Application Brief

New Storage Class Memory Solution Accelerates Non-Relational Database Performance

Aerospike® NoSQL Databases Using KIOXIA FL6 Series Enterprise NVMe® SCM SSDs Deliver Heightened Performance Gains versus TLC SSDs

With the demands of modern data center workloads, new forms of data storage have emerged including scale-out storage, object-based storage and web-based file storage. The combination of traditional and modern data storage approaches, coupled with high data generation that is growing daily, has created new challenges to administrators tasked with ensuring high performance and speedy response times for their enterprise applications.

Historically, a lot of data has been stored in relational databases which rely on data that is represented in tables comprised of rows for storing records and columns, and hold a value for a given attribute (Figure 1). However, many data types do not fit into relational databases, and often there are performance and capacity requirements that relational databases struggle to meet. Traditional database querying and data inserting operations are not as performant in the relational database model, which is no longer sufficient at ensuring large scale high performance past a single server node.

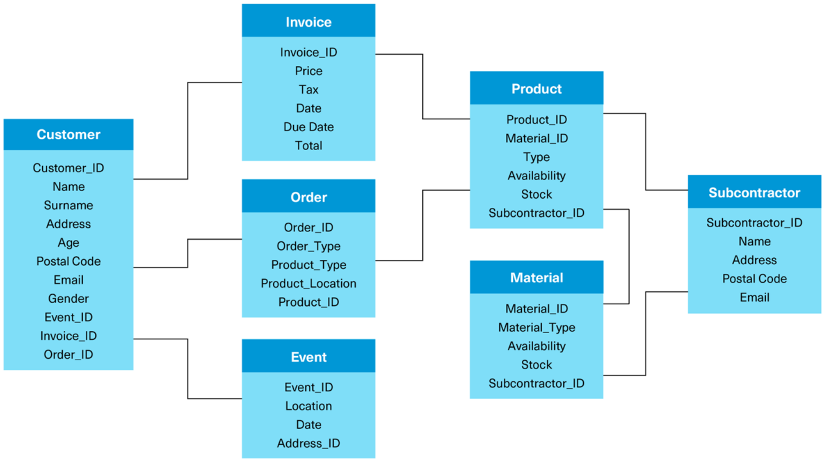


Figure 1: Depicts a sample snapshot of a relational database application

Relational databases are also not well-designed for changes in data and table structures. Data modeling must be performed in advance to ensure that the database tables can accurately store all of the data it needs. Any data changes that occur which require updates to the relational database structure can be time-consuming and resource-intensive to implement. Though advanced data modeling techniques can sometimes remediate these scenarios, the lack of flexibility associated with relational databases hamper these efforts.

To better address large-sized and evolving datasets, IT teams are scaling storage and the databases contained on them. Non-relational NoSQL[1] databases are able to take advantage of many features such as being able to handle large volumes of data at high speed, store many different types of data, update schemas and fields easily, and support a constantly evolving data model. Database sharding, which can break up these large datasets into smaller chunks, allows for scale-out architectures with high performance and scalability. Though there are many NoSQL databases that handle data at scale, most are not optimized specifically for fast PCIe® enterprise NVMe SSDs.

Aerospike is a NoSQL key-value database[2] capable of delivering very fast run time performance for all-sized read and write workloads as its architecture is flash-optimized. It is designed for low latency storage media and able to support PCIe NVMe SSDs with the ability to scale up to petabytes[3] (PB) of data. Aerospike features direct device access that enables high throughput and low latency to be delivered directly on the storage device itself. In order to take advantage of this capability, the underlying SSDs must be fast enough to support this level of performance.

KIOXIA FL6 Series enterprise NVMe SSDs are Storage Class Memory (SCM) solutions that deliver a level of performance well-suited for Aerospike applications. These solutions bridge the gap between DRAM system memory and mainstream Triple-Level Cell (TLC) flash memory SSDs, making them a good fit for latency-sensitive use cases such as databases. The FL6 Series incorporates the XL-FLASH™ BiCS FLASH™ 3D flash memory technology that delivers very low read, update, read-modify-write, and insert latencies, which in turn enables fast application response times. When configured with an Aerospike database, FL6 Series SCM SSDs demonstrate decreased latency and heightened throughput performance with no additional use of system memory required to achieve the performance gains and high CPU utilization.

This application brief presents test results of varying operations that were executed using Aerospike database workloads within an SCM SSD configuration when compared to a leading and currently shipping TLC-based PCIe 4.0 drive.

