

LSI MegaRAID Controller Benchmark Tips

November 6, 2012 v9.0 Preliminary



Benchmark Review Requirements

- Providing the following information will improve our ability to support product evaluations
 - Summary of test cases you will be performing
 - System details and benchmark parameters listed on slide 3
- Questions regarding configuration and benchmark results should be addressed to:
 - LSI Technical Support
 - <u>support@lsi.com</u>
 - (For Product Evaluations)
 - Craig McCombs
 - craig.mccombs@lsi.com
 - James Evans
 - james.evans@lsi.com

Elements that Affect Performance

- System
 - Motherboard, Chip Set, BIOS, Processor, Memory
 - System chip set and memory speed can impact benchmark performance
 - Recommend 8-wide (x8) PCIe Generation-2 slot for all 6 Gb/s SAS benchmarks
 - Operating System with latest Service Pack and Updates
- RAID controller
 - Firmware, BIOS, driver version
 - Disk write cache policy setting
 - RAID level
 - Stripe size
 - Read and write caching policies
- Drives
 - Total number of drives/Drives per channel
 - Enclosure model + firmware
 - Drive interface (SATA, SAS) and Speed (i.e. 1.5Gb/s, 3Gb/s, 6Gb/s)
 - Make, model & firmware of drive
 - Class of the drive (e.g., HDD disk drive, SSD solid state, etc.)
- Benchmark Tool
 - Test profile (Request Size, Sequential Read or Write, Random Read or Write)



Tips

- Use the latest firmware & driver
 - Latest software can be downloaded from <u>MegaRAID Downloads</u>
- To configure the RAID Adapter and create logical arrays use either
 - CTRL-H utility during BIOS POST
 - Use MegaRAID Storage Manager[™] (MSM) running from OS
- To achieve maximum sustained RAID controller throughput
 - For maximum bandwidth (MB/s) use 64KB or larger sequential requests
 - For maximum operations (IO/s) use 0.5KB to 8KB random requests
 - Connect enough drives to saturate the RAID controller to achieve maximum performance.
 - For example use 20 or more 15K SAS drives to achieve maximum bandwidth performance
 - Use all SAS channels connected to drives or drive enclosures
 - With multiple drive enclosures make sure drives evenly distributed across enclosures
 - Drive technology and cache size can significantly impact performance
 - Benchmark queue depth will impact performance. Recommend queue depth of 16 or greater

Standard Benchmark Tool

- IOmeter is a I/O performance analysis tool for servers and workstations
- Obtain latest IOmeter from Source Forge
 - IOmeter User's Guide IOmeter.org
 - Windows 32-bit and 64-bit versions available
 - Recommend using 2006 or later version
 - Different IOmeter versions use different data patterns and compressibility
- Throughput is measured in IO/sec and MB/sec
 - I/Os per second for random/transactional workloads
 - Megabytes per second for sequential/streaming workloads
- IOmeter performance testing article and test script examples
 - LSI Knowledge Base Article on IOMeter scripts
 - Link to knowledge base article http://mycusthelp.info/LSI/_cs/AnswerDetail.aspx?inc=8274
 - Article Includes links to IOmeter script examples
 - See also Appendix for basic IOmeter workload configuration examples



Recommended Settings For HDD Performance Testing

HDD DISK ENVIRONMENTS		RECOMMENDED SETTINGS			
RAID Type	I/O Benchmarking	RAID Write Cache	RAID Read Cache	Stripe Size	
0	Transactional	Enabled	No Read Ahead	64KB - 256KB	
1/10	Transactional	Enabled	No Read Ahead	64KB - 256KB	
5/50	Transactional	Enabled	No Read Ahead	64KB - 256KB	
6/60	Transactional	Enabled	No Read Ahead	64KB - 256KB	
0	Streaming	Disabled	Always Read Ahead	256KB or higher	
1/10	Streaming	Disabled	Always Read Ahead	256KB or higher	
5/50	Streaming	Enabled	Always Read Ahead	256KB or higher	
6/60	Streaming	Enabled	Always Read Ahead	256KB or higher	

Recommended Settings For SSD Performance Testing

SSD DISK ENVIRONMENTS		RE	ECOMMENDED SETTINGS	
RAID Type	I/O Benchmarking	RAID Write Cache	RAID Read Cache	Stripe Size
0	Transactional	Disabled	No Read Ahead	64KB
1/10	Transactional	Disabled	No Read Ahead	64KB
5/50	Transactional	Disabled	No Read Ahead	64KB
6/60	Transactional	Disabled	No Read Ahead	64KB
0	Streaming	Enabled	Always Read Ahead	64KB
1/10	Streaming	Enabled	Always Read Ahead	64KB
5/50	Streaming	Enabled	Always Read Ahead	64KB
6/60	Streaming	Enabled	Always Read Ahead	64KB

Read Policies

Always Read Ahead

 This specifies that the controller uses read-ahead if the two most recent disk accesses occurred in sequential sectors. If all read requests are random, the algorithm does not read ahead, however all requests are continually evaluated for possible sequential operation.

