

# **Tencent Cloud TCHouse-D**

## **Performance Testing**

### **Product Documentation**



## Copyright Notice

©2013-2024 Tencent Cloud. All rights reserved.

Copyright in this document is exclusively owned by Tencent Cloud. You must not reproduce, modify, copy or distribute in any way, in whole or in part, the contents of this document without Tencent Cloud's the prior written consent.

## Trademark Notice



All trademarks associated with Tencent Cloud and its services are owned by Tencent Cloud Computing (Beijing) Company Limited and its affiliated companies. Trademarks of third parties referred to in this document are owned by their respective proprietors.

## Service Statement

This document is intended to provide users with general information about Tencent Cloud's products and services only and does not form part of Tencent Cloud's terms and conditions. Tencent Cloud's products or services are subject to change. Specific products and services and the standards applicable to them are exclusively provided for in Tencent Cloud's applicable terms and conditions.

# Contents

## Performance Testing

### TPC-H Performance Testing

Test Scheme Introduction

Test Results for Reference

### SSB Performance Testing

Test Scheme Introduction

Test Results for Reference

# Performance Testing

## TPC-H Performance Testing

### Test Scheme Introduction

Last updated : 2024-07-31 09:18:34

This document will introduce how to use TPC-H (Business Intelligence Computing Test) to perform performance testing on Tencent Cloud TCHouse-D. Taking the TPC-H query performance of a 16-core cluster under a 100 GB data set as an example, a reference test scheme is given.

## About TPC-H Performance Test

TPC-H is a decision support benchmark that consists of a set of business-oriented ad hoc queries and concurrent data modifications. The data it queries and populates in the database is extensively industry-related. This benchmark test demonstrates the ability of a decision support system to examine large amounts of data, perform highly complex queries, and answer critical business questions. The performance metric reported by TPC-H is called TPC-H Composite Query-per-Hour Performance Metric (QphH@Size), which reflects the system's ability to process multiple queries.

### Note:

TPC-H simulates a data warehouse of a sales system. The benchmark test includes 22 queries in total, and the main evaluation metric is the response time of each query, that is, the time required from submitting the query to returning the result. TPC-H test results can comprehensively reflect the system's ability to process queries.

## Test Scheme Introduction

### Test Environment Preparation

#### Hardware Environment

In the reference scheme given in this document, the tested cluster includes 1 FE and 3 BEs. The FE/BE node processes are deployed separately. The specific specifications are as follows. It should be noted that in actual testing, such a large amount of hardware resources will not be consumed.

Node Type	Node Specifications
1 FE, standard	CPU:4 cores Memory: 16 GB Hard disk: Enhanced SSD Cloud Disk 200 GB

3 BEs, standard	CPU:16 cores Memory: 64 GB Hard disk: Enhanced SSD Cloud Disk 1000 GB
-----------------	---

### Software Version

Tencent Cloud TChouse-D 1.2.7

### Test Script Preparation

Download the TPC-H toolkit from [Toolkit Address](#) and compile it.

### TPC-H 100 G Data Test

**Generate a 100 G data set.**



```
sh gen-tpch-data.sh -s 100 -c 10
```

The data generated is shown in the following table:

TPC-H table name	Number of rows	Remarks
REGION	5	Region Table
NATION	25	Country Table
SUPPLIER	1 million	Supplier Table

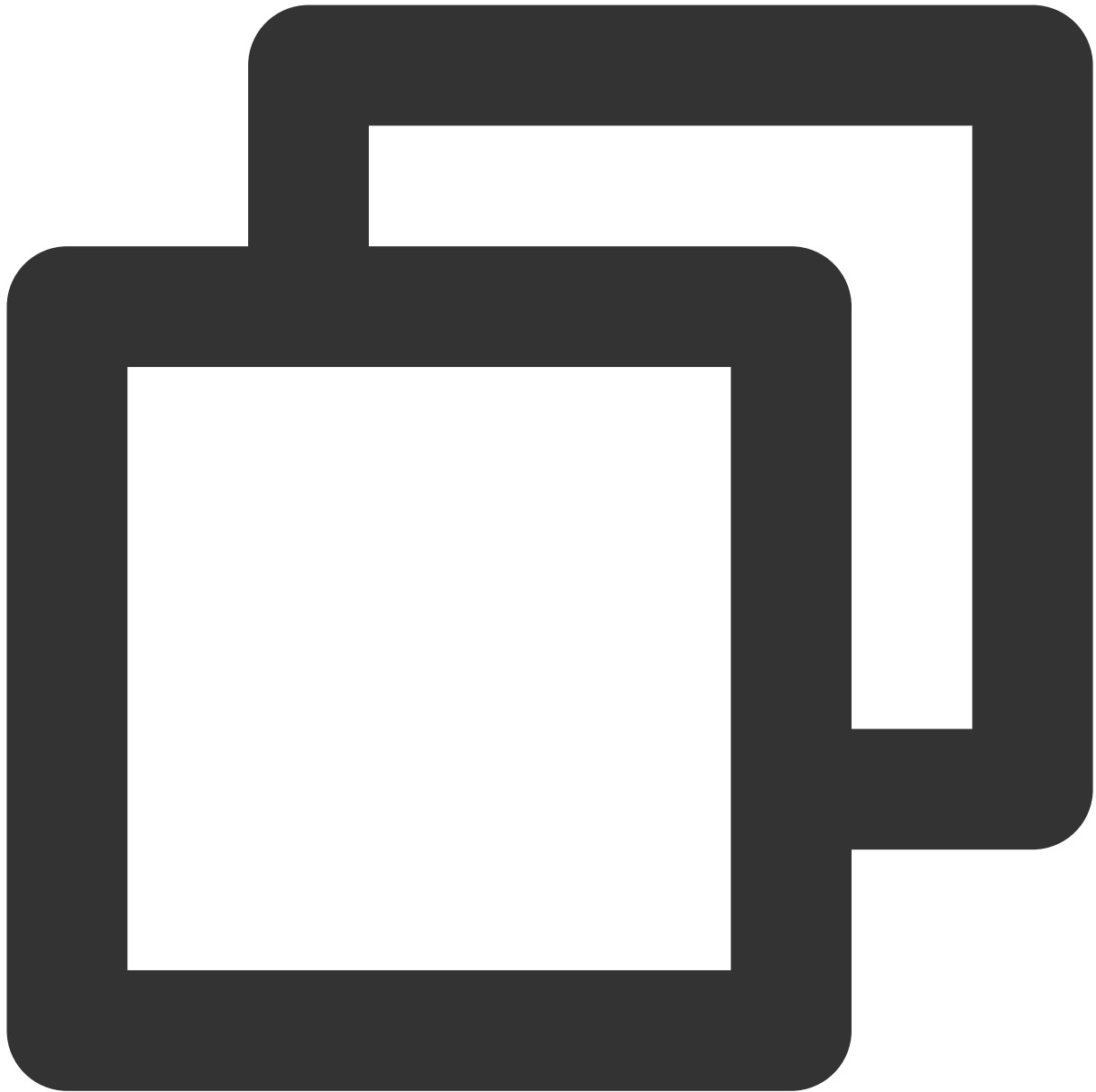
---

PART	20 million	Parts List
PARTSUPP	80 million	Parts Supply List
CUSTOMER	15 million	Customer Table
ORDERS	150 million	Order Table
LINEITEM	600 million	Order Details Table

**Create a table**

Modify the doris-cluster.conf configuration file.

