Evaluation of PMC-Sierra Flashtec NVMe2032 Controller vs. Eight (8) SATA SSDs in Database Workload

PMC-Sierra's NVMe Flashtec NVMe2032 controller outperforms eight SATA SSDs.

Executive Summary

Flash storage is increasingly deployed in the enterprise. For many applications, local "server-side" flash storage is an appropriate way to deploy this very fast storage solution.

There are several types of server-side flash, and PMC-Sierra commissioned Demartek to perform a comparison of the PMC-Sierra Flashtec NVMe2032 PCIe SSD with eight high-performance SATA SSDs in a real-world workload environment. The PMC-Sierra Flashtec NVMe2032 controller can be deployed in a PCIe card form factor or a drive form factor.

Key Findings

The PMC-Sierra Flashtec NVMe2032 solution outperformed a configuration of eight and four high-performance SATA SSDs in every measure of application and storage performance in the database OLTP workload.

- The average Read throughput for the Flashtec NVMe2032 solution was 2.5 times higher than the eight SATA SSD solution and 4.3 times higher than the four SATA SSD solution.
- The average Read IOPS for the Flashtec NVMe2032 solution was 23% higher than the eight SATA SSD solution and 68% higher than the four SATA SSD solution.
- The average latency of the Flashtec NVMe2032 solution was approximately 2/3 the latency of the eight SATA SSD solution and approximately 1/3 the latency of the four SATA SSD solution.

NVMe and SATA

NVM Express, or NVMe, is a new type of streamlined storage protocol specifically designed for storage based on non-volatile memory media, including NAND Flash and other non-volatile memory technologies that are expected to emerge in the future. NVMe currently uses the PCIe bus as the physical interface for this type of storage. As a result of a streamlined storage stack running on a very fast bus, performance is generally higher and latency generally lower for NVMe devices when compared to traditional SAS and SATA interfaces. Performance is limited by the speed and number of lanes of a PCIe bus that are used. NVMe can be used for solid-state storage on PCIe cards and on drive form-factor solid-state storage. NVMe "drives" can have the same performance characteristics and NVMe PCIe cards, given that they use the same generation and number of PCIe bus lanes.

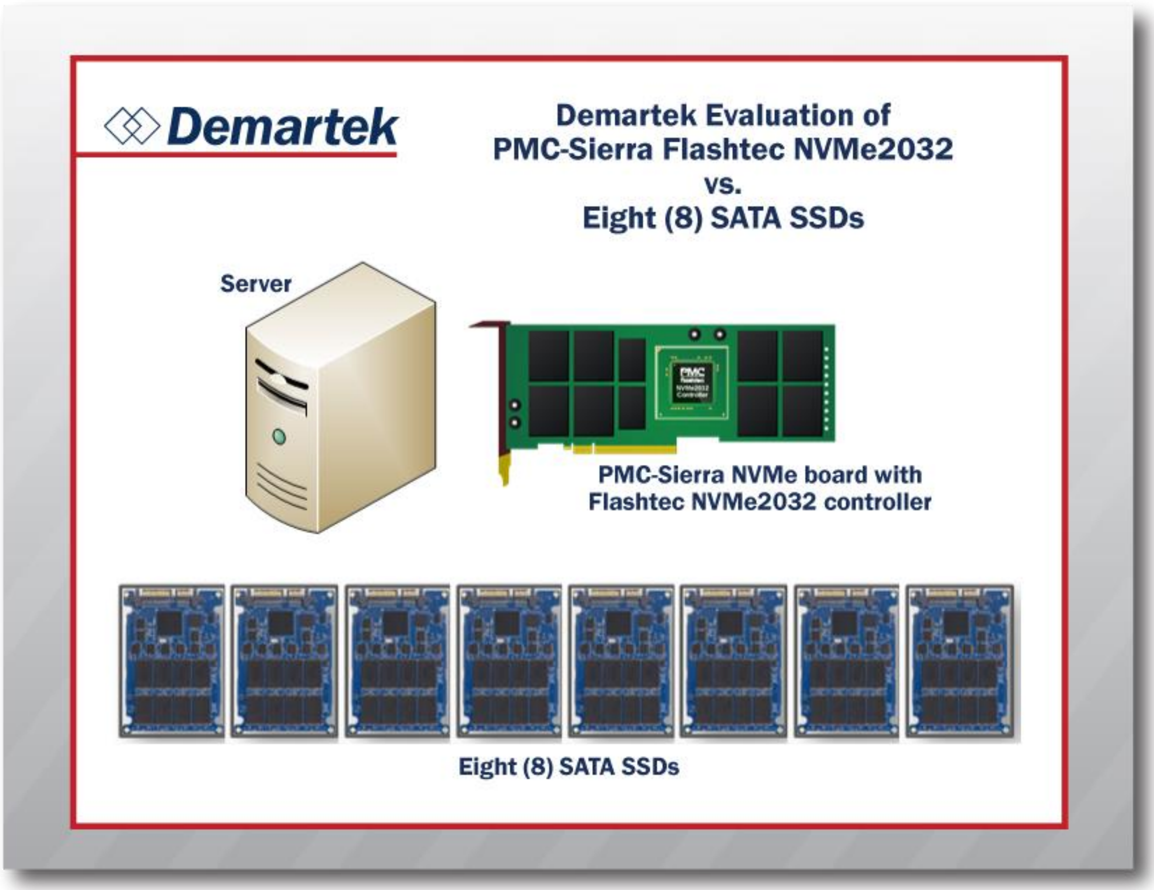
Serial ATA, or SATA, has been available for many years, and is a storage interface that uses a traditional software stack and a storage-specific physical interface. SATA currently runs over a 0.6 Gbps interface and is a point-to-point protocol, meaning that only one storage device can be on one end of the cable, while the other end is connected to a computer. SATA is used for hard disk drives (HDDs) and for solid state drives (SSDs).

