Performance Analysis
July 2011

Performance Analysis Update for the ServeRAID M5025 SAS/SATA Controller Using 6Gbps SAS HDDs

Beverly Clapp
Advisory Engineer
System x Server/Storage Performance Development & Design Guidance
IBM Systems and Technology Group

Introduction

This paper is intended for anyone who are interested in the baseline performance of the ServeRAID® M5025 SAS/SATA Controller, a high-performance RAID controller that supports 6Gb SAS, 3Gb SAS and SATA II hard drive technologies.

This paper is an update to the previous version entitled "Performance Analysis of the IBM ServeRAID-M5025 SAS/SATA Controller," which was published December 2010.[1]

The ServeRAID M5025 SAS/SATA Controller is a cost-effective, enterprise-grade RAID solution for external hard drives (HDDs) that integrates emerging SAS technology into an organization's storage infrastructure. At 6Gbps, the M5025 offers improved performance over its predecessors. The ServeRAID M5025 SAS/SATA Controller supports SAS and SATA hard-drive-redundant configurations for server storage, thereby providing investment protection to our clients. The M5025 is ideal for supporting server mission-critical applications where high levels of sustained read and write operations are required, such as medical imaging, video streaming, Web content, video-on-demand, security and surveillance, fixed content, and reference data storage.

The purpose of this paper is to present the performance results obtained using the Iometer and fio benchmark tools to measure the performance of the ServeRAID M5025 in Microsoft® Windows® Server 2008 R2 and Linux® environments, respectively.[2] The ServeRAID M5025 controller's performance using 6Gbps SAS hard drives is compared to ServeRAID M5025 controller's performance using 3Gbps SAS hard drives. The performance data using the 3Gbps SAS hard drives was taken from the data used in the creation of the previous version of this white paper from December 2010 as stated above.

The paper is organized in four sections. The first section briefly describes the tools used to measure the performance of the ServeRAID M5025 and defines the workloads used in the measurements. The second section describes the hardware and software measurement environment. The third section presents the results of the measurements and explains how the results should be interpreted. Finally, the fourth section presents guidance on how the ServeRAID M5025 product should be positioned from a performance perspective.

Important lessons learned from this performance study are highlighted in boxes at appropriate points throughout the paper.

Questions about the information presented should be directed to the author at bclapp@us.ibm.com.

[1] http://public.dhe.ibm.com/common/ssi/ecm/en/xsw03092usen/XSW03092USEN.PDF

[2] The measurement results in this paper represent data that was written to disks or read from disks. The results do not represent data that was read strictly from RAID controller cache or written strictly to RAID controller cache. While both methods produce valid data, the "out-of cache" or "to-cache" measurements do not fit within the scope of this document.

Measurement Tool and Workloads

Iometer Tool

Iometer is a workload generator and a measurement tool originally developed by Intel ® Corporation. It is now maintained under an Intel Open Source License, and it is available at http://sourceforge.net.

Iometer is designed to generate workloads and record measurement results for server disk and network subsystems—*not* desktop disk and network subsystems. In this context, the use of the words "server" and "desktop" is not a trivial matter. Consider the following example.

The single-threaded utility *copy* is routinely used to test server disk subsystems. The *copy* utility is a suitable benchmark for a laptop or desktop machine, but not for a server. Why then is it used so often for measuring server disk subsystem performance? It is probably used for two reasons. First, *copy* is easy to execute, and does not require large amounts of resources. The second reason is that the differences between server architecture and desktop architecture may not have been understood by the people implementing the benchmark.

Desktop machines are designed to manage one task at a time, and they do this very well. In fact, when *copy* is executed, a desktop machine with a single hard drive will usually perform better than a server with an array of multiple drives. The reason for the performance disparity is based on the design differences of the two machines. Servers are designed to handle multiple tasks in parallel. Since *copy* is single-threaded, each I/O request must be satisfied before another I/O request can be generated. Therefore, the multiple-drive array is not being used efficiently, because only one drive is required to satisfy each I/O request.

One way to measure the performance of a server disk subsystem is to use Iometer. With Iometer, multiple I/O requests can be issued in parallel so that all of the drives in an array can be kept busy in a way similar to how it is done by a high-performance SMP server application. Iometer also provides a configurable parameter called "outstanding I/Os," which can be used to increase the load on a server disk subsystem in a Windows environment. The Windows measurement results contained in this paper were generated by increasing the number of outstanding I/Os queued at the drives up to and beyond what would be typical in a production environment. For a copy of the workload scripts used for these measurements, please contact the author at the e-mail address provided in the Abstract.

Do not use desktop-oriented tools or single-threaded utilities, such as *copy*, to measure the performance of a server's disk subsystem. Iometer is specifically designed to generate workloads on servers that issue I/O requests in parallel to the disk subsystem.

The measurement results in this paper were obtained using Iometer version 2008.06.18-RC2, Copyright 1996-1999 Intel Corporation. Intel does not endorse any Iometer results.

The workloads used to yield the results in this document were the On-Line Transaction Processing workload, Streaming Reads workload, Single-Threaded Sequential Read workload, Streaming Writes workload, Single-Threaded Sequential Write workload, Random Reads workload, and the Random Writes workload. The characteristics for each workload are described in the following sections.

On-Line Transaction Processing Workload

The On-Line Transaction Processing (OLTP) workload emulates a transactional database workload. It is defined as 100% random accesses, 67% reads, and 33% writes. This workload is measured using transfer request sizes of 4K, 8K, 16K, 32K, and 64K. The number of outstanding I/Os linearly steps from 1 to 121 outstanding I/Os per target.

Streaming Reads Workload

The Streaming Reads workload emulates a read-intensive multimedia streaming application. It is defined as 100% sequential accesses and 100% reads. This workload is measured using transfer request sizes of 32K, 64K, 128K, 256K, 512K, 1M, and 2M. The number of outstanding I/Os linearly steps from 1 to 121 outstanding I/Os per target.

Single-Threaded Sequential Reads Workload

The Single-Threaded Sequential Reads workload emulates the read portion of a single-threaded file copy benchmark. It is defined as 100% sequential accesses, 100% reads. This workload is measured using transfer request sizes of 64K, 128K, 256K, 512K, 1M, and 2M. The number of outstanding I/Os is fixed at 1 outstanding I/O per target.

Although single file copy benchmarks do not typically represent server workloads, some customers still run these types of benchmarks, so it is valuable to understand how these products will perform in these benchmarks.

Streaming Writes Workload

The Streaming Writes workload emulates a write-intensive multimedia streaming application. It is defined as 100% sequential accesses and 100% writes. This workload is measured using transfer request sizes of 32K, 64K, 128K, 256K, 512K, 1M and 2M. The number of outstanding I/Os linearly steps from 1 to 121 outstanding I/Os per target.

Single-Threaded Sequential Writes Workload

The Single-Threaded Sequential Writes workload emulates the write portion of a single-threaded file copy benchmark. It is defined as 100% sequential accesses, 100% writes. This workload is measured using transfer request sizes of 64K, 128K, 256K, 512K, 1M, and 2M. The number of outstanding I/Os is fixed at 1 outstanding I/O per target.

Although single file copy benchmarks do not typically represent server workloads, some customers still run these types of benchmarks, so it is valuable to understand how these products will perform in these benchmarks.

Random Reads Workload

The Random Reads workload is defined as 100% random accesses and 100% reads. This workload is measured using transfer request sizes of 4K, 8K and 16K. The number of outstanding I/Os linearly steps from 1 to 121 outstanding I/Os per target

Random Writes Workload

The Random Writes workload is defined as 100% random accesses and 100% writes. This workload is measured using transfer request sizes of 4K, 8K and 16K. The number of outstanding I/Os linearly steps from 1 to 121 outstanding I/Os per target.

The fio Benchmark Tool

The fio tool is an I/O workload generator used to benchmark storage subsystem performance. It works on both block devices as well as file systems. It supports several different types of I/O engines, including synchronous and asynchronous I/O, and multi-thread I/O.

Because fio can be run on a single, stand-alone Linux system, it was used for the Linux measurements instead of Iometer, which requires two networked systems to run under Linux.

For a description of the benchmark, see the Web site:

http://www.linux.com/archive/feature/131063

The latest version of fio can be downloaded from the Web site:

http://freshmeat.net/projects/fio

Shell scripts were developed to automate fio measurement collection. Additionally, multiple job files were created using fio to simulate benchmark workloads on SUSE Linux Enterprise Server 10 SP3 similar to those used for Iometer benchmarking on Microsoft Windows Server 2008 R2 Enterprise Edition.

Those job files included the On-Line Transaction Processing (OLTP) workload, Sequential Reads workload, Sequential Writes workload, Random Reads workload and Random Writes workload.

