T.C.

BAHÇEŞEHİR ÜNİVERSİTESİ

PERFORMANCE ANALYSIS ON RAID LEVELS

Master Thesis

Ali Rıza BALI

İSTANBUL, 2010

T.C.

BAHÇEŞEHİR ÜNİVERSİTESİ

The Graduate School of Natural and Applied Sciences

Title of Master Thesis : Performance Analysis on Raid Levels

Name/Last Name of the Student : Ali Rıza BALI

Date of Thesis Defense :

The thesis has been approved by the Computer Engineering.

Signature

Asst. Prof. Dr. F. Tunç BOZBURA

Acting Director

This is to certify that we have read this thesis and that we find it fully adequate in scope, quality and content, as a thesis for the degree of Master of Science.

Examining Committee Members Signature

Assoc. Prof. Dr. Adem KARAHOCA (Supervisor) ----------

Asst. Prof. Dr. Alper TUNGA ----------

Asst. Prof. Dr. Yalçın ÇEKİÇ ----------

T.C.

BAHÇEŞEHİR ÜNİVERSİTESİ

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

PERFORMANCE ANALYSIS ON RAID LEVELS

Master Thesis

Ali Rıza BALI

Supervisor: ASSOC. PROF. DR. ADEM KARAHOCA

İSTANBUL, 2010

ACKNOWLEDGEMENTS

I dedicate this thesis to my family who was my biggest supporter in my life.

I would like to express my gratitude to my supervisor Assoc. Prof. Dr. Adem KARAHOCA. Without his guidance and persistent help this thesis would not have been possible.

Also I would like to thank my precious friends because they always encouraged me.

ABSTRACT

PERFORMANCE ANALYSIS ON RAID LEVELS

Ali Rıza BALI

M.S. Department of Computer Engineering
Supervisor: Assoc. Prof. Dr. Adem Karahoca

June 2010, 74 pages

Increasing the performance is one of the primary goals of IT departments. For companies which have to run different data types and applications with different sizes on server environments, it is necessary to determine the best raid configuration.

This work has intended to find out most suitable structure on storage systems to achieve optimum performance. Performance need became more critical because of fast technology growth. Deep analysis made due to results of RAID performance tests.

With this thesis array performance will be monitored on different hardware and raid level configurations with different data type and size. Monitoring and testing program will be Hp Library and Tape Tools (Hp L&TT). All collected data will be stored with Microsoft Excel for analyzing. Stored data will be analyzed with SPSS 1.6. Statistical data will be compared to determine best raid level and array configuration for specific data, application type and size.

Key Words: Performance, RAID, I/O, redundancy, striping, mirroring, data, system

ÖZET

RAID SEVİYELERİNDE PERFORMANS ANALİZİ

Ali Rıza BALI

Yüksek Lisans, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi: Doç. Dr. Adem Karahoca

Haziran 2010, 74 sayfa

Performansı arttırmak BT birimlerinin en öncelikli görevlerinden birisidir. Değişik veri tipleri ve uygulamalar ile çalışmak durumunda olan şirketler, uygun değerli performansı elde etmek için en uygun sistemi ve en uygun RAID yapısını belirlemek ihtiyacını duymaktadır.

Bu çalışma, bilgi teknolojilerindeki hızlı gelişmelerden ötürü önem kazanan yüksek performans gereksiniminin depolama birimlerinde nasıl en uygun elde edilebileceğini araştırmak üzere yapılmıştır. Birden fazla diski bir araya getirerek, çoklu diskin faydalarını tek bir yapı halinde kullanan RAID sistemleri üzerinde detaylı testler yapılmış ve bunların sonuçları yorumlanmıştır.

Bu çalışma ile disk grubu performansı farklı RAID seviyeleri üzerinde farklı uygulamalar ve veri tipleri ile gözlemlenecektir. İzleme ve sınama aracı Hp Library and Tape Tools(Hp L&TT) olacaktır. Elde edilen tüm veriler analiz edilmek üzere Microsoft Excel programında saklanan veriler SPSS 1.6 aracı kullanılarak analiz edilecek. Değişik veri ve uygulama tipleri için en uygun RAID seviyelerini tespit etmek için istatistiksel veriler karşılaştırılacaktır.

Anahtar Kelimeler: Performans, RAID, okuma yazma, yedeklilik, şeritleme, aynalama, veri, sistem

TABLE OF CONTENTS

1 Introduction to RAID Technology

1.1 What is RAID?

RAID is a storage architecture which was designed to achieve high performance and redundant disk systems. The RAID stands for "Redundant Array of Independent Disks". This concept was proposed in 1987 when "A Case for Redundant Arrays of Inexpensive Disks (RAID)" was published by David Patterson, Garth Gibson, and Randy Katz at the University of California, Berkeley (Richard G. Krum & Virat Thantrakul, 1998). You can read the original RAID study at: http://techreports.lib.berkeley.edu/accessPages/CSD-87-391.html

Increasing performance of CPUs and memories will be squandered if not matched by a similar performance increase in I/O. While the capacity of Single Large Expensive Disk (SLED) has grown rapidly, the performance improvement of SLED has been modest. Redundant Arrays of Inexpensive Disks (RAID), based on the magnetic disk technology developed for personal computers, offers an attractive alternative to SLED, promising improvements of an order of magnitude in performance, reliability, power consumption, and scalability (UC Berkeley, 1987).

The idea was to combine multiple small, inexpensive physical disks into an array that would function as a single logical drive, but provide better performance and higher data availability than a single large expensive disk drive (SLED) (Hp-WiR, 2007).

With a single small disk drive you can have less capacity compared to large disk drives but with RAID technology grouping small disk drives into an array provides the following additional advantages:

- An array of multiple disks accessed in parallel will give greater throughput than a single disk (High transfer rates, High I/O rates).
- Increased disk capacity.
- Redundant data on multiple disks provides fault tolerance (redundancy).

The original study was proposing to rebuild data from backup tape drives when a disk crashes. Also, replacing the failed disk required a downtime. Now, redundancy ensured by RAID technology and hot swap disk drives itself.

With this study we will analyse performance of 3 levels as following;

RAID 0: Increased disk capacity, high transfer rates and high I/O rates at same cost compared to large disk drives.

RAID 1: Provides full redundancy and improved I/O.

RAID 5: Combining redundancy, high I/O rates and increased disk capacity. (Scott Baderman, 2003)

1.1.1 Which RAID Level is Suitable for Me?

With this work I have intended to determine that which application or environment requires which RAID level. Everybody, who related with information technologies, wonder about that which structure can meet its environment requirements. To determine that, first, we need to understand applications and related file properties. Applications and environments can be categorised as database, multimedia, various files, exchange etc...

1.1.1.1 Database files

A database is an integrated collection of logically-related records or files consolidated into a common pool that provides data for one or more multiple uses. One way of classifying databases involves the type of content, for example: bibliographic, full-text, numeric and image (Ling Liu & Tamer M. Özsu, 2009). A database generally contains small, important and commonly applicable data. This infrastructure needs reliable and fast infrastructure to work on small sized files.

1.1.1.2 Multimedia files

Multimedia is media and content that uses a combination of different content forms. Multimedia includes a combination of text, audio, still images, animation, video, and interactivity content forms (Stewart, C & Kowaltzke, A, 1997).

Multimedia files generally need to be stored on big sized infrastructures. Multimedia files are less important files compared to database files.

1.1.1.3 Exchange Files

It is part of the Microsoft Servers line of server products and is used by enterprises using Microsoft infrastructure solutions. Exchange's major features consist of electronic mail, calendaring, contacts and tasks; support for mobile and web-based access to information; and support for data storage (McBee, Jim & Barry Gerber, 2007).

An Exchange file generally contains small, important and commonly applicable data. This infrastructure needs reliable and fast infrastructure to work on small sized files.

1.1.1.4 High End Workstations

A workstation is a high-end microcomputer designed for technical or scientific applications. Intended primarily to be used by one person at a time, they are commonly connected to a local area network and run multi-user operating systems. The term workstation has also been used to refer to a mainframe computer terminal or a PC connected to a network (Wikipedia-Workstation).

1.1.1.5 Data logger

A data logger (also data logger or data recorder) is an electronic device that records data over time or in relation to location either with a built in instrument or sensor or via external instruments and sensors. Increasingly, but not entirely, they are based on a digital processor (or computer). They generally are small, battery powered, portable, and equipped with a microprocessor, internal memory for data storage, and sensors. Some data loggers interface with a personal computer and utilize software to activate the data logger and view and analyze the collected data, while others have a local interface device (keypad, LCD) and can be used as a stand-alone device (Riva, Marco & Piergiovanni, Schiraldi, 2001).

1.1.1.6 Real Time rendering

Real-time rendering is the one of the interactive areas of computer graphics; it means creating synthetic images fast enough on the computer so that the viewer can interact with a virtual environment. The most common place to find real-time rendering is in animated movies or video games. The rate at which images are displayed is measured in

frames per second (frame/s) or Hertz (Hz). The frame rate is the measurement of how quickly an imaging device produces unique consecutive images. If an application is displaying 15 frame/s it is considered real-time (Möller, Tomas, & Eric Haines, 1999).

1.1.1.7 Operating System

In computing, an operating system (OS) is software (programs and data) that provides an interface between the hardware and other software. The OS is responsible for management and coordination of processes and allocation and sharing of hardware resources such as RAM and disk space, and acts as a host for computing applications running on the OS. An operating system may also provide orderly accesses to the hardware by competing software routines. This relieves the application programmers from having to manage these details (Bic, Lubomur F. & Shaw, Alan C., 2003).

1.1.1.8 Database Transaction

A database transaction comprises a unit of work performed within a database management system (or similar system) against a database, and treated in a coherent and reliable way independent of other transactions (Philip A. Bernstein & Eric Newcomer, 2009). Transactions in a database environment have two main purposes:

1. To provide reliable units of work that allow correct recovery from failures and keep a database consistent even in cases of system failure, when execution stops (completely or partially) and many operations upon a database remain uncompleted, with unclear status.

2. To provide isolation between programs accessing a database concurrently. Without isolation the programs' outcomes are possibly erroneous .

1.1.1.9 Data warehouse

A data warehouse is a repository of an organization's electronically stored data. Data warehouses are designed to facilitate reporting and analysis.

This definition of the data warehouse focuses on data storage. However, the means to retrieve and analyze data, to extract, transform and load data, and to manage the data dictionary are also considered essential components of a data warehousing system. Many references to data warehousing use this broader context. Thus, an expanded

definition for data warehousing includes business intelligence tools, tools to extract, transform, and load data into the repository, and tools to manage and retrieve metadata.
Data warehousing arises in an organisation's need for reliable, consolidated, unique and integrated reporting and analysis of its data, at different levels of aggregation (Inmon, W.H., 1995).

1.1.1.10 Web Server

It has been described by (Wikipedia-WS)
A Web server is a computer program that delivers (serves) content, such as Web pages, using the Hypertext Transfer Protocol (HTTP), over the World Wide Web. The term Web server can also refer to the computer or virtual machine running the program. In large commercial deployments, a server computer running a Web server can be rack-mounted in a server rack or cabinet with other servers to operate a Web farm.

1.1.1.11 Archiving

An archive is a collection of historical records, as well as the place they are located. Archives contain primary source documents that have accumulated over the course of an individual or organization's lifetime.
In general, archives consist of records that have been selected for permanent or long-term preservation on grounds of their enduring cultural, historical, or evidentiary value. Archival records are normally unpublished and almost always unique, unlike books or magazines for which many identical copies exist. This means that archives (the places) are quite distinct from libraries with regard to their functions and organization, although archival collections can often be found within library buildings (Walch, Victoria Irons, 2006).

1.1.2 The RAID Concept

Disk arrays are an integral part of high-performance storage systems, and their importance and scale are growing as continuous access to information becomes critical to the day-to-day operation of modern business. (SelectingRAID, 2002)

By combining multiple physical disks for improved data input/output performance the RAID study proposed a multilevel concept. Study also proposed improved data availability by avoiding the impact of disk drive failures. Five original RAID levels

(RAID 1 through RAID 5), were defined to meet the needs of various computing environments. Data redundancy increases as the five original RAID configurations progress from RAID 1 through RAID 5. (Scott Baderman, 2003)

Main attributes was explained as following by (Hp-WiR, 2007)

Overall, RAID has three main attributes that are exploited in some way by all five original RAID configurations and by most other RAID configurations that have been defined since the 1987 study. These attributes are:

- A set of physical disk drives that can function as one or more logical drives (improved I/O)
- Data distribution across multiple physical disks (striping)
- Data recovery, or reconstruction of data in the event of a physical disk failure (redundancy)

RAID configurations, which we will examine for the performance analyse, are as follows:

RAID 0: Increased disk capacity, high transfer rates and high I/O rates at same cost compared to large disk drives.

RAID 1: Full redundancy and improved I/O.

RAID 5: Combining redundancy, high I/O rates and increased disk capacity.

1.2 Performance and Data Redundancy

1.2.1 Increasing Logical Drive Performance

Connecting extra physical disks to a system without an array controller increases the total storage capacity. However, it has no effect on the efficiency of read/write operations, because data can only be transferred to one physical disk at a time (See figure 1-1).

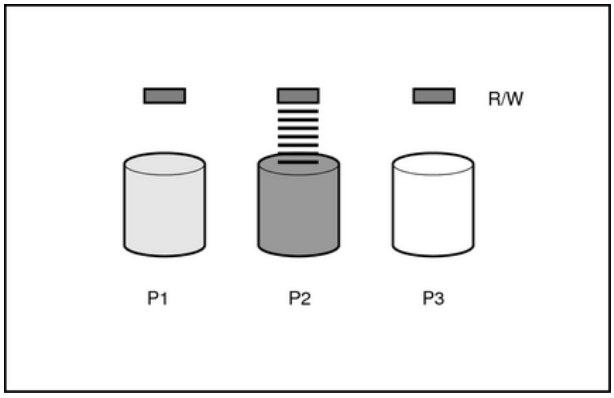


Figure 1-1 Disks without an array controller

Connecting extra physical disks to a system with an array controller increases both the total storage capacity and the read/write efficiency. The capacity of several physical disks is combined into one or more virtual units called logical drives (also called logical volumes).

The read/write heads of all of the physical disks in a logical drive are active simultaneously; improving I/O performance and reducing the total time required for data transfer (See figure 1-2).

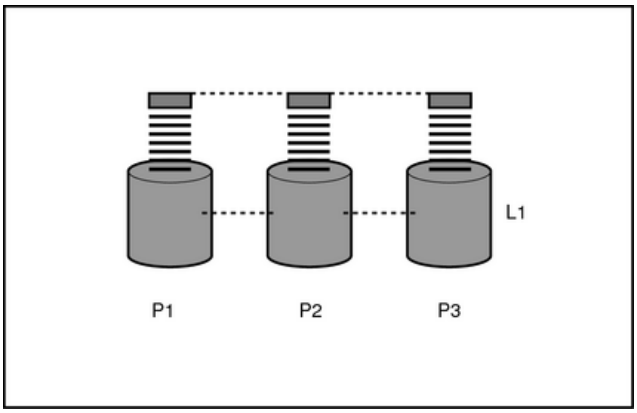


Figure 1-2 Disks Configured into a Logical Drive (L1)

Because the read/write heads for each physical disk are active simultaneously, the same amount of data is written to each disk during any given time interval. Each unit of data is called a block.

Data is striped across an array of physical drives. The granularity at which data is stored on one drive of the array before subsequent data is stored on the next drive of the array is called the stripe-unit size. (IBM, 2006)

The blocks form a set of data stripes that are spread evenly over all the physical disks in a logical drive (See figure 1-3).

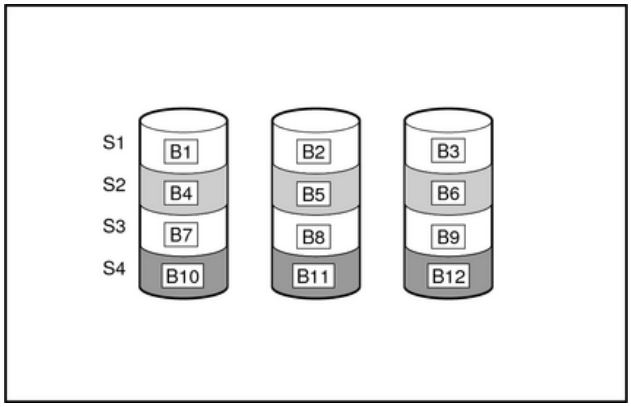


Figure 1-3 Data Striping (S1-S4) of Data Blocks B1-B12

For data in the logical drive to be readable, the data block sequence must be the same in every stripe. This sequencing process is performed by the Smart Array Controller, which sends the data blocks to the physical disk, writing the heads in the correct order.

In a striped array, each physical disk in a logical drive contains the same amount of data. If one physical disk has a larger capacity than other physical disks in the same logical drive, the extra capacity cannot be used.

A logical drive can extend over more than one channel on the same controller, but it cannot extend over more than one controller.

Disk failure, although rare, is potentially catastrophic to an array. If a physical disk fails, the logical drive it is assigned to fails and all of the data on that logical drive is lost. (S. Savage & J. Wilkes, 1996).

1.2.2 Protecting Data with Fault Tolerance and Spare Disks

To protect against data loss due to physical disk failure, logical drives can be configured with fault tolerance. Fault-tolerant RAID configurations, which we will examine for the performance analyse, are as follows:

RAID 1 Data mirroring only (fault tolerant)

RAID 5 Distributed Data guarding (fault tolerant)

For any fault-tolerant configuration, you can create further protection against data loss by assigning a physical disk as an online spare (or "hot spare"). Spare disks contain no data and must be in the same array as the logical drive they are assigned to. Multiple spare physical disks can be assigned to a logical drive, limited only by the availability of unused disks in the array.

When a physical disk in the array fails, the controller automatically rebuilds the information from the failed disk onto an online spare. The system is quickly restored to full RAID-level data protection. In the unlikely event that another disk in the array fails while data is being rewritten to the spare, the logical drive may fail, depending on which RAID configuration is in use.

1.3 RAID Configurations

This section provides details about each of the RAID levels, which we will examine for the performance analyse.

This chapter addresses the following topics:

"RAID 0: No Fault Tolerance"

"RAID 1: Disk Mirroring"

"RAID 5: Distributed Data Guarding"

"Summary of RAID Methods"

1.3.1 RAID 0: No Fault Tolerance

The RAID 0 configuration enhances performance with data striping, but there is no data redundancy to protect against data loss when a physical disk fails. (David C. Stallmo & Randy K. Hall, 1997). RAID 0 is useful for rapid storage of large amounts of non-critical data (for printing or image editing, for example), or when cost is the most important consideration (See figure 1-4).

