

FlashStack Virtual Server Infrastructure Design Guide for VMware vSphere 6.0 U2

Last Updated: November 12, 2017



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>.

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

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

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

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

© 2017 Cisco Systems, Inc. All rights reserved.

Table of Contents

About Cisco Validated Designs	2
Executive Summary.....	4
Solution Overview.....	5
Introduction	5
Audience	5
FlashStack Program Benefits.....	5
Technical Overview	7
FlashStack System Overview.....	7
Validated Components in the Architecture	7
Optional Components Included in the Validated Architecture	10
Alternative Components for the Architecture.....	11
Cisco UCS Differentiators	11
Pure Storage Differentiators.....	14
Solution Architecture.....	16
Design Principles.....	17
Simplicity.....	17
Performance	18
Reliability	19
Solution Design	22
Network Design.....	23
Compute Design	24
Storage Design	25
Design Considerations	25
Management Connectivity	25
Jumbo Frames	26
Cisco UCS Server vSphere Configuration	26
Cisco Nexus 9000 Series vPC Best Practices	28
Direct Attached Storage Design (Alternative Design)	30
Validation.....	31
Validated Hardware and Software.....	31
Summary	33
Reference Sources for Components in this Design	34
Products and Solutions	34
About the Authors.....	36
Acknowledgements	36



Executive Summary

Cisco Validated Designs (CVDs) consist of 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 our customers.

This document discusses the design principles that go into the FlashStack solution, which is a validated Converged Infrastructure(CI) jointly developed by Cisco and Pure Storage. The solution is a predesigned, best-practice data center architecture with VMware vSphere built on the Cisco Unified Computing System (UCS), the Cisco Nexus® 9000 family of switches, Cisco MDS 9000 family of Fibre Channel switches and Pure Storage FlashArray//M all flash array.

The solution architecture presents a robust infrastructure viable for a wide range of application workloads implemented as a Virtual Server Infrastructure(VSI).

Solution Overview

Introduction

In the current industry there is a trend for pre-engineered solutions which standardize the data center infrastructure, offering the business operational efficiencies, agility and scale to address cloud, bi-modal IT and their business. Their challenge is complexity, diverse application support, efficiency and risk. All these are met by FlashStack with:

- Reduced complexity and automatable infrastructure and easily deployed resources
- Robust components capable of supporting high-performance and high bandwidth virtualized applications
- Efficiency through optimization of network bandwidth and in-line storage compression with de-duplication
- Risk reduction at each level of the design with resiliency built into each touch point throughout

Cisco and Pure Storage have partnered to deliver this Cisco Validated Design, which uses best of breed storage, server and network components to serve as the foundation for virtualized workloads, enabling efficient architectural designs that can be quickly and confidently deployed.

In this document we will describe a reference architecture detailing a Virtual Server Infrastructure composed of Cisco Nexus switches, Cisco UCS Compute, Cisco MDS Multilayer Fabric Switches and a Pure Storage FlashArray//M delivering a VMware vSphere 6.0 U2 hypervisor environment.

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.

FlashStack Program Benefits

FlashStack provides a jointly supported solution by Cisco and Pure Storage. Bringing a carefully validated architecture built on superior compute, world class networking, and the leading innovations in all flash storage. The portfolio of validated offerings from FlashStack includes, but is not limited to the following:

- Consistent performance: FlashStack provides higher, more consistent performance than disk-based solutions and delivers a converged infrastructure based on all-flash that provides non-disruptive upgrades and scalability.
- Cost savings: FlashStack uses less power, cooling, and data center space when compared to legacy disk/hybrid storage. It provides industry-leading storage data reduction and exceptional storage density.
- Simplicity: FlashStack requires low ongoing maintenance, and reduces operational overhead. It also scales simply and smoothly in step with business requirements.
- Deployment choices: It is available as a custom-built single unit from FlashStack partners, but organizations can also deploy using equipment from multiple sources, including equipment they already own.
- Unique business model: The Pure Storage Evergreen Storage Model enables companies to keep their storage investments forever, which means no more forklift upgrades and no more downtime.

- Mission-critical resiliency: FlashStack offers best in class performance by providing active-active resiliency, no single point of failure, and non-disruptive operations, enabling organizations to maximize productivity.
- Support choices: Focused, high-quality single-number FlashStack support is available from FlashStack Authorized Support Partners. Single-number support is also available directly from Cisco Systems as part of their Data Center Critical Infrastructure services offering. Support for FlashStack components is available from Cisco, VMware, and Pure Storage individually and leverages TSAnet for resolution of support queries between vendors.

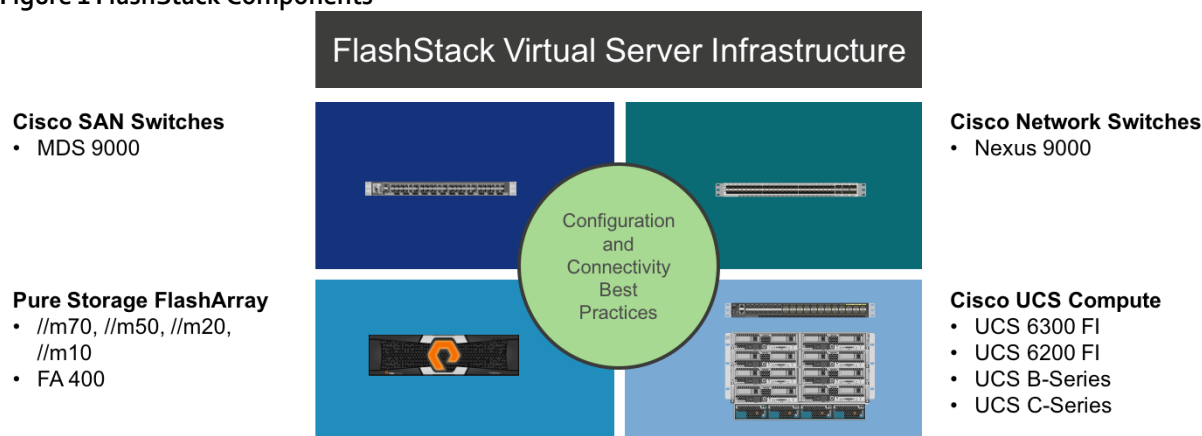


Technical Overview

FlashStack System Overview

The FlashStack architecture is composed of leading edge product lines from Cisco and Pure Storage (Figure 1).

Figure 1 FlashStack Components



This design features a subset of components, centered around the Cisco UCS 6332-16UP and the FlashArray//M70 within a fibre channel implementation for storage communication. A pair of Cisco MDS 9148S SAN switches are positioned between these two main components. This managed compute and storage is delivered to Cisco UCS B200 M4 servers, with all of this extended to the network via a pair of Cisco Nexus 93180YC-EX switches.

Validated Components in the Architecture

[Cisco Nexus 93180-YC-EX Switch](#) – The Cisco Nexus 93180YC-EX is one option among many from the Cisco Nexus 9000 Series Switches viable for FlashStack. These switches are ACI and Tetration ready, offer both modular and fixed 10/25/40/50/100 Gigabit Ethernet switch configurations with scalability up to 60 Tbps of non-blocking performance with less than five-microsecond latency, implementing Layer 2 and Layer 3 Ethernet ports and wire speed VXLAN gateway, bridging, and routing support.



[Cisco MDS 9148S Fabric Switch](#) - The Cisco MDS 9148S 16G Multilayer Fabric Switch is the next generation of options available within the highly reliable Cisco MDS 9100 Series Switches. It includes up to 48 auto-sensing line-rate 16-Gbps Fibre Channel ports in a compact easy to deploy and manage 1-rack-unit (1RU) form factor. In all, the Cisco MDS 9148S is a powerful and flexible switch that delivers high performance and comprehensive Enterprise-class features at an affordable price.



[Cisco UCS 6332-16UP Fabric Interconnect](#) – The Cisco UCS 6332-16UP Fabric Interconnect is the management and communication backbone for Cisco UCS B-Series Blade Servers and C-Series Rack Servers. It implements 20 40-Gigabit Ethernet and Fibre Channel over Ethernet ports, with additional support for 16 unified ports that can be configured to 1 or 10 Gbps Ethernet, or 4/8/16 Gbps Fibre Channel.

The Fabric Interconnect provides high speed upstream connectivity to the network, or converged traffic to servers through its 40Gbps ports, but also allows for Fibre Channel connectivity to SAN switches like the MDS, or alternately directly attached Fibre Channel to storage arrays like the Pure Storage FlashArray through its unified ports.



[Cisco UCS 2304 Fabric Extender](#) – The Cisco UCS 2304 Fabric Extender brings in up to 320 Gbps of converged Ethernet, per pair of fabric extenders, using 40 Gigabit Ethernet connections into the Cisco UCS Chassis. This implements the same form factor as previous generation I/O modules (IOM), supporting the same blade chassis since the release of Cisco UCS.



[Cisco UCS 5108 Blade Server Chassis](#) – The Cisco UCS 5108 Blade Server Chassis supports up to two fabric extenders or Cisco Minis (UCS 6324 Fabric Interconnects). With four hot swappable power supplies and eight half-width blade slots, it allows for power and network redundancy to the blade servers it hosts.



[Cisco UCS B200 M4 Blade Server](#) – The Cisco UCS B200 M4 Blade server is a half-width form factor blade server using the latest Intel® Xeon® E5-2600 v4 Series processor family CPUs with up to 768 GB of RAM (using 32 GB DIMMs), two solid-state drives (SSDs) or hard disk drives (HDDs), and up to 80 Gbps throughput connectivity.



[Cisco UCS 1340 VIC Adapter](#) - The Cisco UCS Virtual Interface Card (VIC) 1340 is a 2-port 40-Gbps Ethernet or dual 4 x 10-Gbps Ethernet, FCoE-capable modular LAN on motherboard (mLOM) designed exclusively for the M4 generation of Cisco UCS B-Series Blade Servers. When used in combination with an optional port expander, the Cisco UCS VIC 1340 capabilities is enabled for two ports of 40-Gbps Ethernet. The Cisco UCS VIC 1340 enables a policy-based, stateless, agile server infrastructure that can present over 256 PCIe standards-compliant interfaces to the host that can be dynamically configured as either network interface cards (NICs) or host bus adapters (HBAs).



