Best Practices for Configuring Data Center Bridging with Windows Server and EqualLogic Storage Arrays

Dell Storage Engineering
January 2014

A Dell Reference Architecture

Revisions

Date	Description
January 2014	Initial release

Table of contents

Acknowledgements

This best practice white paper was produced by the following members of the Dell Storage team:

Engineering: Nirav Shah

Technical Marketing: Guy Westbrook

Editing: Camille Daily

Additional contributors: Mike Kosacek

Feedback

We encourage readers of this publication to provide feedback on the quality and usefulness of this information by sending an email to SISfeedback@Dell.com.

SISfeedback@Dell.com

Executive summary

Data Center Bridging (DCB) is supported by 10GbE Dell EqualLogic storage arrays with firmware 5.1 and later. DCB can be implemented on the host server side by using specialized hardware network adapters or a DCB software feature (as Windows Server 2012 does) and standard Ethernet adapters. This paper outlines the best practice and configuration steps that provide a supportable DCB implementation specific to EqualLogic arrays and the Microsoft Windows operating systems (Windows 2008 R2, Windows Server 2012, and Windows Hyper-V).

1 Introduction

EqualLogic arrays began supporting DCB Ethernet standards with array firmware version 5.1. DCB unifies the communications infrastructure for the data center and supports a design that utilizes a single network with both LAN and SAN traffic traversing the same fabric. Prior to the introduction of DCB, best practices dictated that an Ethernet infrastructure (Server NICs and switches) for SAN traffic was created separate from the infrastructure required for LAN traffic. DCB enables administrators to converge SAN and LAN traffic on the same physical infrastructure.

When sharing the same Ethernet infrastructure, SAN traffic must be guaranteed a percentage of bandwidth to provide consistent performance and ensure delivery of critical storage data. DCB enables this bandwidth guarantee when the same physical network infrastructure is shared between SAN and other traffic types.

The white paper will provide test-based results showing the benefits of DCB in a converged network as well as best practice and configuration information to help network and storage administrators deploy supportable DCB implementations with EqualLogic arrays and Windows 2008 R2, Windows Server 2012 and Hyper-V 3.0 host servers.

1.1 Terminology

Broadcom Advance Controller Suite: (BACS) A Windows management application for Broadcom network adapters.

Converged Network Adapter: (CNA) Combines the function of a SAN host bus adapter (HBA) with a general-purpose network adapter (NIC).

Data Center Bridging: (DCB) A set of enhancements of the IEEE 802.1 bridge specifications for supporting multiple protocols and applications in the data center. It supports converged infrastructure implementations for multiple traffic types on a single physical infrastructure.

Host Bus Adapter: (HBA) A dedicated interface that connects a host system to a storage network.

Link aggregation group: (LAG) Multiple switch ports configured to act as a single high-bandwidth connection to another switch. Unlike a stack, each individual switch must still be administered separately and functions separately.

NIC Partitioning: (NPAR) Ability to separate, or partition, one physical adapter port into multiple simulated adapter port instances within the host system.

VHDX: File format for a Virtual Hard Disk in a Windows Hyper-V 2012 hypervisor environment.

2 DCB overview

Network infrastructure is one of the key resources in a data center environment. It interconnects various devices within and provides connectivity to network clients outside the data center. Network infrastructure provides communication between components hosting enterprise applications and those hosting application data. Many enterprises prefer to have a dedicated network to protect performance sensitive iSCSI traffic from being affected by general Local Area Network (LAN) traffic. DCB provides a way for multiple traffic types to share the same network infrastructure by guaranteeing a minimum bandwidth to specific traffic types.

For more information on DCB, refer to *Data Center Bridging: Standards, Behavioral Requirements, and Configuration Guidelines with Dell EqualLogic iSCSI SANs* at:
http://en.community.dell.com/techcenter/storage/w/wiki/4396.data-center-bridging-standards-behavioral-requirements-and-configuration-guidelines-by-sis.aspx.

and *EqualLogic DCB Configuration Best Practices* at:

http://en.community.dell.com/techcenter/storage/w/wiki/4355.configuring-dcb-with-equallogic-sans.aspx.

2.1 DCB standards

DCB is a set of enhancements made to the IEEE 802.1 bridge specifications for supporting multiple protocols and applications in the data center. It is made up of several IEEE standards as explained below.

Enhanced Transmission Selection (ETS)

ETS allows allocating bandwidth to different traffic classes so that the total link bandwidth is shared among multiple traffic types. The network traffic type is classified using a priority value (0-7) in the VLAN tag of the Ethernet frame. iSCSI traffic by default has a priority value of 4 in the VLAN tag. Likewise, the testing presented in this paper allocated the required SAN bandwidth to iSCSI traffic with a priority value of 4.

Priority-based Flow Control (PFC)

PFC provides the ability to pause traffic types individually depending on their tagged priority values. PFC helps avoid dropping packets during congestion, resulting in a lossless behavior for a particular traffic type. For the network configuration used in this paper, iSCSI traffic was considered crucial and configured to be lossless using PFC settings.

Data Center Bridging Exchange (DCBX)

DCBX is the protocol that enables two peer ports across a link to exchange and discover ETS, PFC, and other DCB configuration parameters allowing for remote configuration and mismatch detection. The willing mode feature allows remote configuration of DCB parameters. When a port is in willing mode it accepts the advertised configuration of the peer port as its local configuration. Most server CNAs and storage ports are in willing mode by default. This enables the network switches to be the source of the

DCB configuration. This model of DCB operation with end devices in willing mode is recommended to minimize mismatches.

iSCSI application TLV

The iSCSI application Type-Length Value (TLV) advertises a capability to support the iSCSI protocol.

ETS, PFC, DCBX and the iSCSI TLV are all required for DCB operation in an EqualLogic storage environment.

Note: Devices claiming DCB support might not support all of the DCB standards, verify the support of each DCB standard. The minimum DCB requirements for EqualLogic iSCSI storage are ETS, PFC and Application priority (iSCSI application TLV).

3 iSCSI initiator types

Windows 2008 and Windows Server 2012 support both Microsoft iSCSI software initiator and hardware iSCSI initiators. This section explains the different iSCSI initiators and applicable DCB implementation.

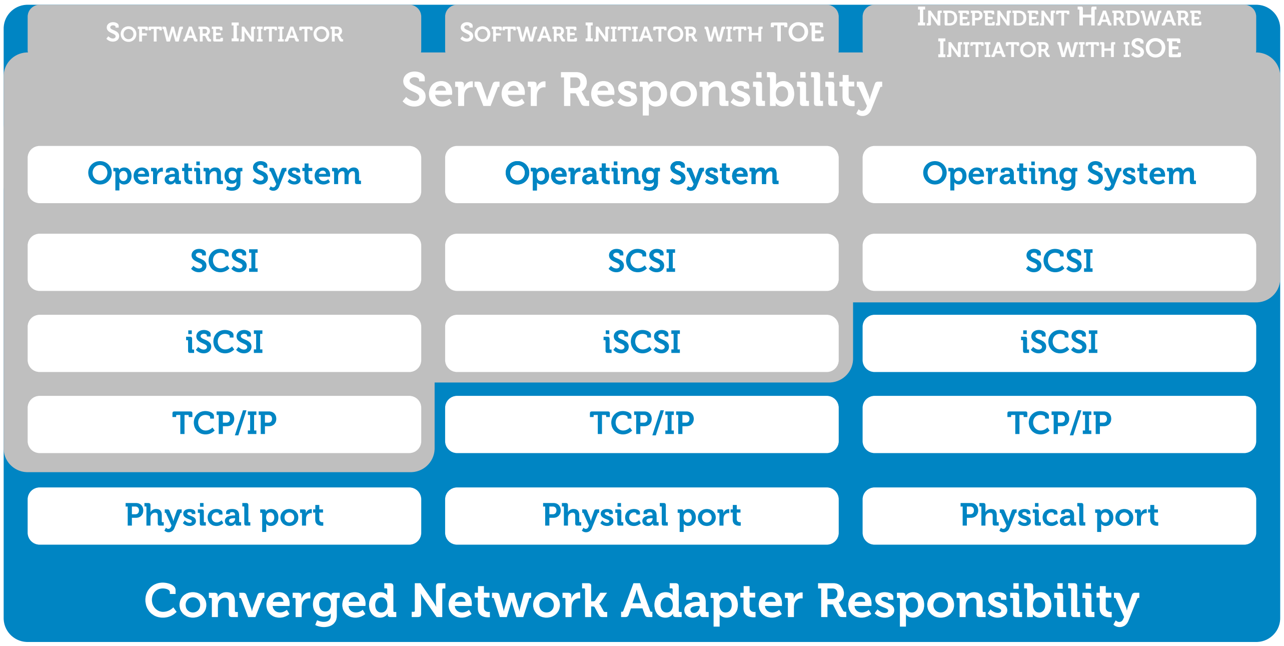


Figure 1 iSCSI initiator types

Software iSCSI initiator

Microsoft Windows Server 2008 R2 and Windows Server 2012 both support the Microsoft iSCSI Initiator. A kernel-mode iSCSI driver moves data between the storage stack and the TCP/IP stack in a Windows operating system. With the Microsoft DCB feature available in Windows Server 2012, DCB can be implemented for iSCSI traffic.

Note: Software controlled DCB for iSCSI is only possible when Windows 2012 is responsible for the entire TCP/IP stack and offload engines (e.g. TOE and iSOE) are not used on the network adapter.

Software iSCSI initiator with TCP Offload Engine

TCP Offload Engine (TOE) is an additional capability provided by specific Ethernet network adapters. These network adapters have dedicated hardware that is capable of processing the TCP/IP stack and offloading it from the host CPU. The OS is only responsible for other network layers above the TCP network layer. This type of network adapter uses the Microsoft software iSCSI Initiator. Software controlled DCB is not available when utilizing an Ethernet interface with the TCP Offload Engine enabled.

Independent hardware iSCSI initiator with iSOE

Network adapters with iSCSI Offload Engine, where the SCSI protocol is implemented in hardware, are available from various manufactures. The Broadcom 57810 converged network adapter (CNA) can act as a hardware iSCSI initiator. Broadcom 57810 Network adapters are capable of implementing DCB for iSCSI traffic. In this case, the operating system was not responsible for implementing DCB or the iSCSI network stack since both were handled by the CNA hardware.

Hardware iSCSI initiators offload the entire iSCSI network stack and DCB which helps to reduce host CPU utilization and may increase performance of the system. Although hardware iSCSI initiators may have a higher cost than standard network adapters that do not have offload capabilities. Typically, hardware iSCSI initiators have better performance.

Note: Windows 2008 R2, unlike Windows Server 2012, does not have a software DCB feature. A DCB capable hardware network adapter is required for DCB implementation on Windows 2008 R2.

4 Broadcom switch independent NIC partitioning

Broadcom 57810 Network adapters are capable of switch independent NIC partitioning (NPAR). This means a single, dual-port network adapter can be partitioned into multiple virtual Ethernet, iSCSI or FCoE adapters. Broadcom NPAR simultaneously supports up to eight (four per port) virtual Ethernet adapters and four (two per port) virtual Host Bus Adapters (HBAs). In addition, these network adapters have configurable weight and maximum bandwidth allocation for traffic shaping and Quality of Service (QoS) control.

NPAR allows configuration of a single 10 Gb port to represent up to four separate functions or partitions. Each partition appears to the operating system or the hypervisor as a discrete NIC with its own driver software functioning independently. Dedicating one of the independent partitions to iSCSI traffic and the other for LAN Ethernet traffic allows the Broadcom CNA to manage the DCB parameters and provide lossless iSCSI traffic. The tests presented in this white paper configured one partition on each physical port dedicated to iSCSI with the iSCSI Offload Engine function enabled (Function 0/1). Another partition on each physical port (Function 2/3 and Function 4/5) was used for Ethernet (NDIS) LAN as shown in Figure 2. VLANs were created on each partition and trunk mode enabled at the physical port. This provided a dedicated virtual LAN for each traffic type. Detailed configuration steps for configuring NPAR is provided in Appendix D.

Note: Using dedicated virtual LANs for different traffic types only provides a logical separation to the network. Unlike DCB, it does not provide guaranteed bandwidth and lossless behavior of traffic and therefore it is not recommended to **only** use VLANs when mixing traffic types on the same switch fabric.

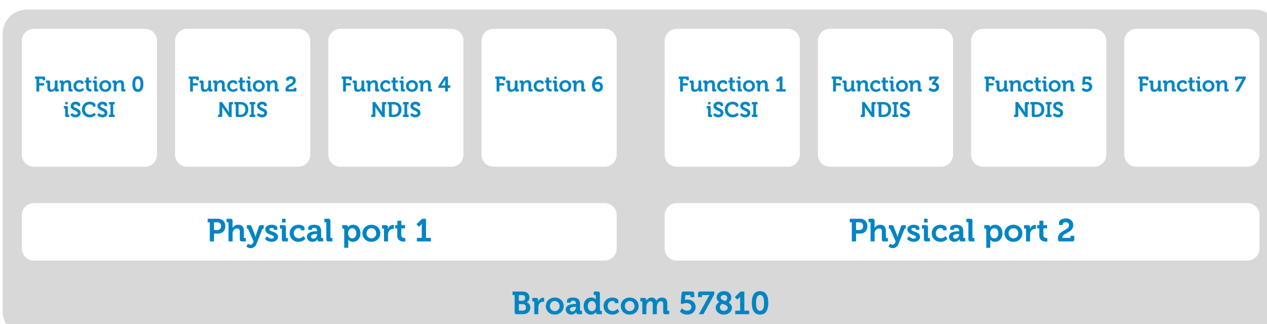


Figure 2 Broadcom Switch Independent Partitioning (This diagram is used as an example only)

5 Hardware test configuration

The presented test configuration consisted of Dell PowerEdge M620 blade servers, Dell Networking MXL blade switches and Dell EqualLogic PS6110 storage arrays. The blade servers and switches were housed in a Dell PowerEdge M1000e blade chassis. M620 servers were installed with Broadcom 57810 network adapters which were internally connected to the MXL blade switch on Fabric A of the M1000e blade chassis. The EqualLogic PS6110 storage arrays were connected to the MXL blade switches using the external 10GbE ports. Figure 3 illustrates this hardware configuration.

