White Paper

Business Case for All-Flash Arrays When Integrating Flash into the Cloud

Sponsored by: Pure Storage

Eric Burgener
November 2016

IDC OPINION

The use of cloud-based services is growing rapidly in today's 3rd Platform computing environment. Datacenters looking to craft the most cost-effective infrastructure to service legacy workloads like relational database and enterprise applications while accommodating next-generation applications in the mobile computing, social media, and big data and analytics areas are increasingly moving to hybrid cloud environments. For performance, security, and/or control reasons, administrators continue to host certain workloads in on-premises infrastructure while moving other workloads to the cloud to take advantage of the economics, agility, and more flexible pricing models that the cloud offers. However, all cloud infrastructures are not webscale, and many cloud-based infrastructures are using all-flash arrays (AFAs) to offer reliable, highly available premium service tiers that must conform to stringent service-level agreements (SLAs). The use of AFAs in this manner is on the rise across the enterprise storage vendors that have all-flash offerings well suited to the requirements of dense, multitenant environments. When choosing AFAs over hyperconverged architectures to build cloud infrastructure, purchasers typically cite one or more of the following:

- **Economics.** Better storage efficiency technologies that drive higher data reduction ratios (and hence a lower effective cost per gigabyte [GB]), combined with the efficiencies overall of highly flash-optimized platforms, drive a compelling total cost of ownership (TCO) argument.
- **Maturity.** The network-based array model is more familiar to many administrators and offers capabilities that are proven over longer periods of time than newer, hyperconverged platforms with workloads that require extremely high reliability and availability.
- **Data services functionality.** These platforms offer a better set of features necessary to reliably manage dense, multitenant infrastructures.
- **Degraded mode performance.** For workloads that require predictable performance, AFAs can support an infrastructure that suffers no performance impact in the event of underlying failures. For hyperconverged offerings, while supporting transparent recovery, performance will vary depending on the type of underlying failure.

With their FlashArray, vendors like Pure Storage meet these requirements and are not just selling these solutions to enterprises; these vendors have a surprisingly large percentage of their businesses coming from service providers and other organizations looking to cloud infrastructures to reliably deliver better economics and increased agility.

IN THIS WHITE PAPER

Over time, hybrid cloud will increasingly become the mainstream deployment model for IT infrastructure. Flash storage brings with it many benefits necessary in hybrid cloud environments, and IDC already views it as a requirement for enterprise workloads that have any performance sensitivity. This IDC white paper discusses the state of enterprise storage with respect to the evolving cloud storage market, explains why flash storage is needed in these environments, and then discusses what Pure Storage, a leading all-flash array vendor, brings to the table in this area. The document concludes with a short service provider case study.

SITUATION OVERVIEW

While the overall enterprise storage systems market is expected to grow only at a compound annual growth rate (CAGR) of 2.9% through 2020 (when it will reach $47.0 billion), there are significant shifts happening between storage spend that are driving strong growth over that same forecast period in certain submarkets. Enterprise vendor growth is strongly dependent upon an ability to effectively target these high-growth segments, and AFAs are clearly one of those areas. As legacy storage systems come up for technology refresh, enterprises and service providers alike are increasingly choosing AFAs to host primary workloads. While the AFA market will grow at a CAGR of 21.4% to crest $7.74 billion in 2020, the hybrid flash array (HFA) market is growing only at a CAGR of 1.9% over that same period, and revenue for external storage arrays built only from spinning disk will actually decline at a CAGR of 11.2%. IDC expects that by 2020, AFAs will drive over 70% of the spend for primary external storage, and as flash costs continue to plummet over the next several years, it will increasingly become an option for other types of file- and object-based workloads.

Cloud storage is another high-growth submarket in the enterprise storage systems arena. As IT organizations look to accommodate rapidly changing business conditions, cloud agility and economics become increasingly attractive. Certain workloads that require guaranteed high performance and/or must meet specific security requirements continue to be hosted in on-premises infrastructure, but new workloads and DevOps projects are often hosted in the cloud. Increasingly, enterprises are using software-as-a-service (SaaS) delivery models for common business applications like enterprise resource planning (ERP), customer relationship management (CRM), and IT service management (ITSM). IDC's recent research indicates that 72% of IT organizations already have a hybrid cloud environment in place and will be moving more workloads to the cloud over time. Cloud spending, made up of public cloud, private cloud off-premises, and private cloud on-premises revenue, is growing at a CAGR of 14.9% and should crest $48.3 billion in revenue by 2020. Overall storage spend is slowly shifting toward the cloud – in 2015, 49% of all enterprise storage spend had a cloud component but that will increase to 54% by 2020.

The grand challenge for enterprises today is how to continue to reliably service existing workload requirements while accommodating new workloads that do not perform well on legacy infrastructure. Virtual infrastructure is the mainstream deployment model for most datacenters – enterprises and service providers alike – and workload consolidation on these platforms creates a very random I/O pattern that is not well handled at all by spinning disk technology. Increasing infrastructure density heightens the impact of failures, driving availability requirements ever higher. Meanwhile, data continues to grow at a rapid rate, causing nondisruptive scalability to become a key purchasing criterion. And much more dynamic business conditions are forcing IT organizations to become more agile to avoid posing as a stumbling block in their firm's quest to maintain competitiveness.

Increases in flash and cloud spending are positively correlated. Relative to legacy storage infrastructure, flash brings much higher and more predictable performance. With 2nd Platform computing (i.e., client/server), storage performance requirements were often more predictable, and dedicated storage silos could be tuned to meet particular requirements. One of the characteristics of next-generation workloads, particularly in cloud-based environments, is a lack of predictability, yet IT organizations are expected to reliably meet SLAs regardless of widely varying I/O patterns. At the same time, consolidating multiple workloads onto the same virtual infrastructure has become the accepted deployment model for most IT organizations. With its ability to burst to tens of thousands of IOPS per device (and hundreds of thousands of IOPS per array) and consistently deliver submillisecond latencies even in the face of extremely random I/O, the use of flash is required to meet performance expectations. Most clouds, regardless of whether they are public or private, have at least a caching tier built from flash, and an increasing number of clouds are deploying persistent flash tiers that give them the flexibility to offer and charge for premium service levels.

The economics of cloud-based offerings are one of the main draws to incorporate the use of cloud storage for enterprises, but competition is fierce among service providers and margins are thin. Efficiency measures concerning infrastructure density, energy consumption, and administrative productivity are top of mind as IT practitioners build cloud-based infrastructure, and availability and reliability concerns are prompting many of them to use AFAs rather than newer, less proven architectures like hyperconverged. Today, solid state drives (SSDs) already offer higher storage densities than hard disk drives (HDDs), and with the performance that flash delivers, AFA vendors can build systems with significantly higher performance and capacity density that take up less floor space and use generally less than half the energy consumption. Inline storage efficiency features like compression, deduplication, and thin provisioning minimize the amount of raw storage capacity required to store and use a given amount of data while still being able to consistently deliver submillisecond latencies against highly varying workloads. Other key features available in AFAs include built-in data integrity checking, redundant designs that support transparent recovery from component failures without performance impact, nondisruptive upgrades and system expansion, replication capabilities that enable the creation of disaster recovery configurations, and support for a number of APIs and scripting interfaces that make integration into datacenter workflows easier.

Even as data sets get larger, they must also be increasingly mobile. Flash performance helps make the workflows that are dependent on data movement, such as backup, disaster recovery, and migrating workloads between on-premises and off-premises locations, faster. Other data services like inline data reduction, space-efficient snapshot and clone creation, and WAN-optimized replication make data movement operations much more efficient by minimizing the amount of data that needs to be moved to meet workflow requirements. In enterprise storage platforms that are purpose built around flash media, these operations can be much more efficient than older designs that were originally architected for use with HDDs.

Pure Storage FlashArray: Proven All-Flash Performance for the Cloud

Pure Storage is an AFA vendor that shipped its initial products in 2011 and has since distinguished itself as an enterprise storage vendor that is changing the industry. Pure Storage's early championing of inline data reduction technology with the company's block-based shared storage array (the FlashArray) has made the feature a baseline requirement for success in the AFA market. Pure Storage's Evergreen Storage program provides, from a customer's point of view, a much better model for ongoing maintenance and technology refresh and has prompted direct responses from the company's major competitors. Pure Storage's focus on the quality of the customer experience,

including everything from product to sales and support, has differentiated the company from the legacy enterprise storage players and has changed customer expectations around how enterprise storage providers should treat their customers. Indeed, Pure Storage continues to produce one of the top net promoter scores (NPSs) among enterprise storage array providers as independent evidence of the customer satisfaction it engenders. NPS is a standardized measure of customer experience, tabulated by third parties, which is used across 220 different industries worldwide. For more information on NPS, see *Why Enterprise Storage Managers Need to Understand the Net Promoter Score* (IDC #US41185416, April 2016). Pure Storage is already a $600+ million company well on its way to a billion in revenue over the course of the next several years and is still growing faster than the AFA market itself.

Pure Storage characterizes its enterprise storage platform offerings as "smart storage that is effortless, efficient, and evergreen." **Effortless** refers to management ease of use — the FlashArray is designed for management simplicity and supports key APIs that enable excellent integration into datacenter workflows, minimizing manual involvement in array management through the use of automation. Pure1, the company's cloud-based predictive analytics offering, comprehensively monitors systems, tracking performance and data reduction statistics and providing input that informs upgrade strategies, best practice implementations, predictive maintenance, and capacity planning. **Efficient** refers to metrics such as performance and storage density, data reduction ratios, and energy consumption as well as flash-optimized data services such as encryption, quality of service (QoS), snapshots, replication, and migration. **Evergreen** refers to the technology refresh model that allows customers to nondisruptively ride technology curves to incorporate next-generation capabilities and preserve investments in hardware and software spend across technology generations.

Despite all the discussion around webscale architectures and the benefits they provide, a surprising number of IT organizations building cloud infrastructures, regardless of whether the infrastructures are private or public, are citing economic, maturity, data services functionality, and degraded mode performance reasons for using AFAs. As one of the top 3 market shareholders by revenue in the AFA space, Pure Storage benefits significantly from this thinking. Cloud infrastructure customers — both enterprises and service providers alike — see the following benefits in using the Pure Storage FlashArray//m products:

- **Economic.** While flash media is still more expensive than spinning disk on a raw cost-per-gigabyte basis, well-architected AFAs actually deliver a TCO, including both capital and operating expenditure, that can be 50–80% lower over a three-year period to deliver the same performance in terms of throughput. Flash latencies (which are roughly 10 times lower than HDD latencies) enable the use of inline storage efficiency technologies like compression, deduplication, and thin provisioning against latency-sensitive primary storage workloads while being able to still consistently deliver predictable, submillisecond latencies (compare this with HDD latencies that can easily vary between 5ms and 20ms or more, with variable workloads such as are experienced in consolidated, virtual infrastructure environments). Data reduction ratios vary on the basis of workload, but even a ratio of 3:1 (which is very conservative for typical mixed virtual workloads) actually makes effective cost per gigabyte much lower with flash. This also means that AFAs need far less raw storage capacity to store any given amount of data, a fact that increases infrastructure density and lowers floor space and energy consumption as well as backup and restore times.

  IT administrators using HDDs with performance-sensitive workloads often spend as much as 10–20 hours per week, fielding customer complaints about storage performance and manually tuning systems to meet performance commitments. Given that AFAs typically deliver way

more performance in terms of latency, throughput, and bandwidth than most datacenters need for mixed virtual workloads, it is very common for administrative time spent on performance issues to go to zero with AFA deployment. For performance-hungry workloads, the use of flash shows a performance improvement that is noticeable by end users. It is typical that, in the wake of AFA deployment, administrators are able to consistently deliver to their SLAs, customers are happier with performance, and datacenters spend far less time dealing with performance issues.

