National University of Computer and Emerging Sciences, Lahore Campus

	Course:	**Data Warehousing**	**Course Code:**
	Program:	**BS**	**Semester:**
	Practice Problem:	**Indexing Techniques - SOLUTION**	

Instruction/Notes:

Consider the following table and statistics which are part of a student system:
Student (RollNo, Name, gpa, DeptID, BatchID, DegreeID, ...);

Assume student table containing 128,000 rows. Each row and index entry takes 256 and 16 bytes space respectively. Data block size is 16KB and available memory size is 100 blocks. Suppose degree= 'MS' has a selectivity of 3%, batch= ('2019' or '2018') has a selectivity of (4% + 2%), and dept= ('CS or 'EE') has a selectivity of (10% + 5%).

Query: *SELECT * FROM student WHERE DegreeID='MS' AND (BatchID='2019' OR BatchID='2018') AND (DeptID='CS' OR DeptID='EE');*

Calculate the I/O cost for the above query for all the indexes specified below.

1) FULL TABLE SCAN
2) SINGLE INDEXING
3) COMBINING MULTIPLE INDEXES
4) DYNAMIC BITMAP INDEX
5) STATIC BITMAP INDEX
6) COMPOSITE INDEX
7) CLUSTERED INDEX

Ans:

Combine selectivity of student is 3% of (6% of (15% of (128000))) = 35 rows.

3% of (128,000 students) = 3840; 6% of (128,000 students) = 7680; 15% of (128,000 students) = 19200;

K=100; **B**=16384; **R**=256; $\mathbf{R_i}$=16; $\mathbf{r_S}$=128,000; **bfr**=64 (i.e. B/R=16K/256); $\mathbf{bfr_i}$=1024 (i.e. B/R_i=16K/16);

$\mathbf{b_S}$=2000 (i.e. r_S/bfr= 128,000/64); $\mathbf{b_{Si}}$=125 (i.e. r_S/bfr_i= 128,000/1024);

1- FTS: b_s= 2000

2- Using Single Index:

Choose highest selectivity column (i.e. degree - 3% of rows)
0.03* 128000 = 3840 (i.e. 3840 students are from MS degree), As 3840 > 2000, so we have to read all the blocks of base table
Total cost = index access cost + base table access cost = 3840/1024 + 2000 = 2004

3- Combining Multiple Indexes:

Degree Index Access Cost (for 3% of MS students): (0.03* 128000)/1024 = 4

Batch Index Access Cost (for 6% of 2018/2019 students): (0.06* 128000)/1024 = 8

Dept Index Access Cost (for 15% of CS/EE students): (0.15* 128000)/1024 = 19

As combine selectivity is 35, so we read only 35 blocks of base table.

Total cost = index access cost + base table access cost = (4 + 8+ 19) +35 = 66

4- Dynamic Bitmap Indexes: Same as combining multiple indexes cost.

5- Static Bitmap Indexes: One Static bitmap size = 128000 / (16*1024*8) = 1 block for each value indexed.

Total cost = index access cost + base table access cost = (1 + 2+ 2) + 35 = 40

6- Composite Index: Assume Size of composite index = 16 bytes.

Order of composite index = degree, batch, dept.

RID list of each combination of interest must be retrieved i.e. (MS, 2019, CS); (MS, 2019, EE); (MS, 2018, CS); (MS, 2018, EE);

I/O cost for 1st combination = index access cost (3% of 4% of 10% of 128000)/1024 + base table access cost = 1 + 16

I/O cost for 2nd combination = index access cost (3% of 4% of 5% of 128000)/1024 + base table access cost = 1 + 8

I/O cost for 3rd combination = index access cost (3% of 2% of 10% of 128000)/1024 + base table access cost = 1 + 8

I/O cost for 4th combination = index access cost (3% of 2% of 5% of 128000)/1024 + base table access cost = 1 + 4

Total cost = index access cost + base table access cost = 4 + 36 = 40

7- Clustered Index: Assume there is a clustered index on best selectivity column 'Degree' - 3% of rows.

0.03* 128000 = 3840 (i.e. 3840 students are from MS degree)

Total cost = index access cost + base table access cost = 3840/1024 + 3840/64 = 4 + 60 = 64