No Read Ahead

 Only the requested data is read and the controller does not read ahead any data

Write Cache Policies

• Write-Through

- Caching strategy where data is committed to disk before a completion status is returned to the host operating system
- Considered more secure, since a power failure will be less likely to cause undetected drive write data loss with no battery-backed cache present
- Data is moved directly from the host to the disks, avoiding copying the data intermediary into cache which can improve overall performance for streaming workloads if Direct IO mode is set.

• Write-Back

- A caching strategy where write operations result in a completion status being sent to the host operating system as soon as data is in written to the RAID cache. Data is written to the disk when it is forced out of controller cache memory.
- Write-Back is more efficient if the temporal and/or spatial locality of the requests is smaller than the controller cache size.
- Write-Back is more efficient in environments with "bursty" write activity.
- Battery backed cache can be used to protect against data loss as a result of a power failure or system crash.



Data Placement Policies

• Direct IO

- All read data is transferred directly to host memory bypassing RAID controller cache. Any Read Ahead data is cached.
- All write data is transferred directly from host memory bypassing RAID controller cache if Write-Through cache mode is set
- Recommended for all configurations

Cached IO

- All read and write data passes through controller cache memory on its way to or from the host (including write data in write-through mode.)
- Required ONLY for CacheCade v1.1 read-only caching, not recommended for CacheCade v2.x and higher or any other configurations



Advanced Software Testing

MegaRAID Fast Path Software

- In order to use Fast Path software, the following requirements must be met:
 - The array must not have any cache enabled
 - No Read Ahead
 - Write Through
 - Direct I/O
 - The array must be in good health
 - Cannot be degraded or currently under rebuild
 - The Fast Path software key must be enabled

MegaRAID CacheCade[™] Software

- In order to use CacheCade software, the following requirements must be met:
 - CacheCade software key must be enabled
 - Use recommended VD settings for drive type and IO workload profile
 - Only exception is use CIO (cached I/O mode) to enabled CacheCade v1.1 readonly caching



Performance Limits & Bottlenecks

		INTERCONNECT	
PCI Express	Lanes	Theoretical Bandwidth (Uni-Directional)	Practical Bandwidth (Uni-Directional)
Gen1	x4	1000	880
Gen1	x8	2000	1760
Gen 2	x4	2000	1600
Gen 2	x8	4000	3200
SAS	PHYs	Theoretical Bandwidth (Uni-Directional)	Practical Bandwidth (Uni-Directional) SAS/SATA
Gen 1	x1	150	105/130
Gen 1	x4	600	420/520
Gen 1	x8	1200	840/1040
Gen 1	x1	300	240/285
Gen 1	x4	1200	990/1140
Gen 1	x8	2400	1975/2280
Gen 2	x1	600	520/570
Gen 2	x4	2400	1975/2280
Gen 2	x8	4800	3950/4560
Disk Interface	Rotational Rate	Random IOPs	Media Rate (Sustained Sequential) MB/s
SATA	7200 RPM	70 - 175	60 - 150
SAS	10K RPM	275 - 300	100 - 175
SAS	15K RPM	350 - 450	125 - 200
SSD	N/A	15K - 180K	30 - 400

____ **~**_ 1 🔊

* Note: Disk IOPs and sustained MB/s are estimates and will change with host system, disk models, technology & drive revisions.

SAS 3Gb/s: 1078 ROC architecture



LSI Proprietary

SAS 6Gb/s: 2108 ROC Architecture (Liberator)



LSI Proprietary



SAS 6Gb/s: 2208 ROC Architecture (Thunderbolt)





Multipath

- Most of SAS drives have 2 SAS ports. Presence of more than one data path between SAS drive and RAID adapter is called multipath.
- SATA drive enclosures either alternate connection between two channels or utilize interposer devices providing multipath.
- Multipath can significantly increase performance for some applications (e.g., VTL/Backup and Media Streaming)

Performance Measurement with and without Multipath I/O (LSI 8888EM2 with 3Gb SAS)

•This chart of 3Gb/s MegaRAID controller benchmark results demonstrates dramatically improved performance in multipath environments.





