Hosted Microsoft® Exchange Server 2003 Deployment Utilizing Network Appliance™ Storage Solutions

Large-Scale, 680,000-Mailbox Exchange Server Proof of Concept

Lee Dorrier, Network Appliance, Inc.

Eric Johnson, Network Appliance, Inc.

May 2007 | TR-3569

Executive Summary

Large, hosted Exchange Server messaging environments require high-performance system architectures in order to meet or exceed customer performance SLAs. Effectively controlling cost is also an important factor in building and maintaining large messaging server configurations that are typical in hosted messaging environments. This technical report describes a "proof of concept" test of a 680,000-user Microsoft Exchange Server 2003 hosted messaging solution for a large European telecommunications and messaging service provider. The solution leverages Network Appliance FAS storage systems in conjunction with NetApp data management software. The configuration utilized the iSCSI storage networking protocol for a cost-effective yet high-performance architecture that met the rigorous performance requirements of the customer.

Table of Contents

1. Introduction

Network Appliance storage systems are widely deployed in messaging solutions of all types and sizes. NetApp technologies and software products such as multiprotocol support (iSCSI, Fibre Channel, NFS), FlexVol® volumes, RAID-DP™, SnapDrive® for Windows™, SnapManager® for Microsoft Exchange, and Single Mailbox Restore (SMBR) for Microsoft Exchange provide a comprehensive solution set for deploying and managing storage in Exchange Server messaging environments. The NetApp Exchange Server solution set is especially applicable to the challenges imposed by large-scale hosted messaging environments:

- High performance
- Economy and scalability
- Rapid data provisioning
- Simplified storage management

Based on the strength of the NetApp solution set, one of the largest European telecommunications and messaging service providers invited NetApp to design a storage solution for a 680,000-user, iSCSI-based, hosted Microsoft Exchange Server messaging environment. The requirements included a "Proof of Concept" (POC) phase, in which the Exchange solution was actually built and tested to verify that it met the performance specifications of the service provider.

The service provider supplied requirements for the Exchange Server architecture, the anticipated user I/O load, and the performance requirements for the storage platform. NetApp engineers analyzed the requirements and designed a solution built around a FAS6030 storage platform. The flexible, modular storage architecture of the FAS6030 storage platform is designed to meet the needs of the largest enterprise applications. NetApp implemented the solution in the NetApp Kilo Client lab in Research Triangle Park, NC. The Kilo Client lab houses over 1,500 IBM BladeCenter server blades and uses NetApp FlexClone® and SAN boot technology to rapidly provision customized test configurations. Refer to http://www.netapp.com/library/tr/3566.pdf for more information on the Kilo Client architecture.

The 680,000-user, iSCSI-based Exchange POC phase culminated with a site visit by the service provider to observe a series of demanding performance tests executed against the FAS6030 over a period of five days. Throughout the testing, the NetApp storage solution successfully met or exceeded all of the performance requirements specified by the service provider. The remainder of this technical report documents the test configuration, test processes, and performance results of the POC phase.

2. Test Configuration

The high-level design requirements of the hosted Exchange Server solution included the following:

- Support for 680,000 Exchange users, split into two groups:
  - 25,000 premium MAPI users (50% concurrency—12,500 active users)
  - 655,000 POP users (10% concurrency—65,500 active users)
- User load consisting of:
  - 0.4 IOPs/mailbox for MAPI users (5,000 sustained MAPI IOPs)
  - 0.2 IOPs/mailbox for POP users (13,100 sustained POP IOPs)
- Average mailbox sizes of:
  - 100MB for MAPI users
  - 8MB for POP users
- iSCSI connectivity between the Exchange servers and NetApp storage systems
- Average database and transaction log latencies of less than 20 milliseconds, per Microsoft best practices. Average latencies above 20ms can negatively affect user response times. As a result, Microsoft recommends that all average database and log latencies be under 20ms for a well-performing Exchange environment.

Building a POC test configuration for a hosted Exchange Server environment requires a complex mixture of Exchange servers, storage systems, domain controllers, global catalog servers, and test clients, in addition to a network infrastructure capable of connecting all of the components. Figure 1 shows the components of the test environment that was used to execute the 680,000-user POC test.

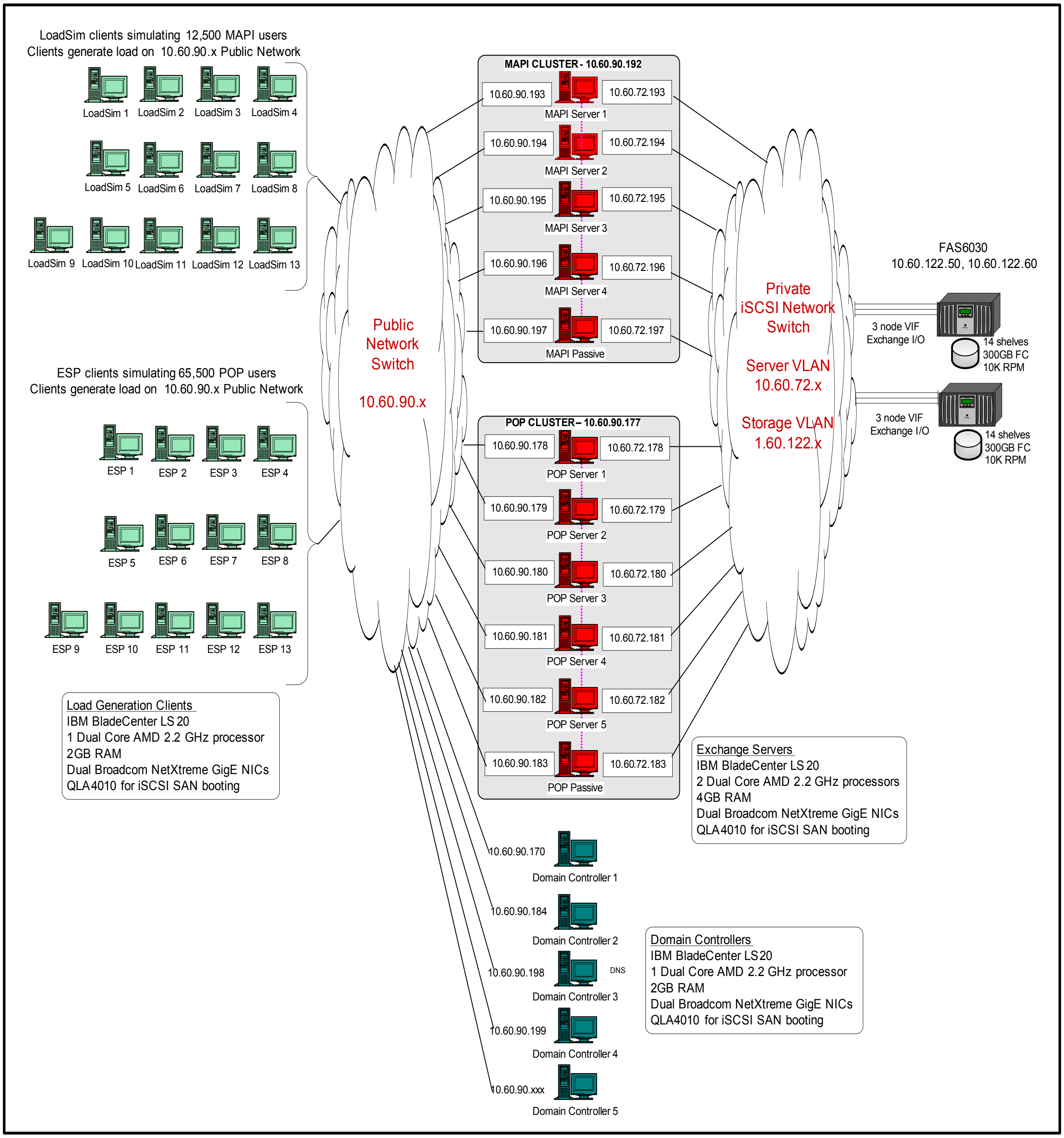

Figure 1) POC test environment.