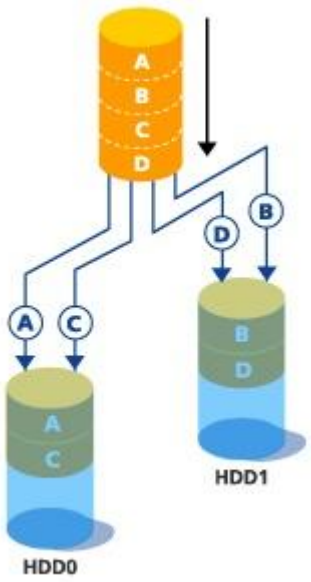


Figure 1-4 RAID 0 structures - data striping

The advantages of RAID 0 are as follows:

• Highest performance configuration for writes

• Lowest cost per unit of data stored

• All disk capacity is used to store data (none needed for fault tolerance)

The disadvantages of RAID 0 are as follows:

• All data on the logical drive is lost if a physical disk fails.

• Online spare disks are not available.

• Data preservation by backing up to external physical disks only.

1.3.2 RAID 1: Disk Mirroring

In spite of its high redundancy level, disk mirroring is a popular RAID paradigm, because replicating data also doubles the bandwidth available for processing read requests, improves the reliability and achieves fault tolerance. (RaidRMS, 2009).

In this configuration, only two physical disks are present in the array. Data is duplicated from one disk onto the other, creating a mirrored pair of disk drives, but there is no striping of data (See figure 1-5). (David C. Stallmo & Randy K. Hall, 1997).

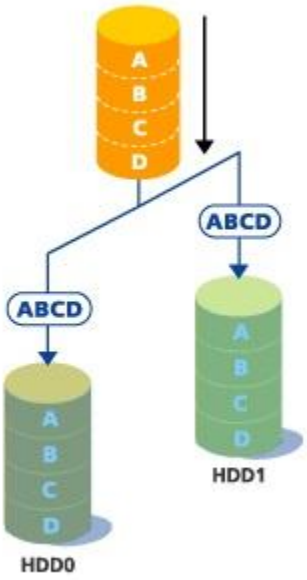


Figure 1-5 RAID 1 structure - data mirroring

The advantages of RAID 1 are as follows:

• No data loss or interruption of service if a disk fails.

• Fast read performance — data is available from either disk.

The disadvantages of RAID 1 are as follows:

• High cost — 50% of disk space is allocated for data protection, so only 50% of total disk drive capacity is usable for data storage.

1.3.3 RAID 5: Distributed Data Guarding

RAID 5 uses a parity data formula to create fault tolerance. In RAID 5, one block in each data stripe contains parity data that is calculated for the other data blocks in that stripe. The blocks of parity data are distributed over the physical disks that make up the logical drive, with each physical disk containing only one block of parity data (See figure 1-6). (David C. Stallmo, Randy K. Hall, 1997). When a physical disk fails, the data that was on the failed disk can be calculated from the parity data in the data blocks on the remaining physical disks in the logical drive. This recovered data is usually written to an online spare in a process called a rebuild.

RAID 5 is useful when cost, performance, and data availability are all equally important.

Figure 1-6 RAID 5 structure – distributed data guarding

The advantages of RAID 5 are as follows:

• High read performance

• No loss of data if one physical disk fails.

• More usable disk capacity than with RAID 1+0; parity information only requires the storage

Space equivalent to one physical disk on the array.

The disadvantages of RAID 5 are as follows:

• Relatively low write performance

• Data loss occurs if a second disk fails before data from the first failed disk is rebuilt.

1.3.4 Summary of RAID Methods

Table 1-1 summarizes the important features of the different RAID configurations.

Table 1-1 Summary of important features of different raid configurations

	RAID 0	RAID1	RAID 5
Alternative name	Striping (no fault tolerance)	Mirroring	Distributed Data Guarding
Usable disk space	100%	50%	67% to 96%
Usable disk space formula	n	n/2	(n-1)/n
Minimum number of physical disks	1	2	3
Tolaretes failure of one physical disk?	No	Yes	Yes
Tolarates simultaneous failure of more than one physical disk?	No	No	No

1.4 Background

With technological improvements on CPU and memory architecture, storage performance became a bottleneck on overall system performance. Increasing storage performance has been very important for decades. RAID technology is a known solution for that bottleneck but there are still some questions about that architecture's levels and applying area. The aim of this study is to find out that which RAID level is the most suitable level for different data and application types. Besides, I have noticed some researches like this study.

1.4.1 Literature Survey

Some previous works on this topic are as follows;

- S. Savage and J. Wilkes worked on an enhanced model on RAID to prevent redundancy write penalty in 1996. They have suggested their enhanced model named which was named as AFRAID. Their work published with name "AFRAID—a frequently redundant array of independent disks". That was a very important work at these years because systems were suffering from redundancy write penalty.

- David C. Stallmo and Randy K. Hall worked for improving RAID performance. Their work published in 1997 with name "Method and apparatus for improving performance in a redundant array of independent disks". They have developed an adaptive system that dynamically determines the RAID configuration used to store host data to maximize response time performance and minimize the loss of disk space used for data protection. That work was a big step for these years.

- Lu Zheng-wu, Xie Chang-sheng and Jiang Guo-song worked on improving RAID read performance in 2008. They gave an international conference on Computer Science and Software Engineering Research of Improving RAID Read Performance. In their paper, the file system principles and RAID algorithm were studied in Linux, current MD driver and RAID performance were anglicized in detail, achieved a RAID algorithm in the field of multimedia applications, this algorithm enhanced the RAID server to read performance effectively, broke i386 operating system's 4K paging limit, used the new pile of management memory, and through the experimental data show that the new algorithm brought about excellent performance. For detail experiment, 64 k-blocks have been proven to the ideal block using in multimedia applications. Their idea to broke operating system's paging limit to improve RAID read performance was resulted new ideas on block and stripe sizes

- Abigail S. Lebrecht, Nicholas J. Dingle, and William J. Knottenbelt worked on "A Response Time Distribution Model for Zoned RAID" and published that work in 2008. Their paper presents a queuing network-based model of RAID systems comprised of zoned disks and operating at RAID level 0-1 or 5. This work showed that distribution of I/O queue to different RAID zones very effective on RAID performance.

- Javad Akbari Torkestania and Mohammad Reza Meybodib worked on "RAID-RMS: A fault tolerant stripped mirroring RAID architecture for distributed systems" and published that work on 2 September 2008. In this paper, writers present a new RAID architecture called RAID-RMS in which a special hybrid mechanism is used to map the data blocks to the cluster. The main idea behind the proposed algorithm is to combine the data block striping and disk mirroring technique with a data block rotation. The resulting architecture improves the parallelism reliability and efficiency of the RAID array. Writers show that the proposed architecture is able to serve many more disk requests compared to the other mirroring-based architectures. Writers also argue that a more balanced disk load is attained by the given architecture, especially when there are some disk failures.

- George Ou published a report about "Comprehensive RAID performance" on May 4th, 2007. The aim of the paper is comparing a large set of RAID performance data and perhaps debunks some storage myths. Writer is trying to determine that what hardware how effects the raid performance and which raid level is the best for what. Most importantly, he says he has worked to clear that two individual raid1 with 2 drives comes beneficial over one raid1+0 with 4 drives.

- Robin Harris published an article about "Chunks: the hidden key to RAID performance "on May 7th, 2007. The aim of the paper is pointing that, what about RAID itself and what is the theory behind RAID performance? Writer points that the RAID concept was born for the cost but not the performance. And raid performance bases on three key concepts. These are Cache, Striping and

Chunk size. He proved that, using Cache memory increases the raid performance dramatically. Also, Big I/Os = small chunks; small I/Os = big chunks.

- Shahid Bokhari, Benjamin Rutt, PeteWyckoff and Paul Buerger published their work on Experimental analysis of a mass storage system on 4 April 2006. The aim of the paper is determining the bottlenecks of a Mass storage system (MSSs) and the ways of improving overall performance. Mass storage systems (MSSs) play a key role in data-intensive parallel computing. Most contemporary MSSs are implemented as redundant arrays of independent/inexpensive disks (RAID) in which commodity disks are tied together with proprietary controller hardware. The performance of such systems can be difficult to predict because most internal details of the controller behaviour are not public. Experiment team present a systematic method for empirically evaluating MSS performance by obtaining measurements on a series of RAID configurations of increasing size and complexity. Experiment team apply this methodology to a large MSS at Ohio Supercomputer Centre that has 16 input/output processors, each connected to four 8 + 1 RAID5 units and provides 128 TB of storage (of which 116.8 TB are usable when formatted). Their methodology permits storage-system designers to evaluate empirically the performance of their systems with considerable confidence. Although experiment team have carried out our experiments in the context of a specific system, our methodology is applicable to all large MSSs. The measurements obtained using our methods permit application programmers to be aware of the limits to the performance of their codes.

2 Tools, Data, Method, Tests and Results

HP Storage Works Library and Tape Tools have used to collect data. Collected data analysed with Statistical Package for the Social Sciences.

HP Storage Works Library and Tape Tools (L&TT) is a robust diagnostic tool for tape storage and magneto-optical storage products. L&TT can perform read and write tests with different data length, and data sequence types.

SPSS (originally, Statistical Package for the Social Sciences) was released in its first version in 1968 after being developed by Norman H. Nie and C. Hadlai Hull. Norman Nie was then a political science postgraduate at Stanford University, and now Research Professor in the Department of Political Science at Stanford and Professor Emeritus of Political Science at the University of Chicago. SPSS is among the most widely used programs for statistical analysis in social science. It is used by market researchers, health researchers, survey companies, government, education researchers, marketing organizations and others. The original SPSS manual (Nie, Bent & Hull, 1970) has been described as one of "sociology's most influential books". In addition to statistical analysis, data management (case selection, file reshaping, creating derived data) and data documentation (a metadata dictionary is stored in the data file) are features of the base software.

Statistics included in the base software:

* Descriptive statistics: Cross tabulation, Frequencies, Descriptives, Explore, Descriptive Ratio Statistics

* Bivariate statistics: Means, t-test, ANOVA, Correlation (bivariate, partial, distances), Nonparametric tests

* Prediction for numerical outcomes: Linear regression

* Prediction for identifying groups: Factor analysis, cluster analysis (two-step, K-means, hierarchical), Discriminant

2.1 Performance Tests and Results

This Section provides details about methodology and concepts, which are parameters of performance tests, and deductions from tests results.

We have run 330 read/write tests to analyse performance. 110 tests for Raid 0 with two 146 GB sized SAS disks, 110 tests for Raid 1 with two 146 GB sized SAS disks and 110 tests for Raid 5 with three 146 GB sized SAS disks (See table 2-1). All disks and logical drivers were formatted with stripe size 128KB. The test platform was Hp Proliant DL380 G5 server and Smart Array E200 controller.

Table 2-1 Test frequency distribution per raid level

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	raid0	110	33,3	33,3	33,3
	raid1	110	33,3	33,3	66,7
	raid5	110	33,3	33,3	100,0
	Total	330	100,0	100,0	

This section addresses the following topics:

“Tests Parameters”

“Tests Aims and Methods”

2.1.1 Tests Parameters

We have 5 ordinal and 3 scalable, totally 8 test parameters. These are “Raid Levels”, “Restore Performance Test File Tree Depth”, “Restore Performance Test File Tree Breadth “, “Write Speed”, “Test Size”, “Backup Performance Test Read Size”, “Backup Performance Test Directory Traverse Method “and “Read Speed”.

2.1.1.1 Raid Levels

Raid Levels is an ordinal parameter because it is a categorical data. In our scenario there are 3 categories, which are Raid0, Raid1 and Raid5. We will analyse our test result within these categories.

2.1.1.2 Restore Performance Test File Tree Depth (RPTDTD)

RPTDTD is an ordinal parameter because we have categorised this parameter with for values, which are 1, 2, 3, 6 (See table 2-2). RPTDTD indicates that, how many folders will be created as a depth nested structure. For example, if we assign 1 to that value, we will only have one test folder and that folder will be the root folder. If we assign 2 to that value, we will have one more folder in our root folder as a nested structure.

Table 2-2 Restore Performance Test File Tree Depth frequency

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	66	20,0	20,0	20,0
	2	66	20,0	20,0	40,0
	3	132	40,0	40,0	80,0
	6	66	20,0	20,0	100,0
	Total	330	100,0	100,0	

If we assign 3 to that value, we will have 3 nested folders (See figure 2-2). This parameter will affect our write speed. We can set that parameter to minimum 1 (See figure 2-1).

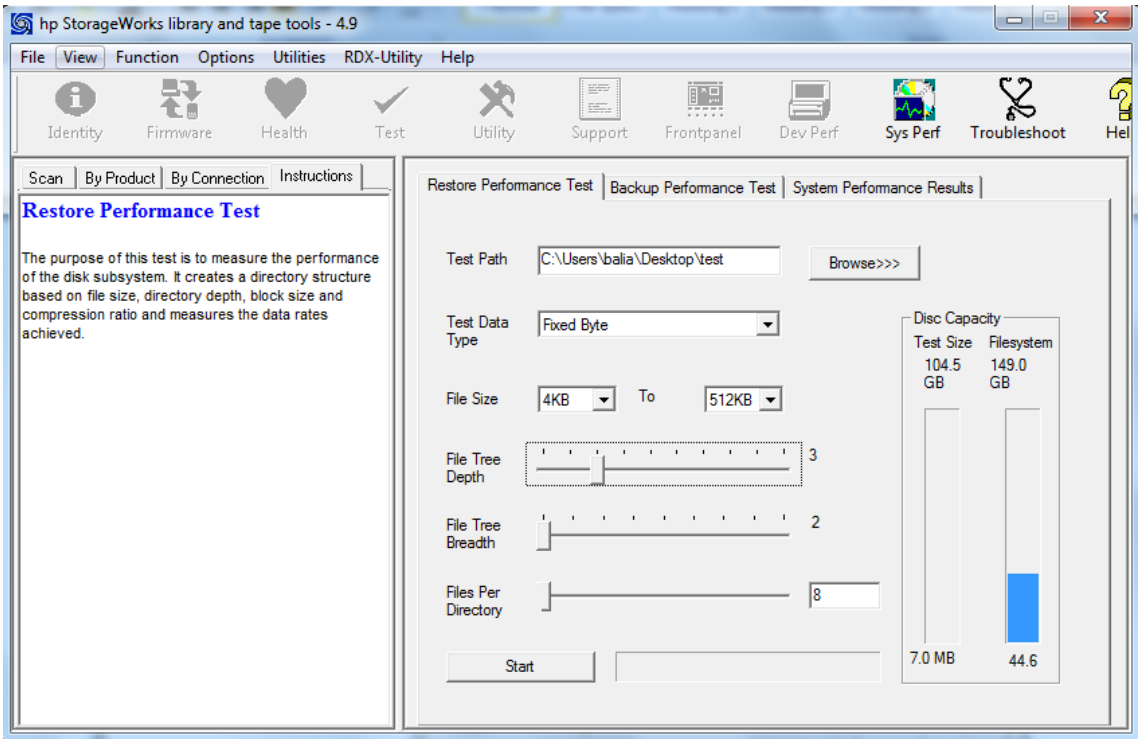


Figure 2-1 Restore Performance Test values set screen on Hp Storage Works Library and Tape Tools

2.1.1.3 Restore Performance Test File Tree Breadth (RPTFTB)

RPTFTB is an ordinal parameter because we have categorised this parameter with for values, which are 1, 2, 3, 7 (See table 2-3).

Table 2-3 Restore Performance Test File Tree Breadth frequency

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	66	20,0	20,0	20,0
	2	132	40,0	40,0	60,0
	3	66	20,0	20,0	80,0
	7	66	20,0	20,0	100,0
	Total	330	100,0	100,0	

RPTFTB indicates that, how many folders will be created in a folder as a breadth nested structure. For example, if we assign 1 to that value, we will only have folders from depth parameter. If we assign 2 to that value, we will have 2 folders in every folder as a breadth nested structure (See figure 2-2). This parameter will affect our write speed.

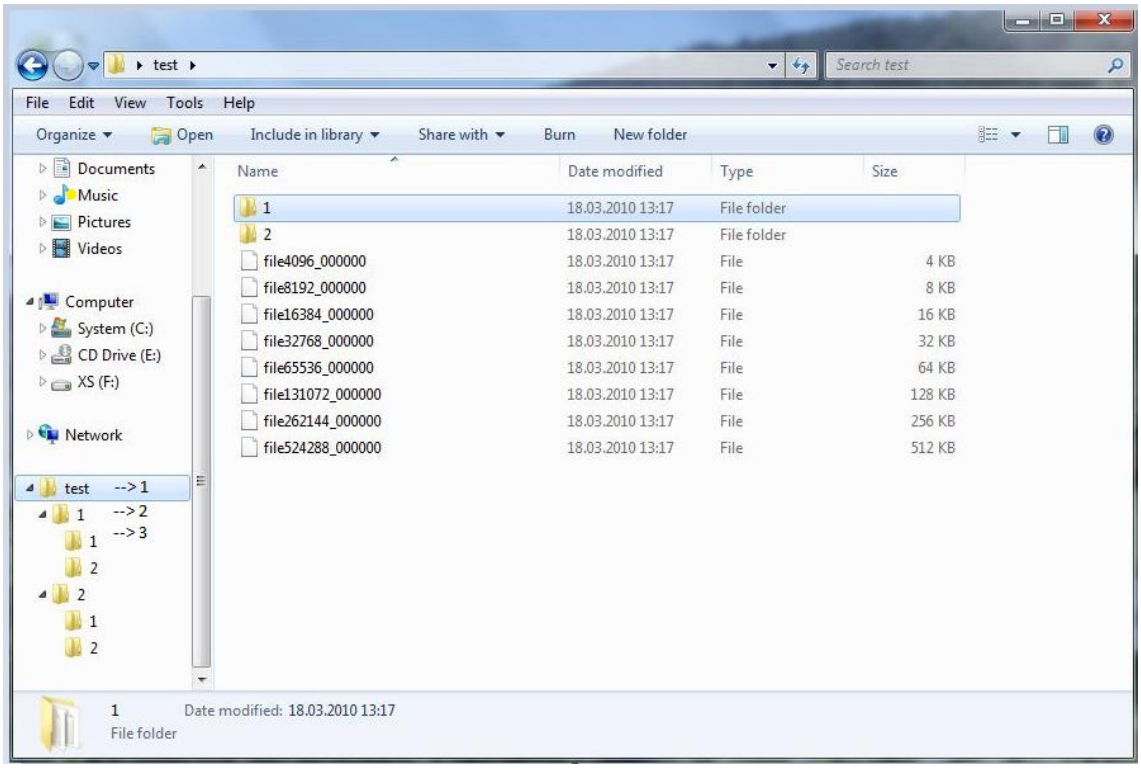


Figure 2-2 an example screen of RPTDTD value results

2.1.1.4 Write Speed

Write speed is a scalable parameter because we will have different values for every different test (See table 2-4). Write speed indicates test data's write speed to test environment. **Table 2-4 Raid Levels * Write Speed (MB/sec) Cross tabulation**

		Write Speed (MB/sec)											
		1	2	13	14	17	29	30	31	51	52	53	Total
Raid Levels	raid0	0	22	0	0	0	0	0	0	22	44	22	110
	raid1	0	22	0	0	0	22	22	44	0	0	0	110
	raid5	22	0	23	44	21	0	0	0	0	0	0	110
	Total	22	44	23	44	21	22	22	44	22	44	22	330

2.1.1.5 Read Speed

Read speed is a scalable parameter because we will have different values for every different test. Read speed indicates test data read speed from test environment.

