EqualLogic Best Practices for using Dell Force10 Leaf-Spine Architecture

A Dell EqualLogic Best Practices Technical White Paper

Dell Storage Engineering
April 2013

Table of contents

Acknowledgements

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

Engineering: Gregory Brent Seaton

Technical Marketing: Guy Westbrook

Editing: Camille Daily

Additional contributors: Matthew Harris, Michael Kosacek, Robert Spear, Steve Williamson, Jaiwant Virk, Raja Jayakumar, and Kirt Gillum

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

1 Introduction

This paper introduces how to properly design and include Dell EqualLogic Storage into the highly modular and scalable Dell Force10 spine-leaf architecture. There is a focus on a simple topology consisting of the high speed spine and two leaf nodes. The configuration of each leaf node is worked through independently, and then the complete topology is shown coming together with the spine to form a distributed core fabric.

1.1 Audience

This white paper is primarily intended for those who are involved in defining, implementing, or supporting an EqualLogic storage network utilizing the Dell Force10 Spine & Leaf architecture. This document assumes the reader is familiar with EqualLogic storage operation and general networking fundamentals.

1.2 Terminology

The following list provides a brief explanation of the terminology used throughout this white paper.

Converged network adapter (CNA) – A network adapter that allows classification of multiple traffic types on a single physical port, in this case, separate classification of iSCSI and other LAN traffic. CNA is usually represented by two different instances of a single physical port within the operating system.

Data Center Bridging (DCB) – Provides a set of extensions to traditional Ethernet with the goal of providing lossless Ethernet.

Jumbo Frame – Any Ethernet frame with an MTU greater than 1500.

Maximum transmission unit (MTU) – The maximum allowed size of an Ethernet frame.

Open Shortest Path First (OSPF) – Standards based dynamic routing protocol.

Virtual Router Redundancy Protocol (VRRP) – A first hop redundancy protocol for layer 3 gateway addresses

2 Dell Force10 leaf-spine architecture overview

The traditional data center architecture based on core chassis switches has served the data center well for traditional north-south traffic scenarios (i.e. Web farms) however it presents a number of concerns in a modern data center with the increase in east-west traffic (i.e. distributed computer clusters).

The Force10 leaf-spine architecture, also called an Active Fabric, provides full bisectional bandwidth between any two nodes in a larger cluster using high capacity low-cost Ethernet switches. This architecture is purpose built for the scenario with interchangeable leaves (1 Gb, 10 Gb, or even 40 Gb) which allows high performance networking fabrics that scale to meet the largest challenges.

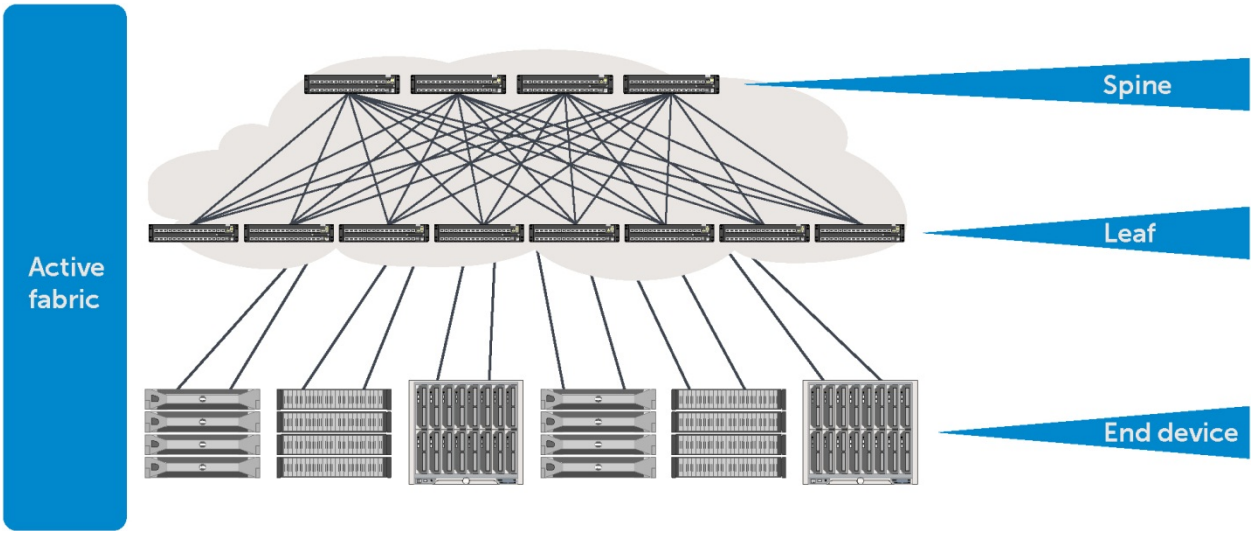

Figure 1 Dell Active Fabric

This white paper will run through a simplified leaf-spine architecture consisting of a high speed 40 Gb spine with two leaf nodes, one 10 Gb leaf and one 1 Gb leaf. The configuration described in Section 3 provides room for expansion in several ways, the two main options are:

- Expansion of current leaf nodes – The leaf nodes configured in this white paper provide room for other devices to be connected. This includes additional EqualLogic storage arrays to grow the iSCSI SAN in that leaf and/or the addition of end devices to utilize the current EqualLogic iSCSI SAN in that leaf.
- Additional leaf nodes – The high speed spine in this white paper consists of Force10 Z9000 40 Gb switches capable of supporting many more leaf nodes. This allows the topology to grow as the need for additional network capacity arises. For example, the addition of Leaf03 could be a high computing performance cluster with many rack servers, Leaf04 could be a fully self-contained leaf consisting of Dell Blade chassis with integrated compute, storage, and switching. The possibilities are endless.

For more information or help designing a Force10 leaf-spine architecture see the Dell Fabric Manager (DFM). A video overview of DFM can be found at http://en.community.dell.com/techcenter/networking/w/wiki/3880.dell-fabric-manager.aspx

3 Test configuration

The networking infrastructure discussed in this paper consists of two Force10 Z9000 switches representing the spine, two Force10 S4810 switches representing a leaf node (Leaf01), and two Force10 S60 switches representing a second leaf node (Leaf02).

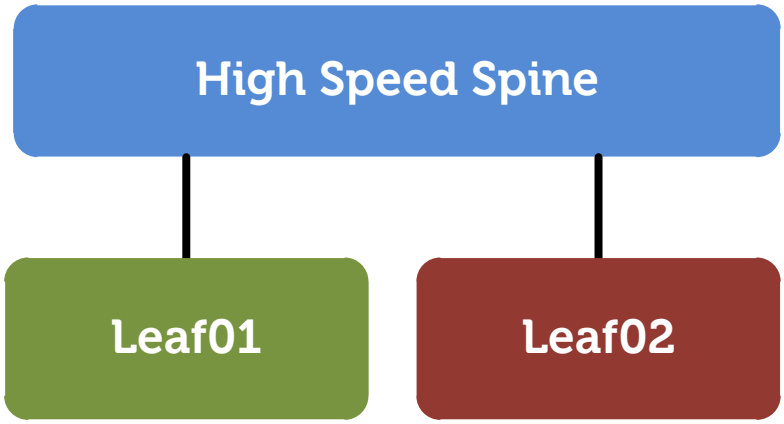

Figure 2 High level topology

Leaf01 and Leaf02 are connected to the spine through multiple high speed 10 Gb and 40 Gb links, the configuration is detailed below.

3.1 iSCSI separation

iSCSI hosts and storage arrays should be physically connected to the same pair of switches when possible. Leaf-spine architecture achieves this by containing iSCSI hosts with the adjoining iSCSI arrays to the same leaf (physical separation) and VLAN (virtual separation, only allowed with DCB support) as demonstrated in the following scenario. This allows for the lowest possible latency and reduces possible congestion issues when traversing the spine.

3.1.1 VLANs and IP address scheme

The following table lists the VLANs and IP subnets used in the leaf nodes and the traffic type each one transported.

VLAN	IP Network	Subnet	Leaf Node	Use
10	192.168.10.0	/24	Leaf01	iSCSI data traffic
11	192.168.11.0	/24	Leaf01	LAN traffic
12	192.168.12.0	/24	Leaf02	iSCSI data traffic

3.2 Leaf01 – Force10 S4810

Leaf01 is connected to three EqualLogic PS6110 iSCSI storage arrays and three Dell™ PowerEdge™ R815 servers with Broadcom® 57810 dual port 10 Gb CNAs.

Refer to Appendix C for information about configuring Broadcom 57810 CNAs.

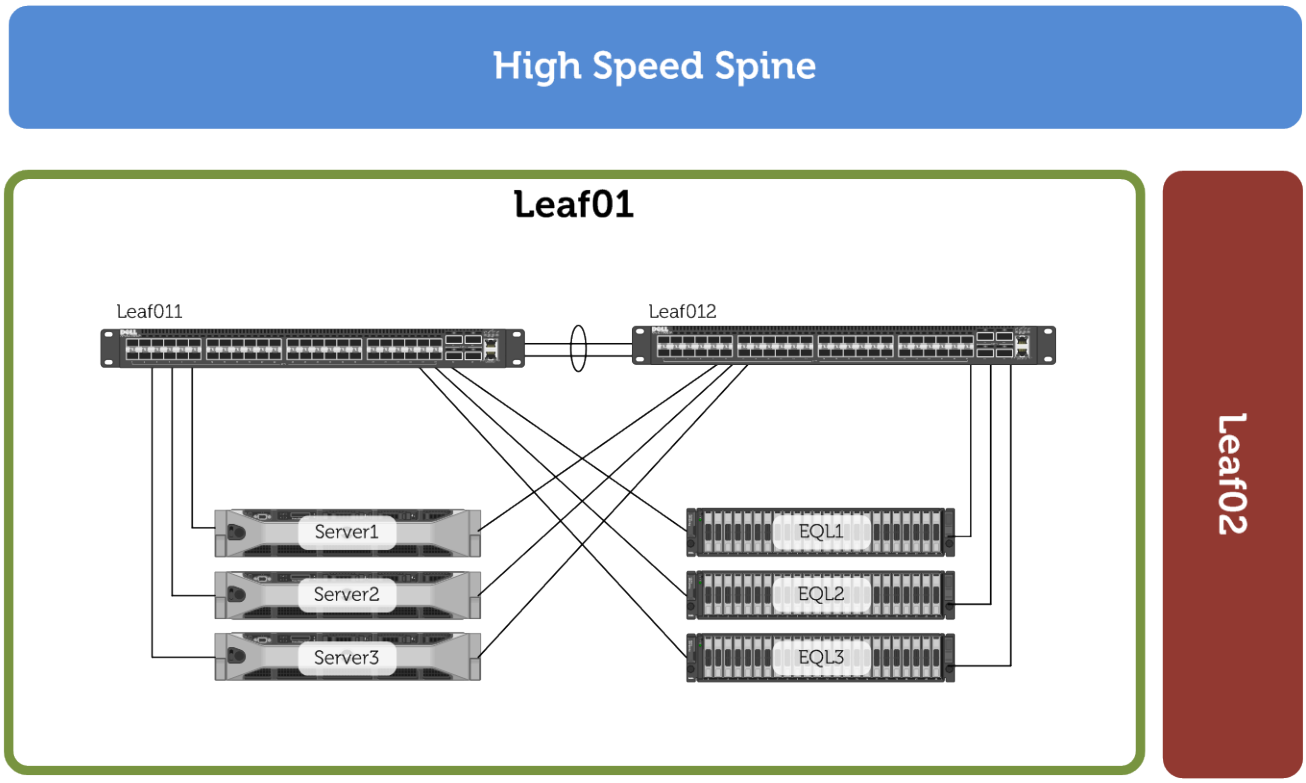

Figure 3 Leaf01 topology

3.2.1 Switch configuration

The following section explains the proper way to cable, enable and configure DCB, set VLAN and IP information, and enable OSPF on a Force10 S4810 series switch.

