Seagate Enterprise SATA SSD with DuraWrite™ Technology Competitive Evaluation

Seagate Enterprise SATA SSDs with DuraWrite Technology have the best performance for compressible Database, Cloud, VDI Software Build, OLTP Database and Microsoft Exchange application workloads.

Executive Summary

As the adoption of flash storage in data centers increases so does the pressure to address the common issue of Write Amplification (WA) which can limit I/O performance. WA, or the ratio of physical media writes to writes requested by the host, increases as free flash memory pages decrease and garbage collection becomes necessary in order to free up additional flash memory pages to perform new writes.

The key to decreasing WA is increasing the unwritten flash memory pages available. Vendors overprovision SSDs in order to increase the free flash memory pages available, thus decreasing WA.

To further increase free flash memory pages, decrease WA, and ultimately increase performance, Seagate has introduced DuraWrite technology. Seagate Enterprise SATA SSDs with DuraWrite Technology take advantage of low data entropy, compressing the data before writing to the SSD. This conserves free flash memory pages and ultimately leads to improved performance.

Seagate commissioned Demartek to evaluate Seagate Enterprise SATA SSDs with DuraWrite Technology (hereafter "Seagate Enterprise SATA SSDs"), comparing their performance and WA with competitive SSDs.

A variety of real-world datasets with varying data entropies were used to test the performance of Seagate, Intel, and Micron enterprise SATA SSDs. With Seagate Enterprise SATA SSDs, we found that all dataset types had performance gains, and the lowest data entropy datasets with the highest workload write percentages had the greatest performance gains.

Key Findings

- > The Seagate Enterprise SATA SSDs achieved 47% higher operations per second for the cloud workload.
- > The Seagate Enterprise SATA SSDs had on average one third less latency when performing the cloud application workload.
- > The Seagate Enterprise SATA SSDs had approximately one third the WA of competitor SSDs when performing the VDI workload.
- > The Seagate Enterprise SATA SSDs had approximately 36% of the latency of the closest competitor and one ninth the WA when performing the Software Build workload.
- > The Seagate Enterprise SATA SSDs achieved 55% of the WA for the SQL OLTP workload while achieving slightly higher database operations per second.
- > The Seagate Enterprise SATA SSDs achieved 55% of the WA for the Exchange Jetstress SQL OLTP workload while achieving 11% higher transactional IOPS.

Seagate Enterprise SATA SSD with DuraWrite™ Technology Competitive Evaluation

Demartek
September 2018

Seagate Enterprise SATA SSDs

Most real-world datasets have low data entropy making them highly compressible. In this document, we will use a simplified definition of data entropy:

$$\textbf{Data Entropy} = \frac{\text{Compressed Size}}{\text{Uncompressed Size}}$$

In previous testing Demartek found that relational databases had data entropies of as little as 10 to 20%. Plaintext and webpages had data entropies of 22-23%. Report files such as csv had data entropies of approximately 11%. These are just a few examples. A more complete study is available on the Demartek website. Storage array vendors have reported average overall data entropies of approximately 20%.

Seagate DuraWrite technology takes advantage of this redundancy in typical data by performing compression before writing to flash. When the data takes up less space, there are more free flash memory pages available. When more free pages are available, the read-erase-modify-write that is necessary to update a flash storage page by making a copy is avoided and instead a simple page erase-write can occur. With more free pages available, there are more writes that are simple page writes to unwritten pages. This reduces WA which also reduces latency, improves throughput, and improves power efficiency. It also allows Seagate to enhance the media usage warranty.

Seagate DuraWrite technology features unique and loss-less data reduction technology designed to:

- Improve performance
- Deliver high power efficiency
- Save cost
- Provide enhanced endurance SSDs

Application Workloads

Application workloads typical of most datacenters were chosen for inclusion in our testing: Cloud, VDI, Software Build, OLTP Database, and Microsoft Exchange. The datasets used by these workloads are similar to most active data in production use cases.

