

Accelerating Server Storage Performance on Lenovo ThinkServer

Lenovo Enterprise Product Group

April 2014

© Copyright Lenovo 2014



LENOVO PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. This information could include technical inaccuracies or typographical errors. Changes may be made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any performance data contained herein was determined in a controlled environment; therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems, and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Any references in this publication to non-Lenovo Web sites are provided for convenience only, and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk.

The following terms are trademarks of Lenovo in the United States, other countries, or both: Lenovo, ThinkServer.

Intel and Xeon are trademarks of Intel Corporation in the U.S. and/or other countries.

LSI, the LSI and Design logo, Syncro, MegaRAID and CacheCade are trademarks or registered trademarks of LSI Corporation in the United States and/or other countries.

Microsoft, Windows Server 2012, SQL Server 2012, and the Windows Logo are trademarks of Microsoft Corporation in the United States and/or other countries.

Contents

1.0	Introduction	5
2.0	The Solution	5
2.1	SSD Storage Arrays.....	5
2.2	SSD Storage Arrays with FastPath Acceleration Software	6
2.3	HDD Arrays using CacheCade Pro 2.0 Software for SSD Caching	6
3.0	Evaluating the Solutions.....	7
3.1	Database Transaction Performance.....	8
3.2	Raw I/O Performance.....	11
3.3	Price Performance.....	13
4.0	Conclusions	14
5.0	Appendix A – Test Configurations.....	16
6.0	Appendix B – References	18

List of Figures

Figure 1 – Lazy Writes per Second	8
Figure 2 – Average Page Life Expectancy	9
Figure 3 – Database Performance (TPS)	9
Figure 4 – SQL Server Page IO Latch Latency	10
Figure 5 – CPU and Database Productivity	11
Figure 6 – Disk IO Performance (IOPs)	11
Figure 7 – CacheCade Performance Scales as Cache Fills with Active Data	12
Figure 8 – Average Disk Throughput (MB/s)	13
Figure 9 – Cost for Performance	14

List of Tables

Table 1 – Storage Configurations	8
Table 2 – Solution Comparison	15
Table 3 – Workload Qualifiers for CacheCade, and FastPath	15
Table 4 – Server Configuration for All HDD and Hybrid Arrays	16
Table 5 – Server Configuration for All SSD Arrays	16
Table 6 – Storage Configurations	17
Table 7 – Volume Policy Settings	17

1.0 Introduction

Server administrators continually face a growing amount of data to manage combined with an ever-increasing demand for better application performance. Modern data centers need storage solutions that can provide applications such as databases and on-line transaction processing (OLTP) with the high-performance data access that they require, while managing costs and generating a quick return on investment (ROI).

While capacity and cost per gigabyte of server hard disk drives (HDD's) has continued to improve at a rapid pace, I/O performance has not. Compared to CPU performance, relatively slow HDD speed results in an "I/O bottleneck" causing ineffective CPU utilization. Dramatic improvements in multicore CPU processing capabilities compared with minimal improvements in HDD performance, result in poor application performance while CPUs spend time waiting for data. Each generational increase in CPU performance exacerbates this problem.

A typical method used to increase the overall Input / Output Operations per Second (IOPS) performance and capacity of a storage array is to increase the number of disks in the array. Since each drive in the array is capable of reading and writing a given amount of data per second, the more drives that are in the overall array, the more aggregate data that can be written to, or read from the array.

For the best performance, administrators may also "short-stroke" each drive. Short stroking is a method by which data is only stored on the outer edge of the platter of each hard drive reducing the time required to physically reposition the read/write heads during use. While this can certainly improve random I/O performance, it also significantly reduces the available capacity of the drive. Capacity loss of 90 percent is common in short-stroke environments.

There are other problems with these approaches, as well. Even though more drives offer better performance and greater capacity, more drives also mean more complexity and increased management costs, and the potential for failure increases as additional system components are added. In some cases, an increasing number of HDDs can put a strain on power constraints of the data center.

2.0 The Solution

There are other approaches to solving this problem using flash-based technologies and solid-state disks (SSDs) in particular. We will explore three alternative configurations:

1. SSD Storage Arrays
2. SSD Storage Arrays with FastPath Acceleration Software
3. HDD Arrays using CacheCade Pro 2.0 Software for SSD Caching

2.1 SSD Storage Arrays

Flash is a chip technology that provides large amounts of non-volatile memory. With no moving parts that slow access to data, flash greatly reduces latency of input / output operations. Elaborate algorithms have been implemented to allow flash to provide consistent performance, reliability, and

scalability. As an alternative to hard drives, SSDs incorporate flash using the same physical enclosure and interface as hard disk drives, thus easily replacing them in most applications.

