KIOXIA

Application Brief

Accelerating Storage Performance for Distributed Data Analytics Workloads

Featuring KIOXIA PM7-R Series 24G SAS SSDs deployed in HPE® ProLiant® DL385 Gen10 Plus v2 servers, running Apache Spark™ and Apache Hadoop®

Introduction

Data analytics can lead to valuable insights that organizations of all kinds use to enhance applications, streamline operations, improve efficiencies and identify key trends. This collection of data is often unstructured or provided in different incompatible structures such as databases, logs, message queues, etc., which then needs to be put into useful forms before analysis begins. The ability to turn data into worthwhile analysis quickly is now vital for many organizations in today's online world.

Transforming data into insights and trends requires efficient data parsing to be able to analyze and report relevant facts from an overwhelming amount of collected data. As data comes from various sources and distributes across a cluster of server nodes, performing transformations on these large datasets is storage intensive. The storage capabilities of the data source must deliver high performance and Quality of Service (QoS), enabling relevant data to be aggregated and presented to end users quickly and efficiently by front-end applications.

This application brief presents accelerated storage performance of distributed data analytics workloads using Apache Spark and Apache Hadoop on a cluster of three HPE ProLiant DL385 Gen10 Plus v2 servers deployed with twelve 3.84 terabyte[1] (TB) KIOXIA PM7-R Series 24G SAS SSDs for read intensive applications. This configuration is tested against twelve leading 6Gb/s SATA SSDs from Vendor A with the same capacities. The drives on each system participated in a hardware RAID 5[2] set. There are eight tests in the comparison. Appendix A covers the hardware and software configuration – Appendix B covers the configuration set-up and test procedures.

The Hadoop Distributed File System (HDFS) stores various data types and is often used for big-data processing applications. It is highly scalable, fault tolerant and delivers high storage throughput. HDFS creates files and divides them into blocks for storing on multiple servers. Through data block replication, the block replicas placed across the cluster increase data availability and performance. In this paper, HDFS was used to provide the persistent storage and data source for Apache Spark.

Apache Spark is an open-source distributed processing engine that is high performing, scalable, and capable of computing large data amounts. It allows workload distribution to the compute and memory portions of multiple systems, performing compute transformations faster. This ability to perform a wide range of analytical tasks with speed occurs by aggregating the performance of many discrete nodes then loading the entire dataset in memory on the cluster to perform computations on it. With high-performance storage, this results in faster time to complete computing and processing of workloads, which in turn allows for faster data analytics insights from huge amounts of data.

Test results show that the KIOXIA PM7-R Series 24G SAS SSDs demonstrate faster completion times, higher throughput, lower latency and better CPU usage of data analytics workloads than Vendor A 6Gb/s SATA SSDs in an HPE ProLiant Gen10 Plus v2 clustered server environment.

Test Results Snapshot

The KIOXIA PM7-R Series 24G SAS SSDs deliver exceptional data analytics performance versus 6Gb/s SATA SSDs as depicted by the test results:

Maximum Read Throughput

190% Higher

Maximum Write Throughput

141% Higher

Average Read Latency

99% Lower

Average Write Latency

82% Lower

Maximum CPU Utilization

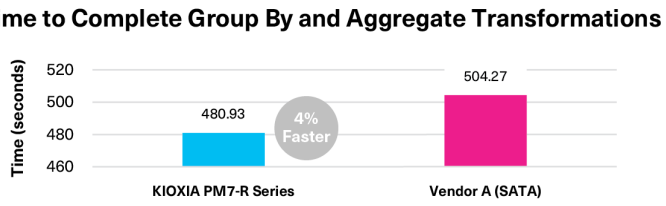
18% Better Usage

The test results presented include a brief description of each workload test, a graphical depiction of the test results and a test analysis.

TEST RESULTS

Workload 1: Time to Complete Group By and Aggregate

Workload 1 is the first of three time to complete tests. This first workload tested the amount of time it took to perform group by and aggregate transformations (in seconds) on the dataset stored on the HDFS across the cluster. Aggregation is a way to group data together performed on a cluster. The lowest result represents the fastest completion time.

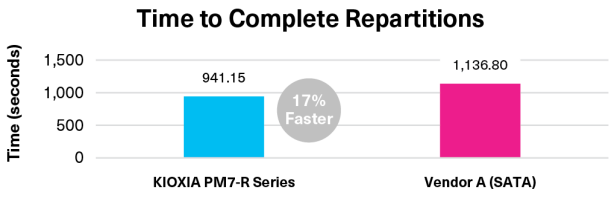


Test Analysis:

The KIOXIA PM7-R Series SSDs demonstrated 4% faster times to complete group by and aggregate transformations versus 6Gb/s SATA SSDs. These results enable analysts to gather specific pieces of data much faster for analysis.