Application Workload Summary

	Test	Operating System
Cloud	Yahoo Cloud Servicing Benchmark (YCSB)	CentOS 7.4
VDI	SPECSFS 2014 SP2	CentOS 7.4
Software Build	SPECSFS 2014 SP2	CentOS 7.4
SQL Server	OLTP	Windows Server 2016
Microsoft Exchange	Jetstress 2013	Windows Server 2016

Yahoo Cloud Serving Benchmark (YCSB)

The Yahoo Cloud Serving Benchmark (YCSB) is a framework that provides a common set of workloads typically found in cloud datacenters.

YCSB includes six workloads that cover many of the workloads found in cloud datacenters. These are:

- Workload A: Update heavy (50% read, 50% write)
- Workload B: Read mostly (95% read, 5% write)
- Workload C: Read only (100% read)
- Workload D: Read latest (new records inserted and then read)
- Workload E: Short ranges (ranges of reads, such as email threads)
- Workload F: Read-modify-write

We used MongoDB version 3.6.5-1.el7 with YCSB 0.5.0 Workload A for our testing. In order to more easily generate datasets of the proper size, the ***fieldlength*** parameter was changed to 10,000, the ***fieldcount*** parameter was changed to 100, and the record count was adjusted. An operation count of 1,000,000 operations was requested. A script was created with a loop to repeatedly execute the same YCSB test until steady state was achieved.

VDI Workload

The VDI benchmark simulates a hypervisor with a set of cloned desktop VMs designed for office-work and was observed to have a read:write application ratio of approximately 1.7 to 1.

SPECSFS 2014 SP2 was used to generate the VDI workload. A typical SPECSFS run collects performance data for a series of different business metrics. The business metrics determine the size of the dataset used as well as the throughput rate the test suite requests from the storage. Single business metric runs were performed as we targeted specific dataset sizes to match our capacity utilization points. This workload was chosen for its filesystem I/O behavior, but not for an official SPECSFS submittal.

Software Build Workload

The software build benchmark is a file system benchmark that mimics the behavior of software builds and traces. 70% of the operations performed by the benchmark are a stat call returning file attributes, or metadata, about the files. As such, the software build workload has a high amount of small block reads with interspersed file I/O. The workload is approximately 80% read, highly random, and small block. SPECSFS 2014 SP2 was also used to generate the Software Build Workload.

Microsoft SQL Server On-Line Transaction Processing (OLTP) Database Workload

Microsoft SQL Server 2017 was deployed on Windows Server 2016 Servers. OLTP databases were generated and an OLTP workload was executed against the databases on our SSDs under test.

The OLTP workload is a variation of an OLTP workload that models a brokerage firm with customers who generate transactions related to trades, account inquiries, and market research. This workload is approximately 90% read and 10% write (not counting the log files that are 100% write). The brokerage firm in turn interacts with financial markets to execute orders on behalf of the customers and updates relevant account information.

This workload emphasizes IOPS (smaller block sizes) and has low, relatively steady latency.

Three partitions were created on the SSDs under test. One for the OLTP database, one for the OLTP log, and one for the tempdb database and log.

Microsoft Exchange JetStress 2013

Microsoft Exchange is one of the dominant email server applications in use today. Many small and medium sized businesses that do not offload their email to third parties run their own Exchange servers. Email is a latency sensitive application (just ask anyone waiting for an important email), and therefore properly designing and scaling storage is critical to an acceptable user experience.

Microsoft Exchange Jetstress 2013 models the I/O loading and patterns that would be generated by an Exchange 2013 server and Exchange 2016 server, including email database reads and writes as well as logging. Its configuration includes the number and size of mailboxes, number of mailbox databases and logs, and intensity of use, among other things. Jetstress uses the same Extensible Storage Engine files used by Microsoft Exchange to ensure that the Jetstress modelled performance is representative of a bona fide Exchange Server of the same version. Jetstress marks a test a failure if any Exchange database experiences an average transactional latency of 20 milliseconds or higher or fails to achieve the desired IOPS.

Microsoft recommends Exchange Jetstress be used to validate email systems before putting those systems into production.

