Scalability & Deployment Best Practices for the Dell EqualLogic FS7500 NAS System as a File Share Solution

A Dell Technical Whitepaper

Storage Infrastructure and Solutions Engineering
Dell Product Group
February 2012

Table of Contents

Acknowledgements

Feedback

We encourage readers of this publication to provide feedback on the quality and usefulness of this information. You can submit feedback as follows:

➢ Use the "Start a new thread" link here:

http://www.delltechcenter.com/page/EqualLogic+FS7500+NAS+Scalability+And+Deployment

Revisions

Date	Status
October 2010	Initial publication
February 2012	Dell Scalable File System changed to Dell Fluid File System

1 Introduction

The Dell™ EqualLogic™ FS7500[1] adds Network Attached Storage (NAS) capabilities to the EqualLogic PS Series[2] product line. Working with PS Series arrays, the FS7500 provides a high performance, scale-out, unified file and block storage solution for small to mid-size environments. It enables easy configuration and management of iSCSI, CIFS, and NFS storage from a single interface – solving the challenges and time consuming inefficiencies IT administrators face when managing NAS and SAN systems through separate management interfaces. The solution is based on the Dell Fluid File system (FluidFS). It enables organizations to scale both capacity and performance incrementally as their unified storage needs grow over time.

The EqualLogic FS7500 NAS service can be quickly and efficiently added to both existing and new PS Series arrays. The FS7500 allows a single file system to use the entire capacity of an EqualLogic pool and the ability to add another FS7500 system into the same namespace to improve file performance.

The performance of an FS7500 system is dependent upon the customer's network and storage infrastructure. It is also dependent upon operational factors, including:

- Number of PS Series arrays and FS7500 systems deployed
- File system protocols in use: NFS, CIFS
- Expected I/O load and the characteristics of the file data being stored

The objective of this paper is to provide:

- Best practices for configuring the supporting network and storage infrastructure when deploying the Dell EqualLogic FS7500 Unified Storage Solution.
- Performance and scalability metrics of the system as a file sharing solution in different operational scenarios.
- Guidelines for system selection based on specific operational requirements

To obtain the results presented in this paper, Dell Labs conducted a series of I/O performance tests to measure how system performance scaled when using various configurations, block sizes, and file system protocols (CIFS vs. NFS). Providing a file serving solution for **end users** was the use case we focused on in our tests. Examples of end user file services include user home shares and departmental file shares. End-user file share access using CIFS and NFS protocols are the two predominant types. The best practice guidelines presented in this paper are based in part on the results of the testing we performed.

1.1 Audience

This paper is intended for Solution Architects, Storage Network Engineers, System Administrators, and IT Managers who need to understand how to design, deploy, and scale the Dell EqualLogic FS7500 NAS solution. We expect the reader to have working knowledge of EqualLogic PS Series SANs and related iSCSI SAN operations.

---

[1] The Dell EqualLogic FS7500 Unified Storage Solution:
http://www.equallogic.com/products/default.aspx?id=10465
[2] EqualLogic PS Series Architecture: http://www.equallogic.com/products/default.aspx?id=5823

1.2 The rest of this paper

The rest of this paper contains the following sections:

2 Overview of the Dell EqualLogic FS7500 Unified Storage Solution

The FS7500 is a high performance unified storage solution that integrates with new and existing EqualLogic PS Series storage arrays and enables customers to easily configure and manage iSCSI, CIFS, and NFS access in a single flexible storage pool using the EqualLogic Group Manager. The unique scale-out architecture of the solution lets customers expand storage capacity and/or system performance as needs change over time. In this section, we outline the benefits of the solution, as well as provide a technical overview of the FS7500 architecture and components.

2.1 Features of the EqualLogic FS7500 Unified Storage Solution

The FS7500 provides many powerful features.

Scale-out unified storage, optimized for the midrange

Unlike many unified storage solutions that only scale in capacity, the EqualLogic FS7500 can scale both capacity and performance. To increase capacity, you add more EqualLogic arrays; to increase performance, you add more FS7500 systems (two controllers and a back-up power supply). As customers scale their FS7500 solution, they can actually increase the overall performance of their storage environment. As customer needs grow and change, they can dynamically modify their block and file storage capacity without disrupting existing applications and storage systems. A FS7500 NAS system can be configured and added to EqualLogic arrays quickly and efficiently.

Expand traditional file share limits

The EqualLogic FS7500 incorporates the Dell Fluid File system[3] (FluidFS). FluidFS is designed to optimize file access performance and hardware utilization while eliminating capacity constraints. A key component in Dell's Fluid Data architecture, FluidFS is a high-performance scale-out file system that presents a single file-system namespace through a virtual IP address, regardless of cluster size. Most NAS solutions have strict limits on the size of a file share. The Dell Fluid File system has no such limitations. With the FS7500, customers can even expand a single file system up to the capacity[4] of an EqualLogic PS Series storage pool.

Integrated management

NAS deployment and management functionality is fully integrated into the EqualLogic Group Manager user interface. The Group Manager facilitates all normal administrative, monitoring and maintenance operations. The Group Manager also automates the initial setup and configuration of your FS7500 system, thus simplifying the process of creating volumes and shares.

Advanced features

The EqualLogic FS7500 includes a file-level snapshot capability (This feature is separate from PS Series array based volume snapshot features). End-users can restore previous versions of files from a directory file snapshot without having to contact system administrators, thus freeing up administrators for other tasks. To help improve storage utilization, the EqualLogic FS7500 features a virtualization layer that lets customers expand the NAS service dynamically. The FS7500 is easily integrated into a

---

[3] For more details about the Dell Fluid File system:
http://www.dell.com/downloads/global/products/pvaul/en/dell-scalable-file-system-white-paper.pdf
[4] Maximum of 509TB usable storage

customer's backup environment with support for Network Data Management Protocol[5] (NDMP) backup. As with all Dell EqualLogic products, the FS7500's NAS feature set, software licensing and future firmware enhancements are included in the base price of the system.

Highly available file serving

A dual active/active controller architecture and large onboard battery protected cache give the EqualLogic FS7500 outstanding performance. Each FS7500 dual controller system includes 48GB of memory for hosting the system cache. The dual active/active controllers automatically balance the load across all controllers. The EqualLogic FS7500 supports all new and existing EqualLogic PS Series array models. The FS7500 allows a single file system using the entire capacity of an EqualLogic pool with the ability to add another FS7500 system into the same namespace to improve file performance.

2.2   Architecture of the EqualLogic FS7500 system

This section provides an overview of the EqualLogic FS7500 architectural components and illustrates the relationships between them.

NAS service

The NAS service is the central component of the unified storage solution. The NAS service provides support for NAS file systems in a PS Series group. UNIX/Linux and Windows clients view a NAS service as one virtual file server that hosts multiple CIFS shares and NFS exports. A NAS service requires at least one EqualLogic FS7500 system in addition to its attached EqualLogic PS Series storage. One EqualLogic FS7500 system includes the following hardware components:

- Two FS7500 Controller units with pre-installed file serving software
- One FS7500 Backup Power Supply (BPS) System unit

Each controller contains memory for the mirrored cache which is protected by a backup power supply. The initial version of the solution scales up to a maximum of two FS7500 systems, each with 2 controllers, with up to a maximum of eight EqualLogic PS Series arrays in the storage pool. The PS Group Manager GUI can be used to discover controllers and configure and manage the NAS service. The NAS service includes the NAS nodes (controllers), network configuration for the service, and the amount of EqualLogic storage pool space for the NAS reserve.

NAS nodes

To ensure maximum performance and availability, the NAS controllers must be added to the NAS service in pairs. When added to a NAS service, a controller appears as a single NAS node. When two NAS nodes are added as pairs in the system they become peer nodes to each other. Peer NAS nodes operate simultaneously. Each node has a high-performance cache that is mirrored to its peer node. In case one node fails, the peer node automatically continues operation with no impact on availability. Deploying a pair of peer nodes increases NAS service performance and improves availability.

NAS reserve

The NAS reserve space is allocated from the space available in the attached EqualLogic PS Series storage pool. The NAS reserve has the following properties:

---

[5] http://www.ndmp.org/info/overview.shtml

- It uses a single EqualLogic storage pool. Only one storage pool in the EqualLogic storage group can be used for allocating the NAS reserve.
- It is formatted with Dell Fluid File system (FluidFS).
- It also stores NAS service metadata.
- The NAS reserve can be expanded later when more storage pool space is available.
- The NAS reserve is comprised of a variable number of automatically created EqualLogic storage volumes. The number of storage volumes created within the pool depends on the reserve size.

NAS file system

Multiple virtual NAS file systems can be created in a NAS service. In the NAS File System:

- Both CIFS shares and NFS exports can be created.
- Clients connect to shares or exports via NAS protocols (CIFS and NFS) using a single service IP address or service name.
- Clients can only access the shares and exports after presenting correct credentials.
- Client connections are load balanced across the available NAS nodes for availability and performance.

The relationship between EqualLogic PS Series storage pool and FS NAS Reserve entities is illustrated in Figure 1 below:
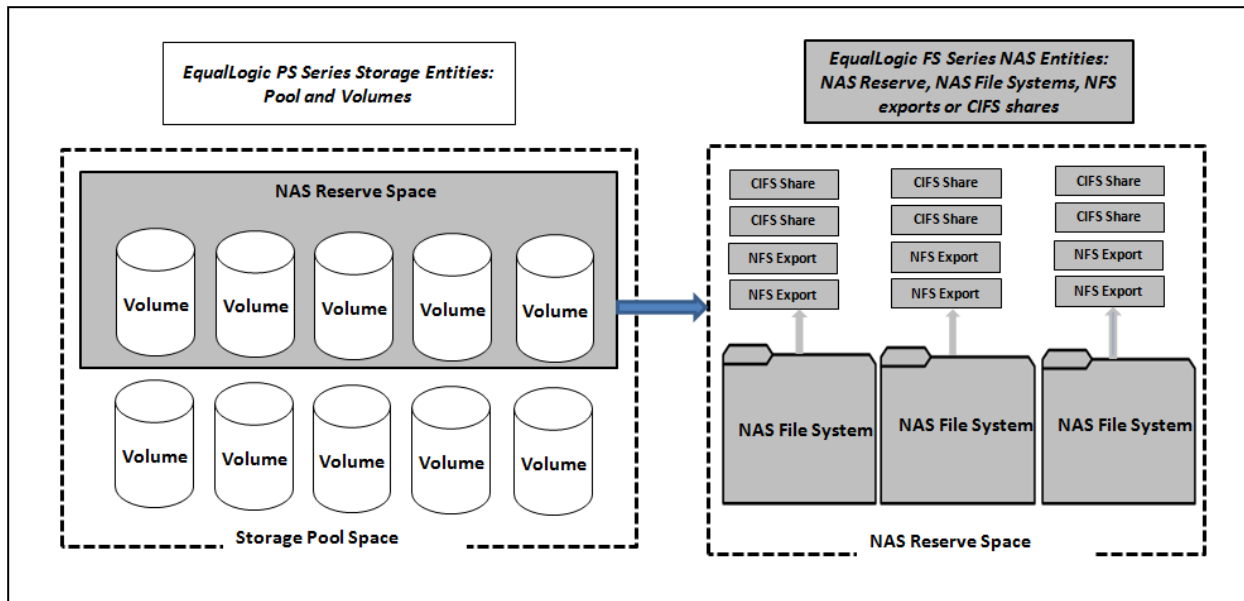

Figure 1: EqualLogic Storage Pool and NAS Reserve

> **Note:** Procedures for EqualLogic NAS configuration and management are provided in the PS online help: http://psonlinehelp.equallogic.com/V5.1/groupmanager.htm

3 Test methodology

The test methodology used to derive the results of this paper was comprised of two parts:

1. The creation of NFS exports or CIFS file shares on the Dell EqualLogic FS7500 Unified Storage Solution
2. Simulating NFS or CIFS file operations respectively from test clients to monitor and record system response times

This section outlines the storage pool configuration, the data set properties, the distribution of NFS and CIFS file system operations, as well as the performance criteria and simulation methodology used in the test studies.

