TPC Express Benchmark™ HS
Full Disclosure Report

Cisco Data Intelligence Platform
(with 1x Cisco UCS C240 M5 Server; 16x Cisco UCS C240 M5 Servers)

Running

Cloudera Enterprise Basic 6.3.0
on
Red Hat Enterprise Linux Server 7.6
**First Edition - December 2019**

*Cisco Systems, Inc. (Cisco)*, the Sponsor of this benchmark test, believes that the information in this document is accurate as of the publication date. The information in this document is subject to change without notice. The Sponsor assumes no responsibility for any errors that may appear in this document.

The pricing information in this document is believed to accurately reflect the current prices as of the publication date. However, the Sponsor provides no warranty of the pricing information in this document.

Benchmark results are highly dependent upon workload, specific application requirements, and system design and implementation. Relative system performance will vary because of these and other factors. Therefore, the TPC Express Benchmark™ V should not be used as a substitute for a specific customer application benchmark when critical capacity planning and/or product evaluation decisions are contemplated.

All performance data contained in this report was obtained in a rigorously controlled environment. Results obtained in other operating environments may vary significantly. No warranty of system performance or price/performance is expressed or implied in this report.

Cisco and the Cisco Logo are trademarks of Cisco Systems, Inc. and/or its affiliates in the U.S. and other countries. Third party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company.

TPC Express Benchmark™ HS, TPCx-HS, and HSph, are registered certification marks of the Transaction Processing Performance Council.

The Cisco products, services or features identified in this document may not yet be available or may not be available in all areas and may be subject to change without notice. Consult your local Cisco business contact for information on the products or services available in your area. You can find additional information via Cisco’s web site at {www.cisco.com}. Actual performance and environmental costs of Cisco products will vary depending on individual customer configurations and conditions.

**Copyright© 2019 Cisco Systems, Inc.**

All rights reserved. Permission is hereby granted to reproduce this document in whole or in part provided the copyright notice printed above is set forth in full text or on the title page of each item reproduced.
Abstract

This document contains the methodology and results of the TPC Express Benchmark™ HS (TPCx-HS) test conducted in conformance with the requirements of the TPCx-HS Standard Specification, Revision 2.0.3.

The benchmark results are summarized below.

<table>
<thead>
<tr>
<th>Measured Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company Name</td>
</tr>
<tr>
<td>Cisco</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TPC Express Benchmark™ HS Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total System Cost</td>
</tr>
<tr>
<td>$1,014,330</td>
</tr>
</tbody>
</table>

Executive Summary

The Executive Summary follows on the next several pages.
Cisco Data Intelligence Platform

Availability Date
Currently Available

TPCx-HS Performance
15.03
HSph@10TB

Price/Performance
$67,487.03
$/HSph@10TB

Total System Cost
$1,014,330 USD

System Under Test Configuration Overview

<table>
<thead>
<tr>
<th>Scale Factor</th>
<th>Hadoop Software</th>
<th>Operating System</th>
<th>Other Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Cloudera Enterprise Basic 6.3.0</td>
<td>Red Hat Enterprise Linux Server 7.6</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Servers:
- 17 (1x Cisco UCS C240 M5; 16x Cisco UCS C240 M5)
- Total Processors/Cores/Threads: 34/680/1,360

Server Configuration:
- Processors:
- Memory:
- Storage Controller:
- Storage Device:
- Network:
  - VIC 1387 Dual Port 40Gb

Connectivity:
- Total Rack Units:
  - 2x Cisco UCS Fabric Interconnect 6332 (32x40G)

