Compelling Flash Economics Driving Strong Flash Deployments in the Enterprise

Sponsored by: HDS
Eric Burgener
May 2014

IDC OPINION

IDC believes that because spinning disk alone cannot cost effectively meet the storage performance requirements of the 3rd Platform of computing, flash storage should be in use in every datacenter in at least some capacity. A critical look at the comprehensive nature of flash economics should make clear why IDC believes this. With the intelligent use of flash, far fewer devices are needed to meet performance requirements, significantly reducing power, cooling, and floor space costs over the useful life of the storage solution. These smaller primary storage configurations require smaller backup infrastructures, resulting in additional cost savings. When enterprises deploy with flash-optimized architectures, they are well positioned to maximize the performance, reliability, and economic benefits that flash storage brings to the table. If enterprises aren't considering how they can integrate flash into their storage infrastructure, they should be.

IN THIS WHITE PAPER

This white paper focuses on highlighting the economic benefits of flash storage deployment for enterprises of all sizes, highlighting optimal approaches for flash utilization that administrators need to consider as they look to maximize their ability to meet 3rd Platform computing requirements most cost effectively. It also considers the types of application workloads where flash is most often used and why. And it provides a brief review of the HDS enterprise flash strategy and flash product offerings.

SITUATION OVERVIEW

The information technology (IT) industry is in the midst of a migration to a new, much more agile computing infrastructure that IDC refers to as the 3rd Platform. Driven by the needs of an increasingly mobile workforce, social media, big data and analytics, and cloud computing, this new platform is built on top of virtual infrastructure, offers significant flexibility to adapt to changing business requirements, and provides the scalability to accommodate explosively high data growth. According to IDC, over the next five years, more than 95% of industry growth will be driven by this 3rd Platform buildout.

May 2014, IDC #248570
The emergence of the 3rd Platform has driven the need for new compute and storage architectures. The traditional spinning disk (hard disk drive [HDD]) has been challenged to cost effectively meet the performance and latency requirements in virtual computing environments, particularly as workload densities per storage solution have increased. Facilities considerations such as floor space and energy consumption are also driving enterprises to rethink how storage requirements are met as administrators struggle with continuing to meet service-level agreements (SLAs) while managing high data growth.

Flash offers significant benefits to administrators facing this new environment, so much so that IDC recommends that all enterprises should be using at least some flash in their datacenters. Flash devices can offer up to 1,000 times the performance of HDDs while consuming only a fraction of the power. Storage performance requirements (input/output operations per second [IOPS]) can be met with far fewer storage devices, making very efficient use of available floor space. And flash makes the use of attractive storage functionality much more viable in high-performance environments, providing additional administrative flexibility and the opportunity for even greater cost savings.

**Objectives of Flash Deployment**

Flash is primarily deployed for performance reasons, but that is not the only reason. In IDC's most recent *Storage Users' Demand Study* (November 2013), respondents across industries identified fast start-up time and fast random access time as the top 2 reasons for flash deployment, although a number of other benefits were listed as well (see Figure 1). Respondents also called out lower power consumption and reduced floor space requirements, resulting in reduced energy costs and more efficient utilization of limited datacenter space. It is telling that 59% of survey respondents saw a reduction in the overall capacity of storage purchases as a result of flash deployment, resulting in smaller storage infrastructures that were more cost effective.
FIGURE 1

Benefits of Flash Deployment

Q. What are the primary benefits of deploying solid state storage in your organization? Please select the top benefits.

Historically, enterprises have deployed flash for use with a specific, critical, performance-sensitive application, such as Oracle, SAP, or Microsoft Exchange. As vendors have increased the ability of their storage solutions to host mixed-workload environments, however, we have seen much broader use of flash across a wider variety of applications. The ability to achieve high workload densities on a given storage platform while still meeting SLAs for performance, availability, and recovery is driving down the effective cost of flash deployments even further. Because of the attractive total cost of ownership (TCO) of these denser deployments, we expect this practice to become even more pervasive in the future.

