THE POSSIBILITIES ARE INFINITE FUJITSU

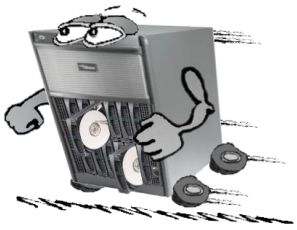
WHITE PAPER

**Version** 1.1
March 2008

# Performance Report
## Modular RAID for PRIMERGY

**Pages** 15


### Abstract



This technical documentation is designed for persons, who deal with the selection of RAID technologies and RAID controllers for storage systems connected to PRIMERGY servers. This document is intended to help find the right RAID configuration for a planned solution at an early stage. Various RAID levels are discussed in connection with “Modular RAID” controllers in a PRIMERGY server system. The performance data of the different “Modular RAID” controllers in the respective RAID array is presented and the impact of the cache settings determined with regard to data throughputs in the environment of the most common RAID configurations and various hard disk types. And providing the conclusion is a comparison of the performance data of “Modular RAID” controllers.


### Contents

# Basic principles

The different PRIMERGY servers are available in various configurations and with varying hard disk and RAID controller configurations. In order to reduce the variety of solutions the specific RAID solutions that used to be normal for individual servers are replaced in the PRIMERGY family by a “Modular RAID” concept for all servers. The comprehensive offering of RAID solutions enables the user to choose the right controller for his application scenario.

## Controller

Three controllers are offered as part of the “Modular RAID” concept:

1. RAID Controller LSI MegaRAID SAS 1068

   The controller is supplied as a PCI Express card. The maximum number of SATA resp. SAS hard disks that can be connected to the controller is eight. Support is provided for RAID levels 0, 1 and 1E. This controller does not offer any controller cache settings.

2. RAID Controller LSI MegaRAID SAS 1064

   This controller has the same features and the same performance as the LSI MegaRAID SAS 1068 controller. However, only four hard disks can be connected to this controller. The LSI MegaRAID SAS controller is available as a PCI Express card and also as an onboard version in a number of PRIMERGY models. The values measured in the following document with the LSI MegaRAID SAS 1068 for configurations with up to four hard disks are also valid for the LSI MegaRAID SAS 1064 controller.

3. RAID Controller LSI MegaRAID SAS 1078

   The controller is supplied as a PCI Express card and offers the user a complete RAID solution. Both SATA and SAS hard disks can be connected. Support is provided for RAID levels 0, 1, 5, 6, 10, 50 and 60. Two different versions of this controller are on offer with either a 256 MB or 512 MB cache. The controller cache can be protected against power failure by an optional battery backup unit (BBU). The controller supports up to 240 hard disks. Depending on the PRIMERGY model and it's configuration with disk cabinets, the number of hard disks can be lower.

### Controller cache

In contrast to the LSI MegaRAID SAS 1068 controller, the LSI MegaRAID SAS 1078 controller provides a controller cache in both versions, which as an optional extra can be protected against power failure by a BBU. The controller cache is used to increase write and read performance and can be influenced by three configuration parameters.

***Write Mode***

The term “Write mode” summarizes the setting options of the controller cache. There are three possible write cache settings: “write-through”, “write-back” and “write cache bad BBU”. The option “write-through” ensures that each write order is only acknowledged by the controller when it has really been written onto the hard disk. With the options “write-back” and “write cache bad BBU” the orders are buffered in the controller cache and immediately acknowledged to the user as done, although in reality they do not exist on the hard disk at all, but are only written there later. This procedure allows the controller resources to be optimally utilized, a faster sequence in the write orders and thus a higher throughput. Any power failures can be bridged by an optional BBU so as to ensure data integrity in the controller cache. The “write cache bad BBU” option also enables the write cache when the battery of the BBU is empty or no BBU has been installed, whereas the “write-back” option automatically switches to “write-through” when the controller cache is not battery-buffered.

***Read Mode***

The parameter “Read mode” can be used to influence the cache behavior during read. Three options are available “No read ahead”, “Read ahead” and “Adaptive”. With “No read ahead” no caching takes place during read. In case a data block is requested by the operating system, “Read ahead” causes other sequential data blocks to already be prophylactically read from the hard disk into the controller cache in the hope that the operating system requests these in subsequent orders. With the setting “Adaptive” the controller itself attempts to determine whether a “Read-ahead” is sensible or not.

***Cache Mode***

The parameter “Cache Mode” - only referred to in short in the web BIOS with “I/O Cache” – also influences the read behavior of the controller cache. The “Direct” option determines that the data to be read is read directly from the hard disk and is also not stored in the controller cache. The consequence of the alternative “Cached” is that an attempt is first made to find the data in the controller cache and to satisfy the read order before the hard disk is accessed, with all the data being written into the controller cache in order to be available for subsequent read orders.

© Fujitsu Technology Solutions 2009 

### Disk cache

In most cases, enabling the disk cache entails an increase in throughput for write access. However, besides the increase in performance, enabling the disk cache also has disadvantages. If a fault occurs in the power supply, important data that has not yet been written from the disk cache to the hard disk may possibly be irrevocably lost. This is why it is advisable to ensure a continuous power supply for the hard disks by activating an uninterruptible power supply (UPS) upstream. If the system is UPS-protected, enabling the disk cache for performance reasons is recommended.

## RAID levels

The best throughput can be achieved with a **RAID 0**. As the number of hard disks in an array increases, so does the throughput. The increase in throughput is achieved through the parallel accesses to the hard disks. RAID 0 makes the entire hard disk capacity available to the user (0% overhead).

The disadvantage is that a RAID 0 has no redundancy whatever to offer. If a RAID 0 hard disk fails, all the data is lost. RAID 0 is typically used when data security plays a subordinate role or the data is otherwise backed up.

A **RAID 1** ensures complete data redundancy on two hard disks. In the best case the read throughput is equivalent to the total of the throughput of the two hard disks. The write throughput is equivalent to the throughput of one hard disk in an array. The disadvantage is that only half the overall array capacity is available to the user (50% overhead).

A **RAID 1E** ensures full data redundancy on more than two hard disks. The read throughput is as with a RAID 1. The advantage is greater flexibility in the configuration, unless the controller offers a higher RAID level. The disadvantage, similar to that of RAID 1, is that only half the overall capacity is available to the user (50% overhead).

