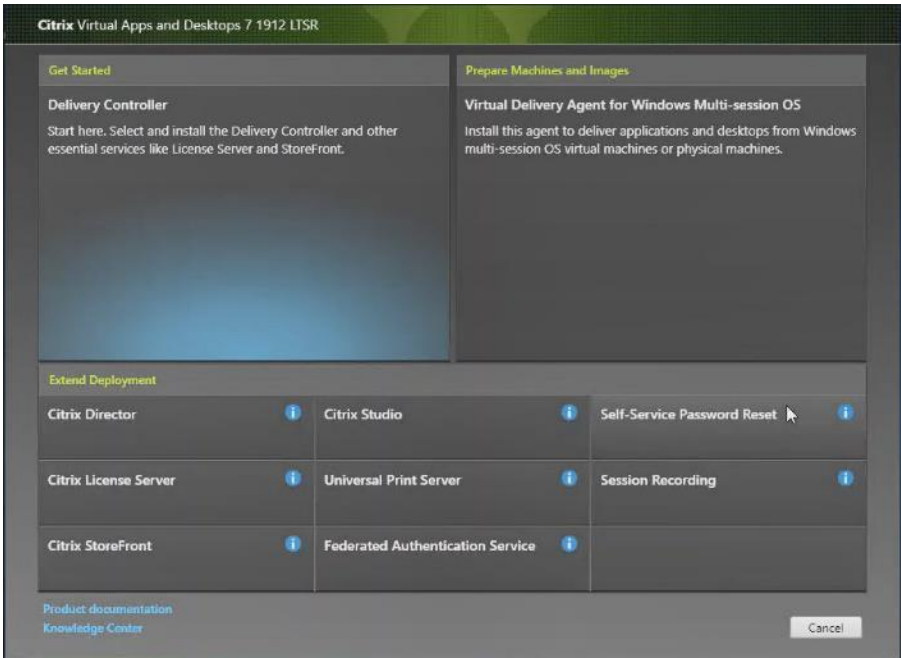
After the first StoreFront server is completely configured and the Store is operational, you can add additional servers.

To configure additional StoreFront server, follow these steps:

1. To begin the installation of the second StoreFront, connect to the second StoreFront server and launch the installer from the Citrix VDI ISO.
2. Click Start.



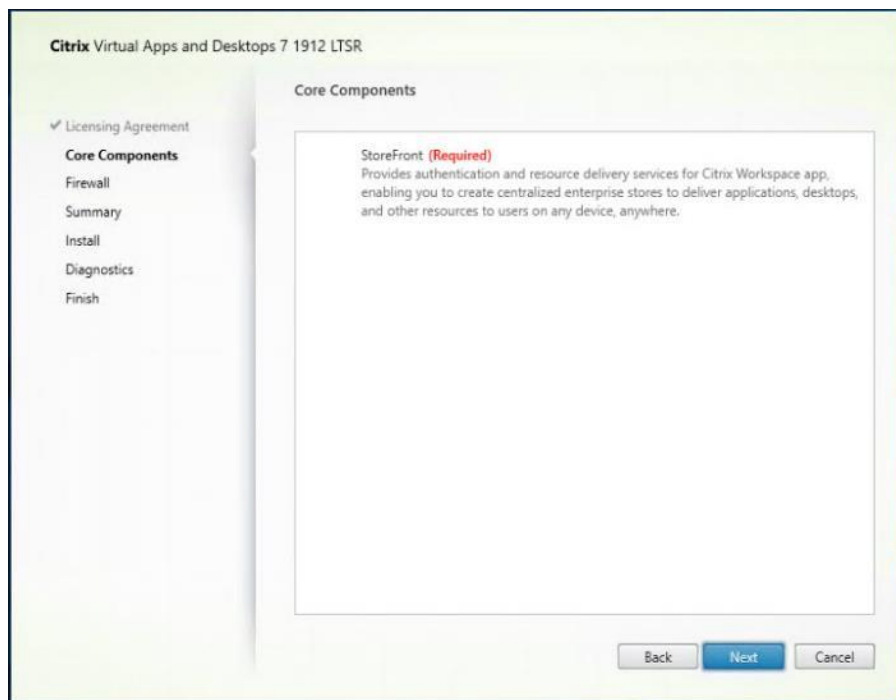
3. Click Extended Deployment Citrix StoreFront.



4. Repeat the same steps used to install the first StoreFront.

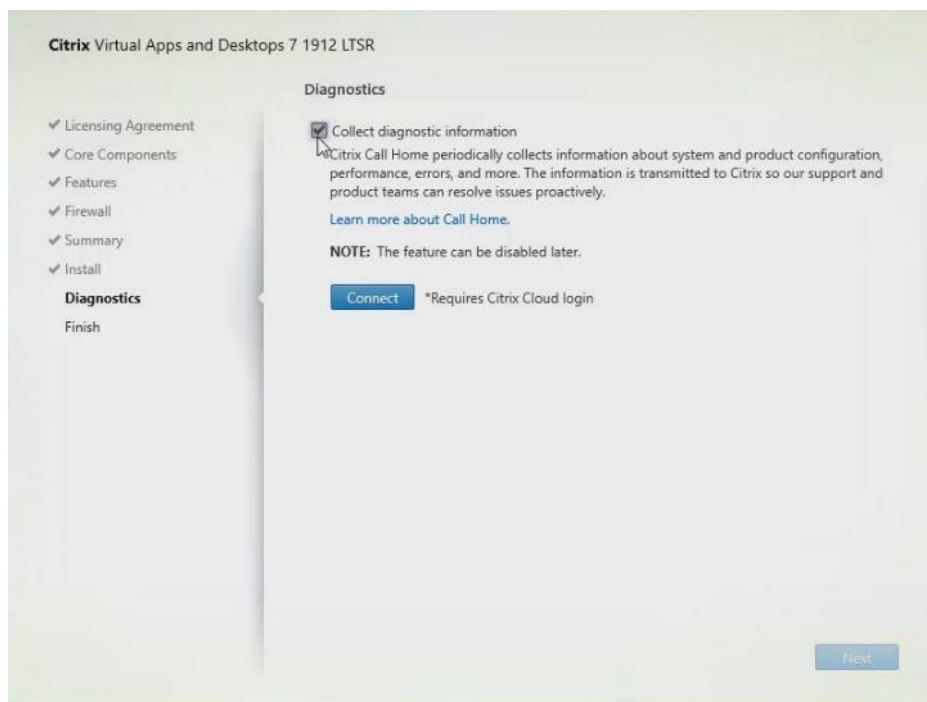
5. Review the Summary configuration.

6. Click Install.



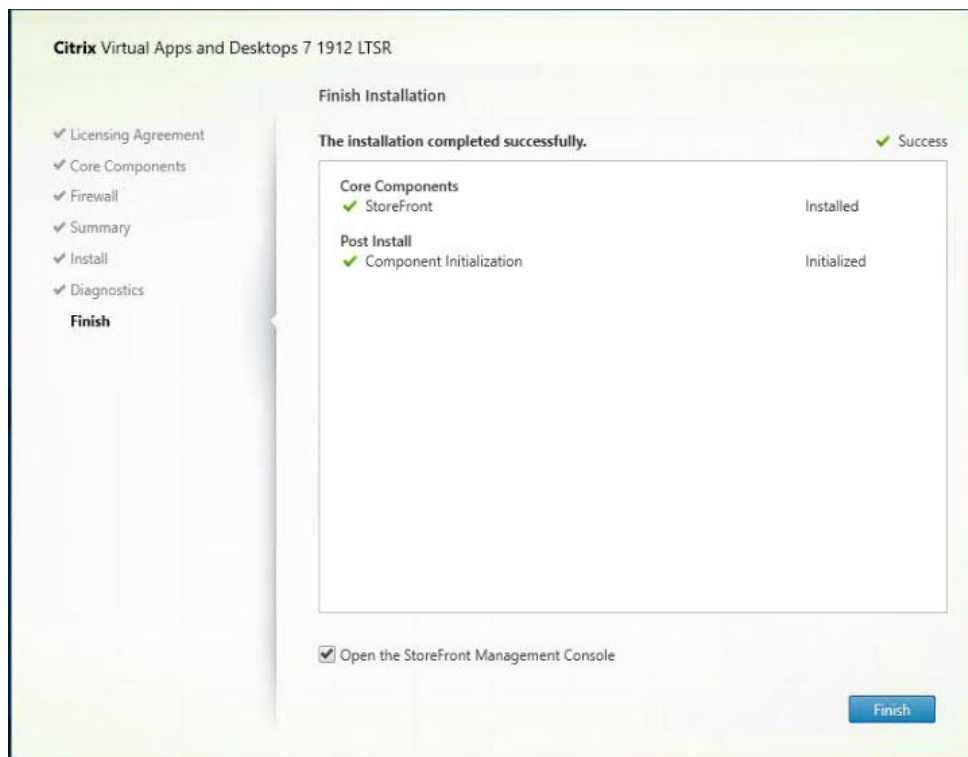
7. (Optional) Click Collect diagnostic information.

8. Click Next.



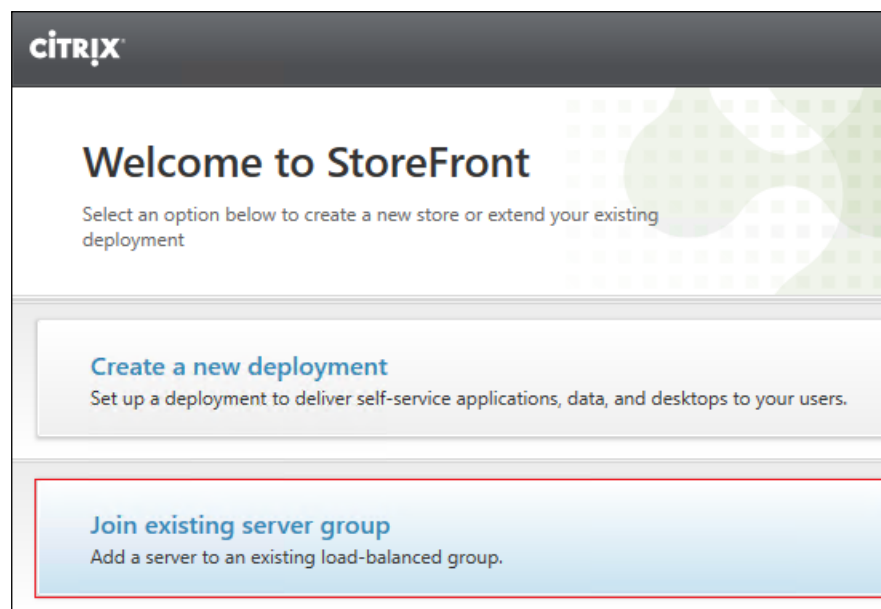
9. Check the box for Open the StoreFront Management Console.

10. Click Finish.



To configure the second StoreFront if used, follow these steps:

1. From the StoreFront Console on the second server click Join existing server group.



2. In the Join Server Group dialog, enter the name of the first Storefront server.

Join Server Group

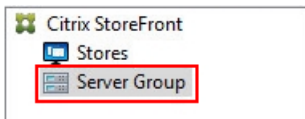
To authorize this server, first connect to a server in the group and choose "Add Server". Enter the provided authorization information here.

Authorizing server: CTX-VDI

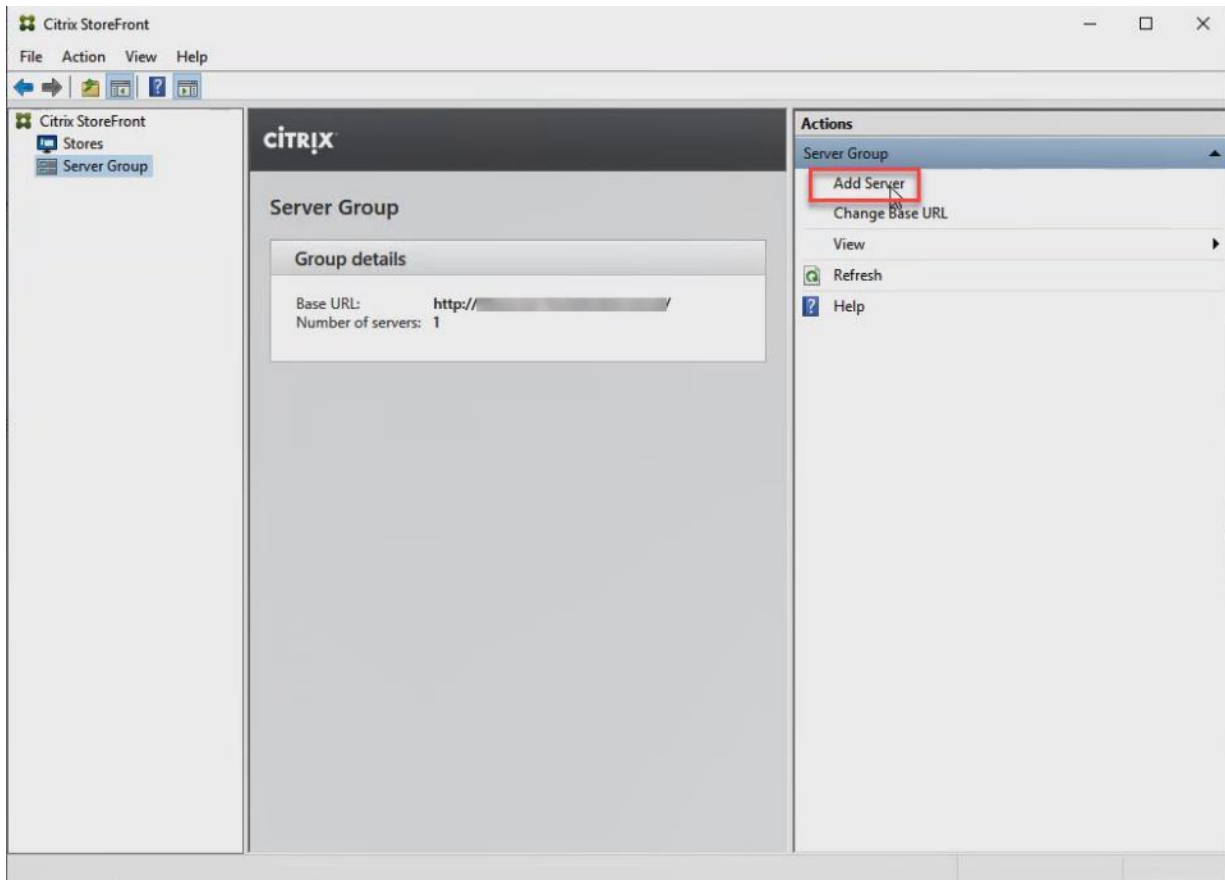
Authorization code:

Join Cancel

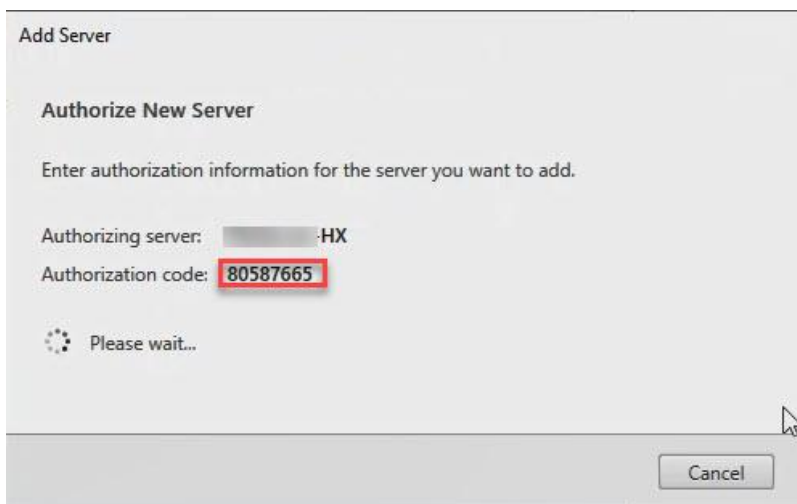
3. Before the additional StoreFront server can join the server group, you must connect to the first Storefront server, add the second server, and obtain the required authorization information.
4. Connect to the first StoreFront server.
5. Using the StoreFront menu on the left, you can scroll through the StoreFront management options.
6. Select Server Group from the menu.



7. To add the second server and generate the authorization information that allows the additional StoreFront server to join the server group, select Add Server.

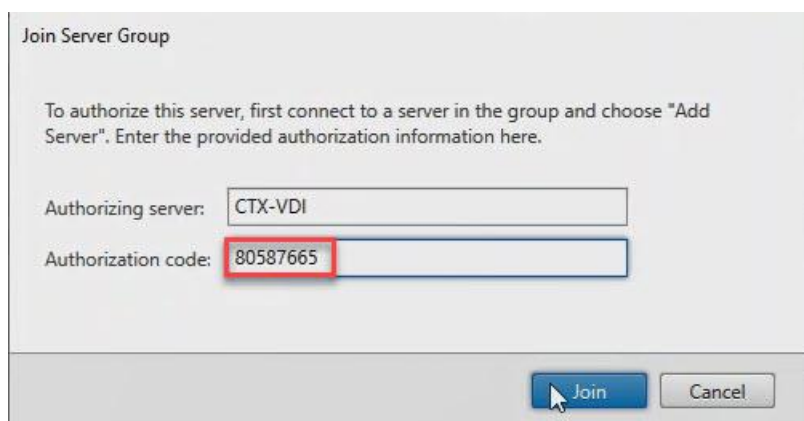


8. Copy the Authorization code from the Add Server dialog.



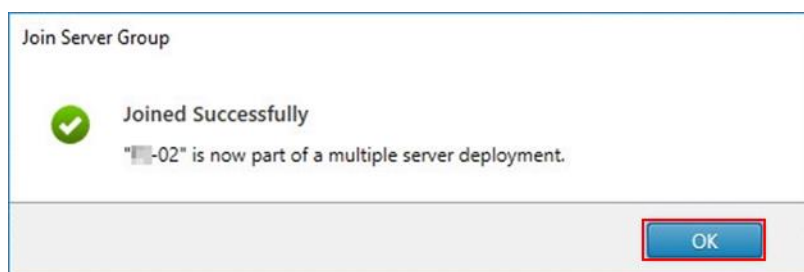
9. Connect to the second Storefront server and paste the Authorization code into the Join Server Group dialog.

10. Click Join.



11. A message appears when the second server has joined successfully.

12. Click OK.



The second StoreFront is now in the Server Group.

Install the Citrix Provisioning Services Target Device Software

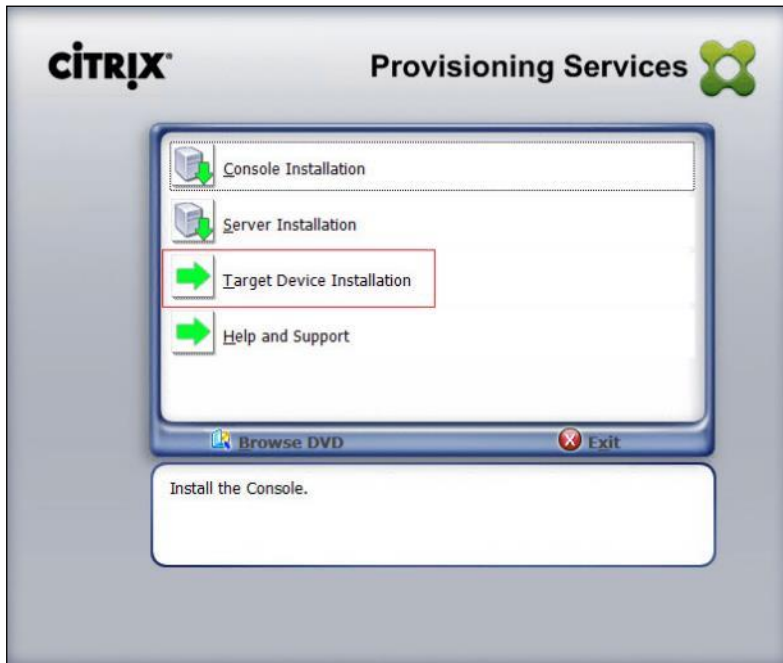
For non-persistent Windows 10 virtual desktops and Server 2019 RDS virtual machines, Citrix Provisioning Services (PVS) is used for deployment. The Master Target Device refers to the target device from which a hard disk image is built and stored on a vDisk. Provisioning Services then streams the contents of the vDisk created to other target devices. This procedure installs the PVS Target Device software that is used to build the RDS and VDI golden images.

To install the Citrix Provisioning Server Target Device software, follow these steps:



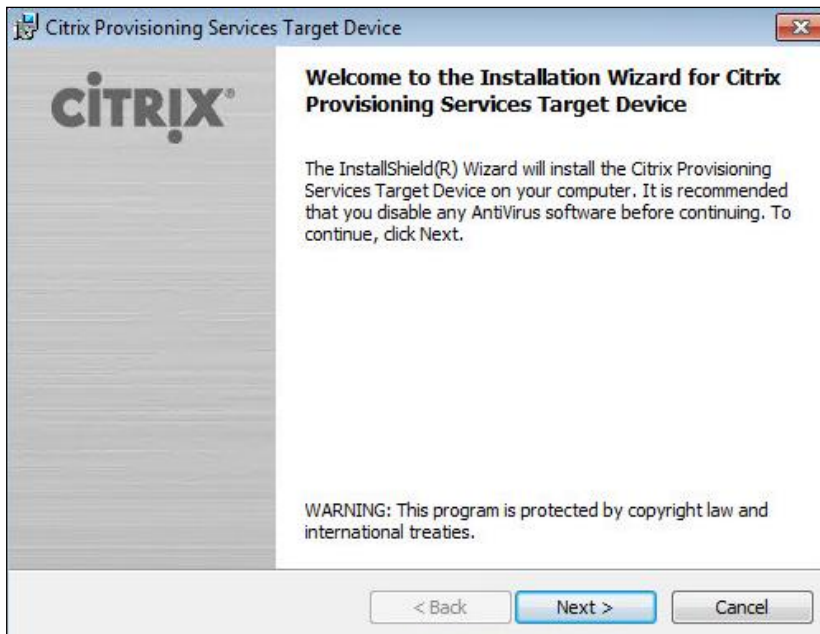
The instructions below outline the installation procedure to configure a vDisk for VDI desktops. When you have completed these installation steps, repeat the procedure to configure a vDisk for RDS.

1. On the Windows 10 Master Target Device, launch the PVS installer from the Provisioning Services ISO.
2. Click Target Device Installation.



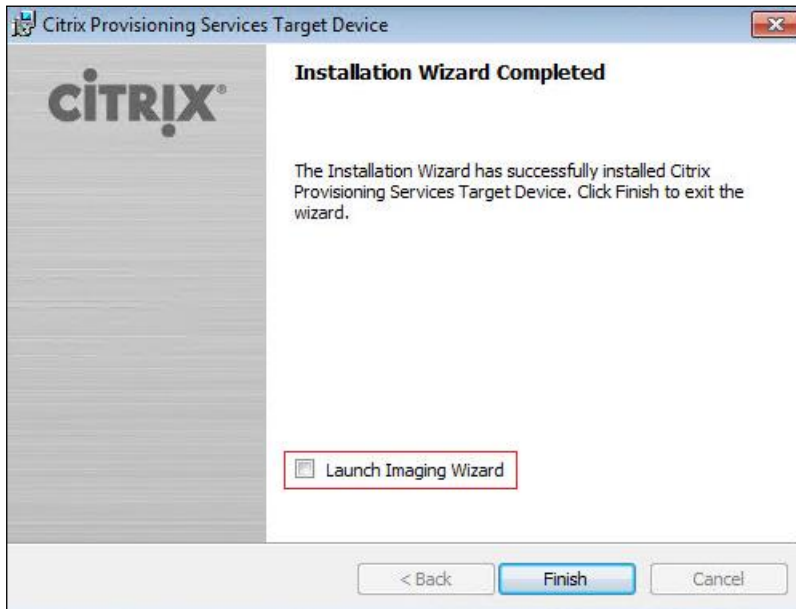
The installation wizard will check to resolve dependencies and then begin the PVS target device installation process.

3. Click Next.



4. Confirm the installation settings and click Install.

5. Deselect the checkbox to launch the Imaging Wizard and click Finish.



6. Reboot the machine.

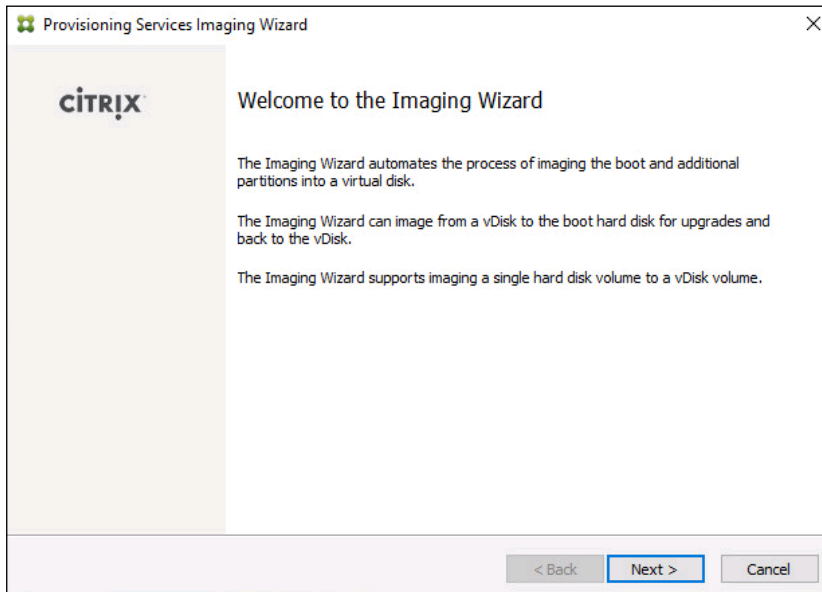
Create Citrix Provisioning Services vDisks

The PVS Imaging Wizard automatically creates a base vDisk image from the master target device. To create the Citrix Provisioning Server vDisks, follow these steps:

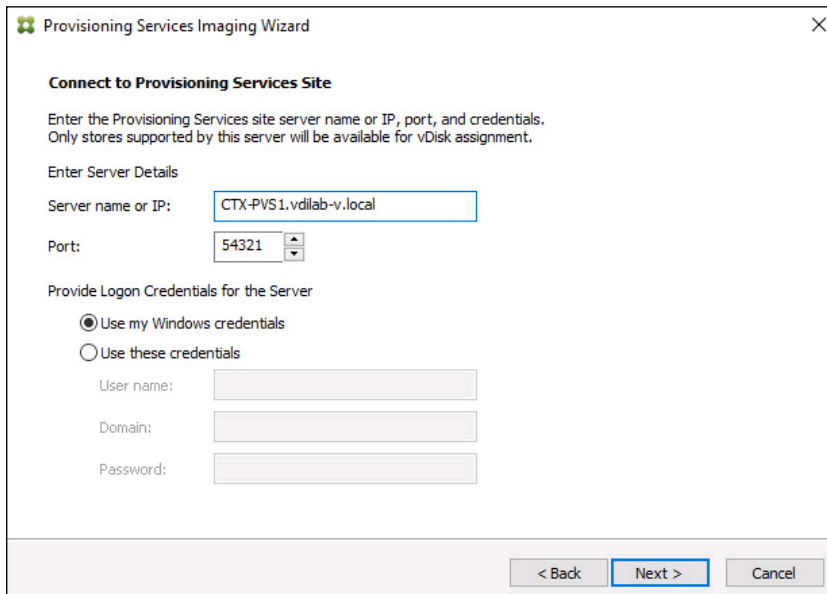


The following procedure explains how to create a vDisk for VDI desktops. When you have completed these steps, repeat the procedure to build a vDisk for RDS.

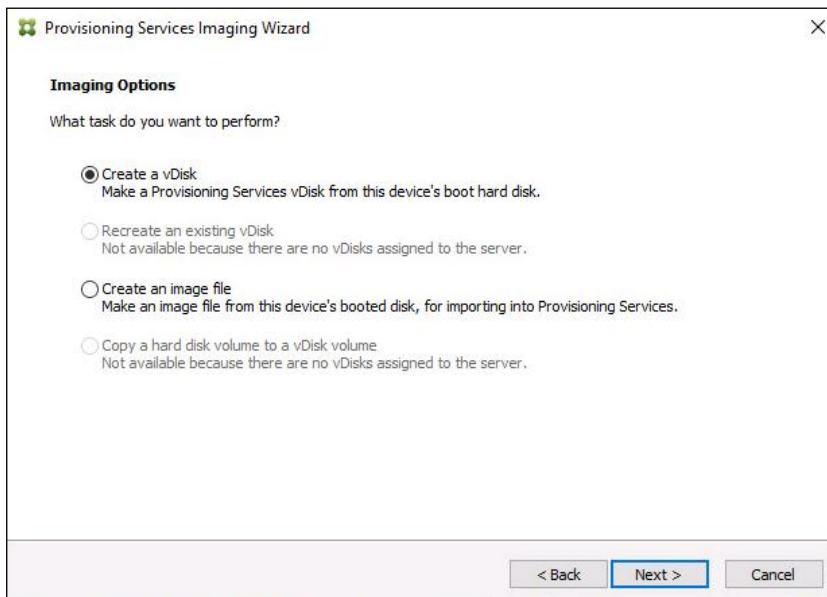
1. The PVS Imaging Wizard's Welcome page appears.
2. Click Next.



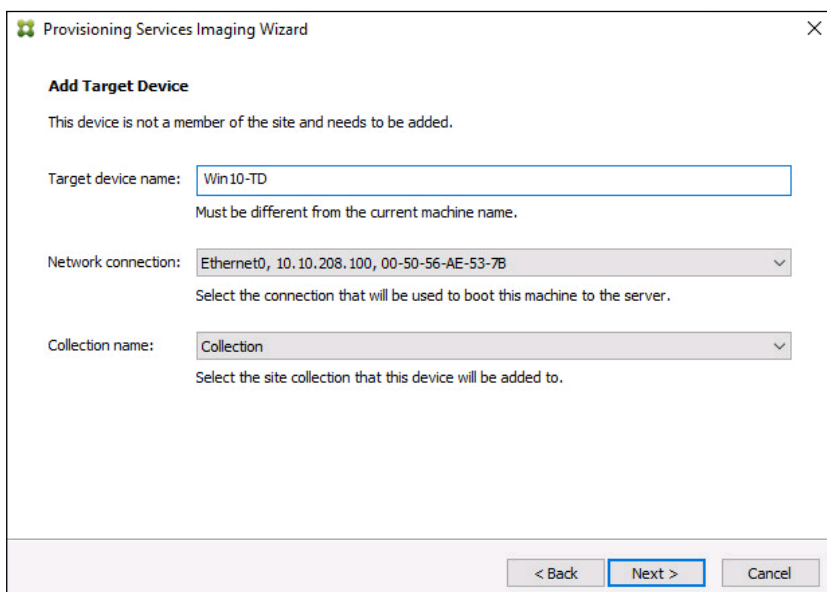
3. The Connect to Farm page appears. Enter the name or IP address of a Provisioning Server within the farm to connect to and the port to use to make that connection.
4. Use the Windows credentials (default) or enter different credentials.
5. Click Next.