3.2.1.1 Layer 1 configuration - switch connections

The following tables show the port mapping for each of the Force10 S4810 switches in Leaf01. All connections within Leaf01 were made with Dell twinax cables. Connections to the spine were made using QSFP+ Twinax and are listed in the tables below with configuration commands detailed in Section 3.4

Table 1 Leaf011-S4810 port mapping

Switch1 (Leaf011-s4810)		
Port 0	connects to	Server 1 port 1
Port 1	connects to	Server 2 port 1
Port 2	connects to	Server 3 port 1
Port 3	connects to	EqualLogic 1 Controller 1 eth0
Port 4	connects to	EqualLogic 2 Controller 1 eth0
Port 5	connects to	EqualLogic 3 Controller 1 eth0
Port 48	connects to	Spine01-z9000 port 0
Port 52	connects to	Spine02-z9000 port 0
Port 56	connects to	Switch2 (Leaf012-s4810) port 56
Port 60	connects to	Switch2 (Leaf012-s4810) port 60

Table 2 Leaf012-s4810 port mapping

Switch2 (Leaf012-s4810)		
Port 0	connects to	Server 1 port 2
Port 1	connects to	Server 2 port 2
Port 2	connects to	Server 3 port 2
Port 3	connects to	EqualLogic 1 Controller 2 eth0
Port 4	connects to	EqualLogic 2 Controller 2 eth0
Port 5	connects to	EqualLogic 3 Controller 2 eth0
Port 48	connects to	Spine01-z9000 port 4
Port 52	connects to	Spine02-z9000 port 4
Port 56	connects to	Switch1 (Leaf011-s4810) port 56
Port 60	connects to	Switch1 (Leaf011-s4810) port 60

3.2.1.2 Layer 2 configuration

Following best practices for Data Center Bridging requires the creation of multiple VLANs when running multiple traffic types on the same physical line connected to the physical host. In this example:

- All 10Gb ports are placed in Layer 2 (switchport) mode.
- An inter-switch port channel is created (VLAN 10 and 11 for iSCSI and OTHER LAN traffic respectively),
- Descriptions are applied (optional but highly recommended), and
- The appropriate ports for VLAN tagging are set.

Note: All Commands below are for switch1 (Leaf011-s4810) and should be repeated for switch2 (Leaf012-s4810), any exceptions are called out.

Set switchport mode

```
Leaf011-S4810#conf
Leaf011-S4810(conf)#interface range tengigabitethernet 0/0 – 47
Leaf011-S4810(conf-if-range-te-0/0-47)#switchport
Leaf011-S4810(conf-if-range-te-0/0-47)#no shutdown
Leaf011-S4810(conf-if-range-te-0/0-47)#exit
Leaf011-S4810(conf)#exit
```

Port-Channel configuration

```
Leaf011-s4810#conf
Leaf011-s4810(conf)#interface fortyGigE 0/56
Leaf011-s4810(conf-if-fo-0/56)#mtu 12000
Leaf011-s4810(conf-if-fo-0/56)#port-channel-protocol lacp
Leaf011-s4810(conf-if-fo-0/56-lacp)#port-channel 3 mode active
Leaf011-s4810(conf-if-fo-0/56-lacp)#no shut
Leaf011-s4810(conf-if-fo-0/56)#exit
Leaf011-s4810(conf)#interface fortyGigE 0/60
Leaf011-s4810(conf-if-fo-0/60)#mtu 12000
Leaf011-s4810(conf-if-fo-0/60)#port-channel-protocol LACP
Leaf011-s4810(conf-if-fo-0/60-lacp)#port-channel 3 mode active
Leaf011-s4810(conf-if-fo-0/60-lacp)#no shutdown
Leaf011-s4810(conf-if-fo-0/60)#exit
Leaf011-s4810(conf)#interface Port-channel 3
Leaf011-s4810(conf-if-po-3)# mtu 12000
Leaf011-s4810(conf-if-po-3)# switchport
Leaf011-s4810(conf-if-po-3)# no shutdown
Leaf011-s4810(conf-if-po-3)#exit
```

VLAN creation and tagging

In the previous section, the port channel is created and placed in layer 2 mode. This section allows tagged traffic over that inter-switch link connection to the two S4810 switches.

```
Leaf011-s4810#conf
Leaf011-s4810(conf)#int vlan 10
Leaf011-s4810(conf-if-vl-10)#description iSCSI_VLAN
Leaf011-s4810(conf-if-vl-10)#mtu 12000
Leaf011-s4810(conf-if-vl-10)#tagged TenGigabitEthernet 0/0-5
Leaf011-s4810(conf-if-vl-10)#tagged Port-channel 3
Leaf011-s4810(conf-if-vl-10)#exit
Leaf011-s4810(conf)#int vlan 11
Leaf011-s4810(conf-if-vl-11)#description LAN_VLAN
Leaf011-s4810(conf-if-vl-11)#mtu 12000
Leaf011-s4810(conf-if-vl-11)#tagged TenGigabitEthernet 0/0-2
Leaf011-s4810(conf-if-vl-11)#tagged Port-channel 3
Leaf011-s4810(conf-if-vl-11)#exit
```

Notice how ports TenGigabitEthernet 0/0-2 are set as tagged in VLAN 10 and 11, this works in conjunction with Data Center Bridging, the iSCSI TLV, and the host network adapter to separate iSCSI traffic from other LAN traffic allowing each to be prioritized separately.

DCB configuration

The following commands configure Data Center Bridging on the Force10 S4810 to enable PFC on priority group 4. iSCSI traffic was placed in the "iSCSI" Priority group with a priority of 4, while the group "OTHER" holds all the remaining priority tags. Bandwidth percentages were allocated to each priority group. To activate this configuration a switch reload (reboot) is required. Save all configuration changes before reloading.

```
Leaf011-s4810#configure
Leaf011-s4810(conf)#dcb-input pfc
Leaf011-s4810(conf-dcb-in)#pfc priority 4
Leaf011-s4810(conf-dcb-in)#exit
Leaf011-s4810(conf)#priority-group iSCSI
Leaf011-s4810(conf-pg)#priority-list 4
Leaf011-s4810(conf-pg)#set-pgid 1
Leaf011-s4810(conf-pg)#exit
Leaf011-s4810(conf)#priority-group OTHER
Leaf011-s4810(conf-pg)#priority-list 0-3,5-7
Leaf011-s4810(conf-pg)#set-pgid 2
Leaf011-s4810(conf-pg)#exit
Leaf011-s4810(conf)#dcb-output ets
Leaf011-s4810(conf-dcb-out)#priority-group iSCSI qos-policy iSCSI
Leaf011-s4810(conf-dcb-out)#priority-group OTHER qos-policy OTHER
Leaf011-s4810(conf-dcb-out)#exit
```

```
Leaf011-s4810(conf)#service-class dynamic dot1p
Leaf011-s4810(conf)#qos-policy-output iSCSI ets
Leaf011-s4810(conf-qos-policy-out-ets)#bandwidth-percentage 30
Leaf011-s4810(conf-qos-policy-out-ets)#exit
Leaf011-s4810(conf)#qos-policy-output OTHER ets
Leaf011-s4810(conf-qos-policy-out-ets)#bandwidth-percentage 70
Leaf011-s4810(conf-qos-policy-out-ets)#exit
Leaf011-s4810(conf)#dcb enable
Leaf011-s4810(conf)#dcb stack-unit all pfc-buffering pfc-ports 56 pfc-queues 2
```

Applying DCB settings to every port in the switch

These commands are used to apply the DCB settings defined in the previous section to all ports on the switch.

```
Leaf011-s4810(conf)#interface range tengigabitethernet 0/0 - 47
Leaf011-s4810(conf-if-range-te-0/0-47)#dcb-policy input pfc
Leaf011-s4810(conf-if-range-te-0/0-47)#dcb-policy output ets
Leaf011-s4810(conf-if-range-te-0/0-47)#exit
Leaf011-s4810(conf)#interface range fortyGigE 0/48 , fortyGigE 0/52 , fortyGigE 0/56 ,
fortyGigE 0/60
Leaf011-s4810(conf-if-range-fo-0/48,fo-0/52,fo-0/56,fo-0...)#dcb-policy input pfc
Leaf011-s4810(conf-if-range-fo-0/48,fo-0/52,fo-0/56,fo-0...)#dcb-policy output ets
Leaf011-s4810(conf-if-range-fo-0/48,fo-0/52,fo-0/56,fo-0...)#exit
```

Note: For more information on configuring the Force10 S4810 for DCB, refer to the *Dell Force10 S4810 Switch Configuration Guide* at http://en.community.dell.com/dell-groups/dtcmedia/m/mediagallery/20220824/download.aspx.

Enable management Link Layer Discovery Protocol (optional)

Many devices are able to gather information about their neighboring devices using Link Layer Discovery Protocol (LLDP). Enabling the switches to advertise the following information simplifies management; however, in security sensitive networks this may not be desired.

```
Leaf011-s4810(conf)#protocol lldp
Leaf011-s4810(conf-lldp)#advertise management-tlv system-capabilities
Leaf011-s4810(conf-lldp)#advertise system-description system-name
Leaf011-s4810(conf-lldp)#advertise interface-port-desc
Leaf011-s4810(conf-lldp)#exit
```

3.2.1.3 Layer 3 configuration

The objective with the Layer 3 commands is to configure a gateway IP address on each VLAN, make it highly available, and then place each subnet into OSPF to be advertised to the other leaf nodes. Advertising each subnet to other leaves will allow IP connectivity between all leaf nodes which, in this example, is required for replication from the remote SAN in Leaf02.

