WHITE PAPER

What's New in VMware vSphere™ 4: Performance Enhancements

vmware®

VMware vSphere™ 4, the industry's first cloud operating system, includes several unique new features that allow IT organizations to leverage the benefits of cloud computing, with maximum efficiency, uncompromised control, and flexibility of choice. The new VMware vSphere 4 provides significant performance enhancements that make it easier for organizations to virtualize their most demanding and intense workloads. These performance enhancements provide VMware vSphere 4 with better:

- **Efficiency**: Optimizations resulting in reduced virtualization overheads and highest consolidation ratios.
- **Control**: Enhancements leading to improved ongoing performance monitoring and management, as well as dynamic resource sizing for better scalability.
- **Choice**: Improvements that provide several options of guest OS, virtualization technologies, comprehensive HCL, integrations with 3rd-party management tools to choose from.

This document outlines the key performance enhancements of VMware vSphere 4, organized into following categories:

- Scalability Enhancements
- CPU, Memory, Storage, Networking
- Resource Management
- Performance Management

Finally, the white paper showcases the performance improvements in various tier-1 enterprise applications as a result of these benefits.

Scalability Enhancements

A summary of the key new scalability improvements of vSphere 4 as compared to VMware's previous datacenter product, VMware Infrastructure 3 (VI3), is shown in the following table:

Feature	VI3	vSphere 4
Virtual Machine CPU Count	4 vCPUs	8 vCPUs
Virtual Machine Memory Maximum	64 GB	255 GB
Host CPU Core Maximum	32 cores	64 cores
Host Memory Maximum	256 GB	1 TB
Powered-on VMs per ESX/ESXi Maximum	128	256

For details see Systems Compatibility Guide and Guest Operating System Installation Guide.

Additional changes that enhance the scalability of vSphere include:

- **64 Logical CPUs and 256 Virtual CPUs Per Host** — ESX/ESXi 4.0 provides headroom for more virtual machines per host and the ability to achieve even higher consolidation ratios on larger machines.
- **64-bit VMkernel** — The VMkernel, a core component of the ESX/ESXi 4.0 hypervisor, is now 64-bit. This provides greater host physical memory capacity and more seamless hardware support than earlier releases.
- **64-bit Service Console** — The Linux-based Service Console for ESX 4.0 has been upgraded to a 64-bit version derived from a recent release of a leading Enterprise Linux vendor.

- **New Virtual Hardware** — ESX/ESXi 4.0 introduces a new generation of virtual hardware (virtual hardware version 7) which adds significant new features including:
  - **Serial Attached SCSI (SAS) virtual device for Microsoft Cluster Service** — Provides support for running Windows Server 2008 in a Microsoft Cluster Service configuration.
  - **IDE virtual device** — Ideal for supporting older operating systems that lack SCSI drivers.
  - **VMXNET Generation 3** — See the Networking section.
  - **Virtual Machine Hot Plug Support**— Provides support for adding and removing virtual devices, adding virtual CPUs, and adding memory to a virtual machine without having to power off the virtual machine.

  Hardware version 7 is the default for new ESX/ESXi 4.0 virtual machines. ESX/ESXi 4.0 will continue to run virtual machines created on hosts running ESX Server versions 2.x and 3.x. Virtual machines that use virtual hardware version 7 features are not compatible with ESX/ESXi releases prior to version 4.0.
- **VMDirectPath for Virtual Machines** — VMDirectPath I/O device access enhances CPU efficiency in handling workloads that require constant and frequent access to I/O devices by allowing virtual machines to directly access the underlying hardware devices. Other virtualization features, such as VMotion™, hardware independence and sharing of physical I/O devices will not be available to the virtual machines using this feature. VMDirectPath I/O for networking I/O devices is fully supported with the Intel 82598 10 Gigabit Ethernet Controller and Broadcom 57710 and 57711 10 Gigabit Ethernet Controller. It is experimentally supported for storage I/O devices with the QLogic QLA25xx 8Gb Fibre Channel, the Emulex LPe12000 8Gb Fibre Channel, and the LSI 3442e-R and 3801e (1068 chip based) 3Gb SAS adapters.
- **Increased NFS Datastore Support** — ESX now supports up to 64 NFS shares as datastores in a cluster.

CPU Enhancements

Resource Management and Processor Scheduling

The ESX 4.0 scheduler includes several new features and enhancements that help improve the throughput of all workloads, with notable gains in I/O intensive workloads. This includes:

- Relaxed co-scheduling of vCPUs, introduced in earlier versions of ESX, has been further fine-tuned especially for SMP VMs.
- ESX 4.0 scheduler utilizes new finer-grained locking that reduces scheduling overheads in cases where frequent scheduling decisions are needed.
- The new scheduler is aware of processor cache topology and takes into account the processor cache architecture to optimize CPU usage.

For I/O intensive workloads, interrupt delivery and the associated processing costs make up a large component of the virtualization overhead. The above scheduler enhancements greatly improve the efficiency of interrupt delivery and associated processing.

Memory Enhancements

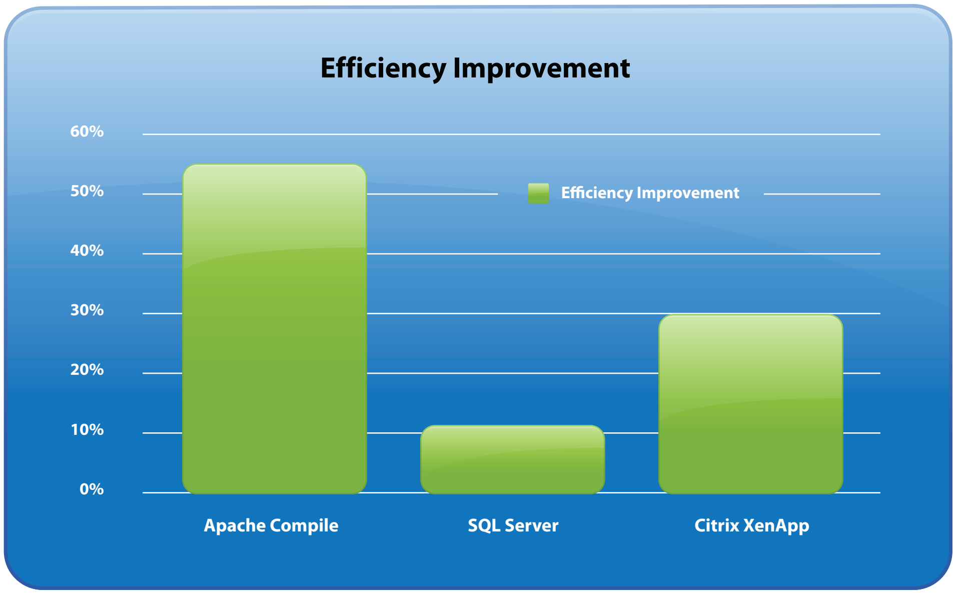
Hardware-assisted Memory Virtualization

Memory management in virtual machines differs from physical machines in one key aspect: virtual memory address translation. Guest virtual memory addresses must be translated first to guest physical addresses using the guest OS's page tables before finally being translated to machine physical memory addresses. The latter step is performed by ESX by means of a set of shadow page tables for each virtual machine. Creating and maintaining the shadow page tables adds both CPU and memory overhead.

Hardware support is available in current processors to alleviate this situation. Hardware-assisted memory management capabilities from Intel and AMD are called EPT and RVI, respectively. This support consists of a second level of page tables implemented in hardware. These page tables contain guest physical to machine memory address translations. ESX 4.0 introduces support for the Intel Xeon processors that support EPT. Support for AMD RVI has existed since ESX 3.5.

Figure 1 illustrates efficiency improvements seen for a few example workloads when using hardware-assisted memory virtualization.

Figure 1 – Efficiency improvements using hardware-assisted memory virtualization



While this hardware support obviates the need for maintaining shadow page tables (and the associated performance overhead) it introduces some costs of its own. Translation look-aside buffer (TLB) miss costs, in the form of increased latency, are higher with two-level page tables than with the one-level table. Using large memory pages, a feature that has been available since ESX 3.5, the number of TLB misses can be reduced. Since TLB miss latency is higher with this form of hardware virtualization assist but large pages reduce the number of TLB misses, the combination of hardware assist and large page support that exists in vSphere yields optimal performance.

Storage Enhancements

A variety of architectural improvements have been made to the storage subsystem of vSphere 4. The combination of the new paravirtualized SCSI driver, and additional ESX kernel-level storage stack optimizations dramatically improves storage I/O performance— with these improvements, all but a very small segment of the most I/O intensive applications become attractive targets for VMware virtualization.

