The Future of High-Performance Storage

Published 24 March 2017 - ID G00325069

A culmination of forces — represented by new nonvolatile memory and better software — are poised to alter the price/performance of next-generation compute and storage environments. We advise I&O leaders on how to both benefit from and mitigate risks for this unproven, fast-evolving market change.

Overview

Impacts

- Storage-class memory technologies could alter price/performance and density of current storage and compute platforms while bringing both benefits and uncertainties to I&O leaders responsible for IT infrastructure planning.
- Faster interfaces between compute and storage that leverage parallel access can reduce storage access latencies, but may require I&O leaders to explore significant additional investments or upgrades, which, in turn, disrupt enterprise IT refresh cycles.
- Memory-centric key value stores and file systems are in early exploratory phases in the enterprise and have potential performance benefits, but I&O leaders have viability concerns for mainstream adoption.

Recommendations

Infrastructure and operations (I&O) leaders responsible for infrastructure agility should:

- Demand emerging memory technology roadmaps with guarantees to ensure scalability and resiliency of the technology and to validate the initial performance and timeline of potential cost reductions. Expect that roadmap timelines will slip.
- Use Nonvolatile Memory Express (NVMe) over Peripheral Component Interface Express (PCIe) to deliver low-latency access to in-memory databases and real-time analytics applications that can't tolerate shared storage latencies.

- Prepare for NVMe over Fabric (NVMeoF) by insisting that your incumbent hardware OEM vendors support it within the next 18 to 24 months.
- Identify target use cases to leverage memory-centric storage software, but use standard front-end interfaces (such as HDFS) to reduce vendor lock-in and lower exit costs.
- Invest in new nonvolatile memory technologies in a tactical manner by focusing on short-term ROI, given that this is an early-stage market where vendor claims are largely unproven in real production deployments.

Strategic Planning Assumptions

3D Xpoint technology will be at least seven times more expensive than NAND flash on a $/GB basis through 2021 and won't be generally available for enterprise applications until late 2018.

The slow pace of innovation in storage software and ecosystem will inhibit adoption of nonvolatile dual in-line memory modules (NVDIMMs) based on new nonvolatile memory technologies through 2020.

Market consolidation will see more than 50% of emerging vendors exiting the NVMeoF appliances or software-defined storage markets through 2022.

Analysis

The CPU advancements of the past three decades, driven by Moore's Law and through the proliferation of multicore processors, have created a significant imbalance in the data center IT infrastructure. While the growth in CPU performance has been near-linear over the past few decades, the disk drives barely grew in performance in the same time period. Even with the advent of new media such as NAND flash for persistent storage, CPU utilization still continues to be low, due to legacy protocols (such as SATA and Serial Attached SCSI [SAS]), as well as additional overheads created by traditional storage software layers that are inhibiting storage performance.

In the past 12 months, there has been tremendous hype about new advancements that include new forms of nonvolatile memory such as Z-NAND and 3D Xpoint, as well as growing adoption of newer interfaces such as NVMe and more modern software that seek to reduce this imbalance by promising low-latency, high-performance data access to applications.

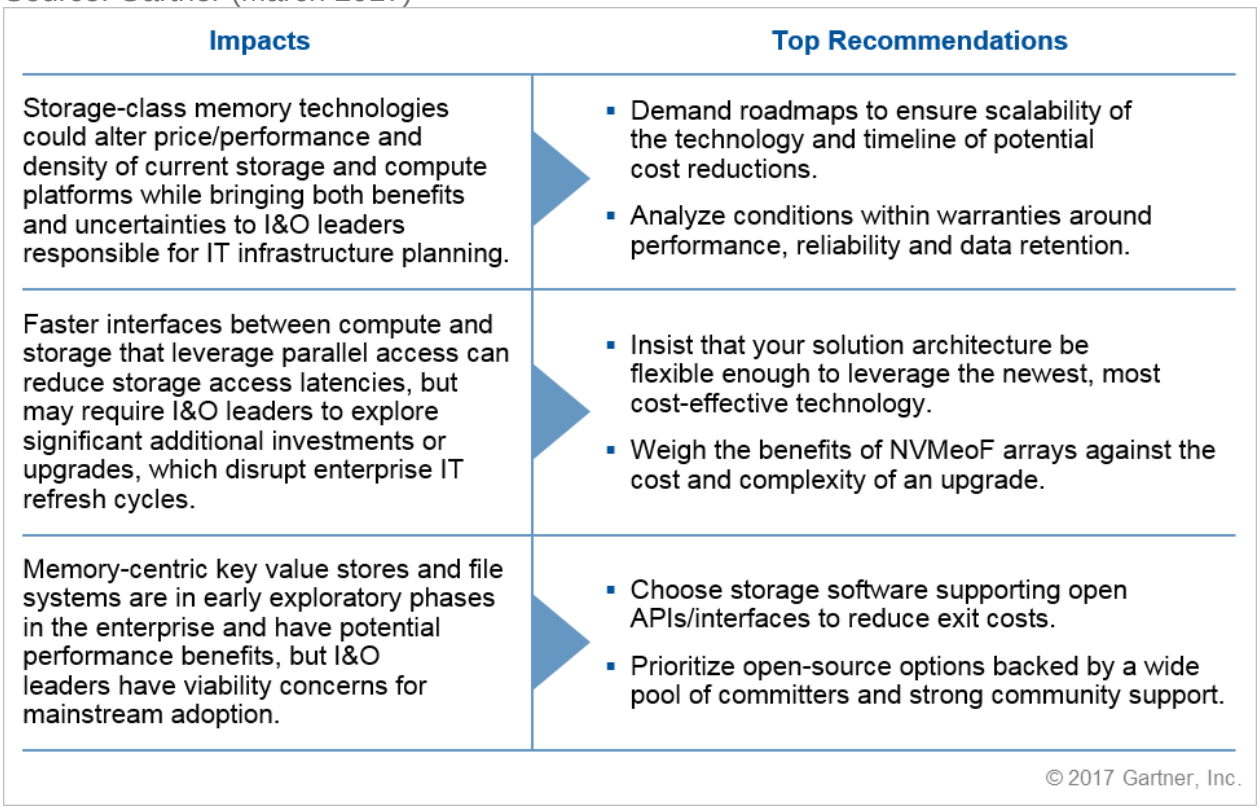
There are three categories of applications where Gartner sees the demand for these innovations:

1. SQL/NoSQL data stores — SQL-based in-memory databases, such as SAP Hana; NoSQL-based in-memory data grids, such as Redis and Memcached; and document and table style databases, such as MongoDB and Apache Cassandra.
2. High-performance computing (HPC) workloads— HPC technical and commercial workloads in academia, research, hyperscale cloud and enterprise organizations.

3. Big data applications — These include Apache Spark-based or Apache Storm-based applications for real-time analytic processing.

While these innovations have the potential to bridge the compute-storage performance gap in the data center, customers need to be wary about the hype surrounding these new technologies, and should strive to protect their investments by seeking warranties from vendors before deploying them in production. The cumulative effects of these technologies have the potential to be far-reaching; however, they are in a nascent stage today, and that immaturity and lack of proven deployments should be carefully weighed against the potential benefits.

Figure 1. Impacts and Top Recommendations for Storage Performance Innovations
Source: Gartner (March 2017)



Impacts and Recommendations

Storage-class memory technologies could alter price/performance and density of current storage and compute platforms while bringing both benefits and uncertainties to I&O leaders responsible for IT infrastructure planning

