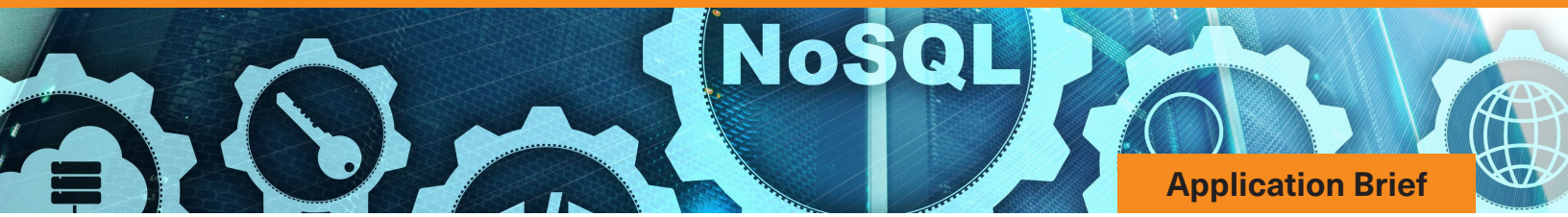
KIOXIA



Application Brief

NoSQL Database Workloads Performance Gains with the Latest PCIe® 5.0 Servers and SSDs

Featuring KIOXIA CD8P Series Data Center NVMe™ SSDs in a Dell™ PowerEdge™ R770 Server, running Aerospike® Database Software

NoSQL databases are a type of database management system that stores data in a non-relational format, enabling the flexibility and scalability to handle large, unstructured datasets. Unlike traditional relational databases which organize data into tables with rows and columns, NoSQL databases utilize different data structures typically in four categories: key-value pairs, documents, graphs and wide columns. This flexibility allows these databases to manage diverse data types and adapt to changing data models efficiently. NoSQL databases are commonly used by large organizations that deal with massive amounts of data and complex data structures.

The role of any database management system is to provide data quickly, accurately and securely across an organization. However, the abundance of data generated in modern data centers has become difficult to manage and access quickly. The need for database technologies to be faster and keep up with the increase in data generation is an architectural priority for many data centers worldwide.

The latest available PCIe 5.0 interface can help meet database objectives with higher data storage performance and lower latency than previous PCIe generations. The PCIe 5.0 interface increases data transfer speeds from 16 gigatransfers per second (GT/s) to 32 GT/s and can move data at approximately 4 gigabytes per second (GB/s) per lane versus about 2 GB/s per lane (PCIe 4.0 interface) and 1 GB/s per lane (PCIe 3.1 interface). SSDs supported by the PCIe 5.0 interface deliver the fastest bandwidth performance to date.

This application brief presents a performance comparison of database operations executed on a new PCIe 5.0 server/SSD configuration versus a previous generation server/SSD configuration. The PCIe 5.0 configuration included a Dell PowerEdge R770 server deployed with eight 3.84 terabyte[1] (TB) KIOXIA CD8P-R Series Data Center NVMe SSDs. The previous generation comparison configuration included a Dell PowerEdge R740xd server deployed with eight 3.84 TB KIOXIA CD6-R Series Data Center NVMe SSDs. Both systems ran an Aerospike NoSQL database driven by synthetic tests from Yahoo!™ Cloud Serving Benchmark (YCSB) software.

Aerospike is a NoSQL key-value database[2] capable of delivering very fast runtime performance for all-sized read and write workloads, as its architecture is flash-optimized. It features direct device access that enables high throughput and low latency delivered directly from the storage device itself. To take advantage of this capability, the underlying SSDs must be fast enough to support it.

There were seven tests in the comparison covering: (1) Runtime; (2) Database Throughput; (3) 50% Read / 50% Update Latency; (4) 95% Read / 5% Update Latency; (5) 100% Read Latency; (6) 95% Read / 5% Insert Latency; and (7) 50% Read / 50% Read-Modify-Write Latency. The test database created on the Aerospike database consisted of 300,000,000 records and 200,000,000 operations. The database throughput workload was especially important to see how many more operations per second that the PCIe 5.0 Intel® Xeon® 6767P CPU and storage backend can deliver over the previous generation configuration.

Test results show that the new PCIe 5.0 configuration demonstrated a faster runtime, higher database throughput and lower latencies of NoSQL workloads when compared with the previous generation configuration. The test results presented include a brief description of each workload, a graphical depiction of the test results and analysis. Appendix A covers the hardware and software test configuration. Appendix B covers the configuration setup and test procedures.

