Hewlett Packard Enterprise

Hewlett Packard Enterprise Company

TPC Express Benchmark™ Big Bench (TPCx-BB)

Full Disclosure Report

for

Hewlett Packard Enterprise ProLiant DL for Big Data

(w/ 18x HPE ProLiant DL380 Gen9, 3x HPE ProLiant DL360 Gen9)

using

Cloudera for Apache Hadoop (CDH) 5.8.0

and

Red Hat Enterprise Linux Server 7.2

First Edition

July 9, 2017

***Hewlett Packard Enterprise Company (HPE)**, 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.

HPE and the HPE Logo are trademarks of Hewlett Packard Enterprise Company 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 HPE and any other company.

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

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

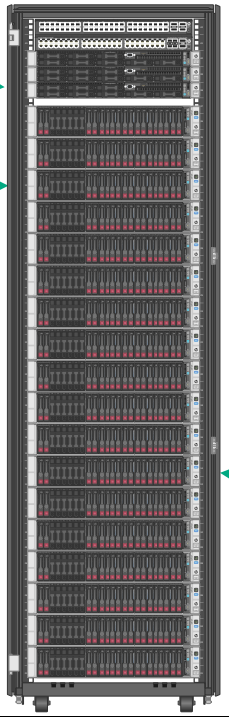
Hewlett Packard Enterprise	**Hewlett Packard Enterprise ProLiant DL for Big Data**	TPCx-BB Rev. v1.2.0 TPC-Pricing Rev. v2.1.1 Report Date: July 9, 2017
Total System Cost	TPCx-BB Performance Metric	Price/Performance
719,290 USD	**830.67** BBQpm@10000	**865.92 USD** $/BBQpm@10000

Framework	Operating System	Other Software	Availability Date	Scale Factor	Streams
Cloudera for Apache Hadoop (CDH) 5.8.0	Red Hat Enterprise Linux Server 7.2	OpenJDK 1.8.0_65	July 9, 2017	10000	2

System Configuration

Ethernet Switch:
HPE 1620-24G Switch (ILO connection)
HP5900AF-48XGT-4QSFP+ (main connection)

Software:
Red Hat Enterprise Linux 7.2
Cloudera Enterprise 5.8



3 Management Node:
Each:
HPE ProLiant DL360 Gen9
2x Intel E5-2640 v4 2.40GHz
1x HPE 800GB 6G SATA SSD (2 nodes)
1x HPE 480GB 6G SATA SSD (1 node)
256GB Memory

18 Worker Nodes:
Each:
HPE ProLiant DL380 Gen9
2x Intel E5-2697A v4 2.60GHz
1x HPE 480GB 6G SATA SSD
1x HPE 1.2TB 6G SATA SSD
16x HPE 1TB 6G SATA 7.2K HDD
256GB Memory

Physical Storage/Scale Factor: 32.03	Scale Factor/Physical Memory: 1.86

Servers:	18x HPE ProLiant DL380 Gen9, 3x HPE ProLiant DL360 Gen9	
Total Processors/Cores/Threads	42/636/1,272	
Server Configuration:	**Per HPE ProLiant DL380 Gen9:**	**Per HPE ProLiant DL360 Gen9:**
Processors	2x Intel Xeon E5-2697A v4 @ 2.60GHz	2x Intel Xeon E5-2640 @ 2.40GHz
Memory	256GB	256GB
Storage Controller	HPE Smart Array P840/4G	HPE Smart Array P440ar Controller
Storage Device	1x HPE 480GB 6G SATA SSD 1x HPE 1.2TB 6G SATA SSD 16x HPE 1TB 6G SATA 7.2K HDD	1x HPE 480GB 6G SATA SSD (1 node) 1x HPE 800GB 6G SATA SSD (2 nodes)
Network	HPE Ethernet 10G 2P Adapter	HPE Ethernet 10G 2P Adapter
Connectivity:	HPE 1620-24G Switch (ILO connection) HPE 5900AF-48XG-4QSFP+ Switch (main connection)	

