



Sizing and Best Practices for Online Transaction Processing Applications with Oracle 11g R2 using Dell EqualLogic PS6210S All Flash and PS6210XS Hybrid Arrays

A Dell EqualLogic Best Practices Technical White Paper

Dell Storage Engineering
August 2014

Revisions

Date	Description
August 2014	Initial release

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.

© 2014 Dell Inc. All rights reserved. Reproduction of this material in any manner whatsoever without the express written permission of Dell Inc. is strictly forbidden. For more information, contact Dell.

PRODUCT WARRANTIES APPLICABLE TO THE DELL PRODUCTS DESCRIBED IN THIS DOCUMENT MAY BE FOUND AT: <http://www.dell.com/learn/us/en/19/terms-of-sale-commercial-and-public-sector> Performance of network reference architectures discussed in this document may vary with differing deployment conditions, network loads, and the like. Third party products may be included in reference architectures for the convenience of the reader. Inclusion of such third party products does not necessarily constitute Dell's recommendation of those products. Please consult your Dell representative for additional information.

Trademarks used in this text:

Dell™, the Dell logo, Dell Boom!™, Dell Precision™, OptiPlex™, Latitude™, PowerEdge™, PowerVault™, PowerConnect™, OpenManage™, EqualLogic™, Compellent™, KACE™, FlexAddress™, Force10™ and Vostro™ are trademarks of Dell Inc. Other Dell trademarks may be used in this document. Cisco Nexus®, Cisco MDS®, Cisco NX-OS®, and other Cisco Catalyst® are registered trademarks of Cisco System Inc. EMC VNX®, and EMC Unisphere® are registered trademarks of EMC Corporation. Intel®, Pentium®, Xeon®, Core® and Celeron® are registered trademarks of Intel Corporation in the U.S. and other countries. AMD® is a registered trademark and AMD Opteron™, AMD Phenom™ and AMD Sempron™ are trademarks of Advanced Micro Devices, Inc. Microsoft®, Windows®, Windows Server®, Internet Explorer®, MS-DOS®, Windows Vista® and Active Directory® are either trademarks or registered trademarks of Microsoft Corporation in the United States and/or other countries. Red Hat® and Red Hat® Enterprise Linux® are registered trademarks of Red Hat, Inc. in the United States and/or other countries. Novell® and SUSE® are registered trademarks of Novell Inc. in the United States and other countries. Oracle® is a registered trademark of Oracle Corporation and/or its affiliates. Citrix®, Xen®, XenServer® and XenMotion® are either registered trademarks or trademarks of Citrix Systems, Inc. in the United States and/or other countries. VMware®, Virtual SMP®, vMotion®, vCenter® and vSphere® are registered trademarks or trademarks of VMware, Inc. in the United States or other countries. IBM® is a registered trademark of International Business Machines Corporation. Broadcom® and NetXtreme® are registered trademarks of Broadcom Corporation. Qlogic is a registered trademark of QLogic Corporation. Other trademarks and trade names may be used in this document to refer to either the entities claiming the marks and/or names or their products and are the property of their respective owners. Dell disclaims proprietary interest in the marks and names of others.



Table of contents

Revisions	2
Acknowledgements	5
Feedback	5
Executive summary	6
1 Introduction	7
1.1 Objective	7
1.1.1 Audience	8
1.2 Terminology	8
2 Solution infrastructure	10
2.1 Physical system test configuration	10
2.2 Storage configuration	11
2.3 Database layout	12
3 Test methodology	14
3.1 Measurement and monitoring	14
3.2 Test criteria	15
4 I/O profiling study	16
4.1 PS6210S - ORION (100% Random I/O)	17
4.2 PS6210XS - ORION (100% random I/O and no data locality)	18
4.3 PS6210XS - VDBench (100% random I/O and 20% data locality)	19
4.4 Number of volumes studies	21
4.5 Summary	22
5 OLTP performance studies using TPC-E like workload	23
5.1 PS6210S: TPC-E Characterization	24
5.2 PS6210XS: TPC-E characterization	25
5.3 Summary	27
5.3.1 Oracle 11g RAC OLTP on PS6210S all-flash array	27
5.3.2 Oracle 11g RAC OLTP on PS6210XS hybrid array	27
6 Best practice recommendations	29
6.1 Storage	29
6.2 Network infrastructure	29
6.3 Operating system	29



6.4	Oracle database application	30
6.4.1	Database volume layout	30
6.4.2	Oracle AMM	30
6.4.3	Oracle ASM configuration on RHEL 6.3	32
7	Conclusion.....	33
A	Configuration details	34
	Additional resources	36



Acknowledgements

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

Engineering: Chidambara Shashikiran

Technical Marketing: Chhandomay Mandal

Editing: Margaret Boeneke

Feedback

We encourage readers of this publication to provide feedback on the quality and usefulness of this information by sending an email to SIFeedback@Dell.com.



SIFeedback@Dell.com



Executive summary

Online transaction processing (OLTP) applications, like web-based e-commerce sites, accounting systems, and customer support programs, are at the heart of today's business functions. As the foundation for a wide range of mission-critical applications, OLTP operations, particularly those using Oracle real application cluster (RAC) database platforms, depend on exceptional storage performance and reliability.

Dell EqualLogic PS6210 series arrays offer scalable, high-performance virtualized storage that is designed for reliability, manageability, and efficiency. The PS6210S series arrays offer flash at the price of a disk which can effectively manage peak performance requirements of the mission critical OLTP database applications.

Moreover, data generated by OLTP applications can grow exponentially over time. Thus, designing and managing storage that can effectively accommodate performance, capacity and future growth requirements for Oracle OLTP databases is a key challenge today.

Dell EqualLogic PS Series hybrid storage arrays that combine solid state drives (SSDs) and hard disk drives (HDDs) within the same chassis with automated, dynamic data tiering are well suited for these OLTP applications. This automated internal data tiering within the array creates a flexible and powerful balance between performance and responsiveness provided by the low latency SSD tier and capacity provided by the HDD tier.

This paper presents sizing guidelines and best practices for deploying Oracle 11g R2 OLTP databases with EqualLogic PS6210S all-flash array and PS6210XS hybrid arrays based on a series of storage I/O performance test results. A variety of OLTP benchmarking tools including Oracle I/O Numbers (ORION), VDBench, and TPC-E were run to demonstrate the efficacy of the PS6210S and PS6210XS array for the OLTP workloads. Key findings include:

- PS6210S all-flash arrays delivered 62,400 IOPS with disk latency less than 2ms while simulating TPC-E like I/O from ORION. These arrays were able to deliver more than 48,000 IOPS even when more than 1,000 concurrent users performed heavy TPC-E style database transactions on the 2-Node RAC database.
- Highly effective, dynamic, automated data tiering by PS6210XS arrays helped to address the key challenge faced by customers of accommodating the exponential increase in capacity and also very strict SLAs related to performance of OLTP database environments (for example, ~300% IOPS increase over time due to tiering between SSDs and HDDs in VDBench test).
- Also the average disk latency on the storage arrays was less than 2ms even when the arrays were delivering the peak performance.

Additionally, this paper presents guidelines for designing Oracle OLTP environments including database volume layout configurations, best practices to be followed while configuring networking and operating system (OS) parameters, and also Oracle database configuration settings.



1 Introduction

Different types of database applications have varying performance and capacity needs. Understanding the models for common database application workloads can be useful in predicting the possible application behavior in a given infrastructure environment. The most common database application workload models are OLTP and Data Warehousing (DW). This paper focuses on OLTP database workloads.

OLTP applications often need to deliver extremely fast response times and rapid access to data that is in high demand. Moreover, significant data growth over time in a typical OLTP system leads to high capacity requirements as well. While the OLTP data set size can grow exponentially over time, the active, frequently-accessed data set size remains relatively small in comparison with the overall OLTP system size.

Storage infrastructure supporting OLTP platforms like Oracle RAC needs to non-disruptively scale both in performance and capacity and also deliver the required high performance in an optimal, cost-effective way. Dell EqualLogic PS6210S all-flash arrays and PS6210XS hybrid arrays, which combine SSDs and HDDs within a single chassis with automated data tiering, are well-suited for these OLTP environments.

OLTP database applications are optimal for managing rapidly changing data. These applications typically have many users who are performing transactions simultaneously. Although individual data requests by users usually reference few records, many of these requests are being made at the same time.

To develop appropriate storage performance and capacity sizing guidelines for OLTP workloads, it is essential to understand the following factors:

- Read I/O and Write I/O percentage
- I/O block sizes
- Data access rate
- Data locality
- Memory configuration on the database server (because this plays a significant role in Oracle database I/O performance)

This whitepaper describes various tests based on these factors and also key guidelines and best practices based on the test results to achieve optimal performance of Oracle 11g RAC OLTP database applications using PS6210 series arrays.

1.1 Objectives

The key objectives of testing were:

- Establish baseline I/O performance characteristics of the EqualLogic PS6210S and PS6210XS arrays when running OLTP-like I/O patterns using the ORION tool without deploying Oracle databases.
- Performance characterization using Benchmark Factory for Databases, simulating Oracle OLTP transactions on the 2-Node Oracle RAC database hosted on PS6210S and PS6210XS arrays with a TPC-E type workload.



- I/O profiling with VDBench to demonstrate performance improvements with the automated data tiering of EqualLogic PS6210XS due to locality of data access in Oracle OLTP environments.
- Determine guidelines and best practices for deploying and achieving optimal performance of Oracle 11g R2 OLTP databases on EqualLogic PS6210 series arrays.

1.1.1 Audience

This paper is intended for system administrators, database administrators and storage architects interested in deploying Oracle OLTP database solutions on Red Hat Enterprise Linux (RHEL) using Dell EqualLogic PS Series SANs. It is assumed that the readers of this document have familiarity with Red Hat Enterprise Linux system administration, EqualLogic storage array configuration and Oracle 11g R2 database installation and administration tasks.

1.2 Terminology

The following terms are used throughout this document.

