WHITE PAPER

Building Data Centers for Today's Data-Driven Economy: The Role of Flash

Sponsored by: SanDisk

Richard L. Villars Eric Burgener

July 2014

EXECUTIVE SUMMARY

The world of IT is in the midst of a massive structural shift from the PC and client/server-based "2nd Platform" that dominated the last 25 years, to what IDC calls the "3rd Platform," which is built on a foundation of mobile computing, social media, big data and analytics, and cloud technologies. Given that the 3rd Platform will dominate IT investments in the next decade, the data center that supports the 3rd Platform's enabling technologies is the foundation for new business models.

In fact, the data center is the most important point of contact with customers, defining the customer experience. Business leaders in financial services, manufacturing, retail, and healthcare expect their data center operators (internal or external) to reliably and dynamically deliver compute, data, and network capacity, on time — with no delays, and no excuses.

When IT asset provisioning is measured in minutes, IT teams must aggressively exploit technologies such as flash that play a key and multifaceted role in higher-performance, more agile, and more efficient data centers. Flash-enabled solid state drives (SSDs) deliver better latency and throughput in many use cases. They consume roughly one-half of the power on a "per device" basis compared to hard disk drives (HDDs) and much less than that on a "per storage solution" basis. SSDs also enable more scalable utilization of Tier 1 data services like compression, deduplication, and thin provisioning that are at the core of enterprise-class systems.

Intelligent use of flash in both servers and storage systems dramatically boosts performance, while also reducing the cost of instance or resource deployment and change. While performance is the key driver, there are a number of secondary economic benefits that dominate total cost of ownership (TCO) as SSD capacities grow. These include the following: far fewer devices required to hit performance requirements, reduced energy and floor space consumption, and lower software licensing costs. In support of the 3rd Platform build-out, IDC expects that spending on data center flash, taking into account server (internal) flash, along with all-flash storage arrays and hybrid storage arrays (combining HDDs and SSDs), will reach $10.91B by 2018.

Organizations should prioritize their use of flash-based solutions to gain major benefits for a range of applications, including:

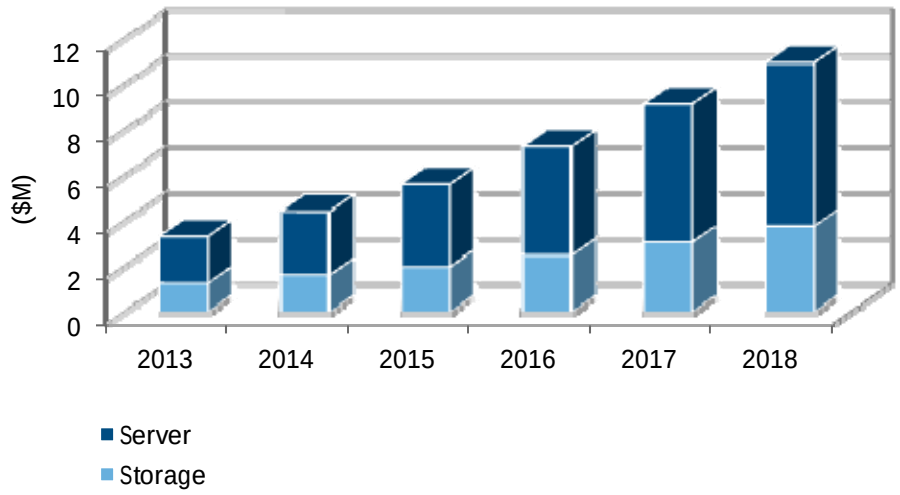
- Mission-critical applications such as transaction-oriented databases
- Compute-intensive workloads for big data/analytics and high-performance computing (HPC)
- Serving of virtual machine (VM) instance images for both servers and desktops (i.e. virtual desktop infrastructure (VDI))
- High-quality, high-throughput, and large-scale delivery of entertainment and medical images, via media-streaming
- Active archives for backup/disaster recovery, regulatory compliance, and ongoing research

SanDisk Corporation, a leading supplier of flash technologies with over 25 years of industry experience, is driving the optimal use of NAND flash technology in storage devices. Its products are designed to deliver high endurance and provide predictable performance as configurations scale, delivering the reliability that enterprise-class solutions require. SanDisk's solutions include sophisticated flash management techniques that recognize and treat flash differently from mechanically-driven HDDs. Innovations in the areas of write minimization, wear-leveling, garbage collection, error detection and correction enable SanDisk's server and storage partners to deliver enterprise system flash-based products.

Figure 1 shows the enterprise SSD revenue forecast for servers, which includes flash attached directly to the memory bus, added in the form of PCIe cards, or integrated as 3 or more SSDs attached to SAS or SATA interfaces, and storage, which includes flash in AFAs and HFAs (regardless of whether it is used as a cache or a tier).

FIGURE 1

Worldwide Enterprise SSD Revenue by Location, 2013-2018



Source: IDC, 2014

NEW BUSINESS CHALLENGES RESHAPE THE MODERN DATA CENTER

From 1985 to 2010, the world of IT was dominated by the PC and client/server-based "2nd Platform" (see Figure 2). Today, IT leaders are extending their data centers into the "3rd Platform" era — an IT environment built on the foundations of mobile computing, social media, big data analytics, and cloud technologies. The three characteristics that define a 3rd Platform-ready data center are:

- **Scale.** Supporting up to 10-fold increases in supported users and/or data sets without comparable increases in data center footprint
- **Speed.** Creating and updating applications and services in weeks/days, not years/months, without increasing IT operations and development staff levels
- **Scope.** Enabling coordination of multiple applications and data sources, internal and external, to deliver new services to customers without sacrificing data integrity and user experience

Business innovation, today, is based on leveraging the 3rd Platform to gain new functionality. IT organizations are extending the value of tens of thousands of existing 2nd platform applications with mobile and analytic add-ons. They are also creating millions of high-value, transforming solutions and services that are based exclusively on this new platform in industries such as financial services, manufacturing, retail, and healthcare.

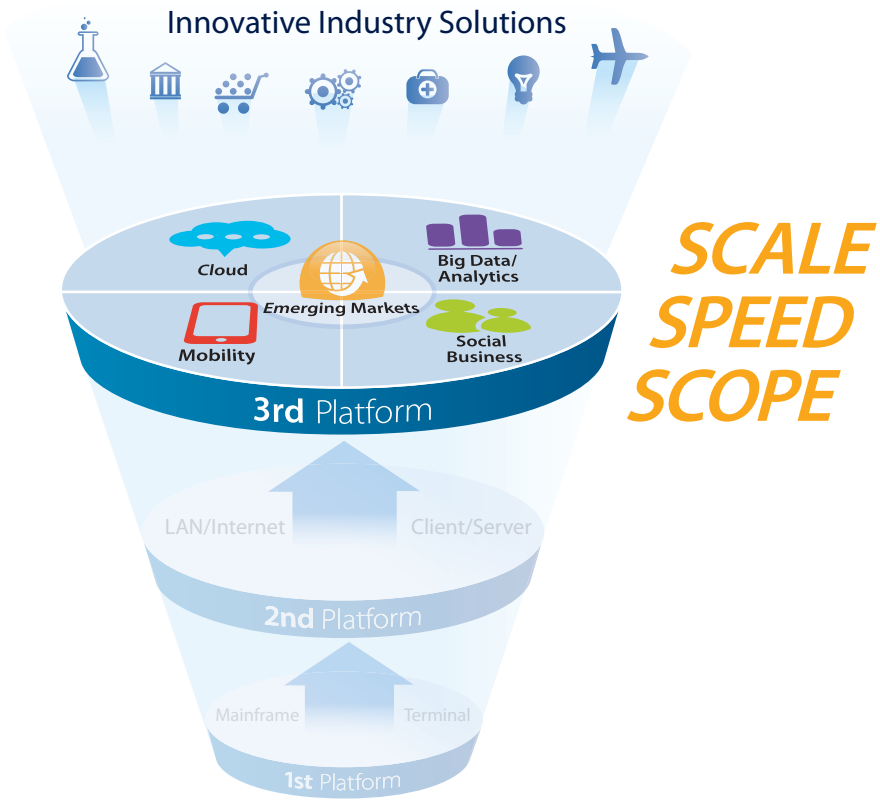
In this new world, where businesses rely upon mobile computing, social media, big data analytics, and cloud to deliver a superior customer experience, the data center can no longer just be the place where an organization keeps its computers. Rather, the data center organization must play the leading role in addressing key business challenges, including:

- **Global expansion:** Businesses have to go where the new customers and emerging, high growth markets are, whether they are in the Americas, EMEA, or Asia/Pacific. To avoid latency caused by network transmission time or to comply with regulations controlling the ways and places where data can be stored, businesses need data centers close to their customers.
- **New service creation/expansion:** Organizations in many industries face an explosion in new 3rd Platform-based services and new applications (e.g., mobile, social, big data and analytics) that take enterprise workloads to the next level, driving new business solutions.
- **Data center obsolescence:** Many companies realize that the existing approach to running data centers isn't keeping pace with changing expectations for better performance and reduced operating costs; they need to change their data center operating model.

Technologies like flash that directly address scale, speed and scope will be greatly valued by data center managers. These enabling technologies will demonstrate that value by dramatically boosting performance, improving reliability and reducing per unit costs for energy and space, taking on a key role in shaping the modern data center.

FIGURE 2

The IT Industry's 3rd Platform for Growth and Innovation



Source: IDC, 2014

Data Center Workloads: Changing the IT Conversation

In this rapidly changing era of the 3rd Platform data center, it's no longer sufficient for IT to talk about workloads solely at the level of individual server and storage systems. Now, IT has to think about the workloads at the level of the data centers themselves.

What are the primary data center workloads?

A key data center workload – the one that is most familiar to most IT organizations – is transactional, involving changes and queries to systems of record that are usually measured in terms of Input/output Operations Per Second (IOPS). In today's organizations, however, three new workloads are driving considerable data center expansion:

- **Ingesting/serving content (e.g., video, audio, and images).** Retailers are ingesting video surveillance streams, and manufacturers are collecting sensor data. Media companies are serving movies to consumers. Healthcare organizations are delivering critical medical records to caregivers. The timely delivery of content (usually measured in terms of throughput) with scalable systems of engagement affects customer satisfaction and the quality of patient care.

- **Analyzing large data sets in real time to drive timely business decisions.** Retailers are responding more quickly to changing consumer purchasing behaviors. Financial services companies are using automated, real-time processing to detect credit-card fraud and money laundering. Atmospheric researchers are improving the timeliness and accuracy of weather forecasts with more flexible high performance computing (HPC) systems. Major boosts in performance for systems of analysis have material impacts on business outcomes.
- **Organizing/preserving information that needs to be archived for a long period.** Research hospitals are preserving genetic sequencing for diagnostic/research purposes. Aerospace companies are preserving flight data and component failure data to improve overall system reliability, Retailers of all kinds are capturing sensor data. Cost-effective storage and fast retrieval of data from systems of data preservation are the keys to owning and monetizing extremely large, long term content depots.

Transitioning to the 3rd Platform — Moving Beyond Transactions Alone

If one looks at workloads in terms of the dollars, time, and staff resources, most organizations believe that their data centers are built, first and foremost, for transactions. This perception is a remnant of the era of the 2nd Platform. IDC's surveys and interviews with leading IT architects make clear, however, that organizations' data centers are actually handling a range of content serving, big data and analytics, and archiving functions as well. These systems of engagement, analysis, and preservation are growing at a much faster rate, consuming more data center space and resources.

In this new IT landscape, IT teams must focus on deploying significant IT assets that are optimized to deliver the performance required across each of these very different application workloads. Flash is one of the key technologies that play a critical role across all of these data center workload environments.

IDC's surveys and interviews with leading IT architects make clear, however, that organizations' data centers are actually handling a range of content serving, big data and analytics, and archiving functions as well. These systems of engagement, analysis, and preservation are growing at a much faster rate, consuming more data center space and resources.

HOW FLASH IS TRANSFORMING DATA CENTERS

With the 3rd computing platform build-out, storage performance requirements are changing significantly. The server consolidation driven by virtual computing presents mixed workloads to storage systems that demand not only more throughput and lower latencies, but very predictable performance. The applications in this build-out – high performance enterprise applications and databases, media-streaming applications that require high sequential performance, VDI, and dense mixed virtual workloads that leverage both structured and unstructured data – need consistently high performance and introduce extremely high I/O spikes. HDD-based configurations are having significant difficulty meeting these requirements cost effectively.

SSDs provide an excellent solution to this problem. Able to consistently deliver at least an order of magnitude better performance than HDDs, SSDs offer a number of benefits to data centers looking to deploy the most cost-effective storage solutions for their combination of 3rd and 2nd platform computing environments.

Impacting Operational Costs

Flash requires far fewer devices (in the form of SSDs) to meet performance requirements and consumes about half the power that HDDs do on a per-device basis.

Data centers that had been adding significant numbers of HDD spindles to meet performance requirements will find that in many cases they need 50% to 80% fewer SSDs to meet performance requirements, resulting in potentially significant floor space savings. The lower device count associated with SSDs reduces power consumption requirements even more, compared with HDDs. Far fewer servers are needed to drive storage performance when flash is in the mix, resulting in license cost savings. These secondary economic benefits of flash deployment really start to kick in as data centers increase flash capacities, driving a "tipping point" after which HDD deployments for primary storage will become more difficult to cost-justify.

Data centers that had been adding significant numbers of HDD spindles to meet performance requirements will find that in many cases they need 50% to 80% fewer SSDs to meet performance requirements, resulting in potentially significant floor space savings.

Flash is Permeating the Data Center

Flash is permeating the data center in a number of different ways. In servers, flash can be attached directly to the memory bus, added in the form of PCIe cards, or integrated as SSDs attached to SAS or SATA interfaces. Server-side flash typically delivers the lowest latencies because there is no network hop – and is a great fit for application environments like big data and analytics that require extremely high performance but no high availability/failover. It is appearing in standalone servers, in blade servers, and in converged and hyper-converged servers as an efficient way to boost performance. Many of the new converged and hyper-converged solutions that leverage scale-out architectures leverage flash to cost-effectively deliver the extremely high performance that the real-time analytics applications for which these platforms are used require.

In the network, flash can appear in caching appliances or in shared storage arrays. All Flash Arrays (AFAs) use only flash – and are generally targeted at application environments that require consistently predictable sub-millisecond response times as well as high availability. Hybrid Flash Arrays (HFAs) use a mix of SSDs and HDDs to deliver much better performance than traditional HDD arrays, but they do so at an acquisition cost lower than AFAs at a particular capacity point. In the long run, as flash-based arrays become more capable of effectively handling mixed workloads while delivering predictable performance across large deployed capacities, they will ultimately become the enterprise storage workhorses for primary storage. Even today, there are many data center environments in which flash-based arrays already offer compelling economic advantage relative to pure HDD-based solutions in terms of total cost of ownership (TCO).

The High Efficiencies of Flash Deployment

IDC has no doubt that the data center is evolving toward greater adoption of architectures for which flash is a key element in meeting growing performance requirements and reducing overall operating costs. The high efficiencies of flash deployment benefit administrators who are contending with data center footprint and local power grid constraints. Operational costs (opex) are impacted by flash-enabled systems. A smaller infrastructure requires less power and cooling, takes up less floor space, and is easier to manage than the larger infrastructures that solely use HDDs. IDC study data shows that fully 51% of data centers have deployed flash already as a way to improve I/O performance, and in the next 12 months the remaining 49% plan to consider it. IDC recommends that all enterprises should be deploying at least some flash in their data center, with the deployment method dependent upon their workload mix.

