LSI™ MegaRAID® Advanced Software Evaluation Guide V3.0

Contents:

Current sightings to be aware of with Evaluation Kits

FastPath Software

- Not identifiable from the management tools unless you use MSM v8.10.-0400 or later.
- Must use the latest OS drivers from the LSI website. Drivers embedded with operating systems are not optimized for FastPath software.

CacheCade Software

- Identified as "Super Sized Cache Disk (SSCD)" in MSM unless you are using version 6.96 or later.
- System re-boot is required after increasing the number of SSDs in a CacheCade volume.
- Must use the latest OS drivers from the LSI website. Drivers embedded with operating systems are not optimized for CacheCade software.
- Make sure FastPath software is also enabled with the current CacheCade software release in order to maximize performance. A change is being made to make this step unnecessary in a future release.

MegaRAID Controllers Supported

- 9280-4i4e
- 9260-4i
- 9260-8i
- 9280-8e (electronic fulfillment only)

Standard volume policy for CacheCade Software

Write Policy: Should be determined based on your write workload characteristics and data safety requirements. Often enabling MegaRAID write-back caching improves performance, especially when drive write caching is disabled.

IO Policy: Cached IO

Read Policy: No Read Ahead

Stripe Size: 64KB

Notes:

1. Cached IO enables CacheCade feature. Use DIO if you do not want a VD's data cached by CacheCade.
2. Parameters above provide the best performance over the broadest range of configurations and workloads.

Standard volume policy for FastPath Software

Write Policy: Write Through

IO Policy: Direct IO

Read Policy: No Read Ahead

Stripe Size: 64KB

Note:

1. Write Through and Direct IO enables FastPath feature for solid-state drives (SSDs). If either Write Back or Cached IO are used, then full FastPath performance can not be achieved.
2. Parameters above provide the best performance over the broadest range of configurations and workloads. Have found in some instances that smaller and or larger strip size can provide some additional performance benefit.

Application Qualifier for CacheCade and FastPath Software

Application Profiles That Benefit from SSDs	Recommended Advanced Software
Web Servers and other transactional, small random read intensive applications	CacheCade Software
File server (including SharePoint)	CacheCade or FastPath Software
E-mail server	CacheCade or FastPath Software
Databases	CacheCade or FastPath Software
OLTP	CacheCade or FastPath Software
E-Commerce	CacheCade or FastPath Software
Large sequential databases	FastPath Software
Streaming read applications	FastPath Software
Streaming write applications	FastPath Software

LSI Flash Technologies Solution Comparison Chart

SOLUTION	SSD Capacity Needed	Existing HDD RAID Data	Volume(s) Accelerated	Targeted Application Workload	Performance Gain	Cost of Capacity	Cost of Ownership (4)
Install 64GB of DDR3 Server Memory	None	No Change Required	All Volumes File System Cached	All workload, Especially Random	Highest	$$$$	$
Add More Short Stroked HDDs	None	Redistribute on HDDs	One HDD Volume	All Workloads	Low (2)	$ (3)	$$$$
CacheCade	Hot Data Only (1)	No Change Required	One or More HDD Volumes	Random Read Intensive	Moderate	$$	$$
FastPath	All Volume Data	Load or Migrate to SSD	One SSD Volume	All Workloads, Especially Random	High	$$$	$

1. CacheCade max. SSD capacity 512GB
2. Depends on the class and number of HDDs added and amount of short stroking
3. Depends on the amount of usable HDD storage when short stroking
4. Power, space, cooling, reliability, etc.

Application Parameters That Affect CacheCade Software Performance

Size of the working data set

The working data set is a subset of the total stored data actively utilized by an application or applications at a specific point in time. The size of working data set varies from a small portion of the total amount of data stored up to all of the stored data depending on the application and typical usage model.

Note: Frequently accessed data or hot data are other terms often used to refer to the working data set.

Size of the SSD cache

1. Performance will scale as more active read data fits in the SSD cache. Tools such as Microsoft XPERF can be used to determine the working data set size so the optimal SSD cache capacity can be determined. Alternatively, adding SSD cache capacity until maximum performance is obtained and when adding additional SSD cache capacity does not significantly improve performance.

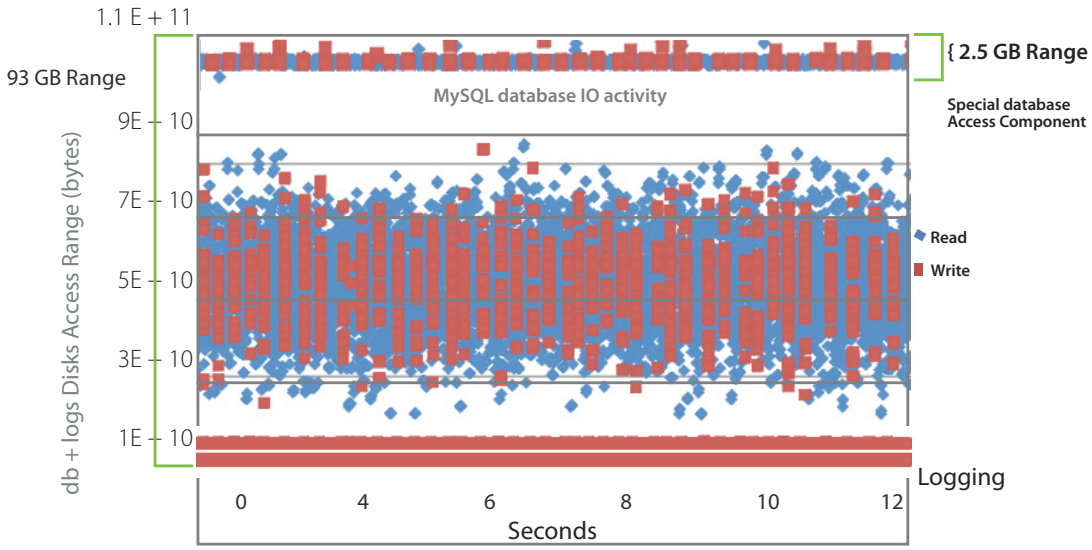
Access patterns (READs vs. WRITEs)

1. Today, CacheCade software is optimized for the READ intensive applications where the majority of working data set fits in SSD cache. Write performance is largely determined by the underlining HDD volume. Because SSD cache off loads the majority of read accesses to the working data set the read and write performance of the underlining HDD volume is also enhanced.