The test environment contained the following components:

- 11 quad-core servers running Microsoft Exchange Server 2003
- 1 NetApp FAS6030 storage system with 2 storage controllers in an active-active configuration
- 5 dual-core servers configured as domain controllers and global catalog servers
- 26 load-generation clients
- Gigabit Ethernet network infrastructure connecting all test components

The following sections describe the configuration of the individual test bed components in more detail.

2.1 Exchange Server Configuration

The hosted Exchange Server environment supported two classes of e-mail users:

- MAPI users: Mailboxes accessed by client applications such as Outlook® or Outlook Web Access
- POP users: Mailboxes accessed by POP e-mail applications such as Outlook Express

The MAPI user class provided additional premium services beyond those offered to POP users, such as larger mailboxes and integrated calendaring and scheduling. Table 1 describes the basic messaging profile requirements of both user classes. These requirements were primary factors in designing and sizing the NetApp storage components in the hosted solution.

Table 1) E-mail profile requirements.

	MAPI User Requirements	POP User Requirements
Number of mailboxes	25,000	655,000
Average mailbox size	100MB	8MB
User concurrency	50%	10%
IOPs per active mailbox	0.4	0.2
Total IOPs	5,000	13,100

Based on the e-mail profile requirements, half of the MAPI users (12,500 of 25,000) and 10% of the POP users (65,500 of 655,500) actively use e-mail resources at any point in time. The average size of the MAPI user mailboxes was 100MB and the average size of the POP user mailboxes was 8MB. The MAPI users averaged 0.4 IOPs/mailbox, or 5,000 total IOPs. The POP users averaged 0.2 IOPs/mailbox, or 13,100 total IOPs. The sum of the IOPs results in approximately 18,100 IOPs generated by the Exchange users to the storage system at any point in time. The main performance requirement was that the latency for the 18,100 IOPs generated by the Exchange servers to storage must be less than 20ms, per Microsoft best practices.

Each user class placed unique demands on the Exchange infrastructure. To accommodate this, the hosted solution used separate clusters of Exchange servers for handling the MAPI and POP users. The clusters were based on the Microsoft Cluster Server (MSCS) application shipped as part of Windows Server 2003 SP1. Table 2 describes the configuration of the two Exchange clusters. The cluster configurations were specified by the customer as part of the high-level POC design requirements.

Table 2) Exchange server configuration.

	MAPI Exchange Cluster	**POP Exchange Cluster**
Number of Exchange servers	4 active, 1 passive	5 active, 1 passive
Exchange server platform	IBM eServer BladeCenter LS20	IBM eServer BladeCenter LS20
Number of Exchange server processors	2-2.2GHz dual core processors per server	2-2.2GHz dual core processors per server
Amount of RAM	4GB per server	4GB per server
Server OS	Windows Server 2003 SP1	Windows Server 2003 SP1
Exchange Server version	Exchange Server 2003 SP2	Exchange Server 2003 SP2

iSCSI connectivity between the Exchange servers and the storage systems was a major design requirement for the POC test. NetApp storage systems include native support for the iSCSI protocol. Each Exchange server included the following software components for providing iSCSI connectivity to the NetApp FAS6030 storage system:

- Broadcom NetXtreme Gigabit Ethernet controller driver version 8.48
- Microsoft iSCSI Software Initiator version 2.02
- NetApp iSCSI Microsoft Windows Initiator Support Kit 2.2
- NetApp SnapDrive for Windows 4.1
- NetApp SnapManager for Microsoft Exchange 3.2

NetApp SnapDrive for Windows reduces the cost and complexity of managing NetApp storage by enabling flexible and efficient utilization of storage resources to improve data and application availability. SnapDrive software offers a rich set of capabilities that virtualizes and enhances storage management for Microsoft Windows environments. It is integrated with NTFS and provides a layer of abstraction between application data and physical storage associated with that data.

SnapManager for Exchange (SME) is a Microsoft Windows-logo'd Exchange Server software backup and recovery management solution that integrates NetApp Snapshot™ backups with Microsoft Exchange Server via the Microsoft standard and approved Volume Shadow Copy Services (VSS) snapshot interface.

Each Exchange server accessed storage on the FAS6030 via a Gigabit Ethernet (GbE) connection. The available throughput of the GbE connection was more than sufficient to handle the Exchange server I/O load generated during the POC test.

Table 3 describes the number of storage groups and databases configured on the Exchange servers.

Table 3) Exchange storage group and database configuration.

	MAPI Exchange Cluster	POP Exchange Cluster
Number of storage groups per Exchange server	4	4
Number of databases per storage group	5	5
Number of mailbox databases per Exchange server	19	20
Number of public store databases per Exchange server	1	0
Average mailbox database size	40GB	50GB

The following tuning changes were made to each Exchange server:

- Optimized Exchange Server 2003 memory usage per KB815372:
    - Added 3GB USERVA=3030 to the `boot.ini` file.
    - Set `HKLM\System\CurrentControlSet\Control\SessionManager\MemoryManagement\SystemPages` to `0`.
    - Set `HKLM\System\CurrentControlSet\Control\SessionManager\HeapDeCommitFreeBlockThreshold` to `0x40000`.
- Used asdiedit to set `msExchESEParamLogBuffers = 9000` for all storage groups. This increased the number of Exchange log buffers to recommended values.

2.2 NetApp Storage Configuration

The storage system handling the Exchange server I/O was a NetApp FAS6030 running Data ONTAP® 7.2 with two storage controllers in an active-active configuration. The FAS6030 contained 28 shelves of 300GB 10K RPM Fibre Channel drives with 14 shelves connected to each storage controller via four Fibre Channel loops.

Table 4 shows how the four active MAPI Exchange servers and the five active POP Exchange servers split their storage resources between the two storage controllers on the FAS6030. The servers are mapped so that there is an even distribution of the Exchange workload between the two storage controllers. Note that the four storage groups for POP Exchange server 5 are split between the storage controllers.

Table 4) Mapping of Exchange servers to FAS6030 storage controllers.

Storage Controller 1	Storage Controller 2
MAPI Exchange server 1	MAPI Exchange server 3
MAPI Exchange server 2	MAPI Exchange server 4
POP Exchange server 1	POP Exchange server 3
POP Exchange server 2	POP Exchange server 4
POP Exchange server 5 – SG1,SG2	POP Exchange server 5 – SG3,SG4

Data ONTAP 7G provides functionality that enables the creation of logical storage volumes (FlexVol volumes, or flexible volumes) for managing data without the need to assign physical drives to the volumes. Instead, these flexible volumes all derive significant performance benefits from a larger pool of physical drives, called an *aggregate.* This results in several notable benefits for Microsoft Exchange Server environments:

- Using FlexVol volumes, the capacity and performance bandwidth of a large collection of fast drives can be made available to all volumes. Even very small FlexVol volumes have the benefit of a very large number of drives. This results in increased performance for Exchange data hosted in FlexVol volumes, because the I/O can be efficiently distributed across a large number of drives.
- A large number of volumes can be created, all with independent Snapshot copy schedules, mirroring events, etc. This allows Exchange administrators to partition Exchange databases and transaction logs into separate FlexVol volumes and independently schedule operations, such as backups, based on business requirements.

The disks on each storage controller are configured into five RAID-DP aggregates, as shown in Table 5. RAID-DP, a RAID 6 implementation developed by NetApp, protects data from double disk failures with negligible performance overhead. RAID-DP is a standard feature of NetApp storage systems that prevents data loss in the event of a second drive failure without excessive redundancy costs. Three of the aggregates provide storage resources for Exchange databases, and one of the aggregates provides storage resources for Exchange transaction logs and other miscellaneous Exchange files.

