Special Study

General-Purpose NVMe-Based All-Flash Arrays to Dominate Revenue as the Enterprise NVMe Market Develops

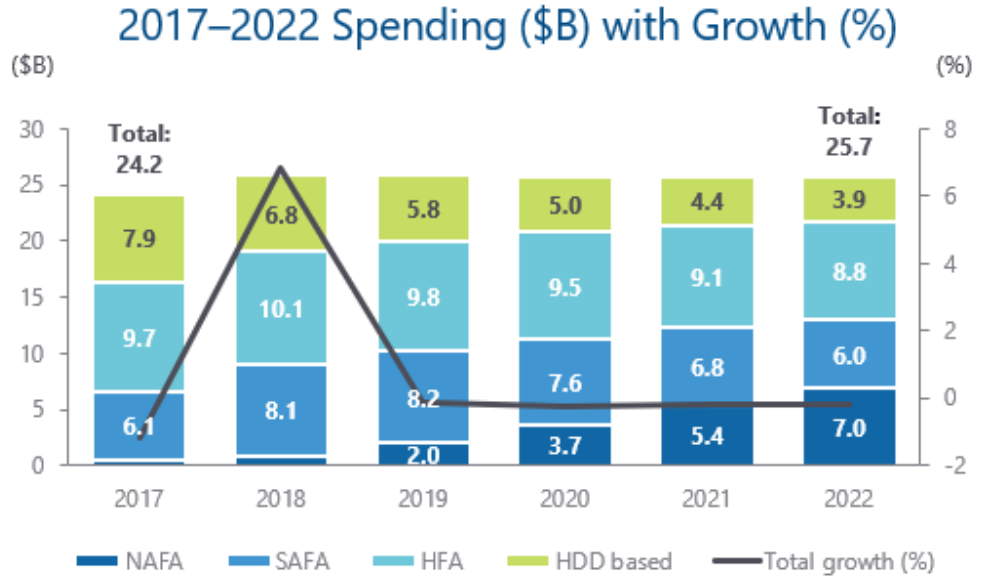
Eric Burgener Jeff Janukowicz Natalya Yezhkova

THIS IDC SPECIAL STUDY EXCERPT WAS CREATED FOR NETAPP

IDC OPINION

FIGURE 1

Worldwide External Enterprise Storage Systems Spending Snapshot



Selected Segment Growth Rate	Total Market CAGR
▲ NAFA CAGR 67.2% ▼ SAFA CAGR -0.4% ▼ HFA CAGR -1.8% ▼ HDD-based CAGR -13.2%	1.2%

Note: Chart legend should be read from left to right.

Source: IDC, 2018

September 2018, IDC #US44265318e

IN THIS STUDY

Information technology (IT) managers in both enterprise and cloud environments are looking to nonvolatile memory express (NVMe) to drive needed storage performance as they add new types of real-time applications and grow their businesses. Significant experimentation with local peripheral component interconnect express (PCIe) and NVMe-based solid state disks (SSDs) over the past two years has driven organizations that need this type of performance at scale toward what IDC is calling NVMe-based all-flash arrays (NAFAs). These solutions can provide the same type of storage latencies as local SSDs but offer the advantages of enterprise arrays: easy scalability to much higher capacities, the ability to share access to high-performance storage across their entire IT infrastructure, and access to enterprise-class data services not available in most local storage configurations (RAID, snapshots, replication, etc.). The first NAFAs shipped in 2016, but the market is expected to grow rapidly, cannibalizing significant revenue from the current all-flash array (AFA) market. IDC expects that, by 2021, NAFA shipments will be dominating primary external AFA revenue. Users largely agree — 78.1% of surveyed users of NVMe technology today think that SCSI will transition to NVMe for primary storage, and 74.8% of those users believe that NVMe will be driving more than 50% of primary external AFA revenue by 2021.

For enterprise storage vendors, NVMe and the other storage technologies it enables like storage-class memory (SCM) represent the next shift in the external storage market. Vendors should already have articulated an NVMe strategy for their customers and may already be somewhat behind the market if they are not already shipping at least NVMe-compatible systems (those that can be nondisruptively upgraded to NVMe). High-performance storage solutions that support NVMe over Fabric (NVMe-oF) host connections are already available from a number of start-ups as well as more established storage vendors like NetApp. End users with performance-sensitive primary storage workloads should be thinking about whether or not a NAFA is right for them on their next enterprise storage platform technology refresh. Demand for NAFAs will be driven not only by very high-performance databases and real-time big data analytics workloads but also by the efficiency advantages of NVMe over SCSI technology for dense mixed workload consolidation. NVMe technology does demand somewhat of a price premium on most enterprise storage platforms, but 81.3% of those customers already using NVMe think that it is worth the price premium. IDC expects NAFAs to be a $2.0 billion market by 2019, at which point it will be generating almost 24% of overall AFA market revenue.

ADVICE FOR TECHNOLOGY SUPPLIERS

- Enterprise storage vendors that expect to remain competitive should already have an NVMe strategy in place and should be shipping at least NVMe-compatible (if not NVMe featuring) storage platforms. Vendors should also already have an SCM strategy in place, although it is too early to be shipping shared storage solutions that leverage SCM at this point. The first shared enterprise storage vendor solutions that feature SCM technology are expected in late 2018 or early 2019.
- Vendors will need to decide which NAFA market they will pursue ("performance oriented" or "data services oriented"), or if they will pursue both, and provide clear and compelling reasons for customers to consider their solutions in each of these markets. Start-up vendors will be challenged to succeed in this market, given that the established enterprise storage market leaders (by revenue) are rapidly entering (and in some cases already pursuing) these markets.

