



# **Databricks, Inc.**

---

TPC Benchmark™ DS

Full Disclosure Report

for

Databricks SQL 8.3

(4X-Large with 256 nodes)

On

Amazon Web Services

---

**First Edition**

**November 2021**

## First Edition - November 2021

This TPC-DS benchmark was conducted by Databricks, Inc. according to standards available at [http://tpc.org/tpc\\_documents\\_current\\_versions/current\\_specifications5.asp](http://tpc.org/tpc_documents_current_versions/current_specifications5.asp). While Databricks believes that comparisons will reflect fair, representative use cases, Databricks cannot and does not guarantee that your experience will match those of other customers or benchmarks, or that comparisons will necessarily be unbiased, or representative of any particular uses of Databricks. A description of our methodology and data is within this document and referenced files.

The Databricks trademarks, trade names, service marks, logos, service names and other distinctive brand features relating to the Databricks products and services are property of Databricks, Inc. and/or its affiliates in the U.S. and other countries.

All content in this document, and in Databricks documents to which this document refers, are the property of Databricks, Inc. and protected by U.S. and international copyright laws. You may not copy, reproduce, distribute, transmit, modify, create derivative works, or in any other way exploit any part of copyrighted material without the prior written permission from Databricks, Inc.

The Databricks Cloud 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 Databricks contact for information on the products or services available in your area. You can find additional information via Databricks international website at <https://www.Databricks.com/>. Actual performance and environmental costs of Databricks products will vary depending on individual customer configurations and conditions.

|  |                                     |
|--|-------------------------------------|
| <b>Abstract</b>  | 5                                   |
| <b>Preface</b>   | <b>Error! Bookmark not defined.</b> |
| TPC-DS™ Benchmark Overview   | 12                                  |
| <b>General Items</b>   | 14                                  |
| 0.1 Test Sponsor   | 14                                  |
| 0.2 Parameter Settings   | 14                                  |
| 0.3 Configuration Diagrams   | 14                                  |
| <b>Clause 2: Logical Database Design Related Items</b>                 | 17                                  |
| 2.1 Database Definition Statements                                     | 17                                  |
| 2.2 Physical Organization  | 17                                  |
| 2.3 Horizontal Partitioning  | 17                                  |
| 2.4 Replication  | 17                                  |
| <b>Clause 3: Scaling and Database Population</b>                       | 18                                  |
| 3.1 Initial Cardinality of Tables                                      | 18                                  |
| 3.2 Distribution of Tables and Logs Across Media                       | 19                                  |
| 3.3 Mapping of Database Partitions/Replications                        | 19                                  |
| 3.4 Implementation of RAID   | 19                                  |
| 3.5 DBGEN Modifications  | 19                                  |
| 3.6 Database Load time   | 20                                  |
| 3.7 Data Storage Ratio   | 20                                  |
| 3.8 Database Load Mechanism Details and Illustration                   | 20                                  |
| 3.9 Qualification Database Configuration                               | 21                                  |
| <b>Clause 4 and 5: Query and Data Maintenance Related Items</b>        | 22                                  |
| 4.1 Query Language   | 22                                  |
| 4.2 Verifying Method of Random Number Generation                       | 22                                  |
| 4.3 Generating Values for Substitution Parameters                      | 22                                  |
| 4.4 Query Text and Output Data from Qualification Database             | 22                                  |
| 4.5 Query Substitution Parameters and Seeds Used                       | 24                                  |
| 4.6 Refresh Setting  | 24                                  |
| 4.7 Source Code of Refresh Functions                                   | 24                                  |
| 4.8 Staging Area   | 24                                  |
| <b>Clause 6: Data Persistence Properties Related Items</b>             | 24                                  |
| <b>Clause 7: Performance Metrics and Execution Rules Related Items</b> | 26                                  |
| 7.1 System Activity  | 26                                  |
| 7.2 Test Steps   | 26                                  |
| 7.3 Timing Intervals for Each Query and Refresh Function               | 27                                  |
| 7.4 Throughput Test Result   | 27                                  |

|  |           |
|--|-----------|
| 7.5 Time for Each Stream                                     | 27        |
| 7.6 Time for Each Refresh Function                           | 27        |
| 7.7 Performance Metrics                                      | 27        |
| <b>Clause 8: SUT and Driver Implementation Related Items</b> | <b>28</b> |
| 8.1 Driver   | 28        |
| 8.2 Implementation Specific Layer (ISL)                      | 28        |
| 8.3 Profile-Directed Optimization                            | 28        |
| <b>Clause 9: Pricing Related Items</b>                       | <b>29</b> |
| 9.1 Hardware and Software Used                               | 29        |
| 9.2 Availability Date  | 29        |
| 9.3 Country-Specific Pricing                                 | 29        |
| <b>Clause 11: Audit Related Items</b>                        | <b>30</b> |
| Auditor's Information and Attestation Letter                 | 30        |
| <b>Supporting Files Index</b>                                | <b>33</b> |
| <b>Appendix A: Provisioning Compute Services</b>             | <b>35</b> |
| <b>Appendix B: Third Party Price Quotes</b>                  | <b>35</b> |

# Abstract

This document contains the methodology and results of the TPC Benchmark™ DS (TPC-DS) test conducted in conformance with the requirements of the TPC-DS Standard Specification, Revision 3.2.0.

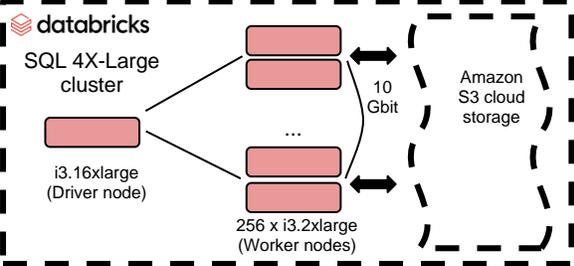
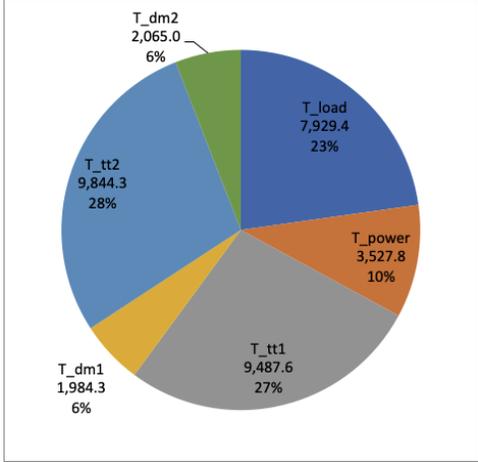
The test was conducted at a Scale Factor of 100,000 GB with 256 Servers running Databricks DBSQL 8.3

## Measured Configuration

| Company Name     | Cluster Node | Database Software  | Operating System |
|------------------|--------------|--------------------|------------------|
| Databricks, Inc. | Amazon EC2   | Databricks SQL 8.3 | Linux            |

## TPC Benchmark™ DS Metrics

| Total System Cost  | TPC-DS Throughput | Price/Performance<br>(USD /<br>kQphDS@100000GB) | Availability Date |
|--------------------|-------------------|---|-------------------|
| <b>\$5,190,345</b> | 32,941,245        | <b>\$157.57</b>                                 | As of Publication |

|   |                                     |  |  |                |
|---|-------------------------------------|--|--|----------------|
|    | <b>Databricks SQL 8.3</b>           |  | TPC-DS: 3.2.0<br>TPC-Pricing: 2.7.0<br>Report Date: 2021-11-02 |                |
| <b>Total System Cost</b>  | <b>TPC-DS Throughput</b>            | <b>Price/Performance</b>   | <b>System Availability Date</b>                                |                |
| <b>\$5,190,345</b><br>USD   | <b>32,941,245</b><br>QphDS@100000GB | <b>\$157.57</b><br>USD/kQphDS@100000GB   | As of Publication  |                |
| <b>Dataset Size</b>   | <b>Database Manager</b>             | <b>Operating System</b>  | <b>Other Software</b>  | <b>Cluster</b> |
| 100,000 GB  | Databricks PhotonEngine 8.3         | Linux  | N/A  | Yes            |
|  <p><b>Benchmarked Configuration</b></p> |                                     |  <p><b>Elapsed Time</b></p> |  |                |
| Load includes backup = No   |                                     | RAID = No  |  |                |
| <b>System Configuration:</b>  |                                     | <b>Databricks DBSQL 4X-Large</b>   |  |                |
| Servers:  |                                     | 1 x i3.16xlarge + 256 x i3.2xlarge   |  |                |
| Total Processors/Cores/Threads:   |                                     | 2,112 vCPU   |  |                |
| Total Memory:   |                                     | 16,104 (GiB)   |  |                |
| Total Storage:  |                                     | 501.6 TB (Nodes) + 26.6 TB (Amazon S3)   |  |                |
| Storage Ratio:  |                                     | 5.40   |  |                |
| <b>Server Configuration:</b>  |                                     | <b>Per Worker Node (256)</b>   | <b>Per Driver Node (1)</b>                                     |                |
| Processors/Cores/Threads:   |                                     | 8 vCPU   | 64 vCPU  |                |
| Memory:   |                                     | 61 GiB   | 488 GiB  |                |
| Network:  |                                     | Up to 10 Gigabit   | Up to 10 Gigabit   |                |
| Storage Device:   |                                     | 1 x 1.9 TB NVMe SSD  | 8 x 1.9 TB NVMe SSD  |                |
| <b>Amazon S3 Standard Storage:</b>  |                                     | 26.6 TB (Total usage)  |  |                |

|  |   | <b>Databricks SQL 8.3</b> |                   |                                 | TPC-DS: 3.2.0<br>TPC-Pricing: 2.7.0<br>Report Date: 2021-11-02 |                          |          |
|---|---|---------------------------|-------------------|---------------------------------|--|--------------------------|----------|
| <i>Description</i>  | <i>Part Number</i>  | <i>Source</i>             | <i>Unit Price</i> | <i>Qty</i>                      | <i>Extended Price</i>  | <i>3 Yr. Maintenance</i> |          |
| <b>Server Hardware</b>  |   |                           |                   |                                 |  |                          |          |
| i3.2xlarge  | i3.2xlarge  | 2                         | \$6,937.92        | 256                             | \$1,776,107.52   |                          |          |
| i3.16xlarge   | i3.16xlarge   | 2                         | \$55,503.36       | 1                               | \$55,503.36  |                          |          |
| Hardware Subtotal   |   |                           |                   |                                 | \$1,831,610.88   | 0                        |          |
| <b>Server Software</b>  |   |                           |                   |                                 |  |                          |          |
| DBU for 4X-Large (528 DBUs, \$0.22/DBU)   | DBU for 4X-Large  | 1                         | \$1,017,561.60    | 3                               | \$3,052,684.80   |                          |          |
| Software Subtotal   |   |                           |                   |                                 | \$3,052,684.80   | 0                        |          |
| <b>External Server Storage</b>  |   |                           |                   |                                 |  |                          |          |
| Amazon S3 standard storage  | S3 Standard storage<br>GB/month                                 | 2                         | \$0.023           | 980,582                         | \$22,553.40  |                          |          |
| PUT, COPY, POST, LIST requests to S3 Standard                                     | PUT requests  | 2                         | \$1,401.521       | 36                              | \$50,454.77  |                          |          |
| GET requests to S3 Standard   | GET requests  | 2                         | \$3,111.186       | 36                              | \$112,002.71   |                          |          |
| Storage Subtotal  |   |                           |                   |                                 | \$185,010.88   | 0                        |          |
| <b>Server Subtotal</b>  |   |                           |                   |                                 | <b>\$5,069,306.56</b>  | <b>0</b>                 |          |
| <b>Client Hardware</b>  |   |                           |                   |                                 |  |                          |          |
| i3en.6xlarge  | i3en.6xlarge  | 2                         | \$26,990          | 1                               | \$26,990   |                          |          |
| Hardware Subtotal   |   |                           |                   |                                 | \$26,990   | 0                        |          |
| <b>Client Software</b>  |   |                           |                   |                                 |  |                          |          |
| Standard All Purpose compute (\$0.40/DBU)   | i3en.6xlarge   24 CPUs  <br>192 GB   8.7 DBU/hr  <br>\$3.480/hr | 1                         | \$31,069.44       | 3                               | \$93,208.32  |                          |          |
| <b>Client Subtotal</b>  |   |                           |                   |                                 | <b>\$120,198</b>   | <b>0</b>                 |          |
| <b>Network/Infrastructure</b>   |   |                           |                   |                                 |  |                          |          |
| Lenovo Chromebook S330 Laptop (Included spare)                                    | <a href="#">B07GLV1VC7</a>                                      | 3                         | \$279.99          | 3                               | \$839.97   |                          |          |
| <b>Network/Infrastructure Subtotal</b>  |   |                           |                   |                                 | <b>\$839.97</b>  | <b>0</b>                 |          |
| Source: 1 = Databricks, 2 = Amazon Web Services, 3 = Amazon.com                   |   |                           |                   | <b>Totals</b>                   |  | <b>\$5,190,344.40</b>    | <b>0</b> |
| <b>Audited by Francois Raab, InfoSizing</b>                                       |   |                           |                   | <b>3-Year Cost of Ownership</b> |  | <b>5,190,345</b>         |          |
|   |   |                           |                   | <b>QphDS@100000GB</b>           |  | <b>32,941,245</b>        |          |
|   |   |                           |                   | <b>\$/kQphDS@100000GB</b>       |  | <b>157.57</b>            |          |