Physical Storage/Scale Factor: 121.46
Scale Factor/Physical Memory: 1.57

Report Date
Dec. 12, 2019
## Description | Part Number | Source | Unit Price | Qty | Extended Price | 3 Yr. Maint. Price
--- | --- | --- | --- | --- | --- | ---
**Hardware**
UCS C240 M5 24 SFF + 2 rear drives w/o CPU, mem, HD, PCIe, PS | UCSC-C240-M5SX | 1 | $4,939.00 | 16 | $79,024.00 | |
SNTE 24X7X4OS UCS C240 M5 24 SFF + 2 rear drives w/o CPU, mem, HD, PCIe, PS | CON-OSP-C240M5SX | 1 | $3,102.75 | 16 | $50,044.00 | |
Riser 1 incl 3 PCIe slots (x8, x16, x8); slot 3 req CPU2 | UCSC-PCI-1-C240M5 | 1 | $199.00 | 16 | $3,184.00 | |
Cisco VIC 1387 Dual Port 40Gb QSFP CNA MLOM | UCSC-MLOM-C40Q-03 | 1 | $2,192.00 | 16 | $35,072.00 | |
240GB SATA M.2 | UCS-MZ-240GB | 1 | $535.00 | 16 | $8,560.00 | |
Ball Bearing Rail Kit for C220 & C240 M4 & M5 rack servers | UCSC-RAILB-M4 | 1 | $220.00 | 16 | $3,520.00 | |
IMC SW (Recommended) latest release for C-Series Servers. | CIMC-LATEST | 1 | $0.00 | 16 | $0.00 | |
Big Data and Analytics Platform (Hadoop/IoT/ITOA/ML) | UCS-SID-INFR-M5 | 1 | $0.00 | 16 | $0.00 | |
Big Data and Analytics (Hadoop/IoT/ITOA) | UCS-SID-WKL-M5 | 1 | $0.00 | 16 | $0.00 | |
Mini Storage carrier for M.2 SATA/NVME (holds up to 2) | UCS-MSTOR-M2 | 1 | $0.00 | 16 | $0.00 | |
Super Cap cable for UCS-RAID-M5HD | CBL-SC-MR12GMSP | 1 | $0.00 | 16 | $0.00 | |
Super Cap for UCS-RAID-M5, UCS-MRAID1GB-KIT | UCSC-SCAP-M5 | 1 | $0.00 | 16 | $0.00 | |
Cisco 12G Modular RAID controller with 4GB cache | UCSC-RAID-M5HD | 1 | $2,900.00 | 16 | $46,400.00 | |
2.4 TB 12G SAS 10K RPM SFF HDD (4K) | UCS-HD24TB10K4KN | 1 | $2,065.00 | 16 | $33,040.00 | |
32GB DDR4-2933-MHz RDIMM/2Rx4/1.2v | UCS-MR-X32G2RT-H | 1 | $2,043.00 | 16 | $32,688.00 | |
C240 Rear UCS-RAID-M5HD SAS cbl (1) | UCS-RAID-M5HD | 1 | $0.00 | 16 | $0.00 | |
C240 M5 Front NVMe cable (1) | UCSC-PCI-1-C240M5 | 1 | $199.00 | 16 | $3,184.00 | |
8TB 2.5in U.2 Intel P4510 NVMe High Perf. Value Endurance | UCS-NVME-R8000 | 1 | $14,675.00 | 32 | $469,600.00 | |
2.4 TB 12G SAS 10K RPM SFF HDD (4K) | UCS-HD24TB10K4KN | 1 | $2,065.00 | 32 | $66,080.00 | |
Riser 2 (Recommended) latest release for C-Series Servers. | CIMC-LATEST | 1 | $0.00 | 16 | $0.00 | |
Cisco UCS 1050W AC Power Supply for Rack Server | UCSC-PSU-1050W | 1 | $729.00 | 16 | $11,664.00 | |
Power Cord, 125VAC 13A NEMA 5-15 Plug, North America | CAB-9K1220-NA | 1 | $0.00 | 16 | $0.00 | |
Heat sink for UCS C240 M5 Rack servers 150W CPUs & below | UCSC-HS-C240M5 | 1 | $0.00 | 16 | $0.00 | |
Intel 6230 2.1GHz/128C 125W 20C/27.50MB  DCP DDR4 2933MHz | UCSC-NVMEHW-I8000 | 1 | $14,675.00 | 32 | $469,600.00 | |
Cisco R42612 standard rack, w/side panels | UCSC-R42612 | 1 | $6,241.00 | 16 | $100,256.00 | |
SNTC 8X5XNBD, Cisco R42612 standard rack, w side panels | CON-SNT-RCK2UCS2 | 1 | $300.00 | 16 | $4,800.00 | |
3rd Gen FI Per port License to connect C-direct only | UCS-LIC-6300-40GC= | 1 | $1,388.00 | 20 | $27,760.00 | |
(continued next page)
TPC Pricing 2.5.0