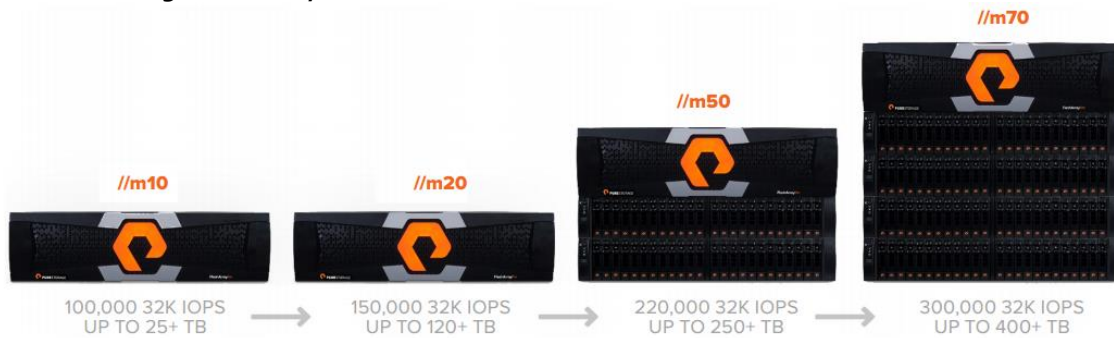
[VMware vCenter Server 6.0](#) – VMware vCenter Server provides unified management of all hosts and VMs from a single console and aggregates performance monitoring of clusters, hosts, and VMs. VMware vCenter Server gives administrators a deep insight into the status and configuration of compute clusters, hosts, VMs, storage, the guest OS, and other critical components of a virtual infrastructure. VMware vCenter manages the rich set of features available in a VMware vSphere environment.

[Pure Storage FlashArray//M](#) - The FlashArray//M expands upon the FlashArray's modular, stateless architecture, designed to enable expandability and upgradability for generations. The FlashArray//M leverages a chassis-based design with customizable modules, enabling both capacity and performance to be independently improved over time with advances in compute and flash, to meet your business' needs today and tomorrow.

Pure Storage FlashArray sets the benchmark for all-flash enterprise storage arrays. It delivers the following:

- Consistent Performance FlashArray delivers consistent <1ms average latency. Performance is optimized for the real-world applications workloads that are dominated by I/O sizes of 32K or larger vs. 4K/8K hero performance benchmarks. Full performance is maintained even under failures/updates.
- Less Cost than Disk Inline de-duplication and compression deliver 5 – 10x space savings across a broad set of I/O workloads including Databases, Virtual Machines and Virtual Desktop Infrastructure. With VDI workloads data reduction is typically > 10:1.
- Mission-Critical Resiliency FlashArray delivers >99.999% proven availability, as measured across the Pure Storage installed base and does so with non-disruptive everything without performance impact.
- Disaster Recovery Built-In FlashArray offers native, fully-integrated, data reduction-optimized backup and disaster recovery at no additional cost. Setup disaster recovery with policy-based automation within minutes and recover instantly from local, space-efficient snapshots, or remote replicas.
- Simplicity Built-In FlashArray offers game-changing management simplicity that makes storage installation, configuration, provisioning and migration a snap. No more managing performance, RAID, tiers or caching. Achieve optimal application performance without any tuning at any layer. Manage the FlashArray the way you like it: Web-based GUI, CLI, VMware vCenter, Windows PowerShell, Python, REST API, or OpenStack.

Figure 2 Pure Storage FlashArray//m Portfolio



[Pure Storage vCenter Plugin for Storage Management](#) - The Pure Storage FlashArray plugin for VMware vSphere provides a free and simple mechanism to manage and provision FlashArray storage from directly inside the vSphere Web Client. As a free and fully supported component of the FlashArray management ecosystem, the FlashArray plugin allows you to perform the following tasks and more in context to the VMware objects they manage on a daily basis:

- Provision new datastores
- Expand datastores
- Delete datastores
- Snapshot datastores
- Restore datastores from snapshots
- Create new datastores from snapshots
- Add datastores to replication groups
- Configure best practices and more

The FlashArray plugin does not require a management appliance and can be directly installed into vCenter from the FlashArrays built-in web management interface. The plugin interacts with the built-in REST API of the FlashArray.

Optional Components Included in the Validated Architecture

[Cisco Nexus 1000V](#) – The Cisco Nexus 1000v is a distributed virtual switch that extends the network edge into the virtual environment of multiple differing hypervisors with an NX-OS CLI implementation. It provides QoS, ERSPAN, VXLAN support and many other features that may not be present in a standard or distributed virtual switch that is native to a hypervisor.

[Cisco Virtual Switch Update Manager \(VSUM\)](#) – The Cisco VSUM is a virtual appliance that is registered as a plug-in to the VMware vCenter Server. The Cisco VSUM simplifies the installation and configuration of the Cisco Nexus 1000v, to include the migration of VMs and hosts.

[Cisco UCS Director](#) – Cisco UCS Director replaces time-consuming, manual provisioning and de-provisioning of data center resources with automated workflows. It supports multiple hardware vendors and multiple hypervisors to deliver heterogeneous infrastructure management. With FlashStack there is a custom adapter to quickly integrate and manage FlashStack resources.

Alternative Components for the Architecture

[Cisco UCS 6200 Fabric Interconnect](#) – The Cisco UCS 6248-UP and 6296-UP Fabric Interconnect models present 10 Gbps alternatives to the UCS 6332-16UP Fabric Interconnect supporting 8 Gbps fibre channel within their unified ports. These fabric interconnects can provide the same management and communication backbone to Cisco UCS B-Series and C-Series as the Cisco UCS 6300 series fabric interconnects.



[Cisco UCS 2204XP/2208XP Fabric Extenders](#) – The Cisco UCS 2204XP and 2208XP Fabric Extenders are four and eight port 10Gbps converged fabric options to the Cisco UCS 2304 Fabric Extender, that deliver combined respective bandwidths of 80Gbps or 160Gbps when implemented as pairs.



[Cisco UCS C220 M4 Rack Mount Server](#) – The Cisco UCS C220 M4 Rack Server is a 1RU rackmount option to the B200 M4. The Cisco UCS C220 M4 also leverages the power of the latest Intel® Xeon® E5-2600 v4 Series processor family CPUs with up to 768 GB of RAM (using 32 GB DIMMs), four large form-factor (LFF) or eight small form-factor (SFF) solid-state drives (SSDs) or hard disk drives (HDDs), and up to 240 Gbps throughput connectivity (with multiple VIC). Cisco UCS C-Series servers added into a FlashStack are easily managed through Cisco UCS Manager with Single Wire Management, using 10G or 40G uplinks to the FI.



[Pure Storage FA400](#) - The FA-400 series from Pure Storage® offers an economical and scalable all-flash storage solution able to handle virtually any workload, from single-application acceleration to data center consolidation. Start at the capacity that fits your business, and then add base and expansion shelves as you grow. The Pure Storage Evergreen Storage model ensures that your FA-400 investment – like every system they sell – will be good for several generations of forklift-free hardware and software upgrades. Performance and capacity can be upgraded independently, and every upgrade will be problem-free: no disruptions, no downtime, and no data migrations.

[Cisco UCS Central](#) – Cisco UCS Central scales policy decisions and management of multiple Cisco UCS domains into a single source. Service Profiles and maintenance can be uniform in their implementation across local and/or remotely deployed fabric interconnects.

[Cisco Cloud Center](#) – Cisco Cloud Center, formerly Cliqr, can securely extend FlashStack resources between private and public cloud. Application blueprints rapidly scale the user's needs, while implementing them in a manner that is secure, and provides the controls for proper governance.

Cisco UCS Differentiators

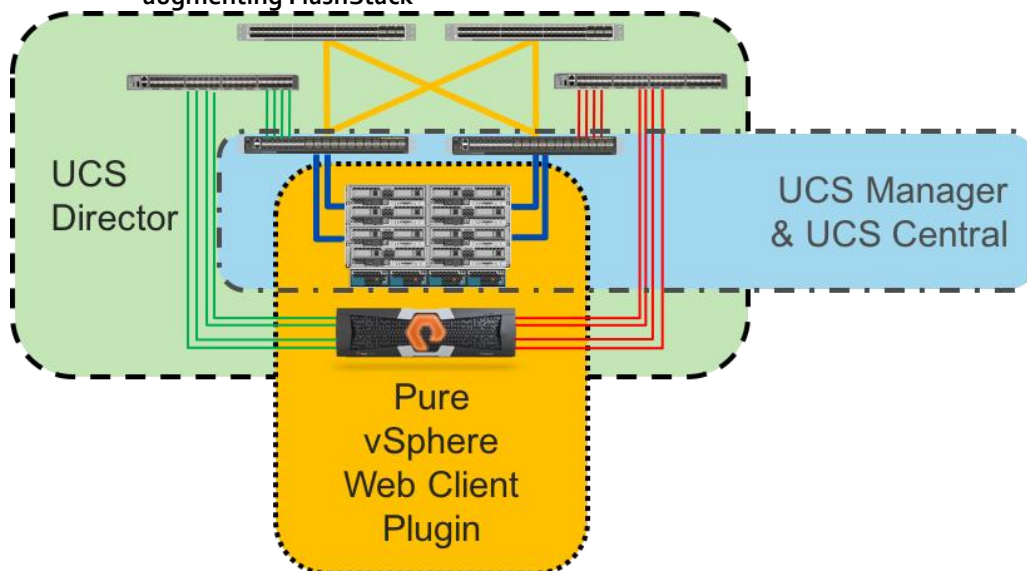
Cisco Unified Compute System has revolutionized the way servers are managed in the datacenter. The following are the unique differentiators of Cisco Unified Computing System and Cisco UCS Manager.

- **Embedded Management** —In Cisco UCS, the servers are managed by the embedded firmware in the Fabric Interconnects, eliminating need for any external physical or virtual devices to manage the servers.

