

TPC Benchmark [™] TPCx-HS Full Disclosure Report DELL PowerEdge R730/730xd Using Cloudera CDH 5.4.2, And Red Hat Enterprise Linux Server 6.5



First Edition

Submitted for Review

October 15, 2015

Dell Inc. PowerEdge R730/R730xd Server with Red Hat Enterprise Linux Server and Cloudera CDH

First Printing October 2015

All rights reserved. Permission is hereby granted to reproduce this document in whole or in part

provided the copyright notice is included on the title page of each item reproduced.

Printed in U.S.A.

DELL believes that the technical, pricing and discounting information in this document is accurate as of

its publication date. The performance information in this document is for guidance only. System

performance is highly dependent on many factors including system hardware, system and user software, and user-

application characteristics. Customer applications must be carefully evaluated before estimating performance. DELL does not warrant or represent that a user can or will achieve similar performance as expressed in this document.

THE TERMS AND CONDITIONS GOVERNING THE SALE OF DELL HARDWARE PRODUCTS AND THE LICENSING OF DELL SOFTWARE CONSIST SOLELY OF THOSE SET FORTH IN THE WRITTEN CONTRACTS BETWEEN DELL AND ITS CUSTOMERS. NO REPRESENTATION OR OTHER AFFIRMATION OF FACT CONTAINED IN THIS DOCUMENT INCLUDING BUT NOT LIMITED TO STATEMENTS REGARDING PRICE, CAPACITY, RESPONSE-TIME PERFORMANCE, SUITABILITY FOR USE, OR PERFORMANCE OF PRODUCTS DESCRIBED HEREIN SHALL BE DEEMED TO BE A WARRANTY BY DELL FOR ANY PURPOSE, OR GIVE RISES TO ANY LIABILITY OF DELL WHATSOEVER.

DELL assumes no responsibility for any errors that may appear in this document. DELL reserves the right to make changes in specifications and other information contained in this document without prior notice, and the reader should in all cases consult DELL to determine whether any such changes have been made.

PowerEdge is an U.S. registered trademark of Dell Inc.

Red Hat Enterprise Linux Server and RHEL are registered trademarks of Red Hat.

CDH is a registered trademark of Cloudera.

Intel and Xeon are registered trademarks of Intel Corporation.

TPC Benchmark TPCx-HS is a trademark of the Transaction Processing Performance Council.

	DELL		_	x-HS Rev. 1.3.0 Pricing Rev. 2.0.0	
DELL		PowerEdge R730/R730xd w/ Cloudera CDH 5.4.2		Report Date: ober 15, 2015	
Total System Cost	TPCx-HS Performance	Metric	Price	e/Performance	
USD 345,578	7.39 HSph@1TI	3	46,762	.93 \$/HSph@1TB	
Apache Hadoop Compatible Software	Operating System	Other S	Software	System Availability Date	
Cloudera CDH 5.4.2 HDFS API Ver 2, Map Reduce API ver 1	Red Hat Enterprise Linux Server 6.5	Se	DK 64-bit rver .0_45	October 15, 2015	
	System	Configu			
IxDell R730 Server (Namenode) w/ 2xIntel E5-2650v3, 8x 1TB,3.5",6G,SATA HDD IzXDell R730xd Servers (Datanodes) each w/ 2xIntel E5-2690v3, 24x 1.2TB,2.5",6G,SAS + 2x300GB,2.5",6G, SAS (OS) 					
Physical Storage	e/Scale Factor	Scale Factor/Physical Memory			
360.80 0.6				0	
Servers: 1 x Dell PowerEdge R730 / 12 x PowerEdge R730xd					
Processors/Cores/Threads : Server Configurat		R730xd		730xd	
Processors	2xIntel Xeon H	2xIntel Xeon E5-2650 v3		Intel Xeon E5-2690 v3	
Memory	2.3GHz, 25MI	3 L3	2.6 128GB	6GHz, 30MB L3	
Storage Controller		Del	1280B 1 PERC H730 M	lini	
Storage Device	8 x 1TB,3.5",7			x1.2TB,2.5",10K,6G,SAS	
Network Interface Card	ds Intel x52 Intel x520 DI	0 DP 10Gb P 10GbE + I350 1GbE	Int	el x520 DP 10GbE + I350 DP bbE	
Switches	2x Dell Force1				

		1	Dell		TPCx-HS Revis	sion 1.3.0	
	Dow		-)vd			
(D¢LL)	PowerEdge 730/730xd with			Report Date: October-15-2015			
	Cloudera CDH 5.4.2				October-15-20	-2012	
Description	Part Number	Кеу	Unit Price	Qty	Extended Price	3 yr. Maint. Price	
HARDWARE COMPONENTS							
PowerEdge R730xd Server	210-ADBC	1	\$25,520.00	12	\$306,240.00		
PowerEdge R730/R730xd Motherboard	591-BBCH	1	\$0.00	12			
R730/xd PCle Riser 2, Center	330-BBCO	1	\$0.00	12			
R730/xd PCle Riser 1, Right	330-BBCR	1	\$0.00	12			
Intel X520 DP 10Gb DA/SFP+, + I350	540-BBBB	1	\$0.00	12			
DP 1Gb Ethernet, Network Daughter Card	340-0000		ŞU.UU	12			
Chassis with up to 24, 2.5 Hard Drives and 2, 2.5" Flex Bay Hard Drives	350-BBFE	1	\$0.00	12			
Performance BIOS Settings	384-BBBL	1	\$0.00	12			
UEFI BIOS	800-BBDM	1	\$0.00	12			
No RAID for H330/H730/H730P including Flex Bay Drives (1-24 HDDs or SSDs)	780-BBLS	1	\$0.00	12			
PERC H730 Integrated RAID	405-AAEG	1	\$0.00	12			
Controller, 1GB Cache							
Intel Xeon E5-2690 v3 2.6GHz,30M Cache,9.60GT/s QPI,Turbo,HT,12C/24T (135W) Max Mem 2133MHz	338-BFFL	1	\$0.00	12			
Upgrade to Two Intel Xeon E5-2690 v3 2.6GHz,30M Cache,9.60GT/s QPI,Turbo,HT,12C/24T (135W)	374-BBGS	1	\$0.00	12			
16GB RDIMM, 2133 MT/s, Dual Rank, x4 Data Width	370-ABUG	1	\$0.00	96			
2133MT/s RDIMMs	370-ABUF	1	\$0.00	12			
Performance Optimized	370-AAIP	1	\$0.00	12			
300GB 10K RPM SAS 6Gbps 2.5in Flex Bay Hard Drive,13G	400-AEOC	1	\$0.00	24			
1.2 TB 10K RPM SAS 6Gbps 2.5in Hot-plug Hard Drive,13G	400-AEFO	1	\$0.00	288			
Electronic System Documentation and OpenManage DVD Kit, PowerEdge R730/xd	631-AAJG	1	\$0.00	12			
ReadyRails Sliding Rails With Cable Management Arm	770-BBBR	1	\$0.00	12			
Dual, Hot-plug, Redundant Power Supply (1+1), 750W	450-ADWS	1	\$0.00	12			
C13 to C14, PDU Style, 12 AMP, 2 Feet (.6m) Power Cord, North America	492-BBDH	1	\$0.00	12			
DIMM Blanks for System with 2 Processors	370-ABWE	1	\$0.00	12			

