Best Practices for Deploying a Mixed 1Gb/10Gb Ethernet SAN using Dell Storage PS Series Arrays

Dell EMC Engineering December 2016

A Dell Best Practices Guide

Revisions

Date	Description
March 2011	Initial publication
April 2012	Correction in Figure 10
December 2016	Updated to include new branding and formatting

Acknowledgements

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

Engineering: Chuck Armstrong

The information in this publication is provided "as is." Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

Copyright © 2016 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, EMC, and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be the property of their respective owners. Published in the USA [12/8/2016] [Best Practices] [BP1008] Dell EMC believes the information in this document is accurate as of its publication date. The information is subject to change without notice.

Table of contents

Re	vision	S	2
Ac	knowl	edgements	2
1	Intro	duction	4
	1.1	Audience	4
2	Integ	gration strategies	5
	2.1	Design considerations for a mixed speed PS Series SAN	6
3	Swite	ch integration configuration	8
	3.1	Method 1	8
	3.2	Method 2	8
	3.3	Method 3	9
4	Pool	architecture and volume placement testing	11
	4.1	Comparing pool design and volume placement: 1Gb host initiator	11
	4.2	Comparing pool design and volume placement: 10Gb host initiator	15
	4.3	Analysis of pool design and architecture test results	18
5	Best	practice recommendations	19
А	1Gb	and 10Gb host SAN tests: full test results	20
	A.1	1Gb host SAN test results	20
	A.2	10Gb host SAN test results	20
В	Pool	/volume placement: full test results	22
С	Perfo	ormance testing configuration	26
D	Tech	nnical Support and resources	27
	D.1	Referenced or recommended documentation	27

1 Introduction

This paper presents network design recommendations and best practices for integration of 10Gb Dell[™] Storage PS Series arrays into an existing 1Gb PS Series iSCSI SAN environment.

By combining PS Series 10Gb storage arrays with industry standard Ethernet switches and the 10Gb Ethernet NIC options available with Dell EMC[™] PowerEdge[™] servers, our customers have the ability to build end-to-end 10Gb high-performance iSCSI Storage Area Network (SAN) infrastructures.

Two methods of introducing 10Gb components into a datacenter environment are end-to-end (10Gb components only) or mixed (1Gb and 10Gb working together). This paper focuses on the mixed method of introduction.

Some of the many reasons to use a mixed integration of 10Gb components into an existing 1Gb SAN environment are:

- Introducing new PS Series arrays equipped with 10Gb controllers into existing 1Gb PS Series SAN
- Introducing new PowerEdge servers equipped with 10Gb NICs, for converged networking projects
- Application I/O throughput requirements are reaching the limits of a 1Gb SAN environment

1.1 Audience

This paper is for Solution Architects, Storage Network Engineers, System Administrators, and IT Managers. Working knowledge by the reader of network switch deployment and operation, network administration and Dell Storage PS Series iSCSI SAN operation is assumed.

2 Integration strategies

Before we begin describing integration strategies in detail, we first want to define the fully redundant iSCSI SAN designs used as building blocks for integration and testing. Figure 1 shows a 1Gb redundant iSCSI SAN design using a pair of stacked Dell Networking N or S Series 1Gb switches connecting to one or more PS Series 1Gb storage arrays. Figure 2 shows a similarly designed 10Gb iSCSI SAN using a pair of stacked Dell Networking N or S one or more Dell Storage PS Series 10Gb storage arrays.

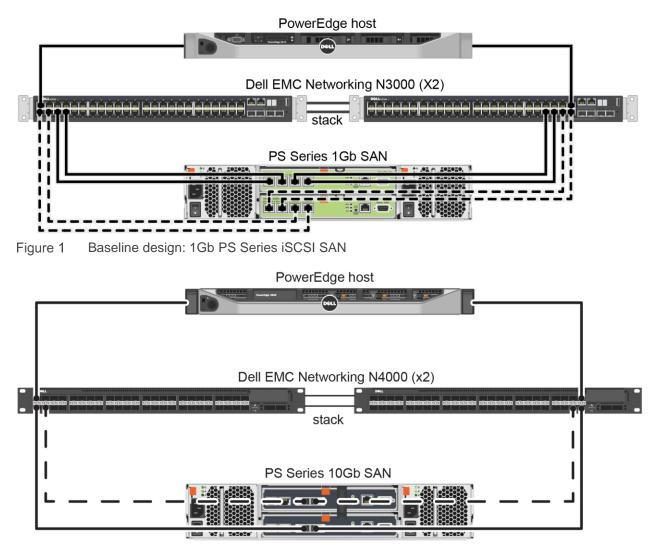


Figure 2 Baseline design: 10Gb PS Series iSCSI SAN

5

The key differences between the 1Gb and 10Gb SAN designs shown in Figure 1 and Figure 2 are:

- The PS Series 1Gb array is equipped with two 4-port 1Gb controllers.
- The PS Series 10Gb array is equipped with two 2-port 10Gb controllers.

