



Sizing and Best Practices for Deploying Microsoft Exchange Server 2013 on VMware vSphere and Dell EqualLogic PS6110E Arrays

A Dell EqualLogic Reference Architecture

Dell Storage Engineering
March 2014

Revisions

Date	Description
March 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 Boomi™, 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.....	4
Feedback.....	4
Executive summary.....	5
1 Introduction.....	6
1.1 Purpose and scope.....	6
1.2 Terminology.....	6
2 Dell EqualLogic PS6110 arrays for Exchange Server 2013 deployments.....	8
3 Virtualization of Exchange Server storage workload.....	9
3.1 Exchange store elements.....	9
4 Solution architecture overview.....	11
4.1 Conceptual system design.....	11
4.2 Physical system configuration.....	12
4.3 Storage and volume layout.....	13
5 Exchange Server deployment factors for storage.....	14
5.1 Mailbox user profile workload.....	15
5.2 Characterize the mailbox size.....	18
5.3 Database volumes layout.....	21
5.4 Scale up the user count per server.....	23
5.5 Characterize the impact of the iSCSI initiator collocation.....	26
6 Best practices recommendations.....	29
A Configuration details.....	32
A.1 Hardware components.....	32
A.2 Software components.....	33
A.3 Network configuration.....	34
A.4 Host hypervisor and VMs configuration.....	35
B Microsoft Jetstress considerations.....	40
C Additional resources.....	41



Acknowledgements

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

Engineering: Danilo Feroce

Technical Marketing: Omar Rawashdeh

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 SISfeedback@Dell.com.



SISfeedback@Dell.com



Executive summary

Virtualization technologies are becoming the backbone of modern datacenters thanks to increased efficiency and reduced costs. A long-established concern about virtualization is the tolerance to heavy storage workloads and the means to correctly predict and properly handle them.

Microsoft Exchange Server is the predominant messaging infrastructure in corporate environments and requires the availability and performance of a mission critical application. It has traditionally been on the list of the top storage workloads to be handled by IT departments. The latest iteration of the product, Microsoft Exchange Server 2013, has an extended reduction in storage data transfers and a steep increment in the supported capacity per mailbox when compared with former versions as seen in Figure 1. The constant investment in redesign and polishing the storage subsystem has shown continuous improvement in recent versions.

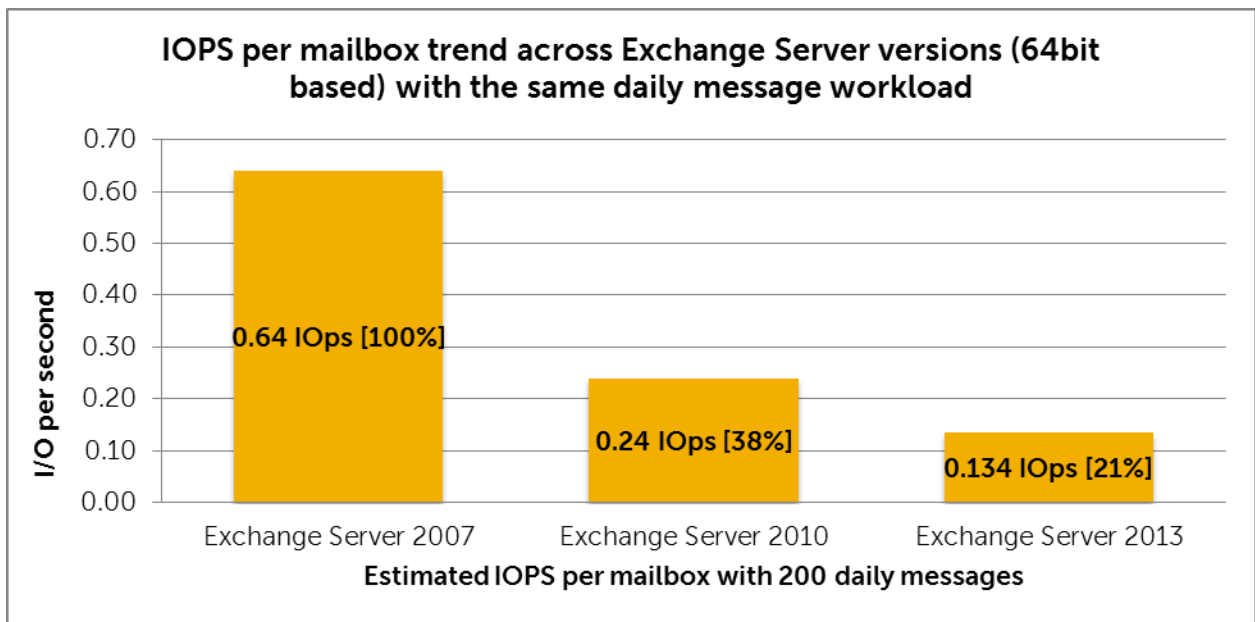


Figure 1 Reduction trend in IOPS per mailbox across Exchange Server versions, considering only the 64bit based generation (data gathered from Microsoft TechNet documentation)

The solution presented in this paper address some of the most relevant variables to take into consideration when planning for the deployment of a virtualized Exchange Server 2013 back-end solution in conjunction with an EqualLogic PS6110E array with 4TB drives:

- The progression of the storage response time when increasing the messaging workload
- The impact of using large mailboxes with an extended capacity allocation on the storage subsystem
- The variation in the composition of the workload when it is distributed across a different number of databases
- The selection of user count per host and array and its impact on performance
- A comparison between a host and a guest iSCSI initiator performance for the virtualized workload



1 Introduction

Messaging solutions based on Microsoft Exchange Server 2013 should follow the changing trends around customer needs and reinforce the new strengths and features available in the product. The support of large disks, up to 4TB, and thus expanded capacity in the EqualLogic PS6110E arrays enables the very large mailboxes required by demanding customers. A natural component of these solutions is a VMware ESXi host that in version 5.5 supports virtual disks greater than 2TB in a VMFS-5 datastore.

The reference architecture validated for the purpose of this paper represents a medium organization with 5,000 mailbox users. An infrastructure running a simulated Exchange Server 2013 is virtualized by the VMware vSphere 5.5.

The responsiveness of this reference architecture has been stretched across several dimensions with the goal of providing the answers to a set of questions that would usually arise at the time of the design and deployment. These exercises assessed a guest versus host iSCSI initiator, the deployment of the same amount of users across different numbers of databases and volumes, the increase of active users to prepare for growth, the impact of increasing mailbox size or the variable load per mailbox.

1.1 Purpose and scope

This paper is primarily intended for IT professionals (IT managers, Solution Architects, Exchange and Storage Administrators, and System and Virtualization Engineers) who are involved in defining, planning, and/or implementing Exchange Server infrastructures and who would like to investigate the benefits of using EqualLogic storage. This document assumes the reader is familiar with Exchange Server functions, EqualLogic SAN operation, and VMware vSphere architecture and system administration. The scope of this paper is restricted to a local datacenter topology and does not include specific or detailed server sizing information.

1.2 Terminology

The following terms are used throughout this document.

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

Member: Identifies a single physical EqualLogic array.

Pool: A logical collection that each member (array) is assigned to after being added to a group and contributes its storage space to the entire pool.

Hypervisor: Denotes the software layer that manages the access to the hardware resources, residing above the hardware, and in between the operating systems, running as guests.

Virtual Machine (VM): An operating system implemented on a software representation of hardware resources (processor, memory, storage, network, and others). VMs are usually identified as guests in



relation with the host operating system that executes the processes to allow them to run over an abstraction layer of the hardware.

Microsoft Exchange Database Availability Group (DAG): A pool of networked Exchange mailbox servers that hosts multiple copies of the same Exchange databases.

Balanced tree (B-Tree): A tree data structure where a node can have a variable number of child nodes, commonly used in databases to maintain data sorted in a hierarchical arrangement. It allows efficient data access to the pages for insertion, deletion, and searches.

Process: An instance of a computer program or application that is being executed. It owns a set of private resources: image or code, memory, handles, security attributes, states, and threads.

Thread: A separate line of execution inside a process with access to the data and resources of the parent process. It is also the smallest unit of instructions executable by an operating system scheduler.

Key performance indicators (KPI): A set of quantifiable measures or criteria used to define the level of success of a particular activity.



2 Dell EqualLogic PS6110 arrays for Exchange Server 2013 deployments

Dell EqualLogic storage is ideal for companies and organizations with growing data and performance needs. EqualLogic's frameless, IP-based peer storage architecture enables live, non-disruptive firmware upgrades and maintenance, as well as linear scaling of capacity and performance with each array added to the group

Dell EqualLogic PS6110 Series arrays are designed to meet the performance and availability needs of application and virtualization environments in medium to large enterprises. These virtualized iSCSI SANs combine intelligence and automation with fault tolerance to provide simplified administration, rapid deployment, enterprise performance and reliability, and seamless scalability using innovative Fluid Data technology. EqualLogic PS6110E with multiple high capacity drive size options, up to 4TB 7.2K RPM drive capacity with Self Encrypting Drive (SED) option, provides both entry-level configuration for affordable capacity and high capacity configurations for data-intensive application, highly consolidated environments and collaboration applications such Microsoft Exchange Server 2013. The PS6110E with 4TB drives offers IT generalists the ability to manage more data with fewer resources, the capability to integrate tightly with Exchange Server 2013 environments and the flexibility to use with various operating systems/hypervisors like VMware vSphere 5.5. The PS6110E offers low total cost of ownership (TCO) that can be tailored to fit your exact messaging needs and grow with you, simply and cost effectively.



3 Virtualization of Exchange Server storage workload

Exchange Server is a product built upon a wide set of components and services that cooperate to support the various requirements needed to design and deploy a messaging infrastructure with advanced capabilities in every organization. The latest Exchange Server version offers a simplified topology, reducing the number of server roles available, while consolidating and optimizing the services in two layers only: the front-end services dealing with the clients (Client Access Service role) and the back-end services in charge of data management and message transport (Mailbox role).

Since the same or more capabilities are now implemented in fewer layers, the first outcome is an increase in the computing processing demand for the servers deployed to support Exchange Server 2013. This is a challenge that should be planned carefully from both capacity and performance standpoints and tailored to each organization in order to be successful with virtualized infrastructures.

The appropriate sizing of the storage subsystem is a key factor in the mailbox role performance and can contribute heavily to easing bottlenecks or administrative overhead for that layer. To better introduce the analysis available in the later sections of this paper, the components underneath the Managed Store of Microsoft Exchange Server 2013 will be examined.

3.1 Exchange store elements

Accessing mailbox databases is the primary activity that generates I/O on a storage subsystem. A database is a logical representation of a collection of user or system mailboxes, and it is also an aggregation of files on the disk which are accessed and manipulated by a set of Exchange services following distinct rules (for example, the Information Store, the Search or the Replication Service).

Database file (*.edb): The container for user mailbox data. Its content, broken into database pages of 32 KB, is primarily read and written randomly as required by the Exchange services running on the mailbox server role. A database has a 1:1 ratio with its own *.edb database file. The maximum supported database size in Exchange Server 2013 is still 16 TB, where the Microsoft guidelines recommend a maximum 200 GB database file in a standalone configuration and 2 TB if the database participates in a replicated DAG environment.

Transaction logs (*.log): The container where all the transactions that occur on the database (create, modify, delete messages, and others) are recorded. Each database owns a set of logs and keeps a one-to-many ratio with them. The logs are written to the disk sequentially, appending the content to the file. The logs are read only when they are in a replicated database configuration within a DAG or in the event of a recovery. The log truncation method is the process to remove old log files preventing them from accumulating. The truncation happens when a consistent backup is performed or when the log configuration is set to circular logging.

Checkpoint file (*.chk): A container for metadata tracking when the last flush of data from the memory cache to the database occurred. Its size is limited to 8 KB and, although repeatedly accessed, its overall amount of I/O is limited and can be ignored. The database keeps a 1:1 ratio with its own checkpoint file and positions it in the same folder location as the log files.



