Improve Distributed Storage System TCO with Host-Managed SMR HDDs from Western Digital and Kalista IO for Performance at Scale

KALISTA IO
Western Digital.

Joint solution brief

Highlights

Western Digital and Kalista IO are collaborating to enable Apache Hadoop® and Ceph® with host-managed SMR HDDs

- Host-managed SMR HDDs offer performance consistency, predictability and a better total cost of ownership than CMR HDDs
- Increased storage density from host-managed SMR HDDs leads to CapEx and OpEx savings
- Host management of device state and data placement helps deliver consistent and predictable performance at scale
- Kalista IO Phalanx Storage System transparently supports host-managed SMR and future storage device technologies without application changes nor new kernel module installs

Introduction

Data is at the center of our information-driven economy. It is the foundation for disruptive technologies such as machine learning (ML) and artificial intelligence (AI)[1]. Data is the catalyst for innovation, enabling businesses to improve and create new products, experience and services.

Data is also growing rapidly. The amount of data created globally will increase from 32 zettabytes (ZB) last year to over 100 ZB by 2023[2]. This explosive growth is fueling an insatiable demand for cost effective and performant data storage solutions.

Western Digital's Ultrastar® host-managed shingled magnetic recording (SMR) HDDs write data sequentially to the platter and overlap new tracks on parts of previously written tracks. This results in higher track density (tracks per inch) to enable higher capacity points. The increase in drive capacity leads to lower total cost of ownership (TCO) through fewer devices and fewer servers as well as reduced maintenance, power and cooling costs.

Latency is further reduced when the host takes responsibility and control of device state and data placement, empowering it to reduce tail latency, increase throughput, and manage performance at scale. Reduced latency is an important differentiator[3] to store, analyze, and retrieve data and can have a significant impact on business revenues[4].

However, host-managed SMR HDDs have a more complex and restrictive usage model compared to conventional HDDs. They require hosts to write sequentially, align IOs to device zone boundaries, and actively monitor and set zone states. In addition, host system software and hardware must be able to recognize and support the newly defined host-managed SMR device type and zone management commands.

These requirements for host-managed SMR HDDs introduce random write performance penalties with current applications and systems, requiring users to modify existing applications and/or install additional kernel modules. These write penalties can be solved using a software solution to sequentialize the incoming data stream.

Kalista IO enables host-managed SMR solutions for Western Digital products that simplify the approach for the deployment of distributed storage. Western Digital and Kalista IO help customers capitalize on the benefits of reliable, performant, and cost effective host-managed SMR storage without the complications of application or kernel changes.

Kalista Phalanx and Ultrastar DC HC620 SMR HDD for Optimal TCO and Performance at Scale

Ultrastar DC HC620 delivers a capacity-optimized data center storage solution with SMR technology. Built for data center class workloads, Ultrastar DC HC620 is ideal for dense scale-out storage systems. It delivers the uncompromising product reliability necessary for private and public cloud enterprise applications. Ultrastar DC HC620 is built on the proven and mature HelioSeal® platform to deliver an exceptional watts/TB power footprint for online storage.

Kalista IO Phalanx Storage System is an intelligent storage system built for software-defined environments and next generation storage devices. It is designed to enable applications and systems to use host-managed SMR devices without modifications and to perform at scale (Figure 1).

Phalanx uses a multilayered architecture (Figure 2):

1. Data access layer provides file, object and block interfaces for applications to store and retrieve data
2. IO engine layer calculates optimal data placement and intelligently prioritizes incoming requests
3. Device management layer optimizes IO requests for each device and manages its state

Designed from the ground up to be device friendly, Phalanx eliminates random writes and evenly distributes data across capacity to prevent hot areas. Accordingly, Phalanx is intrinsically well-suited to work with conventional and host-managed SMR HDDs. Phalanx is engineered to deliver consistent and predictable performance at scale. An innovative row and column architecture scales performance with capacity while minimizing contention. Client requests are intelligently prioritized for fairness and quality of service (QoS). And metadata and data are separated to increase scalability, efficiency and robustness.

Most importantly, Phalanx is designed to fit easily into users' workflows and environments (Figure 3). With a unified file/object/block interface and an implementation that is completely in userspace, Phalanx works without application changes nor additional kernel modules. It can be easily containerized/virtualized to fit within existing orchestration and virtualization frameworks. In sum, Phalanx has minimized the friction and disruption to deploying host-managed SMR and future storage device technologies.

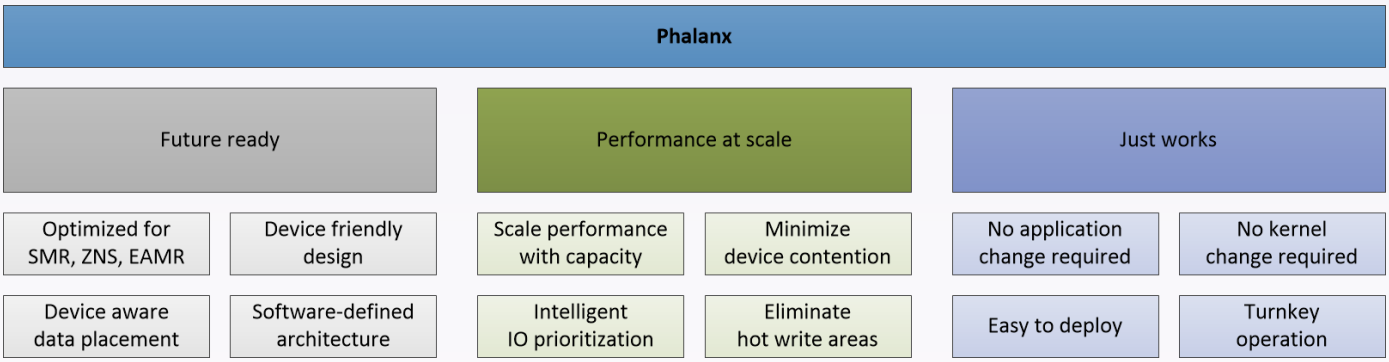


Figure 1 - Phalanx design principles

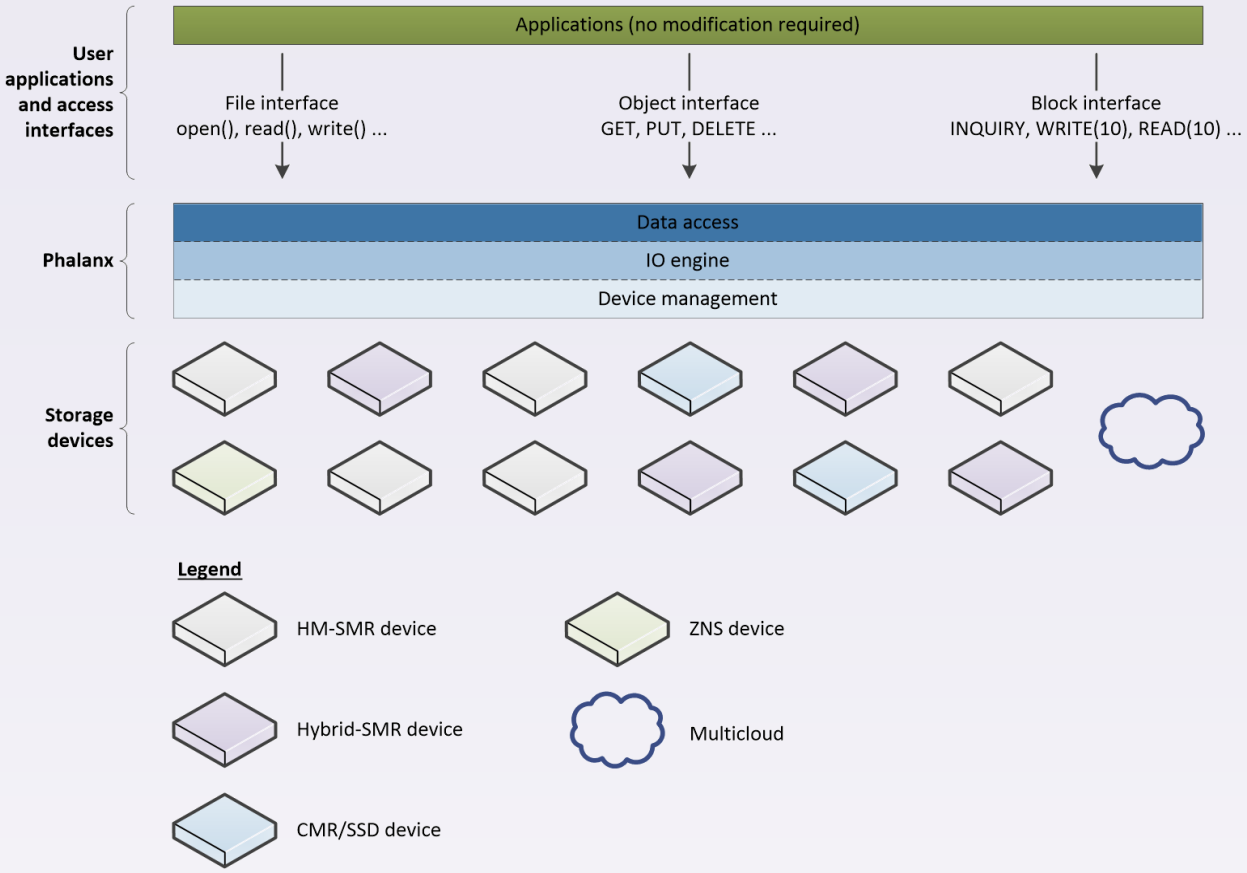


Figure 2 - Phalanx software stack

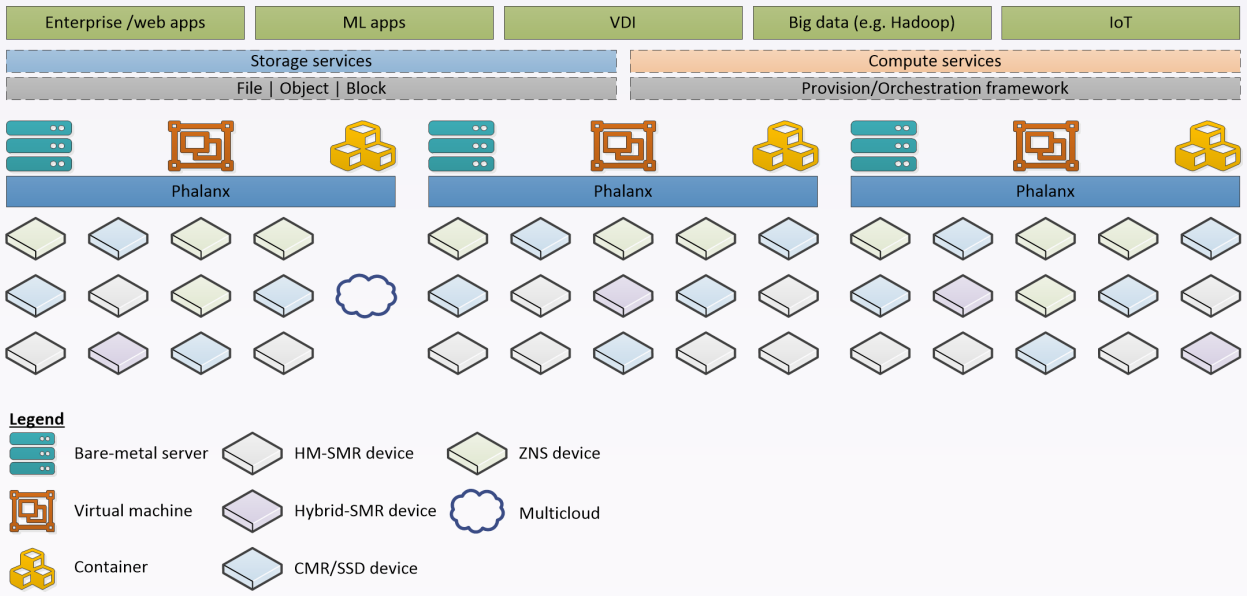


Figure 3 - Phalanx usage scenarios

Best uses

- Workloads requiring robustness and scalability of distributed scale-out storage systems
- Workloads and scenarios needing consistent and predictable performance at scale
- Hyperscale cloud and traditional data center workloads
- Software-defined infrastructures
- Storage scenarios requiring highest density and lowest TCO

Solution description

Distributed storage systems provide fault tolerance, scalability, and performance needed to answer the challenges of big data. Hyperscalers to enterprises use systems like Apache Hadoop and Ceph to rapidly store and process large data sets and make them readily accessible. These systems are perfect use cases to benefit from host-managed SMR's reduced TCO and increased performance consistency.

Kalista IO has successfully enabled distributed storage systems with Western Digital Ultrastar host-managed SMR HDDs while minimizing friction to deployment (Figure 4).