Table 5) Aggregate definitions.

Aggregate Name	Number of Drives	Raid Type	Raid Group Size	Description
aggr0	3	RAID-DP	16	Root aggregate
aggr_db1	58	RAID-DP	20	Database LUNs for 1 MAPI server and 1 POP server
aggr_db2	58	RAID-DP	20	Database LUNs for 1 MAPI server and 1 POP server
aggr_db3	32	RAID-DP	16	Half of the database LUNs for 1 POP server
aggr_logs	36	RAID-DP	12	Transaction log LUNs, SMTP/MTA LUNs, quorum drive LUNs

Each Exchange server contained four storage groups. Each storage group required two LUNS on the FAS6030—one for databases and one for transaction logs. An additional LUN is allocated for Exchange SMTP/MTA storage. All nine LUNs were mapped to the servers using volume mount points. For simplicity, each server had one additional LUN that served as the root for all of the volume mount points.

The 10 LUNs mapped to each Exchange server were contained in their own set of four FlexVol flexible volumes. One flexible volume contained all four database LUNs, and a second flexible volume contained all four transaction log LUNs. The SMTP and volume mount point LUNs each had separate flexible volumes. Table 6 summarizes the volume and LUN configuration for the MAPI and POP Exchange servers. The volume and LUN sizes were chosen to allocate sufficient space for Exchange data and associated Snapshot copies.

Table 6) Volume and LUN definitions.

	MAPI Exchange Servers			POP Exchange Servers		
Flexible Volume Name	**Volume Size (GB)**	**Number of LUNs**	**LUN Size (GB)**	**Volume Size (GB)**	**Number of LUNs**	**LUN Size (GB)**
vol_sg	3584	4	390	6554	4	715
vol_logs	615	4	50	923	4	75
vol_smtp	100	1	50	100	1	50
vol_mntpt	11	1	5	11	1	5

Each storage controller on the FAS6030 was configured with a multimode virtual interface (VIF) to handle iSCSI I/O to/from the Exchange servers. VIFs provide configuration and performance flexibility by aggregating multiple physical network interfaces on the storage platform into a single logical interface. Each VIF contained three GbE ports and was configured with the IP load-balancing policy, which distributed iSCSI traffic across the three ports based on the Exchange server IP address in the iSCSI request.

2.3 Domain Controller Configuration

The POC test environment included 5 servers acting as Windows domain controllers and global catalog servers. Microsoft recommends 1 domain controller processor for every 4 Exchange Server processors. There were 9 active quad-core Exchange servers (36 total processors) and 5 dual-processor domain controllers (10 total processors) in the hosted solution. The Exchange servers distributed their Active Directory lookup requests across all of the domain controllers.

The domain controllers had the following configuration:

- IBM eServer BladeCenter LS20
- One 2.2 GHz dual-core processor
- 2GB of RAM
- Dual-port Broadcom NetXtreme integrated Gigabit Ethernet
- Windows Server 2003 Enterprise Edition SP1
- Global catalog server function enabled
- 3GB switch enabled in `boot.ini`
- Page file size = 4096MB
- Windows native mode enabled
- Used asdiedit to set MaxPageSize to 150000

2.4 Load-Generation Clients Configuration

In order to test the hosted solution, the test environment included 26 systems acting as e-mail load generation clients. These systems simulated MAPI and POP users performing e-mail operations on the Exchange servers. The load generation clients had the following configuration:

- IBM eServer BladeCenter LS20
- One 2.2 GHz dual-core processor
- 2GB of RAM
- Dual-port Broadcom NetXtreme integrated Gigabit Ethernet
- Windows Server 2003 Standard Edition SP1

A group of 13 systems generated e-mail requests to the MAPI Exchange server cluster using the LoadSim 2003 application. LoadSim simulates the activity of MAPI e-mail clients, such as the Microsoft Outlook e-mail client. These systems simulated the e-mail activity of 12,500 active users to the MAPI Exchange server cluster over GbE network connections. (The hosted environment was sized for 25,000 MAPI users with 50% active at any point.) Each system simulated roughly 1,000 users using the built-in MMB3 mail profile.

A second group of 13 systems generated e-mail requests to the POP Exchange server cluster using the Microsoft Exchange Server Stress and Performance 2003 (ESP) tool. This tool supports load generation using various Internet protocols, including POP and SMTP. These systems simulated the e-mail activity of 65,500 active users to the POP Exchange server cluster over GbE network connections. (The hosted environment was sized for 655,000 users with 10% active at any time.) Each system simulated approximately 5,000 users using a custom profile that sent e-mail to random recipients via the SMTP protocol and viewed and downloaded the contents of mailboxes via the POP3 protocol.

3. Test Process

The POC test results in this paper are based on a 12-hour test using all 26 load-generation clients simulating the load of 12,500 active MAPI users and 65,500 active POP users simultaneously performing typical e-mail operations on the MAPI and POP Exchange clusters.

The test process consisted of three phases:

1. Mailbox initialization
2. 12-hour test run
3. Results collection and analysis

The mailbox initialization phase contained the following steps:

1. Used LoadSim to populate 25,000 mailboxes on the MAPI Exchange cluster. The LoadSim initialization process used the MMB3 profile, which created mailboxes with an average size of 100MB. After initialization, the average database size on the MAPI Exchange servers was approximately 40GB.
2. Used a custom script to initialize 655,000 mailboxes on the POP Exchange cluster. The initialization script sent e-mail via the SMTP protocol to each mailbox. After initialization, the average mailbox size was roughly 10MB and the average Exchange database size was 50GB.
3. Took a Snapshot copy after initialization was complete to preserve the state of the initialized Exchange databases.

The 12-hour test run contained the following steps:

1. Rebooted the Exchange servers.
2. Started gathering performance statistics on the Exchange servers, the load-generation clients, and the NetApp FAS6030 storage system.
3. Started LoadSim on the 13 clients generating load to the MAPI Exchange servers.
4. After all LoadSim clients started, started ESP on the 13 clients generating load to the POP Exchange servers.
5. Monitored the load-generation clients periodically throughout the test.
6. Stopped all load-generation clients after the test had run for 12 hours.
7. Stopped gathering performance statistics.

The results collection and analysis phase contained the following steps:

1. Combined the performance statistics captured on each of the Exchange servers into a set of MAPI server and POP server metrics.
2. Removed the 1-hour ramp up and ramp down periods from the combined metrics, resulting in a 10-hour contiguous measurement interval.
3. Extracted disk latency and disk IOPs results from the 10-hour measurement interval.
4. Verified that the Exchange server and storage system performance statistics met the requirement that all Exchange database and transaction log I/O requests have latencies less than 20ms.

4. POC Performance Test Results

The main goal of the hosted Exchange Server POC test was to demonstrate that NetApp storage can support large-scale hosted Exchange Server environments using the iSCSI protocol and successfully meet the following performance requirements:

- Exchange database and transaction disk latencies less than 20ms
- 5,000 sustained IOPs generated by the MAPI Exchange cluster (1,250 average disk IOPs for each MAPI Exchange server)
- 13,100 sustained IOPs generated by the POP Exchange cluster (2,600 average disk IOPs for each POP Exchange server)

Table 7 shows the performance results from the 12-hour test run as measured by Performance Monitor on the Exchange servers with all 26 clients generating load. The Exchange database disk results in Table 7 are averages from all drives on all of the Exchange servers hosting Exchange databases. The Exchange log results are averages from all drives on all of the Exchange servers hosting transaction logs. One-hour ramp up and ramp down times are excluded from the averages, resulting in a contiguous 10-hour measurement period. The table groups the results by the MAPI and POP Exchange servers.

Table 7) Performance results summary.

	MAPI Exchange Servers	POP Exchange Servers
Average Exchange database disk read latency	10.89ms	6.43ms
Average Exchange database disk write latency	5.59ms	6.02ms
Average Exchange log disk write latency	2.09ms	3.29ms
Average Exchange database disk transfers/sec	6,019	13,174
Average Exchange log disk transfers/sec	863	720

The Exchange servers generated an average of more than 20,000 total IOPs using the iSCSI protocol to the FAS6030 storage system. The MAPI Exchange servers averaged 6,019 total IOPs to the drives containing Exchange databases, which was 20% more than the specified requirement. The POP Exchange servers averaged 13,174 total IOPs to the drives containing Exchange databases, which was more than the specified requirement. These results met or exceeded the I/O requirements defined for the POC test environment by the customer.

All of the average disk I/O latencies in Table 7 were well under Microsoft's recommended Exchange latency threshold of 20ms. The database disk average read and write latencies were 10.89ms and 5.59ms respectively for the MAPI Exchange servers, and 6.43ms and 6.02ms respectively for the POP Exchange servers. The transaction log disk average write latency was 2.09ms for the MAPI Exchange servers and 3.29ms for the POP Exchange servers. These results easily met the latency requirements defined for the test environment by the customer.

Figures 2, 3, and 4 show the average read and write latency for volumes containing Exchange databases and transaction logs on the MAPI Exchange servers over the 10-hour measurement period. The results show that the latency results were consistent across the 10-hour measurement period and that all values were well within the recommended latency threshold of 20ms.

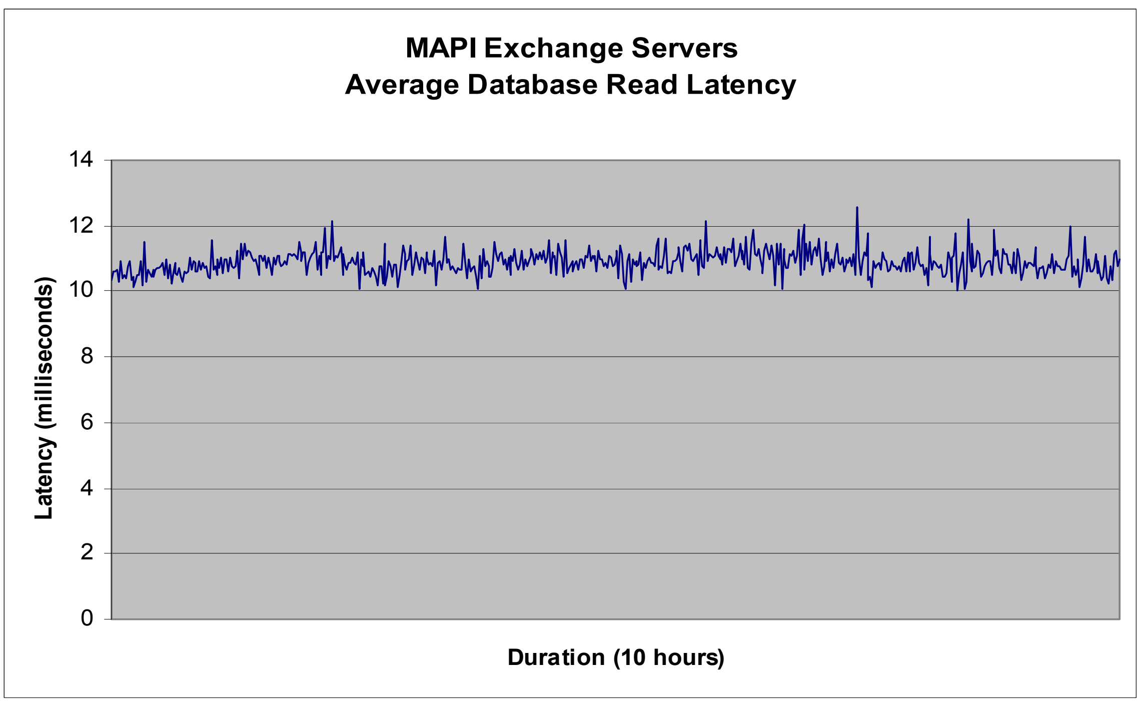

Figure 2) MAPI Exchange Servers average database read latency.

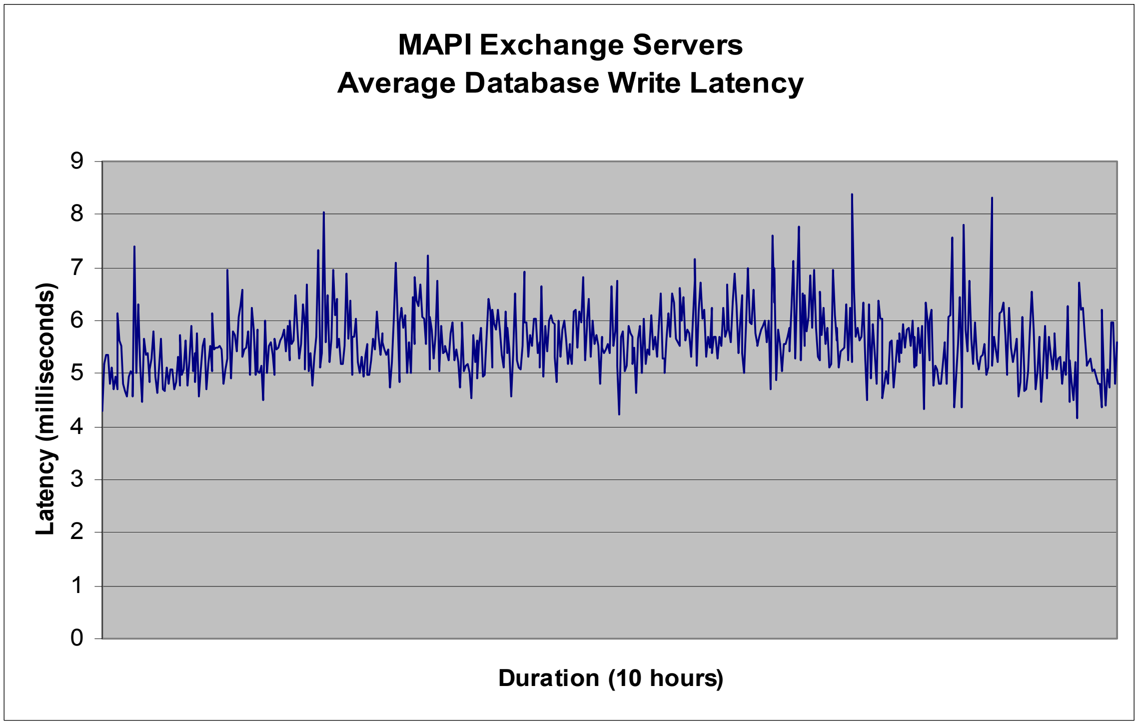

Figure 3) MAPI Exchange Servers average database write latency.

