vmware®

VMware ESX Server 2

Storage Subsystem Performance in VMware ESX Server: BusLogic Versus LSI Logic

Introduction

The storage subsystem is a critical determinant of system performance. The key to good storage performance is to identify factors and system configuration settings that affect performance and understand how to set these in order to achieve the best results.

The basic determinants of performance are the operating system, the data transfer size, and the access pattern. In the virtual machine environment, the drivers for the available virtual adapters are also a factor. VMware ESX Server virtual machines can use virtual BusLogic and virtual LSI Logic SCSI adapters. The default driver for a virtual machine depends on the guest operating system. For example, Windows 2000 guests use the Microsoft-supplied BusLogic adapter by default, while Windows Server 2003 guests use the LSI Logic adapter by default.

This document provides a characterization of storage performance for a VMware ESX Server system with an EMC CX500 SAN as the storage back end. The goal is to provide performance data and system resource utilization at various load levels. Throughput, I/O rate, and response time for various data sizes and access patterns provide sizing guidelines. This baseline data is expected to help debug performance problems and facilitate server consolidation for I/O-intensive workloads.

Executive Summary

The main conclusions that can be drawn from these experiments are:

- Both drivers are equivalent from a performance perspective for a large set of workloads.
- The current version of the BusLogic driver needs additional configuration to reach the performance levels of the LSI Logic driver.

Test Environment

This set of experiments characterizes the storage performance of a uniprocessor virtual machine with Windows 2003 Enterprise Edition as the guest operating system. This virtual machine ran on an ESX Server system and storage was provided by an EMC CX500 SAN. I/O load was generated by Iometer, which is a widely-used tool for evaluating storage subsystems (see Reference 1).

The experimental setup is described in the following subsections, followed by a description of the test cases and relevant Iometer parameter settings.

Hardware Layout

The test machine and SAN storage array were set up as shown in Figure 1. An HP DL580 was the test server. The test server and the storage array connect via a Fibre Channel switch. Detailed specifications of all components appear in Test Environment on page 13.

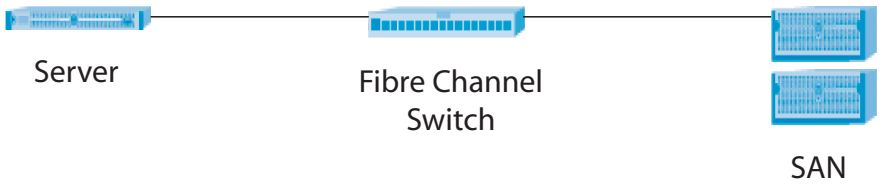

Figure 1 Test Setup

Storage Layout

In the context of this benchmark, storage layout refers to the location and type of the disk used in the tests. Tests were conducted against a 4GB virtual disk located on a five-disk RAID5 LUN. Virtual disks are implemented as files on the underlying storage. From the perspective of the virtual machine, the disk was a physical drive.

Software Configuration

Unless stated otherwise, all ESX Server and guest operating system parameters were left at their default settings.

The BusLogic driver, listed as the VMware SCSI controller driver, can be obtained in the form of a virtual floppy disk (`.flp`) from VMware (see Reference 2) and is also included in the VMware Tools package. See Reference 3 for more details about using this driver with Windows operating systems. LSI Logic drivers for use with Windows 2000 can be downloaded from the LSI Logic Web site. You can find detailed instructions on installing SCSI drivers in virtual machines in VMware knowledge base article 1275 (see Reference 4) and the "Installing Guest Operating Systems" page for the relevant operating system (see Reference 5). The driver can be changed at any point in the life of a virtual machine by following the instructions in the *ESX Server 2.5 Administration Guide* (see Reference 6).

I/O Workload Characteristics

Servers typically run a mix of workloads consisting of different access patterns and I/O data sizes. Within a workload there may be several data transfer sizes and more than one access pattern.

There are a few applications in which access is either purely sequential or purely random. For example, database logs are written sequentially. Reading this data back during database

recovery is done by means of a sequential read operation. Typically, online transaction processing (OLTP) database access is predominantly random in nature.