2.1.1.6 Backup Performance Test Read Size

Backup Performance Test Read Size is an ordinal parameter because we have categorised this parameter with values 1KB, 2KB, 4KB, 8KB, 16KB, 32KB, 64KB, 128KB, 256KB, 512KB, and 1024KB. This value shows us that what will be the read block size per 1 clock time.

2.1.1.7 Test Size

Test size is a scalable parameter because we will have different values for every different test. Test size value will depend on RPTDTD, RPTFTB and files per directory parameter on L&TT (See table 2-5).

Table 2-5 Test Size (MB) frequency

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2047	66	20,0	20,0	20,0
	6143	66	20,0	20,0	40,0
	9804	66	20,0	20,0	60,0
	14335	66	20,0	20,0	80,0
	26623	66	20,0	20,0	100,0
	Total	330	100,0	100,0	

2.1.1.8 Backup Performance Test Directory Traverse Method

Backup Performance Test Directory Traverse Method Size is an ordinal parameter because we have categorised this parameter with values 0=Depth and 1= Breadth (See table 2-6). If value is Depth test program forces system to read data deeply, if value is Breadth test program forces system to read data breadth.

Table 2-6 Backup Performance Test Directory Traverse Method (0=Depth, 1=Breadth) frequency

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	165	50,0	50,0	50,0
	1	165	50,0	50,0	100,0
	Total	330	100,0	100,0	

2.1.2 Tests Aims and Methods

As mentioned before, I have performed 330 tests on same platform with different parameters. All these parameters were used to determine performance changes with different scenarios.

I have used a fixed stripe size value for all logical drives as 128 KB. Despite that, we have selected our test data to be written and read with different stripe sizes. For all raid levels we have selected read/write stripe size as 4KB to 128MB at 88 tests, but remained 22 tests was performed with fixed 4KB test data. This method will show us that, which raid level will provide which performance at which data sizes.

I have also used a depth and breadth file structure to perform performance tests. RPTFTD creates depth nested folder structure and RPTFTB create breath folder structure in depth nested structure. We can assume that RPTFTD creates folders vertically and RPTFTB creates folders horizontally. The minimum complexity and most basic structure will be created with RPTFTD's value1. When RPTFTD's value is 1, we can't assign a value to RPTFTB. Read/write speeds will be different on different folder

structures (See table 2-7).

Table 2-7 Raid Levels * Write Speed (MB/sec) * Restore Performance Test File Tree Depth Cross tabulation

Restore Performance Test File Tree Depth			Write Speed (MB/sec)										
			1	2	13	14	17	29	30	31	51	52	53	Total
1	Raid Levels	raid0			0		0			0			22	22
		raid1			0		0			22			0	22
		raid5			1		21			0			0	22
		Total			1		21			22			22	66
2	Raid Levels	raid0			0			0			22			22
		raid1			0			22			0			22
		raid5			22			0			0			22
		Total			22			22			22			66
3	Raid Levels	raid0				0			0	0		44		44
		raid1				0			22	22		0		44
		raid5				44			0	0		0		44
		Total				44			22	22		44		132
6	Raid Levels	raid0	0	22										22
		raid1	0	22										22
		raid5	22	0										22
		Total	22	44										66

For all written data I have performed 22 read tests. 11 of these 22 tests were for Back up Performance Test Read Size parameter when Backup Performance Test Directory Traverse Method was “Depth” and remaining 11 tests for Back up Performance Test Read Size parameter when Backup Performance Test Directory Traverse Method was “Breadth”.

As mentioned before, stripe size was chosen as128 KB for all logical drives. I have used Back up Performance Test Read Size parameter for read tests to see read performance differences between different red stripe sizes (See figure 2-3).

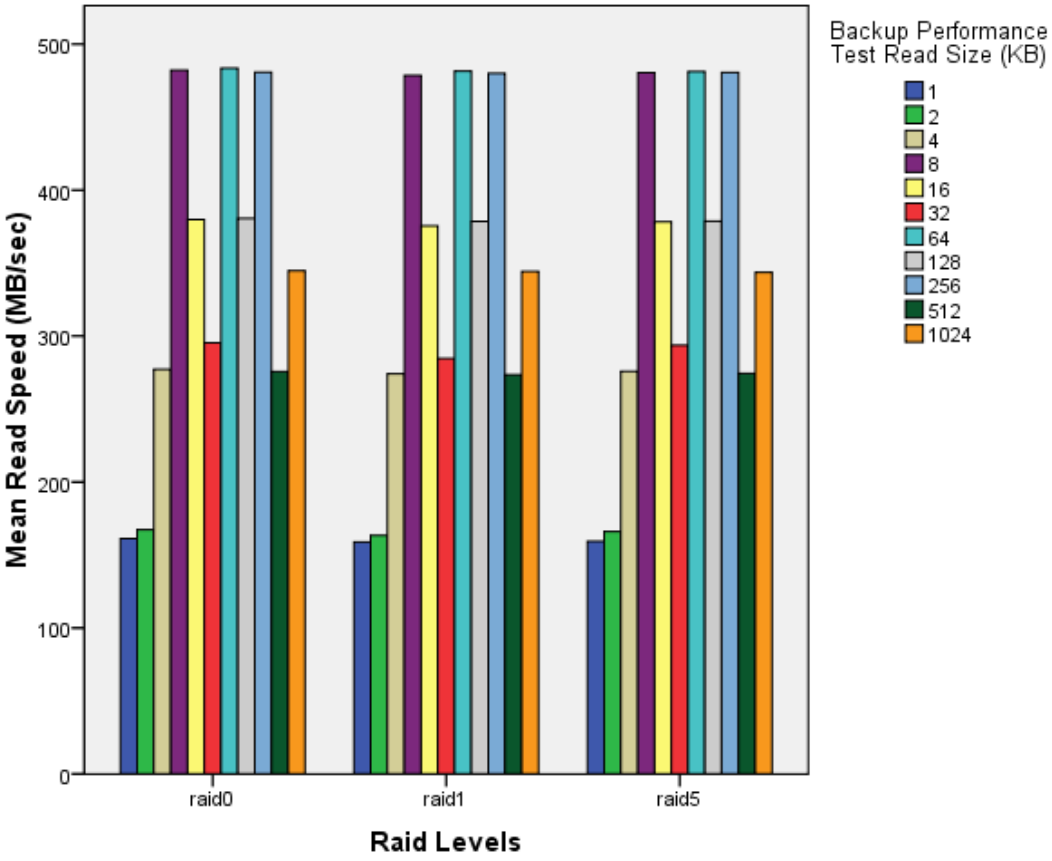


Figure 2-3 Mean read speed per raid levels and test read size cross graphic

Backup Performance Test Directory Traverse Method will allow us to choose that, how the written data will be read. If BPTDTM will be “depth”, than the written data will be read vertically, but if BPTDTM will be “breadth”, than the written data will be read horizontally. These structures also will affect the read performance differently for all raid levels (See figure 2-4)

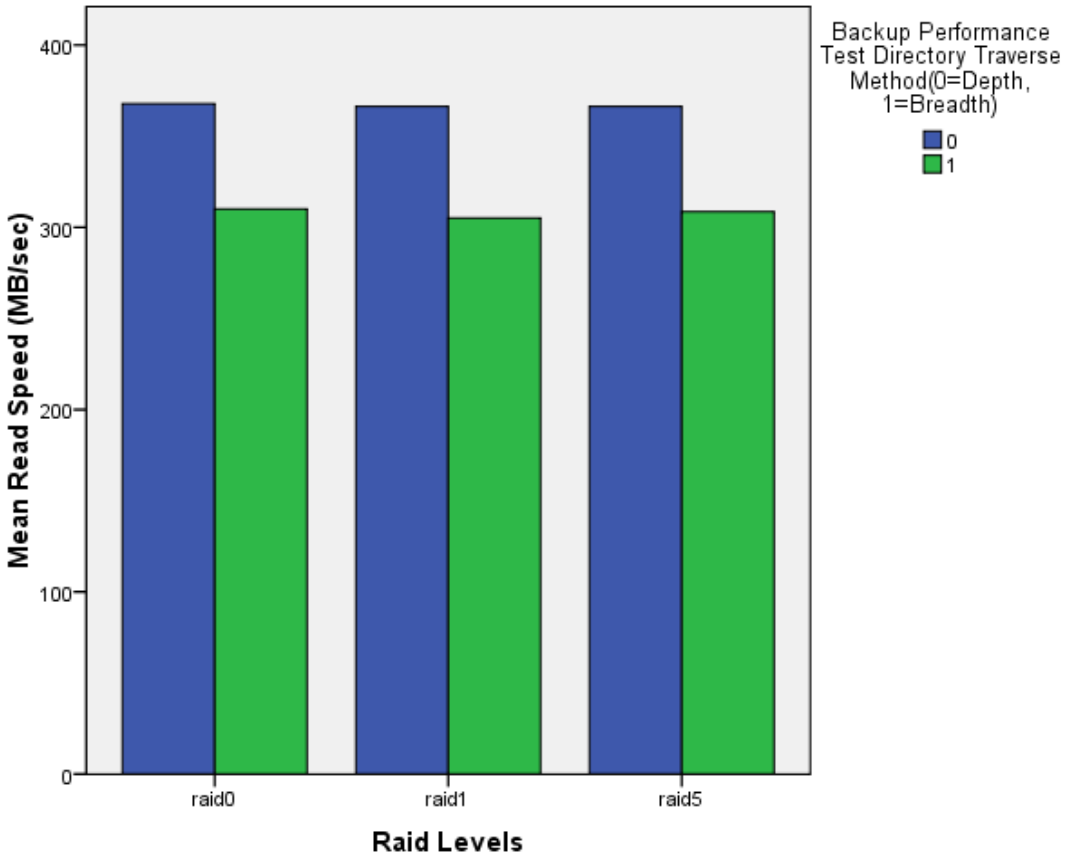


Figure 2-4 Mean read speed per raid levels and BPTDT method cross graphic

2.2 One-Way ANOVA Test

The One-Way ANOVA procedure produces a one-way analysis of variance for a quantitative dependent variable by a single factor (independent) variable. Analysis of variance is used to test the hypothesis that several means are equal. This technique is an extension of the two-sample t test.

In addition to determining that differences exist among the means, you may want to know which means differ. There are two types of tests for comparing means: a priori contrasts and post hoc tests. Contrasts are tests set up before running the experiment and post hoc tests are run after the experiment has been conducted. You can also test for trends across categories.

For each group: number of cases, mean, standard deviation, standard error of the mean, minimum, maximum, and 95% confidence interval for the mean. Levene's test for homogeneity of variance, analysis-of-variance table and robust tests of the equality of means for each dependent variable, user-specified a priori contrasts, and post hoc range tests and multiple comparisons: Bonferroni, Sidak, Tukey's honestly significant difference, Hochberg's GT2, Gabriel, Dunnett, Ryan-Einot-Gabriel-Welsch F test (R-E-G-W F), Ryan-Einot-Gabriel-Welsch range test (R-E-G-W Q), Tamhane's T2, Dunnett's T3, Games-Howell, Dunnett's C, Duncan's multiple range test, Student-Newman-Keuls (S-N-K), Tukey's b, Waller-Duncan, Scheffé, and least-significant difference.

2.2.1 One-Way ANOVA Options

Descriptive: Calculates the number of cases, mean, standard deviation, standard error of the mean, minimum, maximum, and 95% confidence intervals for each dependent variable for each group.

2.2.1.1 Fixed and random effects

Displays the standard deviation, standard error and 95% confidence interval for the fixed-effects model and the standard error, 95% confidence interval and estimate of between-components variance for the random-effects model.

2.2.1.2 Homogeneity of variance test

Calculates the Levene statistic to test for the equality of group variances. This test is not dependent on the assumption of normality.

2.2.1.3 Brown-Forsythe

Calculates the Brown-Forsythe statistic to test for the equality of group means. This statistic is preferable to the F statistic when the assumption of equal variances does not hold.

2.2.1.4 Welch

Calculates the Welch statistic to test for the equality of group means. This statistic is preferable to the F statistic when the assumption of equal variances does not hold.

2.2.1.5 Means plot

Displays a chart that plots the subgroup means (the means for each group defined by values of the factor variable).

2.2.1.6 Missing Values

Controls the treatment of missing values.

2.2.1.7 Exclude cases analysis by analysis

A case with a missing value for either the dependent or the factor variable for a given analysis is not used in that analysis. Also, a case outside the range specified for the factor variable is not used.

2.2.1.8 Exclude cases listwise

Cases with missing values for the factor variable or for any dependent variable included on the dependent list in the main dialog box are excluded from all analyses. If you have not specified multiple dependent variables, this has no effect.

2.2.2 Applying ANOVA to our work

First, we see the descriptive statistics for dependent variables write and read speed based on raid levels.

Table 2-8 Descriptive of dependent variables write and read speed based on raid levels

			N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	Between-Component Variance
							Lower Bound	Upper Bound			
Write Speed (MB/sec)	raid0		110	42,00	20,102	1,917	38,20	45,80	2	53	
	raid1		110	24,60	11,376	1,085	22,45	26,75	2	31	
	raid5		110	11,76	5,568	,531	10,71	12,82	1	17	
	Total		330	26,12	18,467	1,017	24,12	28,12	1	53	
	Model	Fixed Effects			13,717	,755	24,64	27,61			
		Random Effects				8,762	-11,58	63,82			228,584
Read Speed (MB/sec)	raid0		110	338,95	580,922	55,389	229,17	448,72	1	2047	
	raid1		110	335,73	578,360	55,144	226,43	445,02	1	2054	
	raid5		110	337,48	580,369	55,336	227,81	447,16	1	2049	
	Total		330	337,38	578,121	31,825	274,78	399,99	1	2054	
	Model	Fixed Effects			579,885	31,922	274,59	400,18			
		Random Effects				31,922[a]	200,04[a]	474,73[a]			-3054,370
a. Warning: Between-component variance is negative. It was replaced by 0.0 in computing this random effects measure.											

ANOVA descriptive shows us that dependent variable mean write speeds differ based on raid level factor but dependent variable mean read speeds don't differs based on raid levels factor. (See table 2-8). Next we see the results of the Levene's Test of Homogeneity of Variance (See table 2-9).

Table 2-9 Test of Homogeneity of Variances

	Levene Statistic	df1	df2	Sig.
Write Speed (MB/sec)	55,711	2	327	,000
Read Speed (MB/sec)	,002	2	327	,998

This tells us if we have met our second assumption (the groups have approximately equal variance on the dependent variable). If the Levene's Test is significant (the value under "Sig." is less than .05), the two variances are significantly different. If it is not significant (Sig. is greater than .05), the two variances are not significantly different; that is, the two variances are approximately equal. If the Levene's test is not significant, we have met our second assumption. Here, we see that the significance is .998, which is greater than .05. We can assume that the variances are approximately equal for read speed. We have met our second assumption for read speed. If we look the sig value for write speed we will see that the value is less than .05. This means the variances are different for that dependent value.

Finally, we see the results of our One-Way ANOVA (See table 2-10).

Table 2-10 Dependent variables write and read speeds ANOVA test based on raid levels factor

		Sum of Squares	df	Mean Square	F	Sig.
Write Speed (MB/sec)	Between Groups	50664,897	2	25332,448	134,628	,000
	Within Groups	61530,255	327	188,166		
	Total	112195,152	329			
Read Speed (MB/sec)	Between Groups	571,170	2	285,585	,001	,999
	Within Groups	1,100E8	327	336266,290		
	Total	1,100E8	329			

For dependent variable write speed the significance value of the F test in the ANOVA table is .000 (<.05). Thus, you must reject the hypothesis that average assessment scores are equal across raid levels. Despite that, for dependent variable read speed the significance value of the F test in the ANOVA table is 0.999. Thus, you must accept the hypothesis that average assessment scores are equal across raid levels (See table 2-10). Now that you know that the write speed differ based on raid levels in some way, you need to learn more about the structure of the differences.

Table 2-11 Robust Tests of Equality of Means

		Statistic[a]	df1	df2	Sig.
Write Speed (MB/sec)	Welch	154,661	2	180,130	,000
	Brown-Forsythe	134,628	2	191,914	,000
Read Speed (MB/sec)	Welch	,001	2	217,999	,999
	Brown-Forsythe	,001	2	326,995	,999
a. Asymptotically F distributed.					

The significance value of these are both <.05 for write speed value, so we still reject the null hypothesis for write speed tests (See table 2-11). However, this result does not tell us which raid levels are responsible for the difference, so we need the post hoc test results (See table 2-12)

2.2.2.1 One-Way ANOVA Post Hoc Tests

Once you have determined that differences exist among the means, post hoc range tests and pair wise multiple comparisons can determine which means differ. Range tests identify homogeneous subsets of means that are not different from each other. Pair wise multiple comparisons test the difference between each pair of means and yield a matrix where asterisks indicate significantly different group means at an alpha level of 0.05. (See table 2-12).

Tukey's honestly significant difference test, Hochberg's GT2, Gabriel, and Scheffé are

multiple comparison tests and range tests. Other available range tests are Tukey's b, S-N-K (Student-Newman-Keuls), Duncan, R-E-G-W F (Ryan-Einot-Gabriel-Welsch F test), R-E-G-W Q (Ryan-Einot-Gabriel-Welsch range test), and Waller-Duncan. Available multiple comparison tests are Bonferroni, Tukey's honestly significant difference test, Sidak, Gabriel, Hochberg, Dunnett, Scheffé, and LSD.

