

BENEFITS AND BEST PRACTICES FOR DEPLOYING SSDS IN AN OLTP ENVIRONMENT USING DELL EQUALLOGIC PS SERIES

Using Solid State Disks (SSDs) in enterprise storage arrays is one of today's hottest storage trends. Mechanical hard disk drives (HDDs) with a cache memory-based front-end in the disk array controller have been the dominant I/O architecture for high-end application data storage until now. However, with increasing capacities and falling prices of memory chips, and the growing number of performance-hungry applications, many storage vendors have introduced SSDs in their disk arrays to improve application performance. Moreover, since SSDs can consume significantly less power than HDDs, they also help reduce energy consumption in power-hungry data centers.

For the past 25 years, while CPU speeds and HDD capacities have been increasing exponentially, HDD I/O per second (IOPS) has improved only in modest increments with drive spin rates, constraining application performance. As a result, IT managers partition and tier their application data storage from slow, high-capacity SATA HDDs (Tier-2) to fast, more expensive SAS HDDs (Tier-1).

SSDs add another tier (Tier-0) to the application storage architecture. Although currently limited in capacity, SSDs can offer up to an order-of-magnitude IOPS improvement over HDDs. However, real-world, latency-sensitive applications such as banking, electronic trading, web searching, and online reservation systems do not benefit equally from the higher IOPS performance of SSDs. Workload-specific I/O characteristics dictate the performance benefits an application can gain from SSDs. To demonstrate the benefit of SSD arrays with real-world applications, Dell Labs tested the capabilities of SSDs using an Online Transaction Processing (OLTP) workload. The database was implemented in a tiered storage model that includes the latest Dell EqualLogic™ PS6000 arrays with SAS HDD and SSD drives. Dell Labs conducted multiple tests by moving various database elements (including logs, temporary files and data objects) from HDD to SSD drives and measuring the performance deltas. These tests demonstrate that in an EqualLogic iSCSI SAN, depending on your user load, latency-sensitive OLTP database applications can expect up to an estimated 75% increase in transaction throughput or up to a 60% improvement in response times by isolating and moving read-intensive datasets from SAS drives to SSD drives as compared to moving the same datasets to additional SAS drives.

This whitepaper also discusses the best practices for deploying SSDs in an OLTP environment using Dell EqualLogic PS Series arrays.



EQUALLOGIC

www.BDTIC.com/SAMSUNG

Dell EqualLogic PS Series Storage Advantages

- On-line and on-demand scalability — no downtime, no hardware reconfigurations, performance can scale linearly with capacity
- Automatic Management — quick to set up, easy to manage - including thin provisioning, load-balancing, and creation of snapshots, clones and remote replicas
- All-inclusive pricing model — helps eliminate hidden costs, software prices included, simplifies purchasing cycle

DELL EQUALLOGIC PS SERIES TIERED STORAGE

The Dell EqualLogic PS Series redefines simplicity and economics in storage. Its online, on-demand scalability, ease of management, advanced load balancing features, and innovative all-inclusive pricing structure help IT administrators reduce storage costs and complexity. EqualLogic PS Series arrays support RAID 10, 50, 5, and 6 configurations using SSD, SAS and SATA drives, providing multiple storage tiers to meet the vast majority of your performance and capacity needs. Just like HDDs, SSDs appear as standard drives to existing management tools, allowing IT managers to manage SSDs with no special processes, tools, or training.

Dell has introduced SSDs in the EqualLogic PS Series product line at a very affordable price to bring the high-performance and low-power capabilities of SSDs to value-oriented enterprise customers. However, SSD pricing is still higher than HDD for comparable storage capacity, and it may be impractical for storing the entire workload data on SSDs for many customers. The purpose of this whitepaper is to describe the performance gains available when implementing SSD as part of an overall tiered storage deployment strategy.

In this document we first describe the tools and the tests conducted by Dell Labs to demonstrate the performance benefits of PS Series SSD arrays using an OLTP workload running on an Oracle® database, implemented with PS6000XV SAS HDD and PS6000S SSD storage tiers.