The Importance of Flash-Optimized Storage Architectures

Flash can be deployed in the server, in an appliance, or in an array. Server-based flash can deliver the lowest latencies, but it is limited in capacity and poses challenges in environments requiring high availability. RAID can be used to protect data from a device failure, although this limits the usable capacity even further and does not protect against server failures. To provide that protection, data
must be synchronously written to a secondary location outside of the server, an operation that introduces network latency and effectively negates the latency advantages of server-based flash.

In 2013, nearly 61% of flash was deployed into preexisting storage arrays, with the remaining 39% split between entirely new arrays and/or appliances and host-based PCIe flash. Just over 25% of flash capacity was deployed into newer arrays that were specifically designed for flash. The early predominance of flash deployments in existing arrays was driven primarily by ease of integration — it was the easiest and least expensive way to try out flash technology in the datacenter, helping more cost effectively meet performance requirements while extending the usable life of legacy equipment. In the past year, we’ve seen flash deployments in existing arrays drop to 37%, while flash capacity in newer “flash optimized” arrays has increased to 37%.

Arrays that have been designed with flash in mind can easily deliver 10 times more performance out of the same flash device deployed in an array that treats it in exactly the same way as a spinning disk does. These “flash optimized” storage architectures recognize when flash is in use and take steps to optimize its use. Optimized write minimization, wear-leveling, and garbage collection algorithms improve flash media endurance and help ensure consistently predictable performance in the sub-millisecond range, regardless of workload. Storage efficiency technologies like thin provisioning, thin snapshots and clones, and optimized replication are combined with data reduction features to make the most of flash capacities, although we will pass a point in the next year where the single largest solid state disk (SSD) will actually have more capacity than the single largest HDD. As administrators become more aware of the definition and impact of flash-optimized architectures, we expect to see storage purchases skew heavily in this direction.

Leveraging flash in enterprise environments with high write activity poses challenges. Compared with HDDs, flash technology supports a relatively limited number of program/erase (P/E) cycles (i.e., writes). Initial enterprise deployments leveraged a NAND flash technology called single-level cell (SLC). SLC provided better write endurance than multilevel cell (MLC) flash technology but was quite a bit more expensive. The early extensive use of SLC contributed to relatively high price points for flash storage. As vendors introduced techniques to actively manage writes to minimize the number of P/E cycles, along with wear-leveling algorithms that eliminated write hotspots on an SSD, it became possible to introduce MLC-based flash devices whose useful product life exceeded that of the typical enterprise storage array (i.e., five years). The more extensive use of MLC flash is one factor, although not the only one, that has helped bring down flash storage costs significantly.

Write activity against a flash device requires a periodic recycling to remove stale data and free up space for new writes. This is called “garbage collection,” and in early flash designs, real-time I/O performance was often impacted during this process. Enterprise workloads require consistently predictable performance, and newer designs that separate the control and data planes in flash controllers have resolved this issue by splitting the management of real-time I/O and garbage collection across physically separate processors. Flash-optimized architectures are designed to manage garbage collection so as to deliver consistently predictable write performance.

While flash optimization is an important contributor to efficient flash deployment, the ability to consolidate multiple workloads onto a single storage array while still meeting SLAs is arguably even more important. Arrays that can scale to the petabyte range in terms of raw storage capacity, host both block- and file-based storage, support hybrid storage configurations that include both flash and HDD (along with the
caching and/or tiering software necessary to put performance exactly where it is needed), provide scalable and reliable enterprise-class data services (i.e., snapshots, clones, replication), and provide good mixed-protocol support (i.e., SAN/NAS) offer excellent opportunities to achieve high workload densities. As applications are consolidated onto this single platform, older separate storage devices can be retired, lowering energy use and freeing up floor space. And management becomes easier as more of the datacenter workload can be managed from a single user interface.

Shipments of storage arrays with flash-optimized storage architectures are expected to increase at a rapid rate. IDC expects that by 2017, 45% of new flash capacity will be going into flash-optimized products and that ultimately these products will become the enterprise storage workhorses in virtual datacenters worldwide.

The Economic Benefits of Flash Deployment

Enterprises that think flash is still too expensive to deploy have not really looked at the true economics of flash deployment. In IDC’s opinion, the storage performance requirements of virtual infrastructure cannot be cost effectively met with HDD alone, and IDC is advising enterprises of all sizes that at least some flash should already be deployed in their datacenters. The economic benefits of flash deployment include higher performance, lower power consumption and less floor space, and a much broader ability to leverage performant data services (provisioning, data protection [i.e., RAID], snapshots, clones, replication, etc.) to make administration easier. As we discuss these issues in more detail, we assume that flash has been deployed in a flash-optimized storage architecture.

Performance

Because of the far higher IOPS and much lower latencies, it takes far fewer flash devices to meet performance requirements. If you require 40,000 IOPS and 60TB, it would take 200 performance-optimized HDDs (FC, 15,000rpm, 300GB) to meet that requirement, and expected latencies would be in the several millisecond range. Notice that to meet our IOPS requirement, we may be forced to purchase storage capacity that we do not need. A single flash device can meet that IOPS requirement (assuming it is deployed in a flash-optimized storage architecture), and it will consistently deliver response times in the half-a-millisecond range.

Sizing storage configurations is not just about performance, though – both availability and capacity considerations must be taken into account. Most enterprise applications would not deploy in production with a single flash device because that would introduce a potential single point of failure that could put data at risk. A data protection scheme of some kind will likely be implemented (RAID, erasure coding, replicas), with the selected scheme driving some additional device deployment. For the purposes of our example, we’ll assume RAID 5. In our 40,000 IOPS example, a customer may purchase a couple of additional drives just for data protection purposes.

Capacity must also be taken into account. While solid state disks in the terabyte range are quite common today – SanDisk just introduced a 4TB SSD and plans to introduce an 8TB version next year – the largest HDDs today boast 6TB, and because of technological limitations, that number is not expected to increase much at the device level in the coming years. Larger-capacity HDDs tend to deliver fewer IOPS than smaller-capacity HDDs in random workload environments, so virtual infrastructure customers interested in performance will tend to buy smaller-capacity HDDs.
SSDs deliver the same performance, regardless of size, encouraging the purchase of larger-capacity drives when capacity is a concern. If we assume the use of 1TB SSDs in our example, we would need to purchase 60 SSDs (assuming an all-flash configuration).

Hybrid architectures that allow different classes of storage to be simultaneously deployed to create tiered storage configurations may help actually lower the overall acquisition cost. In our example, meeting the performance (40,000 IOPS) and availability (transparent recovery from a single drive failure) requirements may require 4 SSDs. Since those devices deliver more performance than we actually need, we can now afford to buy capacity-optimized HDDs to meet the capacity requirement. Capacity-optimized drives can be larger in size (we don’t care what their performance is since we have flash in the system) – so we do not have to buy very many of them to meet our capacity requirement. If we assume 4TB HDDs, we would have to purchase 15 of them to provide 60TB of capacity. Caching and/or tiering software will ensure that we get flash performance for most I/O, while the HDDs will ensure we have sufficient capacity to meet our storage requirements.

In this example, the hybrid array meets our performance, availability, and capacity requirements with 19 devices, whereas the HDD-only configuration requires 200 devices. Clearly, the flash-enabled configuration will cost significantly less than the HDD-only configuration.

There is one other performance consideration. In the past, administrators have sometimes been hesitant to use technologies like thin provisioned storage and data deduplication because they can introduce unacceptable performance impacts. Even if a storage vendor provides such capabilities in its array, they may not always be used in production environments because of their performance impact. Flash latencies completely overcome this problem, enabling the use of these technologies to reduce the effective dollar per gigabyte of flash deployments even more by making more optimal use of available flash capacity.

Energy and Floor Space Consumption

Each SSD uses roughly half the wattage of a performance-optimized HDD. The real power and cooling savings in deploying flash, however, result primarily from the far lower number of devices that must be deployed to meet the storage requirement.

In our example, a datacenter will have to house, power, and cool 200 HDDs to deliver 40,000 IOPS and 60TB, whereas it will have to house, power, and cool only 19 devices in the hybrid configuration. And interestingly, the hybrid configuration actually delivers significantly more performance – in terms of both IOPS and latencies – and slightly more capacity than the HDD-only configuration.