Hewlett Packard Enterprise

Hewlett Packard Enterprise ProLiant DL for Big Data

TPCx-BB Rev. v1.2.0
TPC-Pricing Rev. v2.1.1

Report Date:
July 9, 2017

Description	Price Key	Part Number	Unit Price	Qty	Extended Price	3 Yr Maint Price
Server Hardware						
HPE DL360 Gen9 8SFF CTO Server	1	755258-B21	$1,797	3	$5,391	
HPE DL360 Gen9 E5-2640v4 Kit	1	842978-B21	$1,359	3	$4,077	
HPE DL360 Gen9 E5-2640v4 FIO Kit	1	842978-L21	$1,359	3	$4,077	
HPE 32GB 2Rx4 PC4-2400T-R Kit	1	805351-B21	$1,059	24	$25,416	
HPE H240ar 12Gb 2-ports Int FIO Smart Host Bus Adapter	1	749976-B21	$249	3	$747	
HPE 500W FS Plat Ht Plg Pwr Supply Kit	1	720478-B21	$309	6	$1,854	
HPE 480GB 6G SATA RI-2 SFF SC SSD	1	804593-B21	$609	1	$609	
HPE 800GB 6G SATA MU-2 SFF SC SSD	1	804671-B21	$1,389	2	$2,778	
HPE Ethernet 10Gb 2P 560FLR-SFP+ Adptr	1	665243-B21	$679	3	$2,037	
HPE 3Y FC 24x7 DL360 Gen9 SVC	1	U7AL9E	$1,404	3		$4,212
HPE iLO Adv incl 3yr TS U E-LTU	1	E6U64ABE	$469	3		$1,407
HP W1972a 18.5-In LED Monitor (1 + 2 spare)	1	B7M13A8#ABA	$80	3	$240	
HP PS/2 Keyboard And Mouse Bundle (1 + 2 spare)	1	B1T13AA#ABA	$28	3	$84	
			Subtotal		**$47,310**	**$5,619**
HPE DL380 Gen9 24SFF CTO Server	1	767032-B21	$2,107	18	$37,926	
HPE DL380 Gen9 High Perf Fan Kit	1	719079-B21	$239	18	$4,302	
HPE DL380 Gen9 2SFF Bay Kit	1	724864-B21	$179	18	$3,222	
HPE DL380 Gen9 Secondary Riser	1	719073-B21	$99	18	$1,782	
HPE XL2x0 Gen9 E5-2697Av4 Kit	1	841179-B21	$4,359	18	$78,462	
HPE XL2x0 Gen9 E5-2697Av4 FIO Kit	1	841179-L21	$4,359	18	$78,462	
HPE 32GB 2Rx4 PC4-2400T-R Kit	1	805351-B21	$1,059	144	$152,496	
HPE 1TB 6G SATA 7.2k 2.5in SC MDL HDD	1	655710-B21	$499	288	$143,712	
HPE 480GB 6G SATA RI-2 SFF SC SSD	1	804593-B21	$609	18	$10,962	
HPE 1.2TB 6G SATA WI-2 SFF SC SSD	1	804677-B21	$2,365	18	$42,570	
HPE Smart Array P840/4G Controller	1	726897-B21	$1,249	18	$22,482	
HPE 12Gb DL380 Gen9 SAS Expander Card	1	727250-B21	$699	18	$12,582	
HPE 800W FS Ti Ht Plg Pwr Supply Kit	1	720482-B21	$409	36	$14,724	
HPE Ethernet 10Gb 2P 560FLR-SFP+ Adptr	1	665243-B21	$679	18	$12,222	
HPE 3Y FC 24x7 DL380 Gen9 SVC	1	U7AE5E	$1,872	18		$33,696
HPE iLO Adv incl 3yr TS U E-LTU	1	E6U64ABE	$469	18		$8,442
			Subtotal		**$615,906**	**$42,138**