2.1 Design considerations for a mixed speed PS Series SAN

To design mixed speed PS Series SANs, two key integration questions require focus:

First: How do you properly design a mixed speed switched Ethernet network fabric?

Proper design and configuration of the Ethernet switching fabric for 1Gb SANs is critical to achieving and maintaining performance and reliability requirements. The same is true for 10Gb SANs. When you begin to integrate 10Gb components into an existing SAN consisting of 1Gb arrays you must first carefully consider how you design the integration points for the 10Gb switches. Integration points are those connections between hosts (NICS), switches and arrays. This design decision can be critical to the performance and reliability of the integrated SAN.

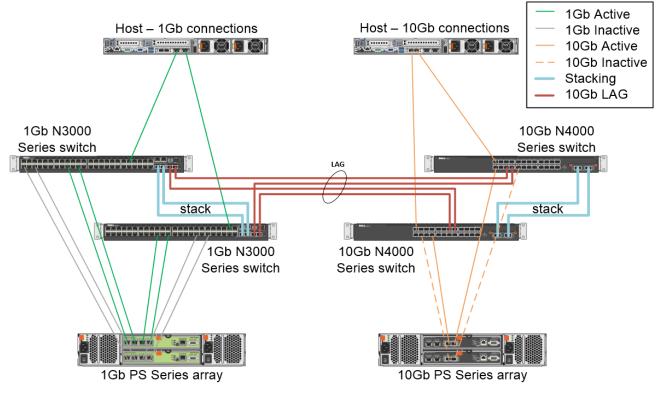
Second: How should you configure storage pools (array membership) and volume layout within those pools when combining 1Gb and 10Gb arrays together in the same SAN?

PS Series Peer Storage functionality allows you to combine multiple arrays (or "members") into a single logical storage pool. Physical storage of volumes is automatically distributed across the pool members. This behavior means that careful consideration to pool membership and volume layout must be taken when mixing 1Gb and 10Gb storage arrays together in the same SAN.

Note: For additional information, see the Dell PS Series Configuration Guide, available at: http://en.community.dell.com/dell-groups/dtcmedia/m/mediagallery/19852516.

There are many ways to design an iSCSI SAN that integrates 1Gb and 10Gb networks together. A significant factor affecting the design of a mixed speed iSCSI SAN is the feature set and port density provided by the Ethernet switches used in the SAN. To create a redundant, mixed speed iSCSI SAN, we recommend that at a minimum you start with dual 1Gb and dual 10Gb switches.

Figure 3 shows a design pattern for combining 1Gb and 10Gb PS Series SANs. The interconnect between the mixed speed components in the SAN is accomplished using a 40Gb LAG group as uplink connections between the stacked 1Gb switches and the stacked 10Gb switches.





Referring to Figure 3, there are some important design considerations to be aware of:

- Each of the 1Gb switches is configured with one dual-port 10Gb uplink module and one stacking module. The stacking module is used to create a single logical switching fabric out of the two 1Gb switches. The 10Gb uplink modules are used for creating a single 40Gb LAG (all four ports from the stack) uplinks to the 10Gb stacked switches.
- The 40Gb LAG uplink between the 1Gb and 10Gb switches are cross-connected so that each 10Gb switch physically connects to both switches in the 1Gb stack for redundancy.
- Each of the 10Gb switches is configured with a stacking module. The stacking module is used to create a single logical switching fabric out of the two 10Gb switches.

You may be wondering, why not just attach my 10Gb arrays directly to the 10Gb uplink ports on the Dell Networking N3000 Series switches? While this looks like a good idea for small installations, because of performance and reliability considerations, we recommend NOT using this connection strategy. One of the primary reasons for this is that the buffering available on the 10Gb uplink ports on 1Gb switches were never designed to have utilization devices (host or storage devices) connecting to these ports.

3 Switch integration configuration

Connecting the 1Gb and 10Gb switch fabrics for iSCSI storage traffic can be accomplished using one of three methods, depending on the capabilities of the switches being used.

Note: Stacking is a method by which multiple switches are connected using a proprietary stacking port and thereafter are managed as, and behave as a single switch. A LAG switch uplink connection uses multiple physical connections, but bundles them together for throughput and individual link redundancy.

3.1 Method 1

The optimal method, if both pair of switches support stacking, is to create a single LAG connection between the stacked 1Gb and the stacked 10Gb fabrics. The individual connections that make up the LAG will connect each of the 10Gb switches to each of the 1Gb switches: providing redundancy in the case of a single switch failure. See Figure 4.

