

TPC-H的22条查询语句分析(Q11-Q15)

2017-05-25 Prof.Guo DatabaseG

先读以下3篇推文：

[数据库基准测试TPC](#)

[TPC-H的22条查询语句分析\(Q1-Q5\)](#)

[TPC-H的22条查询语句分析\(Q6-Q10\)](#)

11 Q11: 库存价值查询

Q11语句是查询库存中某个国家供应的零件的价值。

Q11语句的特点是：带有分组、排序、聚集、子查询操作并存的多表连接查询操作。子查询位于分组操作的HAVING条件中。

Q11的查询语句如下：

```
select
ps_partkey,
sum(ps_supplycost * ps_availqty) as value //聚集操作，商品的总价值
from
partsupp, supplier, nation
where
ps_suppkey = s_suppkey
and s_nationkey = n_nationkey
and n_name = '[NATION]'
group by
ps_partkey having //带有HAVING子句的分组操作
sum(ps_supplycost * ps_availqty) > ( //HAVING子句中包括有子查询
select
sum(ps_supplycost * ps_availqty) * [FRACTION] //子查询中存在聚集操作；FRACTION为
0.0001/SF1
from
partsupp, supplier, nation //与父查询的表连接一致
where //与父查询的WHEWR条件一致
ps_suppkey = s_suppkey
and s_nationkey = n_nationkey
```

```

and n_name = '[NATION]' //指定国家
)
order by //按商品的价值降序排序
value desc;

```

12 Q12: 货运模式和订单优先级查询

Q12语句查询获得货运模式和订单优先级。可以帮助决策：选择便宜的货运模式是否会导致消费者更多的在合同日期之后收到货物，而对紧急优先命令产生负面影响。

Q12语句的特点是：带有分组、排序、聚集操作并存的两表连接查询操作。

Q12的查询语句如下：

```

select
l_shipmode,
sum(case //聚集操作
when o_orderpriority = '1-URGENT' //OR运算，二者满足其一即可，选出URGENT或HIGH的
or o_orderpriority = '2-HIGH'
then 1
else 0
end) as high_line_count,
sum(case
when o_orderpriority <> '1-URGENT' //AND运算，二者都不满足，非URGENT非HIGH的
and o_orderpriority <> '2-HIGH'
then 1
else 0
end) as low_line_count
from
orders,lineitem
where
o_orderkey = l_orderkey
and l_shipmode in ('[SHIPMODE1]', '[SHIPMODE2]') /* 指定货运模式的类型，在TPC-H标准
指定的范围内随机选择，SHIPMODE2必须有别于SHIPMODE1 */
and l_commitdate < l_receiptdate
and l_shipdate < l_commitdate
and l_receiptdate >= date '[DATE]' //从1993年到1997年中任一年的一月一号

```

```

and l_receiptdate < date '[DATE]' + interval '1' year //1年之内
group by //分组操作
l_shipmode
order by //排序操作
l_shipmode;

```

TPC-H标准定义了Q12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。

1 SF，Scale Factor，数据库的比例因子。TPC-H标准规定，测试数据库的比例因子必须从下列固定值中选择：1，10,30,100,1000,3000,10000（相当于1GB，10GB,30GB,100GB,1000GB,3000GB,10000GB）。数据库的大小缺省定义为1（例如：SF = 1；近似于1GB）。

13 Q13: 消费者数量查询

Q13语句查询获得消费者的订单数量，包括过去和现在都没有订单记录的消费者。

Q13语句的特点是：带有分组、排序、聚集、子查询、左外连接操作并存的查询操作。

Q13的查询语句如下：

```

select
c_count, count(*) as custdist //聚集操作，统计每个组的个数
from //子查询
(select
c_custkey,
count(o_orderkey)
from
customer left outer join orders on //子查询中包括左外连接操作
c_custkey = o_custkey
and o_comment not like '%[WORD1]%' //LIKE操作
//WORD1 为以下四个可能值中任意一个：special、pending、unusual、express
//WORD2 为以下四个可能值中任意一个：packages、requests、accounts、deposits
group by //子查询中的分组操作
c_custkey
)as c_orders (c_custkey, c_count)
group by //分组操作