Workload 2: Time to Complete Repartitions

This workload measured the amount of time it took to perform data repartitioning across the cluster (in seconds). Repartitioning is necessary to redistribute data throughout the cluster so that individual worker nodes are able to access data locally, which in turn optimizes other transformations and jobs for future workloads. The lowest result represents the fastest completion time.

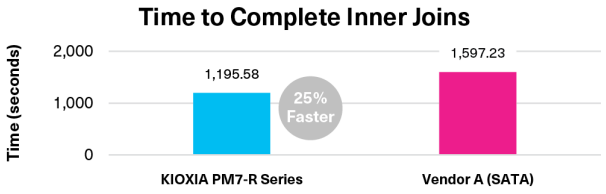


Test Analysis:

The KIOXIA PM7-R Series SSDs demonstrated 17% faster times to complete repartitions versus 6Gb/s SATA SSDs. This enables analysts to take advantage of local data access across the cluster for faster analysis.

Workload 3: Time to Complete Inner Join

This third time to complete workload measures the time it took to perform inner joins on two separate data frames (in seconds). Inner join operations are I/O intensive and can slow down the gathering of data spread out across server nodes within a cluster. The lowest result represents the fastest completion time.

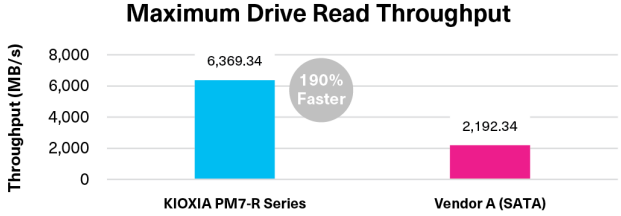


Test Analysis:

The KIOXIA PM7-R Series SSDs demonstrated 25% faster times to complete inner joins versus 6Gb/s SATA SSDs. This enables analysts to gather spread out data across servers in the cluster much faster for speedier insights and analysis.

Workload 4: Maximum Drive Read Throughput

This workload represents the first of two SSD throughput tests and covers the maximum drive read throughput in megabytes per second (MB/s). It shows the maximum throughput achieved from the drives in the RAID set while the workload was running. The bar graph below shows the maximum read throughput from three test runs for each set of drives. The highest result represents the fastest throughput.

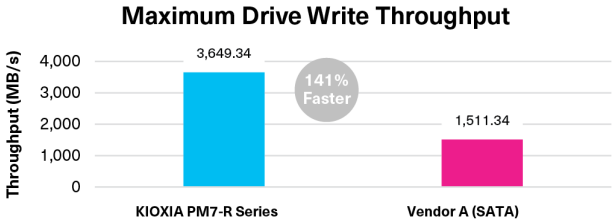


Test Analysis:

The KIOXIA PM7-R Series SSDs demonstrated 190% faster SSD read throughput versus 6Gb/s SATA SSDs. This enables significantly faster data reads into memory for analytic processing, which in turn drives faster insights and analysis.

Workload 5: Maximum Drive Write Throughput

This workload is the second SSD throughput test and covers the maximum drive write throughput (in MB/s). It shows the maximum throughput achieved from the drives in the RAID set while the workload was running. The bar graph below shows the maximum write throughput from three test runs for each set of drives. The highest result represents the fastest throughput.

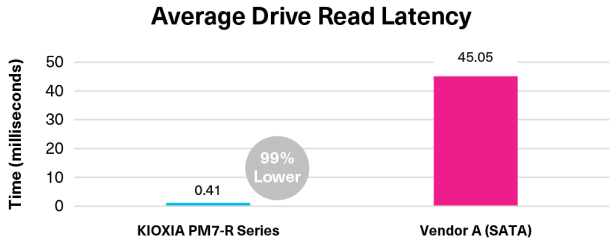


Test Analysis:

The KIOXIA PM7-R Series SSDs demonstrated 141% faster SSD write throughput versus 6Gb/s SATA SSDs. This enables significantly faster data writes from memory especially for rewriting data across the cluster to handle actions such as repartitioning.

Workload 6: Average Drive Read Latency

This workload represents the first of two SSD latency tests and covers the average drive read latency in milliseconds (ms) - the time it took to perform a drive read operation. The bar graph below shows the average read latency from three test runs for each set of drives. The lowest result represents the lowest (or best) latency.