“Prices used in TPC benchmarks reflect the actual prices a customer would pay for a one-time purchase of the stated Line Items. 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 Line Items. For complete details, see the pricing section of the TPC Benchmark Standard. If you find that the stated prices are not available according to these terms, please inform the TPC at [pricing@tpc.org](mailto:pricing@tpc.org). Thank you.”



# Databricks SQL 8.3

TPC-DS: 3.2.0  
TPC-Pricing: 2.7.0  
Report Date: 2021-11-02

## Metrics Details:

| Name              | Value    | Unit    |
|-------------------|----------|---------|
| Scale Factor (SF) | 100,000  | GB      |
| Streams           | 4        | Stream  |
| Queries (Q)       | 396      | Queries |
| T_load            | 7,929.40 | Second  |
| T_ld              | 0.0882   | Hour    |
| T_power           | 3,527.80 | Second  |
| T_pt              | 3.9198   | Hour    |
| T_tt1             | 9,487.60 | Second  |
| T_tt2             | 9,844.30 | Second  |
| T_dm1             | 1,984.30 | Second  |
| T_dm2             | 2,065.00 | Second  |
| T_tt              | 5.37     | Hour    |
| T_dm              | 1.1249   | Hour    |

| Load Step       | Start                    | End                      | (sec.)          | (hh:mm:ss)     |
|-----------------|--------------------------|--------------------------|-----------------|----------------|
| Build           | 08/18/21 19:44:21        | 08/18/21 21:56:31        | 7,929.35        | 2:12:09        |
| Audit           | 08/18/21 21:56:31        | 08/18/21 22:13:41        | 1,030.10        | 0:17:10        |
| Finish          | 08/18/21 22:13:41        | 08/18/21 22:13:41        | 0.02            | 0:00:00        |
| <b>Reported</b> | <b>08/18/21 19:44:21</b> | <b>08/18/21 22:13:41</b> | <b>7,929.37</b> | <b>2:12:09</b> |

| Test      | Start            | End              | (sec.)   | (hh:mm:ss) |
|-----------|------------------|------------------|----------|------------|
| Power     | 08/19/21 0:21:18 | 08/19/21 1:20:05 | 3,527.77 | 0:58:48    |
| Thruput-1 | 08/19/21 1:20:07 | 08/19/21 3:58:15 | 9,487.53 | 2:38:08    |
| DM-1      | 08/19/21 3:58:17 | 08/19/21 4:31:22 | 1,984.24 | 0:33:04    |
| Thruput-2 | 08/19/21 4:31:23 | 08/19/21 7:15:28 | 9,844.25 | 2:44:04    |
| DM-2      | 08/19/21 7:15:40 | 08/19/21 7:50:05 | 2,065.00 | 0:34:25    |

| Stream   | Start            | End              | (sec.)   | (hh:mm:ss) |
|----------|------------------|------------------|----------|------------|
| Pt - 0   | 08/19/21 0:21:18 | 08/19/21 1:20:05 | 3,527.77 | 0:58:48    |
| Tt1 - 1  | 08/19/21 1:20:07 | 08/19/21 3:57:05 | 9,417.35 | 2:36:57    |
| Tt1 - 2  | 08/19/21 1:20:07 | 08/19/21 3:55:11 | 9,303.69 | 2:35:04    |
| Tt1 - 3  | 08/19/21 1:20:07 | 08/19/21 3:58:15 | 9,487.53 | 2:38:08    |
| Tt1 - 4  | 08/19/21 1:20:07 | 08/19/21 3:56:55 | 9,407.67 | 2:36:48    |
| Tt2 - 5  | 08/19/21 4:31:23 | 08/19/21 7:15:26 | 9,842.48 | 2:44:02    |
| Tt2 - 6  | 08/19/21 4:31:23 | 08/19/21 7:15:16 | 9,833.11 | 2:43:53    |
| Tt2 - 7  | 08/19/21 4:31:23 | 08/19/21 7:15:11 | 9,827.77 | 2:43:48    |
| Tt2 - 8  | 08/19/21 4:31:23 | 08/19/21 7:15:28 | 9,844.25 | 2:44:04    |
| DMt1 - 1 | 08/19/21 3:58:17 | 08/19/21 4:14:53 | 995.37   | 0:16:35    |
| DMt1 - 2 | 08/19/21 4:14:53 | 08/19/21 4:31:22 | 988.88   | 0:16:29    |
| DMt2 - 3 | 08/19/21 7:15:40 | 08/19/21 7:32:29 | 1,009.85 | 0:16:50    |
| DMt2 - 4 | 08/19/21 7:32:29 | 08/19/21 7:50:05 | 1,055.16 | 0:17:35    |



