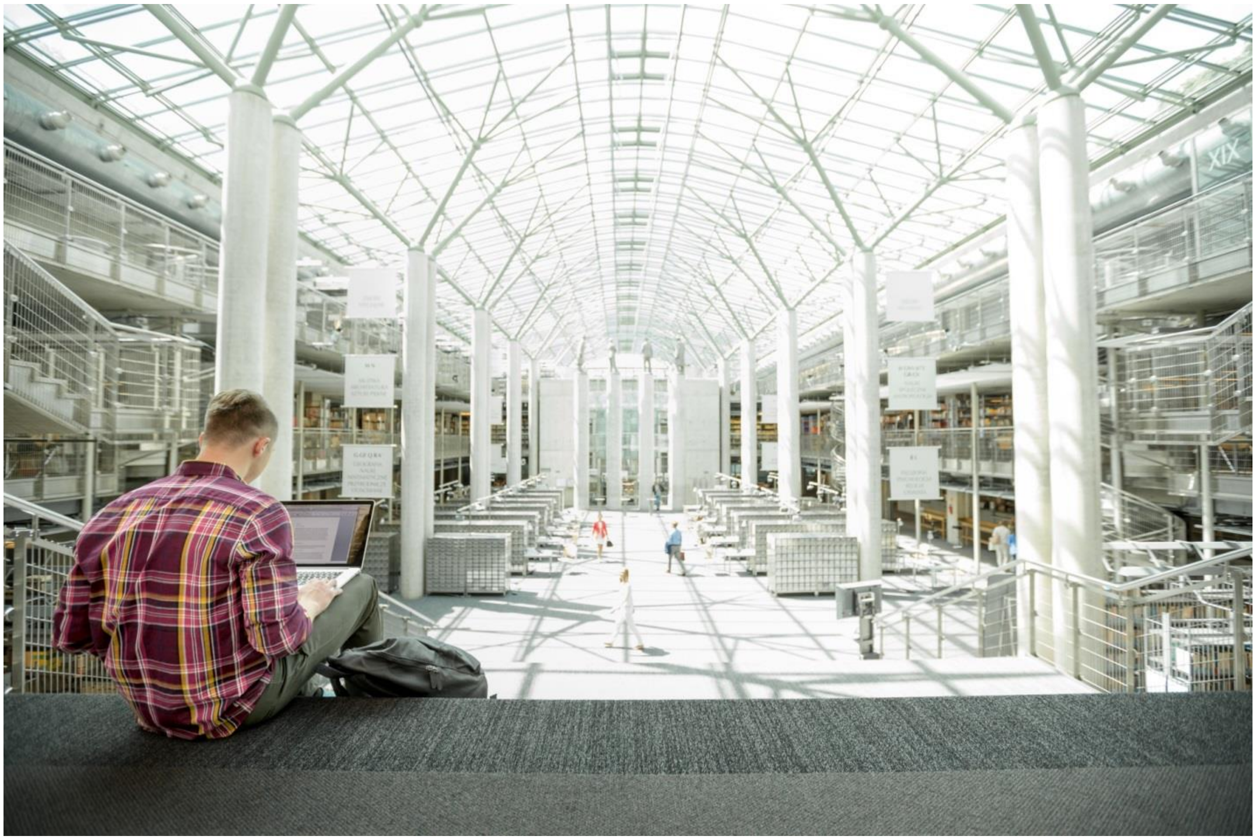


VersaStack with VMware vSphere 6.7, Cisco UCS 4th Generation Fabric, and IBM FS9100 NVMe-accelerated Storage Design Guide

Last Updated: September 23, 2019

CISCO VALIDATED DESIGN

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.

Table of Contents

Executive Summary

Cisco Validated Designs (CVDs) deliver systems and solutions that are designed, tested, and documented to facilitate and improve customer deployments. These designs incorporate a wide range of technologies and products into a portfolio of solutions that have been developed to address the business needs of the customers and to guide them from design to deployment.

The VersaStack solution, described in this CVD, delivers a Converged Infrastructure platform (CI) specifically designed for high-performance Virtual Server Infrastructure (VSI), which is a validated solution jointly developed by Cisco and IBM. In this deployment, IBM® FlashSystem 9100 combines the performance of flash and Non-Volatile Memory Express (NVMe) with the reliability and innovation of IBM FlashCore technology and the rich features of IBM Spectrum Virtualize. With the addition of Cisco UCS M5 servers including 2nd generation Intel Xeon Scalable processors and Cisco UCS 6400 Series Fabric Interconnects, the solution provides superior compute performance and network throughput with 10/25/40/100GbE support for ethernet using Nexus 9000 series switches in the LAN and 32G support for fibre channel connectivity with the Cisco MDS 9000 portfolio of switches in the SAN.

Solution Overview

Introduction

The VersaStack solution is a pre-designed, integrated and validated architecture for the data center that combines Cisco UCS servers, Cisco Nexus family of switches, Cisco MDS fabric switches, IBM Storwize, and FlashSystem Storage Arrays into a single, flexible architecture. VersaStack is designed for high availability, with no single points of failure, while maintaining cost-effectiveness and flexibility in design to support a wide variety of workloads.

The VersaStack design can support different hypervisor options, bare metal servers and can also be sized and optimized based on customer workload requirements. The VersaStack design discussed in this document has been validated for resiliency (under fair load) and fault tolerance during system upgrades, component failures, and partial loss of power scenarios.

This document discusses the design of the high-performance VersaStack with flash and NVMe based solution. The solution is a predesigned, best-practice data center architecture with VMware vSphere built on Cisco Unified Computing System (Cisco UCS). The solution architecture presents a robust infrastructure viable for a wide range of application workloads implemented as a Virtual Server Infrastructure (VSI).

Audience

The intended audience of this document includes, but is not limited to, sales engineers, field consultants, professional services, IT managers, architects, partner engineering, and customers who want to take advantage of an infrastructure built to deliver IT efficiency and enable IT innovation.

What's New in this Release?

The VersaStack with VMware vSphere 6.7 U2 CVD introduces new hardware and software into the portfolio. The following design elements distinguish this version of VersaStack from previous models:

- Support for Cisco UCS 6454 Fabric Interconnects
- Support for VIC 1400 series adapter cards on Cisco UCS M5 servers
- Support for Cisco UCS C480 M5 ML Servers
- Support for the Second Generation Intel® Xeon® Scalable processor (Cascade Lake) refresh and Intel® Optane™ Data Center persistent memory modules on Cisco UCS M5 Intel-based servers
- Improved memory RAS features on Cisco UCS M5 servers
- IBM FlashSystem 9100 release 8.2.1.6
- Support for the Cisco UCS release 4.0(4c)
- Validation of 25GbE IP-based storage design with Nexus NX-OS switches supporting iSCSI-based storage access
- Validation of VMware vSphere 6.7 U2
- 100 Gigabit per second Ethernet Connectivity
- 32 Gigabit per second Fibre Channel Connectivity

For more information on the complete portfolio of VersaStack solutions, please refer to the VersaStack documentation:

http://www.cisco.com/c/en/us/solutions/enterprise/data-center-designs-cloud-computing/versastack-designs.html

VersaStack Program Benefits

Cisco and IBM have carefully validated and verified the VersaStack solution architecture and its many use cases while creating a portfolio of detailed documentation, information, and references to assist customers in transforming their data centers to this shared infrastructure model.

Business Value

VersaStack combines the best-in-breed highly scalable storage controllers from IBM with the Cisco UCS B-Series and C-Series compute servers, and Cisco Nexus and MDS networking components. Quick deployment and rapid time to value allow enterprise clients to move away from disparate layers of compute, network, and storage to integrated stacks.

The CVD for the VersaStack reference architecture with pre-validated configurations reduces risk and expedites the deployment of infrastructure and applications. The system architects and administrators receive configuration guidelines to save implementation time while reducing operational risk.

The complexity of managing systems and deploying resources is reduced dramatically, and problem resolution is provided through a single point of support. VersaStack streamlines the support process so that customers can realize the time benefits and cost benefits that are associated with simplified single-call support.

Cisco Validated Designs incorporate a broad set of technologies, features, and applications to address your business needs.

This portfolio includes, but is not limited to best practice architectural design, Implementation and deployment instructions, Cisco Validated Designs and IBM Redbooks focused on a variety of use cases.

Design Benefits

VersaStack with IBM FlashSystem 9100 overcomes the historical complexity of IT infrastructure and its management.

Incorporating Cisco UCS Servers with IBM FlashSystem 9100 storage, this high-performance solution provides easy deployment and support for existing or new applications and business models. VersaStack accelerates IT and delivers business outcomes in a cost-effective and extremely timely manner.

One of the key benefits of VersaStack is the ability to maintain consistency in both scale-up and scale-down models. VersaStack can scale-up for greater performance and capacity. You can add compute, network, or storage resources as needed; or it can also scale-out when you need multiple consistent deployments such as rolling out additional VersaStack modules. Each of the component families shown in Figure 1 offer platform and resource options to scale the infrastructure up or down while supporting the same features and functionality.

The following factors contribute to significant total cost of ownership (TCO) advantages:

- **Simpler deployment model**: Fewer components to manage
- **Higher performance**: More work from each server due to faster I/O response times
- **Better efficiency**: Power, cooling, space, and performance within those constraints
- **Availability**: Help ensure applications and services availability at all times with no single point of failure
- **Flexibility**: Ability to support new services without requiring underlying infrastructure modifications
- **Manageability**: Ease of deployment and ongoing management to minimize operating costs

- **Scalability**: Ability to expand and grow with significant investment protection
- **Compatibility**: Minimize risk by ensuring compatibility of integrated components
- **Extensibility**: Extensible platform with support for various management applications and configuration tools

Technology Overview

The VersaStack architecture is comprised of the following infrastructure components for compute, network, and storage:

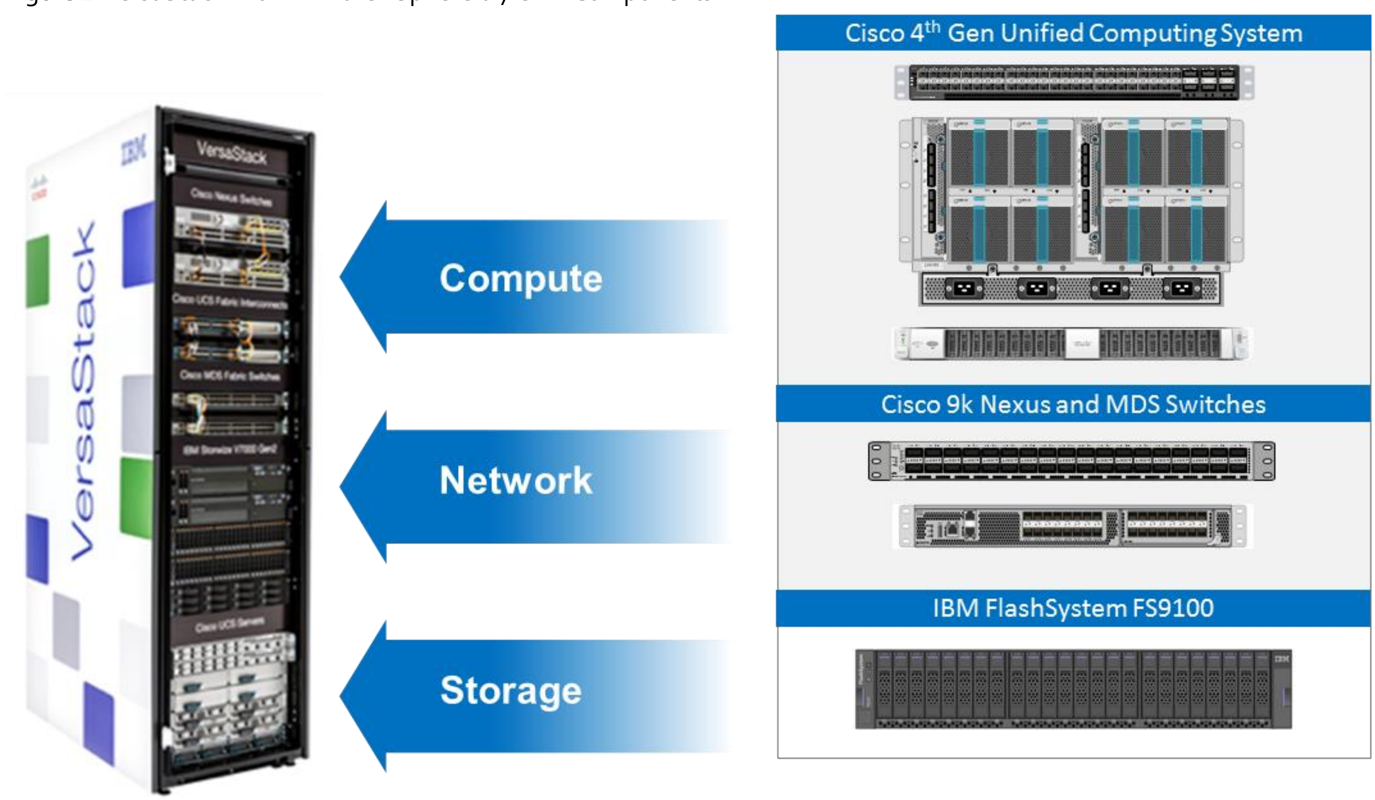
- Cisco Unified Computing System
- Cisco Nexus and Cisco MDS Switches
- IBM SAN Volume Controller, FlashSystem, and IBM Storwize family storage

These components are connected and configured according to best practices of both Cisco and IBM and provide an ideal platform for running a variety of workloads with confidence.

The VersaStack reference architecture covered in this document leverages:

- Cisco UCS 6400 Series Fabric Interconnects (FI)
- Cisco UCS 5108 Blade Server chassis
- Cisco Unified Compute System (Cisco UCS) servers with 2^{nd} generation Intel Xeon scalable processors
- Cisco Nexus 9336C-FX2 Switches running NX-OS mode
- Cisco MDS 9132T Fabric Switches
- IBM FlashSystem 9100 NVMe-accelerated Storage
- VMware vSphere 6.7 Update 2

Figure 1 VersaStack with VMware vSphere 6.7 U2 – Components



The following sections provide a technical overview of the compute, network, storage and management components of the VersaStack solution.

Cisco Unified Computing System

Cisco Unified Computing System (Cisco UCS) is a next-generation data center platform that integrates computing, networking, storage access, and virtualization resources into a cohesive system designed to reduce total cost of ownership (TCO) and to increase business agility. The system integrates a low-latency; lossless unified network fabric with enterprise-class, x86-architecture servers. The system is an integrated, scalable, multi-chassis platform where all resources are managed through a unified management domain.

The Cisco Unified Computing System consists of the following subsystems:

- **Compute** - The compute piece of the system incorporates servers based on latest Intel's x86 processors. Servers are available in blade and rack form factor, managed by Cisco UCS Manager.
- **Network** - The integrated network fabric in the system provides a low-latency, lossless, 10/25/40/100 Gbps Ether-net fabric. Networks for LAN, SAN and management access are consolidated within the fabric. The unified fabric uses the innovative Single Connect technology to lowers costs by reducing the number of network adapters, switches, and cables. This in turn lowers the power and cooling needs of the system.
- **Storage access** - Cisco UCS system provides consolidated access to both SAN storage and Network Attached Storage over the unified fabric. This provides customers with storage choices and investment protection. The use of Policies, Pools, and Profiles allows for simplified storage connectivity management.
- **Management** - The system uniquely integrates compute, network and storage access subsystems, enabling it to be managed as a single entity through Cisco UCS Manager software. Cisco UCS Manager increases IT staff productivity by enabling storage, network, and server administrators to collaborate on Service Profiles that define the desired server configurations.

Cisco UCS Management

Cisco UCS® Manager (UCSM) provides unified, integrated management for all software and hardware components in Cisco UCS. UCSM manages, controls, and administers multiple blades and chassis enabling administrators to manage the entire Cisco Unified Computing System as a single logical entity through an intuitive GUI, a CLI, as well as a robust API. Cisco UCS Manager is embedded into the Cisco UCS Fabric Interconnects and offers comprehensive set of XML API for third party application integration.