Historically, storage overprovisioning has been rampant. By the time administrators meet the IOPS requirement, they have had to buy so many HDDs that they have significant capacity that they do not need. Since all HDDs in one of these configurations are needed to deliver the IOPS, all disks are spinning all the time. The datacenter is housing, powering, and cooling significant capacity that it does not need, effectively increasing power costs for no good reason.
Notice that if our example required 80,000 IOPS but still only 60TB, we would have to buy 400 HDDs to meet the IOPS requirement and end up with an additional 60TB of storage capacity (assuming 300GB drives) that we do not need. With flash as an option, we can meet the 80,000 IOPS requirement with just the 4 SSDs we already have configured, and we do not have to buy any additional devices. Because of their extreme performance capabilities, very few SSDs are needed to meet any given IOPS requirement, and we don’t end up with unneeded capacity that we are forced to house, power, and cool.

**Storage Consolidation**

When you add additional application workloads to a given array, you are generally increasing the requirements for both performance and capacity. Hybrid architectures that support both flash and HDD let you add performance when you need it very efficiently with flash or capacity when you need it very efficiently with the capacity-optimized HDD. In other words, it is very cost effective to add the performance and capacity necessary to host multiple workloads to hybrid configurations compared with HDD-only configurations.

A scalable enterprise-class hybrid array is an excellent platform for storage consolidation because it can cost effectively accommodate both high-performance and high-capacity workloads as needed, ensuring that each application gets what it needs when it needs it, regardless of what other workloads are running on the array. As more workloads are consolidated onto that array, its TCO is spread across more applications. Other equipment that is less power or floor space efficient can potentially be retired. More applications are managed under fewer management interfaces. And there is less wasted storage capacity since this flash-optimized platform provides more granular control to dial in exactly how much performance and capacity are needed – independent of each other. Less primary storage capacity means there is less storage to back up, potentially shortening backup windows and requiring less backup infrastructure.

SSD capacity is still more expensive on a dollar-per-gigabyte basis than HDD capacity, but flash is continuing to come down in price. Flash-optimized architectures that make better use of flash capabilities help lower the effective dollar per gigabyte further, and dense workload consolidation can take that to the point where the average dollar per gigabyte of hybrid configurations can be very close to the effective dollar per gigabyte of HDD-only configurations sized to meet specific performance requirements. Scalable, flash-optimized, enterprise-class arrays that can deliver on the performance, availability, and reliability requirements demanded by the datacenter and include controls to maintain predictable performance at the application level even as configurations scale are the best option today for leveraging workload consolidation to optimize TCO for flash-based storage.

**Total Cost of Ownership**

Although many enterprises are still primarily using dollar per gigabyte to compare the cost of flash-enabled storage with the cost of traditional storage, IDC believes this is outdated. Given flash-optimized architectures, flash-enabled storage that can support a petabyte or more of raw capacity, and solutions that can deliver consistent performance even as configurations scale, combined with the evolving requirements of the 3rd Platform, dollar per gigabyte is no longer the most relevant metric. Metrics like dollar per IOPS, IOPS per terabyte, IOPS per watt, dollar per transaction, and TCO are more accurate ways to quantify the value that this new class of storage provides.
With flash enabling far smaller storage configurations that draw significantly less power and require far less floor space and better facilitate the use of storage efficiency technologies like thin provisioning and deduplication; write minimization and wear-leveling algorithms that enable the use of less expensive MLC flash in large, mission-critical applications; and quality-of-service (QoS) capabilities that allow administrators to increase workload densities without risking performance impacts, TCO is a much better measure of storage array value to the business.

The HDS Accelerated Flash Product Portfolio

HDS is a proven enterprise storage provider with thousands of customers worldwide across all industries. With a large installed base and a strong reputation for reliability, availability, and scalability with its enterprise storage platforms, HDS has historically chosen to take an evolutionary approach to integrating new storage technologies into its portfolio that preserves customer investment. This contrasts sharply with the approach that many point-product start-ups with no installed base have chosen.

