RAID Cache Benefit for Avago® 12Gb/s SAS MegaRAID® Controllers

White Paper

March 2015

DB10-000024-00

Corporate Headquarters	Email	Website
San Jose, CA	globalsupport.pdl@avagotech.com	www.lsi.com

Table of Contents

RAID Cache Benefit for Avago® 12Gb/s SAS MegaRAID® Controllers White Paper

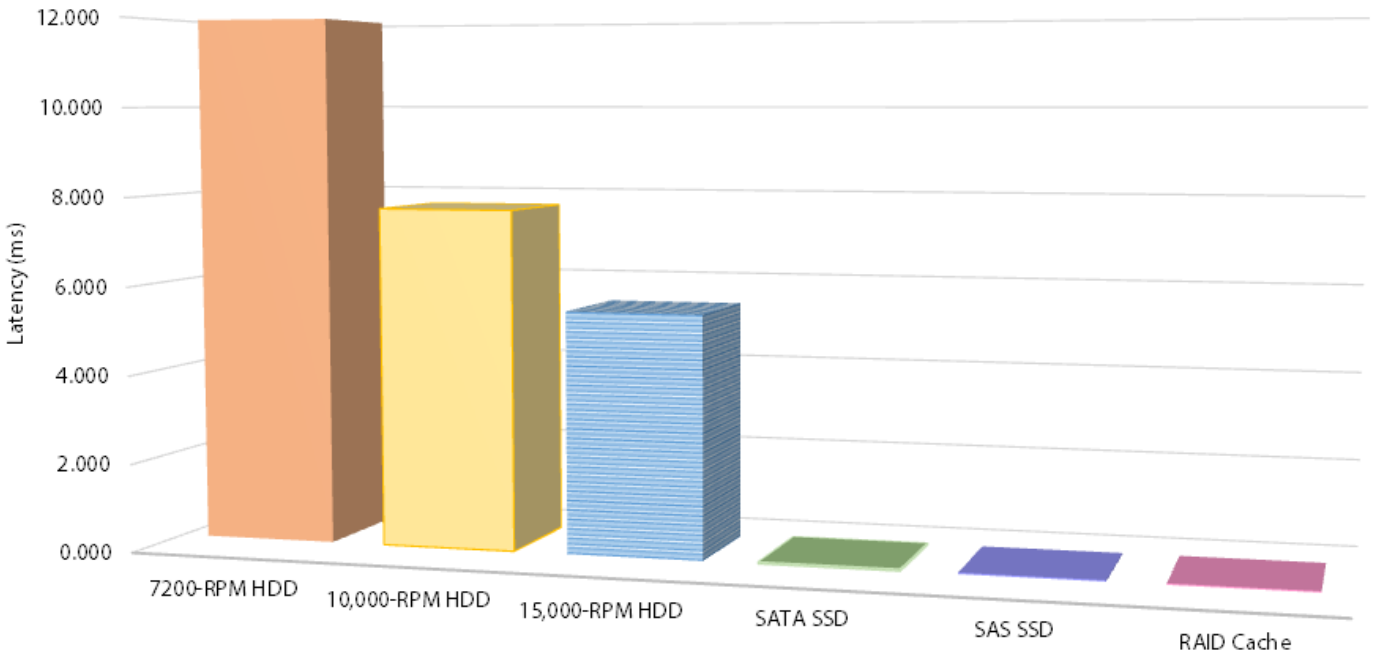
1 Introduction

Avago® 12Gb/s SAS MegaRAID® controllers, featuring the dual-core SAS3108 RAID-on-Chip (RoC) processor, offer significant performance enhancements for solutions architected with 12Gb/s SAS or 6Gb/s SAS drives. The superior read/write performance suits the controllers for a broad range of application workloads:

- Enterprise data center applications, such as email, file servers, transactional databases, and analytical databases
- Cloud computing, software-defined storage systems, and big data (Hadoop®)
- Content applications, such as streaming video and cold storage/archival

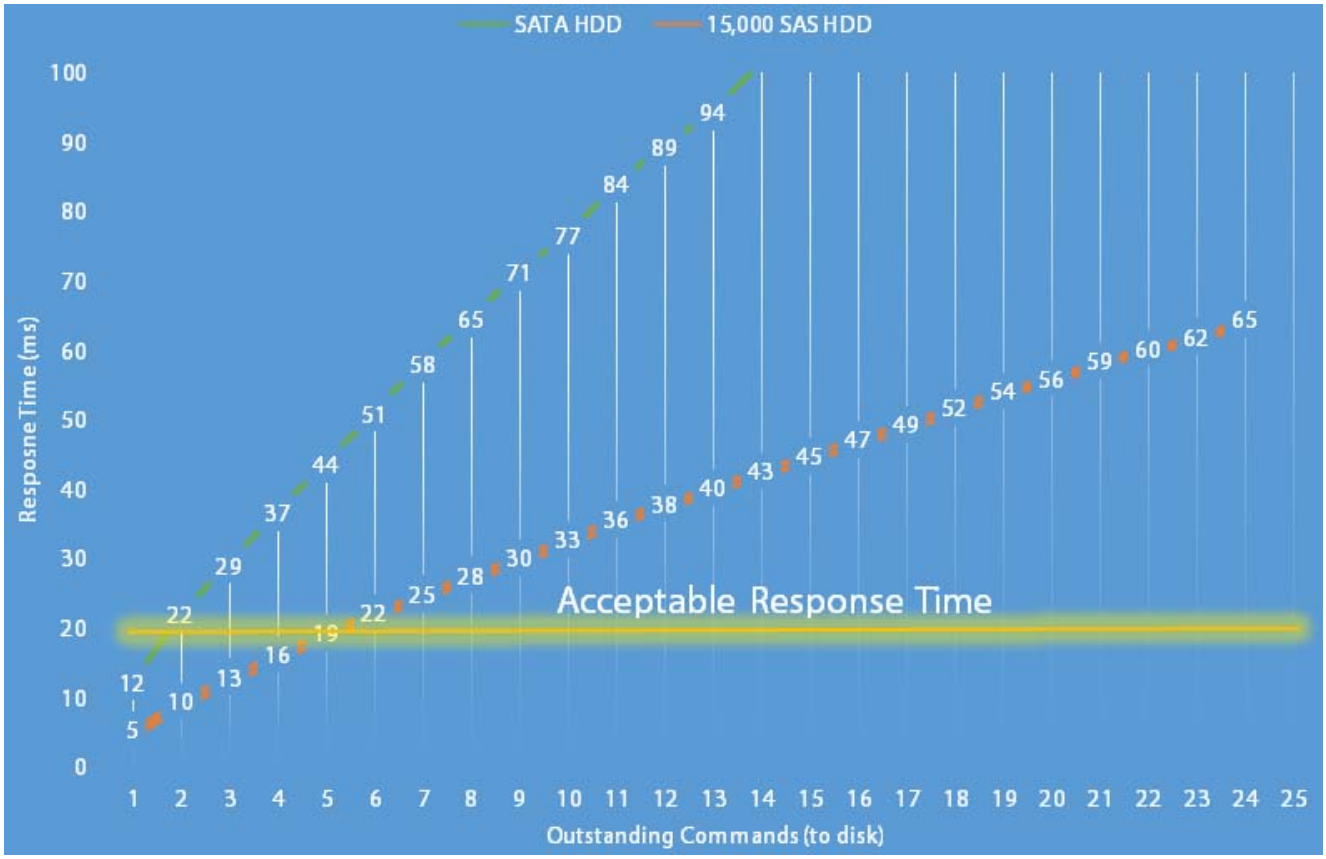
Today's converged storage environments contain many different storage technologies. Traditional hard disks are still most cost effective with choices such as enterprise level 15,000, 10,000 and 7,200 RPM varieties that can be used based on performance and cost requirements. Flash-based SAS and SATA devices add another performance spectrum with IOPs capabilities that exceed 150,000 IOPs per device and latencies less than 35 µs. The following figure shows the typical random write response time for different drive types when queue depth is 1.

