WHITE PAPER

Virtualization with VMware® ESX 4.1: Performance and Best Practices on Intel® processor-based SGI® Rackable™ Servers

October 2010

As we try to drive physical servers to higher utilizations by using multiple virtual machines simultaneously, it becomes critical to have meaningful and precise metrics so that we can effectively compare the suitability and performance of different hardware platforms for virtual environments. Traditional standard benchmarks, consider neither virtual machines nor server consolidation in mind. These metrics focus on achieving maximum system performance for a single workload by driving at least one of the underlying hardware resources into saturation. Clearly, a more sophisticated approach is required to quantify a virtualization environment's performance and ability to run an increasing number of diverse virtual machines as physical resources increase. The VMmark standard benchmark was developed by VMware® to address these goals. This paper describes the performance results achieved with VMmark benchmarks on Intel® Xeon® 5500 processor and Intel® Xeon® 5600 processor based SGI® Rackable™ C2104-TY3 and C2005-TY15 servers respectively, using VMware® ESX 4.1 virtualization software. The paper also recommends best practices for attaining maximum consolidation capacity with the best possible performance level in a virtualized environment.

1. Need for a standard virtualization benchmark

With the increase in the number of powerful multi-core processor servers, the degree of underutilization tends to increase. These realities have led to renewed interest in virtual machines for server consolidation. Virtual machine environments provide a software layer that enables users to create multiple independent virtual machines on the same physical server. By running multiple virtual machines simultaneously, a physical server can be driven to much higher utilizations, albeit with some performance overhead. Although the underlying physical resources are shared, each virtual machine is fully isolated from other virtual machines and executes a separate operating system instance and separate application software. An effective solution is thus needed to measure the performance and scalability of such virtualized environments.

Conventional standard server benchmarks are more designed to measure the performance of a single application running on a single operating system, inside a single physical computer. They cannot accurately measure the performance of a single application running inside a virtual machine, nor can they measure the performance of a single hardware platform used to run several virtual machines at once. These conventional standard benchmarks also cannot stress the physical resources and virtualization layer sufficiently to fully capture the complexity of running multiple different workloads within a server consolidation framework.[1]

A more sophisticated standard benchmark is thus required to quantify a virtualization environment's performance and ability to run an increasing number of diverse virtual machines as physical resources increase, to mimic a consolidated datacenter environment. The benchmark must scale in a controlled fashion to make comparisons between systems meaningful. The benchmark must also exhibit stable, reproducible performance.

1. http://www.vmware.com/products/vmmark/

2. The VMmark benchmark

The VMmark benchmark has been developed by VMware®, Inc., to create a meaningful measurement of virtualization performance across a wide range of hardware platforms. It is a tile-based benchmark consisting of several familiar workloads running simultaneously in separate virtual machines. Each workload component is based upon a single-system benchmark running at less than full utilization. This collection of different workloads is aggregated into a unit of work referred to as a tile. A "Tile" is thus defined as a collection of virtual machines executing a set of diverse workloads on a server platform, running multiple OSes.

The performance of each workload is measured and used to form an aggregate score for the tile. The scores generated when running multiple tiles simultaneously may be summed to increase the overall benchmark score. A system's "consolidation capacity" is a coarse-grain measure of the total number of tiles that it can accommodate with a desirable performance score.

VMmark provides a High-precision Scoring Methodology which allows for the integration of the different component metrics into an overall score. Once a VMmark test completes, each individual workload reports its relevant performance metric.