Standard Heat sink for PowerEdge R730/R730xd	374-BBHM	1	\$0.00	12		
Standard Heat sink for PowerEdge R730/R730xd	374-BBHM	1	\$0.00	12		
INFO QS, 13G HADOOP BUNDLE	379-BBWM	1	\$0.00	12		
4hr Response, 24x7 Maintenance			\$2,529.00	12		\$30,348.00
Package, R730xd						
- Thank you choosing Dell	911-6619	1	\$0.00	12		
ProSupport. For tech support, visit						
http://support.dell.com/ProSupport						
- Dell Limited Hardware Warranty	978-4029	1	\$0.00	12		
Plus Service, Initial Year						
- ProSupport Mission Critical	978-4042	1	\$0.00	12		
Package: Enhanced Services,3 Year	070 4042	1	<u> </u>	12		
- ProSupport Mission Critical: 7X24 HW / SW Tech Support and	978-4043	1	\$0.00	12		
Assistance, 3 Year						
- Mission Critical Package: 4-Hours	978-4044	1	\$0.00	12		
7X24 On-Site Service with	570 1011	-	\$0.00			
Emergency Dispatch, 3 Year						
- On-Site Installation Declined	900-9997	1	\$0.00	12		
- US Order	332-1286	1	\$0.00	12		
PowerEdge R730 Server	210-ACXU	1	\$13,026.00	1	\$13,026.00	
PowerEdge R730/R730xd	591-BBCH	1	\$0.00	1	, ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	
Motherboard		_	+0.00	-		
R730/xd PCIe Riser 2, Center	330-BBCO	1	\$0.00	1		
R730 PCIe Riser 3, Left	330-BBCQ	1	\$0.00	1		
R730/xd PCIe Riser 1, Right	330-BBCR	1	\$0.00	1		
Intel X520 DP 10Gb DA/SFP+ Server	540-BBHY	1	\$0.00	1		
Adapter, Low Profile						
Intel X520 DP 10Gb DA/SFP+, + 1350	540-BBBB	1	\$0.00	1		
DP 1Gb Ethernet, Network						
Daughter Card						
Chassis with up to 8, 3.5" Hard	350-BBEO	1	\$0.00	1		
Drives	204 5551					
Performance BIOS Settings	384-BBBL	1	\$0.00	1		
UEFI BIOS	800-BBDM	1	\$0.00	1		
No RAID for H330/H730/H730P (1- 16 HDDs or SSDs)	780-BBJS	1	\$0.00	1		
PERC H730 Integrated RAID	405-AAEG	1	\$0.00	1		
Controller, 1GB Cache		_	+0.00	-		
Intel Xeon E5-2650 v3 2.3GHz,25M	338-BFFF	1	\$0.00	1		
Cache,9.60GT/s						
QPI,Turbo,HT,10C/20T (105W) Max						
Mem 2133MHz						
Upgrade to Two Intel Xeon E5-2650	374-BBGM	1	\$0.00	1		
v3 2.3GHz,25M Cache,9.60GT/s						
QPI,Turbo,HT,10C/20T (105W)		1	\$0.00	8		
16GB RDIMM, 2133 MT/s, Dual Rank, x4 Data Width	370-ABUG	1	\$0.00	ð		
2133MT/s RDIMMs	370-ABUF	1	\$0.00	1		
Performance Optimized	370-ABOI 370-AAIP	1	\$0.00	1		
1TB 7.2K RPM SATA 6Gbps 3.5in				8		
Hot-plug Hard Drive,13G	400-AEEZ	1	\$0.00	0		
Hot-plug Hald DIIVE,130		1		I	I	

		-	•		1	-
Electronic System Documentation	631-AAJG	1	\$0.00	1		
and OpenManage DVD Kit,						
PowerEdge R730/xd						
DVD+/-RW, SATA, Internal	429-AAPS	1	\$0.00	1		
ReadyRails Sliding Rails With Cable	770-BBBR	1	\$0.00	1		
Management Arm						
Dual, Hot-plug, Redundant Power	450-ADWS	1	\$0.00	1		
Supply (1+1), 750W						
C13 to C14, PDU Style, 12 AMP, 2	492-BBDH	1	\$0.00	2		
Feet (.6m) Power Cord, North						
America						
DIMM Blanks for System with 2	370-ABWE	1	\$0.00	1		
Processors						
Standard Heat sink for PowerEdge	374-BBHM	1	\$0.00	1		
R730/R730xd						
Standard Heat sink for PowerEdge	374-BBHM	1	\$0.00	1		
R730/R730xd						
INFO QS, 13G HADOOP BUNDLE	379-BBWM	1	\$0.00	1		
4hr response, 24x7 Maintenance		1	\$2,049.00	1		\$2,049.00
Package, R730						
- Thank you choosing Dell	911-6619	1	\$0.00	1		
ProSupport. For tech support, visit						
http://support.dell.com/ProSupport						
- Dell Limited Hardware Warranty	978-3603	1	\$0.00	1		
Plus Service, Initial Year						
- ProSupport Mission Critical	978-3617	1	\$0.00	1		
Package: Enhanced Services, 3 Year						
- ProSupport Mission Critical: 7X24	978-3618	1	\$0.00	1		
HW / SW Tech Support and						
Assistance,3 Year						
- Mission Critical Package: 4-Hours	978-3619	1	\$0.00	1		
7X24 On-Site Service with						
Emergency Dispatch, 3 Year						
- On-Site Installation Declined	900-9997	1	\$0.00	1		
- US Order	332-1286	1	\$0.00	1		
Dell Force10 S4810 Switch		1	\$21,156.00	2	\$42,312.00	
Force10, S4810P, 1RU, 48 x 10GbE	225-2479	1	\$0.00	2		
SFP+, 4 x 40GbE QSFP+, 1 x AC PSU,			7			
2 x FM, PSU to IO Panels (Reverse)						
Dell Networking, Jumper Cord,	450-AASX	1	\$0.00	2		
250V, 12A, 2 Meters, C13/C14, US			,			
Force10 Customer not deploying	332-0139	1	\$0.00	2		1
this switch in iSCSI or FCOE						
environment.						
INFO QS, 13G HADOOP BUNDLE	379-BBWM	1	\$0.00	2		
Dell Networking, Cable, SFP+ to	470-AAGP	1	\$0.00	30		
SFP+, 10GbE, Copper Twinax Direct			+			
Attach Cable, 3 Meter						
4hr response, 24x7 Maintenance		1	\$3,153.99	2		\$6,307.98
Package, S4810						
- Force10, User Documentation for	331-6279	1	\$0.00	2		
S4810, DAO/BCC						
- SW Support,Force10 Software ,3	935-0103	1	\$0.00	2		
Years						
		•			•	

- ProSupport: 7x24 HW / SW Tech	935-0143	1	\$0.00	2		
Support and Assistance, 3 Years	000.0400					
- Thank you choosing Dell ProSupport. For tech support, visit http://www.dell.com/support or call 1-800- 945-3355	989-3439	1	\$0.00	2		
- Dell Hardware Limited Warranty Initial Year	996-2670	1	\$0.00	2		
- Dell Hardware Limited Warranty Extended Year(s)	996-2760	1	\$0.00	2		
- On-Site Installation Declined	900-9997	1	\$0.00	2		
- Declined Remote Consulting Service	973-2426	1	\$0.00	2		
Dell Netshelter SX 42U Rack - 600mm Wide x 1070mm Deep	A7545497	1	\$1,279.99	1	\$1,279.99	
Logitech MK120 Keyboard and Mouse	A3974709	1	\$19.99	1	\$19.99	
HARDWARE COMPONENTS				Subtotal	\$362,877.98	\$38,704.98
SOFTWARE COMPONENTS						
Red Hat Enterprise Linux,1-2SKT,3yr Premium Subscription,1 Virtual Guest	421-5721	1	\$3,059.00	13	\$39,767.00	
- Red Hat Enterprise Linux Non Factory Install, x64,Reqs Subscription Selection	421-4727	1	\$0.00		\$0.00	
- Red Hat Enterprise Linux 6.5 Media Only X86_64, No Subscription	421-5737	1	\$0.00		\$0.00	
Cloudera Enterprise Basic Edition, Node License, 24x7 3YR	A8208056	1	6,912	13	\$89,856.00	
SOFTWARE COMPONENTS				Subtotal	\$129,623.00	\$0.00
Total					\$492,500.98	\$38,704.98
Large Purchase Discount (35%)*					-172,375.34	-13,546.74
ViewSonic VA2055Sa LED Monitor 20" (includes spares)	3701841	2	\$97.99	3	293.97	
Pricing: 1 - Dell 2 - CDW (www.cdw	.com)				Three-Year Cost of Ownership:	\$345,578
* Discount based upon total system of regular customer.	cost as purchas	sed by a			HSph@1TB:	7.39
Audited by Doug Johnson, Infosizing (www.sizing.com)	Inc.				\$ / HSph@1TB:	46,762.93

Prices used in TPC benchmarks reflect the actual prices a customer would pay for a one-time purchase of the stated components. 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 components. For complete details, see the pricing sections of the TPC benchmark specifications. If you find that the stated prices are not available according to these terms, please inform the TPC at pricing@tpc.org.

DELL

DELL PowerEdge R730/R730xd w/ Cloudera CDH 5.4.2

TPCx-HS Revision 1.3.0

> Report Date: October 15, 2015

Numerical Quantities Summary

Measurement Results for Performance Run				
Scale Factor	1TB			
Run Start Time	2015/08/24 15:44:54			
Run End Time	2015/08/24 15:52:57			
Run Elapsed Time	487.000			
Start of HSGen	2015/08/24 15:44:57			
End of HSGen	2015/08/24 15:46:57			
HSGen Time	123.444			
Start of HSSort	2015/08/24 15:47:01			
End of HSSort	2015/08/24 15:51:43			
HSSort Time	282.430			
Start of HSValidate	2015/08/24 15:51:49			
End of HSValidate	2015/08/24 15:52:57			
HSValidate Time	71.267			

Measurement Results for Repeatability Run				
Scale Factor	1TB			
Run Start Time	2015/08/24 15:54:07			
Run End Time	2014/08/24 16:02:06			
Run Elapsed Time	482.000			
Start of HSGen	2015/08/24 15:54:09			
End of HSGen	2015/08/24 15:56:07			
HSGen Time	120.380			
Start of HSSort	2015/08/24 15:56:12			
End of HSSort	2015/08/24 16:00:48			
HSSort Time	277.511			
Start of HSValidate	2015/08/24 16:00:54			
End of HSValidate	2015/08/24 16:02:06			
HSValidate Time	74.878			



DELL PowerEdge R730/R730xd w/ Cloudera 5.4.2

TPCx-HS Revision 1.3.0

> Report Date: October 15, 2015

Run Report

Full run report is provided in the SupoportingFiles Archive. Summary lines are shown below. _____ Performance Metric (HSph@SF) Report Test Run 1 details: Total Time = 487 Total Size = 1000000000 Scale-Factor = 1.0000 TPCx-HS Performance Metric (HSph@SF): 7.3964 _____ _____ Performance Metric (HSph@SF) Report Test Run 2 details: Total Time = 482 Total Size = 1000000000 Scale-Factor = 1.0000 TPCx-HS Performance Metric (HSph@SF): 7.4738 _____

Abstract

This report documents the methodology and results of the TPC Benchmark TPCx-HS test conducted on a cluster of 13 Dell 13g PowerEdge Servers using Cloudera CDH 5.4.2 in conformance with the requirements of the TPCx-HS Benchmark Specification. The operating system used on each server for the benchmark was Red Hat Enterprise Linux Server 6.5.

Measured Configuration

Hardware	Software	Virtualization
1x Dell PowerEdge R730 with 10-core 2.30GHz Intel Xeon E5 2650v3	Red Hat Enterprise Linux 6.5 Cloudera CDH 5.4.2	N/A
12x Dell PowerEdge R730xd with 12- core 2.60GHz Intel Xeon E5 2690v3		

TPC Express Benchmark[®] HSMetrics

Total System Cost	HSph@1TB	\$/HSph@1TB	Availability Date
\$345,578	7.39	46,762.93	October 15,2015

The Transaction Processing Performance Council (TPC) developed the TPCx-HS Benchmark. The TPC was founded to define transactions processing benchmarks and to disseminate objective, verifiable performance data to the industry.

In order to verify compliance to the TPCx-HS benchmark specification, Doug Johnson audited the benchmark configuration, environment and methodology used to produce and validate the test results, and the pricing model used to calculate the price/performance.

Table of Contents

Contents

CLAUSE1: General Items1
1.1: Test Sponsor1
1.2: Parameter Settings1
1.3: Disclosure Requirements1
1.4: Measured and Priced Configurations1
1.5: Distribution of Data
1.6: Software Components3
1.7: Distributed File Systems 4
1.8: Map/Reduce
Clause 2: Workload Related Items 5
2.1: Scripts5
2.2: Version Number and Checksums5
2.3: Run Report5
2.4: Benchmark Kit Changes6
Clause 3: SUT Related Items7
3.1: Hardware and Software Options7
3.2: Data Storage and Memory Ratios7
Clause 4: Performance Metric and Execution Rules Related Items
Clause 8: Auditor-Related Items9
Auditor's Report9
SUPPORTING FILES 10
PRICE QUOTATIONS

CLAUSE1: General Items

1.1: Test Sponsor

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

DELL is the sponsor of this TPC Benchmark $^{\rm TM}$ TPCx-HS result.

1.2: Parameter Settings

7.4.2 Settings must be provided for all customer-tunable parameters and options that have been changed from the defaults found in actual products, including but 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.

Details of system and Hadoop configurations and parameters are provided in SupportingFiles Archive.

1.3: Disclosure Requirements

7.4.3 Explicit response to individual disclosure requirements specified in the body of earlier sections of this document must be provided.

Not applicable

1.4: Measured and Priced Configurations

7.4.4 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);

Copyright 2015 Dell Computer Inc.

- 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. •

The following sample diagram illustrates a measured benchmark configuration using Ethernet, an external driver, and four processors each with two cores and four threads per node in the SUT. Note that this diagram does not depict or imply any optimal configuration for the TPCx-HS benchmark measurement.

Depending on the implementation of the SUT the Name Node, Job Tracker, Task Tracker, Data Nodes etc or the functional euqivalants must be specified in the diagram.

Comment: Detailed diagrams for system configurations and architectures can vary widely, and it is impossible to provide exact quidelines suitable for all implementations. The intent here is to describe the system components and connections in sufficient detail to allow independent reconstruction of the measurement environment. This example diagram shows homogeneous nodes. This does not preclude tests sponsors from using heterogeneous nodes as long as the system diagram reflects the correct system configuration.

The System Under Test (SUT) comprises 1 x DELL PowerEdge R730 Server (Namenode), 12 x Dell PowerEdge R730xd Servers (Datanodes) and 2x Dell Force 10 S4810 switches, depicted in the next diagram. The Namenode server is named r3s1. The Datanode servers are named r3s1xd1 through r3s1xd12. Each Server consists of:

- 2 x 2.30GHz Intel ® Xeon E5-2650v3 Processors, each with a 25MiB L3 cache and 10x 256KiB L2 caches (one per core), Hyper-Threading enabled, 40 total hardware threads (Namenode)
- 2x 2.60GHz Intel ® Xeon E5-2690v3 Processors, each with a 30MiB L3 cache and 12x 256KiB L2 caches (one ٠ per core), Hyper-Threading enabled, 48 total hardware threads (Datanodes)
- 128GiB ECC DDR3 2133 MHz RAM (All nodes)
- Local storage controller: Dell PERC H730, RAID bus controller: LSI Logic / Symbios Logic MegaRAID SAS-3 ٠ 3108 [Invader] (All nodes)
- Intel Ethernet X520 DP 10GBASE-T + I350 DP 1GbE. Each 10GbE port is connected to one of the switches (All ٠ nodes)
- Intel Ethernet X520 DP 10GBASE-T: 2-port 10GbE, each port connected to one of the switches (Namenode) •
- 24 x 1.2TB 10K RPM SAS 6Gbps 2.5in + 2 x 300GB, 10K RPM SAS , 2.5in (Datanodes) •
- 8 x 1TB, 7.2K RPM SAS 6GGbps 3.5in (Namenode) • 2
 - Copyright 2015 Dell Inc.

Each Server has Red Hat Enterprise Linux Server 6.5 installed natively in a partition on the "root" disk. Log files are written to this partition. The root disk also holds the swap partition.

The rest of the disks on the Datanodes are configured with a single partition each which is formatted with ext4. These are used for all Hadoop data except the log files. The two network ports are bonded together in Linux.

There are no differences between the priced and measured configurations.

1.5: Distribution of Data

7.4.5 The distribution of dataset across all media must be explicitly described using a format similar to that shown in the following example for both the tested and priced systems.

Table 1.5.1: Layout Description. Measured and priced configurations are the same.

Server	Physical Disk Drive	Description of Content
r3s1	0 (8 HDD, RAID10)	Operating system, root, swap
r3s1xd(1-12)	0 (2 HDD, RAID1)	Operating system, root, swap
	1-24	Data

1.6: Software Components

7.4.6 The distribution of various software components across the system must be explicitly described using a format similar to that shown in the following example for both the tested and priced systems.

Table 1.6: Distribution of Software Components. Measured and priced configurations are the same.

Server	Software Component(s)
r3s1	NameNode, benchmark driver, JobTracker,Secondary Namenode
r3s1xd(1-12)	DataNode, TaskTracker

1.7: Distributed File Systems

7.4.7 Distributed file system implementation (e.g. Apache HDFS, Red Hat Storage, IBM GPFS, EMC Isilon OneFS) and corresponding Hadoop File System API version must be disclosed.

Cloudera Distribution for Apache Hadoop (CDH) 5.4.2. Apache HDFS version 2 was used. This is the only version of HDFS supported by CDH 5.4.2.

1.8: Map/Reduce

7.4.8 Map/Reduce implementation (e,g. Apache Map/Reduce, IBM Platform Symphony) and corresponding version must be disclosed

Cloudera Distribution for Apache Hadoop (CDH) 5.4.2. Apache Map/Reduce version 1 was used.

Clause 2: Workload Related Items

2.1: Scripts

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

The tunable parameters involved in this benchmark are contained in the supporting files.

2.2: Version Number and Checksums

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

Version number of the kit used is 1.3.0

md5sum checksums:

58c13ddb98a2d1228f2df10f4a087a71 BigData_cluster_validate_suite.sh

70ba6b440de47b4e4a902bf4983ee4c1 TPCx-HS-master.sh

4ceaefc51c698c0733b57244b7760808 TPCx-HS-master.jar

2.3: Run Report

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

The full output file is given in the SupportingFiles Archive. The summary lines of the 2 runs from that file are:

Performance Metric (HSph@SF) Report

Test Run 1 details: Total Time = 487

Total Size = 1000000000

Scale-Factor = 1.0000

TPCx-HS Performance Metric (HSph@SF): 7.3964

Performance Metric (HSph@SF) Report

Test Run 2 details: Total Time = 482 Total Size = 1000000000 Scale-Factor = 1.0000

TPCx-HS Performance Metric (HSph@SF): 7.4738

2.4: Benchmark Kit Changes

No modifications were made to the TPC provided kit.

Clause 3: SUT Related Items

3.1: Hardware and Software Options

7.6.1 All hardware and software options must be reported.

Hardware and software options are contained in the supporting files

3.2: Data Storage and Memory Ratios

7.6.2 The data storage ratio must be disclosed. It is computed by dividing the total physical data storage present in the priced configuration (expressed in TB) by the chosen Scale Factor as defined in Clause 4.1. Let r be the ratio. The reported value for r must be rounded to the nearest 0.01. That is, reported value=round(r,2). For example, a system configured with 96 disks of 1TB capacity for a 1TB Scale Factor has a data storage ratio of 96.

The Scale Factor to memory ratio must be disclosed. It is computed by dividing the Scale Factor by the total physical memory present in the priced configuration (see clause 3). Let r be this ratio. The reported ratio must be rounded to the nearest 0.1. That is, reported value=round(r,1). For example, a system configured with 1TB of physical memory for a 10TB Scale Factor has a memory ratio of 10.

Total physical data storage for R730 (r3s1) is 8 disks X 1TB = 8TB.

Total physical data storage for R730xd (r3s1xd(1-12) is 12 servers (24 disks X 1.2TB + 2 disks X 300GB)= 352.80

Total Physical Storage = 8TB + 352.80TB = 360.80

Total physical memory is 13 hosts X 128GB = 1664GB = 1.664TB. Scale factor is 1TB.

Data storage ratio is 360.80/1 = 360.80

Scale Factor to memory ratio is 1/1.664 = 0.60

Clause 4: Performance Metric and Execution Rules Related Items

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

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

7.7.3 The HSValidate time must be disclosed for Run1 and Run2.

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

7.7.5 The performance metric (HSph@SF) must be disclosed for Run1 and Run2. Price-performance metric (\$/HSph@SF) must be disclosed for the performance run. See Clause 2.3 and Clause 4.

	Run1	Run2
HSGen	123.444	120.380
HSSort	282.430	277.511
HSValidate	71.267	74.878
HSDataCheck	9.859	9.231
HSph@SF	7.3964	7.4738
\$/HSph@SF	\$46,762.93	

Clause 8: Auditor-Related Items

Auditor's Report

The auditor's agency name, address, phone number, and Attestation letter with a brief audit summary report indicating compliance 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

SUPPORTING FILES

The following table describes the files contained in the supporting files archive.

Clause	Description	Location
Clause 1	Parameters and options used to configure the system	SupportingFilesArchive\Clause1
Clause 2	Configuration Scripts & Run report	SupportingFilesArchive\Clause2
Clause 3	System configuration details	SupportingFilesArchive\Clause3

PRICE QUOTATIONS

