WHITE PAPER

What's New in VMware® Infrastructure 3: Performance Enhancements

vmware®

Table of Contents

The new features in VMware® Infrastructure 3 makes it even easier for organizations to virtualize their most demanding and intense workloads. The new version of VMware Infrastructure 3 provides significant performance enhancements, including the release of VMware ESX Server 3.5 and a new ultra-thin hypervisor called VMware ESX Server 3i that can significantly improve the performance of virtualized workloads.

This document outlines the key performance benefits of upgrading to get the new features of VMware Infrastructure 3, including:

- Scalability enhancements
- New guest OS support
- Networking enhancements
    - o Enhanced VMXNET
    - o TCP segmentation offload (TSO)
    - o Jumbo frames
    - o 10 Gigabit Ethernet
    - o Support for Intel® I/O Acceleration Technology
- Support for paravirtualization
- Support for large memory pages
- NUMA optimizations
- Infiniband support

Scalability Enhancements

with the latest enhancements to VMware Infrastructure 3,„the memory limit for VMware ESX Server hosts has been increased to 256GB, and organizations can now assign up to 64GB or memory to each virtual machine. ESX Server 3.5 now also make use of 32 logical processors, with experimental support of as many as 64 logical processors on a single host. These increases in memory and CPU enable organizations to host more virtual machines on each ESX Server host, with larger guests running more applications in each virtual machine. Organizations can also get better performance for 64-bit applications such as databases by avoiding expensive storage access because they can use larger application caches.

In addition, VMware Distributed Resource Scheduler (DRS) and VMware High Availability (HA) now support up to 32 ESX Server nodes in a cluster, doubling the number of nodes supported. Furthermore, VMware VirtualCenter 2.5 supports management of up to 200 running hosts and 2,000 powered-on virtual machines.

New Guest Operating System Support

Virtual Infrastructure 3 now provides even greater scalability and compatibility with the addition of support for several new guest operating systems:

- Windows® Server™ 2008
- Windows® Vista®
- Red Hat® Linux 5
- Ubuntu® Linux 7.04 (paravirtualized and fully virtualized)

Furthermore, Virtual Infrastructure 3 now supports any updates available for guest operating systems that were already supported.

Networking Enhancements

Significant changes have been made to the ESX Server network system, delivering dramatic performance improvements. In addition, Virtual Infrastructure 3 enables support for 10 Gigabit Ethernet networking, as well as support for Infiniband through community source collaboration with Mellanox Technologies.

VMXNET Enhancements

VMware Infrastructure 3 provides a new version of the VMXNET virtual device (the VMware paravirtualized virtual networking device for guest operating systems) called Enhanced VMXNET. Enhanced VMXNET includes several new networking I/O performance improvements, such as support for TCP/IP Segmentation Offload (TSO) and jumbo frames. Enhanced VMXNET also includes limited 64-bit guest operating system support. All other networking features, such as teaming/VLANs are fully supported. To enable this functionality in a supported guest, simply select Enhanced VMXNET when installing the device in the guest using the VMware Tools. Organizations can also use command line interface (CLI) to configure MTU (max packet size), as well as TSO enablement/disablement.

TCP Segmentation Offload (TSO)

At the application level, data transmitted from one system to another must be segmented to fit into the network packets. The size of those packets is limited by the Ethernet specification. Historically, segmentation was performed by the operating system (OS) using the CPU. Modern network interface cards (NIC) try to optimize this TCP segmentation by using larger segment size as well as offloading work from the CPU to NIC hardware. ESX Server 3.5 utilizes this concept to provide a virtual NIC with TSO support—without requiring specialized network hardware.

TSO improves networking I/O performance by reducing the CPU overhead involved with sending large amounts of TCP traffic. TSO improves performance for TCP data coming from a VM and for network traffic sent out of the server, such as VMware VMotion traffic. TSO is supported in both the guest operating system and in the ESX Server kernel TCP/IP stack, and is enabled by default in the VMkernel. To take advantage of TSO, you must select "Enhanced VMXNET" or "e1000" as the virtual networking device for the guest. When the guest operating system can utilize TSO, virtual machines running on ESX Server 3.5 will show lower CPU utilization than virtual machines that lack TSO support, when performing the same network activities.

