FCoE boot from SAN with Dell Networking S5000

Dell Networking Solutions Engineering
September 2013

A Dell EMC Deployment/Configuration Guide

Revisions

Date	Description	Authors
September 2013	Initial release	Neal Beard, Victor Teeter, Stephen Swain, Robert Luis, Manjesh Siddamurthy, Paresh Kelvekar, Mike Orenstein,Kevin Locklear, and Ed Blazek

Table of contents

Executive Summary

The Dell Networking S5000 Converged switch is an efficient and powerful switching platform for Ethernet and FCoE protocols. It is important to understand how these S5000 switches will be utilized in any data center design. In many Enterprise environments, the need for a converged infrastructure pushes the need for redundancy, management, and control. There is an opportunity to boot the server's host OS as well as any VM's completely from the remote storage array utilizing the S5000's FCoE capabilities. In this Deployment Guide the topic will be booting from SAN with FCoE networking and FC storage. This guide will explain one possible scenario that will aid in the build out of an FCoE boot from SAN setup. This step by step process will explain how the configuration was validated using Dell Networking's S5000 switch, Brocade 5300 Fiber Channel switches, EMC VNX 5300 Storage set in block storage mode, and Dell PowerEdge M620 blade servers using Broadcom 57810 converged network adapters.

The step by step scenario is intended to cover one validated solution but should be applicable in many cases to other hardware and devices with minor configuration changes based on the technology used. Use this guide as a knowledge base if the hardware is different. Please feel free to provide feedback on this document for further enhancements.

1 Introduction

This Deployment Guide's intention is to lead network administrators and design architects through the steps of an FCoE boot from SAN configuration. It is applicable to the particular hardware mentioned but can be used as guidelines for other boot from SAN implementations.

1.1 Objective

The scope of this document is covering the specific hardware mentioned in the scenario but as mentioned before, this document will also serve as a guideline for hardware that falls outside the scenario covered. It will not be an exact match with other hardware such as different storage, CNA, or networking switches, but will provide the technology guidance to understand what should be achieved.

1.2 Audience

This document is for all respective network engineers, SAN administrators, Server administrators, and end-users that will need to implement FCoE boot from SAN in the infrastructure being built or upgraded. The intended audience should have a general networking, server and storage skill set in order to understand the topics being discussed.

2 Initial Configuration

The configuration steps used to create an end to end Boot from SAN (BFS) FCoE configuration will use the HW in Figure 1. ESXi 5.1 update 1 and Windows 2008 R2 will be the OS's used in this deployment guide.

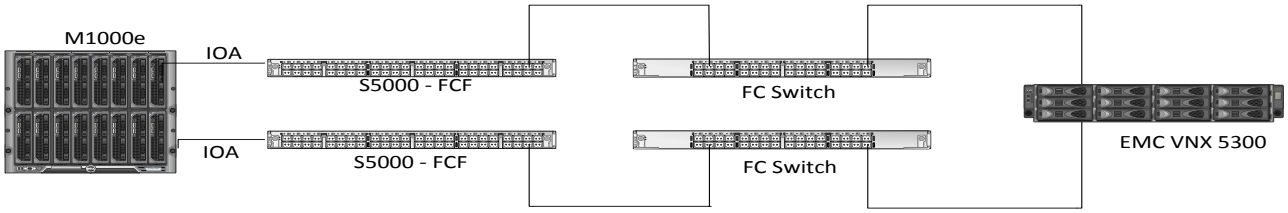

Figure 1 Validated Example Configuration

Table 1 Infrastructure Components

Infrastructure Components	Firmware Revision
Dell M620 Server	1.6.1 (BIOS)
Dell I/O Aggregator	8.3.17.4
Dell S5000	9.0.1.0B2
Brocade B5300	7.1.0a
Broadcom 57810 CNA	7.4.8
EMC VNX 5300	05.32.000.5.008

Hardware needed for installing VMware ESXi 5.1 update 1 and Windows 2008 R2:

- USB DVD external drive
- USB memory stick with Windows 2008 R2 FCoE BFS drivers
- DVD with Windows 2008 R2 image or Dell VMware image.
- A Windows 2008 R2 or VMware .iso image can also be used via Dell's iDRAC Virtual Media

URL on Dell EMC's support site for VMware ESXi 5.1 update 1 image is below:

http://www.dell.com/support/drivers/us/en/04/DriverDetails/Product/poweredge-m820?driverId=RD5JN&osCode=XI51&fileId=3176127681&languageCode=EN&categoryId=EC

Note: The above image has additional support for FCoE BFS drivers

3 Setup:

1. Attach USB DVD external drive/or iDRAC virtual media .iso image to desired Server for FCOE boot from SAN.
2. Login to CMC
3. Select Server Overview
4. Select slot that server resides in
5. Click on WWN/MAC
6. Make a note of the base MAC of the CNA that will be used for FCoE boot from SAN.
7. Also note the fabric it is located on. A1 or A2 etc… This will help identify the adapter when in the Server Adaptor BIOS, especially if using NPAR. The last 2 bytes of the base Mac will be used to identify the initiator in the BIOS.
8. Go to WWN/MAC section. See figure 2 below.

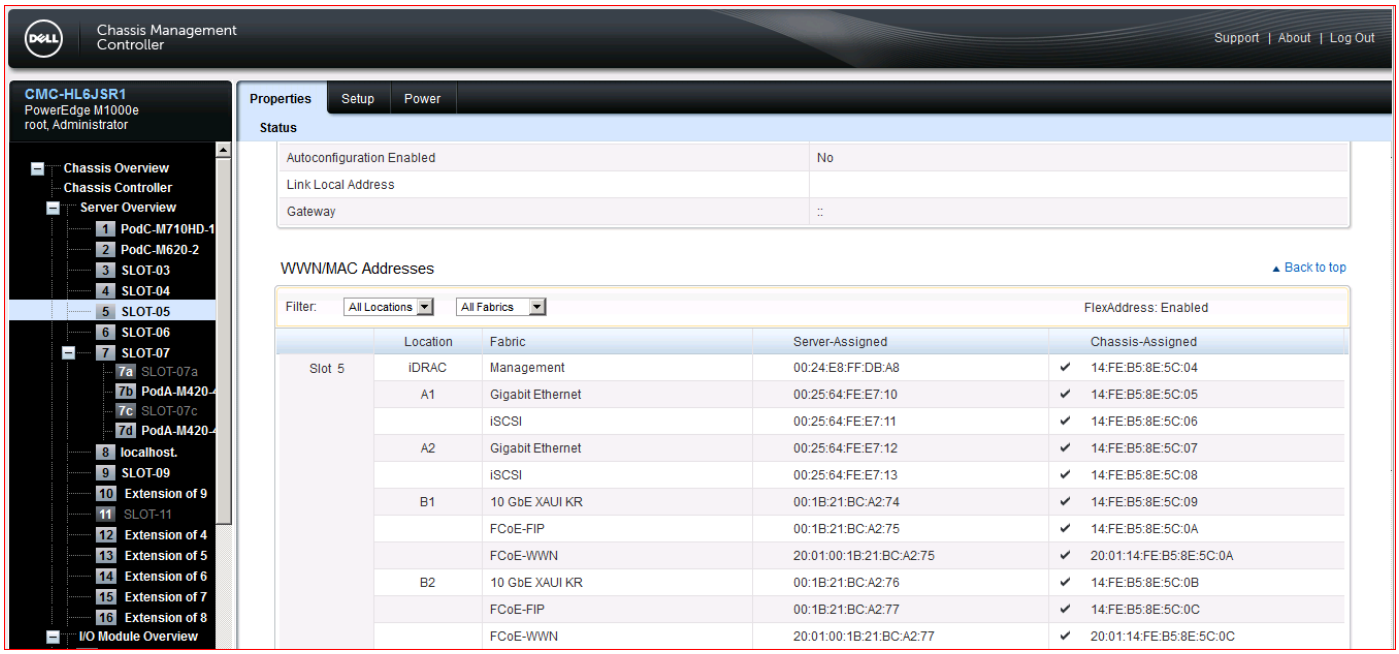

Figure 2 WWN/MAC table from CMC

9. Record addresses (server-assigned or chassis-assigned (Flex Address option)) for the Fabric you are attaching to (e.g. B2 and C2). This information will be used in a later step.
10. Select the server that FCoE BFS will be configured on. Reboot the server to access the System Setup menu (F2 during POST).

3.1 Fibre Channel Switch Configuration:

1. Zoning will need to be setup between the CNA and the Storage Processor
2. Please ensure that only the CNA and the Storage Processor WWPN is in the zone
3. Add the zone to the Zone configuration and enable the configuration.

3.2 Provisioning Storage Access in the SAN

Storage access consists of zone provisioning and storage selective LUN presentation, each of which is commonly provisioned per initiator WWPN. Two main paths are available for approaching storage access:

- Pre-Provisioning
- CTRL+R Method

3.2.1 Pre-Provisioning

With pre-provisioning, note the initiator WWPN and manually modify fabric zoning and storage selective LUN presentation to allow the appropriate access for the initiator.

The initiator WWPN can be seen at the bottom of the screen in the FCoE boot target configuration window.

The initiator WWPN can also be directly inferred from the FIP MAC address associated with the interface(s) planned for boot. Two MAC addresses are printed on stickers attached to the SFP+ cage on your adapter. The FIP MAC ends in an odd digit. The WWPN is 20:00: + <FIP MAC>. For example, if the FIP MAC is 00:10:18:11:22:33, then the WWPN will be 20:00:00:10:18:11:22:33.

Note: The default WWPN is 20:01: + <FIP MAC>. The default WWNN is 10:00: + <FIP MAC>. In Dell FlexAddress configurations, the FCoE/FIP MAC will be overridden by the blade chassis management system.

3.2.2 CTRL+R Method

The “CTRL+R” method allows you to use the boot initiator to bring up the link and login into all available fabrics and targets. Using this method, you can ensure that the initiator is logged into the fabric/target before making provisioning changes, and as such, can provision without manually typing in WWPNs.

1. Configure at least one boot target through CCM (CTRL-S) as described above.
2. Allow the system to attempt to boot through the selected initiator.
3. Once the initiator boot starts, it will commence with DCBX sync, FIP Discovery, Fabric Login, Target Login, and LUN readiness checks. As each of these phases completes, if the initiator is unable to proceed to the next phase, MBA will present the option to press CTRL+R.
4. Once “CTRL+R” has been activated, the boot initiator will maintain a link in whatever phase has most recently succeeded and allow you time to make the necessary provisioning corrections to proceed to the next phase.

5. If the initiator logs into the fabric, but is unable to log into the target, a “CTRL+R” will pause the boot process and allow you to configure fabric zoning.
6. Once zoning is complete, the initiator will automatically log into all visible targets. If the initiator is unable to discover the designated LUN on the designated target as provisioned in step 1, “CTRL+R” will pause the boot process and allow you to configure selective LUN presentation.
7. The boot initiator will periodically poll the LUN for readiness, and once the user has provisioned access to the LUN, the boot process will automatically proceed.

Note: This does not preclude the need to put the boot initiator into one-time disabled mode