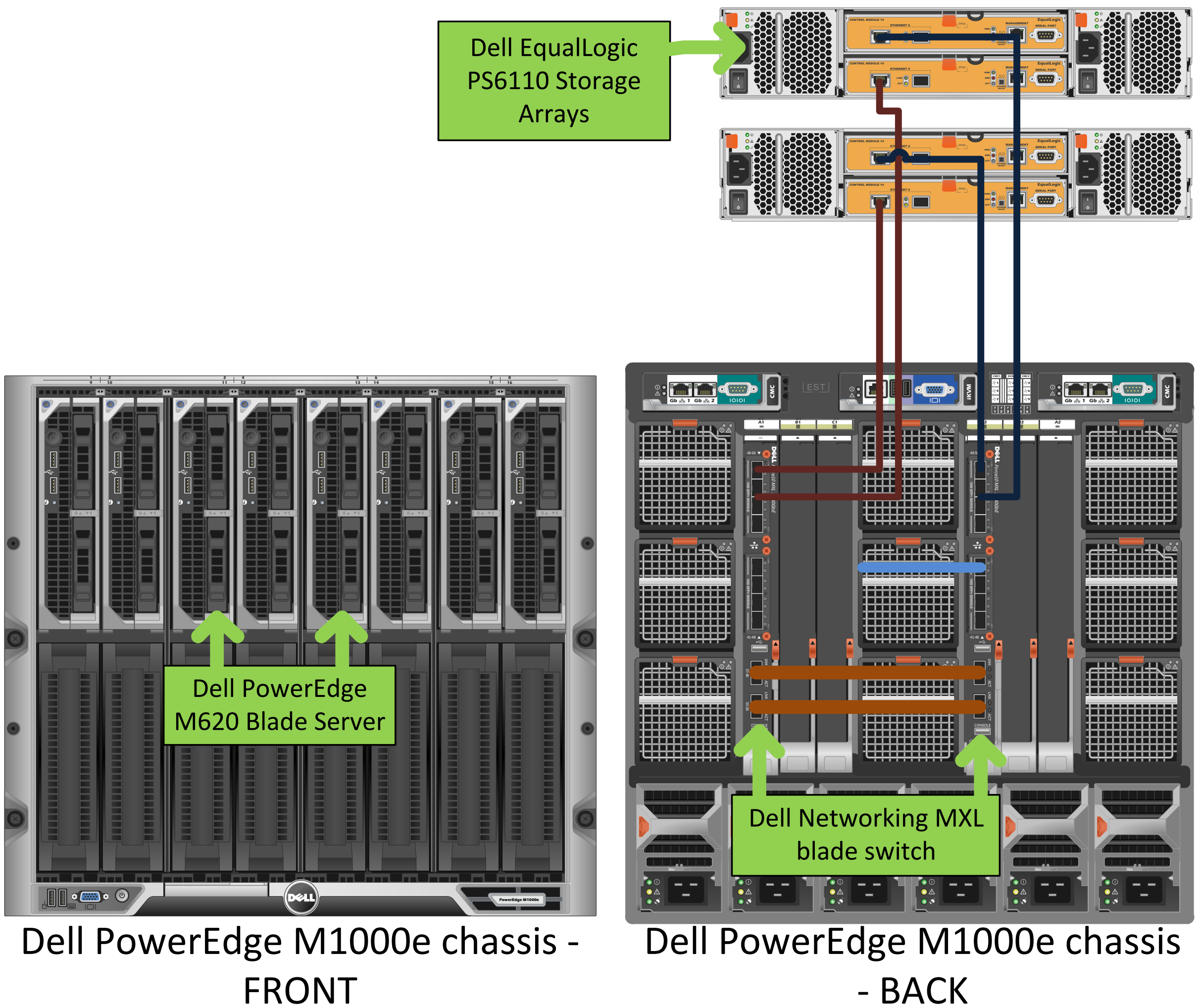


Figure 3 Test setup configuration

5.1 Dell Networking MXL Blade switch configuration.

The Dell Networking MXL switch is a modular switch that is compatible with the PowerEdge M1000e blade chassis and provides 10, and 40 GbE ports to address the diverse needs of most network environments. The MXL switch supports 32 internal 10 GbE ports, as well as two fixed 40 GbE QSFP+ ports and offers two bays for optional FlexIO modules. FlexIO modules can be added as needed to provide additional 10 or 40 GbE ports.

For additional information on the Dell Networking MXL switch, refer to the manuals and documents at: http://www.dell.com/support/Manuals/us/en/04/Product/force10-mxl-blade

The Dell Networking MXL blade switch is DCB capable, enabling it to carry multiple traffic classes. It also allows lossless traffic class configuration for iSCSI. The test environment had DCB enabled with the iSCSI traffic class configured in a lossless priority group to ensure priority of iSCSI SAN traffic. LAN and other types of traffic were in a separate (lossy) priority group. For these tests, ETS was configured to provide a minimum of 50% bandwidth for iSCSI traffic and 50% for the other (LAN) traffic type.

Two blade switches were used in Fabric A of the PowerEdge M1000e chassis for High Availability (HA). The two Dell Networking MXL switches were connected with a LAG using the two external 40GbE ports to enable communication and traffic between them.

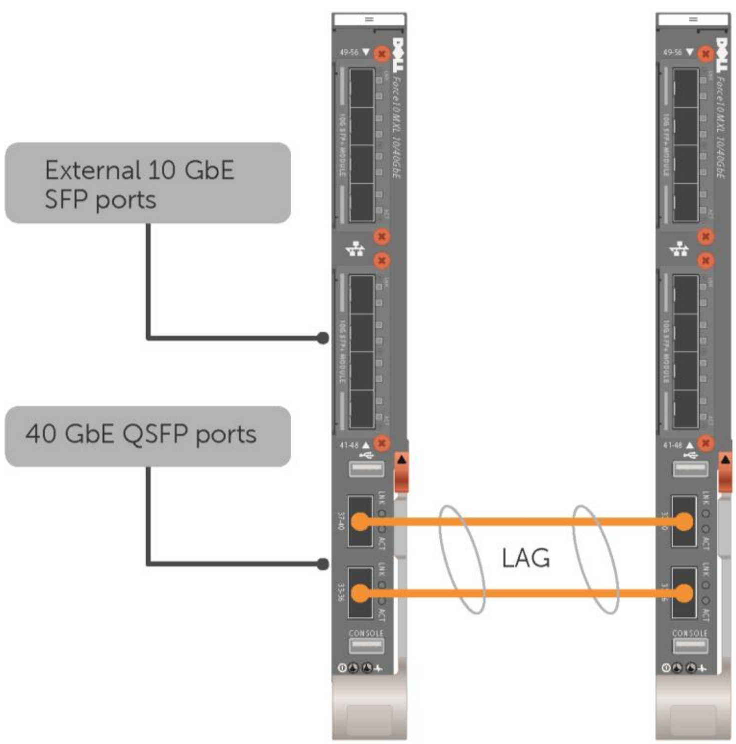


Figure 4 LAG connection of two Dell Networking MXL blade switches in a PowerEdge M1000e

5.1.1 DCB configuration on MXL switch

Configuring DCB on the Ethernet switch is a crucial part of the entire DCB setup. The DCB settings on the Ethernet switch will propagate to the connected devices using DCBX. Dell EqualLogic storage arrays are in DCB willing mode and can accept DCBX configuration parameters from the network switch to configure the DCB settings accordingly. **When DCB is properly configured on the switch, it will remotely configure DCB settings on the storage arrays and host network adapters**. This makes the DCB configuration on the switch the most important step when configuring a DCB network.

This section covers the basic steps for configuring DCB on a Dell Networking MXL network switch. The configuration steps are similar for other switches.

> **Note**: For the most current information this and other switches, refer to the "Switch Configuration Guides" page in Dell TechCenter at http://en.community.dell.com/techcenter/storage/w/wiki/4250.switch-configuration-guides-by-sis.aspx.

These configuration steps were used to configure the Dell Networking MXL switch:

- Enable switch ports and spanning tree on switch ports.
- Enable jumbo frame support.
- Configure QSFP ports for LAG between the switches.
- Enable DCB on the switch and create the DCB policy.
  - Create separate non-default VLANs for iSCSI and LAN traffic.
  - Configure priority group and policies.
  - Configure ETS values.
- Save the switch settings.

Detailed configuration instructions with switch configuration commands are available in Appendix B.

5.2 EqualLogic storage array configuration

Two EqualLogic PS6110XV 10GbE storage arrays were used during testing. The storage arrays were connected to the Dell Networking MXL blade switch external 10GbE ports. The two storage arrays were configured in a single pool and eight volumes were created across the two arrays. Appropriate access to volumes was created for servers to connect to the iSCSI volumes.

> For instructions regarding array configuration and setup refer to the *EqualLogic Configuration Guide* at http://en.community.dell.com/techcenter/storage/w/wiki/2639.equallogic-configuration-guide.aspx

The EqualLogic arrays are always in the DCB willing mode state to accept the DCBX parameters from the network switch on the storage network. The only DCB settings configured on the EqualLogic arrays were to enable DCB (the default) and to set a non-default VLAN. The detailed steps are outlined below.

5.2.1 DCB configuration

This section covers the DCB configuration steps needed for EqualLogic arrays using EqualLogic Group Manager. Other EqualLogic array configuration details can be found in the "Array Configuration" document on the Rapid EqualLogic Configuration Portal at http://en.community.dell.com/techcenter/storage/w/wiki/3615.rapid-equallogic-configuration-portal-by-sis.aspx

1. In Group Manager, select **Group Configuration** on the left.
2. Under the **Advanced** tab, **Enable DCB** should be checked by default.

> **Note:** The Group Manager interface of EqualLogic PS Series firmware version 7.0 and later does not allow disabling of DCB.

3. Change the **VLAN ID** to match the iSCSI VLAN defined on the switch. In this example, the parameter was set to VLAN 10.
4. Save the setting by clicking the disk icon in the top right corner.
   The DCB configuration is complete.

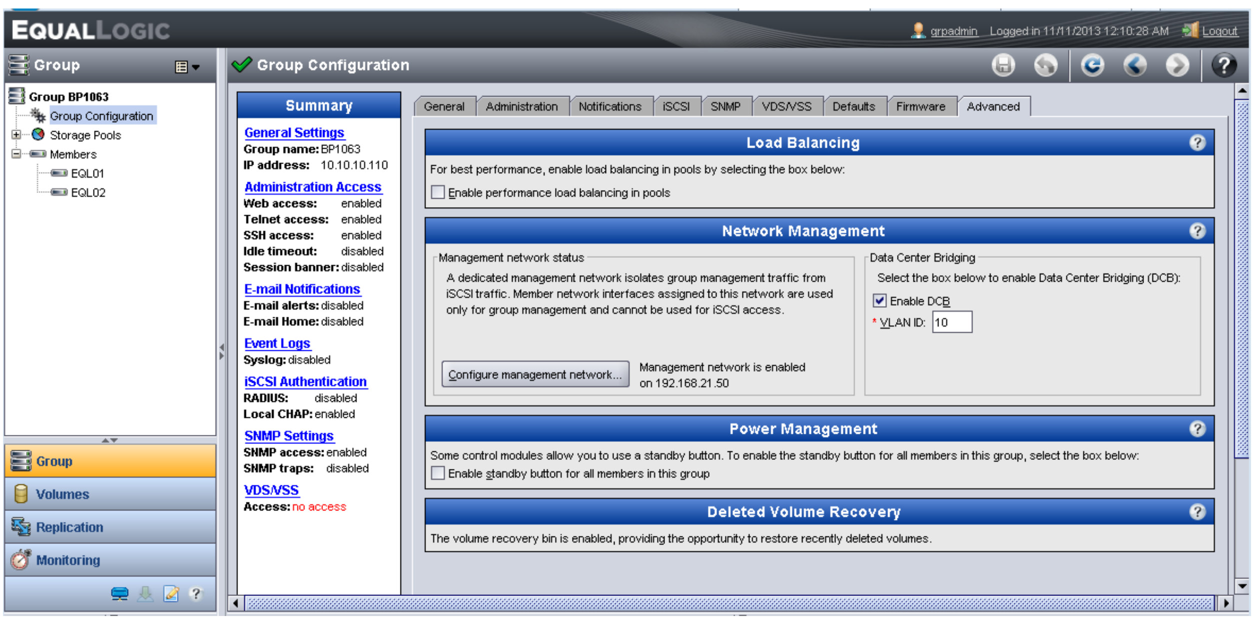


Figure 5 Enable DCB using EqualLogic Group Manager

> **Note**: EqualLogic iSCSI arrays are in a DCB willing state to accept DCBX parameters from the switch. Manually changing the DCB settings in the array is not possible.

6 DCB on Windows 2008 R2 and Windows Server 2012

The Windows 2008 R2 operating system needs a network with a DCB-capable Ethernet network adapter (with hardware offload capabilities or a proprietary device driver) and DCB-capable hardware switches in order to implement an end-to-end DCB solution. As an alternate solution, Windows Server 2012 OS offers a DCB software feature that allows DCB implementation without using specialized network adapters. This section focuses on DCB implementation using the Broadcom iSCSI hardware initiator (iSOE mode) on Windows 2008 R2 and Windows Server 2012 operating systems.

6.1 Solution setup

The PowerEdge M620 blade server was loaded with the Windows 2008 R2 server operating system and the Dell Host integration Tool (HIT) Kit for Windows. Eight 100GB iSCSI volumes were connected to the server using the Broadcom 57810 iSCSI hardware initiator. The test setup diagram is shown in Figure 6.

> NIC Load Balancing and Failover (LBFO) is a new feature for NIC teaming that is natively supported by Windows Server 2012. Dell recommends using the HIT Kit and MPIO for NICs connected to EqualLogic iSCSI storage. The use of **LBFO is not recommended** for NICs dedicated to SAN connectivity since it does not provide any benefit over MPIO. Microsoft LBFO or another vendor specific NIC teaming can be used for non-SAN connected NIC ports. When using NPAR with a DCB enabled (converged) network, LBFO can also be used with partitions that are not used for SAN connectivity.

Broadcom 57810 network adapters were logically partitioned using the Broadcom NPAR feature to enable a separate iSCSI network adapter and Ethernet protocol network adapter. The iSCSI partition was used for the iSCSI storage traffic while the NDIS function/partition was used for the other traffic types (LAN /Data traffic).