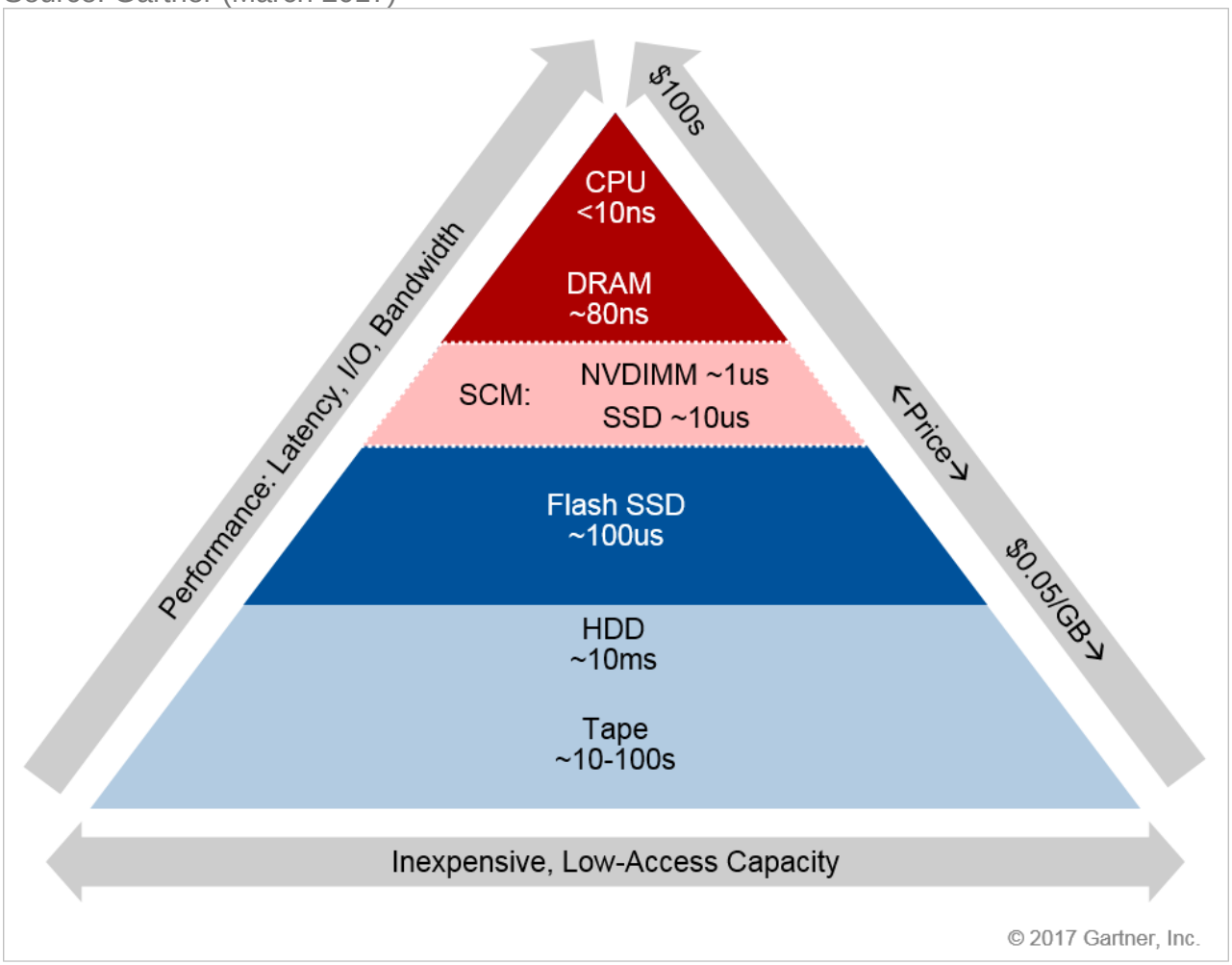
Many emerging memory technologies have strived to marry high performance and cost-efficient density with nonvolatility for persistence in order to be designated as storage-class memory. [1] Most memory technologies fail in this pursuit, and every one of them has taken longer to mature than initial expectations.