Hewlett Packard Enterprise

Hewlett Packard Enterprise ProLiant DL for Big Data

TPCx-BB Rev. v1.2.0
TPC-Pricing Rev. v2.1.1

Report Date:
July 9, 2017

Description	Price Key	Part Number	Unit Price	Qty	Extended Price	3 Yr Maint Price
Network						
HPE 1620-24G Switch	1	JG913A	$299	1	$299	
HPE 5900AF-48XG-4QSFP+ Switch	1	JC772A	$20,990	1	$20,990	
HPE 1m Multi-mode OM3 LC/LC FC Cable	1	AJ834A	$70	21	$1,470	
CAT6 UTP 1G Ethernet Network Cable 7ft (42 cables)	2	C6-UTPSMPVCYL-2M	$2	42	$84	
HPE A58x0AF 650W AC Power Supply	1	JC680A	$749	4	$2,996	
HPE 58x0AF Bck(pwr)-Frt(ports) Fan Tray	1	JC682A	$179	4	$716	
			Subtotal		**$26,555**	**$0**
Rack						
HPE 42U 600x1075mm Adv G2 Kit Pllt Rack	1	P9K07A	$1,179	1	$1,179	
HPE 24A High Voltage Core Only Corded PDU	1	252663-D74	$259	4	$1,036	
			Subtotal		**$2,215**	**$0**
Server Software						
Cloudera Ent Basic Ed 1yr 24x7	1	G7M27A	$2,000	72	$144,000	
RHEL Svr 2 Sckt/2 Gst 3yr 24x7 E-LTU	1	G3J30AAE	$3,702	21	$77,742	
			Subtotal		**$221,742**	**$0**
		Total Extended Price			**$913,728**	**$47,757**
		Total Discounts (35%)			**$242,195**	**$0**
Sales contact: HPE WW Headquarters, 3000 Hanover St., Palo Alto, CA 94304-1185 (650) 857-1501 or HPE: 855-472-5233		**Grand Total**			**$671,533**	**$47,757**

Pricing:1 = {Source 1}; 2 = {Source 2}

(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 of InfoSizing

Three-Year Cost of Ownership	**$719,290**
BBQpm@10000	**830.67**
$/BBQpm@10000	**$ 865.92**

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	10000
Streams	2
SUT Validation Test	PASS

Performance Run (Run 2)

Overall Run Start Time	2017-06-27 00:05:50.957
Overall Run End Time	2017-06-27 21:38:50.865
Overall Run Elapsed Time	77,579.908
Load Test Start Time	2017-06-27 00:05:50.958
Load Test End Time	2017-06-27 00:55:30.300
Load Test Elapsed Time	2,979.342
Power Test Start Time	2017-06-27 00:55:30.302
Power Test End Time	2017-06-27 09:11:59.406
Power Test Elapsed Time	29,789.104
Throughput Test Start Time	2017-06-27 09:11:59.408
Throughput Test End Time	2017-06-27 21:38:50.864
Throughput Test Elapsed Time	44,811.456
Performance Metric (BBQpm@ 10000)	830.67

Repeatability Run (Run 1)

Overall Run Start Time	2017-06-26 01:51:21.978
Overall Run End Time	2017-06-26 22:58:06.590
Overall Run Elapsed Time	76,004.612
Load Test Start Time	2017-06-26 01:51:21.979
Load Test End Time	2017-06-26 02:40:31.683
Load Test Elapsed Time	2,949.704
Power Test Start Time	2017-06-26 02:40:31.685
Power Test End Time	2017-06-26 10:55:37.091
Power Test Elapsed Time	29,705.406
Throughput Test Start Time	2017-06-26 10:55:37.093
Throughput Test End Time	2017-06-26 22:58:06.589
Throughput Test Elapsed Time	43,349.496
Performance Metric (BBQpm@ 10000)	844.88

Performance Run Report (Run 2)

```
************
TPCx-BB
Result
v1.2
************
INFO: T_LOAD = 2979.342
INFO: T_LD = 0.1 * T_LOAD: 297.9342
INFO: T_PT = 20384.2122165116
INFO: T_T_PUT = 44811.456
INFO: T_TT = 22405.728
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@10000 = 830.679180114096
```

Repeatability Run Report (Run 1)

```
************
TPCx-BB
Result
v1.2
************
INFO: T_LOAD = 2949.704
INFO: T_LD = 0.1 * T_LOAD: 294.9704
INFO: T_PT = 20365.0080854201
INFO: T_T_PUT = 43349.496
INFO: T_TT = 21674.748
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@10000 = 844.886181932719
```

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 10000 with 21 nodes (18x HPE ProLiant DL380 Gen9, 3x HPE ProLiant DL360 Gen9) running Cloudera for Apache Hadoop (CDH) 5.8.0 on Red Hat Enterprise Linux Server 7.2.

Measured Configuration

Company Name	Cluster Node	Virtualization	Operating System
Hewlett Packard Enterprise Company	18x HPE ProLiant DL380 Gen9 3x HPE ProLiant DL360 Gen9	n/a	Red Hat Enterprise Linux Server 7.2

TPC Express Benchmark© Big Bench Metrics

Total System Cost	BBQpm@10000	Price/Performance	Availability Date
719,290 USD	830.67	865.92 USD	July 9, 2017

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 Hewlett Packard Enterprise Company

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.