6. Select Create new vDisk.
7. Click Next.

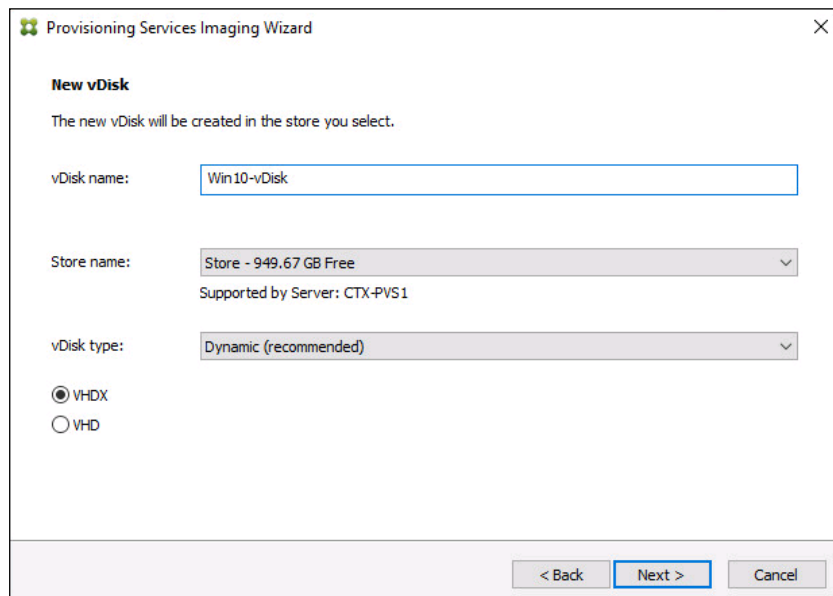


8. The Add Target Device page appears.
9. Select the Target Device Name, the MAC address associated with one of the NICs that was selected when the target device software was installed on the master target device, and the Collection to which you are adding the device.
10. Click Next.



11. The New vDisk dialog displays. Enter the name of the vDisk.
12. Select the Store where the vDisk will reside. Select the vDisk type, either Fixed or Dynamic, from the drop-down list. (This CVD used Dynamic rather than Fixed vDisks.)

13. Click Next.



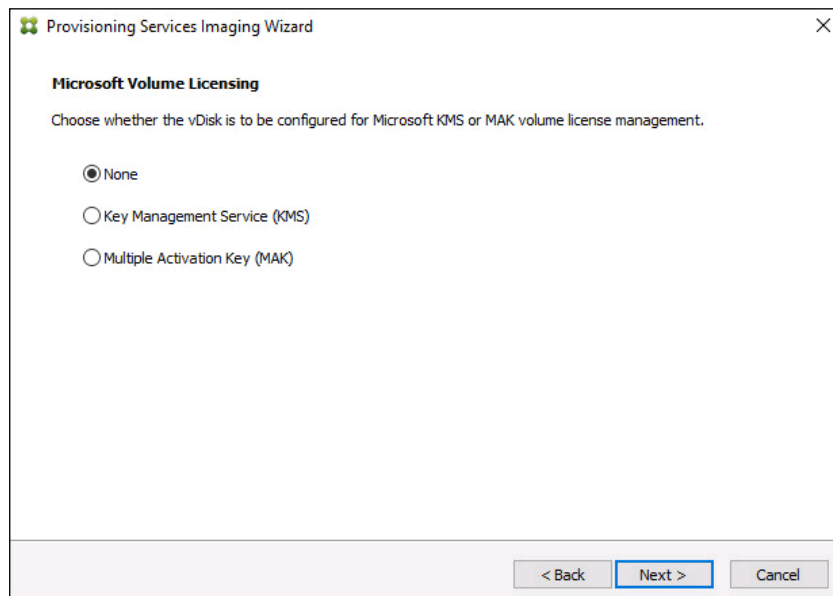
The screenshot shows the 'Provisioning Services Imaging Wizard' window. The title bar reads 'Provisioning Services Imaging Wizard'. The main heading is 'New vDisk'. Below it, a message states: 'The new vDisk will be created in the store you select.' The form contains the following fields and options:

- vDisk name:** A text input field containing 'Win10-vDisk'.
- Store name:** A dropdown menu showing 'Store - 949.67 GB Free'. Below it, the text 'Supported by Server: CTX-PVS1' is displayed.
- vDisk type:** A dropdown menu showing 'Dynamic (recommended)'.
- Radio buttons:** Two options are present: 'VHDX' (which is selected with a filled radio button) and 'VHD' (which is unselected).

At the bottom of the window, there are three buttons: '< Back', 'Next >' (highlighted with a blue border), and 'Cancel'.

14. On the Microsoft Volume Licensing page, select the volume license option to use for target devices. For this CVD, volume licensing is not used, so the None button is selected.

15. Click Next.



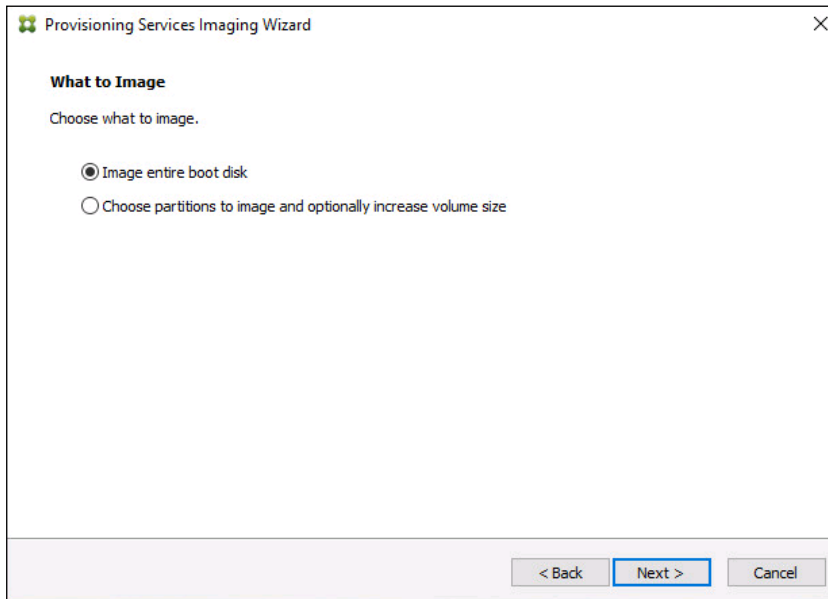
The screenshot shows the 'Provisioning Services Imaging Wizard' window. The title bar reads 'Provisioning Services Imaging Wizard'. The main heading is 'Microsoft Volume Licensing'. Below it, a message states: 'Choose whether the vDisk is to be configured for Microsoft KMS or MAK volume license management.' The form contains the following options:

- Radio buttons:** Three options are present: 'None' (which is selected with a filled radio button), 'Key Management Service (KMS)' (which is unselected), and 'Multiple Activation Key (MAK)' (which is unselected).

At the bottom of the window, there are three buttons: '< Back', 'Next >' (highlighted with a blue border), and 'Cancel'.

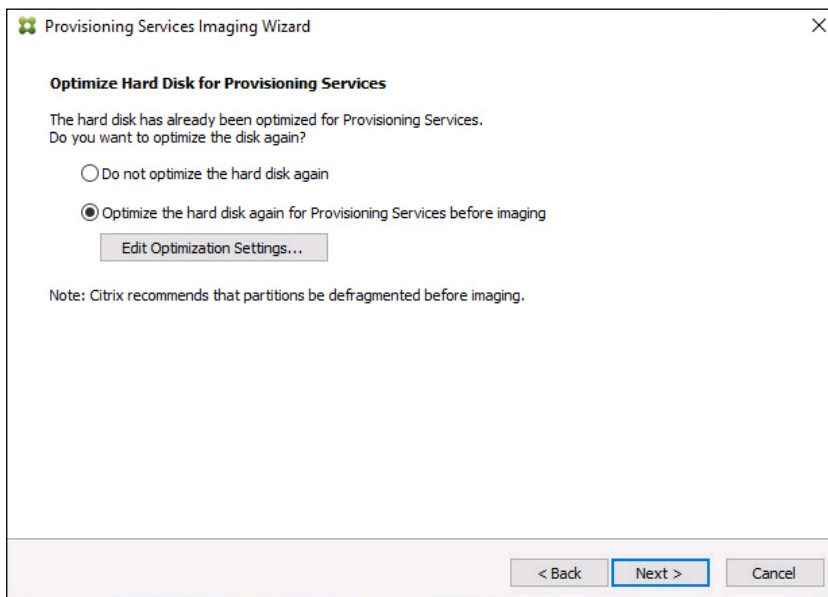
16. Select Image entire boot disk on the Configure Image Volumes page.

17. Click Next.

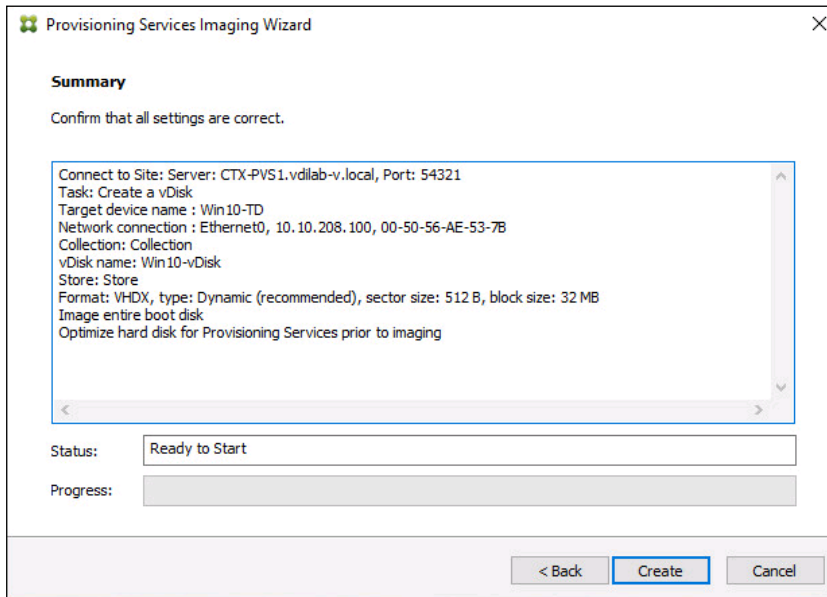


18. Select Optimize for hard disk again for Provisioning Services before imaging on the Optimize Hard Disk for Provisioning Services.

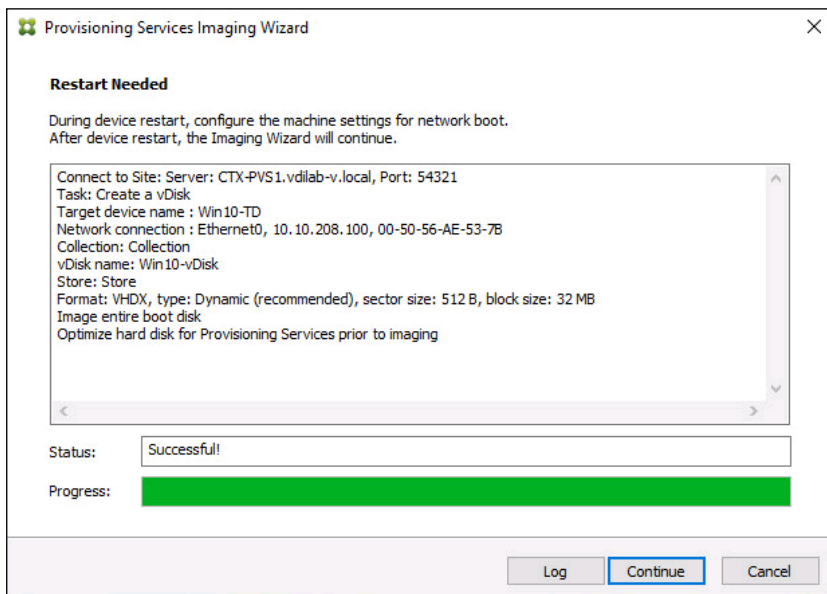
19. Click Next.



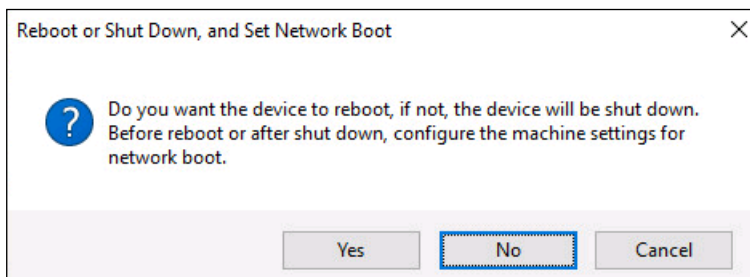
20. Select Create on the Summary page.



21. Review the configuration and click Continue.

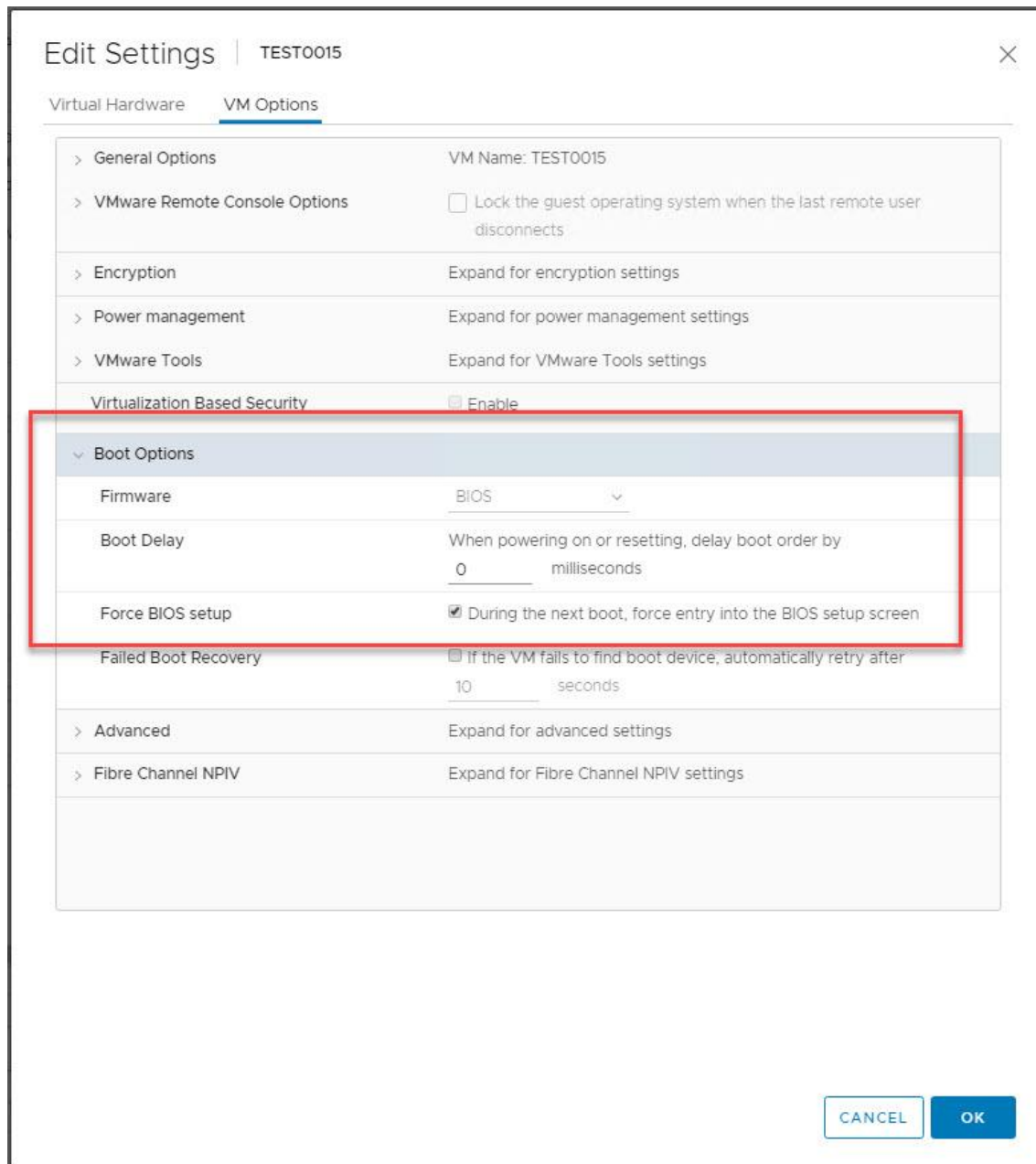


22. When prompted, click No to shut down the machine.



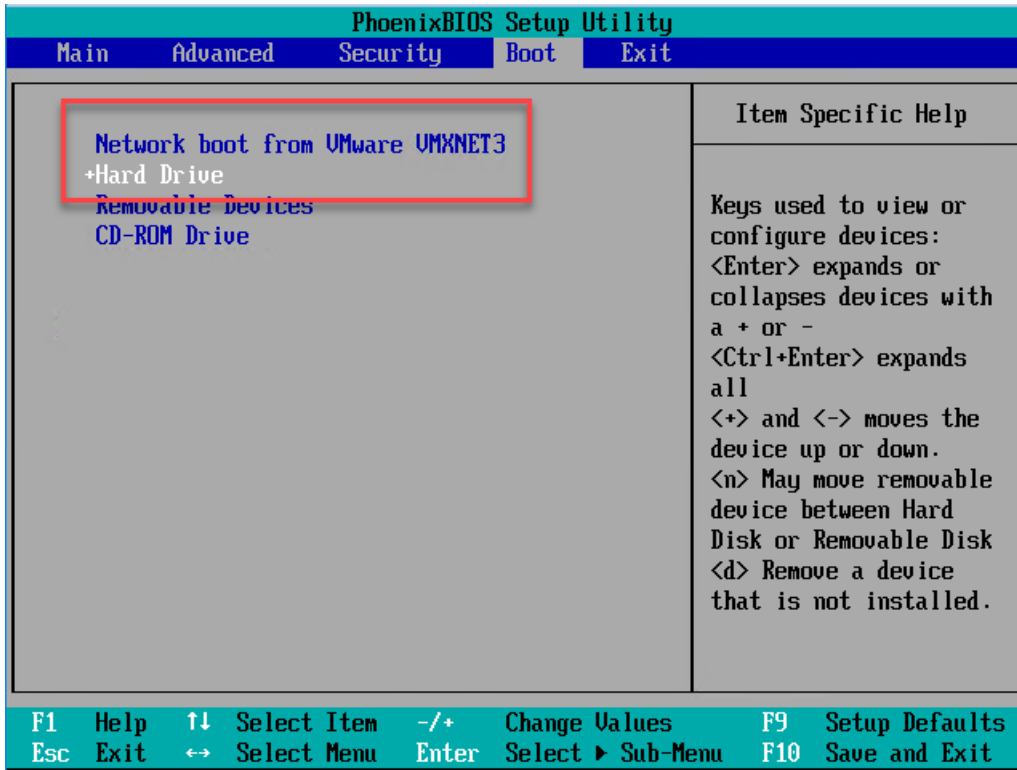
23. Edit the virtual machine settings and select Boot options under VM Options.

24. Select Force BIOS setup.



25. Restart Virtual Machine.

26. When the VM boots into the BIOS, got to Boot menu to move the Network boot from VMware VMXNET3 to the top of the list.

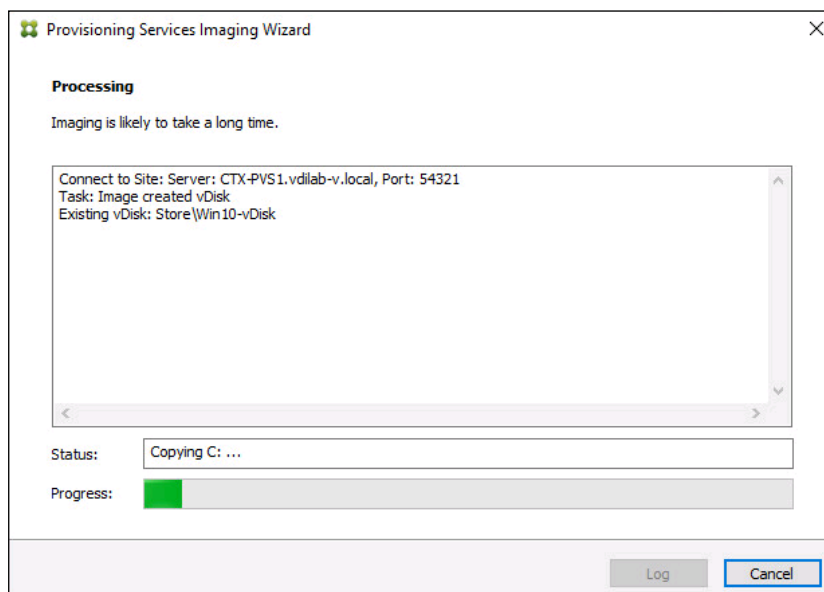


27. Restart Virtual Machine

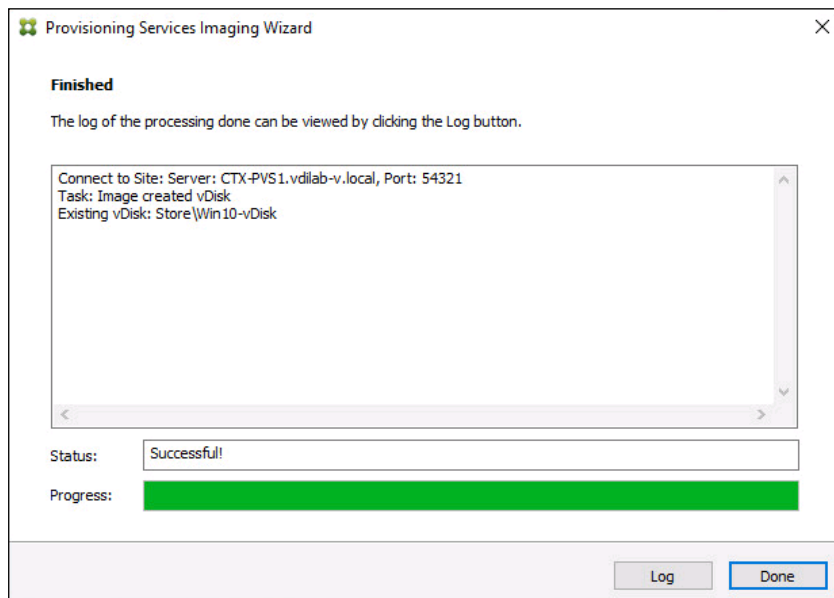


After restarting the virtual machine, log into the VDI or RDS master target. The PVS imaging process begins, copying the contents of the C: drive to the PVS vDisk located on the server.

28. If prompted to Restart select Restart Later.



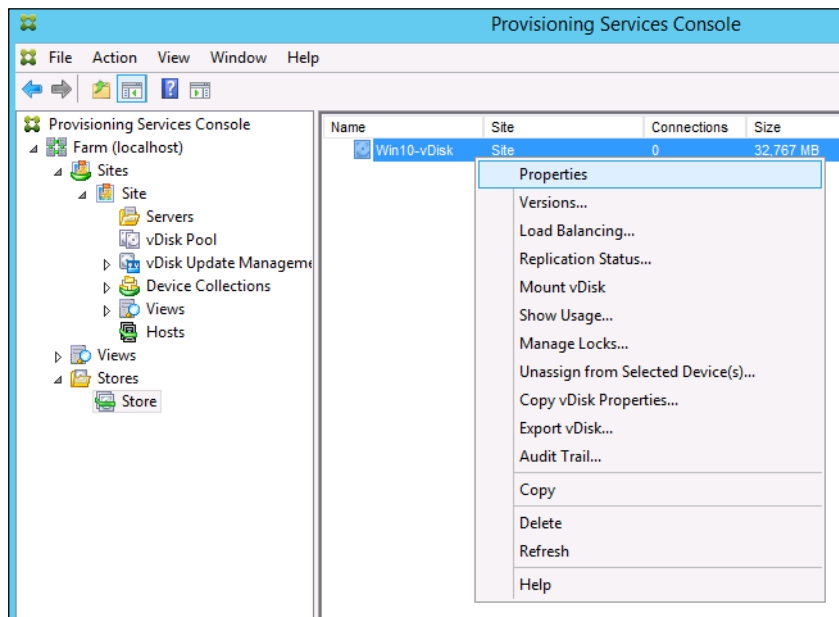
29. A message is displayed when the conversion is complete, click Done.



30. Shutdown the virtual machine used as the VDI or RDS master target.

31. Connect to the PVS server and validate that the vDisk image is available in the Store.

32. Right-click the newly created vDisk and select Properties.



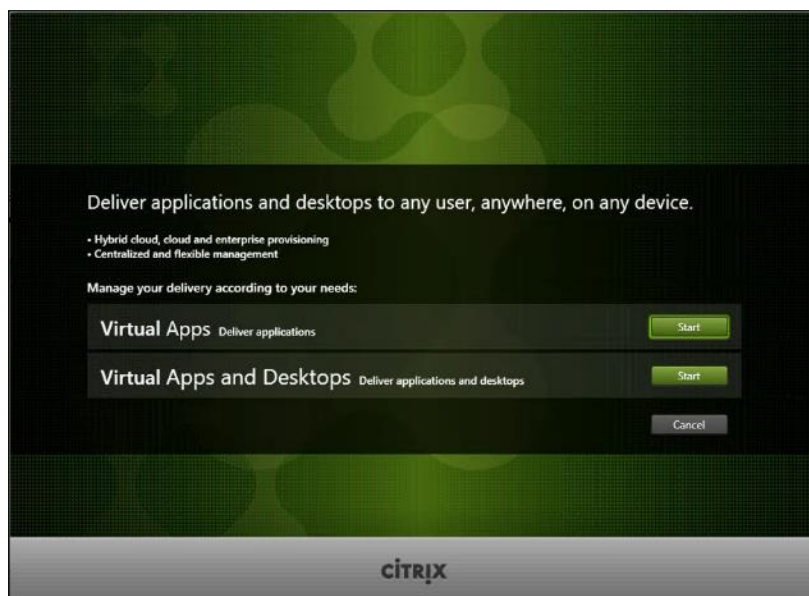
33. On the vDisk Properties dialog, change Access mode to "Private" mode so the Citrix Virtual Desktop Agent can be installed.

Install Citrix Virtual Apps and Desktop Virtual Desktop Agents

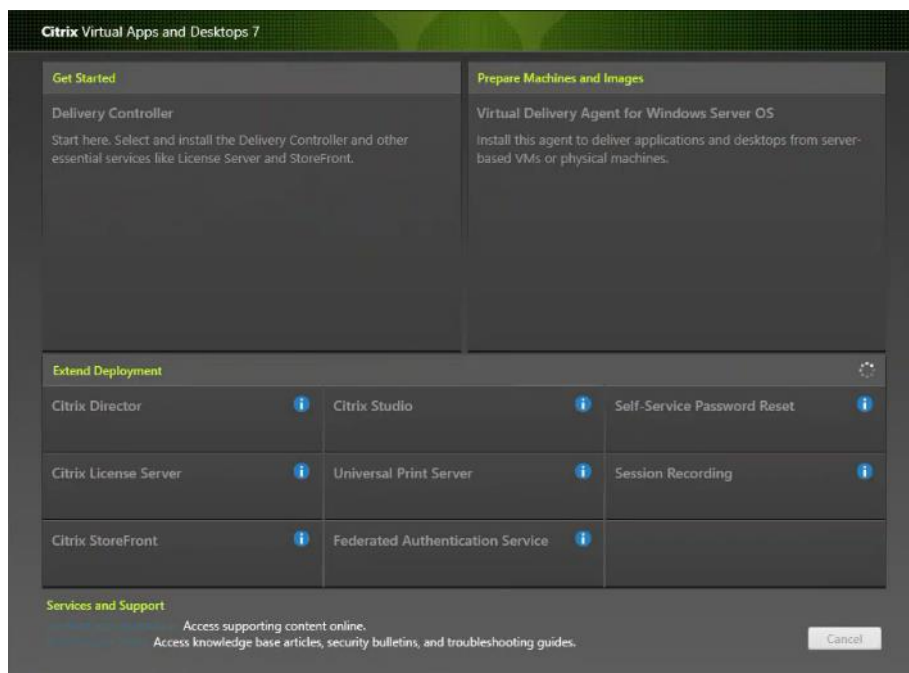
Virtual Delivery Agents (VDAs) are installed on the server and workstation operating systems and enable connections for desktops and apps. The following procedure was used to install VDAs for both HVD and HSD environments.

To install Citrix Desktop Virtual Desktop Agents, follow these steps:

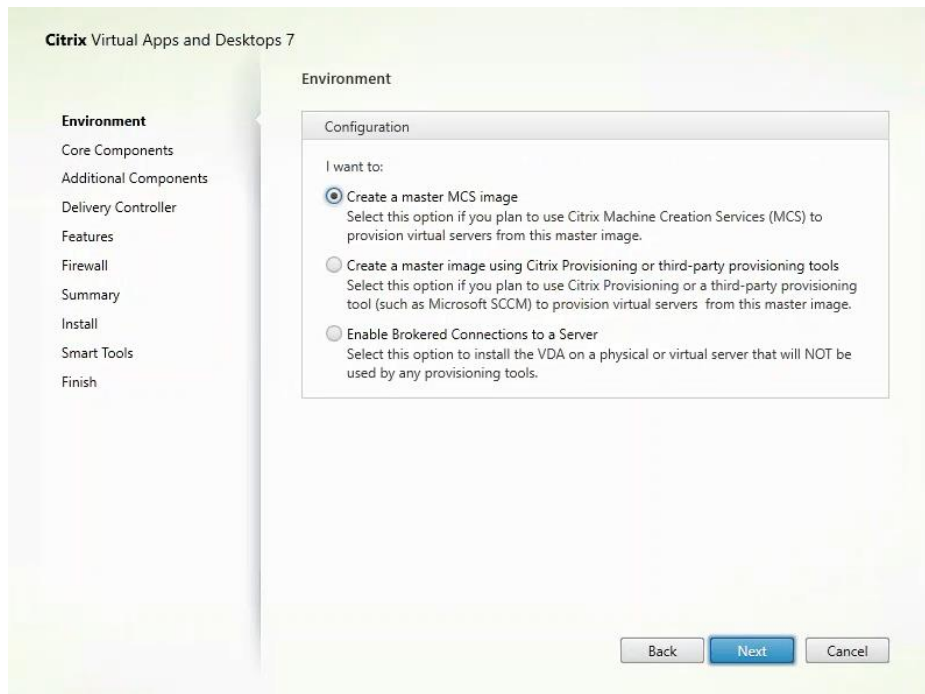
1. Launch the Citrix Desktop installer from the CVA Desktop 1912 LTSR ISO.
2. Click Start on the Welcome Screen.



