# National University of Computer and Emerging Sciences, Lahore Campus

| | | | | |
|---|---|---|---|---|
| | **Course:**<br>**Program:**<br><br><br>**Practice Problem:** | **Data Warehousing**<br>**BS**<br><br><br>**Indexing Techniques - SOLUTION** | **Course Code:**<br>**Semester:** | |

**Instruction/Notes:**

Consider the following statistics:

**Assumptions:**

- Block Size = B = 50KB = 50 × 1024 = 51200 bytes
- Record Size = R = 256 B= 256 bytes
- Total number of Records = r =1280000
- Available Memory= K = 100 blocks
- Index Record Size = *Ri* = 16 B = 16 bytes

**Query1: (High Selectivity)**
*SELECT * FROM employee WHERE Department='Research' AND Gender='Female' AND Salary>300000;*

**Query2: (Low Selectivity)**
*SELECT * FROM employee WHERE Department='Research' AND Gender='Female' AND Salary>50000;*

**Now let's assume following selectivity for the query(s):**

- Department="Research" has 55% selectivity
- Gender="Female" has 12% selectivity
- Salary >300000 has 4% selectivity
- Salary >50000 has 60% selectivity

**Calculate the I/O cost for the above query(s) 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

**Answer:**
**Calculating measures necessary for calculations:**
Blocking Factor = bfr = B/R = 51200 / 256 = 200
Blocking Factor for index= *bfri* = B/ *Ri* = 51200 / 16 = 3200
Blocks required for base table records = b = r / bfr = 1280000 / 200 = 6400
Blocks required for Index = *bi* = r/ *bfri* = 1280000 / 3200= 400

## Query1: (High Selectivity)

*SELECT * FROM employee WHERE Department='Research' AND Gender='Female' AND Salary>300000;*

Combined Selectivity = Sc = 0.55 ×(0.12 ×(0.04 ×1280000) ) = 3380 rows
"Research" Selectivity = S1 = 0.55 × 1280000 = 704000 rows
"Female" Selectivity = S2 = 0.12 × 1280000 = 153600 rows
Salary greater than 300000 Selectivity = *S*3 = 0.04 × 1280000 = 51200 rows

**1. Full Table Scan:**
Since we have to scan the Employee table for once, I/O cost is given as:
I/O cost = number of Base Table (Employee) Blocks = b = 6400 blocks

**2. Single Indexing:**
In this case we choose the highest selectivity index, which is salary> 300000. Its Selectivity is given as:
Qualifying rows = S3 = 51200

Now since Qualifying rows > Base table blocks
i.e. 51200 > 6400
Hence, we have to read all the blocks of base table

Also, Index table access cost = *S*3 / bfri = 51200/3200 = 16 blocks
Total I/O cost = Base table access cost + Index table access cost
Total I/O cost = 6400 + 16 = 6416 blocks

**3. Combining Multiple Indexes:**
Since we are taking all indexes we will consider combined selectivity in this case:
Qualifying rows = Combined Selectivity = *Sc* = 3380 rows

Now since Qualifying rows < Base table blocks
i.e. 3380 < 6400
Hence, we have to read only 3380 blocks of base table

Also, Index access cost = Index1 access cost + Index2 access cost+ Index3 access cost
Total Index access cost = *S*1 / *bfri* + S2 / bfri + S3 / bfri
= (704000/ 3200) + (153600/ 3200) + (51200/ 3200) = 220 + 48+16
= 284

Total I/O cost = Base table access cost +Total Index access cost
Total I/O cost = 3380 + 284 = 3664 blocks

**4. Dynamic Bitmap Index:**
Cost will be same as for combining multiple indexes.

**5. Static Bitmap Index:**

Static Bitmap size is given as:

Static Bitmap Size = r / (B × 8) = 1280000 / (51200 × 8) = 4 blocks for each value indexed

So, we need 4 blocks for each index. i.e.

