Improved Virtualization Performance with 9th Generation Servers

David J. Morse
Dell, Inc.
August 2006

Contents

Executive Summary

Dell's ninth-generation of PowerEdge servers provide an unparalleled level of performance over the previous generation, with Intel Core Microarchitecture processors and fast memory speeds with FB-DIMMs. This paper will quantify these performance gains within a virtualized environment by using VMware Infrastructure 3. VMware ESX Server 3 is the foundation of this infrastructure, and provides enhanced virtual machine performance through several significant architectural improvements. This paper will highlight both the hardware and the software advantages of running VMware on Dell PowerEdge servers using SPECjbb2005, an industry-standard benchmark that measures Java Virtual Machine performance. The results demonstrated a gain of up to 2.2X from a PowerEdge 1850 compared to the Dell PowerEdge 1950.

Introduction

VMware ESX Server 3.0

VMware ESX Server 3.0 is a robust virtualization layer between the hardware and operating system. Each virtual machine has its own set of resources (CPU, memory, networking, and disks) which ESX Server transparently presents to unmodified "guest" operating systems. The advantages of virtualization with VMware are numerous: server consolidation, high availability, dynamic resource allocation, and hosting legacy applications just to name a few. Most of these are beyond the scope of this paper; for more information, visit http://www.dell.com/content/topics/global.aspx/alliances/en/vmware. ESX Server 3.0 was chosen for this performance comparison, because it provides numerous performance improvements over 2.5, including:

- Performance scalability as more virtual machines are added
- Virtual SMP, which allows a guest operating system to use up to four virtual processors simultaneously
- Significant networking enhancements
- Improved memory management unit (MMU) handling
- Linux native POSIX thread library (NPTL) support

The benchmark runs referenced in this paper benefit from the first two improvements listed above.

SPECjbb®2005

SPECjbb2005 is a benchmark designed to measure the performance of a system running a Java server application. The benchmark simulates a wholesaling operation, receiving orders, managing deliveries, and generating reports of various sorts; the database is replaced by in-memory Java Collection objects, and transaction logging is implemented using XML. Several key components that this benchmark stresses include the server's processing power, the underlying memory subsystem, and the Java runtime environment. The benchmark does not stress the network or disks, which is actually desirable, since database benchmarks will often bottleneck on disk subsystem (requiring many external disk arrays); this makes it difficult to get a true measure of the performance of the server itself. Since SPECjbb2005 stresses key internal system components (CPU and memory), it is an obvious choice to showcase the architectural improvements on the improvements in the 9th generation of PowerEdge servers.

BEA JRockit®

BEA JRockit was chosen as the Java Virtual Machine to run the SPECjbb2005 benchmark in these tests, since it has demonstrated record-breaking performance on a Dell PowerEdge server[1]. JRockit is designed for optimal performance for

[1] "BEA Demonstrates Market's Fastest Benchmark Results on BEA JRockit Using Dual-Core Intel Xeon Processors": http://www.bea.com/framework.jsp?CNT=pr01674.htm&FP=/content/news_events/press_releases/2006.

Java-based applications in large-scale, enterprise-wide environments. With it, Java developers can achieve out-of-the-box application performance and scalability for their applications without needing to know JVM internals. Instead, progressive optimization features and dynamic memory management can enable the JVM to automatically deliver constant, optimal application performance.

Hardware/Software Configurations Tested

The following table lists the configurations used for the tests. The configurations of the 8th generation and 9th generation PowerEdge servers were kept as similar as possible, so as not to artificially hinder one platform or the other.

	Dell PowerEdge 1850	**Dell PowerEdge 1950**
Processors	2 x Dual-Core Intel Xeon Processors at 2.8GHz/2x2M cache	2 x Dual-Core Intel Xeon 5160 Processors at 3.0GHz/4M cache
Front-side bus	800MHz	1333MHz
Memory size/type/speed	12GB[2] DDR2 400MHz (6X2GB) single-ranked DIMMs	16GB FBDIMM 667MHz (8x2GB), dual-ranked DIMMs
BIOS	A05	1.1.0
BIOS settings (deviations from defaults)	Logical Processor disabled[3] Sequential Memory access disabled Virtualization Technology not available	Adjacent Cache Line Prefetch disabled Hardware Prefetcher disabled Virtualization Technology enabled for 64-bit tests[4]
Host Operating System	VMware ESX Server 3.0 (managed by VirtualCenter 2.0)	VMware ESX Server 3.0 (managed by VirtualCenter 2.0)
Guest Operating Systems	Microsoft Windows Server 2003 SP1, Enterprise Edition	Microsoft Windows Server 2003 SP1, Enterprise Edition
Guest Java Virtual Machine	BEA JRockit R26.4.0-63 (32-bit)[5]	BEA JRockit R26.4.0-63 (32-bit) BEA JRockit P26.4.0-10 (64-bit)[6]
Disk Controller	PERC4e/Di (embedded)	PERC 5/i (embedded)
Disk Capacities	1 x 36GB 15K RPM SCSI 1 x 73GB 10K RPM SCSI	1 x 36GB 15K RPM SAS 1 x 73GB 15K RPM SAS