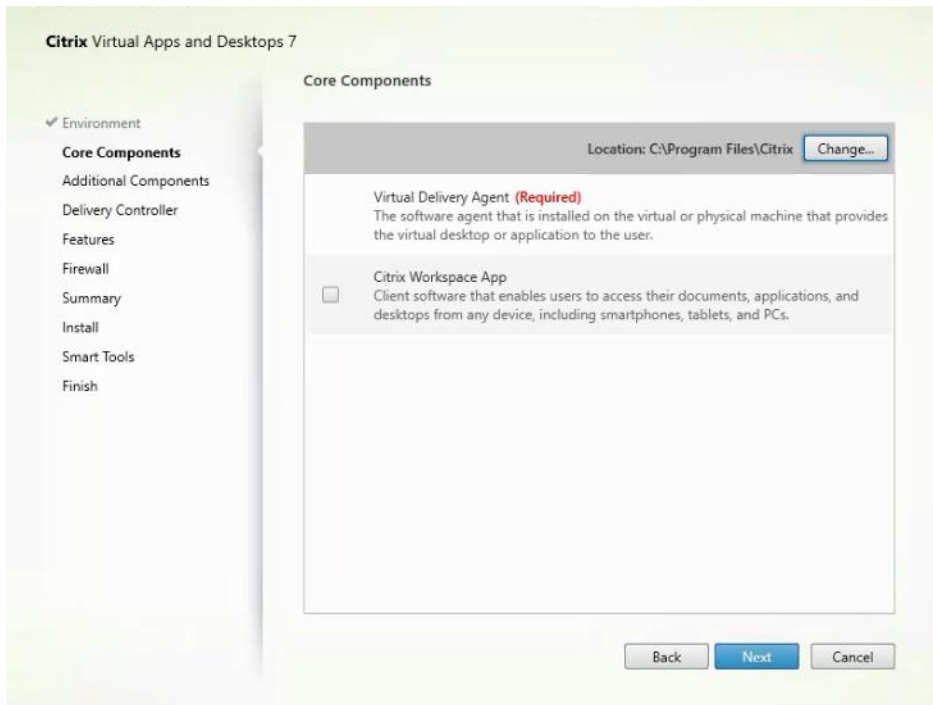
3. To install the VDA for the Hosted Virtual Desktops (VDI), select Virtual Delivery Agent for Windows Desktop OS. After the VDA is installed for Hosted Virtual Desktops, repeat the procedure to install the VDA for Hosted Shared Desktops (RDS). In this case, select Virtual Delivery Agent for Windows Server OS and follow the same basic steps.



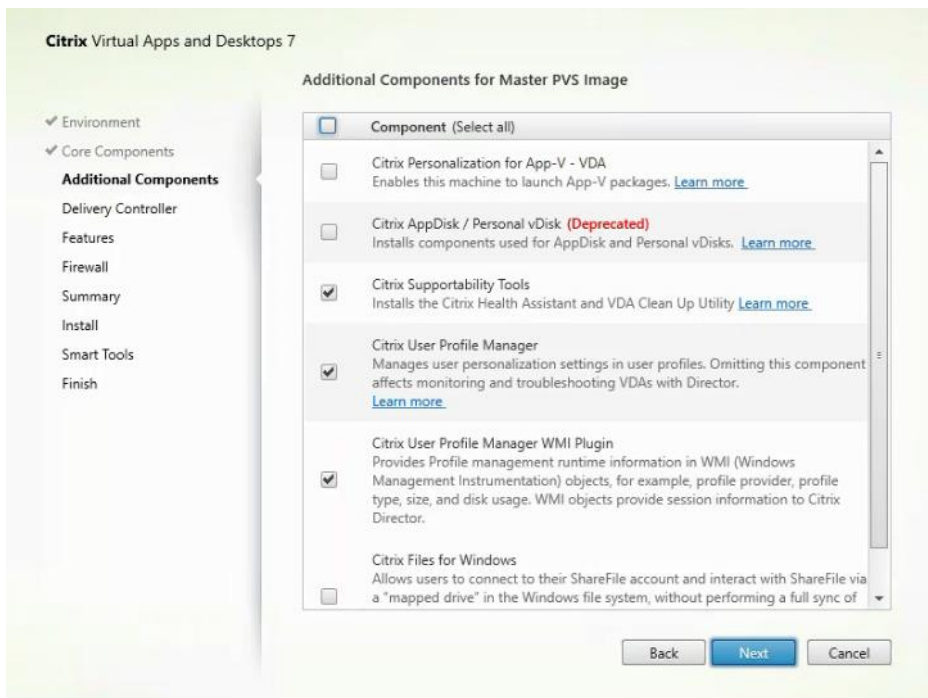
4. Select Create a Master Image.(Be sure to select the proper provisioning technology)
5. Click Next.



6. Optional: Select Citrix Workspace App.
7. Click Next.

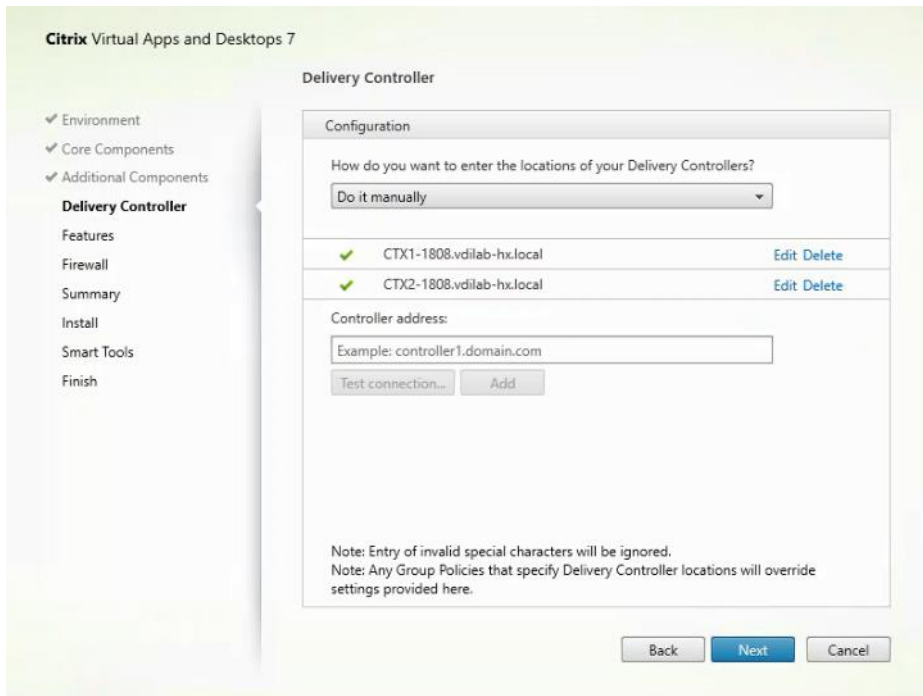


8. Click Next.



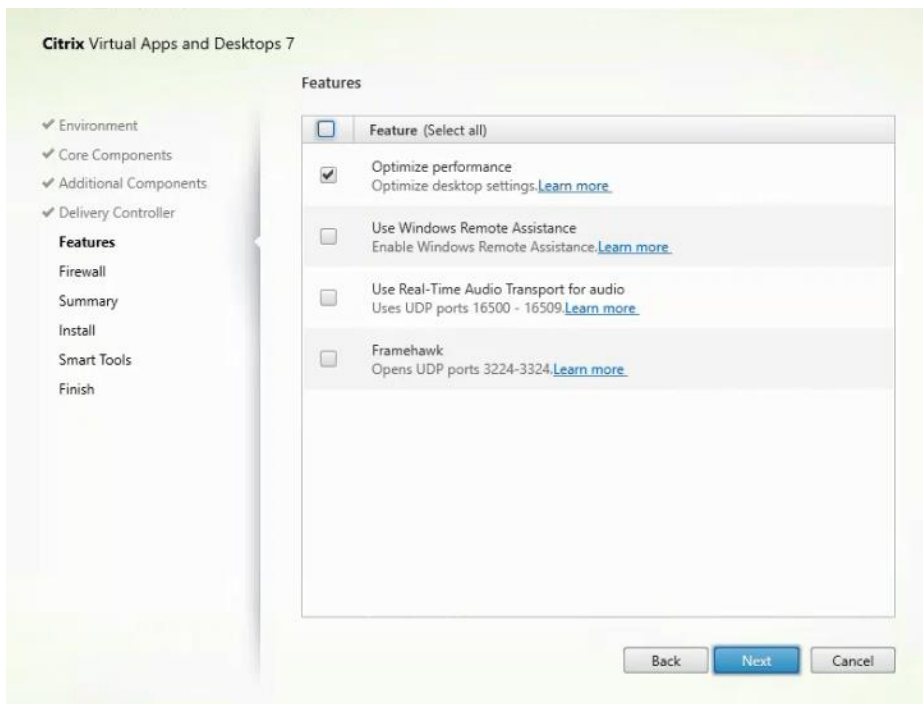
9. Select Do it manually and specify the FQDN of the Delivery Controllers.

10. Click Next.



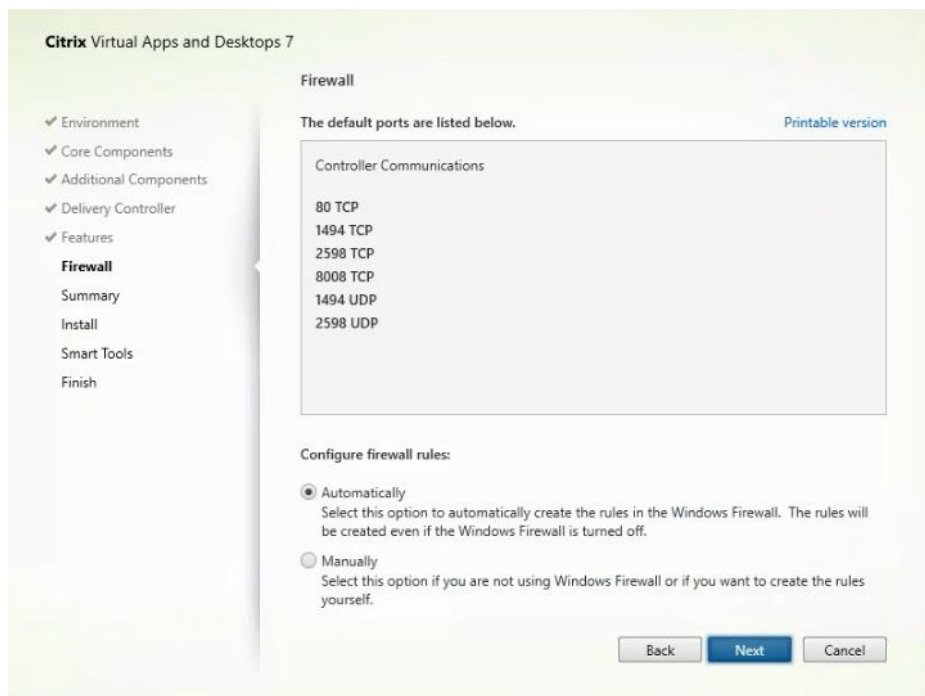
11. Accept the default features.

12. Click Next.



13. Allow the firewall rules to be configured automatically.

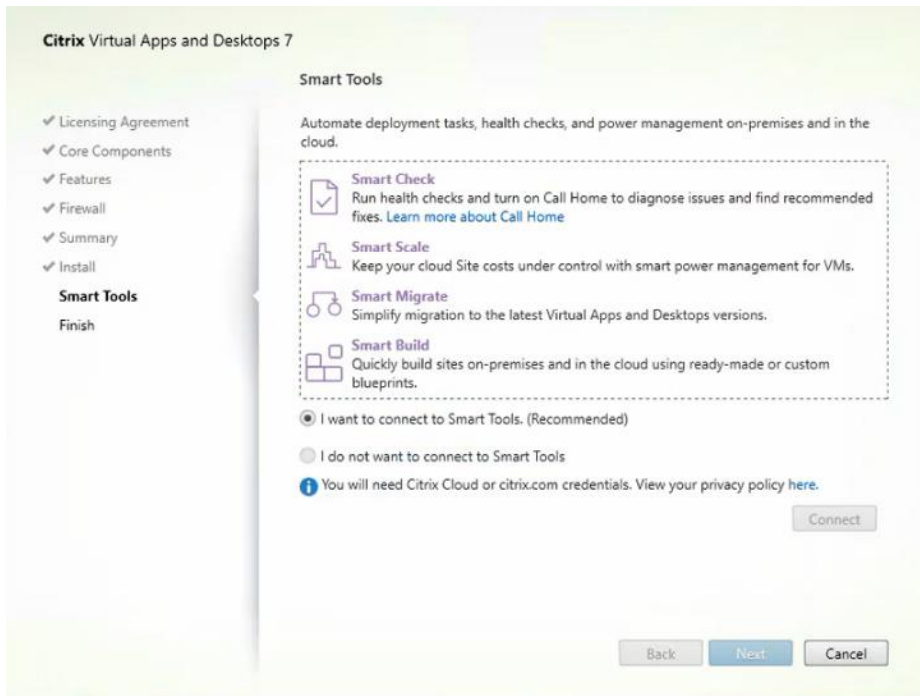
14. Click Next.



15. Verify the Summary and click Install.



16. (Optional) Select Call Home participation.

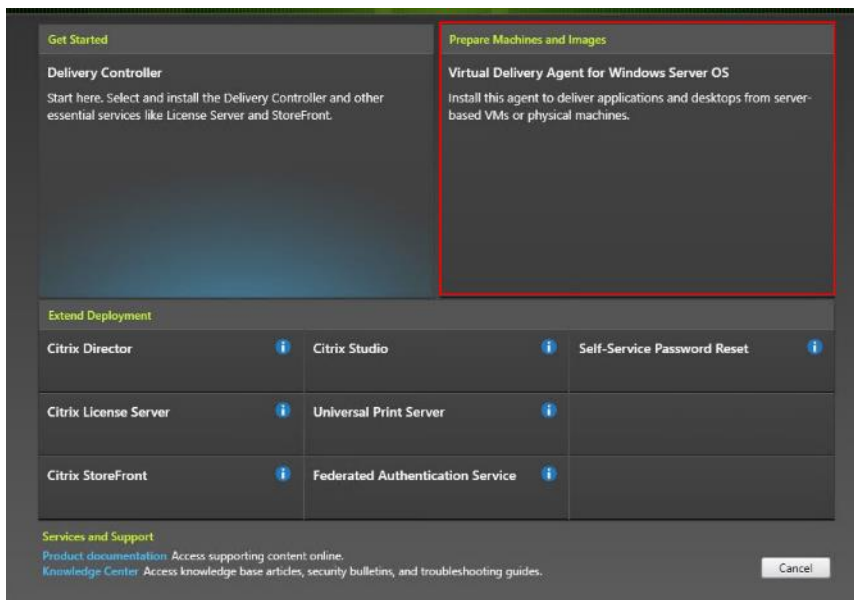


17. (Optional) Check Restart Machine.

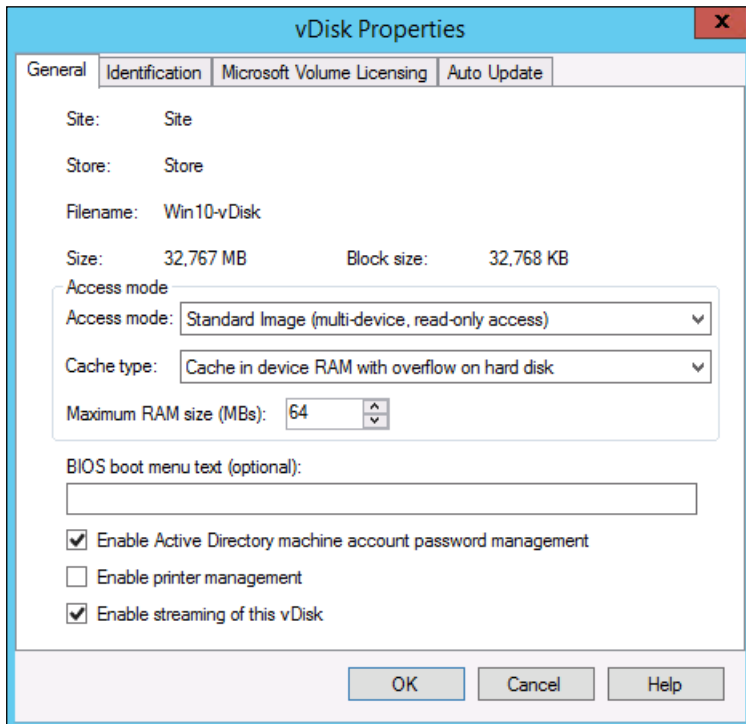
18. Click Finish.

19. Repeat steps 1 - 18 so that VDAs are installed for both HVD (using the Windows 10 OS image) and the HSD desktops (using the Windows Server 2019 image).

20. Select an appropriate workflow for the HSD desktop.



21. Once the Citrix VDA is installed, on the vDisk Properties dialog, change Access mode to “Standard Image (multi-device, read-only access).”
22. Set the Cache Type to Cache in device RAM with overflow on hard disk.
23. Set Maximum RAM size (MBs): 256 for VDI and set 1024 MB for RDS vDisk.

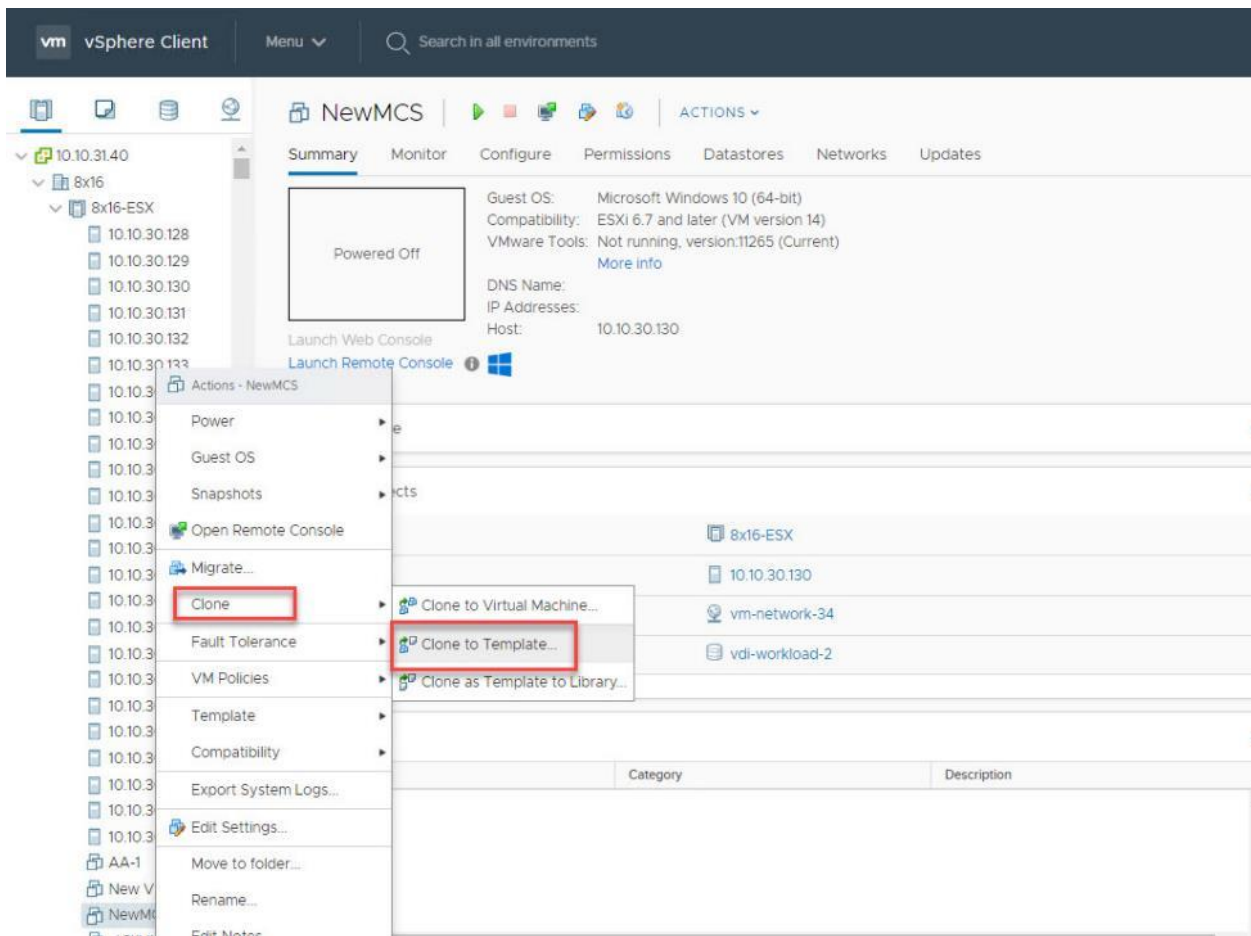


24. Click OK.
25. Repeat steps 1 - 24 to create vDisks for both the Hosted VDI Desktops (using the Windows 10 OS image) and the Hosted Shared Desktops (using the Windows Server 2019 image).

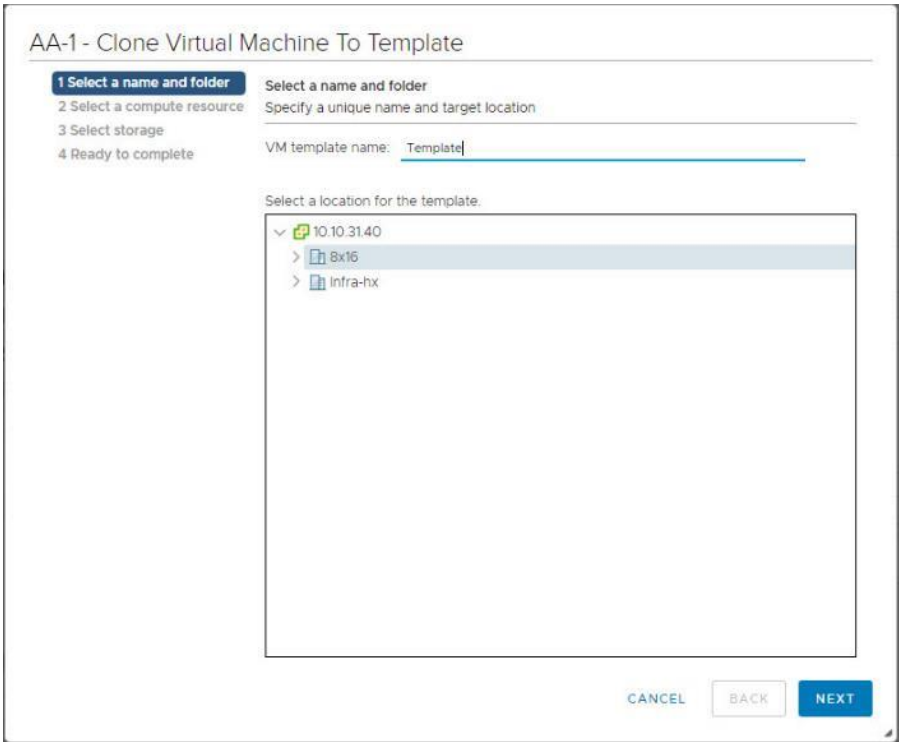
Provision Virtual Desktop Machines

To create VDI and RDS machines, follow these steps:

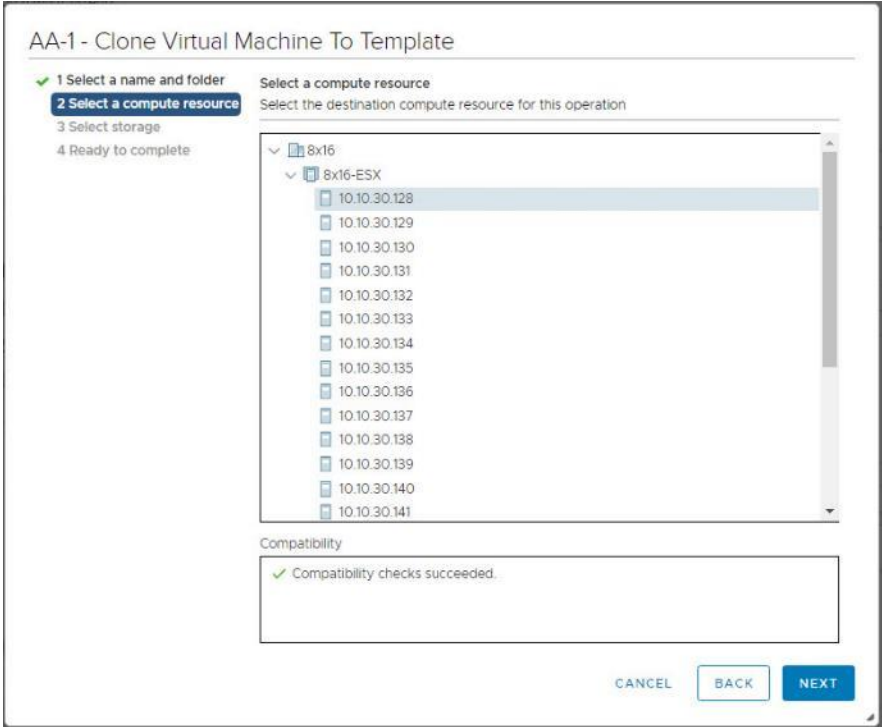
1. Select the Master Target Device virtual machine from the VCenter Client.
2. Right-click the virtual machine and select Clone -> Clone to Template.
3. Name the clone Template.
4. Select the cluster and datastore where the first phase of provisioning will occur.



5. Name the template and click Next.



6. Select a host in the cluster to place the template.



7. Click Next.

AA-1 - Clone Virtual Machine To Template

✓ 1 Select a name and folder
✓ 2 Select a compute resource
3 Select storage
4 Ready to complete

Select storage
Select the storage for the configuration and disk files

Configure per disk

Select virtual disk format: Same format as source

VM Storage Policy: Keep existing VM storage policies

Name	Capacity	Provisioned	Free	Type
esxtop	1 TB	67.5 GB	960.7 GB	NF
SpringpathDS-WZP22121...	216 GB	12.16 GB	203.84 GB	VF
vdi-workload-2	60 TB	46.25 TB	58.2 TB	NF
vdi_workload	40 TB	46.49 TB	38.04 TB	NF

Compatibility

✓ Compatibility checks succeeded.

CANCEL BACK NEXT

8. Click Next.
9. Click Next through the remaining screens
10. Click Finish to create the template.

AA-1 - Clone Virtual Machine To Template

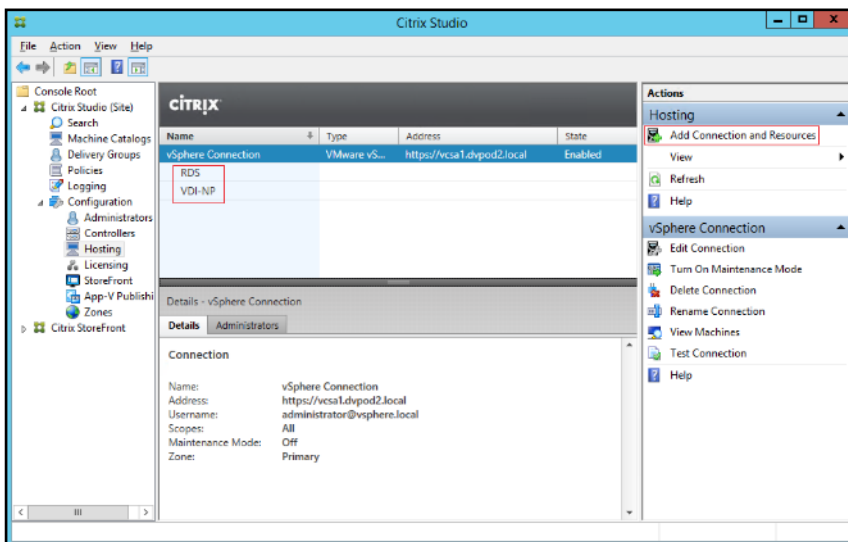
- ✓ 1 Select a name and folder
- ✓ 2 Select a compute resource
- ✓ 3 Select storage
- 4 Ready to complete**

Ready to complete
Click Finish to start creation.

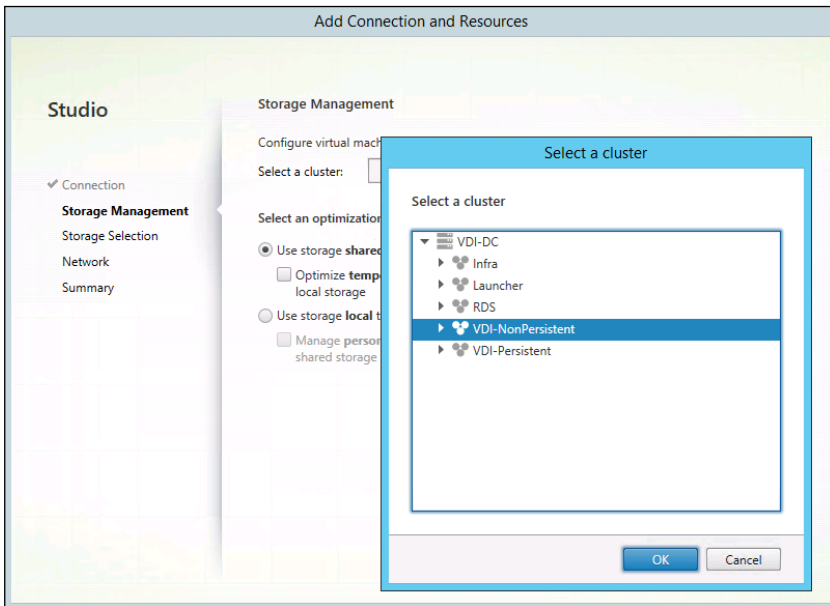
Provisioning type	Clone virtual machine to template
Source virtual machine	AA-1
Template name	Template
Folder	8x16
Host	10.10.30.128
Datastore	vdi_workload
Disk storage	Same format as source

CANCEL BACK FINISH

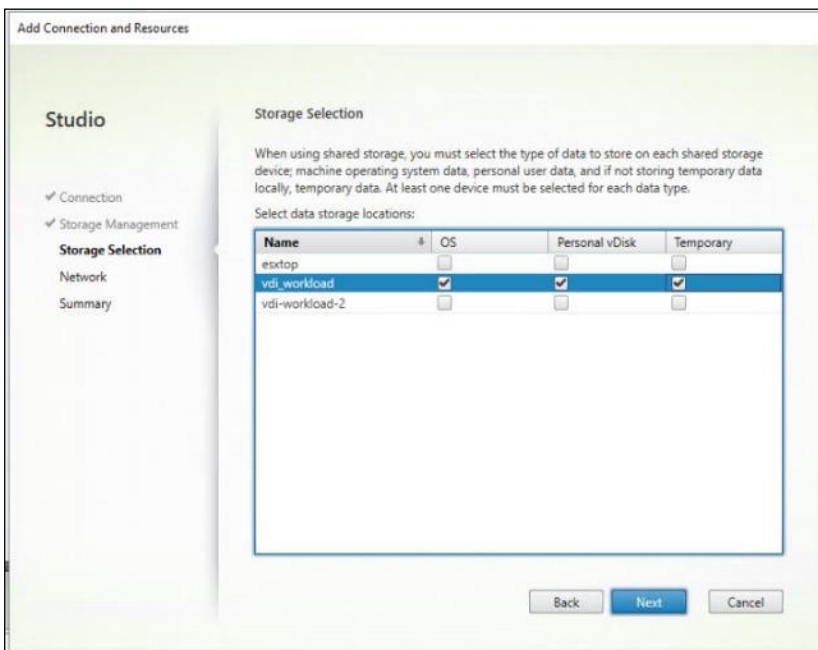
11. From Citrix Studio on the Desktop Controller, select Hosting and Add Connection and Resources.
12. Select Use an existing Connection and click Next.
13. Correspond the name of the resource with desktop machine clusters.