# Databricks SQL 8.3

TPC-DS: 3.2.0  
TPC-Pricing: 2.7.0  
Report Date: 2021-11-02

## Timing Intervals for Each Query (In Seconds)

| Query | Stream 0 | Stream 1 | Stream 2 | Stream 3 | Stream 4 | Min   | 25%tile | Median | 75%tile | Max   | Stream 5 | Stream 6 | Stream 7 | Stream 8 | Min   | 25%tile | Median | 75%tile | Max   |
|-------|----------|----------|----------|----------|----------|-------|---------|--------|---------|-------|----------|----------|----------|----------|-------|---------|--------|---------|-------|
| 1     | 17.9     | 56.7     | 44.1     | 53.5     | 53.1     | 44.1  | 50.9    | 53.3   | 54.3    | 56.7  | 56.6     | 51.7     | 41       | 59.6     | 41    | 49      | 54.2   | 57.4    | 59.6  |
| 2     | 24.1     | 58.6     | 49.2     | 70.9     | 56.2     | 49.2  | 54.5    | 57.4   | 61.7    | 70.9  | 61.2     | 57.6     | 51.6     | 52.7     | 51.6  | 52.4    | 55.2   | 58.5    | 61.2  |
| 3     | 6.9      | 14.3     | 15.4     | 13.7     | 27.5     | 13.7  | 14.2    | 14.9   | 18.4    | 27.5  | 15.7     | 15.6     | 25.3     | 16.7     | 15.6  | 15.7    | 16.2   | 18.9    | 25.3  |
| 4     | 132.5    | 401.8    | 426.4    | 410.1    | 467      | 401.8 | 408     | 418.3  | 436.6   | 467   | 412.2    | 401.9    | 410.2    | 394.4    | 394.4 | 400     | 406.1  | 410.7   | 412.2 |
| 5     | 40       | 127.6    | 117.3    | 58       | 80.2     | 58    | 74.7    | 98.8   | 119.9   | 127.6 | 94.1     | 141.8    | 160      | 109.4    | 94.1  | 105.6   | 125.6  | 146.4   | 160   |
| 6     | 5.8      | 18.3     | 29.8     | 22.9     | 11.5     | 11.5  | 16.6    | 20.6   | 24.6    | 29.8  | 16.2     | 31.2     | 14.8     | 14.5     | 14.5  | 14.7    | 15.5   | 20      | 31.2  |
| 7     | 21.3     | 52.7     | 51.6     | 33.4     | 55.7     | 33.4  | 47.1    | 52.2   | 53.5    | 55.7  | 33.6     | 44.6     | 48.3     | 25       | 25    | 31.5    | 39.1   | 45.5    | 48.3  |
| 8     | 6.4      | 17.3     | 17.3     | 20.8     | 21       | 17.3  | 17.3    | 19.1   | 20.9    | 21    | 17.1     | 15.1     | 16.6     | 17.8     | 15.1  | 16.2    | 16.9   | 17.3    | 17.8  |
| 9     | 40.7     | 150.5    | 104.2    | 142.9    | 101.1    | 101.1 | 103.4   | 123.6  | 144.8   | 150.5 | 160.2    | 156.6    | 129      | 130.5    | 129   | 130.1   | 143.6  | 157.5   | 160.2 |
| 10    | 12.1     | 33.6     | 30.7     | 20.6     | 33.6     | 20.6  | 28.2    | 32.2   | 33.6    | 33.6  | 19.6     | 19.9     | 41.6     | 28.2     | 19.6  | 19.8    | 24.1   | 31.6    | 41.6  |
| 11    | 72.8     | 173.5    | 236.5    | 239.5    | 185.5    | 173.5 | 182.5   | 211    | 237.3   | 239.5 | 172.5    | 162.8    | 244      | 204      | 162.8 | 170.1   | 188.3  | 214     | 244   |
| 12    | 5.7      | 15.2     | 16.4     | 19.7     | 13.5     | 13.5  | 14.8    | 15.8   | 17.2    | 19.7  | 12.1     | 14.5     | 11.5     | 12.4     | 11.5  | 12      | 12.3   | 12.9    | 14.5  |
| 13    | 19       | 48.2     | 52.3     | 42       | 18.5     | 18.5  | 36.1    | 45.1   | 49.2    | 52.3  | 41.8     | 48.5     | 41       | 32.2     | 32.2  | 38.8    | 41.4   | 43.5    | 48.5  |
| 14    | 142      | 395.5    | 406.9    | 392.3    | 413.7    | 392.3 | 394.7   | 401.2  | 408.6   | 413.7 | 404      | 426.4    | 446.6    | 395.8    | 395.8 | 402     | 415.2  | 431.5   | 446.6 |
| 15    | 7.4      | 17       | 18.3     | 15.1     | 31.3     | 15.1  | 16.5    | 17.7   | 21.6    | 31.3  | 17.8     | 13.8     | 22.4     | 21       | 13.8  | 16.8    | 19.4   | 21.4    | 22.4  |
| 16    | 50.7     | 48.5     | 149.8    | 149.6    | 126.4    | 48.5  | 106.9   | 138    | 149.7   | 149.8 | 109.8    | 147.1    | 167.5    | 146.5    | 109.8 | 137.3   | 146.8  | 152.2   | 167.5 |
| 17    | 29.1     | 72.5     | 58.9     | 58.1     | 64.3     | 58.1  | 58.7    | 61.6   | 66.4    | 72.5  | 100.3    | 84.2     | 63.8     | 55.7     | 55.7  | 61.8    | 74     | 88.2    | 100.3 |
| 18    | 26       | 62.9     | 51.3     | 51.2     | 38.7     | 38.7  | 48.1    | 51.3   | 54.2    | 62.9  | 40.8     | 52       | 42.9     | 44.1     | 40.8  | 42.4    | 43.5   | 46.1    | 52    |
| 19    | 11       | 34.9     | 21.4     | 17.3     | 13.1     | 13.1  | 16.3    | 19.4   | 24.8    | 34.9  | 18.8     | 16.3     | 23.7     | 16.9     | 16.3  | 16.8    | 17.9   | 20      | 23.7  |
| 20    | 6.1      | 17.9     | 11.7     | 20.4     | 26       | 11.7  | 16.4    | 19.2   | 21.8    | 26    | 21.5     | 25.6     | 16.9     | 24.7     | 16.9  | 20.4    | 23.1   | 24.9    | 25.6  |
| 21    | 2.6      | 7.7      | 4.5      | 8.3      | 8.2      | 4.5   | 6.9     | 8      | 8.2     | 8.3   | 6.8      | 4.6      | 6.5      | 5.6      | 4.6   | 5.4     | 6.1    | 6.6     | 6.8   |
| 22    | 3.4      | 10.7     | 11.8     | 22.7     | 10.5     | 10.5  | 10.7    | 11.3   | 14.5    | 22.7  | 8.6      | 7.4      | 7.7      | 7.8      | 7.4   | 7.6     | 7.8    | 8       | 8.6   |
| 23    | 182.3    | 512      | 547.7    | 567.5    | 543.9    | 512   | 535.9   | 545.8  | 552.7   | 567.5 | 633.5    | 668.9    | 597.5    | 727.5    | 597.5 | 624.5   | 651.2  | 683.6   | 727.5 |
| 24    | 228.2    | 704.1    | 647.2    | 749      | 712.7    | 647.2 | 689.9   | 708.4  | 721.8   | 749   | 818.4    | 842.6    | 807      | 856.2    | 807   | 815.6   | 830.5  | 846     | 856.2 |
| 25    | 24.6     | 59       | 61.4     | 66.3     | 67       | 59    | 60.8    | 63.9   | 66.5    | 67    | 45.2     | 52.4     | 65.3     | 84.3     | 45.2  | 50.6    | 58.9   | 70.1    | 84.3  |
| 26    | 13.9     | 34       | 36.2     | 24.4     | 30.7     | 24.4  | 29.1    | 32.4   | 34.6    | 36.2  | 26.1     | 35.9     | 23.4     | 23.5     | 23.4  | 23.5    | 24.8   | 28.6    | 35.9  |
| 27    | 19.4     | 46       | 55.5     | 58.8     | 27.5     | 27.5  | 41.4    | 50.8   | 56.3    | 58.8  | 28       | 43.7     | 44.4     | 39.6     | 28    | 36.7    | 41.7   | 43.9    | 44.4  |
| 28    | 111.6    | 333.1    | 258.7    | 386.5    | 308.4    | 258.7 | 296     | 320.8  | 346.5   | 386.5 | 318      | 310.5    | 329.3    | 354.1    | 310.5 | 316.1   | 323.7  | 335.5   | 354.1 |
| 29    | 70.5     | 102.6    | 135.9    | 96.6     | 96.2     | 96.2  | 96.5    | 99.6   | 110.9   | 135.9 | 105.8    | 120.7    | 120.8    | 108.1    | 105.8 | 107.5   | 114.4  | 120.7   | 120.8 |
| 30    | 17.3     | 14.9     | 24.5     | 25.9     | 28.4     | 14.9  | 22.1    | 25.2   | 26.5    | 28.4  | 19.4     | 22.3     | 26.6     | 14.3     | 14.3  | 18.1    | 20.9   | 23.4    | 26.6  |
| 31    | 27.7     | 101      | 50.4     | 53.7     | 56.3     | 50.4  | 52.9    | 55     | 67.5    | 101   | 40.1     | 83.6     | 57.2     | 70.9     | 40.1  | 52.9    | 64.1   | 74.1    | 83.6  |
| 32    | 9.1      | 11.7     | 19.5     | 33.7     | 32       | 11.7  | 17.6    | 25.8   | 32.4    | 33.7  | 17.6     | 22.9     | 15.6     | 27.3     | 15.6  | 17.1    | 20.3   | 24      | 27.3  |
| 33    | 10.7     | 24.8     | 26.8     | 25.8     | 25.5     | 24.8  | 25.3    | 25.7   | 26.1    | 26.8  | 27.1     | 20.8     | 23.5     | 29.8     | 20.8  | 22.8    | 25.3   | 27.8    | 29.8  |
| 34    | 24.5     | 40.2     | 36.9     | 41.4     | 30.8     | 30.8  | 35.4    | 38.6   | 40.5    | 41.4  | 55.8     | 36.8     | 37.8     | 56.1     | 36.8  | 37.6    | 46.8   | 55.9    | 56.1  |
| 35    | 22.5     | 67.3     | 24.7     | 25.5     | 47.4     | 24.7  | 25.3    | 36.5   | 52.4    | 67.3  | 59.4     | 40.6     | 43.2     | 46       | 40.6  | 42.6    | 44.6   | 49.4    | 59.4  |
| 36    | 15.6     | 48.1     | 66.1     | 58.5     | 41.3     | 41.3  | 46.4    | 53.3   | 60.4    | 66.1  | 51.1     | 35.7     | 32.6     | 44.5     | 32.6  | 34.9    | 40.1   | 46.2    | 51.1  |
| 37    | 24.6     | 66.3     | 64.6     | 85.1     | 88.8     | 64.6  | 65.9    | 75.7   | 86      | 88.8  | 78.1     | 102.2    | 72.8     | 43.6     | 43.6  | 65.5    | 75.5   | 84.1    | 102.2 |
| 38    | 45.7     | 90.4     | 79       | 70.7     | 105.2    | 70.7  | 76.9    | 84.7   | 94.1    | 105.2 | 118.6    | 90.4     | 117.3    | 69.9     | 69.9  | 85.3    | 103.9  | 117.6   | 118.6 |

