Cisco UCS Powered by Fusion ioMemory™ Storage Delivers Leading Hadoop Performance

www.SanDisk.com

Table of Contents

Executive Summary

This document describes the performance and scalability benefits of a high performance Hadoop cluster deployment. This deployment uses the Cisco Unified Computing System™ (Cisco UCS®) blade server, UCS fabric interconnect, and UCS Storage Accelerator devices powered by Fusion ioMemory technology from SanDisk®, running the Cloudera Distribution of Apache Hadoop. This combined stack provides a faster time to analytics, with millisecond latency, while offering an unmatched performance advantage. This solution helps to maximize performance while avoiding over-provisioning of hardware resources, which enables optimized deployment of Big Data applications.

Here are some of the competitive advantages of this solution:

- *Unleashing Extreme Hadoop Performance*: Cisco Unified Computing System B200 blade servers with Cisco Storage Accelerators deliver extreme performance for Hadoop-based Big Data applications.

- *Providing Smart Infrastructure Management*: Cisco UCS Manager provides faster resource allocation and this helps in easier, faster Hadoop cluster scaling. Automated server and network policies help to seamlessly scale the Hadoop cluster and lower scaling costs.

- *Enabling High Availability and Higher Network Throughput*: The Cisco UCS fabric interconnect provides lossless 40G Ethernet traffic when clustered within a chassis, delivering both high availability and high performance for Hadoop cluster deployments.

- *Ensuring a Faster Time to Analytics, with Millisecond Latency*: Cisco Storage Accelerator devices deliver the millisecond latency that Hadoop applications need to maintain real-time response when processing tens of terabytes of data.

Big Data Adoption

Big Data – the analysis of massive quantities of data to gain new business insights – has become a new competitive advantage for companies and will be fundamental to business growth and expansion. Big Data adoption is becoming increasingly important across most industries. Retail and healthcare are two prominent industries reaping the benefits of its deployment: retail employs selective ad promotion, while healthcare integrates information from various sources (sensors, X-rays, handwriting, and other medical images) and delivers relevant information in a shorter time, for better patient outcomes. Financial services, communications media, insurance, transportation, and manufacturing are other industries that are capitalizing on the benefits of Big Data.

Big Data Challenges

As various industries adopt Big Data for enterprise-wide solutions, multiple challenges arise:

- Determining how to get the most business value

- Integrating Big Data technology with existing infrastructure

- Keeping the cost of technology infrastructure low (hardware economics)

- Using a shared infrastructure, being scalable without downtime, and offering fault tolerance, all with lower cost (economics of scale)

Solution Stack Advantages

Integrating Big Data solutions with existing infrastructure is an important need. Customers using Cisco UCS infrastructure for relational databases such as Oracle and SQL Server will find it relatively easy to integrate Big Data applications into a solution stack, using Cisco UCS servers for seamless integration and deployment. Below are some of the key advantages of this solution stack.

- Cisco UCS BIOS policies can be customized and automated for large-scale and rapid Hadoop deployment. This saves significant time and effort and improves operational efficiency.

- The Cisco UCS blade server with Cisco Storage Accelerators saves significant rack space, thereby increasing Hadoop cluster density.

- The solution provides savings in power and cooling costs, compared to traditional rack servers with many hard disk drives.