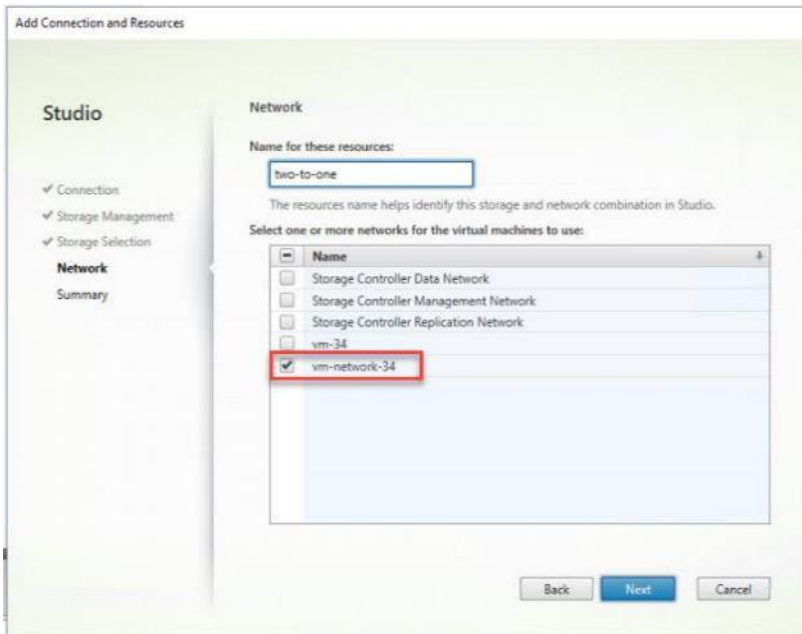
14. Browse and select the VCenter cluster for desktop provisioning and use the default storage method Use storage shared by hypervisors.



15. Select the data storage location for the corresponding resource.



16. Select the VDI networks for the desktop machines and click Next.



17. Click Finish.

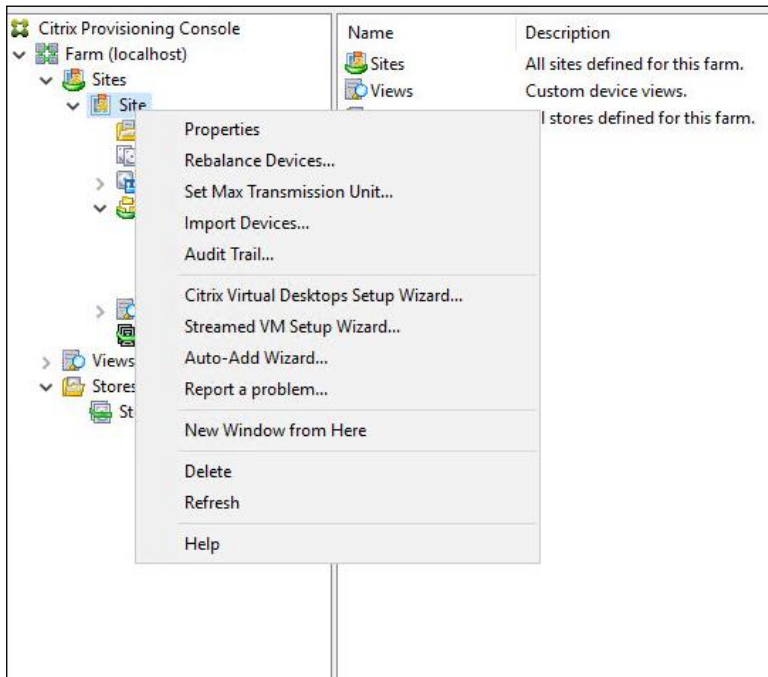


Return to these settings to alter the datastore selection for each set of provisioned desktop machines if you want to create a separate datastore for each image

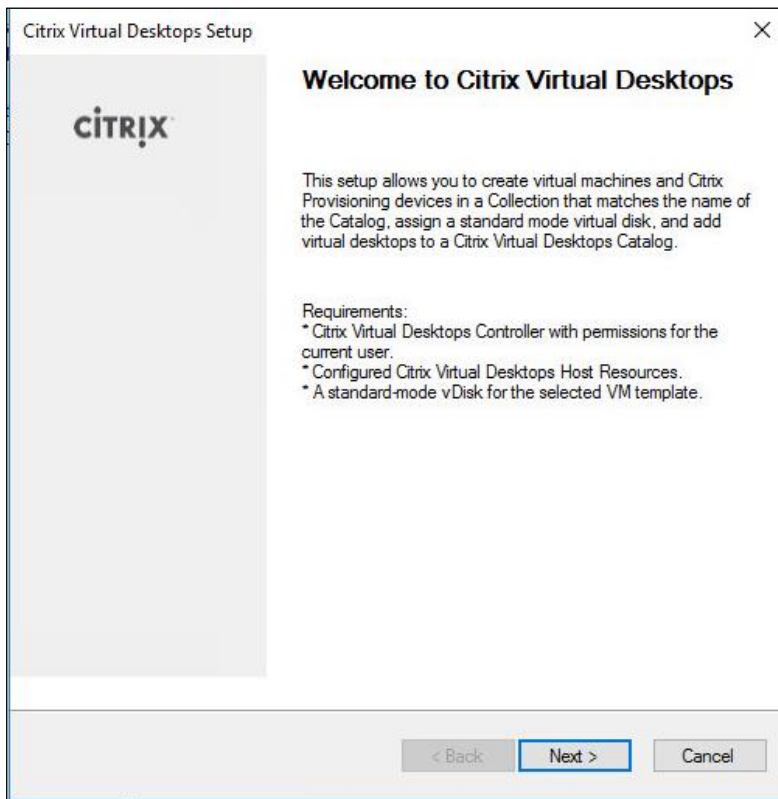
Provision Desktop Machines from Citrix Provisioning Services Console

To provision the desktop machines using the Citrix Provisioning Service Console, follow these steps:

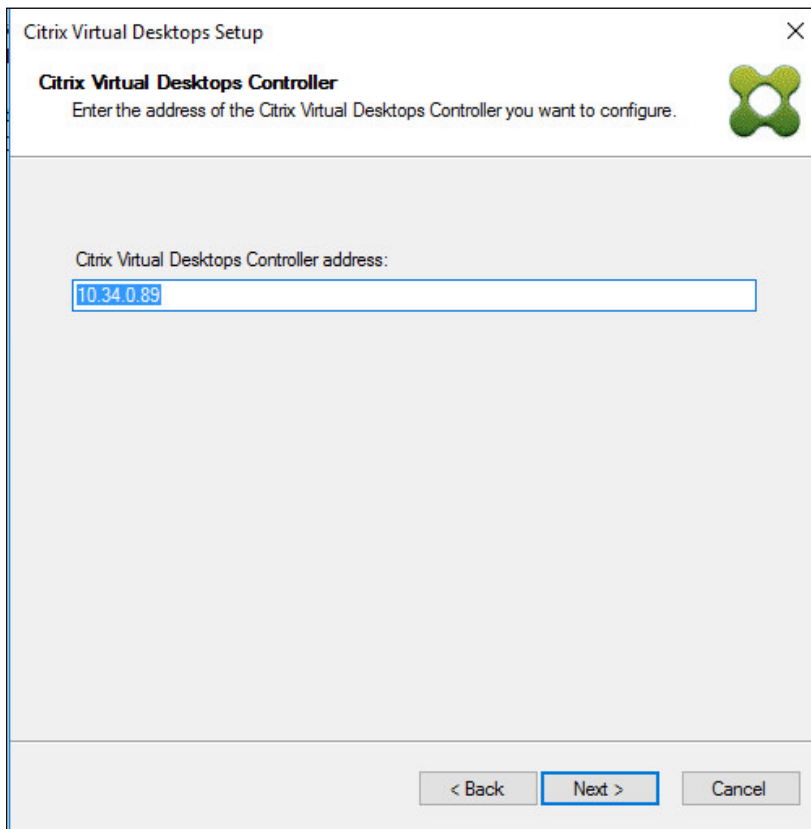
1. Start the Virtual Desktops Setup Wizard from the Provisioning Services Console.
2. Right-click the Site.
3. Choose Virtual Desktops Setup Wizard... from the context menu.



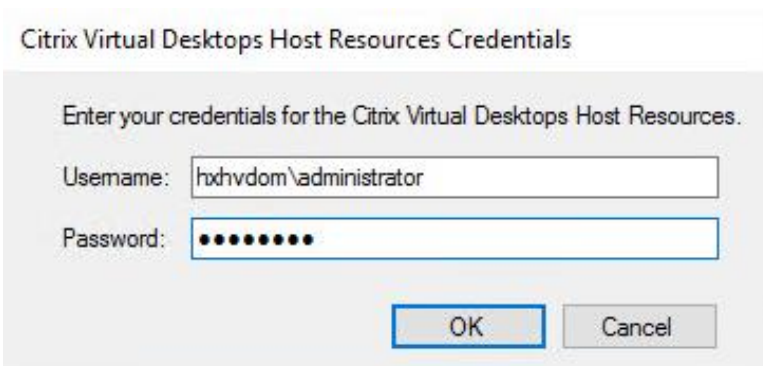
4. Click Next.
5. Enter the Virtual Desktops Controller address that will be used for the wizard operations.
6. Click Next.



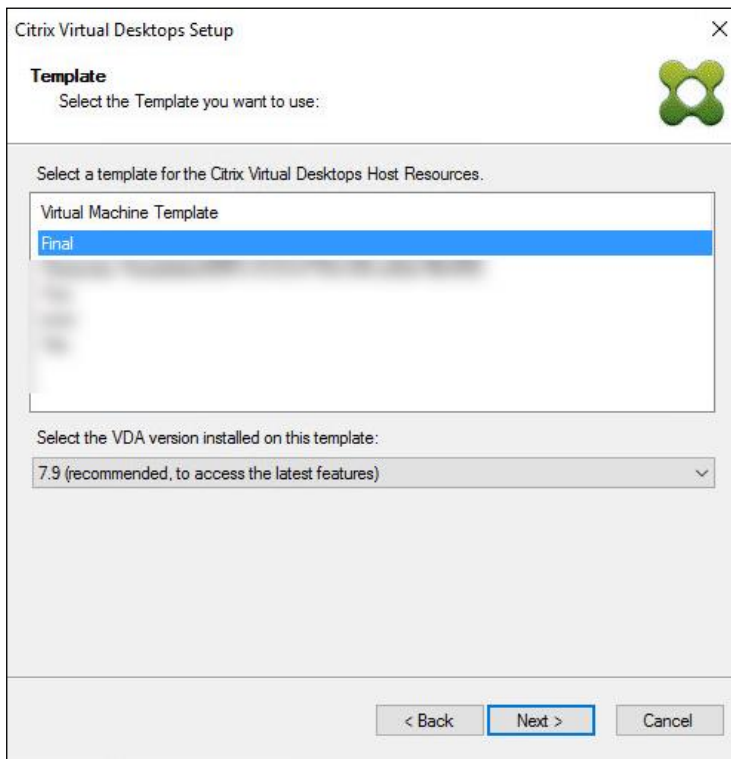
7. Select the Host Resources on which the virtual machines will be created.
8. Click Next.



9. Provide the Host Resources Credentials (Username and Password) to the Virtual Desktops controller when prompted.
10. Click OK.

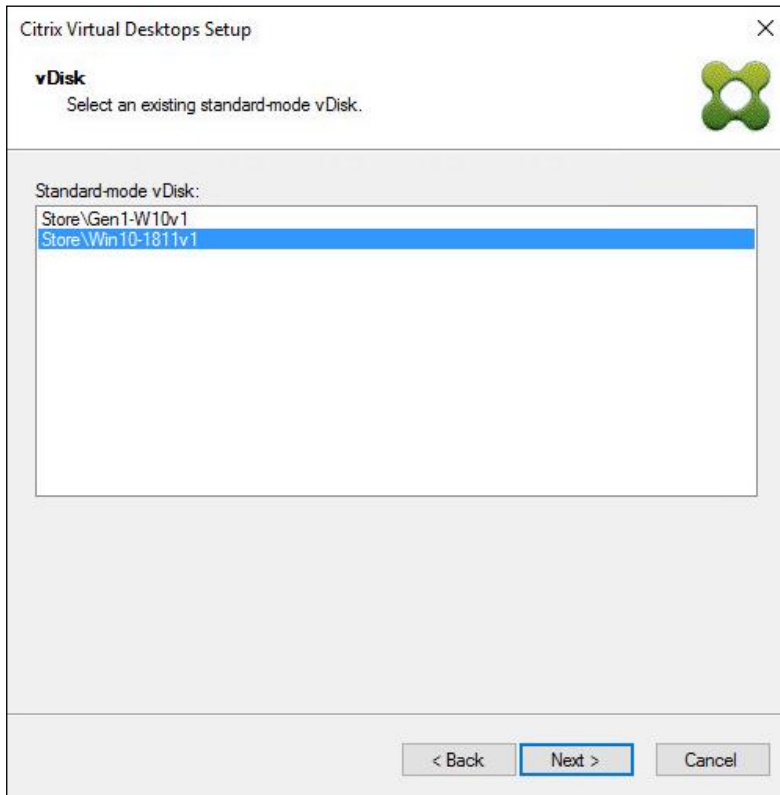


11. Select the Template created earlier.
12. Click Next.



13. Select the vDisk that will be used to stream virtual machines.

14. Click Next.



15. Select Create a new catalog.



The catalog name is also used as the collection name in the PVS site.

16. Click Next.

Citrix Virtual Desktops Setup

Catalog
Select your Catalog preferences.

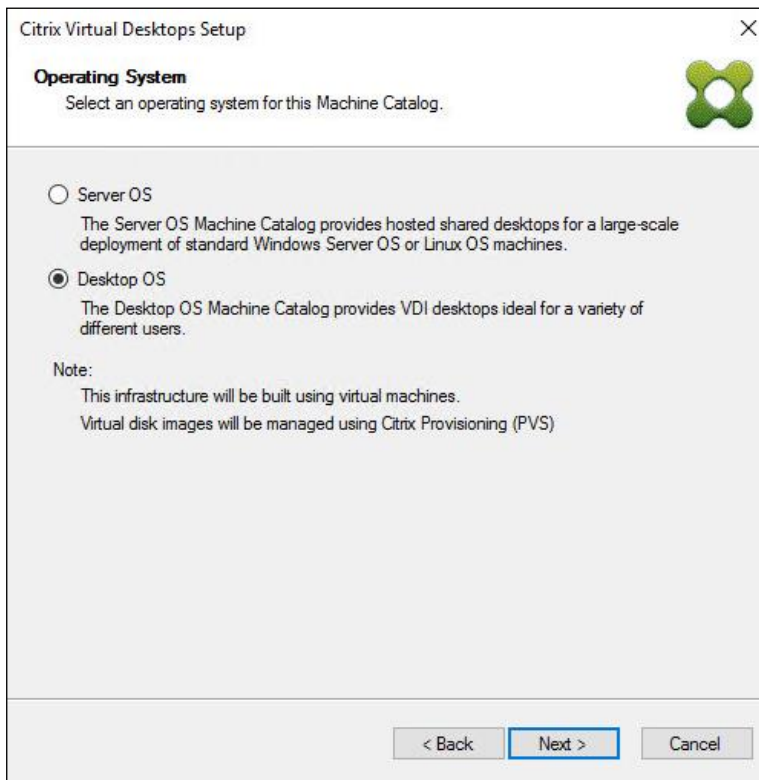
Create a new catalog
 Use an existing catalog

Catalog name:
Description:

< Back Next > Cancel

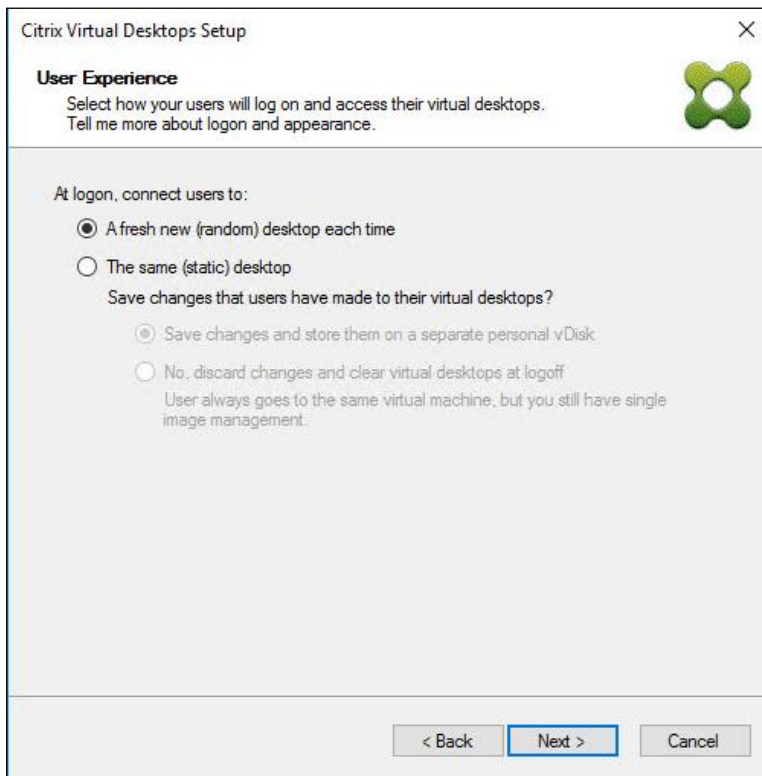
17. On the Operating System dialog, specify the operating system for the catalog. Specify Windows Desktop Operating System for VDI and Windows Server Operating System for RDS.

18. Click Next.



19. If you specified a Windows Desktop OS for VDIs, a User Experience dialog appears. Specify that the user will connect to A fresh new (random) desktop each time.

20. Click Next.



21. On the Virtual machines dialog, specify:

- a. The number of virtual machines to create.
- b. Number of vCPUs for the virtual machine (2 for VDI, 8 for RDS).
- c. The amount of memory for the virtual machine (4GB for VDI, 24GB for RDS).
- d. The write-cache disk size (10GB for VDI, 30GB for RDS).
- e. PXE boot as the Boot Mode.

22. Click Next.

Citrix Virtual Desktops Setup

Virtual machines
Select your virtual machine preferences.

Number of virtual machines to create: 800

vCPUs: 2

Memory: 4096 MB

Local write cache disk: 6 GB

Boot mode:

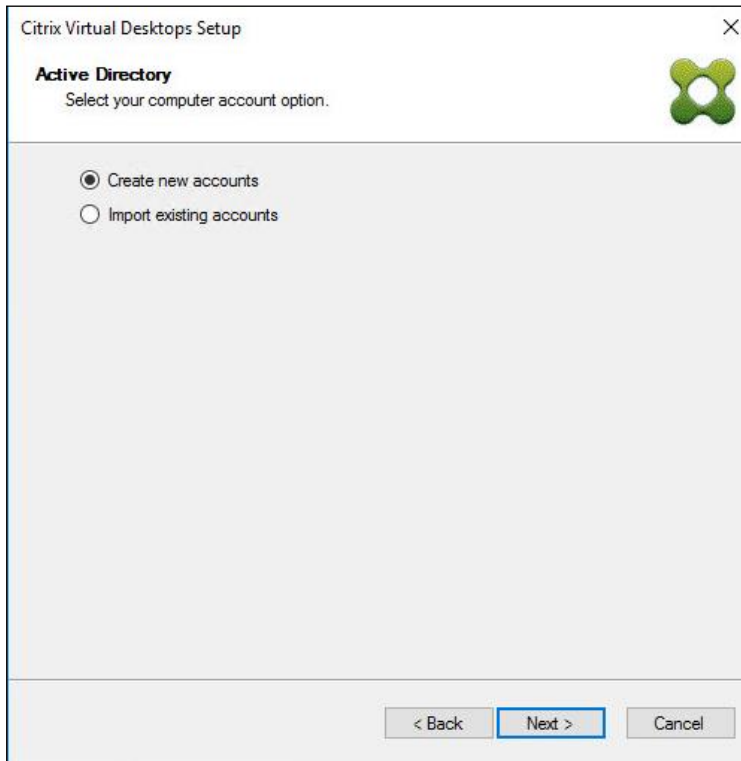
PXE boot (requires a running PXE service)

BDM disk (create a boot device manager partition)

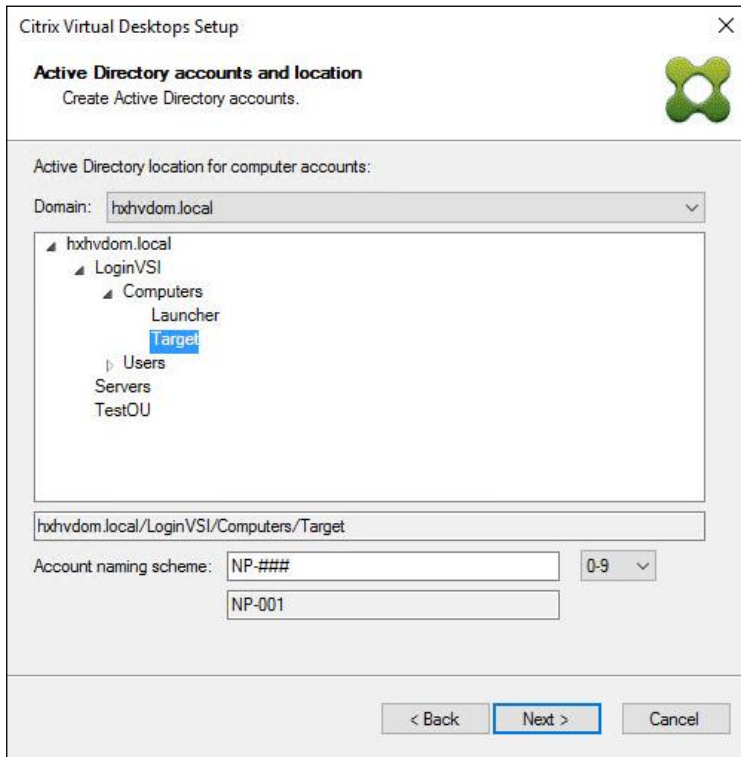
< Back Next > Cancel

23. Select the Create new accounts radio button.

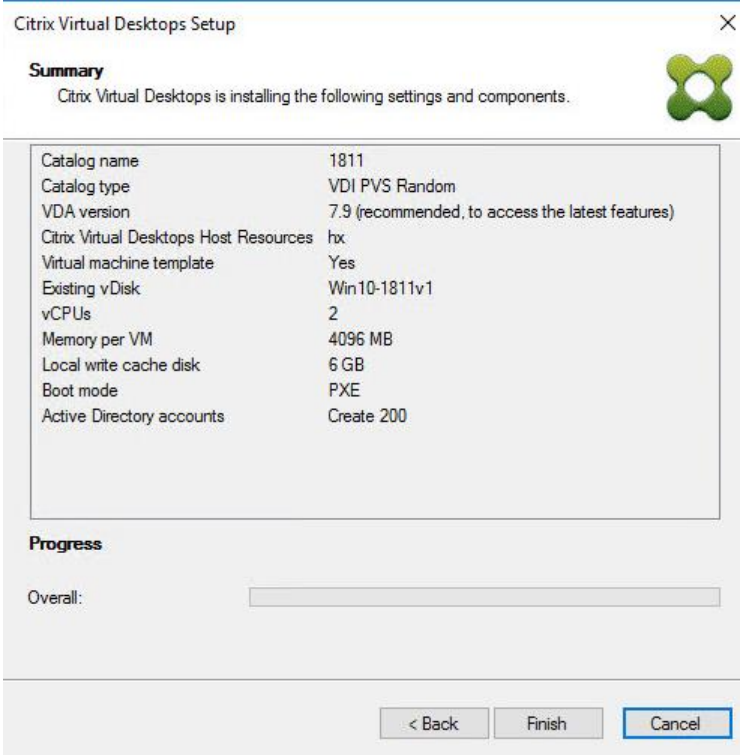
24. Click Next.



25. Specify the Active Directory Accounts and Location. This is where the wizard should create the computer accounts.
26. Provide the Account naming scheme. An example name is shown in the text box below the name scheme selection location.
27. Click Next.



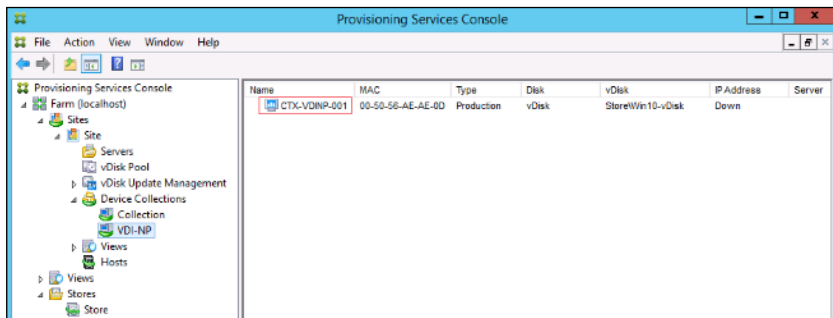
28. Click Finish to begin the virtual machine creation.



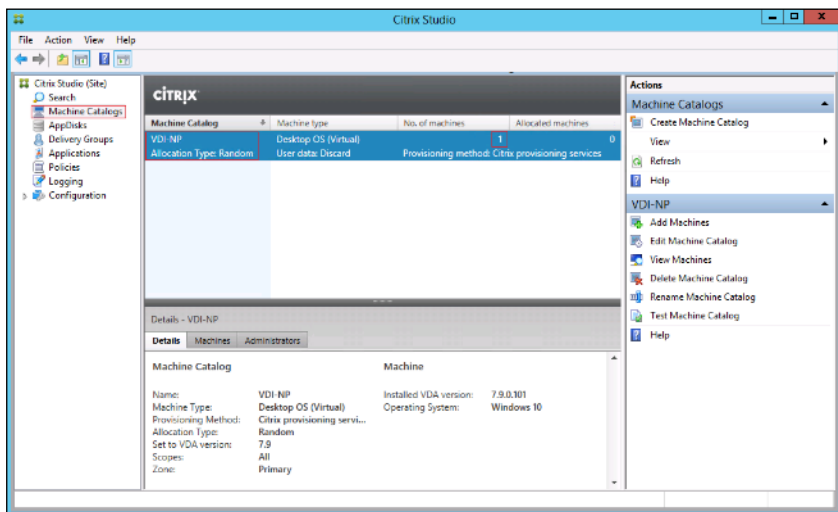
29. When the wizard is done provisioning the virtual machines, click Done.

30. Verify the desktop machines were successfully created in the following locations:

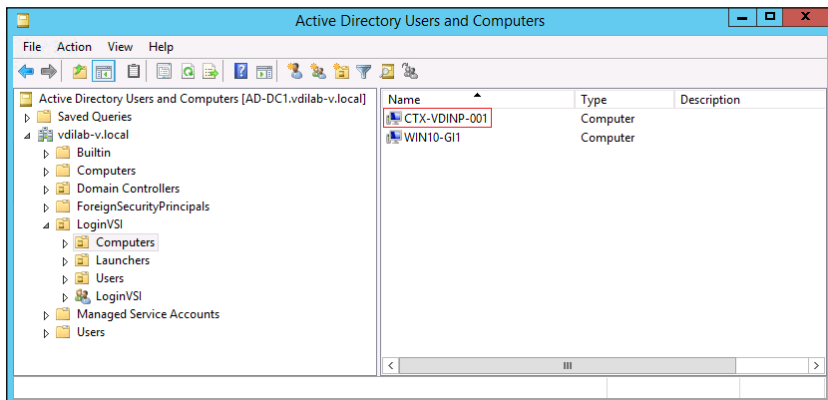
a. PVS1 > Provisioning Services Console > Farm > Site > Device Collections > VDI-NP > CTX-VDI-001



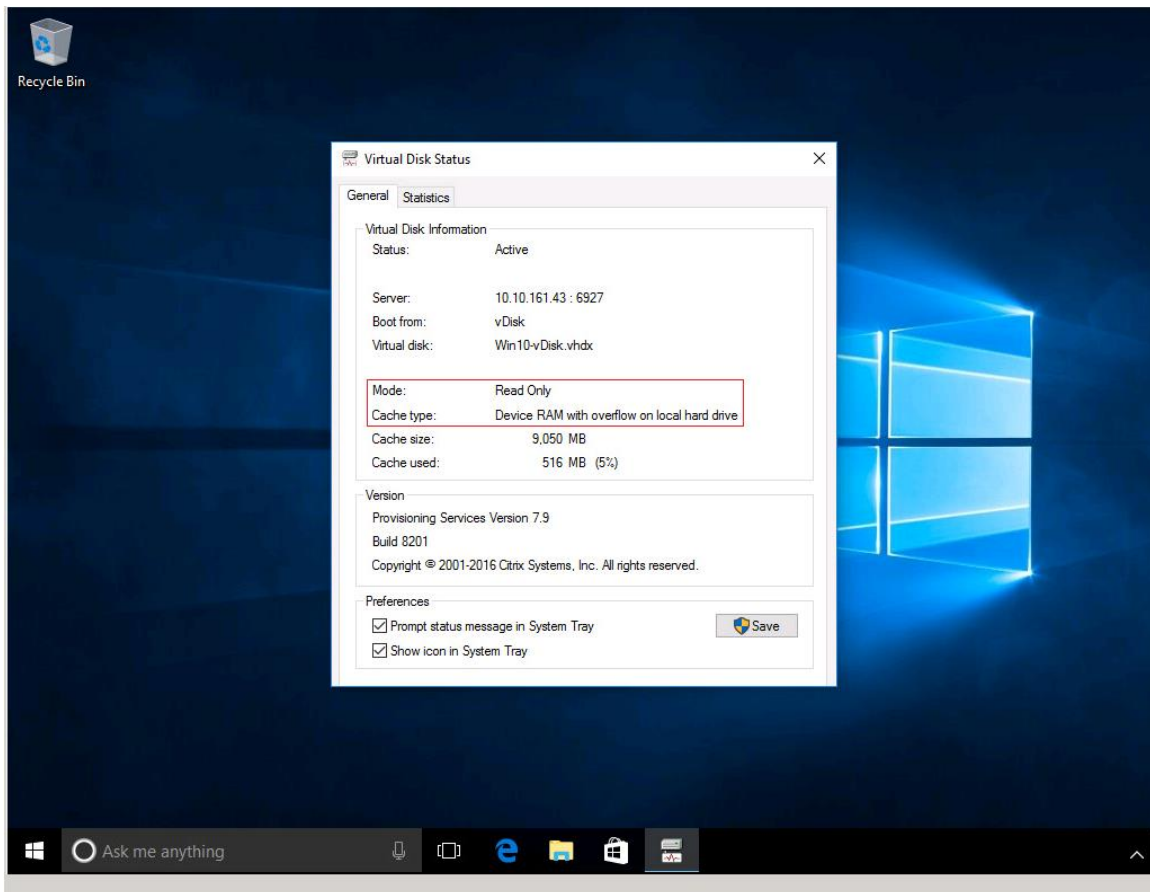
b. CTX-XD1 > Citrix Studio > Machine Catalogs > VDI-NP



c. AD-DC1 > Active Directory Users and Computers > hxxhvd.com.local > Computers > CTX-VDI-001



31. Log into the newly provisioned desktop machine, using the Virtual Disk Status verify the image mode is set to Ready Only and the cache type as Device Ram with overflow on local hard drive.



Install Citrix Virtual Apps and Desktop Virtual Desktop Agents

Virtual Delivery Agents (VDAs) are installed on the server and workstation operating systems and enable connections for desktops and apps. The following procedure was used to install VDAs for both HVD and HSD environments.