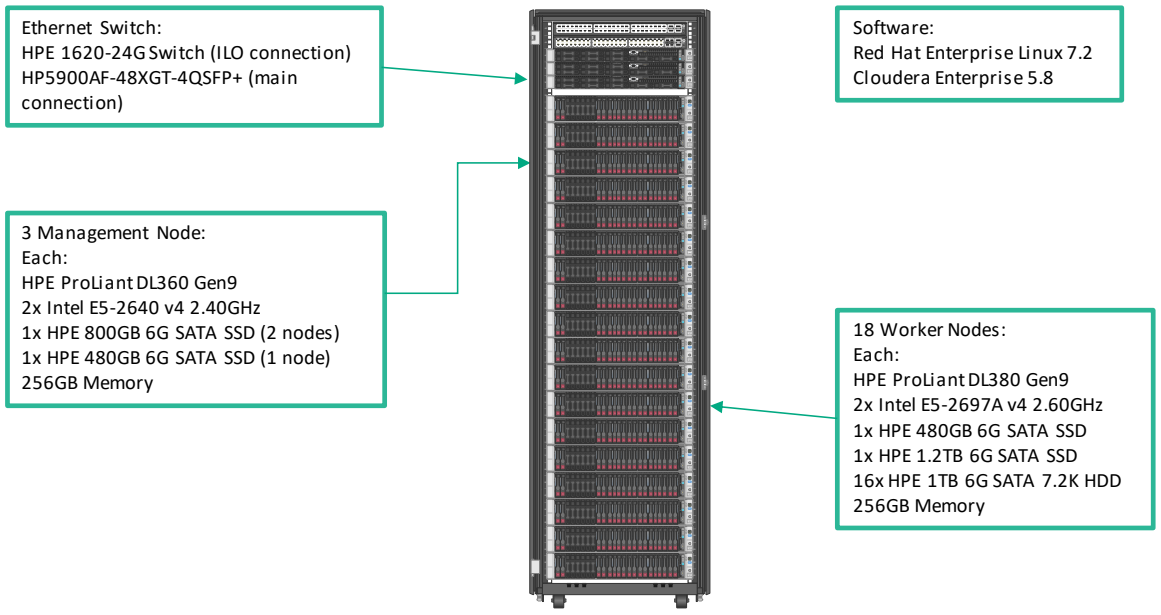
The Supporting Files Archive contains the parameters and options used to configure the components involved in this benchmark.

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: 21
- Total Processors/Cores/Threads: 42/636/1,272
- Total Memory: 5,376
- Total Number of Storage Drives/Devices: 327
- Total Storage Capacity: 320,320

Network connectivity detail:

- HPE 1620-24G Switch (ILO connection), HPE 5900AF-48XG-4QSFP+ Switch (main connection)

Server nodes details:

18x HPE ProLiant DL380 Gen9, each with:

- Processors/Cores/Threads: 2/32/64
- Processor Model: 2x Intel Xeon E5-2697A v4 @ 2.60GHz
- Memory: 256GB
- Controller: 1x HPE Smart Array P840/4G
- Drives:
  - 1x HPE 480GB SSD
  - 1x HPE 1.2TB SSD
  - 16x HPE 1TB 7.2K HDD
- Network: HPE Ethernet 10G 2P Adapter

3x HPE ProLiant DL360 Gen9, each with:

- Processors/Cores/Threads: 2/20/40
- Processor Model: 2x Intel Xeon E5-2640 2.40GHz
- Memory: 256GB
- Controller: HPE Smart Array P440ar Controller
- Drives:
  - 1x HPE 480GB SSD (1 node)
  - 1x HPE 800GB SSD (2 nodes)
- Network: HPE Ethernet 10G 2P Adapter