A **RAID 5** consists of at least three hard disks. The data and additionally calculated parity information are distributed over all existing hard disks. A RAID 5 provides a high degree of data security, but also low throughputs – particularly with write accesses with the result being a capacity overhead of one hard disk in the array, which corresponds to $^{100}/_{\text{number of hard disks in the RAID 5 array}}$ [%].

A **RAID 6** is an enhancement of RAID 5 and ensures that no data is lost if two hard disks fail at the same time. A RAID 6 ensures a high degree of security, but in comparison with a RAID 5 throughput is less with the RAID 6. The capacity overhead is $^{200}/_{\text{number of hard disks in the RAID 6 array}}$ [%].

A **RAID 10** is made up of at least two RAID 1s, which are in turn combined to form a RAID 0 and offers optimal performance with the best possible fail-safety. With a RAID 10 only half the overall capacity is available (capacity overhead 50%).

A **RAID 50** consists of at least two RAID 5s, which are in turn combined to form a RAID 0. Consequently, it is possible to achieve a higher write throughput when compared with a single RAID 5. The capacity overhead is twice as large as with a RAID 5.

The **RAID 60** is a combination of the RAID 6 and RAID 0. The features of the RAID 6 remain unchanged and the fail safety increases in comparison with the RAID 6 and RAID 50. However, throughput is lower than with a RAID 50 and the capacity overhead is twice as large as with the RAID 6.

© Fujitsu Technology Solutions 2009

## Measuring methods

To estimate the capability of disk subsystems Fujitsu Technology Solutions defined a benchmark called StorageBench to compare the different storage systems connected to a server. 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. 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 IO subsystems with a file system, but also subsystems without a file system, 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 systems remains disregarded and operation is in blocks on a single test file.

### Load profile

The manner in which applications access the storage subsystem 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 | Load tool |
|---|---|---|---|---|---|
| | | read | write | | |
| Streaming | sequential | 100% | | 64 kB | Iometer |
| Restore | sequential | | 100% | 64 kB | Iometer |
| Database | random | 67% | 33% | 8 kB | Iometer |
| File-Server | random | 67% | 33% | 64 kB | Iometer |

All four profiles were generated with Iometer.

© Fujitsu Technology Solutions 2009

### 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. The most usual methods of organizing hard disks in arrays are the RAID levels RAID 0, RAID 1, RAID 1E, RAID 5, RAID 6, RAID 10, RAID 50 and RAID 60.
- 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.

### Measuring environment

All the measurements presented in this document were performed with the hardware and software components listed below.

| Component | Details |
|---|---|
| Server | PRIMERGY TX200 S4<br>PRIMERGY RX300 S4 |
| Controller LSI MegaRAID SAS 1068 | Driver Name: lsi_sas.sys, Driver Version: 1.25.05.00, Firmware Version: 01.18.41.00, BIOS Version: 06.12.00.00 |
| Controller LSI MegaRAID SAS 1078 with 256 MB or 512 MB cache | Driver Name: msas2kr.sys, Driver Version: 2.17.0.32, Firmware Package Version: 6.0.1-0081, Firmware Version: 1.11.72-0356, BIOS Version: NT10 |
| Hard disk SATA, 3½", 7.2 krpm | Western Digital WD1600AAJS, 160 GB |
| Hard disk SAS, 2½", 10 krpm | Seagate ST973402SS, 73 GB |
| Hard disk SAS, 2½", 15 krpm | Seagate ST973451SS, 73 GB |
| Hard disk SAS, 3½", 10 krpm | Seagate ST373355SS, 73 GB |
| Hard disk SAS, 3½", 15 krpm | Seagate ST373455SS, 73 GB |
| Operating system | Windows Server 2003, Enterprise Edition, Service Pack 1 |
| File system | NTFS |
| Test tool | Iometer 2006.07.27 |
| Test data | Measurement file of 8 GB |

© Fujitsu Technology Solutions 2009

# Performance analysis

## Disks

In the PRIMERGY models that can be equipped with controllers of the “Modular RAID” family various hard disks are used and support is provided for SATA and SAS hard disks. Depending on the performance required, it is possible to select the appropriate disk subsystem. Below is a comparison of the performance of various hard disk types, regardless of in which PRIMERGY systems they are used. Not all PRIMERGY models with “Modular RAID” controllers support all hard disk types and even the configuration options of the individual models are different. See the data sheets of the PRIMERGY systems for more details.

### SAS disks

As a result of the higher rotational speed of the SAS hard disks (in comparison with the SATA hard disks) the SAS hard disks have shorter access times and achieve higher throughputs. One possible disadvantage of the higher number of revolutions is a higher noise generation and development of heat, which in turn makes additional cooling necessary. 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. Better space utilization should also be mentioned. However, the 2½" hard disks also have disadvantages – such as lower capacity and lower throughput.

### SATA disks

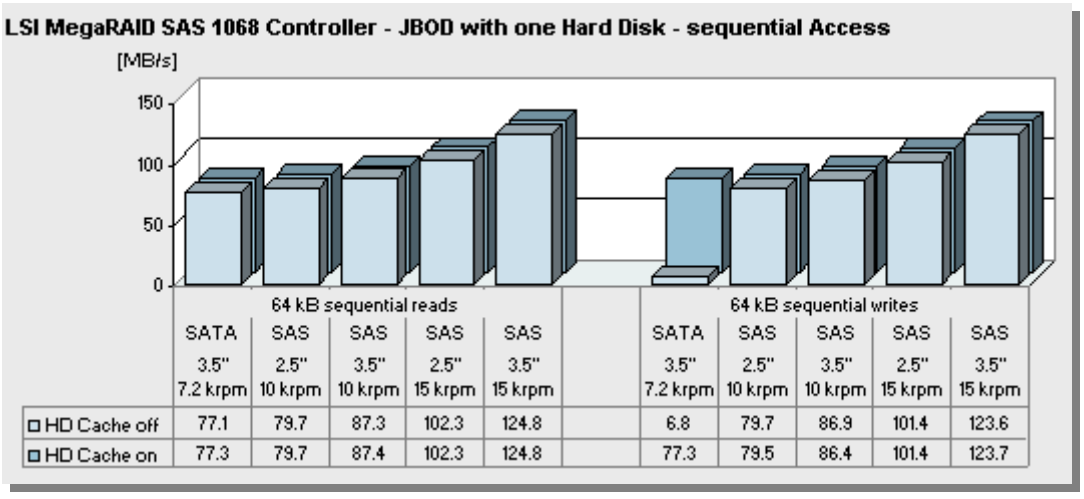
As a result of the lower rotational speed of the SATA hard disks compared with the SAS hard disks, the SATA hard disks have higher access times and achieve lower throughputs. However, the SATA hard disks offer high capacities up to the terabyte range at very favorable costs. They are especially suited for use in secondary storage and backup systems.

### Hard disk comparison

In the test setup only one individual hard disk was connected to the LSI MegaRAID SAS 1068 controller and configured as JBOD. In the measurements all the hard disk types currently available for the PRIMERGY servers were analyzed, i.e. 3½" and 2½" SAS hard disks with a rotational speed of 10 krpm or 15 krpm or 3½" SATA hard disks with a rotational speed of 7.2 krpm. The throughputs of the individual hard disks are compared below with different access patterns.

The hard disk cache has influence on disk I/O performance. Unfortunately, this is frequently seen as a security problem in the event of a power failure and is therefore disabled. On the other hand, it was for a good reason integrated by the hard disk manufacturers to increase write performance. Features, such as Native Command Queuing (NCQ), only function at all when the disk cache is enabled. For performance reasons it is advisable to enable the disk cache for the SATA hard disks in particular, which in comparison with the SAS hard disks rotate slowly. The by far larger cache for I/O accesses and thus a potential security risk for data loss in the event of a power failure is in any case in the main memory and is administered by the operating system. To prevent data losses it is advisable to equip the system with an UPS. The measurements for the hard disk comparison were in each case performed with and without a disk cache.

The diagram shows that as the rotational speed increases, the throughput for sequential reads and writes with a 64 kB block size rises.



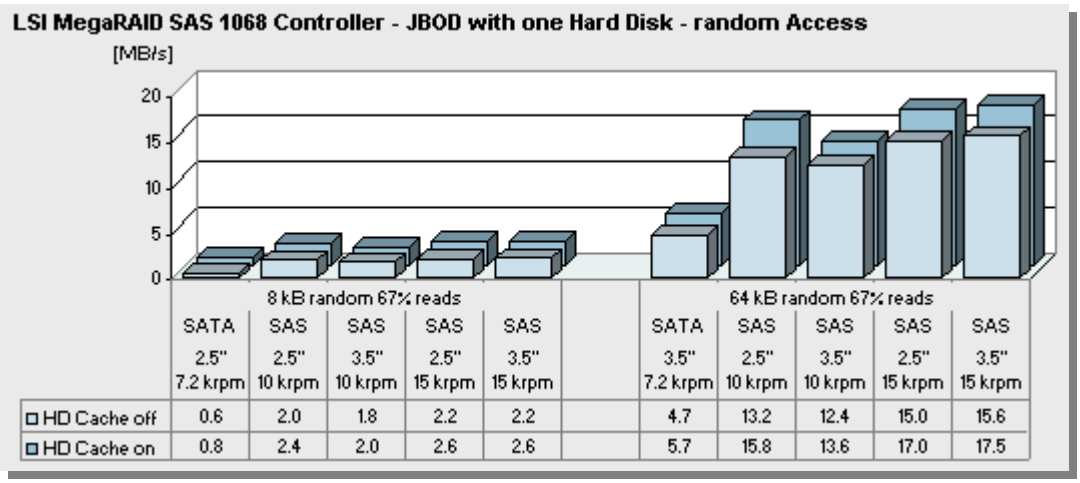

If a hard disk with a rotational speed of 15 krpm is used instead of one with a speed of 10 krpm, the result for the 2½" hard disk is an increase in throughput of about 28% and about 42% for the 3½" hard disk. If you compare the throughputs of the 2½" and 3½" hard disks both with a rotational speed of 10 krpm, you can then see that the throughput for the 3½" hard disk is about 9% higher than for the 2½" hard disk. At a rotational speed of 15 krpm the difference in throughput between the 2½" and 3½" hard disk is even greater and amounts to 22%.

If you compare the 3½" SAS hard disk with the 3½" SATA hard disk, you can then see that the throughput of the SAS hard disk with 10 krpm is about 12% higher than the SATA hard disk with 7.2 krpm for sequential read and write and with

an enabled disk cache. If you compare the 3½" SAS hard disk with 15 krpm with the SATA hard disk, you see that the throughput of the 3½" SAS hard disk with 15 krpm is even 60% higher than with the SATA hard disk.

A special increase in throughput for sequential write, up to eleven-fold, can be achieved with the SATA hard disk by enabling the disk cache. However, enabling the disk cache in the SAS hard disk for sequential read and write does not entail any notable advantages in performance.

For random access with 67% reads you can see that the disk cache in SAS hard disks plays a basically more important role in improving throughput than is the case for sequential read and write. An improvement of up to 20% was determined. An increase in throughput of up to 33% was even achieved in the SATA hard disks. If you compare the hard disks with a rotational speed of 10 krpm with the hard disks with 15 krpm, you can achieve an approximately 8% better throughput with the faster rotating 2½" hard disks and an approximately 30% better performance with the 3½" hard disks with a rotational speed of 15 krpm.




| | SATA 2.5" 7.2 krpm | SAS 2.5" 10 krpm | SAS 3.5" 10 krpm | SAS 2.5" 15 krpm | SAS 3.5" 15 krpm | SATA 3.5" 7.2 krpm | SAS 2.5" 10 krpm | SAS 3.5" 10 krpm | SAS 2.5" 15 krpm | SAS 3.5" 15 krpm |
|---|---|---|---|---|---|---|---|---|---|---|
| HD Cache off | 0.6 | 2.0 | 1.8 | 2.2 | 2.2 | 4.7 | 13.2 | 12.4 | 15.0 | 15.6 |
| HD Cache on | 0.8 | 2.4 | 2.0 | 2.6 | 2.6 | 5.7 | 15.8 | 13.6 | 17.0 | 17.5 |

If you compare the 3½" SAS hard disk with the 3½" SATA hard disk, you can see that the throughput of the SAS hard disk with 10 krpm is about three times higher with random access and enabled disk cache than with the SATA hard disk with 7.2 krpm. If you compare the 3½" SAS hard disk with 15 krpm with the 3½" 7.2 krpm SATA hard disk, you see that the throughput of the 3½" SAS hard disk is 3.35-fold higher than with the SATA hard disk.

