White Paper

Cisco Unified Computing System (UCS) Storage Connectivity Options and Best Practices with NetApp Storage

Chris Naddeo, Cisco UCS Technical Marketing Engineer

Patrick Strick, Technical Marketing Engineer, Datacenter Platforms, NetApp

Contents

About the Authors

Chris Naddeo is a Technical Marketing Engineer for the Cisco Unified Computing System (UCS). His focus is on the storage connectivity aspects of UCS working closely with the storage partner ecosystem on joint projects, Cisco Validated Designs and certifications. His also is responsible for meeting with potential and existing UCS customers and educating Cisco field sales on the UCS platform. Prior to Cisco, Chris worked at NetApp where he was a field sales consulting system engineer focusing on Oracle on NetApp best practices as well as the Data ONTAP GX platform. Prior to NetApp he spent over 8 years at VERITAS/Symantec in a variety of roles including field sales specialist, Technical Product Manager, and Sr. Manager of Product Management. His focus while at VERITAS was on current and next generation volume management and file systems. Chris is based out of the Philadelphia, PA area

Patrick Strick is a Technical Marketing Engineer in NetApp's Datacenter Platforms group. He is currently focusing on ways to integrate NetApp and partner technology through joint best practices and product development. Previously, he led the technical marketing effort for Fibre Channel over Ethernet at NetApp and has over 10 years of experience in IT and storage administration and support.

Introduction

This paper will provide an overview of the various storage features, connectivity options, and best practices when using the Unified Computing System (UCS) with NetApp storage. This document will focus on storage in detail, both block and file protocols and all the best practices for using these features exposed in UCS with NetApp storage. There will not be an application or specific use case focus to this paper. There are existing Cisco Validated Designs that should be referenced for a deeper understanding of how to configure the UCS and NetApp systems in detail for various application centric use cases. These documents treat the combination of UCS and NetApp from a more holistic or, end to end approach and include the design details and options for the various elements of UCS and NetApp systems. The reader is encouraged to review these documents which are referenced below.

Designing Secure Multi-Tenancy into Virtualized Data Centers

http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/Virtualization/securecldg.html

Deploying Enhanced Secure Multi-Tenancy into Virtualized Data Centers

http://www.cisco.com/en/US/docs/solutions/Enterprise/Data_Center/Virtualization/securecldeployg_V2.html

Several FlexPod Cisco Validated Designs can be found at this URL

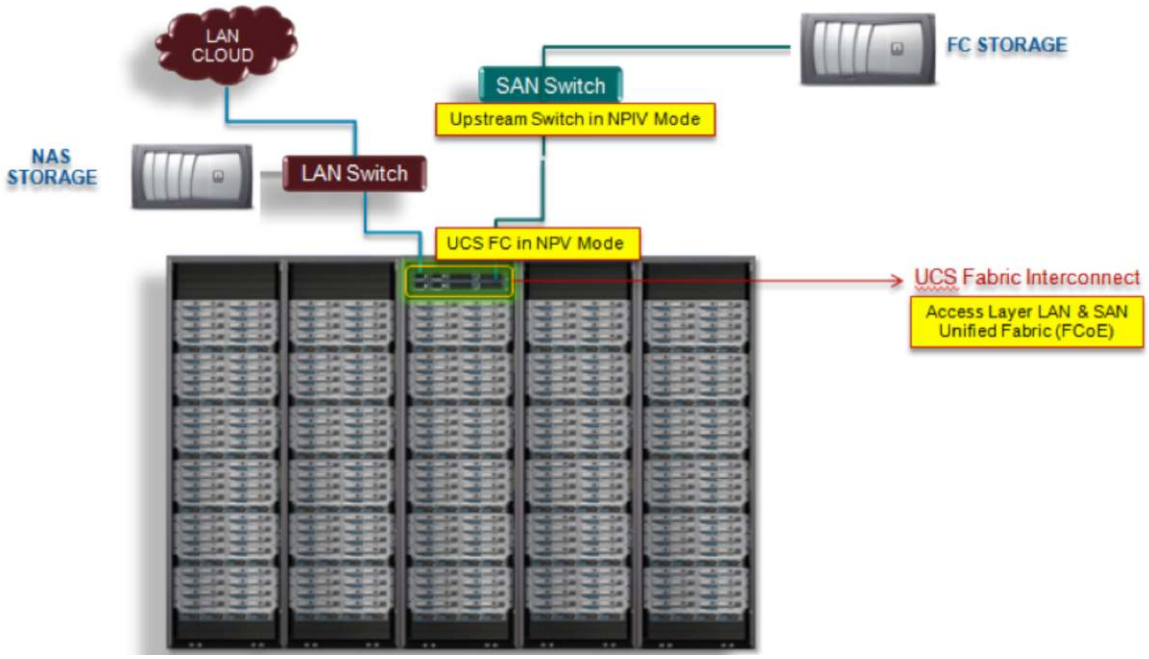
http://tools.cisco.com/search/JSP/search-results.get?strQueryText=flexpod+cisco+validated+design&Search+All+Cisco.com=cisco.com&autosuggest=true

The Cisco Unified Computing System (UCS) provides a flexible environment to host your virtualized applications. It also provides flexibility in the way enterprise storage can be provisioned to it. With this flexibility comes options and with options come questions. This paper will address many of those questions by outlining each of the ways storage can be attached and the recommended way it should be configured. The following examples use NetApp FAS storage appliances that can provide iSCSI, Fibre Channel, and FCoE SAN and CIFS and NFS NAS simultaneously from one unified storage system

UCS storage Overview

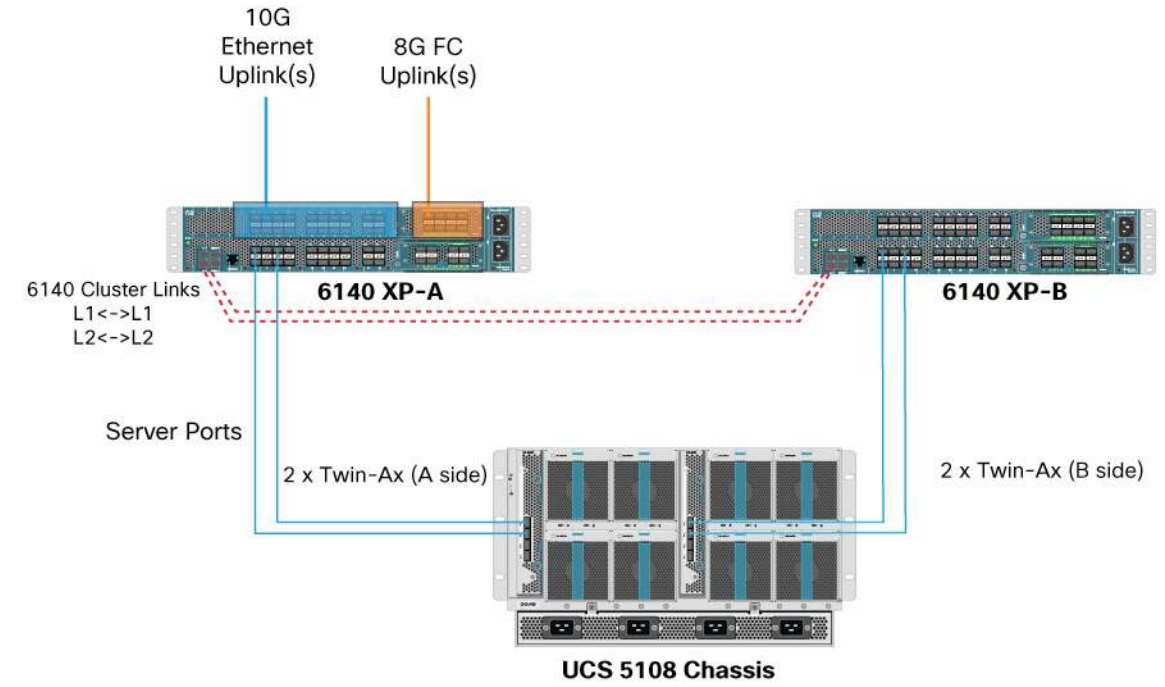
UCS is designed as a revolutionary compute blade architecture but from a storage perspective the entire UCS system appears to the outside world as a collection of Network Interface Cards (NICs) and Fibre Channel Host Bus Adapters (FC HBAs). This is the default mode of operation for UCS and the one that should be employed with the exception of those cases outlined in this paper. The UCS architecture is shown in the diagram below but it is important to realize that the Fabric Interconnects at the top of the UCS architecture are not general purpose Nexus FC or Ethernet switches, rather they are repurposed with special hardware and software to serve as “controllers” for the array of compute nodes behind them. As such they run in an “End Host Mode” of operation for both Ethernet and FC storage options. In the FC domain this is called NPort Virtualization or NPV mode. This mode allows the UCS Fabric Interconnects to act like a server allowing multiple hosts to login to the upstream fabric on the same number of FC uplinks

Figure 1. UCS and External Storage Connectivity Default Mode of Operation



Storage I/O enters and departs the UCS system on the Fabric Interconnect via the use of uplink ports. There are different types of uplink ports and as we shall see special port types when using UCS in a direct attach configuration. The paper will discuss direct connect in detail later, for our current purposes these are the uplink types described and then shown below in Figure 2

Figure 2. UCS Uplink Ports



Fibre Channel and FCoE

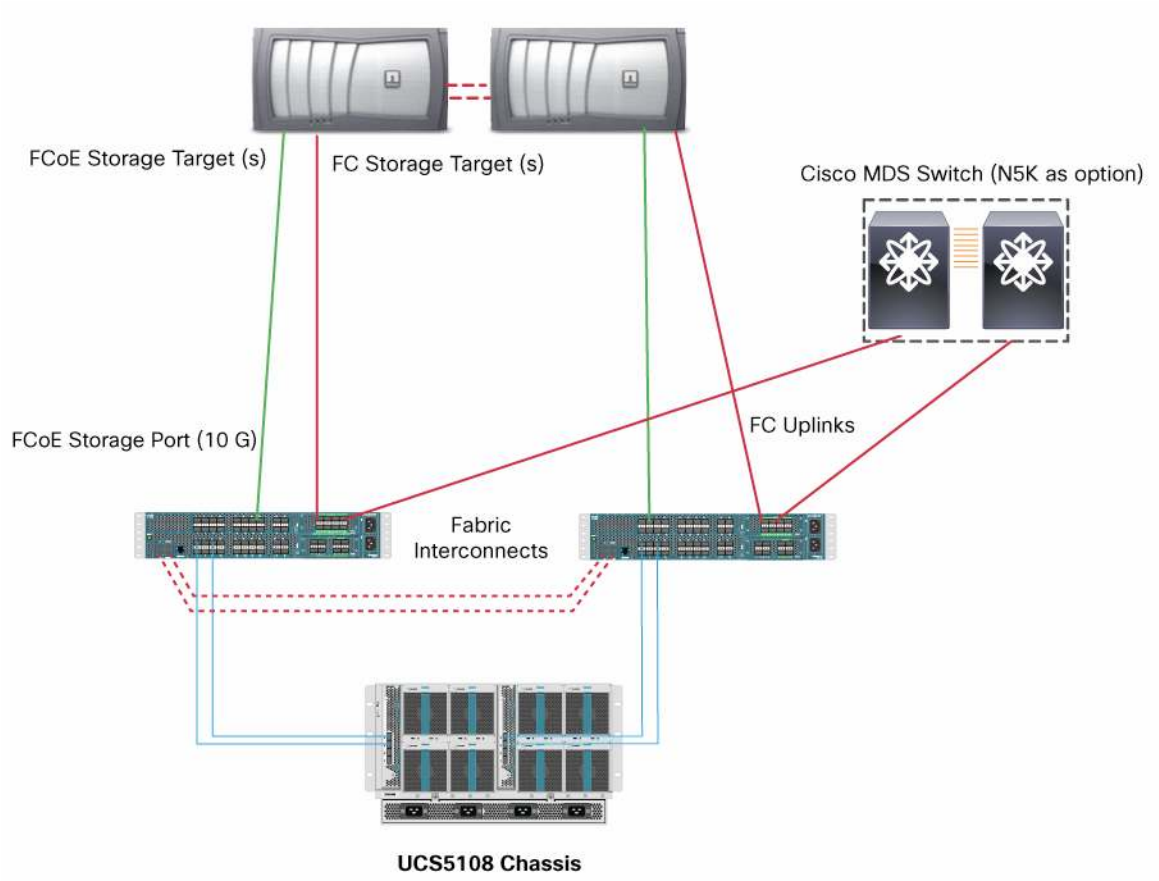
The use of FCoE in UCS is completely transparent to the host operating system. The OS simply sees 10G Ethernet and FC PCIe device handles and associated device drivers. It is possible for UCS to directly address FCoE storage targets via the use of the direct connect topology which will be discussed later. Many customers have deployed FCoE storage with NetApp and UCS using a Nexus 5000 family switch as a FC to FCoE bridge. A future version of UCS will support FCoE northbound out of the standard Ethernet uplinks.

A requirement for the upstream FC switch is that it is running in NPIV mode. Most modern FC switches on the market support this mode of operation. Note that UCS does not require the upstream FC switches to be Cisco as NPIV is an industry standard. Configured as shown here all of the blades in the UCS chassis appear to the upstream Fabric as several HBAs with unique WWPNs

FC and FCoE Direct Connect—FC Switch mode

UCS versions 1.4 and 2.0 do support directly connecting arrays to the Fabric Interconnects so long as an upstream Cisco MDS or Nexus 5000 FC switch is connected via one of the FC uplinks. This upstream FC switch (Cisco only) is necessary as these releases of UCS Manager do not support creating or managing zones on the fabric interconnects. The zones need to be created on the upstream MDS/N5K and the zones will be downloaded automatically to the UCS FIs once they are put into in FC switching mode. The following diagram shows a currently supported direct connect topology.

Figure 3. Supported Direct Connect FC and FCoE Topology



FC Adapter Policies

UCS Manager instruments polices to program the Cisco VIC for key FC parameters. Many customers use host based utilities for such tasks with Emulex or Qlogic based FC adapters. Appendix I discusses the FC adapter polices in detail with each attribute explained as to its function as well as guidance on when to change. It is important to note that in general one does not need change the default settings unless following the instructions indicated in Appendix I.

iSCSI

iSCSI is a protocol that enables the transport of SCSI commands over TCP/IP. It was ratified in 2004 and is documented in RFC3720. iSCSI is similar to Fibre Channel in that it allows multiple hosts to connect to a storage system for disk resources. The biggest difference is that iSCSI runs at the application layer of the networking stack on top of TCP/IP, where FC has its own entire stack and FCoE implements the upper layers of Fibre Channel on top of Ethernet at the network layer.

Because iSCSI runs on top of TCP/IP it is routable and does not require the administrator to have knowledge of zoning, WWPNs, fabric services, or other Fibre Channel SAN specifics. It only requires a stable Ethernet network and knowledge of host and storage system management.