- **Unified Fabric** —In Cisco UCS, from blade server chassis or rack servers to FI, there is a single Ethernet cable used for LAN, SAN and management traffic. This converged I/O results in reduced cables, SFPs and adapters – reducing capital and operational expenses of overall solution.
- **Auto Discovery** —By simply inserting the blade server in the chassis or connecting rack server to the fabric interconnect, discovery and inventory of compute resource occurs automatically without any management intervention. The combination of unified fabric and auto-discovery enables the wire-once architecture of Cisco UCS, where compute capability of Cisco UCS can be extended easily while keeping the existing external connectivity to LAN, SAN and management networks.
- **Policy Based Resource Classification** —When a compute resource is discovered by Cisco UCS Manager, it can be automatically classified to a given resource pool based on policies defined. This capability is useful in multi-tenant cloud computing. This CVD highlights the policy based resource classification of Cisco UCS Manager.
- **Combined Rack and Blade Server Management** —Cisco UCS Manager can manage Cisco UCS B-series blade servers and C-series rack server under the same Cisco UCS domain. This feature, along with stateless computing makes compute resources truly hardware form factor agnostic.
- **Model based Management Architecture** —Cisco UCS Manager architecture and management database is model based and data driven. An open XML API is provided to operate on the management model. This enables easy and scalable integration of Cisco UCS Manager with other management systems.
- **Policies, Pools, Templates** —The management approach in Cisco UCS Manager is based on defining policies, pools and templates, instead of cluttered configuration, which enables a simple, loosely coupled, data driven approach in managing compute, network and storage resources.
- **Loose Referential Integrity** —In Cisco UCS Manager, a service profile, port profile or policies can refer to other policies or logical resources with loose referential integrity. A referred policy cannot exist at the time of authoring the referring policy or a referred policy can be deleted even though other policies are referring to it. This provides different subject matter experts to work independently from each-other. This provides great flexibility where different experts from different domains, such as network, storage, security, server and virtualization work together to accomplish a complex task.
- **Policy Resolution** —In Cisco UCS Manager, a tree structure of organizational unit hierarchy can be created that mimics the real life tenants and/or organization relationships. Various policies, pools and templates can be defined at different levels of organization hierarchy. A policy referring to another policy by name is resolved in the organization hierarchy with closest policy match. If no policy with specific name is found in the hierarchy of the root organization, then special policy named “default” is searched. This policy resolution practice enables automation friendly management APIs and provides great flexibility to owners of different organizations.
- **Service Profiles and Stateless Computing** —A service profile is a logical representation of a server, carrying its various identities and policies. This logical server can be assigned to any physical compute resource as far as it meets the resource requirements. Stateless computing enables procurement of a server within minutes, which used to take days in legacy server management systems.
- **Built-in Multi-Tenancy Support** —The combination of policies, pools and templates, loose referential integrity, policy resolution in organization hierarchy and a service profiles based approach to compute resources makes Cisco UCS Manager inherently friendly to multi-tenant environment typically observed in private and public clouds.
- **Extended Memory** — The enterprise-class Cisco UCS B200 M4 blade server extends the capabilities of the Cisco Unified Computing System portfolio in a half-width blade form factor. The Cisco UCS B200 M4 harnesses the power of the latest Intel® Xeon® E5-2600 v4 Series processor family CPUs with up to 1536 GB of RAM (using 64 GB DIMMs) – allowing huge VM to physical server ratio required in many deployments, or allowing large memory operations required by certain architectures like big data.

- Virtualization Aware Network — Cisco VM-FEX technology makes the access network layer aware about host virtualization. This prevents domain pollution of compute and network domains with virtualization when virtual network is managed by port-profiles defined by the network administrators' team. VM-FEX also off-loads hypervisor CPU by performing switching in the hardware, thus allowing hypervisor CPU to do more virtualization related tasks. VM-FEX technology is well integrated with VMware vCenter, Linux KVM and Hyper-V SR-IOV to simplify cloud management.
- Simplified QoS —Even though Fibre Channel and Ethernet are converged in Cisco UCS fabric, built-in support for QoS and lossless Ethernet makes it seamless. Network Quality of Service (QoS) is simplified in Cisco UCS Manager by representing all system classes in one GUI panel.
- Unified Management & Automation – Cisco UCS Manager provides consistent scalable policy decisions for deployed compute within fabric interconnects. Cisco UCS Central (optional) further scales that providing added management features to multiple Cisco UCS domains handled by those fabric interconnects. Cisco UCS Director (optional) brings automation to deployments of Cisco UCS extending from the virtualization layer and compute, up through IP and storage networks as well as reaching storage partners.

Figure 3 Cisco UCS Management and Automation Solutions with the Pure Storage vSphere Web Client Plugin further augmenting FlashStack

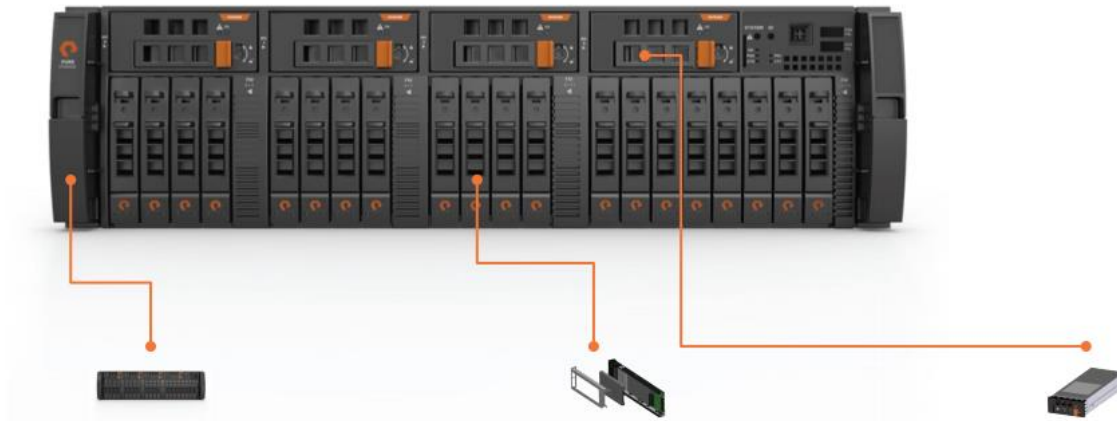


Pure Storage Differentiators

Purity storage software built for flash, uses FlashCare technology to virtualize the entire pool of flash within the FlashArray, and allows Purity to both extend the life and ensure the maximum performance of consumer-grade MLC flash.

- Granular and adaptive—Purity Core is based upon a 512-byte variable block size metadata layer called the Purity Core. This fine-grain metadata enables all of Purity's data and flash management services to operate at the highest efficiency
- Best data reduction available—Purity FlashReduce implements five forms of inline and post process data reduction to offer the most complete data reduction in the industry. Data reduction operates at a 512- byte aligned variable block size, to enable effective reduction across a wide range of mixed workloads without tuning.
- Highly available and resilient—Resiliency is job 1 for Purity. FlashProtect implements high availability, dual-parity RAID-3D, non-disruptive upgrades, and encryption, all of which are designed to deliver full performance to the FlashArray during any failure or maintenance event.
- Disaster recovery built in—FlashRecover combines space saving snapshots, replication, and protection policies into an end-to-end data protection and recovery solution that protects data against loss locally and globally. All FlashProtect services are fully-integrated in the FlashArray and leverage the native data reduction capabilities.

Figure 4 FlashArray//m Front Panel



- //M chassis—The //M Chassis leverages internal PCIe networking for controller HA, as well as both PCIe/NVMe and 12Gb SAS for flash and NV-RAM module connectivity. The result is ultra-fast module connections that enable hotplug and HA of every component.
- Flash modules—Dual-drive flash modules pack two independent SSDs in each slot, doubling the potential performance and density. Drive modules are redundantly connected to both controllers, and can be replaced without performance impact.
- NV-RAM modules NV-DDR4—non-volatile cache modules are used to protect writes when power is lost during I/O processing from power loss. DDR4 memory is backed by super capacitors, and the modules can retain their data for 1 year without power. 2 or 4 redundant modules are networked with hot-swap PCIe/NVMe.

Figure 5 FlashArray//M Back Panel

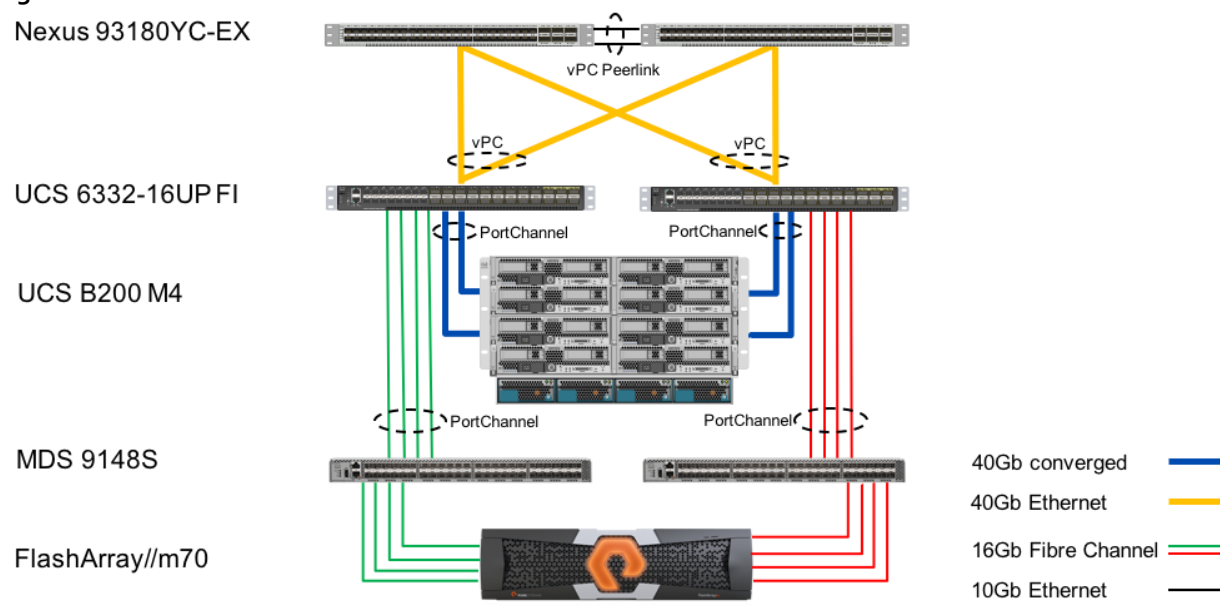


- Controller modules—The //M features four controller options, to allow different tiers of performance and capacity. New controllers are shipped roughly annually, and can be replaced or upgraded without performance impact.
- I/O modules—The //M has onboard SAS, replication, and management ports, and host-based IO ports can be configured to meet a variety of needs. 6 slots per chassis, configurable with 8 or 16 Gb/s FC, or 10 Gb/s Ethernet iSCSI.