Table 2-12 Pair wise multiple comparisons test the difference between each pair of means

Dependent Variable		(I) Raid Levels	(J) Raid Levels	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
							Lower Bound	Upper Bound
Write Speed (MB/sec)	Tukey HSD	raid 0	raid 1	17,400*	1,850	,000	13,05	21,75
			raid 5	30,236*	1,850	,000	25,88	34,59
		raid 1	raid 0	-17,400*	1,850	,000	-21,75	-13,05
			raid 5	12,836*	1,850	,000	8,48	17,19
		raid 5	raid 0	-30,236*	1,850	,000	-34,59	-25,88
			raid 1	-12,836*	1,850	,000	-17,19	-8,48
	Scheffe	raid 0	raid 1	17,400*	1,850	,000	12,85	21,95
			raid 5	30,236*	1,850	,000	25,69	34,78
		raid 1	raid 0	-17,400*	1,850	,000	-21,95	-12,85
			raid 5	12,836*	1,850	,000	8,29	17,38
		raid 5	raid 0	-30,236*	1,850	,000	-34,78	-25,69
			raid 1	-12,836*	1,850	,000	-17,38	-8,29

Read Speed (MB/sec)	Tukey HSD	raid 0	raid 1	3,218	78,192	,999	-180,88	187,32
			raid 5	1,464	78,192	1,000	-182,63	185,56
		raid 1	raid 0	-3,218	78,192	,999	-187,32	180,88
			raid 5	-1,755	78,192	1,000	-185,85	182,34
		raid 5	raid 0	-1,464	78,192	1,000	-185,56	182,63
			raid 1	1,755	78,192	1,000	-182,34	185,85
	Scheffe	raid 0	raid 1	3,218	78,192	,999	-189,06	195,49
			raid 5	1,464	78,192	1,000	-190,81	193,74
		raid 1	raid 0	-3,218	78,192	,999	-195,49	189,06
			raid 5	-1,755	78,192	1,000	-194,03	190,52
		raid 5	raid 0	-1,464	78,192	1,000	-193,74	190,81
			raid 1	1,755	78,192	1,000	-190,52	194,03
*. The mean difference is significant at the 0.05 level.								

Tukey and Scheffe tests can also detect homogeneity subsets. In our work Tukey and Scheffe shows us that there is no homogeneity for Write speed tests. We can see that there are 3 subsets. (See table 2-13)

Table 2-13 Homogeneous Subsets Write Speed (MB/sec)

	Raid Levels	N	Subset for alpha = 0.05		
			1	2	3
Tukey HSD[a]	raid5	110	11,76		
	raid1	110		24,60	
	raid0	110			42,00
	Sig.		1,000	1,000	1,000
Scheffe[a]	raid5	110	11,76		
	raid1	110		24,60	
	raid0	110			42,00
	Sig.		1,000	1,000	1,000
Means for groups in homogeneous subsets are displayed.					
a. Uses Harmonic Mean Sample Size = 110,000. b.					

Despite that, we can see only 1 subset for read speed. That shows us there is homogeneity for read speed tests. (See table 2-14)

Table 2-14 Homogeneous Subsets Read Speed (MB/sec)

	Raid Levels	N	Subset for alpha = 0.05
			1
Tukey HSD[a]	raid1	110	335,73
	raid5	110	337,48
	raid0	110	338,95
	Sig.		,999
Scheffe[a]	raid1	110	335,73
	raid5	110	337,48
	raid0	110	338,95
	Sig.		,999
Means for groups in homogeneous subsets are displayed.			
a. Uses Harmonic Mean Sample Size = 110,000.			

The means plot helps you to "see" this structure. Mean of write speed significantly differs between the 3 raid levels. The difference between the raid0 and raid5 is more highly than the difference between raid0 and raid1 or raid 1 and raid5. (See figure 2-5)

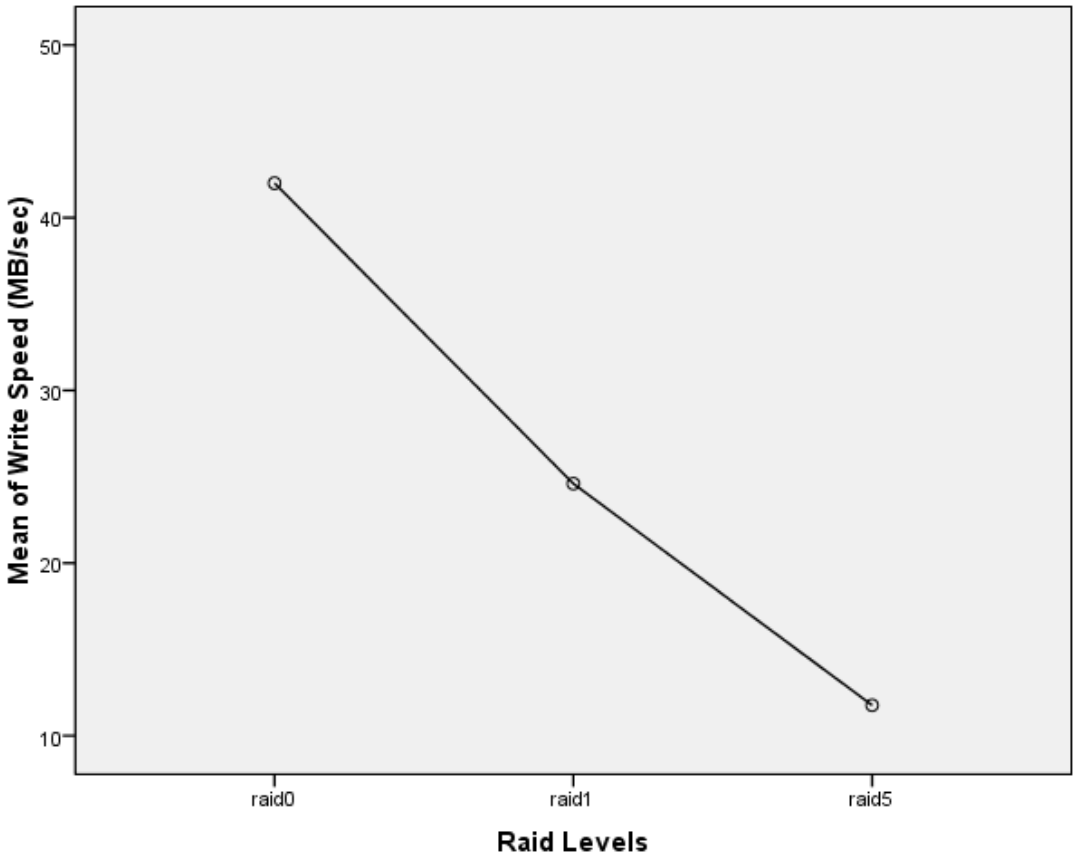


Figure 2-5 Mean plot figure for depended variable write speed based on raid levels factor

Mean of read speed doesn't significantly differ between the 3 raid levels. The difference between the raid0 and raid1 is more highly than the difference between raid 1 and raid5. (See figure 2-6). Still this difference is not statistical significant.

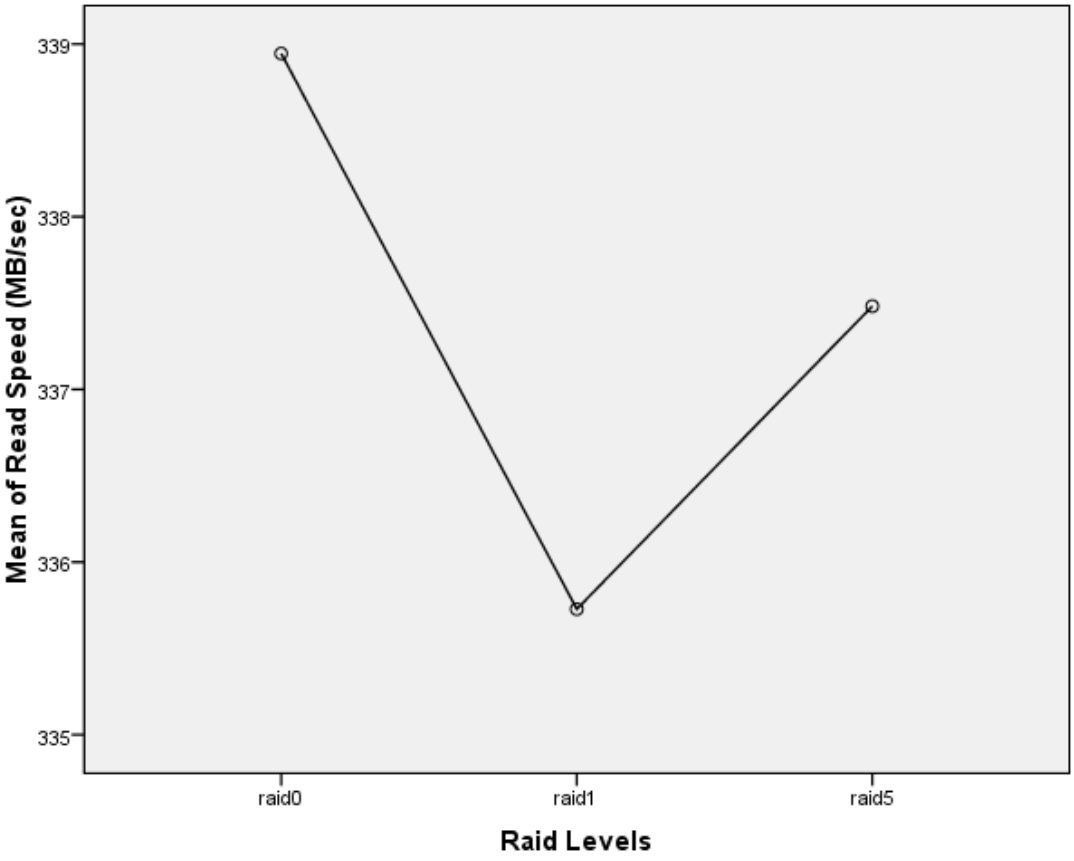


Figure 2-6 Mean plot figure for depended variable read speed based on raid levels factor

3 Discussion and Findings

This chapter contains discussions and findings about our work.

After these entire tests, we have some finding about performance result. These results will help us to identify that which raid level is suitable for which applications.

This chapter addresses the following topics:

“Write Speed Findings and Discussions”

“Read Speed Findings and Discussions”

3.1 Write Speed Findings and Discussions

Write speed is a value that can be affected by raid level, stripe size; used data size and file structure (see-appendix A for all results). Best write performance values were achieved with Raid level 0 (See figure 3-1).

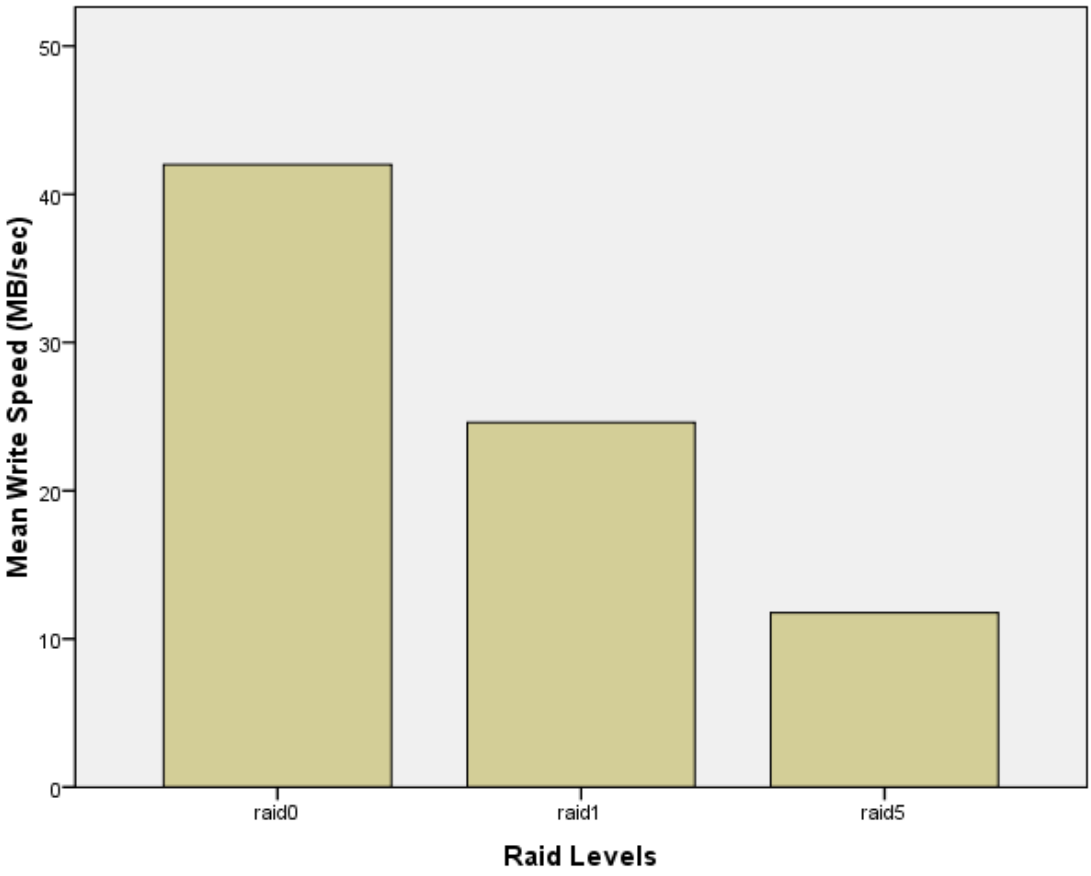


Figure 3-1 Mean write speed per raid levels

Raid 0 has the power of striping. Because of the data striped and distributed on different HDDs without any algorithm, data can be written with best performance. Striping data on multiple disks provides us multiple I/O buses. More I/O buses mean more performance.

Raid 1 has no advantage comparing to Raid 0 with same disk count. Raid 1 also has multiple I/O buses but data written without striping. Same data will be written to 2 disk blocks. If we increase the disk count to 4 disks, write speed can be identical to Raid 0.

Raid 5 has the worst write performance comparing to Raid 0 and Raid 1. This is because of Raid 5's parity data formula to create fault tolerance. Every single data will be processed by hardware based on parity data formula before writing to logical drive. (Jose Luis Gonzalez, Toni Cortes, 2004).

RPTFTD and RPTFTB parameters were highly affected write speed. When these values increased, after a complexity level at test data's folder structure, write speed performance decreased dramatically. (See figure 3-2)

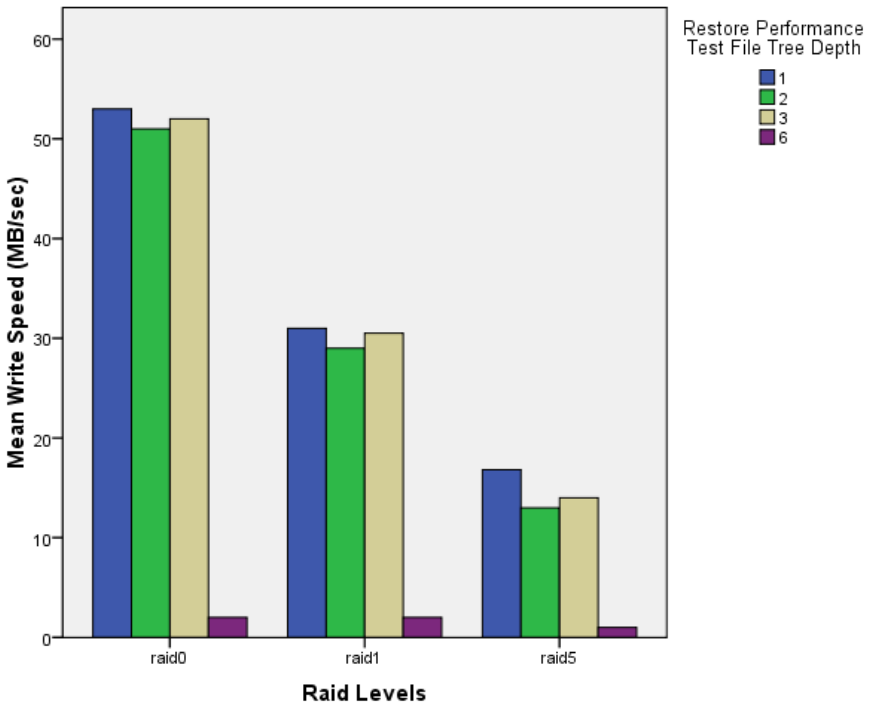


Figure 3-2 Mean write speed per raid levels and RPTFTD cross graphic

If we consider mean write speed on different raid levels, we will notice that RPTFTD and RPTFTB parameters' effect will be similar to all raid levels (See figure 3-3). It means that, if complexity of data and folder structure increases, the write performance decreases.

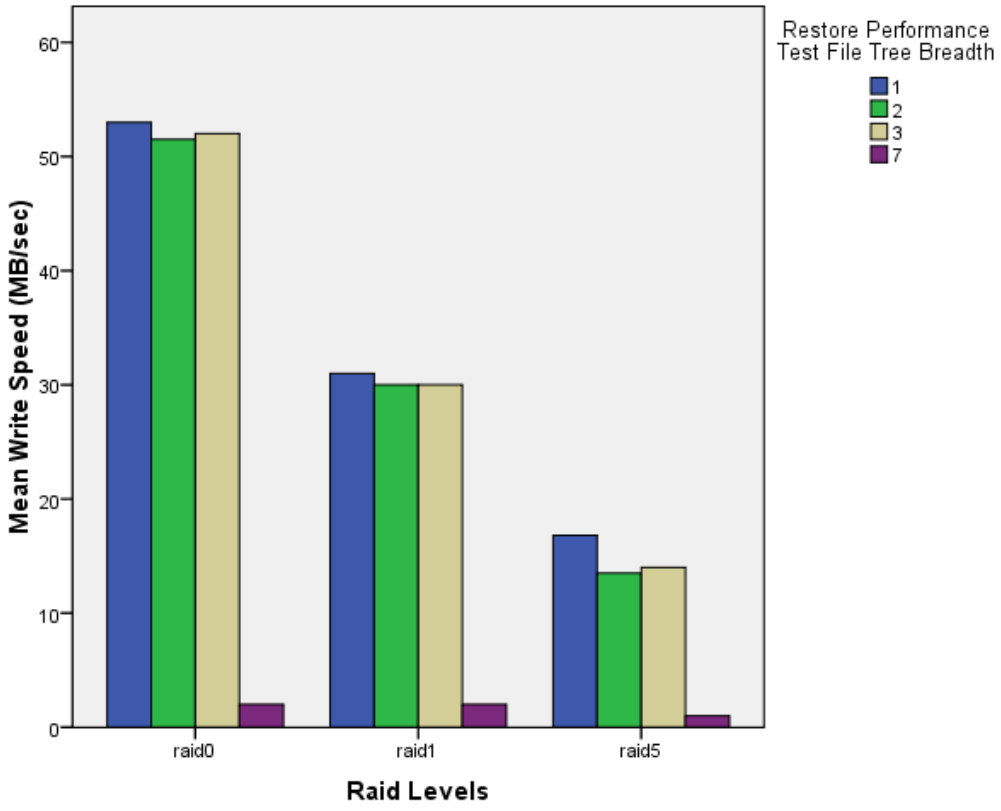


Figure 3-3 Mean write speed per raid levels and RPTFTB cross graphic

3.2 Read Speed Findings

Read speed is a value that can be affected by raid level, stripe size; used data size, file structure and traverse method (see-appendix A for all results). Best read performance values was achieved with Raid level 0 (See figure 3-4).