<table>
<thead>
<tr>
<th>Description</th>
<th>Part Number</th>
<th>Source</th>
<th>Unit Price</th>
<th>Qty</th>
<th>Extended Price</th>
<th>3 Yr. Maint. Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCS C240 M5 24 SFF + 2 rear drives w/o CPU, mem, HD, PCIe, PS</td>
<td>UCSC-C240-M5SX</td>
<td>1</td>
<td>$4,939.00</td>
<td>1</td>
<td>$4,939.00</td>
<td></td>
</tr>
<tr>
<td>Riser 1 incl 3 PCIe slots (x8, x16, x8); slot 3 req CPU2</td>
<td>UCSC-PCI-1-C240M5</td>
<td>1</td>
<td>$199.00</td>
<td></td>
<td>$199.00</td>
<td></td>
</tr>
<tr>
<td>Cisco VIC 1387 Dual Port 40Gb QSFP CNA MLOM</td>
<td>UCSC-MLOM-C40Q-03</td>
<td>1</td>
<td>$2,192.00</td>
<td>1</td>
<td>$2,192.00</td>
<td></td>
</tr>
<tr>
<td>240GB SATA M.2</td>
<td>UCSC-M2-240G</td>
<td>2</td>
<td>$535.00</td>
<td>2</td>
<td>$1,070.00</td>
<td></td>
</tr>
<tr>
<td>Ball Bearing Rail Kit for C220 &amp; C240 M4 &amp; M5 rack servers</td>
<td>UCSC-RAILB-M4</td>
<td>1</td>
<td>$220.00</td>
<td></td>
<td>$220.00</td>
<td></td>
</tr>
<tr>
<td>IMC SW (Recommended) latest release for C-Series Servers.</td>
<td>CIMC-LATEST</td>
<td>1</td>
<td>$0.00</td>
<td></td>
<td>$0.00</td>
<td></td>
</tr>
<tr>
<td>Big Data and Analytics Platform (Hadoop/IoT/ITOA/Al/ML)</td>
<td>UCS-SID-INFR-BD</td>
<td>1</td>
<td>$0.00</td>
<td></td>
<td>$0.00</td>
<td></td>
</tr>
<tr>
<td>Big Data and Analytics (Hadoop/IoT/ITOA)</td>
<td>UCS-SID-WKL-BD</td>
<td>1</td>
<td>$0.00</td>
<td></td>
<td>$0.00</td>
<td></td>
</tr>
<tr>
<td>Mini Storage carrier for M.2 SATA/NVME (holds up to 2)</td>
<td>UCS-MSTOR-M2</td>
<td>1</td>
<td>$0.00</td>
<td></td>
<td>$0.00</td>
<td></td>
</tr>
<tr>
<td>Super Cap cable for UCS-RAID-M5HD</td>
<td>CBL-SC-MR12GMSP</td>
<td>1</td>
<td>$0.00</td>
<td></td>
<td>$0.00</td>
<td></td>
</tr>
<tr>
<td>Super Cap for UCS-RAID-MS, UCS-MRAID1GB-KIT</td>
<td>UCSC-SCAP-M5</td>
<td>1</td>
<td>$0.00</td>
<td></td>
<td>$0.00</td>
<td></td>
</tr>
<tr>
<td>Cisco 12G Modular RAID controller with 4GB cache</td>
<td>UCSC-RAID-M5HD</td>
<td>1</td>
<td>$2,900.00</td>
<td>1</td>
<td>$2,900.00</td>
<td></td>
</tr>
<tr>
<td>32GB DDR4-2933-M1 RDIMM/2Rx4/1.2v</td>
<td>UCSC-MR-X32G2RT-H</td>
<td>1</td>
<td>$2,043.00</td>
<td>12</td>
<td>$24,516.00</td>
<td></td>
</tr>
<tr>
<td>Riser 2C incl 3 PCIe slots (x8) supports front+rear NVMe</td>
<td>UCSC-PCI-2C-240M5</td>
<td>1</td>
<td>$199.00</td>
<td>1</td>
<td>$199.00</td>
<td></td>
</tr>
<tr>
<td>Cisco UCS 1050W AC Power Supply for Rack Server</td>
<td>UCSC-PSU1-1050W</td>
<td>1</td>
<td>$729.00</td>
<td></td>
<td>$1,458.00</td>
<td></td>
</tr>
<tr>
<td>Power Cord, 125VAC 13A NEMA 5-15 Plug, North America</td>
<td>CAB-9K12A-NA</td>
<td>1</td>
<td>$0.00</td>
<td>2</td>
<td>$0.00</td>
<td></td>
</tr>
<tr>
<td>Heat sink for UCS C240 M5 rack servers 150W CPUs &amp; below</td>
<td>UCSC-HS-CS240M5</td>
<td>1</td>
<td>$0.00</td>
<td>2</td>
<td>$0.00</td>
<td></td>
</tr>
<tr>
<td>UCS C-Series M5 SFF drive blanking panel</td>
<td>UCSC-BBLKD-S2</td>
<td>1</td>
<td>$0.00</td>
<td>2</td>
<td>$0.00</td>
<td></td>
</tr>
<tr>
<td>Intel 6230 2.1GHz/125W 20C/27.50MB DCP DDR4 2933 MHz</td>
<td>UCS-CPU-6230</td>
<td>1</td>
<td>$6,500.00</td>
<td>2</td>
<td>$13,000.00</td>
<td></td>
</tr>
<tr>
<td>1.2 TB 12G SAS 10K RPM SFF HDD</td>
<td>UCS-HD12T10K12N</td>
<td>1</td>
<td>$1,533.00</td>
<td>24</td>
<td>$36,792.00</td>
<td></td>
</tr>
</tbody>
</table>

**Software**

- Red Hat Enterprise Linux (1-2 CPU, 1-2 VN); 3-Yr Support Req  
  | RHEL-2S2V-3A | 1 | $0.00 | 17 | $0.00 |
- ISV 24x7 RHEL Server 2Socket-OR-2Virtual; ANNUAL List Price  
  | CON-ISV1-EL2S2V3A | 1 | $3,897.00 | 17 | $66,249.00 |
- Cloudera Ent. Basic, Node LIC, Gold  
  | UCS-BD-CEBN-GD= | 1 | $0.00 | 17 | $0.00 |
- Cloudera Ent. Basic, Node LIC, Gold - 3 Year  
  | UCS-BD-CEBN-GD-3Y | 1 | $10,461.54 | 17 | $177,846.18 |