© Fujitsu Technology Solutions 2009 

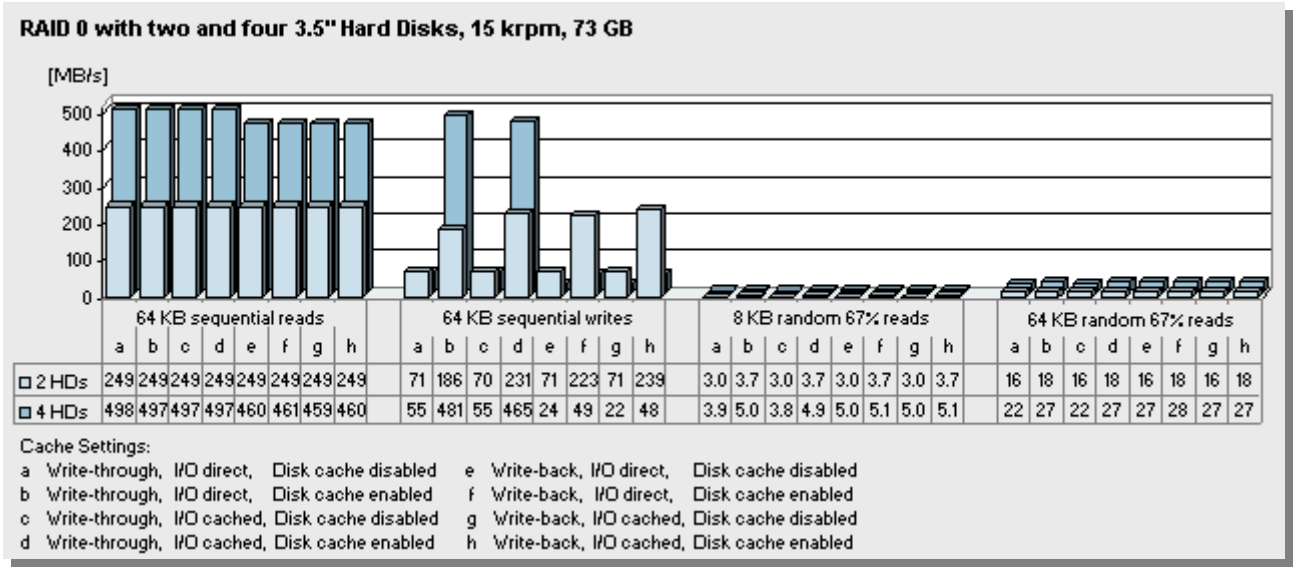
## Controller

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 disk. The number of hard disks configured for the measurements in a RAID array was defined depending on the RAID level. Between two and six hard disks were used. 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, if possible, performed with 3½" hard disks with a rotational speed of 15 krpm. However, the various PRIMERGY models, which can be equipped with a “Modular RAID” controller, can – depending on the system – be equipped with various hard disk models so that the throughputs actually achieved may be lower.

### LSI MegaRAID SAS 1078 Controller

The LSI MegaRAID SAS 1078 controller offers various RAID levels and their performance is analyzed individually below. Throughput can in part be considerably increased through the cache settings, which is why the various cache settings are compared. However, this increase in throughput differs – depending on the data structure and access pattern. All the measurements were performed using the LSI MegaRAID SAS 1078 controller with a 512 MB cache.

The following diagram shows the throughputs in a RAID 0 array with two and four 3½" hard disks for the various load profiles and influences of the cache settings.




*LSI MegaRAID SAS 1078 with 512 MB Cache*

By enabling the cache it is possible to increase the write throughput several times over. In this case enabling the disk cache makes a decisive contribution toward increasing the throughput. However, maximum write throughput is only achieved with the combination of the cache settings “Disk-Cache enabled", "Write-back" and "I/O cached” (see Cache Settings: Setting “h”), which is equivalent to an optimal cache setting for sequential write access. In comparison with write throughput with the cache disabled three-fold throughput is achieved in this way.

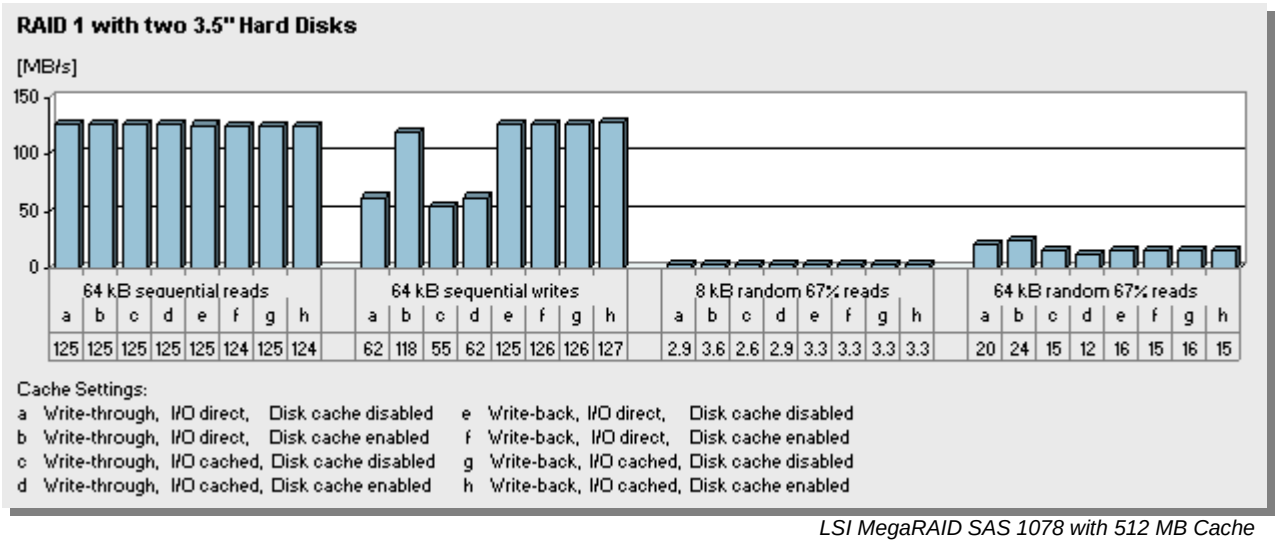
On the other hand, the read throughput in a RAID 0 array with two hard disks cannot be relevantly influenced with cache settings. Read throughput achieves the maximum possible throughput values of about 250 MB/s (2 × 125 MB/s) regardless of the cache setting. The increase in throughput with optimal cache settings is about 15% for random access with 67% read share and 8 kB blocks and about 20% for random access with 67% reads and 64 kB blocks.