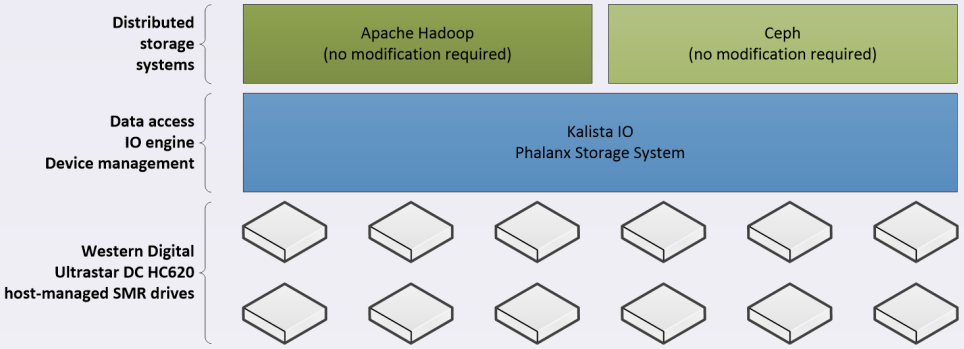


Figure 4 - Enabling Apache Hadoop and Ceph with Western Digital's Ultrastar DC HC620 host-managed SMR drives

The advantage for customers is threefold:

1. Lower TCO and performance optimizations of host-managed SMR
2. Robustness and scalability of distributed scale-out storage systems
3. Simple turnkey operation to use host-managed SMR HDDs

Using Kalista IO Phalanx Storage System, we integrated Western Digital's Ultrastar DC HC620 SMR HDDs into Apache Hadoop and Ceph seamlessly without application nor kernel modifications. And as expected, we found performance enhancements in both Apache Hadoop and Ceph on our test system (Figure 5) running Phalanx with host managed SMR HDDs as compared to legacy file systems (e.g., XFS/ext4) with CMR HDDs. We observed a 58% increase in average input/output operations per second (IOPS) as well as a 10x gain in performance consistency with Ceph Rados write bench, and a 19% higher average throughput with Hadoop TestDFSIO read bench (Figures 6 - 8) [5].

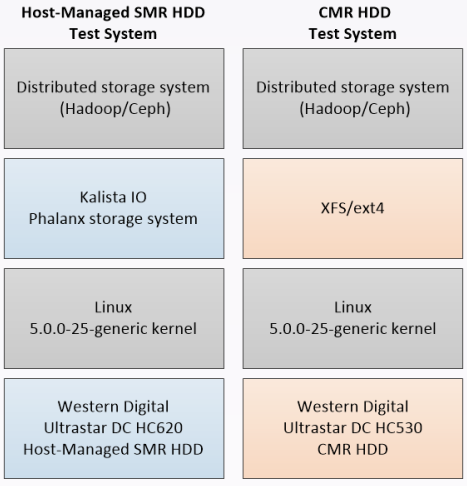
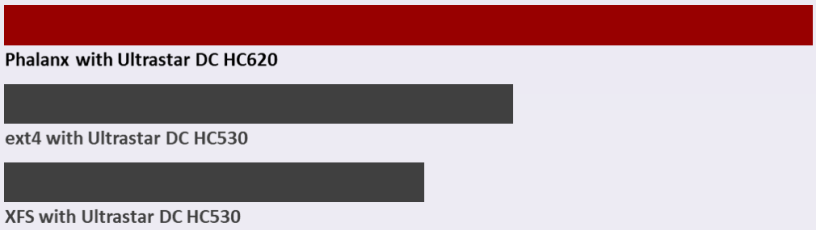


Figure 5 - Test system storage stack

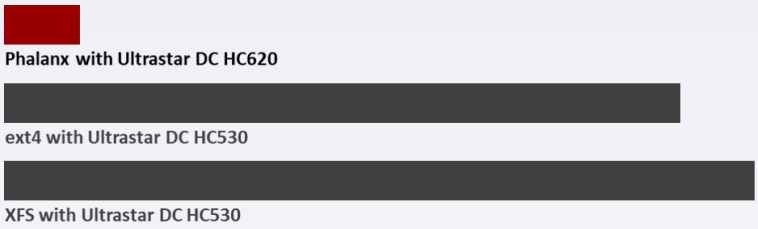
Ceph Rados bench average write IOPS (4MB object and block size, 16 threads)



58%
higher average write IOPS

Figure 6 - Testing conducted by Kalista IO in August 2019 using preproduction Phalanx software with Linux kernel 5.0.0-25-generic, and Intel Core i7-4771 CPU 3.50GHz with 16GiB DDR3 Synchronous 2400 MHz RAM, and Western Digital Ultrastar DC HC620 host managed SMR and Ultrastar DC HC530 CMR drives connected through SATA 3.2, 6.0 Gb/s interface. Tested with Ceph version 13.2.6 Mimic in single node mode with single object replica. Rados write bench ran with 4MB object and block (op) size with 16 concurrent operations for 1800 seconds to capture average and standard deviation IOPS values[6].