When the physical NICs provide TSO functionality, ESX Server 3.5 can leverage the specialized NIC hardware to improve performance. However, performance improvements related to TSO need not require NIC hardware support for TSO.

Figure 1 illustrates the percentage network throughput improvement we observed for various message and socket sizes when using RedHat Enterprise Linux 5 and Windows Server 2003 guest operating systems with TSO-enabled virtual NIC.

Figure 1 illustrates the percentage network throughput improvement we observed for various message and socket sizes when using RedHat Enterprise Linux 5 and Windows Server 2003 guest operating systems with TSO-enabled virtual NIC.

Jumbo Frames

Since the Ethernet specification was developed decades ago, packets have been transmitted over the network in sizes no greater than 1,500 bytes. For each packet, the system has to perform a fixed amount of work to package and transmit the packet. As Ethernet speed increased, so did the amount of work necessary, which resulted in a greater burden on the system.

Recent advances in all areas of the network stack have enabled an increase in the Ethernet packet size to 9,000 bytes. These so-called "jumbo frames" decrease the number of packets requiring packaging compared to previously sized packets. That decrease results in less work for network transactions which frees up resources for other activities.

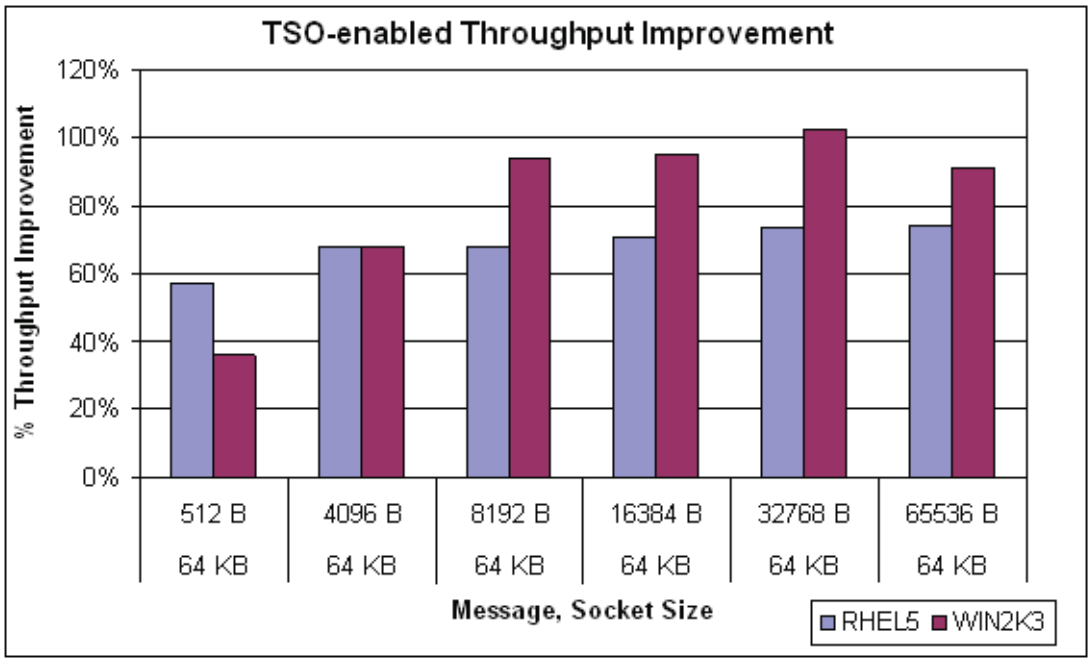

Figure 1: Network throughput improvement with TSO-enabled virtual NIC

ESX Server 3.5 has implemented support for jumbo frames up to 9KB (9,000 bytes). When supported by the system software and hardware, as well as switches and hubs in between client and server, ESX Servers using jumbo frames will realize a decrease in load due to network processing. Like TSO, jumbo frames are supported in both the guest operating system and in the ESX Server kernel TCP/IP stack.

To enable jumbo frames in a virtual machine, configure the guest to use "Enhanced VMXNET" network device using VMware Tools. Jumbo frames support is disabled by default in the VMkernel and needs to be enabled using CLI.

NetQueue

ESX Server 3.5 now supports NetQueue, which improves performance of 10 Gigabit Ethernet network communication. NetQueue requires MSI-X support from the server platform, so support is limited to specific systems and is turned off by default. Check the VMware ESX Server 3.5 hardware compatibility list (HCL) for information on whether support for NetQueue on a particular server is provided.

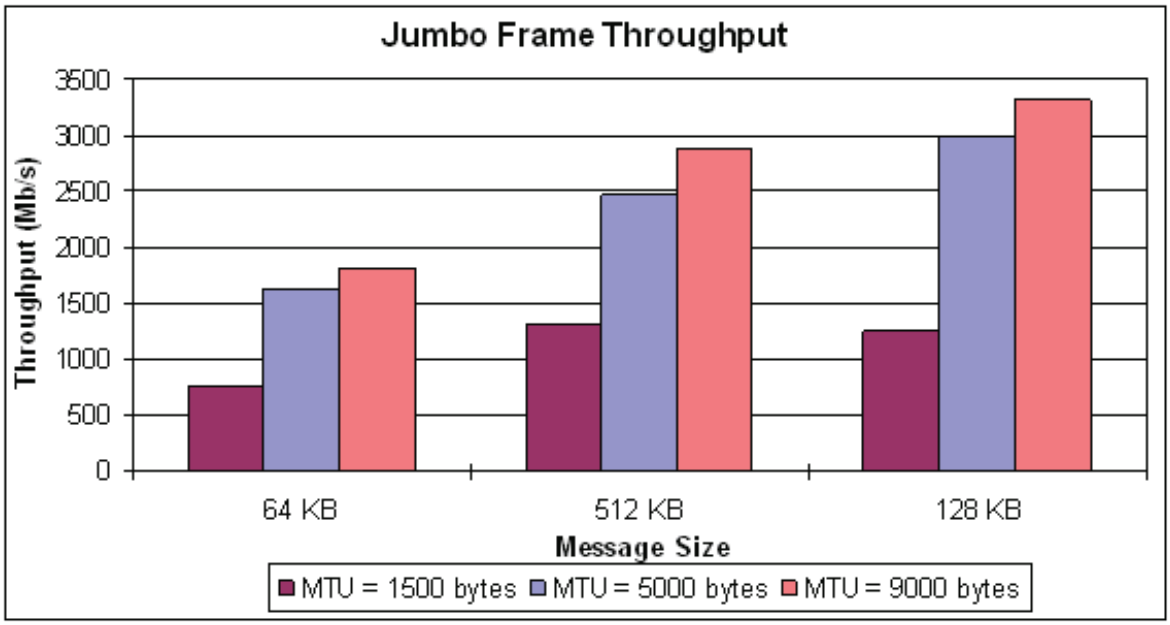

Figure 2 illustrates the network throughput observed for various message sizes using jumbo frames.

10 Gigabit Ethernet

10 Gigabit Ethernet (or 10 GigE) is the result of network hardware manufacturers implementing an IEEE standard for faster networks. In the presence of NICs and interconnecting hardware such as switches that support 10 GigE, performance improvements of an order of magnitude can be realized versus traditional 100 Megabit Ethernet.

ESX Server 3.5 fully supports10 GigE NICs. This means that NICs and network switches can run in between virtual machines on two ESX Server 3.5 hosts supporting 10 GigE.

Intel I/O Acceleration Technology Support (Experimental)

ESX Server 3.5 provides experimental support for Intel I/O Acceleration Technology (I/OATv1). I/OATv1 is a chipset on the motherboard that speeds up memory copies. ESX Server 3.5 takes advantage of this chipset, if present, when dealing with memory copies in the TCP/IP stack implementation and makes improvements in networking performance.