Figure 1 Typical Latencies in the Storage Stack



SATA HDDs are well-suited for some workloads; however, they quickly reach non-linear response times when given heavy workloads. For instance, many vendors recommend an average database latency of no greater than 20 ms to ensure reasonable application response for end users. With only two commands, a SATA disk exceeds this recommendation. Even the fastest 15,000-RPM disks exceed 20-ms latency with only six commands. The following figure shows the 4-KB random writes for single 7200-RPM SATA and 15,000-RPM enterprise SAS disks.

Figure 2 Write Latency versus Queue Depth



Latencies shown in the previous two figures are quickly becoming unacceptable for the demands of today's storage environment. Implementing a RAID cache product can help, as described in following sections.

1.1 Architectural Benefits of RAID Cache

Avago hardware RAID controllers use DDR SDRAM to help buffer writes to a disk and provide a fast read cache. With intelligent cache algorithms, writes can decrease as much as 375 times when compared to a 7200-RPM SATA disk without cache. Bursty writes can post to the RAID cache and be written back to the disk during inactive periods.

For spinning media, seek time dominates the latency for random reads and writes. If you reduce the seek distances, such as by short stroking the drive or grouping commands to optimize seek overhead, a disk can achieve higher performance. The Avago MegaRAID cache algorithms are highly optimized to assist with sorting, which enable the disks to produce more IOPs. For example, an enterprise 15,000-RPM SAS HDD can achieve a maximum of approximately 400 IOPs when given enough commands. You could easily achieve over 800 IOPs by using intelligent RAID caching that provides optimal sort algorithms.

Additionally, when RAID cache handles write requests to the disk, the disks are free to service reads (that are not already in cache). These cache features allow administrators to improve their disk performance. More cache means more writes that can be handled during bursts and more requests that can be sorted. Read cache also provides significant benefits for certain workloads, such as multiple read or write threads when adjacent requests can be grouped to reduce disk seeks.

You can also use flash-based cache, such as MegaRAID CacheCade® Pro 2.0 software, to improve performance and latency improvements for HDD-based RAID volumes. However, flash is susceptible to wear from repeated write and erase cycles, whereas a RAID cache based on DRAM is not.

It can be challenging for administrators and storage designers to understand if RAID cache will benefit their environment. Choosing the correct RAID cache size for workload and performance requirements also presents a challenge. Avago 12Gb/s MegaRAID SAS controllers provide many RAID-cache options, with 1-GB and 2-GB cache sizes, and options to disable read or write cache on a per virtual disk (VD)-based policy. The next section provides performance comparisons between these options to allow informed decisions when designing a storage infrastructure. Keep in mind that DRAM is a volatile memory, and should be protected by using CacheVault™ or batttery backup units designed for your specific RAID controller.

2 Performance Tests and Results

Using IOMeter 1.1, an open-source I/O workload generator, Avago designed I/O workload profiles to demonstrate how real-world applications might perform with various cache settings and values. Sequential read and write benchmarks reveal how streaming applications perform by simulating multiple enterprise workloads that typically include mixed and overlapping reads and writes across request sizes, similar to as the real-world applications in the following table.

Table 1 Real-World Applications That Benefit from Cache

Workload	Typical Applications
File and email servers	Structured file systems
Enterprise databases	Online transaction processing (OLTP), online analytical processing (OLAP) transactions, email applications
Multithreaded reads	Video on demand, Hadoop Framework, Hadoop Distributed File System (HDFS™), cloud content, archival backup
Multithreaded writes	Hadoop source ingest (Apache Sqoop™ or Flume™), video surveillance
Analytics	SQL, NewSQL and NoSQL databases, MapReduce

2.1 Streaming Applications

Read and write cache provides significant improvements to multithreaded sequential environments. Because multiple threads accessing different regions can appear random to the disk, employing read ahead and write combining techniques can greatly increase performance. The following examples use a 24-drive RAID 10 with 15,000-RPM SAS HDDs and show that cache provided up to four times higher read performance and 40 times higher write performance simply by minimizing seek overhead.

Figure 3 Multithreaded Reads

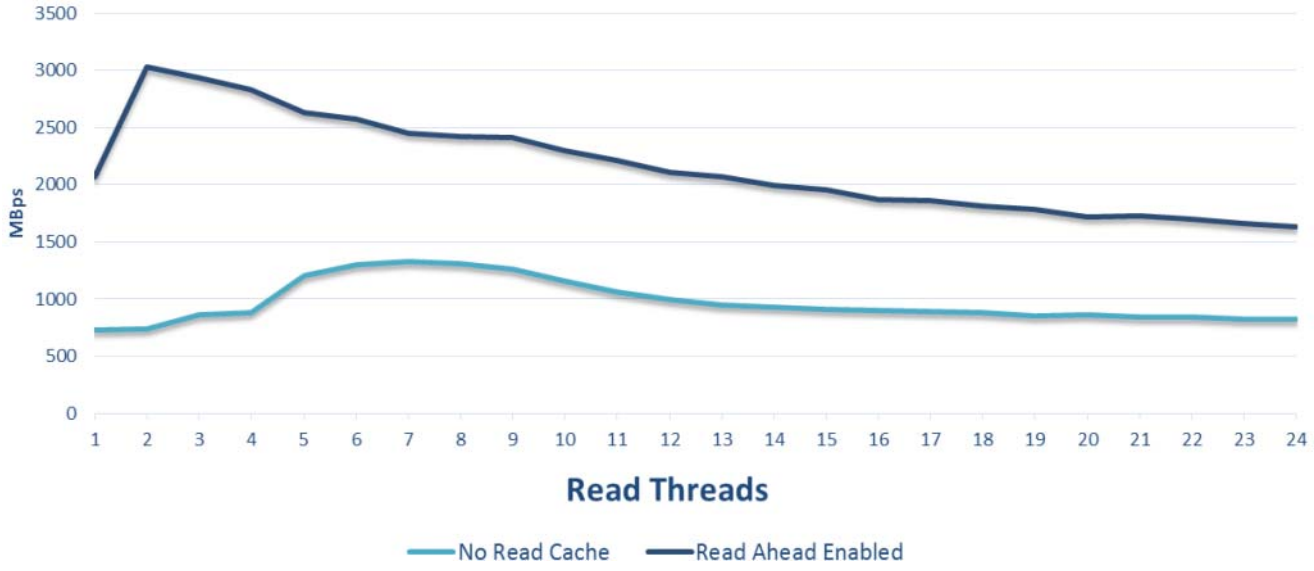
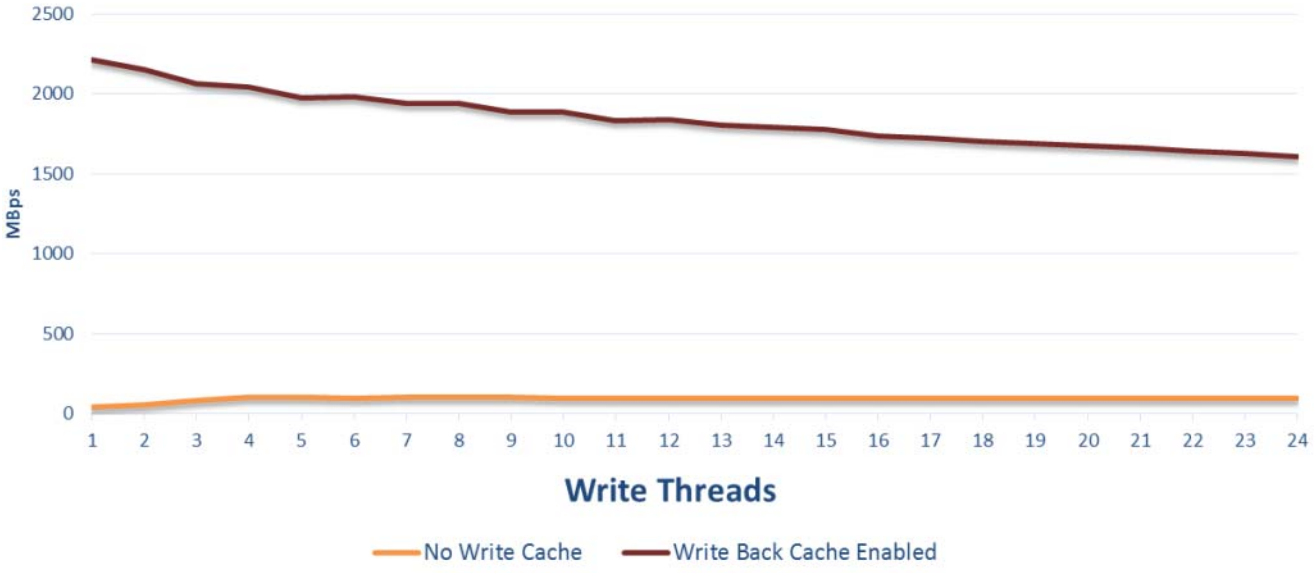


Figure 4 Multithreaded Writes

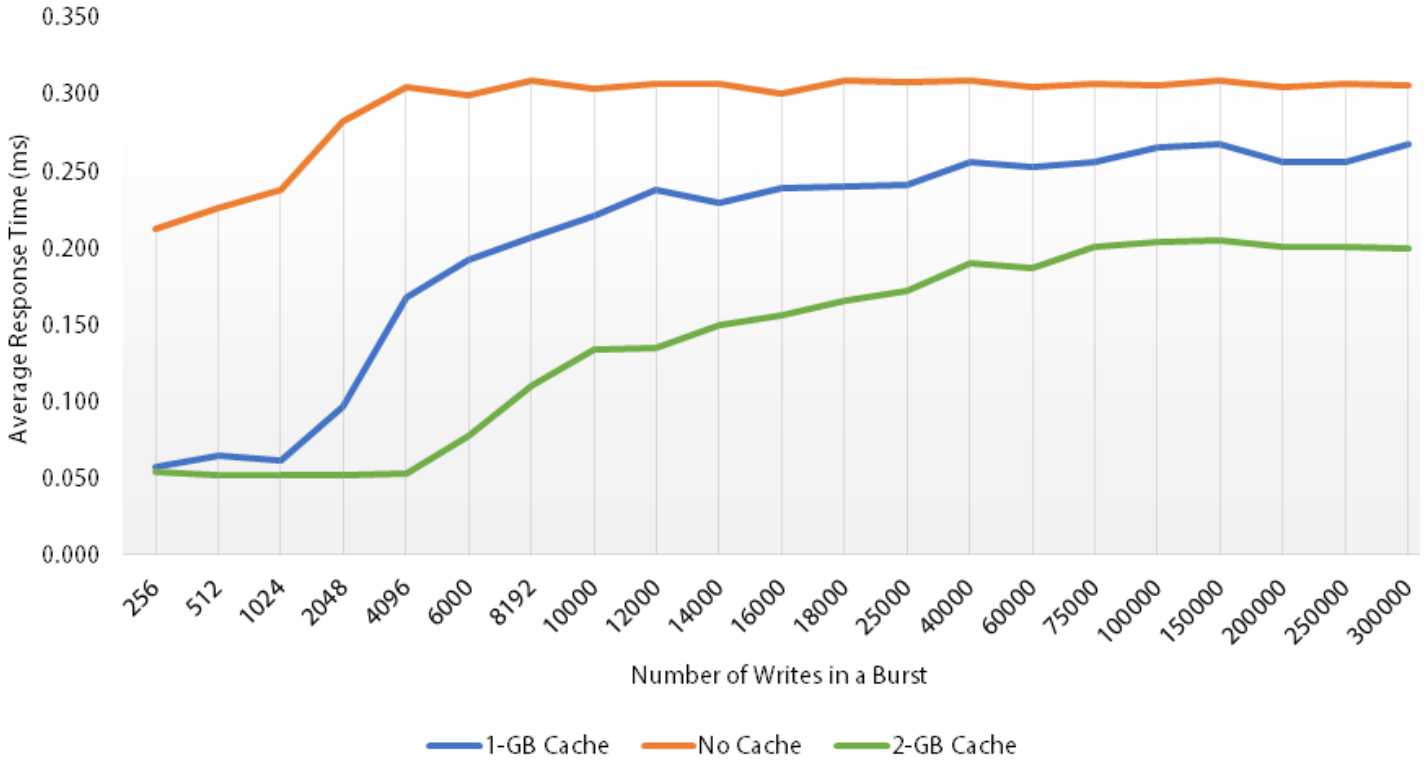


2.2 Transaction-Oriented Applications

Transaction-based applications, such as databases, are very sensitive to latency to make sure SLA requirements are met. While many databases leverage host memory as cache to speed up operations, this cache must be periodically flushed to persistent media (via checkpoints) to reflect changes on the disk. Handling this flurry of writes with the lowest response time helps ensure the highest database performance.

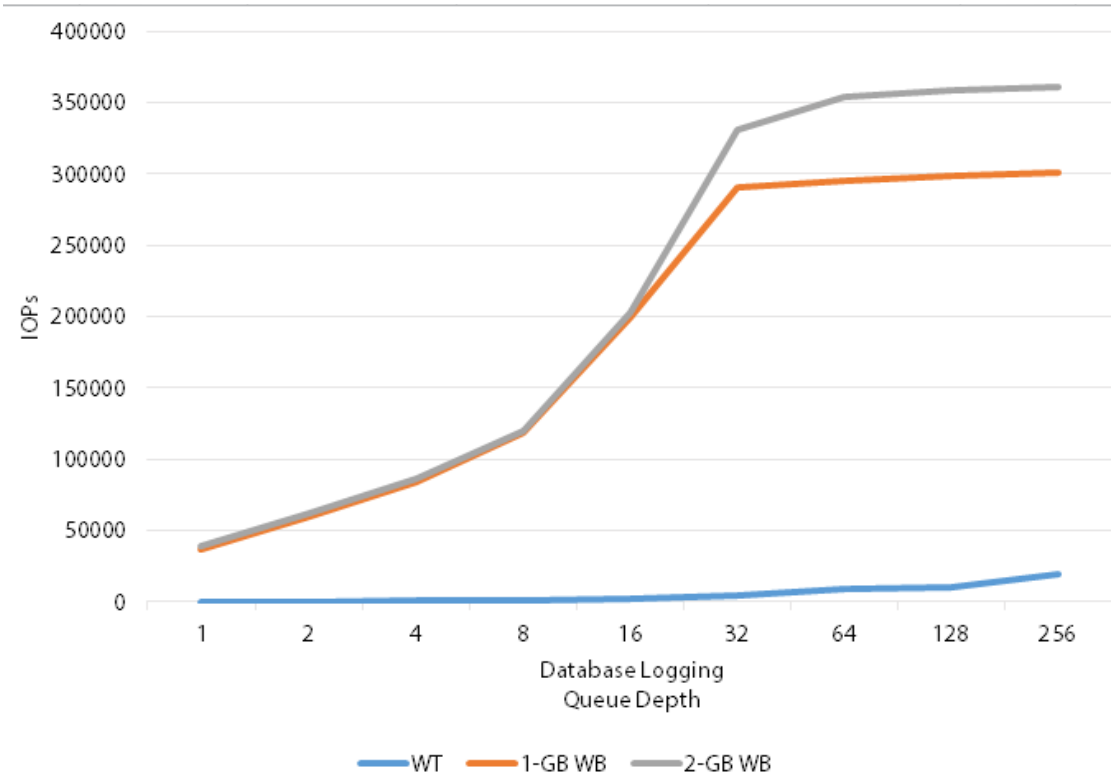
The following chart shows the latency when a burst of 16-KB random writes are sent to an eight-disk array of enterprise class 15,000-RPM disks in five-second sequences. Without cache, the response time is four times higher, and 2-GB cache clearly handles many more bursts than 1-GB cache while maintaining a lower average latency.

Figure 5 Bursty Write Performance versus Response Time (response times averaged over 10 seconds)



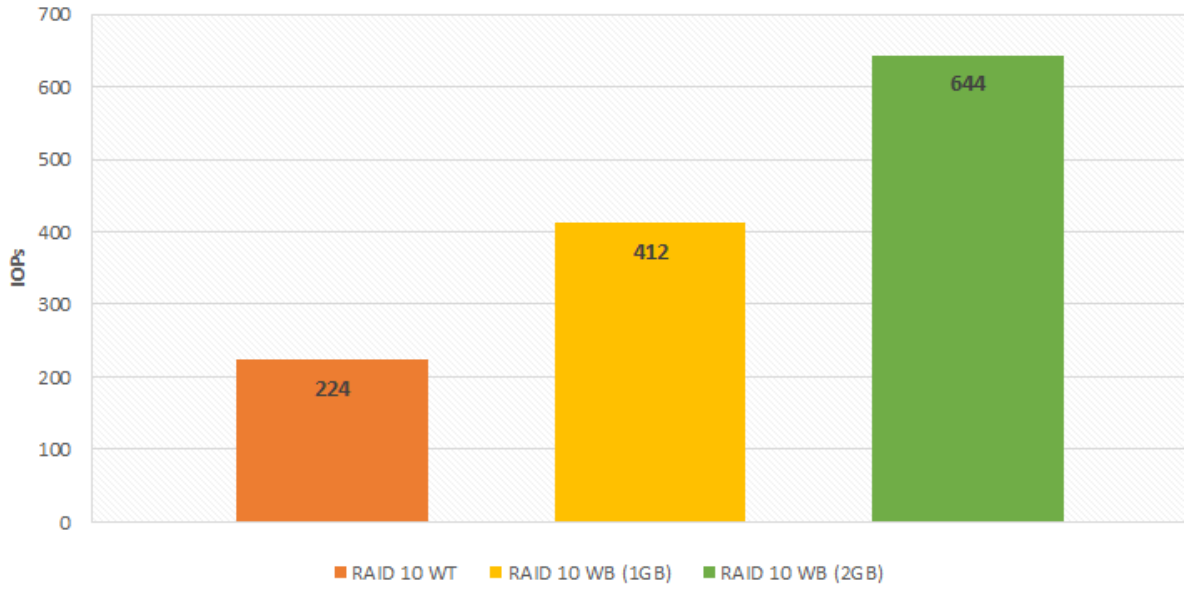
In addition to database requests and checkpoints, databases also use logging (or, journaling) to guarantee ACID properties in case of failure. Maintaining high performing (and low latency) logging is a vital feature of a healthy database. Log files tend to be small and primarily sequential, and we can imitate logging workloads with sequential writes from 512 b to 16 KB. The following chart compares no cache, 1-GB cache, and 2-GB cache log performance benefits. The RAID cache benefits for logging workloads has 142 times higher performance when enabling write back cache.

Figure 6 Log Performance Cache Benefits



Ultimately, database performance depends on how quickly you can execute database commands such as reads, writes, and verifies. Artificial benchmarks that imitate a 2:1 read to write ratio simulate this command sequence. The following chart shows the cache benefit for OLTP in no cache, 1-GB cache, and 2-GB cache environments. Performance nearly doubles with 1-GB cache, while 2-GB cache provided an additional 220 IOPs.

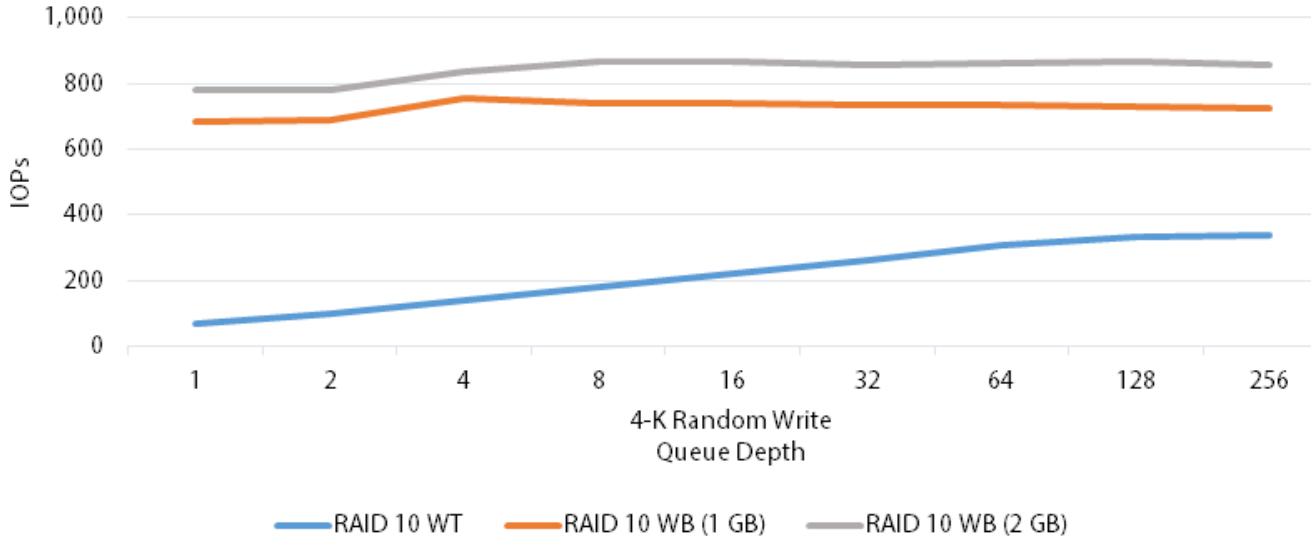
Figure 7 Database Transaction Performance Benefits