3.3 S5000 Top of Rack switch

Figure 3 S5000 IO Side

Figure 4 S5000 PSU side

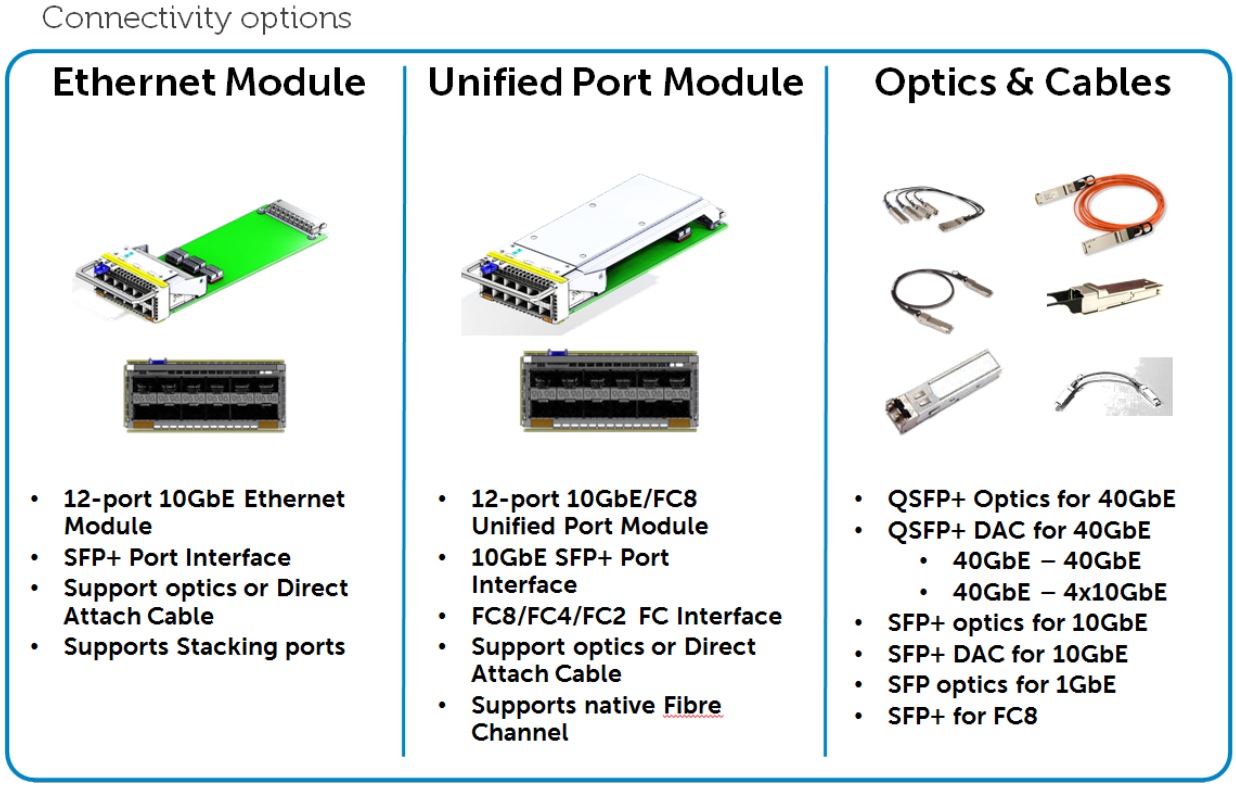

Figure 5 S5000 Connectivity Modules

3.4 S5000 configuration

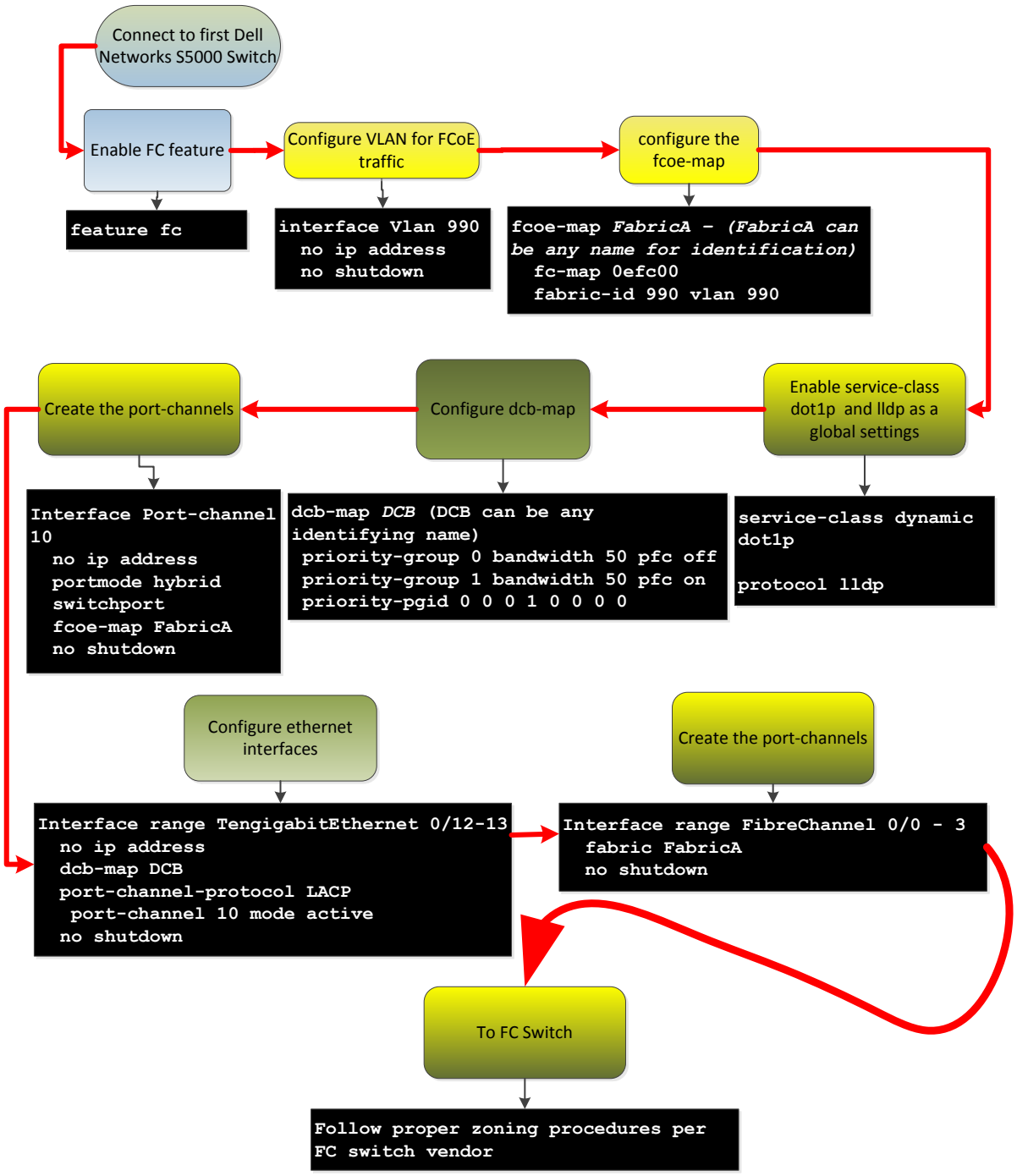

Figure 6 S5000 FCoE configuration flow

3.5 EMC VNX 5300 storage array setup and initiator LUN assignment.

The following steps will configure the EMC VNX 5300

3.5.1 Initial initiator confirmation

1. Confirm EMC storage sees the initiators.
2. Create new Host and Register initiator with storage device, typically done same time on first initiator.
3. Create required number of LUNs
4. Create storage group, listed under the Host Menu
5. Add Host and LUNs to new storage group.
    a. Confirm EMC storage sees the initiators
    b. From tab selection Menu , select “System List”, EMC VNX 53000 link
    c. New tab selection will be presented, select Host tab > initiators
    d. Look for initiator in list, should show as not registered. Select “Not Registered” from Connection status to help filter initiators

6. Create New Host and register initiator.
    a. Click on row with desired initiator
    b. On bottom left corner of screen, select Register
    c. A New window will show up.
        i. Initiator type= CLARiiON VNX (for Windows/VMware)
        ii. Failover Mode=Active-Active (default)
        iii. Select New Host
        iv. Enter host name , you may want to include Rack_Server_Fabric_SAN info
        v. IP address is needed even if not using host agent or Powerpath, IP address can be a fake address in order to create/register the initiator.
        vi. Enter ok
    d. You will see several prompts to continue/ confirm your request.
    e. You will see a final success box, hit OK when prompted
    f. Continue registering the other initiators if multiple paths are available.
    g. Select “existing host” (the one just created)
        i. Browse Host…
        ii. In Filter For: you can include host name to make the search easier if you have many hosts created.
        iii. Select host desired and hit OK
    h. Create desired number of LUNs.
        i. From top menu tabs, select Storage>LUNs
        ii. Bottom Left select “create” to open Create LUNs dialog box
        iii. Storage type: Leave at Pool , for lab purposes
        iv. RAID type: Select default, Lab case is RAID6
        v. Select Storage Pool for new LUN, lab leave at “Pool 0”

vi. LUN Properties
    1. User Capacity: select size, for ESX host at least 20GB minimum
    2. LUN ID : allow system to assign, no selection required
vii. Number of LUNS to Create: Enter a number you want. (Note LUNs assigned to SPA and SPB in round robin) So you need access to SPA and SPB.
viii. LUN Name: leave at default, which automatically assigns LUN IDS.
ix. Select “Apply”

7. Create storage group, listed under the Host Menu.
   a. From top menu tabs, select Host > Storage Groups
   b. Mid to Bottom left , select “Create”
      i. Create Storage box is presented
         1. Enter Storage Name: You may include Rack Server Fabric SAN Hostname
         2. Select OK, (this will show success then step through to add LUNs and Host)

8. Add Host and LUNs to new storage group
   a. From LUNs tab, Select LUNs to add, note SPA and SPB should show LUNs
   b. Select “add”, LUN moved to selected LUNs
   c. Select “Hosts” tab
      i. Select Host from Available Hosts
      ii. Hit right pointing arrow to move Host to “Hosts to be Connected”
      iii. Hit “OK”

3.5.2 Two ways to confirm the Host LUN ID info.

- You can view the storage group and note the Host LUN ID
  - From Storage group list: Host >Storage group
  - Select Storage group desired
  - In the Details window, bottom of web page, select LUNs tab
  - View the LUNs assigned to the storage group
  - Far right column will list the “Host LUN ID” this is the info needed for the initiator Boot from San info.
- From Host > Host List
  - Select Host from List
  - In the Details pane, far right column will show Host LUN ID.

3.6 FCoE Boot from SAN Configuration

The following section details the BIOS setup and configuration of the boot environment prior to the OS install.

The following steps can also be referenced in this document:
http://www.broadcom.com/docs/support/ethernet_nic/Broadcom_NetXtremeII_Server_T7.6.pdf

3.6.1 Preparing System BIOS for FCoE Boot from SAN (BFS)

-Modify System Boot Order: The Broadcom initiator must be the first entry in the system boot order. The second entry must be the OS installation media. It is important that the boot order be set correctly or else the installation will not proceed correctly. Either the desired boot LUN will not be discovered or it will be discovered but marked offline.

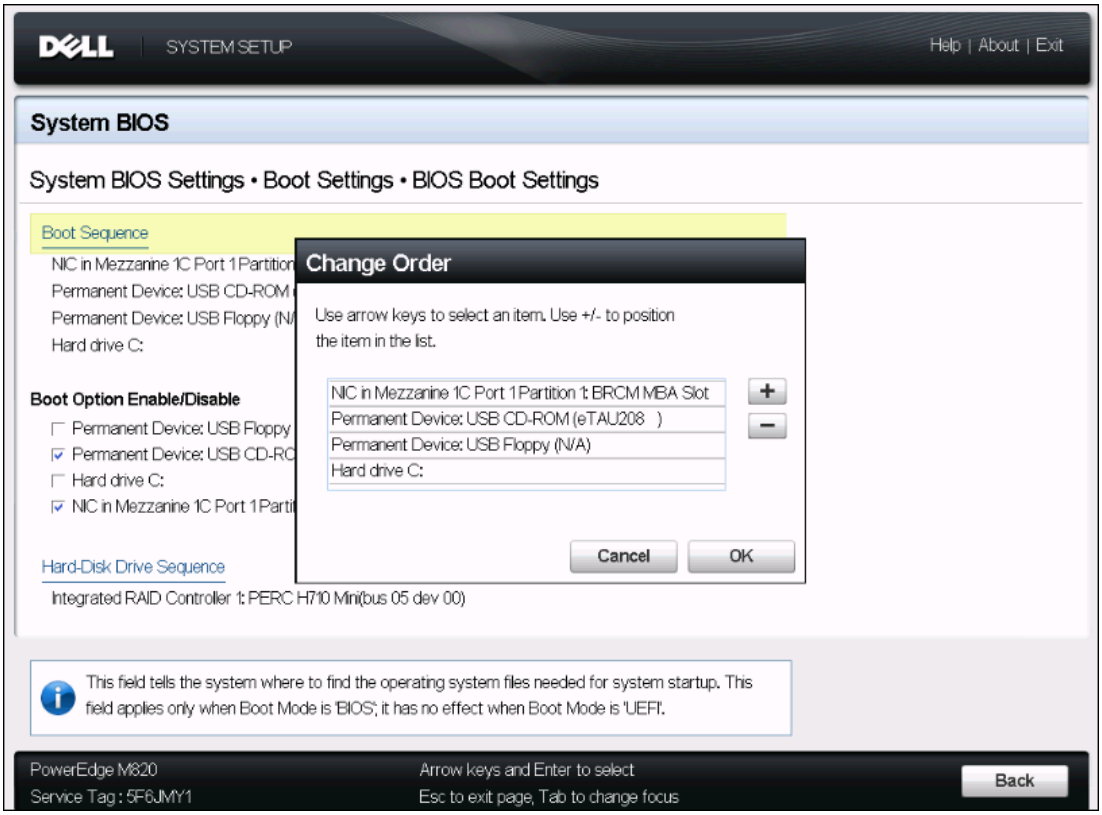

Figure 7 BIOS boot sequence configuration

3.6.2 Prepare Broadcom Multiple Boot Agent for FCoE Boot

1. During POST, press “CTRL+S” at the Broadcom NetXtreme Ethernet Boot Agent banner to invoke the CCM utility.

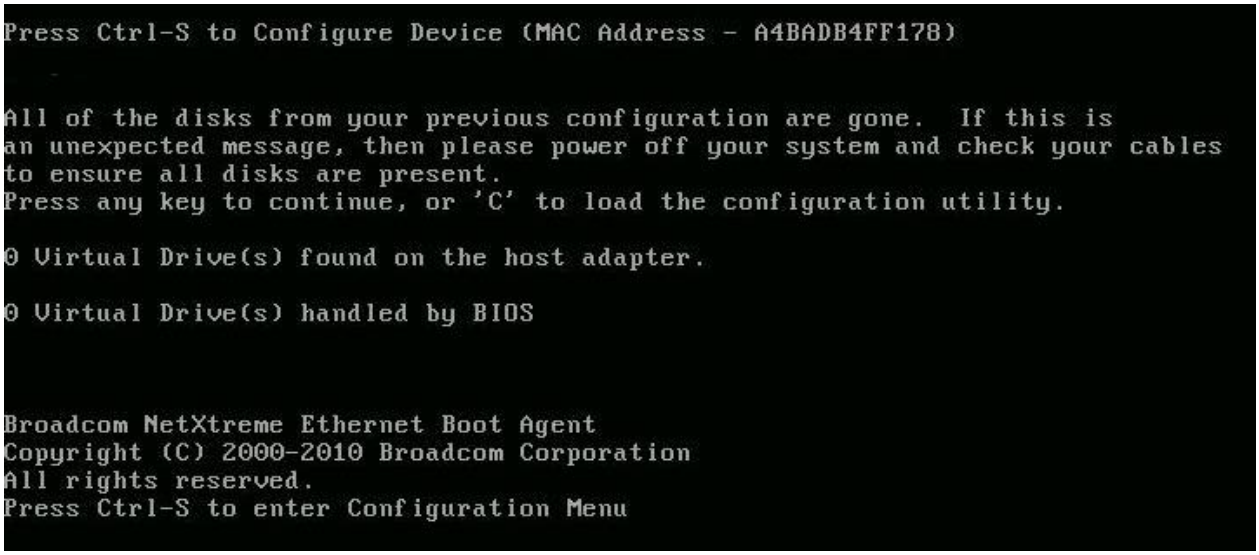

Figure 8 Broadcom CTRL-S at boot up of server

2. Select the device through which FCoE boot from SAN is to be configured.

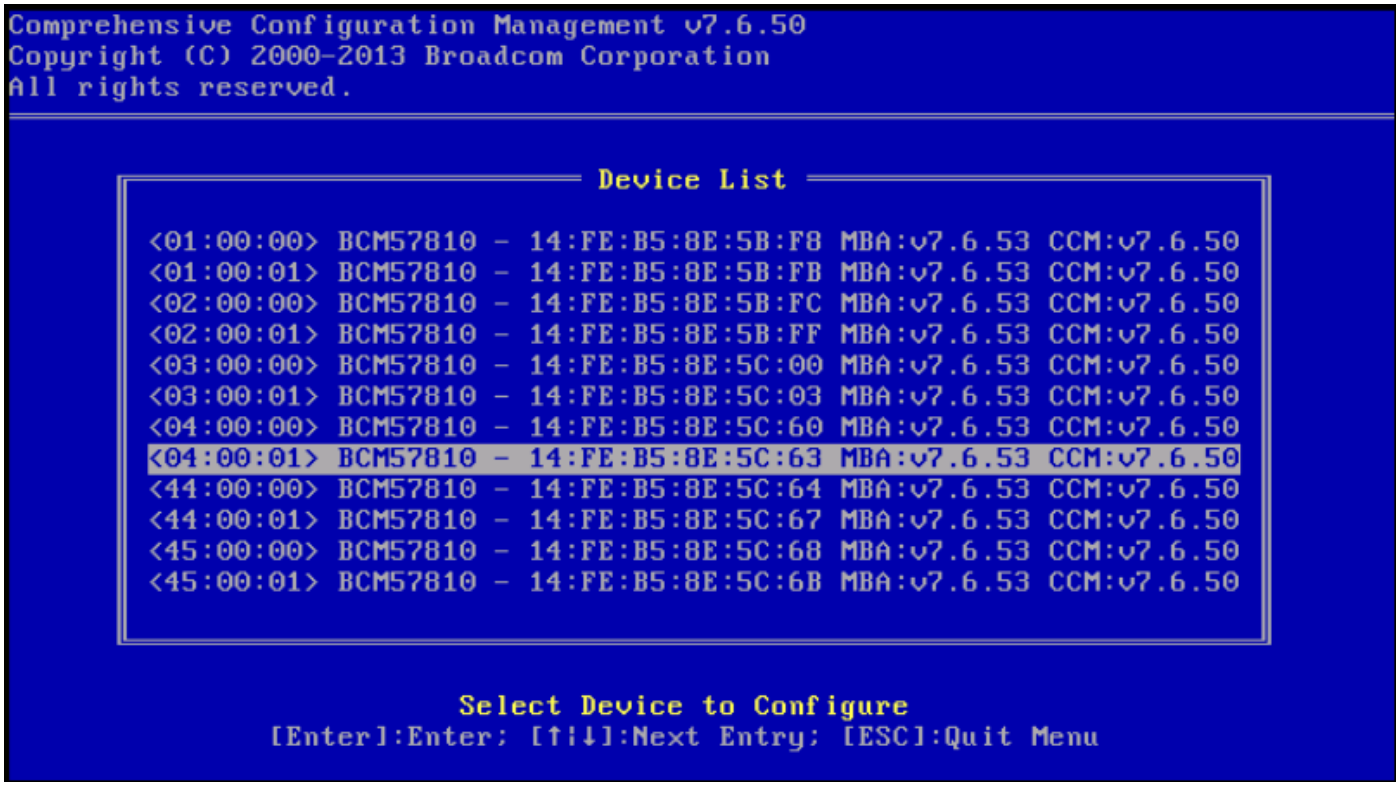

Figure 9 Broadcom Hardware Configuration settings device list

3. Enter the Device Hardware Configuration menu and highlight Multi-Function Mode, hit enter to toggle between SF (single function mode) or NPAR (network partition mode). Highlight DCB Protocol and hit enter to enable.
4. Hit esc and go to the MBA menu