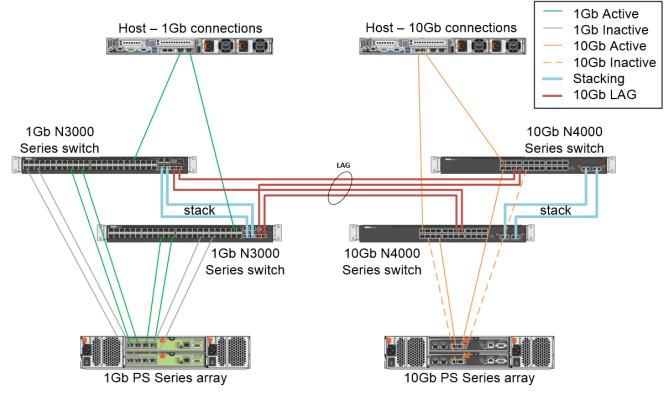


Figure 4 Stacked 1Gb fabric LAGd to stacked 10Gb fabric

3.2 Method 2

If either the 1Gb or the 10Gb switches in use do not support stacking, those switches must be connected using a LAG. In this case, the LAGd switch fabric would connect to the stacked fabric through the use of four LAG connections that can use either split (cross) connect or straight connect. Straight connect LAG

connections can result in loss of access to storage in the event of a switch failure. Figure 5 shows an example of split connect. Additionally, Rapid Spanning Tree (RSTP) would need to be configured appropriately to block the loop path for traffic.

Note: Configuration documentation for specific supported switches can be found at the <u>Rapid EqualLogic</u> <u>Configuration Portal</u>.

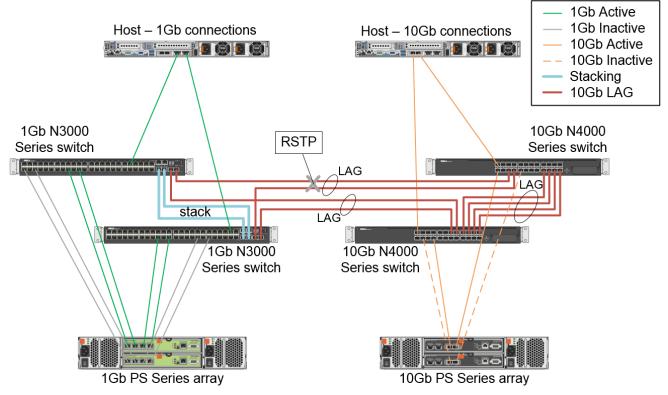


Figure 5 Stacked 1Gb fabric LAGd to a LAG-connected 10Gb fabric – use of RSTP is required

3.3 Method 3

Lastly, if neither the 1Gb or the 10Gb switches support stacking, both the 1Gb and 10Gb switches must be connected together using LAGs to create each switching fabric. In this case, the two switching fabrics (1Gb and 10Gb) can only be connected to each other using the straight connect LAG method as shown in Figure 6. This configuration would also require the use of RSTP to block the loop path for traffic.

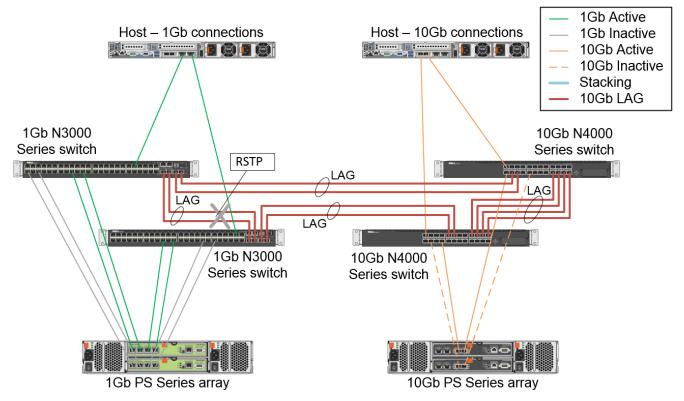


Figure 6 LAG connections for all 1Gb and 10Gb switches – straight-through LAG for each connection

Pool architecture and volume placement testing

When integrating 10Gb arrays into existing 1Gb PS Series SANs you must also consider how you locate volumes within the available pools. This section focuses on answering two key questions:

- Will you get better performance having 1Gb initiators access volumes on 10Gb arrays?
- Do you place 10Gb arrays in the same pools as existing 1Gb arrays?

PS Series arrays are very flexible in regards to their configuration—up to and including the placement of arrays into storage pools. The pool is a very important construct within a PS Series SAN, as they control the load balancing of volumes across multiple arrays, redundancy among those arrays, and inter-array communication.

It is possible to place arrays with different speed controllers in the same pool. But is it advisable to do so? This question has become important, as customers now have the ability to add 10Gb arrays into existing 1Gb SANs.

To provide answers to these questions, three I/O workloads (see Table 1) were run against different configurations of volumes and pools. Each workload was run three times for 15 minutes, then average I/O performance data was calculated. The goal was to measure performance differences between same speed pools and the mixed speed pools.