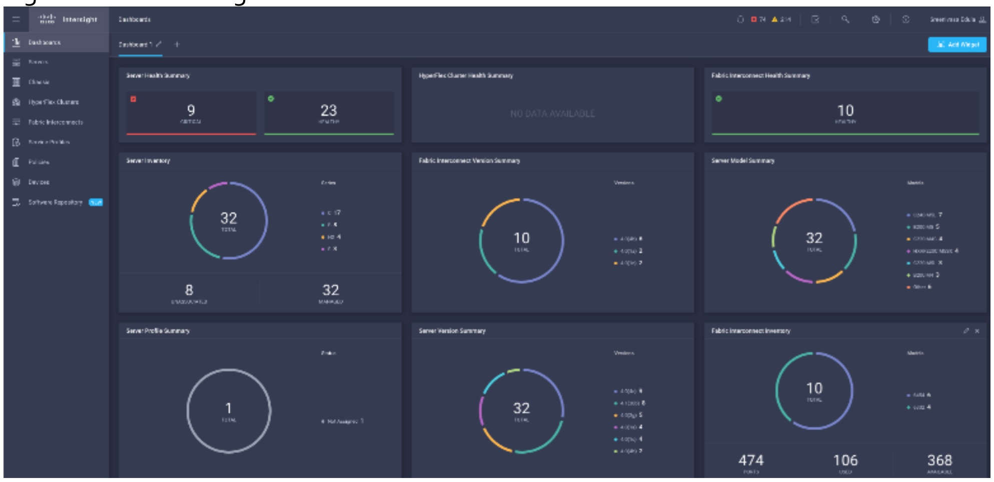
Cisco Intersight (optional)

The Cisco Intersight™ platform provides intelligent cloud-powered infrastructure management for Cisco Unified Computing System™(Cisco UCS®) and Cisco HyperFlex™ platforms. Cisco Intersight is a subscription-based, cloud service for infrastructure management that simplifies operations by providing pro-active, actionable intelligence for operations. Cisco Intersight provides capabilities such as Cisco Technical Assistance Center (TAC) integration for support and Cisco Hardware Compatibility List (HCL) integration for compliance that Enterprises can leverage for all their Cisco HyperFlex and Cisco UCS systems in all locations. Cloud-based delivery enables Enterprises to quickly adopt the new features that are continuously being rolled out in Cisco Intersight.

Each Cisco UCS server or Cisco HyperFlex system automatically includes a Cisco Intersight Base edition at no additional cost when the customer accesses the Cisco Intersight portal and claims the device. In addition, customers can purchase the Cisco Intersight Essentials edition using the Cisco ordering tool.

A view of the unified dashboard provided by Intersight can be seen in Figure 2.

Figure 2 Cisco Intersight Dashboard View



For more information on Cisco Intersight, see:

https://www.intersight.com/help/getting_started#cisco_intersight_overview

Cisco UCS Director (optional)

Cisco UCS Director is a heterogeneous platform for private cloud Infrastructure as a Service (IaaS). It supports a variety of hypervisors along with Cisco and third-party servers, network, storage, converged and hyperconverged infrastructure across bare-metal and virtualized environments. Cisco UCS Director provides increased efficiency through automation capabilities throughout VersaStack components. The Cisco UCS Director adapter for IBM Storage and VersaStack converged infrastructure solution allows easy deployment and management of these technologies using Cisco UCS Director.

Cisco continues to invest and enhance data center automation and private cloud infrastructure as a service (IaaS) platform, Cisco UCS Director. At the same time, we are leveraging the Cisco Intersight platform to deliver additional value and operational benefits when coupled with Cisco UCS Director.

Cisco is implementing a strategy for Cisco UCS Director and Cisco Intersight to help customers transition. Cisco UCS Director can be managed by Cisco Intersight to make updates easier and improve support. The combination of Cisco UCS Director and Intersight will simplify day to day operations and extend private cloud IaaS services.

For more information, see:

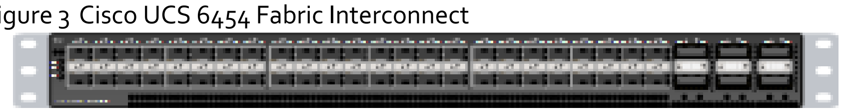
https://www.cisco.com/c/en/us/products/servers-unified-computing/ucs-director/index.html#~stickynav=1

Cisco UCS Fabric Interconnects

The Cisco UCS Fabric Interconnects (FIs) provide a single point for connectivity and management for the entire Cisco Unified Computing System. Typically deployed as an active-active pair, the system's fabric interconnects integrate all components into a single, highly available management domain controlled by the Cisco UCS Manager. Cisco UCS FIs provide a single unified fabric for the system that supports LAN, SAN and management traffic using a single set of cables.

The 4^{th} generation (6454) Fabric Interconnect (Figure 3) leveraged in this VersaStack design provides both network connectivity and management capabilities for the Cisco UCS system. The Cisco UCS 6454 offers line-rate, low-latency, lossless 10/25/40/100 Gigabit Ethernet, Fibre Channel over Ethernet (FCoE), and 32 Gigabit Fibre Channel functions.

Figure 3 Cisco UCS 6454 Fabric Interconnect



Cisco UCS 5108 Blade Server Chassis

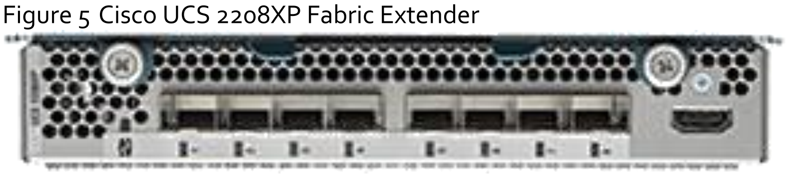
The Cisco UCS 5108 Blade Server Chassis (Figure 4) delivers a scalable and flexible blade server architecture. The Cisco UCS blade server chassis uses an innovative unified fabric with fabric-extender technology to lower total cost of ownership by reducing the number of network interface cards (NICs), host bus adapters (HBAs), switches, and cables that need to be managed. Cisco UCS 5108 is a 6-RU chassis that can house up to 8 half-width or 4 full-width Cisco UCS B-Series Blade Servers. A passive mid-plane provides up to 80Gbps of I/O bandwidth per server slot and up to 160Gbps for two slots (full-width). The rear of the chassis contains two I/O bays to house Cisco UCS Fabric Extenders for enabling uplink connectivity to the pair of FIs for both redundancy and bandwidth aggregation.

Figure 4 Cisco UCS 5108 Blade Server Chassis

Cisco UCS 2208XP Fabric Extender

The Cisco UCS Fabric extender (FEX) or I/O Module (IOM) multiplexes and forwards all traffic from servers in a blade server chassis to the pair of Cisco UCS FIs over 10Gbps unified fabric links. The Cisco UCS 2208XP Fabric Extender (Figure 5) has eight 10 Gigabit Ethernet, FCoE-capable, Enhanced Small Form-Factor Pluggable (SFP+) ports that connect the blade chassis to the FI. Each Cisco UCS 2208XP has thirty-two 10 Gigabit Ethernet ports connected through the midplane to each half-width slot in the chassis. Typically configured in pairs for redundancy, two fabric extenders provide up to 160 Gbps of I/O to the chassis.

Figure 5 Cisco UCS 2208XP Fabric Extender



Cisco UCS Virtual Interface Card 1400

The Cisco UCS Virtual Interface Card (VIC) 1400 Series provides complete programmability of the Cisco UCS I/O infrastructure by presenting virtual NICs (vNICs) as well as virtual HBAs (vHBAs) from the same adapter according to the provisioning specifications within UCSM.

The **Cisco UCS VIC 1440** is a dual-port 40-Gbps or dual 4x 10-Gbps Ethernet/FCoE capable modular LAN On Motherboard (mLOM) adapter designed exclusively for the M5 generation of Cisco UCS B-Series Blade Servers. When used in combination with an optional port expander, the Cisco UCS VIC 1440 capabilities are enabled for two ports of 40-Gbps Ethernet. In this CVD, Cisco UCS B200 M5 blade servers were equipped with Cisco VIC 1440.

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 10/25-Gbps Ethernet or FCoE. The card can present PCIe standards-compliant interfaces to the host, and these can be dynamically configured as either NICs or HBAs. In this CVD, Cisco VIC 1457 was installed in Cisco UCS C240 M5 server.

2nd Generation Intel® Xeon® Scalable Processors

This VersaStack architecture includes the 2nd generation Intel Xeon Scalable processors in all the Cisco UCS M5 server models used in the design. These processors provide a foundation for powerful data center platforms with an evolutionary leap in agility and scalability. Disruptive by design, this innovative processor family supports new levels of platform convergence and capabilities across computing, storage, memory, network, and security resources.

Cascade Lake (CLX-SP) is the code name for the next-generation Intel Xeon Scalable processor family that is supported on the Purley platform serving as the successor to Skylake SP. These chips support up to eight-way multiprocessing, use up to 28 cores, incorporate a new AVX512 x86 extension for neural-network and deep-learning workloads, and introduce persistent memory support. Cascade Lake SP–based chips are manufactured in an enhanced 14-nanometer (14-nm++) process and use the Lewisburg chip set.

Cisco UCS B-Series Blade Servers

Cisco UCS B-Series Blade Servers are based on Intel® Xeon® processors. They work with virtualized and non-virtualized applications to increase performance, energy efficiency, flexibility, and administrator productivity.

Table 1 lists the Cisco UCS B-Series server models available and supported with VersaStack.

Table 1 Cisco UCS B-Series Models

Model	B200 M5	B480 M5
Processors	2	4
Supported Processors	2nd Gen Intel Xeon Scalable and Intel Xeon Scalable processors	2nd Gen Intel Xeon Scalable and Intel Xeon Scalable processors
Maximum Memory	9 TB with Intel Optane DC Persistent Memory	18 TB with Intel Optane DC Persistent Memory
Form Factor	Half-width blade	Full-width blade
Built-in RAID	0, 1	0, 1
Mezzanine I/O Adapter Slots	Up to 3	Up to 5
GPUs	Up to 2	Up to 4
Maximum I/O Throughput per Blade	80 Gbps (2 x 40 Gbps)	160 Gbps (4 x 40 Gbps)
Maximum Servers per Chassis	8	4
Storage		
Internal Storage	Up to 2 SAS/SATA/SSD/NVMe	Up to 4 SAS/SATA/SSD/NVMe
Maximum Internal Storage	20. 5 TB	39 TB

Cisco UCS B200 M5 Servers

The Cisco UCS B200 M5 Blade Server (Figure 6) has been used in this VersaStack architecture. This enterprise-class server offers market-leading performance, versatility, and density without compromise for workloads including Virtual Desktop Infrastructure (VDI), web infrastructure, distributed databases, converged infrastructure, and enterprise applications such as Oracle and SAP HANA. The Cisco UCS B200 M5 server can quickly deploy stateless physical and virtual workloads through programmable, easy-to-use Cisco UCS Manager software and simplified server access through Cisco SingleConnect technology.

Figure 6 Cisco UCS B200 M5 Blade Server

For more information on Cisco UCS B-Series Blade Servers, see:

https://www.cisco.com/c/en/us/products/servers-unified-computing/ucs-b-series-blade-servers/index.html

Cisco UCS C-Series Rack Servers

Cisco UCS C-Series Rack Servers deliver unified computing in an industry-standard form factor to reduce TCO and increase agility. Each server addresses varying workload challenges through a balance of processing, memory, I/O, and internal storage resources.

Table 2 lists the Cisco UCS C-Series server models available and supported with VersaStack.

Table 2 Cisco UCS C-Series Models

Cisco UCS C220 M5 Rack Server	**Cisco UCS C240 M5 Rack Server**	**Cisco UCS C480 ML M5 Rack Server**	**Cisco UCS C480 M5 Rack Server**
• 2nd Gen Intel Xeon Scalable processors or Intel Xeon Scalable processors, 2-socket • 10 SFF or 4 LFF drives • Up to 9 TB of memory • 2 PCIe Generation 3 slots • Intel Optane DC persistent memory	• 2nd Gen Intel Xeon Scalable processors or Intel Xeon Scalable processors, 2-socket • 26 SFF or 12 LFF + 2 SFF drives • Up to 9 TB of memory • 6 PCIe Generation 3 slots • Intel Optane DC persistent memory	• Intel Xeon Scalable processors • 24 SFF drives • Up to 3 TB of memory • 4 PCIe Generation 3 slots with 8 BPUs	• 2nd Gen Intel Xeon Scalable processors or Intel Xeon Scalable processors, 2/4-socket • 32 SFF drives • Up to 18 TB of memory • 12 PCIe Generation 3 slots with up to 6 GPUs • Intel Optane DC persistent memory

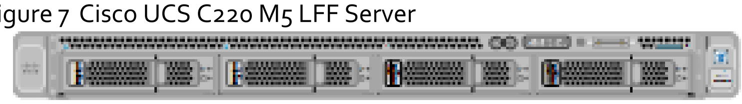
For more information on Cisco UCS C-Series Rack Servers, see:

https://www.cisco.com/c/en/us/products/servers-unified-computing/ucs-c-series-rack-servers/index.html

Cisco UCS C220 M5 LFF Server

The enterprise-class Cisco UCS C220 M5 Rack Server (Figure 7) leveraged in this VersaStack design extends the Cisco UCS portfolio in a 1RU rack server.

Figure 7 Cisco UCS C220 M5 LFF Server



The Cisco C220 M5 delivers outstanding levels of expandability and performance in a compact package, with:

- Latest (second generation) Intel Xeon Scalable CPUs with up to 28 cores per socket
- Supports first-generation Intel Xeon Scalable CPUs with up to 28 cores per socket

- Up to 24 DDR4 DIMMs for improved performance
- Support for the Intel Optane DC Persistent Memory (128G, 256G, 512G)*
- Support for 12-Gbps SAS modular RAID controller in a dedicated slot, leaving the remaining PCIe Generation 3.0 slots available for other expansion cards
- Modular LAN-On-Motherboard (mLOM) slot that can be used to install a Cisco UCS Virtual Interface Card (VIC) without consuming a PCIe slot
- Dual embedded Intel x550 10GBASE-T LAN-On-Motherboard (LOM) ports

* Requires a DDR4 DIMM for deployment

Cisco Nexus Switching Fabric

In the VersaStack design, Cisco UCS Fabric Interconnects and IBM storage systems are connected to the Cisco Nexus 9000 switches. Cisco Nexus 9000 provides Ethernet switching fabric for communications between the Cisco UCS domain, the IBM storage system and the enterprise network.

The validation for this deployment leverages the Cisco Nexus 9000 series switches, which deliver high performance 40/100GbE ports, density, low latency, and exceptional power efficiency in a broad range of compact form factors.

The Nexus 9000 switch leveraged in this CVD is the Nexus 9336C-FX2 (Figure 8) configured in NX-OS standalone mode. NX-OS is a purpose-built datacenter operating system designed for performance, resiliency, scalability, manageability, and programmability at its foundation. It provides a robust and comprehensive feature set that meets the demanding requirements of virtualization and automation in present and future data centers.

The Cisco Nexus 9336C-FX2 Switch is a 1RU switch that supports 36 ports, 7.2 Tbps of bandwidth and over 2.8 bpps. The switch can be configured to work as 1/10/25/40/100-Gbps offering flexible options in a compact form factor. Breakout is supported on all ports.

Figure 8 Cisco Nexus 9336C-FX2 Switch

For more information on Cisco Nexus 9000 series switches, see:

https://www.cisco.com/c/en/us/products/switches/nexus-9000-series-switches/index.html

All Nexus switch models including the Nexus 5000 and Nexus 7000 are supported in this design and may provide additional features such as FCoE or OTV. However, be aware that there may be slight differences in setup and configuration based on the switch used. The switch model also dictates the connectivity options between the devices including the bandwidth supported, transceiver and cable types required.

Cisco MDS 9000 Series Fabric Switch

In the VersaStack design, Cisco MDS switches provide SAN connectivity between the IBM storage systems and Cisco UCS domain.

The Cisco MDS 9000 family of multilayer switches give a diverse range of storage networking platforms, allowing you to build a highly scalable storage network with multiple layers of network and storage management intelligence. Fixed and modular models implement 2/4/8/10/16/32-32Gbps FC,1/10/40*Gbps FCIP, 10/40GE FCoE, and up to 48 Tbps of switching bandwidth.