|    |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 39 | 11.1  | 14.5  | 20.7  | 16.3  | 21.4  | 14.5  | 15.9  | 18.5  | 20.9  | 21.4  | 23.5  | 17.3  | 19.8  | 20.4  | 17.3  | 19.2  | 20.1  | 21.2  | 23.5  |
| 40 | 14.7  | 52.7  | 54.9  | 59.5  | 39.4  | 39.4  | 49.4  | 53.8  | 56.1  | 59.5  | 43.4  | 30.9  | 72.9  | 41.4  | 30.9  | 38.8  | 42.4  | 50.8  | 72.9  |
| 41 | 1     | 2.3   | 1.9   | 3.2   | 2.2   | 1.9   | 2.1   | 2.3   | 2.5   | 3.2   | 2.7   | 1.5   | 3     | 1.7   | 1.5   | 1.7   | 2.2   | 2.8   | 3     |
| 42 | 8.1   | 10.1  | 11    | 14.9  | 16.9  | 10.1  | 10.8  | 13    | 15.4  | 16.9  | 12.9  | 19.4  | 9.4   | 12.5  | 9.4   | 11.7  | 12.7  | 14.5  | 19.4  |
| 43 | 8.8   | 35.1  | 35.7  | 32.1  | 27.4  | 27.4  | 30.9  | 33.6  | 35.3  | 35.7  | 41.4  | 21.9  | 29.5  | 33.8  | 21.9  | 27.6  | 31.7  | 35.7  | 41.4  |
| 44 | 33    | 92.2  | 90.9  | 99.8  | 107.7 | 90.9  | 91.9  | 96    | 101.8 | 107.7 | 60.8  | 117.4 | 123.4 | 121.5 | 60.8  | 103.3 | 119.5 | 122   | 123.4 |
| 45 | 8.2   | 23.4  | 17.6  | 17.1  | 30.9  | 17.1  | 17.5  | 20.5  | 25.3  | 30.9  | 19.8  | 15.3  | 11    | 13.9  | 11    | 13.2  | 14.6  | 16.4  | 19.8  |
| 46 | 17.1  | 35.1  | 39.9  | 26.7  | 30.7  | 26.7  | 29.7  | 32.9  | 36.3  | 39.9  | 28.9  | 22.9  | 39.1  | 24.9  | 22.9  | 24.4  | 26.9  | 31.5  | 39.1  |
| 47 | 36.5  | 132.5 | 108.4 | 98.3  | 125.7 | 98.3  | 105.9 | 117.1 | 127.4 | 132.5 | 104.8 | 97.4  | 97.6  | 80.1  | 80.1  | 93.1  | 97.5  | 99.4  | 104.8 |
| 48 | 12.7  | 41.3  | 23.6  | 43.1  | 32    | 23.6  | 29.9  | 36.7  | 41.8  | 43.1  | 25    | 38.6  | 45.8  | 40.1  | 25    | 35.2  | 39.4  | 41.5  | 45.8  |
| 49 | 26.9  | 25.9  | 58.1  | 28    | 54.9  | 25.9  | 27.5  | 41.5  | 55.7  | 58.1  | 57.8  | 45.7  | 62    | 47.9  | 45.7  | 47.4  | 52.9  | 58.9  | 62    |
| 50 | 96.9  | 360.1 | 356.9 | 269.2 | 385.3 | 269.2 | 335   | 358.5 | 366.4 | 385.3 | 360.5 | 322.8 | 255.6 | 335.6 | 255.6 | 306   | 329.2 | 341.8 | 360.5 |
| 51 | 31.8  | 40.3  | 89.9  | 58.5  | 57.1  | 40.3  | 52.9  | 57.8  | 66.4  | 89.9  | 63.2  | 71.5  | 28.5  | 62.9  | 28.5  | 54.3  | 63.1  | 65.3  | 71.5  |
| 52 | 6.4   | 15.4  | 12.9  | 13.6  | 11.5  | 11.5  | 12.6  | 13.3  | 14.1  | 15.4  | 10.1  | 18.1  | 12.9  | 19.4  | 10.1  | 12.2  | 15.5  | 18.4  | 19.4  |
| 53 | 9     | 35.6  | 27.1  | 32.6  | 21.7  | 21.7  | 25.8  | 29.9  | 33.4  | 35.6  | 28.4  | 28.6  | 18.9  | 49.5  | 18.9  | 26    | 28.5  | 33.8  | 49.5  |
| 54 | 16.9  | 33    | 41    | 46.1  | 35.5  | 33    | 34.9  | 38.3  | 42.3  | 46.1  | 48.9  | 44.6  | 23.3  | 29.2  | 23.3  | 27.7  | 36.9  | 45.7  | 48.9  |
| 55 | 4.8   | 23    | 15.1  | 18.7  | 13.5  | 13.5  | 14.7  | 16.9  | 19.8  | 23    | 14.8  | 11.9  | 14.5  | 12.2  | 11.9  | 12.1  | 13.4  | 14.6  | 14.8  |
| 56 | 13.8  | 17.6  | 20.2  | 32.6  | 38.8  | 17.6  | 19.6  | 26.4  | 34.2  | 38.8  | 23.4  | 30.9  | 22.8  | 18.4  | 18.4  | 21.7  | 23.1  | 25.3  | 30.9  |
| 57 | 28.9  | 63.7  | 62.5  | 80.1  | 79.2  | 62.5  | 63.4  | 71.5  | 79.4  | 80.1  | 53.7  | 103.6 | 88.4  | 48.2  | 48.2  | 52.3  | 71.1  | 92.2  | 103.6 |
| 58 | 5     | 15    | 25.9  | 25.3  | 11.4  | 11.4  | 14.1  | 20.2  | 25.5  | 25.9  | 17.4  | 15.7  | 15.8  | 15.4  | 15.4  | 15.6  | 15.8  | 16.2  | 17.4  |
| 59 | 33.7  | 82.8  | 113.9 | 94.7  | 89.7  | 82.8  | 88    | 92.2  | 99.5  | 113.9 | 99.6  | 85.6  | 105.9 | 106.3 | 85.6  | 96.1  | 102.8 | 106   | 106.3 |
| 60 | 16.3  | 28.7  | 37.6  | 36.2  | 34    | 28.7  | 32.7  | 35.1  | 36.6  | 37.6  | 32.7  | 40.3  | 32.1  | 31.9  | 31.9  | 32.1  | 32.4  | 34.6  | 40.3  |
| 61 | 14.6  | 30.7  | 31.1  | 30.5  | 30.9  | 30.5  | 30.7  | 30.8  | 31    | 31.1  | 35.6  | 27    | 22.8  | 32.3  | 22.8  | 26    | 29.7  | 33.1  | 35.6  |
| 62 | 9     | 22.2  | 28.6  | 26.2  | 29.6  | 22.2  | 25.2  | 27.4  | 28.9  | 29.6  | 36.8  | 31.9  | 30.6  | 29.5  | 29.5  | 30.3  | 31.3  | 33.1  | 36.8  |
| 63 | 8.8   | 20.7  | 28.4  | 23.6  | 20.9  | 20.7  | 20.9  | 22.3  | 24.8  | 28.4  | 19.8  | 24.7  | 31.7  | 16.7  | 16.7  | 19    | 22.3  | 26.5  | 31.7  |
| 64 | 245.9 | 675.2 | 425.8 | 646.6 | 545   | 425.8 | 515.2 | 595.8 | 653.8 | 675.2 | 694.4 | 707.4 | 729.8 | 621.5 | 621.5 | 676.2 | 700.9 | 713   | 729.8 |
| 65 | 62    | 138.5 | 146.4 | 116.5 | 160.6 | 116.5 | 133   | 142.5 | 150   | 160.6 | 88    | 131.8 | 158.2 | 197.4 | 88    | 120.9 | 145   | 168   | 197.4 |
| 66 | 14.8  | 32.4  | 32.7  | 52.6  | 33    | 32.4  | 32.6  | 32.9  | 37.9  | 52.6  | 25.2  | 51.2  | 26.4  | 34.7  | 25.2  | 26.1  | 30.6  | 38.8  | 51.2  |
| 67 | 102.8 | 297.8 | 388.2 | 344   | 369.9 | 297.8 | 332.5 | 357   | 374.5 | 388.2 | 312.1 | 311.7 | 289.1 | 376.3 | 289.1 | 306.1 | 311.9 | 328.2 | 376.3 |
| 68 | 8.2   | 27.4  | 15.1  | 21.5  | 27.6  | 15.1  | 19.9  | 24.5  | 27.5  | 27.6  | 18.1  | 16    | 15.7  | 23.3  | 15.7  | 15.9  | 17.1  | 19.4  | 23.3  |
| 69 | 11.6  | 23.8  | 8.2   | 25    | 24.9  | 8.2   | 19.9  | 24.4  | 24.9  | 25    | 30    | 10.2  | 43.8  | 7.4   | 7.4   | 9.5   | 20.1  | 33.5  | 43.8  |
| 70 | 20.8  | 30.9  | 56.2  | 70.3  | 84.6  | 30.9  | 49.9  | 63.3  | 73.9  | 84.6  | 68    | 74.8  | 40.5  | 58.9  | 40.5  | 54.3  | 63.5  | 69.7  | 74.8  |
| 71 | 19.3  | 40.1  | 55.8  | 38.2  | 37.1  | 37.1  | 37.9  | 39.2  | 44    | 55.8  | 61    | 45.7  | 38    | 58    | 38    | 43.8  | 51.9  | 58.8  | 61    |
| 72 | 41.1  | 90.3  | 91.3  | 89    | 95.6  | 89    | 90    | 90.8  | 92.4  | 95.6  | 102.3 | 116.8 | 79.8  | 144.1 | 79.8  | 96.7  | 109.6 | 123.6 | 144.1 |
| 73 | 6.7   | 24.5  | 28.7  | 12.8  | 13.3  | 12.8  | 13.2  | 18.9  | 25.6  | 28.7  | 14.9  | 14.1  | 14.2  | 16.4  | 14.1  | 14.2  | 14.6  | 15.3  | 16.4  |
| 74 | 57.6  | 109.2 | 148.6 | 151.3 | 123.9 | 109.2 | 120.2 | 136.3 | 149.3 | 151.3 | 120.3 | 144.9 | 122.2 | 138.3 | 120.3 | 121.7 | 130.3 | 140   | 144.9 |
| 75 | 139.5 | 310.3 | 301.2 | 277.9 | 283.8 | 277.9 | 282.3 | 292.5 | 303.5 | 310.3 | 330   | 247.1 | 315.5 | 329   | 247.1 | 298.4 | 322.3 | 329.3 | 330   |
| 76 | 31.5  | 114   | 111.6 | 142.9 | 148.7 | 111.6 | 113.4 | 128.5 | 144.4 | 148.7 | 101.1 | 129.2 | 136.1 | 103.8 | 101.1 | 103.1 | 116.5 | 130.9 | 136.1 |
| 77 | 10.6  | 52.6  | 32.4  | 37.4  | 23.5  | 23.5  | 30.2  | 34.9  | 41.2  | 52.6  | 36.9  | 29.7  | 40.6  | 34.1  | 29.7  | 33    | 35.5  | 37.8  | 40.6  |
| 78 | 104.3 | 250.9 | 251.5 | 282.8 | 262   | 250.9 | 251.4 | 256.8 | 267.2 | 282.8 | 298   | 333.3 | 344.9 | 278.9 | 278.9 | 293.2 | 315.7 | 336.2 | 344.9 |
| 79 | 14.5  | 49.7  | 33.1  | 26.7  | 20.2  | 20.2  | 25.1  | 29.9  | 37.3  | 49.7  | 24.4  | 36.6  | 30.9  | 26.5  | 24.4  | 26    | 28.7  | 32.3  | 36.6  |
| 80 | 44.8  | 132.1 | 125.3 | 128   | 129.1 | 125.3 | 127.3 | 128.6 | 129.9 | 132.1 | 138.2 | 102.1 | 125.3 | 143.6 | 102.1 | 119.5 | 131.8 | 139.6 | 143.6 |
| 81 | 12.7  | 24.9  | 22.2  | 39.3  | 35.7  | 22.2  | 24.2  | 30.3  | 36.6  | 39.3  | 36.4  | 31    | 25.5  | 25.8  | 25.5  | 25.7  | 28.4  | 32.4  | 36.4  |

|    |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 82 | 42.9  | 178   | 123   | 137.6 | 151.7 | 123   | 134   | 144.7 | 158.3 | 178   | 127.8 | 177.9 | 112.3 | 152.8 | 112.3 | 123.9 | 140.3 | 159.1 | 177.9 |
| 83 | 14.6  | 18.8  | 20.9  | 31.1  | 24.8  | 18.8  | 20.4  | 22.9  | 26.4  | 31.1  | 12.9  | 22.1  | 20    | 25.6  | 12.9  | 18.2  | 21.1  | 23    | 25.6  |
| 84 | 9     | 20    | 31.4  | 26.5  | 22.2  | 20    | 21.7  | 24.4  | 27.7  | 31.4  | 22.5  | 23.6  | 21.5  | 23.2  | 21.5  | 22.3  | 22.9  | 23.3  | 23.6  |
| 85 | 19.5  | 30.5  | 47.3  | 31.9  | 62.8  | 30.5  | 31.6  | 39.6  | 51.2  | 62.8  | 35.5  | 32.2  | 42.2  | 46.5  | 32.2  | 34.7  | 38.9  | 43.3  | 46.5  |
| 86 | 10.4  | 42    | 25.4  | 33.5  | 34.5  | 25.4  | 31.5  | 34    | 36.4  | 42    | 35.8  | 17.9  | 25.7  | 30.7  | 17.9  | 23.8  | 28.2  | 32    | 35.8  |
| 87 | 45.5  | 103.8 | 71.3  | 125.8 | 90.5  | 71.3  | 85.7  | 97.2  | 109.3 | 125.8 | 94.5  | 78.5  | 76.5  | 115.1 | 76.5  | 78    | 86.5  | 99.7  | 115.1 |
| 88 | 77.2  | 307.3 | 330.9 | 153.5 | 381.1 | 153.5 | 268.9 | 319.1 | 343.5 | 381.1 | 393.2 | 332   | 377.2 | 277.2 | 277.2 | 318.3 | 354.6 | 381.2 | 393.2 |
| 89 | 10.7  | 39.7  | 43.2  | 20.7  | 24.1  | 20.7  | 23.3  | 31.9  | 40.6  | 43.2  | 24.7  | 22.6  | 34.3  | 20.7  | 20.7  | 22.1  | 23.7  | 27.1  | 34.3  |
| 90 | 8.6   | 31.9  | 36.5  | 32.7  | 43.4  | 31.9  | 32.5  | 34.6  | 38.2  | 43.4  | 29    | 22.7  | 36.4  | 25.5  | 22.7  | 24.8  | 27.3  | 30.9  | 36.4  |
| 91 | 4.6   | 14.9  | 16.2  | 14.9  | 9.6   | 9.6   | 13.6  | 14.9  | 15.2  | 16.2  | 12.9  | 10.6  | 19    | 16.5  | 10.6  | 12.3  | 14.7  | 17.1  | 19    |
| 92 | 5.9   | 23    | 13    | 26.8  | 13.1  | 13    | 13.1  | 18.1  | 24    | 26.8  | 26.4  | 33.4  | 19.3  | 18.9  | 18.9  | 19.2  | 22.9  | 28.2  | 33.4  |
| 93 | 124.8 | 387.6 | 358.1 | 404.6 | 280   | 280   | 338.6 | 372.9 | 391.9 | 404.6 | 440.7 | 295.3 | 381.6 | 383   | 295.3 | 360   | 382.3 | 397.4 | 440.7 |
| 94 | 23.9  | 87.2  | 84.2  | 78.9  | 77.8  | 77.8  | 78.6  | 81.6  | 85    | 87.2  | 74.9  | 91.2  | 70.3  | 47.6  | 47.6  | 64.6  | 72.6  | 79    | 91.2  |
| 95 | 58.7  | 126.5 | 171.4 | 152.3 | 126.3 | 126.3 | 126.5 | 139.4 | 157.1 | 171.4 | 185.7 | 156.4 | 139.8 | 191   | 139.8 | 152.3 | 171.1 | 187   | 191   |
| 96 | 21.3  | 46.6  | 66.1  | 48.1  | 56.5  | 46.6  | 47.7  | 52.3  | 58.9  | 66.1  | 93    | 51.5  | 44.3  | 49.5  | 44.3  | 48.2  | 50.5  | 61.9  | 93    |
| 97 | 32.7  | 117.1 | 102.1 | 120   | 90.8  | 90.8  | 99.3  | 109.6 | 117.8 | 120   | 112.4 | 116.1 | 88.1  | 77.3  | 77.3  | 85.4  | 100.3 | 113.3 | 116.1 |
| 98 | 13.4  | 38.4  | 20.1  | 20.6  | 32.4  | 20.1  | 20.5  | 26.5  | 33.9  | 38.4  | 22.3  | 22.8  | 28.4  | 29.2  | 22.3  | 22.7  | 25.6  | 28.6  | 29.2  |
| 99 | 54.6  | 38.8  | 58.5  | 50.4  | 26.5  | 26.5  | 35.7  | 44.6  | 52.4  | 58.5  | 54.1  | 42.9  | 54.8  | 43.8  | 42.9  | 43.6  | 49    | 54.3  | 54.8  |

