KIOXIA

Performance Brief

Accelerating Relational Database Workloads with KIOXIA CM7 Series PCIe® 5.0 E3.S NVMe™ SSDs and 16G Dell® PowerEdge™ R6625 Servers

Introduction

Relational databases are vital to today's data centers as they store an overwhelming amount of data captured on premises and at the edge of the network. Sales transactions and information relating to customers, vendors, products and financials represent key data.

IT teams need solutions that scale their data center storage platforms to better address large datasets and future growth. As these databases are dependent on fast underlying storage, one way to achieve high performance and scalability is by utilizing servers equipped with enterprise SSDs based on the latest PCIe 5.0 interface and the NVMe 2.0 protocol. The PCIe 5.0 revision can move data through the PCIe interface almost twice as fast when compared with the previous PCIe 4.0 generation. This enables SSDs to deliver input/output (I/O) even faster while requiring fewer servers to achieve the same level of performance.

With the recent availability of Enterprise and Datacenter Standard Form Factor (EDSFF) SSDs, storage performance and total capacity per server can also increase. Servers with EDSFF E3.S slots deployed with E3.S SSDs deliver fast data throughput, fast input/output operations per second (IOPS) performance, low latency, high density and thermally optimized capabilities.

The latest 16G Dell PowerEdge R6625 servers support the PCIe 5.0 interface and the E3.S form factor. They deliver significant performance benefits and an improved system airflow that enhances heat dissipation. This can lead to less thermal throttling and increased lifespans for system components such as CPUs, memory and storage when compared with prior PCIe generations deployed with 2.5-inch[1] form factor SSDs.

This performance brief presents a generational SSD/server performance and power consumption comparison using PostgreSQL® relational database[2] workloads. It compares KIOXIA CM7-R Series E3.S enterprise NVMe SSDs deployed in a 16G Dell PowerEdge R6625 PCIe 5.0 E3.S server with a previous PCIe generation SSD/system configuration.

Five tests were performed on each SSD/server configuration that included database throughput, drive read latency, drive write latency, server power consumption and throughput per watt. The tests included three total runs, and the average of the three runs was calculated and compared with each SSD/server configuration. HammerDB[3] test software was used for both SSD/server configurations, which enabled the TPROC-C[4] online transaction processing (OLTP) workloads to run against the PostgreSQL database.

The test results show that the latest PCIe generation of KIOXIA CM7-R Series SSDs and 16G Dell PowerEdge R6625 server in the test configuration deliver almost twice the relational database transactions using the same amount of power when compared with the previous PCIe generation SSD/server configuration.

The results presented include a brief description of each workload test, a graphical depiction of the test results and an analysis. Appendix A covers the PostgreSQL parameters and operating system tuning parameters. Appendix B covers the hardware and software test configuration. Appendix C covers the configuration set-up and test procedures.

Test Results Snapshot

KIOXIA CM7-R Series SSDs deployed in a 16G Dell PowerEdge R6625 server delivered the following results when compared with a previous PCIe SSD/ server configuration:

Average Database Throughput
(higher is better)
89% Higher

Average Drive Read Latency
(lower is better)
19% Lower

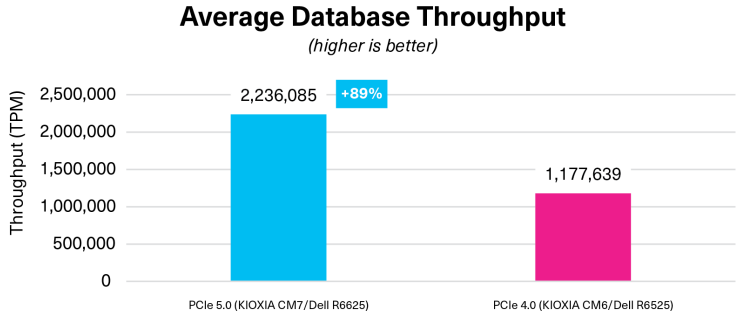
Average Drive Write Latency
(lower is better)
33% Lower

Average Throughput per Watt
(higher is better)
85% Higher

Test Results[5]

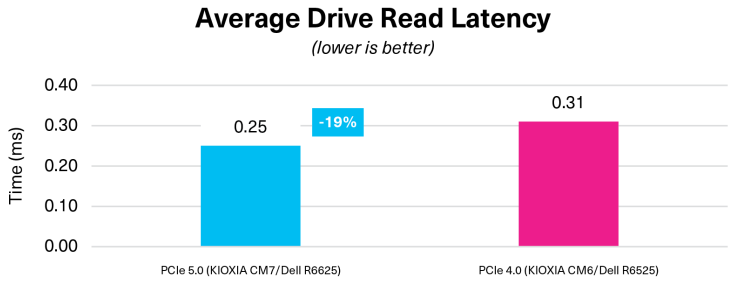
Test 1: Average Database Throughput

This test measured how many transactions in the TPROC-C workload were executed per minute. The HammerDB software, executing the TPROC-C transaction profile, randomly performed new order, payment, order status, delivery and stock level transactions. The benchmark simulated an OLTP environment with a large number of users conducting simple and short transactions (that require sub-second response times and return relatively few records). The following results show the average database throughput from three test runs for each set of drives. The results are in transactions per minute (TPM). The higher result is better.



