VMware vSphere™ 4: Exchange Server® on NFS, iSCSI, and Fibre Channel

vmware®

Table of Contents

Executive Summary

Microsoft Exchange Server 2007 is a storage intensive workload used by a large number of IT organizations to provide email for their users. In this paper, a large installation of 16,000 Exchange users was configured across eight virtual machines (VMs) on a single VMware vSphere™ 4 server. The performance of this configuration was measured when using storage supporting Fibre Channel, iSCSI, and NFS storage protocols. The results show that each protocol achieved great performance with Fibre Channel leading the way, with iSCSI and NFS following closely behind.

Introduction

At VMworld Europe 2008, VMware announced results of a test demonstrating 16,000 Exchange 2007 users on a single ESX Server. The ESX Server was running eight virtual machines each supporting 2,000 users. This multiple virtual machine, building block approach allows for a much higher number of users on a single physical server than had been possible in the past.

In order to test how Fibre Channel, iSCSI, and NFS storage protocols would perform under a high stress load for an extended period of time, VMware conducted a similar Exchange Server test. A NetApp FAS6030 was selected because it supported all three protocols, which would make it possible to compare them using the same storage and servers. A single Fibre Channel or Ethernet storage connection was used in order to include the bandwidth capability as part of the test. This is important because the additional bandwidth capacity of Fibre is often cited as its key advantage over iSCSI or NFS.

Hardware

Server

An HP Proliant DL 580 G5, a four socket rack-mounted server, was configured with four Quad-Core Intel Xeon X7350 2.93 GHz processors for testing. The server had 128GB of RAM, a Qlogic HBA, and an Intel Quad Port Gigabit Ethernet card.

HP ProLiant DL 580 G5	
Processor	4 x Quad-Core Intel Xeon X7350 2.93GHz
Memory	128GB
Disk Controller	Smart Array P400
Fibre Channel HBA	Qlogic QLE2462 4 Gigabit PCI-e
Network Adapters	Intel PRO 1000 Quad Port Gigabit Ethernet

Connectivity between the storage arrays and the host was limited to a single connection of each type. This means that the 4 Gigabit Fibre Channel had higher bandwidth capacity than the single Gigabit Ethernet connection used for both iSCSI and NFS. As this is a key differentiator between the three protocols as commonly deployed, it was also part of the testing. An EMC DS500B switch was used for the Fibre Channel connection and an Extreme Networks Summit 400-48t switch was used for the iSCSI and NFS connection.

NetApp FAS6030

The NetApp FAS6030 supports Fibre Channel, iSCSI, and NFS storage protocols, which made it an excellent choice for comparing the three. The FAS6030 was configured for this test with 114 disks which were split into four aggregates: two 40 disk Data aggregates, one 26 disk Log aggregate, and a 4 disk Root aggregate. All aggregates were created using NetApp's double parity RAID. Two 2.5TB volumes were created in each of the 7.5TB Data aggregates. One of which was used for NFS and the other for FC and iSCSI. The remaining 2.5TB on each aggregate was reserved for snapshot cache. The 5.1TB Log aggregate was divided into two 2.5TB volumes with one for nfs and the other for FC and iSCSI. There was no need to keep snapshots of the logs, so no additional space needed to be reserved.
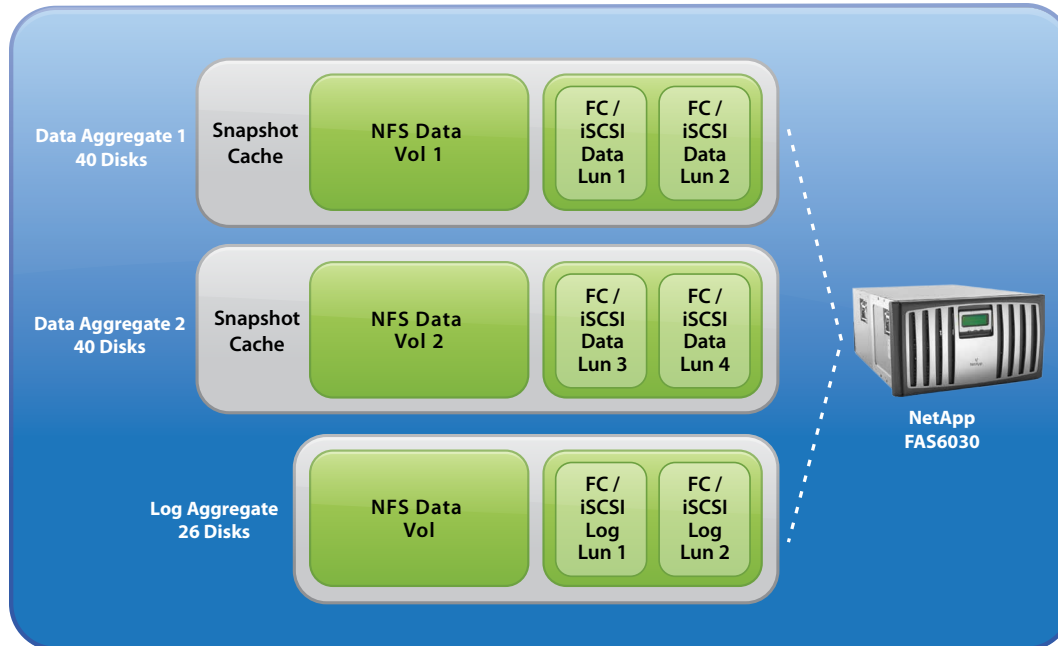
Disks were assigned at the aggregate level, with the volumes and LUNs created within the aggregate spread across evenly. This resulted in good performance within the aggregate as all spindles were actively used.