We then present the relative throughput and transaction response times of various storage configurations obtained by moving database elements such as logs, temporary files, and tablespaces between the arrays. Finally, we offer best-practice recommendations to guide users when introducing PS Series SSD arrays into PS Series HDD array-based database deployments.

ADVANTAGES OF SSD

Unlike mechanical hard disk drives, solid state disks are made of silicon memory chips and have no moving parts. As with hard disks, the data is persistent in SSDs when they are powered down. A computer's operating system treats these devices just like other disks. However, SSDs have near-zero seek time and no rotational delay, drastically reducing response times and latency. Applications that are adversely impacted by storage latency will benefit the most from SSD-based storage.

IT administrators generally implement best practices to tune latency-sensitive applications. A common method is to keep as much application data as possible in server memory, thereby reducing the frequency with which the application must retrieve data from the physical HDDs, as this process has much longer read or write latency than server memory. With exponentially growing application data sets, server memory — with its limited capacity — can end up representing only a small percentage of total application data while the system relies heavily on HDD reads/writes. As a result, IT administrators often implement one or both of the following best practices to reduce HDD access time:

1. Striping application data on a large number of physical disks (wide stripes)
2. Writing application data only on a small portion of a HDD (short stroking)

Unfortunately, these best practices can increase foot-print, costs (hardware, software, power, and maintenance), and complexity. Including SSDs in your solution deployment can help simplify storage configuration and lower costs.

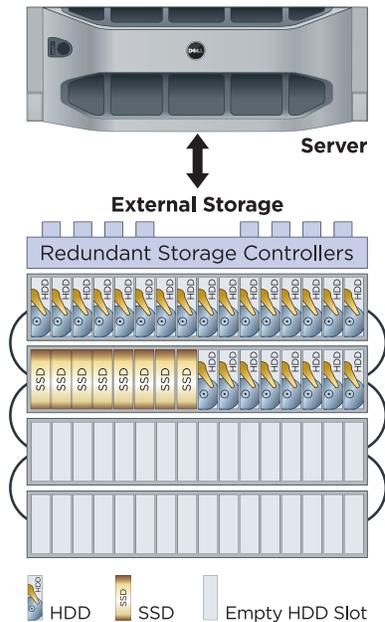


Figure 1 - Solid State Disks in a traditional scale-up frame-based architecture

SSDS IN EQUALLOGIC PS SERIES STORAGE ARCHITECTURE

Most traditional scale-up frame-based disk arrays stack HDD shelves behind a set of redundant storage controllers. When high IOPS SSDs are configured in such storage array architectures (Figure 1), they have to share the array controller resources with the other SAS and SATA HDDs. As a result, such storage arrays can support only a handful of SSDs per frame. On the other hand, the EqualLogic PS Series is based on unique scale-out peer storage architecture (Figure 2). In this context, “peer” describes the collaboration and equal partnership of member arrays that function as peers in a PS Series Group. Individually, each member in a PS Series Group is a fully-functional, high-performance, highly-available storage array with mirrored write-back caches and multiple storage network connections in its own redundant storage controllers. Members in a PS Series Group work together to share resources, evenly distribute loads, and collaborate to help optimize application performance and provide comprehensive

data protection. Moreover, PS Series members can be added to or removed from a PS Series Group with no complex administration tasks or impact on availability. As a result, resources like disks, controllers, caches, and network connections can be easily added to and removed from an EqualLogic PS Series storage solution to scale its capacity and performance.

This scale-out peer architecture of EqualLogic provides an opportunity to unlock the potential of SSDs for higher performance. Affordable PS6000S members can be configured in a separate pool (Tier-0) within the same SAN as PS Series SAS and SATA HDD arrays, but with dedicated controller and network port resources. Moreover, the all-inclusive pricing model of EqualLogic PS Series helps make this SSD solution even more affordable and cost effective. Advanced software features such as snapshots, clones, and remote replication, along with application-specific integration for Microsoft Exchange, SQL Server, and hypervisor-