The VMmark benchmark (version 1.1.1) is based upon a set of relevant datacenter workloads (Table 1):

- Mail server
- Java server
- Standby server
- Web server
- Database server
- File server

The Mail server workload simulates businesses today that provide employees with email as a means for communicating, using the most common mail server, Microsoft Exchange 2003. The Java Server workload is a modified version of SPECjbb2005 which measures Java performance in many modern multi-tiered applications. The Standby Server workload simulates those present in large computing environments, ready for new workloads, or for workloads with bursty behavior and can impact performance of other virtual machines. It is present along with the other workloads, for the VMmark benchmark to be considered valid. A collection of Web Servers, often pervasive in modern data centers are strong candidates for consolidation. The E-commerce profile in SPECweb2005 is used to run this workload. The Database Server is the MySQL database with the open source SysBench benchmark tool simulating users who are repeatedly executing a pre-defined mix of transactions. The File Server workload simulates service requests from Windows 95 clients and is measured by the dbench application. The dbench application is derived from industry-standard NetBench benchmark, which requires a large number of client systems to generate the system load. For more details on VMmark, please refer to: http://www.vmware.com/pdf/vmmark_intro.pdf.

Workload	Benchmark/ Application Name	Virtual Machine Platform	Performance Metric
Mail server	Exchange 2003	Windows 2003 32-bit, 2 CPU, 1GB RAM	Actions/minute
Java server	SPECjbb®2005-based	Windows 2003 64-bit, 2 CPU, 1GB RAM	New orders/minute
Standby Server	None	Windows 2003 32-bit, 1 CPU, 256MB RAM	None
Web server	SPECweb®2005-based	SLES 10 64-bit, 2 CPU, 512MB RAM	Accesses/minute
Database server	MySQL	SLES 10 64-bit, 2 CPU, 2GB RAM	Commits/minute
File server	dbench	SLES 10 32-bit, 1 CPU, 256MB RAM	MB/second

Table 1: VMmark 1.1.1 Workloads and Performance Metrics

3. VMmark Benchmarks on SGI® Rackable™ servers

This section describes the hardware and software configurations used to run the VMmark v1.1.1 benchmarks on Intel® Xeon® 5500 processor based SGI® Rackable™ C2104-TY3 server and Intel® Xeon® 5600 processor based SGI® Rackable™ C2005-TY15 server, as well as the benchmark results.

3.1 Benchmark Configuration

Servers:

- 1x SGI® Rackable™ C2104-TY3 system with
  - 2x Intel® Xeon® 5590 processors 3.33 GHz (Turbo Enabled up to 3.60 GHz), 12x 8GB DIMMS 1066 MHz, HT enabled;
  - Disk Subsystem Type: Internal RAID for OS, FC SAN for VMs;
  - Disk Controller: LSI Logic 1064E Based SAS RAID card.

- 1x SGI® Rackable™ C2005-TY15 system with
  - 2x Intel® Xeon® 5680 3.33 GHz processors (Turbo Enabled up to 3.60 GHz), 12x 16GB DIMMS 1333 MHz, HT enabled;
  - Disk Subsystem Type: Internal RAID for OS, FC SAN for VMs;
  - Disk Controller: Onboard Intel Embedded Server SATA RAID Adapter ICH10R.

Virtualization SW:

- VMware® ESX 4.1 (build 260247);

VM Guest OSes:

- MS Windows Server 2003 EE 32bit and 64bit R2, SP2;
- SUSE Enterprise Linux 10 SP2 32bit and 64bit.

Network:

- 1x QLogic Dual-Port 8Gbit fibre channel HBA (qle2562);
- 1x Onboard Intel 82576 Dual-Port Gigabit Ethernet Controller;
- 2x Single-Port Chelsio N310E 10Gbit Ethernet;
- SMC 8848M Switch with 10GigE card;
- 10Gbps network speed for all VMs except webserver VMs. 1Gbps for webserver VMs.

Storage:

- 1x SGI® InfiniteStorage IS4600, firmware 07.60.31.00
  - 168 disks with 22 LUNs over 8x146GB 15K RPM drives (for C2104-TY3);
  - 256 disks with 32 LUNs over 8x146GB 15K RPM drives (for C2005-TY15);
- 1x Brocade 4800 Director Switch w/ 16port 8Gbit blade.

Clients:

- 18x SGI® Rackable™ C1001-TY5 each with 2x Intel® Xeon® processor L5530 @ 2.4GHz, 12GB (for SGI® Rackable™ C2104-TY3);
- 27x SGI® Rackable™ C1001-TY5 each with 2x Intel® Xeon® processor L5530 @ 2.4GHz, 12GB (for C2005-TY15);
- OS: MS Windows Server 2003 EE 32bit R2, SP2.