For example, the OLTP workload job file looks like this:

```
[global]
rw=randrw
blocksize=${BLKSIZE}
blockalign=${BLKSIZE}
size=45000m
ioengine=libaio
iodepth=${DEPTH}
rwmixread=67
rwmixwrite=33
direct=1
invalidate=1
time_based
runtime=30s
[/dev/sdc1]
```

Measurement Environment

ServeRAID M5025 measurements were conducted using the IBM System x®3690 X5 with one eight-core Intel Xeon® Processor Model X7560 (2.27GHz) and 32GB of system memory.

Microsoft Windows Server 2008 R2 Enterprise x64 Edition was installed on the system for all ServeRAID M5025 Windows measurements.

SUSE Linux Enterprise Server 10 (SLES10) with SP3 (64-bit) was installed on the system for all ServeRAID M5025 Linux measurements.

The system contained one ServeRAID M5025 controller using Windows driver version 4.32.0.64 and Linux driver version 00.00.05.35.

ServeRAID M5025 controller firmware package version 12.12.0-0037 was used for all measurements.

The IBM System Storage™ EXP2524 storage enclosure used Product Revision level/ESM firmware version 546F.

The storage enclosure back-end held one hundred and twenty 15K rpm 6Gbps SAS hard drives.

Two of the EXP2524 storage enclosures were under-utilized with only 12 drives each in order to demonstrate a balanced configuration.

Configuration Diagram for the Measured Hardware

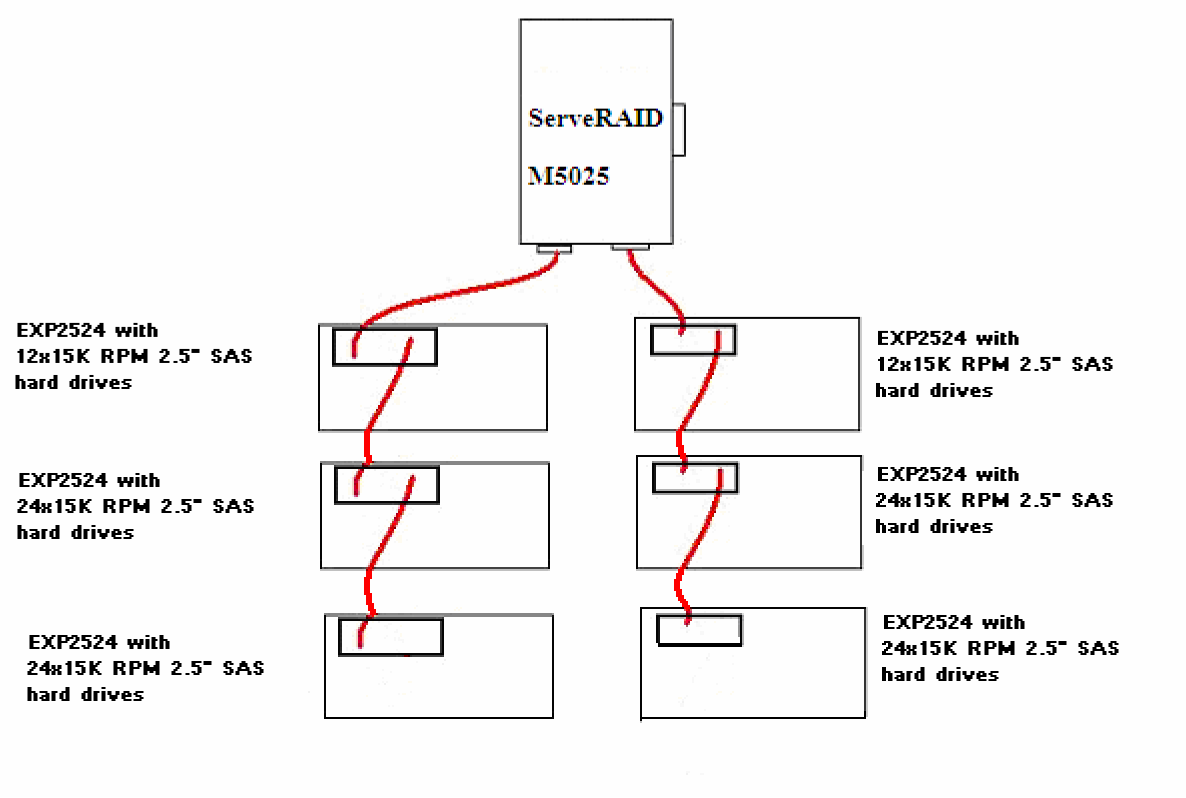


Figure 1. Hardware Configuration Diagram

Windows Measurement Results and Analysis

The performance information contained in this section was derived under specific operating and environmental conditions. The results obtained in your operating environments may vary significantly.

The measurement results in this section represent the maximum sustainable performance for a configuration with either twelve, twenty-four, forty-eight or one-hundred and twenty 15K rpm 6Gbps SAS hard disk drives (HDDs) at both an average and peak response times of approximately 15 milliseconds (ms). Peak performance typically refers to a measurement result with the highest number of IOps or MBps regardless of the average response time associated with that result. However, since most server applications will not wait forever for disk I/O to complete, the 15 ms threshold is a reasonable amount of time for an application to wait for completion of disk I/O before overall performance begins to decrease. For this reason, the measured performance at a 15 ms average response time should be considered as a more accurate representation of real-world performance.

RAID-5 Windows OLTP Workload Results

Table 1 contains RAID-5 measurement results for the OLTP workload for various transfer request sizes. For all configurations, the drives were configured in arrays of 12. In all configurations, only 8% of the total capacity of the drives was used. The workload was simultaneously applied to all arrays. This is true for all of the measurements unless otherwise noted. The write-back cache was enabled for all measurements. I/O policy was set to direct. The default 128K stripe size was used for all measurements.

Workload		OLTP 4K		OLTP 8K		OLTP 16K		OLTP 32K		OLTP 64K	
RAID-5		IOps	MBps	IOps	MBps	IOps	MBps	IOps	MBps	IOps	MBps
12 HDDs	15ms Average Response Time	2896	11.3	2636	20.6	2666	41.7	2396	74.9	2162	135.1
	Peak	3208	12.5	3294	25.7	3041	47.5	2869	89.7	2628	164.2
24 HDDs	15ms Average Response Time	5428	21.2	5344	41.7	5131	80.2	4390	137.2	4262	266.4
	Peak	6149	24.0	6064	47.4	5962	93.2	5550	173.4	5123	320.2
48 HDDs	15ms Average Response Time	9518	37.2	9484	74.1	9088	142.0	8753	273.5	7887	492.9
	Peak	10352	40.4	10400	81.2	9856	154.0	9525	297.7	8731	545.7
120 HDDs	15ms Average Response Time	22763	88.9	22321	174.4	21724	339.4	20513	641.0	18034	1127.1
	Peak	22773	89.0	22324	174.4	21774	340.2	20727	647.7	18090	1130.6

Table 1. ServeRAID M5025 RAID-5 OLTP IOps

Table 1 illustrates that the performance in OLTP workloads will continue to increase as drives are added to the ServeRAID M5025. The ServeRAID M5025 supports up to 216 drives, so for database workloads, drives can continue to be added until the performance is limited by the controller itself and not by the number of drives attached to the controller.

