



Best Practices in Deploying Cisco Nexus 1000V Series Switches on Cisco UCS B and C Series Cisco UCS Manager Servers

White Paper

April 2011



Contents

Overview	3
Audience.....	3
Terminology.....	3
Summary of Recommendations.....	3
Introduction	4
Cisco Unified Computing System Common Configurations	6
Cisco Unified Computing System Switching Functions.....	6
Service Profile Templates and Service Profiles.....	11
Network Policy for Link Status and Cisco Discovery Protocol.....	12
Cisco Nexus 1000V Series Common Configurations	12
Cisco Nexus 1000V Virtual Ethernet Module (VEM) Installation.....	12
Cisco Nexus 1000V Series Redundancy Mechanism.....	13
Cisco Nexus 1000V Series Load Balancing.....	14
Port Profile Preference for Uplink Port Selection.....	14
Dual 10GE Adapter Group Use Cases for the Cisco Nexus 1000V Series	14
High Availability with Dual 10GE Adapters.....	14
Service Profile Template Configuration.....	16
Operations.....	17
Performance and Failover Test.....	18
Dual 10GE and 10G FCoE Adapter Group Use Cases for the Cisco Nexus 1000V Series	19
High Availability with Emulex and QLogic Adapters.....	19
Service Profile Template Configuration.....	21
Operations.....	22
Performance and Failover Test.....	22
Cisco UCS Virtual Interface Card Use Cases for the Cisco Nexus 1000V Series	24
High Availability with Virtual Interface Card.....	24
Service Profile Template Configuration.....	27
Trusted and Untrusted Operations on the Cisco UCS M81KR VIC Adapter.....	27
Operations.....	28
Performance and Failover Test.....	28
Cisco Nexus 1000V Series Virtual Switch Module Placement in the Cisco Unified Computing System	30
Option 1: VSM External to the Cisco Unified Computing System on the Cisco Nexus 1010.....	30
Option 2: VSM Outside the Cisco Unified Computing System on the Cisco Nexus 1000V Series VEM.....	30
Option 3: VSM Outside the Cisco Unified Computing System on the VMware vSwitch.....	30
Option 4: VSM Inside the Cisco Unified Computing System on the VMware vSwitch.....	31
Cisco Unified Computing System and Cisco Nexus 1000V Series Versions Used in This Document.....	31
Conclusion	31
For More Information	31
Appendix A: Troubleshooting VMware ESX Server in the Cisco Unified Computing System	32
VMware ESX and ESXi Commands to Troubleshoot Cisco Nexus 1000V Series Switches in the Cisco Unified Computing System.....	33

Overview

This document provides design guidance and best practices in deploying Cisco Nexus® 1000V Series Switches—software switches embedded within the VMware vSphere hypervisor—on top of the Cisco Unified Computing System™, focusing on deployment options using the various adapter types that the Cisco Unified Computing System supports. For detailed Cisco Nexus 1000V Series configuration documentation, please refer to the respective Cisco® and VMware product configuration guides. Links to the product configuration guides can be found in the “For More Information” section of this document.

Audience

This document is intended for Cisco systems engineers supporting customer server administrators, network administrators, and infrastructure architects (both customers and partners) interested in understanding and deploying Cisco Nexus 1000V Series Switches on the Cisco Unified Computing System. It assumes a basic functional knowledge of the Cisco Nexus 1000V Series, VMware vNetwork Distributed Switching, and Ethernet switching within a VMware environment in general. It also assumes a general understanding of the Cisco UCS, server I/O, virtualization, and networking concepts.

Terminology

This document refers to the virtual machine network interface cards (NICs) that are presented by the VMware ESX hypervisor as virtual NICs (vNICs), and it refers to the service profile–defined interfaces that are created and managed by Cisco UCS Manager as virtual interfaces (VIFs). The Cisco Nexus 1000V Series is the Cisco implementation of what the IEEE 802 working groups call a virtual Ethernet bridge (VEB).

Summary of Recommendations

This document will outline the technical details of these recommendations for most customers deploying the Nexus 1000V on UCS B and C Series servers (C series managed via UCSM), but we will provide an upfront list here to outline the key points:

- Configure the UCS Fabric Interconnects in End Host Mode (default)
- Configure 2 x 10GE adapters in the service profile with replicated configurations
- Configure the first adapter created to UCS fabric A, and the second adapter to fabric B
- Do NOT configure a dynamic vNIC policy on the service profile for the ESX host
- Send all the needed VLANs on these adapter pairs, and define a native management VLAN
- Configure an adapter network control policy in UCS that sends CDP packets to the Nexus 1000V
- Configure a QoS adapter class that trusts the host (the Nexus 1000V) and assign to the adapters
- Do NOT enable fabric failover on the adapters in a UCS Service Profile
- In UCS versions 1.3 and prior, modify the Cisco UCS mac aging timer to a number greater than 14400
- Utilize a stand-alone Nexus 1010 appliance for the supervisory functions
- If you need to use MAC ACLs, ensure the detail in “Cisco Nexus 1000V Series Redundancy Mechanism” below is followed
- On the Nexus 1000V configure the given host VEM module adapter pair for aggregation with “channel-group auto mode on mac-pinning”
- Engineer the Nexus 1000V control traffic (control/packet/management) to a single UCS fabric

- Engineer the Nexus 1000V data traffic (Virtual Machine networks) to the peer UCS fabric
- Ensure all the Nexus 1000V control and VM VLANs are sent from the upstream switch to the UCS
- Ensure all the Nexus 1000V control and VM VLANs are always sent between adjacent upstream switches
- On the upstream switches configure “spanning-tree port type edge trunk” for the interfaces towards the UCS

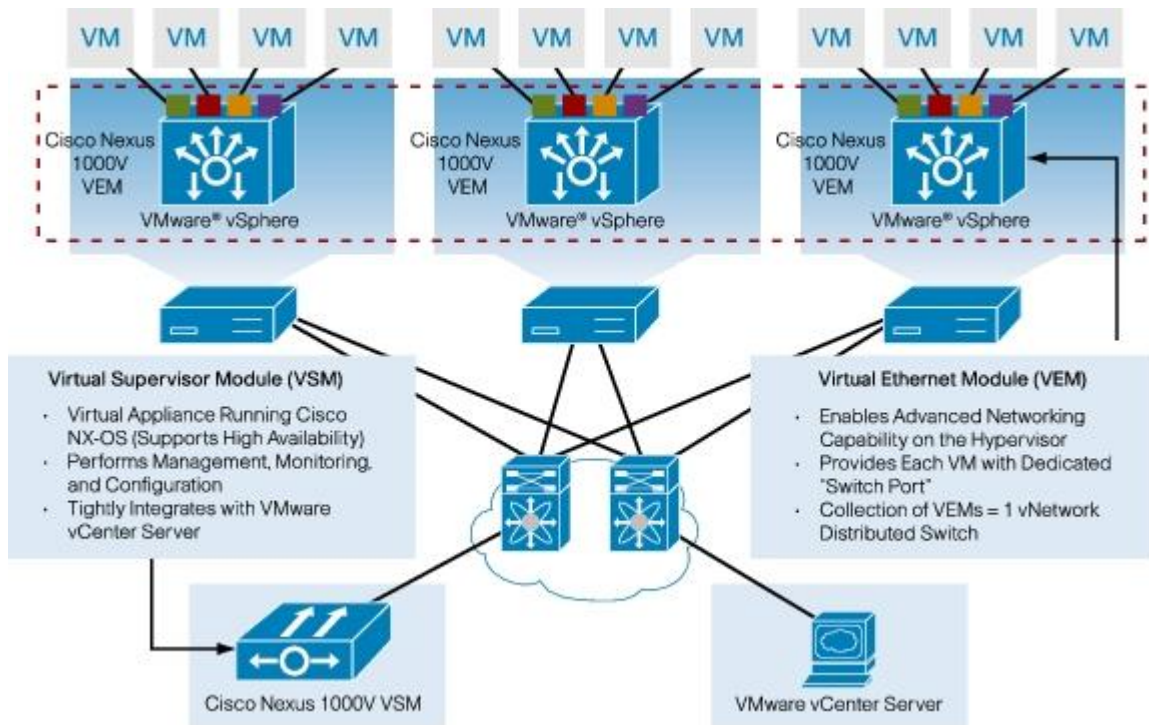
Introduction

Cisco Nexus 1000V Series Switches are virtual machine access switches that are an intelligent software switch implementation for VMware vSphere environments running the Cisco NX-OS Software operating system. Operating inside the VMware ESX hypervisor, the Cisco Nexus 1000V Series supports Cisco VN-Link server virtualization technology to provide:

- Policy-based virtual machine connectivity
- Mobile virtual machine security and network policy
- Non-disruptive operational model for your server virtualization and networking teams

When server virtualization is deployed in the data center, virtual servers typically are not managed the same way as physical servers. Server virtualization is treated as a special deployment, leading to longer deployment time, with a greater degree of coordination required among server, network, storage, and security administrators. With the Cisco Nexus 1000V Series, you can have a consistent networking feature set and infrastructure. Virtual servers can now use the same network configuration, security policy, diagnostic tools, and operational models as their physical server counterparts attached to dedicated physical network ports. Server administrators can access predefined network policy that follows mobile virtual machines to help ensure proper connectivity, saving valuable time for focusing on virtual machine administration. This comprehensive set of capabilities helps you deploy server virtualization faster and achieve its benefits sooner (Figure 1).

Figure 1. The Cisco Nexus 1000V Series Enables a Virtualized Server Operational Model Similar to That of a Physical Server

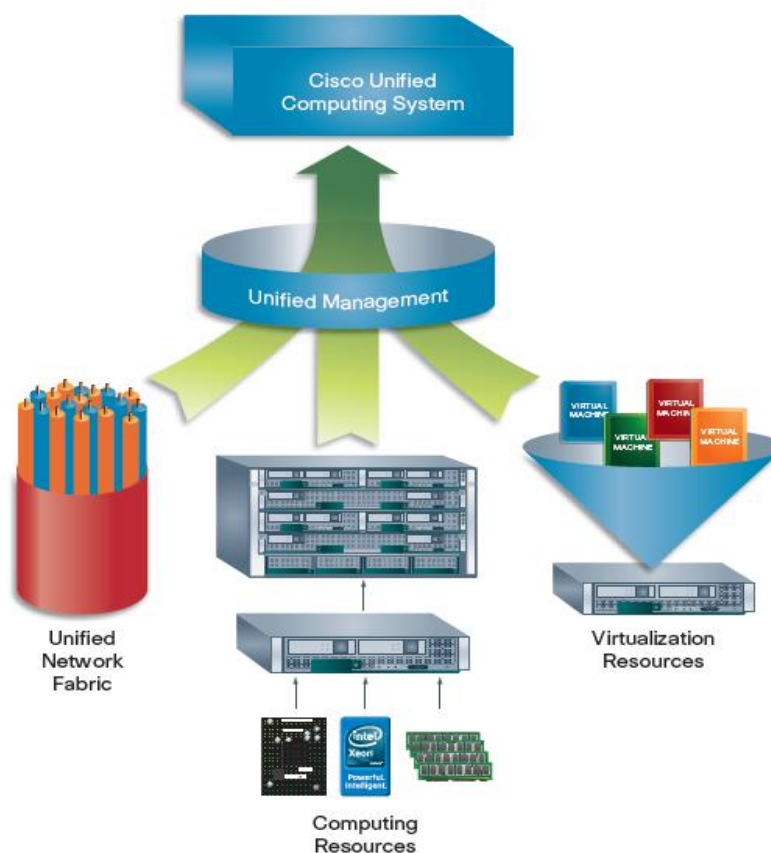


The Cisco Unified Computing System integrates low-latency unified network fabric with enterprise-class, x86-based servers, creating an integrated, scalable, multi-chassis platform in which all resources participate in a unified management domain. A single system scales to up to 40 chassis, 320 compute nodes, and thousands of virtual machines.

The Cisco Unified Computing System enables more dynamic and agile data centers, in which server identity (MAC addresses, worldwide names [WWNs], firmware and BIOS revisions, network and storage connectivity profiles and policies, etc.) can be dynamically provisioned or migrated to any physical server within the system. The Cisco Unified Computing System creates a highly dynamic and cohesive integration with network and storage resources to meet the rapidly changing needs in today's data centers. New computing resources can be deployed in a just-in-time approach, or even pre-provisioned before hardware arrives. Traditional physical and virtual workloads can be easily migrated between servers through remote management, regardless of physical connectivity. The Cisco Unified Computing System improves availability, security, agility, and performance through an integrated architecture (Figure 2).

