



WHITE PAPER

SanDisk's Enterprise-Class SSDs Accelerate eDiscovery Access in the Data Center

June 2014

By
Ravi Naik, Chief Information Officer, SanDisk
Martyn Wiltshire, Director, SanDisk® IT

Table of Contents

- 1. Executive Overview 3**
- 2. The Business Need for Speed 4**
- 3. SanDisk Legal and eDiscovery Economics 6**
- 4. Enter SanDisk IT 6**
- 5. SanDisk Flash and Enterprise Storage: The Data Dilemma 7**
- 6. Data Center Efficiencies: Reducing Operational Costs..... 10**
- 7. SUMMARY: The Future of SSDs in the Enterprise.....12**

1. Executive Overview

Solid State Disks (SSDs), with their inherent high performance, high capacity and support for enterprise workload acceleration, have become a driving force in the enterprise – and they are viewed as critical to the future of the data center and the storage industry.

Traditionally, companies have populated their data centers with Hard Disk Drives (HDDs) using Storage Area Network (SAN) and Network Attached Storage (NAS) technologies. These HDDs tackle day-to-day processing and storage requirements, and have a complementary place in data centers, alongside higher performance all-flash storage arrays with racks of SSDs.¹

However, traditional HDD storage using spinning magnetic media is increasingly proving to be less than ideal for critical business applications that have ultra-high-performance needs. Examples include: time-sensitive Big Data Analytics, VDI and eDiscovery with large volumes of email.

Latency, performance and mean time between failure (MTBF) – all are improved with the use of SSDs to do the same tasks. All-flash storage arrays with racks of enterprise-grade SSDs are used for high-performance, high-throughput processing, with near-instantaneous access, and storage of essential business data.

The pervasiveness of cloud computing has only exacerbated the contrast between traditional HDDs and higher-performance flash storage because of the urgent need for people to access, download or analyze vast amounts of information in milliseconds, anytime and anywhere. As a result, SSDs and the all-flash storage arrays in which they reside are increasingly taking on the tasks of the highest-performance HDDs in the data center...”²

¹ IDC: “Worldwide Enterprise All-Solid State Storage Array, 2013-2016 Forecast”, April 2013

² Gartner: “Solid State Drives Will Complement, Not Replace Hard-Disk Drives in Data Centers,” April 15, 2013

2. The Business Need for Speed

Flash storage is an enterprise game-changer. SSDs offer a powerful combination of high-performance, capacity, endurance and reliability. These attributes are extremely important for certain critical business applications, such as rapidly and cost-effectively indexing, sorting and storing huge volumes of email and other critical data. Such a real business need proved to be the use case at SanDisk, a global leader in flash-storage solutions.

Implementing an eDiscovery Solution at SanDisk

SanDisk IT decided to implement an eDiscovery solution to meet the complex requirements from the company's Legal department, which must comply with Records and Information Management (RIM) Governance, case, management and litigation hold requirements. The solution required an innovative use of SanDisk SSD technology deployed in an all-flash storage array.

After careful consideration of all of the IT and business factors, SanDisk decided to deploy an HP Autonomy solution for eDiscovery of all of its legal documents – and to install a Kaminario all-flash storage array to retain the digitized data for fast access to that data. SanDisk decided to leverage flash SSDs because they offer a powerful combination of high-performance capacity, endurance and reliability.

Before the Deployment

In any company, compliance with retention-of-data regulations is critical because it helps keep information safe, lowers the risk of civil or criminal penalties, saves millions of dollars in unnecessary litigation costs, and reduces storage and backup costs. Until recently, SanDisk Corp. did not limit how long employees are allowed to retain email and, instead of a single company-wide email archive solution, employees used Exchange Outlook PST files and, in some regions, Enterprise Vault (EVault) to store copies of email messages.

SanDisk IT decided to implement an eDiscovery solution to meet the complex requirements from the company's Legal department, which must comply with Records and Information Management (RIM) Governance, case, management and litigation hold requirements. The solution required an innovative use of SanDisk SSD technology deployed in an all-flash storage array.

After careful consideration of all of the IT and business factors, SanDisk decided to deploy an HP Autonomy solution for eDiscovery of all of its legal documents – and to install a Kaminario all-flash storage array to retain the digitized data for fast access to that data. SanDisk decided to leverage flash SSDs because they offer a powerful combination of high-performance capacity, endurance and reliability.