Search Catalog: A collection of flat files (content index files) built by the Microsoft Search Service, having several file extensions and residing in the same folder. The client applications connected to Exchange Server benefit from this catalog by performing faster searches based on indexes instead of full scans.

Exchange Server uses a proprietary format called Extensible Storage Engine (ESE) to access, manipulate, and save data to its own mailbox databases. The same format is now employed on the Exchange HUB server role for the queue databases. ESE technology, previously known as Jet Database Engine, has evolved through several versions of Exchange Server and has been a part of several Microsoft products since its inception (for example, Microsoft Access, Active Directory, File Replication Service, WINS server, and Certificate Services).

The ESE is an Indexed Sequential Access Method (ISAM) technology that organizes database data in B-Tree structures. Ideally, these databases are populated by data kept together or adjacent. When this does not occur, external reorganization or defragmentation processes should be used to restore the optimal data contiguity in these structured databases.

To summarize, an Exchange mailbox database is subject to a subset of tasks performing storage access:

- The regular read and write access required to retrieve and store user mailbox data (according to the Exchange cache policy)
- The online defragmentation and compacting activities due to the B-Tree optimization
- The background database maintenance, which includes recoverable items cleanup, deleted mailboxes purge, and other activities addressing logical object support
- The checksum database scan to verify data block integrity (sequential read activity), which can be set to run constantly in the background or at a scheduled time

Furthermore, Exchange Server offers a specialized offline manual defragmentation task that runs while the database is dismounted by taking advantage of the ESEUTIL.EXE command line tool. The principal goal of this task is to reclaim the empty space left in a database by online defragmentation and to shrink the size of the *.edb file itself. This returns the free space to the operating system volume.

Note: It is not recommended to include offline defragmentation in a regular maintenance plan due to the disruption in the availability of the database, the rupture of the logs chain, and the need for database re-seeding in case of DAG configuration.

Exchange DAG is a pool of up to 16 networked servers that hosts multiple copies of the same Exchange database or databases where only one of the copies is active at a specific point-in-time within the group. The other copies are passive and contain data sourced from replicated and replayed transaction logs.

Log Checkpoint depth refers to the amount of logs written to the disk and that contain transactions not yet flushed to the database file. In Exchange Server 2013, during a DAG failover the database cache is no longer flushed, since it is treated as a persistent object. Thus the log checkpoint for the passive databases is increased to 100 to reduce the write I/O and to reduce the failover time since the passive database has to pre-read less data.



4 Solution architecture overview

The solution presented and evaluated in this paper is built upon a virtual infrastructure supported by VMware vSphere 5.5 and a back-end iSCSI SAN provisioned by Dell EqualLogic Storage. The operating system of the VM simulating the Exchange Server workload is Windows Server 2012, while the remaining VMs running the infrastructure and monitoring are based on Windows Server 2008 R2.

4.1 Conceptual system design

The elements of the infrastructure supporting the simulated environment, their main relationships and connectivity links are represented in the conceptual diagram in Figure 2.

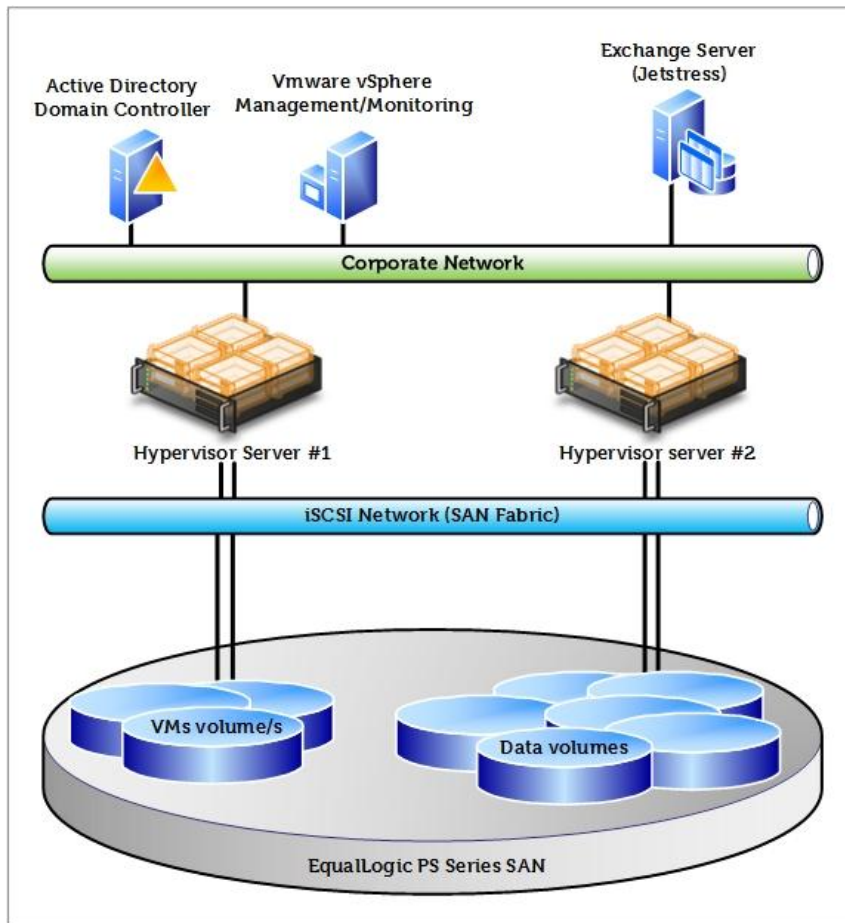


Figure 2 Conceptual system design for the components of the solution

The key elements of this design are:

- Single Active Directory forest, single domain, single site
- Centralized management and monitoring with dedicated resources (both physical and virtual)
- Separated network design to maintain traffic isolation between traditional LAN and iSCSI access
- Building block design approach for mailbox server with Jetstress



4.2 Physical system configuration

The physical components and the connections beneath the virtual infrastructure are shown in Figure 3.

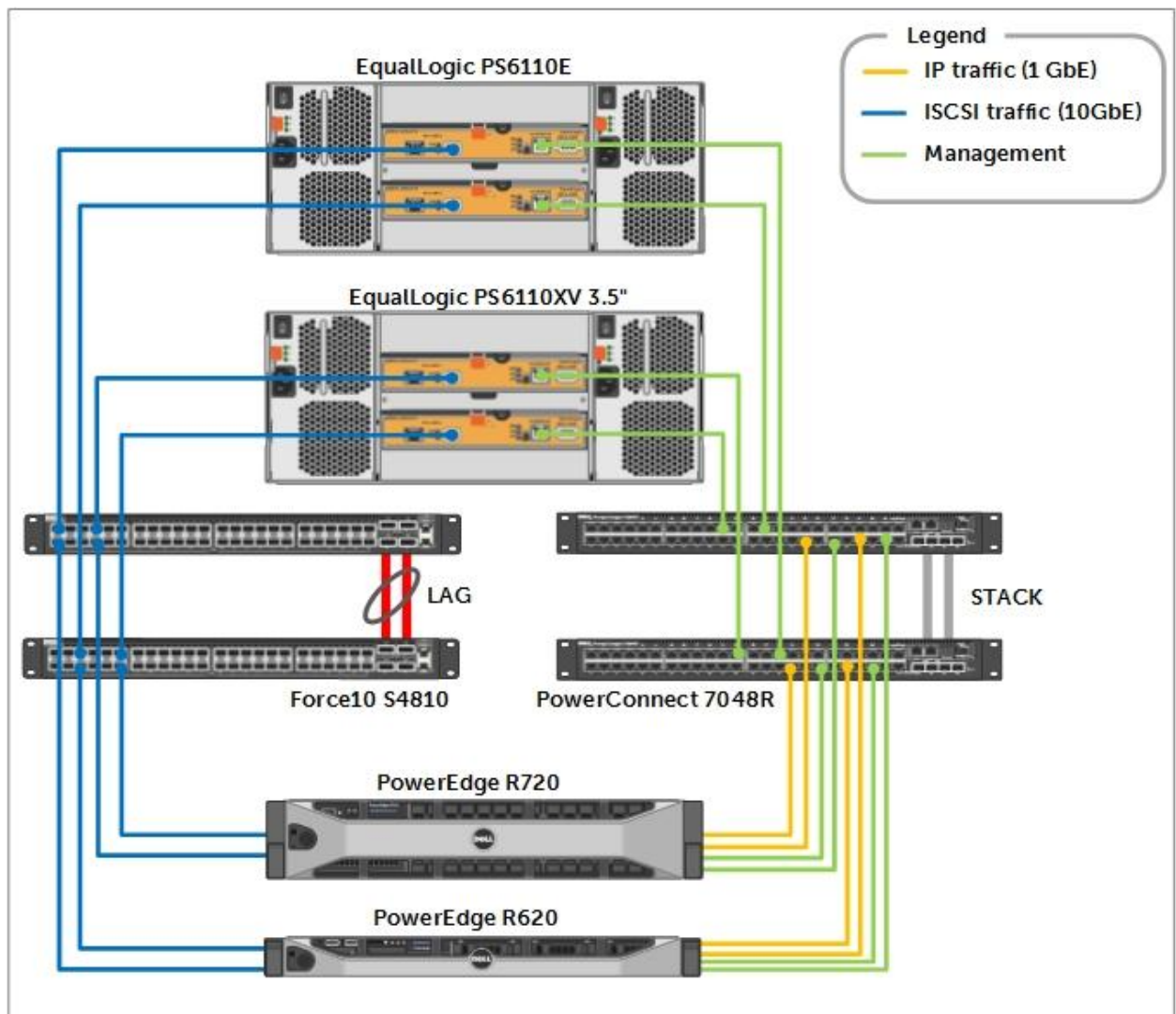


Figure 3 Physical system design for the components of the solution

The solution architecture is deployed on Dell rack servers with top of rack (ToR) Ethernet network switches separately dedicated to IP traffic (traditional client/server, management, and hypervisor communications) and to iSCSI storage access. The hardware elements contributing in the architecture are:

- One PowerEdge R720 rack server to power the hypervisor beneath the simulated Exchange VM
- One PowerEdge R620 rack server to power the hypervisor beneath the centralized management and monitoring infrastructure
- One EqualLogic iSCSI SAN provisioned with one PS6110XV 3.5\" and one PS6110E array (10GbE)
- Two PowerConnect 7048R Ethernet switches to support LAN IP traffic configured in a stack



- Two Force10 S4810 Ethernet switches to support the iSCSI traffic configured in a Link Aggregation Group (LAG) consisting of two connections between the ToR S4810 switches

Note: For more details of the configurations used for the solution infrastructure, including a hardware and software list, SAN array characteristics, hypervisor and VMs relationship, and physical and virtual network connections, refer to Appendix A.

4.3 Storage and volume layout

The EqualLogic SAN arrays and the volumes underlying the Exchange databases are configured with:

- Two EqualLogic groups, one configured on a PS6110XV 3.5" array and one on a PS6110E array.
- Two storage pools, one per each group and including the single member
- RAID 6 policy applied as a reference configuration
- One volume allocated within the PS6110XV pool for each host, to store the file images, configuration and temporary files for the corresponding hosted VMs
- A dedicated set of data volumes allocated within the PS6110E pool, dedicated to the Exchange databases and uniquely assigned to the R720 host or to the Exchange VM hosting the simulation
- Each Exchange mailbox database and its private set of log files maintain a 1:1 ratio with the corresponding SAN volume

Note: The selection of two groups instead of one is because of the need to reinitialize the PS6110E array during each reconfiguration. One group with two pools to isolate the different workloads would have been an easier solution to manage, but would have increased the time for preparation between test cases.

Figure 4 shows the volumes defined on the EqualLogic SAN and the respective pools.

