# Intel® RAID Controllers

## *Competitive Performance White Paper*

**Revision 1.0**

**February 02, 2009**

**Enterprise Platforms and Services Division - Marketing**

# Revision History

| Date | Revision Number | Modifications |
|---|---|---|
| Feb. 02, 2009 | 1.0 | Initial Release |

# Disclaimers

Information in this document is provided in connection with Intel® products. No license, express or implied, by estoppel or otherwise, to any intellectual property rights is granted by this document. Except as provided in Intel's Terms and Conditions of Sale for such products, Intel assumes no liability whatsoever, and Intel disclaims any express or implied warranty, relating to sale and/or use of Intel products including liability or warranties relating to fitness for a particular purpose, merchantability, or infringement of any patent, copyright or other intellectual property right. Intel products are not intended for use in medical, life saving, or life sustaining applications. Intel may make changes to specifications and product descriptions at any time, without notice.

Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined." Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them.

The Intel® RAID Controllers may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Except as permitted by license, no part of this document may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without the express written consent of Intel Corporation.

Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, visit Intel Performance Benchmark Limitations at http://www.intel.com/performance/resources/benchmark_limitations.htm.

Intel, Pentium, Itanium, and Xeon are trademarks or registered trademarks of Intel Corporation.

*Other brands and names may be claimed as the property of others.

Copyright © Intel Corporation 2009. Portions Copyright © LSI Corporation 2009.

# *Table of Contents*

# List of Figures

# List of Tables

# 1. Introduction

RAID (Redundant Array of Independent Disks) technology has been commonly implemented in server usage models as an option to increase server performance while enhancing data protection. RAID solutions are now also found in other computer environments such as desktops, workstations, and external storage devices that support a large number of hard drives.

Performance is often a key parameter used to compare RAID solutions from multiple vendors. Therefore, it is common for vendors to provide performance results that provide maximum value, but yet the results do not reflect real-world applications and the configuration settings appropriate to these applications.

This paper examines various I/O workloads seen in common application storage ecologies, such as data streaming and transactional environments, and provides benchmark results using configuration settings suited to these applications. For information and comparison, the performance of Intel® RAID Controller SRCSASJV is portrayed along side results from two leading competitive products.

In addition, this paper describes several best configuration practices and tuning guidance that can be employed to optimize storage performance and data availability for many servers and application environments.

Ultimately, this paper portrays the unique advantage offered by Intel® RAID controllers, powered by LSI MegaRAID* technology, for addressing real-user environments and applications. System builders deploying Intel® RAID controllers along with Intel® platforms will be assured that when configured properly for the tasks at hand, they have a highly reliable solution that demonstrates best-in-class performance.

# 2. Performance for the Real World

Real world benchmarks should replicate, as much as possible, the I/O profiles that are characteristic of the given application environment. For example, a benchmark for streaming applications, such as Video Surveillance, may be based on random reads and writes reflecting the existence of multiple backup targets and multiple camera feeds.

Within this section, typical server applications and RAID benchmarks appropriate for these applications are examined.

## 2.1 Unstructured Data

The term "unstructured data" is a common buzzword that represents data that is not managed as a data structure or organization by an application. Examples of unstructured data may include audio, video and text such as an email, word processor, or a spreadsheet document. Research suggests that unstructured data represents as much as 80% of corporate data storage. Unstructured data is usually accessed in a random fashion with data requests usually made up of small-size data transfers.  With small-size random data requests, hard drives seek time and the controller's response to these data requests greatly influences performance. The following figure shows the average response time for the Intel® RAID Controller SRCSASJV compared to comparable competitor's products. A lower response time is the goal for this type of performance measurement.

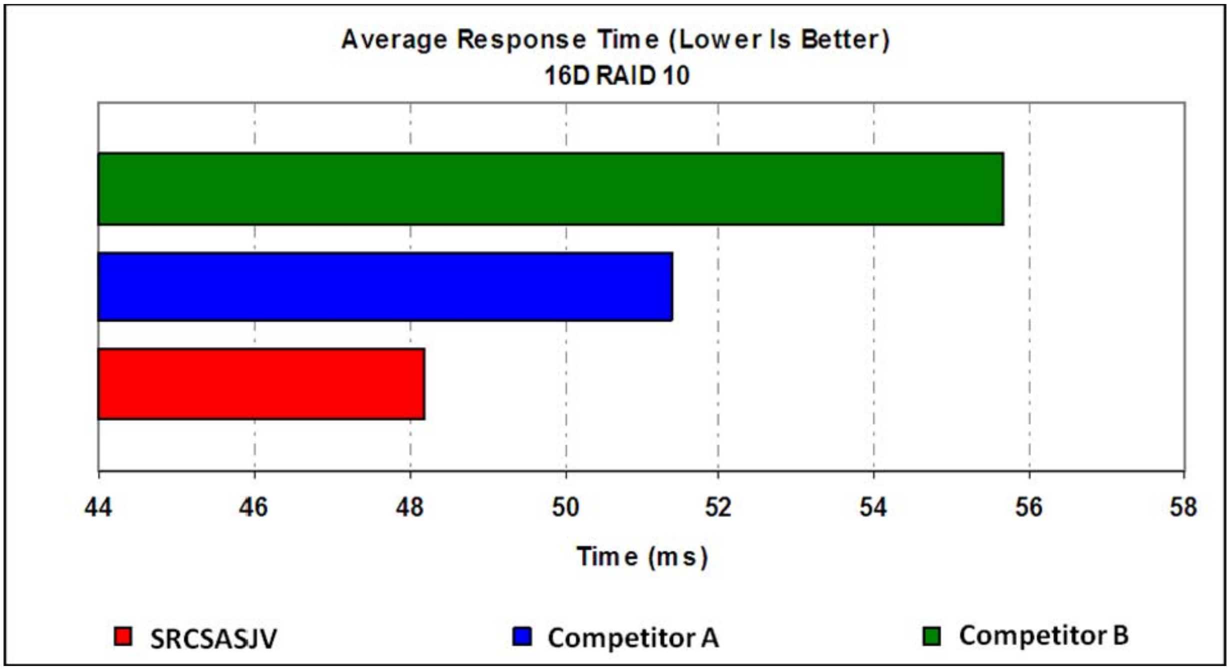
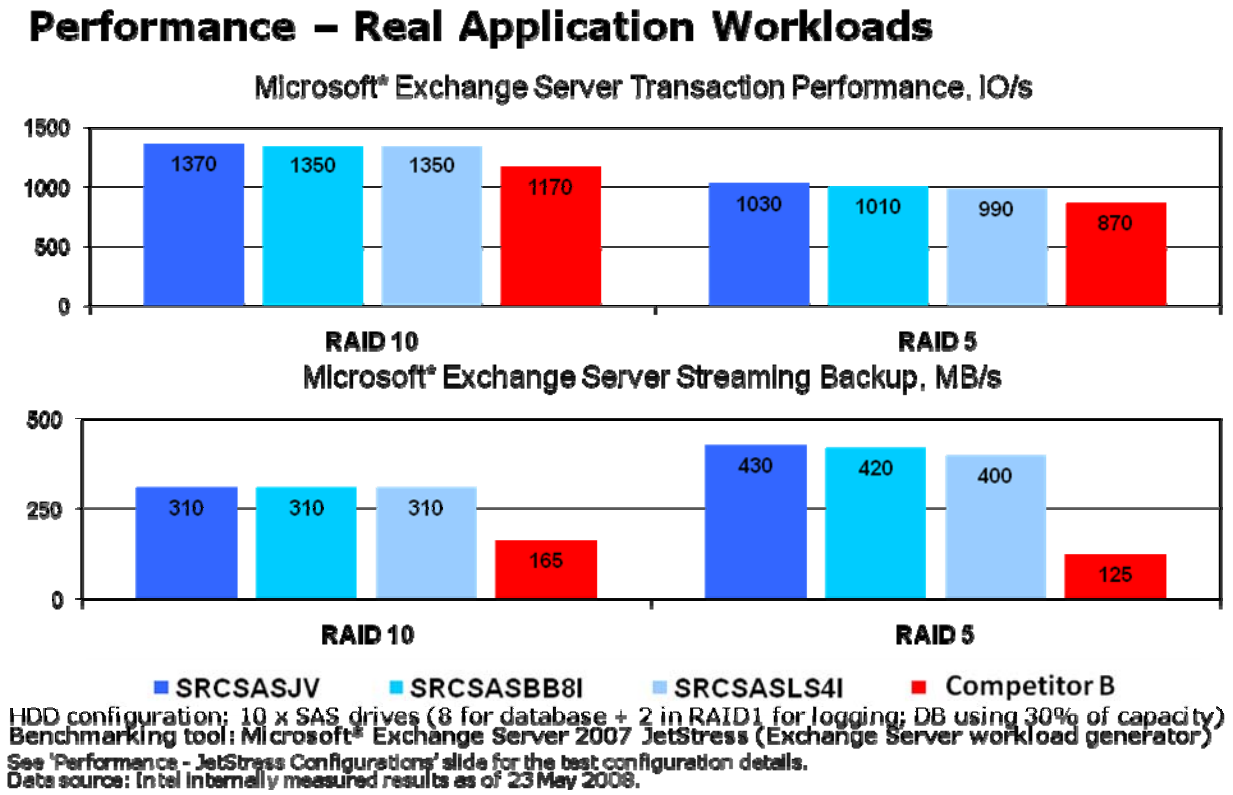