The SATA SSDs chosen for this set of tests, the SanDisk Extreme PRO SATA SSDs, were among the best available in terms of performance for SATA SSDs, based on the publicly available specifications of these and other brands of SATA SSDs.

Performance Comparison Overview

In this test, we compared the performance of a single NVMe PCIe device to the performance of eight and four SATA SSDs in a real-world database application. We measured the usual storage performance metrics of IOPS, throughput and latency. We also measured CPU utilization, SQL transactions per second and the block sizes of the I/O requests.

Test Configuration



Server

- Supermicro X10DRL-i motherboard with 10 internal SATA ports
- Single Intel Xeon E5-2630 v3 processor, 2.4 GHz, 8 cores, 16 threads
- 8 GB RAM
- Windows 8.1 Enterprise
- SQL Server 2012 Standard

Storage

- Boot drive: Samsung 850 PRO 256 GB
- PMC-Sierra Flashtec NVMe2032 board, 4 TB
- 8x SanDisk Extreme PRO SSD, 480 GB (each), 3.8 TB total capacity

Database Storage Configurations

The database workload was an online transaction processing (OLTP) workload that emphasized IOPS. The database workload application required three database volumes and one log volume. Each database and log volume required approximately 350 GB of capacity. The workload was a mix of approximately 90% read and 10% write to the database volumes and mostly writes to the log volume.

Three separate data storage configurations were tested:

1. PMC-Sierra Flashtec NVMe2032 board configured with four simple volumes
2. 8x SanDisk SSDs in a Windows Storage Spaces pool with four logical volumes
3. 4x SanDisk SSDs, one simple volume per device (the entire capacity of each device was allocated)

The same database workload parameters were used for each storage configuration.

Small RAM Cache

Because there was only 8 GB of physical RAM available to the server, SQL Server was forced to perform more I/O than it would if more RAM had been available. This relatively small amount of server memory allowed a relatively equivalent comparison of the storage devices, since the application could not depend very much on caching of its data in RAM.

Test Results

Synthetic Workloads vs. Real-World Workloads

Synthetic workloads are often used to highlight storage I/O performance in "corner cases" that are defined by specific combinations of block sizes, read/write mix and I/O access patterns. These usually include: 100% random read using 4KB block size, 100% sequential read using 256KB block size and other similar patterns. A number of synthetic I/O workload tools are available for this purpose.

Real-world workloads, on the other hand, use variable levels of CPU, memory and I/O resources as the application work progresses. These may use different and multiple I/O characteristics simultaneously for I/O requests, including different block sizes, queue depths, read/write mix and I/O access patterns. As a result, real-world workloads behave differently than synthetic workloads.

In particular, Microsoft SQL Server, the base application used in the testing for this project, has its own methods of attempting to optimize I/O operations. One of these methods is to alter the block size depending on the specific work that needs to be accomplished. In addition, the best practices for SQL Server are to specify a format allocation size of 64KB for the database file volumes.