The size of the data transfer depends on the application and is often a range rather than a single value. For Microsoft® Exchange, the I/O size is generally small (from 4KB to 16KB), Microsoft SQL Server™ database random read and write accesses are 8KB, Oracle accesses are typically 8KB, and Lotus Domino uses 4KB. On the Windows™ platform, the I/O transfer size of an application can be determined using Perfmon.

In summary, I/O characteristics of a workload are defined in terms of the ratio of read operations to write operations, the ratio of sequential accesses to random accesses, and the data transfer size. Often, a range of data transfer sizes may be specified instead of a single value.

Test Cases

The primary objective was to characterize the performance of both drivers for a range of data sizes across a variety of access patterns. The data sizes selected were 4KB, 8KB, 16KB, 32KB, and 64KB. The access patterns were restricted to a combination of 100 percent read or write and 100 percent random or sequential. Each of these four workloads was tested for five data sizes, for a total of 20 data points per workload. The matrix of test cases appears in Table 1.

Table 1 Test Cases

	100 Percent Sequential	100 Percent Random
100% Read	4KB, 8KB, 16KB, 32KB, 64KB	4KB, 8KB, 16KB, 32KB, 64KB
100% Write	4KB, 8KB, 16KB, 32KB, 64KB	4KB, 8KB, 16KB, 32KB, 64KB

All tests were conducted with both BusLogic and LSI Logic drivers. For the best performance, the current implementation of the BusLogic driver requires a workaround which is documented in VMware knowledge base article 1890 (see Reference 7).

Load Generation

The Iometer benchmarking tool, originally developed at Intel and widely used in I/O subsystem performance testing, was used to generate I/O load for these experiments (see Reference 1). A well-designed set of configuration options allow a wide variety of workloads to be emulated and executed. Since this investigation was intended to characterize the relative performance of the two available drivers, only the basic load emulation features were used in these tests.

Iometer configuration options used as variables in these experiments:

- Transfer request sizes: 4KB, 8KB, 16KB, 32KB, and 64KB
- Percent random or sequential distribution: for each transfer request size, 0 percent and 100 percent random accesses were selected
- Percent read or write distribution: for each transfer request size, 0 percent and 100 percent read accesses were selected

Iometer parameters which were held constant for all tests:

- Number of outstanding I/O operations: 16
- Runtime: 3 minutes
- Ramp-up time: 60 seconds
- Number of workers to spawn automatically: 1

Performance Results

This section presents data and analysis for storage subsystem performance in a uniprocessor virtual machine.

Metrics

The primary metric is the throughput rate. The discussion below reports the throughput rate as measured by Iometer. Iometer data from within the virtual machine matches well with the throughput rate measured at the ESX Server system.

The I/O rate and response time are also important criteria for measuring the performance of storage subsystems.

This section also describes the I/O rate and throughput rate achieved per unit of CPU utilization. These derived metrics should be helpful in sizing exercises.

CPU usage data measured within virtual machines is not useful for two reasons. First, it does not always accurately reflect the overhead of virtualization that is incurred by the ESX Server system. Second, the usage data itself may be inaccurate. This is because of the way time is kept within virtual machines (see the section titled "Time Measurements Within a Virtual Machine" in Reference 8 for details). Therefore the CPU utilization reported in this paper is measured at the ESX Server machine level and is the sum of the utilization of all four physical processors. It includes the CPU cost of servicing I/O requests from the virtual machine as well as the CPU overhead of virtualization.

Driver Performance

Performance characteristics of both drivers are summarized in this section. The metrics used are throughput, I/O rate, and average response time.

Figures 2 and 3 show the throughput levels achieved at various data sizes for the LSI Logic and BusLogic drivers, respectively. Both drivers have similar characteristics. The 4KB data point for the sequential-read workload shows a lower throughput than sequential-write for the same data size. This anomaly has also been observed on a physical machine, where the throughput differs by approximately the same amount for the same data points. Similarly, the throughput loss for the 64KB sequential-write data point is also observed on a physical machine.

