PS Series iSCSI Volume Connection Count Maximum Characterization

Dell EMC Engineering
November 2016

A Dell EMC Technical White Paper

Revisions

Date	Description
February 2012	Initial publication
March 2012	Updates to section 5
December 2016	Updated to reflect industry changes

Acknowledgements

Updated by: David Glynn

Table of contents

Revisions....2
Acknowledgements....2
1 Introduction....4
1.1 Audience....4
1.2 Question....4
1.3 Results....4
2 Overview....5
2.1 PS Series SAN architecture....5
2.2 PS Series iSCSI connection maximums....5
2.3 iSCSI connection behavior....5
2.3.1 Native MPIO modules....6
2.3.2 PS Series MPIO modules....6
2.3.3 Volume migration and replication....6
2.3.4 Storage initiator limitations....7
3 Performance testing....8
3.1 Test environment....8
3.2 Results....8
4 Availability testing....12
4.1 Test environment....12
5 Recommended practices....13
5.1 Calculating total iSCSI connections....13
5.2 Tips for reducing iSCSI connections....14
A Hardware diagram....15
B Logical diagram....16
C Solution infrastructure....17
D Vdbench parameters....18
E Additional resources....19

1 Introduction

This section describes the intended audience of the paper, explains the purpose of the paper, and presents a summary of findings.

1.1 Audience

The information in this technical paper is intended for anyone designing, quoting, deploying, administering, or scaling out a Dell EMC PS Series storage solution.

1.2 Question

Because both PS Series and OS-native MPIO modules facilitate multiple, simultaneous connections to iSCSI volumes, the total number of iSCSI connections can be surprisingly large, even with a relatively small number of storage hosts, PS Series member arrays, and storage volume targets. As this paper's topic suggests, there is a limit to the number of iSCSI connections for both a storage group and a storage pool at which a PS Series SAN configuration can operate. This paper will focus on the iSCSI connection maximum for a single pool. It will provide an overview of the factors involved in iSCSI connection creation, define the parameters of the iSCSI connection maximum, and examine whether the SAN performance and availability are affected as the maximum is approached.

1.3 Results

Performance testing results were very consistent, even as the number of iSCSI connections increased from 510 to 1020. The supported maximum for this storage pool configuration is 1024. Consistency was seen in throughput, IOPS, and response time on different pool sizes and under various common workloads.

Availability testing consisted of introducing a common network failure event while the PS Series SAN was under a moderate I/O load generated by Vdbench. This was done at different numbers of iSCSI connections to determine if operating near the supported maximum compromised the fault tolerance of the PS Series SAN. No storage volume timeouts or other anomalies were observed.

From these results, it is clear that the PS Series SAN architecture provides consistent performance and availability while operating within the range of iSCSI connection quantities validated and supported by Dell EMC.

2 Overview

This section provides an overview of the PS Series SAN architecture and gives specific information on when and how iSCSI connections are made by common storage initiators.

2.1 PS Series SAN architecture

The PS Series SAN architecture enables simultaneous access from multiple storage initiators to multiple storage targets distributed across multiple peer storage (PS) member arrays using iSCSI, the Internet Protocol (IP) based storage networking standard. For high availability, each initiator can make multiple iSCSI connections to an individual target volume in a process called multi-pathing. Using MPIO, servers can send multiple I/O streams to a SAN volume concurrently. MPIO routes I/O over redundant paths that connect servers to storage and also manages these paths so that requests can be re-routed in the event of a failure of one of the components. MPIO also provides increased redundancy and improves performance of application data hosted on the SAN.

In addition to the native multi-path functionality in Microsoft Windows Server® and VMware® ESXi®, there are PS Series MPIO extensions that integrate into these native MPIO implementations. This provides advanced features such as intelligent routing of I/O and queue depth load balancing to enhance performance, particularly on multi-member array configurations where individual storage volumes are distributed across more than one PS Series member array creating volume slices. As the number or location of volume slices and the number of available PS Series member array network ports change, PS Series MPIO modules such as the Device Specific Module (DSM) for Windows and the Multipath Extension Module (MEM) for VMware can optimize the iSCSI connection quantity and load balance accordingly.

2.2 PS Series iSCSI connection maximums

Table 1 shows the maximum iSCSI connection counts for different types of PS Series member array pools using different member array firmware versions.

Table 1 Supported maximum iSCSI connection counts by firmware version and PS member array type

PS6xxx pool	PS6xxx group (using four pools)	PS4xxx pool	PS4xxx group (using four pools)
1024	4096	256	512

Testing was completed using PS6000XV arrays, therefore the iSCSI connection maximum in effect for a storage pool was 1024.

2.3 iSCSI connection behavior

This section discusses when and how iSCSI connections are created by different MPIO modules and by volume migration and replication as well as important storage initiator limitations.