Some functions in the Cisco Unified Computing System are similar to those offered by the Cisco Nexus 1000V Series Switches, but with a different set of applications and design scenarios. The Cisco Unified Computing System offers the capability to present adapters to physical and virtual machines directly (with Virtual Machine FEX technology). When presented to the virtual machines, Cisco VM-FEX technology can provide many of the same benefits as those described for the Cisco Nexus 1000V Series.

Figure 2. The Cisco Unified Computing System Integrates Network, Compute, and Virtualization Resources into a Single Cohesive System



Cisco Unified Computing System Common Configurations

This section discusses some of the areas of special interest in the Cisco Unified Computing System that pertain to the configuration and use of the Cisco Nexus 1000V Series. The configurations discussed here apply regardless of adapter type.

Cisco Unified Computing System Switching Functions

The Cisco Unified Computing System offers multiple adapter types, which currently can be grouped into three classes of functionality:

- Dual 10 Gigabit Ethernet Port Adapters:
 - Cisco UCS 82598KR-CI Intel 82598 Based 10 Gigabit Ethernet Adapter (Mezzanine)
 - Cisco UCS M61KR-I Intel 82599 Based 10 Gigabit Ethernet Adapter (Mezzanine)
 - Cisco UCS M51KR-I Broadcom 57711 Based 10 Gigabit Ethernet Adapter (Mezzanine)
 - Cisco UCS Intel-X520 (PCIe)
 - Cisco UCS BCM 57711 (PCIe)
- Dual 10 Gigabit Ethernet Port and Dual 10 Gigabit FCoE Adapters:
 - Cisco UCS M71KR-Q QLogic 2642 Based 4G FCoE Converged Network Adapter (Mezzanine)
 - Cisco UCS M71KR-E Emulex LP21000 Based 4G FCoE CNA (Mezzanine)
 - Cisco UCS M72KR-Q Qlogic 8152 Based 10 Gigabit FCoE CNA (Mezzanine)
 - Cisco UCS M72KR-E Emulex OCe10102-F Based 10 Gigabit FCoE UCNA (Mezzanine)

- Cisco UCS QLE 8152 (PCIe)
- Cisco UCS ELX OCe10102-F (PCIe)
- Virtual Interface Card with User Configurable Ethernet and FCoE ports
- Cisco UCS M81KR Virtual Interface Card (Mezzanine)
- Cisco UCS P81E Virtual Interface Card (PCIe)

Each of these cards has a pair of 10 Gigabit Ethernet connections to the Cisco Unified Computing System backplane or Fabric Interconnect that support the IEEE 802.1 Data Center Bridging (DCB) function to facilitate I/O unification within these adapters. On each mezzanine adapter type, one of these backplane ports is connected through 10GBASE-KR to the A-side I/O module; then that connection goes to the A-side fabric interconnect. 10GBASE-KR is a copper mid-plane technology for interfacing adapters and switching elements through these mid-planes. The other mezzanine connection is 10GBASE-KR to the B-side I/O module; that connection then goes to the B-side fabric interconnect. Figure 3 later in this document shows this connectivity. For the C series UCSM integrated servers, one port is connected to the A-side fabric interconnect directly, and the other is connected to the B-side fabric interconnect.

Within the Cisco UCS M71KR-E, M71KR-Q, M81KR, and P81E adapter types, the Cisco Unified Computing System can enable a fabric failover capability in which loss of connectivity on a path in use will cause remapping of traffic through a redundant path within the Cisco Unified Computing System. In determining the adapter type to use to support the Cisco Nexus 1000V Series, the interface count available to the VMware ESX host is a critical factor. Here are some high-level trade-offs:

With the Cisco UCS M81KR and P81E, multiple VIFs can be identified to match those used by the service console, VMware VMkernel, virtual machine data, Cisco Nexus 1000V Series VLANs, etc. as required. Each of these adapters can be defined within the Cisco Unified Computing System as connected to an individual fabric interconnect and, optionally, enabling a failover to the other. Any adapter configured can be assigned its order of presentation to ESX or ESXi, along with the UCS fabric to connect to.

With the Cisco UCS M71KR-E and M71KR-Q adapters, a maximum of two adapters are presented to the VMware ESX hypervisor running on the blade. Each of these adapters can be defined within the Cisco Unified Computing System as connected to an individual fabric interconnect and, optionally, enabling a failover to the other. This fabric failover enables a model in which the virtual machine data can use one path within the Cisco Unified Computing System by default, and all other connections can go on the other path. For reasons stated below, we do not recommend using this capability when using the Nexus 1000V. With these adapters, the first adapter presented to ESX or ESXi will always be the A-side UCS fabric, and the second presented will be the B-Side UCS fabric.

With the Cisco UCS 82598KR-CI, M51KR-I, M61KR-I, M72KR-Q, M72KR-E, Intel-X520, BCM 57711, QLE 8152, and ELX OCe10102-F adapters, a maximum of two adapters are presented to the VMware ESX hypervisor running on the blade. These interfaces do not support a fabric failover, and the service console must be migrated to the Cisco Nexus 1000V Series Switch along with all these other adapter types if any high-availability requirement exists. The actual migration of this interface during VMware ESX deployment on the blade is discussed in the specific adapter section later in this document, but for more detailed information about how to migrate a user's service console, see the Cisco Nexus 1000V Series documentation. With these adapters, the first adapter presented to ESX or ESXi will always be the A-side UCS fabric, and the second presented will be the B-Side UCS fabric.

Note: The user should assign any VMware service console and VMkernel interface VLANs in the Cisco Nexus 1000V Series Switches as system VLANs along with the management, control, and packet VLANs.

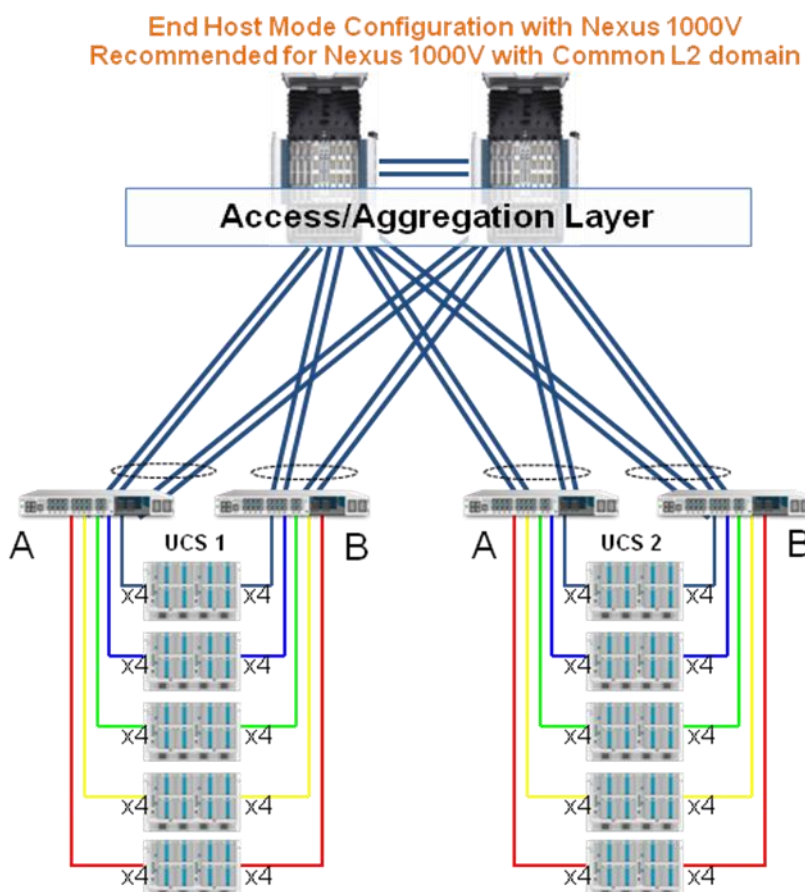
Inside the Cisco UCS 6100 Series Fabric Interconnects, local Layer 2 switching is performed between the server adapters that connect to them, and any traffic that needs other communication will exit on an uplink to the data center switches. Examples of this traffic are Layer 3 communications to the outside network, to other Cisco Unified Computing System instances, and to other VMware nodes, as well as traffic if some servers in a given Cisco Unified Computing System are active on the A side and some are active on the B side (the A-B data communications path is through the upstream switch). The administrator can define which uplink or uplink channel a given server adapter or VIF will use as a point of uplink to get control of the traffic flows.

The Cisco UCS 6100 Series operates in two discrete modes with respect to flows in the Cisco Unified Computing System. The first is assumed to be more common and is called end-host mode; the other is the switched mode, in which the fabric interconnect acts as a normal Ethernet bridge device. Discussion of the differences between these modes is beyond the scope of this document; however, the Cisco Nexus 1000V Series Switches on the server blades will operate regardless of the mode of the fabric interconnects. With respect to a VMware environment running the Cisco Nexus 1000V Series, the preferred solution is end-host mode to help ensure predictable traffic flows.

End-Host Mode Configuration

With the end-host mode configuration (Figure 3), when the Layer 2 communication flows within the Cisco Nexus 1000V Series infrastructure occur, these flows may be either local to a given Cisco UCS 6100 Series Fabric Interconnect or through the upstream data center switch with more hops. Applying a quality-of-service (QoS) policy here to help ensure a minimum bandwidth is recommended. The recommended action in the Cisco Nexus 1000V Series is to assign a class of service (CoS) of 6 to the VMware service console and VMkernel flows and to honor these QoS markings on the data center switch to which the Cisco UCS 6100 Series Fabric Interconnect connects. Marking of QoS values can be performed on the Cisco Nexus 1000V Series Switch in all cases, or it can be performed on a per-VIF basis on the Cisco UCS M81KR or P81E within the Cisco Unified Computing System with or without the Cisco Nexus 1000V Series Switch.

Figure 3. End-Host Mode Configuration with Cisco Unified Computing System



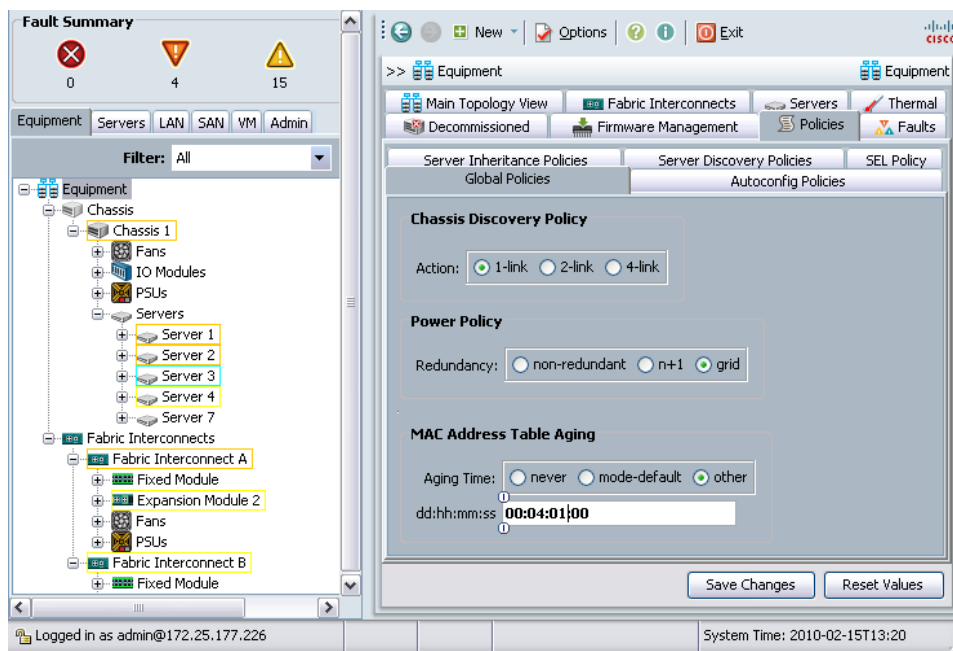
It is also recommended in this scenario that you connect the Cisco UCS 6100 Series Fabric Interconnects with redundant links to a pair of data center devices that are either separate switches or joined through some method for multichassis EtherChannel (MCEC) such as Cisco Virtual Switching System (VSS) or Cisco Virtual PortChannel (vPC). For comparison, the trade-offs in relying on the re-pinning of traffic to uplink ports instead of hashing to another member within the vPC on the uplinked switches will be shown in the later testing results sections. With these redundant uplinks to the data center edge, when the link failure indications are passed to the Cisco Nexus 1000V Series Switch, it will fail traffic over itself. Do not create static pin groups on the system uplinks, which will by default allow the Cisco UCS 6100 Series Fabric Interconnects to dynamically pin the traffic to the uplinks. Again, this configuration reduces the requirements on the Cisco Nexus 1000V Series Switch and increases availability.