CPU Enhancements

While CPU speed has increased significantly since the previous release of VMware Infrastructure 3, software optimization can always be implemented to improve data center productivity. ESX Server 3.5 capitalizes on improvements made to Linux operating systems and AMD processors, reducing the overhead created by the virtualization software itself.

Paravirtualized Linux Guests

Virtual Infrastructure 3 provides support for Virtual Machine Interface (VMI) 3.0, which is the standard adopted in the Linux kernel version 2.6.22. The VMI API standardizes the manner in which hypervisors interact with paravirtualized guests. Unlike the traditional paravirtualization approach that entails modifications to OS components, thus requiring operating system vendors to maintain separate kernel binaries, the VMI API standard allows the same kernel to run natively and in a paravirtualized VM, as well as support multiple hypervisors with the same kernel binary. Such "transparent paravirtualization" enables any off-the-shelf, unmodified Linux distributions using Linux kernel version 2.6.22 or later to be run as a paravirtualized guest on VMware Infrastructure 3.

When VMware Infrastructure 3 was released, Ubuntu Linux 7.04 (Feisty Fawn) – both desktop and server versions – had been certified, and more Linux distributions are now being certified. Novell has announced inclusion of VMI support in SUSE 10 SP2, which is expected early in 2008. When virtual machines have these operating systems installed as guests, near-native performance can be added to the long list of features in VMware Infrastructure 3.

However, VMI is not limited to Linux. OS vendors are free to modify their operating systems to use the VMI interface and deliver improved guest OS performance and timekeeping when running on virtual machines. To enable paravirtualization support for a virtual machine on VMware Infrastructure 3, you must edit the virtual machine hardware settings and select the check box related to paravirtualization support under "Options > Advanced".

Figure 3 illustrates the percentage performance improvement observed for various workloads using VMI-enabled virtual machine.