Test Results Snapshot

The PCIe 5.0 configuration (KIOXIA CD8P Series SSDs / Dell™ PowerEdge™ R770 server) delivers exceptional NoSQL database performance when compared with the previous generation configuration as depicted by the test results below:

Runtime
(lower is better)
up to 52% Lower

Database Throughput
(higher is better)
up to 109% Higher

50% Read / 50% Update Latency
(lower is better)
up to 47% Lower

95% Read / 5% Update Latency
(lower is better)
up to 71% Lower

100% Read Latency
(lower is better)
up to 61% Lower

95% Read / 5% Insert Latency
(lower is better)
up to 75% Lower

50% Read / 50% Read-Modify-Write Latency
(lower is better)
up to 39% Lower

Test Results[3]

Workloads were run using YCSB software against the Aerospike® NoSQL database with metrics recorded for total runtime, database throughput and five database operations that included:

1. *50% Read / 50% Update Latency:*
   Read recent database actions 50% of the time / Update the database actions 50% of the time.

2. *95% Read / 5% Update Latency:*
   Read recent database actions 95% of the time / Update the database actions 5% of the time.

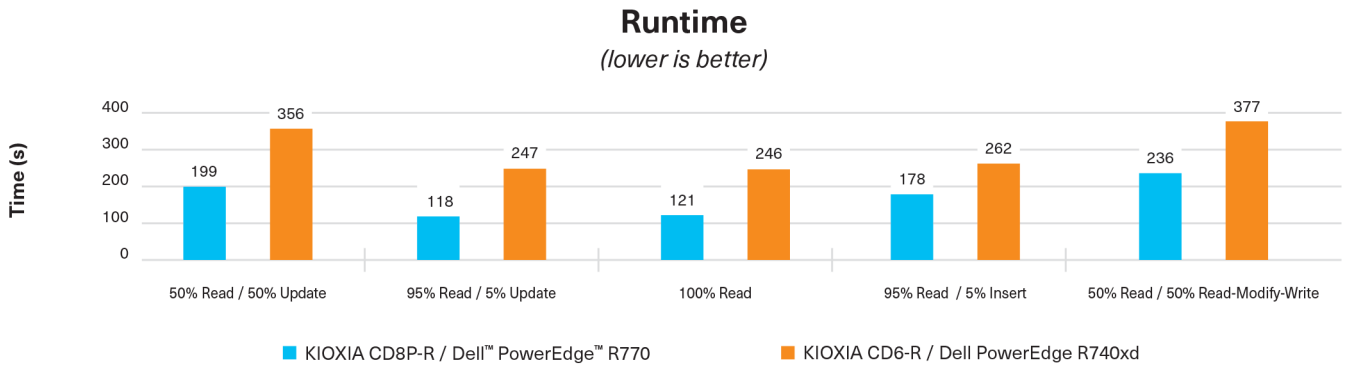
3. *100% Read Latency:*
   Perform a database read operation.

4. *95% Read / 5% Insert Latency:*
   Read recent database actions 95% of the time / Insert new data 5% of the time.

5. *50% Read / 50% Read-Modify-Write Latency:*
   Read recent database actions 50% of the time / Perform database read-modify-write operations 50% of the time.

Test 1: Runtime

This test measured the total time required in seconds (s) to complete each of the five database workloads outlined above. The results are based on an average of three test runs and the lower result is better.



Database Workloads: Runtime Tests *(in seconds)*	PCIe® 5.0 Configuration *(KIOXIA CD8P-R / Dell™ PowerEdge™ R770)*	Previous Generation Configuration *(KIOXIA CD6-R / Dell PowerEdge R740xd)*	PCIe 5.0 Configuration Gains
50% Read / 50% Update	199	356	44%
95% Read / 5% Update	118	247	52%
100% Read	121	246	50%
95% Read / 5% Insert	178	262	32%
50% Read / 50% Read-Modify-Write	236	377	37%

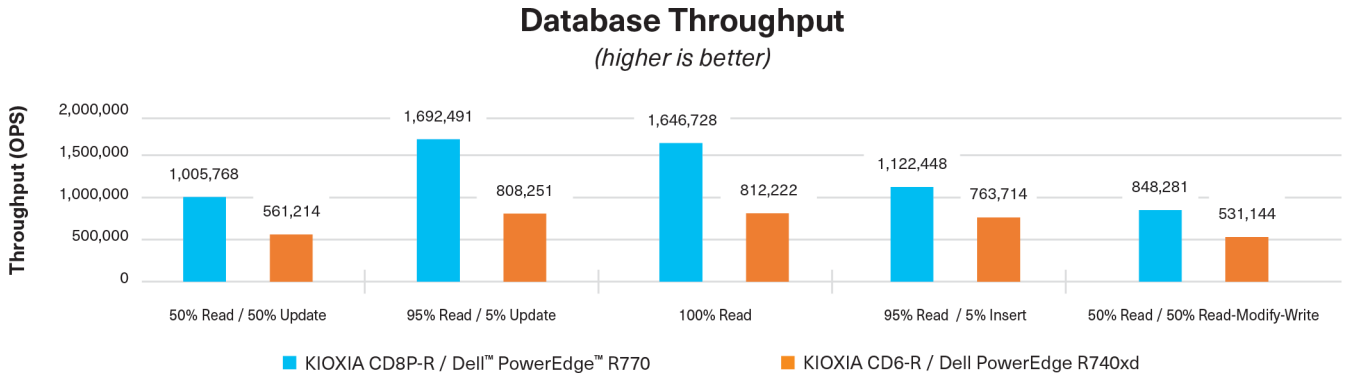
Database runtime results

Analysis:

The test results show that the PCIe 5.0 configuration delivered a lower query time across the five database workloads versus the previous generation configuration. Faster completion time of an individual operation or workload allows the database to handle high volume requests and more users concurrently. The PCIe 5.0 configuration was able to show increased database runtimes required to speed up various workloads when compared to the previous generation configuration.

Test 2: Database Throughput

This test measured the number of operations per second (OPS) a system can complete on average for each of the five database workloads. The results are based on an average of three test runs and the higher result is better.



Aerospike® Operational Workloads: Database Throughput Tests *(in operations per second)*	**PCIe® 5.0 Configuration** *(KIOXIA CD8P-R / Dell™ PowerEdge™ R770)*	**Previous Generation Configuration** *(KIOXIA CD6-R / Dell PowerEdge R740xd)*	**PCIe 5.0 Configuration Gains**
50% Read / 50% Update	1,005,768	561,214	79%
95% Read / 5% Update	1,692,491	808,251	109%
100% Read	1,646,728	812,222	102%
95% Read / 5% Insert	1,122,448	763,714	46%
50% Read / 50% Read-Modify-Write	848,241	531,144	59%

Database throughput results

Analysis:

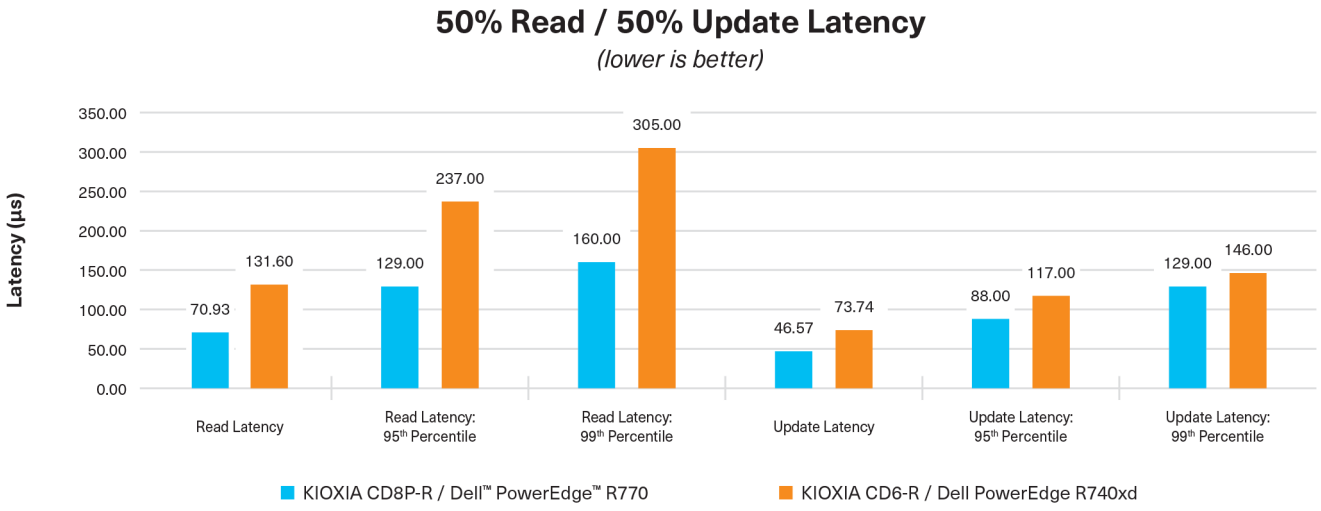
The test results show that the PCIe 5.0 configuration completed more operations per second across the five database workloads than the previous generation configuration. One scenario that can occur is when the number of incoming operations from concurrent users is higher than the achievable database throughput supported by the hardware and software configuration. If this occurs, the server could overload and create longer waiting times per query that could negatively affect application performance and user experience. The PCIe 5.0 configuration was able to achieve higher database throughput and did not encounter this scenario as often as the previous generation configuration. This performance was a result of shorter waiting times per query and more database operations that were performed simultaneously by multiple users.