Function 0/1 was configured with VLAN 10 for iSCSI and Function 2/3 was configured with VLAN 20 for LAN traffic. Table 1 shows the IP address, VLAN configuration of the NDIS and iSCSI logical adapters as well as configured function.

> Broadcom 57810 can also be used in single function mode to implement DCB and to use the iSCSI hardware initiator along with a separate iSCSI adapter and NDIS/Ethernet adapter. The performance of this configuration is expected to be similar to the configuration using NPAR.

> Appendix D describes the configuration of NPAR using BACS and the configuration of iSCSI and NDIS adapters.

Table 1 Parameters on the NDIS and iSCSI partitions

Parameters	**NDIS VDB client**	**iSCSI offload Adapter**
IP address	10.10.20.XXX (LAN IP address)	10.10.10.XXX (iSCSI SAN IP address)
VLAN ID	VLAN 20	VLAN 10
Jumbo frame	No change	MTU 9216
DCB settings	N/A	DCBx in “willing mode”

The DCB parameters on the switches with PFC configured for priority 4 traffic ensured that iSCSI traffic was configured as lossless. The ETS settings were configured with 50% of the bandwidth for iSCSI and 50% for other traffic to give ample bandwidth to iSCSI storage traffic. These settings were configured on the Dell Networking MXL switch to communicate DCB parameters with other components using DCBX. The Broadcom network adapter and EqualLogic storage arrays were set in willing mode to accept the DCB configuration settings propagated by the switch through DCBX.

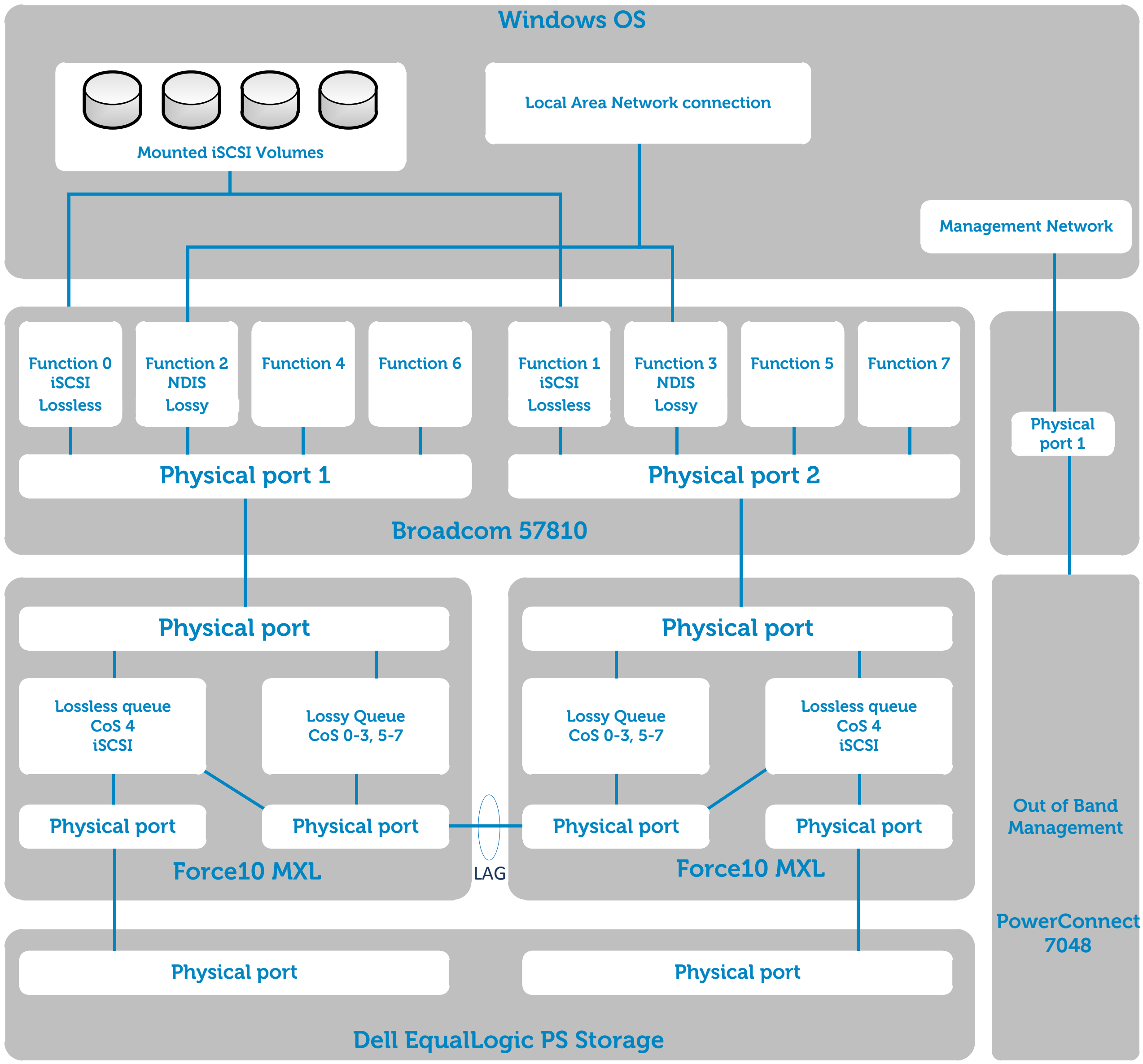


Figure 6 Windows OS test setup using Broadcom iSCSI hardware initiator.

6.2 Performance analysis of converged traffic

The performance analysis of DCB capabilities was tested by simulating iSCSI and LAN traffic on the M620 Windows server. iSCSI traffic was generated by mounting volumes through iSCSI connections and running vdbench to generate iSCSI traffic. LAN traffic was generated by using iPerf to send packets to another M620 server on the network. The DCB settings were verified by monitoring the ETS distribution between iSCSI and LAN traffic. Lossless behavior of iSCSI traffic was enforced with PFC frames.

The graphs in Figure 7 and Figure 8 depict iSCSI and LAN traffic performance on Windows 2008 R2 and Windows Server 2012 respectively, using hardware initiators with offload (iSOE). The charts include a baseline throughput of iSCSI and LAN traffic with converged iSCSI and LAN traffic for **sequential read** (**SR**) and **sequential write** (**SW**) workloads. They also illustrate bandwidth distribution with ETS settings of 50-50 between iSCSI and other traffic types maintained by the DCB implementation.

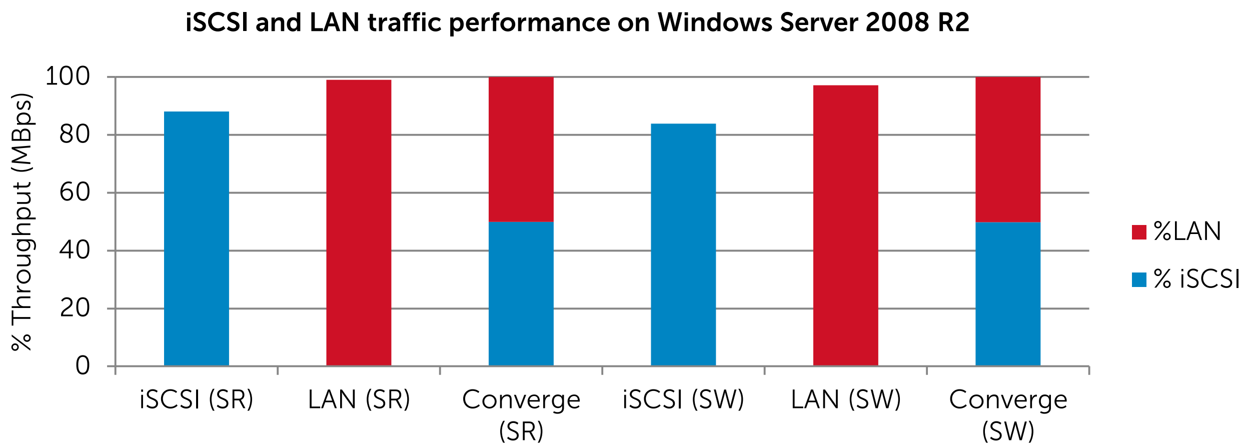


Figure 7 iSCSI and LAN traffic bandwidth distribution on Windows 2008 R2 using iSCSI hardware initiator

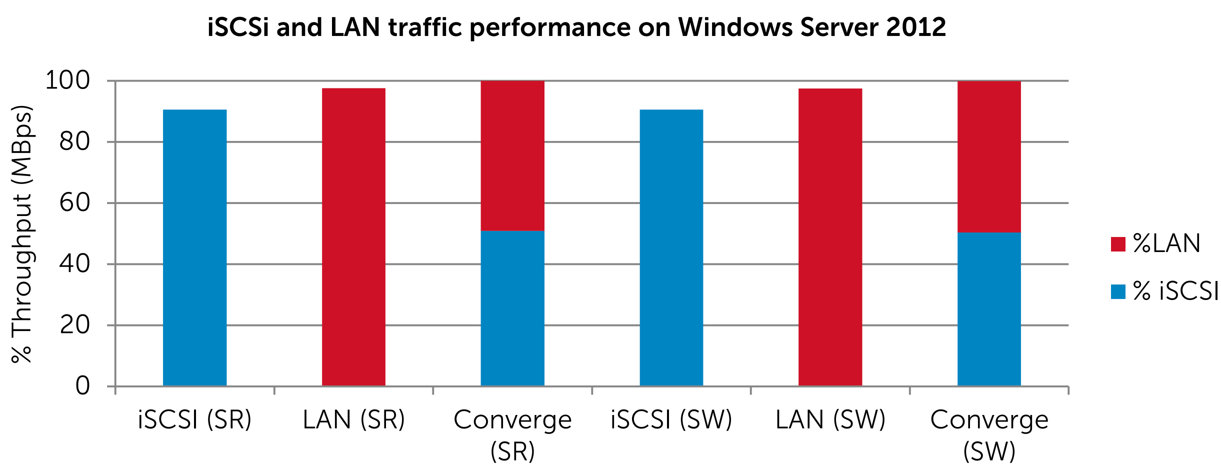


Figure 8 iSCSI and LAN traffic bandwidth distribution on Windows 2012 using iSCSI hardware initiator

7 Windows Server 2012 DCB features using software iSCSI initiator

Windows Server 2012 includes a feature that enables a full DCB solution with standard Ethernet adapters. The OS implements DCB, PFC and ETS by classifying and then tagging the traffic based on priority. Network adapters with offload capabilities (e.g. iSOE, TOE) are not necessary to implement DCB while using Microsoft DCB feature. As discussed in section 3, offload features are also not compatible with the software DCB implementation of Windows Server 2012.

7.1 Solution setup

The Windows Server 2012 operating system and HIT Kit for Windows were all installed on the Dell PowerEdge M620 blade server and the Windows Server 2012 DCB feature was enabled. The latest Broadcom 57810 drivers were also installed on the system. Eight 100GB iSCSI volumes were connected to the server using Broadcom 57810 iSCSI Hardware Initiator. The test setup diagram is shown in Figure 9.

Detailed steps for configuring DCB for Windows Sever 2012 and the Broadcom 57810 adapters are outlined in Appendix C.

Windows Server 2012 natively supports a new feature for NIC teaming called NIC Load Balancing and Failover (LBFO). Dell recommends using the Host Integration Tool kit (HIT Kit) and MPIO for any NIC that is connected to EqualLogic iSCSI storage. The use of **LBFO is not recommended** for NICs dedicated to SAN connectivity since it does not add any benefit over MPIO. Microsoft LBFO or other NIC vendor specific teaming can still be used for non-SAN connected NIC ports. When using NIC partitioning (NPAR) with a DCB enabled (converged) network, LBFO can also be used with partitions that are not used for SAN connectivity.

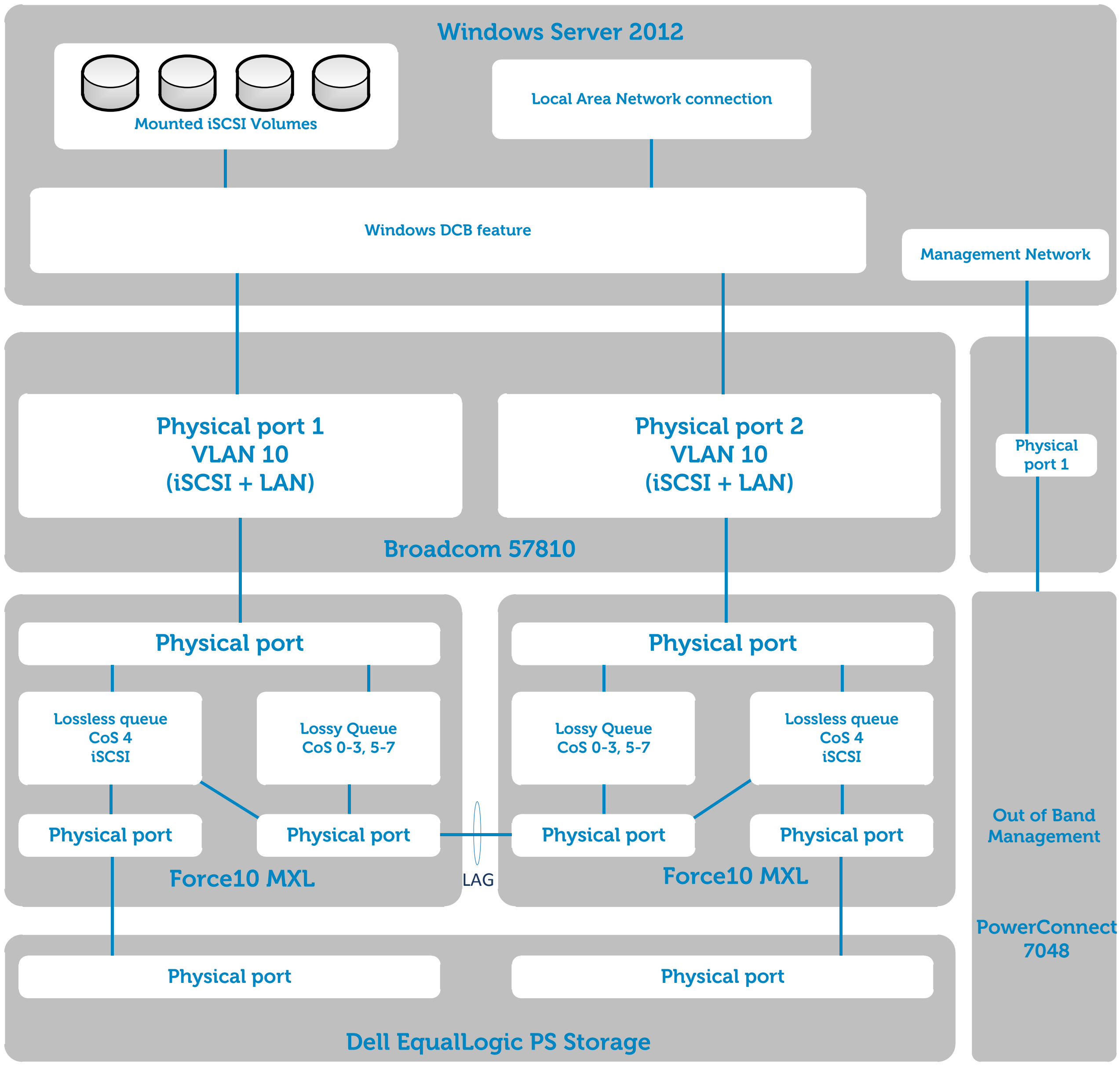


Figure 9 Windows Server 2012 test setup using Windows DCB feature.