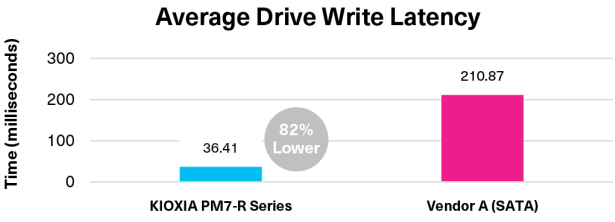


Test Analysis:

The KIOXIA PM7-R Series SSDs demonstrated 99% lower SSD read latency versus 6Gb/s SATA SSDs. This included the average time it took for successful completion of the read operation once the drive received the request issued from the workload generator. This enables analysts to obtain faster insights and analysis.

Workload 7: Average Drive Write Latency

This workload is the second SSD latency test and covers average drive write latency (in ms) - the time it took to perform a drive write operation. The bar graph below shows the average write latency from three test runs for each set of drives. The lowest result represents the lowest (or best) latency.

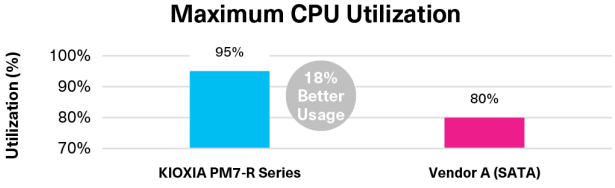


Test Analysis:

The KIOXIA PM7-R Series SSDs demonstrated 82% lower SSD write latency versus 6Gb/s SATA SSDs. This included the average time it took for successful completion of the write operation once the drive received the request issued from the workload generator. This enables analysts to obtain faster insights and analysis.

Workload 8: Maximum CPU Utilization

This workload is the final measurement of this comparison and covers CPU utilization. The test outputted the percentage of the CPU cycles used for a given workload. It was measured to ensure that the server CPUs were being used efficiently while processing the PySpark benchmark[3] workloads. Low CPU utilization indicates that the CPUs were not used efficiently, which could result in an overall underutilization of server capabilities and stranded compute resources. The bar graph below shows the maximum CPU utilization from three test runs for each set of drives. Higher CPU utilization is better.



Test Analysis:

The KIOXIA PM7-R Series SSDs demonstrated 18% better CPU usage versus 6Gb/s SATA SSDs. With high CPU utilization, the benchmark workloads were efficiently processed and the CPUs used their resources better in the PM7 Series configuration.

Summary

TEST METRICS		KIOXIA PM7-R Series 24G SAS SSDs	Vendor A 6Gb/s SATA SSDs
Spark Driver Memory		5 gigabytes[1] (GB)	5 GB
Spark Executor Memory		20 GB	20 GB
No. of Records in Dataset		10 billion	10 billion
Total Size of Dataset Records		2.1 terabytes[1] (TB)	2.1 TB
SSD Time to Complete	*(lower is better)*		
a. Group By and Aggregate Transformations		480.93 sec.	504.27 sec.
b. Repartitions		941.15 sec.	1,136.80 sec.
c. Inner Joins		1,195.58 sec.	1,597.23 sec.
SSD Performance			
a. Maximum Read Throughput	*(higher is better)*	6,369.34 MB/s	2,192.34 MB/s
b. Maximum Write Throughput	*(higher is better)*	3,649.34 MB/s	1,511.34 MB/s
c. Average Read Latency	*(lower is better)*	0.41 ms	45.05 ms
d. Average Write Latency	*(lower is better)*	36.41 ms	210.87 ms
Maximum CPU Utilization	*(higher is better)*	95%	80%

The KIOXIA PM7-R Series lowers the amount of time it takes to perform transformations on large data analytics datasets when compared to 6Gb/s SATA drives. Additionally, these 24G SAS SSDs were able to demonstrate low read and write drive latencies, as well as high read and write throughputs when compared to 6Gb/s SATA drives. The cluster configuration with PM7-R Series SSDs demonstrated improved CPU utilization with a low memory footprint, performing I/O intensive transformations faster than the SATA configuration.

KIOXIA PM7-R Series SSD Product Info

The KIOXIA PM7-R Series of read intensive, dual-port 24G SAS SSDs are available in a 2.5-inch[4] (15 mm Z-height) form factor with capacities up to 30.72 TB at 1 DWPD[5] (Drive Writes Per Day). This 7th generation SAS SSD series features KIOXIA 112-layer BiCS FLASH™ 3D flash memory, power loss protection[6] (PLP) and supports security/encryption options[7].