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 Name(s)	HW/SW Configuration	Storage Setup
Worker	HDFS DataNode/Hive Gateway/YARN Node Manager/Spark Gateway	18	N	bdw21-[04-21]	• HPE DL380 Gen9 • HW/SW Config (Intel Xeon E5-2697A v4, 2, 2.GHz, 64) • Memory: 256GB • Storage: 16 x 1TB SATA HDD, 1 x 480GB, 1x1.2TB SSD • Network: HPE Ethernet 25G NIC • OS: RHEL 7.2 • Cloudera CDH 5.8	OS: HPE 480GB SSD, Intermediate/Shuffle/Temp Data/ Distributed FS: 1 x 1.2TB SSD, 16 x HPE 1TB SATA 7.2K HDD
Cloudera Manager Node #1	HDFS Balancer/HDFS Namenode/Hive Gateway/CMS Alert Publisher/CMS Event Server/CMS Host Monitor/CMS Report Manager/CMS Service Monitor/ZooKeeper Server/Spark Gateway	1	N	bdw21-01	• HPE DL360 Gen10 Server • HW/SW Config (Intel Xeon E5-2640@2.10GHz, 2.4GHz, 40) • Memory: 256GB • Storage: 1 x 800GB SSD • Network: HPE Ethernet 10G NIC • OS: RHEL 7.2 • Cloudera CDH 5.8	OS: HPE 800GB SATA SSD
Cloudera Manager Node #2	Hive Gateway/Hive Metastore Server/HiveServer2/Spark Gateway/Spark History Server/YARN Job History Server/YARN ResourceManager/ZooKeeper Server	1	N	bdw21-02	• HPE DL360 Gen10 Server • HW/SW Config (Intel Xeon E5-2640@2.10GHz, 2.4GHz, 40) • Memory: 256GB • Storage: 1 x 800GB SSD • Network: HPE Ethernet 10G NIC • OS: RHEL 7.2 • Cloudera CDH 5.8	OS: HPE 800GB SATA SSD
Cloudera Manager Node #3	HDFS DataNode/HDFS SecondaryNamenode/Hive Gateway/Spark Gateway/YARN Gateway/Zookeeper Server	1	N	bdw21-03	• HPE DL360 Gen10 Server • HW/SW Config (Intel Xeon E5-2640@2.10GHz, 2.4GHz, 40) • Memory: 256GB • Storage: 1 x 800GB SSD • Network: HPE Ethernet 10G NIC • OS: RHEL 7.2 • Cloudera CDH 5.8	OS: HPE 800GB SATA SSD

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) 5.8.0 (fully HDFS compatible at the API level).

2.3 Engine Implementation

The Engine implementation and corresponding version must be disclosed.

Component	Version
Hive	1.1.0
HDFS	2.6.0
YARN	2.6.0
Spark	1.6.0
MapReduce	2.6.0
Zookeeper	3.4.5

2.4 Frameworks

Frameworks and Engine used in the benchmark should be disclosed.

Framework	Version
CDH	5.8.0
Hive	1.1.0
HDFS	2.6.0
YARN	2.6.0
Spark	1.6.0
MapReduce	2.6.0
Zookeeper	3.4.5

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 = 2949.704
INFO: T_LD = 0.1 * T_LOAD: 294.9704
INFO: T_PT = 20365.0080854201
INFO: T_T_PUT = 43349.496
INFO: T_TT = 21674.748
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@10000 = 844.886181932719
```

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