Because the iSCSI and LAN traffic shared the same network adapter as well as the same VLAN, partitioning the Broadcom 57810 network adapters would have disabled the Windows operating system's ability to manage DCB with the Broadcom NPAR network adapters.

Note: VLAN separation between iSCSI and LAN is not provided in this configuration.

The Broadcom network adapters were configured with VLAN 10 which carried both iSCSI and LAN traffic. Table 2 shows the IP address and VLAN configuration network adapters used during testing.

Appendix D describes configuration of Broadcom Network Adapters using BACS.

Table 2 Parameters on the NDIS and iSCSI partitions

Parameters	Settings
IP address	10.10.10.XXX
VLAN ID	VLAN 10
Jumbo frame	MTU 9216

7.2 Performance analysis of iSCSI and LAN traffic

The DCB feature on Windows Server 2012 was tested by generating converged iSCSI and LAN traffic. iSCSI traffic was generated using vdbench on the iSCSI volumes which were mounted using the Microsoft iSCSI software initiator. LAN traffic was generated using iPerf and sent from one M620 server to another M620 server on the network. The DCB settings were verified by monitoring ETS distribution between iSCSI and LAN traffic. The lossless behavior of iSCSI traffic was confirmed by monitoring the transmission and reception of PFC frames.

Running converged iSCSI and LAN traffic on the network resulted in the proper distribution of bandwidth between iSCSI and LAN traffic. This shows that the configured minimum bandwidth of 50% for storage traffic was maintained by the ETS settings. There were no retransmissions because iSCSI traffic was in a lossless class.

Figure 10 shows a throughput and CPU utilization comparison of the tests that used a dedicated DCB capable hardware initiator and the Windows Server 2012 software DCB feature. Using the Windows software DCB feature resulted in additional workload for the operating system stack and CPU. The network TCP stack and SCSI protocol layer were controlled by the operating system instead of the dedicated iSCSI hardware initiator. Using the Windows DCB feature resulted in slightly higher CPU utilization and slightly lower throughput for iSCSI as shown in Figure 10 and Figure 11. iSCSI **sequential read (SR)** performance was 20% lower and iSCSI **sequential write (SW)** was 5% lower compared to tests performed using the iSCSI hardware initiator on Windows 2012.

This test was run using only two 10Gb ports on a single Windows server. If the number of ports is increased, it is expected that the CPU utilization will increase and as a result performance will continue to be lower than observed with an iSCSI hardware initiator.

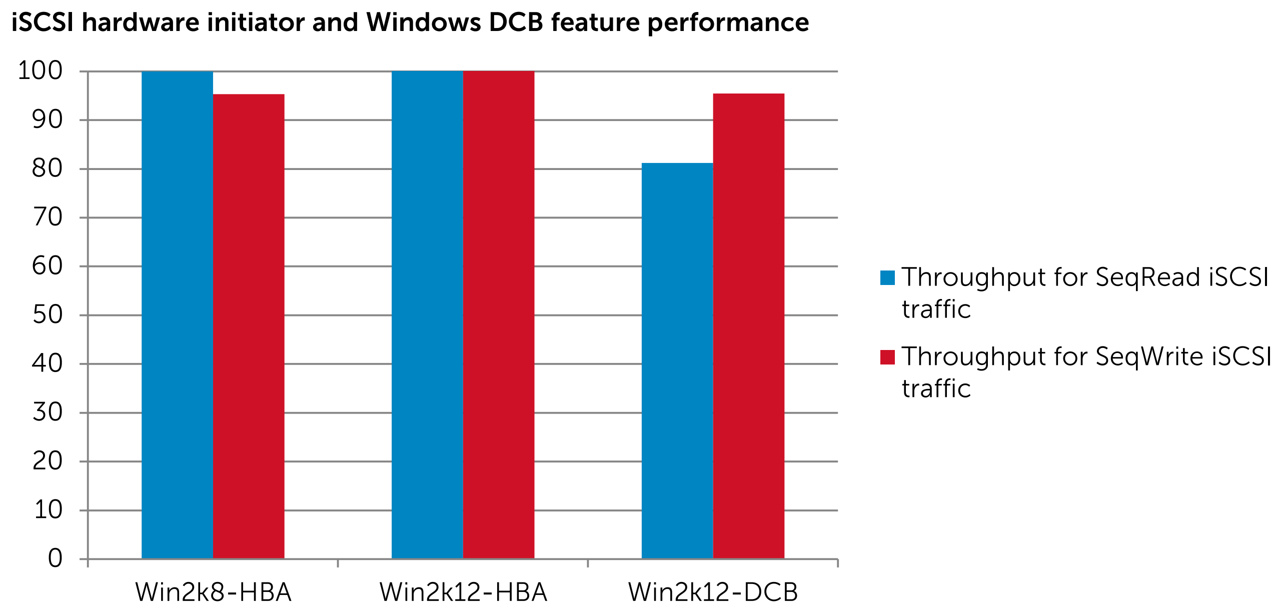


Figure 10 Performance comparison of DCB hardware and the DCB software feature

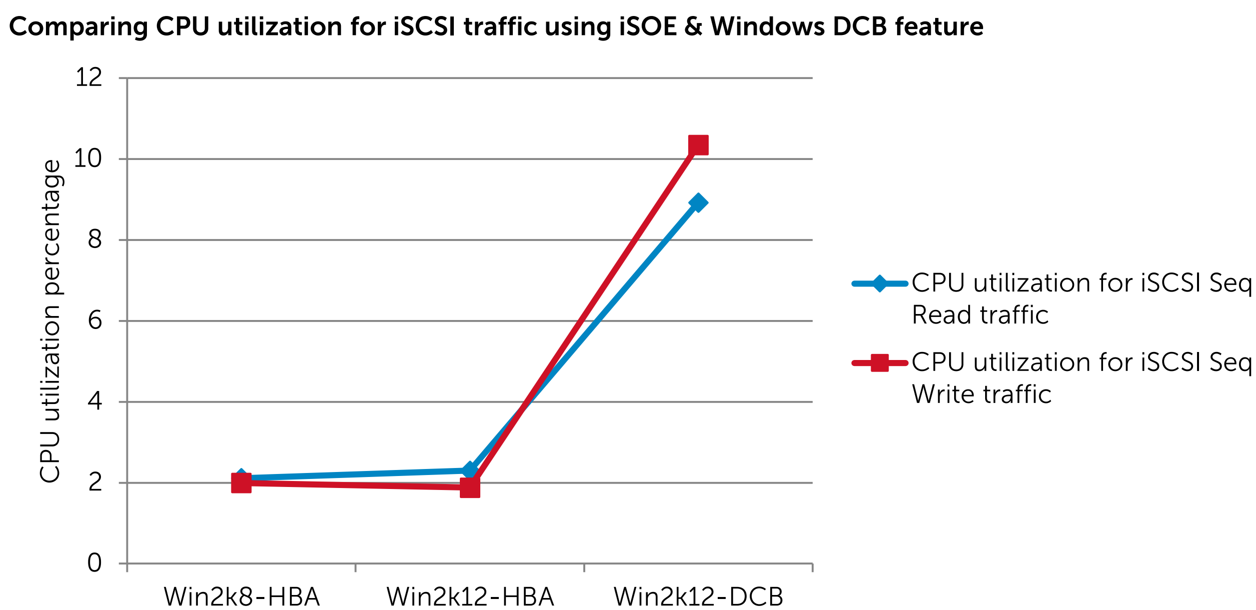


Figure 11 Comparison of CPU utilization between the DCB hardware and DCB software feature

8 DCB on Hyper-V 3.0

Hyper-V 3.0 requires a DCB capable Ethernet network adapter and DCB capable hardware switches in order to implement a fully enabled DCB solution. Figure 12 shows a logical overview of the DCB implementation on Hyper-V 3.0 using the Hardware iSCSI initiator. This section focuses on a DCB implementation using the Broadcom iSCSI hardware initiator with Hyper-V 3.0.

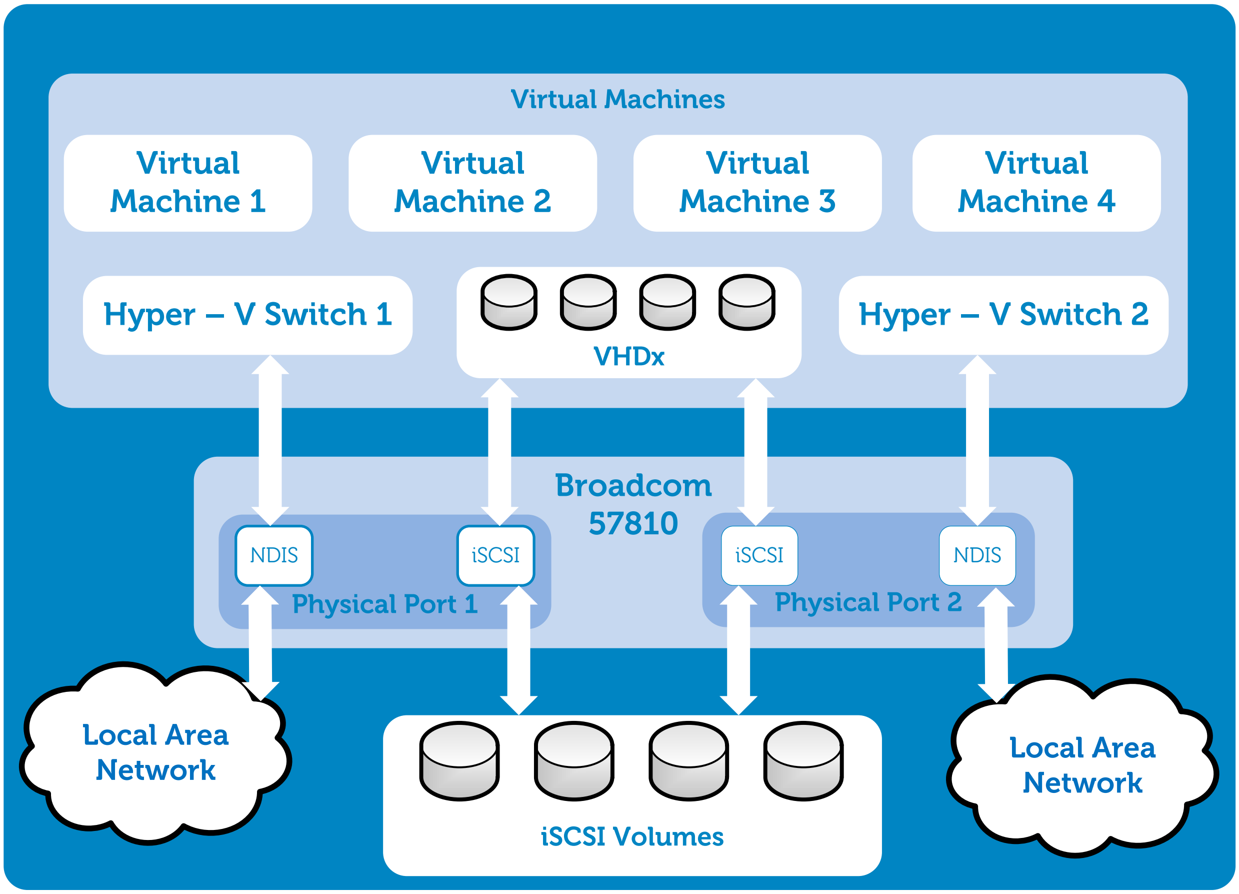


Figure 12 DCB solution logical diagram using Hyper-V 3.0

Note: Although not specifically tested or documented as part of this paper, implementation of Hyper V 2.0 with Windows Server 2008 R2 is very similar to the steps presented for Hyper-V 3.0.

8.1 Solution setup

The Windows Server 2012 operating system is installed on the PowerEdge M620 blade server with the Hyper-V Role enabled. The Windows HIT Kit is also installed on the server since a host initiator was used for iSCSI connections to the EqualLogic iSCSI volumes. The latest Broadcom 57810 drivers were installed on the system. The test setup is illustrated in Figure 13.

This test setup used the DCB capable iSCSI hardware initiator Broadcom 57810 network adapter for iSCSI storage traffic. Using NPAR, the Broadcom 57810 network adapter was logically partitioned into one iSCSI network adapter and two Ethernet protocol network adapters. The iSCSI partition was used for iSCSI traffic, the remaining two partitions were used for VM LAN traffic and live migration traffic

Function 0/1 was configured with VLAN 10 (for iSCSI), Function 2/3 was configured for VLAN 20 (LAN traffic from VM), and Function 4/5 was configured with VLAN 21 (for live migration traffic).

Four Virtual machines were created using Hyper-V manager on Windows Server 2012. All four virtual machines were identical and ran Windows Server 2012 operating system. Virtual switches were created for Function 2/3 resulting in two virtual switches. Virtual NICs were created for each virtual switch on every virtual machine. Each virtual machine had two network adapters for LAN traffic. The second NDIS partition (Function 4/5) was used by the root partition for Live Migration traffic. The iSCSI partition (Function 0/1) was used by the root partition to establish iSCSI connections. The iSCSI volumes are configured as VHDx. Eight VHDx volumes were mounted on each virtual machine.