### Timing Intervals for Each Refresh Function (In Seconds)

| DM Fx | R-Run 1 | R-Run 2 | R-Run 3 | R-Run 4 | Min   | 25%tile | Median | 75%tile | Max   |
|-------|---------|---------|---------|---------|-------|---------|--------|---------|-------|
| DF_CS | 318.9   | 312.3   | 328.7   | 317.4   | 312.3 | 316.2   | 318.2  | 321.3   | 328.7 |
| DF_I  | 16.9    | 16.8    | 18.6    | 17.4    | 16.8  | 16.9    | 17.1   | 17.7    | 18.6  |
| DF_SS | 735     | 722.9   | 724.2   | 735.1   | 722.9 | 723.9   | 729.6  | 735     | 735.1 |
| DF_WS | 402.3   | 493.1   | 508     | 511.3   | 402.3 | 470.4   | 500.5  | 508.8   | 511.3 |
| LF_CR | 36.3    | 30      | 34      | 31.6    | 30    | 31.2    | 32.8   | 34.6    | 36.3  |
| LF_CS | 54.6    | 51.4    | 54.4    | 58.4    | 51.4  | 53.6    | 54.5   | 55.5    | 58.4  |
| LF_I  | 13.4    | 14.5    | 11.2    | 17.2    | 11.2  | 12.9    | 14     | 15.2    | 17.2  |
| LF_SR | 33.2    | 33.3    | 28.8    | 81.6    | 28.8  | 32.1    | 33.3   | 45.4    | 81.6  |
| LF_SS | 52.2    | 59.4    | 75.9    | 57.8    | 52.2  | 56.4    | 58.6   | 63.5    | 75.9  |
| LF_WR | 26.8    | 26.7    | 27.3    | 26.9    | 26.7  | 26.8    | 26.9   | 27      | 27.3  |
| LF_WS | 43.8    | 50.6    | 54      | 46.5    | 43.8  | 45.8    | 48.6   | 51.5    | 54    |

# Preface

## TPC-DS™ Benchmark Overview

The TPC Benchmark™ DS (TPC-DS) is a decision support benchmark that models several generally applicable aspects of a decision support system, including queries and data maintenance. The benchmark provides a representative evaluation of performance as a general purpose decision support system.

This benchmark illustrates decision support systems that:

- Examine large volumes of data;
- Give answers to real-world business questions;
- Execute queries of various operational requirements and complexities (e.g., ad-hoc, reporting, iterative OLAP, data mining);
- Are characterized by high CPU and IO load;
- Are periodically synchronized with source OLTP databases through database maintenance functions.
- Run on “Big Data” solutions, such as RDBMS as well as Hadoop/Spark based systems.

A benchmark result measures query response time in single user mode, query throughput in multi user mode and data maintenance performance for a given hardware, operating system, and data processing system configuration under a controlled, complex, multi-user decision support workload.

The purpose of TPC benchmarks is to provide relevant, objective performance data to industry users. To achieve that purpose, TPC benchmark specifications require benchmark tests be implemented with systems, products, technologies and pricing that:

- A. Are generally available to users;
- B. Are relevant to the market segment that the individual TPC benchmark models or represents (e.g., TPC-DS models and represents complex, high data volume, decision support environments);
- C. Would plausibly be implemented by a significant number of users in the market segment modeled or represented by the benchmark.

In keeping with these requirements, the TPC-DS database must be implemented using commercially available data processing software, and its queries must be executed via SQL interface. 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, which improve benchmark results but not real-world performance or pricing, are prohibited.

TPC benchmark results are expected to be accurate representations of system performance. Therefore, there are specific guidelines that are expected to be followed when measuring those results. The approach or methodology to be used in the measurements are either explicitly described in the specification or left to the discretion of the test sponsor.

When not described in the specification, the methodologies and approaches used must meet the following requirements:

- The approach is an accepted engineering practice or standard;
- The approach does not enhance the result;
- Equipment used in measuring the results is calibrated according to established quality standards;
- Fidelity and candor is maintained in reporting any anomalies in the results, even if not specified in the benchmark requirements.

Further information is available at <http://www.tpc.org/>.

# General Items

## 0.1 Test Sponsor

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

This benchmark was sponsored by Databricks, Inc.

## 0.2 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 but not limited to:*

- *Database Tuning Options*
- *Optimizer/Query execution options*
- *Query processing tool/language configuration parameters*
- *Recovery/commit options*
- *Consistency/locking options*
- *Operating system and configuration parameters*
- *Configuration parameters and options for any other software component incorporated into the pricing structure*
- *Compiler optimization options*

*This requirement can be satisfied by providing a full list of all parameters and options, as long as all those which have been modified from their default values have been clearly identified and these parameters and options are only set once.*

The Supporting File Archive contains the Operating System and DBMS parameters used in this benchmark.

## 0.3 Configuration Diagrams

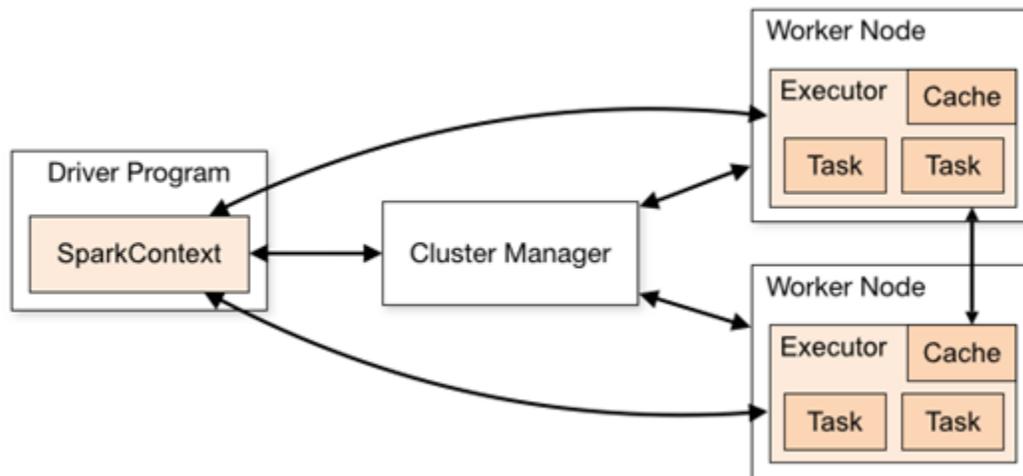
*Diagrams of both measured and priced configurations must be provided, accompanied by a description of the differences. This includes, but is not limited to:*

- *Number and type of processors (including size of L2 cache);*
- *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, including routers, workstations, terminals, etc., that were physically used in the test or are incorporated into the pricing structure;*

- *Type and the run-time execution location of software components (e.g., data processing system, query processing tools/languages, middleware components, software drivers, etc.).*

## Architecture

Following is an illustration of the Databricks Cluster architecture.



- The **Driver Program** receives requests and dispatches them to the Worker Nodes.
- The **Worker Nodes** execute the queries.
- The **Cluster Manager** controls the cloud instances and allocates resources to Spark applications.
- The **Executor** is the process running on each Worker Node that manages the cache and the tasks on that node.

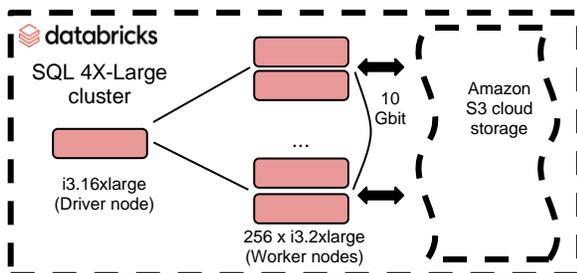
For this benchmark, the Databricks Cluster is deployed on AWS with one instance hosting the Driver Program and the Cluster Manager and 256 instances each hosting the Worker Nodes. Amazon's S3 cloud storage service Storage is hosting the Databricks object store.

## Benchmark Configuration

### Databricks Cluster: 4X-Large

|                             | Driver              | Worker              |
|-----------------------------|---------------------|---------------------|
| <b>AWS Instance Count</b>   | 1                   | 256                 |
| <b>Instance Type</b>        | i3.16xlarge         | i3.2xlarge          |
| <b>vCPU per Instance</b>    | 64                  | 8                   |
| <b>Memory per Instance</b>  | 488 GiB             | 61 GiB              |
| <b>Storage per Instance</b> | 8 x 1.9 TB NVMe SSD | 1 x 1.9 TB NVMe SSD |

|                                |           |
|--------------------------------|-----------|
| <b>Object Storage Type</b>     | Amazon S3 |
| <b>Object Storage Capacity</b> | 26.6 TB   |



### Links to documentation

- Databricks: <https://docs.databricks.com/sql/admin/sql-endpoints.html#cluster-size>
- AWS: <https://aws.amazon.com/ec2/instance-types/i3/>

# Clause 2: Logical Database Design Related Items

## 2.1 Database Definition Statements

*Listings must be provided for the DDL scripts and must include all table definition statements and all other statements used to set up the test and qualification databases.*

The Supporting File Archive contains the table definitions and all other statements used to set up the test and qualification databases.

## 2.2 Physical Organization

*The physical organization of tables and indices within the test and qualification databases must be disclosed. If the column ordering of any table is different from that specified in Clause 2.3 or 2.4, it must be noted.*

Horizontal partitioning is used as described in 2.3.

## 2.3 Horizontal Partitioning

*If any directives to DDLs are used to horizontally partition tables and rows in the test and qualification databases, these directives, DDLs, and other details necessary to replicate the partitioning behavior must be disclosed.*

The following table represents the tables which use horizontal partitioning. Each uses a granularity of 1 day/value per partition.

| <b>table_name</b> | <b>column_name</b> |
|-------------------|--------------------|
| catalog_sales     | cs_sold_date_sk    |
| inventory         | inv_date_sk        |
| store_sales       | ss_sold_date_sk    |
| web_sales         | ws_sold_date_sk    |

## 2.4 Replication

*Any replication of physical objects must be disclosed and must conform to the requirements of Clause 2.5.3.*

No replication was used.

# Clause 3: Scaling and Database Population

## 3.1 Initial Cardinality of Tables

The cardinality (e.g., the number of rows) of each table of the test database, as it existed at the completion of the database load (see Clause 7.1.2) must be disclosed.

Table 3.1 lists the TPC Benchmark DS defined tables and the row count for each table as they existed upon completion of the build.