Modify configuration: FE\_HOST, PASSWORD, DB.



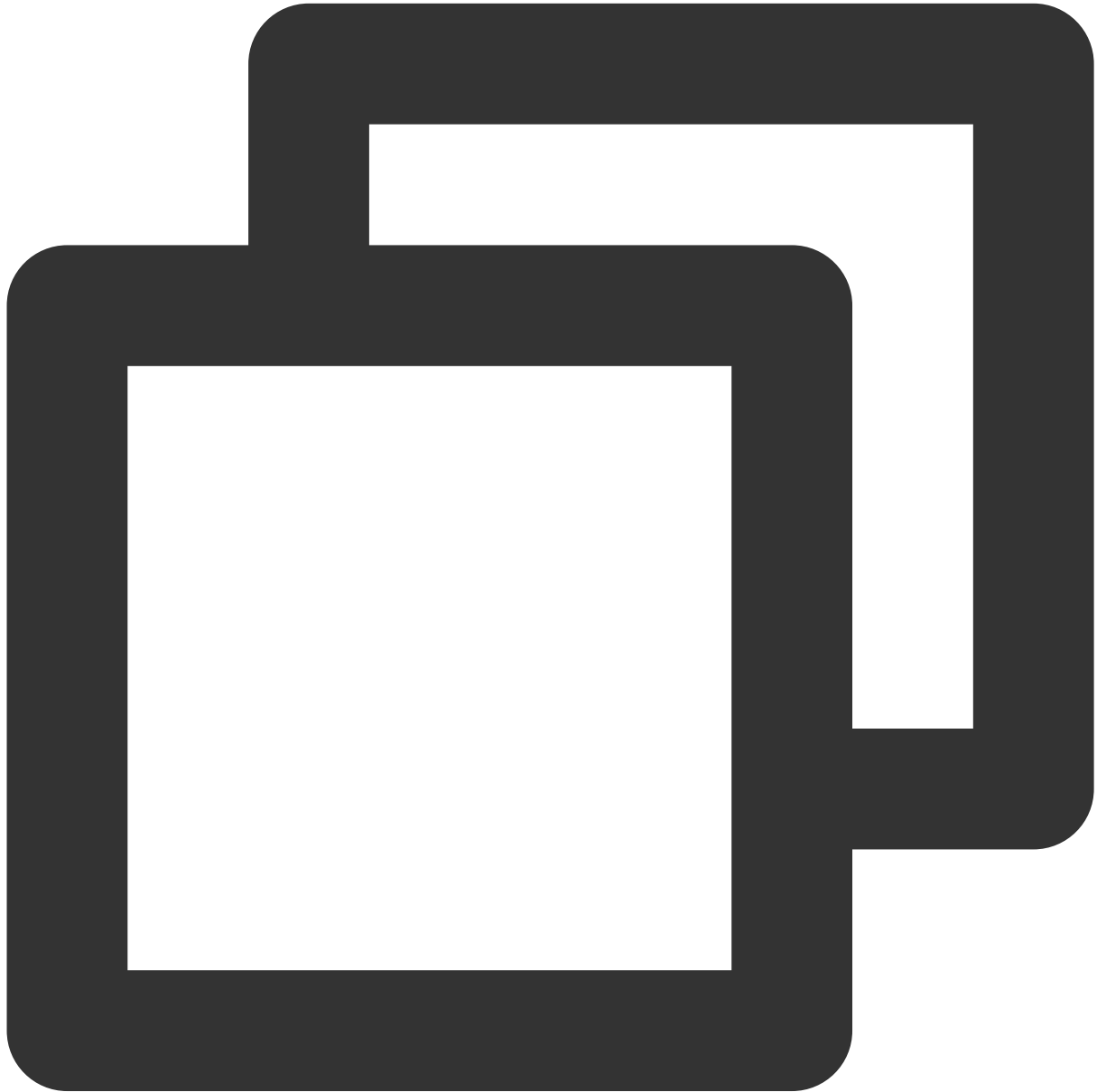
```
# cat doris-cluster.conf

# Any of FE host
export FE_HOST='127.0.0.1'
# http_port in fe.conf
export FE_HTTP_PORT=8030
# query_port in fe.conf
export FE_QUERY_PORT=9030
# Doris username
export USER='root'
# Doris password
```



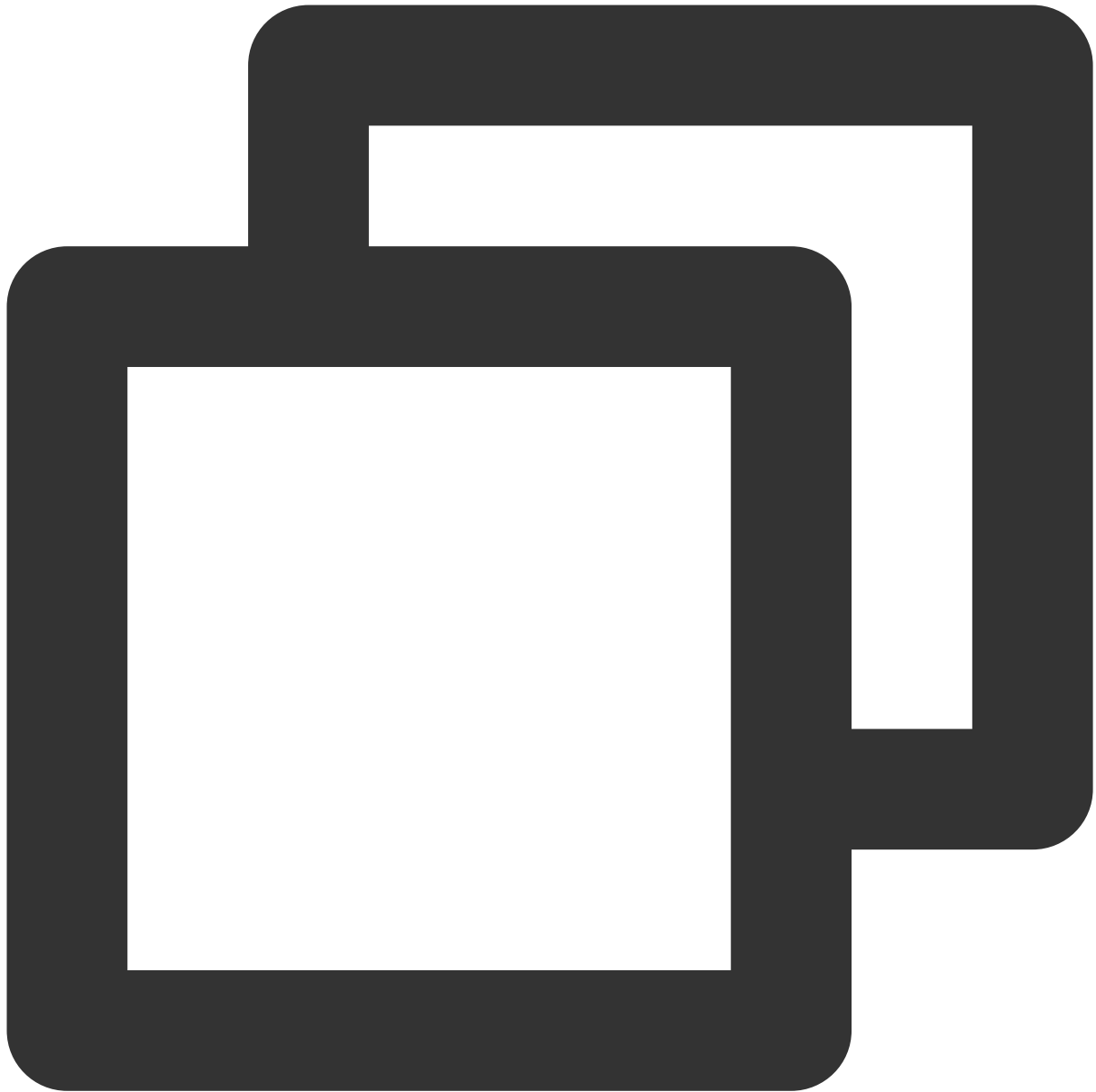
```
export PASSWORD=''
# The database where TPC-H tables located
export DB='tpch_100g_decimalv3'
# The scale of testing data
export SCALE='100g' # only support '100g' or '1t'
```