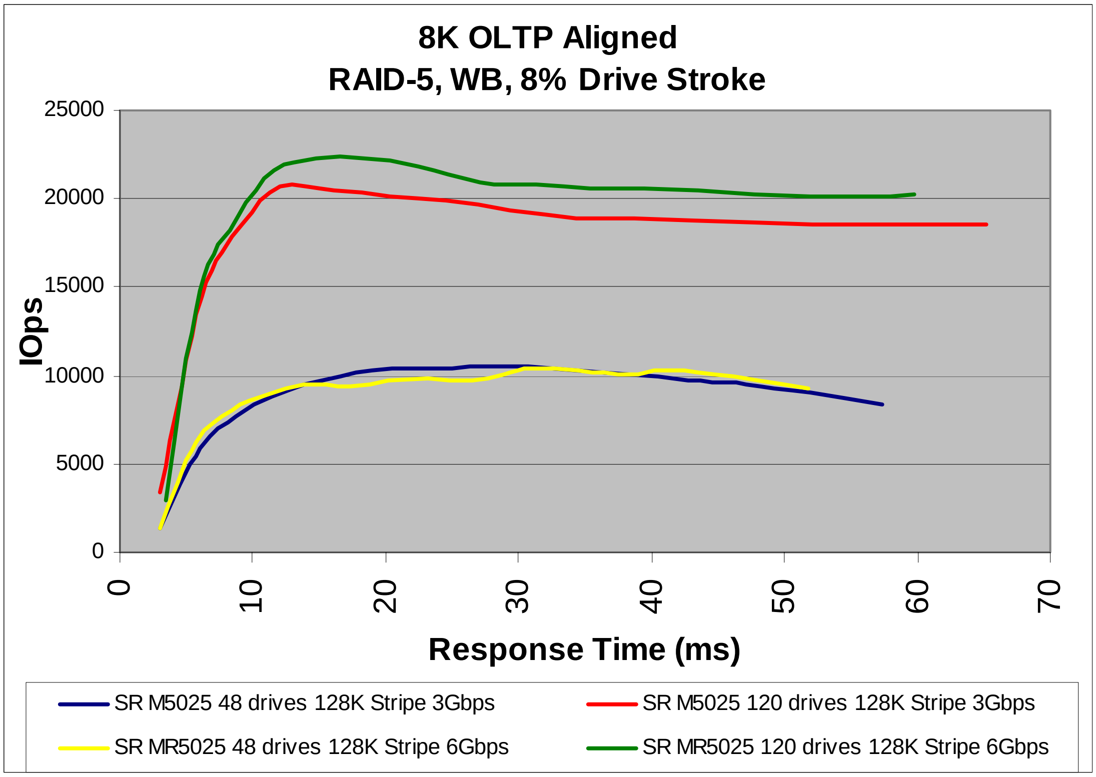


Figure 2. ServeRAID M5025 RAID-5 8K OLTP IOps 3Gbps vs. 6Gbps SAS HDDs

Figure 2 illustrates the relative performance advantage that the ServeRAID M5025 with 6Gbps HDDs has over the ServeRAID M5025 with 3Gbps HDDs in OLTP workloads. In 48-drive configurations, the peak OLTP performance difference is negligible. In 120-drive configurations, the peak OLTP performance of the ServeRAID M5025 with 6Gbps HDDs is 7.5% higher than that of the ServeRAID M5025 with 3Gbps HDDs.

RAID-5 Windows Streaming Reads and Streaming Writes Workload Results

Figure 3 illustrates the peak RAID-5 streaming reads performance of the ServeRAID M5025 with 6Gbps HDDs vs. the ServeRAID M5025 with 3Gbps HDDs as the transfer request size is increased. The peak performance for each transfer size is plotted on the chart without regard to response time.

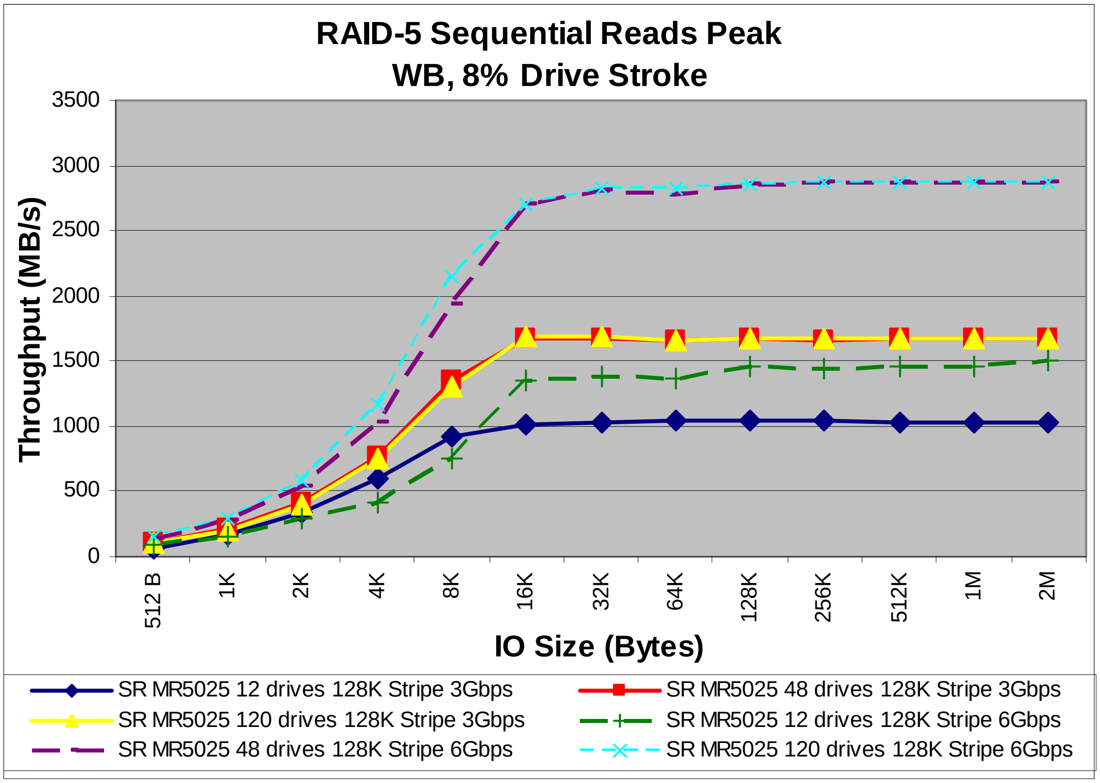


Figure 3. ServeRAID M5025 RAID-5 Streaming Reads Transfer Rate 3Gbps vs. 6Gbps SAS HDDs

One major observation can be made from Figure 3. The ServeRAID M5025 performance results with an end-to-end 6Gb SAS drive infrastructure (6Gb SAS drives and 6Gb SAS expanders) show much higher peak streaming read performance.

First, in the small 12-drive configuration, the peak RAID-5 streaming read performance of the ServeRAID M5025 with 6Gbps HDDs is 44% higher than that of the 3Gbps HDDs configuration. Both 12-drive configurations use x4-wide SAS connections between the ServeRAID M5025 and the storage enclosures. When using 3Gb HDDs, the x4-wide SAS connection is limited to a little more than 1,000 MB/s, so the 12-drive 3Gbps SAS configuration's performance is SAS-link-limited. When using 6Gbps HDDs, the x4-wide SAS connection is limited to a little more than 2,000 MB/s, so the 12-drive 6Gbps SAS configuration's performance is drive-limited.

Second, in the larger 48- and 120-drive configurations, the peak RAID-5 streaming read performance of the ServeRAID M5025 with 6Gbps HDDs is 71% higher than that of the 3Gbps HDDs configuration. Again, the 3Gbps HDDs configuration is SAS-link-limited. The 6Gbps HDDs configurations are controller-limited. For the ServeRAID M5025 controller to achieve peak, 6Gbps HDDs are required.

With 6Gbps drives used in conjunction with the ServeRAID M5025, peak RAID-5 streaming read performance can still be reached with smaller I/O block sizes (requires only a 16K block size to reach peak) as seen with the previous 3Gbps configuration.

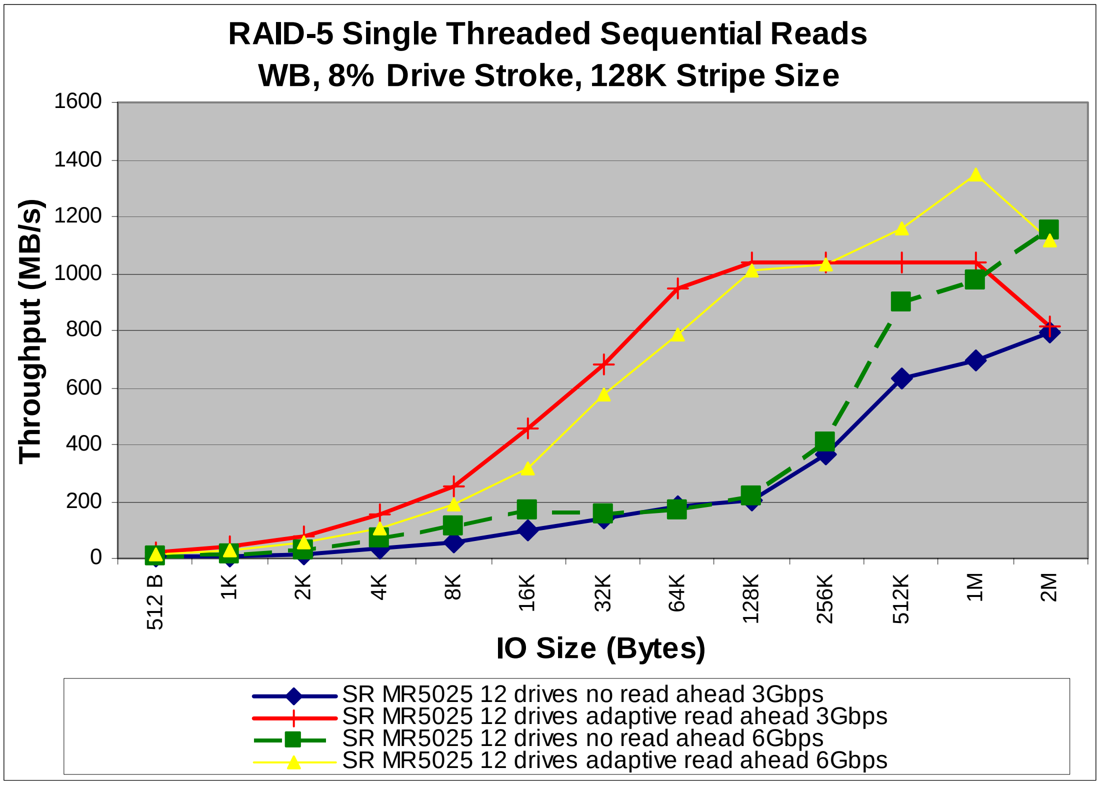


Figure 4. ServeRAID M5025 RAID-5 Single-Threaded Sequential Read Transfer Rate 3Gbps vs. 6Gbps SAS HDDs

Figure 4 compares RAID-5 single-threaded read performance of the ServeRAID M5025 with 6Gbps HDDs to that of the ServeRAID M5025 with 3Gbps HDDs. This comparison is of particular interest to those who use file copy benchmarks to evaluate performance. When read-ahead is disabled (default), the single-threaded sequential read performance of the ServeRAID M5025 with 6Gbps HDDs is 0% to 45% better than the ServeRAID M5025 with 3Gbps HDDs. Using the default read-ahead disabled setting will limit the single-threaded sequential read performance of both configurations.

In both configurations, the adaptive read-ahead setting was used for evaluating single-threaded read performance because it greatly improves single-threaded sequential read performance. When read-ahead is enabled, the single-threaded sequential read performance of the ServeRAID M5025 with 6Gbps HDDs is 0% to 30% better than that of the ServeRAID M5025 with 3Gbps HDDs.

The default non-read-ahead setting is higher-performing for the majority of server workloads. Enabling the adaptive read-ahead setting can dramatically improve performance for the single-threaded sequential read workloads common in file copy benchmarks

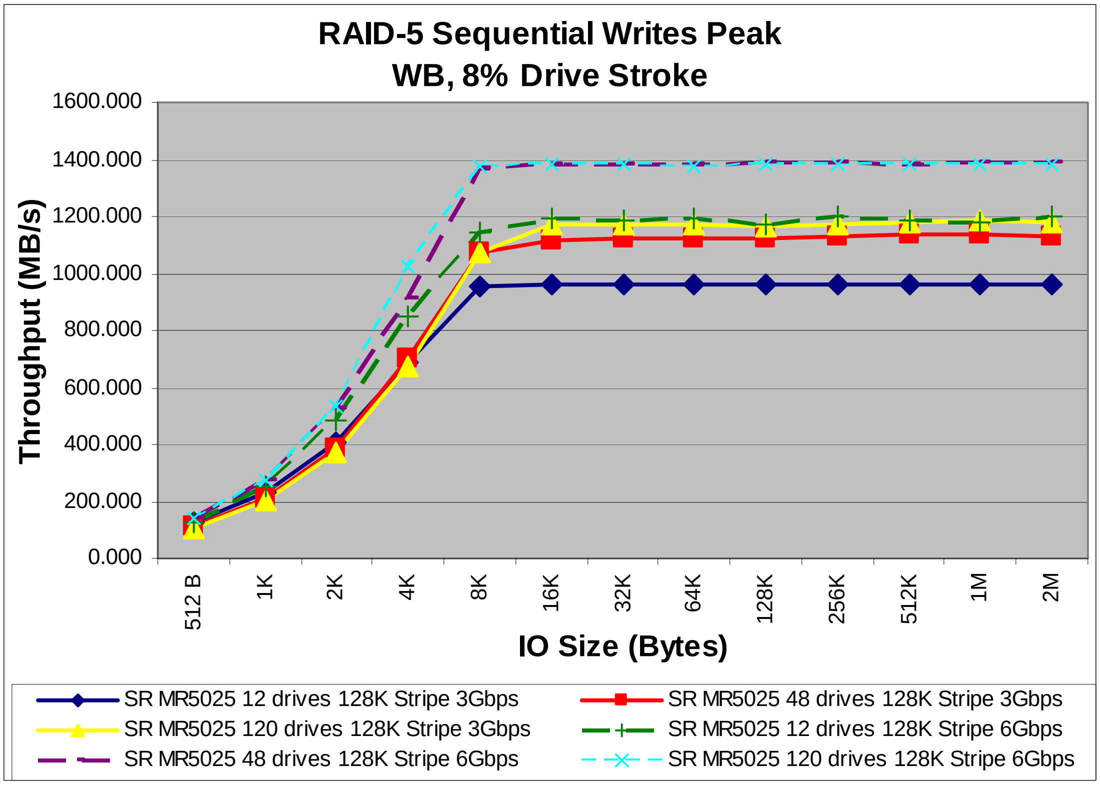


Figure 5. ServeRAID M5025 RAID-5 Streaming Writes Transfer Rate 3Gbps vs. 6Gbps SAS HDDs

Figure 5 illustrates that the ServeRAID M5025 with 6Gbps HDDs peak RAID-5 streaming write performance is up to 25% higher than that of the ServeRAID M5025 with 3Gbps HDDs.

Also note that with the ServeRAID M5025, peak RAID-5 streaming write performance cannot be obtained with twelve 6Gbps SAS HDDs. Due to the throughput limitations of the 6Gbps SAS drives, more than 12 drives are required.

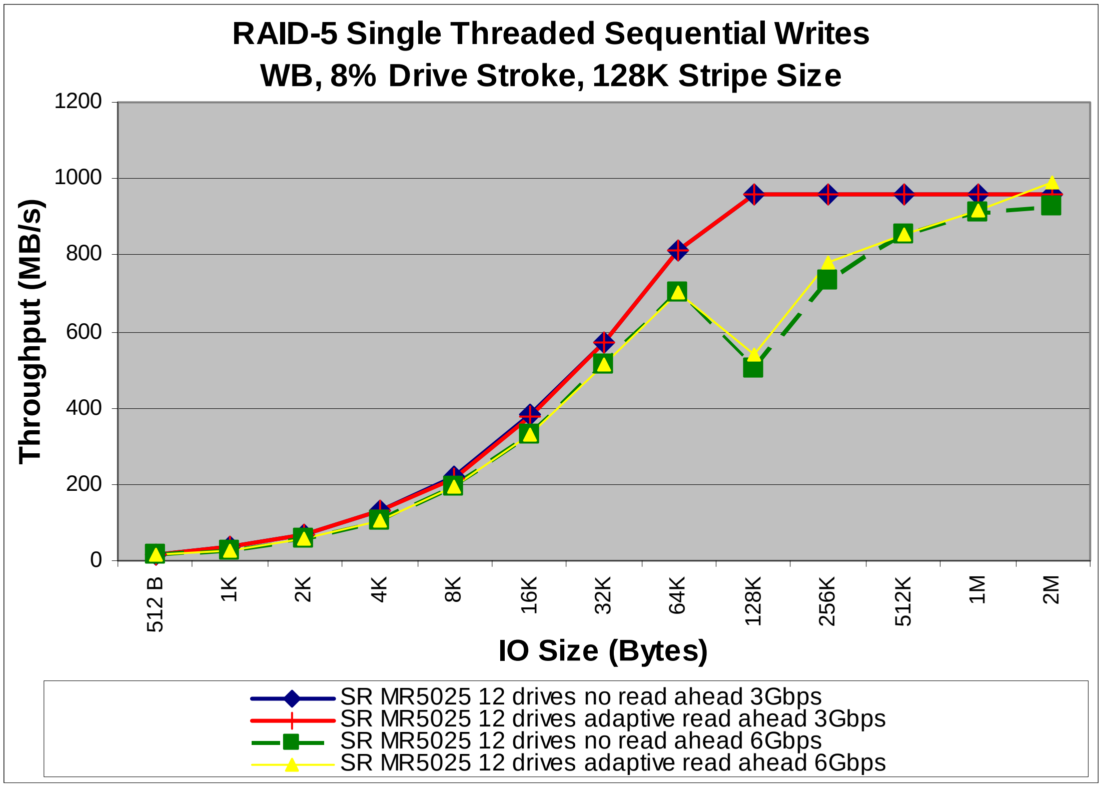


Figure 6. ServeRAID M5025 RAID-5 Single-Threaded Sequential Write Transfer Rate 3Gbps vs. 6Gbps SAS HDDs

Figure 6 illustrates that in single-threaded write workloads, the ServeRAID M5025 with 6Gbps HDDs outperforms the ServeRAID M5025 with 3Gbps HDDs only slightly. This comparison is of particular interest to those who use file copy benchmarks to evaluate performance.

Also note that enabling adaptive read-ahead on the ServeRAID M5025 has no effect on single-threaded write performance with either 3Gbps or 6Gbps SAS HDDs.

RAID-5 Windows Random Reads and Random Writes Workload Results

Table 2 contains the results for the Random Reads 4K, 8K, and 16K workloads, and the results for the Random Writes 4K, 8K, and 16K workloads. These random read and write workloads are used to track product performance and serve as a comparison between similar products. In addition, these workloads are typical of some production environments.

Workload		Random Reads 4K		Random Reads 8K		Random Reads 16K		Random Write 4K		Random Write 8K		Random Write 16K	
RAID-5		IOps	MBps	IOps	MBps	IOps	MBps	IOps	MBps	IOps	MBps	IOps	MBps
12 HDDs	15ms Average Response Time	5278	20.6	5189	40.5	5000	78.1	1731	6.8	1791	14.0	1730	27.0
	Peak	5500	21.5	5409	42.3	5213	81.4	1807	7.1	1801	14.1	1732	27.1
24 HDDs	15ms Average Response Time	10460	40.9	10292	80.4	9972	155.8	3409	13.3	3257	25.4	3152	49.2
	Peak	11133	43.5	10676	83.4	10282	160.7	3440	13.4	3318	25.9	3281	51.3
48 HDDs	15ms Average Response Time	19997	78.1	19640	153.4	18592	290.5	4069	15.9	3999	31.2	3779	59.0
	Peak	21025	82.1	20600	160.9	20355	318.0	4812	18.8	4487	35.1	4410	68.9
120 HDDs	15ms Average Response Time	51354	200.6	50145	391.8	48520	758.1	9941	38.8	9636	75.3	9384	146.6
	Peak	52613	205.5	51625	403.3	49914	779.9	9951	38.9	9654	75.4	9394	146.8

Table 2. Results for Random Reads and Random Writes 4K, 8K and 16K Workloads

Table 2 illustrates that the performance in random read and random write workloads continues to increase as drives are added to the ServeRAID M5025. The ServeRAID M5025 supports up to 216 drives, so for random read and random write workloads, drives can continue to be added until the performance is limited by the controller itself and not by the number of drives attached to the controller.

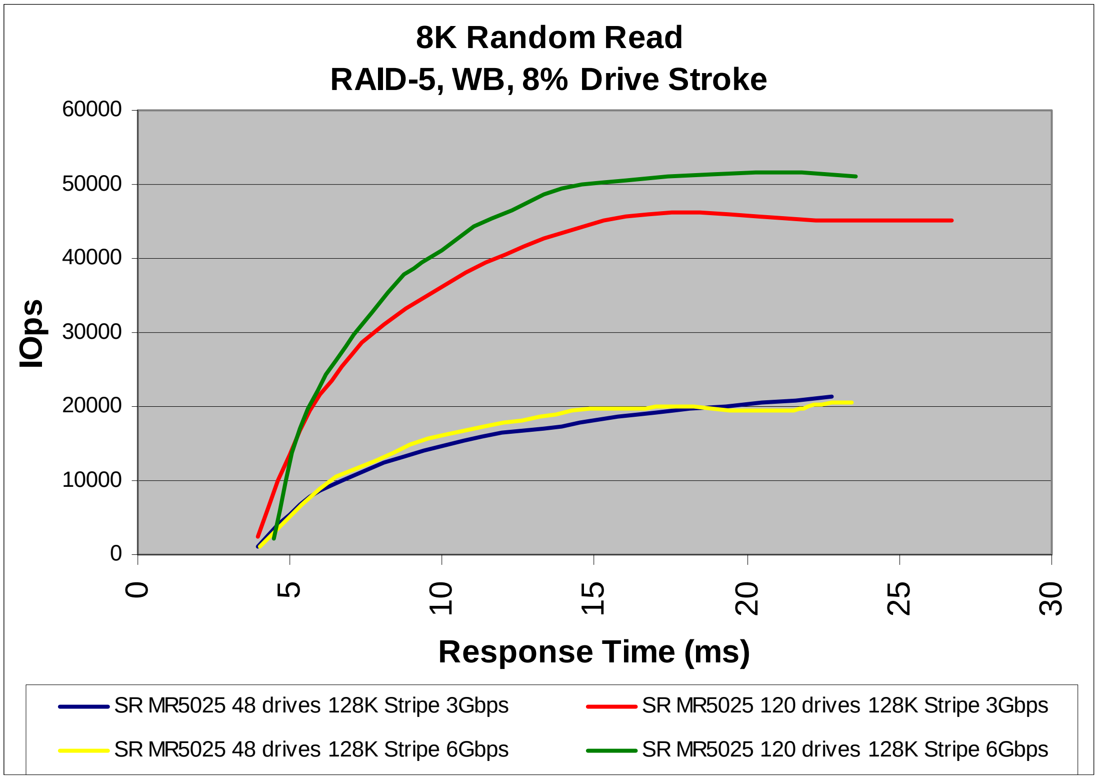


Figure 7. ServeRAID M5025 RAID-5 Random Read 8K IOps 3Gbps vs. 6Gbps SAS HDDs

Figure 7 illustrates that in 48-drive configurations, the ServeRAID M5025 with 6Gbps HDDs has only a minuscule performance advantage over the configuration using 3Gbps HDDs in the 8K random read workload.

However, the ServeRAID M5025 using 6Gbps HDDs has a larger performance advantage over the 3Gbps HDD configuration in random read workloads when more than 48 drives are used. In the 120-drive configuration, the peak random read performance using the 6Gbps SAS HDDs is 12% higher than when using the 3Gbps SAS HDDs.

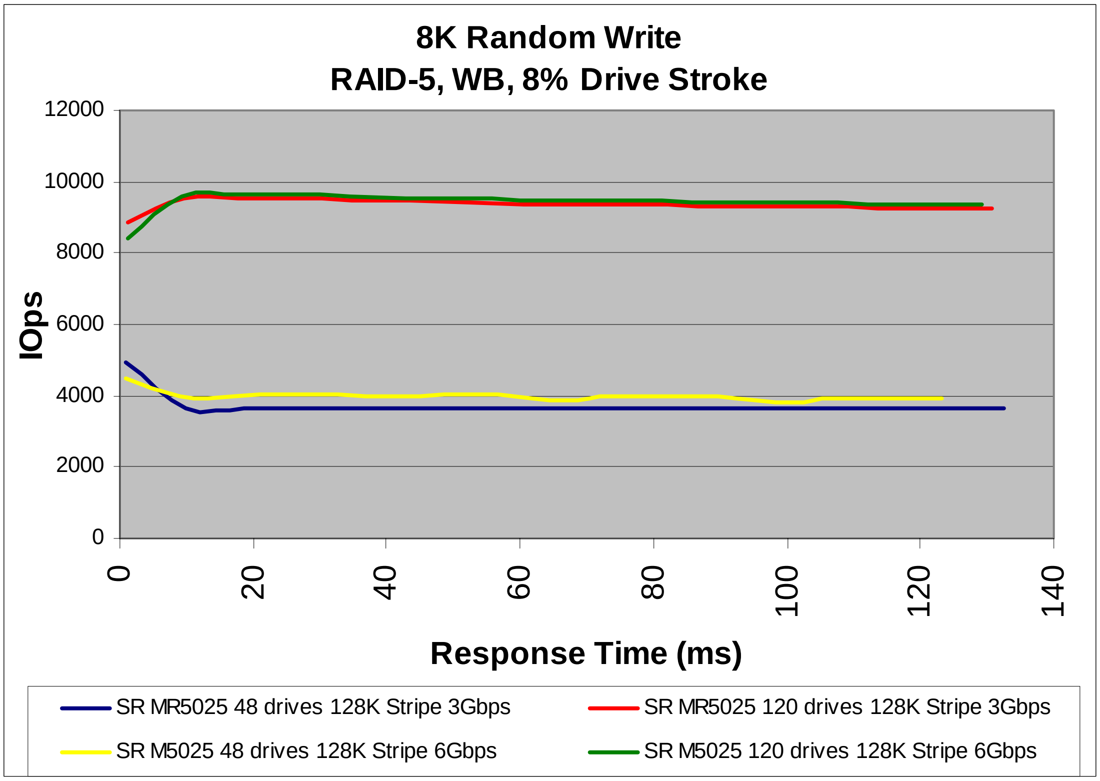


Figure 8. ServeRAID M5025 RAID-5 Random Write 8K IOps 3Gbps vs. 6Gbps SAS HDDs

Figure 8 illustrates that the ServeRAID M5025 with 6Gbps HDDs does not have a large performance advantage over the 3Gbps configuration in random write workloads regardless of the number of drives used. In 48-drive configurations, the peak random write performance when using 6Gbps SAS HDDs is only 5% higher than when using 3Gbps SAS HDDs. In 120-drive configurations, the peak random write performance difference is negligible.

Linux Measurement Results and Analysis

The performance information contained in this section was derived under specific operating and environmental conditions. The results obtained in your operating environments may vary significantly.

The measurement results in this section represent the maximum sustainable performance for a configuration with either twelve, twenty-four, forty-eight or one-hundred and twenty 15K rpm SAS hard disk drives (HDDs) at both an average and peak response times of approximately 15 milliseconds (ms). Peak performance typically refers to a measurement result with the highest number of IOps or MBps regardless of the average response time associated with that result. However, since most server applications will not wait forever for disk I/O to complete, the 15 ms threshold is a reasonable amount of time for an application to wait for completion of disk I/O before overall performance begins to decrease. For this reason, the measured performance at a 15 ms average response time should be considered as a more accurate representation of real-world performance.

RAID-5 Linux OLTP Workload Results

The SUSE Linux Enterprise Server 10 SP3 64-bit operating system was used for all ServeRAID M5025 Linux measurements. The data drives were raw, containing no file system.

Table 3 contains RAID-5 Linux measurement results for the OLTP workload for various transfer request sizes. For all configurations, the drives were configured in arrays of 12. In all configurations, only 8% of the total capacity of the drives was used. The workload was simultaneously applied to all arrays. This is true for all of the measurements unless otherwise noted. The write-back cache was enabled for all measurements. I/O policy was set to direct. A 128K stripe size was used for all measurements.

Workload		**OLTP 4K**		**OLTP 8K**		**OLTP 16K**		**OLTP 32K**		**OLTP 64K**	
RAID-5		**IOps**	**MBps**	**IOps**	**MBps**	**IOps**	**MBps**	**IOps**	**MBps**	**IOps**	**MBps**
12 HDDs	15ms Average Response Time	2819	10.9	2504	19.3	2383	36.7	2073	63.8	1651	101.7
	Peak	3369	12.7	3211	24.1	2912	43.9	2621	78.4	2110	125.8
24 HDDs	15ms Average Response Time	4994	19.4	4744	36.8	4273	66.3	3799	117.8	3106	192.7
	Peak	6423	24.7	6198	47.5	5476	84.2	4968	151.7	3949	241.3
48 HDDs	15ms Average Response Time	10387	40.4	9762	76.0	8758	136.3	7626	237.4	5968	371.6
	Peak	11685	45.3	10577	81.9	10157	157.7	8515	264.3	7059	437.1
120 HDDs	15ms Average Response Time	18158	70.8	17140	133.7	16626	259.4	14828	462.7	12624	787.8
	Peak	18868	73.6	18270	142.6	17069	266.4	15149	472.9	12747	795.6

Table 3. ServeRAID M5025 RAID-5 Linux OLTP IOps

Table 3 illustrates that in Linux, performance in OLTP workloads will continue to increase as drives are added to the ServeRAID M5025, just as it did in Windows.

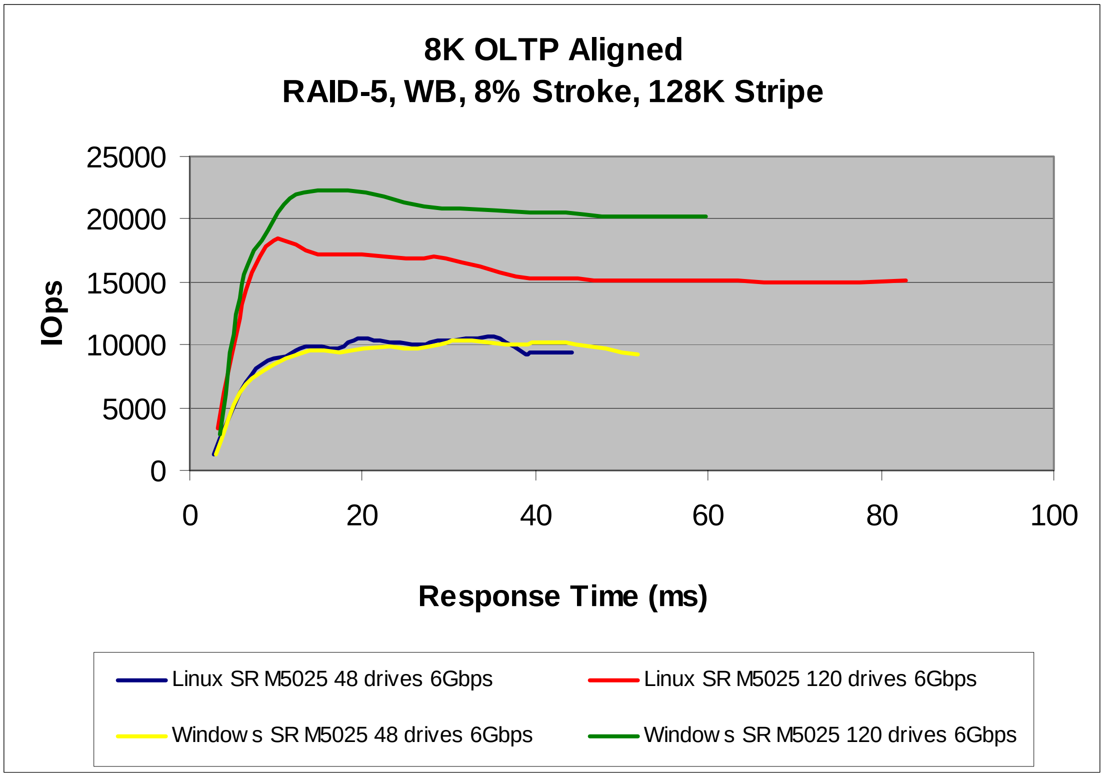


Figure 9. ServeRAID M5025 RAID-5 Linux OLTP 8K IOps vs. Windows OLTP 8K IOps

Figure 9 compares Linux fio OLTP results to Windows Iometer OLTP results as I/O workload increases. When comparing the Linux OLTP results to the Windows OLTP results, the Windows results are slightly higher than the Linux results in the 48-drive configuration. In the 120-drive configuration, the Windows results are 22% higher than the Linux results.

RAID-5 Linux Sequential Reads and Sequential Writes fio Workload Results

Figures 10 and 11 illustrate the RAID-5 Linux sequential performance of the ServeRAID M5025 as the transfer request size is increased. The performance for each transfer size is plotted on the chart without regard to response time.

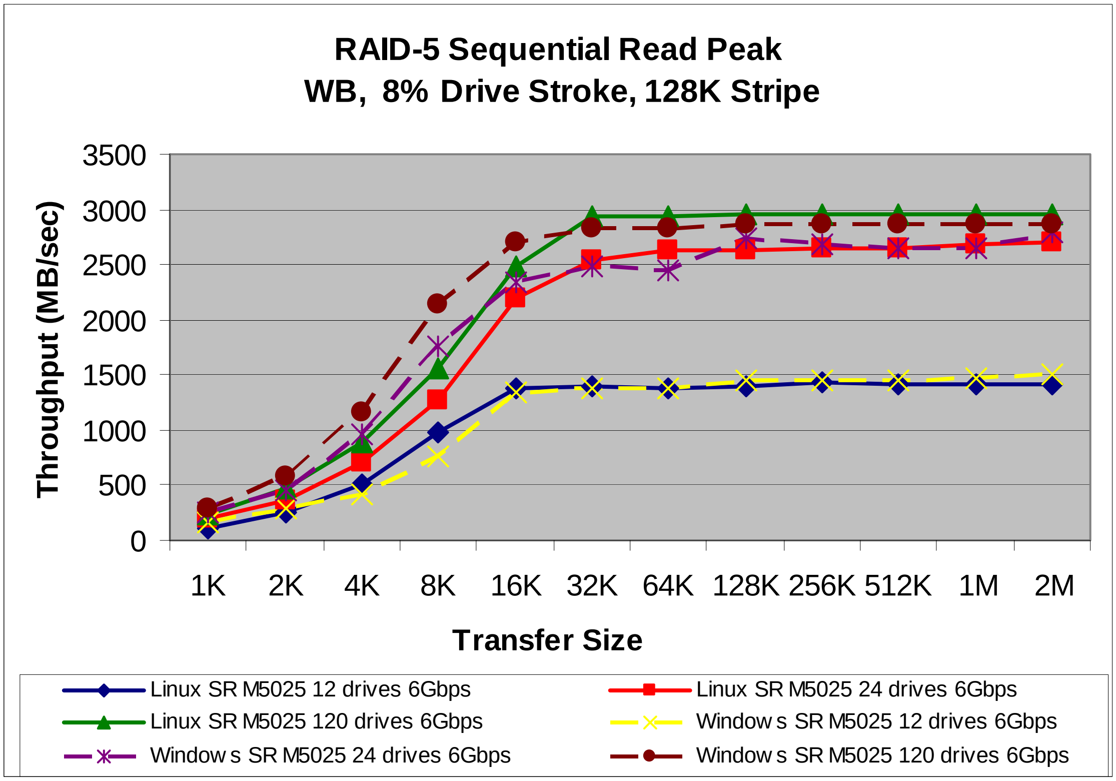


Figure 10. ServeRAID M5025 RAID-5 Linux fio vs. Windows Iometer Sequential Reads Transfer Rate

