

TPC-H的22条查询语句分析(Q16-Q22)

2017-05-26 Prof.Guo DatabaseG

16 Q16: 零件/供货商关系查询

先读以下4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\(Q6-Q10\)](#)

[TPC-H的22条查询语句分析\(Q11-Q15\)](#)

Q16语句查询获得能够以指定的贡献条件供应零件的供货商数量。可用于决定在订单量大，任务紧急时，是否有充足的供货商。

Q16语句的特点是：带有分组、排序、聚集、去重、NOT IN子查询操作并存的两表连接操作。

Q16的查询语句如下：

```
select
  p_brand,
  p_type,
  p_size,
  count(distinct ps_suppkey) as supplier_cnt //聚集、去重操作
from
  partsupp,
  part
where
  p_partkey = ps_partkey
  and p_brand <> '[BRAND]'
// BRAND = Brand # MN , M和N是两个字母，代表两个数值，相互独立，取值在1到5之间
  and p_type not like '[TYPE]%' //消费者不感兴趣的类型和尺寸
  and p_size in ([SIZE1], [SIZE2], [SIZE3], [SIZE4], [SIZE5], [SIZE6], [SIZE7], [SIZE8])
//TYPEX是在1到50之间任意选择的一组八个不同的值
  and ps_suppkey not in ( //NOT IN子查询，消费者排除某些供货商
select
```

```

        s_suppkey
    from
        supplier
    where
        s_comment like '%Customer%Complaints%'
)
group by //分组操作
    p_brand,
    p_type,
    p_size
order by //排序操作
    supplier_cnt desc, //按数量降序排列，按品牌、种类、尺寸升序排列
    p_brand,
    p_type,
    p_size;
```

17 Q17: 小订单收入查询

Q17语句查询获得比平均供货量的百分之二十还低的小批量订单。对于指定品牌和指定包装类型的零件，决定在一个七年数据库的所有订单中这些订单零件的平均项目数量（过去的和未决的）。如果这些零件中少于平均数20%的订单不再被接纳，那平均一年会损失多少呢？所以此查询可用于计算出如果没有小量订单，平均年收入将损失多少（因为大量商品的货运，将降低管理费用）。

Q17语句的特点是：带有聚集、聚集子查询操作并存的两表连接操作。

Q17的查询语句如下：

```

select
    sum(l_extendedprice) / 7.0 as avg_yearly //聚集操作
from
    lineitem, part
where
    p_partkey = l_partkey
    and p_brand = '[BRAND]' /*指定品牌。 BRAND = ' Brand#MN' ，M和N是两个字母，代表两个数值，相互独立，取值在1到5之间 */
    and p_container = '[CONTAINER]' //指定包装类型。在TPC-H标准指定的范围内随机选择
```

```

and l_quantity < ( //聚集子查询
  select
    0.2 * avg(l_quantity)
  from
    lineitem
  where
    l_partkey = p_partkey
);

```

18 Q18: 大订单顾客查询

Q18语句查询获得比指定供货量大的供货商信息。可用于决定在订单量大，任务紧急时，验证否有充足的供货商。

Q18语句的特点是：带有分组、排序、聚集、IN子查询操作并存的三表连接操作。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100行（通常依赖于应用程序实现）。

Q18的查询语句如下：

```

select
  c_name, c_custkey, o_orderkey, o_orderdate, o_totalprice, //基本信息
  sum(l_quantity) //订货总数
from
  customer, orders, lineitem
where
  o_orderkey in ( //带有分组操作的IN子查询
    select
      l_orderkey
    from
      lineitem
    group by
      l_orderkey having
        sum(l_quantity) > [QUANTITY] // QUANTITY是位于312到315之间的任意值
  )
  and c_custkey = o_custkey

```

```

    and o_orderkey = l_orderkey
group by
    c_name,
    c_custkey,
    o_orderkey,
    o_orderdate,
    o_totalprice
order by
    o_totalprice desc,
    o_orderdate;

```

19 Q19: 折扣收入查询

Q19语句查询得到对一些空运或人工运输零件三个不同种类的所有订单的总折扣收入。零件的选择考虑特定品牌、包装和尺寸范围。本查询是用数据挖掘工具产生格式化代码的一个例子。

Q19语句的特点是：带有分组、排序、聚集、IN子查询操作并存的三表连接操作。

Q19的查询语句如下：

```

select
    sum(l_extendedprice * (1 - l_discount) ) as revenue
from
    lineitem, part
where
    (
        p_partkey = l_partkey
        and p_brand = '[BRAND1]' /*特定品牌。BRAND1、BRAND2、BRAND3 = 'Brand #
MN'，M和N是两个字母，代表两个数值，相互独立，取值在1到5之间 */
        and p_container in ( 'SM CASE' , 'SM BOX' , 'SM PACK' , 'SM PKG' ) //包装范围
        and l_quantity >= [QUANTITY1] and l_quantity <= [QUANTITY1] + 10 /* QUANTITY1 是
1到10之间的任意取值 */
        and p_size between 1 and 5 //尺寸范围
        and l_shipmode in ( 'AIR' , 'AIR REG' ) //运输模式，如下带有阴影的粗体表示的条件是
相同的，存在条件化简的可能
        and l_shipinstruct = 'DELIVER IN PERSON'
    )

```

```

or
(
  p_partkey = l_partkey
  and p_brand = '[BRAND2]'
  and p_container in ( 'MED BAG' , 'MED BOX' , 'MED PKG' , 'MED PACK' )
  and l_quantity >= [QUANTITY2] and l_quantity <= [QUANTITY2] + 10 /* QUANTITY2 是
10到20之间的任意取值 */
  and p_size between 1 and 10
  and l_shipmode in ( 'AIR' , 'AIR REG' )
  and l_shipinstruct = 'DELIVER IN PERSON'
)
or
(
  p_partkey = l_partkey
  and p_brand = '[BRAND3]'
  and p_container in ( 'LG CASE' , 'LG BOX' , 'LG PACK' , 'LG PKG' )
  and l_quantity >= [QUANTITY3] and l_quantity <= [QUANTITY3] + 10 /* QUANTITY3 是
20到30之间的任意取值 */
  and p_size between 1 and 15
  and l_shipmode in ( 'AIR' , 'AIR REG' )
  and l_shipinstruct = 'DELIVER IN PERSON'
);

```

20 Q20: 供货商竞争力查询

Q20语句查询确定在某一年内，找出指定国家的能对某一零件商品提供更有竞争力价格的供货货。所谓更有竞争力的供货商，是指那些零件有过剩的供货商，超过供或商在某一年中货运给定国的某一零件的50%则为过剩。

Q20语句的特点是：带有排序、聚集、IN子查询、普通子查询操作并存的两表连接操作。

Q20的查询语句如下：

```

select
  s_name, s_address
from
  supplier, nation
where

```

```

s_suppkey in ( //第一层的IN子查询
select
    ps_suppkey
from
    partsupp
where
    ps_partkey in ( //第二层嵌套的IN子查询
        select
            p_partkey
        from
            part
        where
            p_name like '[COLOR]%' //COLOR为产生P_NAME的值的列表中的任意值
    )
and ps_availqty > ( //第二层嵌套的子查询
    select
        0.5 * sum(l_quantity) //聚集子查询
    from
        lineitem
    where
        l_partkey = ps_partkey
        and l_suppkey = ps_suppkey
        and l_shipdate >= date('[DATE]' ) //DATE为在1993年至1997年的任一年的一月
        and l_shipdate < date('[DATE]' ) + interval '1' year //1年内
    )
)
and s_nationkey = n_nationkey
and n_name = '[NATION]' //TPC-H标准定义的任意值
order by
    s_name;

```

21 Q21: 供货商竞争力查询

Q21语句查询获得不能及时交货的供货商。

Q21语句的特点是：带有分组、排序、聚集、EXISTS子查询、NOT EXISTS子查询操作并存的四表连接操作。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100行（通常依赖于应用程序实现）。

Q21的查询语句如下：

```
select
    s_name, count(*) as numwait
from
    supplier, lineitem l1, orders, nation
where
    s_suppkey = l1.l_suppkey
    and o_orderkey = l1.l_orderkey
    and o_orderstatus = 'F'
    and l1.l_receiptdate > l1.l_commitdate
    and exists ( //EXISTS子查询
        select
            *
        from
            lineitem l2
        where
            l2.l_orderkey = l1.l_orderkey
            and l2.l_suppkey <> l1.l_suppkey
    )
    and not exists ( //NOT EXISTS子查询
        select
            *
        from
            lineitem l3
        where
            l3.l_orderkey = l1.l_orderkey
            and l3.l_suppkey <> l1.l_suppkey
            and l3.l_receiptdate > l3.l_commitdate
    )
    and s_nationkey = n_nationkey
    and n_name = '[NATION]' //TPC-H标准定义的任意值
group by
    s_name
order by
```

```
numwait desc,
s_name;
```

22 Q22: 全球销售机会查询

Q22语句查询获得消费者可能购买的地理分布。本查询计算在指定的国家，比平均水平更持肯定态度但还没下七年订单的消费者数量。能反应出普通消费者的的态度，即购买意向。

Q22语句的特点是：带有分组、排序、聚集、EXISTS子查询、NOT EXISTS子查询操作并存的四表连接操作。

Q22的查询语句如下：

```
select
  cntrycode,
  count(*) as numcust,
  sum(c_acctbal) as totacctbal
from ( //第一层子查询
  select
    substring(c_phone from 1 for 2) as cntrycode,
    c_acctbal
  from
    customer
  where
    // I1...I7是在TPC-H中定义国家代码的可能值中不重复的任意值
    substring(c_phone from 1 for 2) in ('[I1]','[I2]' ,'[I3]','[I4]','[I5]','[I6]','[I7]')
    and c_acctbal > (//第二层聚集子查询
      select
        avg(c_acctbal)
      from
        customer
      where
        c_acctbal > 0.00
        and substr (c_phone from 1 for 2)
          in ('[I1]','[I2]','[I3]','[I4]','[I5]','[I6]','[I7]')
    )
  and not exists (//第二层NOT EXISTS子查询
    select
```



```
        *  
    from  
        orders  
    where  
        o_custkey = c_custkey  
    )  
    ) as custsale  
group by  
    cntrycode  
order by  
    cntrycode;
```

喜欢就点赞👍，爱就转发🌹😊。



长按“识别图中二维码”关注