On a NetApp array, the NFS mount points are configured at the volume level. LUNs (which can be accessed via iSCSI or Fibre Channel) are created within a volume and then assigned to hosts. In order to switch a LUN between iSCSI and Fibre Channel, the group that it is mapped to was simply changed.

Two 1.17TB LUNs were created on each of the Data volumes, which means that four data LUNs were presented to the ESX host. These were mounted as VMFS storage and used to hold the data disks for the Exchange virtual machines. Because NFS is supported directly by ESX, the 2.5TB NFS Data volumes were mounted directly with no VMFS partition used.

Software

ESX 4.0 Release Candidate

A key component of VMware vSphere is ESX Server. ESX Server 4.0 Release Candidate build 140815 was installed on the HP DL580 G5. It was configured to use the NetApp 6030 by way of Fibre Channel, iSCSI, and NFS storage options. This included configuring a virtual network switch to connect the software based iSCSI initiator of ESX to the FAS 6030 by a dedicated Gigabit Ethernet network. This same virtual switch was used for the connection to access NFS mounts on the NetApp array. The Fibre Channel based data connections do not go through the virtual switch, but instead are managed from the Storage Adapters section where the connection between the Fibre Channel HBA and the storage is configured.

Domain Controller Virtual Machine

In order to support the Exchange 2007 testing environment, a new domain controller virtual machine was created. It was configured with two vCPUs and 8GB of RAM. Windows Server 2003 R2 Enterprise Edition was installed with DNS services, and then promoted to domain controller of a new domain with the dcpromo command. In addition to a domain controller, Exchange Server 2007 has three server roles that were required. The Hub Transport and Client Access roles were installed on this domain controller virtual machine. These roles are not heavily stressed in the testing, but must be present for Exchange 2007. The Domain Controller virtual machine was kept on the same HP DL580 G5 server as the Exchange virtual machines for the testing.

Exchange Virtual Machines

In order to support 16,000 users, eight Exchange virtual machines each designed to support 2,000 users were created. Each virtual machine was configured according to Microsoft's best practices with 14GB of RAM and 2 vCPUs. A 20GB operating system disk for each virtual machine was created on the local disks of the DL580 G5.

The virtual machines were installed with Windows Server 2003 R2 64-bit Enterprise Edition, VMware Tools, PowerShell 1.0, .Net Framework 2.0 with SP 1, and other fixes required for Exchange 2007 SP1. The virtual machines were then added to a windows domain created for the purpose of this testing. Exchange 2007 SP1 was installed with just the Mailbox Role and management tools options selected.

Each virtual machine was assigned two 150GB virtual disks on two of the data LUNs and one 300GB virtual disk from the log LUN. Each virtual machine had a total of six virtual disks assigned: 1 for OS, 4 for mailbox databases, and one for logs. The four mailbox database virtual disks were created so that half of the virtual machines used Data Aggregate 1 and the other half used Data Aggregate 2. All virtual disks were attached to the virtual machines using the default LSI Logic Parallel virtual SCSI controller. The partitions were aligned following best practices. The VMFS partitions are aligned automatically because they are created using the VI Client. The partitions inside the windows guests were aligned using the diskpart command before formatting.