VLAN gateway IP address

A gateway IP address is used by an end device to reach addresses beyond the reach of the local subnet.

```
Leaf011-s4810#conf
Leaf011-s4810(conf)#int vlan 10
Leaf011-s4810(conf-if-vl-10)# ip address 192.168.10.2/24
Leaf011-s4810(conf-if-vl-10)#no shut
Leaf011-s4810(conf-if-vl-10)#exit
Leaf011-s4810(conf)#interface Vlan 11
Leaf011-s4810(conf-if-vl-11)#ip address 192.168.11.2/24
Leaf011-s4810(conf-if-vl-11)#no shut
Leaf011-s4810(conf-if-vl-11)#exit
```

Note: In this example Switch2 (Leaf012-S4810) should have IP addresses 192.168.10.**3**/24 and 192.168.11.**3**/24

First hop redundancy protocol – VRRP

VRRP allows for the virtualization of the gateway IP address. This is configured to guarantee that the gateway IP address remains available should either switch1 (Leaf011-S4810) or switch2 (Leaf012-S4810) become unavailable.

```
Leaf011-s4810#conf
Leaf011-s4810(conf)#int vlan 10
Leaf011-s4810(conf-if-vl-10)#vrrp-group 10
Leaf011-s4810(conf-if-vl-10-vrid-10)#virtual-address 192.168.10.1
Leaf011-s4810(conf-if-vl-10-vrid-10)#no shut
Leaf011-s4810(conf-if-vl-10)#exit
Leaf011-s4810(conf)#int vlan 11
Leaf011-s4810(conf-if-vl-11)#vrrp-group 11
Leaf011-s4810(conf-if-vl-11-vrid-11)#virtual-address 192.168.11.1
Leaf011-s4810(conf-if-vl-11-vrid-11)#no shut
Leaf011-s4810(conf-if-vl-11)#exit
```

Dynamic routing

Using dynamic routing allows every router to advertise the subnets it knows how to reach with each of its neighboring routers. This provides each router with a complete picture of the network topology creating a more resilient network. This topology uses the standards based dynamic routing protocol Open Shortest Path First (OSPF).

```
Leaf011-s4810#conf
Leaf011-s4810(conf)#router ospf 1
Leaf011-s4810(conf-router_ospf-1)# network 192.168.10.0/24 area 0
Leaf011-s4810(conf-router_ospf-1)# network 192.168.11.0/24 area 0
Leaf011-s4810(conf-router_ospf-1)#exit
```

3.2.2 Workload

Table 3 below shows the breakdown for the workload applied to the EqualLogic storage in Leaf01. Each test was run for a total of nine hours to ensure the configuration exhibited expected stability and consistent performance over an extended period of time. For each test cycle of Leaf01, each workload was run in the order and for the duration shown.

Table 3 Workload patterns

IO Pattern	Block Size	Read/Write Ratio	Duration
Random	8 KB	67/33	3 Hours
Sequential Read	256 KB	100 / 0	3 Hours
Sequential Write	64 KB	0 / 100	3 Hours

Sample storage load generation configuration files can be found in Appendix D.

3.2.3 Results

Test results of Leaf01 in this configuration showed 0% TCP retransmit rate during the entire run of the test as shown in the graph below. Notice how this does not change when the workload demand placed on the array changes.

Figure 4 Leaf01 - TCP retransmit percentage

The following graph shows the average latency for the random workload of 20 ms which drops below 10 ms for heavy sequential reads, and then lower for sequential write.

Figure 5 Leaf01 - Average latency

The results above show no TCP retransmissions implying a clean running network; the average latency of 20ms or below also corresponds to a network with ample resources to handle the load.

3.3 Leaf02 – Force10 s60

Leaf02 utilized one EqualLogic PS6100 iSCSI storage array and two PowerEdge R815 servers with the onboard Broadcom network ports.

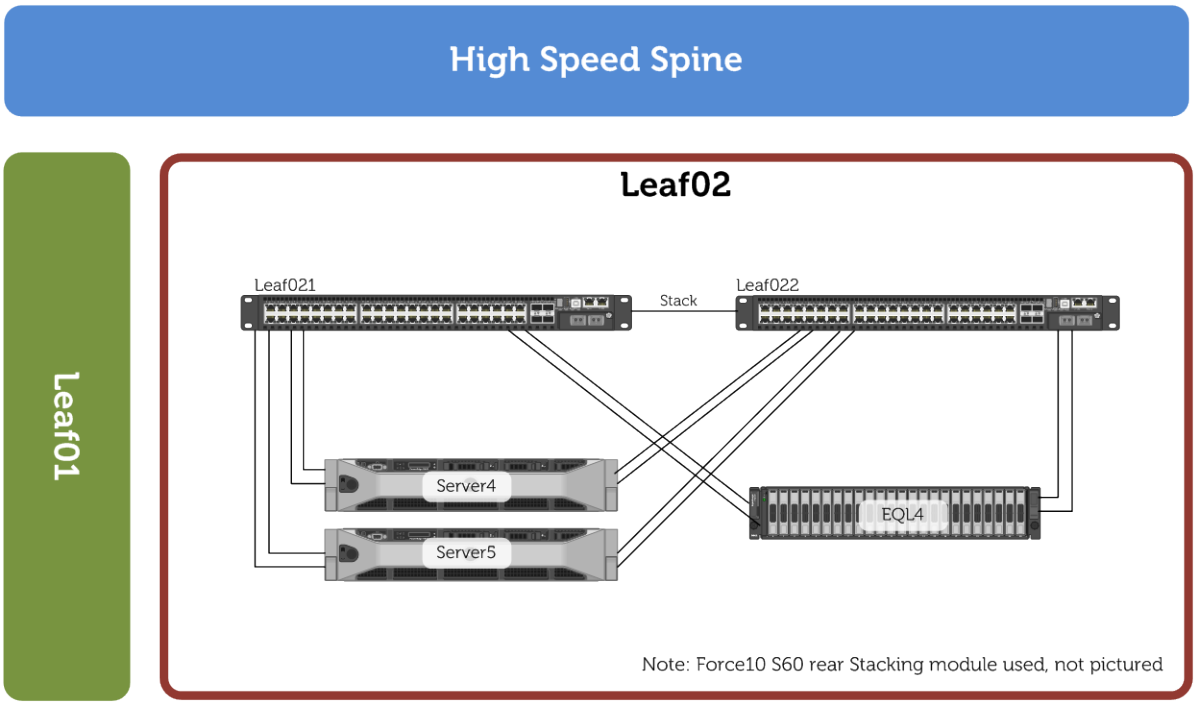

Figure 6 Leaf02 topology

3.3.1 Switch configuration

Leaf02 consisted of two Force10 S60 units stacked together with a high speed 24 Gb interconnect that allowed for the simplification of configuration and management. When configuring Leaf02 the same theme of separation of iSCSI traffic is followed. This calls for the use of dedicated switches and the creation of one VLAN for iSCSI traffic. In this case, VLANs 12 was used.

Note: This configuration could be modified to use an additional dual port uplink module configured in a link aggregation group (LAG) instead of the proprietary stacking module. With future expansion being a concern the high speed stacking module was chosen for its higher interconnect bandwidth of 24 Gb versus 20 Gb with an uplink module.

3.3.1.1 Layer 1 configuration – switch connections

The following tables show the port mapping for each of the Force10 S60 switches used in Leaf02. All end device connections were made using standard category 6 cables. The stacking connections were created using the Force10 proprietary stacking cables and the connections to the spine were made using one 40 Gb to four 10 Gb breakout cables.

Table 4 Leaf021-S60 port mapping

Switch1 (Leaf021-S60)		
Port 0	connects to	Server 4 port 1
Port 1	connects to	Server 4 port 2
Port 2	connects to	Server 5 port 1
Port 3	connects to	Server 5 port 2
Port 4	connects to	EqualLogic Controller 1 eth0
Port 5	connects to	EqualLogic Controller 2 eth1
Port 6	connects to	EqualLogic Controller 1 eth2
Port 7	connects to	EqualLogic Controller 2 eth3
Port 48	connects to	Spine01-z9000 Port 8 - A
Port 49	connects to	Spine02-z9000 Port 8 - A
Stack1	connects to	Switch2 (Leaf022-s60) stack1
Stack2	connects to	Switch2 (Leaf022-s60) stack2

Table 5 Leaf022-S60 port mapping

Switch2 (Leaf022-S60)		
Port 0	connects to	Server 4 port 3
Port 1	connects to	Server 4 port 4
Port 2	connects to	Server 5 port 3
Port 3	connects to	Server 5 port 4
Port 4	connects to	EqualLogic Controller 2 eth0
Port 5	connects to	EqualLogic Controller 1 eth1
Port 6	connects to	EqualLogic Controller 2 eth2
Port 7	connects to	EqualLogic Controller 1 eth3
Port 48	connects to	Spine01-z9000 Port 8 - B
Port 49	connects to	Spine02-z9000 Port 8 - B
Stack1	connects to	Switch1 (Leaf021-s60) stack1
Stack2	connects to	Switch1 (Leaf021-s60) stack2

3.3.1.2 Layer 2 configuration

The following Layer 2 configuration commands were used to create VLAN 12, apply descriptions (optional but highly recommended), and place the iSCSI ports in the correct VLAN.

```
Leaf02x-s60#conf
Leaf02x-s60(conf)#int vlan 12
Leaf02x-s60(conf-if-vl-12)#description iSCSI_VLAN
Leaf02x-s60(conf-if-vl-12)#untagged GigabitEthernet 0/1,3-7
Leaf02x-s60(conf-if-vl-12)#untagged GigabitEthernet 1/1,3-7
Leaf02x-s60(conf-if-vl-13)#exit
```

Note: This configuration differs from the S4810 configuration in Leaf01 by using the untagged command for each port. This is required because DCB is not available on 1 Gb Ethernet. Only one traffic type is sent per physical port (the end device *will not be applying* a VLAN header to every frame).

3.3.1.3 Layer 3 configuration

The objective of the following Layer 3 commands was to configure a gateway IP address on the VLAN and then place the IP subnet into OSPF to be advertised to the other leaf nodes. This allows IP connectivity between every subnet in each leaf and between each leaf which was required for replication back to the main SAN in Leaf01.

Vlan Gateway IP Address

```
Leaf02x-s60(conf)#int vlan 12
Leaf02x-s60(conf-if-vl-12)#ip address 192.168.12.1/24
Leaf02x-s60(conf-if-vl-12)#no shutdown
Leaf02x-s60(conf-if-vl-12)#exit
```

Dynamic Routing

```
Leaf02x-s60(conf)#router ospf 1
Leaf02x-s60(conf-router_ospf-1)#network 192.168.12.0/24 area 0
Leaf02x-s60(conf-router_ospf-1)#exit
```

3.3.2 Workload

The table below shows the breakdown for the workload applied to the EqualLogic storage in Leaf02. Each test was run for a total of nine hours to ensure the configuration exhibited expected stability and consistent performance over an extended period of time. For each test cycle of Leaf02, each workload was run in the order shown, and for the duration shown.

Table 6 Workload pattern

IO Pattern	Block Size	Read/Write Ratio	Duration
Random	8 KB	67/33	3 Hours
Sequential Read	256 KB	100 / 0	3 Hours
Sequential Write	64 KB	0 / 100	3 Hours

Sample storage load generation configuration files can be found in Appendix D of this whitepaper.

3.3.3 Results

When the 1 Gb leaf node is configured as described above, the rate of TCP retransmitted packets in Leaf02 remains nonexistent at 0% as the graph below depicts.

Figure 7 Leaf02 – TCP retransmit rate

The average latency in the 1 Gb leaf node averages around 30 ms during heavy times of read with write latency never above 5 ms.

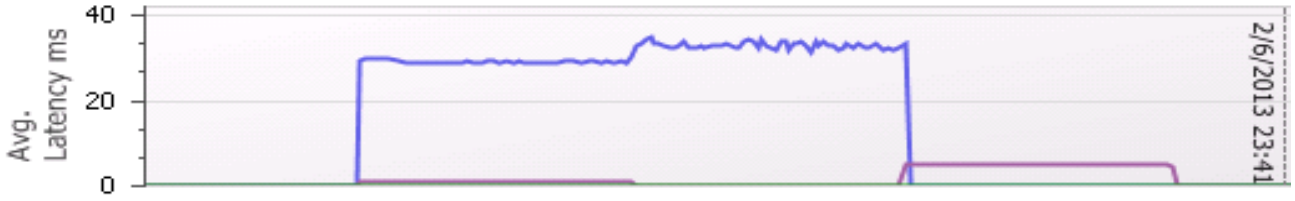

Figure 8 Leaf02 – average latency ms

The results above show no TCP retransmissions implying a clean running network, the average latency of 30ms or below also corresponds to a network with adequate resources to handle the load.

