

TPC-H的22条查询语句分析(Q6-Q10)

2017-05-24 Prof.Guo DatabaseG

6 Q6: 预测收入变化查询

先读以下2篇推文：

[数据库基准测试TPC](#)

[TPC-H的22条查询语句分析\(Q1-Q5\)](#)

Q6语句查询得到某一年中通过变换折扣带来的增量收入。这是典型的“what-if”判断，用来寻找增加收入的途径。预测收入变化查询考虑了指定的一年中折扣在“DISCOUNT-0.01”和“DISCOUNT+0.01”之间的已运送的所有订单，求解把l_quantity小于quantity的订单的折扣消除之后总收入增加的数量。

Q6语句的特点是：带有聚集操作的单表查询操作。查询语句使用了BETWEEN-AND操作符，有的数据库可以对BETWEEN-AND进行优化。

Q6的查询语句如下：

```
select
sum(l_extendedprice*l_discount) as revenue //潜在的收入增加量
from lineitem //单表查询
where
l_shipdate >= date '[DATE]' //DATE是从[1993, 1997]中随机选择的一年的1月1日
and l_shipdate < date '[DATE]' + interval '1' year //一年内
and l_discount between [DISCOUNT] - 0.01 and [DISCOUNT] + 0.01 //between
and l_quantity < [QUANTITY]; // QUANTITY在区间[24, 25]中随机选择
```

7 Q7: 货运盈利情况查询

Q7语句是查询从供货商国家与销售商品的地区之间通过销售获利情况的查询。此查询确定在两国之间货运商品的量用以帮助重新谈判货运合同。

Q7语句的特点是：带有分组、排序、聚集、子查询操作并存的多表查询操作。子查询的父层查询不存在其他查询对象，是格式相对简单的子查询。

Q7的查询语句如下：

```
select
```

```

supp_nation, //供货商国家
cust_nation, //顾客国家
l_year, sum(volume) as revenue //年度、年度的货运收入
from ( //子查询
select
n1.n_name as supp_nation,
n2.n_name as cust_nation,
extract(year from l_shipdate) as l_year,
l_extendedprice * (1 - l_discount) as volume
from
supplier,lineitem,orders,customer,nation n1,nation n2 //六表连接
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 ( // NATION2和NATION1的值不同，表示查询的是跨国的货运情况
(n1.n_name = '[NATION1]' and n2.n_name = '[NATION2]')
or (n1.n_name = '[NATION2]' and n2.n_name = '[NATION1]')
)
and l_shipdate between date '1995-01-01' and date '1996-12-31'
) as shipping
group by supp_nation, cust_nation, l_year
order by supp_nation, cust_nation, l_year;

```

8 Q8: 国家市场份额查询

Q8语句是查询在过去的两年中一个给定零件类型在某国某地区市场份额的变化情况。

Q8语句的特点是：带有分组、排序、聚集、子查询操作并存的查询操作。子查询的父层查询不存在其他查询对象，是格式相对简单的子查询，但子查询自身是多表连接的查询。

Q8的查询语句如下：

```

select
o_year, //年份

```

```
sum(case
when nation = '[NATION]'//指定国家，在TPC-H标准指定的范围内随机选择
then volume
else 0
end) / sum(volume) as mkt_share //市场份额：特定种类的产品收入的百分比；聚集操作
from //子查询
(select
extract(year from o_orderdate) as o_year, //分解出年份
l_extendedprice * (1-l_discount) as volume, //特定种类的产品收入
n2.n_name as nation
from
part,supplier,lineitem,orders,customer,nation n1,nation n2,region //八表连接
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 = '[REGION]' //指定地区，在TPC-H标准指定的范围内随机选择
and s_nationkey = n2.n_nationkey
and o_orderdate between date '1995-01-01' and date '1996-12-31' //只查95、96年的情况
and p_type = '[TYPE]' //指定零件类型，在TPC-H标准指定的范围内随机选择
) as all_nations
group by //按年分组
o_year
order by //按年排序
o_year;
```

TPC-H标准定义了Q8语句等价的变形SQL，与上述查询语句格式上基本相同，主要是目标列使用了不同的表达方式，在此不再赘述。

9 Q9: 产品类型利润估量查询

Q9语句是查询每个国家每一年所有被定购的零件在一年中的总利润。

Q9语句的特点是：带有分组、排序、聚集、子查询操作并存的查询操作。子查询的父层查询不存在其他查询对象，是格式相对简单的子查询，但子查询自身是多表连接的查询。子查询中使用了LIKE操作符，有的查询优化器不支持对LIKE操作符进行优化。

Q9的查询语句如下：

```
select
nation,
o_year,
sum(amount) as sum_profit //每个国家每一年所有被定购的零件在一年中的总利润
from
(select
n_name as nation, //国家
extract(year from o_orderdate) as o_year, //取出年份
l_extendedprice * (1 - l_discount) - ps_supplycost * l_quantity as amount //利润
from
part,supplier,lineitem,partsupp,orders,nation //六表连接
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 '%[COLOR]%' //LIKE操作，查询优化器可能进行优化
) as profit
group by //按国家和年份分组
nation,
o_year
```

```
order by //按国家和年份排序，年份大者靠前  
nation,  
o_year desc;
```

10 Q10: 货物存在问题查询

Q10语句是查询每个国家在某时刻起的三个月内货运存在问题的客户和造成的损失。

Q10语句的特点是：带有分组、排序、聚集操作并存的多表连接查询操作。查询语句没有从语法上限制返回多少条元组，但是TPC-H标准规定，查询结果只返回前10行（通常依赖于应用程序实现）。

Q10的查询语句如下：

```
select  
c_custkey, c_name, //客户信息  
sum(l_extendedprice * (1 - l_discount)) as revenue, //收入损失  
c_acctbal,  
n_name, c_address, c_phone, c_comment //国家、地址、电话、意见信息等  
from  
customer, orders, lineitem, nation  
where  
c_custkey = o_custkey  
and l_orderkey = o_orderkey  
and o_orderdate >= date '[DATE]' // DATE是位于1993年一月到1994年十二月中任一月的一号  
and o_orderdate < date '[DATE]' + interval '3' month //3个月内  
and l_returnflag = 'R' //货物被回退  
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;

喜欢就点赞 ， 爱就转发 。

