

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

TPCx-HS Version Report Edition Report Submitted 2.0.3 First December 13, 2019

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.

Measured Configuration						
Company Name Cluster Node Hadoop Software Operating System						
Cisco	Cisco UCS C240 M5	Cloudera Enterprise Basic 6.3.0	Red Hat Enterprise Linux Server 7.6			

TPC Express Benchmark™ HS Metrics Total System Cost HSph@1TB Price/Performance Availability Date						
\$1,014,330	15.60	\$65,021.16	Currently Available			

Executive Summary

The <u>Executive Summary</u> follows on the next several pages.

ıılıılı cısco	Cisco Intelligence		TPCx-HS TPC Pricing Report Date Dec	2.0.3 2.5.0 c. 13, 2019
Availability Date	TPCx-HS Performance	Price/Performance	Total Syster	m Cost
Currently Available	15.60 HSph@1TB	\$65,021.16 \$ / HSph@1TB	\$1,014,330	USD
S	ystem Under Test Con	figuration Overview		
Scale Factor	Hadoop Software	Operating System	Other Soft	ware
1	Cloudera Enterprise Basic 6.3.0	Red Hat Enterprise Linux Server 7.6	n/a	
1 x Cisco UCS C240 M5 Servers (N 2 x Intel Xeon Gold 6230 CPU 384 GiB Memory 24 x 1.2TB 10k rpm SAS HDD 2 x 240 GB SATA M.2 (BOOT)	lame node) with:	2 x Cisco UCS Fabric I	nterconnects 6332 (32)	«40G)
Physical Storage/Scale	Factor: 1,214.56	Scale Factor/Phy	vsical Memory: (0.16
Servers: Total Processors/Cores/Threa	, i	C240 M5; 16x Cisco U	CS C240 M5)	
Server Configuration: Processors Memory Storage Controller Storage Device Network Connectivity:	1x Cisco UCS C2 2x Intel(R) Xeon(F 384 GiB 1x 12G SAS Mode 2x 240GB SATA F 24x 1.2TB 10k rpr VIC 1387 Dual Pc	R) Gold 6230 2x Internation 384 Gi ular RAID 1x 120 M.2 2x 240 n SAS HDD 24x 2.4 2x 8TE	SAS Modular R GB SATA M.2 4TB 10k rpm SAS Intel NVMe 87 Dual Port 400	d 6230 AID S HDD
Total Rack Units:		$M5) + (2 \times FI 6332) = (1)$		RU