3.1 Storage pool and NAS reserve configuration

For all tests, an EqualLogic storage pool was created with the required number of PS Series arrays being tested. The entire storage pool space was dedicated to the NAS reserve. None of it was used for block I/O access for other applications. The required number of virtual NAS file systems for the test was created. Each NAS file system was configured with only one NFS export, or one CIFS based share, based on which test was being conducted. See Figure 2.
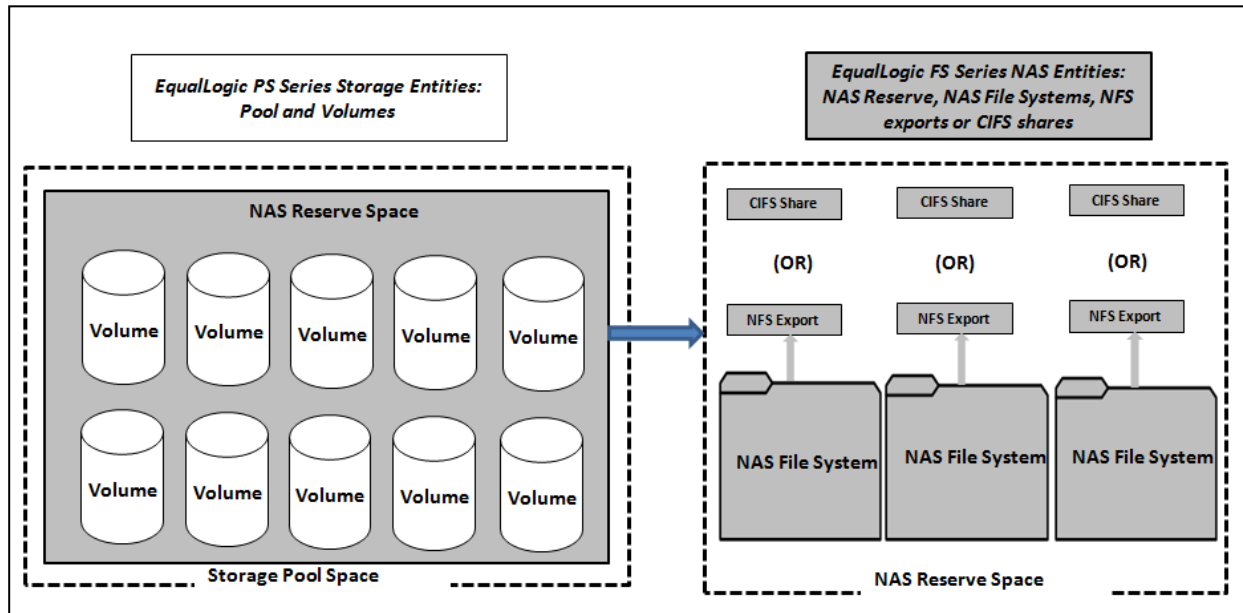
Figure 2: Test configuration - EqualLogic Storage Pool and NAS Reserve

3.2 Data set properties

File data was created to simulate the file operation load. The size of the data set was proportional to the number of operations being simulated, with about 30% of the data set accessed during simulation. The structure of the data set included directories, files, and symbolic links. The file size distribution for the simulation data set is given below. The file size distribution percentage was constant across all simulations.

Table 1 File Size distribution

Size	Percent
Less than 4KB	50%
4KB - 32KB	22%
32KB - 256KB	28%

> **Note:** The design of the data set as well as the distribution of NFS and CIFS file operations during each test was loosely based on the principles of the SPECsfs2008 file server benchmark. The test results presented in this paper should not be considered SPECsfs2008 benchmarks.

3.3   NFS simulation

Several test client virtual machines were used to load the FS7500 Unified Storage Solution with NFS file I/O operations. The NFS client-side file I/O simulation software was custom. The distribution of file system operations for the NFS simulation is given below:

Table 2 NFS Simulation: File System Operation Distribution

Operation Type	Percent
READ	18%
WRITE	10%
OTHER[6]	72%
Distribution: READ Only	
Less than 4KB	25%
4KB - 32KB	65%
32KB - 256KB	10%
Distribution: WRITE Only	
Less than 4KB	40%
4KB - 32KB	50%
32KB - 256KB	10%

3.4   CIFS simulation

Several test client virtual machines were used to load the FS7500 Unified Storage Solution with CIFS operations. The CIFS client-side file I/O simulation software was custom. The distribution of file system operations for the CIFS simulation is given below.

---

[6] The OTHER category above includes primarily NFS meta-data operations such as LOOKUP, GETATTR, READLINK, READDIR, CREATE, REMOVE, FSSTAT, SETATTR, READDIRPLUS, and ACCESS.

Table 3 CIFS Simulation: File System Operation Distribution

Operation Type	Percent
READ_ANDX	20%
WRITE_ANDX	9%
OTHER[7]	71%
Distribution: READ Only	
Less than 4KB	60%
4KB - 32KB	20%
32KB - 256KB	20%
Distribution: WRITE Only	
Less than 4KB	67%
4KB - 32KB	18%
32KB - 256KB	15%

3.5 Performance criteria

The use case considered for simulations in this paper is end-user access of file shares. The response time expectation for end-user file share interaction from a client-side application is typically in the 1-2 seconds range. At approximately 1 second response time, end-users start to notice a delay. But, a 1 second delay should not cause end-users to divert their attention from the application.[8] At 0.1 second (100ms) users typically notice no delay and perceive the application response to be instantaneous. For our file share simulations we chose the response time requirement for file operations to be under 100ms.[9] In fact, the actual measured response times during our simulations were much lower than this arbitrary limit.

Response time is measured from test clients for operations completed on the NAS file shares. Note that enterprise server applications using a NAS as backend storage will need much faster average I/O response times than 100ms.

3.6 Simulation methodology

The simulation workload was applied or requested from test clients on the NAS system in increasing order of file I/O operations per second. The achieved file operations per second was measured and reported. If the achieved operations per second were not close to the requested operations per second, then the system was deemed to be saturated at an earlier requested load point. Also, average response time of the operations was measured across test clients to make sure they stayed within the established criteria. The achieved operations per second close to the saturation level was considered the optimal operations per second achievable for the configuration.

---

[7] The OTHER category includes primarily NFS meta-data operations such as ECHO, LOCKING_ANDX, NT_TRANSACT_QUERY_SECURITY_DESC, TRANS2_QUERY_FS_INFORMATION, TRANS2_QUERY_FILE_INFORMATION, TRANS2_QUERY_PATH_INFORMATION, TRANS2_SET_FILE_INFORMATION, TRANS2_FIND_FIRST2, NT_CREATE_ANDX, CLOSE, and FLUSH.
[8] http://www.useit.com/papers/responsetime.html
[9] Miller, R. B. (1968). Response time in man-computer conversational transactions - www.mattson.com.au/robert/files/Miller1968.pdf

4 Test hardware configurations

This section outlines the major system hardware components involved in the simulations. These are illustrated in the figures below. Figure 3 illustrates one of the smallest configurations tested, which included two EqualLogic PS Series storage arrays and two EqualLogic FS7500 controllers. Figure 4 illustrates one of the largest configurations tested based, which included the maximum number of eight EqualLogic PS Series storage arrays in the storage pool and the maximum number of four EqualLogic FS7500 controllers. A varying number of configurations were tested in certain combinations of the following variables:

1. Number of EqualLogic FS7500 systems: 1 or 2 (2 or 4 controllers)
2. Type of EqualLogic PS Series storage arrays: PS6000XV, PS6500E
3. Number of EqualLogic PS Series storage arrays: 2, 4, or 8
4. RAID types for EqualLogic PS Series storage arrays: RAID 10, 50, 5, or 6

The network configuration remained the same for all test configurations.

Test simulations were conducted from test client machines that generated NFS and CIFS file I/O loads. The test client machines were configured as virtual machines using Windows® Server 2008 R2 Enterprise Edition with Hyper-V™ as the host hypervisor. Two Dell™ PowerEdge™ R815 servers ran Hyper-V™ for hosting the client virtual machines. Configuration details of the test client host and guest machines are given in the appendix section.
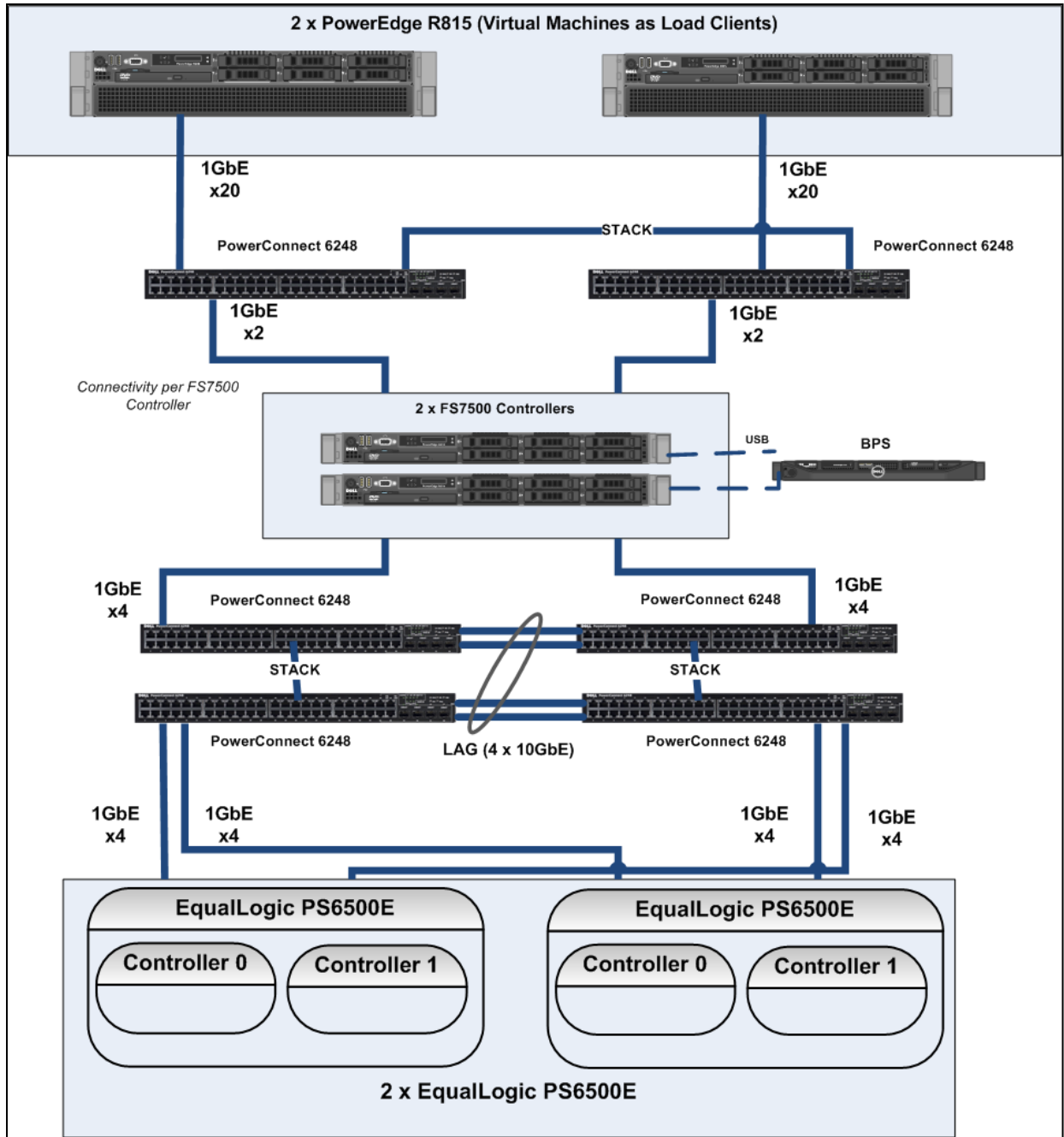
Figure 3: Test setup using two EqualLogic FS7500 Controllers and two EqualLogic PS6500E arrays
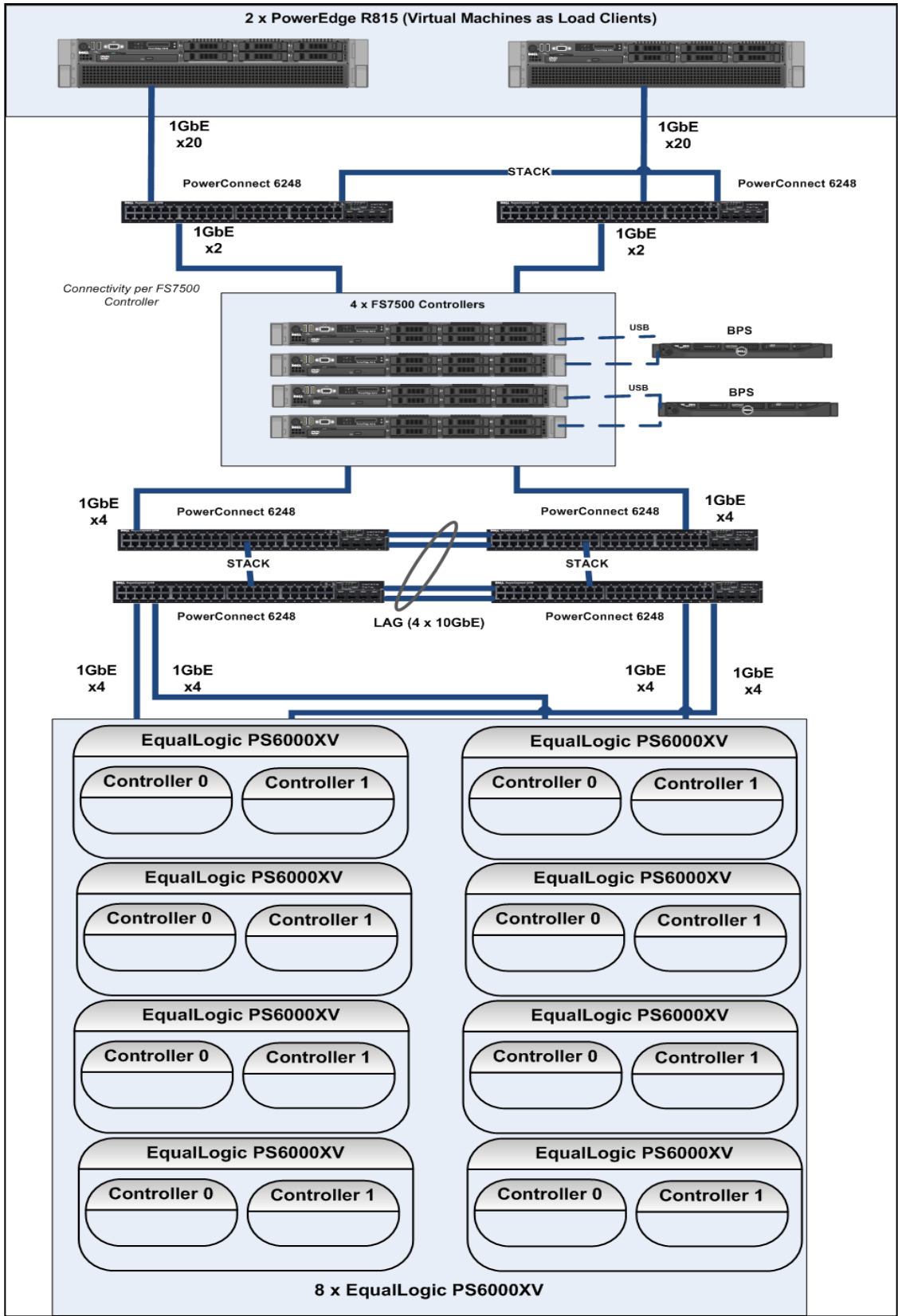
Figure 4: Test setup using four EqualLogic FS7500 Controllers and eight EqualLogic PS6000XV arrays

4.1 Network configuration

The connection paths from a single NAS node (FS7500 controller) to both the LAN and SAN are shown in Figure 5 below. The figure also details the connection paths for the EqualLogic PS Series storage array to the SAN.
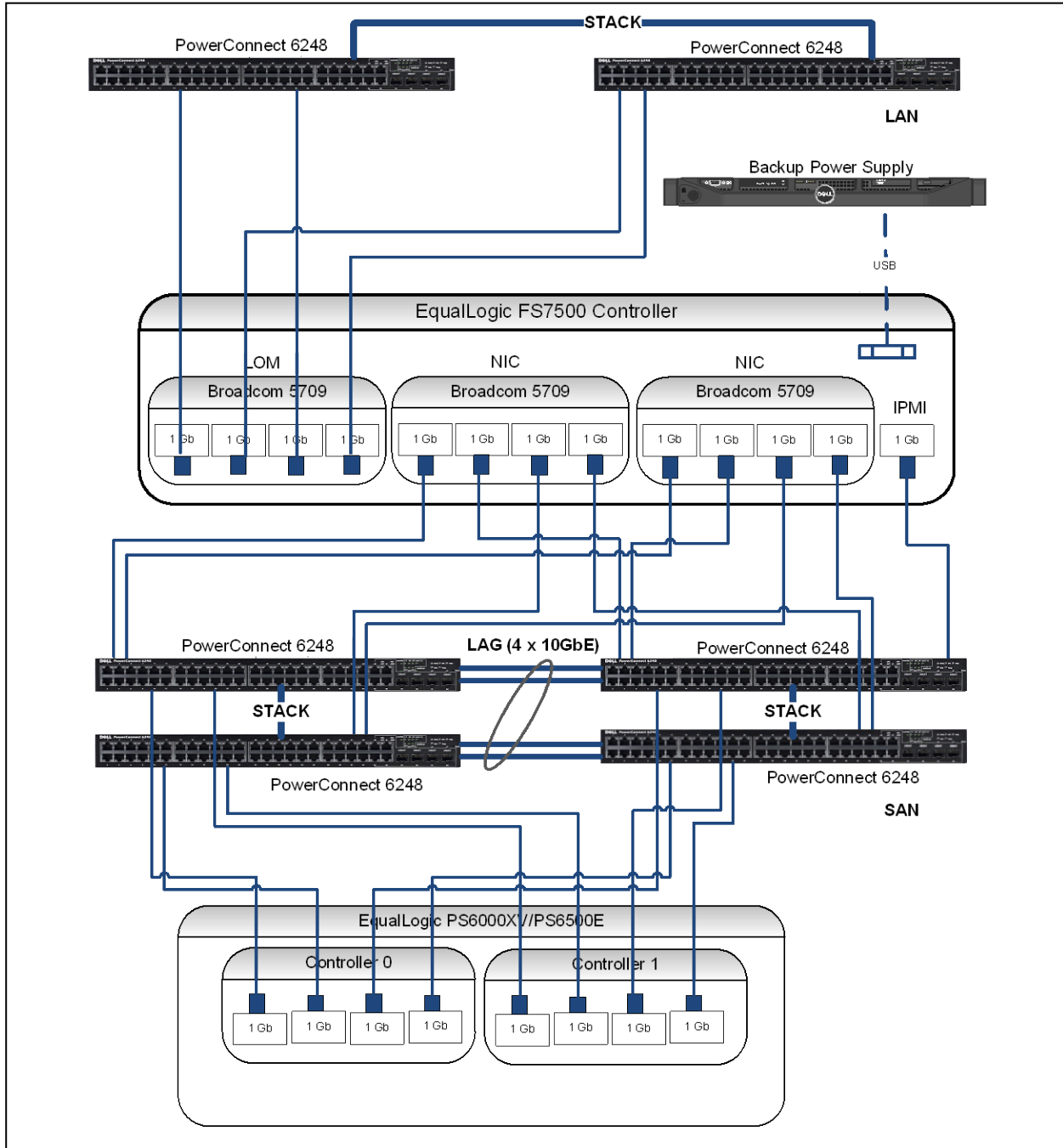
Figure 5: Network connectivity of FS7500 controller and PS Series storage array

4.1.1   NAS node NICs

Four NICs are dedicated on each NAS node for each of the networks – SAN, client and internal. The connectivity of each of these NICS is shown in Figure 5 above. Refer to [PS online help](#) for more information on the [FS7500 network configuration](#) details and procedures.

The FS7500 system LAN on motherboard (LOM) NICs are used for the client network and are shown connecting to the LAN switched stack in Figure 5. Client network NIC load balancing is set on each NAS node using a bonding mode. The default bonding mode is Adaptive Load Balancing (ALB), which requires no special switch settings and supports automatic, transparent network interface failover. The default bonding mode can be modified to use LACP (IEEE 802.3). Changing to LACP mode requires network switch configuration changes. In addition to bonding mode, more NAS service IP addresses can be added to the client network configuration for proper load balancing across multiple client subnets. The number of NAS service IP addresses that can be added ranges from one to four addresses for each NAS node, depending on the bonding mode for the service. For ALB, one to four IP addresses can be specified, multiplied by the number of NAS nodes. For LACP (IEEE 802.3), one IP address can be specified, multiplied by the number of NAS nodes. In our tests the default bonding mode of ALB was used on the NAS nodes for load balancing and one NAS service IP address was used. We did this because the client network was a non-routed single layer 2 network. Requests to the single NAS IP address were load balanced across the NAS nodes by a custom distribution mechanism among the NAS nodes and within a NAS node using ALB.

Each NAS node also required the use of a set of internal and SAN NICs. Two add-on quad port NICs on the FS7500 system are used for the SAN and Internal networks. They are both connected to the SAN. The internal network is used for communication between the NAS nodes. The SAN network is used for connecting to the EqualLogic PS Series storage arrays. Both the internal network and SAN network NICs on the NAS nodes should be attached to the same layer 2 Ethernet network and not be separated by separate switches or VLANs. Each node also includes an IPMI management NIC in addition to the SAN and internal network NICs. The IPMI NIC should also be connected to the same layer 2 network as the SAN/internal NICs, which would be the SAN as shown in figure above.

Further references to the term "LAN" in this document implies the client network. References to the term "SAN" implies the network that includes both the SAN and the internal network used for FS7500 inter-controller communication.

4.1.2   Network switches

The test configuration included 2 separate switching networks for the SAN and LAN. The FS7500 controllers connected to the EqualLogic PS Series storage arrays via the SAN. The FS7500 controllers or NAS nodes were also inter-connected via the SAN to support private internal network communications and IPMI management communications.

LAN

- Two Dell™ PowerConnect™ 6248 switches were configured for the LAN. The switches were inter-connected via the switch stacking mechanism. Each switch was configured with a stack module to support this inter-connection.
- The client network NICs of the FS7500 controllers were connected in a distributed fashion to these switches. The NICs from the PowerEdge 815 servers hosting test client VMs were also connected across these switches.

- The LAN was configured as single layer 2 network using the default VLAN. The default MAC based load-balancing mode (ALB) was used for distributing file system traffic across the FS7500 nodes.
- Flow control and Jumbo frames with MTU size = 9216 were enabled on the switch ports.

SAN

- Four Dell PowerConnect 6248 switches were configured for the SAN. Two switch pairs each were configured in a stack and the two stacks were inter-connected via a 4 x 10GbE Link Aggregation Group (LAG) using LACP. Each switch was installed with a stacking module to connect to its stacking peer. Each switch was also installed with an uplink module with 2 x 10GbE SFP+ modules to connect to a pair of peer switches on the other stack as shown in Figure 5 above.
- The LAG was chosen between the two switch stacks to support higher levels of availability. For instance, using this configuration you could complete firmware upgrades on one switch stack while the second stack continues providing storage connectivity.
- The SAN, internal network and IPMI NICs of the FS7500 controller were connected across the two switch stacks in a distributed fashion. The NICs from each EqualLogic PS Series storage array controller were also distributed across these two switch stacks.
- The SAN was configured as a single layer 2 network using the default VLAN. For proper FS7500 system operation you must have the SAN, internal network and IPMI NICs of the FS7500 controllers connected to the same layer 2 VLAN.
- Flow control and jumbo frames with MTU = 9216 was enabled on all switch ports including the LAG virtual port channels. The spanning tree Portfast setting was enabled on switch ports connecting to FS7500 controllers and EqualLogic PS Series controllers.

The IP and subnet information for the EqualLogic PS Series arrays and FS7500 networks is provided in the appendix.

5 Test simulations

A series of NFS and CIFS simulations were conducted to study the scalability and performance behavior of the NAS solution varying the number of NAS nodes, number of PS Series arrays, the PS Series array type and RAID type. This section describes the conduct of the tests and presents an analysis of the results obtained. These results provide scaling guidelines for designing FS7500 systems.

5.1 EqualLogic PS6000XV RAID 10 scalability

NFS and CIFS simulations were run on the FS7500 using EqualLogic PS6000XV arrays in a RAID10 configuration. The test data set and workload characteristics are described in section 3. The NFS and CIFS simulations were run on four different configurations, as shown below.

Table 4 PS6000XV RAID 10 Scalability Test Configurations

Configuration	Workload Type	File System Size (TB)	File Systems	Total File System Size (TB)	Total Data Set Size (TB)	Test Clients
1 x FS7500 + 2 x PS6000XV RAID10	NFS	1.75	4	7	1.76	16
1 x FS7500 + 4 x PS6000XV RAID10			8	14	2.34	16
2 x FS7500 + 4 x PS6000XV RAID10			8	14	3.40	32
2 x FS7500 + 8 x PS6000XV RAID10			16	28	4.00	32
1 x FS7500 + 2 x PS6000XV RAID10	CIFS	1.75	4	7	1.83	16
1 x FS7500 + 4 x PS6000XV RAID10			8	14	2.56	16
2 x FS7500 + 4 x PS6000XV RAID10			8	14	4.78	32
2 x FS7500 + 8 x PS6000XV RAID10			16	28	5.72	32

Figure 6 below illustrates the NFS achieved operations per second for the configurations in a relative scale. Total average client response time across all configurations remained under 10ms during all tests.
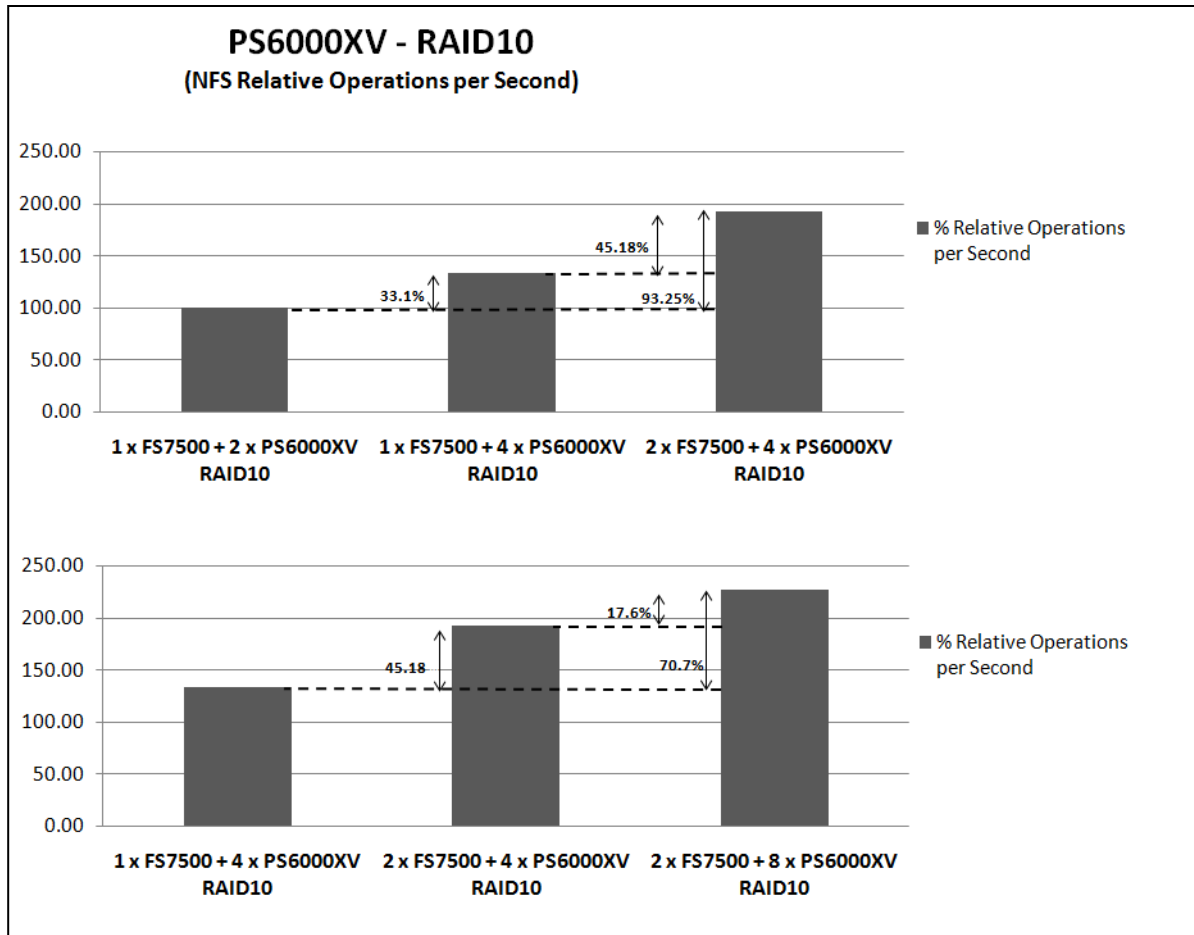
Figure 6: PS6000XV RAID10 NFS scalability

The results presented in Figure 6 show the following scaling characteristics:

- Using a single FS7500 system, four arrays yielded 33% more operations per second than two arrays. In this comparison we increased the number of PS Series arrays from two to four PS6000XV arrays while keeping the number of FS7500 controllers constant at two.
- Using four PS6000XV systems and two FS7500 systems we were able to support 45% more operations per second than when using a single FS7500 system. In this comparison the number of the PS Series arrays (four PS6000XV arrays) remained constant while the number of FS7500 controllers doubled from two to four.
- When using two FS7500 systems and four PS6000XV arrays we were able to achieve almost twice the performance (93% increase) than when using one FS7500 and two PS6000XV arrays. In this comparison we doubled the number of the PS Series arrays from two to four arrays and doubled the number of FS7500 controllers from two to four.
- When using two FS7500 systems and eight PS6000XV arrays we were able to achieve a 17% performance improvement than when using four PS6000XV arrays. In this comparison we increased the number of the PS Series arrays from four to eight PS6000XV arrays while keeping the number of FS7500 controllers constant.
- When using two FS7500 systems and eight PS6000XV arrays we were able to achieve a 70% performance improvement over using one FS7500 and four PS6000XV arrays. In this

comparison we doubled the number of the PS Series arrays from four to eight arrays and doubled the number of FS7500 controllers from two to four.

From these results we can conclude the following:

- Scaling the FS7500 systems along with the PS Series storage arrays can provide maximum system throughput and performance.
- Adding more FS7500 systems will produce more throughput for the same number of PS Series arrays. This is due to the additional memory, processing and network resources that are available as you increase the number of controllers.
- In certain deployments, more PS Series storage arrays may be needed with a single FS7500 system. This would be to enhance capacity even though I/O performance may be sufficient with fewer arrays.

The average client response time for the NFS READ operation alone (18% of the mix) was under 35ms and for the NFS WRITE operation (10% of mix) it was under 23ms across all configurations.

Figure 7 below illustrates the CIFS achieved operations per second for the configurations in a relative scale. Total average client response time across all file operations remained under 11ms during all tests.
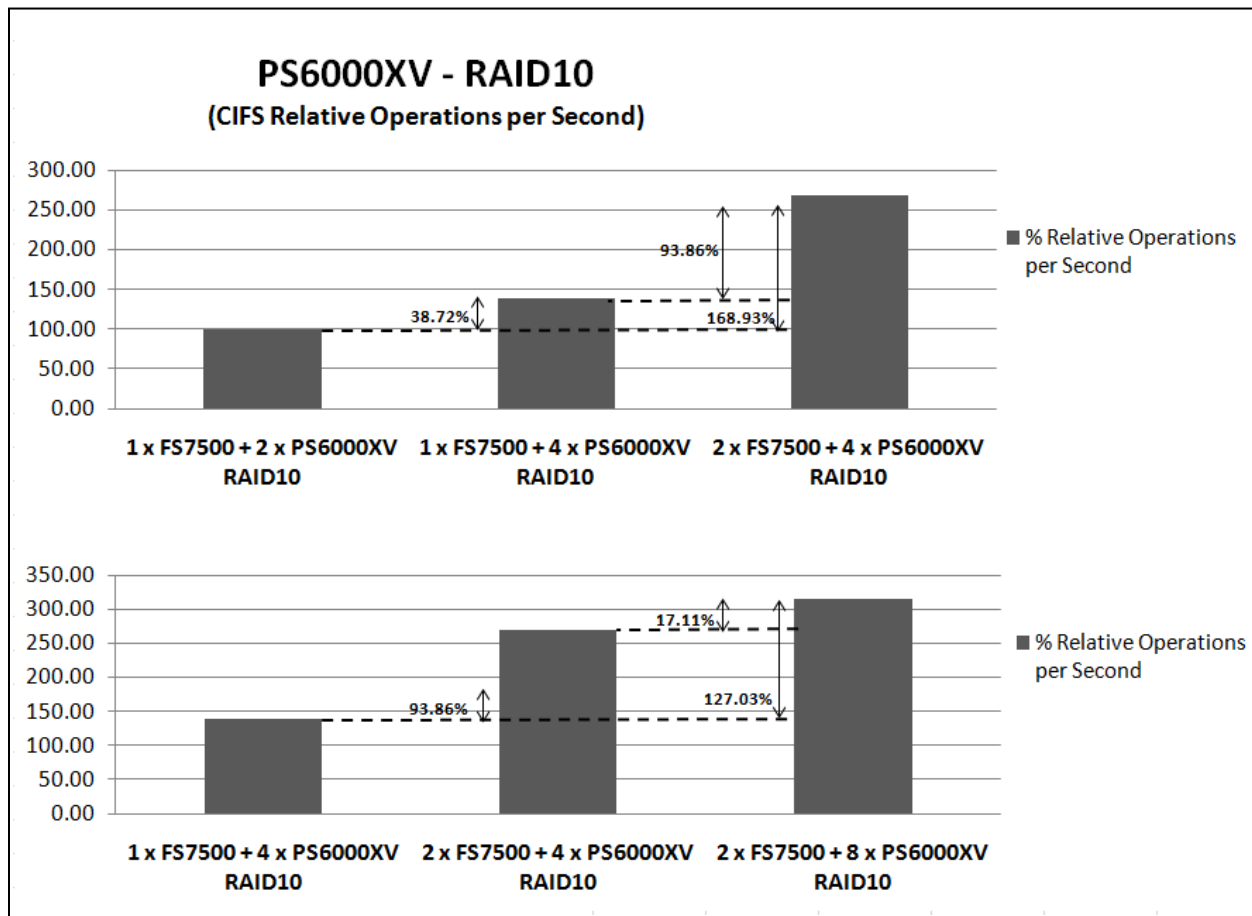

Figure 7: PS6000 RAID10 CIFS scalability

The results presented in Figure 7 show the following scaling characteristics:

- Using a single FS7500 system, four arrays yielded 38% more operations per second than two arrays. In this comparison we increased the number of the PS Series arrays from two to four PS6000XV arrays while keeping the number of FS7500 controllers constant at two.
- Using four PS6000XV systems and two FS7500 systems we were able to support 93% more operations per second than when using a single FS7500 system. In this comparison the number of PS Series arrays (four PS6000XV arrays) remained constant while the number of FS7500 controllers doubled from two to four.
- When using two FS7500 systems and four PS6000XV arrays we were able to achieve more than twice the performance (168% increase) than when using one FS7500 and two PS6000XV arrays. In this comparison we doubled the number of PS Series arrays from two to four arrays and doubled the number of FS7500 controllers from two to four.
- When using two FS7500 systems and eight PS6000XV arrays we were able to achieve a 17% performance improvement than when using four PS6000XV arrays. In this comparison we increased the number of PS Series arrays from four to eight PS6000XV arrays while keeping the number of FS7500 controllers constant.
- When using two FS7500 systems and eight PS6000XV arrays we were able to achieve a 127% performance improvement over using one FS7500 and four PS6000XV arrays. In this comparison we doubled the number of PS Series arrays from four to eight arrays and doubled the number of FS7500 controllers from two to four.

From these results we can conclude the following:

- As we saw with the NFS results, scaling the FS7500 systems along with the PS Series storage arrays can provide maximum system throughput and performance.
- Also, as we saw with the NFS results, adding more FS7500 systems will produce more throughput for the same number of PS Series arrays.
- For the same set of tests, the CIFS performance scaled better than the NFS performance. We surmise that this is due to the different nature of the workloads and their respective server side implementations.

The average client response time of the CIFS READ_ANDX operation alone (18% of the mix) remained less than 39ms and for the CIFS WRITE_ANDX operation (9% of the mix) remained less than 12ms across configurations.

5.2 EqualLogic PS6000XV RAID50 scalability

NFS simulations were run on FS7500 using EqualLogic PS6000XV arrays in a RAID50 configuration. The configurations were scaled similar to the RAID10 study in the previous section. The test data set and workload characteristics are described in section 3.

Table 5 PS6000XV RAID 50 Scalability Test Configurations

Configuration	Workload Type	File System Size (TB)	File System Count	Total File System Size (TB)	Total Data Set Size (TB)	Test Client Count
1 x FS7500 + 2 x PS6000XV RAID50	NFS	2	4	8	1.36	16
1 x FS7500 + 4 x PS6000XV RAID50			8	16	1.83	16
2 x FS7500 + 4 x PS6000XV RAID50			8	16	2.5	32
2 x FS7500 + 8 x PS6000XV RAID50			16	32	3.66	32

Figure 8 below illustrates the NFS achieved operations per second for the configurations in a normalized scale. The RAID50 operations per second are compared with RAID10 operations per second recorded in the previous test. Total average client response time across all file operations remained under 10ms during all tests.
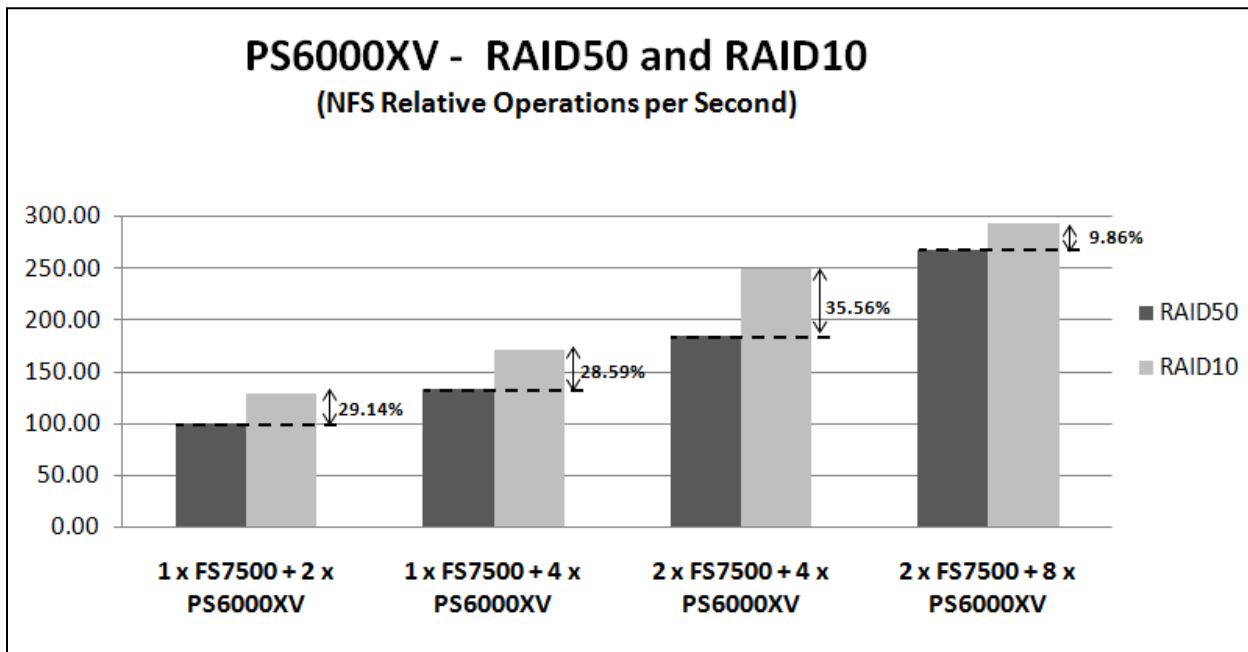

Figure 8: PS6000XV RAID50 and RAID10 NFS Scalability

The results presented in Figure 8 show the following scaling characteristics:

- RAID 10 performed approximately 30% to 35% better than RAID 50 in the first three configurations. This difference was due to the extra I/O write operations necessary to maintain RAID50 parity blocks at the storage layer.
- In the largest configuration, the performance difference decreased to approximately 10% in favor of RAID 10. The lower difference in this case was caused by the FS7500 system reaching its maximum achievable throughput in the RAID10 configuration.

With RAID 50, the average client response time of the NFS READ operation alone (18% of the mix) remained less than 39ms and for the NFS WRITE operation (10% of mix) remained less than 21ms across all configurations.

5.3 EqualLogic PS6500E RAID 50 scalability

NFS and CIFS simulations were run on FS7500 using an EqualLogic PS6500E array in a RAID50 configuration. The configurations were scaled similar to the method in the previous sections. The test data set and workload characteristics are described in section 3.

Table 6 PS6500E RAID 50 Scalability Test Configurations

Configuration	Workload Type	File System Size (TB)	File System Count	Total File System Size (TB)	Total Data Set Size (TB)	Test Client Count
1 x FS7500 + 2 x PS6500E RAID50	NFS	2	4	8	2	16
1 x FS7500 + 4 x PS6500E RAID50			8	16	2.45	16
2 x FS7500 + 4 x PS6500E RAID50			8	16	3.69	32
2 x FS7500 + 8 x PS6500E RAID50			16	32	3.87	32
1 x FS7500 + 2 x PS6500E RAID50	CIFS	2	4	8	2.34	16
1 x FS7500 + 4 x PS6500E RAID50			8	16	3.08	16
2 x FS7500 + 4 x PS6500E RAID50			8	16	5.25	32
2 x FS7500 + 8 x PS6500E RAID50			16	32	5.94	32

Figure 9 below shows the NFS achieved operations per second for the configurations using a % relative scale from the minimal system configuration. Total average client response time across all file operations remained less than 10ms during all tests.
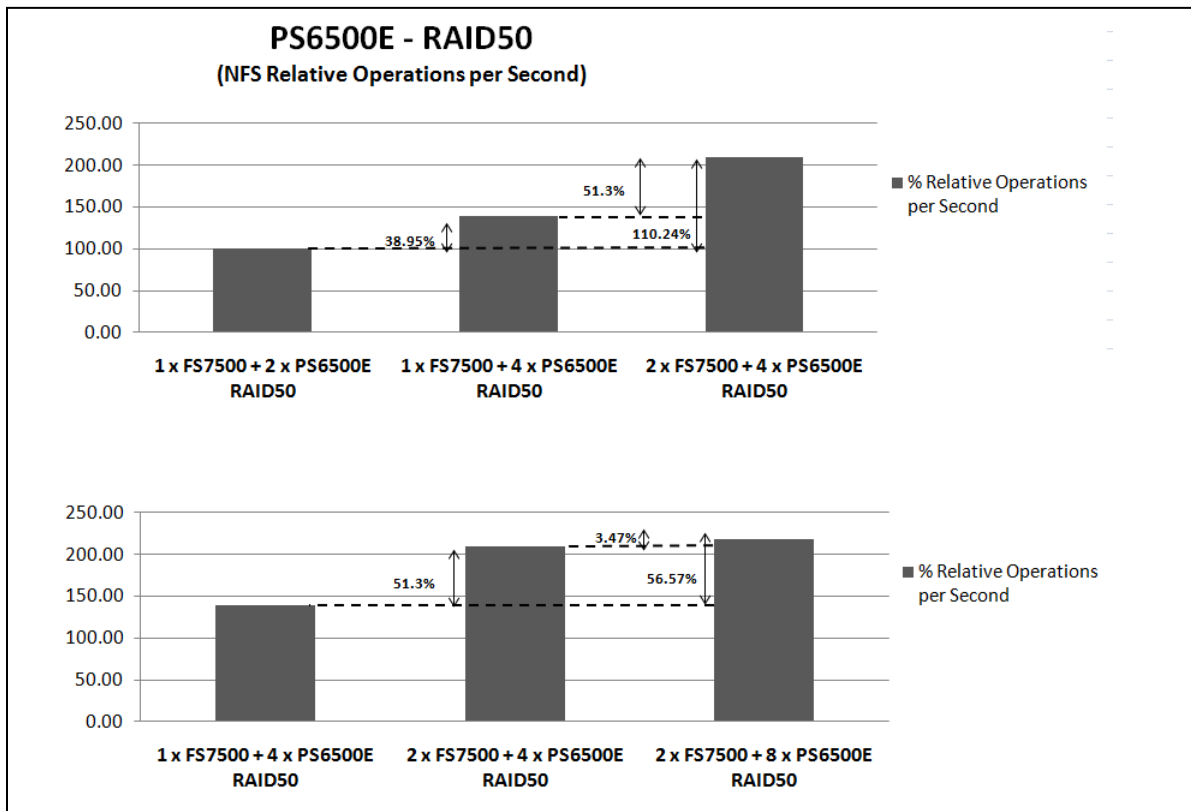

Figure 9: PS6500E RAID50 NFS Scalability

The results presented in Figure 9 show the following scaling characteristics:

- Using a single FS7500 system, four arrays yielded 38% more operations per second than two arrays. In this comparison we increased the number of PS Series arrays from two to four PS6500E arrays while keeping the number of FS7500 controllers constant at two.
- Using four PS6500E systems and two FS7500 systems we were able to support 51% more operations per second than when using a single FS7500 system. In this comparison the number of PS Series arrays (four PS6500E arrays) remained constant while the number of FS7500 controllers doubled from two to four.
- When using two FS7500 systems and four PS6500E arrays we were able to achieve approximately twice the performance (110% increase) than when using one FS7500 and two PS6500E arrays. In this comparison we doubled the number of PS Series arrays from two to four arrays and doubled the number of FS7500 controllers from two to four.
- When using two FS7500 systems and eight PS6500E arrays we were able to achieve a 4% performance improvement than when using four PS6500E arrays. In this comparison we increased the number of the PS Series arrays from four to eight PS6500E arrays while keeping the number of FS7500 controllers constant.

- When using two FS7500 systems and eight PS6500E arrays we were able to achieve a 56% performance improvement over using one FS7500 and four PS6500E arrays. In this comparison we doubled the number of the PS Series arrays from four to eight arrays and doubled the number of FS7500 controllers from two to four.

FS7500 system saturation capped throughput performance for the eight array configuration. This is apparent when you look at the difference between the largest configuration (two FS7500 systems and eight arrays) and the same configuration with four arrays. A small increase (4%) in throughput was measured as compared to using four arrays.

The average client response time of the NFS READ operation alone (18% of the mix) remained less than 33ms and NFS WRITE operation (10% of mix) remained less than 16ms across all configurations.

Figure 10 below illustrates the CIFS achieved operations per second for the configurations using a % relative scale from the minimal system configuration. Total average client response time across all file operations remained less than 10ms during all tests.
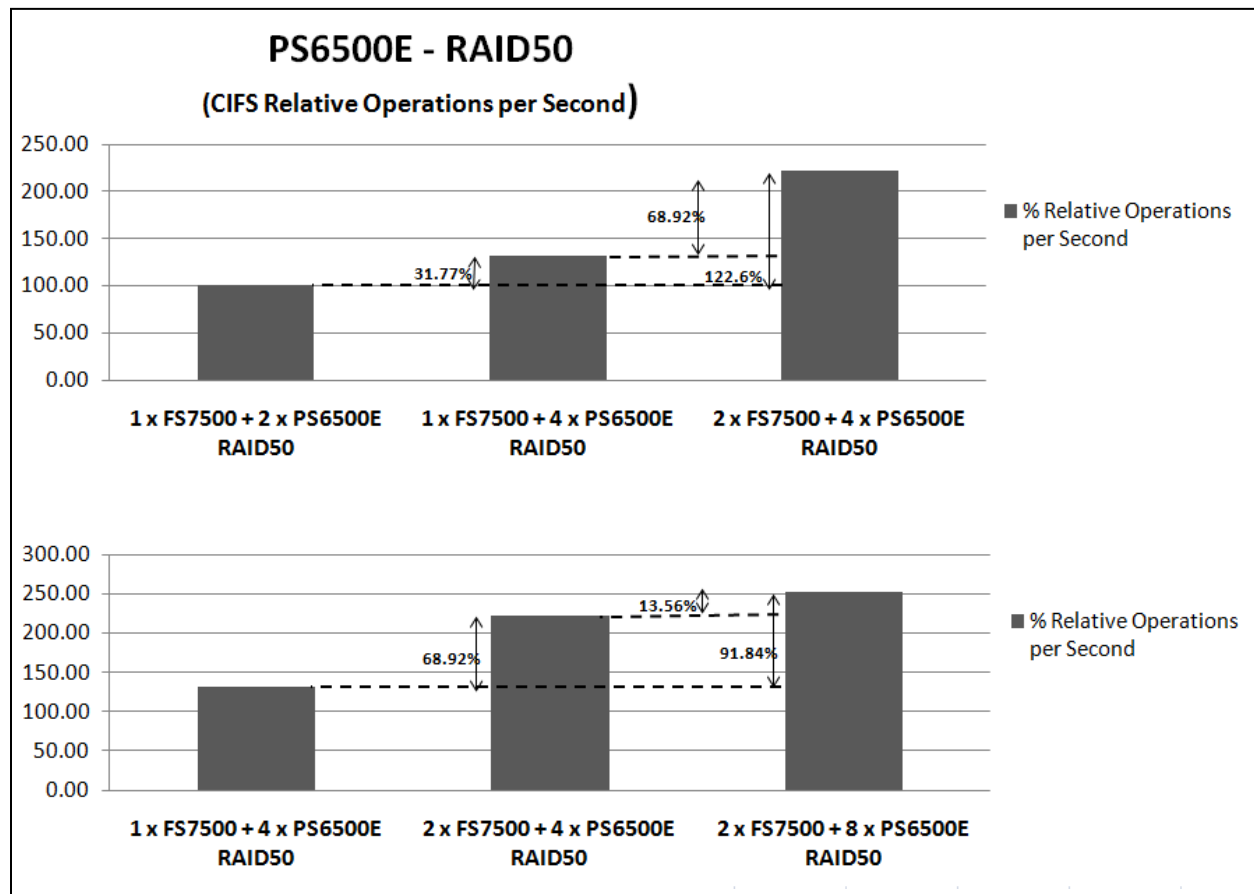

Figure 10: PS6500E RAID50 CIFS Scalability

The results presented in Figure 10 show the following scaling characteristics:
- Using a single FS7500 system, four arrays yielded 31% more operations per second than two arrays. In this comparison we increased the number of PS Series arrays from two to four PS6500E arrays while keeping the number of FS7500 controllers constant at two.

- Using four PS6500E systems and two FS7500 systems we were able to support 69% more operations per second than when using a single FS7500 system. In this comparison the number of PS Series arrays (four PS6500E arrays) remained constant while the number of FS7500 controllers doubled from two to four.
- When using two FS7500 systems and four PS6500E arrays we were able to achieve more than twice the performance (122% increase) than when using one FS7500 and two PS6500E arrays. In this comparison we doubled the number of PS Series arrays from two to four arrays and doubled the number of FS7500 controllers from two to four.
- When using two FS7500 systems and eight PS6500E arrays we were able to achieve a 13% performance improvement than when using four PS6500E arrays. In this comparison we increased the number of PS Series arrays from four to eight PS6500E arrays while keeping the number of FS7500 controllers constant.
- When using two FS7500 systems and eight PS6500E arrays we were able to achieve a 91% performance improvement over using one FS7500 and four PS6500E arrays. In this comparison we doubled the number of PS Series arrays from four to eight arrays and doubled the number of FS7500 controllers from two to four.

The higher scaling values observed with CIFS as compared to NFS across the configurations is due to the different nature of the workloads and their respective server side implementations.

The average client response time of the CIFS READ_ANDX operation alone (18% of the mix) remained less than 38ms and for the CIFS WRITE_ANDX operation (9% of the mix) remained less than 12ms across configurations.

5.4 NFS block size

NFS simulations were conducted to study how changing the NFS block size affects throughput operations per second. The FS7500 supports NFS version 3 (NFSv3) and supports block sizes up to 32KB. Two studies were conducted: one using 8KB block and one using 32KB block settings on the NFS test clients. The configuration is given below:

Table 7 NFS Block Size Test Configurations

Configuration	Workload Type	File System Size (TB)	File System Count	Total File System Size (TB)	Total Data Set Size (TB)	Test Client Count
1 x FS7500 + 2 x PS6000XV RAID50	NFS – 32KB	2	4	8	1.83	16
1 x FS7500 + 2 x PS6000XV RAID50	NFS – 8KB					

Figure 11 below shows the percent difference in NFS achieved operations / second for the two configurations and the average response time during each test.
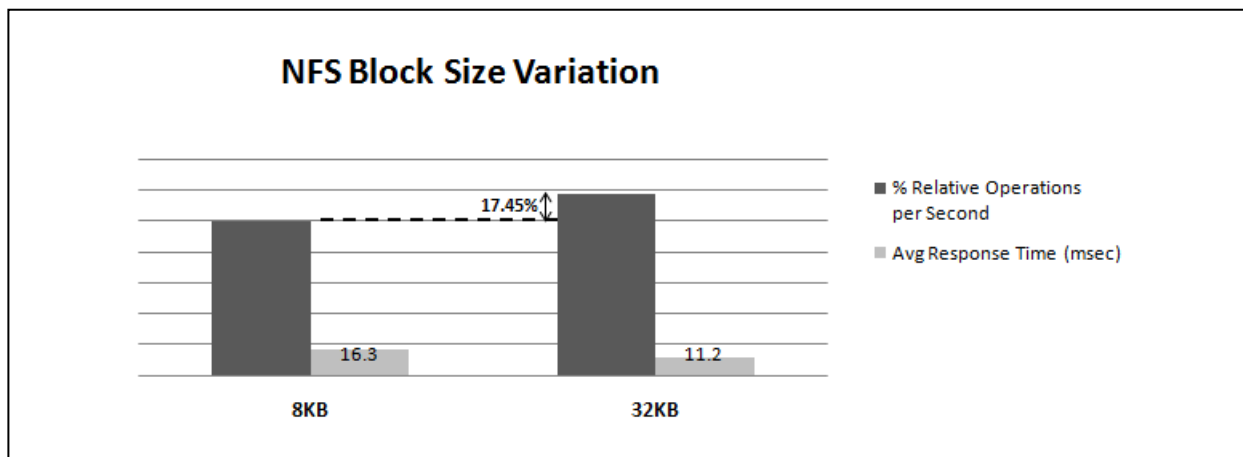
Figure 11: NFS Block Size Variation

As seen in Figure 11, the achieved operations per second with 32KB block size was 17% greater than that for 8KB block size. This difference is due to the processing overhead involved in fragmenting larger sized client file operation requests into smaller sized chunks of 8KB for transmission while the block size was set at 8KB. Also the response time was slightly higher with 8KB block size due to the same factor.

5.5 Client side MTU

NFS simulations were conducted to study the effect of client network MTU size on achieved throughput operations per second. Two studies were conducted: one using a1500 byte MTU and another using a 9000 byte MTU setting on the client network. The MTU settings for each test was set appropriately on the test client NICs, FS7500 client network NICs and the client network switch ports. The SAN network MTU was kept at 9000 bytes during all tests as per general best practice recommendations. The test configuration is given below:

Table 8 Client Side MTU Test Configurations

Configuration	MTU	Workload Type	File System Size (TB)	File System Count	Total File System Size (TB)	Total Data Set Size (TB)	Test Client Count
1 x FS7500 + 2 x PS6000XV RAID50	9000	NFS	2	4	8	1.83	16
1 x FS7500 + 2 x PS6000XV RAID50	1500						

Figure 12 below shows the percent difference in NFS achieved operations / second for the two configurations and the average response time during each test.
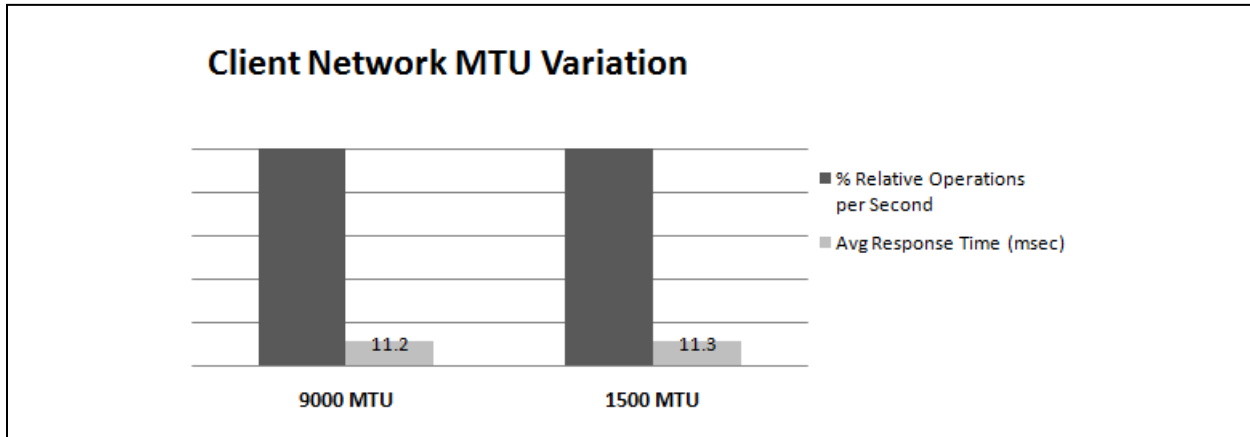
Figure 12: Client network MTU Variation

As seen in Figure 12, there was virtually no difference when varying client side MTU size. This is primarily due to the random nature of the file I/O workload simulated in the file share use case.

5.6 RAID Studies – RAID50, RAID5 and RAID 6

NFS and CIFS simulations were conducted with RAID 50, RAID 5 and RAID 6 on PS6500E. The configurations tested are shown below.

Table 9 PS6500E RAID Type Test Configurations

Configuration	Workload Type	File System Size (TB)	File System Count	Total File System Size (TB)	Total Data Set Size (TB)	Test Client Count
1 x FS7500 +2 x PS6500E RAID50	NFS	2	4	8	2.00	16
1 x FS7500 + 2 x PS6500E RAID6					1.42	
1 x FS7500 + 2 x PS6500E RAID5					1.42	
1 x FS7500 + 2 x PS6500E RAID50	CIFS	2	4	8	2.34	16
1 x FS7500 + 2 x PS6500E RAID6					2.28	
1 x FS7500 + 2 x PS6500E RAID5					2.20	

Figure 13 below shows the percent difference in NFS achieved operations per second for the configurations in a normalized scale. Total average client response time across all file operations remained less than10ms.

Figure 13: PS6500E NFS RAID Studies

The achieved operations per second with RAID 5 and RAID 6 were lower than for RAID 50. RAID 5 and RAID 6 incur penalties for parity calculations during write operations. The block I/O from NAS nodes to storage arrays included significant amounts of writes with the NFS workload. Approximately 20% lower operations per second were measured with RAID5 and RAID6.

Figure 14 below shows the percent difference in CIFS achieved operations per second for the configurations in a normalized scale. Total average client response time across all file operations remained less than 10ms.
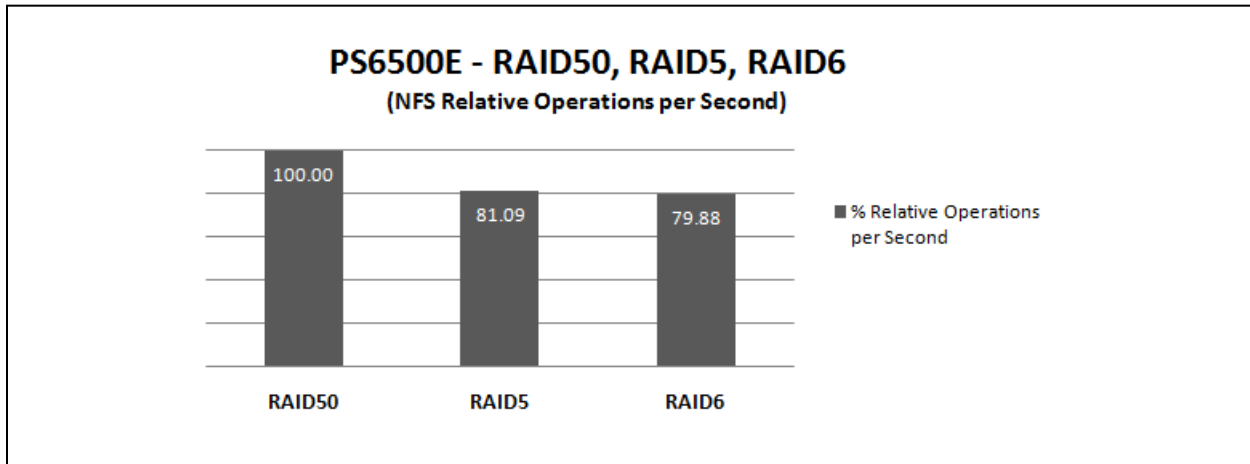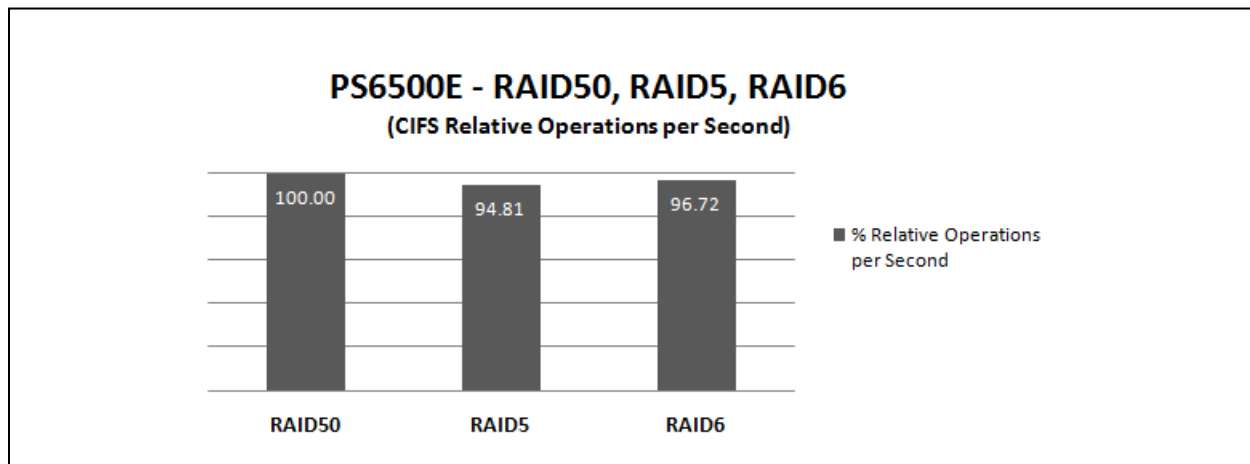

Figure 14: PS6500E CIFS RAID studies

The achieved CIFS operations per second with RAID 5 and RAID 6 were slightly lower than RAID 50. In this test the differences were not as great as for the NFS workload in the previous test. This is because the block I/O from the NAS nodes had a lower write I/O percentage.

5.7 NFS data and meta-data mix

The default mix of NFS operations in our test simulations was 28% data and 72% meta-data. Data operations are the actual read and write I/O operations to files. They incur higher latencies and can be deemed heavy I/O. Meta-data operations are operations on file and file system attributes. They incur lower latencies and can be deemed light I/O. Different application use cases may have differing needs of data and meta-data operations on the file system. For example, a file share dedicated to a departmental workgroup of users working on graphical editing applications may incur more data operations than normal. On the other hand, a file share used as a document repository with periodic file access and heavy search activity may incur more meta-data operations.

NFS simulations were conducted with the following mix of data and meta-data operations:

Table 10 NFS data mix

Configuration	Workload Type	File System Size (TB)	File System Count	Total File System Size (TB)	Total Data Set Size (TB)	Test Client Count
1 x FS7500 + 2 x PS6500E RAID6	NFS (56%D, 44%MD)	2	4	8	0.8	16
	NFS (28%D, 72%MD)				1.42	
	NFS (14%D, 86%MD)				2.34	

Figure 15 below shows the percent difference in NFS achieved operations per second for the three configurations described above.
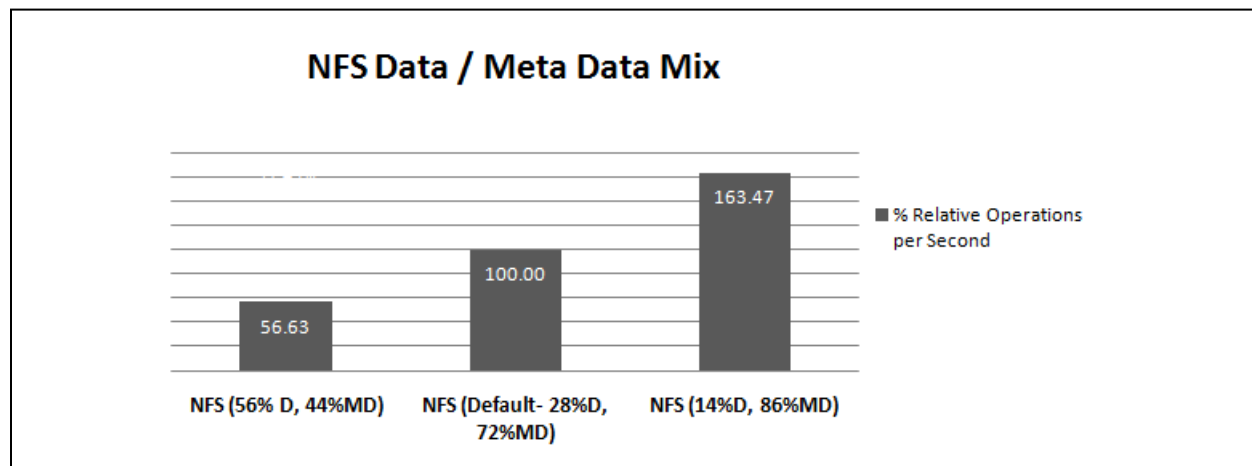
Figure 15: NFS Data and Meta-data mix

The simulation with more data operations yielded lower operations per sec. This is due to the higher processing cost of the data operations with regards to latency and the amount of actual data accessed from physical disks on storage arrays. On the other hand, the simulation with more meta-data operations yielded more operations per second. This is due to the lower processing cost of meta-data operations in terms of latency and data accessed from physical disks.

5.8 Sequential write I/O simulation

Studies were conducted to understand the maximum sequential I/O throughput of the solution using I/O simulation on CIFS shares. The I/O simulation tool used was IOMeter. In order to ensure sequential nature of the client I/O, each NIC on the FS7500 client network served only one client machine and also only one I/O thread per client was simulated. A NAS file system was created for each client with a corresponding CIFS share exposed to the client. Each client machine simulated sequential I/O to its corresponding CIFS share using a single large IOMeter data file. The configurations tested are given below.

Table 11 PS6500E RAID50 Sequential I/O Comparison

Configuration	Workload Type	File System Size (GB)	File System Count	Total File System Size (GB)	Total Data Set Size (GB)	Test Client Count
2 x FS7500 + 8 x PS6500E RAID50 (Clients load balanced to all 4 nodes)	IOMeter – 1MB SeqWrite (CIFS)	100	16	1600	1600	16
2 x FS7500 + 8 x PS6500E RAID50 (Clients load balanced to 2 nodes only)			8	800	800	8

In the first configuration clients were load balanced to all four nodes in the 2 x FS7500 systems and in the second configuration clients were load balanced to only two. IOMeter measured sequential throughput from clients as a percentage of available bandwidth. When using one FS7500 system, there are eight 1GbE NICs available for the client network, corresponding to a total aggregate bandwidth of approximately 950 MB/s. When using two FS7500 systems sixteen 1GbE NICs corresponding to approximately 1900 MB/s is available for client network. The results of the test are shown in figure 16 below.
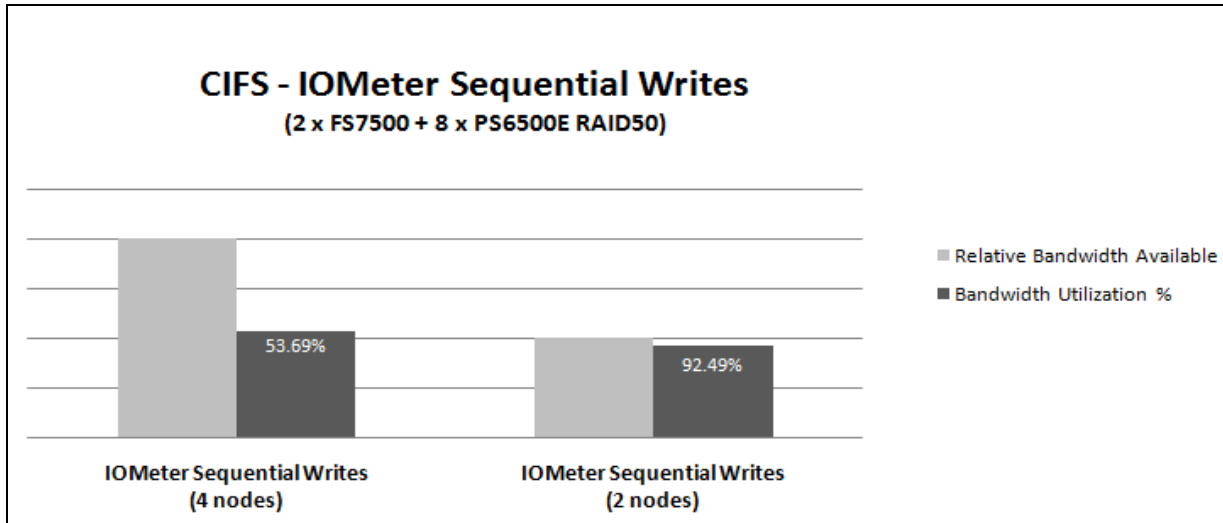
CIFS - IOMeter Sequential Writes
(2 x FS7500 + 8 x PS6500E RAID50)

53.69%

92.49%

■ Relative Bandwidth Available
■ Bandwidth Utilization %

IOMeter Sequential Writes
(4 nodes)

IOMeter Sequential Writes
(2 nodes)

Figure 16: CIFS IOMeter sequential write throughput

With two nodes 92% of the available client network bandwidth was utilized, offering close to line rate utilization. With four nodes, about 54% of the bandwidth was utilized. The primary constraint with the four node configuration was the number of backend disk drives used in testing. The end-to-end total throughput was limited by the number and type of disks utilized. Using a different array model with higher performing drives and drive count would have resulted in higher utilization with four nodes.

6 Best practices

In this section we provide best practice design and deployment recommendations based on the test results presented in this paper.

6.1 Network

You should use separate network infrastructure (network switches) for the client network (LAN) and the SAN+FS7500 Internal Communication network.

6.1.1 SAN design

1. Redundant inter-connected switches or switching fabrics are recommended for the SAN network infrastructure to ensure high-availability. Switch inter-connects (stack or LACP) should be properly sized with to provide the required bandwidth for the PS Series arrays.
2. Flow control should be enabled on the switch ports connecting to the FS7500 and the arrays.
3. Enable jumbo frames (MTU 9216) on the switch ports connecting to the FS7500 and arrays.
4. The FS7500 internal network NIC ports and storage NIC ports should be connected in a way such that any single component failure in the SAN will not disable access to any storage array volumes or peer FS7500 controllers.
5. The FS7500 internal NICs and SAN NICs should be on the same layer 2 network and not separated by VLANs.
6. For all FS7500 controllers, ensure that their corresponding NIC ports are connected on the same switch. For example, the first internal NIC of all FS7500 controllers on a single switch ("switch1") in the network. Second internal NIC of all 7500 controllers on a single switch which is same ("switch1") or different ("switch2").
7. Disable spanning tree on switch ports connecting to end devices like FS7500 ports and storage ports. Portfast should be enabled for these ports.

> **Note:** General recommendations for EqualLogic PS Series array network configuration and performance are provided in the following documents.
>
> ➢ *PS Series Array Network Performance Guidelines:*
>
>   http://www.equallogic.com/resourcecenter/assetview.aspx?id=5229
>
> ➢ *EqualLogic Configuration Guide:*
>
>   http://www.delltechcenter.com/page/EqualLogic+Configuration+Guide

6.1.2 LAN

1. Redundant interconnected switches or switching fabrics are recommended for the LAN network infrastructure to ensure high-availability. Switch inter-connects (stack or LACP) should be properly sized with to provide the required bandwidth for the PS Series arrays.
2. The FS7500 client network NIC ports should be connected in a way such that any single component failure in the LAN switches will not disable access to the file system.
3. Enable flow control on the switches connecting to the FS7500 client network NIC ports.

4. Enable jumbo frames (MTU 9216) on the switch ports connecting to the FS7500 client network NIC ports to assist with sequential file I/O requests. If enabled, ensure jumbo frames is enabled across the network end-to-end from the file client to the FS7500 system.
5. Ensure that the file system client software on the file clients is configured to utilize the maximum block size for the file protocol. This will be 32KB for NFS v3 clients and 64KB for CIFS clients.
6. For routed client side networks, ensure an appropriate load balancing mechanism is chosen to balance incoming client traffic to FS7500 NAS nodes. Follow the FS7500 setup guidelines to choose your load balancing mechanism.
7. Follow the FS7500 setup guidelines for FS7500 client network IP configuration based on the IP load balancing policy mechanism chosen.
8. Ensure that the client network end-to-end latency between file clients and FS7500 is within acceptable limits. Ideally this should be less than 1000ms for the targeted end-user file share use case.

6.2 Storage

EqualLogic PS Series storage array best practice recommendations:

1. You should dedicate a separate storage pool for file system NAS reserve in environments with heavy file I/O needs.
2. When sharing a storage pool for hosting both file system NAS reserve and volumes for other block I/O applications, you need to ensure latencies are within acceptable limits for the block applications while the NAS is running under typical I/O load.
3. For higher file I/O throughput, scale the FS7500 systems along with PS Series arrays. In environments with very high file I/O needs, choose 2 X FS7500 systems when there are four or more PS Series arrays used to host the NAS reserve.
4. Choose the appropriate RAID type for the PS Series arrays hosting the NAS reserve based on capacity and workload performance requirements:
   - For high write data operations in user load: RAID 10 offers best performance at lower capacity levels.
   - For moderate mix of data reads and writes in user load: RAID50 offers good performance at good capacity levels.
   - For high read data operations in user load: RAID5 or RAID6 offer good performance at high capacity levels.
5. Consider the amount of file data and meta-data operations in the user workload. High meta-data operations in the workload will require a lower number of PS Series arrays for the NAS reserve. High file data focused workloads will require more PS Series arrays and may also require two FS7500 systems based on the workload.
6. For workloads performing sequential I/O, ensure end-to-end bandwidth availability between the file clients and FS7500 systems. Also enable end-to-end jumbo frames on the client network to assist with improved throughput. Ensure sufficient number of PS Series arrays are deployed for the NAS reserve to provide sufficient disk bandwidth, network bandwidth and processing resources for sequential file I/O.

Appendix A

A.1   Test system components

The table below provides details for the major hardware and software system components used in the test system configuration.

Table 12 Test System Configuration Details

Hardware Components	
Test Client Servers	• 2 x Dell PowerEdge R815 Server: 4 x 12 Core AMD Opteron TM 6174 2.2 GHz Processors, 96 GB RAM, 4 x Broadcom NetXtreme II 5709 Quad Port Gigabit Ethernet NIC Cards installed
Network	• SAN: 4 x Dell PowerConnect 6248 Switches installed in top of rack • LAN: 2 x Dell PowerConnect 6248 Switches installed in top of rack • Firmware: 3.2.1.3
Storage	• Dell EqualLogic FS7500 • Firmware 1.0 • Dell EqualLogic PS6000XV: 16 x 450GB 15K SAS disks with dual 4-port 1GbE controllers • Firmware 5.1 • Dell EqualLogic PS6500E: 46 x 2048GB 7.2K SATA II disks with dual 4-port 1GbE controllers • Firmware 5.1
Software Components	
Test Clients	• Host: Windows Server 2008 R2 Enterprise Edition with Hyper-V • Guest: Microsoft® Windows Server 2008 R2 Enterprise Edition (virtual machine)
Monitoring	• SAN Headquarters 2.1.0.5342

A.2 IP and Subnet information

Table 13 Test Network IP Configuration

Entity	Subnet 255.255.255.0	IP Address 1	IP Address 2	IP Address 3	IP Address 4
NAS Service IP	172.16.101.x	172.16.101.100			
NAS Node Client Network:					
Node 1	172.16.101.x	172.16.101.111			
Node 2	172.16.101.x	172.16.101.121			
Node 3	172.16.101.x	172.16.101.131			
Node 4	172.16.101.x	172.16.101.141			
NAS Node SAN:					
Node 1	172.16.103.x	172.16.103.11	172.16.103.12	172.16.103.13	172.16.103.14
Node 2	172.16.103.x	172.16.103.21	172.16.103.22	172.16.103.23	172.16.103.24
Node 3	172.16.103.x	172.16.103.31	172.16.103.32	172.16.103.33	172.16.103.34
Node 4	172.16.103.x	172.16.103.41	172.16.103.42	172.16.103.43	172.16.103.44
NAS Node Internal Network	172.16.102.x	172.16.102.0 To 172.16.102.255			
NAS Service Management IP	172.16.103.x	172.16.103.101			
EqualLogic Group IP	172.16.103.x	172.16.103.100			
Member IP:					
Member 1	172.16.103.x	172.16.103.111	172.16.103.112	172.16.103.113	172.16.103.114
Member 2	172.16.103.x	172.16.103.121	172.16.103.122	172.16.103.123	172.16.103.124
Member 3	172.16.103.x	172.16.103.131	172.16.103.132	172.16.103.133	172.16.103.134
Member 4	172.16.103.x	172.16.103.141	172.16.103.142	172.16.103.143	172.16.103.144
Member 5	172.16.103.x	172.16.103.151	172.16.103.152	172.16.103.153	172.16.103.154
Member 6	172.16.103.x	172.16.103.161	172.16.103.162	172.16.103.163	172.16.103.164
Member 7	172.16.103.x	172.16.103.171	172.16.103.172	172.16.103.173	172.16.103.174
Member 8	172.16.103.x	172.16.103.181	172.16.103.182	172.16.103.183	172.16.103.184

Related Publications

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

➢ The Dell EqualLogic FS7500 Unified Storage Solution

http://www.equallogic.com/products/default.aspx?id=10465

➢ EqualLogic PS Series Architecture

http://www.equallogic.com/products/default.aspx?id=5823

➢ Dell Fluid File system White Paper

http://www.dell.com/downloads/global/products/pvaul/en/dell-scalable-file-system-white-paper.pdf

➢ EqualLogic Configuration Guide:

http://www.delltechcenter.com/page/EqualLogic+Configuration+Guide