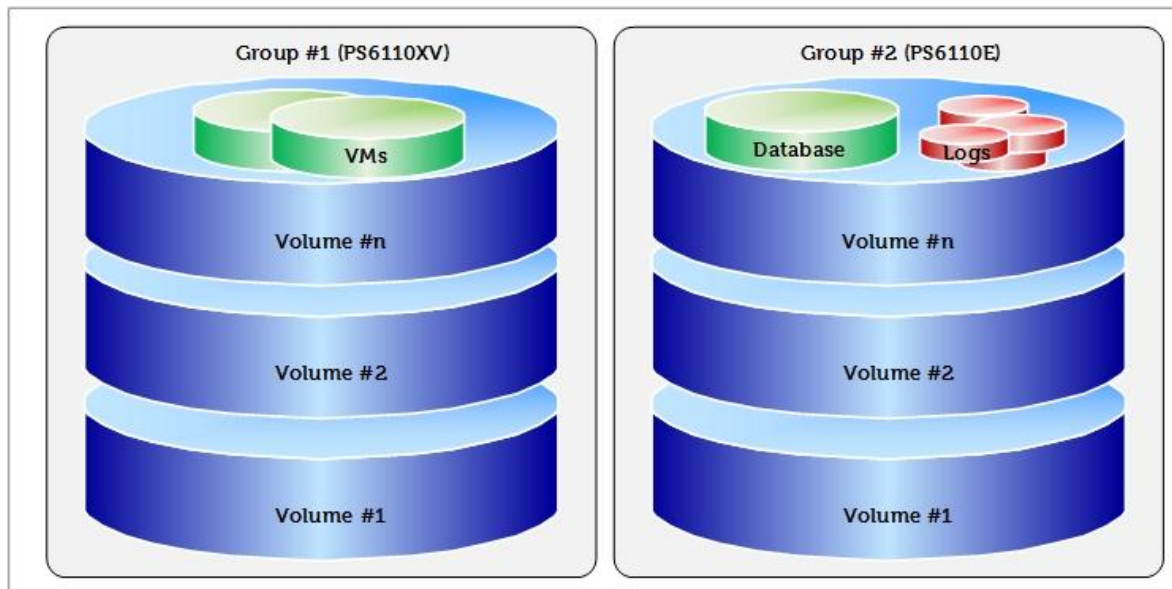


Figure 4 Volumes and databases/logs layout



5 Exchange Server deployment factors for storage

The choices selected at the time of deployment of an Exchange Server infrastructure affect the storage performance and ease of administration of a solution. Some of the most relevant areas where a different approach can make a difference have been evaluated in this paper.

To forecast the behavior of the underlying storage subsystem, the following list of variables has been evaluated to show trends.

- Assessment of the average usage profile for a mailbox user in the organization
- Definition of the mailbox quota or cap the organization plans to enforce
- Average mailbox count per database planned for the deployment
- Number of users per deployment building block and room for future growth
- Methodology to access the mailbox database files on the SAN from a virtual infrastructure (collocation of the iSCSI initiators)

The reference Exchange Server deployment used is detailed in Table 1. For each scenario evaluated, a description of the relevant differences is reported in the corresponding section later in the paper.

Table 1 Reference configuration for Microsoft Jetstress 2013 tests

Reference configuration: factors under study	
Messages per day per mailbox / IOPS per mailbox	200 messages / 0.134 IOPS (with DAG)
Number of simulated users / mailboxes	5,000 concurrent users
Mailbox size	2 GB each
Number of databases	5 databases (active)
Mailbox allocation per database	1,000 mailboxes per each mailbox database
Database size	2 TB each
iSCSI initiator software collocation	Host software (SW) initiator (ESXi 5.5)
Reference configuration: consistent factors across each scenario	
Number of database replica copies	2 (two node DAG)
Background database maintenance	Enabled
Windows Disk/Partition File System	Basic disk, GPT partition, default alignment NTFS, 64KB allocation unit size
Array model, SAN configuration	1x PS6110E (24x 4TB NL-SAS 7.2K disks), one pool
RAID policy	RAID 6
Test duration	2 hours + time required to complete DBs checksum



Below is a list of metrics and pass/fail criteria recorded while completing the tests. Most of this information is outlined by the Jetstress tool report or is verified through the recording of Windows Performance Monitor and Dell EqualLogic SAN Headquarters counters. Microsoft thresholds for Exchange Server storage validation are reported as well.

Database Reads Average Latency (ms) is the average length of time to wait for a database read operation (random reads). It should be less than 20 ms according to Microsoft threshold criteria.

Database Writes Average Latency (ms) is the average length of time to wait for a database write operation (random writes). It should be less than 20 ms according to Microsoft threshold criteria.

Logs Writes Average Latency (ms) is the average length in time to wait for a log file write operation (sequential writes). It should be less than 10 ms according to Microsoft threshold criteria.

Planned Transactional IOPS are the target amount of IOPS for the test (calculated by multiplying the number of users by the IOPS per mailbox).

Achieved Transactional IOPS are the amount of IOPS actually performed by the storage subsystem to address the transactional requests. The result should not diverge more than 5% from the planned IOPS to be considered a successful test iteration according to Microsoft Jetstress.

LOGs IOPS are the IOPS performed against the log files during the transactional test. They are not directly taken into account as part of the transactional IOPS, but are tracked separately instead.

Additional IOPS are the IOPS generated for the database (DB) maintenance, log files replication and all the remaining activities on the storage subsystem, calculated as the difference between the IOPS provisioned by the EqualLogic SAN and the previously reported transactional and logs IOPS.

Total IOPS of the SAN is the sum of the three elements above: achieved transactional IOPS, Logs IOPS and additional IOPS. It represents the entire IOPS footprint performed against the back-end SAN during a test. It is recorded at the SAN level and verified with the Exchange host.

Note: For details about the simulation tool, Microsoft Jetstress 2013, refer to Appendix B.

5.1 Mailbox user profile workload

The first variable to calculate the workload performed by an Exchange server or even a single database is the user profile load in that specific environment. Every organization has a unique set of users with their own way to benefit from the messaging system implemented, although different departments may have uneven approaches due to their communication style or simply their role in the organization.

A mailbox usage profile denotes the usage characteristics of a mailbox (send, receive, open, close and delete items). It is commonly defined by the amount of messages sent and received per day and the average message size (typical Microsoft interpretation is 75KB). It is further illustrated in terms of transactional IOPS per mailbox through considerations made around the size of database cache allocated per each mailbox.



The goal of the following analysis is to establish the Exchange KPI trends and IOPS ratios when increasing the average mailbox user profile in the reference environment defined previously.

Table 2 shows a summary of the configurations used for this scenario (reference and variations).

Table 2 Test parameters: mailbox profile workload

Reference configuration: factors under study	
Messages per day per mailbox / IOPS per mailbox	100 messages / 0.067 IOPS (with DAG)
	200 messages / 0.134 IOPS (with DAG)
	300 messages / 0.201 IOPS (with DAG)
Reference configuration: consistent factors within this scenario	
Number of simulated users / mailboxes	5,000 concurrent users
Mailbox size	2 GB each
Number of databases	5 databases (active)
Mailbox allocation per database	1,000 mailboxes per each mailbox database
Database size	2 TB each
iSCSI initiator software collocation	Host SW initiator (ESXi 5.5)
Number of database replica copies	2 (two node DAG)
Array model, SAN configuration	1x PS6110E (24x 4TB NL-SAS 7.2K disks), one pool
RAID policy	RAID 6

The three profiles selected for evaluation cover a wide spectrum of typical users, from light to intensive, with their respective resulting IOPS per mailbox that are then used to size the storage solution accordingly. Each profile has its corresponding estimated memory footprint in the Exchange DB cache, where intensive users with a higher ratio of data changes require a larger amount of cache.

The Exchange DB cache provides a temporary storage for the data coming in and out from the storage subsystem, allows faster data accessibility to end-users and optimized storage access thanks to algorithms that combine multiple changes to the same blocks of data before flushing them to the storage.

Table 3 Exchange database cache per mailbox profile and estimated per server

Messages per day per mailbox	Cache per user	Number of mailboxes	Exchange DB cache
100 messages	24MB	5,000	120GB
200 messages	48MB		240GB
300 messages	72MB		360GB



Table 3 shows a summary of the memory footprint and requirements of a single mailbox and the corresponding estimate for the Exchange server. The memory allocated for the cache is the largest fraction of the total amount required by the server, which should include the amount for the basic services (evaluated against the number of active databases in the worst availability scenario) and the amount consumed by the new generation of content indexing technology included in Exchange Server 2013.

Figure 5 shows the results collected from the three instances of Exchange Jetstress simulations.

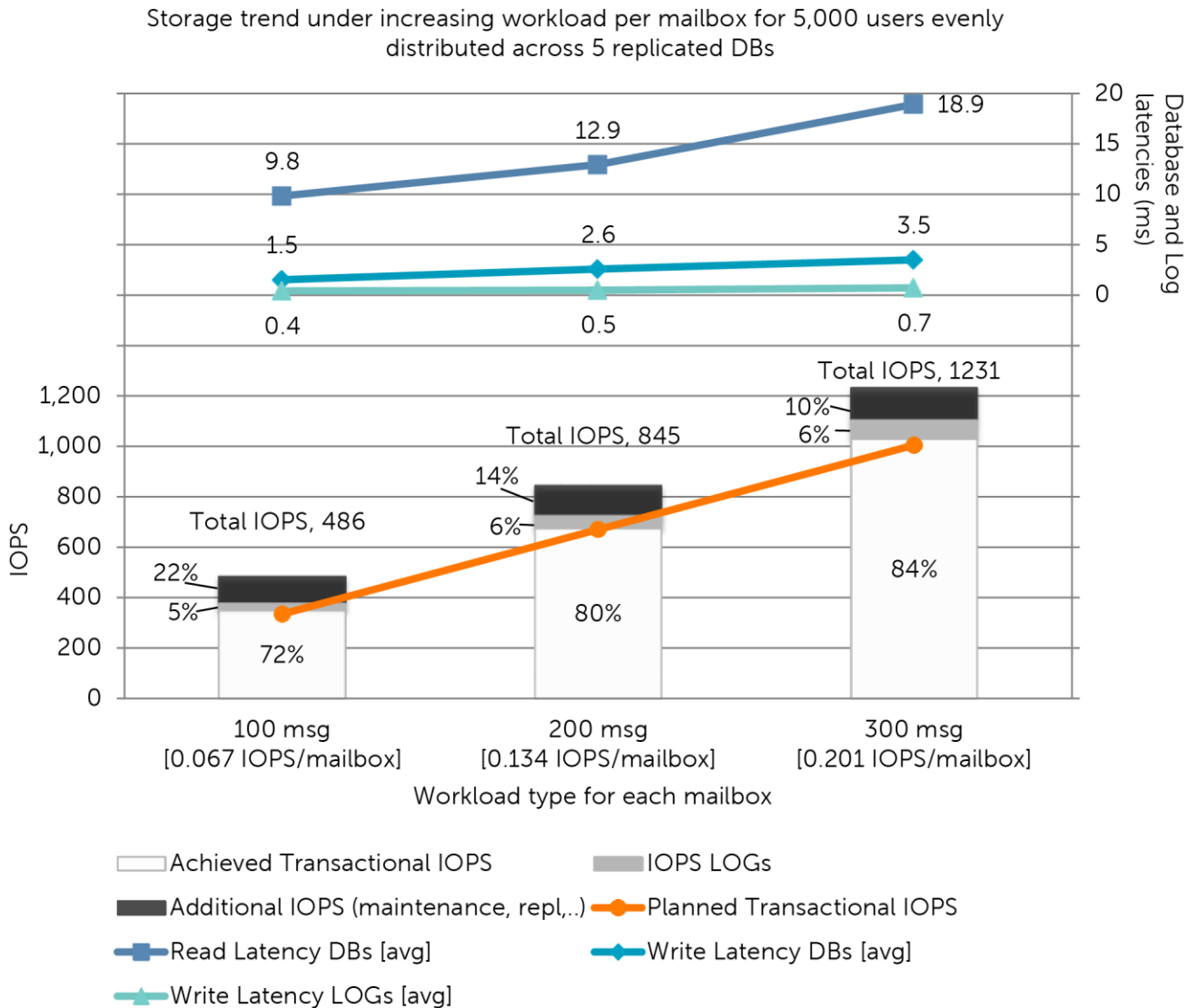


Figure 5 Storage trend under increasing workload per mailbox for 5,000 users evenly distributed across five replicated databases

Note: For graphical representation purposes, the numerical results presented in this paper are rounded to one decimal digit for latencies and zero decimal digits for percentages and IOPS.



