

Alibaba Cloud

TPC Express Benchmark™ Big Bench (TPCx-BB)

Full Disclosure Report

for

Alibaba Cloud ApsaraCompute Big Data Accelerator

(with 1x Aliyun ecs.g7.16xlarge; 10x Aliyun ecs.g7.16xlarge)

using

Horton Works HDP 3.0.1

and

CentOS Linux release 7.7.1908

First Edition

October 19, 2021

Alibaba Cloud (Alibaba), 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.


Alibaba and the Alibaba Logo are trademarks of Alibaba Cloud 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 Alibaba and any other company.

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

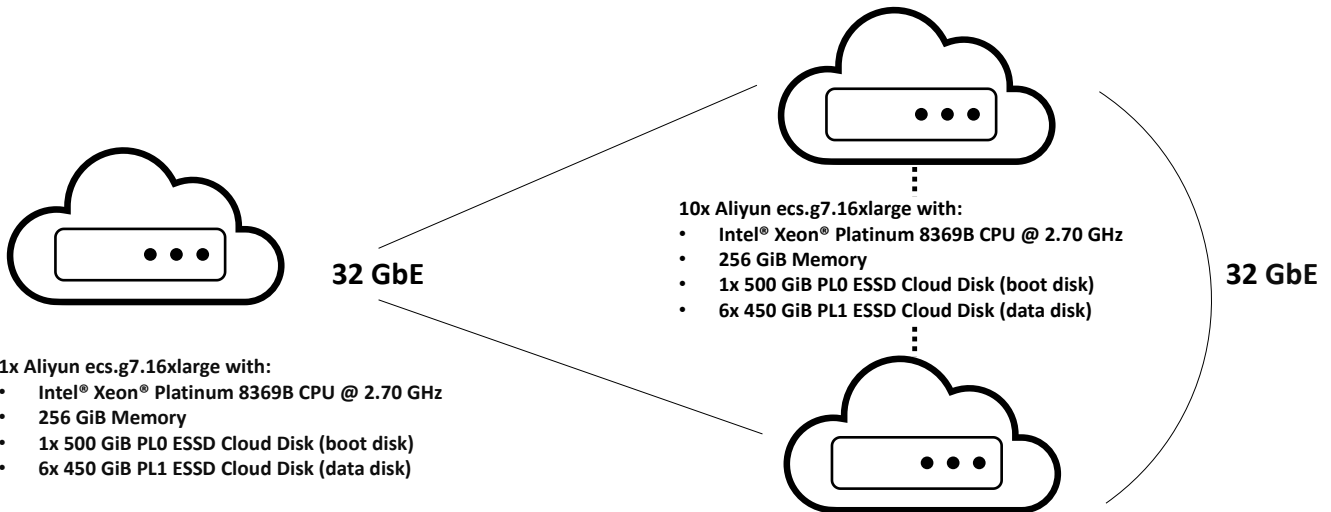
The Alibaba products, services or features identified in this document may not yet be available or may not be available in all areas and may be subject to change without notice. Consult your local Alibaba business contact for information on the products or services available in your area. You can find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

Copyright © 2021 Alibaba Cloud

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.

		Alibaba Cloud ApsaraCompute Big Data Accelerator		TPCx-BB Rev. v1.5.0 TPC-Pricing Rev. v2.7.0	
				Report Date: October 19, 2021	
Total System Cost		TPCx-BB Performance Metric		Price/Performance	
758,001 USD		2,187.42 BBQpm@3000		346.53 USD \$/BBQpm@3000	
Framework	Operating System	Other Software	Availability Date	Scale Factor	Streams
Horton Works HDP 3.0.1	CentOS Linux release 7.7.1908	None	October 19, 2021	3000	8

