Storage in the era of cloud and big data: the advantages of SSDs over HDDs
Executive overview

According to IDC, “From 2005 to 2020, the digital universe will grow by a factor of 300—from 130 exabytes to 40,000 exabytes.” This exponential expansion of data has been the subject of much discussion in recent years. What gets lost in the conversation, however, is that the “data deluge” is fundamentally changing the way that storage is approached. This shift has less to do with the quantity of storage needed and more to do with what that storage will be used for. As a result, storage purchases will stop being a “one-size-fits-all” commodity and will increasingly become application or business-outcome focused.

This white paper explores the causes of the shift and the implications for the legacy technology—the traditional hard disk drive (HDD)—and presents the case for a new technology: the solid state drive (SSD). SSDs aren’t going to be the right technology for every application, but for many scenarios, they are significantly better than traditional HDDs. An understanding of how storage will be used is critically important to making intelligent decisions about the technology.

As storage shifts from a bulk, commodity purchase to a more strategic investment, it will become increasingly important to choose technology partners who have a deep understanding of the changing nature of storage. For a quarter century, SanDisk has displayed innovation and leadership in solid state mass storage technology and is uniquely positioned to help enterprises embrace the new storage model in a way that makes sense for their business.

Clinging to an outdated storage paradigm

Enterprise storage needs are changing but approaches to planning and purchasing aren’t keeping up. HDDs are cheap, plentiful, have lots of capacity and are usually purchased as a bulk commodity. As a result, most of the discussion has focused on cost per gigabyte ($/GB). The advent of tiered storage has introduced the concept of data access needs (and choices between 15K, 10K, 7.2K or 5.4K RPM drives), but the technology for the drives is fundamentally the same: rotating disks and a magnetic head on a moving arm to read and write data.

HDD technology is well suited for many storage scenarios, but has some intrinsic limitations. Data is stored on spinning platters with “read/write” arms that have to physically move across specific sectors of the hard drive. As a result, HDDs are inherently good at reading data that is stored sequentially within and across sectors because the movement of the “read/write” arm is minimized. But the nature of today’s computing makes it increasingly likely that data will be stored wherever there is space on the drive. The increasingly random nature of data storage is being driven by virtualization—and it can cause significant problems for traditional hard drives.

The virtualization conundrum and the random-access challenge

The amount of data that moves back and forth from data centers around the world is vast. Whether it is uploading photos, downloading video or accessing corporate data, the demand for quick response times is stretching HDDs to their limit. Data center operators have sought a myriad of ways to address these needs and extend capabilities into the cloud. One of the primary solutions for this problem is virtualization. In a virtualized
environment, hardware resources (including compute capability and storage) are “abstracted” (or separated) from the software that uses them. As a result, data centers can operate far more efficiently, because unused compute and storage capacity can be applied where needed.

But virtualization creates unique challenges for an HDD-driven storage infrastructure. “Virtual machines” (VMs) create, store and access data across a pool of virtualized storage. That data is stored wherever the capacity exists, in a random fashion. As the number of VMs trying to access data simultaneously increases, it creates a problem known as the I/O blender effect.

The result is that data is stored in an increasingly random fashion, forcing the read/write arm of a traditional HDD to move constantly. And each one of those moves adds time delays to the system. As the number of VMs increases and the amount of data being searched grows, the delays can really add up. This creates a bottleneck in storage and reduces the responsiveness of data center applications.

Response time—or “latency”—is a challenge for businesses of all kinds. From content providers who are delivering media simultaneously to thousands of users, to securities traders who rely on networked resources for split-second analysis of data, slow response time is unacceptable. But these delays are more than just an annoyance: Latency can directly affect the bottom line by driving away customers, missing a window of opportunity or delaying a critical decision. Eliminating split-second delays at the individual user level means that milli-, micro- and increasingly, nanoseconds matter when a storage system is accessing data.

The problem facing many storage administrators is that in a virtualized pool of storage, all of the storage is treated the same. But in today’s cloud-based, data-driven world, not all storage access is the same. In some cases, the ability to read data quickly is most important. In other cases, the ability to write data quickly is paramount. In order to accommodate these differing needs, IT departments have significantly overprovisioned their storage capacity. Buying additional storage can reduce system latency—but at the expense of buying way more storage than needed.

The quantity of data continues to grow, and so do the number of users trying to simultaneously access shared data pools. Overprovisioning can be an effective short-term solution—but it can’t be sustained in the face
of the scalability requirements for enterprise storage. It also creates the need for extra data center space to accommodate banks of hard drives. Further complicating matters, all those spinning drives use power and generate heat, which must then be dissipated with expensive data center cooling systems.

**The SSD solution—delivering real-world impact**

The good news for data center managers and for storage experts is that SSDs present an attractive alternative to HDDs. Because they have no moving parts, SSDs remove one of the major drawbacks of HDDs. SSDs also eliminate sources of latency as well as heat. The physics of how SSDs access data means they’re better suited for today’s random-access data needs. The old conversation was about $/GB with a focus on the cost of the storage itself. SSDs enable a new approach that focuses on the cost of achieving a business outcome, with new measures such as the cost per transaction, the cost of Input/Output Operations Per Second ($/IOPS), and IOPS per Watt.

But the case for SSDs is about more than just changing the technology or changing the conversation. SSDs deliver in very real, tangible ways. The first, and most obvious, has to do with cost. On the surface, the cost debate between HDDs and SSDs seems straightforward. HDDs are cheaper per GB—even though the cost of SSDs is coming down dramatically. But as mentioned earlier, $/GB is an incomplete, and often misleading, picture. Here are several ways in which SSDs wind up actually being less expensive than HDDs:

- **Lower capital expenditures**
  - Fewer SSDs are needed to deliver the same outcome. The performance characteristics of SSDs make them much less subject to the response time issues that can plague HDDs—and make massive overprovisioning a thing of the past.

- **Lower operational expenditures**
  - Reduced data center space requirements. SSD-based systems not only require fewer drives, they accommodate smaller form factors, enabling two SSDs to fit in the space traditionally required for one HDD.
  - SSDs have lower power consumption and generate less heat without all of those spinning disks. This not only saves on powering the drives themselves, there is less heat to dissipate using expensive data center cooling systems.

- **Save time and money by reducing latency and improving quality of service (QoS)**
  - SSDs are fast. Much faster than HDDs for most read and write operations—as much as four to five times faster² on reads/writes per second.
  - SSDs excel at random reads—so they aren’t as susceptible to latency caused by the demands of virtualization. The result is more predictable performance as an infrastructure scales without overprovisioning.
  - SSDs are ideal for weathering events like a “boot storm,” when many users are trying to log in simultaneously. HDDs suffer from boot-up latency that makes them especially susceptible during “storm” events.

- **Increased reliability and longevity**
  - SSDs are more reliable than HDDs. Unlike HDDs, SSDs have no moving parts. That makes them more resistant to shock, vibration and temperature variations. As a result, they have a higher mean time between failures (MTBF)—a standard industry measure for reliability. SSDs have an industry average MTBF of approximately 2 million hours, while HDDs can only claim 1.5 million hours.³ For SSDs, a 2-million-hour MTBF equates to a lower average failure rate (AFR) of 0.44.
SSDs in practice—where do they make sense?

Despite the advantages of SSDs, they are not a one-size-fits-all solution. There are applications where HDDs still make sense. And although the price of SSDs has come down significantly, HDDs still have an acquisition cost advantage for the near future. For applications where data is written and accessed infrequently—it still makes sense to invest in HDDs. When performance is critical, SSDs are the best solution.

The first consideration for enterprise storage is whether or not there is a need for an enterprise-grade drive. IT departments are under constant pressure to cut costs and demonstrate ROI. As a result, it isn’t surprising that some companies explore using consumer-grade drives in the data center. This may make sense in some scenarios, but the performance expectations for consumer drives are considerably different from enterprise drives:

<table>
<thead>
<tr>
<th></th>
<th>Consumer SSDs</th>
<th>Enterprise SSDs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Performance expectation</strong></td>
<td>Typical 40-hour week</td>
<td>24x7 operation</td>
</tr>
<tr>
<td><strong>Operating temperatures</strong></td>
<td>Air-conditioned office (roughly 76°F/24°C)</td>
<td>Data center temperatures (80°F/27°C and up)</td>
</tr>
<tr>
<td><strong>Data protection</strong></td>
<td>Consumer-grade</td>
<td>Enterprise-class reliability and lower uncorrectable bit error rate (UBER)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Capacitor-backed cache (CBC) protection to ensure data protection even in a power outage</td>
</tr>
<tr>
<td><strong>Flash quality</strong></td>
<td>Standard</td>
<td>High-quality</td>
</tr>
</tbody>
</table>

One of the key advantages with SSDs is the ability to make intelligent decisions about which type of drive makes sense based on specific business needs. There is a range of options when selecting solid state drives—and capacity needs are typically balanced with performance requirements. Even within the realm of high-performance drives, it is critical to understand what kind of performance is needed: Greater write speed or greater read speed? A balance of mixed-use (read/write) performance? What level of error and fault tolerance is required?

SSD USES

Data centers storage needs are varied—from a small number of super-high-performance drives, to a larger number of higher capacity but lower cost drives.
Here are some real-world scenarios where SSDs make sense:

- **Write-intensive scenarios**—For Online Transaction Processing (OLTP) or analyzing search logs from online applications, the speed with which data can be written to a storage environment is critical. Delays can result in transactions that “time out” or that simply deliver a poor experience to customers or employees.

- **Mixed-use scenarios**—Email has become the default means of communication for most businesses today. Ensuring that data gets in and out of the system in a timely manner is critically important to keeping business running.

- **Read-intensive scenarios**—SSD technology is ideal for online content providers, especially video-on-demand (VOD) and image-retrieval applications. With thousands of users accessing high-definition (HD) content simultaneously, the customer experience is completely dependent on how quickly data can be accessed and delivered.

- **Emerging applications (big data/Hadoop/cloud)**—Hadoop is an open-source software framework specifically designed to support data-intensive applications. It is being used by many cloud companies as more and more applications are capturing and generating vast amounts of data. SSDs are ideal for these environments, because they enable advanced data mining and deliver the low-latency transactions required for online services.

**SanDisk—storage solutions for the enterprise**

Unlike competitors who approach enterprise storage from a memory perspective, SanDisk has always been a storage company. This is not inconsequential for the enterprise. With memory, the key concern is data accessibility and not necessarily long-term viability. As a storage company, we understand that data protection is paramount in the data center, and that reliability and accessibility must be balanced to suit unique needs. And that protection-first perspective is built into every SanDisk product.

SanDisk also has the advantage of vertical integration, from the semiconductor NAND technology to the NAND die, from the wafer to the controller, and from the SSD to the software that optimizes it. This enables us to be more selective about the components we choose and more meticulous in our production processes. It means that we have the supply chain flexibility to meet changing market demands. It means that we can offer better service and support because we engineer all of our components to work better together—and work seamlessly with existing infrastructures.

Our vertical integration also means that we can bring a high-quality enterprise product to market at a competitive price. For example, our proprietary Guardian Technology™ Platform allows inexpensive MLC flash to be used for enterprise applications by taking a system-level approach to flash management. It improves the endurance of MLC flash, while protecting from data loss or data corruption. This extension of MLC flash into the data center helps bridge the gap between HDD and SSD technology.

SanDisk has been a global leader in solid state storage technology for more than 25 years. We pioneered flash memory mass storage technology—and our heritage of storage innovation has enabled us to become a leader in meeting the growing storage needs of the enterprise data center. SanDisk provides unique hardware and software solutions that deliver predictable performance, trusted reliability and superior value.
Our products are also rigorously tested specifically for enterprise environments and held to the industry’s highest standards. After four product generations, our customer field data and reliability testing proves that we have met or exceeded our design targets time and time again. For example, two million hours MTBF on our SAS SSD products is more than just a promise—it is a mark our products are consistently exceeding in the real world.

The SanDisk product line has become a fixture in several tier-one OEM enterprise products. OEMs really benefit from our vertical integration. Our supply chain flexibility makes it easier for them to respond to changes in product demand. Because we control the entire process, there’s less finger pointing between vendors—so they get better support for their products. And our pricing advantages help them maintain their margins while still delivering an affordable product.

Our SAS SSDs and PCI Express (PCIe) drives are workload-optimized for a range of application solutions, from small to medium businesses (SMB) to hyperscale enterprise environments.

**SSD USES**

With the recent acquisition of SMART Storage™ Systems, we are extending our product line to include enterprise SATA SSDs and flash DIMMs as well. SanDisk is poised to help transform data centers by helping optimize storage to meet application performance requirements.
Embrace the future of storage with less risk—with SanDisk

HDDs still have a place in the modern data center. But SSDs present a growing, compelling alternative to this older technology. SSDs allow for an outcome-focused approach to storage. And there are endless possibilities that can be opened up when current response time and throughput constraints are removed. SanDisk SSDs offer predictable performance, trusted reliability and cost-effective solutions for a variety of data center requirements.

Data centers—whether large or small—need storage that is easily integrated, scalable and simple to manage. Storage must be efficient and reliable, but also cost-effective—saving money, space and power. Enterprises deserve a storage partner who understands the things that matter most—and who can offer solutions that meet your specific needs.

For 25 years, SanDisk has had a singular focus: to expand the possibilities of data storage technology. In the enterprise, it’s no different. We’re committed to helping unlock business potential through high-performance, reliable storage, and delivering real peace of mind along the way.

www.sandisk.com

SanDisk Corporation
951 SanDisk Drive
Milpitas, CA 95035-7933
Phone: 408-801-1000
Fax: 408-801-8657

2 http://www.storagereview.com/ssd_vs_hdd
3 http://www.storagereview.com/ssd_vs_hdd
4 Based on internal testing using Telcordia stress part testing.

© 2013 SanDisk Corporation. All rights reserved. SanDisk is a trademark of SanDisk Corporation, registered in the United States and other countries. Guardian Technology, CloudSpeed, FlashSoft and SMART Storage are trademarks of SanDisk Enterprise IP, LLC. Lighting and Optimus are U.S. registered trademarks of SanDisk Enterprise IP, LLC. Other brand names mentioned herein are for identification purposes only and may be the trademarks of their respective holder(s).