- **Maturity.** Administrative planning is clearly easier when infrastructure capabilities are known, well understood, and predictably reliable. Pure Storage has been delivering enterprise-class storage solutions since 2011, and with several thousand production deployments among Fortune 5000 companies, the company has proven that it delivers high-performance, highly available AFAs for bet-your-business workloads. This is evident in product design that supports easy online upgradability within or across technology generations, guarantees around data reduction ratios, flash endurance and "flat and fair" maintenance costs, and an extremely high repurchase rate, among Pure Storage's customers, that is consistent with the company's high net promoter score. Another example of the maturity of the FlashArray platform is its broad support of orchestration platforms that allow companies to achieve overall infrastructure automation with less custom integration work.
- **Data services functionality.** In many cloud environments, high infrastructure density is desirable. Dense consolidation of workloads helps drive lower costs per workload, but being able to reliably support high infrastructure density is not just about performance and capacity density. It must also support highly consolidated workloads and multitenant environments while providing the necessary management capabilities for data protection, disaster recovery, and other required datacenter workflows. This would include storage efficiency technologies such as compression, deduplication, thin provisioning, space-efficient snapshots, and WAN-optimized replication as well as other capabilities around QoS, encryption, and datacenter integration (for automation purposes).
- **Degraded mode performance.** Failures occur even in the most reliable systems, and differing designs generate different degraded mode performance. In business models that require predictable performance, administrators need to understand the impact on application services of various types of failures (an SSD, a controller, etc.). In the event of an SSD failure, FlashArray's dual-parity RAID implementation transparently recovers without performance impact. In the event of controller failures, the same is true — in normal operation, two controllers share the front-end I/O load, but a single controller can deliver the full-rated IOPS performance of each model. Failed components can be quickly replaced online without impacting application performance, bringing the system back to full redundancy.

Delivering better storage performance is still the single biggest driver of AFA purchases overall, but as IT organizations increasingly move toward the use of AFAs as general-purpose enterprise storage platforms for mixed workloads that can include primary, secondary, and test and development environments, they are expecting more than just performance. Vendors like Pure Storage that are offering flash performance and efficiencies in a platform that also delivers on the other enterprise requirements of scalability, availability, reliability, and manageability have been the most successful to date.

Predictability: A Key Driver of Cornerstone OnDemand's FlashArray Purchases

Cornerstone OnDemand helps organizations recruit, train, and manage their people by delivering cloud-based human capital management (HCM) products and services. The company's solutions are in use across nearly 28 million people in 191 countries and throughout a variety of different industries. The company delivers its applications through a large private cloud/software-as-a-service infrastructure globally. As the company's business grew, continuing to deliver to SLAs was becoming more of a challenge and taking more and more administrative time, and the company began to explore options for integrating flash media into its primarily HDD-based storage infrastructure. Key concerns included an ability to provide consistent and predictable performance for Cornerstone OnDemand's database workloads and isolating the company's customers from any performance impact in the event of component failures. Cornerstone OnDemand uses hyperconverged infrastructure for some of its workloads, but failures in those environments result in performance degradation that limits its ability to deliver on stringent SLAs.

Cornerstone OnDemand brought in its first Pure Storage FlashArrays in 2015. Although the platform was initially selected for performance and reliability reasons, other benefits soon became apparent. The data reduction ratio of FlashArrays, which averages between 3:1 and 4:1 across the company's mixed workloads, was key in making the choice a cost-effective one, and its nondisruptive operations capability allowed the company to easily meet strict uptime requirements in the face of component failures, firmware upgrades, and ongoing integration of next-generation storage technologies. With flash's high performance, time spent dealing with storage performance issues has dropped to near zero. Cornerstone OnDemand uses the snapshots and replication of FlashArrays to integrate into an automated data protection routine that allows the company to meet industry's best-of-breed recovery point objectives and support its ability to deliver on an impressive "four-nines" availability target for its global service delivery. Relative to competitive offerings, Pure Storage offers noticeably better technical support, and its data reduction achieves higher ratios than other network storage arrays and hyperconverged offerings that Cornerstone OnDemand had as part of its infrastructure.