**Table 3.1 Initial Number of Rows**

| table_name             | row_count       |
|------------------------|-----------------|
| call_center            | 60              |
| catalog_page           | 50,000          |
| catalog_returns        | 14,406,204,801  |
| catalog_sales          | 144,014,807,953 |
| customer               | 100,000,000     |
| customer_address       | 50,000,000      |
| customer_demographics  | 1,920,800       |
| date_dim               | 73,049          |
| household_demographics | 7,200           |
| income_band            | 20              |
| inventory              | 1,965,337,830   |
| item                   | 502,000         |
| promotion              | 2,500           |
| reason                 | 75              |
| ship_mode              | 25              |
| store                  | 1,902           |
| store_returns          | 28,792,385,679  |
| store_sales            | 288,002,140,051 |
| time_dim               | 86,400          |

|             |                |
|-------------|----------------|
| warehouse   | 30             |
| web_page    | 5,004          |
| web_returns | 7,199,635,551  |
| web_sales   | 71,995,997,062 |
| web_site    | 96             |

## 3.2 Distribution of Tables and Logs Across Media

Databricks provides an abstraction over the cloud object store with Delta which stores both the data and the change log on the underlying object store (Amazon S3). This provides an equivalent guarantee as the underlying storage and data distribution is handled transparently by S3. Details are included in the [Delta paper](#).

## 3.3 Mapping of Database Partitions/Replications

Neither database partitions nor replications are mapped to specific devices.

## 3.4 Implementation of RAID

*Implementations may use some form of RAID. The RAID level used must be disclosed for each device. If RAID is used in an implementation, the logical intent of its use must be disclosed.*

Amazon S3 Standard, S3 Standard-IA, and S3 Glacier storage classes redundantly store your objects on multiple devices across a minimum of three Availability Zones (AZs) in an Amazon S3 Region before returning SUCCESS. The S3 One Zone-IA storage class stores data redundantly across multiple devices within a single AZ. These services are designed to sustain concurrent device failures by quickly detecting and repairing any lost redundancy, and they also regularly verify the integrity of your data using checksums.

## 3.5 DBGEN Modifications

*The version number (i.e., the major revision number, the minor revision number, and third tier number) of dsdgen must be disclosed. Any modifications to the dsdgen source code (see Appendix B:) must be disclosed. In the event that a program other than dsdgen was used to populate the database, it must be disclosed in its entirety.*

Dsdgen version 2.13.0 was used. No modifications were made.

Due to a TPC-internal error during the production of version 3.2.0 of the TPC-DS kit, the benchmark execution had to use version 2.13 of the kit. It was confirmed by the TPC that the

only changes between these two versions of the kit is the version number set in the tools/release.h parameter file.

### 3.6 Database Load time

*The database load time for the test database (see Clause 7.4.3.7) must be disclosed.*

The database load time was 7,929 seconds.

### 3.7 Data Storage Ratio

*The data storage ratio must be disclosed. It is computed by dividing the total data storage of the priced configuration (expressed in GB) by SF corresponding to the scale factor chosen for the test database as defined in Clause 3.1. The ratio must be reported to the nearest 1/100th, rounded up. For example, a system configured with 96 disks of 2.1 GB capacity for a 100GB test database has a data storage ratio of 2.02.*

Amazon S3 storage used = 26.6 TB

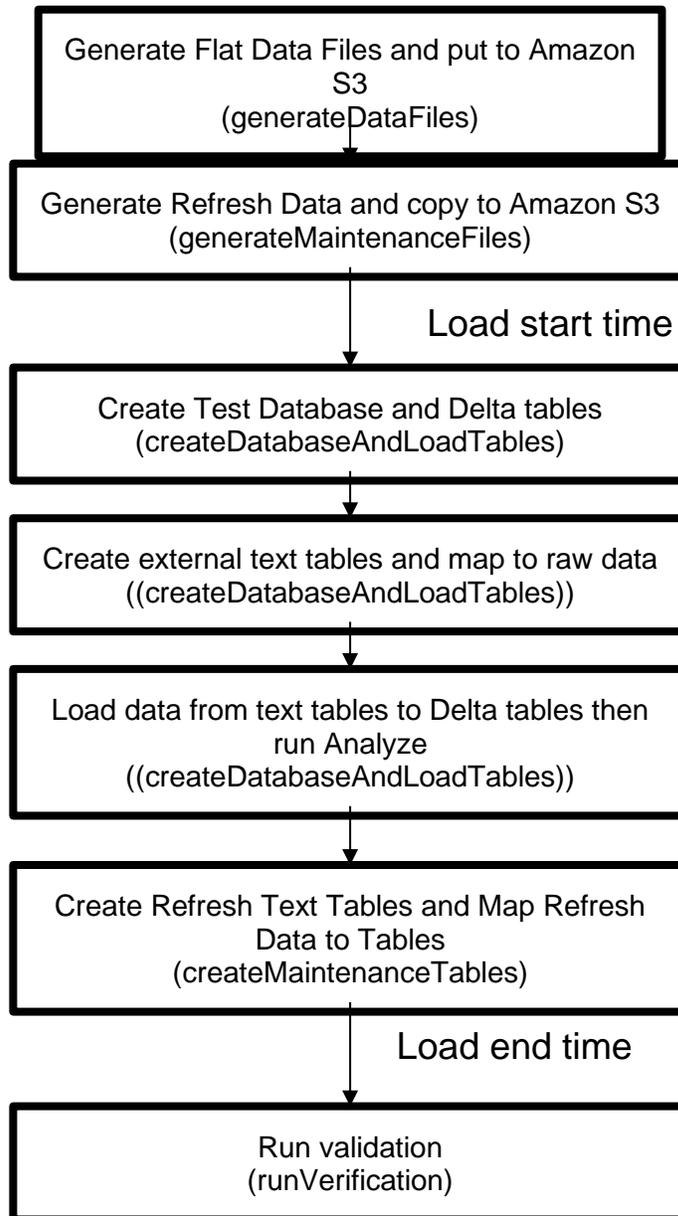
Total local storage Capacity (Disk) =  $(8 \times 1.9)$  (Driver node) +  $(1.9 * 1) * 256$  (Worker nodes) = 513,638 GB

The data storage ratio is  $(513,638 + 26.6 * 1024) / 100,000 = 5.40$ .

### 3.8 Database Load Mechanism Details and Illustration

*The details of the database load must be disclosed, including a block diagram illustrating the overall process. Disclosure of the load procedure includes all steps, scripts, input and configuration files required to completely reproduce the test and qualification databases.*

The tables were loaded as shown in Figure 3.8. All of the related source code and scripts are included in the Supporting Files.



### 3.9 Qualification Database Configuration

*Any differences between the configuration of the qualification database and the test database must be disclosed.*

The qualification database used identical scripts to create and load the data with changes only to adjust for the database scale factor.

# Clause 4 and 5: Query and Data Maintenance Related Items

## 4.1 Query Language

*The query language used to implement the queries must be identified.*

SQL was the query language used to implement the queries.

## 4.2 Verifying Method of Random Number Generation

*The method of verification for the random number generation must be described unless the supplied dsdgen and dsqgen were used.*

The TPC supplied dsdgen version 2.13.0 and dsqgen version 2.13.0 were used.

Due to a TPC-internal error during the production of version 3.2.0 of the TPC-DS kit, the benchmark execution had to use version 2.13 of the kit. It was confirmed by the TPC that the only changes between these two versions of the kit is the version number set in the tools/release.h parameter file.

## 4.3 Generating Values for Substitution Parameters

*The method used to generate values for substitution parameters must be disclosed. The version number (i.e., the major revision number, the minor revision number, and third tier number) of dsqgen must be disclosed.*

TPC supplied dsqgen version 2.13.0 was used to generate the substitution parameters:

```
./dsqgen -DIRECTORY ../../scripts/query_templates/ -INPUT  
../../scripts/query_templates/templates.lst -VERBOSE Y -SCALE 100000 -DIALECT databricks -  
OUTPUT_DIR /tmp/tpcds_100000_dbsql_qual_run_3/Clause_7//sqls -STREAMS 9 -rngseed  
08182213409
```

## 4.4 Query Text and Output Data from Qualification Database

*The executable query text used for query validation must be disclosed along with the corresponding output data generated during the execution of the query text against the qualification database. If minor modifications have been applied to any functional query definitions or approved variants in order to obtain executable query text, these modifications must be disclosed and justified. The justification for a particular minor query modification can*

*apply collectively to all queries for which it has been used. The output data for the power and Throughput Tests must be made available electronically upon request.*

Supporting Files Archive contains the actual query text and query output. Following are the modifications to the query text.

The following MQM are used:

- Use back quote instead of double quote for select-list expression aliases. (MQM e.1)
  - Q16
  - Q32
  - Q50
  - Q62
  - Q92
  - Q94
  - Q95
  - Q99
- Aliasing of column names in SELECT list. (MQM e.5)
  - Q35
  - Q39
- Use vendor-specific date expression syntax. (MQM f.1)
  - Q5
  - Q12
  - Q16
  - Q20
  - Q21
  - Q32
  - Q37
  - Q40
  - Q72
  - Q77
  - Q80
  - Q82
  - Q92
  - Q94
  - Q95
  - Q98
- Scalar functions (such as CAST) whose sole purpose is to affect result precision. (MQM f.6)
  - Q78
- Query results are inserted in a file. (clause 4.2.5)
  - Q64

## 4.5 Query Substitution Parameters and Seeds Used

*All the query substitution parameters used during the performance test must be disclosed in tabular format, along with the seeds used to generate these parameters.*

Supporting Files Archive contains the query substitution parameters and seed used.

## 4.6 Refresh Setting

*All query and refresh session initialization parameters, settings and commands must be disclosed.*

Supporting Files Archive contains the query and scripts.

## 4.7 Source Code of Refresh Functions

*The details of how the data maintenance functions were implemented must be disclosed (including source code of any non-commercial program used).*

Supporting Files Archive contains the Source Code of refresh functions.

## 4.8 Staging Area

*Any object created in the staging area (see Clause 5.1.8 for definition and usage restrictions) used to implement the data maintenance functions must be disclosed. Also, any disk storage used for the staging area must be priced, and any mapping or virtualization of disk storage must be disclosed.*

No staging area was used.

# Clause 6: Data Accessibility Properties Related Items

*A description of how the data accessibility requirements were met must be disclosed.*

In this benchmark, the Data Accessibility requirements are met by providing documentation of the Data Accessibility features supported by the benchmark configuration.

Databricks uses object stores provided by the major cloud provider Amazon AWS. Databricks provides an abstraction over the cloud object store with Delta which stores both the data and the change log on the underlying object store and hence provides an equivalent guarantee as the underlying storage. Details are included in the [Delta paper](#).

## Public Documentation

This benchmark result was produced using the AWS S3 object store. The reliability features of the object stores are provided in the following documents:

- **General AWS S3 Documentation**  
[AWS S3 - 11 9's of durability](#)
- **Information on how data availability is achieved in AWS S3**  
[AWS S3](#)

## TPC-DS Data accessibility requirements

*The Data Accessibility Documentations must describe how data redundancy is accomplished within the SUT. Following are some examples of such description:*

- *Data Objects are stored on redundant devices (e.g. RAID 1, RAID 5)*
- *Data Objects are redundantly stored on multiple storage devices in the same facility.*
- *Data Objects are redundantly stored across multiple facilities.*
- *Data Objects are redundantly stored across data centers in multiple regions.*

*The following features must be supported by the SUT and described in the Data Accessibility Documentation:*

- *Synchronous writes: The redundant writes of multiple copies of Data Objects to multiple storage devices are executed synchronously.*
- *Automatic repair: Any loss of redundancy of a Data Object is automatically repaired without any operator intervention.*

Amazon S3 Standard, S3 Standard-IA, and S3 Glacier storage classes redundantly store your objects on multiple devices across a minimum of three Availability Zones (AZs) in an Amazon S3 Region before returning SUCCESS. The S3 One Zone-IA storage class stores data redundantly across multiple devices within a single AZ. These services are designed to sustain concurrent device failures by quickly detecting and repairing any lost redundancy, and they also regularly verify the integrity of your data using checksums.

