WHITE PAPER

Performance Report PRIMERGY SX940 S1

Version 1.0a
March 2010

Pages 11

Abstract

This document contains a summary of the benchmarks executed for the PRIMERGY SX940 S1. In addition to the benchmark results, an explanation has been included for each benchmark and for the benchmark environment.

Contents

Document history

Version 1.0

First report version including the benchmark chapter

- StorageBench
  Measurements with LSI MegaRAID SAS 1064 controller
  Measurements with LSI MegaRAID SAS 1078 controller

Version 1.0a

- StorageBench (corrections, footnote)

Technical Data

PRIMERGY BX900 S1 Blade Servers are highly scalable 19-inch rack systems occupying 10 height units (U) with 18 front bays that accommodate up to six storage blades or a maximum of 18 server blades. In addition two fan units with two fans each, up to six power supply modules, up to two management blades and up to eight connection blades can be integrated in a PRIMERGY BX900 S1.

The PRIMERGY BX920 S1 dual server blades have an Intel 5520 chip set, two Intel Xeon processors (Dual-Core or Quad-Core), nine DIMM slots for up to 72 GB DDR3-SDRAM, two 2-channel GBit LAN controllers, a SAS RAID controller and two bays for 2.5'' SAS hard disks.

With the PRIMERGY SX940 S1 storage blade the internal storage capacity of the PRIMERGY BX920 S1 server blade can be extended by up to 584 GB. It can be assigned for the connection of either up to two disks to two different, adjacent Server Blades, or up to four disks to one adjacent Server Blade. The PRIMERGY SX940 S1 occupies one slot within the PRIMERGY BX900 S1, positioned at the left or right side of an adjacent PRIMERGY BX920 S1. It houses four SAS hard disks driven by one or two SAS RAID controllers.



See http://ts.fujitsu.com/products/standard_servers/blade/bx900/primergy_bx900_storageblades.html for detailed technical information.

StorageBench

Benchmark description

To estimate the capability of disk subsystems Fujitsu Technology Solutions defined a benchmark called StorageBench to compare the different storage systems connected to a system. To do this StorageBench makes use of the Iometer measuring tool developed by Intel combined with a defined set of load profiles that occur in real customer applications and a defined measuring scenario.

Measuring tool

Since the end of 2001 Iometer has been a project at http://SourceForge.net and is ported to various platforms and enhanced by a group of international developers. Iometer consists of a user interface for Windows systems and the so-called “dynamo” which is available for various platforms. For some years now it has been possible to download these two components under “Intel Open Source License” from http://www.iometer.org/ or http://sourceforge.net/projects/iometer.

Iometer gives you the opportunity to reproduce the behavior of real applications as far as accesses to IO subsystems are concerned. For this purpose, you can among other things configure the block sizes to be used, the type of access, such as sequential read or write, random read or write and also combinations of these. You can also configure the number of simultaneous accesses ("Outstanding IOs"). As a result Iometer provides a text file with comma separated values (.csv) containing basic parameters, such as throughput per second, transactions per second and average response time for the respective access pattern. This method permits the efficiency of various subsystems with certain access patterns to be compared. Iometer is in a position to access not only subsystems with a file system, but also so-called raw devices.

With Iometer it is possible to simulate and measure the access patterns of various applications, but the file cache of the operating system remains disregarded and operation is in blocks on a single test file.

Load profile

The manner in which applications access the mass storage system considerably influences the performance of a storage system. Examples of various access patterns of a number of applications:

Application	Access pattern
Database (data transfer)	random, 67% read, 33% write, 8 KB (SQL Server)
Database (log file)	sequential, 100% write, 64 KB blocks
Backup	sequential, 100% read, 64 KB blocks
Restore	sequential, 100% write, 64 KB blocks
Video streaming	sequential, 100% read, blocks ≥ 64 KB
File server	random, 67% read, 33% write, 64 KB blocks
Web server	random, 100% read, 64 KB blocks
Operating system	random, 40% read, 60% write, blocks ≥ 4 KB
File copy	random, 50% read, 50% write, 64 KB blocks

From this four distinctive profiles were derived:

Load profile	Access	Access pattern		Block size	Outstanding IOs	Load tool
		read	write			
Streaming	sequential	100%		64 KB	3	Iometer
Restore	sequential		100%	64 KB	3	Iometer
Database	random	67%	33%	8 KB	3	Iometer
File server	random	67%	33%	64 KB	3	Iometer

All four profiles were generated with Iometer.

