White Paper

September 2011

Midmarket Data Center Architecture: Cisco Unified Computing System with the Cisco Nexus 1000V Switch

Abstract

The Cisco® midmarket architecture is designed to meet the needs of businesses with 100 to 200 servers per data center. The Cisco Unified Computing System™ (Cisco UCS™) gives midmarket organizations access to the first converged system available anywhere, and one that goes beyond just increasing efficiency: it helps administrators become more effective in their mission, freeing them to spend less time on manual, tedious, repetitive chores and more time addressing strategic business initiatives. This document describes how to configure Cisco UCS as a virtualized platform running VMware vSphere software and the Cisco Nexus® 1000V Switch. It addresses how to integrate the system into a midmarket network architecture, and how to automate server configuration within the system. The resulting architecture results in a server, storage, and network design that is automatically configured with highly reliable, active-active network links and rapid convergence times in the event of a network failure.

Midmarket Data Center Architecture: Cisco Unified Computing System with the Cisco Nexus 1000V Switch

White Paper
September 2011

Introduction

Midmarket organizations have data centers with 100 to 200 servers. Although these organizations operate on a significantly smaller scale than their large-enterprise counterparts, they share some of the same challenges. They need cost-effective solutions that do not sacrifice scalability or investment protection for price. They need solutions that help them untangle the complex infrastructure that has grown over time, simplifying and consolidating so they can ease the burdens of overworked administrative staff. They need efficient, flexible solutions that help them quickly align their IT infrastructure with the goals of the business.

The Cisco® midmarket data center architecture is designed to meet the needs of these medium-sized businesses. It combines a comprehensive network and storage network architecture with the power and flexibility of the Cisco Unified Computing System™ (Cisco UCS™). Combined, the two elements deliver an architecture that is simplified, cost-effective, scalable, and flexible.

- The Cisco midmarket architecture simplifies data center networking by collapsing multiple layers down to two tiers with strict separation between Layer 2 and Layer 3 functions. Multiple redundant IP and storage networks are consolidated into a single unified fabric, reducing capital and operational costs while simplifying administration. The network makes more effective use of bandwidth and increases resiliency through the use of active-active links that are enabled by Cisco virtual PortChannel (vPC) technology. The Cisco Fabric Extender Architecture (FEX Architecture) physically distributes the access layer to individual server racks, while logically remaining part of a single pair of Cisco Nexus® 5000 Series Switches. Cisco FEX Architecture supports all common Ethernet network speeds in use today while easing the transition to 10 Gigabit Ethernet. Cisco VN-Link technology combines the scalability of virtual network environments with the visibility and control of physical ones, making it simple and easy to connect virtual links directly to virtual machines and manage them as if they were physical.
- Cisco UCS is a single converged system that is entirely configurable through unified, model-based management to power enterprise-class applications

and services, whether they run in bare-metal, virtualized, or cloud-computing environments. The system builds on Cisco's strengths in enterprise networking by integrating network and x86-architecture computing resources into a single system that is based on industry standards. For midmarket organizations, Cisco UCS provides a flexible, scalable pool of resources in which any server can be put to use for any workload on demand, with click-of-the-mouse simplicity. The system supports scalability without complexity through a simplified architecture that dramatically reduces the number of network interface cards (NICs), host-bus adapters (HBAs), cables, and upstream switch ports required. The result is a powerful tool for midmarket organizations that increases not just efficiency, but also the effectiveness of every resource, from capital equipment to IT subject-matter experts.

- This white paper discusses how to integrate the Cisco Unified Computing System into the Cisco midmarket architecture. It begins with a brief overview of Cisco UCS. This is followed by a detailed discussion of how the system is connected to the midmarket architecture and how it is configured internally. Although Cisco UCS powers both virtualized and nonvirtualized workloads, this document focuses on virtualized environments supported by VMware vSphere software and the Cisco Nexus 1000V Switch. The configuration described in this document uses Cisco UCS servers equipped with converged network adapters (CNAs). Similar configuration approaches can be used with Cisco virtual interface cards (VICs) to achieve even greater flexibility and performance.

Cisco Unified Computing System

Cisco UCS integrates a low-latency unified network fabric with enterprise-class x86-architecture servers to create an integrated, scalable, multichassis platform in which all resources participate in a unified, model-based management domain. A single system scales up to 40 blade server chassis, 320 computing nodes, and thousands of virtual machines. The innovations embodied in Cisco UCS help midmarket IT departments be more efficient and more effective in their missions.

Industry-Standard x86-Architecture Servers

The system's x86-architecture rack-mount and blade servers are powered by Intel® Xeon® processors. These industry-standard servers deliver world-record performance to power the full complement of enterprise applications and services. Cisco Extended Memory Technology expands the capability of 2-socket servers, supporting higher density for virtualized environments such as virtual desktop environments and improving performance for large database and data warehouse workloads. Cisco servers, combined with the system's simplified architecture, help increase IT productivity and provide superior price-to-performance value for improved total cost of ownership (TCO).

Unified Fabric for Unified Access to I/O Resources

Cisco interconnects the server resources with a standards-based, high-bandwidth, low-latency, virtualization-aware unified fabric. The fabric is wired once to support the desired bandwidth, and it can carry all network, storage, interprocess communication, and virtual machine traffic with security isolation, visibility, and control all the way to individual virtual machines. Like the rest of the Cisco midmarket architecture, the system's unified fabric brings uniform access to

network and storage resources with capabilities such as access to Fibre Channel storage configured through software, without the need to purchase additional network components. The unified fabric meets the I/O demands of today's multicore processors; eliminates costly redundancy; and increases workload agility, reliability, and performance.

Smart Infrastructure for Exceptional Agility

The system is programmable infrastructure, making it like a physical instantiation of a cloud: resources are abstracted to a pool that can be applied dynamically to any workload. Every aspect of the hardware and network configuration is programmable, allowing resources to be configured and provisioned by applying identity and personality to them. This approach gives servers personality and identity, and the parameters that can be programmed include MAC addresses and worldwide names (WWNs), firmware revisions, BIOS settings, RAID controller settings, network profiles, and storage connectivity profiles. This fundamental aspect of the system supports exceptional business agility because it enables faster deployment, redeployment, and scaling, increasing availability as well as IT productivity.

Integrated, Model-Based Management

The systems' integrated, model-based management maintains the identity and personality of every device in the system. Cisco UCS Manager aggregates element management for every device in the hardware and networking stack. It integrates all components into the single system and reliably automates configuration management. Administrators use a GUI to manipulate a model representing the system, and the system is configured as a side effect of associating a model with a physical resource.

The identity and personality for a single server and the description of its network connectivity are embodied in a Cisco service profile: applying a Cisco service profile to a server completely and accurately defines the server's personality, configuration, and network connectivity. This process can be automated through the use of service profile templates that dictate how to use predefined policies to create multiple Cisco service profiles. Subject-matter experts can define standard, compliant policies for different types of servers (web, application, and database servers, for example), and any authorized administrator can invoke the policies to provision specific servers. When service profiles are defined as updating, any subsequent changes to the policies that they invoke result in the application of changes to the running configuration, significantly easing the task of maintaining consistency and compliance in the data center. Adapter templates embody policies for configuring LAN and SAN connectivity in I/O adapters, accelerating the creation of consistent and compliant I/O configurations.

Cisco UCS Manager's role- and policy-based design preserves existing server, storage, and network administrator roles, and its standards-based XML API allows the system to be managed by third-party systems. This integrated, unified approach increases IT staff productivity, improves compliance, and reduces the chance of errors that can cause downtime. It allows the system to be treated as a single entity, allowing 100 servers to be configured with the same ease as configuring a single one.

Cisco Fabric Extender Architecture: Scalability Without Complexity

Cisco UCS scales without complexity through the Cisco FEX Architecture. Traditional blade server chassis require a set of switches, adding cost and complexity by requiring all the infrastructure that would normally be in a rack to support only 8 to 16 blade servers. In contrast, Cisco uses fabric extenders to bring fabric interconnect ports directly to each blade server. Cisco fabric extenders are implemented as virtual remote line cards: they are logically part of the fabric interconnect, yet physically part of the blade chassis or server rack. Cisco fabric extenders forward all traffic up to the fabric interconnects, and they are designed for lossless operation.