These results are gathered from simulations executed in a lab built to the specifications explained in this whitepaper. Other generally available tools or sizing calculators might have results slightly different based on the amount of assumptions made for each particular tool.

The read/write ratio recorded during the three sessions of tests shows a stable 70% reads versus 30% writes for the lighter workload, slightly increasing the percentage of writes with the increase of the workload up to 67% reads and 33% writes for the heavier workload simulated. This is a first proof of the profound changes to the storage access pattern in Exchange Server 2013 versus the former 2010 version where we recorded a 50%/50% split with similar workloads.

Furthermore both the LOGS and the additional IOPS (maintenance, replication, etc.) show a reduction in percentage when compared to Exchange Server 2010, and leave more room for Transactional IOPS dedicated straight to the mailbox user activities.

Overall the outcome of the three sets of tests exhibits a nearly linear trend with all Exchange Server 2013 KPI below the warning thresholds. The multiplication of the workload by a factor of 2x or 3x does not correspond to a similar steep increase of the various latencies tracked, but instead showcases a slower increase that allows the successful completion of all validations.

Note: For additional information and to compare the results of this section of the paper with similar scenarios validated on an Exchange Server 2010 platform, refer to *Sizing Microsoft Exchange 2010 on EqualLogic PS6100 and PS4100 Series Arrays with VMware vSphere 5*, available at: <http://en.community.dell.com/dell-groups/dtcmedia/m/mediagallery/20099453/download.aspx>

5.2 Characterize the mailbox size

The mailbox size defines the maximum capacity to which a mailbox is allowed to grow. It is usually enforced by one or more quota policies at the corporate or more granular levels to avoid database overgrowth and volumes or disks capacity saturation. While this value has the most impact on the end-users, the Exchange administrators should factor other components in the equation of the capacity allocated per each user mailbox, including DB white spaces and recoverable items.

- The DB white spaces are an offshoot of continuous mailbox activities and database maintenance. Since the maximum size of the database is not set, the regular user activity allocates database pages which expands the database file or reuses free pages as needed. The database maintenance frees tombstoned objects, and then the online defragmentation consolidates the user data optimizing the B-tree structure, at the same time increasing the amount of white spaces.
- The recoverable items are the items deleted from the end users but not yet purged from the Exchange system according to a set of retention rules (retention window per mailbox items, per calendar items, single item recovery).



The capacity quota per mailbox also influences the storage performance due to the limited or wider physical disk surface that must be accessed to store and retrieve the data, which affects the response time and latency of the IOPS performed against the storage subsystem.

The goal of the following analysis is to establish the Exchange KPI trends and IOPS ratios when increasing the average mailbox size in the reference environment defined previously.

Table 4 shows a summary of the configurations used for this scenario (reference and variations).

Table 4 Test parameters: mailbox size

Reference configuration: factors under study	
Mailbox size	2 GB / 4GB / 6GB each
Database size	2 TB / 4TB / 6TB each
Reference configuration: consistent factors within this scenario	
Messages per day per mailbox / IOPS per mailbox	200 messages / 0.134 IOPS (with DAG)
Number of simulated users / mailboxes	5,000 concurrent users
Number of databases	5 databases (active)
Mailbox allocation per database	1,000 mailboxes per each mailbox database
iSCSI initiator software collocation	Host SW initiator (ESXi 5.5)
Number of database replica copies	2 (two node DAG)
Array model, SAN configuration	1x PS6110E (24x 4TB NL-SAS 7.2K disks), one pool
RAID policy	RAID 6

The three mailbox storage quotas selected for evaluation cover from 2GB to 6GB for each mailbox. Microsoft supports the striking amount of 100GB per mailbox in Exchange Server 2013, up from 10GB per mailbox in the former Exchange Server 2010 version.

Since, for comparison purposes, the ratio of user mailboxes per database is kept constant across the three tests, a side effect is the comparable growth of the underlying database size. The current database size supported by Microsoft is 16TB, while the Microsoft recommended size is 2TB in a DAG configuration. The DB size selection has a direct impact on replication (initial seeding or reseeding operations), backup/restore and recovery times. Therefore, it is highly recommended to select Exchange DB size(s) that align with your business's service level agreements (SLAs), if any.

Note: The maximum size of the mailboxes and the databases that the Microsoft Exchange Jetstress tool can build are lower than the sizes supported in Exchange Server 2013 (16TB for databases, 100GB for mailboxes).



The set of scenarios selected in this section offers the analysis for a trend of large mailboxes but not of very large mailboxes (usually over 10GB each).

Figure 6 shows the results collected from the three instances of Exchange Jetstress simulations.

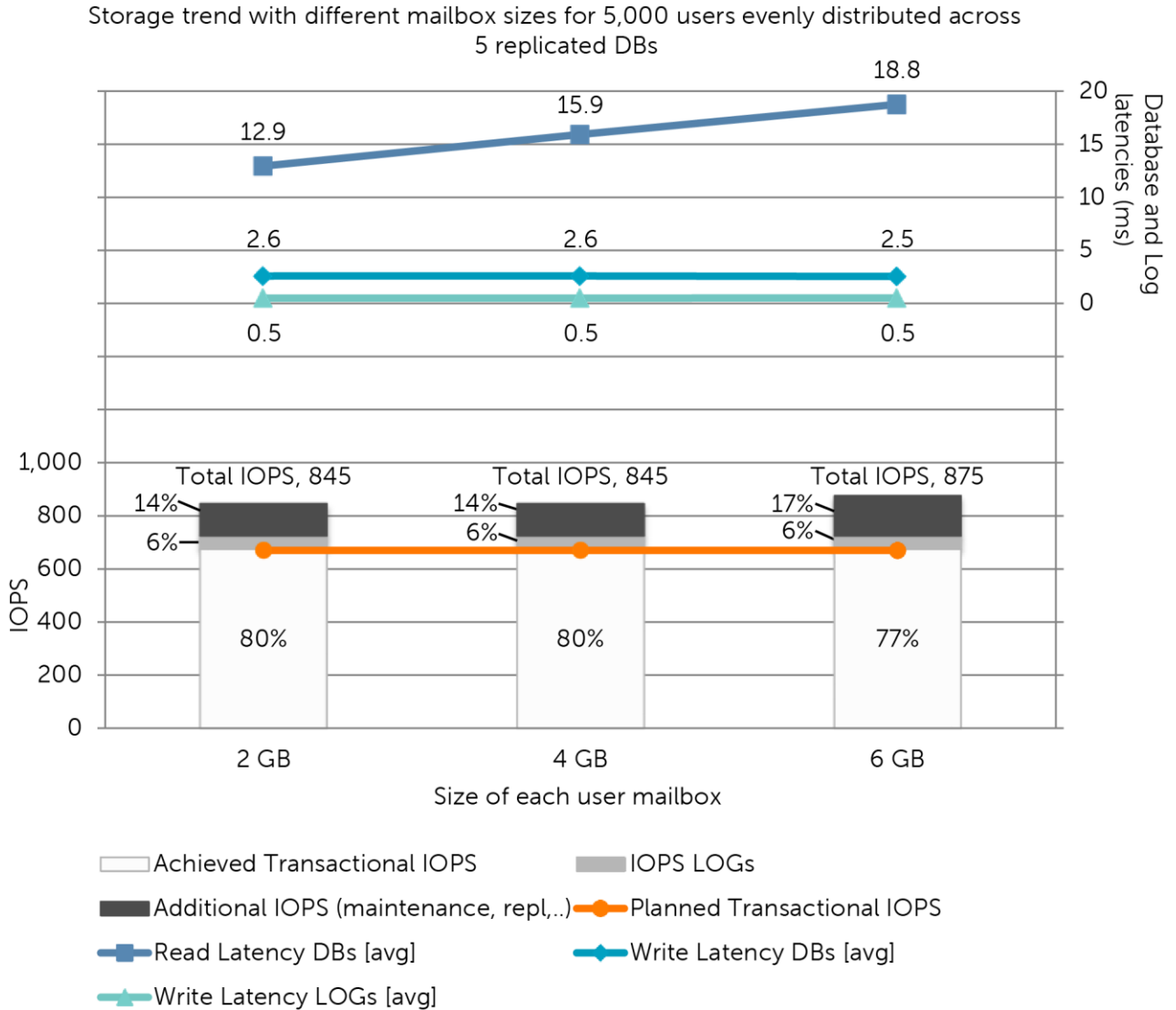


Figure 6 Storage trend with different mailbox size for 5,000 users evenly distributed across five replicated databases

The outcome conveys a steady behavior for the amount of IOPS performed and write latencies at both database and log files level. Alternately, the read latencies show a rise in parallel with the increase of the capacity allocated, although not as relevant as the factor of the mailbox size increase of 2x and 3x. The trend is expected considering the disk surface to be reached by the drive spindle movements is doubled and then tripled.



5.3 Database volumes layout

The Exchange database layout designates the number of databases, the mailbox distributions across the databases and the file placement on the SAN volumes.

- Each Exchange database is managed as a single administrative unit and is assisted by a set of services with a 1:1 ratio (defragmentation, maintenance, logs generation).
- Exchange Server Standard edition scales to five mounted databases per server; Exchange Server Enterprise edition scales to 50 mounted databases in the Release to Manufacturing (RTM) and Cumulative Update 1 (CU1) versions, up to 100 mounted databases in the Cumulative Update 2 (CU2) version.

The mailbox count per database is not bound to a declared limit or range. It is an informed administrative decision to select this count in relation with the administration style and resources of the specific environment.

- The decrease of the mailbox count per database induces the deflation of the average database size with the benefit of streamlining the administrative tasks on a smaller number of users and reducing the timeline for the activities at the file level (replication, recovery). Meanwhile it introduces the drawback of incrementing the number of databases which might negatively reflect on the complexity and overall administration of the messaging infrastructure especially when designed with multi-replicas of the databases.
- The online mailbox move feature allows administrators to seamlessly redistribute mailboxes across databases or mailbox servers as well as postpone or lift constraints imposed by decisions made early in the design and deployment stage of the messaging infrastructure.

The practice of placing both the database and log files within the same volume or disk is now confirmed and consolidated in Exchange Server 2013 with an HA configuration as well as in the former Exchange Server 2010. The two historical reasons to deploy database and log files on split volumes, listed below, have been superseded by a product which is constantly reducing the I/O footprint and which is offering a wide variety of options to protect the data stored.

- Performance: The different I/O pattern of database and log files streams of data (random reads/writes versus sequential writes) and the aim to associate them with the most fitting storage device (for example with the most appropriate disk rotational speed or RAID policy)
- Reliability: The simultaneous loss of both the database and log files could jeopardize the recoverability of user data depending on the data protection tactic in place

Note: For additional information about Exchange Server data protection options with Dell EqualLogic SANs, refer to *Best Practices for Enhancing Microsoft Exchange Server 2010 Data Protection and Availability using Dell EqualLogic Snapshots*, available at: <http://en.community.dell.com/dell-groups/dtcmedia/m/mediagallery/19961113/download.aspx>



The goal of the following analysis is to establish the Exchange KPI trends and IOPS ratios when increasing the number of databases while reducing the ratio of mailboxes per database in the reference environment defined previously.

Table 5 shows a summary of the configurations used for this scenario (reference and variations).

Table 5 Test parameters: database deployment layout

Reference configuration: factors under study	
Number of databases	5 / 10 /20 databases (active)
Mailbox allocation per database	1,000 / 500 / 250 mailboxes per each mailbox DB
Database size	2 TB / 1TB / 500GB each
Reference configuration: consistent factors within this scenario	
Messages per day per mailbox / IOPS per mailbox	200 messages / 0.134 IOPS (with DAG)
Number of simulated users / mailboxes	5,000 concurrent users
Mailbox size	2 GB each
iSCSI initiator software collocation	Host SW initiator (ESXi 5.5)
Number of database replica copies	2 (two node DAG)
Array model, SAN configuration	1x PS6110E (24x 4TB NL-SAS 7.2K disks), one pool
RAID policy	RAID 6

The three deployment scenarios selected specify an increasing number of databases while maintaining the total amount of users per server at 5,000. The progression in the configuration changes includes number of databases, user count per database and the resulting mailbox database size moving from five databases (1,000 users per database, mailbox databases of 2TB), to ten databases (500 users per databases, mailbox databases of 1TB), and then to 20 databases (250 users per databases, mailbox databases of 5000GB).

By default, Exchange Server 2013 activates the maintenance tasks on each new mailbox database created. Most of these activities are important to the health and efficiency of the databases and would not be deactivated. The maintenance overhead must be considered when increasing the number of databases,

The trend of this set of tests shows the tasks running on the additional databases still influence the overall balance between Transactional IOPS and the rest of the IOPS, similar to the previous Exchange Server versions. However the performance decline is contained and limited to the write latency of the databases, while the log files performance is intact and the read latency of the databases shows a small increase when the number of databases grows.



The overall impact of the practice of using many databases is mitigated in this version compared to the past and might be now more feasible in an environment where the very large mailboxes are used and the need to reduce the size of the single databases is vital.

Figure 7 shows the results collected from the three instances of Exchange Jetstress simulations.

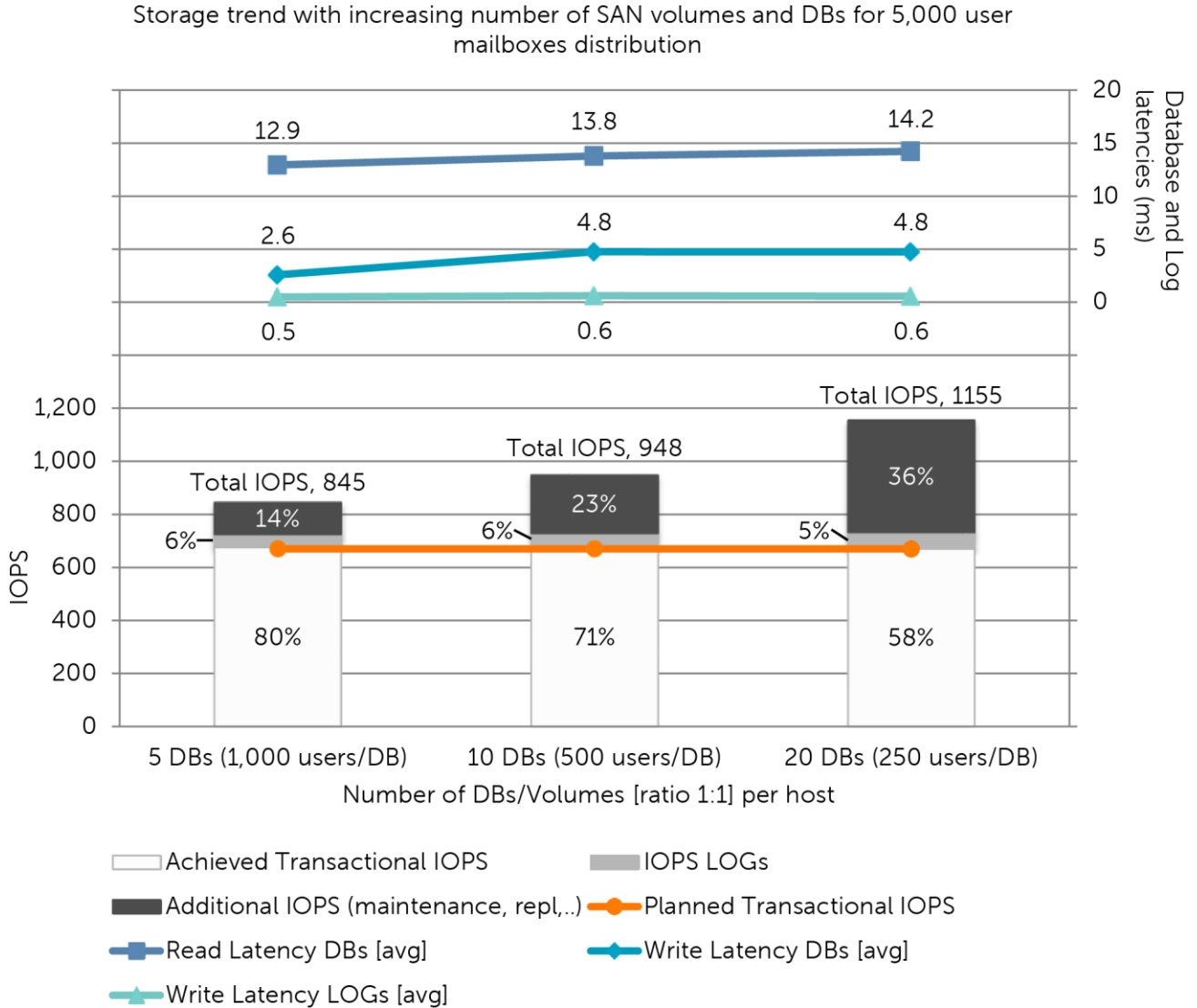


Figure 7 Storage trend with increasing number of SAN volumes and DBs for 5,000 user mailboxes distribution

5.4 Scale up the user count per server

The mailbox count represents the global number of concurrent users hosted by the mailbox server and stored on the corresponding SAN array. When the mailbox count increases, the amount of IOPS, capacity allocated, and log files generated increase.



Additionally, when a large number of mailboxes are hosted by a single server, the demand for a highly available solution is intensified due to the risk of widely distributed loss of the messaging service in case of server failure.

The goal of the following analysis is to establish the Exchange KPI trends and IOPS ratios when increasing the number of concurrent users and thus the global workload on the SAN in the reference environment defined previously.

Table 6 shows a summary of the configurations used for this scenario (reference and variations).

Table 6 Test parameters: variable number of mailbox concurrent users

Reference configuration: factors under study	
Number of simulated users / mailboxes	4,000 / 5,000 / 6,000 concurrent users
Number of databases	4 / 5 / 6 databases (active)
Reference configuration: consistent factors within this scenario	
Messages per day per mailbox / IOPS per mailbox	200 messages / 0.134 IOPS (with DAG)
Mailbox size	2 GB each
Mailbox allocation per database	1,000 mailboxes per each mailbox database
Database size	2 TB each
iSCSI initiator software collocation	Host SW initiator (ESXi 5.5)
Number of database replica copies	2 (two node DAG)
Array model, SAN configuration	1x PS6110E (24x 4TB NL-SAS 7.2K disks), one pool
RAID policy	RAID 6

The building block used to scale up the number of users for this set of tests is the addition of databases, while maintaining the users per database ratio at 1,000:1. The three progressively increasing workloads scale from 4,000 to 6,000 concurrent mailbox users against the SAN built on the single reference array. A database of 2TB, hosting an additional 1,000 users, is added at each subsequent sample.

Gradually provisioning predefined-sized mailbox databases simplifies the administrative burden when the demand to support more users increases and conveniently increments the workload in a linear fashion.

The variation in Exchange database cache estimated for these three scenarios is reported in 0, and again does not consider other factors, services and configurations of the Exchange mailbox server. The estimates for the database cache per mailbox are based on Microsoft published metrics and not on recorded values from the tests.

The outcomes of this series reveal a nearly linear scalability on the entire series for both of the write latencies under consideration. The read latency keeps a good pace on the first scale up step, but then



results in a steeper than expected increase. The additional transactional IOPS, which are random in nature, push the responsiveness of the large 4TB NL-SAS drives up, but still comfortably within the predefined threshold.

Figure 8 shows the results collected from the three instances of Exchange Jetstress simulations.

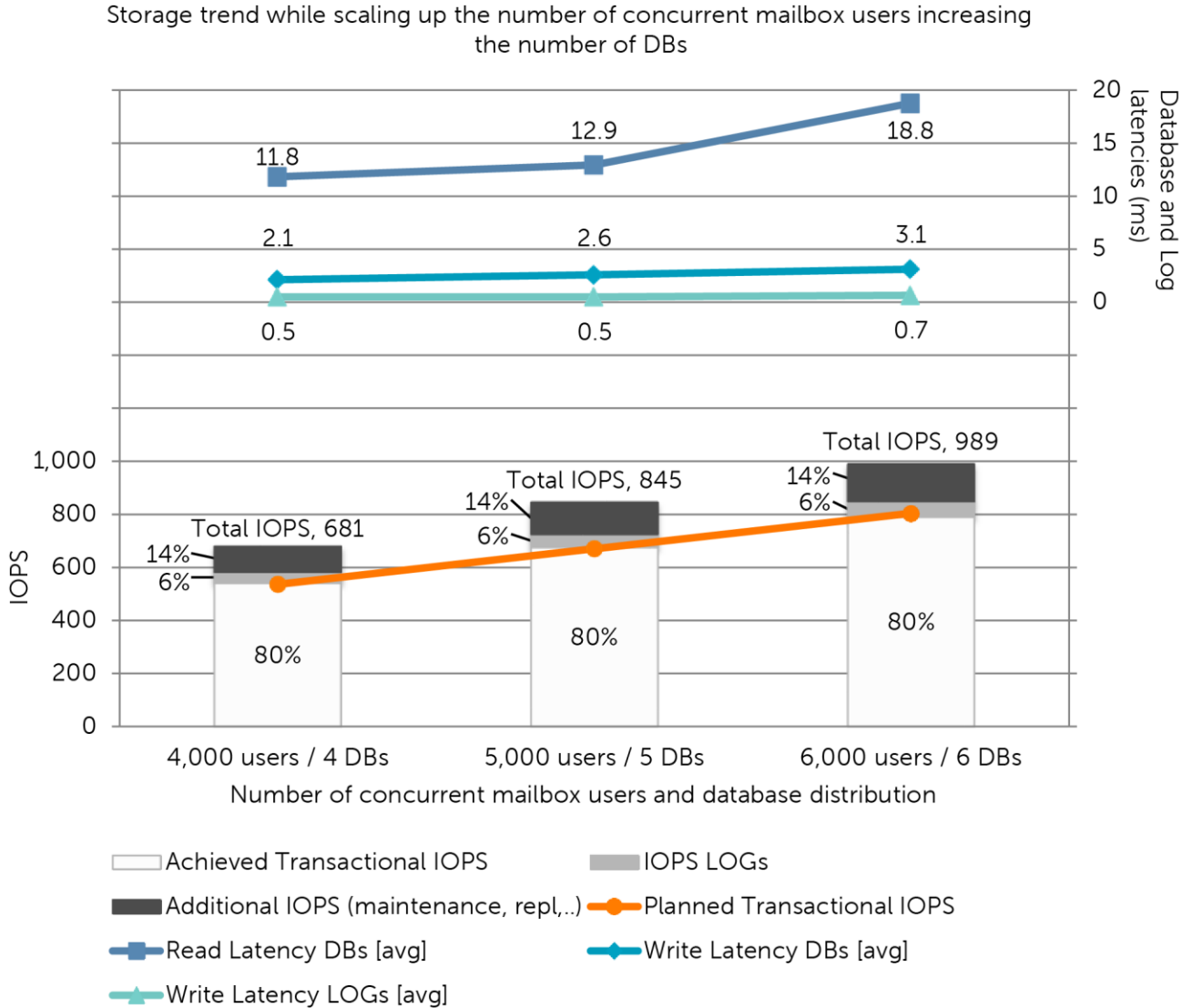


Figure 8 Storage trend while scaling up the number of concurrent mailbox users increasing the number of databases

The variation in Exchange database cache estimated for these three scenarios is reported in 0, and again does not consider other factors, services and configurations of the Exchange mailbox server. The estimates for the database cache per mailbox are based on Microsoft published metrics and not on recorded values from the executed tests.