AMM: The Oracle Automatic Memory Management (AMM) feature automates and simplifies the memory tuning configuration tasks by allowing you to manage the different memory components of Oracle using a single database initialization parameter called "MEMORY_TARGET".

ASM: Oracle Automatic Storage Management (ASM) is a volume manager and a file system that supports Oracle database.

Group: One or more EqualLogic PS Series arrays connected to an IP network that work together to provide SAN resources to host servers.

Member: A single physical EqualLogic array.

OLTP I/O pattern: OLTP workloads tend to select a small number of rows at a time. These transfers happen anywhere in the data, and are each typically between 8K and 64K. This causes the I/O pattern to be random in nature¹. The key metric in measuring performance of OLTP workloads is the number of I/Os per second (IOPS) that can be achieved while maintaining a healthy response time.

ORION: ORION is a popular I/O simulation tool used for understanding the I/O performance behavior of storage systems for Oracle databases. The tool and documentation is included as part of the Oracle 11g installation software.

Quest Benchmark Factory: A database performance tool that can simulate actual database transactions such as TPC-C and TPC-E according to the TPC specifications. More details on this tool and the download location can be found at <http://www.quest.com/benchmark-factory/>.

¹ DW workloads, primarily composed of running database queries and generating reports, are sequential in nature. If the same database is used for both transaction processing and report generation, a mix of OLTP and DW workloads will be observed.



SAN HQ: SAN Headquarters (SAN HQ) monitors one or more PS Series groups. The tool is a client/server application that runs on a Microsoft Windows system and uses simple network management protocol (SNMP) to query the groups. Much like a flight data recorder on an aircraft, SAN HQ collects data over time and stores it on the server for later retrieval and analysis. Client systems connect to the server to format and display the data in the SAN HQ GUI.

TPC-E: A benchmark developed by TPC to simulate OLTP database workloads. The TPC-E benchmark simulates the OLTP workload of a brokerage firm. Although the underlying business model of TPC-E is a brokerage firm, the database scheme, data population, transaction mix and implementation rules broadly represent a modern OLTP database system. More information on TPC-E can be found at <http://www.tpc.org/tpce/default.asp>.

VDBench: VDBench is a disk and tape I/O workload generator specifically created to help engineers and customers generate disk I/O workloads to be used for validating storage performance and storage data integrity. The tool can be downloaded at the following link: <http://www.oracle.com/technetwork/server-storage/vdbench-downloads-1901681.html>



2 Solution infrastructure

The test system used to conduct ORION and TPC-E testing for this paper is shown in Figure 1 and Figure 2.

2.1 Physical system test configuration

The logical test configuration diagram for simulating Oracle OLTP database I/O using the ORION tool is shown below.

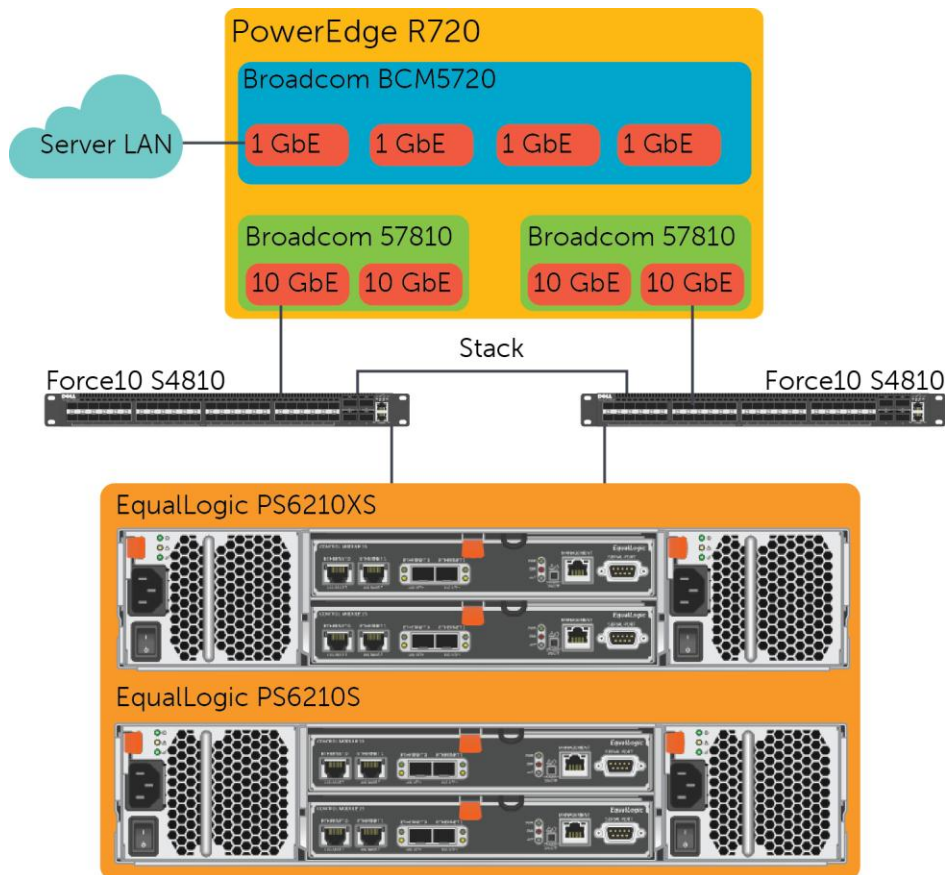


Figure 1 ORION Test configuration - LAN and iSCSI SAN connectivity

The logical test configuration diagram for OLTP database application tests (TPC-E) using Quest Benchmark Factory is shown below.

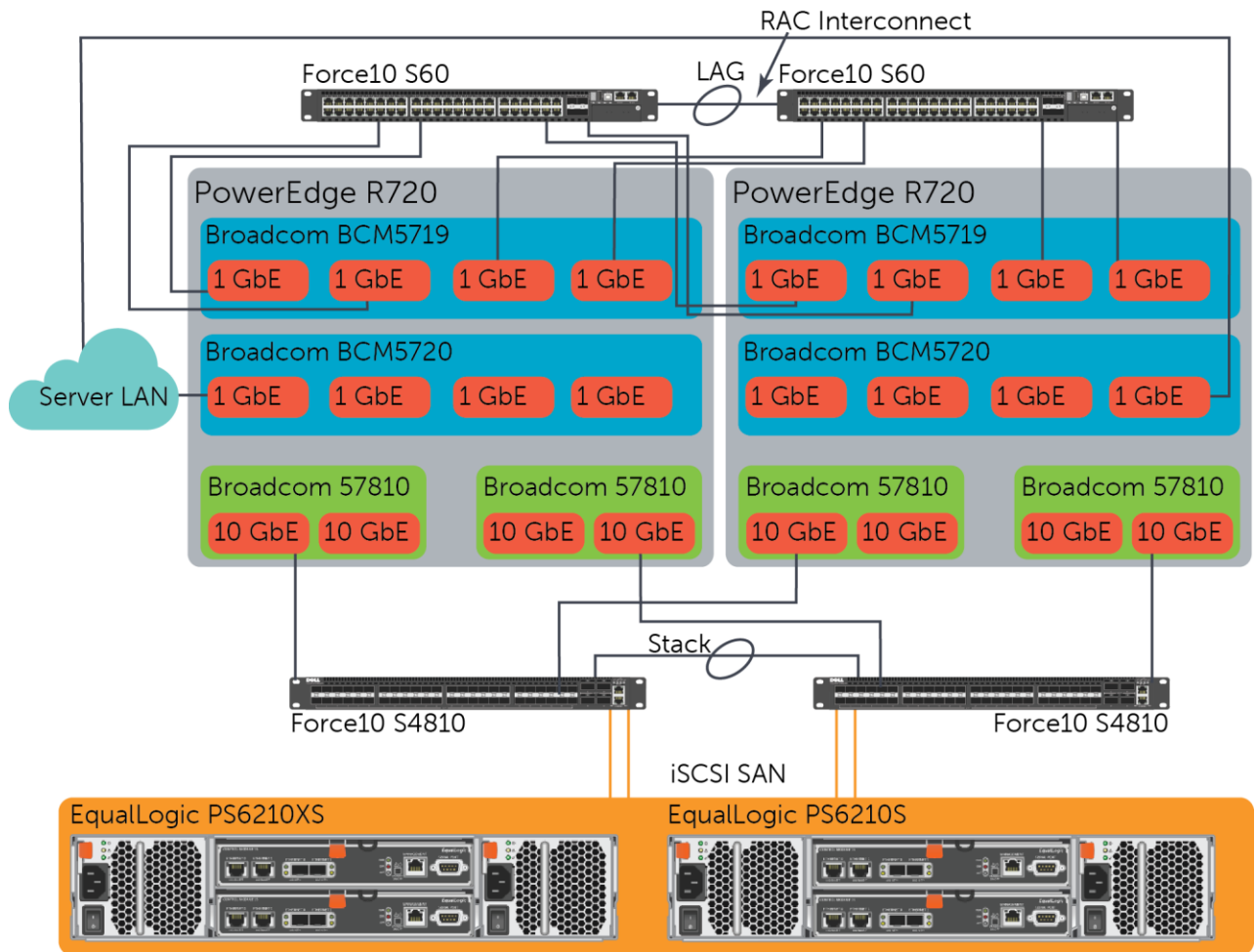


Figure 2 TPC-E test configuration - LAN and iSCSI SAN connectivity

2.2 Storage configuration

The EqualLogic arrays listed below were used for validating the solution.

- **EqualLogic PS6210S**
 - 24 x 800 GB SSDs as RAID 10, with two hot spares
 - 10 GbE controller running firmware version 7.0.4
- **EqualLogic PS6210XS**
 - 7x 800 GB SSDs and 17 x 1.2TB 10K SAS drives as RAID 6 (Accelerated), with two hot spares
 - 10 GbE controller running firmware version 7.0.4
- **EqualLogic PS6210E**
 - 24 x 2 TB SATA drives as RAID 50, with two hot spares

- 10 GbE controller running firmware version 7.0.4

The database files, REDO logs and system related table spaces such as SYSTEM and UNDO were stored on either a PS6210S or PS6210XS array in its own pool. The flashback recovery and archive data was stored on a PS6210E array which was in a separate pool named "BACKUP".