Similarly, Cisco VICs extend fabric interconnect ports directly to virtual machines. These mezzanine-format interface cards are prestandard implementations of IEEE 802.1Qbh. They support a programmable number of virtual devices whose type and number are configured on demand through Cisco UCS Manager. The card can be configured with static interfaces such as NICs and HBAs to support traditional operating systems and hypervisors, and dynamic interfaces can be created on demand to connect directly to virtual machines.

This document describes a software approach that also gives visibility and control of the access layer directly to virtual machines: the Cisco Nexus 1000V Switch. This fully featured Cisco switch is the first to use the VMware Distributed Virtual Switch (DVS) interface. It extends the access layer directly to virtual machines so that their network profiles can be managed, secured, and moved between physical machines with the same level of security as physical links. The switch can be used with CNAs (as described in this document) and with Cisco VICs.

Efficient and Ready for the Future

Cisco UCS is highly efficient; requires fewer servers, switches, adapters, and cables; consumes less power; and generates less heat than traditional, manually assembled systems. It is a system ready for the future, with power and cooling infrastructure prepared to handle processors that dissipate as much as 130 watts (W) each, and the chassis midplane scales to support 80 Gbps of Ethernet bandwidth per server. These features improve TCO while increasing performance and availability and protecting investments.

System Overview

The Cisco Unified Computing System includes both blade and rack-mount servers (Figure 1). The system's 10-Gbps unified fabric and embedded management system is supported by Cisco UCS 6100 and 6200 Series Fabric Interconnects, usually deployed in a redundant pair. Cisco UCS B-Series Blade Servers are integrated into the system through a combined data and management plane. Cisco UCS C-Series Rack-Mount Servers are integrated into the system through separate data and management connections. The testing and validation of the midmarket architecture was performed using blade servers, and the architecture as it pertains to blade servers is described in detail in this section.

Cisco UCS 6100 and 6200 Series Fabric Interconnects

Cisco UCS 6100 and 6200 Series Fabric Interconnects are families of line-rate, low-latency, lossless 10 Gigabit Ethernet and Fibre Channel over Ethernet (FCoE)

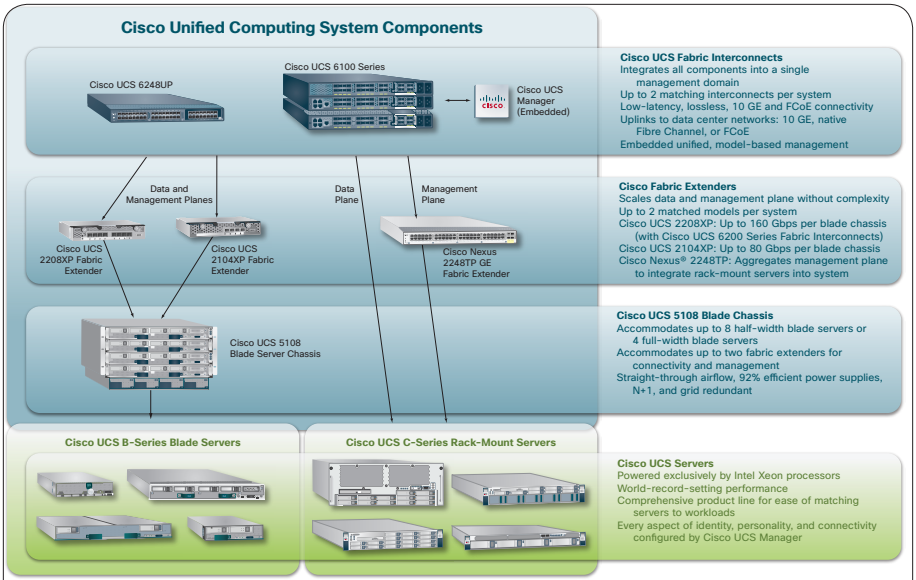

Figure 1. Cisco Unified Computing System

interconnects that consolidate all I/O at the rack level. Based on the same switching technology as the Cisco Nexus 5000 Series Switches, the fabric interconnects provide additional features and management capabilities to support Cisco UCS.

Three versions of the fabric interconnects are available:

- The Cisco UCS 6248UP 48-Port Fabric Interconnect supports 32 fixed ports and a single expansion card that brings the unit's capacity to 48 unified ports in a single rack unit. Each unified port is capable of line-rate, low-latency, lossless 1 and 10 Gigabit Ethernet, FCoE, and up to 8-Gbps Fibre Channel connectivity.
- The Cisco UCS 6120XP 20-Port Fabric Interconnect supports 20 fixed ports and a single expansion module. The Cisco UCS 6100 Series Fabric Interconnects support expansion modules to increase the system's Ethernet connectivity, connect to native Fibre Channel networks (up to 8 Gbps), or both. The four available expansion modules support 8-port 2/4/8-Gbps Fibre Channel, 4-port Ethernet plus 4-port Fibre Channel, 6-port 10 Gigabit Ethernet, and 6-port 2/4/8-Gbps Fibre Channel.
- The Cisco UCS 6140XP 40-Port Fabric Interconnect supports 40 fixed ports and two expansion modules.

Cisco UCS 5108 Blade Server Chassis

The system's blade server chassis is logically part of the system's fabric interconnects, creating a single management domain. The Cisco UCS 5108 Blade Server Chassis extends the system's capacity without complexity because the chassis is logically part of the fabric interconnects and is not a point of management in itself. The chassis has only five basic components, with all but the midplane hot pluggable and user serviceable. The chassis has flexible partitioning with removable dividers that can support any combination of two different server form factors: half-width and full-width blades.

Each of the server's half slots provides power and I/O connectivity. Each half slot is wired with four 10 Gigabit Ethernet connections to each of the two fabric extender slots. The Cisco UCS 2014XP Fabric Extenders provide up to 20-Gbps of bandwidth per half slot. The Cisco UCS 2208XP Fabric Extender brings up to 80 Gbps of bandwidth to each half slot, or up to 160 Gbps per full-width blade.

Cisco Fabric Extenders

Up to two fabric extenders per chassis bring the unified fabric into the blade server chassis. They pass all network traffic up to the fabric interconnects, eliminating blade server switches as a point of management. They are treated by the fabric interconnects as remote line cards, making them logically part of the interconnects but physically distributed across a system's blade server chassis. This design contributes to the system's scalability and reduced complexity: by forwarding all traffic to the parent fabric interconnects, the fabric extenders eliminate a layer of switching in blade server chassis, allowing the interconnects to consistently and uniformly switch all network traffic for the system. The fabric extenders themselves are configured implicitly by the fabric interconnects.

Each Cisco UCS 2208XP Fabric Extender supports up to eight 10-Gbps uplinks to the fabric interconnects and is connected through the chassis midplane to each of the chassis' eight half slots. The Cisco UCS 2100 Series Fabric Extenders support up to four 10-Gbps uplinks each. The system is designed so that it is wired once to support the desired bandwidth, with features such as FCoE managed through software. One fabric extender supports what this document refers to as the A side, and the other supports the B side. Cisco fabric extenders also provide management-plane connectivity that allows the fabric interconnects to manage every aspect of server, chassis, and fabric extender configuration and operation.

Cisco UCS B-Series Blade Servers

Cisco UCS B-Series Blade Servers are industry-standard x86-architecture servers based on powerful Intel Xeon processors. The breadth of Cisco's blade server product line helps organizations provide the best match of the processing power, memory capacity, and I/O configuration of their server resources to the workloads to be supported.

- The Cisco UCS B200 M2 Blade Server is a 2-socket, half-width blade server that hosts up to two Intel Xeon 5600 series processors for up to 12 cores, up to 96 GB of main memory (based on 8-GB DIMMs), and two small-form-factor (SFF) disk drives with integrated RAID.
- The Cisco UCS B250 M2 Extended Memory Blade Server is a 2-socket, full-width blade server. It can be configured with up to two Intel Xeon 5600 series processors for up to 12 cores. This server supports the large memory footprint required by many virtualized and large-data-set environments without imposing the cost of purchasing a 4-socket server when only additional memory capacity is needed. The system supports an economical 192-GB memory footprint using low-cost 4-GB DIMMs, or the largest memory footprint available in any 2-socket server (384 GB) using 8-GB DIMMs. The system supports two SFF disk drives and integrated RAID.
- The Cisco UCS B230 M1 Blade Server is a 2-socket, half-width blade server that hosts up to two Intel Xeon 7500 series processors, up to 256 GB of memory, and up to two SFF disk drives with integrated RAID.