In a RAID 0 array with four 3½" hard disks you can see the negative impact of the “Write-back” setting. In this case, the read throughput falls - in comparison with the maximum possible value - by about 8%.

The write throughput achieved with the optimal cache settings is only just below the maximum possible write throughput. In comparison with the throughput achieved with disabled caches throughput has been improved about 8.7-fold.

Enabling the disk cache makes itself largely felt in random access with 67% reads and the increase in throughput amounts to about 25% with the 8 kB blocks and about 35% with the 64 kB blocks. The direct comparison of the RAID 0 arrays with two and four 3½" hard disks shows that the throughput can also be increased, in this case doubled, with an increasing number of hard disks and an optimal cache setting for sequential read/write access.

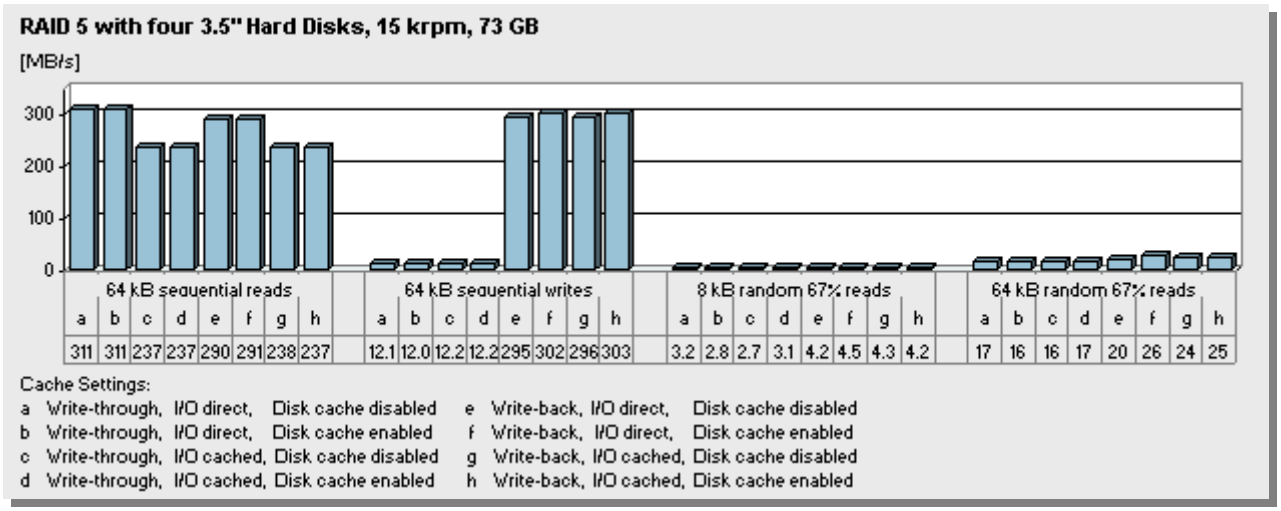
The following diagram shows the throughputs of the 3½" hard disks in the RAID 1 array with two hard disks. The read throughput has reached the maximum possible throughput value of 125 MB/s. The cache settings only have marginal




*LSI MegaRAID SAS 1078 with 512 MB Cache*

influence on the read throughput. On the other hand, write throughput depends very much on the cache settings. To achieve optimal performance you need to use the optimal cache settings "Write-back", "I/O direct" and "Disk-Cache enabled". The improvement achieved in this way is about 60%. The same marked dependence of throughput on the cache settings can also be seen in random access with 64 kB blocks. The optimal cache settings increase throughput by about 50%.

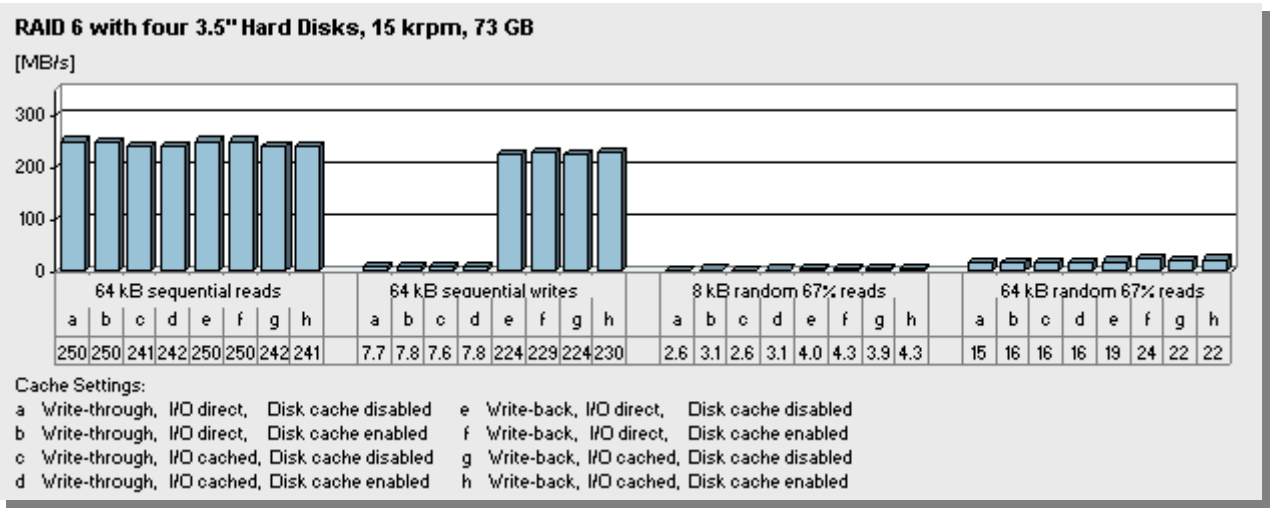
The importance of optimal cache settings can be seen particularly clearly with RAID 5. The diagram shows that sequential write throughput increases considerably as a result of enabling the controller cache with the option "Write-back" and achieves even higher values than with sequential read, although for write accesses an additional parity block has to be calculated and written. On the other hand, the cache settings have less impact on throughput with sequential read. It is interesting to see how counterproductive the effect of enabling the I/O cache is on read throughput particularly for reads.




