

STUDENT ID - 202218061

IT667 - Database Management Systems

Lab Assignment 9 - Query Processing and Optimization

1. QUERY

```
EXPLAIN ANALYZE
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 < '1995-03-15' AND L_SHIPDATE > '1995-03-15'
GROUP BY L_ORDERKEY, O_ORDERDATE, O_SHIPPRIORITY
ORDER BY REVENUE DESC, O_ORDERDATE
LIMIT 10;
```


Query

performance based on optimisation.

:-

```
EXPLAIN ANALYZE
SELECT L_ORDERKEY,
SUM(L_EXTENDEDPRICE*(1-L_DISCOUNT)) AS REVENUE,
O_ORDERDATE,
O_SHIPPRIORITY
FROM CUSTOMER
JOIN ORDERS ON C_CUSTKEY = O_CUSTKEY
JOIN LINEITEM ON L_ORDERKEY = O_ORDERKEY
WHERE C_MKTSEGMENT = 'BUILDING'
AND
O_ORDERDATE < '1995-03-15' AND L_SHIPDATE > '1995-03-15'
GROUP BY L_ORDERKEY, O_ORDERDATE, O_SHIPPRIORITY
ORDER BY REVENUE DESC, O_ORDERDATE
LIMIT 10;
```

202218061_DB/postgres@PostgreSQL 10

Query Query History Data output

```

1 SET SEARCH_PATH TO 'TPCH_DB'
2
3 EXPLAIN ANALYZE
4 SELECT L_ORDERKEY,
5 SUM(L_EXTENDEDPRICE*(1-L_DISCOUNT)) AS REVENUE,
6 O_ORDERDATE,
7 O_SHIPPRIORITY
8 FROM CUSTOMER
9 JOIN ORDERS ON C_CUSTKEY = O_CUSTKEY
10 JOIN LINEITEM ON L_ORDERKEY = O_ORDERKEY
11
12 WHERE C_MKTSEGMENT = 'BUILDING'
13 AND
14 O_ORDERDATE < '1995-03-15' AND L_SHIPDATE > '1995-03-15'
15 GROUP BY L_ORDERKEY, O_ORDERDATE, O_SHIPPRIORITY
16 ORDER BY REVENUE DESC, O_ORDERDATE
17 LIMIT 10;
18
19
20

```

Messages

Successfully run. Total query runtime: 17 secs 342 msec. 31 rows affected.

QUERY PLAN

text

14	Sort Method: quicksort Memory: 1179kB
15	-> Nested Loop (cost=5852.49..134867.31 rows=130008 width=24) (actual tin
16	-> Hash Join (cost=5852.06..40752.96 rows=60449 width=12) (actual tin
17	Hash Cond: (orders.o_custkey = customer.c_custkey)
18	-> Parallel Seq Scan on orders (cost=0.00..34105.50 rows=303000 width
19	Filter: (o_orderdate < '1995-03-15'::date)
20	Rows Removed by Filter: 257565
21	-> Hash (cost=5478.00..5478.00 rows=29925 width=4) (actual time=234.
22	Buckets: 32768 Batches: 1 Memory Usage: 1316kB
23	-> Seq Scan on customer (cost=0.00..5478.00 rows=29925 width=4) (act
24	Filter: (c_mktsegment = 'BUILDING'::bpchar)
25	Rows Removed by Filter: 119858
26	-> Index Scan using lineitem_pkey on lineitem (cost=0.43..1.47 rows=9 w
27	Index Cond: (l_orderkey = orders.o_orderkey)
28	Filter: (l_shipdate > '1995-03-15'::date)
29	Rows Removed by Filter: 4
30	Planning time: 3.449 ms
31	Execution time: 17030.151 ms

Total rows: 31 of 31 Query complete 00:00:17.342

It reduces the planning time after optimization

2. QUERY:-

EXPLAIN ANALYZE

```
SELECT customer.c_name,nation.n_name
FROM "TPCH_DB".customer ,"TPCH_DB".nation
where customer.c_nationkey=nation.n_nationkey
and nation.n_name='ARGENTINA'
```