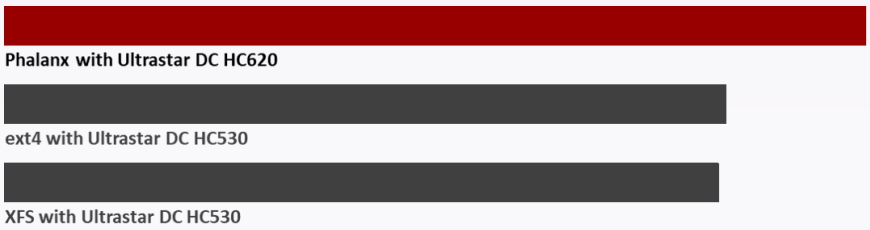
Ceph Rados bench write IOPS standard deviation (4MB object and block size, 16 threads)



10x
better write performance consistency

Figure 7 - Testing conducted by Kalista IO in August 2019 using preproduction Phalanx software with Linux kernel 5.0.0-25-generic, and Intel Core i7-4771 CPU 3.50GHz with 16GiB DDR3 Synchronous 2400 MHz RAM, and Western Digital Ultrastar DC HC620 host managed SMR and Ultrastar DC HC530 CMR drives connected through SATA 3.2, 6.0 Gb/s interface. Tested with Ceph version 13.2.6 Mimic in single node mode with single object replica. Rados write bench ran with 4MB object and block (op) size with 16 concurrent operations for 1800 seconds to capture average and standard deviation IOPS values[6].

Apache Hadoop TestDFSIO average read throughput (32 files, 16GB each)



19%
faster average read throughput

Figure 8 - Testing conducted by Kalista IO in August 2019 using preproduction Phalanx software with Linux kernel 5.0.0-25-generic, and Intel® Core™ i7-4771 CPU 3.50GHz with 16GiB DDR3 Synchronous 2400 MHz RAM, and Western Digital Ultrastar DC HC620 host managed SMR and Ultrastar DC HC530 CMR drives connected through SATA 3.2, 6.0 Gb/s interface. Tested with Apache Hadoop version 3.2.0 in single node pseudo-distributed mode with single block replica, and TestDFSIO version 1.8 on OpenJDK version 1.8.0_222. TestDFSIO read benchmark ran with 32 files, 16GB each for a 512GB dataset. Executed 3 times to capture average and standard deviation throughput values[7].

Conclusion

Western Digital and Kalista IO are collaborating to enable, simplify and optimize host-managed SMR systems. Western Digital's Ultrastar host-managed SMR HDDs offer performance consistency, predictability and a better TCO than conventional HDDs. When integrated with Kalista IO's Phalanx Storage System, performance scales with capacity and users can use existing applications and kernels without modification.

Learn more at http://www.kalista.io and https://www.westerndigital.com/products/data-center-drives/ultrastar-dc-hc600-series-hdd

References

[1] R. Bean, "How Big Data Is Empowering AI and Machine Learning at Scale," MIT Sloan Management Review, 8 May 2017. [Online]. Available: https://sloanreview.mit.edu/article/how-big-data-is-empowering-ai-and-machine-learning-at-scale/.

[2] D. Reinsel and J. Rydning, "Worldwide Global DataSphere Forecast, 2019–2023: Consumer Dependence on the Enterprise Widening," IDC, 2019.

[3] J. Dean and L. A. Barroso, "The Tail at Scale," Communications of the ACM, vol. 56, no. 2, pp. 74-80, 2013.

[4] U. Hoelzle, "The Google Gospel of Speed," Google, January 2012. [Online]. Available: https://www.thinkwithgoogle.com/marketing-resources/the-google-gospel-of-speed-urs-hoelzle/.

[5] Testing conducted by Kalista IO in August 2019 using preproduction Phalanx software with Linux kernel 5.0.0-25-generic, and Intel Core i7-4771 CPU 3.50GHz with 16GiB DDR3 Synchronous 2400 MHz RAM, and Western Digital Ultrastar DC HC620 host managed SMR and Ultrastar DC HC530 CMR drives connected through SATA 3.2, 6.0 Gb/s interface.

[6] Kalista IO, "Phalanx Ceph OSD and Rados benchmarks," [Online]. Available: http://kalista.io/resources/performance/phalanx-ceph-benchmarks.pdf.

[7] Kalista IO, "Phalanx Hadoop TestDFSIO benchmarks," [Online]. Available: http://kalista.io/resources/performance/phalanx-hadoop-benchmarks.pdf.

KALISTA IO

Building intelligent compute and storage systems designed and optimized for software-defined environments and next generation storage devices

http://www.kalista.io
@kalista.io
info@kalista.io