To provide security and supportability iSCSI traffic should be isolated on its own VLAN or physical network. Without this separation, not only is the iSCSI data potentially visible to other devices on the network, but if a issue arises it will be more difficult to isolate the cause. Additionally, if more than one path from the host to the target is configured (MPIO), each path should be on its own network subnet (they can share the same VLAN). This will prevent potential routing issues.

iSCSI with Data ONTAP

For the iSCSI protocol in Data ONTAP 7.x and 8.x 7-mode, each storage controller in an HA pair services only its own LUNs unless in a controller failover. That means during normal operations the host will only access a LUN through the owning controller's network ports. This is different from Fibre Channel where the host can access the LUN through the ports on both controllers in the HA pair. With iSCSI, in the case of a storage failover, the partner node assumes ownership of the LUNs and the identity of the network ports. Once the takeover is complete, the host will reconnect its iSCSI session and any outstanding SCSI requests will be resent.

UCS and iSCSI Overview

UCS has supported iSCSI data LUN access via a variety of adapter models since UCS 1.0. This access was via the operating system iSCSI software initiators when the Cisco Virtual Interface Card (VIC) is used. Cisco optionally offers the Broadcom 57711 adapter which can be configured for hardware iSCSI outside of UCS using the host utilities provided by Broadcom. iSCSI Hardware Offload is not a requirement to support booting, rather the adapter only need support configuring the iSCSI Boot Firmware Table (iBFT) in the OptionROM of the adatper. This allows the booting feature to be supported on the Cisco VIC even though this adapter does not have any hardware iSCSI features. The need for hardware iSCSI on an adapter has been overcome in recent years with the power of the Intel processors. There are more than enough cores and GHz available to process the iSCSI traffic in additional to the application workload. Many customers struggle to drive CPU utilization to reasonable levels, thus fueling the trend towards high density virtualized environments.

UCS 2.0 introduced the ability to boot from iSCSI targets. This is done via a new object in UCS called an "iSCSI vNIC" which is an overlay or child object of the normal UCS "vNIC". The iSCSI vNIC has a variety of properties and consumes various pools and policies all of which are there to allow UCS to properly configure the Option ROM memory of the adapter. Once the server has completed the boot cycle then the iSCSI vNIC is not in the I/O path any longer and all the I/O is done via the normal vNICs and host iSCSI initiator if using the VIC adatper. UCS 2.0 supports the Cisco VIC and the Broadcom adapter models for this feature but NetApp has only formally tested and supports the VIC adapter. Broadcom will only be considered if a customer submits a product variance request (PVR) to NetApp.

The support of iSCSI boot now allows customers to deploy UCS with all NAS storage protocols and still continue to have the service profile mobility use case vs. requiring the OS to be installed on local drivers. There are some notable caveats that need mentioning for using this feature with UCS 2.0 (1)

UCS iSCSI Boot Caveats with the 2.0(1) Release

Not all adapters instantiate iSCSI HBAs to the operating system identically or at all. Contrast with vNICs and vHBAs which are standard across adapters in terms of what the OS "sees" and what UCSM shows in the equipment tab. The net effect of this is that the operating system view of iSCSI HBAs (PCIe devices) will diverge from what is shown in the UCS equipment tab. This has been addressed in the UCSM interface via the use of the Properties Attribute in properties screen to show distinction

- VIC: iSCSI vNICs are not visible to the operating system

- Broadcom: iSCSI vNICs are visible to the operating system

Other caveats of note

- iSCSI boot using Appliance Ports with C-Series servers under UCSM control is not supported until the 2.0(2) release of UCSM. The reader should not deploy this until NetApp completes qualification of this feature and topology.
- iSCSI Qualified Names (iQN) are not pooled, they must be manually entered for each initiator. Note that with software iSCSI initiators there is typically a single iQN per host. The Cisco VIC is not a hardware iSCSI implementation and as such after booting is complete it plays no role in the iSCSI I/O stack as the host initiator takes over. The user may want to use a single iSCSI vNIC or rename the individual iSCSI vNICs to match to accommodate this approach. iQN names are pooled in the UCS 2.0(2) release available in March 2012
- The Cisco VIC does not support non native VLAN boot. This must be selected during configuration.
- Configured boot order when looking at UCSM screens will show storage for the iSCSI targets when using the Cisco VIC. This is due to the SCSI interrupt vector that the VIC employs.
- There is no failover support for the iSCSI vNICs during the boot phase itself. If there is a path failure during the actual boot sequence the blade will have to be rebooted to use the remaining path. There are several ways to accomplish this which will be discussed in a later section on UCS iSCSI and fabric failover.
- You can configure a single or two iSCSI vNICs per service profile. However in the 2.0(1-) release there are some points to consider when choosing which to use.
- When you configure two iSCSI vNICs on two separate physical adapters (full width blade scenario) and if the secondary iSCSI vNIC is discovered first and the primary fails, boot will not succeed. The work around is to change boot policy by swapping the primary/secondary roles and re-associate
- If the primary boot path fails, even with a secondary configured on a separate adapter the boot will fail. Work around is to configure primary and secondary on the same adapter or use the process mentioned in the preceding bullet.
- EUI formatting for initiators is not supported in this release
- There are no statistics for iSCSI vNICs
- Some operating systems required a pingable gateway IP address during the iSCSI booting phase or the boot will fail. If the topology is using Appliance ports and there is no upstream network to supply this address then the solution here is to make the gateway IP address the same as the Storage controller IP.
- Windows 2K8 WFP Lightweight Filter Driver must be disabled manually on the secondary vNIC (not used for iSCSI boot) after the OS is up. If this is not done and primary path fails then subsequent reboot using secondary path will fail.
  - Known MS issue: http://support.microsoft.com/kb/976042
- Remove local drives in blades for Windows iSCSI.

UCS iSCSI Boot Changes with the 2.0(2) Release (Available March 2012) This maintenance release adds some enhancements to the iSCSI feature set. Some of the caveats listed above have been addressed

- The primary and secondary iSCSI vNICs can now span adapters and in all failure cases the iBFT will still get posted without service prolife reconfigurations or reboots.

- iQNs are now pooled and as such uniqueness of iQN names is enforced. If the user still desires 1 iQN per host then the recommendation is to only configure 1 iSCSI vNIC

UCS iSCSI Boot and Fabric Failover

The UCS feature of Fabric Failover (FF) is commonly used for vNICs which are used for Ethernet traffic in UCS. However with iSCSI the vNICs should not use the FF feature as the best practices is to use the host based multi-pathing drivers for all load balancing and failure handling. This is consistent with the FC world where no FF feature exists for UCS vHBAs. Additionally during boot or install time the FF feature is not even active or possible. The temporary management NIC interface created to program OptionROM with the boot parameters does not support the FF feature.

The best practice is to use host multi-pathing driver to handle iSCSI traffic load balancing and failover (analogous to FC protocol) and thus not use the FF feature. A common deployment would be to have two vNICs used for iSCSI traffic and then two or more additional vNICs used for standard Ethernet traffic. One could enable the FF feature on the standard Ethernet traffic vNICs, just not the ones used for iSCSI traffic. Remember all the iSCSI vNIC does is program the VIC to support booting, then it is out of the way and the host is just using operating system iSCSI initiators to access iSCSI LUNs.

Typical iSCSI SAN deployment—Upstream in End Host mode

The standard method for connecting storage to a Cisco UCS is with the Fabric Interconnect in Ethernet End Host mode and the storage target connected through an upstream switch or switched network. Although the Cisco UCS uses 10 GbE internally, the storage target does not have to because it is connected through your existing network. However, it certainly can if both the network and target support 10 GbE.

Whether the upstream network is 1 GbE or 10 GbE, using jumbo frames (e.g. an MTU size of 9000) will improve performance by reducing the number of individual frames that must be sent for a given amount of data and prevent having to break SCSI data blocks up over multiple Ethernet frames. These in turn also lower host and storage CPU utilization. If jumbo frames are used, you must be sure that all network equipment between and including the UCS and storage target are able and configured to support the larger frame size. These considerations can be discussed with the network administrator.

The host's configuration will depend on which CNA is installed in the UCS blade. The Cisco M81KR Virtual Interface Card (VIC) has the ability to have many virtual interfaces presented to the host from a single card. With this capability, vNICs can be created specifically for the iSCSI traffic. In UCS Manager, when creating the vNIC or vNIC template, select the iSCSI VLAN and assign it as the native VLAN. Note that VIC does not support **non native** VLAN booting with iSCSI.

Highlights of Configuring iSCSI boot with UCS Manager

We will take this from the perspective of creating a UCS service profile configured for iSCSI boot. This will show the overall flow as well as point out some necessary or optional steps. A later section will cover the caveats one needs to be aware of.

The overall steps are the following:

- Create vNICs using expert Mode
- Add iSCSI vNICs
- Overlay vNIC is the parent to the iSCSI vNIC you are creating

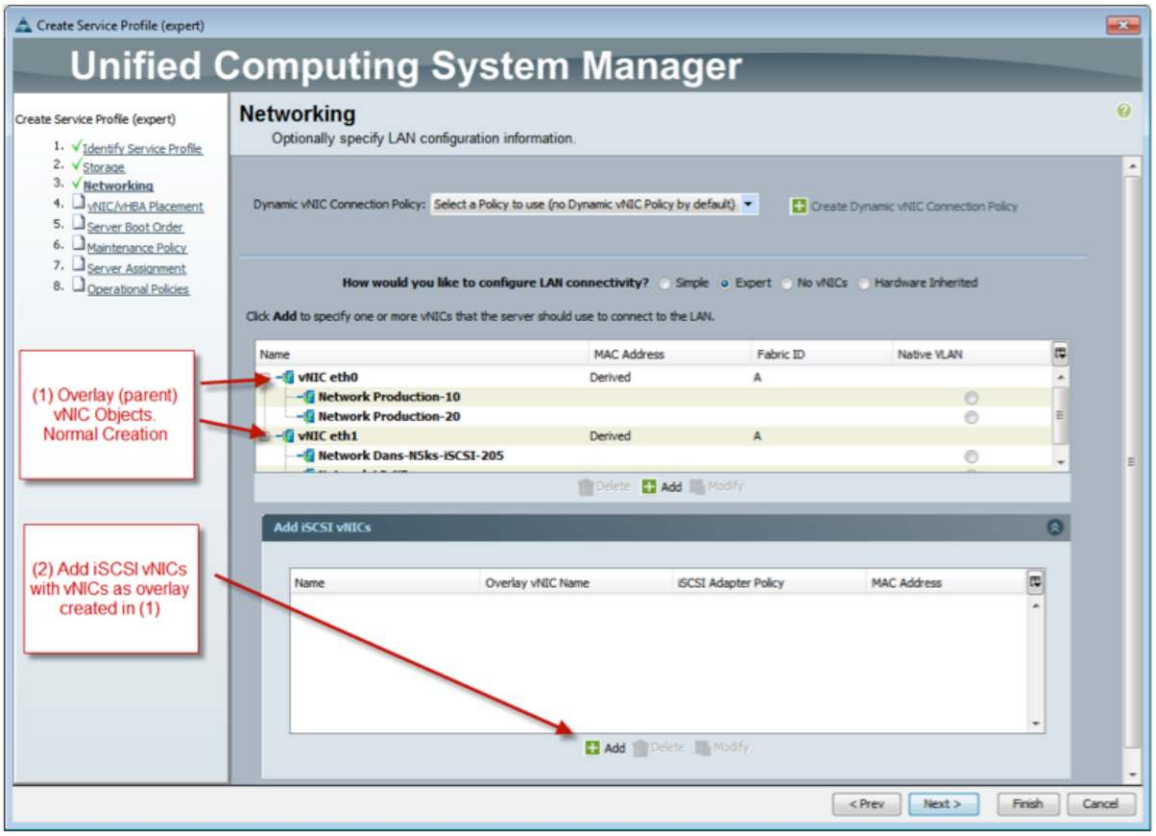
- iSCSI Adapter Policy . If you are using a Broadcom adapter than you will need one policy for install and one for boot or make changes to same policy. This is not necessary with the Cisco VIC
- VLAN Selection (from overlay/parent vNIC VLANs)
- iSCSI MAC Address Assignment

  VIC must be set to “none”

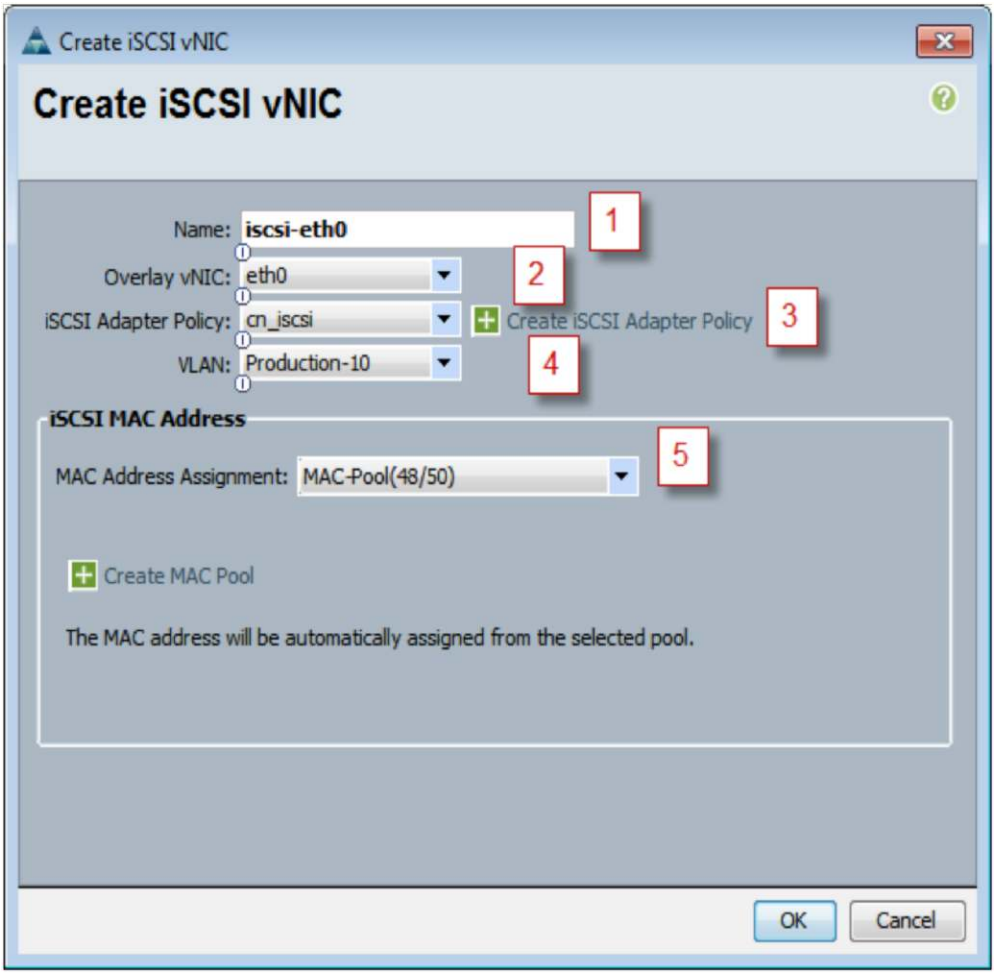
  Broadcom = standard MAC Pool, or user, etc.
- Boot Policy, add iSCSI vNICs

  IQN is manually entered here, these are not pooled until the 2.0(2) release.