Percentage of READs vs. WRITEs

1. Applications with a very large random or sequential access ranges show less improvement with CacheCade software. This is because there are no frequently accessed data hot spots under these conditions. For these situations FastPath software provides the best performance enhancement alternative.
2. For large wording data sets or sequential write intensive workloads, FastPath provides the best performance enhancement alternative.

The figure below illustrates READ and WRITE access patterns of a medium sized (93 GB) transactional database, which is an ideal candidate for CacheCade software.



SSD conditioning to optimize real world performance

IOMeter pre-conditioning: Within IOMeter, perform 128KB sequential writes and then perform 4KB random writes for each SSD. Make sure I/Os are aligned on 4KB sectors. Typically each workload should be run from 30-minutes to one hour depending on the SSD capacity. You can estimate the total run time by monitoring the IOMeter data throughput rate while running 128KB sequential writes. For example take a 50GB SSD and measure 128KB sequential write throughput of 50MB/s the estimated minimum run time for each IOMeter workload is (50 GB * 1024 MB/GB)/(50 MB/s) = 1024 seconds or 17 minutes. If each SSD is not pre-conditioned in some fasion then SSD performance will initially be abnormally high and then slow decrease over time to stable pre-conditioned SSD performance specifications.

Real-world application benchmark performance stabilization: When running real-world application benchmarks, make at least two long runs, typically 1-hour each, to make sure a steady state of performance is reached. If the performance of the second run is > 10% greater than or less than first run then continue to run until performance stabilizes. Some examples of real world benchmark applications include:

- MySQL with SysBench
- Microsoft Exchange with JetStress
- Microsoft SQL with TPC-E
- Microsoft IIS web server with NeoLoad

Using IOMeter to simulate Hot Spot Data IO activity for CacheCade Software

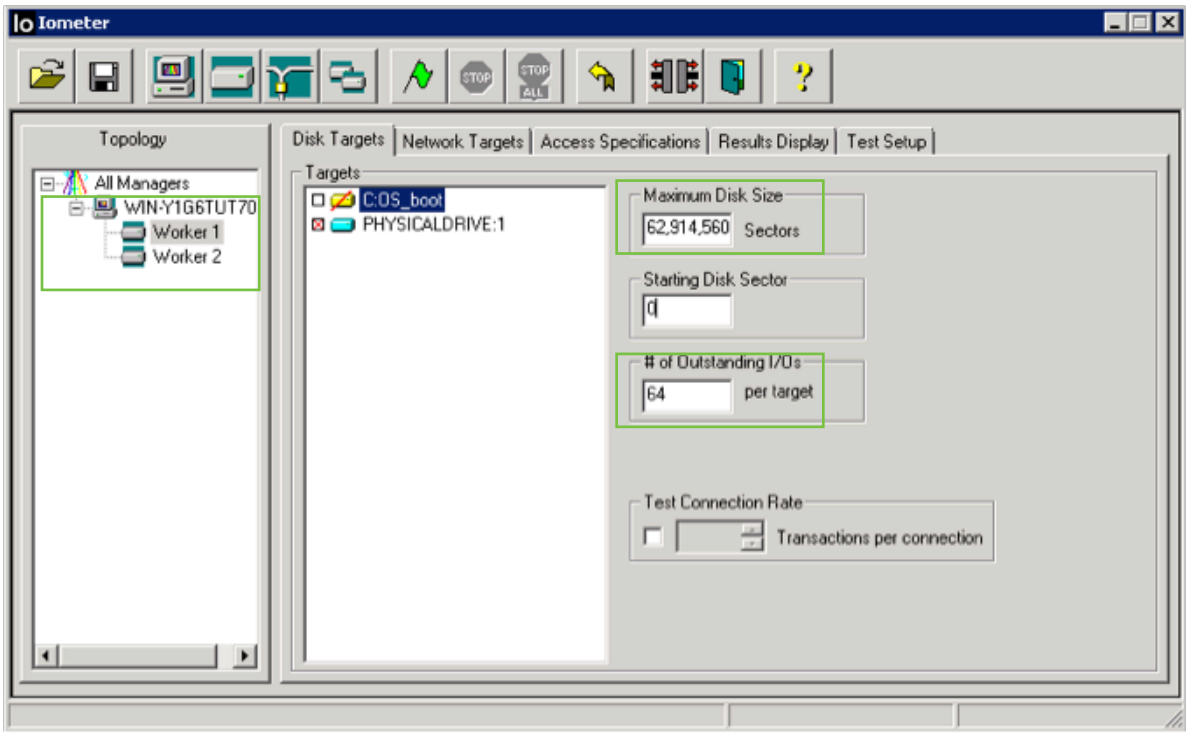
IOMeter generates purely random IOs. So if you do not create a real world workload with a working data set consistent with the available SSD cache capacity, the purely random nature of IOMeter will not be able to leverage read caching on the SSDs. The procedure below will allow you to simulate hot spot activity.