To implement eDiscovery and RIM Governance requirements, the SanDisk Legal department needed a solution to implement the specific retention rules established for Digital communications and emails. The eDiscovery solution, in addition to implementing digital retention rules, provided legal capabilities for legal hold, early case assessment, external case management and extensive analytics. For end- user emails, retention rules established in the eDiscovery system dispose of digital communications automatically at the end of the specific retention period unless it is required by active litigation holds. In these cases, SanDisk Legal preserved copies of the emails before disposal from an employee's mailbox.

This led to the need for SanDisk to establish one email archive global solution to replace EVault and to eliminate local PST file archiving, which would be accessible from the worldwide web and Microsoft Outlook.

To put all of this in perspective, consider that SanDisk employees around the world send and receive more than 500,000 to 1 Million emails per day, according to data compiled by SanDisk's IT organization. To store all that data, SanDisk IT maintains a corporate email archive that stores over 21 Terabytes (TB) of data. At 25,000 pages per gigabyte (GB), that is more than 500 Million pages of documents that reside in SanDisk's email archive. The cost of searching and reviewing an email archive of that size is prohibitive.

Addressing the Pain-points in the Legacy eDiscovery Solution

These indexing, sorting, and archiving processes are recognized as time-consuming and expensive pain points, which is why SanDisk's Legal department adopted HP Autonomy's e-Discovery platform to serve as the company's single, global archive solution. SanDisk, having recently celebrated its 25th anniversary, has accumulated an enormous volume of emails and attachments since it was founded in 1988. Based on RIM Governance policies, SanDisk Legal now will retain emails, plus attachments, with identified business or legal importance for up to 21 years, in the HP Autonomy application. The goal is to provide case management and search capabilities for the Legal team so that they can effectively manage and contain litigation, which is typical for companies with a global footprint and large patent portfolio.

SanDisk is rolling out the HP Autonomy application to provide powerful email search and archiving capabilities to all 5,000-plus employees worldwide. When complete, this effort will allow employees to search and sort two years' worth of their emails, which will put substantially less strain on the IT infrastructure. This initiative also will improve the performance of employee computers by reducing the amount of local storage required on their machines and, most important, provide the Legal department with the ability to rapidly search, process and manage the retention of digital communications.

3. SanDisk Legal and eDiscovery Economics

Using an all-flash array, SanDisk was able to dramatically reduce the processing time for its eDiscovery workloads. With the hard-disk-drive (HDD) solution, it took 50 days to load 250 million email messages – processing 2 million emails per day – including loading them, indexing them and archiving them. With the SanDisk/Kaminario all-flash solution, it took just 12 days to load 100 million emails, processing them at a rate of 7.2 million messages per day.

When a litigation case is presented at SanDisk, the company's attorneys first need to scope and size all aspects of the case. Strict deadlines drive the process and case information is needed immediately. Any delays can impede a case and drive legal fees to extraordinarily high levels. The strategy is essentially therefore one of cost avoidance by eliminating irrelevant emails and documents.

A simplified look at eDiscovery economics quickly demonstrates the importance of accurately scoping the parameters of a case and accessing information as quickly as possible. An attorney generally takes 167 hours to read one GB of email (about 20,000 emails). At two emails per minute, or 120 per hour, the legal cost to read up to one GB of email is estimated by the SanDisk Legal Department to range anywhere from about \$18,000 to \$42,000, depending on variations in legal fees.

With this high cost in mind, even a short delay is unacceptable. Imagine a situation in which a pianist practices a piano piece without hearing the sound for 6 to 12 hours. In this example, it would take multiple iterations over an extraordinarily long period of time to learn a piece. This simply would not work. Likewise, in the real world of eDiscovery and litigation case management, the legal team needs to be able to run repeated data searches, refining and reducing the search results for presentation in trial. This demanding real-world application requires high performance storage, capable of significant continuous data ingestion, de-duplication, and complex indexing and tagging, with the Tier 1 Service Level Agreements (SLAs) on a 24/7 basis, 365 days a year.