Appendix D describes NPAR configuration using BACS as well as iSCSI and NDIS adapter configuration.

Table 3 Parameters on the NDIS and iSCSI partitions

Parameters	**iSCSI offload Adapter (Function 0/1)**	**NDIS adapter for VM traffic (Function 2/3)**	**Root partitions NDIS adapter (Function 4/5)**
IP address	10.10.10.XXX (iSCSI / SAN IP address)	10.10.20.XXX (Configured in the Virtual Machine)	10.10.21.XXX (Live Migration IP address)
VLAN ID	VLAN 10	VLAN 20	VLAN 21
Jumbo frame	MTU 9216	No change	No change
DCB settings	DCBx in “willing mode”	N/A	N/A

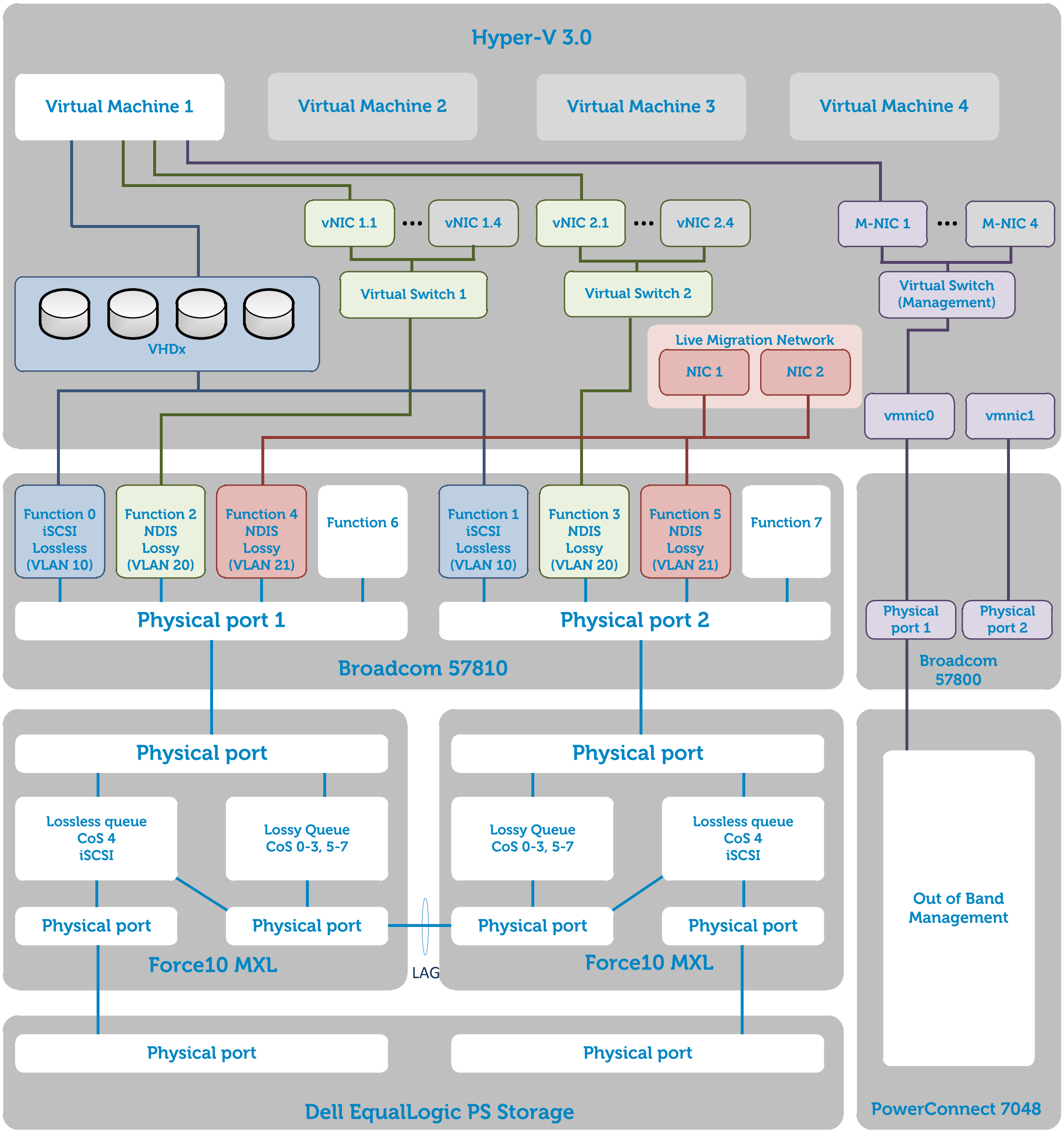


Figure 13 Windows Hyper-V 3.0 test setup using Broadcom iSCSI Hardware Initiator.

DCB parameters on the switches with PFC configured for priority 4 traffic ensured that iSCSI traffic was configured to be lossless. The ETS settings were configured with 50% of the bandwidth for iSCSI and 50% for other traffic to give ample bandwidth to iSCSI storage traffic. These settings were configured on the Dell Networking MXL switch which communicated DCB parameters with other components using DCBX.

The Broadcom network adapter and EqualLogic storage arrays were set in willing mode to accept the DCB configuration settings propagated by the switch through DCBX.

8.2 Performance analysis of iSCSI and LAN traffic

DCB capabilities of the Hyper-V 3.0 solution were tested by generating simulated iSCSI and LAN traffic from the virtual machines. A simulation of Live migration traffic was generated at the root-partition. The iSCSI traffic was generated by using vdbench to simulate storage traffic from VHDx volumes mounted on the virtual machine. iPerf was used to generate the LAN traffic on the virtual machines and live migration traffic between clusters. The DCB settings were verified by monitoring ETS distribution between iSCSI and LAN traffic; and the lossless behavior of iSCSI traffic was verified by monitoring the use of PFC frames.

Figure 14 depicts the performance of the iSCSI and LAN traffic that was generated on the Hyper-V 3.0 solution. The graph shows that the ETS setting maintained the guaranteed 50% of the bandwidth for iSCSI traffic. There were no retransmissions or dropped packets for the storage iSCSI traffic which confirms the lossless behavior of iSCSI storage traffic.

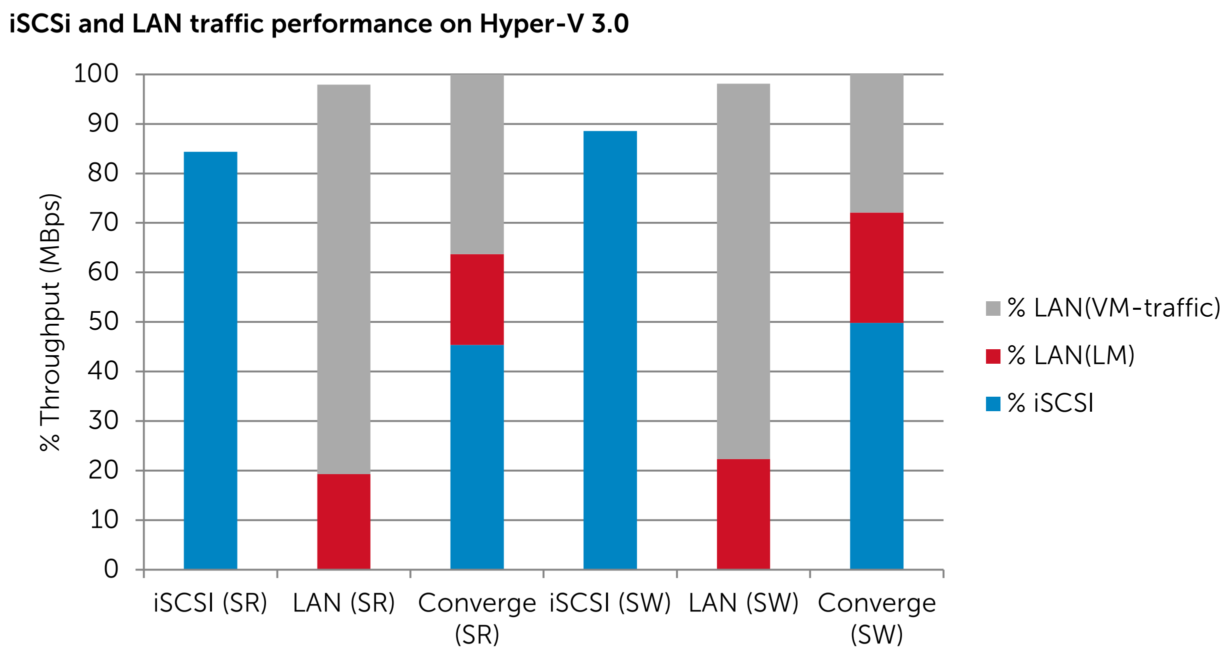


Figure 14 iSCSI, VM LAN and live migration LAN traffic bandwidth distribution on Hyper-V 3.0 using iSCSI hardware initiator.

9 Best practices and conclusions

DCB provides the ability to use a single physical network for multiple traffic types while guaranteeing bandwidth for specific traffic. In addition, it provides a lossless traffic class for storage or other performance sensitive storage traffic. This DCB capability is highly valuable in helping to reduce infrastructure cost by making efficient use of the network.

Windows 2008 R2, Windows Server 2012 and Windows Hyper-V 3.0 require a network adapter capable of supporting DCB in order to successfully deploy DCB on the servers. Windows Server 2012 has a built-in Microsoft software DCB feature and can also implement DCB without a DCB capable network adapter. The DCB capable Dell Networking switches and EqualLogic storage arrays combined with a DCB enabled server enables an end-to-end DCB solution.

The tested configurations presented in Sections 6, 7 and 8 are complete DCB solutions and the test results indicate that lossless behavior and guaranteed bandwidth can be achieved for performance sensitive iSCSI traffic.

Specifically, the iSCSI traffic load in a DCB implementation using dedicated hardware iSCSI initiators can increase overall performance and lower CPU utilization on the host. This happens as the TCP/IP network stack, and SCSI layer protocol are offloaded to the hardware initiator.

On the other hand, software iSCSI initiator implementation with using Microsoft DCB feature will reduce the infrastructure cost, as no dedicated hardware is needed for DCB implementation on the Server. Also one can implement DCB with their existing server hardware as long as the Windows Server 2012 is installed and the existing server hardware supports DCB. The downside is lower iSCSI throughput and higher CPU utilization as the Operating system takes responsibility for implementing the SCSI and TCP/IP network stack. Also, dedicated VLANs cannot be provided for iSCSI and LAN traffic which will not be complaint with existing EqualLogic best practices.

In conclusion, DCB technology provides a lower network hardware cost with more efficient use of network resources. The solutions presented in this paper will help achieve efficiency goals and provide guaranteed bandwidth as well as a lossless traffic class for performance sensitive iSCSI storage traffic.

A Test setup configuration details

Table 4 Test component version table.

Component	Version / Model number
Dell Networking MXL switch	
Firmware Version	9.2.0.0
Dell PowerEdge M620 server components.	
BIOS version	Dell Inc. 1.6.1
OS versions	Microsoft Windows Server 2012 Datacenter (6.2.9200 Build 9200) Microsoft Windows Server 2008 R2 Standard (6.1.7601 Build 7601)
HIT kit version	Version 4.5
Dell PowerEdge M1000e chassis components	
CMC firmware	4.20
Broadcom 57810	
Firmware version	FFV7.6.15
Driver version	7.6.59.0
iSCSI Driver version	7.6.2.0
NDIS Driver version	7.6.51.0
Broadcom Management Suite	
BACS	Broadcom Advance Control Suite 4 – Version 15.6.31.0
Dell EqualLogic PS6110 storage array components	
Firmware version	Storage Array Firmware V6.0.5
RAID configuration	RAID 10

B Dell Networking MXL blade switch DCB configuration

This section outlines the steps to configure DCB on a Dell Networking or Force10 MXL switch. For switches other than this please refer to the Switch Configuration Guides for EqualLogic at http://en.community.dell.com/techcenter/storage/w/wiki/4250.switch-configuration-guides-by-sis.aspx.

Disable 802.3 x flow control:

If 802.3x flow control (link layer flow control) is configured on the switch, it needs to be disabled before enabling DCB on the switch.
This step can be skipped if flow control is not configured.

```
FTOS#configure
FTOS(conf)#interface range tengigabitethernet 0/1 - 32 , tengigabitethernet
  0/41 - 56
FTOS(conf-if-range-te-0/1-32,te-0/41-56)#no flowcontrol rx on tx off
FTOS(conf-if-range-te-0/1-32,te-0/41-56)#exit
FTOS(conf)#interface range fortyGigE 0/33 , fortyGigE 0/37
FTOS(conf-if-range-fo-0/33,fo-0/37)#no flowcontrol rx on tx off
```

Enable DCB:

The Dell Networking MXL switch has DCB enabled by default. Switch buffering is optimized for two queues to accommodate the lossy and lossless traffic classes.

```
FTOS#configure
FTOS(conf)#dcb enable
FTOS(conf)#dcb stack-unit all pfc-buffering pfc-ports 56 pfc-queues 2
FTOS(conf)#exit
FTOS#copy running-config startup-config
FTOS#reload
```

Configure VLANs:

Non-default VLANs are configured for iSCSI and LAN traffic. Creating separate VLANs for iSCSI traffic and LAN traffic provides logical separation between the two traffic types. Follow this EqualLogic best practice except when using the Windows 2012 DCB feature where iSCSI and LAN traffic reside on a single VLAN.

> **Note:** Using dedicated virtual LANs for different traffic types only provides a logical separation to the network. Unlike DCB, it does not provide guaranteed bandwidth and lossless behavior of traffic. Therefore, it is not recommended to **only** use VLANs when mixing traffic types on the same switch fabric.

Multiple VLANs can be configured as needed. For example separate non-default VLANs can be created for Live Migration traffic between physical hosts.

```
FTOS(conf)#interface vlan 10  (For iSCSI/SAN traffic)
FTOS (conf-if-vl-10)#tagged tengigabitethernet 0/1-32
FTOS (conf-if-vl-10)#tagged tengigabitethernet 0/41-56
FTOS (conf-if-vl-10)#tagged port-channel 1
FTOS (conf-if-vl-10)#exit
FTOS(conf)#interface vlan 20  (For LAN traffic)
FTOS(conf-if-vl-20)#tagged tengigabitethernet 0/1-32
FTOS(conf-if-vl-20)#tagged tengigabitethernet 0/41-56
FTOS(conf-if-vl-20)#tagged port-channel 1
FTOS(conf-if-vl-20)#exit
```

Configure Priority group:

Priority group iSCSI and OTHER are created for iSCSI and LAN traffic respectively. Also appropriate traffic type is linked to the priority group. The switch is configured to enforce PFC on iSCSI traffic type and ETS is on both priority groups.

```
FTOS(conf)#dcb-input pfc
FTOS(conf-dcb-in)#pfc priority 4
FTOS(conf-dcb-in)#exit
FTOS(conf)#priority-group iSCSI
FTOS(conf-pg)#priority-list 4
FTOS(conf-pg)#set-pgid 1
FTOS(conf-pg)#exit
FTOS(conf)#priority-group OTHER
FTOS(conf-pg)#priority-list 0-3,5-7
FTOS(conf-pg)#set-pgid 2
FTOS(conf-pg)#exit
FTOS(conf)#dcb-output ets
FTOS(conf-dcb-out)#priority-group iSCSI qos-policy iSCSI
FTOS(conf-dcb-out)#priority-group OTHER qos-policy OTHER
FTOS(conf-dcb-out)#exit
FTOS(conf)#service-class dynamic dot1p
```

Create DCB policies for input and output:

Input PFC policies and output ETS policies are configured and applied to switch ports.

```
FTOS(conf)#interface range tengigabitethernet 0/1 - 32 , tengigabitethernet
 0/41 - 56
FTOS(conf-if-range-te-0/1-32,te-0/41-56)#dcb-policy input pfc
FTOS(conf-if-range-te-0/1-32,te-0/41-56)#dcb-policy output ets
FTOS(conf-if-range-te-0/1-32,te-0/41-56)#protocol lldp
FTOS(conf-if-range-te-0/1-32,te-0/41-56-lldp)#exit
FTOS(conf-if-range-te-0/1-32,te-0/41-56)#exit
FTOS(conf)#interface range fortyGigE 0/33 , fortyGigE 0/37
FTOS(conf-if-range-fo-0/33,fo-0/37)#dcb-policy input pfc
FTOS(conf-if-range-fo-0/33,fo-0/37)#dcb-policy output ets
FTOS(conf-if-range-fo-0/33,fo-0/37)#exit
FTOS(conf)#exit
```

Configure switch port ETS settings:

ETS values will determine the bandwidth percentage allocated to different traffic classes. The iSCSI traffic class was assigned 50% bandwidth and LAN traffic was assigned 50% bandwidth through ETS settings.

```
FTOS#configure
FTOS(conf)#qos-policy-output iSCSI ets
FTOS(conf-qos-policy-out)#bandwidth-percentage 50
FTOS(conf-qos-policy-out)#exit
FTOS(conf)#qos-policy-output OTHER ets
FTOS(conf-qos-policy-out)#bandwidth-percentage 50
FTOS(conf-qos-policy-out)#exit
```

Save switch configuration:

```
FTOS#copy running-config startup-config
```

C Windows Server 2012 DCB and software iSCSI initiator configuration

Windows 2012 introduces a new software DCB feature which will provide DCB capabilities between standard Ethernet network adapters and DCB-capable switches.

Currently, there are three typical configuration modes available for implementing DCB. The first uses a standard 10GbE Ethernet NIC with proprietary driver and firmware support to provide the DCB functionality (such as priority tagging and ETS). This option was not tested as part of this paper. A second mode involves using a CNA hardware initiator or offload (iSoE) adapter where the TCP networking, iSCSI and DCB protocol functionality are all handled by the CNA hardware. A third option again uses a standard 10GbE Ethernet NIC in conjunction with a software DCB feature from Windows Server 2012. The following sections provide detailed configuration information for setting up the software DCB feature with Windows Server 2012 with a Broadcom 57810 CNA.

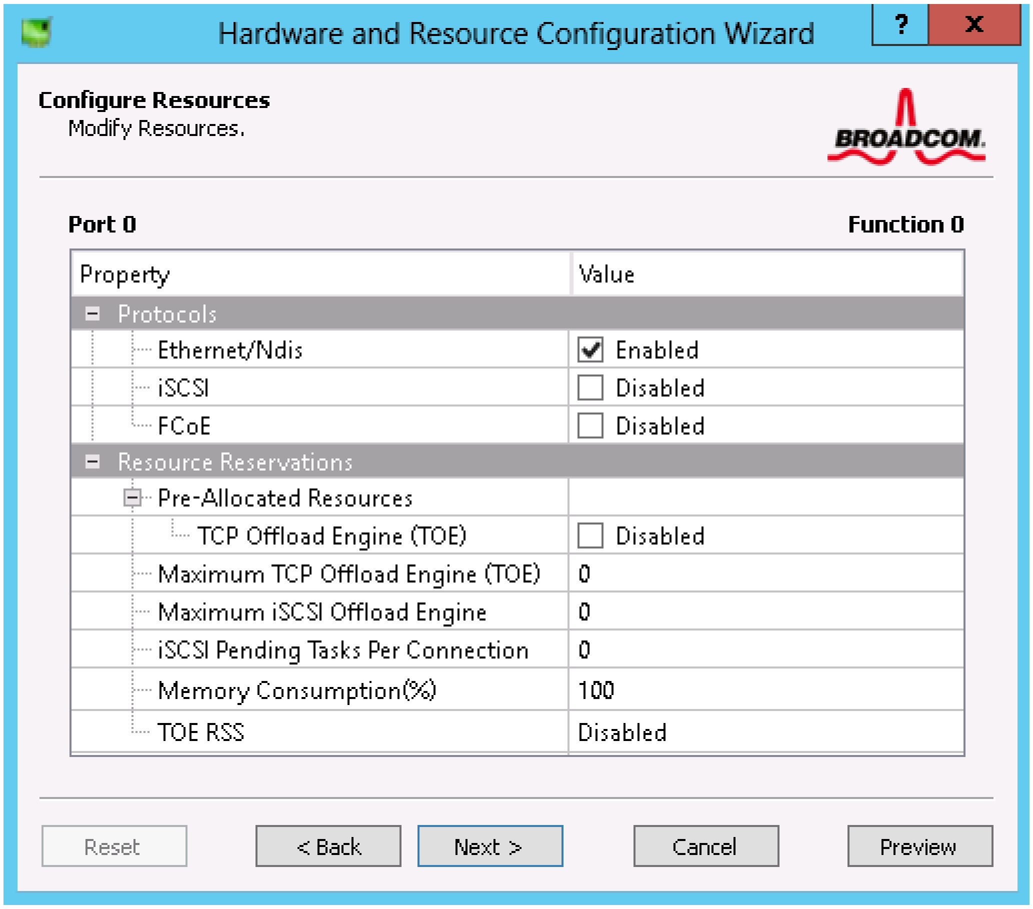
C.1 Adding a Broadcom 57810 to the configuration

The Microsoft software iSCSI initiator is used with the Windows Server 2012 DCB software feature. Full hardware iSCSI initiators are not supported because the Windows operating system needs to detect the entire Ethernet packet in order to implement DCB.

Note: The single function mode was used because the Windows DCB feature cannot manage Broadcom NPAR partitioned NICs.

1. Configure the Broadcom network adapter for single function mode using BACS 4.
   If the Broadcom network adapter is not in single function mode by default, complete the following steps.

   a. Click **Adapter # (BCM57810 B0)** in the BACS window to select the Broadcom network adapter.
   b. In the **Configuration** tab, expand **Multi-function** and click **Configure.**
   c. In the **Hardware and Resource Configuration Wizard**, set the multi-function mode as **Single Function**.
   d. Click **Apply** and then restart the server.
2. Configure the Broadcom network adapter with Ethernet protocol and disable the TCP Offload Engine.
   a. Click **Adapter # (BCM57810 B0)** in the BACS window to select the Broadcom Network adapter.
   b. In the **Configuration** tab, expand **Multi-function** and click **Configure.**
   c. In the **Hardware and Resource Configuration Wizard**, click **Next** to go to port configuration.
   d. Select the physical port to be configured and click **Next**.
   e. Check the **Ethernet/NDIS** protocol check-box and deselect the **TCP Offload Engine (TOE)** checkbox to disable TOE. Click **Next**.

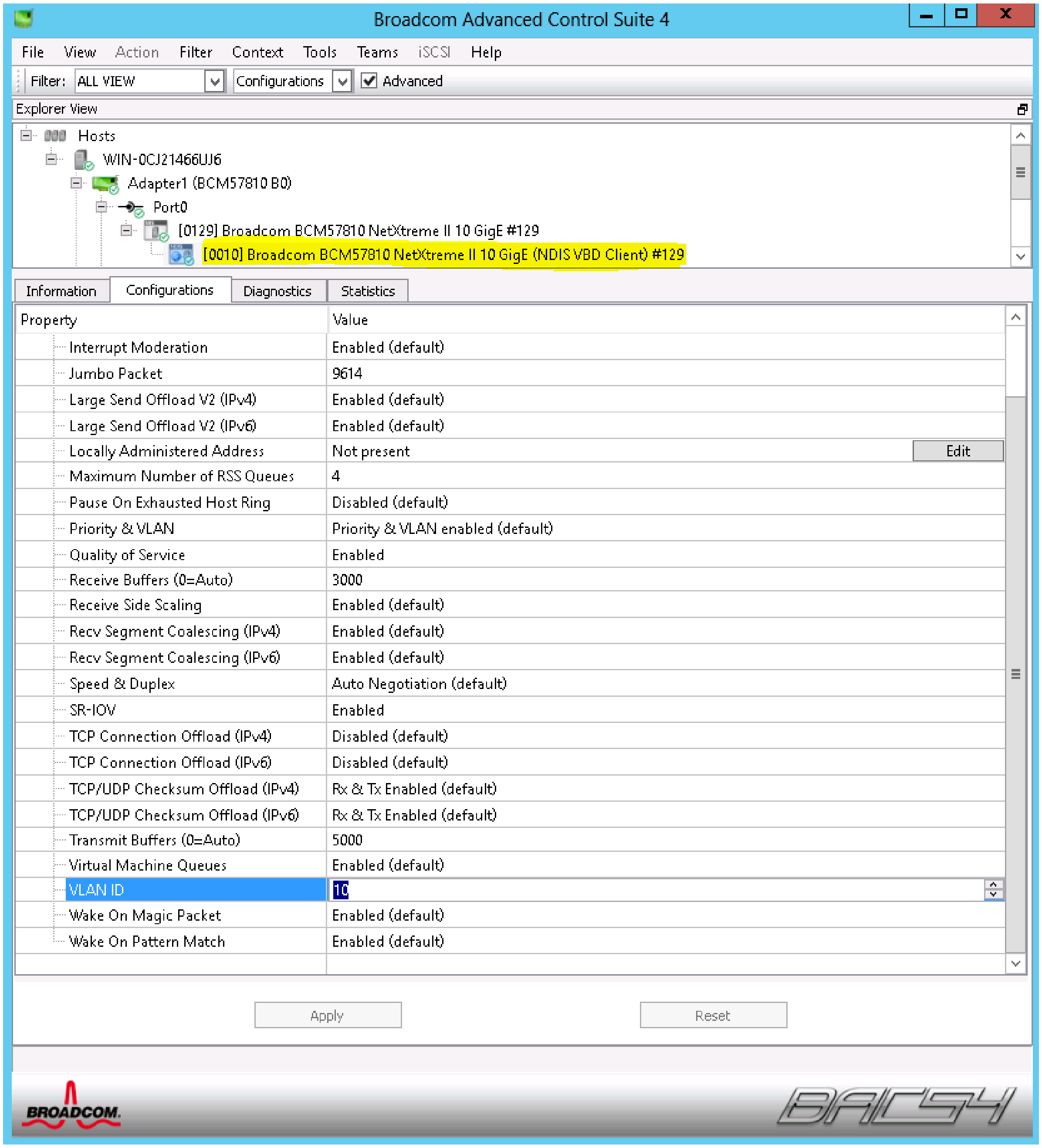


f. Repeat steps d through e for all the ports.

g. Click **Apply**.

3. Continue configuring the network adapter with the following steps.

4. Click **Broadcom BCM57810 NetXtreme II 10 GigE (NDIS VDB Client) #** in the BACS window and then click the **Configuration** tab. Expand the **Advanced** list to configure the settings.



5. Configure the settings as shown in Table 5.

Table 5 Settings for Ethernet network adapter

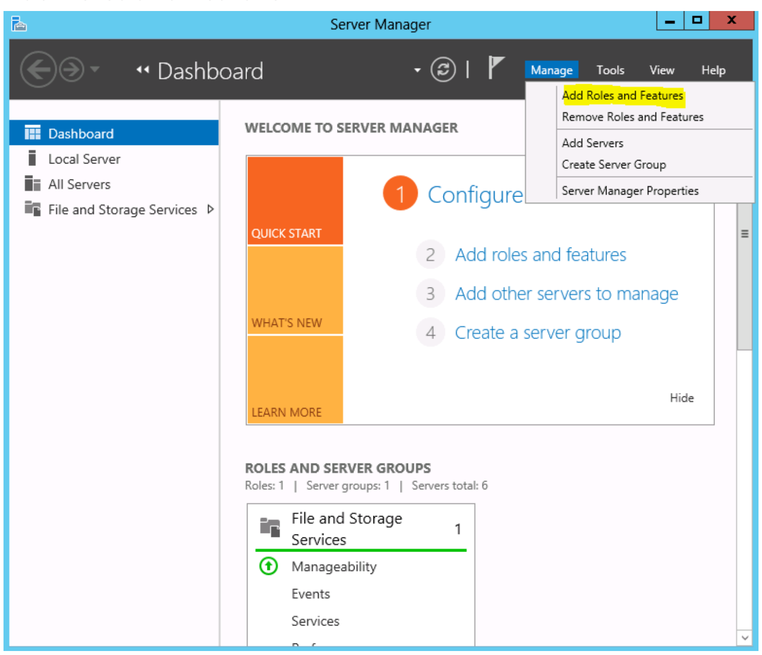
Settings	Values
Jumbo Packet	9614 (Max)
QoS	Enable
Receive Buffer	3000 (Max)
TCP connection Offload (IPv4)	Disable
TCP connection Offload (IPv6)	Disable
Transmit Buffer	5000 (Max)
VLAN ID	10 (VLAN ID of the Storage Network)

6. In the Windows OS Control Panel, click **Network and Internet** > **Network and Sharing Center** and enter the SAN IP addresses.