Test Duration and Parameters

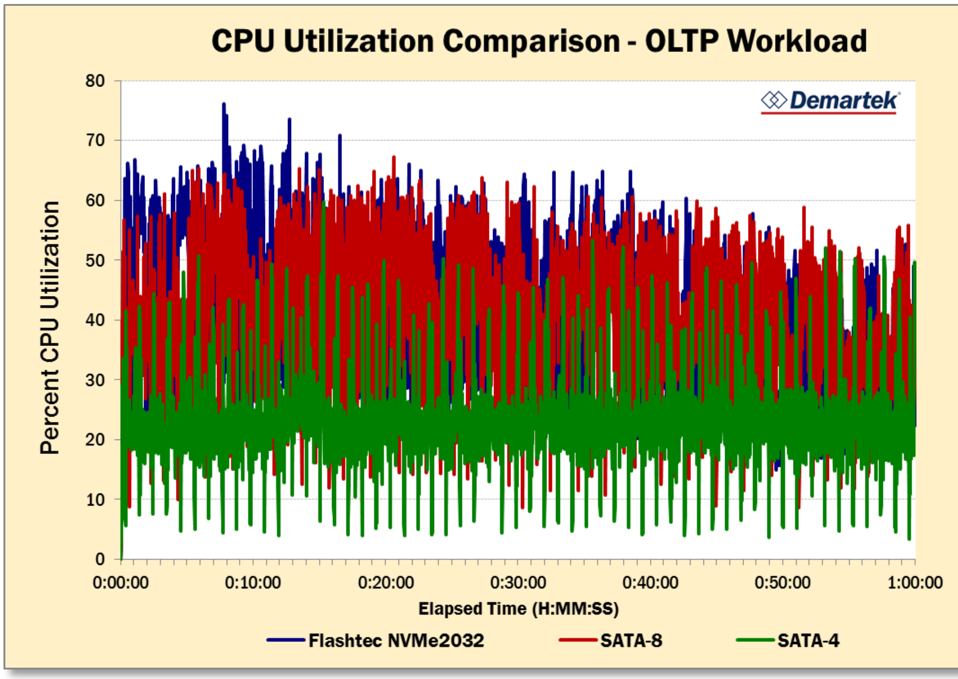
The OLTP application was started and allowed to run for more than one hour for each configuration and then it was stopped. The test results show the first hour of each run.

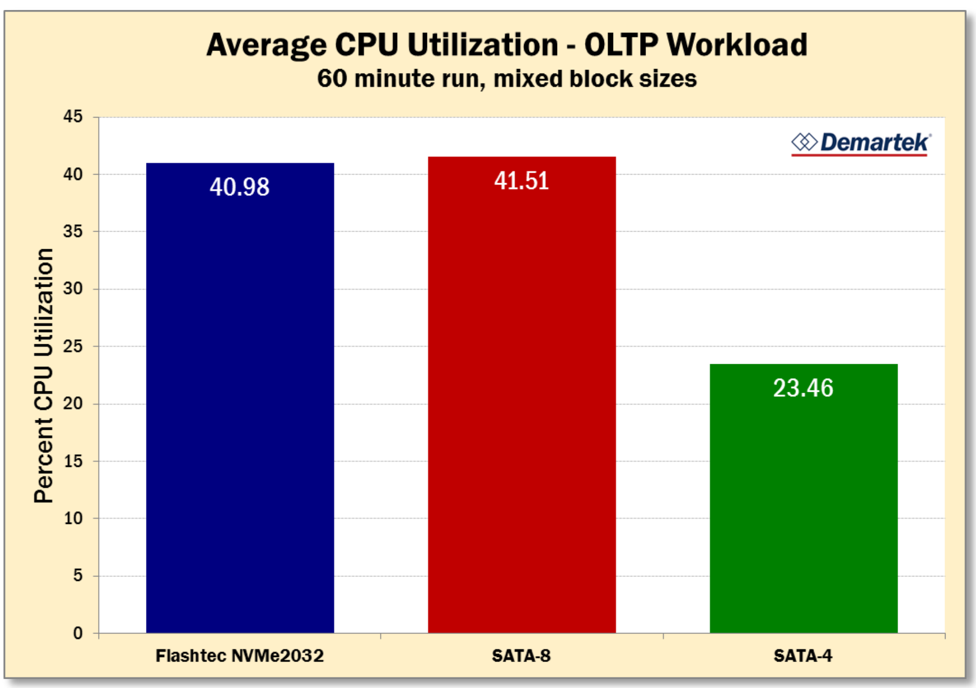
Each test configuration started with an identical copy of all the database and log volumes, so that each started from the same logical point.

Each test configuration used identical test parameters. Changes were observed in internal I/O parameters used by the application, such as block size, but these changes were not parameters that we specified for the tests. For database volume I/O, Microsoft SQL Server uses a minimum of 8 KB block size, but can and often does use a larger block size.

Test Results – CPU Utilization

The CPU utilization results show that the host CPU utilization was approximately the same for the NVMe storage configuration as the eight-drive SATA SSD configuration. This is interesting in light of the results that follow.