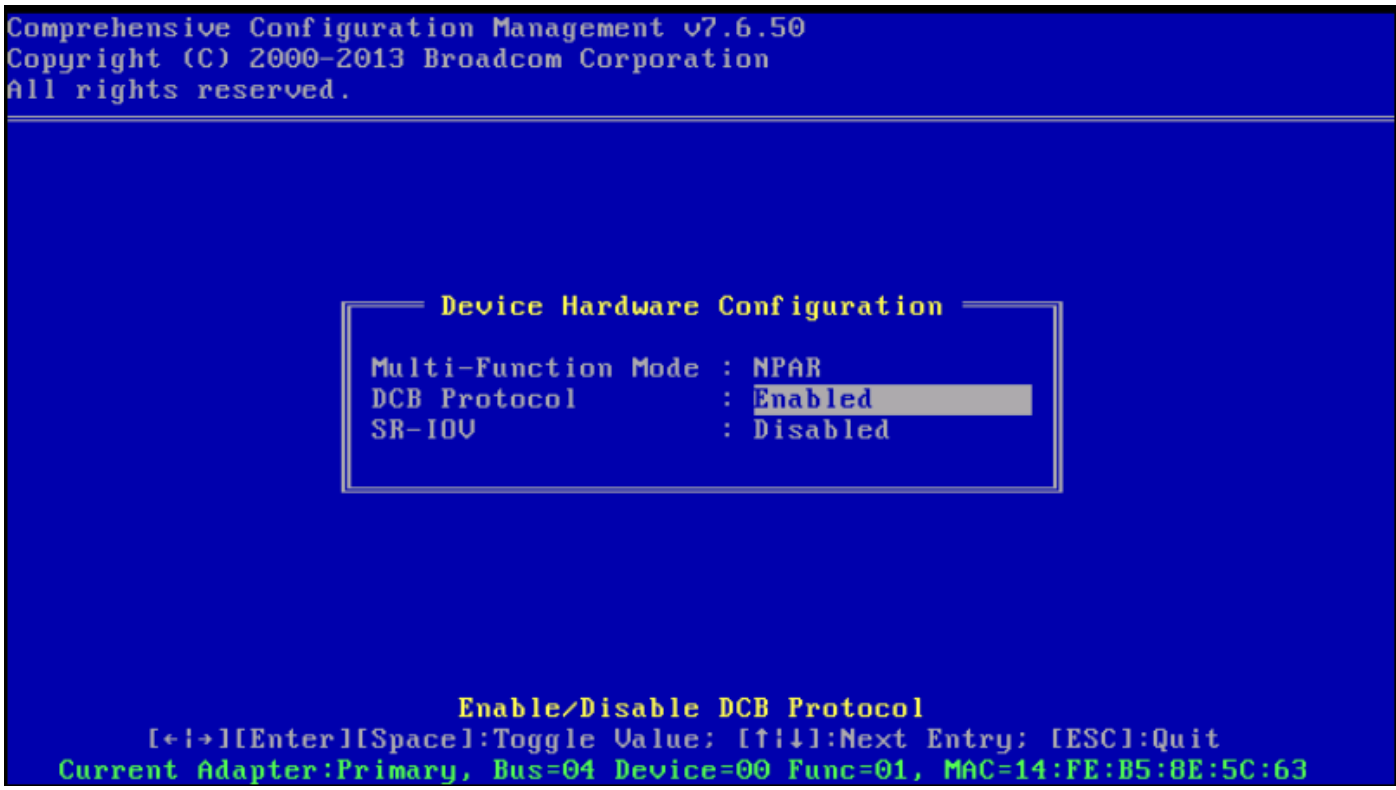

Figure 10 Device Hardware Configuration for NPAR and DCB Protocol

5. In the MBA Configuration Menu set the Option ROM to Enabled. Set the Boot Protocol field to FCoE, all other fields can be left at default values.
6. Hit esc and go to the FCoE Boot Configuration menu

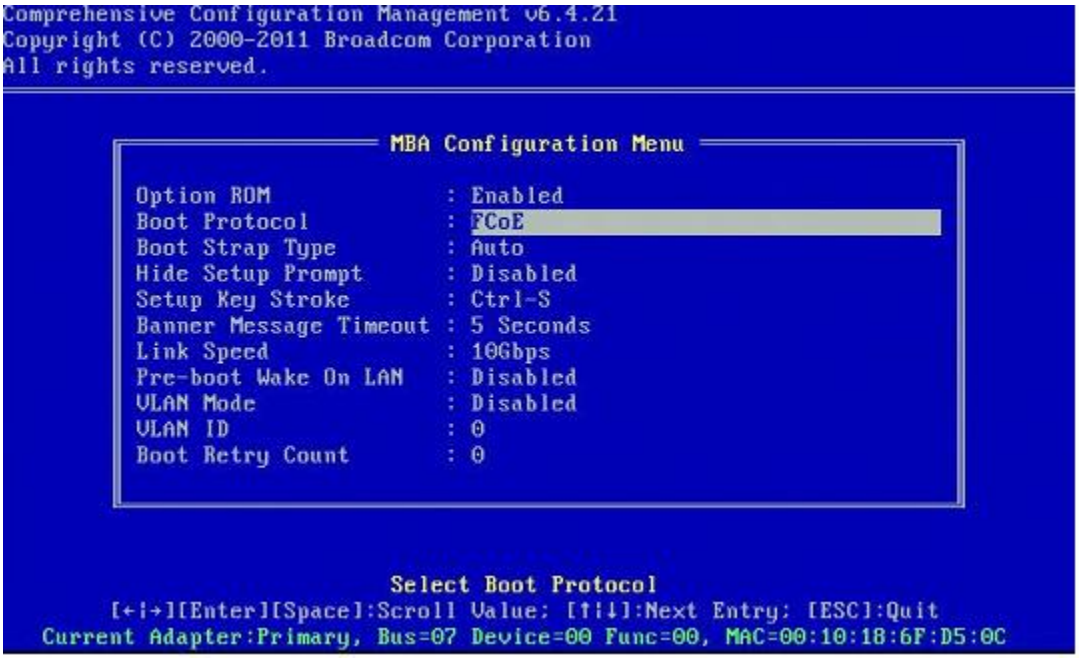

Figure 11 MBA Configuration Screen

7. Enter the FCoE Boot Configuration menu
8. Enter the General Parameters sub-menu
   Set the field “Boot to FCoE Target as “One Time Disabled”
   Set the field “Target as First HDD” to “Enabled”
   Set the field “HBA Boot Mode” to “Enabled”
   All other fields can be left at default settings
9. Hit esc and go to the Target Information Menu

Note: By setting the parameter One Time Disabled, this allows the CNA to bypass booting to the LUN (which has no OS installed) and allows the cdrom or iDRAC virtual media mounted .iso image to be the boot device for loading the OS. This setting will need to be re-enabled after each reboot if a further use of the cdrom is needed.

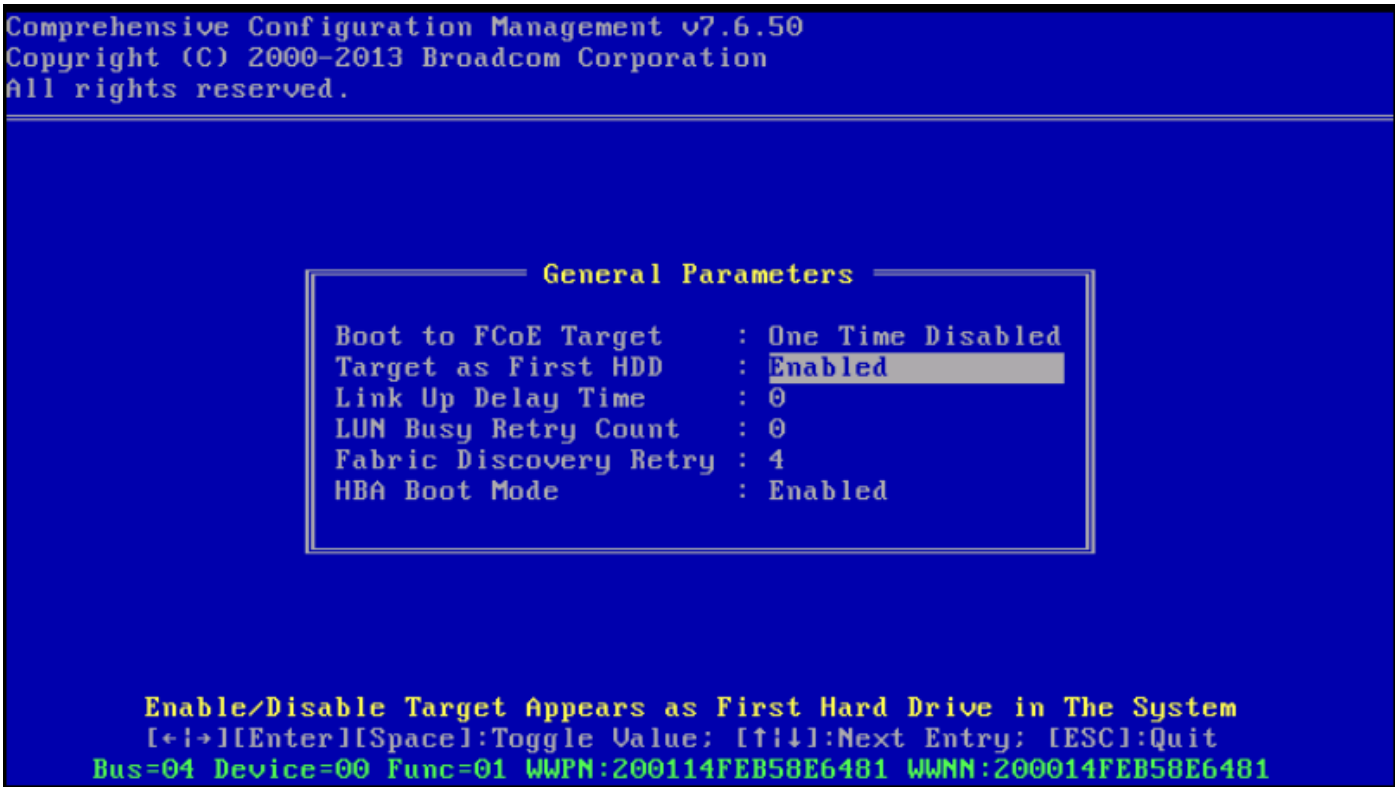

Figure 12 FCoE Boot Main Menu, General Parameters sub-menu

10. From the Target Information Menu select No. 1 Target.

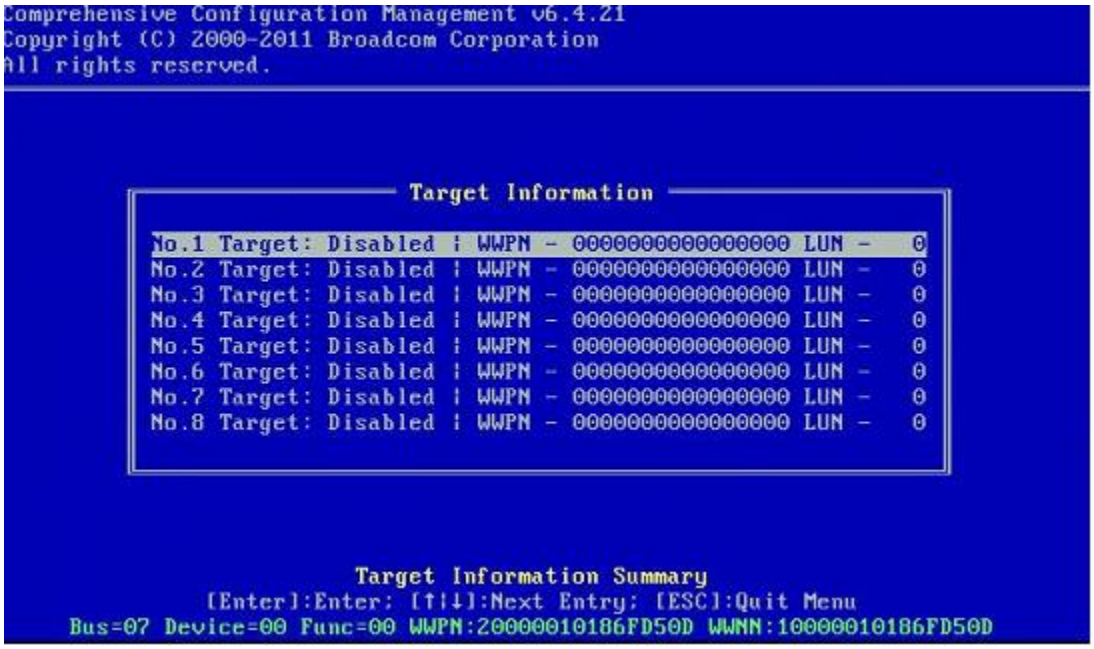

Figure 13 Target Information Screen

11. Press the “Enter” key to enable the “Connect” field.
12. Highlight “WWPN” field and hit enter. Type in the WWPN of the EMC VNX 5300 controller