*LSI MegaRAID SAS 1078 with 512 MB Cache*

A marked dependence of throughput on the cache settings can also be found in random access with 64 kB blocks. The optimal cache settings increase throughput by about 40%.

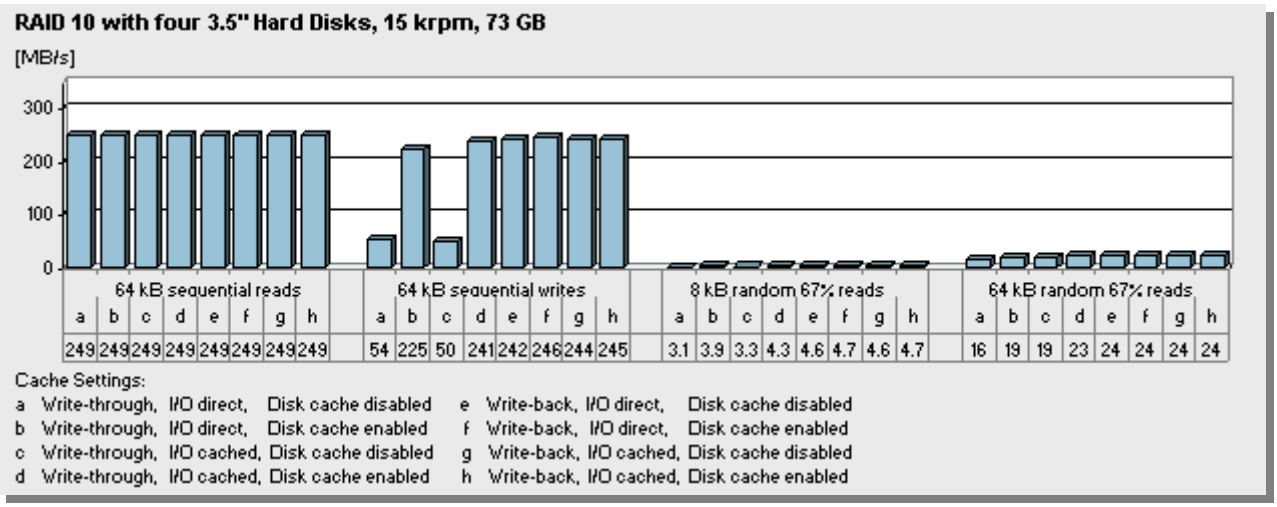
The RAID 6 is used in order to further improve fail-safety in comparison with the RAID 5. This ensures that that no data is lost even if two hard disks in an array fail. The dependence of throughput on the cache settings, as depicted in the diagram, is very similar in the RAID 5 and RAID 6. However, the impact of the enabled I/O cache is not as pronounced




*LSI MegaRAID SAS 1078 with 512 MB Cache*

with the RAID 6 as is the case with the RAID 5. As a result of writing an additional parity block the write throughput with the RAID 6 is somewhat lower than with a RAID 5 array. Whereas the throughput for random access is approximately the same with both arrays, the differences in throughput for sequential read and write accesses are higher. The read throughput with optimal cache settings is about 24% higher with the RAID 5 and write throughput even 40% higher than with the RAID 6.

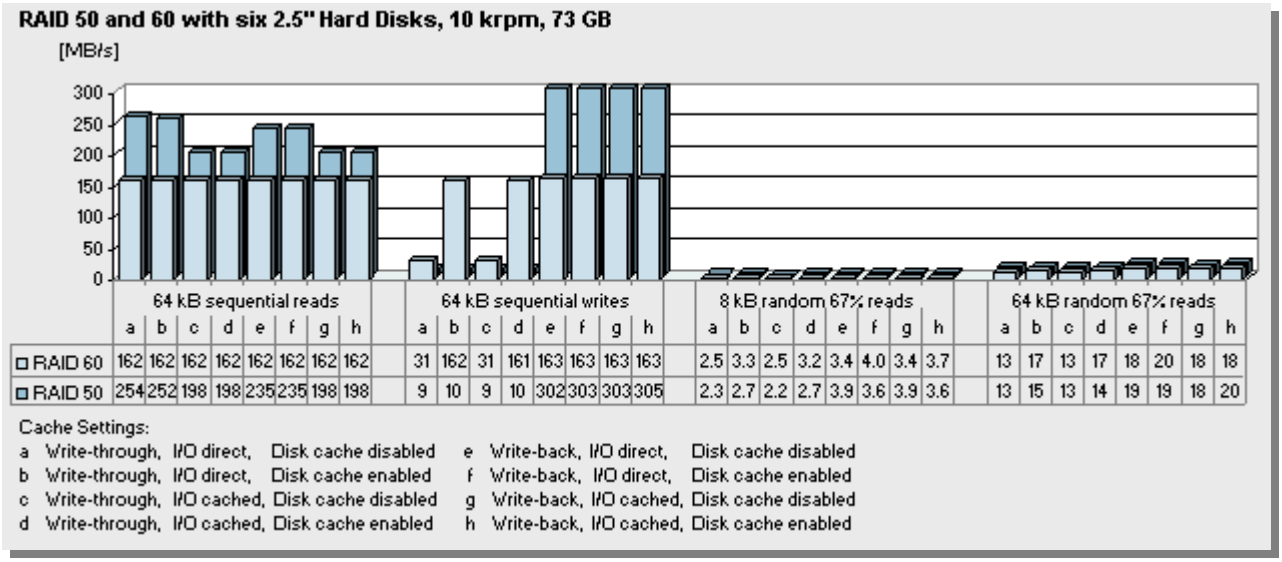
A further solution that ensures high data security and high data throughputs is RAID 10. However, the capacity overhead is 50%. The following diagram compares the throughputs for a RAID 10 with four hard disks.




*LSI MegaRAID SAS 1078 with 512 MB Cache*

The read throughput for sequential read is approximately equivalent to the maximum throughput of the 3½" hard disks. In the event of an optimal cache setting with a “Write-back” option, the same also applies for the throughput for sequential write. On the other hand, for random access with 67% reads the throughputs increase on average by about 20%. At first glance, a RAID 5 and RAID 6 appear to be more efficient in comparison with the RAID 10. However, if you take a closer look at the measurement data, you can see that this only applies for a purely sequential access pattern, which in practice only occurs very seldom in this form. In the important access profiles with mixed write and read access you can see that a RAID 10 provides in comparison with RAID 5 and RAID 6 a throughput that is up to 61% higher.