The HDS storage architecture is designed to allow for expansion and new device integration, as well as firmware updates and hardware component replacements, to be performed nondisruptively. This allows HDS customers to take advantage of newer storage technologies as they become available without disrupting their environment while continuing to benefit from the advantages of HDS storage. In today's increasingly agile world, downtime is not an option for most enterprises whose business depends on the availability of services every minute of every day. As new technologies are integrated, HDS customers continue to enjoy the performance, scalability, and availability of a proven enterprise-class platform while managing those technologies from a familiar user interface.

The HDS storage product line encompasses unified (block/file/object) enterprise storage platforms for both open and mainframe environments, spanning primary and secondary storage use cases from high-performance databases, enterprise applications, and Web/file serving to content archives and other cold storage needs. A software-defined storage layer, managed by Hitachi Command Suite, allows new technologies to be easily integrated and managed under the familiar Command Suite user interface.

HDS supported the integration of flash drives into its existing storage platforms early on, giving customers easy early access to this technology. With the introduction of its Flash Module Drive (FMD) technology, HDS provided a flash-optimized storage architecture that could be easily added to existing installations to deliver the full performance benefit of flash technology. FMDs can be used in any of the HDS primary storage platforms.

Designed specifically with flash in mind and incorporating a variety of flash-specific features, the FMDs meet IDC's definition of a "flash optimized" storage architecture. FMDs include write minimization implemented through zero block write avoidance and write coalescing, separate control and data paths to handle garbage collection without impacting real-time I/O performance to ensure predictable write latencies, and wear-leveling, digital signal processing, and background data sweeping to ensure long device life and data integrity. These flash optimization techniques allow HDS to use lower-cost MLC flash yet still provide useful lives that exceed the standard five-year depreciation cycle.
HDS supports both all-flash and hybrid flash configurations with autotiering software that manages data placement for optimal performance (in hybrid configurations) on an ongoing basis. Data is migrated across tiers according to certain parameters established up front by the administrator. LUN pinning is supported as well to help ensure flash performance is available at all times for selected application environments – with the perfect example being the boot disk in a virtual desktop infrastructure (VDI) installation.

HDS also supports a full complement of enterprise-class data services, including dynamic provisioning, data protection (RAID), snapshots (ShadowImage), clones (Thin Image), and replication (TrueCopy, Replication Manager), to help ease administration. Postprocess data reduction helps make optimal use of available flash capacity. All of these capabilities are managed from a single point of unified management, an increasingly important issue for operators whose administrative span of control is being expanded by budget-conscious enterprises.

HDS has earned its reputation for performance and scalability across thousands of installations. The newest HDS storage platform, the Virtual Storage Platform (VSP) G1000, supports 4 times the internal bandwidth, 3 times the IOPS, 2 times the cache capacity, 4 times the maximum cores, and 10% better power efficiency than the previous high-end VSP. All-flash configurations can now support up to 600TB of raw flash capacity, and hybrid configurations can exceed several petabytes. FMDs have 4 SAS lanes coming out of them and dedicated SAS interfaces for each FMD controller to provide maximum bandwidth to/from flash devices. Quality-of-service controls ensure that performance can be assigned and predictably delivered, no matter what other workloads are running on the platform. High levels of connectivity and scalability, combined with strong multiprotocol support (i.e., SAN and NAS) and the inherent performance capabilities of the architecture, make this an ideal platform for dense workload consolidation.

**FUTURE OUTLOOK**

With the storage performance demands of the 3rd Platform, flash should be a part of every datacenter. Although flash offers significantly better performance than spinning disk, vendors that provide flash-optimized storage architectures will deliver the best economic benefits for flash deployment. The potent combination of high performance, lower latency, far fewer storage devices, and reduced power, cooling, and floor space requirements will make flash the centerpiece in the enterprise storage workhorses of the future.

Flash vendors, regardless of whether they offer all-flash or hybrid flash storage platforms, will likely look to increase the density of workloads their products support with features like increased capacity, multiprotocol support, high connectivity, and QoS controls that completely resolve the “noisy neighbor” problem in mixed-workload environments. These types of features will drive the overall TCO of flash-based storage solutions ever lower, making them that much more attractive relative to pure spinning disk-based configurations.
**CHALLENGES/OPPORTUNITIES**

Flash-based arrays are not the only option customers have to integrate flash technologies. Hyperconverged architectures that combine compute and storage into x86-based nodes can leverage host-based flash, in combination with spinning disk, to create potentially large-capacity storage solutions that are centrally managed. For certain application environments, enterprises are already starting to deploy them. These types of platforms are still relatively new, however, and have yet to prove their ability to reliably host large multiworkload environments while predictably meeting SLAs. How quickly these types of platforms ride up the reliability curve could impact flash array-based growth rates.

Similarly, early market entrants into the all-flash array space usually produced platforms with limited capacities that were generally targeted for use with a single, demanding application. These stand in stark contrast to storage platforms that can scale to a petabyte and beyond and will likely remain niche players until they can provide much larger storage capacities. That alone will not be sufficient, however, to make them viable candidates for the enterprise storage workhorse of the future. As mentioned previously, that requires not only a flash-optimized architecture but also storage efficiency features, enterprise-class data services, broad multiprotocol support, and QoS controls that ensure that applications get the performance they need, even as workload densities increase.

For entrenched storage vendors, an evolutionary approach to implementing newer technologies in a nondisruptive manner will allow their customers to preserve existing investments even as they move forward. When these platforms are flash optimized and offer the necessary features to support dense workload consolidation, they can offer very aggressive TCOs for medium-sized to large enterprise environments. Vendors with a proven reputation for performance, scalability, reliability, and availability can leverage their "trusted vendor status" to compete effectively against start-ups that are still trying to prove themselves.

**CONCLUSION**

Flash is destined to become a key component of enterprise storage in the datacenter. Already, over half of all enterprises have deployed flash in some capacity, with the remaining 49% of enterprises planning integration of flash within the next 12 months. It is clear to those that understand the technical challenges posed by the 3rd Platform that spinning disks alone will not be able to cost effectively meet storage performance requirements. That is why IDC is already recommending that all datacenters should be using flash storage in at least some form today.

Enterprises that aren't deploying flash because they think it is too expensive may not understand the full panoply of economic benefits flash brings to the table. At a device level, flash delivers 1,000 times the IOPS at sub-millisecond latencies and uses roughly half the power of HDDs. When flash is part of the storage solution, far fewer devices can be deployed to meet any given performance and capacity requirements. Significantly less datacenter floor space is required. And the much smaller primary storage infrastructures can reduce backup windows and lower backup infrastructure costs as well.
The use of flash-optimized storage architectures maximizes not only the performance, endurance, and reliability but also the effective economic benefits of flash-based storage. To ease the transition to flash-based storage, end users may choose to integrate SSDs into their existing storage arrays to help boost performance and extend the useful life of existing equipment. But when enterprises are considering new storage array purchases, they should be considering flash-optimized offerings. If they are looking to get the most out of flash technologies, they should consider arrays that scale to the petabyte range and beyond and also support storage efficiency features such as thin provisioning and data reduction, enterprise-class data services, multiple storage protocols, and QoS controls that guarantee performance at an application level even as configurations scale.

Finally, when enterprises are looking for a new enterprise storage workhorse, reliability and availability features are critical. Vendors with an established reputation as an enterprise storage supplier have an advantage they can leverage against newer players. Hosting mission-critical applications requires high-availability features to not only protect data locally but also provide the foundation for solid disaster recovery and business continuity strategies.

When all these considerations are taken into account, dollar per gigabyte is no longer a relevant metric to gauge the value an array brings to the business. Other metrics such as dollar per IOPS, IOPS per terabyte, IOPS per watt, dollar per transaction, and particularly TCO much better reflect the true value, especially if that platform has the capabilities to enable dense workload consolidation. It is not possible to effectively gauge the value of a flash-based storage platform without also taking into account the need for far fewer devices, along with the power, cooling, and floor space savings over the expected life. Those significant savings are not reflected in acquisition costs alone.

Flash economics inevitably win out against HDD-only configurations when it comes to supporting the 3rd Platform buildout. That is why IDC believes that flash should be in use in every datacenter today.
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-insights-community.com
www.idc.com

Copyright Notice

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2014 IDC. Reproduction without written permission is completely forbidden.