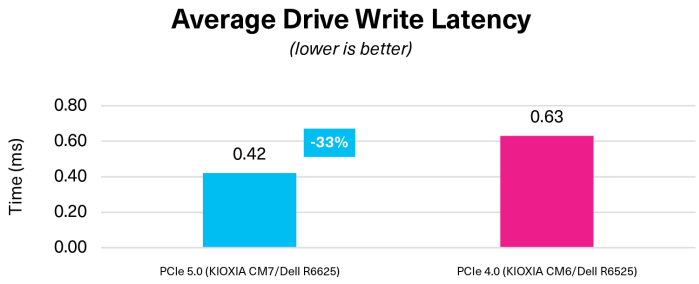
Test 2: Average Drive Read Latency

This test measured the time it took to perform a drive read operation. It included the average time it took to not only issue the read operation, but also the average time it took to complete the operation and receive a 'successfully completed' acknowledgement. The following results show the average read latency from three test runs for each set of drives. The results are in milliseconds (ms). The lower result is better.



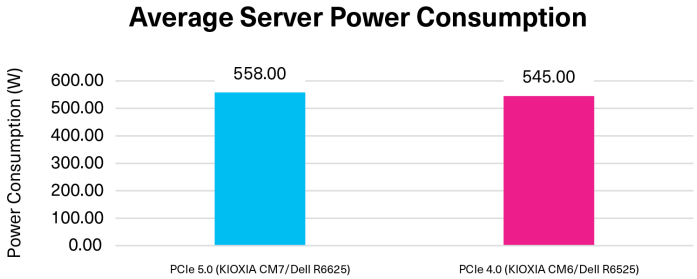
Test 3: Average Drive Write Latency

This test measured the time it took to perform a drive write operation. It included the average time it took to not only issue the write operation, but also the average time it took to complete the operation and receive a 'successfully completed' acknowledgement. The following shows the average write latency from three test runs for each set of drives. The results are in milliseconds (ms). The lower result is better.



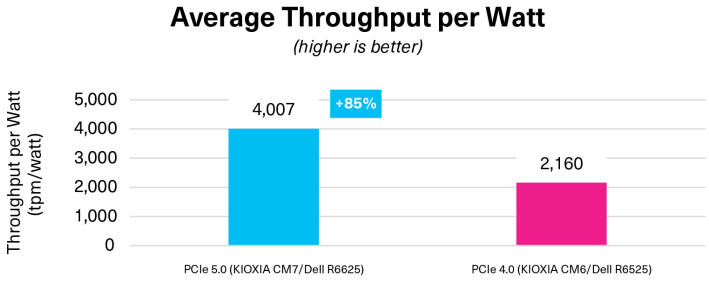
Test 4: Average Server Power Consumption

This test measured the average amount of power drawn by each server system in its entirety including all of the individual components that run from the server's power supply unit (PSU). This included the motherboard, CPU, memory, storage and other server components. The following results were obtained from the Integrated Dell Remote Access Controller (iDRAC). The results are in watts (W).



Although the overall system power draw was slightly higher in the PCIe 5.0 configuration, the solution was able to maintain 89% higher database throughput, 19% lower read latency and 33% lower write latency on average.

Using the average server power consumption results the average database throughput per watt can be easily determined by dividing the average database throughput by the average server power consumption as depicted below. The results are in throughput (transactions per minute) per watt (tpm/watt). The higher result is better.



The PCIe 5.0 configuration was able to deliver 4,007 tpm/watt versus 2,160 tpm/watt delivered by the PCIe 4.0 configuration, nearly doubling database throughput per watt. At the data center level, these results enable administrators to use the same number of servers for nearly double the performance, or converse to this, scale the number of servers to help save on power consumption and total cost of ownership without sacrificing performance.

Analysis

Next generation KIOXIA CM7-R Series PCIe 5.0 E3.S SSDs deployed in Dell PowerEdge R6625 servers show nearly double the database performance when compared with a previous PCI generation SSD/server configuration while lowering SSD latency by performing read/write operations faster. The PCIe 5.0 SSD/server configuration delivered 89% more transactions per minute enabling higher relational database workload densities while reducing the footprint of servers needed to service these workloads.

The KIOXIA CM7-R Series SSD and Dell PowerEdge R6625 server configuration also demonstrated a comparable server power draw when compared with the previous PCI generation SSD/server configuration. Though the active power increased from PCIe 4.0 to PCIe 5.0 by approximately 13 watts, the system was able to process almost twice as many transactions while consuming almost the same amount of power. As such, fewer servers are necessary to achieve the same level of performance without experiencing a power consumption spike.

Summary

This performance brief presented a PCIe generational performance and power consumption comparison of two SSD/server configurations using PostgreSQL relational database workloads. It compared KIOXIA CM7-R Series E3.S enterprise NVMe SSDs deployed in 16G Dell PowerEdge R6625 PCIe 5.0 E3.S servers with a previous PCIe generation SSD/server configuration.

The latest 16G Dell PowerEdge R6625 servers support the PCIe 5.0 interface and the E3.S form factor. They deliver significant performance benefits and an improved system airflow that enhances heat dissipation and can lead to less thermal throttling and increased lifespans for system components such as CPUs, memory and storage when compared with prior PCIe generations deployed with 2.5-inch form factor SSDs.

The test results indicate that the PCIe 5.0 configuration that included the KIOXIA CM7-R Series SSDs and 16G Dell PowerEdge R6625 server delivered almost twice the relational database transactions using approximately the same amount of power when compared with the previous PCIe generation SSD/server configuration.

Products Tested:

KIOXIA CM7 Series SSD Product Info

The latest generation KIOXIA CM7 Series enterprise NVMe SSDs support Enterprise and Datacenter Standard Form Factor (EDSFF) E3.S and 2.5-inch form factors and are compliant with the NVMe 2.0 and PCIe 5.0 specifications. These SSDs are available in two configurations: CM7-R Series for read intensive applications (1 DWPD[6], up to 30.72 TB[7] capacities) and CM7-V Series for higher endurance mixed use applications (3 DWPD, up to 12.80 TB capacities). Additional features include a dual-port design for high availability applications, Flash Die Failure Protection and security options[8]. Additional KIOXIA CM7 Series SSD information is available here.

KIOXIA CM7 Series SSD[9]

Dell PowerEdge 6625 Rack Server Product Info

Specifications: https://www.delltechnologies.com/asset/en-us/products/servers/technical-support/poweredge-r6625-spec-sheet.pdf.

Appendix A

PostgreSQL Parameters/Operating System Tuning Parameters

The PostgreSQL parameters used for this comparison include:

PostgreSQL Parameters (version 8.0.31-0)

For Ubuntu 22.04.2 operating system

shared_buffers	64 GB
temp_buffers	4 GB
work_mem	4 GB
maintenance_work_men	512 MB[6]
max_files_per_process	65,535
effective_io_concurrency	32
wal_buffers	512 MB
max_wal_size	200 GB
min_wal_size	80 MB

Additional tuning parameters performed on the operating system (OS) to optimize system performance were made to /etc/sysctl.conf files and /etc/security/limits.conf files. The /etc/sysctl.conf files override OS default kernel parameter values while the /etc/security/limits.conf files allow resource limits to be set. These tuning parameters include:

/etc/sysctl.conf file changes:

Parameter	Value
vm.swappiness	0
kernel.sem	250 32000 100 128
fs.file-max	6815744
net.core.rmem_default	262144
net.core.rmem_max	4194304
net.core.wmem_default	262144
net.core.wmem_max	1048576
fs.aio-max-nr	1048576
vm.nr_hugepages	35000

/etc/security/limits.conf file changes:

User	Values		
*	soft	nproc	65535
*	hard	nproc	65535
*	soft	nofile	65535
*	hard	nofile	65535
root	soft	nproc	65535
root	hard	nproc	65535
root	soft	nofile	65535
root	hard	nofile	65535
postgres	soft	memlock	100000000
postgres	hard	memlock	100000000

Appendix B

Configuration Set-up/Test Procedures

Server Information		
Server Model	Dell PowerEdge R6625	Dell PowerEdge R6525
No. of Servers	3	3

CPU Information		
CPU Model	AMD EPYC™ 9334	AMD EPYC 7352
No. of Sockets	2	2
No. of Cores	32	24

Memory Information		
Memory Type	DDR5	DDR4
Memory Speed *(in megatransfers per second)*	4,800 MT/s	3,200 MT/s
Memory Size	384 GB[7]	128 GB

SSD Information		
SSD Model	KIOXIA CM7-R Series	KIOXIA CM6-R Series
Form Factor	E3.S	2.5-inch (U.3)
Interface	PCIe 5.0 x4	PCIe 4.0 x4
Interface Speed *(in gigatransfers per second)*	128 GT/s	64 GT/s
No. of SSDs	4	4
SSD Capacity	3.84 TB	3.84 TB
Drive Writes per Day	1	1
Active Power	up to 24 watts	up to 19 watts

Operating System Information		
Operating System (OS)	Ubuntu®	Ubunto
OS Version	22.04.2	22.04.2
Kernel	5.15.0-76-generic	5.15.0-76-generic
RAID	RAID 5[10]	RAID 5
RAID Version	mdadm 4.2	mdadm 4.2

Test Software Information		
Test Software Model	HammerDB	HammerDB
Benchmark	TPROC-C	TPROC-C
Version	4.8	4.8
No. of Virtual Users	128	128

Appendix C

Configuration Set-up/Test Procedures

Configuration Set-up

A Dell PowerEdge 6625 Rack Server was set-up with the Ubuntu 22.04.2 operating system.