- The Cisco UCS B440 M1 High-Performance Blade Server is a 4-socket, full-width blade server that hosts up to four Intel Xeon 7500 series processors, up to 512 GB of memory, and up to four SFF disk drives with integrated RAID.

Each half-width blade server supports a single mezzanine card that interfaces with each of the two 10-Gbps unified fabric connections on the midplane. The full-width blade servers support two mezzanine cards and a total of four midplane connections.

Cisco UCS Network Adapters

Cisco supports a range of mezzanine-format network adapters, including a 10 Gigabit Ethernet network adapter designed for efficiency and performance, the Cisco UCS M81KR VIC designed with Cisco Data Center Virtual Machine Fabric Extender (VM-FEX) technology to deliver the system's most comprehensive support for virtualized I/O, and a set of Cisco UCS M71KR CNAs designed for full compatibility with existing Ethernet and Fibre Channel environments.

Both CNAs and the VIC make the existence of the unified fabric transparent to the operating system. They present both Ethernet NICs and Fibre Channel HBAs to the system's PCIe bus and pass Fibre Channel traffic onto the unified fabric through FCoE protocols. Regardless of the adapter chosen for a Cisco UCS B-Series Blade Server, all configuration and management is handled by Cisco UCS Manager, eliminating the need for administrators to configure individual servers using separate tools for each type of interface. Adapter templates can be created in Cisco UCS Manager to define the policy that governs the way that an adapter is configured, with the act of configuring an adapter using the template essentially guaranteed to create a compliant configuration. When an adapter template is set as updating, a change to the template immediately updates the configuration in all the adapters that use that template.

Using CNAs with VMware vSphere

Converged network adapters place two 10 Gigabit Ethernet NICs and two Fibre Channel HBAs on the server's PCIe bus. The NICs and HBAs are connected to alternate fabrics so that both the A and B sides are accessible, depending on which interface is addressed.

When CNAs are used with VMware vSphere software, the Fibre Channel HBAs provide access to the shared storage that VMware recommends for virtual disk storage. The Ethernet NICs provide network access for both the VMware vSphere hypervisor and virtual machines. Following VMware's recommended best practices for physically separating production and hypervisor traffic, each type of traffic can be passed to a separate fabric. In the midmarket architecture, the A fabric handles all control traffic, and the B fabric handles all virtual machine production traffic.

A software switch is needed to handle traffic from multiple virtual machines and the vSphere hypervisor that passes through the CNA's two NICs. The software switch also handles local switching for inter–virtual machine traffic on the same server. Both the VMware vSwitch and the Cisco Nexus 1000V Switch can handle these functions; however, the latter is a fully featured Cisco switch that implements Cisco VN-Link technology, offering exceptional visibility and control over virtual machine network traffic.

Cisco VN-Link technology extends the network access layer directly to virtual machines so that virtual network traffic can be handled with the same level of ease as for physical links. With the Cisco Nexus 1000V Switch, all network attributes that can be assigned to a physical link can be assigned to a virtual link. Network profiles that support attributes such as quality of service (QoS) and VLAN membership can be assigned to traffic flows from virtual machines. The Cisco Nexus 1000V is a distributed virtual switch in which the switch control plane spans multiple hypervisor instances so that an administrator interacts with a single point of management to control switching across an entire pool of virtualized servers (Figure 2). The switch's virtual supervisor module (VSM) manages multiple virtual Ethernet modules (VEMs), one of which runs in each hypervisor.

One of the benefits of having a single distributed virtual switch that spans all hypervisor instances is that when virtual machines move between servers through

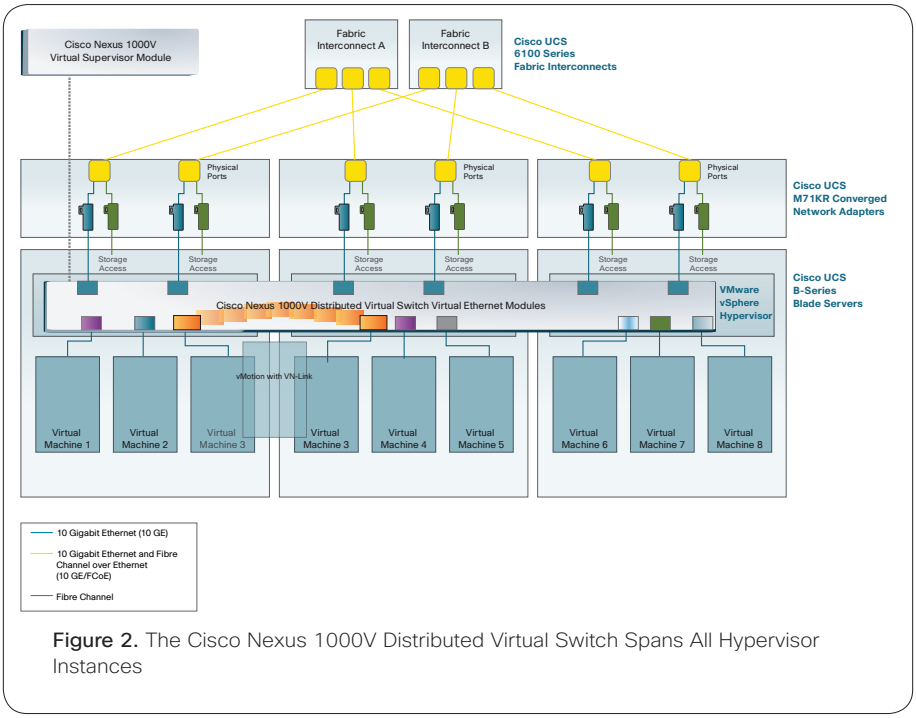

Figure 2. The Cisco Nexus 1000V Distributed Virtual Switch Spans All Hypervisor Instances

VMware vMotion, the port profiles assigned to a virtual machine move with that virtual machine. This behavior allows security attributes to be managed on a per–virtual machine basis, eliminating the need to use least-common-denominator security across the server farm.

Using Cisco Virtual Interface Cards with VMware vSphere

Cisco VICs connect up to 128 virtual devices to the server's PCI bus, with the number and type of interfaces (NIC or HBA) programmed by Cisco UCS Manager. Each virtual device can be configured to connect to either the A or the B-side fabric. The total number of devices is determined by the capabilities of the upstream fabric extender and fabric interconnects. With today's Cisco Nexus 6100 Series Fabric

Interconnects, Cisco UCS M81KR VICs can support up to 58 devices, a number sufficient to dedicate one or two NICs to each virtual machine and have a sufficient number of devices remaining to statically dedicate one to each hypervisor function such as VMware vmkernel, vmconsole, and vMotion.

The VIC can be integrated into virtualized environments so that virtual machines access the network in one of two ways (Figure 3):

- Through a software switch: The card can connect Ethernet NICs to a virtual switch in a way analogous to the CNA configuration discussed in the previous section.
- Through virtual NICs with Cisco Data Center VM-FEX technology: Each virtual machine directly accesses one or more dedicated virtual NICs through Cisco Data Center VM-FEX technology. Depending on the hypervisor, all traffic can avoid any hypervisor intervention and achieve up to a 38 percent increase in network bandwidth.