SSDs can be used to create a complete storage array. SSDs can enable extreme application performance and response time improvements on the server, providing up to 1000 times more I/O transfers per second than HDDs. Power and cooling requirements for SSDs can also be a fraction of that required for HDDs resulting in significantly improved ROI.

Lenovo offers Intel® Data Center Family SSDs that provide consistently high I/O performance, and has lower and more consistent access times by optimizing the quality of service of the drive. ThinkServer SSDs insure a higher steady state performance with a more consistent range of operation. This consistency improves RAID performance because no one drive in a RAID array is holding up the rest.

The SSDs also keep data secure with 256-bit AES encryption, and an end-to-end data protection scheme delivers important data integrity by protecting against possible corruption of in-transit and stored data through every element of the SSD device. In addition, if an unexpected power loss occurs, the drives protect the data in flight by flushing it to the NAND memory.

THINKSERVER SSD KEY FEATURES

- Fast and consistent performance
- High write endurance technology
- Cost and power efficiencies with leading IOPS per Watt

2.2 SSD Storage Arrays with FastPath Acceleration Software

When using SSD arrays with the ThinkServer RAID 710 adapter card, optional ThinkServer RAID FastPath software provides a high performance I/O accelerator that can improve performance even further by dramatically boosting throughput of multiple SSDs attached to the controller. FastPath is used when SSDs are configured as virtual disk volumes, and used as primary storage for files when the highest possible application performance is required. FastPath software is an optimized version of MegaRAID technology designed to get the best possible performance from a pure SSD array. FastPath is particularly useful when used with applications that demonstrate small, random read/write operation workloads that require high throughput, typical of transactional database applications. In this test, we were able to show that SSD volumes with FastPath enabled resulted in better throughput than SSD volumes without FastPath enabled. FastPath can significantly improve the performance of an SSD array in certain configurations.

THINKSERVER RAID FASTPATH KEY FEATURES

- Up to 150,000 IO reads per second for small, random block-size IO activity
- Up to two times faster than not using FastPath software in transactional database environments

2.3 HDD Arrays using CacheCade Pro 2.0 Software for SSD Caching

Another approach to accelerate the performance of existing HDD arrays, without making substantial investments in new hardware, is to deploy optional ThinkServer RAID CacheCade Pro 2.0 software with the ThinkServer RAID 710 adapter card, utilizing the SSDs to create up to 512GB of high-performance read and write controller cache pools to be used with existing HDD-based arrays.

CacheCade Pro 2.0 intelligently and dynamically manages frequently accessed data by copying it from HDD volumes to a higher performance layer of SSD cache. Directing the most frequently accessed data (“hot spots”) to flash cache relieves the primary HDD array from time-consuming I/O transactions, which allows for more efficient hard disk operation, reduced latency, and accelerated read and write speeds. For enterprise data protection, the same data is retained on RAID-protected HDDs.

THINKSERVER RAID CACHECADE PRO 2.0 KEY FEATURES

- Accelerate performance of existing HDD arrays with minimal investment
- Read and write caching of hot-spot data significantly reduces I/O latency
- Optimized for real-world transaction intensive workloads

CacheCade allows very large data sets to be present in cache, enabling performance improvements of two to 12 times that of HDD-only configurations, and up to 50 times greater than regular cache in many read-intensive server applications.

Server workloads such as web, file, online transaction processing (OLTP) database, and data mining, will benefit. Even write intensive workloads such as Exchange server, high performance computing (HPC) applications, Web 2.0 and other IO-intensive OLTP database system workloads, experience dramatic performance improvements.

3.0 Evaluating the Solutions

To evaluate the performance enhancements of each of these solutions against traditional hard drive implementations, we used a real world OLTP database workload similar to an online trading application for a brokerage firm. The benchmark workload simulates central database transactions related to managing a firm’s customer accounts. The workload generator is configurable to simulate users accessing the database and the rate of transactions per user.

The testing allows us to examine key storage performance factors including database transactions per second, database transaction response time, and storage raw I/O operations per second. We also calculated the relative cost of each solution using a measure of solution price per I/O performance.

The workload was tuned to optimize the performance of the hard drive only configuration so that we could reliably compare the improvements that could be achieved using the alternative solutions. This resulted in a profile that simulated 150 users accessing a 563 GB database installed on the Lenovo ThinkServer. The size of memory available to the database was set to insure that the storage subsystem was exercised. Testing consisted of a standard OLTP type workload of 90% reads with SQL Server 2012. The Windows performance monitor (PerfMon counters) framework was used to gather the performance metrics.

“Appendix A – Test Configurations” describes the server and database configurations for each solution. The storage configurations used in the testing are shown in Table 1.

Table 1 – Storage Configurations

Storage Configurations	Boot Volume	Database Volume	Log Volume	CacheCade Cache Volume
HDD Only	2 HDD / RAID 1	8 HDD / RAID 10	4 HDD / RAID 10	N/A
HDDs w/ CacheCade + SSDs	2 HDD / RAID 1	8 HDD / RAID 10	4 HDD / RAID 10	2 SSDs / RAID 1
SSDs Only	2 SSD / RAID 1	4 SSDs / RAID 10	2 SSDs / RAID 1	N/A
SSDs w/ FastPath	2 SSD / RAID 1	4 SSDs / RAID 10	2 SSDs / RAID 1	N/A

3.1 Database Transaction Performance

In order to verify that the test configurations were accessing the storage system, and not just exercising the SQL Server buffer pool, only 32 GB of DRAM was configured for SQL Server. Two particular monitors in the SQL Server Buffer Manager are most often used to understand stress on the storage system. The first is Lazy Writes per Second. Lazy Writes per Second shows how often the database had to write pages from main memory to storage. As shown in Figure 1, the SSD-based solutions were utilized much better than HDDs. In real world scenarios, SSD-based storage systems are far more capable and can handle daily peak loads, backups, database reorganization, or data conversions when the storage stress will be higher.

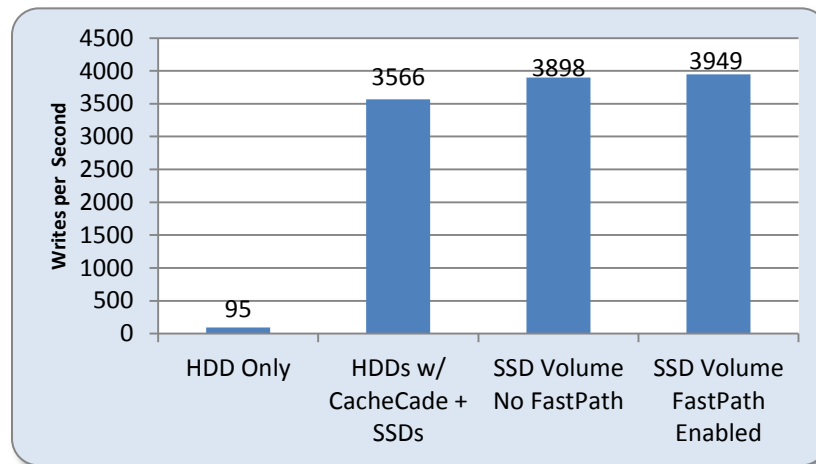


Figure 1 – Lazy Writes per Second

The second monitor, the Average Page Life Expectancy, is a measure of how long a page remains in the buffer pool before being overwritten. Figure 2 shows that the HDD configuration is taking much longer to retrieve pages from the slower storage media. Pages are being overwritten more quickly with all three SSD-based solutions in order to keep up with the demand for SQL server buffer pool memory space limitation during the test.

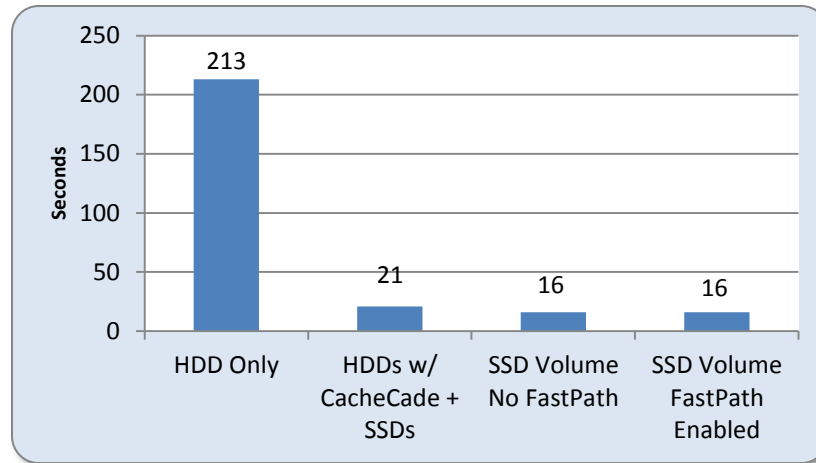


Figure 2 – Average Page Life Expectancy

The baseline performance of the database using the HDD array was 696 database engine transactions per second (tps) as shown in Figure 3. By using CacheCade, database performance increased by a factor of more than 14x to 10,016 tps. An all SSD volume enabled a performance of 10,970 tps, and using SSDs with FastPath improved performance to 11,037 tps.