LoadGen Test System

Microsoft provides Exchange Load Generator, commonly referred to as LoadGen, as a tool to simulate Exchange users. This tool was new with Exchange 2007 and replaces the previous tool used with earlier versions of Exchange, which was called LoadSim. A Dell PowerEdge 1950 installed with Window Server 2003 64-bit Enterprise Edition was used as the LoadGen Server.

Testing

Anatomy of a LoadGen Test

There are two phases to a LoadGen test. The first is the user creation and initialization phase where LoadGen creates the specified number of users and pre-populates their inboxes with a number of items. The provided Heavy User profile is used which means that each user will be created with approximately 250MB of data. The initialization phase of the 16,000 users proceeds at a rate of one user being initialized every .38 minutes, or a total of about 4 days.

To preserve this initialized state for the data, a snapshot of each of the data aggregates was taken on the NetApp array at this point. The snapshot is then used to reset mailbox databases back to the initialized state before each new test. The snapshot restore takes only a few seconds to complete, which is a big time savings over the hours it would take to reinitialize.

The second phase of the LoadGen test is to run a simulation. For these tests, each simulation was run for eight hours and 30 minutes to model a typical workday, with a little bit of added time to account for getting the test started. During this simulation, LoadGen used a range of operations including sending mail, browsing the calendar, making appointments, and deleting items to generate its load against the Exchange servers. At the end of the test, LoadGen reports resulted in terms of response time for each of these operations.

Fibre Channel, iSCSI, and NFS Tests

The initialization of users was done while the virtual machines were configured to access the NetApp array over Fibre Channel. As mentioned earlier, this initialized state was captured with a snapshot of the two data aggregates on the NetApp array.

The NetApp array used connection groups to define how LUNs are accessed. A Fibre Channel and an iSCSI group are created with each having either the Fibre Channel WWN or iSCSI initiator string of the ESX server. To switch from Fibre Channel to iSCSI, the LUNs were simply moved from one group to the other. Switching to NFS requires the VMDK files for the data and log LUNs to be copied to the volume for the NFS mount. To do this, the NFS volumes were added to the ESX server and then a simple copy is performed from within the service console in order to make a copy of the VMDK files on the NFS volumes.

Procedure

The basic procedure for each test was the same regardless of the storage protocol used. The ESX Server was restarted. Next, the Domain Controller virtual machine is booted and given approximately 10 minutes to complete starting. All of the Exchange virtual machines were started at the same time and then given approximately 25 minutes. During this time the LoadGen server was rebooted. The LoadGen tool was then launched on its server and a test is started for all 16,000 users with the initialization phase skipped. The test is monitored with esxtop from the ESX Service console, LoadGen on the LoadGen server, perfmon within the virtual machines, and sysstat on the Netapp array.

Results

Using Microsoft Exchange Load Generator, or LoadGen, with the Outlook 2007 Online Heavy User profile has been the standard way to test. This user profile equates to each user doing 47 tasks per day. In order to test an even higher level of work that might more closely resemble email power users, a double heavy profile was created by editing the Heavy profile and doubling all of the user tasks, resulting in 94 tasks per user per day.

The LoadGen results show response time or latency for a set of tasks that Exchange users performed. These results show that all three storage protocols were able to provide sufficient performance to support all 16,000 Exchange users. Fibre Channel provided the best performance, but iSCSI and NFS were close behind.

Of all of LoadGen's user tasks, Send Mail is most commonly used to indicate overall performance. For Send Mail latencies, a result of greater than 1,000 milliseconds signifies a degraded user experience. The average latency and 95th percentile latencies are shown in the chart below. All results are the average result over the entire eight-hour run.

Figure 1. Average Send Mail Latency
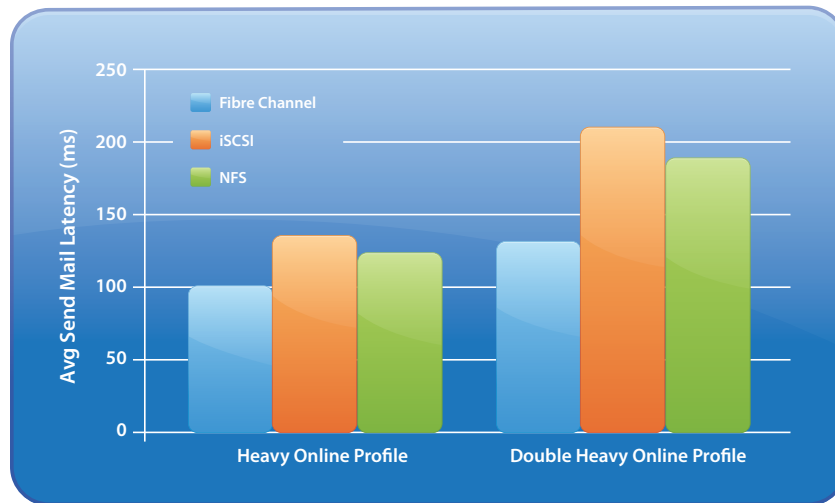
With the Heavy Online Profile, all three protocols were within 40 milliseconds of each other in average Send Mail latency with Fibre Channel reporting the best time. The increased load of the Double Heavy Online Profile better distinguishes performance of the three protocols with Fibre Channel still showing the best performance.

While average latency is a good metric, 95th percentile latency is a better metric for understanding the worst-case performance for an individual user. The 95th percentile metric is often used to capture the Quality of Service and to establish Service Level Agreement metrics. The numbers for 95th percentile Send Mail latency are higher overall but show similar relative results to the average latency. All three protocols are still fairly close in the Heavy Profile, with a little bit more separation on the Double Heavy profile. Fibre Channel once again provided the best performance.

*Figure 2. 95th Percentile Send Mail Latency*
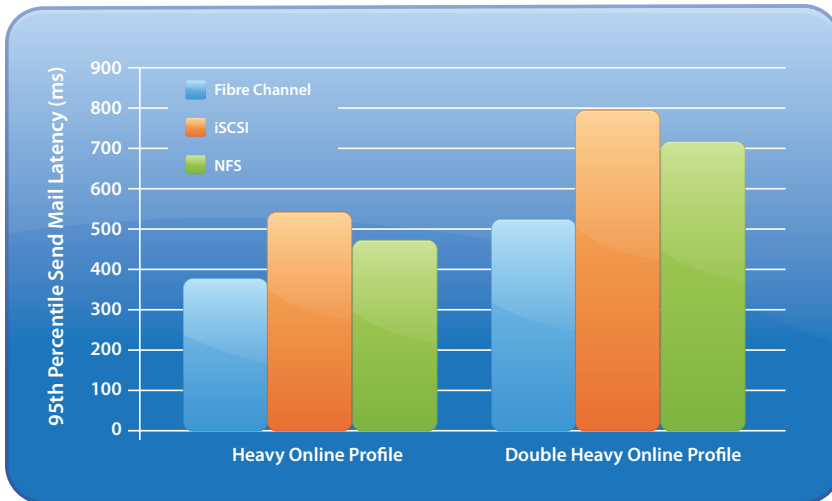
The average I/O operations per second (IOPS) is about 3,700 for the Heavy profile and 5,300 for the Double Heavy profile. This shows how much the increase in tasks affects the test workload. It takes about half of an eight-hour LoadGen test before the test will reach a steady state. The graph below shows the IOPS for all three protocols during the heavy user profile. IOPS initially started off near 6,500, but ended around 3,300 for all three protocols.

