

Microsoft Exchange Server 2003 Performance on VMware ESX Server 3

BY KONG L. YANG
ARAVIND PAVULURI

Before virtualizing critical systems such as Microsoft® Exchange servers, enterprises should understand how those virtual machines might perform. This article describes tests evaluating the performance and scalability of Microsoft Exchange Server 2003 on VMware® ESX Server 3, and compares this virtual system's performance with that of a dedicated physical server.



Virtualization has become a mainstream technology, allowing enterprises to consolidate underutilized servers while helping increase reliability and fault tolerance and simplify load balancing. As organizations embrace virtualization in the data center, many may consider virtualizing Microsoft Exchange software on a platform such as VMware ESX Server,¹ which raises several important questions: How many users can a virtual machine (VM) support compared with a dedicated physical server? How well can a virtualized environment scale when adding virtual processors or VMs?

To help provide guidance on these questions, in the winter of 2006, Dell and VMware engineers conducted tests to compare Exchange Server 2003 performance on physical and virtual systems and evaluate VM scalability. The results indicated that in the test environment, one VM with a single virtual processor could support 1,300 heavy users, and consolidating multiple instances of these single-processor VMs could cumulatively support up to 4,000 heavy users while providing acceptable performance and scaling.

Test environment and methodology

The test environment was based on a Dell™ PowerEdge™ 6850 server, a four-socket rack server that supports Intel® Xeon® 7000 and 7100 series processors and up to 64 GB of RAM, and provides four PCI Express and three PCI Extended I/O slots.

The server was configured with 16 GB of RAM and ran the 32-bit Microsoft Windows Server® 2003 Enterprise Edition OS and Microsoft Exchange Server 2003 with Service Pack 2 (SP2). In the physical-environment tests, the amount of memory used was limited with the boot.ini /MAXMEM option. In the virtual-environment tests, each VM ran the same configuration hosted on a VMware ESX Server 3.0.1 platform. Figure 1 summarizes this server configuration.

The PowerEdge 6850 connected to a Dell/EMC CX500 storage array through a Brocade SilkWorm 3900 switch and to multiple PowerEdge 1855 blade servers through a Dell PowerConnect™ 5224 switch (see Figures 2 and 3). The blade servers in the PowerEdge 1855 system served as benchmark clients, the Microsoft Active Directory® directory service server, the domain controller, and the Domain Name System (DNS) server. Although the server and storage back end used in the test environment were not the latest available models, the tests focused on the Microsoft Exchange and VMware ESX Server software, and the results should extend to later-generation servers and storage.

Test and measurement tools

The test team used Microsoft Exchange Server 2003 Load Simulator (LoadSim) running on the PowerEdge 1855 client servers to simulate a messaging workload for Messaging Application

Related Categories:

Microsoft Exchange

Virtualization

VMware

Visit www.dell.com/powersolutions for the complete category index.

¹When considering the deployment of Microsoft Exchange Server 2003 on a virtualization platform such as VMware ESX Server, please refer to support.microsoft.com/kb/897615 for information about the Microsoft support policy for their applications running in non-Microsoft virtualization software.

Server	Dell PowerEdge 6850
Processors	Four dual-core Intel Xeon 7020 processors at 2.66 GHz (eight total cores)
Memory	16 GB (with 4 GB allotted to the server in the physical-environment tests and to each VM in the virtual-environment tests)
Hard drives	Two 146 GB, 10,000 rpm drives in a RAID-1 configuration
Host bus adapters (HBAs)	Two Emulex Fibre Channel HBAs
OS (physical or virtual)	Microsoft Windows Server 2003 Enterprise Edition
Application (physical or virtual)	Microsoft Exchange Server 2003 with SP2
Virtualization software	VMware ESX Server 3.0.1

Figure 1. Server hardware and software configuration used in the test environment

Storage enclosure	Dell/EMC CX500 with one disk processor enclosure and three disk array enclosures
Hard drives	Fifty-four 73 GB, 15,000 rpm drives
RAID configuration	<ul style="list-style-type: none"> • Eight 6-drive RAID-10 volumes for the Exchange Information Store • One 6-drive RAID-10 volume partitioned into eight logical units (LUNs) for the transaction log