- A fault-tolerant Cisco hardware stack, improved operation efficiency for Hadoop deployment, rack space savings, and reduced power and cooling requirements all reduces the customer total cost of ownership (TCO).
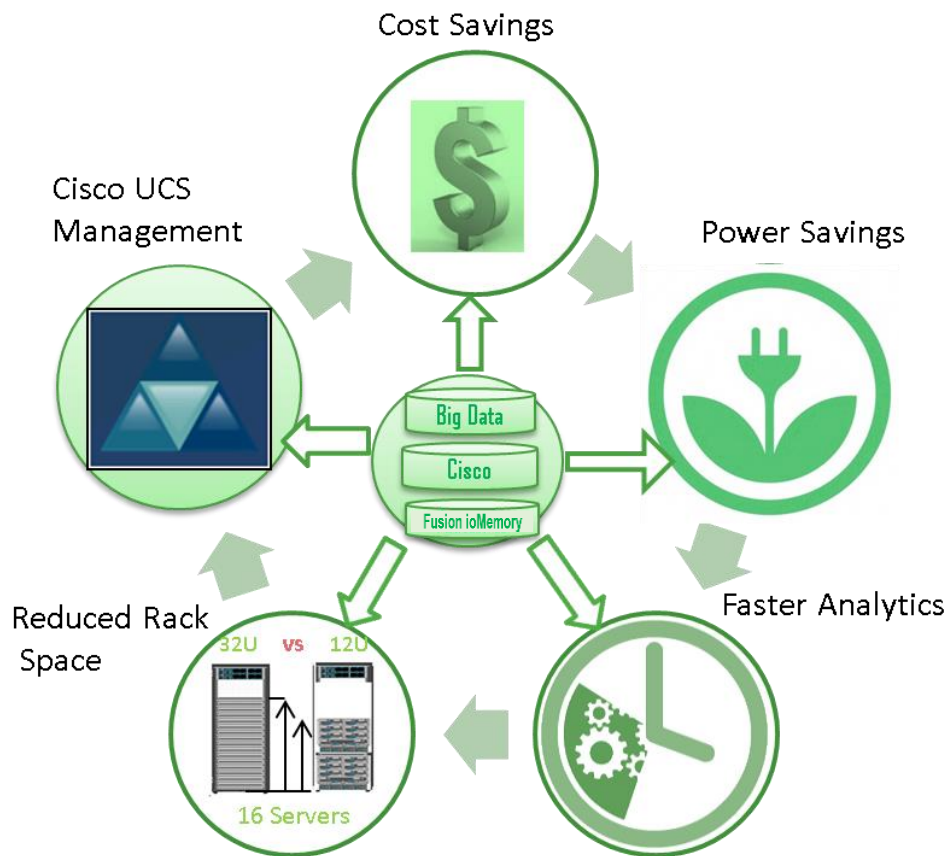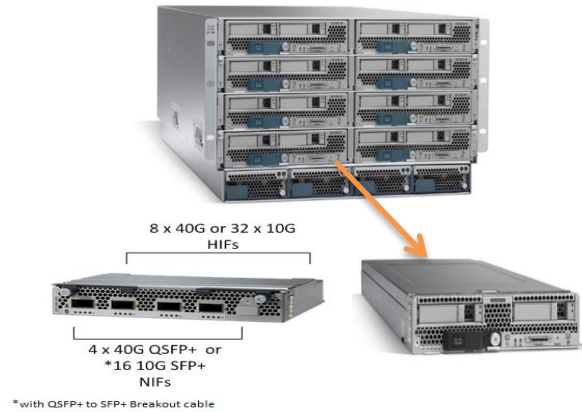
These advantages are illustrated in the figure below.

Figure 1: Competitive advantages of the combined solution

Cisco UCS 5108 Blade Server Chassis

The Cisco UCS 5108 blade server chassis is a 6RU model based on the Intel® Xeon® processor E5 v4 family. It can accommodate up to 8 half-width blades or 4 full-width blades. This chassis provides a single, highly available management domain for all systems. An automated service profile configuration reduces administrative tasks, and a unified fabric helps decrease TCO by reducing the number of network interface cards (NICs), host bus adapters (HBAs), switches, and cables needed. The high-performance chassis mid-plane supports up to two 40 Gb Ethernet links to each half-width blade slot, or up to four 40 Gb links to each full-width slot. This provides 8 blades with 1.2 terabits (Tb) of available Ethernet throughput for future I/O requirements.

8 x 40G or 32 x 10G HIFs

4 x 40G QSFP+ or *16 10G SFP+ NIFs

*with QSFP+ to SFP+ Breakout cable

The chassis does not need switches, which avoids the complex configuration and management typical for switches. This allows a system to scale without unnecessary complexity and cost. The chassis comes with redundant, hot-swappable power supplies and fans, providing high availability in multiple configurations and uninterrupted service during maintenance.