```
**********
TPCx-BB
Result
v1.2
**********
INFO: T_LOAD = 2979.342
INFO: T_LD = 0.1 * T_LOAD: 297.9342
INFO: T_PT = 20384.2122165116
INFO: T_T_PUT = 44811.456
INFO: T_TT = 22405.728
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@10000 = 830.679180114096
```

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.

Query Type	Query Number	Power	Throughput	
		Stream 1	Stream 1	Stream 2
Structured	1	272.027	296.950	272.841
	6	835.341	853.201	871.805
	7	593.981	1,157.091	1,501.498
	9	490.715	504.606	851.316
	11	238.165	322.754	250.499
	13	461.223	1,332.544	1,117.927
	14	145.174	144.479	145.284
	15	158.241	340.889	464.294
	16	1,961.833	2,073.692	2,496.399
	17	374.422	591.890	415.456
	20	438.816	811.119	447.100
	21	1,227.419	1,417.752	2,119.369
	22	481.151	741.938	979.388
	23	433.095	460.293	623.874
	24	367.860	395.748	376.594
	25	726.891	746.305	722.742
	26	729.440	1,548.866	1,013.194
	29	1,128.672	1,305.199	1,759.359
Semi-structured	2	2,457.710	4,342.623	3,432.447
	3	1,367.051	2,222.187	1,448.947
	4	2,161.689	3,911.804	3,886.226
	5	3,423.412	4,809.556	5,303.882
	8	814.021	1,301.727	1,611.961
	12	1,172.399	3,612.252	1,559.745
	30	2,563.193	2,830.496	3,732.554
Unstructured	10	769.596	1,172.483	791.446
	18	2,717.358	3,365.385	3,840.798
	19	835.631	1,687.838	1,190.585
	27	78.676	91.230	79.566
	28	363.837	418.509	376.239

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/userSettings.conf
- bigBench-configs/hive/conf/engineSettings.conf
- bigBench-configs/hive/conf/engineSettings.sql
- bigBench-configs/hive/population/hiveCreateLoad.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 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)
1	1	480	480
2	1	800	1,600
18	1	480	8,640
18	1	1200	21,600
18	16	1000	288,000

Total Storage (GB)	320,320
Scale Factor	10000
Data Storage Ratio	32.03

4.5 Scale Factor to Memory Ratio

The Scale Factor to memory ratio must be disclosed.

Nodes	Memory (GB)	Total (GB)
21	256	5,376

Scale Factor	10000
Total Memory (GB)	5,376
SF / Memory Ratio	1.86

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@10000 830.67

5.2 Repeatability Run Metric

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

Repeatability Run
BBQpm@10000 844.88

5.3 Price-Performance Metric

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

Price / Performance
$BBQpm@10000 865.92

5.4 Scale Factor

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

Scale Factor
10000

5.5 Stream Count