Test Methodology

Six NoSQL database workloads below were driven by the Aerospike database with synthetic tests run through Yahoo! Cloud Serving Benchmark (YCSB) software that served as the database load generator:

Workload	Workload Mix	Use-Case Example
A	50% Read / 50% Update	*Recent actions that are recorded and updated such as cookies in a web browser.*
B	95% Read / 5% Update	*Image tags where users can label images with certain tags to make them more searchable. Creating a tag is the update operation, while searching for images via a tag is the read operation.*
C	100% Read	*User profile caches where profiles are stored at remote locations.*
D	95% Read / 5% Insert	*User status updates where an application might have users posting new content driven by insert operations, while other users want to see that content, which is driven by read operations.*
E	100% Insert	*Multiple files or objects uploaded by a user to an endpoint database.*
F	50% Read / 50% Read-Modify-Write	*User databases where records are read and modified by the user or used to record user activity (can be in the form of user history).*

The synthetic benchmark tests were conducted by KIOXIA in a laboratory environment where eight key metrics were recorded:

Metric	Description
Run Time	*Represents the total time required in minutes to complete the database workload assuming that both drives tested are running the same number of total operations. A lower time is better and represents that the underlying drive was faster at completing the database workload.*
Operations Throughput	*Represents the number of operations per second (ops/s) a system can complete on average. Relating to database throughput, it also measures how quickly a given server solution can process incoming database queries. This is used to discern if the number of incoming queries is much higher than the achievable database throughput. If this occurs, the server can overload and create longer waiting times per query that can negatively impact application performance and the user experience. This metric is especially important when a mix of operations from a large group of users require simultaneous processing at sub-second response times.*
Average Read Latency	*Represents the time it takes in microseconds (µs) to perform a read database operation. It includes the average time it takes for the YCSB workload generator to not only issue the read operation, but also the time it takes for the operation to be successfully completed.*
Average Update Latency	*Represents the time it takes in µs to perform an update database operation. It includes the average time it takes for the YCSB workload generator to not only issue the update operation, but also the time it takes for the operation to be successfully completed.*
Average Read-Modify-Write Latency	*Represents the time it takes in µs to perform a read-modify-write database operation. It includes the average time it takes for the YCSB workload generator to not only issue the read-modify-write operation, but also the time it takes for the operation to be successfully completed.*
Average Insert Latency	*Represents the time it takes in µs to perform an insert database operation. It includes the average time it takes for the YCSB workload generator to not only issue the insert operation, but also the time it takes for the operation to be successfully completed.*
Maximum CPU Utilization	*Represents a percentage of the maximum CPU cycles used for a given workload and measured to ensure that the server CPUs were not incurring extra processing for database workloads. Low utilization means that the CPUs were not being used efficiently, which could result in an underutilization of server capabilities and stranded compute resources. High CPU utilization is better and indicates that the CPU is using all of its resources to gain higher performance.*
Memory Utilization	*Represents a percentage of memory utilization that was used by the system for a given workload. Aerospike provides direct device access to an SSD enabling it to be used for both primary index storage and data storage. Low memory utilization is expected as the test configuration did not implement a full DRAM storage model or a hybrid storage model in which both memory and storage were used to retain data. Both primary index storage and data storage was located on the SSD.*

Average latencies (read, update, read-modify-write and insert) are important metrics as they can positively or negatively affect database performance and application response times, which in turn translates into the user database experience.

The benchmark consisted of the six workloads (A through F) and was performed on the FL6 Series SCM SSD. The results for the eight metrics listed above were recorded. This identical process was then performed on the TLC-based PCIe 4.0 drive with the results also recorded. A comparison of the results and analysis are included.

Test Equipment

The hardware and software equipment used to perform the benchmark test of the six workloads:

- **Supermicro® AS-2114S-WN24RT Server:**
  One (1) socket server with one (1) AMD EPYC™ 7702P processor featuring 64 processing cores, 2.00 GHz frequency, and 512 gigabytes[3] (GB) of DDR4 DRAM
- **Operating System:** Ubuntu® v20.04.3 (Kernel 5.4.0-89-generic)
- **Application:** Aerospike v5.7.0.8:
  Database size = 375.93 GB
  Number of Connections / Threads = 128
  Number of Record = 300M
  Number of Operations = 2B
- **Test Software:** Synthetic tests run through YCSB software (version 0.17.0)
- **Storage Devices (Table 1):**
  One (1) KIOXIA FL6 Series enterprise NVMe SCM SSD with 3.2 terabytes[3] (TB) capacity
  One (1) Vendor A TLC PCIe 4.0 enterprise NVMe SSD with 3.2 TB capacity

Specifications	FL6 Series	Vendor A
Form Factor	U.3	U.2
Interface	PCIe 4.0	PCIe 4.0
Capacity	3.2 TB	3.2 TB
NAND Flash Type	XL-FLASH	V-NAND
Endurance Rating	60 Drive Writes Per Day[4] (DWPD)	3 DWPD
Endurance Time Frame	5 years	5 years
Power	~25W	~25W

Table 1: SSD specifications and set-up parameters

Equipment Set-up

The Supermicro AS-2114S-WN24RT server was configured with the Ubuntu v20.04.3 operating system and YCSB v0.17.0 test software. The YCSB software was used to create a database on the FL6 Series SSD and the Vendor A TLC drive to run workloads A through F as described earlier against the Aerospike database. The A through F workloads were selected as they represent common workload use cases in databases.