Multipath Example



- Multipath provides two paths to each drive
- RAID controller utilizes both channels for transferring data
- RAID controller balances workload across both channels

17



Multipath Impact

Sequential IO

- Significant performance gain on the sequential writes
- Dual path configuration outperformed single path on big block (64K or bigger) sequential reads by 40%

• Random IO

- Dual path configuration outperforms single path one by 5% on small block writes
- No difference is seen on random reads

• Summary

 Dual path implementation increases IO performance for VTL/Backup and Media Streaming applications.

Appendix: IOmeter Workload Profile Examples



1MB Sequential Read Profile

Edit Access Specification	×	1 Use a single worker
Name Default Assignment 1MB 100% reads 0% random All Workers		
Size % Access % Read % Random Delay Burst Alignment 1MB 0KB 0B 100 100 0 1 sector	Reply Insert Before Insert Alter Delete	2. Use IOmeter Maximum Disk Size (sectors) sufficiently large enough to ensure all volume drives are accessed
Transfer Request Size Percent of Access Specification 1 0 100 Percent Megabytes Bytes Percent Random/Sequential Distribution Burstiness Transfer Delay Burst Length	Percent Read/Write Distribution	3. Use 64KB to 1MB transfer size
100% 0% Sequential Random	Megabytes Kilobytes Bytes	lometer
Reply Size No Reply Megabytes Kilobytes Bytes	OK Cancel	Image: Seconds Image
LSI Proprietary		20 LSI 🐩



1MB Sequential Write Profile

Edit Access Specification		×	1 Use a single worker
Name IMB 0% reads 0% random	Default Assignment		
Size % Access % Read % Random Delay Bu 1MB OKB OB 100 0 0 0	turst Alignment Reply 1 sector none	Insert Before Insert After Delete	2. Use IOmeter Maximum Disk Size (sectors) sufficiently large enough to ensure all volume drives are accessed
Transfer Request Size Percent of Access Sp 1 0 • Megabytes Kilobytes Bytes Percent Random/Sequential Distribution Burstiness Transfer Delay Transfer Delay	Percent Read/Write	Distribution 0% Read	3. Use 64KB to 1MB transfer size
100% 0% Sequential Random	1 I/Os Megabytes Kil	bytes Bytes	l Indeter
Reply Size			
Megabytes Kilobytes Bytes	<u>OK</u>	Cancel	Topology Disk Targets Network Targets Access Specifications Results Display Test Setup All Managers Test Description IMB Sequential Write Profile IMB Sequential Write Profile Number of Workers to Spawn Automatically Image: Network Barp Up Time Number of Workers to Spawn Automatically Image: Network Barp Up Time Number of Workers to Spawn Automatically Image: Network Barp Up Time Number of CPUs Image: Network Barp Up Time Number of CPUs Image: Network Barp Up Time Network Image: Network Barp Up Time Number of Workers to Spawn Automatically Image: Network Barp Up Time Number of Workers to Spawn Automatically Image: Network Barp Up Time Network Image: Network Barp Up Time Network Image: Network Barp Up Time Image: Start Image Image: Network Start Image:



4KB Random Read Profile

Edit Access Specification Name Default Assignment [4KB Random Reads] All Workers	<u> </u>	1. Use one or more workers
Size % Access % Read % Random Delay Burst Alignment OMB 4KB 0B 100 100 100 0 1 sector	Reply Insert Before none Insert After Delete	2. Use IOmeter Maximum Disk Size (sectors) sufficiently large enough to ensure all volume drives are accessed
Transfer Request Size Percent of Access Specification 0 ÷ 4 ÷ 0 ÷ Megabytes Kilobytes Bytes Percent Random/Sequential Distribution Burstiness Transfer Delay Burst Length	Percent Read/Write Distribution	
0% 100% Sequential Random	C 0 1 0 1 512 1 Megabytes Kilobytes Bytes	🔲 🔲 kometer
Reply Size No Reply Megabytes Kilobytes Bytes	OK Cancel	Image: Seconds Image: Seconds Image: Seconds Image: Seconds Image: Seconds Image: Seconds </td
		Start 1 End 256 Step 1 Linear Stepping V Exponential Stepping V



4KB Random Write Profile

dit Access Specification			X	1 Use one or more workers
Name 4KB Random Writes	Default Assignment			1. Use one of more workers
Size % Access % Read % Random D OMB 4KB 0B 100 0 100	relay Burst Alignment 0 1 sector	Reply none	Insert Before Insert After Delete	2. Use IOmeter Maximum Disk Size (sectors) sufficiently large enough to ensure all volume drives are accessed
Transfer Request Size 0 + 0 + Percent of Av 0 + 4 + 0 + - Megabytes Kilobytes Bytes - - - -	ccess Specification	Percent Read/Write D	istribution 0% Read	
Percent Random/Sequential Distribution Burstiness J 0% 100% Sequential Random	elay ms 1 Hurst Length 1 1/0s	Align I/Os on C Sector Boundaries C C C C C C C C C C C C C C C C C C C	512 🔆 ytes Bytes	lometer
Reply Size		Or	T Canad	Topology Disk Targets Network Targets Access Specifications Results Display Test Setup
Megabytes Nilobytes bytes				All Managers B B ICTCMCCOMB15: 4KB Random Write Profile
				Run Time Ramp Up Time 0 Hours 30 Seconds 0 Minutes 30 Seconds
				Cycling Options Cycle # Outstanding I/Os ~ run step outstanding I/Os on all disks at a time. Workers Start Start Start Step Power 4
				Linear Stepping Exponential Stepping



8KB Online Transaction Processing (OLTP)

dit Access Specification		×	1. Use one or more workers.
Size % Access % Read % Random Delay E OMB 8KB 0B 100 67 100 0	Default Assignment All Workers Burst Alignment Reply 1 sector none	Insert Before Insert After Delete	2. Use IOmeter Maximum Disk Size (sectors) sufficiently large enough to ensure all volume drives are accessed
Transfer Request Size Percent of Access S 0 • Megabytes Kilobytes Bytes 100 F	Percent Read/Write Di Percent Write Align 1/0s on © Sector Boundaries	stribution 67% Read	
0% 100% Sequential Random	1 I/Os C Megabytes Kiloby	tes Bytes	lometer
Reply Size	OK	Cancel	Image: Start Image: Start <td< td=""></td<>



Email Server Profile

Edit Access Specification		1. Use one or more workers
Name Default Assignment E-Mail Server X Access X Read X Random Delay Burst Alignment Size X Access X Read X Random Delay Burst Alignment OMB 8KB 0B 100 50 100 0 1	Reply Insert Before Insert After Delete	2. Use IOmeter Maximum Disk Size (sectors) sufficiently large enough to ensure all volume drives are accessed
Transfer Request Size Percent of Access Specification 0 100 0 100 Megabytes Bytes 100 Percent 100 Percent 0% 100% 0% 100% 0% 100% 100% 1 100% 1 100% 1	Percent Read/Write Distribution 50% 50% Write Read Align I/Ds on © Sector Boundaries © 0 0 0 512 0 Megabytes Kilobytes Bytes	
Reply Size No Reply Megabytes Kilobytes Bytes	OK Cancel	Image: Seconds Image



Workstation Profile

Edit Access Specification		1. Use one or more workers
Name Default Assignment None		
Size % Access % Read % Random Delay Burst Alignment OMB 8KB 0B 100 80 80 0 1 sector	Reply Insert Before none Insert After Delete	2. Use IOmeter Maximum Disk Size (sectors) sufficiently large enough to ensure all volume drives are accessed
Transfer Request Size Percent of Access Specification 0 8 0 Megabytes Kilobytes Bytes Percent Random/Sequential Distribution Burstiness 20% 80% 20% 80%	Percent Read/Write Distribution	
Reply Size	Megabytes Nilobytes Bytes	
• No Hepy • O	OK Cancel	Topology Disk Targets Network Targets Access Specifications Results Display Test Setup Test Description Workstation Profile Workstation Profile Number of Workers to Spawn Automatically 0 Hours 30 Seconds Vetwork If a of CPUs 0 Minutes Record Results If a of CPUs If a of CPUs 0 Minutes Record Results If a of CPUs If a of CPUs 0 Vorkers Seconds If a of CPUs If a of CPUs 0 Vorkers Seconds If a of CPUs If a of CPUs 0 Vorkers Statt If a of CPUs If a of CPUs 0 Statt Statt If and the of Outstanding I/Os If a of Outstanding I/Os Statt Statt Statt Statt Exponential Stepping Exponential Stepping
LSI Proprietary		26 LSI



Web Server Profile

Edit Access Specification		X	1 Use one or more workers
Name Web Server	Default Assignment Disk Workers		
Size % Access % Read % Read 0MB 0KB 512B 22 100 1 0MB 1KB 0B 15 100 1 0MB 1KB 0B 15 100 1 0MB 2KB 0B 8 100 1 0MB 4KB 0B 23 100 1 0MB 6KB 0B 2 100 1 0MB 16KB 0B 2 100 1 0MB 32KB 0B 6 100 1 0MB 32KB 0B 7 100 1 0MB 126KB 0B 1 100 1 0MB 126KB 0B 1 100 1	andom Delay Burst Alignment 00 0 1 sector 00 0 1 sector	Reply Insert Before none	2. Use IOmeter Maximum Disk Size (sectors) sufficiently large enough to ensure all volume drives are accessed
Transfer Request Size	ercent of Access Specification	Percent Read/Write Distribution	
Percent Random/Sequential Distribution	Transfer Delay 0 ms 1 I/Os	Align I/Os on Sector Boundaries Correct O 512	lo lometer
Reply Size			
Megabytes Kilobytes Bytes		OK Cancel	Topology Disk Targets Network Targets Access Specifications Results Display Test Setup Image: Mail Managers Internet Description Web Server Profile Web Server Profile
			Bun Time Bamp Up Time Number of Worker 0 Hours 30 Seconds 0 Minutes Record Results © # of CPUs 30 Seconds Image: Control of CPUs
			Cycling Options Cycle # Outstanding I/Os ~ run step outstanding I/Os on all disks at a time. Workers Start T Start Step T Step T Power 2
			Linear Stepping V Linear Stepping V Expone



Number of Workers to Spawn Automatically

of Outstanding I/Os Start 1 End 256

Exponential Stepping 💌

Power 2

Network

C # of CPUs • 0

•

_ 🗆 🛛

File Server Profile



1. Use one or more workers

2. Use IOmeter Maximum Disk Size (sectors) sufficiently large enough to ensure all volume drives are accessed

Number of Workers to Spawn Automatically

of Outstanding I/Os-

Exponential Stepping 💌

Start 1

Power 4

Network

. 0

End 256

C # of CPUs

-

 ?

Disk

CO

Seconds

Targets

Start 1

Step 1

-

OS Drive

Edit Access Specification		1. Use one or
Name DS Drive Size % Access % Read % Random Delay OMB 8KB 0B 100 70 100 0	Default Assignment All Workers Burst Alignment Reply 1 sector none	2. Use IOmeter Insert After Delete Delete
Transfer Request Size Percent of Access 0 8 0 • Megabytes Kilobytes Bytes 100 Percent Random/Sequential Distribution Transfer Delay 0% 100% 0% 100% 100% mm	Specification Percent Read/Write Di 30% Write Align 1/Ds on Burst Length I 1/Ds Meaabutes Kiloby	Distribution 70% Read
Reply Size No Reply Megabytes Kilobytes Bytes		Cancel

nore workers

Maximum Disk Size ently large enough to me drives are accessed

lo lometer	
Topology	Disk Targets Network Targets Access Specifications Results Display Test Setup Test Description OS Drive
	Run Time Ramp Up Time Number of Workers to Spawn Automatically 0 Hours 30 Seconds 0 Minutes Record Results © 10 30 Seconds All
	Cycling Options Cycle # Outstanding I/Os run step outstanding I/Os on all disks at a time.
	Workers Targets Start Start Step Step Linear Stepping Linear Stepping



2

OS Paging

Edit Access Specification	1. Use one or more workers
Name Default Assignment IS Pagins All Workers Size % Access % Read % Random Delay Burst Alignment Reply Insert Before OMB 64KB 0B 100 90 0 1 sector none Insert After Delete	2. Use IOmeter Maximum Disk Size (sectors) sufficiently large enough to ensure all volume drives are accessed
Transfer Request Size Percent of Access Specification Percent Read/Write Distribution 0 + 64 + 0 Megabytes Kilobytes Bytes 100 Percent	
Percent Random/Sequential Distribution 100% Sequential Burst Length Transfer Delay Burst Length D ms L/Ds Align I/Os on Sector Boundaries C Megabytes Kilobytes Bytes Setor Boundaries Megabytes Kilobytes Bytes Sector Boundaries Megabytes Kilobytes Setor Boundaries Megabytes Kilobytes Setor Boundaries Megabytes Setor Boundaries Megabytes M	lometer
Reply Size Image: No Reply Image: Discrete structure Image: Discrete structure Megabytes Kilobytes Bytes	Image: Second
	Image: CTDMCCOMBISCS OS Paging Run Time Ramp Up Time Image: Im
LSI Proprietary	30 LSI 💥