Choosing the expert mode for Networking will expose the bottom of the GUI screen and the ability to create the iSCSI vNIC objects.



A detailed review of the iSCSI vNIC is key to understanding this feature.

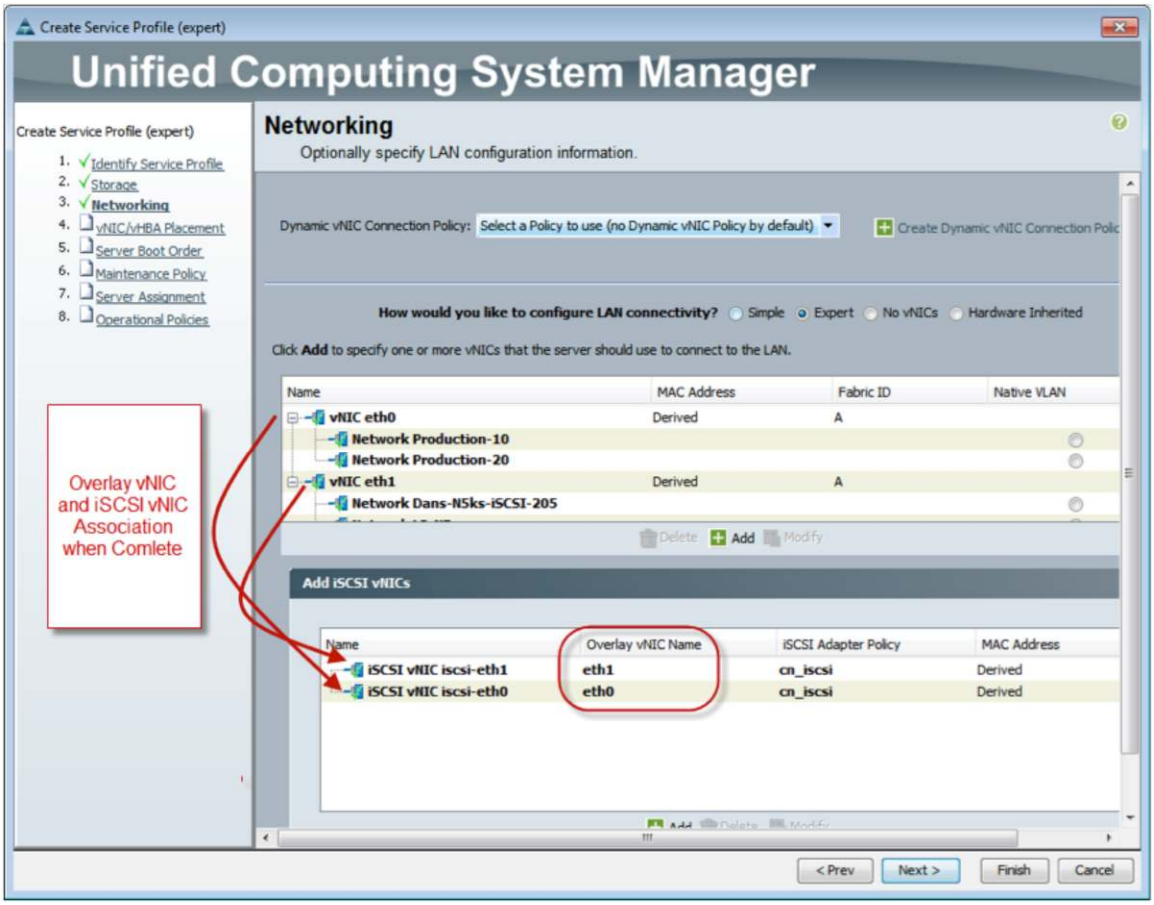


1. Object name, user preference
2. One of the vNICs created in previous step. This is the child object and it inherits some attributes from the parent vNIC you have previously created.
3. Adapter Policy, typically this is created before you create the service profile but optionally can be done here.
4. VLAN, inherited from overlay vNIC, drop down list of available VLANs assigned to overlay vNIC

Note: the Cisco VIC only supports native VLAN boot.

5. **For VIC set to "None"** For Broadcom you can use the Standard MAC Pool

The following shows the relationship between the vNIC and the iSCSI vNICs after you have created all the objects.

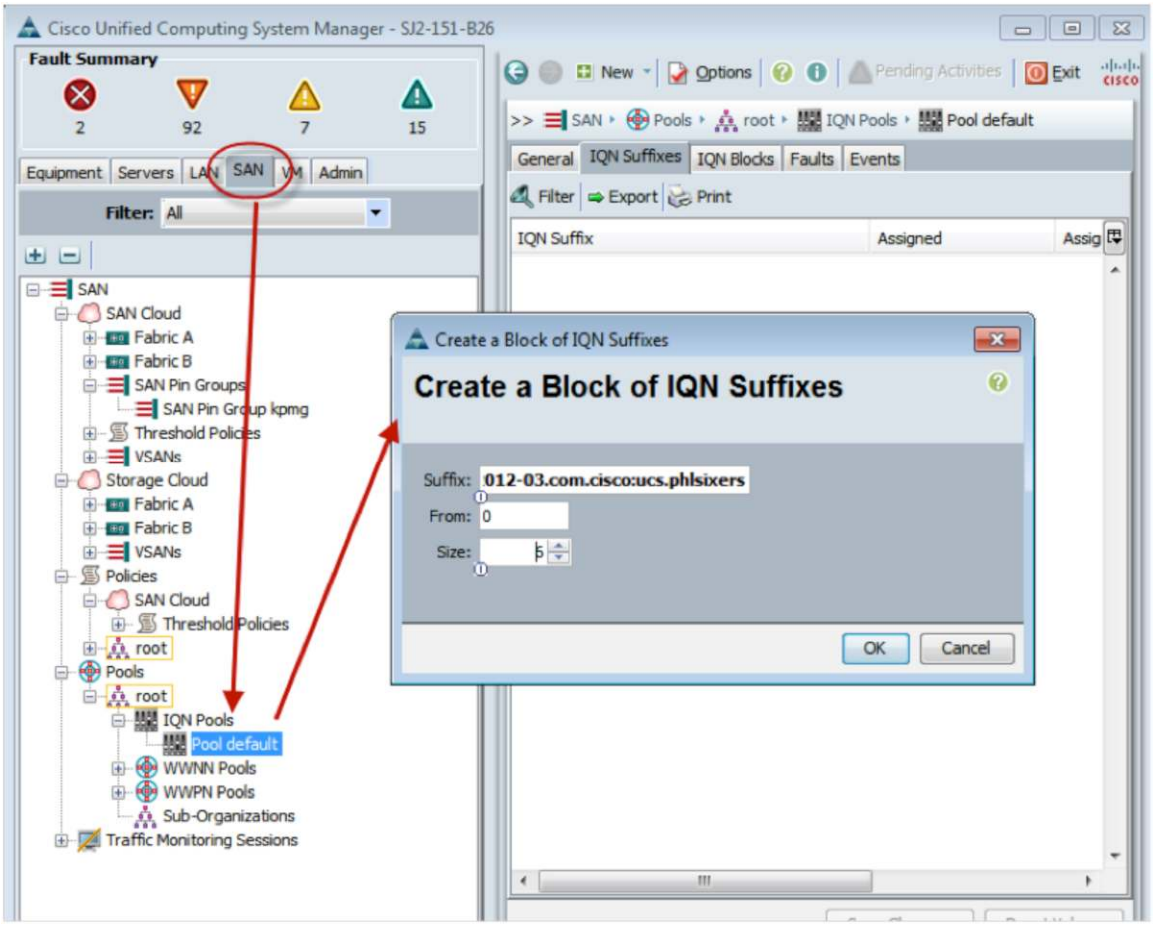


Placement polices is the next step but unlike traditional vNICs there is no ability to configure placement policies for iSCSI vNICs, rather they will just inherit the parent vNIC policy. However always put the primary iSCSI vNIC on the lowest vNIC PCIe device so it is discovered first. This is discussed in the caveats section later in the document as well. The way to handle this is to assign the primary iSCSI vNIC to the lowest numbered vNIC name assuming you will then use a placement policy to ensure this vNIC is discovered first. However if no placement policy is used the first vNIC is usually discovered first.

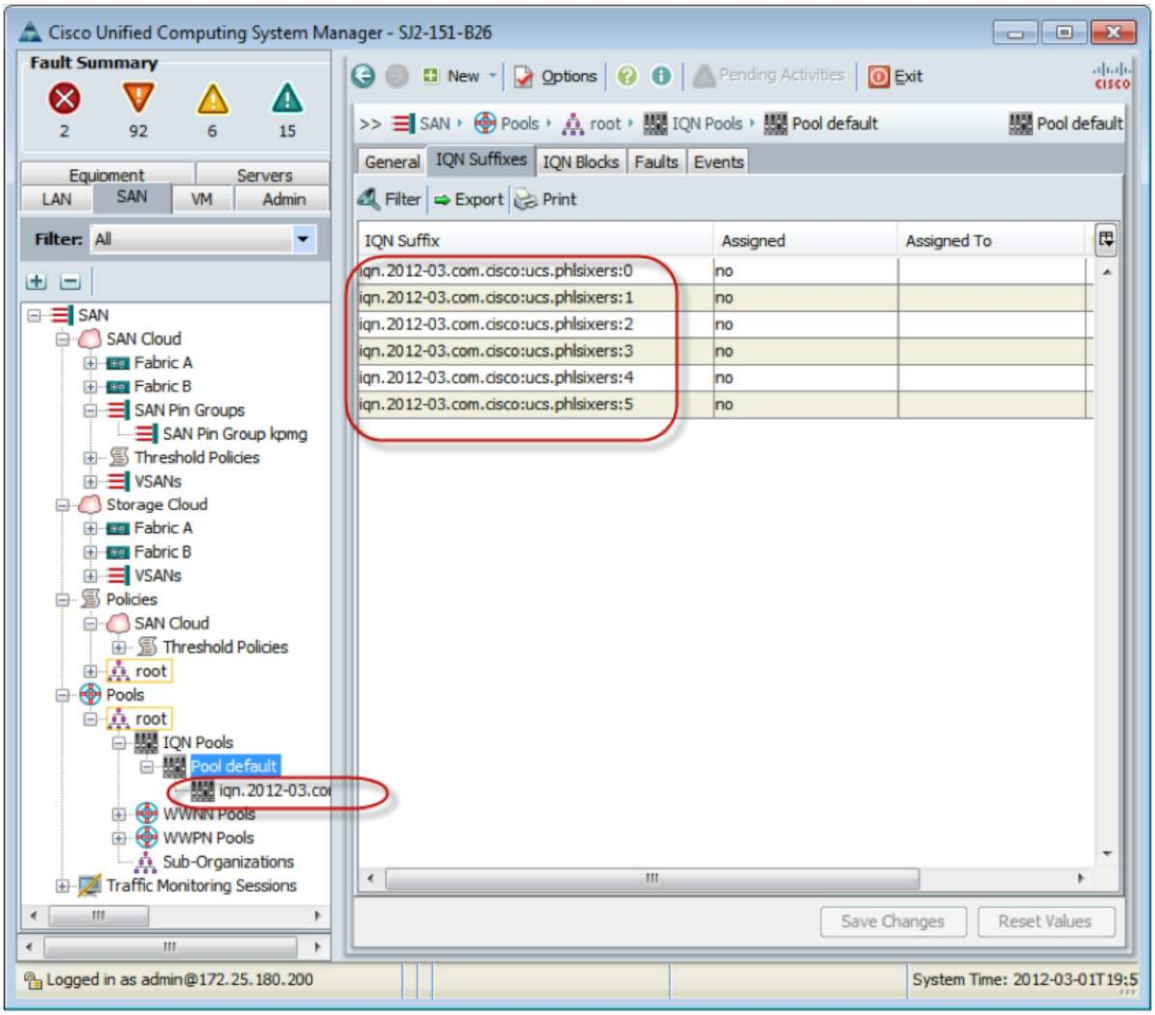
Now we must configure the boot policy and the correct parameters that UCS will use for to properly program the Option ROM to post the iBFT and then boot successfully. Note that in the releases of UCS 2.0(2) and earlier you cannot create a separate iSCSI boot policy “outside” of the service profile. This FC like capability is coming in a future release of UCS.

A Note about Templates and iSCSI Boot Policies: A SP template can be created as it normally would be but you must then create the iSCSI boot policy within the context of the template vs. creating an iSCSI boot policy outside of the template creation. This capability will be introduced in a later release

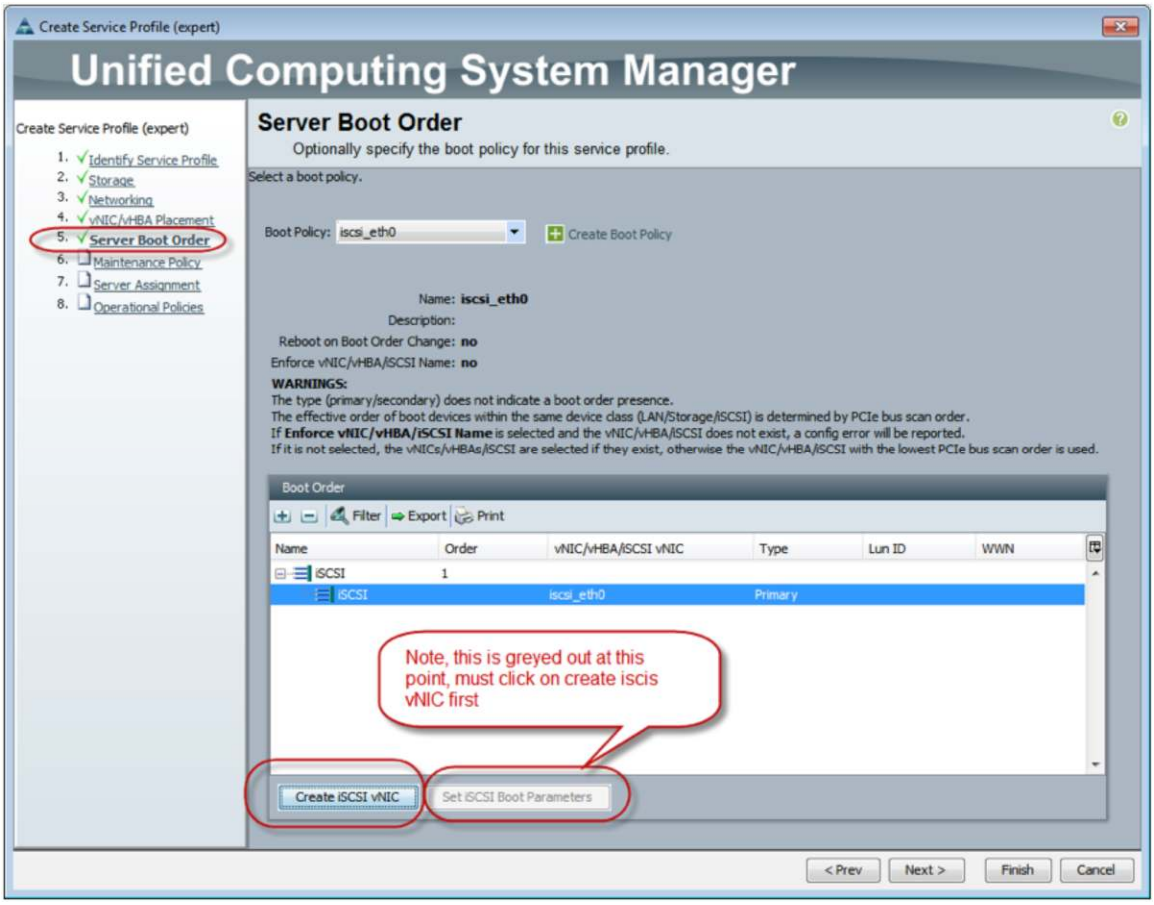
Now you can either manually populate the initiator IQN name in the2.0(1) release or use the new pool feature in the 2.0(2) software. The naming convention enforced to some degree but the user should check with NetApp documentation on specifics of formatting. The screen shot below shows how to configure iQN pools in the 2.0(2) release.