*Figure 3. IOPS during eight-hour Heavy Profile LoadGen Test*
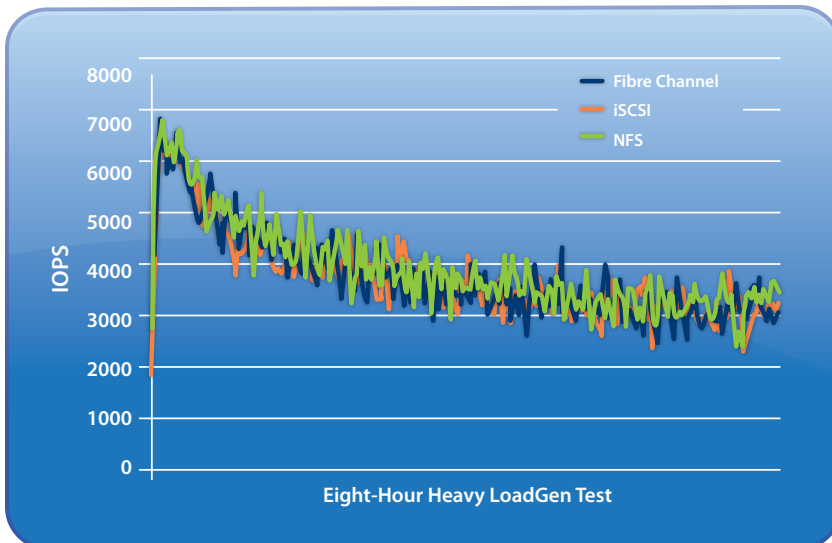
The same graph of the Double Heavy workload explains the separation in the latencies under higher stress. At the beginning of the test, Fibre Channel was able to reach a higher max number of IOPS than iSCSI and NFS. NFS and iSCSI then remained at a higher level of IOPS than Fibre Channel to clear the backlog. After this first part of the test, all three protocols were performing approximately the same number of IOPS much as in the graph from the Heavy profile.

*Figure 4. IOPS during eight-hour Double Heavy LoadGen Test*
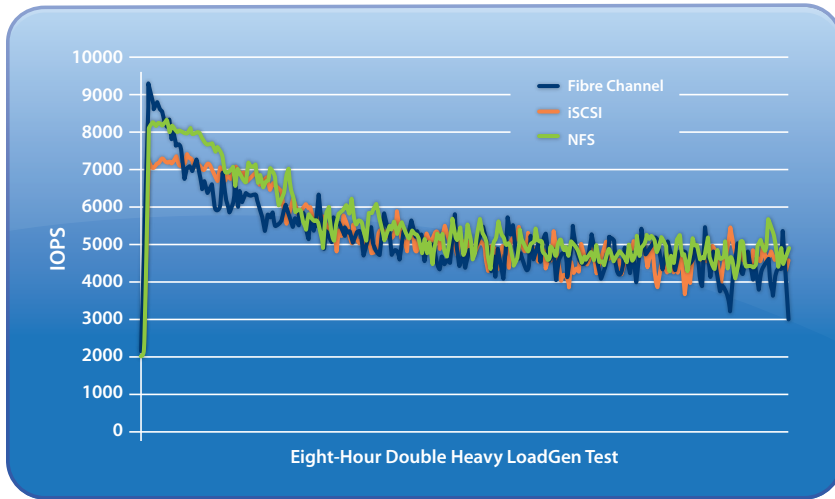
Due to the bursty nature of Exchange I/O, iSCSI and NFS briefly exceeded the maximum capacity of a single Gigabit Ethernet connection, which resulted in higher latency. In order to show that this was the bottleneck, an additional test was run on iSCSI with four network ports, each with a Gigabit Ethernet connection. This configuration more closely compares with the bandwidth available to the four Gigabit Fibre Channel connection. The results show that when bandwidth is not constrained, latency remained lower and the iSCSI connection behaved more like the Fibre Channel. The figure below shows the average disk latency results.

Figure 5. Disk Latency during eight-hour Double Heavy LoadGen Test with Four Connections for iSCSI
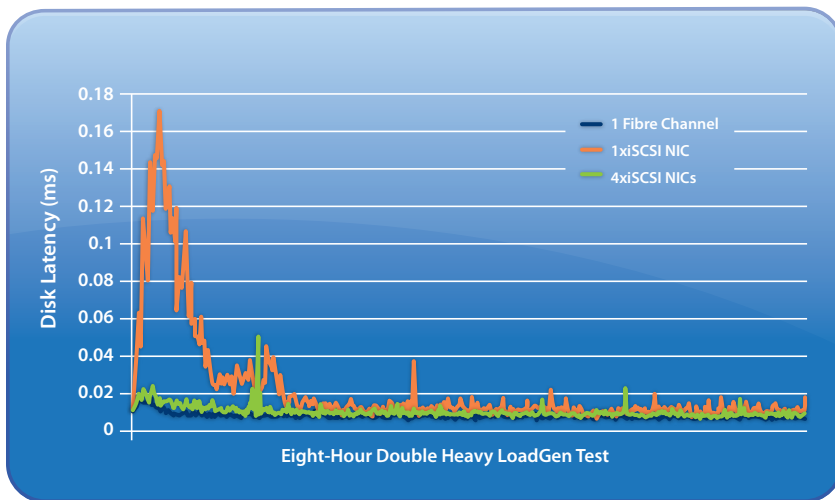
An additional measure of performance for these tests is CPU utilization, which shows the impact that each of the storage protocols had on overall system utilization. Fibre Channel had a lower impact because there is less processing that must be done by the host system, with the majority of the work being done on the HBA. iSCSI and NFS required additional work at the host which resulted in higher ESX CPU utilization.