Small "Synthetic" Working Data Set IOMeter Benchmark

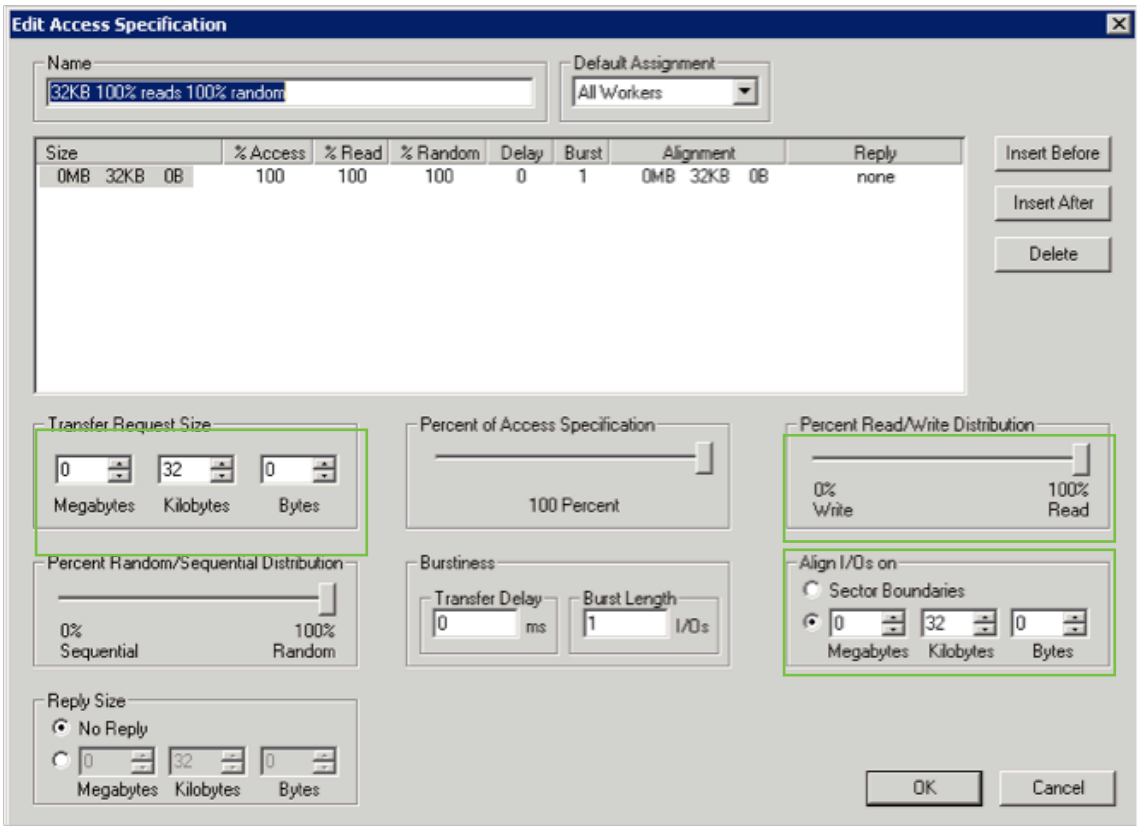
- Create a VD that uses the entire HDD array.
- Load CacheCade with the following cache priming procedure

SSD Cache Priming for Small "Synthetic" Working Data Set IOMeter Benchmark

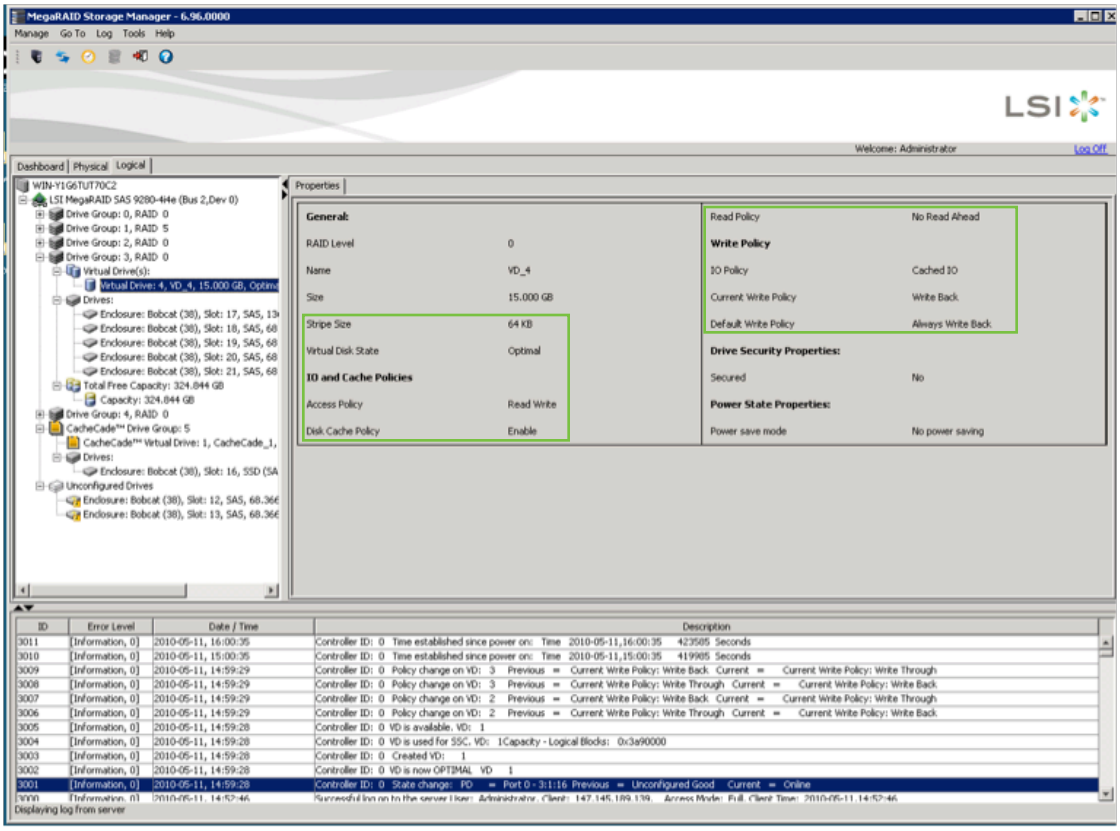
We recommend using two IOmeter workers. In this example we set the maximum disk sector size to 62,914,560, which equals a 30 GB data hot spot: (30 GB) * (1024^3 Bytes per GB)]/(512 bytes per sector). Set up both workers the same for a total command queue depth of 128.