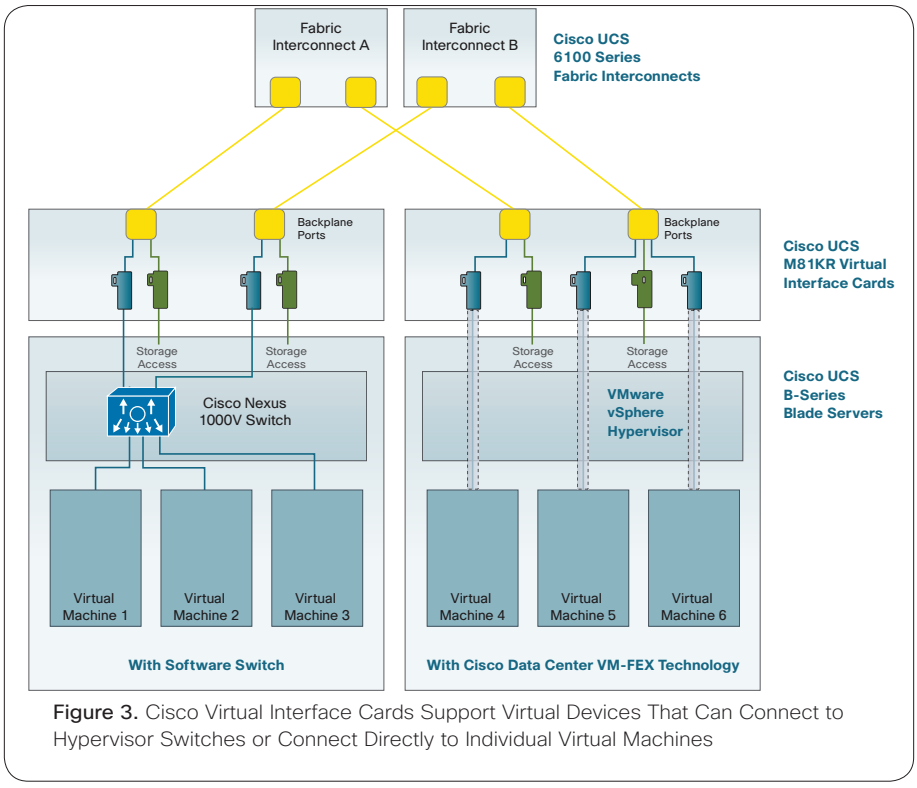

Figure 3. Cisco Virtual Interface Cards Support Virtual Devices That Can Connect to Hypervisor Switches or Connect Directly to Individual Virtual Machines

Interfaces connected to hypervisor functions or to a software switch can be configured statically so that they will always remain associated with the hypervisor regardless of the server on which it is booted. Interfaces connected directly to virtual machines can be configured dynamically so that they move, along with their port profiles, when a virtual machine is moved to another hypervisor instance.

Cisco VICs offer exceptional I/O flexibility, because they enable any server to host any operating system or hypervisor. With central, model-based management, servers can be configured to host any environment on demand. This capability increases flexibility of the virtualization pool: servers can be configured in and out of the pool depending on workload requirements and repurposed for other functions when not supporting virtualized environments.

Implementation Overview

The configuration that was tested and validated by Cisco included two Cisco UCS 6120XP fabric interconnects, each configured with an expansion model with six 4-Gbps Fibre Channel ports for SAN connectivity (Figure 4). A single Cisco UCS 5108 Blade Server Chassis was equipped with two Cisco UCS 2104XP Fabric Extenders. The chassis was loaded with two Cisco UCS B250 M1 Extended Memory Blade Servers, each with one Cisco UCS M71KR-Q QLogic CNA. Each of the servers was configured to boot the VMware vSphere hypervisor from local storage, and each hypervisor instance was configured with a Cisco Nexus 1000V VEM.

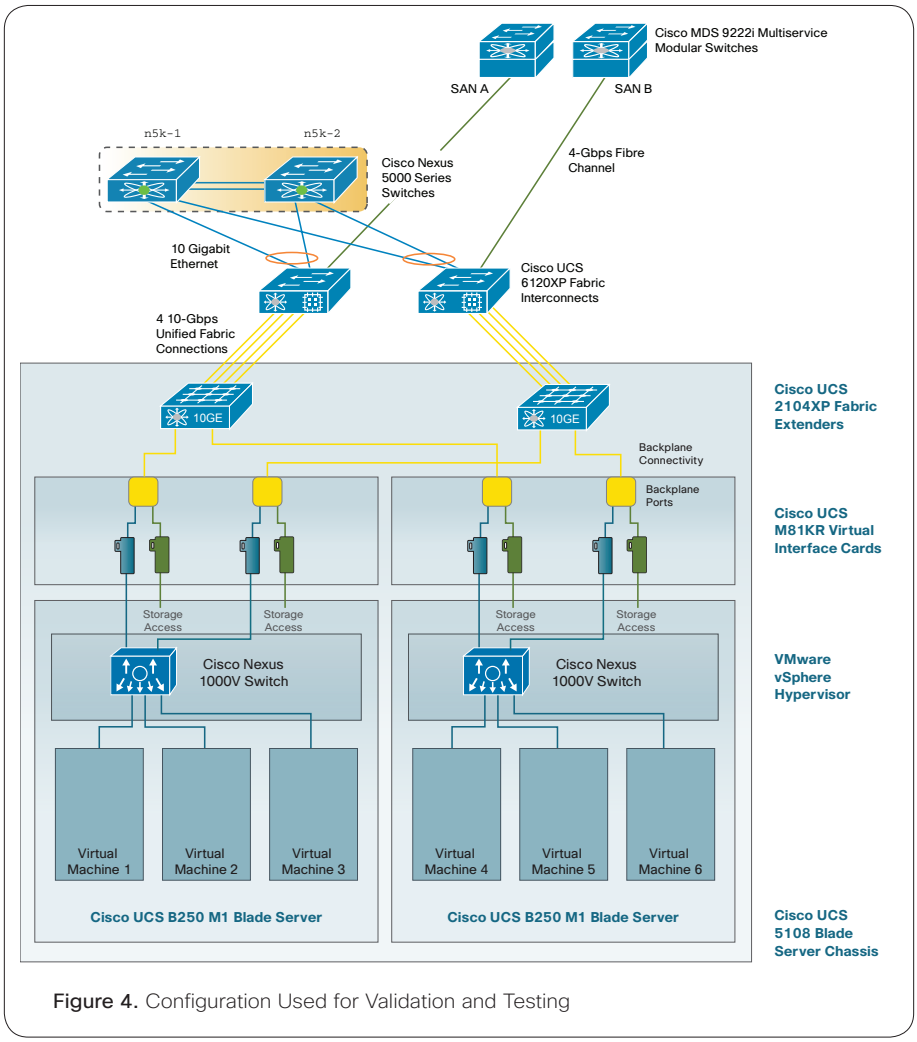

Figure 4. Configuration Used for Validation and Testing

Upstream Physical Connectivity

Each fabric interconnect was connected to the midmarket architecture's access layer through a vPC that connected each fabric interconnect to each Cisco Nexus 5000 Series Switch in an active-active configuration. SAN access was provided by connecting a single 4-Gbps Fibre Channel cable from each fabric interconnect to the Cisco MDS 9222i Multiservice Modular Switches. Following storage networking best practices, one fabric interconnect was connected to SAN A, and the other was connected to SAN B.

Internal System Physical Connectivity

Internally, each of the two blade server's CNAs connected to each of the two Cisco UCS 2104XP Fabric Extenders through the chassis midplane. In the tested configuration, all four of each fabric extender's uplinks were connected to a parent Cisco UCS 6120XP fabric interconnect, providing a total of 160 Gbps of bidirectional bandwidth between the chassis and the fabric interconnects.

Logical System Connectivity

The system's A side consists of all networking components shown on the left side of Figure 4, from the left port of a server's CNA to the left fabric extender and on to the left fabric interconnect. The A side supports VLANs for VMware vmconsole, vmkernel, and vMotion traffic as well as Cisco Nexus 1000V Switch control traffic. Although segregating all system control functions physically is not necessary because of the fine-grained control the unified fabric offers over security and class of service (CoS), this configuration was chosen to demonstrate how recommended best practices for VMware vSphere software can be implemented. The A-side network also hosts a VLAN dedicated to FCoE traffic that is extended to a physical Fibre Channel connection to SAN A.

The system's B side supports VLANs for all virtual machine production and backup traffic.

Connectivity to Midmarket Architecture

Cisco UCS interfaces with the rest of the midmarket architecture through the Cisco UCS 6100 Series Fabric Interconnects. This section provides recommendations for configuring the fabric interconnects for optimal performance, reliability, and network resiliency with respect to upstream connectivity. Recommendations include running the fabric interconnects in end-host mode, connecting Ethernet uplinks to the Cisco Nexus 5000 Series Switches using vPCs, and forwarding the system's FCoE traffic to native Fibre Channel uplinks.