VMware Paravirtualized SCSI (PVSCSI)

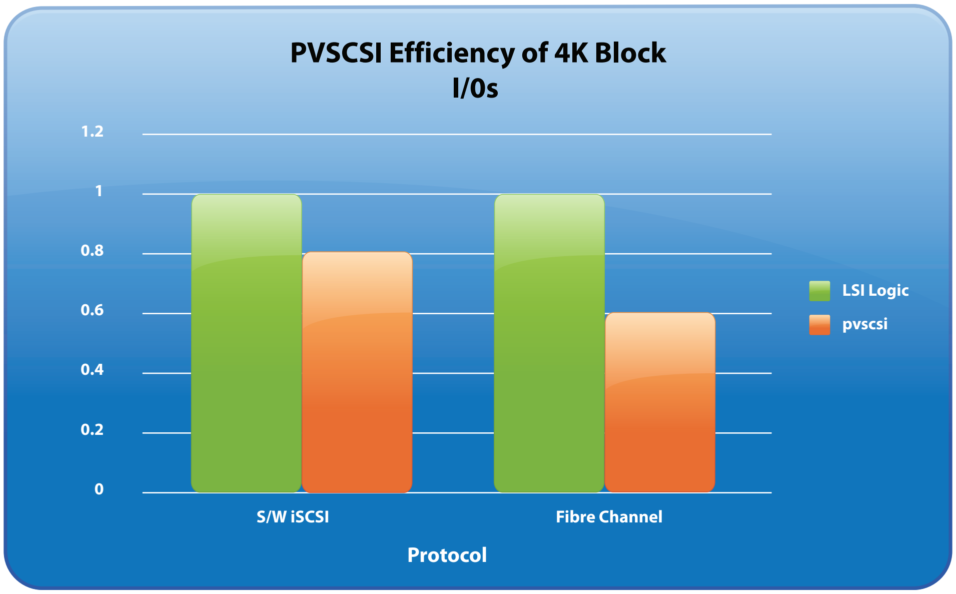
Emulated versions of hardware storage adapters from BusLogic and LSILogic were the only choices available in earlier ESX releases. The advantage of this full virtualization is that most operating systems ship drivers for these devices. However, this precludes the use of performance optimizations that are possible in virtualized environments. To this end, ESX 4.0 ships with a new virtual storage adapter – Paravirtualized SCSI (PVSCSI). PVSCSI adapters are high-performance storage adapters that offer greater throughput and lower CPU utilization for virtual machines. They are best suited for environments in which guest applications are very I/O intensive.

PVSCSI adapter extends to the storage stack performance gains associated with other paravirtual devices such as the network adapter VMXNET available in earlier versions of ESX. As with other device emulations, PVSCSI emulation improves efficiency by:

- Reducing the cost of virtual interrupts
- Batching the processing of I/O requests
- Batching I/O completion interrupts

A further optimization, which is specific to virtual environments, reduces the number of context switches between the guest and Virtual Machine Monitor. Efficiency gains from PVSCSI can result in additional 2x CPU savings for Fibre Channel (FC), up to 30 percent CPU savings for iSCSI.