2.3 Common Enterprise Applications

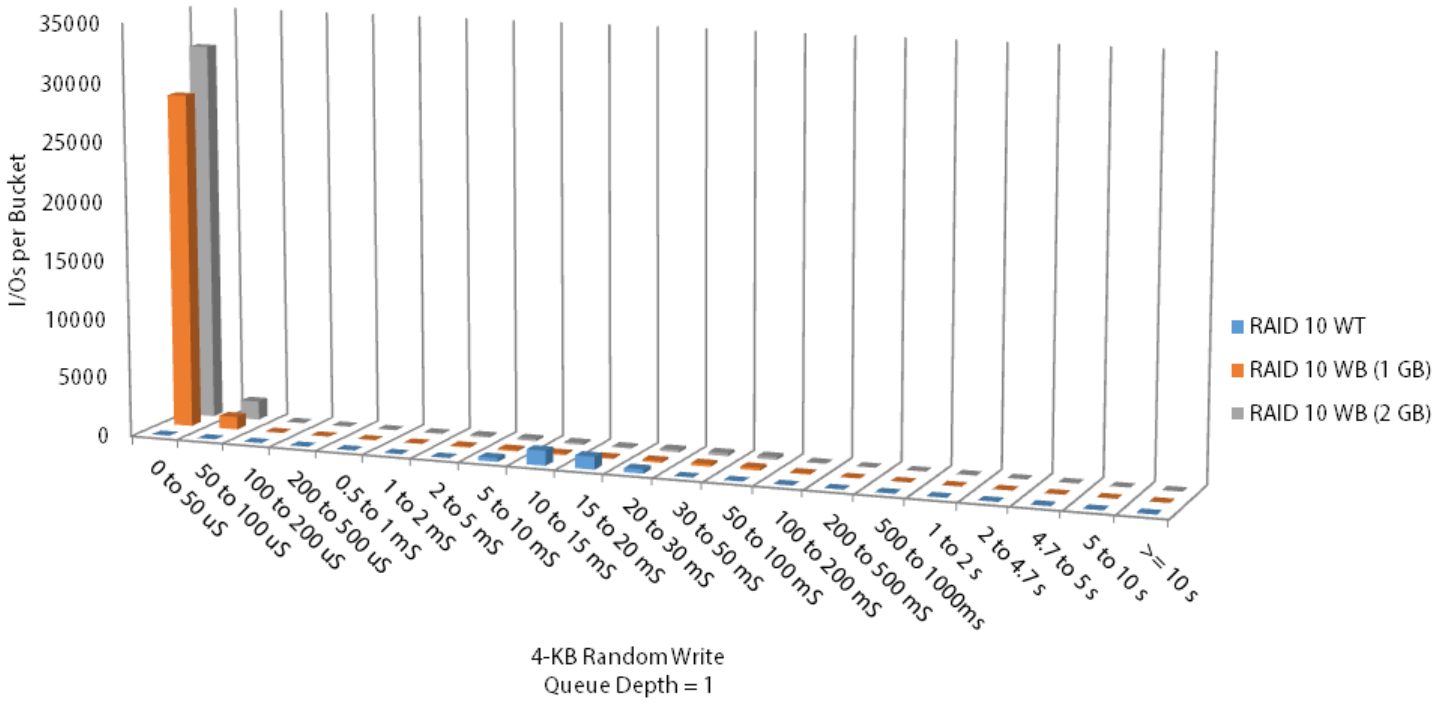
Random writes (small I/O sizes typically less than 64 KB not located adjacent to each other) are a common component of most enterprise disk workloads including email servers, e-commerce servers, database servers, virtualized environments, research, and analytics. Cache provides significant benefit to these workload types by improving the IOPs and reducing the overall latency. These benefits provide a better user experience, allow higher concurrency, and improve worker production. The following chart shows an Avago performance test with 4-KB random writes using no cache, 1-GB cache, and 2-GB cache. With 2-GB cache, Avago measured up to 10 times more IOPs when compared to no cache, and with 1-GB cache Avago achieved up to eight times more IOPs than no cache.

Figure 8 4-KB Random Write Performance Benefits



In addition to the IOPs benefit, implementing cache significantly lowers response times. The following chart shows the response times collected in the same three configurations: Write Through (no cache enabled), 1-GB Write Back cache, and 2-GB Write Back cache.

Figure 9 4-KB Random Write Response Time Histogram



The write back configurations handled many more I/Os during the collection period, and most completed in less than 100 µs (2-GB cache saw 15% more completions in less than 50 µs). The no cache environment completed much fewer I/Os with response times centered around 10 ms to 30 ms.

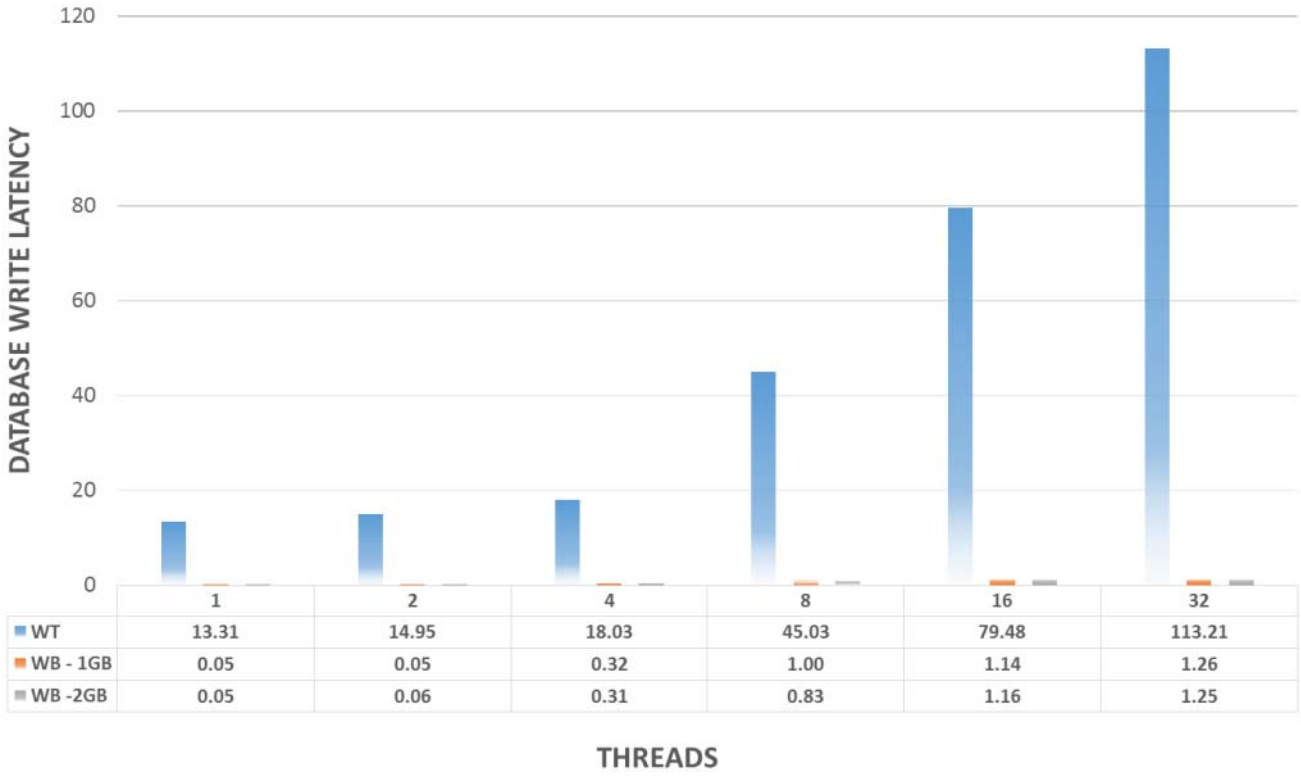
2.4 Microsoft Exchange Server 2013 Benchmarking

The Jetstress tool, provided by Microsoft, simulates Microsoft Exchange disk I/O load on a server to verify the performance and stability of your disk subsystem before you put a server into a production environment. The workload consists primarily of random 16-KB read and write accesses. RAID cache still provides significant benefits by providing low latency log writes and optimized disk writes. The Jetstress benchmark is paced based on the number of threads, the response time, and the maximum average database access times less than 20 ms.