For more information about Cisco MDS 9000 Series Switches, see: https://www.cisco.com/c/en/us/products/storage-networking/product-listing.html

Cisco MDS 9132T Fabric Switch

The Cisco MDS 9132T 32G Multilayer Fabric Switch used in this design is the next generation of the highly reliable, flexible, and low-cost Cisco MDS 9100 Series switches. It combines high performance with exceptional flexibility and cost effectiveness. This powerful, compact one rack-unit (1RU) switch scales from 8 to 32 line-rate 32 Gbps Fibre Channel ports. The Cisco MDS 9132T (Figure 9) delivers advanced storage networking features and functions with ease of management and compatibility with the entire Cisco MDS 9000 Family portfolio for reliable end-to-end connectivity. This switch also offers state-of-the-art SAN analytics and telemetry capabilities that have been built into this next-generation hardware platform.

Figure 9 Cisco MDS 9132T Switch

IBM Storage Networking c-type Family

Based on the Cisco MDS 9000 portfolio, IBM Storage Networking c-type 32G FC directors and switches complement IBM Storage all-flash and hybrid-flash systems through delivering industry-leading performance, scalability, security, and network connectivity in dense form factors. All Cisco sales teams that retire quota for MDS also retire quota for IBM c-type.

For more information on IBM Storage Networking c-type family, see: https://salesconnect.cisco.com/#/program/PAGE-14677

https://www.ibm.com/downloads/cas/Q64WPXKN

Cisco Data Center Network Manager(optional)

Cisco Data Center Network Manager (DCNM) is the comprehensive management solution for all NX-OS network deployments spanning LAN fabrics, SAN fabrics, and IP Fabric for Media (IPFM) networking in the data center powered by Cisco. DCNM provides management, control, automation, monitoring, visualization, and troubleshooting across Cisco Nexus and Cisco Multilayer Distributed Switching (MDS) solutions.

Cisco DCNM is optimized for large deployments with little overhead, but traditional deployments are supported as well for implementations. Fabric deployments can be customized by the user to meet business needs, delivering a range of features including SAN Telemetry, end-to-end topology views, and simplified zoning.

For more information about DCNM, see: https://www.cisco.com/c/en/us/products/collateral/cloud-systems-management/prime-data-center-network-manager/datasheet-c78-740978.html

IBM Spectrum Virtualize

The IBM Spectrum Virtualize™ software stack was first introduced as a part of the IBM SAN Volume Controller (SVC)product released in 2003, offering unparalleled storage virtualization capabilities before being integrated into the IBM Storwize platform and more recently, a subset of the IBM FlashSystem storage appliances.

Since the first release of IBM SAN Volume Controller, IBM Spectrum Virtualize has evolved into the feature-rich storage hypervisor evolving over 34 major software releases, installed and deployed on over 240,000+ Storwize and 70,000 SVC engines. Managing 410,000 enclosures, virtualizing, managing and securing 9.6 Exabytes of data. Exceeding 99.999% availability.

IBM Spectrum Virtualize firmware version 8.2.1.0 provides the following features:

- Connectivity: Incorporating support for increased bandwidth requirements of modern operating systems
  - Both 10GbE and 25GbE ports offering increased iSCSI performance for Ethernet environments.
  - NVMe-over-Fibre Channel on 16 Gb Fibre Channel adapters to allow end-to-end NVMe IO from supported Host Operating Systems.
- Virtualization: Supporting the external virtualization of over 450 (IBM and non-IBM branded) storage arrays over both Fibre Channel and iSCSI.
- Availability: Stretched Cluster and HyperSwap® for high availability among physically separated data centers. Or in a single site environment, Virtual Disk Mirroring for two redundant copies of LUN and higher data availability.
- Thin-provisioning: Helps improve efficiency by allocating disk storage space in a flexible manner among multiple users, based on the minimum space that is required by each user at any time.
- Data migration: Enables easy and nondisruptive moves of volumes from another storage system to the IBM FlashSystem 9100 system by using FC connectivity.
- Distributed RAID: Optimizing the process of rebuilding an array in the event of drive failures for better availability and faster rebuild times, minimizing the risk of an array outage by reducing the time taken for the rebuild to complete.
- Scalability: Clustering for performance and capacity scalability, by combining up-to 4 control enclosures together in the same cluster or connecting up-to 20 expansion enclosures.
- Simple GUI: Simplified management with the intuitive GUI enables storage to be quickly deployed and efficiently managed.
- Easy Tier technology: This feature provides a mechanism to seamlessly migrate data to the most appropriate tier within the IBM FlashSystem 9100 system.
- Automatic re-striping of data across storage pools: When growing a storage pool by adding more storage to it, IBM FlashSystem 9100 Software can restripe your data on pools of storage without having to implement any manual or scripting steps.
- FlashCopy: Provides an instant volume-level, point-in-time copy function. With FlashCopy and snapshot functions, you can create copies of data for backup, parallel processing, testing, and development, and have the copies available almost immediately.
- Encryption: The system provides optional encryption of data at rest, which protects against the potential exposure of sensitive user data and user metadata that is stored on discarded, lost, or stolen storage devices.
- Data Reduction Pools: Helps improve efficiency by compressing data by as much as 80%, enabling storage of up to 5x as much data in the same physical space.
- Remote mirroring: Provides storage-system-based data replication by using either synchronous or asynchronous data transfers over FC communication links:
  - Metro Mirror maintains a fully synchronized copy at metropolitan distances (up to 300 km).

– Global Mirror operates asynchronously and maintains a copy at much greater distances (up to 250 milliseconds round-trip time when using FC connections).

Both functions support VMware Site Recovery Manager to help speed DR. IBM FlashSystem 9100 remote mirroring interoperates with other IBM FlashSystem 9100, IBM FlashSystem V840, SAN Volume Controller, and IBM Storwize® V7000 storage systems.

For more information, go to the IBM Spectrum Virtualize website:
http://www03.ibm.com/systems/storage/spectrum/virtualize/index.html

IBM FlashSystems 9100

For decades, IBM has offered a range of enterprise class high-performance, ultra-low latency storage solutions. Now, IBM FlashSystem 9100 (Figure 10) combines the performance of flash and end-to-end NVMe with the reliability and innovation of IBM FlashCore technology and the rich feature set and high availability of IBM Spectrum Virtualize.

This powerful new storage platform provides:

- The option to use IBM FlashCore modules (FCMs) with performance neutral, inline-hardware compression, data protection and innovative flash management features provided by IBM FlashCore technology, or industry-standard NVMe flash drives.
- The software-defined storage functionality of IBM Spectrum Virtualize with a full range of industry-leading data services such as dynamic tiering, IBM FlashCopy management, data mobility and high-performance data encryption, among many others.
- Innovative data-reduction pool (DRP) technology that includes deduplication and hardware-accelerated compression technology, plus SCSI UNMAP support and all the thin provisioning, copy management and efficiency you'd expect from storage based on IBM Spectrum Virtualize.

Figure 10 IBM FlashSystem FS9100

The FlashSystem FS9100 series is comprised of two models; FS9110 and FS9150. Both storage arrays are dual, Active-Active controllers with 24 dual-ported NVMe drive slots. These NVMe slots cater for both traditional SSD drives, as well as the newly redesigned IBM FlashCore Modules.

The IBM FlashSystem 9100 system has two different types of enclosures: control enclosures and expansion enclosures.

- Control Enclosures

  Each control enclosure can have multiple attached expansion enclosures, which expands the available capacity of the whole system. The IBM FlashSystem 9100 system supports up to four control enclosures and up to two chains of SAS expansion enclosures per control enclosure.

  Host interface support includes 16 Gb Fibre Channel (FC), and 10 Gb or 25Gb Ethernet adapters for iSCSI host connectivity. Advanced Encryption Standard (AES) 256 hardware-based encryption adds to the rich feature set.

The IBM FlashSystem 9100 control enclosure supports up to 24 NVMe capable flash drives in a 2U high form factor.

There are two standard models of IBM FlashSystem 9100: 9110-AF7 and 9150-AF8. There are also two utility models of the IBM FlashSystem 9100: the 9110-UF7 and 9150-UF8.

The FS9110 has a total of 32 cores (16 per canister) while the 9150 has 56 cores (28 per canister). The FS9100 supports six different memory configurations as shown in Table 3 .

Table 3 FS9100 Memory Configurations

Memory per Canister	Memory per Control Enclosure
64 GB	128 GB
128 GB	256 GB
192 GB	384 GB
384 GB	768 GB
576 GB	1152 GB
768 GB	1536 GB

- Expansion Enclosures

  New SAS-based small form factor (SFF) and large form factor (LFF) expansion enclosures support flash-only MDisks in a storage pool, which can be used for IBM Easy Tier®:

  – The new IBM FlashSystem 9100 SFF expansion enclosure Model AAF offers new tiering options with solid-state drive (SSD flash drives). Up to 480 drives of serial-attached SCSI (SAS) expansions are supported per IBM FlashSystem 9100 control enclosure. The expansion enclosure is 2U high.

  – The new IBM FlashSystem 9100 LFF expansion enclosure Model A9F offers new tiering options with solid-state drive (SSD flash drives). Up to 736 drives of serial-attached SCSI (SAS) expansions are supported per IBM FlashSystem 9100 control enclosure. The expansion enclosure is 5U high.

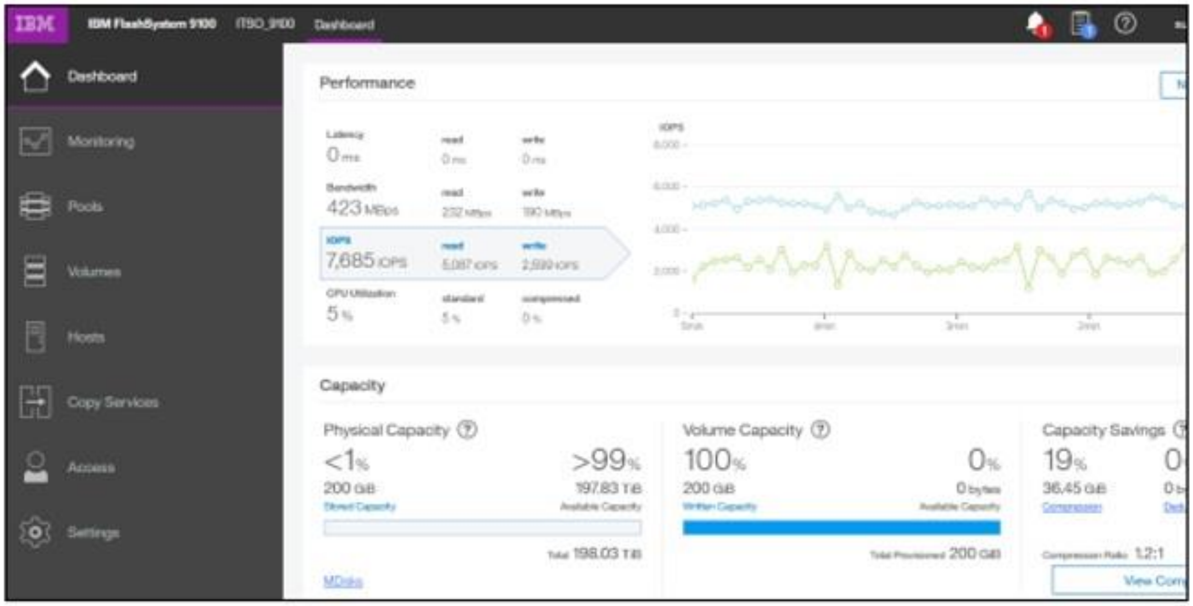
  The FS9100 supports NVMe attached flash drives, both the IBM Flash Core Modules (FCM) and commercial off the shelf (COTS) SSDs. The IBM FCMs support hardware compression and encryption with no reduction in performance. IBM offers the FCMs in three capacities: 4.8 TB, 9.6 TB and 19.2 TB, Standard NVMe SSDs are offered in four capacities, 1.92 TB, 3.84 TB, 7.68 TB, and 15.36 TB.

System Management and the Browser Interface

The IBM FlashSystem 9100 includes a single easy-to-use management graphical user interface (GUI) to help monitor, manage, and configure the system. The IBM FlashSystem 9100 system introduces an improved GUI with the same look and feel as other IBM FlashSystem solutions for a consistent management experience across all platforms. The GUI has an improved overview dashboard that provides all information in an easy-to-understand format and allows visualization of effective capacity.

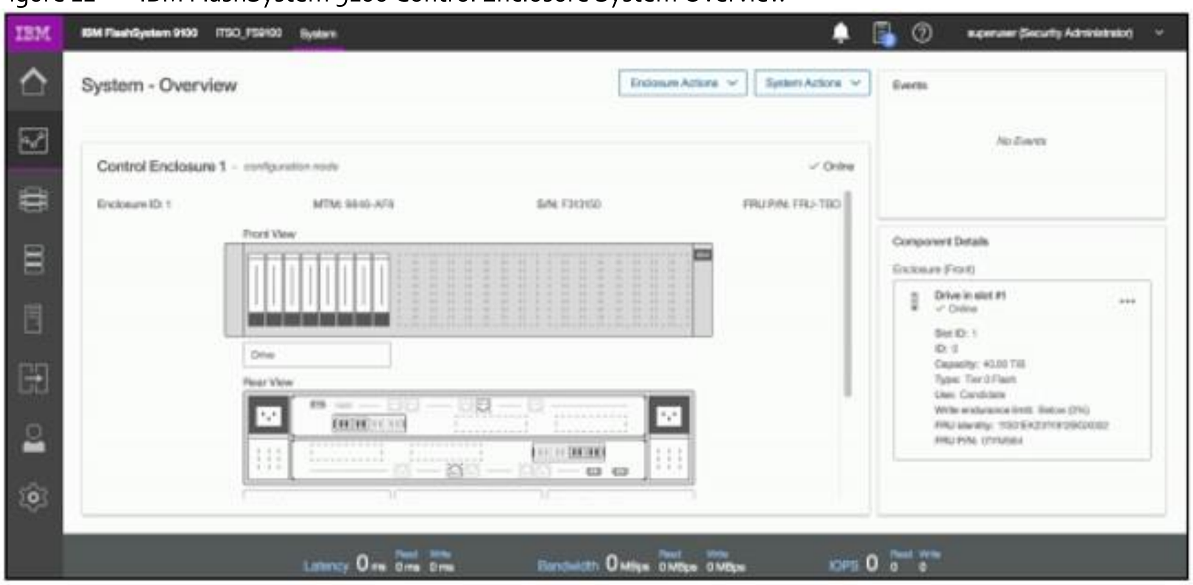
Figure 11 shows the IBM FlashSystem 9100 dashboard view. This is the default view that is displayed after the user logs on to the IBM FlashSystem 9100 system.

Figure 11 IBM FlashSystem 9100 Control Enclosure Dashboard View



In Figure 12, the GUI shows one IBM FlashSystem 9100 Control Enclosure. This is the System Overview window.

Figure 12 IBM FlashSystem 9100 Control Enclosure System Overview



The IBM FlashSystem 9100 system includes a CLI, which is useful for scripting, and an intuitive GUI for simple and familiar management of the product. RESTful API support was recently introduced to allow workflow automation or integration into DevOps environments.

The IBM FlashSystem 9100 system supports Simple Network Management Protocol (SNMP), email forwarding that uses Simple Mail Transfer Protocol (SMTP), and syslog redirection for complete enterprise management access.

VMware vSphere 6.7 Update 2

VMware vSphere is a virtualization platform for holistically managing large collections of infrastructures (resources-CPUs, storage and networking) as a seamless, versatile, and dynamic operating environment. Unlike traditional operating systems that manage an individual machine, VMware vSphere aggregates the infrastructure of an entire data center to create a single powerhouse with resources that can be allocated quickly and dynamically to any application in need.

vSphere 6.7 Update 2 (U2)brings a number of improvements including, but not limited to:

- vSphere Web Client and fully featured HTML-5 client
- Content Library Improvements
- VMware Appliance Management Interface Improvements

Solution Design

This VersaStack design aligns with the converged infrastructure configurations and best practices as identified in the previous VersaStack releases. The solution focuses on integration of IBM FS9100 in to VersaStack architecture with Cisco UCS 4th Generation and support for VMware vSphere 6.7 U2.

Requirements

The VersaStack data center is intended to provide a Virtual Server Infrastructure (VSI) that becomes the foundation for hosting virtual machines and applications. This design assumes existence of management, network and routing infrastructure to provide necessary connectivity, along with the availability of common services such as DNS and NTP, and so on.