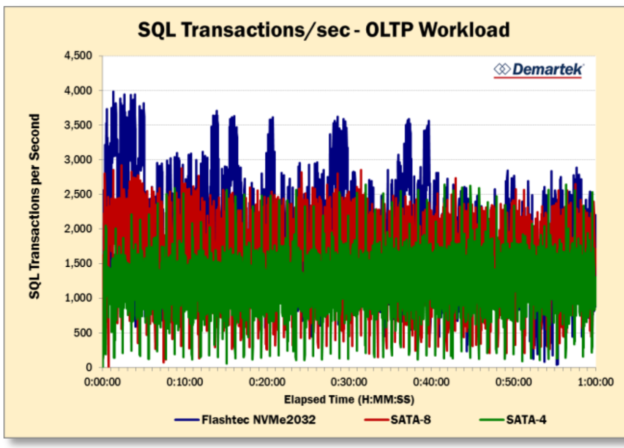


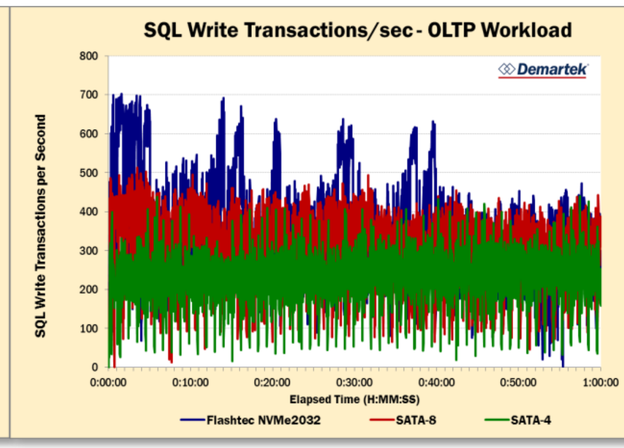
Test Results – SQL Transactions per Second

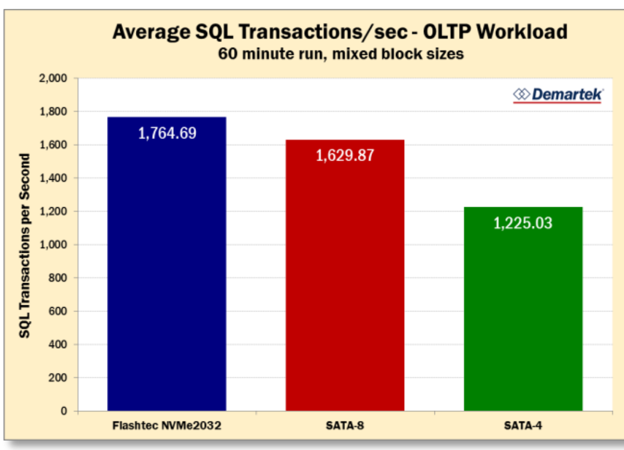
One of the common application metrics for a SQL Server database workload is the SQL transactions per second. This is a measure of the amount of application work completed.

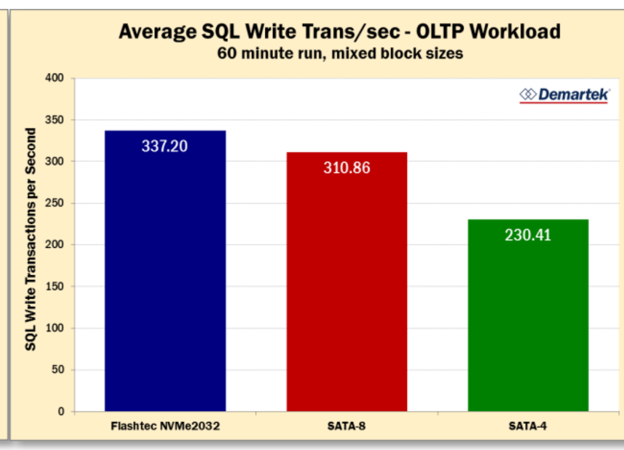
The NVMe storage performed better than the SATA SSD configurations for both overall transactions/second and for write transactions/second.

There are some similarities between the SQL transactions per second and the Read IOPS data shown further in this report.



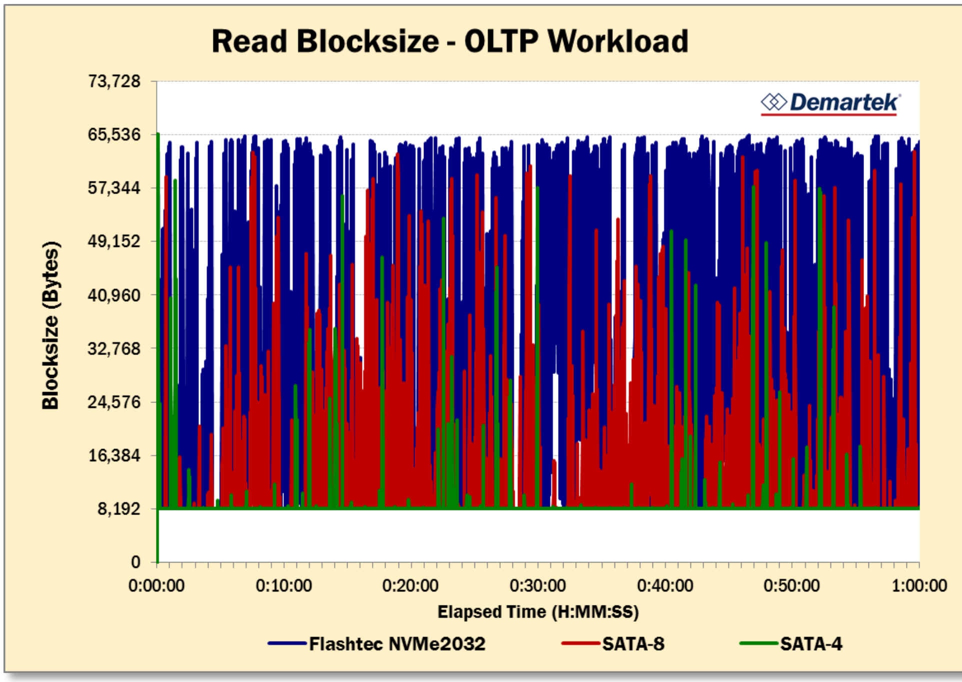


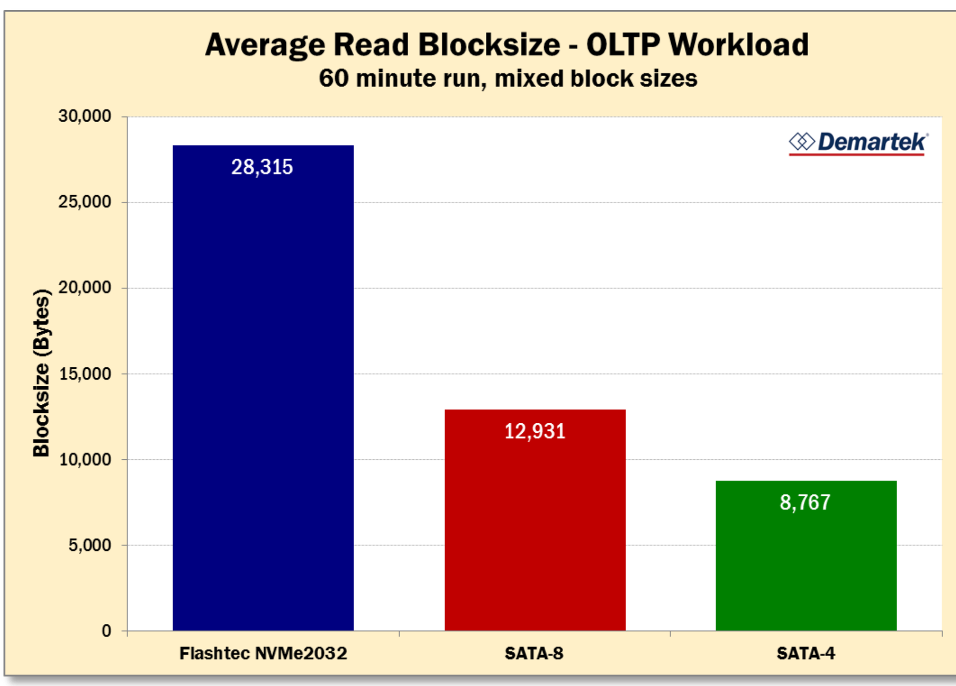




Test Results – Read Block Sizes

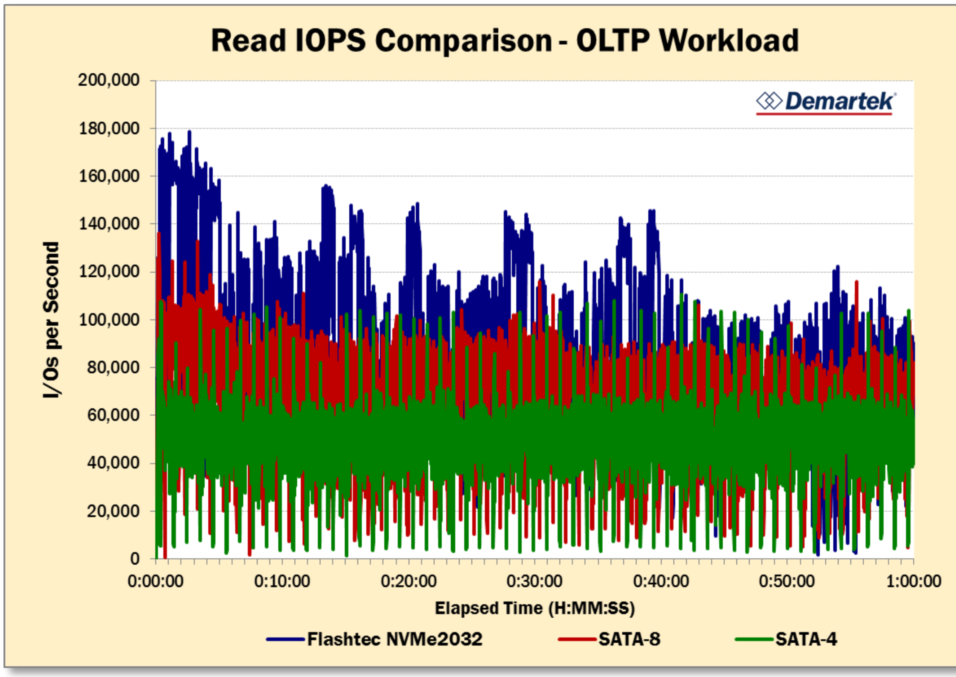
The application workload varied the block size used for reads (90% of the workload) depending on the storage configuration. The block size changes (8 KB – 64 KB) affect the throughput and overall performance.

