**Data Warehousing | Assignment # 4**

Abdul Saboor

20L-1113

BDS-5A

-----------------------------------------------------------------------------------------------------------------------------------------------------------

**Assumptions:**

**Query 1 (high selectivity)**

*select avg( (days(date) - days(customer.birth\_dt)) / 365.25 )*

*from customer*

*where customer.state\_cd in (‘CA’ , MA’)*

*and customer.education\_cd = ‘G’*

*and customer.occupation\_cd = ‘CONSULTANT’*

*and customer.hobby\_cd in (‘VOLLEYBALL’,‘CHESS’)*

*;*

**Query 2 (low selectivity)**

*select \**

*from customer*

*where customer.state\_cd in (‘CA’ , MA’)*

*;*

8% of customers live in California.

4% of customers live in Massachusetts.

4% of customers have completed a graduate degree.

6% of customers are consultants.

2% of customers have a primary hobby of chess.

3% of customers have a primary hobby of volleyball.

**K**=100; **B**=65536; **R**=128; **Ri**=16; **rc**=20,000,000; **bfr**=512 (B/R = 64K/128); **bfri**=4096 (B/Ri = 64K/16);

**bc**=39063 (rc/bfr = 20M/512); **bci**=4883 (rc/bfri = 20M/4096);

**Combined Selectivity of Query 1 is 12% of (4% of (6% of (5% of (20,000,000)))) = 288 rows**

**Combined Selectivity of Query 2 is 12% of (20,000,000) = 2,400,000 rows**

**Full Table Scan**

Same for both queries:

bc=39,063 I/O operations

**Single Indexing**

Same for both queries:

Choose from highest selectivity column (degree is 4% of all data)

0.04\*20,000,000 = 800,000

800,000 > 39,063, so all blocks of base table shall be accessed

Index access cost = 800,000/4096 = 196

Total Cost = index access cost + base table access = 196 + 39063 = 39259

**Combining Multiple Indexes**

For Query 1:

State index cost (12% of 20M): (0.12\*20M)/4096 = 586

Education index cost (4% of 20M): (0.04\*20M)/4096 = 196

Occupation index cost (6% of 20M): (0.06\*20M)/4096 = 293

Hobby index cost (5% of 20M): (0.05\*20M)/4096 = 245

Total cost = index access cost + base table access(combined selectivity) = (586+196+293+245) + 288 = 1608

**Combining Multiple Indexes**

For Query 2:

Here all blocks will be read for 2.4M selected rows

Total cost = index access cost + base table access(combined selectivity) = (586+196+293+245) + 39,063 = 40,383

**Dynamic Bitmap Index**

This will be same as combining multiple indexes. 1608 for Query 1 and 40383 for Query 2.

**Static Bitmap Index**

One Static bitmap size = 20M/(16\*4096\*8) = 39 blocks for each value indexed.

For Query 1:

State index: 2\*39 = 78

Education index: 39

Occupation Index: 39

Hobby Index: 2\*39 = 78

Total cost = index access cost + base table access(combined selectivity) = (78+39+39+78) + 288 = 522

For Query 2:

State index: 2\*39 = 78

Total cost = index access cost + base table access(combined selectivity) = 78 + 39,063 = 39,141

**Clustered Index**

For Query 1:

Assuming clustered index is formed on column of *education* (4% of 20M)

20,000,000\*0.04 = 800,000

Total cost = index access cost + base table access(combined selectivity) = (800,000/4096) + (800,000/512) = 1759

For Query 2:

Assuming clustered index is formed on column of *state* (12% of 20M)

20,000,000\*0.12 = 2,400,000

Total cost = index access cost + base table access(combined selectivity) = (2,400,000/4096) + (2,400,000/512) = 5274