Mailboxes of size 2GB were used in our testing, with the number of mailboxes in our database changed in order to take up different amounts of space. A standard

Jetstress configuration of 50% read and 50% write was performed. 1 database was configured for each drive. Threadcount was adjusted to 1 and IOPS adjusted to 5, simulating a user environment where approximately 4000 messages were received per user per day. (A high traffic environment was chosen to better stress the storage, however a more typical environment should also show performance improvements with Seagate Durawrite.)

Test Preconditioning and Procedures

Real World Preconditioning

Typical NAND flash solid-state devices or drives (SSDs) arrive from the factory in a condition known as “fresh out of the box” or FOB. This is a condition in which all the NAND flash blocks are available to be written without the usual read-erase-modify-write cycle that is required for blocks that already contain user data. During this initial FOB condition, the drive performs at higher than normal performance levels. This performance level will drop to a level known as “steady state performance” as user data is written to the NAND flash blocks and the process of “garbage collection” begins.

In order to achieve steady state performance, workloads need to be written repeatedly until steady state performance is achieved. SSDs in normal service typically maintain steady state performance for the usable life of the drive in normal operating conditions.

The process of deliberately forcing an SSD from FOB condition to steady state condition is known as pre-conditioning. Pre-conditioning occurs when specific workloads are applied to the SSD and the performance is measured with the goal of finding the steady state performance. Pre-conditioning tests are typically configured to read and write specific block sizes with either random or sequential I/O addressing patterns addressing the entire logical block address ranges. These are typically performed by synthetic I/O workload tools so that the various I/O parameters can be controlled.

Real-world application workloads behave differently than synthetic I/O workload generators. First, real-world applications perform a mixture of CPU, memory, network and storage operations, depending on the nature of the application, in order to achieve the application’s function. These typically make variable use of many of the system resources and are often considered more unpredictable in the use of these resources than synthetic I/O workload generating tools.

Secondly, a real-world application may run for a period of time with one usage pattern of system resources then shift to a different usage mixture of these same resources once a certain portion of work is completed. For example, real-world database transactional applications often issue higher I/O block size read requests, such as 64KB or higher, in order to fill the memory cache, then once the cache is full, reduce the I/O block size requests to a more typical block size, such as 8KB for transactional workloads. In another example, the same underlying database application, if running a data warehousing workload, may issue very large block requests, such as 512KB block size, while it is gathering data, then issue very few I/O requests while it performs analysis of that data.

A third factor to consider is that not all real-world applications issue block I/O commands. Many real-world applications run on top of a file system and file systems issue I/O requests differently, in order to maintain file system integrity while servicing the needs of the application. Typically, the application running on top of a file system does not know, or cannot predict, the logical block address issued by the file system for any given amount of data to be stored or read from storage. In addition, file systems may or may not use the same logical block address range when rewriting all of the contents of a file.

In light of these factors, we ran each real-world application multiple times until steady-state performance was achieved. However, the exact I/O sequence used to achieve steady-state performance was different in every case.

Test Procedure

For each of the application workload tests, we sized the application databases relative to the size of the SSD under test to the workload and allowed that workload to fill that space. We allocated a reasonable amount of space to the operating system and file system for typical usage, and left some of the capacity unused. We followed this same pattern with each brand of drive.

The application workloads were run repeatedly, or long enough, until steady state was achieved. The results shown are those achieved after steady state was reached.

TRIM was not enabled for the Linux tests (default setting for EXT4) and was enabled for the Windows SQL Server tests (default setting).

Test Results

In the case of multiple test runs, SSDs were interrogated for their SMART attributes at the beginning and end of each run. In the case of one continuous test run, SSDs were interrogated for their SMART attributes every 5 minutes during the test run. The difference observed in SMART attributes between collection times was used to determine the WA present during each run or 5-minute test segment of run.

In the case of multiple test runs, the final run after steady state had been achieved is used for test result comparisons. In the case of one continuous test run, a 5-minute SMART data interval that occurred after steady state was achieved for all brands was selected for comparison.

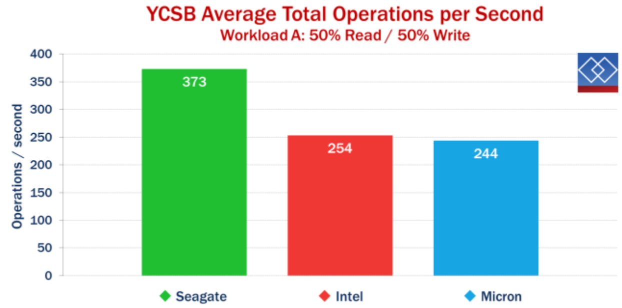
Cloud Workload Application: YCSB Workload A

YCSB results include comparisons of total operations per second, time to complete the workload, latency and write amplification.

When compressed into a tgz file, the YCSB dataset had a data entropy of 79%, which was relatively high.

Total Operations per Second

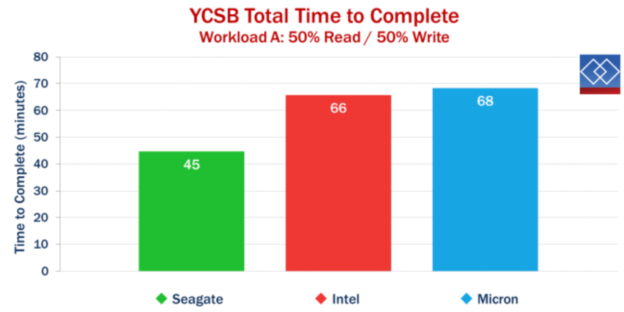
The cloud database operations per second of the Seagate Enterprise SATA SSD were on average 47% higher than those of competitor SSDs.



Total Time to Complete

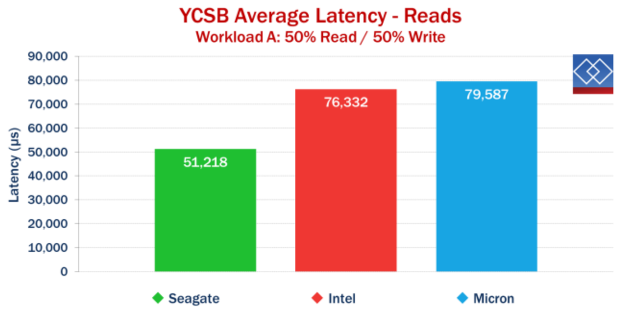
The higher operation rates resulted in faster completion times for the 1,000,000 operation YCSB benchmark.

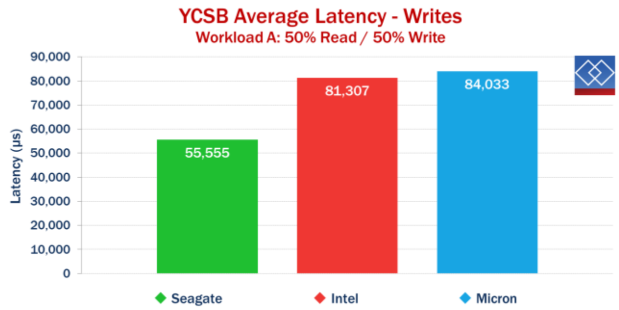
The Seagate Enterprise SATA SSD took one third less time to complete the transactions.



Average Latency

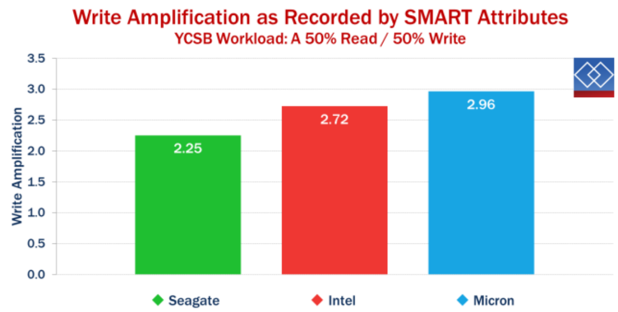
The Seagate Enterprise SATA SSD took on average one third less time to complete cloud operations requests for both reads and writes (updates).