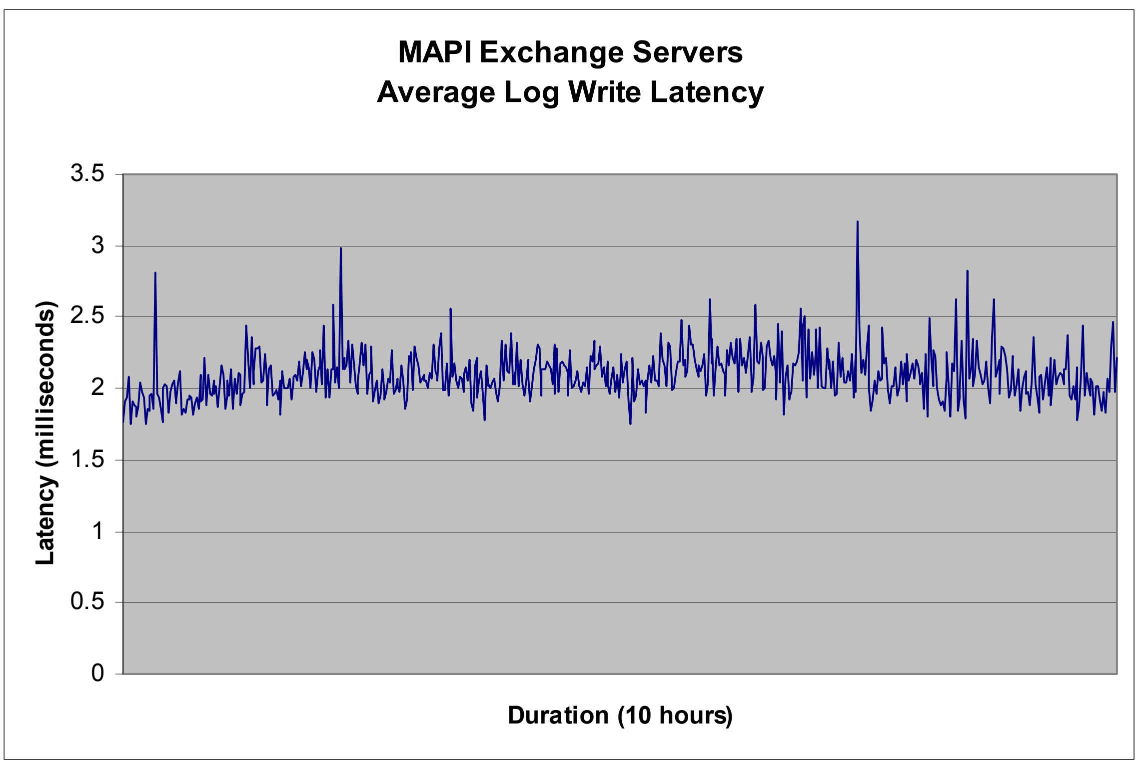

Figure 4) MAPI Exchange Servers average log write latency.

Figure 5 shows the RPC averaged latency measured on the MAPI Exchange servers. RPC averaged latency indicates the response time of the Exchange servers to RPC requests generated by Outlook mail clients. (This metric is not applicable to the POP Exchange servers.) According to Microsoft, average RPC latencies under 50ms represent a responsive Exchange environment. The RPC averaged latency for the POC was approximately 12.5ms over the 10-hour measurement period in the test.

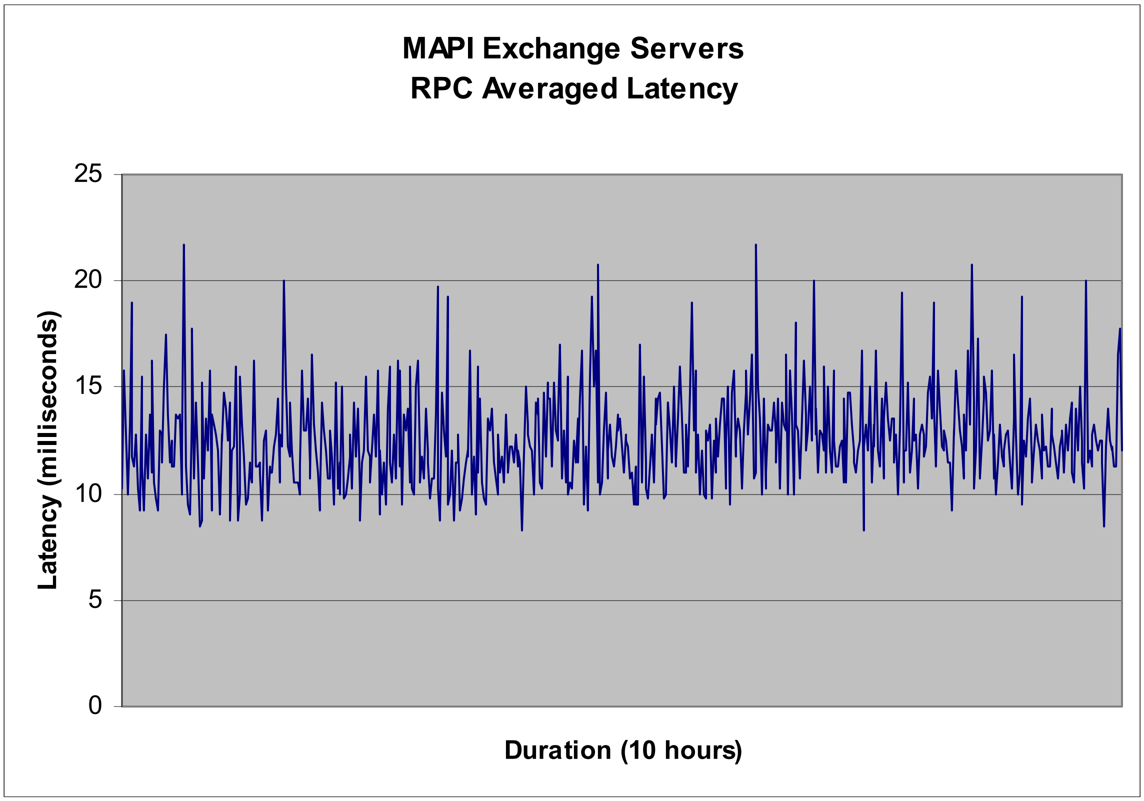

Figure 5) MAPI Exchange servers RPC averaged latency.

Figures 6, 7, and 8 show the average read and write latency to volumes containing Exchange databases and transaction logs on the POP Exchange servers over the 10-hour measurement period. As with the previous data, the latency results were consistent across the 10-hour measurement period and well within the recommended latency threshold of 20ms.

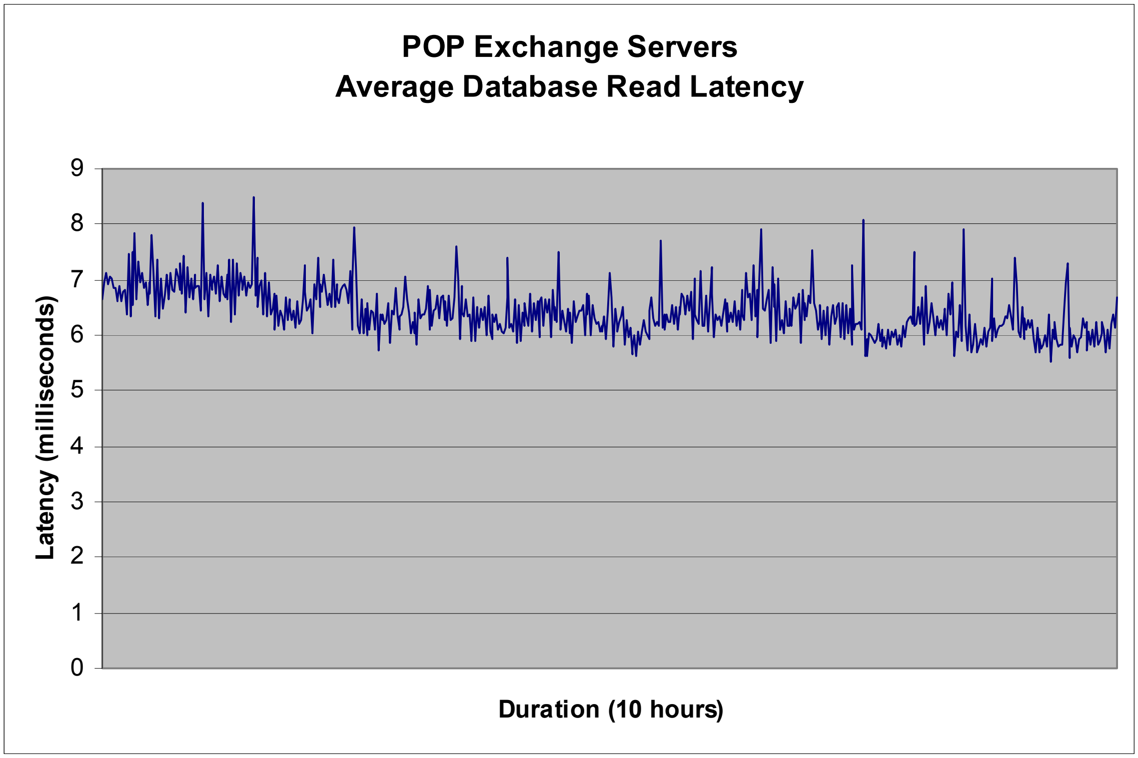

Figure 6) POP Exchange Servers average database read latency.