Table I. Server configurations

[2] The PowerEdge 1850 supports a maximum of 12GB; the extra 4GB on the PE1950 (which can support up to 32GB) was not utilized during these benchmarks.
[3] HyperThreading was disabled since it was not found to be beneficial for this benchmark.
[4] Note that this setting is not needed for 32-bit VMs; it enables ESX 3.0 to run 64-bit VMs (experimental). See http://www.vmware.com/support/kb/enduser/std_adp.php?p_faqid=1901 (N/A on 8th-generation PowerEdge servers).
[5] Available at http://commerce.bea.com/products/weblogicjrockit/5.0/jr_50.jsp.
[6] Available at http://dev2dev.bea.com/jrockit/releaseupdate.html.

VMware Host/Guest Configuration

VMware ESX Server 3.0 was easily installed on both the PowerEdge 1850 and PowerEdge 1950, and both servers are officially supported by VMware[7]. An OptiPlex GX280 with a 3.6GHz CPU and 4GB of RAM running Microsoft Windows Server 2003, Enterprise Edition was configured to act as a VirtualCenter Server. VirtualCenter Server allows easy administration of multiple ESX hosts and their virtual machines via the Virtual Infrastructure Client.

First, Microsoft Windows Server 2003 was installed as a guest operating system within a virtual machine on the PowerEdge 1850. This VM was configured with 1 virtual CPU (vCPU), 4 GB of RAM, and a 10GB virtual SCSI disk. VMware Tools was then installed, which is highly recommended by VMware, as it improves VM performance by installing enhanced device drivers in addition to other helpful features (consult the *VMware Infrastructure 3 Basic Administration Guide* for a complete list). The BEA JRockit Java Runtime Environment was installed, along with the SPECjbb2005 benchmark to stress the VM.

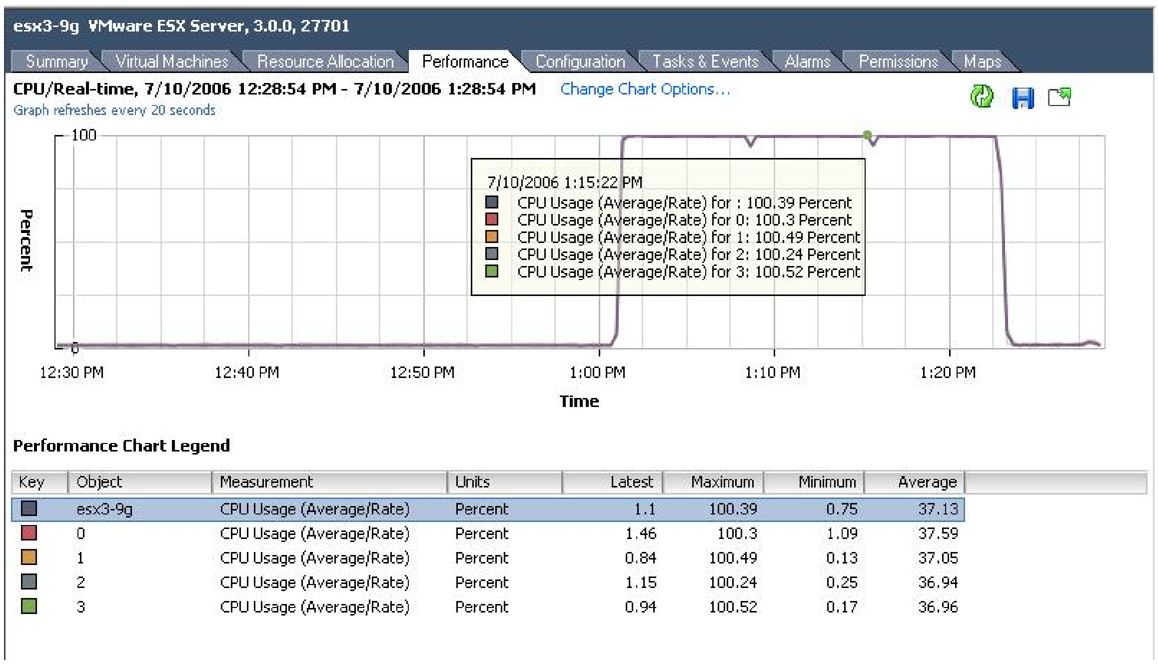
Once this first virtual machine was configured and verified to run the benchmark successfully, it was then cloned until both the PowerEdge 1850 and 1950 each had four VMs. Both physical servers had four cores, so four 1 vCPU virtual machines could saturate the physical server by running the SPECjbb2005 benchmark in parallel. The Virtual Infrastructure Client made the cloning process as easy as right-clicking the source virtual machine and selecting “Clone”. Since all the hardware is virtualized at the guest operating system level, virtual machines can be cloned or migrated between different server models with no reconfiguration needed (regardless of physical server differences such as number of processors, disk controllers, network cards, etc.). The only change necessary after each cloning operation was to assign a unique IP address and computer name to the newly-cloned guest operating system.

To measure the performance benefits of ESX Server 3.0, an additional set of scaling experiments was run with two 2-way Virtual SMP (which stresses two physical cores during the benchmark run). A new (and currently experimental) feature in ESX 3.0, 64-bit guest operating system support, was also tested, since this feature allows operating systems to address larger physical and virtual address spaces. These experiments could only be run on the PowerEdge 1950, as VMware relies upon hardware virtualization support to run 64-bit guests. This is present in the new Intel Xeon 5000 series processors for the 1950, but not on the 1850.

[7] Systems Compatibility Guide for ESX Server 3.0: http://www.vmware.com/pdf/vi3_systems_guide.pdf

Benchmarking Methodology and Performance Results

SPECjbb2005 was designed to run single or multiple instances of a Java virtual machine on one single physical server. Although the multi-JVM support uses TCP sockets to coordinate the multiple JVM instances, it does not currently have the necessary hooks to coordinate JVMs running in parallel across multiple servers (physical or virtual). To facilitate executing the benchmark in parallel, a simple script was written that would iterate through each virtual machine and start the SPECjbb2005 benchmark simultaneously. As the screenshot below from the Virtual Infrastructure Client indicates, the four virtual machines started and stopped their stress on each server's processor at the same time:



CPU Utilization during SPECjbb2005 Benchmark

To determine the physical server's overall performance in a virtualized environment, the benchmark's metric (SPECjbb2005 business operations per second, or bops) from each virtual machine was collected and summed together.

The first set of runs consisted of 4 virtual machines, each with 1 virtual CPU, 3GB RAM, and Windows Server 2003 (32-bit) running BEA JRockit. As the table below indicates, the PowerEdge 1950 with Intel Xeon 5160 dual-core processors was able to achieve over double the throughput of the fastest dual-core processor available on the PowerEdge 1850.

	SPECjbb2005 bops				
Server	**VM1**	**VM2**	**VM3**	**VM4**	**Aggregate**
PowerEdge 1850	8643	8705	8607	8708	34663
PowerEdge 1950	19375	19241	19272	19267	77155
				Improvement:	**2.2X**

To observe the increased virtual machine scalability improvements in ESX Server 3.0, two machines were powered off in each physical server, and the two remaining VMs were assigned two virtual CPUs instead of one:

	SPECjbb2005 bops		
Server	**VM1**	**VM2**	**Aggregate**
PowerEdge 1850	18532	18648	37180
PowerEdge 1950	39615	39713	79328
		Improvement:	**2.1X**

The table above indicates the aggregate throughput is higher relative to the 4-VM runs. This is actually due to the way the SPECjbb2005 metric is calculated; by default, the benchmark calculates the throughput score by averaging the points from N to 2*N warehouses, where N is the number of processors within the guest operating system. The table below shows which warehouses are used to calculate the metric in different virtualized environments:

Virtual CPUs within guest operating system	Warehouse points used to calculate score
1	1,2
2	2,3,4
4	4,5,6,7,8

Unfortunately, the just-in-time (JIT) compilation and optimization performed by the JVM does not complete by the first warehouse measurement point, so the throughput reported in the 1-vCPU tests is taken from data points before the JVM has fully "warmed up". This explains why the total throughput of four 1-vCPU VMs is slightly lower than two 2-vCPU VMs. There is a way to work around this behavior by setting the `input.expected_peak_warehouse` property of the benchmark, but the result must be submitted to SPEC for review[8].

The final test was to measure the performance and stability of the 64-bit guest operating system support on the PowerEdge 1950. Two virtual machines were installed with Windows Server 2003 x64 and the 64-bit version of BEA JRockit. The heap size allocated to the JVM could then be raised from just under 2GB (the default user-space limitation of Windows 2003 Server 32-bit) to almost 4GB. This improved performance over the 32-bit configuration by 3.5%; performance metrics can not be publicly quoted, however, since SPEC run rules requires a

[8] See http://www.spec.org/jbb2005/docs/RunRules.html for more information on SPECjbb2005 run and reporting rules.

recommended, supported, and encouraged configuration. It is expected that a follow-up release to ESX Server 3.0 will add full support for 64-bit operating systems.

Tip for Performance Benchmarking Virtual Infrastructure 3
It was observed while running benchmarks like SPECjbb2005 in a virtualized environment that the Virtual Infrastructure Client would trigger a warning if the host or any guest virtual machine's memory or CPU usage exceeded 75% (which is prudent for a production environment, where utilizations this high are undesirable), and triggered an error if it exceeds 90% (see figure below).

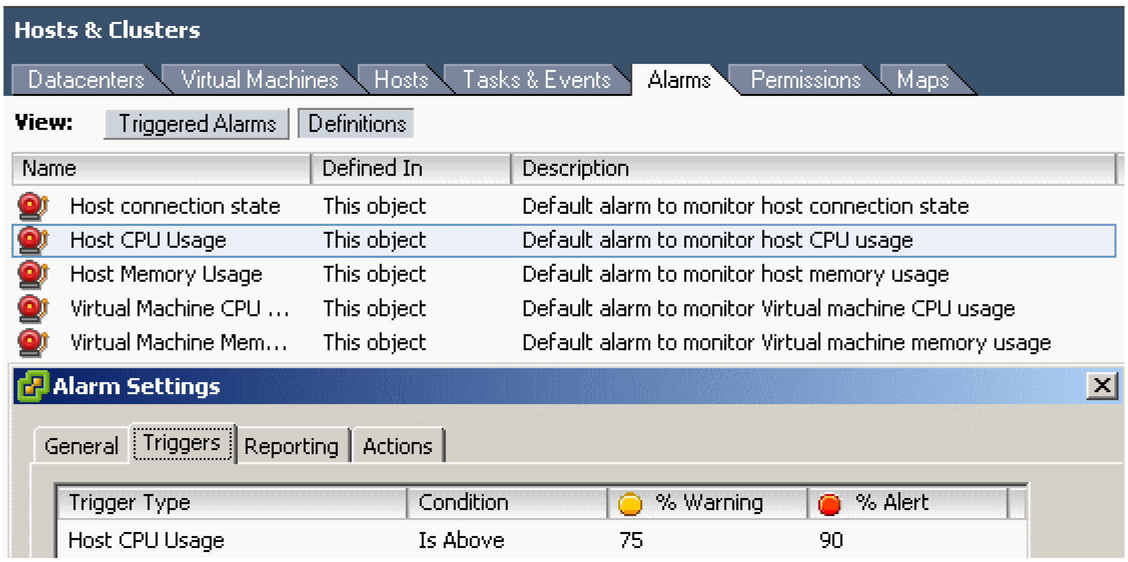


Figure 1. Default Alarms in VirtualCenter 2.0

The nature of performance benchmarking involves stressing these subsystems to their limits, so these default alarms were disabled for these tests to avoid the alarms constantly being triggered. It is important to stress, however, that real-world deployments of VMware Virtual Infrastructure should keep these alarms enabled, as they are a valuable indicator that more resources should be allocated to the host or guest.

Improved Virtualization Performance with 9th Generation Servers

In summary, significant performance gains within a virtualized environment were measured in the new 9th generation of PowerEdge servers as compared to the prior generation in Dell's benchmark testing. SPECjbb2005 aggregate throughput more than doubled in all cases. While customer configurations will undoubtedly vary, SPECjbb2005 is both a processor- and memory-intensive benchmark. These test results were able to demonstrate the benefits of architectural improvements in Intel's new Core Microarchitecture, faster memory with FB-DIMMs, and hardware virtualization to enable 64-bit guest operating system support in a virtualized environment.

Microsoft does not support issues that occur in Microsoft operating systems or programs that run in a virtual machine until it is determined that the same issue can be reproduced outside the virtual machine environment. See Microsoft Knowledge Base article ID 897615 for more information.