Workload type	Block size	Read/Write ratio
Sequential read	256 K	100% / 0%
Sequential write	64 K	0% / 100%
Random read/write	8 K	67% / 33%

Table 1 Workloads

4

The tests performed for this paper use the split connect LAG method (shown in section 3.2) for connecting the 1Gb and 10Gb switch fabrics. The test information and results have been brought forward from the previous version of this paper as there have been no changes in the PS Series storage architecture that would significantly change these results. As such, at the time of testing, the 10Gb switches used did not support stacking.

4.1 Comparing pool design and volume placement: 1Gb host initiator

Note: These test results were brought forward from the previous version of this paper. Best practice recommendations remain the same with update PS Series storage.

Tests for 1Gb initiator IO paths:

• A: One PowerEdge server with two 1Gb SAN ports connecting to four volumes in one **PS Series 1Gb** array through stacked Dell Networking N3000 Series switches.

- B: One PowerEdge server with two 1Gb SAN ports connecting to four volumes in one PS Series 10Gb array via stacked Dell Networking N3000 Series switches and LAGd Dell Networking N4000 Series switches
- C: One PowerEdge server with two 1Gb SAN ports connecting simultaneously to separate speed pools: two volumes in one PS Series 1Gb array and two volumes in one PS Series 10Gb array.
- D: One PowerEdge server with two 1Gb SAN ports connecting simultaneously to four volumes stored in a **mixed speed pool** consisting of one **PS Series 1Gb array** and one **PS Series 10Gb array**.

Parts A and B provide I/O performance data for comparing the difference when connecting 1Gb initiators to 1Gb vs. 10Gb arrays.

Parts C and D provide I/O performance data for comparing I/O performance when a 1Gb initiator simultaneously connects to volumes in separate speed pools vs. connecting to the same set of volumes in a single mixed speed pool.

Figure 7 illustrates the I/O flow paths for the 1Gb host initiator accessing 1Gb storage targets or 10Gb storage targets. Figure 8 shows the comparison of performance between 1Gb storage and 10Gb storage targets where the 1Gb storage target is the baseline.

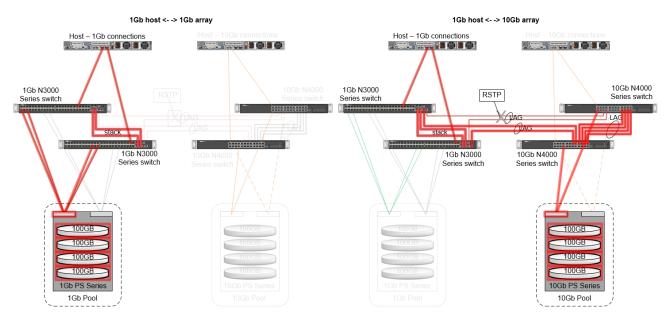
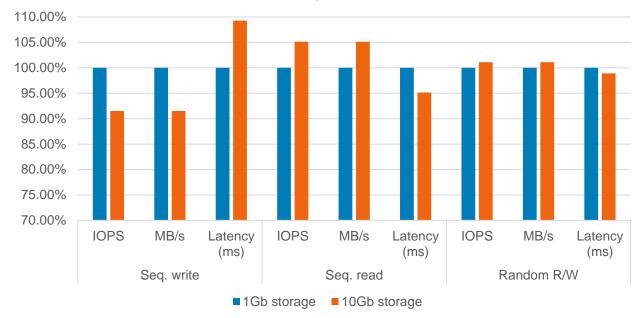


Figure 7 1Gb host performance comparison: Connecting to 1Gb array versus 10Gb array

1Gb host performance





The information shown in Figure 8 identifies some differences in sequential I/O performance. After moving the target volumes over to the 10Gb array, there was a decrease in sequential write performance and an increase in sequential read performance. However, there was negligible performance difference for random I/O.

Figure 9 illustrates the I/O flow paths for the 1Gb initiator accessing volumes in separate speed pools or volumes in a mixed speed pool. Figure 10 shows the comparison of performance between volumes being accessed on separate speed pools and on a single mixed speed pool.

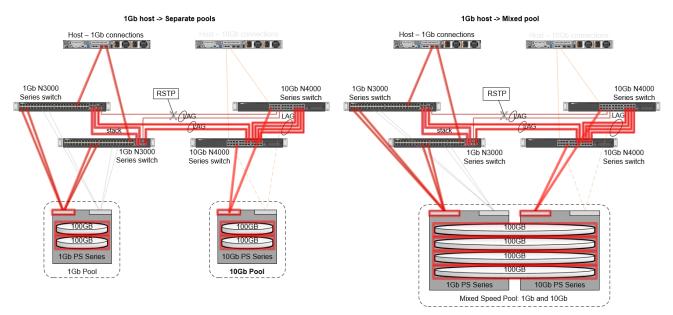
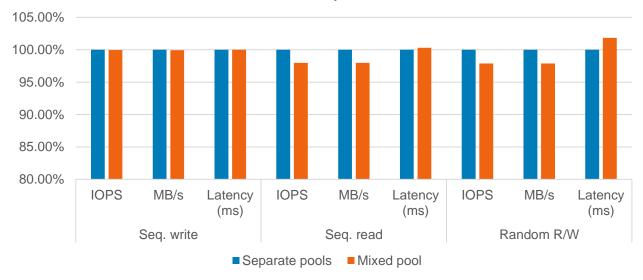


Figure 9 1Gb host pool performance comparison: Mixed speed targets in separate pools versus targets in a single mixed speed pool



1Gb host: Pool performance



The information shown in Figure 10 indicates very little difference in I/O performance when the 1Gb initiator is connecting to separate speed pools versus mixed speed pools.

From this series of 1Gb initiator test results we can conclude that the placement of the volumes within a single pool versus separating them out by speed does not have a significant impact on performance. In all cases, throughput was limited by the 1Gb bandwidth available to the host. Latency and IOPS provided similar results.

4.2 Comparing pool design and volume placement: 10Gb host initiator

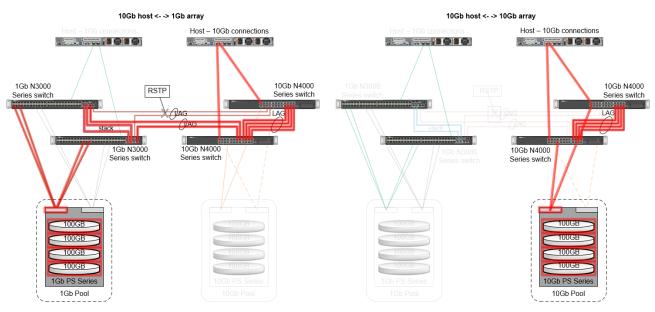
The same test sequences from the previous section were repeated using a 10Gb initiator on the host:

- E. One PowerEdge server with two 10Gb SAN ports connecting to four volumes in one **PS Series 10Gb array** using LAGd Dell Networking N4000 Series switches.
- F. One PowerEdge server with two 10Gb SAN ports connecting to four volumes in one PS Series 1Gb array using LAGd Dell Networking N4000 Series switches and stacked Dell Networking N3000 Series switches.
- **G**. One PowerEdge server with two 10Gb SAN ports connecting simultaneously to **separate speed pools**: two volumes in one **PS Series 1Gb array** and two volumes in one **PS Series 10Gb array**.
- H. One PowerEdge server with two 10Gb SAN ports connecting simultaneously to four volumes stored in a mixed speed pool consisting of one PS Series 1Gb array and one PS Series 10Gb array.

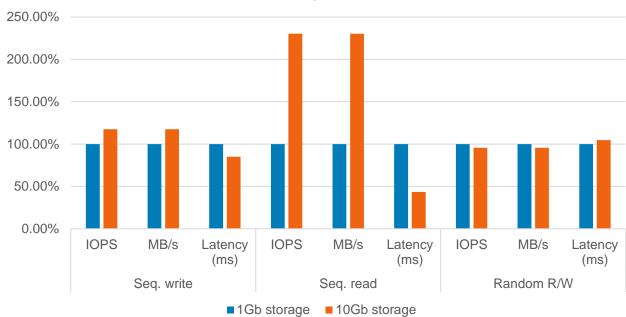
Parts E and F provide I/O performance data for comparing the difference when connecting 10Gb initiators to 1Gb vs. 10Gb arrays.

Parts G and H provide I/O performance data for comparing I/O performance when a 10Gb initiator simultaneously connects to volumes in separate speed pools versus connecting to the same set of volumes in a single mixed speed pool.

Figure 11 illustrates the I/O flow paths for the 10Gb host initiator accessing 1Gb storage targets or 10Gb storage targets. Figure 12 shows the comparison of performance between 1Gb storage and 10Gb storage targets, where the 1Gb storage target is the baseline.







10Gb host performance



The information shown in Figure 12 indicates sequential read performance significantly increases with 10Gb storage targets as compared to 1Gb storage targets. As would be expected for sequential I/O, IOPS and MB/s are considerably higher and latency is considerably lower on the 10Gb storage target as compared to the 1Gb storage target.

Figure 13 illustrates the I/O flow paths for the 10Gb initiator accessing volumes in separate speed pools or volumes in a mixed speed pool. Figure 14 shows the comparison of performance between volumes being accessed on separate speed pools and on a single mixed speed pool.