Raid 0 has the power of striping. Because of the data striped and distributed on different HDDs without any algorithm, data can be read with best performance. Striping data on multiple disks provides us multiple I/O buses. More I/O buses mean more performance.

Raid 1 achieved an approximate read speed value with Raid 0. This can be resulted because both levels have multiple I/O buses because of multiple disks.

Raid 5 achieved an approximate read speed value with Raid 0 and Raid 1. Raid 5 has more I/O buses compared or others when minimum disk requirements satisfied but despite that Raid 5 decodes data to read from logical drive because of parity data formula to create fault tolerance. (Jose Luis Gonzalez & Toni Cortes, 2004).

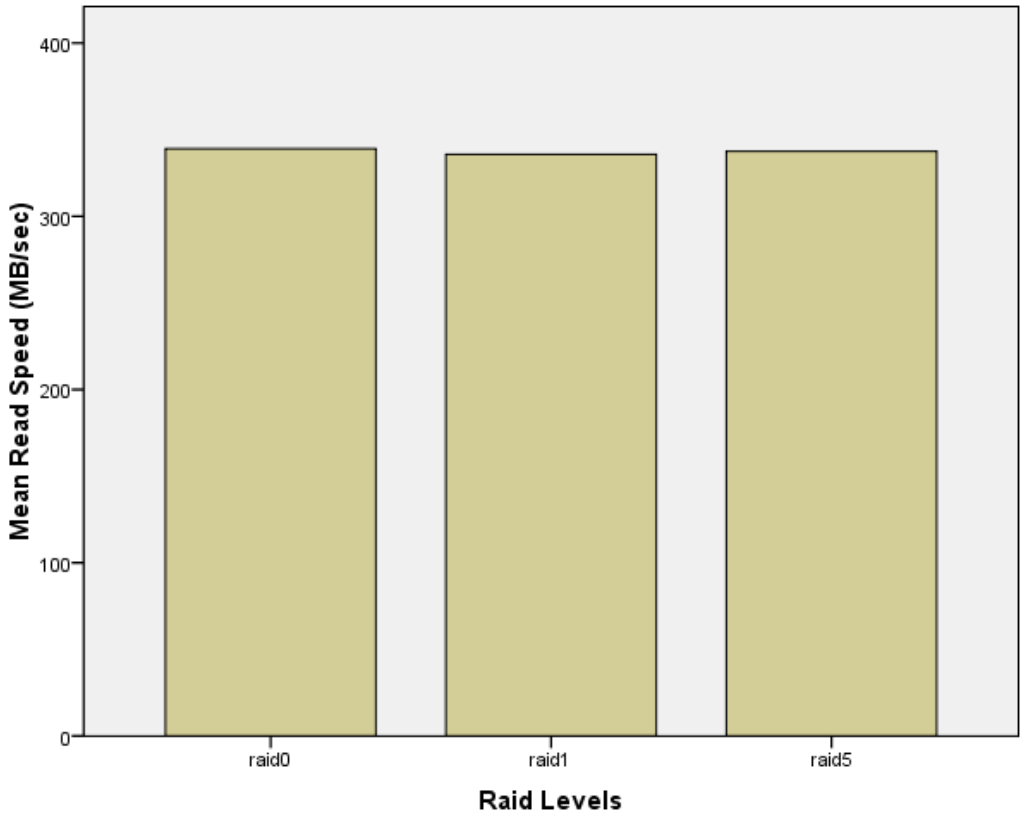


Figure 3-4 Mean read speed per raid levels

RPTFTD and RPTFTB parameters were highly affected read speed. When these values increased, after a complexity level at test data's folder structure, read speed performance decreased dramatically. (See figure 3-5)

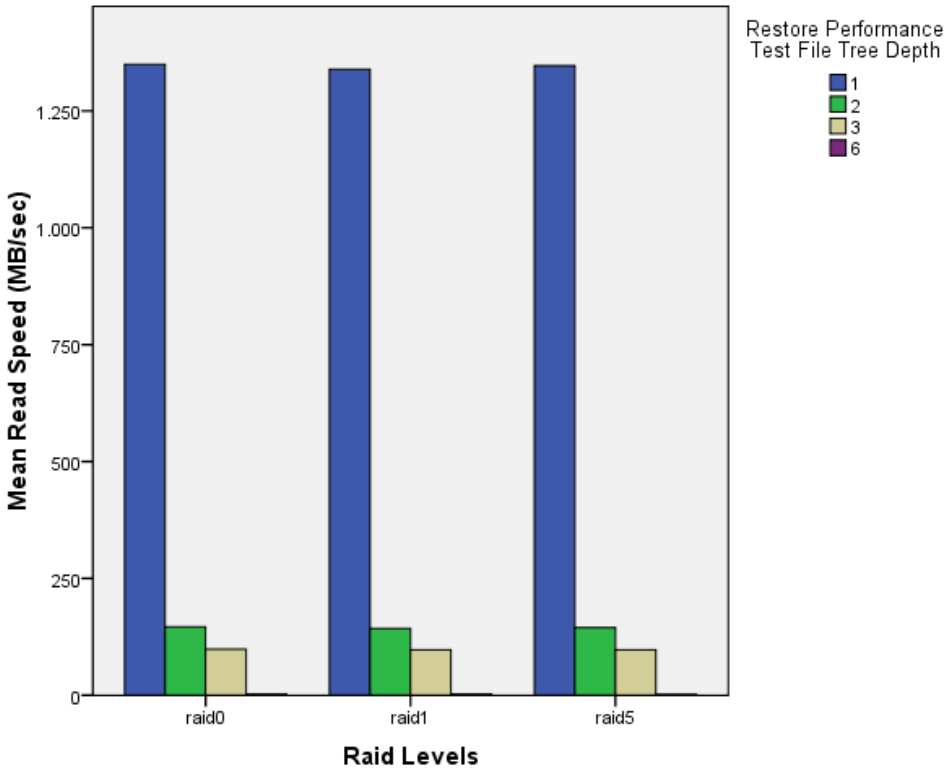


Figure 3-5 Mean Read speed per raid levels and RPTFTD cross graphic

4 Conculusion

The disk transfers per second decreased when additional load was placed on the controller. This performance curve is an indication that the controller throughput is saturated and optimal performance was achieved with less load. (Hp Tests, 2005)

To create an optimal cost-effective RAID configuration, we need to achieve the following goals:

- Maximize the number of disks being accessed in parallel.
- Minimize the amount of disk space being used for redundant data.
- Minimize the overhead required to achieve the above goals.

Raid Level selection explains as following by (WRL, 2005)

To choose the RAID level that's right for you, begin by considering the factors below. Each one of these factors becomes a trade-off for another:

- Cost of disk storage
- Data protection or data availability required (low, medium, high)
- Performance requirements (low, medium, high)

And explain as following by Consensy (1997)

Estimating the theoretical performance of the different RAID types is fairly easy. We rate a single standard disk as 1. Five striped disks have a theoretical performance factor of 5 in reads and writes. Six RAID 5 disks have a read performance factor of 5, but suffer a write performance penalty that brings their overall performance down to about 3. Two mirrored disks have a read performance factor of 2, and a write performance factor of 1, for an overall performance of about 1.8.

For our tests, Raid 0 and Raid 5 showed superior read performance. Raid 1 read performance was also very high, but it was slightly lower than the others. Raid 0 was the winner of write speed performance with a significant difference. Raid 1 was the follower with middle write performance and Raid 5 performed worst. These deductions observed when only minimum number of required disks satisfied.

That shows us that we should implement our structure with Raid 0 if no data protection

but high performance needed. So we can say that using Raid 0 will be suitable for following structures; high end workstations, data logging, real-time rendering and very transitory data (See table 5-1).

We conclude on that we should implement our structure with Raid 1 if data protection and availability is so important. So we can say that using Raid 1 will be suitable for following structures; operating system, transaction databases (See table 5-1).

As you can guess implementing our structure with Raid 5 is suitable if our budget is larger and we need both performance and redundancy on data reading. There is a write penalty for Raid 5, so, we shouldn't use this level for any structure that requires often write jobs. So we can say that using Raid 5 will be suitable for following structures; data warehousing, web serving, archiving (See table 5-1).

Table 5-1 Different raid levesl feature deductions

Features	RAID 0	RAID 1	RAID 5
Minimum Number of Drives	1	2	3
Data Protection	No Protection	Single Drive Failure	Single Drive Failure
Read Performance	Superior	Very High	Superior
Write Performance	Superior	Medium	Low
Read Performance (degreaded)	N/A	Medium	Low
Write Performance (degreaded)	N/A	High	Low
Capacity Utilization	100%	50%	67% - 94%
Typical Applications	High end workstations, data logging, real-time rendering, very transitory data	Operating system, transaction databases	Data warehousing, web serving, archiving

Glossary

Availability Is how well a system can work in times of a failure. If a system is able to work even in the presence of a failure of one or more system components, the system is said to be available.

Array A set of physical disks configured into one or more logical drives. Arrayed disks have significant performance and data protection advantages over non-arrayed disks.

Array capacity expansion See capacity expansion.

Auto-Reliability Monitoring (ARM) Also known as surface analysis. A fault management feature that scans physical disks for bad sectors. Data in the faulty sectors remaps onto good sectors. Also checks parity data consistency for disks in RAID 5 or RAID ADG configurations. Operates as a background process.

Automatic Data Recovery A process that automatically reconstructs data from a failed disk and writes it onto a replacement disk. Automatic Data Recovery time depends on several factors. Also known as rebuild.

Cache A high-speed memory component, used to store data temporarily for rapid access.

Capacity expansion The addition of physical disks to an existing disk array, and redistribution of existing logical drives and data over the enlarged array. The size of the logical drives does not change. Also known as array capacity expansion.

Capacity extension The enlargement of a logical drive without disruption of data. There must be free space on the array before capacity extension can occur. If necessary, create free space by deleting a logical drive or by carrying out a capacity expansion. Also known as logical drive capacity extension.

Data guarding See RAID.

Data striping Writing data to logical drives in interleaved chunks (by byte or by sector). Data triping improves system performance by distributing data evenly across all physical disks in the array, but has no fault tolerance

Drive mirroring Duplicating data from one disk onto a second disk. Mirroring provides fault tolerance, but can only recover from failure of one physical disk per mirrored pair.

Error Correction and Checking (ECC) memory A type of memory that checks and corrects single-bit or multi-bit memory errors (depending on configuration) without causing the server to halt or to corrupt data.

Fault tolerance The ability of a server to recover from physical disk hardware problems without interrupting server performance or corrupting data. Hardware RAID is most commonly used, but there are other types of fault tolerance, including controller duplexing and software-based RAID.

Hot spare See online spare.

Interim data recovery If a disk fails in RAID 1, 1+0, 5 or ADG, the system still processes I/O requests, but at a reduced performance level.

Logical drive A group of physical disks, or part of a group, that behaves as one storage unit. Each constituent physical disk contributes the same storage volume to the total volume of the logical drive. A logical drive has performance advantages over individual physical disks. Also known as a logical volume.

Logical drive capacity extension See capacity extension.

Online spare A fault-tolerant system that normally contains no data. When any other disk in the array fails, the controller automatically rebuilds the data that was on the failed disk onto the online spare. Also known as a hot spare.

Physical disk A random-access storage device. In traditional non-arrayed storage, one physical disk typically contains a single logical drive. In RAID configurations, multiple disks are combined to form a single logical drive.

RAID **R**edundant **A**rray of **I**nexpensive **D**isks

Rebuild See Automatic Data Recovery.

Redundant Array of Independent Disks (RAID) A form of fault-tolerant storage control. See "Introduction to RAID Technology"

Reliability Is how well a system can work without any failures in its components. If there is a failure, the system was not reliable.

Rotational latency Amount of time needed for the desired sector to rotate under the disk head.

Spare See online spare.

Striping See data striping.

Surface analysis See Auto-Reliability Monitoring.

Seek time Amount of time needed to move the head to the correct radial position of the disk.

References

1- Bic, Lubomur F. & Shaw, Alan C. (2003) - Operating Systems
2- Consensy (1997) - http://www.consensys.com/html/rzma1_raid_disk_array__overvie.html
3- David C. Stallmo & Randy K. Hall (1997) - Method and apparatus for improving performance in a reduntant array of independendt disks
4- Fujitsu (2010) - http://storage-system.fujitsu.com/global/services/system/glossary/raid/raid5/
5- Hp Tests (2005) - HP SAS benchmark performance tests, TPC Benchmark™H (TPC-H) test published by HP in August 2005 for the HP ProLiant DL585 4P Opteron Dual Core server.
6- Hp-WiR (2007) – RAID Technology Overview - HP Smart Array RAID Controllers, HP Part Number: J6369-90050, Published: September 2007, Hewlett-Packard Development Company L.P. http://docs.hp.com/en/J6369-90026/ch01.html
7- IBM (2006) - IBM Systems Software Information Center http://publib.boulder.ibm.com/infocenter/eserver/v1r2/index.jsp?topic=/diricinfo/fqy0_cselraid.html
8- Inmon, W.H. (1995) - Tech Topic: What is a Data Warehouse? Prism Solutions. Volume 1.
9- Jose Luis Gonzalez & Toni Cortes (2004) - Increasing the capacity of RAID5 by online gradual assimilation
10- Ling Liu & Tamer M. Özsu (2009) - Encyclopedia of Database Systems
11- McBee, Jim & Barry Gerber (2007) - Mastering Microsoft Exchange Server 2007
12- Möller, Tomas & Eric Haines (1999) – Real-Time Rendering. 1st ed. Natick, MA: A K Peters, Ltd.
13- Philip A. Bernstein & Eric Newcomer (2009) - Principles of Transaction Processing
14- RaidRMS (2009) - RAID-RMS: A fault tolerant stripped mirroring RAID architecture for distributed systems
Javad Akbari Torkestani & Mohammad Reza Meybodi Computer Engineering Department, Islamic Azad University, Arak, Iran
Computer Engineering Department, Amirkabir University of Technology, Tehran, Iran Institute for Studies an Theoretical Physics and Mathematics (IPM), School of Computer Science, Tehran, Iran
15- Richard G. Krum & Virat Thantrakul (1998) - High density redundant array of independent disks in a chassis having a door with shock absorbers held against the disks when the door is closed
16- Riva, Marco & Piergiovanni, Schiraldi (2001) – "Performances of time-temperature indicators in the study of temperature exposure of packaged fresh foods", Packaging Technology and Science 14 (1): 1–39

17- S. Savage & J. Wilkes (1996) - AFRAID—a frequently redundant array of independent disks
18- Scott Baderman (2003) - System and method for handling temporary errors on a redundant array of independendt disks
19- SelectingRAID (2002) - Selecting RAID levels for disk arrays, Conference on File and Storage Technologies (FAST'02), pp. 189–201, 28–30 January 2002, Monterey, CA. (USENIX, Berkeley, CA.)
20- Stewart, C & Kowaltzke, A (1997) - Media: New Ways and Meanings (second edition), JACARANDa, Milton, Sydney. pp.102.)
21- UC Berkeley (1987) – Original study http://techreports.lib.berkeley.edu/accessPages/CSD-87-391.html
22- Umass (2008) - http://www.ecs.umass.edu/ece/koren/architecture/Raid/cp.html
23- Walch, Victoria Irons (2006) - Archival Census and Education Needs Survey in the United States
24- (Wikipedia-Workstation) - Workstation from Wikipedia- http://en.wikipedia.org/wiki/Workstation
25- (Wikipedia-WS) – Web Server from Wikipedia- http://en.wikipedia.org/wiki/Webserver
26- WRL (2005) - Which RAID Level is Right for Me?, Adaptec, Inc. , P/N: 666849-011 http://www.adaptec.com/NR/rdonlyres/874D145E-F64F-4804-9E27-037BC5A9DCE0/0/3994_RAID_WhichOne_v112.pdf

Appendix A: Tests Parameters' Values and Tests Results

Raid Level	Stripe Size	Physical Drives Attached	Test sizes (K-MB)	RPTFTD	RPTFTB	Write Speed	BPTRS	BPTDTM	Read Speed
0	128	2	4-128	1	1	53	1	DEPTH	409
0	128	2	4-128	1	1	53	1	BREADTH	511
0	128	2	4-128	1	1	53	2	DEPTH	682
0	128	2	4-128	1	1	53	2	BREADTH	282
0	128	2	4-128	1	1	53	4	DEPTH	1023
0	128	2	4-128	1	1	53	4	BREADTH	1023
0	128	2	4-128	1	1	53	8	DEPTH	2047
0	128	2	4-128	1	1	53	8	BREADTH	2047
0	128	2	4-128	1	1	53	16	DEPTH	2047
0	128	2	4-128	1	1	53	16	BREADTH	1023
0	128	2	4-128	1	1	53	32	DEPTH	1023
0	128	2	4-128	1	1	53	32	BREADTH	1203
0	128	2	4-128	1	1	53	64	DEPTH	2047
0	128	2	4-128	1	1	53	64	BREADTH	2047
0	128	2	4-128	1	1	53	128	DEPTH	2047
0	128	2	4-128	1	1	53	128	BREADTH	1023
0	128	2	4-128	1	1	53	256	DEPTH	2047
0	128	2	4-128	1	1	53	256	BREADTH	2047
0	128	2	4-128	1	1	53	512	DEPTH	1023
0	128	2	4-128	1	1	53	512	BREADTH	1023
0	128	2	4-128	1	1	53	1024	DEPTH	2047

0	128	2	4-128	1	1	53	1024	BREADTH	1023
0	128	2	4-128	2	2	51	1	DEPTH	139
0	128	2	4-128	2	2	51	1	BREADTH	146
0	128	2	4-128	2	2	51	2	DEPTH	149
0	128	2	4-128	2	2	51	2	BREADTH	149
0	128	2	4-128	2	2	51	4	DEPTH	157
0	128	2	4-128	2	2	51	4	BREADTH	153
0	128	2	4-128	2	2	51	8	DEPTH	157
0	128	2	4-128	2	2	51	8	BREADTH	157
0	128	2	4-128	2	2	51	16	DEPTH	157
0	128	2	4-128	2	2	51	16	BREADTH	153
0	128	2	4-128	2	2	51	32	DEPTH	153
0	128	2	4-128	2	2	51	32	BREADTH	153
0	128	2	4-128	2	2	51	64	DEPTH	157
0	128	2	4-128	2	2	51	64	BREADTH	161
0	128	2	4-128	2	2	51	128	DEPTH	161.
0	128	2	4-128	2	2	51	128	BREADTH	157
0	128	2	4-128	2	2	51	256	DEPTH	149
0	128	2	4-128	2	2	51	256	BREADTH	149
0	128	2	4-128	2	2	51	512	DEPTH	146
0	128	2	4-128	2	2	51	512	BREADTH	149
0	128	2	4-128	2	2	51	1024	DEPTH	87
0	128	2	4-128	2	2	51	1024	BREADTH	84
0	128	2	4-128	3	2	52	1	DEPTH	103
0	128	2	4-128	3	2	52	1	BREADTH	106