```

```
c_count  
order by //排序操作  
custdist desc, //从大到小降序排序  
c_count desc;
```

TPC-H标准定义了Q13语句等价的变形SQL，与上述查询语句格式上不相同，上述语句使用子查询作为查询的对象，变形的SQL把子查询部分变为视图，然后基于视图做查询，这种做法的意义在于有些数据库不支持如上语法，但存在等价的其他语法，如MySQL就不支持如上语法，需要使用如下等价形式。

```
create view orders_per_cust:s (custkey, ordercount) as //创建视图，相当与标准Q13的子查询内容
```

```
select  
c_custkey,  
count(o_orderkey)  
from  
customer left outer join orders on  
c_custkey = o_custkey  
and o_comment not like '%:1%:2%'  
group by  
c_custkey;
```

```
select  
ordercount,  
count(*) as custdist  
from  
orders_per_cust:s //对视图进行查询  
group by  
ordercount  
order by  
custdist desc,  
ordercount desc;
```

```
drop view orders_per_cust:s;
```

14 Q14:促销效果查询

Q14语句查询获得某一个月的收入中有多大的百分比是来自促销零件。用以监视促销带来的市场反应。

Q14语句的特点是：带有分组、排序、聚集、子查询、左外连接操作并存的查询操作。

Q14的查询语句如下：

```
select
100.00 * sum(case
when p_type like 'PROMO%' //促销零件
then l_extendedprice*(1-l_discount) //某一特定时间的收入
else 0
end) / sum(l_extendedprice * (1 - l_discount)) as promo_revenue
from
lineitem, part
where
l_partkey = p_partkey
and l_shipdate >= date '[DATE]' // DATE是从1993年到1997年中任一年的任一月的一号
and l_shipdate < date '[DATE]' + interval '1' month;
```

TPC-H标准定义了Q14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。

15 Q15:头等供货商查询

Q15语句查询获得某段时间内为总收入贡献最多的供货商（排名第一）的信息。可用以决定对哪些头等供货商给予奖励、给予更多订单、给予特别认证、给予鼓舞等激励。

Q15语句的特点是：带有分排序、聚集、聚集子查询操作并存的普通表与视图的连接操作。

Q15的查询语句如下：

create view revenue[STREAM_ID] (supplier_no, total_revenue) as //创建复杂视图（带有分组操作）

select

l_suppkey,

sum(l_extendedprice * (1 - l_discount)) //获取供货商为公司带来的总利润

from

lineitem

where

l_shipdate >= date '[DATE]' //DATE 是从1993年一月到1997年十月中任一月的一号

and l_shipdate < date '[DATE]' + interval '3' month //3个月内

group by //分组键与查询对象之一相同

l_suppkey;

//查询语句

select

s_suppkey,

s_name,

s_address,

s_phone,

total_revenue

from

supplier,revenue[STREAM_ID] //普通表与复杂视图进行连接操作

where

s_suppkey = supplier_no

and total_revenue = (//聚集子查询

select

max(total_revenue)

from

revenue[STREAM_ID] //聚集子查询从视图获得数据

)

```
order by  
s_suppkey;  
//删除视图  
drop view revenue[STREAM_ID];
```

TPC-H标准定义了Q15语句等价的变形SQL，与上述查询语句格式上不相同，上述查询语句首先定义了视图，然后用表与视图连接；变形的SQL定了WITH语句，然后用WITH的对象与表进行连接。变形SQL的语句如下：

```
WITH revenue (supplier_no, total_revenue) as (  
SELECT  
l_suppkey,  
SUM(l_extendedprice * (1-l_discount))  
FROM  
lineitem  
WHERE  
l_shipdate >= date ':1'  
AND l_shipdate < date ':1' + interval '3' month  
GROUP BY  
l_suppkey  
)
```

```
SELECT  
s_suppkey,  
s_name,  
s_address,  
s_phone,  
total_revenue  
FROM  
supplier,  
revenue  
WHERE  
s_suppkey = supplier_no  
AND total_revenue = (  
SELECT
```

```
MAX(total_revenue)
FROM
revenue
)
ORDER BY
s_suppkey;
```

喜欢就点赞👍，爱就转发🌹😊。



长按“识别图中二维码”关注