**Hardware & Software Subtotals** | $2,488,029.18 | $67,070.55 |

**Large Purchase Discount 1**

- 61% for products and 35% for service  
  | 1 | $1,517,697.80 | $23,474.69 |

**Hardware & Software Totals** | $970,331.38 | $43,595.86 |

**Infrastructure**

- ViewSonic VA2246M 22" Monitor  
  | 2 | $83.39 | 3 | $250.17 |
- Logitech MK120 USB Wired Keyboard/mouse  
  | 2 | $12.25 | 3 | $36.75 |

**Totals** | $970,618.30 | $43,595.86 |

**Pricing:** 1 = Cisco; 2 = CDW

- Discount applies to all line items where Key = 1. Discount based upon total system cost as purchased by a regular customer.
- S: One or more components of the measured configuration have been substituted in the priced configuration. See the FDR for details.

**Three-Year Cost of Ownership:** $1,014,330

**HSph@10TB:** 15.03

$ / HSph@10TB: $67,487.03

**Audited by Doug Johnson, InfoSizing**

Prices used in TPC benchmarks reflect the actual prices a customer would pay for a one-time purchase of the stated Line Items. Individually negotiated discounts are not permitted. Special prices based on assumptions about past or future purchases are not permitted. All discounts reflect standard pricing policies for the listed Line Items. For complete details, see the pricing section of the TPC Benchmark Standard. If you find that the stated prices are not available according to these terms, please inform the TPC at pricing@tpc.org. Thank you.
### Numerical Quantities

#### Performance Run – Run 2

<table>
<thead>
<tr>
<th>Scale Factor</th>
<th>10TB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run Start Time</td>
<td>2019-11-21 14:44:30.000</td>
</tr>
<tr>
<td>Run End Time</td>
<td>2019-11-21 15:24:23.000</td>
</tr>
<tr>
<td>Run Elapsed Time</td>
<td>2,394.000</td>
</tr>
<tr>
<td>HSGen Start Time</td>
<td>2019-11-21 14:44:31.000</td>
</tr>
<tr>
<td>HSGen End Time</td>
<td>2019-11-21 14:54:47.000</td>
</tr>
<tr>
<td>HSGen Elapsed Time</td>
<td>617.606</td>
</tr>
<tr>
<td>HSSort Start Time</td>
<td>2019-11-21 14:54:49.000</td>
</tr>
<tr>
<td>HSSort End Time</td>
<td>2019-11-21 15:18:02.000</td>
</tr>
<tr>
<td>HSSort Elapsed Time</td>
<td>1,393.772</td>
</tr>
<tr>
<td>HSValidate Start Time</td>
<td>2019-11-21 15:18:05.000</td>
</tr>
<tr>
<td>HSValidate End Time</td>
<td>2019-11-21 15:24:23.000</td>
</tr>
<tr>
<td>HSValidate Elapsed Time</td>
<td>378.862</td>
</tr>
</tbody>
</table>

#### Repeatability Run – Run 1

<table>
<thead>
<tr>
<th>Scale Factor</th>
<th>10TB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run Start Time</td>
<td>2019-11-21 14:04:29.000</td>
</tr>
<tr>
<td>Run End Time</td>
<td>2019-11-21 14:43:55.000</td>
</tr>
<tr>
<td>Run Elapsed Time</td>
<td>2,368.000</td>
</tr>
<tr>
<td>HSGen Start Time</td>
<td>2019-11-21 14:04:30.000</td>
</tr>
<tr>
<td>HSGen End Time</td>
<td>2019-11-21 14:14:10.000</td>
</tr>
<tr>
<td>HSGen Elapsed Time</td>
<td>581.143</td>
</tr>
<tr>
<td>HSSort Start Time</td>
<td>2019-11-21 14:14:12.000</td>
</tr>
<tr>
<td>HSSort End Time</td>
<td>2019-11-21 14:37:54.000</td>
</tr>
<tr>
<td>HSSort Elapsed Time</td>
<td>1,422.483</td>
</tr>
<tr>
<td>HSValidate Start Time</td>
<td>2019-11-21 14:37:56.000</td>
</tr>
<tr>
<td>HSValidate End Time</td>
<td>2019-11-21 14:43:55.000</td>
</tr>
<tr>
<td>HSValidate Elapsed Time</td>
<td>359.853</td>
</tr>
</tbody>
</table>
Run Reports

Run Report for Performance Run – Run 2
===========================================
TPCx-HS Performance Metric (HSph@SF) Report