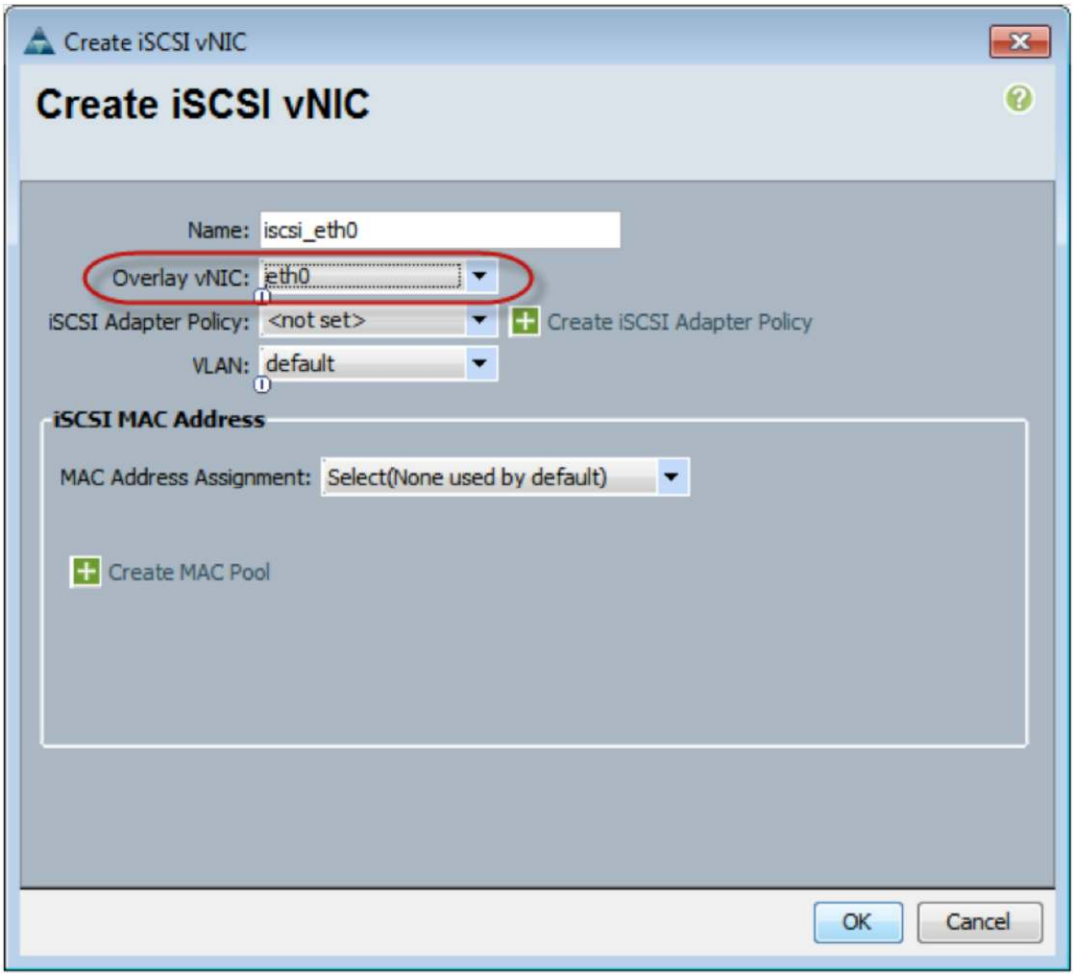
Once you select OK, you will see a block of iQNs created with the numerical suffix appended to your suffix that you used as shown here



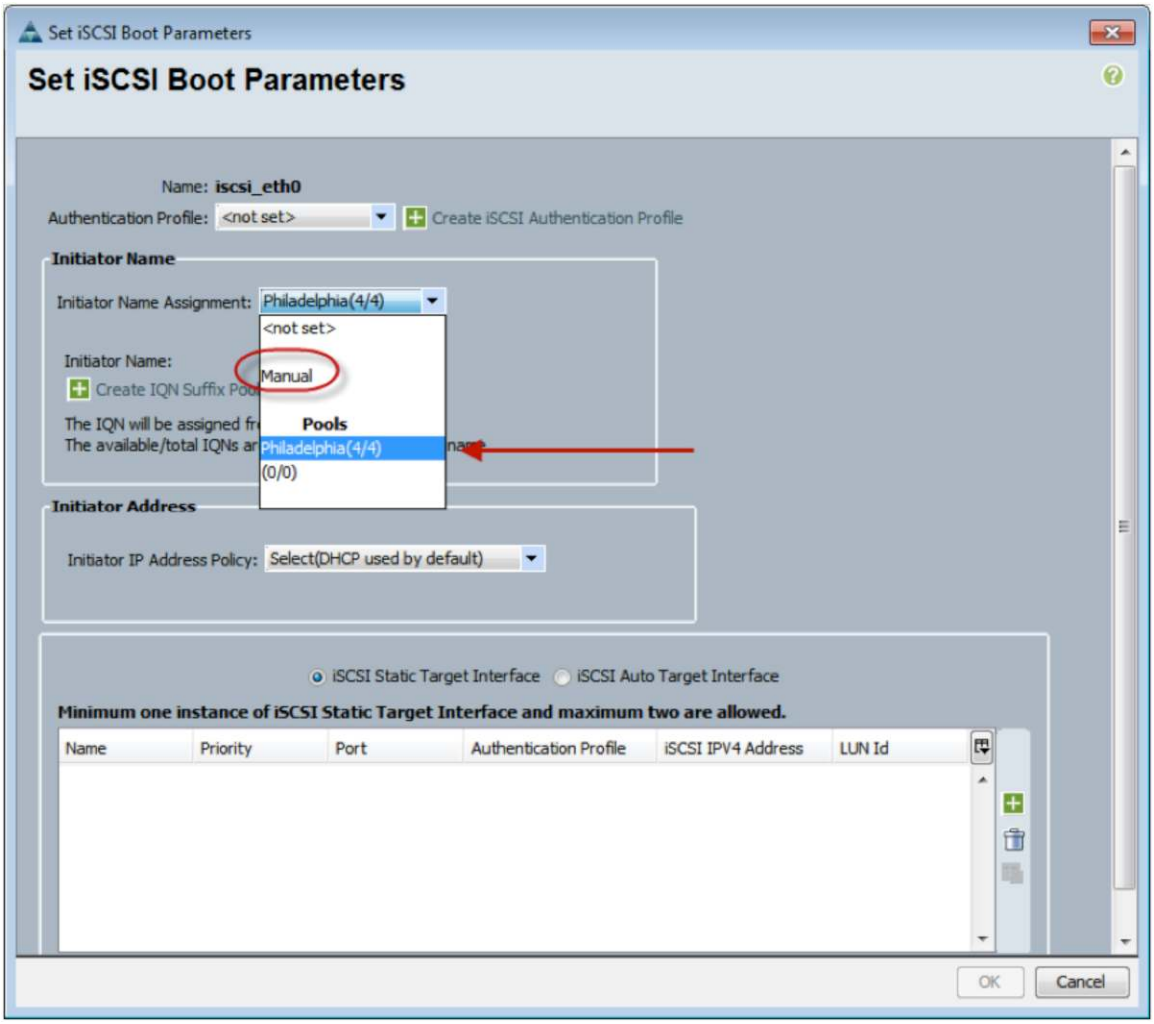
Now when you are modifying the boot policy for the iSCSI target you need to do a few steps. The Modify boot parameters is grayed out until you click on the “create iSCSI vNIC” button at the bottom as shown here.



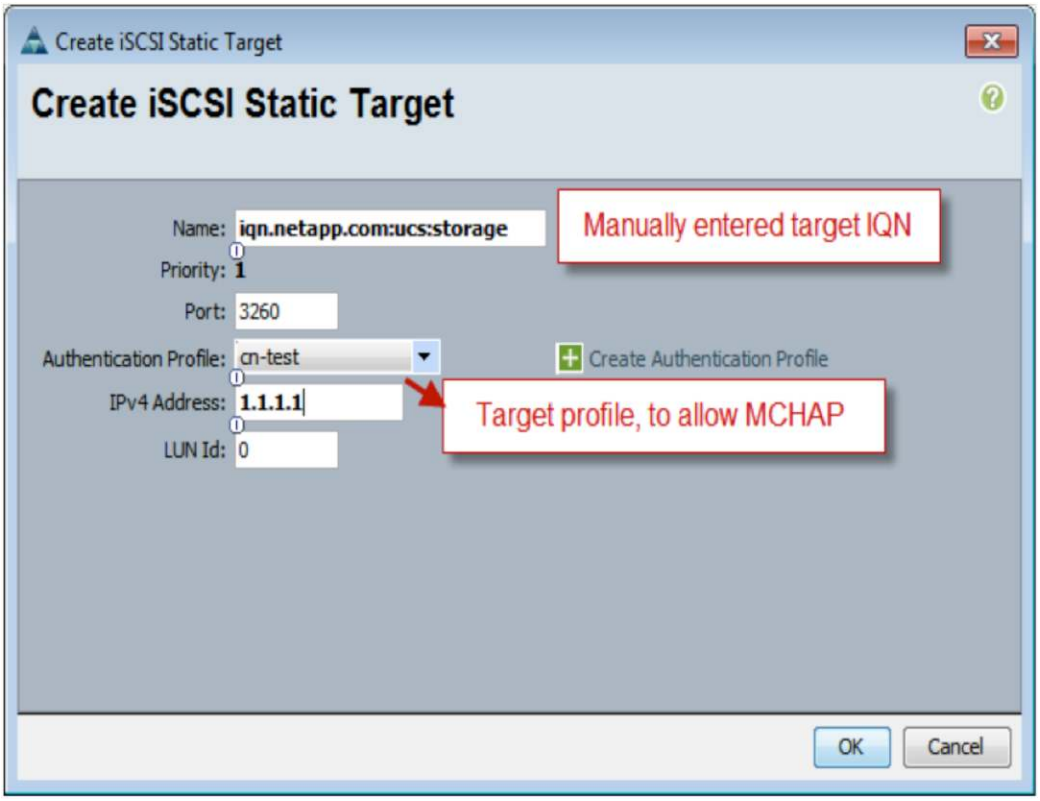
Clicking “Create iSCSI vNIC” really just allows UCSM to choose the overlay vNIC at this step. You need not create a new iSCSI vNIC but this is necessary in case the user wants to change the assignment of iSCSI vNIC to vNIC (overlay). So as shown here you must choose the overly vNIC to use



Now the boot parameters button is active and upon clicking that you are presented with the following screen where you can manually enter an iQN or just use a previously created pool.



Next we will need parameters for the target that the initiator is logging into



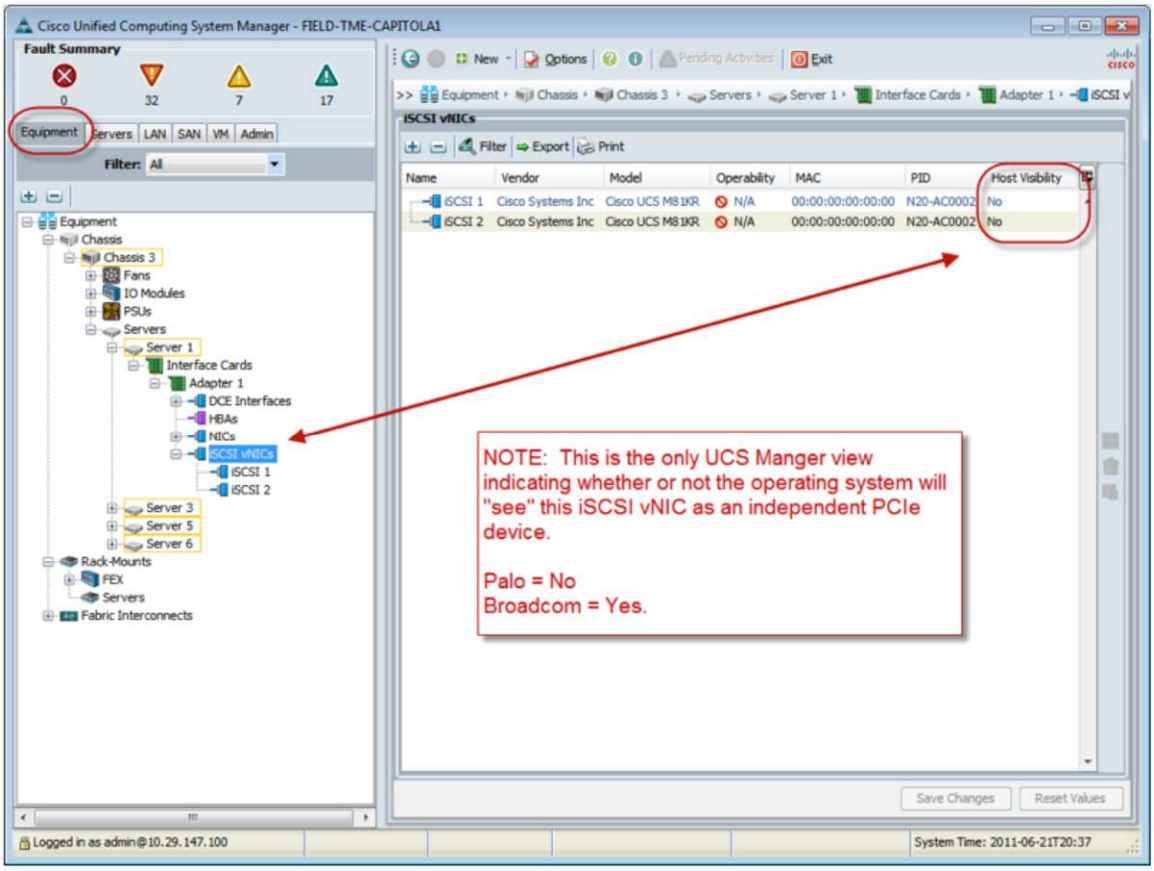
Manually populate the target IQN name which is obtained from the NetApp controller management software

Authentication profile here is for the target, cannot be the same as initiator profile. This allows MCAP

The authentication credentials set on a Target profile are used by the Initiator to verify the challenge string sent to it by that target

The IP target address is manually entered.