IDC study data shows that fully 51% of data centers have deployed flash already as a way to improve I/O performance, and in the next 12 months the remaining 49% plan to consider it.

Flash Benefits for Specific Workloads

As noted previously, data centers are handling an expanding set of new workloads that require support for a growing variety of application types. Flash-based solutions offer a number of benefits across a number of application types, including:

- Mission-critical applications such as transaction-oriented databases
- Compute-intensive workloads for big data/analytics and high-performance computing (HPC)
- Serving of VM instance images for both servers and desktops (i.e. VDI)
- High-quality, high-throughput, and large-scale delivery of entertainment and medical images
- Active archives for backup/disaster recovery, regulatory compliance, and ongoing research

Transaction-Oriented Databases

In enterprise database environments, throughput and latency are particularly important. Over the years, databases were optimized to increase performance in HDD-based environments with features such as transaction logging and write coalescing. Time-consuming complex storage administration techniques like "short stroking" and striping were used to improve performance, requiring very sophisticated administrators. Flash delivers performance very simply, providing a storage solution for database environments that is much easier to manage and offers much higher performance. For many enterprises, the ability to complete more transactions per second creates new business opportunities or drives higher revenue.

Database environments will no longer have to be significantly over-provisioned to provide performance — and energy and floor space costs will be reduced on a per-database basis. Database software licensing costs will also be reduced, because fewer servers will be needed to drive required storage performance levels.

Big Data and Analytics and High-Performance Computing

With the increasing importance of big data and analytics to business operations, scale-out architectures are becoming widely adopted in enterprise data centers. Instead of moving data to compute resources across networks, these architectures effectively bring compute to the data to increase performance. Because big data applications crunch through enormous amounts of data, very high storage performance is critical to performing these types of operations in near-real-time, for quick response to changing conditions.

It is common for hyper-converged architectures to include processors, memory, an SSD, and one or more HDDs — all housed within the same chassis or rack. The addition of flash, combined with intelligent caching and/or tiering software, allows these compact configurations to maintain extremely high performance for big data and analytics operations. Because of IT considerations for infrastructure power, cooling, and floor space requirements, flash-based storage configurations are also much more cost effective, resulting in significant TCO reductions over time relative to legacy storage architectures utilizing only HDDs.

High-performance computing (HPC) includes engineering and data-intensive workloads like seismic analysis, oil and gas exploration, weather forecasting, market simulations, and other big data and analytics applications. These HPC workloads often require high-volume processing of large data sets, which greatly benefit from the use of flash technology.

Agile and Efficient Virtual Computing

Virtual computing (virtualization) is the mainstream deployment infrastructure for the 3rd Platform, allowing much higher utilization of computing resources. Workload consolidation onto virtual infrastructure creates a storage challenge called the "I/O blender effect." Because of the way hypervisors work, the I/O patterns from virtual hosts are much more random than those from physical servers that were generally running only a single application. As VM density increases, this problem gets worse. When computing environments experience extremely random I/O patterns, latency for HDDs is increased on a "per I/O" basis, while reducing overall throughput by as much as 40% to 60%.

Given the IOPS level provided by flash-enabled systems, storage efficiency technologies like compression, data deduplication, thin provisioning, and space-efficient snapshots and clones can now be used without impacting application performance to further reduce raw storage capacity requirements for enterprise and cloud systems. Flash high capacity and density results in smaller infrastructure requirements, with all of the reduced power and cooling costs that come along with that. Those characteristics drive compelling economic reasons to use flash in virtual environments.

Virtual environments improve IT efficiency and business agility, allowing new VMs to be quickly spun up to meet increasing demand for data services. For certain businesses, the ability to bring up a new VM almost instantly is an important aspect of the customer experience. Online on-demand gaming and VDI environments are workloads that leverage flash to store templates or linked clones. This enables the creation of new VMs within seconds (rather than minutes), providing a meaningful performance differentiator for customers and/or end users.

VDI environments have other needs that require flash. VM densities are much greater on the servers that are hosting virtual desktops, exacerbating the "I/O blender effect" mentioned previously. Certain tasks, like booting desktops, logging in and logging out, produce I/O spikes, referred to as boot, login and logout storms, which can increase the IOPS requirements by a factor of 10 over the steady-state IOPS requirements of Windows desktops. Flash-based technology is the perfect medium to cost-effectively handle these types of widely varying I/O requirements, delivering performance that is equal to or better than physical desktops at a storage cost per desktop better than what HDD could provide. This is why, for many enterprises, VDI is the application environment where flash is first deployed.

Content- and Archive-Oriented Storage Applications

Newer, but very fast-growing, storage applications like content archives, backup/restore, and disaster recovery typically do not require the type of performance that transactional applications do, but they tend to require significantly more capacity. For that reason, $/GB has historically been an important consideration for these "secondary" storage use cases.

The largest-capacity SSD products on the market today support 4TB, whereas the largest HDDs support 6TB. For extremely high-capacity applications that require very little performance, HDDs may be a more cost-effective storage solution today, but within the next year or so, the capacity of available SSDs will outstrip the single-drive capacity of the largest HDDs.

IDC forecasts that by 2017, the raw $/GB of enterprise-class SSDs and HDDs will be very close. The reasons for this include: advances in flash technology, growing volumes, and competition in the marketplace that is reducing pricing differences with HDDs. With all of these trends in mind, we note that SSDs are becoming increasingly cost competitive with HDDs, even for workloads that have historically been considered to be capacity-intensive.

There are areas in secondary storage today for which high performance offers needed value. Many content archives will write a file or an object once, but will read it many times. SSDs excel at providing blazingly fast read performance for both random and sequential access patterns. Media archives, for example, that play movies or other video content back on-demand require some level of performance to deliver media-streams with consistent quality of service (QoS), and they are a good fit for flash. In backup and archive environments, flash-based metadata indexes can be used to speed search operations.

For many enterprises, recovery time objectives (RTOs) are measured in seconds or just a few minutes. Most recoveries, particularly file- or object-level recoveries, are done from the most recent backup, which represents a relatively small percentage of the required overall backup capacity. Flash is often used in purpose-built backup appliances to serve as a very high performance landing pad for backups. Maintaining the latest backup on flash storage in these appliances can also help to support extremely low RTOs (because the data can be restored from flash much faster than from disk).

SANDISK'S ENTERPRISE FLASH PORTFOLIO

SanDisk, with over 25 years of experience in the flash industry and over 5,000 related patents, is a leading flash storage vendor. SanDisk's close relationship with Toshiba, a flash component fabrication partner, gives SanDisk significant advantages in delivering compelling flash solutions to the market, including lower cost, strict quality control, and an ability to design features in at the die level that it can leverage upstream for better performance and longevity in its flash solutions.

SanDisk's enterprise flash portfolio consists of memory channel and PCI Express (PCIe) attached flash as well as SSD drives with SAS, SATA and memory channel interfaces; FlashSoft software; and innovative SanDisk foundation technologies that are leveraged across all enterprise flash products:

- **ULLtraDIMM SSDs** connect flash memory directly to the memory channel via standard DIMM slots, delivering ultra-low latency for the most demanding applications.
- **PCI Express Solid State Accelerators (SSAs)** offer SanDisk's proven enterprise SSD technology in an add-in card form factor with a PCIe interface that eliminates RAID controller delay, has a minimal impact on CPU utilization and reduces the need for additional DRAM.
- **The Optimus Family of SAS SSDs** provide both SLC- and MLC-based NAND options and are available in capacities ranging from 200GB to 4TB, offering high-performance solutions for demanding enterprise environments, I/O-intensive applications, and low-latency data access. The 4 TB Optimus drive, introduced earlier this year, is the highest capacity SAS SSD in the market, as of July 2014.
- **The CloudSpeed Family of SATA SSDs** provide cost-effective flash performance for a variety of read-intensive, mixed-use and write-intensive workloads. They are available in capacities ranging from 100GB-960GB.
- **ZetaScale software** enables extremely efficient access to flash for NoSQL and in-memory computing (IMC) applications, opening up opportunities to accelerate the performance of big data and real time analytics in enterprise data centers. A ZetaScale SDK is also available for qualified developers to provide this capability out-of-the-box with their applications.
- **FlashSoft software** provides optimized server-side data caching algorithms that are compatible with any type of flash-based storage to deliver maximum performance, utilization, efficiency, and value. SanDisk also provides the **Guardian** software technology, which manages flash cell wear to provide endurance on par with the depreciation life cycle of traditional enterprise storage, which is 5 years.
- **SanDisk Innovations** are foundation flash management technologies, applicable to all SanDisk hardware flash products, which improve endurance, performance, and capacity utilization.

SanDisk recognizes that optimal use of NAND flash technology requires sophisticated flash management techniques that treat flash differently from HDDs. SanDisk Innovations, including SanDisk's patented Guardian Technology, implement wear-leveling algorithms, bad block management techniques, error detection and correction, and write minimization to deliver flash products with high reliability, long service life, high performance, and excellent data integrity. These types of capabilities provide the foundation layer around which to build the most cost-effective flash-optimized storage systems architectures. SanDisk stands behind its SSD products with a five-year warranty.

Challenges for SanDisk

Although SanDisk has a long, successful history with flash technologies, the company is a relatively new entrant to the enterprise space. SanDisk's strategy to generate the same reputation for quality product in the enterprise space that it has in the consumer space is a good one, but reputations are not made overnight, particularly in the enterprise space.

SanDisk delivers flash-based storage devices that meet the functional requirements for flash-optimized storage in host, appliance, and array-based solutions, but SanDisk will need to be clear and succinct in its marketing of these products with the right business and technical benefits. It will also need to be forceful in efforts to get the industry to evaluate the economics of flash solutions based around more relevant metrics than $/GB, making it easier for customers to cost-justify their purchases with the finance and business units that approve IT purchases. Providing a holistic view of capital expense (CAPEX) and operational expense (OPEX) should help show the business benefits of flash adoption through significant reductions in power, cooling, floor space and software licensing requirements.

The density of flash storage products will be a key determinant in continuing to ratchet down the $/GB cost of flash over time, as will the density of mixed data center workloads that flash-based arrays will be able to successfully support. The industry is on the cusp of a major advance forward in this area with the coming of 3D NAND technology, and Toshiba, SanDisk's flash component fabrication partner, is on the forefront of that evolution. SanDisk needs to ensure a seamless transition for its customer base to this new technology while it leverages the improved cost-effectiveness of its products to continue to drive down the cost of flash-based solutions.

ESSENTIAL GUIDANCE: FINAL THOUGHTS

In today's world, which is dominated by the need to extend services to customers via mobile solutions and the need to use quick analysis of data to change business outcomes, the data center has to be more than the place where an organization keeps its computers. In the world of 3rd Platform, the data center is the foundation for new business. IT executives need solutions that ensure their organizations have access to reliable and dynamically variable (rapidly expanding and contracting) amounts of transactional, content serving, archiving, and analytic capacity – and all of this has to be done on time, with no delays and no excuses.

Flash is already playing a key and multifaceted role in the evolution of IT infrastructure that enables higher-performance, more agile, and more efficient data centers. Server-based flash is appropriate for data center workloads that demand extremely low latency. As workloads have evolved, flash-based arrays are rapidly maturing to support a broader set of applications; exhibiting the same reliability, availability, scalability, and mature data management capabilities associated with today's enterprise-class arrays. The stage will be set for flash-based arrays to become the enterprise storage workhorses of the future, providing a path to the all-flash data center for primary storage for those customers that want it. Flash should already be present, in at least some capacity, in every data center that is using virtual computing, to improve efficiency and reduce operational costs.

Flash is clearly a transformative technology that will enable the next generation data center required to cost-effectively handle the requirements of 3rd Platform computing. In the very near future, flash-based solutions will be at the core of both enterprise and service provider data centers. Aggressive exploitation of flash technologies and SSDs will be one of the key determinants of market leadership in the world of the 3rd Platform.

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