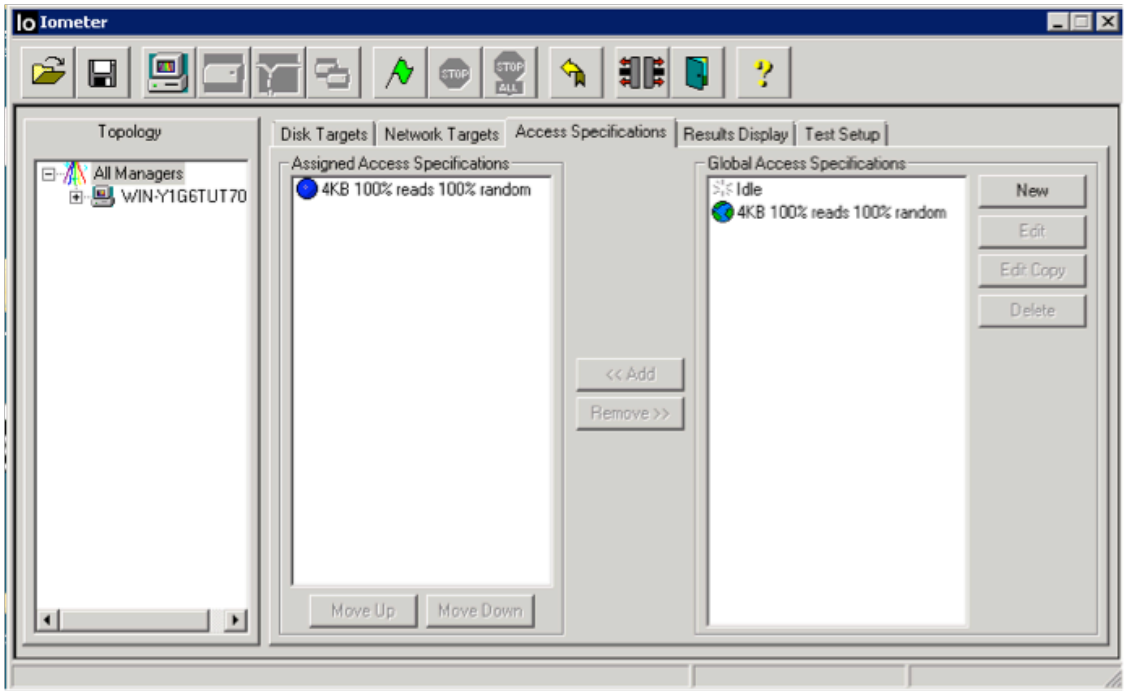
For the access specification transfer size, select 32KB, 100% sequential reads aligned on 32KB boundaries.



Next, Create SAS or SATA HDD volume configured to Write Back/Cached IO/No Read Ahead/Drive Write Cache Enabled/Stripe Size 64KB.



Now perform 4KB Random Reads (QD=64) across the 30GB region with two workers. Align the I/Os on 4Kb sectors. IOmeter should measure roughly 39,000 IOps for single Intel X25-E SSD. Steady state performance will not be reached until the SSD cache is fully loaded. 140,000 IOps is the absolute maximum MegaRAID controller 4KB random read performance for CacehCade. Typically, with more capacity added to the Cache-Cade pool, you should be able to measure approximately 60,000 to 100,000 IOps.



Other Tips:

- Pre-load CacheCade software and run IOMeter before demoing with customers.
- Configuration option for a larger number of SSDs in the CacheCade solution: LSI 620J enclosure with 16 each 2.5" SAS HDDs with room for up to 8 each 2.5" SSDs, allowing connection via two x4 SAS cables.

Full Disk Access IOMeter Benchmark for CacheCade Software

This benchmark exercise simulates a stressful real world workload emulating one or more applications accessing a large data set at different rates, creating graduated regions of access frequency. It provides an excellent test of CacheCade read caching technology and demonstrates the effectiveness of hot read data detection and retention.

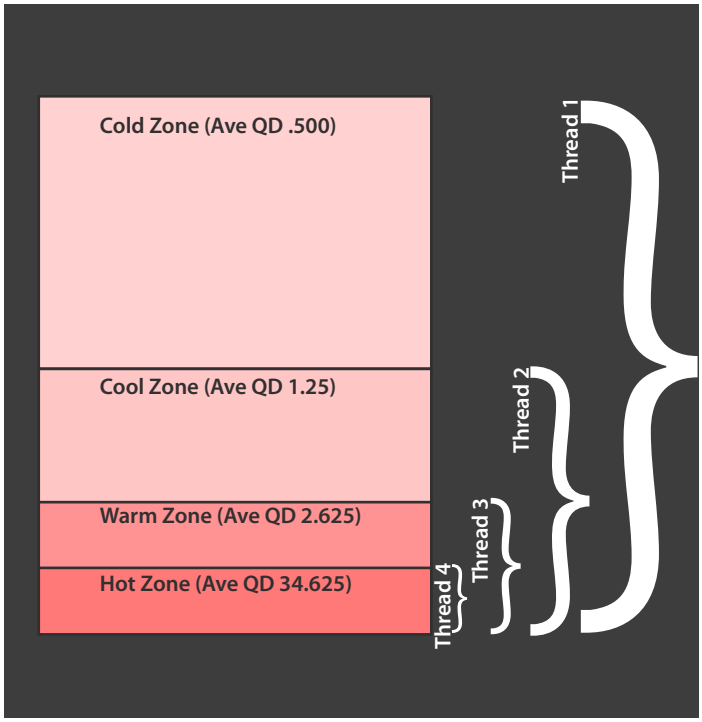
- Use multiple workers accessing overlapping HDD VD sector ranges
- Emulates different users and/or applications accessing different but overlapped HDD VD ranges simultaneously
- Emulates a small hot data range with progressively larger warm, cool and cold data access ranges.
- More frequent accesses to smaller hot data range and progressively less frequent accesses to warm, medium and cold ranges.
- Steady state performance will not be reached until CacheCade cache is fully loaded.