**Documentation:** [Data redundancy within Amazon S3](#)

Amazon S3 Standard, S3 Standard-IA, S3 Intelligent-Tiering, S3 One Zone-IA, S3 Glacier, and S3 Glacier Deep Archive are all designed to provide 99.999999999% (11 9's) of data durability of objects over a given year. This durability level corresponds to an average annual expected loss of 0.000000001% of objects. For example, if you store 10,000,000 objects with Amazon S3, you can on average expect to incur a loss of a single object once every 10,000 years. S3 on Outposts is designed to durably and redundantly store data across multiple devices and servers on your Outpost. In addition, Amazon S3 Standard, S3 Standard-IA, S3 Glacier, and S3 Glacier Deep Archive are all designed to sustain data in the event of an entire S3 Availability Zone loss.

**Documentation:** [Data accessibility features supported by Amazon S3](#)

## Screen-capture

Aug 26, 2021

The screenshot shows the AWS S3 FAQs page. The navigation bar includes the AWS logo, a search bar, and links for Contact Us, Support, English, My Account, and Sign In. The main content area is titled 'Durability & Data Protection' and contains a 'Q: How durable is Amazon S3?' section. The page content is organized into a sidebar with categories like General, Security, Durability & Data Protection, Storage Classes, Storage Management, Storage Analytics & Insights, Query in Place, Replication, and Data processing. The main text explains that Amazon S3 Standard, S3 Standard-IA, S3 Intelligent-Tiering, S3 One Zone-IA, S3 Glacier, and S3 Glacier Deep Archive are all designed to provide 99.999999999% (11 9's) of data durability of objects over a given year. It also mentions that S3 on Outposts is designed to durably and redundantly store data across multiple devices and servers on your Outpost. A second question asks 'How are Amazon S3 and Amazon S3 Glacier designed to achieve 99.999999999% durability?' and the answer states that Amazon S3 Standard, S3 Standard-IA, and S3 Glacier storage classes redundantly store your objects on multiple devices across a minimum of three Availability Zones (AZs) in an Amazon S3 Region before returning SUCCESS.

## References

- <https://aws.amazon.com/s3/faqs/>
- <https://aws.amazon.com/s3/sla/>
- <https://docs.aws.amazon.com/AmazonS3/latest/userguide/disaster-recovery-resiliency.html>

# Clause 7: Performance Metrics and Execution Rules Related Items

## 7.1 System Activity

*Any system activity on the SUT that takes place between the conclusion of the load test and the beginning of the performance test must be fully disclosed including listings of scripts or command logs.*

The only activity between the end of the load test and the beginning of the performance test was the generation of the executable query text.

## 7.2 Test Steps

*The details of the steps followed to implement the performance test must be disclosed.*

Supporting Files Archive contains the scripts and logs.

## 7.3 Timing Intervals for Each Query and Refresh Function

*The timing intervals defined in Clause 7 must be disclosed.*

See the Executive Summary at the beginning of this report.

## 7.4 Throughput Test Result

*For each Throughput Test, the minimum, the 25th percentile, the median, the 75th percentile, and the maximum times for each query shall be reported.*

See the Executive Summary at the beginning of this report.

## 7.5 Time for Each Stream

*The start time and finish time for each query stream must be reported.*

See the Executive Summary at the beginning of this report.

## 7.6 Time for Each Refresh Function

The start time and finish time for each data maintenance function in the refresh run must be reported for the Throughput Tests

See the Executive Summary at the beginning of this report.

## 7.7 Performance Metrics

*The computed performance metric, related numerical quantities and the price/performance metric must be reported.*

QphDS@100,000GB = 32,941,245

See the Executive Summary at the beginning of this report for more detail.

# Clause 8: SUT and Driver Implementation Related Items

## 8.1 Driver

*A detailed textual description of how the driver performs its functions, how its various components interact and any product functionalities or environmental settings on which it relies must be provided. All related source code, scripts and configuration files must be disclosed. The information provided should be sufficient for an independent reconstruction of the driver.*

Databricks JDBC Driver [2.6.17](#) is used to connect to Databricks DBSQL endpoints, [SparkJDBC42.jar](#) is installed on the Client Driver. JDBC connections are created against the target 4X-Large endpoint, then the JDBC connections are used from the scala code in runQueryJdbc to run queries against the 4X-Large endpoint. Entry point for the benchmark is defined in the Scala function named fullRun.

The Scala code is executed from a Notebook, which is invoked from a browser.

The Supporting Files Archive contains all the commands, scripts, and logs.

## 8.2 Implementation Specific Layer (ISL)

*If an implementation specific layer is used, then a detailed description of how it performs its functions, how its various components interact and any product functionalities or environmental setting on which it relies must be provided. All related source code, scripts and configuration files must be disclosed. The information provided should be sufficient for an independent reconstruction of the implementation specific layer.*

No Implementation Specific Layer was used.

## 8.3 Profile-Directed Optimization

*If profile-directed optimization as described in Clause 7.2.10 is used, such use must be disclosed. In particular, the procedure and any scripts used to perform the optimization must be disclosed.*

Profile-directed optimization was not used.

# Clause 9: Pricing Related Items

## 9.1 Hardware and Software Used

*A detailed list of hardware and software used in the priced system must be reported. The rules for pricing are included in the current revision of the TPC Pricing Specification located on the TPC website (<http://www.tpc.org>)*

A detailed list of all licensed services, hardware and software, is provided in the Executive Summary of this report.

## 9.2 Availability Date

*The System Availability Date (see Clause 7.6.5) must be the single availability date reported on the first page of the executive summary. The full disclosure report must report Availability Dates individually for at least each of the categories for which a pricing subtotal must be. All Availability Dates required to be reported must be disclosed to a precision of 1 day, but the precise format is left to the test sponsor.*

The total system availability date is available as of publication date.

## 9.3 Country-Specific Pricing

*Additional Clause 7 related items may be included in the full disclosure report for each country specific priced configuration.*

The configuration is priced for the US market.

## 9.4 Storage access pricing

Storage access is calculated based on running the reported workload continuously for 3 years. Pricing includes 50% more access than reported by the metrics.

Storage cost/month

Unit conversions

- S3 Standard storage: 26.6 TB per month x 1024 GB in a TB = 27238.4 GB per month

Pricing calculations

- Tiered price for: 27238.4 GB
- 27238.4 GB x 0.0230000000 USD = 626.48 USD
- Total tier cost = 626.4832 USD (S3 Standard storage cost)

- 224,243,438 PUT requests for S3 Storage x 0.000005 USD per request = 1,121.2172 USD (S3 Standard PUT requests cost)
- 6,222,372,868 GET requests in a month x 0.0000004 USD per request = 2,488.9491 USD (S3 Standard GET requests cost)
- 626.4832 USD + 2,488.9491 USD + 1,121.2172 USD = 4,236.65 USD (Total S3 Standard Storage, data requests, S3 select cost)
- **S3 Standard cost (monthly): 4,236.65 USD**

## Clause 11: Audit Related Items

### Auditor's Information and Attestation Letter

This benchmark was audited by: Francois Raab, of InfoSizing.



Benchmark sponsor: Mostafa Mokhtar  
Databricks Inc.  
Sr. Staff Engineer  
160 Spear St 13th floor  
San Francisco, CA 94105

September 2, 2021

I verified the TPC Benchmark™ DS (TPC-DS™ v3.2.0) performance of the following configuration:

Platform: Databricks SQL 8.3 4X-Large  
on 256+1 AWS Instances  
Operating System: Linux  
Database Manager: Databricks Delta Engine 8.3

The results were:

**Performance Metric** 32,941,245QphDS@100000GB  
Database Load Time 2h 12m 9s

**Servers** **Amazon EC2 Instances**

**Workers: 256 x i3.2xlarge, each with:**

|         |            |             |             |
|---------|------------|-------------|-------------|
| CPU     | 8 x vCPU   |             |             |
| Memory  | 61 GiB     |             |             |
| Storage | <b>Qty</b> | <b>Size</b> | <b>Type</b> |
|         | 1          | 1.9 TB      | NVMe SSD    |

**1 x i3.16xlarge, with:**

|         |            |             |             |
|---------|------------|-------------|-------------|
| CPU     | 64 x vCPU  |             |             |
| Memory  | 488 GiB    |             |             |
| Storage | <b>Qty</b> | <b>Size</b> | <b>Type</b> |
|         | 8          | 1.9 TB      | NVMe SSD    |

**Storage** **Amazon S3**

**S3 Standard Storage, with:**

|               |         |
|---------------|---------|
| Used Capacity | 26.6 TB |
|---------------|---------|

In my opinion, these performance results were produced in compliance with the TPC requirements for the benchmark.

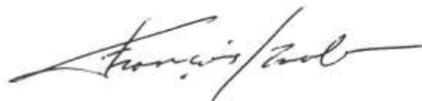
The following verification items were given special attention:

- The database records were defined with the proper layout and size
- The database population was generated using Dsdgen
- The database was properly scaled to 100,000 GB and populated accordingly
- The database load time was correctly measured and reported
- The query templates were produced using approved minor query modifications and query variants
- The query input variables were generated by Dsqgen
- The execution of the queries against the qualification database produced compliant output
- The tests were driven and sequenced according to the requirements
- The throughput tests involved 4 query streams
- The execution times for queries and data maintenance functions were correctly measured and reported
- The data accessibility requirements were satisfied through documentation
- The system pricing was verified for major components and maintenance
- The major pages from the FDR were verified for accuracy

Additional Audit Notes:

Due to a TPC-internal error during the production of version 3.2.0 of the TPC-DS kit, the benchmark execution had to use version 2.13 of the kit. It was confirmed by the TPC that the only changes between these two versions of the kit is the version number set in the tools/release.h parameter file.

