



WHITE PAPER

CloudSpeed™ SATA SSDs Support Faster Hadoop Performance and TCO Savings

August 2014

Table of Contents

1. Executive Summary	3
Summary of Results	3
Summary of Flash-Enabled Hadoop Cluster Testing: Key Findings	3
2. Apache™ Hadoop®	4
3. Hadoop and SSDs - When to Use Them	4
4. The TestDFSIO Benchmark.....	5
5. Test Design.....	5
Test Environment.....	6
Technical Component Specifications	6
Hardware.....	6
Software	7
Compute Infrastructure	7
Network Infrastructure.....	7
Storage Infrastructure	7
Cloudera® Hadoop Configuration.....	8
Operating System (OS) Configuration.....	9
6. Test Validation	9
Test Methodology	9
Results Summary.....	10
7. Results Analysis	12
Performance Analysis.....	12
TCO Analysis (Cost/Job Analysis)	12
8. Conclusions	14
9. References.....	14

1. Executive Summary

The IT industry is increasingly adopting Apache™ Hadoop® for analytics in big data environments. SanDisk® investigated how the performance of Hadoop applications could be accelerated by taking advantage of solid state drives (SSDs) in Hadoop clusters.

For more than 25 years, SanDisk® has been transforming digital storage with breakthrough products and ideas that improve performance and reduce latency, pushing past the boundaries of what's been possible with hard disk drive (HDD)-based technology. SanDisk flash memory technologies can be found in many of the world's largest data centers. (See product details at www.sandisk.com.)

This technical paper describes the TestDFSIO benchmark testing for a 1TB dataset on a Hadoop cluster. The intent of this paper is to show the benefits of using SanDisk SSDs within a Hadoop cluster environment.

CloudSpeed Ascend™ SATA SSDs from SanDisk were used in this TestDFSIO testing. CloudSpeed Ascend SATA SSDs are designed to run read-intensive application workloads in enterprise server and cloud computing environments. The CloudSpeed™ SATA SSD product family offers a portfolio of workload-optimized SSDs for mixed-use, read-intensive and write-intensive data center environments with all of the features expected from an enterprise-grade SATA SSD.



Summary of Results

- Replacing all the HDDs with SSDs reduced 1TB TestDFSIO benchmark runtime for read and write operations, **completing the job faster** than an all-HDD configuration.
- The higher IOPS (input/output operations per second) delivered by this solution led to faster job completion on the Hadoop cluster. That translates to a much more efficient use of the Hadoop cluster by running more number of jobs in the same period of time.
- The ability to run more jobs on the Hadoop cluster translates into total cost of ownership (TCO) savings of the Hadoop cluster in the long-run (over a period of 3-5 years), even if the initial investment may be higher due to the cost of the SSDs themselves.

Summary of Flash-Enabled Hadoop Cluster Testing: Key Findings

SanDisk tested a Hadoop cluster using the Cloudera® Distribution of Hadoop (CDH). This cluster consisted of one NameNode and six DataNodes. The cluster was set up for the purpose of measuring the performance when using SSDs within a Hadoop environment, focusing on the TestDFSIO benchmark.

For this testing, SanDisk ran the standard Hadoop TestDFSIO benchmark on multiple cluster configurations to examine the results under different conditions. The runtime and throughput for the TestDFSIO benchmark for a 1TB dataset was recorded for the different configurations. These results are summarized and analyzed in the Results Summary and Results Analysis sections.

The results revealed that SSDs can be deployed in Hadoop environments to provide **significant performance improvement (> 70% for the TestDFSIO benchmark¹) and TCO benefits (> 60% reduction in cost/job for the TestDFSIO benchmark²)** to organizations that deploy them. This means that customers will be able to add SSDs in clusters that have multiple servers with HDD storage. In summary, these tests showed that the SSDs are beneficial for Hadoop environments that are storage-intensive.

2. Apache™ Hadoop®

Apache Hadoop is a framework that allows for the distributed processing of large datasets across clusters of computers using simple programming models. It is designed to scale up from a few servers to several thousands of servers, if needed, with each server offering local computation and storage. Rather than relying on hardware to deliver high-availability, Hadoop is designed to detect and handle failures at the application layer, thus delivering a highly available service on top of a cluster of computers, any one of which may be prone to failures.

Hadoop Distributed File System (HDFS) is the distributed storage used by Hadoop applications. A HDFS cluster consists of a NameNode that manages the file system metadata and DataNodes that store the actual data. Clients contact NameNode for file metadata or file modifications and perform actual file I/O directly with DataNodes.

Hadoop MapReduce is a software framework for processing large amounts of data (multi-terabyte data-sets) in-parallel on Hadoop clusters. These clusters leverage inexpensive servers – but they do so in a reliable, fault-tolerant manner, given the characteristics of the Hadoop processing model.

A MapReduce job splits the input data-set into independent chunks that are processed by the map tasks in a completely parallel manner. The framework sorts the outputs of the maps, which are then input to the reduce tasks. Typically, both the input and output of a MapReduce job are stored within the HDFS file system. The framework takes care of scheduling tasks, monitoring them and re-executing the failed tasks. The MapReduce framework consists of a single (1) “master” JobTracker and one “slave” TaskTracker per cluster-node. The master is responsible for scheduling the jobs component tasks on the slaves, monitoring them and re-executing the failed tasks. In this way, the slave server nodes execute tasks as directed by the master node.

¹Please see Figure 2: Runtime Comparisons, page 11

²Please see Figure 4: \$/Job for Read-Intensive Workloads and Figure 5: \$/Job for Write-Intensive Workloads, page 13

3. Hadoop and SSDs - When to Use Them

Typically, Hadoop environments use commodity servers, with SATA HDDs used as the local storage, as cluster nodes. However SSDs, used strategically within a Hadoop environment, can provide significant performance benefits due to their NAND flash technology design (no mechanical parts) and their ability to reduce latency for computing tasks. These characteristics affect operational costs associated with using the cluster in an ongoing basis.

Hadoop workloads have a lot of variation in terms of their storage access profiles. Some Hadoop workloads are compute-intensive, some are storage-intensive and some are in between. Many Hadoop workloads use custom datasets, and customized MapReduce algorithms to execute very specific analysis tasks on the datasets.

SSDs in Hadoop environments will benefit storage-intensive datasets and workloads. This technical paper discusses one such storage-intensive benchmark called TestDFSIO.

4. The TestDFSIO Benchmark

TestDFSIO is a standard Hadoop benchmark. It is an I/O-intensive workload and it involves operations that are 100% file-read or file-write for Big Data workloads using the MapReduce paradigm. The input and output data for these workloads are stored on the HDFS file system. The TestDFSIO benchmark requires the following input:

1. **Number of files:** This component indicates the number of files to be read or written.
2. **File Size:** This component gives the size of each file in MB.
3. **-read/-write flags:** These flags indicate whether the I/O workload should do 100% read operations or 100% write operations. Note that for this benchmark, both read and write operations cannot be combined.

The testing discussed in this paper used the enhanced version of the TestDFSIO benchmark, which is available as part of the Intel® HiBench Benchmark suite.

In order to attempt stock TestDFSIO on a 1TB dataset, with 512 files, each with 2000MB of data, execute the following command as 'hdfs' user with number of files, file size and type of operation (whether to perform read or write operations). Note that the unit of file size is MB:

```
# hadoop jar /opt/cloudera/parcels/CDH-4.6.0-1.cdh4.6.0.p0.26/lib/hadoop-  
mapreduce/hadoop-mapreduce-client-jobclient-2.0.0-cdh4.6.0-tests.jar TestDFSIO  
[-write | -read ] -nrFiles 512 -fileSize 2000
```

5. Test Design

A Hadoop cluster using the Cloudera Distribution of Hadoop (CDH) consisting of one NameNode and six DataNodes was set up for the purpose of determining the benefits of using SSDs within a Hadoop environment, focusing on the TestDFSIO benchmark. The testing consisted of using the standard Hadoop TestDFSIO benchmark on different cluster configurations (described in the Test Methodology section). The runtime for the TestDFSIO benchmark on a 1TB dataset (with 512 files, each with 2000MB of data) was recorded for the different configurations that were tested. The enhanced version of TestDFSIO provides average aggregate throughput for the cluster for read and writes, which was also recorded for the different configurations under test. The results of these tests are summarized and analyzed in the Results Analysis section.

Test Environment

The test environment consists of one Dell® PowerEdge R720 server being used as a single NameNode and six Dell PowerEdge R320 servers being used as DataNodes in a Hadoop cluster. The network configuration consists of a 10GbE private network interconnect that connects all the servers for internode Hadoop communication and a 1GbE network interconnect for management network. The Hadoop cluster storage is switched between HDDs and SSDs, based on the test configurations. Figure 1 below shows a pictorial view of the environment, which is followed by a listing of the hardware and software components used within the test environment.

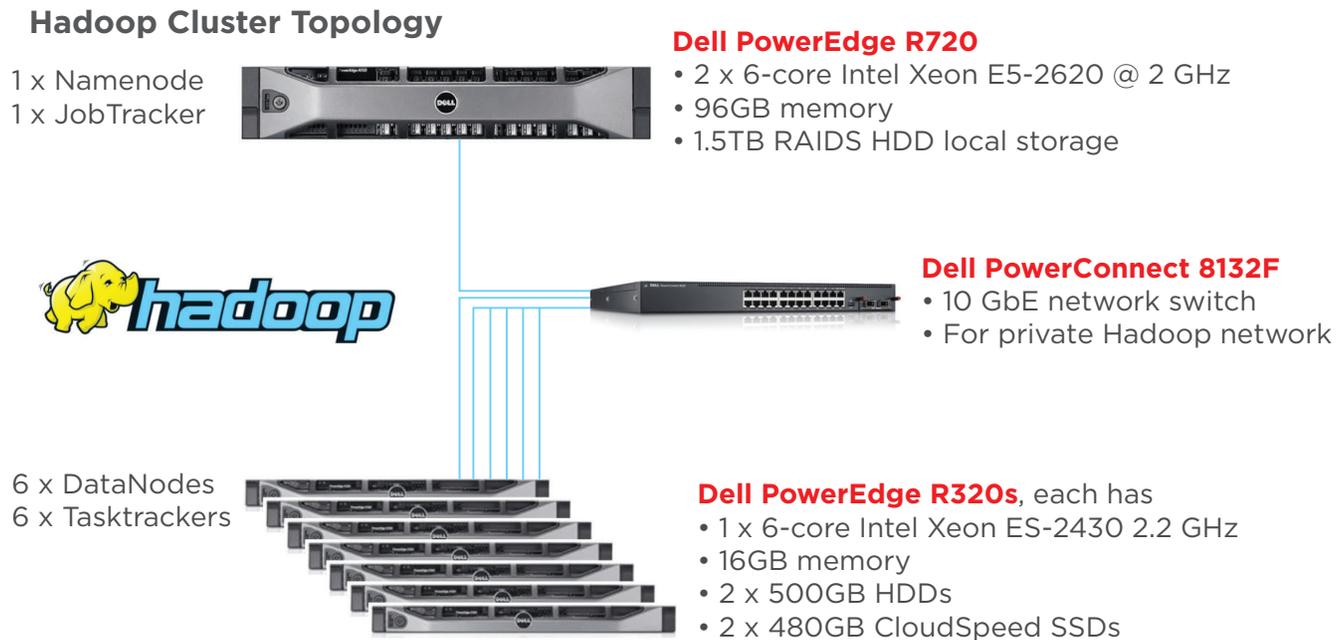


Figure 1.: Cludera Hadoop Cluster Environment

Technical Component Specifications

Hardware

Hardware	Software if applicable	Purpose	Quantity
Dell® PowerEdge R320	<ul style="list-style-type: none"> • RHEL 6.4 • Cludera® Distribution of Hadoop 4.6.0 	DataNodes	6
Dell® PowerEdge R720	<ul style="list-style-type: none"> • RHEL 6.4 • Cludera Distribution of Hadoop 4.6.0 • Cludera Manager 4.8.1 	NameNode, Secondary NameNode	1
Dell® PowerConnect 2824 24-port switch	1GbE network switch	Management network	1
Dell® PowerConnect 8132F 24-port switch	10GbE network switch	Hadoop data network	1
500GB 7.2K RPM Dell SATA HDDs	Used as Just a bunch of disks (JBODs)	DataNode drives	12
480GB CloudSpeed Ascend™ SATA SSDs	Used as Just a bunch of disks (JBODs)	DataNode drives	12
Dell® 300GB 15K RPM SAS HDDs	Used as a single RAID 5 (5+1) group	NameNode drives	6

Table 1.: Hardware components

Software

Software	Version	Purpose
Red Hat® Enterprise Linux	6.4	Operating system for DataNodes and NameNode
Cloudera® Manager	4.8.1	Cloudera Hadoop cluster administration
Cloudera® Distribution of Hadoop (CDH)	4.6.0	Cloudera's Hadoop distribution

Table 2.: Software components

Compute Infrastructure

The Hadoop cluster NameNode is a Dell PowerEdge R720 server with two hex-core Intel® Xeon E5-2620 2GHz CPU (hyper-threaded) and 96GB of memory. This server has a single 300GB SSD as a boot drive. The server has dual power-supplies for redundancy.

The Hadoop cluster DataNodes are six Dell PowerEdge R320 servers, each with one hex-core Intel® Xeon E5-2430 2.2GHz CPUs (hyper-threaded) and 16GB of memory. Each of these servers has a single 500GB 7.2K RPM SATA HDD as a boot drive. The servers have dual power supplies for redundancy.

Network Infrastructure

All cluster nodes (NameNode and all DataNodes) are connected to a 1GbE management network via the onboard 1GbE Network Interface Card (NIC). All cluster nodes are also connected to a 10GbE Hadoop cluster network with an add-on 10GbE NIC. The 1GbE management network is via a Dell PowerConnect 2824 24-port 1GbE switch. The 10GbE cluster network is via a Dell PowerConnect 8132F 10GbE switch.

Storage Infrastructure

The NameNode uses six 300GB 15K RPM SAS HDDs in RAID5 configuration for the Hadoop file system. RAID5 is used on the NameNode to protect the HDFS metadata stored on the NameNode in case of disk failure or corruption. This NameNode setup is used across all the different testing configurations. The RAID5 logical volume is formatted as an ext4 file system and mounted for use by the Hadoop NameNode.

Each DataNode has one of the following storage environments depending on the configuration being tested. The specific configurations are discussed in detail in the Test Methodology section:

1. 2 x 500GB 7.2K RPM Dell SATA HDDs, OR
2. 2 x 480GB CloudSpeed Ascend SATA SSDs

In each of the above environments, the disks are used in a JBOD (just a bunch of disks) configuration. HDFS maintains its own replication and striping for the HDFS file data and therefore a JBOD configuration is recommended for the data nodes for better performance as well as better failure recovery.

Cloudera® Hadoop Configuration

The Cloudera Hadoop configuration consists of the HDFS configuration and the MapReduce configuration. The specific configuration parameters that were changed from their defaults are listed in Table 3. All remaining parameters were left at the default values, including the replication factor for HDFS (3).

Configuration parameter	Value	Purpose
dfs.namenode.name.dir	/data1/dfs/nn /data1 is mounted on the RAID5 logical volume on the NameNode.	Determines where on the local file system the NameNode should store the name table (fsimage).
dfs.datanode.data.dir	/data1/dfs/dn, /data2/dfs/dn Note: /data1 and /data2 are mounted on HDDs or SSDs depending on which storage configuration is used.	Comma-delimited list of directories on the local file system where the DataNode stores HDFS block data.
mapred.job.reuse.jvm.num.tasks	-1	Number of tasks to run per Java Virtual Machine (JVM). If set to -1, there is no limit.
mapred.output.compress	Disabled	Compress the output of MapReduce jobs.
MapReduce Child Java Maximum Heap Size	512 MB	The maximum heap size, in bytes, of the Java child process. This number will be formatted and concatenated with the 'base' setting for 'mapred.child.java.opts' to pass to Hadoop.
mapred.tasktracker.map.tasks.maximum	12	The maximum number of map tasks that a TaskTracker can run simultaneously.
mapred.tasktracker.reduce.tasks.maximum	6	The maximum number of reduce tasks that a TaskTracker can run simultaneously.
mapred.local.dir (job tracker)	/data1/mapred/jt, /data1 is mounted on the RAID5 logical volume on the NameNode.	Directory on the local file system where the JobTracker stores job configuration data.
mapred.local.dir (task tracker)	/data1/mapred/local, which is hosted on HDD or SSD depending on which storage configuration is used.	List of directories on the local file system where a TaskTracker stores intermediate data files.

Table 3.: Cloudera Hadoop configuration parameters

Operating System (OS) Configuration

The following configuration changes were made to the Red Hat Enterprise Linux OS parameters.

1. As per Cloudera recommendations, swapping factor on the OS was changed to 20 from the default of 60 to avoid unnecessary swapping on the Hadoop DataNodes. `/etc/sysctl.conf` was also updated with this value.

```
# sysctl -w vm.swappiness=20
```

2. All file systems related to the Hadoop configuration were mounted via `/etc/fstab` with the 'noatime' option as per Cloudera recommendations. With the 'noatime' option, the file access times aren't written back, thus improving performance. For example, for one of the configurations, `/etc/fstab` had the following entries.

```
/dev/sdb1    /data1    ext4    noatime    0 0
/dev/sdc1    /data2    ext4    noatime    0 0
```

3. The open files limit was changed from 1024 to 16384. This required updating `/etc/security/limits.conf` as below,

```
* Soft nofile 16384
* Hard nofile 16384
```

And `/etc/pam.d/system-auth`, `/etc/pam.d/sshd`, `/etc/pam.d/su`, `/etc/pam.d/login` were updated to include:

```
session include system-auth
```

6. Test Validation

Test Methodology

The purpose of this technical paper is to showcase the benefits of using SSDs within a Hadoop environment. To achieve this, SanDisk tested two separate configurations of the Hadoop cluster with the TestDFSIO Hadoop benchmark on a 1TB dataset (with 512 files, each with 2000MB of data). The two configurations are described in detail as follows. Note that there is no change to the NameNode configuration and it remains the same across all configurations.

1. All-HDD configuration

- a. The Hadoop DataNodes use HDDs for the Hadoop distributed file system as well as Hadoop MapReduce.
- b. Each DataNode had two HDDs setup as JBODs. The devices were partitioned with a single partition and formatted as ext4 file systems. These were then mounted in `/etc/fstab` to `/data1` and `/data2` with the noatime option. `/data1` and `/data2` were used within the Hadoop configuration for DataNodes (`dfs.datanode.data.dir`) and `/data1` was used for task trackers directories (`mapred.local.dir`).

2. All-SSD configuration

In this configuration, the HDDs of the first configuration were replaced with SSDs.

- a. Each DataNode had two SSDs setup as JBODs. The devices were partitioned with a single partition with a 4K divisible boundary as follows.

```
[root@hadoop2 ~]# fdisk -S 32 -H 64 /dev/sdd
WARNING: DOS-compatible mode is deprecated. It's strongly recommended to
switch off the mode (command 'c') and change display units to sectors
(command 'u').
Command (m for help): c
DOS Compatibility flag is not set
Command (m for help): u
Changing display/entry units to sectors
Command (m for help): n
Command action
  e  extended
  p  primary partition (1-4)
p
Partition number (1-4): 1
First sector (2048-937703087, default 2048):
Using default value 2048
Last sector, +sectors or +size{K,M,G} (2048-937703087, default
937703087):
Using default value 937703087
Command (m for help): w
The partition table has been altered!
Calling ioctl() to re-read partition table.
Syncing disks.
```

- b. These drives were then formatted as ext4 file systems using mkfs.
- c. They were mounted via /etc/fstab to /data1 and /data2 with the noatime option. /data1 and /data2 were used within the Hadoop configuration for DataNodes (dfs.datanode.data.dir) and /data1 is used for task trackers (mapred.local.dir).

For each of the above configurations, TestDFSIO read/write was executed for a 1TB dataset (with 512 files, each with 2000MB of data) with 100% read and 100% write options. The time taken to run the benchmark was recorded, along with the average throughput seen across the cluster for read/write.

Results Summary

TestDFSIO benchmark runs were conducted on the two configurations described in the Test Methodology section. The runtime for completing the benchmark on a 1TB dataset was collected. The runtime results are summarized in Figure 2. The X-axis on the graph shows the results of the 100% read runs and the 100% write runs — and the Y-axis shows the runtime in seconds. The runtimes are shown for the all-HDD configuration via the gray columns and the all-SSD configurations are shown via the red columns.

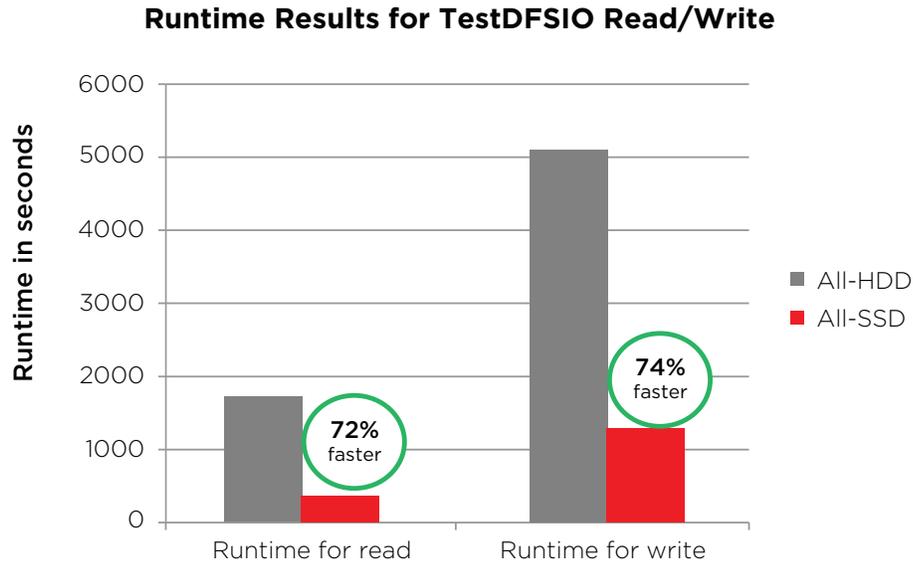


Figure 2.: Runtime comparisons

The average cluster throughput in MB/s (megabytes per second) was also collected for the 100% read benchmarks and the 100% write benchmarks. The throughput results are summarized in Figure 3. The X-axis on the graph shows the 100% read and 100% write runs, and the Y-axis shows the throughput value in terms of MB/s. The throughput for the all-HDD configuration is shown via the gray columns and the all-SSD configuration is shown via the red columns.

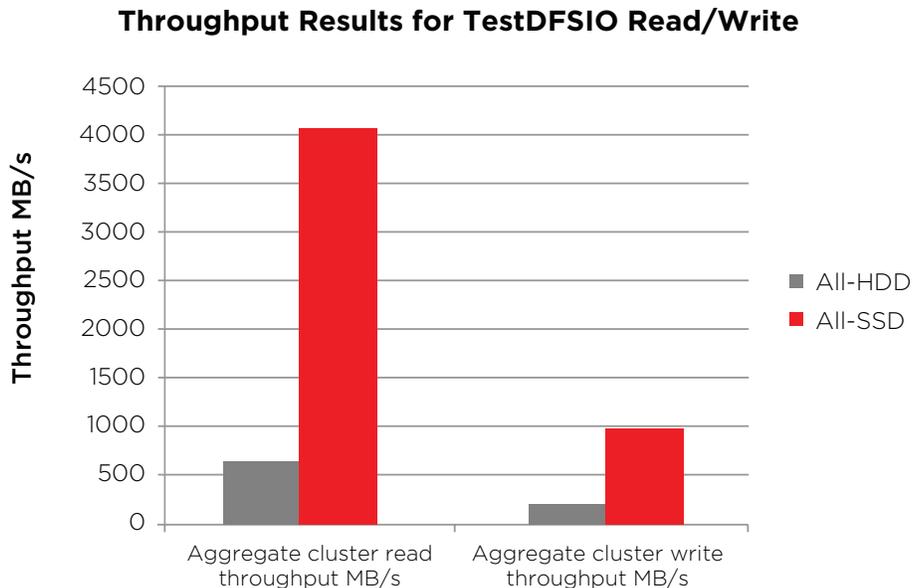


Figure 3.: Throughput comparisons

The results shown (runtime and throughput) above in graphical format are also shown in tabular format in Table 4:

Configuration	1TB read runtime (secs)	1TB write runtime (secs)	1TB read average cluster throughput (MB/s)	1TB write average cluster throughput (MB/s)
All-HDD configuration	1756	5131	722.56	218.36
All-SSD configuration	485	1319	4086.78	952.35

Table 4.: Results summary

7. Results Analysis

Performance Analysis

Observations from the runtime results summarized in Figure 2 and Table 4 are as follows.

This testing showed that replacing all the HDDs on the DataNodes with SSDs can reduce the 1TB TestDFSIO benchmark runtimes for read and write by ~72% and 74% respectively, therefore completing the job faster compared to an all-HDD configuration. The 1TB dataset was comprised of 512 files, each of which had 2000MB of data.

1. There are significant performance benefits when replacing the all-HDD configuration with an all-SSD configuration from SanDisk. SanDisk provides read-intensive CloudSpeed Ascend SATA SSDs 70K IOPS for random read and 14K IOPS for random writes. The higher IOPS translate to faster job completion on the Hadoop cluster, which effectively translates to a much more efficient use of the Hadoop cluster by running more number of jobs in the same period of time.
2. A greater number of jobs on the Hadoop cluster will translate to savings in the TCO of the Hadoop cluster in the long-run (for example over a period of 3-5 years), even if the initial investment may be higher due to the cost of the SSDs. This is discussed further in the next section.

TCO Analysis (Cost/Job Analysis)

Hadoop has been architected for deployment with commodity servers and hardware. So, typically, Hadoop clusters use inexpensive SATA HDDs for local storage on cluster nodes. However, it is important to consider SSDs when planning out a Hadoop environment, especially if the workload is storage-intensive and has a higher proportion of random I/O patterns.

SanDisk SSDs can provide compelling performance benefits, likely at a lower TCO over the lifetime of the infrastructure.

Consider the TestDFSIO performance for the two configurations, in terms of the total number of 1TB read or write I/O jobs that can be completed over three years. This is determined using the total runtime of one job to determine the number of jobs per day ($24 \times 60 \times 60 / \text{runtime_in_seconds}$) and then over three years ($\# \text{jobs per day} \times 3 \times 365$). The total number of jobs for the two configurations over three years is shown in Table 5.

Configuration	Runtime for a single 1TB read job in seconds	#read jobs in 1 day	#read jobs in 3 years	Runtime for a single 1TB write job in seconds	#write jobs in 1 day	#write jobs in 3 years
All-HDD configuration	1756	49.2	53876	5131	16.84	18438
All-SSD configuration	485	178.15	195068	1319	65.50	71727

Table 5.: Number of jobs over 3 years

Also consider the total price for the Test Environment described earlier in this paper. The pricing includes the cost of the one NameNode, six DataNodes, local storage disks on the cluster nodes, and networking equipment (Ethernet switches and the 10 GbE NICs on the cluster nodes). Pricing, as shown in these tables, has been determined via Dell’s Online Configuration tool for rack servers and Dell’s pricing for accessories, as listed in Table 6 below.

Configuration	Discounted Pricing from http://www.dell.com
All-HDD configuration	\$37,561
All-SSD configuration	\$46,567

Table 6.: Pricing the configurations

Now consider the cost of the environment per job when the environment is used over three years (cost of configuration / number of jobs in three years).

Configuration	Discounted Pricing from www.dell.com	# read-intensive jobs in 3 years	\$/ read-intensive job	# write-intensive jobs in 3 years	\$/ write-intensive job
All-HDD configuration	\$37,561	53876	\$0.69	18438	\$2.03
All-SSD configuration	\$46,567	195068	\$0.23	71727	\$0.64

Table 7.: \$ Cost/Job

Table 7 shows the cost per job (\$/job) across the two Hadoop configurations and how the SanDisk all-SSD configuration compares with the all-HDD configuration. These results are graphically summarized in Figure 4 and Figure 5 below.

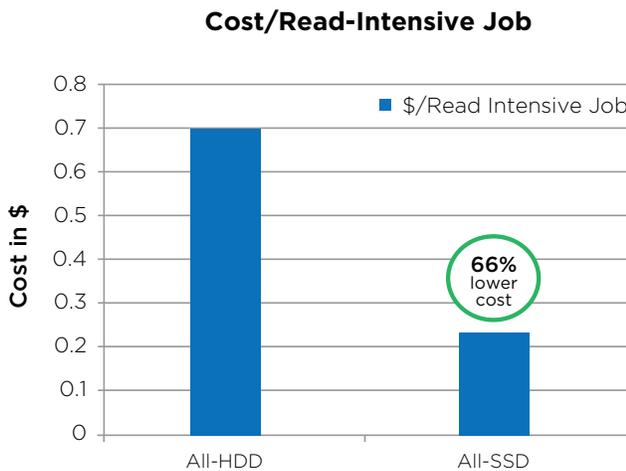


Figure 4.: \$/job for read-intensive workloads

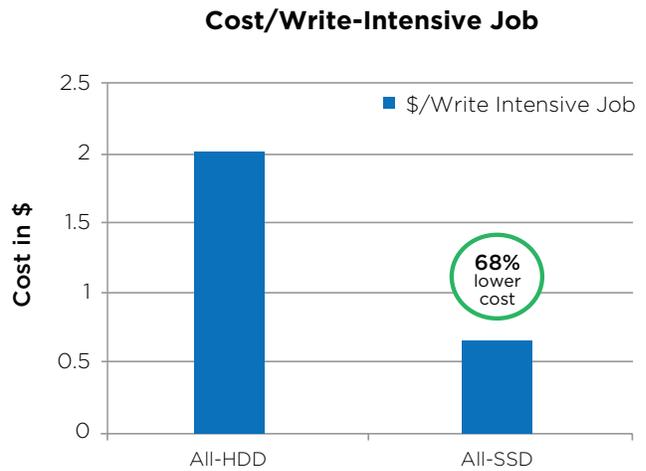


Figure 5.: \$/job for write-intensive workloads

Observations from these analysis results:

1. SanDisk's all-SSD configuration **reduces the cost/job by more than 60%** when compared to the all-HDD configuration for both read-intensive and write-intensive workloads.
2. As a result, over the lifetime of the infrastructure, the **TCO is significantly lower** for the SanDisk SSD Hadoop configurations when considering the total number of completed jobs.
3. SanDisk SSDs also have a much lower power consumption profile than HDDs. As a result, the all-SSD configuration will likely see the TCO reduced even further, due to reduced power use.

8. Conclusions

SanDisk SSDs can be deployed strategically in Hadoop environments to provide significant performance improvement and TCO benefits. Performance was over 70% for the TestDFSIO benchmark³ and TCO benefits include over a 60% reduction in \$/job for the TestDFSIO benchmark⁴.

SanDisk SSDs are beneficial for Hadoop environments, especially when the workloads are storage-intensive. This technical paper discussed one such deployment of CloudSpeed Ascend SATA SSDs within Hadoop and provided a proof-point for the SSD performance and SSD TCO benefits within the Hadoop ecosystem.

It is important to understand that Hadoop applications and workloads vary significantly in terms of their I/O access patterns. Therefore all Hadoop workloads may not benefit equally from the use of SSDs. It is necessary for customers to develop a proof-of-concept (POC) for their custom Hadoop applications and workloads, so that they can evaluate the benefits that SSDs can provide to those applications and workloads.

9. References

1. SanDisk website: www.sandisk.com/
2. Apache Hadoop: <http://hadoop.apache.org/>
3. Cloudera Distribution of Hadoop: www.cloudera.com/content/cloudera/en/home.html
4. Dell Configuration and pricing tools
 - a. www.dell.com/us/business/p/poweredge-r320/pd
 - b. www.dell.com/us/business/p/poweredge-r720/pd
5. Intel HiBench Benchmark suite: <https://github.com/intel-hadoop/HiBench>

³Please see Figure 2: Runtime Comparisons, page 11

⁴Please see Figure 4: \$/Job for Read-Intensive Workloads and Figure 5: \$/Job for Write-Intensive Workloads, page 13

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. CloudSpeed and CloudSpeed Ascend are trademarks 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). Faster_Hadoop_Performance - 07.01.2016

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