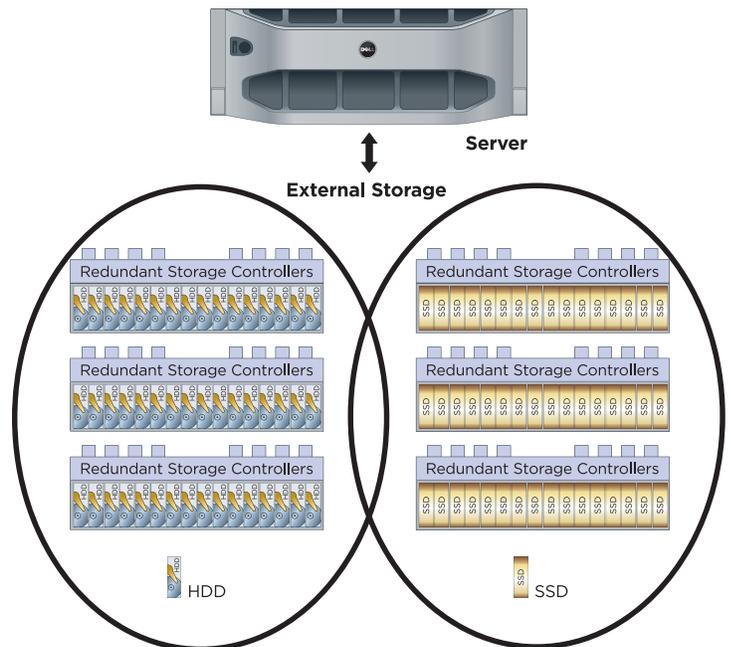


Figure 2-Solid State Disks in an EqualLogic scale-out peer architecture

based VMware and Microsoft Hyper-V virtual machine environments are included with the EqualLogic PS Series arrays at no additional cost.

USE CASES

Based on tests conducted by Dell Labs, SSDs work best when the I/O pattern is random and I/Os are mostly read with small transfer block size (for example, 8KB). These I/O characteristics are very typical in OLTP workloads. Traditional HDD latency consists of seek time, rotational delay, and actual data transfer time; in random transfers, seek and rotational latency delays are disproportionately larger than the data transfer time. SSD storage greatly reduces the seek times, and eliminates the rotational delay (since there are no moving parts) ultimately reducing the amount of time an OLTP transaction waits for physical I/Os to complete. Dell Labs conducted tests with the EqualLogic PS6000S array in various configurations. These tools, tests and results are outlined below.

TEST TOOLS AND CONFIGURATIONS

To measure the performance of EqualLogic PS6000S arrays, Dell Labs used two tools —Oracle I/O Numbers (ORION) and Quest® Software’s Benchmark Factory®.

ORION (Oracle I/O Numbers)

ORION is a tool for simulating Oracle I/O workloads. In this study, the small random I/O workload was tested; it simulated the typical OLTP applications in which most I/O operations are random reads and writes with an I/O size equivalent to the database block size, typically 8 KB. Test outputs include throughput (measured in IOPS) and I/O response times, key criteria for comparing the performance of storage subsystems.

ORION can run tests at different levels of I/O load to measure performance metrics such as IOPS and I/O latency. Load level is expressed in terms of the number of outstanding asynchronous I/Os.

An ORION test was conducted with one PS6000S array and one PS6000XV array, and simulated a typical OLTP workload by using 70% small, random read I/Os and 30% small, random write I/Os.

The test, run against three 70 GB raw volumes, simulates the effect of striping performed by Oracle’s Automatic Storage Management (ASM), which allows the Oracle database to optimize reads and writes to multiple storage volumes. The ORION test configuration is summarized in Table 1A.

Benchmark Factory

Quest Software’s Benchmark Factory TPC-C is a load-generating utility that simulates OLTP users and transactions on a database for a given number of users. The database configuration used in this test was a 2-node Oracle 11 g (11.1.0.7) Real Application Cluster (RAC). The total database schema size was 130 GB, which was populated by Benchmark Factory. The test outputs include metrics such as the average transaction response time and transactions per second (TPS). The Benchmark Factory test configuration is summarized in Table 1B.