By default, when you install the Virtual Delivery Agent, Citrix User Profile Management is installed silently on master images.



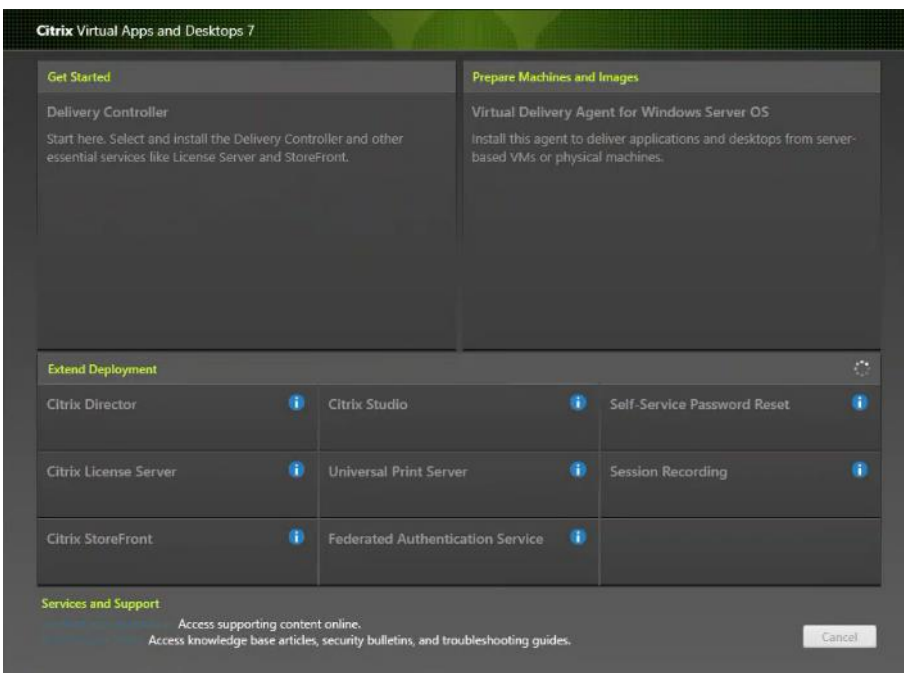
Using profile management as a profile solution is optional but was used for this CVD and is described in a subsequent section.

To install Citrix Desktop Virtual Desktop Agents, follow these steps:

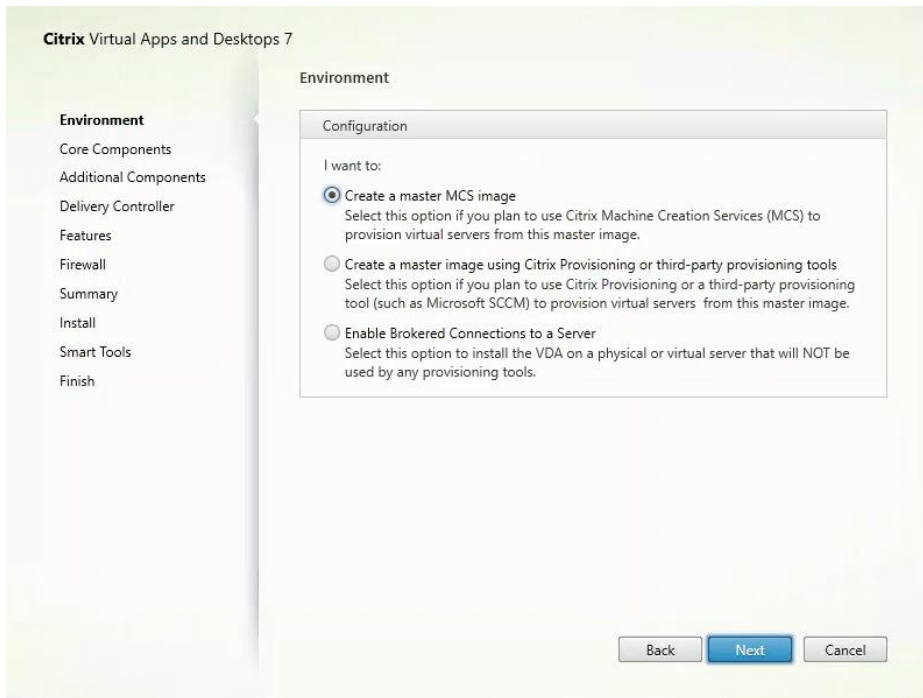
1. Launch the Citrix Desktop installer from the CVA Desktop 1912 LTSR ISO.
2. Click Start on the Welcome Screen.



- To install the VDA for the Hosted Virtual Desktops (VDI), select Virtual Delivery Agent for Windows Desktop OS. After the VDA is installed for Hosted Virtual Desktops, repeat the procedure to install the VDA for Hosted Shared Desktops (RDS). In this case, select Virtual Delivery Agent for Windows Server OS and follow the same basic steps.

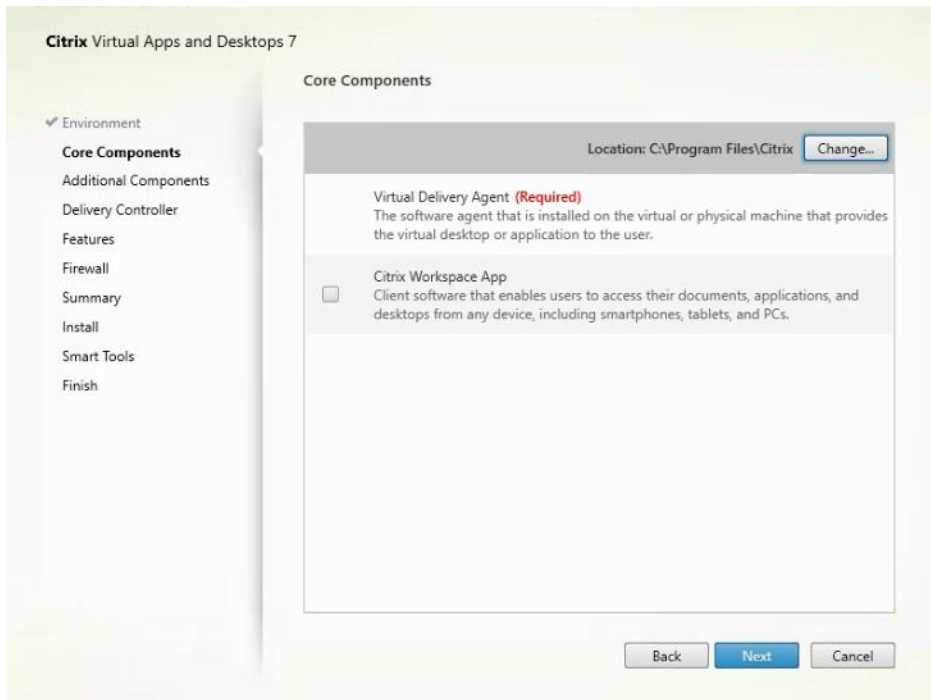


- Select Create a Master Image. (Be sure to select the proper provisioning technology)
- Click Next.

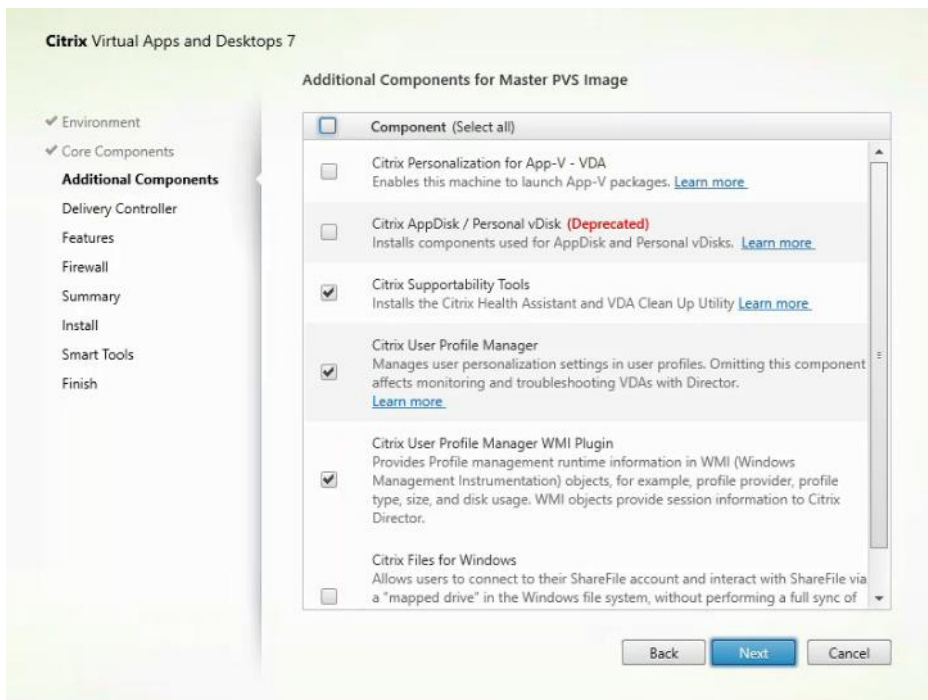


6. Optional: Select Citrix Workspace App.

7. Click Next.

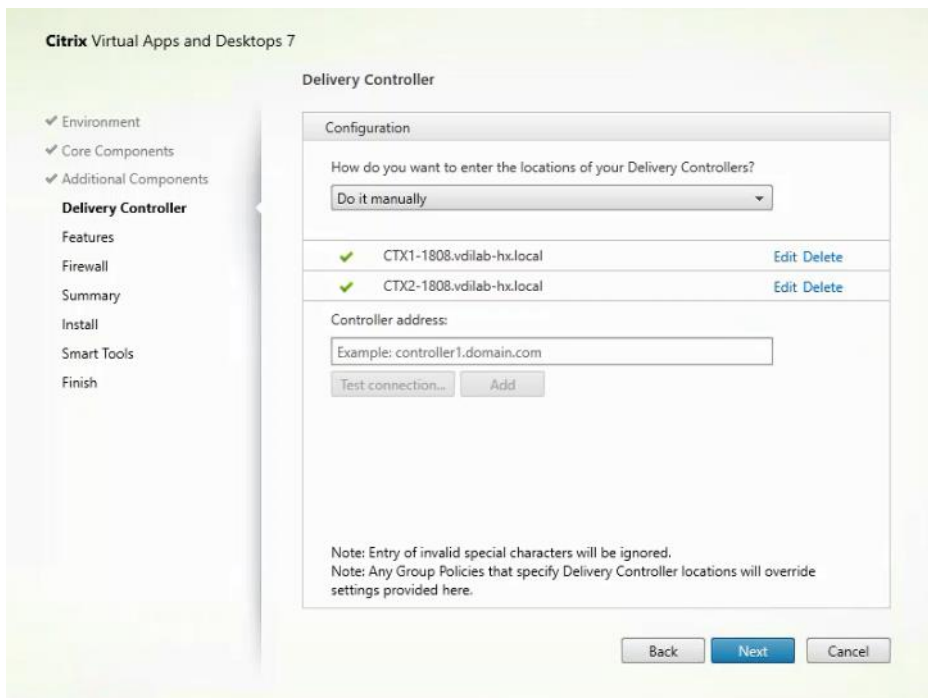


8. Click Next.



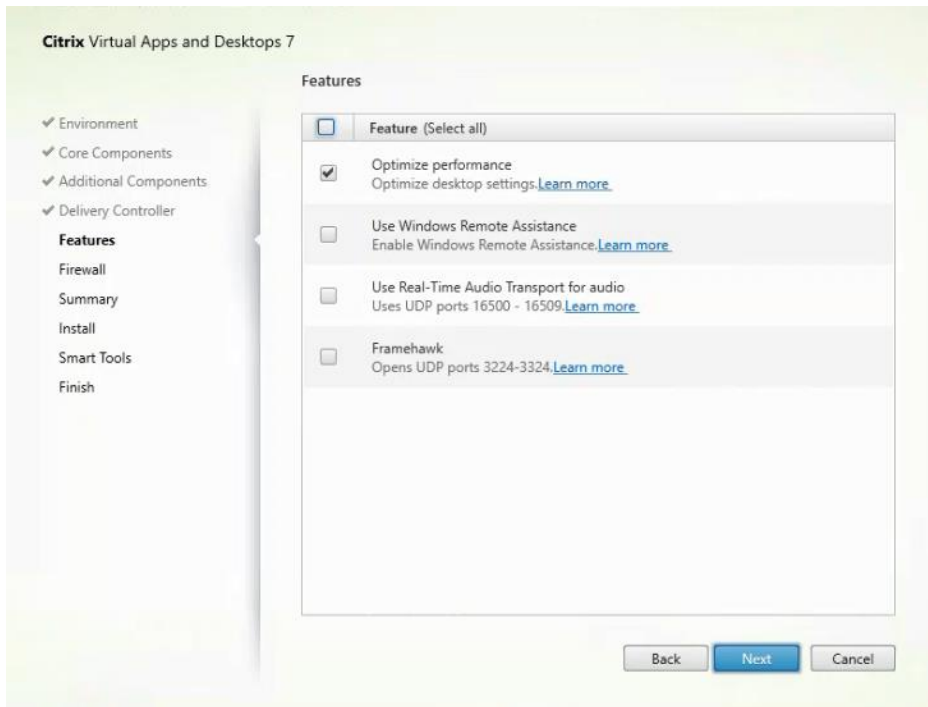
9. Select Do it manually and specify the FQDN of the Delivery Controllers.

10. Click Next.



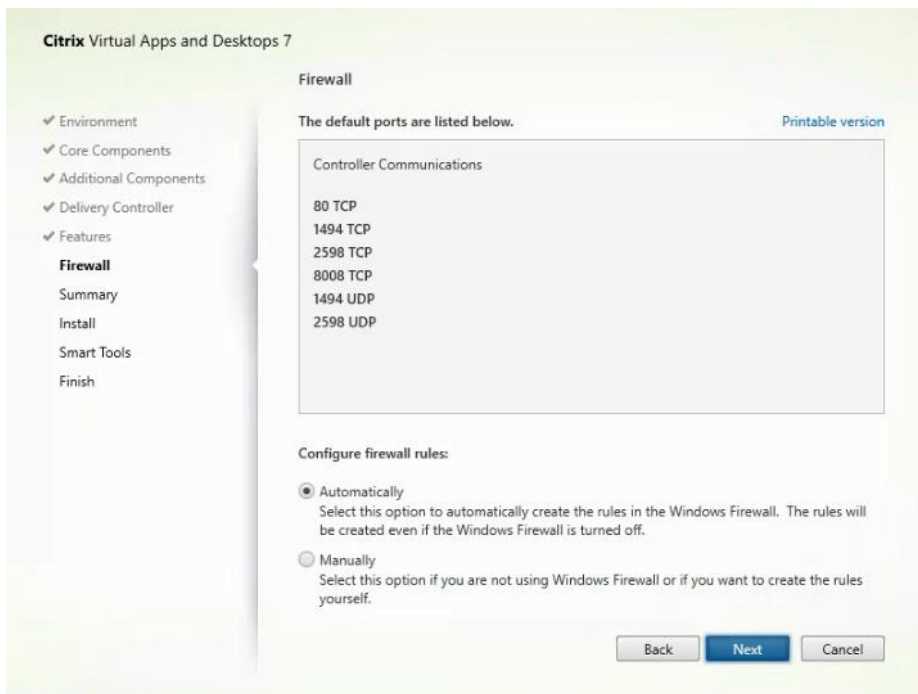
11. Accept the default features.

12. Click Next.

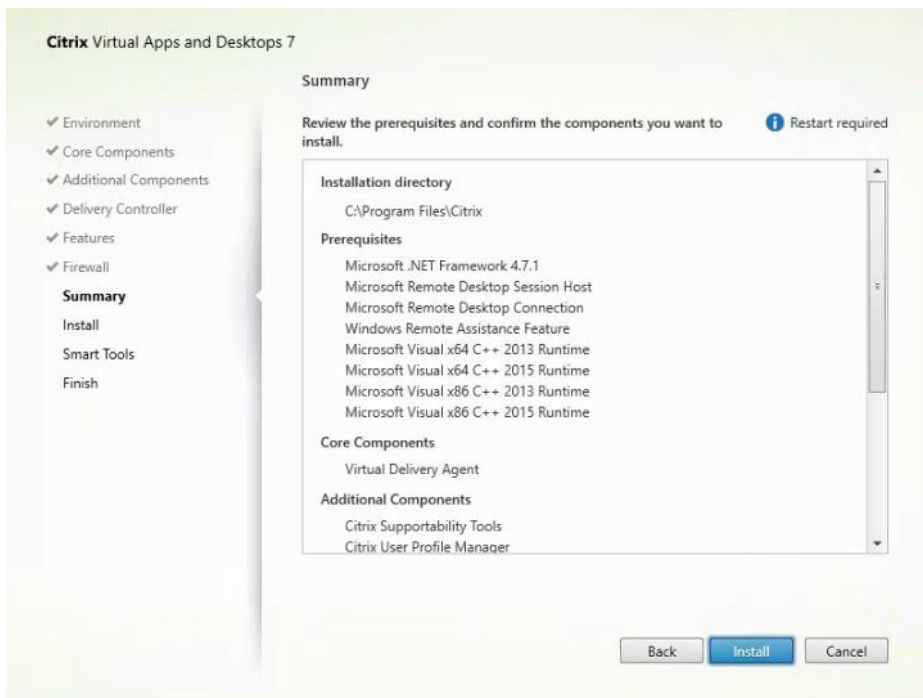


13. Allow the firewall rules to be configured automatically.

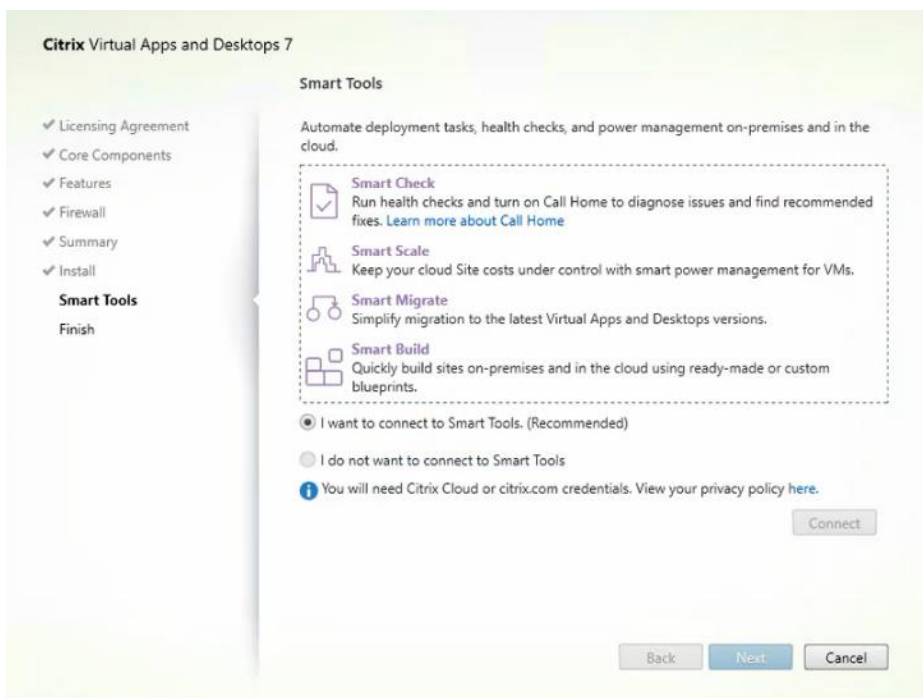
14. Click Next.



15. Verify the Summary and click Install.



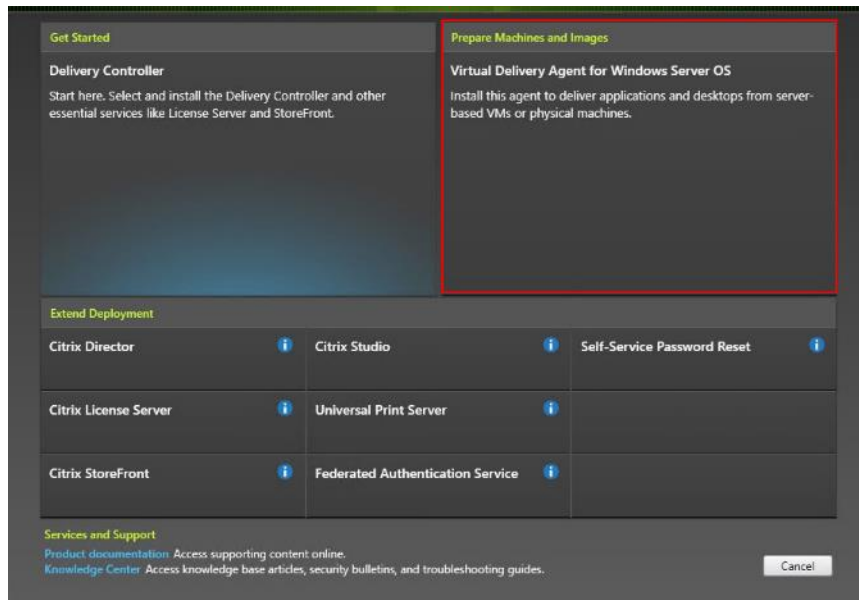
16. (Optional) Select Call Home participation.



17. (Optional) Check Restart Machine.

18. Click Finish.

19. Repeat steps 1 - 18 so that VDAs are installed for both HVD (using the Windows 10 OS image) and the HSD desktops (using the Windows Server 2019 image).
20. Select an appropriate workflow for the HSD desktop.



Create Delivery Groups

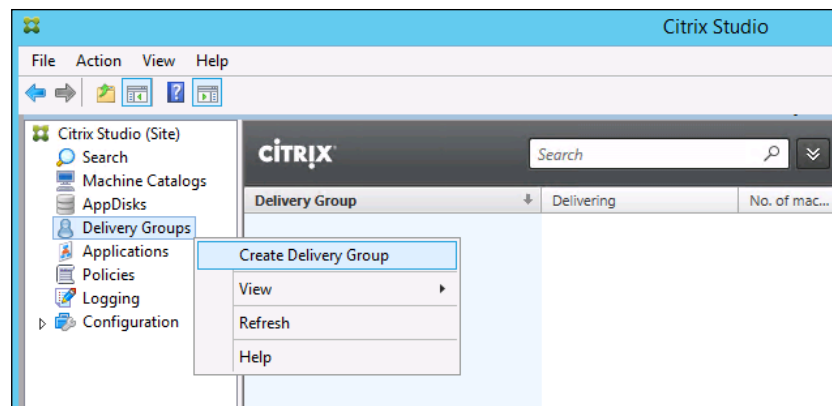
Delivery Groups are collections of machines that control access to desktops and applications. With Delivery Groups, you can specify which users and groups can access which desktops and applications.

To create delivery groups, follow these steps:

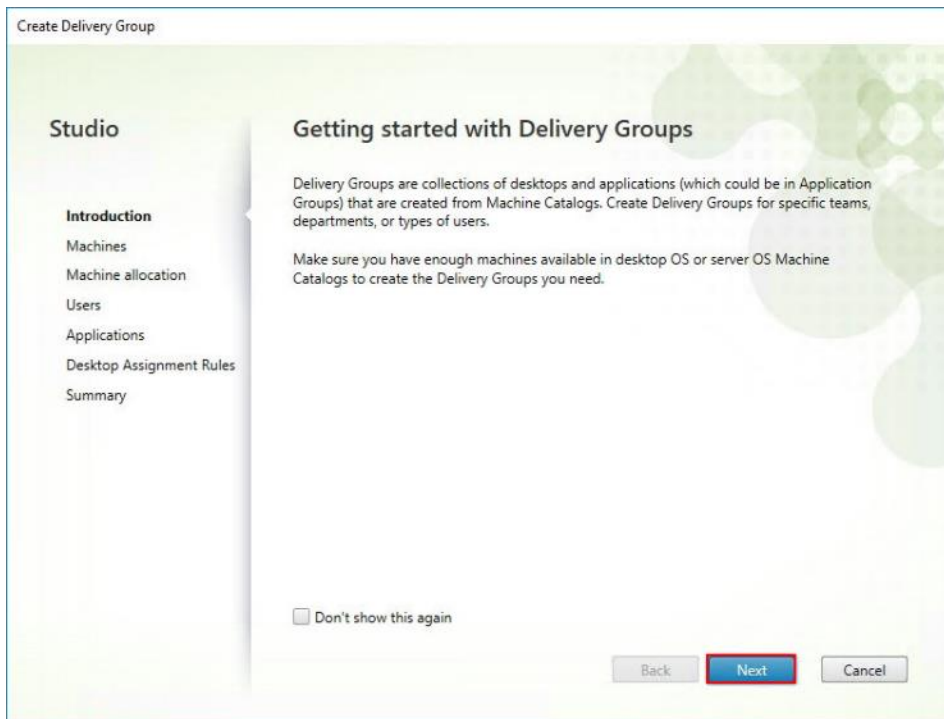


The instructions below outline the steps to create a Delivery Group for VDI desktops. When you have completed these steps, repeat the procedure to a Delivery Group for HVD desktops.

1. Connect to a Virtual Desktops server and launch Citrix Studio.
2. Choose Create Delivery Group from the drop-down list.



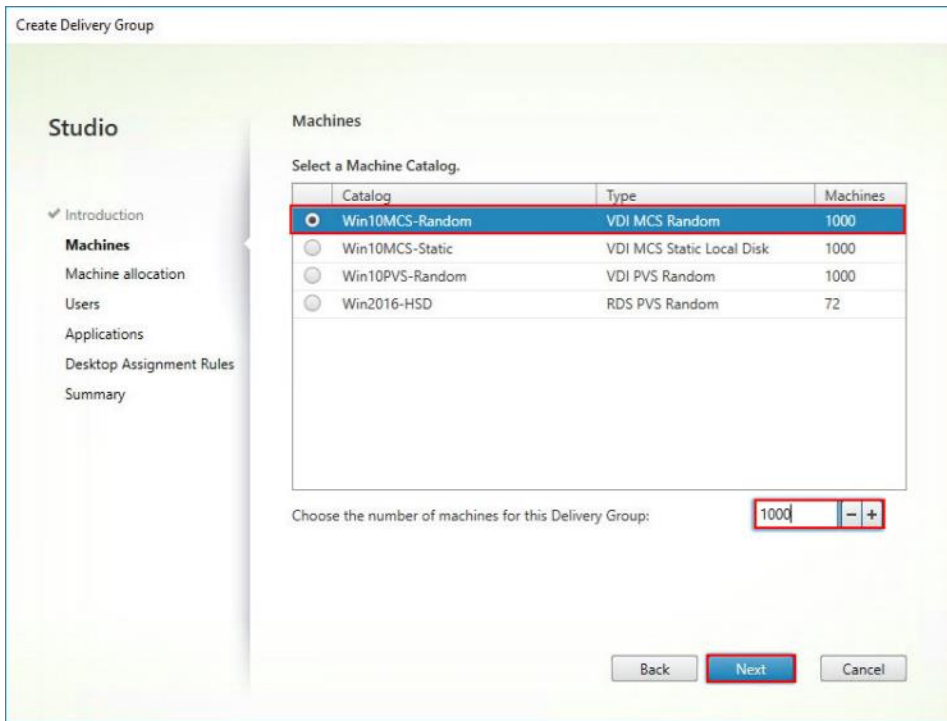
3. Click Next.



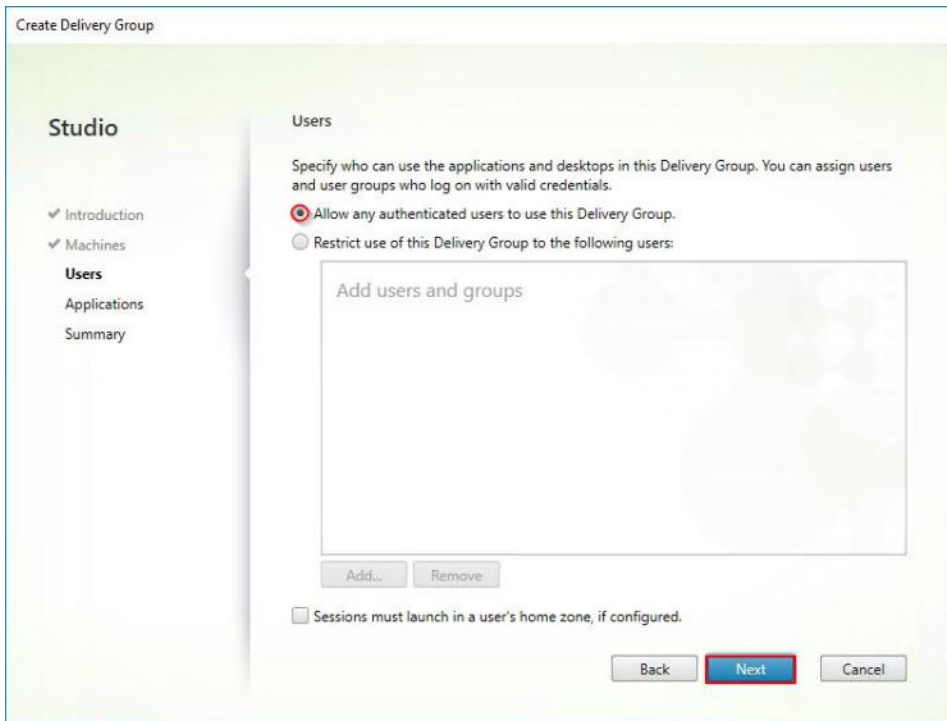
4. Select Machine catalog.

5. Provide the number of machines to be added to the delivery Group.

6. Click Next.



7. To make the Delivery Group accessible, you must add users, select Allow any authenticated users to use this Delivery Group.
8. Click Next.

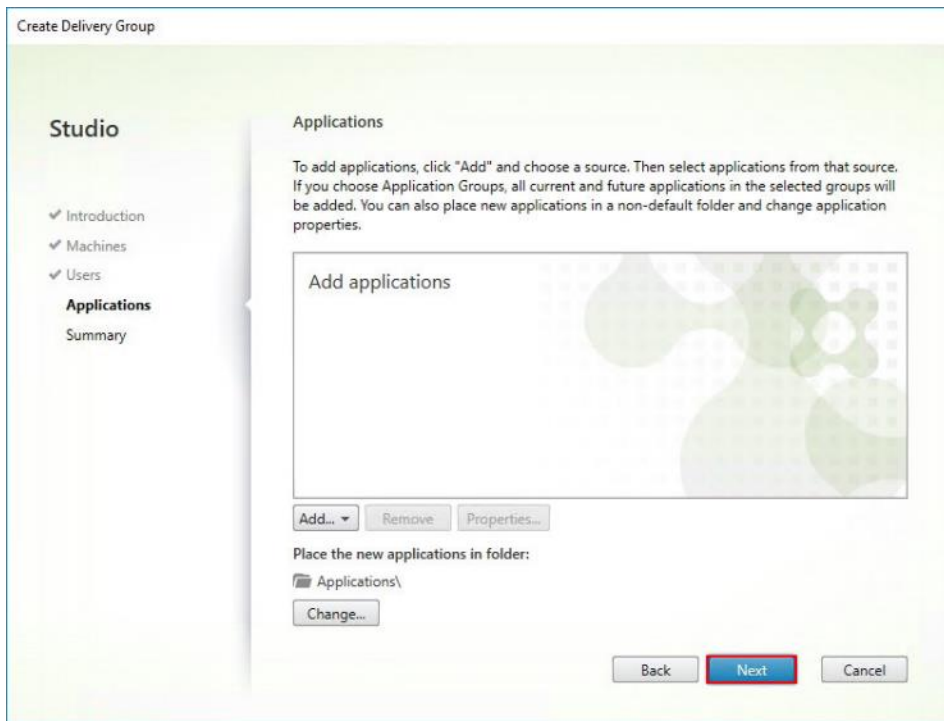




User assignment can be updated any time after Delivery group creation by accessing Delivery group properties in Desktop Studio.