When the service profile is complete you will see the iSCSI vNIC objects in the UCSM object tree hierarchy. However with the Cisco VIC the OS does not see these additional interfaces as we discussed in the caveat section. The host visibility field was created to address any confusion.



The diagram below shows a typical iSCSI configuration with a NetApp FAS 3200 series HA pair. The blue lines represent the logical layout of the iSCSI connections. Each UCS blade in this example has 2 vNICs for iSCSI, one for each UCS fabric with failover disabled. The Fabric Interconnects use virtual PortChannels to connect to the upstream top of rack switches, which in turn connect to the rest the LAN. It should be noted that the storage controllers could also be connected directly to the upstream top of rack switches instead of through the LAN cloud. Here, each storage controller is connected to their own pair of top of rack switches with four 1 GbE links. 1 GbE is used to show a common example; two alternatives would be using 10 GbE instead of 1 GbE, or only using a single link per switch instead of a PortChannel. In any of these cases, the initiator is configured with two paths to the target, one on each subnet with both on the iSCSI VLAN.

How many Ethernet links from your storage to switch you use will depend on the bandwidth requirements, whether they are being shared with other applications, and the speed of the storage network interfaces.

Figure 4. Flow Chart for Using Interface Groups on NetApp Controllers

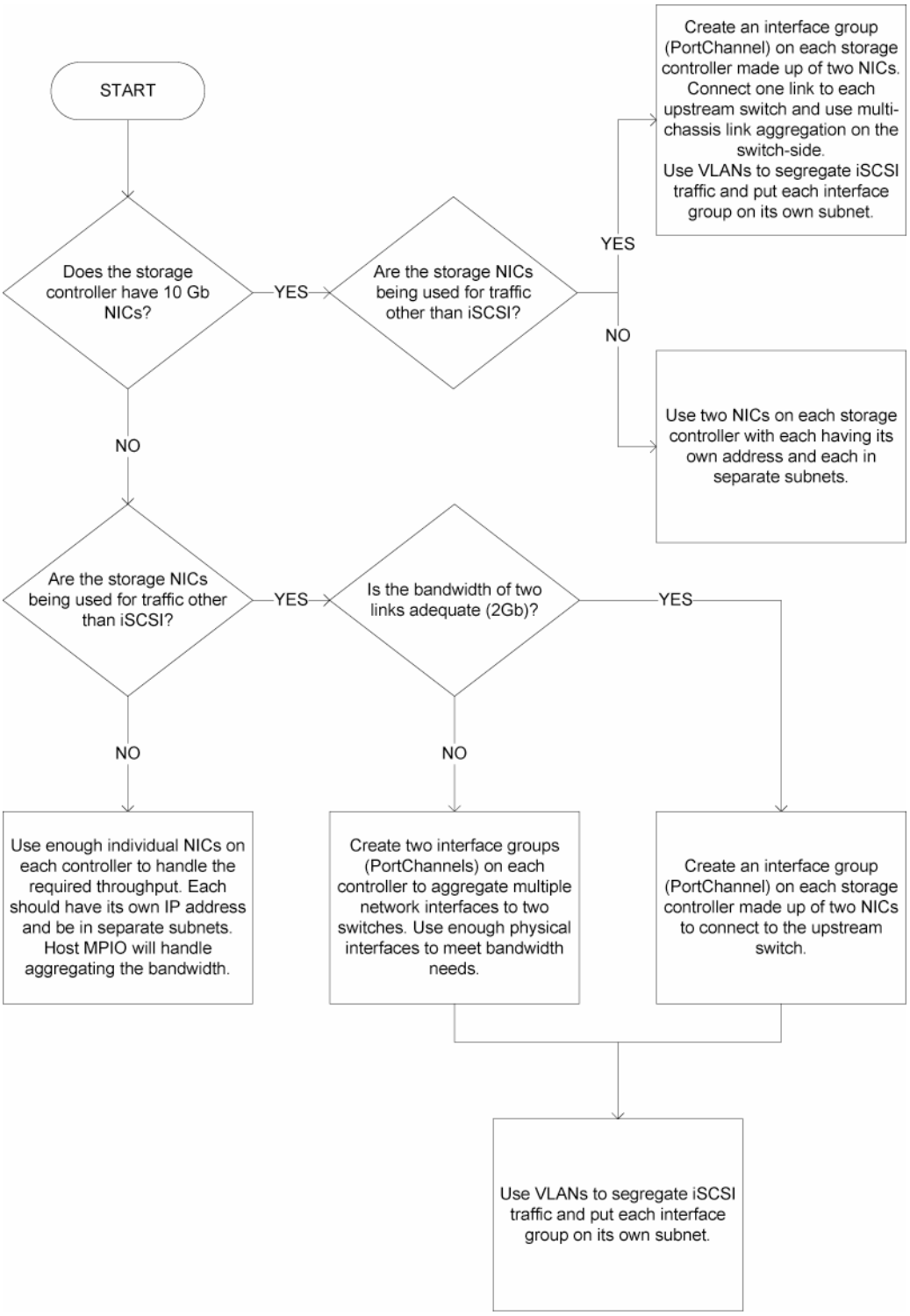
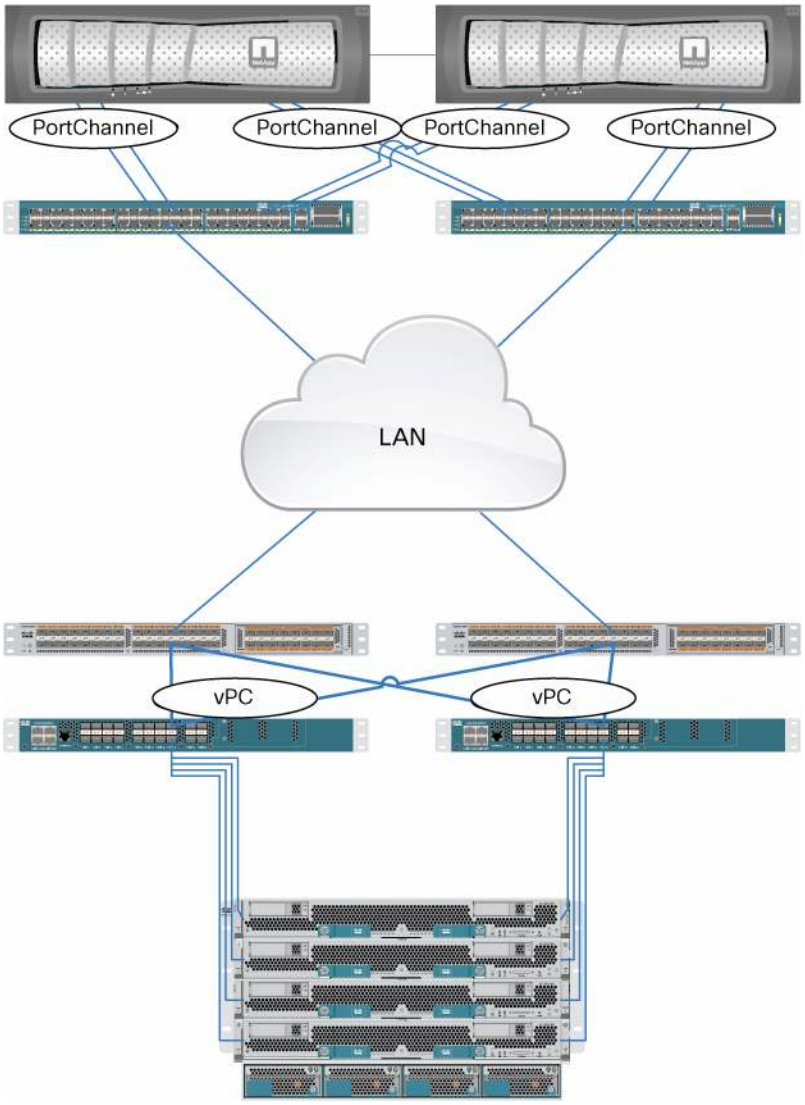


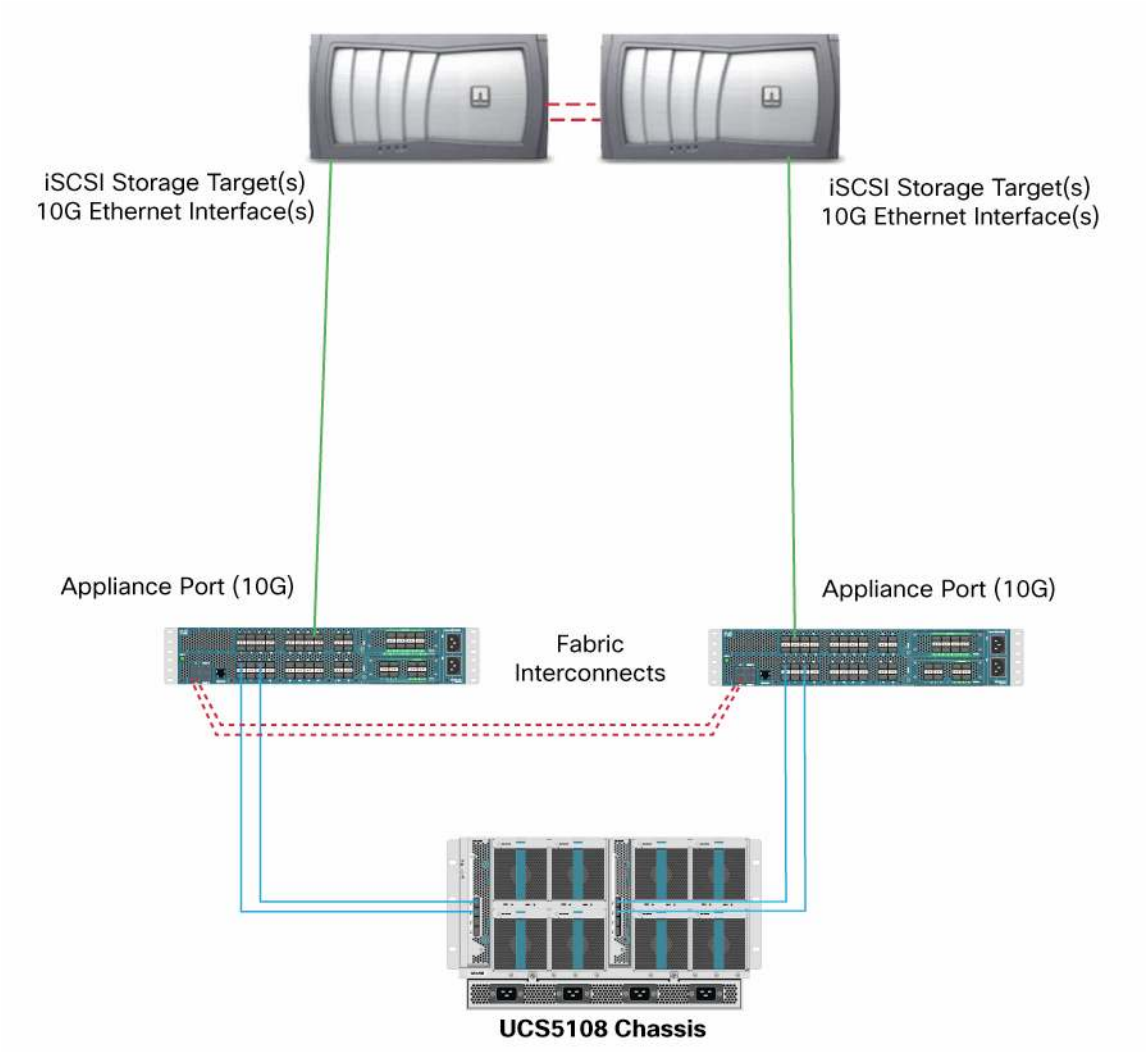
Figure 5. Reference or Typical iSCSI Configuration



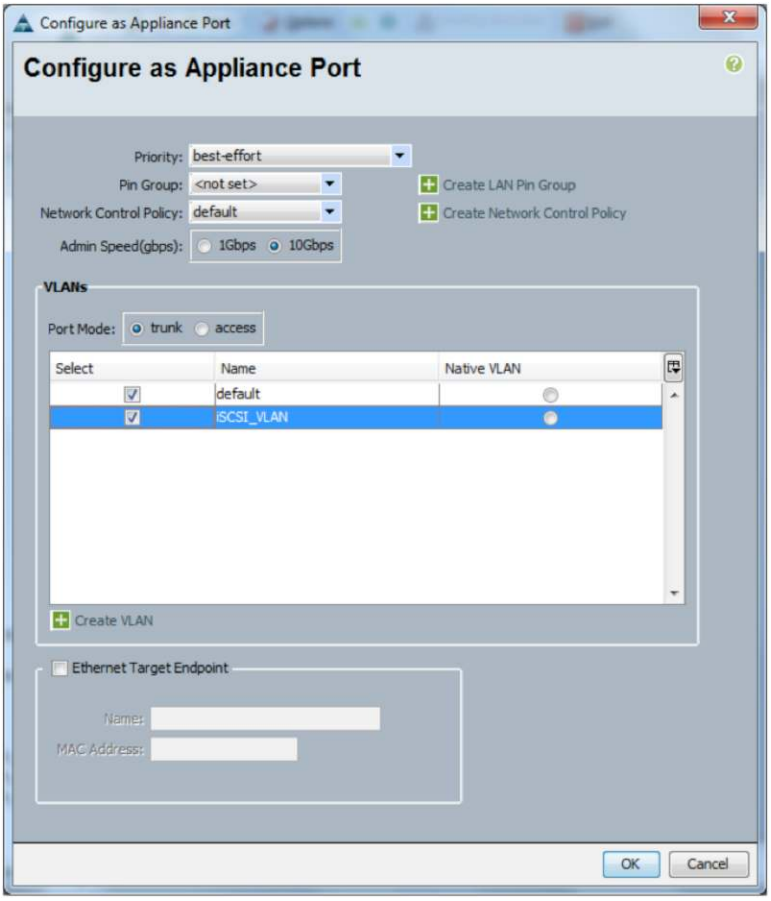
iSCSI SAN Direct Connect—Appliance Port in End Host made

With Cisco UCS firmware version 1.4 came the ability to directly connect storage to the Fabric Interconnect. This removes the need for an intermediate network and can be useful in a pod-like scenario. There are specific requirements and considerations for directly connected IP storage that must be observed.

Figure 6. Direct Connect iSCSI Topology with UCS



When using an Appliance port to connect iSCSI storage, VLANs must be configured on each Appliance port. Unless the storage controller is only going to be serving iSCSI on the network interfaces connected to the UCS, the Appliance ports should be configured as trunk ports with all desired VLANS available. This allows the storage controller's network adapter to serve iSCSI traffic on one VLAN and other data on separate VLANs. The default VLAN, VLAN 1, is also the default native VLAN.



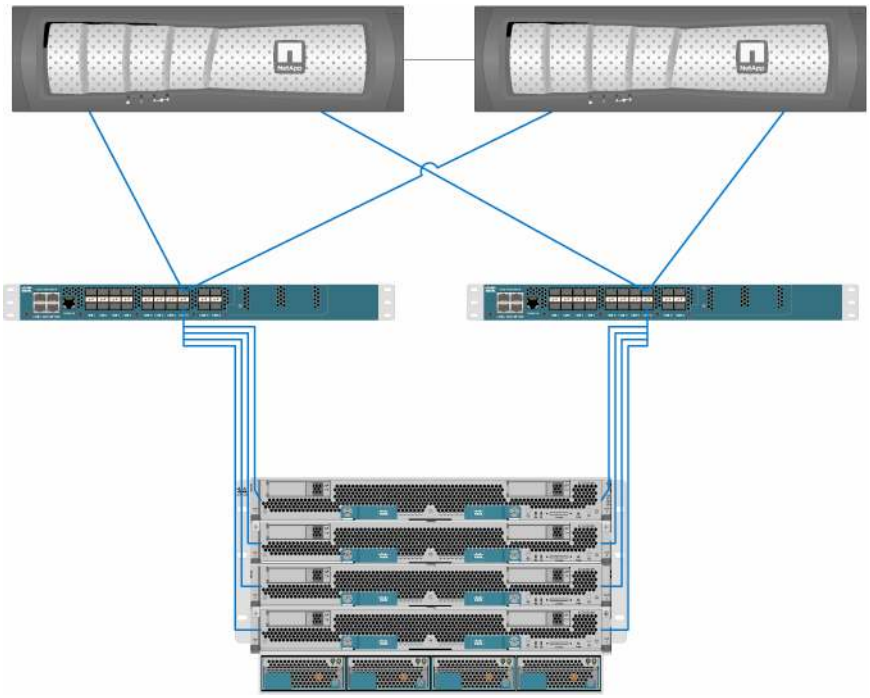
On the storage controller, the network ports connected to the Fabric Interconnect need to have VLAN interfaces configured for each non-native VLAN including the one used for iSCSI. An example of this is shown in the example below.

```
FAS3240-A> vlan create e1a 100
Ethernet e1a: Link being reconfigured.
vlan: e1a-100 has been created
Ethernet e1a: Link up.


FAS3240-A> vlan create e1b 100
Ethernet e1b: Link being reconfigured.
vlan: e1b-100 has been created
Ethernet e1b: Link up.


FAS3240-A> ifconfig e1a-100 192.168.101.105 netmask 255.255.255.0 mtusize 9000
partner e1a-100


FAS3240-A> ifconfig e1b-100 192.168.102.105 netmask 255.255.255.0 mtusize 9000
partner e1b-100


FAS3240-A> ifconfig –a
e1a: flags=0x80f0c867<BROADCAST,RUNNING,MULTICAST,TCPCKSUM,VLAN> mtu 9000
        ether 00:c0:dd:11:40:2c (auto-10g_twinax-fd-up) flowcontrol full
e1b: flags=0x80f0c867<BROADCAST,RUNNING,MULTICAST,TCPCKSUM,VLAN> mtu 9000
        ether 00:c0:dd:11:40:2e (auto-10g_twinax-fd-up) flowcontrol full
```

```
e1a-100: flags=0xb4c867<UP,BROADCAST,RUNNING,MULTICAST,TCPCKSUM> mtu 9000
        inet 192.168.101.105 netmask 0xffffff00 broadcast 192.168.101.255
        ether 00:c0:dd:11:40:2c (auto-10g_twinax-fd-up) flowcontrol full
e1b-100: flags=0xb4c867<UP,BROADCAST,RUNNING,MULTICAST,TCPCKSUM> mtu 9000
        inet 192.168.102.105 netmask 0xffffff00 broadcast 192.168.102.255
        ether 00:c0:dd:11:40:2e (auto-10g_twinax-fd-up) flowcontrol full
```

In the image below a NetApp FAS 3200 series HA pair is connected to the Fabric Interconnect with one 10 GbE link to each fabric. The blue lines represent the logical layout of the iSCSI connections. Each UCS blade has two vNICs with the iSCSI VLAN set as native. One vNIC goes to Fabric A and one to Fabric B; failover is disabled on both.

Figure 7. NetApp Direct Connect iSCSI



vNIC fabric failover is not needed and should not be used with iSCSI because the host OS will use MPIO to determine which paths to storage are active and handle any path failures. The best practices is to not use FF at all with iSCSI traffic in UCS unless a host multi-pathing driver cannot be used for some reason.

NAS and UCS Appliance Ports

When using NFS or CIFS with Appliance ports it is essential to have upstream L2 Ethernet switching to allow traffic to flow in certain failure scenarios. This is not required for iSCSI only traffic given the use of the host MPIO stack which manages path failures and recovery end to end.

Appliance ports were introduced in UCS 1.4 and were designed to allow a direct connection between the UCS Fabric Interconnect and the NetApp storage controller. An appliance port is essentially a server port under the covers that also does MAC learning. Given the appliance port is a server port the same policies apply to them in terms of uplink or border ports. They have border ports associated with them and there is a network control policy that determines what do to in the event the last available border or uplink port goes down.

Appliance ports like server ports have an uplink or border port assigned either via static or dynamic pinning. By default the loss of last uplink port will result in the appliance port being taken down. One can change the network control policy under the appliance cloud to have this be a warning only. For NFS configurations the default should be used which will down the appliance port if the last uplink port is lost. This ensure more deterministic failover in the topology.

UCS Fabric Interconnects cannot run at vPC peers so from a UCS FI standpoint there is a active/passive data path for NFS traffic to the same IP address. You can of course do active/active I/O from both FIs to different backend NetApp volumes and controllers.

Once should utilize the Interface Group (ifgrp) feature of managing Ethernet ports on the NetApp controller. This feature was called Virtual Interface (VIF) in earlier releases of Data ONTAP. There are single mode ifgrps and multimode ifgrps. Single mode are active-standby while multi-mode can be static or dynamic supporting LACP port channels. A detailed description of this feature can be found in the Data ONTAP Network Management Guide

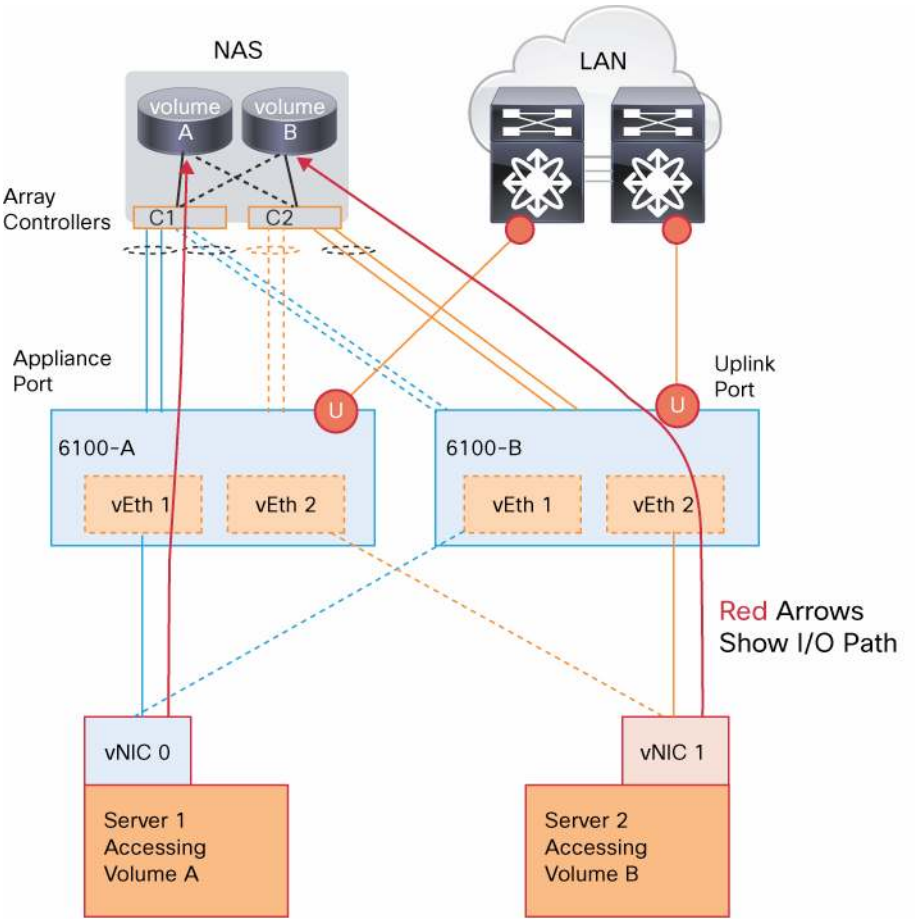
Figure shows the best practice topology for using UCS appliance ports with NFS/CIFS traffic. This is the steady state or reference topology from which we will examine various failure cases in subsequent sections.

There some key principals one must understand when using appliance ports in such a configuration.

- The 2nd level VIF provides further grouping of multiple multimode VIFs and it provides a standby path in the event the primary multimode VIF fails. The primary multimode VIF will be constructed of a LACP port channel thus this complete failure is unlikely.
- Failure processing within UCS and within the NetApp controllers are not coordinated. This means what one tier might see as a failure the other will not. This is the fundamental reason for the upstream L2 switching infrastructure we will see in subsequent diagrams
- NetApp VIF failovers are based on link state only. This means that if something happens within the UCS system and it does not translate to the link down being sent to the NetApp controller than no failure processing will occur on the NetApp controller. Conversely a NetApp interface group port migration may not translate to any failures on the UCS tier which would trigger a vNIC migration.

Now let us walk through a few failure cases and diagrams where these principals will be evident. Here is the steady state traffic flow between the UCS and NetApp systems. The red arrows show the traffic path from the UCS vNICs to the NetApp exported volumes.

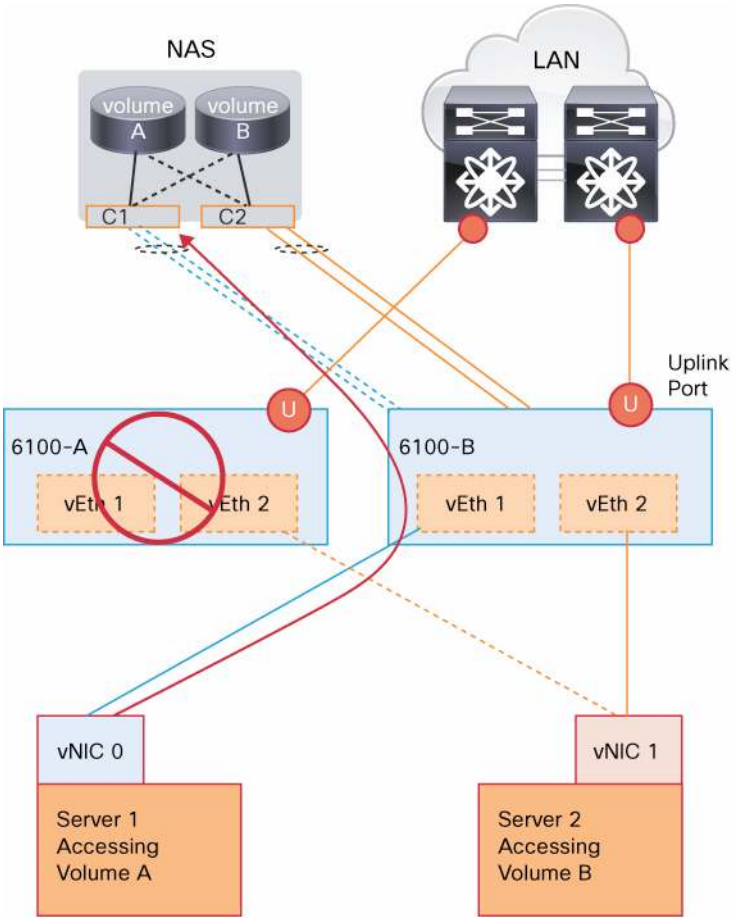
Figure 8. Steady State Direct Connect Appliance Ports with NFS



Failure Scenario 1: UCS FI Fails (Simplest failure case)

In this case the UCS FI on the left fails. UCS will see this as a situation where the vNIC is moved from the A side to the B side by the Fabric Failover feature. No issues here as the traffic fails over to the standby VIF since NetApp would see the FI failure as a “link down” event on the primary interface and thus expect to start seeing traffic on the standby and accepting it.

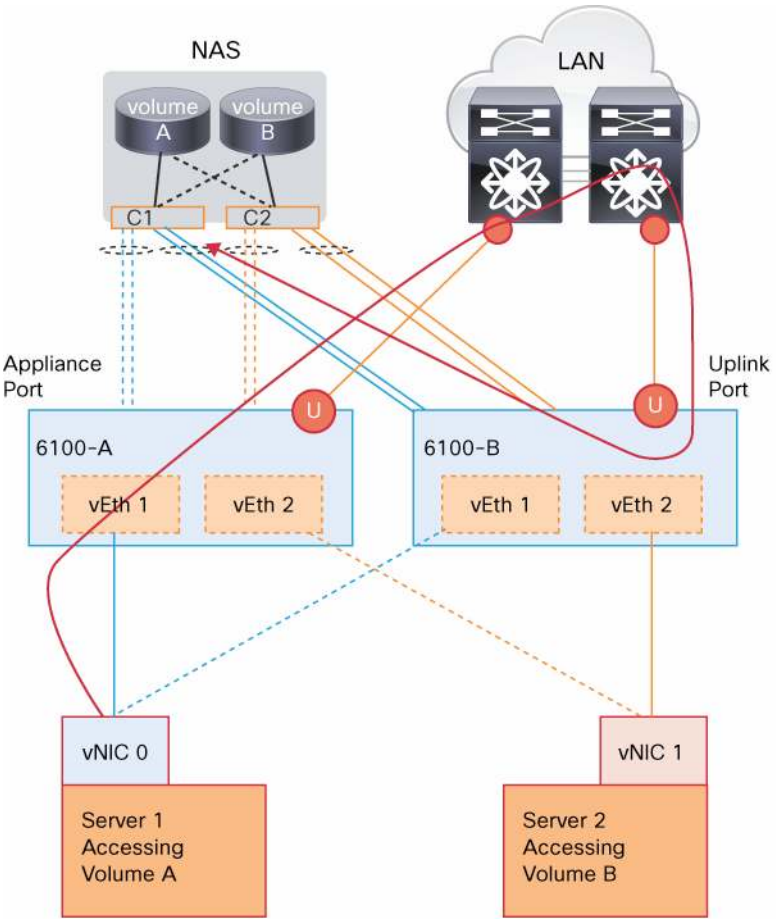
Figure 9. Failure Case 1: FI is Lost



Recovery From Failure Scenario 1: FI is repaired, rebooted

When the FI on the left comes back on-line the traffic will not automatically go back to the steady state unless one uses the NetApp “FAVOR” option when configuring the VIF interfaces. This is considered a best practice and should always be done. If it is not done then the traffic will be flowing through the upstream L2 infrastructure as shown in the next figure. Recall the Appliance ports do MAC learning and have associated uplink ports which is how this traffic pattern is able to flow.

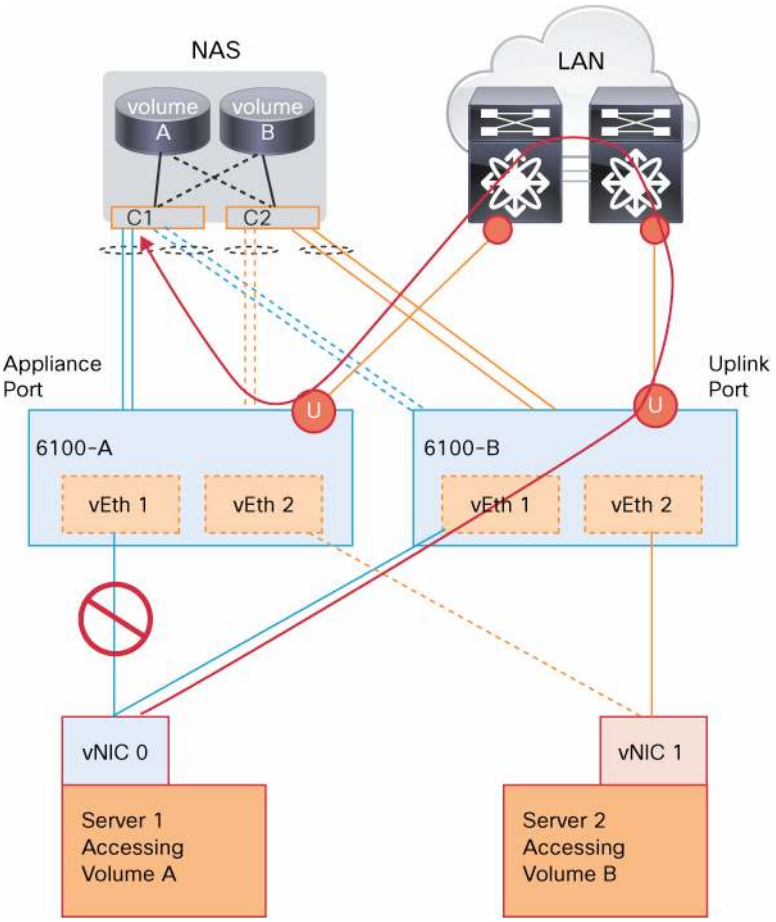
Figure 10. Data Path Upon Recovery of FI



Failure Scenario 2: UCS Failure of Chassis IOM or all uplinks to FI.

This is the most complex and confusing failure condition. UCS will see this as an event upon which to move the vNIC to Fabric B however this will NOT appear as link down even from the NetApp perspective. Thus in this scenario to allow traffic to continue to flow an upstream L2 device must be present as shown in the following diagram. Traffic will not be able to reach the NetApp volumes without the use of the upstream L2 network.

Figure 11. Failure Case 2: Lost of IOM



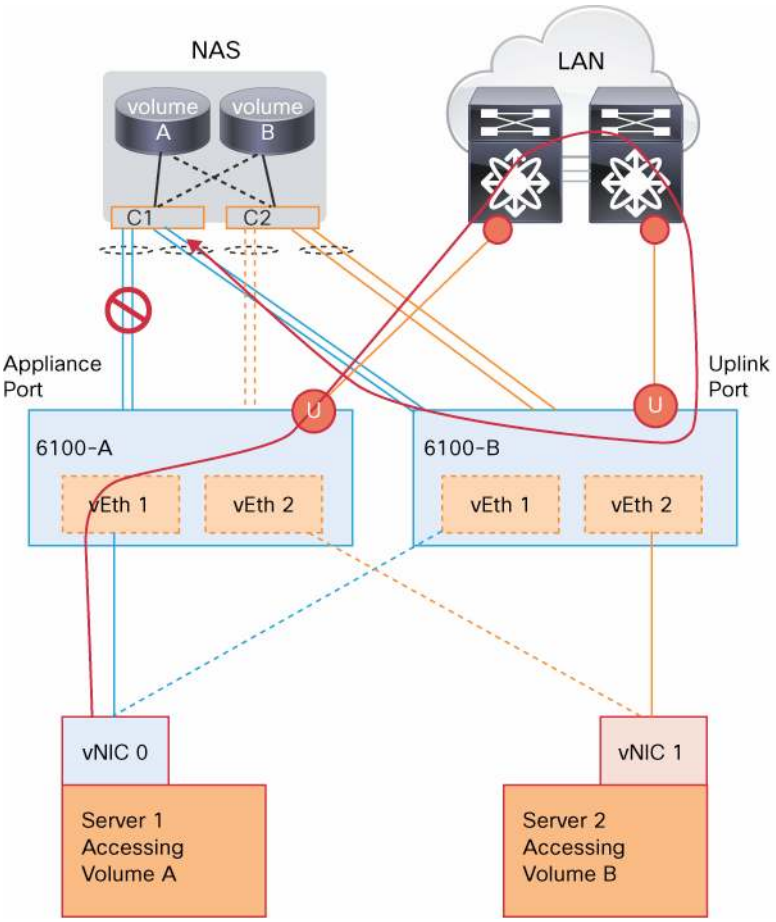
A natural question to ask at this point is the following: “If you need to have upstream L2 devices anyway, then why use Appliance ports in the first place?” The answer is that the upstream is only used in this very rare failure condition and the remainder of the time your storage traffic is going directly from UCS to the array.

Now when the IOM or links are restored in the diagram above traffic flows normally as in the steady state since the NetApp controller never made any interface primary changes.

Failure Scenario 3: Underlying Multi-mode VIF Failure (Appliance port)

This scenario shows a case where the appliance port link or port channel fails. NetApp would see this as al link down event and now expect to see traffic on the standby link. However UCS does not see this as a link down event for the vNIC and would thus keep the vNIC assigned to Fabric A. The uplink port on the FI would enable the traffic to go out the FI to the upstream L2 network back to Fabric B and then to the standby VIF on the NetApp. This is shown below.

Figure 12. Loss of Appliance Port Links

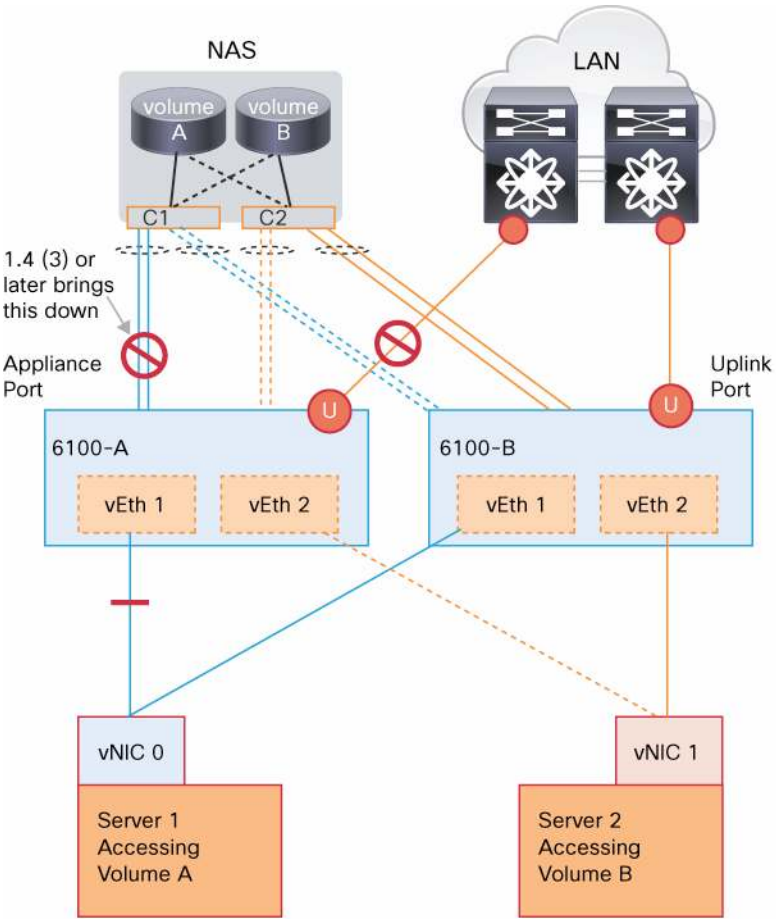


Recovery from this is similar to what has been discussed before. Ensure to use the FAVOR option in the VIF configuration to allow steady state to return.

Failure Scenario 4: Last Uplink on UCS FI fails

As of the UCS 1.4(3) releases this problem is identical the failure of a FI itself which we saw in Failure Scenario 1. The default policy is to bring down the appliance ports if the last uplink is down.

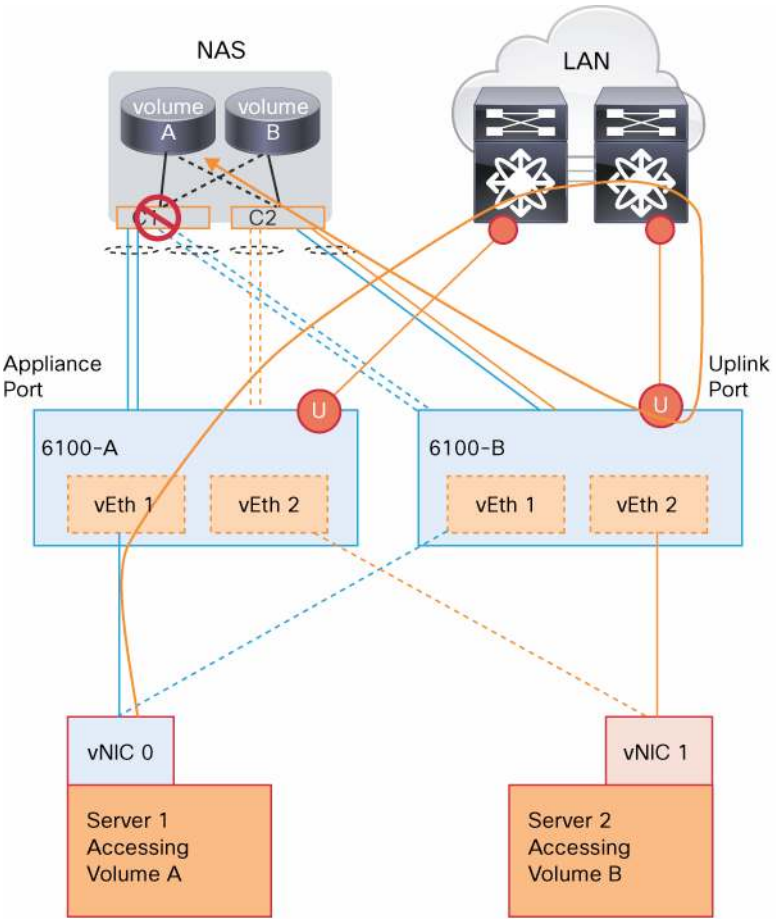
Figure 13. Loss of Last Uplink Port on the FI



Failure Scenario 5: NetApp Controller Failure

This case examines what happens when the one of the two NetApp controllers fail. A NetApp Controller Failover (CFO) event will take place assigning ownership of the volumes to the remaining controller. However from a UCS perspective nothing happened so the vNIC will stay on FI-A. Again the only way traffic can continue to flow is from the upstream L2 network since the failures are not coordinated. The following diagram shows this.

Figure 14. Loss of NetApp Controller



Recovery from this event is a manual process as a user must initiate the controller giveback command to return everything to the steady state.

Here is a summary of the key principals when using Appliance ports for NFS traffic with NetApp arrays:

- Must have L2 Ethernet upstream for failure cases, one cannot just deploy a Storage array and UCS hardware.
- Always enable IP storage VLAN on the FI uplinks for data traffic failover scenarios
- Provision the uplink bandwidth with consideration of different failure cases discussed.
- Minimum of two uplink ports per FI required. Port Channel ideal configuration.
- Do not deploy 1 Gig uplinks with IP storage, lest performance suddenly decrease during a failure event and the user community is ok with this scenario.
- At a minimum pair of 10 Gig links per multimode VIF is recommended.

Conclusion

There are a wide variety of connecting external storage to the UCS system. The long standing industry best practices for FC, FCoE, iSCSI and NFS/CIFS apply with no changes for UCS unless the user wishes to deploy a

"direct connect" topology with UCS and the storage. Using both NetApp and UCS in a combined best practices manner results in reliable, flexible and scalable compute and storage infrastructures.

References

Cisco UCS Manager GUI Configuration Guides
http://www.cisco.com/en/US/partner/products/ps10281/products_installation_and_configuration_guides_list.html

Cisco UCS B-series OS Installation Guides
http://www.cisco.com/en/US/products/ps10280/products_installation_and_configuration_guides_list.html

NetApp Interoperability Matrix (Requires NOW account)
http://now.netapp.com/matrix/mtx/login.do

NetApp Fibre Channel and iSCSI Configuration Guide (Requires NOW account)
https://now.netapp.com/knowledge/docs/san/fcp_iscsi_config/config_guide_80/frameset.html

NetApp Data ONTAP Documentation (Requires NOW account)
https://now.netapp.com/NOW/knowledge/docs/ontap/ontap_index.shtml

Appendix I: FC Adapter Polices Explained

Scope

This document will explain the exposed settings for the Unified Computing System Manager (UCSM) FC adapter policy settings as well as outline those parameters which are hardcoded and not configurable by the user. The paper will also outline which settings are applicable to which adapter model.

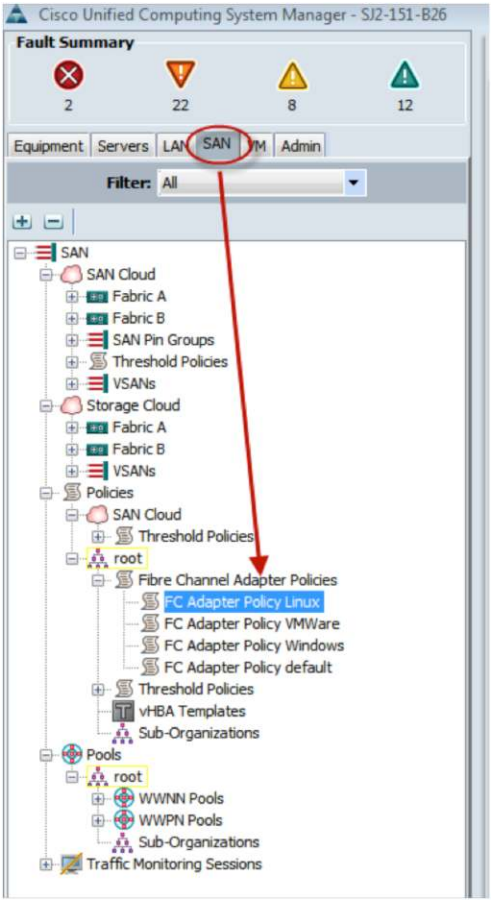
It is important to note that the default values shown in this document and UCSM have been carefully chosen by Cisco engineering and should only be changed after careful consideration following the guidance for that particular parameter.