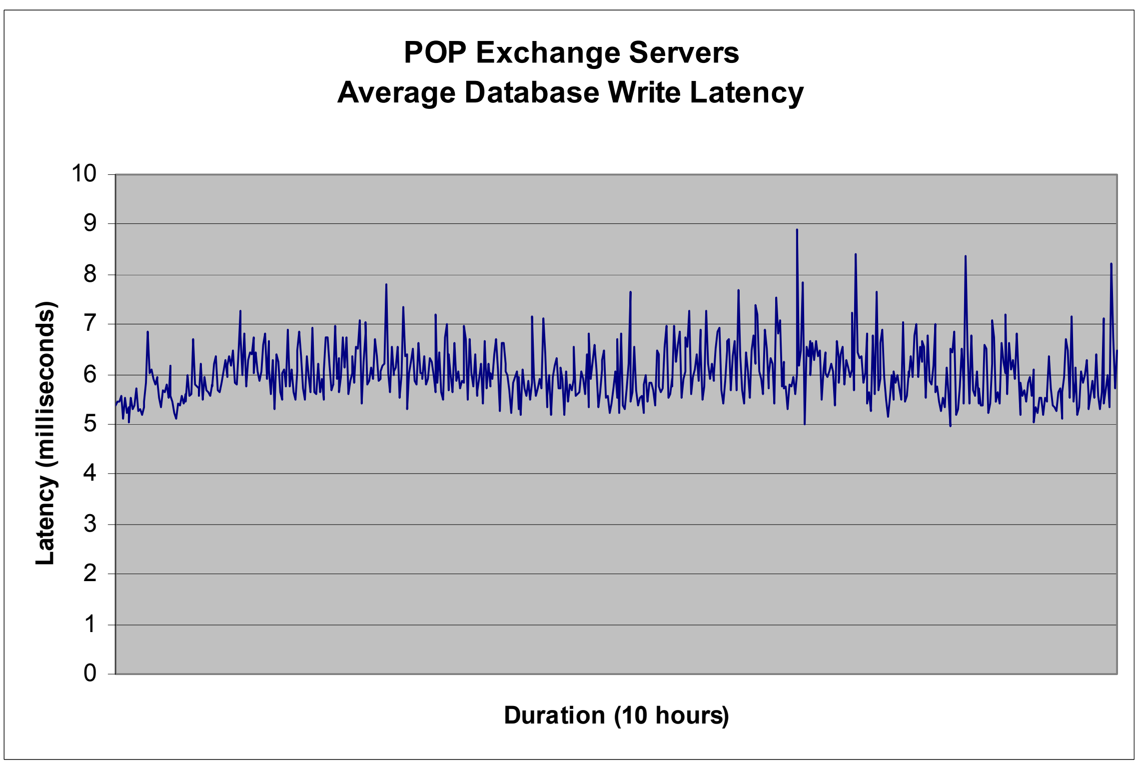

Figure 7) POP Exchange Servers average database write latency.

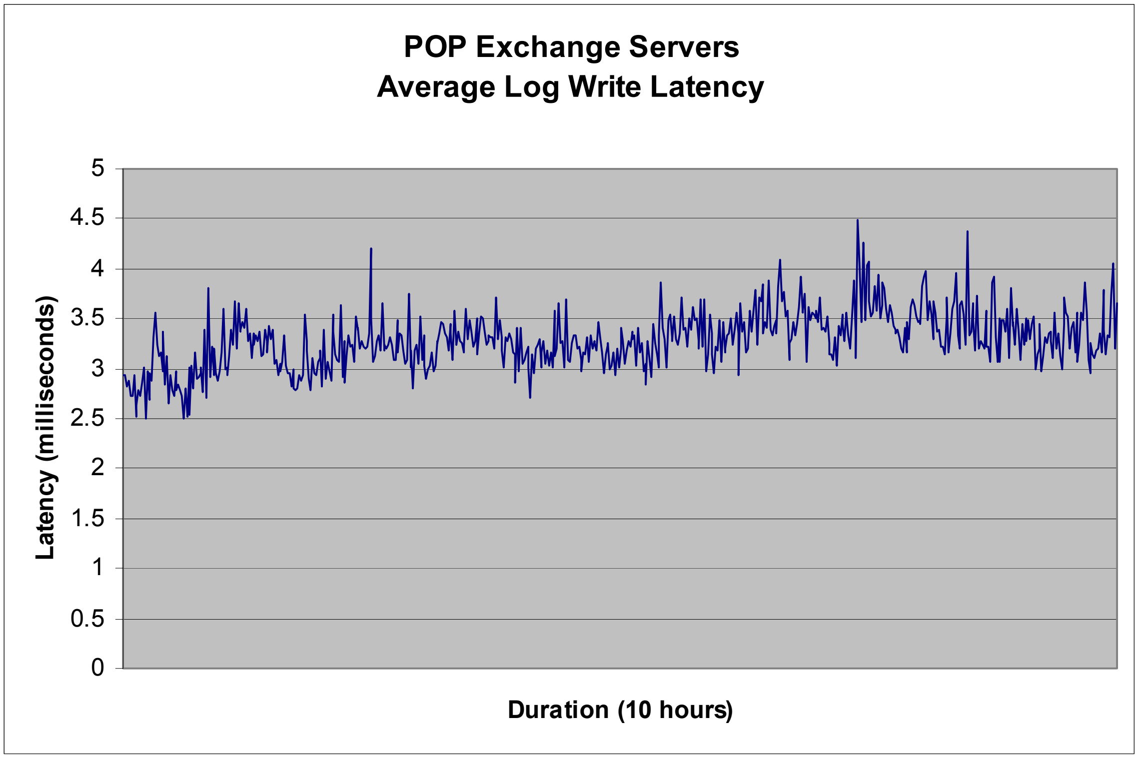

Figure 8) POP Exchange servers average log write latency.

Figures 9 and 10 show the average CPU utilization for the Exchange servers in the MAPI and POP clusters respectively. The servers in the MAPI Exchange cluster had an average CPU utilization between 20% and 25%. The servers in the POP Exchange cluster were busier, with an average CPU utilization between 60% and 70%.