End-Host Mode

Cisco UCS fabric interconnects can be configured to run in end-host mode or switched mode. The system default is end-host mode. End-host mode simplifies the way in which the system appears to the Layer 2 network in comparison to switched mode, in which the fabric interconnect acts as a normal Ethernet bridge. End-host mode is the default, and it is recommended configuration for the following reasons:

- Layer 2 is simplified from the perspective of the upstream switches because end-host mode hides the details of the system's internal network. This configuration

allows a larger Layer 2 size because of lower memory requirements on the upstream Cisco Nexus 5000 Series Switches. Likewise, fabric interconnect operation is simplified because the fabric interconnects must keep track of only the end hosts in the system. Any traffic that is not internal to the system is passed up to the access layer.

- End-host mode eliminates the need to run Spanning Tree Protocol on the fabric interconnects, limiting its use to the upstream pair of Cisco Nexus 5000 Series Switches.
- Links can be pinned upstream of the fabric interconnects as necessary to direct traffic flows rather than relying on Spanning Tree Protocol to do so.

Verify that the system is configured in end-host mode by selecting the Equipment tab in the navigation pane. Select one of the two fabric interconnects. Select the General tab in the work pane, and under Status verify that Ethernet Mode is set to end-host (Figure 5). Repeat this process for the second fabric interconnect.

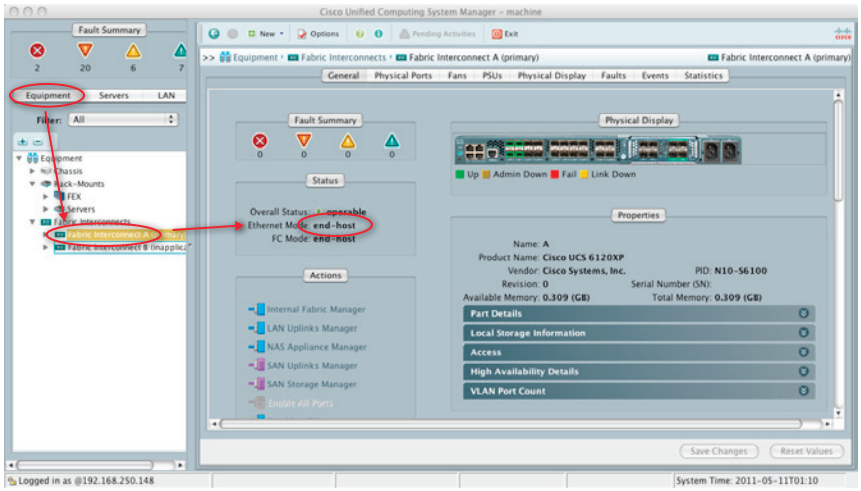

Figure 5. Verify End-Host Mode Settings

Configuration recommendations related to end-host mode include the following:

- Connect the fabric interconnects to the upstream switches that are joined through some method for multichassis EtherChannel (MCEC) such as vPC as described in the next section.
- Do not create static pin groups on the system uplinks. These allow the fabric interconnects to dynamically pin traffic on the uplinks by default. This approach reduces the requirements on the Cisco Nexus 1000V Switch and increases availability.
- With end-host mode, a single port is chosen on the fabric interconnects as a listening port for broadcast and multicast traffic. Make sure that all uplink ports are connected to the same Layer 2 domain, as they are in this configuration, so

that the single port will receive all broadcast and multicast traffic for the Layer 2 domain.

- A higher CoS for vmconsole and vmkernel traffic can be enabled within Cisco UCS to help guarantee minimum bandwidth allocation to these two traffic flows. In Cisco UCS Manager, select the LAN tab in the navigation pane and navigate to QoS System Class. In the work pane, click the Enabled check box next to the Platinum priority (Figure 6). The hypervisor must be configured to use this CoS tag for the vmconsole and vmkernel flows, and the upstream switch must also be set to accept the settings.

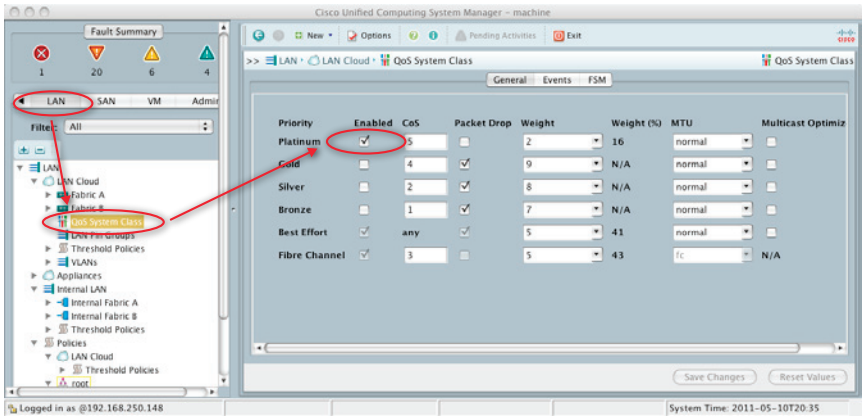

Figure 6. Enabling QoS for vmkernel and vmconsole Traffic

Virtual PortChannels

Using vPCs to manage the multiple connections from the Cisco UCS 6100 Series Fabric Interconnects to the Cisco Nexus 5000 Series Switches establishes a more robust network with better convergence times in the event of a link failure. vPCs also help scale the number of VLANs in the system. vPCs support dynamic load balancing across multiple uplinks, allowing their full aggregate bandwidth to be used, in comparison to active-standby approaches, in which standby links are used only in the event of a failure.

The Cisco UCS 6100 Series Fabric Interconnects are unaware of the existence of the vPC configured on the upstream pair of Cisco Nexus 5000 Series Switches. To the fabric interconnects, the upstream switches appear as a single switch, and the only configuration needed is establishment of a PortChannel that unites the multiple uplinks to what appears as a single switch. Link Aggregation Control Protocol (LACP) is the recommended configuration for the PortChannel because it supports dynamic load balancing on a per-flow basis. Static pinning is not needed because LACP dynamically assigns the uplink for each flow. In the event of a failure, flows are not repinned to a different uplink; instead, the LACP hashing algorithm simply moves the flow to a remaining link. This approach uses fewer switch resources and provides faster reconvergence in response to a failure.

Because either of the fabric interconnects can pass traffic from any VLAN to either of the upstream switches, the Cisco Nexus 5000 Series Switches and the Cisco UCS 6100 Series Fabric Interconnects must be prepared to receive traffic from any of the VLANs used in the system configuration.

Fibre Channel Connectivity

Within Cisco UCS, all Fibre Channel traffic destined for SAN A is passed using FCoE on a VLAN established on the system's A-side fabric, and correspondingly for SAN B.

Setting up this configuration requires setting up the Cisco UCS 6100 Series Fabric Interconnects to pass FCoE traffic from the A side to native Fibre Channel supported by an expansion module on the fabric interconnect. The same is true for the B-side FCoE traffic.

Cisco UCS is set up by default to route Fibre Channel traffic as described. To check these settings, select the Equipment tab and expand a fabric interconnect's Expansion Module 2 inventory (Figure 7). Under Uplink FC Ports, select Port 1 and observe in the work pane that Fabric dual/vsan default(1) is selected next to VSAN, under Properties. Repeat this process for the other fabric interconnect.

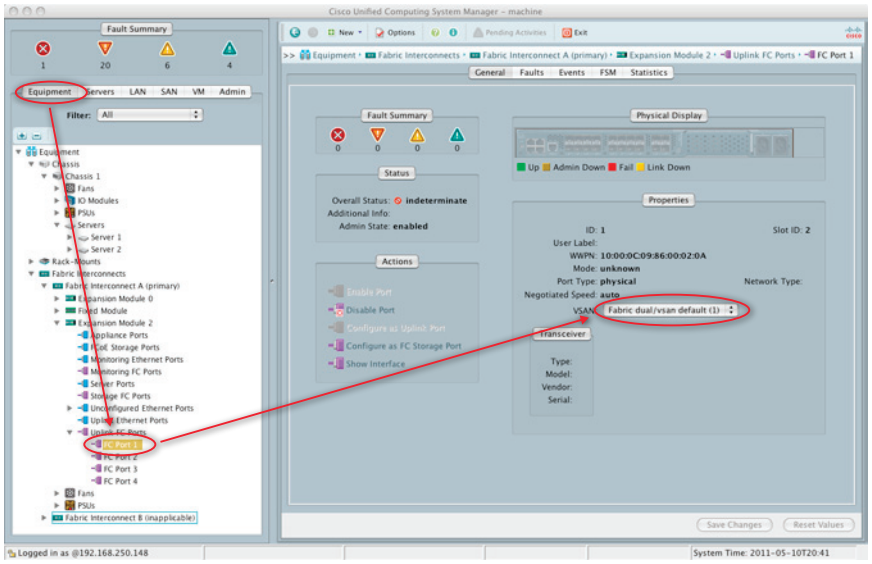

Figure 7. Associate SAN A with the System's A-Side Fabric

Cisco UCS Internal Configuration

The internal network configuration for Cisco UCS is designed to physically separate traffic streams according to recommended VMware best practices and provide high availability and performance through redundant network paths. The midmarket architecture was designed for fast failover and failback times in the event of a link failure.

Failover can be handled at three levels in the system. The fabric interconnects can redistribute traffic within the LACP-based EtherChannel in the event of a link or

upstream switch failure. Traffic for the fabric (A or B) can fail over to the other fabric interconnect in the pair if one fails. The CNAs in the blade servers can fail over in the event of a failure that brings down either the A-side or B-side fabric. In addition, the Cisco Nexus 1000V Switch can fail over connectivity between its own uplinks in the event of a NIC or fabric failure.

The goal in configuring the system's internal network for failover is to make the Cisco Nexus 1000V Switch aware of any failures so that it can use its intelligence to change the way it directs traffic flows to upstream links. This section discusses how to configure the system's internal networks to accomplish this goal.

Link-Down Network Control Policy

The system uses a link-down control policy to push notification of failures down to the Cisco Nexus 1000V Switch. This is the system's default operation. To verify the appropriate settings, select the LAN tab in the navigation pane, open Policies to reveal the default network control policy, and verify that Action on Uplink Fail in the work pane is set to link-down (Figure 8).

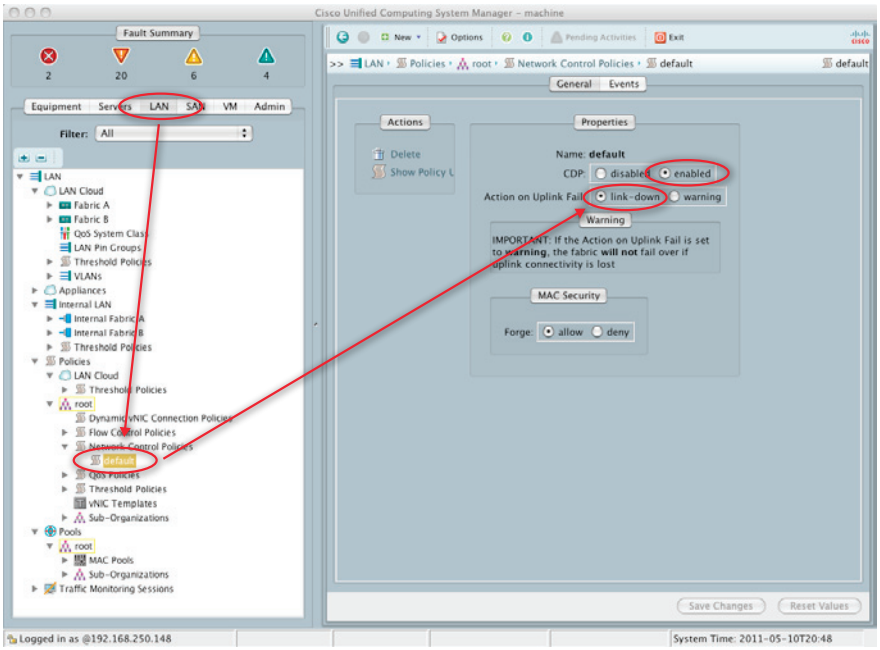

Figure 8. Setting the Link-Down Network Control Policy and Cisco Discovery Protocol

If a single link in a fabric interconnect's vPC fails, the interconnect simply moves traffic flows onto the remaining links, and no notification is pushed down. If, however, all links in a vPC fail, the fabric can no longer pass traffic to an upstream Cisco Nexus 5000 Series Switch. The link-down control policy pushes this failure notification to the blade server's can, which signals a link-down situation on the virtual NIC (vNIC) connected to the failed fabric. The Cisco Nexus 1000V Switch recognizes this situation, and it moves traffic flows to the second vNIC. The upstream Cisco Nexus 5000 Series Switch recognizes and routes traffic onto

this new network path after it starts receiving traffic from the other fabric's fabric interconnect.

The result of this policy is that production and management traffic are separated onto two different fabrics during normal operation, but in the event of a complete loss of connectivity on one fabric, the other fabric can carry both types of traffic.

Cisco Discovery Protocol can be enabled to help troubleshoot any connectivity problems on both physical and virtual adapters. Figure 8 shows Cisco Discovery Protocol enabled in Cisco UCS Manager. Administrators can assign this network control policy to adapters when they are defined in service profiles. Note that the MAC Security feature should always be set to "allow" for Cisco Nexus 1000V Switches (and any virtual switch) that will be using an adapter with this network control policy.

No Fabric Failover in CNAs

The CNA vNIC definitions in Cisco UCS Manager should not include fabric failover. As virtual machines are powered on and pass traffic, the receiving fabric interconnect learns the virtual machines' MAC addresses and stores them in its MAC address table. In the event of a link failure, the fabric interconnect gaining the additional traffic would have to learn of the new MAC addresses and send gratuitous Address Resolution Protocol (ARP) messages to the upstream Cisco Nexus 5000 Series Switch to notify it of the new path.

With fabric failover enabled, the Cisco Nexus 1000V Switch would not see any link outage and would be unaware of the reconfiguration that occurred upstream. Disabling fabric failover on the blade server's vNICs eliminates this problem and allows the traffic engineering features of the Cisco Nexus 1000V Switch to handle the failover function. These settings are described in "CNA Configuration" later in this document.

MAC Pinning

Cisco Nexus 1000V Switches can aggregate multiple uplink ports without running a bidirectional link-aggregation protocol such as Port Aggregation Protocol (PaGP) or LACP. This feature is called MAC pinning, and it is recommended when using the Cisco Nexus 1000V Switch with Cisco UCS.

With MAC pinning, the Cisco Nexus 1000V VEM assigns traffic to a group of uplink interfaces that pass the correct VLAN, and it will assign the vNIC of a given virtual machine to one of the uplinks on a per-vNIC basis. The upstream fabric interconnects are not aware of this configuration on the Cisco Nexus 1000V Switch and see only the appropriate MAC addresses on the appropriate blade servers. If a link fails, the traffic fails over to a surviving link to mitigate the outage, and the traffic is returned to the original link after it is restored (following an appropriate delay to help ensure link stability).

To configure MAC pinning on the Cisco Nexus 1000V Switch, a single option must be specified in the Cisco Nexus 1000V Series Ethernet port profile—for example:

```
port-profile type ethernet Qlogic-no-ha-uplink
…
```

```
  channel-group auto mode on mac-pinning
  …
```

In the case of the configuration tested for the purpose of this document, VMware vmconsole, vmkernel, and vMotion and Cisco Nexus 1000V Switch control traffic were assigned to VLANs associated with the A fabric, and virtual machine production traffic was assigned to a VLAN associated with the B fabric.

Preferred interfaces can be specified by assigning them as in the following configuration. This example calls for the VMware vmkernel traffic to use vmnic0 (or the A fabric) as its uplink. If the chosen vmnic is not available or does not carry the appropriate network, the switch will choose a vmnic that can meet the port profile requirements.

port-profile type vethernet VMKernelNet

```
  …
  pinning id 0
  …
port-profile type vethernet VMDataNet
  …
  pinning id 1
  …
```

The selection mechanism defines the pinning ID in a virtual Ethernet (vEthernet) port profile, which represents the virtual machine NIC (VMNIC) as reported by VMware ESX Server. Different VLANs can be specified for different traffic types: for example, for control traffic and for production traffic in the Ethernet port profiles for the physical uplink adapters. Under normal circumstances, the Cisco Nexus 1000V Switch can apply the preferences specified in port profiles. In the event of a link failure, the switch moves all traffic to the surviving uplink.

Port Profile Configuration

The Cisco Nexus 1000V uplink port profile assigned to the CNAs has a descriptive name so that, when it appears in VMware vCenter, it is easy to map the Cisco Nexus 1000V Series Ethernet ports to the appropriate uplink port groups in VMware vCenter. This port profile summarizes the recommendations in the prior sections in its specification of MAC pinning and VLAN use. The relevant portion of the Cisco Nexus 1000V Switch configuration is shown here:

```
port-profile type ethernet Qlogic-no-ha-uplink
  vmware port-group
  switchport mode trunk
  switchport trunk allowed vlan 1-6
  channel-group auto mode on mac-pinning
  no shutdown
  system vlan 2-5
  state enabled

interface Ethernet5/1
  inherit port-profile Qlogic-no-ha-uplink
```

```
interface Ethernet5/2
  inherit port-profile Qlogic-no-ha-uplink
```

QoS Configuration

QoS settings should be configured in the Cisco Nexus 1000V Switch so that QoS is established at the lowest level and is then carried up through the vNICs and the fabric interconnects.

CNA Configuration

CNAs are configured by creating a model of the desired configuration in a service profile. When a service profile is created and applied to a physical server, the server's CNA will be configured implicitly. The service profile should be created using the following guidelines:

- Define the VLANs that are initially passed to the adapter and thus to the Cisco Nexus 1000V Switch (Figure 9). Select the LAN tab in the navigator pane, right-click VLANs, and add the VLANs that are shown in Figure 9. The native (untagged) VLAN is the management network. In normal operation, VLANs 2 through 5 are used for control traffic and are supported on fabric A, and VLAN 6 is used for virtual machine traffic and is supported on fabric B.

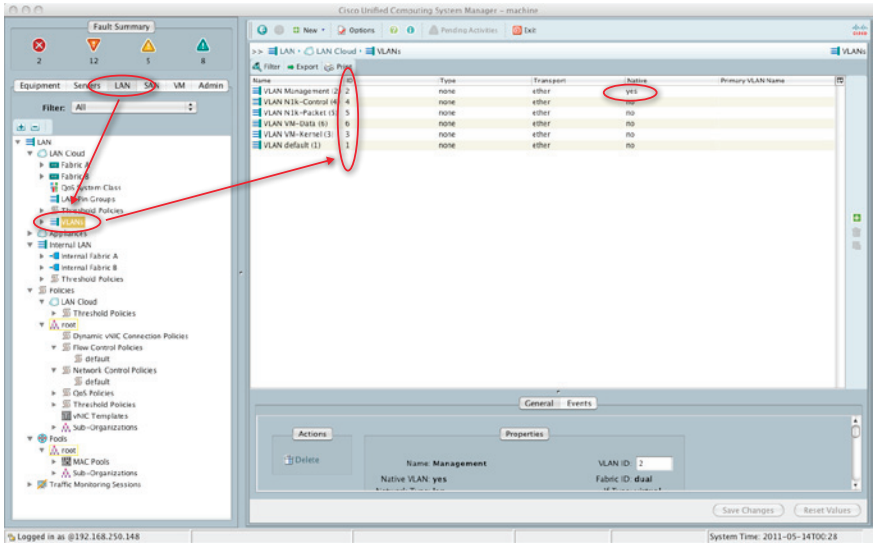

Figure 9. VLAN Definitions Used in the Test Configuration

- After the VLANs are defined, they can be assigned to the individual vNICs defined in each server's service profile. Select the Servers tab in the navigation pane, and then select the desired service profile. Select a vNIC definition and choose Modify. Use the Modify vNIC dialog box to configure each of the server's two vNICs (Figure 10):
  - Select a MAC address assignment.

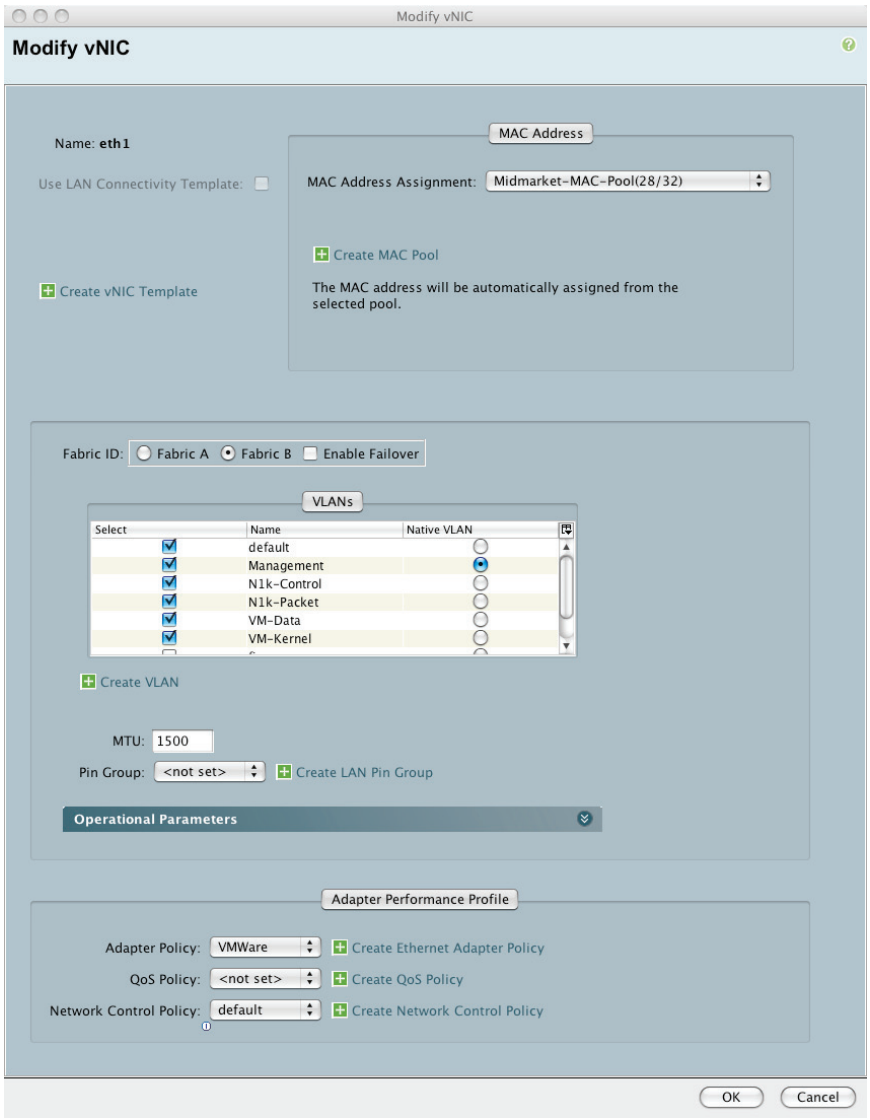

Figure 10. Use the Modify vNIC Dialog to Configure Each CNA's vNICs

- Connect vNIC eth0 to fabric A, connect vNIC eth1 to fabric B, and disable fabric failover.
- Select all the VLANs defined in the previous step and make sure that the Management VLAN is selected as the default. Note that both eth0 and eth1 should allow all VLANs so that either fabric can handle all VLANs during a failure scenario.
- Use the predefined VMware adapter policy.
- QoS is set in the Cisco Nexus 1000V Switch, so no QoS setting is required. If there is traffic for which QoS settings are not set in the Cisco Nexus 1000V Switch, a QoS profile can be selected here.

- Select the default network control policy to inherit the link-down control policy and Cisco Discover Protocol settings defined earlier.
- Cisco Discovery Protocol should be selected on the adapter policy if a management connectivity view is desired.

The network portion of the resulting service profile is shown in Figure 11. When more than one server is to be configured, the more expedient approach is to define a service profile template that embodies all the settings described in the previous sections and use the service profile template to generate a unique service profile for each server configured to run VMware vSphere software and the Cisco Nexus 1000V Switch.

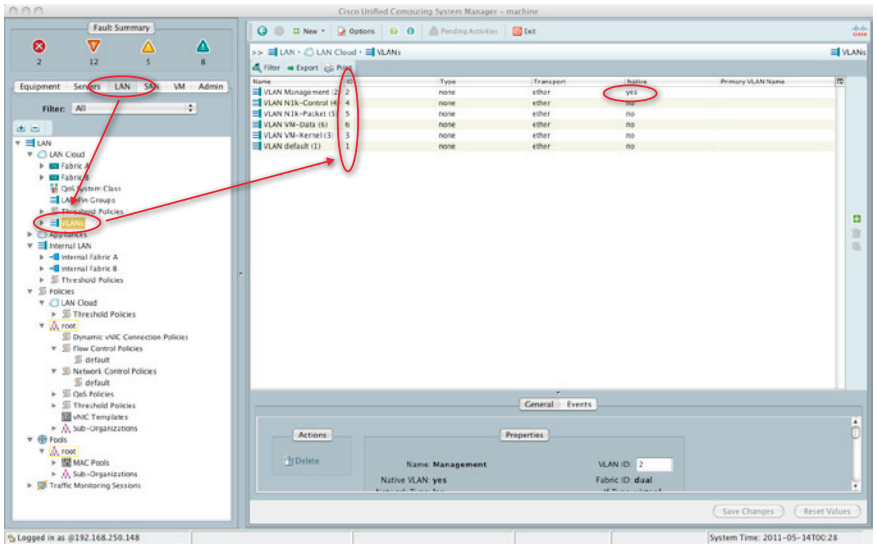

Figure 11. Network Portion of the Server's Service Profile

Network Reliability and Convergence Times

Cisco configured systems as described in this document, thoroughly testing them to be sure that the network failed over properly when errors were induced and failed back when the error condition was restored. This check was not intended to be an exhaustive test suite, but rather a verification that the expected behavior actually occurs in response to a failure, and that traffic that should not be affected by the failure was indeed not affected.

Figure 12 shows the failures that were induced, and Table 1 summarizes the observations.

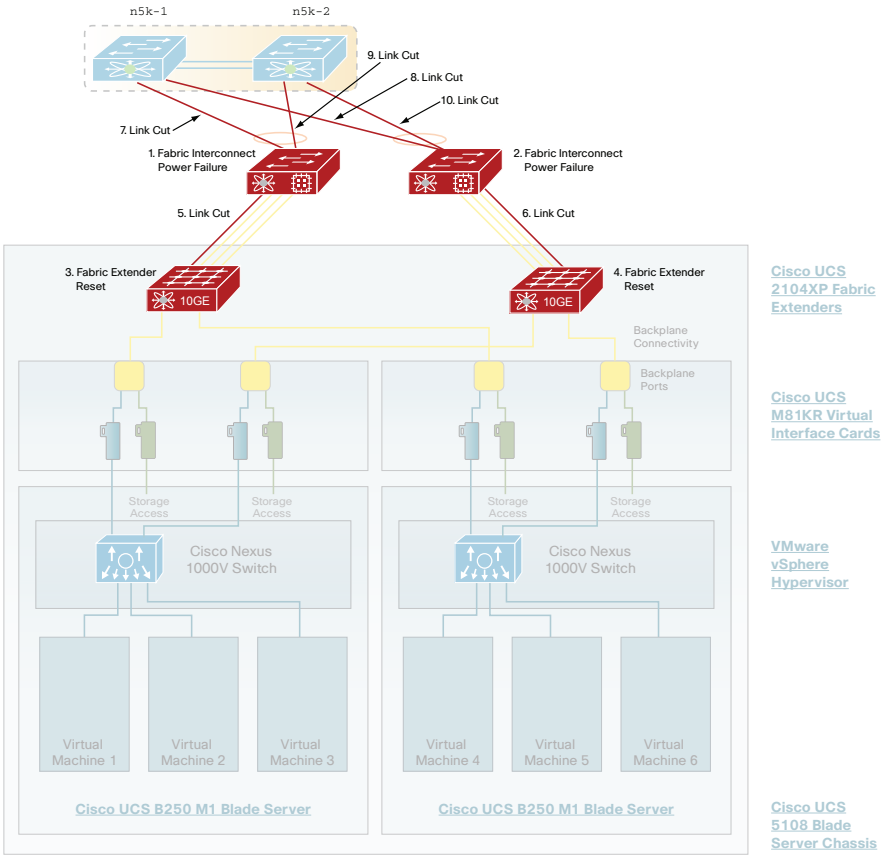

Figure 12. Location of Failures Induced for Reliability Testing

Table 1. Summary of Failure Testing Observations and Measurements

Induced Failure	A Side Carrying Control Traffic	B Side Carrying Virtual Machine Production Traffic
Fabric interconnect power failure	1. No loss of connectivity occurred between the external Cisco Nexus 1000V Switch VSM and the VEMs. No packets were lost for any virtual machine production traffic.	2. Traffic was interrupted for 1.66 seconds until the network converged onto the A side.

Induced Failure	A Side Carrying Control Traffic	B Side Carrying Virtual Machine Production Traffic
Fabric extender reset (simulating failure)	3. No control or production traffic was lost when failure was induced or restored.	4. A convergence time of 477 milliseconds (ms) was measured for virtual machine traffic that was not locally switched. No VSM-to-VEM connections were interrupted.
Link cut between fabric extender and fabric interconnect (simulated by pulling cable)	5. No VSM-to-VEM connections were lost, and no loss occurred for production traffic, demonstrating the lossless nature of Cisco UCS internal networks.	6. A convergence time of 445 ms was measured for virtual machine traffic that was not locally switched. When the link was restored, the network reconverged in 155 ms. No VSM-to-VEM connections were interrupted.
Link cut between fabric interconnect and Cisco Nexus 5000 Series Switch 1	7. No VSM-to-VEM connections were lost, and no loss occurred for production traffic. The network reconverged properly when the failure was remedied.	8. A convergence time of 138 ms was measured for virtual machine traffic that was not locally switched. When the link was restored, the network reconverged in 7.2 ms. No VSM-to-VEM connections were interrupted.
Link cut between fabric interconnect and Cisco Nexus 5000 Series Switch 2	9. No VSM-to-VEM connections were lost, and no loss occurred for production traffic. The network reconverged properly when the failure was remedied.	10. The network converged properly when the failure was induced and reconverged properly when the failure was restored. No virtual machine traffic was lost, and no VSM-to-VEM connections were interrupted.

Conclusion

Cisco Unified Computing System is a single converged system that is entirely programmable through unified, model-based management to simplify and speed deployment of enterprise-class applications and services running in bare-metal, virtualized, and cloud-computing environments. The system integrates into the Cisco midmarket architecture through active-active uplinks supported by vPC technology. The system integrates into the midmarket architecture's Ethernet network as a single system. End-host mode makes the system appear not as a hierarchy of networks, but as a single system, simplifying Layer 2 connectivity and eliminating the need for Spanning Tree Protocol. The system consolidates network and storage I/O into a single network, saving the cost of multiple Ethernet and Fibre Channel adapters, transceivers, cables, and upstream switch ports. The system connects to native Fibre Channel at a single point, simplifying connectivity to existing storage networks.

The system is configured internally so that it is consistent with VMware vSphere best practices for separating management and production traffic. Production and management traffic is directed onto the two physical network fabrics in Cisco UCS, but with the capability for all traffic to be directed onto a single network in the event of a network failure. This configuration and all failover scenarios were tested and found to show that the system responds rapidly and gracefully to failures, using the system's redundancy quite effectively.

Unlike traditional blade server environments, Cisco UCS is designed so that its configuration can be programmed, from server firmware and BIOS settings to its network connectivity. The system's unified, model-based management uses Cisco service profiles to automate configuration, making the provisioning of existing and new servers fast and accurate, eliminating many of the sources of errors that can cause downtime. This document discussed how to configure servers equipped with CNAs. Today's best practices use Cisco VICs to connect vNICs directly to individual virtual machines, making management of virtual machines equivalent to management of physical servers, further simplifying the administration of systems in midmarket environments.

For More Information

- For more information about Cisco Nexus 5000 Series Switches, Cisco Nexus 2000 Series Fabric Extenders, and the Cisco Nexus 1000V Switch, please visit http://www.cisco.com/go/nexus.
- For more information about Cisco UCS, including Cisco UCS B-Series Blade Servers and Cisco UCS C-Series Rack-Mount Servers, please visit http://www.cisco.com/go/ucs.
- For more information about configuring the Cisco Nexus 1000V Switch in Cisco UCS, please see Best Practices in Deploying Cisco Nexus 1000V Series Switches on Cisco UCS B Series Blade Servers at http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps9902/white_paper_c11-558242.pdf.

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

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

LE-30602-00 09/11