Cisco UCS B200 M4 Blade Server

Delivering performance, versatility, and density without compromise, the Cisco UCS B200 M4 blade server addresses the broadest set of workloads, from IT and web infrastructure through distributed database. The enterprise-class Cisco UCS B200 M4 blade server extends the capabilities of Cisco's Unified Computing System portfolio in a half-width blade form factor. The Cisco UCS B200 M4 harnesses the power of the latest Intel Xeon E5-2600 v3 and v4 Series processor family of CPUs. It features up to 1536 GB of RAM (using 64 GB DIMMs), two solid-state drives (SSDs) or hard disk drives (HDDs), and up to 80 Gbps throughput connectivity.

Cisco developed the 1200 Series and 1300 Series Virtual Interface Cards (VICs) to provide the flexibility to create multiple NIC and HBA devices. The VICs also support Fabric Extender and Virtual Machine Fabric Extender technologies for adapters. It has two Converged Network Adapter (CNA) ports, supporting both Ethernet and FCoE delivers 80 Gbps total I/O throughput to the server. It can create up to 256 fully functional unique and independent PCIe adapters and interfaces (NICs or HBAs) without requiring single-root I/O virtualization (SR-IOV) support from operating systems or hyper-visors.

Cisco UCS Storage Accelerator adapters are designed specifically for the Cisco UCS B Series M4 blade servers and integrate seamlessly to allow improvement in performance and relief of I/O bottlenecks.

Cisco UCS 6300 Fabric Interconnect

The Cisco UCS 6200 and 6300 Series Fabric Interconnects are a core part of the Cisco UCS and provide the management and communication backbone for Cisco UCS B-Series Blade Servers, the UCS 5100 Series Blade Server Chassis, UCS C-Series Rack Servers, and the UCS Mini.

The Cisco UCS 6300 series Fabric interconnects offer high-performance ports capable of

- Line-rate, low-latency, lossless 1-, 10-, and 40-Gigabit Ethernet (varies by model)
- Fibre Channel over Ethernet (FCoE)
- 16-, 8-, and 4-Gbps and 8-, 4-, and 2-Gbps Fibre Channel.

Specifications for the Cisco UCS FI 6332UP (32-Port Fabric Interconnect) and Cisco UCS FI 6332-16UP (40 Port Fabric Interconnects) models are shown below.

Cisco UCS FI 6332UP

- 32 x 40 GbE QSFP+ ports
- 2.56 Tbps switching performance
- 1 RU fixed form factor, two power supplies, four fans
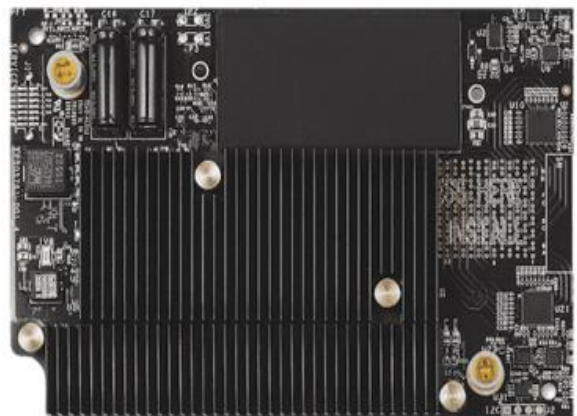
Cisco UCS FI 6332-16UP

- 24 x 40 GbE QSFP+, 16 UP ports (1/10 GbE or 4/8/16 Gb FC)
- 2.43 Tbps switching performance
- 1 RU Fixed form factor, two power supplies, four fans

Cisco Storage Accelerators (UCSB-F-FIO-1300MP)

The Cisco Storage Accelerators are designed to provide performance-driven application environments with ultra-low latency, superior reliability, and maximum business value. Cisco Storage Accelerator devices in Cisco UCS B200 blade servers reduce infrastructure footprint as well as power and cooling costs, thereby lowering the TCO. These devices are available in capacities from 1.3 TB up to 1.6 TB. They offer ultra-low 92µs/15µs read/write data access latency, superior reliability with an UBER of 10^{20}, outstanding random read/write performance of up to 235K/375K IOPS, and sequential read/write speed of up to 2.7/1.7 GB/s.