While Cornerstone OnDemand is not moving toward a 100% all-flash datacenter, its strategy does include putting all performance-sensitive workloads (particularly where there are SLAs involved) on flash. The cloud provider places all its mission-critical, performance-intensive database workloads on Pure Storage, has over 260PB of storage capacity spread across multiple FlashArrays, and is looking to move more workloads to these platforms as its business continues to grow. Lower-performance tiers are still serviced by the company's legacy HDD-based infrastructure, but as these systems come up for technology refresh and flash costs continue to drop, the company expects that it will be moving some of these workloads to FlashArrays as well as for overall efficiency, infrastructure density, and ease of management reasons.

FUTURE OUTLOOK

The attraction of cloud infrastructure use for IT organizations is based on the economics, ease of scalability, agility, and granular consumption models it enables. As enterprises and cloud providers build out their own cloud infrastructures, IDC expects there to be a mix of storage architectures used. While the monolithic, legacy storage arrays of yesteryear that were built around HDD-based technology will not often be used because of their expense and complexity, purpose-built AFAs will be widely used to create high-performance tiers in cloud-based environments. These systems are optimized to get the most out of flash media in terms of performance, storage density, reliability,

energy consumption, and cost, and they provide a more familiar model that offers better degraded mode performance than the other storage architecture that will be widely used in cloud environments — hyperconverged. Along with cost considerations, how well various vendor implementations of each of these architectures can be integrated into the automated operations strategies of enterprises and cloud providers will be another strong determinant of their adoption.

All-flash offerings are moving quickly to become the mainstream deployment options for performance-sensitive workloads. However, HDDs are not expected to go away anytime soon and will continue to be widely used for lower-performance workloads that are particularly cost sensitive. While flash costs are coming down at a much faster rate than HDD costs, they are not likely to drop down to $0.03 per gigabyte that 7,200rpm HDDs will be offering in 2017 for quite a while yet (if ever). The use of 15,000rpm and 10,000rpm HDDs will continue to recede at a rapid rate as flash costs drop to parity over the next two to three years while offering significantly better latency, throughput, bandwidth, performance consistency, reliability, storage density, server CPU utilization, and lower energy consumption. But most cloud infrastructures, which will ultimately be hosting both performance- and cost-sensitive workloads, will be made up of a mix of both flash and spinning disk technology.

CONCLUSION

For on-premises infrastructure, AFAs are rapidly replacing HFAs for performance-sensitive primary workloads. But as AFAs have continued to evolve, improving their efficiencies and multitenant support capabilities and lowering their cost, IT organizations are deploying them with a wider set of workloads, leveraging the infrastructure density they enable to streamline datacenter operations. Over the past 12–18 months, however, IDC is seeing the use of AFAs in more and more cloud infrastructures to handle performance-sensitive workloads and enable the creation of premium service tiers that provide additional incentives for administrators to look at moving workloads to the cloud.

When considering the use of AFAs in cloud-based infrastructures, economic, maturity, data services functionality, and degraded mode performance considerations are key purchase criteria. Relative to hyperconverged platforms (another architecture used in cloud infrastructures), these characteristics are where AFAs differentiate themselves. Not all AFAs can meet these requirements, but vendors like Pure Storage marry flash performance with a compelling cost and functionality proposition that makes the use of their enterprise storage platforms in cloud infrastructures a rapidly growing business for them. In cloud environments hosting performance-sensitive workloads, flash is an absolute requirement to deliver needed performance while enabling optimum infrastructure density — the question for organizations building cloud-based infrastructures is how to integrate flash in a manner that best meets their own requirements as well as the requirements of their customers.

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