Create a Table:



```
sh create-tpch-tables.sh
```

**Import Data**



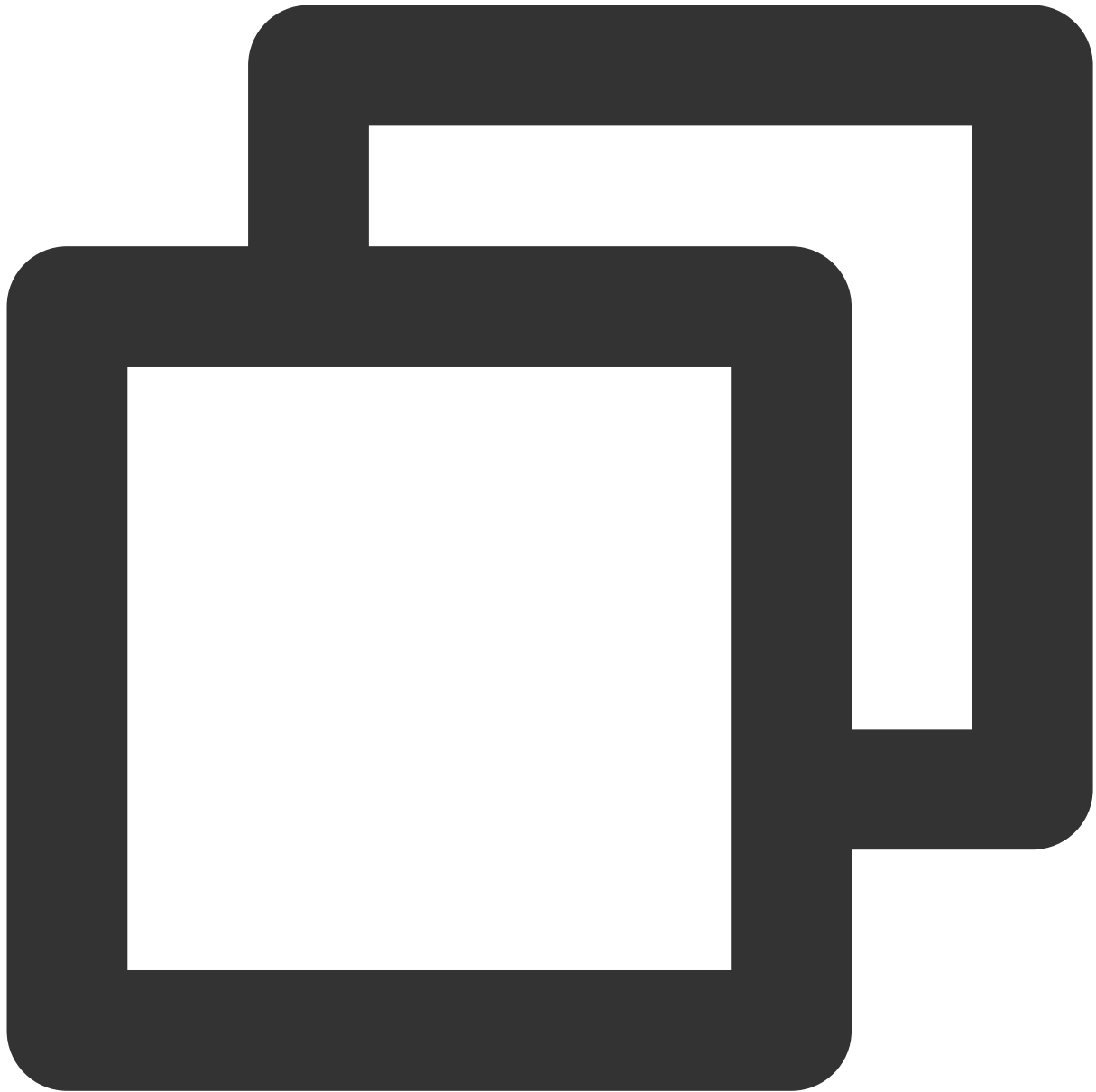
```
sh load-tpch-data.sh
```

```
MySQL [tpch100g]> show data;
```

```
+-----+-----+-----+
| TableName | Size          | ReplicaCount |
+-----+-----+-----+
| customer  | 1.317 GB     | 24           |
| lineitem  | 20.880 GB    | 96           |
| nation    | 2.571 KB     | 1            |
| orders    | 6.302 GB     | 96           |
```

```
| part      | 752.470 MB | 24      |
| partsupp | 4.375 GB   | 24      |
| region   | 1.090 KB   | 1       |
| supplier | 85.528 MB  | 12      |
| Total    | 33.693 GB  | 278     |
| Quota    | 1024.000 TB | 1073741824 |
| Left     | 1023.967 TB | 1073741546 |
+-----+-----+-----+
11 rows in set (0.00 sec)
```

## Query



```
[root@9 tpch-tools]# sh bin/run-tpch-queries.sh
q1: 2103
q2: 305
q3: 792
q4: 516
q5: 1036
q6: 60
q7: 493
q8: 954
q9: 4411
q10: 870
```

```
q11: 183
q12: 1847
q13: 2886
q14: 165
q15: 255
q16: 398
q17: 520
q18: 1665
q19: 468
q20: 347
q21: 1741
q22: 412
total time: 22427 ms
```

Thus, TCP-H data generation, table creation, import, and query under the 100 GB data set scene is completed.

# Test Results for Reference

Last updated : 2024-07-31 09:18:47

This document will provide a reference for the results of the performance test of Tencent Cloud TCHouse-D using TPC-H (Business Intelligence Computing Test).

## About TPC-H Performance Test

TPC-H is a decision support benchmark that consists of a set of business-oriented ad hoc queries and concurrent data modifications. The data it queries and populates in the database is extensively industry-related. This benchmark test demonstrates the ability of a decision support system to examine large amounts of data, perform highly complex queries, and answer critical business questions. The performance metric reported by TPC-H is called TPC-H Composite Query-per-Hour Performance Metric (QphH@Size), which reflects the system's ability to process multiple queries.

## Test Environment

### Hardware Environment

In this document, two model clusters are tested, each containing 1 FE and 3 BEs. The FE/BE node processes are deployed separately. It should be noted that the principle for selecting a cluster of models is to be close to common user configurations, and in actual testing, such a large amount of hardware resources will not be consumed.

Cluster Specifications	Node Type	Specification
Specification 1 (small and medium-sized data scenes)	1 FE	CPU:4 cores Memory: 16 GB Hard disk: Enhanced SSD Cloud Disk 200 GB
	3 BEs	CPU:16 cores Memory: 64 GB Hard disk: Enhanced SSD Cloud Disk 1500 GB
Specification 2 (large-scale data scenes)	1 FE	CPU:16 cores Memory: 64 GB Hard disk: Enhanced SSD Cloud Disk 200 GB
	3 BEs	CPU:64 cores Memory: 256 GB Hard disk: Enhanced SSD Cloud Disk 1500 GB

## Software Version

Tencent Cloud TCHouse-D 1.2.7

## Test Results for Reference

### Test Data

The test is conducted using two data sets, Scale 100 and Scale 1000. The descriptions and data volume of the created tables are as follows:

TPC-H table name	Number of rows - Scale 100	Number of rows - Scale 1000	Remarks
REGION	5	5	Region Table
NATION	25	25	Country Table
SUPPLIER	1 million	1 million	Supplier Table
PART	20 million	200 million	Parts List
PARTSUPP	80 million	800 million	Parts Supply List
CUSTOMER	15 million	150 million	Customer Table
ORDERS	150 million	1.5 billion	Order Table
LINEITEM	600 million	6 billion	Order Details Table

### Performance Test Results

#### Note:

The test result is the average time of three queries of the corresponding SQL file, in seconds (s).

Query Number	Specification 1-Scale 100 data set	Specification 2-Scale 100 data set	Specification 2-Scale 1000 data set
SQL-1	2.05	1.00	9.67
SQL-2	0.23	0.24	2.05
SQL-3	0.71	0.62	30.46
SQL-4	0.5	0.38	9.74
SQL-5	1.01	0.72	11.10

SQL-6	0.06	0.05	0.58
SQL-7	0.48	0.40	32.13
SQL-8	0.86	0.61	16.00
SQL-9	4.2	3.19	76.98
SQL-10	0.84	0.64	11.36
SQL-11	0.18	0.16	2.14
SQL-12	1.76	1.47	17.03
SQL-13	2.85	1.60	19.02
SQL-14	0.16	0.15	1.73
SQL-15	0.25	0.20	1.66
SQL-16	0.39	0.35	3.56
SQL-17	0.51	0.42	12.38
SQL-18	1.72	1.07	19.64
SQL-19	0.48	0.28	6.75
SQL-20	0.35	0.34	12.90
SQL-21	1.74	0.82	14.61
SQL-22	0.42	0.39	9.60
<b>Total Time</b>	<b>21.74</b>	<b>15.09</b>	<b>321.12</b>



# SSB Performance Testing

## Test Scheme Introduction

Last updated : 2024-07-31 09:19:12

This document introduces how to use the Star Schema data set to perform performance testing on Tencent Cloud TCHouse-D, and provides a reference scheme for data import and performance testing.

## About SSB Performance Test

SSB (Star Schema Benchmark) is a lightweight performance test set for data warehouse scenes. SSB provides a simplified star model data set based on TPC-H, which is mainly used to test the performance of multi-table join queries under the star model. In addition, the industry practice is to flatten SSB into a wide table model (hereinafter referred to as: SSB FLAT) to test the performance of the query engine.

## Test Scheme Introduction

### Test Environment Preparation

#### Hardware Environment

In the reference scheme given in this document, the tested cluster includes 1 FE and 3 BEs. The FE/BE node processes are deployed separately. The specific specifications are as follows. It should be noted that in actual testing, such a large amount of hardware resources will not be consumed.

Node Type	Node Specifications
1 FE, standard	CPU:4 cores Memory: 16 GB Hard disk: Enhanced SSD Cloud Disk 200 GB
3 BEs, standard	CPU:16 cores Memory: 64 GB Hard disk: Enhanced SSD Cloud Disk 1000 GB