**Figure 1. Sixteen-drive RAID 10, 4 KB random reads, average response time utilizing 15K RPM Seagate SAS disks**

Response time is defined as the amount of time required to complete a request submitted by the application; lower is better.

## 2.2 Databases

Databases are employed in many applications, everything from online transaction processing (OLTP), e-mail (Exchange, Notes), to Enterprise Resource Planning. Though they vary significantly in complexity, all databases have common characteristics, such as a structured collection of records, strict in-order processing (to ensure transactions are processed reliably), and logging to maintain the integrity of the database. The number of I/Os that a RAID controller can perform is a vitally important attribute for measuring the performance of a database transaction, as well as the throughput of the RAID solution in MB/s. As shown in the following figure, the combination of I/Os capability and throughput of the Intel® RAID Controller SRCSASJV make it an excellent selection for database applications.



**Performance – Real Application Workloads**

Microsoft* Exchange Server Transaction Performance, IO/s

Microsoft* Exchange Server Streaming Backup, MB/s

HDD configuration: 10 x SAS drives (8 for database + 2 in RAID1 for logging; DB using 30% of capacity)
Benchmarking tool: Microsoft* Exchange Server 2007 JetStress (Exchange Server workload generator)
See 'Performance - JetStress Configurations' slide for the best configuration details.
Data source: Intel internally measured results as of 23 May 2008.

**Figure 2. Intel RAID Controllers – Database Workload**

## 2.3 E-commerce

Today, e-commerce lies at the core of modern society providing services such as banking, online retail shopping and travel reservation systems. Other common real-world applications include web servers, email servers, and workstations.

I/Os per seconds are typically used to measure transactional based storage performance. Figures 3 and 4 below depict the performance of the Intel® RAID Controller SRCSASJV using a benchmark appropriate for judging E-commerce effectiveness.

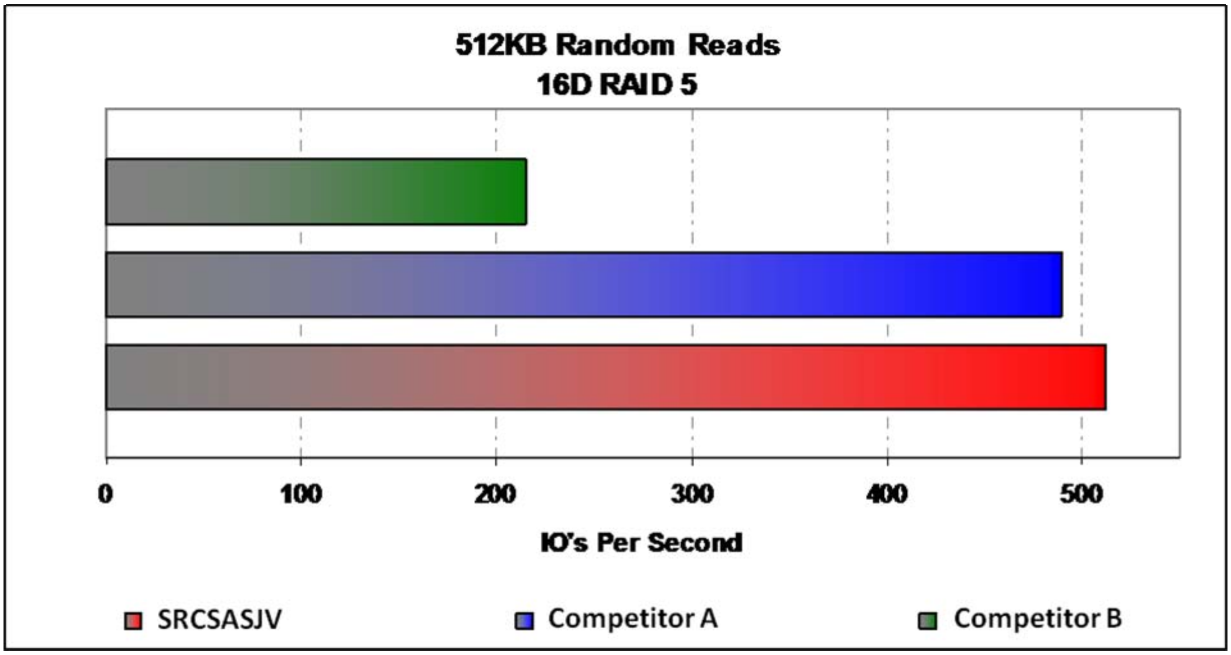


Figure 3. Sixteen-drive RAID 5, 512 KB random read performance at queue depth of 256

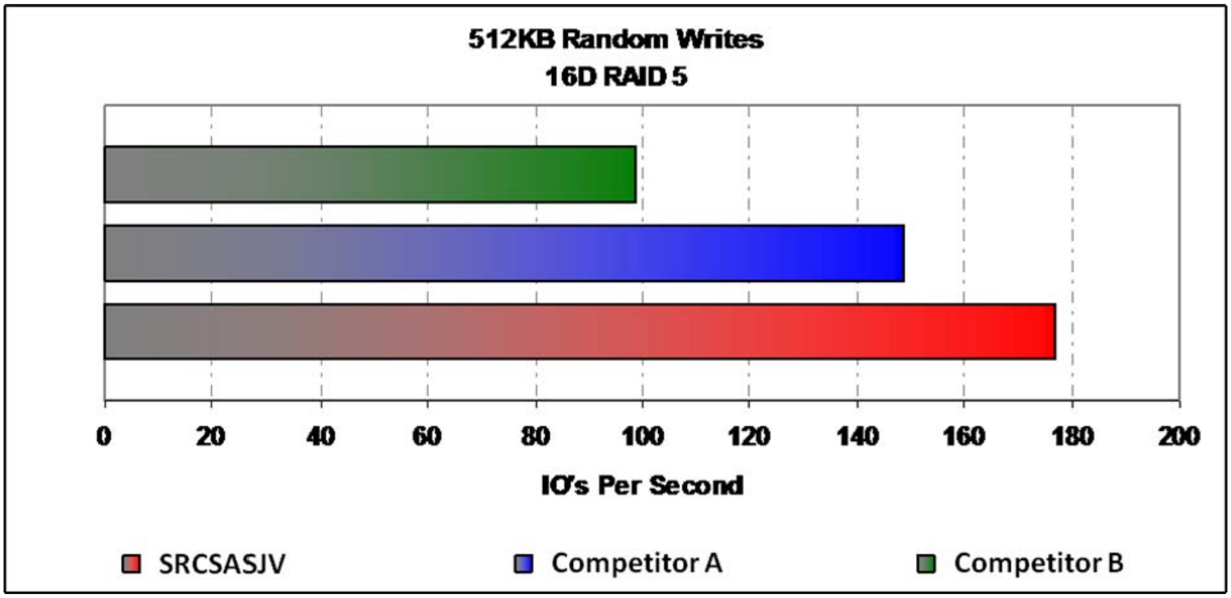


Figure 4. Sixteen-drive RAID 5, 512 KB random write performance at queue depth of 256

## 2.4 Steaming Applications (Example: Video Surveillance)

With the declining cost of hard disks and the proliferation of IP Storage, digital technology has made video surveillance practical and economical. Historically, video quality was low resolution, and was stored as an analog signal on magnetic tape. SATA disks have now become the medium of choice for media-based applications, replacing cumbersome and delicate magnetic tape.