<u>Write Amplification</u>

The WA for the Seagate Enterprise SATA SSD was noticeably lower than the competitive drives. This showed that while the cloud dataset had high entropy, there was still significant WA savings. A high entropy dataset can still give large performance benefits with DuraWrite technology.



The lower WA conserves power and avoids excess writes.

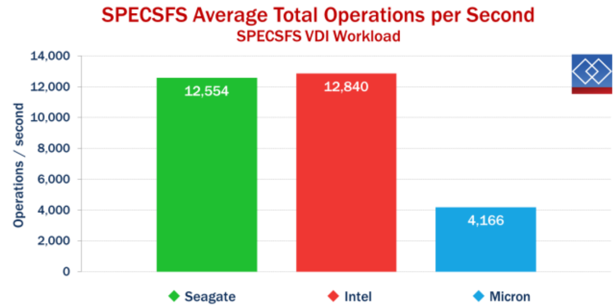
Fileserver Workload Application: SPECSFS 2014 SP2 VDI Workload

SPECSFS 2014 VDI workload results include comparisons of total operations per second, latency and write amplification.

When compressed into a tgz file, the VDI dataset had a data entropy of approximately 37%, which was approximately half the data entropy of the cloud dataset.

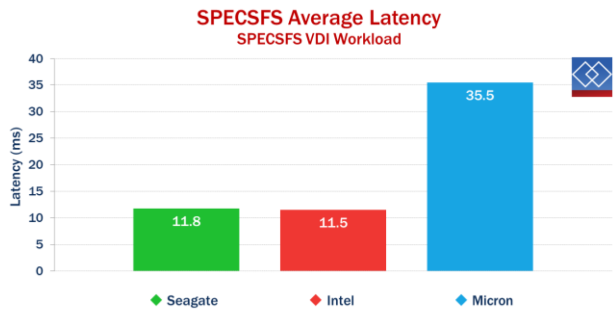
<u>Total Operations per Second</u>

SPECSFS seeks a specific rate of operations per second. For this test, the operations rate of 14,800 operations/second was requested of all SSDs. No SSD was able to keep up with the demand. The Seagate Enterprise SATA SSD and Intel SSD were both able to achieve similar operations throughput rates of 12,554 and 12,840, respectively, when pushed as hard as possible.



<u>Average Latency</u>

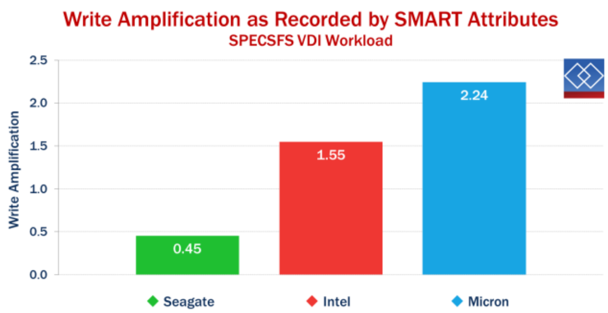
The Seagate Enterprise SATA SSD and the Intel SSD had similar application latencies for this test. The Micron SSD was already pushed well past the optimal operations point.



Seagate Enterprise SATA SSD with DuraWrite™ Technology Competitive Evaluation

Write Amplification

This workload was less write-intensive, and therefore suffered less from WA than the cloud workload. In addition, the dataset had low data entropy, heightening the performance and WA savings from the Seagate DuraWrite technology. The Seagate Enterprise SATA SSD had approximately one third the WA of competitor SSDs.



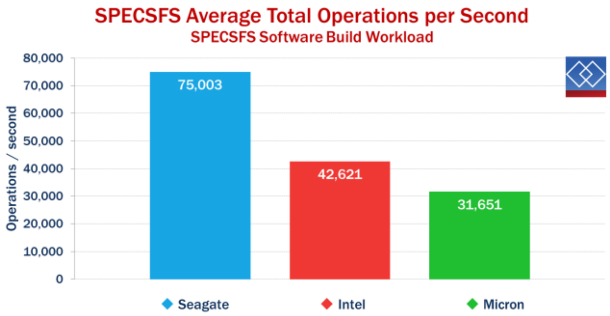
Fileserver Workload Application: SPECSFS 2014 SP2 SWBUILD Workload

SPECSFS 2014 SWBUILD workload results include comparisons of total operations per second, latency and write amplification.