This VersaStack solution meets the following general design requirements:

- Resilient design across all layers of the infrastructure with no single point of failure.
- Scalable design with the flexibility to add compute capacity, storage, or network bandwidth as needed.
- Modular design that can be replicated to expand and grow as the needs of the business grow.
- Flexible design that can support components beyond what is validated and documented in this guide.
- Simplified design with ability to automate and integrate with external automation and orchestration tools.
- Extensible design with support for extensions to existing infrastructure services and management applications.

The system includes hardware and software compatibility support between all components and aligns to the configuration best practices for each of these components. All the core hardware components and software releases are listed and supported on both the Cisco compatibility list:

http://www.cisco.com/en/US/products/ps10477/prod_technical_reference_list.html

and IBM Interoperability Matrix:

http://www-03.ibm.com/systems/support/storage/ssic/interoperability.wss

The following sections explain the physical and logical connectivity details across the stack including various design choices at compute, storage, virtualization and network layers.

Physical Topology

The VersaStack infrastructure satisfies the high-availability design requirements and is physically redundant across the network, compute and storage stacks. Figure 13 provides a high-level topology of the system connectivity.

To provide the compute to storage system connectivity, this design guide highlights two different storage connectivity options:

- Option 1: iSCSI based storage access through Cisco Nexus Fabric
- Option 2: FC based storage access through Cisco MDS Fabric

This VersaStack design utilizes Cisco UCS platform with Cisco UCS B200 M5 half-width blades and Cisco UCS C220 M5 servers connected and managed through Cisco UCS 6454 Fabric Interconnects and the integrated Cisco UCS Manager (UCSM). These high-performance servers are configured as stateless compute nodes where ESXi 6.7 U2 hypervisor is loaded using SAN (iSCSI and FC) boot. The boot disks to store ESXi hypervisor image and configuration along with the block based datastores to host application Virtual Machines (VMs) are provisioned on the IBM FS9100 storage array.

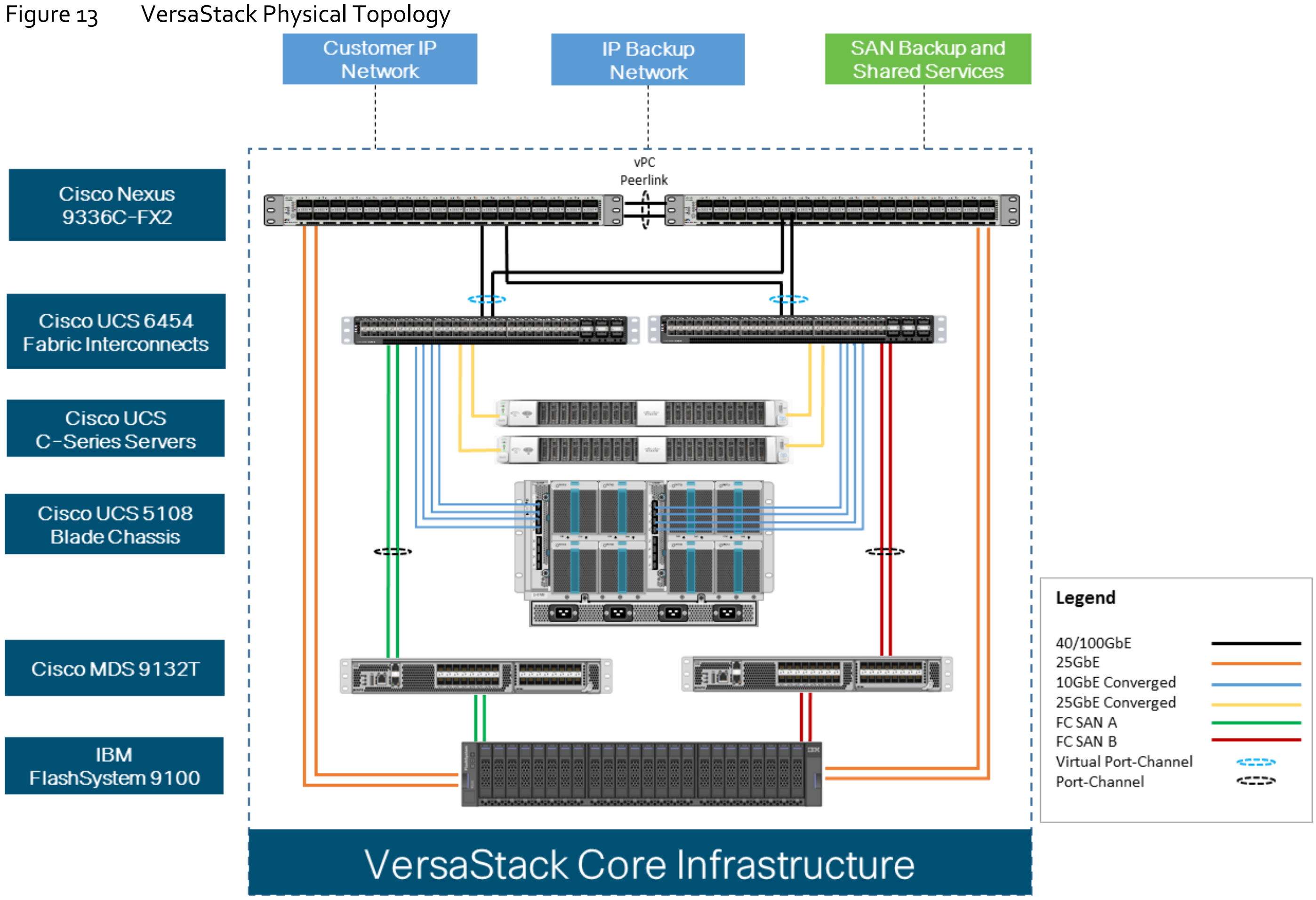
This design has following physical connectivity between the components of VersaStack:

- 4 X 10 Gb Ethernet connections port-channeled between the Cisco UCS 5108 Blade Chassis and the Cisco UCS Fabric Interconnects
- 25 Gb Ethernet connections port-channeled between the Cisco UCS C-Series rackmounts and the Cisco UCS Fabric Interconnects
- 100 Gb Ethernet connections port-channeled between the Cisco UCS Fabric Interconnect and Cisco Nexus 9000s
- 32 Gb Fibre Channel connections port-channeled between the Cisco UCS Fabric Interconnect and Cisco MDS 9132T
- 16 Gb Fibre Channel connections between the Cisco MDS 9132T and IBM FS9100 storage array for fibre channel block storage access
- 25 Gb Ethernet connections between the Cisco Nexus 9000s and IBM FS9100 storage array for iSCSI block storage access

---

Any supported connectivity to existing customer IP and SAN Networks from the VersaStack core infrastructure is allowed.

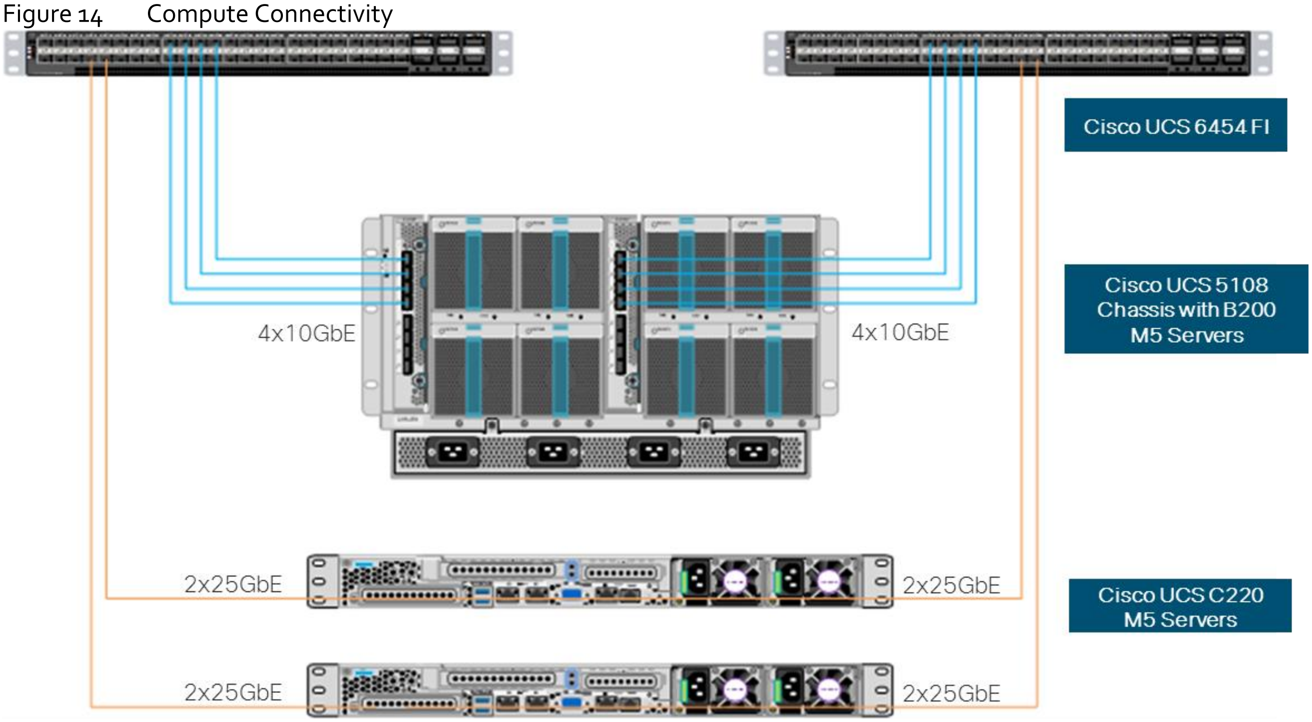
Figure 13 VersaStack Physical Topology



Compute Connectivity

The VersaStack compute design supports both Cisco UCS B-Series and C-Series deployments. Cisco UCS supports the virtual server environment by providing robust, highly available, and integrated compute resources centrally managed from Cisco UCS Manager in the Enterprise or from the Cisco Intersight Software as a Service (SaaS) cloud offering. In this validation effort, multiple Cisco UCS B-Series and C-Series ESXi servers are booted from SAN using iSCSI and FC (depending on the storage design used).

Figure 14 Compute Connectivity



The 5108 chassis in the design is populated with Cisco UCS B200 M5 blade servers and each of these blade servers contain one physical network adapter (Cisco VIC 1440) that passes converged fibre channel and IP traffic through the chassis mid-plane to the 2208XP FEXs. The FEXs are redundantly connected to the managing fabric interconnects with 4X 10GbE ports per FEX to deliver an aggregate bandwidth of 80Gbps to the chassis, full population of the 2208XP FEX can support 8x 10Gbps ports, allowing for an aggregate bandwidth of 160Gbps to the chassis.

These connections from the Cisco UCS Fabric Interconnects to the FEXs are automatically configured as port channels by specifying a Chassis/FEX Discovery Policy within UCSM.

Each Cisco UCS C-Series rack server in the design is redundantly connected to the managing fabric interconnects with at least one port connected to each FI to support converged traffic as with the Cisco UCS B-Series servers. Internally the Cisco UCS C-Series servers are equipped with a Cisco VIC 1457 network interface card (NIC) with quad 10/25 Gigabit Ethernet (GbE) ports. The Cisco VIC is installed in a modular LAN on motherboard (MLOM) slot. The standard practice for redundant connectivity is to connect port 1 of each server's VIC card to a numbered port on FI A, and port 3 of each server's VIC card to the same numbered port on FI B. The use of ports 1 and 3 are because ports 1 and 2 form an internal port-channel, as does ports 3 and 4. This allows an optional 4 cable connection method providing an effective 50GbE bandwidth to each fabric interconnect. The upstream ethernet traffic and the fibre channel traffic from the IBM Storage array is converged and transmitted to the Cisco UCS servers.

Cisco UCS Server Configuration for VMware vSphere

The Cisco UCS servers are stateless and are deployed using Cisco UCS Service Profiles (SP) that consists of server identity information pulled from pools (WWPN, MAC, UUID, and so on) as well as policies covering connectivity, firmware and power control options, and so on. The service profiles are provisioned from the Cisco UCS Service Profile Templates that allow rapid creation, as well as guaranteed consistency of the hosts at the Cisco UCS hardware layer.

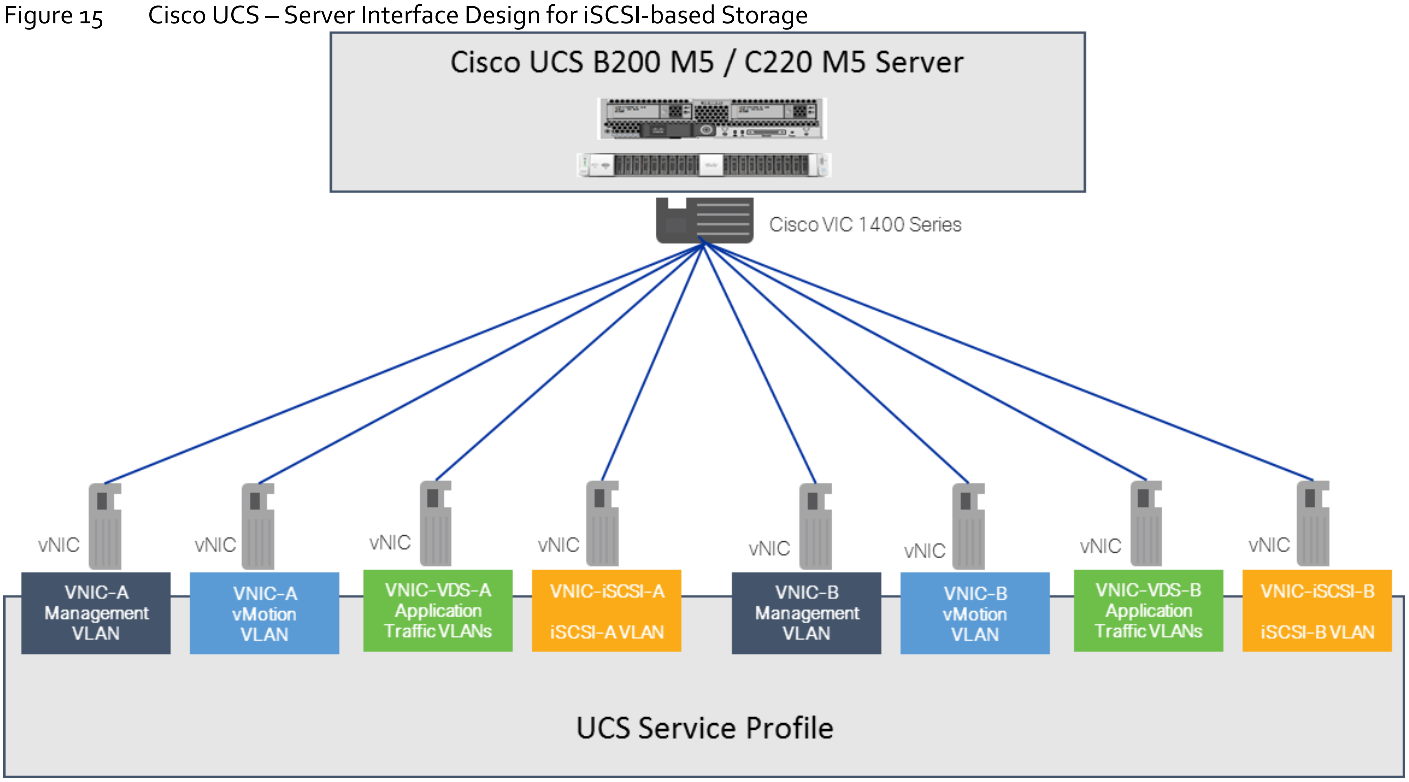
The ESXi nodes consist of Cisco UCS B200 M5 blades or Cisco UCS C220 M5 rack servers with Cisco UCS 1400 series VIC. These nodes are allocated to a VMware High Availability cluster to support infrastructure services and applications. At the server level, the Cisco 1400 VIC presents multiple virtual PCIe devices to the ESXi node and the vSphere environment identifies these interfaces as vmnics or vmhbas. The ESXi operating system is unaware of the fact that the NICs or HBAs are virtual adapters.