Additional OS level parameters were adjusted to help increase system performance (to adjust these parameters, refer to Appendix A).

The system was rebooted.

Four 3.84 TB KIOXIA CM7 Series SSDs were placed in a RAID 5 set (via mdadm) to hold the PostgreSQL database in the server.

An XFS® file system was placed on top of the RAID 5 set and was mounted with noatime[11] and discard[12] flags.

PostgreSQL relational database was installed in the server and the service was started.

HammerDB test software was installed on the server for the KIOXIA CM7 Series SSDs, enabling the TPROC-C online transaction processing (OLTP) workloads to run against the PostgreSQL database.

Test Procedures

The following metrics were recorded when the TPROC-C workload was run against each configuration:

- Average Database Throughput
- Average Drive Read Latency
- Average Drive Write Latency
- Average Server Power Consumption

For each individual metric, three total runs were performed and the average of the three runs were calculated and compared with each configuration.

NOTES:

[1] 2.5-inch indicates the form factor of the SSD and not the drive's physical size.

[2] PostgreSQL is a powerful, open source object-relational database system with over 35 years of active development and a reputation for reliability, feature robustness and performance.

[3] 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.

[4] TPROC-C is the OLTP workload implemented in HammerDB derived from the TPC-C™ specification with modification to make running HammerDB straightforward and cost-effective on any of the supported database environments. The HammerDB TPROC-C workload is an open source workload derived from the TPC-C Benchmark Standard and as such is not comparable to published TPC-C results, as the results comply with a subset rather than the full TPC-C Benchmark Standard. TPROC-C means Transaction Processing Benchmark derived from the TPC "C" specification.

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

[6] DWPD or Drive Write(s) Per Day (DWPD) 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] 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.

[8] Optional security feature compliant drives are not available in all countries due to export and local regulations.

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

[10] RAID 5 is a redundant array of independent disks configuration that uses disk striping with parity - Data and parity are striped evenly across all of the disks, so no single disk is a bottleneck.

[11] The noatime option turns off access time recording so that the file system will ignore access time updates on files. If the file system is used for database workloads, specifying noatime can reduce writes to the file system.

[12] The discard option allows the file system to inform the underlying block device to issue a TRIM command when blocks are longer used. KIOXIA makes no warranties regarding the test results and performance can vary due to system configuration usage and other factors.

TRADEMARKS:

AMD EPYC and combinations thereof are trademark of Advanced Micro Devices, Inc. Dell and PowerEdge are registered trademarks or trademarks of Dell Inc. NVMe is a registered or unregistered trademark of NVM Express, Inc. in the United States and other countries. PCIe is a registered trademark of PCI-SIG. PostgreSQL is a registered trademark of the PostgreSQL Community Association of Canada. TPC-C is a trademark of the Transaction Processing Performance Council. Ubuntu is a registered trademark of Canonical Ltd. XFS is a registered trademark of Silicon Graphics International Corp. or its subsidiaries in the United States and/or other countries. All other company names, product names and service names may be trademarks or registered trademarks of their respective companies.

DISCLAIMERS:

KIOXIA America, Inc. may make changes to specifications and product descriptions at any time. The information presented in this performance brief is for informational purposes only and may contain technical inaccuracies, omissions and typographical errors. Any performance tests and ratings are measured using systems that reflect the approximate performance of KIOXIA America, Inc. products as measured by those tests. Any differences in software or hardware configuration may affect actual performance, and KIOXIA America, Inc. does not control the design or implementation of third party benchmarks or websites referenced in this document. The information contained herein is subject to change and may render inaccuracies for many reasons, including but not limited to any changes in product and/or roadmap, component and hardware revision changes, new model and/or product releases, software changes, firmware changes, or the like. KIOXIA America, Inc. assumes no obligation to update or otherwise correct or revise this information.

KIOXIA America, Inc. makes no representations or warranties with respect to the contents herein and assumes no responsibility for any inaccuracies, errors or omissions that may appear in this information.

KIOXIA America, Inc. specifically disclaims any implied warranties of merchantability or fitness for any particular purpose. In no event will KIOXIA America, Inc. be liable to any person for any direct, indirect, special or other consequential damages arising from the use of any information contained herein, even if KIOXIA America, Inc. are advised of the possibility of such damages.