EqualLogic™ PS Series
Reference Architecture for
Arista™ 7148sx
Four-Switch SAN Reference

Validation Engineer: Gregory (Brent) Seaton, Tony Ansley

Storage Infrastructure and Solutions Engineering
Dell™ Product Group
December 2010

Table of Contents

Table of Tables

Table of Figures

1 Executive Summary

1.1 Product Description

The Arista™ 7148sx Datacenter Ethernet switch features a high density 10 Gigabit Ethernet switching solution and an extensible, modular network operating system. The Arista 7148sx is the top end of the Arista 7100s Series switch family and is optimized for non-blocking applications with up to 960Gbps throughput while providing 48x 1Gbs/10Gbs SFP+ ports.

Testing has been performed that proves the value of the Arista 7148sx when used to build a storage area network (SAN) using the Dell™ EqualLogic™ series of virtual storage arrays. Based on this testing, the Arista 7148sx is an excellent switch for use when building SANs that consist of one or more 10Gb/s PS 6x10 arrays.

This reference design paper illustrates how to build a medium to large scale SAN consisting of Four (4) Arista 7148sx switches. It also provides data to support its belief that the Arista 7148sx can support a medium sized SAN using a four-switch solution, and can support twelve (12) arrays in a fully redundant configuration while offering excellent overall SAN performance for a variety of workloads. The actual ratio of arrays to hosts is based solely on the available number of ports when using four switches.

2 Objectives

This document provides a reference design for a SAN solution using the Arista 7148sx switch validation with EqualLogic Storage.

The goal of this exercise is not to provide a comprehensive set of possible configurations, but is an illustration of one possible solution using all of the accepted common practices recommendations for EqualLogic SANs. This document is designed to focus solely on the switch settings and interconnect methodology and not on the configuration of the host or arrays. For this reason, a standardized host system and EqualLogic Group configuration was used for this reference architecture. The overall solution is designed such that it is easy to deploy – many settings are based on defaults values – and allow the customer to quickly deploy a SAN solution based on these switches.

The following overall guidelines were used for developing this reference solution:

- All hosts have two (2) Ethernet ports attached to the SAN per EqualLogic documented best practices for a redundant storage network.
- All SAN attached ports will be configured based on default, "out of the box" settings with the exception of the use of Jumbo Frames, and Flow Control, which was enabled for the solution.
- Microsoft Windows Server was the operating system used on all hosts
- EqualLogic's Host Integration Toolkit was used for all hosts. In particular, Dell's EqualLogic MPIO Device Specific Module was used to provide best practices, EqualLogic aware multi-path management.
- Host connections to the SAN equaled the number of active array ports connected to the SAN.
    - Since each PS6x00 series array has four (4) active array ports, for each array in the test configuration, there will be two (2) hosts connected to the SAN

  - Since each PS6x10 series array has two(2) active array ports, for each array in the test configuration, there will be one (1) host connected to the SAN
  - While this may not meet each customer's requirements, this decision was used to simplify the testing by allowing us to insure that there were an adequate number of hosts to run the various workloads used during testing and to generate enough load to ensure that the PowerConnect 8024F could manage the traffic for a maximum number of arrays. It is expected that customers who deploy fewer arrays, has the flexibility to deploy more hosts as the available ports allow.
- The SAN was a single subnet and non-routed.
- The SAN was based on standard IPv4 addressing
- All testing used three pre-defined standardized workloads that reflect various types of real-world SAN utilization.

Note: For more information on EqualLogic SAN design common practices recommendations, consult the *EqualLogic Configuration Guide* that can be found at www.delltechcenter.com/equallogic/pages/equallogic