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

Streams
2

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	21:06:44.612	76,004.612
Run 2	21:32:59.908	77,579.908

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	2,979.342	2,949.704
Power Test	29,789.104	29,705.406
Throughput Test	44,811.456	43,349.496

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

Mr. Paul Cao
Hewlett Packard Enterprise
11445 Compaq Center Dr West
Houston, TX 77070

July 7, 2017

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

Platform:	Hewlett Packard Enterprise ProLiant DL for Big Data (w/ 18x HPE ProLiant DL380 Gen 9, 3x HPE ProLiant DL360 Gen9)
Operating System:	Red Hat Enterprise Linux Server 7.2
Apache Hadoop Compatible Software:	Cloudera for Apache Hadoop (CDH) 5.8.0

The results were:

Performance Metric	**830.67 BBQpm@10000GB**
Run Elapsed Time	21:32:59.908 (77,579.908 Seconds)

Cluster	**18x HPE ProLiant DL380 Gen 9 (Data nodes), 3x HPE ProLiant DL360 Gen9 (Management nodes)**
CPUs	2 x Intel Xeon E5-2697A v4 (2.60 GHz, 16-core, 40 MB L3) (Data nodes) 2 x Intel Xeon E5-2640 (2.40 GHz, 10-core, 25 MB L3) (Mgmt. nodes)
Memory	256GB (all nodes)
Storage	

Qty	*Size*	*Type*
1	480GB	6G SATA SSD (OS, 1 Mgmt. node)
1	800GB	6G SATA SSD (OS, 2 Mgmt. nodes)
1	480GB	6G SATA SSD (OS, Data nodes)
1	1.2TB	6G SATA SSD (Temp, Data nodes)
16	1TB	6G SATA 7.2K HDD (Data, 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

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

- The tested Scale Factor (10000GB) 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:

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.

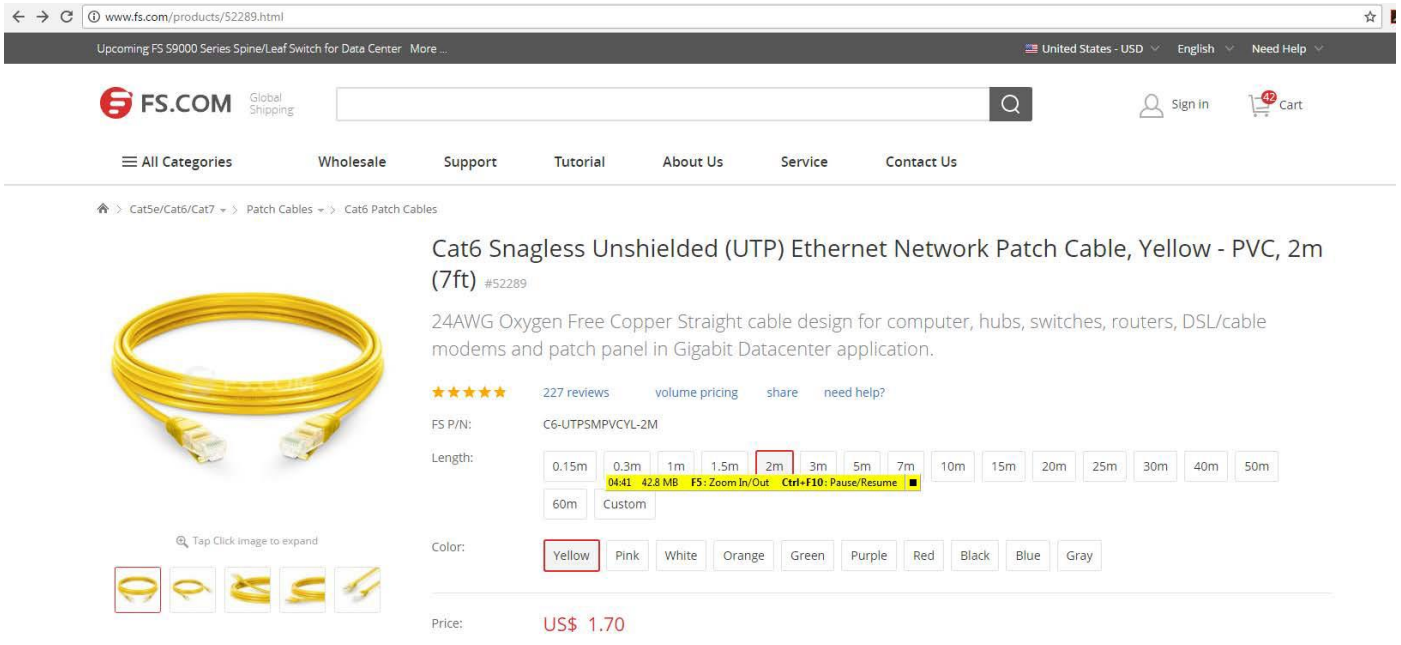
Query 28 failed automated Engine Validation. A manual validation was performed as part of this audit to confirm the only differences were due to white space.

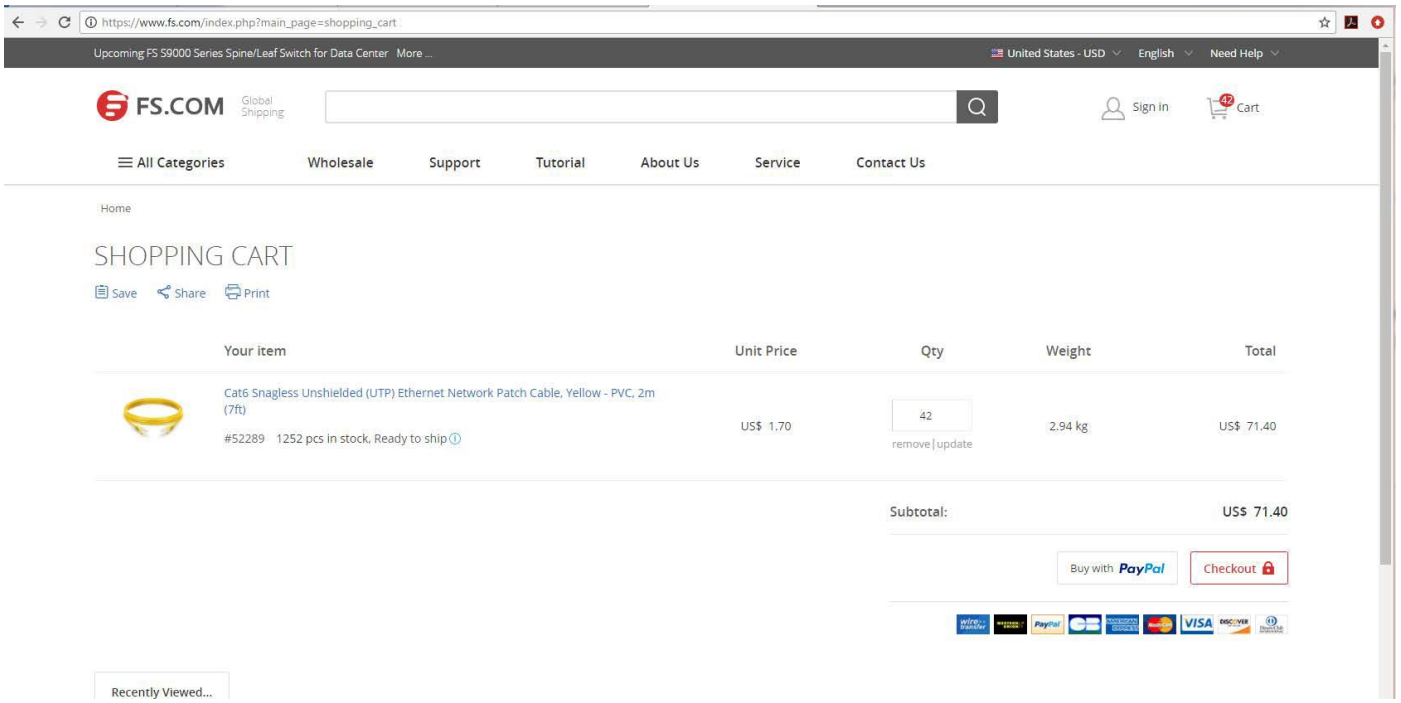
Respectfully Yours,

Doug Johnson, TPC Auditor

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

Third Party Price Quotes





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-10TB-Broadwell-7-2017\
Validation Run Files	Supporting-Files-10TB-Broadwell-7-2017\Validation-Run-Broadwell-logs-20170626-005911-hive-sf10000
Performance Run Files	Supporting-Files-10TB-Broadwell-7-2017\Performance-run-Broadwell-logs-20170627-214217-hive-sf10000
Repeatability Run Files	Supporting-Files-10TB-Broadwell-7-2017\Repeatability-run-Broadwell-logs-20170626-230136-hive-sf10000
