Lenovo

TPC Express Benchmark™ Big Bench (TPCx-BB)

Full Disclosure Report

for

Lenovo Cluster

(with 36x ThinkSystem SR650, 3x ThinkSystem SR630)

using

Cloudera for Apache Hadoop (CDH) 5.12.1

and

Red Hat Enterprise Linux Server 7.6

First Edition

July 12, 2019

***Lenovo**, 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 as a result of these and other factors. Therefore, the TPC Express Benchmark™ BB 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.

Lenovo and the Lenovo Logo are trademarks of Lenovo 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 Lenovo and any other company.

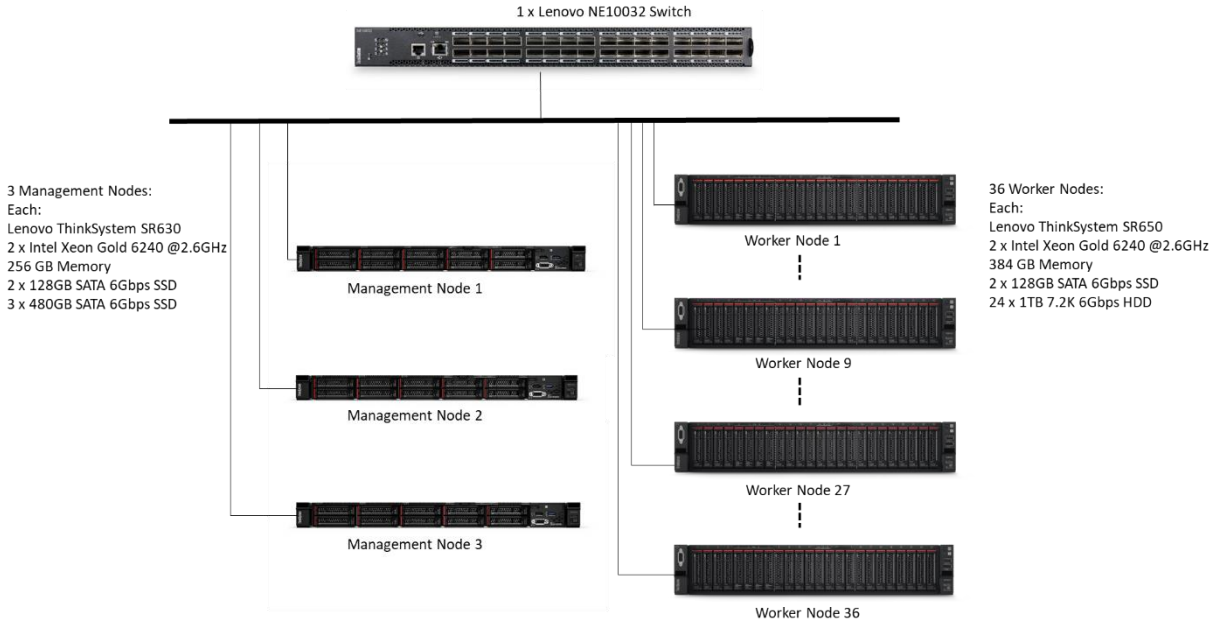
TPC Benchmark™, TPCx-BB and BBQpm, are registered certification marks of the Transaction Processing Performance Council.

The Lenovo 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 Lenovo business contact for information on the products or services available in your area. You can find additional information via Lenovo's web site at www.lenovo.com. Actual performance and environmental costs of Lenovo products will vary depending on individual customer configurations and conditions.

Lenovo	**Lenovo Cluster**	TPCx-BB Rev. v1.2.0 TPC-Pricing Rev. v2.4.0 Report Date: July 12, 2019
Total System Cost	TPCx-BB Performance Metric	Price/Performance
1,433,851 USD	**3,767.91** BBQpm@30000	**380.55 USD** $/BBQpm@30000

Framework	Operating System	Other Software	Availability Date	Scale Factor	Streams
Cloudera for Apache Hadoop (CDH) 5.12.1	Red Hat Enterprise Linux Server 7.6	None	July 12, 2019	30000	4

System Configuration



Physical Storage/Scale Factor: 29.28	Scale Factor/Physical Memory: 2.06

Servers: Total Processors/Cores/Threads	36x ThinkSystem SR650, 3x ThinkSystem SR630 78/1,404/2,808	
Server Configuration: Processors Memory Storage Controller Storage Device Network	**Per ThinkSystem SR650:** 2x Intel Xeon Gold 6240 @ 2.60GHz 384GiB 3x ThinkSystem RAID 430-8i 2x 128GB SATA 6Gbps SSD 24x 1TB 7.2K rpm SATA 6Gbps HDD 1x ThinkSystem 10Gb 4-port	**Per ThinkSystem SR630:** 2x Intel Xeon Gold 6240 @ 2.60 GHz 256GiB ThinkSystem RAID 930-16i 2x 128GB SATA 6Gbps SSD 3x 480GB SATA 6Gbps SSD 1x ThinkSystem 10Gb 4-port
Connectivity:	1x NE 10032 100GbE Switch	