Test Run 2 Details
Total Size = 100000000000
Scale-Factor = 10

TPCx-HS Performance Metric (HSph@SF): 15.0375

Run Report for Repeatability Run – Run 1
===========================================
TPCx-HS Performance Metric (HSph@SF) Report

Test Run 1 Details
Total Size = 100000000000
Scale-Factor = 10

TPCx-HS Performance Metric (HSph@SF): 15.2045
Revision History

<table>
<thead>
<tr>
<th>Date</th>
<th>Edition</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>December 12, 2019</td>
<td>First</td>
<td>Initial Publication</td>
</tr>
</tbody>
</table>
Table of Contents

Abstract ................................................................................................................................. 3
Executive Summary ............................................................................................................... 3
Table of Contents .................................................................................................................. 10
Clause 0 – Preamble .............................................................................................................. 11
  0.1 TPCx Express Benchmark™ HS Overview ................................................................. 11
Clause 1 – General Items ..................................................................................................... 12
  1.1 Test Sponsor .................................................................................................................. 12
  1.2 Parameter Settings ....................................................................................................... 12
  1.3 Configuration Diagrams .............................................................................................. 12
    1.3.1 Measured Configuration ....................................................................................... 13
    1.3.2 Priced Configuration ........................................................................................... 13
  1.4 Dataset Distribution .................................................................................................... 14
  1.5 Software Components Distribution ........................................................................... 14
Clause 2 – Workload Related Items ................................................................................... 15
  2.1 Hardware & Software Tunables .................................................................................. 15
  2.2 Run Report .................................................................................................................. 15
  2.3 Benchmark Kit Identification ..................................................................................... 15
  2.4 Benchmark Kit Changes ............................................................................................ 15
Clause 3 – SUT Related Items ............................................................................................ 16
  3.1 Data Storage Ratio ..................................................................................................... 16
  3.2 Memory Ratio ............................................................................................................. 16
Clause 4 – Metrics Related Items ...................................................................................... 17
  4.1 HSGen Time ............................................................................................................... 17
  4.2 HSSort Time ............................................................................................................... 17
  4.3 HSValidate Time ......................................................................................................... 17
  4.4 HSDataCheck Times ................................................................................................. 17
  4.5 Performance & Price-Performance .......................................................................... 17
Auditor’s Information & Letter of Attestation ..................................................................... 18
Supporting Files Index ......................................................................................................... 21
Third-Party Price Quotes .................................................................................................... 22
CDW .................................................................................................................................... 22
Clause 0 – Preamble

0.1 TPC Express Benchmark™ HS Overview

The TPC Express Benchmark™ HS (TPCx-HS) was developed to provide an objective measure of hardware, operating system and commercial Apache Hadoop File System API compatible software distributions, and to provide the industry with verifiable performance, price-performance and availability metrics. The benchmark models a continuous system availability of 24 hours a day, 7 days a week.

Even though the modeled application is simple, the results are highly relevant to hardware and software dealing with Big Data systems in general. TPCx-HS stresses both hardware and software including Hadoop run-time, Hadoop File-system API compatible systems and MapReduce layers. This workload can be used to assess a broad range of system topologies and implementation of Hadoop clusters. TPCx-HS can be used to assess a broad range of system topologies and implementation methodologies in a technically rigorous and directly comparable and vendor-neutral manner.

The TPCx-HS kit is available from the TPC (See www.tpc.org/tpcx-hs for more information). Users must sign-up and agree to the TPCx-HS User Licensing Agreement (ULA) to download the kit. Re-distribution of the kit is prohibited. All related work (such as collaterals, papers, derivatives) must acknowledge the TPC and include TPCx-HS copyright. The TPCx-HS Kit includes: TPCx-HS Specification document, TPCx-HS Users Guide documentation, shell scripts to set up the benchmark environment and Java code to execute the benchmark load.

The purpose of TPC benchmarks is to provide relevant, objective performance data to industry users. To achieve that purpose, TPC benchmark specifications require that benchmark tests be implemented with systems, products, technologies and pricing that:

- Are generally available to users;
- Are relevant to the market segment that the individual TPC benchmark models or represents (e.g., TPCx-HS models and represents Hadoop run-time and Hadoop File-system API compatible systems);
- Would plausibly be implemented by a significant number of users in the market segment the benchmark models or represents.

The use of new systems, products, technologies (hardware or software) and pricing is encouraged so long as they meet the requirements above. Specifically prohibited are benchmark systems, products, technologies or pricing (hereafter referred to as "implementations") whose primary purpose is performance optimization of TPC benchmark results without any corresponding applicability to real-world applications and environments. In other words, all "benchmark special" implementations that improve benchmark results but not real-world performance or pricing, are prohibited.

The rules for pricing are included in the TPC Pricing Specification and rules for energy measurement are included in the TPC Energy Specification. Further information is available at www.tpc.org.
Clause 1 – General Items

1.1 Test Sponsor
A statement identifying the benchmark sponsor(s) and other participating companies must be provided.

This benchmark was sponsored by Cisco Systems, Inc.