The following diagram compares the throughputs of a RAID 50 and a RAID 60, using as an example a RAID array consisting of six 2½" hard disks with a rotational speed of 10 krpm. With optimal cache settings the RAID 50 achieves an approximately 55% better throughput for sequential read than the RAID 60. For sequential write, and if all caches are

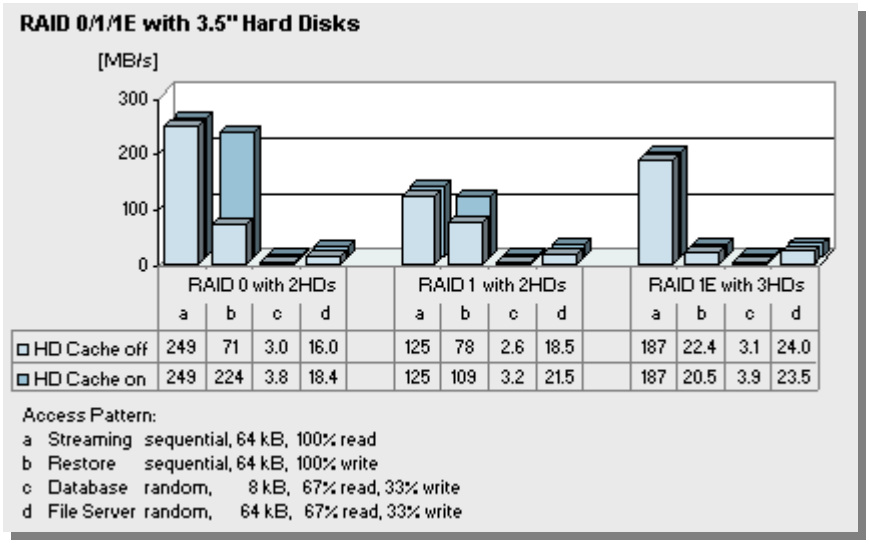



*LSI MegaRAID SAS 1078 with 512 MB Cache*

enabled, throughput is even 88%. For random access with 67% reads the two RAID arrays are on average approximately level. The parity overhead of the RAID 50 is lower than with the RAID 60. The capacity overhead for the RAID 50 in percent is $2 \times {}^{100}/_{\text{number of hard disks in the RAID 50 array}}$ and for the RAID 60 $2 \times {}^{200}/_{\text{number of hard disks in the RAID 60 array}}$. The RAID 60 compensates for the lower throughput and less memory capacity with better data security. Despite the simultaneous failure of up to two hard disks in each RAID 6 subset the data remains unaffected. In a RAID 50 one hard disk at most may fail at the same time in each RAID 50 subset.

## LSI MegaRAID SAS 1064/1068 Controller

The LSI MegaRAID 1064 and 1068 SAS controllers do not have a controller cache. Support is merely provided for RAID levels RAID 0, 1 and 1E. Depending on the configurability of the PRIMERGY system used, up to four (LSI SAS 1064 controller) and eight (LSI SAS 1068 controller) SAS or SATA hard disks can be connected to such a controller. In some PRIMERGY models the LSI MegaRAID 1064 SAS controller is also used as an onboard version.




| | RAID 0 with 2HDs | | | | RAID 1 with 2HDs | | | | RAID 1E with 3HDs | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | a | b | c | d | a | b | c | d | a | b | c | d |
| HD Cache off | 249 | 71 | 3.0 | 16.0 | 125 | 78 | 2.6 | 18.5 | 187 | 22.4 | 3.1 | 24.0 |
| HD Cache on | 249 | 224 | 3.8 | 18.4 | 125 | 109 | 3.2 | 21.5 | 187 | 20.5 | 3.9 | 23.5 |

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

The diagram compares the throughputs of the supported RAID arrays. You can see with sequential read that all hard disks achieve the maximum possible throughput, regardless of in which RAID array they are configured or whether the disk cache is enabled or disabled.

The write throughput for sequential write in the RAID 0 array and with a disabled disk cache of the hard disks is considerably lower than the read throughput. By enabling the disk cache it is possible to more than treble the throughput for sequential write. With random access in the RAID 0 the throughput can also be improved by enabling the disk caches – namely by 26% for random access with 8 kB blocks and by 15% for random access with 64 kB blocks.

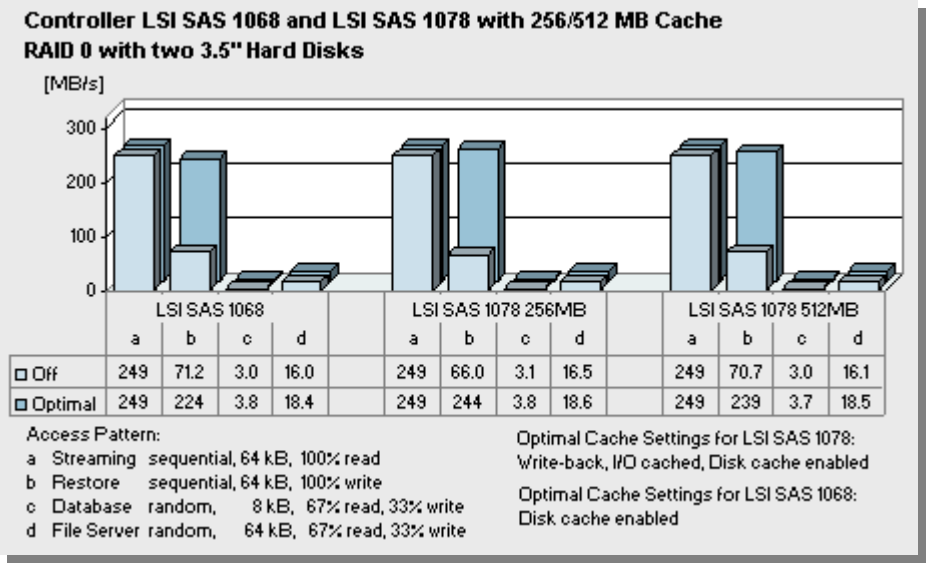
The throughput of the RAID 1 for sequential write with a disabled disk cache achieves approximately the same throughput value as with the RAID 0. Throughput improvement of about 40% is achieved by enabling the disk cache.