Description	Part Number	Price Source	Unit Price	Quantity	Extended Price	3-Yr. Maint. Price
Server Hardware - Management Nodes						
ThinkSystem SR630 Configure-To-Order, includes:	7X02CTO1WW	1	26,575	3	79,725	
ThinkSystem SR630 2.5" Chassis with 10 Bays	AUW1			3		
Intel Xeon Gold 6240 18C 150W 2.6GHz Processor	B4HH			6		
ThinkSystem 32GB TruDDR4 2933MHz (2Rx4 1.2V) RDIMM	B4H3			24		
ThinkSystem SR630/SR570 2.5" AnyBay 10-Bay Backplane	AUW9			3		
ThinkSystem RAID 930-16i 8GB Flash PCIe 12Gb Adapter	B31E			3		
ThinkSystem 2.5" Intel S4500 480GB Entry SATA 6Gb Hot Swap SSD	B0YZ			9		
ThinkSystem M.2 with Mirroring Enablement Kit	AUMV			3		
ThinkSystem M.2 CV3 128GB SATA 6Gbps Non-Hot Swap SSD	AUUV			6		
ThinkSystem SR530/SR570/SR630 x16 PCIe LP Riser 2 Kit	AUWA			3		
ThinkSystem SR530/SR570/SR630 x8/x16 PCIe LP+LP Riser 1 Kit	AUWC			3		
Lenovo ThinkSystem 1U LP+LP BF Riser Bracket	AUWQ			3		
ThinkSystem 10Gb 4-port SFP+ LOM	AUKK			3		
ThinkSystem 1100W (230V/115V) Platinum Hot-Swap Power Supply	AVWB			6		
2.8m, 10A/100-250V, C13 to C14 Jumper Cord	6311			6		
ThinkSystem Toolless Slide Rail	AXCA			3		
ThinkSystem SR630 Refresh MB	B4NK			3		
FBU345 SuperCap	AUNP			3		
RAID Configuration	2302			3		
Lenovo ThinkSystem 1U LP Riser Bracket	AUWN			3		
Essential Service - 3Yr 24x7 4Hr Response + YourDrive YourData	5PS7A01504	1	1,539	3		4,617
				Subtotal	79,725	4,617
Server Hardware - Worker Nodes						
ThinkSystem SR650 Configure-To-Order, includes:	7X06CTO1WW	1	37,964	36	1,366,704	
ThinkSystem SR650 2.5" Chassis with 8, 16 or 24 bays	AUVV			36		
Intel Xeon Gold 6240 18C 150W 2.6GHz Processor	B4HH			72		
ThinkSystem 32GB TruDDR4 2933MHz (2Rx4 1.2V) RDIMM	B4H3			432		
ThinkSystem 2U/Twr 2.5" SATA/SAS 8-Bay Backplane	AURA			72		
ThinkSystem 2U/Twr 2.5" AnyBay 8-Bay Backplane	AUR5			36		
ThinkSystem 430-8i SAS/SATA 12Gb HBA	AUNL			108		
ThinkSystem 2.5" 1TB 7.2K SATA 6Gb Hot Swap 512n HDD	AUUE			864		
ThinkSystem M.2 with Mirroring Enablement Kit	AUMV			36		
ThinkSystem M.2 CV3 128GB SATA 6Gbps Non-Hot Swap SSD	AUUV			72		
ThinkSystem SR550/SR590/SR650 x16/x8 PCIe FH Riser 1 Kit	AUR3			36		
ThinkSystem 10Gb 4-port SFP+ LOM	AUKK			36		
ThinkSystem 1100W (230V/115V) Platinum Hot-Swap Power Supply	AVWF			72		
2.8m, 13A/100-250V, C13 to C14 Jumper Cord	6400			72		
ThinkSystem Toolless Slide Rail	AXCA			36		
ThinkSystem 2U left EIA Latch Standard	AURD			36		
ThinkSystem SR650 Refresh MB	B4NL			36		
Lenovo ThinkSystem 2U 3FH Riser Bracket	AURQ			36		
Essential Service - 3Yr 24x7 4Hr Response + YourDrive YourData	5PS7A01558	1	2,079	36		74,844
				Subtotal	1,366,704	74,844

(continued on next page)

(continued from previous page)

Description	Part Number	Price Source	Unit Price	Quantity	Extended Price	3-Yr. Maint. Price
Server Software						
RHEL Server Physical or Virtual Node, 2 Skt Prem RH Sup 3Yr	7S0FCTO1WW	1	3,703	39	144,417	
Cloudera Enterprise Data Engineering Edition Node License Gold Support 1yr 24x7	CEDEN-GOLD	2	4,000	117	468,000	
				Subtotal	612,417	0
Network						
NE10032 100GbE RackSwitch : Lenovo ThinkSystem NE10032 RackSwitch	7159HE1	1	28,656	1	28,656	
Lenovo ThinkSystem NE10032 RackSwitch (Rear to Front)	AV17			1		
Adjustable 19" 4 Post Rail Kit	A3KP			1		
Lenovo 3m 100G QSFP28 to 4x25G SFP28 Breakout DAC Cable	AV23			10		
2.8m, 10A/100-250V, C13 to C14 Jumper Cord	6311			2		
Essential Service - 3Yr 24x7 4Hr Response	5WS7A01109	1	3,599	1		3,599
				Subtotal	28,656	3,599
Infrastructure						
Infrastructure Rack : NetBAY S2 42U Rack Cabinet	9307RC4	1	2,349	2	4,698	
0U 36 C13/6 C19 24A/200-240V 1 Phase PDU with NEMA L6-30P line cord	00YJ776	1	479	3	1,437	
ThinkSystem Keyboard w/ Int. Pointing Device USB - US Eng 103P RoHS v2	7ZB7A05470	1	99	1	99	
Lenovo L22e-20 21.5-inch LED Backlit LCD Monitor	65DEKCC1US	1	100	1	100	
Essential Service - 3Yr 24x7 4Hr Response	41L2760	1	315	2		630
				Subtotal	6,334	630
				Total	2,093,836	83,690
Dollar Volume Discount (See Note 1)	43.50%	1			743,675	

Pricing:1 = Lenovo 1-877-782-7134; 2 = Cloudera

(1) Discount applies to all line items where Pricing=1; pricing is for these or similar quantities. Discounts for similarly sized configurations will be similar to what is quoted here but may vary based on the specific components priced.