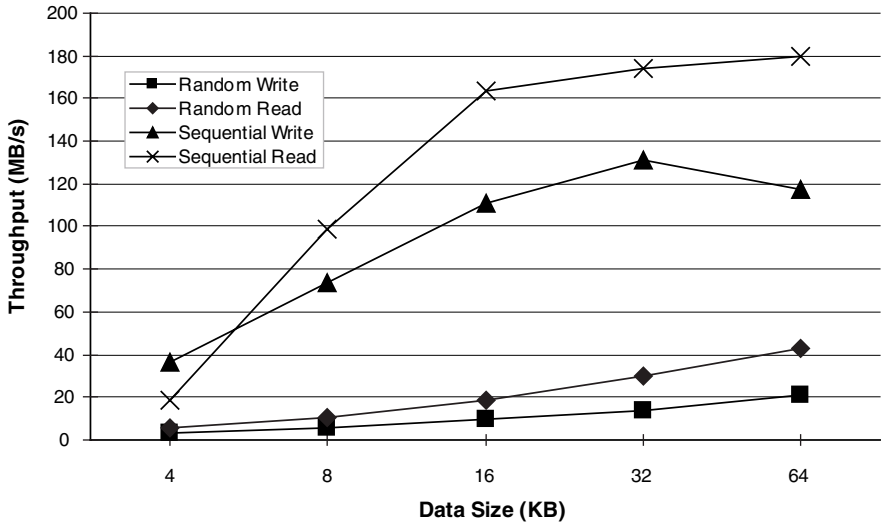

Figure 2 LSI Logic: Throughput

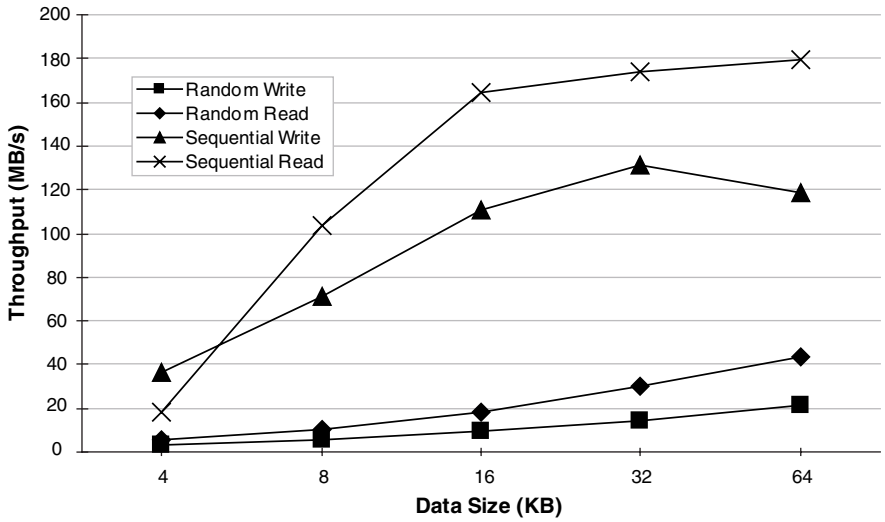

Figure 3 BusLogic: Throughput

The I/O rates corresponding to the above throughput rates are shown in Tables 2 and 3 for the two drivers. The I/O rates mirror the throughput rates and follow predictable trends.

vmware®

Table 2 LSI Logic: I/O Operations per Second

	Workload			
Data Size	**Random Write**	**Random Read**	**Sequential Write**	**Sequential Read**
4KB	817	1424	9377	4759
8KB	725	1360	9412	12618
16KB	597	1182	7122	10471
32KB	445	960	4199	5566
64KB	340	690	1878	2878

Table 3 BusLogic: I/O Operations per Second

	Workload			
Data Size	**Random Write**	**Random Read**	**Sequential Write**	**Sequential Read**
4KB	813	1449	9402	4683
8KB	731	1358	9058	13264
16KB	595	1185	7095	10511
32KB	448	967	4195	5557
64KB	338	690	1896	2877

Figures 4 and 5 show the average response times at various data sizes for the LSI Logic and BusLogic drivers, respectively. With the exception of the sequential-read 4KB data point, all workloads show a gradual increase in response time as data sizes increase. As expected, the random workloads have higher response times. In addition, the *rate* at which the response times increase as data size increases is higher for random workloads than for sequential workloads.