Bandwidth requirements vary based on the resolution and frame rate. However, security cameras can now easily transport high-resolution images from several cameras, requiring hundreds of megabytes per second in streaming writes in order to avoid dropping frames. Conversely, video servers require similar performance by placing streaming read demands on the storage. While each of these streams tends to be fairly sequential in nature, the instantaneous logical block address (LBA) appears random in nature because of the location of the media on the volume. An ideal way to simulate this workload is by applying a large random read or write workload. Figure 1 and Figure 3 above also demonstrate how well the Intel® RAID Controller SRCSASJV can handle this simulated workload.

Usually, streaming applications including video surveillance, Virtual Tape Libraries (VTL), video serving and disk-to-disk-to-tape (D2D2T) are characterized by I/O that is less structured than pure sequential and more structured than pure random I/O – lying somewhere in between. Therefore, multiple benchmarks should be considered to predict performance given the multi-threaded non-overlapping I/O patterns of multi-user requests. Strong performance at both ends of the spectrum ensures that your RAID controller will function well in your environment. (See "Chapter 4. Intel RAID Maximum Throughput" for sequential read and write performance of the Intel® RAID Controller SRCSASJV).

## 2.5 Hardware RAID Advantage

Hardware-based RAID controllers offer a distinct performance advantage over software-based RAID solutions that require host resources. At first glance, one could perceive a software-based RAID solution as leveraging much more powerful host CPU and larger system memory resources, thereby producing a more powerful RAID system. However, there are other variables to consider when looking at solutions tied to host resources, such as host interrupt delays, array management, data integrity, and processor intensive server applications that can burden host resources. RAID reliability can also be compromised with software RAID, as the RAID system is vulnerable to an operating system corruption or crash.

Hardware RAID avoids these compromises by offloading RAID calculations and array management to a dedicated I/O processor and XOR engine with dedicated battery-backed cache memory. All of the RAID functions are performed by the controller hardware, freeing the host CPU and other resources for user applications.

Some hardware RAID controllers are more effective at offloading the host system resources. Intel® RAID controllers are designed to maximize the offload capabilities built into the on-board processor. The following charts outline the measured benefits of Intel® RAID controller, which enables reduced privilege mode CPU time, interrupt mode CPU time, CPU effectiveness (which is the amount of CPU cycles attributed to each I/O), and interrupts per I/O. All of these reduce available resources for handling your mission critical applications.

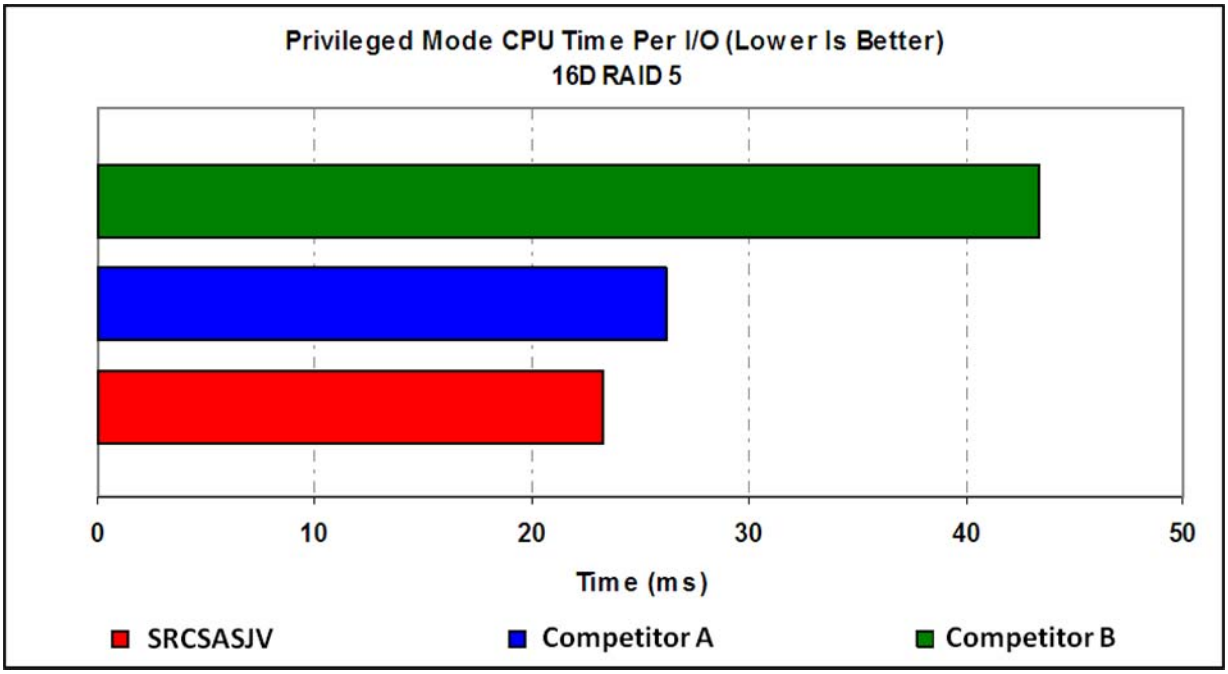


Figure 5. Sixteen-drive RAID 5, 64 KB sequential writes, privileged mode CPU Time per I/O (in milliseconds)

Privileged mode CPU time is the amount of time spent in a high-priority privileged mode for each I/O; lower is better*.*

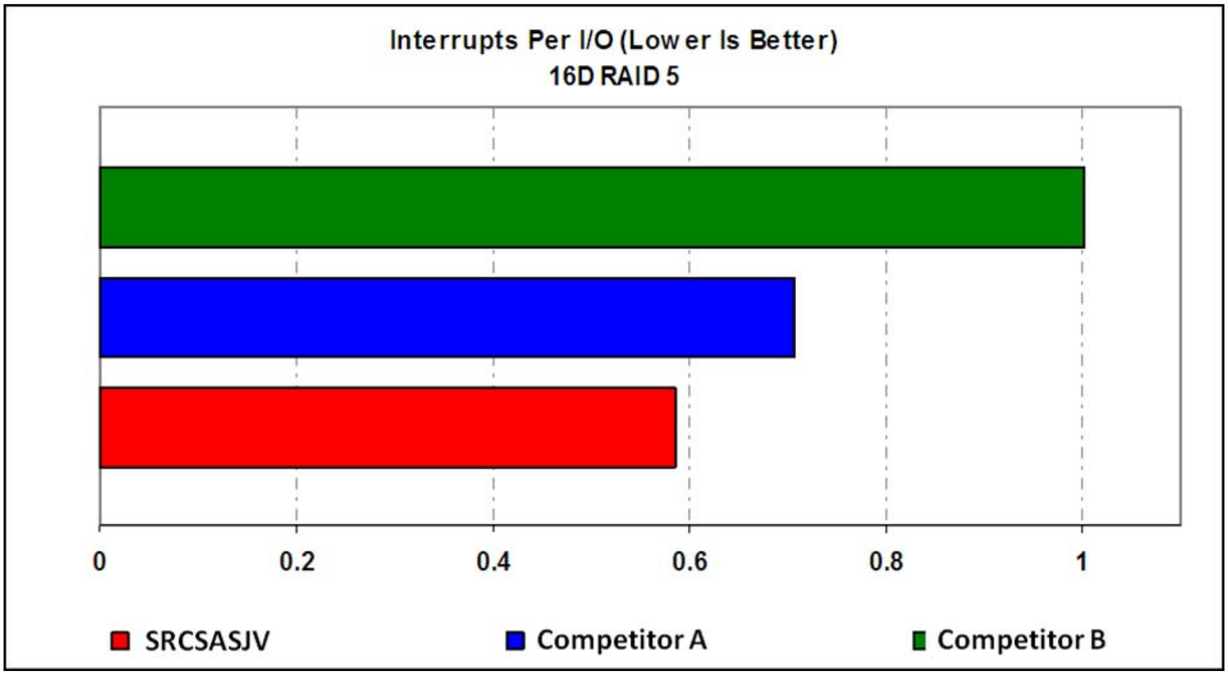


Figure 6. Sixteen-drive RAID 5, 64 KB sequential writes, Interrupts per I/O

Interrupts require additional overhead by causing the processor to discontinue its current work to save its state, incur a context switch and execute the interrupt handler; lower is better.