Today's challenge is to rival the technology and massive manufacturing scale of the DRAM market (server memory) and NAND flash markets (performance storage) by introducing a new tier of storage-class memory that capably blends the attributes of DRAM and NAND with minimal compromise. The goal is to provide a scalable, persistent memory technology that can achieve access speeds closer to CPUs for only a few dollars per GB, as shown in Figure 2.

Figure 2. Price/Performance Hierarchy

HDD: hard-disk drive; I/O: input/output; SSD: solid-state drive

Source: Gartner (March 2017)



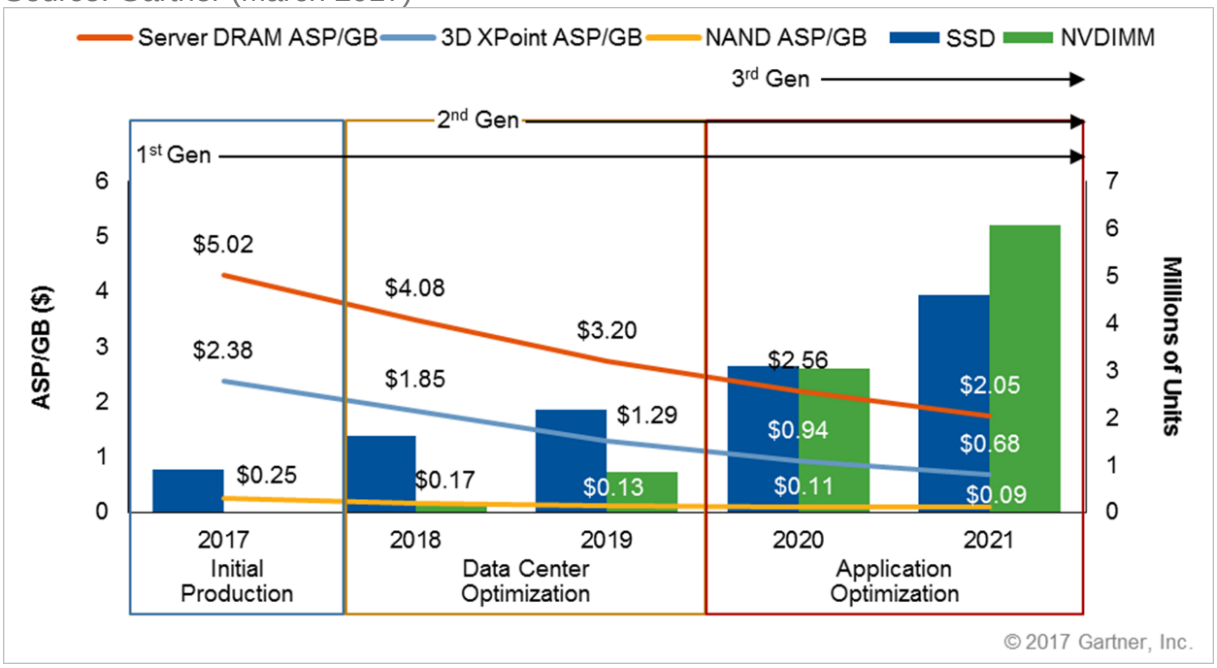
In 2015, Intel and Micron Technology announced a new memory technology called 3D XPoint (cross point), and after initial delays in 2016, both companies have stated that the technology will recognize revenue in 2017. Intel/Micron claims 3D XPoint technology can achieve up to 1,000 times write performance improvement and considerable endurance benefits over NAND flash technology, and with lower-power consumption. While 3D XPoint technology details have been kept mostly secret, the bit storage is based on a change of bulk resistance technology that is done in a grid-access array capable of being stacked in layers (two layers have been announced as of this publication). This transistorless approach allows for cells to be individually programmed, leading to fast read/write speeds in a dense 16GB die (3D structure with two layers).