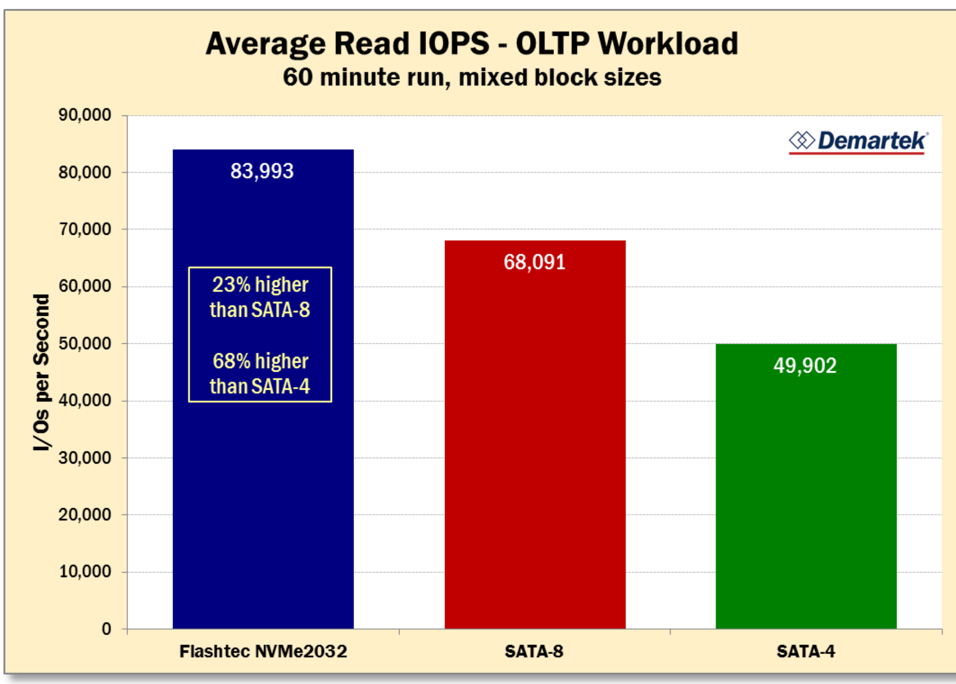




Test Results – Read IOPS

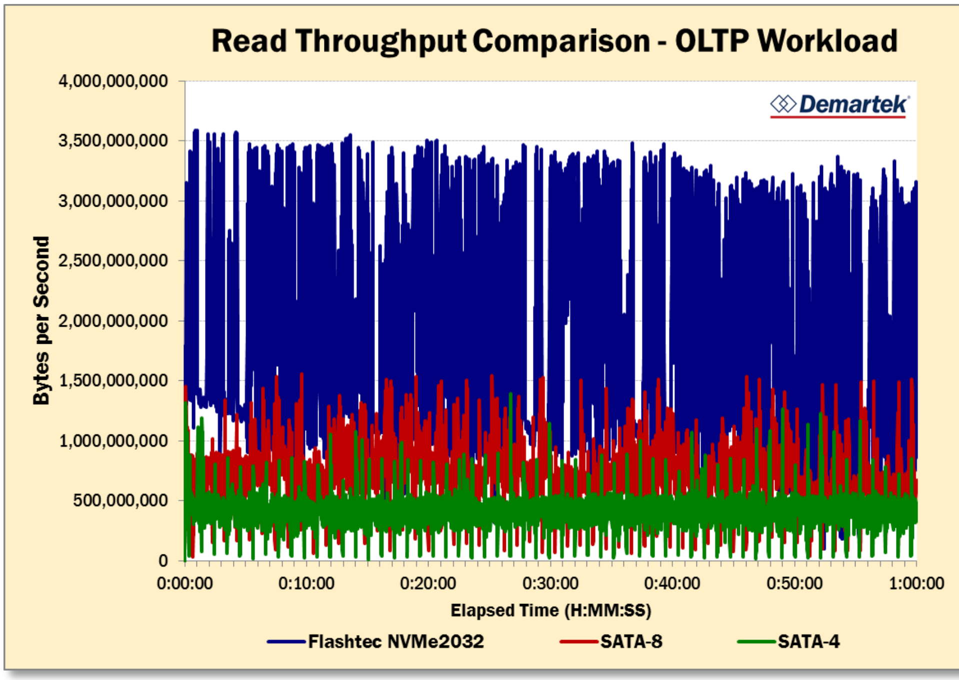
The Read IOPS were somewhat higher for the NVMe storage than for the other storage configurations. This suggests that the NVMe protocol and PCIe interface allow more I/O requests for approximately the same amount of CPU cycles.