#### Software Version

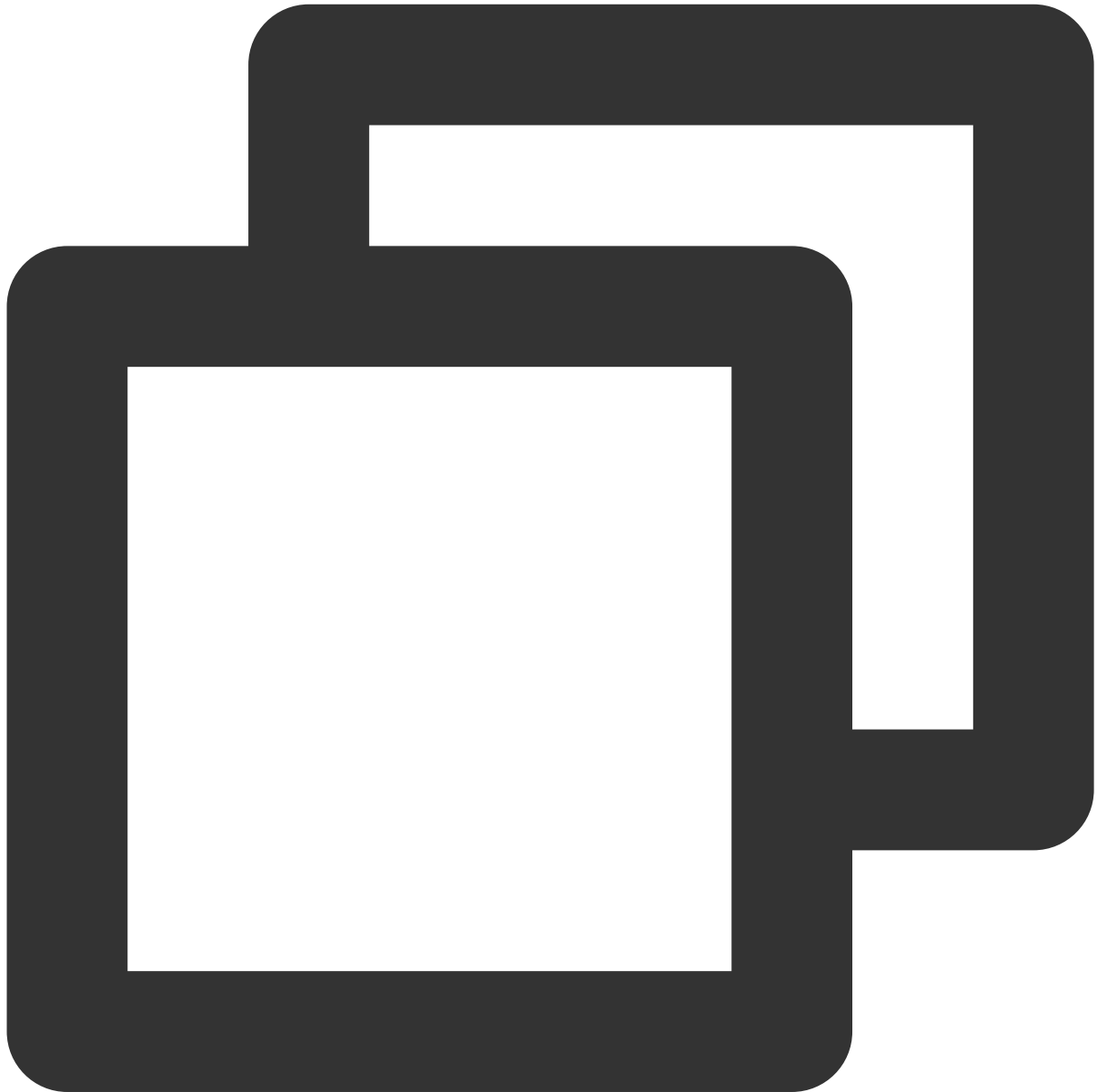
Tencent Cloud TChouse-D 1.2.7

### Test Script Preparation

Download the SSB-H toolkit from [Toolkit Address](#) and compile it.

### Test SSB data set.

### Generate data set.



```
sh bin/gen-ssb-data.sh -s 100 -c 100
```

```
277M    /data/ssb-tools/bin/ssb-data//customer.tbl  
228K    /data/ssb-tools/bin/ssb-data//date.tbl  
5.9G    /data/ssb-tools/bin/ssb-data//lineorder.tbl.1
```

```

7.7G    /data/ssb-tools/bin/ssb-data//lineorder.tbl.10
7.6G    /data/ssb-tools/bin/ssb-data//lineorder.tbl.2
7.2G    /data/ssb-tools/bin/ssb-data//lineorder.tbl.3
6.0G    /data/ssb-tools/bin/ssb-data//lineorder.tbl.4
6.0G    /data/ssb-tools/bin/ssb-data//lineorder.tbl.5
6.0G    /data/ssb-tools/bin/ssb-data//lineorder.tbl.6
6.0G    /data/ssb-tools/bin/ssb-data//lineorder.tbl.7
7.4G    /data/ssb-tools/bin/ssb-data//lineorder.tbl.8
7.5G    /data/ssb-tools/bin/ssb-data//lineorder.tbl.9
116M    /data/ssb-tools/bin/ssb-data//part.tbl
17M     /data/ssb-tools/bin/ssb-data//supplier.tbl

```

Under the `-s 100` parameter, the size of the data set generated is:

SSB Table Name	Number of rows	Remarks
LINEORDER	600,037,902	Product Order Details Table
CUSTOMER	3,000,000	Customer Information Table
PART	1,400,000	Parts Information Table
SUPPLIER	200,000	Supplier Information Table
DATE	2,556	Date Table

### Create a table

Modify the `doris-cluster.conf` configuration file `vim conf/doris-cluster.conf`.

Modify configuration: `FE_HOST`, `PASSWORD`, `DB`.

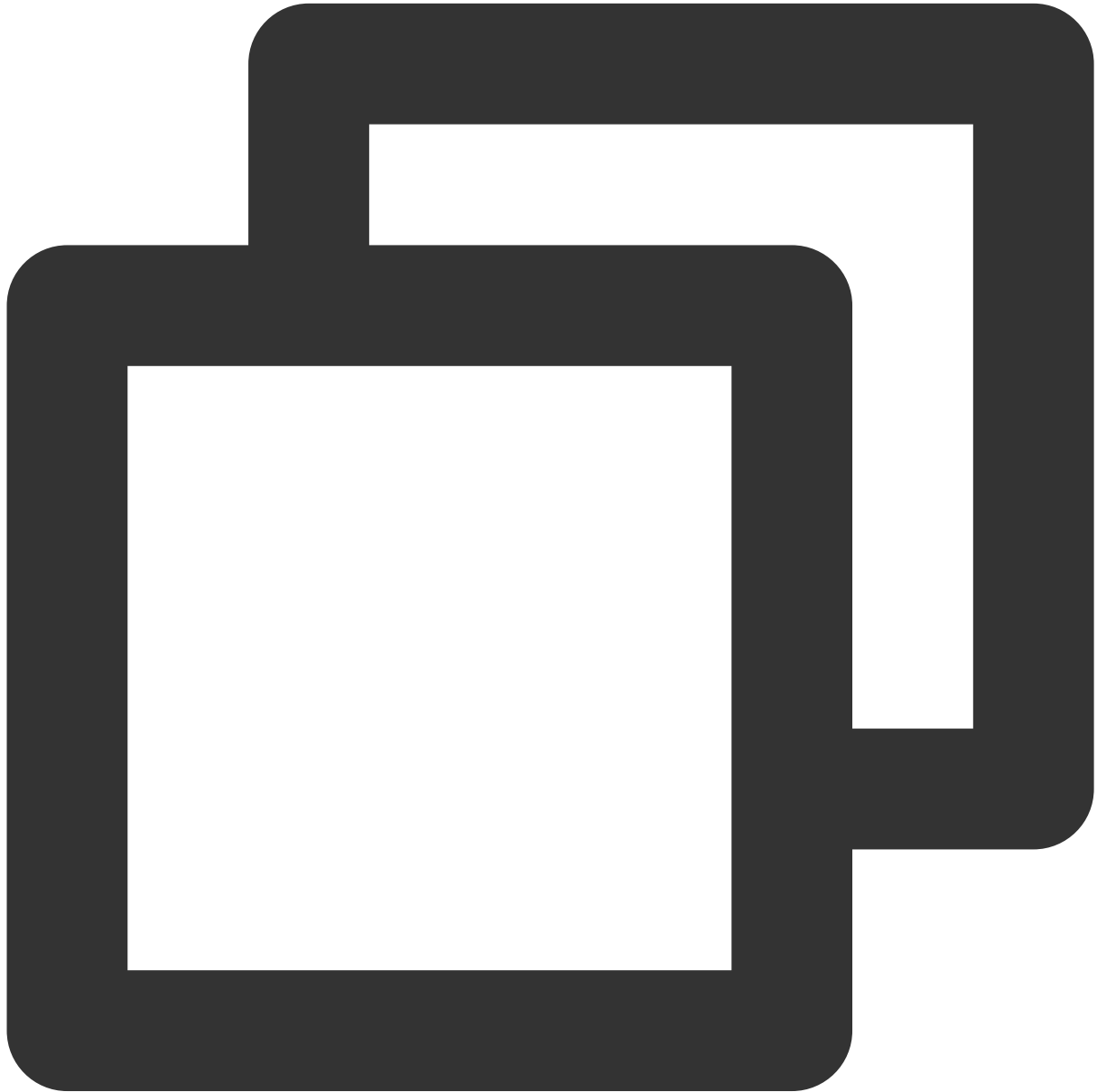


```
# cat doris-cluster.conf

# Any of FE host
export FE_HOST='127.0.0.1'
# http_port in fe.conf
export FE_HTTP_PORT=8030
# query_port in fe.conf
export FE_QUERY_PORT=9030
# Doris username
export USER='root'
# Doris password
```