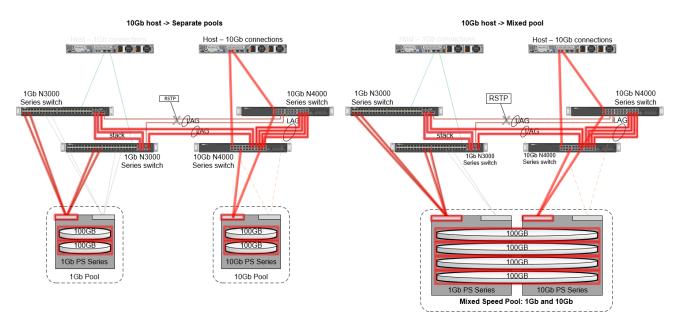
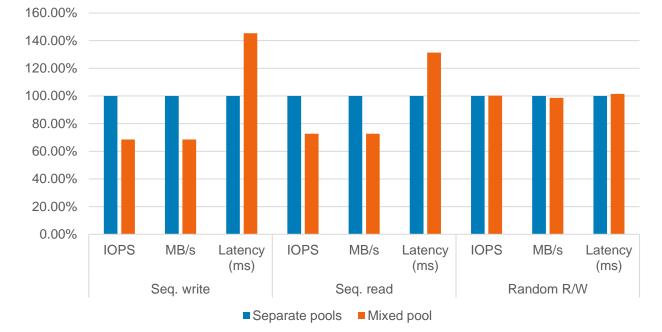


Figure 13 10Gb host pool performance comparison: Mixed speed targets in separate pools versus targets in a single mixed speed pool



10Gb host: Pool performance

Figure 14 Comparison of 10Gb initiator connecting to separate pools and mixed pools

The information shown in Figure 14 indicates a significant decrease in performance with a 10Gb initiator when using mixed speed pools for sequential IO workloads versus separate speed pools.

4.3 Analysis of pool design and architecture test results

For 1Gb host initiators, it is apparent that the placement of volumes in either separate, single speed pools or a single mixed speed pool results in very little performance change. However, 1Gb host initiators that will be migrated to 10Gb host initiators in the future need to be considered. For hosts that will be migrated, the recommendations for 10Gb initiators should be followed. The performance at the 1Gb level would not be affected for these migrated hosts, but configuring for 10Gb connectivity initially would avoid the need to either recreate volumes on or migrate volumes to 10Gb arrays when the host initiator migration occurs.

The second question of whether 1Gb initiators will benefit from accessing 10Gb volumes is also answered positively. No matter the decision on placement of volumes, performance increases can be achieved by accessing new volumes on the 10Gb arrays.

For 10Gb initiators, the placement of those volumes in separate speed pools is much more critical. As seen in the test results, accessing data located only on 1Gb arrays from a 10Gb initiator results in much lower performance. Because of this, accessing volumes on separate speed pools is less affected than accessing volumes on mixed speed pools. Additionally, performance penalties from accessing the lower speed arrays would likely be worsened when multiple 10Gb initiators are in play. This is especially true of latency and TCP retransmits.

5 Best practice recommendations

The results presented in this paper indicate that attention must be paid to some important SAN design considerations when integrating of 10Gb PS Series arrays into an existing 1Gb PS Series SAN.

- When integrating 10Gb switches into your existing 1Gb environment, utilize stacking on both sets of switches if possible, and connect the stacked switches using split uplink LAG to create a highly available switch fabric.
- When connecting 1Gb switches and 10Gb switches together where either or both speed switches do not support stacking, be aware of where Rapid Spanning Tree is enabled to block links, ensuring that 10Gb traffic (i.e. PS Series inter-array data flow) never crosses the 1Gb switch.
- You must configure pools and volumes in a way that minimizes impact to I/O performance.
 - If you have predominately 1Gb initiators, start upgrading your arrays to 10Gb for comparable or better performance across almost all situations.
 - If you have predominately 10Gb initiators, you should only access data and volumes residing on 10Gb arrays (from those initiators). You may see high latency and retransmit rates when 10Gb initiators connect to 1Gb targets.
 - When adding 10Gb arrays, place them in separate pools from your 1Gb arrays.

A 1Gb and 10Gb host SAN tests: full test results

A.1 1Gb host SAN test results

The baseline 1Gb I/O performance data in Figure 15 corresponds to two 1Gb host ports accessing four 100GB volumes on a single 1Gb array. During the baseline test, all traffic was isolated to the Dell Networking N3000 stack, as all the 1Gb ports were directly connected – no traffic crossed the uplink ports to the Dell Networking N4000 fabric.

The comparison 1Gb I/O performance data in Figure 15 corresponds to two 1Gb host ports accessing four 100GB volumes on a single 10Gb array. During the comparison test, all traffic flowed through the Dell Networking N3000 stack, the LAG uplink ports, and then to the Dell Networking N4000 fabric to reach the 10Gb PS Series array.

