KIOXIA

CD6-R Series PCIe® Gen4 NVMe™ Data Center SSD Performance Gains in MySQL Database Environments over PCIe Gen3 SSDs

KIOXIA is one of the first storage vendors to deliver data center SSDs based on the PCIe 4.0 and NVMe specifications. These PCIe 4.0 SSDs better saturate PCIe lanes to continually feed data to today's fast CPUs, which helps achieve better CPU utilization. In comparison to Gen3, Gen4 performance delivers twice the data transfer speed from 8 gigatransfers per second (GT/s) to 16 GT/s, and can move data at approximately 2 gigabytes per second (GB/s) per lane (versus about 1GB/s per lane maximum with the PCIe 3.0 interface). The end result is that the PCIe 4.0 interface enables the doubling of performance and deliver a 4-lane (x4) bandwidth of approximately 8GB/s (Table 1).

Specification				Throughput			
PCIe Revision	Introduced by PCI-SIG	Transfer Rate (GT/s)	x1 (GB/s)	x2 (GB/s)	x4 (GB/s)	x8 (GB/s)	x16 (GB/s)
3.0	2010	8.0	0.9846	1.969	3.94	7.88	15.75
4.0	2017	16.0	1.969	3.938	7.88	15.75	31.51

Table 1: PCIe Gen4 can double GT/s and by-lane performance versus PCIe Gen3 (Source: PCI-SIG)

In enterprise applications, PCIe 4.0 technology is well-suited for workloads that need the utmost SSD performance. Enterprise SSDs typically include high-end features such as dual-port capabilities, larger capacities, high endurances and data protection schemes. For data center workloads, scale-out and hyperscale environments require SSDs to deliver high read-intensive performance and focus on quality of service, but do not have a need for a high availability, enterprise feature-set. These large data center environments manage data integrity and availability at the node or rack level, and require an SSD that is designed for high performance, without the unnecessary feature set typically found in higher end enterprise SSDs. For these applications, the KIOXIA CD6-R Series is a good solution that delivers Gen4 performance in a robust and optimized design.

To validate the performance increases that can be achieved when moving from PCIe Gen3 data center SSDs to the latest PCIe Gen4 specification, KIOXIA conducted a performance analysis that included its CD6-R Series of PCIe Gen4 NVMe data center SSD against a leading and currently available PCIe Gen3 data center SSD. Testing measured both transactions per minute (TPM) and average read/write latency of a mainstream server platform running a relational database management system (RDBMS). Supporting information includes a description of the benchmark tests (criteria, set-up and procedures), and test results (visual representation and analysis).

Test results provide a real-world expectation of the TPM performance and average combined latency that can be achieved when running a MySQL database using comparable equipment and performing queries against it.

Description of Benchmark

Benchmark tests were conducted by KIOXIA in a lab environment that compared TPM performance and transactional read and write latency of a mainstream server platform. The storage configuration included a CD6-R Series PCIe Gen4 NVMe data center SSD and a currently shipping PCIe Gen3 NVMe data center SSD from a leading vendor (referred to as Vendor A). The tests utilized an operational, high-performance online transaction processing (OLTP) MySQL database workload based on comparable TPC-C™ benchmarks created by HammerDB[1] software.

Test Conditions

The hardware and software equipment used for these benchmark tests included:

- **AMD Based QCT S5K Server:** One (1) dual socket server with two (2) AMD EPYC™ 7702P processors, featuring a total of 128 processing cores, 2.0 GHz frequency, and 128 gigabytes[2] (GB) of DDR4 RAM (3,200 speed)
- **Operating System:** CentOS™ v8.2
- **Application:** MySQL v8.0
- **Benchmark Software:** Comparable TPC-C benchmark tests generated through HammerDB v3.3 test software
- **Storage Devices (Table 2):**
  One (1) KIOXIA CD6 Series PCIe Gen4 SSD with 7.68 terabyte[2] (TB) capacity
  One (1) Vendor A PCIe Gen3 SSD with 7.68 TB capacity