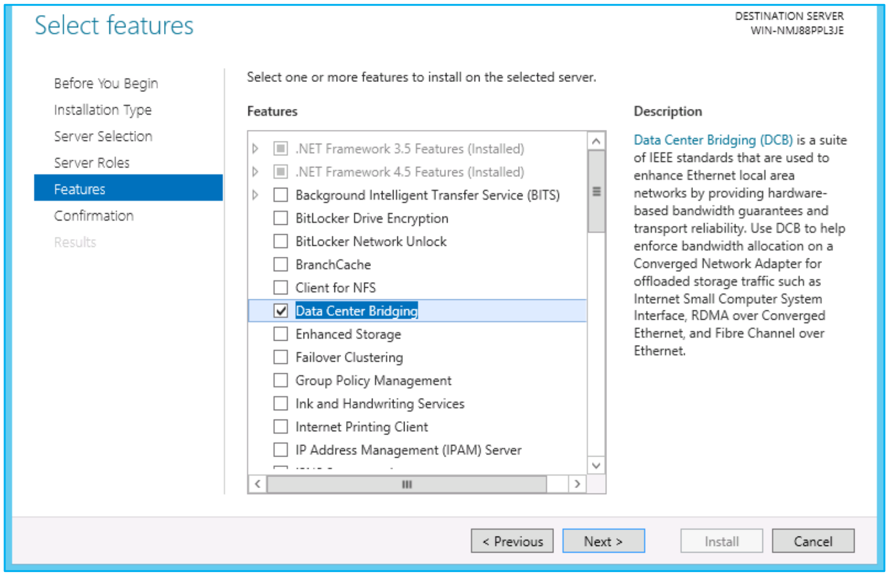
C.2 Installing the DCB software feature on Windows Server 2012

This section describes the setup needed to configure the DCB software feature on Windows Server 2012. Windows Server 2012 must be installed on the system before proceeding with the steps below.

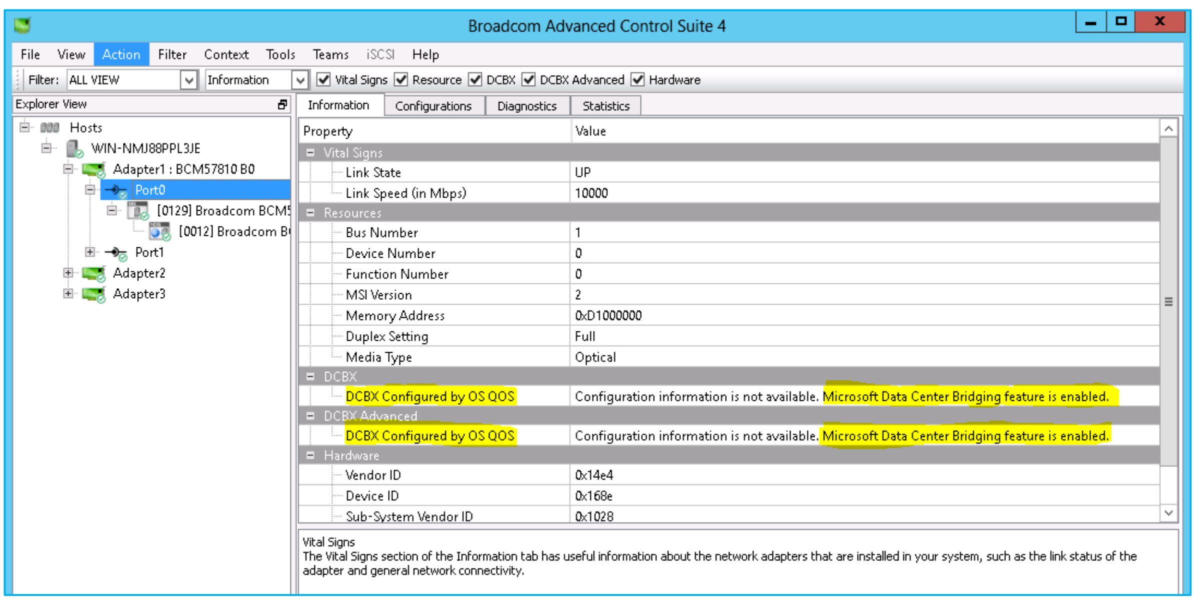
1. From Windows server 2012, go to the Server Manager. On the Dashboard, click **Manage** and select **Add Roles and Feature.**



2. In the **Add Roles and Feature** wizard, click **Installation Type** and then select a **Role-based** or **feature based** installation. Click **Next**.
3. Select the appropriate server and then click **Next.**
4. Skip the **Server Role** section by clicking **Next**.
5. In **Features**, check **Data Center Bridging** and click **Next.**



6. Click **Install** to start the DCB installation.
7. **Close** this window once the installation is complete.
8. To verify the Windows Server 2012 DCB feature is activated and detected by BACS, click the **Information** tab of the network Port. DCBX should indicate **DCBX Configured by OS QOS** as shown.



C.3 Configuring the Windows Server 2012 DCB feature

Once the DCB feature is installed, configure the settings for DCB on the Windows Server 2012 to match the switch.

> The Windows Server 2012 DCB feature cannot add priority tags to the traffic while operating in DCBX **Willing Mode**. Complete DCB functionality can only be achieved by disabling DCBX Willing mode using the PowerShell command.

1. Disable the Windows DCB Willing mode using:
   `Set-NetQosDcbxSetting -Willing 0`

```
PS C:\Users\Administrator> Set-NetQosDcbxSetting -Willing 0

Confirm
Are you sure you want to perform this action?
Set-NetQosDcbxSetting -Willing $false
[Y] Yes  [A] Yes to All  [N] No  [L] No to All  [S] Suspend  [?] Help (default is "Y"): y
PS C:\Users\Administrator>
```

> **Note**: Verify that the ETS settings of Windows Server 2012 DCB are consistent with the settings on the network switch.

2. Create an **iSCSI** traffic class and set the ETS value to 50 percent for priority 4 traffic type.
   `New-NetQosTrafficClass -Name iSCSI -BandwidthPercentage 50 -Algorithm ETS -Priority 4`
3. Create an **OTHER** Traffic Class and set the ETS value to 50 percent for priority 0, 1, 2, 3, 5, 6, and 7 traffic types.
   `New-NetQosTrafficClass -Name OTHER -BandwidthPercentage 50 -Algorithm ETS -Priority 0,1,2,3,5,6,7`

```
PS C:\Users\Administrator> New-NetQosTrafficClass -Name iSCSI -BandwidthPercentage 50 -Algorithm ETS -Priority 4

Name                      Algorithm Bandwidth(%) Priority
----                      --------- ------------ --------
iSCSI                     ETS       50           4

PS C:\Users\Administrator> New-NetQosTrafficClass -Name OTHER -BandwidthPercentage 50 -Algorithm ETS -Priority 0,1,2,3,5
,6,7

Name                      Algorithm Bandwidth(%) Priority
----                      --------- ------------ --------
OTHER                     ETS       50           0-3,5-7
```

4. Enable PFC for the iSCSI traffic.
   `Enable-NetQosFlowControl -Priority 4`
5. Create a QoS Policy for iSCSI priority class 4.
   `New-NetQosPolicy -Name iSCSI -iSCSI -PriorityValue8021Action 4`

6. Run the PowerShell command **Get-NetAdapterQos** and verify the following settings as shown below.
   - Operational traffic class: 50% of the ETS bandwidth is allocated to priority class 4 and 50% of the bandwidth is allocated to priority classes 0-3 and 5-7.
   - Operational flow control: Enable for Priority 4.
   - Also verify that the operational traffic class matches with the remote traffic class.
   - Refer to Table 6 for a list of other PowerShell commands that can be used as needed.

```
PS C:\Users\Administrator> Get-NetAdapterQos


Name                        : NIC1
Enabled                     : True
Capabilities                :                        Hardware     Current
                                                     --------     -------
                              MacSecBypass         : NotSupported NotSupported
                              DcbxSupport          : CEE          CEE
                              NumTCs(Max/ETS/PFC)  : 3/3/3        3/3/3

OperationalTrafficClasses   : TC TSA    Bandwidth Priorities
                              -- ---    --------- ----------
                               1 ETS    50%       4
                               2 ETS    50%       0-3,5-7

OperationalFlowControl      : Priority 4 Enabled
OperationalClassifications  : Protocol  Port/Type Priority
                              --------  --------- --------
                              Default             0
                              TCP       3260      4

RemoteTrafficClasses        : TC TSA    Bandwidth Priorities
                              -- ---    --------- ----------
                               1 ETS    50%       4
                               2 ETS    50%       0-3,5-7

RemoteFlowControl           : Priority 4 Enabled
RemoteClassifications       : Protocol  Port/Type Priority
                              --------  --------- --------
                              TCP/UDP   3260      4
```

Table 6 PowerShell commands to verify DCB

PowerShell Command	Comments
`Get-NetQosTrafficClass`	Verifies the traffic classes and Bandwidth percentage allocated
`Get-NetQosFlowControl`	Verifies that PFC is enabled for Priority 4
`Get-NetQosPolicy`	Verifies the QoS policy for Priority 4
`Get-NetAdapterQos`	Verifies the settings on all network adapters

D Configuring Broadcom network adapter as iSCSI hardware initiator using BACS

This section outlines the steps to configure NPAR on a Broadcom 57810 network adapters using BACS, as well as configuring the iSCSI and NDIS network adapters.

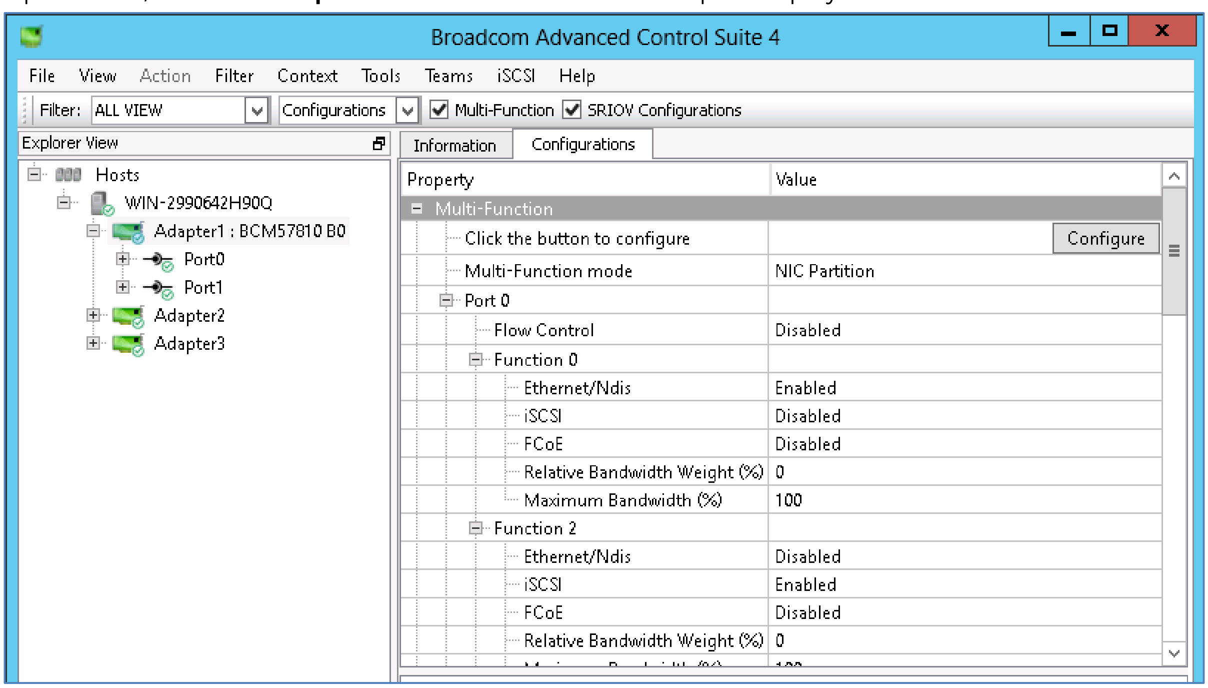
Prerequisites:

- Server has Broadcom network adapter 57810 or similar installed.
- Latest Broadcom firmware and drivers are installed on the system.
- Latest Broadcom Advance Controller Suite for Windows is installed.

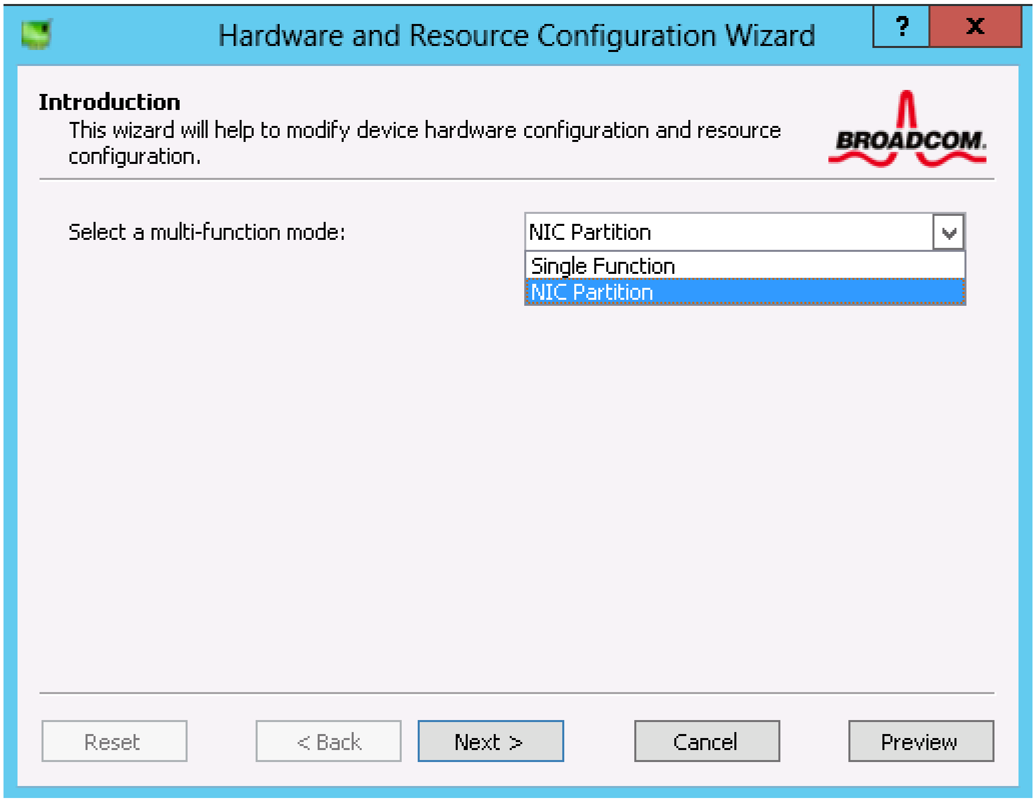
D.1 Configuring NPAR using BACS

Note: NPAR settings can also be configured using the BIOS settings on a Dell PowerEdge server.

1. Open BACS, click the **Adapter #: BCM57810** network adapter displayed in the BACS widow.

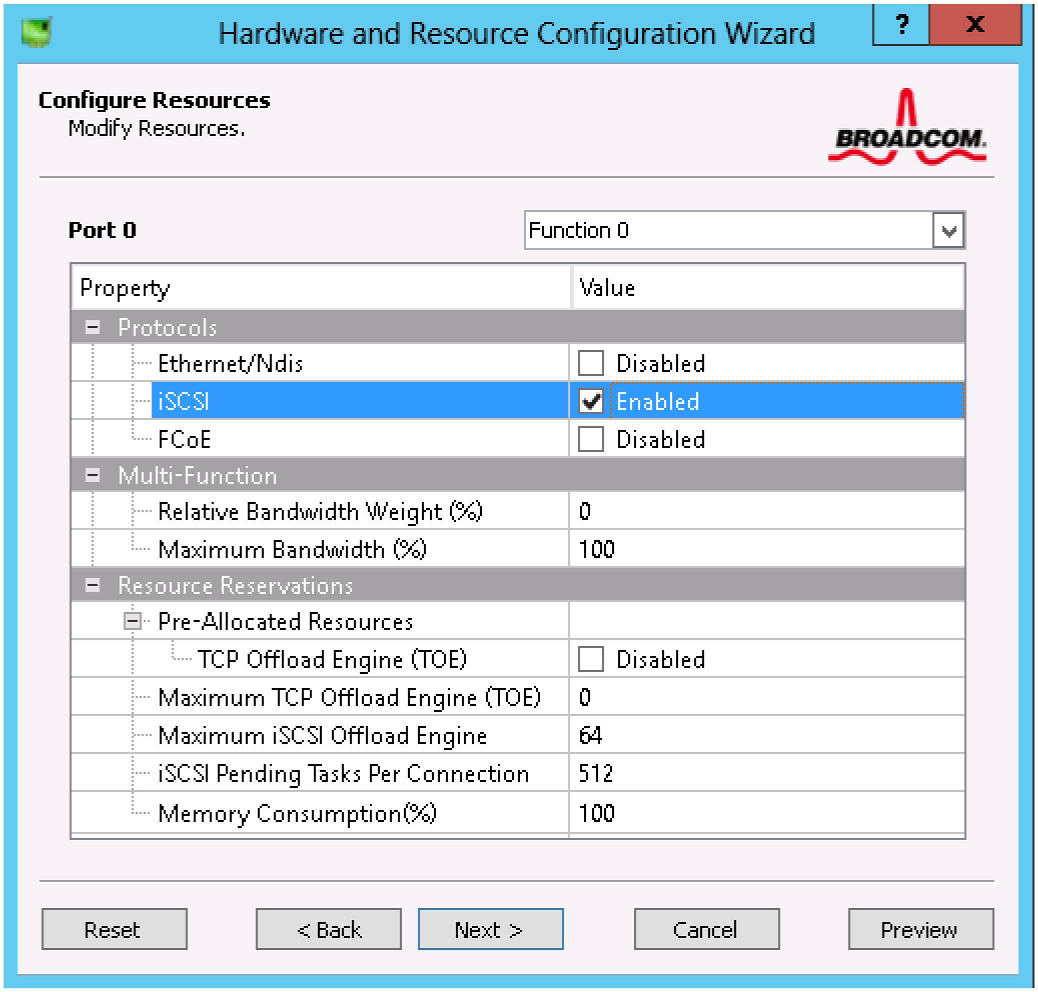


2. Configure NPAR by clicking the **Configuration** tab, and then the **Configure** button under **Multi-Function**.



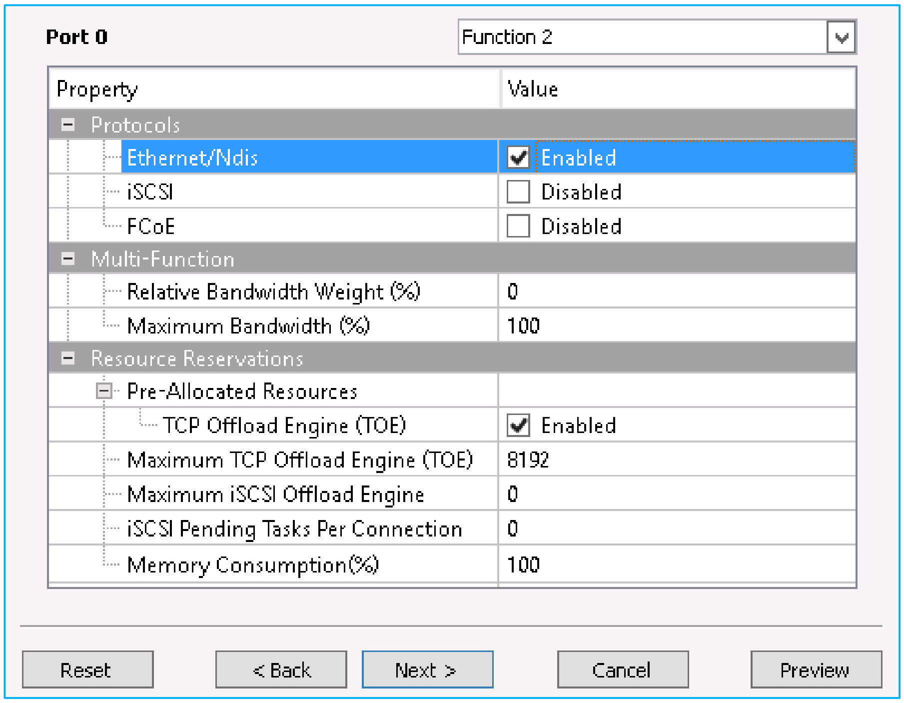
3. In **Hardware and Resource Configuration Wizard**, select **NIC Partition** as the Multi-function mode to create a logical partition of the network port. Click **Next**.

4. Select the appropriate port number and click **Next**. Select function 0 for port 0 and function 1 for port 1, and then chose iSCSI protocol to enable the iSCSI Offload Engine for the iSCSI hardware initiator.



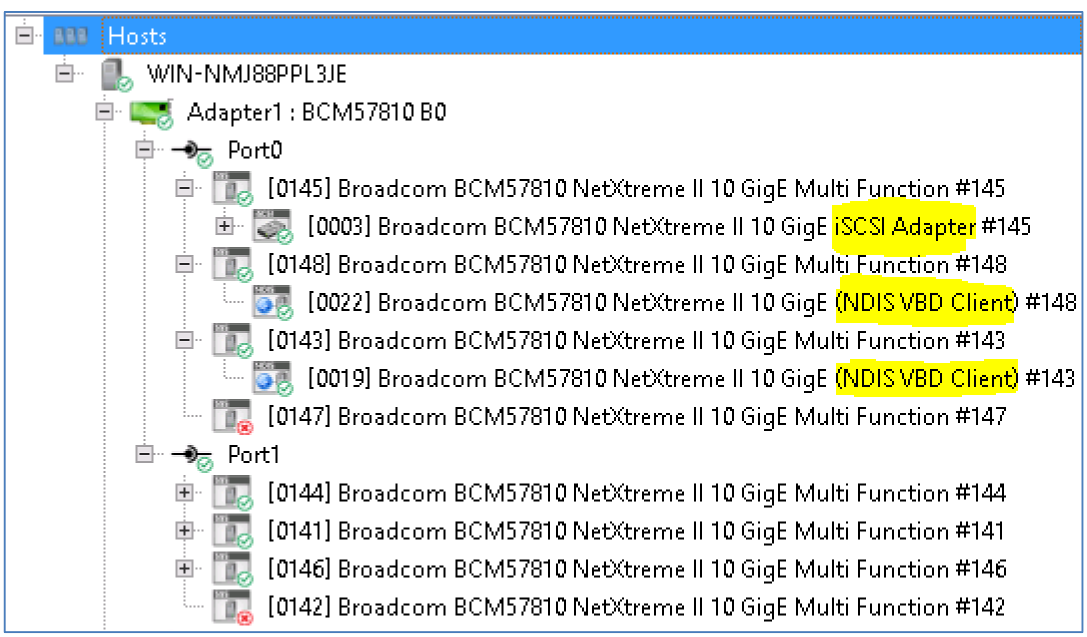
5. Select functions 2/3, 4/5 and 6/7 and set the protocols as needed. The testing presented in this paper used functions 2/3 and 4/5 as NDIS partitions for Ethernet network adapters with TOE enabled.

Note: The test configuration used for the Hyper-V 3.0 in section 8 uses two NDIS functions (2/3 and 4/5) with Ethernet protocol. The Windows 2008 R2 and Windows 2012 configurations in section 6 only used one function with Ethernet protocol.



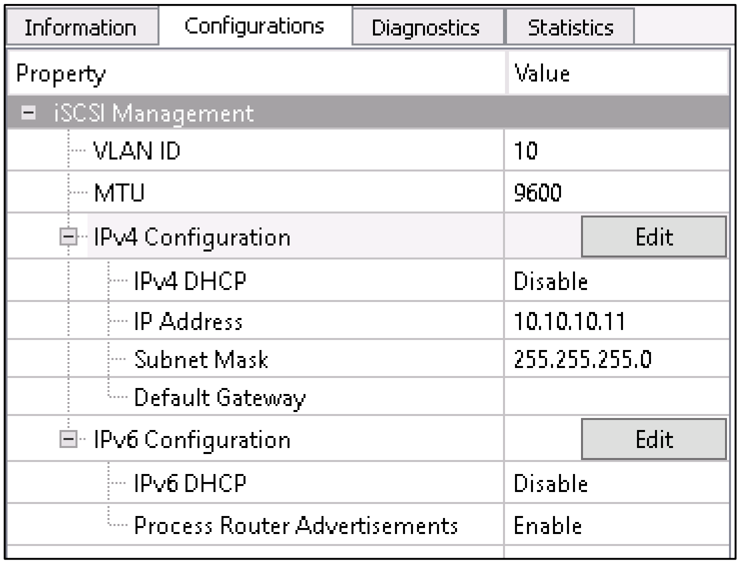
6. Click **Next** and select another port.
7. Once completed, click **Apply** to save the configuration.
8. A reboot might be needed for the changes to take effect.

Once the changes take effect, Port 0 and Port 1 display four Functions. Function 0/1 was configured as an iSCSI adapter, functions 2/3 and 4/5 as NDIS Ethernet adapters, and function 6/7 was not configured.



D.2 Configure iSCSI and NDIS adapters

1. To configure function 0/1 (iSCSI adapter), click **Broadcom BCM57810 NetXtreme II 10 GigE iSCSI Adapter #** in the left window of BACS.
2. Click the **Configurations** tab. Configure **VLAN ID** (VLAN 10), **MTU** size (9600) and static **IP address** for the iSCSI adapter.
3. The Broadcom 57810S CNA defaults to willing mode for DCBx and will accept the configuration from the MXL switch.



4. To configure functions 2/3 and 4/5 (NDIS VDB Client), click on **Broadcom BCM57810 NetXtreme II 10 GigE (NDIS VDB Client) #** in left window of BACS, and then select the **Configuration** tab. Expand the **Advanced** list and configure **VLAN ID** (VLAN 20 for function 2/3 and VLAN 21 for function 4/5). Change the **Receive Buffer** and **Transmit Buffers** to maximum. All other settings can be left as default.
5. Go to the **Control Panel** on the Windows OS and click **Network and Internet.** Go to the **Network and Sharing Center** to configure the **IP address** for the adapters used for LAN traffic.

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 Customers and Dell employees for the purpose of sharing knowledge, best practices, and information about Dell products and your installations.

Referenced or recommended Dell publications:

- Dell EqualLogic Configuration Guide:
  http://en.community.dell.com/dell-groups/dtcmedia/m/mediagallery/19852516/download.aspx
- Dell Networking MXL Switch Configuration Guide
  http://en.community.dell.com/techcenter/storage/w/wiki/4250.switch-configuration-guides-by-sis.aspx
- Best Practices for Configuring DCB with VMware ESXi 5.1 and Dell EqualLogic Storage
  http://en.community.dell.com/techcenter/extras/m/white_papers/20437313/download.aspx
- Dell EqualLogic Storage with Heterogeneous Virtualized Workloads on Microsoft Windows Server 2012 with Hyper-V
  http://en.community.dell.com/techcenter/extras/m/white_papers/20437010/download.aspx
- Data Center Bridging: Standards, Behavioral Requirements, and Configuration Guidelines
  http://en.community.dell.com/techcenter/storage/w/wiki/4396.data-center-bridging-standards-behavioral-requirements-and-configuration-guidelines-by-sis.aspx

For EqualLogic best practices white papers, reference architectures, and sizing guidelines for enterprise applications and SANs, refer to Storage Infrastructure and Solutions Team Publications at:

- http://dell.to/sM4hJT

Other Resources:

- **Windows Server 2012 DCB windows PowerShell user scripting guide.**
  http://technet.microsoft.com/en-us/library/jj573093.aspx

This white paper is for informational purposes only. The content is provided as is, without express or implied warranties of any kind.