KIOXIA

Performance Brief

Getting More Out of Data Analytics Workloads with 24G SAS SSDs

Performance and Compute Resource Gains Using KIOXIA PM7 Series SSDs Compared with Enterprise SATA SSDs

The rate at which data is created and collected continues to increase. Organizations are gaining valuable insights and further intelligence from the massive amount of data they collect within a data center. Analyzing these large datasets allows businesses to find patterns, trends and correlations which can then provide insights for making critical decisions, or becoming more efficient, or optimizing processes.

Collecting and analyzing data to make more informed decisions has become paramount. With vast amounts of data being generated, traditional data storage hardware has difficulties and limitations in keeping pace with the advanced queries requested of it. There is a need for faster storage media that can support these queries with high data throughput and low latency response times, which in turn can save valuable compute time and resources, and enable the system to be available for other processing tasks.

This performance brief focuses on the performance gains in data analytics workloads using 24G SAS SSDs compared with enterprise SATA SSDs, which in turn enables valuable compute/storage resources to be available for additional application tasks.

Data Analytics Focus

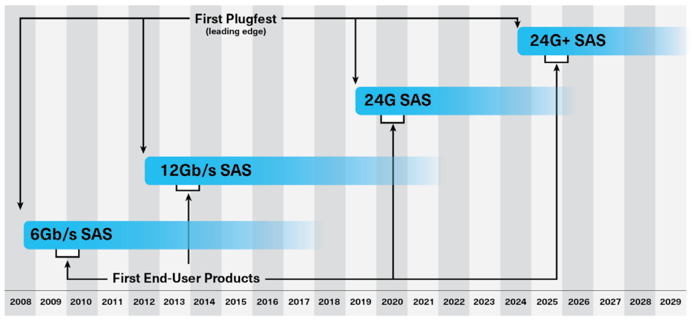
There are many types of data analytics used to analyze manufacturing operational data, market data, competitive data, financial data and enterprise security data, to name a few. The focus of this performance brief, however, is on accelerating queries against transactional data stored in a Microsoft® SQL Server™ database. It includes performance testing for one virtual user that conducts advanced analytic queries in succession against the database under test. Separate tests also include seven virtual users who conduct the analytic queries in parallel. For both the one and seven virtual user test comparisons, drive read throughput and latency performance results are also provided. KIOXIA PM7 Series SSD performance is compared with leading and available competitive SATA SSDs for comparing 24G SAS database performance to SATA.

24G SAS Overview

The SAS interface moves data into and out of storage devices using the SCSI protocol. It has a reputation for reliability and is used for applications that require high availability and data protection. SAS is one of the main storage interfaces between computing and storage subsystems in data centers worldwide, and with a defined roadmap (Figure 1).

Most servers today ship with a SAS infrastructure and a backplane that also connects to SATA SSDs and/or HDDs, enabling both SAS and SATA drives to be used in the same drive bay. As future storage requirements change, SATA drives can be easily replaced with higher performing and larger capacity 24G SAS SSDs without any changes to the server or SAS infrastructure. Since the SAS interface can support up to 65,535 devices through expanders, it is well-positioned for large data center topologies where thousands of drives are required to support a range of applications. Moreover, the 24G SAS interface is backwards-compatible with earlier infrastructure generations (12 Gb/s SAS and 6 Gb/s SAS).

SAS Technology Roadmap



Source: SCSI Trade Association, September 2021

Figure 1: SAS roadmap from 6 Gb/s to 24G and beyond to 24G+
Reproduced with permission from SCSI Trade Association, September 2021

KIOXIA PM7 Series 24G SAS SSD Overview

PM7 Series SSD[3]

KIOXIA was the first storage vendor to introduce SSDs based on the 24G SAS interface[1] with its PM6 Series introduction in 2020. The KIOXIA PM7 Series is the 2nd generation of 24G SAS SSDs that leverages 112-layer BiCS FLASH™ 3D flash memory technology in a 2.5-inch[2] form factor. The PM7 Series features flash die failure recovery protection, endurance and security capabilities, and represents KIOXIA's 7th SAS SSD generation that builds on the company's successes as a leading SAS SSD vendor.

PM7 Series SSDs are available in two configurations:

- *PM7-R Series for read-intensive applications supports 1.92 terabyte[4] (TB) up to 30.72 TB capacities at 1 Drive Write Per Day[5] (DWPD)*
- *PM7-V Series for mixed-use applications supports 1.6 TB up to 12.8 TB capacities at 3 DWPD*

Test Steps / Objectives

- *Create a SQL Server database on an HPE® ProLiant® DL385 Gen10 Plus server platform using a Microsoft Windows® Server 2022 Datacenter operating system*
- *Use SQL Server to store its data and log files on four PM7-R SSDs, as well as four competitive SATA SSDs from a leading vendor*
- *Test SQL Server database performance using HammerDB[6] benchmark software with a TPROC-H[7] workload on four PM7-R SSDs and four competitive SATA SSDs*
- *Collect the performance metrics for both SSD configurations covering drive read throughput, drive read latency, geometric mean of query times and completion time of query sets*
- *Demonstrate that PM7-R SSDs deliver higher overall performance and lower overall latency than equivalent capacity SATA drives*

Test Configuration

The hardware and software equipment used to run the database performance tests described above includes:

Database Server Information	
Model	HPE ProLiant DL385 Gen10 Plus v2
Number of CPU Sockets	2
CPU	AMD EPYC™ 7543
No. of CPU Cores	32
CPU Frequency	2.8 GHz
Total Memory	256 GB DDR4 DRAM
Memory Frequency	3,200 megatransfers per second (MT/s)
Operating System Information	
Operating System	Microsoft Windows Server 2022 Datacenter
Version	21H2

Software Information	
Database Software	Microsoft SQL Server 2022
Version	16.0.1000.6
Test Software	HammerDB
Version	4.6

SSD Information		
Interface	24G SAS	SATA
Interface Speed	22.5 Gb/s	6 Gb/s
Model	KIOXIA PM7-R Series	Vendor A
Number of Devices	4	4
Form Factor	2.5-inch	2.5-inch
Capacity	7.68 TB	7.68 TB
DWPD	1 (5 years)	1 (5 years)
Active Power	5 W	2.2 W

Test Setup and Procedure

Test Setup:

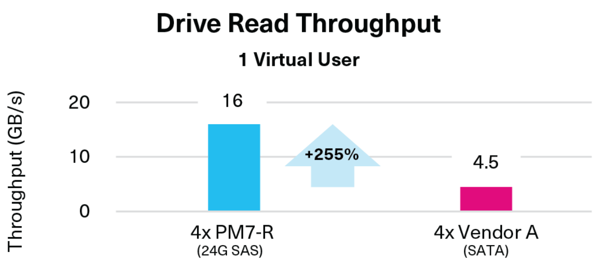
- *The HPE ProLiant DL385 Gen10 Plus v2 server was loaded with four KIOXIA PM7-R SSDs (each with 7.68 TB capacity) in support of read-intensive applications.*
- *Microsoft Windows Server 2022 was installed.*
- *The four PM7-R SSDs were configured with Windows Storage Spaces™ for presenting a four drive mirrored volume.*
- *Microsoft SQL Server 2022 was installed, and then HammerDB software was used to run a TPROC-H workload against the Microsoft SQL Server instance.*
- *To emulate a typically very large database in which analytics would be performed on, a pre-determined database size (or scale factor) of 1,000 was used.*

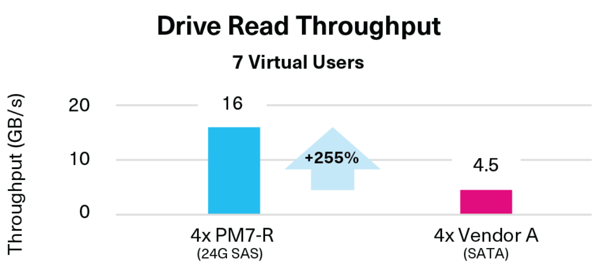
Test Procedure:

- *The HammerDB TPROC-H workload was then run against the Microsoft SQL Server instance using the four KIOXIA PM7-R SSDs with 7.68 TB capacities. The first phase included workload tests for one virtual user with a focus on query processing throughput in a single stream, performing advanced analytic queries in succession against the SQL Server database under test. The tests included: (1) Drive Read Throughput; (2) Drive Read Latency: (3) Geometric Mean of Query Times: and (4) Query Set Completion Time.*
- *The metric for each test was then recorded.*
- *The HammerDB TPROC-H workload tests were then run for seven virtual users with a focus on total query throughput from multiple concurrent users. In this test case, seven virtual users are a recommended HammerDB configuration when running a scale factor of 1,000[8]. The tests also included: (1) Drive Read Throughput; (2) Drive Read Latency; (3) Geometric Mean of Query Times; and (4) Query Set Completion Time.*
- *The metric for each test was then recorded.*
- *Using four Vendor A SATA SSDs with 7.68 TB capacities each for both one and seven virtual users, the HammerDB TPROC-H workload was then run against the SQL Server instance.*
- *The metric for each test was then recorded.*

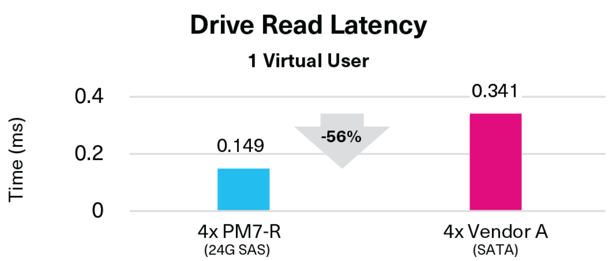
Test Results

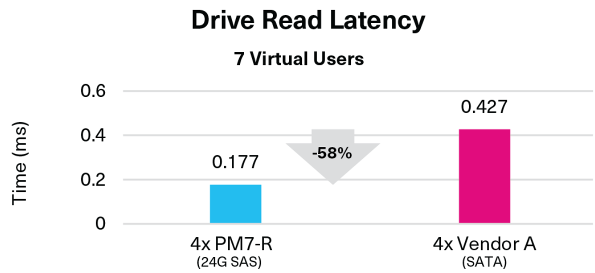
Drive Read Throughput *(in gigabytes per second (GB/s)*:
The drive read throughput test represents the total query throughput of all database queries. For one virtual user, this test performed analytic queries in succession against SQL Server. For seven virtual users, this test performed analytic queries of these users in parallel. For these tests, the higher result is better.



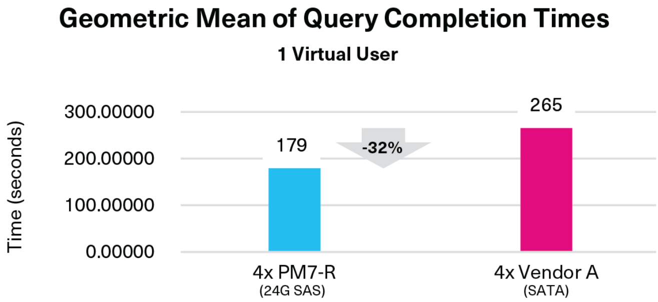


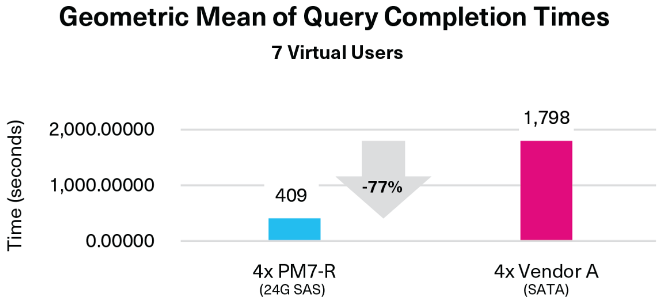
Drive Read Latency *(in milliseconds, or ms)*:
The drive read latency test represents the time it takes to perform a drive read operation. It includes the average time it takes for the read operation to successfully complete once the drive has received the read operation request issued from the workload generator. For one virtual user, this test performed analytic queries in succession against SQL Server. For seven virtual users, this test performed analytic queries in parallel. For these tests, the lower result is better.



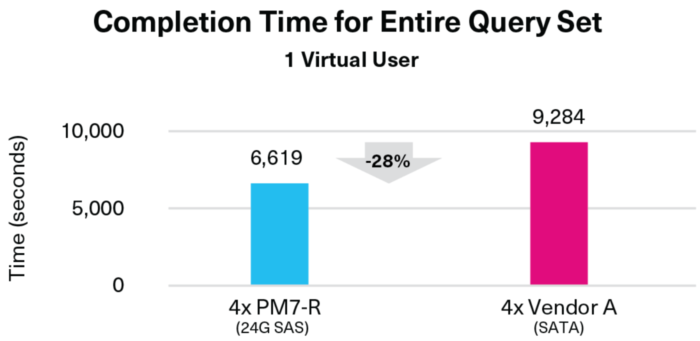


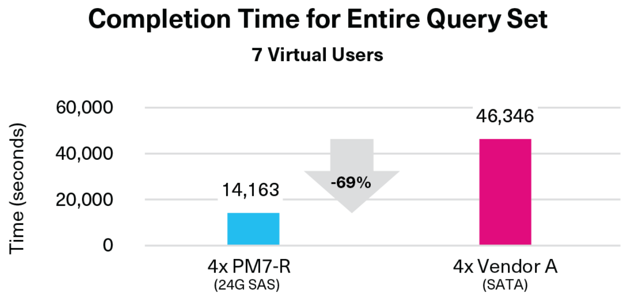
Geometric Mean of Query Times *(in seconds)*:
The geometric mean of query times test is the average time it takes to complete an individual query in the query set. For one virtual user, this test performed analytic queries in succession against SQL Server. For seven virtual users, this test performed analytic queries of these users in parallel. For these tests, the lower result is better.





Completion Time for Entire Query Set *(in seconds)*:
The completion time for the entire query set is the total amount of time in which it takes to complete every query in the query set. For one virtual user, this test performed analytic queries in succession against SQL Server. For seven virtual users, this test performed analytic queries of these users in parallel. For these tests, the lower result is better.





Test Analysis

The test results indicate that the KIOXIA PM7-R Series 24G SAS SSD configuration accelerated database performance in all four tested performance areas, whether the test was on one virtual user with query processing in a single stream or seven virtual users with query processing in parallel. Since both one virtual user and seven virtual users have the same total throughput, both the KIOXIA PM7-R Series SSDs and Vendor A SATA drives were being fully utilized in terms of throughput in both tests, but the PM7-R SSDs were able to process 3x the amount of data.

For the tests conducted, the scale factor was set to 1,000; the Maximum Degree of Parallelism[9] (MAXDOP) was set to 7; and the database table size was 3.27 TB. The following is a summation of the test results:

1 Virtual User

Test Specifics	KIOXIA PM7-R 24G SAS	Vendor A SATA	KIOXIA PM7-R Advantages
Drive Read Throughput (higher is better)	16 GB/s	4.5 GB/s	**255%**
Drive Read Latency (lower is better)	0.149 ms	0.341 ms	**56%**
Geometric Mean of Query Completion Times (lower is better)	179 seconds	265 seconds	**32%**
Completion Time for Entire Query Set (lower is better)	6,619 seconds (~110 minutes)	9,284 seconds (~154 minutes)	**28%**