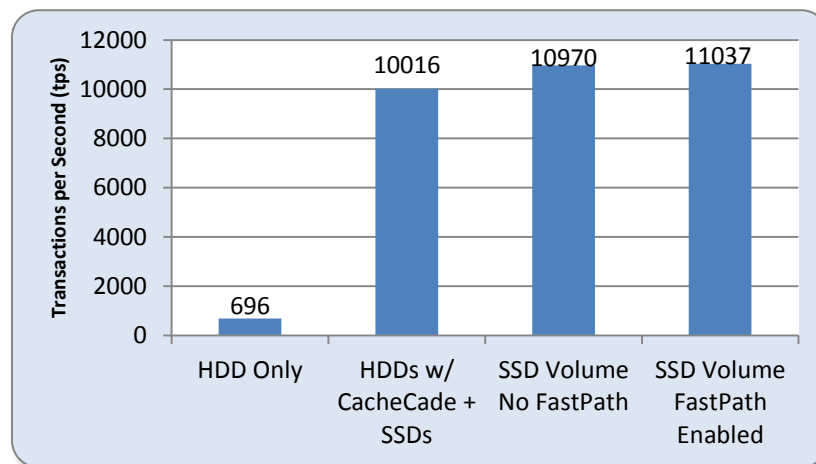


Figure 3 – Database Performance (TPS)

Another key measure of database performance is response times for user requests. The baseline OLTP workload was optimized to take best advantage of the performance limits of the hard drive-only configuration, while keeping the average wait time (SQL Server page IO latch) below 50 milliseconds (ms). SQL Server Page IO Latch Latency is a key metric of database performance and a component used to understand the overall impact on query performance and thus user experience. High SQL Server Page IO Latch Latency can be an indicator of an IO-bound storage subsystem, which is critical for database

administrators in designing servers for database applications. An SQL Server Page IO Latch occurs when a page is being retrieved from storage into the SQL Server buffer pool, and the time this takes is critical to the performance of a query or user interaction. Your goal as a server administrator is to lower the SQL Server Page wait time to as low as possible to also lower user request times when Cache misses happen in the buffer pool of SQL Server.

Figure 4 below shows the average wait time per SQL Server page IO latch. Database response times on the hard drive array are significantly longer than the solutions using SSDs. A goal of one millisecond, as the SSD solutions provide, will result in a highly efficient design.

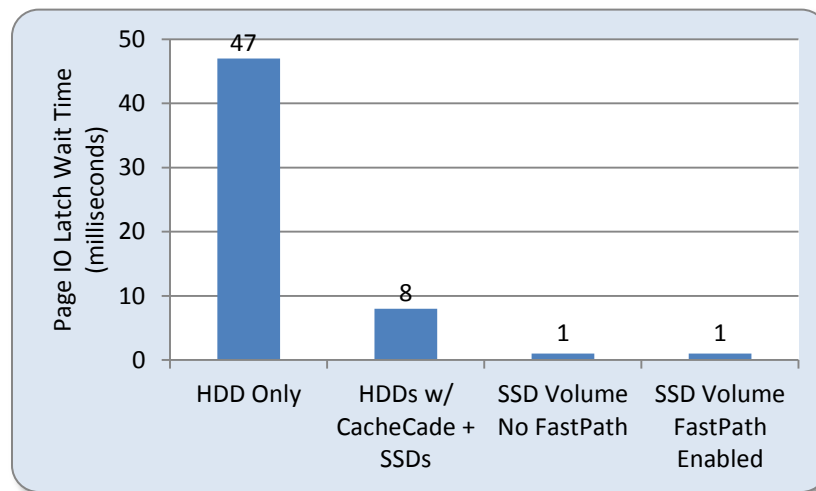


Figure 4 – SQL Server Page IO Latch Latency

Another key measure for database administrators is how efficiently the server is performing with a given database workload. One critical metric is the CPU utilization over the course of the benchmark. With the faster SSD-based storage I/O performance in each scenario, the CPU is not waiting for long requests on the database disk volumes, and a greater number of real database transactions are performed in the same period. This extends the ability of your server, and most importantly the useful database transactions that the server can perform. As shown in Figure 5, the three SSD-based solutions provide efficient use of the CPU between 76 and 92 percent, and the scalability of the transactions per second is more than 10x the HDD Only test.