System Configuration



Physical Storage/Scale Factor: 12.60 Scale Factor/Physical Memory: 1.07		Main Data Redundancy Model 3-Way Replication	
Servers: Total Processors/Cores/Threads		1x Aliyun ecs.g7.16xlarge / 10x Aliyun ecs.g7.16xlarge 11/352/704	
Server Configuration: Processors Memory Storage Device Network Controller	1x Aliyun ecs.g7.16xlarge (Admin): 1x Intel® Xeon® Platinum 8369B CPU @ 2.70GHz 256 GiB 1x 500 GiB PL0 ESSD Cloud Disk 6x 450 GiB PL1 ESSD Cloud Disk Bandwidth: 32 Gbps, Packet forwarding rate 12.000.000	10x Aliyun ecs.g7.16xlarge (Worker): 1x Intel® Xeon® Platinum 8369B CPU @ 2.70GHz 256 GiB 1x 500 GiB PL0 ESSD Cloud Disk 6x 450 GiB PL1 ESSD Cloud Disk Bandwidth: 32 Gbps, Packet forwarding rate 12.000.000	
Connectivity:		Cloud Infrastructure	



Alibaba Cloud Apsara Compute Big Data Accelerator

TPCx-BB Rev. v1.5.0
TPC-Pricing Rev. v2.7.0

Report Date:
October 19, 2021

Description	Part Number	Source	Unit Price	Qty	Ext. Price	3-Year Maint.
Licensed Compute Services						
ECS Instance ecs.g7.16xlarge	ecs.g7.16xlarge (China Beijing)		1 \$49,245.26	11	\$541,697.86	included
ECS System Disk (PL0 ESSD Cloud Disk 500 GiB)	Option		1 \$1,393.95	11	\$15,333.45	included
ECS Data Disk (PL1 ESSD Cloud Disk 450 GiB)	Option		1 \$2,509.10	66	\$165,600.60	included
Licensed Compute Services Subtotal					\$722,631.91	
Software Components						
Horton Works HDP Enterprise Plus, 24x7 1yr	N/A		2 \$11,200.00	33	\$369,600.00	
Software Components Subtotal					\$369,600.00	
Total Cost					\$1,092,232.00	
LCS Discount					-\$334,231.06	

Pricing: 1 = Alibaba; 2 = Cloudera

⁽¹⁾ All discounts are based on US list prices and for similar quantities and configurations. The discounts are based on the overall specific components pricing from respective vendors in this single quotation. Discounts for similarly sized configurations will be similar to those quoted here but may vary based on the components in the configuration.

Audited by Doug Johnson, InfoSizing

Three-Year Cost of Ownership	\$758,001
BBQpm@3000	2,187.42
\$/BBQpm@3000	\$ 346.53

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	3000
Streams	8
SUT Validation Test	PASS

Performance Run (Run 2)

Overall Run Start Time	2021-10-05 19:32:01.748
Overall Run End Time	2021-10-06 01:29:23.181
Overall Run Elapsed Time	21,441.433
Load Test Start Time	2021-10-05 19:32:01.749
Load Test End Time	2021-10-05 19:44:29.009
Load Test Elapsed Time	747.260
Power Test Start Time	2021-10-05 19:44:29.010
Power Test End Time	2021-10-05 20:55:07.083
Power Test Elapsed Time	4,238.073
Throughput Test Start Time	2021-10-05 20:55:07.083
Throughput Test End Time	2021-10-06 01:29:23.181
Throughput Test Elapsed Time	16,456.098
Performance Metric (BBQpm@ 3000)	2,187.42

Repeatability Run (Run 1)

Overall Run Start Time	2021-10-05 13:13:02.979
Overall Run End Time	2021-10-05 19:09:18.932
Overall Run Elapsed Time	21,375.953
Load Test Start Time	2021-10-05 13:13:02.980
Load Test End Time	2021-10-05 13:25:39.379
Load Test Elapsed Time	756.399
Power Test Start Time	2021-10-05 13:25:39.380
Power Test End Time	2021-10-05 14:35:51.122
Power Test Elapsed Time	4,211.742
Throughput Test Start Time	2021-10-05 14:35:51.122
Throughput Test End Time	2021-10-05 19:09:18.932
Throughput Test Elapsed Time	16,407.810
Performance Metric (BBQpm@ 3000)	2,196.83

Performance Run Report (Run 2)

```
*****
TPCx-BB
Result
v1.5.0
*****
INFO: T_LOAD = 747.26
INFO: T_LD = 0.1 * T_LOAD: 74.726
INFO: T_PT = 2786.02808204629
INFO: T_T_PUT = 16456.098
INFO: T_TT = 2057.01225
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@3000 = 2187.42634689913
```

Repeatability Run Report (Run 1)

```
*****
TPCx-BB
Result
v1.5.0
*****
INFO: T_LOAD = 756.399
INFO: T_LD = 0.1 * T_LOAD: 75.6399
INFO: T_PT = 2767.4761023159
INFO: T_T_PUT = 16407.81
INFO: T_TT = 2050.97625
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@3000 = 2196.83525587891
```

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

Table of Contents

ABSTRACT	8
PREFACE	9
CLAUSE 1: GENERAL ITEMS.....	10
1.1 TEST SPONSOR.....	10
1.2 PARAMETER SETTINGS	10
1.3 CONFIGURATION DIAGRAMS.....	10
CLAUSE 2: SOFTWARE COMPONENTS AND DATASET DISTRIBUTION	12
2.1 ROLES AND DATASET DISTRIBUTION.....	12
2.2 DISTRIBUTED FILE SYSTEM IMPLEMENTATION	12
2.3 ENGINE IMPLEMENTATION	13
2.4 FRAMEWORKS	13
2.5 APPLIED PATCHES	13
CLAUSE 3: WORKLOAD RELATED ITEMS	14
3.1 HARDWARE & SOFTWARE TUNABLE	14
3.2 KIT VERSION	14
3.3 RUN REPORT.....	14
3.4 QUERY ELAPSED TIMES	15
3.5 VALIDATION TEST OUTPUT.....	16
3.6 GLOBAL FRAMEWORK PARAMETERS.....	16
3.7 KIT MODIFICATIONS.....	16
CLAUSE 4: SUT RELATED ITEMS.....	18
4.1 SPECIALIZED HARDWARE/SOFTWARE	18
4.2 FRAMEWORK CONFIGURATION FILES	18
4.3 SUT ENVIRONMENT INFORMATION	18
4.4 DATA STORAGE TO SCALE FACTOR RATIO.....	18
4.5 SCALE FACTOR TO MEMORY RATIO	18
CLAUSE 5: METRICS AND SCALE FACTORS.....	19
5.1 PERFORMANCE RUN METRIC	19
5.2 REPEATABILITY RUN METRIC	19
5.3 PRICE-PERFORMANCE METRIC	19
5.4 SCALE FACTOR.....	19
5.5 STREAM COUNT.....	19
5.6 ELAPSED RUN TIMES.....	20
5.7 ELAPSED TEST TIMES.....	20
AUDITORS' INFORMATION AND ATTESTATION LETTER.....	21
THIRD PARTY PRICE QUOTES.....	24
CLOUDERA	24
SUPPORTING FILE INDEX.....	25

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 v1.5.0.

The test was conducted at a Scale Factor of 3000 with 11 nodes (1x Aliyun ecs.g7.16xlarge, 10x Aliyun ecs.g7.16xlarge) running Horton Works HDP 3.0.1 on CentOS Linux release 7.7.1908.

Measured Configuration

Company Name	Cluster Node	Virtualization	Operating System
Alibaba Cloud	1x Aliyun ecs.g7.16xlarge 10x Aliyun ecs.g7.16xlarge	n/a	CentOS Linux release 7.7.1908

TPC Express Benchmark© Big Bench Metrics

Total System Cost	BBQpm@3000	Price/Performance	Availability Date
758,001 USD	2,187.42	346.53 USD	October 19, 2021

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 Alibaba Cloud

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.

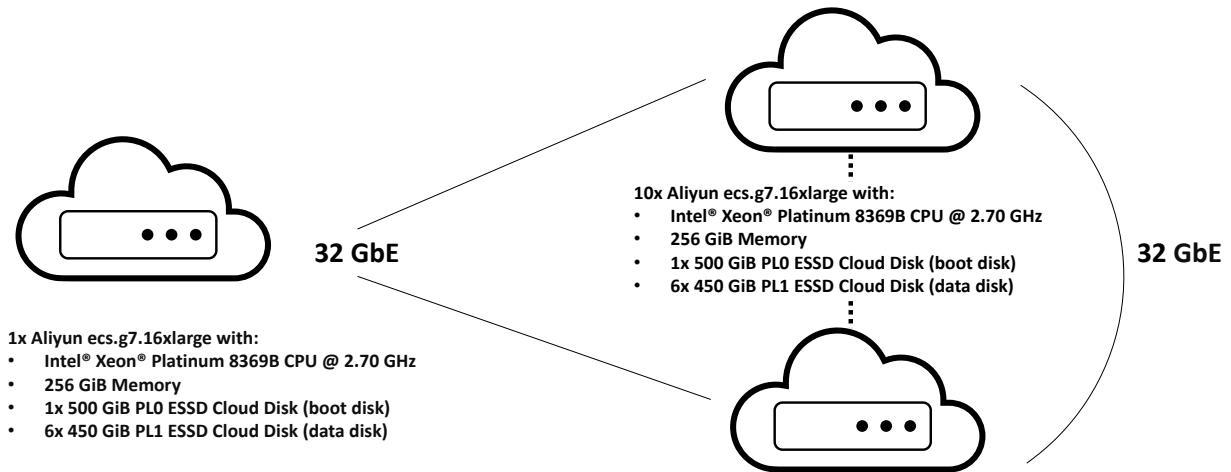
The Supporting Files Archive contains the parameters and options used to configure the components involved in this benchmark.

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:	11
Total Processors/Cores/Threads:	11/352/704
Total Memory:	2,816
Total Number of Storage Devices:	77
Total Storage Capacity:	37,796

Network: Cloud Infrastructure

	1x Aliyun ecs.g7.16xlarge (Admin):	10x Aliyun ecs.g7.16xlarge (Worker):
Processors/Cores/Threads:	1/32/64	1/32/64
Processor Model:	1x Intel® Xeon® Platinum 8369B CPU @ 2.70GHz	1x Intel® Xeon® Platinum 8369B CPU @ 2.70GHz
Memory:	256 GiB	256 GiB
Storage Devices:	1x 500 GiB PL0 ESSD Cloud Disk 6x 450 GiB PL1 ESSD Cloud Disk	1x 500 GiB PL0 ESSD Cloud Disk 6x 450 GiB PL1 ESSD Cloud Disk
Network Controller:	Bandwidth: 32 Gbps, Packet forwarding rate 12.000.000	Bandwidth: 32 Gbps, Packet forwarding rate 12.000.000

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
Worker	DataNode NodeManager HDFS Client Hive Client Spark2 Client Yarn Client Mapreduce2 Client	10	Y	hdp-tpcworker0[01-10]	Intel® Xeon® Platinum 8369B CPU @ 2.70GHz 256GB Memory 6 * 450 GiB PL1 ESSD Cloud Disk (data disk) 1 * 500 GiB PL0 ESSD Cloud Disk (boot disk) Bandwidth: 32 Gbps, Packet forwarding rate 12.000.000 OS: CentOS Linux release 7.7.1908	OS: 1 * 500GiB PL0 ESSD Cloud Disk Data Drive: 6 * 450 GiB PL1 ESSD Cloud Disk
Admin	DataNode NodeManager Hive Metastore HiverServer2 NameNode ResoureManager Spark2History Server HDFS NFSGateway HDFS Client Hive Client Spark2 Client Yarn Client Mapreduce2 Client	1	Y	tpc-testhdp001	Intel® Xeon® Platinum 8369B CPU @ 2.70GHz 256GB Memory 6 * 450 GB PL1 ESSD Cloud Disk (data disk) 1 * 500 GB PL0 ESSD Cloud Disk (boot disk) Bandwidth: 32 Gbps, Packet forwarding rate 12.000.000 OS: CentOS Linux release 7.7.1908	OS: 1 * 500GiB PL0 ESSD Cloud Disk Data Drive: 6 * 450 GiB PL1 ESSD Cloud Disk

2.2 Distributed File System Implementation

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

Horton Works HDP 3.0.1 (fully HDFS compatible at the API level).

2.3 Engine Implementation

The Engine implementation and corresponding version must be disclosed.

Component	Version
Hive	3.1.0
HDFS	3.1.1
YARN	3.1.1
Spark	2.3.1
MapReduce2	3.1.1

2.4 Frameworks

Frameworks and Engine used in the benchmark should be disclosed.