7 Virtual Users

Test Specifics	KIOXIA PM7-R 24G SAS	Vendor A SATA	KIOXIA PM7-R Advantages
Drive Read Throughput (higher is better)	16 GB/s	4.5 GB/s	**255%**
Drive Read Latency (lower is better)	0.177 ms	0.427 ms	**58%**
Geometric Mean of Query Completion Times (lower is better)	409 seconds	1,798 seconds	**77%**
Completion Time for Entire Query Set (lower is better)	14,163 seconds (~236 mins. / ~ 3.9 hrs.)	46,346 seconds (~772 mins. / ~12.8 hrs.)	**69%**

When one virtual user performed the query set, the PM7-R SSDs were able to complete each individual query in the set in about 179 seconds versus the Vendor A SATA SSDs which took about 265 seconds to complete (or an 86 second improvement). The single virtual user also completed the entire analytics query set in about 6,619 seconds with PM7-R SSDs compared to the Vendor A SATA set at 9,284 seconds. This represents a 44-minute savings in compute resources when PM7-R SSDs are deployed, allowing the system to be available for other tasks. The PM7-R SSDs were able to utilize upwards to 16 GB/s of read throughput compared to only 4.5 GB/s on the Vendor A SATA set (or 11.5 GB/s faster). The drive read latency for the PM7-R SSDs was only 0.149 ms, or 0.192 ms better than the Vendor A SATA set at 0.341 ms.

When seven virtual users performed the query set, the KIOXIA PM7-R SSDs were able to complete each individual query in the set in about 409 seconds versus the Vendor A SATA SSDs which took about 1,798 seconds to complete (or 23 minutes faster). The seven virtual users also completed the entire analytics query set in about 14,163 seconds with PM7-R SSDs compared to the Vendor A SATA set at 46,346 seconds. This represents a significant, almost 9-hour savings in compute resources when PM7-R SSDs are deployed, allowing the system to be available for other tasks. The PM7-R SSDs were able to utilize upwards to 16 GB/s of read throughput compared to only 4.5 GB/s on the Vendor A SATA set (or 11.5 GB/s faster). The drive read latency for the PM7-R SSDs was only 0.177 ms, or 0.250 ms better than the Vendor A SATA set at 0.427 ms.

Summary

KIOXIA PM7 Series 24G SAS SSDs can accelerate data-intensive computational workloads, such as data analytics, and meet performance requirements better when compared to enterprise SATA drives. Large analytics data sets require more storage, and with high storage capacities, KIOXIA PM7 Series SSDs can hold large databases in which these analytics can be performed on. The high sequential read throughput of KIOXIA PM7 Series SSDs enables multiple users to concurrently perform advanced query analytics. They also provide lower latency than SATA, allowing for faster querying times that can lead to a decrease in the overall time waiting for data to be gathered. KIOXIA PM7 Series SSDs also enable valuable compute resources to be available for other processing tasks. Its high performance allows businesses to analyze data, discover meaningful patterns and trends, and provide effective and decision making in a timely manner.

When compared to the SATA interface, 24G SAS delivers about four times the bandwidth (22.5 Gb/s vs 6 Gb/s), and about eight times the bandwidth when you consider that SAS is full-duplex versus SATA at half-duplex. In a data center, with data streaming constantly, the KIOXIA PM7 Series SSDs can save users time throughout the entire data analytics process, from storing and writing the data, to indexing the data, to running/reading analytics against the data, all in near real time. KIOXIA PM7 Series 24G SAS SSDs deliver fast data analytics results when users need them.

Notes:

[1] Source: KIOXIA America, Inc., News Release, June 16, 2020, https://americas.kioxia.com/en-us/business/news/2020/ssd-20200616-1.html.

[2] 2.5-inch indicates the form factor of the SSD and not the drive's physical size.

[3] Product image may represent a design model.

[4] 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 (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^{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.

[5] Drive Write(s) 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, under the specified workload for the specified lifetime. Actual results may vary due to system configuration, usage, and other factors.

[6] 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.

[7] 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.

[8] Source: HammerDB, https://hammerdb.com/docs/ch11s02.html.

[9] The Maximum Degree of Parallelism (MAXDOP) 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, such as the analytics queries outlined in this benchmark. For this benchmark, seven cores were used.