Test 3: 50% Read / 50% Update Latency

This test measured the time it took in microseconds (µs) to perform an individual read or update operation within the 50% Read / 50% Update workload. It includes the time it took for the YCSB workload generator to not only issue the read or update operation, but also the time it took to complete the operation and receive a 'successfully completed' acknowledgement. The results are based on an average of three test runs and included the following metrics:

- *Average Read Latency:*
  The average time it took to perform a database read operation.
- *Average Read Latency at the 95th Percentile:*
  The average time it took to perform a database read operation at the 95th percentile (or where 95% of the operations had this read latency or lower).
- *Average Read Latency at the 99th Percentile:*
  The average time it took to perform a database read operation at the 99th percentile (or where 99% of the operations had this read latency or lower).
- *Average Update Latency:*
  The average time it took to perform a database update operation.
- *Average Update Latency at the 95th Percentile:*
  The average time it took to perform a database update operation at the 95th percentile (or where 95% of the operations had this update latency or lower).
- *Average Update Latency at the 99th Percentile:*
  The average time it took to perform a database update operation at the 99th percentile (or where 99% of the operations had this update latency or lower).

The results are in 'µs' and the lower result is better.



Database Workloads: 50% Read / 50% Update Latency Tests *(in microseconds)*	PCIe® 5.0 Configuration *(KIOXIA CD8P-R / Dell™ PowerEdge™ R770)*	Previous Generation Configuration *(KIOXIA CD6-R / Dell PowerEdge R740xd)*	PCIe 5.0 Configuration Gains
Read Latency	70.93	131.60	46%
Read Latency: 95th Percentile	129.00	237.00	45%
Read Latency: 99th Percentile	160.00	305.00	47%
Update Latency	46.57	73.74	36%
Update Latency: 95th Percentile	88.00	117.00	24%
Update Latency: 99th Percentile	129.00	146.00	11%

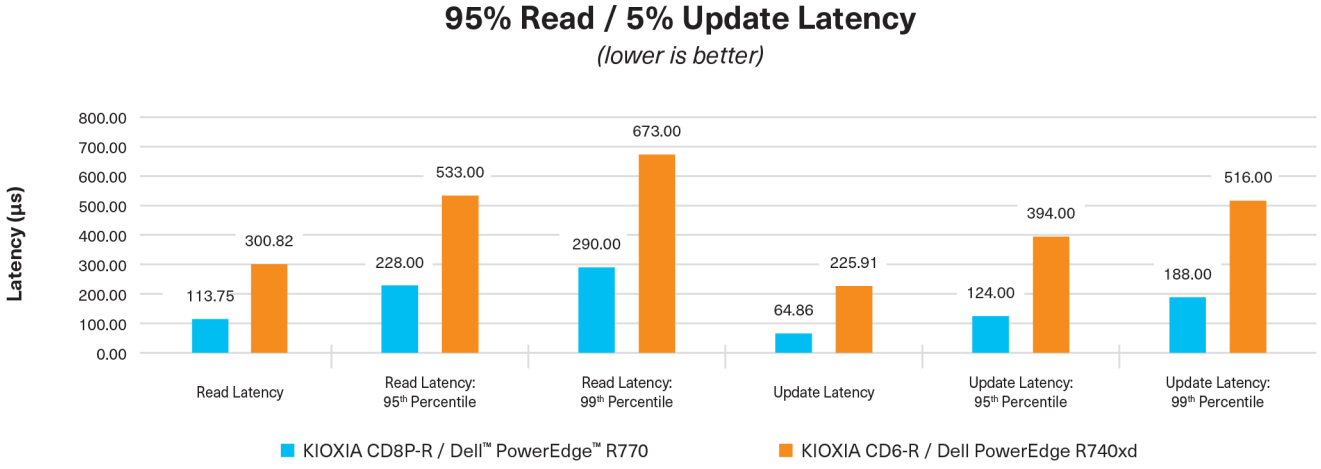
50% read / 50% update latency results

Analysis:

The test results show that the PCIe 5.0 configuration completed read and update operations as measured by each of the six metrics in less time than the previous generation configuration. One example of an application use case that benefits from this performance is key-value storage for social media sites where fast data streaming represents the 50% read and posting the data represents the 50% update.

Test 4: 95% Read / 5% Update Latency

This test measured the time it took in microseconds (µs) to perform an individual read or update operation within the 95% Read / 5% Update workload. It includes the time it took for the YCSB workload generator to not only issue the read or update operation, but also the time it took to complete the operation and receive a 'successfully completed' acknowledgement. The read and update latency metric descriptions are the same as Test 3 above and the results are based on an average of three test runs. The results are in 'µs' and the lower result is better.