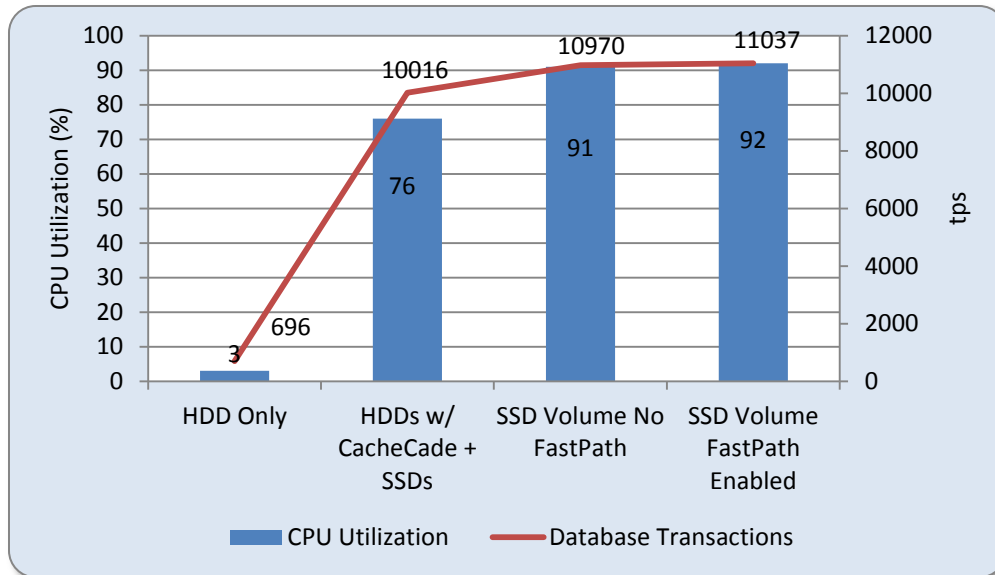


Figure 5 – CPU and Database Productivity

3.2 Raw I/O Performance

The average storage performance of the database during the course of the testing is shown in the Figure 6. The baseline hard drive-only configuration averaged 3,979 raw read IOPS. Using SSDs as cache with CacheCade, performance improved significantly to 45,230 IOPS.

All SSD array performance reached 50,294 IOPS, and increased even further to 50,880 IOPS when utilizing FastPath.

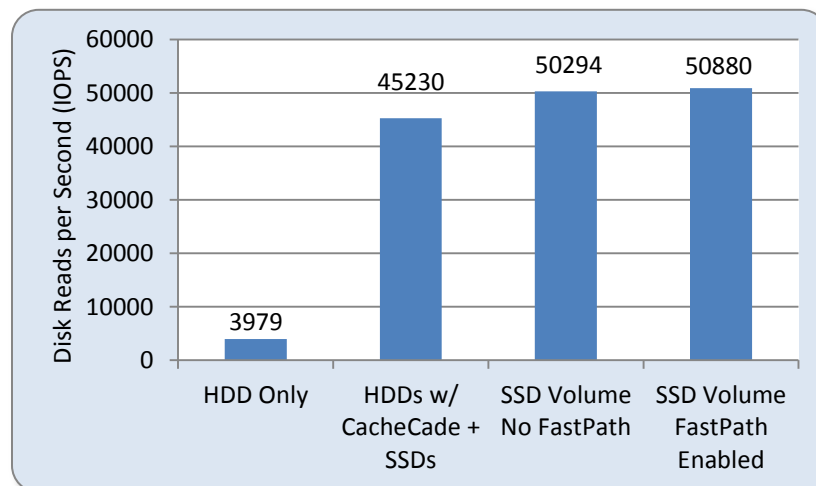


Figure 6 – Disk IO Performance (IOPS)

As shown in Figure 7, note that CacheCade performance scaled as more active data fit into the SSD cache. Before all active (hot-spot) data was moved to cache, performance was basically that of the HDD array, however, the CacheCade performance reached its maximum within a half-hour of beginning the benchmark without any user intervention. Due to database checkpoint activity, some peaks and dips are observed.

