

# 腾讯云数据仓库 TCHouse-D

## 性能测试



腾讯云

## 【 版权声明 】

©2013–2024 腾讯云版权所有

本文档（含所有文字、数据、图片等内容）完整的著作权归腾讯云计算（北京）有限责任公司单独所有，未经腾讯云事先明确书面许可，任何主体不得以任何形式复制、修改、使用、抄袭、传播本文档全部或部分内容。前述行为构成对腾讯云著作权的侵犯，腾讯云将依法采取措施追究法律责任。

## 【 商标声明 】



及其它腾讯云服务相关的商标均为腾讯云计算（北京）有限责任公司及其关联公司所有。本文档涉及的第三方主体的商标，依法由权利人所有。未经腾讯云及有关权利人书面许可，任何主体不得以任何方式对前述商标进行使用、复制、修改、传播、抄录等行为，否则将构成对腾讯云及有关权利人商标权的侵犯，腾讯云将依法采取措施追究法律责任。

## 【 服务声明 】

本文档意在向您介绍腾讯云全部或部分产品、服务的当时的相关概况，部分产品、服务的内容可能不时有所调整。您所购买的腾讯云产品、服务的种类、服务标准等应由您与腾讯云之间的商业合同约定，除非双方另有约定，否则，腾讯云对本文档内容不做任何明示或默示的承诺或保证。

## 【 联系我们 】

我们致力于为您提供个性化的售前购买咨询服务，及相应的技术售后服务，任何问题请联系 4009100100或 95716。

---

## 文档目录

### 性能测试

#### TPC-H 性能测试

##### 测试方案介绍

#### SSB 性能测试

##### 测试方案介绍

# 性能测试

## TPC-H 性能测试 测试方案介绍

最近更新时间：2023-09-28 15:27:31

本文将为您介绍如何使用 TPC-H（商业智能计算测试）对腾讯云 TCHouse-D 进行性能测试，以测试16核集群在100 GB数据集下的 TPC-H 查询性能为例，给出参考测试方案。

### 关于 TPC-H 性能测试

TPC-H 是一个决策支持基准（Decision Support Benchmark），它由一套面向业务的临时查询和并发数据修改组成，其查询和填充数据库的数据具有广泛的行业相关性。该基准测试展示了决策支持系统在检查大量数据、执行高度复杂的查询以及解答关键业务问题的能力。TPC-H 报告的性能指标称为 TPC-H Composite Query-per-Hour Performance Metric (QphH@Size)，能够反映系统处理查询的多种能力。

#### 说明：

TPC-H 模拟了一套销售系统的数据仓库。其共包含8张表，数据量可设定从1 GB~3 TB不等。其基准测试共包含了22个查询，主要评价指标为各查询的响应时间，即从提交查询到结果返回所需的时间。TPC-H 测试结果可以综合反映系统处理查询时的能力。

### 测试方案介绍

#### 测试环境准备

##### 硬件环境

在本文给出的参考方案中，所测试的集群包含 FE 1台、BE 3台，FE/BE 节点进程均为单独部署，具体规格如下所示。需要特别说明的是，实际测试中不会消耗如此多的硬件资源。

节点类型	节点规格
FE 1台 标准型	CPU: 4 cores 内存: 16 GB 磁盘: 增强型SSD云硬盘200 GB
BE 3台 标准型	CPU: 16 cores 内存: 64 GB 磁盘: 增强型SSD云硬盘1000 GB

##### 软件版本

腾讯云 TChouse-D 1.2.7

## 测试脚本准备

从[工具包地址](#)下载 TPC-H 工具包并编译

## TPC-H 100G数据测试

### 生成100G数据集

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

所生成的数据如下表所示：

TPC-H 表名	行数	备注
REGION	5	区域表
NATION	25	国家表
SUPPLIER	100万	供应商表
PART	2000万	零部件表
PARTSUPP	8000万	零部件供应表
CUSTOMER	1500万	客户表
ORDERS	1.5亿	订单表
LINEITEM	6亿	订单明细表

## 建表

修改 doris-cluster.conf 配置文件。

修改配置：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'
```

建表:

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

导入数据

```
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)
```

查询

```
[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
```

由此，即完成了针对100 GB数据集场景下的 TCP-H 数据生成、建表、导入和查询。

# SSB 性能测试

## 测试方案介绍

最近更新时间：2023-10-31 15:45:21

本文为您介绍如何使用 Star Schema 数据集对腾讯云 TCHouse-D 进行性能测试，给出数据导入及性能测试的参考方案。

### 关于 SSB 性能测试

SSB (Star Schema Benchmark) 是一个轻量级的数仓场景下的性能测试集。SSB 基于TPC-H 提供了一个简化版的星型模型数据集，主要用于测试在星型模型下，多表关联查询的性能表现。另外，业界内通常也会将 SSB 打平为宽表模型（以下简称：SSB FLAT），来测试查询引擎的性能。

### 测试方案介绍

#### 测试环境准备

##### 硬件环境

在本文给出的参考方案中，所测试的集群包含 FE 1台、BE 3台，FE/BE 节点进程均为单独部署，具体规格如下所示。需要特别说明的是，实际测试中不会消耗如此多的硬件资源。

节点类型	节点规格
FE 1台 标准型	CPU: 4 cores 内存: 16 GB 磁盘: 增强型 SSD 云硬盘200 GB
BE 3台 标准型	CPU: 16 cores 内存: 64 GB 磁盘: 增强型 SSD 云硬盘1000 GB

##### 软件版本

腾讯云 TChouse-D 1.2.7

#### 测试脚本准备

从[工具包地址](#)下载 SSB 工具包并编译

#### 测试 SSB 数据集

##### 生成数据集



```
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
```

在 `-s 100` 参数下，生成的数据集大小为：

SSB 表名	行数	备注
LINEORDER	600,037,902	商品订单明细表
CUSTOMER	3,000,000	客户信息表
PART	1,400,000	零件信息表
SUPPLIER	200,000	供应商信息表
DATE	2,556	日期表

## 建表

修改 `doris-cluster.conf` 配置文件 `vim conf/doris-cluster.conf`。

修改配置：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'
```

建表:

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

## 导入数据

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

## 检查导入数据

```
selectcount(*)from part;  
selectcount(*)from customer;  
selectcount(*)from supplier;  
selectcount(*)fromdate;  
selectcount(*)from lineorder;  
selectcount(*)from lineorder_flat;
```

数据量应和生成数据的行数一致。

## 查询

### 查询 SSB 表

```
# 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
```

## 查询 FLAT 表

```
# 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
```

由此，即完成了 SSB 数据生成、建表、导入和查询。