Table 7 Exchange database cache estimated per server while scaling up the number of users

Number of concurrent mailbox users	Cache per user	Exchange database cache estimated
4,000	24MB	96GB
5,000		120GB
6,000		144GB

Note: Microsoft Exchange Jetstress memory and processors utilization performs differently from an Exchange Server. Some details of Microsoft Jetstress tool are reported in Appendix B.

5.5 Characterize the impact of the iSCSI initiator collocation

In networked storage systems, the iSCSI communications happen between two layers or end nodes through the wire. The iSCSI initiator functions as a client, which accesses the storage resources located on the target. The target acts as a server hosting the data. Messages and commands, similar to SCSI, are sent over the IP network between the two end nodes.

A software initiator is an implementation of these functionalities by code. It usually runs as a device driver and reuses the network cards available in the operating system to emulate SCSI devices. While implementing a virtual infrastructure, a choice for the connectivity with the EqualLogic SAN through a software based iSCSI initiator would be the collocation of the initiator itself.

- Guest initiator: software located and running on the guest VMs, which allows direct connections to the volumes residing on the SAN through the virtual network adapters of the guest. The settings of the VMs include additional virtual network adapters dedicated to SAN traffic. The host hypervisor is not aware of the type of traffic traversing the virtual network adapters.
- Host initiator: software located and running on the host hypervisor, which allows you to connect to the volumes residing on the EqualLogic SAN from the ESXi host through physical network adapters dedicated to the SAN traffic. VMDK files are created on the SAN volumes attached to the host and then added as SCSI disks to the settings of the VMs. The VMs are not aware of where their disks reside, either on local storage or the SAN.

The goal of the following analysis is to establish the Exchange KPI trends and IOPS ratios when comparing a deployment implemented through host iSCSI initiator against one implemented with a guest initiator in the reference environment defined previously.

Table 8 shows a summary of the configurations used for this scenario (reference and variations).



Table 8 Test parameters: iSCSI software initiator collocation

Reference configuration: factors under study	
iSCSI initiator software collocation	Host SW initiator (ESXi 5.5)
	Guest initiator (Microsoft Windows Server 2012)
Reference configuration: consistent factors within this scenario	
Messages per day per mailbox / IOPS per mailbox	200 messages / 0.134 IOPS (with DAG)
Number of simulated users / mailboxes	5,000 concurrent users
Mailbox size	2 GB each
Number of databases	5 databases (active)
Mailbox allocation per database	1,000 mailboxes per each mailbox database
Database size	2 TB each
Number of database replica copies	2 (two node DAG)
Array model, SAN configuration	1x PS6110E (24x 4TB NL-SAS 7.2K disks), one pool
RAID policy	RAID 6

The comparison shows a noticeable performance penalty for the guest iSCSI initiator. These tests demonstrate the good compatibility of the VMDK file format and the VMFS-5 file system with the access pattern required by Exchange Server 2013 databases and log files. Although the storage access must traverse multiple logical layers, the performance is kept higher than when a virtualized network deployment is in use. The future implementation of network features (for example SR-IOV) exploiting more direct access to the physical network adapters would definitely benefit a solution implemented with guest initiator in mind.

Figure 9 shows the results collected from the two instances of Exchange Jetstress simulations.



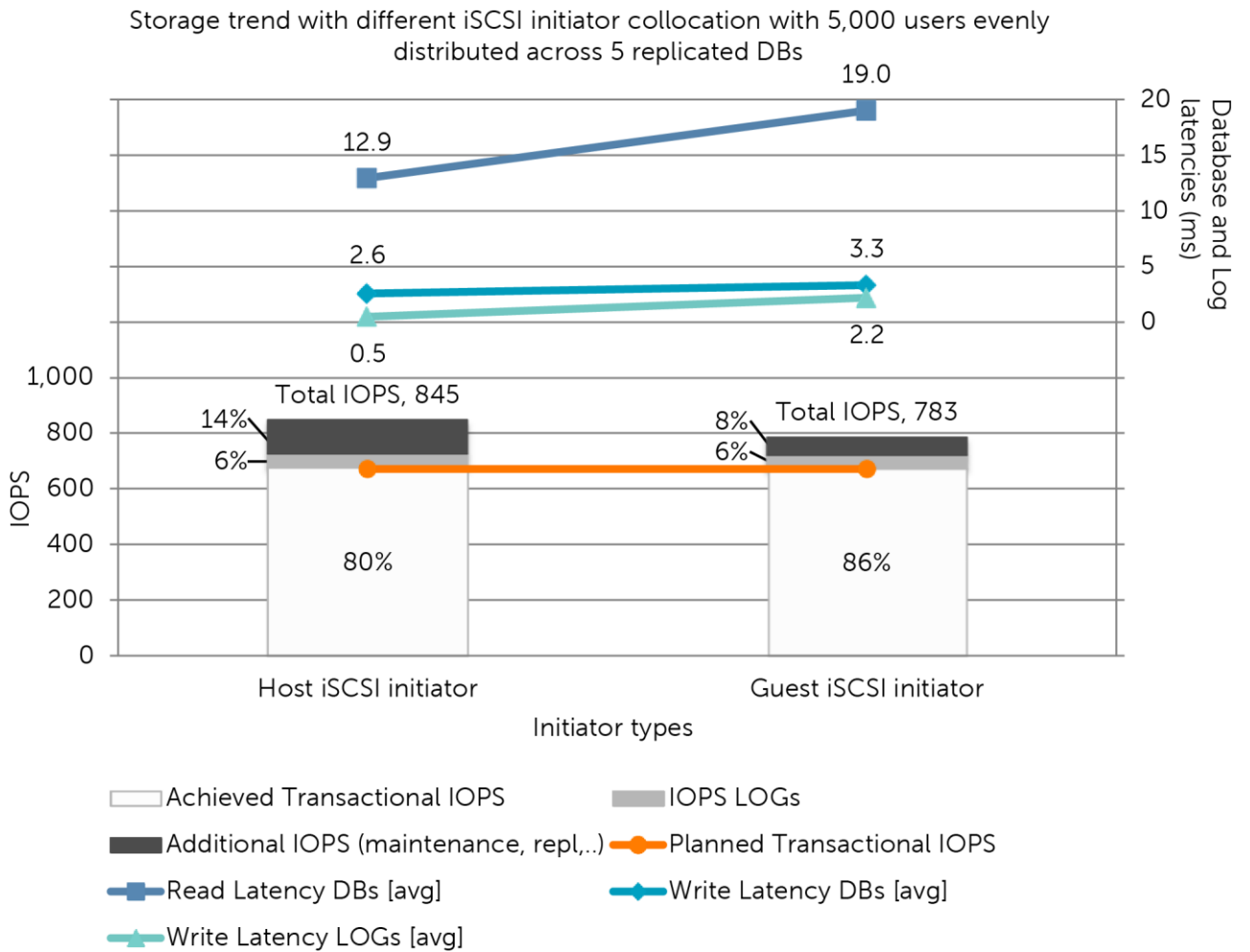


Figure 9 Storage trend with different iSCSI initiator collocation with 5,000 users evenly distributed across 5 replicated databases

Note: The approach to select one initiator versus another is not driven only by performance results. Considerations about how to protect the data from within the guest or from the host have to be weighed together with the level of granularity of the backup/restore offered in either scenario.



6 Best practices recommendations

Refer to these best practices to plan and configure EqualLogic arrays, Exchange Server 2013 and VMs.

Storage best practices

- Distribute network port connections on the controllers according to the port failover guidelines and the redundancy implemented on the network switches.
- Maintain a close as possible to a 1:1 ratio between the number of network ports on the active controller of the array and the number of host network adapters to maximize the utilization of the available bandwidth and to minimize oversubscription.
- Use a Multipath I/O (MPIO) extension provided by EqualLogic to improve iSCSI performance and reliability. This is a Device Specific Module (DSM) in Windows or Extension Module (MEM) in ESXi.
- Carefully choose the most appropriate RAID policy when designing the environment according to the performance, capacity, and tolerance to failure requirements of your environment.
- Do not share the disk drives for active and replicated copies of an Exchange mailbox database in a DAG environment. If there is a failure of a set of drives with multiple copies of the same data, the resilience or the perceived availability of the applications would be affected. Dedicate separate EqualLogic pools for each replicated instance of the data instead.
- For environments with a very large number of volumes attached to the servers, monitor the amount of iSCSI connections per host generated.

Network best practices

- Design separated network infrastructures to isolate the LAN traffic from the SAN traffic (iSCSI).
- Secure the networks, including the SAN, to protect your resources from external threats such as unwanted packet sniffing or connection hijacking. Security mechanisms should include physical security, authentication and application level encryption.
- Implement redundant components (switches, ISLs, and network adapters) to provision a resilient infrastructure between the endpoints (stack, LAG, load balancing, or network card teaming).
- Disable spanning tree for the switch ports hosting the PS Series array controllers' connections and enable Rapid Spanning Tree Protocol instead.
- Enable flow control for the switch ports hosting the PS Series array controller connections.
- Enable jumbo frames (large MTU) for the switch ports hosting the PS Series array controller connections (in conjunction with the following host and guest related jumbo frames configuration).

ESXi and VMs best practices

- The use of NUMA is advised to address the management of VMs with large and very large memory settings. Verify the number of NUMA nodes available in the system, based on the number of processors, and then design and size the VMs to have their memory resources entirely contained in a single NUMA node. Spanning a VM memory across multiple NUMA nodes can result in less efficient usage of the memory and can decrease performance.



- Exchange Server 2013 is a memory intensive workload. Do not overcommit memory on the ESXi host to avoid guests' contention for memory resources, which could penalize the overall VMs performance and storage I/O execution. Lock all memory assigned to the Exchange servers in the resource allocation settings, and size all the hosts accordingly in case of an HA environment with vMotion enabled.
- Carefully reserve the size for the volumes hosting the hard disk files of the VMs (*.vmdk), and include the space required for swap (*.vswp) and log files (*.log). The higher the memory configuration, the larger the swap files will be.
- Remember the supportability constraints of Exchange Server 2013 when deployed in a virtualized environment around suspended state, VM snapshots, and non-persistent disk image files.
- Do not mix VMware thin provisioned virtual disk drives and thin provisioned volumes in the same deployment. In cases with space usage challenges, thin provisioned volumes are recommended.
- Avoid mixing LAN and iSCSI traffic on the same virtual switch and enforce it by the LAN and iSCSI network isolation design.
- Configure Port Group and VLAN tagging to logically segregate different LAN traffic and/or host management traffic.
- Isolate the virtual switches dedicated to host and guest iSCSI initiator from each other.
- Configure the virtual switch dedicated to host iSCSI initiator with the support of the EqualLogic MEM wizard.
- Configure a dedicated Port Group for each virtual network adapter connected to the SAN traffic you plan to have in the VM, and add one physical network adapter for every Port Group isolating it by the override switch failover option.
- Reserve at least two network adapters for each virtual switch to achieve resiliency.
- Select performance optimized network adapters of VMXNET3 type for guest network adapters connected to the SAN network (VMware tools required in the guest operating system).
- Enable jumbo frames (large MTU) for the virtual switches, port groups and VM network adapters assigned to SAN traffic (iSCSI).
- Evaluate jumbo frames (large MTU) for the virtual switches, port groups and VM network adapters assigned to LAN traffic (limited by the type of devices the traffic traverses).
- Enable flow control on the host network adapters and on the VM virtual adapters dedicated to SAN traffic (iSCSI).
- Enable Large Send Offload, IP, TCP, and UDP Checksum Offload for both RX and TX on the VM network adapters connected to the SAN traffic (iSCSI).