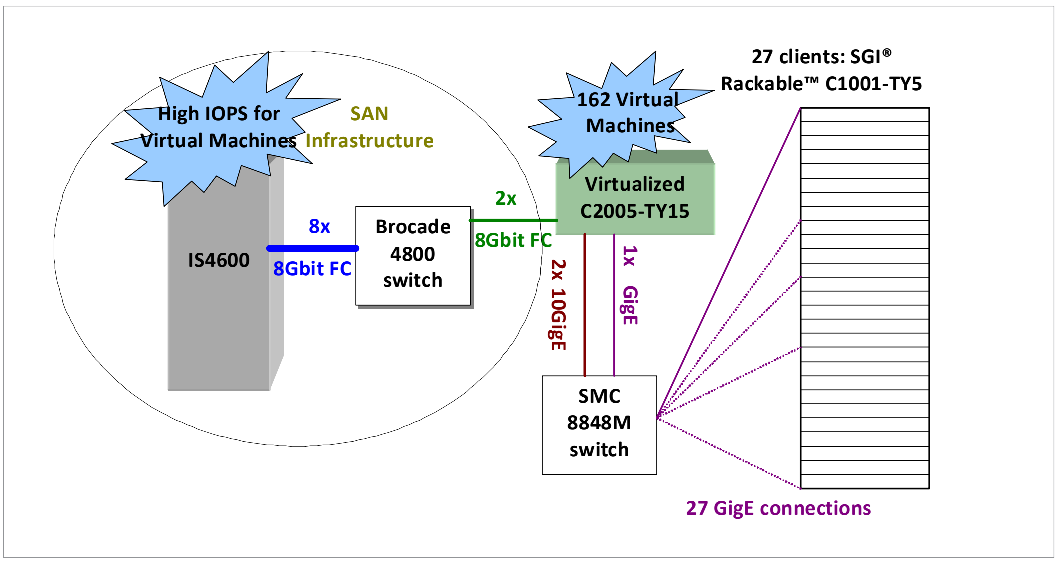

Figure 1: VMmark benchmark on SGI® Rackable™ C2005-TY15 server: Schematics

3.2 Benchmark Results

The VMmark scoring methodology involves test runs, where each individual workload reports its relevant performance metric (Table 1).

Metrics are collected at frequent intervals during the course of the run. The standard VMmark workload is designed to run for 3 hours with workload metrics reported every 60 seconds. This means that rather than having a single number upon completion of a test run, the user will have a series of numbers for each of the five workloads. However, each workload score is defined as a single number: the average of a consecutive subset of the series of data points for that workload. The resulting per-tile scores are then summed to create the final metric. Normalization allows the integration of the different component metrics into an overall score. The benchmark helps measure the virtualization overheads of the individual workloads as well as the scalability of the entire system. Therefore, results for multi-tile runs are reported as the aggregate score for all tiles, the individual scores for each of the tiles, and the scores for the workloads within the tiles.

NOTE: The standby server workload does not produce a metric that affects the benchmark score. However, the standby server is required to answer to a periodic heartbeat request in order for the VMmark test to be considered valid.

For more information on VMmark scoring methodology, please refer to: http://www.vmware.com/products/vm-mark/overview.html.