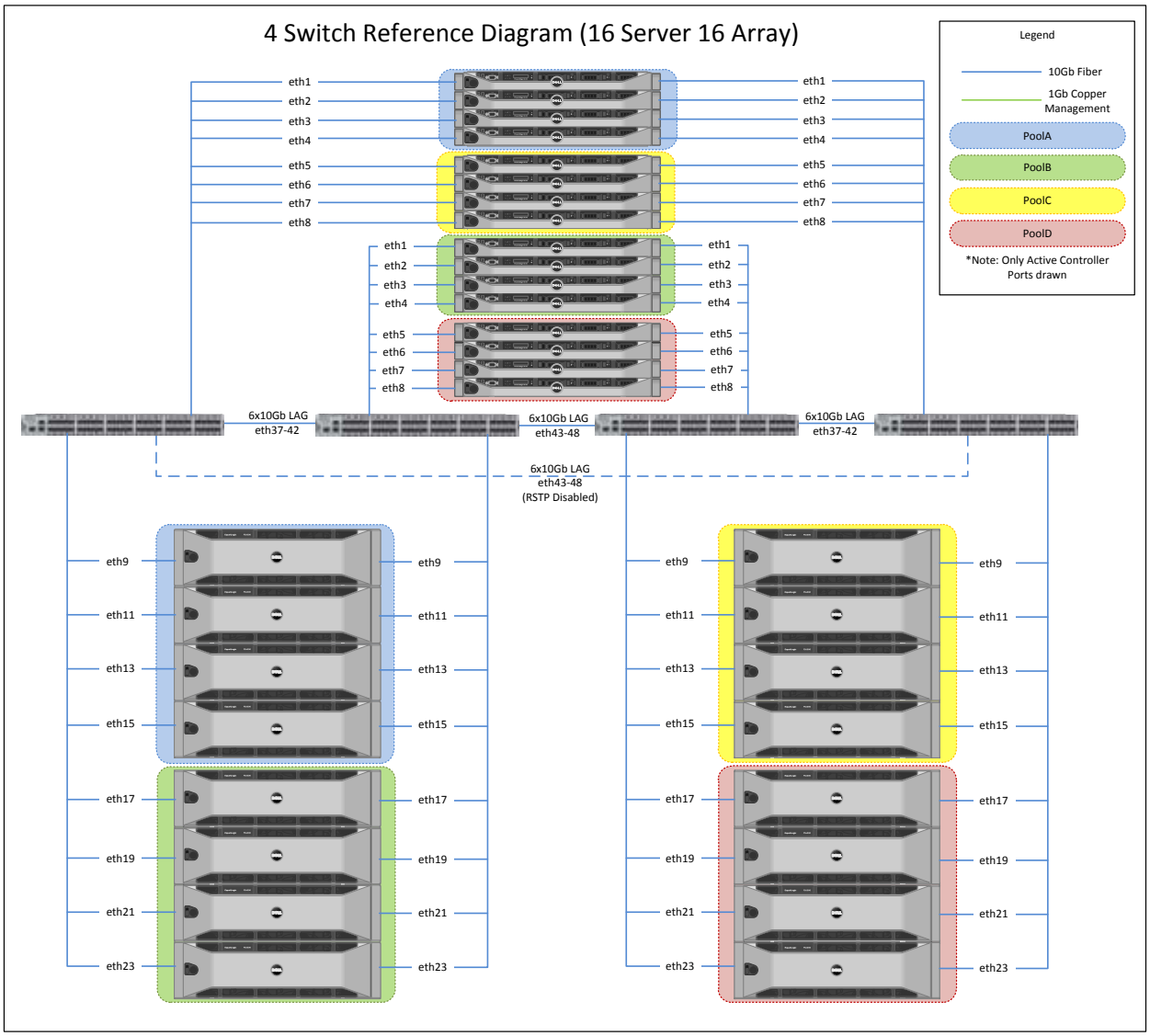
3 Reference SAN Design Configuration Details

The resulting reference design documented in this paper consists of four (4) Arista 7148SX switches. The switches are inter-connected using a multi-link aggregation group based on industry standard protocols to provide adequate bandwidth between switches to support the resulting configuration.

Connected to these switches are a total of sixteen (16) PS6010 arrays and sixteen (16) PowerEdge servers. Based on lab tests and the number of ports available for use to connect hosts, arrays, and to inter-connect the four switches, it was determined that this was the maximum recommended SAN when using this switch.

These sixteen arrays were configured into four (4) storage pools. Each pool was dedicated to a different workload type to ensure maximum performance of volumes for each workload profile. The resulting SAN reference configuration is illustrated in Figure 1 below.

Figure 1- Four Switch / Sixteen Array Reference Configuration



3.1 Server Configuration

Nine PowerEdge R610 rack servers were used in this test. Each server is described in Table 1 below.