Solution Architecture

The FlashStack architecture brings together the proven data center strengths of the Cisco UCS and Cisco Nexus network switches with the Fibre Channel storage of the leading visionary in all flash arrays. This collaboration creates a simple, yet powerful and resilient data center footprint for the modern enterprise.

Figure 6 FlashStack Architecture



The components of this integrated architecture shown in Figure 6 above are:

- Cisco Nexus 93180YC-EX – 100Gb capable, LAN connectivity to the Cisco UCS compute resources.
- Cisco UCS 6332-16UP Fabric Interconnect – Unified management of Cisco UCS compute, and that compute's access to storage and networks.
- Cisco UCS B200 M4 – High powered, versatile blade server, which was conceived for virtual computing.
- Cisco MDS 9148S – 16Gb Fibre Channel connectivity within the architecture, as well as interfacing to resources already present in the environment.
- Pure Storage FlashArray//M70 – All flash storage implemented with inline compression and deduplication in a simple and resilient manner.

Virtualization layer components and managers of the architecture additionally include:

- Cisco UCS Manager – Management delivered through the Fabric Interconnect, providing stateless compute, and policy driven implementation of the servers managed by it.
- Cisco Nexus 1000v (optional) – Access layer distributed virtual switch providing a continuity of features, found within Nexus switches, at the virtualization layer.
- Cisco Virtual Switch Update Manager (optional) – vSphere Web Client driven management and installation of the Nexus 1000v.

- Cisco UCS Director (optional) – Automation of the deployment of Cisco infrastructure, complete provisioning of servers as vSphere resources, and accompanying storage from the Pure Storage FlashArray.
- VMware vSphere and VMware vCenter – Hypervisor and Virtual Machine Manager.
- VMware vDS – Distributed Virtual Switch for the vSphere environment (if Cisco Nexus 1000v is not used).
- Pure Storage vSphere Web Client Plugin – Easy to use management of volumes within the vSphere Web Client.

Design Principles

FlashStack is built around three basic design principles:

- Simplicity – Delivering a solution easily implemented for scale and management.
- Performance – Best of breed components creating an ideal architecture for Enterprise application needs
- Reliability – Redundancy at each level of the design.

These design principles are brought into each layer of the design with features and implementation options described below.

Simplicity

Network Simplicity

Simplicity in the Network is first touched by the greatly reduced cabling and complexity with the converged fabric within Cisco UCS.

- VLANs brought in through the Nexus switches have minimal configuration touchpoints within the Cisco UCS fabric interconnects, to deliver as needed, to the Cisco UCS compute.
- Additions and changes to VLANs can be non-disruptively adjusted within the Cisco UCS vNIC templates and scale automatically to configured servers.
- Optional quick and consistent deployment with Cisco UCS Director automation.

Compute Simplicity

Cisco UCS Service Profiles scale easily from templates that produce consistent server configurations on demand.

- Adjustments within the template will propagate across deployed systems, using maintenance acknowledgements to assure against unanticipated downtime.
- Policy based management implements desired options from firmware management and BIOS settings, to disk layout and server qualification.
- Large scale, uniform deployment control, available across multiple geographic locations brought in with the option of Cisco UCS Central added on as an over-arching solution.
- As with the Nexus, quick and consistent deployment available through Cisco UCS Director automation.

Storage Simplicity

FlashArrays automatically adapt to changing conditions in real time. For example, when write load is heavy, they defer inline data reduction to remain responsive. When rebuilding data segments, they may slow down deep reduction that minimizes the footprint of long-lived data. When SSDs are busy writing, they often use RAID rebuilds to satisfy read requests. In extreme cases, when free space becomes low, they may throttle writes to keep

resources available for recycling (also called garbage collecting) storage that holds obsolete content. Arrays make these, and dozens of similar adjustments, automatically to continually optimize utilization of their internal resources

The net effect is that many common administrative functions don't exist in FlashArray environments. For example:

- **No RAID groups:** FlashArrays manage storage as a single pool. They persist incoming data in approximate arrival order rather than in a fixed Logical Block Address (LBA)-physical location relationship. Administrators do not decide which SSDs to virtualize as which LUNs, how to RAID-protect data, or what spares to reserve. Arrays themselves allocate space for storage and rebuilding, transparently and optimally.
- **No tiers:** Because arrays treat all storage the same, there is no tiering, as is typically found in hybrid arrays. Administrators do not associate specific devices with specific clients, applications, or data sets. Similarly, there is no I/O performance tuning. Administrators connect clients and volumes, and arrays manage their resources to respond fairly to all. Arrays are better-positioned to rebalance resources in real time than an administrator observing and responding to changing conditions at human speed could possibly be.
- **No cache management:** Typical SSD read access times are in the range of a couple hundred microseconds, so there is little need for read cache in an AFA. The majority of FlashArray DRAM is used to buffer data waiting to be persisted, and to cache metadata. Arrays manage their own DRAM; there are no associated administrative tasks.
- **No alignment constraints:** Aligning blocks of client data with physical media locations is vital for disk array performance. But because SSDs do not seek or rotate, alignment need not be a consideration. But some legacy designs align virtual and physical block addresses, and even the more modern designs that use content addressing depend on logical-to-physical alignment, neither of which is relevant with flash.

Performance

Network Performance

Enabling 40G within Cisco UCS, and up to 100G for Nexus, allows for incredibly high levels of bandwidth and reduces the number of uplinks to provide that bandwidth.

- Virtual Port Channels provide non-blocking performance that can be expanded as needed.
- QoS configurable at each touch point of the network to guarantee service prioritization as needed.

Compute Performance

Cisco UCS uses the latest family of Intel processors with high memory capacity geared toward virtualized environments.

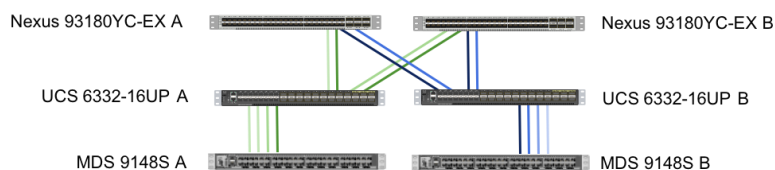
Bandwidth coming into the UCS Fabric Interconnect from the network and storage environment is expandable as port channels to allow for extremely large throughput (Figure 7).

Figure 7 Bandwidth options between the Cisco Nexus 9318oYC-EX, Cisco UCS 6332-16UP Fabric Interconnect, and the MDS 9148S Multilayer Fabric Switch

Nexus vPC	FI A	FI B	FI Total
one port member	80G	80G	160G
two port members *	160G	160G	320G

MDS	FI A	FI B	FI Total
one port	16G	16G	32G
two ports	32G	32G	64G
three ports	48G	48G	96G
four ports *	64G	64G	128G

* Both of these sets of options can expand beyond what is listed as needed



Connections within the Fabric Interconnect to the Cisco UCS Blade Chassis can additionally be expanded to meet high traffic needs as it connects into the Cisco UCS 2304 Fabric Extender (Figure 8).

Figure 8 Bandwidth options between the Cisco UCS 6332-16UP Fabric Interconnect and the Cisco UCS 2304 Fabric Extender

IOM	FI A	FI B	FI Total
one uplink	40G	40G	80G
two uplinks	80G	80G	160G
four uplinks	160G	160G	320G



Storage Performance

The FlashArray//M70 offers purpose-built hardware designed for the performance of flash. And delivers consistent <1ms average latency. Performance is optimized for the real-world applications workloads that are dominated by I/O sizes of 32K or larger vs. 4K/8K hero performance benchmarks. Full performance is maintained even under failures/updates.

The FlashArray offers the following:

- Dual-socket 16-core processor
- 512 GB DDR4-2133 memory
- 4GB NVRAM
- 8x10 GbE iSCSI ports or 8 16 Gb Fibre Channel
- 4x10 GbE ports for replication
- Up to 11.5 GB/s bandwidth

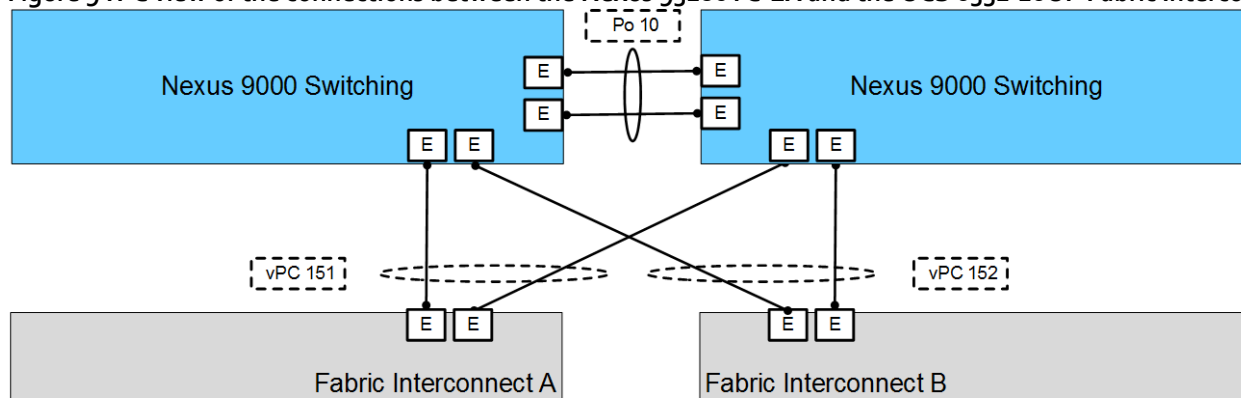
Reliability

Network Reliability

Network reliability is attained through the configuration of Port Channels within the design.

- Virtual port channels (vPC) are used between FIs and Cisco Nexus switches. vPCs provide higher availability than port channels as it can continue to forward traffic even if one of the Cisco Nexus switches fail because vPCs distribute member links of port-channel across different Cisco Nexus switches.

Figure 9 vPC view of the connections between the Nexus 9318oYC-EX and the UCS 6332-16UP Fabric Interconnect



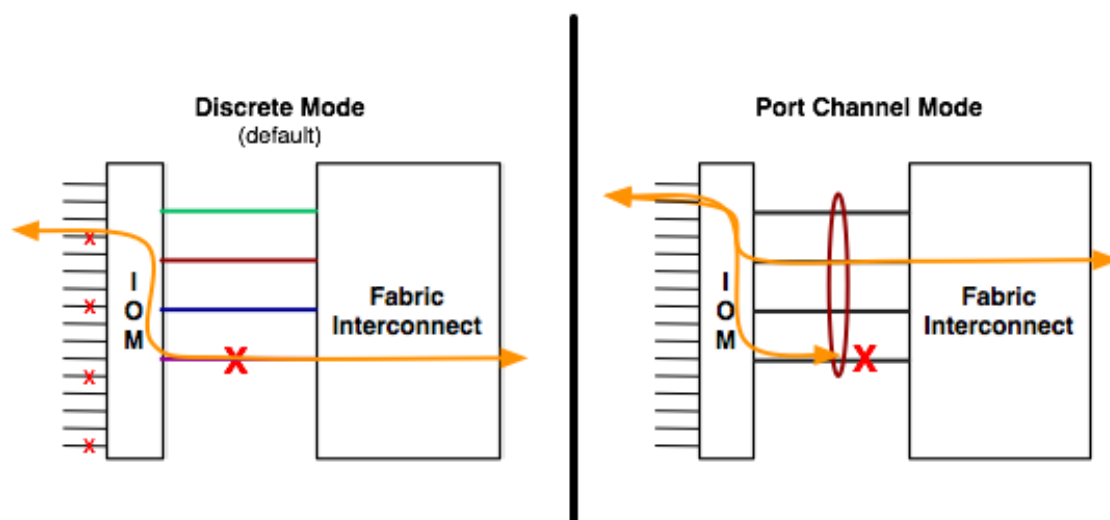
- In-Service Software Upgrade (ISSU) allows for operation without interruption during upgrades and patching of the Nexus 9500 and 9300 switches.
- The Nexus switches are fully redundant and have hot swappable components.
- Within the connecting Fabric Interconnects, these upstream connections coming in from the Nexus are configured as Link Aggregation Control Protocol (LACP) Port Channels. The Fabric Interconnect treats the connected vPC as a single switch and automatically distributes traffic between member interfaces.

Compute Reliability

Reliability at the compute layer is by default delivered within the virtualization layer taking advantage of VMware HA, but extends into multiple areas of Cisco UCS compute configuration.

- Cisco UCS 5108 blade server platform is highly redundant with redundant power supplies, fans and fabric extender modules. C-series rack mount servers also have redundant power supplies and fans. The rack mount servers are directly connected and managed by the fabric interconnects in the FlashStack design.
- The UCS servers are implemented as stateless systems through consistent service policies used to generate UCS Service Profiles. These Service Profiles are configured to be booted from SAN, and can be re-associated between servers delivering the same compute instance in the result of a hardware failure.
- The fabric extenders for the Cisco UCS 5108 blade chassis are deployed with up to 4x40Gb to 6300 series fabric interconnects and up to 8x10Gb to 6200 series fabric interconnects. The links are automatically configured as port channels by the Chassis/FEX Discovery Policy within Cisco UCS Manager, which ensures rerouting of flows to other links in the event of a link failure.

Figure 10 Automatic configuration of Port Channels through the Chassis/FEX Discovery Policy setting



Storage Reliability

Array Resilience

"Always on" 24x7 availability is an absolute requirement for enterprise storage. FlashArray AFAs provide upwards of "five nines" (99.999+%) of uptime, based on installed base data collected by the CloudAssist array monitoring and proactive service facility. Pure Storage has introduced an innovative support model that leverages cloud technology, called CloudAssist, for real-time monitoring of deployed arrays in order to provide a proactive support experience.

With CloudAssist, Pure Storage can remotely monitor arrays, diagnose problems, provide early warnings and reduce issue resolution time. Once connected to CloudAssist, the FlashArray sends diagnostic information every 30 seconds, and logs every 60 minutes to the cloud. The diagnostics and logs are sent to the cloud over a secure HTTPS channel. The 30-second diagnostics are light-weight logs that include information about capacity, performance and health diagnostics of the FlashArray. The 60-minute logs are operational logs from Purity Operating Environment that can be used for advanced global analysis of the FlashArray connected to the cloud, in addition to providing a historical view needed for problem diagnosis and resolution. The logs do not contain any end-user data other than array configuration information. The diagnostic and log processing engine in the cloud regularly monitors and analyzes the health of each FlashArray by parsing the logs received. The engine helps to proactively find potential issues on the FlashArray and alert the end-user along with Pure Storage Support for appropriate action.

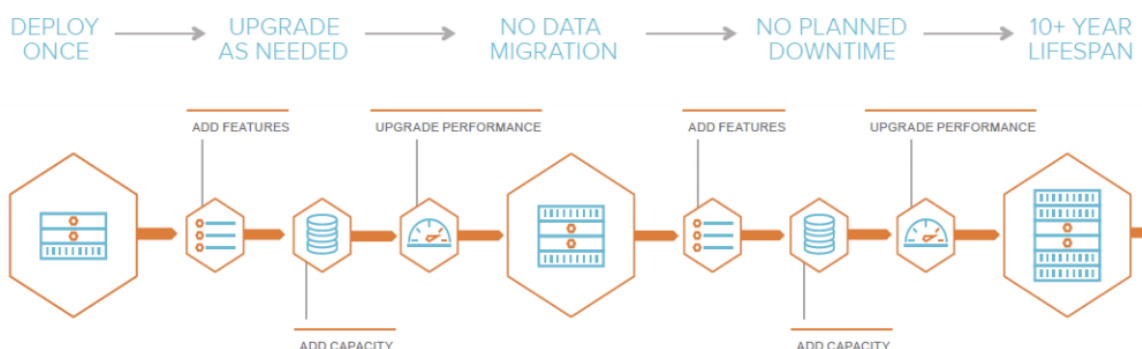
Additionally, the cloud-driven monitoring engine enables Pure Storage to analyze large amounts of data from our entire customer-base. Pure Storage uses these global analytics to cross reference logs between a large set of customers and proactively work on fixes for many customers simultaneously, thus improving the overall customer service experience. But availability isn't just about keeping an array running—it's about resilience—maintaining data integrity and accessibility at sustained levels of performance in the face of failures, both in the array itself and in the data center environment.

FlashArray resilience starts with redundant hardware components and interconnects. Field-replaceable components, up to and including entire controllers, can be replaced or upgraded while an array is serving clients. Multiple storage network ports in each controller make it possible to connect an array to two independent fabrics for end-to-end redundancy. Both of an array's controllers accept client commands, but all back-end processing is done by one primary controller. With this design, a controller failure results in rapid failover (in seconds), and, perhaps more important, negligible reduction in I/O performance. The software is designed to operate "24x7," including non-disruptive upgrade (NDU). New versions can be installed and activated while an array is serving clients. The software cooperates closely with both hardware and CloudAssist to deliver the quality of service that enterprise applications require. It generates alerts both for abnormal hardware conditions and for conditions such as low free space that might require administrative action. Alerts are transmitted to user-designated electronic mail accounts, to designated SNMP servers, and

with customer permission, to CloudAssist servers, where they are archived, analyzed, and, where action is indicated, automatically routed to the Pure Storage Support organization as support cases.

With non-disruptive upgrades, Pure allows for seamless adoption of new hardware guaranteed by the Evergreen Model. In the Evergreen Model, arrays are never replaced in their entirety, so there is no data migration. You can plan on downtime-free, risk-free rapid online upgrade cycles for the life of an array. Figure 11 illustrates the main features of the Evergreen storage model.

Figure 11 The Pure Evergreen Model depicted over the lifespan of a FlashArray



For more information on the Pure Evergreen Model, see: <https://www.purestorage.com/why-pure/evergreen.html>

The FlashArray media layout and metadata structures are designed for robustness and data integrity. Arrays maintain complete metadata and data state in a combination of SSDs and NVRAMs, including that of “in-flight” data (data not yet persisted on SSD storage). FlashArray controllers are therefore stateless, making failover rapid, and non-disruptive replacement and upgrade possible. Indeed, an array’s SSDs and NVRAMs can be moved as a group to a different controller pair, and the entire configuration and all data remain intact. Arrays implement RAID-3D technology, both to supplement SSDs’ own read error detection and to protect against a minimum of two concurrent media or device failures.

All field-replaceable FlashArray components, including controllers, are designed for non-disruptive replacement—both in terms of continuous client service and sustained performance. Non-disruptive controller and software replacement includes upgrades. An array’s controllers can be replaced one-by-one with more powerful models while it is operating and likewise, Purity software can be upgraded non-disruptively while an array is operating.

Solution Design

The FlashStack design leverages a long history of Cisco data center best practices, while also introducing new features that are specific to Pure Storage and new product developments from Cisco.

Figure 12 FlashStack Architecture

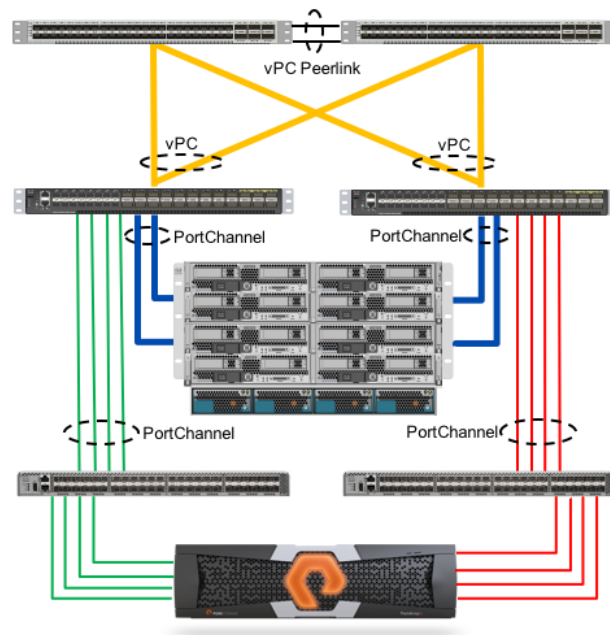
Nexus 93180YC-EX

UCS 6332-16UP FI

UCS B200 M4

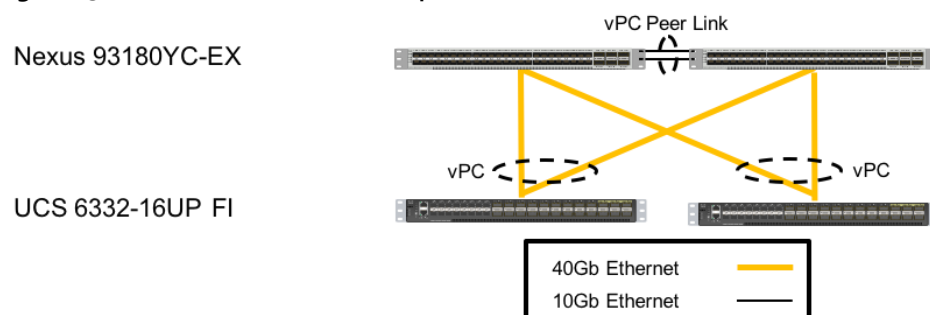
MDS 9148S

FlashArray//m70



Network Design

The network begins at the 100G capable Nexus 93180YC-EX connecting into the Cisco UCS 6332-16UP Fabric Interconnects.

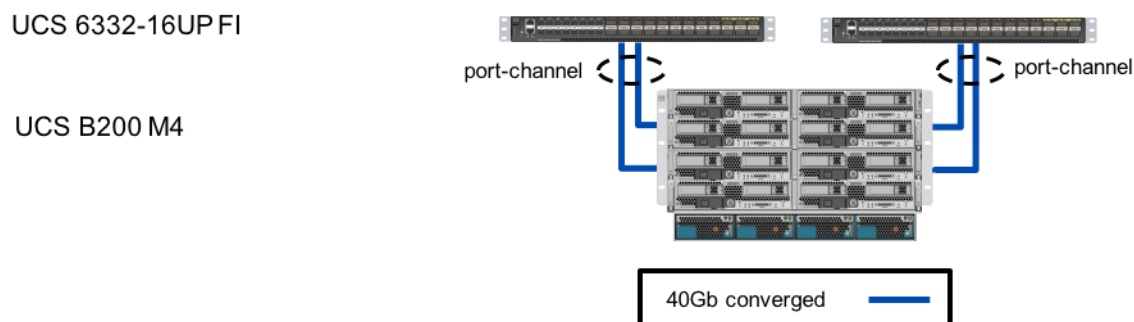
Figure 13 FlashStack Network Components

The connections between these layers are configured as Virtual Port Channels for maximum non-blocking performance of allocated ports. 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 PortChannel (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. vPC addresses aggregate bandwidth, link, and device resiliency.

Each step in the network design brings redundancy, and continuity of implemented features. As a VLAN network segment is brought into the top layer Nexus, continues into the Cisco UCS Fabric Interconnect, expressed out to the vNIC adapters of the Cisco UCS Service Profile before reaching the virtual edge having a clear aggregate path of up to 80Gbps to each server in the Cisco UCS domain. With either the Nexus 1000v or the VMware vDS, the QoS prioritization is guaranteed through the preceding physical components, whether it is enforced at the virtual layer.

Compute Design

Cisco UCS B200 M4 blade servers were used for their high bandwidth, high CPU, and high memory capacity perfect for virtualized environments.

Figure 14 FlashStack Compute Components

Driven through the Cisco UCS 6332-16UP Fabric Interconnect, these blades are stateless through Cisco UCS Service Profiles, creating a hardware footprint that is easily applied to a replacement blade. Each server was configured to boot from SAN volumes shared from the Pure Storage FlashArray to complete the stateless deployment.

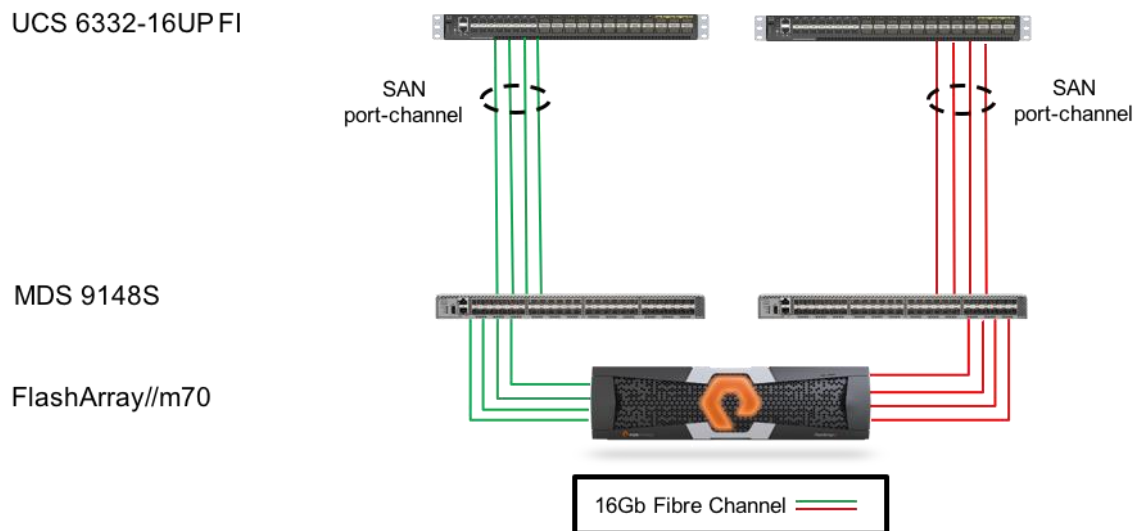
Each B-Series Blade server was equipped with Cisco UCS VIC 1340 adapters that, with the optional port expander, can implement 80Gbps of converged Ethernet and Fibre Channel over Ethernet traffic to the blade. With the VIC, multiple virtual adapters are configured to pass differing traffics like management vs application data that may need different prioritization needs. The Cisco UCS 2304 Fabric Extenders connect into the backplane of Cisco UCS 5108 chassis with up to 320 Gbps of bandwidth. Links coming into the Fabric Extenders are automatically configured as resilient port-channels through policy configuration setting of the Link Grouping Preference.

vNIC Templates are configured as redundancy pairs, to assure that VLANs stay synchronized between the two paths taken over the respective fabric interconnects. As the interfaces are brought into the virtual environment, they are configured as active/active pairs with the exception of a set of dedicated vMotion adapters that are set as active/standby to prevent vMotion traffic from needing to traverse between fabric interconnects, generating unnecessary traffic on the Nexus switch.

Storage Design

The storage aspect of the design uses 16Gb fibre channel connections to extend the Pure Storage FlashArray to the Cisco UCS compute through a pair of Cisco MDS switches as shown in Figure 15.

Figure 15 FlashStack Storage Components



Deployment of the Pure Storage FlashArray is quick and simple after the entry of a few pieces of information relevant to the datacenter environment. The FlashArray supports up to three dual port 16G fibre channel adapters, delivering up to 192 Gbps of aggregate bandwidth across the two controllers in the FlashArray. Volume and datastore deployment are easily put in place with the Pure vSphere Web Plugin enabled.

The FlashArray connects up into a pair of MDS 9148S Multilayer Switches for the SAN switching and zoning. Devices are given alias names for easier identification within zoning and flogi connectivity. Zones set up within the fabric switch provide added security and help optimize traffic to travel exclusively to the intended target. Connections coming out of the MDS 9148S to the UCS Fabric Interconnect are configured as port-channels on the MDS and within the FI for added resiliency.

Within the Fabric Interconnect it is set as end-host mode or NPV mode for the Unified Ports that are connecting into these MDS port-channels. Internally this allows for the connection of the Virtual Host Based Adapters (vHBA) configured from the Cisco UCS VIC 1340 adapters within each Cisco UCS B-Series Blade Server. Each of these vHBAs have unique World Wide Port Numbers (WWPN) that have been previously set up within the zoning in the MDS switch, and the host groups of the Pure Storage Web Portal, allowing for access to the boot volumes and datastore volumes shared by the FlashArray.

Design Considerations

Management Connectivity

Out-of-band management is handled by an independent switch that could be one currently in place in the customer's environment. Each FlashStack physical device had its management interface carried through this Out-of-band switch, with in-band management carried as a differing VLAN within the solution for ESXi, vCenter and other virtual management components.

Out-of-band configuration for the components configured as in-band could be enabled, but would require additional uplink ports on the 6332-16UP Fabric Interconnects if the out of band management is kept on a separate out of band switch. A disjoint layer-2 configuration can then be used to keep the management and data plane networks completely separate. This would require 2 additional vNICs (for example, OOB-Mgmt-A, OOB-Mgmt-B) on each server, which are associated with the management uplink ports.

Jumbo Frames

Jumbo frames are a standard recommendation across Cisco designs to help leverage the increased bandwidth availability of modern networks. To take advantage of the bandwidth optimization and reduced consumption of CPU resources gained through jumbo frames, they were configured at each network level to include the virtual switch and virtual NIC.

This optimization is relevant for VLANs that stay within the pod, and do not connect externally. Any VLANs that are extended outside of the pod should be left at the standard 1500 MTU to prevent drops from any connections or devices not configured to support a larger MTU.

Cisco UCS Server vSphere Configuration

Cisco UCS B-Series servers are installed with ESXi 6.0 U2 using Cisco VIC 1340 adapters to provide separate virtual NICs for the combined management and infrastructure traffic versus application virtual NICs. Within vSphere these hosts were further divided into differing clusters that supported either the virtual infrastructure management components, or the production application virtual machines. For each of these clusters, both VMware High Availability (HA) and VMware Distributed Resource Scheduler (DRS) are enabled.

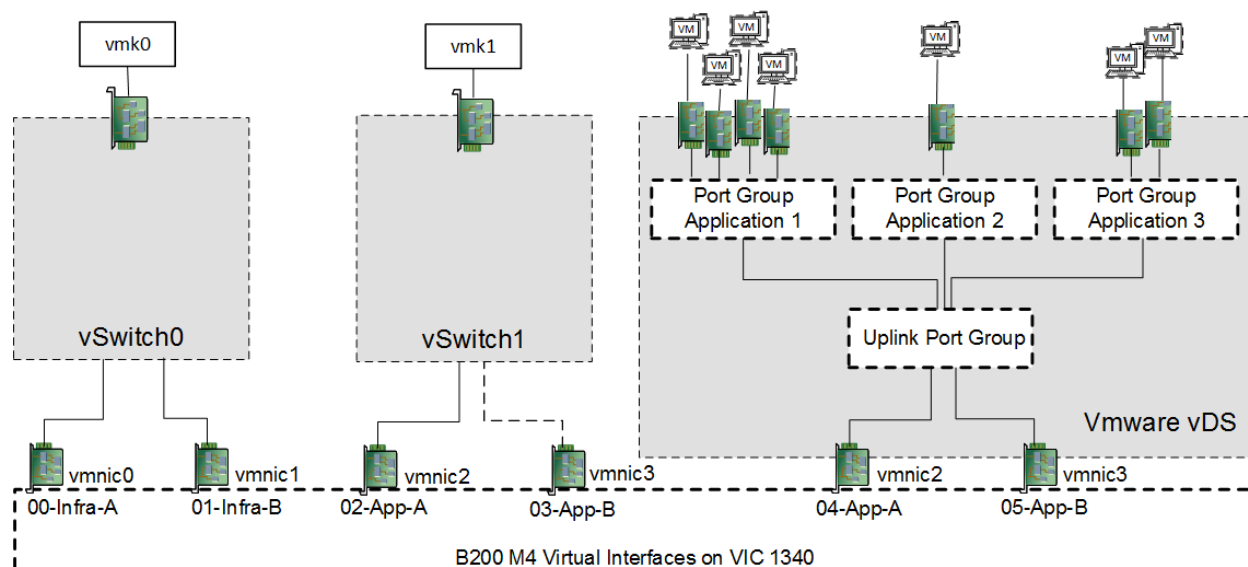
VMware HA is turned on for the clusters to allow automated recovery of active VMs in the event of a physical failure in the underlying ESXi host it resides upon. Depending upon application priority or being up versus having resource guarantees, HA Admission Control might be turned off to allow power-on of VMs in failure scenarios where resources may be constrained.

For VMware DRS, Automation Level should be set as comfortable to the customer. Further into DRS configuration certain infrastructure and application VMs should be set up under DRS Groups Manager and Rules to have placement rules applied for them. These rules can be set to include:

- Keep Virtual Machines Together – For VMs that work with each other that can take advantage of increased performance by being adjacent to each other within the same hypervisor.
- Separate Virtual Machines – For VMs with some form of application level high availability to guarantee that member VMs are not impacted by the same hardware fault.
- Virtual Machines to Hosts - VM to host association that may be relevant for reasons such as licensing.

VMware vDS configured for Application networks, with the management network and vMotion left on separate independent vSwitches as shown in Figure 16.

Figure 16 vDS shown on a Cisco UCS B200 M4 server



Additional standard tasks and best practices include the following:

- A vMotion vmkernel interface is added to each host during the initial setup.
- NTP is set on each ESXi server.
- Shared storage added to each ESXi host in FlashStack using the Pure Storage vSphere Web Plugin.
- Cisco ESXi enic network and fnic fibre channel adapter drivers were applied to each server.
- An ESXi swap datastore is specified to easily allow for separation of VM swapfiles, to make them excludable from snapshots and backup purposes.
- Pure Storage recommended ESXi tuneables were applied using the vSphere PowerCLI.

Cisco UCS Server Virtual Switching using Cisco Nexus 1000V (Optional)

The optional Cisco Nexus 1000v was deployed using the Cisco Virtual Switch Update manager to take advantage of the added features it brings. These features include the following:

- System VLANs
- QOS Marking at the VLAN Level
- QOS Priority Queuing
- VXLAN Support
- ERSPAN
- NX-OS CLI
- NX-OS Show Commands / Monitoring

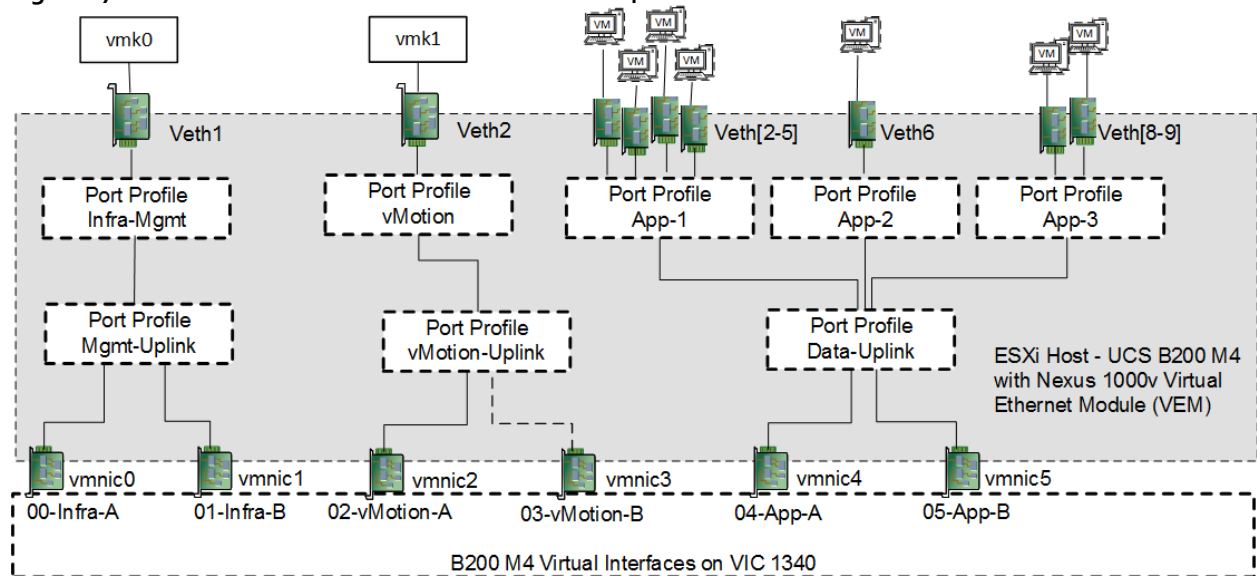
The Cisco Nexus 1000v comprises of the following components and operationally emulates a modular switch where:

- Virtual Supervisor Module (VSM) – is the control and management plane of the virtual switch. VSM is deployed as an external virtual machine and runs NX-OS to manage multiple Virtual Ethernet Modules as one logical modular switch.

- Cisco Virtual Ethernet Module (VEM) – virtual line card or module within the virtual switch that VMs connect into. VEM is embedded in each VMware vSphere host and replaces the VMware Virtual Switch (vSwitch) functionality.
- Operating inside the VMware ESXi hypervisor, Cisco Nexus 1000V VEM uses the VMware vNetwork Distributed Switch (vDS) API, jointly developed by Cisco and VMware to provide policy-based VM connectivity, Layer 2 switching and advanced networking functions. The tight integration makes Cisco Nexus 1000V fully aware of key virtualization features such as VMware vMotion and Distributed Resource Scheduler (DRS). VEM provides switching functionality based on the configuration information it receives from the VSM. In the event of a communication loss with the VSM, VEM continues to forward traffic based on last known configuration or Nonstop Forwarding (NSF). VEM therefore provides reliability and advanced switching functionality.

Figure 17 illustrates a view of the Cisco Nexus 1000v as interpreted through the VEM running on the ESXi host. With the Nexus 1000v Distributed Virtual Switch, multiple pairs of differing uplinks are permitted. Port-profiles and port-profile uplinks are uniform across all VEMs added to the Virtual Supervisor Module VSM.

Figure 17 Nexus 1000v view of the VEM on a B200 M4 server



Best practices for the Cisco Nexus 1000v include:

- Deploying the VSM as an active/standby pair, set to some anti-affinity rule to not run each VSM on the same host.
- Setting System VLANs for network that will be needed during ESXi boot up to include management or iSCSI networks.
- Hosting the VSM on a physical appliance like the Cisco CSP 2100 or a separate virtualized cluster independent of the ESXi hosts registering VEM modules with the VSM.

Cisco Nexus 9000 Series vPC Best Practices

The following Cisco Nexus 9000 design best practices and recommendations were used in the FlashStack design.

vPC Peer Keepalive Link Considerations

- It is recommended to have a dedicated 1Gbps layer 3 link for vPC peer keepalive, followed by out-of-band management interface (mgmt0) and lastly, routing the peer keepalive link over an existing Layer3 infrastructure between the existing vPC peers.

- vPC peer keepalive link should not be routed over a vPC peer-link.
- The out-of-band management network is used as the vPC peer keepalive link in this FlashStack design.

vPC Peer Link Considerations

- Only vPC VLANs are allowed on the vPC peer-links. For deployments that require non-vPC VLAN traffic to be exchanged between vPC peer switches, deploy a separate Layer 2 link for this traffic.
- Only required VLANs are allowed on the vPC peer links and member ports – prune all others to minimize internal resource consumption.
- Ports from different line cards should be used to provide redundancy for vPC peer links if using a modular switch model.

vPC General Considerations

- vPC peer switches deployed using same bridge-id and spanning tree VLAN priority by configuring the peer-switch command on both vPC peer switches. This feature improves convergence and allows peer switches to appear as a single spanning-tree root in the Layer 2 topology.
- vPC role priority specified on both Cisco Nexus peer switches. vPC role priority determines which switch will be primary and which one will be secondary. The device with the lower value will become the primary. By default, this value is 32677. Cisco recommends that the default be changed on both switches. Primary vPC devices are responsible for BPDU and ARP processing. Secondary vPC devices are responsible for shutting down member ports and VLAN interfaces when peer-links fail.
- vPC convergence time of 30s (default) was used to give routing protocol enough time to converge post-reboot. The default value can be changed using delay-restore <1-3600> and delay-restore interface-VLAN <1-3600> commands. If used, this value should be changed globally on both peer switches to meet the needs of your deployment.
- vPC peer switches enabled as peer-gateways using peer-gateway command on both devices. Peer-gateway allows a vPC switch to act as the active gateway for packets that are addressed to the router MAC address of the vPC peer allowing vPC peers to forward traffic.
- vPC auto-recovery enabled to provide a backup mechanism in the event of a vPC peer-link failure due to vPC primary peer device failure or if both switches reload but only one comes back up. This feature allows the one peer to assume the other is not functional and restore the vPC after a default delay of 240s. This needs to be enabled on both switches. The time to wait before a peer restores the vPC can be changed using the command: auto-recovery reload-delay <240-3600>.
- Cisco NX-OS can synchronize ARP tables between vPC peers using the vPC peer links. This is done using a reliable transport mechanism that the Cisco Fabric Services over Ethernet (CFS over E) protocol provides. For faster convergence of address tables between vPC peers, ip arp synchronize command was enabled on both peer devices in this FlashStack design.

vPC Member Link Considerations

- LACP used for port channels in the vPC. LACP should be used when possible for graceful failover and protection from misconfigurations
- LACP mode active-active used on both sides of the port channels in the vPC. LACP active-active is recommended, followed by active-passive mode and manual or static bundling if the access device does not support LACP. Port-channel in mode active-active is preferred as it initiates more quickly than port-channel in mode active-passive.
- LACP graceful-convergence disabled on port-channels going to Cisco UCS FI. LACP graceful-convergence is ON by default and should be enabled when the downstream access switch is a Cisco Nexus device and disabled if it is not.