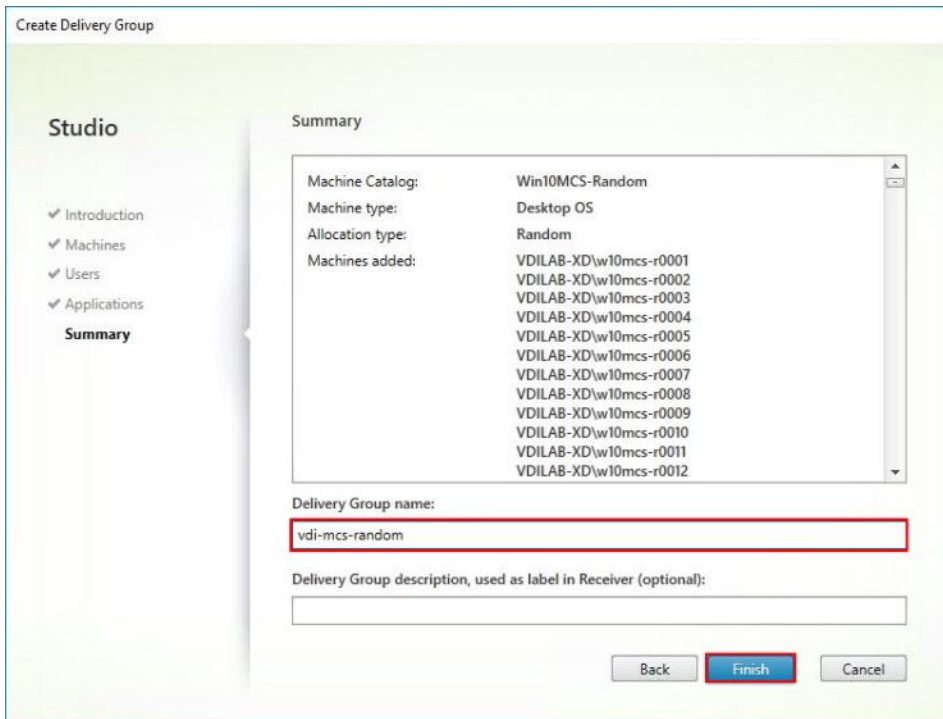
9. (Optional) Specify Applications catalog will deliver.

10. Click Next.

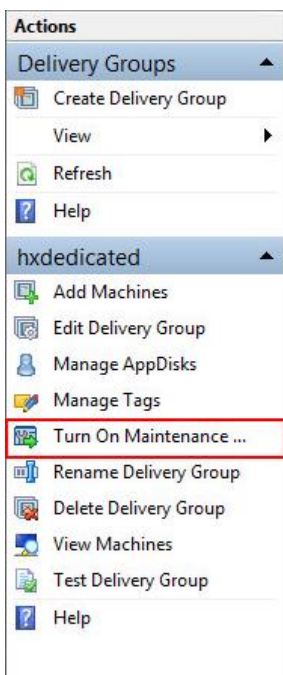


11. On the Summary dialog, review the configuration. Enter a Delivery Group name and a Display name (for example, HVD or HSD).

12. Click Finish.



13. Citrix Studio lists the created Delivery Groups and the type, number of machines created, sessions, and applications for each group in the Delivery Groups tab. Select Delivery Group and in Action List, select Turn on Maintenance Mode.



Citrix Virtual Desktops Policies and Profile Management

Policies and profiles allow the Citrix Virtual Desktops environment to be easily and efficiently customized.

Configure Citrix Virtual Desktops Policies

Citrix Virtual Desktops policies control user access and session environments, and are the most efficient method of controlling connection, security, and bandwidth settings. You can create policies for specific groups of users, devices, or connection types with each policy. Policies can contain multiple settings and are typically defined through Citrix Studio. (The Windows Group Policy Management Console can also be used if the network environment includes Microsoft Active Directory and permissions are set for managing Group Policy Objects). [Figure 25](#) shows policies for Login VSI testing in this CVD.

Figure 25. Virtual Desktops Policy

Policies		Testing Policy	
		Overview	Settings
1	Unfiltered		
2	Testing Policy		<ul style="list-style-type: none">▶ Auto connect client drives User setting - ICA\File Redirection Disabled (Default: Enabled)▶ Auto-create client printers User setting - ICA\Printing\Client Printers Do not auto-create client printers (Default: Auto-create all client printers)▶ Client printer redirection User setting - ICA\Printing Prohibited (Default: Allowed)▶ Concurrent logons tolerance Computer setting - Load Management Value: 4 (Default: Value: 2)▶ CPU usage Computer setting - Load Management Disabled (Default: Disabled)▶ CPU usage excluded process priority Computer setting - Load Management Disabled (Default: Below Normal or Low)▶ Flash default behavior User setting - ICA\Adobe Flash Delivery\Flash Redirection Disable Flash acceleration (Default: Enable Flash acceleration)▶ Memory usage Computer setting - Load Management Disabled (Default: Disabled)▶ Memory usage base load Computer setting - Load Management Disabled (Default: Zero load: 768 MBs)
3	VDI Policy		
4	RDS Policy		

Configure User Profile Management

Profile management provides an easy, reliable, and high-performance way to manage user personalization settings in virtualized or physical Windows environments. It requires minimal infrastructure and administration and provides users with fast logons and logoffs. A Windows user profile is a collection of folders, files, registry settings, and configuration settings that define the environment for a user who logs on with a particular user account. These settings may be customizable by the user, depending on the administrative configuration. Examples of settings that can be customized are:

- Desktop settings such as wallpaper and screen saver
- Shortcuts and Start menu setting
- Internet Explorer Favorites and Home Page

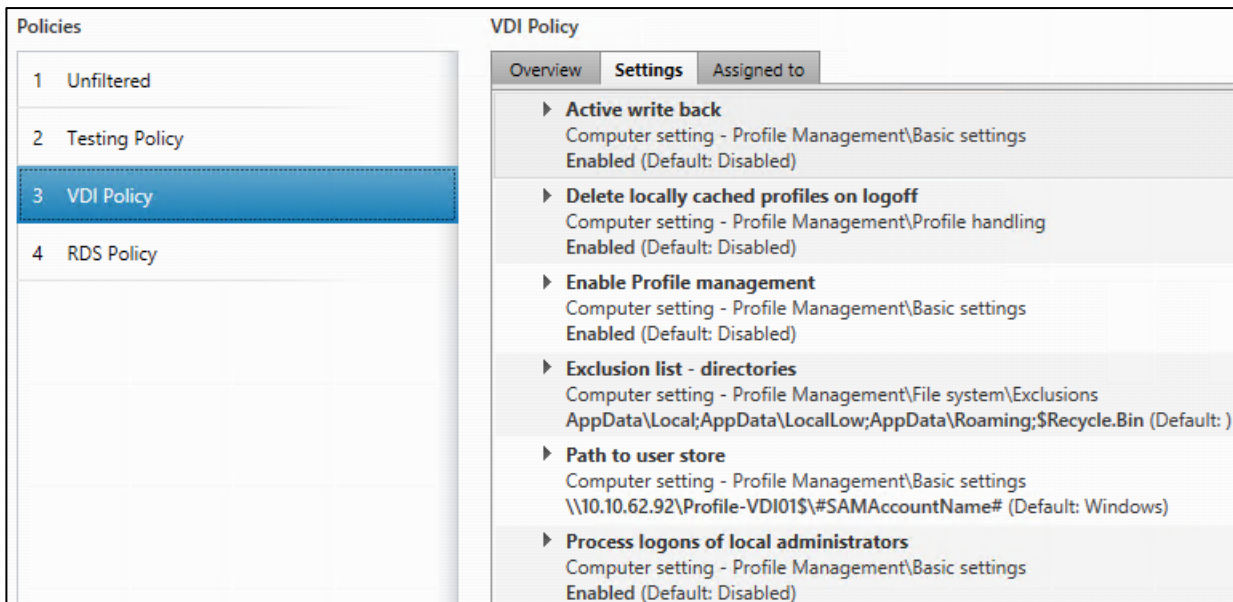
- Microsoft Outlook signature
- Printers

Some user settings and data can be redirected by means of folder redirection. However, if folder redirection is not used these settings are stored within the user profile.

The first stage in planning a profile management deployment is to decide on a set of policy settings that together form a suitable configuration for your environment and users. The automatic configuration feature simplifies some of this decision-making for Virtual Desktops deployments. Screenshots of the User Profile Management interfaces that establish policies for this CVD’s RDS and VDI users (for testing purposes) are shown below.

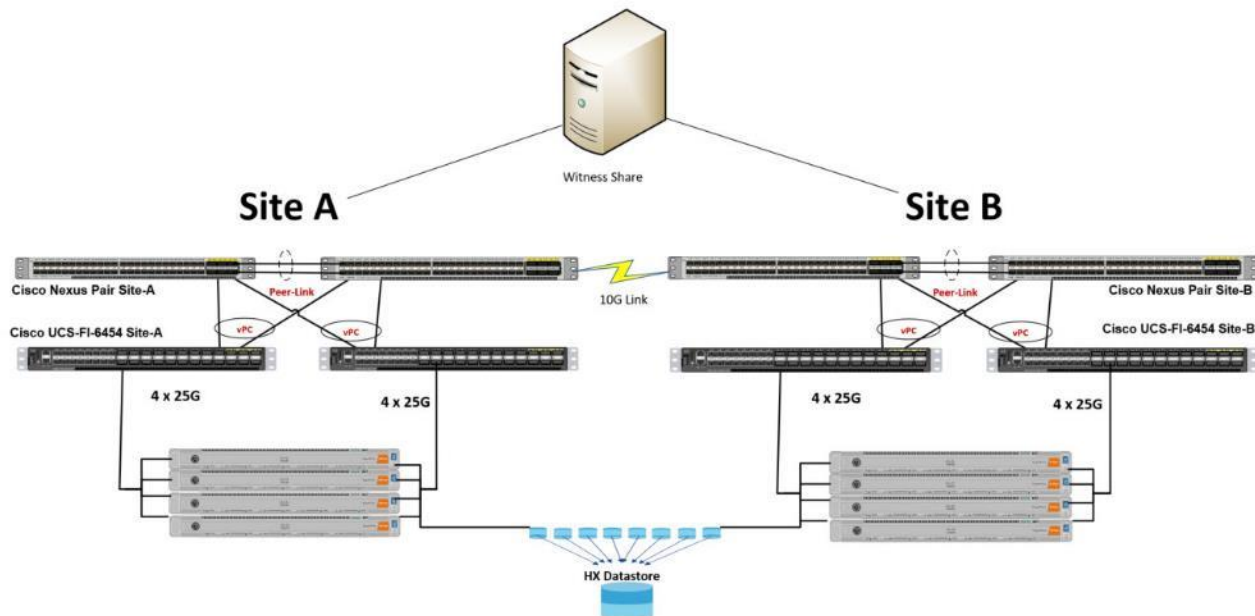
Basic profile management policy settings are documented here: <https://docs.citrix.com/en-us/citrix-virtual-apps-desktops>

Figure 26. VDI User Profile Manager Policy



Test Setup and Configurations

In this project, we tested a single Cisco HyperFlex stretch cluster running four Cisco UCS HXAF220C-M5SX Rack Servers in two separate Cisco UCS domains for a combined 8 node HX cluster. This solution is tested to illustrate linear scalability for each workload studied. We tested failover and resiliency of the stretch cluster, as well as VDI performance.



Hardware Components:

- 4 x Cisco UCS 6454 Fabric Interconnects
- 4 x Cisco Nexus 93108YCPX Access Switches
- 8 x Cisco UCS HXAF220c-M5SX Rack Servers (2 Intel Xeon Gold 6230 scalable family processor at 2.3 GHz, with 768 GB of memory per server [32 GB x 24 DIMMs at 2666 MHz])
- Cisco VIC 1457 mLOM
- 12G modular SAS HBA Controller
- 240GB M.2 SATA SSD drive (Boot and HyperFlex Data Platform controller virtual machine)
- 240GB 2.5" 6G SATA SSD drive (Housekeeping)
- 400GB 2.5" 6G SAS SSD drive (Cache)
- 8 x 960GB 2.5" SATA SSD drive (Capacity)
- 1 x 32GB mSD card (Upgrades temporary cache)

Software Components:

- Cisco UCS firmware 4.0(4g)

-
- Cisco HyperFlex Data Platform 4.0.2b
 - Citrix Virtual Desktops 1912 LTSR
 - Citrix User Profile Management
 - Microsoft SQL Server 2016
 - Microsoft Windows 10
 - Microsoft Windows 2019
 - Microsoft Office 2016
 - Login VSI 4.1.40

Test Methodology and Success Criteria

All validation testing was conducted on-site within the Cisco labs in San Jose, California.

Along with regular performance testing for VDI workloads, we also tested disaster recovery and failover functionality of a stretched cluster environment.

The testing results focused on the entire process of the virtual desktop lifecycle by capturing metrics during the desktop boot-up, user logon and virtual desktop acquisition (also referred to as ramp-up,) user workload execution (also referred to as steady state), and user logoff for the Hosted Shared Desktop Session under test.

Test metrics were gathered from the virtual desktop, storage, and load generation software to assess the overall success of an individual test cycle. Each test cycle was not considered passing unless all of the planned test users completed the ramp-up and steady state phases (described below) and unless all metrics were within the permissible thresholds as noted as success criteria.

Three successfully completed test cycles were conducted for each hardware configuration and results were found to be relatively consistent from one test to the next.

You can obtain additional information and a free test license from <http://www.loginvsi.com>.

Test Procedure

The following protocol was used for each test cycle in this study to insure consistent results.

Pre-Test Setup for Testing

Windows 10 virtual machines for VDI and Windows 2019 Server for RDS were deployed on both sites of the stretch cluster. Fifty percent of the total number on each site. All machines were shut down utilizing the Citrix Virtual Desktops 1912 LTSR Administrator Console.

All Launchers for the test were shut down. They were then restarted in groups of 10 each minute until the required number of launchers was running with the Login VSI Agent at a “waiting for test to start” state.

Test Run Protocol

To simulate severe, real-world environments, Cisco requires the log-on and start-work sequence, known as Ramp Up, to complete in 48 minutes. Additionally, we require all sessions started, whether 60 single server users or 500 full scale test users to become active within two minutes after the last session is launched.

In addition, Cisco requires that the Login VSI Benchmark method is used for all single server and scale testing. This assures that our tests represent real-world scenarios. For each of the three consecutive runs on single server tests, the same process was followed. To run the test protocol, follow these steps:

1. Time 0:00:00 Start esxtop Logging on the following systems:
 - a. Infrastructure and VDI Host Blades used in test run
 - b. All Infrastructure virtual machines used in test run (AD, SQL, Citrix Connection brokers, image mgmt., and so on)
2. Time 0:00:10 Start Storage Partner Performance Logging on Storage System.

-
3. Time 0:05: Boot VDI Machines using Citrix Virtual Desktops 1912 LTSR Administrator Console.
 4. Time 0:06 First machines boot.
 5. Time 0:35 Single Server or Scale target number of VDI Servers registered on Desktop Studio.



No more than 60 Minutes of rest time is allowed after the last desktop is registered and available on Citrix Virtual Desktops 1912 LTSR Administrator Console dashboard. Typically, a 20-30 minute rest period for Windows 10 desktops and 10 minutes for RDS virtual machines is sufficient.

6. Time 1:35 Start Login VSI 4.1.40 Knowledge Worker Benchmark Mode Test, setting auto-logoff time at 900 seconds, with Single Server or Scale target number of desktop virtual machines utilizing sufficient number of Launchers (at 20-25 sessions/Launcher).
7. Time 2:23 Single Server or Scale target number of desktop virtual machines desktops launched (48-minute benchmark launch rate).
8. Time 2:25 All launched sessions must become active.



All sessions launched must become active for a valid test run within this window.

9. Time 2:40 Login VSI Test Ends (based on Auto Logoff 900 Second period designated above).
10. Time 2:55 All active sessions logged off.
11. All sessions launched and active must be logged off for a valid test run. The Citrix Virtual Desktops 1912 LTSR Administrator Dashboard must show that all desktops have been returned to the registered/available state as evidence of this condition being met.
12. Time 2:57 All logging terminated; Test complete.
13. Time 3:15 Copy all log files off to archive; Set virtual desktops to maintenance mode through broker; Shut-down all Windows 7 machines.
14. Time 3:30 Reboot all hypervisors.
15. Time 3:45 Ready for new test sequence.

Stretch Cluster Failure Modes

One of the main precautions required for using a stretch or metropolitan (multisite) single cluster is the need to avoid a split-brain scenario. A split-brain condition indicates data or availability inconsistencies originating from the maintenance of two separate data sets with overlap in scope, either because of the loss of a site or a failure condition based on servers not communicating and synchronizing their data with each other (site link loss). The witness exists to prevent this scenario, and it is discussed in the various failure modes presented here.

Because a stretch cluster is a single cluster, for most failure situations you can simply ask yourself: How would a single cluster with a replication factor of 2 behave here? It is when you experience site losses (or more than two

simultaneous node failures on a single site) that the behavior diverges from that of the single-location RF 2 cluster because you actually have the advantage of an effective RF4.

To appreciate the failover mechanics of a stretch cluster, take a closer look at ZooKeeper. Architecturally, a stretch cluster contains five instances of ZooKeeper: two at each site and one on the witness server. So, in total there is one master ZK node and 4 followers. Only a storage node can be ZK nodes. Compute-only nodes will never be created with a ZK instance. The function of ZooKeeper is to maintain the cluster membership and a consistent cluster-wide file system configuration. So, if there are eight nodes at each site (a 16-node cluster), there will still be two ZooKeeper instances running on two nodes at each site and one more on the witness server.

Whenever a failure occurs, at least three ZooKeeper instances must be present to re-create the cluster membership and help ensure a consistent file system configuration. ZooKeeper achieves this behavior by using its built-in voting algorithm (based on the well-known Paxos algorithm).

If the witness goes down, then one ZooKeeper instance is lost. However, four more ZooKeeper instances are still running (2 at each site), which is more than the minimum of three ZooKeeper instances needed. Hence, the cluster will not be affected (no virtual machine failover or internal I/O hand-off occurs).

If a site goes offline, two ZooKeeper instances will go down. However, three more ZooKeeper instances are still running, which again is greater than or equal to the minimum of three ZooKeeper instances required. Hence, the cluster will not be affected. Virtual machines will automatically failover to the surviving site because of the presence of VMware HA. This failure will be treated as if half (minus one since the witness is still online) the number of nodes are lost in a single cluster.

If a ZK node at a site goes down that was hosting the ZooKeeper master, the ZooKeeper algorithm will elect another ZK node to be promoted to ZooKeeper master. The promotion of another ZK node happens only if the failed node is a master ZK node and the failover target is part of the ZK ensemble (which is always the case for a ZK master). If the failed node was a ZK follower (i.e. a stand-alone ZK node), then no election occurs, and you are running with one less ZK instance. Either way, four more ZooKeeper instances are still running, which is more than the minimum of three ZooKeeper instances required. The site and the cluster will still be online. Only the affected virtual machines will be restarted on the surviving nodes at the same site (with stretch cluster DRS rules managing the movement). The failure will be treated like a node lost in a single cluster.

It is expected that you will recover or replace any failed nodes. New ZK nodes will not be automatically created. If those nodes happen to be ZK nodes, then there is a manual process to reassign ZK membership if the node needs to be completely replaced. See your support representative for assistance. A recovered node will simply resume its previous role unless it was master (since a new master is now elected) in which case it will join as a follower.

Survivability while maintaining online status after node losses requires a majority zookeeper quorum and more than 50% of any nodes (the witness counts as an active zookeeper node). If one site has suffered multiple losses, it is possible that the surviving site could tolerate a node or disk loss (in a cluster greater than 2+2) if that node is not a zookeeper node, but it is not guaranteed.

Zookeeper has a notable dependence on NTP from the nodes to maintain cluster synchronization. The allowable ZK time drift between nodes is 300ms. If the skew exceeds this the cluster is subject to ZK errors and may not function properly. It is advisable to monitor NTP skew between CVMs using the HX APIs and alert on time drift issues.

Recovery of ZK Nodes that have Failed

As mentioned above, there are two types of ZK nodes: a single ensemble master and 4 followers. Zookeeper nodes that fail have to undergo a special process to be replaced. If they are recovered, then they resume their previous role unless they were the master since a new master is now in place. This node will become an ensemble follower.

- If the master ZK node fails, zookeeper will automatically elect a new master from the remaining ZK follower nodes. This will leave you with a cluster having 4 ZK instances.
- If a follower fails, a new follower is not created on rebuild. If the node that failed is recovered, your ZK instance will return, and you will be back to normal (5 ZK instances). If you are unable to recover the node, the manual node-remove workflow and node replacement will result in a new follower being created. Contact support for assistance.

Types of Failures

The types of failures and the responses to each are summarized here:

- Disk loss
 - Cache disk: This failure is treated the same way as in a normal cluster. Other cache disks in the site service requests, and overall cache capacity is reduced until the failed component is replaced.
 - Persistent disk: This failure is treated the same way as in a normal cluster. After a 2-minute timeout interval, the data from the failed disk is rebuilt using the remaining capacity.
- Node loss
 - 1x: The site will rebuild the failed node after a 2-hour timeout or earlier through manual intervention.
 - Nx: If the node losses are simultaneous and are not all the nodes in a site (e.g., lose 3 nodes out of 5 on a site), the site will remain online, and site failover will not occur. For example, if you have a 3+3 cluster, and you lose 2 nodes on site 1 (regardless of ZK type), then site 1 will still be active, VMs will migrate to the surviving node and the site will still function. There may not be enough resources on the surviving 1 node to restart all the VMs from the 2 failed nodes on the site. In that case, since the host affinity rules are “should run” and not “must run”, DRS will restart the VMs that exceed the site capacity at the other site.
- Fabric interconnect loss
 - 1x: The redundant fabric interconnect at the site will handle data until its partner is recovered.
 - 2x: The site will be offline, and site failover will occur.
- Witness loss
 - Nothing happens; the cluster is not affected. Bring the witness back online after it is repaired.
 - Since the Witness is a ZK node, you are guaranteed to survive one node failure at either site (since in worst case that failed node will be ZK as well). This leaves 3 ZK's left. You cannot be guaranteed another failure because if that is a ZK node as well, you no longer have majority ZK surviving. You ***may*** survive these failures if you get lucky with no additional ZK node failures, but you are not guaranteed this condition. Only worst-case survivability is reported.
- Accidental deletion of the witness virtual machine

- Restore from backup with identical networking. The cluster will discover the witness and resynchronize.
- Contact the Cisco TAC for a recovery process otherwise.
- Switch loss (single site)
 - 1x: For redundant switches at a site, the partner switch will handle data until the failure is repaired. If there is a single uplink switch per site, site failover will occur.
 - 2x: The site will be offline, and site failover will occur.
- Site loss
 - The site will be offline, and site failover will occur.
- Site link loss

For a scenario in which a fault occurs in the network between the two sites (a cable is damaged, a network port on either site fails, etc.) but the nodes on the two sites are still alive, the following process is implemented:

- When a stretch cluster is created, one site is biased to establish a ZooKeeper master. This is done by assigning a higher node ID. For the purpose of this discussion, the quorum site is site A.
- When the network disconnect occurs, the witness and the nodes of the site that have the ZooKeeper master form the quorum.
- The nodes at the other site (site B) will still stay powered on, and I/O operations from the local IO Visor instance from this site (site B) will not be able to perform write I/O operations successfully, which this will guarantee the isolated site's consistency. The `stcli cluster-info` command will show these nodes as unavailable in the cluster, even though physically they may be powered on.
- Because site A is the ZooKeeper quorum site, the updates to ZooKeeper will eventually (after a failure-detection timeout) be visible to site B. Eventually, the IO Visor on ESX at site B will see that it needs to talk to a different node, which is the actual I/O primary node (which is in the ZooKeeper quorum at site A). Because there is no network connection, site B will keep retrying those I/O operations and will eventually see "All Paths Down" (APD), assuming that there are still user virtual machines on this site (site B). Your intervention should verify that eventually no virtual machines remain on this site (because they have been failed over to other ESX hosts).
- Virtual machines fail over to the site having the ZooKeeper leader. VMware HA and DRS are responsible for the failover of virtual machines.
- If the network is restored, the nodes of site B that were fenced out will become available again in the cluster. Automatic resynchronization between the sites should occur. However, virtual machine failback is not automatic.

Scenario Walk Through - Failure of Multiple Nodes in a Site

If you have an 8-node cluster (4 nodes on each site) and you lose 2 nodes on site A, what happens to the remaining nodes? Do the VMs still continue to run on site A's remaining nodes?

In this scenario, the remaining nodes on site A restart the VMs that were running on the failed nodes. The site is still online and serving data, however, since the site is locally RF2 (globally RF4) you will have lost some part of the distributed local primary write logs that were running on the 2 nodes that failed. You will also have lost some local persistent data. HX will recognize this and switch the primaries over to Site B. This will incur a read penalty

for these VMs. Note also, depending on how heavily loaded the system is, the surviving nodes on Site A may be either at capacity or unable to restart all of the VMs. In that case, DRS can ignore the affinity rules (HX uses “should” rules for affinity not “must” rules) and restart the VMs on available resources in Site B. If there is capacity in Site A, rebuild will begin and attempt to reestablish local RF2.

This behavior is the same if you have a 10+10 cluster and lose, say, 5 nodes on Site A.

If you were to suffer multiple node failure at each site simultaneously, that would constitute a catastrophic loss and your cluster would be offline pending recovery.

Scenario Walk Through - Failure of a Site

In the event of site failover, operations should continue as intended after the virtual machines from the failed site boot on the surviving site. Virtual machine and IO Visor behavior is as described for site link loss in the preceding discussion. For example:

If you have an 8-node cluster (4 nodes on each site) and you lose site A, vCenter HA initiates a restart all of Site A’s VMs on nodes in site B. Failback is not automatic. If sized at maximum capacity (50% per site) then Site B is now running at maximum capacity (100%).

After your downed site has been recovered and communications with the remaining site and the witness have been reestablished, you can move your virtual machines back to their original compute resource (based on site affinity) at the recovered site. Use vMotion for this process so that affinity and proper IO Visor routing occurs after the virtual machines are moved back to their preferred locations. Storage vMotion is not required, since the datastore is mounted on all nodes. Only a migration of the compute resource is needed to re-establish site storage affinity and compute resource parity. In short, during a site failover the affinity rule will not change so that you can quickly migrate back once the impacted site is recovered. However, if you manually Storage vMotion (SVMotion) VMs around outside of a failover event, the affinity rules will automatically be updated to reflect residence in the datastore with the correct rules.

Failure Response Summary

[Table 10](#) lists the failure modes discussed previously, with some additional information for particular situations. Note that double, separate catastrophic failures are not considered here (for example, both site loss and witness loss) because such failures always result in a cluster offline status.

Table 10. Failure Responses

Component Failure	Cluster Behavior	Quorum Update	Virtual Machine Restart	Site Status	Cluster Status
Single site single cache disk	Site is online, with diminished cache capacity.	No	No	Online	Online
Single site single persistent disk	Site is online, with diminished capacity, and is rebuilt after 2 minutes using the remaining capacity.	No	No	Online	Online
Single site double cache	Site is online, with diminished cache capacity.	No	No	Online	Online

Component Failure	Cluster Behavior	Quorum Update	Virtual Machine Restart	Site Status	Cluster Status
disk					
Single site double persistent disk	If failure is simultaneous and on different nodes. The site is still online but some VMs will switch primary write logs to the opposite site.	No	No	Online	Online
Single site single node loss	If the failure is not simultaneous on different nodes at different times, then the cluster behaves as with a single-disk failure with reduced capacity.	No	No	Online	Online
Single site multiple node loss	If the failure is simultaneous on the same node.	No	No	Online	Online
Single site single fabric interconnect loss	Node is rebuilt after 2 hours or through manual intervention.	Maybe	Yes	Online	Online
Single site double fabric interconnect loss	Site is online and will restart VMs on surviving nodes in the site.	Maybe	Yes	Online	Online
Double site single fabric interconnect loss	No impact on the site; recover the fabric interconnect.	No	No	Online	Online
Double site double fabric interconnect loss	Site is offline.	Yes	Yes	Offline	Online
Witness loss	No impact on the site; recover the fabric interconnects.	No	No	Online	Online
Single site single switch loss	Both sites are offline.	No	No	Offline	Offline
Single site double switch loss	No impact on the site; recover the witness.	No	No	Online	Online
Double site single switch loss	If redundant switching exists at the site, there is no impact; recover the switch.	No	No	Online	Online

Component Failure	Cluster Behavior	Quorum Update	Virtual Machine Restart	Site Status	Cluster Status
Double site double switch loss	If the site has only a single switch, site is offline.	Yes	Yes	Offline	Online
Site loss	Site is offline.	Yes	Yes	Offline	Online
Site link loss - sites still online and witness is reachable, but site-to-site link is down	If redundant switching exists at the sites, there is no impact; recover the switches.	No	No	Online	Offline
Site loss and disk or node loss on the surviving site in a 2+2 node cluster	If the sites have only a single switch, the sites are offline.	No	No	Offline	Offline
Site loss and disk or node loss on the surviving site in a n+n node cluster	Both sites are offline.	No	No	Offline	Online
Site Loss and witness loss (connectivity, VM failure and so on)	ZooKeeper instance maintains information about cluster groups and forms the quorum. When a site is lost, ZooKeeper communications disappear, site fencing is enforced, and the cluster quorum is redefined. ZooKeeper with DRS rules (affinity, groups, and so on) makes sure that the same virtual machine is never running on both sites simultaneously.	Yes	Yes	Offline	Online

Witness Failure and Restore from Backup

To increase the resiliency of a stretch cluster deployment, many users will back up their witness VM. If the witness were to fail, you can restore from back up (retaining identical network settings). The witness ZK instance will be stale but will re-synchronize with the surviving site(s). As mentioned in the failure scenario section, in the event that the witness fails after a site goes offline and the cluster fails over, the system will be offline. If the witness is subsequently restored from backup and communication is available between witness and either site, the cluster will synchronize zookeeper and come back online.

It is also possible in this scenario to maintain a cold witness stand-by VM and promote it when needed. In order to properly integrate with the cluster, it will need to be an identical copy of the original witness and will have to retain the same networking. It will synchronize with the cluster when brought online.

Failure Response Times

Failure response times for disk loss are near-instantaneous. It is the same for active/passive standby links to the FIs. Node failures in a site are they typical node timeout values (approximately 17 seconds). For a site failover to occur, the timeout must happen for multiple nodes at the same time. Since connectivity loss to a site typically involves multiple simultaneous losses, the site time out is at best the same as a node time out. This value can increase due to other factors, such as heavy workloads (affecting ZK updates), latency to the witness, and inter-site link latency.

