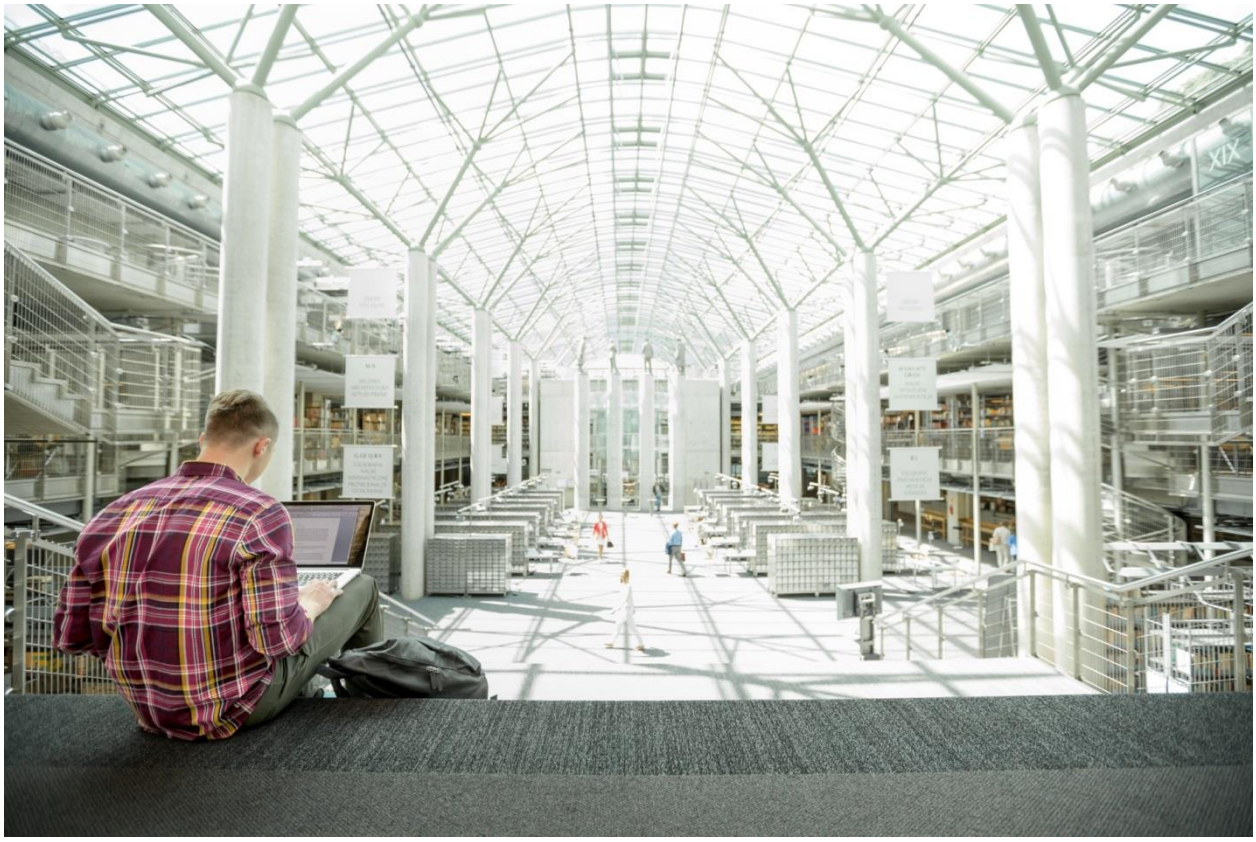


VersaStack for Data Center with Cisco Application Centric Infrastructure Design Guide

Cisco ACI and IBM FlashSystem V9000 and Storwize V7000 Unified with vSphere 6.0

Last Updated: April 20, 2017

CISCO VALIDATED DESIGN

About Cisco Validated Designs

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

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

Table of Contents

Executive Summary

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.

Customers looking to deploy applications using shared data center infrastructure face a number of challenges. A recurrent infrastructure challenge is to achieve the levels of IT agility and efficiency that can effectively meet the business objectives. Addressing these challenges requires having an optimal solution with the following key characteristics:

- Availability: Helps 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
- Efficiency: Facilitate efficient operation of the infrastructure through re-usable policies
- 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

Cisco and IBM have partnered to deliver a series of VersaStack solutions that enable strategic data center platforms with the above characteristics. VersaStack solution delivers an integrated architecture that incorporates compute, storage and network design best practices thereby minimizing IT risks by validating the integrated architecture to ensure compatibility between various components. The solution also addresses IT pain points by providing documented design guidance, deployment guidance and support that can be used in various stages (planning, designing and implementation) of a deployment.

The Cisco Application Centric Infrastructure (ACI) based VersaStack solution, covered in this CVD, delivers a converged infrastructure platform specifically designed for software defined networking (SDN) enabled data centers. The design showcases:

- ACI enabled Cisco Nexus 9000 switching architecture
- Cisco Unified Compute System (UCS) servers with Intel Broadwell processors
- Storage designs covering both IBM Storwize V7000 Unified and IBM FlashSystem V9000 storage systems
- VMware vSphere 6.0U1b hypervisor
- Cisco MDS Fibre Channel (FC) switches for SAN connectivity

Solution Overview

Introduction

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 and 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.

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

VersaStack with ACI solution is designed to simplify the data center evolution to a shared cloud-ready infrastructure based on an application driven policy model. With the integration of Cisco ACI to the VersaStack platform, the solution delivers an application centric architecture with centralized automation that combines software flexibility with the hardware performance.

Audience

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

What's New?

The following design elements distinguish this version of VersaStack from previous models:

- Validation of the Cisco ACI release 1.3 with IBM FlashSystem V9000 and IBM Storwize V7000 Unified
- Support for the Cisco UCS release 3.1 and Intel Broadwell based M4 servers
- Support for latest IBM software release 7.6.0.4 and IBM File Module software release 1.6.1
- Validation of IP-based storage design supporting both NFS and iSCSI based storage access
- Support for Fiber Chanel storage utilizing Cisco MDS 9148S
- Application design guidance for multi-tiered application using Cisco ACI application profiles and policies
- Support for application segregation utilizing ACI multi-tenancy
- Integration of Cisco ASA firewall appliance for enhanced application security

For more information on previous VersaStack models, please refer the VersaStack guides at:

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. This portfolio will include, but is not limited to the following items:

- Best practice architectural design
- Implementation and deployment instructions
- Technical specifications (rules for what is, and what is not, a VersaStack configuration)
- Cisco Validated Designs (CVDs) and IBM Redbooks focused on a variety of use cases

Cisco and IBM have also built a robust and experienced support team focused on VersaStack solutions. The team includes customer account and technical sales representatives as well as professional services and technical support engineers. The support alliance between IBM and Cisco provides customers and channel services partners direct access to technical experts who are involved in cross vendor collaboration and have access to shared lab resources to resolve potential multi-vendor issues.

Technology Overview

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

- Cisco Unified Computing System (Cisco UCS)
- Cisco Nexus and Cisco MDS Switches
- IBM FlashSystem and IBM Storwize family storage

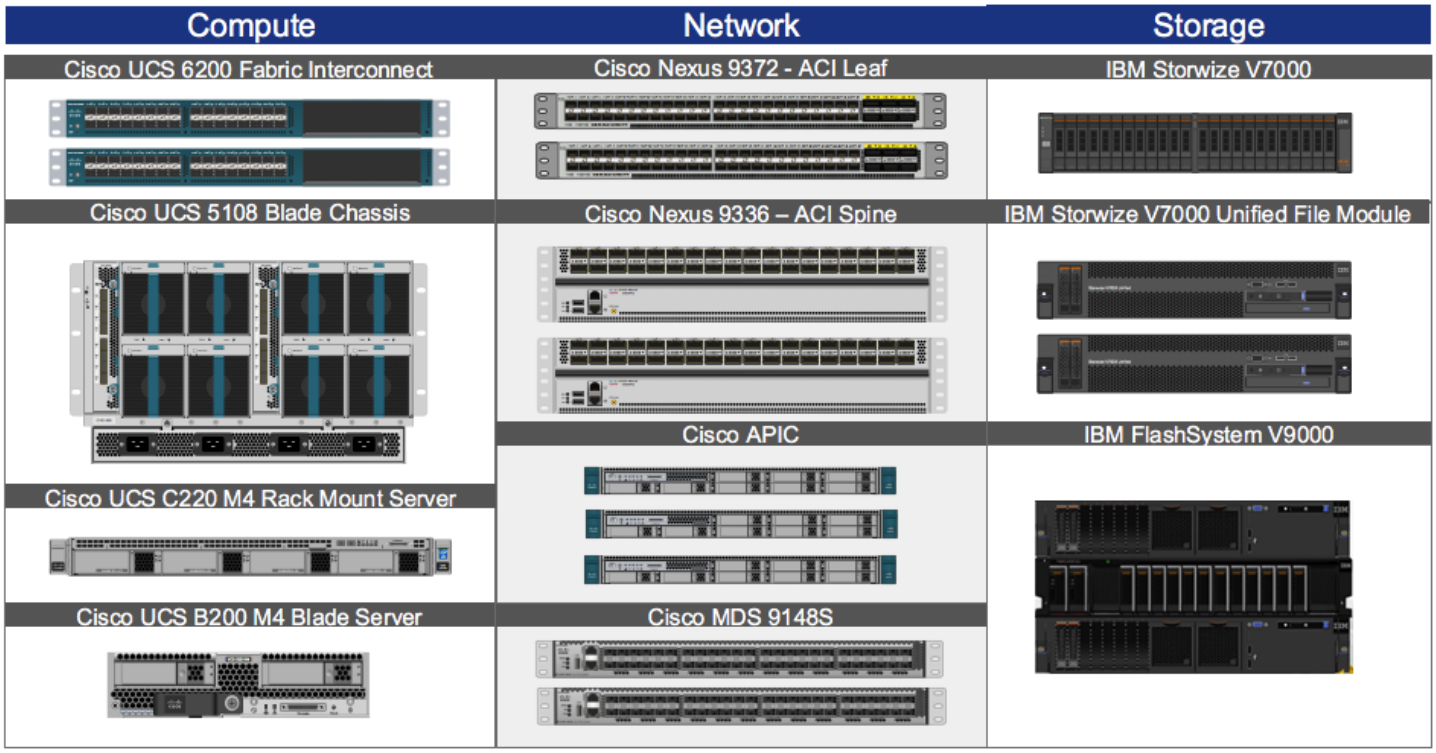
These components are connected and configured according to best practices of both Cisco and IBM and provide an ideal platform for running a variety of workloads with confidence. The reference architecture covered in this document leverages:

- Cisco UCS 5108 Blade Server chassis with Cisco UCS 2200 Series Fabric Extenders (FEX)
- Cisco UCS B-Series Blade servers
- Cisco UCS C-Series Rack Mount servers
- Cisco UCS 6200 Series Fabric Interconnects (FI)
- Cisco Application Policy Infrastructure Controllers (APIC)
- Cisco Nexus 9336 ACI Spine Switches
- Cisco Nexus 9372 ACI Leaf Switches
- Cisco MDS 9148S Fabric Switches
- IBM FlashSystem V9000*
- IBM Storwize V7000 Unified*
- VMware vSphere 6.0
- Cisco Application Virtual Switch (AVS) **

* This design guide covers two unique storage configuration options: Option 1 utilizes IBM Storwize V7000 unified to support a hybrid Fibre Channel and NFS connectivity design; Option 2 utilizes IBM FlashSystem V9000 to support an iSCSI based IP-only storage design.

** This design guide covers both VMware Virtual Distributed Switch (VDS) and Cisco AVS.

Figure 1 VersaStack with ACI – Components



One of the key benefits of VersaStack is the ability to maintain consistency at both scale up and scale down models. VersaStack can scale up for greater performance and capacity. In other words, you can add compute, network, or storage resources as needed; or it can also scale out where you need multiple consistent deployments like rolling out additional VerstaStack modules. Each of the component families shown in Figure 1 (Cisco Unified Computing System, Cisco Switches, and IBM storage arrays) offer platform and resource options to scale the infrastructure up or down while supporting the same features and functionality.

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

Cisco Unified Compute System

The Cisco Unified Computing System (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 10 Gigabit Ethernet 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 in the VersaStack architecture utilizes the following components:

- Cisco UCS Manager (UCSM) provides unified management of all software and hardware components in the Cisco UCS to manage servers, networking, and storage configurations. The system uniquely integrates all the system components, enabling the entire solution to be managed as a single entity through Cisco UCSM software. Customers can interface with Cisco UCSM through 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 6200 Series Fabric Interconnects is a family of line-rate, low-latency, lossless, 10-Gbps Ethernet, Fibre Channel and Fibre Channel over Ethernet interconnect switches providing the management and communication backbone for the Cisco UCS. The unified fabric lowers costs by reducing the number of network adapters, switches, and cables, and by decreasing the power and cooling requirements.
- Cisco UCS 5108 Blade Server Chassis supports up to eight blade servers and up to two fabric extenders in a six-rack unit (RU) enclosure.
- Cisco UCS B-Series Blade Servers provide performance, efficiency, versatility and productivity with the latest Intel based processors.
- Cisco UCS C-Series Rack Mount Servers deliver unified computing innovations and benefits to rack servers with performance and density to support a wide range of workloads.
- Cisco UCS Network Adapters provide wire-once architecture and offer a range of options to converge the fabric, optimize virtualization and simplify management.

Cisco UCS Management

Cisco's Unified Compute System has revolutionized the way servers are managed in the data center. Some of the unique differentiators of Cisco UCS and Cisco UCS Manager are:

- Embedded Management –In Cisco UCS, the servers are managed by the embedded firmware in the Fabric Interconnects, eliminating need for any external physical or virtual devices to manage the servers.
- Auto Discovery –By simply inserting the blade server in the chassis or connecting rack server to the fabric interconnect, discovery and inventory of compute resource occurs automatically without any management intervention.
- Policy Based Resource Classification –Once a compute resource is discovered by Cisco UCS Manager, it can be automatically classified based on defined policies. This capability is useful in multi-tenant cloud computing.
- Combined Rack and Blade Server Management –Cisco UCS Manager can manage B-series blade servers and C-series rack server under the same Cisco UCS domain. This feature, along with stateless computing makes compute resources truly hardware form factor agnostic.
- Model based Management Architecture –Cisco UCS Manager architecture and management database is model based and data driven. An open XML API is provided to operate on the management model. This enables easy and scalable integration of Cisco UCS Manager with other management systems.
- Service Profiles and Stateless Computing –A service profile is a logical representation of a server, carrying its various identities and policies. Stateless computing enables procurement of a server within minutes compared to days in legacy server management systems.
- Policies, Pools, Templates –The management approach in Cisco UCS Manager is based on defining policies, pools and templates, instead of cluttered configuration, which enables a simple, loosely coupled, data driven approach in managing compute, network and storage resources.
- Built-in Multi-Tenancy Support –The combination of policies, pools and templates, organization hierarchy and a service profiles based approach to compute resources makes Cisco UCS Manager inherently friendly to multi-tenant environment typically observed in private and public clouds.

Cisco UCS 6200 Series Fabric Interconnects

The Cisco UCS Fabric interconnects (FI) provide a single point for connectivity and management for the entire system by integrating all compute components into a single, highly available management domain controlled by Cisco UCS Manager. Cisco UCS FIs support the system's unified fabric with low-latency, lossless, cut-through switching that supports IP, storage, and management traffic using a single set of cables. Cisco UCS FIs are typically deployed in redundant pairs. The Cisco UCS 6248UP model utilized in this CVD is a 1-RU form factor that features up to 48 universal ports that can support 10 Gigabit Ethernet or Fibre Channel over Ethernet, or 8/4/2 native Fibre Channel connectivity.

Figure 2 Cisco UCS 6248UP Fabric Interconnect

For more information on various models of the Cisco UCS 6200 Fabric Interconnect, visit http://www.cisco.com/c/en/us/products/servers-unified-computing/ucs-6200-series-fabric-interconnects/index.html

Latest generation of Cisco UCS 6300 Fabric Interconnect is not covered in this design guide.

Cisco UCS 5108 Blade Server Chassis

The Cisco UCS 5108 Blade Server Chassis is a fundamental building block of the Cisco Unified Computing System, delivering a scalable and flexible blade server architecture. A Cisco UCS 5108 Blade Server chassis is six rack units (6RU) high and can house up to eight half-width or four full-width Cisco UCS B-series blade servers.

For a complete list of blade servers supported, see: http://www.cisco.com/c/en/us/products/servers-unified-computing/ucs-b-series-blade-servers/index.html

The Cisco UCS 5108 chassis contains two I/O bays in the rear that can support Cisco UCS 2200 Series Fabric Extenders. The two fabric extenders can be used for both redundancy and bandwidth aggregation. A passive mid-plane provides up to 80 Gbps of I/O bandwidth per server slot and up to 160 Gbps of I/O bandwidth for two slots (full width) blades. The chassis is also capable of supporting 40 Gigabit Ethernet. Cisco UCS 5108 blade server chassis uses a unified fabric and fabric-extender technology to simplify and reduce cabling by eliminating the need for dedicated chassis management and blade switches. The unified fabric also reduces TCO by reducing the number of network interface cards (NICs), host bus adapters (HBAs), switches, and cables that need to be managed, cooled, and powered. This architecture enables a single Cisco UCS domain to scale up to 20 chassis with minimal complexity.

For more information, see:
http://www.cisco.com/c/en/us/products/servers-unified-computing/ucs-5100-series-blade-server-chassis/index.html

Figure 3 Cisco UCS 5108 Blade Server Chassis

Cisco UCS 2200 Series Fabric Extenders

The Cisco UCS 2200 Series Fabric Extender multiplexes and forwards all traffic from servers in a chassis to a parent Cisco UCS Fabric Interconnect over from 10-Gbps unified fabric links. All traffic, including traffic between servers on the same chassis, or between virtual machines on the same server, is forwarded to the parent fabric interconnect, where network profiles and polices are maintained and managed by the Cisco UCS Manager. Up to two fabric extenders can be deployed in a Cisco UCS chassis. The Cisco UCS 2200 Series Fabric Extenders come in two flavors:

- The Cisco UCS 2204XP Fabric Extender has four 10 Gigabit Ethernet, FCoE-capable, SFP+ ports that connect the blade chassis to the fabric interconnect. Each Cisco UCS 2204XP has sixteen 10 Gigabit Ethernet ports connected through the mid-plane to each half-width slot in the chassis. Typically configured in pairs for redundancy, two fabric extenders provide up to 80 Gbps of I/O to the chassis.
- The Cisco UCS 2208XP Fabric Extender has eight 10 Gigabit Ethernet, FCoE-capable, SFP+ ports that connect the blade chassis to the fabric interconnect. Each Cisco UCS 2208XP has thirty-two 10 Gigabit Ethernet ports connected through the mid-plane 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.

For more information, see: http://www.cisco.com/c/en/us/products/collateral/servers-unified-computing/ucs-6200-series-fabric-interconnects/data_sheet_c78-675243.html

Figure 4 Cisco UCS 2204XP Fabric Extender

Latest generation of Cisco UCS 2300 Fabric Extenders is not covered in this design guide.

Cisco UCS B200 M4 Servers

The enterprise-class Cisco UCS B200 M4 Blade Server extends the capabilities of Cisco's Unified Computing System portfolio in a half-width blade form factor. The Cisco UCS B200 M4 uses the power of the latest Intel® Xeon® E5-2600 v3 and v4 series processor family CPUs providing up to 44 processing cores, up to 1536 GB of RAM (using 64 GB DIMMs), two solid-state drives (SSDs) or hard disk drives (HDDs), and up to 80 Gbps throughput connectivity. The Cisco UCS B200 M4 Blade Server mounts in a Cisco UCS 5100 Series blade server chassis or Cisco UCS Mini blade server chassis. It has 24 total slots for

registered ECC DIMMs (RDIMMs) or load-reduced DIMMs (LR DIMMs) for up to 1536 GB total memory capacity. It supports one connector for Cisco's VIC 1340 or 1240 adapter, which provides Ethernet and FCoE connectivity.

For more information, see: http://www.cisco.com/c/en/us/products/servers-unified-computing/ucs-b200-m4-blade-server/index.html

Figure 5 Cisco UCS B200 M4 Blade Server

Cisco UCS C220 M4 Servers

The enterprise-class Cisco UCS C220 M4 server extends the capabilities of the Cisco Unified Computing System (UCS) portfolio in a one rack-unit (1RU) form-factor. The Cisco UCS C220 M4 uses the power of the latest Intel® Xeon® E5-2600 v3 and v4 Series processor family CPUs with up to 1536 GB of RAM (using 64 GB DIMMs), 8 Small Form-Factor (SFF) drives or 4 Large Form-Factor (LFF) drives, and up to 80 Gbps throughput connectivity. The Cisco UCS C220 M4 Rack Server can be used standalone, or as integrated part of the Unified Computing System. It has 24 DIMM for up to 1536 GB total memory capacity. It supports one connector for the Cisco VIC 1225, 1227 or 1380 adapters, which provide Ethernet and FCoE.

For more information, see: http://www.cisco.com/c/en/us/products/collateral/servers-unified-computing/ucs-c220-m4-rack-server/datasheet-c78-732386.html

Figure 6 Cisco UCS C220 M4 Rack Server

Cisco UCS Network Adapters

The Cisco Unified Computing System supports converged network adapters (CNAs) to provide connectivity to the blade and rack mount servers. CNAs obviate the need for multiple network interface cards (NICs) and host bus adapters (HBAs) by converging LAN and SAN traffic in a single interface. While Cisco UCS supports wide variety of interface cards, this CVD utilizes following two models: Cisco Virtual Interface Card (VIC) 1340 and Cisco VIC 1227. Further discussion around Cisco UCS adapters will be limited to these two models.

Cisco UCS Virtual Interface Card 1340

The Cisco UCS Virtual Interface Card (VIC) 1340 is a 2-port 40-Gbps Ethernet or dual 4 x 10-Gbps Ethernet, FCoE-capable modular LAN on motherboard (mLOM) designed exclusively for the M4 generation of Cisco UCS B-Series Blade Servers. The Cisco 1340 VIC supports an optional port-expander which enables the 40-Gbps Ethernet capabilities of the card. The Cisco UCS VIC 1340 supports a policy- based, stateless, agile server infrastructure that can present over 256 PCIe standards-compliant interfaces to the host that can be dynamically configured as either network interface cards (NICs) or host bus adapters (HBAs). In addition, the Cisco UCS VIC 1340 supports Cisco Virtual Machine Fabric Extender (VM-FEX) technology, which extends the Cisco UCS Fabric interconnect ports to virtual machines, simplifying server virtualization deployment and management.

For more information, see: http://www.cisco.com/c/en/us/products/interfaces-modules/ucs-virtual-interface-card-1340/index.html

Figure 7 Cisco VIC 1340

Cisco UCS Virtual Interface Card 1227

The Cisco UCS VIC 1227 is a dual-port Enhanced Small Form-Factor Pluggable (SFP+) 10-Gbps Ethernet and FCoE-capable PCI Express (PCIe) mLOM adapter designed exclusively for Cisco UCS C-Series Rack Servers. The mLOM slot can be used to install a Cisco VIC without consuming a PCIe slot, which provides greater I/O expandability. Just like Cisco UCS VIC 1340, the Cisco UCS VIC 1227 enables a policy-based, stateless, agile server infrastructure that can present up to 256 PCIe standards-compliant interfaces to the host that can be dynamically configured as either NICs or HBAs. The Cisco UCS VIC 1227 also supports VM-FEX technology, which extends the Cisco UCS fabric interconnect ports to virtual machines, simplifying server virtualization deployment.

For more information, see: http://www.cisco.com/c/en/us/products/interfaces-modules/ucs-virtual-interface-card-1227/index.html

Figure 8 Cisco VIC 1227

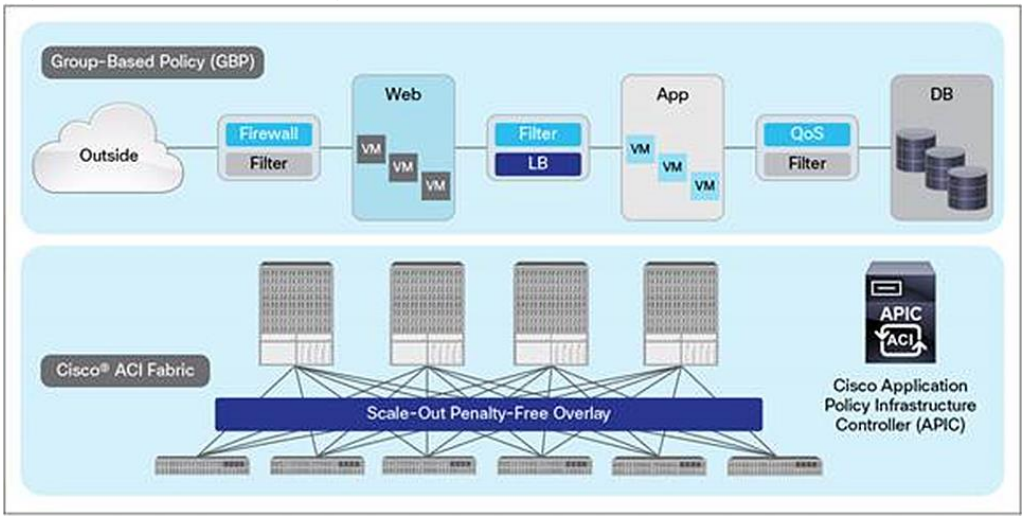
Cisco Nexus 9000 based Application Centric Infrastructure

The Cisco Nexus 9000 family of switches supports two modes of operation: NX-OS standalone mode and Application Centric Infrastructure (ACI) fabric mode. In standalone mode, the switch performs as a typical Cisco Nexus switch with increased port density, low latency and 40G connectivity. In fabric mode, the administrator can take advantage of Cisco Application Centric Infrastructure (ACI).

Cisco ACI is a new data center architecture designed to address the requirements of today's traditional networks, as well as to meet emerging demands that new computing trends and business factors are placing on the network. Software-defined networking (SDN) has garnered much attention in the networking industry over the past few years due to its promise of a more agile and programmable network infrastructure. Cisco ACI not only addresses the challenges of agility and network programmability that software-based overlay networks are trying to address, but it also presents solutions to the new challenges that SDN technologies are currently unable to address.

Cisco ACI leverages a network fabric that employs industry proven protocols coupled with innovative technologies to create a flexible, scalable, and highly available architecture of low-latency, high-bandwidth links. This fabric delivers application instantiations using profiles that house the requisite characteristics to enable end-to-end connectivity. The ACI fabric is designed to support the management automation, programmatic policies, and dynamic workload provisioning. The ACI fabric accomplishes this with a combination of hardware, policy-based control systems, and closely coupled software to provide advantages not possible in other vendor solutions.

Figure 9 Cisco ACI - High Level Architecture

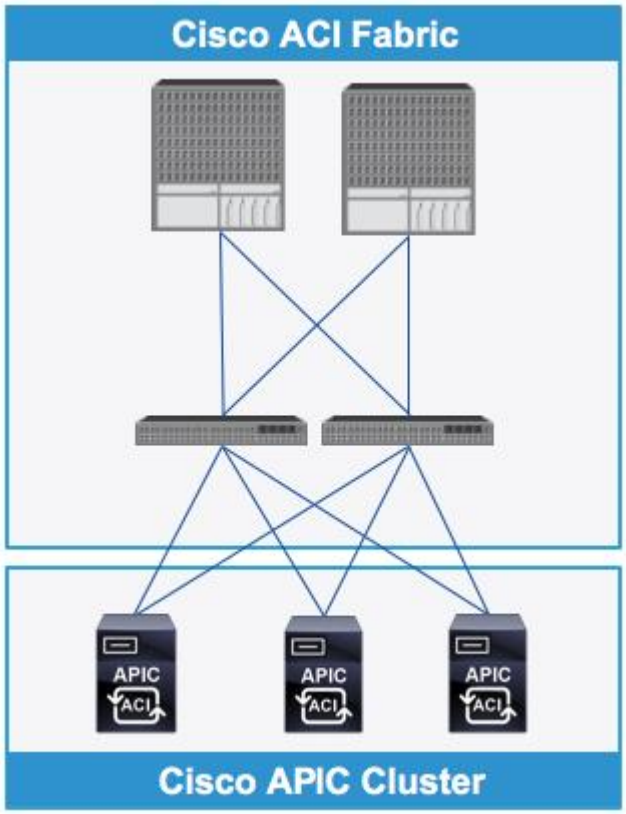


The ACI switching architecture is presented in a leaf-and-spine topology where every leaf connects to every spine using 40G Ethernet interfaces. At a high-level, the Cisco ACI fabric consists of three major components:

- The Application Policy Infrastructure Controller (APIC)
- Spine switches
- Leaf switches

Cisco Nexus 9000-based VersaStack design with Cisco ACI consists of Cisco Nexus 9336 PQ based spine and Cisco 9372 PX based leaf switching architecture controlled using a cluster of three Application Policy Infrastructure Controllers (APICs).

Figure 10 VersaStack ACI Design



Application Policy Infrastructure Controller (APIC)

The software controller, APIC, is delivered as an appliance and three or more such appliances form a cluster for high availability and enhanced performance. APIC is responsible for all tasks enabling traffic transport including fabric activation, switch firmware management and network policy configuration and instantiation. Though the APIC acts as the centralized point of configuration for policy and network connectivity, it is never in-line with the data path or the forwarding topology and the fabric can still forward traffic even when communication with the APIC is disrupted. APIC provides both a command-line interface (CLI) and graphical-user interface (GUI) to configure and control the ACI fabric. APIC also exposes a northbound API through XML and JavaScript Object Notation (JSON) and an open source southbound API.

ACI 9300 based Spine and Leaf Switches

The Cisco Nexus 9300 Series Switches include both spine and leaf switches. Cisco Nexus 9300 platform leaf switches are Layer 2 and 3 non-blocking 10 and 40 Gigabit Ethernet switches with up to 2.56 terabits per second (Tbps) of internal bandwidth.

Cisco Nexus 9336 PQ Spine

The Cisco Nexus 9336PQ ACI Spine Switch is a 2-rack-unit (2RU) spine switch for Cisco ACI that supports 2.88 Tbps of bandwidth across 36 fixed 40 QSFP+ ports as shown in Figure 11.

Figure 11 Cisco Nexus 9336 PQ Switch

Cisco Nexus 9372 PX Leaf

The Cisco Nexus 9372PX and 9372PX-E Switches are 1RU switches that support 1.44 Tbps of bandwidth and over 1150 mpps across 48 fixed 10-Gbps SFP+ ports and 6 fixed 40-Gbps QSFP+ ports as shown in Figure 12. The Cisco Nexus 9372PX-E is a minor hardware revision of the Cisco Nexus 9372PX.

Figure 12 Cisco Nexus 9372 PX Switch

For detailed information on the Cisco Nexus 9000 product line, refer to http://www.cisco.com/c/en/us/products/collateral/switches/nexus-9000-series-switches/datasheet-c78-729405.html

Cisco MDS 9100 Series Fabric Switches

The Cisco MDS 9148S 16G Multilayer Fabric Switch 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 12 to 48 line-rate 16 Gbps Fibre Channel ports. Cisco MDS 9148S is powered by Cisco NX-OS and 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. Cisco MDS 9148S provides up to 256 buffer credits per group of 4 ports and supports some of the advanced functions such as Virtual SAN (VSAN), Inter-VSAN routing (IVR), port-channels and multipath load balancing and flow and zone based QoS.

Figure 13 Cisco MDS 9148S

For more information, refer to: http://www.cisco.com/c/en/us/products/storage-networking/mds-9000-series-multilayer-switches/index.html

IBM Spectrum Virtualize

In the era of cloud, big data and analytics, mobile and social computing, organizations need to meet ever-changing demands for storage, while also improving data economics. IT must deliver more services faster and more efficiently, enable real-time insight and support more customer interaction. The right infrastructure allows clients to share information, secure transactions and drive real-time insights.

Built with IBM Spectrum Virtualize software–part of the IBM Spectrum Storage family–the IBM FlashSystem V9000 & IBM Storwize V7000 Unified help organizations achieve better data economics by supporting these new work-loads that are critical to their success. The FlashSystem V9000 and Storwize V7000 Unified systems can handle the massive volumes of data from mobile and social applications, enable rapid and flexible cloud services deployments, and deliver the performance and scalability needed to gain insights from the latest analytics technologies.

Version 7.6 of IBM Spectrum Virtualize enriches the exiting offering of features with the following functionality across both the IBM FlashSystem V9000 and IBM Storwize V7000 Unified storage systems.

Improve data security and reduce capital and operating expense

- Single point of control for encrypting data on heterogeneous storage simplifies management
- Eliminates need to buy new storage systems to enable encryption

Improve data protection and performance with lower cost storage

- Distributed RAID improves drive rebuild time 5-10x: enables use of large drives with more confidence
- All drives are active, which improves performance especially with flash drives

Reduce cost and complexity for high availability configurations

- New GUI makes HyperSwap easier to configure and use
- IP quorum support eliminates need for extra storage and fibre channel networking to third site

Reduce cost of storage: store up to 5x as much data in same space

- Quicker, easier view of potential compression benefits with integrated Comprestimator

Existing features of IBM Spectrum Virtualize product offering

- FlashCopy for near-instant data backups
- IBM Real-time Compression Accelerators
- IBM EasyTier for automated storage tiering
- Thin provisioning
- Synchronous data replication with Metro Mirror
- Asynchronous data replication with Global Mirror
- Data virtualization
- HyperSwap Split-Clusters
- Highly available configurations
- External storage virtualization and data migration

IBM FlashSystem V9000

IBM FlashSystem V9000 is a virtualized, flash-optimized, enterprise-class storage system that provides the foundation for implementing an effective storage infrastructure with simplicity and by transforming the economics of data storage. Designed to complement virtual server environments, these modular storage systems deliver the flexibility and responsiveness required for changing business needs.

IBM FlashSystem V9000 uses a fully featured and scalable all-flash architecture that performs at up to 2.5 million input/output operations per second (IOPS) with IBM MicroLatency. The system is scalable to 19.2 gigabytes per second (GBps), and delivers an effective capacity of up to 2.28 petabytes (PB). Using its flash-optimized design, FlashSystem V9000 can provide response times of around 200 microseconds. The system delivers better acquisition costs than a high-performance spinning disk for the same effective

capacity while achieving five times the performance, making it ideal for environments that demand extreme performance.

Figure 14 IBM V9000 Storage Array

Each IBM FlashSystem V9000 node canister has up to 128GB internal cache to accelerate and optimize writes, and hardware acceleration to boost the performance of Real-time Compression.

IBM FlashSystem V9000 with Real-time Compression

IBM's Real-time Compression is a key differentiator in the industry. Unlike other approaches to compression, Real-time Compression is designed to work on active primary data, by harnessing dedicated hardware acceleration thereby s achieving extraordinary efficiency on a wide range of candidate data. Certain workloads, such as production databases or email systems etc., can store up to five times the data in the same physical disk space. FlashSystem V9000 also provides a built in, non-intrusive compression analysis tool, that measures the yield due to compression of an existing volume and therefore providing accurate workload capacity planning

IBM FlashSystem V9000 Easy-to-Use Management GUI

The IBM FlashSystem V9000 built-in user interface (Figure 15) hides configuration complexity and makes it possible for administrators to quickly and easily complete common storage tasks such as creating and deploying volumes, mapping hosts to volumes as well as managing system performance from the same GUI. The IBM FlashSystem V9000 management interface provides customers with an upgrade wizard to keep easily upgrade to the latest software release with just a few mouse clicks. The interface provides auto-discovery and presets that help the admin greatly reduce setup time and help them easily implement a successful deployment. The interface is web-accessible and built into the product, removing the need for the administrator to download and update management software.

Figure 15 IBM FlashSystem V9000 Management GUI Example

For more information on the IBM FlashSystem V9000 refer to the IBM Redbook, IBM FlashSystem V9000 Product Guide: http://www.redbooks.ibm.com/abstracts/redp5317.html?Open

IBM Storwize V7000 Unified

IBM Storwize V7000 Unified is a virtualized, enterprise-class hybrid storage system that provides the foundation for implementing an effective storage infrastructure and transforming the economics of data storage. Designed to complement virtual server environments, these modular storage systems deliver the flexibility and responsiveness required for changing business needs.

The hardware platform of Storwize V7000 Unified is designed to deliver both high performance and dramatically improved data economics. A control enclosure contains dual redundant controllers, each with an 8-core 1.9 GHz Intel Xeon processor with 32 GB or 64 GB of cache. Each controller contains a hardware compression accelerator based on Intel QuickAssist technology with an available second accelerator. Flexible host interface options include 16 Gbps and 8 Gbps Fibre Channel, 1 Gbps iSCSI, and 10 Gbps iSCSI or Fibre Channel over Ethernet. This powerful platform delivers up to twice as much throughput as previous systems. Each control enclosure supports up to 20 expansion enclosures attached using high-performance 12 Gbps SAS for maximum expansion of 504 drives or approximately 2 PB of capacity. Control enclosures support up to 24 2.5-inch drives and two models of expansion enclosure support up to 24 2.5-inch or 12 3.5-inch drives.

Clustered systems provide scale-out growth in performance and capacity with up to four control enclosures and associated expansion enclosures operating as a single storage system with 64 processor cores, up to 512 GB of cache, supporting up to 1,056 drives and 7.87 PB of total capacity. Storwize V7000 Unified

systems also include dual redundant File Modules with 1 Gbps and 10 Gbps interfaces for network-attached storage (NAS) capability.

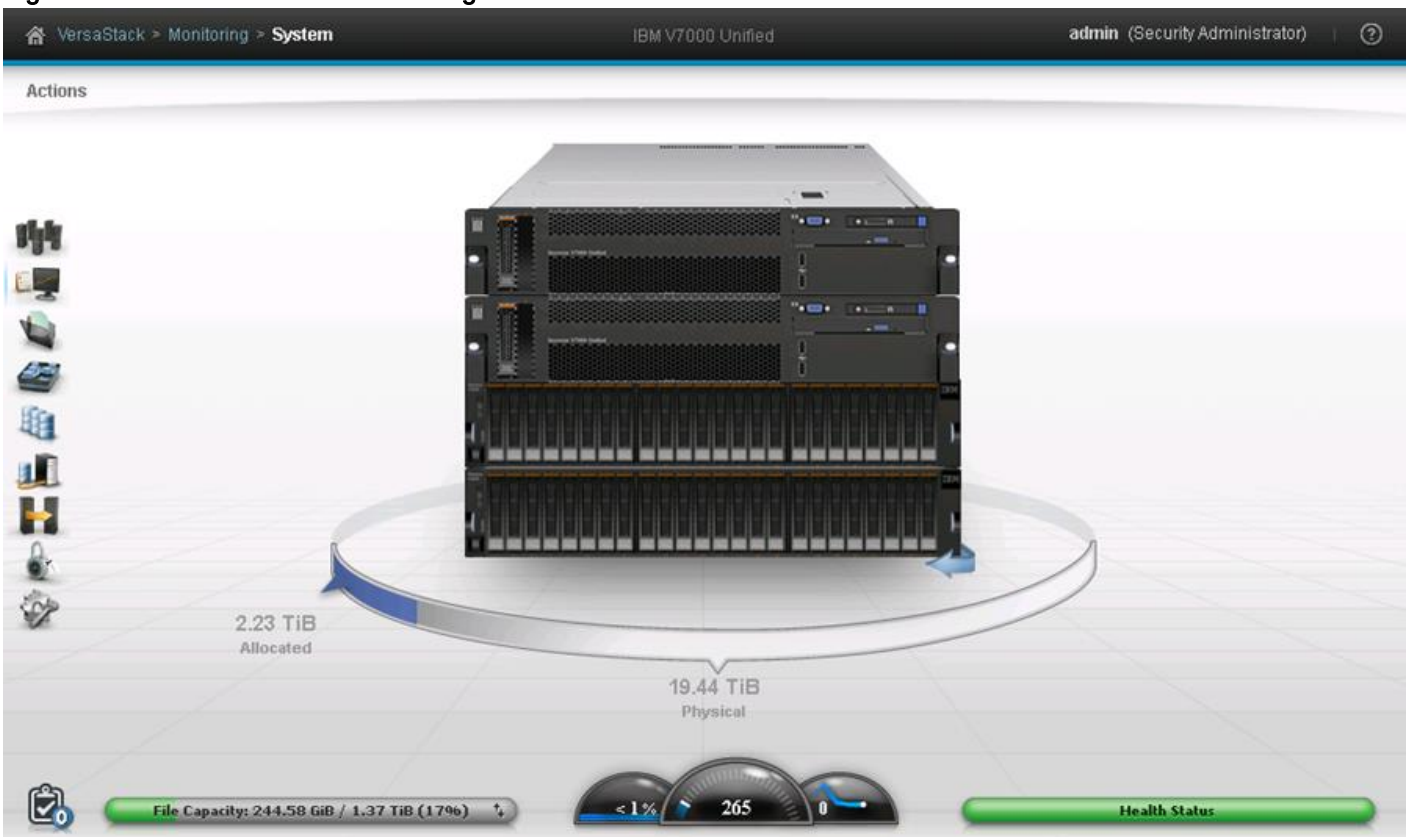
IBM Spectrum Virtualize capabilities unique to V7000 Unified storage system

- Spectrum Scale – Global Namespace with unique Active File Management
- IBM Active Cloud Management technology delivers automated storage efficiency capabilities
- Built-in IBM Spectrum Protect client simplifies backup
- Local Active File Management (ILM)
- Global Active File Management (WAN Caching)
- Support file access protocols NFS, CIFS, FTP, HTTPS, SCP
- Antivirus Support
- Data Protection with File Cloning, Replication & Snapshot

V7000 Unified Management GUI showing single interface for block and file management

An intuitive management interface enables administrators to easily manage both block and file data in the same system.

Figure 16 IBM V7000 Unified Management GUI



For further reading on the IBM Storwize V7000 Unified refer to the IBM Redbook, Implementing the IBM Storwize V7000 Unified Disk System

VMware vCenter Server

VMware vCenter is the simplest and most efficient way to manage VMware vSphere hosts at scale. It provides unified management of all hosts and virtual machines from a single console and aggregates performance monitoring of clusters, hosts, and virtual machines. VMware vCenter Server gives administrators a deep insight into the status and configuration of compute clusters, hosts, virtual machines, storage, the guest OS, and other critical components of a virtual infrastructure. A single administrator can manage 100 or more virtualization environment workloads using VMware vCenter Server, more than doubling the typical productivity in managing physical infrastructure. VMware vCenter manages the rich set of features available in a VMware vSphere environment.

For more information, refer to: http://www.vmware.com/products/vcenter-server/overview.html

Cisco Application Virtual Switch

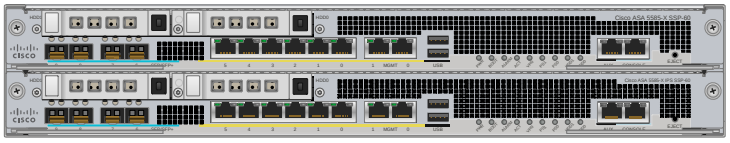
Cisco Application Virtual Switch (AVS) is a hypervisor-resident virtual network switch that is specifically designed for the ACI architecture. Based on the Cisco Nexus 1000V virtual switch, Cisco AVS provides feature support for the ACI application policy model, full switching capabilities, and more advanced telemetry features. Instead of using hypervisor-specific management stations, Cisco AVS provides cross-consistency in features, management, and control directly through Cisco APIC. Some of the key features of Cisco AVS include:

- A purpose-built, virtual network edge for ACI fabric architecture
- Integration with the ACI management and orchestration platform to automate virtual network provisioning and application services deployments
- Integrated visibility of both physical and virtual workloads and network paths
- Open APIs to extend the software-based control and orchestration of the virtual network fabric
- Optimal traffic steering to application services and seamless workload mobility
- Support for a consistent operational model across multiple hypervisors for simplified operations in heterogeneous data centers

Cisco Adaptive Security Appliance (ASA)

The Cisco ASA Family of security devices protects corporate networks and data centers of all sizes. Cisco ASA delivers enterprise-class firewall capabilities for ASA devices in an array of form factors. ASA Software also integrates with other critical security technologies to deliver comprehensive solutions that meet continuously evolving security needs. Cisco ASA delivers high availability for high resiliency applications thereby meeting the unique requirements in the data center. Cisco ASA supports multiple contexts for a multi-tenant deployment. This design guide uses Cisco ASA 5585 platform (shown in Figure 17) to provide firewall functionality.

Figure 17 Cisco ASA 5585



Solution Design

VersaStack with Cisco ACI architecture aligns with the converged infrastructure configurations and best practices as identified in the previous VersaStack releases. 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 system supports high availability at network, compute and storage layers such that no single point of failure exists in the design. The system utilizes 10 and 40 Gbps Ethernet jumbo-frame based connectivity combined with port aggregation technologies such as virtual port-channels (VPC) for non-blocking LAN traffic forwarding. A dual SAN 8/16 Gbps environment provides redundant storage access from compute devices to the storage controllers.

The VersaStack Datacenter with Cisco ACI solution, as covered in this design guide, conforms to a number of key design and functionality requirements. Some of the key features of the solution are highlighted below.

- The system is able to tolerate the failure of compute, network or storage components without significant loss of functionality or connectivity
- The system is built with a modular approach thereby allowing customers to easily add more network (LAN or SAN) bandwidth, compute power or storage capacity as needed
- The system supports stateless compute design thereby reducing time and effort required to replace or add new compute nodes
- The system provides network automation and orchestration capabilities to the network administrators using Cisco APIC GUI, CLI and restful API
- The systems allow the compute administrators to instantiate and control application Virtual Machines (VMs) from VMware vCenter
- The system permits storage administrators to easily manage the storage using IBM management GUI
- The solution supports live VM migration between various compute nodes and protects the VM by utilizing VMware HA and DRS functionality
- The system can be easily integrated with optional Cisco (and third party) orchestration and management application such as Cisco UCS Central and Cisco UCS Director
- The system showcases layer-3 connectivity to the existing enterprise network and provides firewall based system security by utilizing Cisco Adaptive Security Appliance (ASA)

Physical Topology

This VersaStack with Cisco ACI solution utilizes Cisco UCS platform with Cisco B200 M4 half-width blades and Cisco UCS C220 M4 rack mount servers connected and managed through Cisco UCS 6248 Fabric

Interconnects and the integrated Cisco UCS manager. These high performance servers are configured as stateless compute nodes where ESXi 6.0 U1b hypervisor is loaded using SAN (iSCSI and FC) boot. The boot disks to store ESXi hypervisor image and configuration along with the block and file based datastores to host application Virtual Machines (VMs) are provisioned on the IBM storage devices.

As in the non-ACI designs of VersaStack, link aggregation technologies play an important role in VersaStack with ACI solution providing improved aggregate bandwidth and link resiliency across the solution stack. The IBM Storwize V7000 Unified File Modules, Cisco UCS, and Cisco Nexus 9000 platforms support active port channeling using 802.3ad standard Link Aggregation Control Protocol (LACP). In addition, the Cisco Nexus 9000 series features virtual Port Channel (vPC) capability which allows links that are physically connected to two different Cisco Nexus devices to appear as a single "logical" port channel. Each Cisco UCS FI is connected to both the Cisco Nexus 9372 leaf switches using virtual port-channel (vPC) enabled 10GbE uplinks for a total aggregate bandwidth of 40GBps. Additional ports can be easily added to the design for increased bandwidth. Each Cisco UCS 5108 chassis is connected to the FIs using a pair of 10GbE ports from each IO Module for a combined 40GbE uplink. Each of the Cisco UCS C-220 servers connects directly into each of the FIs using a 10Gbps converged link for an aggregate bandwidth of 20Gbps per server.

To provide the storage system connectivity, this design guides covers two different storage connectivity options:

- Option 1: iSCSI based storage design validated for IBM FlashSystem V9000
- Option 2: FC and NFS based storage design validated for IBM Storwize V7000 Unified

While both IBM FlashSystem V9000 and IBM Storwize V7000 Unified support iSCSI and FC connectivity, the storage connectivity models presented in this design guide are only intended to provide a reference connectivity model. All the possible storage connectivity designs on different controllers are not covered.

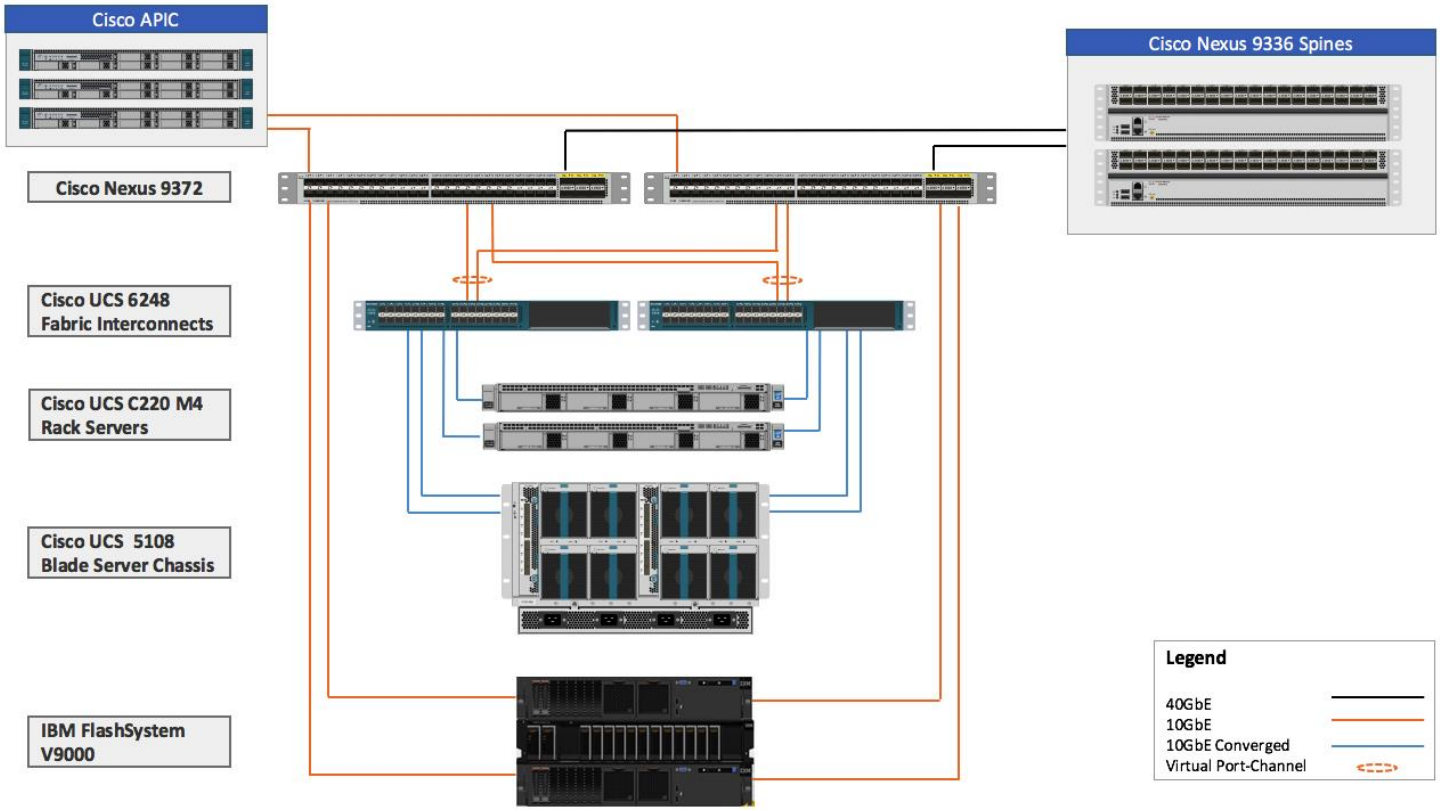
Figure 18 and Figure 19 below provide a high level topology of the two storage connectivity options. The VersaStack infrastructure in both these option satisfy the high-availability design requirements and is physically redundant across the network, compute and storage stacks. The integrated compute stack design presented in this document can withstand failure of one or more links as well as the failure of one or more devices.

iSCSI based Storage Design with IBM FlashSystem V9000

IBM FlashSystem V9000 based VersaStack design option is shown in Figure 18 below. This design utilizes an all-IP based storage access model where IBM FlashSystem V9000 is connected directly to the Cisco Nexus 9372 leaf switches without requiring Fiber based switching infrastructure*. A10GbE port from each IBM FlashSystem V9000 controller is connected to each of the two Cisco Nexus 9372 leaf switches providing an aggregate bandwidth of 40Gbps.

* Depending on the storage layout, Fibre Channel switching infrastructure may still be required for IBM V9000 controller and storage enclosure connectivity.

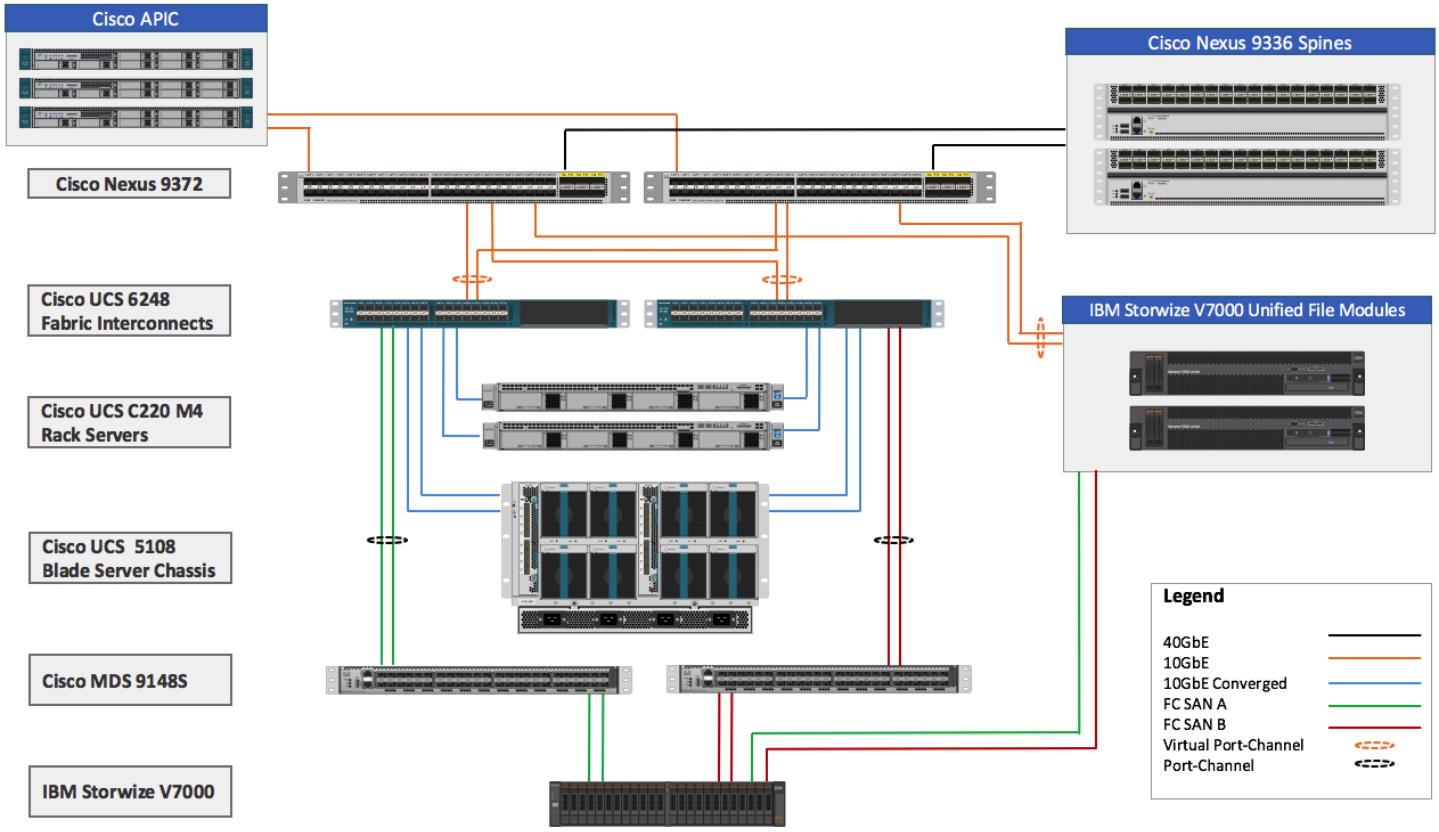
Figure 18 iSCSI based Storage Design with IBM FlashSystem V9000



FC and NFS based Storage Design with IBM Storwize V7000 Unified

IBM Storwize V7000 unified based VersaStack design option is shown in Figure 19 below. This design utilizes an FC based storage access model where IBM Storwize V7000 controller is connected to the Cisco UCS Fabric Interconnects through a dedicated Cisco MDS 9148S based redundant FC fabric. This design also covers NFS based storage connectivity by utilizing IBM Storwize V7000 Unified File Modules. The IBM Storwize V7000 Unified File Modules are connected to the Cisco Nexus 9372 leaf switches using 10GbE connectivity as shown below.