3.4 Spine – Force10 Z9000

The Force10 leaf-spine architecture calls for a high speed spine to interconnect all leaf nodes. In this environment the smallest possible configuration is demonstrated utilizing two Force10 Z9000 switches running the high speed, dynamic routing protocol OSPF to distribute IP routing information between leaf nodes.

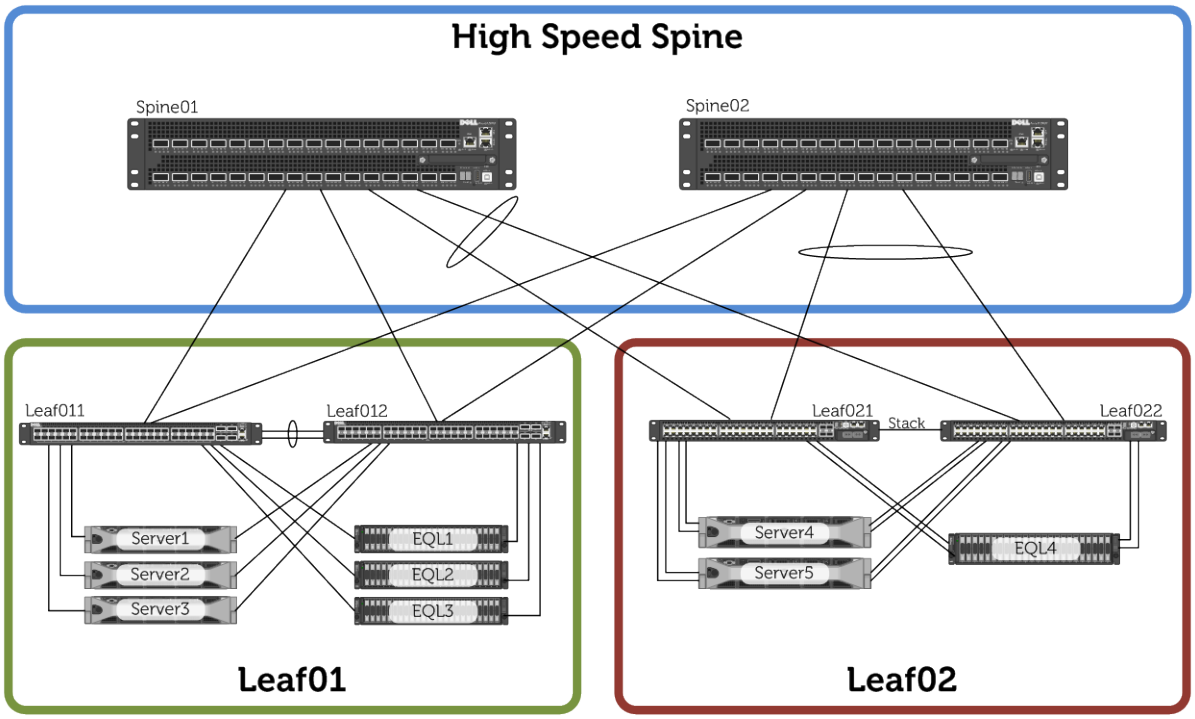

Figure 9 Spine topology

3.4.1 Switch configuration

The high speed spine consists of two Force10 Z9000 units. The prime objective for the spine is to create a series of equal cost, high speed distribution of connections. This is achieved using multiple Layer 3 links with OSPF all in area 0.

3.4.1.1 Layer 1 configuration – switch connections

The following tables show the port mapping for each of the Force10 Z9000 switches in the spine. All 40 Gb to 40 Gb connections were made using QSFP+ Twinax cables. The Z9000 to S60 connections were created using Force10 QSFP+ breakout cables; each provided one 40 Gb to four 10 Gb connections.

Table 7 Spine01-z9000 port mapping

Switch1 (Spine01-Z9000)		
Port 0	connects to	Leaf011-s4810 port 48
Port 4	connects to	Leaf012-s4810 port 48
Port 8 (40Gb to 4 x 10Gb breakout)		
Port 8 - A	connects to	Leaf021-s60 port 48
Port 8 - B	connects to	Leaf022-s60 port 48
Port 8 - C	-	-
Port 8 - D	-	-

Table 8 Spine02-z9000 port mapping

Switch2 (Spine02-Z9000)		
Port 0	connects to	Leaf011-s4810 port 52
Port 4	connects to	Leaf012-s4810 port 52
Port 8 (40Gb to 4 x 10Gb breakout)		
Port 8 - A	connects to	Leaf021-s60 port 49
Port 8 - B	connects to	Leaf022-s60 port 49
Port 8 - C	-	-
Port 8 - D	-	-

3.4.1.2 Layer 2 configuration

There is minimal configuration work needed on the Spine for Layer 2. The following commands show setting the correct MTU on the 40Gb ports in use.

Note: All Commands below are for switch1 (Spine01-Z9000) and should be repeated for switch2 (Spine02-Z9000), any exceptions are called out below.

```
Spine01-z9000#conf
Spine01-z9000(conf)#interface fortyGigE 0/0
Spine01-z9000(conf-if-fo-0/0)#mtu 12000
Spine01-z9000(conf-if-fo-0/0)#exit
Spine01-z9000(conf)#interface fortyGigE 0/4
Spine01-z9000(conf-if-fo-0/4)#mtu 12000
Spine01-z9000(conf-if-fo-0/4)#exit
```

3.4.1.3 Layer 3 configuration

The spine switches provide high speed equal cost connections to each node by relying on OSPF and its equal cost multipathing functionality.

IP Subnetting

Up to this point every subnet shown has been a class C, or /24, subnet providing up to 254 (256 minus 1 for network address and minus 1 for broadcast address) host IP addresses. The underutilization of each subnet has been intentional to allow for the addition and growth of each leaf within the environment. A more resource conscientious approach has been taken for the spine to leaf switch interconnect links, each link will be its own /30 subnet providing one network address, one broadcast address, and two host addresses (one for each end of the link) as shown in the tables below.

Table 9 Spine01-Z9000 connection IP address map

Switch1 (Spine01-z9000) Connections				
Network Address	Host Address 1	Host Address 2	Broadcast Address	Subnet mask
192.168.9.0	192.168.9.1	192.168.9.2	192.168.9.3	255.255.255.252
192.168.9.4	192.168.9.5	192.168.9.6	192.168.9.7	255.255.255.252
192.168.9.8	192.168.9.9	192.168.9.10	192.168.9.11	255.255.255.252

Table 10 Spine01-Z9000 connection IP address map

Switch2 (Spine02-z9000) Connections				
Network Address	Host Address 1	Host Address 2	Broadcast Address	Subnet mask
192.168.9.16	192.168.9.17	192.168.9.18	192.168.9.19	255.255.255.252
192.168.9.20	192.168.9.21	192.168.9.22	192.168.9.23	255.255.255.252
192.168.9.24	192.168.9.25	192.168.9.26	192.168.9.27	255.255.255.252

Setting IP address and creating Layer 3 port channel

The following commands configure the Layer 3 Port-channel and IP Addresses utilized on the Spine in this topology.

```
Spine01-z9000#conf
Spine01-z9000(conf)#interface fortyGigE 0/0
Spine01-z9000(conf-if-fo-0/0)#ip address 192.168.9.1/30
Spine01-z9000(conf-if-fo-0/0)#no shutdown
Spine01-z9000(conf-if-fo-0/0)#exit
Spine01-z9000(conf)#interface fortyGigE 0/4
Spine01-z9000(conf-if-fo-0/4)#ip address 192.168.9.5/30
Spine01-z9000(conf-if-fo-0/4)#no shutdown
Spine01-z9000(conf-if-fo-0/4)#exit
Spine01-z9000(conf)#interface TenGigabitEthernet 0/8
Spine01-z9000(conf-if-te-0/8)#port-channel-protocol LACP
Spine01-z9000(conf-if-te-0/8-lacp)#port-channel 1 mode active
Spine01-z9000(conf-if-te-0/8-lacp)#no shutdown
Spine01-z9000(conf-if-te-0/8)#exit
Spine01-z9000(conf)#interface TenGigabitEthernet 0/9
Spine01-z9000(conf-if-te-0/9)#port-channel-protocol LACP
Spine01-z9000(conf-if-te-0/9-lacp)#port-channel 1 mode active
Spine01-z9000(conf-if-te-0/9-lacp)#no shutdown
Spine01-z9000(conf-if-te-0/9)#exit
Spine01-z9000(conf)#interface Port-channel 1
Spine01-z9000(conf-if-po-1)#ip address 192.168.9.9/30
Spine01-z9000(conf-if-po-1)#no shutdown
Spine01-z9000(conf-if-po-1)#exit
```

Note: The Spine02-Z9000 IP addresses need to match the configuration as shown in Table 10.

Dynamic routing

```
Spine01-z9000(conf)#router ospf 1
Spine01-z9000(conf-router_ospf-1)# network 192.168.9.0/30 area 0
Spine01-z9000(conf-router_ospf-1)# network 192.168.9.4/30 area 0
Spine01-z9000(conf-router_ospf-1)# network 192.168.9.8/30 area 0
Spine01-z9000(conf-router_ospf-1)#exit
```

Note: The Spine02-Z9000 network addresses need to match its configuration. In this example, 192.168.9.16, 192.168.9.20, and 192.168.9.24 were used.

Configuring Leaf01 and Leaf02 uplinks

At this point, IP addresses have been configured for Leaf01 uplinks and placed in OSPF using the commands below.