Specifications	CD6-R Series	Vendor A
Interface	PCIe 4.0	PCIe 3.0
Capacity	7.68 TB	7.68 TB
Form Factor	2.5-inch (15mm)	2.5-inch (7mm)
NAND Flash Type	BiCS4 3D NAND (96-layer)	3D TLC (64-layer)
Drive Writes per Day[3] (DWPD)	1 (5 yrs)	1.3 (3 yrs)

Key Setup Parameters	CD6-R Series	Vendor A
Data Warehouses[4]	5,000	5,000
Virtual User Count	64	64
InnoDB Buffer Pool Size	64 GB	64 GB

Table 2: SSD specifications and set-up parameters

Set-up & Test Procedures

The test system was configured using the hardware and software equipment outlined above. The MySQL application was configured to store both the log and database on the SSD that was tested. Additionally, HammerDB software was run locally on the server as there was excess CPU available to not cause bottlenecks in performance. It was configured with a test schema based on the TPC-C benchmark (to emulate a MySQL OLTP database environment). The MySQL application was then loaded with 5,000 data warehouses that comprised about 442 GB of the server's storage capacity. The MySQL InnoDB storage engine buffer pool size was set to 64 GB, which corresponds to the amount of memory that is allocated to the MySQL application.

Additionally, the test tool was configured for 64 virtual users to simultaneously send query threads in order to obtain responses. The query response time was also set to one millisecond (ms), demonstrating the ability to achieve very fast responses.

TPM and latency tests were run once for each tested SSD series and the latency results were gathered while the TPM test was ongoing.

The objective of the tests was to showcase how a PCIe Gen4 NVMe data center SSD provides significantly higher performance and lower latency for common database applications when compared to a PCIe Gen3 SSD.

Test Results

TPM and latency benchmarks were conducted with the result recorded. For TPM, the higher the value, the better the result. For read and write latency, the lower the value, the better the result.

Transactions Per Minute

In an OLTP database environment, TPM is a measure of how many transactions in the TPC-C transaction profile are being executed per minute. The HammerDB software, executing the TPC-C transaction profile, randomly performs new order transactions and randomly executes additional transaction types such as payment, order status, delivery and stock levels. This benchmark simulates an OLTP environment where there are a large number of users that conduct simple, yet short transactions that require sub-second response times and return relatively few records.

The TPM test results (Figure 1):

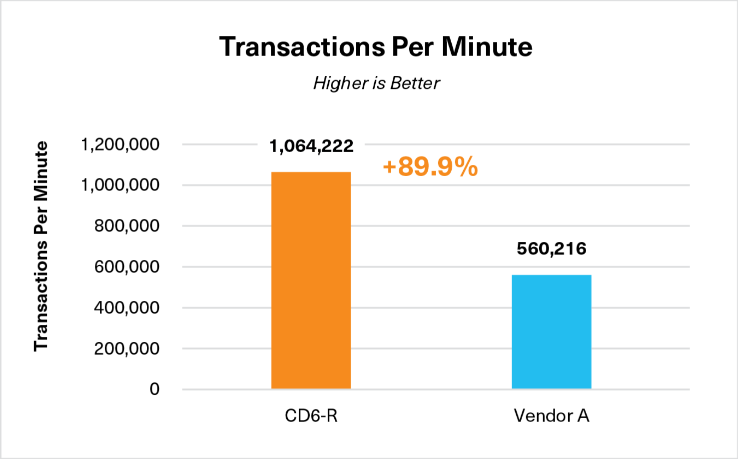
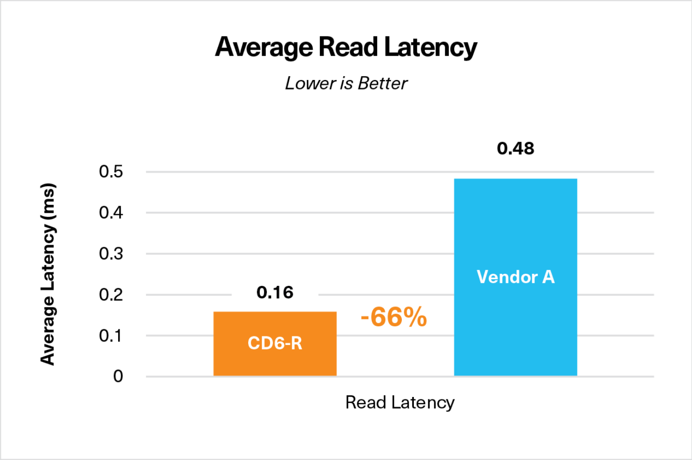