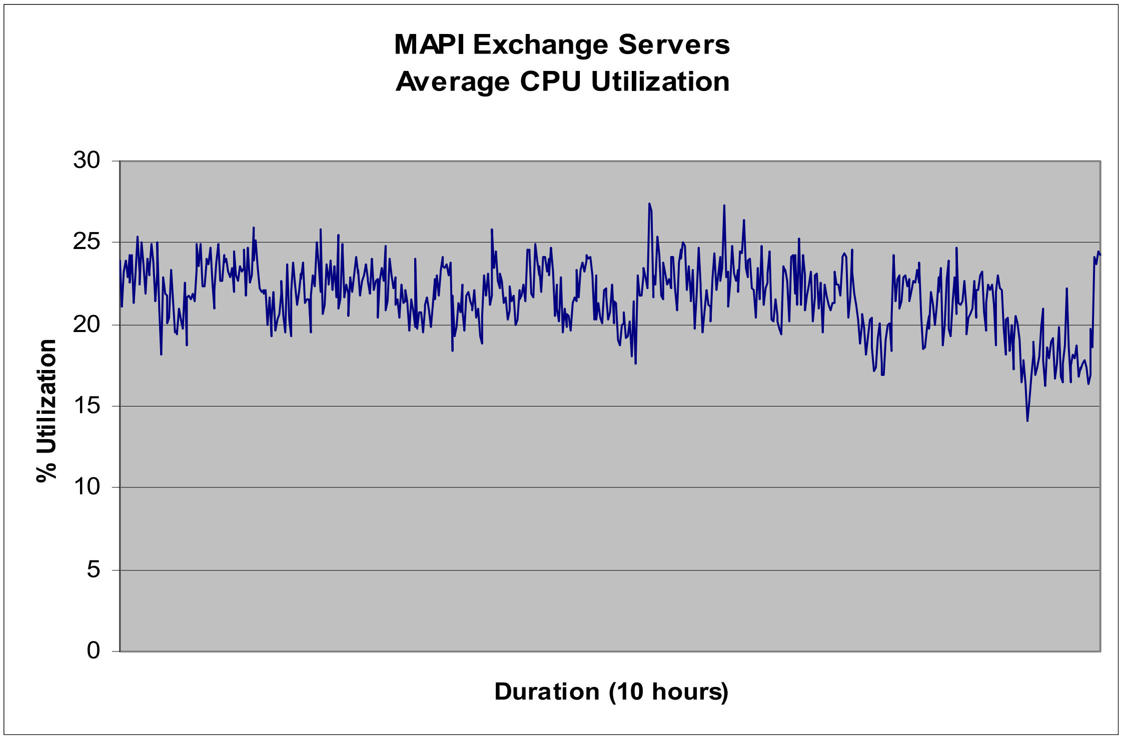

Figure 9) MAPI Exchange Servers average CPU utilization.

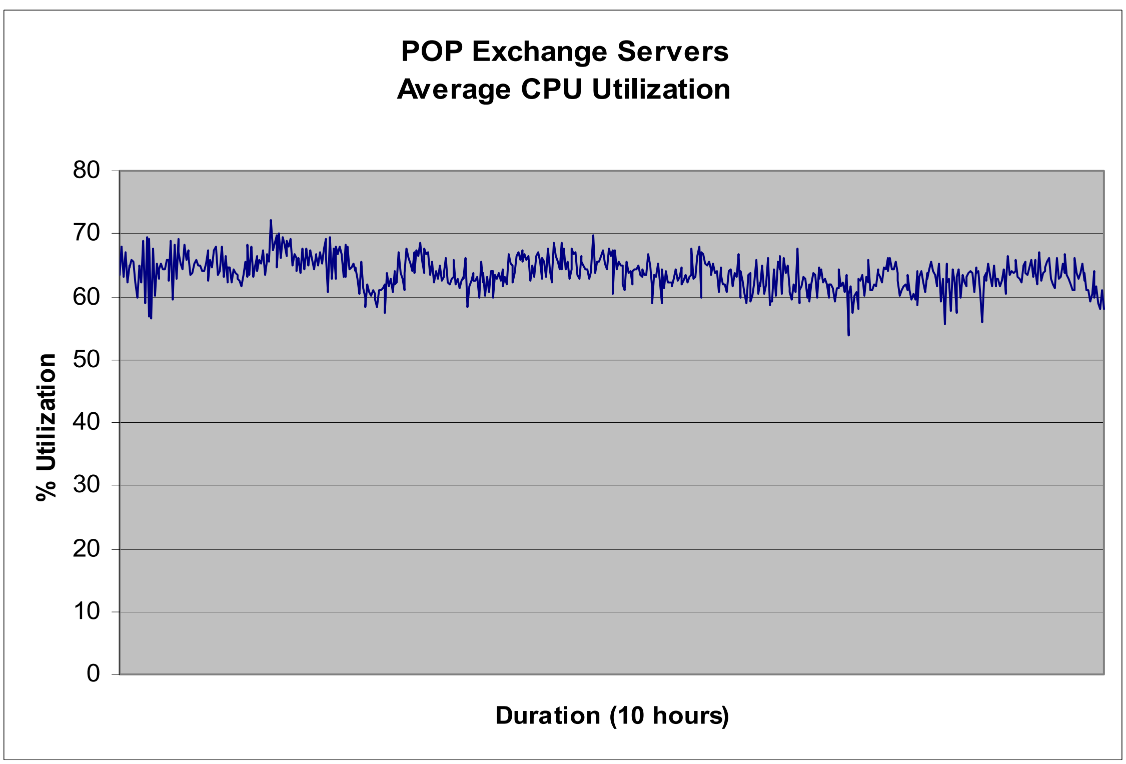

Figure 10) POP Exchange Servers average CPU utilization.

Figure 11 shows the average number of iSCSI IOPs executed by the FAS6030 storage system during the 10-hour measurement interval. The FAS6030 averaged approximately 21,800 iSCSI IOPs, which correlated closely to the number of disk transfers/second measured on the Exchange servers.

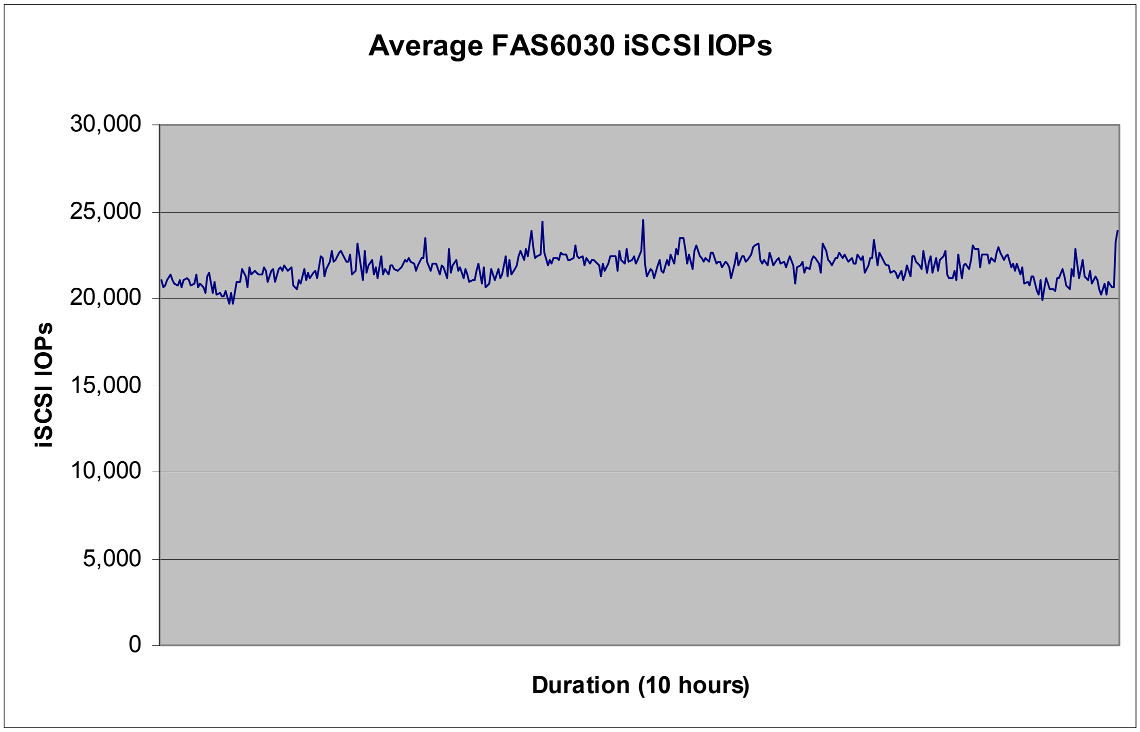

Figure 11) Average FAS6030 iSCSI IOPs.