When compressed into a tgz file, the SWBUILD dataset had a data entropy of 17%, which was approximately half the data entropy of the VDI dataset.

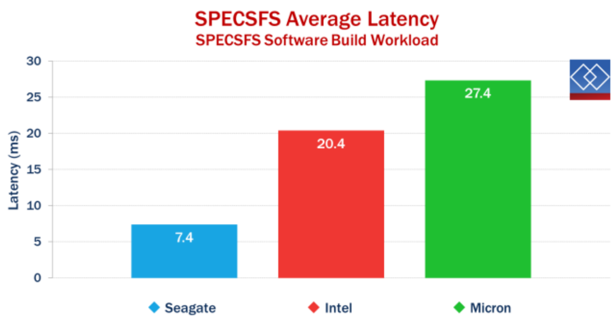
Total Operations per Second

SPECSFS seeks a specific rate of operations per second. An operations rate of 86,500 operations/second was requested of all SSDs. No SSD was able to keep up with the demand. Competitor SSDs were so far past the optimal performance point that Seagate Enterprise SATA SSDs had close to double the performance of their closest competitor.



Average Latency

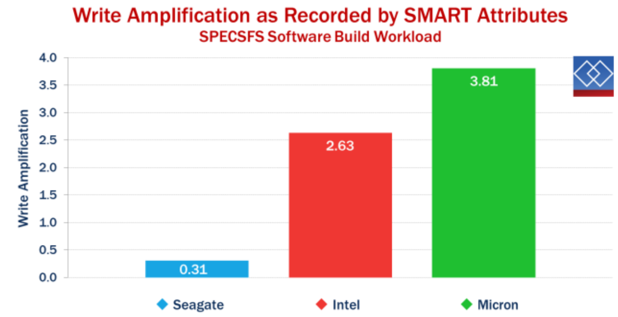
For this test, we were pushing the SSDs to perform at a level of which they are not capable, causing performance degradation. However, the Seagate Enterprise SATA SSD still had half the latency of the closest competitor.



Write Amplification

This dataset had very low data entropy, and had a large number of small metadata updates, which heightened the performance and WA savings from the Seagate DuraWrite technology. For our tests, the Seagate Enterprise SATA SSD WA was approximately one ninth of the Intel SSD. The WA reduction and performance benefits are more evident with this extremely low data entropy dataset.

Seagate Enterprise SATA SSD with DuraWrite™ Technology Competitive Evaluation



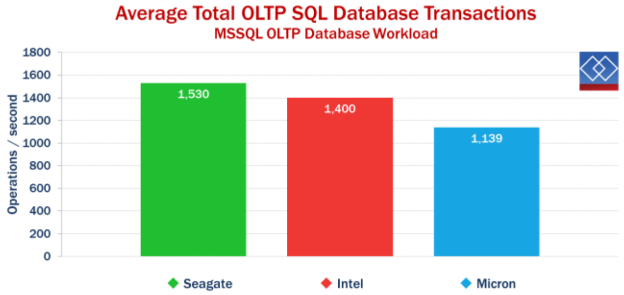
Database Workload Application: Microsoft SQL Server On-Line Transaction Processing (OLTP) Database

SQL OLTP results include comparisons of total operations per second, latency and write amplification.

When compressed into a zip file, the OLTP dataset had a data entropy of 19%, which is a little higher than that of the Software Build Workload, yet still fairly low.