Exchange Server installation best practices

- Use Basic disk type for all EqualLogic volumes.
- Use GUID partition table (GPT) for Exchange volumes.
- Use default disk alignment provided by Windows 2008 or greater.
- Use NTFS file system with 64 KB allocation unit for Exchange database and log partitions.
- Evaluate the use of mount points for all the SAN volumes or the attached virtual disks to increase management flexibility and database portability. Mount points are required when the number of volumes is greater than the number of available drive letters in the servers.



- Prevent Windows Server from assigning drive letters to volumes by disabling the auto-mount option to minimize unwanted volume letter assignment in a mount point managed environment.
- Deploy Windows operating system and Exchange data in physically separated disk drives, or in separated pools for an EqualLogic storage environment.
- Database and log file isolation is not required when deployed in a DAG environment.
- Leave background database maintenance (BDM) enabled (24x7) and account for the additional load. The BDM is activated by default on every replica copy of your DAG configuration.

Know your workload

- Do not begin a deployment without having a solid understanding of your current messaging workload. In the case of a greenfield deployment, collect estimates based on business cases matching your organization size and drive conservative figures for the average user profiles.
- Forecast the workload gap between your current version of Exchange Server and Exchange Server 2013 when you plan to design a storage solution jointly with a migration.

Distribution of databases and user mailbox count per database

- Plan carefully the number of databases and the mailbox count per database in the entire organization. The right balance of the number of mailbox databases to support a defined number of users is mostly based on administrative policies.
- Larger databases fitted with a high number of users have a smaller impact on the storage subsystem, but it provides less flexibility because of large database files maintenance and replication.
- Reduced mailbox count per database provides a more agile environment to administer when using a traditional backup application or when administrative tasks require temporarily dismounting a database causing a downtime. Deploying many users in a single database is perceived as a single point of failure or as a constraint.
- Microsoft Windows operating system supports a precise maximum amount of iSCSI targets and connections: when a guest initiator is in use together with a great number of volumes, monitor the number of connections and do not pass the limits to avoid an unsupported deployment scenario.

Growth and overhead

- Understand your wider business requirements and company strategy to predict the organization growth instead of following it.
- Design your environment based on building blocks that are easily replicable to address sudden changes in business requirements.
- Always account for at least 20% overhead of the maximum sizes calculated to address unforeseen growth.



A Configuration details

A.1 Hardware components

Table 9 lists the details of the hardware components used for the configuration.

Table 9 Hardware components

Component	Description
Servers	<p>Dell PowerEdge R620 server, Firmware 2.0.19</p> <ul style="list-style-type: none">• 2x Eight Core Intel Xeon E5-2665 Processors, 2.4 Ghz, 20M Cache• RAM 32 GB (4x 8GB)• iDRAC7 Enterprise, Firmware 1.46.45• PERC H710 Mini RAID controller, Firmware 21.2.0-0007• 4x 146 GB 15K SAS (2x RAID-1, stripe 1MB)• 4x Broadcom NetXtreme 5720 Quad Port 1GbE Base-T onboard, Firmware 7.6.15• 2x Broadcom NetXtreme II 57810 Dual Port 10GbE Base-T, Firmware 7.6.15 <p>Dell PowerEdge R720 servers, Firmware 2.0.19</p> <ul style="list-style-type: none">• 2x Eight Core Intel Xeon E5-2665 Processors, 2.4 Ghz, 20M Cache• RAM 192 GB (24x 8GB)• iDRAC-7 Enterprise, Firmware 1.46.45• PERC H710 Mini RAID controller, Firmware 21.2.0-0007• 4x 146 GB 15K SAS (2x RAID-1, stripe 1MB)• 4x Broadcom NetXtreme 5720 Quad Port 1GbE Base-T onboard, Firmware 7.6.15• 2x Broadcom NetXtreme II 57810 Dual Port 10GbE Base-T, Firmware 7.6.15
Network	<p>2x Dell Force10 S4810 Ethernet switches, Firmware 9.2.0.0</p> <ul style="list-style-type: none">• 48x 10GbE interfaces• 4x 40GbE interfaces• Installed top of the rack• Connected by 2x 40GbE redundant uplinks (LAG) <p>Dell PowerConnect 7048R Ethernet switch, Firmware 5.1.1.7</p> <ul style="list-style-type: none">• 48x 1GbE interfaces• 4x 10GbE interfaces• Installed top of the rack• Connected by 2x 16GbE redundant uplinks (STACK)
Storage	<p>1x Dell EqualLogic PS6110XV 3.5"</p> <ul style="list-style-type: none">• Storage Array Firmware 6.0.7• Dual 1 port 10GbE controllers• Dual 1 port 1GbE management interface• 24x 600GB 15K 3.5" SAS disk drives, raw capacity 14.4 TB <p>1x Dell EqualLogic PS6110E</p> <ul style="list-style-type: none">• Storage Array Firmware 6.0.7



	<ul style="list-style-type: none"> • Dual 1 port 10GbE controllers • Dual 1 port 1GbE management interface 24x 4TB 7.2K 3.5" NL-SAS disk drives, raw capacity 96 TB
--	--

A.2 Software components

The environment required to perform the simulations described in this paper included the following software components:

- Hypervisor: VMware 5.5 ESXi on every physical host
- Operating System: Windows Server 2012 on workload simulation VM, Windows Server 2008 R2 on management VM
- Dell EqualLogic Host integration Toolkit to provide Dell MPIO access to the back-end SAN on the VM directly accessing the SAN (guest initiator scenario)
- Dell EqualLogic Multi-pathing Extension Module to provide Dell MPIO access to the back-end SAN on the hypervisor directly accessing the SAN (host initiator scenario)
- Dell EqualLogic SAN Headquarters to monitor the health and performance of the SAN
- Microsoft Exchange Jetstress to simulate the access to the storage subsystem from the mailboxes store simulated VM

The following software components are installed and configured to simplify the management of the environment and to support the failover cluster configuration, while they were not strictly required to accomplish the tests:

- Active Directory Domain Services and DNS Server roles for the domain controller VM
- VMware vSphere vCenter Server and client for the management VM

Table 10 lists the details of the software components used for the configuration.



Table 10 Software components

Component	Description
Operating systems	<p>Host servers:</p> <ul style="list-style-type: none"> VMware ESXi 5.5.0, build 1331820 Dell Multipathing Extension Module (MEM) 1.2 <p>Guest VMs:</p> <ul style="list-style-type: none"> Microsoft Windows Server 2012 Standard Edition (build 9200) Microsoft Windows Server 2008 R2 Datacenter Edition Service Pack 1 (build 7601) VMware Tools for Windows 9.4.0, build 1280544 Dell EqualLogic Host Integration Toolkit 4.6.0 (only for guest initiator connectivity) MPIO enabled using EqualLogic DSM for Windows
Applications	<p>VMware vSphere vCenter Server 5.5.0, build 1312298</p> <p>VMware vSphere vCenter Client 5.5.0, build 1281650</p>
Monitoring tools	<p>Dell EqualLogic SAN Headquarters 3.0 EPA (build 2.6.0.6947)</p> <p>Microsoft Performance Monitor from the Windows Operating System</p>
Simulation tools	<p>Microsoft Exchange Jetstress 2013 (build 15.00.0775.000)</p> <ul style="list-style-type: none"> Exchange 2013 Server Database Storage Engine and Library CU2 (build 15.00.0712.008)

A.3 Network configuration

Two physical networks provide full isolation between regular IP traffic and iSCSI data storage traffic. Also, each regular IP network is segregated from the others by VLANs with tagged traffic. In order to achieve network resiliency to hardware faults, at least two physical switches are linked for the iSCSI data storage network using redundant uplinks (LAG) between the switches. Some relevant configuration aspects are:

- Flow control enabled for every port on S4810 switches
- Rapid Spanning Tree Protocol enabled for every edge port on S4810 switches
- Jumbo frames enabled for every port on S4810 and 7048R switches

Table 11 and Table 12 summarize the aspects of the physical and logical networks implemented in the reference architecture and their purpose.



Table 11 Network configuration: network switch, purpose, and networks

Network Switch	Placement	Purpose	VLAN ID
PowerConnect 7048R	Top of rack	IP Traffic - LAN	100
		IP Traffic - Management	200
Force10 S4810 #1	Top of rack	iSCSI data storage traffic	1 (default)
Force10 S4810 #2	Top of rack	iSCSI data storage traffic	1 (default)

Table 12 Network configuration: host to switch connections

Server	Interface	NIC port	Purpose	
PowerEdge R620	BCM5720 #1	1x 1GbE	VM traffic and Management	IP traffic
	BCM5720 #2	1x 1GbE		
	BCM5720 #3	1x 1GbE		
	BCM5720 #4	1x 1GbE		
	BCM57810 #1	1x 10GbE	SAN management (guest VM)	iSCSI data storage traffic
	BCM57810 #2	1x 10GbE		
Total ports = 6 (4 onboard 1GbE, 2 slot 10GbE)				
PowerEdge R720	BCM5720 #1	1x 1GbE	VM traffic and Management	IP traffic
	BCM5720 #2	1x 1GbE		
	BCM5720 #3	1x 1GbE		
	BCM5720 #4	1x 1GbE		
	BCM57810 #1	1x 10GbE	SAN access (host initiator)	iSCSI data storage traffic
	BCM57810 #2	1x 10GbE		
	BCM57810 #3	1x 10GbE	SAN access (guest initiator)	
	BCM57810 #4	1x 10GbE		
Total ports = 8 (4 onboard 1GbE, 4 slot 10GbE)				

A.4 Host hypervisor and VMs configuration

A virtual infrastructure built on VMware ESXi hosted all the components of the test infrastructure. The primary elements of the virtual infrastructure configuration are:



- VMware ESXi deployed on all hosts, managed by the vSphere Client or centrally by the vCenter server
- All guests deployed from two image templates (one per OS version) of Windows Server 2008 R2 and Windows Server 2012 operating systems
- Host iSCSI initiator to access the system disks (VMDK) of the VMs, which reside on the EqualLogic PS6110XV array
- Host or guest iSCSI initiator, depending on the appropriate scenario, used to access the volumes containing data and hosted on the EqualLogic PS6110E array

Table 13 lists the relation between each hypervisor host and its respective set of VMs, with a brief summary of the virtual resources allocated for each VM.

Table 13 Configuration: guest to host placement

Host	VM	Purpose	vCPU	Memory	Storage	Network adapters
R620	DC01	Active Directory Domain Controller	2	4GB	25GB VMDK	1x E1000
	vCENTER	Virtual Infrastructure management	2	8GB	25GB VMDK	2x E1000
R720	MBX01	Exchange Server mailbox server simulation (Jetstress)	4	64GB	100GB VMDK	1x E1000
						2x VMXNET3 (iSCSI traffic)

VMware ESXi configuration of NUMA

Non-Uniform Memory Access (NUMA) capabilities are enabled on the PowerEdge R720 physical hosts (Node Interleaving disabled in the server BIOS) to allow memory access across CPUs. This R720s has two NUMA nodes each managing 96GB of memory. No specific affinity is defined for the VM hosting the simulated Exchange Server workload, shown in Figure 10, since no contention with other VMs is present in this testing scenario and the VM resources are furthermore fully contained in one NUMA node.