0	128	2	4-128	3	2	52	2	DEPTH	104
0	128	2	4-128	3	2	52	2	BREADTH	107
0	128	2	4-128	3	2	52	4	DEPTH	106
0	128	2	4-128	3	2	52	4	BREADTH	107
0	128	2	4-128	3	2	52	8	DEPTH	106
0	128	2	4-128	3	2	52	8	BREADTH	108
0	128	2	4-128	3	2	52	16	DEPTH	107
0	128	2	4-128	3	2	52	16	BREADTH	109
0	128	2	4-128	3	2	52	32	DEPTH	109
0	128	2	4-128	3	2	52	32	BREADTH	107
0	128	2	4-128	3	2	52	64	DEPTH	107
0	128	2	4-128	3	2	52	64	BREADTH	109
0	128	2	4-128	3	2	52	128	DEPTH	107
0	128	2	4-128	3	2	52	128	BREADTH	108
0	128	2	4-128	3	2	52	256	DEPTH	106
0	128	2	4-128	3	2	52	256	BREADTH	107
0	128	2	4-128	3	2	52	512	DEPTH	106
0	128	2	4-128	3	2	52	512	BREADTH	107
0	128	2	4-128	3	2	52	1024	DEPTH	51
0	128	2	4-128	3	2	52	1024	BREADTH	53
0	128	2	4-128	3	3	52	1	DEPTH	96
0	128	2	4-128	3	3	52	1	BREADTH	100
0	128	2	4-128	3	3	52	2	DEPTH	100
0	128	2	4-128	3	3	52	2	BREADTH	98
0	128	2	4-128	3	3	52	4	DEPTH	101

0	128	2	4-128	3	3	52	4	BREADTH	98
0	128	2	4-128	3	3	52	8	DEPTH	98
0	128	2	4-128	3	3	52	8	BREADTH	97
0	128	2	4-128	3	3	52	16	DEPTH	101
0	128	2	4-128	3	3	52	16	BREADTH	97
0	128	2	4-128	3	3	52	32	DEPTH	101
0	128	2	4-128	3	3	52	32	BREADTH	101
0	128	2	4-128	3	3	52	64	DEPTH	101
0	128	2	4-128	3	3	52	64	BREADTH	102
0	128	2	4-128	3	3	52	128	DEPTH	101
0	128	2	4-128	3	3	52	128	BREADTH	98
0	128	2	4-128	3	3	52	256	DEPTH	100
0	128	2	4-128	3	3	52	256	BREADTH	99
0	128	2	4-128	3	3	52	512	DEPTH	99
0	128	2	4-128	3	3	52	512	BREADTH	100
0	128	2	4-128	3	3	52	1024	DEPTH	48
0	128	2	4-128	3	3	52	1024	BREADTH	51
0	128	2	4	6	7	2	1	DEPTH	1
0	128	2	4	6	7	2	1	BREADTH	1
0	128	2	4	6	7	2	2	DEPTH	1
0	128	2	4	6	7	2	2	BREADTH	1
0	128	2	4	6	7	2	4	DEPTH	2
0	128	2	4	6	7	2	4	BREADTH	2
0	128	2	4	6	7	2	8	DEPTH	2
0	128	2	4	6	7	2	8	BREADTH	2

0	128	2	4	6	7	2	16	DEPTH	2
0	128	2	4	6	7	2	16	BREADTH	2
0	128	2	4	6	7	2	32	DEPTH	2
0	128	2	4	6	7	2	32	BREADTH	2
0	128	2	4	6	7	2	64	DEPTH	2
0	128	2	4	6	7	2	64	BREADTH	2
0	128	2	4	6	7	2	128	DEPTH	2
0	128	2	4	6	7	2	128	BREADTH	2
0	128	2	4	6	7	2	256	DEPTH	2
0	128	2	4	6	7	2	256	BREADTH	2
0	128	2	4	6	7	2	512	DEPTH	2
0	128	2	4	6	7	2	512	BREADTH	2
0	128	2	4	6	7	2	1024	DEPTH	2
0	128	2	4	6	7	2	1024	BREADTH	2
1	128	2	4-128	1	1	31	1	DEPTH	399
1	128	2	4-128	1	1	31	1	BREADTH	506
1	128	2	4-128	1	1	31	2	DEPTH	663
1	128	2	4-128	1	1	31	2	BREADTH	282
1	128	2	4-128	1	1	31	4	DEPTH	1023
1	128	2	4-128	1	1	31	4	BREADTH	1001
1	128	2	4-128	1	1	31	8	DEPTH	2054
1	128	2	4-128	1	1	31	8	BREADTH	2016
1	128	2	4-128	1	1	31	16	DEPTH	2045
1	128	2	4-128	1	1	31	16	BREADTH	1002
1	128	2	4-128	1	1	31	32	DEPTH	1026

1	128	2	4-128	1	1	31	32	BREADTH	1100
1	128	2	4-128	1	1	31	64	DEPTH	2046
1	128	2	4-128	1	1	31	64	BREADTH	2046
1	128	2	4-128	1	1	31	128	DEPTH	2046
1	128	2	4-128	1	1	31	128	BREADTH	1019
1	128	2	4-128	1	1	31	256	DEPTH	2046
1	128	2	4-128	1	1	31	256	BREADTH	2046
1	128	2	4-128	1	1	31	512	DEPTH	1019
1	128	2	4-128	1	1	31	512	BREADTH	1016
1	128	2	4-128	1	1	31	1024	DEPTH	2045
1	128	2	4-128	1	1	31	1024	BREADTH	1019
1	128	2	4-128	2	2	29	1	DEPTH	135
1	128	2	4-128	2	2	29	1	BREADTH	146
1	128	2	4-128	2	2	29	2	DEPTH	145
1	128	2	4-128	2	2	29	2	BREADTH	142
1	128	2	4-128	2	2	29	4	DEPTH	153
1	128	2	4-128	2	2	29	4	BREADTH	152
1	128	2	4-128	2	2	29	8	DEPTH	156
1	128	2	4-128	2	2	29	8	BREADTH	156
1	128	2	4-128	2	2	29	16	DEPTH	156
1	128	2	4-128	2	2	29	16	BREADTH	146
1	128	2	4-128	2	2	29	32	DEPTH	150
1	128	2	4-128	2	2	29	32	BREADTH	150
1	128	2	4-128	2	2	29	64	DEPTH	150
1	128	2	4-128	2	2	29	64	BREADTH	156

1	128	2	4-128	2	2	29	128	DEPTH	150
1	128	2	4-128	2	2	29	128	BREADTH	156
1	128	2	4-128	2	2	29	256	DEPTH	146
1	128	2	4-128	2	2	29	256	BREADTH	145
1	128	2	4-128	2	2	29	512	DEPTH	141
1	128	2	4-128	2	2	29	512	BREADTH	143
1	128	2	4-128	2	2	29	1024	DEPTH	90
1	128	2	4-128	2	2	29	1024	BREADTH	81
1	128	2	4-128	3	2	31	1	DEPTH	101
1	128	2	4-128	3	2	31	1	BREADTH	104
1	128	2	4-128	3	2	31	2	DEPTH	103
1	128	2	4-128	3	2	31	2	BREADTH	101
1	128	2	4-128	3	2	31	4	DEPTH	105
1	128	2	4-128	3	2	31	4	BREADTH	107
1	128	2	4-128	3	2	31	8	DEPTH	108
1	128	2	4-128	3	2	31	8	BREADTH	99
1	128	2	4-128	3	2	31	16	DEPTH	102
1	128	2	4-128	3	2	31	16	BREADTH	103
1	128	2	4-128	3	2	31	32	DEPTH	108
1	128	2	4-128	3	2	31	32	BREADTH	109
1	128	2	4-128	3	2	31	64	DEPTH	104
1	128	2	4-128	3	2	31	64	BREADTH	108
1	128	2	4-128	3	2	31	128	DEPTH	107
1	128	2	4-128	3	2	31	128	BREADTH	103
1	128	2	4-128	3	2	31	256	DEPTH	106

1	128	2	4-128	3	2	31	256	BREADTH	108
1	128	2	4-128	3	2	31	512	DEPTH	109
1	128	2	4-128	3	2	31	512	BREADTH	106
1	128	2	4-128	3	2	31	1024	DEPTH	52
1	128	2	4-128	3	2	31	1024	BREADTH	54
1	128	2	4-128	3	3	30	1	DEPTH	97
1	128	2	4-128	3	3	30	1	BREADTH	99
1	128	2	4-128	3	3	30	2	DEPTH	99
1	128	2	4-128	3	3	30	2	BREADTH	97
1	128	2	4-128	3	3	30	4	DEPTH	100
1	128	2	4-128	3	3	30	4	BREADTH	97
1	128	2	4-128	3	3	30	8	DEPTH	97
1	128	2	4-128	3	3	30	8	BREADTH	96
1	128	2	4-128	3	3	30	16	DEPTH	100
1	128	2	4-128	3	3	30	16	BREADTH	96
1	128	2	4-128	3	3	30	32	DEPTH	100
1	128	2	4-128	3	3	30	32	BREADTH	100
1	128	2	4-128	3	3	30	64	DEPTH	100
1	128	2	4-128	3	3	30	64	BREADTH	101
1	128	2	4-128	3	3	30	128	DEPTH	102
1	128	2	4-128	3	3	30	128	BREADTH	99
1	128	2	4-128	3	3	30	256	DEPTH	99
1	128	2	4-128	3	3	30	256	BREADTH	100
1	128	2	4-128	3	3	30	512	DEPTH	98
1	128	2	4-128	3	3	30	512	BREADTH	99

1	128	2	4-128	3	3	30	1024	DEPTH	47
1	128	2	4-128	3	3	30	1024	BREADTH	51
1	128	2	4	6	7	2	1	DEPTH	1
1	128	2	4	6	7	2	1	BREADTH	1
1	128	2	4	6	7	2	2	DEPTH	1
1	128	2	4	6	7	2	2	BREADTH	1
1	128	2	4	6	7	2	4	DEPTH	1
1	128	2	4	6	7	2	4	BREADTH	2
1	128	2	4	6	7	2	8	DEPTH	2
1	128	2	4	6	7	2	8	BREADTH	2
1	128	2	4	6	7	2	16	DEPTH	2
1	128	2	4	6	7	2	16	BREADTH	2
1	128	2	4	6	7	2	32	DEPTH	2
1	128	2	4	6	7	2	32	BREADTH	2
1	128	2	4	6	7	2	64	DEPTH	2
1	128	2	4	6	7	2	64	BREADTH	2
1	128	2	4	6	7	2	128	DEPTH	2
1	128	2	4	6	7	2	128	BREADTH	2
1	128	2	4	6	7	2	256	DEPTH	2
1	128	2	4	6	7	2	256	BREADTH	2
1	128	2	4	6	7	2	512	DEPTH	2
1	128	2	4	6	7	2	512	BREADTH	2
1	128	2	4	6	7	2	1024	DEPTH	2
1	128	2	4	6	7	2	1024	BREADTH	2
5	128	3	4-128	1	1	17	1	DEPTH	405

5	128	3	4-128	1	1	17	1	BREADTH	507
5	128	3	4-128	1	1	17	2	DEPTH	679
5	128	3	4-128	1	1	17	2	BREADTH	281
5	128	3	4-128	1	1	17	4	DEPTH	1019
5	128	3	4-128	1	1	17	4	BREADTH	1019
5	128	3	4-128	1	1	17	8	DEPTH	2043
5	128	3	4-128	1	1	17	8	BREADTH	2042
5	128	3	4-128	1	1	17	16	DEPTH	2041
5	128	3	4-128	1	1	17	16	BREADTH	1018
5	128	3	4-128	1	1	17	32	DEPTH	1019
5	128	3	4-128	1	1	17	32	BREADTH	1201
5	128	3	4-128	1	1	17	64	DEPTH	2043
5	128	3	4-128	1	1	17	64	BREADTH	2044
5	128	3	4-128	1	1	17	128	DEPTH	2046
5	128	3	4-128	1	1	17	128	BREADTH	1021
5	128	3	4-128	1	1	17	256	DEPTH	2049
5	128	3	4-128	1	1	17	256	BREADTH	2047
5	128	3	4-128	1	1	17	512	DEPTH	1022
5	128	3	4-128	1	1	17	512	BREADTH	1021
5	128	3	4-128	1	1	17	1024	DEPTH	2046
5	128	3	4-128	1	1	13	1024	BREADTH	1021
5	128	3	4-128	2	2	13	1	DEPTH	137
5	128	3	4-128	2	2	13	1	BREADTH	145
5	128	3	4-128	2	2	13	2	DEPTH	148
5	128	3	4-128	2	2	13	2	BREADTH	145

5	128	3	4-128	2	2	13	4	DEPTH	156
5	128	3	4-128	2	2	13	4	BREADTH	152
5	128	3	4-128	2	2	13	8	DEPTH	156
5	128	3	4-128	2	2	13	8	BREADTH	155
5	128	3	4-128	2	2	13	16	DEPTH	156
5	128	3	4-128	2	2	13	16	BREADTH	152
5	128	3	4-128	2	2	13	32	DEPTH	150
5	128	3	4-128	2	2	13	32	BREADTH	152
5	128	3	4-128	2	2	13	64	DEPTH	155
5	128	3	4-128	2	2	13	64	BREADTH	159
5	128	3	4-128	2	2	13	128	DEPTH	158
5	128	3	4-128	2	2	13	128	BREADTH	155
5	128	3	4-128	2	2	13	256	DEPTH	147
5	128	3	4-128	2	2	13	256	BREADTH	146
5	128	3	4-128	2	2	13	512	DEPTH	144
5	128	3	4-128	2	2	13	512	BREADTH	147
5	128	3	4-128	2	2	13	1024	DEPTH	85
5	128	3	4-128	2	2	13	1024	BREADTH	82
5	128	3	4-128	3	2	14	1	DEPTH	101
5	128	3	4-128	3	2	14	1	BREADTH	105
5	128	3	4-128	3	2	14	2	DEPTH	102
5	128	3	4-128	3	2	14	2	BREADTH	106
5	128	3	4-128	3	2	14	4	DEPTH	105
5	128	3	4-128	3	2	14	4	BREADTH	107
5	128	3	4-128	3	2	14	8	DEPTH	105

5	128	3	4-128	3	2	14	8	BREADTH	107
5	128	3	4-128	3	2	14	16	DEPTH	106
5	128	3	4-128	3	2	14	16	BREADTH	108
5	128	3	4-128	3	2	14	32	DEPTH	107
5	128	3	4-128	3	2	14	32	BREADTH	104
5	128	3	4-128	3	2	14	64	DEPTH	104
5	128	3	4-128	3	2	14	64	BREADTH	106
5	128	3	4-128	3	2	14	128	DEPTH	104
5	128	3	4-128	3	2	14	128	BREADTH	106
5	128	3	4-128	3	2	14	256	DEPTH	105
5	128	3	4-128	3	2	14	256	BREADTH	107
5	128	3	4-128	3	2	14	512	DEPTH	105
5	128	3	4-128	3	2	14	512	BREADTH	104
5	128	3	4-128	3	2	14	1024	DEPTH	51
5	128	3	4-128	3	2	14	1024	BREADTH	53
5	128	3	4-128	3	3	14	1	DEPTH	93
5	128	3	4-128	3	3	14	1	BREADTH	99
5	128	3	4-128	3	3	14	2	DEPTH	99
5	128	3	4-128	3	3	14	2	BREADTH	98
5	128	3	4-128	3	3	14	4	DEPTH	100
5	128	3	4-128	3	3	14	4	BREADTH	98
5	128	3	4-128	3	3	14	8	DEPTH	97
5	128	3	4-128	3	3	14	8	BREADTH	97
5	128	3	4-128	3	3	14	16	DEPTH	101
5	128	3	4-128	3	3	14	16	BREADTH	97

5	128	3	4-128	3	3	14	32	DEPTH	100
5	128	3	4-128	3	3	14	32	BREADTH	99
5	128	3	4-128	3	3	14	64	DEPTH	98
5	128	3	4-128	3	3	14	64	BREADTH	99
5	128	3	4-128	3	3	14	128	DEPTH	98
5	128	3	4-128	3	3	14	128	BREADTH	96
5	128	3	4-128	3	3	14	256	DEPTH	101
5	128	3	4-128	3	3	14	256	BREADTH	100
5	128	3	4-128	3	3	14	512	DEPTH	98
5	128	3	4-128	3	3	14	512	BREADTH	99
5	128	3	4-128	3	3	14	1024	DEPTH	48
5	128	3	4-128	3	3	14	1024	BREADTH	47
5	128	3	4	6	7	1	1	DEPTH	1
5	128	3	4	6	7	1	1	BREADTH	1
5	128	3	4	6	7	1	2	DEPTH	1
5	128	3	4	6	7	1	2	BREADTH	1
5	128	3	4	6	7	1	4	DEPTH	1
5	128	3	4	6	7	1	4	BREADTH	1
5	128	3	4	6	7	1	8	DEPTH	1
5	128	3	4	6	7	1	8	BREADTH	2
5	128	3	4	6	7	1	16	DEPTH	2
5	128	3	4	6	7	1	16	BREADTH	2
5	128	3	4	6	7	1	32	DEPTH	2
5	128	3	4	6	7	1	32	BREADTH	2
5	128	3	4	6	7	1	64	DEPTH	2

5	128	3	4	6	7	1	64	BREADTH	2
5	128	3	4	6	7	1	128	DEPTH	2
5	128	3	4	6	7	1	128	BREADTH	2
5	128	3	4	6	7	1	256	DEPTH	2
5	128	3	4	6	7	1	256	BREADTH	2
5	128	3	4	6	7	1	512	DEPTH	2
5	128	3	4	6	7	1	512	BREADTH	2
5	128	3	4	6	7	1	1024	DEPTH	2
5	128	3	4	6	7	1	1024	BREADTH	2

Appendix B: Exchange Server 2003 test

Jetstress test results

Microsoft's Jetstress 2004 utility was used to simulate Exchange I/0 against a storage subsystem to test the performance and determine the maximum number of Exchange IOPs that the subsystem can support. Engineers performed Jetstress tests on each controller-disk configuration at RAID levels 0, 1+0, 5, and 6. For each configuration, engineers increased the number of Jetstress threads until the I/O latency exceeded the acceptable threshold limit of 20 ms, as recommended by Microsoft. In each Jetstress test performed by HP, the read latency was the first metric to exceed 20 ms. Therefore, engineers used read latency as the key metric in determining the pass/fail status of each configuration.