Additional details on Cisco Storage Accelerators can be found in this data sheet:

https://www.sandisk.com/content/dam/sandisk-main/en_us/assets/resources/enterprise/data-sheets/cisco-ucs-fusion-iomemory-pcie-flash-adapters-for-b-series-blade-servers-datasheet.pdf

Performance Validation

To evaluate performance of the Hadoop cluster solution, we carried out various performance benchmarks, starting with basic Flexible I/O (FIO), Hadoop Distributed File System I/O (DFSIO), and an application workload based on the industry-standard benchmark TPC Express Benchmark HS (TPCx-HS)[1]. The goal was to assess the various performance scenarios and generate a ready reference of performance data points for customer deployments. These data points help shorten the customer evaluation and deployment cycle.

fio Read and Write Throughput Charts

It's important to validate raw disk drive performance before installing the software applications. Various disk performance assessment tools are available, such as fio and Iometer; for our testing purposes we chose the FIO tool. A text-based CLI testing tool, fio provides the flexibility of measuring I/O for random and sequential read, write, and mixed workloads. Because Hadoop jobs tend to execute a high percentage of large-block sequential writes and reads, the fio script is designed to test similar large block sequential workloads.

The following fio script evaluates the raw disk performance for a single server, with a sequential write/read workload. This script was invoked concurrently against all eight servers to measure aggregate performance.

```
# fio --name=writebw --filename=/data/disk1/fio_writetest -size=1024M --direct=1 --
rw=write --bs=512m --numjobs=4 --iodepth=16 --direct=1 --runtime=300 --ramp_time=5 --
time_based --ioengine=libaio --group_reporting > Hadoop_seqwrite_fio_test-Cisco-Fusion-
512M-block.out


# fio --name=readbw --filename=/data/disk1/fio_readtest -size=1024M --direct=1 --rw=read
--bs=512m --numjobs=4 --iodepth=16 --direct=1 --runtime=300 --ramp_time=5 --time_based -
-ioengine=libaio --group_reporting > Hadoop_seqread_fio_test-Cisco-Fusion-512M-block.out
```

The chart below shows the outcome of the sequential reads and writes workloads. (All eight Cisco B200 blade servers were loaded with Cisco Storage Accelerators.) The chart emphasizes the following points:

- The fio script output shows uniform performance throughput, whether the script was invoked individually on each server or concurrently on all servers.

- On average, each server delivered sequential read throughput of 2.8 GB/s and sequential write throughput of 1.8 GB/s.

---

[1] Workload based on TPCx-HS but not audited or published. No comparisons were made with published TPCx-HS results.

512MB Block Size IO Throughput Chart

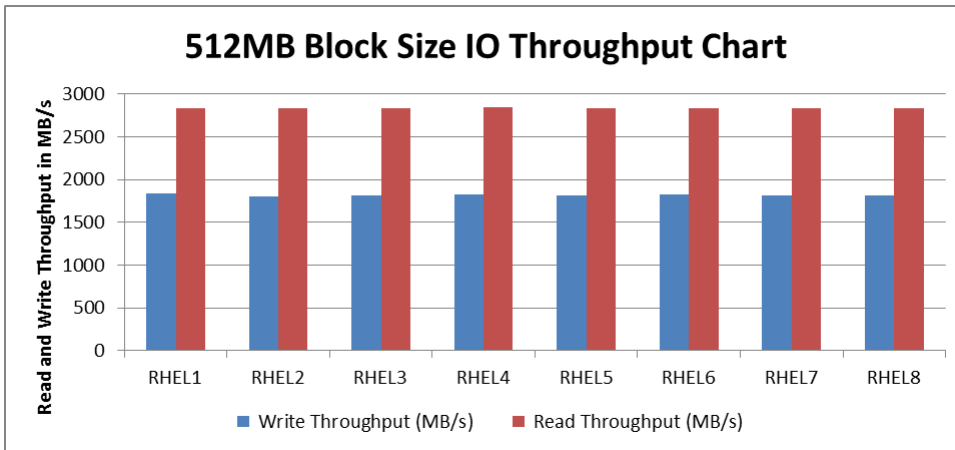Figure 2: 512 MB block size I/O throughput

DFSIO Write Throughput Charts

TestDFSIO is a distributed filesystem test for HDFS (Hadoop Distributed File System) that evaluates Hadoop cluster throughput performance. The test measures HDFS I/O for write and read throughput.

The TestDFSIO write benchmark generates a write-intensive workload by creating a large number of files. The test benchmark involved creating a 1 TB dataset with 192 files, each file with 5.4 GB of storage. The number of files equals the number of map jobs created in the cluster, and the resource manager distributes these 192 jobs equally to all seven data nodes of the cluster. The Hadoop replication factor was configured at the default of three, so with three-way replication the total dataset generated was 3 TB. The total time needed to generate the 1 TB of test data and replicate it three times was just under 300 seconds.

The chart below shows the TESTDFSIO write performance of the solution stack, with seven data nodes generating an average of 10.15 GB/s write throughput and 13.7 GB/s of average network I/O throughput. The Cisco fabric interconnect provides excellent network throughput performance as it replicates the data to all data nodes in the cluster.

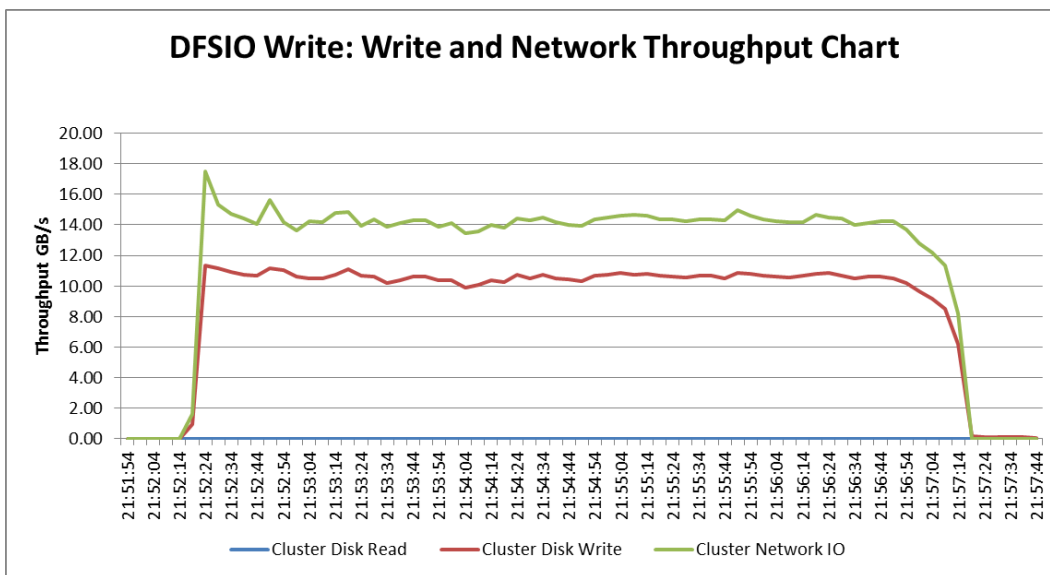**DFSIO Write: Write and Network Throughput Chart**

Figure 3: Write and network throughput chart

DFSIO Read Throughput Charts

The DFSIO-Read benchmark is a read-intensive test, reading the corresponding files that were generated by the DFSIO write workload. This read benchmark test initiates 192 map jobs to read 192 files. The TestDFSIO read benchmark finished under 60 seconds. During this test, the Hadoop cluster generated an average read throughput of 15.72 GB/s and average network I/O throughput of 19.67 GB/s.
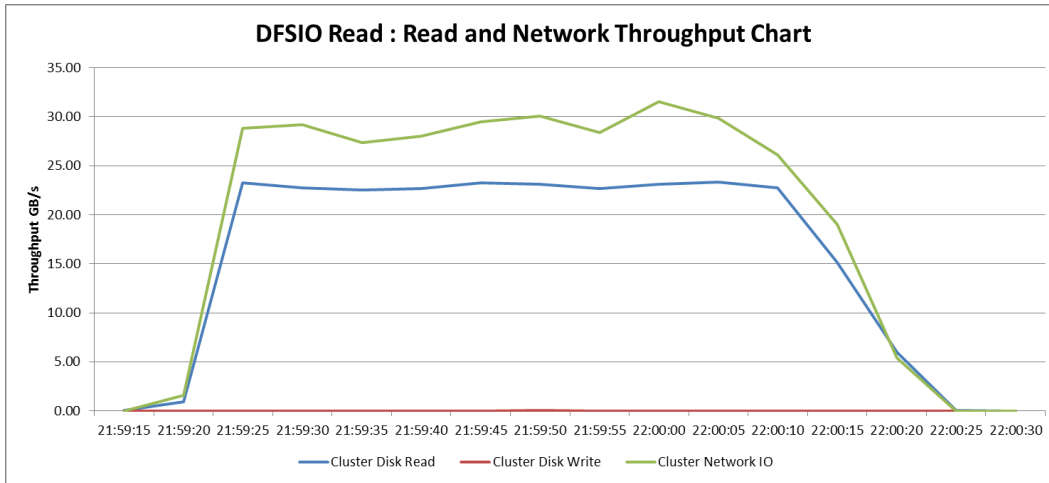

Figure 4: Read and network throughput chart

Measuring Performance with Application Benchmark Suite

Over the past quarter-century, industry standard benchmarks have had a significant impact on the computing industry. Vendors use benchmark standards to illustrate performance competitiveness for their existing products, as well as to improve and monitor the performance of their products under development.

Demonstrating the Transaction Processing Performance Council's commitment to bring relevant benchmarks to industry, TPCx-HS becomes the first standard that provides verifiable performance, price/performance, and energy consumption metrics for Big Data systems. TPCx-HS can be used to assess a broad range of system topologies and implementation methodologies for Hadoop, in a technically rigorous and directly comparable, vendor-neutral manner. And while modeling is based on a simple application, the results are highly relevant to Big Data hardware and software systems.

This benchmark is executed in three phases:

1. *HSGen* generates the test workload and replicates it three times to all data nodes of the cluster. This workload determines the disk writes and network performance throughput of the Hadoop cluster.

2. *HSSort* samples the input data and sorts the data. The sorted data must be replicated three ways and written on a durable storage. This workload evaluates the CPU and disk performance for sorting and merging.

3. *HSValidate* verifies the cardinality, size, and replication factor of the generated data. This phase evaluates the network and read throughput performance of the Hadoop cluster.

The workload used in this experiment was based on TPCx-HS but not audited or published. No comparisons were made with published TPCx-HS results. The run report shows the duration of execution for various phases of the test. As shown in the chart below, the 1 TB dataset benchmark completed in 13 minutes.
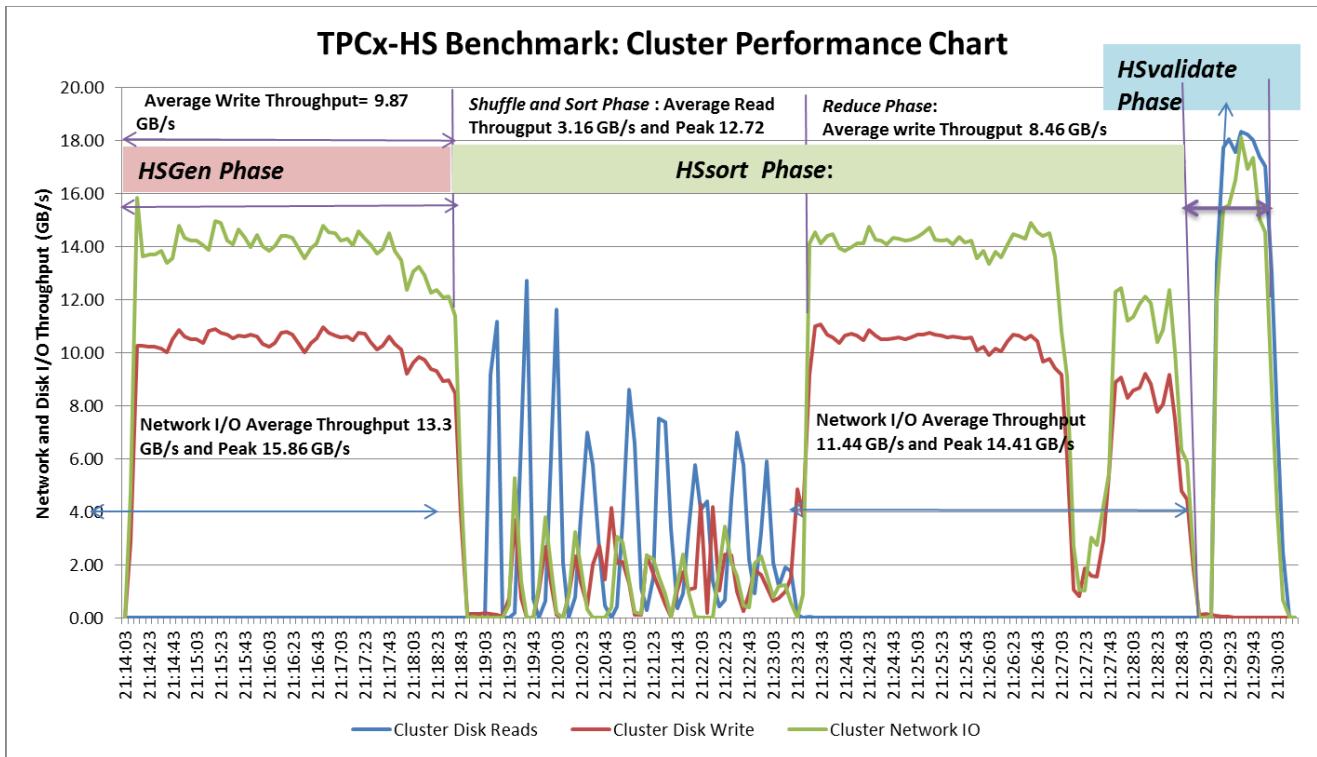
Figure 5: Application performance chart

Both the cluster-level and single-data-node performance chart exhibit uniform performance behavior for various stages of the benchmark. For example, in the HSGen phase, the Hadoop cluster shows an average write throughput of 9.87 GB/s. This equates to 1.41 GB/s per data node on a seven-data-node cluster, figure 6 shows a single data node delivering similar disk write throughput of 1.44 GB/s. These figures demonstrate uniform performance scalability, from a few data nodes to large number of nodes, without sacrificing performance. In the HSSort phase, similar performance scalability was realized from a single data node to a seven-data-node cluster.