The manufacturing process is believed to have considerable overlap with existing NAND flash manufacturing processes, which will be important to consistently deliver high-volume manufacturing at a compelling cost structure of around $1 to $2 per GB, initially, if sold as a discrete chip. However, most 3D XPoint shipments will be sold in a usable product form, which will command substantial margins and would be marked up further if sold within a solution to an end user. While commercialization has only just started and yields are not known, we believe that mainstream adoption is really in 2019 and beyond, as the coming years will be foundational for most customers. 3D XPoint technology will come in two forms: first, as storage in the form of NVMe PCIe SSD technology, then followed by NVDIMM. Claiming to be fast storage in data centers is much easier than being directly addressable as a NVDIMM, considering the significant optimizations that are required within the host server and application layers. Figure 3 shows Gartner's estimation of the evolution of 3D XPoint through 2021 in terms of usable price per GB to customers and corresponding volume shipments of SSDs and NVDIMMs.

Figure 3. Evolution of 3D XPoint

Note: 3D XPoint ASP/GB is a Gartner estimate of usable capacity as sold to OEM customers.

Source: Gartner (March 2017)



3D XPoint is not the only emerging memory technology under development in recent years:

- In August 2016, Samsung announced its ZNAND technology as a flash-based derivative to rival 3D XPoint in the near term.
- Memristors, also known as resistive RAM (ReRAM), have been under development for years, with memory makers (including Western Digital, Toshiba and SK hynix) licensing the technology from HPE. While memristors appear compelling, high-volume manufacturing is always a high barrier to overcome. Thus, this is more likely a 2020 technology.

Regardless of which storage class memory they select, I&O leaders are advised to:

- Demand emerging memory technology roadmaps with guarantees to ensure scalability of the technology and timeline of potential cost reductions. Expect that roadmap timelines will slip until proven otherwise.
- Test products through diverse and extreme performance workloads and over extended periods of time.
- Assess security requirements in terms of encryption and secure erase due to the persistence of data in both the NVDIMM and the SSD.
- Analyze any conditions within warranties around performance, reliability and data retention. Request case studies and references that are similar to your real-world application workloads.

Faster interfaces between compute and storage that leverage parallel access can reduce storage access latencies, but may require I&O leaders to explore significant additional investments or upgrades, which, in turn, disrupt enterprise IT refresh cycles