Storage array configurations

Engineers performed Jetstress tests on one SAS array configuration and two SCSI array configurations connected to a ProLiant DL380 G4 server (Figure B-1). The controller-disk subsystem configurations were as follows:

• Smart Array P600 controller attached to an MSA50 enclosure populated with ten 36-GB, 10K RPM SAS drives

• Smart Array 6402 controller attached to an MSA30 enclosure with ten 10K RPM, 36-GB, U320 SCSI drives

• Smart Array

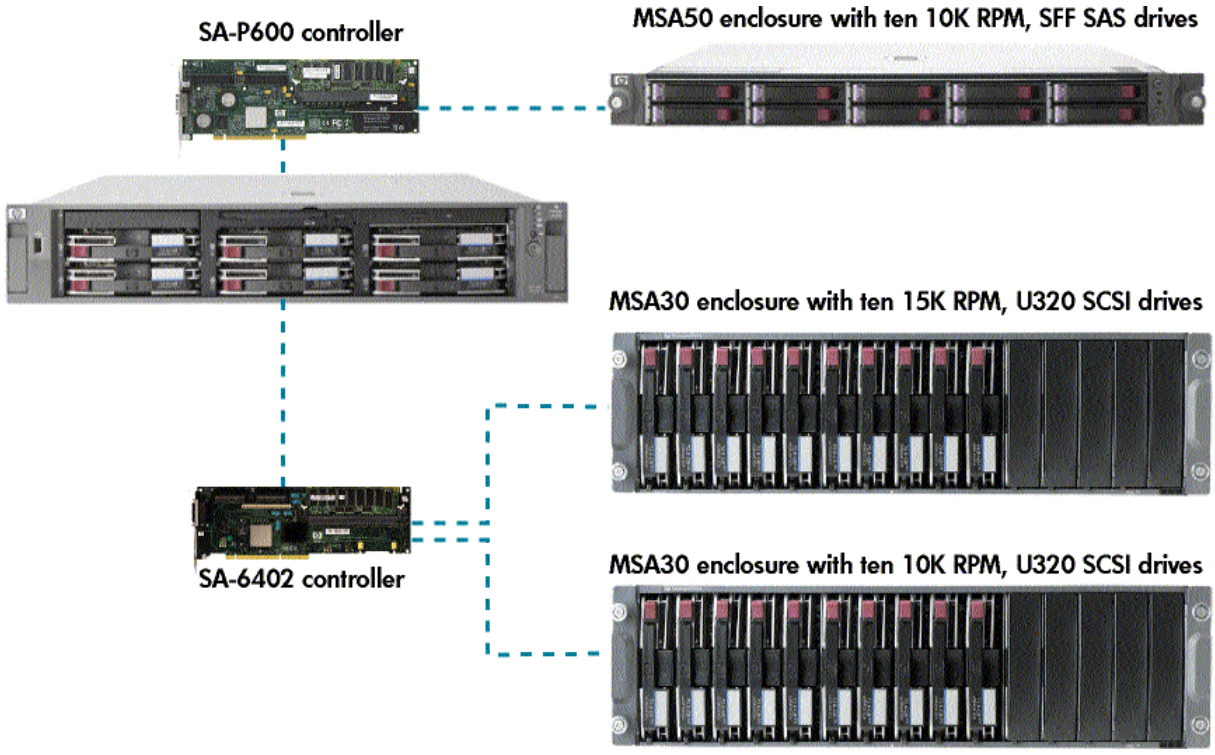


Figure B-1. Controller-disk subsystem configurations for Jetstress tests

The tests were performed using RAID levels 0, 1+0, 5, and 6. Engineers used the default cache settings—50 percent read and 50 percent write—as well as the default stripe sizes indicated in table B-1.

Table B-1. Default stripe sizes used in Jetstress tests

RAID level	Default stripe size
0	128 KB
1+0	128 KB
5	64 KB
6	16 KB

The Jetstress GUI (Figure B-2) was used to configure the test settings shown in Table B-2.

Table B-2. Test settings used in Jetstress tests

Test parameter	Setting
Test duration	2 hr
Estimated IOPs per mailbox	1
Mailbox size limit	100 MB
No. of mailboxes on server	1,000
Hardware storage cache	256 MB
No. of storage groups	1

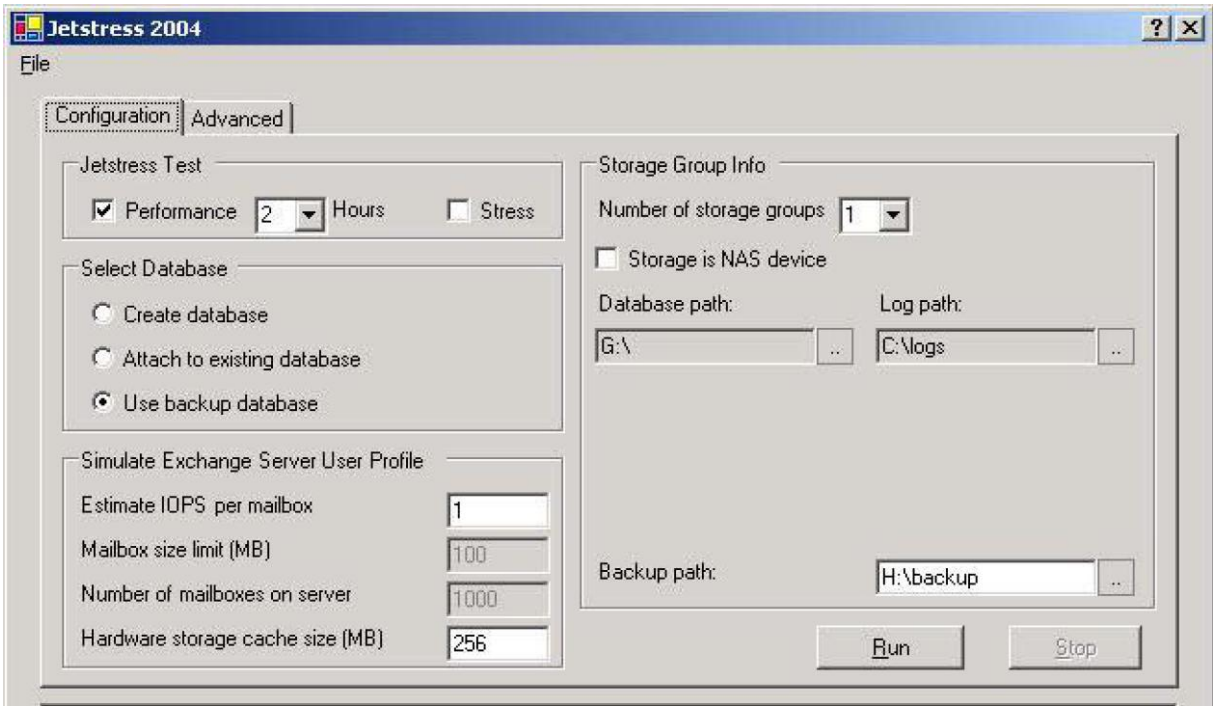


Figure B-2. Jetstress GUI displaying test configuration settings

The Jetstress GUI Advanced screen (Figure B-3) allowed engineers to further customize the test parameters. The default configuration allows Jetstress to self-tune to determine the maximum number of I/Os that a storage subsystem can support. Allowing Jetstress to self-tune would result in configuration parameters that varied between tests, which does not allow an accurate comparison of controller and disk configurations. Therefore, engineers specified a fixed workload for all tests and varied only the number of threads until the average latency for disk reads or writes exceeded 20 ms. The parameters in Advanced Settings were configured as shown in Table B-3.

Table B-3. Test parameters defined in Jetstress advanced settings

*** The number of threads was varied between tests to increase the I/O load.**

Test setting	Parameter
Threads	Variable*
Log buffer	64
Inserts	20
Replaces	75
Deletes	5
Lazy commits	93

The values shown in Table B-3 and Figure B-3 result in an approximate 65:35 read/write ratio workload. This read/write ratio is typical of a corporate workload.

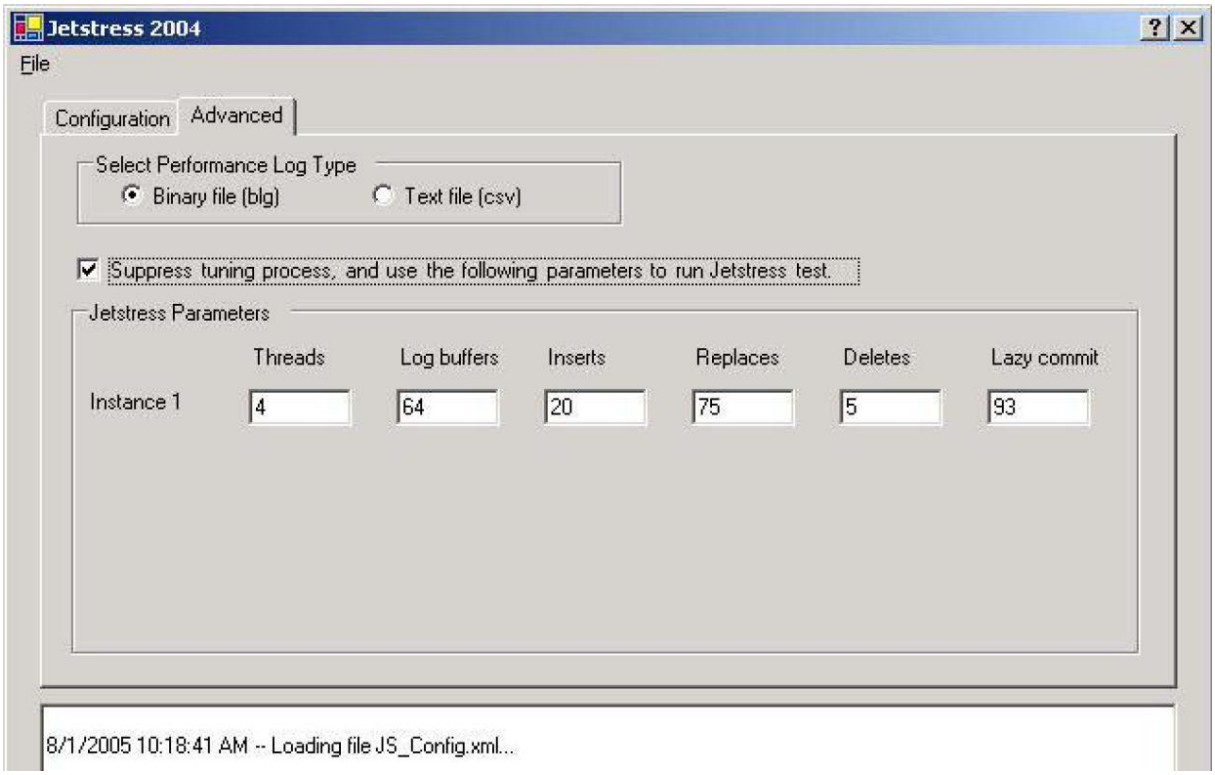


Figure B-3. Jetstress Advanced settings

RAID 0 test results

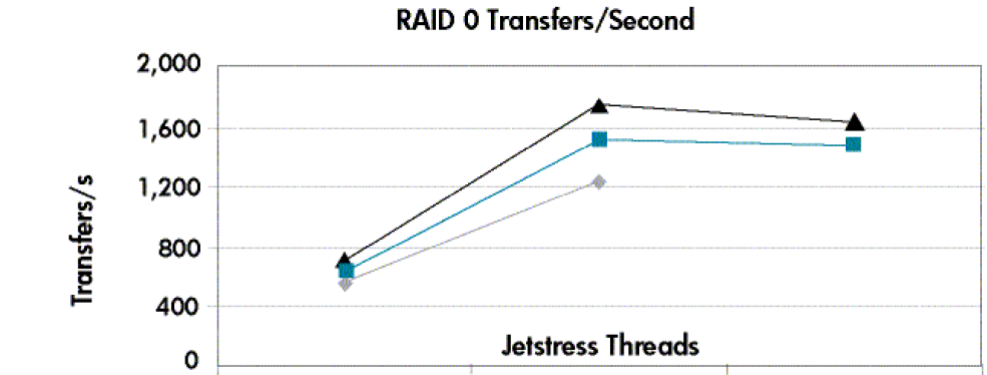
Raid 0 is the highest performing RAID level configuration because it performs striping across all members of the array. However, RAID 0 provides no fault tolerance

capabilities and is susceptible to catastrophic data loss in the event of a single disk failure. Due to possible data loss, RAID 0 is seldom used for production deployments. The results of the Jetstress tests at RAID 0 are shown in Figure B-4 and summarized below.

• The SA-6402 with 15K RPM U320 SCSI drives averaged 1757 transfers per second at 16 Jetstress threads with an average read latency of 0.013 seconds and an average write latency of 0.0017 seconds.

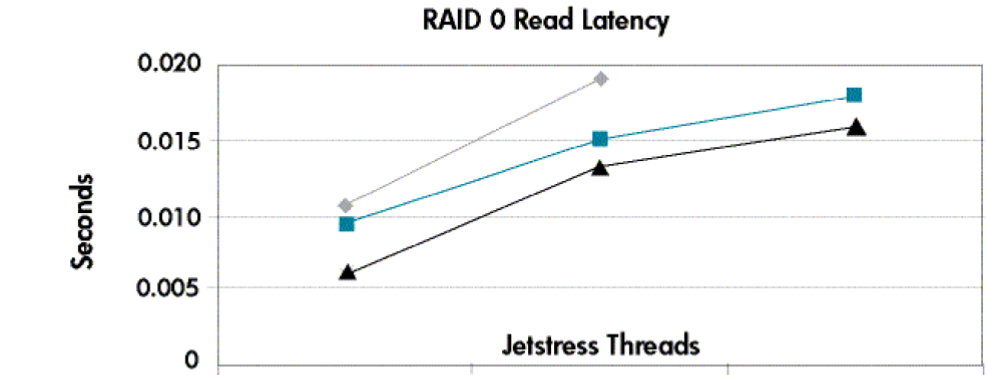
• SA-P600 with 10K RPM SAS drives averaged 1526 transfers per second at 16 Jetstress threads with an average read latency of 0.015 seconds and an average write latency of 0.00018.

• The SA-6402 with the 10K RPM Ultra 3 SCSI drives averaged 1241 transfers per second at 16 Jetstress threads with an average read latency of 0.019 and an average write latency of 0.00167.



Drives	4	16	18
10K RPM SCSI	576.94	1241.53	*
10K RPM SAS	653.28	1526.20	1493.67
15K RPM SCSI	704.59	1757.24	1637.11

*Read latency > 20 ms (see RAID 0 Read Latency chart below)



Drives	4	16	18
10K RPM SCSI	0.0103	0.01900	*
10K RPM SAS	0.0090	0.01540	0.0178
15K RPM SCSI	0.0063	0.01334	0.0161

*Read latency > 20 ms

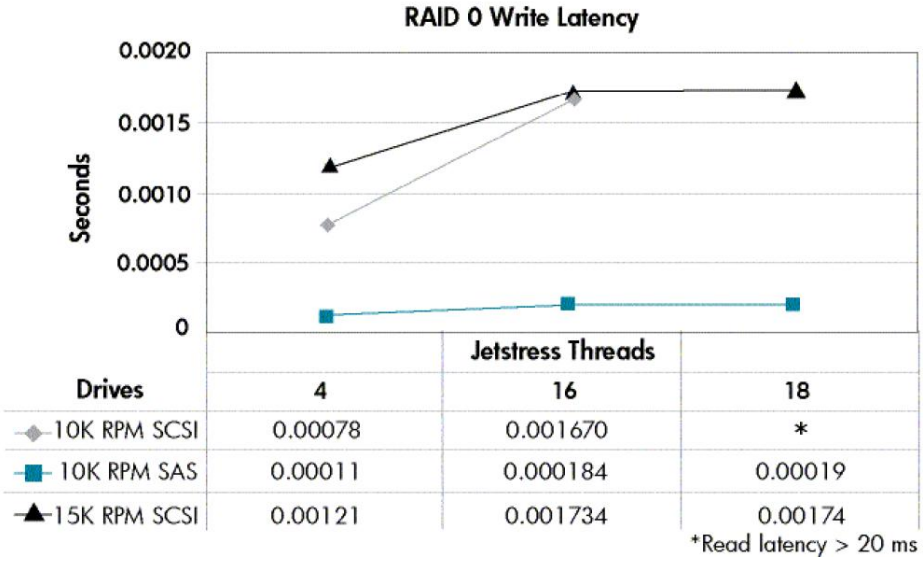


Figure B-4. Jetstress test data at RAID 0

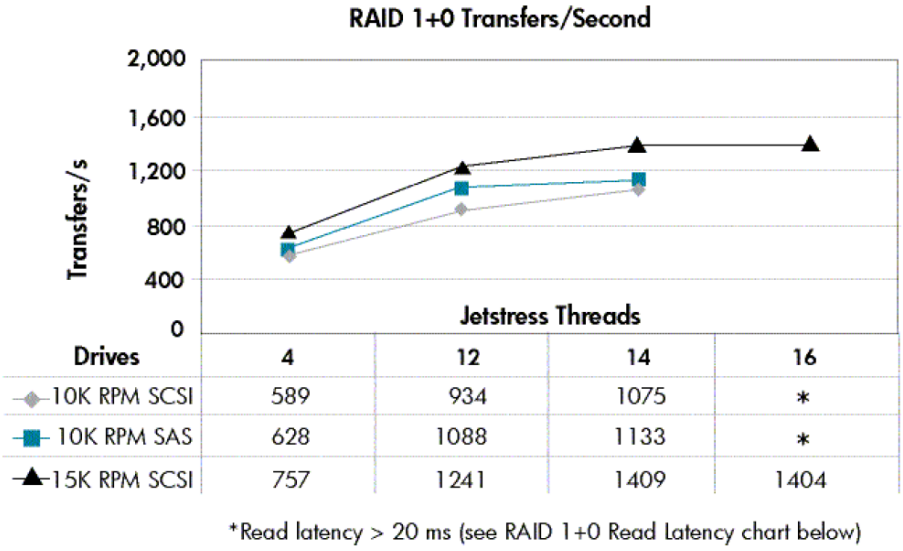
RAID 1+0 test results

RAID 1+0 uses disk striping and data mirroring to provide the highest level of performance with the highest level of fault tolerance to protect against data loss in the event of a hard disk failure. In a RAID 1+0 logical drive, 50 percent of the available disk capacity is required for data mirroring. RAID 1+0 is commonly used in production environments when both performance and fault tolerance are required. The results of the Jetstress tests at RAID 1+0 are shown in Figure B-5 and summarized below.

• The SA-6402 with 15K RPM U320 SCSI drives averaged 1409 transfers per second at 14 Jetstress threads with an average read latency of 0.015 seconds and an average write latency of 0.0019 seconds.

