Ceng352 - Database Management Systems Written Assignment 2

Furkan Doğan / 2098937

1.

a.

For this query it is possible to evaluate the query with the index-only plan since it just projects and selects with the given indexes.

b.

For this query it is not possible since it has gender = “Female” which is not shown at the index.

2.

First let give a number to methods like below:

1- Use a heap file (i.e. an unsorted file) storing relation R.

2- Use an unclustered B+ tree index on attribute R.A.

3- Use a (clustered) B+ tree index on attribute R.A

4- Use a hash index on attribute R.A.

a.

Since it is point query and there is no range query 2 will have least cost and if index files are in memory, then 3 and 4 will have almost same effect.

b.

Since it is range query 3 will have least cost.

c.

Since it is range query 3,4 will have least costs.

d.

Since it is just not selection to one number, 4 of them will have same cost unless clustering or indexing is also cost if it is 1 will have least cost.

3.

A picture containing text, table, photo, white

Description automatically generated a.

b.

A close up of a map

Description automatically generated

4.

a.

For block nested loops join algorithms read the outer relation in blocks, and for each block it scans the inner relation for matching tuples. So the outer relation is still read once, but the inner relation is scanned only once for each outer block, of which there are ceiling ( # pages in outer / (B−2) ) = ceil (2000/40) = 50.

TotalCost = B(R) + B(S)∗ceil(B(R) / (M−2)) = 27000

b.

For block nested loops join algorithms read the outer relation in blocks, and for each block it scans the inner relation for matching tuples. So the outer relation is still read once, but the inner relation is scanned only once for each outer block, of which there are ceiling ( # pages in outer / (B−2) ) = ceil (500/40) = 13.

TotalCost = B(S) + B(R)∗ceil(B(S) / (M−2)) = 26500

c.

The cost of merge join is :

2B(R) ⌈logM B(R)⌉ + 2B(S) ⌈logM B(S)⌉ − B(R) − B(S)

2000\*2\*3 + 500\*2\*2 - 2000 - 500 = 11500

All pages of memory is used first for sorting and second for merging.

Sorted buckets is written to the disk.

d.

Since B(R) > M we cannot apply one-pass algorithm so we need to apply partitioned Hash join and the cost of partitioned Hash Join is 3B(R)+3B(S) if min(B(R),B(S))<=M2. Since √B(S) ≈ 14, we can assume that this condition is met. We will also assume uniform partitioning from our hash function.

All available buffer pages in memory is used one for input buffer rest for buckets before writing the disk and 1 used for input 1 used for output and rest used for buckets while reading from memory.

This algorithm writes hashed buckets into a disk when the bucket at the memory fills up.

TotalCost = 3B(R)+3B(S) = 3600.

e.

Since they have indexes, we should use Index-based algorithms to join two relations.

Cost of index nested loop join is:

– If index on S is clustered: B(R) + T(R)B(S)/V(S,b) = 2000 + 20000\*500/5000 = 4000 (since b is primary key)

– If index on S is unclustered: B(R) + T(R)T(S)/V(S,b) = 2000 + 20000 = 22000

5.

a.

tswim\*(m10+m11+m12)/N.

b.

I assume that propotion of each type is same for each mont so cardinality of tswim for each month is tswim/N. But if it is not and the all of swimsuits are sold in first 9 months these assumptions will be wrong.