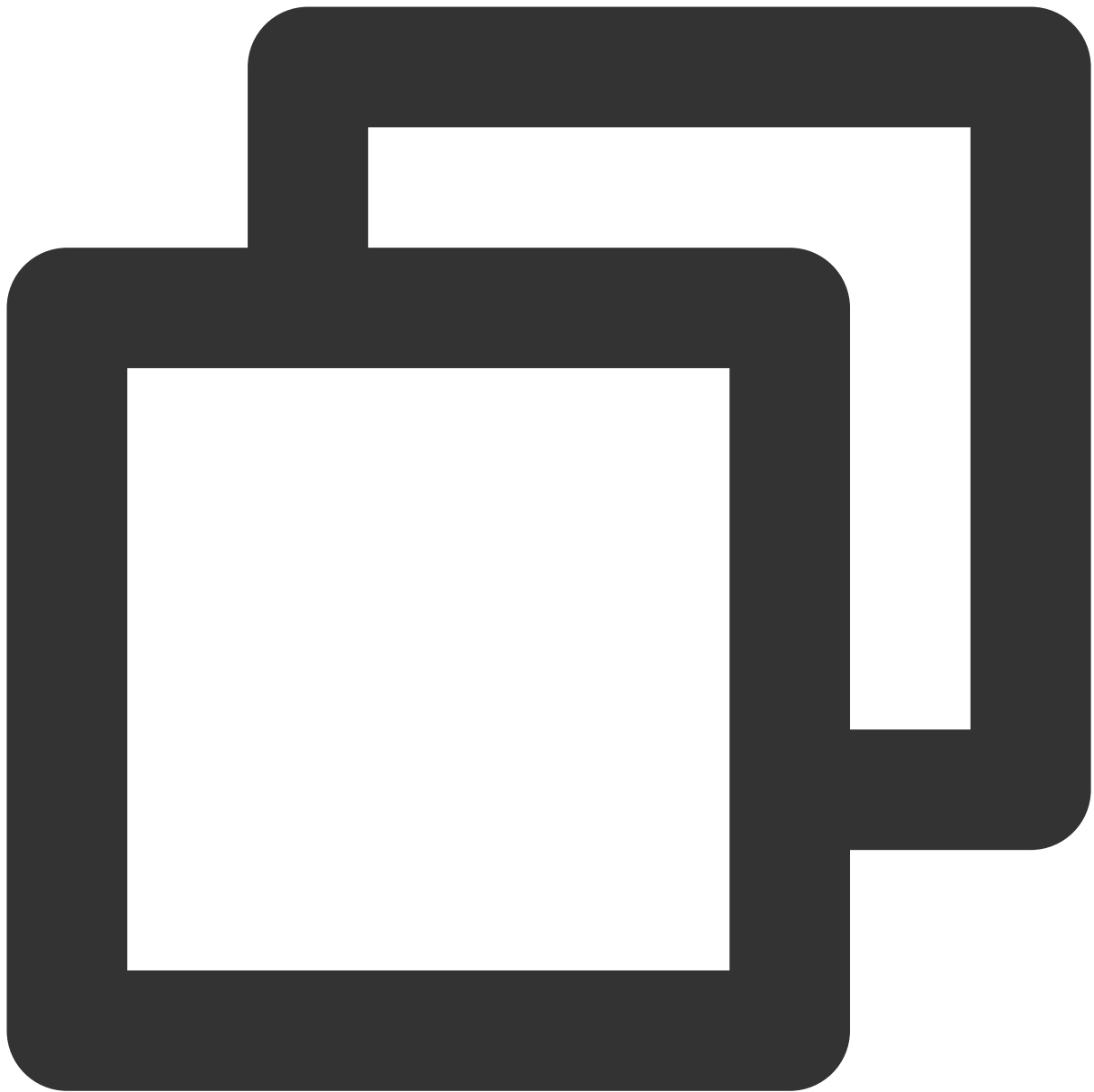
```
export PASSWORD=''
# The database where SSB tables located
export DB='ssb_100g'
# The scale of testing data
export SCALE='100g' # only support '100g' or '1t'
```

Create a Table:



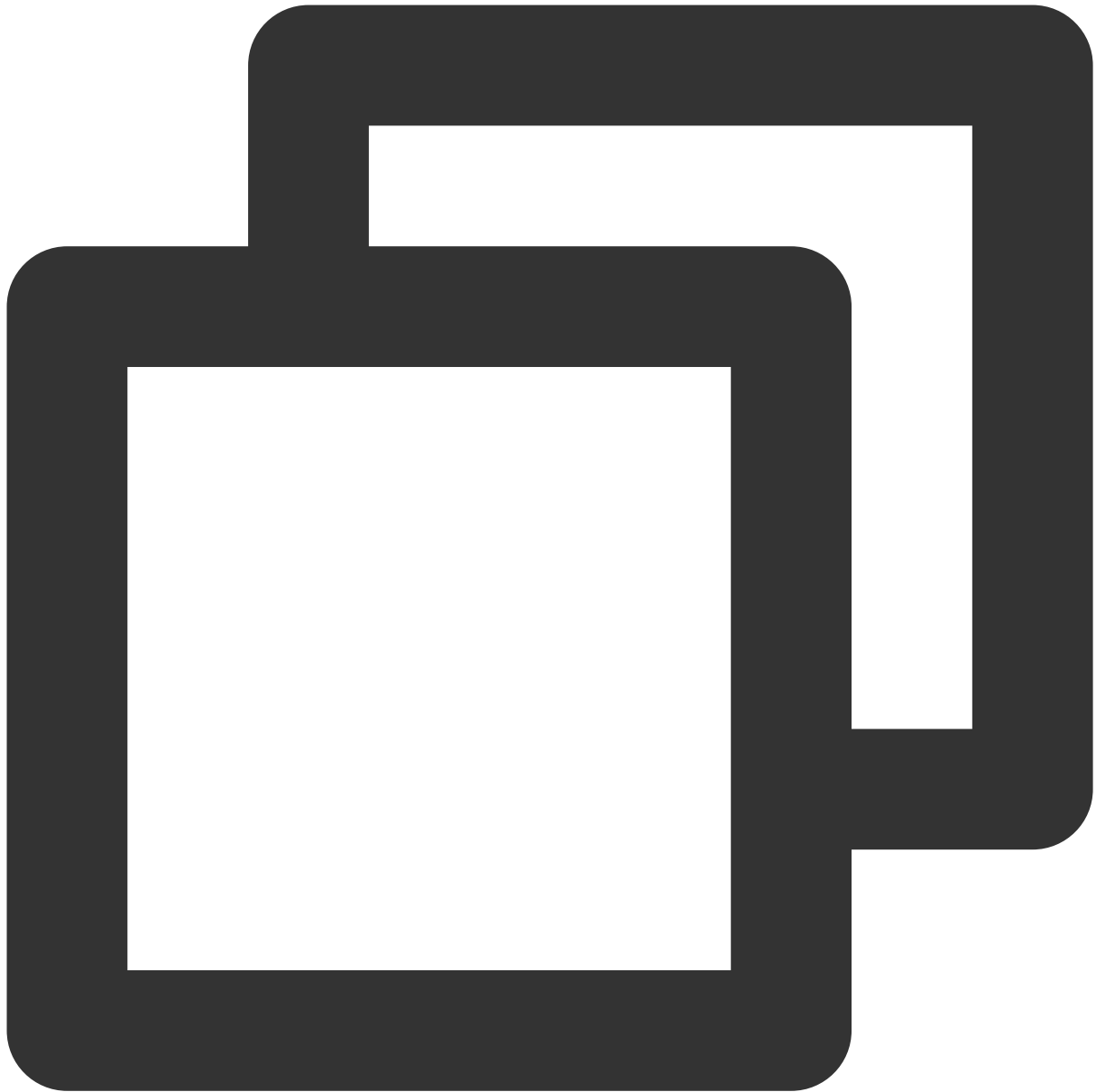
```
sh bin/create-ssb-tables.sh
```

**Import Data**



```
sh bin/load-ssb-data.sh
```

### Check Imported Data

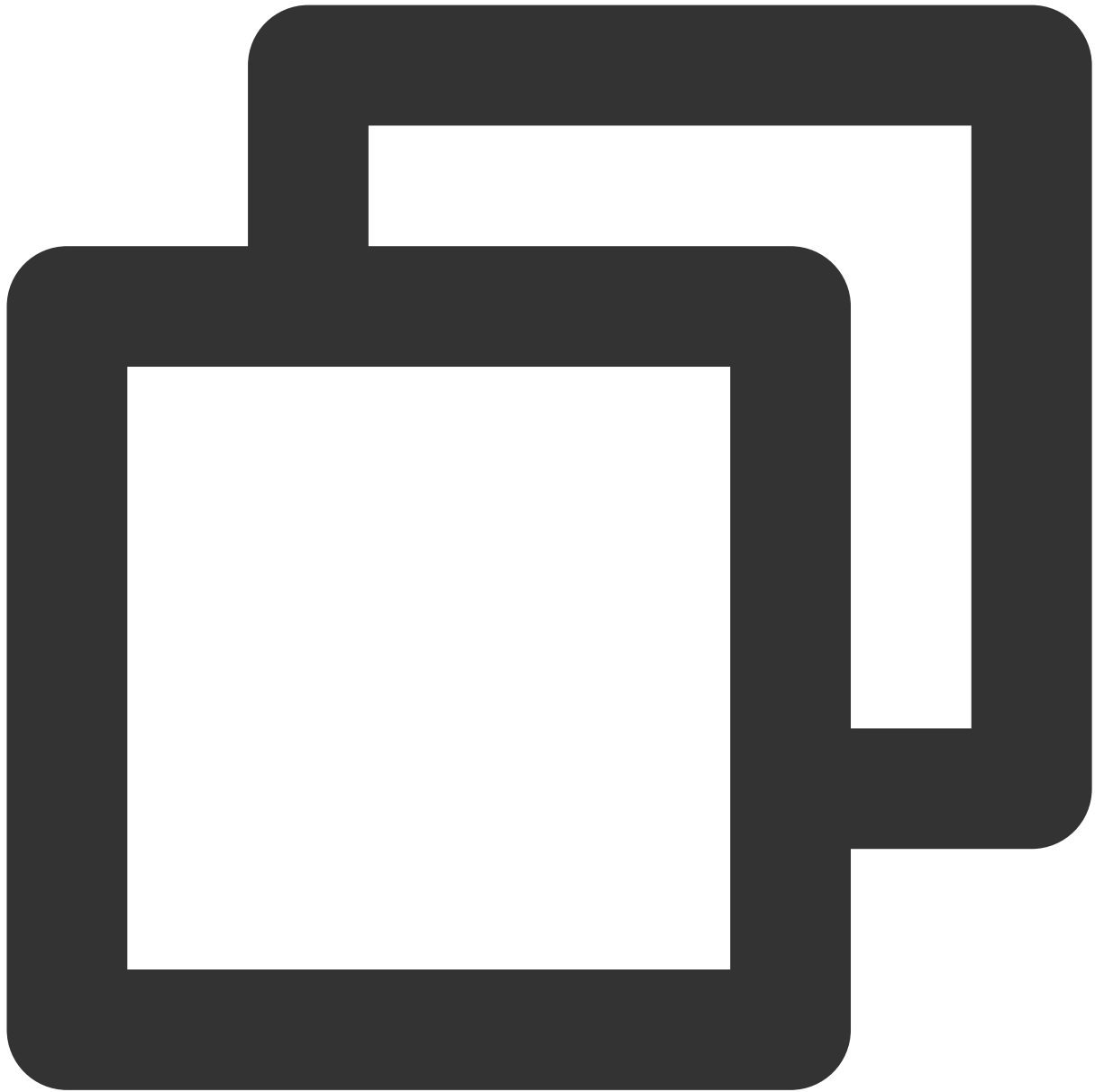


```
select count(*) from part;  
select count(*) from customer;  
select count(*) from supplier;  
select count(*) from date;  
select count(*) from lineorder;  
select count(*) from lineorder_flat;
```

The amount of data shall be consistent with the number of rows of generated data.

## Query

Query SSB tables.

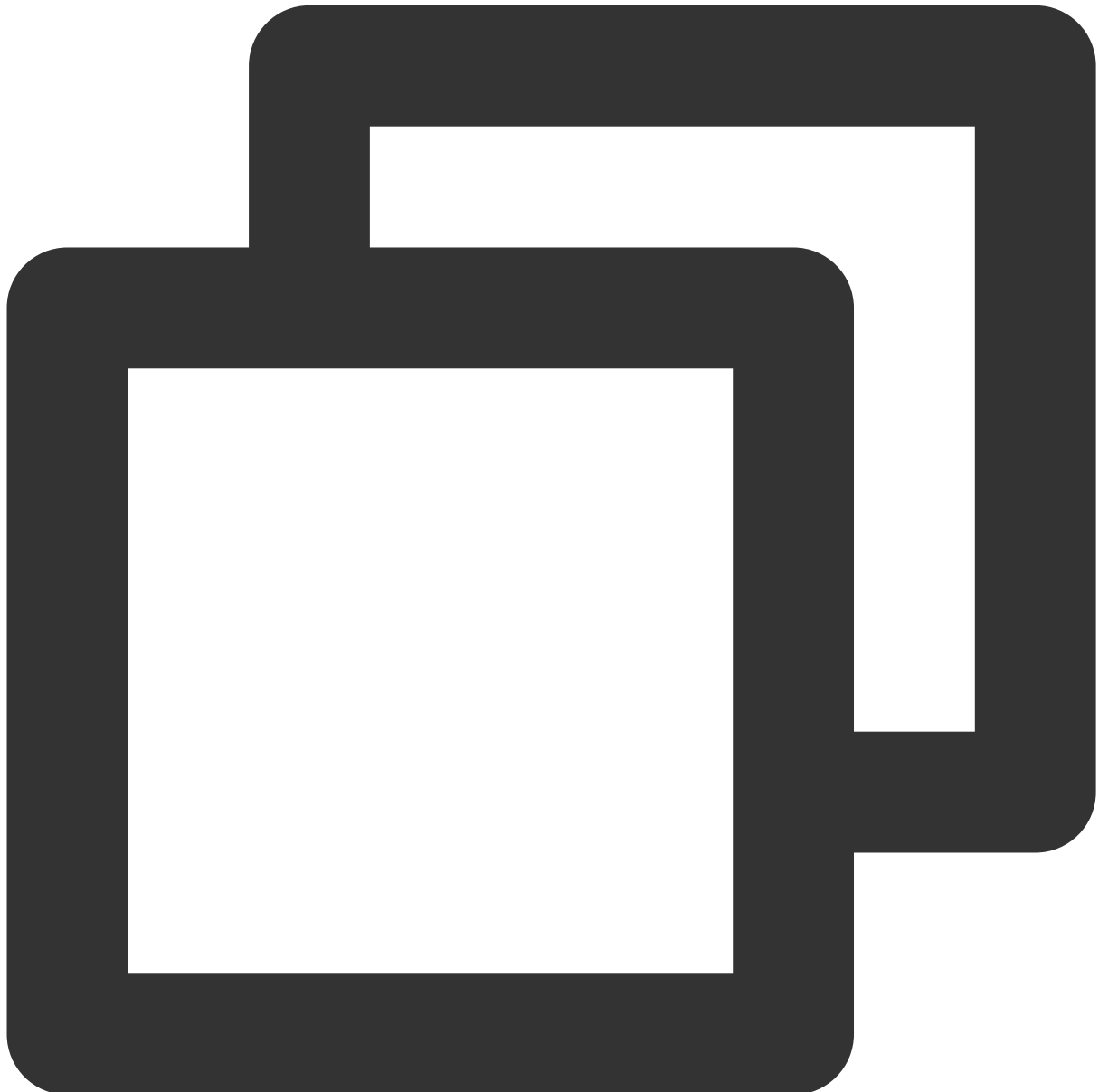


```
# sh bin/run-ssb-queries.sh
q1.1: 46
q1.2: 29
q1.3: 26
q2.1: 340
q2.2: 273
q2.3: 257
q3.1: 542
q3.2: 237
```



```
q3.3: 297
q3.4: 57
q4.1: 732
q4.2: 372
q4.3: 483
total time: 3691 ms
```

### Query FLAT tables.



```
# sh bin/run-ssb-flat-queries.sh
q1.1: 26
```

```
q1.2: 10
q1.3: 35
q2.1: 85
q2.2: 83
q2.3: 60
q3.1: 157
q3.2: 78
q3.3: 75
q3.4: 13
q4.1: 131
q4.2: 49
q4.3: 33
total time: 835 ms
```

Thus, SSB data generation, table creation, import, and query are completed.

# Test Results for Reference

Last updated : 2024-07-31 09:19:42

This document will provide a reference for the results of performance testing of Tencent Cloud TCHouse-D using SSB.

## About SSB Performance Test

SSB (Star Schema Benchmark) is a lightweight performance test set for data warehouse scenes. SSB provides a simplified star model data set based on TPC-H, which is mainly used to test the performance of multi-table join queries under the star model. In addition, the industry practice is to flatten SSB into a wide table model (hereinafter referred to as: SSB FLAT) to test the performance of the query engine.

## Test Environment

### Hardware Environment

In this document, two model clusters are tested, each containing 1 FE and 3 BEs. The FE/BE node processes are deployed separately. It should be noted that the principle for selecting a cluster of models is to be close to common user configurations, and in actual testing, such a large amount of hardware resources will not be consumed.

Cluster Specifications	Node Type	Specification
Specification 1 (small and medium-sized data scenes)	1 FE	CPU:4 cores Memory: 16 GB Hard disk: Enhanced SSD Cloud Disk 200 GB
	3 BEs	CPU:16 cores Memory: 64 GB Hard disk: Enhanced SSD Cloud Disk 1500 GB
Specification 2 (large-scale data scenes)	1 FE	CPU:16 cores Memory: 64 GB Hard disk: Enhanced SSD Cloud Disk 200 GB
	3 BEs	CPU:64 cores Memory: 256 GB Hard disk: Enhanced SSD Cloud Disk 1500 GB

### Software Version

Tencent Cloud TCHouse-D 1.2.7

## Test Results for Reference

### Test Data

The test is conducted using two data sets, Scale 100 and Scale 1000. The descriptions and data volume of the created tables are as follows:

SSB Table Name	Number of rows - Scale 100	Number of rows - Scale 1000	Remarks
LINEORDER	600,037,902	5,999,989,709	Product Order Details Table
CUSTOMER	3,000,000	30,000,000	Customer Information Table
PART	1,400,000	2,000,000	Parts Information Table
SUPPLIER	200,000	2,000,000	Supplier Information Table
DATE	2,556	2,556	Date Table

### Performance Test Results

#### Note:

The test result is the average time of three queries of the corresponding SQL file, in seconds (s).

### SSB Test Results

Query Number	Specification 1- Scale 100 data set	Specification 2- Scale 100 data set	Specification 1- Scale 1000 data set	Specification 2-Scale 1000 data set
Q1.1	0.05	0.04	0.27	0.26
Q1.2	0.03	0.03	0.14	0.13
Q1.3	0.03	0.03	0.13	0.13
Q2.1	0.33	0.15	5.32	1.29
Q2.2	0.27	0.12	4.89	1.19
Q2.3	0.26	0.12	4.67	1.14
Q3.1	0.53	0.39	9.52	7.27
Q3.2	0.23	0.12	4.33	1.06

Q3.3	0.30	0.12	3.58	0.71
Q3.4	0.06	0.05	0.34	0.24
Q4.1	0.73	0.41	10.84	5.59
Q4.2	0.38	0.34	4.15	2.56
Q4.3	0.47	0.54	4.37	4.22
<b>Total time</b>	<b>3.66</b>	<b>2.43</b>	<b>52.54</b>	<b>25.79</b>

### SSB FLAT Test Results

Query Number	Specification 1- Scale 100 data set	Specification 2- Scale 100 data set	Specification 1- Scale 1000 data set	Specification 2-Scale 1000 data set
Q1.1	0.03	0.02	0.13	0.10
Q1.2	0.01	0.01	0.02	0.02
Q1.3	0.03	0.04	0.23	0.22
Q2.1	0.08	0.05	3.61	1.04
Q2.2	0.08	0.05	2.86	0.37
Q2.3	0.06	0.04	2.54	0.28
Q3.1	0.15	0.09	4.30	1.50
Q3.2	0.08	0.05	3.45	1.09
Q3.3	0.07	0.04	1.93	0.26
Q3.4	0.01	0.02	0.03	0.02
Q4.1	0.13	0.08	4.96	2.22
Q4.2	0.05	0.04	1.59	0.23
Q4.3	0.03	0.03	1.23	0.14
<b>Total time</b>	<b>0.82</b>	<b>0.54</b>	<b>26.87</b>	<b>7.49</b>