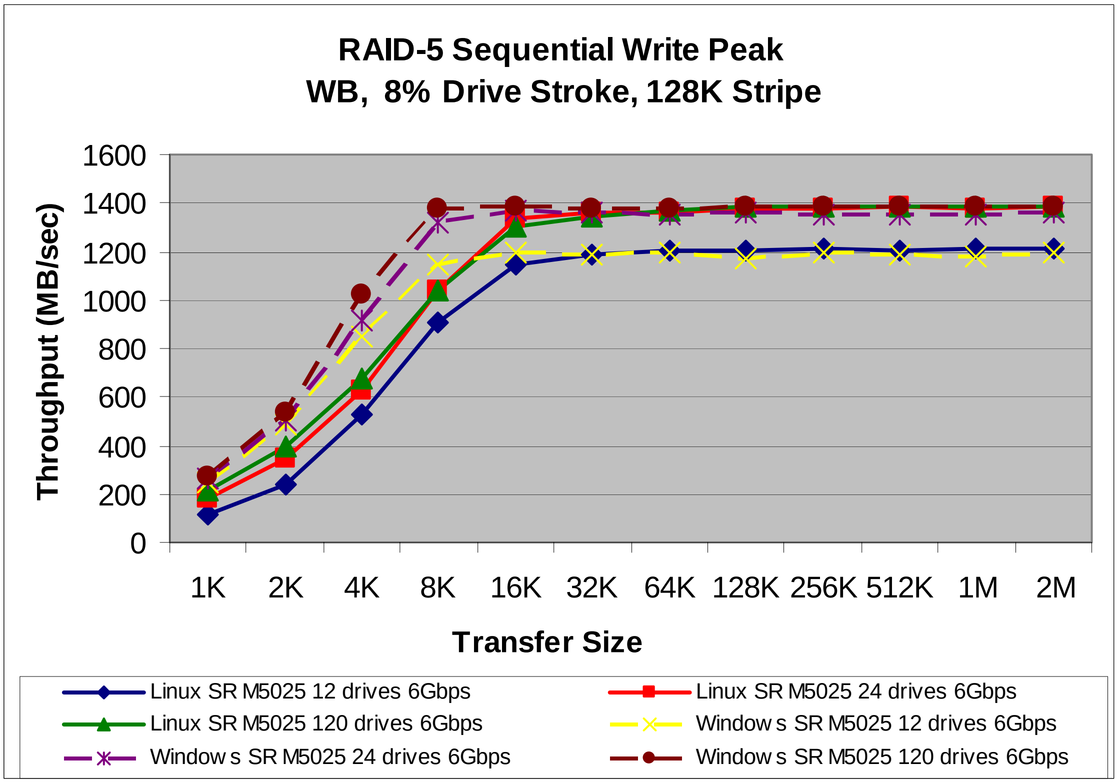


Figure 11. ServeRAID M5025 RAID-5 Linux fio vs. Windows Iometer Sequential Write Transfer Rate

Figures 10 and 11 illustrate that the Linux fio peak sequential read and peak sequential write performance results were almost identical to the Windows Iometer results. The Linux fio results were generally less than 5% higher than the Windows Iometer results in all cases.

RAID-5 Linux Random Reads and Random Writes Workload Results

Table 4 contains the results for the Random Reads 4K, 8K and 16K workloads, and the results for the Random Writes 4K, 8K and 16K workloads. These random read and write workloads are used to track product performance and serve as a comparison between similar products. In addition, these workloads are typical of some production environments.

Workload		Random Reads 4K		Random Reads 8K		Random Reads 16K		Random Write 4K		Random Write 8K		Random Write 16K	
RAID-5		IOps	MBps	IOps	MBps	IOps	MBps	IOps	MBps	IOps	MBps	IOps	MBps
12 HDDs	15ms Average Response Time	4647	17.9	4439	34.2	3936	60.6	1674	6.4	1615	12.4	1521	23.4
	Peak	5172	19.7	4738	36.3	4531	68.9	1869	7.1	1745	13.0	1604	24.4
24 HDDs	15ms Average Response Time	9868	38.3	9480	73.5	8427	130.7	3195	12.4	3102	24.1	2879	44.7
	Peak	10273	39.7	10346	79.9	9216	142.5	3406	12.9	3281	24.9	2950	44.7
48 HDDs	15ms Average Response Time	20463	79.6	19479	151.6	17939	279.3	3910	15.2	3816	29.6	3531	54.9
	Peak	21652	84.1	20111	156.4	19130	297.0	5273	20.6	4994	39.0	4450	69.5
120 HDDs	15ms Average Response Time	46218	180.3	44277	345.4	40196	627.2	7822	30.5	7555	58.9	7136	111.3
	Peak	46659	182.0	45453	354.5	43145	672.9	8209	32.1	7873	61.5	7357	114.9

Table 4. ServeRAID M5025 RAID-5 Linux Random Reads IOps

Table 4 illustrates that the Linux performance in random read and random write workloads continues to increase as drives are added to the ServeRAID M5025 just as it did in Windows.

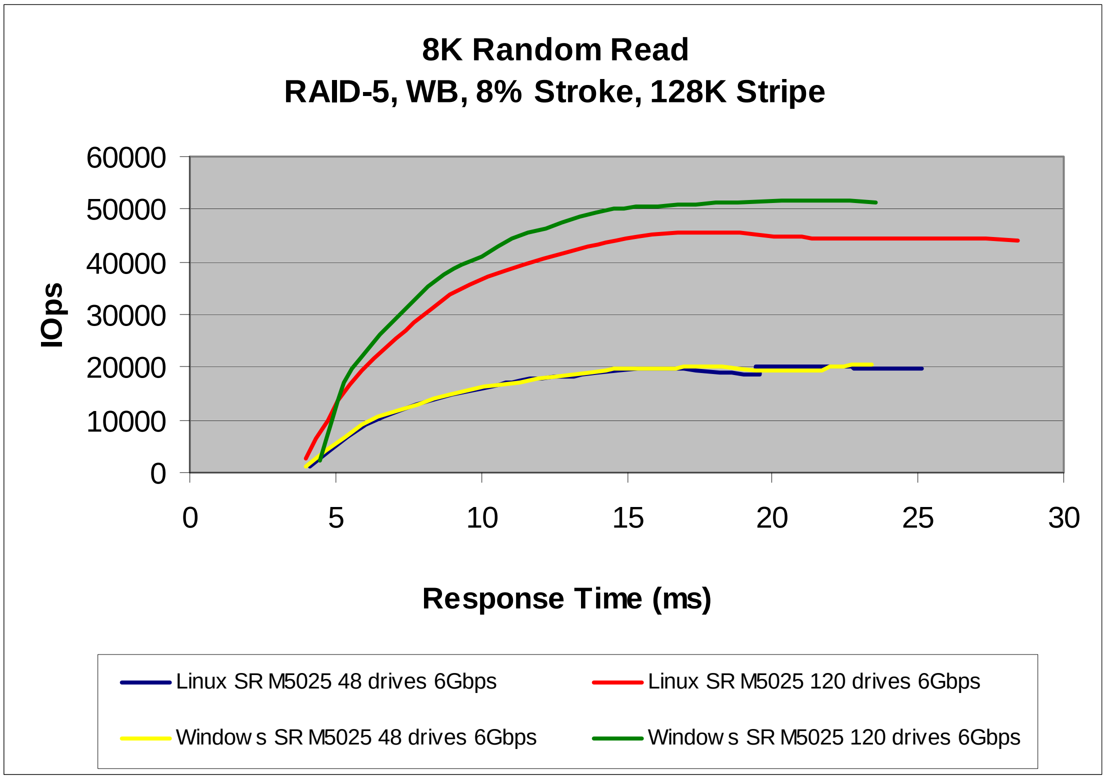


Figure 12. ServeRAID M5025 RAID-5 Linux fio vs. Windows Iometer 8K Random Read Transfer Rate

Figure 12 illustrates that with a 48-drive configuration, the Linux fio random read performance results were virtually identical to the Windows Iometer results. However, with a 120-drive configuration, the Windows Iometer results were up to 14% higher than the Linux fio results.