Framework	Version
HDP	3.0.1
Hive	3.1.0
HDFS	3.1.1
YARN	3.1.1
Spark	2.3.1
MapReduce2	3.1.1

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

v1.5.0

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
v1.5.0
*****
INFO: T_LOAD = 756.399
INFO: T_LD = 0.1 * T_LOAD: 75.6399
INFO: T_PT = 2767.4761023159
INFO: T_T_PUT = 16407.81
INFO: T_TT = 2050.97625
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@3000 = 2196.83525587891
```

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

```
*****
TPCx-BB
Result
v1.5.0
*****
INFO: T_LOAD = 747.26
INFO: T_LD = 0.1 * T_LOAD: 74.726
INFO: T_PT = 2786.02808204629
INFO: T_T_PUT = 16456.098
INFO: T_TT = 2057.01225
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@3000 = 2187.42634689913
```

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	Stream 5	Stream 6	Stream 7	Stream 8
Structured	1	46.784	109.982	179.339	273.148	141.700	493.531	143.246	160.731	107.673
	6	94.108	256.258	307.192	313.446	198.482	412.603	261.885	453.923	606.676
	7	42.403	123.806	523.443	225.288	138.554	388.719	65.140	208.737	97.386
	9	41.629	232.854	100.153	141.082	97.181	40.659	114.084	100.203	66.540
	11	40.330	109.164	195.101	92.017	76.457	140.246	110.509	85.412	392.509
	13	57.792	185.309	117.598	201.040	116.304	755.705	251.458	146.269	221.252
	14	41.180	592.309	90.537	70.491	154.441	185.041	206.504	90.085	92.390
	15	39.213	55.722	124.044	132.030	266.517	38.640	147.084	60.288	185.504
	16	80.376	490.268	308.284	140.869	126.271	566.074	570.052	400.192	271.108
	17	43.295	186.498	118.115	104.919	95.207	211.409	124.324	237.523	99.837
	20	92.309	206.057	214.724	726.533	287.855	352.470	686.797	664.998	276.265
	21	75.186	154.613	229.011	597.665	535.900	80.546	172.374	705.182	154.562
	22	44.491	198.878	427.268	258.439	197.511	222.817	310.557	87.220	88.999
	23	51.715	371.443	185.370	94.816	725.546	228.321	75.768	98.839	454.288
	24	46.021	80.971	93.234	366.408	121.396	174.060	296.596	162.697	657.715
25	102.807	508.115	657.148	560.551	285.226	338.648	1,009.242	262.183	397.294	
26	90.814	404.463	737.518	206.461	220.263	300.427	489.940	292.577	214.903	
29	67.185	174.088	352.557	194.930	253.365	338.231	293.206	178.320	136.871	
Semi-structured	2	623.032	2,530.230	2,516.353	2,221.795	1,947.221	2,772.813	1,979.037	1,344.737	2,221.617
	3	266.706	1,123.346	1,217.529	837.066	1,279.450	1,256.777	1,386.612	797.501	1,071.555
	4	433.373	1,840.533	1,614.320	1,297.511	1,734.217	1,192.984	1,098.966	1,669.149	1,454.983
	5	122.245	338.352	593.447	400.960	398.229	632.346	597.981	558.340	426.115
	8	123.501	308.227	325.208	1,022.637	309.885	424.769	811.868	386.733	320.414
	12	59.798	303.065	106.931	527.228	146.039	114.833	134.511	255.401	165.229
	30	489.803	1,773.063	1,490.544	1,362.757	1,915.996	1,890.749	738.047	2,143.574	2,121.244
Unstructured	10	105.274	442.454	605.620	371.854	523.500	257.924	510.469	512.063	423.795
	18	318.192	1,378.464	1,571.031	1,076.755	1,763.462	1,128.895	1,418.413	1,574.809	1,689.211
	19	454.923	1,400.203	908.583	1,379.898	1,397.904	968.242	810.552	1,483.908	1,511.662
	27	52.244	207.169	143.269	96.120	201.953	93.521	170.282	161.197	114.066
28	91.319	234.394	271.414	344.583	640.332	454.087	819.165	223.886	225.366	

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/spark_sql/conf/engineSettings.conf
- bigBench-configs/spark_sql/queries/q02/engineLocalSettings.sql
- bigBench-configs/spark_sql/queries/q03/engineLocalSettings.sql
- bigBench-configs/spark_sql/queries/q04/engineLocalSettings.sql
- bigBench-configs/spark_sql/queries/q05/engineLocalSettings.sql
- bigBench-configs/spark_sql/queries/q10/engineLocalSettings.sql
