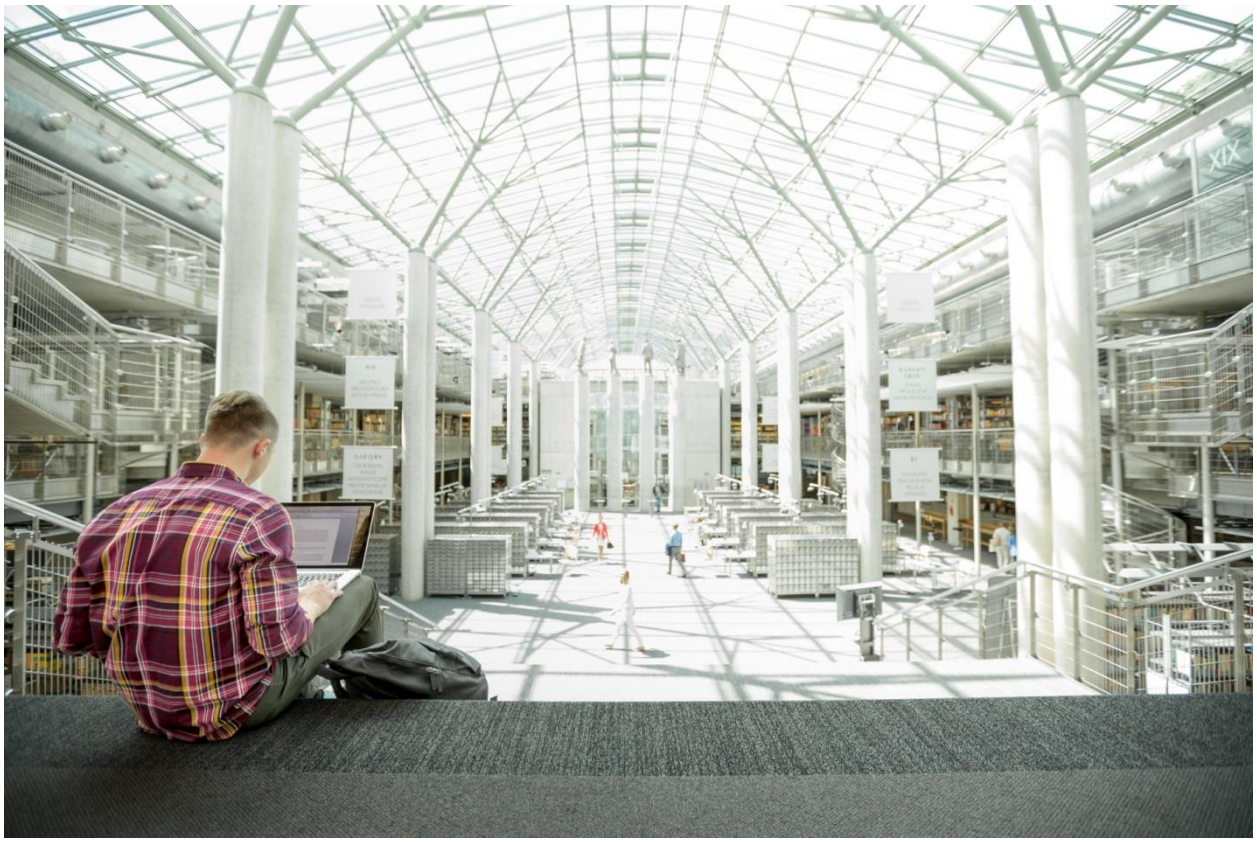


FlexPod Datacenter with Cisco ACI and VMware vSphere 6.0 U1 Design Guide

Last Updated: July 25, 2016

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 include 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 customers.

This document describes the Cisco and NetApp® FlexPod® Datacenter with NetApp All Flash FAS (AFF), Cisco Application Centric Infrastructure (ACI), and VMware vSphere 6.0 Update 1b. Cisco ACI is a holistic architecture that introduces hardware and software innovations built upon the new Cisco Nexus 9000® Series product line. Cisco ACI provides a centralized policy-driven application deployment architecture that is managed through the Cisco Application Policy Infrastructure Controller (APIC). Cisco ACI delivers software flexibility with the scalability of hardware performance.

FlexPod Datacenter with NetApp AFF and Cisco ACI is a predesigned, best-practice data center architecture built on the Cisco Unified Computing System (UCS), the Cisco Nexus® 9000 family of switches, and NetApp AFF. Key design details and best practices for this new architecture are covered in the following sections.

Solution Overview

Introduction

Industry trends indicate a vast data center transformation toward shared infrastructure and cloud computing. Business agility requires application agility, so IT teams need to provision applications in hours instead of months. Resources need to scale up (or down) in minutes, not hours.

To simplify the evolution to a shared cloud infrastructure based on an application driven policy model, Cisco and NetApp have developed this solution called VMware vSphere® on FlexPod with Cisco Application Centric Infrastructure (ACI). Cisco ACI in the data center is a holistic architecture with centralized automation and policy-driven application profiles that delivers software flexibility with hardware performance.

Audience

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

Changes in FlexPod

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

- Validation of the latest version of Cisco ACI with the latest version of the NetApp All Flash FAS storage array
- Validation of the Cisco ACI 1.3 Release on Cisco Nexus 9000 Series Switches
- Support for the Cisco UCS 3.1 release and Cisco UCS B200-M4 and C220-M4 servers with Intel E5-2600 v4 Series processors
- Support for the latest release of NetApp Data ONTAP® 8.3.2
- A storage design supporting both NAS datastores and iSCSI and Fibre Channel over Ethernet (FCoE) SAN LUNs

FlexPod Program Benefits

Cisco and NetApp have carefully validated and verified the FlexPod solution architecture and its many use cases and created a portfolio of detailed documentation, information, and references to assist customers in transforming their data centers to this shared infrastructure model. This portfolio includes, but is not limited to the following items:

- Best practice architectural design
- Workload sizing and scaling guidance
- Implementation and deployment instructions

- Technical specifications (rules for what constitutes a FlexPod configuration)
- Frequently asked questions (FAQs) and answers
- Cisco Validated Designs (CVDs) and NetApp Validated Architectures (NVAs) covering a variety of use cases

Cisco and NetApp have also built a robust and experienced support team focused on FlexPod solutions, from customer account and technical sales representatives to professional services and technical support engineers. The support alliance between NetApp and Cisco provides customers and channel services partners with direct access to technical experts who collaborate with cross vendors and have access to shared lab resources to resolve potential issues.

FlexPod supports tight integration with virtualized and cloud infrastructures, making it the logical choice for long-term investment. FlexPod also provides a uniform approach to IT architecture, offering a well-characterized and documented shared pool of resources for application workloads. FlexPod delivers operational efficiency and consistency with the versatility to meet a variety of SLAs and IT initiatives, including:

- Application rollouts or application migrations
- Business continuity and disaster recovery
- Desktop virtualization
- Cloud delivery models (public, private, hybrid) and service models (IaaS, PaaS, SaaS)
- Asset consolidation and virtualization

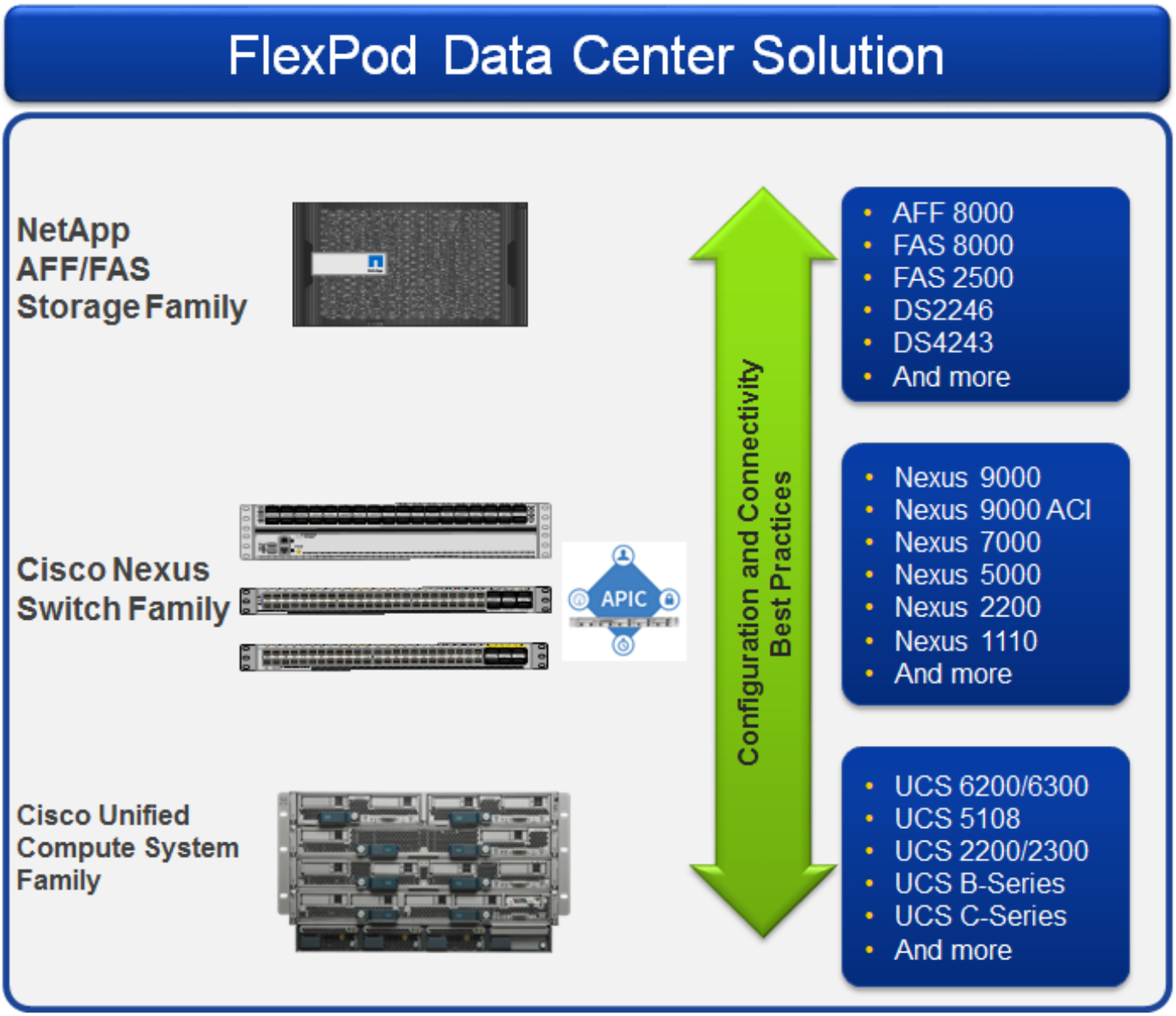
Technology Overview

FlexPod System Overview

FlexPod is a best practice datacenter architecture that includes three components:

- Cisco Unified Computing System (Cisco UCS)
- Cisco Nexus switches
- NetApp fabric-attached storage (FAS) systems

Figure 1 FlexPod Component Families



These components are connected and configured according to best practices of both Cisco and NetApp and provide the ideal platform for running a variety of enterprise workloads with confidence. FlexPod can scale

up for greater performance and capacity (adding compute, network, or storage resources individually as needed), or it can scale out for environments that require multiple consistent deployments (rolling out additional FlexPod stacks). The reference architecture covered in this document leverages the Cisco Nexus 9000 for the switching element.

One of the key benefits of FlexPod is the ability to maintain consistency at scale. Each of the component families shown (Cisco UCS, Cisco Nexus, and NetApp FAS) offers platform and resource options to scale the infrastructure up or down, while supporting the same features and functionality that are required under the configuration and connectivity best practices of FlexPod.

FlexPod Design Principles

FlexPod addresses four primary design principles: scalability, flexibility, availability, and manageability. These architecture goals are as follows:

- Application availability. Makes sure that services are accessible and ready to use.
- Scalability. Addresses increasing demands with appropriate resources.
- Flexibility. Provides new services or recovers resources without requiring infrastructure modification.
- Manageability. Facilitates efficient infrastructure operations through open standards and APIs.

Note: Performance is a key design criterion that is not directly addressed in this document. It has been addressed in other collateral, benchmarking, and solution testing efforts; this design guide validates the functionality.

FlexPod and Application Centric Infrastructure

The Cisco Nexus 9000 family of switches supports two modes of operation: NxOS standalone mode and Application Centric Infrastructure (ACI) fabric mode. In standalone mode, the switch performs as a typical Nexus switch with increased port density, low latency and 40G/100G connectivity. In fabric mode, the administrator can take advantage of Cisco ACI. Cisco Nexus 9000-based FlexPod design with Cisco ACI consists of Cisco Nexus 9500 and 9300 based spine/leaf switching architecture controlled using a cluster of three Application Policy Infrastructure Controllers (APICs).

Cisco ACI delivers a resilient fabric to satisfy today's dynamic applications. 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 industry trends of 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 architectures.

Cisco ACI Fabric

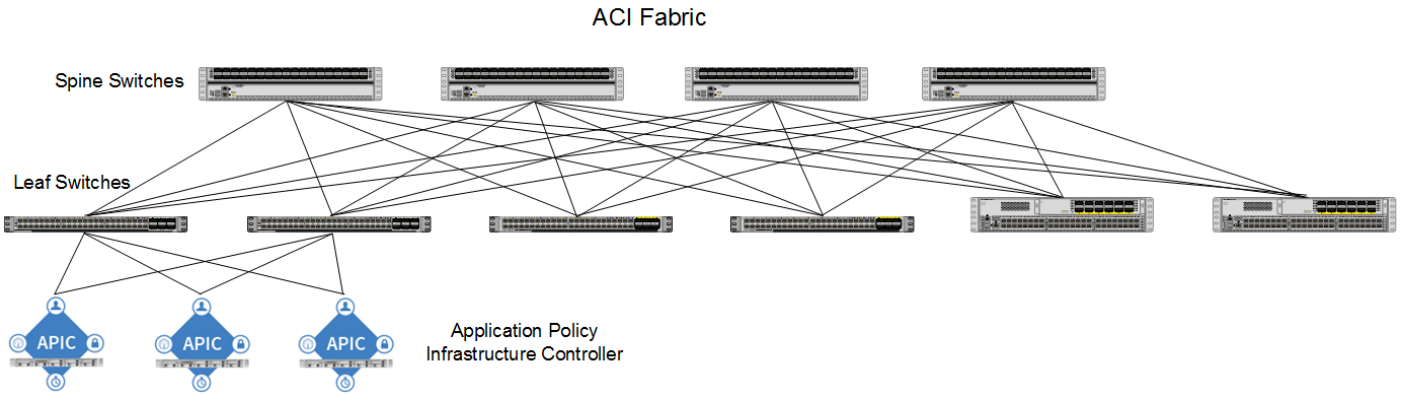
The Cisco ACI fabric consists of three major components:

- The Application Policy Infrastructure Controller (APIC)

- Spine switches
- Leaf switches

The ACI switching architecture is presented in a leaf-and-spine topology where every leaf connects to every spine using 40G Ethernet interface(s). The ACI Fabric Architecture is outlined in Figure 2.

Figure 2 Cisco ACI Fabric Architecture



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
- Network policy configuration and instantiation

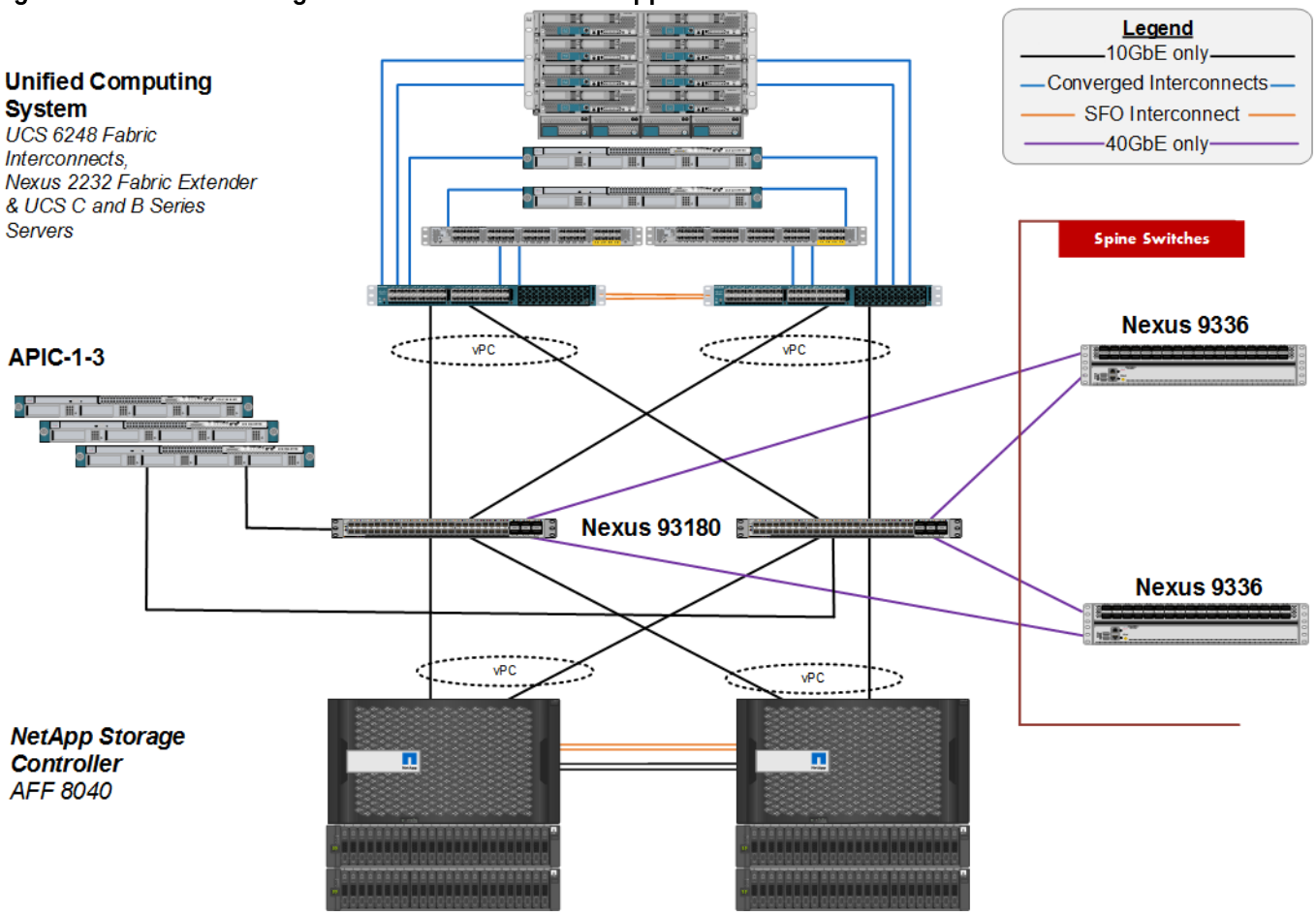
Though the APIC acts as the centralized point of configuration for policy and network connectivity, it is never in line with the data path or the forwarding topology. The fabric can still forward traffic even when communication with the APIC is lost.

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.

FlexPod with Cisco ACI–Components

FlexPod with ACI is designed to be fully redundant in the compute, network, and storage layers. There is no single point of failure from a device or traffic path perspective. Figure 3 shows how the various elements are connected together.

Figure 3 FlexPod Design with Cisco ACI and NetApp Clustered Data ONTAP

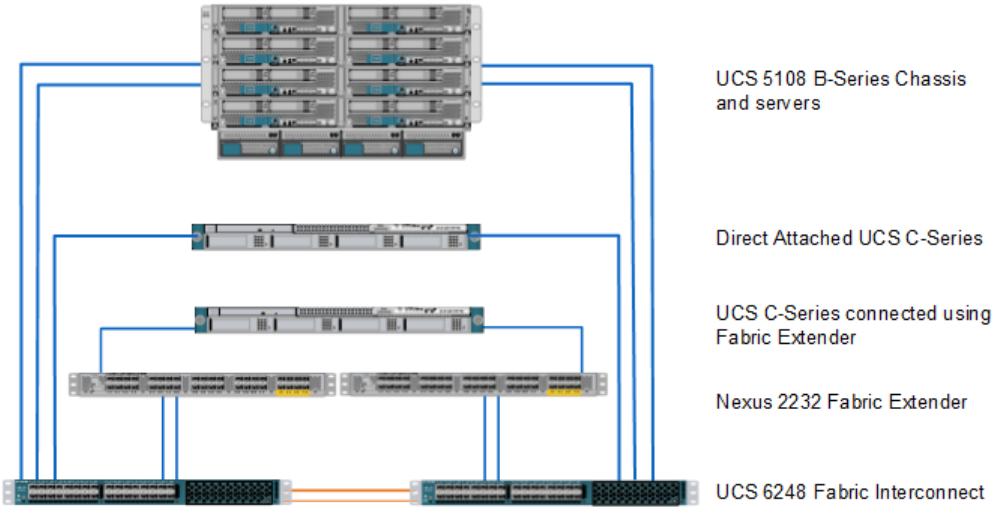


Fabric: As in the previous designs of FlexPod, link aggregation technologies play an important role in FlexPod with ACI providing improved aggregate bandwidth and link resiliency across the solution stack. The NetApp storage controllers, Cisco Unified Computing System, and Cisco Nexus 9000 platforms support active port channeling using 802.3ad standard Link Aggregation Control Protocol (LACP). Port channeling is a link aggregation technique offering link fault tolerance and traffic distribution (load balancing) for improved aggregate bandwidth across member ports. In addition, the Cisco Nexus 9000 series features virtual Port Channel (vPC) capabilities. vPC allows links that are physically connected to two different Cisco Nexus 9000 Series devices to appear as a single "logical" port channel to a third device, essentially offering device fault tolerance. Note in the Figure above that vPC peer links are no longer needed. The peer link is handled in the leaf to spine connections and any two leaves in an ACI fabric can be paired in a vPC. The Cisco UCS Fabric Interconnects and NetApp FAS controllers benefit from the Cisco Nexus vPC abstraction, gaining link and device resiliency as well as full utilization of a non-blocking Ethernet fabric.