Figure 1: TPM comparison of data center PCIe Gen4 and PCIe Gen3 NVMe SSDs (higher is better)

Average Read and Write Latency

Latency is the delay in time before a storage device completes a data transaction following an instruction from the host for that request. The time it takes for the data requests to complete can greatly affect application performance and the user experience. Application response time often has built-in latency between the user and the server, so maintaining low-latency within the server will usually translate into a better user experience.

The read and write latency test results (Figure 2):



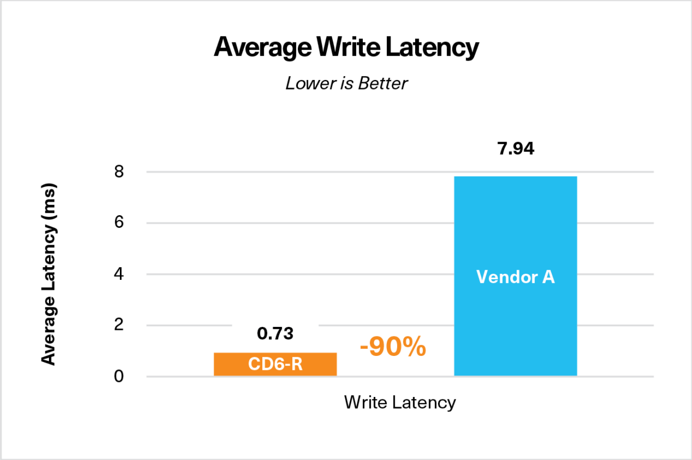


Figure 2: Latency comparison of data center PCIe Gen4 vs PCIe Gen3 NVMe SSDs (lower is better)

Test Analysis

From the benchmark test results, the PCIe Gen4 NVMe data center SSD (KIOXIA CD6-R Series) enabled the server to deliver almost 90% more transactions per minute when compared to the PCIe Gen3 NVMe SSD. The CD6-R Series' TPM performance enables a mainstream server platform to support significantly more transactions, all while servicing transactions with over 66% better read latency.

Better storage performance also has a domino effect on overall system performance, to the point of potentially reducing DRAM requirements or other factors. To demonstrate the advantages of the CD6-R PCIe Gen4 NVMe data center SSD as it relates to memory allocation, additional tests were conducted that limited the MySQL application to 32 GB of available DRAM (versus 64 GB as set in the original performance test). For this DRAM-specific test, the CD6-R Series SSD delivered 662,975 transactions per minute, while the Vendor A drive delivered a TPM of 337,419. This resulted in a 96% performance advantage for the CD6-R Series SSD.

The CD6-R Series SSD result of 662,975 TPM when under a DRAM constraint of 32 GB outperformed the Vendor A system configuration when 64 GB of DRAM is available, delivering 560,216 transactions per minute. Even with half of the DRAM allocated, the CD6-R Series with 32 GB of DRAM delivered over 18% better TPM when compared to Vendor A's performance with 64 GB of DRAM allocated (Figure 3). This result demonstrated that increased SSD performance impacts system performance regardless of a scale-up or a scale-down.