HARDWARE AND SOFTWARE CONFIGURATIONS FOR ORION TEST	
Server	One Dell PowerEdge 2950 server with: <ul style="list-style-type: none"> • Two Intel® Xeon® quad-core 3.16 GHz CPU • 32 GB of RAM • 4 One Gigabit Intel PRO/1000 NIC ports for iSCSI traffic
External Storage	Dell EqualLogic PS6000XV OR PS6000S <ul style="list-style-type: none"> • RAID 10 with 2 spare disks in each member • 15K RPM SAS disks in PS6000XV; 50 GB SSD disks in PS6000S • Firmware: Highland Park Beta Gold 4.1.1 (R88972)
Volume Configuration	Three volumes of 70 GB each
OS and Device Driver	Microsoft Windows 2003 Servers x64 R2 Enterprise Edition with SP2 <ul style="list-style-type: none"> • Microsoft iSCSI initiator 2.0.8 • EqualLogic Multipath I/O Device Specific Module (DSM) Version 3.2 beta
Storage Network	2 stacked Dell PowerConnect 6248 Gigabit Ethernet switches for iSCSI SAN
Test Software	ORION Version: 10.2.0.1.0

Table 1A: ORION Test Configuration

Storage Network Configuration

For both ORION and Benchmark Factory tests, two Dell™ PowerConnect™ 6248 Gigabit Ethernet switches were implemented to connect the host servers to the

HARDWARE AND SOFTWARE CONFIGURATIONS FOR BENCHMARK FACTORY TEST	
Server	Two Dell PowerEdge M710 blade servers with: <ul style="list-style-type: none"> • Two Intel® Xeon® quad-core 2.67 GHz CPU • 24 GB of RAM • 4 of 1 Gb Broadcom NetXtreme II NIC ports for iSCSI traffic
External Storage	Dell EqualLogic PS6000XV OR PS6000S <ul style="list-style-type: none"> • RAID 10 with 2 spare disks in each member • 15K RPM SAS disks in PS6000XV; 50 GB SSD disks in PS6000S • Firmware: Highland Park Beta Gold 4.1.1 (R88972)
Volume Configuration	One 170 GB volume; One 100 GB volume; One 80 GB volume
OS and Device Driver	Microsoft Windows 2003 Servers x64 R2 Enterprise Edition with SP2 <ul style="list-style-type: none"> • Microsoft iSCSI initiator 2.0.8 • EqualLogic Multipath I/O Device Specific Module (DSM) Version 3.2 beta
Storage Network	2 stacked Dell PowerConnect 6248 Gigabit Ethernet switches for iSCSI SAN
Test Software	Quest Benchmark Factory 5.7.1 with Oracle 64 bit 11.1.0.7 EE RAC

Table 1B: Hardware and Software Configurations for Benchmark Factory Test

storage subsystem, and to segregate the iSCSI SAN traffic from the public and private LAN traffic. The two Gigabit Ethernet switches were stacked to provide high bandwidth for the inter-switch network. The following EqualLogic network best practices were implemented:

- Enable flow control
- Enable Spanning Tree Protocol (STP) port fast feature
- Enable broadcast and multicast storm control
- Disable unicast storm control
- Enable jumbo frames

TEST RESULTS

ORION OLTP Workload Results

Figures 3 and 4 show results from the ORION test for a typical OLTP I/O workload represented by 70% small, random read I/Os and 30% small, random write I/Os. Figure 3 illustrates the IOPS at different load levels. Figure 4 illustrates the I/O latency at different load levels. All results are normalized and are provided here solely for the purpose of comparison between the

PS6000S and PS6000XV. They do not portray the maximum capacities of either storage system.

As shown in Figures 3 and 4, for small random I/O mixed read/write OLTP workload, the PS6000S delivered approximately 2.5 to 3 times better IOPS as compared to PS6000XV at higher load levels, and up to 12 times better IOPS at lower load levels.

Benchmark Factory TPCC Results

Dell Labs conducted multiple tests on the EqualLogic iSCSI SAN, varying the location of various Oracle database components. An architectural overview of the 11g RAC database is shown in Figure 5.