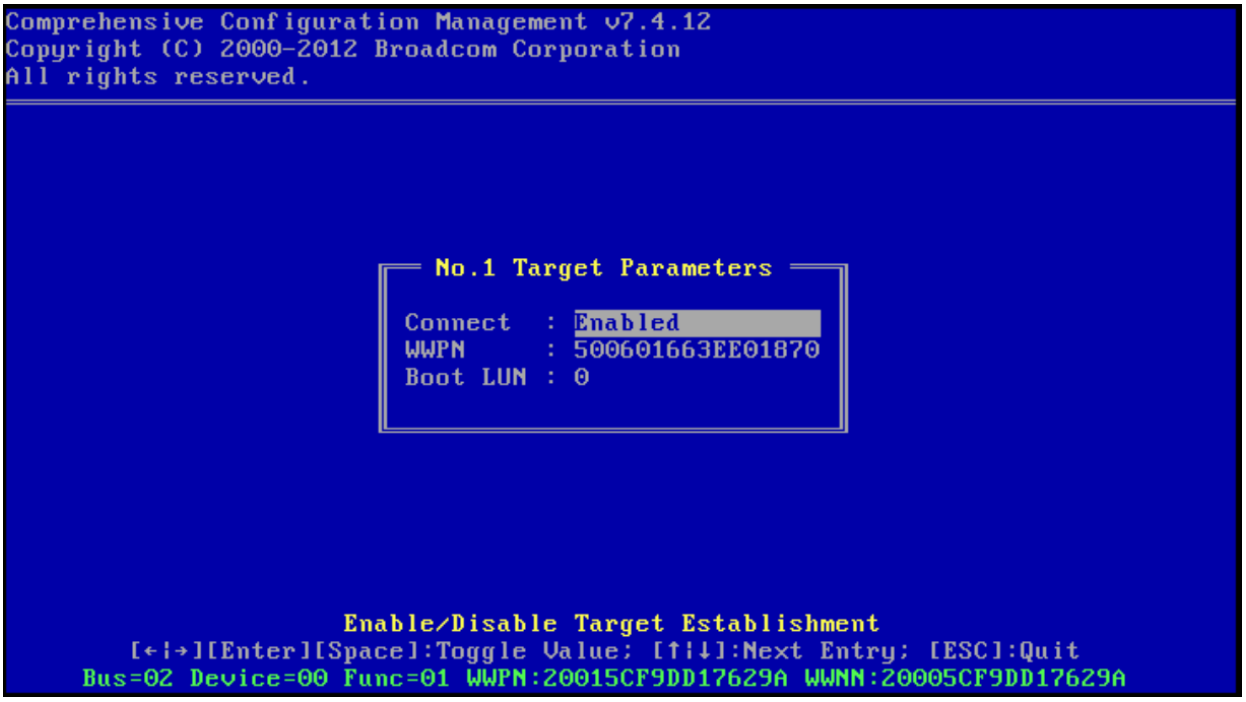

Figure 14 Target Parameters

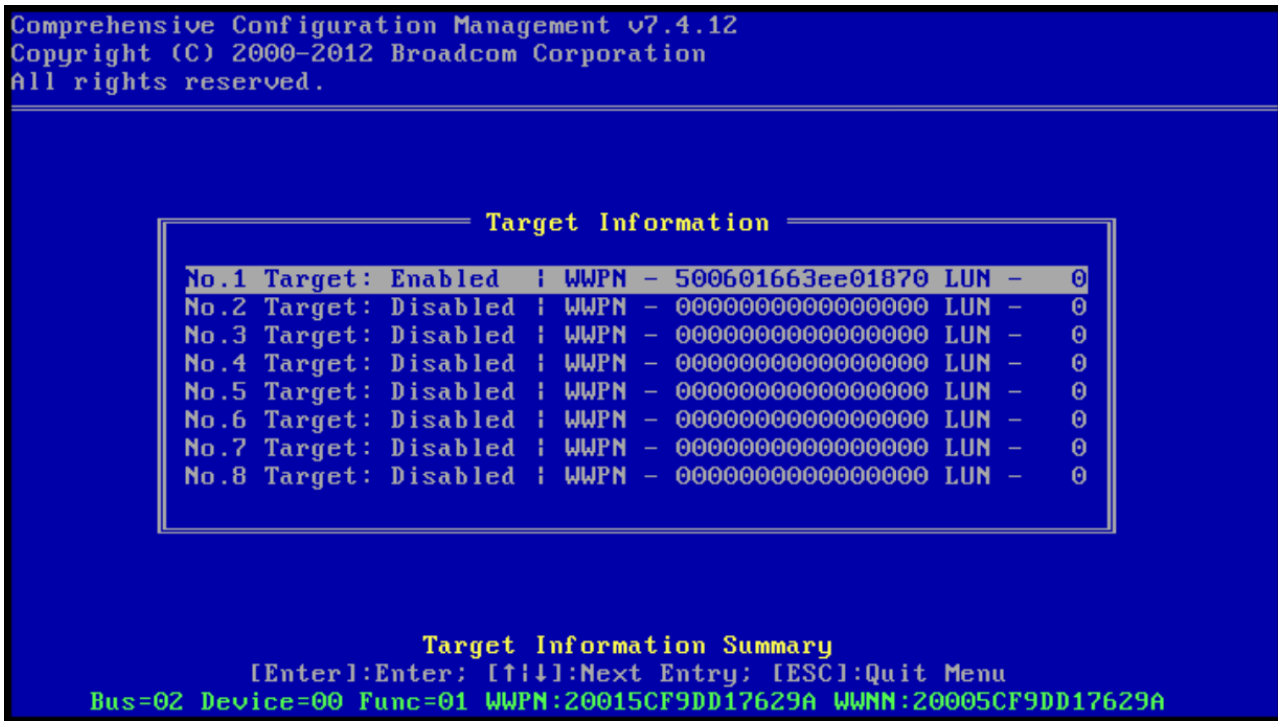

Figure 15 Target Information

13. Press ESC twice and hit enter on NIC Partition Configuration menu
Highlight Partition #1 and hit enter
Enable Ethernet Protocol and FCoE Offload Protocol
Ensure that iSCSI Offload Protocol is disabled
Hit ESC and ensure that the next 3 Partitions have all protocols disabled

Note: By disabling the other protocols this will ensure that only one path will be presented to the EMC VNX 5300 array. It will also ensure easier trouble shooting should an issue arise.

Note: IMPORTANT: When running in NIC Partitioning (NPAR) mode, FCoE boot is supported only when the first function on each port is assigned an FCoE personality. FCoE boot is not supported when the FCoE personality is assigned to any other function.

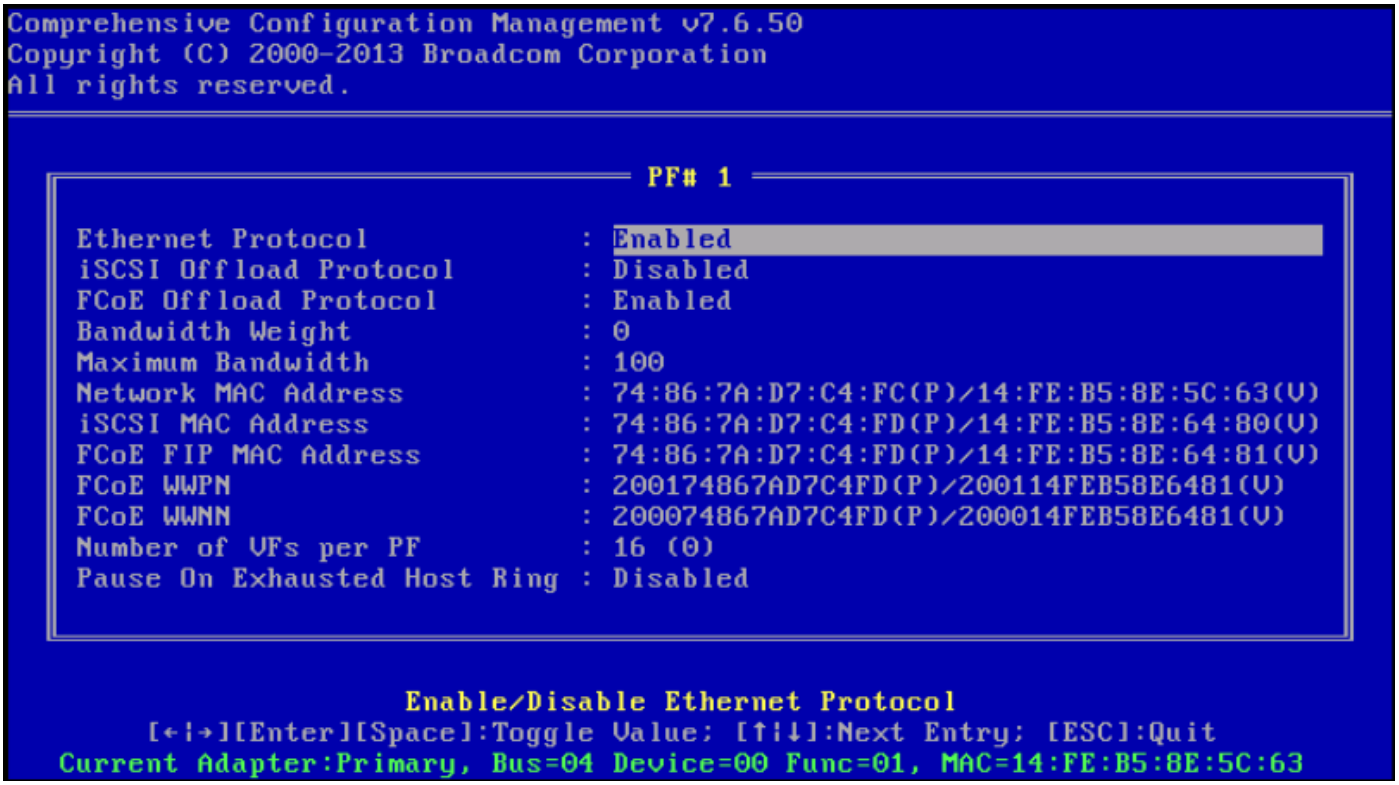

Figure 16 Partition #1 protocol configuration

14. Press ESC until prompted to exit and save changes.
15. The system will then reboot

16. At this point the Broadcom FCoE boot screen should appear and succeeded should be seen for each step of the identification and login process.

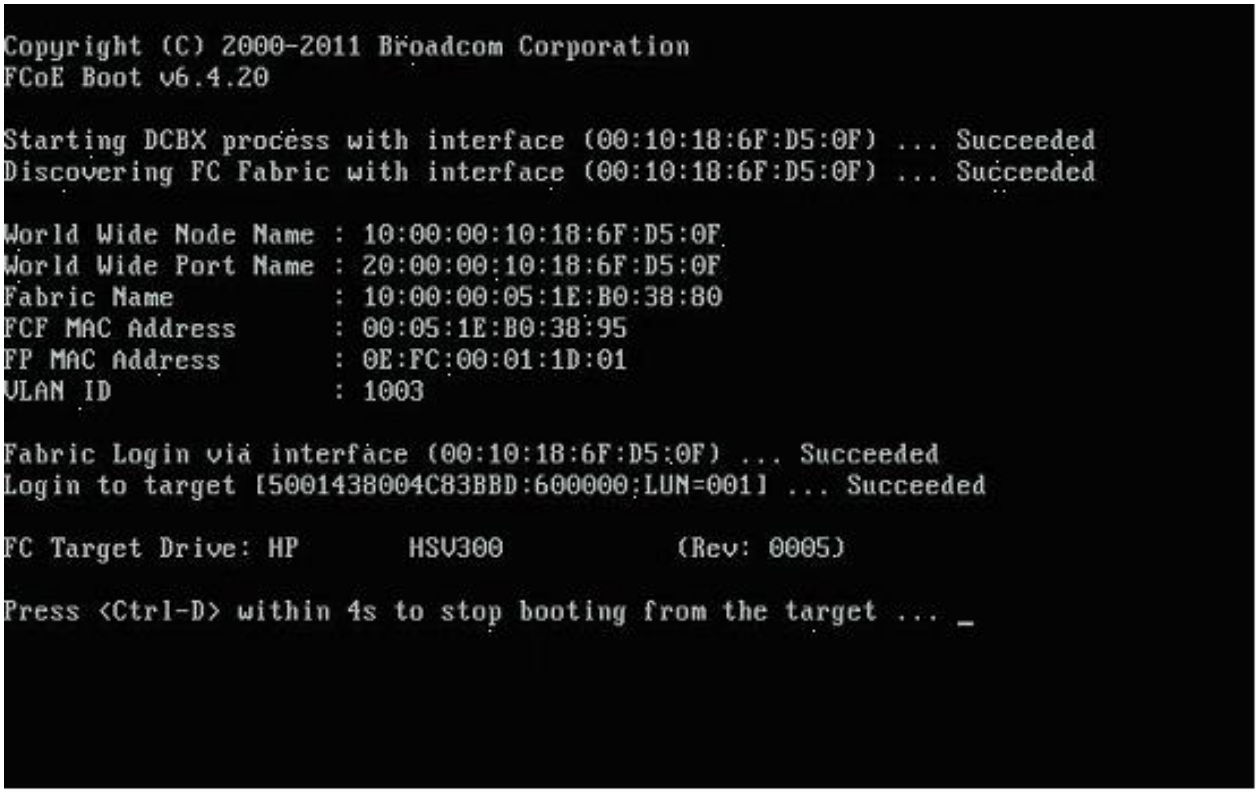

Figure 17 Successful FCoE BFS login and device discovery screen

4 VMware ESXi 5.1 FCoE Boot Installation

FCoE Boot from SAN requires that the latest Broadcom NetXtreme II async drivers be included into the ESXi 5.1 install image. Refer to *Image_builder_doc.pdf* from VMware on how to slipstream drivers.

Note: Refer to section 2.0 for a link to Dell's custom ESXi 5.1 Update 1 image with all current CNA drivers included.