Results as of the date this paper was written (Oct 6, 2010) show that SGI® Rackable™ C2104-TY3 with 2x Intel® Xeon® 5590 processors 3.33 GHz (8 cores/16 threads), 96GB 1066MHz has the highest VMmark score of 25.67 compared to competitive servers based on Intel® Xeon® 5500 processors (Figure 2). One SGI® Rackable™ C2005-TY15 server with 2x Intel® Xeon® 5680 3.33 GHz (12 cores/24 threads), 192GB, 1333MHz has the third highest VMmark score of 36.76 compared to competitive servers based on Intel® Xeon® 5600 processors (Figure 3). A single SGI® Rackable™ C2104-TY3 server and a single SGI® Rackable™ C2005-TY15 server can consolidate 108 and 162 virtual machines on 18 and 27 tiles respectively. The tiles are running different types of workloads on different OSes (Figure 4). The high IOPS from the SGI® InfiniteStorage IS4600 subsystem that enable low access time/latency for virtual machines; the fast memory bandwidth of Intel® Xeon® 5590 processor-based SGI® Rackable™ C2104-TY3 server and Intel® Xeon® 5680 processor-based SGI® Rackable™ C2005-TY15 server and the network configured for high bandwidth from the overall topology - all contributed to high VMmark scores with a consolidation capacity of more than 100 virtual machines on a single node. For full SGI® disclosures, please refer to: http://www.vmware.com/products/vmmark/results.html.