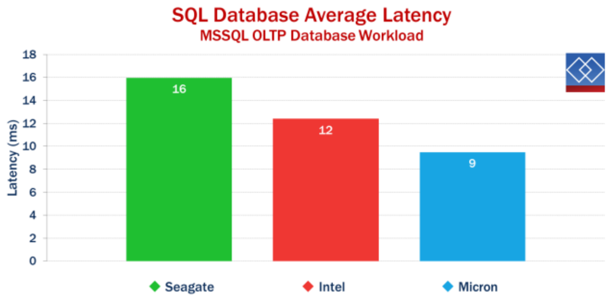
Total OLTP SQL Transactions per Second

For our test case, all drives are struggling to keep up, yet Seagate Enterprise SATA SSDs still have 9% more SQL Transactions per second than the closest competitor.



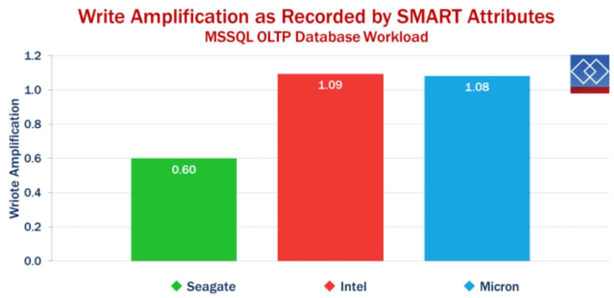
Average Latency

For these tests, the workload was overtaxing all of the drives, which increased the latencies across the board.



Write Amplification

The WA for the Seagate drive was approximately 55% of the WA of the other drives, contributing to the higher performance.



Microsoft Exchange Application: Microsoft Exchange Jetstress 2013

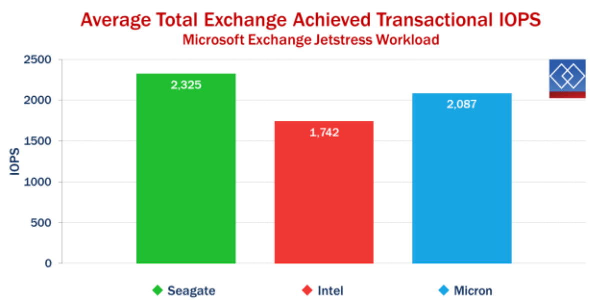
Exchange Jetstress results include comparisons of total transactional IOPS, latency and write amplification.

When compressed into a zip file, the Jetstress dataset had a data entropy of 1.26%, which is the lowest seen for any dataset. Most likely the low data entropy has to do with mailboxes or individual e-mails in the Jetstress generated Exchange databases being very similar or identical. For comparison, a backup copy of a production Exchange Server database was zipped and found to have a data entropy of 44%. As such, the Jetstress results should be considered as the low data entropy bookend to our data entropy spectrum.

Seagate Enterprise SATA SSD with DuraWrite™ Technology Competitive Evaluation

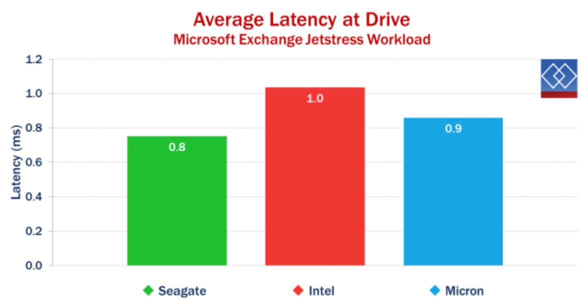
<u>Total OLTP SQL Transactions per Second</u>

For our test case, Jetstress had a target IOPS of 2175. Intel was unable to achieve this rate and failed the test. Seagate and Micron were both able to pass the test when given a 5% allowable IOPS margin, but Seagate Enterprise SATA SSDs still had 11% more IOPS than the closest competitor.



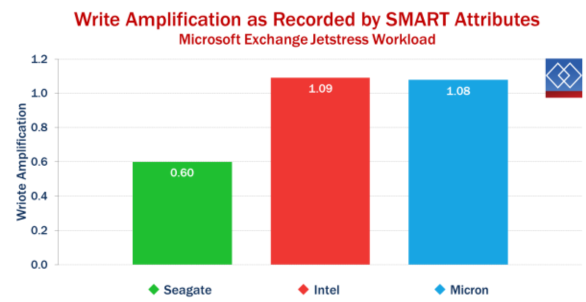
<u>Average Latency</u>

The Intel SSD had the hardest time keeping up, and this is shown in the higher latency.