IIIIIICisco Data IntelligenceTPCx-HSCISCOPlatformReport Date

S 2.0.3 cing 2.5.0

Report Date Dec. 13, 2019

Description	Part Number	Source	Unit Price	Qty	Extended Price	3 Yr. Maint. Price
lardware						
JCS C240 M5 24 SFF + 2 rear drives w/o CPU,mem,HD,PCIe,PS	UCSC-C240-M5SX		\$4,939.00	16	\$79,024.00	
NTC 24X7X4OS UCS C240 M5 24 SFF + 2 rear drives w/o CPU,mem	CON-OSP-C240M5SX		\$3,102.75	17		\$52,746.7
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		\$2,192.00	16	\$35,072.00	
40GB SATA M.2	UCS-M2-240GB	1	\$535.00	32	\$17,120.00	
Ball Bearing Rail Kit for C220 & C240 M4 & M5 rack servers	UCSC-RAILB-M4	1	\$220.00	16	\$3,520.00	
MC SW (Recommended) latest release for C-Series Servers.	CIMC-LATEST	1	\$0.00	16	\$0.00	
ig Data and Analytics Platform (Hadoop/IoT/ITOA/AI/ML)	UCS-SID-INFR-BD	1	\$0.00	16	\$0.00	
Big Data and Analytics (Hadoop/IoT/ITOA)	UCS-SID-WKL-BD	1	\$0.00	16	\$0.00	
/ini Storage carrier for M.2 SATA/NVME (holds up to 2)	UCS-MSTOR-M2	1	\$0.00	16	\$0.00	
uper Cap cable for UCSC-RAID-M5HD	CBL-SC-MR12GM5P	1	\$0.00	16	\$0.00	
uper Cap for UCSC-RAID-M5, UCSC-MRAID1GB-KIT	UCSC-SCAP-M5	1	\$0.00	16	\$0.00	
isco 12G Modular RAID controller with 4GB cache	UCSC-RAID-M5HD	1	\$2,900.00	16	\$46,400.00	
.4 TB 12G SAS 10K RPM SFF HDD (4K)	UCS-HD24TB10K4KN	1S	\$2,065.00	352	\$726,880.00	
2GB DDR4-2933-MHz RDIMM/2Rx4/1.2v	UCS-MR-X32G2RT-H	1	\$2,043.00	192	\$392,256.00	
240 Rear UCS-RAID-M5HD SAS cbl(1)kitinclfan,bkpln	UCSC-RSAS-240M5X	1	\$0.00	16	\$0.00	
240 M5 Front NVMe cable (1)	CBL-NVME-C240SFF	1	\$0.00	16	\$0.00	
TB 2.5in U.2 Intel P4510 NVMe High Perf. Value Endurance	UCSC-NVMEHW-18000	1	\$14,675.00	32	\$469,600.00	
.4 TB 12G SAS 10K RPM SFF HDD (4K)	UCS-HD24TB10K4KN	1	\$2,065.00	32	\$66,080.00	
liser 2C incll 3 PCIe slots (3 x8) supports front+rear NVMe	UCSC-PCI-2C-240M5	1	\$199.00	16	\$3,184.00	
Cisco UCS 1050W AC Power Supply for Rack Server	UCSC-PSU1-1050W	1	\$729.00	32	\$23,328.00	
ower Cord, 125VAC 13A NEMA 5-15 Plug, North America	CAB-9K12A-NA	1	\$0.00	32	\$0.00	
leat sink for UCS C240 M5 rack servers 150W CPUs & below	UCSC-HS-C240M5	1	\$0.00	32	\$0.00	
ntel 6230 2.1GHz/125W 20C/27.50MB DCP DDR4 2933 MHz	UCS-CPU-16230	1	\$6,500.00	32	\$208,000.00	
Cisco R42612 standard rack, w/side panels	RACK2-UCS2	1	\$6,241.00	1	\$6,241.00	
NTC 8X5XNBD, Cisco R42612 standard rack, w side panels	CON-SNT-RCK2UCS2	1	\$300.00	1		\$300.0
Not sold standalone) UCS 6332 1RU FI/12 QSFP+	UCS-SP-FI6332	1	\$24,400.00	2	\$48,800.00	
0NSITE 24X7X4 (Not sold standalone) UCS 6332 1RU FI/No PSU/3	CON-OSP-SPFI6332	1	\$7,011.90	2		\$14,023.8
JCS 6332/ 6454 Power Supply/100-240VAC	UCS-PSU-6332-AC	1	\$0.00	4	\$0.00	
ower Cord, 125VAC 13A NEMA 5-15 Plug, North America	CAB-9K12A-NA	1	\$0.00	4	\$0.00	
OGBASE-CR4 Passive Copper Cable, 3m	QSFP-H40G-CU3M	1	\$0.00	16	\$0.00	
QSFP40G BiDi Short-reach Transceiver	QSFP-40G-SR-BD	1	\$0.00	8	\$0.00	
JCS Manager v3.2	N10-MGT015	1	\$0.00	2	\$0.00	
ICS 6332/ 6454 Fan Module	UCS-FAN-6332	1	\$0.00	8	\$0.00	
JCS 6332/ 6454 Chassis Accessory Kit	UCS-ACC-6332	1	\$0.00	2	\$0.00	
Brd Gen FI Per port License to connect C-direct only	UCS-LIC-6300-40GC=	1	\$1,388.00	20	\$27,760.00	
					. ,	

	Cisco Da	ata I	ntolli		nco	TPCx-	HS		2.0.
				90		TPC F	ricing		2.5.
CISCO	P	latt	orm			Repor	t Date	Dec	. 13, 201
·	(contir	nued from	previous page)						
Descriptior	1		Part Numbe	r Source	Unit Price	Qty	Extended P	rice 3	Yr. Maint. Prio
JCS C240 M5 24 SFF + 2 rear drives w/c	CPU,mem,HD,PCIe,PS	UCSC-C	240-M5SX	1	\$4,939.00	1	\$4,93	9.00	
Riser 1 incl 3 PCIe slots (x8, x16, x8); sl	ot 3 req CPU2	UCSC-F	PCI-1-C240M5	1	\$199.00	1	\$19	9.00	
Cisco VIC 1387 Dual Port 40Gb QSFP CN	IA MLOM	UCSC-N	/LOM-C40Q-03	1	\$2,192.00	1	\$2,19	2.00	
240GB SATA M.2		UCS-M	2-240GB	1	\$535.00	2	\$1,07	0.00	
3all Bearing Rail Kit for C220 & C240 M	4 & M5 rack servers	UCSC-F	RAILB-M4	1	\$220.00	1	\$22	0.00	
MC SW (Recommended) latest releas	e for C-Series Servers.	CIMC-L	ATEST	1	\$0.00	1	\$	0.00	
Big Data and Analytics Platform (Hado	op/IoT/ITOA/AI/ML)	UCS-SII	D-INFR-BD	1	\$0.00	1	\$	0.00	
Big Data and Analytics (Hadoop/IoT/IT	OA)	UCS-SII	D-WKL-BD	1	\$0.00	1	\$	0.00	
Mini Storage carrier for M.2 SATA/NV	VIE (holds up to 2)	UCS-M	STOR-M2	1	\$0.00	1	\$	0.00	
Super Cap cable for UCSC-RAID-M5HD		CBL-SC	-MR12GM5P	1	\$0.00	1		0.00	
Super Cap for UCSC-RAID-M5, UCSC-M			CAP-M5	1	\$0.00	1		0.00	
Cisco 12G Modular RAID controller wit			AID-M5HD		\$2,900.00	1			
32GB DDR4-2933-MHz RDIMM/2Rx4/1.			R-X32G2RT-H	1	\$2,043.00	12			
Riser 2C incll 3 PCIe slots (3 x8) suppor			PCI-2C-240M5	1	\$199.00	1		9.00	
Cisco UCS 1050W AC Power Supply for			SU1-1050W	1	\$729.00	2			
Power Cord, 125VAC 13A NEMA 5-15 P	-		(12A-NA	1	\$0.00	2		0.00	
Heat sink for UCS C240 M5 rack servers			IS-C240M5	1	\$0.00	2		0.00	
JCS C-Series M5 SFF drive blanking pa			BLKD-S2	1	\$0.00	2		0.00	
ntel 6230 2.1GHz/125W 20C/27.50MB	DCP DDR4 2933 MHz		PU-16230	1	\$6,500.00	2			
2 TB 12G SAS 10K RPM SFF HDD		UCS-HL	012TB10K12N	1	\$1,533.00	24	\$36,79	2.00	
Software			C2) / 2 A		ć0.00	47	ć	0.00	
Red Hat Enterprise Linux (1-2 CPU,1-2			S2V-3A	1	\$0.00 \$3,897.00	17 17		0.00	
SV 24X7 RHEL Server 2Socket-OR-2Vir Cloudera Ent. Basic, Node LIC, Gold	lual, ANNOAL LIST PHILE		V1-EL2S2V3A D-CEBN-GD=	1	\$0.00 \$0.00	17		9.00 0.00	
Cloudera Ent. Basic, Node LIC, Gold - 3	3 Year		O-CEBN-GD-3Y		\$10,461.54	17			
			На	rdware 8	Software S	Subtotals	\$2,488,02	9.18	\$67,070.
arge Purchase Discount 1									
51% for products and 35% for service				1 Hardwai	re & Softwa	re Totals	\$1,517,69 \$970,33		\$23,474. \$43,595.
nfrastructure /iewSonic VA2246M 22'' Monitor				2	\$83.39	2	¢2Ε	0.17	
.ogitech MK120 USB Wired Keyboard/	moure			2	\$03.39 \$12.25	3		0.17 6.75	
	mouse			2	<i>¥12.25</i>				
						Totals			\$43,595.
Pricing: 1 = Cisco; 2 = CDW	where Key A. D's		Three-	Year C	Cost of C		-	\$1	,014,33
Discount applies to all line items pon total system cost as purchas						ISph@			15.6
: One or more components of the een substituted in the priced con etails.					\$/H	lSph@	1TB:	\$6	5,021. 1
Audited by Doug Jo	hnson, InfoSizina								
Prices used in TPC benchmarks r ndividually negotiated discounts a permitted. All discounts reflect sta	eflect the actual prices a are not permitted. Speci	al prices l	based on assi	umption	s about pa	ast or fut	ure purch	ases	are not

		Intelligence	TPCx-HS TPC Pricing	2.0.3 2.5.1
CISCO	Plat	tform		Dec. 13, 201
	Numerica	l Quantities		
	Dorformono	e Run – Run 2		
Scale	Factor		1TB	
Run B	Start Time End Time Elapsed Time	2019-12-13 06:29:28 2019-12-13 06:33:18 231		
HSG6 HSG6	en Start Time en End Time en Elapsed Time	2019-12-13 06:29:30 2019-12-13 06:30:38	0.000	
HSSc HSSc	ort Start Time ort End Time ort Elapsed Time	2019-12-13 06:30:41 2019-12-13 06:32:54	.000	
HSVa	alidate Start Time alidate End Time alidate Elapsed Time	2019-12-13 06:32:58 2019-12-13 06:33:18 21		
	Repeatabilit	y Run – Run 1		
Scale	Factor	,,	1TB	
Run E	Start Time End Time Elapsed Time	2019-12-13 06:25:07 2019-12-13 06:28:54 229		
HSGe	en Start Time en End Time en Elapsed Time	2019-12-13 06:25:09 2019-12-13 06:26:16 68		
HSSo	ort Start Time ort End Time ort Elapsed Time	2019-12-13 06:26:19 2019-12-13 06:28:30 132		
HSVa	alidate Start Time alidate End Time alidate Elapsed Time	2019-12-13 06:28:34 2019-12-13 06:28:54 21		

1	Cisco Data Intelligenc	
cisco	Platform	TPC Pricing Report Date Dec. 13,
		Report Date Dec. 13,
	Run Reports	
Run Report	for Performance Run – Run 2	
TPCx-HS Pe	erformance Metric (HSph@SF) Report	
Test Run 2 I		231
	Total Size = Scale-Factor =	1000000000 1
TPCx-HS Pe	erformance Metric (HSph@SF):	15.6006
=========		
Run Report	for Repeatability Run – Run 1	
TPCx-HS Pe	erformance Metric (HSph@SF) Report	
Test Run 1 I		229
	Total Size = Scale-Factor =	1000000000 1
TPCx-HS Pe	erformance Metric (HSph@SF):	15.7232

Cisco Data Intelligence Platform TPCx-HS 2.0.3 TPC Pricing 2.5.0 Report Date Dec. 13, 2019
Revision History
Revision History
Revision History
Revision History
Date Edition Description
December 13, 2019 First Initial Publication

Table of Contents

Abstract	3
Executive Summary	3
Table of Contents	10
Clause 0 – Preamble	11
0.1 TPC 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 BenchmarkTM 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 asses 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 <u>www.tpc.org/tpcx-hs</u> 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

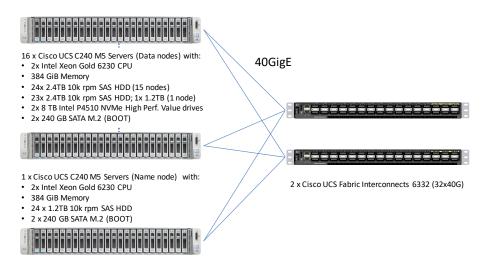


Figure 1-1 Measured Configuration

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: 6.38TiB
- 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 Interconnect 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.

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

Table 1-1Dataset 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.

	Map/R	Map/Reduce		HDFS		Spark
Node	Resource Manager	Node Manager	NameNode	DataNode	QuorumPeer	HistoryServer
1	Х		Х		Х	
2		Х		Х	Х	Х
3		Х		Х	Х	
4-17		Х		Х		

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 = Total Size = Scale-Factor =	229 10000000000 1				
TPCx-HS Performance	ce Metric (HSph@SF):	15.7232				
Run Report for Run 2	- Performance Run					
TPCx-HS Performance	ce Metric (HSph@SF) Report					
Test Run 2 Details	Total Time = Total Size = Scale-Factor =	231 10000000000 1				
TPCx-HS Performance	ce Metric (HSph@SF):	15.6006				

2.3 Benchmark Kit Identification

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

2.0.3

	21010
File	MD5
BigData_cluster_validate_suite.sh	57f7cd68251a9aba0feb6648630ff5da
HSDataCheck.sh	faeff3091759aac98080be4e39f7896a
TPCx-HS-master_Spark.jar	19f3ce092066e056b884a85ee92fb7fc
TPCx-HS-master.sh	c619a0819571ecd00cd75d2b76ba8c64

2.4 Benchmark Kit Changes

Kit Version

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.

Quantity	Capacity	Total (TB)
34	240 GB	8.16
384	2.4 TB	921.60
32	8.0 TB	256.00
24	1.2 TB	28.80
Total Storage (TB)		1,214.56

Table 3-1 Storage Device Capacities

Scale Factor = 1

Data Storage Ratio = (Total Storage (TB) / SF) = 1,214.56

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)) = 0.16

Clause 4 – Metrics Related Items

4.1 HSGen Time

The HSGen time must be disclosed for Run1 and Run2.

	Run 1	Run 2
HSGen	68.743	69.724

Table 4-1 HSGen Times

4.2 HSSort Time

The HSSort time must be disclosed for Run1 and Run2.

	Run 1	Run 2
HSSort	132.994	134.753

Table 4-2 HSSort Times

4.3 HSValidate Time

The HSValidate time must be disclosed for Run1 and Run2.

	Run 1	Run 2
HSValidate	21.575	21.524

Table 4-3 HSValidate Times

4.4 HSDataCheck Times

Both HSDataCheck times must be disclosed for Run1 and Run2.

	Run 1	Run 2
HSDataCheck (pre-sort)	3.000	3.000
HSDataCheck (post-sort)	4.000	4.000

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.

	Run 1	Run 2
HSph@1TB	15.72	15.60

Table 4-5 Performance Metrics

Run 2 Price-Performance: 65,021.16 \$/ HSph@1TB

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 <u>www.tpc.org</u>.

A copy of the auditor's Letter of Attestation follows.

Cisco Data Intelligence Platform (17x Cisco UCS C240 M5 Servers) Red Hat Enterprise Linux Server 7.6 Cloudera Enterprise Basic v6.3.0		
15.60 HSph@1TB 231.00 Seconds		
17x Cisco UCS C240 M5 with:		

63 Lourdes Dr. | Leominster, MA 01453 | 978-343-6562 | www.sizing.com

- 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,

talins

Doug Johnson, Certified TPC Auditor

63 Lourdes Dr. | Leominster, MA 01453 | 978-343-6562 | www.sizing.com

Supporting Files Index

Clause	Description	Archive File Pathname
Clause 1	Parameters and options used to configure the system	SupportingFiles/Clause1
Clause 2	Configuration scripts and Run Report	SupportingFiles/Clause2
Clause 3	System configuration details	SupportingFiles/Clause3

Third-Party Price Quotes CDW