Figure 6. ESX Host CPU Utilization
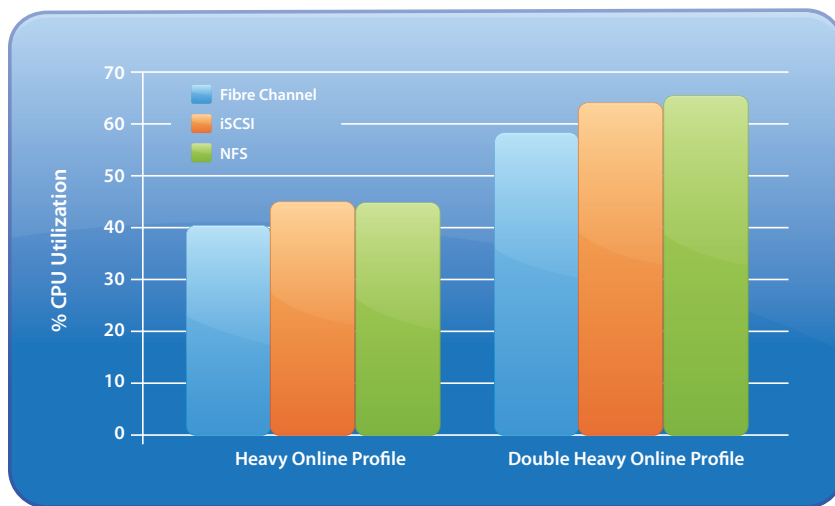
Performance Comparison of ESX 3.5 and ESX 4.0

There are a lot of configuration differences between the test described in this paper and the 16,000-user Exchange test published a little over a year ago. These differences make direct comparison difficult. The storage arrays are different with a different number of total disks that were configured with different types of RAID. Additionally, the Exchange 2007 architecture is different. In the earlier test, each virtual machine was loaded with hub transport, client access, and mailbox roles and had a dedicated domain controller virtual machine on another ESX host. In the tests described in this paper, each virtual machine only had the mailbox role, but shared a common domain controller that also has the hub transport and client access roles. This domain controller/hub transport/client access virtual machine was also running on the same server as the Exchange virtual machines. These are significant differences.

There are also quite a few similarities. Both tests were with four-socket quad-core Intel servers with 128GB of RAM. The virtual machines were configured the same with two vCPUs and 14 Gigabit of RAM running Windows Server 2003 x64 and Exchange 2007 with SP1. LoadGen was used with the same test profile of Heavy Online Outlook 2007 for 16,000 users and an eight-hour test.

The remaining major difference between the two tests is the version of ESX Server, allowing a loose comparison of ESX 3.5 and ESX 4.0. The performance of the 16,000-user configuration on ESX 4.0 was better than the ESX 3.5 configuration. On the Heavy workload with Fibre Channel storage, the average latency was 135 milliseconds for Send Mail with ESX Server 3.5 in the previous test and just 101 milliseconds for ESX Server 4.0 in this test. The 95th percentile latency comparison shows a similar advantage for ESX 4.0: 440 milliseconds versus 374 milliseconds for ESX 3.5 and ESX 4.0, respectively.

Conclusion

All three protocols tested have shown to be able to support 16,000 Exchange users in both the Heavy and Double Heavy workloads. These workloads stressed a 16-core HP DL 580 server to approximately 60 percent CPU utilization and pushed an average of over 5,000 IOPS to the attached NetApp storage array. This throughput was maintained while the average latency stayed under 220 milliseconds and 95th percentile latency remained under one second. Fibre Channel provided the best performance with the lowest CPU utilization. NFS and iSCSI performance was comparable with slightly longer latencies, and slightly higher CPU utilization.

vmware®