To enable advanced MPIO functionality, the EqualLogic Host Integration Tools for Linux (HIT/Linux) was installed on the RHEL server with the following settings:

- Default connection parameters of two iSCSI sessions per volume slice on an array
- Maximum of six iSCSI sessions for the entire volume
- MPIO default load balancing policy and I/Os per path
- The *eqltune* utility, installed as part of the HIT/Linux package, was run and its OS parameter recommendations were implemented

2.3 Database layout

The EqualLogic arrays were placed in different pools, and the following Oracle ASM disks were configured using the ASM disk layout described below.

DB1DATA	Database files; temporary table space; online REDO logs; system related table spaces such as SYSTEM and UNDO
DB1LOG1/DB1LOG2	REDO logs
DB1FLASH	Archive logs and Flash recovery area
OCRVOTE	ASM Cluster file system (ACFS) for storing clusterware related information such as the OCR and voting disks

Table 1 shows the ASM disk and diskgroup configuration.

Table 1 Oracle ASM disk and diskgroup configuration

DB1DATA	DB1LOG1	DB1LOG2	DB1FLASH	OCRVOTE
ORADB1 (250GB) ORADB2 (250GB) ORADB3 (250GB) ORADB4 (250GB) ORADB5 (250GB) ORADB6 (250GB)	ORALOG1 (100GB)	ORALOG2 (100GB)	ORAFLASH1 (1.5 TB) ORAFLASH2 (1.5 TB)	ORACRS (20GB)

The following figure shows the containment model and relationships between the EqualLogic pool and volumes and the Oracle ASM disk groups and disks.



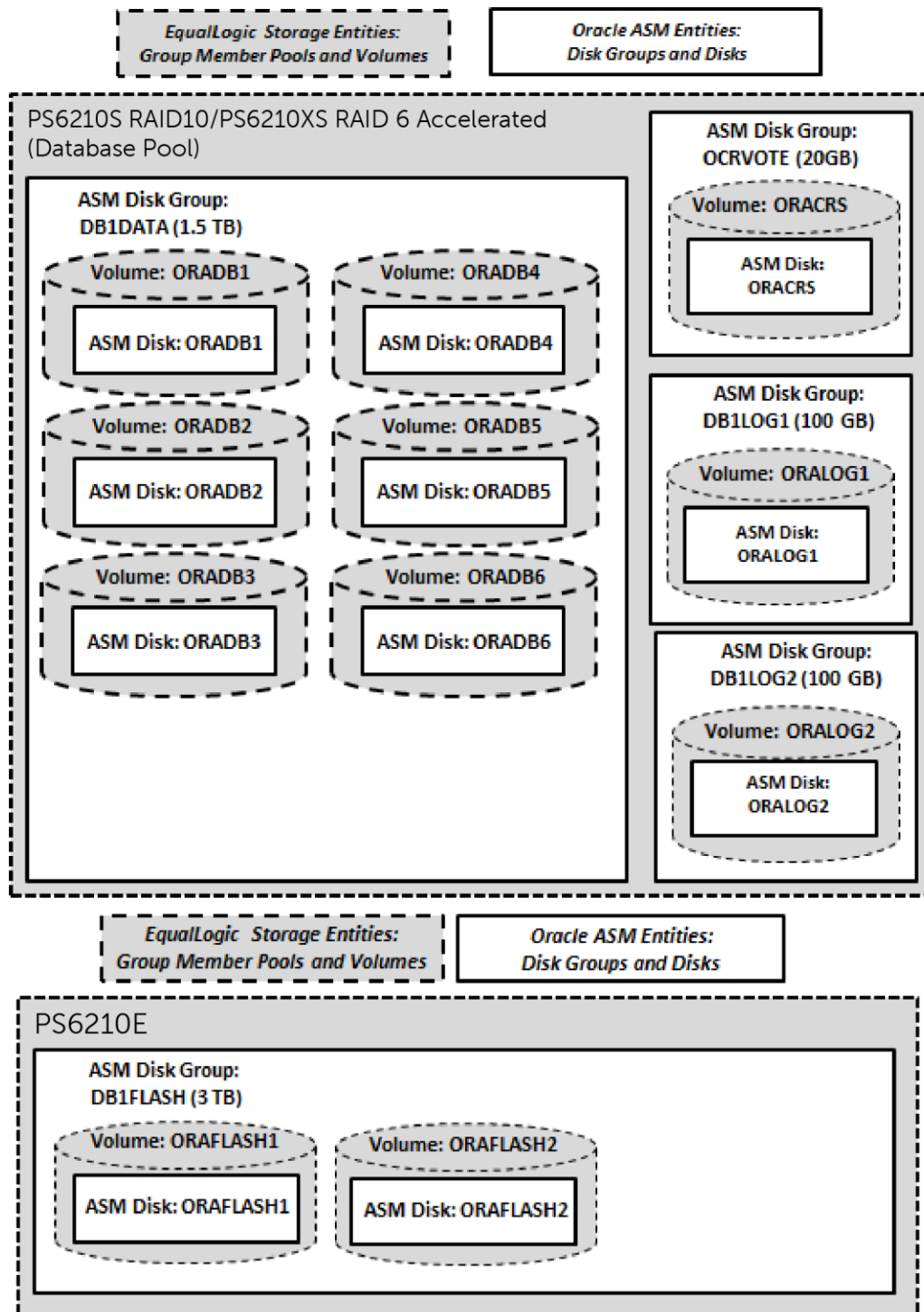


Figure 3 EqualLogic volume and Oracle ASM disk configuration

3 Test methodology

A series of I/O simulations were conducted using ORION to understand the performance characteristics of the PS6210S all-flash and PS6210XS hybrid storage arrays.

The ORION tool was installed on the OS (Red Hat Enterprise Linux in this case) and used to simulate I/O on the external EqualLogic storage arrays. The EqualLogic storage volumes were presented to the OS as block I/O devices. I/O simulations were conducted with small sized 8K I/O block requests simulating Oracle OLTP database I/O to understand the IOPS offered by the storage arrays. The VDBench I/O tool was used to evaluate the performance characteristics of the PS6210XS array when a specific locality of data on the disks was frequently accessed, which is observed in a typical database environment.

OLTP I/O is typically represented as 90% read/10% writes, specifically for workloads simulating financial organizations like banks and brokerage firms (More detailed information about TPC-E characteristics is available at: <http://www.cs.cmu.edu/~chensm/papers/TPCE-sigmod-record10.pdf>).

The I/O test simulating TPC-E style I/O workload was executed using the ORION tool to determine the baseline performance and the actual database performance characterization was performed by executing TPC-E style database transactions using Quest Benchmark Factory.

3.1 Measurement and monitoring

Detailed performance metrics were captured from the storage, OS and database application layers as listed below.

- **ORION**

Total IOPS and average latency as part of ORION standard test results were captured.

- **Quest Benchmark Factory**

The following metrics were captured from Quest Benchmark Factory while executing TPC-E database transactions.

- **TPS:** Transactions per second, based on the simulated load.
- **Response time:** End-to-end SQL execution time, measured from SQL submission until the end of the fetch of the SQL result set.
- **Transaction time:** Average time to run a single transaction.

- **EqualLogic SAN HeadQuarters (SAN HQ)**

EqualLogic storage array statistics for various objects including network, disk, I/O performance, and I/O capacity were captured. A dedicated server was used to run this tool.

- **Oracle AWR reports**

Oracle Automatic Workload Repository (AWR) reports were collected at one minute intervals while executing TPC-E transactions. The top five events by total wait time were analyzed to determine if there were any bottlenecks related to RAC, network or other factors.



3.2 Test criteria

The test criteria used for the study was:

- The storage array disk access latencies (read and write) remain below 20 ms per volume
- Database server CPU utilization remains below an 80% average
- TCP retransmits on the storage network remain below 0.5%
- Application response times remain below two seconds, which is the industry standard acceptable latency for Oracle OLTP database applications.
- No RAC related bottlenecks (such as database locking, user and system I/O wait, RAC interconnect response, or others) while executing the peak I/O load



4 I/O profiling study

I/O profiling tests were run using ORION on the PS6210S all-flash array to establish baseline I/O performance characteristics while simulating a TPC-E like I/O workload.

The storage array performance in a specific OLTP environment also depends on factors such as the amount of data stored on the array, frequency of access and most importantly, the locality of the data referenced.

Next, I/O profiling tests were run using ORION on PS6210XS arrays to establish baseline I/O performance characteristics of the hybrid storage array under test. An I/O test using VDBench evaluated the performance benefits offered by PS6210XS arrays due to automated tiering.

The tests listed in the table below were run with 8K I/O block size and at increasing queue depths (number of outstanding I/Os) to determine the maximum IOPS the storage array would sustain within the 20 ms latency (read and write latencies measured separately).

The configuration parameters for the tests are shown in Table 2.

Table 2 Test parameters: I/O workload studies

Configuration Parameters	
EqualLogic SAN	One PS6210XS (7x 800GB SSDs and 17 x 1.2TB 10K SAS drives) One PS6210S (24x 800GB SSDs as RAID 10)
Test configuration #1 PS6210S - ORION (100% Random I/O)	ORION I/O mix: 8K 90% read / 10% write (Simulate TPC-E)
Test configuration #2 PS6210XS - ORION (100% random I/O and no data locality)	ORION (100% Random I/O locality) I/O mix: 8K 90% read / 10% write (Simulate TPC-E)
Test configuration #3 PS6210XS - VDBench (100% random I/O and 20% data locality)	VDBench (20% random I/O locality) I/O mix: 90% read / 10% write (Simulate TPC-E I/O with 20% I/O locality)

The VDBench I/O generation tool was used for simulating 20% I/O locality test as observed in a typical database environment. The ORION tool was not appropriate for this study because it does not provide an option to specify locality of data reference.

The results collected from the above tests are illustrated in the following sections. Additional I/O simulation tests were run to determine the volume configuration which produces optimal performance. These test results are discussed in Section 4.4 and all the key findings are summarized in Section 4.5.