Figure one below illustrates a series of workers, or threads, that are accessing various regions of VD that encompasses the entire disk group. These overlapping access regions represent different degrees of disk activity ranging from hot (thread 4) to cold (thread 1). The cold data range is set to 100% of the HDD VD capacity (thread 1) down to 12.5% for hot data range (thread 4). The queue depths (QD) represent disk IO activity (QD=1 is the lowest and QD=32 is the highest). The hot data range should be set to less than the SSD cache capacity.

- Thread 1 = 100%, QD=1
- Thread 2 = 50%, QD=2
- Thread 3 = 25%, QD=4
- Thread 4 = 12.5%, QD=32

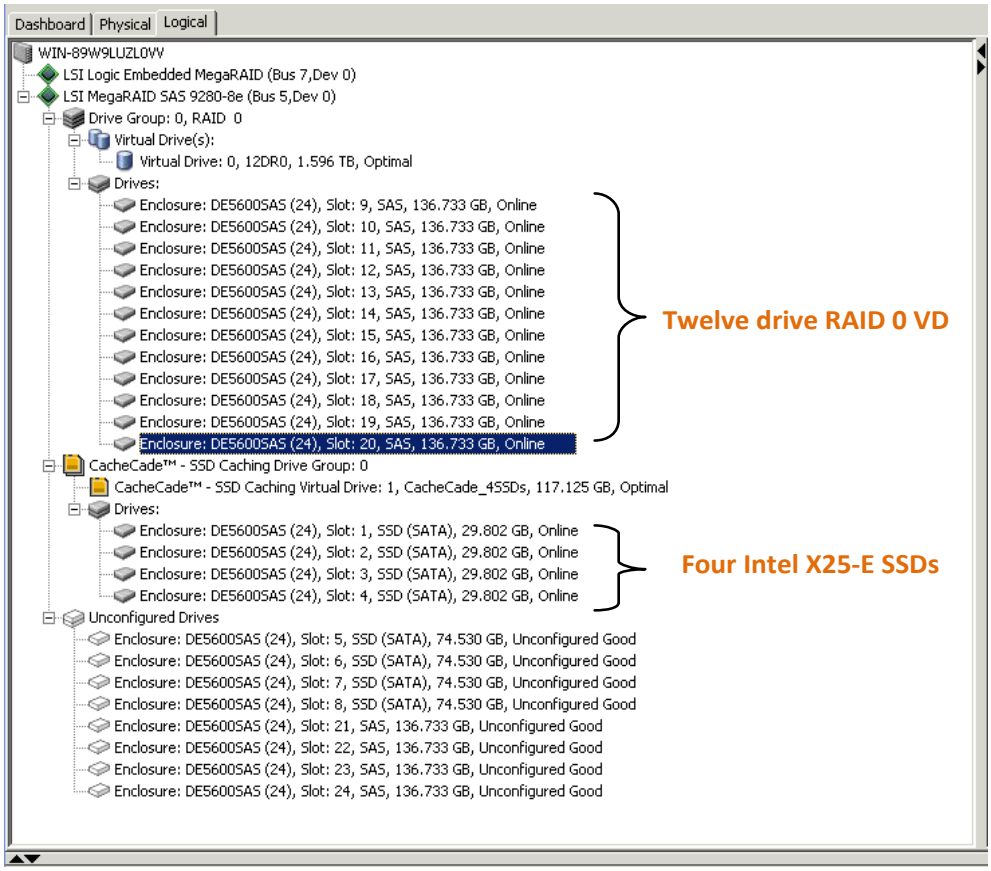
Make sure SSD capacity is greater than hot data or greater then hot and warm data if you choose to cache both. I/O is run for 5 minutes to prime the cache, then performance is captured for 1 minute.

Figure 1: Full disk access illustration.



Full Disk Access IOMeter Setup

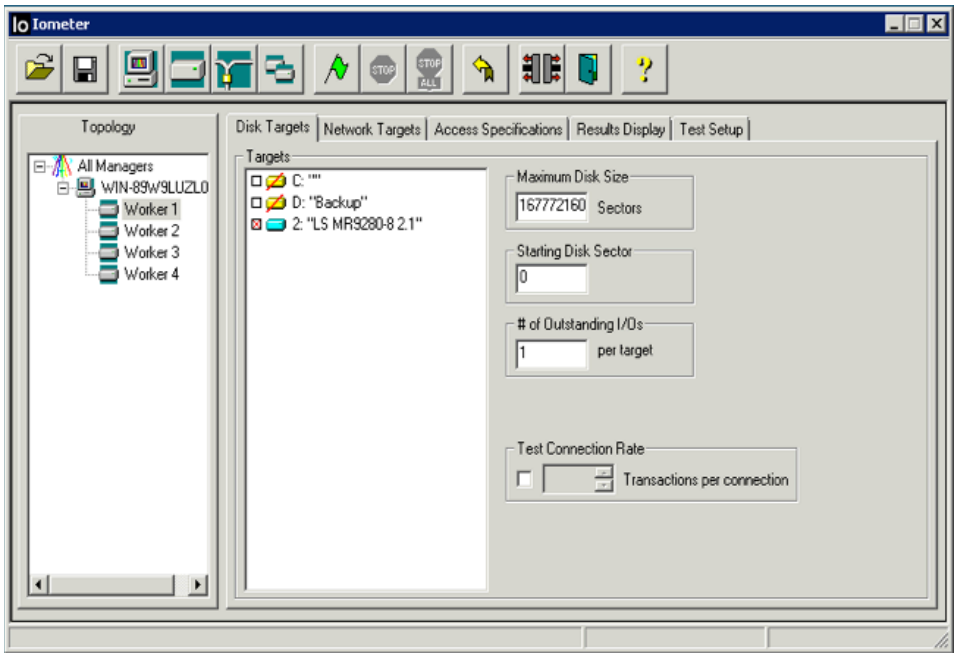
The system configuration used for this example is based on a RAID-0 volume consisting of twelve (12) 137GB 6Gb/s SAS HDDs. Four (4) Intel X25-E SATA Solid State Drives with approximately 29.8GB of usable capacity for total of 119GB. The SSDs had been previously used for several months of strenuous testing, thus they were very worn in, which is ideal for this exercise.



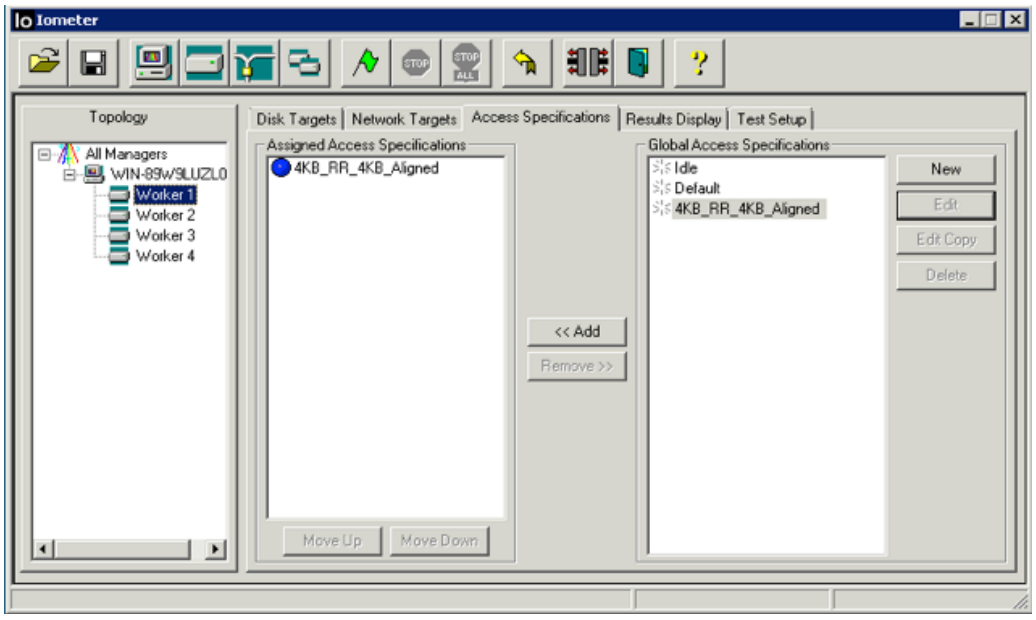
The Table below illustrates the queue depths and disk sector ranges that were used for each IOMeter worker or thread. As you can see, the Hot read data region sized to fit within the available CacheCade SSD cache capacity indicated above.

REGION	QUEUE DEPTH	START DISK SECTOR	MAX DISK SECTOR SIZE	RANGE CAPACITY (GB)
Cold	1	0	1677721600	800GB
Cool	2	0	838860800	400GB
Warm	4	0	419430400	200GB
Hot	32	0	209715200	100GB

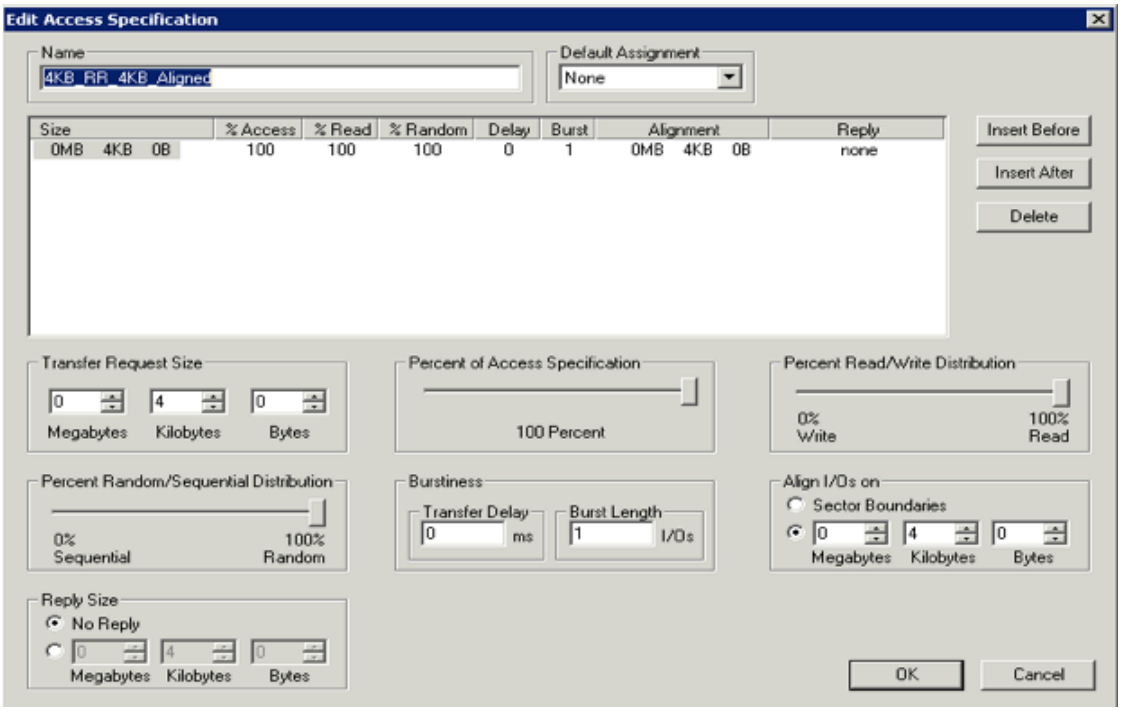
From the "Disk Targets" tab, select the appropriate raw storage device. Enter the "Maximum Disk Size" for each worker, which represents the Cold, Cool, Warm and Hot regions