Database Workloads: 95% Read / 5% Update Latency Tests *(in microseconds)*	PCIe® 5.0 Configuration *(KIOXIA CD8P-R / Dell™ PowerEdge™ R770)*	Previous Generation Configuration *(KIOXIA CD6-R / Dell PowerEdge R740xd)*	PCIe 5.0 Configuration Gains
Read Latency	113.75	300.82	62%
Read Latency: 95th Percentile	228.00	533.00	57%
Read Latency: 99th Percentile	290.00	673.00	56%
Update Latency	64.86	225.91	71%
Update Latency: 95th Percentile	124.00	394.00	68%
Update Latency: 99th Percentile	188.00	516.00	63%

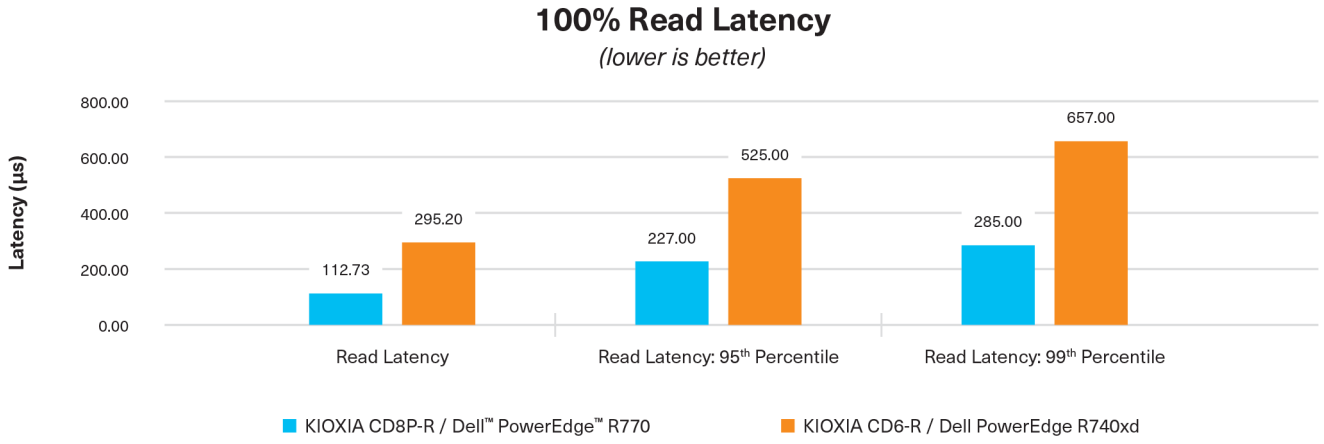
95% read / 5% update latency results

Analysis:

The test results show that the PCIe 5.0 configuration completed read and update operations as measured by each of the six metrics in less time than the previous generation configuration. One example of an application use case that benefits from this performance are those that need to tag and label records stored in a database so that a filtering system can be used to sort through them. Applications that utilize these filters show increased performance when reading and sorting through records, then updating a page with relevant records.

Test 5: 100% Read Latency

This test measured the time it took in microseconds (µs) to perform an individual read operation within the 100% Read workload. It includes the average time it took for the YCSB workload generator to not only issue the read operation, but also the time it took to complete the operation and receive a 'successfully completed' acknowledgement. The read latency metric descriptions are the same as presented in Test 3 and based on an average of three test runs. The results are in 'µs' and the lower result is better.



Database Workloads: 100% Read Latency Tests *(in microseconds)*	PCIe® 5.0 Configuration *(KIOXIA CD8P-R / Dell™ PowerEdge™ R770)*	Previous Generation Configuration *(KIOXIA CD6-R / Dell PowerEdge R740xd)*	PCIe 5.0 Configuration Gains
Read Latency	112.73	295.20	61%
Read Latency: 95th Percentile	227.00	525.00	56%
Read Latency: 99th Percentile	285.00	657.00	56%

100% read latency results

Analysis:

The test results show that the PCIe 5.0 configuration completed read operations as measured by each of the three metrics in less time than the previous generation configuration. This performance can replace memcached[4] distributed caching systems where the low latency cache layer requires persistence. High performance workloads, such as data analytics, represent an application use case that can benefit from low 100% read latency.