1. Boot from the updated ESXi 5.1 installation image and select **ESX 5.1 installer** when prompted.

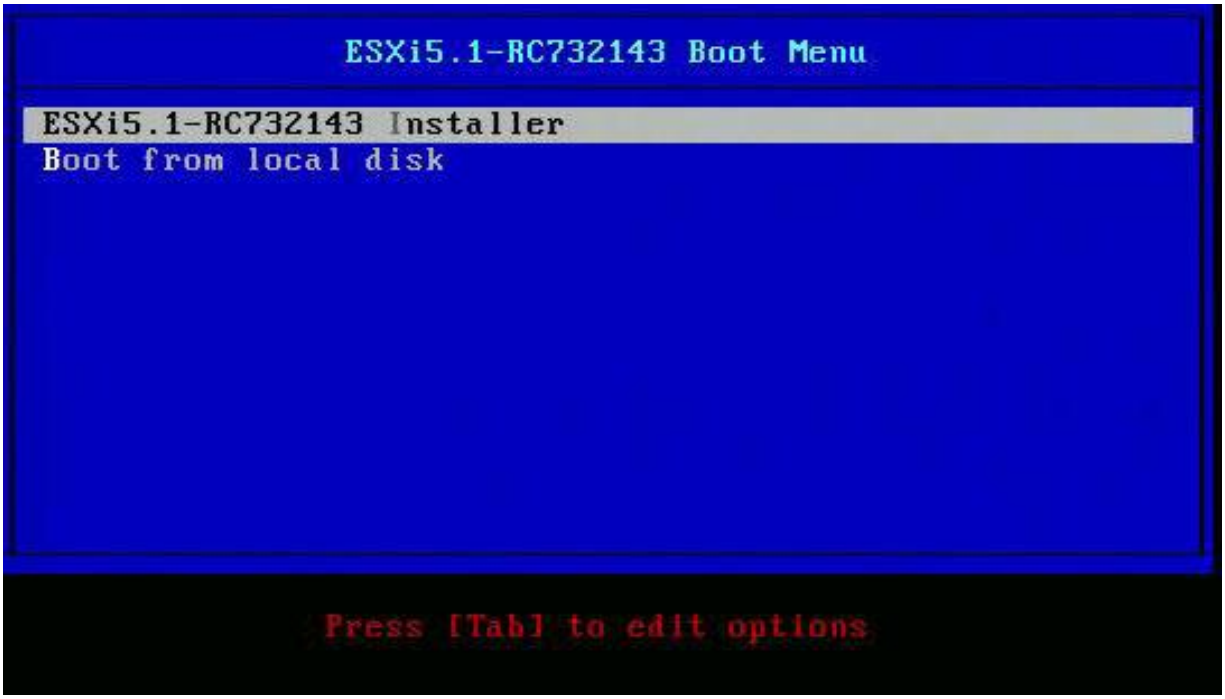

Figure 18 ESXi 5.1 Boot Menu

2. Press Enter to continue.

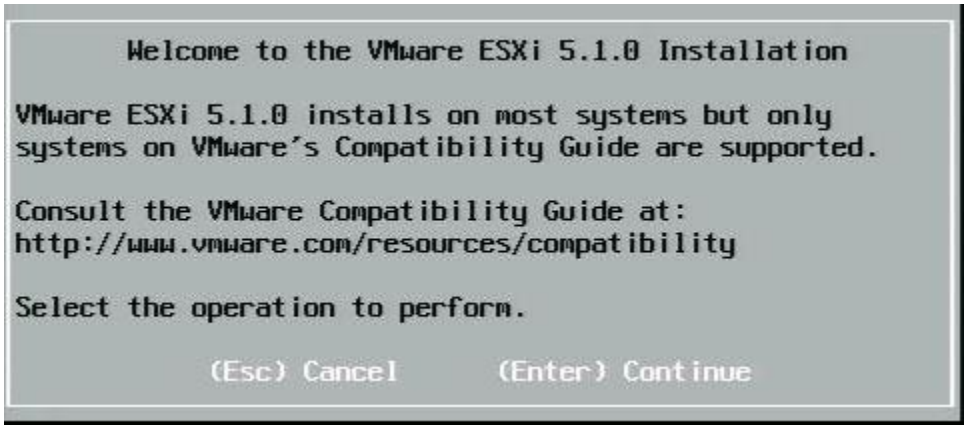

Figure 19 ESXi Installation screen

3. Press **F11** to accept the agreement and continue.

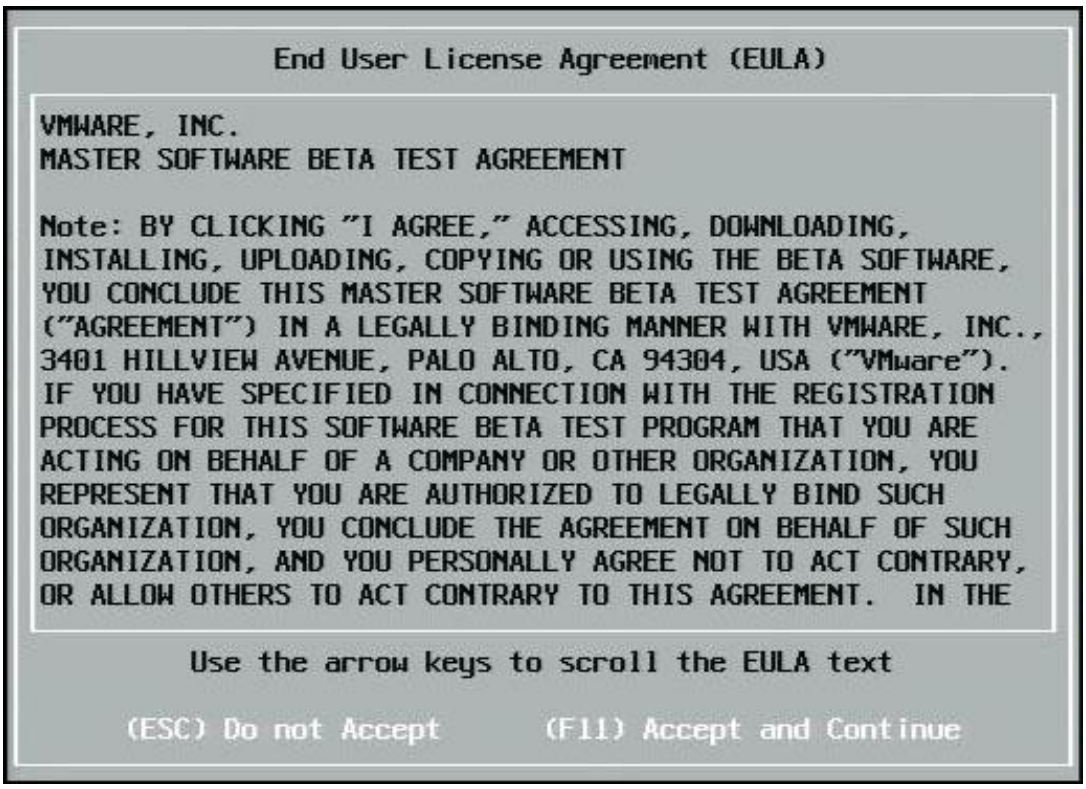

Figure 20 End User License Agreement

4. Select the boot LUN for installation and press **Enter** to continue.

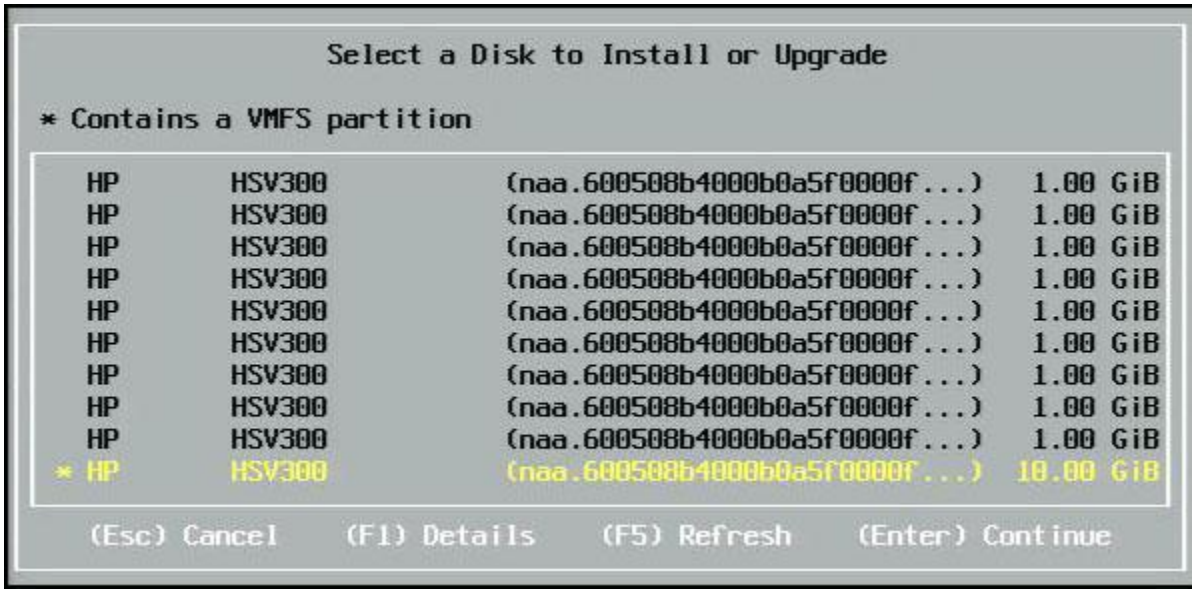

Figure 21 Disk selection menu

5. Select the desired installation method.

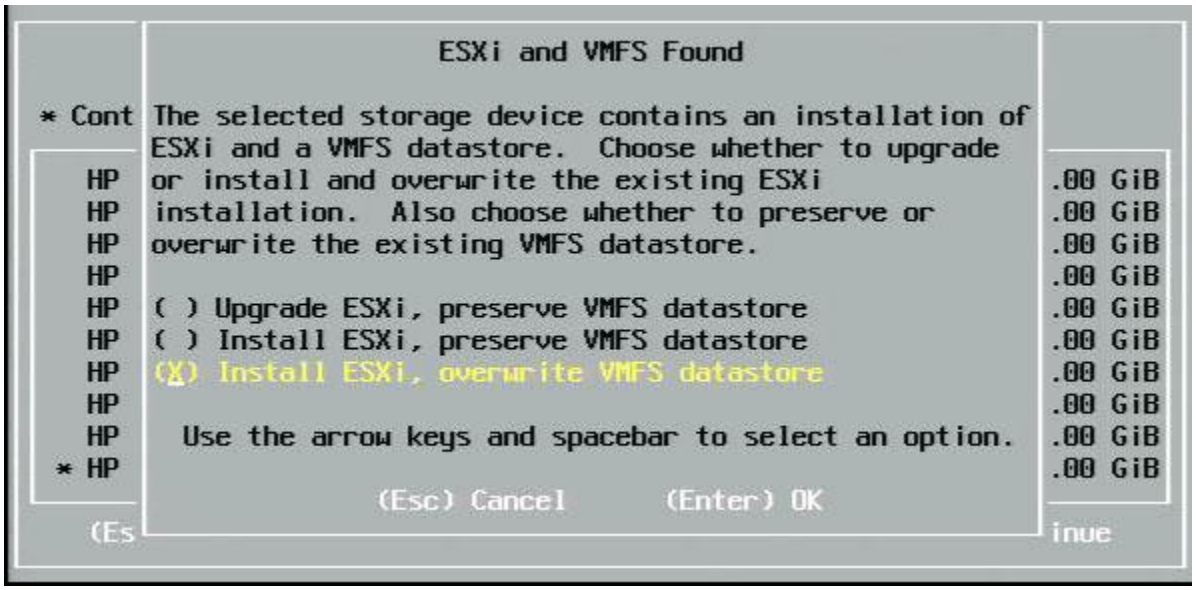

Figure 22 VMFS data store and ESXi installation

6. Select the keyboard layout.

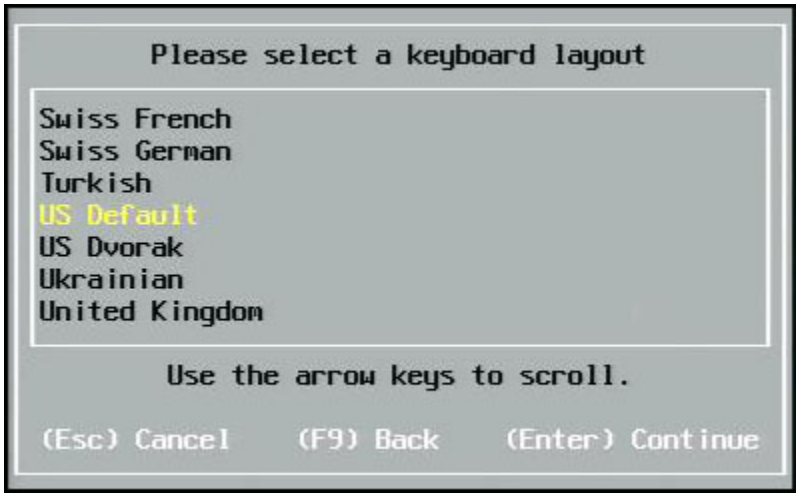

Figure 23 Keyboard layout menu

7. Enter a password.

Figure 24 Password

8. Press **F11** to confirm the install.

Figure 25 Install confirmation

9. Press **Enter** to reboot after installation.

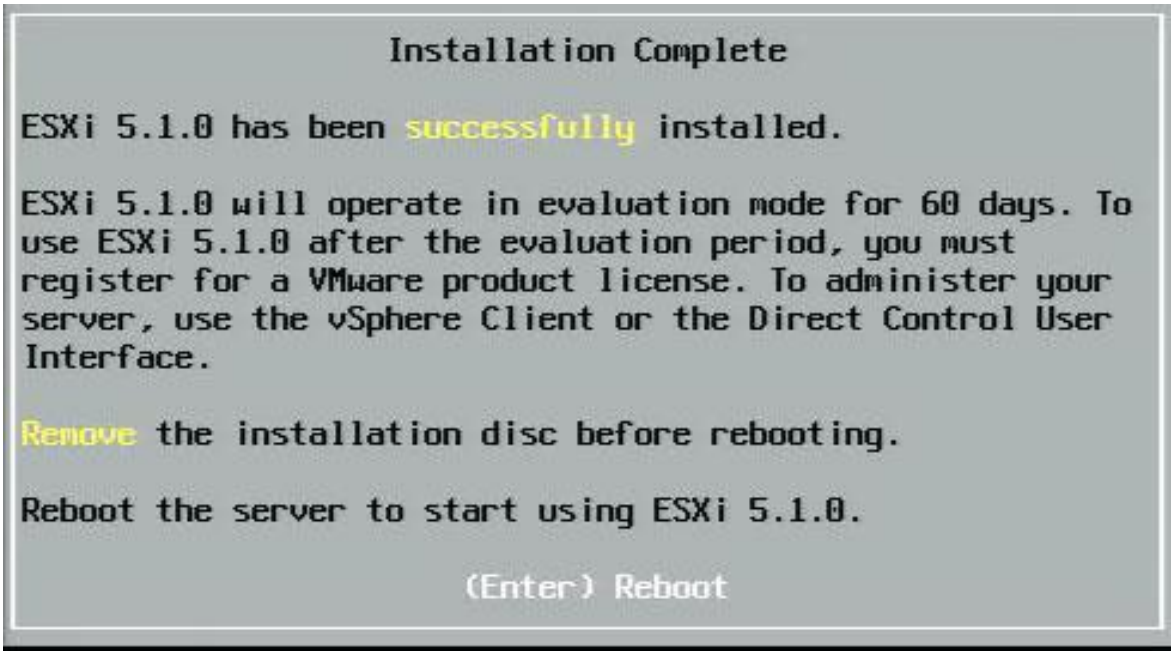

Figure 26 Installation Complete

10. On 57800 and 57810 boards, the management network is not vmnic0. After booting, open the GUI console and display the **configure management network** > **network adapters** screen to select the NIC to be used as the management network device.

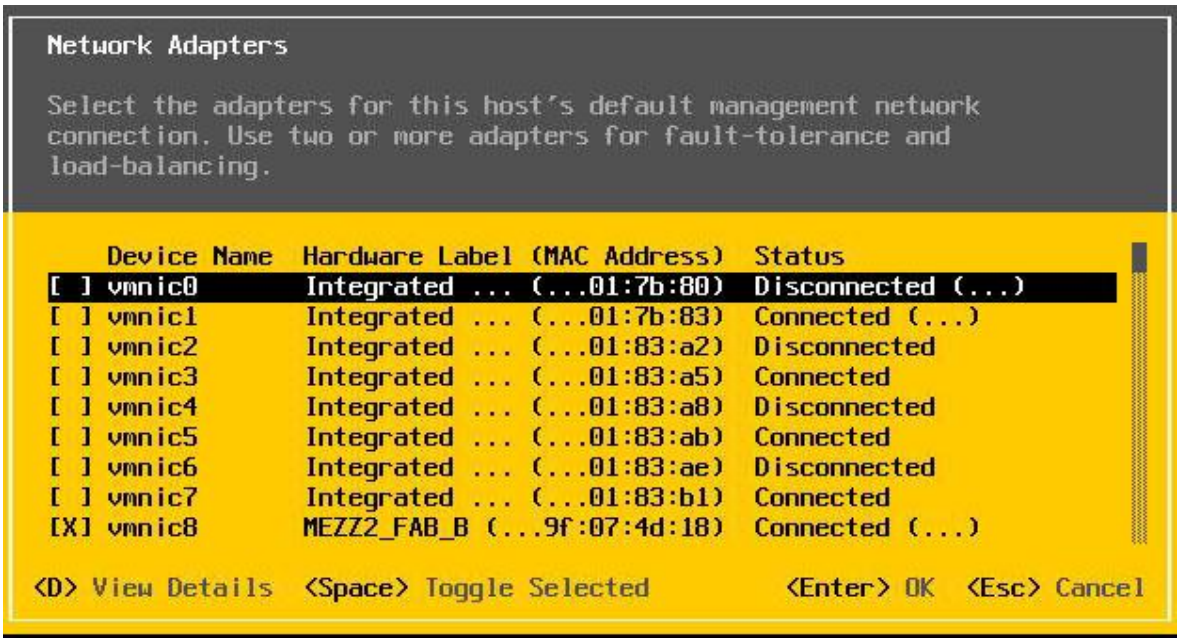

Figure 27 Network Adaptors

11. For BCM57800 and BCM57810 boards, the FCoE boot devices need to have a separate vSwitch other than vSwitch0. This allows DHCP to assign the IP address to the management network rather than to the FCoE boot device. To create a vSwitch for the FCoE boot devices, add the boot device vmnics in vSphere Client under Networking.

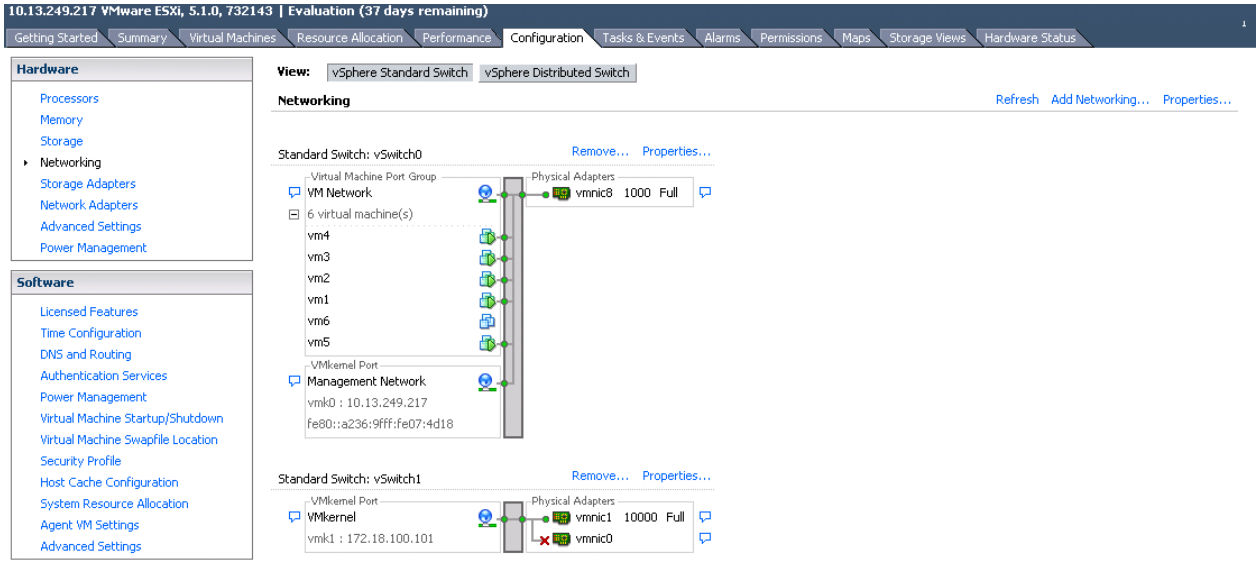

Figure 28 FCoE vSwitch Configuration Tab

Note: ESXi 5.1 has a limitation in that a VLAN ID change for a boot device is not possible. It works only for non-boot devices.

5 Windows 2008 R2 and Windows 2008 SP2 FCoE Boot Installation

Ensure that no USB flash drive is attached before starting the OS installer. The EVBD and OFC/BXFOE drivers need to be loaded during installation. Go through the normal procedures for OS installation. When no disk devices are found, Windows will prompt you to load additional drivers. At this point, connect a USB flash drive containing the full contents of the provided EVBD and OFC boot driver folders. After all appropriate drivers are loaded; setup will show the target disk(s). Disconnect the USB flash drive before selecting the disk for installation.

1. Load the EVBD driver first.

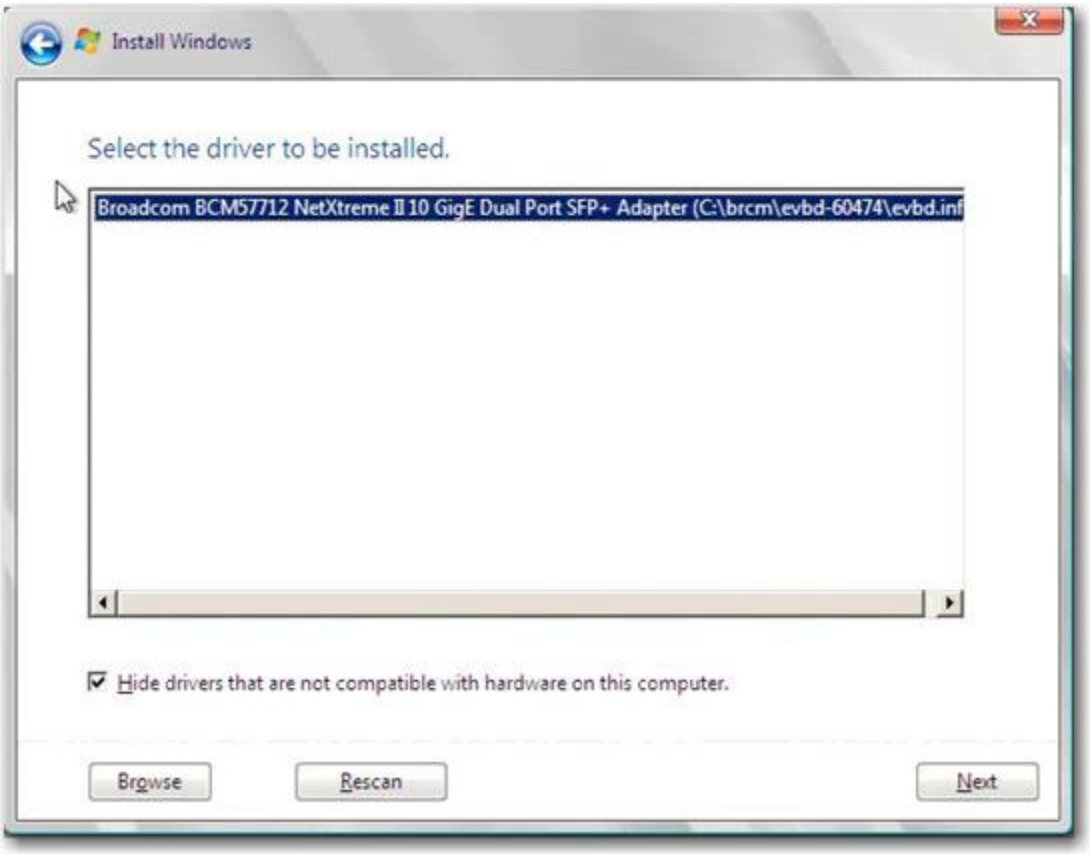

Figure 29 Driver selection

2. Then load the bxfcoe (OFC) driver.

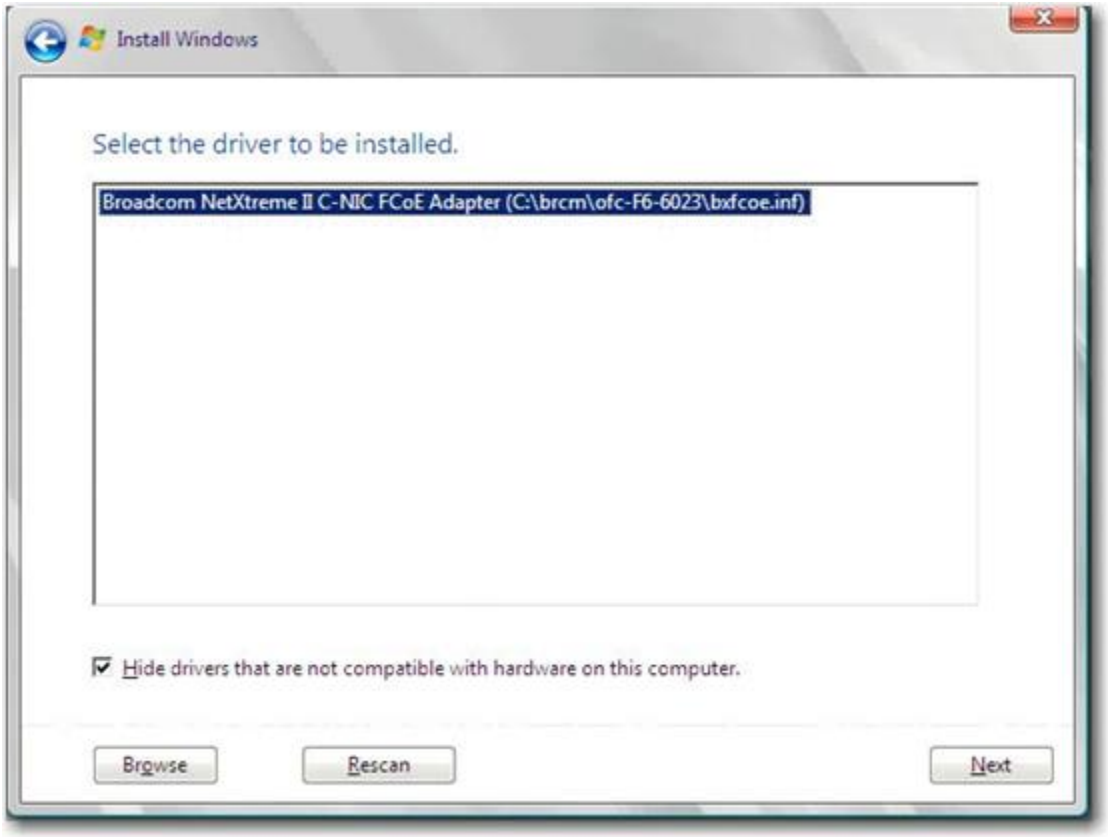

Figure 30 Bxfcoe.inf selection

3. Select the boot LUN to be installed:

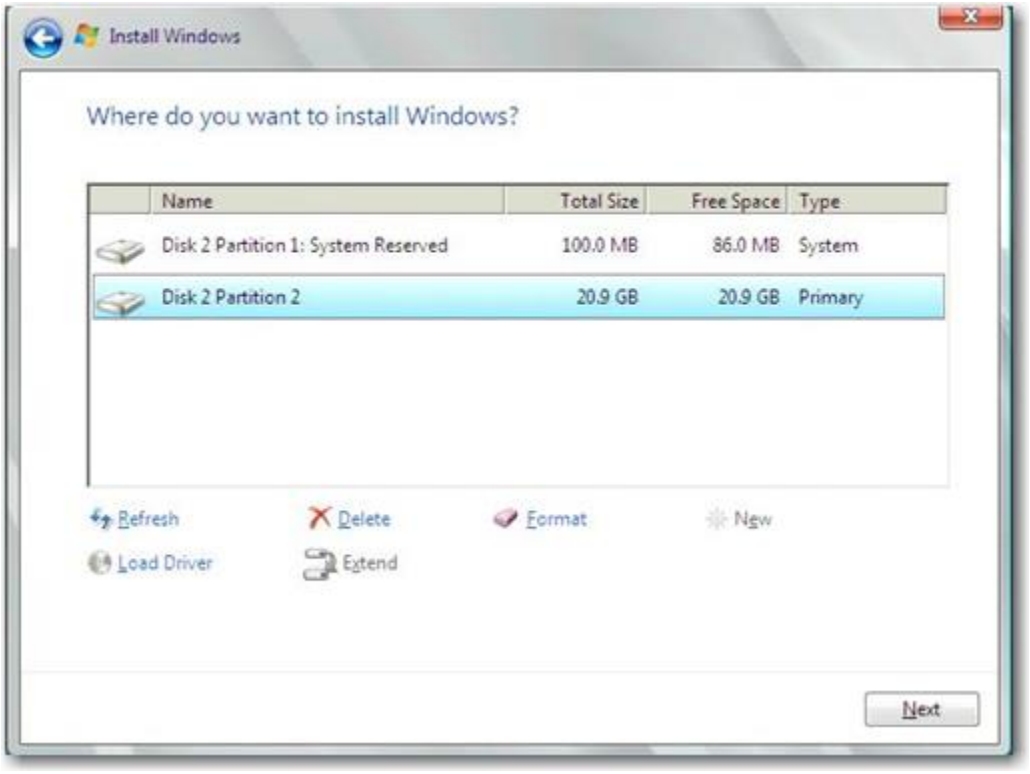

Figure 31 Partition selection

4. Continue with the rest of the installation. Installation should complete normally.

6 FCoE BFS HA Configuration

A High-Availability configuration whether it is FCoE, FC, or iSCSI, requires a mind-set during the design phase of NSPOF (no single point of failure). This NSPOF mind-set is directed towards attaining the “5 9’s” or 99.999% uptime. This 99.999% uptime equivocates to less than 5 minutes of “downtime” per year. Figure 1 shows that a single M620 server with (2) Broadcom 57810 CNA’s are being directed through (2) MXL I/0 modules > (2) S5000 converged switches > (2) Fibre Channel switches > (1) EMC VNX 5300 with (2) ports provisioned for the FCoE traffic. In the configuration being described in this section of the deployment guide, the single point of failure to be addressed is the failure of a Broadcom 57810 CNA.

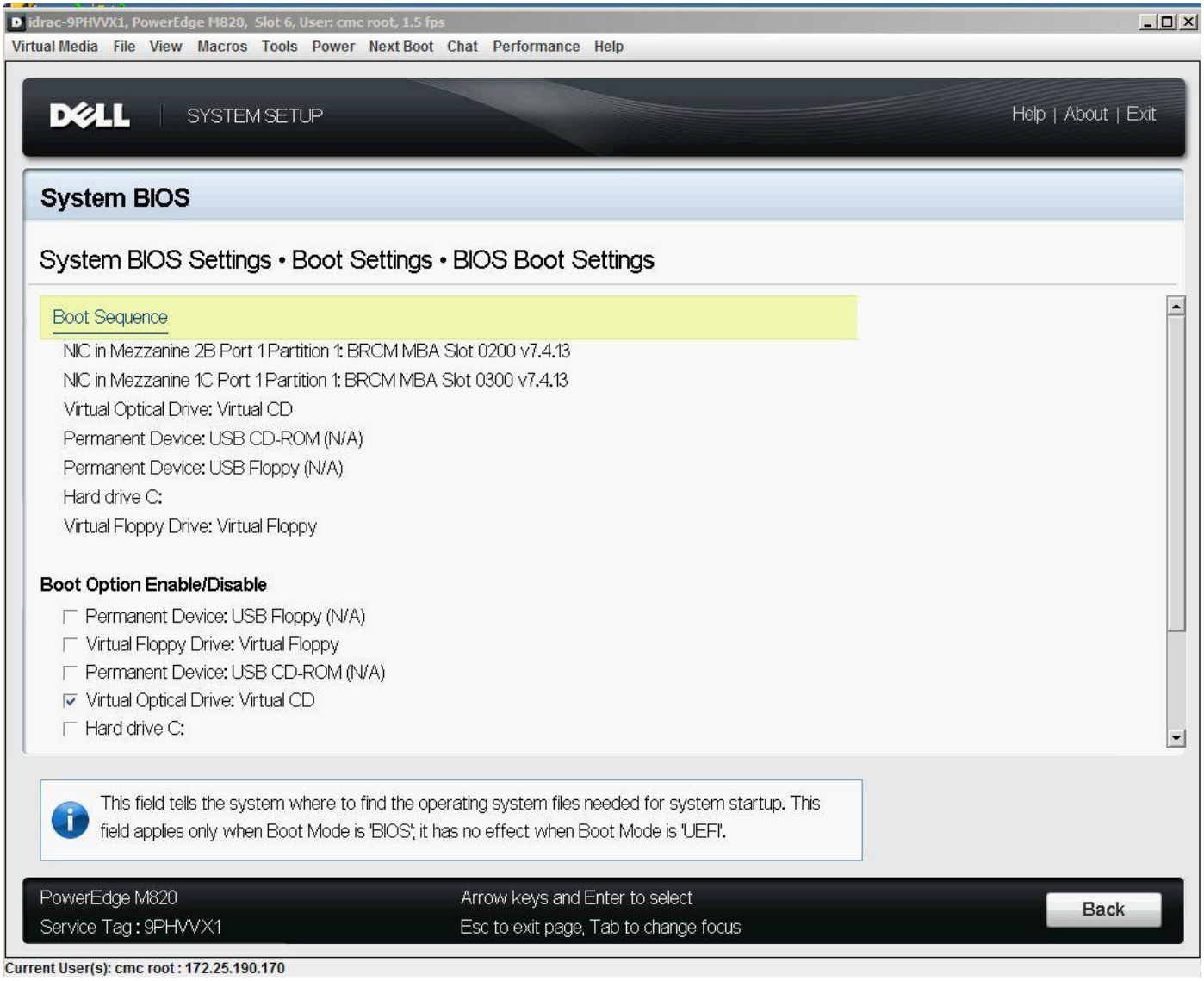

Figure 32 Broadcom 57810 CNA HA Boot Sequence

In the above Boot Sequence image, the first NIC (CNA) is our primary adaptor for booting the OS from the remote LUN. The second NIC in the boot sequence will take over all functions that the first NIC was performing if it should fail. The NIC's in figure 32 have been configured exactly the same.

6.1 VMware Multi-CNA MPIO Confirmation

VMware automatically loads a Multi-Path I/0 (MPIO) plug-in when a primary and secondary boot adaptor are both configured for the same LUN. This section will walk through the confirmation of the automatic MPIO configuration for both Broadcom 57810 CNA's.

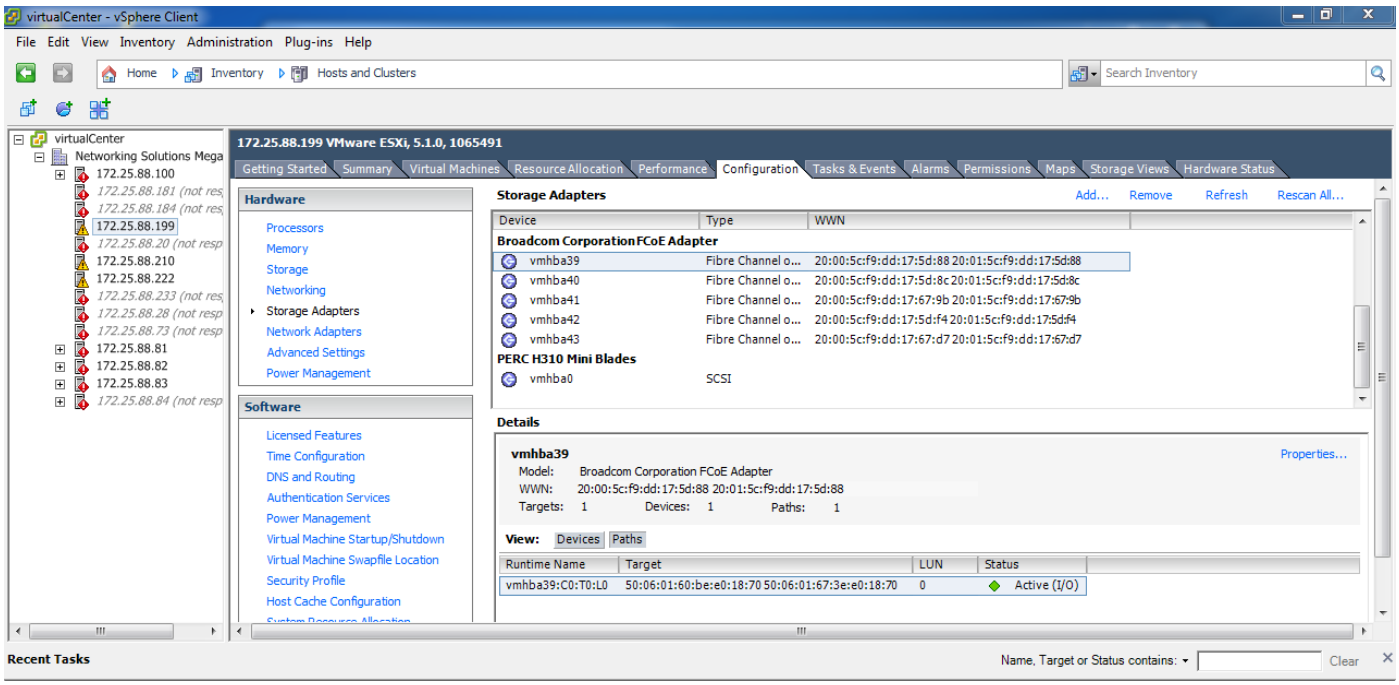

Figure 33 Primary CNA MPIO Details

The primary FCoE BFS CNA can be confirmed in a number of fields. First, under Storage Adaptors, in the WWN field it can be seen that 5D:88 is the last two bytes of the primary FCoE BFS CNA. Second, in the Details section under Status it says "Active (I/0)". This status confirms that 5D:88 is our primary path to the remote LUN.