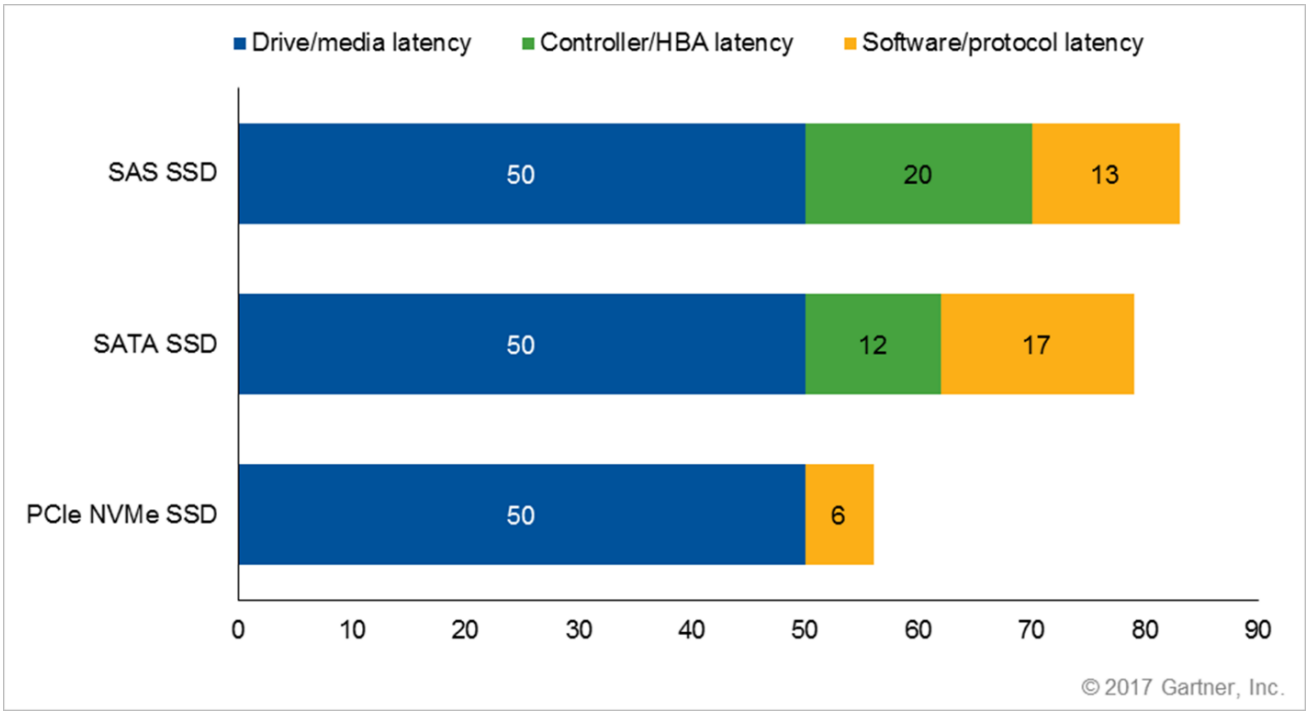
Nonvolatile Memory Express (NVMe or NVM Express) is the most recent host controller interface specification that allows solid-state storage devices to take full advantage of high-speed PCIe bus. NVMe specifications are developed by an industry consortium, NVM Express, with more than 80 members. They are now working on version 1.3, utilizing a standards-based approach that enables broad ecosystem adoption, compatibility and interoperability. SSDs are inherently capable of much higher transfer speeds at lower latencies, but without an optimized software interface and drivers, they could not reach their full potential. NVMe has specifically emerged to take advantage of the parallelism and low latency of PCIe as the physical interface. It offers optimized command and completion paths for use with NVMe-based storage devices.

Figure 4 provides a comparative overview of SSD latencies across different interface types.

Figure 4. I/O Request Latency Breakdown

HBA: host bus adapter

Source: Gartner (March 2017)



While NVMe introduces significant performance improvements in back-end storage systems, it is limited to the distance of PCIe bus and can only address a small number of devices connected to it. I&O leaders have already begun adopting NVMe drives in servers and are exploring extending the NVMe technology to remote storage subsystems. NVMeoF is a new data center protocol that can run over Ethernet, InfiniBand or Fibre Chanel, and extends low-latency NVMe connected storage to create large scale-out clusters of NVMe-enabled storage systems and hosts. The goal of NVMe over Fabrics is to provide connectivity to remote NVMe devices with no more than 10 microseconds (µs) of additional latency over a native NVMe device inside a server node. NVMeoF specifications were released in June 2016, and now provide industry standards to build enterprise products with large pools of shared NVMe storage scaling independently of compute resources with centralized management and significantly lower network latency when compared to traditional storage area networks (SANs, such as Fibre Channel or internet SCSI). Eventually, all hybrid storage arrays, solid state arrays and just a bunch of flash (JBOF), which use SSDs, will move to NVMeoF to exploit the performance improvements. The cost of NVMe SSDs is a major barrier to adoption today, but at the end of 2017, we expect dual-port NVMe drives to be comparable in price to their SAS counterparts. I&O leaders should expect NVMeoF-enabled storage products to become more mainstream in the next five years as a stand-alone hardware product (that is, as a solid-state array [SSA]), as software-defined storage, or as part of a converged or hyperconverged integrated systems (HCIS) infrastructure offering. [2](See Note 1 for examples of vendors in this area.)

I&O leaders should:

- Insist that the solution architectures that you are choosing are flexible enough to leverage the newest and most cost-effective technology.

- Leverage NVMe SSAs for low-latency applications, but expect them to be priced at a premium until the economics make sense for the broader range of enterprise applications.
- Weigh the benefits of NVMeoF arrays against the cost and complexity of performing an upgrade of your existing network infrastructures.

