All-Flash Buyer's Guide

Tips For Evaluating Solid State Arrays

JUNE 2017

NetApp®

Think Flash-First for Primary Storage

With the introduction of enterprise-grade all-flash arrays, conventional thinking about storage has been turned on its head. For example, a recent Gartner study[1] found that all-flash storage pays for itself in just five to six months on average, the result of dramatic improvements in total cost of ownership (TCO). Gartner reported that operating expenses for administration, power, space, cooling, and maintenance were all significantly reduced. Reductions in software licensing costs and improvements in IT productivity were also contributors.

The question is no longer when to choose all-flash storage. You now must ask yourself, "When would I *not* choose all-flash storage?"

If you're running a traditional storage architecture with spinning disks, you're probably aware of the growing challenges. Many IT teams struggle to meet performance SLAs. You may be spending increasing amounts of time trying to tune performance, a problem made more difficult by new applications with different I/O profiles and performance demands. As you shuffle data around from one storage system to another to improve performance, it increases the difficulty of maintaining security and compliance, adding to your team's workload.

Perhaps your current storage systems are behind the technology curve or approaching the end of their warranty, or perhaps the future plans of an incumbent storage vendor are uncertain. Whatever the reason, there has never been a better time to consider switching to all-flash storage.

This guide will help you understand what's important and help you evaluate storage options based on the most important selection criteria.

[1] Gartner, *Solid-State Array TCO Reality Check*, January 22, 2016.

Evaluating All-Flash Storage

It's important to recognize that all-flash storage solutions may vary widely in terms of features and capabilities. For most IT organizations evaluating flash storage, the primary considerations are:

- Performance
- Capacity and scalability
- Resiliency and availability
- Support for future needs
- Total cost of ownership

Each of these topics is covered in the sections that follow. As you consider these purchase criteria, it will be helpful to evaluate the strengths and limitations of your current storage solution. This can help you focus on the criteria that are most important based on your organization's needs.

"Solid-state arrays deliver superior performance and are increasingly replacing legacy and hybrid external controller-based storage systems."

—Gartner Solid-State Array TCO Reality Check, January 22, 2016

All-Flash Performance

Most all-flash storage systems on the market utilize solid-state drives (SSDs) available from a number of different vendors (although a few vendors persist in building storage systems using custom NAND Flash modules rather than commercial SSDs).

SSDs deliver superior performance versus hard disk drives (HDDs) in terms of the three most common storage performance metrics:

- I/O Operations per Second (IOPS)
- Throughput, usually measured in MB/sec
- Latency, usually measured in milliseconds or microseconds

Different applications generate different I/O workloads. When you consider all-flash storage options, it's important to think about the I/O characteristics of the workload or workloads you need to support. All-flash storage systems that offer quality of service (QoS) controls can provide significant advantages in situations where you are consolidating multiple applications on a single storage system, preventing any one application from consuming too many resources and allowing you to prioritize the most important applications.

IOPS

Transaction-oriented applications such as databases generate small, random reads and writes. This type of storage performance is measured in IOPS and is the most common metric reported for all-flash storage. For IOPS comparisons to be meaningful, it's important to know the operation size and the mix of read and write operations.

This is the metric where SSDs and all-flash arrays are most differentiated from traditional HDD-based storage. A single SSD can deliver IOPS measured in the tens of thousands where a single HDD can only deliver hundreds of IOPs, which means it requires lots of HDDs to achieve the performance of a single SSD. If you're used to provisioning hundreds of spindles to storage performance, this will be a major difference from what you're used to.

Most vendors publish the IOPS ratings for their all-flash arrays. Tips for comparing IOPS performance can be found on page 7.

Latency

Latency is a measure of how long it takes a storage system to satisfy an I/O. It is an important metric for time-critical applications such as real-time trading and online transaction processing (OLTP), which are extremely latency sensitive. Interactive users notice and react to differences in latency. According to Amazon.com, 100ms in increased latency corresponds to a 1% drop in sales.