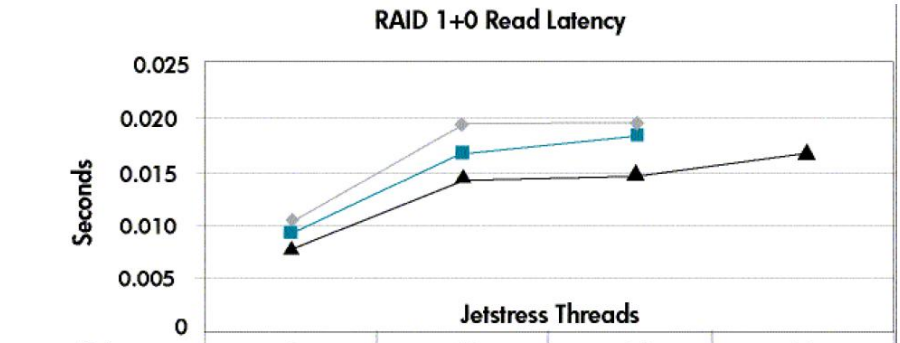
• SA-P600 with 10K RPM SAS drives averaged 1133 transfers per second at 14 Jetstress threads with an average read latency of 0.0186 seconds and an average write latency of 0.000197.

• The SA-6402 with the 10K RPM Ultra 3 SCSI drives averaged 1075 transfers per second at 14 Jetstress threads with an average read latency of 0.0196 and an average write latency of 0.00178.



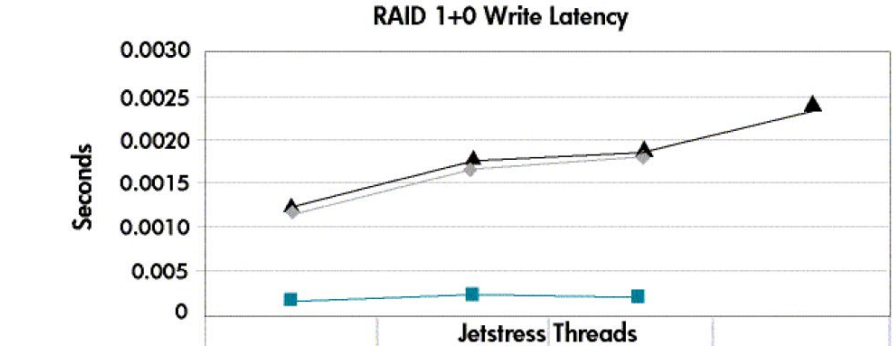
Drives	4	12	14	16
10K RPM SCSI	589	934	1075	*
10K RPM SAS	628	1088	1133	*
15K RPM SCSI	757	1241	1409	1404

*Read latency > 20 ms (see RAID 1+0 Read Latency chart below)



Drives	4	12	14	16
10K RPM SCSI	0.01010	0.01945	0.0196	*
10K RPM SAS	0.00951	0.01659	0.0186	*
15K RPM SCSI	0.00782	0.01440	0.0148	0.017

*Read latency > 20 ms



Drives	4	12	14	16
10K RPM SCSI	0.001170	0.001642	0.001780	*
10K RPM SAS	0.000150	0.000199	0.000197	*
15K RPM SCSI	0.001261	0.001797	0.001880	0.002411

*Read latency > 20 ms

Figure B-5. Jetstress test data at RAID1+0

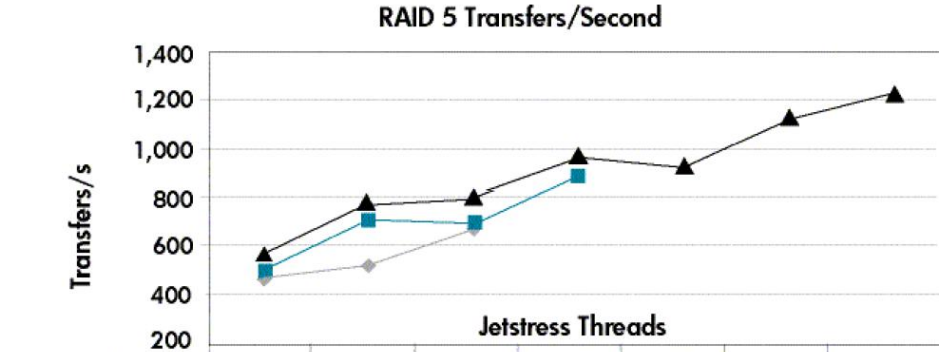
RAID 5 test results

RAID 5 uses striping with parity to provide increased performance and fault tolerance. RAID 5 logical drives do not provide the same level of performance achieved with a RAID 1+0 logical drives. However, RAID 5 logical drives only require 1/n (n = number of drives in the logical drive) of the total disk capacity to provide a level of fault tolerance for data protection. In general, a RAID 5 array is less costly than a RAID 1+0 array, which requires 50 percent of the total disk capacity of a logical drive for data mirroring. The results of the Jetstress tests at RAID 5 are shown in Figure B-6 and summarized below.

• The SA-6402 with 15K RPM U320 SCSI drives averaged 1218.5 transfers per second at 16 Jetstress threads with an average read latency of 0.0197 seconds and an average write latency of 0.00618 seconds.

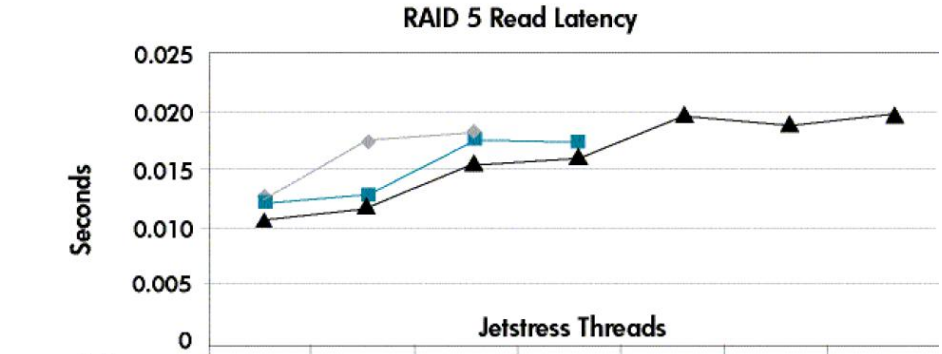
• SA-P600 with 10K RPM SAS drives averaged 873.8 transfers per second at 10 Jetstress threads with an average read latency of 0.0173 seconds and an average write latency of 0.00018.

• The SA-6402 with the 10K RPM Ultra 3 SCSI drives averaged 665.3 transfers per second at 8 Jetstress threads with an average read latency of 0.0182 and an average write latency of 0.0017.



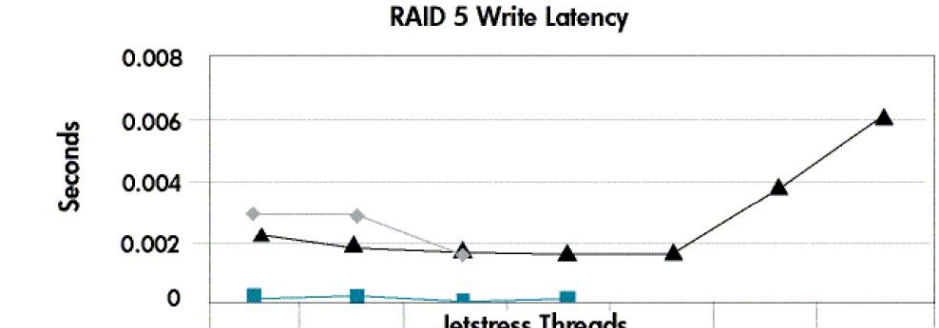
Drives	4	6	8	10	12	14	16
10K RPM SCSI	475	518	665	*	*	*	*
10K RPM SAS	496	705	687	1088	873	*	*
15K RPM SCSI	565	774	793	966	926	1127	1219

*Read latency > 20 ms (see RAID 5 Read Latency chart below)



Drives	4	6	8	10	12	14	16
10K RPM SCSI	0.01270	0.01763	0.01820	*	*	*	*
10K RPM SAS	0.01217	0.01283	0.01760	0.01731	*	*	*
15K RPM SCSI	0.01064	0.01164	0.01526	0.01560	0.0196	0.01869	0.0197

*Read latency > 20 ms



Drives	4	6	8	10	12	14	16
10K RPM SCSI	0.00297	0.00293	0.00170	*	*	*	*
10K RPM SAS	0.00016	0.00025	0.00014	0.00018	*	*	*
15K RPM SCSI	0.00234	0.00194	0.00186	0.00177	0.00176	0.00373	0.00618

*Read latency > 20 ms

Figure B-6. Jetstress test data at RAID 5

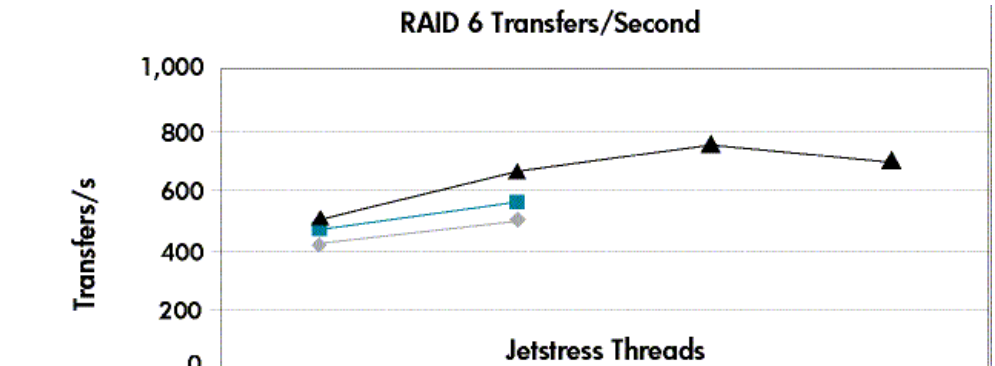
RAID 6 test results

RAID 6, unique to HP Smart Array controllers, provides additional data protection by recording two independent sets of parity data. The results of the Jetstress tests with RAID 6 are shown in Figure B-7 and summarized below.

• The SA-6402 with 15K RPM U320 SCSI drives averaged 825.95 transfers per second at 10 Jetstress threads with an average read latency of 0.018 seconds and an average write latency of 0.0033 seconds.

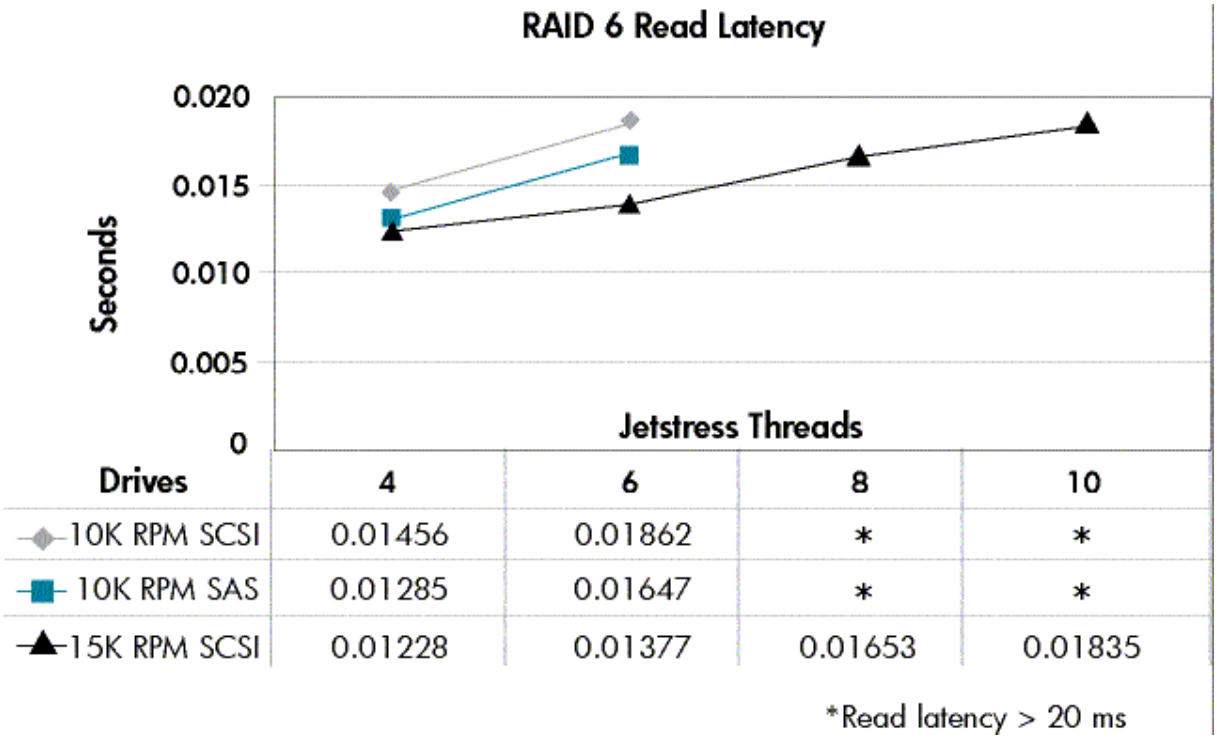
• SA-P600 with 10K RPM SAS drives averaged 544 transfers per second at 6 Jetstress threads with an average read latency of 0.016 seconds and an average write latency of 0.00029.

• The SA-6402 with the 10K RPM Ultra 3 SCSI drives averaged 492 transfers per second at 6 Jetstress threads with an average read latency of 0.018 and an average write latency of 0.003.



Drives	4	6	8	10
10K RPM SCSI	416	492	*	*
10K RPM SAS	470	544	*	*
15K RPM SCSI	492	660	733	826

*Read latency > 20 ms (see RAID 6 Read Latency chart below)



Drives	4	6	8	10
10K RPM SCSI	0.01456	0.01862	*	*
10K RPM SAS	0.01285	0.01647	*	*
15K RPM SCSI	0.01228	0.01377	0.01653	0.01835

*Read latency > 20 ms

Figure B-7. Jetstress test data with RAID 6

Conclusions

• The SA-6402 controller configured with the 15K RPM U320 SCSI drives achieved the highest number of disk transfers per second below the 20 ms threshold.

• The SA-P600 controller with 10K RPM SAS drives consistently outperformed the SA-6402 controller configured with the 10K RPM U320 SCSI drives.

• The SA-P600 controller with 10K SAS drives had significantly lower write latencies that the SCSI storage configurations.

LoadSim 2003 test results

HP engineers performed LoadSim tests to validate the performance of SAS and SCSI storage subsystems connected to a ProLiant DL380 G4 server running Exchange 2003. The LoadSim tests were configured to run 1,200 simulated MMB-3 client profiles. The ntroller-disk subsystems configurations included the following:

• SA-P600 controller attached to an MSA50 enclosure populated with ten 36-GB, 10K RPM SAS drives.

• SA-6402 controller attached to an MSA30 enclosure with ten 10K RPM, 36-GB, U320 SCSI drives.

• SA-6402 controller attached to an MSA30 enclosure with ten 15K RPM, 36-GB, U320 SCSI drives.

Engineers performed LoadSim tests on each configuration at RAID 1+0 and RAID 5. These two RAID levels are prevalent in Microsoft Exchange deployments because they provide the best combination of performance and fault tolerance.

Given a fixed number of simulated users (1,200), the engineers used client response time as the key metric in evaluating the relative performance of the controller-disk subsystems at each RAID level. Figure B-8 shows that at each RAID level, the 15K RPM U320 SCSI drives achieved the fastest client response time and the 10K RPM U320 SCSI drives had the slowest response time.

• During several tests, the disk transfers per second decreased when additional load was placed on the controller. This performance curve is an indication that the controller throughput is saturated and optimal performance was achieved with less load.

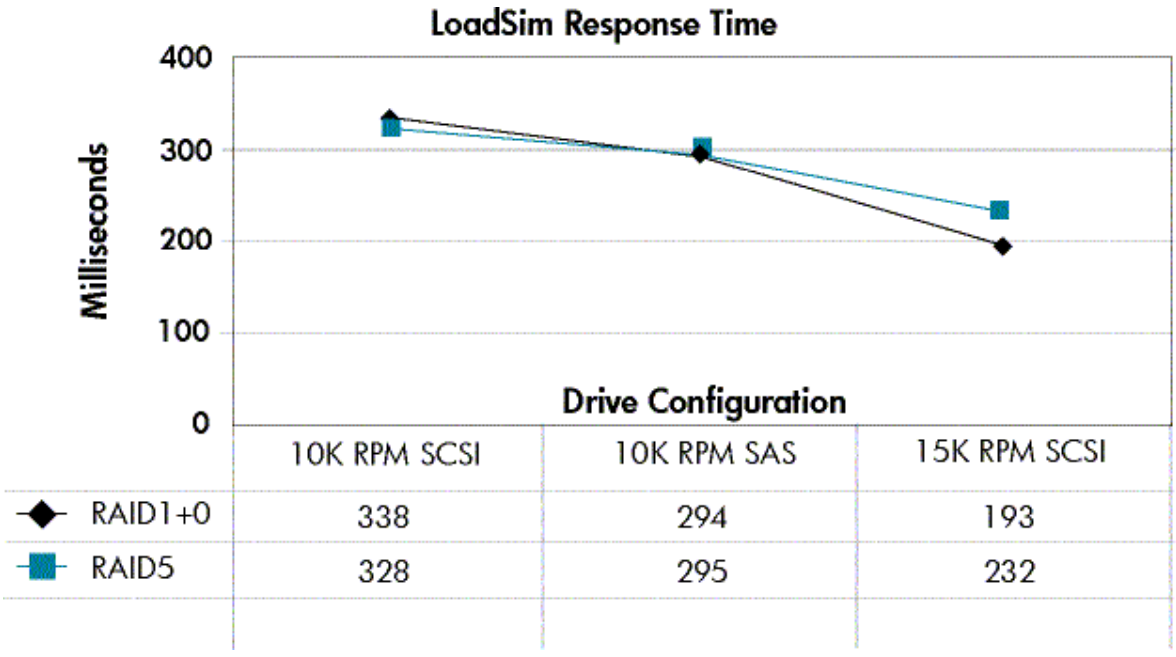


Figure B-8. LoadSim MMB-3 response time for 1200 simulated clients

The LoadSim MMB-3 tests simulating 1,200 user profiles generated an average of 580 transfers per second during each test. Reviewing the results from the Jetstress tests at RAID 1+0, this average transfer rate is well within the capabilities of the SA-P600 and the SA-6402 controllers. For example, during the Jetstress test at RAID 1+0, the SA-P600 controller successfully supported 1133 transfers per second with an average latency below the 20 ms threshold.

VITAE

Name Surname: Ali Rıza BALI

Address: Murat Reis Mah. Hatmiyan Sk. Melek Apt. No: 8/17 Uskudar / ISTANBUL

Birth place and year: Bursa 1984

Foreign Language: English

Under Graduate: Marmara University 2006

Graduate: Bahcesehir University

Institute Name: Institute of Science

Program Name: Computer Engineering

Working Life: Metis A.Ş. 2006-2008, Hewlett-Packard 2008-Still employee