1.2 Parameter Settings
Settings must be provided for all customer-tunable parameters and options which have been changed from the defaults found in actual products, including by not limited to:

- Configuration parameters and options for server, storage, network and other hardware component incorporated into the pricing structure;
- Configuration parameters and options for operating system and file system component incorporated into the pricing structure;
- Configuration parameters and options for any other software component incorporated into the pricing structure;
- Compiler optimization options.

Comment 1: In the event that some parameters and options are set multiple times, it must be easily discernible by an interested reader when the parameter or option was modified and what new value it received each time.

Comment 2: This requirement can be satisfied by providing a full list of all parameters and options, as long as all those that have been modified from their default values have been clearly identified and these parameters and options are only set once.

The supporting files contain the parameters and options used to configure the components involved in this benchmark.

1.3 Configuration Diagrams
Diagrams of both measured and priced configurations must be provided, accompanied by a description of the differences. This includes, but is not limited to:

- Total number of nodes used;
- Total number and type of processors used/total number of cores used/total number of threads used (including sizes of L2 and L3 caches);
- Size of allocated memory, and any specific mapping/partitioning of memory unique to the test;
- Number and type of disk units (and controllers, if applicable);
- Number of channels or bus connections to disk units, including their protocol type;
- Number of LAN (e.g., Ethernet) connections and speed for switches and other hardware components physically used in the test or are incorporated into the pricing structure;
- Type and the run-time execution location of software components.
1.3.1 Measured Configuration

The measured configuration consisted of:

- Total Nodes: 17 (1x Cisco UCS C240 M5; 16x Cisco UCS C240 M5)
- Total Processors/Cores/Threads: 34/680/1,360
- Total Memory: 6.38 TiB
- Total Number of Storage Drives/Devices: 474
- Total Storage Capacity: 1,214.56TB

Server node details:

- 17x Cisco UCS C240 M5 Servers, each with:
  - Processors/Cores/Threads: 2/40/80
  - Processor Model: Intel(R) Xeon(R) Gold 6230
  - Memory: 384 GiB
  - Controller: 1x 12G SAS Modular RAID
  - Drives:
    - 2x 240GB SATA M.2 (all nodes)
    - 24x 1.2TB 10k rpm SAS HDD (name node)
    - 2x 8TB Intel NVMe (16 data nodes)
    - 24x 2.4TB 10k rpm SAS HDD (15 data nodes)
    - 23x 2.4TB 10k rpm SAS HDD; 1x 1.2TB 10k rpm SAS HDD (1 data node)
  - Network: VIC 1387 Dual Port 40Gb

Network connectivity detail:

- 2x Cisco UCS Fabric Interconnects 6332 (32x40G)

The distribution of software components over server nodes is detailed in section Error! Reference source not found..
1.3.2 Priced Configuration
The priced configuration substituted one 2.4TB 10k rpm SAS HDD for one 1.2TB 10K rpm SAS HDD in the measured configuration on one of the data nodes.

1.4 Dataset Distribution
*The distribution of dataset across all media must be explicitly described.*

Table 1-1 describes the distribution of the dataset across all media in the system.

<table>
<thead>
<tr>
<th>Server Node</th>
<th>Controller</th>
<th>Disk Drive</th>
<th>Description of Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cisco 12G Modular RAID controller with 4GB cache</td>
<td>1-24 (HDD, RAID-1)</td>
<td>RAID-1 Disk for Data and Temp</td>
</tr>
<tr>
<td>1</td>
<td>Embedded RAID PCH SATA</td>
<td>0 (2 x SSD, RAID-1)</td>
<td>Boot Disk for Operating System, Root, Swap, Hadoop Master</td>
</tr>
<tr>
<td>2-17</td>
<td>Cisco 12G Modular RAID controller with 4GB cache</td>
<td>1-24(HDD, RAID-0)</td>
<td>Data</td>
</tr>
<tr>
<td>2-17</td>
<td>Embedded RAID PCH SATA</td>
<td>0 (2 x SSD, RAID-1)</td>
<td>Boot Disk for Operating system, Root, Swap, Hadoop Master</td>
</tr>
<tr>
<td>2-17</td>
<td>NVME-Direct-Attached</td>
<td>25-26 (2 x NVMe)</td>
<td>Temp (NodeManager Local Directories, NodeManager Container Log Directories)</td>
</tr>
</tbody>
</table>

**Table 1-1 Dataset Distribution**

1.5 Software Components Distribution
*The distribution of various software components across the system must be explicitly described.*

Table 1-2 describes the distribution of the software components across the system.

<table>
<thead>
<tr>
<th>Map/Reduce</th>
<th>HDFS</th>
<th>ZooKeeper</th>
<th>Spark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node</td>
<td>Resource Manager</td>
<td>Node Manager</td>
<td>NameNode</td>
</tr>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>4-17</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

**Table 1-2 Software Component Distribution**

*Distributed file system implementation and corresponding Hadoop File System API version must be disclosed.*

Cloudera Enterprise Basic 6.3.0 (fully HDFS compatible at the API level).
Map/Reduce implementation and corresponding version must be disclosed.