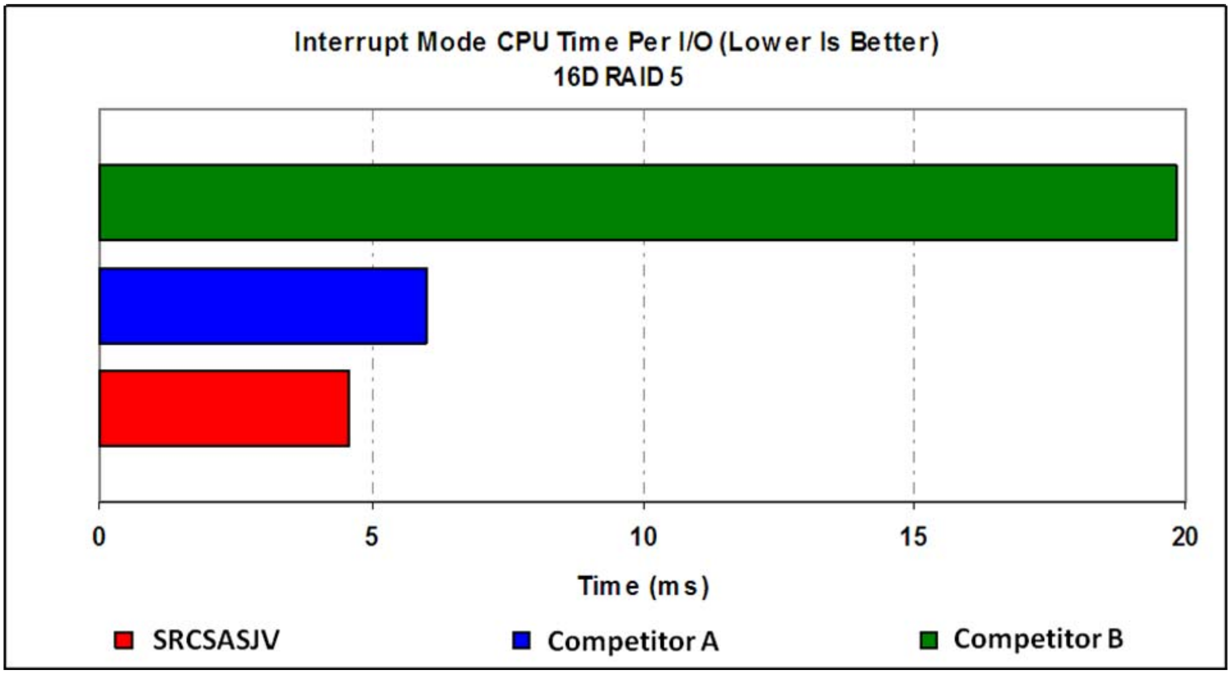


**Figure 7. Sixteen-drive RAID 10, 4 KB random Interrupt mode CPU Time per I/O (in ms)**

Interrupt mode CPU time is the amount of time that the processor spends handling an interrupt; lower is better.

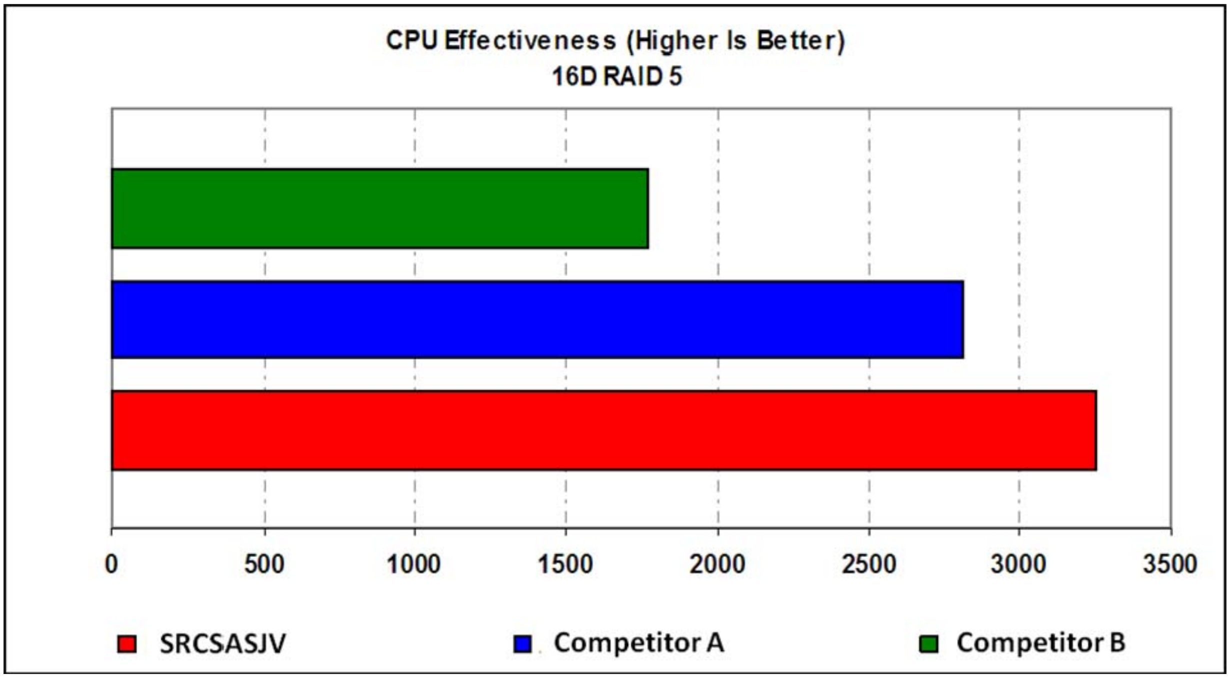


**Figure 8. Sixteen-drive RAID 10, 64 KB sequential writes, CPU effectiveness**

CPU effectiveness is measured as the % CPU utilization divided by the I/Os per second. A higher number indicates that I/O requests can be handled with fewer CPU cycles; higher is better.

# 3. Best Practices

Because each server environment is unique, configuring your RAID storage is not a one-size-fits-all recipe. Determining the best RAID type to create an effective storage subsystem requires careful consideration of your requirements and multiple tradeoffs. The following sub-sections provide guidance into architecting a RAID solution.

## 3.1 Balancing Capacity, Fault Tolerance, and Performance

One of the first things to consider is capacity requirements. Fault tolerant RAID arrays require forfeiting some of the usable storage capacity to store either redundant data or parity data. The following table shows how to calculate the usable capacity for the most common RAID types.

| **RAID Type** | **1** | **10** | **5** | **6** |
|---|---|---|---|---|
| **Minimum Disks Needed** | 2 | 4 | 3 | 4 |
| **Usable Capacity of Array with N disks** | $\frac{N}{2}$ | $\frac{N}{2}$ | $N - 1$ | $N - 2$ |

**Table 1. Usable capacity calculations for common RAID types (assumes equally sized disk capacity)**

The ideal RAID solution consists of balancing trade-offs between fault tolerance, cost, and performance requirements. A RAID 10 may require forfeiting 50% of your capacity; however, it can sustain a loss of one disk from each redundant leg, thus increasing the Mean Time To Data Loss (MTTDL) and minimizing server downtime (see Figure 8). If data availability is your primary concern, the lower performance of a RAID 6 or the lower MTTDL of a RAID 5 may be worth it, especially for applications such as Video Surveillance systems where sequential performance is almost indistinguishable. Given the diverse options available in the Intel® RAID products, it is useful to have a clear understanding of the storage workload environment in order to make the best decision.