As illustrated in Figure 5, the EqualLogic storage group includes two members: one PS6000XV and one PS6000S, with a separate pool created for each array. Initially, the Oracle RAC database resided on the following three volumes within the SAS storage pool:

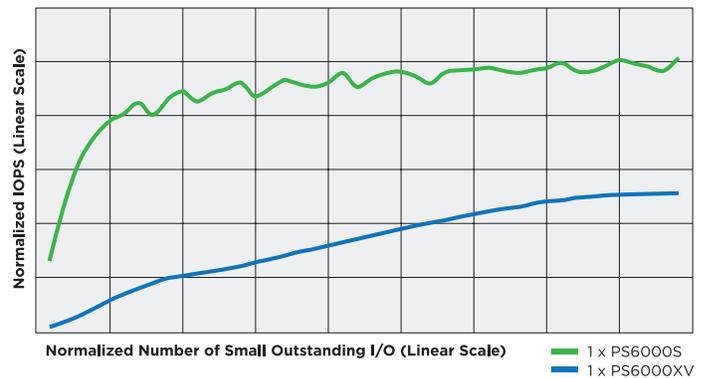


Figure 3 - IOPS vs. Load for OLTP workload

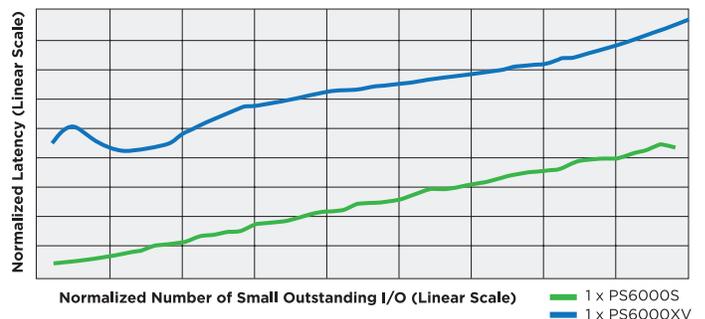


Figure 4 - I/O Latency vs. Load for OLTP workload

- 1-GB volume hosting Oracle Clusterware files including Oracle Cluster Registry (OCR) and Cluster Synchronization Services (CSS) Voting Disk. This volume is formatted as a RAW partition in the database server operating system.
- 170 GB volume hosting database files including online redo log files, control files, and the temporary tablespaces. This volume is formatted as an Oracle Automatic Storage Management (ASM) diskgroup.
- 20-GB volume hosting the Oracle Flash Recovery Area which stores the archived redo log files. This volume is also formatted as an Oracle ASM diskgroup.

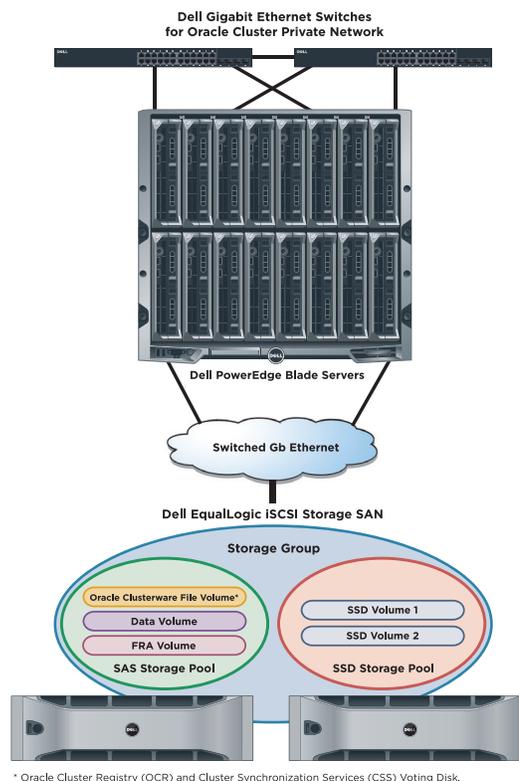
A Benchmark Factory TPC-C schema was created and populated with approximately 130 GB of data, including tables and indexes. With all database files initially residing in the SAS storage pool of the PS6000XV, the baseline performance characteristics were obtained. By using the statistics collected during this initial test, we identified the read-intensive database objects suitable for the SSD disks.