2.3.1 Native MPIO modules

Table 2 shows the number of iSCSI connections created by default in Windows Server and ESXi when using the native MPIO modules and software initiator of each OS to connect to a storage target. Additional paths may be configured manually depending on the OS. Note that the connection count is not affected by the number of member arrays (and volume slices). In ESXi, one iSCSI connection is made for each VMkernel port assigned to the software initiator.

Table 2 Default number of native MPIO iSCSI connections created when connecting to a storage target

Storage initiator OS	One member array in pool	Three member arrays in pool
Windows Server	2	2
VMware ESXi	One for each VMkernel port assigned to the iSCSI software initiator	One for each VMkernel port assigned to the iSCSI software initiator

2.3.2 PS Series MPIO modules

Table 3 shows the number of iSCSI connections created by default in Windows Server and VMware ESXi when using the PS Series MPIO modules and software initiator of each OS to connect to a storage target. Within the respective MPIO module interfaces, the number of connections per volume slice and the maximum number of connections per volume can be changed. The default values are two connections per volume slice for a maximum of six connections per volume. Under normal circumstances, volumes are not spanned across more than three member arrays, so the number of volumes slices will not be more than three. Note that contrary to native MPIO, the number of member arrays (and volume slices) affects the connection count; however the number of network interfaces on the SAN network available to the OS does not.

Table 3 Default number of MPIO iSCSI connections created when connecting to a storage target

Storage initiator OS	One member array in pool	Three member arrays in pool
Windows Server	2	6
VMware ESXi	2	6

2.3.3 Volume migration and replication

Two PS Series SAN administrative processes that can temporarily affect the total number of iSCSI connections are volume migration and volume replication.

Standard volume migration within a storage group is handled by internal connections among the PS Series member arrays and will not result in an increase in iSCSI connections which count against the maximum. However, one way that volume migration does increase iSCSI connections is by temporarily increasing the number of slices for a volume. During volume migration, the number of volume slices in existence will temporarily increase when a single volume is migrated from one member array to another member array as for a period of time the volume will reside on higher number of member arrays. If the storage initiator is running a PS Series MPIO module that is dynamically tracking volume slice quantity and location, then the total number of iSCSI connections will temporarily increase before returning to the original quantity.

In the case of volume replication, iSCSI connections are temporarily created during replication at the destination storage group where the volume is being replicated at the rate of one iSCSI connection for each source volume slice. The automated volume replication functions in a way that does not require large numbers of iSCSI connections to be opened simultaneously, but the administrator might notice a small number of connections coming into existence during the actual volume replications.

2.3.4 Storage initiator limitations

During testing, the following iSCSI connection limitations were observed in the host OS.

In Windows Server, the maximum number of iSCSI connections that the host can initiate is 255. This is a hard limit within the Windows OS that cannot be changed.

In ESXi, the maximum number of iSCSI connections that the software iSCSI initiator supports is 1024. The PS Series MEM limits the maximum number of iSCSI connections it will initiate to 512 by default. This limit can be increased to 1024 if necessary, though this is rarely the case, using the following MEM installation script command:

```
esxcli equallogic param set --name=TotalSessions --value=1024
```

Note: For information on the latest version of MEM, see the *Multipathing Extension Module Installation and User Guide* on eqlsupport.dell.com (login required).

3 Performance testing

This section covers in greater detail the methods and results of the testing used to determine the effect of larger numbers of iSCSI connections on storage pool performance.

3.1 Test environment

In order to determine whether or not SAN performance is affected as the total number of storage pool iSCSI connections increases towards the officially supported maximum, we used the performance tool Vdbench to capture throughput, IOPS, and application latency. This tool is a disk and tape I/O workload generator for testing data integrity and assessing storage performance.

Note: Vdbench is available at: http://sourceforge.net/projects/vdbench/

A single physical server running Windows Server with the PS Series MPIO module installed ran Vdbench workloads on 34 volumes residing on one-, two-, and three-member array pools.

The server had four 1 GbE network interfaces on the SAN network and two iSCSI connections per volume slice.

The performance metrics of IOPS and Vdbench response time were captured at the server while each pool was accommodating total connections of 510, 750, and 1020.

The exact storage pool connection totals were reached using a second server running vSphere ESXi hosting multiple Windows Server virtual machines to connect to the volumes under load. In order to isolate the effect of the PS Series SAN handling greater numbers of iSCSI connections, the overall load on the SAN was kept consistent by not generating I/O over those additional connections as those connections were added, but by maintaining the I/O load generated by the single physical Windows server on the 34 volumes. Both the vSphere ESXi server and the Windows virtual machines were also running PS Series MPIO modules. The following three workloads were defined:

- **Test 1:** 8 KB transfer size, random I/O, 67% read
- **Test 2:** 64 KB transfer size, sequential I/O, 100% write
- **Test 3:** 256 KB transfer size, sequential I/O, 100% read

Each workload was run for 30 minutes and the I/O rate was not capped (the Vdbench iorate parameter was set to **max**).