4.1 PS6210S - ORION (100% random I/O)

ORION tests were run with a range of user loads using eight 250 GB volumes. More than 50% of the array capacity was utilized and the I/O access was completely random in nature (100% random with no locality of I/O access).

This test on the array produced approximately 62,400 IOPS for an 8K block size with a 90%/10% read/write mix workload while the disk latency remained less than 2 ms (for both read and write IOPS measured separately).

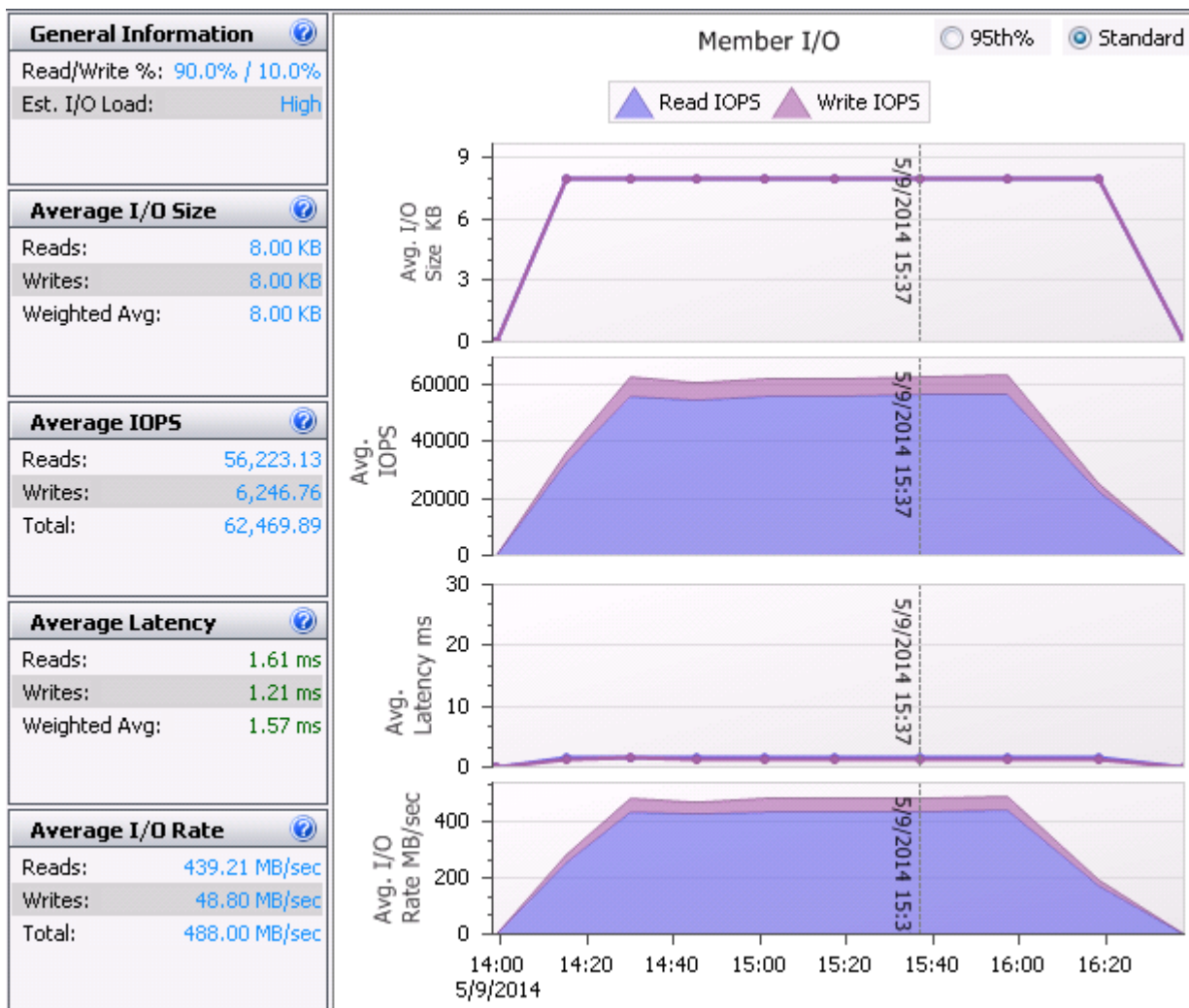


Figure 4 PS6210S - IOPS and latency reported by SAN HQ

The raw I/O performance from this simulation test is compared with the I/O performance when the TPC-E like workload was run on the actual database in Section 5.1.

4.2 PS6210XS - ORION (100% random I/O and no data locality)

In this test, eight 500 GB volumes were used to run the ORION test. The simulated I/O was completely random with no locality of data access. This test on the array produced approximately 13,000 IOPS for an 8K block size with a 90/10 read/write mix workload while staying within the generally accepted disk latency limit of 20 ms (for both read and write IOPS measured separately).

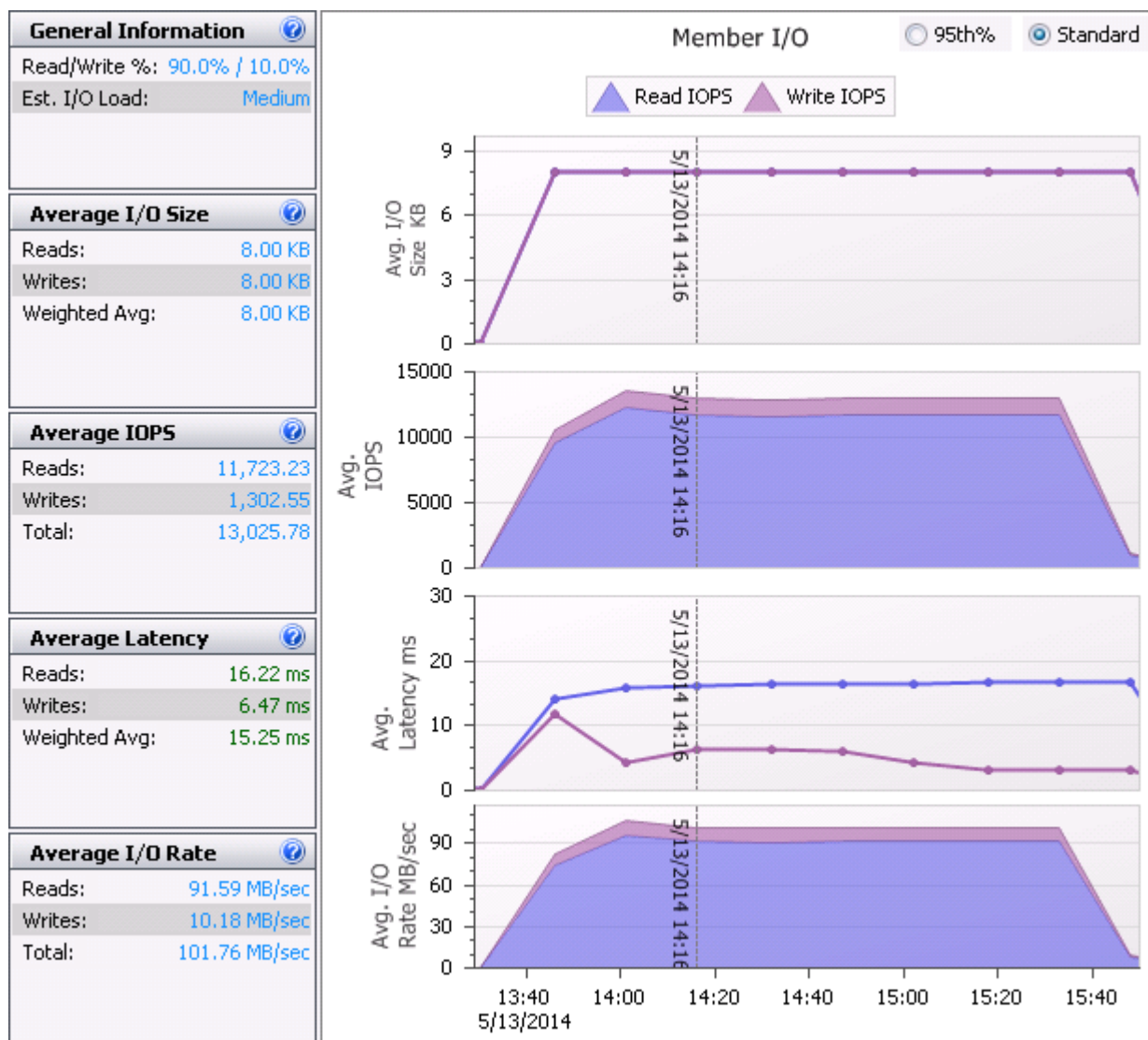


Figure 5 PS6210XS - IOPS and latency reported by SAN HQ (No data locality)

The total capacity of the volumes utilized for this test was more than the entire usable capacity of SSDs on PS6210XS storage array. Also the I/O simulated using ORION for this test was completely random in nature with no locality of data. As a result, I/O was completely spread across both SSD and 10K SAS drives and it is shown in the chart below.

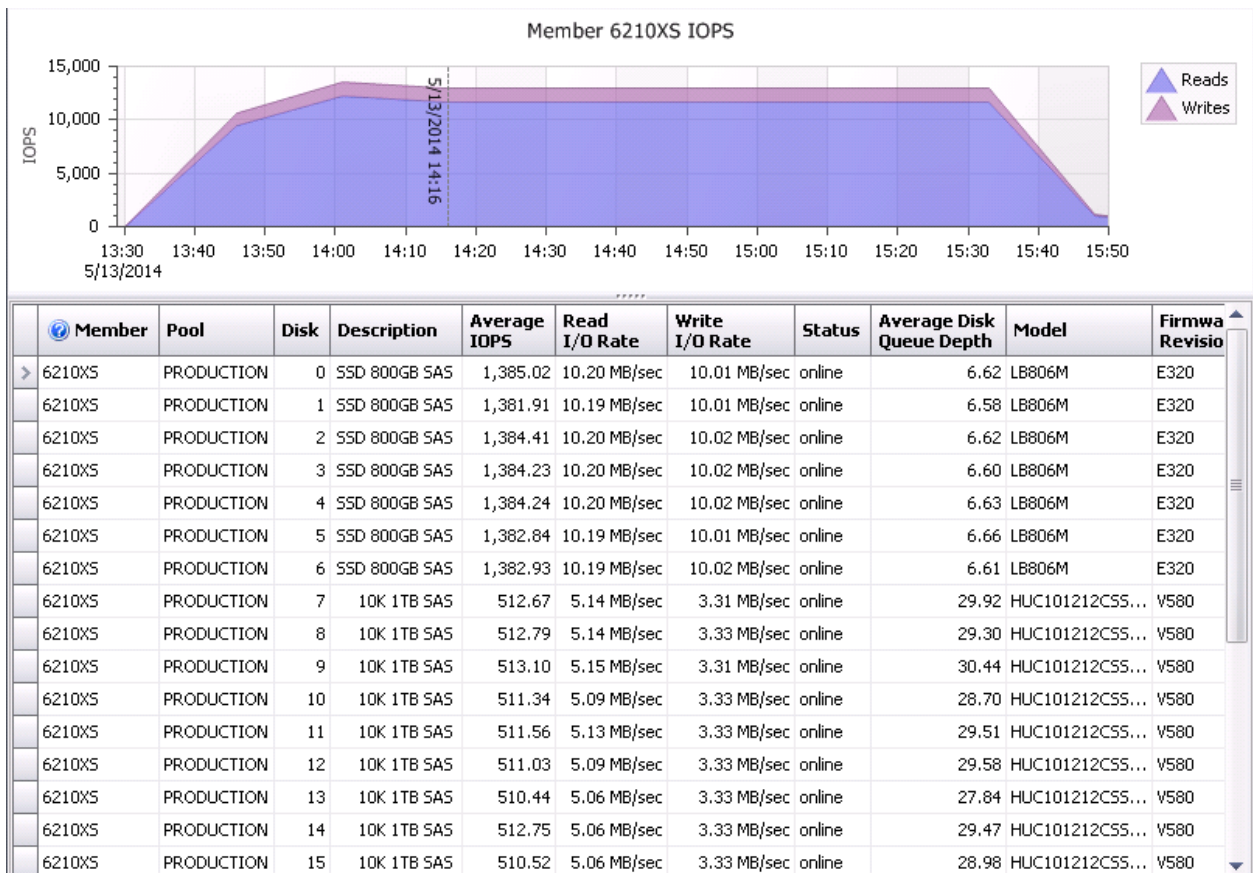


Figure 6 PS6210XS – I/O activity on disks (No data locality)

The raw I/O performance from this simulation test is compared with the I/O performance when the TPC-E like workload was run on the actual database in Section 5.2.

4.3 PS6210XS - VDBench (100% random I/O and 20% data locality)

Applications such as Oracle OLTP databases typically contain both infrequently accessed and highly dynamic data within a single volume. The advanced tiering feature on PS6210XS arrays categorizes and operates on data workloads within each volume.

These workloads are automatically categorized and placed as appropriate onto either SSD or conventional hard-drive tiers based on the performance requirements on a PS6210XS array. Typically, in database applications, there are always specific areas on the disks such as REDO logs which are accessed more frequently than the other database objects. These frequently accessed objects usually need higher performance than other database objects.

In order to test the I/O scenario with a specific locality and frequency of I/O access, the VDBench I/O generation tool was used to simulate I/O. The ORION tool was not appropriate for this study because



there is no option to specify that the I/O workload access a specific locality of data on the disk. A VDBench test was run with 20% locality of reference to evaluate the tiering benefits.

The test result showed approximately 11,800 IOPS at the beginning and then gradually increasing to 46,500 over the test timespan. As the test was run, specific portions of the disks were accessed more frequently, and as a result, more hot pages got created and automatically moved to the SSD tier. This resulted in significant performance improvement.

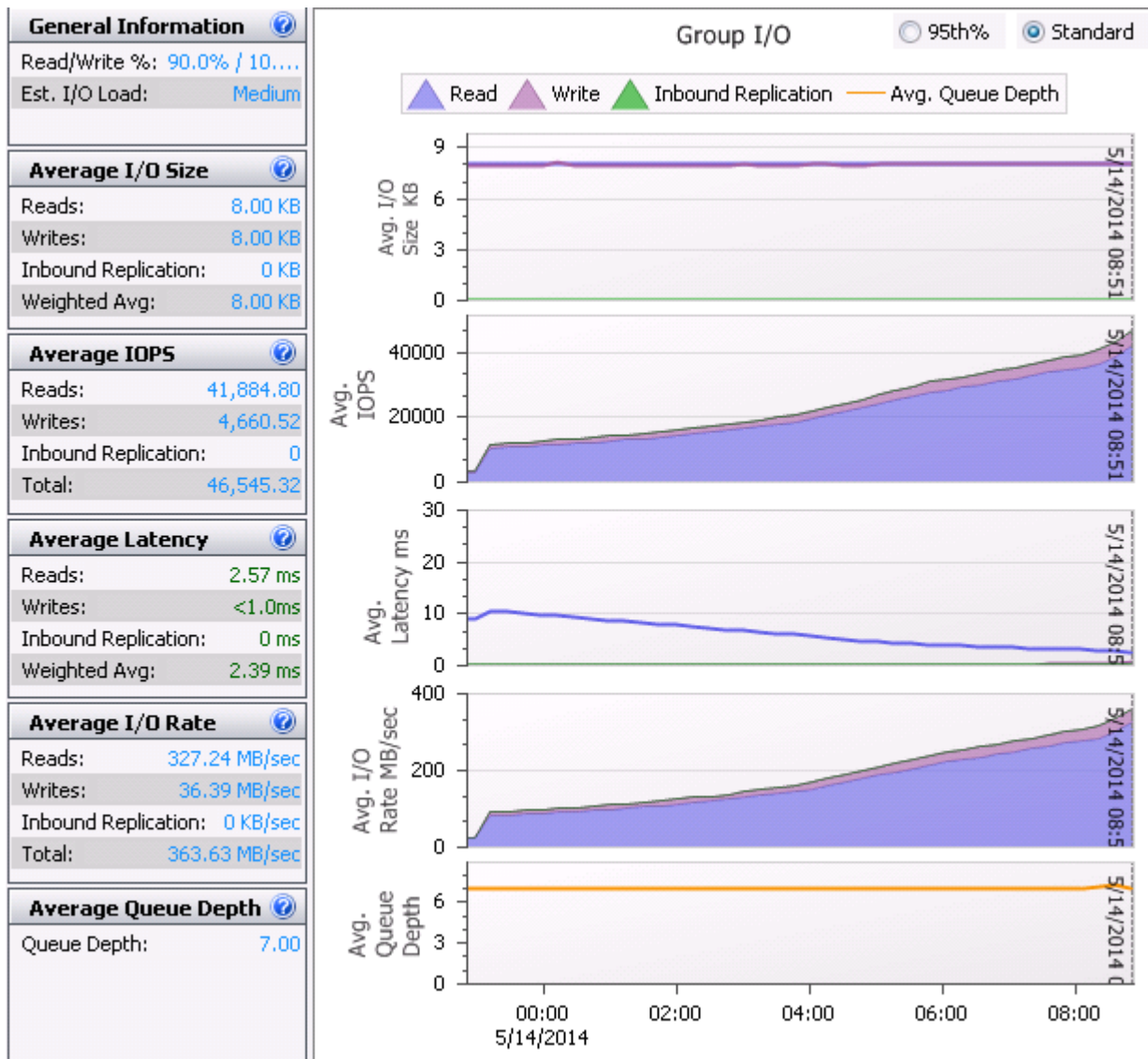


Figure 7 Increase in IOPS due to PS6210XS internal tiering (As reported by SAN HQ)

The SAN HQ chart above shows the gradual increase in IOPS from 11,800 to 46,500 over the test timespan due to the automatic tiering feature of PS6210XS arrays.

4.4 Number of volumes studies

The primary objective of these tests was to determine the EqualLogic volume configuration which produces the optimal performance for the Oracle OLTP type of workload.

These tests were run with 8K I/O block size to determine the maximum IOPS the storage array would sustain within the 20 ms latency (read and write latencies measured separately). Several tests were run using the same array capacity with a varying number of volumes to determine the optimal volume configuration.

The volume configurations used for the tests are shown in the following table.

Table 3 Test parameters: Volume configuration studies

Configuration Parameters	
Volume configuration #1	Two volumes, 1 TB each
Volume configuration #2	Four volumes, 500 GB each
Volume configuration #3	Eight volumes, 250 GB each

There was a more than 6% increase in performance when the number of volumes was increased from two to four. Similarly, almost a 23% increase in performance was observed when eight volumes were configured on the array.