Respectfully Yours,



François Raab, TPC Certified Auditor

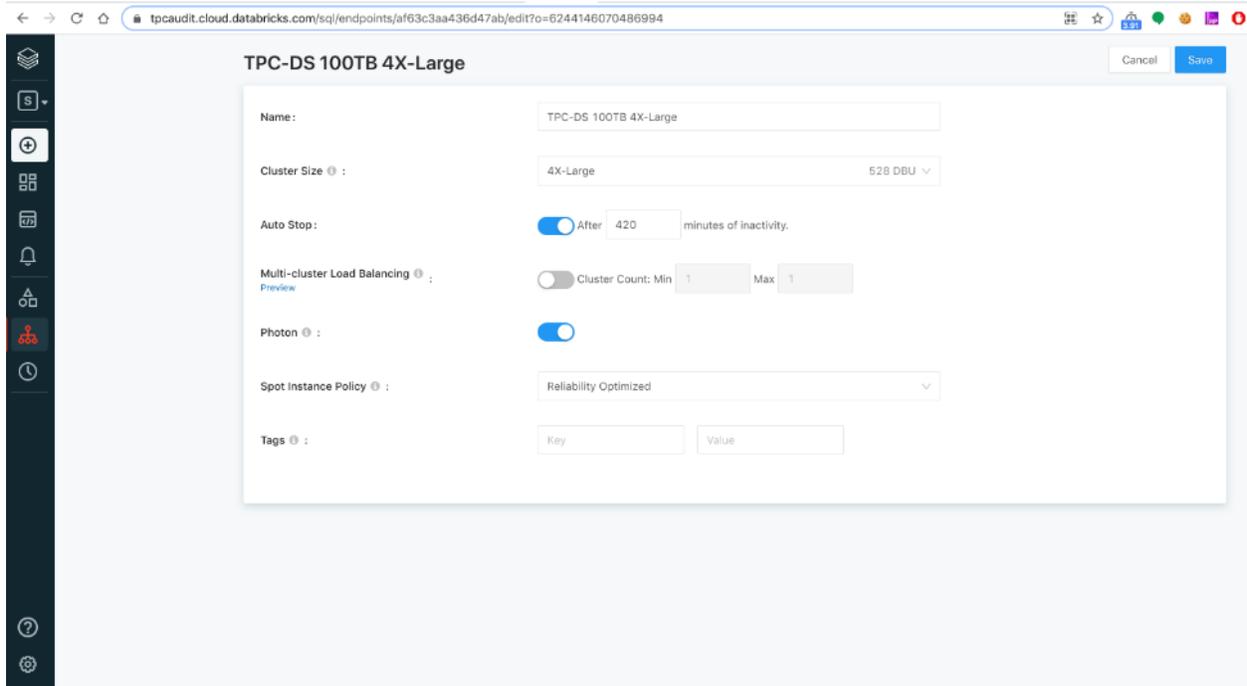
20 KREG LANE • MANITOU SPRINGS, CO 80829 • 719-473-7555 • WWW.SIZING.COM

# Supporting Files Index

| Clause   | Description  | Archive File Pathname  |
|----------|--|--|
| Clause 3 | Database create and load scripts and log files                               | SupportingFiles/Clause_3/create_database.sql<br>SupportingFiles/Clause_3/build_logs/<br>SupportingFiles/Clause_3/create_external/<br>SupportingFiles/Clause_3/create_table/<br>SupportingFiles/Clause_3/insert_into/ |
|          | Query generation logs  | SupportingFiles/Clause_3/querygen.sh<br>SupportingFiles/Clause_3/logs/   |
|          | Database validation scripts and log files                                    | SupportingFiles/Clause_3/validate_data.sh<br>SupportingFiles/Clause_3/validation/<br>SupportingFiles/Clause_3/verification_logs/   |
| Clause 4 | The script to generate queries for the qualification test and execution logs | SupportingFiles/Clause_4/generate_qualification_query.sh<br>SupportingFiles/Clause_4/qualification_run.log<br>SupportingFiles/Clause_4/logs/   |
|          | Output from qualification query execution                                    | SupportingFiles/Clause_4/output/   |
|          | Query templates and MQM documentation  | SupportingFiles/Clause_4/patches/  |
|          | Qualification queries SQL  | SupportingFiles/Clause_4/queries/  |
|          | Creation script for qualification database                                   | SupportingFiles/Clause_4/sqls/   |
| Clause 5 | Scripts for data maintenance execution                                       | SupportingFiles/Clause_5/mtsqls_[s]/   |

|          |  |  |
|----------|--|--|
|          | and verification for each stream [s]                                   |  |
|          | Logs from execution of data maintenance functions                      | SupportingFiles/Clause_5/logs/   |
|          | Verification of data maintenance functions                             | SupportingFiles/Clause_5/logs/mt_verify/   |
| Clause 6 | Data accessibility documentation                                       | SupportingFiles/Clause_6/DA_Documentation.docx   |
| Clause 7 | Full run execution script and log                                      | SupportingFiles/Clause_7/full_run.scala<br>SupportingFiles/Clause_7/full_run.log   |
|          | Query streams execution logs   | SupportingFiles/Clause_7/logs/   |
|          | Query text in stream [s]   | SupportingFiles/Clause_7/stream_[s]_queries/   |
|          | Output of queries in Power test and in each stream of Throughput tests | SupportingFiles/Clause_7/output/Power_0/<br>SupportingFiles/Clause_7/output/tt_1_1_1/<br>SupportingFiles/Clause_7/output/tt_1_2_2/<br>SupportingFiles/Clause_7/output/tt_1_3_3/<br>SupportingFiles/Clause_7/output/tt_1_4_4/<br>SupportingFiles/Clause_7/output/tt_2_1_5/<br>SupportingFiles/Clause_7/output/tt_2_2_6/<br>SupportingFiles/Clause_7/output/tt_2_3_7/<br>SupportingFiles/Clause_7/output/tt_2_4_8/ |
|          | Full SQL query set for each stream                                     | SupportingFiles/Clause_7/stream_[s]/sqls/  |

# Appendix A: Provisioning Compute Services



# Appendix B: Third Party Price Quotes

- Driver Amazon EC2 pricing

Services (4)

| Amazon EC2  | Region: US East (N. Virginia) | Edit     | Action        |
|---|-------------------------------|----------|---------------|
| <b>Quick estimate</b>   |                               |          |               |
| Operating system (Linux), Quantity (1), Pricing strategy (EC2 Instance Savings Plans 3 Year All Upfront), Storage amount (0 GB), Instance type (i3en.6xlarge) |                               | Monthly: | 0.00 USD      |
|   |                               | Upfront: | 26,989.56 USD |

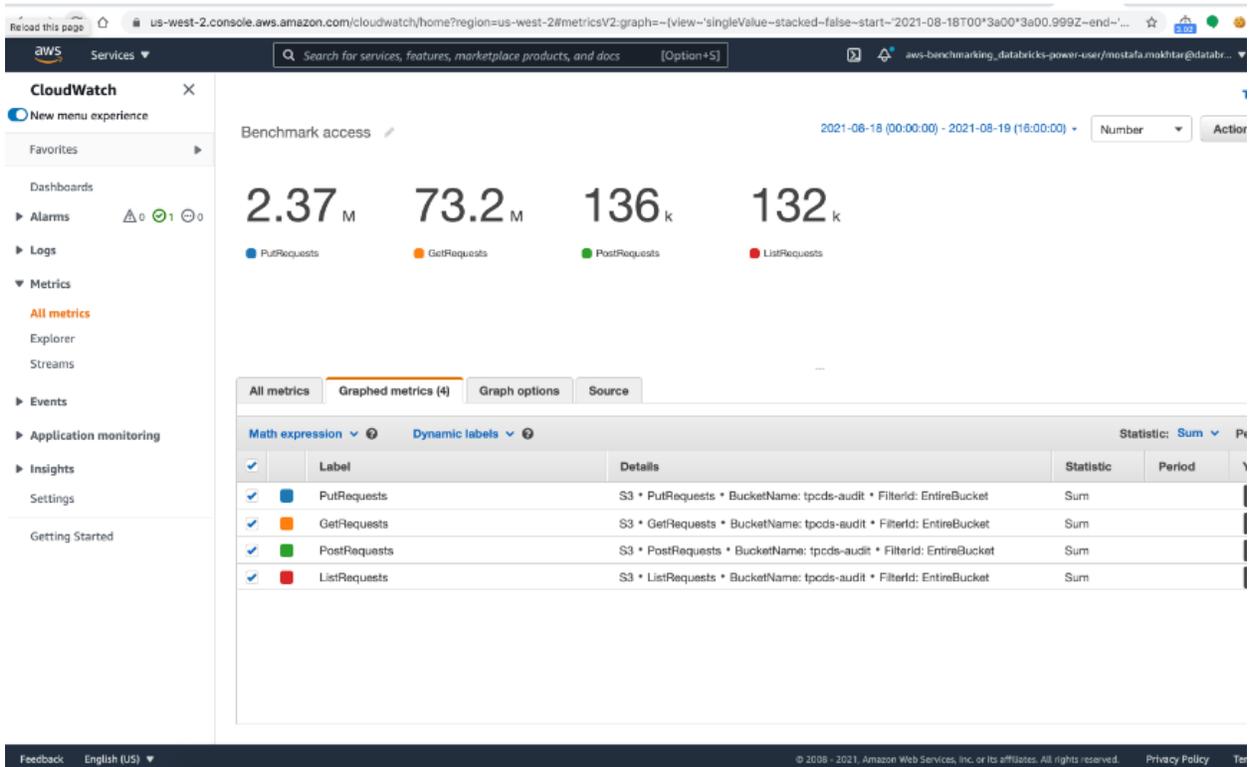
- Workers Amazon EC2 pricing

|   |  |          |                  |
|---|--|----------|------------------|
| <b>Quick estimate</b>   |  |          |                  |
| Operating system (Linux), Quantity (256), Pricing strategy (EC2 Instance Savings Plans 3 Year All UpFront), Storage amount (0 GB), Instance type (i3.2xlarge) |  | Monthly: | 0.00 USD         |
|   |  | Upfront: | 1,776,107.52 USD |

- Amazon EC2 Client Hardware pricing

|   |  |          |               |
|---|--|----------|---------------|
| <b>Quick estimate</b>   |  |          |               |
| Operating system (Linux), Quantity (1), Pricing strategy (EC2 Instance Savings Plans 3 Year All Upfront), Storage amount (0 GB), Instance type (i3en.6xlarge) |  | Monthly: | 0.00 USD      |
|   |  | Upfront: | 26,989.56 USD |

- Amazon S3 data access



- Amazon S3 data access cost calculation

### Amazon S3 data access

S3 Standard storage  
 TB per month

PUT, COPY, POST, LIST requests to S3 Standard

GET, SELECT, and all other requests from S3 Standard

Data returned by S3 Select  
 GB per month

Data scanned by S3 Select  
 GB per month

▼ Show calculations

Unit conversions  
 S3 Standard storage: 26.6 TB per month x 1024 GB in a TB = 27238.4 GB per month

Pricing calculations  
 Tiered price for: 27238.4 GB  
 27238.4 GB x 0.0230000000 USD = 626.48 USD  
 Total tier cost = 626.4832 USD (S3 Standard storage cost)  
 224,243,438 PUT requests for S3 Storage x 0.000005 USD per request = 1,121.2172 USD (S3 Standard PUT requests cost)  
 6,222,372,868 GET requests in a month x 0.0000004 USD per request = 2,488.9491 USD (S3 Standard GET requests cost)  
 626.4832 USD + 2,488.9491 USD + 1,121.2172 USD = 4,236.65 USD (Total S3 Standard Storage, data requests, S3 select cost)  
**S3 Standard cost (monthly): 4,236.65 USD**

- Pricing for laptop

Computers   Laptops   Desktops   Monitors   Tablets   Computer Accessories   PC Component

Subscribe & Save **Up to 15% off auto-deliveries** shop now

Back to results



**Lenovo Chromebook S330 Laptop, 14-Inch FHD (1920 x 1080) Display, MediaTek MT8173C Processor, 4GB LPDDR3, 64GB eMMC, Chrome OS, 81JW0000US, Business Black**

Visit the [Lenovo Store](#)  
★★★★☆ 2,209 ratings | 280 answered questions

Amazon's Choice for "b07glv1vc7"

List Price: ~~\$279.99~~ Details  
Price: **\$204.53** ✓ Prime FREE One-Day & FREE Returns  
You Save: \$75.46 (27%)

Pay **\$34.09/month for 6 months**, interest-free upon approval for the [Amazon Prime Rewards Visa Card](#)

May be available at a lower price from [other sellers](#), potentially without free Prime shipping.

|                           |              |
|---------------------------|--------------|
| Series                    | Ideapad S330 |
| Brand                     | Lenovo       |
| Specific Uses For Product | Personal     |
| Screen Size               | 14 Inches    |
| Operating System          | Chrome OS    |