- bigBench-configs/spark_sql/queries/q16/engineLocalSettings.sql
- bigBench-configs/spark_sql/queries/q18/engineLocalSettings.sql
- bigBench-configs/spark_sql/queries/q19/engineLocalSettings.sql
- bigBench-configs/spark_sql/queries/q21/engineLocalSettings.sql
- bigBench-configs/spark_sql/queries/q22/engineLocalSettings.sql
- bigBench-configs/spark_sql/queries/q23/engineLocalSettings.sql
- bigBench-configs/spark_sql/queries/q29/engineLocalSettings.sql
- bigBench-configs/spark_sql/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 from every role in the server.

All envinfo.log files are included 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)
11	1	537	5,906
11	6	483	31,890

Total Storage (GB) 37,796

Scale Factor 3000

Data Storage Ratio 12.60

4.5 Scale Factor to Memory Ratio

The Scale Factor to memory ratio must be disclosed.

Nodes	Memory (GiB)	Total (GiB)
11	256	2,816

Scale Factor 3000

Total Memory (GiB) 2,816

SF / Memory Ratio 1.07

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@3000 2,187.42

5.2 Repeatability Run Metric

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

Repeatability Run
BBQpm@3000 2,196.83

5.3 Price-Performance Metric

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

Price / Performance
\$BBQpm@3000 346.53

5.4 Scale Factor

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

Scale Factor
3000

5.5 Stream Count

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

Streams
8

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 05:56:15.953	21,375.953
Run 2	00 05:57:21.433	21,441.433

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	747.260	756.399
Power Test	4,238.073	4,211.742
Throughput Test	16,456.098	16,407.810

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.

Li Tengfei
Building 9, Block 4, Wangjing East Park
Chaoyang District, Beijing, 100102
China

October 18, 2021

I verified the TPC Express Benchmark™ BB v1.5.0 performance of the following configuration:

Platform: Alibaba Cloud ApsaraCompute Big Data Accelerator
(w/ 1x Admin Node, 10x Compute Nodes)
Operating System: CentOS Linux release 7.7.1908
Framework: Horton Works HDP 3.0.1

The results were:

Performance Metric **2,187.42 BBQpm@3000**
Run Elapsed Time 00 05:57:21.433 (21,441.443 Seconds)

Cluster **1x Admin Node, 10x Compute Nodes**

CPU	1x Intel® Xeon® Platinum 8369B (2.70 GHz, 32-core, 48 MB L3)		
Memory	256GiB		
Storage	Qty	Size	Type
	1	500 GiB	PL0 ESSD Cloud Disk
	6	450 GiB	PL1 ESSD Cloud Disk

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 v1.5.0
- 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 (3000GB) was confirmed to be valid for publication
- 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:

None.

Respectfully Yours,

A handwritten signature in cursive script that reads "Doug Johnson". The signature is written in black ink and has a long, sweeping horizontal line extending to the right.

Doug Johnson, TPC Auditor

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

Third Party Price Quotes

Cloudera

Overview CDP Public Cloud CDP Private Cloud CDP support Additional offerings [Contact sales](#)

DATA SCIENCE WORKBENCH Delivers fast, easy, and secure self-service data science for the enterprise. \$5,000/user Annual subscription Learn more >	HDP ENTERPRISE PLUS Securely store, process, and analyze all your structured and unstructured data at rest. \$10,000/node + variable ¹ Annual subscription Learn more >	ENTERPRISE DATA HUB Transform complex data into clear and actionable insights with an integrated suite of analytic engines. \$10,000/node + variable ¹ Annual subscription Learn more >