Figure 19 FC and NFS based Storage Design with IBM Storwize V7000 Unified



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

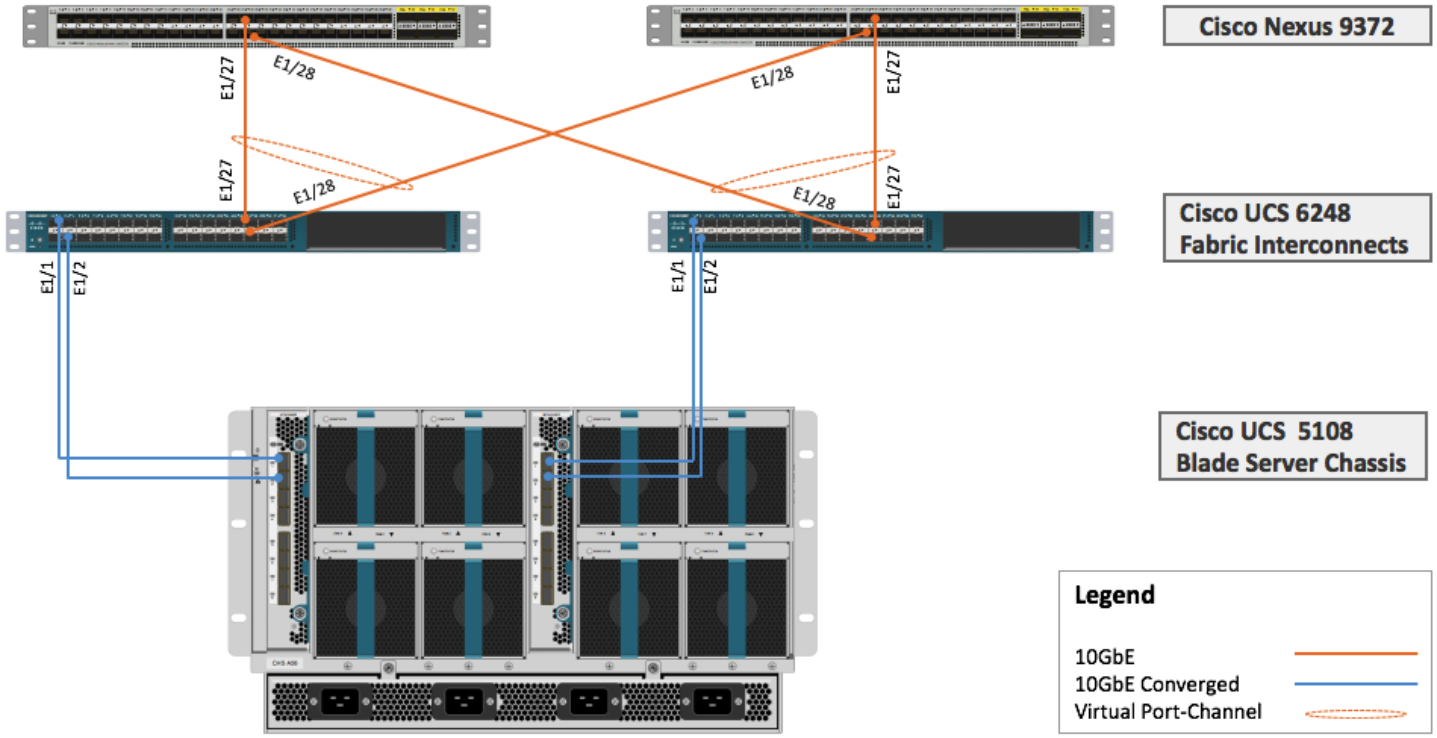
Cisco Unified Computing System

The VersaStack compute design supports both Cisco UCS B-Series and C-Series deployments. The Cisco UCS supports the virtual server environment by providing robust, highly available, and extremely manageable compute resources. In this validation effort, multiple Cisco UCS B-Series and C-Series ESXi servers are booted from SAN using iSCSI or FC (depending on the storage design option).

Cisco UCS LAN Connectivity

Cisco UCS Fabric Interconnects are configured with two port-channels, one from each FI, to the Cisco Nexus 9372 leaf switches. These port-channels carry all the data and IP-based storage traffic originated on the Cisco Unified Computing System. Virtual Port-Channels (vPC) are configured on the Cisco Nexus 9372 to provide device level redundancy. The validated design utilized two uplinks from each FI to the leaf switches for an aggregate bandwidth of 40GbE (4 x 10GbE). The number of links can be increased based on customer data throughput requirements.

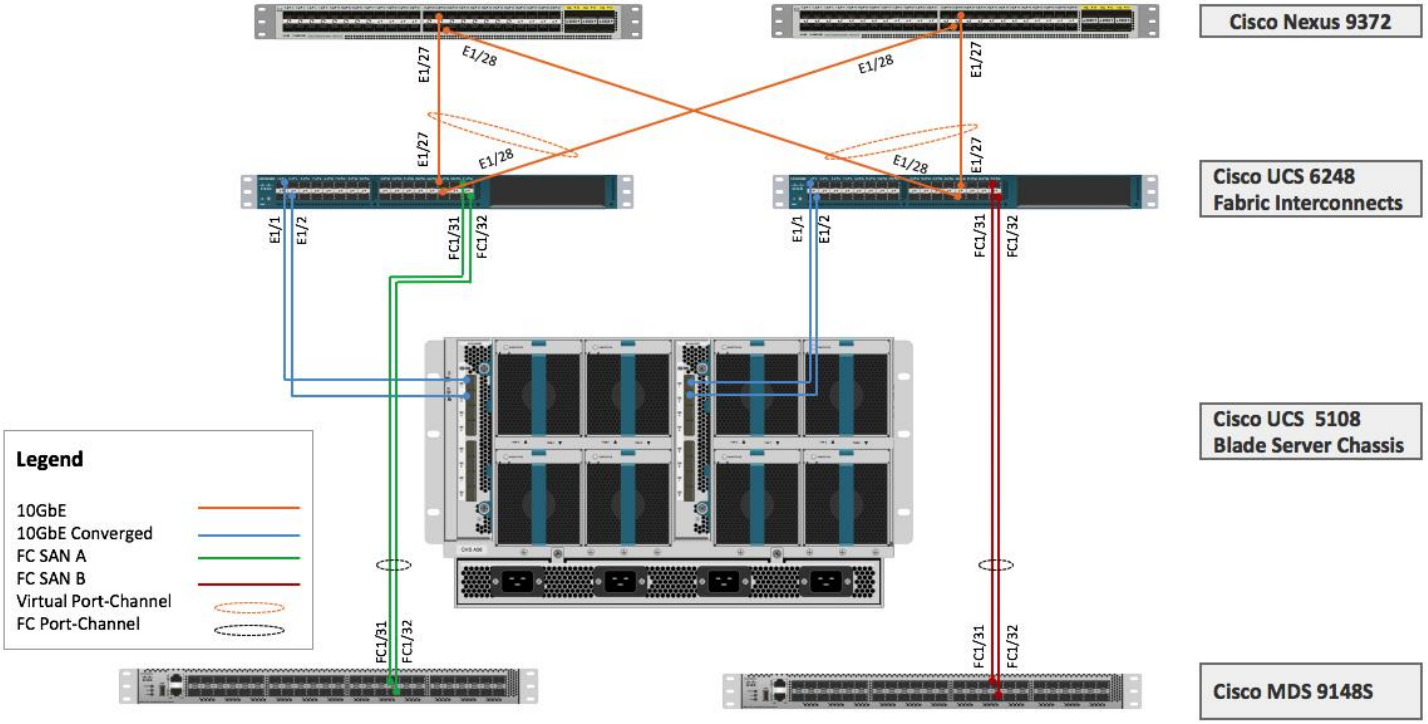
Figure 20 Cisco UCS - Physical Connectivity for iSCSI based Storage Designs



Cisco UCS SAN Connectivity

Cisco UCS Fabric Interconnects (FI) are connected to a dedicated redundant Cisco MDS 9148S based SAN fabric to provide Fibre Channel storage connectivity. In addition to the LAN connections covered in Figure 20, two 8Gbps FC ports from each of the FIs are connected to a single Cisco MDS fabric switch to support a SAN-A/B design. These ports are configured as FC port-channel to provide 16Gbps effective BW from each FI to each fabric switch. This design is shown in Figure 21:

Figure 21 Cisco UCS - Physical Connectivity for FC based Storage Designs

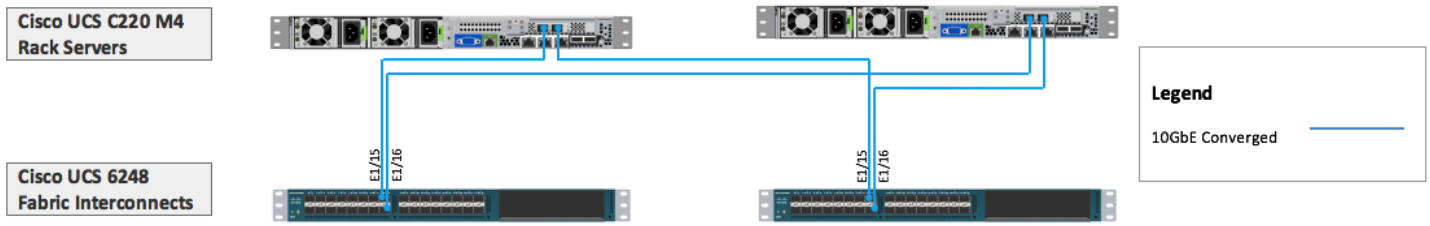


Cisco MDS switches are deployed with N-Port ID Virtualization (NPIV) enabled to support the virtualized environment running on Cisco UCS blade and rack servers. To support NPIV on the Cisco UCS servers, the Cisco UCS Fabric Interconnects that connect the servers to the SAN fabric, are enabled for N-Port Virtualization (NPV) mode by configuring to operate in end-host mode. NPV enables Cisco FIs to proxy fabric login, registration and other messages from the servers to the SAN Fabric without being a part of the SAN fabric. This is important for keeping the limited number of Domain IDs that Fabric switches require to a minimum. FC port-channels are utilized for higher aggregate bandwidth and redundancy. Cisco MDS also provide zoning configuration to enable single initiator (vHBA) to talk to multiple targets.

Cisco UCS C-series Server Connectivity

In all VersaStack designs, Cisco UCS C-series rack mount servers are always connected via the Cisco UCS FIs and managed through Cisco UCS Manager to provide a common management look and feel. Cisco UCS Manager 2.2 and later versions allow customers to connect Cisco UCS C-Series servers directly to Cisco UCS Fabric Interconnects without requiring a Fabric Extender (FEX). While the Cisco UCS C-Series connectivity using Cisco Nexus 2232 FEX is still supported and recommended for large scale Cisco UCS C-Series server deployments, direct attached design allows customers to connect and manage Cisco UCS C-Series servers on a smaller scale without buying additional hardware. In the VersaStack with ACI design, two Cisco UCS C220-M4 servers were directly attached to Cisco UCS FI using two 10Gbps converged connections (one connection to each FI) as shown in Figure 22.

Figure 22 Cisco UCS C220 Connectivity



Cisco UCS Server configuration for VSphere

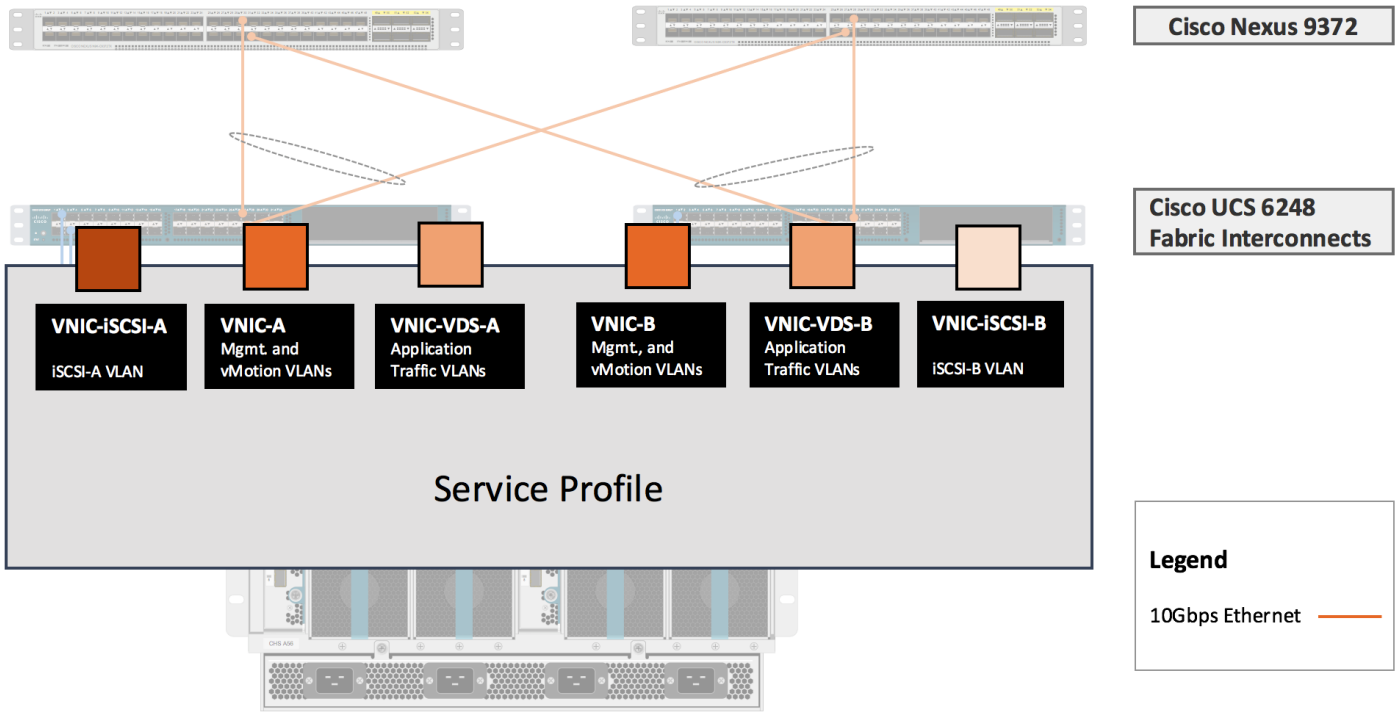
The ESXi nodes consist of Cisco UCS B200-M4 series blades with Cisco 1340 VIC or Cisco UCS C220-M4 rack mount servers with Cisco 1227 VIC. These nodes are allocated to a VMware High Availability (HA) cluster to support infrastructure services and applications. At the server level, the Cisco 1227/1340 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 for iSCSI storage, six vNICs are created and utilized as follows (Figure 23):

- One vNIC for iSCSI-A traffic
- One vNIC fo iSCSI-B traffic
- Two vNICs for infrastructure traffic including management and vMotion traffic
- Two vNICs for application related data including storage access if required. These vNICs are assigned to APIC controlled distributed switch (vDS or AVS)

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

Figure 23 Cisco UCS - Server Interface Design for iSCSI based Storage

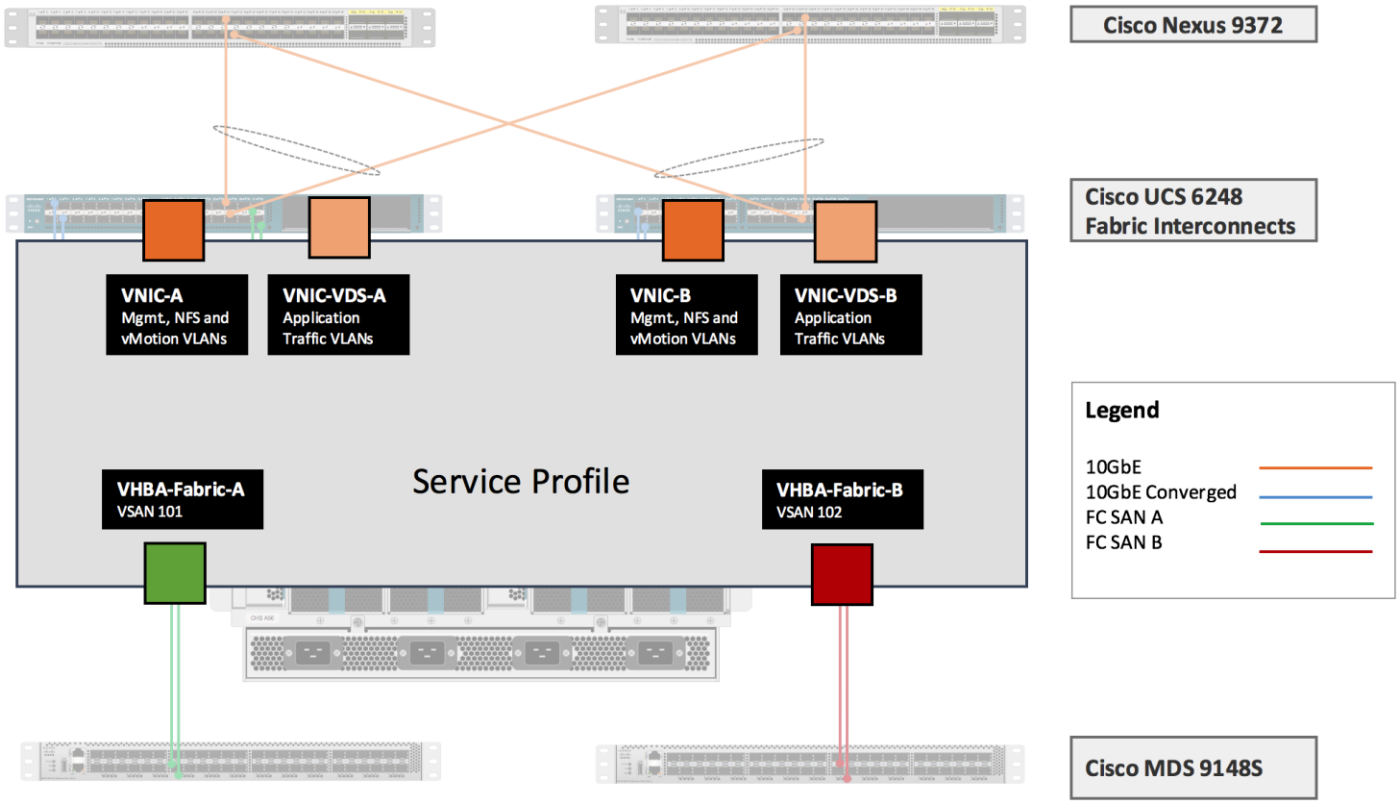


In the VersaStack design for FC and NFS storage, four vNICs and two vHBAs are created and utilized as follows (Figure 24):

- Two vNICs for infrastructure traffic including management, NFS and vMotion traffic
- Two vNICs for application related data including application storage access if required. These vNICs are assigned to APIC controlled distributed switch

- One vHBA for VSAN-A FC traffic
- One vHBA for VSAN-B FC traffic

Figure 24 Cisco UCS - Server Interface Design for FC and NFS based Storage



IBM Storage Systems

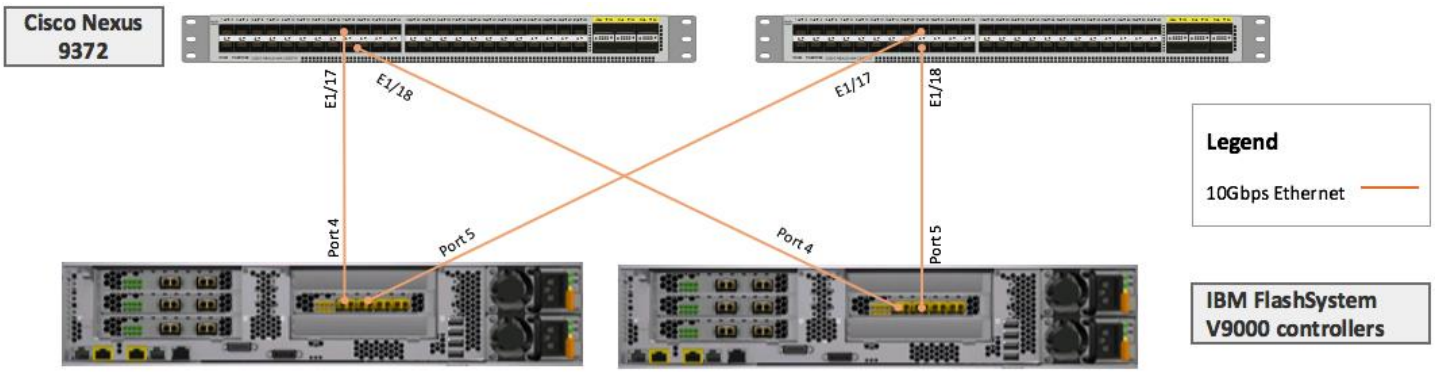
IBM FlashSystem V9000 and IBM Storwize V7000 Unified, covered in this VersaStack with ACI design, are deployed as high availability storage solutions. IBM storage systems support fully redundant connections for communication between control enclosures, external storage, and host systems.

Each storage system provides redundant controllers and redundant iSCSI, FC and NFS 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 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 V9000 – iSCSI Connectivity

To support iSCSI based IP-only storage connectivity, each IBM FlashSystem V9000 controller is connected to each of the Cisco Nexus 9372 leaf switch. The physical connectivity is shown in Figure 25. In this design, each 10Gbps Ethernet port between the storage controller and the ACI fabric is configured for either iSCSI-A or iSCSI-B path. Additional ports can be easily added for additional bandwidth.

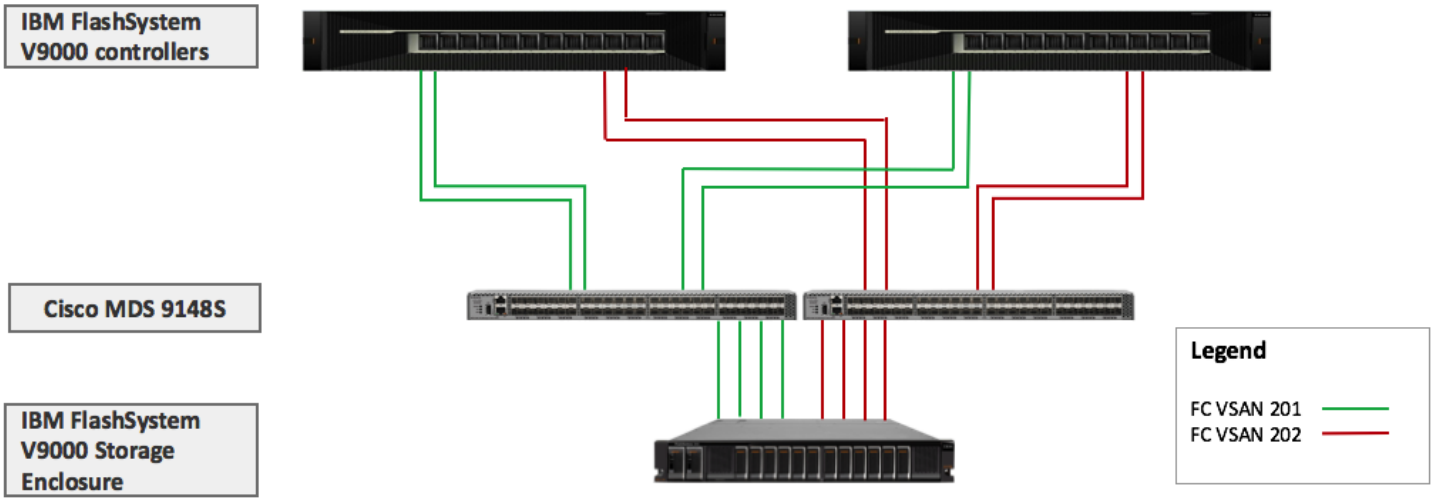
Figure 25 IBM FlashSystem V9000 - iSCSI based Storage Access



IBM FlashSystem V9000 – Connectivity to the Storage Enclosure

Figure 26 illustrates connectivity between FlashSystem V9000 Controllers and the Storage Enclosure using Cisco MDS 9000 switches. All connectivity between the Controllers and the Storage Enclosure is 16 Gbps End-to-End. Since the ESXi to storage system connectivity uses iSCSI protocol, in this particular design guide Cisco MDS switches were only configured with storage system VSANs.

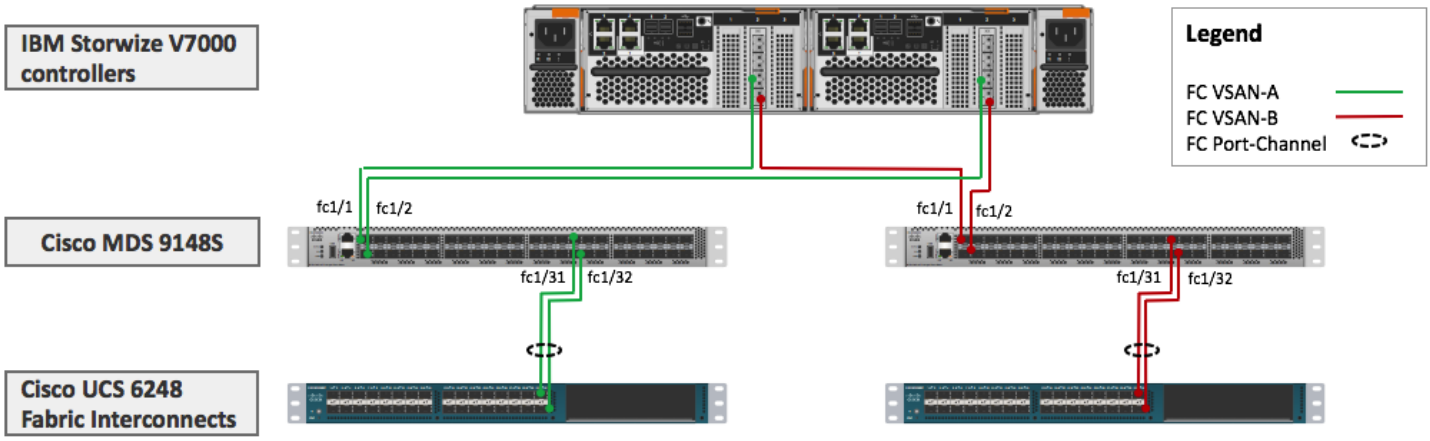
Figure 26 IBM FlashSystem V9000 – Connectivity to the Storage Controller



IBM Storwize V7000 Unified - FC Connectivity

To support FC based storage connectivity, each IBM Storwize V7000 controller is connected to a redundant Cisco MDS 9148S fabric switch. The physical connectivity is shown in Figure 27. In this design, each of the V7000 controllers utilizes redundant SAN-A and SAN-B paths and can tolerate link or device failures.