Leaf011-S4810

```
Leaf011-s4810(conf)#interface fortyGigE 0/48
Leaf011-s4810(conf-if-fo-0/48)#ip address 192.168.9.2/30
Leaf011-s4810(conf-if-fo-0/48)#no shutdown
Leaf011-s4810(conf-if-fo-0/48)#exit
Leaf011-s4810(conf)#interface fortyGigE 0/52
Leaf011-s4810(conf-if-fo-0/52)#ip address 192.168.9.18/30
Leaf011-s4810(conf-if-fo-0/52)#no shutdown
Leaf011-s4810(conf-if-fo-0/52)#exit
Leaf011-s4810(conf)#router ospf 1
Leaf011-s4810(conf-router_ospf-1)#network 192.168.9.0/30 area 0
Leaf011-s4810(conf-router_ospf-1)#network 192.168.9.16/30 area 0
Leaf011-s4810(conf-router_ospf-1)#exit
```

Leaf012-S4810

```
Leaf012-s4810(conf)#interface fortyGigE 0/48
Leaf012-s4810(conf-if-fo-0/48)#ip address 192.168.9.6/30
Leaf012-s4810(conf-if-fo-0/48)#no shutdown
Leaf012-s4810(conf-if-fo-0/48)#exit
Leaf012-s4810(conf)#interface fortyGigE 0/52
Leaf012-s4810(conf-if-fo-0/52)#ip address 192.168.9.22/30
Leaf012-s4810(conf-if-fo-0/52)#no shutdown
Leaf012-s4810(conf-if-fo-0/52)#exit
Leaf012-s4810(conf)#router ospf 1
Leaf012-s4810(conf-router_ospf-1)#network 192.168.9.4/30 area 0
Leaf012-s4810(conf-router_ospf-1)#network 192.168.9.20/30 area 0
Leaf012-s4810(conf-router_ospf-1)#exit
```

Leaf02x-S60

After Leaf02 completes the port channel configuration, IP addresses are assigned and those networks placed in OSPF using the following commands.

```
Leaf02x-s60(conf)#interface TenGigabitEthernet 0/48
Leaf02x-s60(conf-if-te-0/48)#port-channel-protocol LACP
Leaf02x-s60(conf-if-te-0/48-lacp)#port-channel 1 mode active
Leaf02x-s60(conf-if-te-0/48-lacp)#no shutdown
Leaf02x-s60(conf-if-te-0/48)#exit
Leaf02x-s60(conf)#interface TenGigabitEthernet 0/49
Leaf02x-s60(conf-if-te-0/49)#port-channel-protocol LACP
Leaf02x-s60(conf-if-te-0/49-lacp)#port-channel 2 mode active
```

```
Leaf02x-s60(conf-if-te-0/49-lacp)#no shutdown
Leaf02x-s60(conf-if-te-0/49)#exit
Leaf02x-s60(conf)#interface TenGigabitEthernet 1/48
Leaf02x-s60(conf-if-te-1/48)#port-channel-protocol LACP
Leaf02x-s60(conf-if-te-1/48-lacp)#port-channel 1 mode active
Leaf02x-s60(conf-if-te-1/48-lacp)#no shutdown
Leaf02x-s60(conf-if-te-1/48)#exit
Leaf02x-s60(conf)#interface TenGigabitEthernet 1/49
Leaf02x-s60(conf-if-te-1/49)#port-channel-protocol LACP
Leaf02x-s60(conf-if-te-1/49-lacp)#port-channel 2 mode active
Leaf02x-s60(conf-if-te-1/49-lacp)#no shutdown
Leaf02x-s60(conf-if-te-1/49)#exit
Leaf02x-s60(conf)#interface Port-channel 1
Leaf02x-s60(conf-if-po-1)#ip address 192.168.9.10/30
Leaf02x-s60(conf-if-po-1)#no shutdown
Leaf02x-s60(conf-if-po-1)#exit
Leaf02x-s60(conf)#interface Port-channel 2
Leaf02x-s60(conf-if-po-2)#ip address 192.168.9.26/30
Leaf02x-s60(conf-if-po-2)#no shutdown
Leaf02x-s60(conf-if-po-2)#exit
Leaf02x-s60(conf)#router ospf 1
Leaf02x-s60(conf-router_ospf-1)#network 192.168.9.8/30 area 0
Leaf02x-s60(conf-router_ospf-1)#network 192.168.9.24/30 area 0
Leaf02x-s60(conf-router_ospf-1)#exit
```

3.4.2 Workload

Following the test configuration presented in this paper provides best practices that contain all iSCSI SAN traffic to the respective leaf nodes. The result is that there is not a load applied to the spine switches for each of the above scenarios. This is not to say IP connectivity has been limited in the configuration above, every subnet has the capability to route through the spine should the need arise as would be expected with other LAN traffic.

Since the previous SAN workloads do not generate a workload on the spine, the load is provided by replicating data from the remote iSCSI SAN in Leaf02 to the main iSCSI SAN in Leaf01.

Replication configuration

The replication configuration for this white paper consists of eight 50 GB volumes in Leaf02 replicated to the larger EqualLogic SAN in Leaf01.

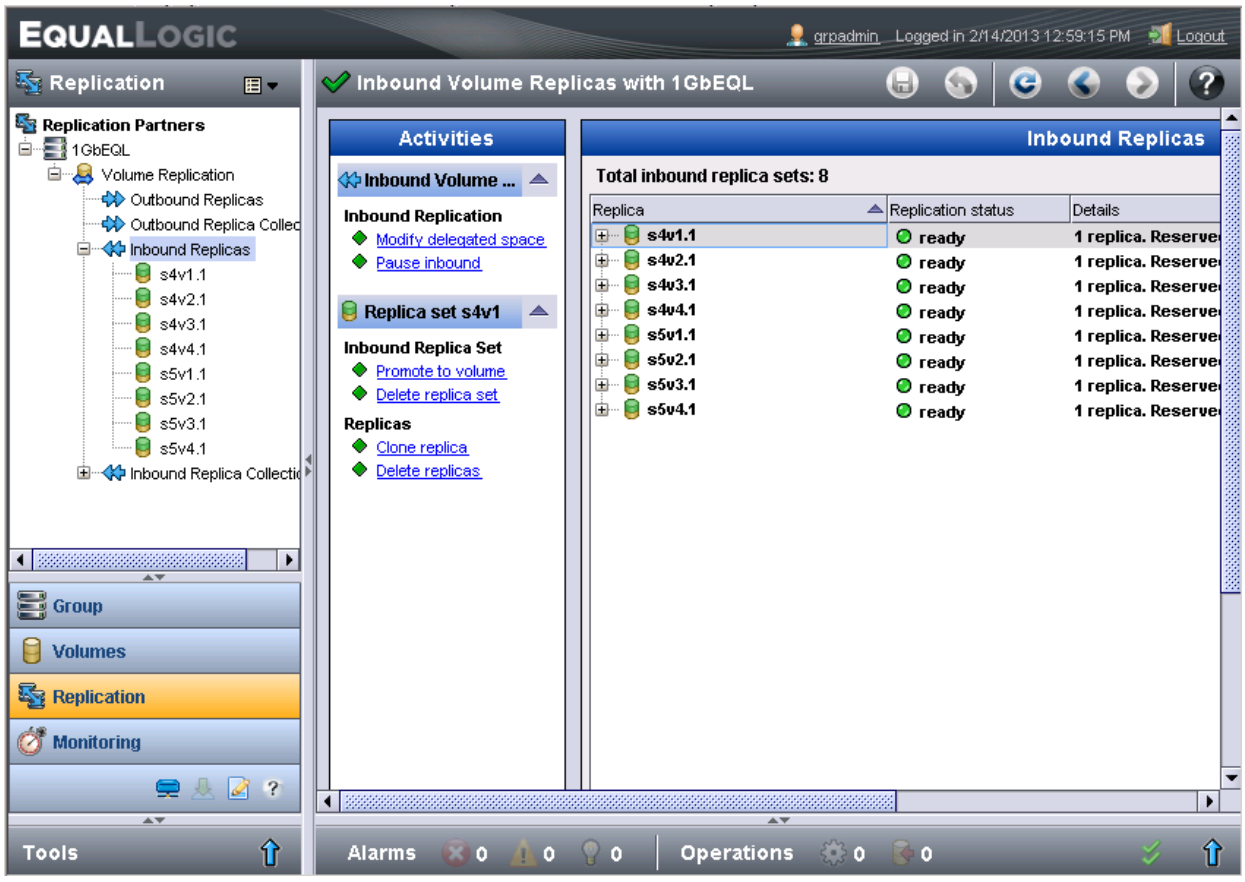

Figure 10 EqualLogic Replication setup

Note: This paper focuses purely on the network aspect of replicating from one leaf to another; for guidance on configuring and sizing replication appropriately please see *Dell EqualLogic Replication Best Practices and Sizing Guide* at http://en.community.dell.com/dell-groups/dtcmedia/m/mediagallery/19854181/download.aspx.

3.4.3 Results

Sending replication traffic through the high speed spine was not a challenge for the 40 Gb Dell Force10 Z9000 switches.

All network measurements were taken while a replica set was in process as shown in Figure 11. No other traffic was running during the test.

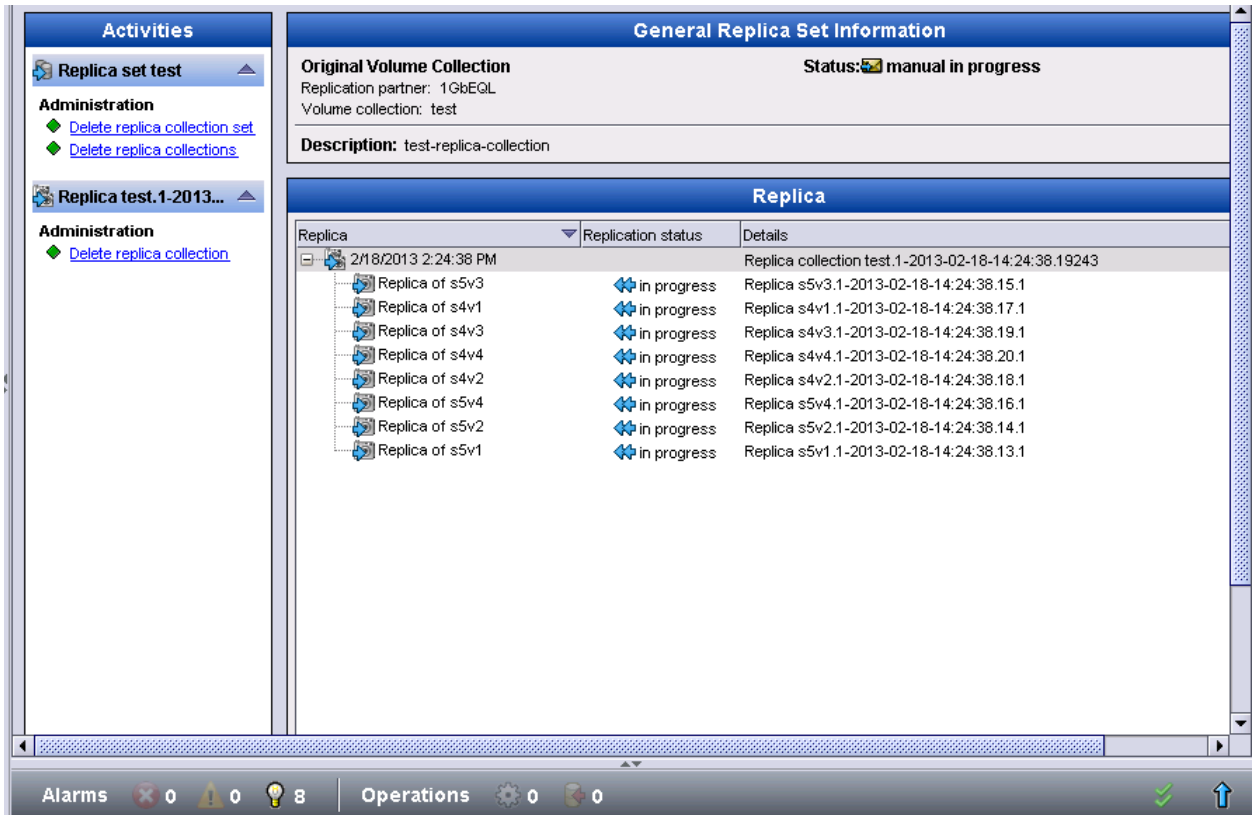

Figure 11 Replica in progress

As shown in the chart below, the TCP retransmits was 0% for both groups during this time.