Although many all-flash storage systems can support latencies as low as 1ms, applications that require extreme performance may require consistent, sustained response times that are measured in hundreds of microseconds rather than milliseconds. These applications do not tolerate latencies that fluctuate due to backend storage services or that increase rapidly as the IOPS load increases.

Throughput

Applications such as data warehouses and video rely on sequential access to data in large blocks, resulting in a workload that's dramatically different from transaction-oriented applications. Throughput is a measure of the amount of data that can be moved in or out of a storage system and is typically reported in MB/sec or GB/sec.

Not all vendors report throughput numbers. If you know you have a throughput-oriented application, you should make sure to evaluate throughput performance.

High-Performance Connections

As the performance and density of all-flash storage systems increases, the likelihood of network bottlenecks also increases, making network performance an important consideration. In a recent study of all-flash arrays and network performance, Gartner found that I/O performance bottlenecks have already moved to the network, concluding that network upgrades can double or quadruple application performance.[2]

[2] Gartner, *Slow Storage Replication Requires the Redesign of Disaster Recovery Infrastructures*, October 10, 2016.

Comparing Performance

There can be wide performance variations across vendor systems, and even wide variability within a single system, depending on a number of factors. For example, some architectures may have markedly better read performance than write performance.

In general, you should seek solutions that can demonstrate consistent, scalable IOPS performance at latencies under 1ms based on third-party benchmarks that simulate real-world workloads, such as SPC-1 and TPC-E. You should also take claims about "maximum IOPS" with a grain of salt unless a vendor can provide you with more details. For example, without the answers to the following questions, you can't make a valid apples-to-apples comparison of IOPS performance among competing solutions:

- What was the I/O size used to run the test?
- How does this compare with the I/O size for my applications?
- What was the read/write mix?
- Were the I/Os random or sequential?
- How many IOPS are supported at 1ms of latency?
- What QoS features are available to make sure that IOPS are available when you need them?

IS YOUR NETWORK FAST ENOUGH TO KEEP UP WITH GROWING ALL-FLASH STORAGE DENSITY?

Just a handful of SSDs have the potential to saturate typical 8Gb and 16Gb Fibre Channel SANs or 10 Gigabit Ethernet networks. When you increase the density and capacity of an all-flash storage system, expect the network bandwidth requirement for that storage system to increase at the same rate. Unless you plan ahead, network bottlenecks are the inevitable result. Look for all-flash storage that supports multiple network connections and the latest high-speed network options, including 32Gb Fibre Channel and 40 Gigabit Ethernet.

Density and Scale

IT organizations are updating to all-flash storage to achieve:

- **Increased density.** Replacing older equipment with the latest all-flash systems can free up valuable space and decrease the money spent on power and cooling.
- **Greater flexibility.** With the IT environment evolving rapidly, you have to make storage choices that give you the greatest flexibility to move in different directions in the future.
- **Infrastructure standardization.** By eliminating specialized storage infrastructure to support different applications and standardizing your hardware selections, you create a simpler, more flexible data center that adapts to new requirements more readily.

The storage density of SSDs is growing faster than HDDs. In 2016, the first 15TB SSDs were released, exceeding the capacity of the densest HDDs available at that time by more than 50%. Since then, even higher-density SSDs have been announced.

From a practical standpoint, this rapid density increase means that a few rack units of all-flash storage can take the place of multiple racks of HDD storage from just a few years ago in terms of both capacity and performance, allowing data centers to recover space and significantly reduce power and cooling expenses.

However, not all storage systems can accommodate the latest high-density SSDs. If maximizing capacity is your goal, this should be part of your buying decision.

Effective Capacity and Storage Efficiency

When you begin comparing all-flash storage, you'll notice that many vendors report two capacity metrics: raw capacity and effective capacity. The effective capacity is a measure of how much data the vendor expects the storage system to hold after deduplication and compression are applied.

Most all-flash arrays on the market today provide inline deduplication and compression, reducing the total amount of storage you must purchase to store data. There are

differences between the effectiveness of storage efficiency technologies from one vendor to another, but, as with most things, be skeptical of vendor claims that seem too good to be true.

While inline deduplication and compression are often considered table stakes when evaluating all-flash systems, exceptions exist, as noted in the performance section above. For applications that demand the lowest possible latency, eliminating inline storage efficiency technologies reduces the data path, lowering latency and making I/O delivery more predictable.

All-Flash Scaling

There are two approaches you can take when scaling all-flash storage: scale-up and scale-out. With the familiar scale-up method, you add SSDs to a dual-controller storage system until you reach the capacity or performance limit. When that happens, you must either replace the existing storage system with a more powerful one or add another separately managed system.

With the scale-out method, storage controllers are clustered together such that additional controllers can be added to scale performance and capacity beyond the limits of a conventional dual controller system, offering a single storage pool and simplified management versus multiple independent systems. Scale-out architectures can eliminate the complexity and disruption that may result when scaling capacity.

Resiliency and Availability

SSDs have been available in the market for a number of years, and industry data indicates they are inherently more reliable than HDDs. SSDs do fail, however, so standard forms of data protection still apply. Most all-flash storage systems incorporate some form of redundant array of independent disks (RAID).

Given their much higher performance, if an SSD that is part of a RAID group fails, rebuilds happen much more quickly, limiting exposure to a second failure. In many situations, flash provides enough performance cushion that your users and customers are unlikely to see a change in performance during a rebuild.

Avoiding Planned and Unplanned Downtime

Just because SSDs are reliable, doesn't mean that an all-flash storage system isn't subject to both hardware and software failures that can cause unplanned downtime. It's always good advice to choose a vendor with a well-designed architecture, mature processes, a proven track record of reliability, and first-class support and professional services.

In most data centers, planned downtime for maintenance, upgrades, etc., is almost a thing of the past. As you consider upgrading to all-flash, look for an architecture

IMPROVED RELIABILITY

Some IT professionals continue to be concerned about the durability of flash media, or flash "wear-out," even though a wealth of real-world data demonstrates that SSDs are extremely reliable. The good news is that most vendors are so confident in the long-term reliability of their all-flash systems that they're happy to provide you with an extended warranty. Other options, such as free controller upgrades, may also be available if you are willing to commit to a long-term service contract.

You'll need to factor all of the costs into your TCO calculations to make a direct comparison among vendors. For example, you may discover that the potential savings from some "free" controller upgrade options are more than offset by higher annual service charges.

that supports nondisruptive operations for all common maintenance activities.

Data Protection and Disaster Recovery

No matter how reliable the architecture you choose is, you still need backup and disaster recovery (DR) to protect against user errors, bugs in application software, widespread power outages, and other natural and manmade disasters.

Modern storage architectures typically integrate snapshot and replication functions that make backup, recovery, and DR functions faster and more efficient. A mature all-flash storage system should include data protection and DR features, including snapshots, asynchronous and synchronous replication, application-level integration, and support for an ecosystem of data protection partners.

Streamlined Management

IT infrastructure is supposed to be a means to an end, not an end in itself. The continuous monitoring, management, and maintenance required with traditional storage systems, however, may be keeping your IT team from focusing enough attention on the applications and services that move your business forward.

Look for all-flash storage that:

- Reduces or eliminates routine management tasks
- Enables maintenance and upgrades to be performed without planned downtime
- Offers fast and efficient data management capabilities including snapshot copies, replication, and cloning
- Provides intuitive user interfaces with role-based access control
- Integrates with your existing management and monitoring framework
- Provides REST APIs, PowerShell cmdlets, plug-ins and/or other integrations that make it easy to automate and orchestrate storage in your environment

IS YOUR ALL-FLASH STORAGE NVME-READY?

Non-Volatile Memory Express, or NVMe, is an alternative to the SAS and SATA drive interfaces, providing a PCIe interface optimized for flash SSDs and storage class memory (SCM) devices. NVMe greatly increases parallel I/O, delivering higher performance and lower latency while reducing the load on storage system CPUs.

NVMe over Fabrics (NVMe-oF) is a new interconnect that extends the PCIe bus between servers and storage. New storage systems are being readied with NVMe SSDs and others will include an option to deploy NVMe over Fabrics, so it will be important to understand where each storage vendor stands with regard to this next-generation technology.

Learn more about NVMe.

Future-Proof Architecture

The all-flash storage you choose should be able to adapt to your future needs. The storage you purchase today should be able to support the changing needs of your existing applications, new cloud applications, or both.

Existing Applications

An all-flash solution must fit easily with your current application environment. Out-of-the-box integration with enterprise applications, such as those from Citrix, Microsoft, Oracle, VMware, and SAP, will facilitate provisioning and application-consistent data protection.

It should also be able to accommodate the future needs of your enterprise applications. This could include: the capability to support faster network technologies or other infrastructure upgrades, the capability to support a different storage protocol or add NAS support to an existing SAN system, or to replicate data from flash storage to other media, either on premises or in the cloud.

Cloud Applications

Next-generation cloud applications and services are typically optimized to run in the public cloud or on top of a private cloud or enterprise cloud infrastructure. These environments may also require the all-flash capabilities described in the previous section. In addition, they will benefit from scale-out, easy automation, full REST APIs, and QoS.

If your organization is developing next-generation applications, you may need to be able to run them both on premises and in the cloud. Storage that offers cloud integration can simplify the migration of applications to and from the cloud. You should consider all-flash systems that can facilitate the following functions:

- Move workloads to and from the cloud
- Tier data between on-premises and cloud storage
- Backup on-premises storage to the cloud
- Provide DR in the cloud

Total Cost of Ownership

For most IT organizations, cost is a mitigating factor in purchasing decisions. As noted earlier, Gartner and other studies have found that all-flash storage can significantly lower TCO relative to traditional storage systems. You may need to do a TCO comparison of a proposed all-flash array versus your existing disk-based or hybrid flash systems to justify the purchase. Be sure to factor in the following items as part of your analysis:

Effective capacity. This is the amount of logical capacity available to store application data after storage efficiencies are applied. At a minimum, your calculations should include the inline efficiencies available with an all-flash system, such as data compression and deduplication. You should also include snapshot copies and clones if you intend to use them.

Application integration. The management costs for provisioning, protecting, and recovering application data are often overlooked when evaluating primary storage. However, the capabilities and costs for managing application-consistent data copies can vary widely across all-flash offerings and should be included when calculating TCO.

Data center operational costs. Power, cooling, and data center real estate are all important operational costs that must be considered. According to one user survey, all-flash storage systems provide over 70% savings in power and cooling when compared with disk-based systems.

IT management. When storage performance no longer needs to be treated as a scarce resource, IT personnel are freed up across the entire organization. Database, application, and infrastructure specialists can all reclaim time that was previously spent architecting and managing storage performance.

Software licenses. The consolidation of workloads onto an all-flash platform can result in software savings. Database software licenses and maintenance fees are often reduced because the improved I/O performance of all-flash storage systems reduces the number of CPU cores or servers needed to achieve the same level of application performance.

NetApp All-Flash Solutions

The NetApp® portfolio of all-flash solutions includes several product families, each optimized to deliver maximum benefits for particular workload profiles.

NetApp AFF

NetApp All Flash FAS (AFF) systems provide a high-performance solution for consolidating multiple SAN and NAS workloads on a single, unified scale-out architecture. Designed specifically for flash, AFF all-flash systems deliver industry-leading performance, capacity density, scalability, security, and network connectivity. As the first all-flash arrays to provide both 40 Gigabit Ethernet (40GbE) and 32Gb Fibre Channel connectivity, AFF A series systems eliminate the network bottlenecks that can occur as flash storage systems get faster and with denser storage capacities.

AFF offers a number of leading innovations including the latest SSD technologies. It is the first all-flash array to support 15TB SSDs and Multi-Stream-Write (MSW) SSDs, dramatically changing data center economics.

Built on ONTAP® storage software, AFF systems provide a common set of tools for data management and integrated data protection either on premises or in the cloud, enabling seamless movement of applications and data to wherever they run best.

NetApp AFF systems enable you to:

- Consolidate and accelerate a wide range of workloads on a single platform. Ranked in the top three for performance in the latest SPC-1 benchmark testing with 2.4M SPC-1 IOPS at 0.69ms of latency, AFF systems are ideal for supporting the demands of a shared environment of enterprise applications and data analytics with millions of IOPS and hundreds of petabytes of effective capacity in a single cluster.
- Simplify IT operations while transforming data center economics, reducing power consumption by up to 11 times and rack space by up to 19 times, and slashing support and performance-tuning costs to a third compared with hybrid systems.

AFF

Enterprise-grade flash, ready for cloud

- **Consolidation and virtualization**
- **Enterprise applications**
- **Engineering and design workloads**
- **Big data analytics**

EF-Series

Field proven platform for low-latency performance

- **Data analytics**
- **Databases**
- **Dedicated applications**
- **High-performance computing**

SolidFire

Scale-out, service-oriented flash for innovative architectures

- **Virtualization**
- **Private cloud**
- **DevOps and SaaS**
- **Managed cloud and hosting**

- Future-proof your infrastructure and investment with maximum flexibility resulting from a scale-out architecture, best-in-class cloud integration, and NVMe and NVM-oF-ready technologies.

NetApp SolidFire

SolidFire® systems combine the performance and economics of all-flash storage with a web-scale architecture that radically simplifies data center operations and enables rapid deployment of new applications using a scale-out all flash storage platform designed for large-scale infrastructure.

SolidFire systems offer a unique capability to manage storage performance independent of capacity and guarantee that performance not just to one application, but to hundreds of applications within a single storage platform. These capabilities, combined with granular scale-out and the capability to automate every aspect of the storage

platform, enable you to streamline operations and respond to changing business and application demands faster and more predictably.

These systems are optimized for virtualized private cloud and service provider environments where predictable application performance, on-demand scalability, and orchestrated operations are top-level requirements to meet escalating demands for flexible and efficient IT service, delivery and consumption. SolidFire systems provide:

- The capability to start with exactly what you need and quickly scale node by node to address growing business demands
- User-defined minimum, maximum, and burst QoS settings so you can guarantee performance SLAs for applications, workloads, and tenants across your infrastructure
- A 100% programmable platform that automates complex storage management tasks and delivers deep management integration into next-generation frameworks including OpenStack, VMware, and Docker

NetApp EF-Series

The EF-Series all-flash arrays are designed specifically for high-speed transactional applications that demand high IOPS and consistent low latency. A single EF-Series all-flash system in a 2U enclosure can deliver the performance of over one thousand 15K RPM hard disk drives while requiring just a fraction of the rack space, power, and cooling. The extreme performance provided by EF-Series arrays can:

- Deliver faster, actionable results from OLTP and OLAP environments
- Improve the performance of data analytics applications by more than 50%
- Enhance the user experience for customer-facing and decision support systems that require consistent microsecond responses

The NetApp E2800 supports all-flash configurations in a system optimized for small enterprises and remote offices that require mixed-workload performance and reliability for dedicated apps.

For More Information

NetApp AFF A700 Performance with Microsoft SQL Server 2014 (TR-4580)
http://www.netapp.com/us/media/tr-4580.pdf

NetApp AFF A700 Performance with Oracle Database (TR-4582)
http://www.netapp.com/us/media/tr-4582.pdf

Introduction to NetApp EF560 Flash Array (TR-4546)
http://www.netapp.com/us/media/tr-4546.pdf

SolidFire: Storage Built for the Next Generation Data Center
http://www.netapp.com/us/media/ds-3773-solidfire.pdf

SolidFire: The Five Principles of the Next Generation Data Center
http://www.netapp.com/us/media/ds-solid-fire-article-five-principles-next-generation-data-center.pdf

NetApp All-Flash Customer Case Studies
http://www.netapp.com/us/company/customer-stories

NetApp®