Another important point to remember when using end-host mode is that a single port will be chosen on each fabric interconnect as a listening port for broadcast and multicast traffic. Therefore, make sure that all uplinks ports are attached to the same Layer 2 domain. If they are not, then switch mode operations on the fabric interconnect will be required.

When using end-host mode, you should also remember that the Cisco Unified Computing System never listens for unknown unicast traffic from the uplink segments. Following the usual data switching rules, the Cisco Unified Computing System uses the default MAC address aging timer setting (7200 seconds prior to version 1.3, and 14500 for version 1.3 and later), at which time the dynamic MAC address entries will be flushed. However, in some cases upstream devices may still have an Address Resolution Protocol

(ARP) entry, and these devices may cache for 14400 seconds (the typical setting on many Cisco products) by default. The recommendation here is to modify the Cisco Unified Computing System MAC address aging timer when using UCS software versions prior to 1.3 as shown in Figure 4.

Figure 4. Configuring MAC Address Aging Timer for Dynamically Learned MAC Address Entries on the Fabric Interconnects



Another recommendation is that the customers consolidate the Cisco UCS 6100 Series Fabric Interconnect links from the A and B sides on the same line cards within the aggregation layer. This action allows the user to take advantage of local switching performance on that device to offer better scale because the flows need not traverse the fabric. Note that the channel bundling is performed through Link Aggregation Control Protocol (LACP), and the Cisco UCS 6100 Series Fabric Interconnects will appear to the network to be running in LACP active mode.

In addition, the Cisco Nexus 1000V Series has management and data VLANs, along with a pair of special VLANs called control and packet VLANs. All of these VLANs must be allowed between the data center aggregation switches to which the Cisco UCS 6100 Series Fabric Interconnects connect (along with any other hosts of interest outside the Cisco Unified Computing System).

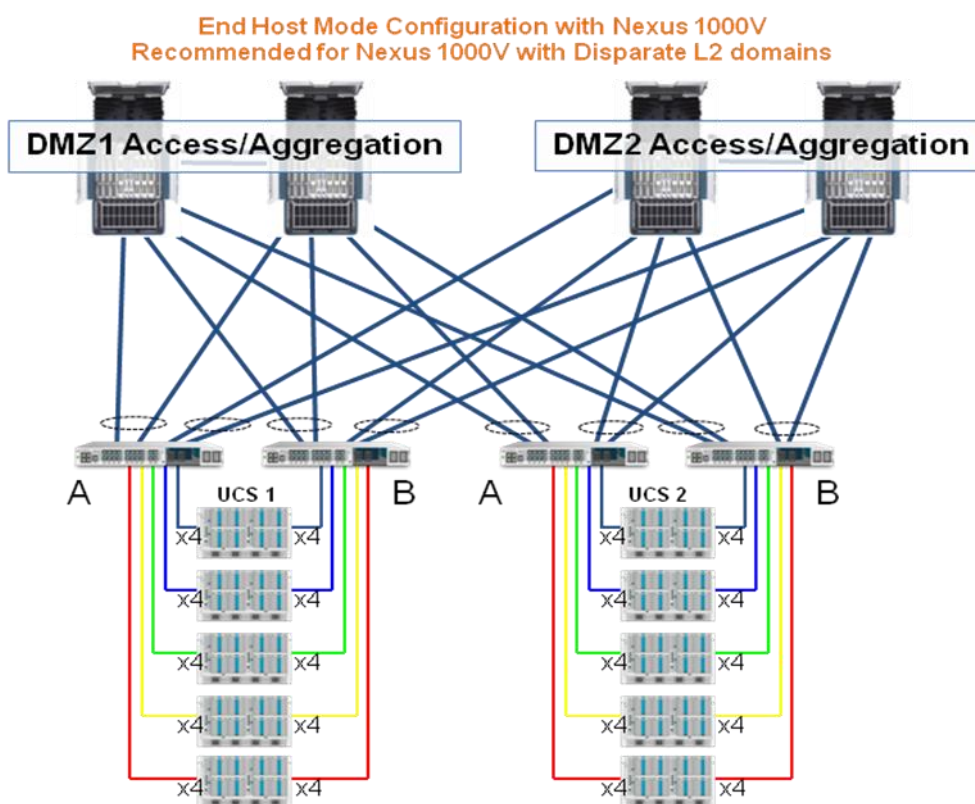
Switch Mode Configuration

When using switch mode (Figure 5), the Layer 2 communication flows within the Cisco Nexus 1000V Series infrastructure can be local to a given Cisco UCS 6100 Series Fabric Interconnect or through uplinks within the given Cisco Unified Computing System. This design is not recommended for the following reasons:

- With Virtual Switching System (VSS) or Virtual Port Channel (vPC) designs, all paths will be forwarding from the Cisco UCS 6100 Series Fabric Interconnects.
- If a Layer 2 link exists between the aggregation switches without VSS or vPC, a spanning-tree block will form between some fabric interconnects and upstream devices.

- Customers wishing to leverage the traffic engineering capabilities within the UCS will now have to rely on spanning tree determining forwarding paths
- Layer 2 domain sizing issues come into play, as all MAC addresses within the Layer 2 domain are now stored within the memory of the UCS Fabric Interconnect (as opposed to end host mode where only the registered and learned downstream MAC addresses are stored)
- Deterministic traffic failover and failback can be harder to achieve

Figure 5. Switch Mode Configuration with Cisco Unified Computing System



A First-Hop Routing Protocol (FHRP) is used to give the default gateway visibility to the servers and with switch mode this can be one or two physical hops from the server.

Because of the restrictions of the switch mode noted here, the end-host mode of operation is the general recommended best practice when implementing Nexus 1000V on UCS. The key exception is when separate layer 2 domains are to be sharing a UCS.

Service Profile Templates and Service Profiles

For the Cisco Nexus 1000V Series, some general best practices exist for the definitions of the service profile templates and service profiles in Cisco UCS Manager. When the adapters are defined and used within a service profile, they become available according to the availability of the fabric to which they are assigned. The definitions available to the VMware ESX hypervisor depend on the order in which the adapter is defined within the service profile (which adapter is interface 0, interface 1, etc.). The general goal is a correlation between the adapter-creation function that the Cisco Unified Computing System administrator performs and the VMware ESX hypervisor network assignment. This function is similar to the physical server function of determining which RJ-45 port on the server plugs into which network and then identifying this adapter on the server. For this reason, the recommended best practice is to invoke all

service profiles from a common template to help prevent adapters and VIFs from being assigned in different orders on the server PCI bus. This practice has far more importance for the Cisco UCS M81KR and P81E, where these can be set within a service profile. With the other adapter types, the A side adapter will be the first one discovered in ESX or ESXi.

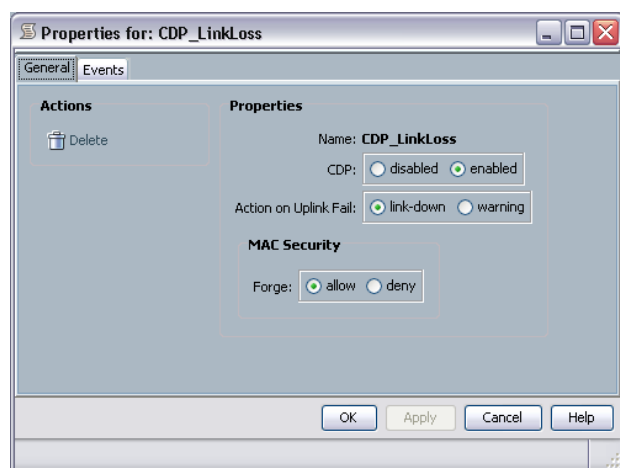
In this same way, the administrator can help ensure a common baseline of adapter firmware to the common servers, along with QoS, VLAN settings, etc.

Network Policy for Link Status and Cisco Discovery Protocol

Within the Cisco Unified Computing System, the uplink from the Cisco Unified Computing System to the customer data center network can be an individual link, multiple links, or a PortChannel. When the individual link or the last remaining member of a group or channel of links to the upstream network loses connectivity, it is desirable to have this outage signaled back to the blade and to the Cisco Nexus 1000V Series virtual Ethernet module (VEM) running on it. This action can be accomplished by having the Cisco Unified Computing System fabric take down the link to the adapter (regardless of whether it is a physical adapter or a virtualized adapter). The switching logic on the VEM will then be applied to reestablish connectivity on the surviving adapter.

Another management feature is the capability to enable Cisco Discovery Protocol on these physical and virtual adapters to help troubleshoot any connectivity problems. Figure 6 shows how to set this policy in Cisco UCS Manager. The administrator can then assign this network control policy to adapters upon their definition in the service profile. Note that the MAC Security feature should always be set to “allow” for Cisco Nexus 1000V Series Switches (and any virtual switch) that will be using an adapter with this network control policy.

Figure 6. Configuring Network Control Policy for Cisco Discovery Protocol Capabilities and Uplink Failure Actions



Cisco Nexus 1000V Series Common Configurations

This section discusses some of the areas of special interest with respect to the Cisco Nexus 1000V Series when it is employed in the Cisco Unified Computing System.

Cisco Nexus 1000V Virtual Ethernet Module (VEM) Installation

A fundamental task to adding an ESX host to the Nexus 1000V deployment includes the installation of the VEM onto the chosen UCS blade. This function can be repetitive and sometimes errors can be introduced with the repetition. While the customer can manually download the VEM file from the VSM or

from www.cisco.com to the host, and run the “esxupdate ...” commands to install the VEM manually, the customer then needs to ensure the compatibility with the ESX version.

VMware provides a service called VMware Update Manager (VUM) which can automate this process of installing the correct VEM modules on a given host and ensuring the correct versioning. With the usage of VUM, an administrative action of adding a new host to a Nexus 1000V becomes a single operation of pointing to the host to add. VUM will perform operations to stage (position the files on the hosts), and remediate (install and activate the modules) – then the user has no additional tasks to bring these UCS servers into the vCenter cluster. While the steps to configure VUM are outside the scope of this document, a key item is to ensure the following tasks are performed under the Update Manager Tab for the vCenter:

- Select configuration tab from VUM
- Patch download settings
- Add Patch source
- In the Custom Patch Type enter:

<http://hostupdate.vmware.com/software/VUM/PRODUCTION/cisco-main-index.xml>

MAC Pinning

Cisco Nexus 1000V Series running Cisco NX-OS Software Release 4.1(2) and later provides the capability to 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 in all adapter types when running a Cisco Nexus 1000V Series Switch on the Cisco Unified Computing System. The configuration is described in the appropriate adapter group section of this document.

The general theory is that the Cisco Nexus 1000V Series VEM will assign traffic to the 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 switches (in this case, the Cisco UCS 6100 Series Fabric Interconnects) are not aware of the configuration on the Cisco Nexus 1000V Series Switch and see only the appropriate MAC addresses on the appropriate blades. If a link fails, the traffic fails over to the surviving link to mitigate the outage, and the traffic is returned to the original link (after an appropriate timeframe to ensure stability) after the link is restored.