Memory-centric key value stores and file systems are in early exploratory phases in the enterprise and have potential performance benefits, but I&O leaders have viability concerns for mainstream adoption

Significant progress is being made in storage-class memory and high-speed interfaces that can deliver latencies in tens of microseconds. However, the traditional storage software stack continues to be a limiting factor in harnessing the full capabilities of the media and interface innovation due to its optimization for serial devices. Traditional storage software uses inefficient locking mechanisms for multithreading, which limit its ability to do parallel I/O tasks at scale. Moreover, they add complex, higher-level abstractions, such as blocks and pages, rather than addressing data at a byte level (which 3D XPoint is capable of). For storage-class memory applications to function effectively, the storage software needs to become more lightweight with low overhead and optimized for parallel access at better read/write granularity.

The OS, hypervisor and storage software to support storage-class memory is in early stages of evolution, which will further inhibit market adoption. The full potential of these innovations is expected to be harnessed over the next two years. We see the following developments as holding initial promise:

- Microsoft has announced that it is working on supporting NVDIMM devices natively through modifications to its file system in Windows. Creation of a new NVM library, pmem.io, for Linux, which would provide a byte addressable interface for storage class memory.
- Emergence of memory-centric file systems, such as Alluxio (formerly Tachyon) that can persist and protect data in memory for fast read/write access.
- Creation of the Apache open-source Distributed Asynchronous Object Storage (DAOS) project that will exploit NVM devices and NVMeoF.

Recommendations:

- Formulate risk mitigation steps, given that storage software support for new NVM technologies is in a nascent stage, with an unclear timeline of availability and maturity.
- Choose storage software that supports open APIs/interfaces (such as Hadoop Distributed File System [HDFS]) to reduce exit costs and to eliminate the risks of vendor viability.

- Prioritize open-source options that are backed by a wide pool of committers and a strong community over proprietary options and single-vendor-backed open-source projects.
- Seek guarantees around timelines and roadmaps from your incumbent OS and storage vendors on their support for storage-class memory media and interfaces.

Evidence

This research is based on more than 50 inquiries handled by the authors with device manufacturers, OEMs and IT leaders. Detailed discussions were conducted with Intel and several other OEMs represented in this research.

Intel: "DAOS: An Architecture for Extreme Scale Storage"

Intel and Micron Technology: "A Revolutionary Breakthrough in Memory Technology"

Gartner: "Emerging Technology Analysis: The Future and Opportunities for Next-Generation Memory"

[1] Storage-class memory is a classification of emerging memory technologies that has read/write performance closer to DRAM, but is persistent and thus able to store information without power. While storage-class memory has greater endurance than NAND flash technology, it also comes with a premium on a cost-per-GB basis, but below that of volatile memory such as DRAM. Storage-class memory can be addressed at either the byte or block level, allowing it to be treated as a storage device or, perhaps, as memory in the future.

[2] HCIS systems will now be able to benefit from low-overhead NVMeoF protocol between cluster nodes to create low latency and high-performance backplane storage pools. This connects NVMe direct-attached storage drives with less than 5% of performance degradation, compared to local NVMe SSDs. While adoption today is very low, within two years, NVMe combined with NVMeoF will be in the top tier of storage, as it takes full advantage of the inherent speed and low latency of flash while complementing the potential of multicore processors that can support the latest applications. The most prominent use cases for shared accelerated storage will be online transaction processing (OLTP) databases, data mining, real-time analytics, high-performance computing applications for video editing, financial processing and analysis, online trading, oil and gas exploration, genomic research, and fraud detection.

Note 1 Sample Vendors of NVMeoF Arrays or Software

Examples of vendors delivering NVMeoF arrays or software include:

- Apeiron Data Systems
- Cray

- DDN
- E8 Storage
- FlashGrid
- Dell EMC
- Excelero
- Mangstor
- Pavilion Data Systems
- Plexistor
- Weka.IO
- X-IO Technologies
- Zsto