The screenshot shows a PostgreSQL query execution interface. The query is as follows:

```
1 SET SEARCH_PATH TO 'TPCH_DB'
2
3 EXPLAIN ANALYZE
4 SELECT customer.c_name,nation.n_name
5 FROM "TPCH_DB".customer ,"TPCH_DB".nation
6 where customer.c_nationkey=nation.n_nationkey
7 and nation.n_name='ARGENTINA'
8
9
10 EXPLAIN ANALYZE
11 SELECT customer.c_name,nation.n_name
12 FROM "TPCH_DB".nation
13 join "TPCH_DB".customer
14 on customer.c_nationkey=nation.n_nationkey
15 where nation.n_name='ARGENTINA'
16
17
18
19
20
```

The query plan is as follows:

Step	Operation
1	Hash Join (cost=1.33..5564.82 rows=6000 width=123) (actual time=0.476..1024.260 rows=5975 loops=1)
2	Hash Cond: (customer.c_nationkey = nation.n_nationkey)
3	-> Seq Scan on customer (cost=0.00..5103.00 rows=150000 width=23) (actual time=0.410..972.073 rows=150000 loops=1)
4	-> Hash (cost=1.31..1.31 rows=1 width=108) (actual time=0.049..0.049 rows=1 loops=1)
5	Buckets: 1024 Batches: 1 Memory Usage: 9kB
6	-> Seq Scan on nation (cost=0.00..1.31 rows=1 width=108) (actual time=0.029..0.037 rows=1 loops=1)
7	Filter: (n_name = 'ARGENTINA'::bpchar)
8	Rows Removed by Filter: 24
9	Planning time: 3.125 ms
10	Execution time: 1026.929 ms

The Messages pane shows:

```
Successfully run. Total query
runtime: 1 secs 84 msec.
10 rows affected.
```

Total rows: 10 of 10 Query complete 00:00:01.084

2.Query

performance based on optimisation.

```

EXPLAIN ANALYZE
SELECT customer.c_name,nation.n_name
FROM "TPCH_DB".nation
join "TPCH_DB".customer
on customer.c_nationkey=nation.n_nationkey
where nation.n_name='ARGENTINA'

```

202218061_DB/postgres@PostgreSQL 10	
No limit	
Data output	
<div> <div> <div>+</div> <div> <div> <div>QUERY PLAN</div> <div>text</div> </div> </div> </div> <div> <div>1</div> <div>Hash Join (cost=1.33..5564.82 rows=6000 width=123) (actual time=0.115..83.989 rows=5975 loops=1)</div> </div> <div> <div>2</div> <div>Hash Cond: (customer.c_nationkey = nation.n_nationkey)</div> </div> <div> <div>3</div> <div>-> Seq Scan on customer (cost=0.00..5103.00 rows=150000 width=23) (actual time=0.045..40.071 rows=150000)</div> </div> <div> <div>4</div> <div>-> Hash (cost=1.31..1.31 rows=1 width=108) (actual time=0.037..0.038 rows=1 loops=1)</div> </div> <div> <div>5</div> <div>Buckets: 1024 Batches: 1 Memory Usage: 9kB</div> </div> <div> <div>6</div> <div>-> Seq Scan on nation (cost=0.00..1.31 rows=1 width=108) (actual time=0.024..0.033 rows=1 loops=1)</div> </div> <div> <div>7</div> <div>Filter: (n_name = 'ARGENTINA'::bpchar)</div> </div> <div> <div>8</div> <div>Rows Removed by Filter: 24</div> </div> <div> <div>9</div> <div>Planning time: 0.469 ms</div> </div> <div> <div>10</div> <div>Execution time: 84.474 ms</div> </div> </div>	
Total rows: 10 of 10	Query complete 00:00:00.253