IDENTIFYING CANDIDATES FOR SSD DEPLOYMENT

An Oracle database has elements with high I/O demands, including the online redo log files, the undo tablespaces, and the temporary tablespace. In addition, read-intensive database objects are also prime candidates to be moved to SSDs. Oracle's Automatic Workload Repository (AWR), a performance gathering tool, is available with the current Oracle database release. The reports from this tool were used to identify the most read-intensive database objects.

From the AWR report generated during the initial Benchmark TPC-C test, the segment I/O statistics section reported the information used to isolate specific objects that would benefit from being placed on SSDs. The segments with the most logical reads and physical reads are presented in Tables 2 and 3. These segments should be considered as possible candidates to be placed on SSDs.

For our specific example, the three indexes involving the largest number of reads (C_ORDER_LINE_I1, C_ORDER_I1, and C_STOCK_I1) were selected to



* Oracle Cluster Registry (OCR) and Cluster Synchronization Services (CSS) Voting Disk.

Figure 5 - Two Node Oracle 11g R1 RAC Database Architecture for Benchmark Factory TPC-C Testing

move to SSD. The total size of the three indexes is approximately 12 GB, which is about 10% of the total schema size.

As shown in Figure 5, two volumes were created in the SSD storage pool residing in the PS6000S array:

- The 100 GB volume was created on the PS6000S array and formatted as an ASM diskgroup. Then the online redo log files, undo datafiles, and temporary files were moved from the PS6000XV diskgroup to the newly created diskgroup.
- Another 80 GB volume was created and formatted as an ASM diskgroup. The three read intensive indexes identified above were then moved to this 80 GB ASM diskgroup.

The following TPC-C test runs were conducted:

1. Run 1: All database files reside in one PS6000XV.
2. Run 2: Online redo log files, undo tablespaces, temporary tablespaces and three indexes were moved to PS6000XV, with the rest of the database residing on a second PS6000XV array within its own pool.
3. Run 3: Online redo log files, undo tablespaces, temporary tablespaces and three indexes were moved to PS6000S, with the rest of the database residing in PS6000XV.

Normalized results from the above runs are graphed in Figures 6 and 7.

The blue line in Figures 6 and 7 represents the results from Run 1, in which all Oracle files reside on one PS6000XV. The red line in these Figures represents the results from Run 2, in which the temporary tablespaces, undo tablespaces, online redo log files, and three read-intensive indexes reside on one PS6000XV while the rest of the Oracle files/datasets reside on the second PS6000XV. Finally, the green line in these Figures represents the results from Run 3, in which the temporary tablespaces, undo tablespaces, online redo log files, and three read-intensive indexes reside on the PS6000S and the rest of the Oracle files/datasets reside on the PS6000XV.

From Figure 6 we can conclude the following:

Isolating and moving read-intensive datasets in Oracle OLTP applications from SAS disks in a PS6000XV to SSD disks in an additional PS6000S, as compared to

Owner	Tablespace Name	Object Name	Object Type	Physical Reads	%Total
QUEST	QUESTDATA	C_ORDER_LINE_I1	INDEX	32,623,712	57%
QUEST	QUESTDATA	C_ORDER_I1	INDEX	11,026,784	19%
QUEST	QUESTDATA	C_STOCK_I1	INDEX	6,391,040	11%
QUEST	QUESTDATA	C_STOCK	TABLE	1,804,912	3%
QUEST	QUESTDATA	C_ITEM_I1	INDEX	970,832	1%

Table 2: Segments by Logical Reads

Owner	Tablespace Name	Object Name	Object Type	Physical Reads	%Total
QUEST	QUESTDATA	C_ORDER_I1	INDEX	2,758,009	45%
QUEST	QUESTDATA	C_STOCK_I1	INDEX	1,635,640	27%
QUEST	QUESTDATA	C_STOCK	TABLE	787,727	13%
QUEST	QUESTDATA	C_CUSTOMER	TABLE	142,056	2%
QUEST	QUESTDATA	C_ORDER	TABLE	123,568	2%

Table 3: Segments by Physical Reads

SAS disks in an additional PS6000XV, can improve application transaction response times by up to 60% depending on user load

From Figure 7 we can conclude the following:

Isolating and moving read-intensive datasets in Oracle OLTP applications from SAS disks in a PS6000XV to SSD disks in an additional PS6000S, as compared to SAS disks in an additional PS6000XV, can increase application transaction throughput by up to an estimated 75% depending on response time.