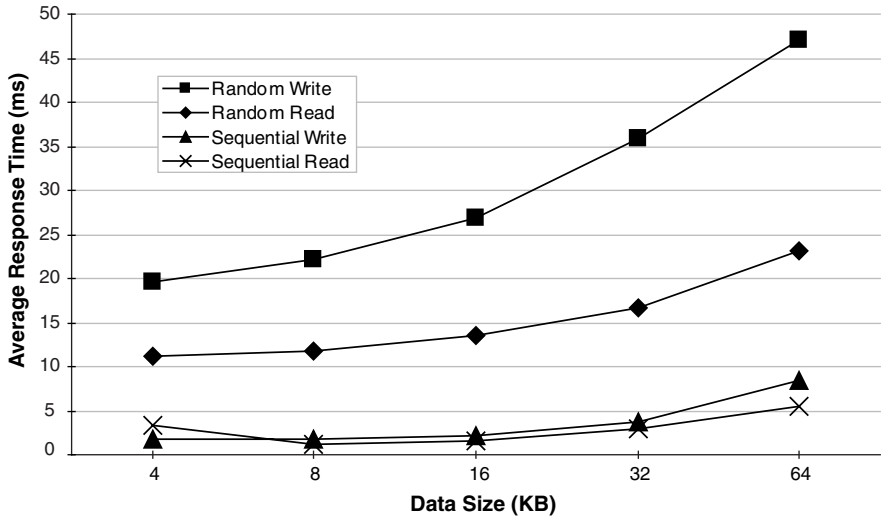

Figure 4 LSI Logic: Response Time

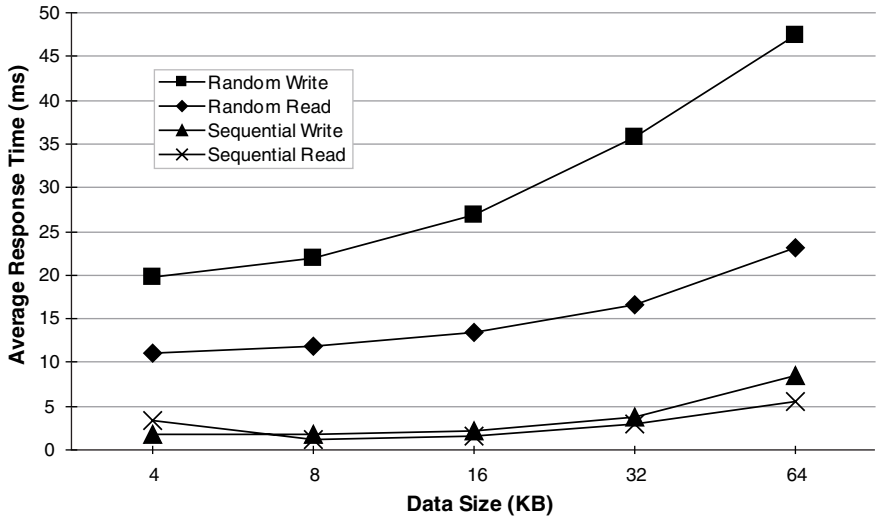

Figure 5 BusLogic: Response Time

Tables 4 and 5 show the percentage CPU utilization measured at the ESX Server machine level for the two drivers. CPU utilization is found to be a function of the I/O rates at the respective load levels. For the sequential workloads, CPU utilization is marginally lower for the BusLogic driver. The outcome is mixed for the random workloads.

Note: Because the CPU utilization data collected at the ESX Server is the sum of the utilization across all four physical CPUs, some of the utilization data in Tables 4 and 5 is higher than 100 percent.