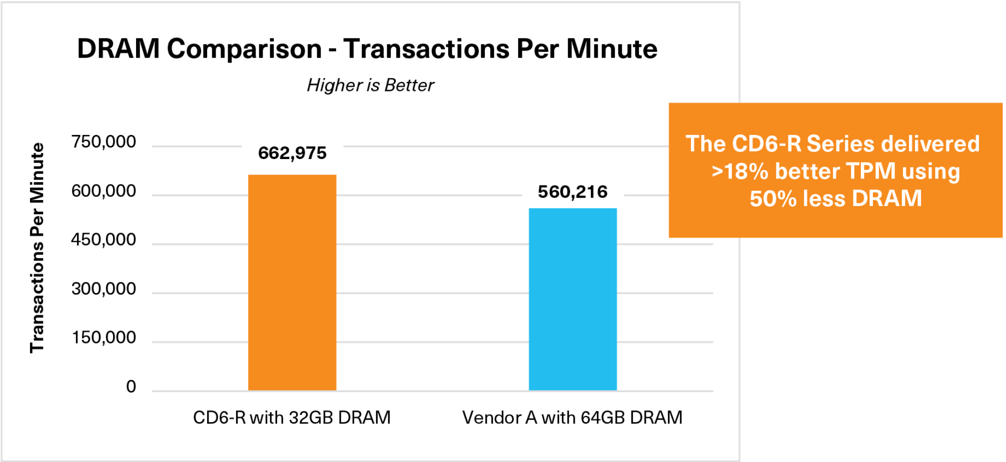


Figure 3: TPM comparison of data center PCIe Gen4 vs PCIe Gen3 NVMe SSDs using 50% less DRAM (higher is better)

Systems deployed with PCIe Gen4 NVMe data center SSDs can deliver more TPM, allowing clustered systems to run many more transactions per node. Higher transactions per node enable a large scale out of MySQL clusters that utilize fewer servers for hosting MySQL databases, and significantly lower acquisition and operating cost per transaction served. It also serves more users per node which ultimately translates to fewer nodes that are required to meet cluster capacity or performance targets. The lower latency per transaction can also help to lower response times, which in turn improve the user experience.

AMD EPYC Nodes Per Socket[5] Setting

It should also be noted that the AMD EPYC-based QCT S5K server processor settings were changed from a default Nodes per Socket (NPS) that equaled '4' to an NPS setting that equaled '1.' The NPS settings impact how memory is allocated to the CPU where an 'NPS=1' setting enables interleaving of all eight (8) channels for each socket, and creates a single NUMA node per socket. Therefore, it is recommended to change the 'NPS' setting from '4' to '1' which increases per-core memory bandwidth (at the expense of memory latency), which in turn, increases the CPU's ability to move data from memory to the SSD across the PCIe bus. Since all memory channels are available to all cores, the CPU is able to process transactions to disk faster, resulting in overall higher performance from both the system and the KIOXIA CD6-R SSD.

With this change from 'NPS=4' to 'NPS=1', the CD6-R performance improved from 993,593 transactions per minute ('NPS=4') to 1,064,222 transactions per minute ('NPS=1'), and had an over 7% improvement in TPM performance (Figure 4).

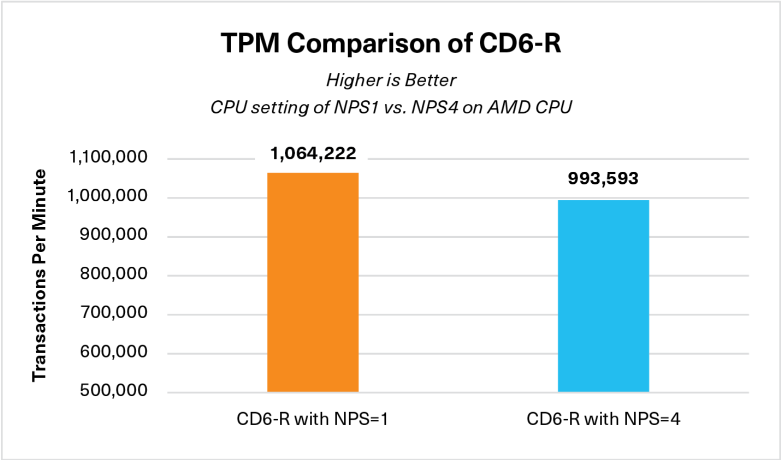


Figure 4: TPM comparison of a CD6-R Series PCIe Gen4 NVMe data center SSD with the CPU setting at '1' and '4' (higher is better)