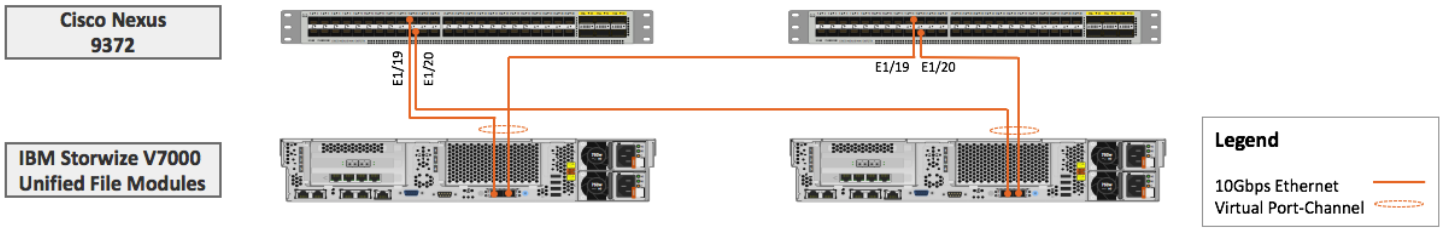
Figure 27 IBM V7000 Unified - FC Connectivity



IBM Storwize V7000 Unified - NFS Connectivity

To support NFS based storage access, redundant IBM V7000 Unified File Modules are deployed in this design. Figure 28 shows the connectivity between the File Modules and the Cisco Nexus 9372 leaf switches. Each of the File Module is configured to connect to the Cisco Nexus 9372 using a port-channel. The Cisco Nexus 9372 leaf is configured for vPC. The design can withstand failure of one or more links as well as failure of a device.

Figure 28 IBM V7000 Unified - File Module Connectivity



Network Design

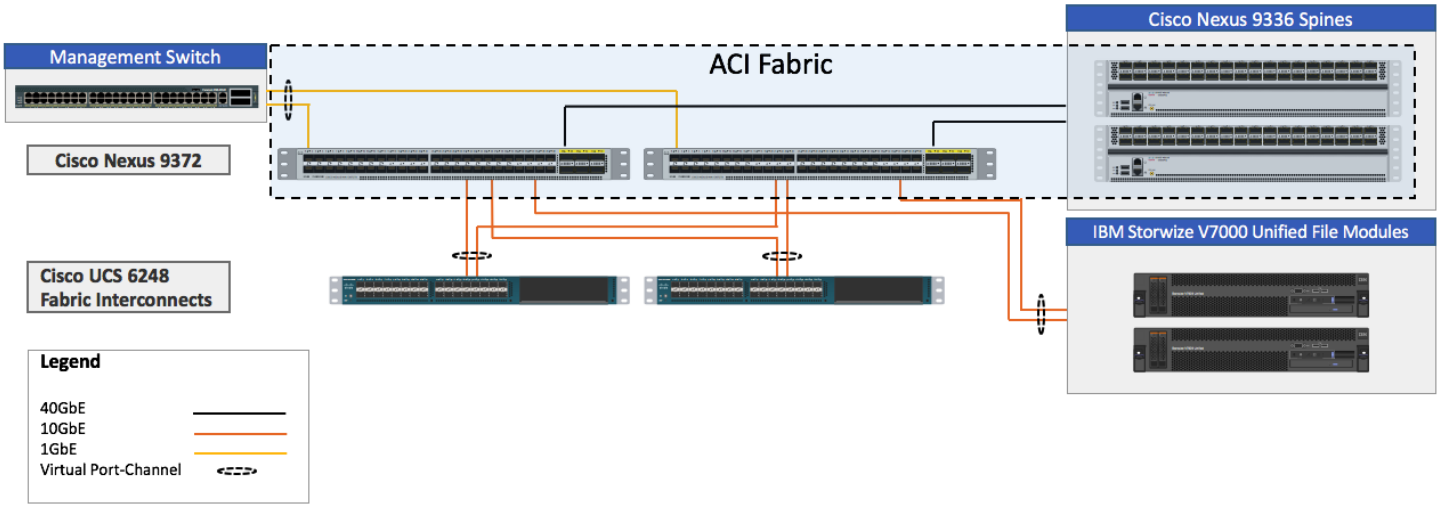
In this VersaStack with ACI design, a pair of redundant Cisco Nexus 9372 leaf switches provide ACI based Ethernet switching fabric for communication between the virtual machine and bare metal compute, NFS and iSCSI based storage and the existing traditional (non-ACI) enterprise networks. Similar to previous 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

In the current VersaStack design, following devices are connected to the ACI fabric using a vPC (Figure 29):

- Cisco UCS FIs
- IBM Storwize V7000 Unified File Modules
- Connection to In-Band management infrastructure switch

Figure 29 Network Design – vPC Enabled Connections



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 foundation services include:

- iSCSI VLANs to provide access to iSCSI datastores including boot LUNs
- NFS VLAN(s) to access VM datastores used by vSphere environment
- A pool of VLANs associated with ACI Virtual Machine Manager (VMM) domain. VLANs from this pool are dynamically allocated by APIC to application end point groups (covered later in the document)

These VLAN configurations are covered below.

VLANs in ACI

VLANs in an ACI do not have the same meaning as VLANs in a regular switched infrastructure. The VLAN tag for a VLAN in ACI is used purely for classification purposes. In ACI, data traffic is mapped to a bridge domain that has a global scope therefore local VLANs on two ports might differ even if they belong to the same broadcast domain. Rather than using forwarding constructs such as addressing or VLANs to apply connectivity and policy, ACI utilizes End Point Groups (EPGs) to establish communication between application endpoints. The VLAN to EPGs mapping is covered under the ACI section later in the document.

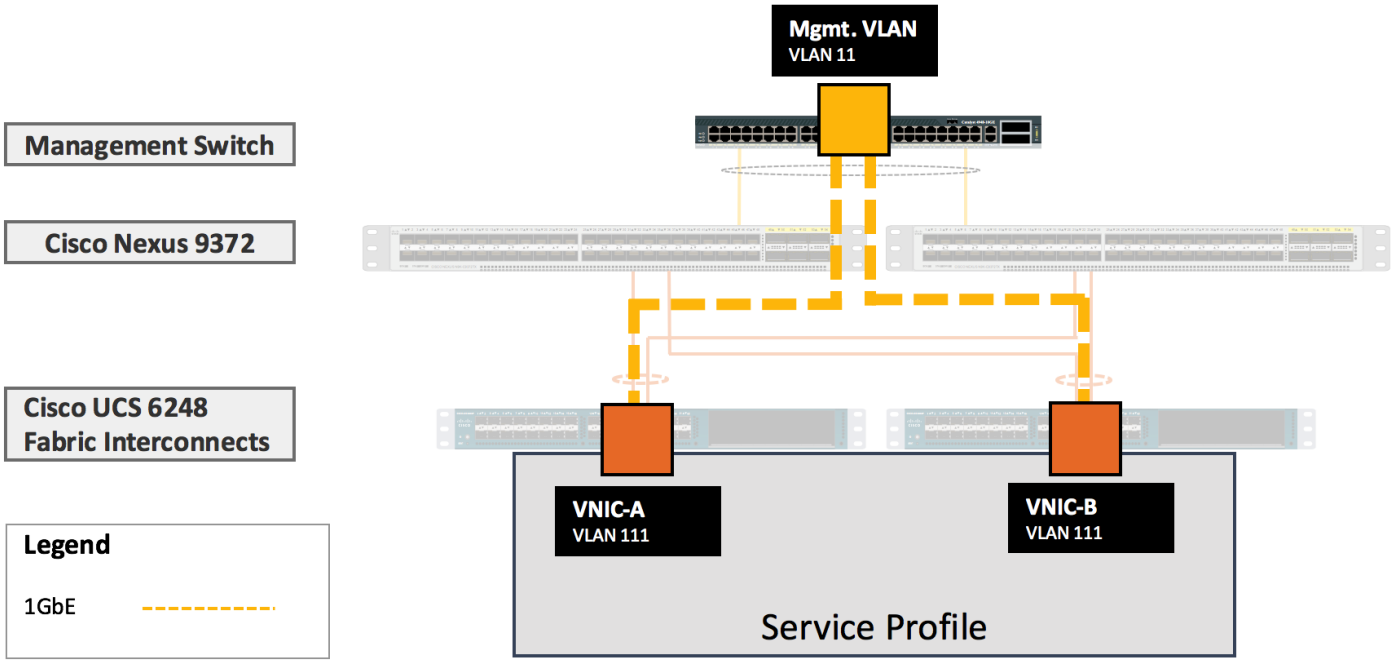
In the VersaStack with ACI design, different VLANs are utilized when two or more paths are mapped to the same EPG. For example, when an ESXi host is setup for iSCSI-A connectivity, the VLAN tag used to identify this traffic on Cisco UCS FI is 3030 but the VLAN tag used on the IBM FlashSystem V9000 is 3031. However, since the ACI fabric provides the necessary VLAN stitching, the iSCSI-A VMkernel port on ESXi host and the interface on the IBM V9000 are in the same IP subnet. Figure 31 highlights the VLAN usage for the two iSCSI paths.

In-Band Management VLAN Configuration

To provide in-band management access to the ESXi hosts and the infrastructure/management Virtual Machines (VMs), the existing management infrastructure switch is connected to both the Cisco Nexus 9372 leaf switches using a vPC. In this design, VLAN 11 is the pre-existing management VLAN on the infrastructure management switch. VLAN 111 is configured as the management VLAN in Cisco UCS

configuration. Within Cisco UCS service profile, VLAN 111 is enabled on vNIC-A and vNIC-B interfaces as shown in Figure 30.

Figure 30 Network Design – VLAN Mapping for in-band Management

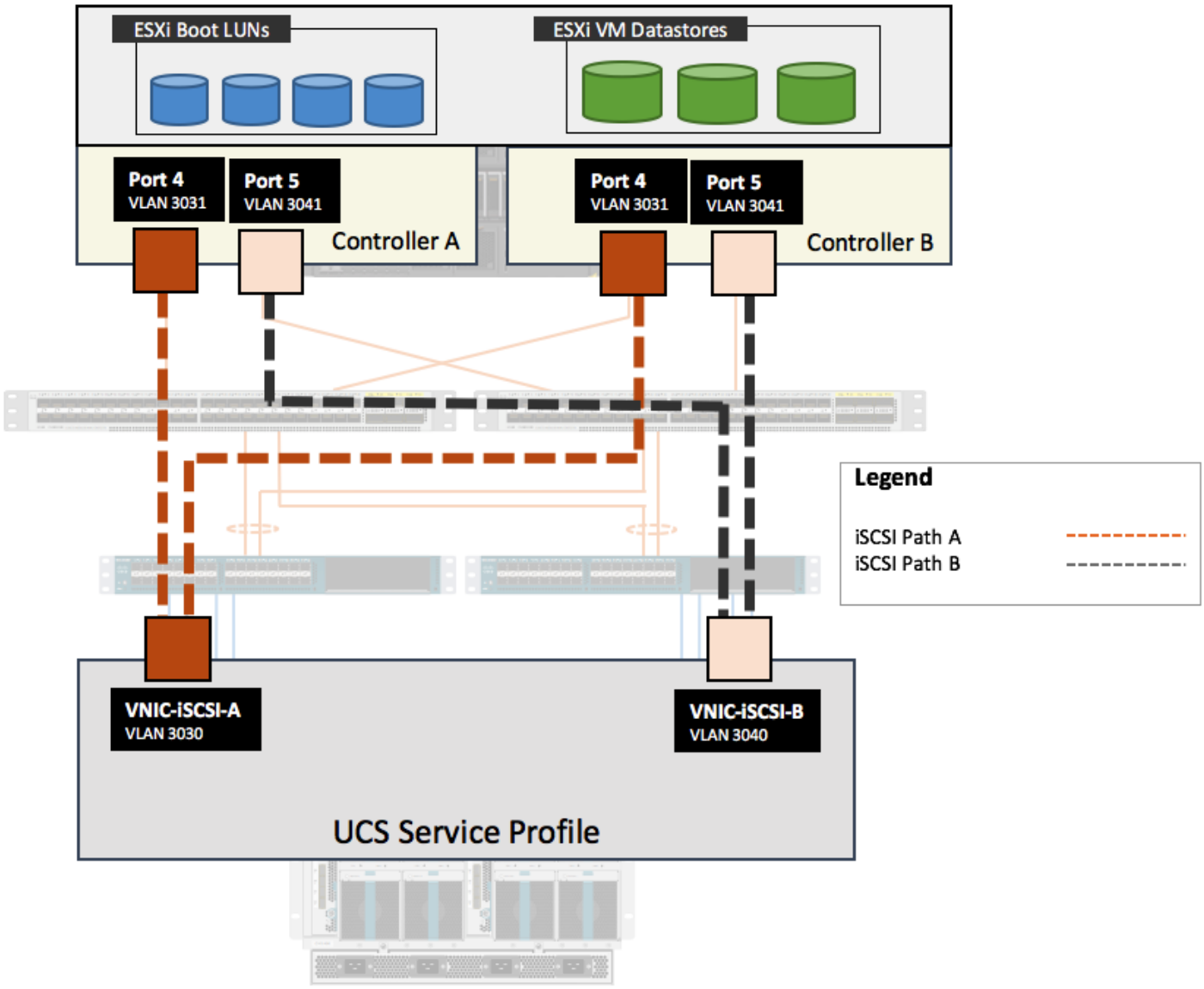


iSCSI VLAN Configuration

To provide redundant iSCSI paths, two VMkernel interfaces tied to separate NICs are configured for host to storage connectivity. In this configuration, each VMkernel port becomes 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 FlashSystem V9000 controllers, VLAN 3030 is configured on the Cisco UCS and VLAN 3031 is configured on the IBM FlashSystem V9000 controller interfaces. To setup iSCSI-B path between the ESXi hosts and the IBM FlashSystem V9000 controllers, VLAN 3040 is configured on the Cisco UCS and VLAN 3041 is configured on the appropriate IBM FlashSystem V9000 controller interfaces. Within Cisco UCS service profile, these VLANs are enabled on vNIC-iSCSI-A and vNIC-iSCSI-B interfaces respectively as shown in Figure 31. The iSCSI VLANs are set as native VLANs on the vNICs to enable boot from SAN functionality.

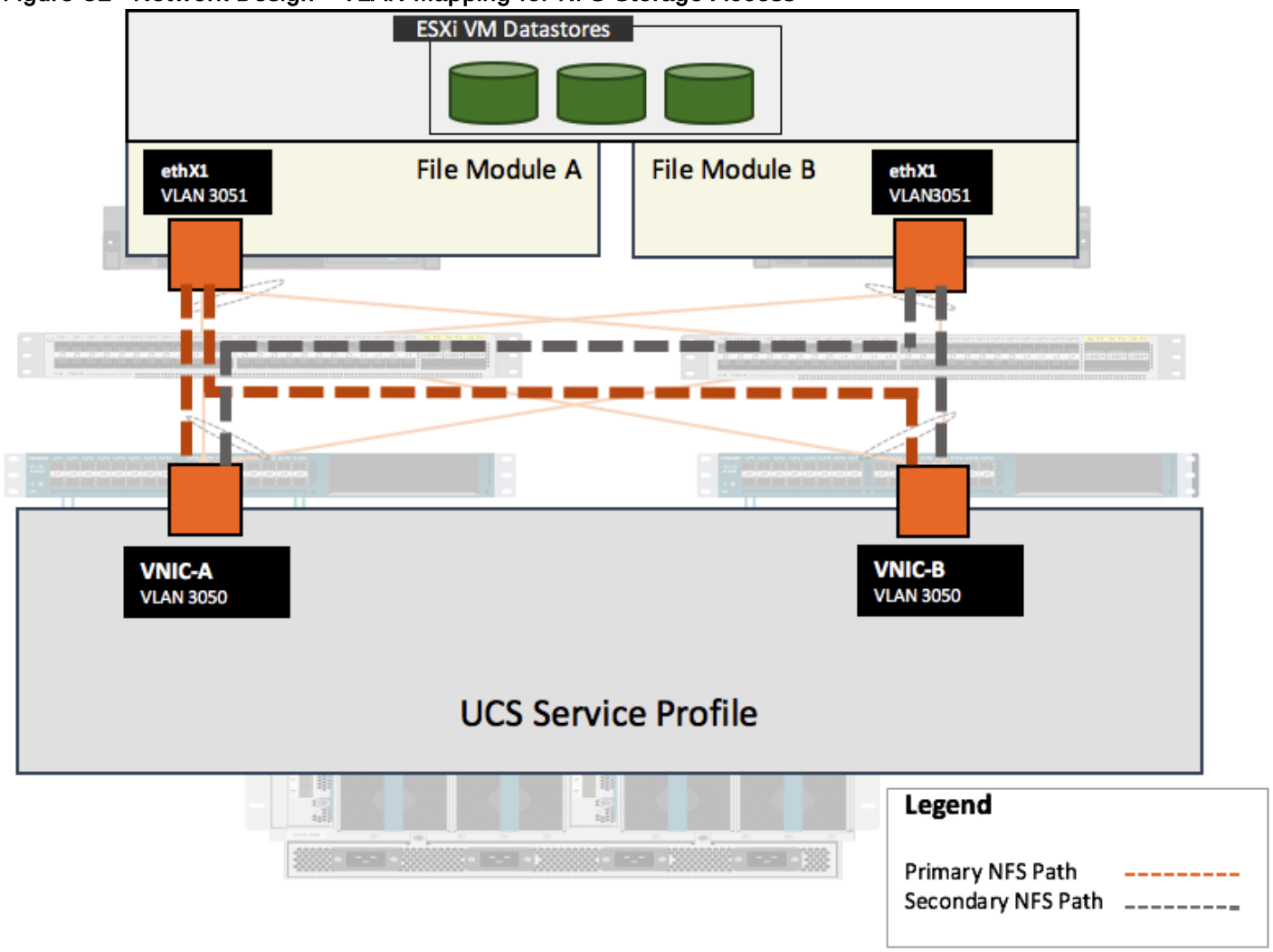
Figure 31 Network Design – VLAN mapping for iSCSI Storage Access



NFS VLAN Configuration

To setup NFS connectivity between the ESXi hosts and the IBM Storwize V7000 Unified File Modules, VLAN 3050 is configured on the Cisco UCS and VLAN 3051 is configured on the File Modules. Within Cisco UCS service profile, VLAN 3050 is enabled on both vNIC-A and vNIC-B interfaces as shown in Figure 32.

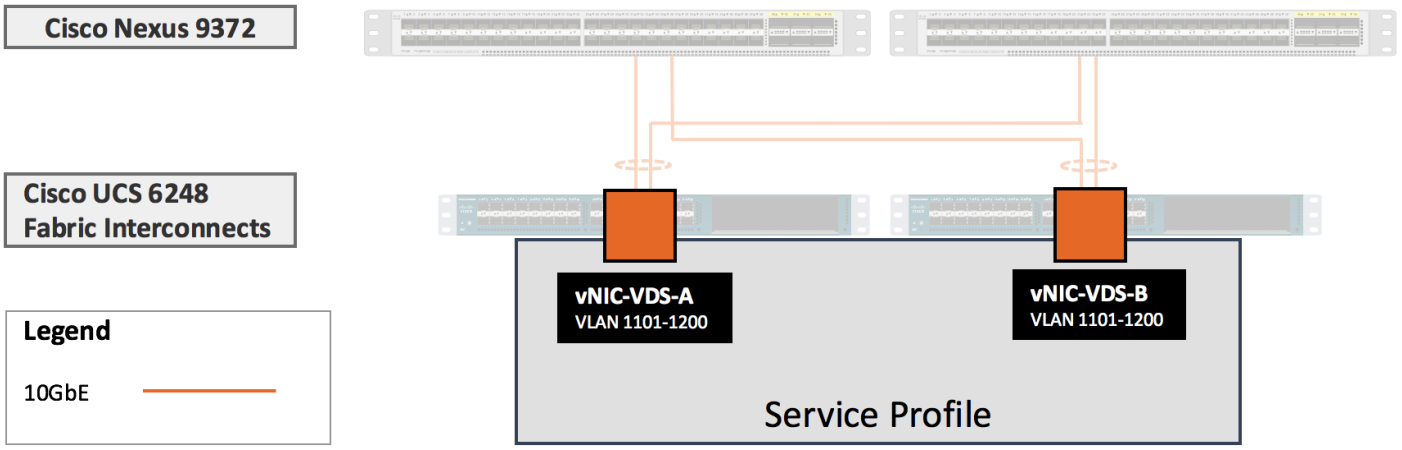
Figure 32 Network Design – VLAN Mapping for NFS Storage Access



Virtual Machine Networking VLANs for VMware vDS

When using VMware VDS in an ACI setup, a pool of 100 VLANs,1101-1200, is defined to be used on-demand by the VM Networking. VLANs from this pool are dynamically assigned to the EPGs mapped to the Virtual Machine Manager (VMM) domain. Since Cisco APIC does not manage or configure the Cisco UCS FIs, FIs are treated as unmanaged switches in the middle and the pool of VLANs needed to enable VM networking also has to be defined on Cisco UCS and enabled on the vNICs vNIC-VDS-A and vNIC-VDS-B as shown in Figure 33.

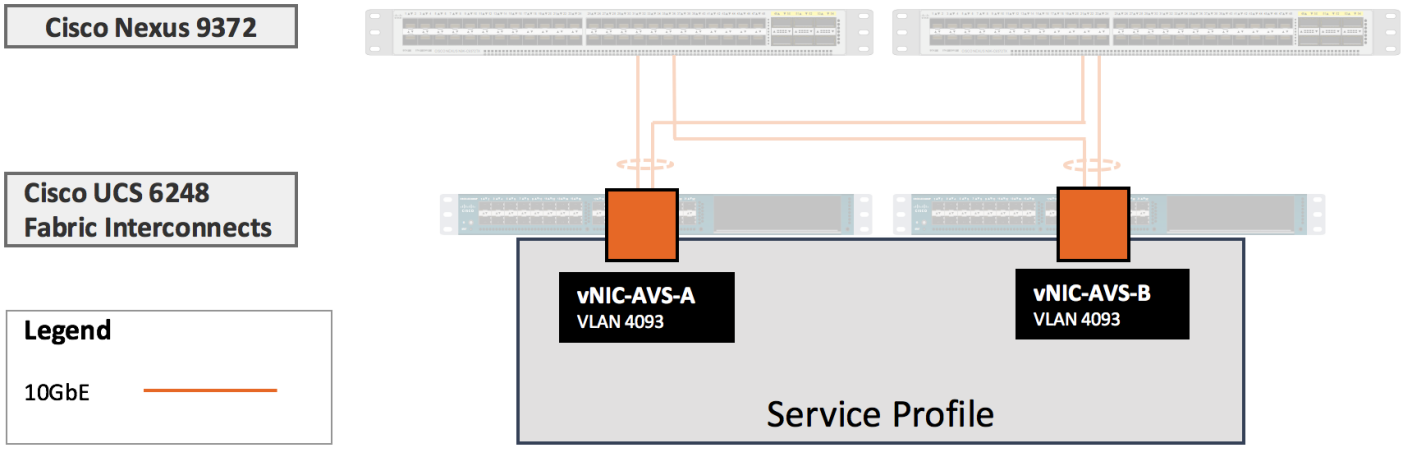
Figure 33 Network Design – VLANs for VMM Domain Associated with VMware VDS



Virtual Machine Networking VLANs for Cisco AVS

When using Cisco Application Virtual Switch (AVS) VxLAN mode in an ACI setup, EPGs associated with the VMM domain utilize VxLAN and a single carrier VLAN, 4093, needs to be defined in the Cisco UCS and enabled on the vNIC-AVS-A and vNIC-AVS-B interfaces as shown in Figure 34.

Figure 34 Network Design – VLANs for VMM Domain Associated with Cisco AVS



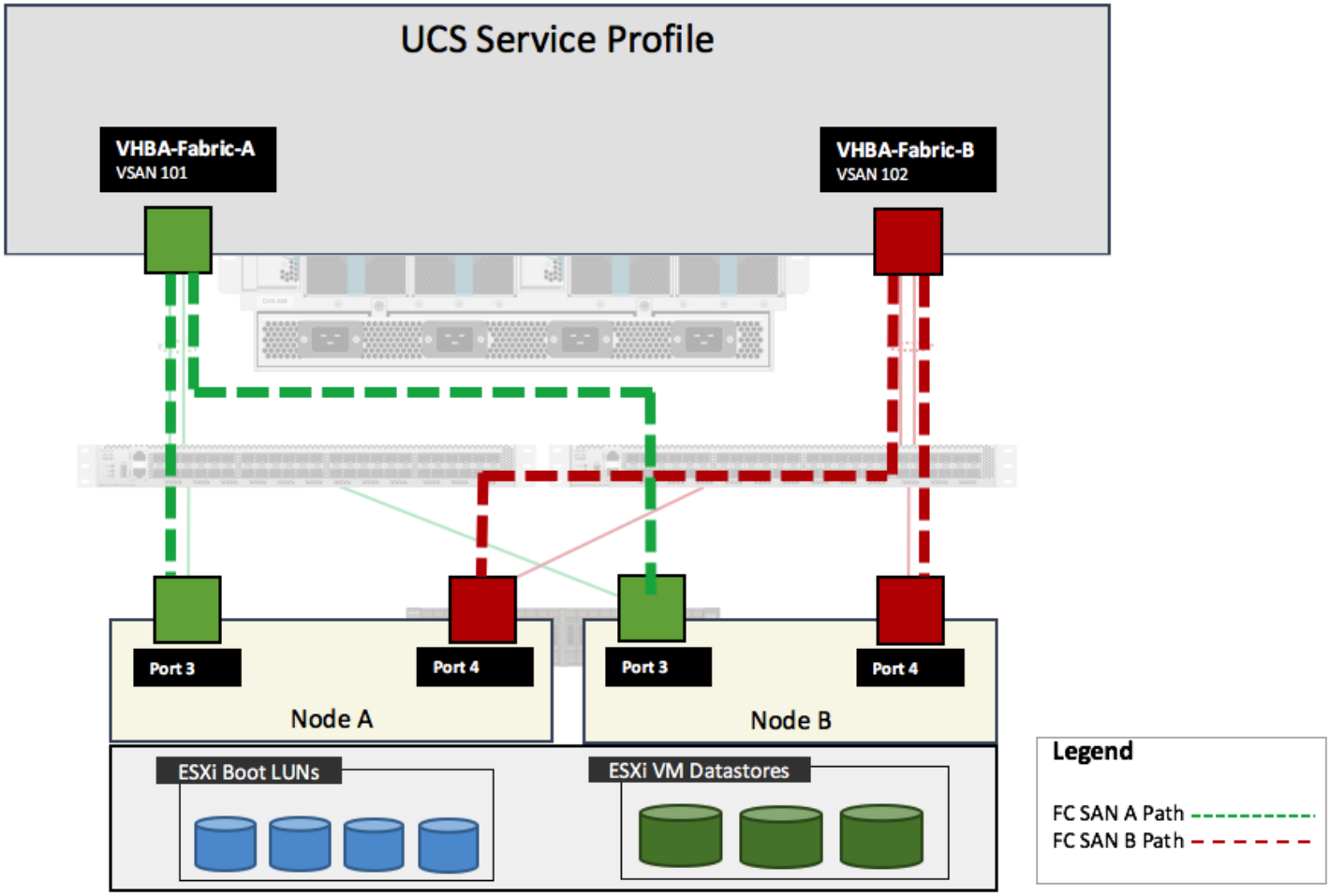
VSAN Design

In the VersaStack with ACI design, isolated fabric topologies are created on the Cisco MDS 9148S switches using VSANs to support Host and Cluster interconnect traffic. Use of VSANs allows the isolation of traffic within specific portions of the storage area network.

Redundant host VSANs support host or server traffic to the V7000 Controllers through independent fabrics. Cisco UCS FIs connect to the Cisco MDS switches using Port Channels while IBM V7000 controllers are connected using independent FC ports. VSAN 101 and VSAN 102 provide the dual SAN paths as shown in Figure 35.

Additional VSANs might need to be defined on the Cisco MDS switches for connectivity between IBM storage controllers and the IBM disk enclosures

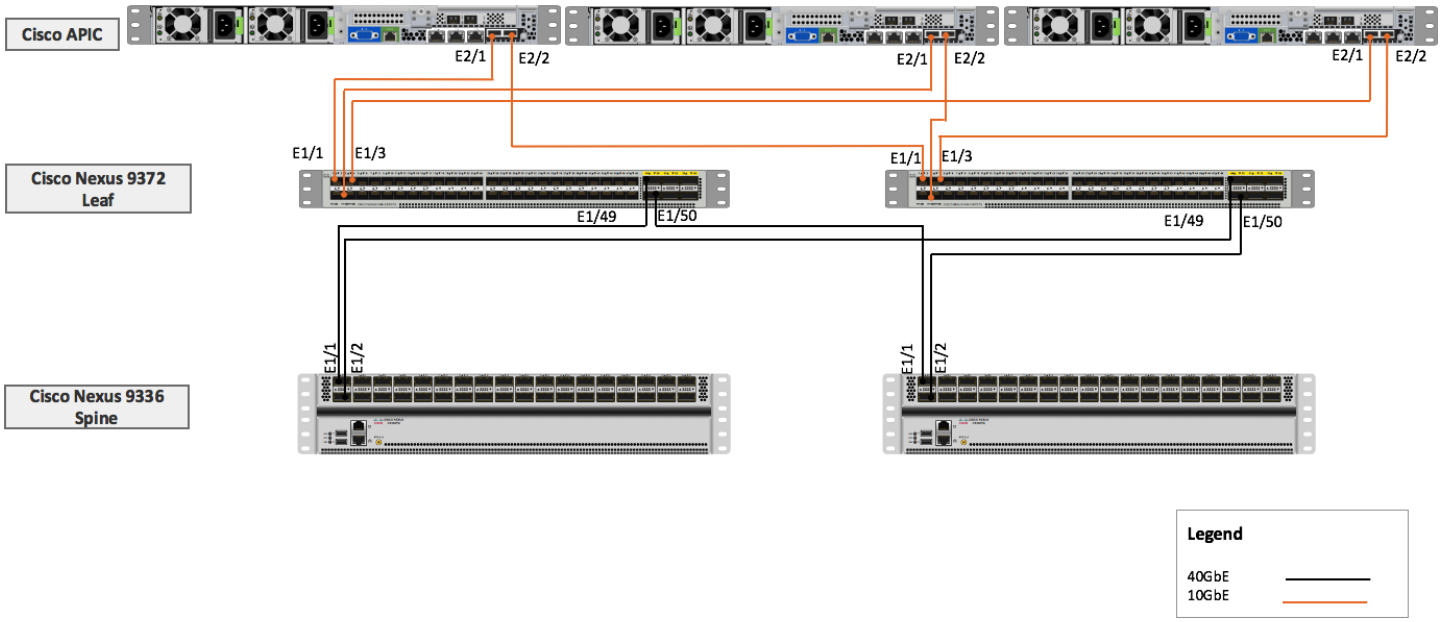
Figure 35 VSAN Design – Cisco UCS Server to IBM Storwize V7000 Unified SAN Connectivity



Application Centric Infrastructure Design

The Cisco Application Centric Infrastructure (ACI) fabric consists of discrete components that operate as routers and switches but are provisioned and monitored as a single entity. These components and the integrated management allow Cisco ACI to provide advanced traffic optimization, security, and telemetry functions for both virtual and physical workloads. This CVD utilizes Cisco ACI fabric based networking configurations as defined in the upcoming sections. Cisco ACI fabric is deployed in a leaf-spine architecture controller by a minimum of three Application Policy Infrastructure Controllers (APIC) as shown in Figure 36.

Figure 36 Cisco ACI Fabric – Physical Layout



Since the network provisioning in ACI based VersaStack is quite different from traditional Cisco Nexus 9000 NxOS based VersaStack and requires a basic knowledge of some of the core ACI concepts.

ACI Components

Leaf switches: The ACI leaf provides physical connectivity for servers, storage devices and other network elements as well as enforces ACI policies. A leaf typically is a fixed form factor switch such as the Cisco Nexus 9372PX switch used in the current design. Leaf switches also provide a connection point to the existing enterprise or service provider infrastructure. The leaf switches provide both 10G and 40G Ethernet ports for connectivity.

In the VersaStack with ACI design, Cisco UCS FI, IBM storage devices and Cisco Nexus 7000 based WAN/Enterprise routers are connected to both the leaves for high availability.

Spine switches: The ACI spine provides the mapping database function and the connectivity among leaf switches. A spine can be the modular Cisco Nexus 9500 series switch equipped with ACI ready line cards or fixed form-factor switch such as the Cisco Nexus 9336PQ (used in this design). Spine switches provide high-density 40 Gigabit Ethernet connectivity between leaf switches.

Tenant: A tenant is a logical container which can represent an actual tenant, an organization, an application or a construct to easily organize information. From a policy perspective, a tenant represents a unit of isolation. All application configurations in Cisco ACI are part of a tenant. Within a tenant, one or more VRF contexts, one or more bridge domains, and one or more EPGs can be defined according to application requirements.

VersaStack with ACI design requires creation of an infrastructure tenant called "Foundation" to provide compute to storage connectivity for iSCSI based SAN environment as well as to connect the ESXi servers to the NFS datastores. The design also utilizes the predefined "common" tenant to provide in-band management infrastructure connectivity and to host core services required by all the tenants such as DNS, AD etc. In addition, each subsequent application deployment requires creation of a dedicated tenant.

Contexts: Tenants can be further divided into contexts, which directly relate to Virtual Routing and Forwarding (VRF) instances (separate IP spaces). Contexts provide a way to further separate the

organizational and forwarding requirements for a given tenant. Because contexts use separate forwarding instances, IP addressing can be duplicated in across contexts for multitenancy. In the current design, each tenant typically used a single VRF.

Application Profile: An application profile models application requirements and contains one or more End Point Groups (EPGs) as necessary to provide the application capabilities. Depending on the application and connectivity requirements, VersaStack with ACI design uses multiple application profiles to define multi-tier applications as well as storage connectivity.

Bridge Domain: A bridge domain represents a L2 forwarding construct within the fabric. One or more EPG can be associated with one bridge domain or subnet. In ACI, a bridge domain represents the broadcast domain and the bridge domain might not allow flooding and ARP broadcast depending on the configuration. The bridge domain has a global scope, while VLANs do not. Each endpoint group (EPG) is mapped to a bridge domain. A bridge domain can have one or more subnets associated with it and one or more bridge domains together form a tenant network

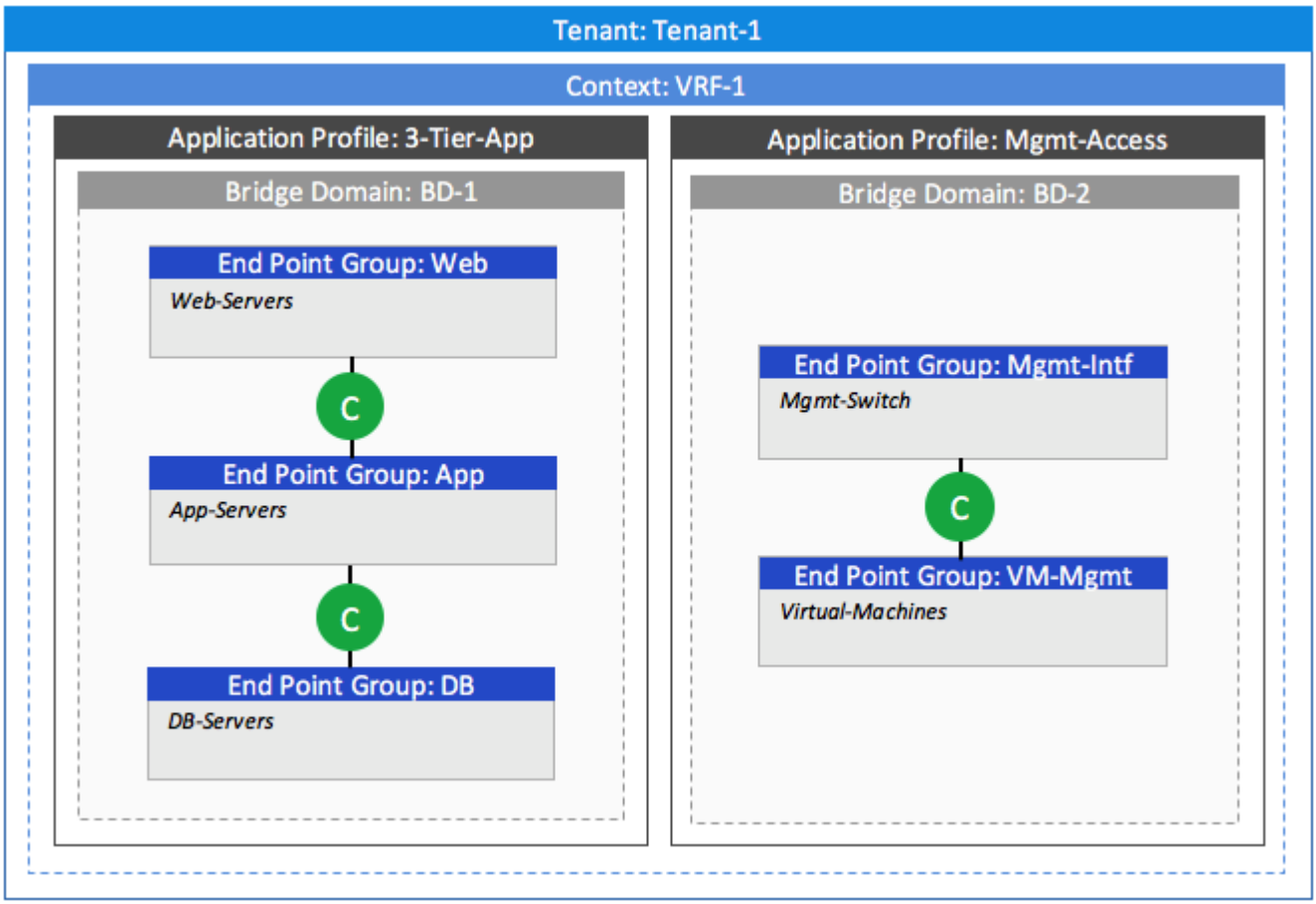
End Point Group: An End Point Group (EPG) is a collection of physical and/or virtual end points that require common services and policies. An EPG example is a set of servers or VMs on a common VLAN segment providing a common function or service. While the scope of an EPG definition is much wider, in the simplest terms an EPG can be defined on a per VLAN basis where all the servers or VMs on a common LAN segment become part of the same EPG.

In the VersaStack with ACI design, various application tiers, ESXi VMkernel ports for iSCSI, NFS and vMotion, and interfaces on IBM storage devices are mapped to various EPGs. The design details are covered in the later sections.

Contracts: Contracts define inbound and outbound traffic filter, QoS rules and Layer 4 to Layer 7 redirect policies. Contracts define the way an EPG can communicate with other EPGs depending on the application requirements. Contracts are defined using provider-consumer relationships; one EPG provides a contract and another EPG(s) consume that contract. Contracts utilize filters to limit the traffic between the applications to certain ports and protocols.

Figure 37 covers relationship between various ACI elements. As shown in the figure, a Tenant can contain one or more application profiles and an application profile can contain one or more end point groups (EPGs). The devices in the same EPG can talk to each other without any special configuration. Devices in different EPGs can talk to each other using contracts and associated filters. A tenant can also contain one or more VRFs and bridge domains. Different application profiles and EPGs can utilize the same VRF or the bridge domain.

Figure 37 ACI - Relationship Between Major Components



End Point Group (EPG) Mapping in a VersaStack Environment

In the VersaStack with ACI infrastructure, traffic is associated with an EPG in one of the following two ways:

- Statically mapping a Path/VLAN to an EPG (Figure 38).
- Associating an EPG with a Virtual Machine Manager (VMM) domain thereby allocating a VLAN or VxLAN dynamically from a pre-defined pool in APIC (Figure 39).

Figure 38 ACI - Static Path Binding

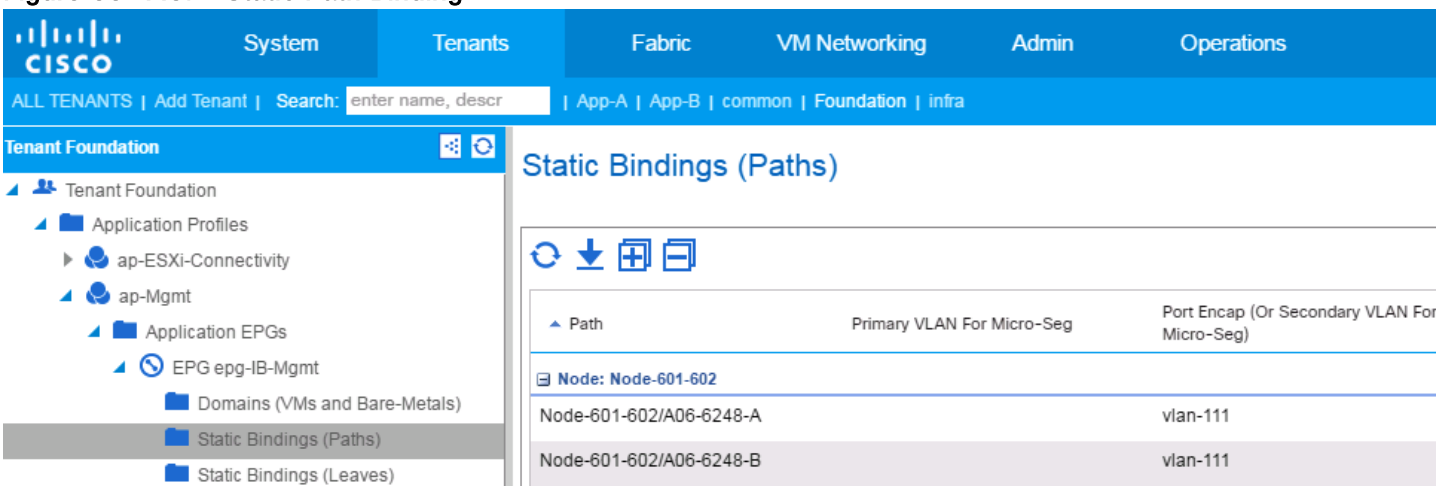
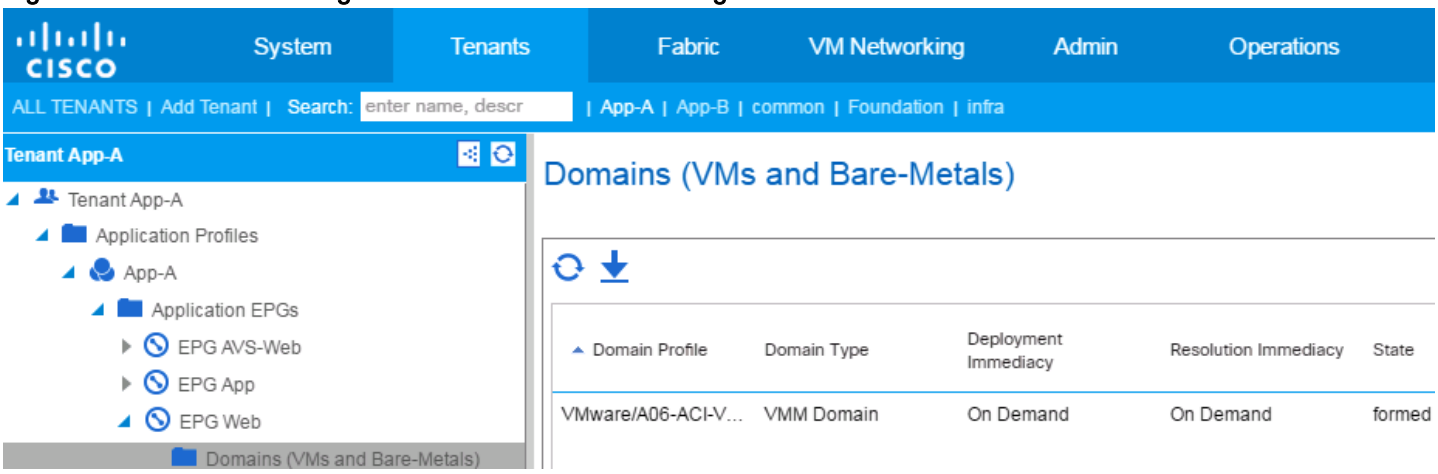


Figure 39 ACI – EPG Assigned to Virtual Machine Manager



Statically mapping of Path/VLAN to an EPG is useful for:

- Mapping iSCSI and NFS VLANs on both the Cisco UCS and the IBM storage systems to appropriate EPGs
- Mapping bare metal servers to an EPG
- Mapping vMotion VLANs on the Cisco UCS/ESXi Hosts to an EPG
- Mapping the management VLAN(s) from the existing infrastructure to an EPG in the common tenant. This EPG is utilized for in-band management access by both ESXi hosts and the VMs

Dynamically mapping a VLAN to an EPG by defining a VMM domain is useful for:

- Deploying VMs in a multi-tier Application requiring one or more EPGs
- Deploying application specific IP based storage access within the tenants

Virtual Machine Networking

The Cisco APIC automates the networking for all virtual and physical workloads including access policies and L4-L7 services. When connected to the VMware vCenter, APIC controls the VM related virtual distributed switching as detailed in the following sections.

Virtual Machine Manager (VMM) Domains

In a VMware vCenter environment, Cisco APIC controls the creation and configuration of the Virtual Distributed Switch (VDS) or the Cisco Application Virtual Switch (AVS). Once the virtual distributed switches are deployed, APIC communicates with the switches to publish network policies that are applied to the virtual workloads including creation of port groups for VM association. A VMM domain contains multiple EPGs and hence multiple port groups. To position an application, the application administrator deploys the VMs and places the VM NIC into the port group defined for the appropriate application tier.

Virtual Switching Architecture

As stated previously, a tenant application deployment utilizes port groups on APIC controlled distributed switch (Cisco AVS or VDS). However, for some of the core connectivity such as out of band management access, storage LUN access using iSCSI, and infrastructure datastore access using NFS, vSphere vSwitches are deployed. To support the multi-virtual-switch requirement, multiple vNIC interfaces are setup for each Cisco UCS services profile. Storage, management and VM data VLANs are then enabled on the appropriate

vNIC interfaces. Figure 40 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 a Foundation Services vSwitch and iSCSI-A and iSCSI-B traffic is handled by two dedicated iSCSI vSwitches. The resulting ESXi host configuration therefore has a combination of 3 vSwitches and a single APIC-Controlled distributed switch which handles tenant (application) specific traffic.

Figure 40 ACI – ESXi Host vNIC and vmk Distribution for iSCSI based Storage Access

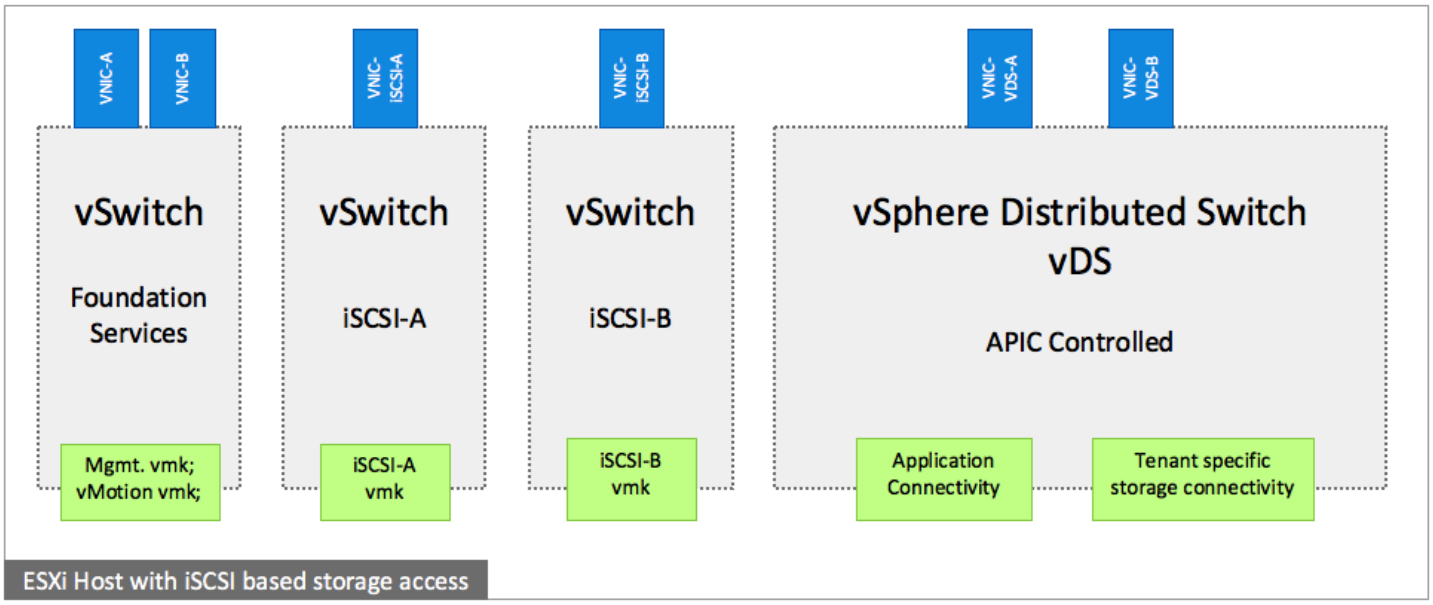
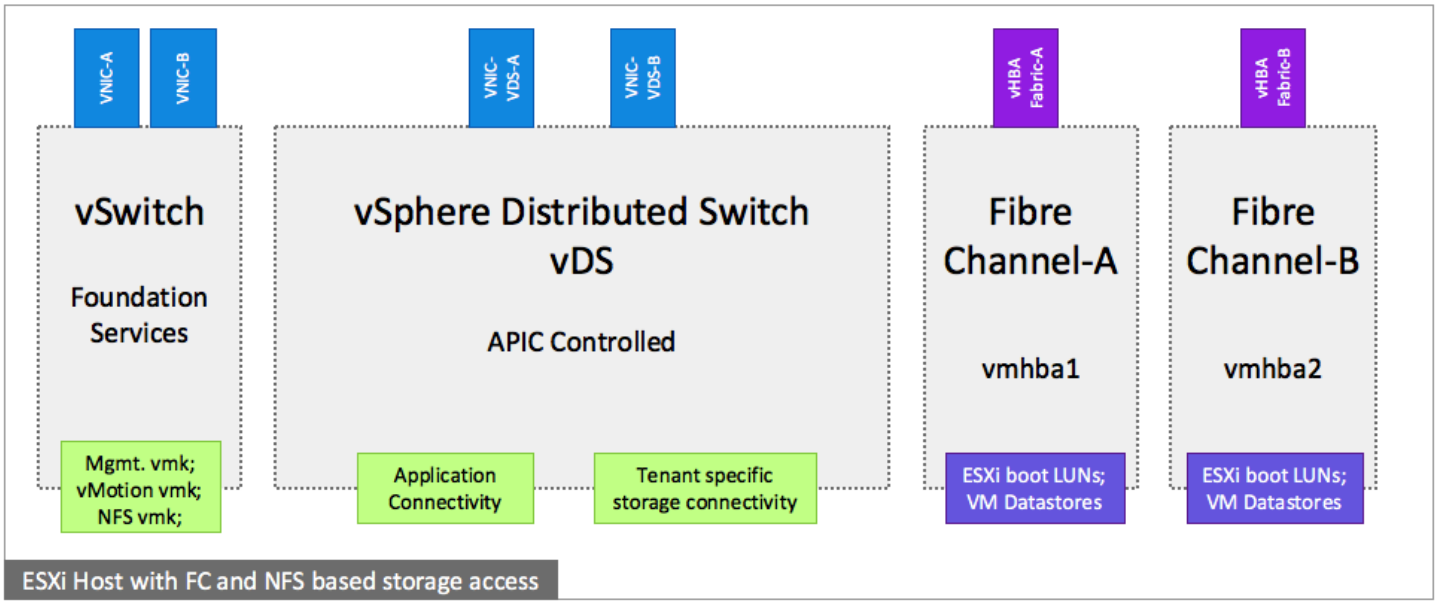


Figure 41 below shows similar configuration for an FC and NFS connected ESXi host. In this case, In-band management, NFS and vMotion traffic is handled by a Foundation Services vSwitch and the Fibre Channel SAN-A and SAN-B traffic is handled by two dedicated vHBAs. The resulting ESXi host configuration therefore has a combination of 1 vSwitch and one APIC controlled distributed switch.

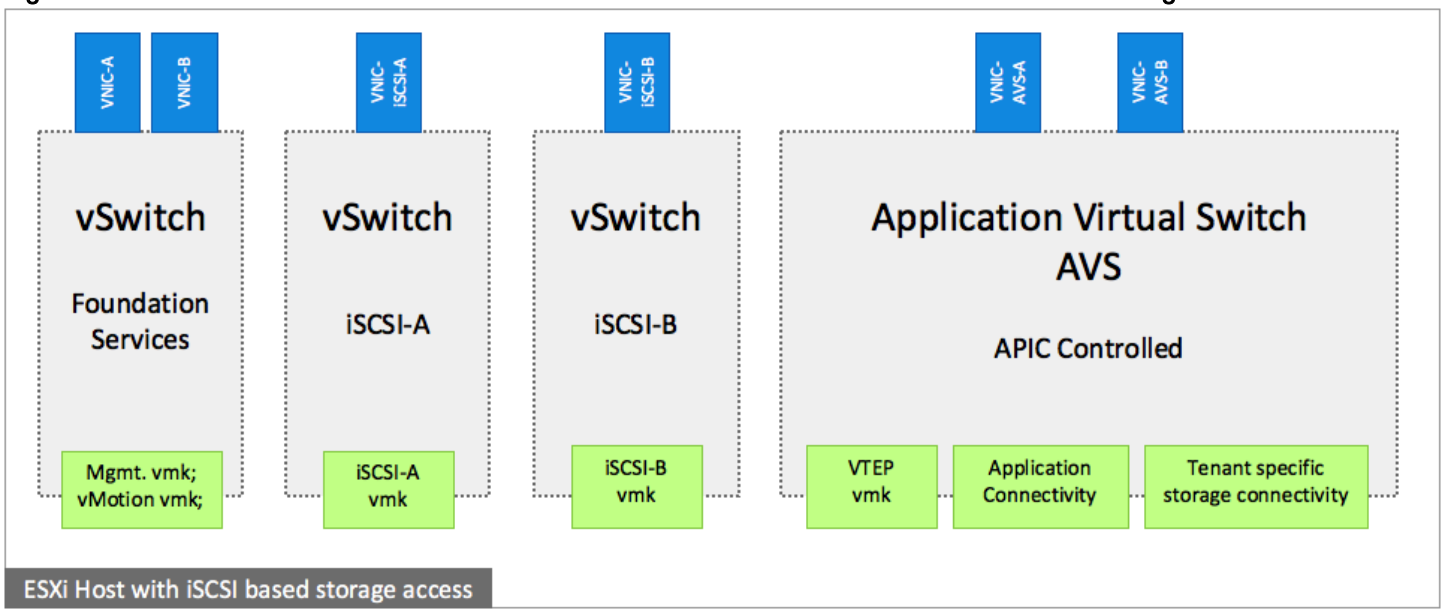
On ESXi servers where VMware vCenter VM is hosted on the NFS datastore, it is critical that NFS VLAN is placed on the vSwitch and not on the distributed switch to prevent loss of access to vCenter. In the event of an environment failure, a distributed switch might not be operational to provide access to NFS datastore where vCenter resides. It is also important to place all iSCSI VMK ports on the iSCSI vSwitches to prevent loss of mapped boot LUNs.

Figure 41 ACI – ESXi Host vNIC, vHBA and vmk Distribution for FC/NFS based Storage Access



When the Cisco AVS in VXLAN mode is used instead of the VMware vDS, a VTEP vmk port is also provisioned on the Cisco AVS as shown in Figure 42. The VTEP interface provides necessary VxLAN encapsulation. In VxLAN mode, Cisco AVS only requires the infrastructure VLAN, 4093, to be enabled on the vNIC interfaces.

Figure 42 ACI – ESXi Host vNIC and vmk Distribution for Cisco AVS and iSCSI Based Storage



Onboarding Infrastructure Services

In an ACI fabric, all the applications, services and connectivity between various elements are defined within the confines of tenants, application profiles, bridge domains and EPGs. The tenant configured to provide the infrastructure services is named *Foundation*. The *Foundation* tenant enables compute to storage connectivity for accessing iSCSI and NFS datastores and provides ESXi hosts and VMs access to existing management infrastructure. The Foundation tenant comprises of a single bridge domain called *bd-foundation-internal*. This bridge domain is shared by all the EPGs in the *Foundation* tenant. Since there are no overlapping IP address space requirements, *Foundation* tenant consists of a single context (VRF) called *Foundation.*

Foundation tenant is configured with two different Application Profiles:

- *ap-ESXi-Connectivity*: This application profile contains EPGs to support compute to storage connectivity as well as VMware vMotion. The three EPGs defined under this application profile are: "iSCSI", "NFS" and "vMotion"
- *ap*-Mgmt: This application profile provides ESXi host and VMs connectivity to existing management segment through the *Common* tenant (details covered later in this section)

As outlined before, this design guide covers two different IP based storage connectivity models:

- An iSCSI based storage connectivity design showcasing IBM FlashSystem V9000
- An NFS based storage connectivity designs showcasing IBM Storwize V7000 Unified File Modules

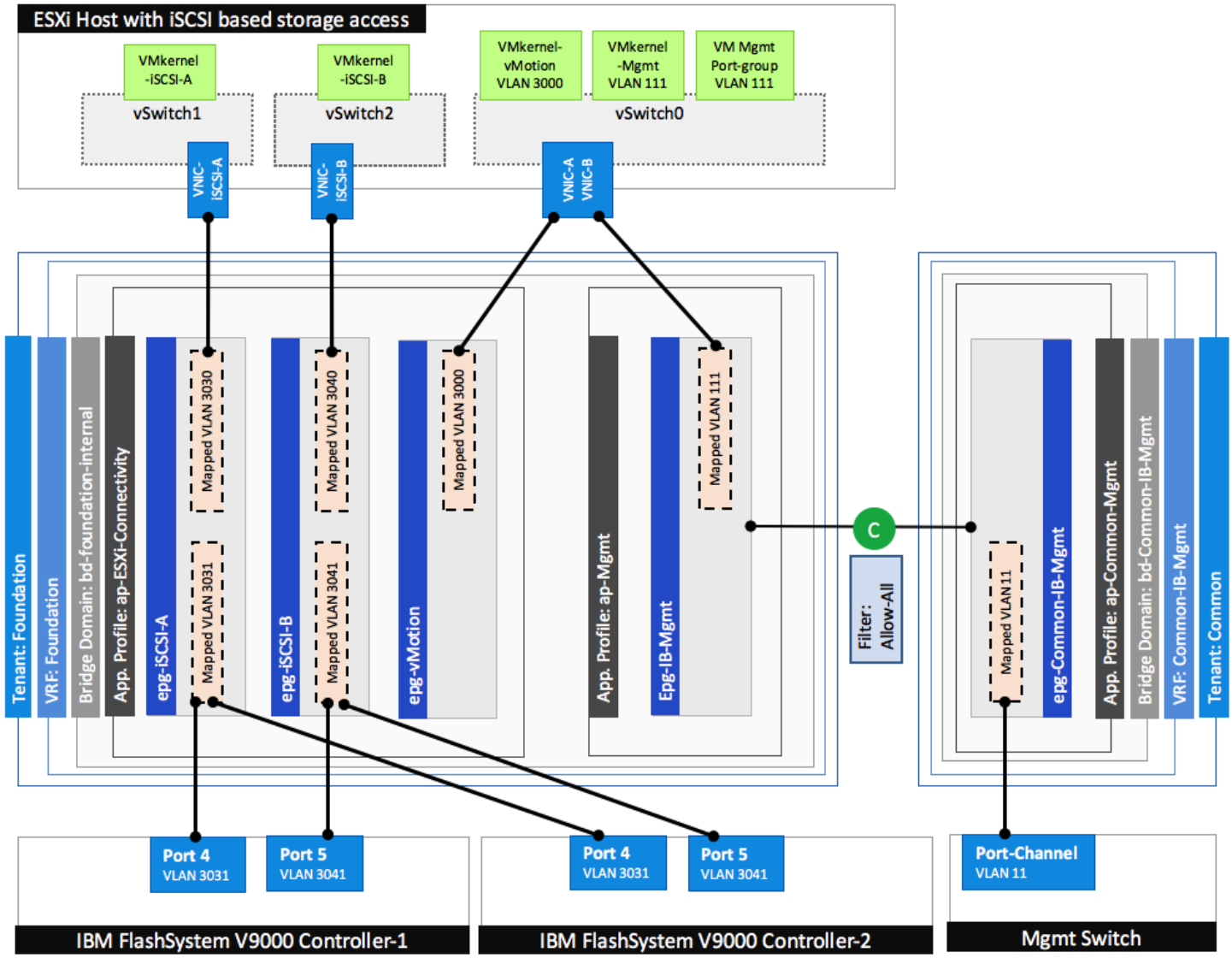
In design option 1, ESXi boot LUNs are accesses using iSCSI connectivity to IBM FlashSystem V9000 while the option 2 relies on FC connectivity to IBM Storwize V7000 controllers for the ESXi boot LUN access

The details of the ACI constructs for core infrastructure services and the IP bases storage access are discussed below.

Foundation Tenant EPG Design for iSCSI based Storage

Figure 43 shows an overview of ACI deign covering connectivity details and the relationship between various ACI elements for an iSCSI based storage design.

Figure 43 ACI – Foundation Tenant EPG Design for iSCSI Storage



Following ACI constructs defined the *Foundation* Tenant configuration for the iSCSI based storage access:

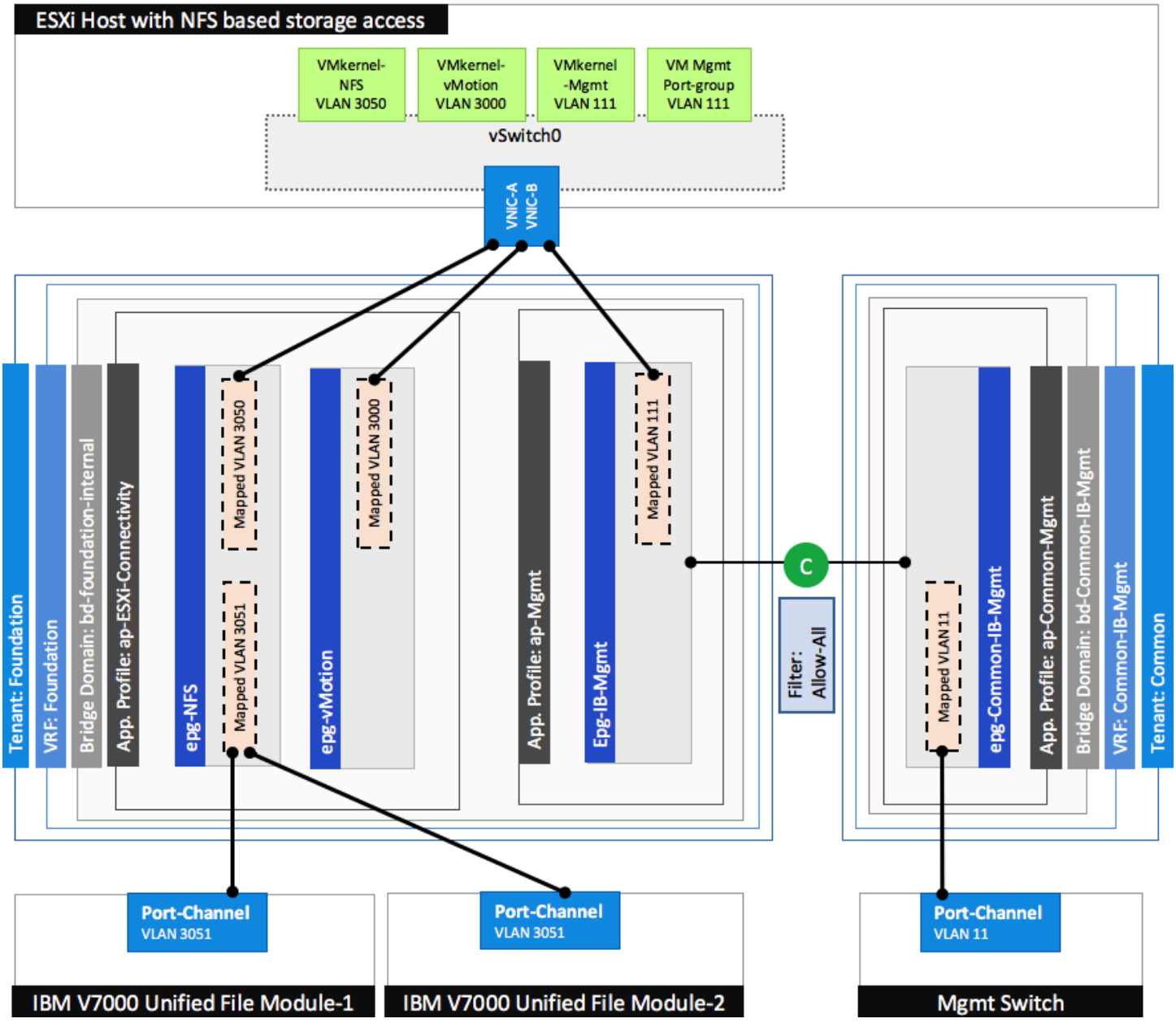
- Tenant: Foundation
- VRF: Foundation
- Bridge Domain: bd-foundation-internal
- Application Profile *ap-ESXi-Connectivity* consist of three EPGs:
  - *epg-iSCSI-A* statically maps the VLANs associated with iSCSI-A interfaces on the IBM storage controllers (VLAN 3031) and Cisco UCS Fabric Interconnects (3030)
  - *epg-iSCSI-B* statically maps the VLANs associated with iSCSI-B interfaces on the IBM storage controllers (VLAN 3041) and Cisco UCS Fabric Interconnects (3040)
  - *epg-vMotion* statically maps vMotion VLAN (3000) on the Cisco UCS Fabric Interconnects
- Application Profile *ap-Mgmt* consist of one EPGs:

— *epg-IB-Mgmt* statically maps the management VLAN (111) on the Cisco UCS Fabric Interconnects. This EPG is configured to provide VMs and ESXi hosts access to the existing management network as covered in the next section.

Foundation Tenant EPG Design for NFS based Storage

An overview of ACI design covering connectivity details and the relationship between various ACI elements of an NFS based storage design are covered in Figure 44.

Figure 44 ACI – Foundation Tenant EPG Design for NFS Storage



Following ACI constructs defined the *Foundation* Tenant configuration for NFS based storage access:

- Tenant: Foundation
- VRF: Foundation
- Bridge Domain: bd-foundation-internal

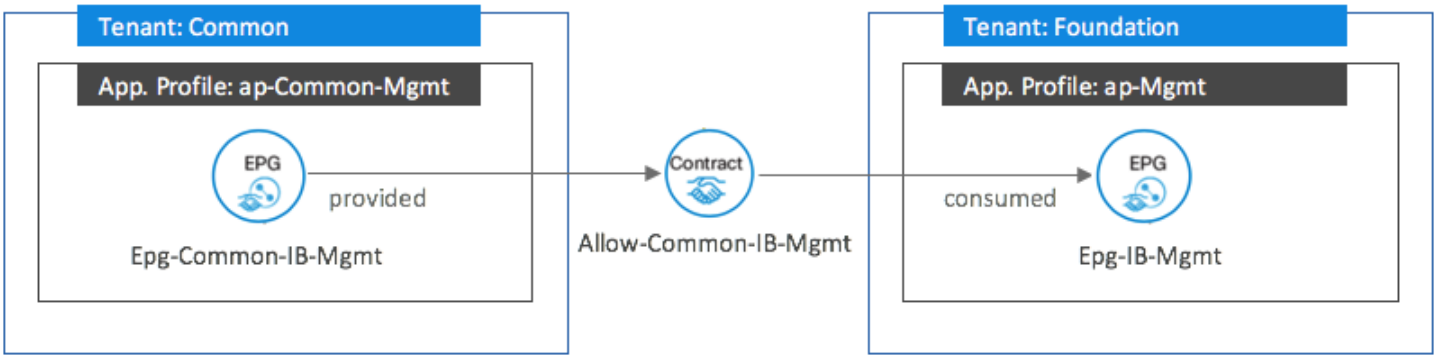
- Application Profile *ap-ESXi-Connectivity* consist of two EPGs:
  - *epg-NFS* statically maps the VLANs associated with NFS interfaces on the IBM File Modules (VLAN 3051) and Cisco UCS Fabric Interconnects (3050)
  - *epg-vMotion* statically maps vMotion VLAN (3000) on the Cisco UCS Fabric Interconnects
- Application Profile *ap-Mgmt* consist of one EPGs:
  - *epg-IB-Mgmt* statically maps the management VLAN (111) on the Cisco UCS Fabric Interconnects. This EPG is configured to provide VMs and ESXi hosts access to the existing management network as covered in the next section.

Enabling Management Access through Common Tenant

To provide ESXi hosts and VMs access to management segment and common services such as Active Directory (AD), Domain Name Services (DNS), management and monitoring software etc., inter-tenant contracts are utilized. Cisco ACI fabric provides a predefined tenant named *common* to host the services and connectivity that can be easily shared by other tenants in the system. The policies defined in the *common* tenant are usable by all the tenants without any special configurations. By default, in addition to the locally defined contracts, all the tenants in ACI fabric can "consume" the contracts defined in the "common" tenant.

In the VersaStack environment, access to the management segment is provided as shown Figure 45.

Figure 45 ACI – Providing Management Access through the Common Tenant



To provide this access:

- EPG *epg-Common-IB-Mgmt* is defined in the *common* tenant.
- *epg-Common-IB-Mgmt* statically maps the management VLAN (11) on the current management switch
- *epg-Common-IB-Mgmt* "provides" a contract *Allow-Common-IB-Mgmt*
- ESXi hosts and infrastructure related VMs become part of the EPG *epg-IB-Mgmt* and access the management segment by "consuming" the *Allow-Common-IB-Mgmt* contract.
- Tenant VMs can also access the common management segment by "consuming" the same contract
- The contract filters can be configured to only allow specific services related ports

Onboarding Multi-Tier Application

The ACI constructs for a multi-tier application deployment include defining a new tenant, VRF(s), bridge domain(s), application profile(s), end point group(s), and the contract(s) to allow communication between various tiers of the application. Figure 46 provides an overview of the constructs required for deploying a sample 2-tier application. To deploy a sample 2-Tier application, following elements are configured:

- A new Tenant called *App-A* is defined to host the application
- A VRF called *vrf-App-A* is defined under the tenant to provide the tenant IP address space
- A bridge domain *bd-App-A-Internal* is associated with the tenant
- An application profiles, *ap-App-A* is utilized to deploy the application.
- Two EPGs, Web and App are associated with the VMM domain to host Web and App/DB tiers of the application
- A contract to allow communication between the two application tiers is defined. This contract is "provided" by the EPG *App* and "consumed" by the EPG *Web.*

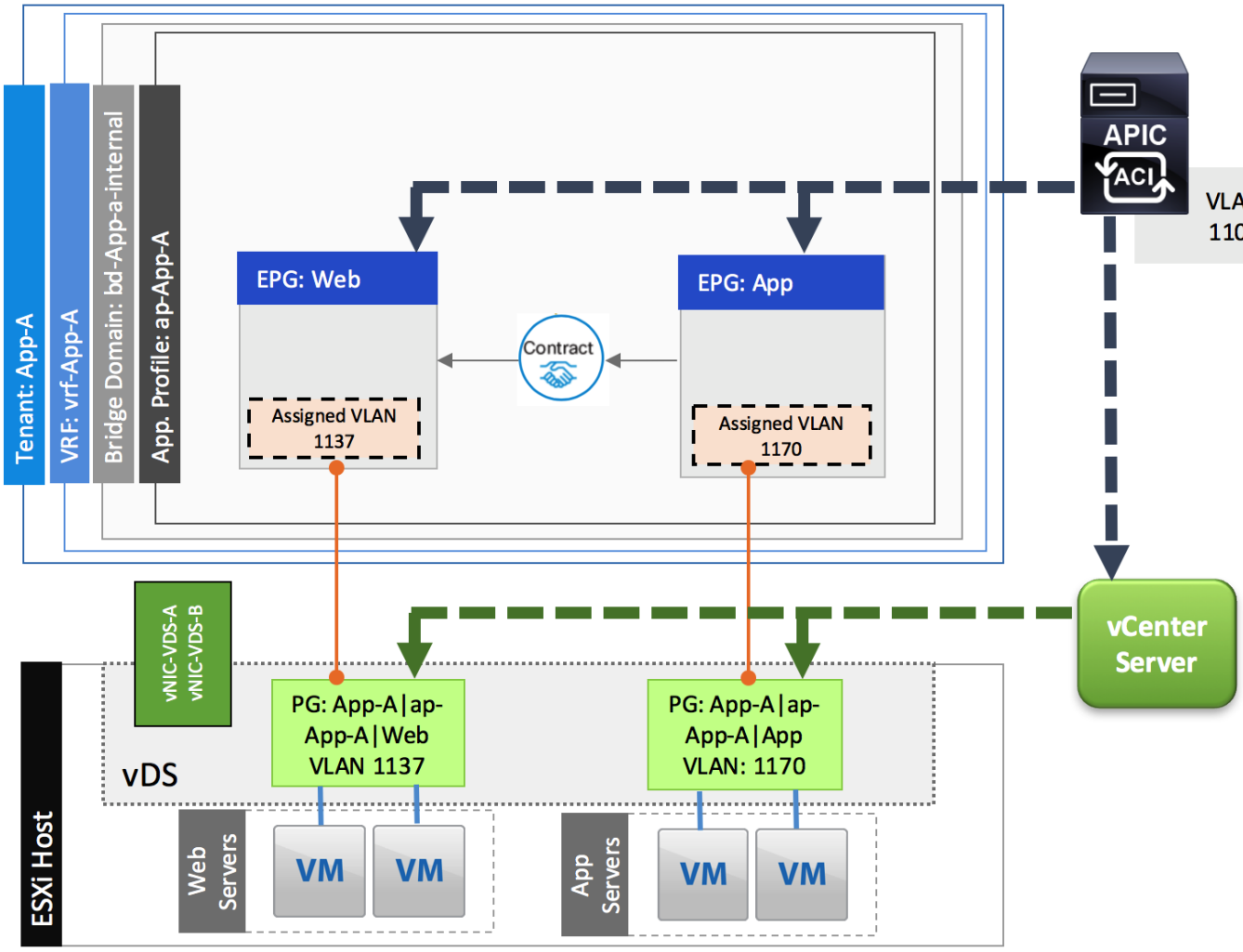
The VLAN or VxLAN allocation differs slightly depending on the virtual distributed switch in use. The following two sub-sections cover the deployment details for VMware vDS and Cisco AVS.

Port Group creation for VMware vDS

When application EPGs are attached to a VMware vDS based VMM domain, Cisco APIC assigns VLANs from a pre-defined pool and uses its connection to the VMware vCenter to create a new port groups on the VMware vDS. These port groups are used to deploy application VMs in the appropriate application tier. The port group name is determined using following format: "Tenant_Name | Application Profile_Name | EPG_Name".

For example, as shown in Figure 46 below, when a new EPG *Web* is defined under application profile *ap-App-A* (that belongs to tenant *App-A*), VLAN 1137 gets assigned to this EPG and a new port group named *App-A|ap-App-A|Web* is dynamically created on the VMware vDS. When a virtualization administrator assigns a VM NIC to this port group, all the network policies including security (contracts), L4-L7 and QoS policies automatically get applied to the VM communication.

Figure 46 ACI – Attaching an Application EPG with VMware vDS

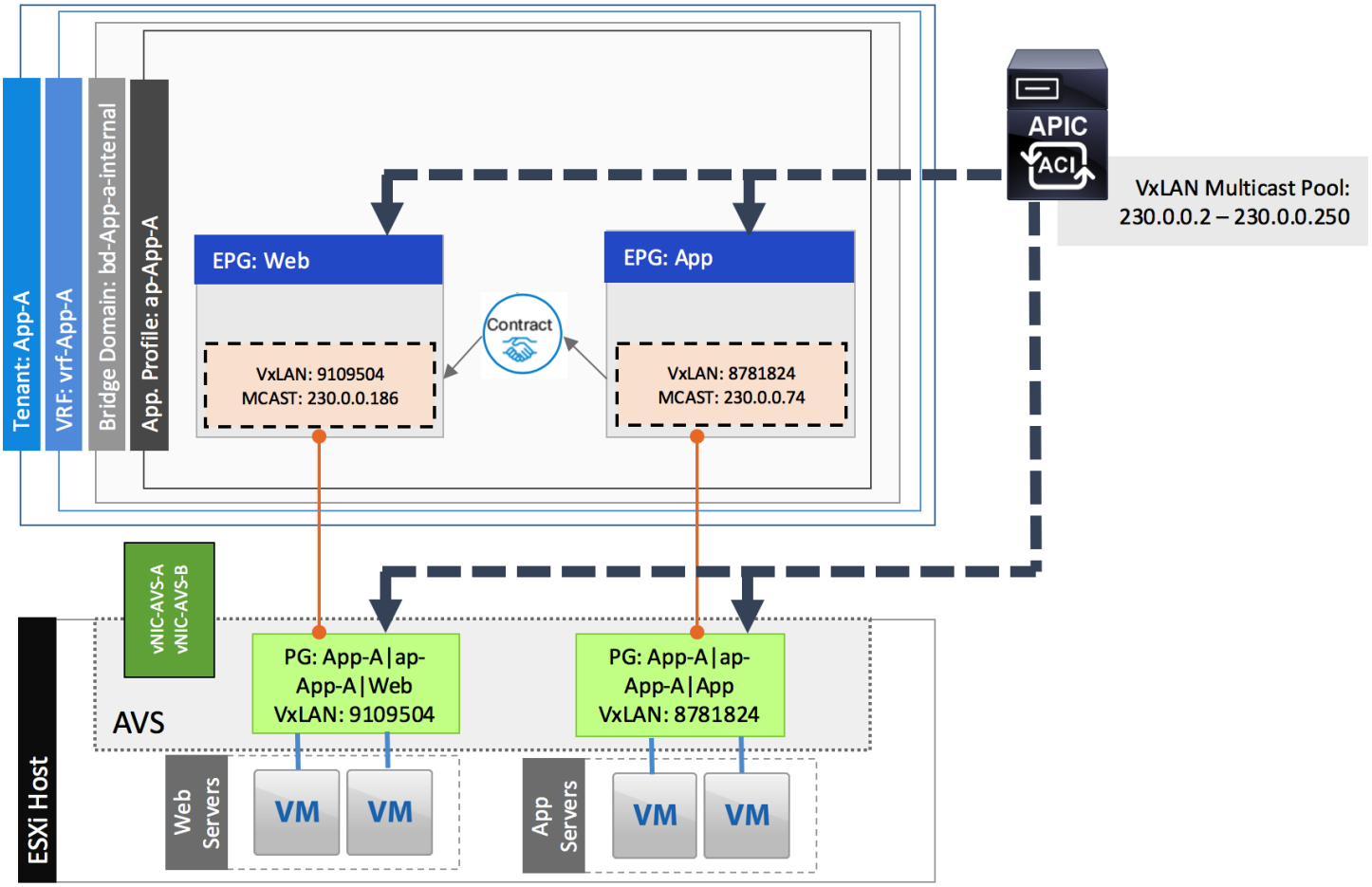


Port Group creation for Cisco AVS

When application EPGs are attached to a Cisco AVS (VxLAN mode) based VMM domain, new VxLAN segments are used for creating port groups on the Cisco AVS. These port groups are then used to deploy application VMs in the appropriate application tiers.

For example, as shown in Figure 47 below, when a new EPG *Web* is defined under application profile *ap-App-A* that belongs to tenant *App-A*, VxLAN 9109504 (number generated automatically) gets assigned to this EPG using the multicast address of 230.0.0.186 from the pre-defined pool. A new port-group named *App-A|ap-App-A|Web* is also configured on the Cisco AVS. When an application admin attached a VM NIC to this port group, all the network policies including security (contracts), L4-L7 and QoS policies get applied to the VM.

Figure 47 ACI – Attaching an Application EPG with Cisco AVS



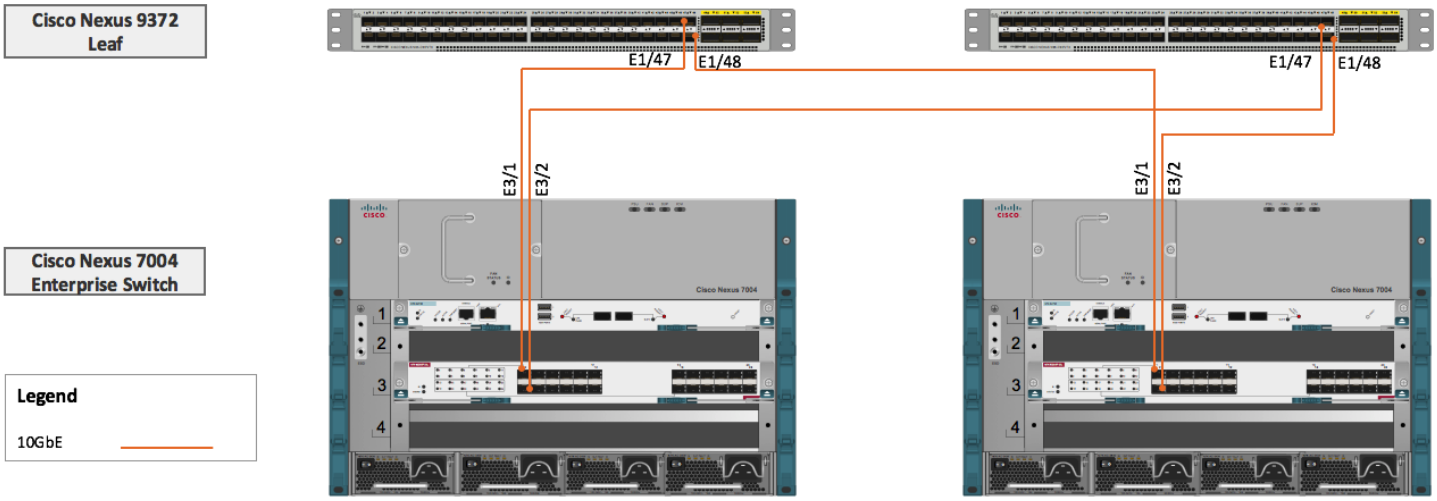
To provide layer-3 connectivity to the existing enterprise infrastructure, the application EPGs can "consume" contracts from the *common* tenant EPGs. Some examples of these contracts are covered in the upcoming sections.

External Network Connectivity - Shared Layer 3 Out

In order to connect ACI fabric to existing infrastructure, the ACI leaf nodes are connected to the existing enterprise core routers/switches. In this design, a Cisco Nexus 7000 was configured as the enterprise core router. Figure 48 shows the physical connectivity details. As shown in the figure, each of the leaf switches is physically connected to each of the core router for redundancy using a 10GbE connection.

In this design guide, a single pair of Cisco Nexus 9000 based leaf switches provides all the VersaStack connectivity including the layer 3 connectivity to existing infrastructure (border leaf functionality). In large scale configurations, customers are encouraged to deploy a dedicated pair of border leaf switches.

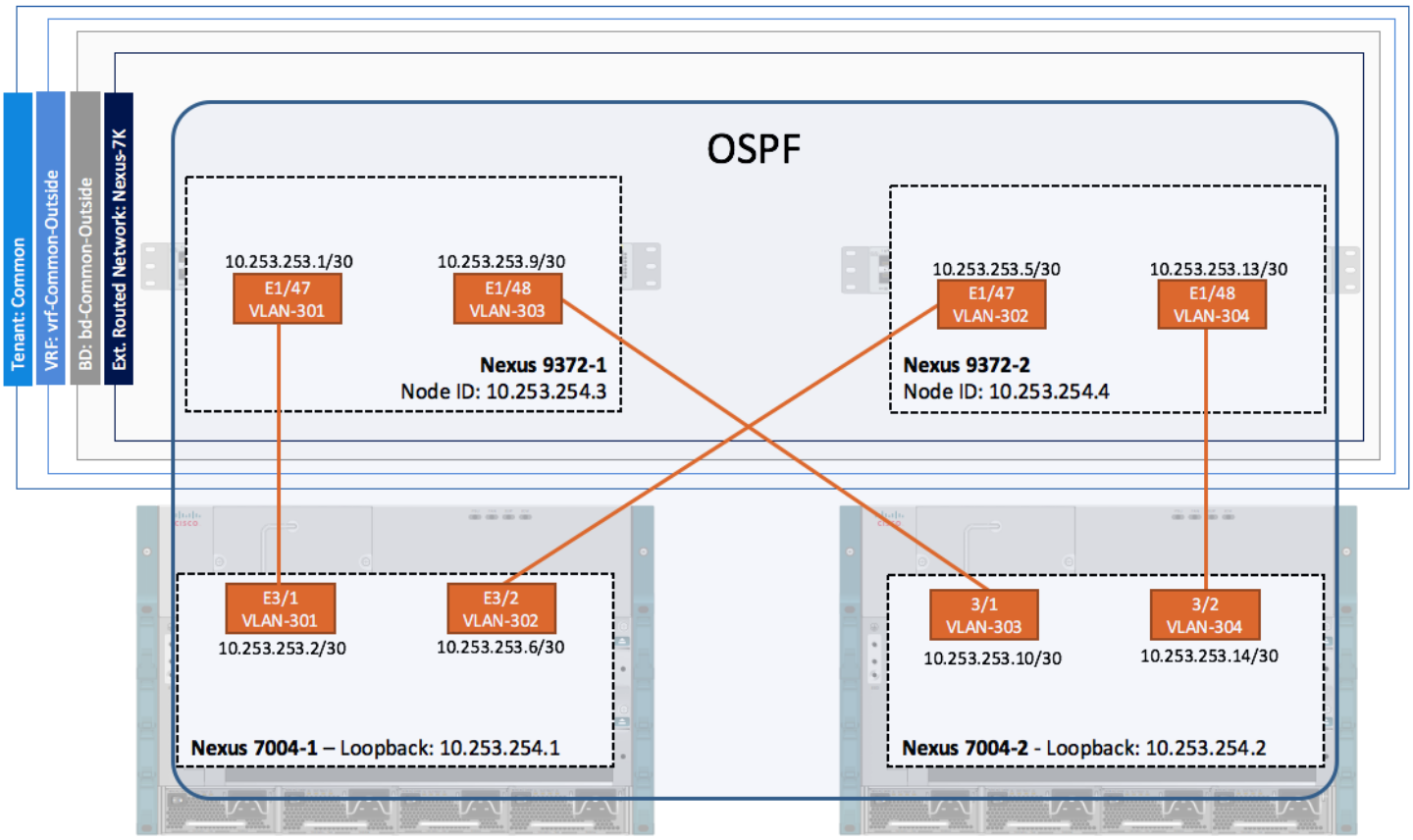
Figure 48 ACI – Physical Connectivity to Existing Infrastructure



A shared Layer 3 Out configuration, a new feature in ACI 1.2 and later, provides routed connectivity to external networks as a shared service. Shared Layer 3 Out functionality can be provisioned as a shared service in any tenant. In the VersaStack with ACI validated design, this functionality in configured in the *common* tenant. As shown in Figure 49, a single External Routed Network is configured under tenant *common* to connect ACI infrastructure to Cisco Nexus 7000s using OSPF. Some of the ACI constructs used in this design are:

- A unique private network and a dedicated external facing bridge domain is defined under the *common* tenant. This private network (VRF) is setup with OSPF to provide connectivity to external infrastructure. As shown in Figure 49, the private network configured under the tenant is called *vrf-Common-Outside* is and the bridge domain is called *bd-Common-Outside*.
- Four unique VLANs (sub-interfaces) are configured between ACI leaf switches and the core router; one for each of the four physical paths. The VLANs utilized are 301-304 as seen in Figure 49.
- OSPF routing is enabled on all the four paths between the Cisco Nexus 9000 and the Cisco Nexus 7000 core router
- On ACI fabric, *common* tenant learns a default route from the core router and advertises a "public" routable subnets to the core infrastructure.
- Core routers can optionally use OSPF metrics to configure path preferences.

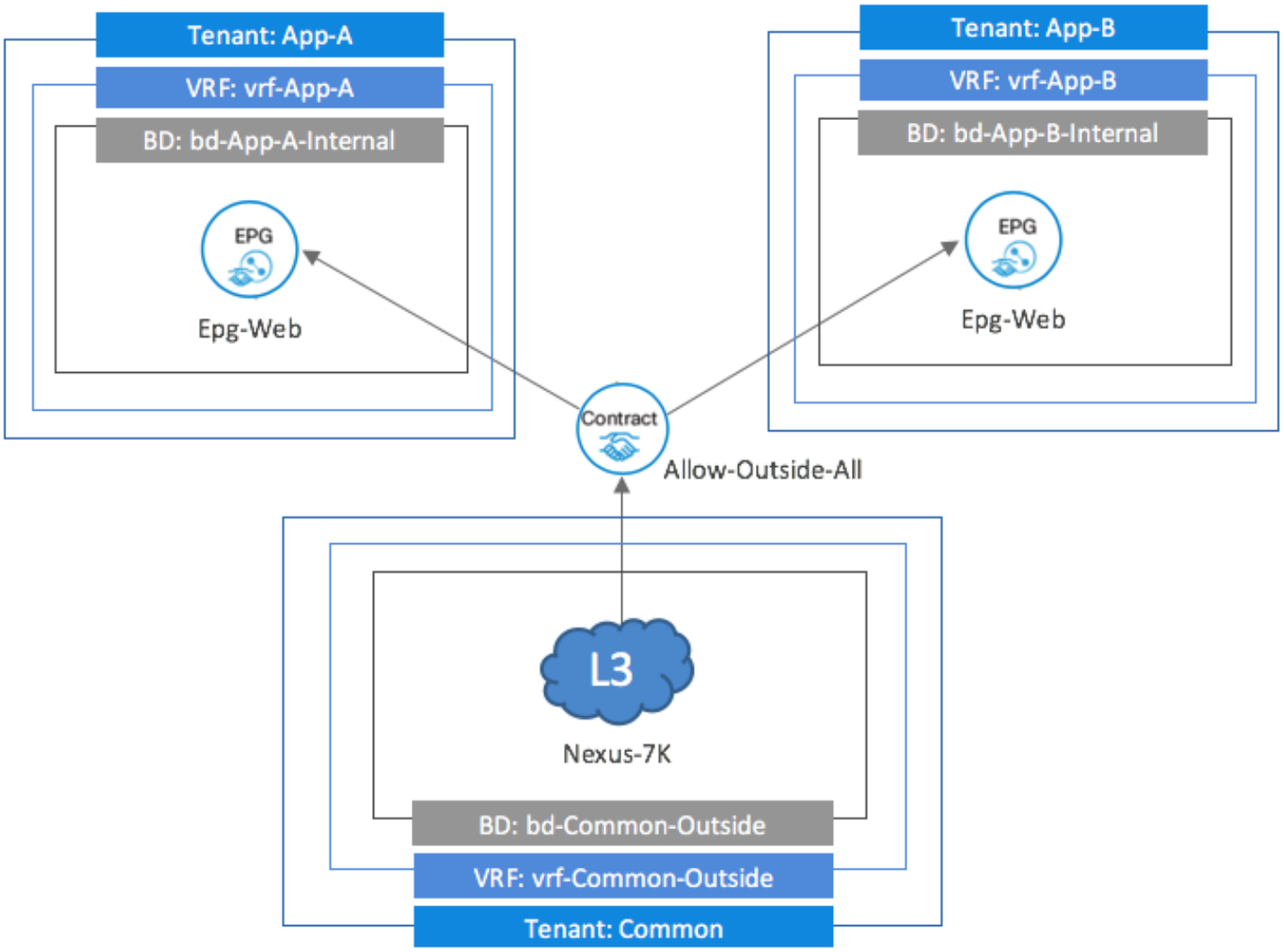
Figure 49 ACI - Logical Connectivity to Existing Infrastructure



When *common* tenants is configured with the Layer-3 connectivity, the remaining tenants can share this connection through contracts to access existing enterprise infrastructure as shown in Figure 50. The external routed network, *Nexus-7K*, "provides" a contract called *Allow-Outside-All*. When application tenant EPGs "consume" this contract, the "public" IP subnet defined under the application tenant EPGs get advertised to the core routers. The application EPGs also learns the default route from the tenant *common*. The filters under the contract control the traffic that can be sent and received from the shared L3 out. In this design, each tenant is configured with its own VRF as well as the bridge domain.

Tenant advertised prefixes for a shared Layer 3 out must to be unique; overlapping "public" subnets are not supported

Figure 50 ACI – Tenant Contracts for Shared L3 Out

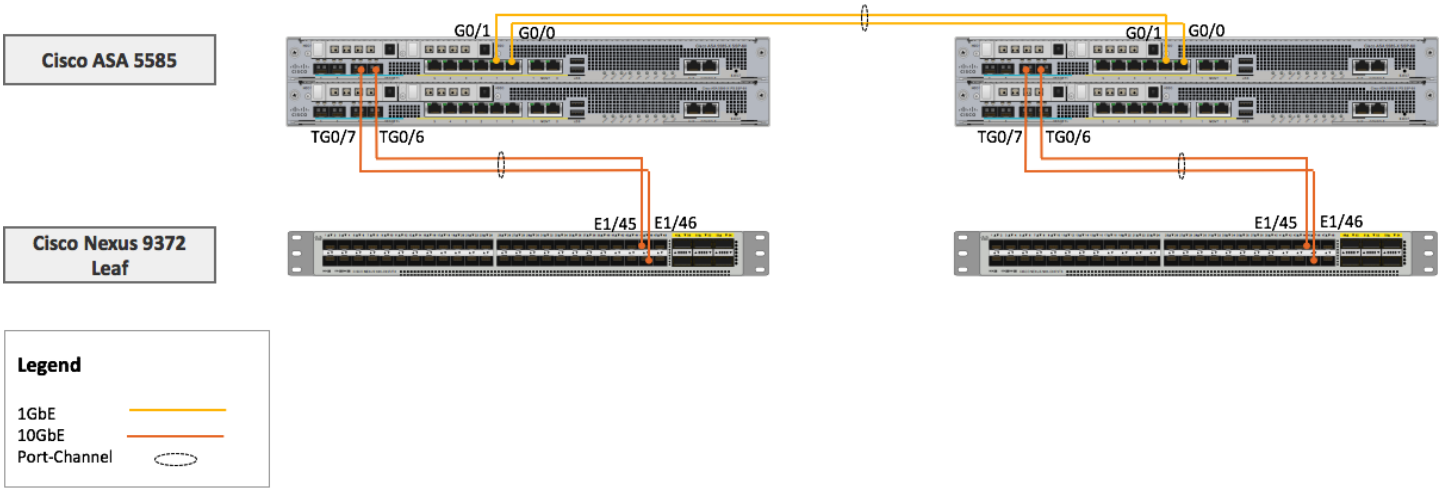


Integrating Firewall Services using Network-Only-Stitching (Unmanaged Firewall)

The network-only-stitching, or unmanaged, service graph is a relatively new option for configuring service insertion. With this option, Cisco ACI does not configure the L4-L7 device hence a device package is not needed. With network-only stitching, a L4-L7 device is configured as unmanaged on the APIC but the information about the connectivity including physical ports and VLANs, device cluster settings, and the device operation mode (routed vs transparent) is defined in APIC to setup the fabric correctly.

In this VersaStack with ACI design, a pair of redundant Cisco ASA5585 firewall devices are attached to Cisco Nexus 9372 leaf switches as shown in Figure 51. Each device is connected to single Cisco Nexus 9372 leaf using a 2 port 10GE port-channel. The two ASA devices are configured for high availability in case of link or device failure.

Figure 51 ACI – Physical Connectivity to Cisco ASA5585 Devices



The Shared Layer 3 connectivity section covered the layer 3 connection details such that the traffic to and from various tenants is controlled using contract filters. To provide additional security to various application tiers, an ASA firewall (context) can be easily integrated into each tenant by using the contracts as shown in Figure 52. A contract called *ASAFW* is provided from the application tenant EPG *Web* and consumed by the shared Layer 3 EPG called *Nexus-7K*. This contract is used to redirect the web traffic to an ASA firewall context before sending the traffic out to core infrastructure.

Figure 52 ACI – Adding L4-7 Services

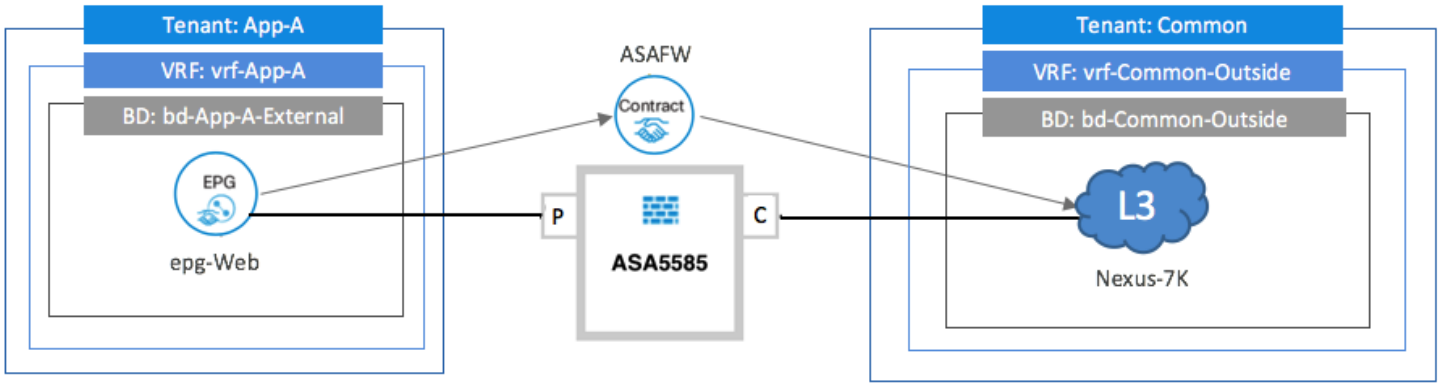
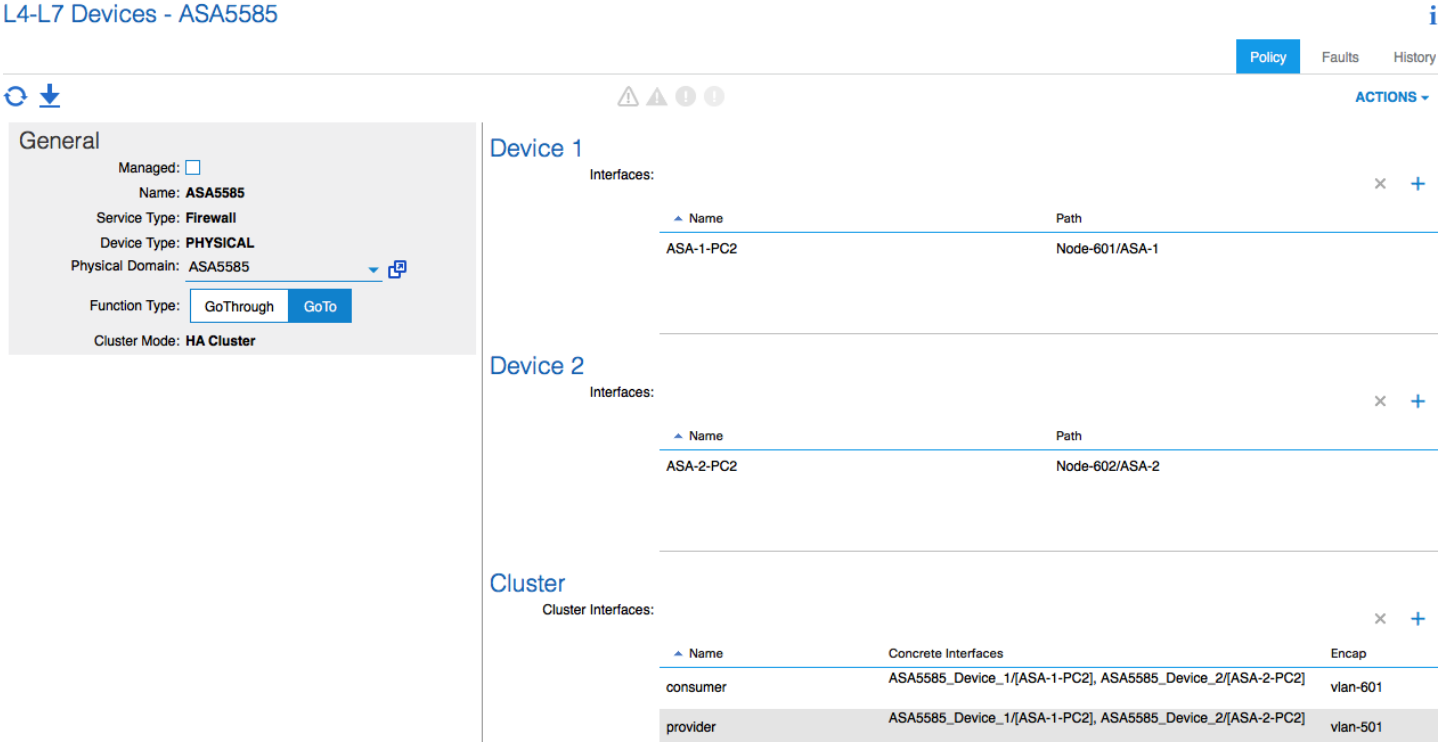


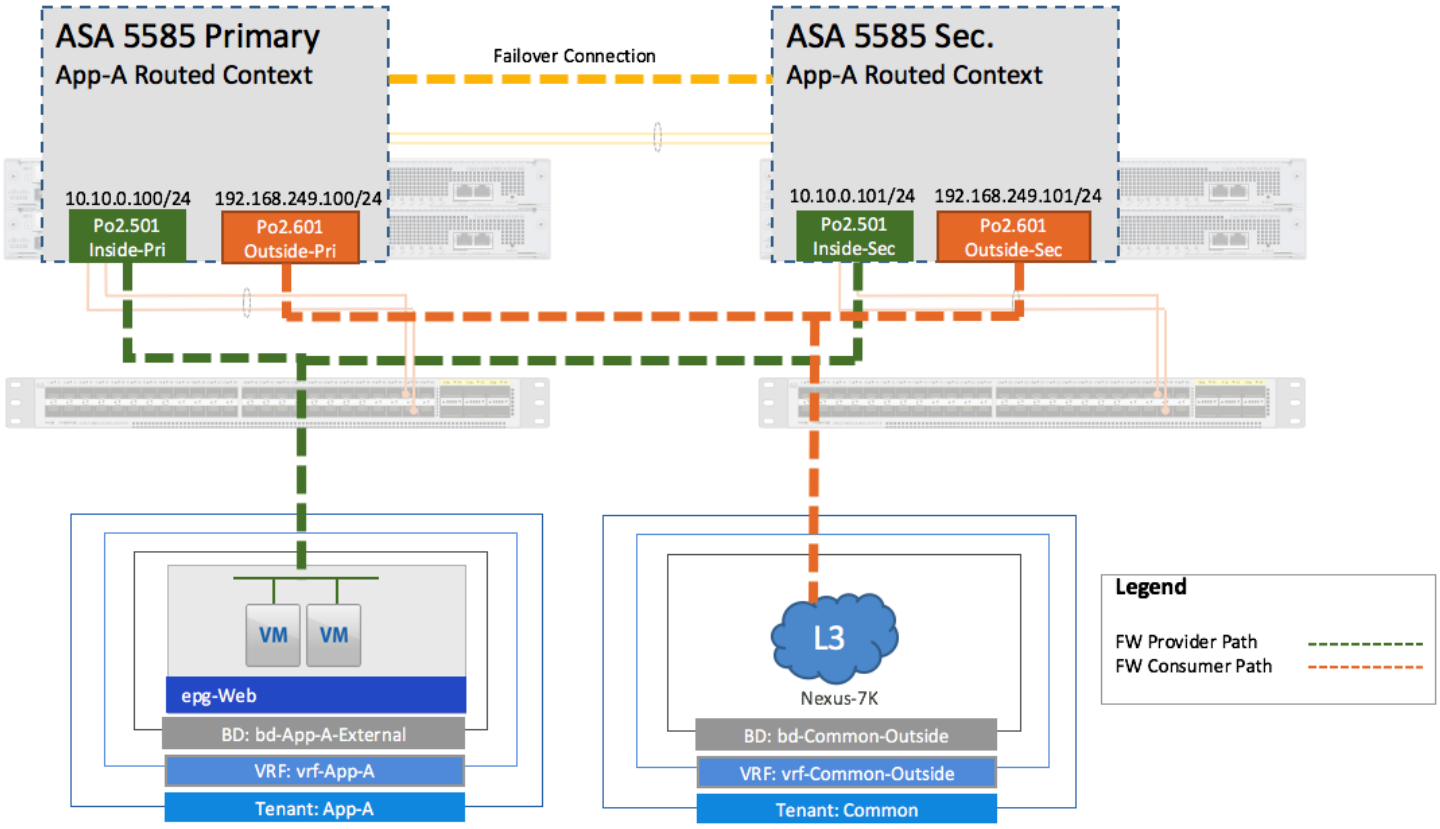
Figure 53 shows ASA device definition under the application tenant App-A. As seen from the figure, the two ASA devices are configured as Firewall GoTo (routed) devices and are deployed as an HA cluster. VLAN 501 is configured as the provider (protected) link to the firewall and VLAN 601 is configured as the consumer (unprotected) link.

Figure 53 ACI – L4-L7 Device Definition under Application Tenant



In addition to defining the L4-L7 devices, a service graph template is also configured and a service graph is deployed under the application tenant App-A. The contract provided by the EPG e*pg-Web* and consumed by Shared Layer 3 Out results in the traffic path as shown in Figure 54.

Figure 54 ACI – ASA FW Logical Configuration



In this configuration, the VMs deployed in the *epg-Web* have their gateway set to inside interface of the ASA FW context (10.10.0.100) and the ASA FW's default route is set to SVI defined in the common tenant bridge domain (192.168.249.254). The ASA is configured such that the VM traffic gets NAT-ed to outside interface IP address of the ASA FW (192.168.249.100) before being sent out to the core router using Shared L3 Out connection.

Design Considerations

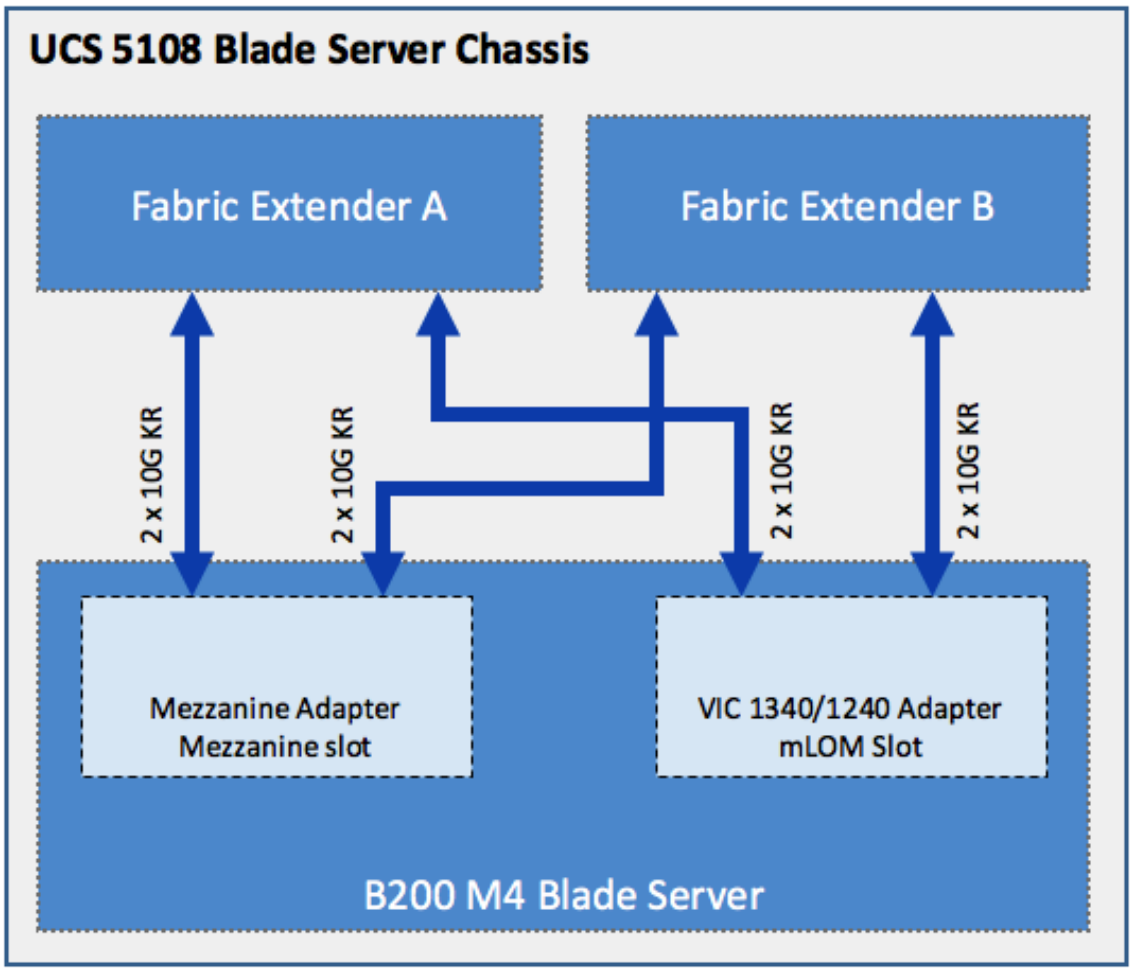
VersaStack designs incorporate connectivity and configuration best practices at every layer of the stack to provide a highly availability best performing system. VersaStack is a modular architecture that allows customers to adjust the individual components of the system to meet their particular scale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

Cisco Unified Computing System I/O Component Selection

Selection of I/O components has a direct impact on scale and performance characteristics when ordering the Cisco components. Figure 55 illustrates the available backplane connections in the Cisco UCS 5100 series chassis. As shown, each of the two Fabric Extenders (I/O module) has four 10GBASE KR (802.3ap) standardized Ethernet backplane paths available for connection to the half-width blade slot. This means that each half-width slot has the potential to support up to 80Gb of aggregate traffic depending on selection of the following:

- Cisco UCS Fabric Extender model (2204XP or 2208XP)
- Modular LAN on Motherboard (mLOM) card
- Mezzanine Slot card

Figure 55 Cisco UCS B-Series M4 Server Chassis Backplane Connections



Fabric Extender Modules (FEX)

Each Cisco UCS chassis is equipped with a pair of Cisco UCS Fabric Extenders. The fabric extenders have two different models, 2208XP and 2204XP. Cisco UCS 2208XP has eight 10 Gigabit Ethernet, FCoE-capable ports that connect the blade chassis to the fabric interconnect. The Cisco UCS 2204XP has four external ports with identical characteristics to connect to the fabric interconnect. Each Cisco UCS 2208XP has thirty-two 10 Gigabit Ethernet ports connected through the mid-plane to the eight half-width slots (four per slot) in the chassis, while the 2204XP has 16 such ports (two per slot).

Table 1 Number of Network and Host Facing Interface Fabric Extenders

	Network Facing Interface	Host Facing Interface
Cisco UCS 2204XP	4	16
Cisco UCS 2208XP	8	32

MLOM Virtual Interface Card (VIC)

The B200 M4 server accommodates different types of adapter cards. The mLOM slot only accommodates the Cisco VIC 1240 and Cisco VIC 1340 adapters, while the mezzanine slot accommodates all other adapters (such as port expander, VIC 1280/1380, and Cisco adapters). The VersaStack with ACI solution has been validated using Cisco VIC 1340. The Cisco VIC 1340 is a 2-port 40 Gigabit Ethernet FCoE-capable

mLOM adapter designed for both Cisco UCS B200 M3 and M4 generation of Cisco UCS B-Series Blade Servers. When used in combination with an optional Cisco VIC 1380, the Cisco UCS VIC 1340 capabilities can be expanded to support 80Gbpst Ethernet with the use of Cisco UCS 2208XP fabric extender.

Mezzanine Slot Card

A Cisco VIC 1380 is a two-port 40 Gigabit Ethernet, FCoE-capable mezzanine adapter designed exclusively for Cisco UCS B200 M3 and m4 Blade Servers.

Validated I/O Component Configurations

The validated I/O component configurations for this VersaStack with ACI design are:

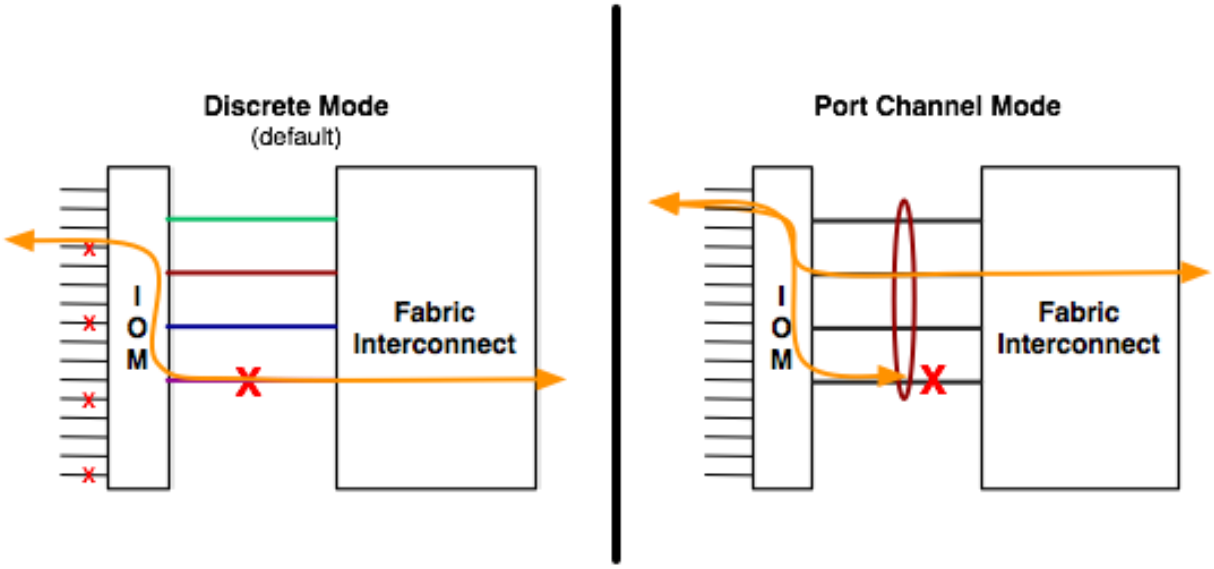
- Cisco UCS B200M3 with VIC 1240 and FEX 2208XP
- Cisco UCS B200M4 with VIC 1340 and FEX 2208XP

Cisco Unified Computing System Chassis/FEX Discovery Policy

The Cisco UCS chassis/FEX discovery policy determines how the system reacts when you add a new chassis or FEX. Cisco UCS Manager uses the settings in the chassis/FEX discovery policy to determine the minimum threshold for the number of links between the chassis or FEX and the fabric interconnect and whether to group links from the IOM to the fabric interconnect in a fabric port channel. Cisco UCS Manager cannot discover any chassis that is wired for fewer links than are configured in the chassis/FEX discovery policy.

Cisco Unified Computing System can be configured to discover a chassis using Discrete Mode or the Port-Channel mode (Figure 56). In Discrete Mode, each FEX KR connection and therefore server connection is tied or pinned to a network fabric connection homed to a port on the Fabric Interconnect. In the presence of a failure on the external "link" all KR connections are disabled within the FEX I/O module. In Port-Channel mode, the failure of a network fabric link allows for redistribution of flows across the remaining port channel members. Port-Channel mode therefore is less disruptive to the fabric and hence recommended in the VersaStack designs.

Figure 56 Cisco UCS Chassis Discovery Policy - Discrete Mode vs. Port Channel Mode



Storage Design and Scalability

IBM FlashSystem V9000 consists of 'building blocks', and optional additional storage expansion enclosures, all interconnected on a 16G FC interconnect provided by the Cisco MDS 9148S fabrics. Multiple 16G links provide high availability for inter-cluster communications, as well as parallelism when accessing block data. The building block consists of two FlashSystem V9000 Control Enclosures, and one FlashSystem V9000 Storage Enclosure. Within the building block, the controllers act as active-active data path to the volumes. Each Storage Enclosure provides RAID5 protection for its capacity. The current VersaStack design leveraged single scalable building block with potential to scale in two dimensions: performance, and capacity. Each building block provides linear scale of performance, that is, N building blocks are capable of providing N times the performance of a single building block. Each building block also carries the choice of its capacity, ranging from as little as 4TB to as much as 57TB per flash enclosure. Once the performance needs are satisfied, up to 4 additional Storage Enclosure can be added, each again ranging in 4-57TB.

The IBM Storwize V7000 Unified clustered systems provide scale-out growth in performance and capacity with up to four control enclosures and up to 20 expansion enclosures with 12 Gbps SAS connectivity. The scale-out design operates as a single storage system with 64 processor cores, up to 512 GB of cache, supporting up to 1,056 (SFF) drives and 7.87 PB of total capacity. V7000 Unified systems also include dual redundant File Modules with 1 Gbps and 10 Gbps interfaces for network-attached storage (NAS) capability. Leveraging the Cisco MDS switches, additional building blocks, storage enclosures and control enclosures can be added to the IBM FlashSystem V9000 and IBM V7000 Unified non-disruptively allowing your VersaStack solution to grow or take on new/additional workload.

Management Network Design

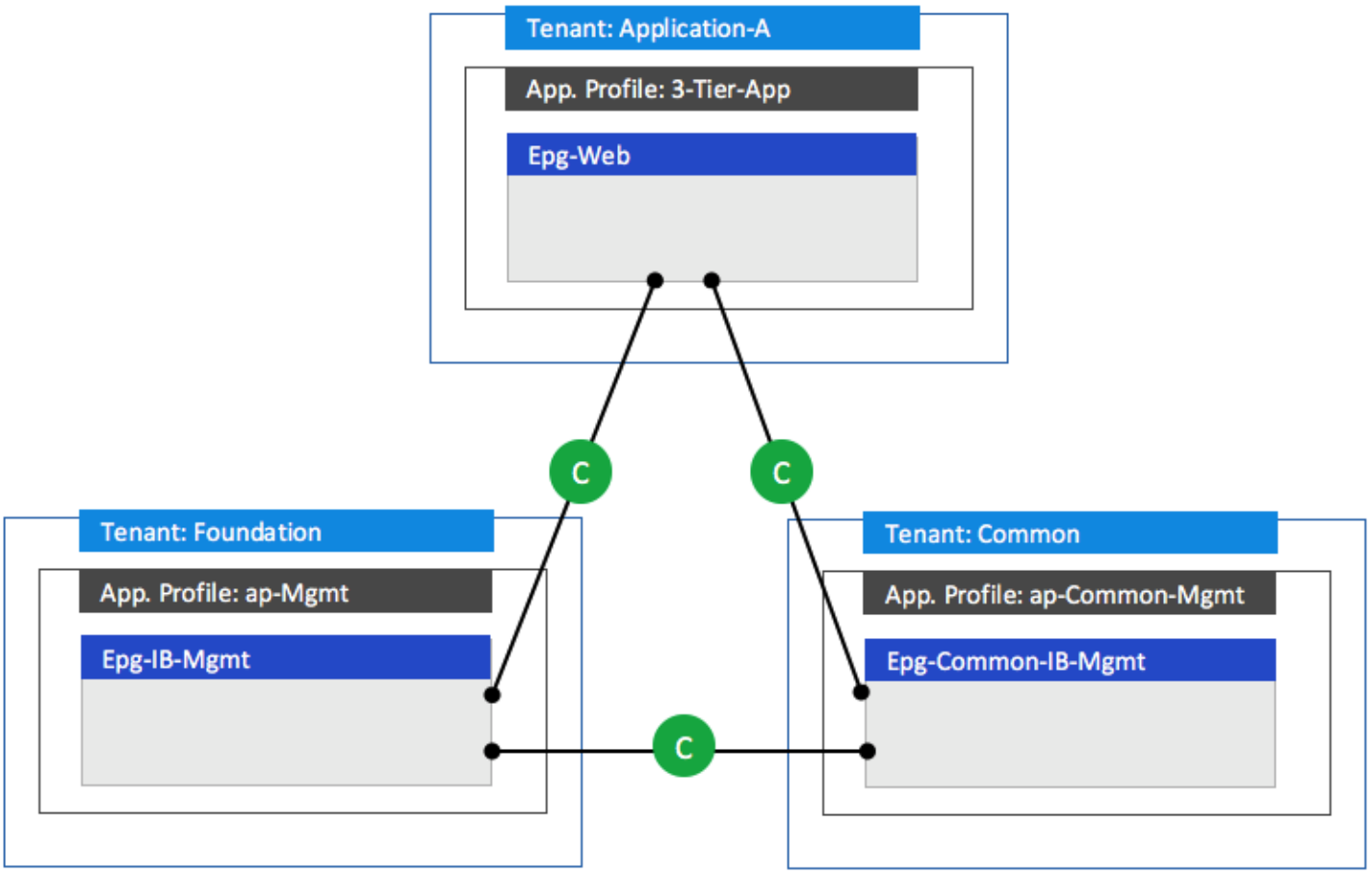
This VersaStack with ACI design uses two different networks to manage the solution:

- 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 configure ESXi servers and VMs (including infrastructure VMs hosted on VersaStack). If out-of-band management of these components are required, the disjoint layer 2 feature available in Cisco UCS can be used. This however would require additional uplink port(s) on the Cisco UCS FIs 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 withmore than one uplink.

In certain datacenters designs, there is a single management segment which requires combining both out-of-band and in-band networks into one. The current VersaStack with ACI design assumes infrastructure VMs such as vCenter, AD, DNS etc. are all running within VersaStack environment. The design also supports a separate management pod connected to the ACI fabric to host the infrastructure VMs but the design details are not covered in this document.

As discussed in "Enabling Management Access through Common Tenant", the contract between EPG *epg-IB-Mgmt* in the *Foundation* tenant and the *epg-Common-IB-Mgmt* in *common* tenant enables infrastructure VMs and ESXi hosts to talk to pre-existing management network. For VMs hosted in a application tenants requiring access to infrastructure VMs or ESXi servers within the VersaStack environment as well as management hosts outside VersaStack environment, separate contracts must be defined to both *epg-IB-Mgmt* in the *Foundation* tenant and *epg-Common-IB-Mgmt* in the *common* tenant as shown in Figure 57.

Figure 57 Providing Management Access to Application VMs



Virtual Port Channel Configuration

Virtual Port Channel (vPC) allows Ethernet links from that are physically connected to two different Cisco Nexus 9000 Series devices 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. Unlike an NxOS based design, a vPC configuration in ACI does not require a vPC peer-link to be explicitly connected and configured between the peer-devices (leaf switches). The peer communication is carried over the 40G connections through the spines. In the VersaStack with ACI design, when possible, vPC is a preferred mode of Port Channel configuration.

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, etc.) that can benefit from a larger frame size. Cisco UCS and Cisco ACI policies enable and deliver the jumbo frame functionality. 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. In this validation effort, the VersaStack was configured to support jumbo frames with an MTU size of 9000.

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.

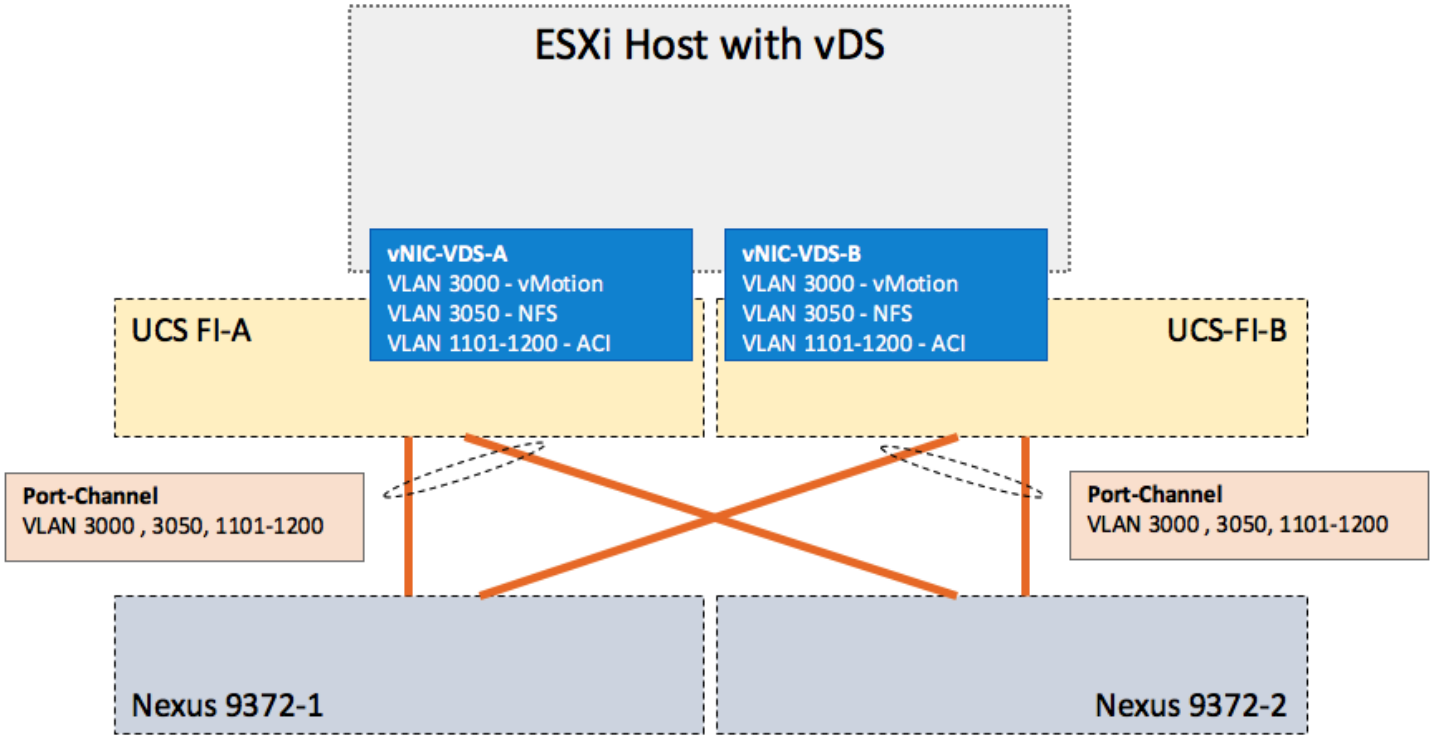
Distributed Switch – VLAN vs VxLAN encapsulation

Cisco ACI supports both VMware vDS and Cisco AVS and the VersaStack with ACI design was validated using both the switches. Both VMware vDS and Cisco AVS support VLAN encapsulation mode. In addition, Cisco AVS also supports VxLAN encapsulation mode. In this design guide, VLAN encapsulation mode for vDS and VxLAN encapsulation mode for Cisco AVS is used.

When choosing VLAN encapsulation for vDS (or AVS), a range of VLANs must be available for use by APIC. A pool of 100 VLANs is configured and assigned to VMM domain supporting vDS. These VLANs also need to be configured on the Cisco UCS FIs and assigned to the appropriate vNIC interfaces as shown in Figure 58. While Cisco APIC enables a single VLANs for every new EPG, Cisco UCS FIs act as unmanaged switches in the middle and are therefore not configured by APIC*.

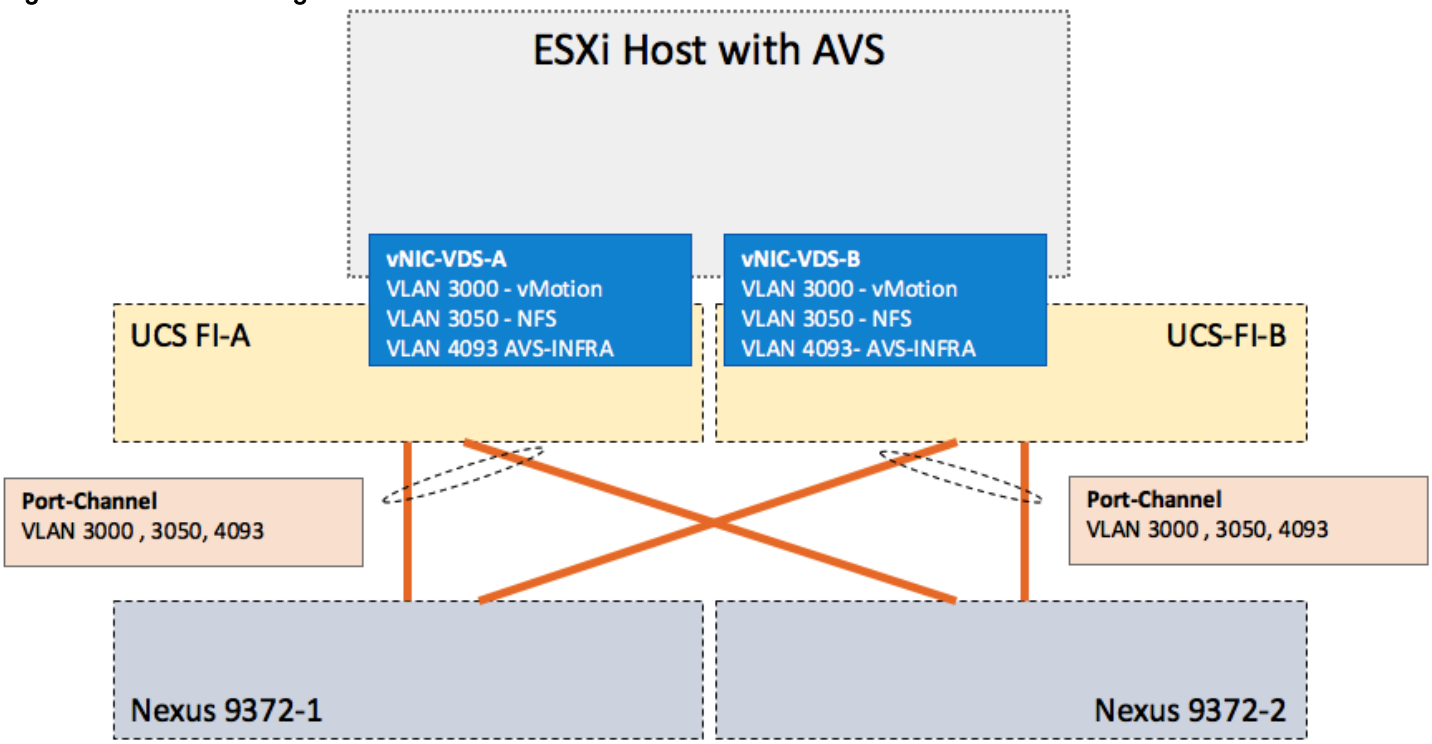
* Cisco UCS Director based ACI deployment is capable of configuring Cisco UCS FIs dynamically and pre-defining the VLAN pool is therefore not needed on the Cisco UCS

Figure 58 VLAN Configuration for VMware vDS



When choosing VXLAN encapsulation for AVS, only the infra-VLAN needs to be available between the Cisco AVS and the leaf as shown in Figure 59. This results in a simplified configuration and therefore with Cisco AVS and Cisco UCS FIs, VxLAN is the recommended encapsulation type.

Figure 59 VLAN Configuration for Cisco AVS



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) along a path fails, all storage traffic will reroute to an alternate path. When both paths are active, traffic is load balanced.

Deployment Hardware and Software

Hardware and Software Revisions

Table 2 below outlines 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 refer to the following links for more information:

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

Table 2 Hardware and Software Revisions Validated

Layer	Device	Image	Comments
Compute	Cisco UCS Fabric Interconnects 6200 Series, Cisco UCS B-200 M4, Cisco UCS C-220 M4	3.1(1g)	Includes the Cisco UCS-IOM 2208XP, Cisco UCS Manager, and Cisco UCS VIC 1340
	Cisco ESXi eNIC	2.3.0.7	Ethernet driver for Cisco VIC
	Cisco ESXi fnic Driver	1.6.0.25	FCoE driver for Cisco VIC
Network	Cisco Nexus Switches	11.3(2f)	iNXOS
	Cisco APIC	1.3(2f)	ACI release
	Cisco MDS 9148S	6.2(13b)	FC switch firmware version
Storage	IBM FlashSystem V9000	7.6.0.4	Software version
	IBM Storwize V7000 Unified	7.6.0.4	Software version
	IBM Storwize V7000 Unified File Modules	1.6.1	Software version
Software	VMware vSphere ESXi	6.0 update1b	Software version
	VMware vCenter	6.0 update 1b	Software version
	Cisco AVS	5.2(1)SV3(1.25)	Software version

Validation

Test Plan

The VersaStack with ACI solution was validated for by deploying multiple VMs running IOMeter tool. The system was validated for resiliency by failing various aspects of the system under the load. The types of tests executed on the system are as follows:

Cisco UCS Validation

- Failure and recovery of links from Cisco UCS Chassis (IOM) to FI-A and FI-B, one at a time
- Failure and recovery of links from Cisco UCS C220 to FI-A and FI-B, one at a time
- Rebooting Cisco UCS FI, one at a time
- Removing the physical cables between FI and Cisco Nexus 9372 leaf switches to simulate path failure scenarios

Network and ACI Validation

- Fail/power off both Cisco 9336 spine switches, one after other
- Fail/power off both Cisco 9372 leaf switches, one after other
- Fail/Isolate single APIC controller
- Fail/Isolate APIC controller cluster
- Remove multiple links between Cisco Nexus 9372 leaf switches and Cisco Nexus 7004 switches

Storage Validation

- Failure and recovery of the Cisco MDS switches, one at a time
- Failure and recovery of the links between Cisco MDS and IBM storage controllers
- Failure and recovery of links between Cisco UCS FI and Cisco MDS
- Failure and recovery of the links between Cisco Nexus 9372 leaf switches and IBM V7000 File Modules

vSphere Validation

- Failure and recovery of ESXi hosts in a cluster (rebooting of hosts, shutting down of hosts etc.)
- In case of a host failure, VM auto restart within the HA cluster
- VM vMotion across both B-Series and C-series servers
- Storage vMotion between iSCSI, NFS and FC datastores
- Restart/Isolate vCenter server

Bill of Materials

To find various components of VersaStack system:

- Go to the Main CCW page: https://apps.cisco.com/Commerce/home
- Under Find Products and Solutions, Click on the Search for solutions
- Type VersaStack. System will pull all the VersaStack variations
- Select one of the solutions and click View Components

Summary

VersaStack with ACI solution is designed to simplify the data center evolution to a shared cloud-ready infrastructure by using an application driven policy model. With the integration of Cisco ACI to the VersaStack platform, the solution delivers an application centric architecture with centralized automation that combines software flexibility with the hardware performance.

VersaStack Datacenter with Cisco ACI 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. VersaStack design discussed in this document has been validated for resiliency (under fair load) and fault tolerance during system upgrades, component failures, and partial as well as complete loss of power scenarios.

References

Products and Solutions

Cisco Unified Computing System:

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

Cisco UCS 6200 Series Fabric Interconnects:

http://www.cisco.com/en/US/products/ps11544/index.html

Cisco UCS 5100 Series Blade Server Chassis:

http://www.cisco.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 Nexus 9000 Series Switches:

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

Cisco Application Centric Infrastructure:

http://www.cisco.com/c/en/us/solutions/data-center-virtualization/application-centric-infrastructure/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 V9000:

http://www-03.ibm.com/systems/storage/flash/v9000/index.html

IBM Storwize V7000:

http://www-03.ibm.com/systems/storage/disk/storwize_v7000/

Interoperability Matrixes

Cisco UCS Hardware Compatibility Matrix:

http://www.cisco.com/c/en/us/support/servers-unified-computing/unified-computing-system/products-technical-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

IBM System Storage Interoperation Center:

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

About the Authors

Haseeb Niazi, Technical Marketing Engineer, Computing Systems Product Group, Cisco Systems, Inc.

Haseeb Niazi has over 17 years of experience at Cisco in the Data Center, Enterprise and Service Provider solutions and technologies. As a member of various solution teams and Advanced Services, Haseeb has helped many enterprise and service provider customers evaluate and deploy a wide range of Cisco solutions. As a technical marking engineer at Cisco UCS solutions group, Haseeb currently focuses on network, compute, virtualization, storage and orchestration aspects of various Compute Stacks. Haseeb holds a master's degree in Computer Engineering from the University of Southern California and is a Cisco Certified Internetwork Expert (CCIE 7848).

Adam H. Reid, Test Specialist, Systems & Technology Group, IBM

Adam H. Reid is a published author with more than 15 years of Computer Engineering experience. Focused more recently on IBM's Spectrum Virtualize, he's been deeply involved with the testing and configuration of virtualized environments pivotal to the future of software defined storage. Adam has designed, tested and validated systems to meet the demands of a wide range of mid-range and enterprise environments.

Acknowledgements

Authors would like to thank Sreenivasa Edula, Technical Marketing Engineer, Cisco Systems, Inc. for his contributions in building this solution and his valuable input in developing this design document.