Throughput for sequential write in the RAID 1E is of a low level and is lower than with a RAID 1. The write throughput can only be slightly improved by enabling the disk cache. However, for random access a substantial improvement in throughput of about 26% was achieved through enabling the disk cache and the achieved throughput is on the same level as with the RAID 0 and RAID 1. With random access the RAID 1E throughput rate is about 30% above that of the RAID 1 and 50% above that of the RAID 0.

© Fujitsu Technology Solutions 2009

### Controller comparison

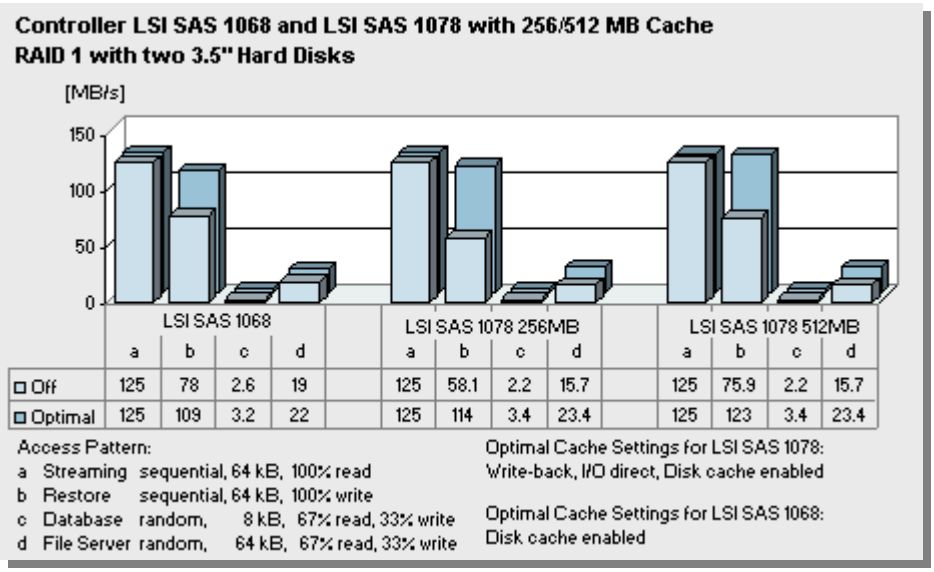
The comparison depicts the throughputs of the various controllers. The measurements were made with the same hard disk types in the same RAID array. The diagram shows the throughputs achieved with disabled caches (Off) and with optimal cache settings (Optimal).




The diagram shows that the cache settings do not have any impact on the throughput for sequential read access in the RAID 0 array with two hard disks. The throughput achieved is very close to the maximum possible values. All three controllers are equally efficient for sequential read with 64 kB blocks. On the other hand, for sequential write with 64 kB blocks it is possible to achieve a relevant increase in throughput through the optimal cache settings. With the LSI MegaRAID SAS 1068 controller the throughput increases 3.1-fold. With the LSI MegaRAID SAS 1078 controller with a 512 MB cache the throughput increases 3.4-fold and with the LSI MegaRAID SAS controller with a 256 MB cache throughput even increases 3.7-fold. The LSI MegaRAID SAS 1078 controllers achieve approximately the same throughputs, regardless of whether the version with the larger or smaller controller cache is used and they offer an up to 9% better performance than an LSI MegaRAID SAS 1068 controller.

For random access with 67% reads a substantial increase in throughput can in part also be achieved through optimal cache settings, but the absolute throughputs in the RAID 0 with two hard disks are approximately identical.

Similar to the RAID 0 array, the cache settings for sequential read also do not have any or only a very minor influence on throughput, regardless of which controller is used. The throughput values achieved are equivalent to the maximum possible values. For sequential write access it is also possible to achieve an increase in throughput with the optimal cache setting. However, it is not as pronounced as in the RAID 0 array. Throughput increases by about 40% with the LSI MegaRAID SAS 1068 controller. The best throughput values were achieved with the LSI MegaRAID SAS 1078 controller with a 512 MB controller cache. The difference in performance to the LSI MegaRAID 1078 controller with a 256 MB controller cache is about 8% with optimal cache settings and sequential write access. The difference in performance to the LSI MegaRAID SAS 1068 controller is approximately 13%.



# Conclusion

The “Modular RAID” concept of PRIMERGY Server offers a plethora of opportunities to meet the various requirements of different application scenarios.

The entry-level controller, represented by the LSI MegaRAID SAS 1068 controller, offers the basic RAID solutions RAID 0, RAID 1 and RAID 1E and at the same time supports these RAID levels with optimal performance.

The high-end controller, represented by the LSI MegaRAID 1078 controller, offers all today’s current RAID solutions RAID 0, 1, 5, 6, 10, 50 and 60. This controller is supplied with a 256 MB or 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, a RAID 10 for optimal performance and security is recommended.

“Modular RAID” is used in various PRIMERGY models, which differ with regard to hardware configuration and configuration options with controllers and hard disks. Depending on the PRIMERGY model, a choice between SATA and SAS, 2½" hard disks and 3½" hard disks is offered and also for SAS disks different speeds of 10 krpm or 15 krpm. Depending on the performance required, a decision must be taken as to whether hard disks with which rotational speed are to be used. Hard disks with 15 krpm offer an approximately 42% better performance. As a result of using 2½" hard disks it is possible – depending on the RAID level – to achieve higher parallelism through the use of more hard disks in the RAID array.

For maximum performance it is advisable, particularly with SATA hard disks or when using a controller without a controller cache, to enable the hard disk cache. Depending on the disk type used and access pattern, the increase in performance is 11-fold. When the hard disk cache is activated we recommend the use of an UPS.

© Fujitsu Technology Solutions 2009

# Literature

| | |
|---|---|
| PRIMERGY Systems | http://ts.fujitsu.com/primergy |
| PRIMERGY Performance | http://ts.fujitsu.com/products/standard_servers/primergy_bov.html |
| Information about Iometer | http://www.iometer.org |

# Contact

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

***PRIMERGY Performance and Benchmarks***
PRIMERGY Performance and Benchmarks
mailto:primergy.benchmark@ts.fujitsu.com

Delivery subject to availability, specifications subject to change without notice, correction of errors and omissions excepted.
All conditions quoted (TCs) are recommended cost prices in EURO excl. VAT (unless stated otherwise in the text). All hardware and software names used are brand names and/or trademarks of their respective holders.

Copyright © Fujitsu Technology Solutions GmbH 2009

Published by department:

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

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

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