Figure 2. Storage configuration used in the test environment

Server	Dell PowerEdge 1855 with four blade servers
Processors	Two Intel Xeon processors at 3.6 GHz per blade server
Memory	2 GB per blade server
OS	Microsoft Windows Server 2003 Enterprise Edition
Benchmark	Microsoft Exchange Server 2003 LoadSim

Figure 3. Client server hardware and software configuration used in the test environment

Programming Interface (MAPI) clients using the heavy workload profile. LoadSim attempts to closely model the normal daily e-mail usage profile of real users and estimate the number of users a system can support. It provides performance metrics indicating how an Exchange server would typically respond when supporting a given number of users. LoadSim tests do not necessarily match real-world deployments, but they can provide consistent, repeatable data for comparison.

The test team used Microsoft Windows® Performance Monitor (perfmon) to monitor performance in the physical environment and esxtop to do so in the virtual environment. The team configured perfmon, a performance monitoring tool for Windows operating systems, to log relevant processor, memory, disk, network, and storage

counters as well as Exchange-specific counters. These counters were logged to a comma-separated value (CSV) performance log, which was later imported to a spreadsheet for analysis. The team

configured esxtop, a performance monitoring tool for ESX Server 3.0.1, to log processor, memory, disk, network, and storage counters for both ESX Server and the VMs.

Test methodology

The objective of the tests was to evaluate how physical Exchange systems compare with virtual Exchange systems, and to determine how well VMs scale out when using multiple VMs and scale up when using multiple virtual processors. In both the physical and virtual environments, the PowerEdge 6850 ran with all four dual-core processors and 16 GB of memory, and the test team used the boot.ini /PAE and /3GB options to extend the physical address and allow Exchange to use up to 3 GB of memory. In the physical environment, the test team used the boot.ini /NUMPROC option to limit the number of processors, and the /MAXMEM option to limit the amount of memory available to the OS. Figure 4 summarizes the different configurations, comparisons, and objectives in the test environment.

The primary performance metric was the number of LoadSim heavy users that each configuration could support while meeting two basic quality-of-service requirements: maintaining an average host server processor utilization of 80 percent, and keeping 95 percent of the response times under 500 milliseconds (to help ensure that the simulated end users experienced a reasonable quality of service). Each test was performed twice to help verify the results' consistency and repeatability.

Configuration 1	Configuration 2	Objective
Physical system with one processor	VM with one virtual processor	Comparing physical and virtual system performance
Physical system with two processors	VM with two virtual processors	Comparing physical and virtual system performance
VM with one virtual processor	Two and four VMs with one virtual processor each	Evaluating how virtual systems scale out
VM with two virtual processors	Two VMs with two virtual processors each	Evaluating how virtual systems scale out
VM with one virtual processor	VM with two virtual processors	Evaluating how virtual systems scale up

Figure 4. Different configurations compared in the test environment

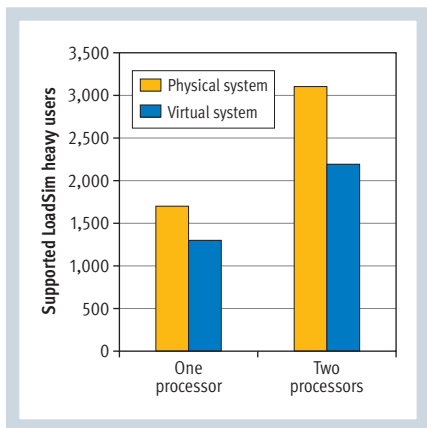


Figure 5. Number of supported users on physical and virtual systems

Test results: Evaluating physical and virtual system performance and VM scalability

The test team used the tests described in the preceding section to compare physical and virtual system performance and evaluate VM scalability. To assess the systems’ responsiveness to end users, they also measured the response times for different configurations.

Comparing physical and virtual systems

To compare physical and virtual systems, the test team used LoadSim with the heavy workload profile on single- and dual-processor configurations of each type of system. The single-processor physical and virtual systems supported 1,700 and 1,300 LoadSim heavy users, respectively. The multiprocessor physical and virtual systems supported 3,100 and 2,200 users, respectively (see Figure 5).

Evaluating VM scalability

In addition to comparing physical and virtual systems, the test team also evaluated the ability of virtual systems to scale out (by adding additional VMs to a host) and scale up (by adding additional virtual processors to a single VM). Figure 6 shows the number of supported users for different numbers of VMs and virtual processors. With one virtual processor each, one VM supported 1,300 users, two VMs supported 2,400 users, and four VMs supported 4,000 users; the additional VMs provided increases of

85 percent and 208 percent over a single VM, respectively. With two virtual processors each, one VM supported 2,200 users and two VMs supported 4,000 users; the additional VM provided an increase of 82 percent. Finally, comparing the results for single- and dual-processor VMs shows that one single-processor VM supported 1,300 users and one dual-processor VM supported 2,200 users; the additional virtual processor provided an increase of 69 percent.

The test team also compared multiple single-processor VMs with half the number of dual-processor VMs. Although two single-processor VMs supported 2,400 users, compared with 2,200 users for one dual-processor VM, four single-processor VMs and two dual-processor VMs both supported 4,000 users. Because single- and dual-processor VMs performed equally well in these tests, factors such as software licensing costs, ease of management (by using fewer VMs), or corporate standards (from standardizing on a single- or dual-processor image) may dictate whether single-processor VMs or half as many dual-processor VMs are more appropriate.

Measuring response times

The processor, memory, network, and storage virtualization overhead resulted in higher LoadSim response times for VMs than for physical systems. The test team measured 95th-percentile response

times for a fixed number of users on both physical and virtual single- and dual-processor systems. With one processor and 1,300 users, the physical system had a response time of 292 milliseconds, compared with 365 milliseconds for a VM; with two processors and 2,200 users, the physical system had a response time of 260 milliseconds, compared with 440 milliseconds for a VM. All four configurations, however, were well within the 500-millisecond quality-of-service limit defined for the LoadSim tests.

System component performance in the test environment

In addition to measuring benchmark throughputs for different physical and virtual configurations, the test team also analyzed the impact of those throughputs on system components, including the processors, memory, network, and storage, as well as on ESX Server resource management.

Processor performance

As described in the “Test environment and methodology” section in this article, the test team maintained the average host processor utilization at approximately 80 percent during tests of both physical and virtual systems. Differences in performance can be caused partly by the processor overhead required in virtual systems. To

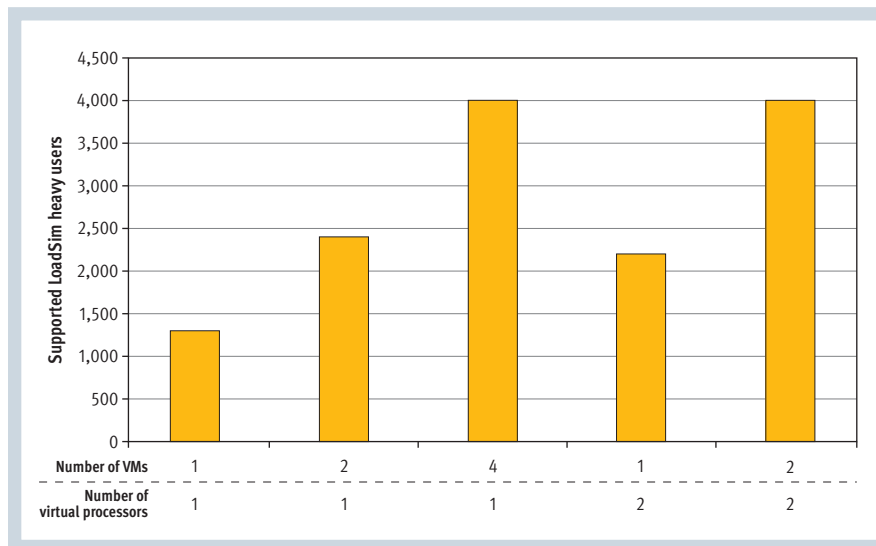


Figure 6. Number of supported users for different numbers of virtual machines and virtual processors

help gauge this overhead, the test team retested the single- and dual-processor physical systems with the maximum number of users supported by the single- and dual-processor VMs at the 80 percent processor utilization level—1,300 and 2,200 users, respectively. With one processor and 1,300 users, the physical processor utilization was 51 percent; with two processors and 2,200 users, the physical processor utilization was 49 percent. The higher utilization in the virtual systems is attributable to virtualization overhead.

Enterprises should keep in mind that the processor virtualization overhead for Exchange is typically not a problem in real-world deployments. A survey conducted by VMware and RapidApp of small, midsize, and large businesses revealed that the weighted average for peak processor utilization in production Exchange servers is approximately 16 percent, leaving substantial resources for virtualization overhead.²

Memory performance

In the Dell tests, even with 16 GB of RAM on the physical host and 4 GB allotted to each of four VMs, the virtual system did not use all of the available memory, leaving approximately 3.5 GB free. This result can be attributed to the transparent page sharing of ESX Server, which shares common memory pages across VMs and allows the overcommitment of VM memory without reducing performance.

An Exchange Server 2003 instance on a physical server running a 32-bit Windows OS can only use up to 4 GB of memory. Because Exchange is memory intensive, this limit on memory can also limit the number of users a server can support. However, using virtualization to run multiple instances of these 32-bit operating systems on the same physical server allows that server to increase memory utilization and support more users than it could otherwise.

Network performance

In the Dell tests, both the physical and virtual systems generated the same amount of network

traffic: with one processor and 1,300 users, they received data at approximately 1 Mbps and transmitted data at approximately 2 Mbps, and with two processors and 2,200 users, they received data at 1 Mbps and transmitted data at 4 Mbps. Adding VMs and virtual processors, and thereby increasing the number of supported users, increased the amount of network traffic both received and transmitted, but even at a high of 4,000 users for four single-processor VMs or two dual-processor VMs, the system received data at a maximum of 3 Mbps and transmitted it at a maximum of 6 Mbps, for a total of less than 10 Mbps—a fraction of the throughput available on the underlying Gigabit Ethernet network.

Storage performance


In the Dell tests, both the physical and virtual systems generated the same amount of storage I/O: with one processor and 1,300 users, they read data at approximately 1 MB/sec and wrote data at approximately 2 MB/sec, and with two processors and 2,200 users, they read data at 3 MB/sec and wrote data at 4 MB/sec. Adding VMs and virtual processors, and thereby increasing the number of supported users, increased the amount of storage I/O both read and written, but even at a high of 4,000 users for four single-processor VMs or two dual-processor VMs, the highest cumulative I/O was approximately 14 MB/sec, a fraction of the throughput available on the 2 Gbps Fibre Channel link. The lack of queued commands reported by esxtop in all tested configurations indicated the absence of any storage bottlenecks.

Resource management in virtual systems

In the Dell tests, the esxtop Memory Overcommitment counter indicated that the virtual systems were not overcommitted, and were not requiring any ESX Server swap activity that might have otherwise reduced performance. As mentioned in the “Memory performance” section in this article, ESX Server hosts use transparent page sharing to help eliminate redundant copies of memory pages, a method that is typically

preferable to swapping and ballooning. Increasing the number of VMs in the Dell tests also increased the amount of memory reclaimed through page sharing; in the test configuration with four VMs, this sharing reclaimed approximately 3.5 GB of memory, an amount that can potentially allow the consolidation of more physical servers onto a virtual platform than would otherwise be possible.

Microsoft Exchange performance and scalability on physical and virtual systems

Understanding how Microsoft Exchange, a communication lifeline for many enterprises, performs in virtualized environments is key to planning data center deployments. In the Dell tests described in this article, a single-processor VM on a Dell PowerEdge 6850 server supported up to 1,300 LoadSim heavy users, and four single-processor or two dual-processor VMs supported up to 4,000 LoadSim heavy users, while still providing acceptable quality of service. Because the performance differences between multiple single-processor VMs and half as many dual-processor VMs were not significant, enterprises can base their choice of VM configuration on factors such as software licensing costs, ease of management, and corporate standards. Although LoadSim results are not necessarily representative of production deployments, and performance will vary depending on the specific infrastructure, Exchange configuration, and workload profile, these results can provide a reference point for this particular configuration using this specific workload profile. 

Kong L. Yang is a systems engineer on the Dell Virtualization Solutions Engineering team. He has an M.S. in Electrical Engineering from the University of Illinois at Urbana-Champaign.

Aravind Pavuluri is a software performance engineer at VMware. He holds a master's degree in Information Networking from Carnegie Mellon University.

²“Deploying Microsoft Exchange in VMware Infrastructure,” by VMware, www.vmware.com/pdf/exchange_best_practices.pdf.