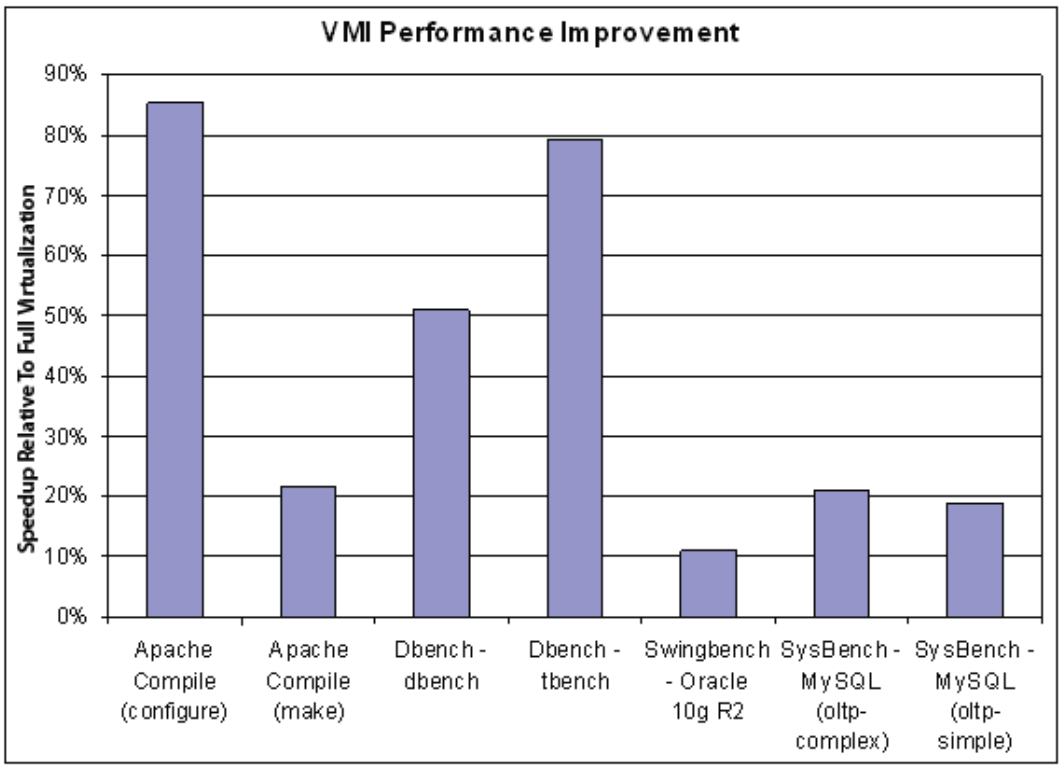

Figure 3: Performance improvements with a VMI-enabled VM

Memory Enhancements

Large Memory Pages

The virtual memory of traditional operating systems maps addresses from the virtual memory space to the physical memory space. Instead of mapping each individual byte, groups of addresses—or "pages"—are mapped between these spaces, and offsets allow for access of bytes within each page. As memory usage grows, the number of pages that must be maintained also increases. On x86 processors, operating systems typically manage memory in 4KB blocks. Using and managing small (4KB) pages can be inefficient for applications with large code and working sets. By increasing the size of the memory pages, fewer virtual-to-physical mappings are needed to process execution. Support for large (2MB) pages in modern CPUs helps to reduce this overhead. Operating system and application support for large memory pages has shown measurable improvement in performance.

The ESX Server 3.5 can now assign 2MB machine pages to guest operating systems that request them. Guest operating systems that support 2MB pages will achieve higher performance when backed with 2MB machine memory pages. ESX Server 3.5 therefore delivers a commensurate performance increase as a result of reduced memory management overhead. With ESX Server 3.5, the performance benefits observed by using large memory pages in the native environments are now possible in virtual environments, as well.

NUMA Improvements

Today's powerful servers commonly use multiple processors to increase system performance. While some architectures provide a single bank of memory to be shared by all processors, another technique for system design is to dedicate a bank of memory to each processor. With this design, processors get faster, unshared access to their local memory but pay a penalty in access latency when accessing memory dedicated to the other processor. This design is called Non-uniform Memory Access (NUMA). Making the best use out of NUMA systems means scheduling processes to run on processors with access to the node's local memory. While ESX Server 3 supported NUMA systems, significant improvements have been made to the NUMA scheduling algorithms in ESX Server 3.5. These changes will greatly benefit virtual machines running on NUMA platforms.

Storage Enhancements

A variety of architectural improvements have been made to the storage subsystem of ESX Server 3.5. Each of these changes alone could increase throughput and reduce CPU effort for virtualization. Combined, they can greatly improve the performance of all virtual machines running in a VMware Infrastructure 3 environment. However, this document focuses only on the performance-related Infiniband support feature.

Infiniband Support

Virtual Infrastructure 3 now supports Infiniband-based Host Channel Adapters (HCA) from Mellanox Technologies. ESX Server 3.5 includes extensions that allow an OFED (OpenFabrics Enterprise Distribution) InfiniBand stack to be plugged in and used. The kernel extensions should work for any vendor, but have been certified only with Mellanox Technologies at the release time. We expect that these extensions will be available and generally supported for all VMware co-development partners in the future. Until then, the extensions are supported for Mellanox Technologies, and potentially for others who work directly with VMware to test their offerings. This feature is a result of the VMware Community Source co-development effort with Mellanox Technologies. Support for this feature is provided by Mellanox Technologies as part of the VMware Third-Party Hardware and Software Support Policy.

Infiniband support is completely hidden from the view of the IT manager and completely transparent to the virtual machines. There are no changes required in the virtual machines when installed in the VMware environment. The virtual machines continue to see networking NICs and Fibre channel HBAs. VirtualCenter works as before along with VMware VMotion and other VMware features. Instead of using the NIC and Fibre Channel adapters on your servers, you would install Infiniband HCA adapters. The virtual machines never know that Infiniband HCAs exist; instead, they behave as if they are interacting with Fibre Channel HBAs and Ethernet NICs.

Summary

VMware innovations continue to make VMware Infrastructure 3 the industry standard for computing in data centers of all sizes and across all industries. The numerous performance enhancements in VMware ESX Server 3.5 and other components of the new VMware Infrastructure 3 enable organizations to get even more out of their virtual infrastructure and further reinforce the role of VMware as industry leader in virtualization.

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

vmware®