Clause 3 - Workload Related Items	
Benchmark Generic Parameters	Supporting-Files-10TB-Broadwell-7-2017\Performance-run-Broadwell-logs-20170627-214217-hive-sf10000\bigBench-configs\conf\userSettings.conf
Query Parameters used in the benchmark execution Settings	Supporting-Files-10TB-Broadwell-7-2017\Performance-run-Broadwell-logs-20170627-214217-hive-sf10000\bigBench-configs\hive\conf\queryParameters.sql
Benchmark Global Framework Parameters Settings	Supporting-Files-10TB-Broadwell-7-2017\Performance-run-Broadwell-logs-20170627-214217-hive-sf10000\bigBench-configs\hive\conf\engineSettings.sql
Benchmark Global Framework Parameters Settings	Supporting-Files-10TB-Broadwell-7-2017\Performance-run-Broadwell-logs-20170627-214217-hive-sf10000\bigBench-configs\hive\conf\engineSettings.conf
Load Test script	Supporting-Files-10TB-Broadwell-7-2017\Performance-run-Broadwell-logs-20170627-214217-hive-sf10000\bigBench-configs\hive\population\hiveCreateLoad.sql
Queries specific optimization parameters settings	Supporting-Files-10TB-Broadwell-7-2017\Performance-run-Broadwell-logs-20170627-214217-hive-sf10000\bigBench-configs\hive\queries\q[01-30]\engineLocalSettings.conf
Queries specific optimization parameters settings	Supporting-Files-10TB-Broadwell-7-2017\Performance-run-Broadwell-logs-20170627-214217-hive-sf10000\bigBench-configs\hive\queries\q[01-30]\engineLocalSettings.sql
Clause 4 - SUT Related Items	
Data Redundancy report	Supporting-Files-10TB-Broadwell-7-2017\hdfs-data-redundancy-report.txt
Benchmark execution script	Supporting-Files-10TB-Broadwell-7-2017\publication-run.sh
Hardware and Software Report from a representative node	Supporting-Files-10TB-Broadwell-7-2017\Performance-run-Broadwell-logs-20170627-214217-hive-sf10000\run-logs\envInfo-skl21-04\envInfo.log
All Framework configuration files are included in the Supporting Files Archive	Supporting-Files-10TB-Broadwell-7-2017\Performance-run-Broadwell-logs-20170627-214217-hive-sf10000\bigBench-configs\conf
	Supporting-Files-10TB-Broadwell-7-2017\Performance-run-Broadwell-logs-20170627-214217-hive-sf10000\bigBench-configs\hive
Clause 5 - Metric and Scale Factor Related Items	
Benchmark Performance Report	Supporting-Files-10TB-Broadwell-7-2017\Performance-run-Broadwell-logs-20170627-214217-hive-sf10000\run-logs\BigBenchResult.log
Validation Test Report	Supporting-Files-10TB-Broadwell-7-2017\Validation-Run-Broadwell-logs-20170626-005911-hive-sf10000\run-logs\BigBenchResult.log