In the VersaStack design with iSCSI storage, eight vNICs are created and utilized as follows (Figure 15):

- One vNIC (iSCSI-A) for iSCSI SAN traffic
- One vNIC (iSCSI-B) for iSCSI SAN traffic
- Two vNICs for in-band management traffic
- Two vNICs for vMotion traffic
- Two vNICs for application virtual machines hosted on the infrastructure including storage access. These vNICs are assigned to a distributed switch (vDS)

These vNICs are pinned to different Fabric Interconnect uplink interfaces and are assigned to separate vSwitches and vSphere distributed switches (VDS) based on type of traffic. The vNIC to vSwitch and vDS assignment is covered later in the document.

Figure 15 Cisco UCS – Server Interface Design for iSCSI-based Storage

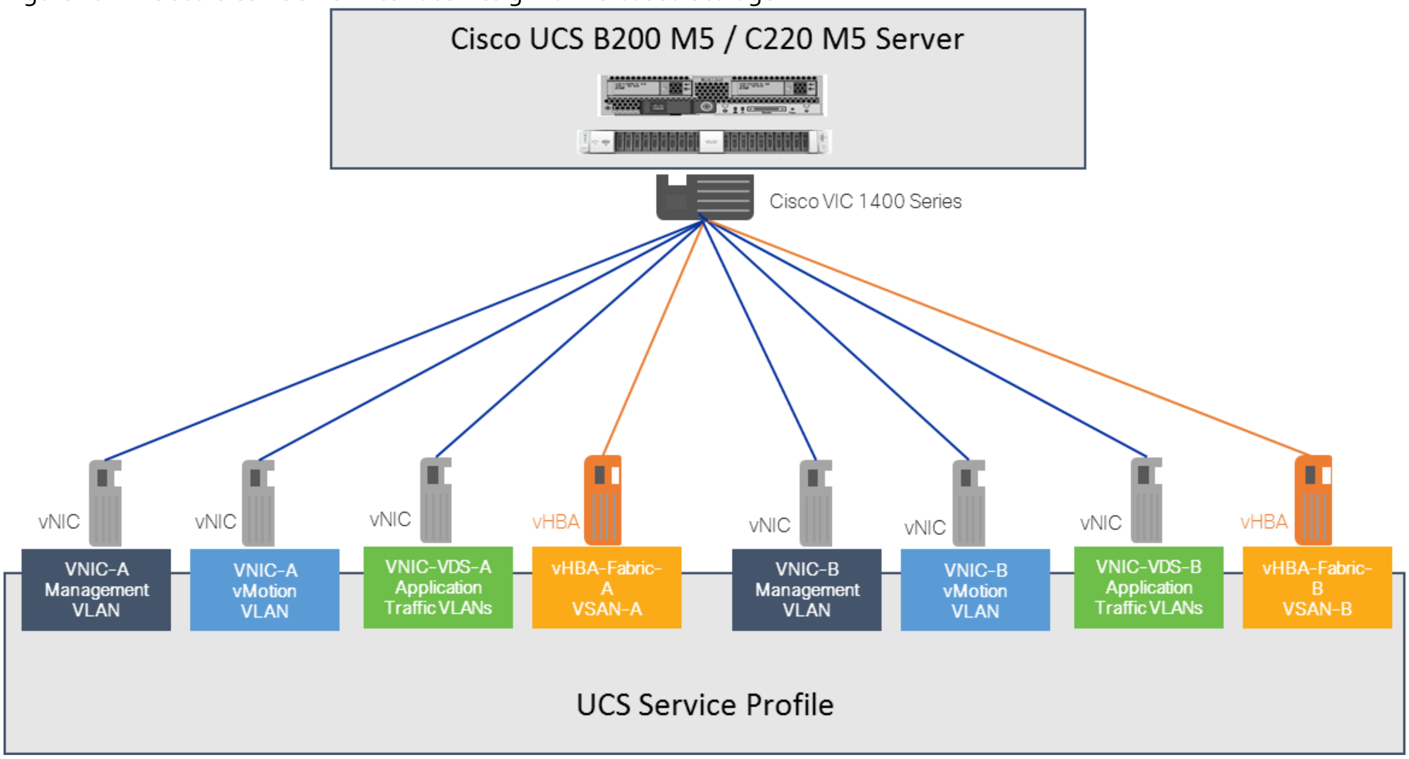


In the VersaStack design with FC storage, six vNICs and two vHBAs are created and utilized as follows (Figure 16):

- Two vNICs for in-band management traffic
- Two vNICs for vMotion traffic

- Two vNICs for application related data including application storage access if required
- One vHBA for VSAN-A FC traffic
- One vHBA for VSAN-B FC traffic

Figure 16 Cisco UCS - Server Interface Design for FC-based Storage

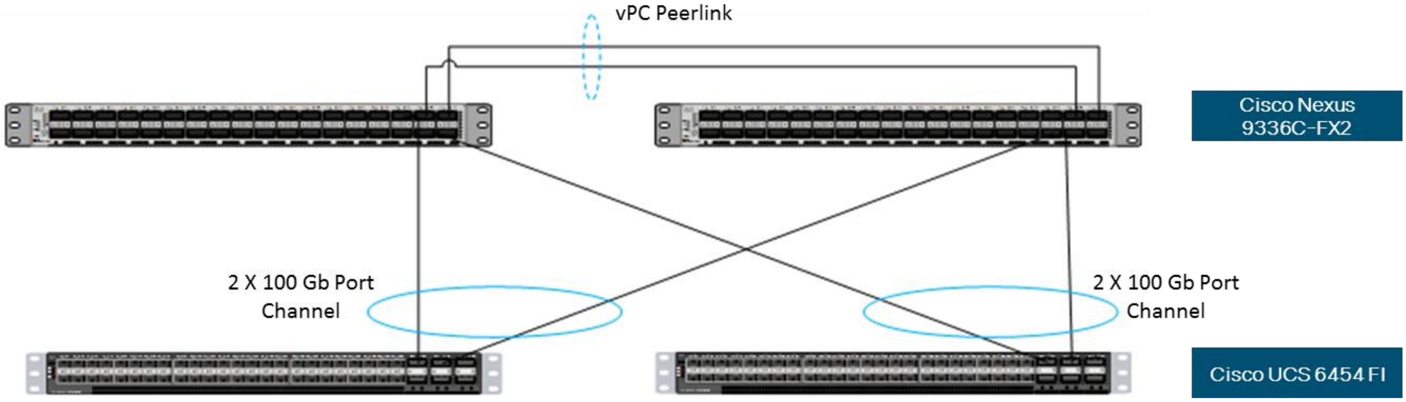


Network Connectivity

In this design, a pair of redundant Cisco Nexus 9336C-FX2 switches provide Ethernet switching fabric for iSCSI storage and application communication including communication with the existing enterprise networks. Similar to previous versions of VersaStack, the core network constructs such as virtual port channels (vPC) and VLANs plays an important role in providing the necessary Ethernet based IP connectivity.

Virtual Port-Channel Design Network reliability is achieved through the configuration of virtual Port Channels within the design as shown in Figure 17.

Figure 17 Network Connectivity



Virtual Port Channel allows Ethernet links that are physically connected to two different Cisco Nexus 9336C-FX2 switches to appear as a single Port Channel. vPC provides a loop-free topology and enables fast convergence if either one of the physical links or a device fails. In this design, the 100G ports from the 40/100G ports on the 6454 (1/49-54) were used for the virtual port channels.

vPC on Nexus switches running in NXOS mode requires a peer-link to be explicitly connected and configured between peer-devices (Nexus 9000 switch pair). In addition to the vPC peer-link, the vPC peer keepalive link is a required component of a vPC configuration. The peer keepalive link allows each vPC enabled switch to monitor the health of its peer. This link accelerates convergence and reduces the occurrence of split-brain scenarios. In this validated solution, the vPC peer keepalive link uses the out-of-band management network.

VLAN Design

To enable connectivity between compute and storage layers of the VersaStack and to provide in-band management access to both physical and virtual devices, several VLANs are configured and enabled on various paths.

The VLANs configured for the infrastructure services include:

- iSCSI VLANs to provide access to iSCSI datastores including boot LUNs
- Management and vMotion VLANs used by compute and vSphere environment
- Application VLANs used for application and virtual machine communication

Table 4 lists various VLANs configured for setting up the VersaStack environment.

Table 4 VersaStack VLAN Usage

VLAN ID Name Usage	VLAN ID Name Usage	VLAN ID Name Usage
2	Native VLAN	VLAN 2 used as Native VLAN instead of default VLAN (1)
11	IB-MGMT-VLAN	Management VLAN to access and manage the servers
3174	Data-Traffic	VLAN to carry data traffic for both VM and bare-metal Servers
3173	vMotion	VMware vMotion traffic
3161 (Fabric-A)	iSCSI-A	iSCSI-A path for booting both B Series and C Series servers and datastore access

3162 (Fabric-B)	iSCSI-B	iSCSI-B path for booting both B Series and C Series servers and datastore access

IBM Storage Systems

IBM FlashSystem 9100 covered in this VersaStack design, is deployed as high availability storage solution. IBM storage systems support fully redundant connections for communication between control enclosures, external storage, and host systems.

Each storage system provides redundant controllers and redundant iSCSI and FC paths to each controller to avoid failures at path as well as hardware level. For high availability, the storage systems are attached to two separate fabrics, SAN-A and SAN-B. If a SAN fabric fault disrupts communication or I/O operations, the system recovers and retries the operation through the alternative communication path. Host (ESXi) systems are configured to use ALUA multi-pathing, and in case of SAN fabric fault or node canister failure, the host seamlessly switches over to alternate I/O path.

IBM FlashSystem 9100 Storage

A basic configuration of an IBM FlashSystem 9100 storage platform consists of one IBM FlashSystem 9100 Control Enclosure. For a balanced increase of performance and scale, up to four IBM FlashSystem 9100 Control Enclosures can be clustered into a single storage system, multiplying performance and capacity with each addition.

The IBM FlashSystem 9100 Control Enclosure node canisters are configured for active-passive redundancy. The node canisters run a highly customized Linux -based OS that coordinates and monitors all significant functions in the system. Each Control Enclosure is defined as an I/O group and can be visualized as an isolated appliance resource for servicing I/O requests.

In this design guide, one pair of FS9100 node canisters (I/O Group 0) were deployed with in a single FS9100 Control Enclosure. The storage configuration includes defining logical units with capacities, access policies, and other parameters.

> Based on the specific storage requirements and scale, the number of I/O Groups in customer deployments will vary.

IBM FlashSystem 9100 – iSCSI Connectivity

To support iSCSI-based IP storage connectivity, each IBM FS9100 node canister is connected to each of the Cisco Nexus 9336C-FX2 switch for iSCSI boot and VMware datastore access. The physical connectivity is shown in Figure 18. Two 25GbE ports from each IBM FS9100 are connected to each of the two Cisco Nexus 9336CFX2 switches providing an aggregate bandwidth of 100Gbps. In this design, 25Gbps Ethernet ports between the FS9100 I/O Group and the Nexus fabric are utilized by redundant iSCSI-A and iSCSI-B paths, providing redundancy for link and device failures. Additional ports can be easily added for additional bandwidth if needed.

The Nexus 9336C-FX2 switches used in this design support 10/25/40/100 Gbps on all the ports. The switch supports breakout interfaces, each 100Gbps port on the switch can be split in to 4 X 25Gbps interfaces. The QSFP breakout cable has been leveraged to connect 25Gbps iSCSI ethernet ports on the FS9100 storage array to the 100Gbps QSFP port on the switch end. With this connectivity, IBM SFP transceiver on the FS9100 are not required.

> Connectivity between the Nexus switches and IBM FS9100 for iSCSI access depends on the Nexus 9000 switch model used within the architecture. If any supported Nexus switch with 25Gbps capable SFP ports is used, breakout cable is not required and ports from the switch to IBM FS9100 can be connected directly using the SFP transceivers on both sides.

Figure 18 IBM FS9100 - iSCSI Connectivity with Nexus 9336C-FX2 Switch

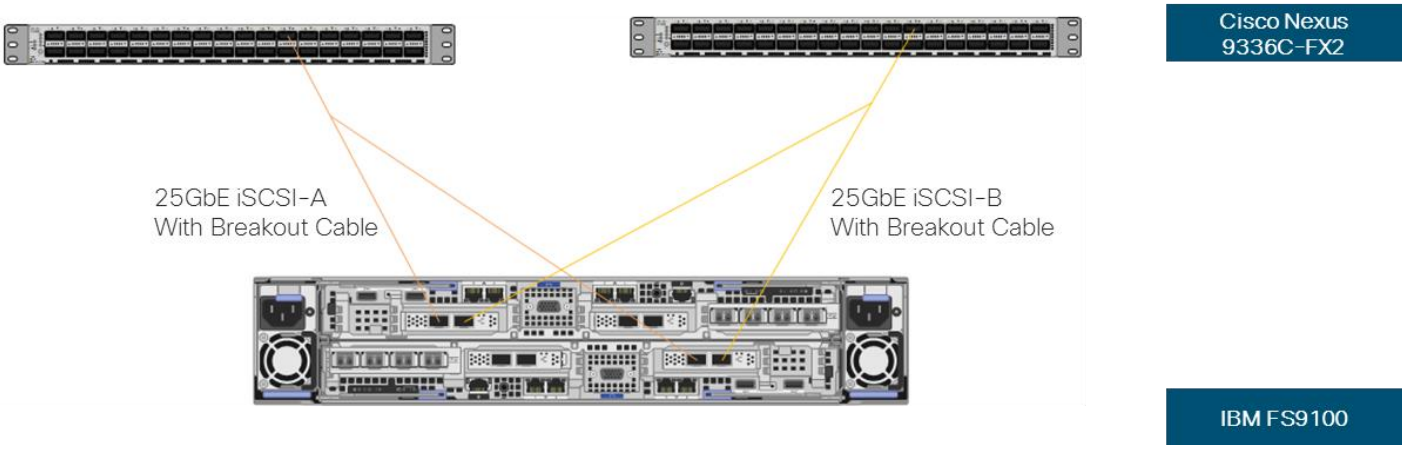
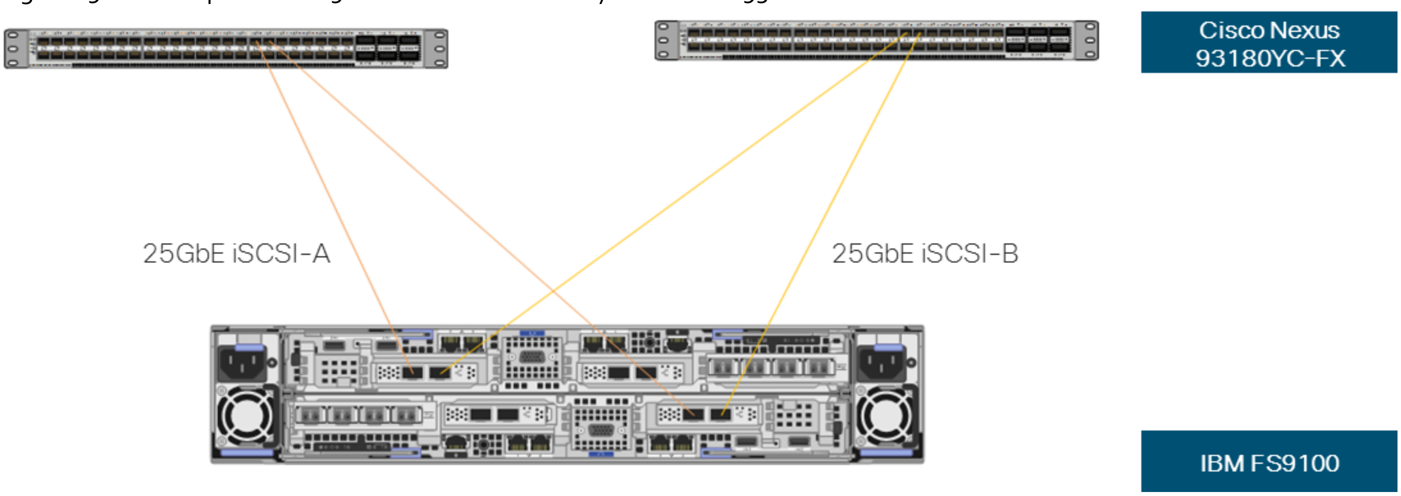


Figure 19 illustrates direct connectivity using SFP transceivers between 93180YC-FC switches and IBM FS9100. Other models of Nexus 9000 series switches with SFP ports can also be used for direct connectivity with the FS9100 storage array.

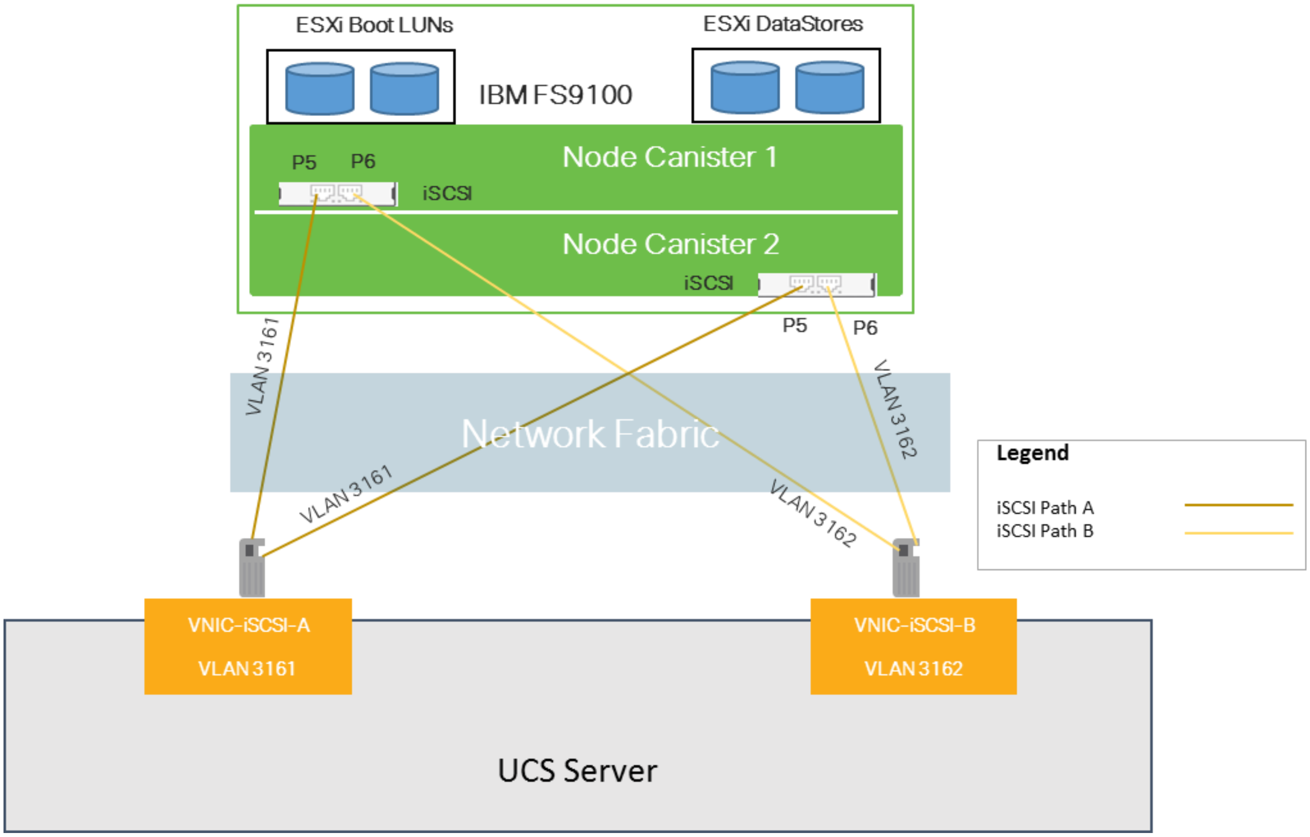
Figure 19 Example: IBM FS9100 – iSCSI Connectivity with Nexus 93180YC-FX Switch



To provide redundant iSCSI paths, two VMkernel interfaces are configured to use dedicated NICs for host to storage connectivity. In this configuration, each VMkernel port provided a different path that the iSCSI storage stack and its storage-aware multi-pathing plug-ins can use.

To setup iSCSI-A path between the ESXi hosts and the IBM FS9100 node canisters, VLAN 3161 is configured on the Cisco UCS, Cisco Nexus and on the IBM FS9100 interfaces. To setup iSCSI-B path between the ESXi hosts and the IBM FS9100, VLAN 3162 is configured on the Cisco UCS, Cisco Nexus and on the appropriate IBM FS9100 node interfaces. Within Cisco UCS service profile, these VLANs are enabled on vNIC-iSCSI-A and vNIC-iSCSI-B interfaces respectively. The iSCSI VLANs are set as native VLANs on the vNICs to enable boot from SAN functionality. Figure 20 shows the iSCSI connectivity between the UCS server and IBM storage array.

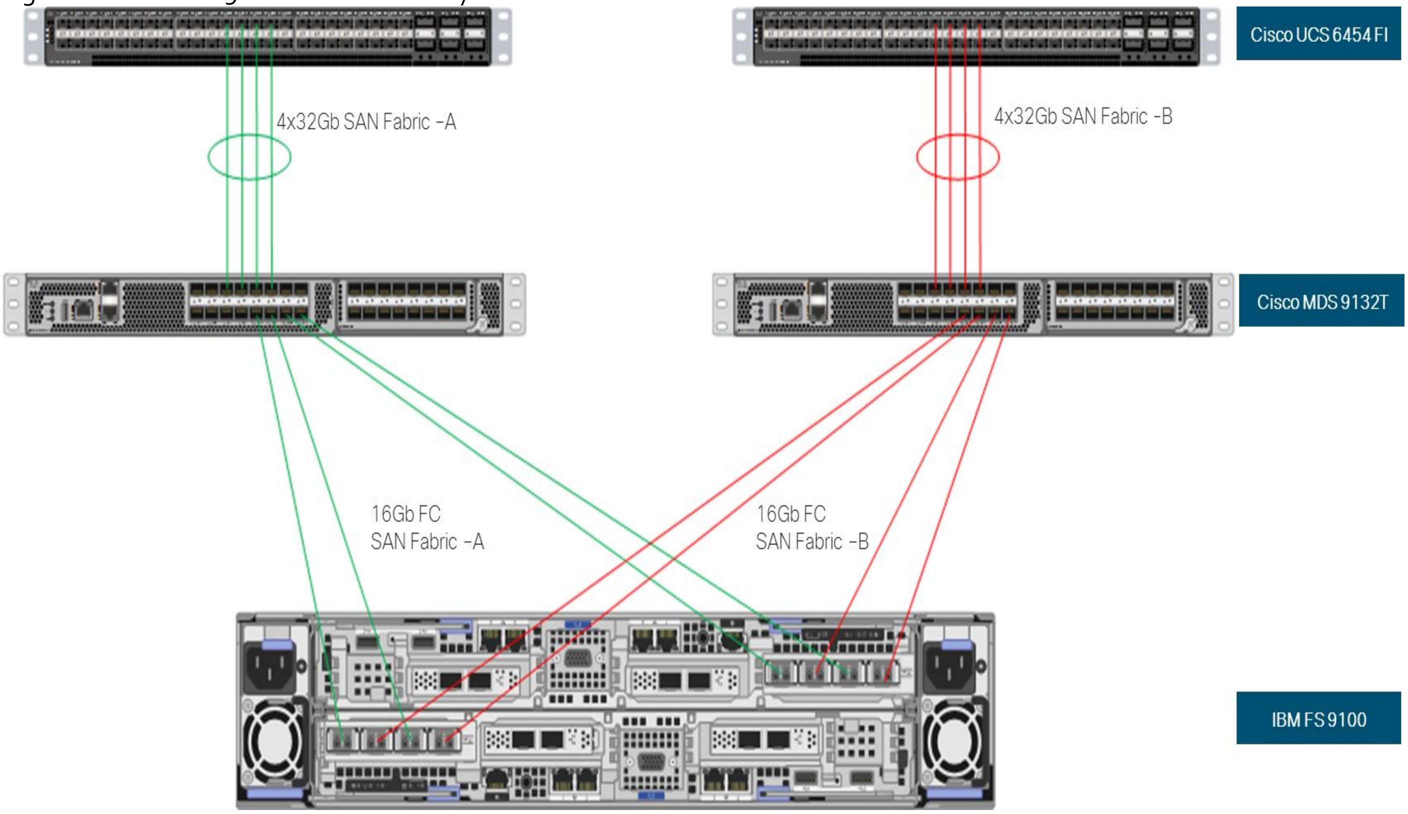
Figure 20 Network Design – VLAN Mapping for iSCSI Storage Access



IBM FlashSystem 9100 - FC Connectivity

To support FC based storage connectivity, each IBM FS9100 node canister is connected to Cisco UCS Fabric Interconnect using redundant SAN Fabric configuration provided by two Cisco MDS 9132T switches. Figure 21 shows the resulting FC switching infrastructure.

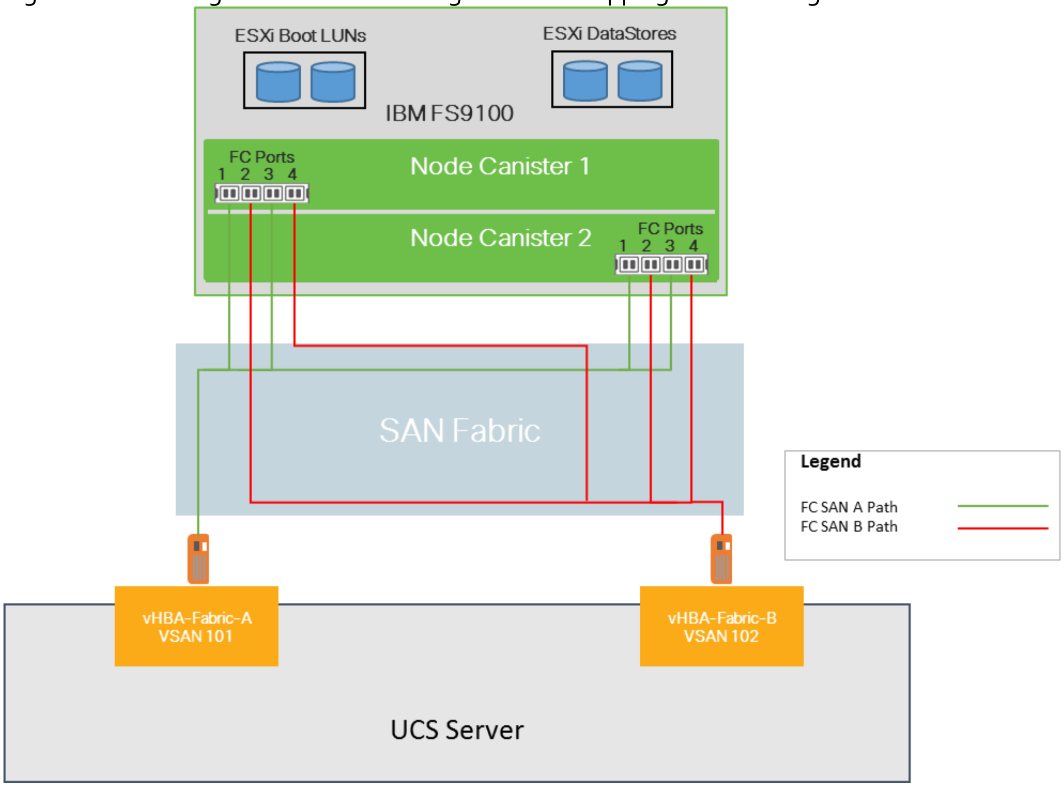
Figure 21 IBM FS9100 - FC Connectivity



In the FC design, the storage controller is connected to a Cisco MDS 9132T SAN switching infrastructure for FC boot and VMware data store access, isolated fabrics (A and B) are created on the Cisco MDS 9132T switches, VSAN 101 and VSAN 102 provide the dual SAN paths. Use of SAN zoning on Cisco MDS switches allows the isolation of traffic within specific portions of the storage area network. Cisco UCS FIs connect to the Cisco MDS switches using Port Channels while IBM FS9100 controllers are connected using independent FC ports.

The IBM FS9100 use NPIV to virtualize WWPNs, which increases redundancy during firmware updates and scheduled maintenance where WWPNs transparently move to the controller that is not being maintained. As a consequence, FC-attached hosts experience zero path reduction during controller outages. The IBM FS 9100 system has NPIV enabled by default. For the FC design, the FC HBA ports operate in FC target mode and the ports are connected to two SAN fabrics. This method provides increased redundancy to make sure that the paths from the host to its storage LUNs are always available. Figure 22 shows the fibre channel connectivity between the UCS server and IBM storage array.

Figure 22 Storage Area Network Design – VSAN Mapping for FC Storage Access



Host Clusters

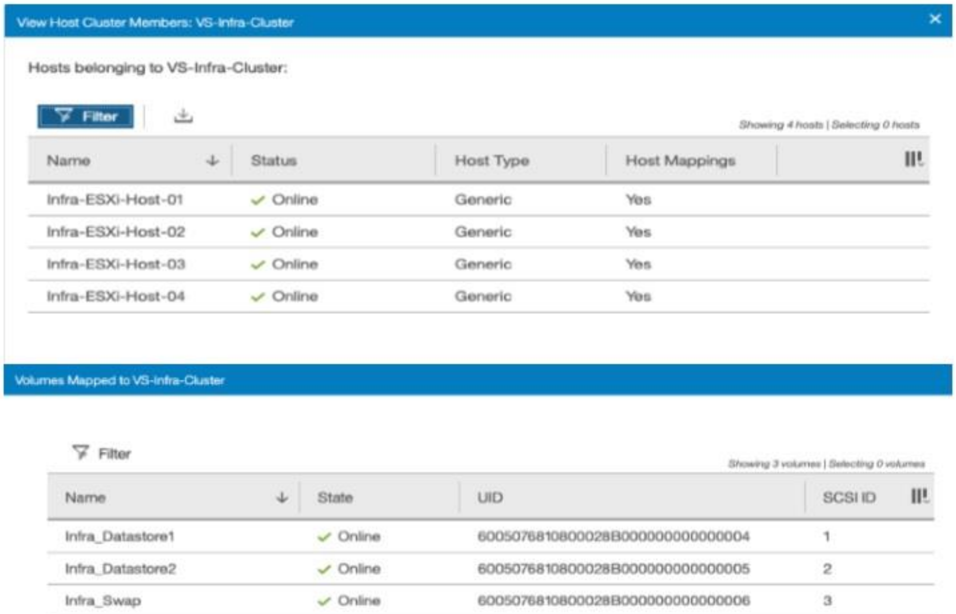
When managing how volumes are mapped to Hosts, IBM Spectrum Virtualize incorporates the concept of Hosts and Host Clusters. In VersaStack configuration, each VMware ESXi (or physical server) instance should be defined as an independent Host object within FS9100. If each VMware ESXi host has multiple associated FC WWPN ports (when using FibreChannel) or IQN ports when using iSCSI, it is recommended that all ports associated with each physical host be contained within a single host object.

When using vSphere clustering where storage resources (data stores) are expected to be shared between multiple VMware ESXi hosts, it is recommended that a Host Cluster be defined for each vSphere cluster. When mapping volumes from the FS9100 designed for VMFS Datastores, shared Host Cluster mappings should be used. The benefits are as follows:

- All members of the vSphere cluster will inherit the same storage mappings
- SCSI LUN IDs are consistent across all members of the vSphere cluster
- Simplified administration of storage when adding/removing vSphere cluster members
- Better visibility of the host/hostcluster state if particular ports/SAN become disconnected

However, when using SAN boot volumes, ensure that these are mapped to the specific host via private mappings. This will ensure that they remain accessible to only the corresponding VMware ESXi host.

Figure 23 Host Cluster and Volume Mappings



View Host Cluster Members: VS-Infra-Cluster

Hosts belonging to VS-Infra-Cluster:

Filter

Showing 4 hosts | Selecting 0 hosts

Name	Status	Host Type	Host Mappings
Infra-ESXi-Host-01	Online	Generic	Yes
Infra-ESXi-Host-02	Online	Generic	Yes
Infra-ESXi-Host-03	Online	Generic	Yes
Infra-ESXi-Host-04	Online	Generic	Yes

Volumes Mapped to VS-Infra-Cluster

Filter

Showing 3 volumes | Selecting 0 volumes

Name	State	UID	SCSI ID
Infra_Datastore1	Online	6005076810800028B000000000000004	1
Infra_Datastore2	Online	6005076810800028B000000000000005	2
Infra_Swap	Online	6005076810800028B000000000000006	3

ESXi Host Virtual Switching Architecture

Application deployment utilizes port groups on VMware distributed switch (VDS). However, for some of the core connectivity such as out of band management access, vSphere vMotion and storage LUN access using iSCSI, vSphere vSwitches are deployed. To support this multi-vSwitch requirement, multiple vNIC interfaces are setup in Cisco UCS services profile and storage, management and VM data VLANs are then enabled on the appropriate vNIC interfaces. Figure 24 shows the distribution of VMkernel ports and VM port-groups on an iSCSI connected ESXi server. For an ESXi server, supporting iSCSI-based storage access, In-band management and vMotion traffic is handled by dedicated vSwitches, iSCSI-A and iSCSI-B traffic are also handled by two dedicated iSCSI vSwitches. The resulting ESXi host configuration is therefore a combination of four vSwitches and a single distributed switch for handles application specific traffic.

Figure 24 ESXi Host vNIC and vmk Distribution for iSCSI Based Storage Access

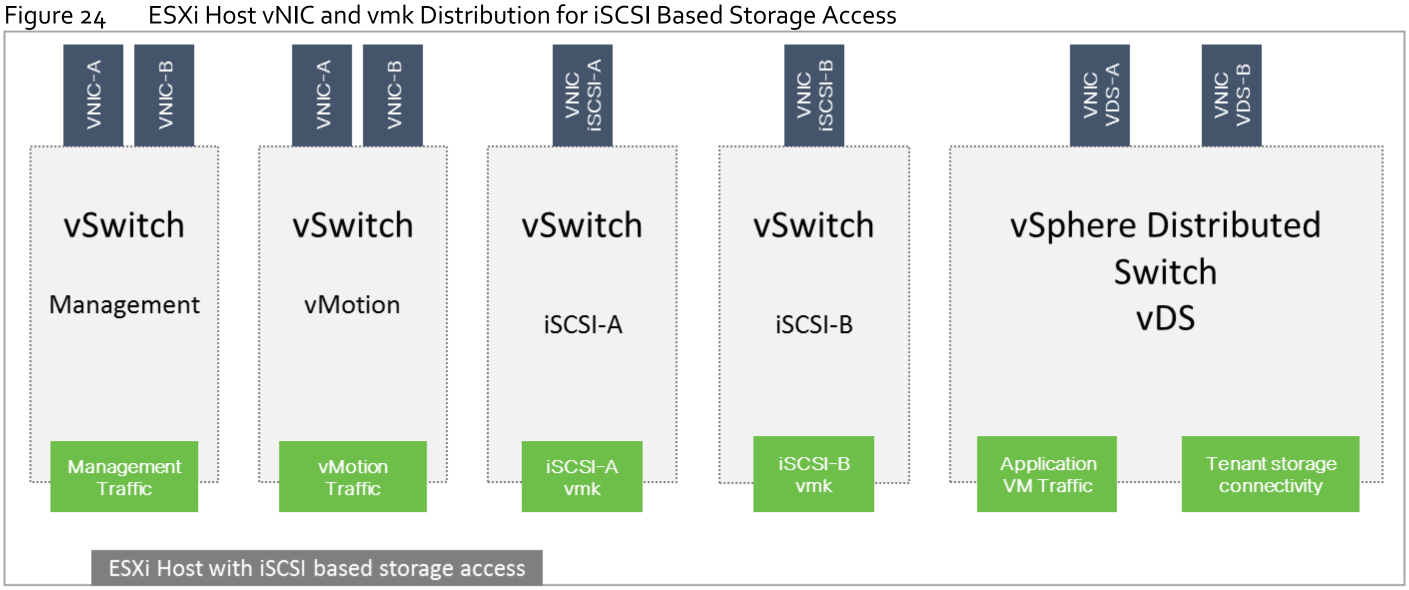
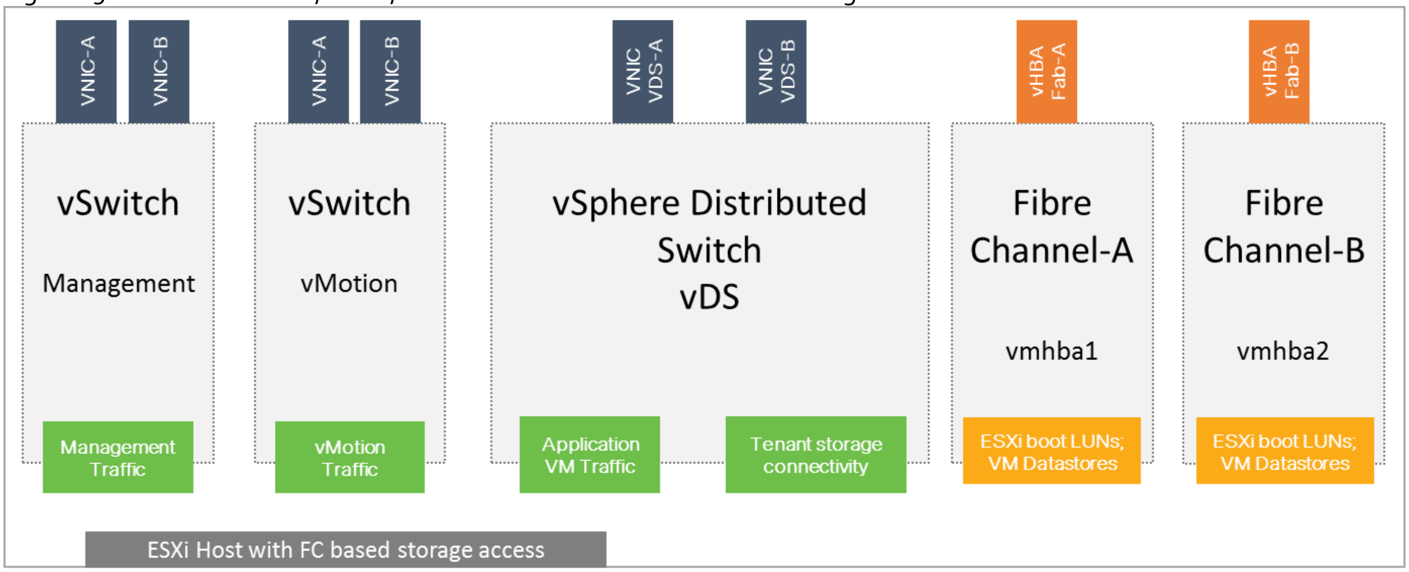


Figure 25 shows the virtual networking on an FC connected ESXi host. In this case, In-band management and vMotion traffic is handled by two dedicated vSwitches but the two vSwitches for iSCSI traffic are not needed. The Fibre Channel SAN-A and SAN-B traffic is handled by two dedicated vHBAs. The resulting ESXi host configuration therefore has a combination of one vSwitch and one (vDS) distributed switch.

Figure 25 ESXi Host vNIC, vHBA, and vmk Distribution for FC Based Storage Access

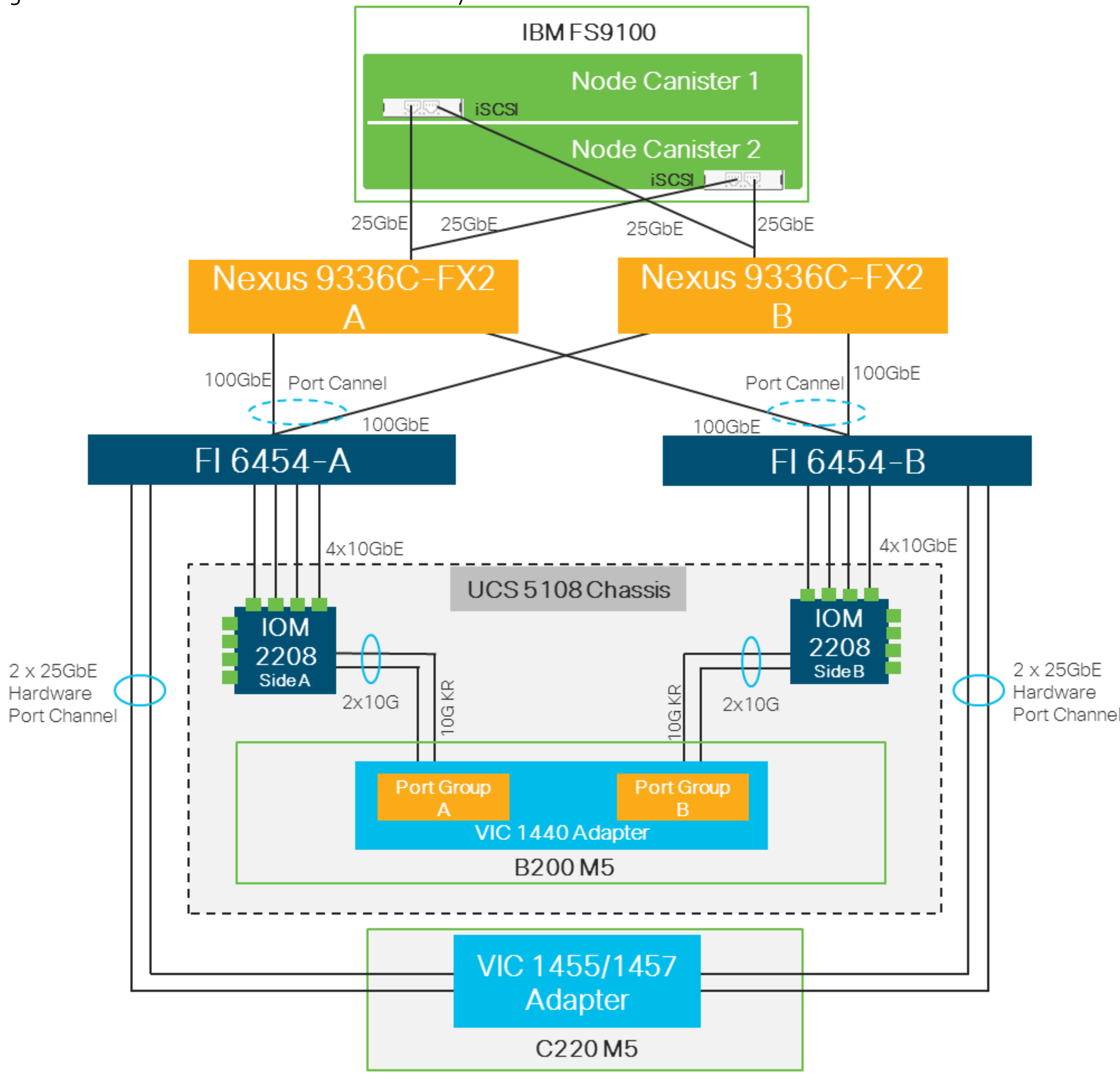


End-to-End IP Network Connectivity

The Cisco Nexus 9336C-FX2 used in this design supports flexible port speeds and provides 25Gbps connectivity to the IBM FS9100 storage controllers for iSCSI traffic and 100Gbps connectivity to the Cisco UCS FIs for iSCSI storage and application traffic.

- The Cisco UCS M5 Servers are equipped with a VIC 1400 Series adapter
- In the Cisco B200 M5 server, a VIC 1440 provides 2x10Gbps to IOM A and 2x10Gbps to IOM B via the Cisco UCS Chassis 5108 chassis backplane
- In the Cisco C220 M5 server, a VIC 1457 is used with 2x 25Gbps connections port-channeled to FI-A and 2x 25Gbps connections port-channel to FI-B.
- Each IOM is connected to its respective Cisco UCS 6454 Fabric Interconnect using a port-channel for 4-8 links.
- Each Cisco UCS 6454 FI connects to the Nexus 9336C-FX2 through 2X 100Gbps virtual port channels.
- The IBM FS9100 is connected to both Nexus 9336C-FX2 switches using QSFP100G-4SFP25G cable with 4X 25 Gbps connections to provide redundant paths.

Figure 26 End-to-End IP Network Connectivity



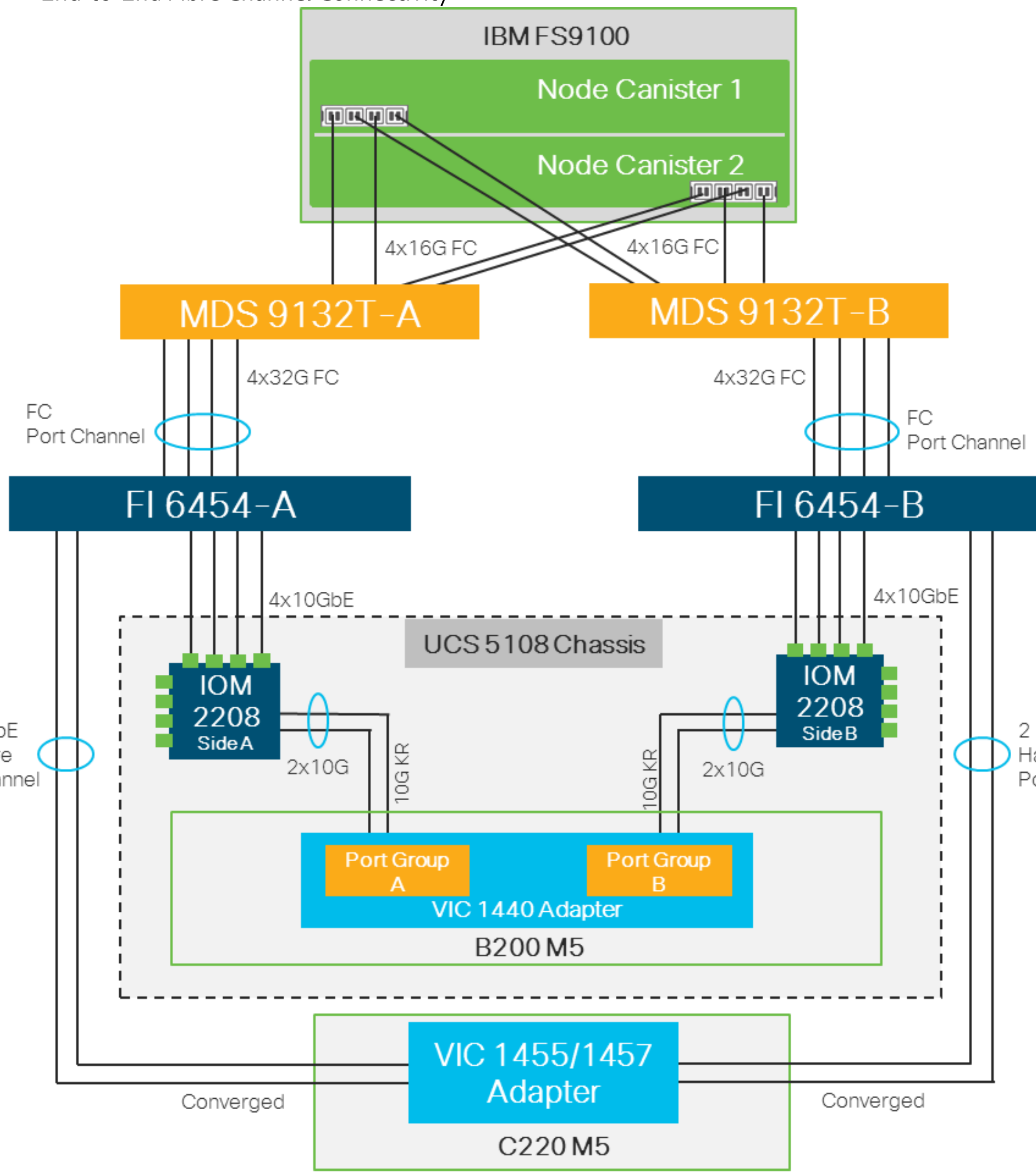
End-to-End Fibre Channel Connectivity

The Cisco MDS 9132T provides high-speed Fibre Channel connectivity within the design with redundant Fibre Channel links between the MDS 9132Ts and the Cisco UCS 6454 FIs as detailed below:

- The Cisco UCS M5 Servers are equipped with a VIC 1400 Series adapter
- In the Cisco B200 M5 server, a VIC 1440 provides 2x10Gbps to IOM A and 2x10Gbps to IOM B via the Cisco UCS Chassis 5108 chassis backplane
- In the Cisco C220 M5 server, a VIC 1457 is used with 2x 25Gbps connections port-channeled to FI-A and 2x 25Gbps connections port-channel to FI-B
- Each IOM is connected to its respective Cisco UCS 6454 Fabric Interconnect using a port-channel for 4-8 links

- Each Cisco UCS 6454 FI connects to the MDS 9132T for the respective SAN fabric using an F-Port channel
- The IBM FS9100 is connected to both MDS 9132T switches to provide redundant paths through both fabrics

Figure 27 End-to-End Fibre Channel Connectivity



Design Considerations

VersaStack designs incorporate connectivity and configuration best practices at every layer of the stack to provide a highly available best performing integrated system. VersaStack is a modular architecture that allows customers to adjust the individual components of the system to meet their particular scale or performance requirements. This section describes some of the design considerations for the current design and a few additional design selection options available to the customers.

VersaStack Scalability Considerations

VersaStack can be scaled-up or scaled-out depending on the needs of your business, applications, workloads, and users.

Cisco UCS

Cisco UCS policy-based management allows faster deployment of new capacity for rapid and accurate scalability. Scaling up is a huge benefit to many customers, due to the ease of the Cisco UCS Service Profiles. When a new generation, faster, more powerful server is needed, the previous server's service profile need only be applied to the new server. The new server appears to both the OS and hypervisor exactly as the older server, except more resources are available.

Cisco UCS scales linearly, with no performance degradation, the limits of scaling are found only in the port count of the Fabric Interconnects. UCS environment in VersaStack with a pair of Cisco UCS 6454 Fabric Interconnects can scale up to 20 Chassis and up to 160 blades in a single Cisco UCS management domain. Cisco UCS can scale to multiple Cisco UCS Domains with Cisco UCS Central or Cisco Intersight within and across data centers globally.

Cisco Network Fabric

Cisco Nexus 9000 series and Cisco MDS 9000 switches are built on modern system architectures designed to provide high performance and meet the evolving needs of highly scalable data centers and growing enterprises.

The Cisco Nexus 9300-FX2 switches used in this design offer a variety of interface options to transparently migrate and scale as needed from 1Gbps, and 10Gbps speeds to 25Gbps at the server and storage, and from 10Gbps and 40Gbps speeds to 50Gbps and 100Gbps at the aggregation layer.

The next-generation Cisco® MDS 9000 32Gbps Fibre Channel Switches provides high-speed Fibre Channel connectivity for All-Flash arrays. As with the Nexus 9000 switches, the MDS switch ports support lower-speed (4- or 8- or 16-Gbps) giving you the option to start with lower bandwidth connections and upgrade to 32Gbps connectivity as needed using the same ports. Within VersaStack design, the MDS switch deployed in N_Port ID Virtualization (NPIV) core mode with Cisco UCS Fabric Interconnects connecting to it in N_Port Virtualization (NPV) mode. This enables the device ports to scale very cost-effectively with time without adding the burden of managing the additional NPV switches (Cisco UCS FI's) in SAN Fabric.

Link aggregation using port channels and virtual port channels have been used throughout the design for higher bandwidth and availability. This helps LAN and SAN networks scale seamlessly with the same VersaStack fundamental design. When more bandwidth is required between the devices to support workloads, additional ports can be easily added to the existing port channels.

IBM FS9100 Storage

IBM FlashSystem 9100 system has a scalable architecture that enables flash capacity to be added (scaled up) to support multiple applications. The virtualized system can also be expanded (scaled-out) to support higher IOPS and bandwidth, or the solution can be simultaneously scaled up and out to improve capacity, IOPS, and bandwidth while maintaining MicroLatency.

A single IBM FlashSystem 9100 storage system consists of one control enclosure with internal storage, representing a 2U building block. The FlashSystem 9100 control enclosure can support multiple attached expansion enclosures. Expansion enclosures can be dynamically added with virtually no downtime.

For balanced increase of performance and scale, up to four IBM FlashSystem 9100 control enclosures can be clustered into a single storage system, multiplying performance and capacity with each addition. Clustering FlashSystem 9100 will scale the performance with additional NVMe storage. With four-way system clustering, the size of the system can be increased to a maximum of 3,040 drives.

Management Network Design

This VersaStack design uses two different networks to manage the solution:

- An out-of-band management network to configure and manage physical compute, storage and network components in the solution. Management ports on each physical device (Cisco UCS FI, IBM Storage Controllers, Cisco Nexus and Cisco MDS switches) in the solution are connected to a separate, dedicated management switch.
- An in-band management network to manage and access ESXi servers and VMs (including infrastructure VMs hosted on VersaStack). Cisco UCS does allow out-of-band management of servers and VMs by using disjoint layer 2 feature but for the current design, In-Band manageability is deemed sufficient.

If a Disjoint Layer-2 feature is used, additional uplink port(s) on the Cisco UCS FIs are required to connect to the management switches and additional vNICs have to be associated with these uplink ports. The additional vNICs are necessary since a server vNIC cannot be associated with more than one uplink.

Jumbo Frame Configuration

Enabling jumbo frames in a VersaStack environment optimizes throughput between devices by enabling larger size frames on the wire while reducing the CPU resources to process these frames. VersaStack supports wide variety of traffic types (vMotion, NFS, iSCSI, control traffic, and so on) that can benefit from a larger frame size. In this validation effort the VersaStack was configured to support jumbo frames with an MTU size of 9000. In VMware vSphere, the jumbo frames are configured by setting MTU sizes at both vSwitches and VMkernel ports. On IBM storage systems, the interface MTUs are modified to enable the jumbo frame.

When setting the Jumbo frames, it is important to make sure MTU settings are applied uniformly across the stack to prevent fragmentation and the negative performance.

Compute and Virtualization High Availability Considerations

In the VersaStack solution, each ESXi server is deployed using vNICs and vHBAs that provide redundant connectivity to the unified fabric. All of the server NICs and HBAs are configured to use both the Cisco UCS FIs to avoid traffic disruptions.

VMware vCenter is used to deploy VMware HA clusters to allow VMs to failover in the event of a server failure. VMware vMotion and VMware HA are enabled to auto restart VMs after a failure. Host Monitoring is enabled to monitor heartbeats of all ESXi hosts in the cluster for faster detection. Admission Control is also enabled on the blade servers to ensure the cluster has enough resources to accommodate a single host failure.

VMware vSphere hosts use SAN multi-pathing to access LUNs on the IBM storage devices. If any component (NIC, HBA, FEX, FI, Cisco MDS, IBM controller, cables, and so on) along a path fails, all storage traffic will reroute to an alternate path. When both paths are active, traffic is load balanced.

IBM FS9100 Storage Considerations

Based on the workload requirements, the FS9100 system should be sized with appropriate cache and I/O cards and ports. Configure a balanced system with performance and capacity targeted volumes and spread the resources by using multiple volumes and combining them at the host. If you're running many workloads, then a single volume might be good enough for each workload. If the balance of the system is leaning towards DRP, consider two Data Reduction Pools.

Port Bandwidth

A single Fibre Channel port can deliver over 1.5 Gbps (allowing for overheads) and an FC card in each canister with 8 ports can deliver more than 12 Gbps. An NVMe device can perform at over 1 Gbps.

A single Fibre Channel port can deliver 80,000 - 100,000 IOPS with a 4 Kb block size. An FC card in each canister with 8 ports can deliver up to 800,000 IOPS. An IBM FlashSystem 9100 can support over 1.1 million 4 Kb read miss IOPS.

So, if you have more than 12 NVMe devices, use two Fibre Channel cards per container, and a third Fibre Channel card enables you to achieve up to 33 Gbps. If you want to drive more than 600,000 IOPS, use two Fibre Channel cards per container.

Cache

256 GB per system (128 GB base plus a 128 GB upgrade) is a good starting point. If you're using data reduction pool or making heavy use of copy services, add a further 128 GB per system. As your capacity increases (especially with the 19.2 TB FCM devices) add more cache to accommodate more of the working set (most accessed workloads, excluding snapshots, backups, and so on). A truly random working set might not benefit from a right-sized cache. If you're consolidating from multiple controllers, consider at least matching the amount of cache across those controllers.

Data Reduction Pools

VersaStack with the inclusion of IBM FlashSystem 9100 supports Data Reduction Pools. Data Reduction Pools are a new type of storage pool that implement several techniques, such as thin-provisioning, compression, and deduplication, to reduce the amount of physical capacity required to store data. Savings in storage capacity requirements translate into reduction in the cost of storing the data.

Compression and deduplication are not mutually exclusive, one or both or neither features can be enabled. If the volume is deduplicated and compressed, data is deduplicated first, and then compressed. Therefore, deduplication references are created on the compressed data stored on the physical domain.

Data Reduction Pools are useful if the underlying media does not have hardware acceleration, or if the goal is to ensure the most data reduction possible, by enabling deduplication.

Multiple Volumes

IBM FlashSystem FS9100 is optimized for multiple volumes, and around 30 volumes are required to unlock the maximum performance. A workload can become unnecessarily limited when backed by fewer volumes, and a single volume is limited to up to 10% of the ultimate performance. This is due to the relationship between the host interface adapter port(s), and how internal resources within Spectrum Virtualize are allocated to CPU threads/cores. Adding volumes initially scales performance linearly and enables the workload to be balanced across the ports and canisters.

Deployment Hardware and Software

Hardware and Software Revisions

Table 5 lists the hardware and software versions used for the solution validation. It is important to note that Cisco, IBM, and VMware have interoperability matrices that should be referenced to determine support for any specific implementation of VersaStack. Please visit the following links for more information:

- IBM System Storage Interoperation Center
- Cisco UCS Hardware and Software Interoperability Tool
- VMware Compatibility Guide

Table 5 Validated Hardware and Software Revisions

Layer	Device	Image	Comments
Compute	Cisco UCS Fabric Interconnects 6400 Series, Cisco UCS B200 M5 Cisco UCS C220 M5	4.0 (4c)	Includes the Cisco UCS-IOM 2208XP, Cisco UCS Manager, Cisco UCS VIC 1440 and Cisco UCS VIC 1457
	Cisco nenic Driver	1.0.29.0	Ethernet driver for Cisco VIC
	Cisco fnic Driver	4.0.0.40	FCoE driver for Cisco VIC
Network	Cisco Nexus Switches	7.0(3)I7(6)	NXOS
	Cisco MDS 9132T	8.4(1)	FC switch firmware version
Storage	IBM FlashSystem 9110	8.2.1.6	Software version
Software	VMware vSphere ESXi	6.7 update 2	Software version
	VMware vCenter	6.7 update 2	Software version

Validation

Test Plan

The solution was validated by deploying multiple VMs running tools such as IOMeter and Vdbench. The system was validated for resiliency by failing various aspects of the system under the load. The following sections provide examples of the tests executed for this solution.

Cisco UCS Validation

- Failure and recovery of links from Cisco UCS Chassis (IOM) and Cisco UCS C-Series servers to FI-A and FI-B
- Rebooting Cisco UCS FI, one at a time
- Removing the physical cables between FI and Cisco Nexus 9336C-FX2 switches to simulate path failure

Network Validation

- Fail/power off both Cisco 9336C-FX2 switches, one after other
- Failure and recovery of links with in vPC from UCS FI and Cisco Nexus 9336C-FX2 switches
- Failure and recovery of physical links from Cisco Nexus 9336C-FX2 switches and the management switch

Storage Validation

- Failure and recovery of the Cisco MDS switches, one at a time
- Failure and recovery of the links between Cisco MDS, Cisco Nexus 9336C-FX2 switches and IBM FS9100 Nodes and IBM storage controllers
- Failure and recovery of links between Cisco UCS FI and Cisco MDS

vSphere Validation

- Failure and recovery of ESXi hosts in a cluster (rebooting of hosts, shutting down of hosts, and so on)
- In case of a host failure, verify VM auto restart within the high availability cluster
- VM vMotion across ESXi servers

Summary

VersaStack delivers a platform for enterprise and cloud datacenters using Cisco UCS Blade and Rack Servers, Cisco Fabric Interconnects, Cisco Nexus 9000 switches, Cisco MDS switches, and Fibre Channel or iSCSI attached IBM Storage Arrays.

VersaStack with VMware vSphere 6.7 U2, Cisco UCS 4th Generation and IBM FS9100 NVMe-accelerated Storage architecture aligns with the converged infrastructure configurations and best practices as identified by previous VersaStack releases. The system includes hardware and software compatibility support between all components and aligns to the configuration recommendations for each of these components. The VersaStack design described in this document has been validated for resiliency (under fair load) and fault tolerance during system upgrades, component failures, and power outage scenarios.

References

Products and Solutions

Cisco Unified Computing System:

http://www.cisco.com/en/US/products/ps10265/index.html

Cisco UCS 6400 Series Fabric Interconnects:

https://www.cisco.com/c/en/us/support/servers-unified-computing/ucs-6400-series-fabric-interconnects/tsdproducts-support-series-home.html

Cisco UCS 5100 Series Blade Server Chassis:

http://www.isco.com/en/US/products/ps10279/index.html

Cisco UCS B-Series Blade Servers:

http://www.cisco.com/c/en/us/products/servers-unified-computing/ucs-b-series-blade-servers/index.html

Cisco UCS C-Series Rack Servers:

http://www.cisco.com/c/en/us/products/servers-unified-computing/ucs-c-series-rack-servers/index.html

Cisco UCS Adapters:

http://www.cisco.com/en/US/products/ps10277/prod_module_series_home.html

Cisco UCS Manager:

http://www.cisco.com/en/US/products/ps10281/index.html

Cisco Intersight:

https://www.cisco.com/c/en/us/products/servers-unified-computing/intersight/index.html

Cisco Nexus 9000 Series Switches:

http://www.cisco.com/c/en/us/support/switches/nexus-9000-series-switches/tsd-products-support-serieshome.html

Cisco Application Centric Infrastructure:

http://www.cisco.com/c/en/us/solutions/data-center-virtualization/application-centric-infrastructure/index.html

Cisco Data Center Network Manager:

https://www.cisco.com/c/en/us/products/cloud-systems-management/prime-data-center-networkmanager/index.html

Cisco UCS Director:

https://www.cisco.com/c/en/us/products/servers-unified-computing/ucs-director/index.html

VMware vCenter Server:

http://www.vmware.com/products/vcenter-server/overview.html

VMware vSphere:

https://www.vmware.com/tryvmware_tpl/vsphere-55_evalcenter.html

IBM FlashSystem 9100:

https://www.ibm.com/us-en/marketplace/flashsystem-9100

Interoperability Matrixes

Cisco UCS Hardware Compatibility Matrix:

http://www.cisco.com/c/en/us/support/servers-unified-computing/unified-computing-system/productstechnical-reference-list.html

VMware and Cisco Unified Computing System:

http://www.vmware.com/resources/compatibility

IBM System Storage Interoperation Center:

http://www-03.ibm.com/systems/support/storage/ssic/interoperability.wss

About the Authors

Sreenivasa Edula, Technical Marketing Engineer, UCS Data Center Solutions Engineering, Cisco Systems, Inc.

Sreeni is a Technical Marketing Engineer in the UCS Data Center Solutions Engineering team focusing on converged and hyper-converged infrastructure solutions, prior to that he worked as a Solutions Architect at EMC Corporation. He has experience in Information Systems with expertise across Cisco Data Center technology portfolio, including DC architecture design, virtualization, compute, network, storage and cloud computing.

Warren Hawkins, Virtualization Test Specialist for IBM Spectrum Virtualize, IBM

Working as part of the development organization within IBM Storage, Warren Hawkins is also a speaker and published author detailing best practices for integrating IBM Storage offerings into virtualized infrastructures. Warren has a background in supporting Windows and VMware environments working in second-line and third-line support in both public and private sector organizations. Since joining IBM in 2013, Warren has played a crucial part in customer engagements and, using his field experience, has established himself as the Test Lead for the IBM Spectrum Virtualize™ product family, focusing on clustered host environments

Acknowledgements

For their support and contribution to the design, validation, and creation of this Cisco Validated Design, the authors would like to thank:

- Haseeb Niazi, Technical Marketing Engineer, Cisco Systems, Inc.
- Archana Sharma, Technical Marketing Engineer, Cisco Systems, Inc.