Compute: Each Fabric Interconnect (FI) is connected to both the leaf switches and the links provide a robust 40GbE connection between the Cisco Unified Computing System and ACI fabric. Figure 4 illustrates the use of vPC enabled 10GbE uplinks between the Cisco Nexus 9000 leaf switches and Cisco UCS FI. Additional ports can be easily added to the design for increased bandwidth as needed. Each Cisco UCS 5108 chassis is connected to the FIs using a pair of ports from each IO Module for a combined 40G uplink. Current FlexPod design supports Cisco UCS C-Series connectivity both for direct attaching the Cisco UCS C-Series servers into the FIs or by connecting Cisco UCS C-Series to a Cisco Nexus 2232 Fabric Extender hanging off of the Cisco UCS FIs. The Fabric Extenders are used when using many UCS C-Series servers and the number of available ports on the Fabric Interconnects becomes a concern. FlexPod designs mandate Cisco

UCS C-Series management using Cisco UCS Manager to provide a uniform look and feel across blade and standalone servers.

Figure 4 Compute Connectivity



Storage: The ACI-based FlexPod design is an end-to-end IP-based storage solution that supports SAN access by using iSCSI. The solution provides a 10/40GbE fabric that is defined by Ethernet uplinks from the Cisco UCS Fabric Interconnects and NetApp storage devices connected to the Cisco Nexus switches. Optionally, the ACI-based FlexPod design can be configured for SAN boot or application LUN access by using Fibre Channel (FC) or Fibre Channel over Ethernet (FCoE). FC/FCoE access is provided by directly connecting the NetApp FAS controller to the Cisco UCS Fabric Interconnects with separate ports as shown in Figure 5. Whether the access is FC or FCoE is determined by which port and SFP type is used on the storage controller and Fabric Interconnect. Note that the UCS server access to the Fabric Internconnect is always FCoE, but the connections to the storage controllers can be either FC or FCoE. Also note that although FC and FCoE are supported, only FCoE connections to storage are validated in this CVD.

Figure 5 FC/FCoE Connectivity - Direct Attached SAN

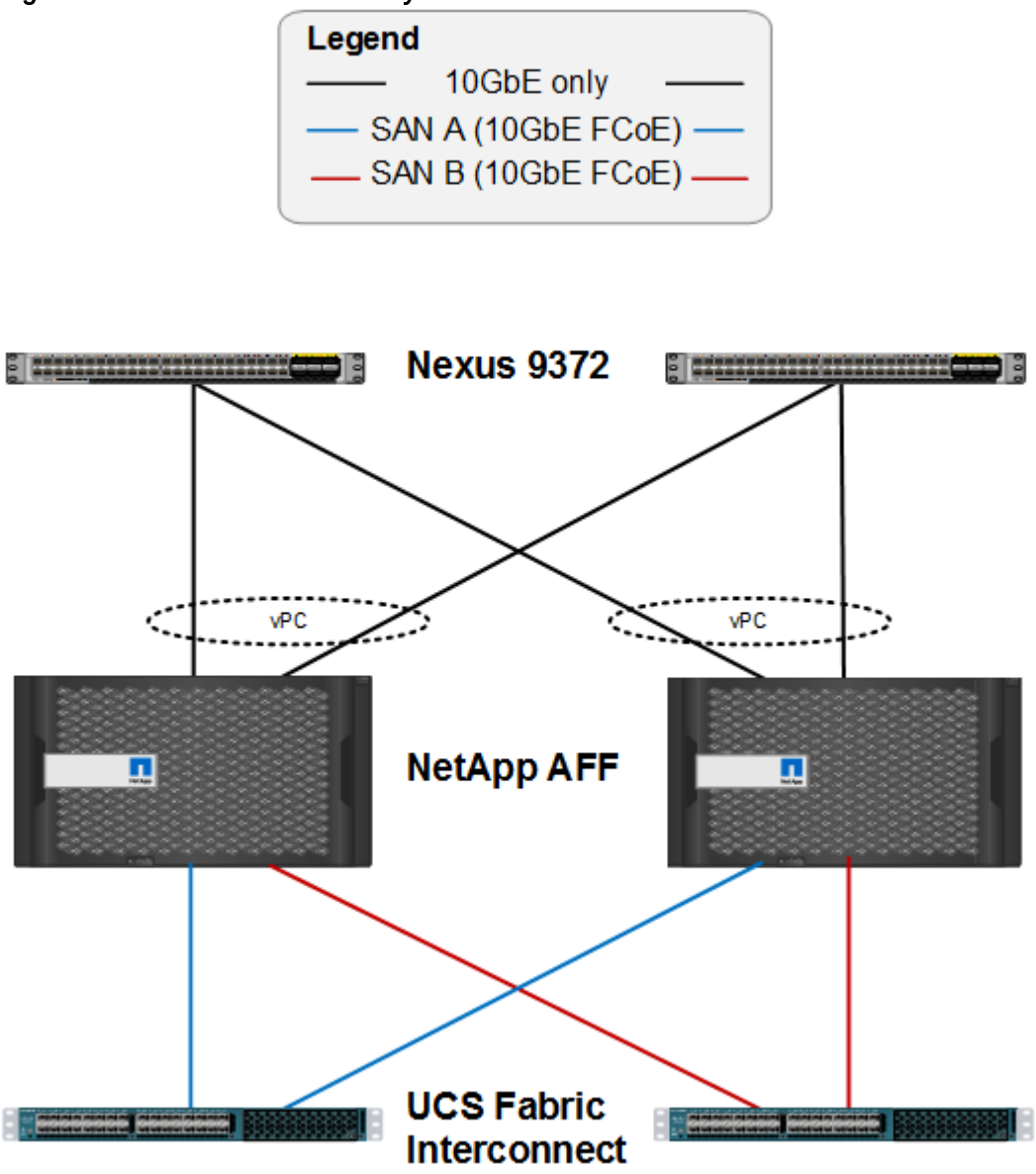


Figure 5 shows the initial storage configuration of this solution as a two-node high availability (HA) pair running clustered Data ONTAP in a switchless cluster configuration. Storage system scalability is easily achieved by adding storage capacity (disks and shelves) to an existing HA pair, or by adding more HA pairs to the cluster or storage domain.

Note: For SAN environments, NetApp clustered Data ONTAP allows up to 4 HA pairs or 8 nodes. For NAS-only environments, it allows 12 HA pairs or 24 nodes to form a logical entity.

The HA interconnect allows each node in an HA pair to assume control of its partner's storage (disks and shelves) directly. The local physical HA storage failover capability does not extend beyond the HA pair. Furthermore, a cluster of nodes does not have to include similar hardware. Rather, individual nodes in an HA pair are configured alike, allowing customers to scale as needed, as they bring additional HA pairs into the larger cluster.

For more information about the virtual design of the environment that consists of VMware vSphere, Cisco Application Virtual Switch (AVS) or VMware virtual distributed switching, and NetApp storage controllers, refer to the section FlexPod Infrastructure Physical Build.

Validated System Hardware Components

The following components were used to validate this Cisco Nexus 9000 ACI design:

- Cisco Unified Computing System
- Cisco Nexus 2232 Fabric Extender (optional)
- Cisco Nexus 9396 Series Leaf Switch
- Cisco Nexus 9372 Series Leaf Switch
- Cisco Nexus 93180YC-EX Series Leaf Switch
- Cisco Nexus 9336 Spine Switch
- Cisco Application Policy Infrastructure Controller (APIC)
- NetApp All-Flash FAS Unified Storage

Cisco Unified Computing System

The Cisco Unified Computing System is a next-generation solution for blade and rack server computing. 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 in which all resources participate in a unified management domain. The Cisco Unified Computing System accelerates the delivery of new services simply, reliably, and securely through end-to-end provisioning and migration support for both virtualized and non-virtualized systems.

The Cisco Unified Computing System consists of the following components:

- Cisco UCS Manager (http://www.cisco.com/en/US/products/ps10281/index.html) provides unified, embedded management of all software and hardware components in the Cisco Unified Computing System.
- Cisco UCS 6200 Series Fabric Interconnects (http://www.cisco.com/en/US/products/ps11544/index.html) is a family of line-rate, low-latency, lossless, 10-Gbps Ethernet, Fibre Channel over Ethernet and Fibre Channel interconnect switches providing the management and communication backbone for the Cisco Unified Computing System.
- Cisco UCS 5100 Series Blade Server Chassis (http://www.cisco.com/en/US/products/ps10279/index.html) 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 (http://www.cisco.com/c/en/us/products/servers-unified-computing/ucs-b-series-blade-servers/index.html) increase performance, efficiency, versatility and productivity with these Intel based blade servers.

- Cisco UCS C-Series Rack Mount Servers (http://www.cisco.com/en/US/products/ps10493/index.html) deliver unified computing in an industry-standard form factor to reduce total cost of ownership and increase agility.
- Cisco UCS Adapters (http://www.cisco.com/en/US/products/ps10277/prod_module_series_home.html) wire-once architecture offers a range of options to converge the fabric, optimize virtualization and simplify management.

Cisco Nexus 2232PP 10GE Fabric Extender

The Cisco Nexus 2232PP 10G provides 32 10 Gb Ethernet and Fiber Channel Over Ethernet (FCoE) Small Form-Factor Pluggable Plus (SFP+) server ports and eight 10 Gb Ethernet and FCoE SFP+ uplink ports in a compact 1 rack unit (1RU) form factor.

When a Cisco UCS C-Series Rack-Mount Server is integrated with Cisco UCS Manager through the Cisco Nexus 2232 platform, the server is managed using the Cisco UCS Manager GUI or Cisco UCS Manager CLI. The Cisco Nexus 2232 provides data and control traffic support for the integrated Cisco UCS C-Series server.

For more information, refer to: http://www.cisco.com/c/en/us/support/switches/nexus-2232pp-10ge-fabric-extender/model.html.

Cisco Nexus 9000 Series Switch

The Cisco Nexus 9000 Series Switches offer both modular and fixed 10/25/40/100 Gigabit Ethernet switch configurations with scalability up to 60 Tbps of non-blocking performance with less than five-microsecond latency, 10/25/40/50/100 Gigabit Ethernet non-blocking Layer 2 and Layer 3 ports and wire speed VXLAN gateway, bridging, and routing support.

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

NetApp FAS and Data ONTAP

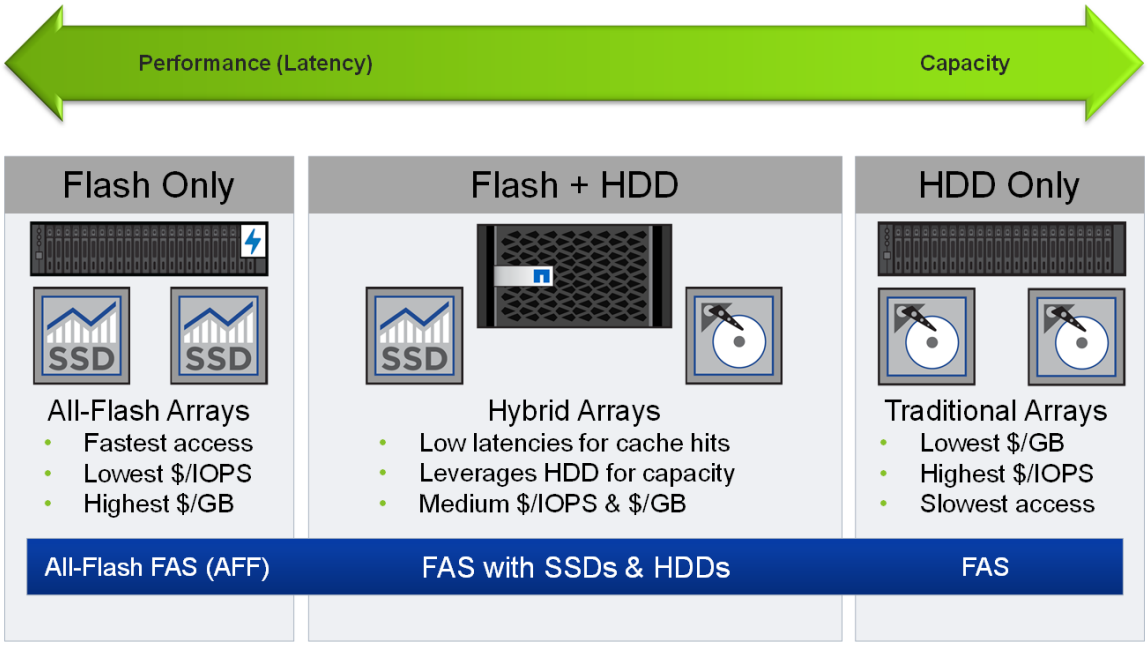
NetApp solutions offer increased availability and efficiency while consuming fewer IT resources. A NetApp solution includes hardware in the form of Fabric Attached Storage (FAS) controllers and disk storage, running the NetApp ONTAP operating system on the controllers. Disk storage is offered in two configurations: FAS with serial attached SCSI (SAS), serial ATA (SATA), or solid state drives (SSD) disks and All Flash FAS (AFF) with only SSD disks. The NetApp portfolio offers flexibility for selecting the controller and disk storage that best fits customer requirements. The storage efficiency built into ONTAP provides substantial space savings, allowing more data to be stored at a lower cost.

NetApp offers unified storage architecture, which simultaneously supports storage area network (SAN), network-attached storage (NAS), and iSCSI across many operating environments, including VMware, Windows®, and UNIX®. This single architecture provides access to data with industry-standard protocols, including NFS, CIFS, iSCSI, and FC/FCoE. Connectivity options include standard Ethernet (10/100/1000MbE or 10GbE) and Fibre Channel (4, 8, or 16Gb/sec).

In addition, all systems can be configured with high-performance SSD or SAS disks for primary storage applications, low-cost SATA disks for secondary applications (such as backup and archive), or a mix of different disk types, as illustrated in Figure 6. Note that the All Flash FAS configuration can only support

SSDs. Also supported is a hybrid cluster with a mix of All Flash FAS high availability (HA) pairs and FAS HA pairs with HDDs and/or SSDs.

Figure 6 NetApp Disk Options



For more information, see the Clustered Data ONTAP Storage Platform Mixing Rules and the NetApp ONTAP Software Overview page:

- https://library.netapp.com/ecm/ecm_get_file/ECMP1644424
- http://www.netapp.com/us/products/platform-os/data-ontap-8/index.aspx.

NetApp All Flash FAS

NetApp All Flash FAS addresses enterprise storage requirements with high performance, superior flexibility, and best-in-class data management. Built on the ONTAP software, All Flash FAS speeds up your business without compromising on efficiency, reliability, or the flexibility of your IT operations. As true enterprise-class, all-flash arrays, these systems accelerate, manage, and protect your business-critical data, now and in the future. With All Flash FAS systems, you can:

Accelerate the speed of business

- The ONTAP data management software employs the NetApp WAFL® (Write Anywhere File Layout) system, which is natively enabled for flash media
- NetApp FlashEssentials enables consistent sub-millisecond latency and up to 4 million IOPS
- The All Flash FAS system delivers 4 to 12 times higher IOPS and 20 times faster response time for databases than traditional hard disk drive (HDD) systems

Reduce costs while simplifying operations

- High performance enables server consolidation and can reduce database licensing costs by up to 50%

- As the industry's only unified all-flash storage that supports synchronous replication, All Flash FAS supports all of your backup and recovery needs with a complete suite of integrated data-protection utilities
- Data-reduction technologies can deliver space savings of 5 to 10 times on average
  - Newly enhanced inline compression delivers near-zero performance effect. Incompressible data detection eliminates wasted cycles.
  - Always-on deduplication runs continuously in the background and provides additional space savings for use cases such as virtual desktop deployments
  - Inline deduplication accelerates virtual machine (VM) provisioning by 20% to 30%
  - Advanced SSD partitioning increases usable capacity by almost 20%

Future-proof your investment with deployment flexibility

- All Flash FAS systems are ready for the data fabric. Data can move between the performance and capacity tiers on premises or in the cloud
- All Flash FAS offers application and ecosystem integration for virtual desktop integration (VDI), database, and server virtualization
- Without silos, you can non-disruptively scale out and move workloads between flash and HDDs within a cluster

All-Flash Performance Powered by NetApp FlashEssentials

NetApp FlashEssentials is behind the performance and efficiency of All Flash FAS and encapsulates the flash innovation and optimization technologies in ONTAP. Although ONTAP is well known as a leading storage operating system, it is not widely known that this system is natively suited for flash media due to the WAFL file system. FlashEssentials encompasses the technologies that optimize flash performance and media endurance, including:

- Coalesced writes to free blocks, maximizing the performance and longevity of flash media
- A random read I/O processing path that is designed from the ground up for flash
- A highly parallelized processing architecture that promotes consistent low latency
- Built-in quality of service (QoS) that safeguards SLAs in multi-workload and multi-tenant environments
- Inline data deduplication and compression innovations

For more information on All Flash FAS, click the following link:

http://www.netapp.com/us/products/storage-systems/all-flash-fas

NetApp ONTAP

With ONTAP, NetApp provides enterprise-ready, unified scale-out storage. Developed from a solid foundation of proven technology and innovation, ONTAP is the basis for large virtualized shared storage infrastructures that are architected for non-disruptive operations over the system lifetime. Controller nodes are deployed in HA pairs in a single storage domain or cluster.

ONTAP scale out is a way to respond to growth in a storage environment. As the storage environment grows, additional controllers are added seamlessly to the resource pool residing on a shared storage infrastructure. Host and client connections as well as datastores can move seamlessly and non-disruptively anywhere in the resource pool. Therefore, existing workloads can be easily balanced over the available resources, and new workloads can be easily deployed. Technology refreshes (for example, replacing disk shelves or adding or completely replacing storage controllers) are accomplished while the environment remains online and continues serving data. ONTAP is the first product to offer a complete scale-out solution, and it offers an adaptable, always-available storage infrastructure for today's highly virtualized environment.

NetApp Storage Virtual Machines

A cluster serves data through at least one and possibly multiple storage virtual machines (SVMs; formerly called Vservers). An SVM is a logical abstraction that represents the set of physical resources of the cluster. Data volumes and network logical interfaces (LIFs) are created and assigned to an SVM and may reside on any node in the cluster to which the SVM has been given access. An SVM may own resources on multiple nodes concurrently, and those resources can be moved non-disruptively from one node to another. For example, a flexible volume can be non-disruptively moved to a new node and aggregate, or a data LIF can be transparently reassigned to a different physical network port. The SVM abstracts the cluster hardware and is not tied to any specific physical hardware.

An SVM is capable of supporting multiple data protocols concurrently. Volumes within the SVM can be junctioned together to form a single NAS namespace, which makes all of an SVM's data available through a single share or mount point to NFS and CIFS clients. SVMs also support block-based protocols, and LUNs can be created and exported using iSCSI, Fibre Channel, or FCoE. Any or all of these data protocols can be configured for use within a given SVM.

Because it is a secure entity, an SVM is only aware of the resources that are assigned to it. It has no knowledge of other SVMs and their respective resources. Each SVM operates as a separate and distinct entity with its own security domain. Tenants can manage the resources allocated to them through a delegated SVM administration account. Each SVM can connect to unique authentication zones such as Active Directory®, LDAP, or NIS.

VMware vSphere

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

The VMware vSphere environment delivers a robust application environment. For example, with VMware vSphere, all applications can be protected from downtime with VMware High Availability (HA) without the complexity of conventional clustering. In addition, applications can be scaled dynamically to meet changing loads with capabilities such as Hot Add and VMware Distributed Resource Scheduler (DRS).

For more information, click the following link:

http://www.vmware.com/products/datacenter-virtualization/vsphere/overview.html

Domain and Element Management

This section provides general descriptions of the domain and element managers used during the validation effort. The following managers were used:

- Cisco UCS Manager
- Cisco APIC
- Cisco AVS
- Cisco VSUM
- VMware vCenter™ Server
- NetApp OnCommand® System Manager
- NetApp OnCommand® Unified Manager
- NetApp Virtual Storage Console (VSC)
- NetApp OnCommand Performance Manager
- NetApp Snap Manager and Snap Drive

Cisco Unified Computing System Manager

Cisco UCS Manager provides unified, centralized, embedded management of all Cisco Unified Computing System software and hardware components across multiple chassis and thousands of virtual machines. Administrators use the software to manage the entire Cisco Unified Computing System as a single logical entity through an intuitive GUI, a command-line interface (CLI), or an XML API.

The Cisco UCS Manager resides on a pair of Cisco UCS 6200 Series Fabric Interconnects using a clustered, active-standby configuration for high availability. The software gives administrators a single interface for performing server provisioning, device discovery, inventory, configuration, diagnostics, monitoring, fault detection, auditing, and statistics collection. Cisco UCS Manager service profiles and templates support versatile role- and policy-based management, and system configuration information can be exported to configuration management databases (CMDBs) to facilitate processes based on IT Infrastructure Library (ITIL) concepts. Service profiles benefit both virtualized and non-virtualized environments and increase the mobility of non-virtualized servers, such as when moving workloads from server to server or taking a server offline for service or upgrade. Profiles can be used in conjunction with virtualization clusters to bring new resources online easily, complementing existing virtual machine mobility.

For more information on Cisco UCS Manager, click the following link:

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

Cisco Application Policy Infrastructure Controller (APIC)

The Cisco Application Policy Infrastructure Controller (APIC) is the unifying point of automation and management for the ACI fabric. The Cisco APIC provides centralized access to all fabric information, optimizes the application lifecycle for scale and performance, and supports flexible application provisioning across physical and virtual resources. Some of the key benefits of Cisco APIC are:

- Centralized application-level policy engine for physical, virtual, and cloud infrastructures

- Detailed visibility, telemetry, and health scores by application and by tenant
- Designed around open standards and open APIs
- Robust implementation of multi-tenant security, quality of service (QoS), and high availability
- Integration with management systems such as VMware, Microsoft, and OpenStack

Cisco APIC exposes northbound APIs through XML and JSON and provides both a command-line interface (CLI) and GUI that utilize the APIs to manage the fabric holistically. For redundancy and load distribution, three APIC controllers are recommended for managing ACI fabric.

For more information on Cisco APIC, click the following link:

http://www.cisco.com/c/en/us/products/cloud-systems-management/application-policy-infrastructure-controller-apic/index.html

Cisco Application Virtual Switch (AVS)

Cisco Application Virtual Switch (AVS) is an optional hypervisor-resident virtual network switch that is specifically designed for the Application Centric Infrastructure (ACI) architecture. Based on the highly successful Cisco Nexus 1000V virtual switch, AVS provides feature support for the ACI application policy model, full switching capabilities, and more advanced telemetry features. Some of the key benefits of Cisco AVS are:

- 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
- High performance and throughput
- 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

As a virtual leaf in the ACI Fabric, the Cisco AVS in VXLAN mode only requires extending the ACI fabric system VLAN through the Cisco UCS Fabric Interconnects to the VXLAN tunneling endpoint (VTEP) vmkernel (VMK) ports on the ESXi hosts.

For more information on Cisco AVS, click the following link:

http://www.cisco.com/c/en/us/products/switches/application-virtual-switch/index.html.

Cisco Virtual Switch Update Manager (VSUM)

Cisco VSUM is a virtual appliance that is registered as a plug-in to the VMware vCenter Server. Cisco VSUM simplifies the installation and configuration of the Cisco AVS. Some of the key benefits of Cisco VSUM are:

- Install the Cisco AVS vSphere Installation Bundle (VIB) to the ESXi host
- Add hosts to the Cisco AVS
- Upgrade the Cisco AVS

For more information on Cisco VSUM, click the following link:

http://www.cisco.com/c/en/us/td/docs/switches/datacenter/nexus1000/vsum/2-0/release_notes/n1K-vsum-2-0.html

VMware vCenter Server

VMware vCenter Server is the simplest and most efficient way to manage VMware vSphere, irrespective of the number of VMs you have. It provides unified management of all hosts and VMs from a single console and aggregates performance monitoring of clusters, hosts, and VMs. VMware vCenter Server gives administrators a deep insight into the status and configuration of compute clusters, hosts, VMs, 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 typical productivity in managing physical infrastructure. VMware vCenter manages the rich set of features available in a VMware vSphere environment.

For more information, click the following link:

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

NetApp OnCommand System and Unified Manager

NetApp OnCommand® System Manager allows storage administrators to manage individual storage systems or clusters of storage systems. Its easy-to-use interface simplifies common storage administration tasks such as creating volumes, LUNs, qtrees, shares, and exports, saving time and helping to prevent errors. System Manager works across all NetApp storage systems. NetApp OnCommand Unified Manager complements the features of System Manager by enabling the monitoring and management of storage within the NetApp storage infrastructure.

This solution uses both OnCommand System Manager and OnCommand Unified Manager to provide storage provisioning and monitoring capabilities within the infrastructure.

NetApp OnCommand Performance Manager

OnCommand Performance Manager provides performance monitoring and incident root-cause analysis of systems running ONTAP®. It is the performance management part of OnCommand Unified Manager and is integrated into OnCommand Unified Manager's User Interface. Performance Manager helps you identify workloads that are overusing cluster components and decreasing the performance of other workloads on the cluster. It alerts you to these performance events, called incidents, so that you can take corrective action and return the system back to normal operation. You can view and analyze incidents in the Performance Manager GUI or view them on the Unified Manager Dashboard.

NetApp Virtual Storage Console

The NetApp Virtual Storage Console (VSC) delivers storage configuration and monitoring, datastore provisioning, virtual machine (VM) cloning, and backup and recovery of VMs and datastores. VSC also includes an application-programming interface (API) for automated control.

VSC is a single VMware vCenter Server Web Interface plug-in that provides end-to-end VM lifecycle management for VMware environments that use NetApp storage. VSC is available to all VMware vSphere clients that connect to the vCenter server. This availability is different from a client-side plug-in that must be installed on every VMware vSphere client. The VSC software can be installed either on the Microsoft Windows-based vCenter server or on a separate Microsoft Windows Server® instance or VM.

NetApp SnapManager and NetApp SnapDrive

NetApp SnapManager® storage management software and NetApp SnapDrive® data management software are used to provision and back up storage for applications under Cisco ACI in this solution. The portfolio of SnapManager products is specific to the particular application. SnapDrive is used with all SnapManager products.

To create a backup, SnapManager interacts with the application so that application data is placed in a state such that a consistent NetApp Snapshot® copy of that data can be made. It then causes SnapDrive to interact with the storage system SVM to create the Snapshot copy, effectively backing up the application data. In addition to managing Snapshot copies of application data, SnapDrive can be used to accomplish the following tasks:

- Provisioning application data LUNs in the SVM as mapped disks on the application VM
- Managing Snapshot copies of application mapped disks or VMDK disks on NFS or VMFS datastores

Snapshot copy management of application data LUNs is handled by the interaction of SnapDrive with the SVM management LIF.

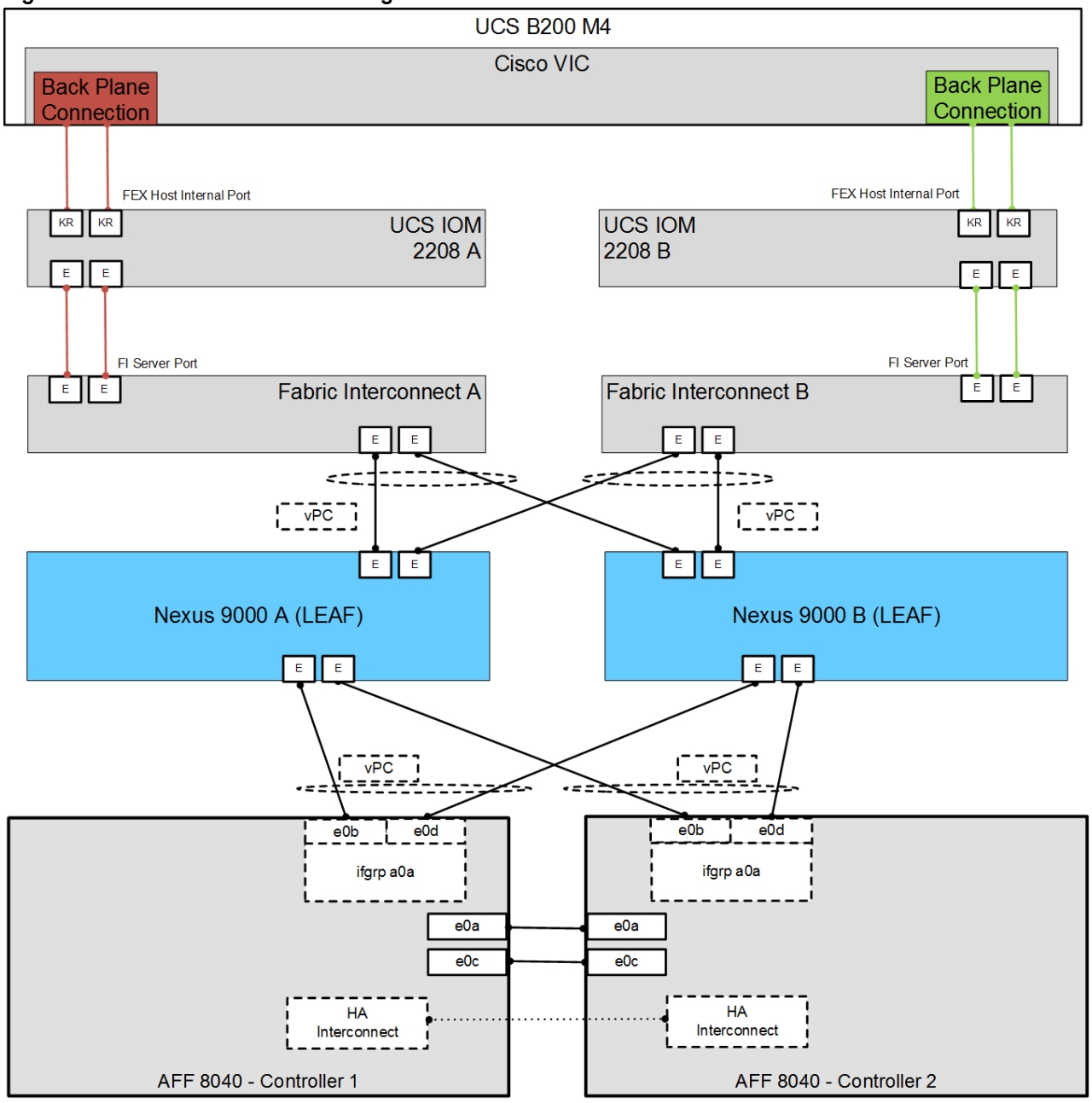
Solution Architecture

FlexPod Infrastructure Physical Building Blocks

Physical Topology

Figure 7 illustrates the new ACI connected FlexPod design. The infrastructure is physically redundant across the stack, addressing Layer 1 high-availability requirements where the integrated stack can withstand failure of a link or failure of a device. The solution also incorporates additional Cisco and NetApp technologies and features that further increase the design efficiency. Figure 7 illustrates the compute, network and storage design overview of the FlexPod solution. The individual details of these components will be covered in the upcoming sections.

Figure 7 Cisco Nexus 9000 Design for Clustered Data ONTAP



Cisco Unified Computing System

The FlexPod compute design supports both Cisco UCS B-Series and C-Series deployments. The components of the Cisco Unified Computing System offer physical redundancy and a set of logical structures to deliver a very resilient FlexPod compute domain. In this validation effort, multiple Cisco UCS B-Series and C-Series ESXi servers are booted from SAN using either iSCSI or FCoE.

Cisco UCS Physical Connectivity

Cisco UCS Fabric Interconnects are configured with two port-channels, one from each FI, to the Cisco Nexus 9000s. These port-channels carry all the data and storage traffic originated on the Cisco Unified Computing System. The validated design utilized two uplinks from each FI to the leaf switches for an

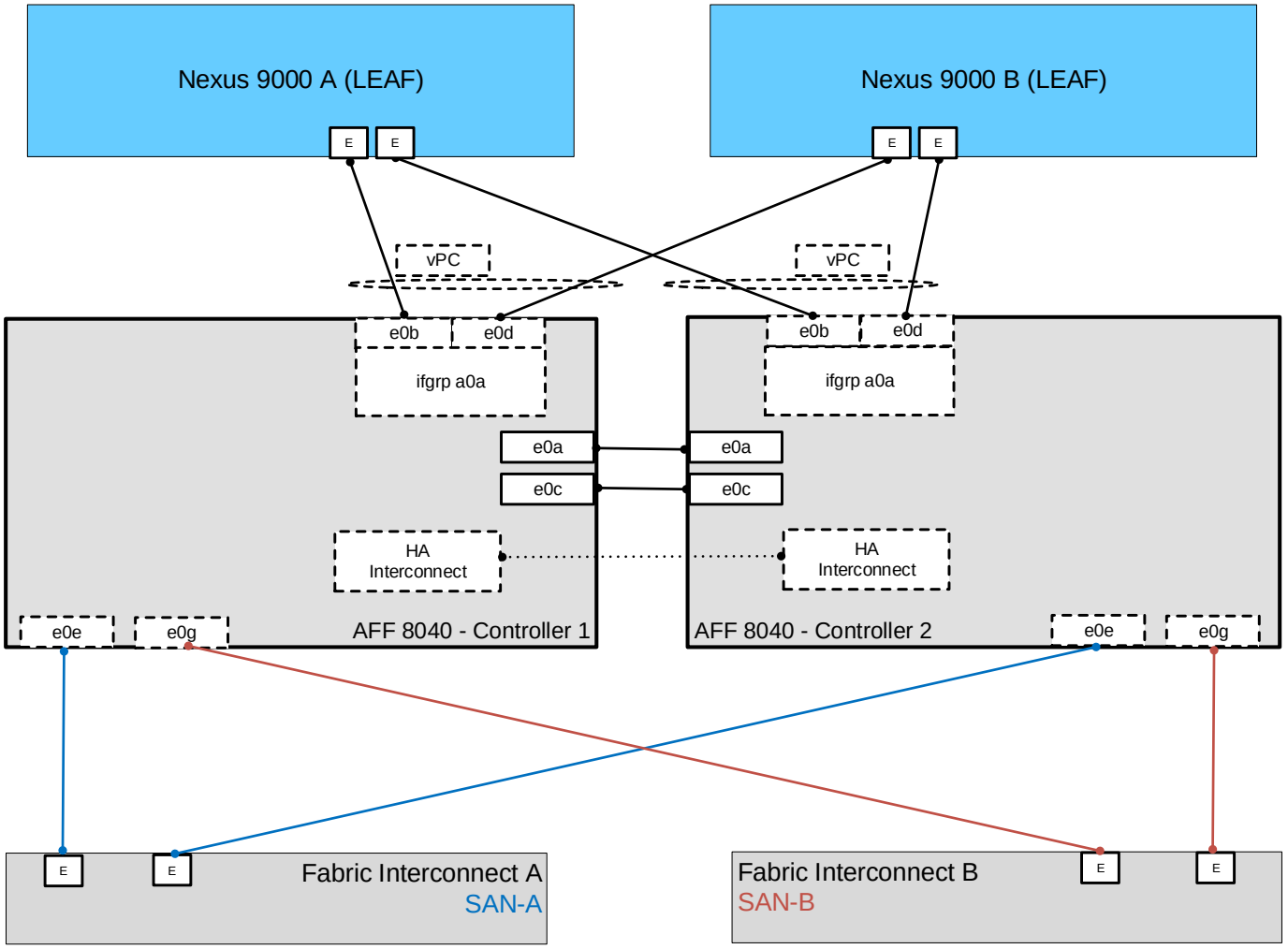
aggregate bandwidth of 40GbE (4 x 10GbE). The number of links can be easily increased based on customer data throughput requirements.

FCoE Connectivity

The FlexPod with ACI design also supports SAN using FCoE by directly connecting the NetApp controllers to the Cisco UCS Fabric Interconnects. The updated physical design changes are covered in Figure 8.

In the FCoE design, zoning and related SAN configuration is configured on Cisco UCS Manager and the Fabric Interconnects provide the SAN-A and SAN-B separation. On NetApp, the Unified Target Adapter is needed to provide physical connectivity.

Figure 8 Boot from SAN using FCoE (Optional)

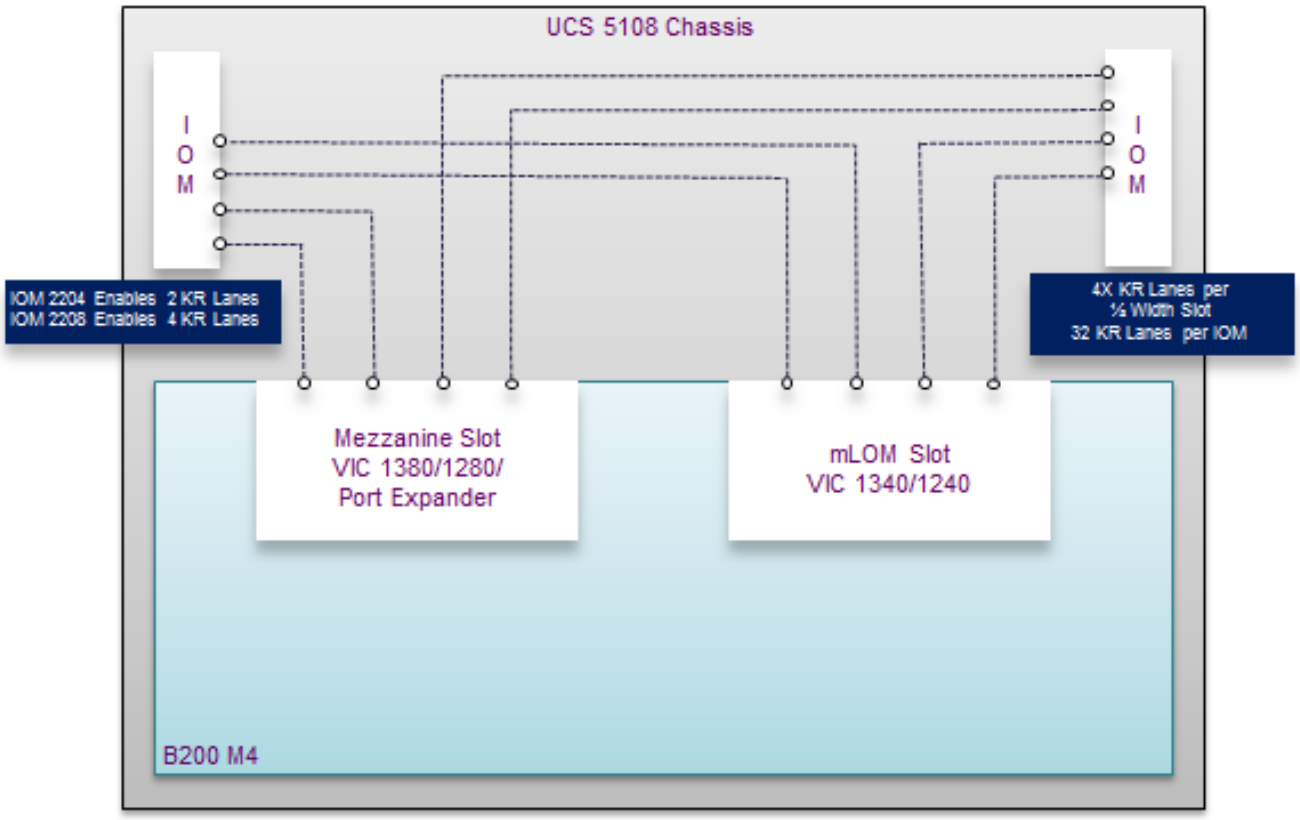


Cisco Unified Computing System I/O Component Selection

FlexPod allows customers to adjust the individual components of the system to meet their particular scale or performance requirements. Selection of I/O components has a direct impact on scale and performance characteristics when ordering the Cisco components. Figure 9 illustrates the available backplane connections in the Cisco UCS 5100 series chassis. As shown, each of the two Fabric Extenders (I/O module) has either two or four 10GBASE KR (802.3ap) standardized Ethernet backplane KR lanes available for connection to each 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:

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

Figure 9 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. The 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 midplane 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
UCS 2204XP	4	16
UCS 2208XP	8	32

MLOM Virtual Interface Card (VIC)

The FlexPod solution with B-Series is typically validated using Cisco VIC 1340 or Cisco VIC 1380. The Cisco VIC 1340, the next generation 2-port, 40 Gigabit Ethernet, Fibre Channel over Ethernet (FCoE)-capable modular LAN on motherboard (mLOM) mezzanine adapter is designed for both Cisco UCS B200 M3 and M4

generations of Cisco UCS B-Series Blade Servers. When used in combination with an optional Port Expander, the Cisco UCS VIC 1340 capabilities can be expanded to eight ports of 10 Gigabit Ethernet with the use of Cisco UCS 2208XP fabric extenders.

Mezzanine Slot Card

A Cisco VIC 1380 is an eight-port 10 Gigabit Ethernet, Fibre Channel over Ethernet (FCoE)-capable mezzanine card designed exclusively for Cisco UCS B-Series Blade Servers.

Server Traffic Aggregation

Selection of the FEX, VIC and Mezzanine cards plays a major role in determining the aggregate traffic throughput to and from a server. Figure 9 shows an overview of backplane connectivity for both the I/O Modules and Cisco VICs. The number of KR lanes indicates the 10GbE paths available to the chassis and therefore blades. As shown in Figure 9, depending on the models of I/O modules and VICs, traffic aggregation differs. 2204XP enables two KR lanes per half-width blade slot while the 2208XP enables all four. Similarly, the number of KR lanes varies based on selection of VIC 1340/1240, VIC 1340/1240 with Port Expander and VIC 1380/1280.

Validated I/O Component Configurations

Two of the most commonly validated I/O component configurations in FlexPod designs are:

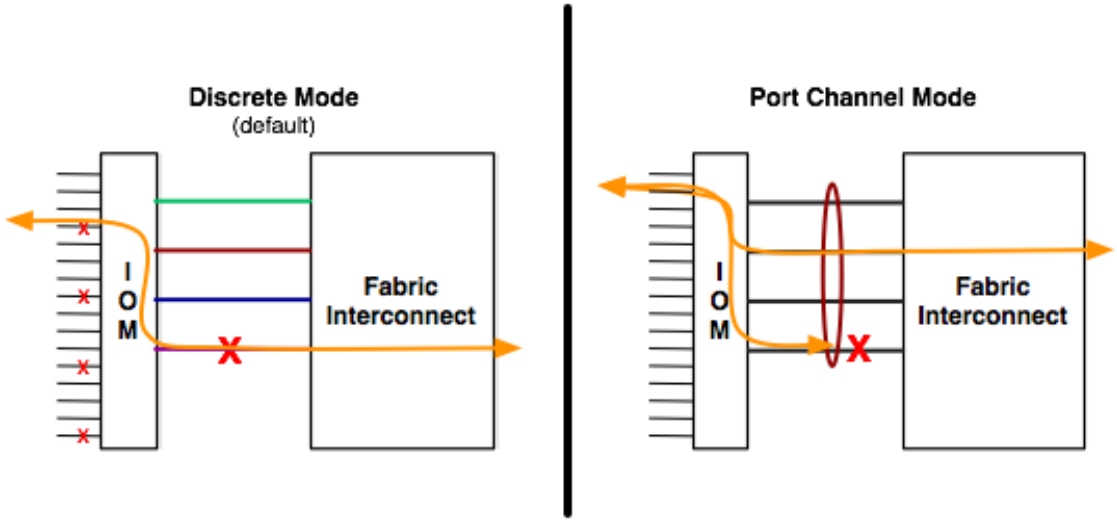
- Cisco UCS B200M4 with VIC 1340, Port Expander and FEX 2208XP
- Cisco UCS B200M4 with VIC 1340 and FEX 2208XP

In the configuration with Cisco UCS B200M4 with VIC 1340, Port Expander and FEX 2208XP, the FEX 2208XP enables 8 KR lanes to the half-width blade (4 to each IOM), while the UCS global discovery policy dictates the formation of fabric port channels. The VMware vmnics will appear as 40 Gb/s NICs. In the configuration with Cisco UCS B200M4 with VIC 1340 and FEX 2208XP, the VMware vmnics will appear as 20 Gb/s NICs.

Cisco Unified Computing System Chassis/FEX Discovery Policy

Cisco Unified Computing System can be configured to discover a chassis using Discrete Mode or the Port-Channel mode (Figure 10). 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 FlexPod designs.

Figure 10 Cisco UCS Chassis Discovery Policy—Discrete Mode vs. Port Channel Mode



Cisco Unified Computing System—QoS and Jumbo Frames

FlexPod accommodates a myriad of traffic types (vMotion, NFS, FCoE, control traffic, etc.) and is capable of absorbing traffic spikes and protects against traffic loss. Cisco UCS and Nexus QoS system classes and policies deliver this functionality. In this validation effort, the FlexPod was configured to support jumbo frames with an MTU size of 9000. Enabling jumbo frames allows the FlexPod environment to optimize throughput between devices while simultaneously reducing the consumption of CPU resources.

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

Cisco Unified Computing System—Cisco UCS C-Series Server Design

Fabric Interconnect—UCS C-Series Server Direct Attached Design

Cisco UCS Manager 3.1 allows 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.

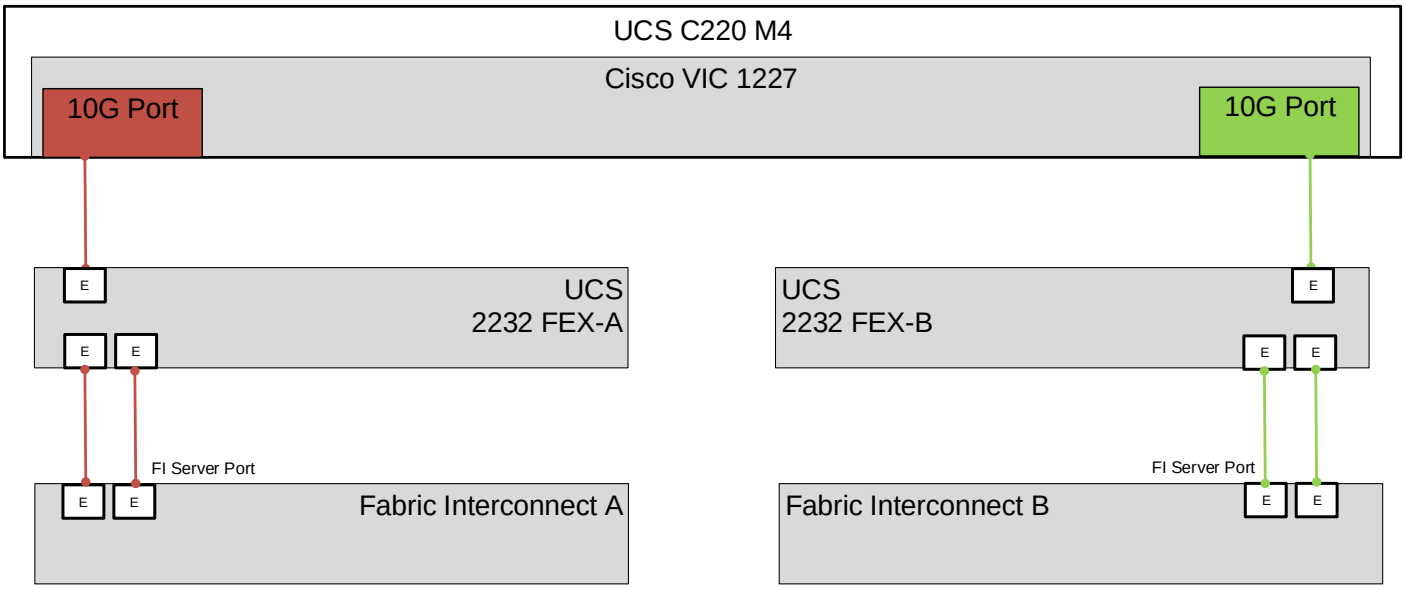
Note: For detailed connectivity requirements, refer to: http://www.cisco.com/c/en/us/td/docs/unified_computing/ucs/c-series_integration/ucsm3-1/b_C-Series-Integration_UCSM3-1/b_C-Series-Integration_UCSM3-1_chapter_01.html

Fabric Interconnect—Fabric Extender Attached Design

Figure 11 illustrates the connectivity of the Cisco UCS C-Series server into the Cisco UCS domain using a Fabric Extender. Functionally, the one RU Nexus FEX 2232PP is an external device that replaces the Cisco UCS 2204 or 2208 IOM (located within the Cisco UCS 5108 blade chassis). Each 10GbE VIC port connects to Fabric A or B through the FEX. The FEX and Fabric Interconnects form port channels automatically based on the chassis discovery policy providing a link resiliency to the Cisco UCS C-series server. This is identical to the behavior of the IOM to Fabric Interconnect connectivity. Logically, the virtual circuits formed within the

Cisco UCS domain are consistent between B and C series deployment models and the virtual constructs formed at the vSphere level are unaware of the platform in use.

Figure 11 Cisco UCS C-Series with VIC 1227



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 supporting infrastructure services such as vSphere Virtual Center, Microsoft Active Directory and NetApp Virtual Storage Console (VSC).

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. The ESXi operating system is unaware of the fact that the vNICs are virtual adapters. In the FlexPod with Cisco ACI design, six vNICs are created and utilized as follows:

- Two vNICs carry in-band management traffic, infrastructure NAS storage traffic, and Core Services Virtual Machine (VM) traffic
- One vNIC carries infrastructure and tenant iSCSI-A traffic (SAN A)
- One vNIC carries infrastructure and tenant iSCSI-B traffic (SAN B)
- Two vNICs carry data traffic including storage traffic for infrastructure and tenants and are connected to either the APIC-controlled VMware vSphere Distributed Switch (vDS) or the Cisco Application Virtual Switch (AVS)

These vNICs are pinned to the Fabric Interconnect uplink interfaces.

Figure 12 ESXi Server–vNICs and vmnics

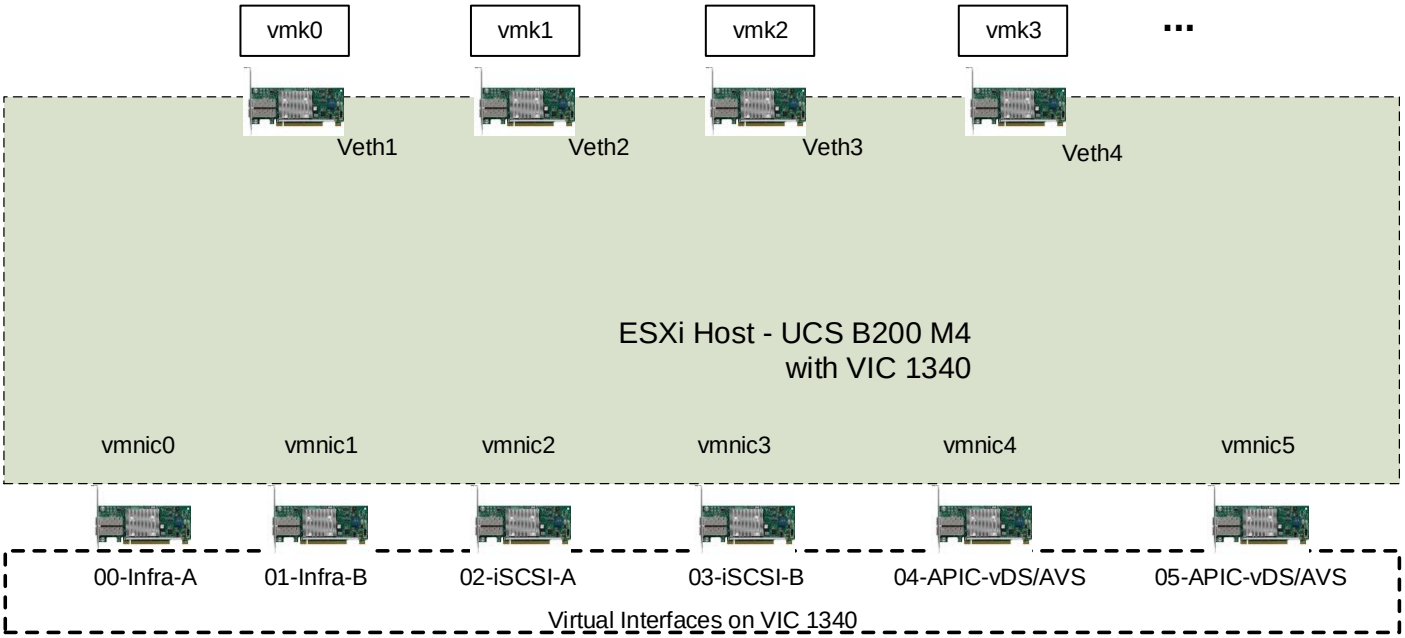


Figure 12 details the ESXi server design showing both virtual interfaces and VMkernel ports. All the Ethernet adapters vmnic0 through vmnic5 are virtual NICs created using Service Profile.

NetApp Storage Design

The FlexPod storage design supports a variety of NetApp FAS controllers such as the AFF8000, the FAS 2500 and the FAS 8000 in addition to legacy NetApp storage. This Cisco Validated Design leverages a pair of NetApp AFF8040 controllers, deployed with ONTAP software.

In the ONTAP architecture, all data is accessed through secure virtual storage partitions known as storage virtual machines (SVMs). You can have a single SVM that represents the resources of the entire cluster or multiple SVMs that are assigned specific subsets of cluster resources for given applications, tenants or workloads. In the current implementation of ACI, SVMs provide storage for each application that has ESXi hosts booted from SAN. SVMs also provide storage for application data presented as iSCSI, FCoE, CIFS or NFS traffic.

For more information about the AFF8000 product family, click the following links:

http://www.netapp.com/us/products/storage-systems/all-flash-fas

For more information about the FAS 8000 product family, see:

http://www.netapp.com/us/products/storage-systems/fas8000/

For more information about the FAS 2500 product family, see:

http://www.netapp.com/us/products/storage-systems/fas2500/index.aspx

For more information about clustered Data ONTAP, see:

http://www.netapp.com/us/products/platform-os/data-ontap-8/index.aspx

Network and Storage Physical Connectivity

NetApp AFF8000 storage controllers are configured with two port channels, one per controller, connected to the Cisco Nexus 9000 leaf switches. These port channels carry all IP-based ingress and egress data traffic for the NetApp controllers. This validated design uses two physical ports from each NetApp controller, configured as a LACP interface group (ifgrp). The number of ports used can be readily increased depending on the application requirements.

An ONTAP storage solution includes the following connections or network types:

- HA interconnect. A dedicated interconnect between two nodes to form HA pairs. These pairs are also known as storage failover pairs.
- Cluster interconnect. A dedicated high-speed, low-latency, private network used for communication between nodes. This network can be implemented through the deployment of a switchless cluster or by leveraging dedicated cluster interconnect switches.

Note: NetApp switchless cluster is only supported for two node clusters.

- Management network. A network used for the administration of nodes, the cluster, and SVMs.
- Data network. A network used by clients to access data.
- Ports. A physical port such as e0a or e1a or a logical port such as a virtual LAN (VLAN) or an interface group.
- Interface groups. A collection of physical ports to create one logical port. The NetApp interface group is a link aggregation technology that can be deployed in single (active/passive), multiple ("always on"), or dynamic (active LACP) mode.

This validation uses two storage nodes configured as a switchless, two-node storage failover pair through an internal HA interconnect direct connection. The FlexPod design uses the following port and interface assignments:

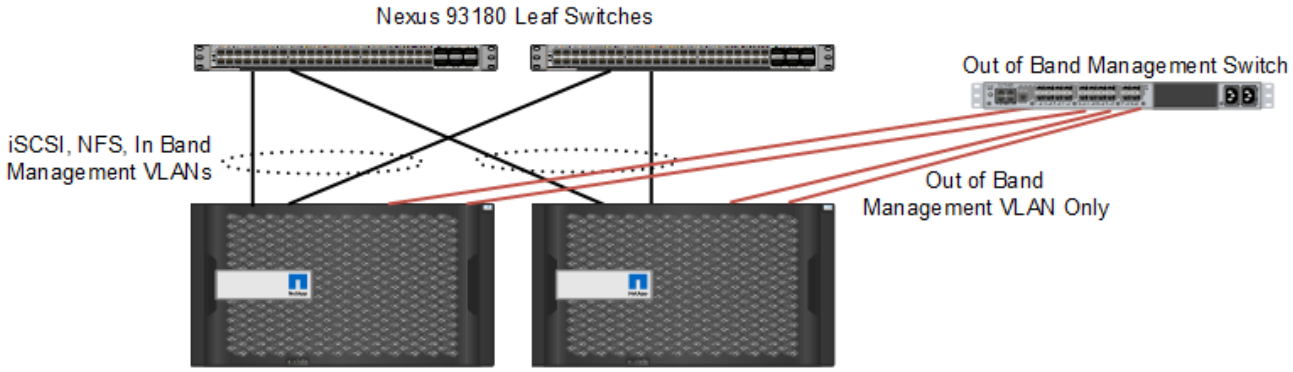
- Ethernet ports e0b and e0d on each node are members of a multimode LACP interface group for Ethernet data. This design leverages an interface group that has LIFs associated with it to support NFS,iSCSI, and SVM management traffic.
- Ethernet ports e0a and e0c on each node are connected to the corresponding ports on the other node to form the switchless cluster interconnect.
- Port e0M on each node supports a LIF dedicated to node management. Port e0i is defined as a failover port supporting the “node_mgmt” role.
- Port e0i supports cluster management data traffic through the cluster management LIF. This port and LIF allow for administration of the cluster from the failover port and LIF if necessary.

Out of Band Network Connectivity

FlexPod leverages out-of-band management networking. Port e0M on each node supports a LIF dedicated to node management. Port e0i is defined as a failover port for node management. To support out-of-band management connectivity, the NetApp controllers are directly connected to out-of-band management switches as shown in Figure 13. The e0M and e0i ports can be connected to two different out-of-band management switches for hardware redundancy.

Note: The AFF8040 controllers are sold in a single-chassis, dual-controller option only. Figure 13 represents the NetApp storage controllers as a dual-chassis dual-controller deployment. Figure 13 shows two AFF8040 controllers side by side, for visual purposes only.

Figure 13 Storage Out of Band Management Network Connectivity

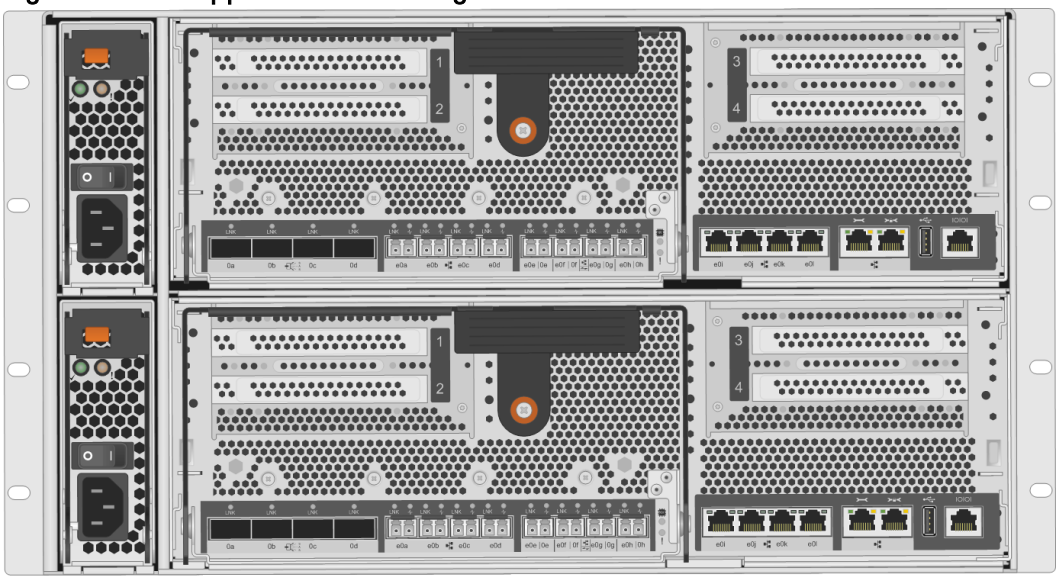


NetApp FAS I/O Connectivity

One of the main benefits of FlexPod is that it gives you the ability to right-size your deployment. This effort can include the selection of the appropriate protocol for your workload as well as the performance capabilities of various transport protocols. The AFF 8000 product family supports FC, FCoE, iSCSI, NFS, pNFS, and CIFS/SMB. The AFF8000 comes standard with onboard UTA2, 10GbE, 1GbE, and SAS ports. In addition, the AFF8000 offers up to 24 PCIe expansion ports per HA pair.

Figure 14 depicts the rear of the AFF8040 chassis. The AFF8040 is configured in a single HA enclosure that contains controllers housed in a single chassis. External disk shelves are connected through onboard SAS ports, data is accessed through the onboard UTA2 ports, and cluster interconnect traffic is over the onboard 10GbE ports.

Figure 14 NetApp AFF 8000 Storage Controller



NetApp ONTAP and Storage Virtual Machines Overview

ONTAP allows the logical partitioning of storage resources in the form of SVMs. The following components comprise an SVM:

- Logical interfaces: All SVM networking is performed through logical interfaces (LIFs) that are created within the SVM. As logical constructs, LIFs are abstracted from the physical networking ports on which they reside.

- Flexible volumes: A flexible volume is the basic unit of storage for an SVM. An SVM has a root volume and can have one or more data volumes. Data volumes can be created in any aggregate that has been delegated by the cluster administrator for use by the SVM. Depending on the data protocols used by the SVM, volumes can contain either LUNs for use with block protocols, files for use with NAS protocols, or both concurrently.

- Namespace: Each SVM has a distinct namespace through which all of the NAS data shared from that SVM can be accessed. This namespace can be thought of as a map to all of the junctioned volumes for the SVM, no matter on which node or aggregate they might physically reside. Volumes can be junctioned at the root of the namespace or beneath other volumes that are part of the namespace hierarchy.

- IPspaces: The IPspace feature enables a single storage system to be accessed by clients from more than one disconnected network, even if those clients are using the same IP address. An IPspace defines a distinct IP address space and routing domain in which SVMs can participate. IP addresses defined for an IPspace are applicable only within that IPspace. A distinct routing table is maintained for each IPspace. No cross-IPspace traffic routing happens.

- Storage QoS: Storage Quality of Service (QoS) helps manage the risks associated with meeting performance objectives. You can use storage QoS to limit the throughput to workloads and to monitor workload performance. You can reactively limit workloads to address performance problems and you can proactively limit workloads to prevent performance problems. You can also limit workloads to support SLAs with customers. Workloads can be limited on the basis of either workload IOPs or bandwidth in MB/s.

 Note: Storage QoS is supported on clusters that have up to eight nodes.

A workload represents the input/output (I/O) operations to one of the following storage objects:

- An SVM with flexible volumes

- A flexible volume

- A LUN

- A file (typically represents a VM)

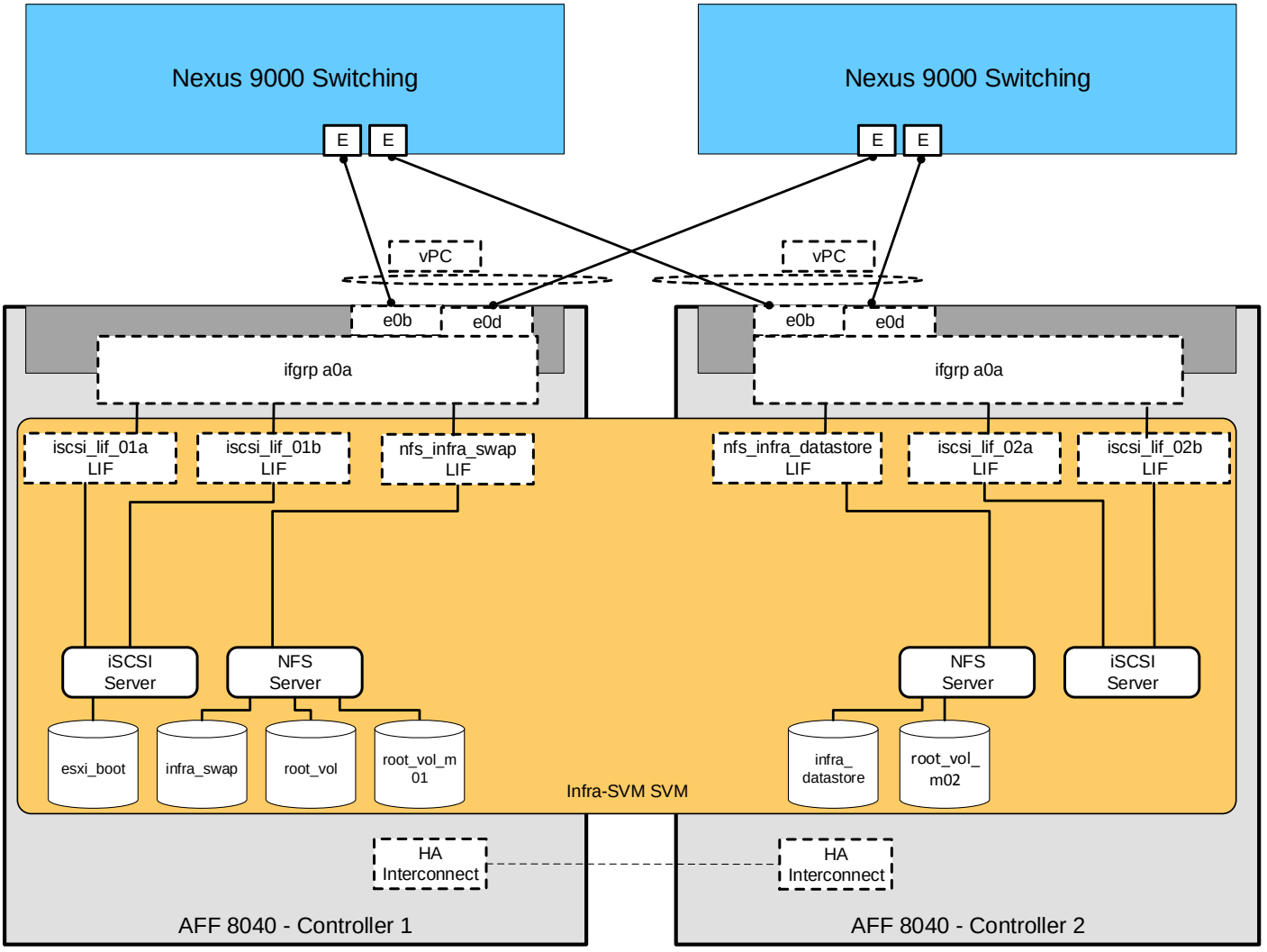
In the ACI architecture, because an SVM is usually associated with an application, a QoS policy group is normally applied to the SVM, setting up an overall storage rate limit for the workload. Storage QoS is administered by the cluster administrator.

Storage objects are assigned to a QoS policy group to control and monitor a workload. You can monitor workloads without controlling them in order to size the workload and determine appropriate limits within the storage cluster.

For more information about managing workload performance by using storage QoS, see the section "Managing system performance" in the Clustered Data ONTAP 8.3 System Administration Guide for Cluster Administrators. ONTAP Logical Topology

Figure 15 details the logical configuration of the clustered Data ONTAP environment used for validation of the FlexPod solution. The physical cluster consists of two NetApp storage controllers (nodes) configured as an HA pair and a switchless cluster.

Figure 15 NetApp Storage Controller—Clustered Data ONTAP



The following key components enable connectivity to data on a per application basis:

LIF: A logical interface that is associated with a physical port, an interface group, or a VLAN interface. More than one LIF can be associated with a physical port at the same time. There are three types of LIFs:

- NFS LIFs
- iSCSI LIFs
- Fibre Channel Protocol (FCP) LIFs

LIFs are logical network entities that have the same characteristics as physical network devices and yet are not tied to physical objects. LIFs used for Ethernet traffic are assigned specific Ethernet-based details such as IP addresses. They are then associated with a specific physical port capable of supporting Ethernet traffic. NAS LIFs can be non-disruptively migrated to any other physical network port throughout the entire cluster at any time, either manually or automatically (by using policies).

In this Cisco Validated Design, LIFs are layered on top of the physical interface groups and are associated with a given VLAN interface. LIFs are then consumed by the SVMs and are typically associated with a given protocol and data store.

SVM: An SVM is a secure virtual storage server that contains data volumes and one or more LIFs, through which it serves data to the clients. An SVM securely isolates shared virtualized data storage and network resources and appears as a single dedicated server to its clients. Each SVM has a separate administrator authentication domain and can be managed independently by an SVM administrator.

ONTAP Configuration for vSphere

This solution defines a single infrastructure SVM to own and export the data necessary to run the VMware vSphere infrastructure. This SVM specifically owns the following flexible volumes:

- Root volume. A flexible volume that contains the root of the SVM namespace.
- Root volume load-sharing mirrors. Mirrored volumes of the root volume created to accelerate read throughput. In this instance, they are labeled `root_vol_m01` and `root_vol_m02`.
- Boot volume. A flexible volume that contains ESXi boot LUNs. These ESXi boot LUNs are exported through either iSCSI or FCoE to the Cisco UCS servers.
- Infrastructure datastore volume. A flexible volume that is exported through NFS to the ESXi host and is used as the infrastructure NFS datastore to store VM files.
- Infrastructure swap volume. A flexible volume that is exported through NFS to each ESXi host and used to store VM swap data.

The NFS datastores are mounted on each VMware ESXi host in the VMware cluster and are provided by ONTAP through NFS over the 10GbE network. The SVM has a minimum of one LIF per protocol per node where the volume needs to be accessed to maintain volume availability across the cluster nodes. It has also been a best practice in FlexPod with VMware to use one LIF per NFS datastore. This allows workloads to more easily be moved around in the cluster both for load balancing and to take advantage of hybrid clusters, and can ease cluster troubleshooting by having the ability to list in one place what datastores are mounted on what LIFs. The LIFs use failover groups generated from broadcast domains. Failover groups are network polices defining the ports or interface groups available to support a single LIF migration or a group of LIFs migrating within or across nodes in a cluster. Multiple LIFs may be associated with a network port or interface group.

In addition to failover groups, the ONTAP system uses failover policies. Failover polices define the order in which the ports in the failover group are prioritized. Failover policies define migration policy in the event of port failures, port recoveries, or user-initiated requests. The most basic storage failover scenarios possible in this cluster are as follows:

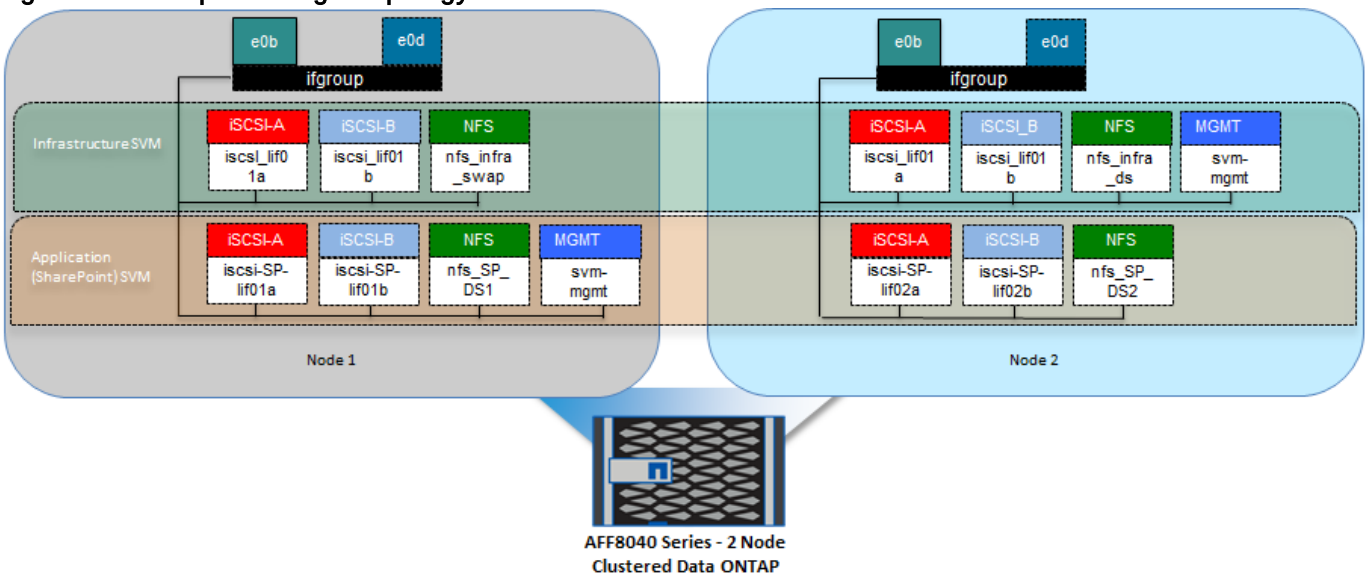
- Node 1 fails, and Node 2 takes over Node 1's storage.
- Node 2 fails, and Node 1 takes over Node 2's storage.

The remaining node network connectivity failures are addressed through the redundant port, interface groups, and logical interface abstractions afforded by the ONTAP system.

Storage Virtual Machine Layout

Figure 16 highlights the storage topology showing SVM and associated LIFs. There are two storage nodes and the SVM's are layered across both controller nodes. Each SVM has its own LIFs configured to support SVM specific storage protocols. Each of these LIFs are mapped to end-point groups on the ACI fabric.

Figure 16 Sample Storage Topology



Cisco Nexus 9000 ACI

In the current Cisco Validated Design, the Cisco Nexus 9336 Spine and the Cisco Nexus 9396, 9372, or 93180 leaf switches provide an ACI based Ethernet switching fabric for communication between the virtual machine and bare metal compute, NFS and iSCSI based storage and the existing traditional enterprise networks. Similar to previous versions of FlexPod, the virtual port channel plays an important role in providing the necessary connectivity.

Virtual Port Channel (vPC) Configuration

A virtual PortChannel (vPC) allows a device's Ethernet links that are physically connected to two different Cisco Nexus 9000 Series devices to appear as a single PortChannel. In a switching environment, a vPC provides the following benefits:

- Allows a single device to use a PortChannel across two upstream devices
- Eliminates Spanning Tree Protocol blocked ports and uses all available uplink bandwidth
- Provides a loop-free topology
- Provides fast convergence if either one of the physical links or a device fails
- Helps ensure high availability of the overall FlexPod system

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 (leaves). The peer communication is carried over the 40G connections through the Spines.

Compute and Storage Connectivity

Cisco UCS Fabric Interconnects and NetApp storage systems are connected to the Cisco Nexus 9000 leaf switches using vPCs. The Port Channels connecting NetApp controllers to the ACI fabric are configured with three types of VLANs:

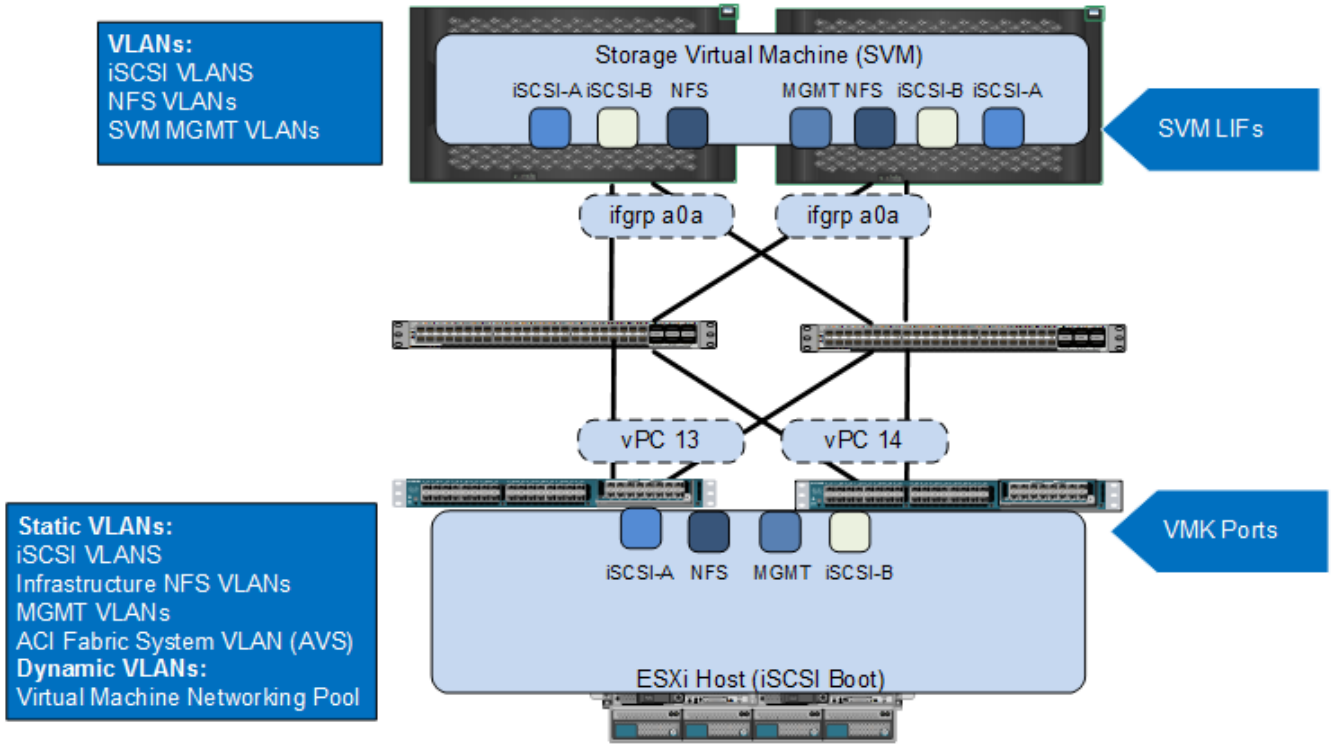
- iSCSI VLANs to provide storage LUN access including boot LUNs
- NFS VLANs to access Infrastructure and swap datastore volumes
- Management VLAN(s) to provide access to Tenant SVMs

The Port Channels connecting the Cisco UCS Fabric Interconnects to the ACI fabric are configured with up to six types of VLANs:

- iSCSI VLANs to provide ESXi hosts access to boot, application data, and datastore LUNs
- NFS VLANs to access virtual machine and swap datastores to be used by the vSphere environment to host virtual machine services
- The In Band Management VLAN for management access to the ESXi hosts and management VMs.
- A pool of VLANs associated with an ACI Virtual Machine Manager (VMM) domain for the VMware vSphere Distributed Switch (vDS). VLANs from this pool are dynamically allocated by the APIC to newly created end point groups (EPGs) that become port-profiles in the VMware vDS
- The ACI system VLAN to connect the Cisco AVS VXLAN tunnel endpoints (VTEPs) to the ACI fabric VTEPs when the Cisco AVS is being used in VXLAN mode
- The Cisco UCS port channels also include the vMotion VLAN, but since that VLAN is not mapped to storage, it is not shown the the next subsections.

These VLAN configurations are covered in detail in the next subsections.

Figure 17 Compute and Storage Connectivity to Cisco Nexus 9000 ACI



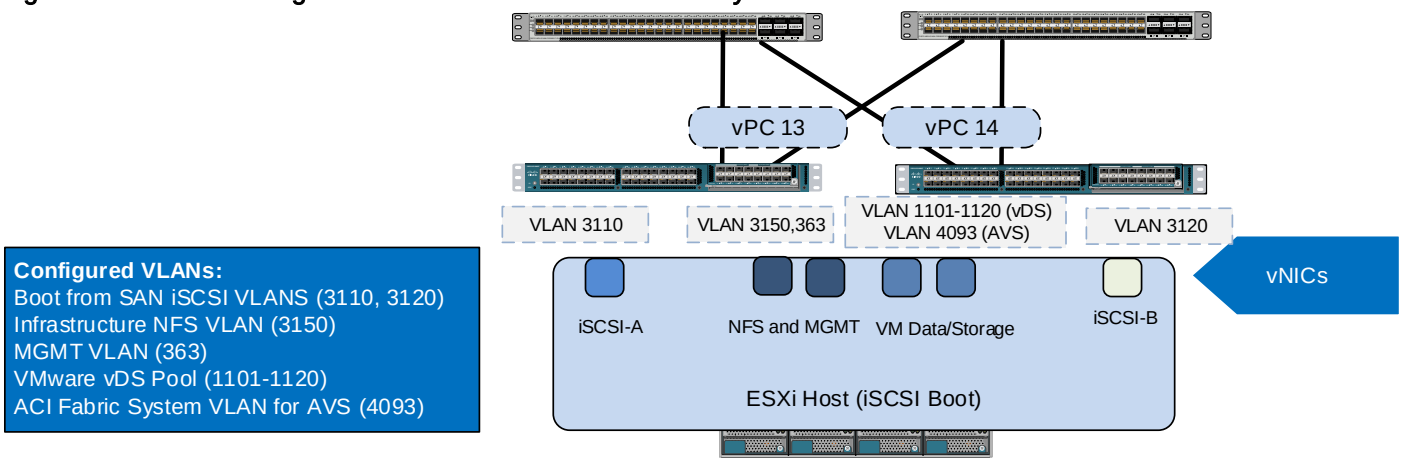
VLAN Configuration for Cisco Unified Computing System

For the Cisco Unified Computing System to Cisco Nexus 9000 connectivity, iSCSI VLANs associated with boot from SAN configuration and the NFS VLANs used by the infrastructure ESXi hosts are pre-configured on the Cisco UCS Fabric Interconnect. In 0, VLANs 3110 and 3120 are the Infrastructure Tenant iSCSI-A and iSCSI-B VLANs that are assigned to individual virtual NICs (vNICs) and enabled on the uplink ports. The Infrastructure NFS (3150) and In Band Management (363) VLANs are assigned to a pair of vNICs also enabled on the uplink ports. Note that any tenant-based iSCSI VLANs would be assigned to the iSCSI-A and iSCSI-B vNICs.

In an ACI based configuration, when the VMware vDS is being used, the Cisco APIC connects to VMware vCenter and automatically configures port-groups on the vDS based on the user-defined End Point Group (EPG) configuration. These port-groups are associated with a dynamically assigned VLAN from a pre-defined pool in Cisco APIC. Since Cisco APIC does not configure the Cisco UCS Fabric Interconnect, this range of pool VLANs has to be pre-configured on the uplink vNIC interfaces of the ESXi service profiles. In 0, VLAN 1101-1120 is part of the APIC defined VLAN pool.

If the Cisco AVS in VXLAN mode is used, Cisco APIC connects to VMware vCenter and also automatically configures port-groups on the AVS based on the user-defined EPG configurations. The Cisco AVS is treated as a vendor specific distributed virtual switch (DVS). These port-groups are associated with VXLANs assigned by Cisco APIC. Each ESXi host that is a member of the AVS has one VXLAN tunnel end point (VTEP) vmkernel (VMK) port for each AVS network uplink on the host. For all AVS network traffic that leaves the ESXi host, the VTEPs enscapsulate and send the traffic using the ACI fabric system VLAN (4093) in Figure 18. The ACI fabric system VLAN is the only VLAN that has to be extended from the ACI fabric through the Cisco UCS fabric interconnects to the ESXi hosts for the Cisco AVS.

Figure 18 VLAN Configuration for Cisco UCS Connectivity

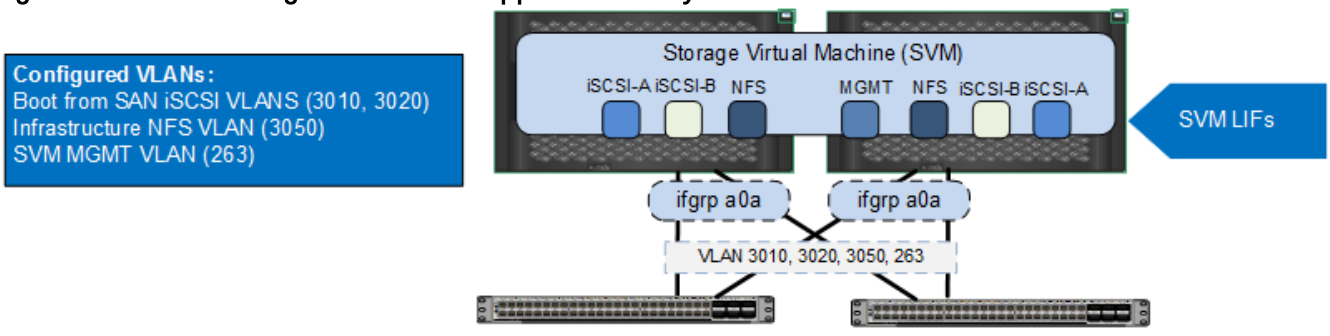


VLAN Configuration for NetApp

When configuring NetApp controllers for Cisco Nexus 9000 connectivity, iSCSI VLANs used for boot from SAN and application data LUN access, NFS VLANs for the ESXi host datastore access and SVM management LIFs are defined on the NetApp controllers. In Figure 19, VLANs 3050, 3010 and 3020 are the NFS, iSCSI-A and iSCSI-B VLANs for the infrastructure tenant. VLAN 263 is the SVM management interface.

Note: Currently in ACI, a VLAN can only be associated with a single physical domain when the L2 VLAN Scope is Global instead of Port Local, therefore the NFS and iSCSI VLAN IDs used on Cisco Unified Computing System and NetApp controllers are different. In 0 and Figure 19, VLANs 3150, 3110, 3120 and 363 are defined on Cisco Unified Computing System whereas VLANs 3050, 3010, 3020 and 263 are defined on NetApp controllers for the same storage path. The ACI fabric provides the necessary VLAN translation to enable communication between the VMkernel and the LIF EPGs. For additional information about EPGs and VLAN mapping, refer to the Application Centric Infrastructure (ACI) Design section.

Figure 19 VLAN Configuration for NetApp Connectivity



Application Centric Infrastructure (ACI) Design

The Cisco 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 ACI to provide advanced traffic optimization, security, and telemetry functions for both virtual and physical workloads. The Cisco ACI fabric is deployed in a leaf-spine architecture. The network provisioning in an

ACI-based FlexPod is quite different from traditional FlexPod and requires a basic knowledge of some of the core concepts of ACI.

ACI Components

Leaf switches: The ACI leaf provides physical server and storage connectivity as well as enforces ACI policies. A leaf typically is a fixed form factor switch such as the Cisco Nexus N9K-C9396PX, the N9K-C9372PX and N9K-C93180YC-EX switches. 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 FlexPod with ACI design, Cisco UCS Fabric Interconnect, NetApp Controllers and WAN/Enterprise routers are connected to both the leaves for high availability.

Spine switches: The ACI spine provides the mapping database function and connectivity among leaf switches. A spine can be the Cisco Nexus® N9K-C9508 switch equipped with N9K-X9736PQ line cards or fixed form-factor switches such as the Cisco Nexus N9K-C9336PQ ACI spine switch. Spine switches provide high-density 40 Gigabit Ethernet connectivity between leaf switches.

Tenant: A tenant (0) is a logical container or a folder for application policies. This container can represent an actual tenant, an organization, an application or can just be used for the convenience of organizing information. A tenant represents a unit of isolation from a policy perspective. All application configurations in Cisco ACI are part of a tenant. Within a tenant, you define one or more Layer 3 networks (VRF instances), one or more bridge domains per network, and EPGs to divide the bridge domains.

FlexPod with ACI design requires creation of a tenant called "Foundation" for providing compute to storage connectivity to setup the boot from SAN environment as well as for accessing the Infrastructure datastores using NFS. The design also utilizes the predefined "common" tenant to host core services (such as DNS, AD etc.) required by all the tenants. In most cases, each subsequent application deployment will require creation of a dedicated tenant.

Application Profile: Modern applications contain multiple components. For example, an e-commerce application could require a web server, a database server, data located in a storage area network, and access to outside resources that enable financial transactions. An application profile (0) models application requirements and contains as many (or as few) End Point Groups (EPGs) as necessary that are logically related to providing the capabilities of an application. Depending on the tenant requirements, in the FlexPod with ACI design, an application profile will be used to define a multi-tier application (such as Microsoft SharePoint) as well as to define storage connectivity using different storage protocols (NFS and iSCSI).

Bridge Domain: A bridge domain represents a L2 forwarding construct within the fabric. One or more EPGs can be associated with one bridge domain or subnet. A bridge domain can have one or more subnets associated with it. One or more bridge domains together form a tenant network.

For FlexPod design, setting up a bridge domain is an important consideration. A bridge domain in ACI is equivalent to a broadcast layer-2 domain in traditional Ethernet networks. When a bridge domain contains endpoints belonging to different VLANs (outside of the ACI fabric), a unique MAC address is required for every unique endpoint. NetApp storage controllers, however, use the same MAC address for an interface group and all the VLAN interface ports defined for that interface group on that storage node. As a result, all the LIFs on a NetApp interface group end up sharing a single MAC address even though these LIFs belong to different VLANs.

```
a01-aff8040::> network port show -fields mac
```

```
node           port        mac
-------------- ----        -----------------
a01-aff8040-01 a0a         02:a0:98:5b:46:fe
a01-aff8040-01 a0a-263     02:a0:98:5b:46:fe (MGMT)
a01-aff8040-01 a0a-3010    02:a0:98:5b:46:fe (iSCSI-A)
a01-aff8040-01 a0a-3020    02:a0:98:5b:46:fe (iSCSI-B)
a01-aff8040-01 a0a-3050    02:a0:98:5b:46:fe (NFS)
```

To overcome potential issues caused by overlapping MAC addresses, multiple bridge domains need to be deployed for correct storage connectivity. The details of the required bridge domains are covered in the design section below.

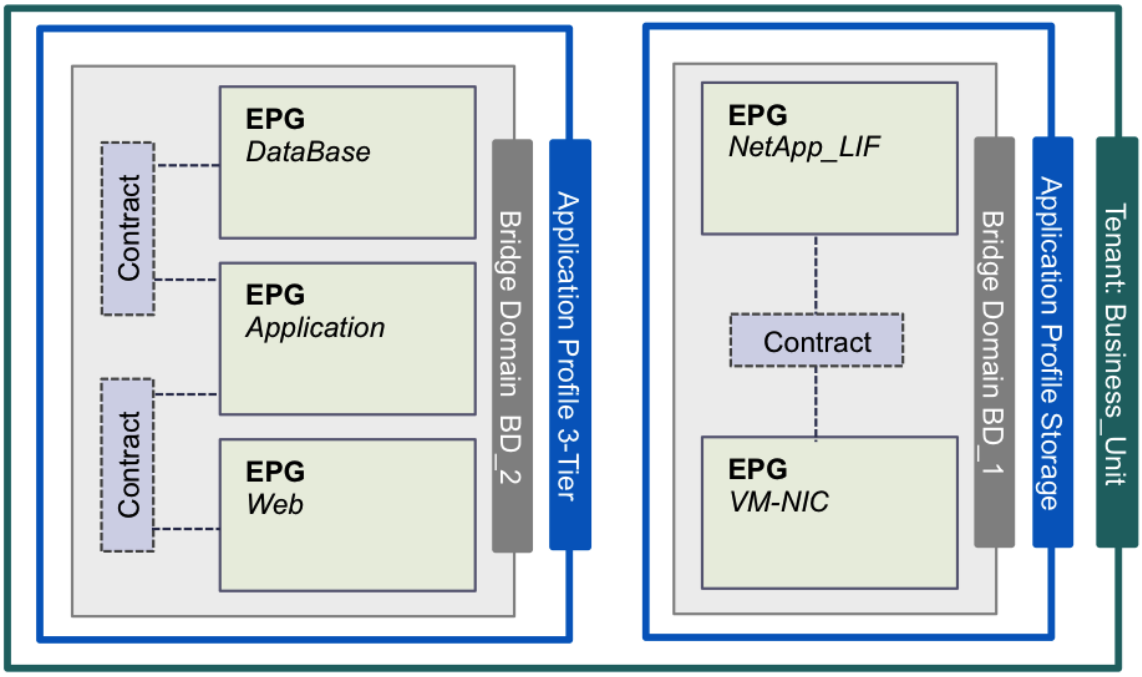
End Point Group (EPG): An End Point Group (EPG) is a collection of physical and/or virtual end points that require common services and policies. An End Point Group example is a set of servers or storage LIFs on a common VLAN providing a common application function or service. While the scope of an EPG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

In the FlexPod design, various application tiers, ESXi VMkernel ports for iSCSI, NFS and vMotion connectivity, and NetApp LIFs for SVM-Management and NFS and iSCSI datastores are placed in separate EPGs. The design details are covered in the following sections.

Contracts: A service contract can exist between two or more participating peer entities, such as two applications running and talking to each other behind different endpoint groups, or between providers and consumers, such as a DNS contract between a provider entity and a consumer entity. Contracts utilize filters to limit the traffic between the applications to certain ports and protocols.

Figure 20 covers the relationship between the ACI elements defined above. 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. 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 bridge domains and multiple application profiles and end point groups can utilize the same bridge domain.

Figure 20 ACI—Relationship between Major Components



End Point Group (EPG) Mapping in a FlexPod Environment

In the FlexPod with ACI infrastructure, traffic is associated with an EPG in one of the three following ways.

- Statically mapping a VLAN to an EPG (Figure 21)
- Associating an EPG with a Virtual Machine Manager (VMM) domain for VMware vDS and allocating a VLAN dynamically from a pre-defined pool in APIC (Figure 22)
- Associating an EPG with a Virtual Machine Manager (VMM) domain for Cisco AVS and APIC allocating a VXLAN (Figure 23).

Figure 21 EPG—Static Binding to a Path

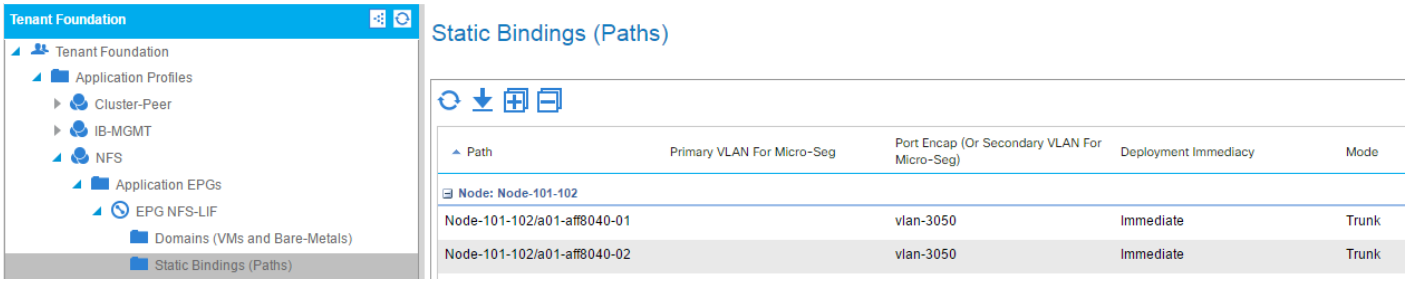


Figure 22 EPG–Virtual Machine Manager Domain Binding

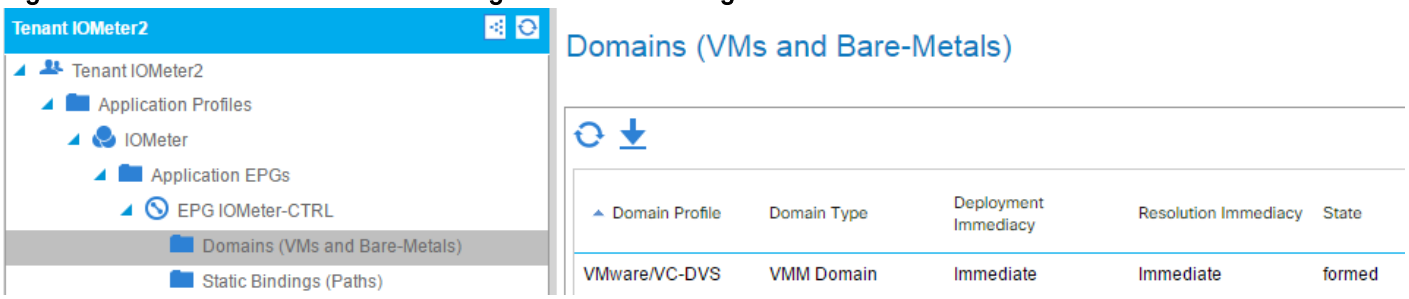
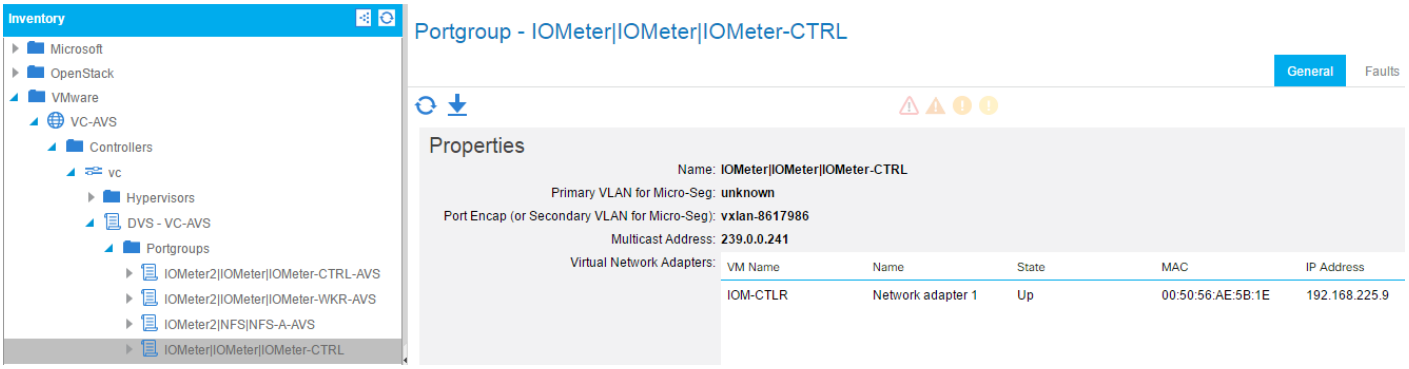


Figure 23 EPG–APIC Virtual Machine Manager VXLAN Assignment



The first method of statically mapping a VLAN is useful for the following:

- Mapping storage VLANs on the NetApp Controllers to storage protocol-related EPGs. These storage EPGs become the storage "providers" and are accessed by the ESXi host either by being in the same EPG or separate EPGs through contracts as shown in Figure 20.
- Connecting an ACI environment to an existing layer-2 bridge domain, such as an existing management segment. A VLAN on an out of band management switch is statically mapped to a management EPG in the common tenant to provide management services to VMs across all the tenants.
- Mapping iSCSI and NFS datastores VLANs on Cisco UCS to EPGs that consume the NetApp storage EPGs. 0 illustrates this mapping.
- Mapping the VMware vMotion VLAN on Cisco UCS to an EPG as shown in Figure 24.
- Mapping tenant iSCSI for tenant VMs or ESXi servers to access iSCSI LUNs on storage.

Note: iSCSI VLANs mapped to VMware ESXi hosts should not be mapped on either the VMware vDS or Cisco AVS, because the mapped storage will not be automatically available on ESXi reboot. It is fine to map iSCSI VLANs directly to VMs using either the VMware vDS or Cisco AVS.

The second method of dynamically mapping a VLAN/VXLAN to an EPG by defining a VMM domain is used for the following:

- Deploying VMs in a multi-tier Application as shown in Figure 28
- Deploying NFS related storage access for the application Tenant as shown in Figure 28

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 configuration of the VM switching as detailed in the following sections.

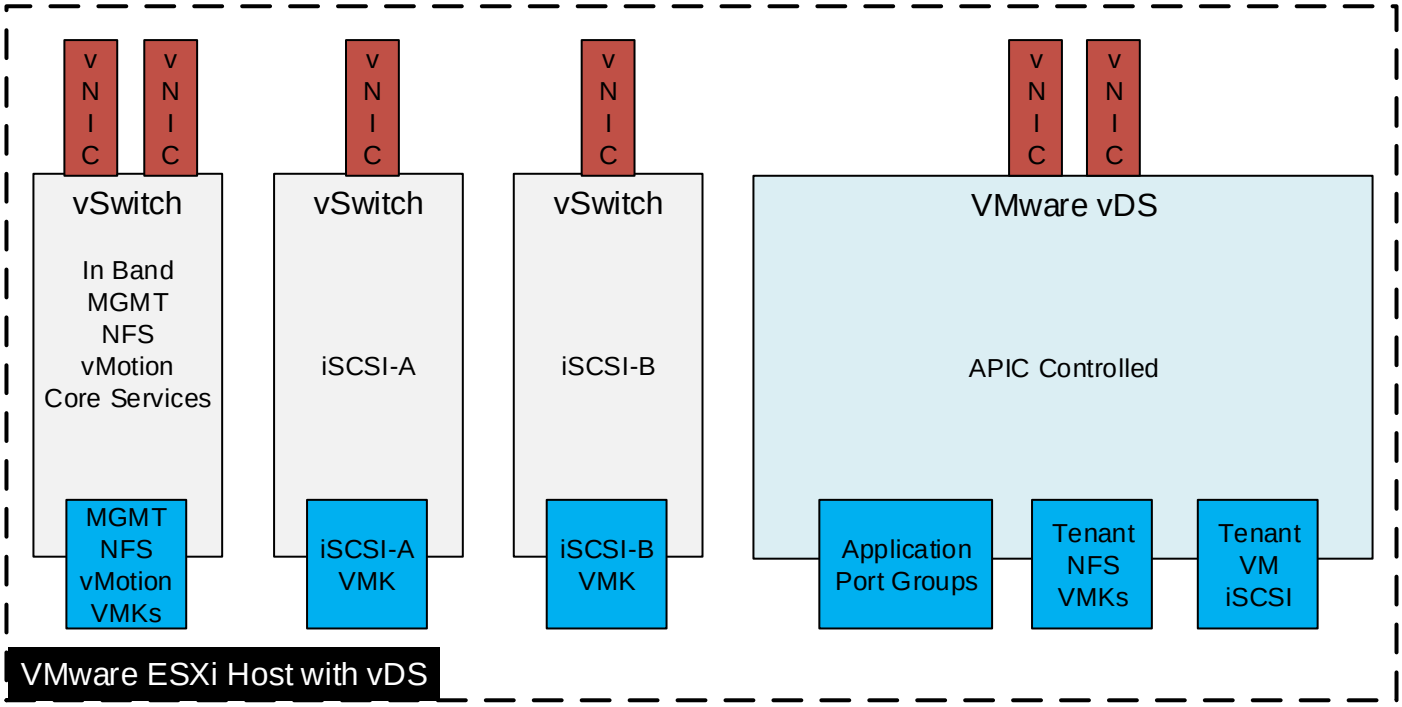
Virtual Machine Manager (VMM) Domains

For a VMware vCenter, the creation and network configuration of the VMware vSphere Distributed Switch (vDS) or Cisco AVS and the set up of port groups are performed by the APIC. The APIC communicates with the vDS to publish network policies that are applied to the virtual workloads. To position an application, the application administrator deploys the VMs and places the VM NIC(s) into the appropriate port group(s) for various application tiers. A VMM domain contains multiple EPGs and hence multiple port groups.

vSwitch and Distributed Virtual Switch (DVS)

While a tenant application deployment utilizes port groups on an APIC controlled DVS, some of the core functionality such as in-band management access, and iSCSI access utilizes vSphere vSwitches. The resulting distribution of VMkernel ports and VM port-groups on an ESXi server with the VMware vDS is shown in Figure 24. In the Cisco UCS service profile for ESXi hosts, storage, management and VM data VLANs are defined on the vNIC interfaces used by appropriate vSwitches. Note that it is important that Infrastructure NFS be placed on the VMware vSwitch along with management and vMotion. On servers where VMware vCenter is running, it is critical that Infrastructure NFS is placed on this vSwitch and not on a DVS to prevent loss of access to vCenter in the event of a host reboot. It is also important to place all iSCSI VMK ports on the iSCSI vSwitches to prevent loss of mapped LUNs in the event of a host reboot.

Figure 24 ESXi Host Server Design with VMware vDS

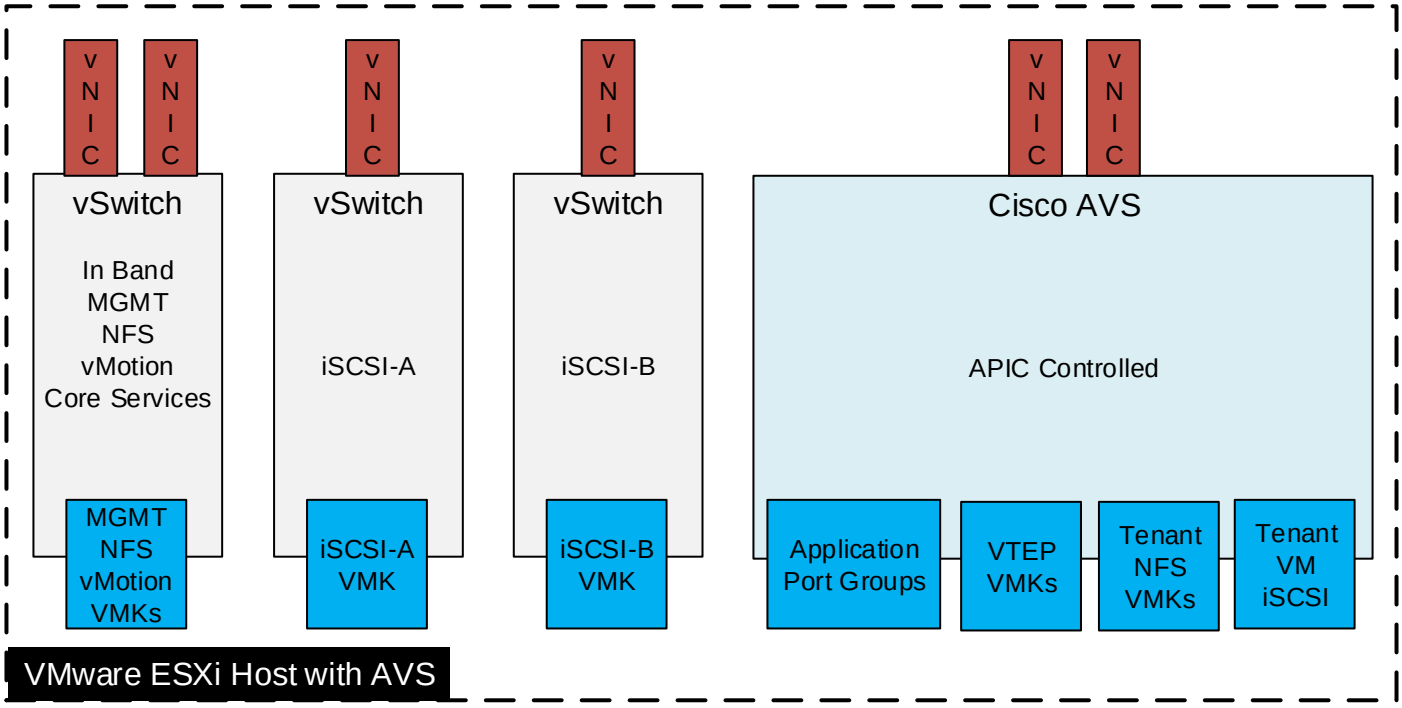


When the Cisco AVS in VXLAN mode is used as the DVS instead of the VMware vDS, one VTEP VMK port per vNIC uplink is provisioned on the AVS as shown in Figure 25. The VTEPs are used in the VXLAN encapsulation process. All other rules used for setting up vNICs and VMK ports on the ESXi Hosts with the

VMware vDS apply on the ESXi Hosts with the Cisco AVS. It is also possible to have both a VMware vDS and a Cisco AVS on the same ESXi host.

When using Cisco AVS in VXLAN mode with Jumbo Frames, the NetApp LIF, VMware VMK port, and VM NIC port MTUs that would normally be set to 9000 need to be set to 8950 to account for the extra 50 bytes in the VXLAN encapsulation.

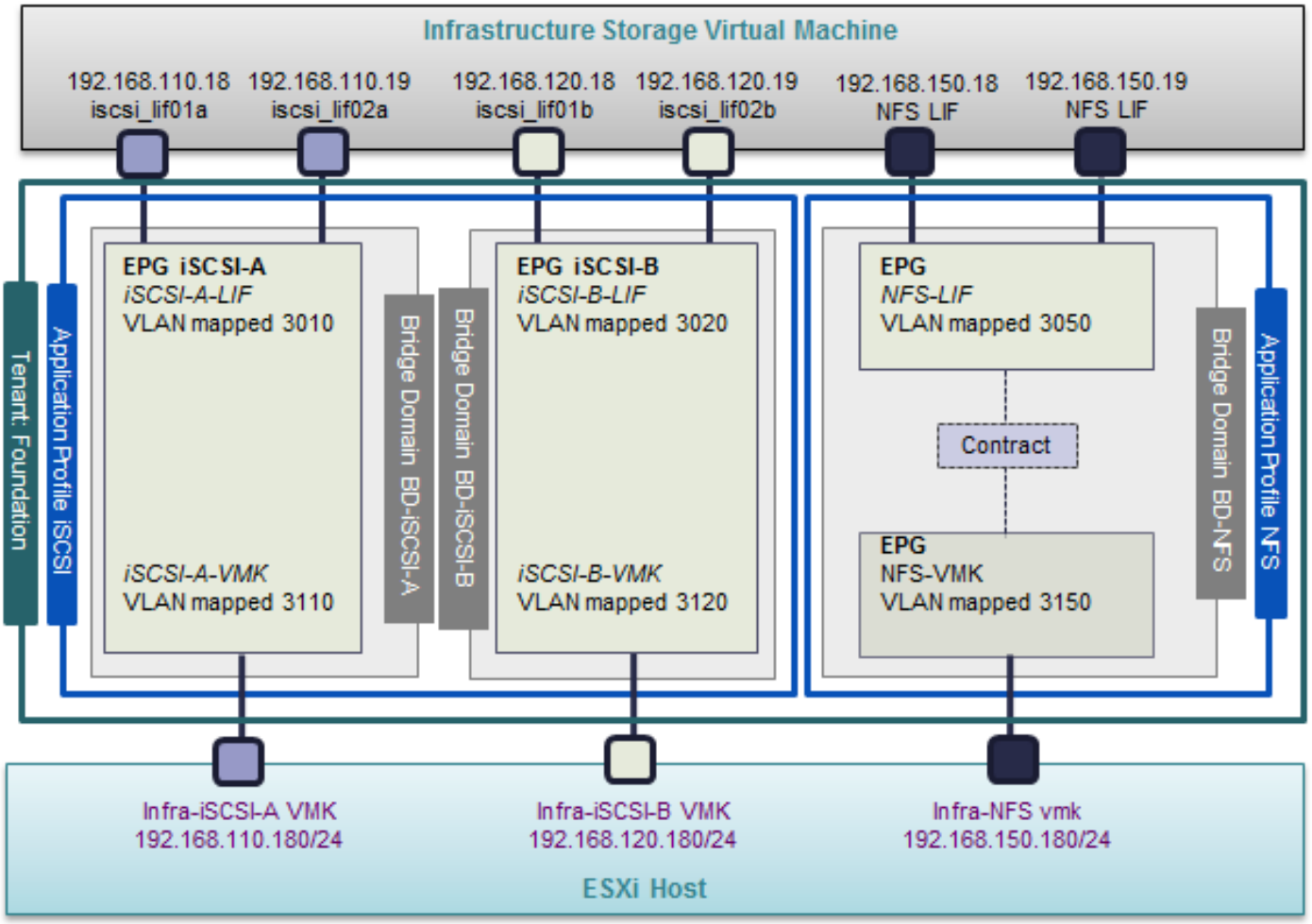
Figure 25 Esxi Host Server Design with AVS



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 ACI constructs for core infrastructure services including an overview of the connectivity and relationship between various ACI elements is covered in Figure 26.

Figure 26 Design Details of the Foundation (Infrastructure) Tenant



ACI Design for Foundation Tenant

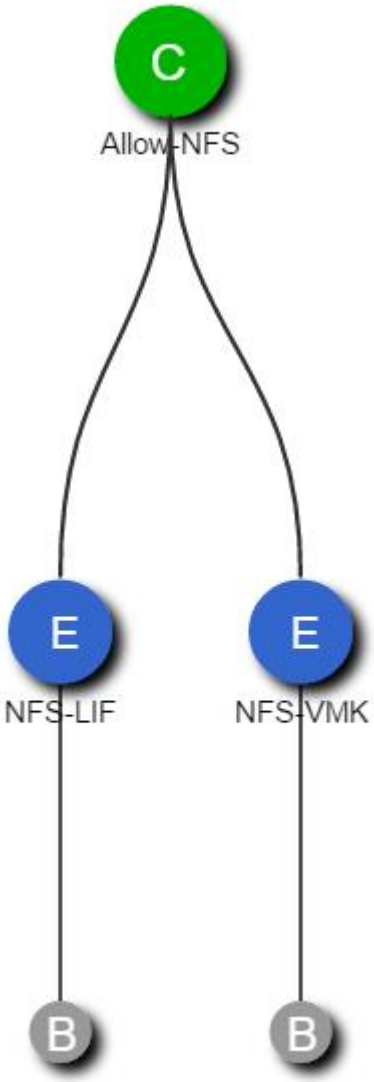
- Tenant Role: To enable the compute to storage connectivity for accessing iSCSI boot LUNs and NFS datastores. The boot LUNs enable stateless compute functionality while the NFS datastore hosts all the Infrastructure VMs.
- VRF: Each tenant in this implementation was assigned a separate Virtual Routing and Forwarding (VRF) instance, providing each tenant a separate routing table. Additionally, with application tenants, each tenant was assigned a separate Storage Virtual Machine (SVM) with its own IPspace in the NetApp Storage, allowing tenants the capability of using overlapping IP address spaces. The Foundation Tenant's storage SVM was assigned to the Default IPspace in the NetApp Storage.
- Application Profile, EPGs and Contracts: The foundation tenant comprises of four application profiles, "iSCSI", "NFS", IB-MGMT (In Band Management), and "vMotion".

IB-MGMT and vMotion are not shown in the diagrams for this section.

- Application Profile "NFS" comprises of two EPGs, "NFS-LIF" and "NFS-VMK" as shown in Figure 27.

- The first EPG and static path mapping maps the VLAN associated with the NFS LIF interfaces on the NetApp Infrastructure SVM (VLAN 3050). This mapping provides NFS storage access to the compute environment.
- The second EPG and static path mapping maps the VLAN associated with the NFS VMkernel port for the infrastructure ESXi host (VLAN 3150).
- An NFS contract connects these two EPGs, and can include a filter to only allow NFS traffic to pass.

Figure 27 Foundation Tenant–Application Profile NFS



Note: Each entry point into the ACI environment is mapped to a unique VLAN, because Global L2 VLANs are being used. Even though the VMkernel ports on the ESXi hosts and the NFS LIF interfaces on the NetApp SVM are part of the same layer-2 domain, two different VLANs (3050 and 3150) are configured for this EPG. Also, note that flexibility exists to map the two entry points into one EPG, or to map the two entry points into two EPGs and use a restrictive contract to connect them.

- Application Profile "iSCSI" is comprised of two EPGs, "iSCSI-A", "iSCSI-B.

- Static path mappings in the iSCSI-A EPG statically map the VLANs associated with iSCSI-A (VLAN 3010 and 3110) to the LIF interface on the NetApp Infrastructure SVM and the VMK interface in the Cisco UCS.
- Static path mappings in the iSCSI-B EPG statically map the VLANs associated with iSCSI-BA (VLAN 3020 and 3120) to the LIF interface on the NetApp Infrastructure SVM and the VMK interface in the Cisco UCS.

Since the two static mapped endpoints for the two iSCSI networks are in the same EPGs, access between the iSCSI LIFs and corresponding VMK port is unrestricted.

- Bridge Domains: While all the EPGs in a tenant can theoretically share the same bridge domain, overlapping MAC address usage by NetApp storage controllers on the interface groups across multiple VLANs determines the actual number of bridge domains required. As shown in Figure 26, the "Foundation" tenant connects to two iSCSI LIFs and one NFS LIF to provide storage connectivity to the infrastructure SVM. Since these three LIFs share the same MAC address, a separate BD is required for each LIF. The "Foundation" tenant therefore comprises of four bridge domains: BD-iSCSI-A, BD-iSCSI-B, BD-NFS, and BD-Internal.
  - BD-iSCSI-A is the bridge domain configured to host EPGs for iSCSI-A traffic
  - BD-iSCSI-B is the bridge domain configured to host EPGs for iSCSI-B traffic
  - BD-NFS is the bridge domain configured to host EPGs for NFS traffic
  - BD-Internal is the bridge domain configured to host EPGs for all other Tenant Foundation traffic. This bridge domain is also utilized for hosting EPGs related to vMotion and application traffic since there is no MAC address overlap with these functions

Onboarding a 3-Tier Application

The ACI constructs for a 3-tier application deployment are a little more involved than the infrastructure tenant "Foundation" covered in the last section. In addition to providing ESXi to storage connectivity, various tiers of the application also need to communicate amongst themselves as well as with the storage and common services (DNS, AD etc.). Figure 28 provides an overview of the connectivity and relationship between various ACI elements for a sample 3-tier Application.

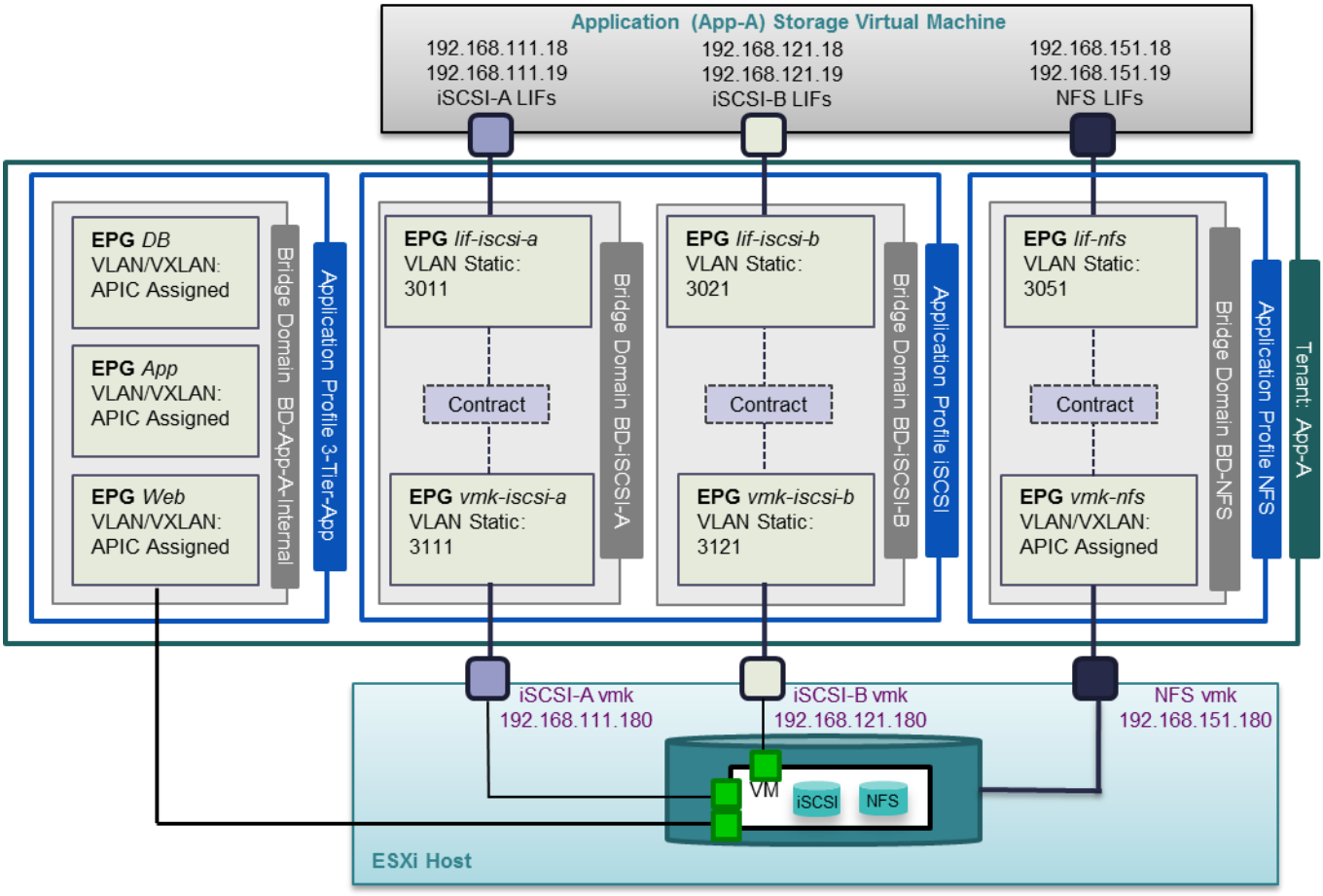
Some of the key highlights of the sample 3-Tier Application deployment are as follows:

- Four application profiles, NFS, iSCSI, SVM-MGMT and 3-Tier-App are utilized to deploy the application. SVM-MGMT is not shown in Figure 28.
- ESXi servers will map an NFS datastore from a dedicated Application SVM on NetApp controllers. This datastore hosts all the application VMs.
- The VMkernel port-group for mounting NFS datastores is managed and deployed by APIC on either the VMware vDS or Cisco AVS.
- To provide ESXi host-based access to iSCSI storage LUNs either through iSCSI datastores or Raw Device Mapped (RDM) LUNs, two iSCSI VMK port-groups are deployed in the appropriate VLANs on

the two iSCSI vSwitches on the ESXi host. This deployment is necessary to prevent the iSCSI mappings from not coming back up after an ESXi reboot.

- To provide VMs a direct access to storage LUNs using a software iSCSI iniator, two iSCSI port-groups can be deployed using APIC on either the VMware vDS or Cisco AVS. These port groups can also be deployed on the iSCSI vSwitches on the ESXi host.
- The SVM-MGMT EPG can be tied by contract to any of the three application tiers where access to the SVM for storage provisioning and backup is necessary.
- Four unique bridge domains are needed to host iSCSI, NFS and VM traffic.

Figure 28 Design Details of the 3-Tier Application



ACI Design for 3-Tier Application Tenant

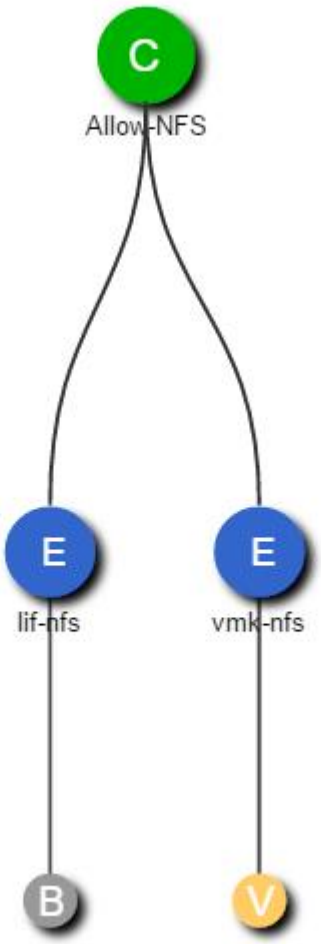
- Tenant Role: To host a multi-tier application (App-A in this design) and to provide application specific compute to storage connectivity, a tenant named "App-A" is configured.
- VRF: Each tenant in this implementation was assigned a separate Virtual Routing and Forwarding (VRF) instance, providing each tenant a separate routing table. Additionally, with application tenants, each tenant was assigned a separate Storage Virtual Machine (SVM) with its own IPspace in the NetApp Storage, allowing tenants the capability of using overlapping IP address spaces. Two tenants were set up in the lab with separate VRFs and IPspaces. Although different VLANs/VXLANs were used for each tenant, the same IP subnets were used for all storage connections, i.e. 192.168.111.0/24 for iSCSI-A

in both tenants. A separate VMare ESXi host cluster had to be used for each tenant and each datastore had a unique name.

- Application Profile and EPGs: The "App-A" tenant comprises of four application profiles, "3-Tier-App", "iSCSI", "NFS", and “SVM-MGMT”.
- Application Profile "NFS" comprises of two EPGs, "lif-NFS" and "vmk-NFS" as shown in Figure 29.
  - EPG "lif-nfs" statically maps the VLAN associated with NFS LIF on the App-A SVM (VLAN 3051). This EPG "provides" NFS storage access to the tenant environment.
  - EPG "vmk-nfs" is attached to the VMM domain to provide an NFS port-group in the vSphere environment. This port-group is utilized by the tenant (App-A) ESXi servers.

A contract "Allow-NFS" is defined to allow NFS traffic. This contract is "Provided" by EPG lif-nfs and is "Consumed" by EPG vmk-nfs.

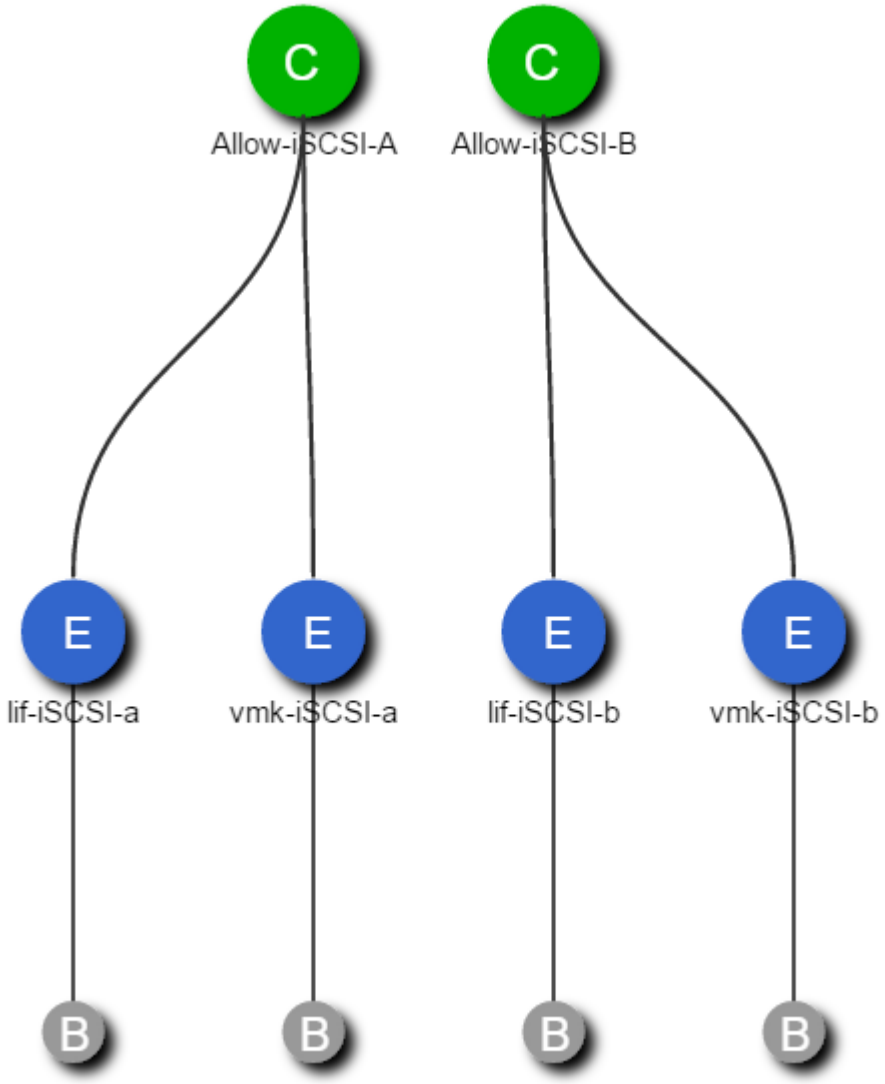
Figure 29 App-A–Application Profile NFS



- Application Profile "iSCSI" is comprised of four EPGs, "lif-iSCSI-a", "lif-iSCSI-b", "vmk-iSCSI-a" and "vmk-iSCSI-b" as shown in Figure 30.
  - EPGs "lif-iSCSI-a" and "lif-iSCSI-b" statically maps the VLANs associated with iSCSI-A and iSCSI-B LIF interfaces on the NetApp Infrastructure SVM (VLAN 3011 and 3021). These EPGs "provide" LUN access to VMs.
  - EPGs "vmk-iSCSI-a" and "vmk-iSCSI-b" are statically mapped to the Cisco UCS vPCs (VLAN 3111 and 3121) and provide iSCSI port-groups on the iSCSI vSwitches. These port-groups are utilized by VMs that require direct raw device access to storage LUNs.

  Two contracts "Allow-iSCSI-A" and "Allow-iSCSI-B" are defined to allow iSCSI traffic. These contracts are"Provided" by EPGs lif-iSCSI-a and lif-iSCSI-b and are "Consumed" by EPGs vmk-iSCSI-a and vmk-iSCSI-b.

Figure 30 App-A–Application Profile iSCSI



- Application Profile "3-Tier-App" comprises of four EPGs, "Web", "App", "DB" and "External"

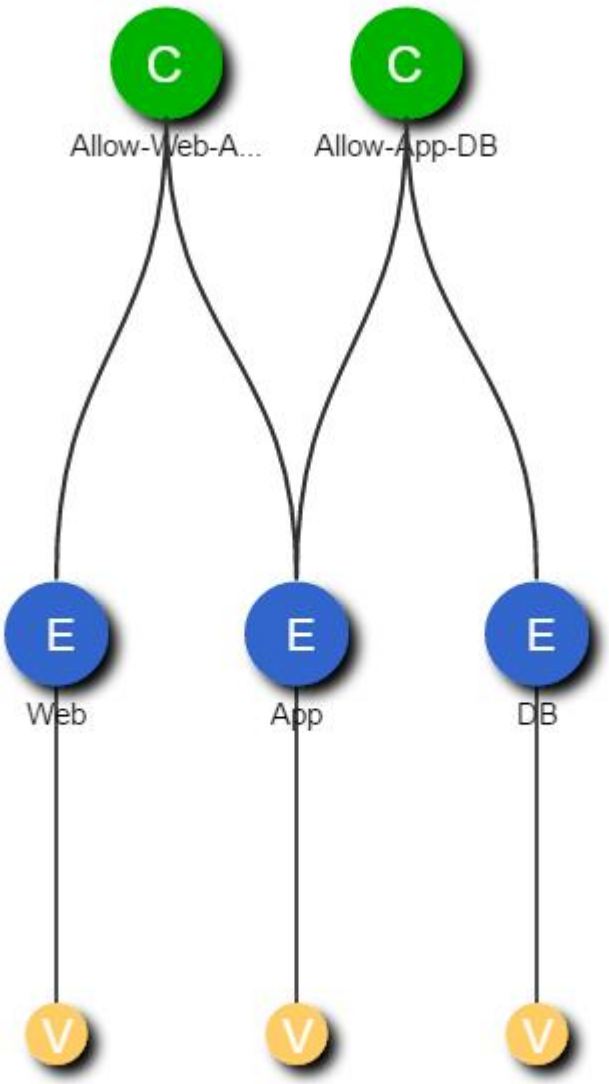
— EPG "Web" is attached to the VMM domain and provides a port-group on VDS to connect the web servers.

— EPG "App" is attached to the VMM domain and provides a port-group on VDS to connect the application servers.

— EPG "DB" is attached to the VMM domain and provides a port-group on VDS to connect the database servers.

— Appropriate contracts are defined to allow traffic between various application tiers.

Figure 31 App-A–3-Tier-App Application Profile



- Bridge Domain: The "App-A" tenant comprises of four bridge domains, BD-iSCSI-A, BD-iSCSI-B, BD-NFS, and BD-Internal. As explained before, overlapping MAC addresses on NetApp Controllers require iSCSI-A, iSCSI-B and NFS traffic to use separate bridge domains.

— BD-iSCSI-A is the bridge domain configured to host EPGs configured for iSCSI-A traffic

- BD-iSCSI-B is the bridge domain configured to host EPGs configured for iSCSI-B traffic
- BD-NFS is the bridge domain configured to host EPGs configured for NFS traffic
- BD-Internal is the bridge domain configured for hosting EPGs related to SVM-MGMT and application traffic since there is no MAC address overlap with the application VMs

Core Services and Storage Management

Accessing Core Services

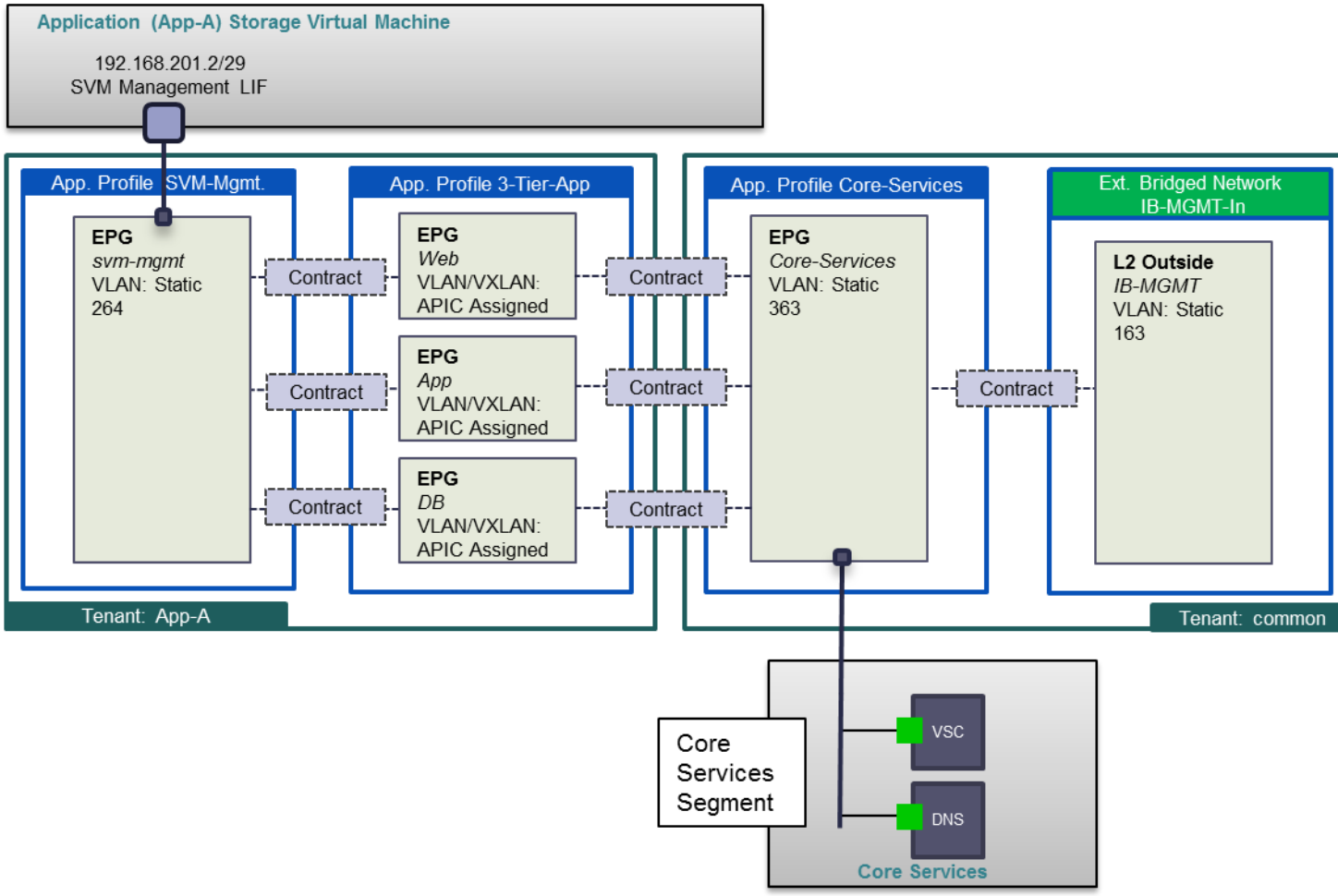
To provide application servers access to common services such as Active Directory (AD), Domain Name Services (DNS), management and monitoring software etc., inter-tenant contracts are utilized. The Cisco ACI fabric provides a predefined tenant named "common" to host the shared services. The policies defined in the "common" tenant are usable by all the other tenants. In addition to the locally defined contracts, all the tenants in the ACI fabric have access to the contracts defined in the "common" tenant.

In the FlexPod environment, access to core services is provided as shown in Figure 32. To provide this access:

- A common services segment is defined where core services VMs connect. The core services port-group is defined on the Infrastructure vSwitch on the VMware ESXi hosts. A separate services segment ensures that the access from the tenant VMs is limited to only core services' VMs
- The EPG for the core services segment "Core-Services" is defined in the "common" tenant
- The tenant VMs access the core services segment by consuming contracts from the "common" tenant
- The contract filters can be configured to only allow specific services related ports
- The tenant VMs access the core services segment using their EPG gateway.
- Since the tenant VMs reside in separate subnets than the Core-Services VMs, routes must be configured in the Core-Services VMs to reach the Application tenant VMs. For this lab implementation a supernet route with destination 172.16.0.0/16 was put into each Core-Services VM. Routes needed to be shared across VRFs since the two EPGs were in different tenants.
- Unique IP subnets have to be used for each EPG connecting to Core-Services.

Figure 32 shows both "provider" EPG "Core-Services" in the tenant "common" and the consumer EPGs "Web", "App" and "DB" in tenant "App-A".

Figure 32 Core Services and Storage Management



Accessing SVM Management

Some applications such as NetApp Snap Drive require direct connectivity from the application (SharePoint, Exchange etc.) VMs to the management LIF on the tenant SVM. To provide this connectivity securely, a separate VLAN is dedicated for each tenant to define the management LIF. This VLAN is then statically mapped to an EPG in the application tenant as shown in Figure 32. Application VMs can access this LIF by defining and utilizing contracts.

Note: When an application tenant contains mappings for NetApp LIFs for storage access (iSCSI, NFS etc.), a separate bridge domain is required for the SVM management LIF because of the overlapping MAC addresses. Make sure that only one type of storage VLAN interface is accessed using a given bridge domain.

NetApp SnapManager and SnapDrive with Cisco ACI

NetApp SnapDrive and the SnapManager portfolio of products greatly simplify storage provisioning and the backup of application data. In this design, the SVM management LIF is placed on a VLAN within the application tenant and a contract is built linking the application VM's management interface with the SVM management LIF. This interaction takes place through HTTPS on TCP port 443 by default.

The interaction of SnapDrive and the SVM management LIF also handle LUN provisioning. If VMware RDM LUNs are being used for the application data, SnapDrive must interact with VMware vCenter to perform the RDM LUN mapping. The SnapDrive interaction with NetApp VSC handles the Snapshot copy management of

application VMDK disks on NFS or VMFS datastores. In this design, the VMware vCenter and VSC VM network interfaces are placed in the ACI "common" tenant in the Core-Services EPG. Application VMs from multiple tenants can then consume contracts to access the vCenter and VSC VMs simultaneously while not allowing any communication between tenant VMs. The vCenter interaction takes place through HTTPS on TCP port 443 by default, while the VSC interaction takes place on TCP port 8043 by default. Specific contracts with only these TCP ports can be configured.

These ACI capabilities are combined with the Role-Based Access Control (RBAC) capabilities of vCenter, NetApp VSC, and NetApp clustered Data ONTAP to allow multiple-tenant administrators and individual-application administrators to simultaneously and securely provision and back up application data storage while taking advantage of the NetApp storage efficiency capabilities.

FlexPod Connectivity to Existing Infrastructure (Shared Layer 3 Out)

In order to connect the ACI fabric to existing infrastructure, the leaf nodes are connected to a pair of core infrastructure routers/switches. In this design, a Cisco Nexus 7000 was configured as the core router. Figure 33 shows the connectivity details from the Shared_L3_Out External Routed Domain in the “common” tenant and the “common/default” VRF. Figure 34 shows how tenants with other VRFs are connected to the Shared_L3_Out via contracts. Tenant network routes can be shared with the Cisco Nexus 7000s using OSPF and external routes from the Cisco Nexus 7000s can be shared with the tenant VRFs. Routes can also be shared across VRFs.

Figure 33 ACI Connectivity to Existing Infrastructure

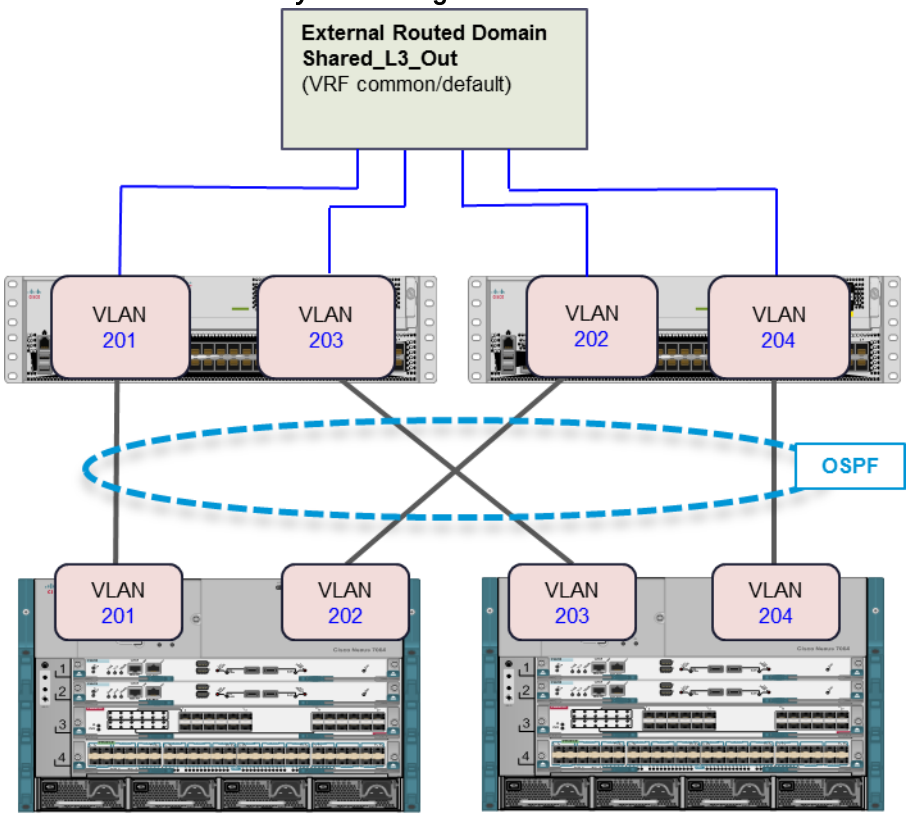
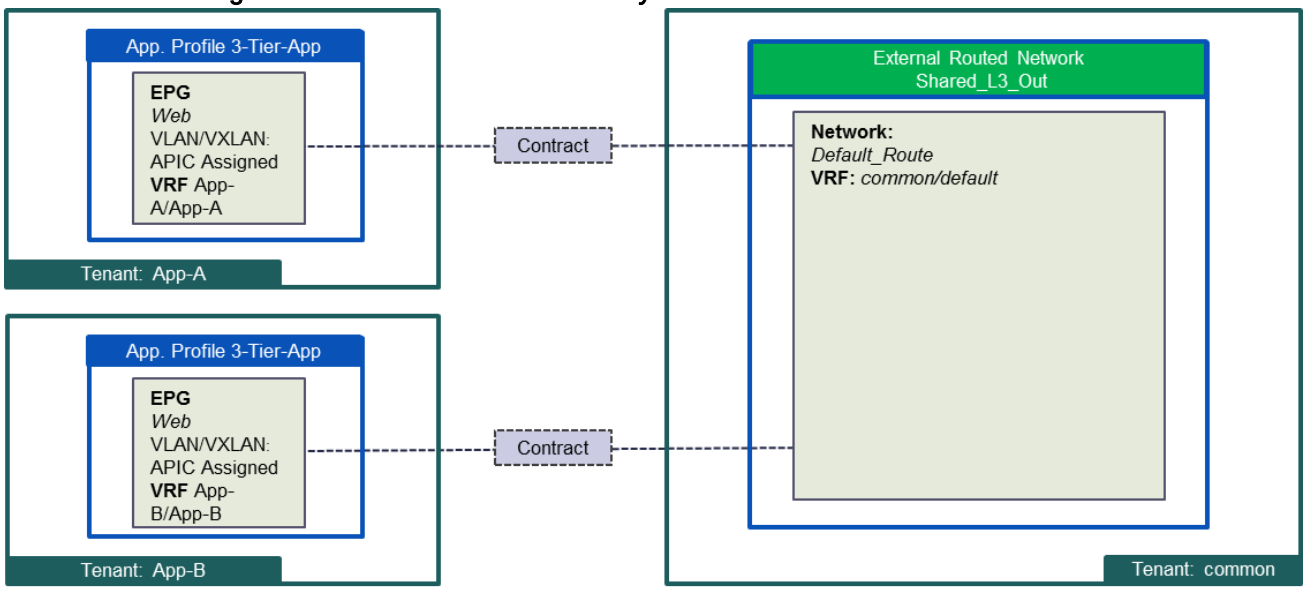


Figure 34 Connecting Tenant Networks to Shared Layer 3 Out



Some of the design principles for external connectivity are as follows:

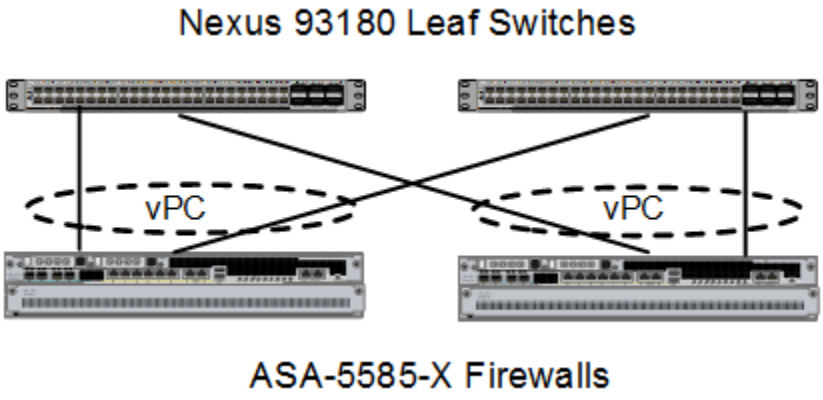
- Each Leaf switch is connected to both Cisco Nexus 7000 switches for redundancy.
- A unique VRF is defined for every tenant. App-A/App-A and App-B/App-B are two such VRFs shown in Figure 34.
- Unique VLANs are configured to provide multi-path connectivity between the ACI fabric and the core infrastructure. VLANs 201-204 are configured as shown.
- On the ACI fabric, OSPF is configured to share routes. The ACI fabric learns a default route from the core router and each tenant advertises one or more "public" routable subnets to the core infrastructure across VRFs.

L4-L7 Services VLAN Stitching

ACI Release1.2 introduced the concept of L4-L7 Services VLAN Stitching. In previous ACI releases, L4-L7 services were provided with Device Packages where supported services devices were configured and controlled by the APIC. With the introduction of VLAN Stitching the services device is configured and controlled using the device user interface meaning virtually any services device can be used. The services device can be inserted into a contract using inside (provider) and outside (consumer) VLANs. A services device is interfaced to two leaves and can be inserted into a contract using a Service Graph between two EPGs within a tenant, or between an EPG and an External Network. If the services device supports multiple virtual devices or contexts, each virtual device can be placed into a Service Graph within a tenant, giving each tenant its own unique virtual device.

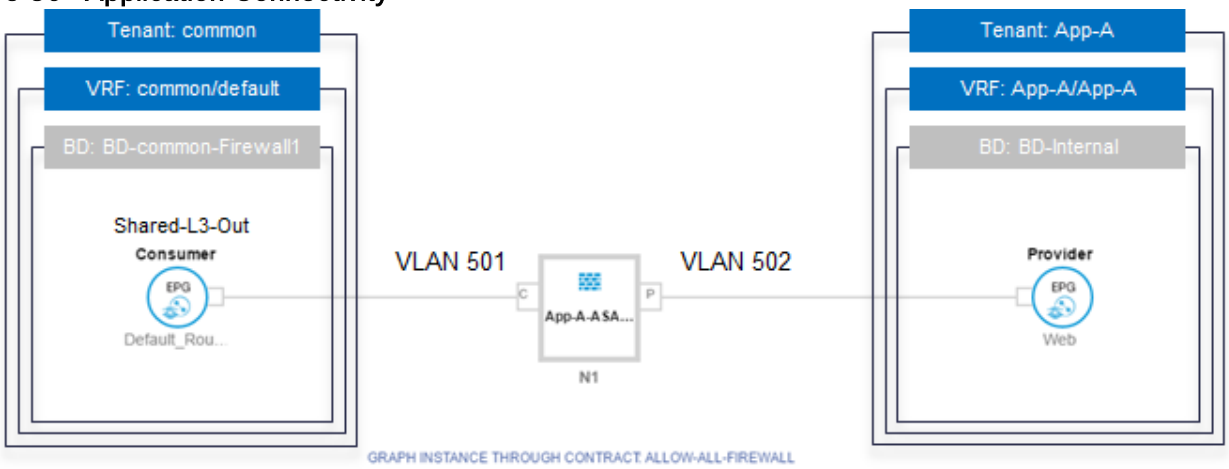
In this FlexPod with ACI environment, a pair of Cisco ASA-5585-X firewalls in a high availability configuration were used to validate L4-L7 Services VLAN Stitching. Each firewall was connected with 10 GE to two leaf switches with a vPC as shown in Figure 35. The inside and outside firewall interfaces were then configured and multiple ASA contexts were used to assign virtual firewalls to multiple tenants.

Figure 35 ASA Firewall Deployment in FlexPod with ACI



When a services device is deployed based on the design described, the resulting configuration looks like Figure 36. In this example, a Cisco ASA firewall context is placed between the App-A tenant Web EPG and the Shared L3 Out Default_Route EPG, providing firewall services between the outside world and the App-A tenant's Web front-end.

Figure 36 Application Connectivity



Validation

Validation Testing

A high level summary of the validation done for the FlexPod with Cisco ACI design is provided in this section. The solution was built in the lab using the prescribed methods in the Deployment Guide. The solution level high availability components were tested using the IOMeter tool setup as a tenant in this architecture. IOMeter provided a load on the system while the validation tests were run. Examples of the types of tests executed are as follows:

- Failure and recovery of ACI Spines and Leaves.
- Failure and recovery of UCS Fabric Interconnects
- Failure and recovery of various network links in the Infrastructure
- Failure and recovery of redundant NetApp AFF Storage Controllers
- Failure and recovery of SSDs in the NetApp AFF Storage Controllers
- Failure and recovery of ESXi hosts in a host cluster
- Use of NetApp QOS to limit load to keep CPU utilization at a recommended level for Storage Controller failure and recovery tests
- Validation of Layer 4-7 Service VLAN Stitching with Multiple Tenants and Firewall Contexts
- Validation of Shared L3 Out with Multiple Tenants

The IOMeter traffic profile used for these tests was 64K size, 100% Random, 80% Read, 16 Outstanding IOs.

Minimum Hardware List for Validation

Table 2 lists a mimimum set of hardware needed to complete the lab validation. The solution can be scaled and many different models in the hardware product families are supported.

Table 2 Minimum Hardware List for Validation

Layer	Device	Quantitiy
Compute	Cisco UCS B200M4 with 128 GB RAM, 2 CPUs, and VIC 1340	4
	Cisco UCS C220M4 with 128 GB RAM, 2 CPUs, and VIC 1227	2
	Cisco UCS 5108AC2 Chassis	1
	Cisco UCS 2204XP FEX	2
	Cisco UCS 6248 Fabric Interconnect	2
Network	Cisco APIC-M1	3

Layer	Device	Quantitiy
	Cisco Nexus 9372PX Leaf	2
	Cisco Nexus 9336 Spine	2
Storage	NetApp AFF 8040	2
	NetApp DS2246 Disk Shelf with 24-400GB SSDs	2

Hardware and Software Revisions

Error! Reference source not found.describes the hardware and software versions used during solution validation. It is important to note that Cisco, NetApp, and VMware have interoperability matrixes that should be referenced to determine support for any specific implementation of FlexPod. Click the following links for more information:

- NetApp Interoperability Matrix Tool: http://support.netapp.com/matrix/
- Cisco UCS Hardware and Software Interoperability Tool: http://www.cisco.com/web/techdoc/ucs/interoperability/matrix/matrix.html
- VMware Compatibility Guide: http://www.vmware.com/resources/compatibility/search.php

Table 3 Validated Software Versions

Layer	Device	Image	Comments
Compute	Cisco UCS Fabric Interconnects 6200 Series, UCS B-200 M4, UCS C-220 M4	3.1(1h)	Includes the Cisco UCS-IOM 2208XP, Cisco UCS Manager, UCS VIC 1340 and UCS VIC 1227
	Cisco eNIC	2.3.0.7	
	Cisco fNIC	1.6.0.25	
Network	Cisco APIC	1.3(2f)	
	Cisco Nexus 9000 iNX-OS	11.3(2f)	
	Cisco Virtual Switch Update Manager (VSUM)	2.0	
	Cisco AVS	5.2(1)SV3(1.25)	
Storage	NetApp AFF 8040	Data ONTAP 8.3.2	
Software	VMware vSphere ESXi	6.0u1b	
	VMware vCenter	6.0u1b	
	OnCommand Unified Manager for clustered Data ONTAP	6.4	
	NetApp Virtual Storage Console (VSC)	6.2	

Layer	Device	Image	Comments
	OnCommand Performance Manager	2.1	

Summary

Conclusion

FlexPod with Cisco ACI is the optimal shared infrastructure foundation to deploy a variety of IT workloads. Cisco and NetApp have created a platform that is both flexible and scalable for multiple use cases and applications. From virtual desktop infrastructure to SAP®, FlexPod can efficiently and effectively support business-critical applications running simultaneously from the same shared infrastructure. The flexibility and scalability of FlexPod also enable customers to create a right-sized infrastructure that can grow with and adapt to their evolving business requirements.

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:

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

Cisco UCS C-Series Rack Servers:

https://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/products/switches/nexus-9000-series-switches/index.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:

http://www.vmware.com/products/vsphere/

NetApp Data ONTAP:

http://www.netapp.com/us/products/platform-os/ontap/index.aspx

NetApp All Flash FAS:

http://www.netapp.com/us/products/storage-systems/all-flash-fas/index.aspx

NetApp FAS8000:

http://www.netapp.com/us/products/storage-systems/fas8000/

NetApp OnCommand:

http://www.netapp.com/us/products/management-software/

NetApp VSC:

http://www.netapp.com/us/products/management-software/vsc/

NetApp SnapManager:

http://www.netapp.com/us/products/management-software/snapmanager/

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

NetApp Interoperability Matrix Tool:

http://support.netapp.com/matrix/mtx/login.do

About Authors

John George, Technical Marketing Engineer, Cisco UCS Data Center Solutions Engineering, Cisco Systems, Inc.

John George recently moved to Cisco from Netapp and is focused on designing, developing, validating, and supporting the FlexPod Converged Infrastructure. Before his roles with FlexPod, he supported and administered a large worldwide training network and VPN infrastructure. John holds a Master's degree in Computer Engineering from Clemson University.

Lindsey Street, Solutions Architect, Infrastructure and Cloud Engineering, NetApp

Lindsey Street is a Solutions Architect in the NetApp Infrastructure and Cloud Engineering team. She focuses on the architecture, implementation, compatibility, and security of innovative vendor technologies to develop competitive and high-performance end-to-end cloud solutions for customers. Lindsey started her career in 2006 at Nortel as an interoperability test engineer, testing customer equipment interoperability for certification. Lindsey has a Bachelors of Science degree in Computer Networking and her Masters of Science in Information Security from East Carolina University.

Acknowledgements

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

- Haseeb Niazi, Cisco Systems, Inc.
- Chris O' Brien, Cisco Systems, Inc.
- Ramesh Isaac, Cisco Systems, Inc.
- Melissa Palmer, NetApp
- Nabil Fares, NetApp