Summary

The testing demonstrated that a single KIOXIA CD6-R PCIe Gen4 NVMe data center SSD with less DRAM can outperform a currently shipping PCIe Gen3 NVMe SSD from a leading vendor with more DRAM, and can have a dramatic impact on system sizing. If extra performance is needed, then CD6-R Series SSDs can be utilized to get greater throughput out of the system.

KIOXIA CD6-R Series PCIe Gen4 NVMe data center SSDs demonstrate both value and high performance, which are key benefits that come to mind when selecting storage capabilities for data center environments. The series demonstrates significantly improved performance and latency compared to the PCIe Gen3 NVMe SSD and offers a range of storage capacities up to 15.36 TB. The CD6-R Series delivers enterprise-class quality with support of one and three DWPD options for read-intensive and mixed use applications respectively.

Driven by the PCIe 4.0 interface and NVMe 1.4 specification, CD6-R Series SSDs deliver more storage bandwidth and input/output operations per second (IOPS) than SSDs based on the PCIe 3.0 interface, enabling them to complete over 90% more database transactions per minute, and service more users simultaneously. These data center SSDs represent KIOXIA's commitment to quality and innovation, and deliver an improved user database experience.

The CD6-R Series is included in the KIOXIA CD Series of PCIe Gen4 NVMe data center SSDs. Additional CD6-R Series PCIe Gen4 NVMe data center SSD information is available here.

NOTES:

[1] HammerDB is benchmarking and load testing software that is used to test popular databases. It simulates the stored workloads of multiple virtual users against specific databases to identify transactional scenarios and derive meaningful information about the data environment, such as performance comparisons. TPC Benchmark C is a supported OLTP benchmark that includes a mix of five concurrent transactions of different types, and nine types of tables with a wide range of record and population sizes and where results are measured in transactions per minute.

[2] Definition of capacity - KIOXIA Corporation defines a megabyte (MB) as 1,000,000 bytes, a gigabyte (GB) as 1,000,000,000 bytes and a terabyte (TB) as 1,000,000,000,000 bytes. A computer operating system, however, reports storage capacity using powers of 2 for the definition of 1Gbit = 2^{30} bits = 1,073,741,824 bits, 1GB = 2^{30} bytes = 1,073,741,824 bytes and 1TB = 2^{40} bytes = 1,099,511,627,776 bytes and therefore shows less storage capacity. Available storage capacity (including examples of various media files) will vary based on file size, formatting, settings, software and operating system, and/or pre-installed software applications, or media content. Actual formatted capacity may vary.

[3] Drive Write(s) per Day: One full drive write per day means the drive can be written and re-written to full capacity once a day, every day, for the specified lifetime. Actual results may vary due to system configuration, usage, and other factors.

[4] The 5,000 data warehouses, in combination with the 64G buffer pool size, represents a real-world database configuration that creates a normal database size. In the TPC-C benchmarks created by HammerDB, 5,000 data warehouses equates to a database size of 442 GB, with the ability to cache 14.48% of the database with 64 GB of RAM.

[5] At a Nodes per Socket (NPS) setting of 4, and 2 CPUs, the system regards this set-up as 8 logical CPUs, with multiple cores, and assigns 1/8 of memory to each logical node. The memory split means that less processing cores can access each segment of memory, however, memory access within that segment is faster. When the Nodes per Socket (NPS) setting was changed from default '4' to a '1' setting, TPM performance increased. General NPS information can be found at https://developer.amd.com/wp-content/resources/56745_0.75.pdf, section 2.2.2, page 10.

KIOXIA