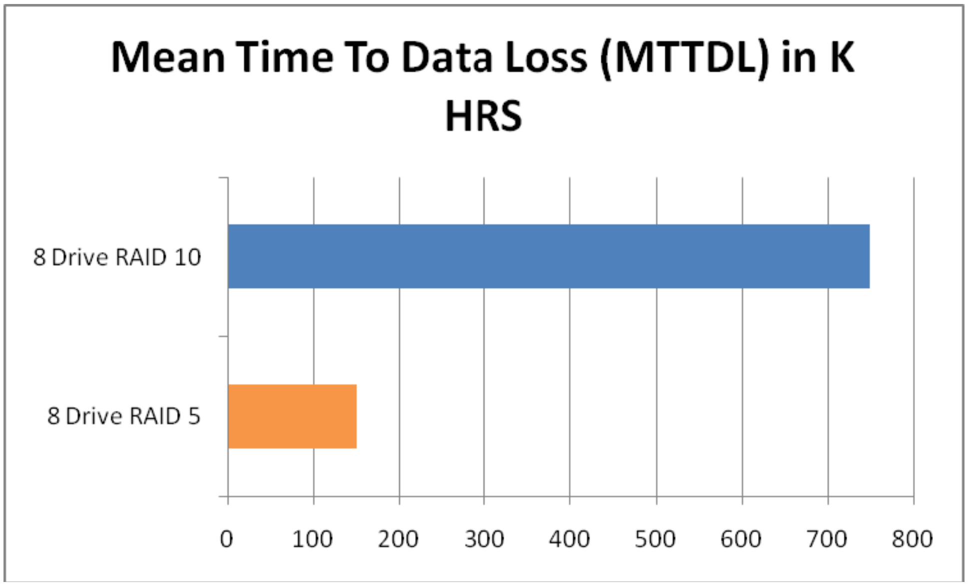


**Figure 9. Relative Mean Time To Data Loss (MTTDL) comparisons for RAID 10 and RAID 5 RAID types**

**Note:** Eight drives assume ideal rebuild times, immediate failed detect times, (disk MTBF of 1.4 million hours).

Because the typical database transaction produces twice as many read requests than write requests, RAID 10 is the ideal RAID type to use for database storage appliances requiring redundancy and high availability. Although mirror-based RAID types require a copy of each write, requiring two disk I/O's for every write, it still offers higher performance as it does not incur the read-modify-write penalty associated with RAID 5 and RAID 6.

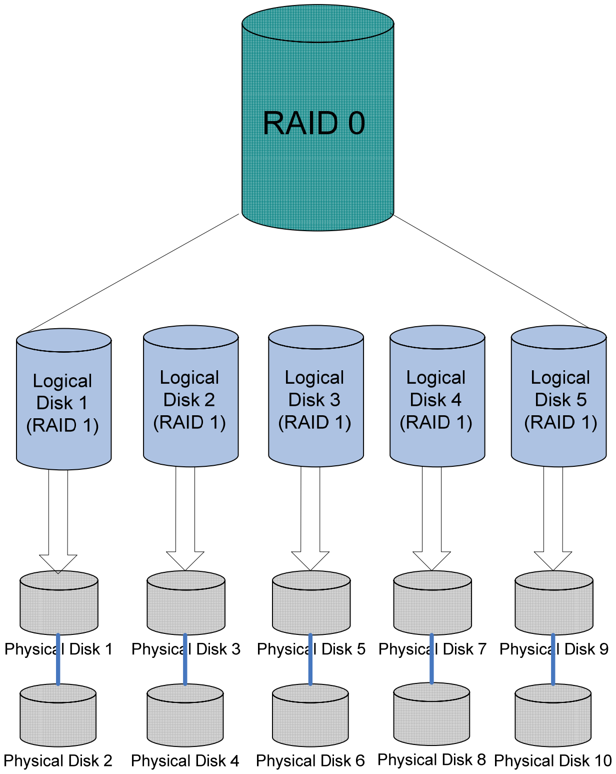


**Figure 10. Physical to logical data layout for RAID 0+1**

Stripe size can affect overall performance, and most RAID vendors will make suggestions on selecting the optimal size. For primarily random applications, in general it is best to avoid stripe breaks, which are defined as I/O requests that cross the stripe boundary between physical disks. When this happens, the RAID controller must dispatch two separate disk requests resulting in two disk seeks in order to complete the command. Because of this, it is generally better to use a stripe size sufficiently large enough to reduce the probability of stripe breaks.

Intel® RAID controllers provide other ways to tune and optimize your RAID environment by offering two data placement policies – Cached and Direct I/O. Cached I/O will move all reads through the cache, while direct I/O will transfer read requests directly from host memory to disk. For workloads with sufficient temporal and spatial localities, cached I/O can provide significant performance benefits.

Intel® RAID controllers offer read cache options, which allow you to customize your read cache policies for your application. By disabling read cache for applications that never or rarely generate predictable sequential read patterns, you can preserve your cache and reduce the overhead. ‘Adaptive’ and ‘Always’ are two other options that help you to fine tune the amount of data read into cache.

When read ahead always is selected, it enables read ahead capability, which allows the controller to read sequentially ahead of requested data and to store the additional data in cache memory, anticipating that the data will be needed soon. This speeds up reads for sequential data, but there is little (or no) improvement when accessing random data.

Adaptive read ahead allows the controller to use read ahead if the two most recent drive accesses occurred in sequential sectors. If the read requests are primarily random, the controller dynamically reverts to Normal (no read ahead).

Enabling cache on disk drives can significantly improve performance, but it is important to note that the potential for data loss occurs if power to the disk drives is lost while data written to the disk has not yet been committed to the medium.

Native command queuing (NCQ) for SATA and tagged command queuing (TCQ) for SAS hard drives also increases performance by allowing individual hard disks to optimize the order in which read and write commands are executed. As additional I/O operations arrive from different sources in a multi-stream, multi-I/O environment, more disk head "thrashing" can occur. To help mitigate this issue, NCQ and TCQ algorithms allow I/O operations to be performed out of order to optimize and leverage disk read/write head positioning, thus ultimately increasing overall performance.

As mentioned in section “2.5 Hardware RAID Advantage”, dedicated cache on the RAID controller dramatically improves disk write performance, particularly when the controller is configured with Write Back Cache enabled. However, Intel recommends battery backup for controller cache when this mode is selected in order to protect data still in controller cache in the event of a power failure.

## 3.2 Tuning Controller Cache Options

Optimizing the overall performance of a RAID subsystem requires careful consideration of several factors that can affect performance, including the controller and disk drive cache settings and the interaction of these settings with system applications. The following sections provide a limited discussion of some of these factors.  For a full review of performance tuning, please refer to the *Intel® RAID Controller Performance Optimization White Paper*, available at http://support.intel.com/go/raid.

**Note:** There are a variety of factors that can affect the performance of the RAID subsystem, including PCI bus bandwidth, logical drive cache settings, stripe size, hard disk drive cache settings, RAID level, ratio of read versus write operations, ratio of sequential versus random operations, and the number of disks in an array.

Tuning cache memory options on the RAID controller can improve performance. There are three settings available in the controller cache to allow fine tuning:

- Read Ahead Option
- Write Back Option
- Cached I/O Option

The following table provides a quick reference for RAID settings. The information is simplified and may not be accurate with some applications or tests.  For detailed performance tuning information, please refer to the *Intel® RAID Controller Performance Optimization White Paper*, available at http://support.intel.com/go/raid.

| | |
|---|---|
| **Read Cache Policy** | Direct I/O |
| **Read Ahead Policy** | Adaptive Read Ahead |
| **Write Cache Policy** | Write Back* |

* A RAID controller battery should be used whenever virtual drive write-back cache is enabled and data is mission critical.

## 3.3 Hard Disk Cache

Disk drive cache can be enabled in the virtual drive properties page of the RAID configuration utility. There is a risk of data loss using a hard drive cache; an overview is provided below.

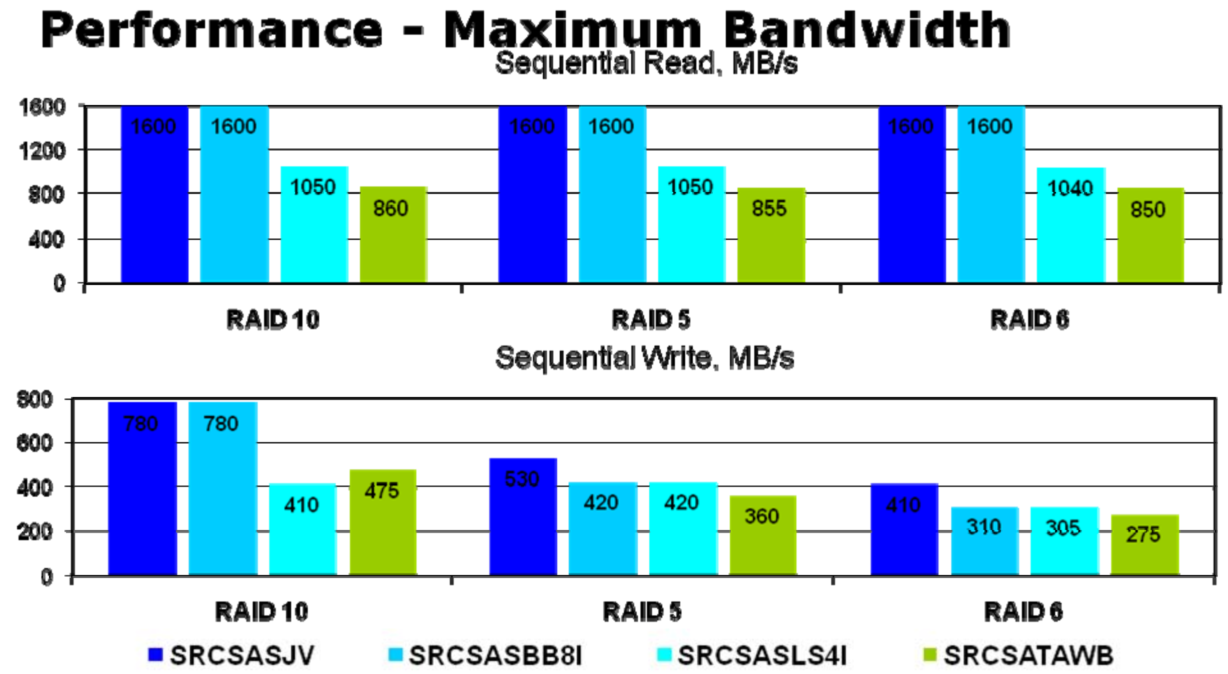
Hard disk drive cache is located within the logic of the hard drive. Cache provides enhanced performance for sequential read access by retrieving adjacent data on the drive into the data buffer in case the host computer requests it. This process allows the data to be directly transferred from the drive’s memory when it is requested rather than waiting for a disk access, which results in lower latency. Enabling the hard drive cache can also improve write performance by providing additional memory space for queued data. Write data can be queued in the disk cache and reported as written even though the data will not move from memory to the disk until disk access is available. This reduces the delay during disk I/O operations.

There is an inherent risk in holding data in the drive cache when a write has been acknowledged as complete but has not been written to the disk. If the drive loses power, the data in the cache will be lost before it is written to the disk. This can cause a “hole” in the data file, which makes the file unusable. Using a UPS will mitigate this risk but not eliminate it.

**Note:** A soft or hard reset (<Ctrl> + <Alt> + <Del> or the reset button) does not affect the completion of a disk write operation because the disk cache will be flushed as long as drive power is maintained.

# 4. Intel RAID Maximum Throughput

Traditionally, customers have turned to RAID 5 to give them the data protection and performance that they required and have thought of performance in terms of maximum throughput in megabytes per second through the controller. These measurements usually center in sequentially reading from or writing to the virtual drive and while closely matching the data transfer sizes with the optimal operating system transfer size, and closely aligning the stripe size of the RAID controller. The following figure depicts the maximum throughput of four of Intel's popular RAID controllers in megabytes per second.



Industry leading read performance across all RAID levels!
Industry leading RAID 10 performance – read and write!

Data source: Intel internally measured results as of 23 May 2008. See 'Test Configurations' slide for test configuration details.

**Figure 11. Maximum Bandwidth**

Although, most applications are comprised of a significant random component, performance in these applications rely heavily on the I/O handling capability of the controller. These measurements are measured in I/O per second (IO/s). Applications such as Microsoft Exchange, Web and SQL servers, can take advantage of the speed and data protection of RAID 10. Since RAID 10 combines striping and mirroring to produce large virtual disks with high performance and fault tolerance, the performance gain comes from striping across mirror sets without the need for parity calculations. The following figures depict test results from these workloads.

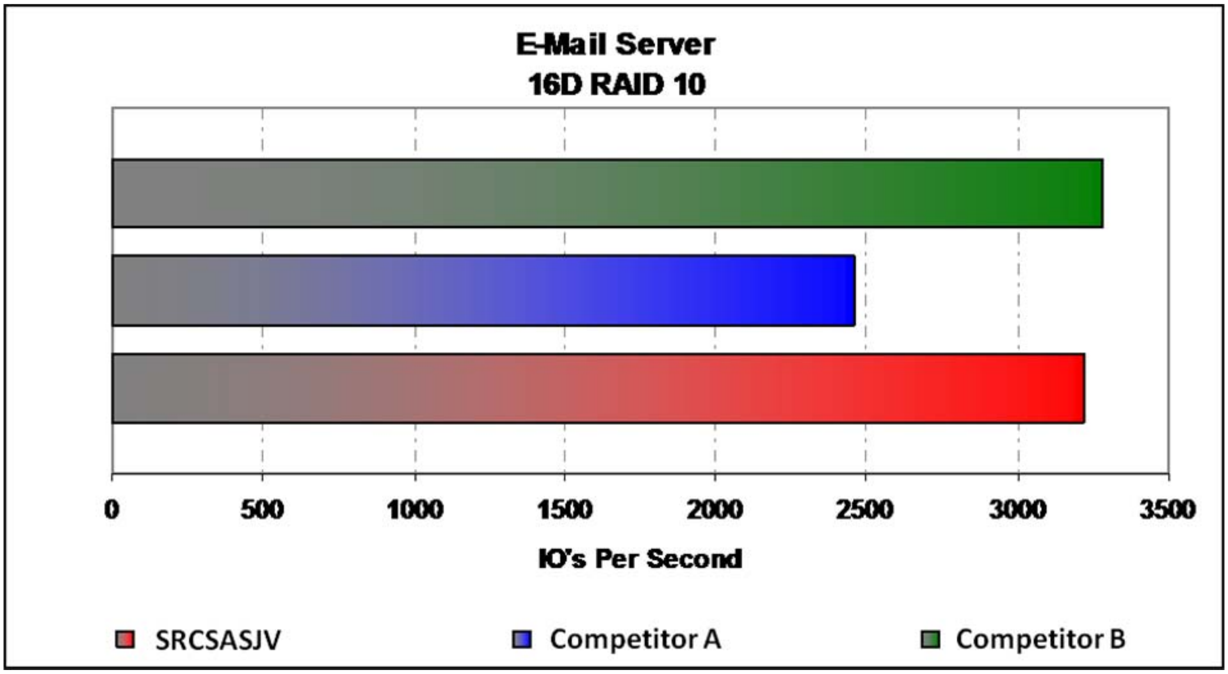


Figure 12. Sixteen-drive RAID 10, E-mail server performance at queue depth of 256

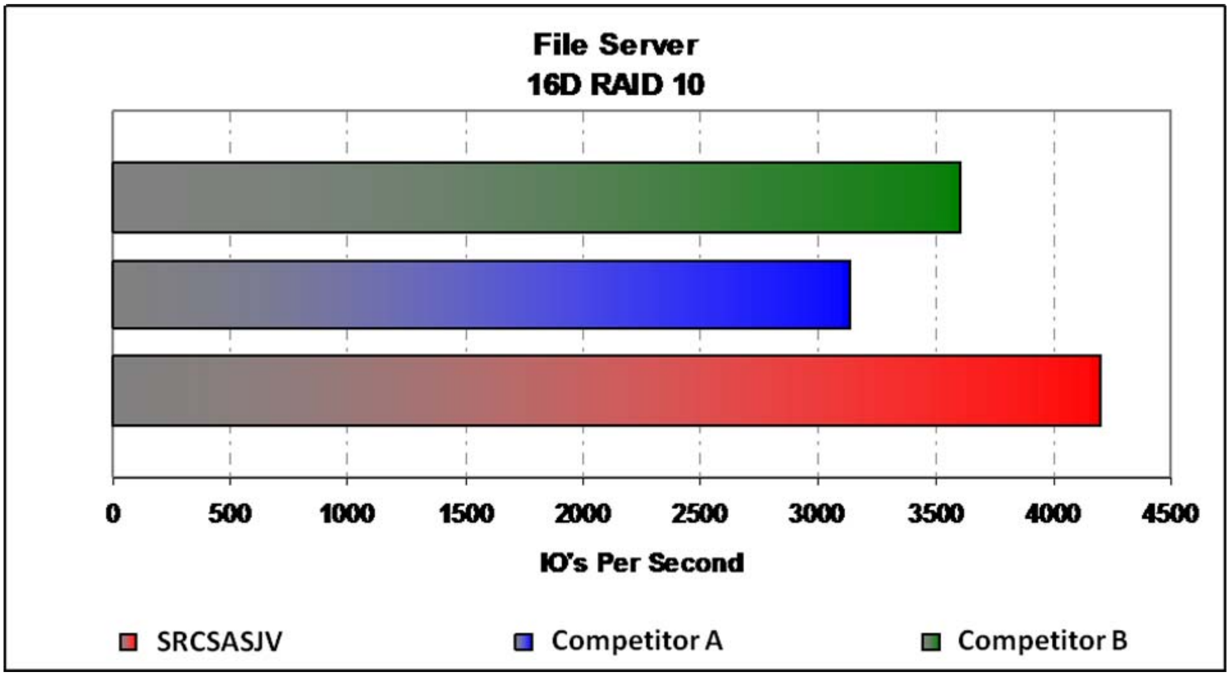


Figure 13. Sixteen-drive RAID 10, File server performance at queue depth of 256

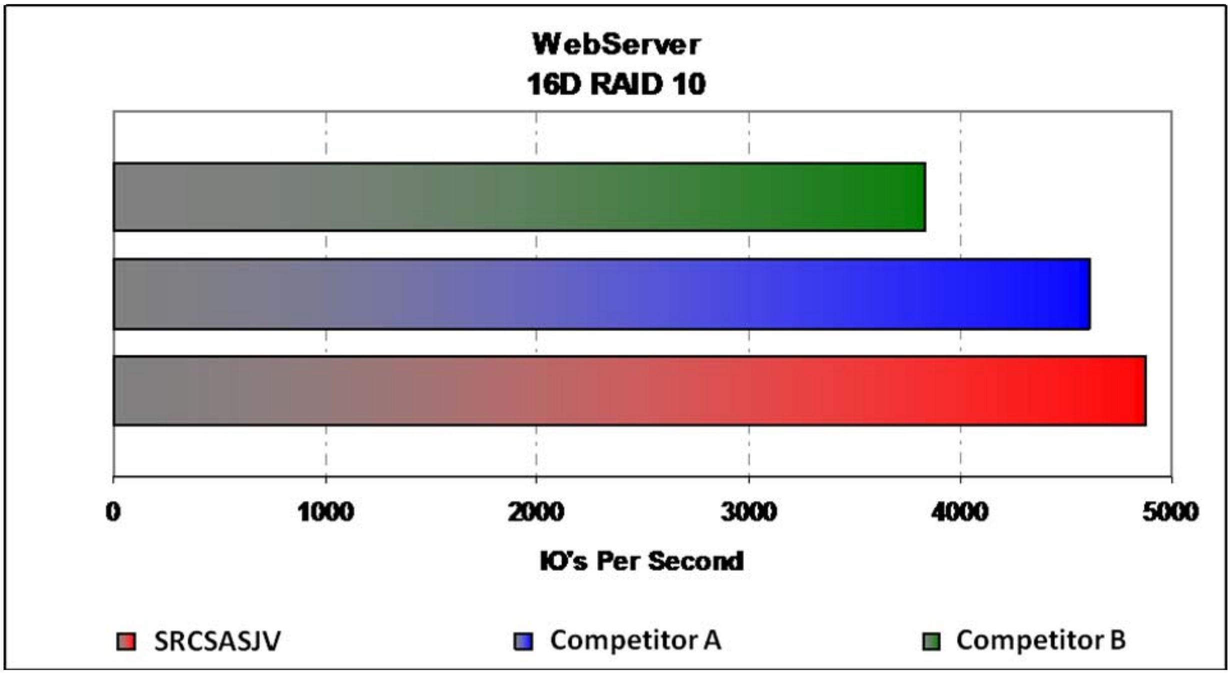


Figure 14. Sixteen-drive RAID 10, Web server performance at queue depth of 256

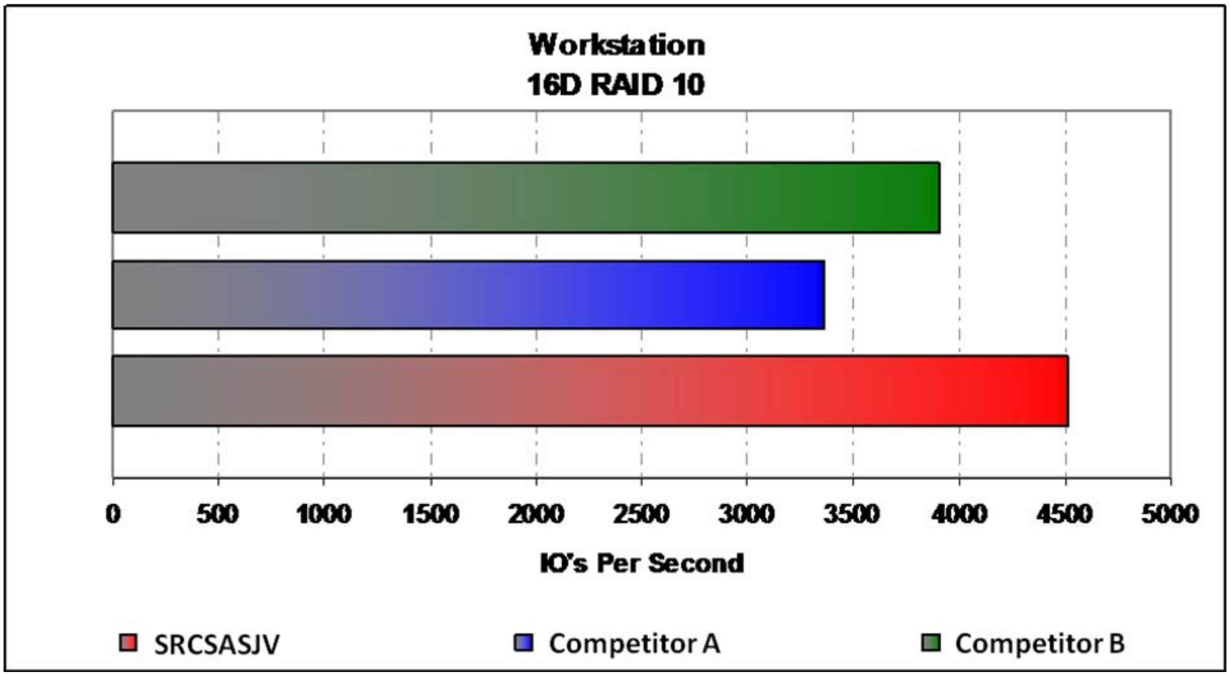


Figure 15. Sixteen-drive RAID 10, Workstation performance at queue depth of 256

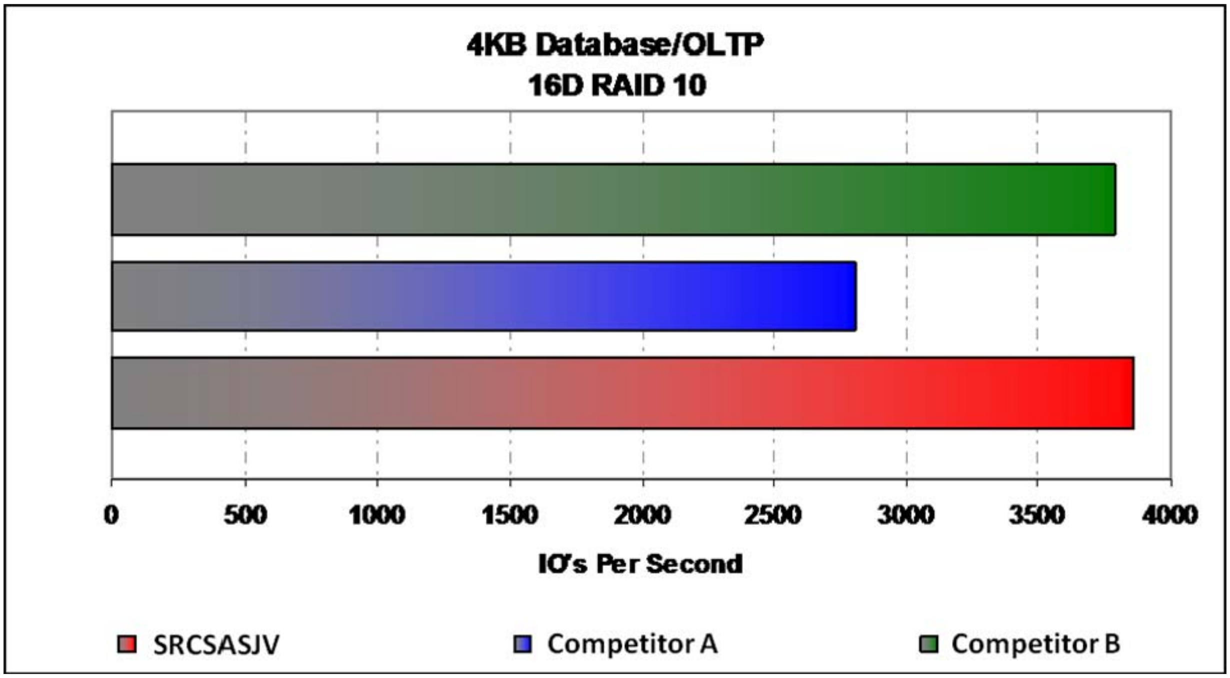


Figure 16. Sixteen-drive RAID 10, Online transaction database performance at queue depth of 256

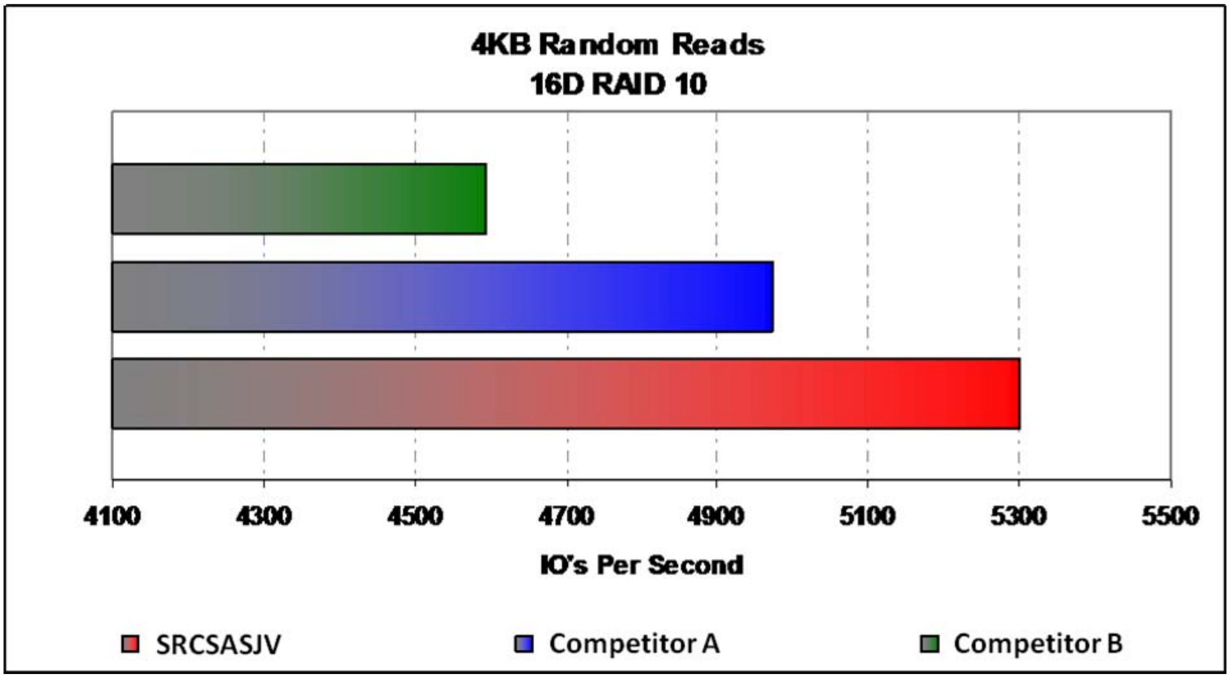


Figure 17. Sixteen-drive RAID 10, 4 KB random read performance at queue depth of 256

RAID 6 also offers a large benefit, especially in an environment utilizing large SATA drives that contain critical data. Due to the long rebuild times of these large SATA drives and the risk that a second drive may fail during this rebuild time (an event that would be catastrophic in a RAID 5 configuration), RAID 6 provides a good alternative. The benefit of RAID 5 is that only one drive is lost to redundancy while RAID 10 requires the loss of 50% of capacity for redundancy. RAID 6 offers the loss of the capacity of two drives to redundancy, which provides an efficient and safe alternative to RAID 10 and RAID 5 solutions. The drawback to both RAID 5 and RAID 6 is slower write performance due to the time it takes to calculate the parity stripes, with RAID 6 adding an additional loss of performance compared to RAID 5.

Utilizing the formulas found in table 1 above, the following is an example of capacity available in a configuration using six drives with 1TB capacity:

- RAID 10 would yield 3 TB of capacity
- RAID 5 would yield 5 TB of capacity
- RAID 6 would yield 4 TB of capacity

# 5. Summary

When architecting a RAID solution, careful consideration should be given to factors that will allow for the best performance of the intended application. This document has described several “Real World” scenarios and benchmarks for consideration as good representatives of performance within these scenarios.  Using the Intel® RAID Controller SRCSASJV for benchmarks and comparison, this white paper conveys the message that Intel® RAID Controllers powered by LSI MegaRAID* technology offer exceptional performance.

This paper also offers guidance for selecting RAID levels and configuring your RAID controller for a balance of performance and fault tolerance.

Intel highly values the importance of protecting your business data, optimizing your server resources and enabling the highest performance for your RAID solution. Intel offeres a wide range of RAID products making it possible to architect a highly optimized and scalable SAS solution specifically for your environment. By utilizing advanced technology, world-class features, and a SAS design that allows for both SAS and SATA hard disks, Intel RAID allows you to balance performance, reliability and cost requirements.

# 6. Testing Disclosure and Methodology

IOmeter version 7.26.2006 was used as a workload generator. All RAID controllers were used in an identical system as described in the following table. Logical RAID arrays were allowed to fully initialize prior to benchmarking. All software packages were updated to the latest publicly available versions. File systems were not created on the logical volumes under test (this eliminates file system overhead and caching). System boot drive is controlled by the on-board SATA. When performing storage benchmarking, all screensavers and power energy modes were disabled. Current production controller drivers were used, disk cache was disabled, and controller cache options were enabled for consistency.

| System Configuration | |
|---|---|
| Processor | Intel® Xeon® DP 5080 CPU 3.73 GHz |
| Motherboard | Intel® Server BoardS5000XVN/XSL |
| Memory | 4 GB |
| System Interconnect | Slot 6 PCI Express* X8 |
| System BIOS | S5000.86B.10.00.0087 |
| Type: | Seagate SAS |
| Drive Model: | ST373455SS (write cache disabled) |
| Drive FW: | 2 |
| Drives Size: | 73 GB |
| # of drives | 16 |
| Type: | 2 External 12 drive Enclosures |
| Enclosure Model | SYM-3600 |
| Enclosure FW | 193 |
| OS | Windows |
| Version | 2003 SP2 |
| Type | 32-bit Enterprise Edition |
| IOmeter Version | 7.26.2006 |

**Table 2. System Configuration**