- Technology refresh cycles, with customers upgrading from older SCSI-based all-flash arrays (SAFAs), will drive the major portion of NAFA market revenue over the five-year forecast period. Both market segments ("performance oriented" and "data services oriented"), however, represent sizable and interesting opportunities for vendors.
- As NVMe matures, there will be fewer reasons for vendors to develop custom hardware and software in solutions targeted at the NAFA market. Vendors that do decide to continue to leverage proprietary designs and technologies should revisit that on every technology refresh and make sure they clearly understand how the use of custom content delivers meaningful differentiation to target customers.

MARKET FORECAST

Table 1 and Figure 2 provide a top-level snapshot of the worldwide NAFA market forecast. All established storage providers are expected to be shipping NAFAs with support for NVMe-oF by 2020 (some of these are already shipping NVMe-oF solutions today). Additional start-up players will focus their sales efforts more on the performance-oriented segment – a segment that will drive smaller revenue over the forecast period than the "data services oriented" segment – although several of these players will add data services and attempt to move into the larger segment as well. Component revenue will be generated by other vendors around NVMe technology, like Attala Systems, Broadcom, Cavium, Intel, Micron, Samsung, Seagate, SK Hynix, and Toshiba, but those revenues are not included as part of the NAFA forecast. This forecast only covers NAFA systems that ship with an enterprise-class storage operating system and are supported as standalone storage platforms by vendors to align with IDC's storage system forecast.

TABLE 1

Worldwide NAFA Revenue, 2017-2022

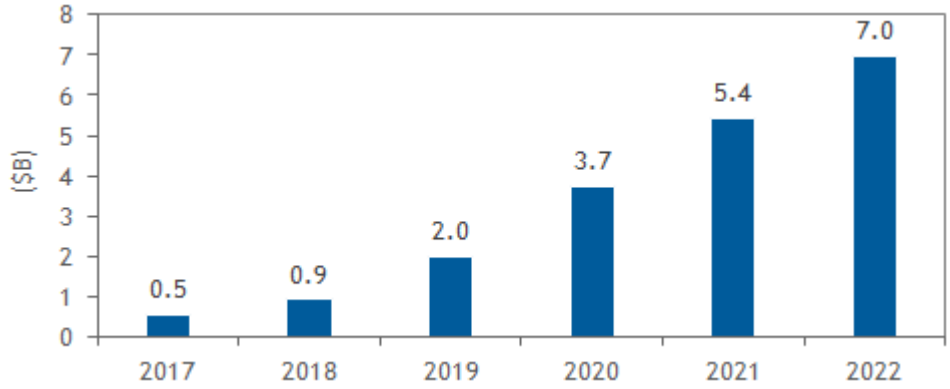
	2017	2018	2019	2020	2021	2022	2017–2022 CAGR (%)
Revenue ($B)	0.5	0.9	2.0	3.7	5.4	7.0	67.2

Note: All revenue forecasts in this document are based off of IDC's June Enterprise Storage Tracker release.

Source: IDC, 2018

FIGURE 2

Worldwide NAFA Revenue, 2017–2022

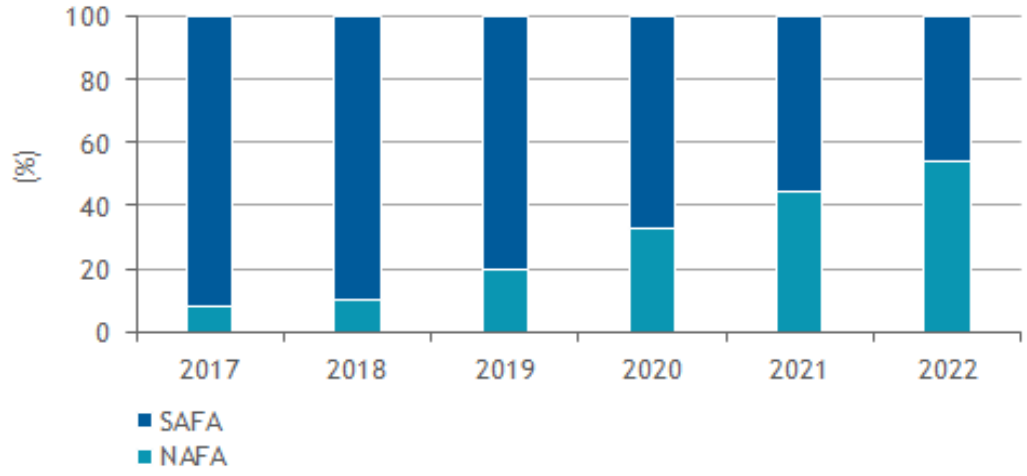


Source: IDC, 2018

As NVMe matures as a standard, prices will drop, product volumes and multivendor availability will increase, and the use of NVMe technology in production environments will become more mainstream. Customers that IDC surveyed in June 2018 (see *Customers Highly Satisfied with the Use of Enterprise NVMe Technology in Production Environments,* IDC #US44157618, August 2018) think the transition from SCSI to NVMe as the core technology in enterprise storage platforms will happen rapidly. Furthermore, 78.1% agree that SCSI will transition to NVMe for primary storage, and 74.8% think it will happen by 2021. IDC is more bullish on this market growth than end users and believes that by 2021 NAFAs will drive over 60% of primary external array revenue (assuming primary storage revenue are 35% of all storage revenue). The NAFA market will grow from 8.0% of the overall AFA market in 2017 to 53.7% of it by 2022 (see Figure 3 and Table 2).

FIGURE 3

Worldwide SAFA and NAFA Revenue Share, 2017–2022



Source: IDC, 2018

TABLE 2

Worldwide External Enterprise Storage Systems Revenue by Segment, 2017–2022 ($B)

	2017	2018	2019	2020	2021	2022	2017–2022 CAGR (%)
NAFA	0.5	0.9	2.0	3.7	5.4	7.0	67.2
SAFA	6.1	8.1	8.2	7.6	6.8	6.0	-0.4
HFA	9.7	10.1	9.8	9.5	9.1	8.8	-1.8
HDD only	7.9	6.8	5.8	5.0	4.4	3.9	-13.2
Total	24.2	25.9	25.9	25.8	25.8	25.7	1.2

Source: IDC, 2018

MARKET CONTEXT

It's been a short six years since the AFA market started to take off. Evolving datacenter workloads were creating new storage performance requirements that were becoming very difficult to address with hard disk drives (HDDs), and the use of flash as a persistent storage technology resolved these challenges. In 2018, IDC expects the AFA market to generate $9.0 billion in revenue, representing roughly 35% of the overall external enterprise storage systems market but well over 70% of the primary storage spend in this space. The lion's share of AFAs purchased over the past six years has been based around core SCSI technologies.

Application performance benefited significantly from flash, but there were other positive impacts from the use of flash as persistent storage that have become extremely important in the past several years. Low flash latencies enabled the use of inline data reduction (compression and deduplication) with latency-sensitive primary workloads, and the resulting data reduction ratios (which for most mixed enterprise workloads averaged between 2:1 and 6:1) helped to significantly lower the effective cost per gigabyte of flash storage, making it that much more attractive relative to HDD-only systems. The fact that data reduction could be used in line without undue performance impacts helped AFAs penetrate the primary storage markets that much more rapidly. This additional performance also meant that other enterprise-class data services, such as additional storage efficiency technologies (thin provisioning, write minimization, pattern recognition, space-efficient snapshots, and delta differential-based replication), various data protection schemes (RAID and erasure coding), and encryption, could now be used in line with latency-sensitive primary storage workloads as well, making for a compelling total-cost-of-ownership story around AFAs.

Most AFA purchases were driven by what IDC refers to as 3rd Platform computing workloads. In the 3rd Platform computing era, CIOs are tasked with providing support for new workloads like mobile

computing, social media, big data analytics, and cloud while continuing to support legacy workloads like relational databases, collaboration platforms, and enterprise applications. Virtual computing had datacenters consolidating these workloads onto virtual infrastructure, and the resulting I/O profiles could not be cost effectively handled by HDDs, driving strong demand for flash as a persistent storage medium. Just within the past several years, storage performance requirements have again been undergoing significant change, driven by customer expectations about real-time response and the increasing use of big data analytics to help provide competitive differentiation and drive better business decisions. The use of artificial intelligence/machine learning (AI/ML) is becoming popular across many different industries, and the desire to leverage these technologies against rapidly growing multi-petabyte data sets has driven demand for more performance than the SCSI I/O interface could deliver, even when coupled with persistent flash storage.

As an I/O interface, SCSI was specifically designed for use with HDDs. Flash is accessed very differently than spinning disk media, and while the SCSI interface could be used with it, it did not result in a very efficient use of flash performance or capacity. A new interface specification called NVMe that was specifically designed and optimized for use with nonvolatile memory, such as NAND flash media, was initially introduced in 2011, and NVMe solid state drives that could plug directly into the PCIe bus in x86 servers were shipping by 2014 (Samsung introduced the first one). As a protocol, NVMe was optimized for efficient nonvolatile memory access and can be deployed within a server used in shared storage solutions or in switched fabrics. While SCSI-based flash access latencies in a SAN could be well below a millisecond, NVMe access latencies across an NVMe over fabric connection could be below 100 microseconds — up to a 10x reduction in latency. NVMe also supported one to two orders of magnitude higher throughput and bandwidth than SCSI on a per-device basis, as well as two orders of magnitude better parallelism (NVMe supports a queue depth of 64,000).

With other persistent memory technologies like SCM on the horizon, the move to NVMe provides a future-proof growth path for customers and will also be required if customers want to support SCM products like Intel Optane. SCM should start to become more generally available in 2019 and will be in demand among customers that feel they are already pushing the performance envelope with real-time big data analytics and other workloads with extremely low latency or high throughput and bandwidth requirements. This makes it very important for enterprise storage vendors that do not want to be perceived as laggards to not only provide a well-articulated growth path to NVMe and SCM but also make those upgrades nondisruptive.

History of NAFA Market Development

The NAFA market grew directly out of both enterprise customers and cloud providers traversing a similar path from web-scale infrastructure leveraging internal storage to a more disaggregated model. As workload performance requirements outran the ability of SAFAs, many storage managers looked to internal storage options, such as SSDs, to address their challenges. As local devices internal to x86 servers, they were directly attached either to SCSI or PCIe busses, incurring no network latencies for storage access. With local NVMe SSDs, many customers were enjoying sub-100 microsecond latencies. Storage managers liked the latency they got from these approaches, and end users absolutely noticed the performance advantages. Most of these workloads were growing at double-digit rates, however, and end users regularly petitioned their IT organizations for "larger SSDs" (i.e., increased capacity).

As IT managers considered scaling these environments to meet growing data demands, concerns arose around capacity scalability, capacity utilization (and the inability to efficiently share this high-cost

and high-performance storage), and the lack of features typically desired in enterprise storage (storage efficiency technologies, snapshots, replication, etc.). It became clear that what was needed was an enterprise-class shared storage solution that delivered the same performance as the local NVMe storage but addressed these concerns. The fact that such a solution should be disaggregated, allowing customers to scale compute and storage resources independently, is also considered important by many of these customers (that include both enterprises and cloud providers).

NVMe is looking to displace the use of traditional protocols, such as Fibre Channel (FC) and internet SCSI (iSCSI), and promises all the performance of PCIe, which was primarily only in use in local storage, but with support for enterprise features that promote better scalability, more efficient capacity utilization, and improved availability in shared storage configurations. NVMe over Fabric (NVMe-oF) offers the ability to run the Direct Memory Access (DMA) capability of NVMe over a switched fabric, providing what is referred to as Remote Direct Memory Access (RDMA). With the high performance promise of RDMA, NVMe-oF is and will continue to be a critical enabler for NAFA systems of both types (performance- and data services-oriented). Today, NVMe-oF can run over three types of network transports – FC, Ethernet, and InfiniBand. Practitioners tend to want to deploy using a transport with which they are already familiar and is already running in their environment. IDC has noted that installed FC customers want FC, while installed Ethernet customers and many greenfield deployments prefer Ethernet. FC customers will generally want to consider NVMe-oF deployments using Gen 6 (16/32Gb) FC, while Ethernet customers will generally want to use at least 40GbE and will be able to choose from three different approaches (RoCE, iWARP, or TCP). InfiniBand as a transport is mostly used in high-performance computing environments, although there is a bit of it in enterprises. As a standard, NVMe-oF is less mature than NVMe, and customers using NVMe in their environments today look forward to the standard maturing as a way to lower NVMe deployment costs, increase the volume availability of NVMe components, and move away from the need for the custom content necessary to create an end-to-end enterprise-class NVMe solution today.

NAFA Market Forecast Segmentation

As the NAFA market grows, it will cannibalize the SAFA market. Throughout the forecast period, IDC expects that all of the revenue through 2020 will be generated from products sold as primary storage platforms. We assume that there will be a very small percentage of NAFAs sold in 2021 and 2022 that use very high-capacity SSDs and are targeted at some more performance-sensitive secondary storage workloads like backup and disaster recovery (to help reduce recovery point and recovery time objectives). Those few systems targeted at secondary storage environments will be of the data services-oriented type. In the early years of market development, there will be two market segments that IDC is referring to as the "performance oriented" and the "data services oriented" segments, although this segmentation will blur in the later years of the forecast. Established vendors introducing NVMe on their flagship platforms already bring with them a full complement of data services, making the systems eligible to not only host very performance-sensitive workloads but also handle dense mixed workload consolidation more efficiently than SAFAs. Start-ups have focused more on the performance-oriented segment to date, but many of these vendors will add data services to their systems over time. Early indications are that there is a slight price premium (roughly 15% on a cost per-gigabyte basis) for the lower latency performance-oriented systems despite their lack of software functionality (relative to the data services-oriented systems).

It is still early, but discussions with customers that have tested NAFAs of both types indicate that there is a performance difference between the two of somewhere between 10% and 20% (in terms of latency and throughput) on a per-controller basis. For some workloads, this performance differential is enough

to justify a purchase of the higher-performance option, but for many workloads, that difference will not be meaningful. For a vendor to target the "performance oriented" segment over time and be successful, the vendor will need to focus on workloads where the performance differential (whatever it may be for its platform) is meaningful to customers or pursue some other offering features (e.g., vertical market expertise) that will allow them to set themselves apart from "data services oriented" platforms. Systems that leverage SCM, either as a cache or as a persistent storage, will be able to deliver performance that is truly differentiating from data services–oriented NAFAs (at least until those platforms start to leverage SCM technologies as well). One of the factors contributing to lower performance with data services-oriented NAFAs is the additional latency incurred by running inline data services such as data reduction and encryption. Systems that allow data services to be enabled on a per volume (or file system) basis give customers more flexibility in meeting different performance requirements.

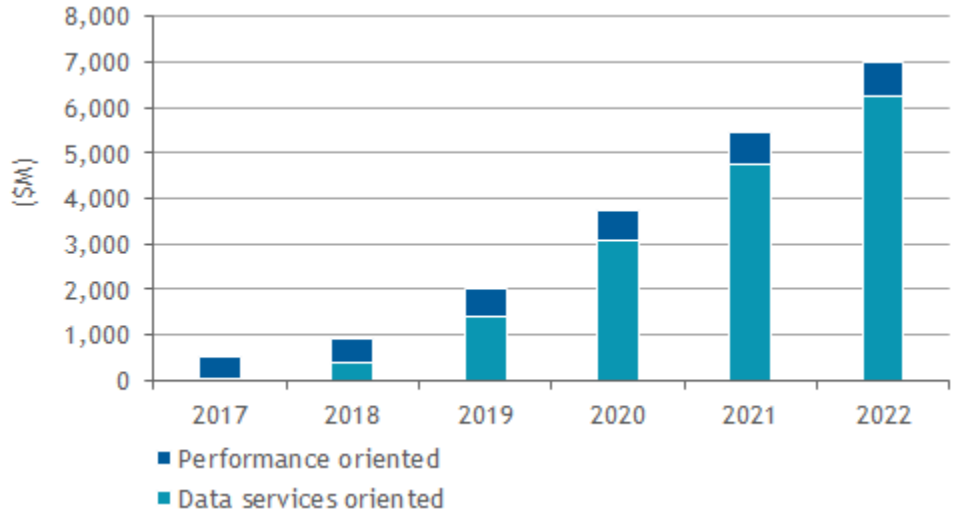
Note that the 10–20% performance differential that IDC has seen to date assumes an apples-to-apples comparison in terms of the number of controllers across different platforms. Some performance-oriented systems support as many as 20 controllers within a single system image, going beyond what data services–oriented platforms offer and thereby creating a throughput differential that can be much greater than the 20% cited previously. Note also that while performance scalability is generally not linear (it can be workload dependent), there will be some variability in the extent of linear scalability based on software designs between platforms. Latency will also be a differentiator, since running data services in line will definitely impact it. The ability of performance-oriented NAFAs to provide more consistent and predictable latencies, particularly at scale, will be a reason to buy them, but note that even data services–oriented NAFAs, when coupled with NVMe-oF, will be able to deliver latencies much lower and more consistent than those available from SAFAs. The key for these vendors, again, will be to find customers and workloads where this difference matters.

Performance density (in terms of how many IOPS can be driven in a certain number of Us) is an issue for some customers, and customers have commented on how much more compact their newer NAFA solutions are relative to the systems they may have been running very high-performance workloads on before. This can be an issue of importance to customers buying in either segment (performance or data services oriented).

The key assumption that underlies how the performance- and data services–oriented submarkets develop is that there will not be much top-line revenue growth in the performance-oriented segment over the five-year forecast period (see Figure 4 and Table 3). The anticipated performance differential between the two types of platforms (performance and data services oriented) is not expected to drive that much differentiation between the two types of platforms. So while there are some workloads that will absolutely want the highest-performance option available, most workloads will be more than satisfied with the slightly lower performance of the data services–oriented platforms. At the same time, however, many of the performance-oriented start-ups will be growing their revenue as well, replacing some of this revenue loss to the overall segment. During the forecast period, some of the performance-oriented vendors will begin to straddle these two markets as they add data services to their storage operating systems, beginning to compete somewhat more in the more general-purpose market for mixed enterprise workload consolidation (but also lowering their average latencies with the addition of inline data services). The end result is what appears to be a relatively low growth rate over the five-year forecast period, but there is actually a shift of revenue between platform types driving that.

FIGURE 4

Worldwide NAFA Revenue by Segment, 2017–2022



Source: IDC, 2018

TABLE 3

Worldwide NAFA Revenue by Segment, 2017–2022 ($M)

	2017	2018	2019	2020	2021	2022	2017–2022 CAGR (%)
Performance oriented	488.0	536.8	587.8	634.8	685.6	740.5	8.7
Data services oriented	46.0	387.2	1,420.2	3,082.2	4,745.4	6,243.5	167.0
Total	534.0	924.0	2,008.0	3,717.0	5,431.0	6,984.0	67.2

Source: IDC, 2018

Note that vendors have already started to incorporate NAFAs into converged infrastructure (CI) offerings, particularly those that include NVIDIA GPUs and are specifically targeted at AI/ML workloads. Both Pure Storage and NetApp have announced products in this space, and IDC expects other established storage vendors to follow suit by the end of the year. The NAFA forecast revenue does include revenue generated from CI offerings that include NAFAs.

Drivers and Inhibitors

Drivers

Tier 0 Workloads

- **Assumption:** In a number of industries, there are workloads where increased performance translates to improved business results.

- As SAFAs first started to penetrate enterprise storage, and then later with NVMe technology, IDC noted that these are the workloads where extreme performance-oriented storage vendors target first. This will be no different with NAFAs. IDC had an opportunity to speak with many of the early adopters in this category (see *NVMe-Based All-Flash Arrays Achieving an Early Beachhead with Business-Critical Workloads,* IDC #US44151118, July 2018) and their deployment patterns confirm this observation. With the increasingly real-time expectations of customers across all industries, the percentage of workloads that fit the definition of "tier 0" is creeping up. Dropping NVMe costs, relative to alternatives, can encourage customers to lump additional workloads into their defined "premium performance" tier that resides (or will reside) on NAFAs.
- **Impact:** Over time, because of digital transformation and dropping NVMe costs, a higher percentage of enterprise workloads will be deemed to need NVMe technology deployment, driving broader NVMe penetration of external enterprise storage.

Increased Penetration of Real-Time Big Data Analytics

- **Assumption:** In the era of digital transformation, data has become a strategic asset and more IT organizations are depending on fast and easy access to data (often processed through AI-/ML-driven workloads) to enable better, more informed business decisions. Given the scale of data used in these real-time processes, the benefits of NVMe will be in increasing demand. Real-time big data analytics workloads will increasingly be deemed mission critical by IT organizations, driving higher capacity and availability requirements that are better met by NAFAs (as opposed to local storage solutions).
- **Impact:** As a natural business evolution, more IT organizations will outrun the performance capabilities of SAFAs and be dissatisfied with the limitations of local storage configurations, driving increasing NAFA penetration. This will be a key driver of the rapid transition from SCSI to NVMe for general-purpose primary storage.

Technology Refresh and Mixed Workload Consolidation

- **Assumption:** With higher-performance storage technologies like storage-class memory that require NVMe in the near future, the ease of integrating these technologies into production environments will impact enterprise storage platform purchase decisions. NAFAs clearly enable this better than most SAFAs. By 2020, buying a SAFA on technology refresh for primary storage will be less likely than buying a NAFA, but until then, IT organizations will need to carefully consider the pros and cons of moving to a NAFA as they retire legacy primary storage arrays. The fact that the transition from SCSI to NVMe will be much easier and simpler than the prior transition from HDD to flash will help drive strong migration to NVMe on technology refresh. Purchase decision makers will be weighing the higher cost of NAFA acquisition against the better performance, increased efficiencies, and better ability to integrate future storage technologies that those platforms offer relative to SAFAs.
- **Impact:** The lack of a NAFA option will put those vendors that do not have one at a competitive disadvantage relative to those vendors that can let their customers choose between either SAFAs or NAFAs. This will be very important for the larger "data services oriented" market but much less so for the "performance oriented" segment. This will represent a net add to the revenue of those vendors with both options and a net decrease for those vendors that offer only one option. Inhibitors

Higher Price Premium of NVMe Solutions

- **Assumption:** Even though NAFAs offer increased performance and efficiencies relative to SAFAs, some workloads just do not require their performance, and the improved efficiencies are not enough to drive NAFA acquisition. While 81.3% of enterprises that have already

deployed NVMe technology think it is worth the price (see *Customers Highly Satisfied with the Use of Enterprise NVMe Technology in Production Environments,* IDC #US44157618, August 2018), clearly most IT organizations have not yet made the leap to NVMe. In that same survey, 89.8% of current NVMe users unsurprisingly stated that lowered costs would drive increased NVMe penetration across more workload types.

- **Impact:** The cost premium for NVMe will decrease over the next five years, but it will not reach cost parity with SCSI. Still, as the gap between the two narrows, IT organizations will feel cost justified in deploying NVMe for a broader set of workloads.

Delayed Maturation of NVMe and NVMe-oF Standards

- **Assumption:** As technologies approach commodity status, products based on them decrease in cost, are available from multiple sources, and tend to become more reliable. Many current users of NVMe technology are anxious for the technology standard to mature, not only to offer the previously stated benefits but also to add additional support for enterprise-class features (dual-ported NVMe SSDs and access to better SSD status information through APIs were two features specifically called out by customers). As a standard, NVMe is more mature than NVMe-oF, but NVMe-oF is required for the full-performance advantages of NVMe to impact applications. Customers with strong "tier 0" storage performance requirements are deploying NVMe-oF technology, even with custom HBA and driver content, but look forward to the opportunity to deploy NVMe-oF without recourse to custom products. This will result not only in lower cost, wider sharing of the high-performance NVMe storage and streamlined management (not having to deal with upgrades of nonstandard products) but also in more reliable solutions.
- **Impact:** Faster standards maturation will promote faster NVMe technology penetration, while delayed maturation will slow deployments down.

Significant Market Developments

Availability of Very High-Capacity Storage Devices

In the HDD-based era, performance-sensitive customers deployed smaller device sizes and used other tuning techniques (like short stroking) to coax more performance out of spinning disks. The use of flash as a storage medium opened up much broader use of higher-capacity devices for latency-sensitive workloads, since access times (for reads) were impacted less by the physical location of the data within an SSD and more by free space management (i.e., garbage collection) algorithms that were running at any given point in time. Just within the past 12–18 months, commodity SSDs have surpassed the capacity of the largest HDDs, and with the introduction of newer flash media technologies (e.g., 3D triple-level cell [TLC] and quad-level cell [QLC]), commodity SSDs will only widen that gap going forward. With the significantly higher parallelism enabled by NVMe, these larger capacity devices will be viable for use in primary storage workloads (although possibly not the most read latency-sensitive environments). Large capacity HDDs, limited to using the SCSI interface, have typically been limited to secondary storage workloads as they do not perform well on random reads. The largest enterprise HDDs available today are 14TB, while the largest enterprise-class SSDs are 30TB (Samsung's PM1643 boasts an impressive 30.72TB of raw capacity and has been shipping since 2Q18). For any given storage medium, larger capacity devices will offer a lower cost per gigabyte than smaller capacity devices, all other things being equal, although the acquisition cost on a per-device basis can be much higher.

The wide availability in volume of 32TB and 64TB commodity SSDs will provide options to lower the cost of AFAs yet again, although there will be challenges that systems vendors will need to overcome

to use them (e.g., the lower endurance of QLC technology, RAID rebuild times, and other data protection concerns). While these devices will help drive the cost of NAFAs targeted at primary storage workloads above a certain level of scale down, they will also open up opportunities to sell NAFAs into certain secondary storage environments. As a result, IDC's NAFA revenue forecast for 2021 and 2022 assume small revenue generated from secondary storage environments.

Movement Away from Custom Designs

Storage vendors often justify custom development based on their claim that they can deliver higher performance, increased density, better endurance, and/or lower cost per gigabyte from these designs relative to what is available in commodity hardware. Over the past five years, there were a number of vendors shipping AFAs that included custom hardware: HPE uses a custom ASIC in its 3PAR StoreServ system architecture, and Hitachi Vantara, IBM, Pure Storage, Skyera (acquired by HGST in 2014), and Violin Memory have used custom PCI-based hardware (i.e., custom flash modules instead of SSDs) to deliver NVMe performance well ahead of the availability of commodity NVMe hardware. Nimbus Data and Pure Storage have used custom packaging, along with commodity SSDs, to increase the storage capacity that a single drive slot could support (Nimbus Data offers its 100TB ExaDrive product that fits into a 3.5in. drive slot).

The maturation of the NVMe standard has had an impact on these approaches. Skyera disappeared into what eventually became Western Digital, which now sells the IntelliFlash N-Series NAFA (a product that does not use any Skyera IP). Hitachi Vantara and IBM recently announced new storage systems that can use either their custom flash modules or commodity SSDs, signaling that these vendors feel they are able to meet high-performance customer requirements with off-the-shelf technology and will be moving away from custom designs. Violin Memory has ceased operations, and Nimbus Data remains a very small niche player in the AFA market. Only Pure Storage seems to be moving in the direction of more custom content — first with the introduction of its FlashBlade product, based around custom-designed PCIe hardware, and then with the introduction of its DirectFlash Modules (which it uses instead of the commodity SSDs that it used to use in its FlashArray product). FlashBlade has always used the DirectFlash Modules, but now the FlashArray uses them too. When standard NVMe technology — available from multiple vendors in high volumes that drive strong price competition and based on well-accepted standards — delivers comparable performance, density, and endurance at lower cost, it is difficult to justify continued custom hardware development.

NVMe-oF, the last piece of the end-to-end NVMe puzzle, is still an evolving standard. While NVMe drivers started to ship with the first major operating system distributions earlier this year, early adopters of NVMe-oF using FC and Ethernet (RoCE and iWARP) transports confirm that it is still a bit difficult to get RDMA to work over a switched fabric. First-mover NAFA vendors like E8 Storage and Excelero still ship custom HBAs and drivers, which need to be installed on x86 servers to be used with their NAFAs, and Apeiron Data Systems entirely avoided the question of NVMe-oF maturity by implementing its own proprietary NVMe over Ethernet protocol (which does not use RDMA but still requires vendor-specific HBAs and host drivers). There is very strong customer interest in NVMe-oF that uses TCP as a transport, primarily because this requires no custom content of any kind, leverages proven flow control and error recovery routines, and promises to connect to any commodity x86 server through its ability to use widely available, off-the-shelf Ethernet HBAs and host drivers. This interest is there despite the fact that this approach does not support RDMA and will likely deliver less performance than RDMA-based NVMe-oF (but much more performance than standard iSCSI). As the NVMe-oF standard solidifies, vendors should move rapidly to it.

Innovation Accelerators Driving the Increase of Workloads that Require NVMe

As part of 3rd Platform computing, IDC has identified six innovation accelerators that will drive significant next-generation application (NGA) development: Internet of Things (IoT), cognitive systems (AI/ML), next-generation security, augmented and virtual reality, robotics, and 3D printing. While not directly related to storage, these innovation accelerators either produce and/or use massive amounts of data. Increasingly, real-time expectations on the part of customers using these accelerators to help fuel business growth will help drive faster NVMe penetration. IDC has predicted that, by 2020, 60–70% of Fortune 2000 companies will also have at least one real-time big data analytics workload that is considered mission critical. The need for higher availability will impact application design (those workloads that will be run on web-scale infrastructure will have to take on more of the responsibility for delivering high availability) and storage architecture deployment choices (the entire IT infrastructure must support nondisruptive operations over its entire useful life, which shared storage can offer without requiring specially developed applications). NVMe performance will be needed to meet service-level agreements (SLAs), enable rapid data mobility, and speed recovery from storage device failures.

Changes from Prior Forecast

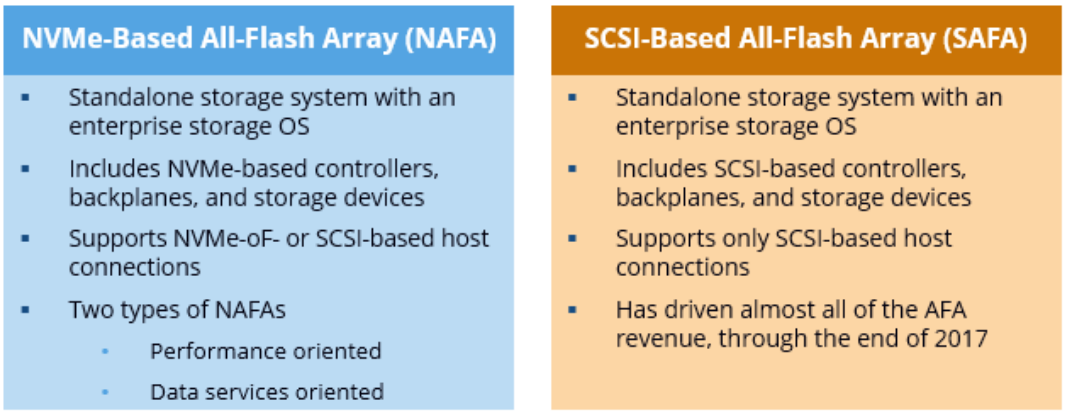
This is the first NAFA forecast IDC has produced.

MARKET DEFINITION

The NAFA market initially arose as a segment devoted to customers that wanted the highest performance available for performance-sensitive primary storage workloads. While performance is the key driver for NVMe technology usage, this aspect has multiple facets. Customers tend to be most interested in performance consistency with large and growing data sets, given that they will deliver latencies that are at least comparable to that achievable with internal SSD storage. Predictable performance is followed, in order, by increased throughput and/or bandwidth, increased infrastructure density (performance and/or capacity per U), tilting the IOPS/capacity ratio more in favor of IOPS, and lower latencies. The NAFA market forecast grows mostly at the expense of the SAFA market and for some workloads that are focused more around throughput and/or bandwidth specifically, the scale-out NAS market. As NVMe costs decrease, the standard matures, and the NAFA market grows, the use of NVMe will move toward significant additional usage in general-purpose storage platforms for mixed workload consolidation.

FIGURE 5

NAFA and SAFA Taxonomy



Source: IDC, 2018

At this point, there are two distinct segments of the NAFA market. The first, populated primarily by the start-ups, are systems designed to deliver optimal performance and offer a limited set of enterprise-class data services. These are used for classic "tier 0" workloads, and IDC refers to these systems as "performance oriented." The second, populated primarily by the established vendors, delivers most of the performance of "performance oriented" NAFAs but includes a much more mature storage operating system with a comprehensive set of proven enterprise-class data services. IDC refers to these latter systems as "data services oriented." This distinction is likely to be relatively short lived, however, as the performance-oriented vendors add enterprise-class data services. Even as the two system types converge, there will be a distinct difference in the maturity of features on the "data services oriented" NAFAs as compared with those of the "performance oriented" NAFAs.

The "performance oriented" NAFAs are aimed at the high-performance primary storage market, targeting very latency-sensitive workloads like super high-performance databases and real-time big data analytics. In the financial services arena, these systems are used in online trading, fraud analytics, and other workloads where low latency provides direct bottom-line business benefits. In healthcare, these systems are used for genomics, diagnostics, and other workloads where they improve the quality of patient care while lowering associated costs. Cloud providers also purchased many of these early systems, looking to better meet SLAs for premium storage tiers as their cloud infrastructures continued to grow. Oracle, Oracle Reliable Application Cluster (RAC), and SQL workloads are very commonly in use on NAFAs, as are NoSQL databases like MongoDB, Cassandra, and Redis. In most cases, customers use these platforms as dedicated storage platforms for very latency-sensitive workloads, and IDC survey data indicates that the majority of them (55.9%) are running two or fewer discrete workloads.

The "data services oriented" NAFAs address somewhat of a different market opportunity. For the most part, these systems are NVMe-based versions of proven SAFA designs that have been shipping for at least five years. The storage operating environments of these systems are mature and proven, given that many of them are based on the software designs of HDD-based storage platforms that had been shipping for, in some cases, decades. As the established vendors incorporated newer storage technologies – first persistent flash storage based on SCSI, then NVMe – they flash-optimized the software platforms. These systems bring with them all the multitenant management capabilities

necessary for mixed enterprise workload consolidation, and given that they also feature NVMe performance, they can support very dense consolidation. To date, many of these systems are sold into existing customer bases as replacements for the vendors' prior SAFA platforms, and customers are moving workloads from those older platforms over time onto NAFAs. What is interesting about these systems, however, is that they can potentially support one or more workloads that absolutely require NVMe performance along with a collection of other workloads which do not, making "data services oriented" NAFAs desirable platforms for efficient workload consolidation of all types.

Although the market for "performance oriented" NAFAs starts out larger than the "data services oriented" market, by 2019, "data services oriented" NAFAs will be generating more revenue. As the market evolves, the revenue contribution (percentagewise) of "data services oriented" NAFAs will quickly increase. But for many customers, these systems will be running at least one or two "tier 0" workloads that might otherwise have driven the purchase of a "performance oriented" NAFA. Customers that desire the absolute most performance or for some reason want to keep a very latency-sensitive workload separate from other general-purpose workloads will be the ones that buy the "performance oriented" NAFAs.

While the market between performance- and data services–oriented NAFAs will blur over time, absolute performance and maturity differences between the platforms will still drive relevant differentiation.

METHODOLOGY

The five-year revenue forecast published in this study is based around IDC's Worldwide Quarterly Enterprise Storage Systems forecast released in June 2018, which provides a detailed forecast by region and quarter for 1Q18–4Q19 and annually thereafter through 2022. The quarterly forecast covers the AFA market segment from which this NAFA forecast is drawn.

The tables and figures in this study are generated from a proprietary IDC database and analytical tools. Our census process researches enterprise storage information on a product-, vendor-, and geography-specific basis. IDC's methodology provides customers with a nearly unlimited ability to analyze the enterprise storage market from many perspectives. IDC's Continuous Intelligence Services and consulting services can provide additional insights beyond the scope of this study using the database supporting it.

Note: All numbers in this document may not be exact due to rounding.

LEARN MORE

Synopsis

This IDC study presents the 2018–2022 forecast for the NVMe-based all-flash array market.

"The NVMe-based all-flash array market will grow very quickly as it cannibalizes revenue from the SCSI-based all-flash arrays that today are dominating primary storage spend," said Eric Burgener, research vice president, Storage. "The rapid growth of new workloads that demand NVMe performance and the fact that the transition to NVMe will be much easier than the transition from spinning disk to persistent flash was over the past decade are two key factors driving the strong growth of this market."

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

5 Speen Street
Framingham, MA 01701
USA
508.872.8200
Twitter: @IDC
idc-community.com
www.idc.com