Next, select the "Access Specifications" tab to define the IO profile to be tested.

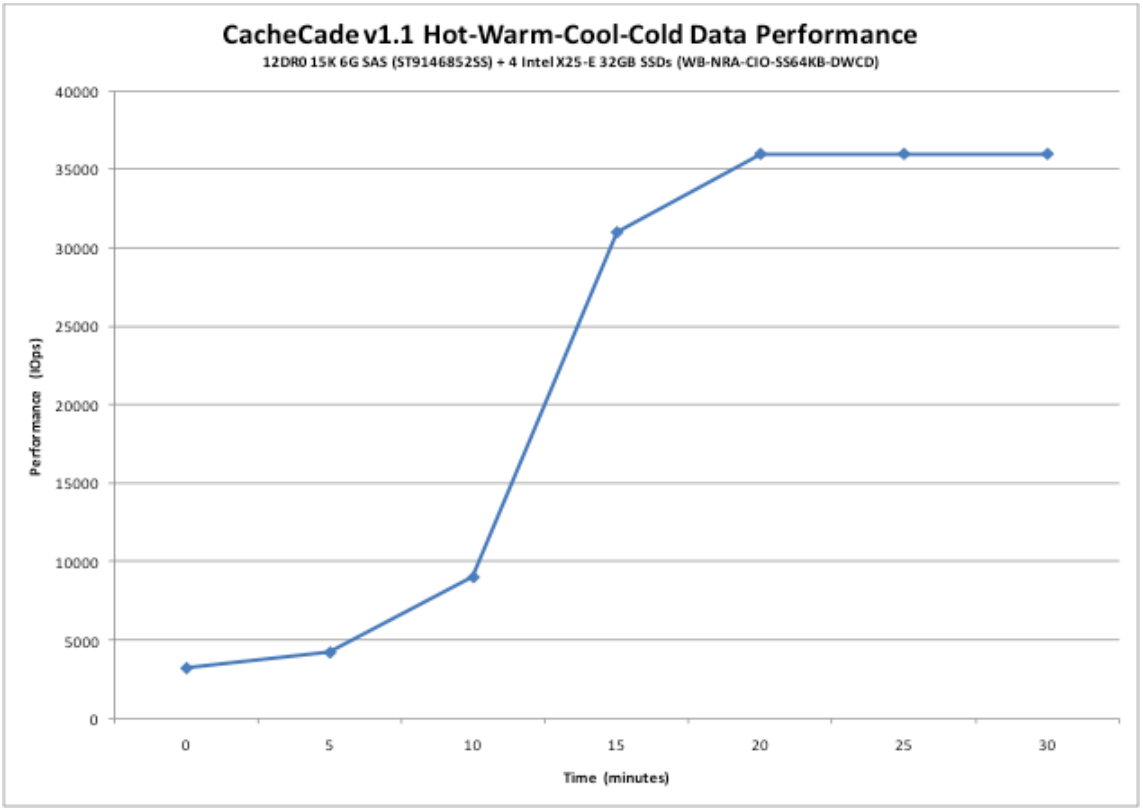


Run IOMeter 4K 100% random read access profile to obtain maximum IOPs performance. You can ether edit one of the pre-defined IO profiles, or select the "New" button. When editing the access specification, make sure that IOs are aligned to the transfer request size. In this example the transfer size 4KB and the IOs are aligned to 4KB as well.



Next select the Test Setup tab. Make sure you select a run time that is long enough to allow CacheCade to fully populate the SSDs with hot data.

The first few minutes of the test will indicate the read performance of the HDD array as data is being read and re-read for the first time. Once the CacheCade technology identifies data as hot and copies it in to the SSD cache pool, IOPs performance will dramatically scale, as indicated in the chart below.



Using IOMeter to Demonstrate Optimum FastPath Software Performance

Note: SSD pre-conditioning procedure as described above should be performaned on a FastPath SSD VD especially before demonstrating random write performance.

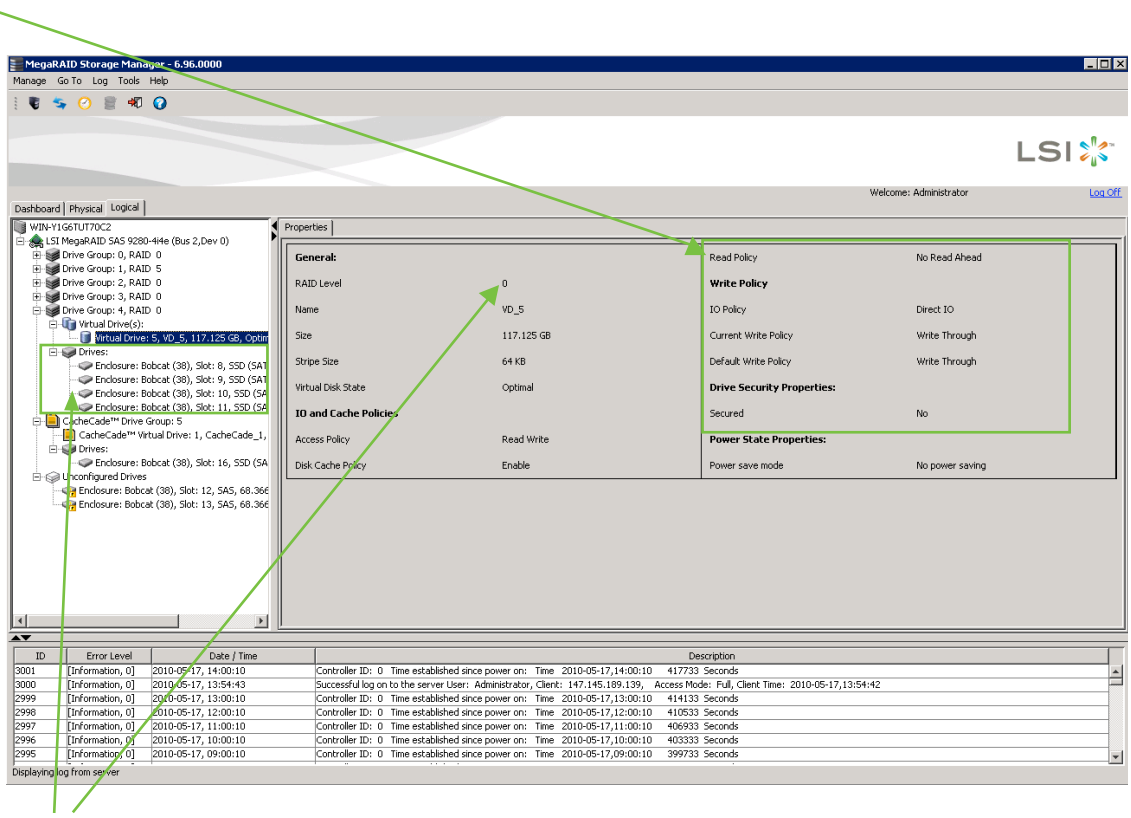
Optimum Controller settings for FastPath

Write Policy: Write Through

IO Policy: Direct IO

Read Policy: No Read Ahead

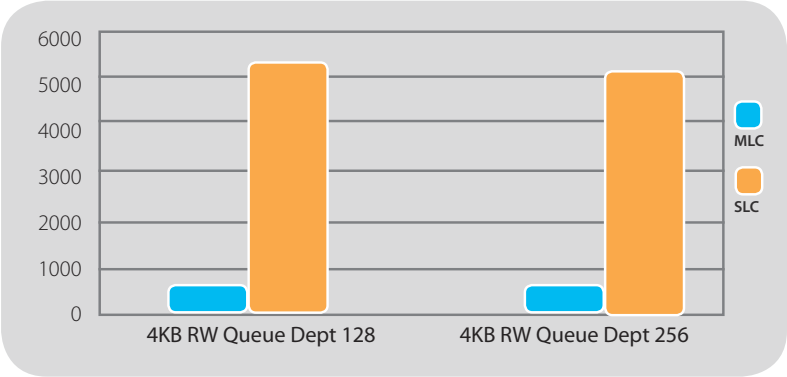
Stripe Size: 64KB



4 SSDs in a RAID 0 volume.

Run IOMeter 4K 100% random read access profile to obtain maximum IOPs performance. We also recommend running 4KB random write performance. While 4KB random write performance is much lower than 4KB random reads, it is still significantly better than HDD performance.

Note that CacheCade and FastPath performance depends on SSD performance. SSD performance depends on make, model, firmware, technology, specifications, and number of SSDs used. It is important to select the right SSD to achieve customer's performance requirements. The chart below displays IOMeter benchmark results on an MLC and an SLC SSD. The single drives were configured as RAID 0 arrays. 4KB random writes were performed on the SSDs over two hours.



MegaCLI Commands to Enable Fast Path and CacheCade Software

(Commands below assume CacheCade and Fast Path feature keys already installed.)

MegaCli -LDSetProp {-Name LdNamestring} | -RW|RO|Blocked | WT|WB [-Immediate] |RA|NORA|ADRA | Cached|Direct | -EnDskCache|DisDskCache | CachedBadBBU|NoCachedBadBBU -Lx|-L0,1,2|-Lall -aN|-a0,1,2|-aALL

Use the following commands on the LD number you want to modify. So if you had only one LD, that being LD0, the command would be:

FastPath software:

megacli –LDSetProp WT Direct NORA L0 a0

Or you can issue these one at a time such as:

Megacli –LDSetProp WT L0 a0

Megacli –LDSetProp Direct L0 a0

Megacli –LDSetProp NORA L0 a0

CacheCade software:

Megacli –LDSetProp WB NORA Cached CachedBadBBU L0 a0

Or you can issue these one at a time such as:

Megacli –LDSetProp WB L0 a0

Megacli –LDSetProp NORA L0 a0

Megacli –LDSetProp Cached L0 a0

Megacli – CachedBadBBU Cached L0 a0 (Cached Bad BBU command will make sure the controller stays in Write Back mode with no battery connected.)

The stripe size is set at creation time, and cannot be changed with LDSetProp.

** A 14GB volume on a two drive RAID 1 array was used to simulate web server re-read hot spot activity. Using IOMeter version 2006.07.27, a MegaRAID SAS 9260-8e with configuration settings: No Read Ahead, Write Back Cache, 64KB Stripe Size and Direct IO, using firmware version 2.60.03-0778, reached 14,896 IOs per second with CacheCade enabled, compared to 273 IOs per second with CacheCade disabled. This represents an increase of more than 50x.*

For more information and sales office locations, please visit the LSI web sites at:
lsi.com lsi.com/channel

Feb 2011