<u>Write Amplification</u>

The WA for the Seagate drive was again approximately 55% of the WA of the other drives, contributing to the higher performance.



Seagate Enterprise SATA SSD with DuraWrite™ Technology Competitive Evaluation

Summary and Conclusion

The Seagate Enterprise SATA SSD utilizes data redundancy to increase the free flash memory pages available on SATA SSDs. Even for the highest data entropy dataset free flash memory page savings goes a long way to increase throughput, decrease latency, and decrease WA. The lower the data entropy, the greater the performance gains with Seagate Enterprise SATA SSDs. Working our way downwards from the highest data entropy to lowest data entropy workload datasets:

> With the YCSB workload data entropy of 79%, Seagate Enterprise SATA SSD latency was 33% lower and application operations per second was 47% higher. This high entropy workload was able to benefit so much due to the high percentage of writes.

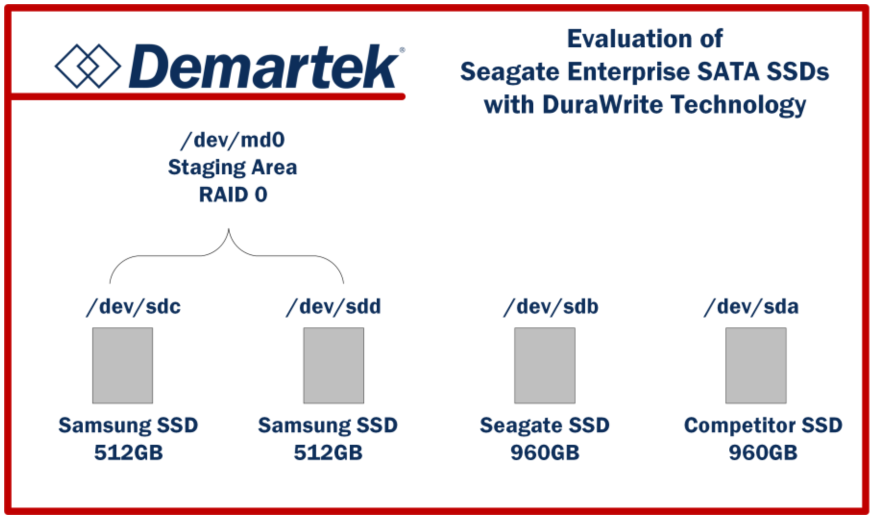
> For the VDI workload at a data entropy of 37%, Seagate Enterprise SATA SSD had one third the WA of competitors.

> For the OLTP Database workload at a data entropy of 19%, Seagate Enterprise SATA SSDs achieved slightly more SQL Transactions per second, with 55% of the WA, even though this workload was only 10% write.

> For the Software Build workload at a data entropy of 17%, Seagate Enterprise SATA SSDs had approximately 36% the latency of the closest competitor and one ninth the WA.

> For the Exchange Jetstress workload at a data entropy of 1.26%, Seagate Enterprise SATA SSDs had achieved 11% more transactional IOPS than their closest competitor with 55% of the WA.

Appendix 1: Test Configuration



Server

Two identical servers were used for SSD testing:

- 2x Supermicro X9DRE-TF+
- 2x Intel Xeon E5-2690 v2 @ 3.0 GHz, 40 total cores, 80 total threads, 256GB RAM
- CentOS Linux 7.4.1708 (Core), kernel 3.10.0-693.21.1.el7.x86_64

Storage Tested (6 Gb/s SATA, 2.5-inch SFF)

- Seagate Nytro 1351, 960GB
- Intel SSD DC S4500 Series, 960GB
- Micron 5200 ECO, 960GB

The most current version of this report is available at https://www.demartek.com/Seagate-DuraWrite/ on the Demartek website.