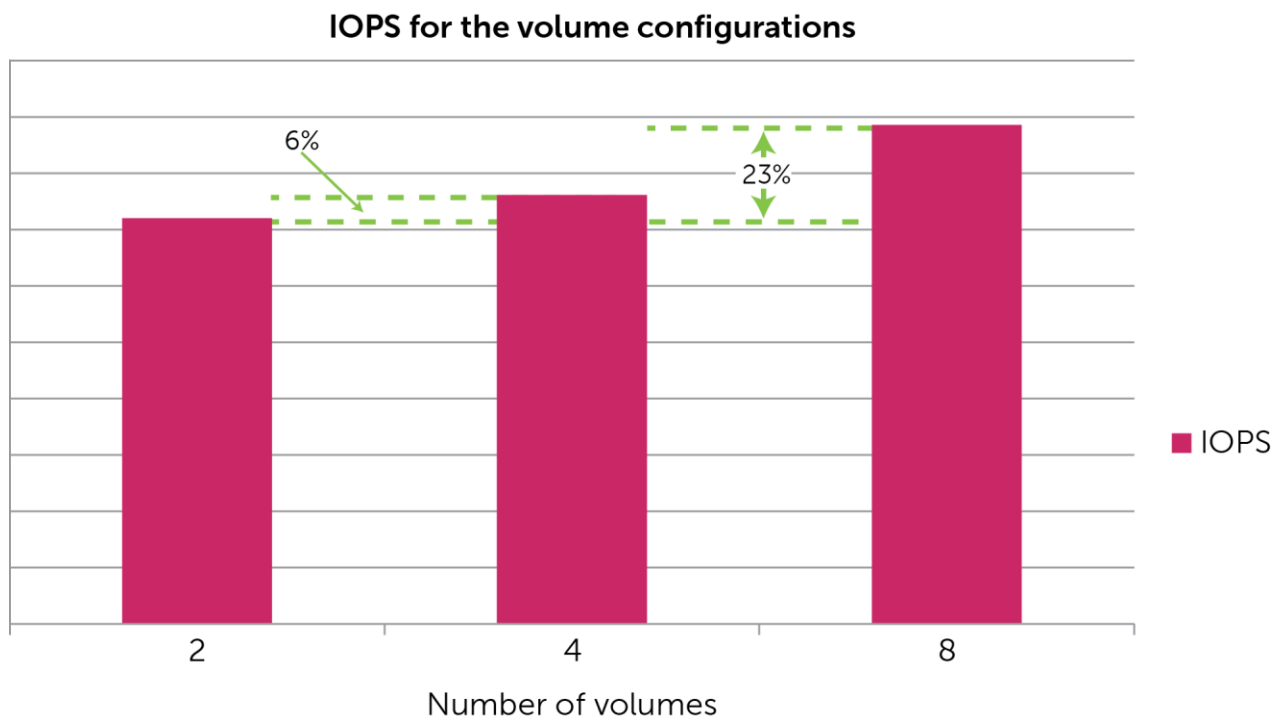


Figure 8 IOPS for each Volume configuration

These results confirm higher performance when there are eight or more volumes. Also, increasing the number of volumes beyond eight did not result in any significant increase in IOPS based on the test results.

A configuration with more small volumes is preferred over a low number of large volumes for better performance and simpler management of OLTP databases. As illustrated in Figure 88, there was 23% improvement in performance when the number of volumes was increased from 2 to 8.

4.5 Summary

The tests described in Section 4.1 clearly show the significant performance benefits from the PS6210S all-flash array where more than 62,000 IOPS were achieved with less than 2 ms disk latency on the storage array.

The test results in Section 4.2 clearly show that the array tries to use the SSDs first to provide the best performance. If the entire database does not fit into SSDs, then the data is spread across both SSDs and 10K SAS drives.

However in typical OLTP database applications, most of the active data set usually accessed frequently is contained within a relatively small locality on disks. In these scenarios, the internal tiering feature of PS6210XS will identify the frequently accessed data and move the hot pages automatically to SSDs. This can create significant performance improvements.

To illustrate the unique tiering benefits of PS6210XS arrays, the VDBench I/O generation test was executed as explained in Section 4.3. The test was run to simulate 8K block sized I/Os with 90%/10% read/write mix and 20% locality of I/O access. When a specific locality of I/O got accessed frequently, it was clear that the PS6210XS array was providing tiering benefits and the IOPS increased from 11,800 to 46,500 (refer to Section 4.3 for more details).

Additional tests were run to determine the volume configuration which produces optimal performance. Test results from Section 4.4 confirmed that using eight or more volumes per PS6210 series array member is recommended for achieving optimal performance for Oracle OLTP database workloads.



5 OLTP performance studies using TPC-E like workload

ORION and VDBench are synthetic I/O tools that provide an estimate of Oracle OLTP performance without actually installing Oracle database, whereas TPC-E provides estimates of Oracle OLTP performance by executing the workload on actual Oracle databases.

The ORION and VDBench test results described in Section 4 helped in determining the baseline performance of a typical Oracle OLTP database on the PS6210S and PS6210XS arrays. This section discusses Quest Benchmark Factory tests for databases simulating a real TPC-E style Oracle OLTP database workload using an actual 2-Node Oracle 11g R2 RAC OLTP database.

Some of the key configuration details for the database are described below:

- Used a single PS6210S/PS6210XS array for hosting Oracle database and REDO log files. One more EqualLogic PS6210E array was used to configure the archive and flashback recovery area.
- Installed Oracle 11g R2 clusterware on the two Dell PowerEdge R720 “database” servers and the logical test configuration is described Section 2.1.
- The database elements, REDO logs, archive and flash data were managed using Oracle ASM. More details on disk layout are available in Section 2.3.
- A TPC-E like database was populated with a scaling factor of 40 using Quest Benchmark Factory.

The TPC-E tests were also run on two storage array configurations (PS6210S and PS6210XS) to evaluate the performance benefits. The two test configuration details are listed in the table below.



Table 1 Test parameters: Database file and volume layout studies

Configuration Parameters	
EqualLogic SAN	<p>One PS6210S (24x 800 GB SSDs) RAID 10 - Hosting the Database files and REDO logs.</p> <p>One PS6210XS (7x 800 GB SSDs and 17x 1.2 TB 10K SAS drives) RAID 6 Accelerated - Hosting the Database files and REDO logs.</p> <p>One PS6110E (24x 2 TB SATA drives as RAID 50) – Hosting the ARCH logs and flash recovery area.</p>
Test configuration details	
Test configuration #4 PS6210S: TPC-E Characterization	TPC-E test with scaling factor of 40. Simulated 1,000 users using four agents.
Test configuration #5 PS6210XS: TPC-E characterization	TPC-E test with scaling factor of 40. Simulated 500 users using four agents.
Database Workload Parameters	
Database	2-Node RAC running on two R720 servers
User transactions	TPC-E from Benchmark Factory for Databases
Test duration	5 Hours
Database size	~900 GB (including tables, indexes)

The results from these two tests are described in the following Sections 5.1 and 5.2.

5.1 PS6210S: TPC-E characterization

In this test, the entire database (database files, redo logs and index) was hosted on the PS6210S array. The database was loaded with a TPC-E style database using Quest Benchmark Factory with a scaling factor of 40. The initial set of tests was executed with varying loads to determine the maximum users which could be achieved on this configuration while keeping storage array disk latency less than 20 ms and also the average response time less than 2 seconds.

It was determined that 1,000 users is the maximum user load for this configuration based on the initial test results and test criteria as explained in Section 3.2.

Quest Benchmark Factory software simulated TPC-E transactions with 1,000 users using four agents for an extended period of time to evaluate the storage array performance. The storage array produced 48,422 IOPS with less than 2 ms disk latency as shown in the SAN HQ chart below.



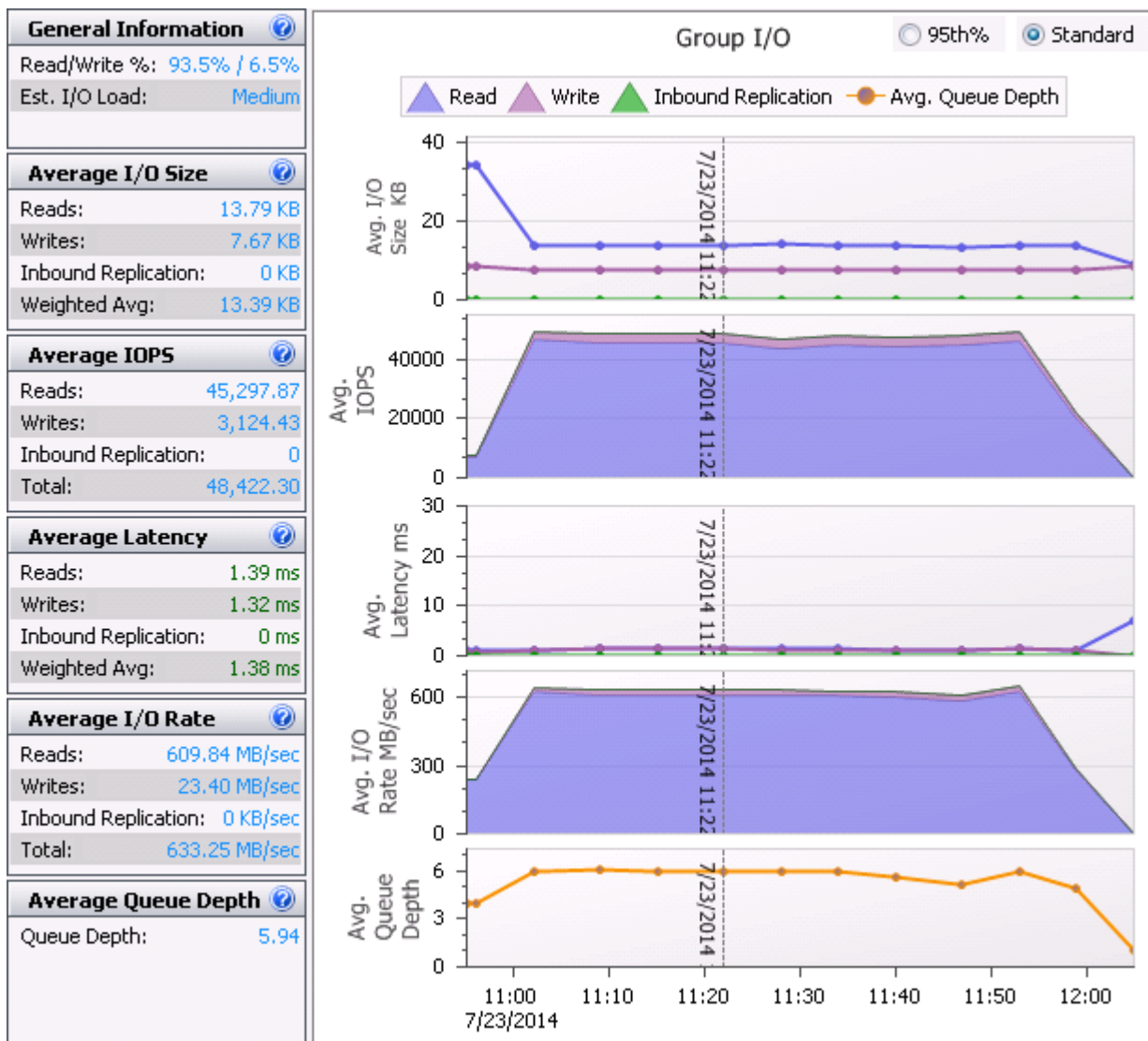


Figure 9 PS6210S TPC-E I/O (IOPS and latency reported by SAN HQ)

Quest Benchmark Factory simulated heavy real-world database transactions (TPC-E workload) using 1,000 concurrent users spread among four agents. More than 48,000 IOPS were achieved on a PS6210S array with average disk latency less than 2 ms.

5.2 PS6210XS: TPC-E characterization

In this test, the entire database (database files, redo logs and index) was hosted on the PS6210XS array. Tests were executed at varying loads using Quest Benchmark Factory to determine the maximum user load which could provide optimal performance adhering to the test criteria (Disk latency less than 20ms and also the average response time less than 2 seconds).

It was determined that 600 users is the optimal user load for this configuration based on the initial test results and test criteria as explained in Section 3.2. The storage array produced 42,638 IOPS with less than 2 ms disk latency as shown in the SAN HQ chart below.

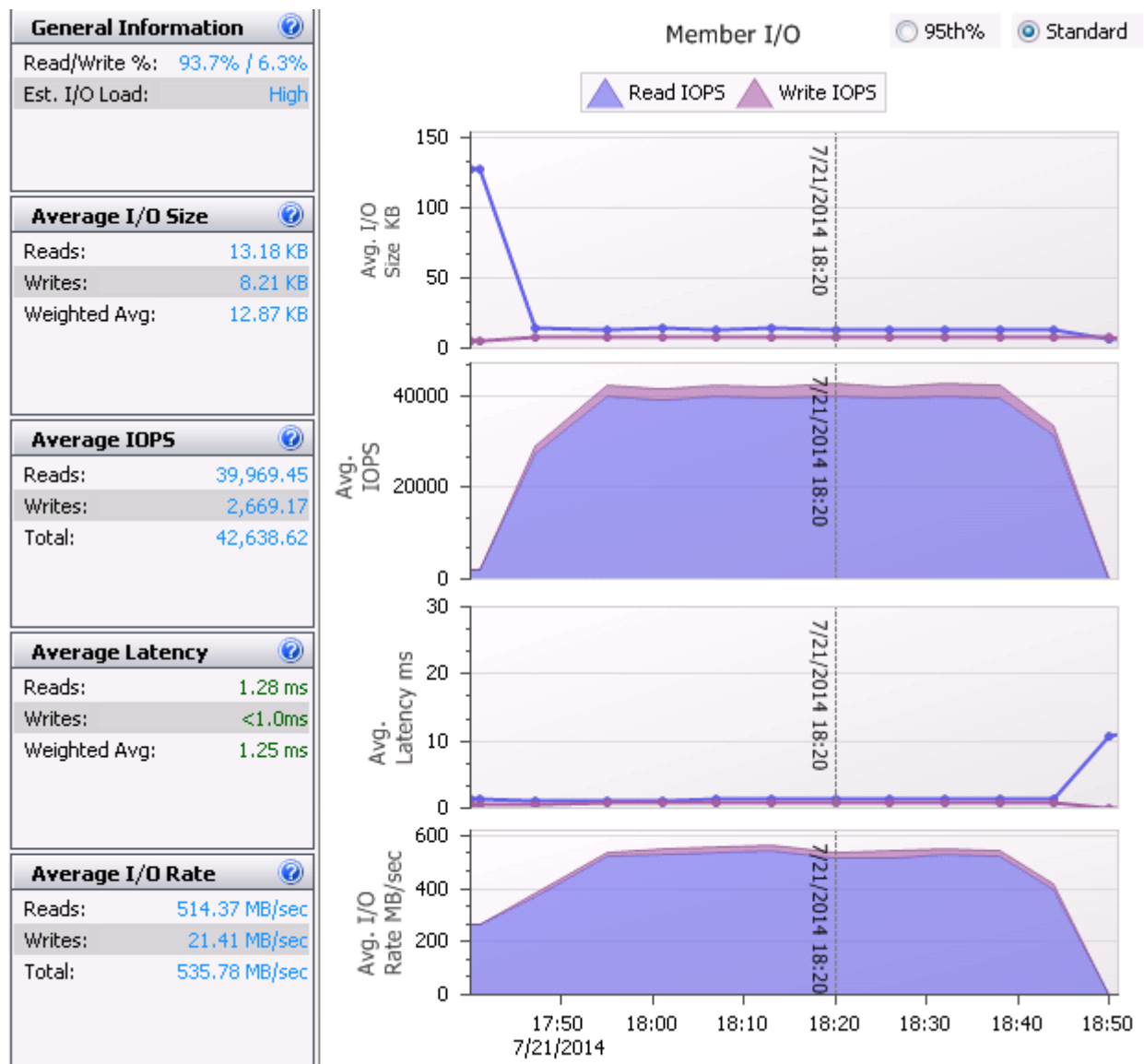


Figure 10 PS6210XS TPC-E I/O (IOPS and latency reported by SAN HQ)

Quest Benchmark Factory simulated heavy real-world database transactions (TPC-E workload) using 600 concurrent users spread among four agents. More than 42,000 IOPS were achieved on a PS6210XS array with average disk latency less than 2 ms.

5.3 Summary

5.3.1 Oracle 11g RAC OLTP on PS6210S all-flash array

More than 62,000 IOPS were observed while executing the ORION I/O simulation test. TPC-E like database transactions were executed by running 1,000 concurrent users spread among four agents using Quest Benchmark Factory. Even while executing these heavy database transactions, more than 48,000 IOPS were observed on the storage array with disk latency less than 2 ms for the entire duration of test.

Slightly reduced IOPS were observed on the storage array while executing TPC-E database transactions compared to the raw I/O numbers using ORION. The reasons for this behavior are:

- 48 GB of memory was allocated to the Oracle process on each node and this enabled caching a significant amount of read I/O operations at the server memory.
- ORION I/O tool simulated raw I/O and does not use Oracle stack. As a result, server memory does not play a role while executing this workload.
- The average read I/O block size while executing TPC-E style database transactions was around 13 KB (Refer to Figure 9 in Section 5.1) because I/O operations were cached at the server memory. But the I/O block size for both read and write operations was 8 KB for ORION (Refer to Figure 4 in Section 4.1).
- No difference in the block sizes for write I/O was seen because TPC-E is more read intensive with a 9.7:1 I/O read to write ratio, (More detailed information about TPC-E characteristics is available at: <http://www.cs.cmu.edu/~chensm/papers/TPCE-sigmod-record10.pdf>).

5.3.2 Oracle 11g RAC OLTP on PS6210XS hybrid array

PS6210XS arrays offer superior performance and also huge usable capacity due to the unique combination of SSD and 10K SAS drives within the same chassis. The PS6210XS arrays consist of seven SSDs and 17 10K SAS drives and have built-in intelligence to automatically move the data between SSD and SAS tiers within a volume based on the performance requirements.

The objective of the test described in Section 4.2 was to evaluate the behavior of the PS6210XS array when the database does not fit within the SSD capacity of the array and also I/O is 100% random in nature without any locality of reference. This is the worst case scenario because typically there is always a set of data which is accessed more frequently in a database environment. In this worst case scenario, the SSDs had to wait for the I/O activity to complete on the 10K SAS drives. Even in this worst case scenario, the PS6210XS arrays offered more than 13,000 IOPS with disk latency less than 2 ms.

The following chart compares the performance numbers from PS6210XS for the following scenarios:

- ORION I/O simulating TPC-E style workload with 100% random I/O and no data locality
- ORION I/O simulating TPC-E style workload with 100% random I/O and 20% data locality
- TPC-E style transactions executed on the actual Oracle 11g 2-Node RAC database using Quest Benchmark Factory



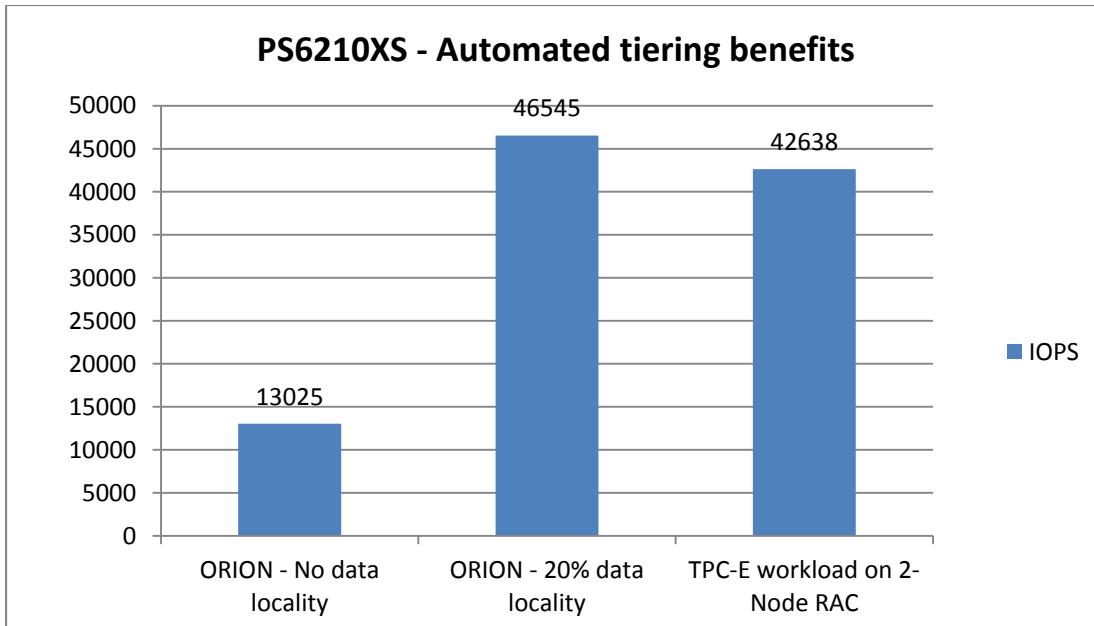


Figure 11 PS6210XS: Automated internal tiering benefits for OLTP database applications

When a specific locality is accessed frequently, the PS6210XS internal tiering mechanism identifies the hot pages and then moves them to the SSD tier as demonstrated by the test results in Section 5.3. In this test, locality of data reference was specified as 20%.

The test result showed approximately 11,800 IOPS at the beginning and then gradually increasing to 46,500 over the test time span. These results clearly demonstrated how the internal data tiering mechanism built in within PS6210XS hybrid array can create Oracle OLTP database performance improvements while cost-effectively satisfying the capacity requirements simultaneously.

6 Best practice recommendations

6.1 Storage

Use the following best practices when configuring Dell EqualLogic storage arrays for a data protection solution.

- A configuration with more small volumes is preferred over a low number of large volumes for better performance and simpler management of OLTP databases. It is recommended to use at least eight or more volumes per PS6110XS array member for achieving optimal performance. Increasing the number of volumes beyond eight may not result in significant performance improvements.

6.2 Network infrastructure

Use the following network infrastructure design best practices:

- Design separate network infrastructures to isolate the LAN traffic from the SAN traffic (iSCSI).
- Design redundant SAN component architectures. This includes the NICs on the servers and switches for the storage network (including server blade chassis switches and external switches).
- Make sure that the server NIC ports and storage NIC ports are connected so that any single component failure in the SAN will not disable access to any storage array volumes.
- Enable flow control on both the server NICs and switch ports connecting to the server and storage ports.
- Enable jumbo frames on the server ports and switch ports.
- On iSCSI SAN switches, spanning tree should be disabled on switch ports connecting to end devices like server and storage ports. The 'Portfast' setting should be enabled in the switch configuration for these ports.

Note: General recommendations for EqualLogic PS Series array network configuration and performance is provided in the document titled *EqualLogic Configuration Guide* at <http://en.community.dell.com/dell-groups/dtcmmedia/m/mediagallery/19852516/download.aspx>.

6.3 Operating system

- Enable flow control and jumbo frames on the Server NIC ports used for connecting to storage and match those settings with your switch and array controller port settings. Jumbo frames provide a more efficient transport for data transfers.
- It is recommended to use EqualLogic-aware MPIO functionality provided by the Host Integration Tools for the server OS. HIT/Linux can be obtained from EqualLogic support at www.equallogic.com/support. Configuring EqualLogic aware MPIO provides enhanced performance benefits and also enables easier configuration by automating configuration of multiple required iSCSI sessions (connections) to the target volumes. Since MPIO will be in use, disable NIC teaming or bonding on all ports used for iSCSI connections. Also, to simplify system



management and reduce the chances of making errors during system configuration, disable any unused NIC ports on the system.

- It is recommended to run the utility 'eqltune' included with the HIT/Linux and follow the OS specific parameter recommendations listed by the utility.

6.4 Oracle database application

6.4.1 Database volume layout

- Oracle recommends using ASM for simplified administration and optimal I/O balancing across disk groups.
- ASM External redundancy is preferred while configuring ASM disk groups. There is no need to use other types of redundancy because the EqualLogic array will provide the required redundancy.
- Preferably use same sized ASM disks within ASM disk groups. This helps Oracle ASM to optimally perform load balancing.
- Map storage volumes to individual ASM disks and create separate ASM Disk Groups for database data files, temporary data files and transactions logs.
- Create a separate tablespace for database data files and temporary database files, ensuring hosting on respective ASM disk groups.

6.4.2 Oracle AMM

- The Oracle AMM feature assigns MEMORY_TARGET to be 40% of the total physical memory on the server. Based on the test results, setting MEMORY_TARGET to a value 50% to 60% of the total server memory can be used as a guideline for TPC-E like workloads. The memory setting can vary based on various database factors, so careful monitoring and analysis should be done before tuning this parameter.
- The Oracle AMM feature requires the /dev/shm file system size to be at least equal to the value of the MEMORY_TARGET parameter. By default, the size of this file system will be set to 50% of the total physical memory.
- You need to change the size of the /dev/shm file system to support custom MEMORY_TARGET settings. Make sure that the database instances are shut down before making any changes to the size of /dev/shm.
- You can use the **mount** command as shown below. The following command will allow the MEMORY_TARGET value to be set up to 64 GB.
[root@oranode1 oswbb]# mount -t tmpfs shmfs -o size=65g /dev/shm
- Verify that the size of the /dev/shm file system is modified using the command shown below.
[root@oranode1 oswbb]# **df -k**

Filesystem	1K-blocks	Used	Available	Use%	Mounted on
/dev/mapper/vg_oranode1-lv_root	51606140	26696252	22288448	55%	/



```

tmpfs                68157440          0  68157440    0% /dev/shm
/dev/sda1            495844        68132    402112   15% /boot
/dev/mapper/vg_oranode1-lv_home
                    201585632    195108 191150524    1% /home
shmfs                68157440          0  68157440    0% /dev/shm

```

- The ASM instance on nodes will terminate as soon as the /dev/shm size is modified. Complete the following step to bring up ASM instances on database nodes.

- Login to each node as “grid” user and export ORACLE_SID with ASM instance. Enter the commands as shown below to bring up the ASM instance. Repeat the same steps on all database nodes.

```
[grid@oranode1 bin]$ export ORACLE_SID=+ASM1
```

```
[grid@oranode1 bin]$ pwd
```

```
/u01/app/11.2.0/grid/bin
```

```
[grid@oranode1 bin]$ ./asmcmd
```

```
Connected to an idle instance.
```

```
ASMCMD> startup
```

```
ASM instance started
```

```
Total System Global Area 1135747072 bytes
```

```
Fixed Size                2260728 bytes
```

```
Variable Size            1108320520 bytes
```

```
ASM Cache                25165824 bytes
```

```
ASM diskgroups mounted
```

```
ASM diskgroups volume enabled
```



```
ASMCMD> ls

DB1DATA/

DB1FLASH/

DB1LOG1/

DB1LOG2/

OCRVOTE/

ASMCMD> exit
```

6.4.3 Oracle ASM configuration on RHEL 6.3

ASMLib is a free, optional software feature from Oracle that simplifies the management and discovery of ASM disks, but Oracle changed the policy for RHEL 6 and provided ASMLib software and support only for Unbreakable Enterprise Kernel and the Red Hat compatible kernel for Oracle Linux.

Red Hat released an open-source package 'kmod-oracleasm' beginning with RHEL 6.4 that can be installed from the RHEL server supplementary channel. But, updates and support for this module must be received from Red Hat and this is not supported by Oracle.

To deploy Oracle ASM on RHEL 6.x kernels without ASMLib support from Oracle, use the procedure described in the link below. The following procedure uses *udev* (device manager for Linux which manages all device nodes in /dev) rules on a RHEL 6 server to configure Oracle ASM.

<http://en.community.dell.com/techcenter/b/techcenter/archive/2013/10/04/configure-oracle-asm-without-asmlib-on-rhel-6-x.aspx>



Conclusion

The test results illustrate that the PS6210S all-flash arrays are a good choice when performance is a key concern for running Oracle OLTP database applications. The PS6210S all-flash arrays offered significant performance benefits specifically for small random I/O, which typically represents OLTP database applications. This was clearly demonstrated by the ORION I/O simulation and execution of TPC-E style database I/O transactions on the Oracle 2-Node RAC database.

Additionally, as database sizes grow significantly over time, enterprises need to plan for increased capacity without compromising the performance. In an Oracle OLTP environment, the working set – frequently accessed data that needs fast access time – is typically a relatively small percentage of the total OLTP system capacity. Therefore, the performance and capacity balance provided by the EqualLogic PS6210XS arrays is exceptionally well suited for Oracle OLTP environments. This was clearly demonstrated by the test results discussed in Sections 4 and 5.

Optimal database performance can be achieved when the applicable best practices laid out in this paper are adhered to. It must be ensured that the entire ecosystem, including server resources and network switches, are sized and configured appropriately to meet the workload performance requirements. The administrator must also ensure that the OS and databases are configured with the optimal settings and design recommendations mentioned in this paper.



A Configuration details

This section contains an overview of the configurations used for the testing described in this document.

Table 2 Test configuration: Hardware components

[illegible]

Table 3 Test configuration: Software components

Test configuration: Software components	
Database Servers	Red Hat Enterprise Linux 6.3 Host Integration Tools for Linux 1.2.0 Oracle Database 11g R2 Enterprise Edition (Version 11.2.0.4) Two-Node Oracle RAC database ASM for Clusterware and Database
Quest Benchmark Factory	2x Virtual machines were used (One as the Benchmark Factory console and one for launching agents) 2x Windows Server 2008 R2 Enterprise Edition Quest Benchmark Factory 6.9.2 2 agents from each VM Oracle 11.2.0.3 client installed VMware vSphere ESX version 5.1 Bare metal hypervisor directly installed on R810 servers Managed by vCenter.
Monitoring Tools:	EqualLogic SAN HQ version 2.6.0 Oracle OS Watcher utility (installed on database servers) Oracle 11g R2 Automatic Workload Repository (AWR)



Additional resources

Support.dell.com is focused on meeting your needs with proven services and support.

DellTechCenter.com is an IT Community where you can connect with Dell Customers and Dell employees for the purpose of sharing knowledge, best practices, and information about Dell products and your installations.

Referenced or recommended Dell publications:

- Dell EqualLogic Configuration Guide:
<http://en.community.dell.com/dell-groups/dtcmedia/m/mediagallery/19852516/download.aspx>

Referenced or recommended Oracle publications:

- Oracle 11g R2 installation guide:
http://docs.oracle.com/cd/E11882_01/install.112/e47689.pdf

For EqualLogic best practices white papers, reference architectures, and sizing guidelines for enterprise applications and SANs, refer to Storage Infrastructure and Solutions Team Publications at:

- <http://dell.to/sM4hJT>