Audited by Doug Johnson of InfoSizing

Three-Year Cost of Ownership	**$1,433,851**
BBQpm@30000	**3,767.91**
$/BBQpm@30000	**$ 380.55**

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 at pricing@tpc.org. Thank you.

Numerical Quantities

Scale Factor	30000
Streams	4
SUT Validation Test	PASS

Performance Run (Run 2)

Overall Run Start Time	2019-06-26 09:02:12.549
Overall Run End Time	2019-06-27 09:09:55.926
Overall Run Elapsed Time	86,863.377
Load Test Start Time	2019-06-26 09:02:12.550
Load Test End Time	2019-06-26 09:50:33.098
Load Test Elapsed Time	2,900.548
Power Test Start Time	2019-06-26 09:50:33.100
Power Test End Time	2019-06-26 16:58:38.782
Power Test Elapsed Time	25,685.682
Throughput Test Start Time	2019-06-26 16:58:38.782
Throughput Test End Time	2019-06-27 09:09:55.926
Throughput Test Elapsed Time	58,277.144
Performance Metric (BBQpm@ 30000)	3,767.91

Repeatability Run (Run 1)

Overall Run Start Time	2019-06-25 08:07:16.312
Overall Run End Time	2019-06-26 08:05:37.862
Overall Run Elapsed Time	86,301.550
Load Test Start Time	2019-06-25 08:07:16.312
Load Test End Time	2019-06-25 08:56:15.453
Load Test Elapsed Time	2,939.141
Power Test Start Time	2019-06-25 08:56:15.454
Power Test End Time	2019-06-25 16:08:04.584
Power Test Elapsed Time	25,909.130
Throughput Test Start Time	2019-06-25 16:08:04.585
Throughput Test End Time	2019-06-26 08:05:37.862
Throughput Test Elapsed Time	57,453.277
Performance Metric (BBQpm@ 30000)	3,782.64

Performance Run Report (Run 2)

****
TPCx-BB
Result
v1.2
****
INFO: T_LOAD = 2900.548
INFO: T_LD = 0.1 * T_LOAD: 290.0548
INFO: T_PT = 13532.8125368354
INFO: T_T_PUT = 58277.144
INFO: T_TT = 14569.286
INFO: === Checking validity of the final result ===
INFO: OK: All required BigBench phases were performed.
INFO: OK: All 30 queries were running in the power test.
INFO: OK: All 30 queries were running in the first throughput test.
INFO: OK: Pretend mode was inactive. All commands were executed.
INFO: === Final result ===
INFO: VALID BBQpm@30000 = 3767.91226157961

Repeatability Run Report (Run 1)

****
TPCx-BB
Result
v1.2
****
INFO: T_LOAD = 2939.141
INFO: T_LD = 0.1 * T_LOAD: 293.9141
INFO: T_PT = 13610.4037837776
INFO: T_T_PUT = 57453.277
INFO: T_TT = 14363.31925
INFO: === Checking validity of the final result ===
INFO: OK: All required BigBench phases were performed.
INFO: OK: All 30 queries were running in the power test.
INFO: OK: All 30 queries were running in the first throughput test.
INFO: OK: Pretend mode was inactive. All commands were executed.
INFO: === Final result ===
INFO: VALID BBQpm@30000 = 3782.64938631678

Summary details of the run reports are shown above. For the complete run reports, see the Support Files Archive.

Table of Contents

Abstract

This document contains the methodology and results of the TPC Express Benchmark™ Big Bench (TPCx-BB) test conducted in conformance with the requirements of the TPCx-BB Standard Specification, Revision v1.2.0.

The test was conducted at a Scale Factor of 30000 with 36 nodes (ThinkSystem SR650) running Cloudera for Apache Hadoop (CDH) 5.12.1 on Red Hat Enterprise Linux Server 7.6.

Measured Configuration

Company Name	Cluster Node	Virtualization	Operating System
Lenovo	36x ThinkSystem SR650 3x ThinkSystem SR630	n/a	Red Hat Enterprise Linux Server 7.6

TPC Express Benchmark© Big Bench Metrics

Total System Cost	BBQpm@30000	Price/Performance	Availability Date
1,433,851 USD	3,767.91	380.55 USD	July 12, 2019

Preface

TPC Express Benchmark™ Big Bench Overview

Big data analytics is a growing field of research and business. The significant decrease in the overall cost of hardware, the emergence of Open Source based analytics frameworks, along with the greater depth of data mining capabilities allows new types of data sources to be correlated with traditional data sources. For example, online retailers used to record only successful transactions on their website, whereas modern systems are capable of recording every interaction. The former allowed for simple shopping basket analysis techniques, while the current level of detail in monitoring makes detailed user modeling possible. The growing demands on data management systems and the new forms of analysis have led to the development of a new type of ***Big Data Analytics Systems (BDAS)****.*

Similar to the advent of ***Database Management Systems****, there is a vastly growing ecosystem of diverse approaches to enabling Big Data Analytics Systems. This leads to a dilemma for customers of* ***BDAS****, as there are no realistic and proven measures to compare different* ***BDAS*** *solutions. To address this, TPC has developed TPCx-BB (BigBench), which is an express benchmark for comparing* ***BDAS*** *solutions. The TPCx-BB Benchmark was developed to cover essential functional and business aspects of big data use cases. The benchmark allows for an objective measurement of* ***BDAS*** *System under Test, and provides the industry with verifiable performance, price/performance, and availability metrics.*

The TPCx-BB kit is available from the TPC website (see www.tpc.org for more information). Users must sign-up and agree to the TPCx-BB End User Licensing Agreement (EULA) to download the kit. All related work (such as collaterals, papers, derivatives) must acknowledge the TPC and include the TPCx-BB copyright. The TPCx-BB kit includes: TPCx-BB Specification document (this document), TPCx-BB Users Guide documentation, shell scripts to set up the benchmark environment, Java code to execute the benchmark workload, Data Generator, ***Query*** *files, and Benchmark Driver.*

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-BB models and represents a Big Data Analytics System such as Hadoop ecosystem or 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 Lenovo

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 components used by the SUT.*
- *Configuration parameters and options for Operating System and file system components used by the SUT.*
- *Configuration parameters and options for any other software components (e.g compiler optimization options) used by the SUT.*

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.

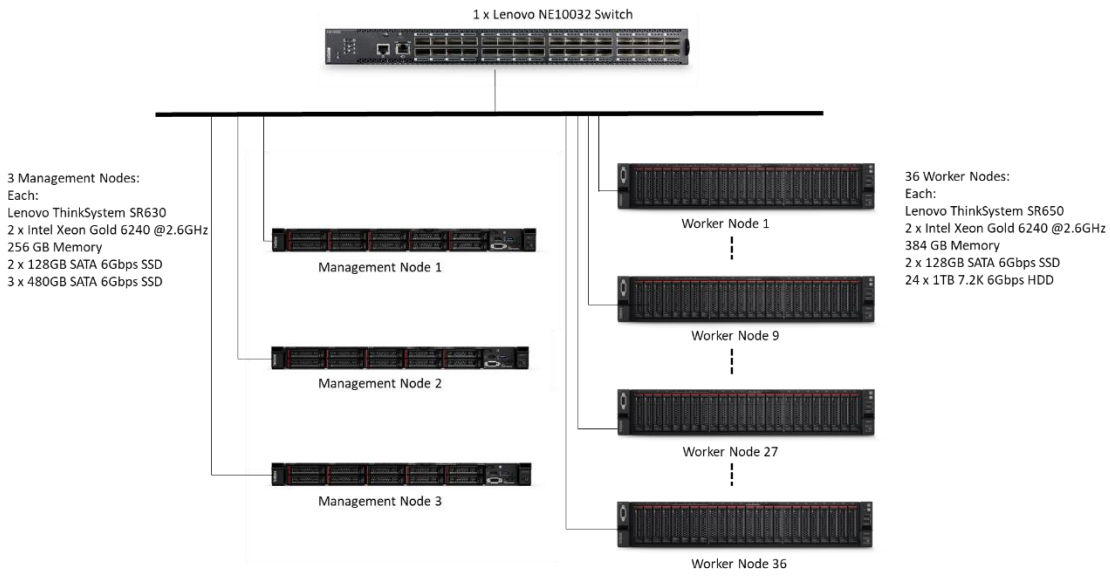
The Supporting Files Archive contains the parameters and options used to configure the components involved in this benchmark.

1.3 Configuration Diagrams

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);*
- *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.*

Measured Configuration



The measured configuration consisted of:

- Total Nodes: 39
- Total Processors/Cores/Threads: 78/1,404/2,808
- Total Memory: 14,592
- Total Number of Storage Drives/Devices: 951
- Total Storage Capacity: 878,304

Server node details:

36x ThinkSystem SR650, each with:

- Processors/Cores/Threads: 2/36/72
- Processor Model: 2x Intel Xeon Gold 6240 @ 2.60GHz
- Memory: 384GiB
- Controller: 3x ThinkSystem RAID 430-8i
- Drives:
  - 2x 128GB SATA 6Gbps SSD
  - 24x 1TB 7.2K rpm SATA 6Gbps HDD
- Network: 1x ThinkSystem 10Gb 4-port

3x ThinkSystem SR630, each with:

- Processors/Cores/Threads: 2/36/72
- Processor Model: 2x Intel Xeon Gold 6240 @ 2.60GHz
- Memory: 256GiB
- Controller: 1x ThinkSystem RAID 930-16i
- Drives:
  - 2x 128GB SATA 6Gbps SSD
  - 3x 480GB SATA 6Gbps SSD
- Network: 1x ThinkSystem 10Gb 4-port

Network connectivity details:

- 1x Lenovo NE 10032 Switch

The distribution of software components over server nodes is detailed in section 2.1.

Priced Configuration

There are no differences between the priced and measured configurations.

Clause 2: Software Components and Dataset Distribution

2.1 Roles and Dataset Distribution

The distribution of dataset across all media must be explicitly described.
The distribution of various software components across the system must be explicitly described.

Table 1.4 describes the distribution of the dataset across all media in the system.

Table 1.4: Software Components and Dataset Distribution

Server	Role(s)	Count	Virtual	Host Names	HW/SW Configuration	Storage Setup
Cloudera Manager 01	HDFS Balancer/ HDFS NameNode/ Hive Gateway/ Hive Metastore Server/ Spark Gateway/ Spark History Server/ YARN JobHistory Server/YARN ResourceManager/ ZooKeeper Server	1	N	bb-mn01	Lenovo ThinkSystem SR630 HW/SW Config (Intel Xeon Gold 6240, 2, 2.60GHz, 72) Memory: 256 GB Storage: 2 x 128 GB SSD, 3 x 480 GB SSD, Network: 1x Thinksystem 10Gb 4-port SFP+ LOM OS: RHEL 7.6 Cloudera CDH: 5.12.1	OS:2x 128GB SATA 6Gbps SSD Data Drive: 3x 480GB SATA 6Gbps SSD
Cloudera Manager 02	/HDFS Secondary NameNode/ Hive Gateway/ HiveServer2/ Spark Gateway/ ZooKeeper Server	1	N	bb-mn02	Lenovo ThinkSystem SR630 HW/SW Config (Intel Xeon Gold 6240, 2, 2.60GHz, 72) Memory: 256 GB Storage: 2 x 128 GB SSD, 3 x 480 GB SSD, Network: 1x Thinksystem 10Gb 4-port SFP+ LOM OS: RHEL 7.6 Cloudera CDH: 5.12.1	OS:2x 128GB SATA 6Gbps SSD Data Drive: 3x 480GB SATA 6Gbps SSD
Cloudera Manager 03	Activity Monitor/Alert Publisher/Event Server/Host Monitor/Reports Manager/Service Monitor/ZooKeeper Server	1	N	bb-mn03	Lenovo ThinkSystem SR630 HW/SW Config (Intel Xeon Gold 6240, 2, 2.60GHz, 72) Memory: 256 GB Storage: 2 x 128 GB SSD, 3 x 480 GB SSD, Network: 1x Thinksystem 10Gb 4-port SFP+ LOM OS: RHEL 7.6 Cloudera CDH: 5.12.1	OS and Data Drive: 2x 128GB SATA 6Gbps SSD
Cloudera Manager Worker Node	HDFS Data Node/ Hive Gateway/ Spark Gateway/ Yarn Node Manager	1	N	bb-wn[01-36]	Lenovo ThinkSystem SR650 HW/SW Config (Intel Xeon Gold 6240, 2, 2.60GHz, 72) Memory: 384 GB Storage: 2 x 128 GB SSD, 24 x 1TB 7200 rpm 6Gb HDD Network: 1x Thinksystem 10Gb 4-port SFP+ LOM OS: RHEL 7.6 Cloudera CDH: 5.12.1	OS:2x 128GB SATA 6Gbps SSD Data Drive: 24x 1TB 7200 rpm 6Gbps HDD

2.2 Distributed File System Implementation

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

Cloudera for Apache Hadoop (CDH) 5.12.1 (fully HDFS compatible at the API level).

2.3 Engine Implementation

The Engine implementation and corresponding version must be disclosed.

Component	Version
HDFS	2.6.0
Hive	1.1.0
Spark	1.6.0
ZooKeeper	3.4.5
MapReduce	2.6.0

2.4 Frameworks

Frameworks and Engine used in the benchmark should be disclosed.

Framework	Version
CDH	5.12.1
HDFS	2.6.0
Hive	1.1.0
Spark	1.6.0
YARN	2.6.0
ZooKeeper	3.4.5
MapReduce	2.6.0

2.5 Applied Patches

Any additional vendor supported patches applied to the SUT should be disclosed.

No additional patches were applied.

Clause 3: Workload Related Items

3.1 Hardware & Software Tunable

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

The Supporting Files Archive contains all configuration scripts.

3.2 Kit Version

Version number of the TPCx-BB kit must be included in the Report.

TPCx-BB Kit Version
v1.2

3.3 Run Report

The run report generated by TPCx-BB benchmark kit must be included in the Report.

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

- **Run1 Report Summary (Repeatability Run)**

```
**********
TPCx-BB
Result
v1.2
**********
INFO: T_LOAD = 2939.141
INFO: T_LD = 0.1 * T_LOAD: 293.9141
INFO: T_PT = 13610.4037837776
INFO: T_T_PUT = 57453.277
INFO: T_TT = 14363.31925
INFO: === Checking validity of the final result ===
INFO: OK: All required BigBench phases were performed.
INFO: OK: All 30 queries were running in the power test.
INFO: OK: All 30 queries were running in the first throughput test.
INFO: OK: Pretend mode was inactive. All commands were executed.
INFO: === Final result ===
INFO: VALID BBQpm@30000 = 3782.64938631678
```

- **Run2 Report Summary (Performance Run)**

```
**********
TPCx-BB
Result
v1.2
**********
INFO: T_LOAD = 2900.548
INFO: T_LD = 0.1 * T_LOAD: 290.0548
INFO: T_PT = 13532.8125368354
INFO: T_T_PUT = 58277.144
INFO: T_TT = 14569.286
INFO: === Checking validity of the final result ===
INFO: OK: All required BigBench phases were performed.
INFO: OK: All 30 queries were running in the power test.
INFO: OK: All 30 queries were running in the first throughput test.
INFO: OK: Pretend mode was inactive. All commands were executed.
INFO: === Final result ===
INFO: VALID BBQpm@30000 = 3767.91226157961
```

3.4 Query Elapsed Times

Elapsed times of all power and throughput Queries needs to be reported from the Performance Run, grouped respectively as Structured, semi-structured and unstructured buckets.

Type	Query	Power	Stream 1	Stream 2	Stream 3	Stream 4
Structured	1	111.757	164.546	1,407.434	124.899	476.917
	6	403.985	1,655.260	1,476.360	407.570	1,302.527
	7	322.128	956.454	491.791	325.305	825.665
	9	288.779	452.949	1,060.516	3,409.270	1,531.206
	11	124.863	200.763	563.473	512.259	307.267
	13	188.961	226.896	1,093.780	505.986	190.204
	14	248.783	1,037.134	1,571.110	1,376.497	1,170.593
	15	93.574	156.782	97.501	89.184	296.949
	16	465.812	2,239.663	907.095	1,547.736	491.533
	17	167.000	565.037	840.168	311.536	192.567
	20	310.982	1,003.875	994.999	1,165.284	839.048
	21	764.476	2,315.017	1,812.201	2,395.293	2,638.617
	22	75.335	105.154	93.250	413.202	151.243
	23	289.510	824.050	686.829	725.914	883.392
	24	125.276	534.332	613.987	923.301	393.391
	25	475.239	1,260.125	1,410.171	1,611.727	1,614.055
	26	326.506	922.120	974.769	929.126	692.653
	29	189.327	498.291	450.944	191.213	1,117.017
Semi-structured	2	2,585.434	7,242.557	3,752.121	3,803.061	6,630.597
	3	1,580.468	2,480.495	2,244.705	1,737.644	4,842.220
	4	2,396.387	6,177.037	7,571.718	4,505.806	6,364.091
	5	576.125	1,539.887	720.957	1,177.258	791.052
	8	697.132	1,350.241	776.394	3,500.604	833.173
	12	1,149.141	2,961.983	2,297.517	4,363.913	1,452.010
	30	2,349.461	6,429.384	6,047.651	7,375.283	3,004.193
Unstructured	10	1,656.928	3,122.356	3,080.528	3,178.597	2,935.479
	18	5,099.544	7,144.782	10,947.452	7,872.702	7,264.553
	19	1,503.418	2,293.284	1,540.169	1,911.927	1,923.089
	27	89.245	90.220	341.488	76.935	124.606
	28	1,030.078	1,454.521	1,649.301	1,808.097	2,407.827

3.5 Validation Test Output

Output report from successful SUT Validation test must be included in the Report.

Query Number	Query Execution	Output Validation
1	PASS	PASS
2	PASS	PASS
3	PASS	PASS
4	PASS	PASS
5	PASS	PASS
6	PASS	PASS
7	PASS	PASS
8	PASS	PASS
9	PASS	PASS
10	PASS	PASS
11	PASS	PASS
12	PASS	PASS
13	PASS	PASS
14	PASS	PASS
15	PASS	PASS
16	PASS	PASS
17	PASS	PASS
18	PASS	PASS
19	PASS	PASS
20	PASS	PASS
21	PASS	PASS
22	PASS	PASS
23	PASS	PASS
24	PASS	PASS
25	PASS	PASS
26	PASS	PASS
27	PASS	PASS
28	PASS	PASS
29	PASS	PASS
30	PASS	PASS

3.6 Global Framework Parameters

Global Framework parameter settings files must be included in the Report.

The Supporting File Archive contains the global framework parameter settings files.

3.7 Kit Modifications

Test Sponsor kit modifications files must be included in the Report.

The following files were modified by the Test Sponsor to facilitate system, platform and Framework differences.

- bigBench-configs/conf/bigBench.properties
- bigBench-configs/conf/userSettings.conf
- bigBench-configs/hive/conf/engineSettings.conf
- bigBench-configs/hive/conf/engineSettings.sql
- bigBench-configs/hive/queries/q01/engineLocalSettings.sql
- bigBench-configs/hive/queries/q02/engineLocalSettings.sql
- bigBench-configs/hive/queries/q03/engineLocalSettings.sql
- bigBench-configs/hive/queries/q04/engineLocalSettings.sql
- bigBench-configs/hive/queries/q05/engineLocalSettings.sql
- bigBench-configs/hive/queries/q06/engineLocalSettings.sql
- bigBench-configs/hive/queries/q07/engineLocalSettings.sql
- bigBench-configs/hive/queries/q08/engineLocalSettings.sql
- bigBench-configs/hive/queries/q09/engineLocalSettings.sql
- bigBench-configs/hive/queries/q10/engineLocalSettings.sql
- bigBench-configs/hive/queries/q11/engineLocalSettings.sql
- bigBench-configs/hive/queries/q12/engineLocalSettings.sql
- bigBench-configs/hive/queries/q13/engineLocalSettings.sql
- bigBench-configs/hive/queries/q14/engineLocalSettings.sql
- bigBench-configs/hive/queries/q15/engineLocalSettings.sql
- bigBench-configs/hive/queries/q16/engineLocalSettings.sql
- bigBench-configs/hive/queries/q17/engineLocalSettings.sql
- bigBench-configs/hive/queries/q18/engineLocalSettings.sql
- bigBench-configs/hive/queries/q19/engineLocalSettings.sql
- bigBench-configs/hive/queries/q20/engineLocalSettings.sql
- bigBench-configs/hive/queries/q21/engineLocalSettings.sql
- bigBench-configs/hive/queries/q22/engineLocalSettings.sql
- bigBench-configs/hive/queries/q23/engineLocalSettings.sql
- bigBench-configs/hive/queries/q24/engineLocalSettings.sql
- bigBench-configs/hive/queries/q25/engineLocalSettings.sql
- bigBench-configs/hive/queries/q26/engineLocalSettings.sql
- bigBench-configs/hive/queries/q27/engineLocalSettings.sql
- bigBench-configs/hive/queries/q28/engineLocalSettings.sql
- bigBench-configs/hive/queries/q29/engineLocalSettings.sql
- bigBench-configs/hive/queries/q30/engineLocalSettings.sql

Clause 4: SUT Related Items

4.1 Specialized Hardware/Software

Specialized Hardware/Software used in the SUT must be included.

No specialized hardware or software was used.

4.2 Framework Configuration Files

All Framework configuration files from SUT, for the performance run.

All Framework configuration files are included in the Supporting Files Archive.

4.3 SUT Environment Information

SUT environment info in form of envinfo.log from a representative worker node form every role in the server.

All envinfo.log files are include in the Supporting Files Archive.

4.4 Data Storage to Scale Factor Ratio

The data storage ratio must be disclosed.

Nodes	Disks	Size (GB)	Total (GB)
39	2	128	9,984
3	3	480	4,320
36	24	1,000	864,000

Total Storage (GB)	878,304
Scale Factor	30000
Data Storage Ratio	29.28

4.5 Scale Factor to Memory Ratio

The Scale Factor to memory ratio must be disclosed.

Nodes	Memory (GB)	Total (GB)
3	256	768
36	384	13,824

Scale Factor	30000
Total Memory (GB)	14,592
SF / Memory Ratio	2.06

Clause 5: Metrics and Scale Factors

5.1 Performance Run Metric

The Reported Performance Metric (BBQpm@SF for the Performance Run) must be disclosed in the Report.

Performance Run
BBQpm@30000 3,767.91

5.2 Repeatability Run Metric

The Performance Metric (BBQpm@SF) for the Repeatability Run must be disclosed in the Report..

Repeatability Run
BBQpm@30000 3,782.64

5.3 Price-Performance Metric

The Reported Performance Metric (BBQpm@SF for the Performance Run) must be disclosed in the Report.

Price / Performance
$BBQpm@30000 380.55

5.4 Scale Factor

The Scale Factor used for the Result must be disclosed in the Report.

Scale Factor
30000

5.5 Stream Count

The number of streams in the throughput run used for the Result must be disclosed in the Report.

Streams
4

5.6 Elapsed Run Times

The total elapsed time for the execution of the Performance Run and Repeatability Run must be disclosed in the Report.

Run	Elapsed Time	Seconds
Run 1	00 23:58:21.550	86,301.550
Run 2	01 00:07:43.377	86,863.377

5.7 Elapsed Test Times

The total time for each of the three tests must be disclosed for the Performance Run and the Repeatability Run.

Test	Performance Run	Repeatability Run
Load Test	2,900.548	2,939.141
Power Test	25,685.682	25,909.130
Throughput Test	58,277.144	57,453.277

Auditors' Information and Attestation Letter

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 01453
978-343-6562.

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

A copy of the auditor's attestation letter is included in the next two pages.

InfoSizing
The Right Metric For Sizing IT

TPC Transaction Processing Performance Council
Certified Auditors

William Milani
Director DCG Product Development
Lenovo Data Center Group
7001 Development Drive
Morrisville, NC 27560

July 12, 2019

I verified the TPC Express Benchmark™ BB v1.2.0 performance of the following configuration:

Platform:	Lenovo Cluster (w/ 36x ThinkSystem SR650, 3x ThinkSystem SR630)
Operating System:	Red Hat Enterprise Linux Server 7.6
Apache Hadoop Compatible Software:	Cloudera for Apache Hadoop (CDH) 5.12.1

The results were:

Performance Metric	**3,767.91 BBQpm@30000**
Run Elapsed Time	01 00:07:43.377 (86,863.377 Seconds)

Cluster	**36x ThinkSystem SR650 (Data nodes), 3x ThinkSystem SR630 (Management nodes)**
CPUs	2 x Intel Xeon Gold 6240 (2.60 GHz, 18-core, 24.75 MB L3) (All nodes)
Memory	384GiB (Data nodes), 256GiB (Mgmt. nodes)
Storage	***Qty*** ***Size*** ***Type***
	2 128GB 6G SATA SSD (All nodes)
	3 480GB 6G SATA SSD (Mgmt. nodes)
	24 1TB 6G SATA 7.2K HDD (Data nodes)

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 v1.2.0
- No modifications were made to any of the Java code
- Any and all modifications to shell scripts were reviewed for compliance
- The tested Scale Factor (30000GB) was confirmed to be valid for publication

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

- All validation queries executed successfully and produced compliant results
- No errors were reported during the run
- 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:

From the TPCx-BB Kit's README:

Q28 Depending on the Hadoop distribution version can fail automated Engine Validation due to empty space characters when the output is written to HDFS. Manually open the result file and validate the reference values and written values.

Query 28 failed automated Engine Validation. A manual validation was performed as part of this audit to confirm the only differences were due to white space.