GPU ACCELERATION Leverage GPUs to accelerate Cloudera Data Engineering as well as Machine Learning Spark workloads. \$7,500/CGU Annual subscription Learn more >		

¹Variable compute price: \$75 per CCU over 16 cores, 128 GB RAM node cap; variable storage price: \$25 per TB over 48 TB node cap.

Variable and CCU Description

¹ Variable compute price: \$75 per CCU over 16 cores / 128 GB RAM node cap; variable storage price: \$25 per TB over 48 TB node cap.

Definitions

- Cloudera Compute Unit (CCU): 1 physical core and 8GB RAM

CDP Private Cloud pricing reflects Business-level Support.

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	Supporting-Files-3TB-Alibaba-ApsaraCompute-Big-Data-Accelerator-10-2021\
Validation Run Files	Supporting-Files-3TB-Alibaba-ApsaraCompute-Big-Data-Accelerator-10-2021\Validation-Run logs
Performance Run Files	Supporting-Files-3TB-Alibaba-ApsaraCompute-Big-Data-Accelerator-10-2021\Performance-run logs
Repeatability Run Files	Supporting-Files-3TB-Alibaba-ApsaraCompute-Big-Data-Accelerator-10-2021\Repeatability-run logs
Clause 3 - Workload Related Items	
Benchmark Generic Parameters	Supporting-Files-3TB-Alibaba-ApsaraCompute-Big-Data-Accelerator-10-2021\Performance-run logs\bigBench-configs\conf\userSettings.conf
Query Parameters used in the benchmark execution Settings	Supporting-Files-3TB-Alibaba-ApsaraCompute-Big-Data-Accelerator-10-2021\Performance-run logs\bigBench-configs\spark_sql\conf\queryParameters.sql
Benchmark Global Framework Parameters Settings	Supporting-Files-3TB-Alibaba-ApsaraCompute-Big-Data-Accelerator-10-2021\Performance-run logs\bigBench-configs\spark_sql\conf\engineSettings.sql
Benchmark Global Framework Parameters Settings	Supporting-Files-3TB-Alibaba-ApsaraCompute-Big-Data-Accelerator-10-2021\Performance-run logs\bigBench-configs\spark_sql\conf\engineSettings.conf
Load Test script	Supporting-Files-3TB-Alibaba-ApsaraCompute-Big-Data-Accelerator-10-2021\Performance-run logs\bigBench-configs\spark_sql\population\sparkSqlCreateLoad.sql
Queries specific optimization parameters settings	Supporting-Files-3TB-Alibaba-ApsaraCompute-Big-Data-Accelerator-10-2021\Performance-run logs\bigBench-configs\spark_sql\queries\q[01-30]\engineLocalSettings.conf
Queries specific optimization parameters settings	Supporting-Files-3TB-Alibaba-ApsaraCompute-Big-Data-Accelerator-10-2021\Performance-run logs\bigBench-configs\spark_sql\queries\q[01-30]\engineLocalSettings.sql
Clause 4 - SUT Related Items	
Data Redundancy report	Supporting-Files-3TB-Alibaba-ApsaraCompute-Big-Data-Accelerator-10-2021\Performance-run logs\run-logs\data_redundancy_report.log
Benchmark execution script	Supporting-Files-3TB-Alibaba-ApsaraCompute-Big-Data-Accelerator-10-2021\TPCxBB_Benchmarkrun.sh
Hardware and Software Report from a representative node	Supporting-Files-3TB-Alibaba-ApsaraCompute-Big-Data-Accelerator-10-2021\Performance-run logs\envInfo-hdp-tpcworker001\envInfo.log
All Framework configuration files are included in the Supporting Files Archive	Supporting-Files-3TB-Alibaba-ApsaraCompute-Big-Data-Accelerator-10-2021\Performance-run logs\envInfo-hdp-tpcworker001\hadoop
	Supporting-Files-3TB-Alibaba-ApsaraCompute-Big-Data-Accelerator-10-2021\Performance-run logs\envInfo-hdp-tpcworker001\hive
Clause 5 - Metric and Scale Factor Related Items	
Benchmark Performance Report	Supporting-Files-3TB-Alibaba-ApsaraCompute-Big-Data-Accelerator-10-2021\Performance-run logs\run-logs\BigBenchResult.log
Validation Test Report	Supporting-Files-3TB-Alibaba-ApsaraCompute-Big-Data-Accelerator-10-2021\Validation-Run logs\run-logs\BigBenchResult.log