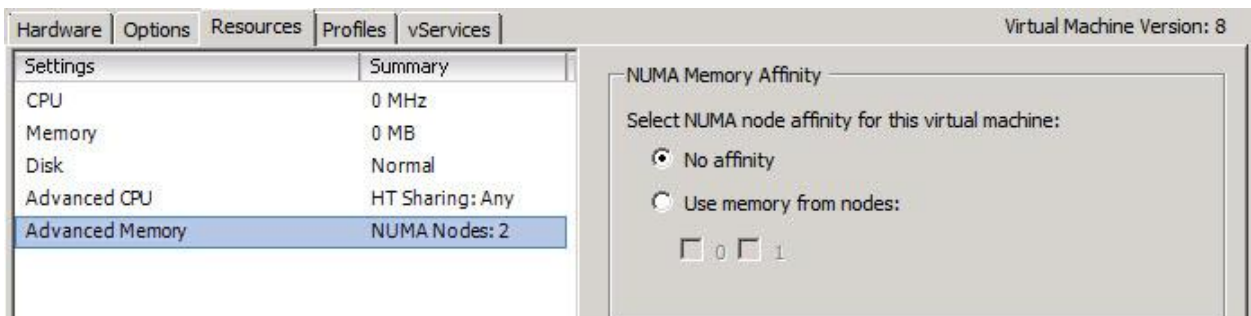


Figure 10 ESXi setting for NUMA



Datstores and VM disks settings

All the datstores used on the hypervisor are natively formatted with a file system version 5 (VMFS-5) with the maximum available space of the LUN allocated on the first instance.

All the virtual machine disks (VMDK) are provisioned as 'Thick' and stored in the according LUN on the SAN. The VMDK files representing the boot disk of the operating systems are hosted on the PS6110XV array, the remaining VMDK files created to support the Exchange data (for host initiator use case) are hosted on the PS6110E. The use of vSphere Web Client was required in order to create VMDK files larger than 2TB.

Host network adapters and virtual network configuration

The host network adapters providing connectivity for the hosts and the VMs are configured as listed.

- One standard virtual switch aggregating all the network adapters designated for IP traffic (onboard Broadcom 1GbE ports), with the default load balancing policy 'Route based on the originating virtual switch port ID' selected
 - 'Management Network' port for all hypervisors (VMkernel port type)
 - 'vCenters' port group (Virtual Machine type), only for the R620 hypervisor
 - 'VM LAN traffic' port group for all hypervisors (Virtual Machine type)
 - Each port group traffic segregated by tagged packets (VLANs)
- One standard virtual switch aggregating two network adapters designated for iSCSI traffic-host initiator (Broadcom 10GbE ports), configured interactively with EqualLogic Multipathing Extension Module, which leaves the default load balancing policy 'Route based on the originating virtual switch port ID' selected as seen in Figure 11
 - One 'iSCSI traffic' port group for each network adapter in the v-switch (VMkernel type), for the R720 hypervisor
 - Port groups with Load balancing and Failback disabled, and the switch failover default order overridden allowing only one active adapter while the other/s are kept unused (Figure 13)
 - Moreover port groups are bound to the iSCSI Software adapter of the ESXi software initiator

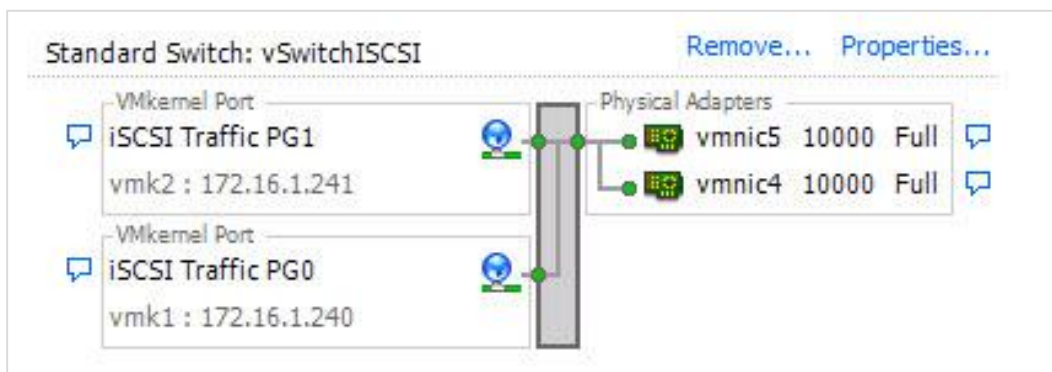


Figure 11 Configuration – vSwitch dedicated to iSCSI traffic-host initiator on R720



- One standard virtual switch aggregating two network adapters designated for iSCSI traffic-guest initiator (Broadcom 10GbE ports), with the default load balancing policy 'Route based on the originating virtual switch port ID' selected (Figure 12)
 - One 'iSCSI traffic' port group for each network adapter in the v-switch (VM type) for the R720 hypervisor
 - Port groups with Load balancing and Failback disabled, and the switch failover default order overridden allowing only one active adapter while the other/s are kept unused (Figure 13)

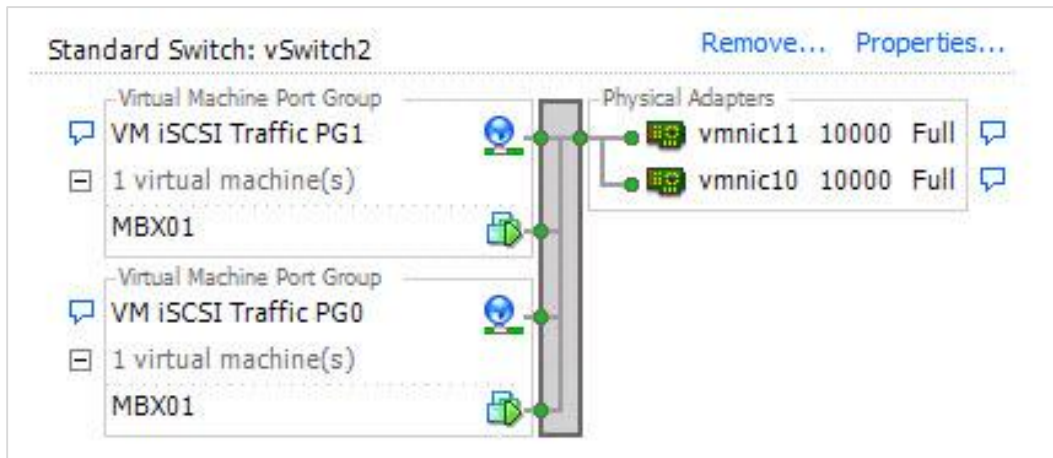


Figure 12 Configuration – vSwitch dedicated to iSCSI traffic-guest initiator on R720



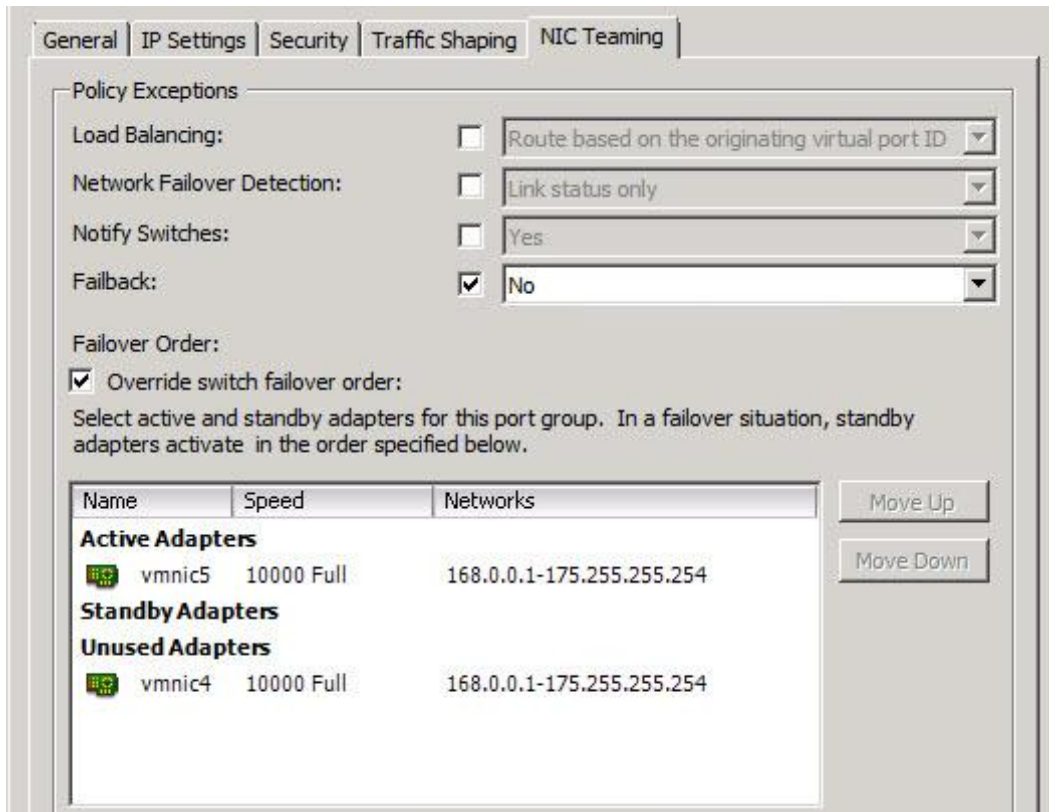


Figure 13 Configuration – Port group details snapshot

Furthermore, the following details are configured across network adapters and virtual switches:

- Jumbo frames are enabled on each virtual switch and/or port group of the hypervisors
- Flow control on received and transmitted traffic (RX/TX) is enabled on each physical network adapter
- Virtual adapters within VMs used for guest iSCSI initiators have
 - MPIO (Multi-path I/O) enabled and provided by EqualLogic DSM module
 - Jumbo frames enabled
 - Flow control (both RX/TX) enabled
 - Receive and transmit buffers maximized
 - Large Send Offload enabled
 - IP, TCP and UDP Checksum Offload enabled (both RX/TX)



B Microsoft Jetstress considerations

Microsoft Exchange Server Jetstress 2013 is a simulation tool able to reproduce the database and logs I/O workload of an Exchange mailbox role server. It is usually used to verify and validate the conformity of a storage subsystem solution before the full Exchange software stack is deployed. Some elements worth being considered about Microsoft Jetstress are:

- Does not require and should not be hosted on a server where Exchange Server is running
- Performs only Exchange storage access and not host processes simulations. It does not contribute in assessing or sizing the Exchange memory and processes footprints
- Is an ESE application requiring access to the ESE dynamic link libraries to perform database access. It takes advantage of the same API used by the Exchange Server application software stack and as such it is a reliable simulation application
- Runs on a single server. When a multiple servers simulation is required, the orchestration of the distributed instances has to be fostered by external management tools
- Requires, and provides, an initialization step to create and populate the database/s that will be used for the subsequent test phases. The database/s should be planned of the same capacity as the one/s planned for the Exchange Server future deployment
- Its topology layout includes number and size of simulated mailboxes, number and placement of databases and log files, number of database replica copies (it simulates only active databases)
- While carrying out a mailbox profile test, it executes a pre-defined mix of insert, delete, replace and commit operations against the database objects during the transactional step, then it performs a full database checksum
- Collects Application and System Event Logs, performance counter values for the criteria metrics of both operating system resources and ESE instances during transactional and DB checksum phases. It then generates a detailed HTML-based report
- Throttles the disk I/O generation using the assigned IOPS per mailbox, thread count (global per all databases) and SluggishSessions threads property (fine tuning for threads execution pace)



C 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 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 Microsoft publications:

- Tools for performance and Scalability Evaluation:
<http://technet.microsoft.com/en-us/library/dd335108.aspx>