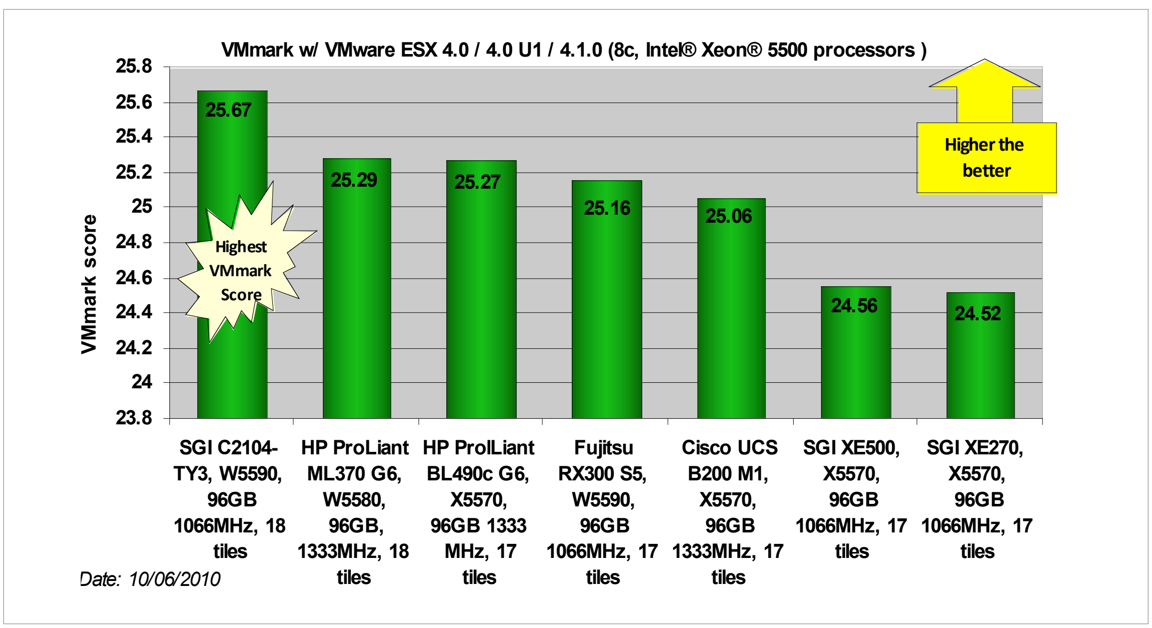

Figure 2: VMmark Scores: Intel® Xeon® 5500 processors, 96GB

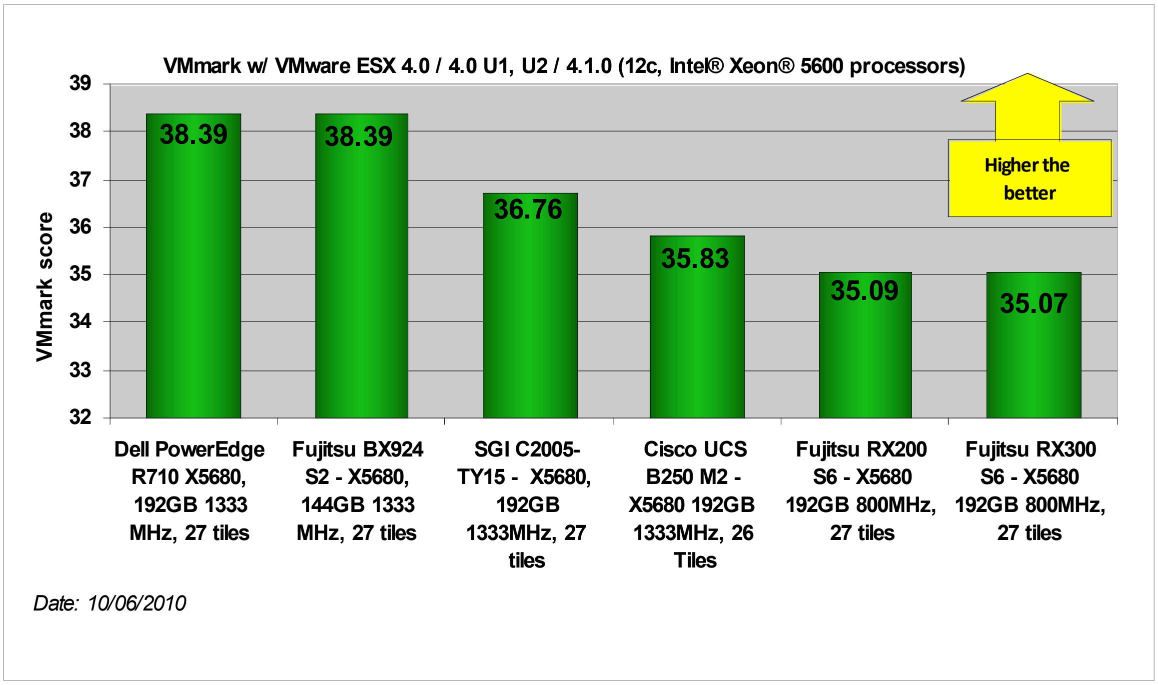

Figure 3: VMmark Scores: Intel® Xeon® 5680 processors

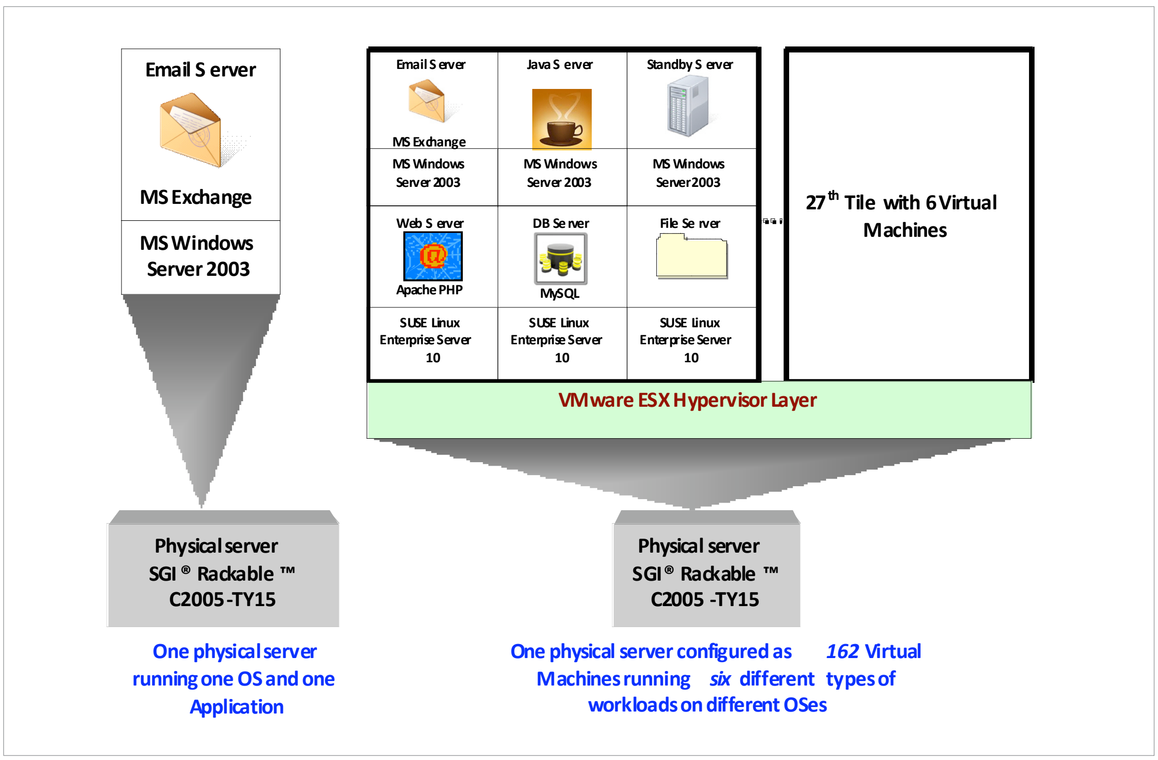

Figure 4: VMmark: Using a consolidation strategy with VMware® ESX 4.1

4. Memory Size and Bandwidth

This section describes the various tests performed to evaluate performance of a virtualized system with a) varied memory sizes and b) a large memory size versus high memory bandwidth.