KIOXIA PM7-R Series[8]

PM7-R Series (24G SAS) SSDs (in dual-port mode)[9]

SPECIFICATIONS		Units	30.72 TB (up to)	15.36 TB (up to)	7.68 TB (up to)	3.84 TB (up to)	1.92 TB (up to)
Sequential Read	*(128 KB; QD=32; 18W)*	MB/s	4,150	4,200	4,200	4,200	4,200
Sequential Write	*(128 KB; QD=32; 18W)*	MB/s	3,200	4,100	4,100	3,650	3,400
Random Read	*(4 KB; QD=256; 18W)*	KIOPS	720	720	720	720	720
Random Write	*(4 KB; QD=32; 18W)*	KIOPS	80	160	175	155	155
Random Read Latency	*(QD=1; 18W)*	µs	80	80	80	80	80
Random Write Latency	*(QD=1; 18W)*	µs	15	15	15	15	15

Additional information on KIOXIA PM7 Series enterprise 24G SAS SSDs is available here.

Appendix A

Hardware/Software Test Configuration

Server Information	
Server	HPE ProLiant DL385 Gen10 Plus v2
No. of Servers	3
No. 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	3.2 GHz
Operating System Information	
Operating System	Ubuntu®
Version	22.04.2
Kernel	5.15.0-73-generic

SSD Information		
Model	KIOXIA PM7-R Series	Vendor A
Interface	24G SAS	6Gb/s SATA
No. of Devices	12	12
Form Factor	U.2 (2.5-inch)	U.2 (2.5-inch)
Capacity	3.84 TB	3.84 TB
DWPD	1 (5 years)	1 (5 years)
Active Power	18 W	9 W

HDFS Information	
Software	Apache Hadoop
Version	3.3.5
Total Capacity	29.1 TB

Apache Spark Information	
Software	Apache Spark
Version	3.3.1

Appendix B

Configuration Set-up/Test Procedures

Set-up

- The HPE ProLiant DL385 Gen10 Plus v2 server was set-up – twelve KIOXIA PM7-R SSDs (3.84 TB capacity each) were also deployed into three servers that formed the cluster (four drives per server).
- The four PM7-R Series SSDs were set-up with a RAID 5 set using the HPE SR932i-p Gen10+Hardware RAID card on each node in the cluster.
- The Ubuntu 22.04.2 OS was set-up on each server.
- The PM7 Series SSDs in the RAID 5 set had a XFS file system placed on top of them for all of the server nodes in the cluster.
- Apache Hadoop software was set-up to use all of these mounted volumes across the cluster to enable participation in the HDFS using a replication factor of three. This resulted in a total capacity size of 29.1 TB for HDFS storage.
- Apache Spark software was set-up with one master and three nodes across the cluster. One node served as both the master and a worker, while the other two servers in the clusters were worker nodes.
- The PySpark lightweight benchmarking tool served as the workload generator to run on the Apache Spark cluster.
- Ten billion records across one thousand partitions was the created dataset - HDFS capacity was 2.1 TB.

Procedures

- The three 'time to complete' workloads were performed including group by/aggregate operations, repartitioning and inner join operations against the large dataset on the HDFS and distributed across the Apache Spark cluster - test results were then recorded.
- The two SSD throughput workloads were tested three times and recorded to determine the maximum drive read and write throughput achieved in the RAID 5 set (while the workload was running).
- The two SSD latency workloads were tested three times and the average time to complete a drive read or drive write operation was determined.
- In conjunction with the throughput and latency tests, the maximum CPU utilization was measured and recorded to ensure that the server AMD EPYC CPUs were efficiently processing the benchmark datasets.
- The entire set-up and procedures were replicated for the twelve Vendor A 6Gb/s SATA SSDs (3.84 TB capacity each) when deployed into the three HPE ProLiant Gen10 Plus v2 servers that formed the cluster.

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] RAID 5 set is a redundant array of independent disks configuration that uses disk striping with parity. The data and parity stripes evenly across all of the disks so no single disk becomes a bottleneck. This type of striping also enables users to reconstruct data in case of a disk failure.

[3] PySpark benchmark is a lightweight benchmarking tool used as a workload generator to run on an Apache Spark cluster - https://github.com/DIYBigData/pyspark-benchmark.

[4] 2.5-inch indicates the form factor of the SSD and not the drive's physical size.

[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, for the specified lifetime. Actual results may vary due to system configuration, usage, and other factors.

[6] Power Loss Protection (PLP): records data in buffer memory to NAND flash memory, utilizing backup power of the solid capacitor in case of a sudden shut down or power outage.

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

[8] The KIOXIA PM7-R Series SSD shown is a representation of the design model and not an accurate product depiction.

[9] KIOXIA PM7-R specifications provided by KIOXIA Corporation and are accurate as of this publication date.

TRADEMARKS:

DISCLAIMERS: