HYPERION RESEARCH

Technology Spotlight

Flash Storage Trends and Impacts

Sponsored by HPE

Bob Sorensen, Earl Joseph, Steve Conway, and Alex Norton
April 2018

EXECUTIVE SUMMARY

Advances in 3D NAND chips and subsequent NAND chip architectures, coupled with the flash-based solid-state disks (SSDs) devices that use them, bode well for the overall SSD sector and particularly in high performance storage system. Flash-based SSDs offer significant advantages over hard disk drives (HDDs) from a speed, power, and increasingly, total cost of ownership perspective.

Flash-based SSDs will continue to evolve and be an increasingly critical element in driving performance gains in the memory/storage hierarchy of systems that have a wide range of use cases and that can accelerate performance when operating either on-prem, in the cloud, or as part of a hybrid system architecture.

The growth of enterprise SSD use over the past five years has been impressive. Between 2012 and 2017, the value of enterprise-class SSD revenues grew from $3.0 billion to $11.4 billion with a CAGR of over 30%, based on recent industry data from Gartner, The Register, and Stifel. The rise of enterprise SSD usage is impressive when compared with counterpart HDD storage. Based on recent industry data available from Gartner, The Register, and Stifel in 2017, the value of enterprise-class SSDs sold exceeded the value of enterprise HDDs sold for the first time.

The rapid emergence of SSDs in recent years has led to the development of a wide range of memory/storage innovations that are working their way into data centers of both on-prem and in the cloud. In addition, the availability of flash-based SSDs has enabled a wide range of innovative solutions that add to the overall value and applicability of the flash-based SSD. These include:

- Intel's' Optane memory accelerator/expander scheme that seeks to offer DRAM-like performance with the power and non-volatility of flash
- Software Defined Storage, such as that from WekaIO, that are designed to seamlessly integrate SSD and HDD to optimize both performance and cost
- NVMe over Fabric that seeks to expand the domain and reach of SSD devices across wide networks

These and other NAND/SSD developments will contribute to the growing adoption of SSD and ensure that computer designers and users have the best available options when choosing an efficient memory/storage hierarchy.

INTRODUCTION

NAND Flash and SSDs: Bridging the Gap between DRAM and Spinning Disks

Computer designers spanning the range of IT equipment from desktop PCs to leadership-class HPCs are increasingly looking for ways to improve system performance through the use of more sophisticated memory/storage hierarchies. These designers, however, face considerable design complexities in price/performance trade-offs.

In general, lower memory capacity devices with faster access times and higher data bandwidth, such as DRAMs, are the most costly elements in a memory/storage hierarchy but can have one of the most direct impacts on the overall performance of a computer. Conversely, magnetic hard disk drives (HDDs) are significantly cheaper per unit of storage than semiconductor-based counterparts and can offer significantly more storage capability than DRAMs, albeit with slower data access times and reduced data bandwidth capabilities.

In the past few years, technology advances in the field of NAND flash devices have engendered a new memory/storage technology (solid state disks aka SSDs), that from a systems designer's perspective, can bridge the gap between DRAM and HDDs in terms of both performance and cost metrics.

- At its most basic level, NAND flash is a type of nonvolatile (NV) semiconductor memory device where data is safely stored even when the memory device is not electrically powered. In contrast, DRAMs need a continual supply of power in order to keep stored data intact. NAND chips and the schemes they are packaged in directly relate to whether they are used as SSDs that work in concert with HDDs or as augmented main memory used as a complement to DRAM-based memory.
- NAND devices are not, however, the only device option currently used in SSDs. One notable example is the Intel Optane SSD, which uses a proprietary high-performance technology called 3D XPoint that is non-volatile like NAND counterparts but that enables much better endurance rates and faster data access times. In addition, each storage cell in a 3D XPoint chip, unlike the case in DRAMs, does not require a switching transistor at each storage location, allowing for an 8X to 10X storage density improvement over DRAMs.