Table 4 LSI Logic: Percent CPU Utilization

	Workload			
Data Size	**Random Write**	**Random Read**	**Sequential Write**	**Sequential Read**
4KB	10.7	20.4	82.6	49.7
8KB	9.8	19.5	82.3	106.3
16KB	9.7	17.7	72.9	107.2
32KB	7.7	16.1	53.2	87.2
64KB	7.2	12.3	29.2	51.5

Table 5 BusLogic: Percent CPU Utilization

	Workload			
Data Size	**Random Write**	**Random Read**	**Sequential Write**	**Sequential Read**
4KB	11.1	20.1	81.7	47.9
8KB	10.9	20.6	80.2	107.8
16KB	9.8	18.4	72.4	106.2
32KB	8.3	16.4	51.6	81.7
64KB	6.6	11.9	27.9	50.0

CPU Efficiency

I/O efficiency can be expressed in terms of either the I/O rate per percent CPU utilization or the throughput rate per percent CPU utilization, depending on which metric is more relevant to the application running in the virtual machine.

I/O rates and throughput rates measured at the ESX Server machine level are highly relevant from a capacity-planning perspective. This section presents the cost per I/O request and the cost per KB/s in terms of percent CPU utilization measured at the ESX Server machine level.

For many workloads, the I/O rate efficiency is between 60 and 70 I/O operations per second per percent CPU utilization. The 64KB data transfer size data point for the random-write workload is slightly less efficient, with approximately 50 I/O operations per second per percent CPU utilization. The system processes sequential workloads with data sizes of 4KB, 8KB, and 16KB most efficiently. These workloads are relevant to a large class of applications based on Microsoft SQL Server and Exchange Server. For these cases, the achievable I/O rate efficiency is 100-110 I/O operations per second per percent CPU utilization. Both drivers are comparable in this regard. The I/O rate efficiency is shown in Figures 6 and 7.

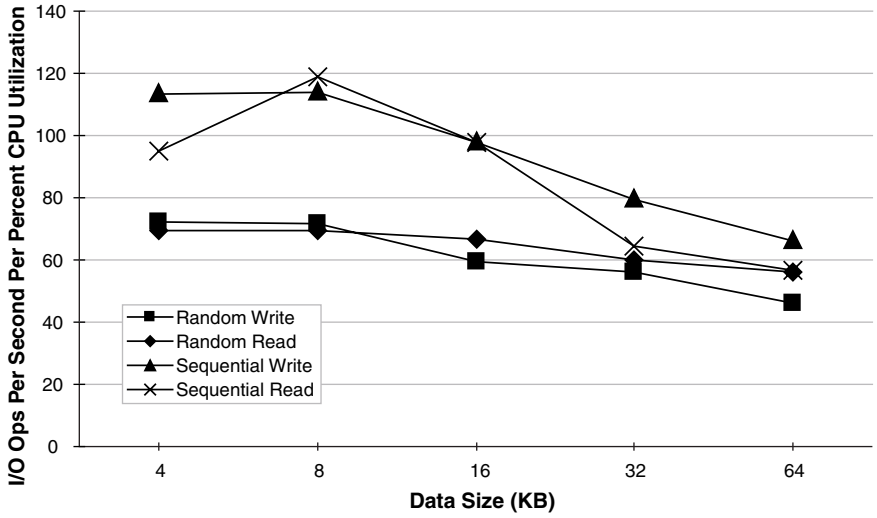

Figure 6 LSI Logic: I/O Rate Efficiency

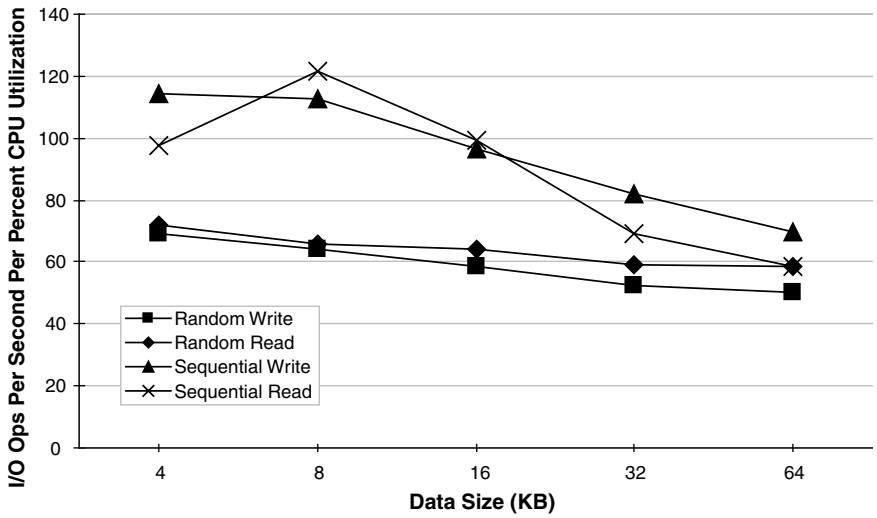

Figure 7 BusLogic: I/O Rate Efficiency

The CPU cost of data transfer is inversely proportional to the data transfer size. This relationship can be seen in Figures 8 and 9. Both drivers have equivalent throughput transfer efficiency.

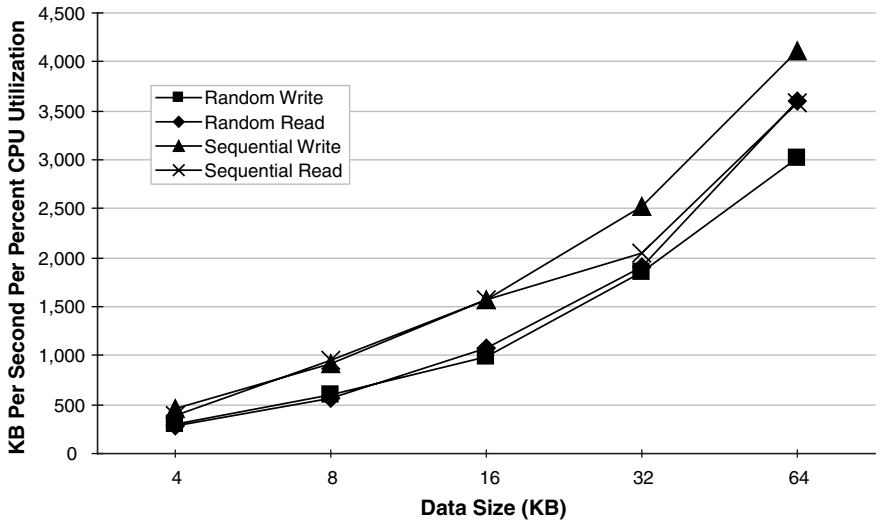

Figure 8 LSI Logic: Throughput Rate Efficiency

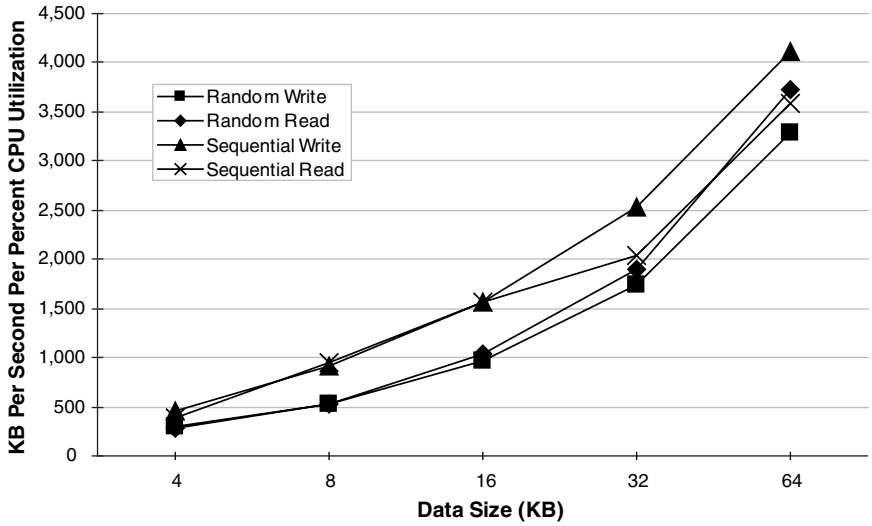

Figure 9 BusLogic: Throughput Rate Efficiency

Conclusion

The relative performance of random and sequential workloads, both read and write, follow a predictable pattern. Throughput levels for sequential workloads are substantially higher than random access workloads. Similarly, sequential workloads exhibit lower average response times. Overall, both drivers are equivalent from a performance perspective.

