KIOXIA

Performance Brief

Improving Data Analytics Performance Using 12 Gb/s KIOXIA RM7 Series Value SAS SSDs Compared with SATA SSDs

Introduction

Collecting and analyzing data has become paramount to organizations who seek valuable insights about their business operations, such as customer preferences, market trends, business strategies and performance. Data analytics provides tools that uncover intelligence for informed decision-making and sustained growth. The rate of data creation and collection is growing rapidly, making data analytics increasingly essential.

With vast amounts of data generation, traditional data storage hardware encounters performance challenges, especially in keeping pace with the throughput required for more advanced analysis and larger datasets. High throughput is critical for efficiently handling the complex queries needed for data analysis. As such, there is a need for faster storage solutions that can support these complex queries with high data throughput and low latency response times. These performance capabilities enable key compute resources to be freed up for additional processing tasks.

There are many types of analytical workloads used in today's data centers - however, this performance brief focuses on accelerating analytical queries against transactional data stored in a Microsoft® SQL Server™ database. With the need for high data throughput, low latency response times and compute resource optimization, KIOXIA RM7 Series Value SAS SSDs are a great drop-in replacement option for new or existing servers that deploy enterprise SATA SSDs. Value SAS SSDs are interchangeable with SATA SSDs and use the same CPUs, system memory and RAID[1]/Dell® PERC[2] cards within the server.

This performance brief presents an SSD performance comparison using data analytics workloads. It compares KIOXIA RM7-R Series Value SAS SSDs with SATA SSDs from a leading provider (Vendor A). A Dell PowerEdge™ R660 rack server was used for testing.

Data analytics test runs were performed on both SSD configurations and the recorded metrics included drive read throughput, drive read latency, geometric mean of query completion times and the completion time for the entire query set. The four metrics were recorded from two tests for each SSD configuration - one test with one virtual user and one test with seven virtual users – and each virtual user performed twenty-one different query analytical tests. For one virtual user, advanced analytic queries were conducted in succession against the database, and for seven virtual users advanced analytic queries were conducted in parallel against the database. HammerDB[3] test software was used for each SSD configuration that enabled the TPROC-H[4] data analytics workloads to run against the Microsoft SQL Server database.

The test results show that with a maximum SAS interface throughput of 12 gigabits per second (Gb/s), the KIOXIA RM7-R Series Value SAS SSDs delivered demonstrable performance improvements when compared with the Vendor A SSDs limited by the 6 Gb/s SATA interface. These Value SAS SSDs delivered a significant boost when accelerating data intensive computational tasks such as data analytics, facilitating seamless performance for advanced query analytics, even when multiple users sent queries concurrently.

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

Test Results Snapshot

KIOXIA RM7-R Series Value SAS SSDs delivered the following performance results when compared with SATA SSDs from Vendor A:

Average Database Throughput
(higher is better)

1 Virtual User	7 Virtual Users
92% Higher	108% Higher

Average Drive Read Latency
(lower is better)

1 Virtual User	7 Virtual Users
12% Lower	37% Lower

Geometric Mean of Query Completion Times
(lower is better)

1 Virtual User	7 Virtual Users
47% Lower	52% Lower

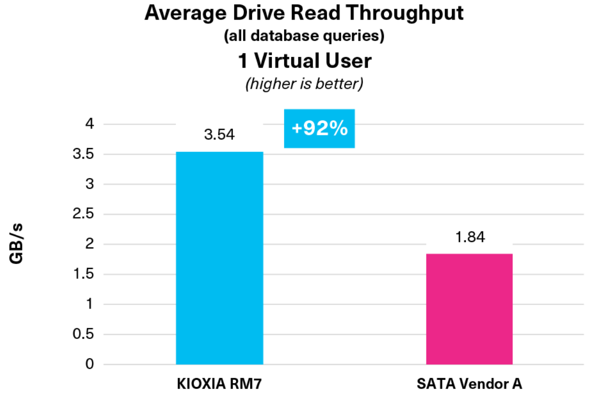
Completion Time for Entire Query Set
(lower is better)

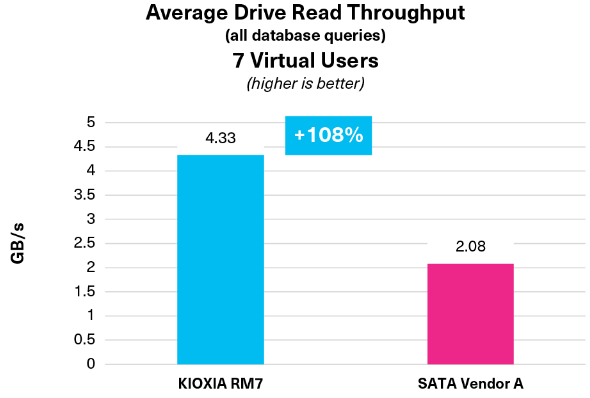
1 Virtual User	7 Virtual Users
38% Lower	53% Lower

