目录

[1，测试环境 1](#_Toc485993629)

[2，测试场景 1](#_Toc485993630)

[3，精确查询 2](#_Toc485993631)

[4，tpch语句查询 3](#_Toc485993632)

[5，kudu/impala参数调整 10](#_Toc485993633)

# 1，测试环境

● 节点:

1台主节点，5台计算节点

● 机器配置:

32个物理核

128G内存

12\*4T磁盘

● 操作系统:

redhat 6.4

● 版本:

CDH 5.11.0

# 2，测试场景

主要关注两方面的测试：

1，首先满足精确查询，比如查询身份证号码，找到和这个人相关的信息

2，其次，TPCH数据集，是用于评测决策支持系统（或数据仓库）的标准SQL测试集。这个测试集包含对大数据集的统计／报表生成／联机查询／数据挖掘等复杂应用，测试用的数据和值是有倾斜的，与真实数据一致。可以说TPCH是与真实场景非常接近的一个测试集，也是难度较大的一个测试集。

本次测试生成了100G数据，函数如下：

|  |  |  |
| --- | --- | --- |
| 表名 | 行数 | 行数 |
| customer | 15000000 | 1千五百万 |
| lineitem | 600037902 | 6亿 |
| nation | 25 | -- |
| orders | 150000000 | 15亿 |
| part | 20000000 | 2千万 |
| partsupp | 80000000 | 8千万 |
| region | 5 | -- |
| supplier | 1000000 | 1百万 |

后续会扩展到1TB的测试

# 3，精确查询

Customer：

|  |  |  |  |
| --- | --- | --- | --- |
| Count | Sql | Time | 说明 |
| 1千五百万 | select count(\*) from customer; | 0.23 |  |
| select \* from customer where c\_name="Customer#013299353"; | 0.24s | 非主键，string |
| select \* from customer where c\_custkey=13299353; | 0.26s | 主键，int |

Partsupp：

|  |  |  |  |
| --- | --- | --- | --- |
| Count | Sql | Time | 说明 |
| 8千万 | select count(\*) from partsupp; | 0.43s |  |
| select \* from partsupp where ps\_comment="riously pending packages eat furiously deposits. final braids sleep quick" | 1.12s | 非主键，string |
| select \* from partsupp where ps\_partkey=8697341; | 0.22s | 主键，int |

lineitem：

|  |  |  |  |
| --- | --- | --- | --- |
| Count | Sql | Time | 说明 |
| 6亿 | select count(\*) from lineitem; | 1.24s |  |
| select \* from lineitem where l\_comment ="efully special accoun" | Fetched 577 row(s) in 8.16s | 非主键，string |
| select \* from lineitem where ps\_partkey=8697341; | 0.27s | 主键，int |

lineitem：

|  |  |  |  |
| --- | --- | --- | --- |
| Count | Sql | Time | 说明 |
| 15亿 | select count(\*) from orders; | 0.42s |  |
| select \* from orders where l\_comment =" equests are furiously. carefu" | 0.33s | 非主键，string |
| select \* from orders where o\_orderkey= 2946; | 0.31s | 主键，int |

# 4，tpch语句查询

1，**tpch\_query1.sql**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | select | |  | l\_returnflag, | |  | l\_linestatus, | |  | sum(l\_quantity) as sum\_qty, | |  | sum(l\_extendedprice) as sum\_base\_price, | |  | sum(l\_extendedprice \* (1 - l\_discount)) as sum\_disc\_price, | |  | sum(l\_extendedprice \* (1 - l\_discount) \* (1 + l\_tax)) as sum\_charge, | |  | avg(l\_quantity) as avg\_qty, | |  | avg(l\_extendedprice) as avg\_price, | |  | avg(l\_discount) as avg\_disc, | |  | count(\*) as count\_order | |  | from | |  | lineitem | |  | where | |  | l\_shipdate <= '1998-09-16' | |  | group by | |  | l\_returnflag, | |  | l\_linestatus | |  | order by | |  | l\_returnflag, | |  | l\_linestatus; | |

2，**tpch\_query5.sql**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | select | |  | n\_name, | |  | sum(l\_extendedprice \* (1 - l\_discount)) as revenue | |  | from | |  | customer, | |  | orders, | |  | lineitem, | |  | supplier, | |  | nation, | |  | region | |  | where | |  | c\_custkey = o\_custkey | |  | and l\_orderkey = o\_orderkey | |  | and l\_suppkey = s\_suppkey | |  | and c\_nationkey = s\_nationkey | |  | and s\_nationkey = n\_nationkey | |  | and n\_regionkey = r\_regionkey | |  | and r\_name = 'AFRICA' | |  | and o\_orderdate >= '1993-01-01' | |  | and o\_orderdate < '1994-01-01' | |  | group by | |  | n\_name | |  | order by | |  | revenue desc; | |

3，**tpch\_query10.sql**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | select | |  | c\_custkey, | |  | c\_name, | |  | sum(l\_extendedprice \* (1 - l\_discount)) as revenue, | |  | c\_acctbal, | |  | n\_name, | |  | c\_address, | |  | c\_phone, | |  | c\_comment | |  | from | |  | customer, | |  | orders, | |  | lineitem, | |  | nation | |  | where | |  | c\_custkey = o\_custkey | |  | and l\_orderkey = o\_orderkey | |  | and o\_orderdate >= '1993-07-01' | |  | and o\_orderdate < '1993-10-01' | |  | and l\_returnflag = 'R' | |  | and c\_nationkey = n\_nationkey | |  | group by | |  | c\_custkey, | |  | c\_name, | |  | c\_acctbal, | |  | c\_phone, | |  | n\_name, | |  | c\_address, | |  | c\_comment | |  | order by | |  | revenue desc | |  | limit 20; | |

4, **tpch\_query3.sql**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | select | |  | l\_orderkey, | |  | sum(l\_extendedprice \* (1 - l\_discount)) as revenue, | |  | o\_orderdate, | |  | o\_shippriority | |  | from | |  | customer, | |  | orders, | |  | lineitem | |  | where | |  | c\_mktsegment = 'BUILDING' | |  | and c\_custkey = o\_custkey | |  | and l\_orderkey = o\_orderkey | |  | and o\_orderdate < '1995-03-22' | |  | and l\_shipdate > '1995-03-22' | |  | group by | |  | l\_orderkey, | |  | o\_orderdate, | |  | o\_shippriority | |  | order by | |  | revenue desc, | |  | o\_orderdate | |  | limit 10; | |

5, **tpch\_query4.sql**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | select | |  | o\_orderpriority, | |  | count(\*) as order\_count | |  | from | |  | orders as o | |  | where | |  | o\_orderdate >= '1996-05-01' | |  | and o\_orderdate < '1996-08-01' | |  | and exists ( | |  | select | |  | \* | |  | from | |  | lineitem | |  | where | |  | l\_orderkey = o.o\_orderkey | |  | and l\_commitdate < l\_receiptdate | |  | ) | |  | group by | |  | o\_orderpriority | |  | order by | |  | o\_orderpriority; | |

6, **tpch\_query6.sql**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | select | |  | sum(l\_extendedprice \* l\_discount) as revenue | |  | from | |  | lineitem | |  | where | |  | l\_shipdate >= '1993-01-01' | |  | and l\_shipdate < '1994-01-01' | |  | and l\_discount between 0.06 - 0.01 and 0.06 + 0.01 | |  | and l\_quantity < 25; | |

7, **tpch\_query7.sql**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | select | |  | supp\_nation, | |  | cust\_nation, | |  | l\_year, | |  | sum(volume) as revenue | |  | from | |  | ( | |  | select | |  | n1.n\_name as supp\_nation, | |  | n2.n\_name as cust\_nation, | |  | year(l\_shipdate) as l\_year, | |  | l\_extendedprice \* (1 - l\_discount) as volume | |  | from | |  | supplier, | |  | lineitem, | |  | orders, | |  | customer, | |  | nation n1, | |  | nation n2 | |  | where | |  | s\_suppkey = l\_suppkey | |  | and o\_orderkey = l\_orderkey | |  | and c\_custkey = o\_custkey | |  | and s\_nationkey = n1.n\_nationkey | |  | and c\_nationkey = n2.n\_nationkey | |  | and ( | |  | (n1.n\_name = 'KENYA' and n2.n\_name = 'PERU') | |  | or (n1.n\_name = 'PERU' and n2.n\_name = 'KENYA') | |  | ) | |  | and l\_shipdate between '1995-01-01' and '1996-12-31' | |  | ) as shipping | |  | group by | |  | supp\_nation, | |  | cust\_nation, | |  | l\_year | |  | order by | |  | supp\_nation, | |  | cust\_nation, | |  | l\_year; | |

8，**tpch\_query8.sql**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | select | |  | o\_year, | |  | sum(case | |  | when nation = 'PERU' then volume | |  | else 0 | |  | end) / sum(volume) as mkt\_share | |  | from | |  | ( | |  | select | |  | year(o\_orderdate) as o\_year, | |  | l\_extendedprice \* (1 - l\_discount) as volume, | |  | n2.n\_name as nation | |  | from | |  | part, | |  | supplier, | |  | lineitem, | |  | orders, | |  | customer, | |  | nation n1, | |  | nation n2, | |  | region | |  | where | |  | p\_partkey = l\_partkey | |  | and s\_suppkey = l\_suppkey | |  | and l\_orderkey = o\_orderkey | |  | and o\_custkey = c\_custkey | |  | and c\_nationkey = n1.n\_nationkey | |  | and n1.n\_regionkey = r\_regionkey | |  | and r\_name = 'AMERICA' | |  | and s\_nationkey = n2.n\_nationkey | |  | and o\_orderdate between '1995-01-01' and '1996-12-31' | |  | and p\_type = 'ECONOMY BURNISHED NICKEL' | |  | ) as all\_nations | |  | group by | |  | o\_year | |  | order by | |  | o\_year; | |

9，**tpch\_query8.sql**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | select | |  | o\_year, | |  | sum(case | |  | when nation = 'PERU' then volume | |  | else 0 | |  | end) / sum(volume) as mkt\_share | |  | from | |  | ( | |  | select | |  | year(o\_orderdate) as o\_year, | |  | l\_extendedprice \* (1 - l\_discount) as volume, | |  | n2.n\_name as nation | |  | from | |  | part, | |  | supplier, | |  | lineitem, | |  | orders, | |  | customer, | |  | nation n1, | |  | nation n2, | |  | region | |  | where | |  | p\_partkey = l\_partkey | |  | and s\_suppkey = l\_suppkey | |  | and l\_orderkey = o\_orderkey | |  | and o\_custkey = c\_custkey | |  | and c\_nationkey = n1.n\_nationkey | |  | and n1.n\_regionkey = r\_regionkey | |  | and r\_name = 'AMERICA' | |  | and s\_nationkey = n2.n\_nationkey | |  | and o\_orderdate between '1995-01-01' and '1996-12-31' | |  | and p\_type = 'ECONOMY BURNISHED NICKEL' | |  | ) as all\_nations | |  | group by | |  | o\_year | |  | order by | |  | o\_year; | |

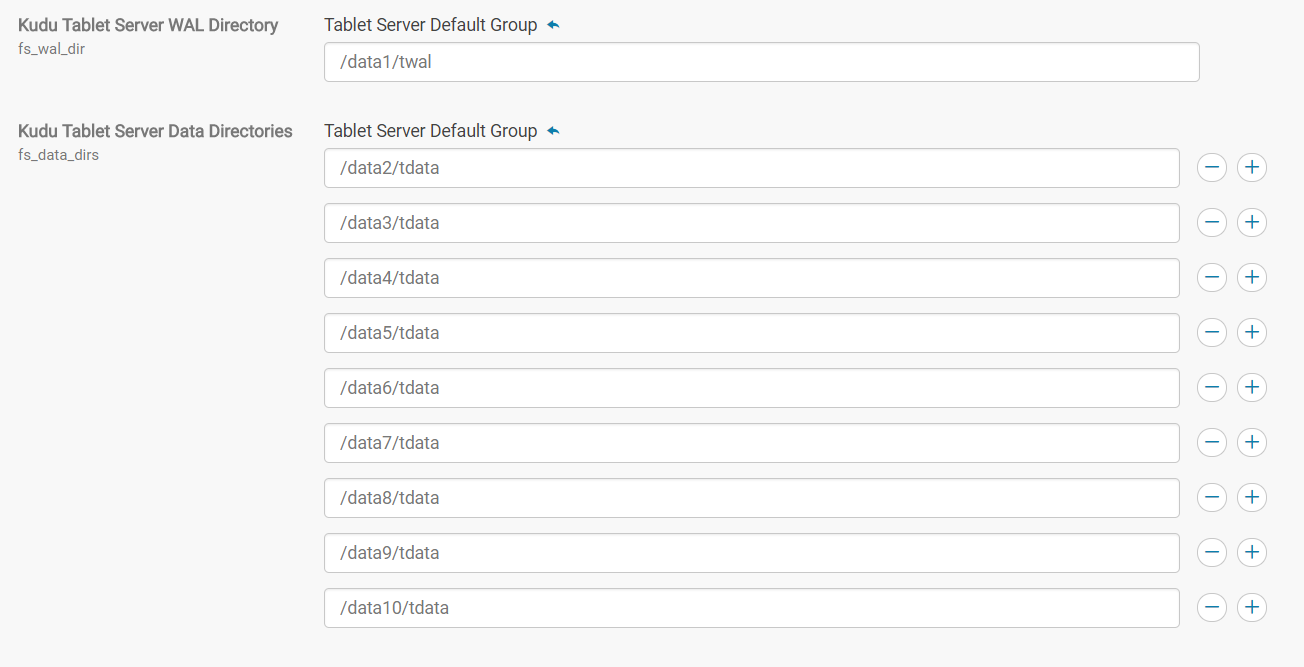
10，**tpch\_query9.sql**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | select | |  | nation, | |  | o\_year, | |  | sum(amount) as sum\_profit | |  | from | |  | ( | |  | select | |  | n\_name as nation, | |  | year(o\_orderdate) as o\_year, | |  | l\_extendedprice \* (1 - l\_discount) - ps\_supplycost \* l\_quantity as amount | |  | from | |  | part, | |  | supplier, | |  | lineitem, | |  | partsupp, | |  | orders, | |  | nation | |  | where | |  | s\_suppkey = l\_suppkey | |  | and ps\_suppkey = l\_suppkey | |  | and ps\_partkey = l\_partkey | |  | and p\_partkey = l\_partkey | |  | and o\_orderkey = l\_orderkey | |  | and s\_nationkey = n\_nationkey | |  | and p\_name like '%plum%' | |  | ) as profit | |  | group by | |  | nation, | |  | o\_year | |  | order by | |  | nation, | |  | o\_year desc; | |

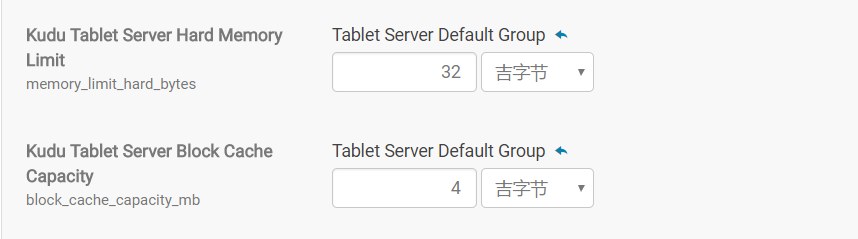
|  |  |  |
| --- | --- | --- |
| Sql | Hive时间 | Kudu时间（impala） |
| **tpch\_query1** | Time taken: 76.664 seconds, Fetched: 4 row(s) | Fetched 4 row(s) in 9.48s |
| **tpch\_query5** | Time taken: 310.823 seconds, Fetched: 5 row(s) | Fetched 5 row(s) in 194.38s |
| **tpch\_query3** | Time taken: 344.762 seconds, Fetched: 10 row(s) | Fetched 10 row(s) in 79.49s |
| **tpch\_query4** | Time taken: 137.316 seconds, Fetched: 5 row(s) | Fetched 5 row(s) in 167.58s |
| **tpch\_query6** | Time taken: 156.265 seconds, Fetched: 1 row(s) | Fetched 1 row(s) in 2.27s |
| **tpch\_query7** | Time taken: 400.269 seconds, Fetched: 4 row(s) | Fetched 4 row(s) in 134.75s |
| **tpch\_query8** | 超过半小时 | Fetched 2 row(s) in 219.91s |
| **tpch\_query9** | 超过半小时 | Fetched 175 row(s) in 255.17s |
| **tpch\_query10** | Time taken: 233.292 seconds, Fetched: 20 row(s) | Fetched 20 row(s) in 43.46s |

# 5，kudu/impala参数调整

1. Kudu： wal日志和数据存放日志分开，wal放在第一块磁盘上，如果有条件可以放在ssd上，数据目录放在第二块到第十二块上面



1. Kudu：memory\_limit\_hard\_bytes参数调整为32G，默认是4G，block\_cache\_capacity\_mb参数调整为4G，默认是512M



3，impala内存调整