Failure Capacity Considerations and Example

Failure of a node in a site reduces the overall capacity by the free space on the lost node symmetrically; that is, times two. In other words, the cluster capacity in general equals twice the minimum site cluster capacity at the cluster RF (4). This can be expressed as $2[\min(\text{site A capacity}, \text{site B capacity})]/4$.

Take the following example for capacity after node loss on a 5+5 stretch cluster:

siteA/siteB

5 nodes/5 nodes = 10 nodes

At 10TB/node useable that is 100 TB usable total

50TB usable/ 50 TB usable per site

RF 2+2 usable / RF 2+2 usable data protection (RF 4 equivalent)

25TB usable after data protection (100 TB/4)

50 TB/4=12.5TB per site usable.

If site A loses a node, it has dropped by 20% capacity from 12.5 TB to 10 TB. Since the total cluster usable capacity is defined as $2[\min(\text{site A capacity}, \text{site B capacity})]/4$. Site A is now the cluster site minimum at 10 TB (site B is still at 12.5TB) which means the total cluster capacity is now $2(10\text{TB})/4 = 20$ TB usable.

In the event of a site failure, the surviving site capacity is the total cluster capacity divided by two, however the remaining capacity is filled in a RF 2 manner until the failed site is recovered so the total free capacity remains constant before and after a site failure. Once the failed site is recovered changes made since the failure are synchronized across the cluster and RF 4 is re-established. Capacity reporting during this transitional interval (site loss, surviving site stabilization, surviving site production usage, failed site recovery) is in flux. The reported capacity will be variable as things like the file system cleaner, rebuilds, and transitions to temporary RF 2 for all VMs take place. In the steady state where the failed site is not recovered in a timely fashion, the surviving capacity will approach RF 2 for the free capacity that was available on the surviving site before the secondary site failure occurred

Success Criteria

Our “pass” criteria for this testing is as follows: Cisco will run tests at a session count levels that effectively utilize the server capacity measured by CPU, memory, storage, and network utilization. We use Login VSI version 4.1.25 to launch Knowledge Worker workload sessions. The number of launched sessions must equal active sessions within two minutes of the last session launched in a test as observed on the VSI Management console.

The Citrix Virtual Desktops Studio will be monitored throughout the steady state to make sure of the following:

- All running sessions report In Use throughout the steady state
- No sessions move to unregistered, unavailable, or available state at any time during steady state

Within 20 minutes of the end of the test, all sessions on all launchers must have logged out automatically and the Login VSI Agent must have shut down. Cisco’s tolerance for Stuck Sessions is 0.5 percent (half of one percent.) If the Stuck Session count exceeds that value, we identify it as a test failure condition.

Cisco requires three consecutive runs with results within +/-1 percent variability to pass the Cisco Validated Design performance criteria. For white papers written by partners, two consecutive runs within +/-1 percent variability are accepted. (All test data from partner run testing must be supplied along with proposed white paper.)

We will publish Cisco Validated Designs with our recommended workload following the process above and will note that we did not reach a VSImax dynamic in our testing.

The purpose of this testing is to provide the data needed to validate Citrix Virtual Desktops 1912 LTSR Hosted Shared Desktop with Citrix Virtual Desktops 1912 LTSR Composer provisioning using Microsoft Windows Server 2016 sessions on Cisco UCS HXAF220c-M4S, Cisco UCS 220 M4 and Cisco UCS B200 M4 servers.

The information contained in this section provides data points that a customer may reference in designing their own implementations. These validation results are an example of what is possible under the specific environment conditions outlined here and do not represent the full characterization of Citrix and Microsoft products.

Four test sequences, each containing three consecutive test runs generating the same result, were performed to establish system performance and linear scalability.

All of these standard Login VSI CVD Testing results for VDI will be evaluated against each failure scenario in the Stretch Cluster. Each test should pass in each failure scenario for this to be considered a successful test.

VSImax 4.1.x Description

The philosophy behind Login VSI is different to conventional benchmarks. In general, most system benchmarks are steady state benchmarks. These benchmarks execute one or multiple processes, and the measured execution time is the outcome of the test. Simply put: the faster the execution time or the bigger the throughput, the faster the system is according to the benchmark.

Login VSI is different in approach. Login VSI is not primarily designed to be a steady state benchmark (however, if needed, Login VSI can act like one). Login VSI was designed to perform benchmarks for SBC or VDI workloads through system saturation. Login VSI loads the system with simulated user workloads using well known desktop applications like Microsoft Office, Internet Explorer, and Adobe PDF reader. By gradually increasing the number of simulated users, the system will eventually be saturated. Once the system is saturated, the response time of the applications will increase significantly. This latency in application response times show a clear indication

whether the system is (close to being) overloaded. As a result, by nearly overloading a system it is possible to find out what is its true maximum user capacity.

After a test is performed, the response times can be analyzed to calculate the maximum active session/desktop capacity. Within Login VSI this is calculated as VSI_{max}. When the system is coming closer to its saturation point, response times will rise. When reviewing the average response time, it will be clear the response times escalate at saturation point.

This VSI_{max} is the “Virtual Session Index (VSI)”. With Virtual Desktop Infrastructure (VDI) and Terminal Services (RDS) workloads this is valid and useful information. This index simplifies comparisons and makes it possible to understand the true impact of configuration changes on hypervisor host or guest level.

Server-Side Response Time Measurements

It is important to understand why specific Login VSI design choices have been made. An important design choice is to execute the workload directly on the target system within the session instead of using remote sessions. The scripts simulating the workloads are performed by an engine that executes workload scripts on every target system and are initiated at logon within the simulated user’s desktop session context.

An alternative to the Login VSI method would be to generate user actions client side through the remoting protocol. These methods are always specific to a product and vendor dependent. More importantly, some protocols simply do not have a method to script user actions client side.

For Login VSI the choice has been made to execute the scripts completely server side. This is the only practical and platform independent solutions, for a benchmark like Login VSI.

Calculating VSI_{max} v4.1.x

The simulated desktop workload is scripted in a 48-minute loop when a simulated Login VSI user is logged on, performing generic Office worker activities. After the loop is finished it will restart automatically. Within each loop the response times of sixteen specific operations are measured in a regular interval: sixteen times in within each loop. The response times of these five operations are used to determine VSI_{max}.

The five operations from which the response times are measured are:

- Notepad File Open (NFO)

Loading and initiating VSINotepad.exe and opening the openfile dialog. This operation is handled by the OS and by the VSINotepad.exe itself through execution. This operation seems almost instant from an end-user’s point of view.

- Notepad Start Load (NSLD)

Loading and initiating VSINotepad.exe and opening a file. This operation is also handled by the OS and by the VSINotepad.exe itself through execution. This operation seems almost instant from an end-user’s point of view.

- Zip High Compression (ZHC)

This action copy’s a random file and compresses it (with 7zip) with high compression enabled. The compression will very briefly spike CPU and disk IO.

- Zip Low Compression (ZLC)

This action copy's a random file and compresses it (with 7zip) with low compression enabled. The compression will very briefly disk IO and creates some load on the CPU.

- CPU

Calculates a large array of random data and spikes the CPU for a short period of time.

These measured operations within Login VSI do hit considerably different subsystems such as CPU (user and kernel), Memory, Disk, the OS in general, the application itself, print, GDI, etc. These operations are specifically short by nature. When such operations become consistently long: the system is saturated because of excessive queuing on any kind of resource. As a result, the average response times will then escalate. This effect is clearly visible to end-users. If such operations consistently consume multiple seconds the user will regard the system as slow and unresponsive.

Figure 27. Sample of a VSI Max Response Time Graph, Representing a Normal Test

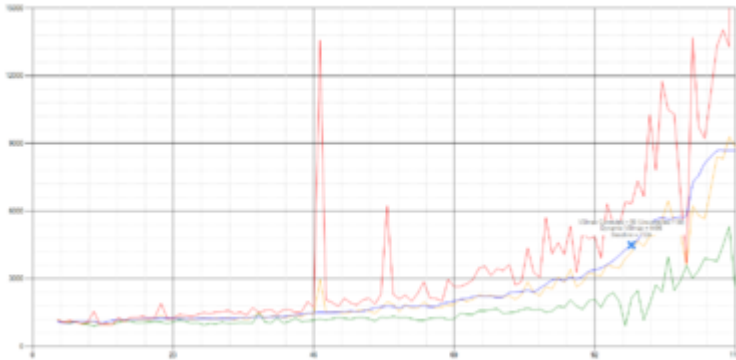
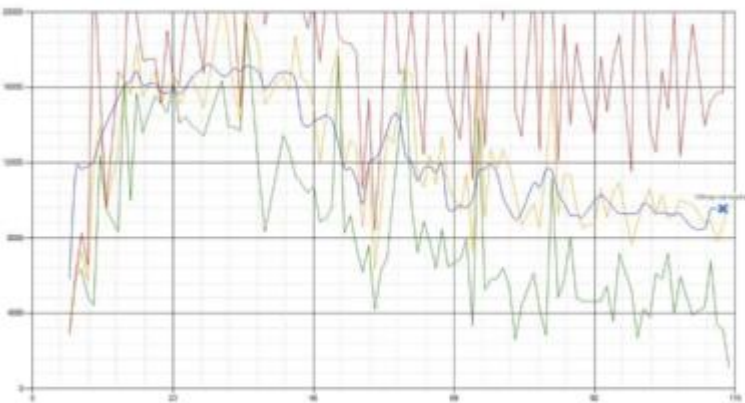


Figure 28. Sample of a VSI Test Response Time Graph with a Clear Performance Issue



When the test is finished, VSImax can be calculated. When the system is not saturated, and it could complete the full test without exceeding the average response time latency threshold, VSImax is not reached, and the number of sessions ran successfully.

The response times are very different per measurement type, for instance Zip with compression can be around 2800 ms, while the Zip action without compression can only take 75ms. This response time of these actions are

weighted before they are added to the total. This ensures that each activity has an equal impact on the total response time.

In comparison to previous VSImax models, this weighting much better represents system performance. All actions have very similar weight in the VSImax total. The following weighting of the response times are applied.

The following actions are part of the VSImax v4.1 calculation and are weighted as follows (US notation):

- Notepad File Open (NFO): 0.75
- Notepad Start Load (NSLD): 0.2
- Zip High Compression (ZHC): 0.125
- Zip Low Compression (ZLC): 0.2
- CPU: 0.75

This weighting is applied on the baseline and normal Login VSI response times.

With the introduction of Login VSI 4.1 we also created a new method to calculate the base phase of an environment. With the new workloads (Taskworker, Powerworker, and so on) enabling 'base phase' for a more reliable baseline has become obsolete. The calculation is explained below. In total 15 lowest VSI response time samples are taken from the entire test, the lowest 2 samples are removed, and the 13 remaining samples are averaged. The result is the Baseline. The calculation is as follows:

- Take the lowest 15 samples of the complete test
- From those 15 samples remove the lowest 2
- Average the 13 results that are left is the baseline

The VSImax average response time in Login VSI 4.1.x is calculated on the number of active users that are logged on the system.

Always a 5 Login VSI response time samples are averaged + 40 percent of the number of “active” sessions. For example, if the active sessions are 60, then latest 5 + 24 (=40 percent of 60) = 31 response time measurement are used for the average calculation.

To remove noise (accidental spikes) from the calculation, the top 5 percent, and bottom 5 percent of the VSI response time samples are removed from the average calculation, with a minimum of 1 top and 1 bottom sample. As a result, with 60 active users, the last 31 VSI response time sample are taken. From those 31 samples the top 2 samples are removed and lowest 2 results are removed (5 percent of 31 = 1.55, rounded to 2). At 60 users the average is then calculated over the 27 remaining results.

VSImax v4.1.x is reached when the VSIbase + a 1000 ms latency threshold is not reached by the average VSI response time result. Depending on the tested system, VSImax response time can grow 2 - 3x the baseline average. In end-user computing, a 3x increase in response time in comparison to the baseline is typically regarded as the maximum performance degradation to be considered acceptable.

In VSImax v4.1.x this latency threshold is fixed to 1000ms, this allows better and fairer comparisons between two different systems, especially when they have different baseline results. Ultimately, in VSImax v4.1.x, the

performance of the system is not decided by the total average response time, but by the latency it has under load. For all systems, this is now 1000ms (weighted).

The threshold for the total response time is average weighted baseline response time + 1000ms.

When the system has a weighted baseline response time average of 1500ms, the maximum average response time may not be greater than 2500ms (1500+1000). If the average baseline is 3000 the maximum average response time may not be greater than 4000ms (3000+1000).

When the threshold is not exceeded by the average VSI response time during the test, VSImax is not hit, and the number of sessions ran successfully. This approach is fundamentally different in comparison to previous VSImax methods, as it was always required to saturate the system beyond VSImax threshold.

Lastly, VSImax v4.1.x is now always reported with the average baseline VSI response time result. For example: "The VSImax v4.1 was 125 with a baseline of 1526ms". This helps considerably in the comparison of systems and gives a more complete understanding of the system. The baseline performance helps to understand the best performance the system can give to an individual user. VSImax indicates what the total user capacity is for the system. These two are not automatically connected and related:

When a server with a very fast dual core CPU, running at 3.6 GHZ, is compared to a 10 core CPU, running at 2,26 GHZ, the dual core machine will give an individual user better performance than the 10 core machine. This is indicated by the baseline VSI response time. The lower this score is, the better performance an individual user can expect.

However, the server with the slower 10 core CPU will easily have a larger capacity than the faster dual core system. This is indicated by VSImax v4.1.x, and the higher VSImax is, the larger overall user capacity can be expected.

With Login VSI 4.1.x a new VSImax method is introduced: VSImax v4.1. This methodology gives much better insight in system performance and scales to extremely large systems.

Test Results

Eight Node Cisco HXAF220c-M5 Stretch Cluster

For Citrix Virtual Apps & Desktops, the recommended maximum workload was determined based on both Login VSI Knowledge Worker workload end user experience measures and HXAF220c-M5SX server operating parameters.

This recommended maximum workload for stretch clustering allows you to determine the cluster N+1 fault tolerance load the sites can successfully support in the event of a server/site outage for maintenance or upgrade.

Our recommendation is that the Login VSI Average Response and VSI Index Average should not exceed the Baseline plus 2000 milliseconds to ensure that end-user experience is outstanding. Additionally, during steady state, the processor utilization should average no more than 90-95 percent.



Memory should never be oversubscribed for Desktop Virtualization workloads.

Test Phase	Description
Boot	Start all RDS and/or VDI virtual machines at the same time.
Login	The Login VSI phase of test is where sessions are launched and start executing the workload over a 48 minutes duration.
Steady state	The steady state phase is where all users are logged in and performing various workload tasks such as using Microsoft Office, Web browsing, PDF printing, playing videos, and compressing files.
Logoff	Sessions finish executing the Login VSI workload and logoff.



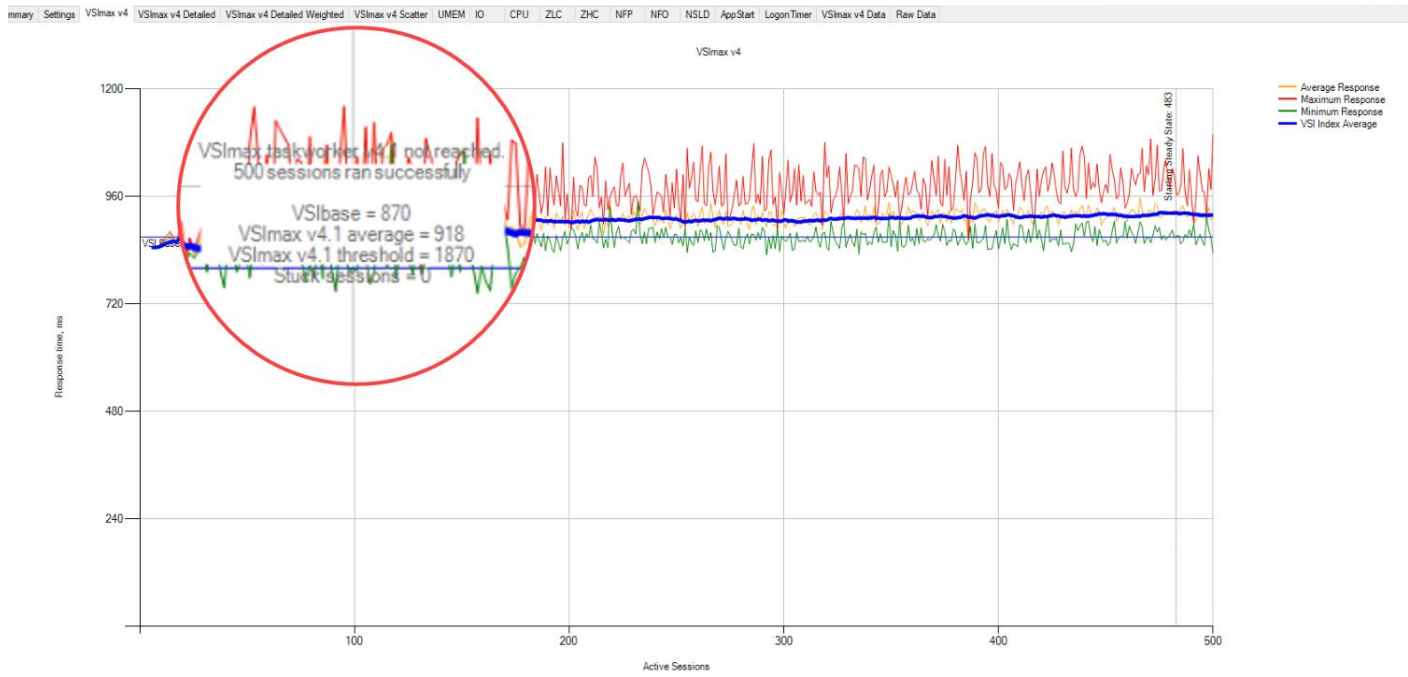
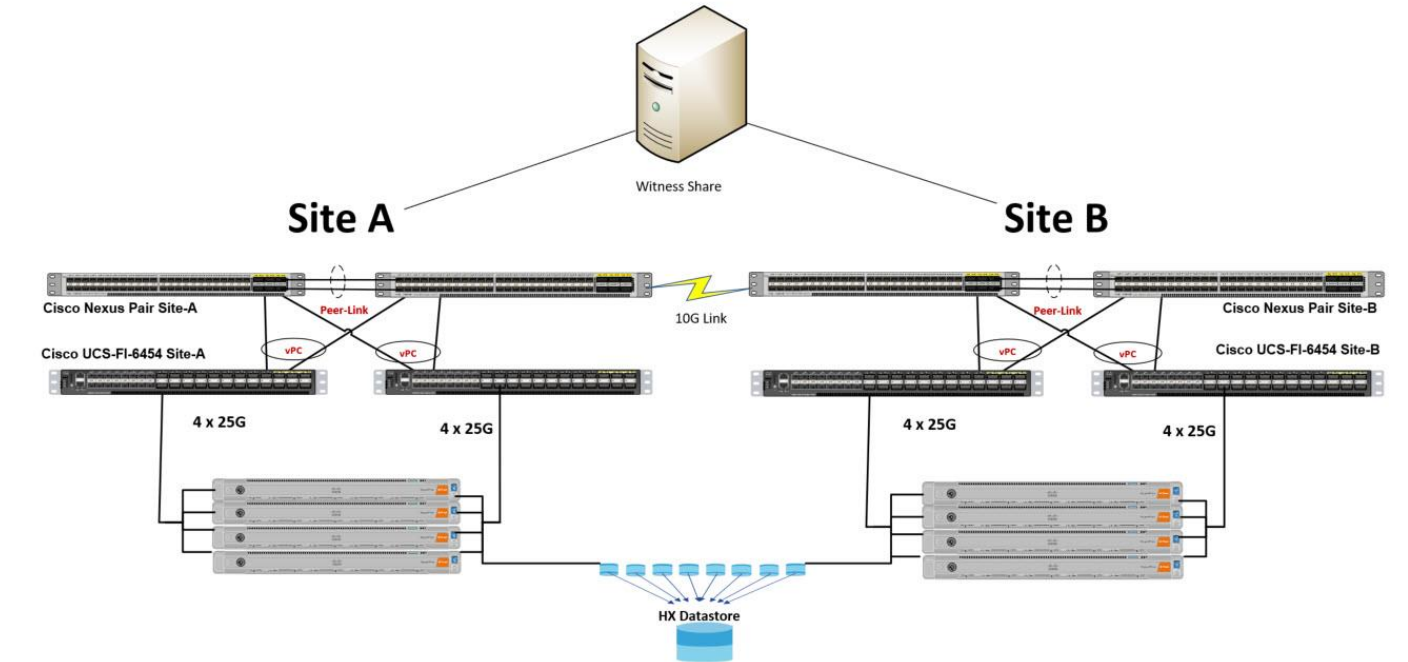
The recommended maximum workload for a Cisco HyperFlex cluster configured on Cisco HXAF220c-M5SX with Intel Xeon Gold 6230 scalable family processors and 768GB of RAM for Windows 10 desktops with Office 2016 is 500 virtual desktops.



The recommended maximum workload for a Cisco HyperFlex cluster configured on Cisco HXAF220c-M5SX with Intel Xeon Gold 6230 scalable family processors and 768GB of RAM for Windows Server 2019 RDS desktop sessions with Office 2016 is 500 virtual desktops.

500 Windows 10 VDI Workstations Testing on an 8 Node Cisco HyperFlex Stretch Cluster in Various Failure Scenarios

Figure 29. Test Results for 500 User Sessions with Both Sites Fully Functional



Summary Settings VSImax v4 VSImax v4 Detailed VSImax v4 Detailed Weighted VSImax v4 Scatter UMEM IO CPU ZLC ZHC NFP NFO NSLD A

SCC2-4

Successfully completed Login VSI test with **500 taskworker** sessions. VSImax (system saturation) was not reached. All Login VSI users completed the test.

Test result review

500 sessions were configured to be launched in **2880** seconds.

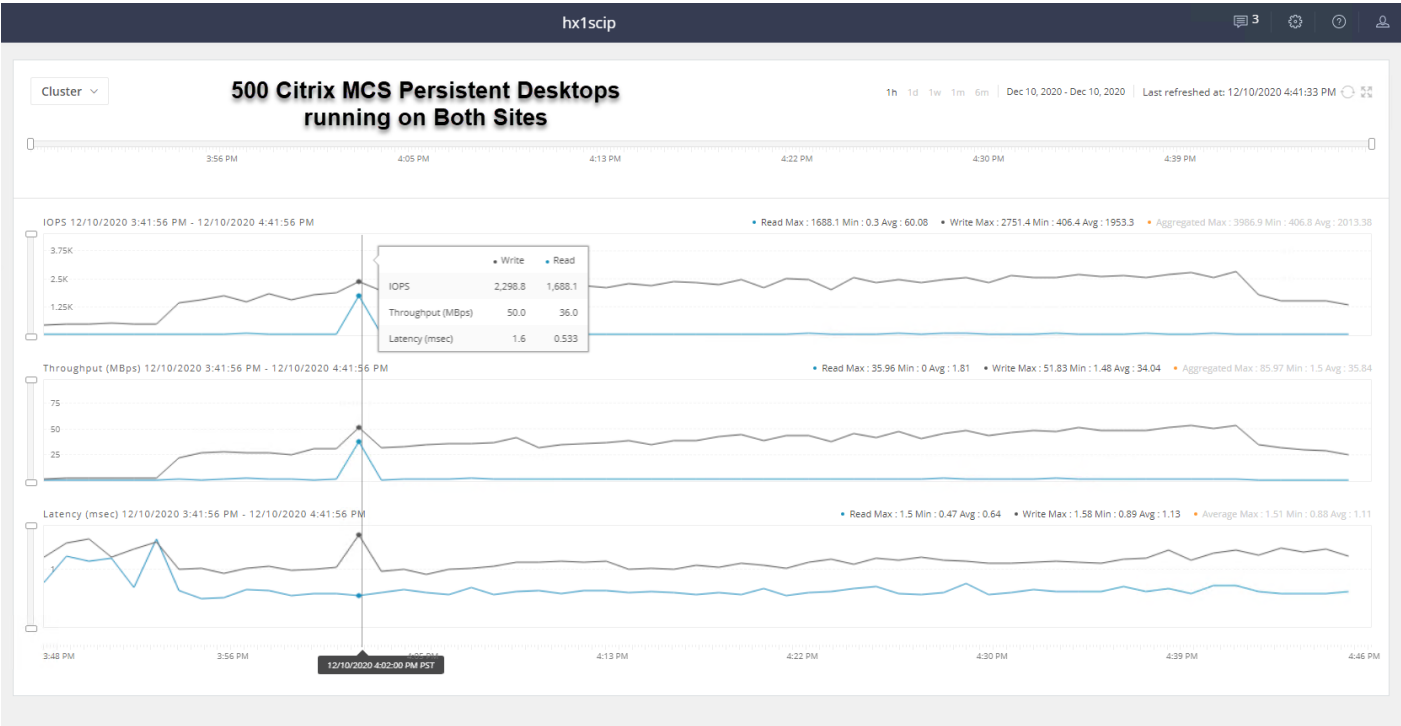
In total **0** sessions failed during the test:

- 0** sessions was/were not successfully launched
- 0** launched sessions failed to become active
- 500** sessions were active during the test
- 0** sessions got stuck during the test (before VSImax threshold)

With **500** sessions the maximum capacity VSImax (v4.1) **taskworker** was not reached with a Login VSI baseline performance score of **870**

Login VSI index average score is **958** lower than threshold. It might be possible to launch more sessions in this configuration.

Baseline performance of **870** is: **Good**



500 Windows 10 VDI Workstations Testing on an 8 Node Cisco HyperFlex Stretch Cluster in Various Failure Scenarios

Figure 30. Test Results for 500 User Sessions with a Witness Failure

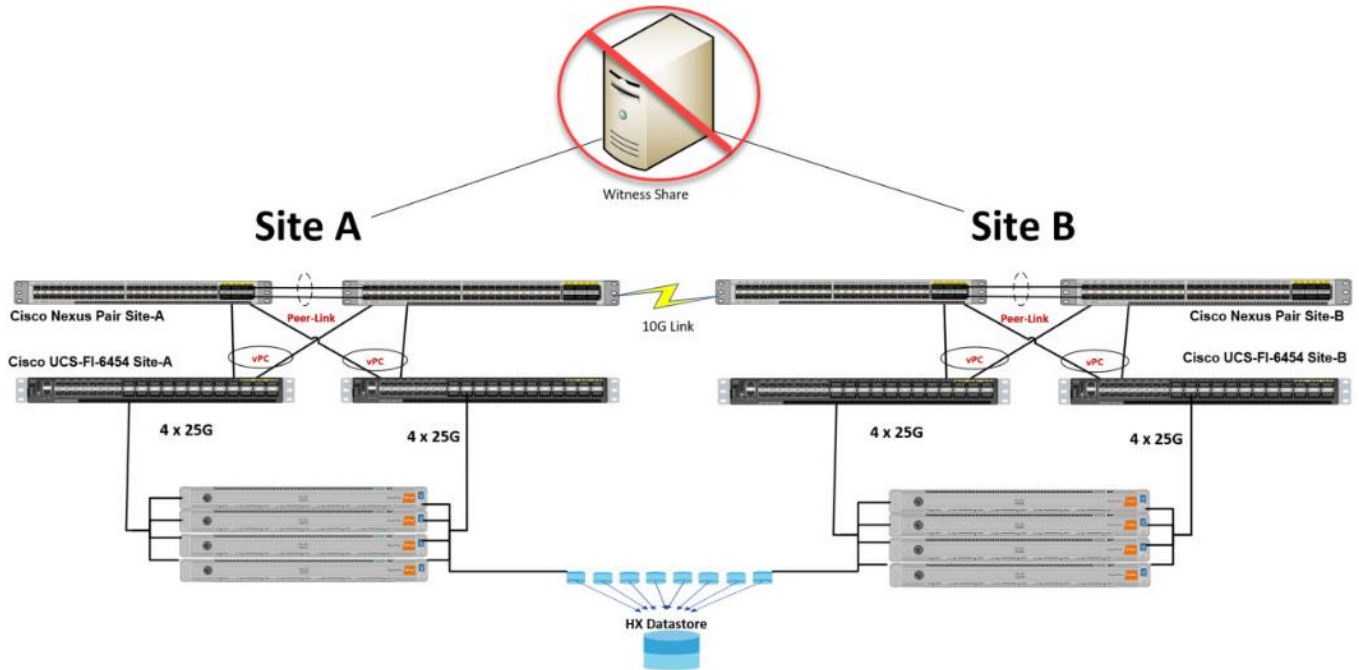
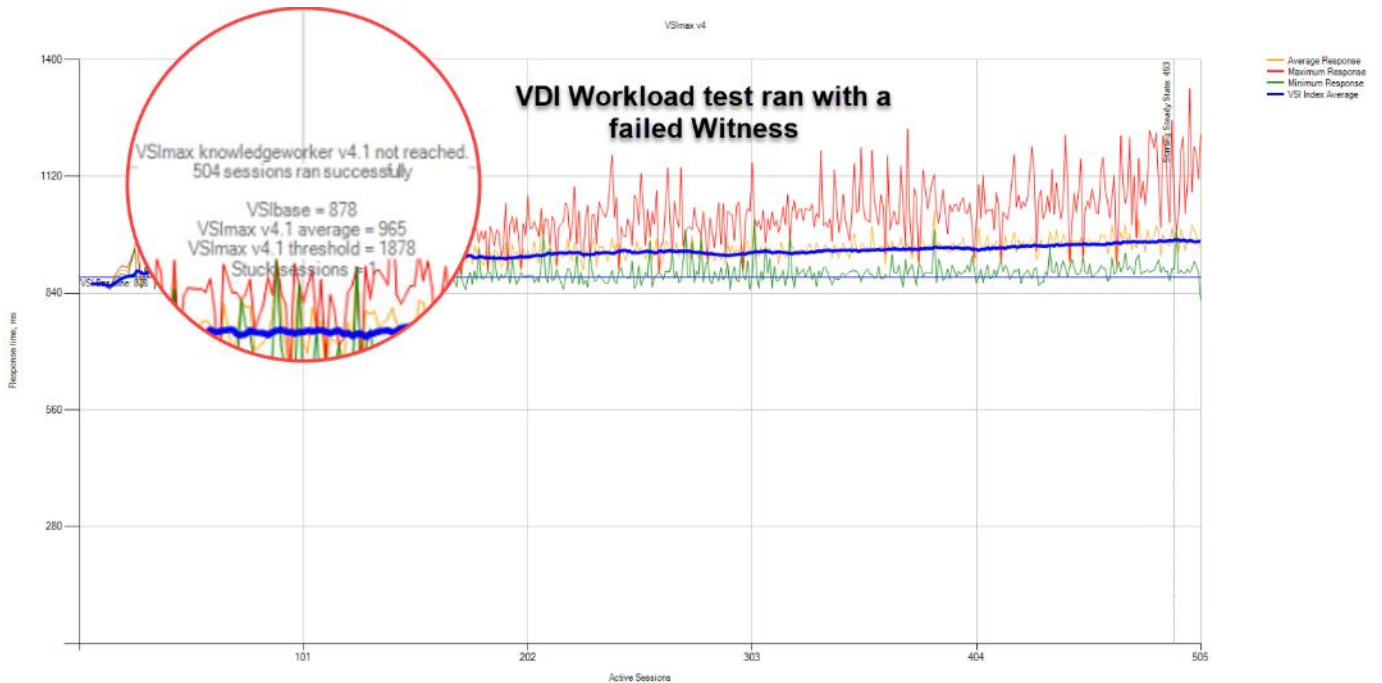


Figure 31. Login VSI Analyzer Chart for 500 Windows 10 Citrix VDI Desktops



MCS-KW-WITNESS FAIL-01a

Successfully completed Login VSI test with **504 knowledgeworker** sessions. VSImax (system saturation) was not reached.