Single Data Node

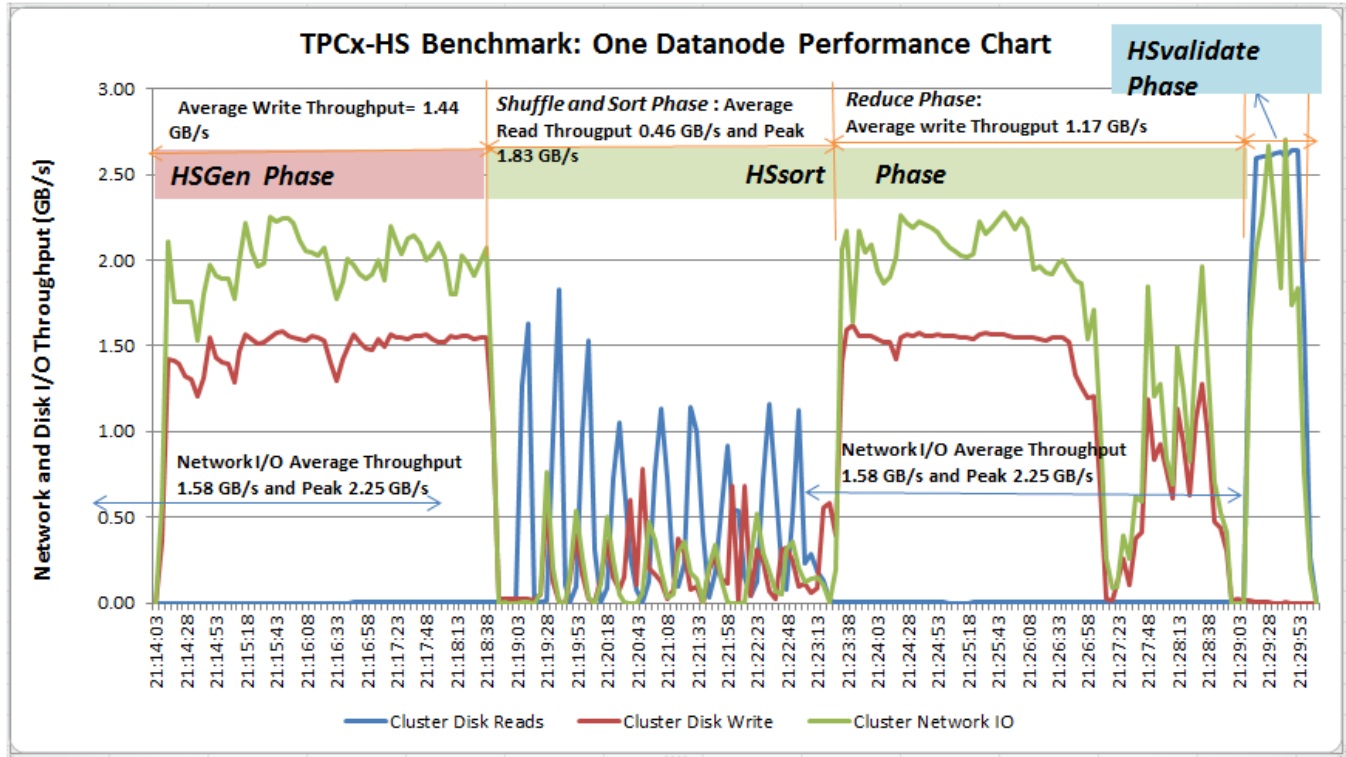The performance results for the single-node data are shown in the chart below.

Figure 6: Single-node-data performance chart

Conclusion

With Big Data becoming increasingly important for gaining a business advantage, it's important to understand the challenges with its implementation. The solution offered in this document explains these challenges and offers ways to mitigate them, as well as describing the performance and scalability advantages for Hadoop-based Big Data deployment. Cisco UCS servers and Cisco Storage Accelerators benefit Hadoop cluster deployment with improved operation efficiency, faster analytics with millisecond latency – and at a lower cost. This solution stack can be seamlessly integrated with existing infrastructure using Cisco UCS servers, which enables customers to confidently engage in their Big Data plans.

Acknowledgments

Prasad Venkatachar, Technical Marketing Manager, SanDisk, and Manan Trivedi, Technical Marketing Engineer, Cisco, were instrumental in the design, development, and delivery of this Big Data solution.

Resources

Customers can find more information about Cisco Big Data solutions and SanDisk products at the links listed below:

Cisco

Design Zone for Big Data :

 http://www.cisco.com/c/en/us/solutions/enterprise/data-center-designs-cloud-computing/bigdata.html

Cisco Servers – Unified Computing :

http://www.cisco.com/c/en/us/products/servers-unified-computing/product-listing.html

SanDisk

SanDisk Data Center Products: http://www.sandisk.com/enterprise/

Fusion ioMemory Products:

https://www.sandisk.com/business/datacenter/products/flash-devices/pcie-flash/mezzanine

SanDisk®
a Western Digital brand