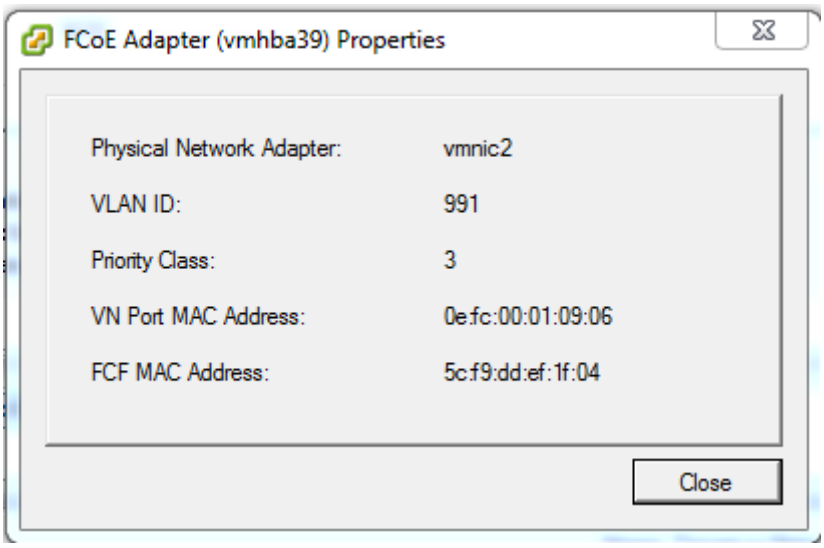

Figure 34 Primary FCoE BFS CNA Properties

This information was gathered by clicking on the Properties link in the right hand corner of the Details section. The properties show that the VLAN id is 991, the Priority Group ID (PGID) is 3, the Fabric Provided MAC Address is 0E:FC:00:01:09:06, and the FCF MAC Address is 5C:F9:DD:EF:1F:04. All of these values can be confirmed on the Primary S5000 with the command “show fip-snooping sessions”.

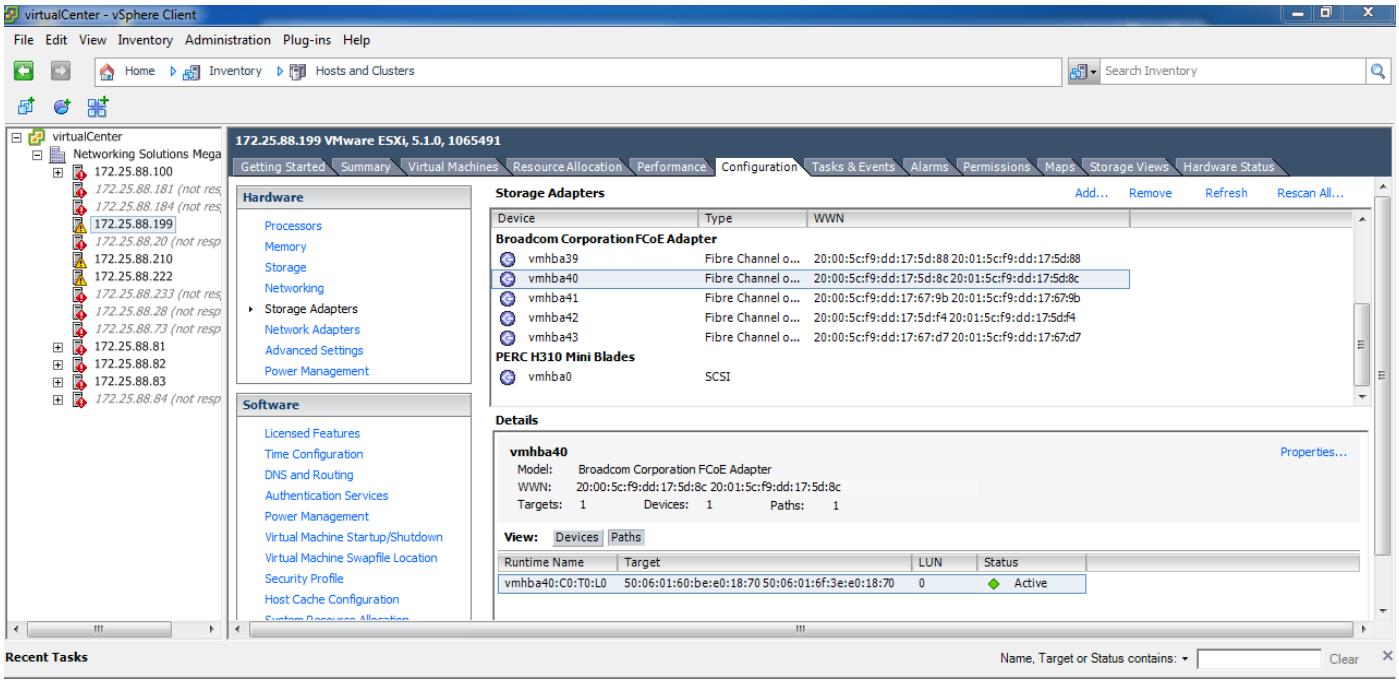

Figure 35 Secondary CNA MPIO Details

The secondary FCoE BFS CNA can be confirmed in a number of fields. First, under Storage Adaptors, in the WWN field it can be seen that 5D:8C is the last two bytes of the secondary FCoE BFS CNA. Second, in the Details section under Status it says “Active”. Since there is no designation of (I/0), this status confirms that 5D:8C is our secondary path to the remote LUN.

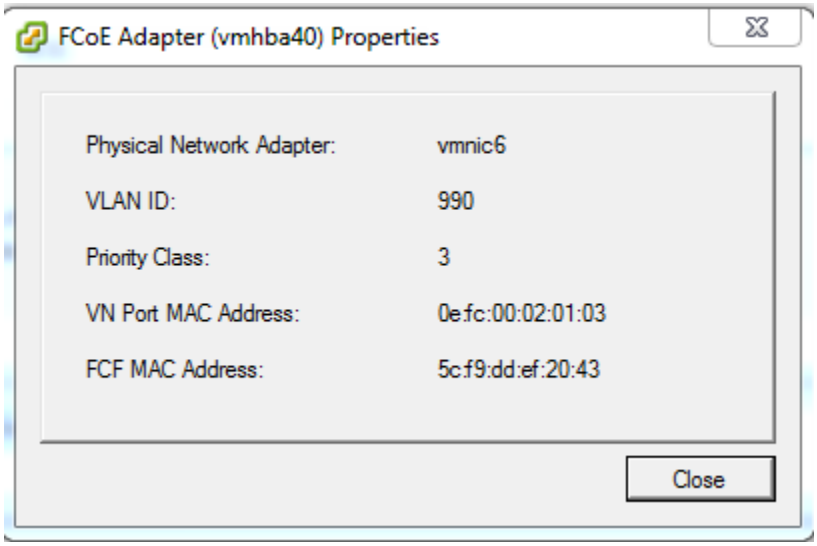

Figure 36 Secondary FCoE BFS CNA Properties

This information was gathered by clicking on the Properties link in the right hand corner of the Details section. The properties show that the VLAN id is 990, the Priority Group ID (PGID) is 3, the Fabric Provided MAC Address is 0E:FC:00:02:01:03, and the FCF MAC Address is 5C:F9:DD:EF:20:43. All of these values can be confirmed on the Secondary S5000 with the command “show fip-snooping sessions”.

The above information has confirmed that we have:

- (2) Broadcom CNA’s with the ability to access the same LUN
- (2) FCF MAC address’s that confirm primary and secondary S5000’s
- (2) Fabric Provided MAC address’s that can be confirmed with the command “show fip-snooping sessions”

6.2 VMware Dual-Port CNA MPIO HA Configuration

There are two steps that need to be taken for the second port of the dual port CNA to participate in an MPIO HA configuration. In this MPIO HA scenario port 0 is connecting to the remote LUN and booting the OS. Port 1 is configured exactly like port 0 and can take over the remote LUN presentation if port 0 fails. In order for port 1 to be visible to VMware's MPIO list of available FCoE resources, multiple esxcli commands will need to be executed. The commands that will be executed are exemplified in the VMware KB articles at the end of this section.

6.2.1 Adding a Broadcom dual port CNA to VMware's SATP Mappings

VMware maintains a list of array types to Storage Array Type Plugin (SATP) mappings.

Because there is no default SATP for FCoE, the default SATP is set to VMW_SATP_LOCAL. The first discovered path to a storage device is mapped, all subsequent discovered paths to the same storage device are unavailable for failover or load balancing. This secondary path unavailability is due to the fact that VMW_SATP_LOCAL is intended for local storage and uses only one path.

The work around for this single path issue is to configure the active/active SATP plug-in VMW_SATP_DEFAULT_AA. The steps to change the SATP are below:

1, Run the following command: exscli storage nmp satp rule add –transport fcoe –satp VMW_SATP_DEFAULT_AA –description "Fibre Channel over Ethernet Devices"

2. Migrate all VM's and put the ESXi host in maintenance mode

3. Restart the ESXi host to ensure that all paths are refreshed with the update SATP

6.2.2 Activating the second port of a Broadcom dual port CNA in VMware

By default VMware will not list or display FCoE capable devices as storage adaptors unless the FCoE port or adaptor has been the primary boot device. In order for the second port of a dual port CNA to be recognized the FCoE port must be activated via the esxcli commands below.

1. esxcli fcoe nic list (this will determine which interfaces are FCoE capable) (Output below)

    vmnic4 (make note of this interface enumeration)
    User Priority: 3
    Source MAC: FF:FF:FF:FF:FF:FF
    Active: false (this interface has not been activated, but, is FCoE capable)
    Priority Settable: false
    Source MAC Settable: false
    VLAN Range Settable: false

2. exscli fcoe nic discover –n vmnic4 (FCoE activation of the second port. Please note that no output is produced with this command)

3. esxcli fcoe adaptor list (interface activation verification command. Output below)

vmhba35
Source MAC: bc:30:5b:01:82:3c
Active: true (the interface has been activated)
FCF MAC: 00:05:73:bf:48:33
VNPort MAC: 0e:fc:00:47:02:24
Physical NIC: vmnic4
User Priority: 3
VLAN id: 2008

To verify proper MPIO management, launch the VSphere client and go to the configure tab and click on Storage Adaptors. The ports will be listed based on their WWPN's with the primary port having the words I/O next to it.

Verification of failover can be completed by rebooting the I/O module that the primary CNA port is going through. Active I/O will transfer to the second port of the CNA.

7 S5000 FCoE Connectivity Verification Commands

During the inital setup of the S5000 switchs FCoE configuration, its essential to understand which commands can give insight into the FCoE functionality.

```
FTOS#show npiv devices

ENode[0]:
ENode MAC    :   14:fe:b5:8e:5f:f9
ENode Intf   :   Po 10
FCF MAC      :   5c:f9:dd:ef:20:43
Fabric Intf  :   Fc 0/0
FCOE Vlan    :   990
Fabric Map   :   fabrica
ENode WWPN   :   20:01:14:fe:b5:8e:5f:f9
ENode WWNN   :   20:00:14:fe:b5:8e:5f:f9
FCoE MAC     :   0e:fc:00:02:01:01
FC-ID        :   02:01:01
LoginMethod  :   FLOGI
Secs         :   80972
Status       :   LOGGED IN
```

Figure 37 Show NPIV Devices

The command “Show NPIV Devices” displays information about the:

- Enode – CNA that is logged in
- FCF – FCoE switch
- FCoE VLAN – VLAN carrying FCoE traffic
- FC-ID – Fibre Channel ID that was assigned to the CNA when it performed a Fabric Login (FLOGI) to the Fibre Channel switch
- Status – Confirms that the CNA is logged in

```
FTOS#
FTOS#
FTOS#
FTOS#show fip-snooping sessions
Enode MAC           Enode Intf     FCF MAC            FCF Intf       VLAN
FCoE MAC            FC-ID     Port WWPN                 Port WWNN

---------           ----------     -------            --------       ----
--------            -----     ---------                 ---------

14:fe:b5:8e:5f:f9   Po 10          5c:f9:dd:ef:20:43   Fc 0/0         990
0e:fc:00:02:01:01   02:01:01  20:01:14:fe:b5:8e:5f:f9   20:00:14:fe:b5:8e:5f:f9
```

Figure 38 Show FIP-Snooping Sessions

The command “Show FIP-Snooping Sessions” can quickly prove whether the CNA has been able to create a virtual link. If the command output is empty the next step is to confirm that:

- LUN Masking and Zoning are correct
- The S5000 configuration flow-chart on page 15 has been followed
- The CNA’s Option ROM BIOS has been setup for FCoE

```
FTOS#show mac-address-table

Codes: *N - VLT Peer Synced MAC
VlanId      Mac Address            Type           Interface         State
 1       14:fe:b5:8e:5c:0d       Dynamic          Po 10             Active
 1       14:fe:b5:8e:5c:75       Dynamic          Po 10             Active
 1       d0:67:e5:c1:68:e7       Dynamic          Po 10             Active
 990     0e:fc:00:02:01:01       Static           Po 10             Active
 990     14:fe:b5:8e:5f:f9       Static           Po 10             Active
```

Figure 39 Show MAC-Address-Table

The command “Show MAC-Address-Table” can be used to determine if the MAC address of the CNA is assigned to the FCoE VLAN and the correct port-channel interface.

```
FTOS#show fip-snooping statistics interface vlan 990
Number of Vlan Requests                                 :0
Number of Vlan Notifications                            :0
Number of Multicast Discovery Solicits                  :59
Number of Unicast Discovery Solicits                    :0
Number of FLOGI                                         :59
Number of FDISC                                         :0
Number of FLOGO                                         :0
Number of Enode Keep Alive                              :31105
Number of VN Port Keep Alive                            :2751
Number of Multicast Discovery Advertisement             :179185
Number of Unicast Discovery Advertisement               :96
Number of FLOGI Accepts                                 :59
Number of FLOGI Rejects                                 :0
Number of FDISC Accepts                                 :0
Number of FDISC Rejects                                 :0
Number of FLOGO Accepts                                 :0
Number of FLOGO Rejects                                 :0
Number of CVL                                           :47
Number of FCF Discovery Timeouts                        :0
Number of Enode Mac Timeouts                            :47
Number of VN Port Session Timeouts                      :0
Number of Session failures due to Hardware Config       :0
```

Figure 40 Show FIP-Snooping Statistics Interface VLAN 990

The command “Show FIP-Snooping Statistics Interface VLAN 990” can show if there were any rejects or failures for the CNA’s using FCoE VLAN 990. The command “Clear FIP-Snooping Statistics” should be run in order to understand if the failures are happening now or occured in the past.

```
FTOS#Show interface status
Port      Description  Status Speed      Duplex Vlan
Fc 0/0                 Up     8000 Mbit  Auto   --
Fc 0/1                 Down   Auto       Auto   --
Fc 0/2                 Down   Auto       Auto   --
Fc 0/3                 Down   Auto       Auto   --
Fc 0/4                 Down   Auto       Auto   --
Fc 0/5                 Down   Auto       Auto   --
Fc 0/6                 Down   Auto       Auto   --
Fc 0/7                 Down   Auto       Auto   --
Fc 0/8                 Down   Auto       Auto   --
Fc 0/9                 Down   Auto       Auto   --
Fc 0/10                Down   Auto       Auto   --
Fc 0/11                Down   Auto       Auto   --
Te 0/12                Down   Auto       Auto   --
Te 0/13                Up     10000 Mbit Full    --
Te 0/14                Down   Auto       Auto   --
Te 0/15                Down   Auto       Auto   --
```

