

CS 460/660 PA3 Answer Sheet

Name: Tyler Waltze

4.1 Statistics of the tables

1) Actual number of tuples: 180,000; estimated number by the optimizer: 180,000;

How do you find this information?

To find the actual number of tuples, I just did:

```
SELECT COUNT(*) FROM partsupp
```

To find the estimated number by the optimizer, I did:

```
SELECT reltuples::bigint AS estimate FROM pg_class where relname='partsupp';
```

2)

Attribute Name	# of Distinct Values
ps_pskey	-1
ps_partkey	399
ps_suppkey	199
ps_availqty	1000
ps_placed	165
ps_ship	264

```
SELECT attname, n_distinct
FROM pg_stats
WHERE table name = 'partsupp';
```

4.2 Index on perfect match query

1) Estimated total cost is 3397;
The cost of a plan means

Cost is the estimated time, measured in arbitrary units, it will take to complete the operation.

2) The estimated result cardinality is 174;
How does the query optimizer get this value?

Is it a reasonable value? YES

3) The access method is Sequential Scan.

4) Order of the tuples returned by the plan:

The order of the rows in the table.

Create index.

5) The access method now is Bitmap Index Scan.

6) Explanation

Without an index, the only way to run the query was to sequentially go through each row. With an index, you know where all the ps_availqty values are, so you can go straight to them.

4.3 Index on range select

1) 125,666 tuples will be returned by this plan; the total cost is 3397;

2) Explanation

3) The access method is Sequential Scan;

Disable the access method in 3)

4) The total cost now is 5080.1;

Order:

It now must use bitmap index scan. The index is set to ps_availqty, so it'll return results ordered by ps_availqty (1, 2, 3, etc). Before it would return the values in the order of their primary keys.

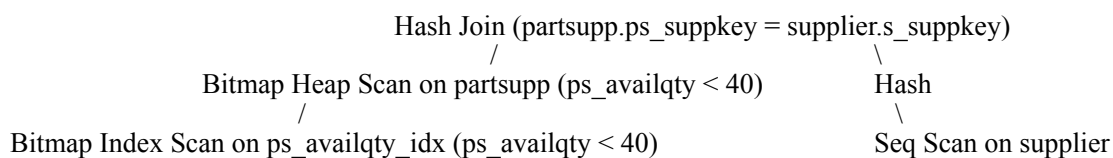
Is it the same as step 1)? NO

5) Explanation:

4.4 Join algorithm

1) Estimated total cost is 1489.96;

Plan Tree:



2) Use Hash join algorithm;

3) Number of tuples that will be retrieved from partsupp is 199;

Disable the join algorithm in 2)

4) Now the join algorithm is Merge; the total cost is now 1947.35

Disable the join algorithm in 2) and 4)

5) Now the join algorithm is Nested Loop; the total cost is now 3349.99

4.5 Three Way Join algorithm

1) Estimated total cost is 7916.50;

Plan Tree:

Hash Join (cost=19.50..7916.50 rows=180000 width=67)

Hash Cond: (partsupp.ps_suppkey = supplier.s_suppkey)

-> Hash Join (cost=12.00..5434.00 rows=180000 width=25)

Hash Cond: (partsupp.ps_partkey = part.p_partkey)

-> Seq Scan on partsupp (cost=0.00..2947.00 rows=180000 width=8)

-> Hash (cost=7.00..7.00 rows=400 width=25)

-> Seq Scan on part (cost=0.00..7.00 rows=400 width=25)

-> Hash (cost=5.00..5.00 rows=200 width=50)

-> Seq Scan on supplier (cost=0.00..5.00 rows=200 width=50)

2) After adding the given condition, what are the differences in the plan,
What are the advantages of the new plan?

With the new plan, the query optimizer can first do a heap scan and then an index scan, which is faster than doing a sequential scan through all rows.