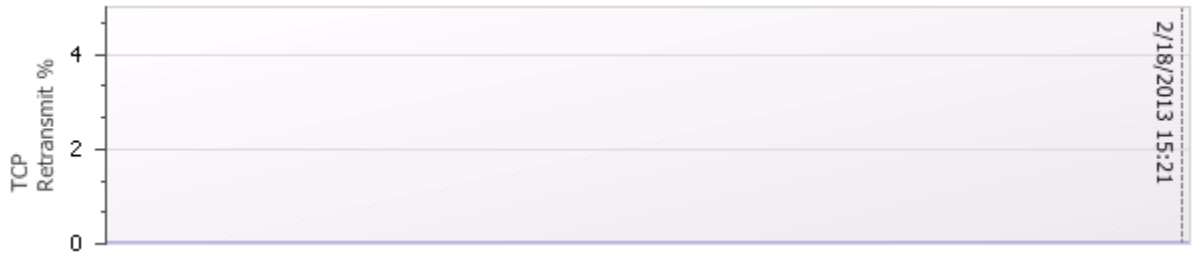

Figure 12 Replication TCP retransmit rate

The latency measurements show why collocating iSCSI Hosts and iSCSI storage to the same leaf node is important. Notice while the average latency is very low, there are occasional spikes in latency that could cause latency sensitive iSCSI traffic to seem unresponsive. This is why replication traffic is the only iSCSI traffic allowed to traverse the spine.

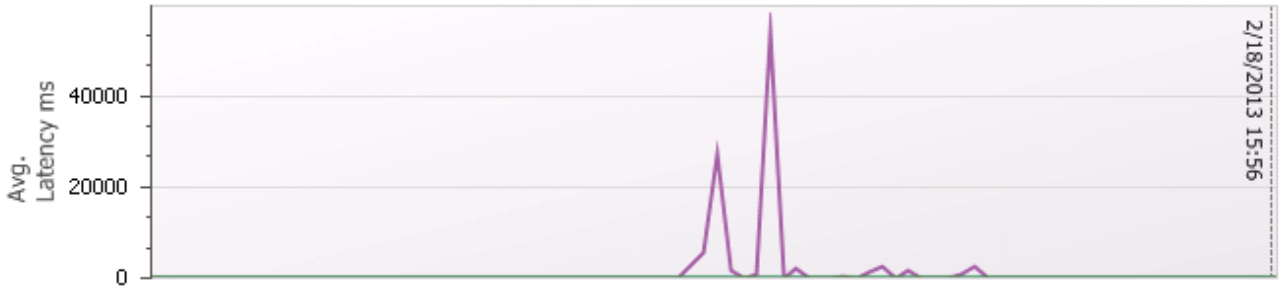

Figure 13 Leaf01 – 10 Gb iSCSI group average latency ms

4 Conclusion

The Force10 leaf-spine architecture is highly customizable depending on the challenges and requirements presented. The topology outlined in this paper represents a very basic configuration that highlights some key points when integrating EqualLogic storage into the distributed core architecture.

- iSCSI traffic separation – Following best practices for Data Center Bridging requires the creation of multiple VLANS when running multiple traffic types on the same physical line. This allows iSCSI to be prioritized appropriately with guaranteed bandwidth, thereby protecting the data traffic.
- iSCSI traffic collocation – Keeping iSCSI storage arrays and iSCSI hosts physically connected to the same leaf node allows for reliably lower latency, lower hop count, and increased simplicity.

This white paper focused on a simple topology consisting of the high speed spine and two leaf nodes. It worked through the configuration of each leaf node independently then showed how the complete topology comes together with the spine to form a very capable distributed core fabric.

A Hardware Configuration details

A.1 Server hardware

Table 11 Test Server (Leaf01)

Dell PowerEdge R815 Configuration (x3)	
Function	Physical load generation servers
Processors	4 x AMD Opteron™ Processor 6174 2.20GHz, L2/L3 Cache: 6MB/10MB
Memory	128 GB
BIOS	1.5.2
Onboard NIC	Broadcom 5709 Quad Port 1GbE NIC TCP Offload Enabled iSCSI Offload Enabled Family Firmware version 7.4.8
Network interface	Broadcom 57810 10Gb Dual Port CNA TCP Offload Enabled iSCSI Offload Enabled Family Firmware version 7.4.8
Embedded Management	iDRAC6 Enterprise
Operating System	Windows Server 2008 R2 with SP1

Table 12 Test Server (Leaf02)

Dell PowerEdge R815 Configuration (x2)	
Function	Physical load generation servers
Processors	4 x AMD Opteron™ Processor 6174 2.20GHz, L2/L3 Cache: 6MB/10MB
Memory	128 GB
BIOS	1.5.2
Onboard NIC	Broadcom 5709 Quad Port 1GbE NIC TCP Offload Enabled iSCSI Offload Enabled Family Firmware version 7.4.8
Embedded Management	iDRAC6 Enterprise
Operating System	Windows Server 2008 R2 with SP1

A.2 Network hardware

Table 13 Spine hardware

Dell Force10 Z9000 (x2)	
Function	Spine interconnect
Expansion Card	N/A
FTOS version	8.3.11.0

Table 14 Leaf01 network hardware

Dell Force10 S4810 (x2)	
Function	10Gb Leaf (Leaf01)
Expansion Card	N/A
FTOS version	8.3.12.0

Table 15 Leaf02 network hardware

Dell Force10 S60 (x2)	
Function	1Gb Leaf (Leaf02)
Expansion Card	Dual port 10Gb SFP+ (front), stacking port (back)
FTOS version	8.3.3.7

A.3 Storage hardware

Table 16 Leaf01 storage hardware

Dell EqualLogic PS6110 (x3)	
Function	10Gb iSCSI SAN
Disks (x24)	SAS HDD Model:ST9146853SS
Firmware	V6.0.2

Table 17 Leaf02 storage hardware

Dell EqualLogic PS6100 (x1)	
Function	1Gb iSCSI SAN
Disks (x24)	SAS HDD Model:ST9146853SS
Firmware	V6.0.2

B Running configurations

The following is the results of running the "`show running-config`" command on the following switches, for brevity interfaces not used in the white paper have been omitted.

B.1 Leaf011-S4810

```
Current Configuration ...
! Version 8.3.12.0
! Last configuration change at Mon Feb 18 13:26:46 2013 by admin
!
boot system stack-unit 0 primary system: B:
boot system stack-unit 0 secondary system: A:
boot system stack-unit 0 default system: A:
boot system gateway 0.0.0.0
!
redundancy auto-synchronize full
!
hardware watchdog
!
hostname Leaf011-s4810
!
username admin password 7 f207d5c095fafa41 privilege 15
!
protocol spanning-tree rstp
 no disable
!
dcb enable
!
stack-unit 0 provision S4810
!
interface TenGigabitEthernet 0/0
 no ip address
 mtu 12000
 switchport
 spanning-tree rstp edge-port
 dcb-policy input pfc
 dcb-policy output ets
!
 protocol lldp
 no shutdown
!
interface TenGigabitEthernet 0/1
 no ip address
 mtu 12000
```

```
 switchport
 spanning-tree rstp edge-port
 dcb-policy input pfc
 dcb-policy output ets
!
 protocol lldp
 no shutdown
!
interface TenGigabitEthernet 0/2
 no ip address
 mtu 12000
 switchport
 spanning-tree rstp edge-port
 dcb-policy input pfc
 dcb-policy output ets
!
 protocol lldp
 no shutdown
!
interface TenGigabitEthernet 0/3
 no ip address
 mtu 12000
 switchport
 spanning-tree mstp edge-port
 spanning-tree rstp edge-port
 spanning-tree 0 portfast
 spanning-tree pvst edge-port
 dcb-policy input pfc
 dcb-policy output ets
!
 protocol lldp
 no shutdown
!
interface TenGigabitEthernet 0/4
 no ip address
 mtu 12000
 switchport
 spanning-tree mstp edge-port
 spanning-tree rstp edge-port
 spanning-tree 0 portfast
 spanning-tree pvst edge-port
 dcb-policy input pfc
 dcb-policy output ets
!
 protocol lldp
 no shutdown
!
```

```
interface TenGigabitEthernet 0/5
 no ip address
 mtu 12000
 switchport
 spanning-tree mstp edge-port
 spanning-tree rstp edge-port
 spanning-tree 0 portfast
 spanning-tree pvst edge-port
 dcb-policy input pfc
 dcb-policy output ets
!
 protocol lldp
 no shutdown
!

… [un-configured interfaces omitted] …

interface fortyGigE 0/48
 ip address 192.168.9.2/30
 mtu 12000
 dcb-policy input pfc
 dcb-policy output ets
 no shutdown
!
interface fortyGigE 0/52
 ip address 192.168.9.18/30
 mtu 12000
 dcb-policy input pfc
 dcb-policy output ets
 no shutdown
!
interface fortyGigE 0/56
 no ip address
 mtu 12000
 dcb-policy input pfc
 dcb-policy output ets
!
 port-channel-protocol LACP
  port-channel 3 mode active
 no shutdown
!
interface fortyGigE 0/60
 no ip address
 mtu 12000
 dcb-policy input pfc
 dcb-policy output ets
!
```

```
 port-channel-protocol LACP
  port-channel 3 mode active
 no shutdown
!
interface ManagementEthernet 0/0
 ip address 192.168.2.143/24
 no shutdown
!
interface ManagementEthernet 1/0
 no shutdown
!
interface ManagementEthernet 2/0
 no shutdown
!
interface ManagementEthernet 3/0
 no shutdown
!
interface ManagementEthernet 4/0
 no shutdown
!
interface ManagementEthernet 5/0
 no shutdown
!
interface ManagementEthernet 6/0
 no shutdown
!
interface ManagementEthernet 7/0
 no shutdown
!
interface ManagementEthernet 8/0
 no shutdown
!
interface ManagementEthernet 9/0
 no shutdown
!
interface ManagementEthernet 10/0
 no shutdown
!
interface ManagementEthernet 11/0
 no shutdown
!
interface Port-channel 3
 no ip address
 mtu 12000
 switchport
 no shutdown
!
```

```
interface Vlan 1
!
interface Vlan 10
 description iSCSI_VLAN
 ip address 192.168.10.2/24
 mtu 12000
 tagged TenGigabitEthernet 0/0-5
 tagged Port-channel 3
!
 vrrp-group 10
  virtual-address 192.168.10.1
 no shutdown
!
interface Vlan 11
 description LAN_VLAN
 ip address 192.168.11.2/24
 mtu 12000
 tagged TenGigabitEthernet 0/0-2
 tagged Port-channel 3
!
 vrrp-group 11
  virtual-address 192.168.11.1
 no shutdown
!
router ospf 1
 network 192.168.9.0/30 area 0
 network 192.168.10.0/24 area 0
 network 192.168.9.16/30 area 0
 network 192.168.11.0/24 area 0
!
service-class dynamic dot1p
!
qos-policy-output OTHER ets
 bandwidth-percentage 70
!
qos-policy-output iSCSI ets
 bandwidth-percentage 30
!
dcb stack-unit all pfc-buffering pfc-ports 56 pfc-queues 2
!
dcb-input pfc
 pfc priority 4
!
priority-group OTHER
 priority-list 0-3,5-7
 set-pgid 2
!
```

```
priority-group iSCSI
 priority-list 4
 set-pgid 1
!
dcb-output ets
 priority-group OTHER qos-policy OTHER
 priority-group iSCSI qos-policy iSCSI
!
protocol lldp
 advertise management-tlv system-capabilities system-description system-name
 advertise interface-port-desc
!
line console 0
line vty 0
 exec-timeout 0 0
line vty 1
line vty 2
line vty 3
line vty 4
line vty 5
line vty 6
line vty 7
line vty 8
line vty 9
!
end
```

B.2 Leaf012-S4810

```
Current Configuration ...
! Version 8.3.12.0
! Last configuration change at Mon Feb 18 13:27:49 2013 by admin
!
boot system stack-unit 0 primary system: B:
boot system stack-unit 0 secondary system: B:
boot system stack-unit 0 default system: A:
boot system gateway 0.0.0.0
!
redundancy auto-synchronize full
!
hardware watchdog
!
hostname Leaf012-s4810
!
username admin password 7 f207d5c095fafa41 privilege 15
!
```

```
protocol spanning-tree rstp
 no disable
!
dcb enable
!
stack-unit 0 provision S4810
!
interface TenGigabitEthernet 0/0
 no ip address
 mtu 12000
 switchport
 spanning-tree rstp edge-port
 dcb-policy input pfc
 dcb-policy output ets
!
 protocol lldp
 no shutdown
!
interface TenGigabitEthernet 0/1
 no ip address
 mtu 12000
 switchport
 spanning-tree rstp edge-port
 dcb-policy input pfc
 dcb-policy output ets
!
 protocol lldp
 no shutdown
!
interface TenGigabitEthernet 0/2
 no ip address
 mtu 12000
 switchport
 spanning-tree rstp edge-port
 dcb-policy input pfc
 dcb-policy output ets
!
 protocol lldp
 no shutdown
!
interface TenGigabitEthernet 0/3
 no ip address
 mtu 12000
 switchport
 spanning-tree mstp edge-port
 spanning-tree rstp edge-port
 spanning-tree 0 portfast
```

```
 spanning-tree pvst edge-port
 dcb-policy input pfc
 dcb-policy output ets
!
 protocol lldp
 no shutdown
!
interface TenGigabitEthernet 0/4
 no ip address
 mtu 12000
 switchport
 spanning-tree mstp edge-port
 spanning-tree rstp edge-port
 spanning-tree 0 portfast
 spanning-tree pvst edge-port
 dcb-policy input pfc
 dcb-policy output ets
!
 protocol lldp
 no shutdown
!
interface TenGigabitEthernet 0/5
 no ip address
 mtu 12000
 switchport
 spanning-tree mstp edge-port
 spanning-tree rstp edge-port
 spanning-tree 0 portfast
 spanning-tree pvst edge-port
 dcb-policy input pfc
 dcb-policy output ets
!
 protocol lldp
 no shutdown
!

… [un-configured interfaces omitted] …


interface fortyGigE 0/48
 ip address 192.168.9.6/30
 mtu 12000
 dcb-policy input pfc
 dcb-policy output ets
 no shutdown
!
interface fortyGigE 0/52
```

```
 ip address 192.168.9.22/30
 mtu 12000
 dcb-policy input pfc
 dcb-policy output ets
 no shutdown
!
interface fortyGigE 0/56
 no ip address
 mtu 12000
 dcb-policy input pfc
 dcb-policy output ets
!
 port-channel-protocol LACP
  port-channel 3 mode active
 no shutdown
!
interface fortyGigE 0/60
 no ip address
 mtu 12000
 dcb-policy input pfc
 dcb-policy output ets
!
 port-channel-protocol LACP
  port-channel 3 mode active
 no shutdown
!
interface ManagementEthernet 0/0
 ip address 192.168.2.144/24
 no shutdown
!
interface ManagementEthernet 1/0
 no shutdown
!
interface ManagementEthernet 2/0
 no shutdown
!
interface ManagementEthernet 3/0
 no shutdown
!
interface ManagementEthernet 4/0
 no shutdown
!
interface ManagementEthernet 5/0
 no shutdown
!
interface ManagementEthernet 6/0
 no shutdown
```

```
!
interface ManagementEthernet 7/0
 no shutdown
!
interface ManagementEthernet 8/0
 no shutdown
!
interface ManagementEthernet 9/0
 no shutdown
!
interface ManagementEthernet 10/0
 no shutdown
!
interface ManagementEthernet 11/0
 no shutdown
!
interface Port-channel 3
 no ip address
 mtu 12000
 switchport
 no shutdown
!
interface Vlan 1
!
interface Vlan 10
 ip address 192.168.10.3/24
 mtu 12000
 tagged TenGigabitEthernet 0/0-5
 tagged Port-channel 3
!
 vrrp-group 10
  virtual-address 192.168.10.1
 no shutdown
!
interface Vlan 11
 ip address 192.168.11.3/24
 tagged TenGigabitEthernet 0/0-2
 tagged Port-channel 3
!
 vrrp-group 11
  virtual-address 192.168.11.1
 no shutdown
!
router ospf 1
 network 192.168.10.0/24 area 0
 network 192.168.9.4/30 area 0
 network 192.168.9.20/30 area 0
```

```
 network 192.168.11.0/24 area 0
!
service-class dynamic dot1p
!
qos-policy-output OTHER ets
 bandwidth-percentage 70
!
qos-policy-output iSCSI ets
 bandwidth-percentage 30
!
dcb-input pfc
 pfc priority 4
!
priority-group OTHER
 priority-list 0-3,5-7
 set-pgid 2
!
priority-group iSCSI
 priority-list 4
 set-pgid 1
!
dcb-output ets
 priority-group OTHER qos-policy OTHER
 priority-group iSCSI qos-policy iSCSI
!
protocol lldp
 advertise management-tlv system-capabilities system-description system-name
 advertise interface-port-desc
!
line console 0
line vty 0
 exec-timeout 0 0
line vty 1
line vty 2
line vty 3
line vty 4
line vty 5
line vty 6
line vty 7
line vty 8
line vty 9
!
end
```

B.3 Leaf02x-S60

```
Current Configuration ...
! Version 8.3.3.7
!
boot system stack-unit 0 primary system: A:
boot system stack-unit 0 secondary system: A:
boot system stack-unit 0 default system: A:
boot system stack-unit 1 primary system: A:
boot system stack-unit 1 secondary system: A:
boot system stack-unit 1 default system: A:
boot system gateway 192.168.130.1
!
redundancy auto-synchronize full
!
hardware watchdog
!
hostname Leaf02x-s60
!
username admin password 7 f207d5c095fafa41 privilege 15
!
protocol spanning-tree rstp
!
stack-unit 0 provision S60
!
interface GigabitEthernet 0/0
 no ip address
 mtu 9216
 switchport
 flowcontrol rx on
 spanning-tree rstp edge-port
 no shutdown
!
interface GigabitEthernet 0/1
 no ip address
 mtu 9216
 switchport
 flowcontrol rx on
 spanning-tree rstp edge-port
 no shutdown
!
interface GigabitEthernet 0/2
 no ip address
 mtu 9216
 switchport
 flowcontrol rx on
 spanning-tree rstp edge-port
 no shutdown
!
```

```
interface GigabitEthernet 0/3
 no ip address
 mtu 9216
 switchport
 flowcontrol rx on
 spanning-tree rstp edge-port
 no shutdown
!
interface GigabitEthernet 0/4
 no ip address
 mtu 9216
 switchport
 flowcontrol rx on
 spanning-tree rstp edge-port
 no shutdown
!
interface GigabitEthernet 0/5
 no ip address
 mtu 9216
 switchport
 flowcontrol rx on
 spanning-tree rstp edge-port
 no shutdown
!
interface GigabitEthernet 0/6
 no ip address
 mtu 9216
 switchport
 flowcontrol rx on
 spanning-tree rstp edge-port
 no shutdown
!
interface GigabitEthernet 0/7
 no ip address
 mtu 9216
 switchport
 flowcontrol rx on
 spanning-tree rstp edge-port
 no shutdown
!
interface GigabitEthernet 0/8
 no ip address
 mtu 9216
 switchport
 flowcontrol rx on
 spanning-tree rstp edge-port
 no shutdown
```

```
!
… [un-configured interfaces omitted] …
interface TenGigabitEthernet 0/48
 no ip address
!
 port-channel-protocol LACP
  port-channel 1 mode active
 no shutdown
!
interface TenGigabitEthernet 0/49
 no ip address
!
 port-channel-protocol LACP
  port-channel 2 mode active
 no shutdown
!
stack-unit 1 provision S60
!
interface GigabitEthernet 1/0
 no ip address
 mtu 9216
 switchport
 flowcontrol rx on
 spanning-tree rstp edge-port
 no shutdown
!
interface GigabitEthernet 1/1
 no ip address
 mtu 9216
 switchport
 flowcontrol rx on
 spanning-tree rstp edge-port
 no shutdown
!
interface GigabitEthernet 1/2
 no ip address
 mtu 9216
 switchport
 flowcontrol rx on
 spanning-tree rstp edge-port
 no shutdown
!
interface GigabitEthernet 1/3
 no ip address
 mtu 9216
 switchport
 flowcontrol rx on
```

```
 spanning-tree rstp edge-port
 no shutdown
!
interface GigabitEthernet 1/4
 no ip address
 mtu 9216
 switchport
 flowcontrol rx on
 spanning-tree rstp edge-port
 no shutdown
!
interface GigabitEthernet 1/5
 no ip address
 mtu 9216
 switchport
 flowcontrol rx on
 spanning-tree rstp edge-port
 no shutdown
!
interface GigabitEthernet 1/6
 no ip address
 mtu 9216
 switchport
 flowcontrol rx on
 spanning-tree rstp edge-port
 no shutdown
!
interface GigabitEthernet 1/7
 no ip address
 mtu 9216
 switchport
 flowcontrol rx on
 spanning-tree rstp edge-port
 no shutdown
!
interface GigabitEthernet 1/8
 no ip address
 mtu 9216
 switchport
 flowcontrol rx on
 spanning-tree rstp edge-port
 no shutdown
!
… [un-configured interfaces omitted] …
interface TenGigabitEthernet 1/48
 no ip address
!
```

```
 port-channel-protocol LACP
  port-channel 1 mode active
 no shutdown
!
interface TenGigabitEthernet 1/49
 no ip address
!
 port-channel-protocol LACP
  port-channel 2 mode active
 no shutdown
!
interface ManagementEthernet 0/0
 ip address 192.168.2.145/24
 no shutdown
!
interface ManagementEthernet 1/0
 ip address 192.168.2.146/24
 no shutdown
!
interface ManagementEthernet 2/0
 no shutdown
!
interface ManagementEthernet 3/0
 no shutdown
!
interface ManagementEthernet 4/0
 no shutdown
!
interface ManagementEthernet 5/0
 no shutdown
!
interface ManagementEthernet 6/0
 no shutdown
!
interface ManagementEthernet 7/0
 no shutdown
!
interface ManagementEthernet 8/0
 no shutdown
!
interface ManagementEthernet 9/0
 no shutdown
!
interface ManagementEthernet 10/0
 no shutdown
!
interface ManagementEthernet 11/0
```

```
 no shutdown
!
interface Port-channel 1
 ip address 192.168.9.10/30
 no shutdown
!
interface Port-channel 2
 ip address 192.168.9.26/30
 no shutdown
!
interface Vlan 1
!untagged GigabitEthernet 0/8-47
!untagged GigabitEthernet 1/8-47
!
interface Vlan 12
 description iSCSI_VLAN
 ip address 192.168.12.1/24
 untagged GigabitEthernet 0/1,3-7
 untagged GigabitEthernet 1/1,3-7
 no shutdown
!
interface Vlan 13
 description LAN_VLAN
 ip address 192.168.13.1/24
 untagged GigabitEthernet 0/0,2
 untagged GigabitEthernet 1/0,2
 no shutdown
!
router ospf 1
 network 192.168.9.8/30 area 0
 network 192.168.9.24/30 area 0
 network 192.168.12.0/24 area 0
 network 192.168.13.0/24 area 0
!
ntp server 192.168.2.1
!
clock timezone 14:10:00 -6
!
line console 0
line vty 0
 exec-timeout 0 0
line vty 1
line vty 2
line vty 3
line vty 4
line vty 5
line vty 6
```

```
line vty 7
line vty 8
line vty 9
!
end
```

B.4 Spine01-Z9000

```
Current Configuration ...
! Version 8.3.11.0
!
boot system stack-unit 0 primary system: A:
boot system stack-unit 0 secondary system: B:
boot system stack-unit 0 default system: A:
boot system gateway 172.27.1.254
!
redundancy auto-synchronize full
!
hardware watchdog
!
hostname Spine01-z9000
!
enable password 7 b125455cf679b208e79b910e85789edf
!
username admin password 7 f207d5c095fafa41 privilege 15
!
protocol spanning-tree pvst
 no disable
!
stack-unit 0 port 8 portmode quad
!
stack-unit 0 provision Z9000
!
interface fortyGigE 0/0
 ip address 192.168.9.1/30
 mtu 12000
 no shutdown
!
interface fortyGigE 0/4
 ip address 192.168.9.5/30
 mtu 12000
 no shutdown
!
interface TenGigabitEthernet 0/8
 no ip address
!
```

```
 port-channel-protocol LACP
  port-channel 1 mode active
 no shutdown
!
interface TenGigabitEthernet 0/9
 no ip address
!
 port-channel-protocol LACP
  port-channel 1 mode active
 no shutdown
!
… [un-configured interfaces omitted] …
interface ManagementEthernet 0/0
 ip address 192.168.2.141/24
 no shutdown
!
interface ManagementEthernet 1/0
 no shutdown
!
interface Port-channel 1
 ip address 192.168.9.9/30
 no shutdown
!
interface Vlan 1
!
router ospf 1
 network 192.168.9.0/30 area 0
 network 192.168.9.4/30 area 0
 network 192.168.9.8/30 area 0
!
protocol gvrp
!
line console 0
line vty 0
 exec-timeout 0 0
line vty 1
line vty 2
line vty 3
line vty 4
line vty 5
line vty 6
line vty 7
line vty 8
line vty 9
!
end
```

B.5 Spine02-Z9000

```
Current Configuration ...
! Version 8.3.11.0
!
boot system stack-unit 0 primary system: A:
boot system stack-unit 0 secondary system: B:
boot system stack-unit 0 default system: A:
boot system gateway 172.27.1.254
!
redundancy auto-synchronize full
!
hardware watchdog
!
hostname Spine02-z9000
!
username admin password 7 f207d5c095fafa41 privilege 15
!
protocol spanning-tree rstp
!
stack-unit 0 port 8 portmode quad
!
stack-unit 0 provision Z9000
!
interface fortyGigE 0/0
 ip address 192.168.9.17/30
 mtu 12000
 no shutdown
!
interface fortyGigE 0/4
 ip address 192.168.9.21/30
 mtu 12000
 no shutdown
!
interface TenGigabitEthernet 0/8
 no ip address
!
 port-channel-protocol LACP
  port-channel 2 mode active
 no shutdown
!
interface TenGigabitEthernet 0/9
 no ip address
!
 port-channel-protocol LACP
  port-channel 2 mode active
```

```
 no shutdown
!
… [un-configured interfaces omitted] …
interface ManagementEthernet 0/0
 ip address 192.168.2.142/24
 no shutdown
!
interface ManagementEthernet 1/0
 no shutdown
!
interface Port-channel 2
 ip address 192.168.9.25/30
 no shutdown
!
interface Vlan 1
!
router ospf 1
 network 192.168.9.16/30 area 0
 network 192.168.9.20/30 area 0
 network 192.168.9.24/30 area 0
!
line console 0
line vty 0
 exec-timeout 0 0
line vty 1
line vty 2
line vty 3
line vty 4
line vty 5
line vty 6
line vty 7
line vty 8
line vty 9
!
End
```

C Configuring Broadcom 57810 for DCB on Windows 2008 R2 with SP1

This section will run through the steps used to configure the Broadcom 57810 for DCB enabled Ethernet for iSCSI storage.

This section assumes the following prerequisite steps have been completed

- Installation and configuration of Dell Host Integration Toolkit
  - Default installation
- Configuration of Microsoft iSCSI Initiator
  - Addition of Group IP in Discovery Portal

Configure Broadcom 57810

1. Enable iSCSI Offload
   a. Open Broadcom Advanced Control Suite
   b. Navigate to Hosts > Hostname > Adapter # > Port #
   c. Select Configuration on the right hand tab
   d. Expand Resource Reservations
   e. Select Configure
   f. Expand Protocols in the Hardware and Resource Configuration Wizard popup window
   g. Check the box next to iSCSI
   h. Select Next
   i. Select Apply

Notice that there are now two adapters listed under the Port #

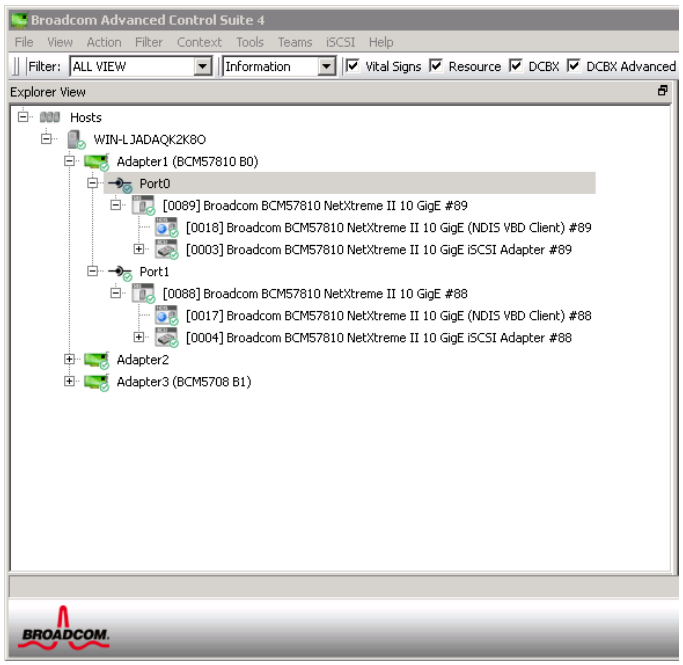

Figure 14 iSCSI offload enabled

NDIS VBD client – This is the adapter as listed under windows networking, utilized for all non-storage traffic that utilizes this adapter

iSCSI adapter – This is the representation of the storage adapter, this is not visible under the Network and sharing center in windows. It is used for all iSCSI traffic traversing this adapter when DCB is enabled.

2. Configure iSCSI Adapter
    a. Select the iSCSI Adapter listed
    b. Select Configuration Tab on right hand pane
    c. Set VLAN ID
        i. Match iSCSI VLAN ID set on Switch
    d. Set MTU to desired setting
        i. Set to 9000 in this scenario
    e. Select Edit next too IPv4 Configuration
        i. Select Enable DHCP if DHCP is desired
        ii. Select Disable DHCP if Static IP information is desired
3. Connect to iSCSI volumes in Microsoft iSCSI initiator

Once configured correctly the iSCSI volumes should appear under the iSCSI Portal list in the BACS4 tool as shown below.

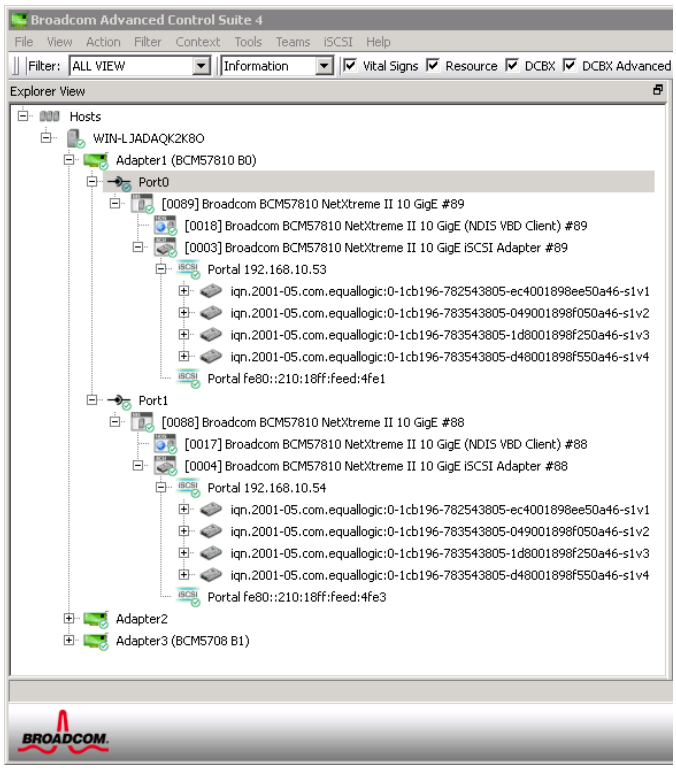

Figure 15 Completed 57810 configuration

D Load generation configuration file

Multiple tools were utilized for load generation. To provide storage traffic to the disks, VDBench was used for its simplicity and script ability.

VDBench Config Server01 (Physical Windows Host)

```
sd=A-a,lun=\\.\PhysicalDrive1
sd=A-b,lun=\\.\PhysicalDrive2
sd=A-c,lun=\\.\PhysicalDrive3
sd=A-d,lun=\\.\PhysicalDrive4
sd=B-a,lun=\\.\PhysicalDrive1,range=(30m,60m)
sd=B-b,lun=\\.\PhysicalDrive2,range=(30m,60m)
sd=B-c,lun=\\.\PhysicalDrive3,range=(30m,60m)
sd=B-d,lun=\\.\PhysicalDrive4,range=(30m,60m)

wd=wd1,sd=A-*,seekpct=100,rdpct=67,xfersize=8k,iorate=9999999,priority=1
wd=wd2,sd=B-*,seekpct=0,rdpct=100,xfersize=256k,iorate=9999999,priority=1
wd=wd3,sd=B-*,seekpct=0,rdpct=0,xfersize=64k,iorate=9999999,priority=1

rd=rd1,wd=wd1,elapsed=10800,interval=30,forthreads=20
rd=rd2,wd=wd2,elapsed=10800,interval=30,forthreads=5
rd=rd3,wd=wd3,elapsed=10800,interval=30,forthreads=5
```

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:

- 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
- Dell Force10 s4810 Switch Configuration Guide (SCG) at http://en.community.dell.com/techcenter/storage/w/wiki/4250.switch-configuration-guides-by-sis.aspx
- Dell EqualLogic Replication Best Practices and Sizing Guide at http://en.community.dell.com/techcenter/storage/w/wiki/2641.aspx
- Dell Fabric Manager video overview at http://en.community.dell.com/techcenter/networking/w/wiki/3880.dell-fabric-manager.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