Figure 41 Show Interface Status

The command “Show Interface Status” should show “Up” for the ports that are connected to the CNA and to the FC Switch. If they show “Down” then the next step would be to:

- Trace cables and ensure they are properly connected to the ports that were configured
- Check the S5000 port’s and ensure that “no shut” is part of each ports configuration

```
FTOS#show interfaces tengigabitethernet 0/13 dcbx detail

E-ETS Configuration TLV enabled                 e-ETS Configuration TLV disabled
R-ETS Recommendation TLV enabled                r-ETS Recommendation TLV disabled
P-PFC Configuration TLV enabled                 p-PFC Configuration TLV disabled
F-Application priority for FCOE enabled         f-Application Priority for FCOE disabled
I-Application priority for iSCSI enabled        i-Application Priority for iSCSI disabled
-------------------------------------------------------------------------------

Interface TenGigabitEthernet 0/13
   Remote Mac Address d0:67:e5:c1:68:e7
    Port Role is Manual
   DCBX Operational Status is Enabled
   Is Configuration Source? FALSE
   Local DCBX Compatibility mode is IEEEv2.5
   Local DCBX Configured mode is AUTO
   Peer Operating version is IEEEv2.5
   Local DCBX TLVs Transmitted: ERPFi
   5 Input PFC TLV pkts, 11 Output PFC TLV pkts, 0 Error PFC pkts
   566500 PFC Pause Tx pkts, 106 Pause Rx pkts
   5 Input ETS Conf TLV Pkts, 11 Output ETS Conf TLV Pkts, 0 Error ETS Conf TLV Pkts
   0 Input ETS Reco TLV pkts, 11 Output ETS Reco TLV pkts, 0 Error ETS Reco TLV Pkts
   0 Input Appln Priority TLV pkts, 0 Output Appln Priority TLV pkts, 0 Error Appln Priority TLV Pk
     Total DCBX Frames transmitted 12
     Total DCBX Frames received 5
     Total DCBX Frame errors 0
     Total DCBX Frames unrecognized 0
```

Figure 42 Show Interfaces Tengigabitethernet 0/13 DCBX Detail

The command “Show Interfaces Tengigabitethernet 0/13 DCBX Detail” can show whether:

- DCBX operational status is enabled – If it is not enabled then check if the DCB map’s PFC and ETS parameters have been created correctly to.
- Peer Operating version – ensure that both Peer Operating version and DCBX compaibility mode match. If there is a mis-match then change the DCBX version on the switch that the CNA supports.
- Local DCBX TLV’s Transmitted – Confirm that “ERPF” are uppercase which confirms they have been transmited. If “erpf” are lowercase then check that the port on the switch that the CNA is connected to is “Up”. Check that the CNA could see the remote LUN during the servers POST routine.

In summary the above commands are but a few that can be used to interpret whether an out of boundary condition has occured or that the FCoE session is healthy.

A Terminology

BFS
Boot from SAN

Bridging
Bridging—commonly called switching—is frame-by-frame layer-2 forwarding of Ethernet traffic with forwarding decisions generally based on each frame's source and destination MAC address. A simple bridge has three available actions to perform on received traffic—filter, forward, or flood. An Ethernet bridge is very simple and relies on bridging protocols like spanning-tree to avoid sending data frames in a loop, causing a storm.

Channel Group
See LAG.

CLI
Command Line Interface (CLI) is the text-based console interface that is used for entering management and configuration commands into devices like the Dell Force10 MXL switch. The MXL's CLI can be accessed via telnet, SSH, an externally-accessible serial connection, and also from the CMC's CLI.

CMC
Chassis Management Controller (CMC) is the embedded management interface of the Dell PowerEdge™ M1000e blade server chassis. Among other functions, the CMC provides network and console access to installed IO modules including the Dell Force10 MXL switch.

E_port
This is a general fibre channel term. An E_port or sometimes referred to as an ISL port is an expansion port or inner switch link between FC switches.

ETS
Enhanced Transmission Selection (ETS) is defined in the IEEE 802.1Qaz standard (IEEE, 2011). ETS supports allocation of bandwidth amongst traffic classes. It then allows for sharing of bandwidth when a particular traffic class doesn't fully utilize the allocated bandwidth. The management of the bandwidth allocations is done with bandwidth-allocation priorities which coexist with strict priorities (IEEE, 2011).

F_Port
This is a general fibre channel term which relates to the FC switch devices ports. F_port's on an FC switch will connect via cable to an N_port on an end device (HBA). The F_port on an FC switch will provide the N_port it is connected with access to fabric services, such as Fabric Login Server, Fabric Name Server, etc.

FIP-Snooping
With FIP-Snooping enabled on the Dell products, FIP logins, solicitations, and advertisements are monitored. In this monitoring or snooping process the switch gathers information pertaining to the ENode and FCF addresses. With this information the switch will then place filters that only allow access to ENode devices that have logged-in successfully. This enables the FCoE VLAN to deny all other traffic except this lossless FCoE storage traffic.

The filtering process also secures the end-to-end path between the ENode device and the FCF. The ENode will only be able to talk with the FCF in which it has logged into.

FIP-Snooping bridge (FSB)
With a switch configured to performing FIP-Snooping the industry term for this switch is FSB or FIP-Snooping bridge. It is performing FIP-Snooping as described in the previous term.

FCF
FCoE forwarders (FCFs) act as an Ethernet and FC switch combined. All typical termination functions that would occur on a FC switch occur on the FCF. FCF's give VF_Ports and VE_Ports for their virtual FC interfaces.

Filter, Flood, Forward
Data frames received by Ethernet switches may be filtered (meaning discarded or dropped) according to defined behavior, automatic protocols, or administrative configuration; flooded (meaning sent out all other links) if the data frame's destination MAC address is unknown; or forwarded (meaning sent to one other link) if the destination MAC address is already learned. A destination for a MAC address may be manually configured on a link's interface; but, most MAC address destinations are learned dynamically based on source MAC addresses of already received data frames—when a bridge receives a data frame, it remembers the source MAC address of the frame on the received link for five minutes and will then selectively forward data frames destined for that MAC address over that link as a learned destination.

IOM
IO module (IOM) refers to the modules at the rear of the Dell PowerEdge M1000e chassis that will receive and transmit I/O (Ethernet, FC, Infiniband, etc.) from the blade servers located at the front of the chassis. The Dell Force10 MXL switch is as an IOM for the M1000e blade server chassis.

LACP
Link Aggregation Control Protocol (LACP) is the protocol used to ensure that the multiple links in a LAG do not form loops due to misconfiguration or device misbehavior. It is recommended practice to always use LACP on configured LAGs.

LAG
Link Aggregation Group (LAG) is a configured bundle of Ethernet links that are treated as the same logical Ethernet link. There are multiple terms that apply to LAGs including channel group, port channel, trunk, and even some server Ethernet interface teaming involves a collection of links that would be considered a LAG.

LAN
A Local Area Network (LAN) is a term for a network that services a limited area from the size of a single table to a large as an office building. They are generally interconnected using Ethernet switches. The term is sometimes applied to a network involving a single broadcast domain and sometimes applied to a network involving multiple broadcast domains separated into VLANs (and often rejoined via routing).

Link
Link is a term in networking that refers to a connection made between two nodes in a network. In Ethernet networking it is generally used to refer to a direct connection between two ports.

MAC Address
Media Access Control Address (MAC Address) is a layer-2 node identifier. In Ethernet bridging, MAC addresses are used for source and destination identification. They can also be used as system identifiers since vendor-assigned (or burned-in) MAC addresses are globally unique. An Ethernet MAC address is 48 bits long and generally written in groupings of two hexadecimal digits often separated by colons or hyphens like this: 00:1e:c9:00:cb:01.

MLAG
Multi-chassis Link Aggregation Group (MLAG) is a LAG implementation that connects across multiple switches on one side of the logical aggregated link. This requires shared management of the logical link between the switches sharing a side and situational forwarding of frames that differs from the forwarding that would be present across the multiple LAGs that an MLAG replaces. There is no standard method for implementing an MLAG but multiple vendor specific methods. Nexus vPC and Force10 VLT are examples of MLAG implementations.

MSTP
Multiple Spanning-Tree Protocol (MSTP) is a standards-based modified version of the rapid spanning tree protocol that carries multiple spanning tree instances within its rapid spanning-tree protocol packet. Spanning-tree instance 0 is assigned to the common rapid spanning-tree instance and additional instances above 0 may be configured. For each spanning-tree instance, a root switch is elected and unique active and backup links can be chosen providing potentially unique traffic paths on the network per instance. Each VLAN can then be assigned to a spanning-tree instance allowing active traffic on separate VLANs to potentially utilize separate paths across the network. In order for interconnected switches to participate together in more than the common spanning-tree instance, they must have the same MSTP configuration (a checksum of this configuration is included the spanning-tree protocol packet and must be the same between switches in order for them to participate). MSTP is originally defined in the IEEE 802.1s standard and is included in 802.1q IEEE Virtual LANs standard.

N_Port
This is a general fibre channel term which relates to the end device, typically the Server/Storage HBA. The N_port is usually referring to the port connection on this end device. The other end of the connection will typically be a F_port on an FC switch.

Out-of-Band
An out-of-band interface provides management connectivity to a device without participating in or relying on a device's in-band (normal-use) data interfaces. An out-of-band interface does not send or receive traffic from the switched links—neither bridged nor routed. Common out-of-band interface types are Ethernet and serial console—often both are presented with RJ-45 (8P8C) connectors although on IO modules in the Dell PowerEdge M1000e chassis the serial connector is sometimes a physical USB type-A port requiring a special cable.

NPIV
N-port identifier virtualization which enables multiple N-port fabric logins at the same time on the same physical FC link (Cisco Systems, Inc., 2011).This term is in reference to the Cisco Nexus 5000 series switches implementation of NPIV. NPIV must be enabled to share multiple logins across a single port/link or a port-channel/multiple-line link.

NPV
N-port virtualizer is a FC aggregation method which passes traffic through to end devices, while eliminating the need to use a domain ID for this device (Cisco Systems, Inc., 2011). This term is used in reference to configuration settings on the Cisco Nexus 5000 series switches.

PFC
Priority Flow Control (PFC), or Per-Priority Pause is defined in the IEEE 802.1Qbb standard. PFC is flow control based on priority settings and adds additional information to the standard pause frame. The additional fields added to the pause frame allow devices to pause traffic on a specific priority instead of pausing all traffic. (IEEE, 2009) Pause frames will be initiated by the FCF in most cases when its receive buffers are starting to reach a congested point. With PFC traffic is paused instead of dropped and retransmitted. This provides the lossless network behavior necessary for FC packets to be encapsulated and passed along the Ethernet paths.

Port Channel
See LAG.

PVST
Per-VLAN Spanning-Tree (PVST) is a vendor specific implantation of the spanning tree protocol that maintains separate instances for each VLAN, passing that instance's protocol frames within the VLAN it manages. This method simplifies deployment of multiple VLANs and is popularly used where available.

RSTP
Rapid Spanning-Tree Protocol (RSTP) is a standards-based modified version of the basic spanning tree protocol that allows for much faster convergence times of spanning tree instances and provides for special administratively-assigned port states that improve behavior in certain circumstances. RSTP is originally defined in the IEEE 802.1w standard and is included in 802.1d IEEE Ethernet bridging standard.

Spanning Tree
Spanning Tree refers to a family of layer-2 management protocols used by Ethernet bridges to establish a loop-free forwarding topology. At layer-2, Ethernet is a very simple technology that without intervening protocols or configuration can easily forward traffic in endless loops—see Bridging for an explanation. The spanning-tree protocols provide a standard way for an interconnected set of Ethernet bridging devices to only use links that will not cause traffic flows to loop (Ethernet switches—being very fast bridges—can forward traffic very quickly and looping traffic flows will rapidly saturate all available bandwidth with unwanted, repeated traffic). In a spanning tree, a single bridge is elected the root bridge—either by lowest assigned priority or by having the lowest of the presented MAC addresses. Once a root bridge is elected, every other bridge keeps its one link with the lowest path cost to the root bridge active. Links with redundant paths are then blocked by switches that don't have the lowest path cost for that link. As a result, with spanning tree each non-root bridge effectively has only one active link between it and the root bridge and the topology of these unblocked links draws a tree to the root bridge. Spanning tree has a long and involved history on Ethernet and there are many different implementations with different timings and protocol feature sets—see RSTP, MSTP, and PVST.

Storm
Ethernet Storms—often called broadcast storms—are a descriptive term for excessive (and generally unwanted) data frames being continuously sent to all available links. Because Ethernet bridges dynamically learn destinations based on recently received traffic, the switches do not limit the number of times that data frames can be forwarded across bridges, and do not recognize when data frames have been forwarded back to the same bridge; without the oversight provided by bridging management protocols, looped flows of data frames can quickly cause a storm. A storm can quickly saturate all available bandwidth—catastrophically affecting a network's performance and reliability. To keep potential storms from happening, Ethernet bridging generally relies on spanning-tree protocols to establish loop-free topologies and MAC learning to keep traffic selectively forwarded.

STP
Spanning-Tree Protocol (STP)—see Spanning Tree.

Switching
Switching, in an Ethernet context, is a specific technology but the term has largely been generalized to mean layer-2 Ethernet bridging. Ethernet switching is the use of ASIC technologies to implement Ethernet traffic forwarding and filtering in specialized circuits and memory structures designed for high throughput, low latency, and low cost performance. Generally an Ethernet switch will at least function as a layer-2 bridge but more advanced models have multilayer capabilities including layer-3 routing and multilayer filtering, logic, and frame modification.

Switchport
Switchport is a configuration term used to denote an Ethernet switch's link interface that is configured for layer-2 bridging (participating in one or more VLANs).

ToR
Top of Rack (ToR) is a term for a switch that is actually positioned at the top of a server rack in a data center.

Trunk
Trunk is an ambiguous term in Ethernet networking that can apply to a LAG—a group of multiple links acting as one or to a switchport interface of an Ethernet switch configured in trunk mode to pass multiple VLANs across the one link.

VLAN
Virtual Local Area Network (VLAN) is a single layer-2 network (also called a broadcast domain as broadcast traffic does not escape a VLAN on its own). Multiple VLANs can be passed between switches using switchport trunk interfaces. When passed across trunk links, frames in a VLAN are prefixed with the number of the VLAN to which they belong —a twelve bit value that allows just over 4000 differently numbered VLANs.

VSAN
Virtual SAN is a logical partitioning of physical connections to provide for fabric or SAN separation. VSAN is a term that is particular to the Cisco Nexus series switches.

WWPN
WWPN refers to Worldwide Port_Name. WWPN is very similar to the burned-in MAC address that exists on Ethernet devices but typically applies to FC or FCoE devices. Each FC HBA or FCoE CNA has a unique WWPN that will used in the FC infrastructure for unique identification, zoning, and LUN masking.

Zones
In the FC environment zones are set up as boundaries which allow connections between specific devices. This connection is usually an end device such as an HBA that can connect to a specific LUN on a storage device.

B Additional resources

Support.dell.com is focused on meeting your needs with proven services and support.

DellTechCenter.com is an IT Community where you can connect with Dell EMC Customers and Dell EMC employees for the purpose of sharing knowledge, best practices, and information about Dell EMC products and installations.

Referenced or recommended Dell EMC publications:

- Dell VMware Troubleshooting Guide http://www.dell.com/support/troubleshooting/us/en/19/KCS/KcsArticles/ArticleView?c=us&l=en&s=dhs&docid=597082

Referenced or recommended VMware publications:

- http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=2032363