Cloudera Enterprise Basic 6.3.0 (compatible equivalent to Hadoop 3.0.0).
Clause 2 – Workload Related Items

2.1 Hardware & Software Tunables

*Script or text used to set for all hardware and software tunable parameters must be reported.*

The Supporting File Archive contains all configuration scripts.

2.2 Run Report

*The run report generated by TPCx-HS benchmark kit must be reported.*

The Supporting File Archive contains the full run report. Following are extracts from the run report that lists the performance summary for both runs.

### Run Report for Run 1 – Repeatability Run

```
TPCx-HS Performance Metric (HSph@SF) Report

Test Run 1 Details
Total Time = 2368
Total Size = 100000000000
Scale-Factor = 10

TPCx-HS Performance Metric (HSph@SF): 15.2045
```

### Run Report for Run 2 – Performance Run

```
TPCx-HS Performance Metric (HSph@SF) Report

Test Run 2 Details
Total Time = 2394
Total Size = 100000000000
Scale-Factor = 10

TPCx-HS Performance Metric (HSph@SF): 15.0375
```

2.3 Benchmark Kit Identification

*Version number of TPCx-HS kit and checksum for HSGen, HSSort and HSValidate Programs must be reported.*

<table>
<thead>
<tr>
<th>Kit Version</th>
<th>MD5</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0.3</td>
<td></td>
</tr>
</tbody>
</table>

- BigData_cluster_validate_suite.sh: 57f7cd6b251a9aba0f6b648630f55da
- HSDataCheck.sh: faeff3091759a980880be4e39f7896a
- TPCx-HS-master_MR2.jar: 492cc51a1a60c28b43d96c79d08683d
- TPCx-HS-master.sh: c619a0819571ed00cd75d2b76ba8c64

2.4 Benchmark Kit Changes

No modifications were made to the TPC-provided kit.
Clause 3 – SUT Related Items

3.1 Data Storage Ratio

The data storage ratio must be disclosed.

Table 3-1 describes the details of the storage devices configured on the system and their capacity.

<table>
<thead>
<tr>
<th>Quantity</th>
<th>Capacity</th>
<th>Total (TB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>240 GB</td>
<td>8.16</td>
</tr>
<tr>
<td>384</td>
<td>2.4 TB</td>
<td>921.60</td>
</tr>
<tr>
<td>32</td>
<td>8.0 TB</td>
<td>256.00</td>
</tr>
<tr>
<td>24</td>
<td>1.2 TB</td>
<td>28.80</td>
</tr>
<tr>
<td><strong>Total Storage (TB)</strong></td>
<td></td>
<td><strong>1,214.56</strong></td>
</tr>
</tbody>
</table>

Scale Factor = 10

**Data Storage Ratio** = (Total Storage (TB) / SF) = **121.46**

3.2 Memory Ratio

The Scale Factor to memory ratio must be disclosed.

Total Configured Memory (TiB) = 6.38

**Scale Factor to Memory Ratio** = (SF / Total Memory(TiB)) = **1.57**
Clause 4 – Metrics Related Items

4.1 HSGen Time

*The HSGen time must be disclosed for Run1 and Run2.*

<table>
<thead>
<tr>
<th></th>
<th>Run 1</th>
<th>Run 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSGen</td>
<td>581.143</td>
<td>617.606</td>
</tr>
</tbody>
</table>

*Table 4-1 HSGen Times*


4.2 HSSort Time

*The HSSort time must be disclosed for Run1 and Run2.*

<table>
<thead>
<tr>
<th></th>
<th>Run 1</th>
<th>Run 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSSort</td>
<td>1,422.483</td>
<td>1,393.772</td>
</tr>
</tbody>
</table>

*Table 4-2 HSSort Times*


4.3 HSVValidate Time

*The HSVValidate time must be disclosed for Run1 and Run2.*

<table>
<thead>
<tr>
<th></th>
<th>Run 1</th>
<th>Run 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSVValidate</td>
<td>359.853</td>
<td>378.862</td>
</tr>
</tbody>
</table>

*Table 4-3 HSVValidate Times*


4.4 HSDataCheck Times

*Both HSDataCheck times must be disclosed for Run1 and Run2.*

<table>
<thead>
<tr>
<th></th>
<th>Run 1</th>
<th>Run 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSDataCheck (pre-sort)</td>
<td>2.000</td>
<td>2.000</td>
</tr>
<tr>
<td>HSDataCheck (post-sort)</td>
<td>2.000</td>
<td>3.000</td>
</tr>
</tbody>
</table>

*Table 4-4 HSDataCheck Times*


4.5 Performance & Price-Performance

*The performance metric (HSph@SF) must be disclosed for Run 1 and Run 2. Price-performance metric ($/HSph@SF) must be disclosed for the performance run.*

<table>
<thead>
<tr>
<th></th>
<th>Run 1</th>
<th>Run 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSph@10TB</td>
<td>15.20</td>
<td>15.03</td>
</tr>
</tbody>
</table>

*Table 4-5 Performance Metrics*

Run 2 Price-Performance: 67,487.03 $/ HSph@10TB
Auditor’s Information & Letter of Attestation

The auditor’s agency name, address, phone number, and Attestation letter must be included in the full disclosure report. A statement should be included specifying who to contact in order to obtain further information regarding the audit process.

