

iSCSI Performance Factors

Table of Contents

Introduction.....	1
General Testing Parameters.....	1
iSCSI on SLES 10 SP1.....	2
Hardware Configuration.....	2
Test Setup	2
iSCSI Write Performance Test Runs.....	3
Cabling Issues.....	4
iSCSI Read Performance Test Setup.....	4
iSCSI Read Performance Test Runs.....	4
Varying Block Size on Write	5
Sequential Access Versus RandomAccess.....	5
iSCSI on NetWare 6.5 SP7.....	6
Hardware Configuration.....	6
Test Setup	7
iSCSI Read Performance Test Results.....	7
Isolating the LAN Driver Bottleneck.....	8
Sequential Access Versus RandomAccess.....	9
Summary and Conclusions.....	10
Additional iSCSI Information.....	10

Introduction

An iSCSI Initiator and an iSCSI Target are shipped with both NetWare 6.5 and with SUSE Linux Enterprise Server (SLES) 10. Both the Initiator and the Target have been upgraded for SLES 10 SP1. The Initiator has been upgraded for NetWare 6.5 SP7. From time to time, questions arise regarding the variables that affect iSCSI performance. The purpose of this paper is to identify some of the factors that impact iSCSI performance.

As defined later in this paper, we used a specific set of hardware and software to illustrate performance variables. The set up that we used provides performance reference points that help identify factors that influence performance; this configuration is not intended to be used as a formal benchmark.

Some obvious things that affect iSCSI performance include the TCP/IP network topology and throughput, the speed of the processors, and the speed of the target disks. This paper illustrates the significant performance impacts of additional factors including the LAN adapter driver, the cabling, the size of the data being read or written, and how the data is fragmented on the disk.

General Testing Parameters

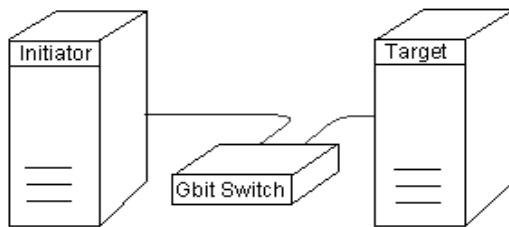
To separate disk performance from iSCSI performance, we used RAM drives on the Target to minimize the mechanical disk overhead from seeks. To separate network performance from iSCSI performance, we isolated the LAN. The test environment was completely standalone. Only TCP packets associated with iSCSI were placed on the wire. No other network traffic was enabled. Using this approach, the speed of the LAN and base iSCSI components could be measured.

iSCSI on SLES 10 SP1

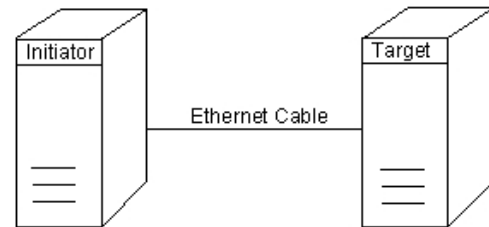
SLES 10 SP1 contains both an iSCSI Initiator and an iSCSI Target. The Initiator can be used with the Target included with SLES, or it can be used with other iSCSI targets. Likewise, the Target in SLES works with both the SLES-provided Initiator and with other initiators. The configuration combinations are numerous. This section of the case study involves the interaction of only the iSCSI Initiator and the iSCSI Target included with SLES 10 SP1.

Hardware Configuration

- 2 Dell Precision 370 machines
 - cpu: 4 - 3.2 Ghz processors
 - nic cards: 1 Gigabit Ethernet – NetXtreme BCM5751 pci card
 - disk: ST380013As – 80 Gbyte / 7200 rpm drive
- 1 Netgear GSM7212 Switch – 1 Gigabit <http://www.novell.com/linux>
- 1 crossover cable
- 2 10' Cat 5 cables



Basic iSCSI Test Configuration
Figure 1



iSCSI Test Configuration using direct cable
Figure 2

Test Setup

We installed a standard distribution of SLES 10 SP1 using default parameters on each Dell Precision machine and created both the Target or the Initiator without changing default configurations.

Preparing the Target

We first created a ram disk as follows:

```
mknod -m 660 /dev/ram b 1 1  
chown root:root /dev/ram
```

We created the targets one at a time using YaST->Miscellaneous->iSCSI Target. Finish and restart the YaST iSCSI Target application after creating each target.