Cisco Nexus 1000V Series Redundancy Mechanism

If a physical uplink fails, the Cisco Nexus 1000V Series Switch will send packets upstream of a surviving link to notify upstream switches of the new path to reach these virtual machines. These notifications are sent to the Cisco UCS 6100 Series Fabric Interconnect, which updates its MAC address tables and sends gratuitous ARP messages on the uplink ports so the data center network can learn of the new path. If the customer chooses to implement MAC address access control lists (ACLs) to restrict layer 2 traffic from virtual machines to a single device (a default gateway is a common example), the ACL must be configured to allow these packets from the Cisco Nexus 1000V Series Switch to the upstream Cisco Unified Computing System to help ensure proper failover. The following ACL shows a sample gateway entry followed by the entry for the Cisco Nexus 1000V Series MAC address notification packets:

```
mac access-list JustExternalGateway
10 permit any 0022.1991.7e08 0000.0000.0000
```

```
20 permit any 0100.0ccd.cdcd 0000.0000.0000
```

```
30 deny any any
```

Cisco Nexus 1000V Series Load Balancing

A fundamental feature of traffic forwarding in the Cisco Unified Computing System is that traffic on a given interface for a host (physical or virtual) in the Cisco Unified Computing System traverses one of the two available fabrics and can fail over to the other. While the Cisco Nexus 1000V Series supports many granular load-balancing mechanisms for traffic from a given host to the network, the Cisco Unified Computing System supports only the port ID and source MAC address mechanisms.

Port Profile Preference for Uplink Port Selection

In a Cisco Nexus 1000V Series Switch's port profile, the administrator can set a preference for the uplink path to use. If that uplink fails, another uplink will be chosen. The selection mechanism defines the pinning ID in a virtual Ethernet (vEthernet) port profile, which will represent the virtual machine NIC (VMNIC) as reported by the VMware ESX server (vmnic0=0, vmnic1=1, etc.). The user can also define a desired uplink for the packet and control VLANs, by defining a pinning control VLAN number and a pinning packet VLAN number under the Ethernet port profiles (the ones for the physical uplink adapters). These define the links to be preferred when multiple links are operational. If a link failure occurs on the active link, the Cisco Nexus 1000V Series Switch will move the traffic to a surviving link. This operation is a fundamental reason for making sure that a link-down-notification policy is used: to signal the Cisco Nexus 1000V Series VEM of the outage so that failover can occur.

An important advantage here is the capability to separate production virtual machine traffic from some of the management operations such as access, reporting, VMware vMotion migration, and backup.

Dual 10GE Adapter Group Use Cases for the Cisco Nexus 1000V Series

High Availability with Dual 10GE Adapters

The dual 10GE adapters run either a classic Ethernet connection or a VNTag connection to the I/O module within UCS. At the adapter or the I/O Module the system QoS policies are enforced and the virtual network tag (VNTag) field is added. No Cisco microcode is executed on this adapter type natively. Therefore, this adapter has no fabric failover capability, so the use case is a single-use scenario: a pair of vNICs, one each to the A- and B-side fabric interconnects. The first discovered adapter (vmnic0 in ESX and ESXi) will be the A-side, and the next (vmnic1) will be the B-side. Another important point is that Fiber Channel over Ethernet (FCoE) is not supported from the Cisco Unified Computing System point of view (the software FCoE is at the OS level), so no SAN boot capabilities are available; therefore, the system needs to rely on the local disk. This factor affects the stateless use cases within the Cisco Unified Computing System, although customers can move the physical disks along with the service profiles.

For the Cisco Nexus 1000V Series Switch to receive Cisco Discovery Protocol packets to help provide a management connectivity view on these adapters, the configuration in the Cisco UCS service profile needs to mark the pair of adapters with the network policy to receive the Cisco Discovery Protocol information.

To achieve the goal of high availability for the virtual machines within the VMware ESX host, use of uplink MAC pinning on the Cisco Nexus 1000V Series Switch is recommended. To configure MAC pinning on the Ethernet type (uplink) port profile, the administrator needs to enter only the following under the Cisco Nexus 1000V Series Ethernet port profile:

```
port-profile type ethernet Intel82599-uplink
...
channel-group auto mode on mac-pinning
...
```

The Cisco Nexus 1000V Series uplink port profile that is to be assigned to the Cisco UCS adapter should be given a descriptive name, because VMware vCenter will display these names, and the administrator will map the Cisco Nexus 1000V Series Ethernet ports to the appropriate uplink port groups. The interesting parts of the Cisco Nexus 1000V Series configurations are shown here. The VLAN mappings used here were:

- VLAN 2: Management and console
- VLAN 3: VMware vMotion
- VLAN 4: Cisco Nexus 1000V Series control VLAN
- VLAN 5: Cisco Nexus 1000V Series packet VLAN
- VLAN 6: Virtual machine data network 1

These mappings are used throughout this document. An important requirement is that these VLANs all must be allowed from the Cisco UCS 6100 Fabric Interconnects to the upstream switches in the customer data center (all VLANs are allowed on all uplinks from the UCS perspective). In many cases, interfaces using these VLANs can be on the A- or B-side fabric in the Cisco Unified Computing System, and to communicate between them, the path will be through the access or aggregation layer in the data center when in end-host mode. The first four VLANs outlined above (mgmt, vMotion, N1k Control, and N1k Packet) must be defined as system VLANs in the uplink port profiles. The relevant portion of the Cisco Nexus 1000V Series configuration is shown here:

<config-snippet>

```
port-profile type ethernet Intel82599-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 Ethernet3/1
    inherit port-profile Intel82599-uplink

interface Ethernet3/2
```

```
inherit port-profile Intel82599-uplink
</config-snippet>
```

Service Profile Template Configuration

The following rules apply in the creation of any Cisco dual port 10 GE adapter based service profile for hosting a VMware ESX instance that will be included in a Cisco Nexus 1000V Series system:

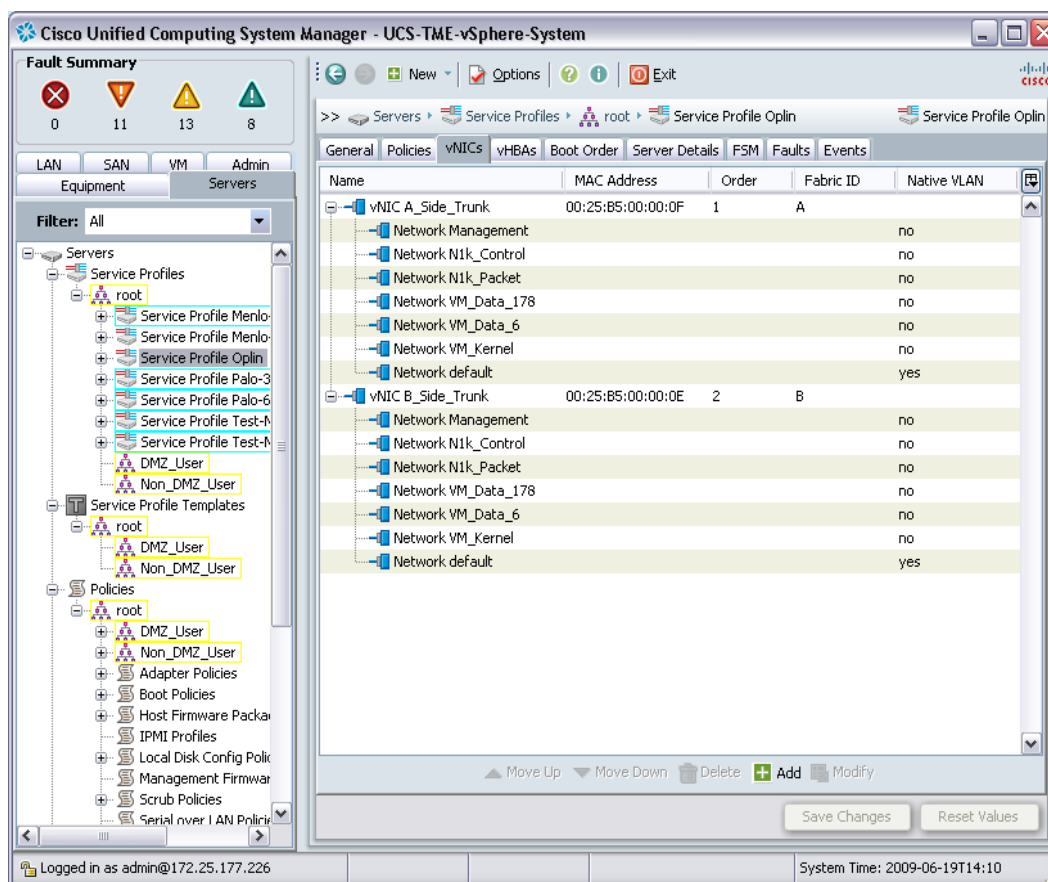
The network administrator must define the VLANs that will initially be passed to the adapter and thus to the Cisco Nexus 1000V Series VEM.

The network administrator should mark the adapter policy for Cisco Discovery Protocol if a management connectivity view on the VMware ESX host is desired.

The network administrator should define a default QoS profile for any traffic that the Cisco Nexus 1000V Series Switch does not mark.

Figure 7 shows an example of a service profile with the pertinent service profile networking configuration.

Figure 7. Cisco UCS Dual 10GE Adapter Service Profile: vNIC Definitions for the Cisco Nexus 1000V Series



Operations

As discussed earlier, the design requires trade-offs. Since multiple adapters cannot be mapped to the host with this dual port 10GE grouping, the user will need to rely on the QoS settings from the Cisco Nexus 1000V Series Switch to help ensure smooth operation. This solution allows the user to define a primary and a backup path for the console, kernel, production, or other interfaces to the fabric. This is accomplished by assigning the preferred interface with the following code:

```
port-profile type vethernet vMotionNet
...
pinning id 0
...
```

As discussed earlier, the pinning ID mappings can control which profiles use which paths to the data center network. In this example, the design calls for VMware vMotion traffic to use vmnic0 as its uplink. If the chosen vmnic is not available, or does not carry the appropriate network, then a vmnic that can meet the profile requirements will be selected.

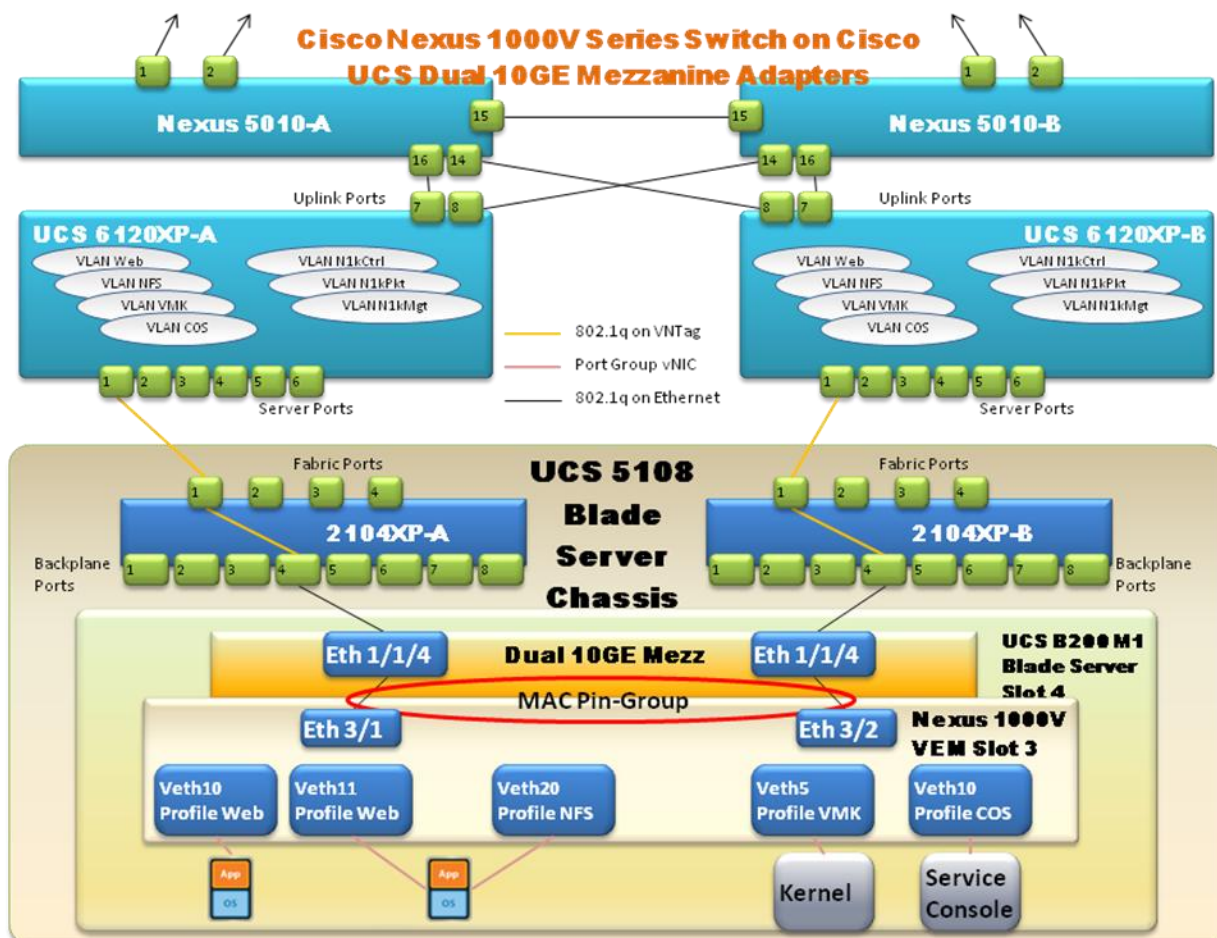
Within this grouping, the Cisco UCS 82598/82599 adapter types use Intel Virtual Machine Device Queue (VMDq) technology to assist in CPU operations to switch packets from the adapter to the virtual machines and to define multiple transmit queues to reduce any head-of-line blocking on transmission of data. This

document does not explore these details; for more information, see the Intel document at http://www.intel.com/network/connectivity/vtc_vmdq.htm.