3.2 Results

The following three figures show the results of the Vdbench performance testing. Each figure includes the total IOPS and Vdbench response time for one-, two-, and three-member array storage pools as the total number of iSCSI connections to the storage pool increased from 510 to 750 to 1020. Each figure shows a different workload type.

Test 1: 8K, random I/O, 67% read workload configurations at 510, 750, and 1020 iSCSI connections during an 8K, random I/O, 67% read workload.

Figure 1 shows the IOPS and response time as a percentage of the baseline value measured at 510 iSCSI connections by Vdbench for one-, two- and three-member array storage pools.

A negligible decrease in total IOPS and increase in response time of less than 5 percent is observed at 1020 iSCSI connections, but overall performance remains consistent as the number of iSCSI connections increase.

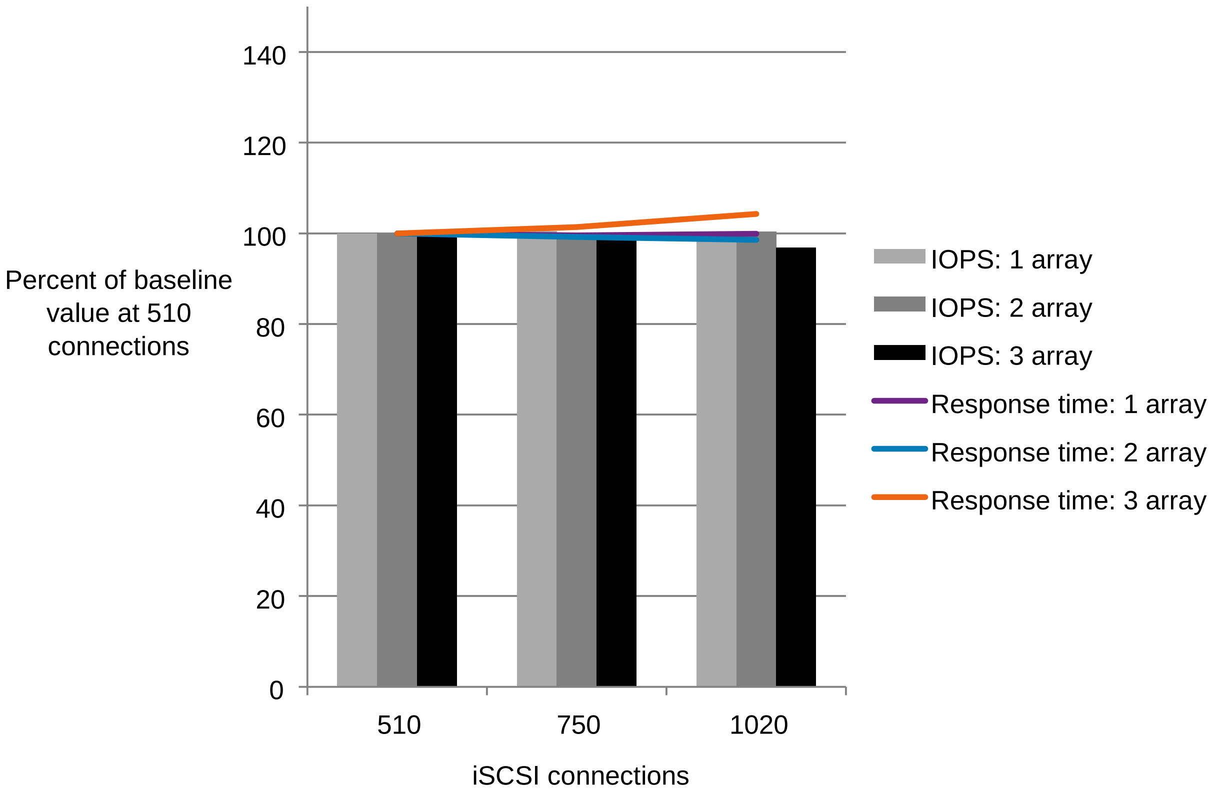

Figure 1 8K, random I/O, 67% read workload

Test 2: 64K, sequential I/O, write workload configurations at 510, 750, and 1020 iSCSI connections during a 64K, sequential I/O, write workload.

Overall performance remains consistent as the number of iSCSI connections increase.

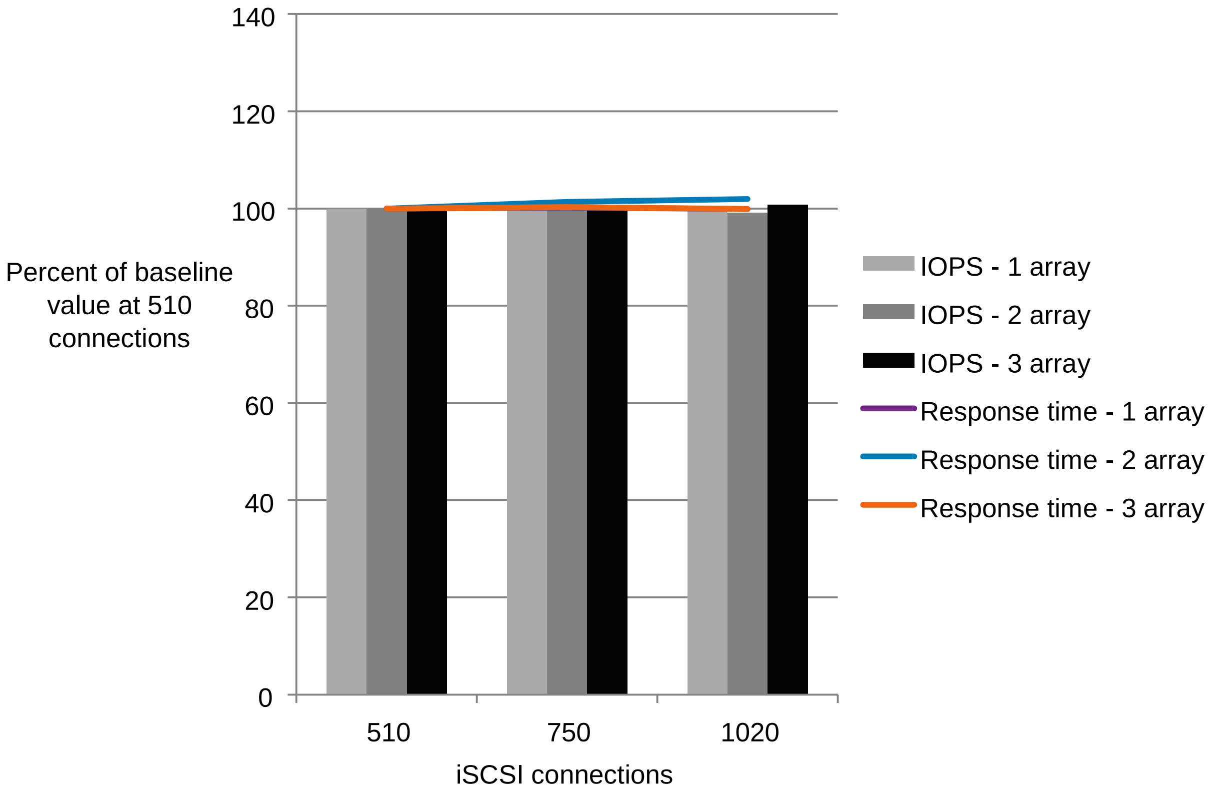

Figure 2 64K, sequential I/O, write workload

Test 3: 256K, sequential I/O, read workload configurations at 510, 750, and 1020 iSCSI connections during a 256K, sequential I/O, read workload.

Overall performance remains consistent as the number of iSCSI connections increase.

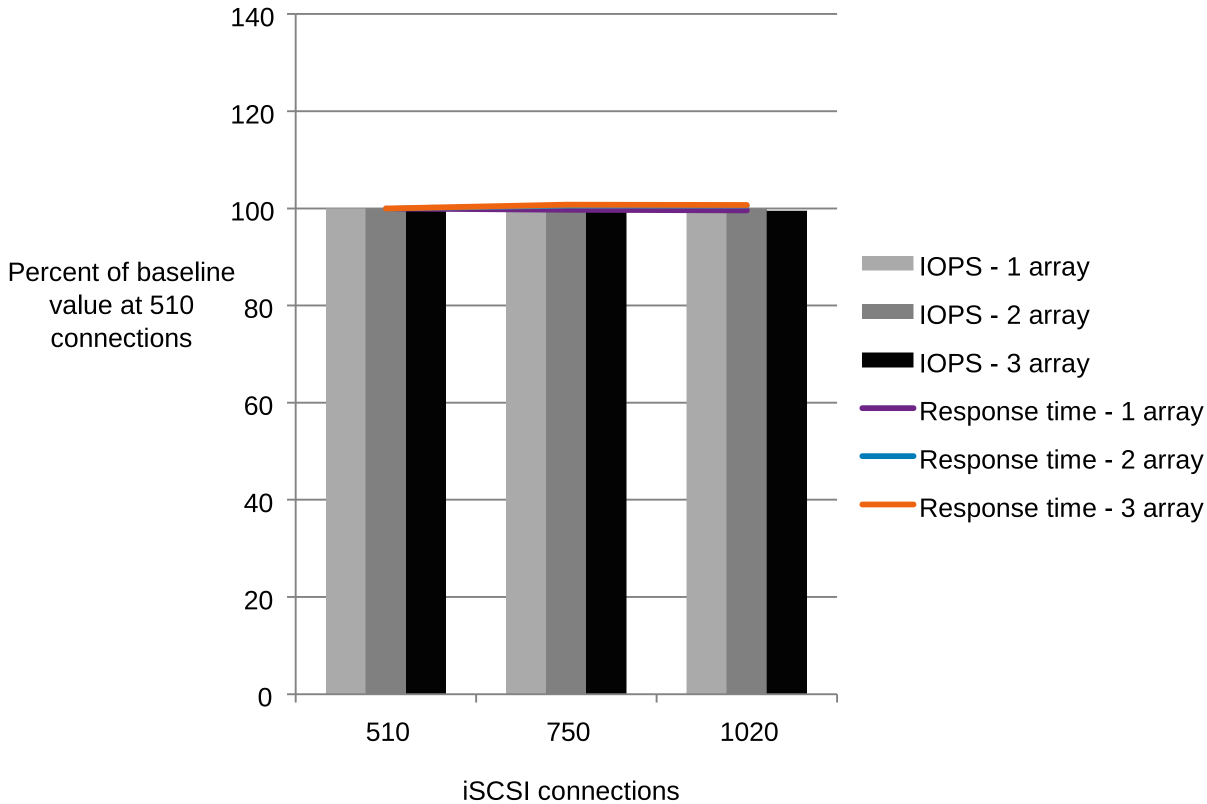

Figure 3 256K, sequential I/O, read workload

4 Availability testing

This section describes the test environment used to determine if SAN availability was affected when operating near the maximum number of iSCSI connections to the storage pool.

4.1 Test environment

In order to determine whether the high availability of the SAN is affected as the total number of storage pool iSCSI connections increases towards the officially supported maximum, a network failure was introduced to the SAN configuration while storage targets sustained a Vdbench workload.

A single physical vSphere ESXi server with five Windows Server virtual machines ran a moderate Vdbench workload on 17 volumes residing on one- and three-member array pools. Both host and guests connected to the SAN using the PS Series MPIO module and software initiator for each OS.

The vSphere ESXi host server had four 1 GbE network interfaces associated with four vmkernel iSCSI ports connected to a virtual switch dedicated to the host.

The five virtual machines each had two vmxnet3 1 GbE network interfaces connected to a virtual switch dedicated to the guest machines. Four 1 GbE network interfaces were dedicated to the guest virtual switch.

Four of the five Windows Server virtual machines connected to the target volumes through the SAN network using iSCSI and the PS Series MPIO module.

One of the five Windows Server virtual machines used raw device mappings from the vSphere ESXi host to connect to the target volumes. This was done to test the vSphere ESXi host software initiator and the PS Series MPIO module for vSphere ESXi, MEM.

The workload was run on 17 volumes consistently while each pool size was accommodating total connections of 510 and 1020. The exact workload was 8 KB transfer size, random I/O, 67-percent read workload. The total workload against each pool size was set to an arbitrary 5,000 IOPS (the Vdbench iorate parameter was set to 1,000 for each of the five virtual machines initiating the workload).

The exact connection totals were reached by creating additional volumes and connecting to them using the vSphere ESXi storage initiator and Windows Server virtual machines storage initiators.

While the SAN pool was under Vdbench load, we introduced a network failure by powering down one of the redundant network switches which comprised the SAN network.

This simulated switch failure event was repeated four times for each combination of iSCSI connection total and storage pool member array quantity. In all cases, no disk timeouts or read/write errors were reported by Vdbench while it maintained its disk I/O during and after the switch failure event.

5 Recommended practices

The following section explains how to calculate the total number of iSCSI connections to expect from a given PS Series SAN configuration along with options for reducing the total number of connections.

5.1 Calculating total iSCSI connections

By default, the PS Series MPIO module creates two connections per volume slice to a storage volume, up to a maximum of six connections. To calculate the number of iSCSI connections that would be made to a particular volume the following formula can be used:

Multiply the number of member arrays in the storage pool where the volume resides (up to three member arrays because a volume will not be spanned across more than three arrays under normal circumstances) by the number of hosts connecting to the volume by the two connections the MPIO module will make to each volume slice.

iSCSI connections per volume = (Member array quantity (up to 3) x host quantity x 2 connections)

For example, a volume spanned across two member arrays in a storage pool and being accessed by five hosts results in 20 iSCSI connections (2 member arrays x 5 hosts x 2 connections).

In another example, a storage pool with five member arrays and two hosts connecting to a single volume will result in 12 iSCSI connections (the 3 member arrays over which the volume spans x 2 hosts x 2 connections).

If all hosts are connecting to all storage volumes in a storage pool, then we can calculate the total number of iSCSI connections by multiplying the per-volume amount calculated above by the number of volumes.

Total iSCSI connections = (Member array quantity (up to 3) x host quantity x volume quantity x 2 connections)

If all hosts are not connecting to every volume, the total number of iSCSI connections will be the sum of all per-volume amounts calculated above for each volume.

In the case of virtualized environments, the vSphere ESXi server should be counted as a single host. Virtual machines that reside on the ESXi server do not make additional iSCSI connections to the storage target if they are accessing the volume as a VMDK virtual disk or using a raw device mapping. In those cases, the virtual machines rely on the existing iSCSI connections established by the ESXi host initiator. However, if the virtual machines are accessing the storage target directly over the SAN network using the software initiator inside the guest OS, then each virtual machine should be counted as an additional host for the purposes of calculating iSCSI connections using the formula above.

The formulas above do not account for the short-lived iSCSI connections that can result from volume migration and replication hosts. See section 2.3.3 for more information about iSCSI connections created by volume migration and replication.

5.2 Tips for reducing iSCSI connections

There are several ways to reduce the number of overall iSCSI connections required for a given SAN installation.

The simplest action, which does not require any change to the underlying SAN design or administrative strategy, is to change the **Max Sessions per Volume Slice** setting of the PS Series MPIO module from two to one on each storage initiator, effectively cutting in half the number of iSCSI connections that will be required for a given SAN configuration. By having the MPIO module create only one connection per volume slice, the formulas from section 5.1 now change to the following:

iSCSI connections per volume = (Member array quantity (up to 3) x host quantity)

Total iSCSI connections = (Member array quantity (up to 3) x host quantity x volume quantity)

Because the MPIO module monitors the network interface link status and reopens an iSCSI connection in the event of network interface failure, SAN availability is not affected by this setting change. From a performance perspective, this change will have minimal impact unless I/O patterns are highly sequential with a larger than average block size. It is recommended that the administrator evaluate any performance differences observed before committing this change on all hosts.

Another option to consider is creating fewer volumes of a larger size. PS Series member arrays support volume sizes up to 15 TB. Currently, Windows Server supports disk partitions up to 16 EB (maximum tested size is 16 TB) and since the release of VMware vSphere 5, ESXi now supports datastores up to 64 TB. Note that larger volume size can increase the time required for backup and recovery procedures.

If every host does not require access to every available volume, you could also consider splitting the hosts into smaller administrative clusters based on volume access requirements. Using the access control list (ACL) of the volume, volume access can be granted to only those hosts which require it. Each PS Series volume ACL has a maximum of 16 entries, and each entry can limit access using CHAP accounts, IQN strings, or IP addresses.

The last common option is to use up to four storage pools. Each new storage pool requires at least one member array dedicated to that pool. A storage pool brings with it support for 1,024 more iSCSI connections (up to the storage group maximum of 4,096). And since a host discovers volumes at the storage group level, having multiple storage pools does not affect volume discovery and access. One thing to note is that volumes cannot be load-balanced across storage pools, so to take advantage of PS Series multimember load balancing for a specific volume, there must be more than one member array in the pool where the volume resides.

A Hardware diagram

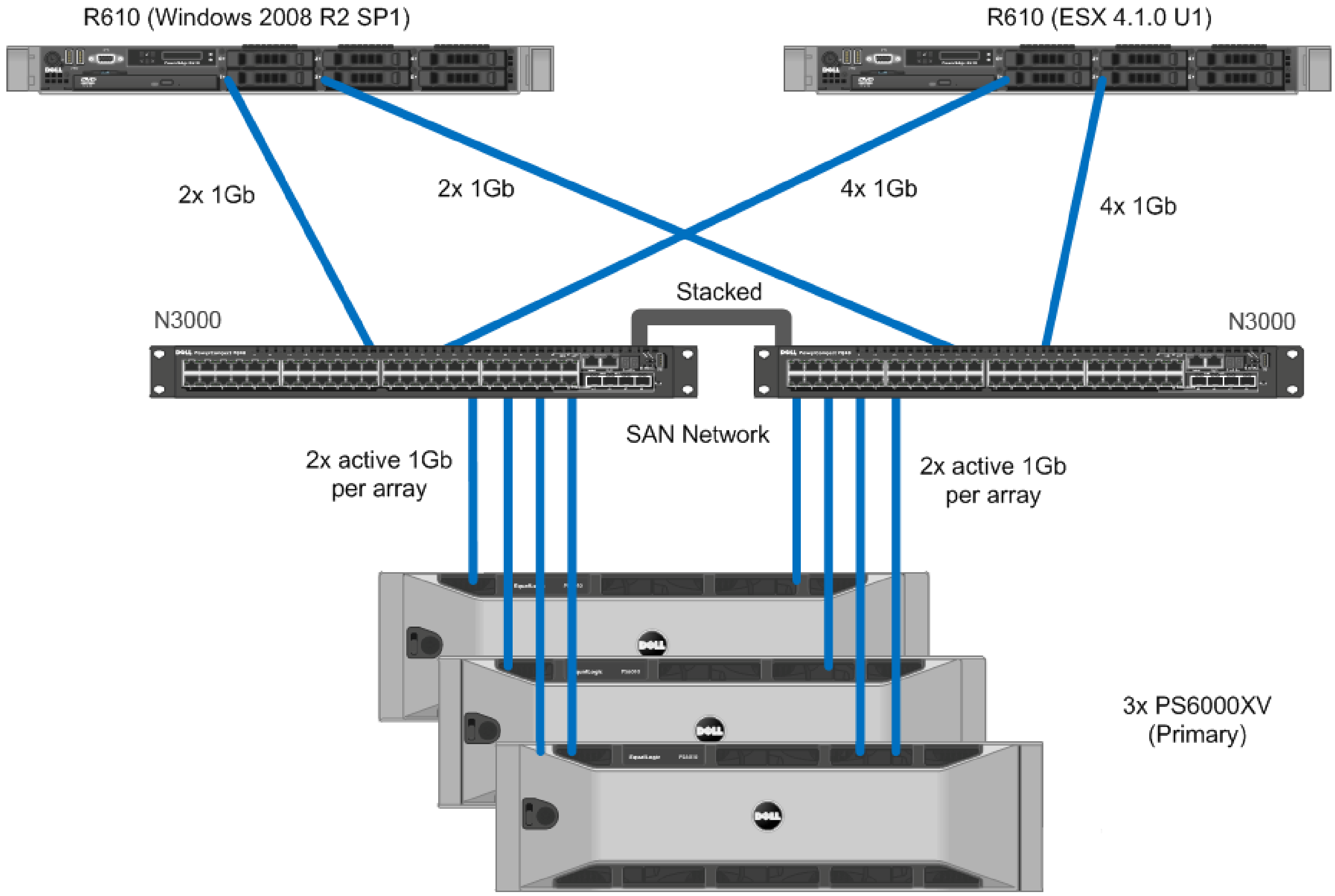

Figure 4 A hardware diagram of the servers, switches, and storage used in the test environment

B Logical diagram

The following diagram shows the allocation of the networking resources in the test environment.

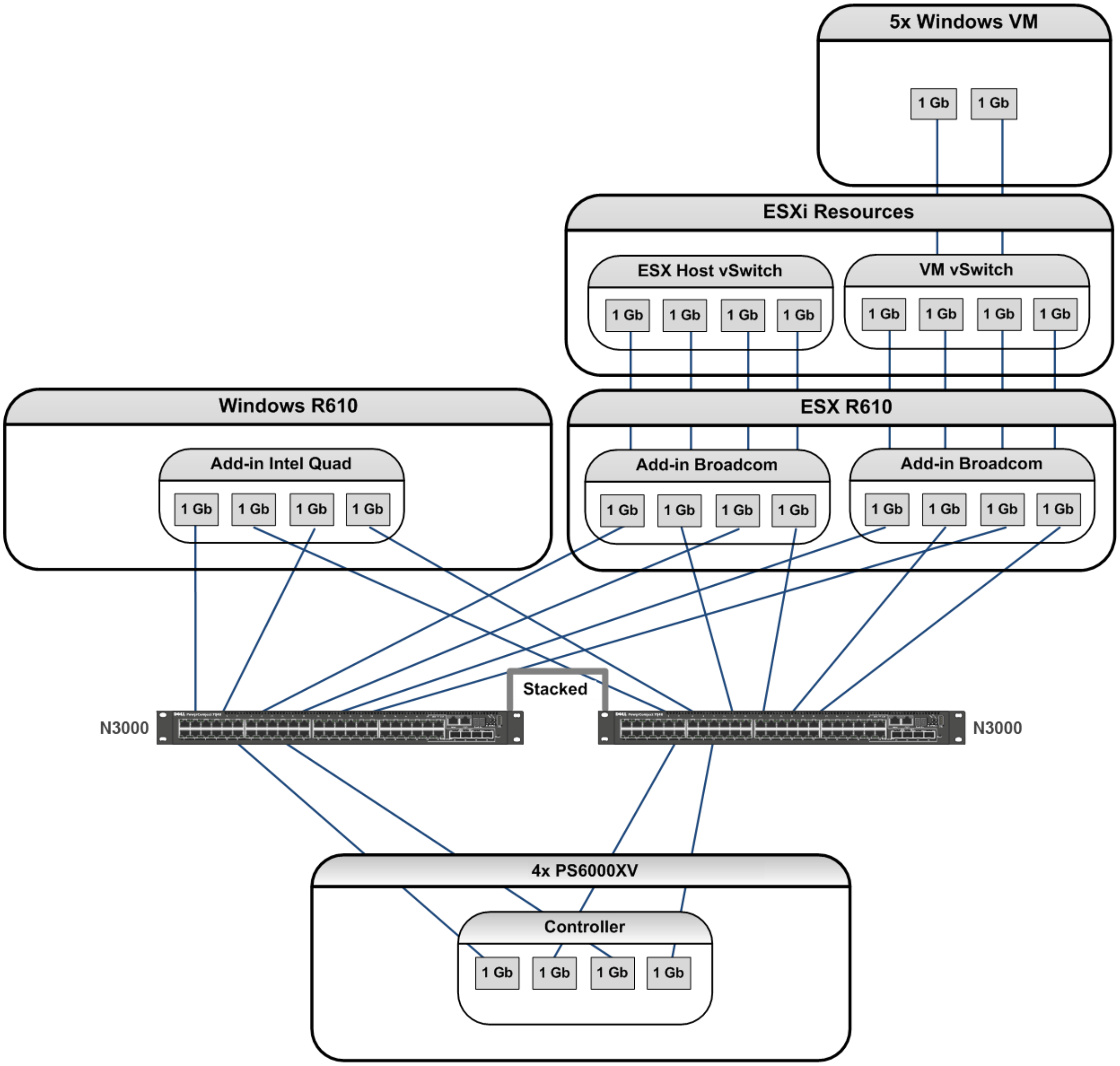

Figure 5 A logical diagram of the allocation of networking resources in the test environment

C Solution infrastructure

Table 4 provides a detailed inventory of the hardware and software configuration in the test environment.

Table 4 Configuration for test environment

Hardware components		Description
Windows Server 2008 R2 server	Dell EMC PowerEdge R610 Server: • Windows Server 2008 R2 SP1 • BIOS version: 3.0.0 • 2 x Intel® Xeon® Processor X5650 • 32GB RAM • PERC 6/i – v1.22.12-0952 • 2 x Fujitsu 300GB SAS – vD809 • 1 x Intel VT Quad 1GbE NIC – driver v11.4.7.0 • EqualLogic Host Integration Toolkit v3.5.1	Storage host
ESXi 4.1 server	Dell EMC PowerEdge R610 Server: • ESXi 4.1 Update 1 • BIOS version: 3.0.0 • 2 x Intel Xeon Processor X5650 • 48GB RAM • PERC 6/i – v1.22.12-0952 • 2 x Fujitsu 300GB SAS – vD809 • 1 x Broadcom 5709 Quad 1GbE NIC – driver v2.1.6b.v41.4 • EqualLogic Multipathing Extension Module v1.0.1	Storage Host
5x ESXi 4.1 Server Virtual Machines	Windows Server 2008 R2 SP1: • 2 x virtual CPU • 4GB virtual RAM • 2 x vmxnet3 virtual NIC	
SAN Network	• 2 x Dell EMC Networking N3000 1 Gb Ethernet Switch – v4.1.0.19 • 2 x 10GbE stacking cables	
Primary Storage	3 x PS6000XV: • 16 x 300GB 15K SAS disks – vEN03 • 2 x quad-port 1GbE controllers • Firmware: 5.1.1 R189834 H2	No dedicated management interface
Replication Storage	PS6000XV: • 16 x 300GB 15K SAS disks – vXRS0 • 2 x quad-port 1GbE controllers • Firmware: 5.1.1 R189834 H2	No dedicated management interface

D Vdbench parameters

Vdbench workloads were executed using the following parameters in the parameter file, where *N* is the number of iSCSI volumes under load.

Common parameters:

```
hd=default
hd=one,system=localhost
```

iSCSI volumes:

```
sd=sd1,host=*,lun=\\.\PhysicalDrive1,size=10240m
sd=sd2,host=*,lun=\\.\PhysicalDrive2,size=10240m
... sd=sd3,host=*,lun=\\.\PhysicalDriveN,size=10240m
```

8K, 67% read, random I/O workload:

```
wd=wd1,sd=(sd1-sdN),xfersize=8k,rdpct=67
```

64K, 100% write, sequential I/O workload:

```
wd=wd1,sd=(sd1-sdN),xfersize=65536,rdpct=0,seekpct=sequential
```

256K, 100% read, sequential I/O workload:

```
wd=wd1,sd=(sd1-sdN),xfersize=262144,rdpct=100,seekpct=sequential
```

Runtime options:

```
rd=rd1,wd=wd1,iorate=max,elapsed=86400,interval=1
```

E Additional resources

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

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

Storage Solutions Technical Documents on Dell TechCenter provide expertise that helps to ensure customer success on Dell EMC storage platforms.

The following Dell EMC publications are referenced in this document or are recommended sources for additional information.

- *Dell EqualLogic PS Arrays – Scalability and Growth in Virtual Environments*
- *Dell PS Series Configuration Guide*