Test 6: 95% Read / 5% Insert Latency

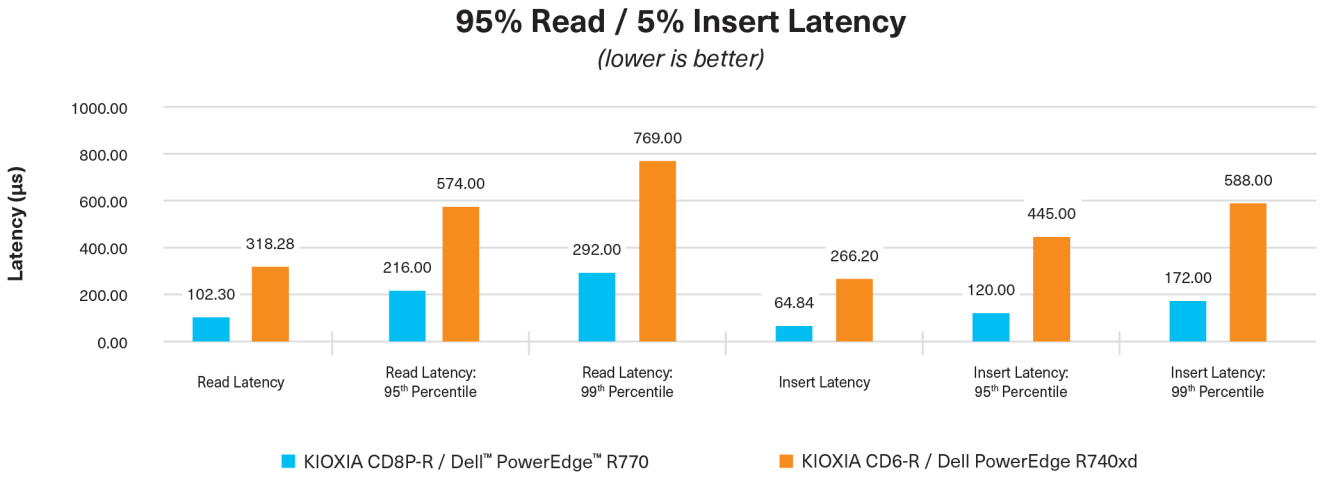
This test measured the time it took in microseconds (µs) to perform an individual read or update operation within the 95% Read / 5% Insert workload. It includes the average time it took for the YCSB workload generator to not only issue the read or insert operation, but also the time it took to complete the operation and receive a 'successfully completed' acknowledgement. The read latency metric descriptions are the same as presented in Test 3. The insert latency metrics included:

- *Average Insert Latency:*
  The average time it took to perform a database insert operation.

- *Average Insert Latency at the 95th Percentile:*
  The average time it took to perform a database insert operation at the 95th percentile (or where 95% of the operations had this insert latency or lower).

- *Average Insert Latency at the 99th Percentile:*
  The average time it took to perform a database insert operation at the 99th percentile (or where 99% of the operations had this insert latency or lower).

The results are in 'µs' and based on an average of three test runs. The lower result is better.



Database Workloads: 95% Read / 5% Update Latency Tests *(in microseconds)*	**PCIe® 5.0 Configuration** *(KIOXIA CD8P-R / Dell™ PowerEdge™ R770)*	**Previous Generation Configuration** *(KIOXIA CD6-R / Dell PowerEdge R740xd)*	**PCIe 5.0 Configuration Gains**
Read Latency	102.30	318.28	67%
Read Latency: 95th Percentile	216.00	574.00	62%
Read Latency: 99th Percentile	292.00	769.00	62%
Update Latency	64.84	266.20	75%
Update Latency: 95th Percentile	120.00	445.00	73%
Update Latency: 99th Percentile	172.00	588.00	70%

95% read / 5% insert latency results

Analysis:
The test results show that the PCIe 5.0 configuration completed more read and insert operations as measured by each of the six metrics in less time than the previous generation configuration. One example of an application use case that benefits from this performance is when users insert new data to a database and other users read and load that content within their application. The database enables users to have the most updated feeds while enabling other users to post new content faster.

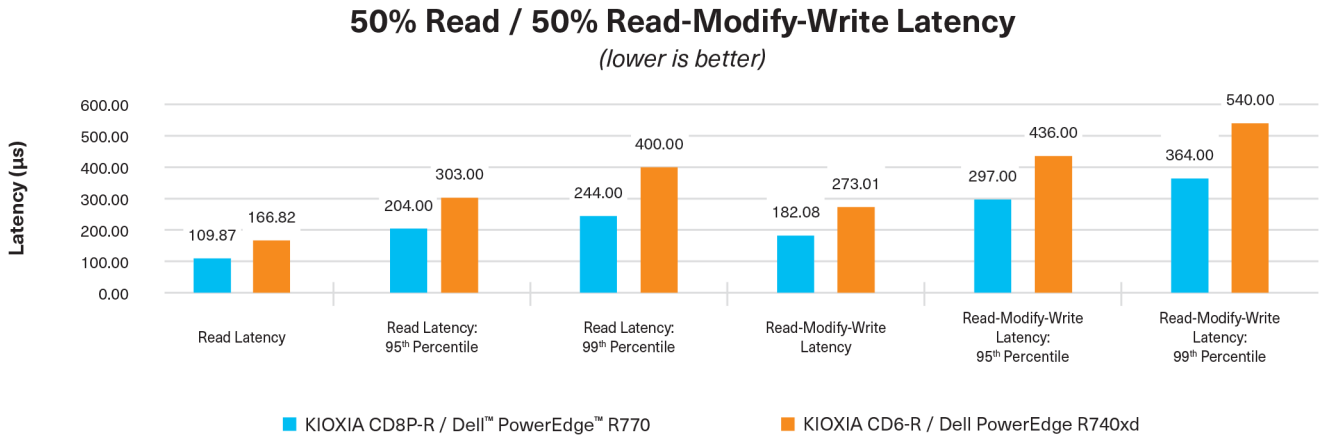
Test 7: 50% Read / 50% Read-Modify-Write Latency
This test measured the time it took in microseconds (µs) to perform an individual read or read-modify-write database operation. It includes the average time it took for the YCSB workload generator to not only issue the read or read-modify-write operation, but also the time it took to complete the operation and receive a 'successfully completed' acknowledgement. The read latency metric descriptions are the same as presented in Test 3. The read-modify-write latency metrics included:

- *Average Read-Modify-Write Latency:*
  The average time it took to perform a database read-modify-write operation.

- *Average Read-Modify-Write Latency at the 95th Percentile:*
  The average time it took to perform a database read-modify-write operation at the 95th percentile (or where 95% of the operations had this read-modify-write latency or lower).

- *Average Read-Modify-Write Latency at the 99th Percentile:*
  The average time it took to perform a database read-modify-write operation at the 99th percentile (or where 99% of the operations had this read-modify-write latency or lower).

The results are in 'µs' and based on an average of three test runs. The lower result is better.



Database Workloads: 95% Read / 5% Update Latency Tests *(in microseconds)*	**PCIe® 5.0 Configuration** *(KIOXIA CD8P-R / Dell™ PowerEdge™ R770)*	**Previous Generation Configuration** *(KIOXIA CD6-R / Dell PowerEdge R740xd)*	**PCIe 5.0 Configuration Gains**
Read Latency	109.87	166.82	34%
Read Latency: 95th Percentile	204.00	303.00	32%
Read Latency: 99th Percentile	244.00	400.00	39%
Update Latency	182.08	273.01	33%
Update Latency: 95th Percentile	297.00	436.00	31%
Update Latency: 99th Percentile	364.00	540.00	32%

50% read / 50% read-modify-write latency results

Analysis:
The test results show that the PCIe 5.0 configuration completed more read and read-modify-write operations as measured by the six metrics in less time than the previous generation configuration. One example of an application use case that benefits from this performance is when concurrent users read and modify records, which are then written into a user activity log. When concurrent users try to read and then modify the same record, potential stalls can occur especially when operations require locks to perform any modification to the record. In the worst case, an operation can be lost if another client updates the record without reading the update operations, as various read-modify-write operations interleave. By increasing the speed at which these operations occur, applications can perform faster, and operations are less likely to be lost.

Summary

Query completion time and throughput are important database performance metrics that determine how fast a system/SSD configuration can complete the database workload and its operations. High performance for these metrics is important to support large concurrent user databases for applications and becomes even more important as the hardware configurations horizontally scale.

Latency performance is another important database metric used to determine the end user experience. Since many NoSQL databases are the backend for web services, faster responses from the database make the application appear snappier. The end user experience is a critical measure as a key revenue component for purchase-based websites.

The Dell™ PowerEdge™ R770 server was able to achieve up to 109% more database throughput (OPS) when compared with the previous generation configuration due to significant performance increases from the CPU and storage backend. The latest generation Intel® Xeon® 6767P CPU delivered an uplift in instructions per cycle and provided more CPU cores (64) for handling many more concurrent queries to the database when compared to previous generations. The key takeaway is that the number of servers needed to run critical database applications can be reduced while still increasing overall performance.

KIOXIA CD8P-R Series Data Center NVMe™ SSDs and Dell PowerEdge R770 servers are advancing database applications by enabling lower runtimes, higher throughputs and lower database operational latencies. IT users who upgrade their prior PCIe® generation systems with the latest PCIe 5.0 interface, can experience significantly improved database speed with faster response times for end users and can reduce the number of servers needed to meet existing workload and service level agreement requirements.

KIOXIA CD8P-R Series Data Center NVMe SSDs

Read-intensive KIOXIA CD8P-R Series Data Center NVMe SSDs include optimizations to support a broad range of scale-out and cloud applications. They use the PCIe 5.0 x4 interface at 32 gigatransfers per second delivering significant performance that includes up to 12,000 megabytes per second (MB/s) for sequential reads, up to 5,500 MB/s for sequential writes, up to 2,000,000 input/output operations per second (IOPS) for random reads and up to 200,000 IOPS for random writes[4].

The KIOXIA CD8P-R Series is available in a 2.5-inch[5] and E3.S form factors. They deliver 1 Drive Write Per Day[6] (DWPD) endurance with storage capacities up to 30.72 TB, making them well-suited for hyperscale data centers and virtualized environments and for big data, IoT and online transaction processing applications, to name a few. Additional KIOXIA CD8P Series SSD specifications and information available here.

KIOXIA CD8P Series SSDs[7]

Appendix A
Database Tests

Server Information		
Server Model	Dell™ PowerEdge™ R770	Dell PowerEdge R740xd
No. of Servers	1	1
No. of CPU Sockets	2	2
CPU	Intel® Xeon® 6767P	Intel Xeon Silver 4214
No. of CPU Cores	64	12
CPU Frequency	2.4 MT/s*	2.2 MT/s*
Memory Type	DDR5	DDR4
Total Memory	2,048 GB[1]	384 GB
Memory Frequency	5,200 MT/s*	2,400 MT/s*

*MT/s = megatransfers per second

Operating System Information		
Operating System	Red Hat® Enterprise Linux®	Red Hat Enterprise Linux
Version	Linux® release 9.5 (Plow)	Linux release 9.5 (Plow)
Kernel	5.14.0-503.35.1.el9_5.x86_64	5.14.0-503.35.1.el9_5.x86_64

SSD Information		
Model	KIOXIA CD8P-R Series[8]	KIOXIA CD6-R Series[8]
Interface	PCIe® 5.0 x4	PCIe 4.0 x4
No. of Devices	8	8
Form Factor	E3.S	2.5-inch (U.2)
Capacity	3.84 TB	3.84 TB
Drive Write(s) Per Day	1 (5 years)	1 (5 years)
Active Power	19 W typ.	15 W typ.

Application Information	
Application	Aerospike® Enterprise
Version	8.0.0.5

Load Generator Information	
Software	Yahoo!™ Cloud Serving Benchmark
Version	0.18.0
Record Count	300,000,000
Operation Count	200,000,000

Appendix B
Configuration Set-up / Test Procedures

Configuration Setup

A Dell™ PowerEdge™ R770 server was set up with the Red Hat® Enterprise Linux® operating system.

Eight KIOXIA CD8P-R Series SSDs were installed in the server.

The Aerospike® Enterprise NoSQL database application was installed.

YCSB software was installed and a database of 300,000,000 records was created on the KIOXIA CD8P-R Series SSDs.

Test Procedures

Seven YCSB workloads and associated tests were run where each workload consisted of 200,000,000 operations to complete. The seven workloads included (1) Runtime; (2) Database Throughput; (3) 50% Read / 50% Update Latency; (4) 95% Read / 5% Update Latency; (5) 100% Read Latency; (6) 95% Read / 5% Insert Latency; and (7) 50% Read / 50% Read-Modify-Write Latency.

The test results for each of the seven YCSB workloads were recorded.

A Dell PowerEdge R740xd server was set up with the same Red Hat Enterprise Linux operating system as prior.

Eight KIOXIA CD6-R Series SSDs were installed in the server.

The same seven YCSB workloads and associated tests were run where each workload also consisted of 300,000,000 records and 200,000,000 operations to complete.

The test results for each of the seven YCSB workloads were recorded.

The PCIe® 5.0 configuration results consisting of the Dell PowerEdge R770 server and KIOXIA CD8P Series SSDs were compared with this configuration (see Test Results).

NOTES:

[1] Definition of capacity - KIOXIA Corporation defines a megabyte (MB) as 1,000,000 bytes, a gigabyte (GB) as 1,000,000,000 bytes, a terabyte (TB) as 1,000,000,000,000 bytes and a petabyte (PB) as 1,000,000,000,000,000 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, 1TB = 2^{40} bytes = 1,099,511,627,776 bytes and 1PB = 2^{40} bytes = 1,125,899,906,842,624 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.

[2] A key-value database is a type of non-relational database that uses a simple key-value method to store data as a collection of key-value pairs in which a key serves as a unique identifier.

[3] Actual read and write speed may vary depending on the host device, read and write conditions, and file size.

[4] Memcached is an open source, general-purpose, distributed memory-caching system used to speed up dynamic database-driven websites by caching data and objects in RAM to reduce the number of times a storage device needs to be read.

[5] 2.5-inch indicates the form factor of the SSD and not its physical size.

[6] Drive Write Per Day (DWPD): 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.

[7] The product image shown is a representation of the design model and not an accurate product depiction.

[8] KIOXIA CD8P-R Series SSD and KIOXIA CD6-R Series SSD performance specifications provided by Kioxia Corporation and are accurate as of this publication date but subject to change.

TRADEMARKS:

DISCLAIMERS: