



WHITE PAPER

# Accelerating Cassandra Workloads using SanDisk® Solid State Drives

February 2015

**SanDisk®**  
a Western Digital brand

Western Digital Technologies, Inc.  
951 SanDisk Drive, Milpitas, CA 95035

[www.SanDisk.com](http://www.SanDisk.com)

**Table of Contents**

Executive Summary .....3

Overview: Apache Cassandra .....3

Why Cassandra for Web-Based Data Workloads? ..... 3 YCSB

Testing Tool .....4

Test Design .....4

    Test Environment .....4

    Technical Component Specifications ..... 4

    Compute Infrastructure .....5

    Network Infrastructure .....5

    Storage Infrastructure .....5

    Cassandra Configuration .....5

Test Validation .....5

    Test Methodology .....5

    Results Summary ..... 6

    Transactions Per Second (TPS) Comparisons ..... 7

    Latency Comparisons .....9

    Results Analysis and Conclusion .....10

Summary .....11

References .....11

### Executive Summary

In today's scale-out data centers the ability to support high data-transfer rates is absolutely important to successful cloud-computing workloads that show "peaks and valleys", requiring NoSQL databases to perform well under demanding conditions. Hyperscale data centers see these workload patterns all the time, and they need consistent performance to ensure quality of service for customers accessing cloud services .

In cloud computing, many of the data requests are for random reads and writes, something that is difficult for mechanically driven hard-disk drives (HDDs) to fulfill .That's why solid-state drives (SSDs) from SanDisk show a significant performance improvements over HDDs when running mixed workloads (mixing reads and writes) that are I/O-intensive workloads . Using the YCSB cloud-computing benchmark, this performance improvement translates directly into higher throughput and lower latency for cloud-computing data centers .

This technical paper describes workload testing with 64GB, 256GB and 1TB datasets conducted on a single-node Cassandra system, using SanDisk SSDs and HDDs . The primary goal of this paper is to show the performance benefits of using SanDisk SSDs within a Cassandra environment . Testing for both Uniform- and Zipfian-style data distributions were including in the YCSB benchmarking protocols . See more details of this testing on the SanDisk website at [www.sandisk.com](http://www.sandisk.com) .

The testing tool used is the Yahoo! Cloud Serving Benchmark (YCSB) framework, with the goal of facilitating performance comparisons of the different types of workloads . For additional information on YCSB, refer to the References section at the end of this paper .

### Overview: Apache Cassandra

Apache Cassandra is a highly scalable, eventually consistent, distributed, structured key-value data-store . Cassandra brings together the distributed systems technologies from Amazon Dynamo and the data model from Google's Big Table technology . Like Dynamo, Cassandra is eventually consistent . Like Big Table, Cassandra provides a column family-based data model (columnar data model) that is richer than typical key/value systems, and typically faster than traditional row-based SQL database systems .

### Why Cassandra for Web-Based Data Workloads?

We should note here that Cassandra's data model offers the convenience of column indexes with the performance of log-structured updates; strong support for denormalization and materialized views; and powerful built-in caching for workload optimization .

Cassandra is a NoSQL column family implementation supporting the BigTable data model, and using the architectural aspects introduced by Amazon Dynamo . Some of the main features of Cassandra are that it:

- 1 . Provides a highly scalable and highly available software platform, with no single point of failure
- 2 . Has a NoSQL column family implementation
- 3 . Has a very high write throughput and good read throughput
- 4 . Supports a SQL-like query language and supports search through secondary indexes
- 5 . Has tunable consistency and support for replication
- 6 . Has flexible schema

### YCSB Testing Tool

YCSB consists of two components:

1. The client, which generates the load according to a workload type, and analyzes latency and throughput
2. Workload files, which define a single benchmark by describing the size of the dataset, the total amount of requests, and the ratio of read and write queries

There are six major workload types in YCSB:

1. Workload A, 50/50 update/read ratio, size of the dataset is 200 000 key/value pairs
2. Workload B, 5/95 update/read ratio, the same size of the dataset
3. Workload C, 100% read-only
4. Workload D, 5/95 insert/read ratio, the read load is skewed towards the end of the key range
5. Workload E, 5/95 ratio of insert/reads over a range of 10 records
6. Workload F, 95% read/modify/write, 5% read

This technical paper only presents the first three types of workloads that are described in the YCSB portfolio: Workloads A, B and C.

### Test Design

A standard Cassandra database was set up for the purpose of determining the benefits of using SSDs within a Cassandra environment, focusing on the YCSB benchmark. The testing consisted of using different YCSB workload types. These are labelled as Workloads A, B and C, with dataset sizes scaling from 64GB to 256GB to 1TB. Results for all three database sizes are provided in this paper.

The performance of each system that was tested was plotted to see which systems performed best. The performance was measured by comparing latency versus throughput for each of these workloads, results of which were summarized and analyzed.

Finally, this paper provides recommendations, based on this testing, for using SSDs within a hardware/software configuration supporting Cassandra workloads.

### Test Environment

The test environment consisted of one Dell PowerEdge R720 with 24 Intel® Xeon® processor cores and 94.4GB DRAM hosting the Cassandra database, and one Dell PowerEdge R720 that serves as a client using YCSB. A 1GbE network interconnect was used to link the server and the client. The dataset size of the YCSB tests was configured at 64GB, 256GB and 1TB.

**Technical Component Specifications**

Hardware	Software if applicable	Purpose	Quantity
Dell Inc. PowerEdge R720 • CPU and OS both 64-bit • 24 Intel® Xeon® CPU E5-2620 0 @ 2GHz • 94GB memory	• Linux • CentOS 5.10 • Cassandra 1.2.2	Server	1
Dell Inc. PowerEdge R720 • CPU and OS both 64-bit • 24 Intel Xeon CPU E5-2620 0 @ 2GHz • 94GB memory	• Linux • CentOS 5.10 • YCSB 0.1.4	Client	1
Dell PowerConnect 2824 24-port switch	1GbE network switch	Mgmt network	1
500GB 7.2K RPM Dell SATA HDDs	Used as JBODs	Data node drives	6
480GB CloudSpeed® Ascend SATA SSDs	Used as JBODs	Data node drives	6

**Figure 1: Hardware components**

Software	Version	Purpose
CentOS Linux	5.10	Operating system for server and client
Apache Cassandra	1.2.2	Database server
YCSB	0.1.4	Client test tool

**Figure 2: Software components**

**Compute Infrastructure**

The server was a Dell PowerEdge R720 with 24 Intel Xeon processor cores with a 2GHz E5-2620 CPU and 94GB of memory . The client’s compute infrastructure is the same as that of the server .

**Network Infrastructure**

The client and the server are connected to a 1GbE management network via the onboard 1GbE NIC .

**Storage Infrastructure**

The server had 94GB of DRAM and used 650GB 7.2K RPM Dell SATA HDDs running in a RAID0 configuration for the HDD tests. The HDDs were then replaced by six 480GB CloudSpeed Ascend SATA SSDs for the SSD tests .

**Cassandra Configuration**

Cassandra default configurations were used during the test, and all the workloads were running on the same Cassandra configurations .

## Test Validation

### Test Methodology

The primary goal of this technical paper is to showcase the benefits of using SSDs within a Cassandra production environment. To achieve this goal, SanDisk tested 36 separate configurations of single-node Cassandra, using the standard YCSB benchmark workloads A, B and C and running that benchmark for three different dataset sizes.

The difference in the YCSB workload results for SSDs and HDDs was dramatic, and varied according to the percentage of reads and writes in the "mixed workload."

The YCSB workload types were as follows (according to the percent of read and write operations being performed in each workload):

1. Workload A: 50/50 update/read ratio, where the size of the dataset is 200K key/value pairs
2. Workload B: A 5%/95% update/read ratio, with the same size of the dataset being tested, as above
3. Workload C: 100% read-only

The two storage media described in detail were as follows:

1. HDD configuration  
The server node uses HDDs for the single-node Cassandra deployment.
2. SSD configuration  
In this configuration, the HDDs of the first configuration were swapped with SSDs.

The dataset types were as follows:

1. 64GB: This tests the case of the dataset that is smaller than the amount memory in the system; in addition, all of the data can be stored in the memory.
2. 256GB: This test the case in which the amount of data exceeds that of the local memory, out of memory case 1.
3. 1TB: This tests the case in which the amount of memory exceeds that of local memory, out of memory case 2.

The YCSB workload types were as follows (according to the operation distribution):

1. Uniform: In which each data has the same frequency to be operated, and thus has equal access to the processor.
2. Zipfian: In which part of the data is operated upon more frequently than other data, meaning that access to the processor is unequal)

Combining all these different types, a total of 36 (3\*2\*3\*2) configurations were tested.

The YCSB workload data schema was as follows:

	field0	field1	...	field9
Rowkey	val0	val1	...	val9
	timestamp0	timestamp1	...	timestamp9

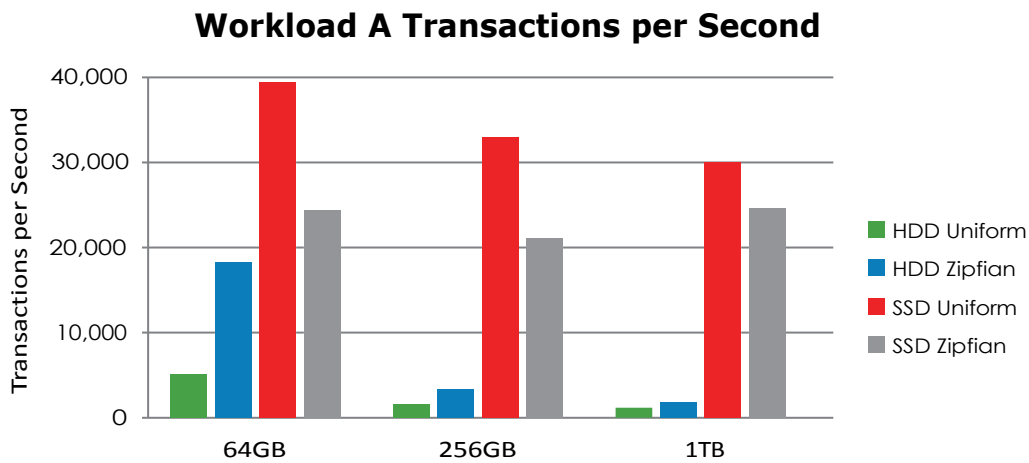
**Results Summary**

YCSB benchmark runs were conducted on a total of 36 hardware/software configurations, as mentioned in the previous section . The throughput and latency of different dataset sizes and workload types on HDD and SSD were collected and analyzed, to see how they compared .

The throughput results are summarized in figures 3, 5 and 7, with respect to the transactions per second (TPS) comparisons .

In these figures, the X-axis on the graph shows the different configurations, and the Y-axis shows the TPS . The latency results are summarized in figures 9, 11 and 13 . The X-axis on the graph shows the different configurations, and the Y-axis shows the latency . The TPS and latency are shown for HDD Uniform (green columns), HDD Zipfian (blue columns), SSD Uniform (red columns) and SSD Zipfian for the entire run (gray columns) .

**Transactions Per Second (TPS) Comparisons**



**Figure 3: TPS comparisons of Workload A**

YCSB Workload Types	Drive Type	YCSB Workload Types	64GB	256GB	1TB
Workload A (50r/50w)	HDD	Uniform	5,213	1,705	1,253
Workload A (50r/50w)	HDD	Zipfian	18,380	3,448	1,916
Workload A (50r/50w)	SSD	Uniform	39,528	32,970	29,975
Workload A (50r/50w)	SSD	Zipfian	24,554	21,104	24,700

**Figure 4: TPS results summary of Workload A**

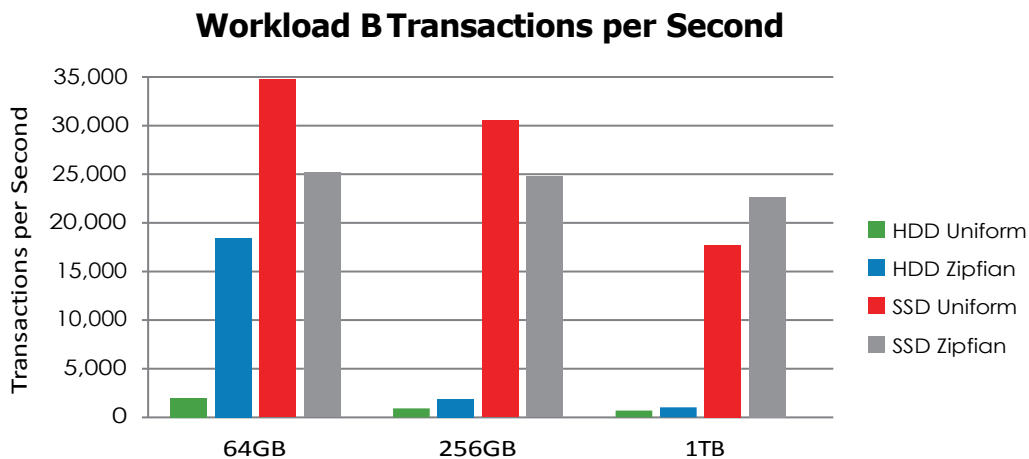


Figure 5: TPS comparisons of Workload B

YCSB Workload Types	Drive Type	YCSB Workload Types	64GB	256GB	1TB
Workload B (95r/5w)	HDD	Uniform	2,036	886	673
Workload B (95r/5w)	HDD	Zipfian	18,417	1,859	992
Workload B (95r/5w)	SSD	Uniform	34,678	30,469	17,673
Workload B (95r/5w)	SSD	Zipfian	25,108	24,734	22,677

Figure 6: TPS results summary of Workload B

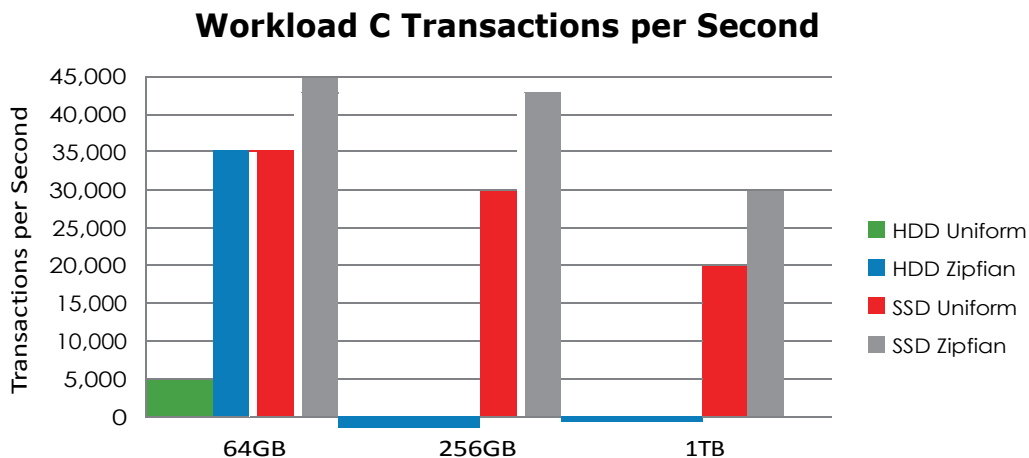


Figure 7: TPS comparisons of Workload C

YCSB Workload Types	Drive Type	YCSB Workload Types	64GB	256GB	1TB
Workload C (100r/0w)	HDD	Uniform	1,922	842	640
Workload C (100r/0w)	HDD	Zipfian	36,305	1,733	941
Workload C (100r/0w)	SSD	Uniform	36,227	29,634	16,636
Workload C (100r/0w)	SSD	Zipfian	43,410	42,680	29,902

Figure 8: TPS results summary of Workload C



Latency Comparisons

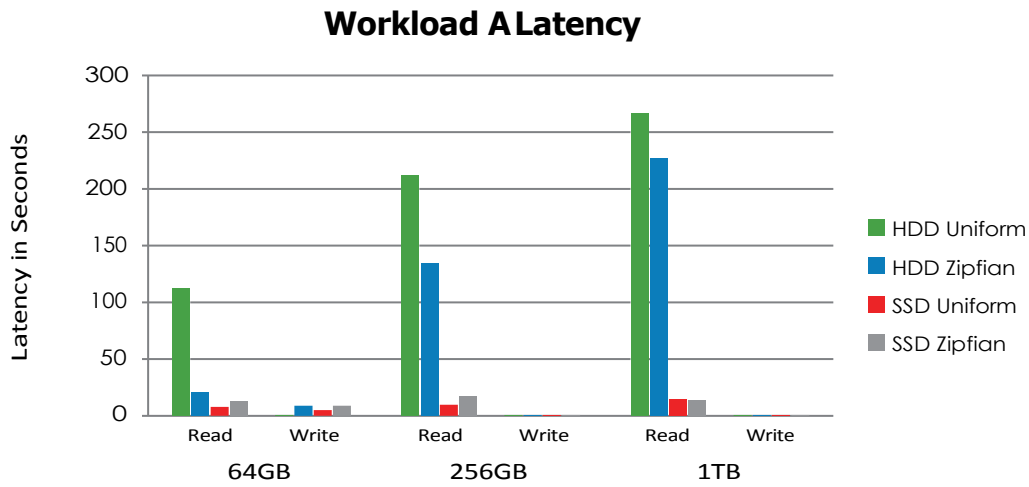


Figure 9: Latency comparisons of Workload C

YCSB Workload Types	Drive Type	YCSB Workload Types	64GB		256GB		1TB	
			Read	Write	Read	Write	Read	Write
Workload A (50r/50w)	HDD	Uniform	112	0	212	0	266	0
Workload A (50r/50w)	HDD	Zipfian	21	9	134	0	227	0
Workload A (50r/50w)	SSD	Uniform	8	5	10	0	15	0
Workload A (50r/50w)	SSD	Zipfian	13	9	17	0	14	0

Figure 10: Latency Results summary of Workload A

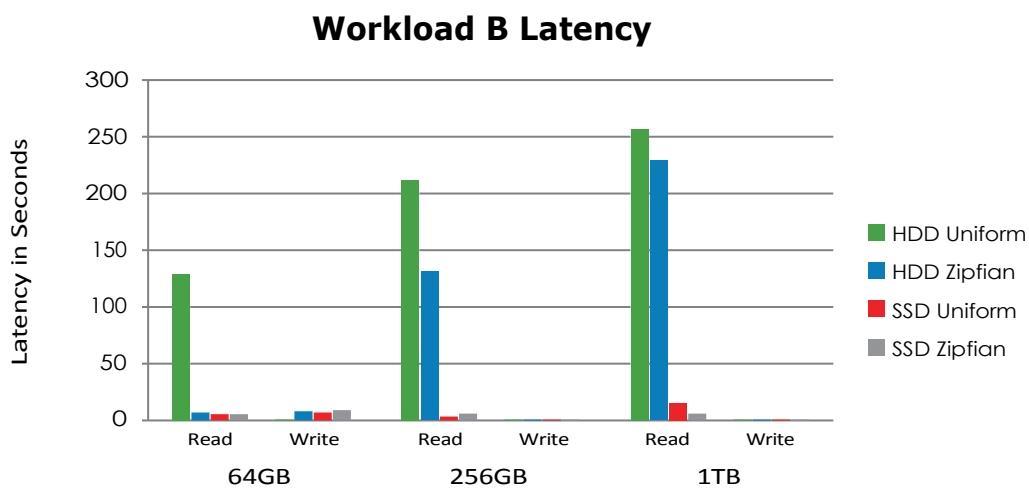


Figure 11: Latency comparisons of Workload B

YCSB Workload Types	Drive Type	YCSB Workload Types	64GB		256GB		1TB	
			Read	Write	Read	Write	Read	Write
Workload B (50r/50w)	HDD	Uniform	131	0	214	0	259	0
Workload B (50r/50w)	HDD	Zipfian	9	10	133	0	231	0
Workload B (50r/50w)	SSD	Uniform	7	9	5	0	17	0
Workload B (50r/50w)	SSD	Zipfian	7	11	8	0	8	0

Figure 12: Latency Results summary of Workload A

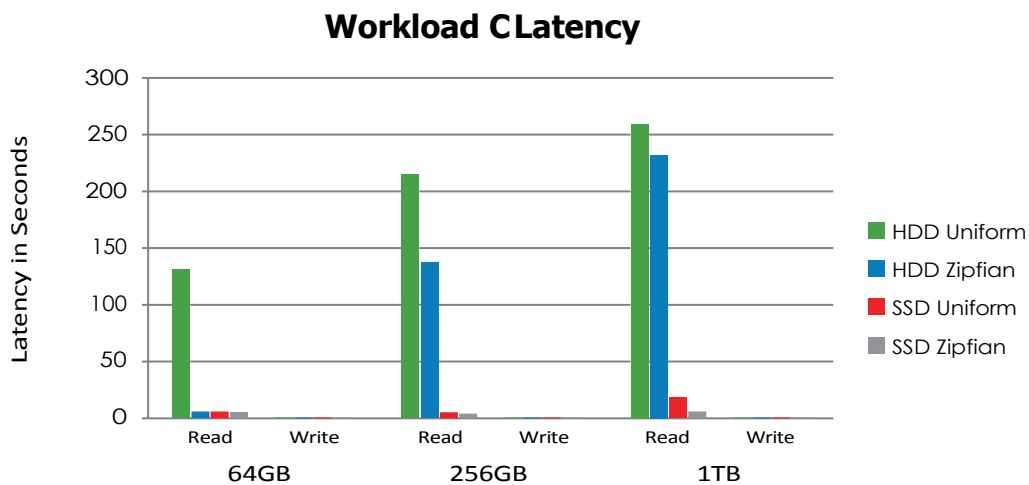


Figure 13: Latency comparisons of Workload C

YCSB Workload Types	Drive Type	YCSB Workload Types	64GB		256GB		1TB	
			Read	Write	Read	Write	Read	Write
Workload C (50r/50w)	HDD	Uniform	131	0	214	0	259	0
Workload C (50r/50w)	HDD	Zipfian	6	0	137	0	231	0
Workload C (50r/50w)	SSD	Uniform	6	0	5	0	18	0
Workload C (50r/50w)	SSD	Zipfian	5	0	4	0	6	0

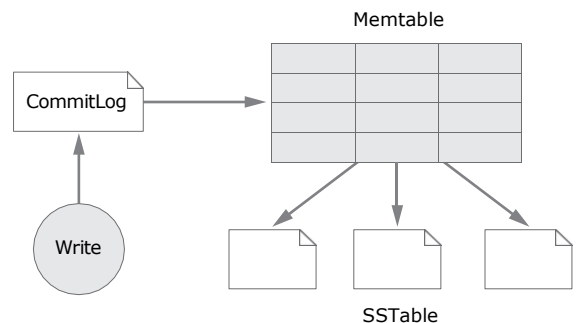
Figure 14: Latency Results summary of Workload C

Results Analysis and Conclusion

Cassandra storage architecture

Write workflow:

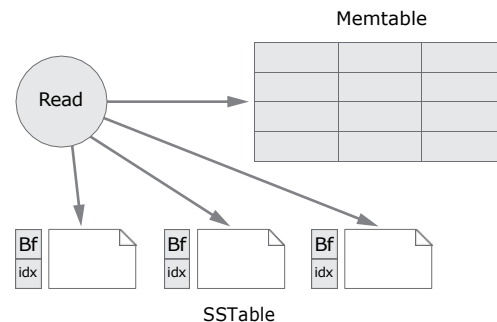
1. Write commit log
2. Write data into memtable
3. Flush memtable to sstable (on disk)



In the Cassandra write scenario, an “update” will update two kinds of files: commitlog and sstable . Both of these types of updates are labelled as “sequential write” .

Read workflow:

- 1 . If data was in memtable (in memory), return
- 2 . Check whether data was in one sstable or not, use “Bloom Filter” algorithm
- 3 . If data was in one sstable, find the data offset by using 'idx' file .
- 4 . Read data from sstable, return



In the Cassandra read scenario, most read requests are random read . As shown, SSDs have a significant advantage over HDDs on the random read .

From the results summary in the previous section, the following observations can be made:

1. SSDs support better TPS rates. SSD-enabled systems supported TPS rates that were up to 80 times higher than the TPS rates for workloads that run on HDD-enabled systems. We should also note that the latency of workloads that use HDDs is 20 times longer than the latency associated with SSDs.
2. Regarding memory, when the Cassandra workload set size is smaller than the memory size, then it's possible to show that the workload has better throughput per second (TPS) rates, and better (smaller) latency than when the workload set size is larger than the memory size. This means that as workloads grow larger in size, they tend to show lower TPS, and longer latency.

## Summary

Based on these observations, the following conclusions can be made:

The performance of Cassandra workloads that use SSDs is much better than for workloads that use HDDs, in terms of higher TPS and less latency . Those results clearly show that SSDs can provide a significant performance improvement, compared with traditional HDDs, when running I/O-intensive workloads, especially when running mixed workloads having random read and write data accesses .

For customers running data-intensive workloads in cloud computing data centers, or hyperscale data centers, these findings present a clear message: Leveraging SSDs improves performance, and reduces latency, for Cassandra workloads. The YCSB benchmark is well-known, and used across the industry – making these findings even more impactful for customers considering SSDs for their data center infrastructure .

## References

SanDisk website: [www.sandisk.com](http://www.sandisk.com) Apache

Cassandra: <http://cassandra.apache.org/>

Yahoo LABS: [http://research.yahoo.com/Web\\_Information\\_Management/YCSB/](http://research.yahoo.com/Web_Information_Management/YCSB/)

Specifications are subject to change. ©2015 - 2016 Western Digital Corporation or its affiliates. All rights reserved. SanDisk and the SanDisk logo are trademarks of Western Digital Corporation or its affiliates, registered in the U.S. 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 respective holder(s).  
20160623

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