With the emergence of lower cost and greater storage capacity flash-based devices, IT designers and users are increasingly turning to NV flash-based SSDs to optimize data access and transfer times within their overall memory/storage hierarchy. Flash-based SSDs offer a number of benefits:

- <u>Speed</u>: the primary benefit of using SSDs is speed of up to 100 times the performance of HDDs in terms of latency and IOPs (or I/O per second). This translates to faster boot times, quicker file transfers, and greater bandwidth for enterprise computing.
- <u>Lifetime Cost or Total Cost of Ownership (TCO):</u> Though the cost for hardware is lower with HDDs, cost for bandwidth (in terms of $/IO) can be much lower with SSDs. This is often a more important measure for enterprises than simple drive capacity per dollar ($/GB). Enterprise SSDs are also expected to have a longer lifespan, making the cost savings add up in the long run.
- <u>Power Consumption</u>: Typically, SSDs consume less power than HDDs. This leads to cost savings, as well as decreased heat generation during operation when compared with HDDs.

The Rise of NVMe

In the early days of SSDs, between 2005 and 2010, most SSDs used established communication buses such as SATA or SAS for interfacing with the rest of the computer system. However, because those protocols were designed for the performance specifications of HDDs, over time they became increasingly inadequate to meet the faster access times and higher data rates typical for SSDs.

Efforts to develop a new standard specifically targeted for SSDs was released in 2011. This new host control interface standard, called Non-Volatile Memory Express (NVMe), was an open logical device interface designed from the ground up to capitalize on the low latency and internal parallelism of flash-based storage devices. In essence, NVMe was a specification that allowed an SSD to make effective use of a high-speed PCI Express bus in a computer. Currently, NVMe works on a variety of form factors.

The NVMe power range spans full-power enterprise devices down to low-power mobile devices and has wide support across most major operating systems including Linux, Windows, Unix, and VMware. The performance advantage of NVMe over both spinning disk and SATA-based SSDs is impressive, as can be seen in Table 1.

TABLE 1

HDD vs. SATA vs. NVMe

Interface	PCIe		SATA 6 Gb/s
Protocol	NVMe Protocol	AHCI Protocol	
	(Optimized for flash SSDs)	(Optimized for mechanical HDDs)	
Bandwidth	PCIe x2 or x4 lane	PCIe x2 or x4 lane	SATA 6 Gb/s
Form Factor	M.2/PCIe Expansion Card/U.2	M.2/PCIe Expansion Card	M.2/2.5" SSD
Max. Read Performance	>3000MB/s	>2000MB/s	>500MB/s
Max. Write Performance	>2000MB/s	>1500MB/s	>500MB/s

Source: Hyperion Research 2018

STRONG MARKET GROWTH OF FLASH-BASED SSD STORAGE

The growth of enterprise SSD over the past five years has been impressive. Between 2012 and 2017, the value of enterprise-class SSD revenues grew from $3.0 billion to $11.4 billion with a CAGR of over 30%, according to recent industry data available from Gartner, The Register, and Stifel. This growth was driven by a combination of factors:

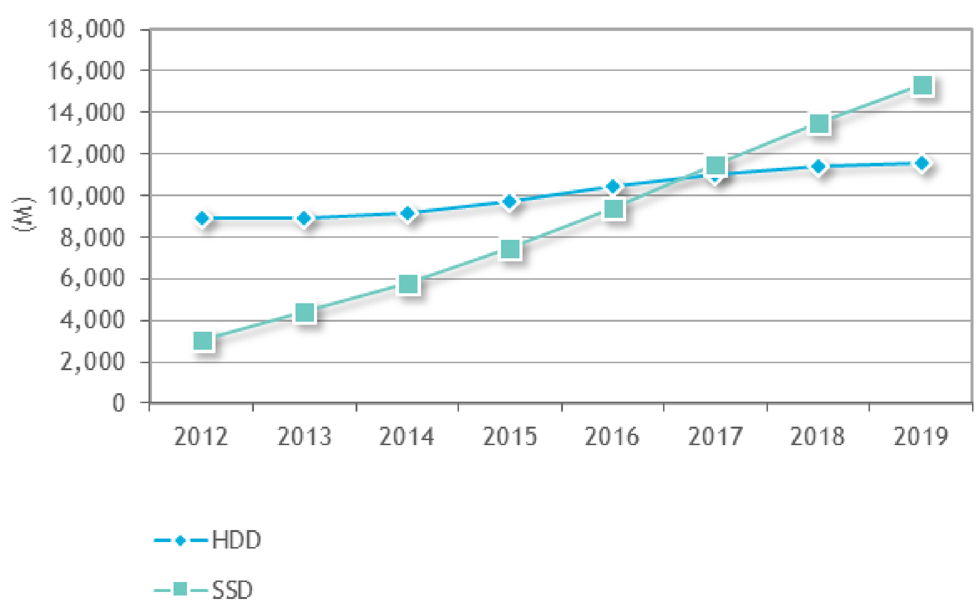
- A strong and highly competitive supplier base for the NAND chips essential to SSD storage units.
- Significant technology advancements from both NAND flash vendors and SSD developers that addressed some of the inherent technical weaknesses of NAND chips (long write times and a limited lifetime for read/write cycles).
- A growing set of use cases within both the technical and enterprise user base for faster, more responsive memory/storage hierarchies to provide data to increasingly demanding multicore CPU and many core GPU processors designs.
- And the rise of big data analytics applications, many with real-time, unique, and both structured and unstructured data access and analysis requirements.

The rise of enterprise SSD usage is impressive when compared with counterpart HDD storage, as seen in Figure One. Based on recent industry data available from Gartner, The Register, and Stifel in 2017, the value of enterprise-class SSDs sold exceeded the value of enterprise HDDs sold for the first time. That year, enterprise SSD sales reached about $11.5 billion, surpassing HDDs with sales worth just a little over $11 billion.

- In 2012, enterprise HDD revenues were almost three times that of counterparts SSD ($8.8 billion vs. $3.0 billion respectively).
- Between 2012 and 2017, HDD revenues increased by a CAGR of 4.4% while SSD drives grew during the same interval at a CAGR of over 30%.

FIGURE 1

Revenues from Enterprise HDDs and SSDs Worldwide

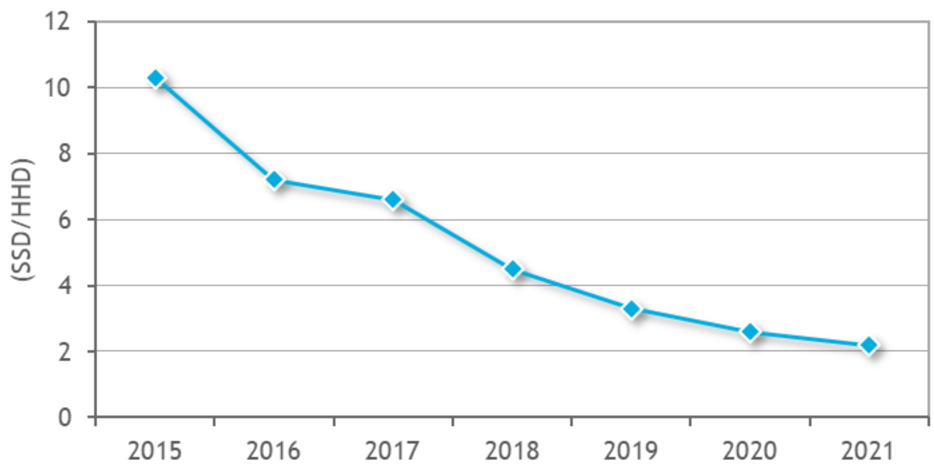


Source: Hyperion Resources, Gartner; The Register; and Stifel 2018

Recent estimates from IDC and Stifle show that SSD/HDD pricing ratios per GB will continue to decline as seen in Figure Two. Based on this projection, by 2021 the cost per GB of SSD will be only twice that of a counterpart HDD.

FIGURE 2

SDD vs HDD Price Per GB Ratio



Source: Hyperion Resources, IDC, Stifel 2018

Implementation of Flash-based SSD in the Data Center

The rapid emergence of SSD in recent years has led to the development of a wide range of memory/storage innovations that are already working their way into data centers both on-prem and in the cloud. These innovations include:

Data Cache for HDD File Systems

In a common scenario, an enterprise or technical computing system stores temporary copies of the most active data in a flash-based SSD and a permanent copy of that same data on a counterpart HDD. The choice of what data is stored in the cache is critical, as the cache will have much smaller storage capacity than the overall HDD system. The cache data-selection algorithm attempts to ensure that the most active data is available in the cache when needed. This selection process is a critical determinant of the overall performance of such a cache system.

The advantages here are myriad: the SSD offers significantly improved access times and data bandwidth capabilities over HDDs, while the HDD can offer a larger overall storage capacity at a lower price per bit stored. In addition, the use of such cache allows designers to optimize the system's interconnect scheme to support the smaller cache system instead of having to develop a high-performance and more costly network to access the entire HDD storage infrastructure.

Hybrid SSD/HDD Arrays

In a hybrid HDD/SSD array, the computer's overall storage system is composed of a mix of HDDs and SSDs. The determination of HDD/SSD mix can vary significantly based on factors that include overall budget, data access and bandwidth requirements, interconnect schemes, power and cooling capabilities, and the predictability and impact of data access patterns of the expected system workload. Here, the tradeoffs of SSD vs HDD price/performance require a thorough understanding of which data needs to reside on each type of storage hardware in order to take advantage of the best features of each.

- An advantage of a hybrid array scheme is that new storage, SSD, HDD, or a combination of both, can be straightforwardly added to the overall storage system based on new or changing information about job mix, budget increase, or even new requirements for performance improvements of key applications.
- Despite these advantages, hybrid arrays require a certain amount of data shepherding, accomplished by control software that continually oversees the placement and migration of user data among SSD and HDD units to optimize overall performance and cost. Data that is misplaced can result in a minimal or even degraded performance impact.

Burst Buffers

Burst buffers are one particular application of flash-NAND based SSD that reside on specialized nodes that bridge the internal interconnect of a given compute system and the overall fabric of the storage system accessed through system I/O nodes.

In essence, a burst buffer is a collection of distributed nodes that contains the on-board processing capability and memory capacity to be tightly integrated with a system's memory (vice storage) system that can manage fast storage across user jobs, provide caching services, support in transit analysis and processing, and enable accelerated application support for small block-size transfers.

All Flash Storage

As its name implies, all flash storage is the exclusive use of flash-based SSD for a system's storage without the use of any HDD. However, despite the many advantages of SSD from a performance perspective, in most cases today, the cost of SSD compared with the price per bit of counterpart HDD has limited the use of all-flash storage systems.

The general trend is that SSD will be an increasingly widespread storage option in many enterprise systems, initially in those with stringent and well-defined performance requirements, but expanding their presence as the cost differences between SSD and HDD narrows.

IV. FLASH STORAGE TRENDS WORTH WATCHING

Within the overall flash-based SSD sector, a number of trends/developments promise additional performance gains for a growing set of computer architectures, data center configurations, and user workloads. Each of these developments, whether working alone or in concert with other SSD technologies, can help to further drive the widespread adoption of flash memory into the overall IT server ecosystem.

3D NAND: Driving Gains in Flash-Based SSD

NAND flash is the largest contributor to improvements in cost, durability, and read/write performance for SSDs, and the advent of 3D NAND technology, which vertically stacks memory cells in multiple layers, has contributed significantly to recent improvements in SSD capacity and capability. 3D NAND flash offers the potential for higher capacity in a smaller physical space than 2D NAND. In comparison with planar NAND, 3D NAND can lower the cost per gigabyte, improve electrical use to reduce power consumption, boost reliability, and provide higher data write performance.

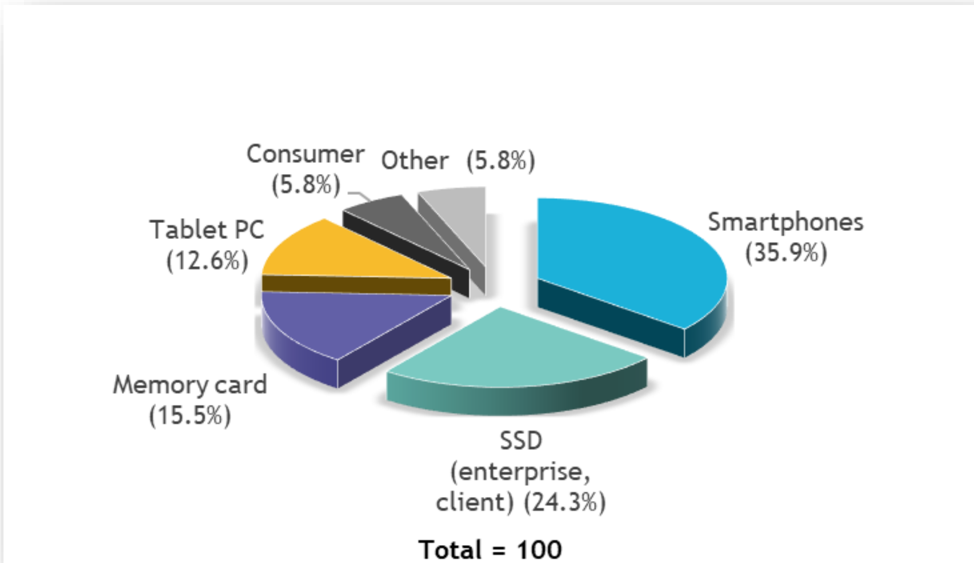
Recent leading-edge enterprise-class 3D NAND-based SSD devices have impressive features that include:

- SSD storage capacities of more than 11 TBytes
- 5.5GB/sec of sequential reads and up to 3.5GB/sec sequential writes.
- A mean time to failure (MTTF) of 3 million device hours
- Power consumption below 6 watts
- The capacity to replace a rack of 24 HDD drives (based on IOPS performance)
- A 2.5" form factor

As can be seen in Figure Three, about 25% of all NAND bits produced worldwide are used in SSDs (both client and enterprise), second only to smart phones but well ahead of memory cards, according to industry data from IC Insights. However, many in the industry expect that the combination of increased price/performance options for flash-based SSDs, new demands for fast memory in enterprise-class storage schemes, and a projected slowdown in the growth of smart phone sales could propel SSD's to the forefront in the next few years. The NAND flash market is supplied by a number of technologically aggressive supplier firms that include Intel, Micron, Samsung, SanDisk and SK Hynix, each with a strong R&D effort that extends their NAND flash development road map well into the next decade. The emergence of SSD as the largest consumer of flash devices ensures that flash makers will concentrate their development efforts on chips best suited to SSD requirements.

FIGURE 3

Distribution of NAND Bits Worldwide in 2018 by Application



Source: Hyperion Resources, IC Insights 2018

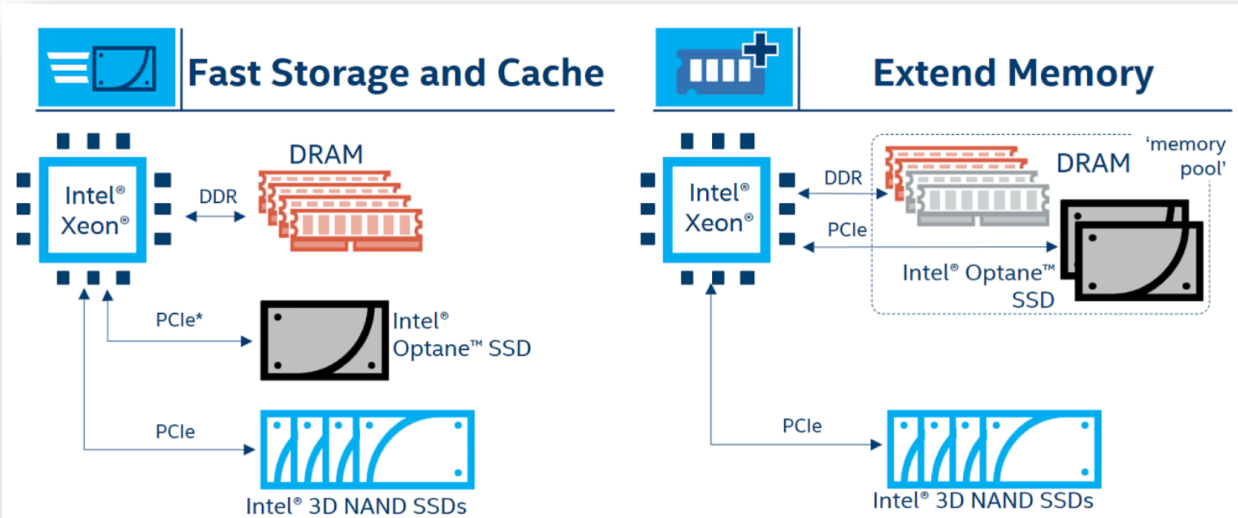
Intel Optane Memory System: Expanding the Universe of Fast Memory

Intel's Optane development is designed to address the performance gap between DRAM memory and disk, but one that takes full advantage of Intel's advanced semiconductor design and production capabilities. along with its ability to communicate directly with its own processors. The Optane memory is targeted not as a replacement for flash-based SSDs, but instead as a DRAM-based memory extender or fast cache device that can be used in concert with more traditional 3D NAND SSDs, as can be seen in Figure Four.

Optane is an effort to combine DRAM's fast access time and long endurance with NAND's density and non-volatility. Introduced in 2015, Optane memory modules are based on a new semiconductor storage design called 3D XPoint that uses a cross point structure of perpendicular wires to access individual storage cells.

FIGURE 4

Intel Optane SSD Use Cases



Source: Hyperion Research, Intel 2018

Advantages of the 3D XPoint chips are many: they are non-volatile like NAND counterparts but have much higher endurance rates and data access times. In addition, each storage cell in a 3D XPoint chip, unlike the case in DRAMs, does not require a switching transistor at each storage location, allowing for an 8X to 10X storage density improvement over DRAMs.

In addition, Intel can take full advantage of offering an integrated processor/memory subsystem by supporting interconnect options directly between its processors and Optane SSDs though either a PCIe bus or through the considerably faster DDR bus. However, because the technology is relatively new and complex to manufacture, the 3D XPoint chips currently have total storage capability much smaller than leading-edge NAND chips, currently Optane offerings are only in the range of 16 and 32 GB, well below the capacity of SSDs that have capacities in the TB range.

NVMe Over Fabric

As mentioned earlier, NVMe is a specification that allowed an SSD to make effective use of a high-speed PCI Express bus in a computer. That specification, albeit useful within a single computer system, does not support SSD access across networks. To address that issue, NVMe over Fabrics (NVMeoF) was created to enable low-latency NVMe SSDs to scale out to the network in the most efficient way possible, simplifying design, reducing overhead, and improving performance.

About 90% of the NVMeoF protocol is the same as local NVMe that includes NVMe namespaces, I/O and administrative commands, register constructs and asynchronous event management. Key differences center on identifier naming conventions and data transfer configurations.

Software Defined Storage

Software-defined storage (SDS) is a relatively new storage paradigm whereby server-based software manages data storage resources and functionality and has no dependencies on the underlying physical storage hardware schemes. Unlike monolithic SAN and NAS systems, software-defined storage products enable users to separately (and more optimally) manage both storage hardware and software.

- Common characteristics of SDS products include the ability to aggregate storage resources, scale out the system across a server cluster, manage the shared storage pool and storage services through a single administrative interface, and set policies to control storage features and functionality.
- A key feature of software defined storage is that it can offer an integrated data management capacity that is centrally managed across an overall file system. This capability can reduce the need for data caches, such as the burst buffers that are common to legacy file system such as Lustre or GPFS, while offering superior performance on small file, metadata, or random I/O operations.

There are a number of cases where SDS suppliers are looking to integrate SSDs into the overall memory/storage hierarchy as a way to combine the superior access times and data bandwidth of SSDs with the flexibility and economic advantages of cloud-based HDDs. One example is provided by WekaIO, which offers a flash-native parallel file system that is a software-only solution targeted to optimize flash-performance on NVMe, SAS, or SATA SSD, as well as manages data migrations to cloud-based HDD storage systems running either on-prem or in a public cloud.

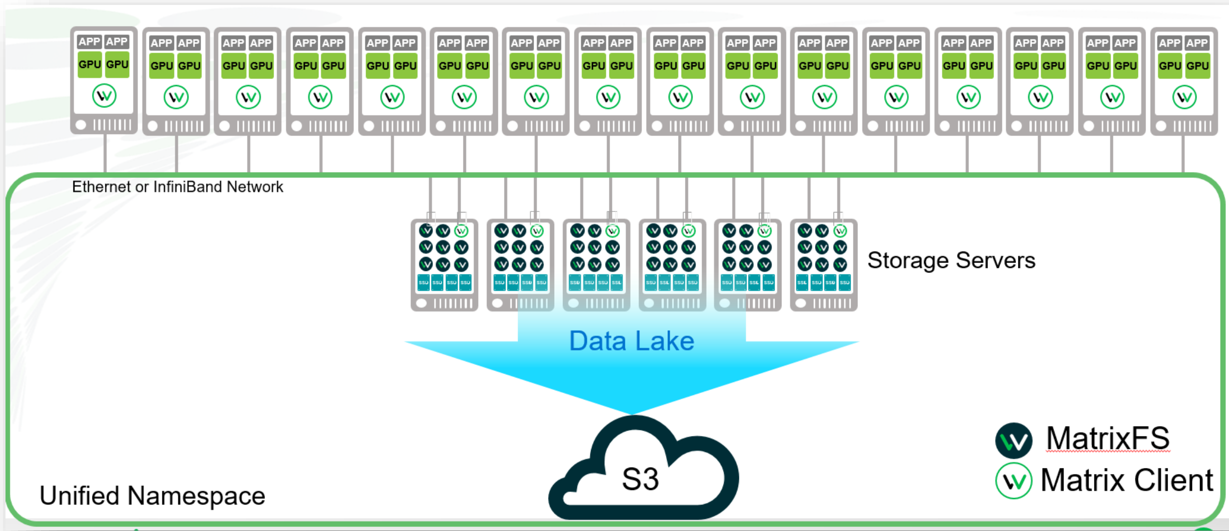
WekaIO's Matrix 3.0 software-based scale-out storage solution is optimized to leverage the performance of flash technology to support both large and small file access, either randomly or sequentially. It can be deployed natively in the cloud with compute and storage integrated into the application cluster or as a dedicated storage server (appliance model), eliminating data center cost and complexity.

The WekaIO Matrix v3.0 software that manages data movement and facilitates high-speed data looks to offer significant performance gains across a wide range of use cases that include big data analytics, life sciences, financial services, and critical national-level applications such as climate change simulation, earthquake modeling, and space research.

One particularly compelling use case for WekaIO SDS is in the deep learning area where the issues of data transfer of training data from storage to a computer's complement of GPUs is a critical performance bottleneck. In this case, WekaIO software is deployed on a dedicated cluster of servers interconnected by either an Ethernet or InfiniBand low latency network. Server cluster resources - CPU cores, memory, network connections, and SSDs - are allocated to the WekaIO software. In this set-up, the WekaIO SDS can deliver a full 5Gbytes/s to each GPU while leveraging S3 compatible HDD based object storage for cost effective storage at massive scale. Figure Five provides an example architecture for a WekaIO storage solution for a deep learning GPU-based solution.

FIGURE 5

WekaIO SSD Storage Solution for GPU Clusters



Source: Hyperion Research, WekaIO 2018

FUTURE OUTLOOK

Continued advances in 3D NAND chips - and subsequent NAND chip architectures – coupled with the flash-based solid-state disks (SSDs) devices that use them, bode well for the overall SSD sector, and particularly in the enterprise-class storage. These flash-based SSDs offer significant advantages over hard disk drives (HDDs) from a speed, power, and increasingly a total cost of ownership perspective.

Flash-based SSDs will continue to evolve, and they likely will be an increasingly critical element in driving performance gains in the memory/storage hierarchy of systems that have a wide range of use cases and that can accelerate performance when operating either on-prem, in the cloud, or as part of a hybrid system architecture.

The availability of flash-based SSDs has enabled a wide range of innovative solutions that include Intel's' Optane memory accelerator/expander that seeks to offer DRAM-like performance with the power and non-volatility of flash, software defined storage such as that from WekaIO that seamlessly integrates SSD and HDD to optimize both performance and cost; and NVMe over Fabric that seeks to expand the domain and reach of SSD devices across wide networks. These, and other development in the works will almost certainly contribute to the growing adoption of SSD and ensure that computer designers and users have the best available options when considering an efficient memory/storage hierarchy.

About Hyperion Research, LLC

Hyperion Research provides data-driven research, analysis and recommendations for technologies, applications, and markets in high performance computing and emerging technology areas to help organizations worldwide make effective decisions and seize growth opportunities. Research includes market sizing and forecasting, share tracking, segmentation, technology and related trend analysis, and both user and vendor analysis for multi-user technical server technology used for HPC and HPDA (high performance data analysis). We provide thought leadership and practical guidance for users, vendors and other members of the HPC community by focusing on key market and technology trends across government, industry, commerce, and academia.

Headquarters

365 Summit Avenue
St. Paul, MN 55102
USA
612.812.5798
www.hpcuserforum.com and www.HyperionResearch.com