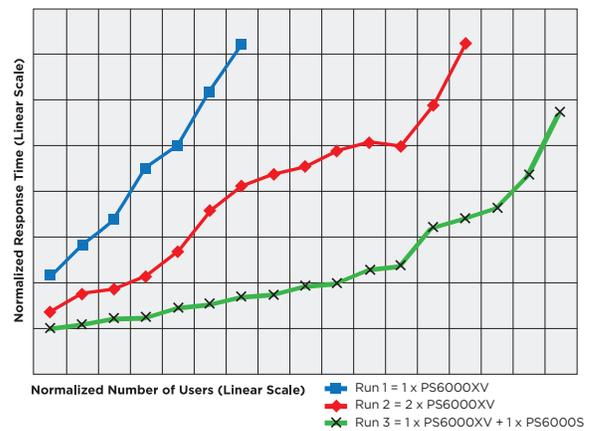


Figure 6 - Transaction Response Time vs. User Load

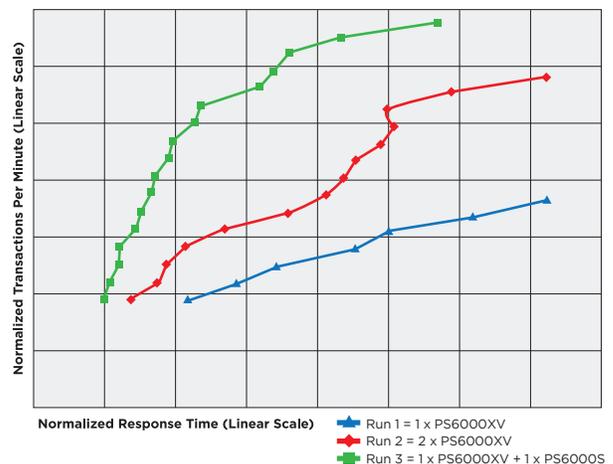


Figure 7 - Transactions Per Minute vs. Response Time

CONCLUSION

An exponential increase in CPU power has shifted the bottleneck of many performance-hungry application systems to disk I/O, as disk performance improvements have not been able to keep up with CPU speeds. With Tier-0 SSD-based arrays in an EqualLogic iSCSI SAN, depending on user load, latency-sensitive OLTP database applications can gain up to an estimated 75% increase in transaction throughput or up to a 60% decrease in response times by isolating and moving read-intensive datasets from SAS disks to SSDs as compared to moving the same datasets to additional SAS disks.

REFERENCES

1. "PS Series Architecture", a Dell EqualLogic Whitepaper <http://www.equallogic.com/resourcecenter/assetview.aspx?id=4711>
2. "Deploying Oracle Database on Dell EqualLogic PS5000XV iSCSI Storage", a Dell Technical Whitepaper. http://www.dell.com/downloads/global/solutions/oracle_ps5000xv_ref_config.pdf?c=us&cs=555&l=en&s=biz
3. "This is your database on Flash", Insights from Oracle <http://www.oracle.com/technology/deploy/performance/pdf/OracleFlash15.pdf>

SIMPLIFY YOUR STORAGE AT WWW.DELL.COM/PSSeries



EQUALLOGIC

THIS WHITE PAPER IS FOR INFORMATIONAL PURPOSES ONLY, AND MAY CONTAIN TYPOGRAPHICAL ERRORS AND TECHNICAL INACCURACIES. THE CONTENT IS PROVIDED AS IS, WITHOUT EXPRESS OR IMPLIED WARRANTIES OF ANY KIND.

www.BDTIC.com/SAMSUNG

© 2009 Dell Inc.
WP917_USA_040809