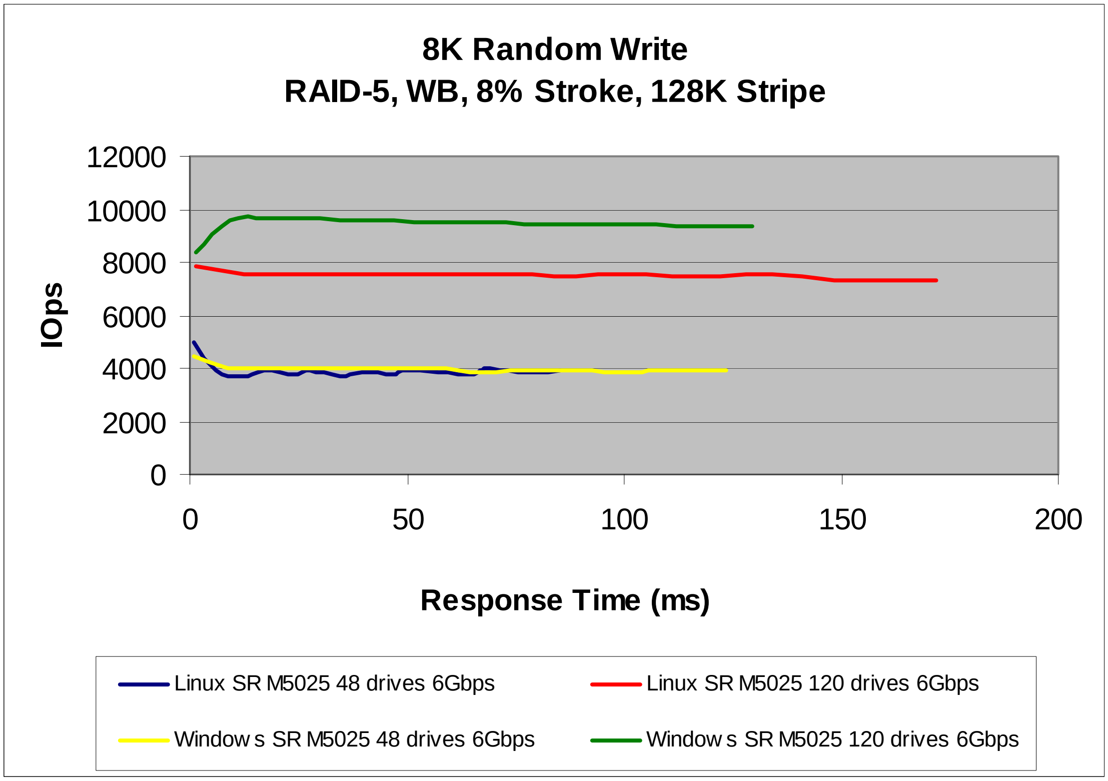


Figure 13. ServeRAID-M5025 RAID-5 Linux fio vs. Windows Iometer 8K Random Write Transfer Rate

Figure 13 illustrates that with a 48-drive configuration, the Linux fio random write performance results were virtually identical to the Windows Iometer results. However, with a 120-drive configuration, the Windows Iometer results were up to 23% higher than the Linux fio results.

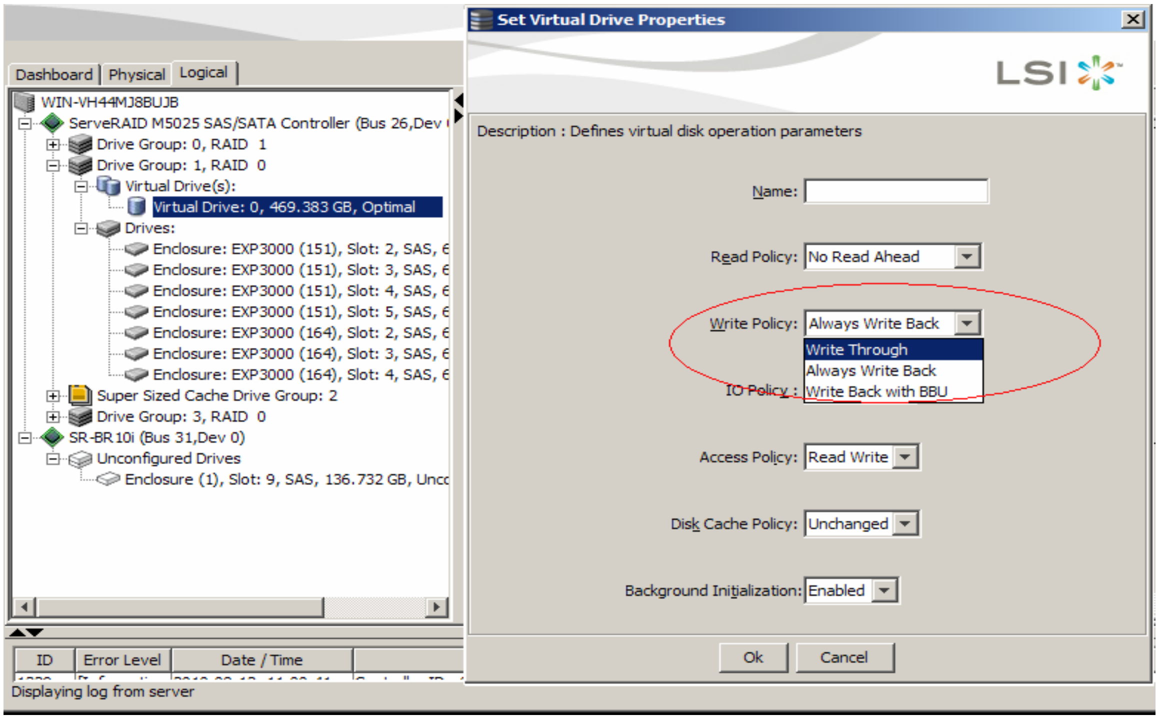
Getting the Best Out-of-the-Box Write Performance

Like any other electronic device that uses battery cells, the ServeRAID M5025's battery must be trained and charged before the battery is fully operational. The initial out-of–the-box training and charging session includes a full charge of the battery, a full discharge of the battery and an additional charge of the battery. As one would expect, this process will take several hours to complete.

The performance impact of the initial battery training and charging session is that the measured write performance can be lower during this period because by default, the write-back cache is disabled until the battery is fully operational. To measure the highest write performance during the initial evaluation period, the user has one of two choices:

- Configure the ServeRAID M5025 in the system; let the battery fully charge overnight and collect performance measurements the next morning. The write-back cache will automatically be enabled once that battery is fully charged, so the write performance results will be back to acceptable levels.

- Use the "Always Write Back" parameter in the MegaRAID Storage Manager (see Figure 14). Using this method will allow the use of the write-back cache while the battery is being trained and charged. This may be acceptable for performance benchmarking purposes, but the setting should be re-enabled before the adapter is used in production to ensure that the write cache is only enabled when the battery has enough charge to protect the data in the cache.



WIN-VH44MJ8BUJB
ServeRAID M5025 SAS/SATA Controller (Bus 26,Dev
Drive Group: 0, RAID 1
Drive Group: 1, RAID 0
Virtual Drive(s):
Virtual Drive: 0, 469.383 GB, Optimal
Drives:
Enclosure: EXP3000 (151), Slot: 2, SAS, 6
Enclosure: EXP3000 (151), Slot: 3, SAS, 6
Enclosure: EXP3000 (151), Slot: 4, SAS, 6
Enclosure: EXP3000 (151), Slot: 5, SAS, 6
Enclosure: EXP3000 (164), Slot: 2, SAS, 6
Enclosure: EXP3000 (164), Slot: 3, SAS, 6
Enclosure: EXP3000 (164), Slot: 4, SAS, 6
Super Sized Cache Drive Group: 2
Drive Group: 3, RAID 0
SR-BR10i (Bus 31,Dev 0)
Unconfigured Drives
Enclosure (1), Slot: 9, SAS, 136.732 GB, Unco

Displaying log from server

Description : Defines virtual disk operation parameters

Name:

Read Policy: No Read Ahead

Write Policy: Always Write Back

IO Policy :

Access Policy: Read Write

Disk Cache Policy: Unchanged

Background Initialization: Enabled

Figure 14. MegaRAID Storage Manager Write Cache Settings

Conclusions

First, the ServeRAID M5025 is designed to support a variety of business applications, including databases, e-mail, file serving and Web serving. The ServeRAID M5025 also performs well in streaming media applications.

Second, the ServeRAID M5025 with 6Gbps SAS HDDs offers a moderate to significant performance improvement over the same adapter used with 3Gbps SAS HDDs in certain workloads. Some examples of the approximate gains include:

- Up to 7.5% for 8K OLTP workload
- Up to 71% for Streaming Reads workload
- Up to 30% for Single-Threaded Sequential Read workload
- Up to 25% for Streaming Writes workload
- Up to 12% for Random Reads workload
- Up to 5% for Random Writes workload

Last, with respect to Windows versus Linux performance, the ServeRAID M5025 with 6Gbps SAS HDDs performs similarly in both OS environments; however, the Windows results were somewhat higher than the Linux results in most cases.

IBM Systems and Technology Group
Department MX5A
Research Triangle Park NC 27709

Produced in the USA

July 2011

IBM reserves the right to change specifications or other product information without notice. References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates. I BM PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions; therefore, this statement may not apply to you.

IBM System x servers are assembled in the U.S., Great Britain, Japan, Australia and Brazil and are composed of U.S. and non-U.S. parts.

This publication may contain links to third party sites that are not under the control of or maintained by IBM. Access to any such third party site is at the user's own risk and IBM is not responsible for the accuracy or reliability of any information, data, opinions, advice or statements made on these sites. IBM provides these links merely as a convenience and the inclusion of such links does not imply an endorsement.