Figure 2 – Efficiency gains with PV SCSI adapter

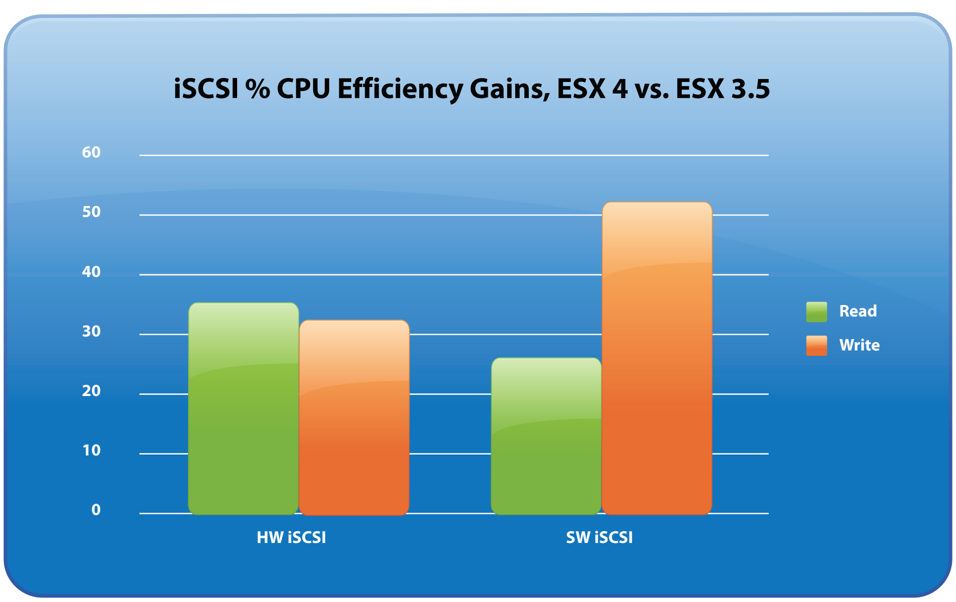


VMware recommends that you create a primary adapter for use with a disk that will host the system software (boot disk) and a separate PVSCSI adapter for the disk that will store user data, such as a database or mailbox. The primary adapter will be the default for the guest operating system on the virtual machine. For example, for virtual machines with Microsoft Windows 2008 guest operating systems, LSI Logic is the default primary adapter.

iSCSI Support Improvements

vSphere 4 includes significant updates to the iSCSI stack for both software iSCSI (that is, in which the iSCSI initiator runs at the ESX layer) and hardware iSCSI (that is, in which ESX leverages a hardware-optimized iSCSI HBA). These changes offer dramatic improvement of both performance as well as functionality of both software and hardware iSCSI and delivering significant reduction of CPU overhead for software iSCSI. Efficiency gains for iSCSI stack can result in 7-26 percent CPU savings for read, 18-52 percent for write.

Figure 3 – iSCSI% CPU Efficiency Gains, ESX 4 vs. ESX 3.5



Software iSCSI and NFS Support with Jumbo Frames

vSphere 4 adds support for Jumbo Frames with both NFS and iSCSI storage protocols on 1Gb as well as 10Gb NICs. The 10Gb support for iSCSI allows for 10x I/O throughput – more details in networking section below.

Improved I/O Concurrency

Asynchronous I/O execution has always been a feature of ESX. However, ESX 4.0 has improved the concurrency of the storage stack with an I/O mode that allows vCPUs in the guest to execute other tasks after initiating an I/O request while the VMkernel handles the actual physical I/O. In VMware's February 2009 announcement on Oracle DB OLTP performance the gains attributed to this improved concurrency model were measured at 5 percent.

Networking Enhancements

Significant changes have been made to the vSphere 4 network subsystem, delivering dramatic performance improvements.

VMXNET Generation 3

vSphere 4 includes, VMXNET3, the third generation of paravirtualized NIC adapter from VMware. New VMXNET3 features over previous version of Enhanced VMXNET include:

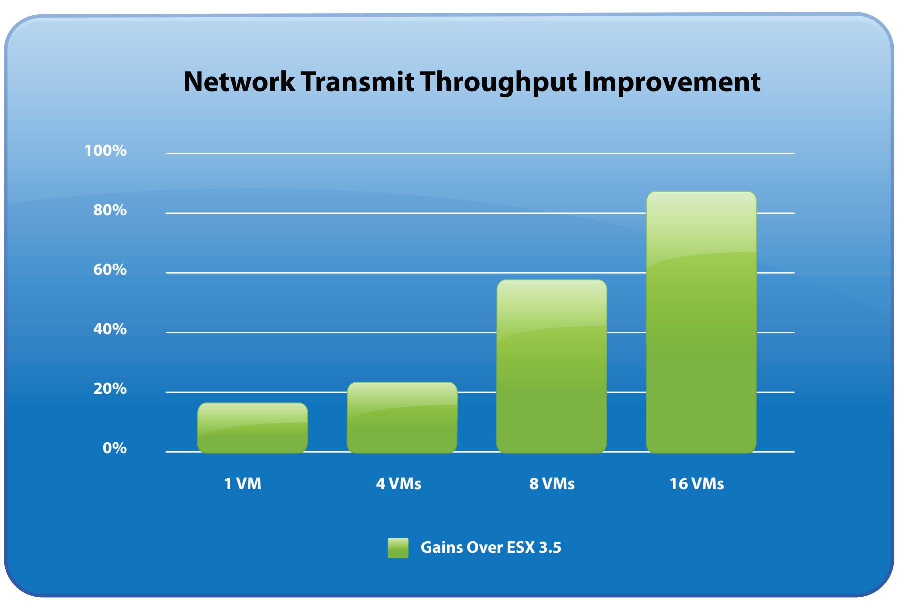
- MSI/MSI-X support (subject to guest operating system kernel support)
- Receive Side Scaling (supported in Windows 2008 when explicitly enabled through the device's Advanced configuration tab)
- IPv6 checksum and TCP Segmentation Offloading (TSO) over IPv6
- VLAN off-loading
- Large TX/RX ring sizes (configured from within the virtual machine)

Network Stack Performance and Scalability

vSphere 4 includes optimizations to the network stack that can saturate 10Gbps links for both transmit and receive side network I/O. The improvements in the VMkernel TCP/IP stack also improve both iSCSI throughput as well as maximum network throughput for VMotion.

vSphere 4 utilizes transmit queues to provide 3X throughput improvements in transmit performance for small packet sizes.