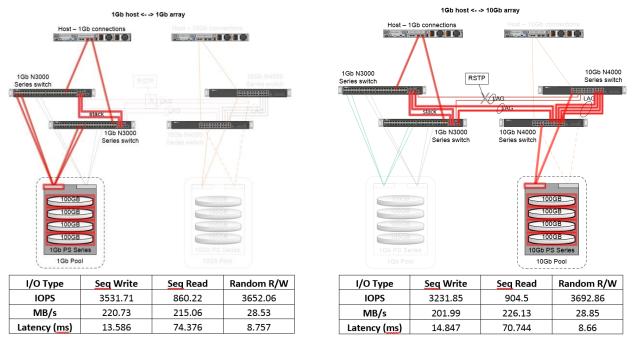


Figure 15 Full test results: 1Gb initiator performance comparison

A.2 10Gb host SAN test results

The baseline 10Gb I/O performance data in Figure 16 corresponds to two 10Gb host ports accessing four 100GB volumes on a single 1Gb array. During the baseline test, all traffic flowed through the Dell Networking N4000 fabric, the LAG uplink ports, and then to the Dell Networking N3000 stack to reach the 1Gb PS Series array.

The comparison 10Gb I/O performance data in Figure 16 corresponds to two 10Gb host ports accessing four 100GB volumes on a single 10Gb array. During the comparison test, all traffic was isolated to the Dell

Networking N4000 switches, as all of the 10Gb ports were directly connected, no traffic crossed the uplink ports to the Dell Networking N3000 stack.

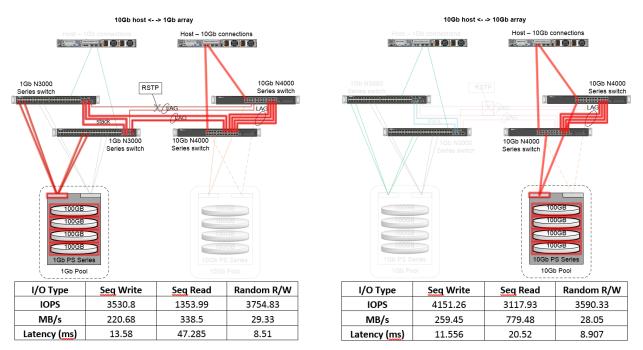
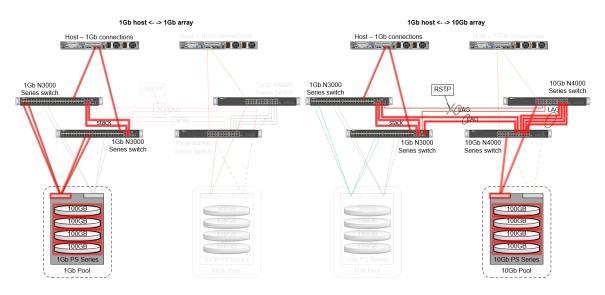


Figure 16 Full test results: 10Gb initiator performance comparison

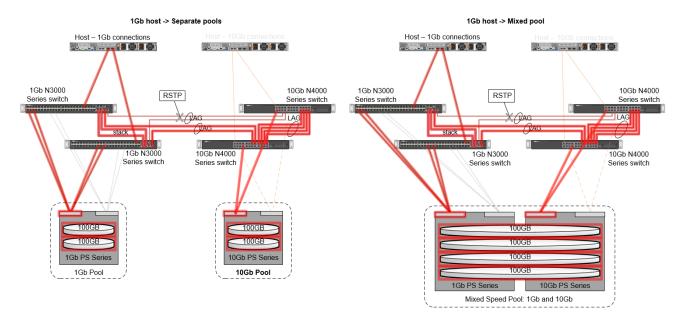
Pool/volume placement: full test results

Β



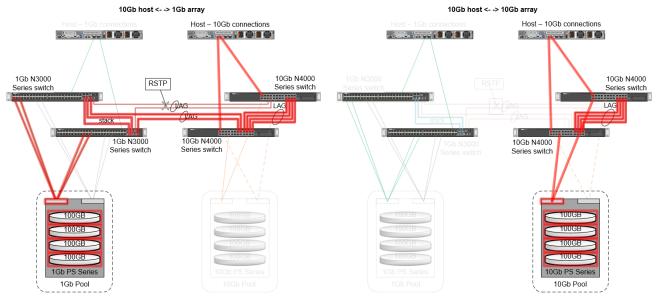
I/O Type		1Gb Array	10Gb Array
	IOPS	3531.71	3231.85
Seq. Write	MB/s	220.73	201.99
	Latency (<u>ms</u>)	13.586	14.847
	IOPS	860.22	904.5
Seq. Read	MB/s	215.06	226.13
	Latency (<u>ms</u>)	74.376	70.744
	IOPS	3652.06	3692.86
Random R/W	MB/s	28.53	28.85
	Latency (<u>ms</u>)	8.757	8.66

Figure 17 Full test results: 1Gb initiator connecting to 1Gb targets versus 10Gb targets



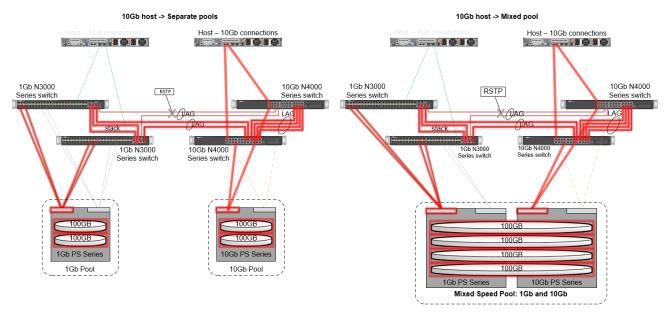
I/O Type		Separate Pools	Mixed Pool
	IOPS	3687.7	3685.55
Seq. Write	MB/s	230.49	230.35
	Latency (ms)	13.01	13.01
	IOPS	923.25	904.57
Seq. Read	MB/s	230.81	226.14
	Latency (<u>ms</u>)	70.57	70.78
	IOPS	5344.48	5231
Random R/W	MB/s	41.76	40.87
	Latency (ms)	6	6.11

Figure 18 Full test results: 1Gb initiator connecting to separate speed versus mixed speed pools



I/O Type		1Gb Array	10Gb Array
	IOPS	3530.8	4151.26
Seq. Write	MB/s	220.68	259.45
	Latency (<u>ms</u>)	13.58	11.55
	IOPS	1353.99	3117.93
Seq. Read	MB/s	338.5	779.48
	Latency (<u>ms</u>)	47.285	20.52
	IOPS	3754.83	3590.33
Random R/W	MB/s	29.33	28.05
	Latency (<u>ms</u>)	8.51	8.9

Figure 19 Full test results: 10Gb initiator connecting to 1Gb targets versus 10Gb targets



I/O Type		Separate Pools	Mixed Pool
	IOPS	7670.07	5253.69
Seq. Write	MB/s	479.38	328.35
	Latency (ms)	6.28	9.13
	IOPS	3427.19	2492.79
Seq. Read	MB/s	856.79	623.2
	Latency (ms)	19.75	25.945
	IOPS	5394.02	5408.58
Random R/W	MB/s	42.14	41.55
	Latency (<u>ms</u>)	5.92	6.01

Figure 20 Full test results: 10Gb initiator connecting to separate speed versus mixed speed pools

C Performance testing configuration

Although the testing was performed previously, the recommendations remain the same as the architecture and functionality remain unchanged with regard to these tests. As a result, it was not necessary to re-run the tests in order to revise this document.

The major hardware components for the test system configuration used throughout this document are:

Test configuration	Hardware components
1 Gigabit iSCSI initiator	Dell PowerEdge R710 Server: • 2 x Quad Core Intel® Xeon® X5570 Processors • 8 M Cache, 2.93 GHz • 32 GB RAM, • 2 x 146GB 10K SAS internal disk drives • Broadcom 5709 1Gb quad-port NIC (LAN on motherboard)
10 Gigabit iSCSI initiator	Dell PowerEdge R710 Server: • 2 x Quad Core Intel® Xeon® X5570 Processors • 8 M Cache, 2.93 GHz • 32 GB RAM • 2 x 146 GB 10K SAS internal disk drives • Broadcom NetXtreme II 57711 10Gb NIC, Dual-Port
Network	2 x Dell Dell Networking N3000 Series 1Gb Ethernet Switch 2 x Dell Dell Networking N4000 Series 10Gb Ethernet Switch
Storage	 Dell PS Series PS6000E: 16 x 450GB SATA disks with dual 4 port 1 Gb controllers Configured as RAID 10, with two hot spares Dell PS Series PS6010E: 16 x 450GB SATA disks with dual 2 port 10 Gb controllers Configured as RAID 10, with two hot spares

Table 2 Test configuration hardware components

The major software components for the test system configuration used throughout this document are:

Table 5 Test connunation soltware components	Table 3	Test configuration	software	components
--	---------	--------------------	----------	------------

Test configuration	Software components
Initiators	Windows 2008 R2 Server • MS iSCSI Initiator • PS Series Host Integration Toolkit • VDBench version 5.0.2 IO Generation software
Monitoring tools	 PS Series SAN Headquarters version 2.0 EQL_Monitor: Dell developed PERL-based monitoring script PerfMon

D Technical Support and resources

Dell.com/support is focused on meeting customer needs with proven services and support.

<u>Dell TechCenter</u> is an online technical community where IT professionals have access to numerous resources for Dell software, hardware and services.

<u>Storage Solutions Technical Documents</u> on Dell TechCenter provide expertise that helps to ensure customer success on Dell Storage platforms.

D.1 Referenced or recommended documentation

Dell publications:

- Integrating PS Series PS6x10 Arrays with Existing SANs <u>http://www.equallogic.com/resourcecenter/assetview.aspx?id=9447</u>
- PS Series Configuration Guide
 <u>http://www.delltechcenter.com/page/EqualLogic+Configuration+Guide</u>
- Dell PS Series PS Series Network Performance Guidelines http://www.equallogic.com/resourcecenter/assetview.aspx?id=5229