Respectfully Yours,

Doug Johnson, TPC Auditor

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

Third Party Price Quotes

Cloudera

cloudera

Quote #:	Q-XXXXX
Date:	X-X-20XX
Offer Expires On:	X-X-20XX
End User:	XXXXXXX
Subscription Bill Terms:	

395 Page Mill Road
Palo Alto, CA 94306 US
Phone: +1 (650) 843-0595
Email: mashburn@cloudera.com

Bill To
XXXXXXXX
XXXXXXXX
XXXXXXXX
United States

Ship To
XXXXXXXX
XXXXXXXX
XXXXXXXX
United States

Product Name	Product Code	Description	List Price	Qty	Start Date	End Date	Discount (%)	Net Total
Cloudera Enterprise Data Engineering Edition, Node License, Gold Support	CEDEN-GOLD	Subscription (per node) for the Data Engineering Edition of the Cloudera Enterprise platform. Includes a commercial license and Gold Support for "Core Hadoop" (i.e. HDFS, YARN/ Mapreduce, Hive, Pig, Hue, Sentry, Flume, Sqoop, Kafka), Spark, Search, Cloudera Manager, Cloudera Director, and Cloudera Navigator (Audit & Lineage, Encryption & Key Trustee). Gold-Level Support provides technical support 24 hours per day, 7 days per week. The price listed is for an annual (12-month) term.	USD 4,000.00	39	5-30-2019	5-29-2022	0.00	USD 468,000.00
							TOTAL:	USD 468,000.00

This Order Form, when signed by Customer or referenced on Customer's purchase order, incorporates and is governed by the applicable terms and conditions located at the following URL for the specific products and/or services set forth herein: http://www.cloudera.com/legal/commercial-terms-and-conditions.html.

The governing agreement referenced herein shall prevail over any additional, conflicting or inconsistent terms and conditions which may appear on any purchase order furnished by Customer, and any additional terms and conditions in any such purchase order shall have no force and effect, notwithstanding Cloudera's acceptance or execution of such purchase order.

This Order Form may be executed by exchange of signature by electronic means through facsimile or scanned and emailed signature, or by electronic signature service where legally permitted. For clarity, an electronic, digital, machine-generated or image of a signature will create a valid and binding obligation of Customer.

Signature: _____ **Effective Date:** _____

Name (Print): _____ **Title:** _____

Page 1

Supporting File Index

The following index outlines the information included in the supporting files archive.

Description	Archive File Pathname
Clause 1 - General Items	
The Supporting Files Archive contains the parameters and options used to configure the components involved in this benchmark	Support-Files-for-Lenovo-39nodes-30TB
Validation Run Files	Support-Files-for-Lenovo -39nodes-30TB\Validation-logs-20190625-071939-hive-sf30000
Performance Run Files	Support-Files-for-Lenovo -39nodes-30TB\Performance-logs-20190627-091149-hive-sf30000
Repeatability Run Files	Support-Files-for-Lenovo -39nodes-30TB\Repeatability-logs-20190626-080731-hive-sf30000
Clause 3 - Workload Related Items	
Benchmark Generic Parameters	Support-Files-for-Lenovo -39nodes-30TB\Performance-logs-20190627-091149-hive-sf30000\bigBench-configs\conf\userSettings.conf
Query Parameters Used in the benchmark execution Settings	Support-Files-for-Lenovo -39nodes-30TB\Performance-logs-20190627-091149-hive-sf30000\bigBench-configs\hive\conf\queryParameters.sql
Benchmark Global Framework Parameters Settings	Support-Files-for-Lenovo -39nodes-30TB\Performance-logs-20190627-091149-hive-sf30000\bigBench-configs\hive\conf\engineSettings.sql
Benchmark Global Framework Parameters Settings	Support-Files-for-Lenovo -39nodes-30TB\Performance-logs-20190627-091149-hive-sf30000\bigBench-configs\hive\conf\engineSettings.conf
Load Test Script	Support-Files-for-Lenovo -39nodes-30TB\Performance-logs-20190627-091149-hive-sf30000\bigBench-configs\hive\population\hiveCreateLoad.sql
Queries specific optimization parameters settings	Support-Files-for-Lenovo -39nodes-30TB\Performance-logs-20190627-091149-hive-sf30000\bigBench-configs\hive\queries\q[01-36]\engineLocalSettings.conf
Queries specific optimization parameters settings	Support-Files-for-Lenovo -39nodes-30TB\Performance-logs-20190627-091149-hive-sf30000\bigBench-configs\hive\queries\q[01-36]\engineLocalSettings.sql
Query #28 validation output	Support-Files-for-Lenovo -39nodes-30TB\query28_validation_run_output_result
Clause 4 - SUT Related Items	
Data Redundancy report	Support-Files-for-Lenovo -39nodes-30TB\hdfs-data-redundancy-report.txt
Benchmark execution script	Support-Files-for-Lenovo -39nodes-30TB\publication-runs.sh
All Framework configuration files are included in the Supporting Files Archive	Support-Files-for-Lenovo -39nodes-30TB\Performance-logs-20190627-091149-hive-sf30000\run-logs\envInfo-bb-wn01\hadoop
	Support-Files-for-Lenovo -39nodes-30TB\Performance-logs-20190627-091149-hive-sf30000\run-logs\envInfo-bb-wn01\hive
	Support-Files-for-Lenovo -39nodes-30TB\Performance-logs-20190627-091149-hive-sf30000\run-logs\envInfo-bb-wn01\spark
Clause 5 - Metric and Scale Factor Related Items	
Benchmark Performance Report	Support-Files-for-Lenovo -39nodes-30TB\Performance-logs-20190627-091149-hive-sf30000\run-logs\BigBenchResult.log
Validation Test Report	Support-Files-for-Lenovo -39nodes-30TB\Validation-logs-20190625-071939-hive-sf30000\run-logs\BigBenchResult.log