Figures 12 and 13 show the average iSCSI read and write throughput measured on the FAS6030 storage system. The throughput includes all iSCSI traffic sent and received over the multimode VIFs on the FAS6030. The FAS6030 averaged 310MB/sec of read throughput and 80MB/sec of write throughput.

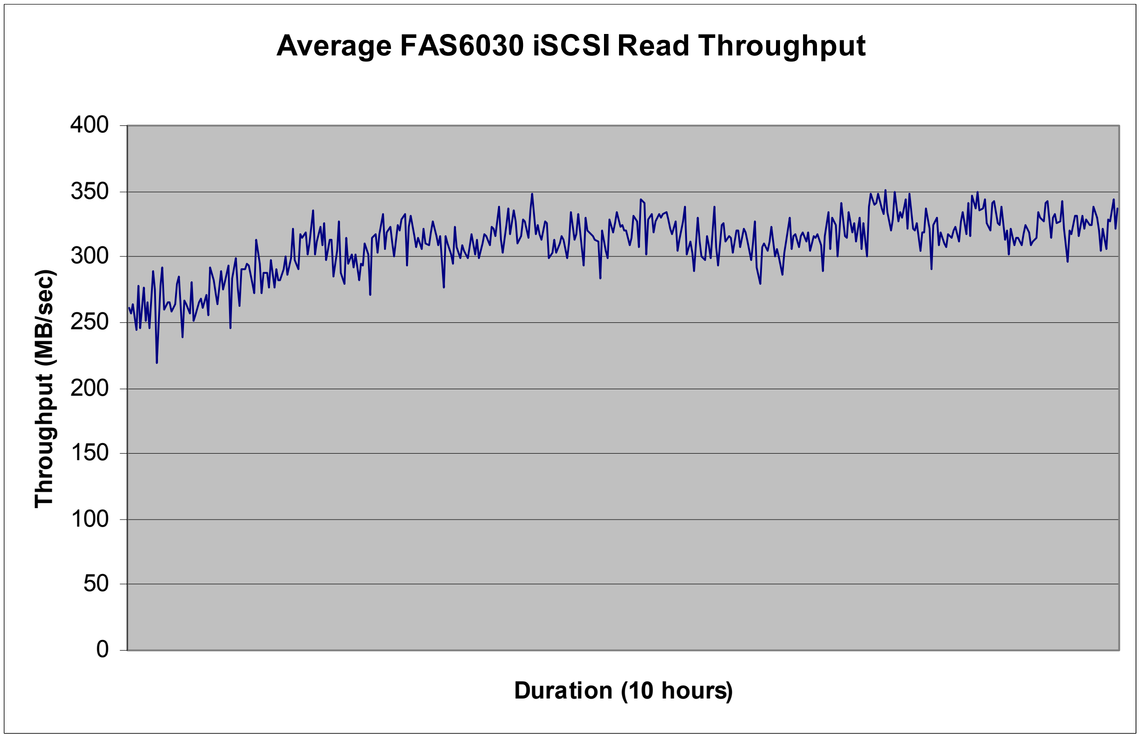

Figure 12) Average FAS6030 iSCSI read throughput.

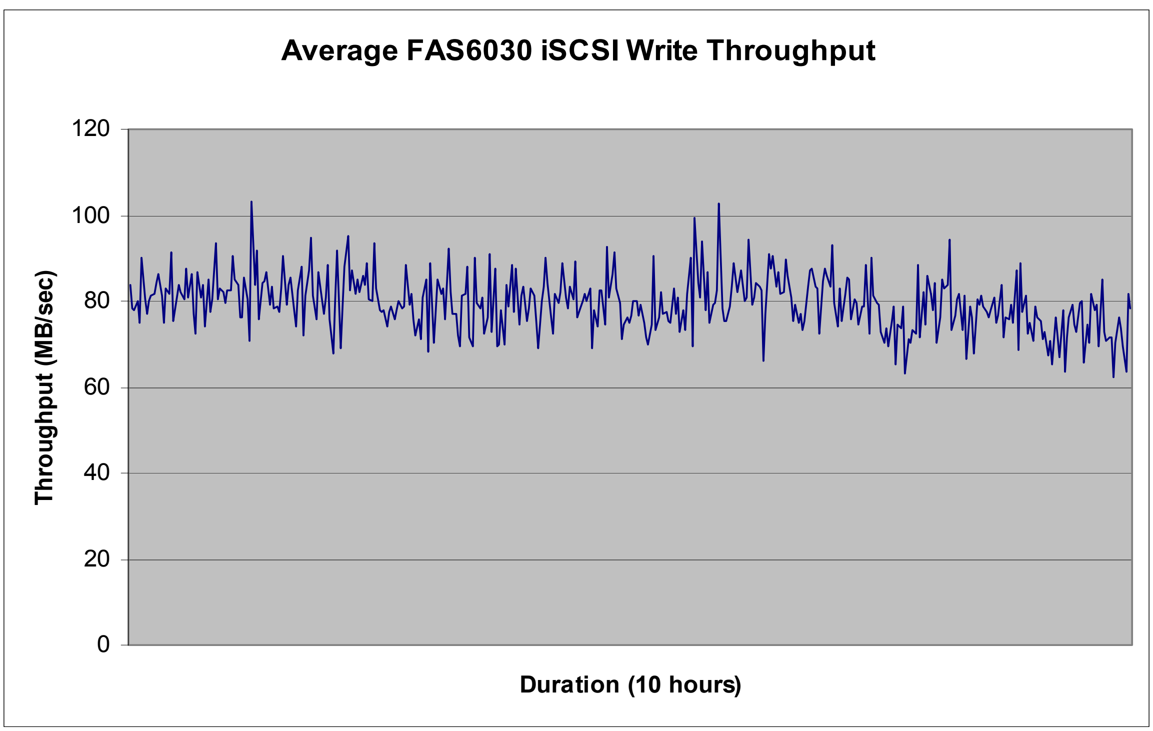

Figure 13) Average FAS6030 iSCSI write throughput.

5. Summary

The performance results of the 680,000-user, hosted Exchange Server proof-of-concept test clearly demonstrate that NetApp storage systems using the iSCSI storage networking protocol easily deliver the performance required for large-scale messaging environments. The storage for the entire hosted messaging environment was handled by a pair of NetApp storage systems, resulting in a small-footprint solution that minimized valuable data-center rack space and power consumption. Furthermore, this POC clearly demonstrates that the iSCSI storage networking protocol, when used with NetApp storage systems, is a cost-effective, viable storage interconnect option for large Exchange Server environments.

In addition to providing excellent performance, the flexibility of the NetApp Exchange Server solution set significantly simplified deployment of the hosted messaging environment. The combination of NetApp aggregates and FlexVol volumes reduced the complexity of storage provisioning by automatically distributing Exchange Server data across the disks in the storage system.

The simple point-and-click interface of NetApp SnapDrive for Windows reduced the time required to configure iSCSI connections between the Exchange servers and the NetApp storage systems.

For more information on the total cost of ownership for Exchange Server when using NetApp storage, see the Mercer Exchange Server TCO Study: http://www.netapp.com/promo/mercer/exchange.

For more information on NetApp solutions for Microsoft Exchange Server, see http://www.netapp.com/solutions/applications/messaging-collaboration/exchange.html.

NetApp

www.netapp.com