Test Results[5]

Test 1: Average Drive Read Throughput

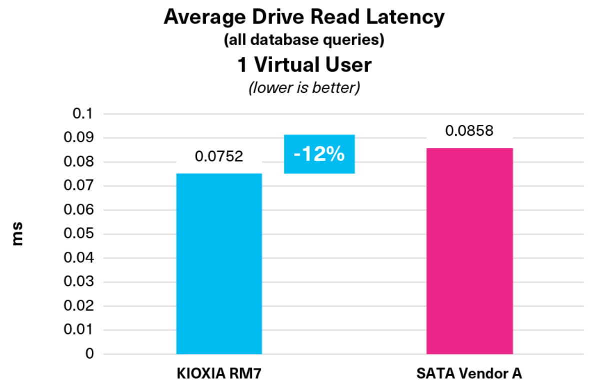
This test represents the average SSD read throughput of all database queries and shows the test server's ability to service these incoming queries. If the data returned by the query is large, it can exceed the throughput capabilities of the storage subsystem resulting in high I/O waits. When this scenario occurs, the CPU is waiting for data instead of performing the requested analytics resulting in longer query times that can drastically impact performance. This test highlights each SSD configuration's adequacy in handling the throughput needed. For one virtual user, this test performed a series of analytic queries against the Microsoft® SQL Server™ database. For seven virtual users, the same series of analytic queries were performed in parallel. The results are in gigabytes per second (GB/s). The higher result is better.

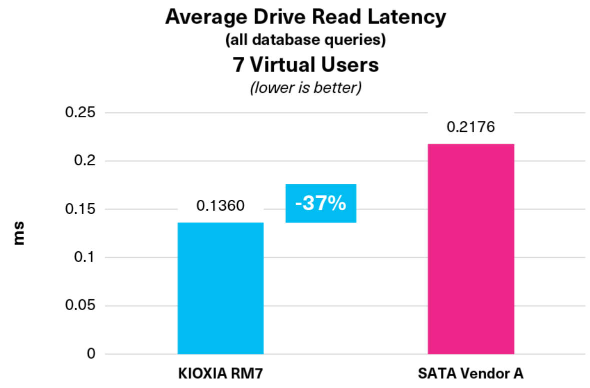




Test 2: Average Drive Read Latency

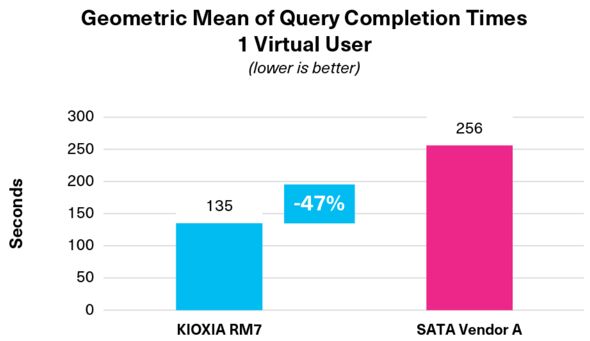
This test represents the average time it took to perform an SSD read operation during the test run. It includes the average time it took for the read operation to successfully complete, once the drive received the read operation request issued from the workload generator. For one virtual user, this test performed a series of analytic queries against the Microsoft SQL Server database. For seven virtual users, the same series of analytic queries were performed in parallel. The results are in milliseconds (ms). The lower result is better.

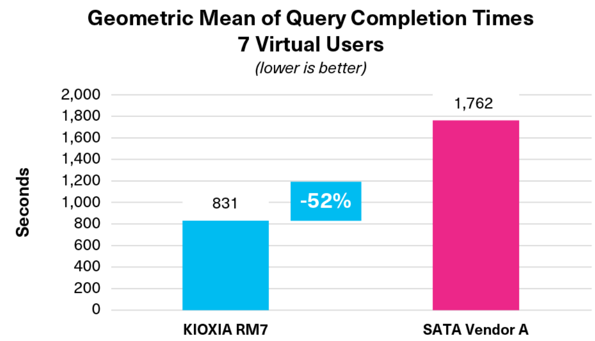




Test 3: Geometric Mean of Query Completion Times

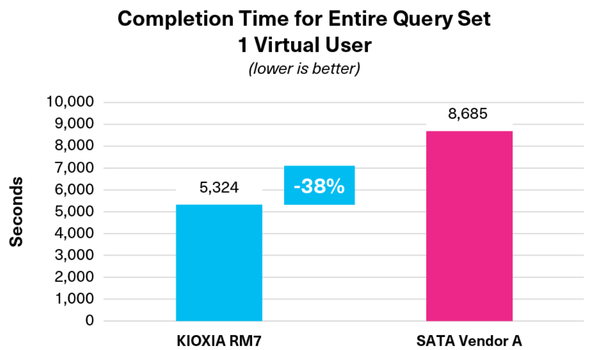
This test represents the average time it took to complete all of the queries in the query set. For one virtual user, this test performed a series of analytic queries against the Microsoft® SQL Server™ database. For seven virtual users, the same series of analytic queries were performed in parallel. As the workload ramped up, the KIOXIA RM7 Series Value SAS SSDs delivered low latency emphasizing that as more analytics were run in parallel, these drives returned fast query results when compared with SATA SSDs. The results are in seconds. The lower result is better.

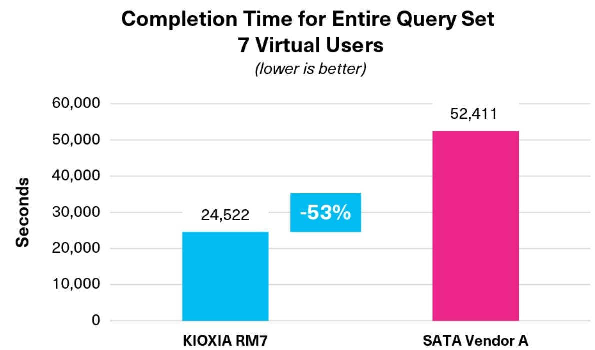




Test 4: Completion Time for Entire Query Set

This test represents the total amount of time it took for each user to complete every query in the query set. For one virtual user, this test performed a series of analytic queries against the Microsoft SQL Server database. For seven virtual users, the same series of analytic queries were performed in parallel. The results are in seconds. The lower result is better.





Analysis

For seven virtual users, the KIOXIA RM7-R Series Value SAS SSDs delivered over 2x faster drive read throughput. This enables organizations to obtain valuable insights twice as fast when compared with SATA SSDs while doubling the analytics runs that can be performed to gather additional data. For a single virtual user, the KIOXIA RM7-R Series Value SAS SSDs delivered almost twice the drive read throughput. A summation of these test results follows:

1 Virtual User

Test		KIOXIA RM7-R 12 Gb/s SAS	Vendor A SATA	KIOXIA RM7-R Gains
Average Drive Read Throughput	*(higher is better)*	3.54 GB/s	1.84 GB/s	92%
Average Drive Read Latency	*(lower is better)*	0.0752 ms	0.0858 ms	12%
Geometric Mean of Query Completion Times	*(lower is better)*	135 seconds	256 seconds	47%
Completion Time for Entire Query Set	*(lower is better)*	5,324 seconds *(~88 minutes)*	8,685 seconds *(~144 minutes)*	38%

7 Virtual Users

Test		KIOXIA RM7-R 12 Gb/s SAS	Vendor A SATA	KIOXIA RM7-R Advantages
Average Drive Read Throughput	*(higher is better)*	4.33 GB/s	2.08 GB/s	108%
Average Drive Read Latency	*(lower is better)*	0.1360 ms	0.2176 ms	37%
Geometric Mean of Query Completion Times	*(lower is better)*	831 seconds	1,762 seconds	52%
Completion Time for Entire Query Set	*(lower is better)*	24,522 seconds *(~408 mins. / ~ 6.8 hrs.)*	52,411 seconds *(~873 mins. / ~14.5 hrs.)*	53%

When compared with Vendor A SATA SSDs for one virtual user, the KIOXIA RM7-R Series Value SAS SSDs:

- *Completed each query in the query set in about 135 seconds, or a 121 second improvement*
- *Completed the entire analytics query set in about 5,324 seconds, or a 56 minute savings in compute time and resources*
- *Delivered 3.54 GB/s of drive read throughput on average, or 1.7 GB/s higher*
- *Delivered 0.0752 ms of drive read latency on average, or 0.0106 ms lower*

When compared with Vendor A SATA SSDs for seven virtual users, the KIOXIA RM7-R Series Value SAS SSDs:

- *Completed each query in the query set in about 831 seconds, or a 15 minute improvement*
- *Completed the entire analytics query set in about 24,522 seconds, or a 7.7 hour savings in compute time and resources*
- *Delivered 4.33 GB/s of drive read throughput on average, or 2.25 GB/s higher*
- *Delivered 0.1360 ms of drive read latency on average, or 0.0816 ms lower*

The HammerDB TPROC-H data analytics workload generated for each test configuration was identical. The workload ran twenty-one different queries for each virtual user assigned, in a random order. In a single virtual user configuration, each of the queries were run in series, one at time. However, this type of sequential query testing can result in a workload that does not saturate SSD throughput capabilities. Since each of the twenty-one queries were different, a single query may not reach the full throughput potential of the drives.

When the testing was run with more than one virtual user, each user exhibited the same behavior of executing queries that were picked at random. However, additional virtual users executing different queries running in parallel can increase the potential to meet the throughput maximum of the drives. When more than one virtual user was added to the configuration, an increase in drive read throughput occurred. The reason for using seven virtual users for these tests was based on Microsoft® recommendations and has no correlation to how many virtual users would be needed to saturate any SSD configuration tested.

Though the overall system power draw was slightly higher than the SATA SSD configuration (up to 5.8 watts), the KIOXIA RM7-R Series Value SAS SSDs were able to deliver significant performance gains. At the data center level, switching to Value SAS SSDs enables administrators to use the same number of servers for about twice the performance, or conversely, scale down the number of servers needed while still meeting throughput requirements. These actions can lead to savings on both power consumption and total cost of ownership, without sacrificing performance.

Summary

This performance brief presented an SSD performance comparison using a HammerDB TPROC-H data analytics workload with KIOXIA RM7-R Series Value SAS SSDs and SATA SSDs from a leading provider. Testing showed that KIOXIA RM7-R Series Value SAS SSDs were able to deliver twice the throughput performance when compared with the SATA SSD configuration.

With their high read throughput, KIOXIA RM7-R Series Value SAS SSDs can effectively process large analytics datasets, facilitating seamless performance for advanced query analytics even when multiple users are accessing them concurrently. Compared with enterprise SATA SSDs, these SSDs deliver lower read latency, resulting in faster querying completion times and reduced waiting periods for data retrieval. Organizations would be provided with further meaningful insights quickly using KIOXIA RM7-R Series Value SAS SSDs, which in turn enable better informed decisions to be made when compared with SATA SSDs. With twice the bandwidth versus SATA, KIOXIA RM7-R Series Value SAS SSDs deliver timely and actionable analytical results, meeting the demands of today's data driven businesses.

KIOXIA RM7 Series Value SAS SSD Product Info

The latest generation KIOXIA RM7 Series Value SAS SSDs support the 12 Gb/s SAS-3 interface and are available in two 2.5-inch[6] configurations: RM7-R Series for read intensive applications (1 DWPD[7], up to 7.68 terabyte[8] (TB) capacities) and RM7-V Series for higher endurance mixed use applications (3 DWPD, up to 3.84 TB capacities). Security options[9] are available for both configurations.

Additional KIOXIA RM7 Series SSD information is available at the KIOXIA 'Life After SATA' site.

KIOXIA RM7 Series SSD[10]

Appendix A

Hardware/Software Test Configuration

Server Information	
Server Model	Dell® PowerEdge™ R660
No. of Servers	1
BIOS Version	1.6.6
PERC Card	Dell PERC H965i Front (Embedded)

CPU Information	
CPU Model	Intel® Xeon® Gold 6430
No. of Sockets	2
No. of Cores	32
Frequency (in gigahertz)	2.1 GHz

Memory Information	
Memory Type	DDR5
Memory Speed	DDR5-4400
Memory Size	32 GB
No. of DIMMs	4
Total Memory Size	128 GB

SSD Information		
SSD Model	KIOXIA RM7-R Series	Vendor A
Form Factor	2.5-inch	2.5-inch
Interface	SAS-3	SATA-3
Interface Speed	12 Gb/s	6 Gb/s
No. of SSDs	4	4
SSD Capacity	3.84 TB	3.84 TB
Drive Writes per Day	1	1
Active Power	up to 9 watts	up to 3.2 watts

Operating System Information	
Operating System (OS)	Microsoft Windows® Server® 2022 Datacenter
OS Version	10.0.20348
Database Model	Microsoft® SQL Server™ 2022
Database Version	16.0.1000.6

Test Software Information	
Test Software Model	HammerDB
Benchmark	TPROC-H
Version	4.9
No. of Virtual Users	1 or 7

Appendix B

Configuration Set-up

A Dell® PowerEdge™ R660 rack server was installed with the Microsoft® Windows® Server® 2022 Datacenter operating system.