This benchmark was audited by Doug Johnson, InfoSizing.

www.sizing.com
63 Lourdes Drive
Leominster, MA 10453
978-343-6562

This benchmark’s Full Disclosure Report (FDR) can be downloaded from www.tpc.org.

A copy of the auditor’s Letter of Attestation follows.
Karthik Krishna  
Cisco Systems Inc.  
3800 Zanker Road  
San Jose, CA 95134  

December 11, 2019

I verified the TPC Express Benchmark™ H5 v2.0.3 performance of the following configuration:

<table>
<thead>
<tr>
<th>Platform:</th>
<th>Cisco Data Intelligence Platform (17x Cisco UCS C240 M5 Servers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating System:</td>
<td>Red Hat Enterprise Linux Server 7.6</td>
</tr>
<tr>
<td>Apache Hadoop:</td>
<td>Cloudera Enterprise Basic v6.3.0</td>
</tr>
<tr>
<td>Compatible Software:</td>
<td></td>
</tr>
</tbody>
</table>

The results were:

**Performance Metric**: 15.03 HSph@10TB  
**Run Elapsed Time**: 2,394.00 Seconds

<table>
<thead>
<tr>
<th>Cluster</th>
<th>17x Cisco UCS C240 M5 with:</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPUs</td>
<td>2x Intel Xeon Gold 6230 (2.10 GHz, 20-core, 27.5 MB L3)</td>
</tr>
<tr>
<td>Memory</td>
<td>384 GB (all nodes)</td>
</tr>
<tr>
<td>Storage</td>
<td>Qty  Size Type</td>
</tr>
<tr>
<td></td>
<td>2  240GB SATA M.2 (all nodes)</td>
</tr>
<tr>
<td></td>
<td>24 2.4TB 10k rpm SAS HDD (16 data nodes)</td>
</tr>
<tr>
<td></td>
<td>2  8TB Intel NVMe (16 data nodes)</td>
</tr>
<tr>
<td></td>
<td>24 1.2TB 10k rpm SAS HDD (name node)</td>
</tr>
</tbody>
</table>

In my opinion, these performance results were produced in compliance with the TPC requirements for the benchmark.

The following verification items were given special attention:

- All TPC-provided components were verified to be v2.0.3
- No modifications were made to any of the Java code
- Any and all modifications to shell scripts were reviewed for compliance
- All checksums were validated for compliance
- The generated dataset was properly scaled to 10TB
- The generated dataset and the sorted dataset were replicated 3-ways
• The elapsed times for all phases and runs were correctly measured and reported
• The Storage and Memory Ratios were correctly calculated and reported
• The system pricing was verified for major components and maintenance
• The major pages from the FDR were verified for accuracy

Additional Audit Notes:

The measured configuration included (1) 1.2TB 10k rpm SAS HDD disk on one data node that was substituted by (1) 2.4TB 10k rpm disk in the priced configuration. Based on the specifications of these disks, it is my opinion that this substitution has no significant effect on performance.

Respectfully Yours,

[Signature]

Doug Johnson, Certified TPC Auditor
## Supporting Files Index

<table>
<thead>
<tr>
<th>Clause</th>
<th>Description</th>
<th>Archive File Pathname</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clause 1</td>
<td>Parameters and options used to configure the system</td>
<td>SupportingFiles/Clause1</td>
</tr>
<tr>
<td>Clause 2</td>
<td>Configuration scripts and Run Report</td>
<td>SupportingFiles/Clause2</td>
</tr>
<tr>
<td>Clause 3</td>
<td>System configuration details</td>
<td>SupportingFiles/Clause3</td>
</tr>
</tbody>
</table>
Third-Party Price Quotes

CDW

ViewSonic VA2246M 22" LED-backlit LCD - Black

- Screen Size: 22 in
- Display Resolution: 1920 x 1080
- Refresh Rate: 75 Hz
- Response Time: 5 ms
- Aspect Ratio: 16:9

Price: $83.39

Logitech MK120 USB Wired Keyboard/Mouse Set

- Connectivity: Wired
- Interface Type: USB
- Compatibility: PC

Price: $12.25

CDW

Full Disclosure Report

Cisco

Cisco Data Intelligence Platform

Report Date

December 12, 2019