The disk based target used /dev/sdb for its path and the ram based target used /dev/ram for its path.

Verify that the LUNS are available by using the `cat /proc/net/iet/volume` command. The path should also be shown for each target created.

Preparing the Initiator

We used YaST->Miscellaneous->iSCSI Initiator to discover and log into the Targets. Finish and restart the YaST iSCSI Initiator application after discovering and logging into each target.

The `iscsiadm -m session` command can be used to verify that an iSCSI connection is available and that two luns are available from the target. Additionally, the `sfdisk -l` command can be used to display these luns (disk devices).

The target device based on the hard drive appeared on the initiator as `/dev/sdb`.
The target device based on the ram drive appeared on the initiator as `/dev/sdc`

Note: When an Initiator logs in to a Target, they negotiate the size of the communication parameters. The values can be viewed on the Initiator with the following command: `iscsiadm -m session -i`

The results of this command in our test evaluation was:

```
HeaderDigest: None
DataDigest: None
MaxRecvDataSegmentLength 65536
MaxXmitDataSegmentLength 8192
FirstBurstLength 65636
MaxBurstLength 262144
ImmediateData: Yes
InitialR2t : Yes
MaxOutstandingR2T: 1
```

iSCSI Write Performance Test Runs

The purpose of this test is to separate disk performance from the iSCSI protocol performance. In each run, commands were issued on the Initiator and results measured on the Initiator.

Test run #1:

```
dd if=/dev/zero of=/dev/sdb oflag=direct bs=64k count=1000 hard drive
```

Test run #2:

```
dd if=/dev/zero of=/dev/sdc oflag=direct bs=64k count=1000 RAM drive
```

Test run #3: Replace the Netgear Switch with a crossover cable.

```
dd if=/dev/zero of=/dev/sdc oflag=direct bs=64k count=1000 RAM drive
```

Test one (hard drive): 40 MB/sec

Test two (RAM drive): 71 MB/sec

Test three (RAM drive w/crossover cable): 10 MB/sec

The results in Test Run #3 show that something as simple as a cable can kill iSCSI performance.

Cabling Issues

Better performance was expected in test three using the crossover cable. The test was re-run using a non-crossover cable as shown in Figure 2. These particular NIC cards will automatically crossover with straight cables.

		Ramdisk	Hard drive
Cat 5	5ft	73.5 MB/sec	41.0 MB/sec
Cat 6	25ft	73.5 MB/sec	41.0 MB/sec

The results indicate that the crossover cable may be bad or that the NIC cards don't perform well with crossover cables. The results with a straight cable were faster than with the switch connected indicating that cabling issues can have a tremendous effect on performance.

If iSCSI performance isn't what you expect, check your cables.

iSCSI Read Performance Test Setup

We created a mount point for the RAM drive on `/mnt/ramdisk` and for the hard drive on `/mnt/iscsidisk`. Using the YaST partitioner, we created a partition on both the RAM drive and the hard drive with an Ext3 file system and the partitions were mounted on their respective mount points.

Data was written to the RAM drive using:

```
dd if=/dev/zero of=/mnt/iramdisk/data.file bs=4k count=1000
```

Data was written to the hard drive using:

```
dd if=/dev/zero of=/mnt/iscsidisk/data.file bs=4k count=1000
```

We dismounted and remounted the partition before starting the test to remove any cached data:

```
umount /mnt/iscsidisk  
mount -t ext3 /dev/sdc1 /mnt/iscsidisk
```

iSCSI Read Performance Test Runs

Data was read from the RAM drive using:

```
dd if=/mnt/iramdisk/data.file of=/dev/null bs=64k count=1000
```

Data was read from the hard drive using:

```
dd if=/mnt/iscsidisk/data.file of=/dev/null bs=64k count=1000
```

The write tests in the previous section were run with a 64K block size. We ran the read test using the same 64K block size and then repeated the test with a 4K block size. As shown in Table 1, the smaller block size has a significant negative impact on performance.

Block Size	Ram disk (MB/sec)	Hard drive (MB/sec)
64K	74.2	65.6
4K	21.3	21.2

Table 1

Varying Block Size on Write

Since block size has a great effect on performance results, we re-ran write tests at various block sizes with results listed in Table 2.