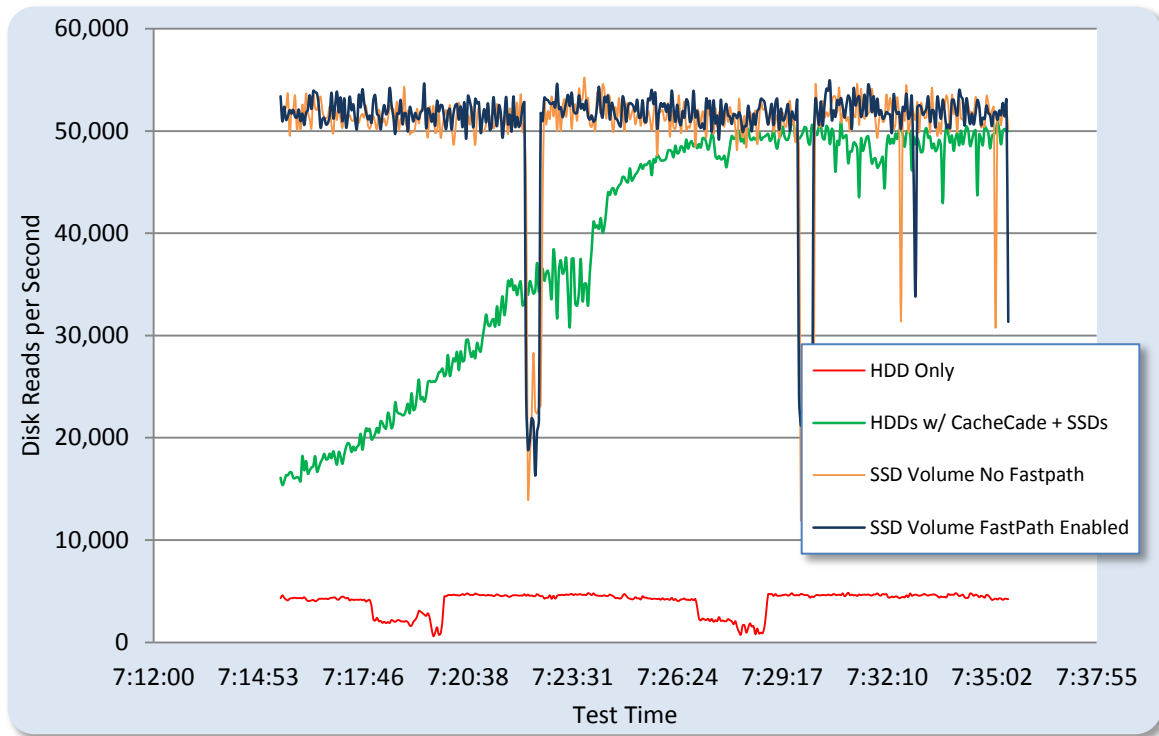


Figure 7 – CacheCade Performance Scales as Cache Fills with Active Data

The tests also showed (see Figure 8) that disk throughput improves as well using each of the alternate solutions.

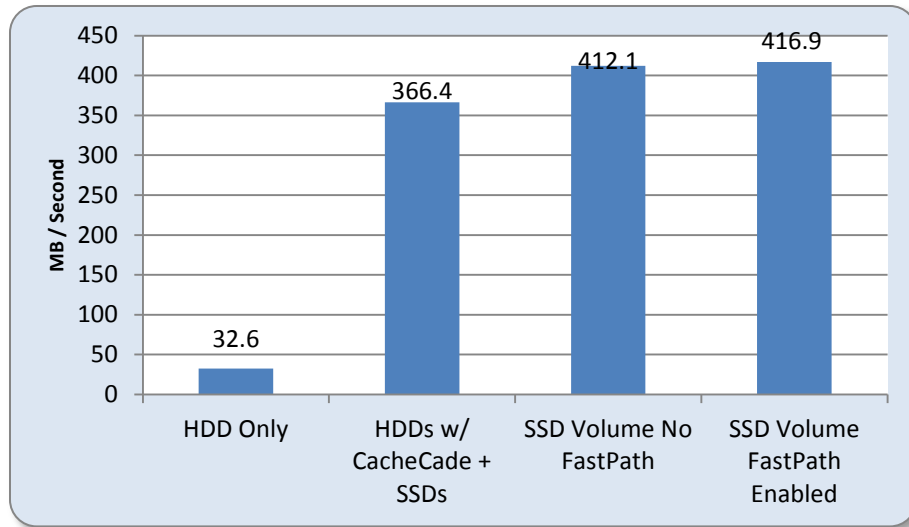


Figure 8 – Average Disk Throughput (MB/s)

3.3 Price Performance

Clearly in terms of overall cost per capacity, the HDD solution wins, however, as an application owner, the primary objective of a solution should be cost for performance. You should focus on I/O building blocks, such as SSDs, that most closely matches the performance of the CPUs you are purchasing. Performance relates significantly to user experience and productivity goals for the server. In Figure 5, we saw that the CPU could not be utilized efficiently with HDDs, thus leaving lots of potential for an underutilized server. In Figure 9, we calculated the cost¹ of each solution measured against the performance gains they achieved. This shows that the SSD Solutions can provide the best returns when performance is a key requirement.

¹ Cost data is as of April 2014, and represents retail list prices for systems and available options combined to create the configurations represented in this paper.

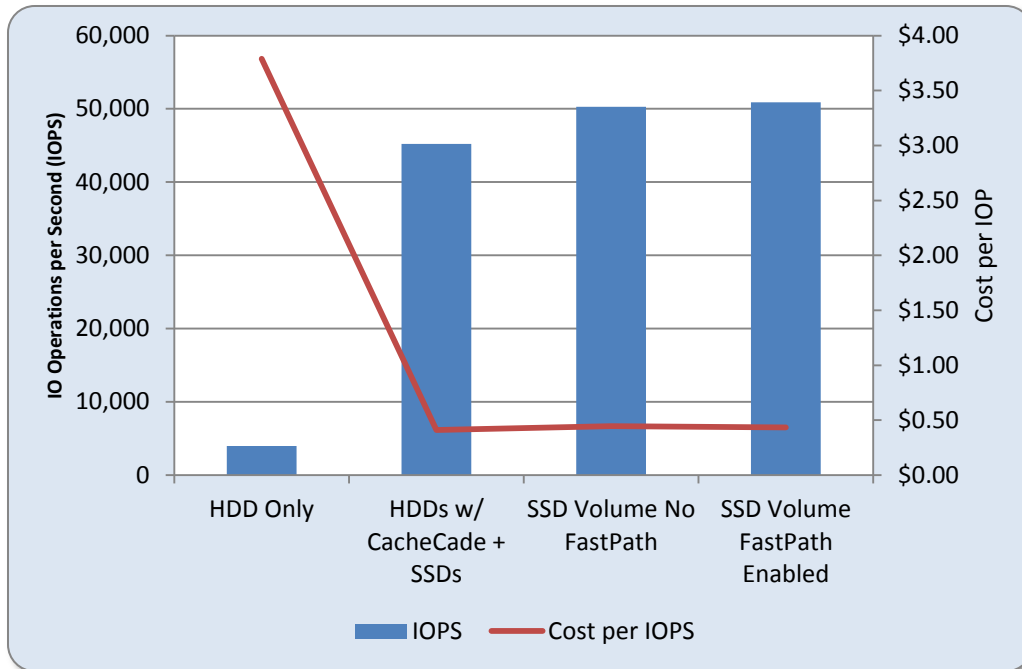


Figure 9 – Cost for Performance

4.0 Conclusions

All three SSD-based solutions achieved dramatic improvements in server storage I/O latency and throughput performance. The price and reliability of SSDs has improved and should no longer be an inhibitor. We have shown that when comparing \$/IOP, SSDs can provide true value when compared to HDDs for the same hardware configuration, allowing you to extend the value of your total server investment. Intel Data Center Family SSDs offered by Lenovo provide the level of endurance, reliability, and data protection that administrators need and expect.

It is important to consider that results presented in this paper are based on testing that resulted in more predictable cached data than may be realistic in a production database. In general, results with all caching software will vary depending on the uniqueness of the workload and the resulting number of cache hits the CacheCade volume can service. New and unique database page requests that must go to hard disk will obviously suffer much higher latencies than requests that can be served out of the SSD Cache.

If the highest levels of performance, long-term reliability, and efficiency of the server are required, applications built entirely on SSD storage will achieve these goals. We have also shown that database transaction performance of when using SSDs can be further enhanced by leveraging FastPath.

A hybrid approach using CacheCade can help balance cost and performance where SSDs are used to cache frequently accessed data, but HDDs are used for primary data storage at a lower cost. CacheCade can minimize investments in SSD technology compared to potentially significant investments in

additional hard drives required to attain comparable performance gains achieved by implementing CacheCade, as well as achieving acceptable performance gains compared to all SSD arrays.

We have also shown that there are additional benefits when making IO more efficient in that CPU utilization is improved and more workload can be performed since the application is not waiting for IO operations to complete.

A comparison of the solutions is summarized in Table 2.

Table 2 – Solution Comparison

Solution	Targeted Workload	SSD Capacity Needed	Volumes Accelerated	Performance Gain	Cost of Capacity	Cost of Ownership
Add more short-stroked HDDs	All workloads	None	One HDD volume	Lowest	\$	\$\$\$\$
HDDs w/ CacheCade + SSDs	Random read and write intensive	Hot data only	One or more HDD volumes	Moderate	\$\$	\$\$
SSD array	All workloads, especially random	All volume data	One or more SSD volumes	High	\$\$\$	\$
SSD array + FastPath	All workloads, especially random	All volume data	One or more SSD volumes	Highest	\$\$\$	\$

Also when choosing a solution, consider that application workloads with small and random I/O patterns requiring high transactional throughput, such as OLTP, will benefit most from SSD acceleration techniques.

Hybrid arrays with CacheCade will be applicable with these types of I/O profiles, but CacheCade is optimized small block random intensive applications, and is best suited when the majority of the working data set can be contained in the SSD cache. Applications with much larger active data sets that will not all fit in cache, or more sequential write intensive I/O profiles show limited improvement with CacheCade software, and performance improvements will be greater with all SSD arrays and FastPath software.

Table 3 provides guidance for selecting an acceleration solution based on server workload.

Table 3 – Workload Qualifiers for CacheCade, and FastPath

Application Workloads that benefit from SSDs	Recommended Software Solution
Web servers and other transactional, small read intensive applications	CacheCade
File server (including SharePoint)	CacheCade or FastPath
E-mail server	CacheCade or FastPath
Databases	CacheCade or FastPath
OLTP	CacheCade or FastPath
E-commerce	CacheCade or FastPath
Large sequential databases	FastPath
Streaming read applications	FastPath
Streaming write applications	FastPath

5.0 Appendix A – Test Configurations

Our tests were run using ThinkServer RD640s. Table 4 shows the hardware configuration of the server used for the all HDD storage solution, as well as the hybrid SSD/HDD with CacheCade solution. Table 5 shows the hardware configuration of the server used for all SSD storage solution.

Table 4 – Server Configuration for All HDD and Hybrid Arrays

Components	Description
Server	ThinkServer RD640
Drive Bays	16x 2.5"
CPU	Dual Intel Xeon E5-2650 v2; 2.6 GHz; 20MB cache; 8.0GT; 8 core CPU
Memory	32 GB total with 4x 8GB ECC PC3 1600 MHz (2Rx8) RDIMMs
Operating System	Microsoft Windows Server 2012 Standard (6.2.9200 Build 9200)
Database	Microsoft SQL Server 2012 11.0.3128.0 Enterprise Edition
RAID Controller	ThinkServer RAID 710 Firmware Version 3.240.85-2745; Driver Version 5.2.122.0
RAID Controller Option	ThinkServer RAID CacheCade Pro 2.0 Software Key

Table 5 – Server Configuration for All SSD Arrays

Components	Description
Server	ThinkServer RD640
Drive Bays	8x 2.5"
CPU	Dual Intel Xeon E5-2650 v2; 2.6 GHz; 20MB cache; 8.0GT; 8 core CPU
Memory	32 GB total with 4x 8GB ECC PC3 1600 MHz (2Rx8) RDIMMs
Operating System	Microsoft Windows Server 2012 Standard (6.2.9200 Build 9200)
Database	Microsoft SQL Server 2012 11.0.3128.0 Enterprise Edition
RAID Controller	ThinkServer RAID 710 Firmware Version 3.240.85-2745; Driver Version 5.2.122.0
RAID Controller Option	ThinkServer RAID FastPath Software Key

The tests used a 563 GB Microsoft SQL Server 2012 Enterprise Edition SP1 (11.0.3128.0) database stored on a dedicated volume ("Database Volume"). The size of the database was chosen to insure that the entire database would not fit in system memory insuring the storage subsystem was being exercised. A 20 GB Log file was stored on a separate volume ("Log Volume"). The workload was random for the data volume and sequential for the log file volume. The database volume was configured as RAID 10 in order to maximize the performance of the all HDD array baseline tests. The log file volume was configured as RAID 1. The storage configurations are summarized in Table 6. Table 7 summarizes the volume policies used.

The OLTP benchmark simulated 150 users accessing the database. The workload pattern for the tests was typically 8 KB reads and writes, with a read/write ratio of 9 to 1.

The SSDs were pre-conditioned to insure that their performance had reached steady-state.

Table 6 – Storage Configurations

	Boot Volume	Database Volume	Log Volume	CacheCade Cache Volume
All HDD Arrays				
Drive Type	300GB 15k rpm, 6Gb/s SAS HDD			N/A
# of Drives	2	8	4	
RAID Configuration	RAID 1	RAID 10	RAID 10	
Volume Size (GB)	136GB	544GB	272GB	
HDD Arrays using CacheCade Pro 2.0 Software for SSD Caching				
Drive Type	300GB 15k rpm, 6Gb/s SAS HDD			ThinkServer 600GB Value Read-Optimized 6Gb/s SATA SSD
# of Drives	2	8	4	2
RAID Configuration	RAID 1	RAID 10	RAID 10	RAID 1
Volume Size (GB)	136GB	544GB	272GB	512GB ²
All SSD Arrays (including FastPath)				
Drive Type	ThinkServer 600GB Value Read-Optimized 6Gb/s SATA SSD			N/A
# of Drives	2	4	2	
RAID Configuration	RAID 1	RAID 10	RAID 1	
Volume Size (GB)	136GB	600GB	600GB	

Table 7 – Volume Policy Settings

Volume Policy	All HDD Array	HDD Array using CacheCade	All SSD Array	All SSD Array with FastPath
Stripe Size	256KB	256KB	256KB	256KB
Disk Cache Policy	Default	Default	Default	Default
Read Policy	Always Read Ahead	Always Read Ahead	Always Read Ahead	Always Read Ahead
IO Policy	Direct IO	Direct IO	Direct IO	Direct IO
Current Write Policy	Write Back	Write Back	Write Back	Write Back
Current Access Policy	Read Write	Read Write	Read Write	Read Write

² Maximum allowable CacheCade capacity

6.0 Appendix B – References

This section provides additional information that can be helpful in understanding the nature of the tests performed in this paper, and in particular, understanding SQL Buffer Management.

Understanding SQL Buffer Management

[http://technet.microsoft.com/en-us/library/aa337525\(v=SQL.105\).aspx](http://technet.microsoft.com/en-us/library/aa337525(v=SQL.105).aspx)

Understanding SQL Buffer Manager Monitors

<http://technet.microsoft.com/en-us/library/ms189628.aspx>

SQL Server's key Performance Counter

<http://www.databasejournal.com/features/mssql/article.php/3932406/Top-10-SQL-Server-Counters-for-Monitoring-SQL-Server-Performance.htm>