Resources

1. Obtain the Iometer Load Generator at
   http://sourceforge.net/projects/iometer
2. You can obtain the BusLogic driver (listed as the VMware SCSI Controller driver) in the form of a virtual floppy disk (`.flp`) from VMware's download site,
   http://www.vmware.com/download/esx/#drivers
3. See "Using the VMware SCSI Disk Driver for Windows Guest Operating Systems," at
   http://www.vmware.com/support/reference/common/guest_win_scsidrv.html
4. See VMware knowledge base article 1275, "Using VMware SCSI Driver in Windows 2000 Guest Operating Systems," at
   http://www.vmware.com/support/kb/enduser/std_adp.php?p_faqid=1275
5. See the "Installing Guest Operating Systems" page for the relevant operating system at
   http://www.vmware.com/support/guestnotes/doc/index.html
6. See the *VMware ESX Server 2.5 Administration Guide* at
   http://www.vmware.com/pdf/esx25_admin.pdf
7. See VMware knowledge base article 1890, "Limited Disk Throughput from Windows with BusLogic Adapter Causes Performance Problems," at
   http://www.vmware.com/support/kb/enduser/std_adp.php?p_faqid=1890
8. See "Timekeeping in VMware Virtual Machines" at
   http://www.vmware.com/pdf/vmware_timekeeping.pdf

vmware®

Test Environment

This section details the hardware and software environment in which the tests described in this paper were run.

Server

Server Hardware

HP ProLiant DL 580
Processor (four-way):
- x86 Family 15 Model 2 Stepping 6 GenuineIntel 2.2GHz
- L2 Cache 512KB

Memory: 16GB
Internal storage: Compaq Computer Corporation Smart Array 5i/532 (rev 01)
Network interface cards (four): Intel Corporation 8254NXX Gigabit Ethernet Controller (rev 01)
HBA: QLogic Corp QLA231x/2340 (rev 02)

Server Software

VMware software: ESX Server 2.5.0, Build 11343
Virtual machine configurations:
- CPU: UP
- Memory: 3.6GB
- Guest operating system: Windows 2003 Enterprise Edition
- Connectivity: vmxnet
- Virtual disk: 4GB five-disk, RAID5 LUN presented as a physical drive

Storage

SAN Storage:

EMC CLARiiON CX500 SAN; DPE and DAE with 15 disks each
SPA: cx500-03a
SPB: cx500-03b
Total (30) 146GB 10K RPM disks
Flare operating system revision: 02.07.500.010

Fibre Channel Switch

EMC DS-8B2 (Brocade Silkworm 3200)
Kernel: 5.4
Fabric operating system: v3.1.2a
Link speed: 2Gb/sec.

SCSI Controllers

LSI Logic:
- LSI Logic PCU-X Ultra320 SCSI host adapter
- Driver provider: Microsoft
- Driver date: 10/1/2002
- Driver version: 5.2.3790.0

BusLogic:
- VMware SCSI controller
- Driver provider: VMware, Inc.
- Driver date: 2/13/2004
- Driver version: 1.2.0.2

VMware, Inc., 3145 Porter Drive, Palo Alto, CA 94304
Tel 650-475-5000 Fax 650-475-5001 www.vmware.com
Copyright © 1998-2006 VMware, Inc. All rights reserved. Protected by one or more of U.S. Patent Nos. 6,397,242, 6,496,847, 6,704,925, 6,711,672, 6,725,289, 6,735,601, 6,785,886, 6,789,156 and 6,795,966; patents pending. VMware, the VMware "boxes" logo and design, Virtual SMP and VMotion are registered trademarks or trademarks of VMware, Inc. in the United States and/or other jurisdictions. Microsoft, SQL Server, Windows and Windows NT are registered trademarks or trademarks of Microsoft Corporation in the United States and/or other jurisdictions. Linux is a registered trademark of Linus Torvalds. All other marks and names mentioned herein may be trademarks of their respective companies. Revision: 20060403 Item: ESX-ENG-Q206-201