Results on SGI® Rackable™ C2005-TY15 show that the total system memory plays an important role in VMware® performance. With 192 GB (12x 16GB), the throughput from a virtualized system improves with the increase in the number of tiles, peaking at 27th tile. With 96 GB (12x 8 GB), the throughput drops after 23rd tile and the benchmark begins to fail during runtime (Figure 5).

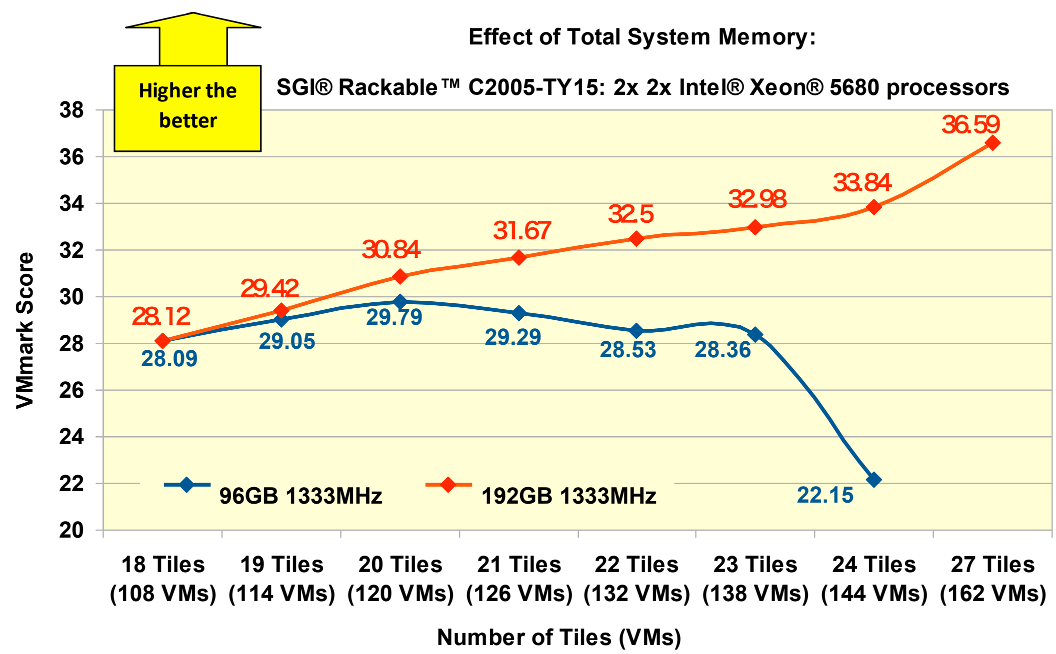

Figure 5: Effect of Total System Memory Size on performance of a virtualized system

Results also show that the VMmark score is highly memory-bandwidth sensitive. On a server with 1066 MHz memory bandwidth (96GB RAM), the performance score is always higher than 800 MHz (144GB RAM), while maintaining the best scalability up to 17 tiles. After 17 tiles (102 VMs), the performance of each individual VM also drops. With 800MHz, one may consolidate more tiles with a trade-off in performance. So it is important to understand the goal of virtual server(s) in production - whether it is to yield the highest performance or highest consolidation capacity. One must then configure the server(s) to meet those goals. Figure 6 shows the results from the memory bandwidth tests performed on an SGI® Altix® XE270 (Intel® Xeon® 5570 processor based) server.

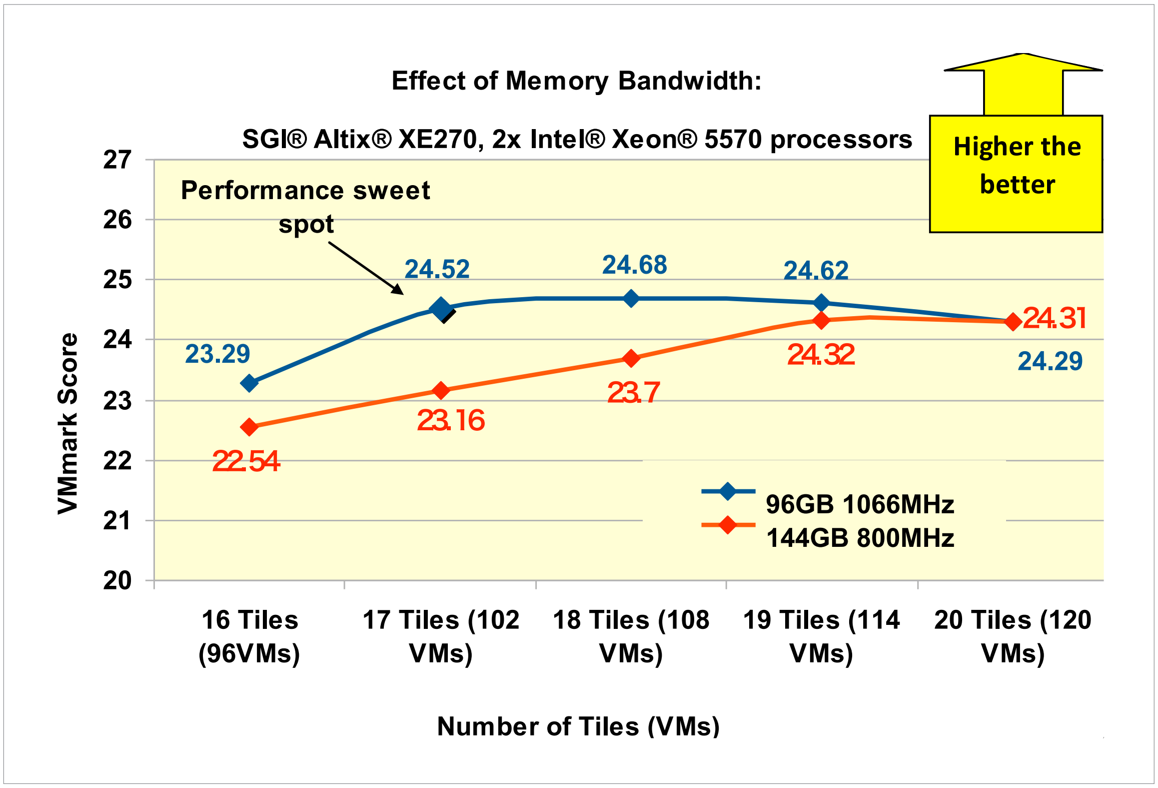

Figure 6: Effect of Memory Size vs. Memory Bandwidth on performance and consolidation capacity of a virtualized system

5. Virtualization with VMware® ESX 4.1: Best Practices

This section describes some of the best practices derived from the VMmark benchmarks on Intel® Xeon® 5500 processor based and Intel® Xeon® 5600 processor based SGI® Rackable™ C2104-TY3 and C2005-TY15 servers, respectively. These recommendations are subject to change while moving to a different platform supporting a different motherboard and processor SKU with variations of BIOS options, or while using different storage subsystem or network hardware/interface. It is recommended to derive a new set of best practices on each virtual platform under test.

The best practices for configuring a virtualized Intel® Xeon® 5500 processor based and Intel® Xeon® 5600 processor based SGI® Rackable™ servers using VMware® ESX 4.1 are as follows:

- With an increased core count and larger memory capabilities of Intel® Xeon® 5600 processor-based systems, it is important to choose a higher memory capacity for a system planned for virtualization. Tests on a 12-core Intel® Xeon® 5680 processor-based SGI® Rackable™ C2005-TY15 server show that with 192 GB (12x 16GB), the throughput of the VMs improves with the increase in the number of tiles, whereas with 96 GB (12x 8GB), throughput drops at a lower number of tiles and benchmark begins to fail during runtime (Memory speed in both cases were 1333 MHz).

- Understanding customer goals is important – whether it is a) the performance of individual virtual machines or b) the consolidation capacity of the system HW within an acceptable trade-off in performance.

One can achieve a higher consolidation factor with more tiles using a larger memory size, but with a trade-off in performance due to reduced memory bandwidth. Using a faster memory bandwidth with a reduced memory size may limit the total number of virtual machines that can possibly run on the system, thus reducing its consolidation factor.

For example, tests on an 8-core Intel® Xeon® 5570 processor-based SGI® Altix® XE270 server show that using 1066 MHz memory speed (96GB) yields the performance sweet spot with 17 tiles (102 Virtual Machines), whereas, using 800 MHz (144 GB) can consolidate more virtual machines within the server but with a lower performance.

- One needs to maintain the default settings for Hyper-threading and Intel Turbo Boost Technology as Enabled.
- For VMware®, the most important performance criterion is access time/latency/IOPS, versus raw bandwidth. For I/O bandwidth, one may use 8Gbit Fibre Channel cards and 10 Gbit Ethernet cards. Special care needs to be taken to insure that latency for each individual Virtual Machine is low enough to yield a low response time from the overall virtualized server.
- When setting up external storage, configure storage for higher IOPS, rather than with I/O bandwidth.
- Segregate storage LUNs appropriately between the VMs to maintain IOPS/latency.
- Tuning the BIOS options for the server platform and network switch and tuning the parameters for the VMware® ESX server, play a very important role in performance of the virtual machines.
- Certain VMware® Advance Setting switches may help to maintain performance of the system when under heavy load. Also, disabling some memory scanning and page reallocation features of ESX, may help to enhance performance, as these tasks on larger memory systems can use up more CPU time. Refer to SGI® VMmark publication for more details on tuning.

6. Summary

As depicted in this paper, SGI® Rackable™ C2104-TY3 and C2005-TY15 servers maintain both high performance and consolidation capacity of more than 100 virtual machines on a single node. This enables adding more of these server hosts into a vCenter cloud to get the best consolidation capacity and performance level in a Data Center.

The recommendations for best practices mentioned in this paper are based on results derived from tests on the above SGI® Rackable™ servers. These are subject to variations with motherboard, BIOS, processor SKUs, and memory options supported for other Intel® Xeon® 5500 processor-based and Intel® Xeon® 5600 processor-based SGI® servers. Additionally, performance may vary using different storage subsystem or network hardware/interface.

Virtualized SGI® systems, small or large, are best suitable for Defense, Federal, Internet/Cloud, Financial Trading, Business Intelligence and Commercial markets, where efficient consolidation of large data centers with minimal performance overhead is the primary goal.

Sanhita Sarkar
Performance Engineering
SGI
46555 Landing Parkway
Fremont, CA 94538
sanhita@sgi.com