Performance and Failover Test

To see the basic performance data and failover times that can be expected, an iPerf test was performed from a virtual machine running behind a Cisco Nexus 1000V Series Switch to a host external to the Cisco Unified Computing System. Figure 8 shows the testing topology that is shown for the dual adapter case.

Figure 8. Cisco Nexus 1000V Series Switch on Cisco UCS Dual 10GE Mezzanines



Not shown in Figure 8 is a traffic target server with a Gigabit Ethernet connection to Cisco Nexus 5010 Switch A.

- Baseline observations:
 - A series of 90-second iPerf flow tests were performed.
 - Bidirectional throughput on a single virtual machine behind the Cisco Nexus 1000V Series VEM in the Cisco Unified Computing System and a single external Gigabit Ethernet host was 490 Mbps.
 - All MAC addresses were visible on the Cisco Nexus 1000V Series Switches, Cisco Nexus 5010 Switches, and Cisco UCS 6120XP 20-Port Fabric Interconnects.

- The B fabric was the side that was actively passing traffic by default. This verifies the pinning ID of 1 (or vmnic1, which was a B-fabric adapter).
- Three threads on the Cisco UCS blade (out of the 16 in the pair of E5520 CPUs) hovered around 30 percent for the duration of the test.
- Observations about the failure of the link from Cisco UCS 6120XP B to Cisco Nexus 5010 A (active path for flow: Cisco Unified Computing System repinning to another uplink will occur):
 - A 90-second iPerf unidirectional (virtual machine to external host) flow test was performed.
 - On failure, the MAC address still remained on the B fabric.
 - Cisco Unified Computing System repinning time for this flow was 1.1 seconds.
 - The Cisco Nexus 1000V Series Switch did not detect a link outage.
 - When this same test was run in the opposite direction (external host to virtual machine), the repinning time was measured at 2.2 seconds.
 - When the Cisco Nexus 5010 Switches were configured in a vPC and this link failure occurred, the PortChannel selected the surviving member for the flow (instead of Cisco Unified Computing System repinning). The failover time in this scenario was 0.09 second, and failback was 0.03 second.
- Observations about failure of the link from Cisco 6120XP B to Cisco Nexus 5010 B (current active path for flow: a Cisco Nexus 1000V Series Switch failover of the VMNICs will occur):
 - A 90-second iPerf unidirectional (virtual machine to external host) flow test was performed.
 - On failover, the MAC address appeared on the A fabric.
 - Failover time for this flow was 2.5 seconds.
 - When this same test was run in the opposite direction (external host to virtual machine), the failover time was 2.5 seconds.
 - On link restoration (to restore the test bed), the Cisco Nexus 1000V Series Switch took less than 0.15 second to restore traffic to the preferred fabric (due to the pinning ID setup).
 - Restoration of the second uplink did cause Cisco Unified Computing System repinning to the original path, and this took 1.03 seconds.
 - The Cisco Nexus 1000V Series MAC pinning provided high availability.

Dual 10GE and 10G FCoE Adapter Group Use Cases for the Cisco Nexus 1000V Series

High Availability with Emulex and QLogic Adapters

From the perspective of high availability in the Cisco Unified Computing System, there is no difference between the Cisco UCS M71KR-E and M71KR-Q adapters, nor between the M72KR-E and M72KR-Q adapters. The former pair of these adapters function differently within the Cisco Unified Computing System compared to the PCI Express (PCIe) form-factor cards available today. The Cisco UCS M71KR-E and M71KR-Q adapters (on the Cisco application-specific integrated circuit [ASIC]) run a firmware that is maintained by Cisco, as opposed to the initial versions of the PCIe cards from Emulex and QLogic (which maintain the Cisco firmware themselves). As a result, some functions are different:

- IEEE 802.1 Data Center Bridging mode
- Failover options of the adapters

The Cisco UCS M71KR-E and M71KR-Q adapters in the Cisco Unified Computing System run a VNTag-based communication flow to the Cisco UCS Cisco UCS 6100 Series Fabric Interconnects rather than a mode called I/O consolidation (IOC) on the PCIe adapters. This VNTag mode enables an additional fabric failover mode (in which the adapter registers with both the Cisco UCS 6100 Series Fabric Interconnects), using one of the paths as an active path and the other as a backup path.

Note: VNTag is not discussed in detail in this document; please refer to other documents for details about the technology.

Cisco firmware is executed on the M71KR-Q and M71KR-E adapters, but not on the M72KR-Q and M72KR-E adapters. The adapter performs a registration process to the fabric interconnects at server startup. This adapter type is its own communications endpoint, and since the VNTag mechanism uniquely identifies the adapter, the registration process registers the adapter's MAC address as defined in the Cisco Unified Computing System service profile. When source traffic is seen on the fabric interconnects with a source MAC address behind these adapters (such as when a Cisco Nexus 1000V Series VEM is running on the VMware ESX server), the MAC address table on the active link's fabric interconnect will be dynamically updated. This information is replicated to the backup fabric interconnect, however this document still recommends the disabling of this fabric failover on the UCS.

For the Cisco Nexus 1000V Series Switch to receive the Cisco Discovery Protocol packets to help display a management connectivity view on this adapter, the configuration on the Cisco UCS dual 10GE and dual 10G FCoE service profiles need to mark the pair of adapters with the network policy to receive the Cisco Discovery Protocol information.

To achieve the goal of high availability for the virtual machines on the VMware ESX host using this type of adapter, the use of uplink MAC pinning on the Cisco Nexus 1000V Series Switch is recommended. To configure MAC pinning on the Ethernet type (uplink) port profile, the administrator needs to enter only the following under the Cisco Nexus 1000V Series Ethernet port profile:

```
port-profile type ethernet Qlogic-no-ha-uplink
...
channel-group auto mode on mac-pinning
...
```

The Cisco Nexus 1000V Series uplink port profile that is to be assigned to the Cisco UCS servers adapters should have a descriptive name, because VMware vCenter will display these names, and the administrator will map the Cisco Nexus 1000V Series Ethernet ports to the appropriate uplink port groups.

The relevant portion of the Cisco Nexus 1000V Series configuration is shown below:

```
<config-snippet>
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
</config-snippet>
```

Service Profile Template Configuration

The following rules apply in the creation of a Cisco UCS M71KR-E and M71KR-Q service profile for hosting a VMware ESX instance that will be included in a Cisco Nexus 1000V Series system:

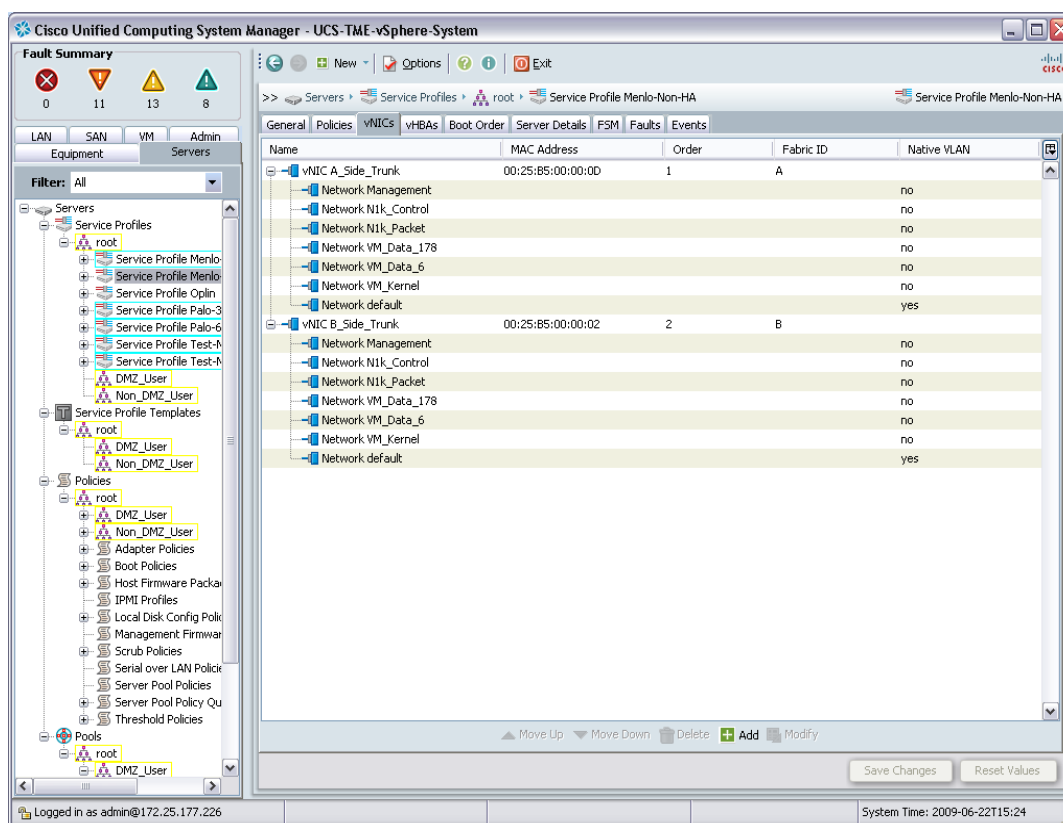
The administrator must define the VLANs that will initially be passed to the adapter and thus to the Cisco Nexus 1000V Series VEM.

The network administrator should mark the adapter policy for Cisco Discovery Protocol if a management connectivity view on the ESX host is desired.

The network administrator should define a default QoS profile for any traffic that the Cisco Nexus 1000V Series Switch does not mark.

Figure 9 shows an example of a service profile with the pertinent service profile networking configuration.

Figure 9. Cisco UCS Dual 10GE and 10G FCoE Service Profile: vNIC Definitions for the Cisco Nexus 1000V Series



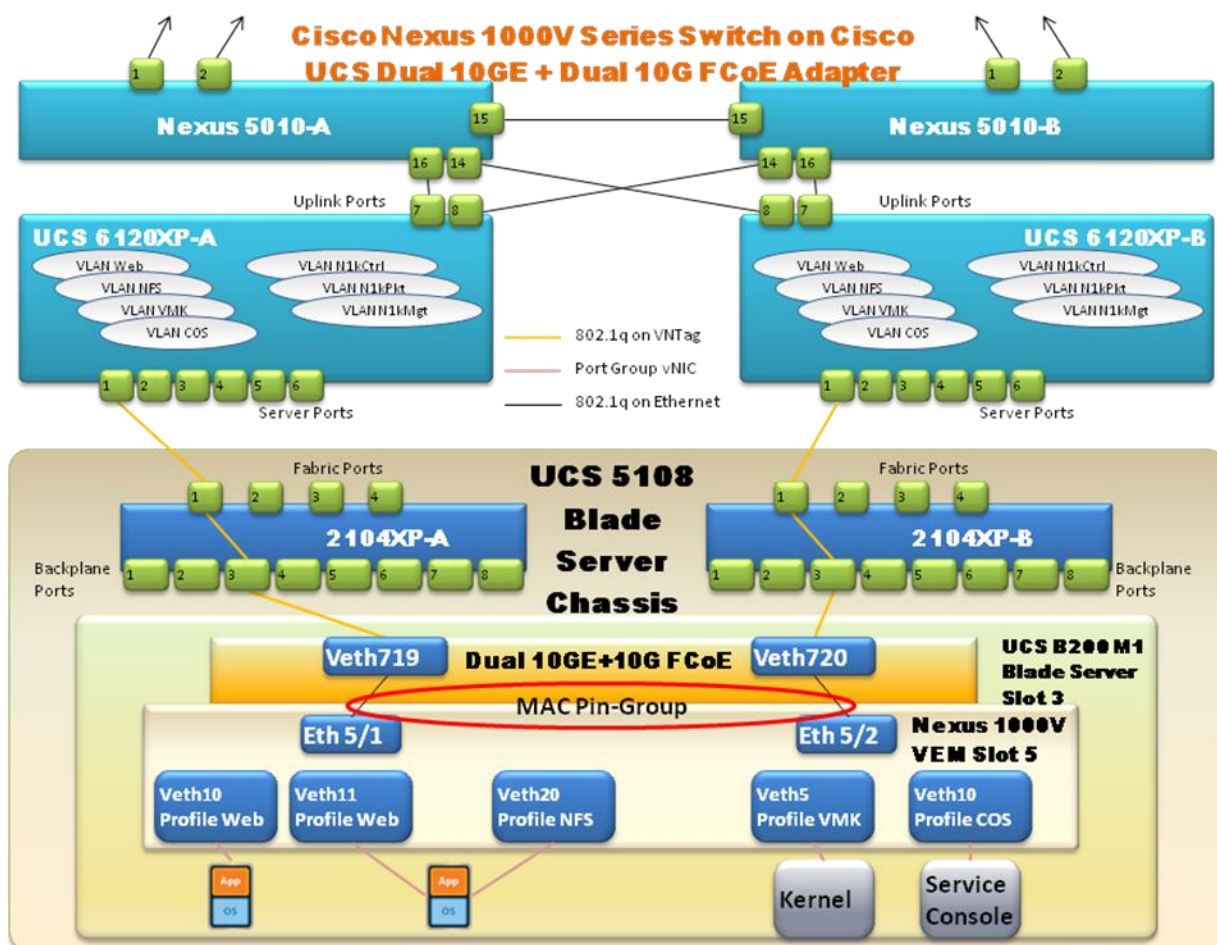
Operations

As discussed earlier, the design requires some trade-offs. Since multiple adapters cannot be mapped to the host, the user will need to rely on the QoS settings from the Cisco Nexus 1000V Series Switch to help ensure smooth operation. This solution allows the user to define a primary and a backup path for the console, kernel, production, or other interfaces to the fabric. This is accomplished by assigning the preferred interface in the same manner as for the Cisco UCS dual 10GE adapter family.

Performance and Failover Test

To see the basic performance data and failover times that can be expected, an iPerf test was performed from a virtual machine running behind the Cisco Nexus 1000V Series Switch to a host external to the Cisco Unified Computing System. Figure 10 shows the testing topology with a 10GE and 10G FCoE adapter.

Figure 10. Cisco Nexus 1000V Series Switch on Cisco UCS Dual 10GE and Dual 10G FCoE Mezzanines



Not shown in Figure 10 is a traffic target server with a Gigabit Ethernet connection to Cisco Nexus 5010 Switch A.

- Baseline observations:
 - A series of 90-second iPerf flow tests were performed.
 - Bidirectional throughput on a single virtual machine behind the Cisco Nexus 1000V Series VEM in the Cisco Unified Computing System and a single external Gigabit Ethernet host was 490 Mbps.
 - All MAC addresses were visible on the Cisco Nexus 1000V Series, Cisco Nexus 5010, and Cisco UCS 6120XP.
 - The B fabric was the side that was actively passing traffic by default. This verifies the pinning ID of 1 (or vmnic1, which was a B-fabric adapter).
 - Three threads on the Cisco UCS blade (out of the 16 in the pair of E5520 CPUs) hovered around 30 percent for the duration of the test.
- Observations about failure of the link from Cisco 6120XP B to Cisco Nexus 5010 A (active path for flow: Cisco Unified Computing System repinning to another uplink will occur):
 - A 90-second iPerf unidirectional (virtual machine to external host) flow test was performed.
 - On failure, the MAC address still remained on the B fabric.

- Cisco Unified Computing System repinning time for this flow was 1.16 seconds.
- The Cisco Nexus 1000V Series Switch did not detect a link outage.
- When this same test was run in the opposite direction (external host to virtual machine), the repinning time was measured at 2.3 seconds.
- When the Cisco Nexus 5010 Switches were configured in a vPC and this link failure occurred, the PortChannel selected the surviving member for the flow (instead of Cisco Unified Computing System repinning). The failover time in this scenario was 0.10 second, and failback was 0.03 second.
- Observations about failure of the link from Cisco 6120XP B to Cisco Nexus 5010 B (current active path for flow: a Cisco Nexus 1000V Series Switch failover of the VMNICs will occur):
 - A 90-second iPerf unidirectional (virtual machine to external host) flow test was performed.
 - On failover, the MAC address appeared on the A fabric.
 - Failover time for this flow was 2.54 seconds.
 - When this same test was run in the opposite direction (external host to virtual machine), the failover time was 2.5 seconds.
 - On link restoration (to restore the test bed), the Cisco Nexus 1000V Series Switch took less than 0.15 second to restore traffic to the preferred fabric (due to the pinning ID setup).
 - Restoration of the second uplink did cause Cisco Unified Computing System repinning to the original path, and this took 1.03 second.
 - The Cisco Nexus 1000V Series MAC pinning provided high availability.

Cisco UCS Virtual Interface Card Use Cases for the Cisco Nexus 1000V Series

High Availability with Virtual Interface Card

The Cisco UCS M81KR and P81E adapters function in a unique way in that they can be fully virtualized and presented not only to the CPU and BIOS as multiple adapters, but it can also be defined on the Cisco UCS 6100 Series Fabric Interconnects as different adapters. The management and control of these definitions are within the capability of the Cisco UCS Manager service profile. This adapter fully utilizes the VNTag technology and an additional protocol called the Virtual Interface Control Protocol for adapter operations. This VNTag mode enables a fabric failover mode (in which the virtual adapters register with both the Cisco UCS 6100 Series Fabric Interconnects), using one of the paths as an active path and the other as a backup path. The host will see multiple 10 Gigabit Ethernet Cisco adapters that are mapped to either the A side or the B side in the Cisco Unified Computing System. The Cisco Unified Computing System administrator has the option of running these adapters in a fabric failover mode; however, high availability would then be maintained within the fabric and not on the host.

As with the Cisco UCS M71KR ASIC, Cisco microcode is executed on this adapter, and the adapter performs a registration process to the fabric interconnects at server startup. Note that the Cisco UCS VIC can operate in a pair of discrete modes as mentioned earlier. With the static VIF mode, the idea is that the administrator will create and maintain all server interfaces. In this mode, the administratively defined adapter is its own endpoint, and since the VNTag mechanism uniquely identifies the adapter, the registration process will register a unicast MAC address for the adapter. When source traffic is seen on the fabric interconnects with a source MAC address behind the Cisco UCS VIC adapters, the MAC address table on the active link's fabric interconnect will be dynamically updated. This information is replicated to the backup fabric interconnect.

The Cisco UCS M81KR or P81E dynamic VIF mode provides an alternative operating model in which each virtual interface is created, deleted, and maintained by the VMware vCenter communications path during virtual machine operations within the cluster. The goal is for these adapters to be coordinated and presented directly to the virtual machines, with no additional virtual switch (vSwitch), distributed virtual switch (DVS), or Cisco Nexus 1000V Series management overhead needed to help ensure port-group consistency as these machines are migrated within a cluster. This mode is referred to as VM-FEX and runs on the Cisco UCS M81KR and P81E adapters only, and the host cannot be a blade within the Cisco Nexus 1000V Series. Since this document focuses on the Cisco Nexus 1000V Series interaction with the Cisco Unified Computing System, these modes are not discussed further here.

The key to this document however, is the fact that if a service profile includes dynamic vNIC definitions, then the installed Nexus 1000V VEM module will switch to a mode to support the VN-Link in hardware and will no longer operate as a traditional Nexus 1000V.

For the Cisco Nexus 1000V Series Switch to receive the Cisco Discovery Protocol packets to help display a management connectivity view on this adapter, the configuration on the Cisco UCS VIC service profiles need to mark the pair of adapters with the network policy to receive the Cisco Discovery Protocol information.

To achieve the goal of high availability for the virtual machines on the VMware ESX host using Cisco UCS M81KR, the use of uplink MAC pinning is recommended. To configure MAC pinning on the Ethernet type (uplink) port profile, the administrator needs to enter only the following under the Cisco Nexus 1000V Series Ethernet port profile:

```
port-profile type ethernet VIC-vm-data6-no-ha-uplink
...
channel-group auto mode on mac-pinning
...
```

The Cisco Nexus 1000V Series uplink port profile that is to be assigned to the Cisco UCS VIC should be given a descriptive name because VMware vCenter will display these names, and the administrator will map the Cisco Nexus 1000V Series Ethernet ports to the appropriate uplink port groups.

The relevant portion of the Cisco Nexus 1000V Series configuration is shown below:

```
<config-snippet>
port-profile type ethernet VIC-nlk-aipc-no-ha-uplink
vmware port-group
switchport mode trunk
switchport trunk allowed vlan 4-5
channel-group auto mode on mac-pinning
no shutdown
system vlan 4-5
state enabled
port-profile type ethernet VIC-vm-data6-no-ha-uplink
```

```
vmware port-group
switchport mode trunk
switchport trunk allowed vlan 1,6
channel-group auto mode on mac-pinning
no shutdown
state enabled
port-profile type ethernet VIC-vmk-sc-no-ha-uplink
vmware port-group
switchport mode trunk
switchport trunk allowed vlan 2-3
channel-group auto mode on mac-pinning
no shutdown
system vlan 2-3
state enabled
interface Ethernet4/1
inherit port-profile VIC-n1k-aipc-no-ha-uplink
interface Ethernet4/2
inherit port-profile VIC-n1k-aipc-no-ha-uplink
interface Ethernet4/3
inherit port-profile VIC-vm-data6-no-ha-uplink
interface Ethernet4/4
inherit port-profile VIC-vm-data6-no-ha-uplink
interface Ethernet4/5
inherit port-profile VIC-vmk-sc-no-ha-uplink
interface Ethernet4/6
inherit port-profile VIC-vmk-sc-no-ha-uplink
</config-snippet>
```

One advantage of the Cisco Nexus 1000V Series Switch on the Cisco UCS M81KR and P81E should be apparent: the capability to define unique interfaces for uplink to specific areas of the network, such as to constrain and pass VMware VMkernel traffic on alternative links to the location where the virtual machine production traffic normally passes. Another is the capability to run a Cisco Nexus 1000V Series Switch alongside a typical vSwitch or virtual distributed switch (vDS) on the same blade. The general recommendation is to still run a pair of 10GE interfaces for simplicity, however in this example we will break this out into 3 pairs to show this capability.

Service Profile Template Configuration

The following rules apply in the creation of a Cisco UCS VIC service profile for hosting a VMware ESX instance that will be included in a Cisco Nexus 1000V Series system:

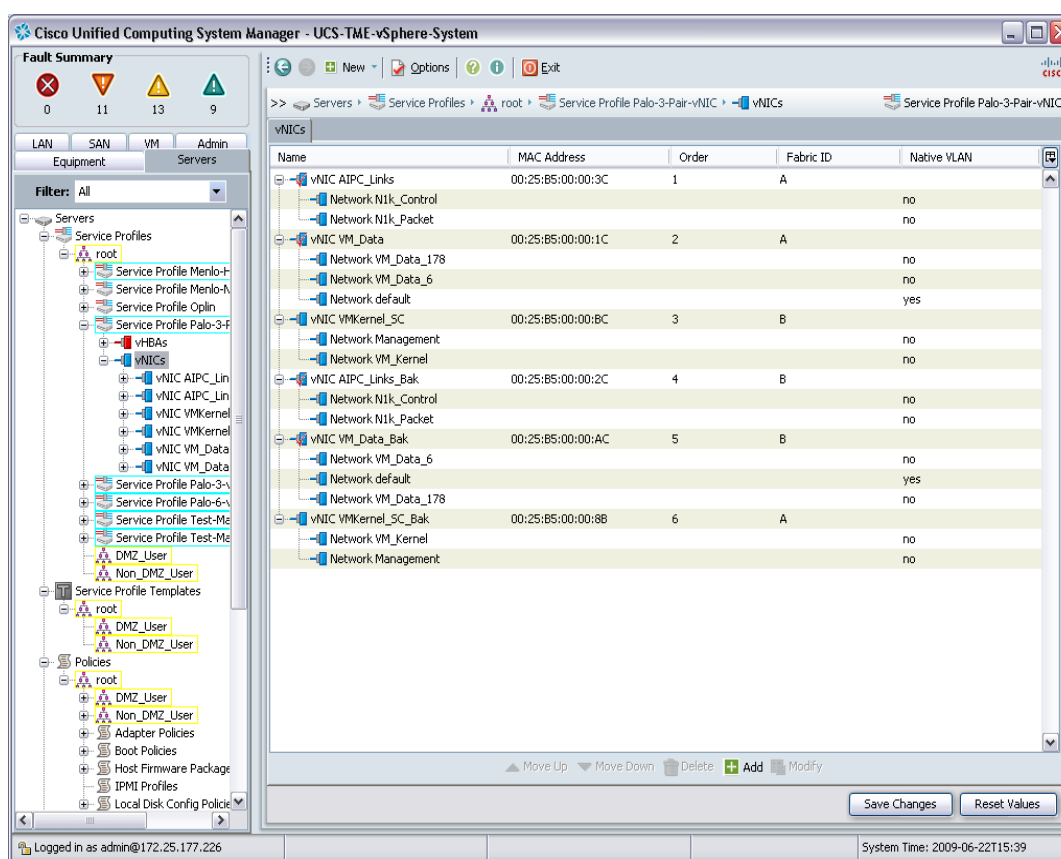
The network administrator must define the VLANs that will initially be passed to the adapter and thus to the Cisco Nexus 1000V Series VEM.

The network administrator should mark the adapter policy for Cisco Discovery Protocol if a management connectivity view on the VMware ESX host is desired.

The network administrator should define a default QoS profile for any traffic that the Cisco Nexus 1000V Switch does not mark.

Figure 11 shows an example of a service profile with the pertinent service profile networking configuration.

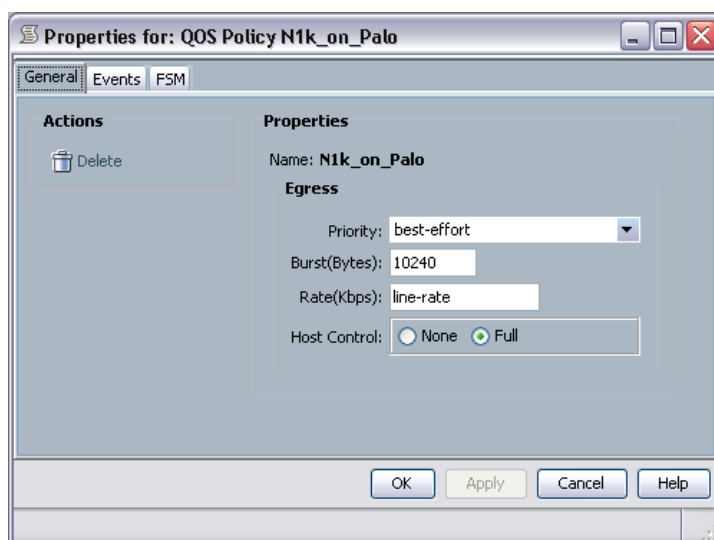
Figure 11. Cisco UCS M81KR Service Profile: Three Pairs of vNIC Definitions for the Cisco Nexus 1000V Series



Trusted and Untrusted Operations on the Cisco UCS M81KR VIC Adapter

Within the service profile for a Cisco UCS M81KR enabled server, the administrator can choose to enable or disable the trusted mode of operation. The trusted mode of operation is the recommendation for situations in which the Cisco Nexus 1000V Series Switch will be marking traffic and expecting QoS identifiers to be used and honored throughout the fabric.

Figure 12 shows how to set this mode.

Figure 12. Cisco UCS M81KR Rate-Limiting and QoS Trust Selection

The main item here is the Host Control selection, where you can set the adapter to impose the QoS marking on a per-adapter basis or to use the marking that is set by the server OS or the Cisco Nexus 1000V Series VEM. An important note here is that if you select the trusted mode (Host Control = Full), the entire physical VIC will use this mode, and all other vNIC instantiations on this adapter will also then be trusted. This model closely resembles the operation (from a QoS perspective) of the Cisco UCS M71KR, M72KR, and dual 10GE adapters. In untrusted mode, the Cisco UCS M81KR VIC interfaces will always set the desired QoS marking based on the adapter's QoS profile. When the administrator makes changes to the Host Control selection, the VMware ESX service profile must be rebooted for this function to operate correctly. This is also the location where the bandwidth of the vmnic within the ESX server can be set (in 1Mbps increments).

Operations

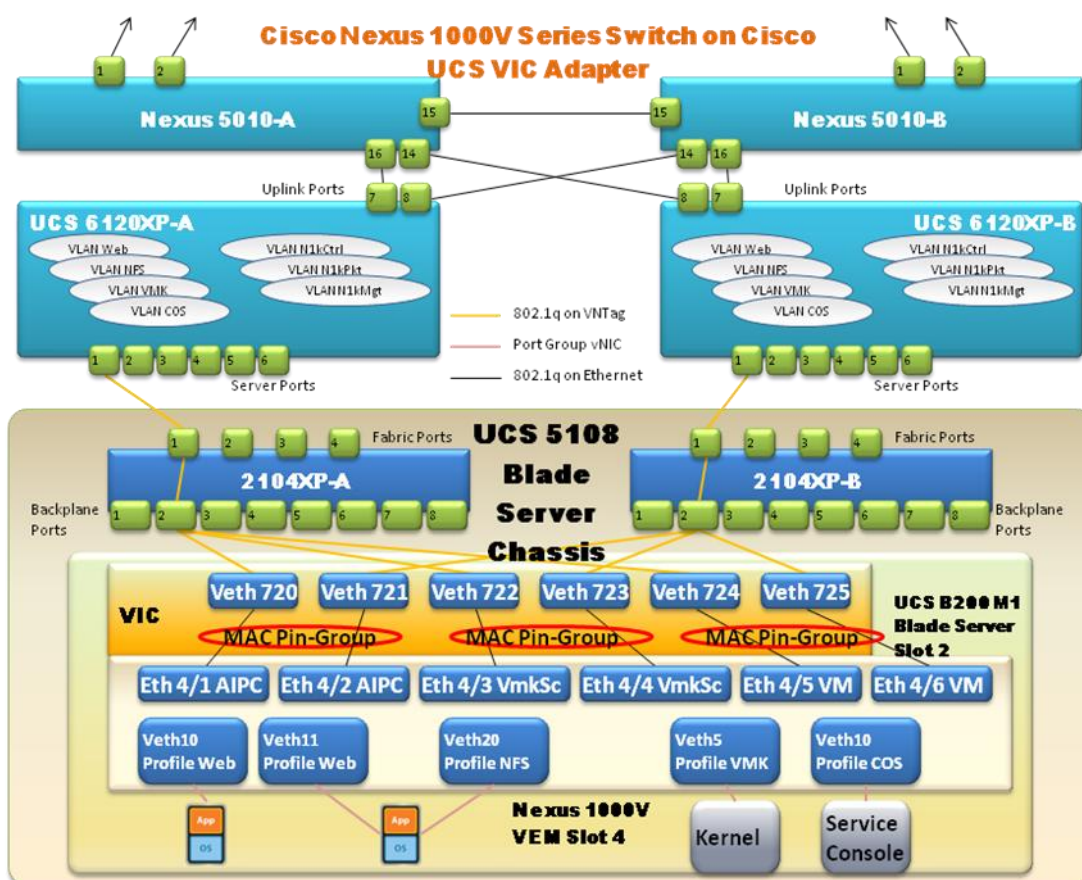
As discussed earlier, the design requires some trade-offs. The CoS markings from the Cisco Nexus 1000V Series are only used when the VIC is configured for full host control. When no host control is selected, the user can select multiple adapters to be defined on the host, and the user will need to map the system QoS settings on each adapter in the Cisco Unified Computing System to help ensure smooth operation. This solution allows the user to define a primary and a backup path for the console, kernel, production, and other interfaces to the fabric.

Note that the Cisco UCS VIC can support enough interfaces so that migrating the service console to the VEM may be unnecessary, because these additional adapters can also be used by VMware's integrated software switch if that customer use case is required.

Performance and Failover Test

To see the basic performance data and failover times that can be expected, an iPerf test was performed from a virtual machine running behind the Cisco Nexus 1000V Series Switch to a host external to the Cisco Unified Computing System. Figure 13 shows the testing topology.

Figure 13. Cisco Nexus 1000V Series Switch on Cisco UCS M81KR



Not shown in Figure 13 is a traffic target server with a Gigabit Ethernet connection to Cisco Nexus 5010 Switch A.

- Baseline observations:
 - A series of 90-second iPerf flow tests were performed.
 - Bidirectional throughput on a single virtual machine behind the Cisco Nexus 1000V Series VEM in the Cisco Unified Computing System and a single external Gigabit Ethernet host was 490 Mbps.
 - All MAC addresses were visible on the Cisco Nexus 1000V Series, Cisco Nexus 5010, and Cisco UCS 6120XP.
 - The A fabric was the side that was actively passing traffic by default. This verifies the pinning ID of 1 (or vmnic1, which was an A-fabric adapter in this profile).
 - Three threads on the Cisco UCS blade (out of the 16 in the pair of E5520 CPUs) hovered around 30 percent for the duration of the test.
- Observations about the failure of the link from Cisco 6120XP B to Cisco Nexus 5010 A (active path for flow: Cisco Unified Computing System repinning to another uplink will occur):
 - A 90-second iPerf unidirectional (virtual machine to external host) flow test was performed.
 - On failure, the MAC address still remained on the A fabric.
 - Cisco Unified Computing System repinning time for this flow was 1.15 seconds.

- The Cisco Nexus 1000V Series Switch did not detect a link outage.
- When this same test was run in the opposite direction (external host to virtual machine), the repinning time was measured at 2.2 seconds.
- When the Cisco Nexus 5010 Switches were configured in a vPC and this link failure occurred, the PortChannel selected the surviving member for the flow (instead of Cisco Unified Computing System repinning). The failover time in this scenario was 0.12 second, and failback was 0.04 second.
- Observations about the failure of the link from Cisco 6120XP B to Cisco Nexus 5010 B (current active path for flow: a Cisco Nexus 1000V Series failover of the VMNICs will occur):
 - A 90-second iPerf unidirectional (virtual machine to external host) flow test was performed.
 - On failover, the MAC address appeared on the A fabric.
 - Failover time for this flow was 2.6 seconds.
 - When this same test was run in the opposite direction (external host to virtual machine), the failover time was 2.6 seconds.
 - On link restoration (to restore the test bed), the Cisco Nexus 1000V Series Switch took less than 0.15 second to restore traffic to the preferred fabric (due to the pinning ID setup).
 - Restoration of the second uplink did cause Cisco Unified Computing System repinning to the original path, and this took 1.0 second.
 - The Cisco Nexus 1000V Series MAC pinning provided high availability.

Cisco Nexus 1000V Series Virtual Switch Module Placement in the Cisco Unified Computing System

This section provides recommendations for the placement of the Virtual Supervisor Module (VSM) relative to the Cisco Unified Computing System and the physical Ethernet interface counts. The options are listed in order of recommended customer use in a Cisco Unified Computing System environment.

Option 1: VSM External to the Cisco Unified Computing System on the Cisco Nexus 1010

In this scenario, management operations of the virtual environment is accomplished in an identical method to existing non-virtualized environments. With multiple VSM instances on the Nexus 1010 multiple vCenter data centers can be supported.

Option 2: VSM Outside the Cisco Unified Computing System on the Cisco Nexus 1000V Series VEM

This model allows for centralized management of the virtual infrastructure and has proven very stable..

Option 3: VSM Outside the Cisco Unified Computing System on the VMware vSwitch

This model provides isolation of the managed devices, and it migrates well to the appliance model of the Cisco Nexus 1010 Virtual Services Appliance. A possible concern here is the management and operational model of the networking links between the VSM and the VEM devices.