Purpose of the FC Adapter Policies

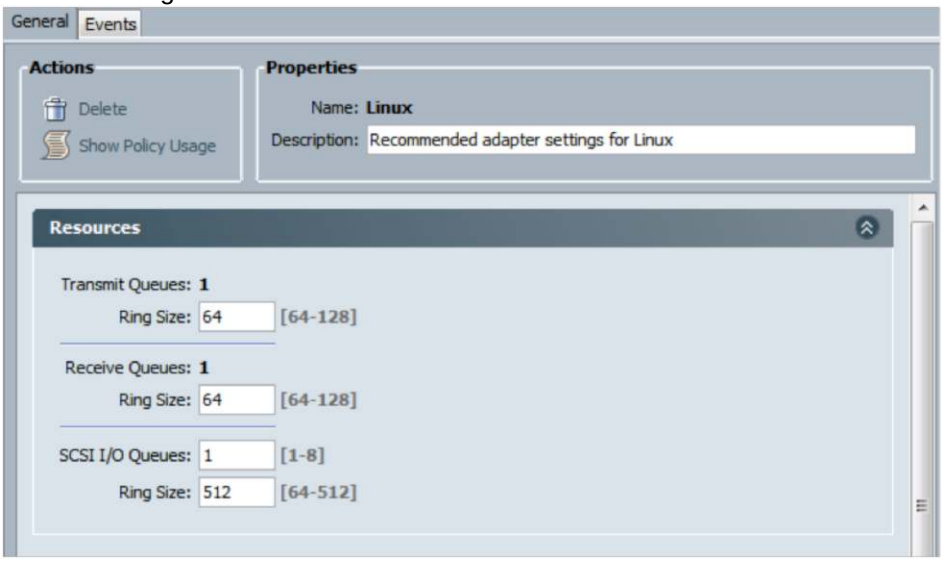
UCS Manager provides the ability to create specific adapter policies for different combinations of adapters, operating systems and storage arrays. The policies should viewed as a way of ensuring the different error handing parameters are configured correctly for the previous mentioned combinations as the main objective the policy. There are very few performance based parameters exposed in the FC adapter polices as of the 2.0(1q) release of UCS.

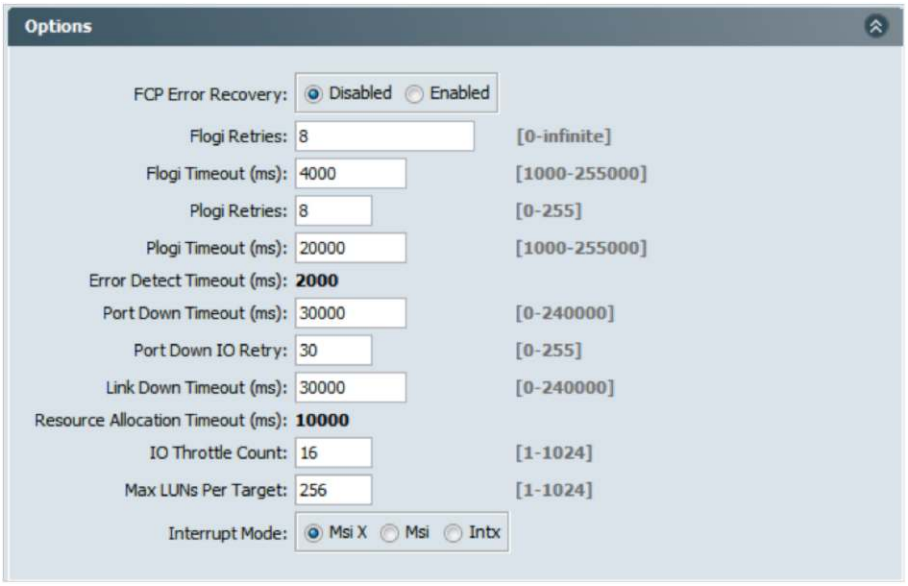
Finding and Creating New FC Adapter Policies

The operational steps to view and create adapter policies are shown below.



Default Settings





Exposed Parameters Detailed Review

Parameter Name	Transmit Queues
Description	Number of transmit queue resources to allocate
Valid Adapter Models and Operating Systems	Cisco Virtual Interface Card (VIC), All Supported Operating Systems
Default Value	1
Valid Range of Values	1
Recommendations on Changing Default Values	No need to change

Parameter Name	Transmit Queue Ring Size
Description	Number of descriptors in each transmit queue. This has to do with Extended Link Services (ELS) and Common Transport (CT) FC frames for generic services. This will not affect performance at all.
Valid Adapter Models and Operating Systems	Cisco Virtual Interface Card (VIC), All Supported Operating Systems
Default Value	64
Valid Range of Values	64 - 128
Recommendations on Changing Default Values	No need to change.

Parameter Name	Receive Queues
Description	Number of receive queue resources to allocate
Valid Adapter Models and Operating Systems	Cisco Virtual Interface Card (VIC), All Supported Operating Systems
Default Value	1
Valid Range of Values	1
Recommendations on Changing Default Values	No need to change.

Parameter Name	Receive Queue Ring Size
Description	Number of descriptors in each receive queue. This has to do with Extended Link Services (ELS) and Common Transport (CT) FC frames for generic services. This will not affect performance at all.

Valid Adapter Models and Operating Systems	Cisco Virtual Interface Card (VIC), All Supported Operating Systems
Default Value	64
Valid Range of Values	64 - 128
Recommendations on Changing Default Values	No need to change

Parameter Name	SCSI I/O Queues
Description	Number of SCSI I/O queue resources to allocate.
Valid Adapter Models and Operating Systems	Cisco Virtual Interface Card (VIC), All Supported Operating Systems
Default Value	1
Valid Range of Values	1 (GUI is misleading showing this can go to 8)
Recommendations on Changing Default Values	Do not change

Parameter Name	SCSI I/O Queue Ring Size
Description	Number of descriptors in each SCSI I/O queue. This value can affect performance
Valid Adapter Models and Operating Systems	Cisco Virtual Interface Card (VIC), All Supported Operating Systems
Default Value	512
Valid Range of Values	64-512
Recommendations on Changing Default Values	This value is already maximized by default thus yielding the optimal performance.

Parameter Name	FCP Error Recovery
Description	Enables or disables the “FC-TAPE” protocol for sequence level error recovery with tape devices. This enables or disables the Read Exchange Concise (REC) and Sequence Retransmission Request (SRR) functions on the VIC firmware.
Valid Adapter Models and Operating Systems	Cisco Virtual Interface Card (VIC), All Supported Operating Systems
Default Value	Disabled
Valid Range of Values	Enabled or Disabled
Recommendations on Changing Default Values	Change to Enabled when connecting to tape drive libraries

Parameter Name	Flogi Retries
Description	Number of times the Flogi is retried before operation is aborted
Valid Adapter Models and Operating Systems	Cisco Virtual Interface Card (VIC), M71KR-E/Q, M72KR-E/Q on all supported operating systems
Default Value	8
Valid Range of Values	Any, -1 means infinite number of retries
Recommendations on Changing Default Values	Check with the storage array vendor’s documentation on a recommended value for this parameter.

Parameter Name	Flogi Timeout (ms)
Description	Number of milliseconds to wait before timing out of the FLOGI exchange
Valid Adapter Models and Operating Systems	Cisco Virtual Interface Card (VIC), M71KR-E/Q, M72KR-E/Q on all supported operating systems
Default Value	4000
Valid Range of Values	1000 - 255000
Recommendations on Changing Default Values	Check with the storage array vendor’s documentation on a recommended value for this parameter.

Parameter Name	Plogi Retries
Description	Number of times the Plogi is retried before operation is aborted
Valid Adapter Models and Operating Systems	Cisco Virtual Interface Card (VIC) on all supported operating systems
Default Value	8
Valid Range of Values	0 - 255
Recommendations on Changing Default Values	Check with the storage array vendor's documentation on a recommended value for this parameter.

Parameter Name	Plogi Timeout (ms)
Description	Number of milliseconds to wait before timing out of the PLOGI exchange
Valid Adapter Models and Operating Systems	Cisco Virtual Interface Card (VIC) on all supported operating systems
Default Value	20000
Valid Range of Values	1000 - 255000
Recommendations on Changing Default Values	Check with the storage array vendor's documentation on a recommended value for this parameter.

Parameter Name	Error Detect Timeout (ms)
Description	Retry interval, how many ms to wait, for various FC commands, part of the FC standard being depreciated moving forward.
Valid Adapter Models and Operating Systems	Cisco Virtual Interface Card (VIC), M71KR-E/Q, M72KR-E/Q on all supported operating systems
Default Value	2000
Valid Range of Values	1000 - 100000
Recommendations on Changing Default Values	Check with the storage array vendor's documentation on a recommended value for this parameter.

Parameter Name	Port Down Timeout (ms)
Description	Number of milliseconds a remote FC port should be offline before informing the SCSI upper layer that the port has failed. This is important for host multi-pathing drivers and it is one of the key indicators that are used for error processing.
Valid Adapter Models and Operating Systems	Cisco Virtual Interface Card (VIC) on all supported operating systems
Default Value	30000 **(NOTE: ESX recommended value is 10000)**
Valid Range of Values	0 - 240000
Recommendations on Changing Default Values	Check with the storage array vendor's documentation on a recommended value for this parameter.

Parameter Name	Port Down I/O Retry
Description	Number of times I/O is sent back to upper SCSI layer when a remote port is down before failing the I/O.
Valid Adapter Models and Operating Systems	Cisco Virtual Interface Card (VIC) on Windows only.
Default Value	8
Valid Range of Values	0 - 255
Recommendations on Changing Default Values	Check with the storage array vendor's documentation on a recommended value for this parameter.

Parameter Name	Link Down Timeout (ms)
Description	Number of milliseconds the uplink should stay offline before informing the

	SCSI layer that the uplink is down and fabric connectivity is lost
Valid Adapter Models and Operating Systems	Cisco Virtual Interface Card (VIC) on Windows only.
Default Value	30000
Valid Range of Values	0 - 240000
Recommendations on Changing Default Values	Check with the storage array vendor's documentation on a recommended value for this parameter.

Parameter Name	Resource Allocation Timeout (ms)
Description	Resource allocation timeout value as part of the general FC specification.
Valid Adapter Models and Operating Systems	Cisco Virtual Interface Card (VIC), M71KR – E/Q, M72KR – E/Q on all supported operating systems
Default Value	10,000
Valid Range of Values	Hardcoded value.
Recommendations on Changing Default Values	Check with the storage array vendor's documentation on a recommended value for this parameter.

Parameter Name	IO Throttle Count
Description	Maximum number of outstanding I/O requests, data or control, per HBA. Note that is *not* LUN queue depth that will be discussed later in this document. If this value is exhausted then I/O will wait in the queue for servicing
Valid Adapter Models and Operating Systems	Emulex and Qlogic CNA models as currently the Cisco Virtual Interface Card (VIC) and driver ignores this parameter.
Default Value	16.
Valid Range of Values	1 - 1024
Recommendations on Changing Default Values	Check with the storage array vendor's documentation on a recommended value for this parameter.

Parameter Name	Max LUNs Per Target
Description	Number of LUNs behind a target controller that the FC driver will export/show.
Valid Adapter Models and Operating Systems	VIC and Qlogic and Emulex adatpers on all supported operating systems
Default Value	256 **Note: Recommended value for ESX and Linux is 1024**
Valid Range of Values	1 - 1024
Recommendations on Changing Default Values	This value will vary with each operating system and the user should consult the latest documentation on this from their respective OS vendor.

Parameter Name	Interrupt Mode
Description	Method used to send interrupts to the operating system from the driver. Message Signaled Interrupt - Extended (MSI-X). MSI-X is part of the PCIe 3.0 specification and results in better performance.
Valid Adapter Models and Operating Systems	Cisco Virtual Interface Card (VIC), M71KR/M72KR E-Q on all supported operating systems except Windows where this is ignored.
Default Value	MSI-X
Valid Range of Values	MSI-X, MSI, INTX
Recommendations on Changing Default Values	Do not change unless the operating system cannot support this method.

Hardcoded Parameters

LUN Queue Depth

This value affects performance in a FC environment when the host throughput is limited by the various queues that exist in the FC driver and SCSI layer of the operating system

This Cisco VIC adapter sets this value to 32 per LUN on ESX and Linux and 255 on Windows and does not expose this parameter in the FC adapter policy. Emulex and Qlogic expose this setting using their host based utilities. Many customers have asked about how to change this value using the Cisco VIC adapter. Cisco is considering this request as an enhancement for a future release. However FC performance with the VIC adapter has been excellent and there no cases in evidence (that the author is aware of) indicating that this setting is not optimal at its current value. It should be noted that this is the default value recommended by VMware for ESX and other operating systems vendors.

Array Vendor Considerations

Most storage array vendors will document the key error handling values they require for their controller for a specific adapter model and operating system. These values have been determined after significant testing on the performance aspect as well as the various failure scenarios. These values should be consulted in the documentation and support matrices of the respective storage vendor.

Appendix II: Cisco and NetApp Support Matrices and References

Many customers ask both Cisco and NetApp engineers how to navigate the various support matrices the two companies maintain. This section will attempt to provide links to the key resources as well as describe how a combination gets officially supported by both companies.

Cisco HCLs

The following is a public link to the Cisco server HCL. These documents cover operating systems versions, blade models, adapter models and the necessary Ethernet and FC driver versions for these combinations. A common error is that the customer does not verify the operating system driver is at the correct level for a given OS version. The Cisco VIC drivers (enic and fnic) are the ONLY UCS host components that exist but they must be carefully adhered to.
http://www.cisco.com/en/US/products/ps10477/prod_technical_reference_list.html

Cisco also maintains a public link showing which FC switches and storage arrays have been tested and supported against different UCS versions. This document will show the FC switch code levels required for a given release of UCS and generally whether or not an array vendor is supported on this combination. For the details from NetApp on what exactly they support with a given UCS combination one must consult the NetApp Interoperability Matrix website and associated tools.
http://www.cisco.com/en/US/docs/switches/datacenter/mds9000/interoperability/matrix/Matrix8.html

As of the UCS 2.0(2) Release the information in this matrix will be moved into the UCS HCL referenced above.

NetApp Interoperability Matrix Tool

NetApp's interoperability group runs hosts, adapters, and switches at varying firmware and OS revisions through a rigorous set of tests to see that all of the components in a customer's SAN will work together correctly. For this reason it is important to ensure that your configuration and firmware revisions match one of the rows in the

NetApp IMT.
http://support.netapp.com/NOW/products/interoperability/

Process Flow to Check Your Configuration

1. Check the Cisco Server HCL to ensure your blade/adapter/OS combination is even supported
2. Use the Cisco Server HCL to ensure you have the right driver versions in your OS image.
3. Check the Cisco Storage HCL (2nd URL provided above) to make sure the UCS version you want to deploy is supported against the version of your FC switching infrastructure. You should also check here that the array vendor is supported against verson.
4. Use the NetApp IMT to confirm the ONTAP and UCS versions and to review any caveats for a UCS + Switch + NetApp Array combination which may be listed as a note or alert on the IMT row matching your configuration.

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.**

Printed in USA C11-702584-01 05/12