Measurement scenario

In order to obtain comparable measurement results it is important to perform all the measurements in identical, reproducible environments. This is why StorageBench is based, in addition to the load profile described above, on the following regulations:

- Since real-life customer configurations work only in exceptional situations with raw devices, performance measurements of internal disks are always conducted on disks containing file systems. NTFS is used for Windows and ext3 for Linux, even if higher performance could possibly be achieved with other file systems or raw devices.
- Hard disks are among the most error-prone components of a computer system. This is why RAID controllers are used in server systems in order to prevent data loss through hard disk failure. Here several hard disks are put together to form a "Redundant Array of Independent Disks", known as RAID in short – with the data being spread over several hard disks in such a way that all the data is retained even if one hard disk fails – except with RAID 0. The most usual methods of organizing hard disks in arrays are the RAID levels RAID 0, RAID 1, RAID 5, RAID 6, RAID 10, RAID 50 and RAID 60. Information about the basics of various RAID arrays is to be found in the paper Performance Report - Modular RAID for PRIMERGY.

  Depending on the number of disks and the installed controller, the possible RAID configurations are used for the StorageBench analyses of the PRIMERGY servers. For systems with two hard disks we use RAID 1 and RAID 0, for three and more hard disks we also use RAID 1E and RAID 5 and, where applicable, further RAID levels – provided that the controller supports these RAID levels.
- Regardless of the size of the hard disk, a measurement file with the size of 8 GB is always used for the measurement.
- In the evaluation of the efficiency of I/O subsystems, processor performance and memory configuration do not play a significant role in today's systems - a possible bottleneck usually affects the hard disks and the RAID controller, and not CPU and memory. Therefore, various configuration alternatives with CPU and memory need not be analyzed under StorageBench.

Measurement results

For each load profile StorageBench provides various key indicators: e.g. "data throughput" in megabytes per second, in short MB/s, "transaction rate" in I/O operations per second, in short IO/s, and "latency time" or also "mean access time" in ms. For sequential load profiles data throughput is the normal indicator, whereas for random load profiles with their small block sizes the transaction rate is normally used. Throughput and transaction rate are directly proportional to each other and can be calculated according to the formula

Data throughput [MB/s]	*= Transaction rate [Disk-I/O s^{-1}] × Block size [MB]*
Transaction rate [Disk-I/O s^{-1}]	*= Data throughput [MB/s] / Block size [MB]*

Benchmark results

The PRIMERGY storage blade SX940 S1 was designed as an additional storage subsystem for the PRIMERGY BX920 S1 server blade. This compact solution enables you to flexibly extend the existing storage capacity, with the optional choice of either connecting 2 × up to two 2.5" SAS drives to two different, adjacent server blades, or to connect up to four 2.5" SAS disk drives to one adjacent server blade. The storage blade can be equipped with one or two controllers from the »Modular RAID« family. The variety of the RAID solutions enables the user to choose the right controller for his application scenario.

The PRIMERGY SX940 S1 has the following RAID solutions to offer:

1. RAID Controller LSI MegaRAID SAS 1064

   The controller is supplied as a PCI Express card. Support is provided for RAID levels 0, 1 and 1E. This controller does not have a cache. The maximum number of hard disks that can be connected to the LSI MegaRAID SAS 1064 controller is four.

2. RAID Controller LSI MegaRAID SAS 1078

   The controller is supplied as a PCI Express card and offers the user a complete RAID solution. Support is provided for RAID levels 0, 1, 5, 6, 10, 50 and 60. This controller is on offer with a 512 MB controller cache. This controller cache can be protected against power failure by an optional battery backup unit (BBU). The controller supports up to 240 hard disks.

Various hard disks can be connected to these controllers. Depending on the performance required, it is possible to select the appropriate disk subsystem.

The following hard disks can be chosen for the PRIMERGY SX940 S1:

- 2½" SAS hard disks with a capacity of 73 GB and 146 GB (10 krpm)
- 2½" SAS hard disks with a capacity of 36 GB and 73 GB (15 krpm)

The 2½" drives have great advantages to offer, because they can reduce both power consumption and heat development and thus cut the costs for device cooling. The better space utilization and consequently the more compact design of the storage blade only became at all possible through the use of 2½" hard disks.

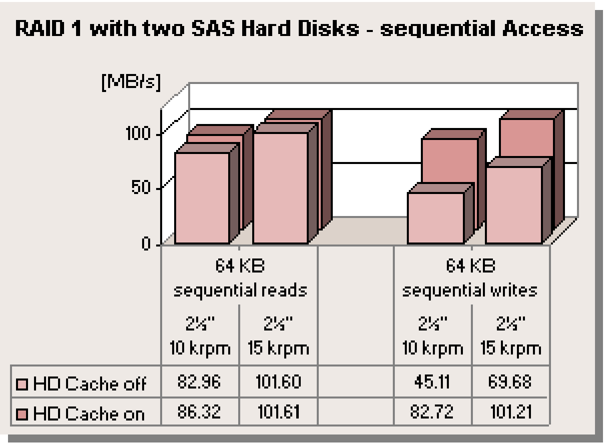
LSI MegaRAID SAS 1064

The performance of the available hard disk types on the LSI MegaRAID SAS 1064 controller is compared below. This controller does not have a controller cache. This is why only the impact of the disk cache parameters was examined in the measurements and the measurements for the hard disk comparison were in each case performed with and without a disk cache.

The hard disk cache has influence on disk I/O performance. This is frequently seen as a safety problem in the event of a power failure and is therefore disabled. Nevertheless, it was for a good reason integrated by the hard disk manufacturers to increase write performance. The by far larger cache for I/O accesses and thus a potential safety risk (data loss) in the event of a power failure is situated in the main memory and is administered by the operating system. To prevent data losses it is advisable to equip the system with an uninterruptible power supply (UPS).

In the test setup two hard disks were connected to the controller and configured as a RAID 1. In the measurements all hard disk types currently available for the PRIMERGY SX940 S1 were analyzed. The throughputs of the individual hard disk types in RAID 1 are compared below with different access patterns.

The diagram shows that as the rotational speed increases, the throughput for sequential reads and writes with a 64 KB block size rises. If for sequential read a hard disk with a rotational speed of 15 krpm is used instead of one with a speed of 10 krpm, the result is an increase in throughput of about 18%.



LSI MegaRAID SAS 1064

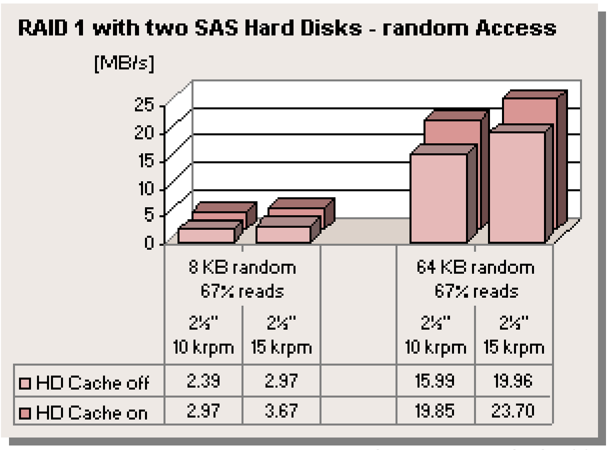
If for sequential write with enabled disk cache a hard disk with a rotational speed of 15 krpm is used instead of one with a speed of 10 krpm, the result for the 2½" hard disk is an increase in throughput of about 22%. If the disk cache is not enabled, the increase is actually 54%.

As can be seen in the diagram above, a relevant increase in throughput with sequential write can be achieved by enabling the disk cache: for hard disks with 10 krpm the throughput increases by about 83% and for hard disks with 15 krpm the throughput increases by about 45%.

The following diagram shows that the disk cache plays an important role in throughput improvement even for random access with 67% read share. For random access with 8 KB blocks with 10 krpm hard disks in RAID 1 the throughput improves through the enabling of the disk cache by about 24%. If 15 krpm hard disks are used instead of 10 krpm ones, the result is an increase in throughput of about 23%.

For random access with 64 KB blocks with 10 krpm hard disks in RAID 1 the throughput improves through the enabling of the disk cache by about 24%. If 15 krpm hard disks are used instead of 10 krpm ones, the result is an increase in throughput of about 19%.

The faster rotating hard disk also performs better with random access, and the rotational speed of 15 krpm makes itself felt with a gain of just above 21%.



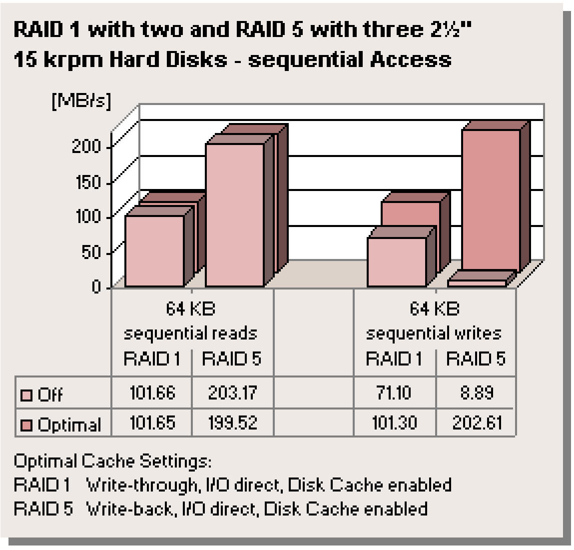
LSI MegaRAID SAS 1064

LSI MegaRAID SAS 1078

The RAID array defines the way in which data is treated as regards availability. How quickly the data is transferred in the respective RAID array context depends largely on the data throughput of the hard disks. The number of hard disks configured for the measurements in a RAID array was defined depending on the RAID level. To ensure that the hard disks do not represent a bottleneck when determining the efficiency of the controller under various cache settings, the measurements were performed with hard disks with a rotational speed of 15 krpm.

The throughput can in certain cases be considerably increased through the cache settings. However, these increases in throughput differ – depending on the data structure and access pattern. For the measurements the controller cache option “Read-Mode” is always set to “No Read-ahead” and the option ”I/O cache” is always set to “I/O direct”. The options “Write-Mode” and “Disk cache” were varied.

The following diagram shows the throughputs for sequential read and write with 64 KB blocks and for different cache settings in RAID 1 with two and in RAID 5 with three 2½" hard disks.



LSI MegaRAID SAS 1078 with 512 MB Cache

The read throughput is in the range of the maximum possible throughput of over 100 MB/s in RAID 1 and 200 MB/s in RAID 5.

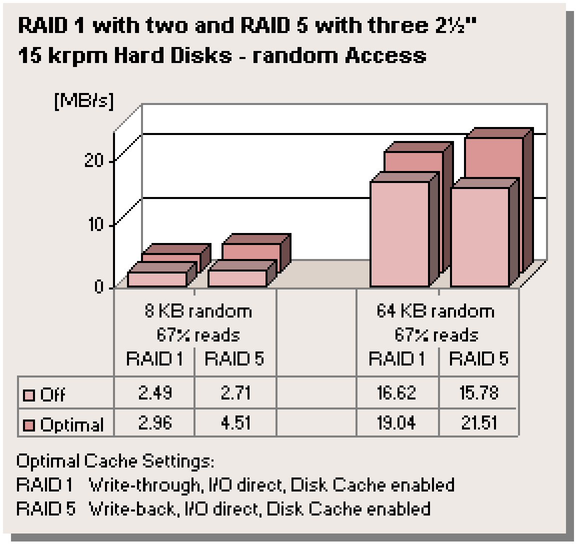
In contrast, the write throughput depends on the cache settings. In order to achieve optimal performance with RAID 1, it is necessary to use the “Disk cache enabled” option as the optimal cache setting. In our case the throughput was improved by a factor 1.4 using sequential write.

The importance of optimal cache settings for a good performance can be seen particularly clearly with RAID 5. The diagram shows that sequential write throughput increases considerably, by a factor of 22, as a result of enabling the controller cache with the option “Write-back” and the disk cache with the option "enabled".

To achieve optimal throughput for random access with RAID 1 it is important to set the write mode option of the controller cache to "Write-through" and to enable the disk cache of the hard disk. As a result of these optimal cache settings, improvements in throughput of 19% and 15% are achieved depending on whether blocks of 8 KB or 64 KB are used for random access.

To achieve optimal throughput for random access with RAID 5 it is important to set the write mode option of the controller cache to "Write-back" and to enable the disk cache of the hard disk. Due to these optimal cache settings, improvements in throughput of 66% and 36% are achieved depending on block size.

More detailed information about this topic is available in the paper Performance Report - Modular RAID for PRIMERGY.

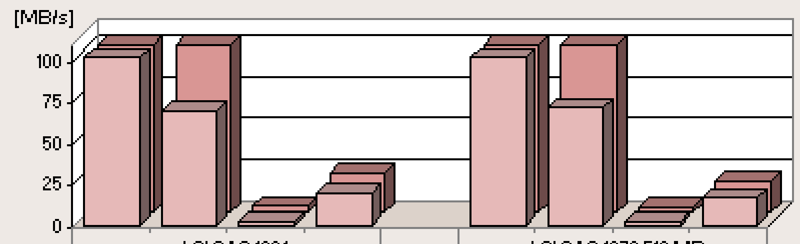


LSI MegaRAID SAS 1078 with 512 MB Cache

Controller comparison

The following comparison depicts the throughputs of the various controllers. The measurements were made with the same hard disk types in the same RAID 1 array. The diagram shows the throughputs achieved with disabled caches (Off) and with optimal cache settings (Optimal).

Controller LSI SAS 1064 and LSI SAS 1078 with 512 MB Controller Cache
RAID 1 with two 2½" 15 krpm Hard Disks



	LSI SAS 1064 a	LSI SAS 1064 b	LSI SAS 1064 c	LSI SAS 1064 d		LSI SAS 1078 512 MB a	LSI SAS 1078 512 MB b	LSI SAS 1078 512 MB c	LSI SAS 1078 512 MB d
Off	101.60	69.68	2.97	19.96		101.66	71.10	2.49	16.62
Optimal	101.61	101.21	3.67	23.70		101.65	101.30	2.96	19.04

Access Pattern:
a Streaming sequential, 64 KB, 100% read
b Restore sequential, 64 KB, 100% write
c Database random, 8 KB, 67% read, 33% write
d File Server random, 64 KB, 67% read, 33% write

Optimal Cache Settings for LSI SAS 1078:
Write-through, IO direct, Disk Cache enabled

Optimal Cache Settings for LSI SAS 1064:
Disk Cache enabled

The performance differences between the used controllers are minimal when looking at purely sequential access. With sequential read all the controllers achieve maximum throughput regardless of the cache settings. All the controllers are also within the same performance range for sequential write and it is possible to increase data throughput by up to 45% through optimal cache settings.

For random access in RAID 1 the entry-level LSI MegaRAID SAS 1064 controller shows a somewhat higher data throughput for this load profile than the LSI MegaRAID SAS 1078 controller, which with its controller cache and extended functionality is optimally equipped for the higher RAID levels and also performs well in RAID 1.

Conclusion

With the “Modular RAID” concept, the PRIMERGY SX940 S1 offers a plethora of opportunities to meet the various requirements of different application scenarios.

The entry-level controller, represented by the LSI MegaRAID SAS 1064 controller, offers the basic RAID solutions RAID 0, RAID 1 and RAID 1E and supports these RAID levels with a very good performance.

The “high-end” controller, represented by the LSI MegaRAID SAS 1078 controller, offers all today’s current RAID solutions; for the PRIMERGY SX940 S1, which can be expanded with up to four internal hard disks, this can be RAID levels 0, 1, 5, 6 and 10. This controller is supplied with a 512 MB controller cache and can as an optional extra be secured with a BBU. Various options for setting the use of the cache enable controller performance to be flexibly adapted to suit the RAID levels used.

Use of RAID 5 or RAID 6 enables the existing hard disk capacity to be utilized economically for a good performance. However, we recommend a RAID 10 for optimal performance and security.

The PRIMERGY SX940 S1 offers 2½ SAS hard disks with rotational speeds of 10 krpm or 15 krpm. Depending on the performance required, a decision must be taken as to which rotational speed is to be used. Hard disks with 15 krpm offer an up to 54% better performance.

For maximum performance it is advisable, particularly when using a controller without a controller cache to enable the hard disk cache. Depending on the disk type used and access pattern, this leads to a performance increase of up to 83%. When the hard disk cache is enabled we recommend the use of a UPS.

Benchmark environment*

All the measurements presented here were performed with the hardware and software components listed below.

Component	Details
Storage Blade	PRIMERGY SX940 S1
Server	PRIMERGY BX920 S1
Operating system	Windows Server 2008, Enterprise Edition Version: 6.0.6001 Service Pack 1 Build 6001
File system	NTFS
Measuring tool	Iometer 27.07.2006
Measurement data	Measurement file of 8 GB
Controller LSI MegaRAID SAS 1064	Product: LSI RAID 0/1 SAS 1064 Driver Name: lsi_sas.sys, Driver Version: 1.29.03.00 Firmware version: 1.27.00.00 BIOS version: 06.26.00.00
Controller LSI MegaRAID SAS 1078	Product: LSI RAID 5/6 SAS 1078 Driver name: megasys.sys, Driver version: 3.9.0.64 Firmware package: 11.0.1-0008 Firmware version: 1.40.32-0580 BIOS version: 2.06.00 Controller cache: 512 MB
Hard Disk SAS, 2½", 10 krpm	Seagate ST973402SS, 73 GB
Hard Disk SAS, 2½", 15 krpm	Seagate ST973451SS, 73 GB

* Some components may not be available in all countries / sales regions.

Literature

PRIMERGY Systems

http://ts.fujitsu.com/primergy

PRIMERGY BX920 S1

http://ts.fujitsu.com/products/standard_servers/blade/bx900/primergy_bx900_serverblades.html

Performance Report – PRIMERGY BX920 S1
http://docs.ts.fujitsu.com/dl.aspx?id=9e5c11e1-9462-4974-a6ff-69d915644867

PRIMERGY SX940 S1

http://ts.fujitsu.com/products/standard_servers/blade/bx900/primergy_bx900_storageblades.html

PRIMERGY Performance

http://ts.fujitsu.com/products/standard_servers/primergy_bov.html

StorageBench

Performance Report - Modular RAID for PRIMERGY
http://docs.ts.fujitsu.com/dl.aspx?id=8f6d5779-2405-4cdd-8268-1f948ba050e6

iometer
http://www.iometer.org

Contact

PRIMERGY Hardware

PRIMERGY Product Marketing
mailto:Primergy-PM@ts.fujitsu.com

PRIMERGY Performance and Benchmarks

mailto:primergy.benchmark@ts.fujitsu.com

Published by department:

Enterprise Products
PRIMERGY Server
PRIMERGY Performance Lab
mailto:primergy.benchmark@ts.fujitsu.com

Internet:
http://www.ts.fujitsu.com/primergy

Extranet:
http://partners.ts.fujitsu.com/com/products/servers/primergy