Test Procedures

The benchmark test was performed on the FL6 Series SSD consisting of workloads A through F. The results included eight metrics: (1) Run Time; (2) Operations Throughput; (3) Average Read Latency; (4) Average Update Latency; (5) Average Read-Modify-Write Latency; (6) Average Insert Latency; (7) Maximum CPU Utilization; and (8) Memory Utilization. The results of the eight metrics were recorded.

This identical process covering the benchmark test and eight metrics was then performed on the Vendor A TLC PCIe 4.0 drive with the results also recorded. A comparison of each metric covering Workloads A thru F is presented[5], as well as an analysis for each test result.

Test Results

Summary of the metrics required for each workload test:

Metric	Workload					
	A	B	C	D	E	F
Run Time	X	X	X	X	X	X
Operations Throughput	X	X	X	X	X	X
Average Read Latency	X	X	X	X		X
Average Update Latency	X	X				X
Average Read-Modify-Write Latency						X
Average Insert Latency				X	X	
Maximum CPU Utilization	X	X	X	X	X	X
Memory Used	X	X	X	X	X	X

Test results for each workload:

Workload A: 50% Read / 50% Update				
Tested Metrics:		**FL6 Series**	**Vendor A**	**FL6 Gains**
Run Time	*(lower result is better)*	30.65 min.	33.66 min.	~8%
Operations Throughput	*(higher result is better)*	1,087,472 ops/s	990,315 ops/s	~9%
Average Read Latency	*(lower result is better)*	100.83 µs	147.24 µs	~31%
Average Update Latency	*(lower result is better)*	101.37 µs	125.53 µs	~19%
Max. CPU Utilization	*(higher result is better)*	77.00%	76.50%	~0.65%
Memory Used	*(lower result is better)*	6.04%	6.04%	0%

Workload B: 95% Read / 5% Update				
Tested Metrics:		**FL6 Series**	**Vendor A**	**FL6 Gains**
Run Time	*(lower result is better)*	28.98 min.	37.64 min.	~23%
Operations Throughput	*(higher result is better)*	1,150,170 ops/s	885,575 ops/s	~29%
Average Read Latency	*(lower result is better)*	107.48 µs	142.84 µs	~24%
Average Update Latency	*(lower result is better)*	79.80 µs	81.33 µs	~1%
Max. CPU Utilization	*(higher result is better)*	79.80%	75.80%	~5%
Memory Used	*(lower result is better)*	6.23%	6.23%	0%

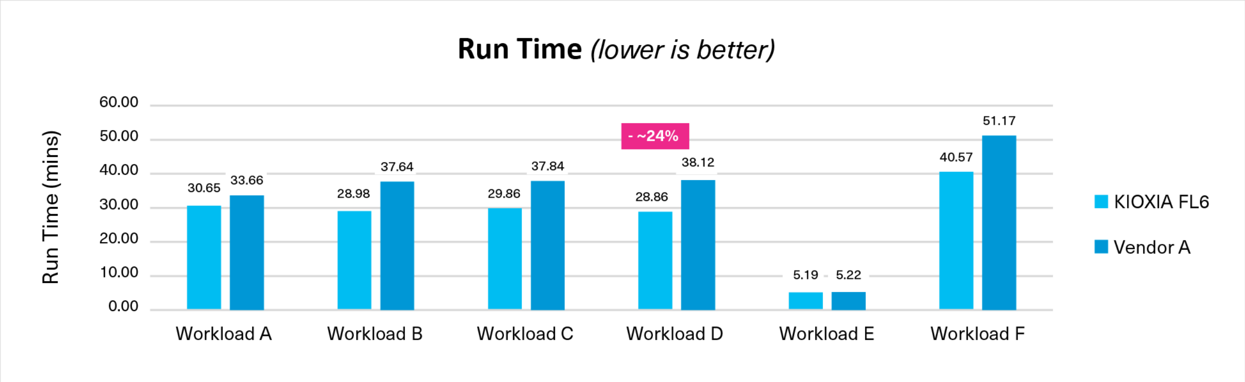
Workload C: 100% Read				
Tested Metrics:		**FL6 Series**	**Vendor A**	**FL6 Gains**
Run Time	*(lower result is better)*	29.86 min.	37.84 min.	~21%
Operations Throughput	*(higher result is better)*	1,116,311 ops/s	881,013 ops/s	~26%
Average Read Latency	*(lower result is better)*	110.19 µs	141.09 µs	~21%
Max. CPU Utilization	*(higher result is better)*	79.20%	75.60%	~4%
Memory Used	*(lower result is better)*	6.23%	6.23%	0%

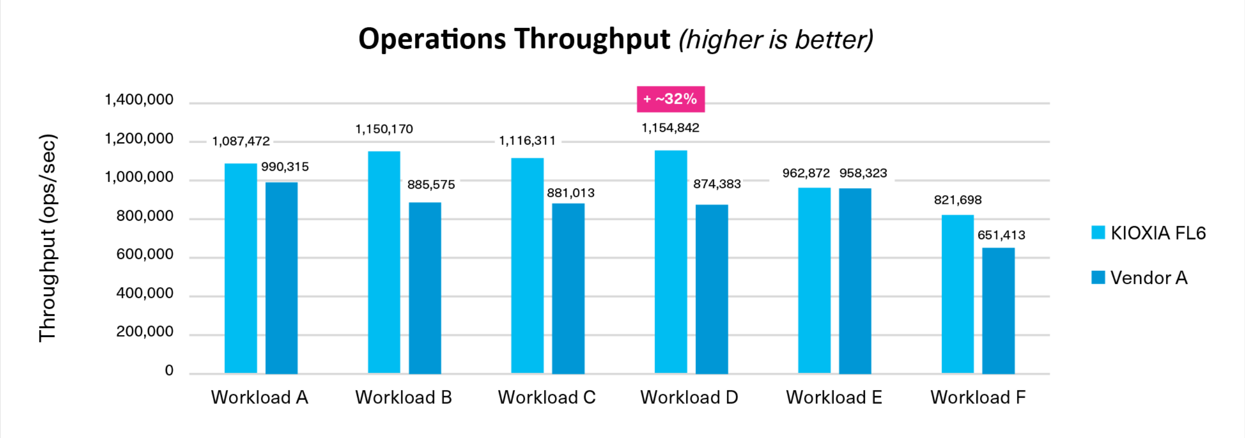
Workload D: 95% Read / 5% Insert				
Tested Metrics:		**FL6 Series**	**Vendor A**	**FL6 Gains**
Run Time	*(lower result is better)*	28.86 min.	38.12 min.	~24%
Operations Throughput	*(higher result is better)*	1,154,842 ops/s	874,383 ops/s	~32%
Average Read Latency	*(lower result is better)*	106.68 µs	144.48 µs	~26%
Average Update Latency	*(lower result is better)*	90.83 µs	95.70 µs	~5%
Max. CPU Utilization	*(higher result is better)*	79.90%	75.70%	~5%
Memory Used	*(lower result is better)*	6.02%	6.02%	0%

Workload E: 100% Insert				
Tested Metrics:		**FL6 Series**	**Vendor A**	**FL6 Gains**
Run Time	*(lower result is better)*	5.19 min.	5.22 min.	~1%
Operations Throughput	*(higher result is better)*	962,872 ops/s	958,323 ops/s	~1%
Average Read Latency	*(lower result is better)*	128.66 µs	129.40 µs	~1%
Max. CPU Utilization	*(higher result is better)*	87.00%	87.00%	0%
Memory Used	*(lower result is better)*	4.29%	4.29%	0%

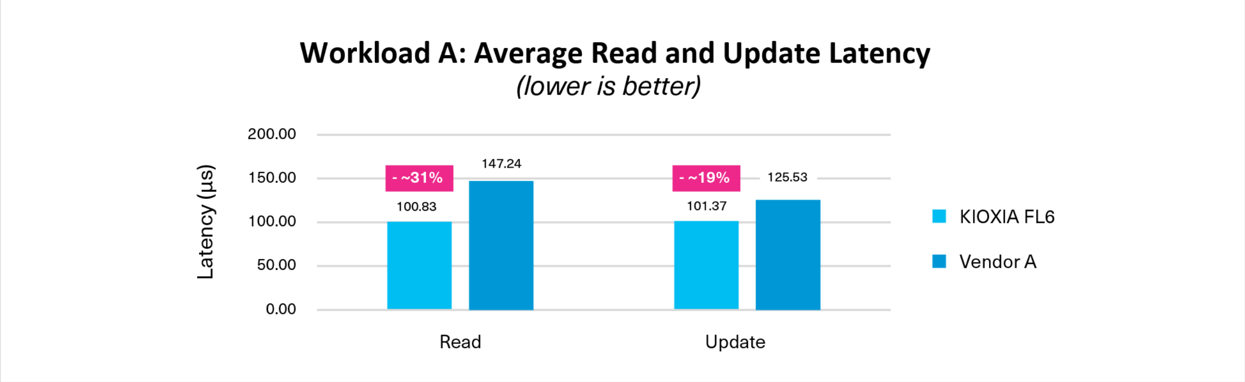
Workload F: 50% Read / 50% Read-Modify-Write				
Tested Metrics:		**FL6 Series**	**Vendor A**	**FL6 Gains**
Run Time	*(lower result is better)*	40.57 min.	51.17 min.	~20%
Operations Throughput	*(higher result is better)*	821,698 ops/s	651,413 ops/s	~26%
Average Read Latency	*(lower result is better)*	104.86 µs	145.34 µs	~27%
Average Read-Modify-Write Latency	*(lower result is better)*	196.10 µs	234.92 µs	~16%
Average Update Latency	*(lower result is better)*	86.04 µs	87.56 µs	~1%
Max. CPU Utilization	*(higher result is better)*	80.00%	77.50%	~3%
Memory Used	*(lower result is better)*	6.23%	6.23%	0%

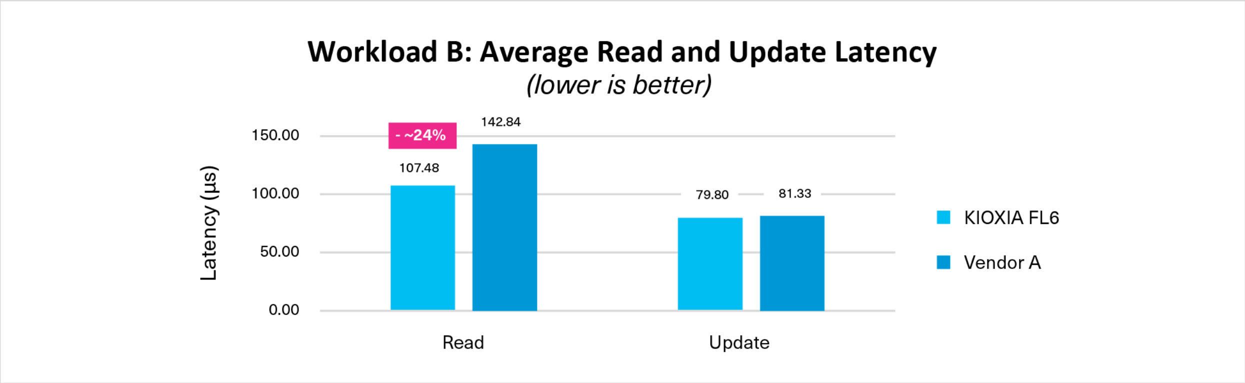
Test results for Run Time and Operations Throughput for the six workloads:

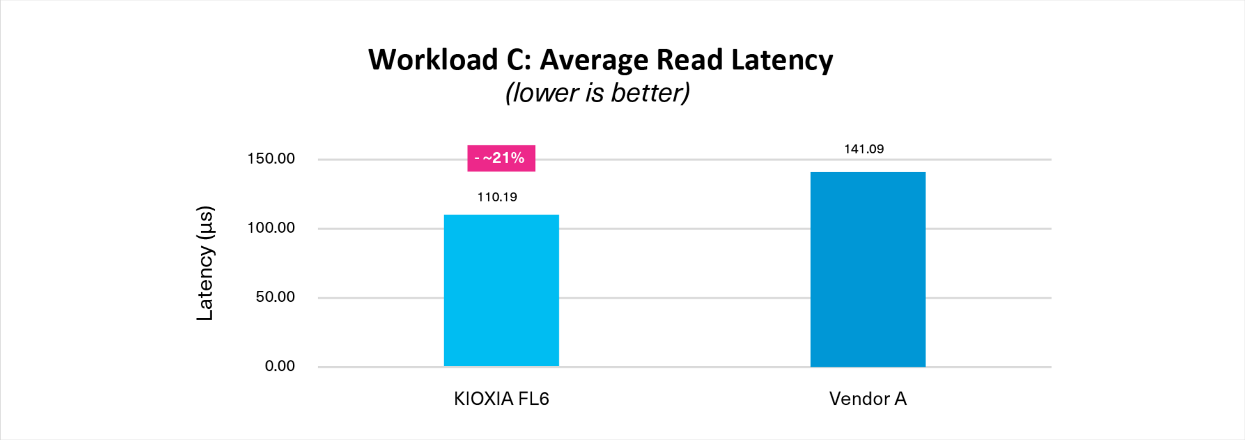


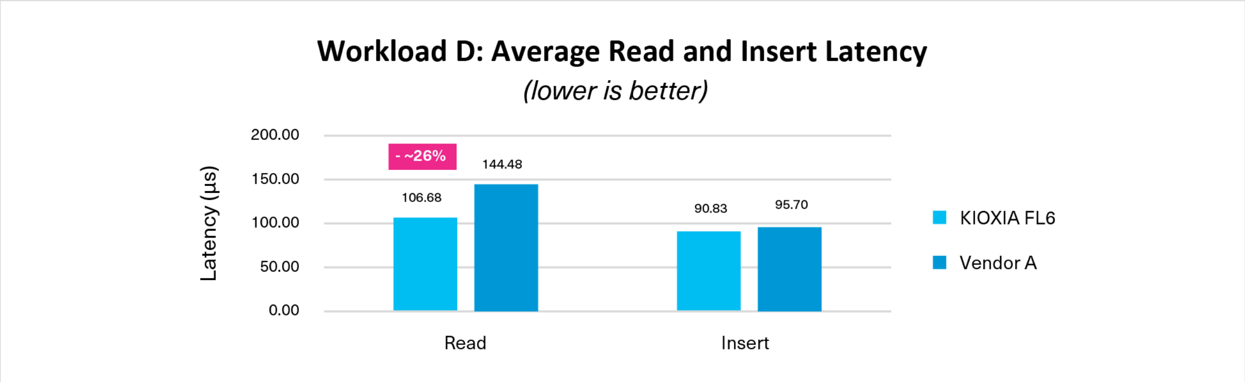


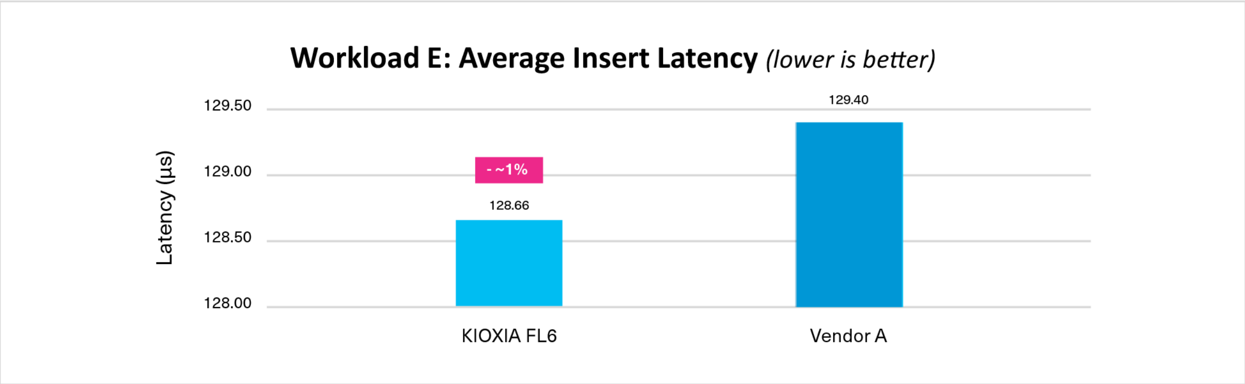
Latency test results for the six workloads:

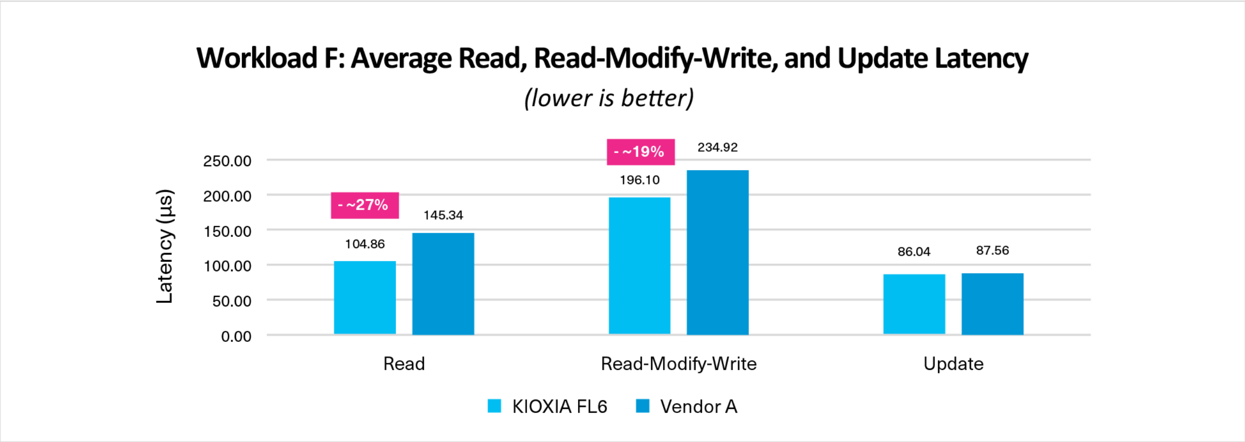



Workload B: Average Read and Update Latency
(lower is better)
Latency (µs)
150.00
100.00
50.00
0.00
- ~24%
107.48
142.84
79.80
81.33
Read
Update
KIOXIA FL6
Vendor A

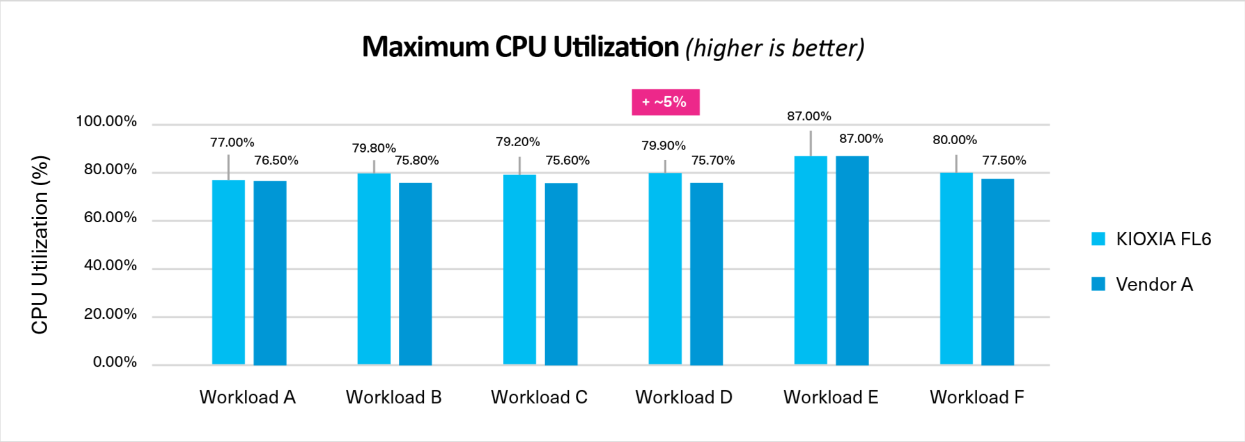

Workload C: Average Read Latency
(lower is better)
Latency (µs)
150.00
100.00
50.00
0.00
- ~21%
110.19
141.09
KIOXIA FL6
Vendor A

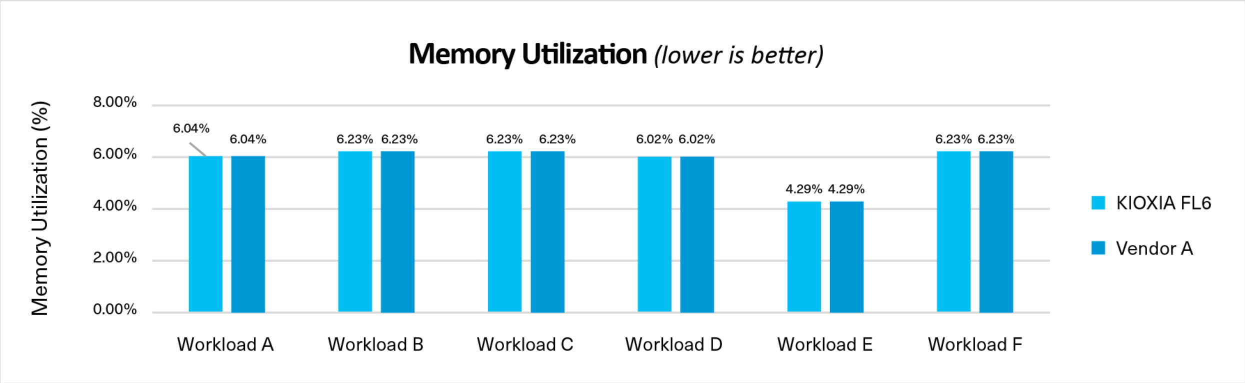

Workload D: Average Read and Insert Latency
(lower is better)
Latency (µs)
200.00
150.00
100.00
50.00
0.00
- ~26%
106.68
144.48
90.83
95.70
Read
Insert
KIOXIA FL6
Vendor A


Workload E: Average Insert Latency (lower is better)
Latency (µs)
129.50
129.00
128.50
128.00
- ~1%
128.66
129.40
KIOXIA FL6
Vendor A



Maximum CPU Utilization and Memory Utilization test results for the six workloads:





Analysis

The test results of the eight recorded operations for each of six workloads include the following analysis:

Operation	Test Result Analysis
Run Time	*The results demonstrate that the SCM solution was faster at completing the same number of total operations for the six workloads tested when compared to the TLC-based PCIe 4.0 drive. In some cases, it was more than 20% faster (24% was the fastest).*
Operations Throughput	*The results demonstrate that the SCM solution completed more operations per second for the six workloads tested when compared to the TLC-based PCIe 4.0 drive. In some cases, the performance was as high as 32% faster.*
Average Read Latency	*The results demonstrate that the SCM solution was 21% to 31% faster at performing read database operations for the five workloads that included read operations tested when compared to the TLC-based PCIe 4.0 drive. Faster data access positively affects database performance and application response times that translates into a heightened user experience.*
Average Update Latency	*The results demonstrate that the SCM solution was 1% to 19% faster at performing update database operations for the three workloads that included update operations tested when compared to the TLC-based PCIe 4.0 drive. Faster data access positively affects database performance and application response times that translates into a heightened user experience.*
Average Read-Modify-Write Latency	*The results demonstrate that the SCM solution was 16% faster at performing read-modify-write database operations for the one workload that included read-modify-write operations tested when compared to the TLC-based PCIe 4.0 drive. Faster data access positively affects database performance and application response times that translates into a heightened user experience.*
Average Insert Latency	*The results demonstrate that the SCM solution was 1% to 5% faster at performing insert database operations for the two workloads that included insert operations tested when compared to the TLC-based PCIe 4.0 drive. Faster data access positively affects database performance and application response times that translates into a heightened user experience.*
Max. CPU Utilization	*The results demonstrate that the SCM solution delivered between 77% and 87% of CPU utilization for the six workloads tested, with small gains when compared to the TLC-based PCIe 4.0 drive. High CPU utilization in the 80% range indicates that the CPU is using all of its resources to gain higher performance.*
Memory Utilization	*The results demonstrate that the SCM solution delivered between 4.29% and 6.23% of memory utilization for the six workloads tested, with small gains when compared to the TLC-based PCIe 4.0 drive. Low memory utilization was expected since the Aerospike database enables direct device access to the SCM SSD, bypassing system memory.*

Summary

Traditional and modern data storage approaches, coupled with high data generation that is growing daily, has created new challenges to the administrators tasked with ensuring high performance and speedy response times for their enterprise applications. SCM SSDs are able to perform better than TLC-based PCIe 4.0 drives as it relates to database operation latency and throughput of large datasets. When the Supermicro AS-2114S-WN24RT Server was configured with an Aerospike database, the KIOXIA FL6 Series SCM SSDs demonstrated an ***average read database latency that was 21% to 31% faster***, and ***completed up to 32% more database operations per second*** when compared to the TLC-based PCIe 4.0 drive. The results also indicate that high CPU utilization was achieved with no additional use of system memory required to obtain these performance gains.

Fast data access and throughput, coupled with high CPU utilization, positively impacts database performance and application response times that can translate into a heightened user experience. KIOXIA FL6 Series SSDs based on innovative XL-FLASH SCM flash memory technology delivers the low latency and high performance required for enterprise database applications. While volatile memory solutions, such as DRAM, provide high database access speed, the ensuing costs can be expensive. FL6 Series SCM SSDs are uniquely architected to cost-effectively fill the gap between DRAM and traditional TLC-based flash SSDs.

Additional Supermicro AS-2114S-WN24RT server information is available here.
Additional FL6 Series SSD information is available here.

NOTES:

[1] A NoSQL database stores data in a format other than relational tables.

[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] Definition of capacity - KIOXIA Corporation defines a kilobyte (KB) as 1,000 bytes, 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 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^{50} 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.

[4] 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.

[5] Certain failures can occur when many database threads try to concurrently access the same key or record for read or write operations. This can result from the 'transaction-pending-limit' parameter which controls how many concurrent read or write operations can occur for one key at a given time. If the number of operations on a specific key is over the 'transaction-pending-limit,' any subsequent read or write requests for that record could fail. As it relates to this paper, the Aerospike application was set to a 'transaction-pending-limit' equal to 0 which means that an unlimited number of operations can occur on any single key. Depending on the application, it is recommended to set the 'transaction-pending-limit' parameter accordingly to accommodate the objective or requirement of the application.

TRADEMARKS:

Aerospike is a registered trademark of Aerospike, Inc. AMD EPYC is a trademark of Advanced Micro Devices, Inc. NVMe is a registered trademark of NVM Express, Inc. PCIe is a registered trademark of PCI-SIG. Supermicro is a registered trademark of Super Micro Computer, Inc. or its subsidiaries in the United States and other countries. Ubuntu is a registered trademark of Canonical Ltd. 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 application 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 be rendered inaccurate 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. is expressly advised of the possibility of such damages.

KIOXIA