Table 1 - Server Configuration Parameters

Server Model	PowerEdge™ R610
BIOS	2.1.9
Intel™ 5500-5520 Chipset	A02
OS	Windows Server® 2008 R2
Host Network Interface Configuration (x2 per server)	
Model	Dell Broadcom™ BCM57711 – Dual Port

iSCSI Offload Engine (iSOE)	Enabled
TCP Offload Engine (ToE)	Disabled
NDIS Mode	Disabled
iSCSI Initiator	Microsoft Windows Server 2008R2
Broadcom™BCM57711 Software	• Firmware: 5.2.7 • iSOE driver: 5.2.11 • NDIS driver: 5.2.14 • VBD driver: 5.2.22
MPIO Configuration	
Dell EqualLogic Host Integration Toolkit	Version 3.4
Dell EqualLogic MPIO Device Specific Module	• Maximum Sessions per Slice(array): 2 • Maximum Sessions per Volume: 6

3.2 Array Configuration

Six EqualLogic PS Series arrays were used in this test. All arrays in the Storage Group were the same model and using the same Array Software version. Array configuration information is described in Table 2 and Table 3 below.

Table 2 - Array Configuration Information

EqualLogic Storage	
Array Model	EqualLogic PS6010XV
Firmware	4.3.6

Table 3 - Pool Definitions

Pool	Array	RAID Policy	Primary IO Profile
A	EQL01	RAID 10	Small block Random I/O
	EQL02	RAID 10	
	EQL03	RAID 10	
B	EQL04	RAID 10	Small block Random I/O
	EQL05	RAID 10	
	EQL06	RAID 10	
C	EQL07	RAID10	Small block Random I/O
	EQL08	RAID10	
	EQL09	RAID10	
D	EQL10	RAID10	Medium Block Sequential I/O Large Block Sequential I/O
	EQL11	RAID10	
	EQL12	RAID10	

3.3 Switch Configuration

The Arista 7148sx is designed as a non-blocking, low latency, high-density 10Gb/s switch. It achieves these capabilities by leveraging a series of multiple application specific integrated circuits (ASICs) in a "Spine and Leaf" design that interconnects each leaf ASIC to each spine ASIC to provide non-blocking data paths between any two ports in the switch. This design, along with industry leading low latency connections and cut-through frame management allows the Arista 7148sx switch to provide excellent performance for high network utilization scenarios such as experienced with storage area networks.

3.3.1 Overview of Switch Settings

Table 4 provides an overview of the settings used to configure the switches for this SAN.

Table 4 - Switch Settings Used

Switch Model	Arista 7148sx
Switch inter-connection	• Dynamic Link Aggregation Protocol Link Aggregation Group (LACP - LAG) • Flow Control enabled on each port channel group • MTU=9212 on each port channel group (default) • Rapid Spanning Tree enabled on each port channel group
Global Switch Settings	Spanning-tree mode: rstp
Individual Port Settings	• MTU= 9212 (default) • "spanning-tree portfast" enabled on all edge ports • "flowcontrol send on" enabled on all edge ports • "flowcontrol receive on" enabled on all edge ports
Switch Firmware	EOS-4.5.1
Switch Hardware Rev	Hardware Rev 3.5 and 4.0
Host-Switch Cable Type	Dell SFP+ SR Optical Transceiver (DP/N 0N743D) (host); Arista SFP+ SR Optical Transceiver () (switch); LC-LC Fiber Optic Cable
Array-Switch Cable Type	Dell SFP+ SR Optical Transceiver (DP/N 0N743D) (array); Arista SFP+ SR Optical Transceiver () (switch); LC-LC Fiber Optic Cable
Switch-Switch Cable Type	SFP+ Copper Twinax (37-0960-01 Rev A0) 1 meter cable

3.3.2 Global Switch Settings

Spanning Tree mode is a global configuration setting for the Arista 7148sx.

Spanning Tree is enabled by default on the Arista 7148sx.

Rapid Spanning Tree (rstp) is enabled with the command

```
switch-1>enable
switch-1#config
switch-1(config)#spanning-tree mode rstp
```

3.3.3 Array and Server Port Configuration

All ports determined to connect to server or array interfaces are configured identically.

For all of these ports, "portfast" is enabled and flowcontrol send and receive are enabled. For this reference architecture, switch port 1 through port 36 on each switch were used for either server or array connections.

Jumbo frames are enabled by default on Arista switches.

The following commands were used to configure the array and server ports on each switch:

```
admin>enable
admin#config t
admin(config)#interface ethernet 1-36
admin(config-if-Et1-36)#flowcontrol send on
admin(config-if-Et1-36)#flowcontrol receive on
admin(config-if-Et1-36)#spanning-tree portfast
admin(config-if-Et1-36)#exit
admin(config)#
```

3.3.4 Link Aggregation Group Configuration

The Link Aggregation Group (LAG) consists of four (4) 10Gb/s links providing an aggregate bandwidth between each switch of 40Gb. Ports 45-48 of each switch was configured into a dynamic Link Aggregation Group (LAG) port channel using the IEEE Link Aggregation Control Protocol (LACP) standard.

The following commands will configure the single link aggregation group used to interconnect the four Arista 7148sx switches:

3.3.4.1 Switch-1:

```
switch-1>enable
switch-1#config t
switch-1(config)#interface ethernet 43-48
switch-1(config-if-Et43-48)#channel-group 1 mode active
switch-1(config-if-Et43-48)#exit
switch-1(config)#interface ethernet 37-42
switch-1(config-if-Et37-42)#channel-group 2 mode active
switch-1(config-if-Et37-42)#exit
switch-1(config)#interface port-Channel 1
switch-1(config-if-Po1)#spanning-tree cost 2000000
switch-1(config-if-Po1)#exit
switch-1(config)#
```

3.3.4.2 Switch-2:

```
switch-2>enable
```

switch-2#*config t*
switch-2(config)#*interface ethernet 43-48*
switch-2(config-if-Et43-48)#*channel-group 1 mode active*
switch-2(config-if-Et43-48)#*exit*
switch-2(config)#*interface ethernet 37-42*
switch-2(config-if-Et37-42)#*channel-group 2 mode active*
switch-2(config-if-Et37-42)#*exit*
switch-2(config)#*interface port-Channel 1*
switch-2(config-if-Po1)#*spanning-tree cost 2000000*
switch-2(config-if-Po1)#*exit*
switch-2(config)#

3.3.4.3 Switch-3:

switch-3>*enable*
switch-3#*config t*
switch-3(config)#*interface ethernet 43-48*
switch-3(config-if-Et43-48)#*channel-group 1 mode active*
switch-3(config-if-Et43-48)#*exit*
switch-3(config)#*interface ethernet 37-42*
switch-3(config-if-Et37-42)#*channel-group 2 mode active*
switch-3(config-if-Et37-42)#*exit*
switch-3(config)#

3.3.4.4 Switch-4:

switch-4>*enable*
switch-4#*config t*
switch-4(config)#*interface ethernet 43-48*
switch-4(config-if-Et43-48)#*channel-group 1 mode active*
switch-4(config-if-Et43-48)#*exit*
switch-4(config)#*interface ethernet 37-42*
switch-4(config-if-Et37-42)#*channel-group 2 mode active*
switch-4(config-if-Et37-42)#*exit*
switch-4(config)#

4 Conclusions

Lab Testing showed the following characteristics of this SAN solution:

- TCP retransmissions from arrays, as polled periodically from array counters and SANHQ were low (< 0.1%) across all test configurations
- Sequential Write performance scaled 98% in terms of throughput as measured at the host indicating that there are no performance issues or bottlenecks associated with the SAN infrastructure components or design.
- Sequential Read Performance scaled to approximately 95% in terms of throughput as measured at the host.
- The Random Read/Write performance in terms of IOs per second scaled 110% on a per server performance basis.

- TCP retransmissions from arrays, as polled periodically from array counters and SAN Headquarters (SANHQ) were low (< 0.1%) across all test configurations. This is another indicator that no bottlenecks or design issues within the switch that limited the ability of the switch to support the reference configuration.

In conclusion, the only limitation this reference design had with respect to scaling an EqualLogic SAN was in the number of available ports that four switches provide. In fact, due to our test methodology, there were unused ports on these switches that could potentially support additional arrays, though depending on the workload profile, some of these ports might need to be used to create larger inter-switch connections.

Appendix A Switch Port Mappings

1	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35	37	39	41	43	45	47	
H1P1	H3P1	H5P1	H7P1	A1C1P0	A3C1P0	A5C1P0	A7C1P0	A9C1P0	A11C1P0	A13C1P0	A15C1P0							S4P37	S4P39	S4P41	S2P43	S2P45	S2P47	SW1
H2P1	H4P1	H6P1	H8P1	A1C2P0	A3C2P0	A5C2P0	A7C2P0	A9C2P0	A11C2P0	A13C2P0	A15C2P0							S4P38	S4P40	S4P42	S2P44	S2P46	S2P48	
2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48	

1	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35	37	39	41	43	45	47	
H1P2	H3P2	H5P2	H7P2	A2C1P0	A4C1P0	A6C1P0	A8C1P0	A10C1P0	12C1P0	A14C1P0	A16C1P0							S3P37	S3P39	S3P41	S1P43	S1P45	S1P47	SW2
H2P2	H4P2	H6P2	H8P2	A2C2P0	A4C2P0	A6C2P0	A8C2P0	A10C2P0	12C2P0	A14C2P0	A16C2P0							S3P38	S3P40	S3P42	S1P44	S1P46	S1P48	
2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48	
1	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35	37	39	41	43	45	47	
H9P1	S11P1	H13P1	H15P1	A2C1P1	A4C1P1	A6C1P1	A8C1P1	A10C1P1	12C1P1	A14C1P1	A16C1P1							S2P37	S2P39	S2P41	S2P47	S2P45	S2P47	SW3
H10P1	H12P1	H14P1	H16P1	A2C2P1	A4C2P1	A6C2P1	A8C2P1	A10C2P1	12C2P1	A14C2P1	A16C2P1							S2P38	S2P40	S2P42	S2P48	S2P46	S2P48	
2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48	

1	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35	37	39	41	43	45	47	
H9P2	S11P2	H13P2	H15P2	A1C1P1	A3C1P1	A5C1P1	A7C1P1	A9C1P1	A11C1P1	A13C1P1	A15C1P1							S1P37	S1P39	S1P41	S2P47	S1P45	S1P47	SW4
H1OP2	H12P2	H14P2	H16P2	A1C2P1	A3C2P1	A5C2P1	A7C2P1	A9C2P1	A11C2P1	A13C2P1	A15C2P1							S1P38	S1P40	S1P42	S2P48	S1P46	S1P48	
2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	40	42	44	46	48	

Array Connections:

AxCyPz = Array x, Controller y, Port z (on that controller)

Link Aggregation Connections:

SxPy = Switch x, Port y

Host Connections

HxPy = Host x, NIC Port y

Appendix B IP Address Mappings

B.1 Array IP Addresses

Array:	Port 0	Port 1
EQL01	192.168.1.10	192.168.1.20
EQL02	192.168.1.11	192.168.1.21
EQL03	192.168.1.12	192.168.1.22
EQL04	192.168.1.13	192.168.1.23
EQL05	192.168.1.14	192.168.1.24
EQL06	192.168.1.15	192.168.1.25
EQL07	192.168.1.16	192.168.1.26
EQL08	192.168.1.17	192.168.1.27
EQL09	192.168.1.18	192.168.1.28
EQL10	192.168.1.19	192.168.1.29
EQL11	192.168.1.30	192.168.1.40
EQL12	192.168.1.31	192.168.1.41
EQL13	192.168.1.32	192.168.1.42
EQL14	192.168.1.33	192.168.1.43
EQL15	192.168.1.34	192.168.1.44
EQL16	192.168.1.35	192.168.1.45

B.2 Host IP Addresses

Server:	NIC 1	NIC 2
SVR01	192.168.1.210	192.168.1.211
SVR02	192.168.1.212	192.168.1.213
SVR03	192.168.1.214	192.168.1.215
SVR04	192.168.1.216	192.168.1.217
SVR05	192.168.1.218	192.168.1.219
SVR06	192.168.1.220	192.168.1.221
SVR07	192.168.1.222	192.168.1.223
SVR08	192.168.1.224	192.168.1.225
SVR09	192.168.1.226	192.168.1.227
SVR10	192.168.1.228	192.168.1.229
SVR11	192.168.1.230	192.168.1.231
SVR12	192.168.1.232	192.168.1.233
SVR13	192.168.1.234	192.168.1.235
SVR14	192.168.1.236	192.168.1.237
SVR15	192.168.1.238	192.168.1.239
SVR16	192.168.1.240	192.168.1.241