Figure 4 – Network Transmit Throughput Improvement for vSphere 4



vSphere 4 supports Large Receive Offload (LRO), a feature that coalesces TCP packets from the same connection to reduce CPU utilization. Using LRO with ESX provides 40 percent improvement in both throughput and CPU costs.

Resource Management Enhancements

VMotion
Performance enhancements in vSphere 4 reduce time to VMotion a VM by up to 75 percent.

Storage VMotion Performance
Storage VMotion is now fully supported (experimental before) and has much improved switchover time. For very I/O intensive VMs, this improvement can be 100x. Storage VMotion leverages a new and more efficient block copy mechanism called Changed Block Tracking, minimizing CPU and memory resource consumption on the ESX host up to two times.

Figure 5 – Decreased Storage VMotion Time

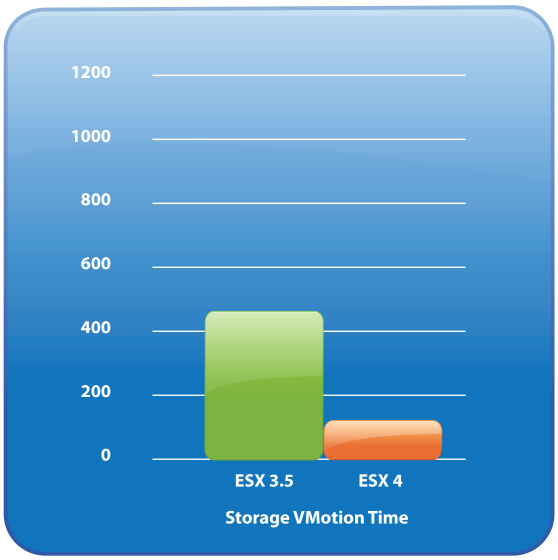


Figure 6 – Improved VMFS Performance

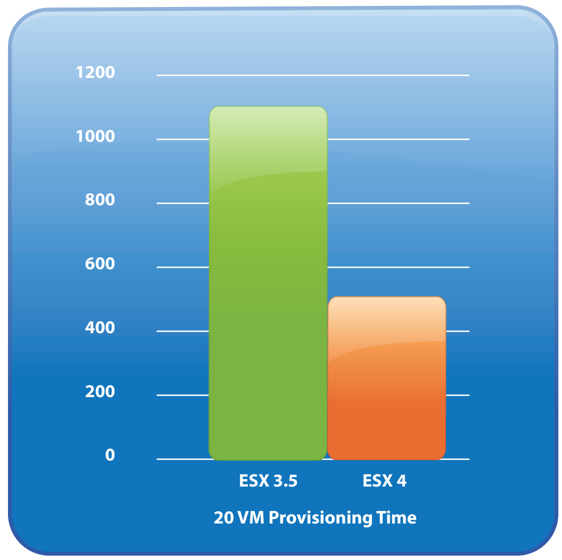


Figure 7 – Performance Enhancements Lead to a Reduced Time to VMotion

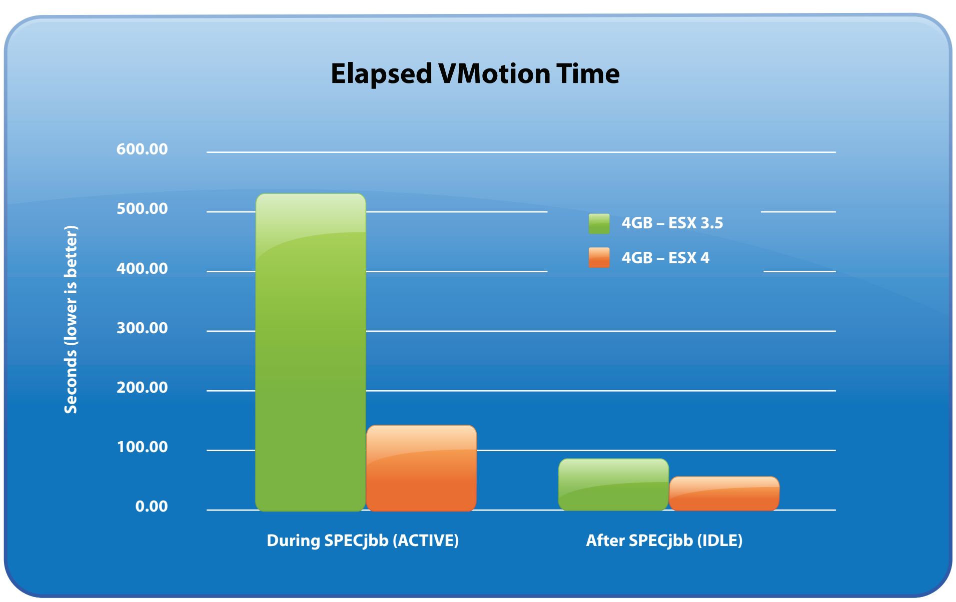
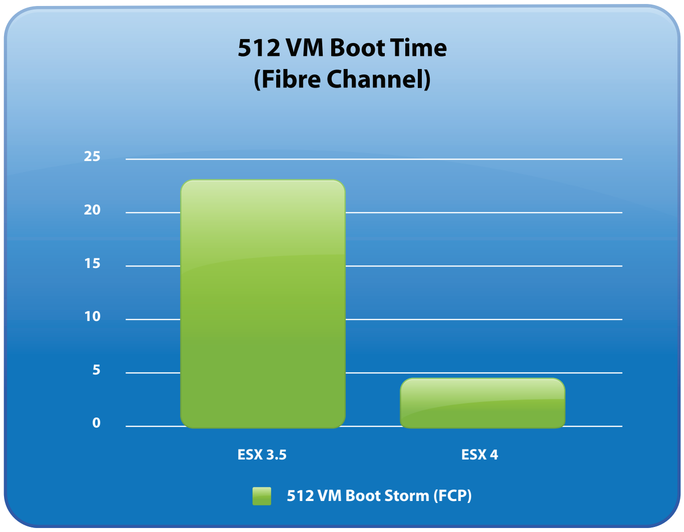


Figure 8 – Time to Boot 512 VDI VMS



VM Provisioning

VMFS performance improvements offer more efficient VM creation and cloning. This use case is especially important with vSphere's more ambitious role as a Cloud operating system.

Performance Management Enhancements

Enhanced vCenter Server Scalability

As organizations adopt server virtualization at an unprecedented level, the need to manage large scale virtual data centers is growing significantly. To address this, vCenter Server, included with vSphere 4, has been enhanced to manage up to 300 hosts and 3000 virtual machines. You also have the ability to link many vCenter Servers in your environment with vCenter Server Linked Mode to manage up to 10,000 virtual machines from a single console.

vCenter Performance Charts Enhancements

Performance charts in vCenter have been enhanced to provide asingle view of all performance metrics such as CPU, memory, disk, and network without navigating through multiple charts. In addition, the performance charts also include the following improvements:

- Aggregated charts show high-level summaries of resource distribution that is useful to identify the top consumers.
- Thumbnail views of hosts, resource pools, clusters, and data stores allow for easy navigation to the individual charts.
- Drill down capability across multiple levels in the inventory helps in isolating the root cause of performance problems quickly.
- Detailed data store level views show utilization by file type and unused capacity.

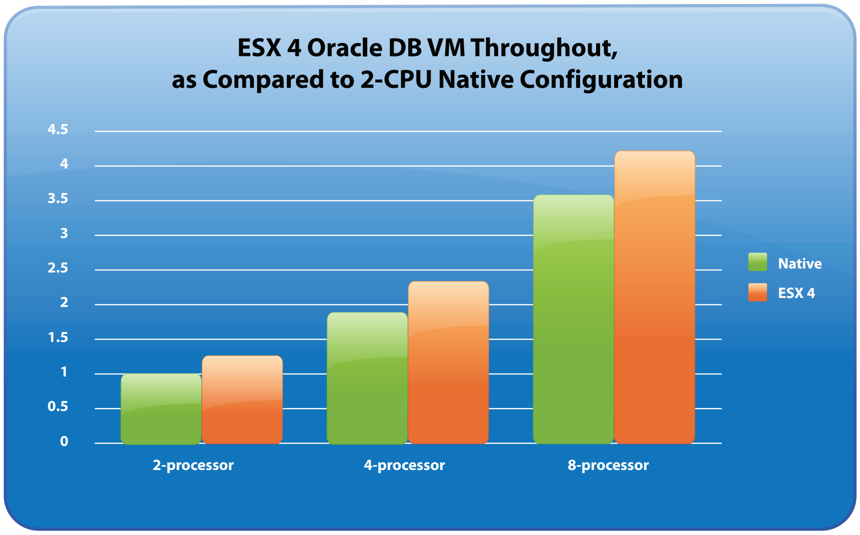
Application Performance

Oracle

VMware testing has shown that running a resource-intensive OLTP benchmark, based on a non-comparable implementation of the TPC-C* workload specification, Oracle DB in an 8-vcpu VM with vSphere 4 achieved 85 percent of native performance. This workload demonstrated 8,900 database transactions per second and 60,000 disk input/outputs per second (IOPS). The results demonstrated in this proof point represent the most I/O-intensive application-based workload ever run in an X86 virtual environment to date.

*The benchmark was a fair-use implementation of the TPC-C business model; these results are not TPC-C compliant results, and not comparable to official TPC-C results. TPC Benchmark is a trademark of the TPC.

Figure 9 – Comparison of Oracle DB VM Throughput vs. 2-CPU Native Configuration

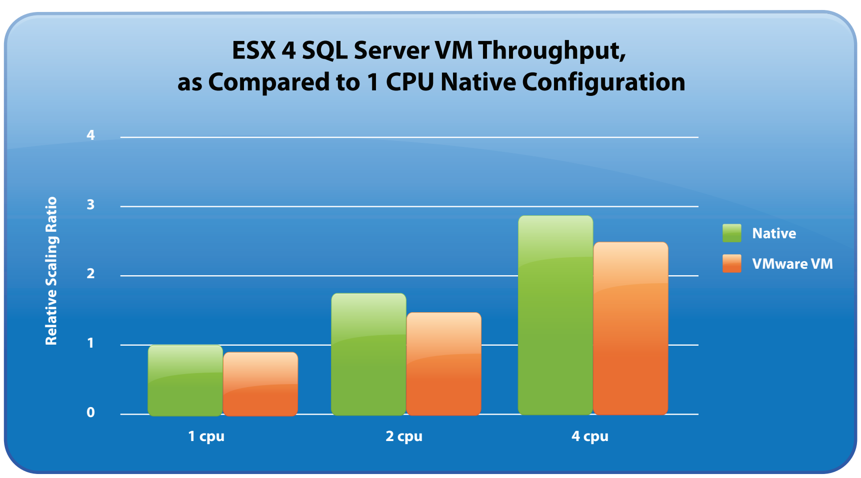


The results above were run on a server with only eight physical cores, resulting in an 8-way VM configuration that was not under-committing the host. The slightly less committed four vCPU configuration ran at 88 percent of native.

SQL Server

Running an OLTP benchmark based on a non-comparable implementation of the TPC-E* workload specification, a SQL Server virtual machine with four virtual CPUs on vSphere 4.0 showed 90 percent efficiency with respect to native. The SQL Server VM with a 500 GB database performed 10,500 IOPS and 50 Mb/s of network throughput.

Figure 10 – Comparison of vSphere 4 SQL Server VM Throughput vs. Native Configuration

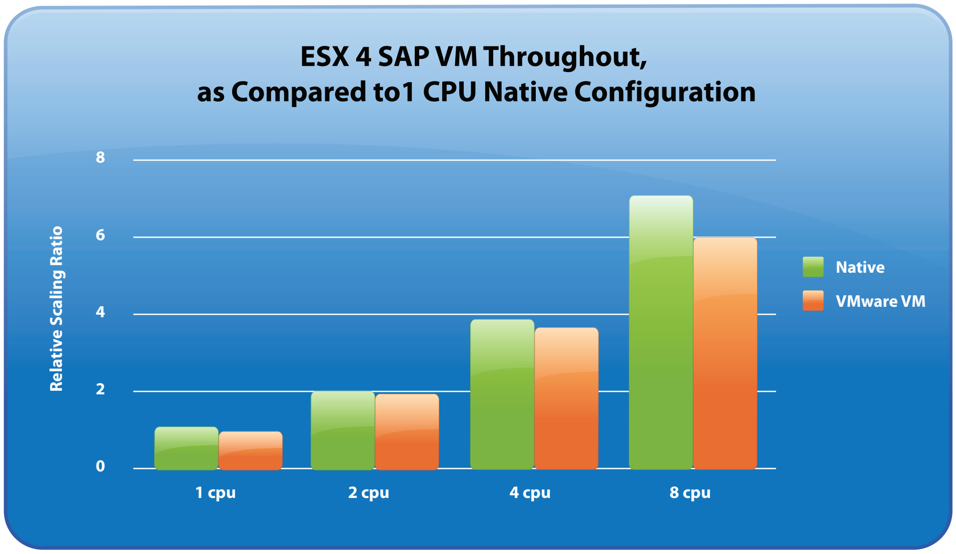


SAP

VMware testing demonstrated that running SAP in a VM with vSphere 4 scaled linearly from one to eight vCPUs per VM and achieved 95 percent of native performance on a standard 2-tier SAP benchmark. This multi-tiered application architecture includes the SAP application tier and back-end SQL Server database instantiated in a single virtual machine.

*The benchmark was a fair-use implementation of the TPC-C business model; these results are not TPC-C compliant results, and not comparable to official TPC-C results. TPC Benchmark is a trademark of the TPC.

Figure 11 – Comparison of ESX 4 SAP VM Throughput vs. Native Configuration

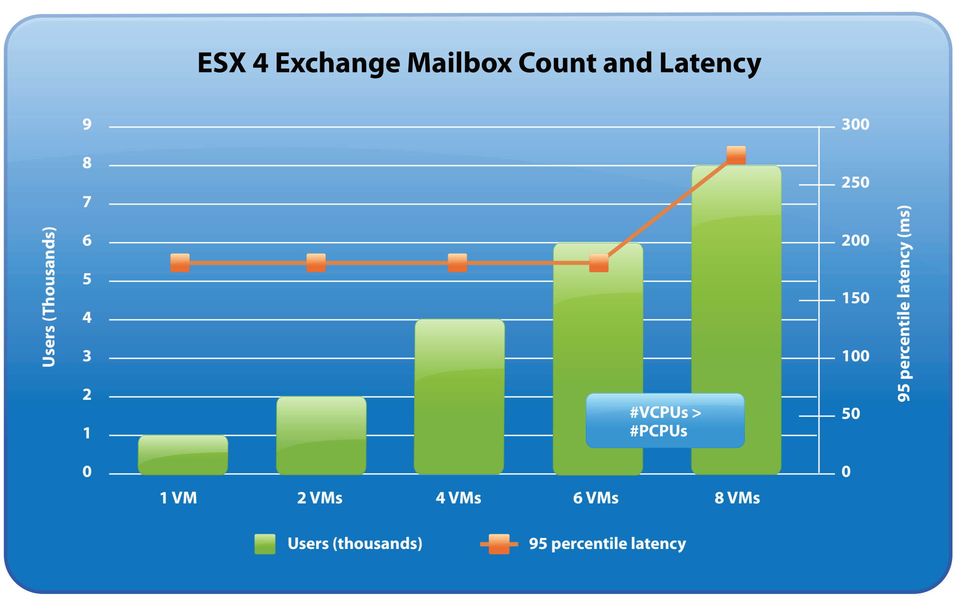


Exchange

Microsoft Exchange Server is one of the most demanding applications in today's datacenters, save the very largest databases being deployed. Previous work on virtual Exchange deployments showed VMware's ability to improve performance from native configurations by designing an Exchange architecture with a greater number of mailbox instances running fewer mailboxes per instance.

With the performance enhancements added to vSphere 4 single VM Exchange mailboxes have been demonstrated at up to 8,000 mailboxes per instance. This means that Exchange administrators will have the option of choosing the higher performing smaller mailboxes or the more cheaply licensed large mailbox servers.

Figure 12 – vSphere performance enhancements with Microsoft Exchange



Summary

VMware innovations continue to make VMware vSphere 4 the industry standard for computing in data centers of all sizes and across all industries. The numerous performance enhancements in VMware vSphere 4 enable organizations to get even more out of their virtual infrastructure and further reinforce the role of VMware as industry leader in virtualization.

vSphere represents dramatic advances in performance compared to VMware Infrastructure 3 to ensure that even the most resource intensive and scale out applications such as large databases and Microsoft Exchange email systems can run on private clouds powered by vSphere.

References

Performance Evaluation of AMD RVI Hardware Assist
http://www.vmware.com/pdf/RVI_performance.pdf

Performance Evaluation of Intel EPT Hardware Assist
http://www.vmware.com/pdf/Perf_ESX_Intel-EPT-eval.pdf

VMware, Inc. 3401 Hillview Ave Palo Alto CA 94304 USA Tel 877-486-9273 Fax 650-427-5001 www.vmware.com