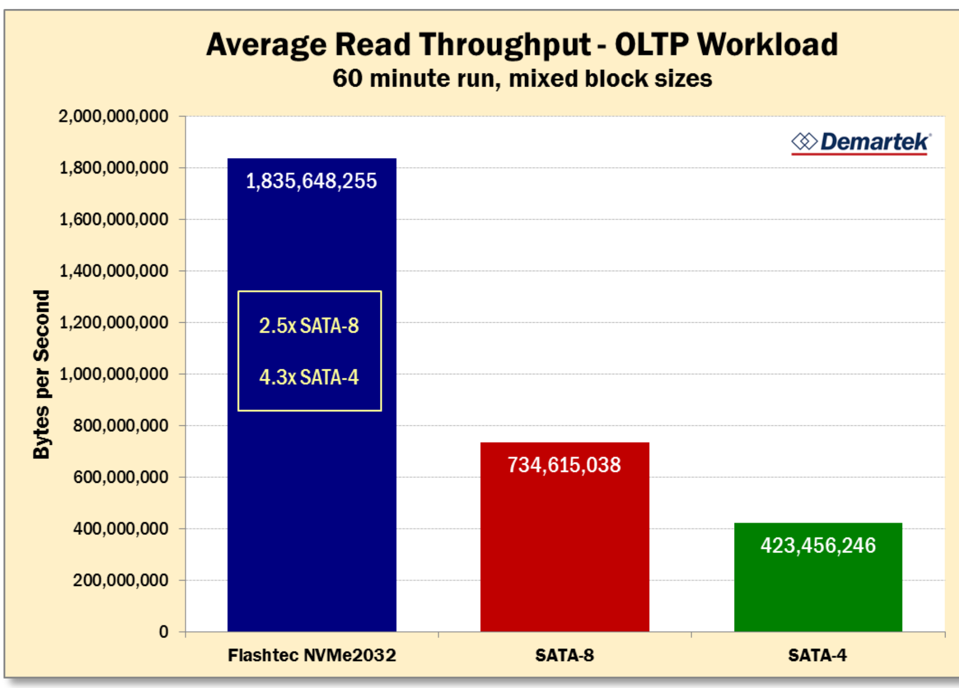




Test Results – Read Throughput

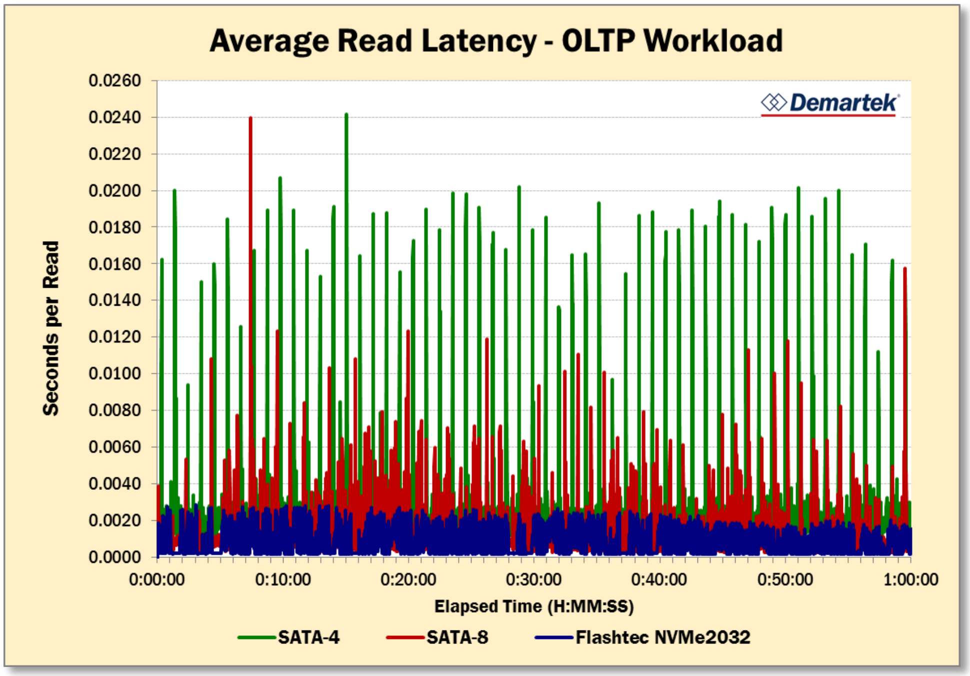
The difference in throughput is significantly larger for the NVMe device than for either of the configurations of the SATA drives grouped together. This speaks to the efficiency of the NVMe protocol, both in terms of its speed and because larger block sizes were used. The peak throughput for the NVMe solution was approximately 3.5 Gigabytes/sec.

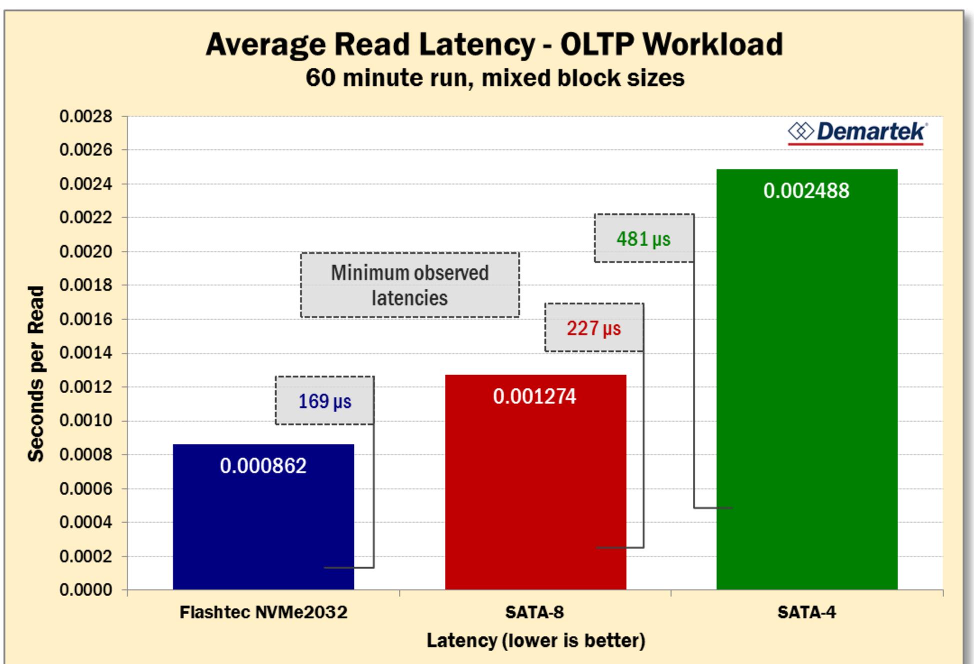




Test Results – Read Latency

Latency, or round-trip time, as it is sometimes called, is a measure of the time it takes to complete I/O operations. Because it refers to the time to complete a unit of work, lower is better. The PMC-Sierra NVMe solution had consistently lower read latency (< 1 millisecond) than the other solutions. It is not unusual to see spikes in latency measurements over a long run. The minimum observed latencies are also indicated.





Summary and Conclusion

The PMC-Sierra Flashtec NVMe2032 controller and solid-state storage device provided better performance and latency than a combined configuration of eight SATA SSDs and four SATA SSDs. We also found that the CPU utilization for the NVMe solution was approximately the same as with the eight-drive SATA SSD configuration, yet the NVMe solution performed more useful work for that CPU utilization, making more efficient use of the CPU.

The most current version of this report is available at http://www.demartek.com/Demartek_PMC-Sierra_Flashtec_NVMe2032_Evaluation_2015-09.html on the Demartek website.