In this simulated environment, caching provides up to 300 times lower log and database write latencies with a 50% increase in the number of transactions per second processed (in less than 20 ms) with 2-GB cache, and a 40% increase in the number of transactions per second processed with 1-GB write back cache.

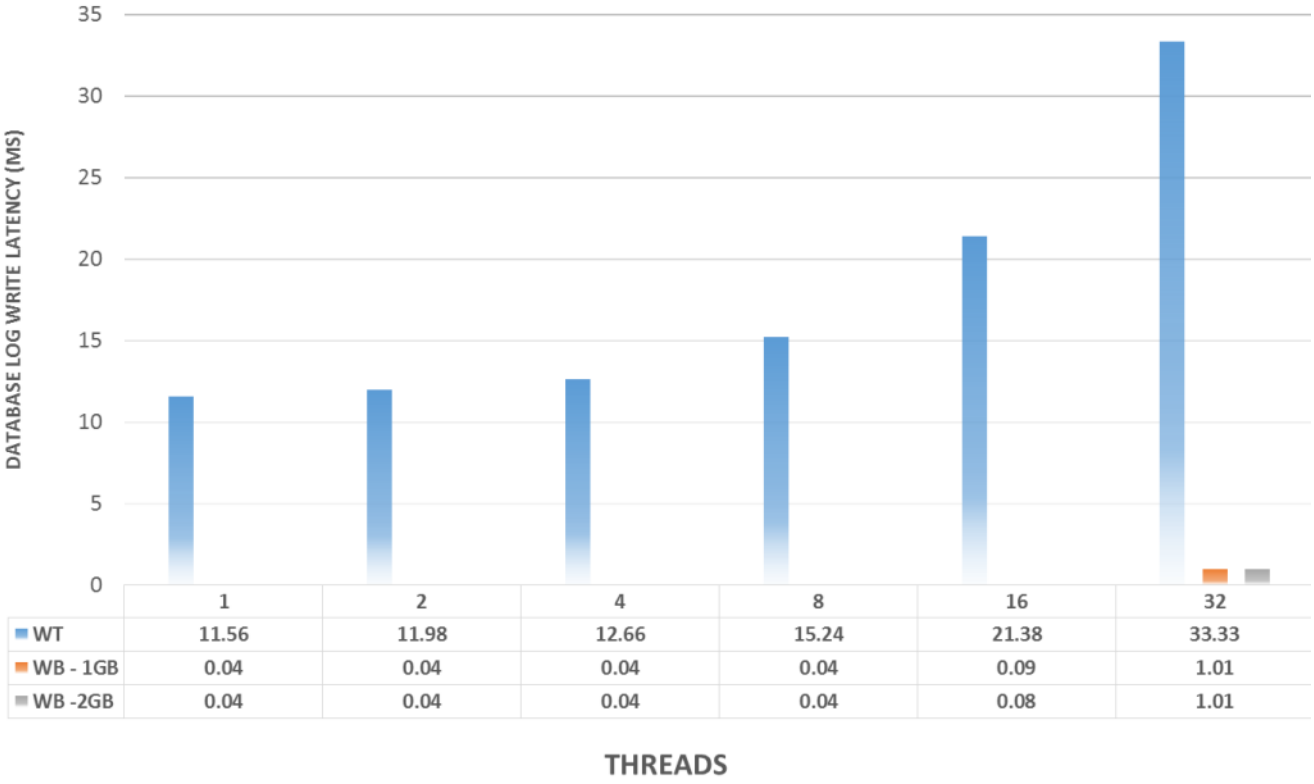
The following chart shows response times in no cache, 1-GB cache, and 2-GB cache environments that implement six single-disk RAID0 1-TB 7200-RPM SATA disks. The Jetstress test used 3000 mailboxes with 200-MB per mailbox and six Jetstress databases.

Figure 10 Exchange Server Database Write Response Times



The following chart shows response times in no cache, 1-GB cache, and 2-GB cache environments that implement six single-disk RAID0 1-TB SATA disks. The Jetstress test used 3000 mailboxes with 200-MB per mailbox and six Jetstress databases.

Figure 11 Exchange Server Database Log Write Response Times



	1	2	4	8	16	32
WT	11.56	11.98	12.66	15.24	21.38	33.33
WB - 1GB	0.04	0.04	0.04	0.04	0.09	1.01
WB -2GB	0.04	0.04	0.04	0.04	0.08	1.01