Test result review

505 sessions were configured to be launched in 2880 seconds.

In total 1 sessions failed during the test:

- 0 sessions was/were not successfully launched
- 0 launched sessions failed to become active
- 505 sessions were active during the test
- 1 sessions got stuck during the test (before VSImax threshold) > [Click Here](#)

VDI Workload test ran with a failed Witness

With 504 sessions the maximum capacity VSImax (v4.1) **knowledgeworker** was not reached with a Login VSI baseline performance score of 878

Login VSI index average score is 933 lower than threshold. It might be possible to launch more sessions in this configuration.

Baseline performance of **878** is: **Good**



Figure 32. Test Results for 500 Citrix VDI Desktops with a Single Node Failure

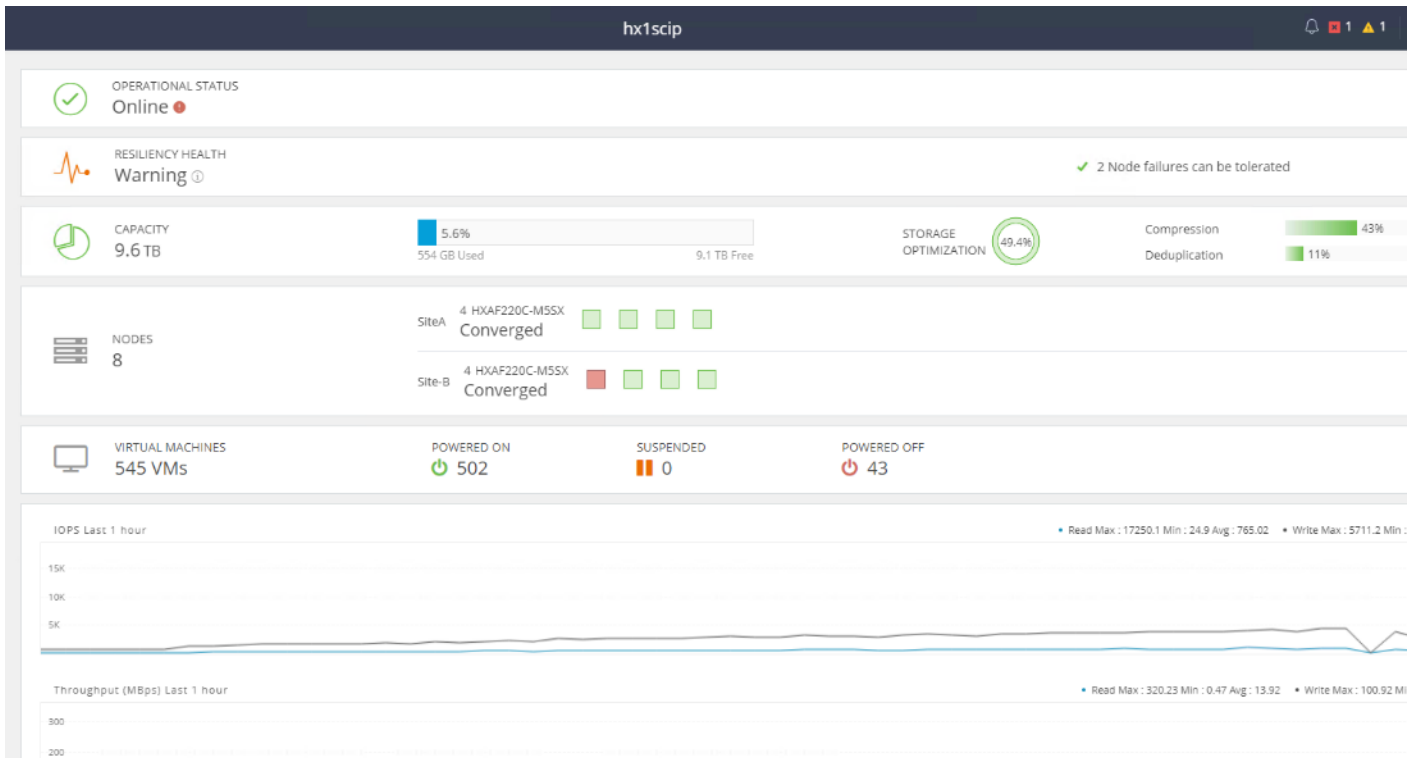
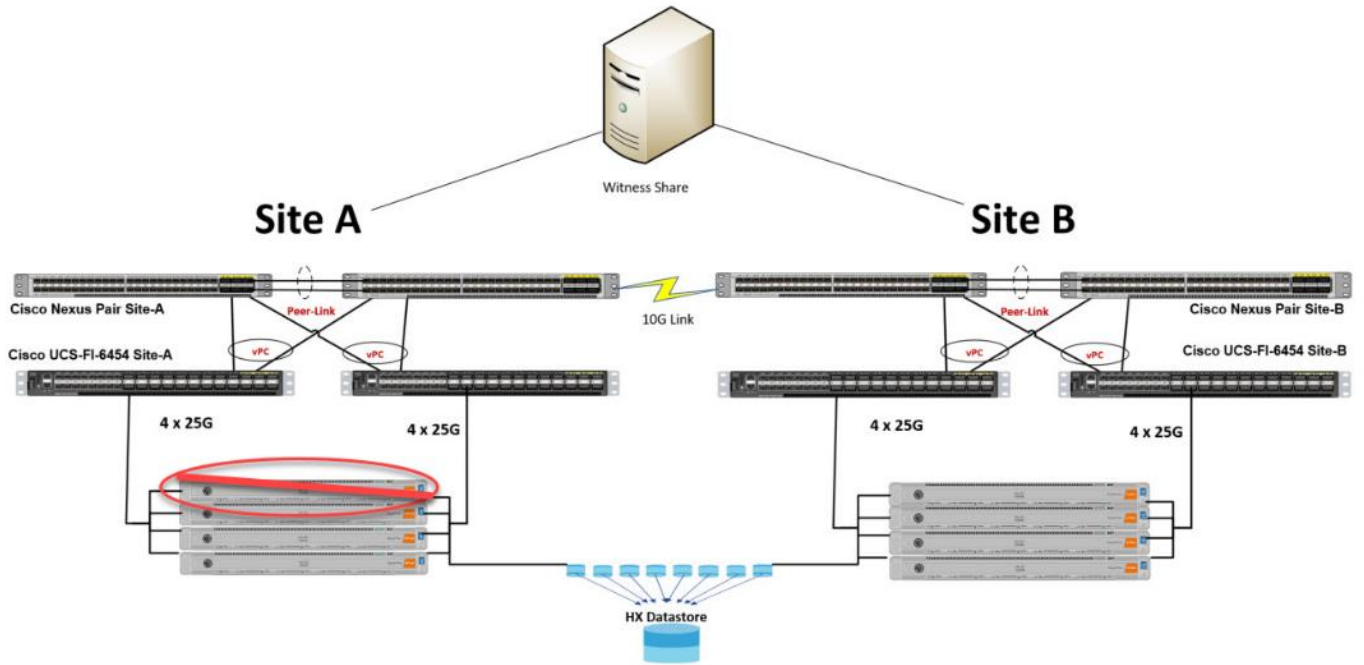
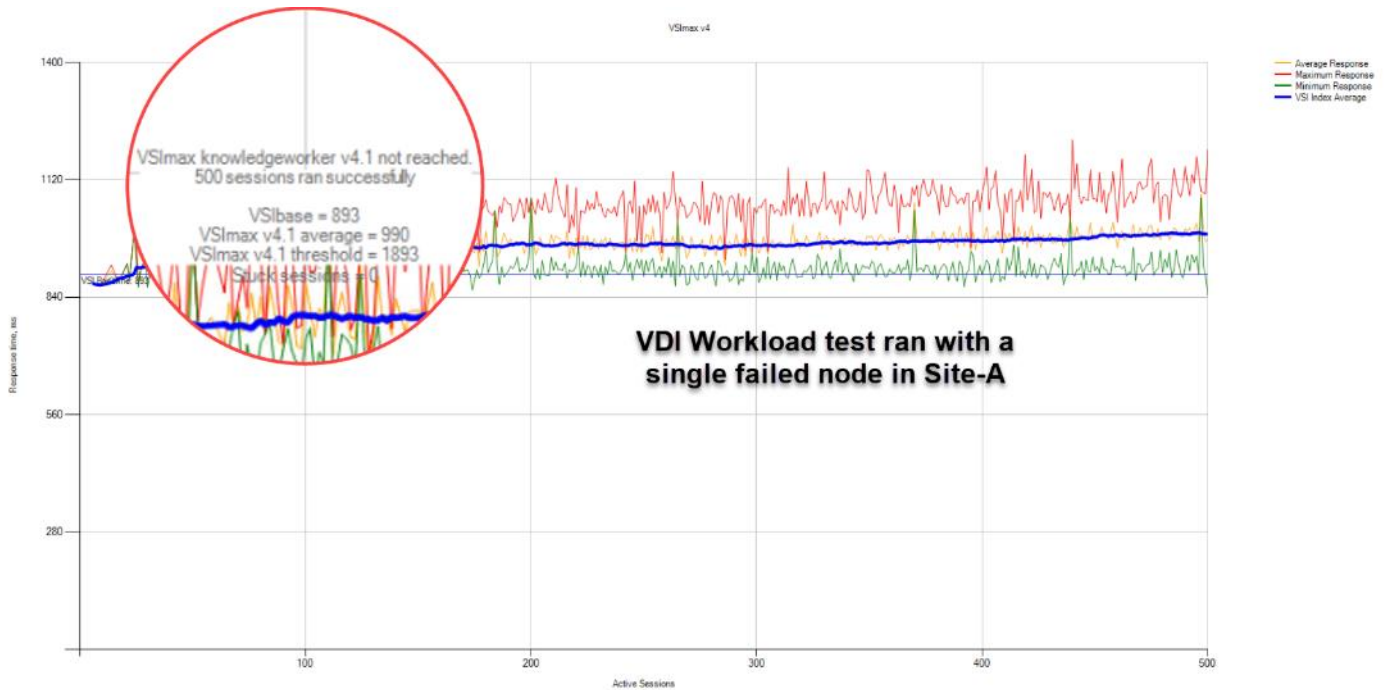


Figure 33. Login VSI Analyzer Chart for 4000 Windows 10 Citrix Virtual Desktops



MCS-Stretch-! node failure

Successfully completed Login VSI test with **500 knowledgeworker** sessions. VSImax (system saturation) was not reached. All Login VSI users completed the test.

Test result review

500 sessions were configured to be launched in **2880** seconds.

In total **0** sessions failed during the test:

- **0** sessions was/were not successfully launched
- **0** launched sessions failed to become active
- **500** sessions were active during the test
- **0** sessions got stuck during the test (before VSImax threshold)

VDI Workload test ran with a single failed node in Site-A

With **500** sessions the maximum capacity VSImax (v4.1) **knowledgeworker** was not reached with a Login VSI baseline performance score of **893**

Login VSI index average score is **915** lower than threshold. It might be possible to launch more sessions in this configuration.

Baseline performance of **893** is: **Good**

Figure 34. Test Results for 500 Citrix VDI with a Whole Site Failure

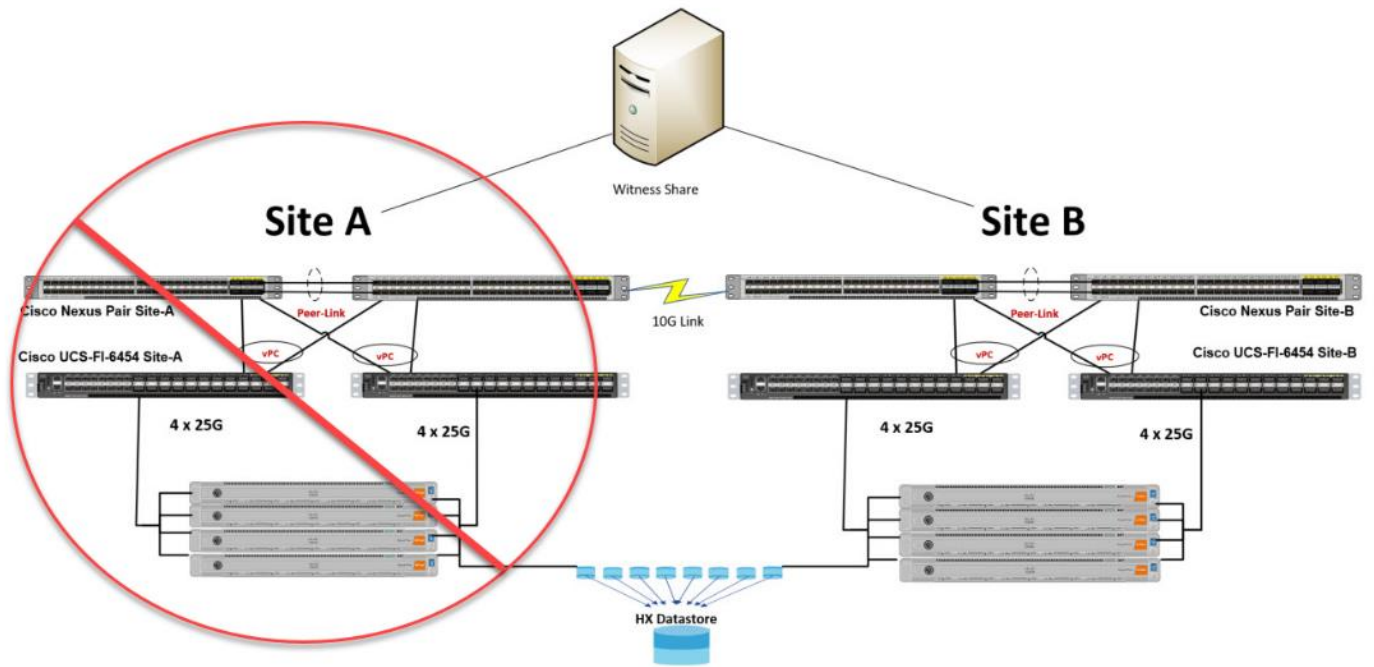


Figure 35. HX Connect Dashboard with Site A in a Failed State

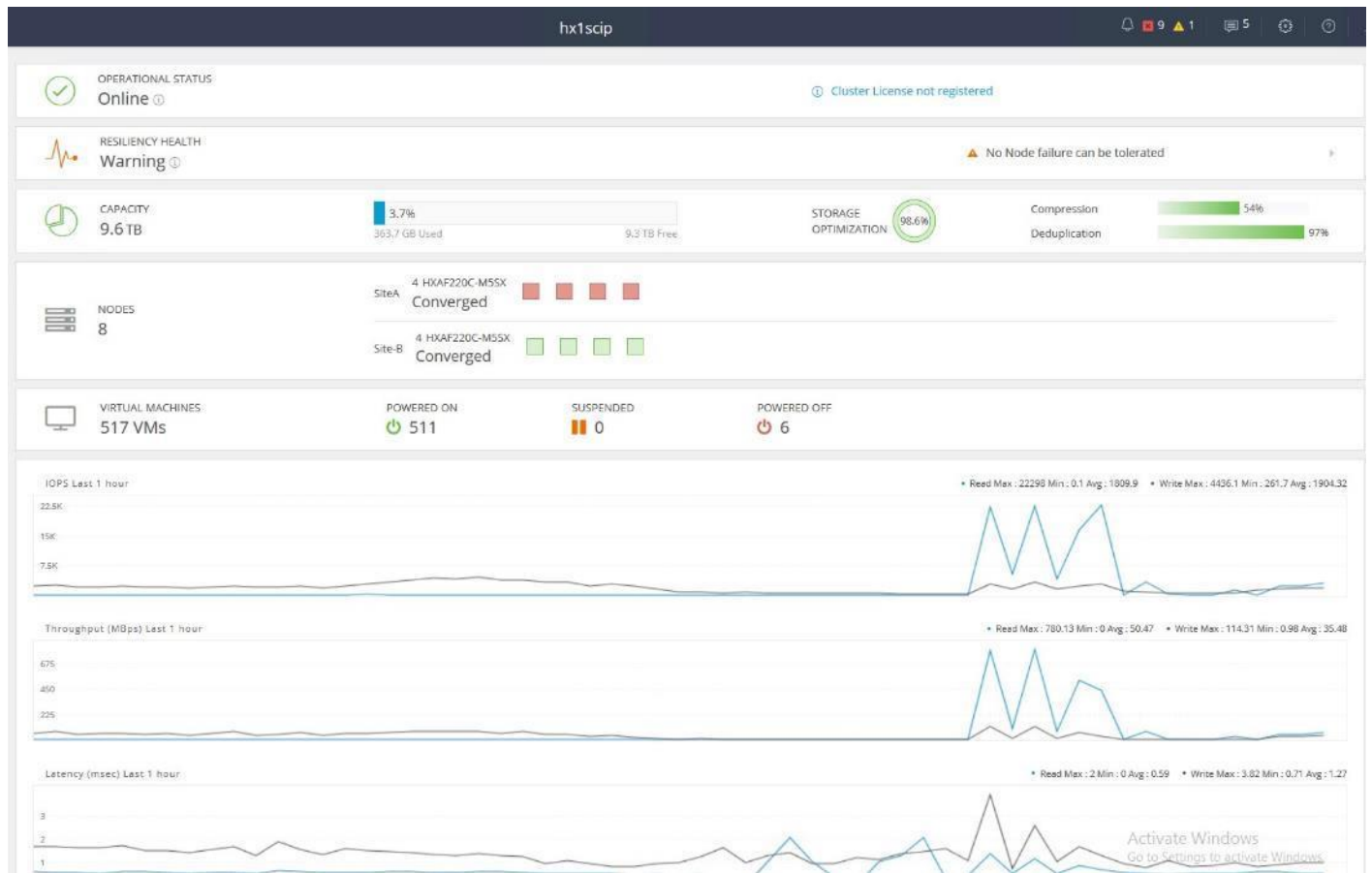
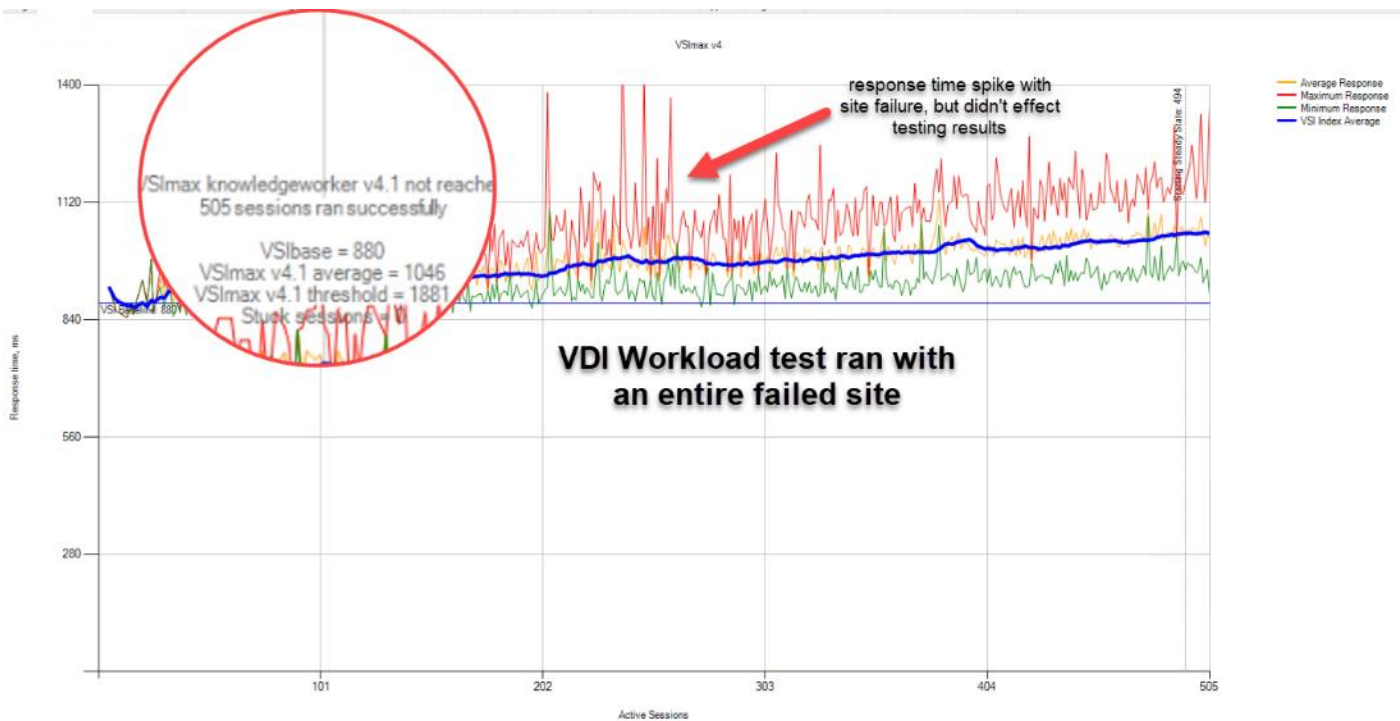


Figure 36. Login VSI Analyzer Chart for 4000 Windows 10 Citrix Virtual Desktops using MCS Persistent



MCS-KW-POSTFAIL-02

Successfully completed Login VSI test with **505 knowledgeworker** sessions. VSImax (system saturation) was not reached. All Login VSI users completed the test.

Test result review

505 sessions were configured to be launched in 2880 seconds.

In total 0 sessions failed during the test:

- 0 sessions was/were not successfully launched
- 0 launched sessions failed to become active
- 505 sessions were active during the test
- 0 sessions got stuck during the test (before VSImax threshold)

VDI Workload test ran with an entire failed site

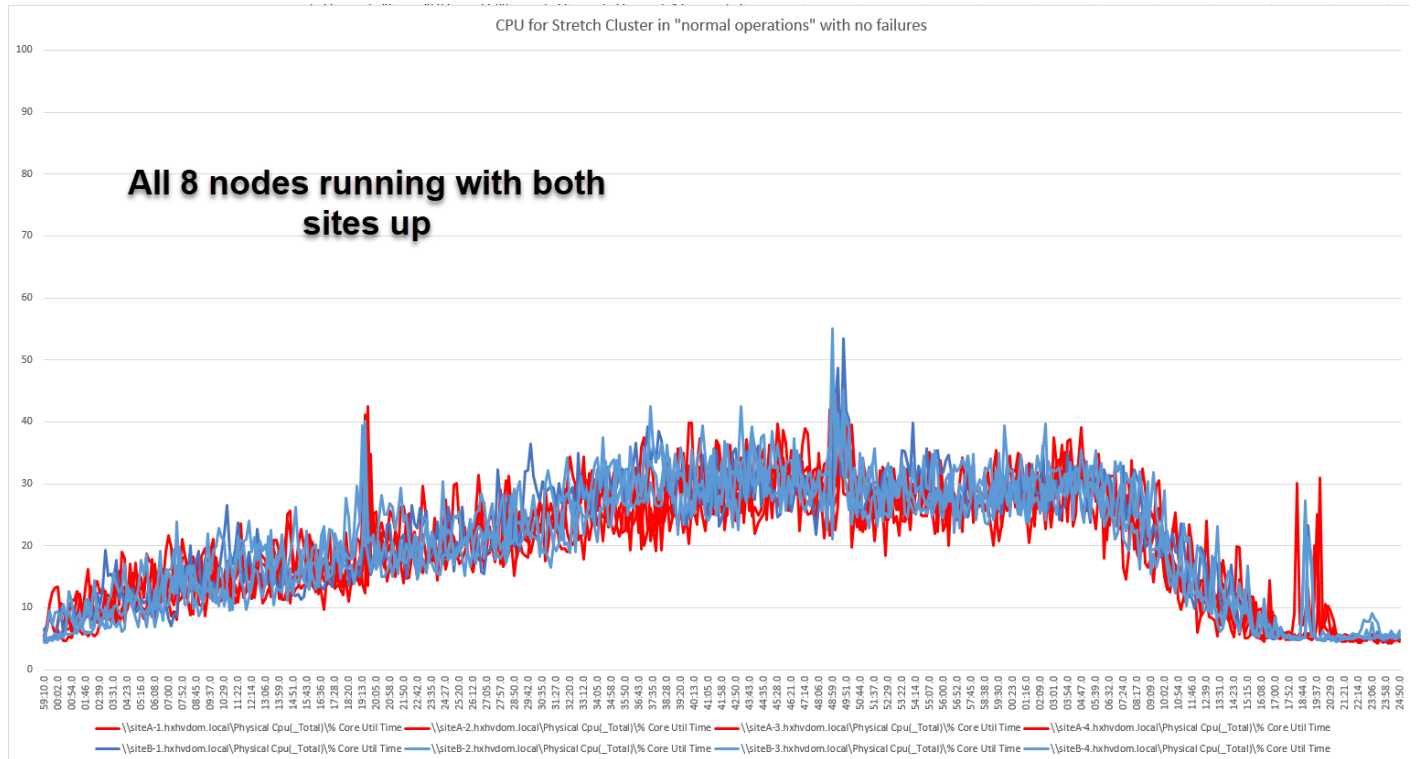
With 505 sessions the maximum capacity VSImax (v4.1) **knowledgeworker** was not reached with a Login VSI baseline performance score of 880

Login VSI index average score is 845 lower than threshold. It might be possible to launch more sessions in this configuration.

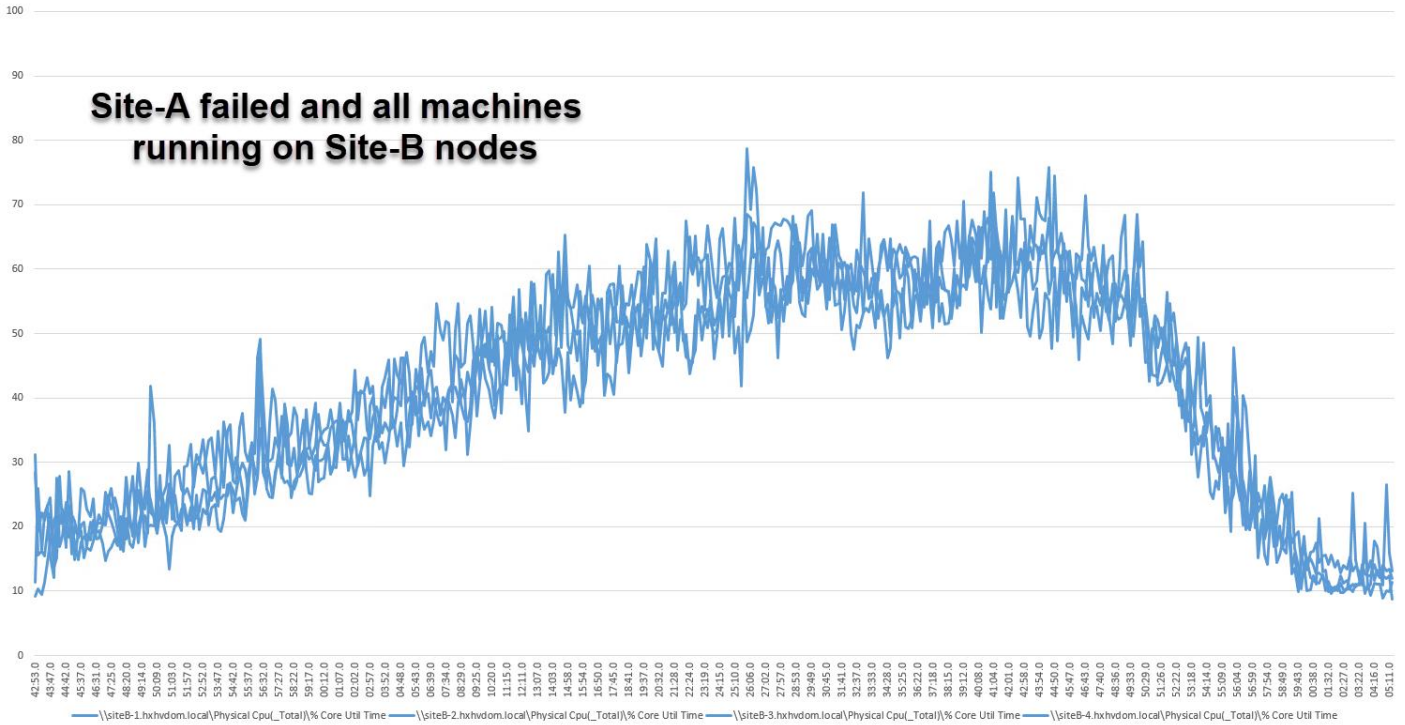
Baseline performance of **880** is: **Good**

ESX Host Performance Counters

When running a VMware ESXi environment for our Citrix Virtual Desktop workloads, it's important to monitor a few key performance counters to ensure the best end-user experience. CPU Performance: With VMware ESXi, using esxtop, our main counter is % Core Utilization.



CPU for complete failure of Site-A (Site-B hosts running ALL Virtual Desktops)



Summary

This Cisco HyperFlex solution addresses urgent needs of IT by delivering a platform that is cost effective and simple to deploy and manage. The architecture and approach used provides for a flexible and high-performance system with a familiar and consistent management model from Cisco. In addition, the solution offers numerous enterprise-class data management features to deliver the next-generation hyper-converged system.

This solutions offers flexibility with creating site resiliency and adds an extra high availability component to ensure VDI workload resiliency.

Only Cisco offers the flexibility to add compute only nodes to a true hyper-converged cluster for compute intensive workloads like desktop virtualization. This translates to lower cost for the customer since no hyper-convergence licensing is required for those nodes.

Delivering responsive, resilient, high-performance Citrix Virtual Desktops provisioned Microsoft Windows 10 Virtual Machines and Microsoft Windows Server for hosted Apps or desktops has many advantages for desktop virtualization administrators.

Virtual desktop end-user experience, as measured by the Login VSI tool in benchmark mode, is outstanding with Intel Xeon scalable family processors and Cisco 2666Mhz memory. In fact, we have set a new industry standard in performance for Desktop Virtualization on a hyper-converged platform.

About the Author

Jeff Nichols, Technical Marketing Manager, Desktop Virtualization and Graphics Solutions, Cisco Systems, Inc.

Jeff Nichols is a Cisco Unified Computing System architect, focusing on Virtual Desktop and Application solutions with extensive experience with Microsoft ESX/Hyper-V, Virtual Desktops, Virtual Apps and Microsoft Remote Desktop Services. He has expert product knowledge in application, desktop, and server virtualization across all three major hypervisor platforms and supporting infrastructures including but not limited to Windows Active Directory and Group Policies, User Profiles, DNS, DHCP and major storage platforms.

Feedback

For comments and suggestions about this guide and related guides, join the discussion on [Cisco Community](https://cs.co/en-cvds) at <https://cs.co/en-cvds>.

Americas Headquarters

Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters

Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters

Cisco Systems International BV Amsterdam,
The Netherlands

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

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