- Only required VLANs are allowed on the vPC peer links and member ports – prune all others to minimize internal resource consumption.
- Source-destination IP, L4 port and VLAN are used load-balancing hashing algorithm for port-channels. This improves fair usage of all member ports forming the port-channel. The default hashing algorithm is source-destination IP and L4 port.

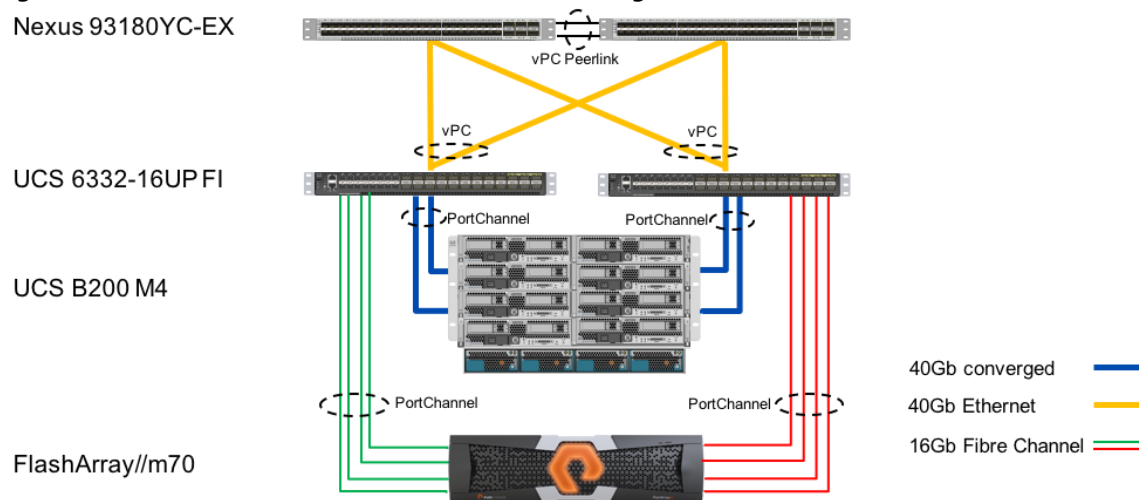
vPC Spanning Tree Considerations

- The spanning tree priority was not modified. Peer-switch (part of vPC configuration) is enabled which allows both switches to act as root for the VLANs
- Loopguard is disabled by default
- BPDU guard and filtering are enabled by default
- Bridge assurance is only enabled on the vPC Peer Link.

Direct Attached Storage Design (Alternative Design)

The Cisco UCS 6300 and 6200 Fabric Interconnect models with Unified Ports can support direct connectivity to a Pure Storage FlashArray in greenfield or existing infrastructure deployments that don't have dependent SAN network needs. This model will remove the Cisco MDS sitting between the Fabric Interconnects and FlashArray, resulting in the removal of a hop for the fibre channel traffic, but will limit options for bringing in access to other SAN resources or extending the FlashArray to systems outside of the Cisco UCS domain managed by the Fabric Interconnects.

Figure 18 FlashStack Direct Attached Alternative Design



A direct attached solution is enabled by configuring the Fabric Interconnect to FC Switching Mode, which will result in fibre channel switching occurring on the Fabric Interconnect. With FC Switching Mode in place, Fabric Interconnect Unified Ports that have been set to be FC Ports for directly connecting to the FlashArray will be set as FC Storage Port and configured to the VSAN appropriate to the fabric.

The Fabric Interconnect will automatically create zones and zonesets within the Fabric Interconnect based upon Service Profile association of the vHBA Initiator Group or SAN Connectivity Policy.

Validation

A high-level summary of the FlashStack validation is provided in this section. The solution was validated for basic data forwarding by deploying virtual machine running IOMeter tool. The system was validated for resiliency by failing various aspects of the system under load. Examples of the types of tests executed include:

- Failure and recovery FC booted ESXi hosts in a cluster
- Rebooting of FC booted hosts
- Service Profile migration between blades
- Failure of partial and complete IOM links
- Failure and recovery of redundant links to FlashArray controllers
- SSD removal to trigger an aggregate rebuild
- Storage link failure between one of the FlashArray controllers and the fabric interconnect
- Load was generated using IOMeter tool and different IO profiles were used to reflect the different profiles that are seen in customer networks

Validated Hardware and Software

The table below lists the hardware and software versions used during solution validation. Each of versions have been used have been certified within interoperability matrixes supported by Cisco, Pure Storage, and VMware. For more current supported version information, consult the following sources:

- Cisco UCS Hardware and Software Interoperability Tool: <http://www.cisco.com/web/techdoc/ucs/interoperability/matrix/matrix.html>
- Pure Storage Interoperability(note, this interoperability list will require a support login form Pure): https://support.purestorage.com/FlashArray/Getting_Started/Compatibility_Matrix
- VMware Compatibility Guide: <http://www.vmware.com/resources/compatibility/search.php>

Additionally, it is also strongly suggested to align FlashStack deployments with the recommended releases for Cisco MDS 9000 Series Switches and Cisco Nexus 9000 switches used in the architecture:

- MDS: https://www.cisco.com/c/en/us/td/docs/switches/datacenter/mds9000/sw/b_MDS_NX-OS_Recommended_Releases.html
- Nexus: https://www.cisco.com/c/en/us/td/docs/switches/datacenter/nexus9000/sw/recommended_release/b_Minimum_and_Recommended_Cisco_NX-OS_Releases_for_Cisco_Nexus_9000_Series_Switches.html

Layer	Device	Image
Compute	Cisco UCS 6332-16UP Fabric Interconnect	3.1(2b)
	Cisco UCS B-200 M4	3.1(2b)
	Cisco eNIC	2.3.0.10
	Cisco fNIC	1.6.0.28
Network	Cisco Nexus 93180YC-EX NX-OS	7.0(3)I4(2)

Layer	Device	Image
	Cisco Nexus 1000v	5.2(1)SV3(2.1)
Storage	Cisco MDS 9148S	7.3(0)DY(1)
	Pure Storage FlashArray //m70	4.7.4
Software	Cisco UCS Manager	3.1(2b)
	Cisco UCS Director	5.5
	Cisco Virtual Switch Update Manager	2.0
	VMware vSphere ESXi	6.0 U2
	VMware vCenter	6.0 U2
	Pure Storage vSphere Web Client Plugin	4.7.4

Summary

FlashStack delivers a platform for Enterprise and cloud datacenters using Cisco UCS Blade Servers, Cisco Fabric Interconnects, Cisco Nexus 9000 switches, Cisco MDS switches and fibre channel-attached Pure Storage FlashArray//M. FlashStack is designed and validated using compute, network and storage best practices for high performance, high availability, and simplicity in implementation and management.

This CVD validates the design, performance, management, scalability, and resilience that FlashStack provides to customers.

Reference Sources for Components in this Design

Products and Solutions

Cisco Unified Computing System:

<http://www.cisco.com/c/en/us/products/servers-unified-computing/index.html>

Cisco UCS 6300 Series Fabric Interconnects:

<http://www.cisco.com/c/en/us/products/servers-unified-computing/ucs-6300-series-fabric-interconnects/index.html>

Cisco UCS 5100 Series Blade Server Chassis:

<http://www.cisco.com/c/en/us/products/servers-unified-computing/ucs-5100-series-blade-server-chassis/index.html>

Cisco UCS B-Series Blade Servers:

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

Cisco UCS Adapters:

<http://www.cisco.com/c/en/us/products/interfaces-modules/unified-computing-system-adapters/index.html>

Cisco UCS Manager:

<http://www.cisco.com/c/en/us/products/servers-unified-computing/ucs-manager/index.html>

Cisco Nexus 1000v:

<http://www.cisco.com/c/en/us/products/switches/nexus-1000v-switch-vmware-vsphere/index.html>

Cisco Nexus 9000 Series Switches:

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

Cisco MDS 9000 Series Multilayer Switches:

<http://www.cisco.com/c/en/us/products/storage-networking/mds-9000-series-multilayer-switches/index.html>

Cisco UCS Director:

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

VMware vCenter Server:

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

VMware vSphere:

<https://www.vmware.com/products/vsphere>

About the Authors

Ramesh Isaac, Technical Marketing Engineer, Cisco Systems, Inc.

Ramesh Isaac is a Technical Marketing Engineer in the Cisco UCS Data Center Solutions Group. Ramesh has worked in data center and mixed-use lab settings since 1995. He started in information technology supporting UNIX environments and focused on designing and implementing multi-tenant virtualization solutions in Cisco labs before entering Technical Marketing. Ramesh holds certifications from Cisco, VMware, and Red Hat.

Cody Hosterman, Technical Director for Virtualization Ecosystem Integration, Pure Storage

Cody Hosterman focuses on the core VMware vSphere virtualization platform, VMware cloud and management applications and 3rd party products. He has a deep background in virtualization and storage technologies, including experience as a Solutions Engineer and Principal Virtualization Technologist. In his current position, he is responsible for VMware integration strategy, best practices, and developing new integrations and documentation. Cody has over 7 years of experience in virtualization and storage in various technical capacities. He is a VMware vExpert, and holds a bachelor's degree from Pennsylvania State University in Information Sciences and Technology.

Acknowledgements

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

- John George, Cisco Systems, Inc.
- Archana Sharma, Cisco Systems, Inc.