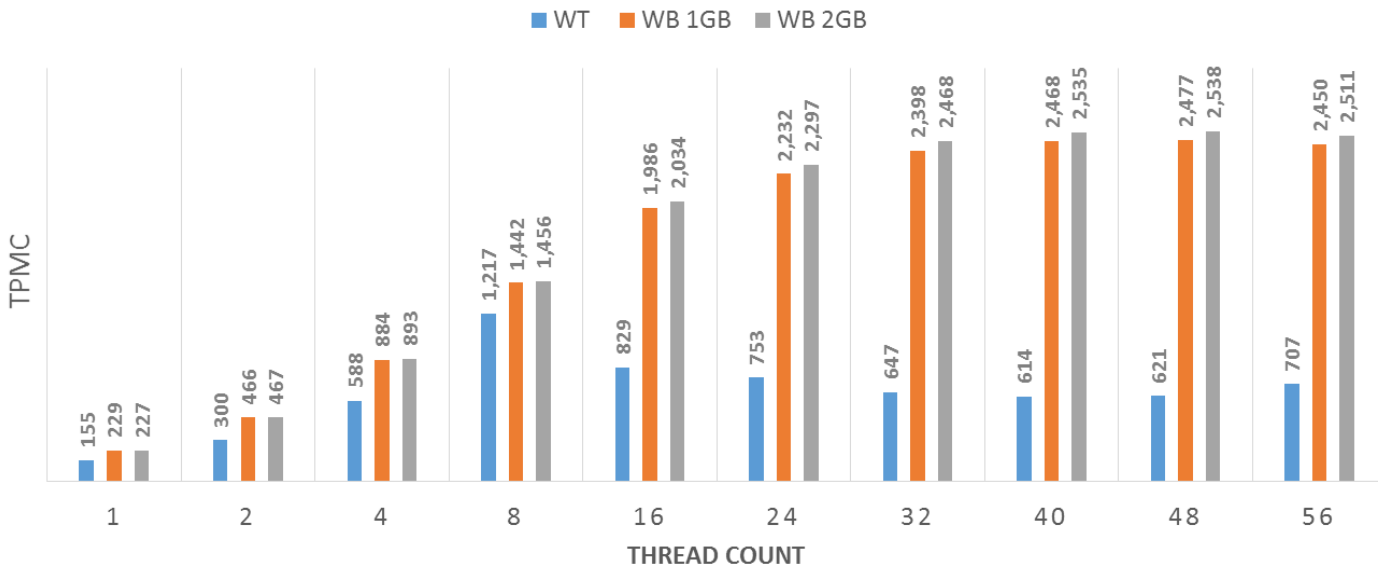
2.5 SQL Server OLTP Benchmarking

Using a standard SQL based application benchmark, Avago measured the transactional performance of 2,500 warehouse database scaling threads from 1 to 56 to determine RAID cache performance in a real database environment. The number of transactions per second increase by over 300%, and increased the concurrency capability from 8 threads to more than 50 threads.

In addition to the increase in completed transactions per second, the response time decreased by 60% with consistent and reliable response times. Without RAID cache, response times vary significantly over the cycle of checkpoints.

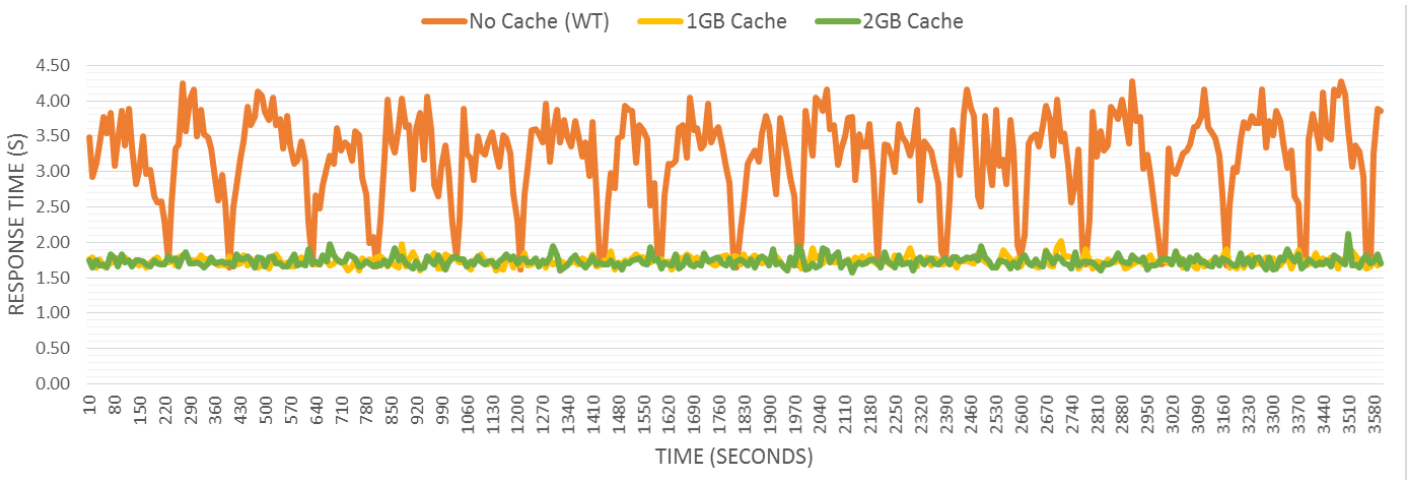
The following chart shows the 2,500 warehouse database using no cache, 1-GB cache, and 2-GB cache.

Figure 12 Transactions Processed per Minute versus Thread Count



The following chart shows the data response time in 10-second intervals for no cache, 1-GB cache, and 2-GB cache environments.

Figure 13 Database Response Time



3 Summary

A primary reason most system administrator and designers use RAID is to improve reliability, availability, and capacity. Adding RAID cache to a storage environment improves on those features with increased IOPs and consistency across many enterprise applications. Avago benchmarking shows significant improvements in IOPs and latency by using RAID cache for virtual disks comprised of spinning media. In addition, RAID cache can be used with flash-based cache such as Avago CacheCade Pro 2.0 software to accelerate performance of even larger datasets.

Today's small and medium business (SMB) and enterprise applications rely on caching to improve performance and increase production. Battery-protected RAID cache from Avago can provide significant enhancements to today's

applications by increasing overall productivity and reliability. The following table summarizes the benefit increase of 1-GB cache and 2-GB cache when compared to a no cache environment.

Table 2 Workload Cache Benefits

Workload	Metric	1-GB Cache	2-GB Cache	Detriment of No RAID Cache
Multithreaded Reads	MB/s	40x	40x	Disk seek overhead limits performance.
Multithreaded Writes	MB/s	400x	400x	Disk seek overhead limits performance.
Bursty Writes	Number of 16-K write commands that a single burst can absorb	> 2,000	> 4,000	All writes incur disk seek overhead.
Log Writes	IOPs	15x	20x	Low IOPs, unacceptable for most database applications.
OLTP	IOPs	85%	190%	Disk-limited performance.
Small Random Writes	IOPs	8x	10x	Disk-limited performance.
Small Random Writes	Response time (µs)	30 to 100	30 to 100[a]	Very high response times from 10,000 ms to 30,000 ms and dominated by disk-seek overhead.
Jetstress Database Write Latency	Decrease in response time (ms)	250x	250x	Low IOPS, unacceptable response times for SMB and Enterprise Exchange Servers.
Jetstress Database Log Latency	Decrease in response time (ms)	300x	300x	Low IOPS, unacceptable response times for SMB and Enterprise Exchange Servers.
Jetstress Transactions per Second	Increase in TPS	40%	50%	Only small thread counts (less than 3 per Jetstress database) generate acceptable database read response times.
Transaction Database Application Benchmarking	Increase in TpmCs	300%	300%	Scales only to 8 threads where response times throttle workload.
Transaction Database Application Benchmarking	Decrease in response times	60%	60%	Inconsistent response times, disks cannot handle database checkpoints which causes an unacceptable response time increase.

a. 15% more completions in less than 50 µs compared to 1-GB cache.


AVAGO
TECHNOLOGIES
STORAGE
BY LSI™