4. Enter SanDisk IT

In order to implement the HP Autonomy eDiscovery solution, SanDisk's attorneys turned to SanDisk IT to migrate the data from Microsoft Exchange and the legacy EVault archiving platforms into HP Autonomy. This involved downloading and processing emails and their attachments, and eliminating all duplicates so that only a single copy of an email message is stored. Finally, and most important, this also involved tagging messages and attachments with metadata and building a powerful index to meet the demanding search requirements of SanDisk Legal. In total, more than 250 Million messages were processed – and then eventually were reduced to 100 Million unique indexed messages (11TB data) on the HP Autonomy platform.

Based on trial runs, SanDisk IT estimated that to load, index and archive these quarter of a billion emails (plus attachments) into traditional HDD storage, the task would take 50 days at a processing rate of 2 million emails per day. This was considered an unacceptable outcome because it would jeopardize the release of the eDiscovery capabilities to the Legal team and, eventually, the release of the powerful email search and archive functionality to the rest of the company.

In a bid to accelerate the timetable, SanDisk's CIO assigned a team of engineers to evaluate the performance of several all-flash storage arrays that would serve as an alternative to HDDs for this particular ultra-high performance application. The hypothesis that drove this decision was that low latency (<0.5ms) flash storage would virtually eliminate the immense email indexing challenge.

5. SanDisk Flash and Enterprise Storage: The Data Dilemma

SanDisk IT embarked on lab test evaluations using HP Autonomy software to test all-flash array storage offerings since existing traditional HDD storage had maximized its available capacity. Of all the systems tested, an all-flash array storage solution from Kaminario, a flash storage company, met SanDisk's rigorous specifications for endurance, performance and reliability.

At the heart of the all-flash storage array is the 800GB capacity Optimus solid-state drive (SSD) from SanDisk. This enterprise-class SSD offers predictably sustained high performance, reliability and endurance while also lowering the cost of enterprise-grade solid-state storage. The Kaminario architecture is scalable and features proprietary technology, including N + 1 redundancy to ensure the high resilience and uptime of SSDs (see figure 1).

The SanDisk Optimus SAS SSD Advantages

- Interface: SAS 6Gb/s
- 24nm MLC Capacity*: 800GB
- Sustained Read/Write (MB/s)**: 500/500 MB/s ; 1 GB/s Wide-Port
- Random Read/Write: 95K/40K IOPS
- Mean Time Between Failures (MTBF)***: 2.5 Million Hours
- Endurance: Drive Writes per Day (random/sequential)****: 50/50 DWPD

Kaminario Technical Advantages

- Kaminario K-Block has N+1 Redundancy. Any failure of a component inside a K-Node (SSD, PCB etc) will cause a K-Node to failover to a management node.
- For a single node failure in a K-Block this N+1 architecture is acceptable without any significant performance degradation
- For a multi-node failure in a K-Block before remediation has completed, this N+1 architecture will shut down the storage array. In this scenario any data written after the first node failure is at risk for data loss

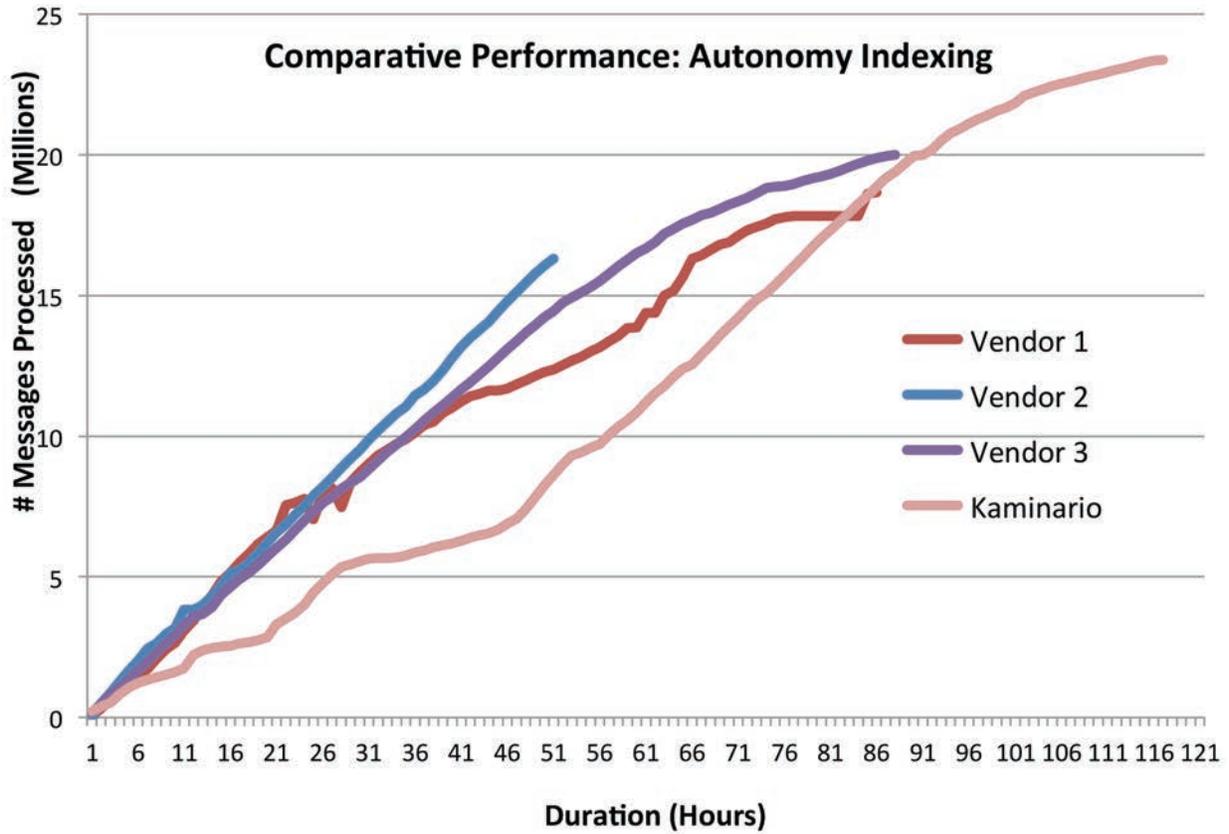


Figure 1: SanDisk IT Test Lab Performance

With the SanDisk/Kaminario all-flash solution, the data dilemma was solved. Against a production workload, HP Autonomy indexed 100 million emails and attachments in just 12 days, averaging 7.2 Million email messages per day—a result that surpassed even the most optimistic estimates and was a fraction of the 50 days (two million emails a day) that it took conventional HDDs to complete the same task. The all-flash array storage solution achieved blazing performance, high IOPS, low latency and high resilience (see figure 2). The architecture was optimized for data protection and proved to have high reliability and availability.

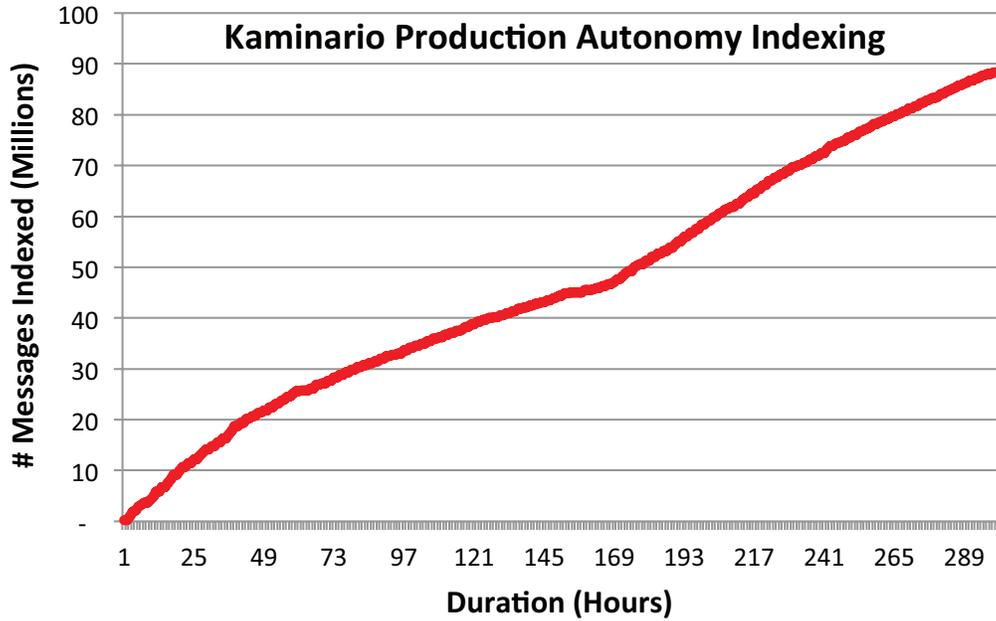
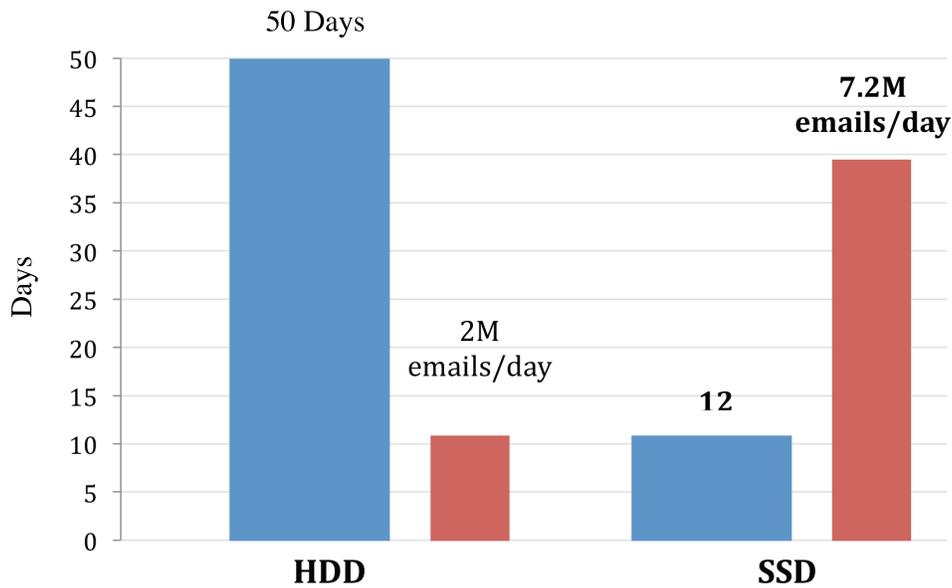


Figure 2: SanDisk IT Production Performance



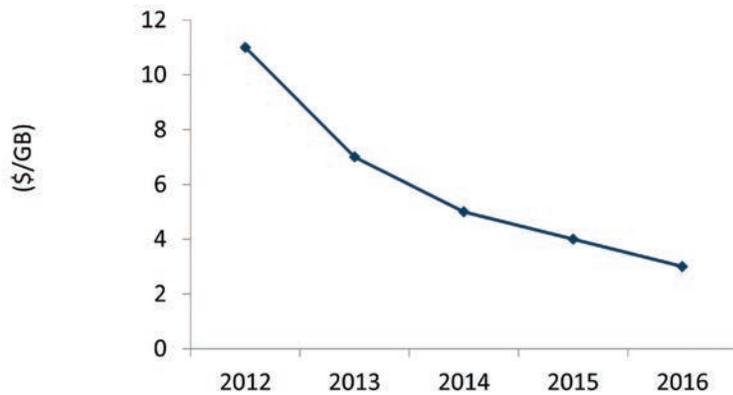
Another benefit related to SSDs is their ability, as compared to HDDs, to dramatically reduce system latency and vastly improve the utilization of Central Processing Units (CPUs) in a highly-virtualized data center environment. In some cases, the rate at which CPUs can “crunch” data when coupled with SSDs increases by an order of magnitude (10^x),³ enhancing the performance of the entire infrastructure - including storage, servers and networking equipment.

³ Gartner: “Solid State Drives Will Complement, Not Replace Hard-Disk Drives in Data Centers,” April 15, 2013

6. Data Center Efficiencies: Reducing Operational Costs

Traditional HDD storage can be less expensive in the short-run than flash storage, but the price gap is narrowing and the cost of All-Flash Solid State Storage Arrays is declining rapidly (see figure 3). In the long run, the Total Cost of Ownership (TCO) of SSDs makes them a highly desirable cost-effective solution in the high-density data center. The TCO of SSDs must also take into account considerations beyond the cost per gigabyte. Those include much higher performance of SSDs, lower power consumption, reduced cooling costs, fewer chassis racks, and a smaller data center footprint⁴ (see figure 4).

**Worldwide Enterprise All-Solid State Storage Array
Average Cost per Gigabyte, 2012 – 2016**



Source: IDC, April 2013

Figure 3

Advantage: SSD vs. HDD in Enterprise Storage

Performance

- RW Speed** 5x Faster
- Data Access Time** <1%
- Concurrent Access** 900% Better
- IOPS** 475% Better

Operating Time/Support

- Reduction in:
- ✓ Boot up
- ✓ Virus Scan
- ✓ Defrag
- ✓ RAID Build,
- ✓ Patching
- ✓ Data Restoration

Environmental

- Shock/Vibration** Higher Resistance to
-Shock 8x Better, Anti-Vibration 16x Better

Operating Temp

- Spec'd at 2x Wider Operating Temp Range

Noise

- 30dB Lower

Space

- Savings in Space \$ at \$/sq ft Office

Weight

- 50% Less Weight

Power

- 92% Less Power, 38% Less Temp

Total Cost of Ownership

- 50% of HDD

Source: IMEX Research SSD Industry Report 2011 for Enterprise Storage Systems

Figure 4

⁴ Gartner: "Solid State Drives Will Complement, Not Replace Hard-Disk Drives in Data Centers," April 15, 2013

Another benefit related to SSDs is their ability, as compared to HDDs, to dramatically reduce system latency and vastly improve the utilization of Central Processing Units (CPUs) in a highly-virtualized data center environment. In some cases, the rate at which CPUs can “crunch” data when coupled with SSDs increases by an order of magnitude (10^x),⁵ enhancing the performance of the entire infrastructure – including storage, servers and networking equipment.

CIOs, therefore, have the opportunity to leverage investments in applications, hardware and software. In addition, they can reduce and/or defer the need to make financial outlays to buy additional equipment to upgrade data center infrastructure to overcome the latency shortcomings of traditional storage (see figure 5).

Flash Drives Savings in the Enterprise

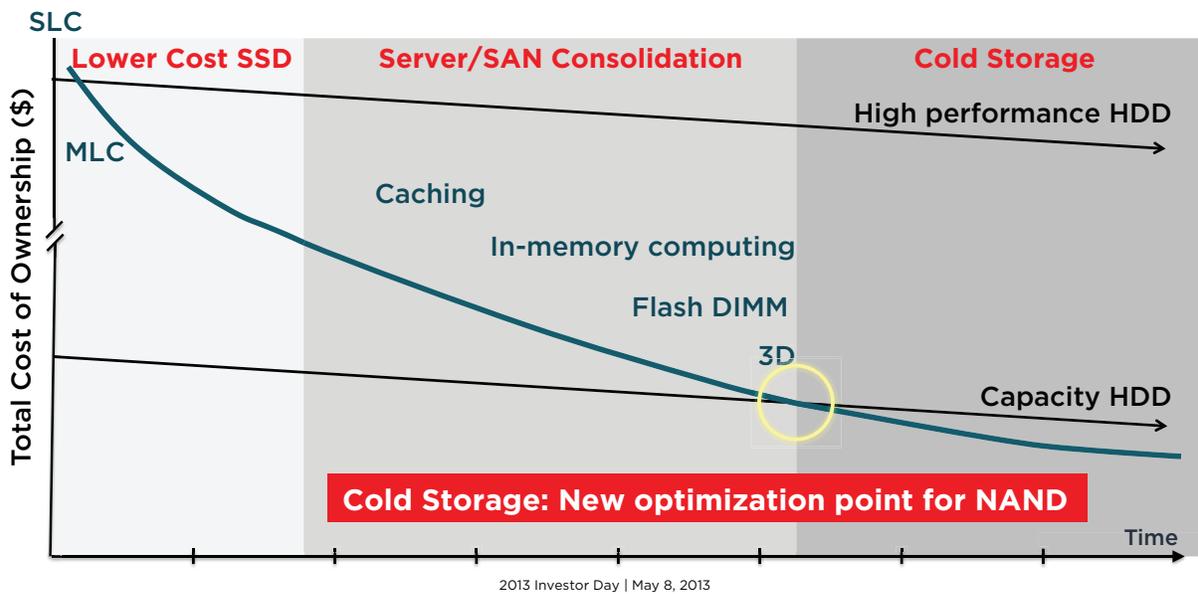


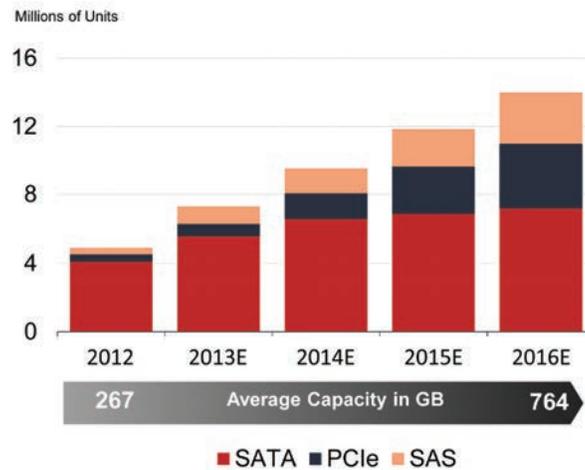
Figure 5

Market research has come to a similar conclusion about the disparity in performance between traditional storage (HDDs) and server-based CPUs, and it has been noted that this performance disparity has created a huge market opportunity for SSDs to meet the ongoing datacenter challenges (see figure 6) related to latency and input/output operations per second (IOPS).⁶

⁵ Gartner: “Solid State Drives Will Complement, Not Replace Hard-Disk Drives in Data Centers,” April 15, 2013

⁶ IDC: “Worldwide Enterprise All-Solid State Storage Array, 2013-2016 Forecast”, April 2013

Enterprise Flash: A Growth Driver



Source: SanDisk estimates

Figure 6

7. SUMMARY: The Future of SSDs in the Enterprise

With billions of emails being generated each year worldwide, CIOs constantly wrestle with the problem of efficiently storing that information and making it immediately accessible to the business.

Executives around the world in data-intensive industries are demanding near-instantaneous response times. Because of their high performance and low latency attributes, SSDs are being broadly adopted by CIOs in various industries including legal, finance, banking, high technology, airlines, automotive, manufacturing and many more. SSDs are ideal for a number of common data-intensive storage applications, including on-line transaction processing (OLTP), high performance computing, data warehousing, Big Data real-time analytics and cloud computing.⁷

SSDs in the Data Center

- Thin, dense, lightweight semiconductor storage devices
- Extremely high performance with low read/write latency
- Very low power consumption and low heat dissipation
- Small physical footprint with fewer storage racks and chassis
- High endurance to process high-performance workloads
- High resistance to shock and vibration
- Generates substantially less noise than a traditional HDD
- Low failure rates and low operating maintenance
- Ideal in virtualized data center environments

⁷ IDC: "Worldwide Enterprise All-Solid State Storage Array, 2013-2016 Forecast", April 2013

As a result, SSDs are transforming the data center. The day is not far off when all-flash storage arrays will become ubiquitous in corporate and government data centers, as well as those managed by third-party service providers.

SanDisk IT, recognizing these trends, set out to prove the sustained performance advantages of SSDs in SanDisk's real-time enterprise operating environment. HP Autonomy software, running on SanDisk Optimus SSDs and Kaminario's advanced storage architecture, survived numerous regression lab tests and also demonstrated world-class resilience in a demanding production environment. Based on the phenomenal results, SanDisk IT foresees the SanDisk SSD/Kaminario combination as a focused high-value solution for HP Autonomy software and for select high-performance applications in a broad cross-section of industries.

*1 GB = 1,000,000,000 bytes. Actual user capacity less.

**Up to stated speed. Based on internal testing; performance may vary depending upon drive capacity, host device, OS and application.1
megabyte (MB) = 1 million bytes.

***MTBF - Mean Time Between Failures based on parts stress analysis.

****Warranty/DWPD - The lesser of 5 years from the date of manufacture of the product or the date on which the product's relevant endurance thresholds set forth in the product specifications are reached.

Specifications are subject to change. © 2014 - 2016 Western Digital Corporation. All rights reserved. SanDisk and the SanDisk logo are trademarks of Western Digital Corporation or its affiliates, registered in the United States and other countries. Optimus is a trademark of Western Digital Corporation or its affiliates. Other brand names mentioned herein are for identification purposes only and may be the trademarks of their holder(s). Enterprise-Class SSDs v2 - 06.21.2016.

Western Digital Technologies, Inc. is the seller of record and licensee in the Americas of SanDisk® products