Block Size	Ram disk (MB/sec)	Hard drive (MB/sec)
4 K	21.5	11.5
8 K	37.0	19.1
16 K	40.4	24.4
32 K	57.1	33.6
64 K	73.0	41.1
128 K	84.7	45.5
256 K	95.0	47.3
512 K	96.0*	51.0*
1 M	107.0*	53.7*
2 M	111.0*	49.9*
4 M	114.0*	48.2*

Table 2

*Note : Count was reduced to 100 because the RAM drive would not hold more data.

Testing indicates that data size has a major effect on performance. This is due primarily to the overhead in writing each chunk of data. Setting the block size at 4 MB and writing to a RAM drive, which eliminates the slowness of the hard drive, results in writes of 114 MB/sec.

Sequential Access Versus Random Access

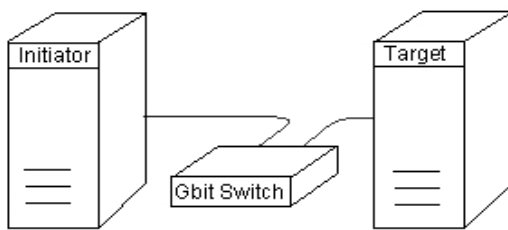
Another significant performance factor is the effect of how the data is laid down on the hard disk. If writes are sequential, which minimize seek times on the hard drive, the data can be written much faster. If access is random, resulting in many seeks across the hard drive, these times can go down dramatically. We have seen access to a local hard drive without iSCSI drop from MB/sec into KB/sec doing random access. Using an appropriate RAID configuration can reduce the effect of the seeks on random access.

iSCSI on NetWare 6.5 SP7

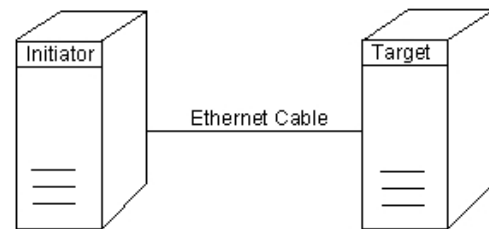
Hardware Configuration

Initial tests on NetWare used the same configuration used for testing on SLES:

- 2 Dell Precision 370 machines
 - cpu: 4 - 3.2 Ghz processors
 - nic cards: 1 Gigabit Ethernet – NetXtreme BCM5751 pci card
 - disk: ST380013As – 80 Gbyte / 7200 rpm drive
- 1 Netgear GSM7212 Switch – 1 Gigabit
- 1 5' Cat 5 cable
- 1 25' Cat 6 cable



Basic iSCSI Test Configuration
Figure 1



iSCSI Test Configuration using direct cable
Figure 2

Additional testing reviewed performance on a second Target machine and a second Initiator machine.

Dell PowerEdge 2900

- cpu: 8 - 3.0 GHz Xenon processors
- nic card: Dell NetXtreme II BCM5708 pci by Broadcom
- disk: Maxtor Atlas 10K V 73GB SAS 10,000 rpm drive

HP ML150

- cpu: 8 - 3.0 GHz Xenon processors
- nic card: HP NC1020 NetXtreme BCM5705_2 Gigabit pci by Broadcom
- nic card 2: Intel PRO 1000 /T
(used in the Dell Precision 370 to obtain the results in the middle column of Table 5)
- disk: Seagate ST3500641AS 500GB SATA 7200 rpm drive
- switch: 1 - Dell PowerConnect 2724 Switch - 1 Gigabit
- cables: 2 - 25" Cat 6

Test Setup

NetWare 6.5 SP7 was installed on each Dell Precision machine. The default parameters were used. No adjustments were made to either the Target or the Initiator configurations.

We used the DPMETERM disk performance test tool on the Initiator machine and configured it to use Device Request Simulation, Read Only, Sequential Single Request, and varied the block size from 8 sectors (4 K) to 1024 sectors (512 K). (One sector equals 512 bytes.)

The iSCSI Control Parameters for NetWare on the Dell Precision 370 machines (extracted from the "_Admin:\Novell\iSCSI\ parameters" file)

```
MaxRecvDataSegmentLength 131072
FirstBurstLength 131072
MaxBurstLength 131072
ImmediateData: yes
InitialR2t : Yes
MaxOutstandingR2T: 1
```

iSCSI Read Performance Test Results

For data reads, we saw the following results:

Block Size	RAM drive (MB/sec)	Hard drive (MB/sec)
4 K	4.6	3.8
8 K	8.6	7.2
16 K	8.0	7.4
32 K	9.0	8.1
64 K	10.0	8.8
128 K	10.2	9.2
256 K	10.2	9.2
512 K	10.2	9.2

Table 3

The performance values shown in Table 3 appear to be quite slow for this system, indicating a possible bottleneck limiting the throughput. As you'll see in the tests that follow, the bottleneck turns out to be the LAN driver.

Isolating the LAN Driver Bottleneck

Direct tests using the DPMETERM disk performance test tool were run on the hard drive and on the RAM drive on the Target system to verify the read performance of the storage subsystem without iSCSI. The same settings were used on this test as were used on the Initiator to verify the capability of the storage system through iSCSI.

As the data in Table 4 illustrates, the slow performance of the iSCSI read tests does not appear to be related to the storage system components on the Target machine.

Block Size	RAM drive (MB/sec)	Hard drive (MB/sec)
4 K	19.844	9.536
8 K	39.888	19.112
16 K	79.520	36.672
32 K	159.328	51.360
64 K	316.848	58.560
128 K	632.192	58.624
256 K	1,265.664	58.624
512 K	2,525.644	58.880

Table 4

We connected the Initiator to a different Target machine and re-ran the tests but the results were almost identical to the results in Table 3 for the old Target machine.

We then added a new LAN adapter to the Initiator and moved the Ethernet cable from the on-board adapter to the new LAN adapter. We re-ran the hard drive tests to the new Target using the new LAN adapter.

Block Size	Alternate Adapter: Hard drive (MB/sec)	Original Adapter: Hard drive (MB/sec)
4 K	7.2	3.8
8 K	10.0	7.2
16 K	15.2	7.4
32 K	20.7	8.1
64 K	25.3	8.8
128 K	27.9	9.2
256 K	28.2	9.2
512 K	28.7	9.2

Table 5

The results for the alternate adapter compared to the results for the original indicate that the LAN driver was limiting the performance in NetWare. If iSCSI performance isn't what you expect, consider trying alternate LAN adapters/drivers.

We ran the tests again, this time using a different machine for the Initiator and the new Target used in the previous test. The new Initiator used a different LAN adapter than the old Initiator.

This test shows that entirely different machines may have different performance curves as illustrated by the 4K run in Table 5, 7.2 MB/sec, vs. Table 6 , 4.1 MB/sec (almost half), and yet the 512K run shows slightly better disk performance in Table 6 than Table 5.

Block Size	RAM drive (MB/sec)	Hard drive (MB/sec)
4 K	5.2	4.1
8 K	9.9	6.5
16 K	16.2	11.7
32 K	30.2	18.0
64 K	38.7	23.5
128 K	38.3	26.0
256 K	34.1	29.0
512 K	41.8	29.2

Table 6

Sequential Access Versus Random Access

Another factor in iSCSI performance is how the data is written or accessed from the hard drive on the Target. If the data is written to the hard drive sequentially, it is much faster than random access, which requires the head to seek to access the data. Table 7 shows the performance differences in tests that were run directly on the hard drive (without iSCSI) using random and sequential access.

Block Size	Random disk access (MB/sec)	Sequential disk access (MB/sec)
4 K	8.5	19.8
8 K	14.7	39.6
16 K	22.8	78.6
32 K	32.9	86.4
64 K	38.9	86.6
128 K	44.2	84.6
256 K	56.8	86.3

Table 7

If iSCSI performance isn't what you expect, use an appropriate RAID configuration to reduce the effect of the seeks on random file access.

Summary and Conclusions

iSCSI lets you create a low-cost Storage Area Network (SAN) using commodity high-speed Ethernet hardware. iSCSI provides significant cost savings when compared to the costs required to create a fibre channel SAN. If you find yourself with iSCSI performance concerns, look first to your network.

Your network configuration and equipment will have the greatest impact on iSCSI performance. Having a high performance switch that does network buffering and flow control, having good cable between the Target and Initiators, and having quality NIC cards is critical to achieving optimal performance.

Additionally, larger data block sizes and sequential data access will greatly improve performance, in many cases doubling rates. Using an appropriate RAID configuration can minimize the effects of the seeks.

The iSCSI software is capable of achieving fast rates-- very near the theoretical limit of 120 MB/second. But so much of the performance depends on the environment in which iSCSI is running.

Additional iSCSI Information

For more information about installing and using iSCSI, see <http://www.novell.com/documentation/oes2/storage.html#iscsi>