Four 3.84 TB[8] KIOXIA RM7-R Series Value SAS SSDs and four 3.84 TB Vendor A SATA SSDs were each placed in a hardware RAID 10[9] set in the BIOS, using the default RAID controller settings.

Using Disk Management on the operating system, the RAID volume was initialized on the OS and a new volume was created.

Microsoft® SQL Server™ was then installed on the newly created volume.

HammerDB test software was installed on the same volume that Microsoft SQL Server resided enabling the TPROC-H data analytics workloads to run against the database. For the schema, to simulate a very large typical database in which data analytics would be able to perform on it, a scale factor of 1,000 was used. The schema was built with MAXDOP[12] and set to '7' with virtual users set to '1.' With a scale factor of 1,000, the size of the database was 3,270 GB[8].

Under Schema, Options need to be edited/changed as follows:

- *Unselect Encrypt Connection*
- *For Microsoft SQL Server ODBC Driver, change 18 to 17 (for easier compatibility)*
- *Select Microsoft SQL Server Authentication - enter the password from the Microsoft SQL Server set-up*
- *Set MAXDOP to 7 – the scale factor to 1,000 – VU to 1*
- *Select Use BCP Option*

Once the test schema is built, the Driver Script needs to be edited/changed as follows:

- *Select Encrypt Connection*
- *For Microsoft SQL Server ODBC Driver, change 18 to 17*
- *Select Microsoft SQL Server Authentication - enter the password from the Microsoft SQL Server set-up*
- *Set MAXDOP to 7 – and Total Query Sets Per User to 1*

Select Virtual User and change the Options as follows:

- *Change Virtual Users to 1 or 7 depending on the test*
- *Change User Delay (in ms) to 1 and Repeat Delay (in ms) to 1*
- *Select Show Output, Log Output to Temp and Log Timestamps*
- *Select Run to initiate the test*

The HammerDB TPROC-H workload tests are now initiated.

Appendix C

Test Procedures

The HammerDB TPROC-H workload was run against the Microsoft® SQL Server™ instance using four 3.84 TB[8] KIOXIA RM7-R Series Value SAS SSDs and four 3.84 TB SATA SSDs from Vendor A in RAID 10 using default settings.

The first test phase included HammerDB TPROC-H workloads for one virtual user and a focus on query processing in a single stream. The advanced analytic queries were performed in succession against the Microsoft SQL Server database.

The metrics that were measured during the test run included:

- *Drive Read Throughput*
- *Drive Read Latency*
- *Geometric Mean of Query Completion Times*
- *Completion Time for Entire Query Set*

The results for each test were recorded.

The second test phase included HammerDB TPROC-H workloads for seven virtual users and a focus on query processing from multiple concurrent users. The advanced analytical queries were performed in parallel against the Microsoft SQL Server database.

The metrics that were measured during the test run included (same as with one virtual user):

- *Drive Read Throughput*
- *Drive Read Latency*
- *Geometric Mean of Query Completion Times*
- *Completion Time for Entire Query Set*

The results for each test were recorded.

NOTES:

[1] RAID (redundant array of independent disks) is a way of storing the same data in different places on multiple SSDs to protect data in the case of a drive failure. There are different RAID levels depending on the redundancy required.

[2] Dell PERC (Dell PowerEdge™ RAID Controller) is a series of RAID disk array controllers manufacturered by Dell for its PowerEdge servers. The controllers support SAS and SATA hard disk drives (HDDs) and solid-state drives (SSDs).

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

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

[6] 2.5-inch indicates the form factor of the SSD and not the drive's physical size.

[7] DWPD: 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.

[8] 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.

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

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

[11] RAID 10 is a redundant array of independent disks configuration that combines disk mirroring and disk striping to protect data.

[12] MAXDOP (Maximum Degree of Parallelism) is a setting in most databases, such as Microsoft® SQL Server™. The metric set by the user represents the maximum number of CPU cores that the user wants the database to use for processing queries.

TRADEMARKS:

Dell and PowerEdge are either registered trademarks or trademarks of Dell Inc. Intel and Xeon are registered trademarks of Intel Corporation or its subsidiaries in the U.S. and/or other countries. Microsoft, Microsoft SQL Server, Microsoft Windows and Windows Server and are registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries. TPC-H is a trademark of the Transaction Processing Performance Council. 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.