Option 4: VSM Inside the Cisco Unified Computing System on the VMware vSwitch

This model also was stable in the test deployments. A possible concern here is the management and operational model of the networking links between the VSM and the VEM devices, and having duplicate switching infrastructures within your Cisco Unified Computing System.

Cisco Unified Computing System and Cisco Nexus 1000V Series Versions Used in This Document

The Cisco Unified Computing System software for this test was Version 1.4(1m). The Cisco Nexus 1000V Series software used in this testing was Version 1000v.4.0.4.SV1.3a.

Conclusion

When using the Cisco Nexus 1000V Series within the Cisco Unified Computing System, always do so without any adapter fabric failover for the reasons discussed earlier, and leave the hashing to the Cisco Nexus 1000V Series MAC pinning function. This recommendation also applies in a scenario with Cisco UCS VIC without dynamic VIF policies using VN-Link in Hardware.

For More Information

For more information about the Cisco Nexus 1000V Series, please refer to www.cisco.com/go/nexus1000.

For more information about the Cisco Unified Computing System, please refer to www.cisco.com/go/unifiedcomputing.

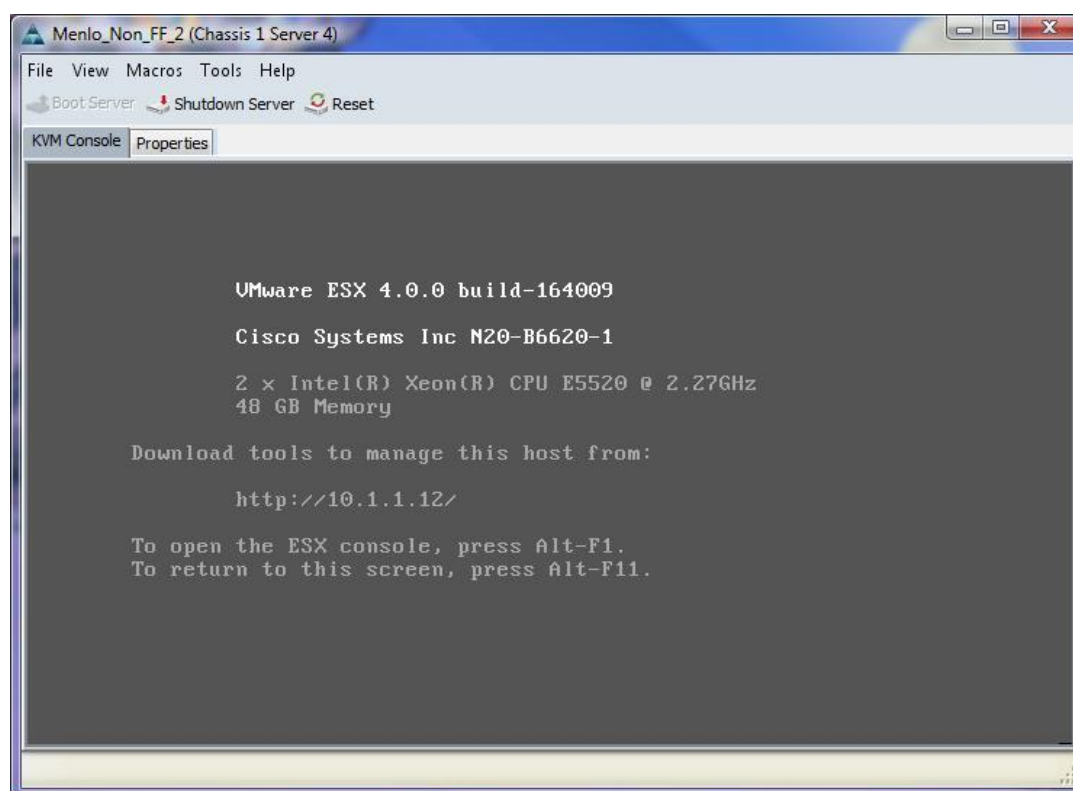
For more information about the IEEE work on virtual Ethernet bridging, please refer to <http://www.ieee802.org/1/pages/dcbridges.html>.

For more information about Intel VMDq technology, please refer to http://www.intel.com/network/connectivity/vtc_vmdq.htm.

Appendix A: Troubleshooting VMware ESX Server in the Cisco Unified Computing System

You may need to troubleshoot when using the Cisco Nexus 1000V Series in a Cisco Unified Computing System environment if the system loses communication with the VMware ESX server, preventing further operations. The first step in resolving this situation is to troubleshoot the connection to the console port directly in the blade running VMware ESX. In the Cisco Unified Computing System, when a service profile is instantiated on a blade, an option will be presented to open a KVM console session to that particular blade. This option will open a Java plug-in to perform the remote KVM console operations and to share media directly from the administrator's machine. A screen similar to the one in Figure 14 will appear. From the KVM console, you can press ALT-F1 to open a login screen, which you can use to log in to VMware ESX directly.

Figure 14. KVM Console to a Cisco UCS Blade Running VMware ESX



One of the most common troubleshooting situations arises when an error occurs in migrating a VMware ESX host service console or VMware VMkernel interface to the Cisco Nexus 1000V Series during installation. Although the use of Cisco Nexus 1000V Series system VLANs can help here, this document discusses a recovery mechanism. The sequence of recovery is to list the port on the DVS to which the vswif interface is connected, remove the vswif from the DVS, add a service console port group back into the vSwitch, and add a vswif interface to that port group. The commands for this process are listed in the following section.

VMware ESX and ESXi Commands to Troubleshoot Cisco Nexus 1000V Series Switches in the Cisco Unified Computing System

In VMware ESX with a service console OS, these commands are in the /usr/sbin directory. Press ALT-F1 to display the service console login prompt, and press ALT-F11 to leave the prompt. To get to a misconfigured Cisco UCS blade for recovery, click the KVM item under the server view. This procedure assumes that the KVM IP addresses have been set up and assigned.

In VMware ESXi, the esxcfg-vswitch functions are all present as in VMware ESX. Since VMware ESXi does not have a service console OS, use ALT-F2 and ALT-F1 key presses to go to and from the VMware ESXi command line (Assuming local tech support mode is enabled). All the commands listed here for esxcfg-vswif will now be replaced with esxcfg-vmknic, and you can update specific details as required.

Most Common Desire: Recover A Host New Mgmt I/F (like ESXi does)

```
esxcfg-vswitch -l | more          (find the DVPort ID of the vmnic0 and vswif0,
and the DVS name)

esxcfg-vswitch -Q vmnic0 -V [vmnic0 port#] DVS_NAME

esxcfg-vswitch -L vmnic0 vSwitch0

esxcfg-vswitch -A "Service Console" vSwitch0

esxcfg-vswitch -p "Service Console" -v 2 vSwitch0          (just an example of
putting it on VLAN2)

esxcfg-vswif -d vswif0 -P [vswif0 port#] -V DVS_NAME

esxcfg-vswif -a vswif0 -p "Service Console" -i 10.1.1.10 -n 255.255.255.0
```

To manage the instance of a VEM on the ESX server:

```
vem-remove -d                    To remove a currently installed VEM

esxupdate -bundle=ftp://Administrator:password@FTP_IP/VEM_Path/VEM410-
201007311.zip update            To download and install the VEM from an FTP
server
```

To remove the DVS extension from the vCenter:

<http://IP-of-vCenter/mob>

content -> extensions

remove extension without quotation marks

To list information on current configuration:

```
esxcfg-vswif -l                To list information on the current vswif's
esxcfg-vswitch -l              To list information on the current vswitches and DVS'
esxcfg-vmknic -l               To list information on the current vmk's
```

To manage the existence of a vSwitch:

`esxcfg-vswitch -a vSwitch0` To add a vSwitch named vSwitch0 to the server
`esxcfg-vswitch -d vSwitch0` To delete a vSwitch named vSwitch0 from server

To manage the MTU of a vSwitch:

`esxcfg-vswitch -m 9000 vSwitch0` To set Jumbo Frames on a vswitch and underlying vmnics

To manage the port-groups on a vSwitch:

`esxcfg-vswitch -A "Service Console" vSwitch0`

To add a port-group named Service Console to vSwitch0

`esxcfg-vswitch -D "Service Console" vSwitch0`

To remove a port-group named Service Console from vSwitch0

To assign VLAN membership to a port-group on a vSwitch:

`esxcfg-vswitch -p "Service Console" -v 102 vSwitch0`

To set a port-group named Service Console to use VLAN 102 on vSwitch0

To view available NICs that ESX can see:

`esxcfg-nics -l` To see list of vmnics and the properties (MTU, etc.)

To manage uplinks from the DVS or N1k:

First, you run the list commands (first in this section) to get the port numbers for the existing uplinks (vmnic0, vmnic1, etc.) – call this port#.

`esxcfg-vswitch -Q vmnic0 -V [port#] myN1kDVS`

To remove the vmnic0 uplink from the DVS named myN1kDVS

`esxcfg-vswitch -P vmnic0 -V [port#] myN1kDVS`

To add the vmnic0 uplink to the DVS named myN1kDVS

To add the vmnic's back into a DVS, we suggest you do this via vCenter and not the command line

To manage uplinks to a vSwitch:

`esxcfg-vswitch -L vmnic0 vSwitch0`

To add vmnic0 as an uplink to vSwitch0

`esxcfg-vswitch -U vmnic0 vSwitch0`

To remove vmnic0 as an uplink to vSwitch0. To remove a vmnic from a vSwitch, we suggest you do this via vCenter and not the command line while migrating from vSwitch to N1k DVS

To delete a vmk from a ESX server port group:

`esxcfg-vmknic -d -p "port_group"`

To delete a vmk from the port-group on vSwitch.

`esxcfg-vmknic -d -v xxx -s myN1kDVS`

To delete a vmk from the port-group on the Nexus1000V.

To delete a vswif from a ESX server port group:

```
esxcfg-vswif -d vswif0
```

To delete a vswif

```
esxcfg-vswif -d vswif0 -p "Service Console" -P xxx -V myN1kDVS
```

To delete a vswif from the port-group on the Nexus1000V. This is a common task if an error situation arises, and you lose connectivity to the ESX server when migrating to Nexus 1000V. You will have to add the port-group to the vSwitch0, then add the vswif on it.

To add a vswif to the ESX server:

```
esxcfg-vswif -a vswif0 -p "Service Console" -i 10.1.1.10 -n 255.255.255.0
```

To add a vswif0 with IP 10.1.1.10/24 to port-group Service Console

```
esxcfg-vswif -a vswif0 -p "Service Console" -i 10.1.1.10 -n 255.255.255.0 -P [port#]-V myN1kDVS
```

To add a vswif0 with IP 10.1.1.10/24 to port-group Service Console to the port = port# on the DVS myN1kDVS. The port# is usually just the last unused entry to the list of available port numbers when listing the DVS configuration with esxcfg-vswitch -l.

To add a vmk to the ESX server:

```
esxcfg-vmknic -a -i 10.1.3.10 -n 255.255.255.0 -m 9000 "VMkernel"
```

To add a vmk with IP 10.1.3.10/24 to the vSwitch with Jumbo Frames

```
esxcfg-vmknic -a -i 10.1.3.10 -n 255.255.255.0 -m 9000 -s myN1kDVS
```

To add a vmk with IP 10.1.3.10/24 to the DVS with Jumbo Frames

To set the jumbo frames on a vmk on a ESX server:

```
esxcfg-vmknic -m 9000 -v xxx -s myN1kDVS
```

To manage a default gateway for the ESX server:

On the vCenter GUI, you can change this directly, but on the CLI, you must edit the /etc/sysconfig/network file and update the vswif's default gateway. The file will look similar to:

```
NETWORKING=yes
```

```
HOSTNAME=ESXServerBlade4.cisco.com
```

```
GATEWAY=10.1.1.1
```

After saving the file, the ESX server must be rebooted to accomplish the change.

Troubleshooting Commands

To Ping test out a VMkernel interface:

```
esxcfg-route -l          To get the routes on the VMKernel network
```

```
esxcfg-vmknic -l        To see the parameters and MTU on vmk interfaces
```

```
vmkping -s pkt_size -f X.X.X.X      To ping an IP on the VMKernel network
with packet size and do not fragment.
```

To test for the VEM installation on an ESX host:

```
vmkload_mod -l | grep vem      To ensure the appropriate VEM is loaded (should
return the modules)
```




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.

 Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: www.cisco.com/go/trademarks. Third party trademarks mentioned 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. (1110R)