Department Research = 4 blocks Gender Female = 4 blocks

Salary (>300000) = 4 blocks

Total Index access cost = 4+4+4 = 12 blocks

Qualifying rows = Combined Selectivity = *Sc* = 3380 rows

Now since Qualifying rows <Base table blocks

i.e. 3380 < 6400

Hence, we have to read only 3380 blocks of base table

Total I/O cost = Base table access cost +Total Index access cost

Total I/O cost = 3380 + 12 = 3392 blocks

**6. Composite Index:**

Let's assume that size of the composite index is given as:

Composite index size = 16 bytes

Also consider that order of composite index is: Salary, Gender, and Department.

Then,

Combined Selectivity = Sc = 0.55 ×(0.12 ×(0.04 ×1280000) ) = 3380 rows

Now since Qualifying rows < Base table blocks

i.e. 3380 < 6400

Hence, we have to read only 3380 blocks of base table.

Blocking Factor for composite index= *bfri* = B/ *Ri* = 51200 / 16 = 3200

Index access cost = *Sc* / *bfri* = 3380 /3200 = 2 blocks

Total I/O cost = Base table access cost + Index access cost

Total I/O cost = 3380 + 2 = 3382 blocks

**7. Clustered Index:**

Let's assume that clustered index is on Salary attribute. Its Selectivity is given as:

Qualifying rows = S3 = 51200

Now since it is clustered index, these rows will be co-located in the memory.

Number of Blocks for Base Table = *S3* / bfr = 51200 / 200 = 256 blocks

Also, Index table access cost = S3 / bfri = 51200/3200 = 16 blocks

Total I/O cost = Base table access cost + Index table access cost

Total I/O cost = 256 + 16 = 272 blocks

## Query2: (Low Selectivity)

*SELECT * FROM employee WHERE Department='Research' AND Gender='Female' AND Salary>50000;*

Combined Selectivity = Sc = 0.55 ×(0.12 ×(0.60 ×1280000) ) = 50688 rows
"Research" Selectivity = S1 = 0.55 × 1280000 = 704000 rows
"Female" Selectivity = S2 = 0.12 × 1280000 = 153600 rows
Salary greater than 50000 Selectivity = *S*3 = 0.60 × 1280000 = 768000 rows

**1. Full Table Scan:**

Since we have to scan the Employee table for once, I/O cost is given as:
I/O cost = number of Base Table (Employee) Blocks = b = 6400 blocks

**2. Single Indexing:**

In this case we choose the highest selectivity index, which is Gender="Female". Its Selectivity is given as:
Qualifying rows = S2 = 153600
Now since Qualifying rows > Base table blocks
i.e. 153600> 6400
Hence, we have to read all the blocks of base table

Also, Index table access cost = S2 / bfri = 153600/3200 = 48 blocks

Total I/O cost = Base table access cost + Index table access cost
Total I/O cost = 6400 + 48 = 6448 blocks

**3. Combining Multiple Indexes:**

Since we are taking all indexes we will consider combined selectivity in this case:
Qualifying rows = Combined Selectivity = *Sc* = 50688 rows
Now since Qualifying rows > Base table blocks
i.e. 50688 >6400
Hence, we have to read all the blocks of base table

Also, Index access cost = Index1 access cost + Index2 access cost+ Index3 access cost
Total Index access cost = *S*1 / *bfri* + S2 / bfri + S3 / bfri
= (704000/ 3200) + (153600/ 3200) + (768000/ 3200) = 220 + 48+240
= 508

Total I/O cost = Base table access cost +Total Index access cost
Total I/O cost = 6400 + 508 = 6908 blocks

**4. Dynamic Bitmap Index:**

Cost will be same as for combining multiple indexes.

**5. Static Bitmap Index:**

Static Bitmap size is given as:
Static Bitmap Size = r / (B × 8) = 1280000 / (51200 × 8) = 4 blocks for each value indexed
So,
Department Research = 4 blocks Gender Female = 4 blocks
Salary (>50000) = 4 blocks
Total Index access cost = 4+4+4 = 12 blocks
Qualifying rows = Combined Selectivity = *Sc* = Now since Qualifying rows > Base table blocks i.e. 50688 > 6400
50688 rows
Hence, we have to read all the blocks of base table
Total I/O cost = Base table access cost +Total Index access cost Total I/O cost = 6400 + 12 = 6412 blocks

**6. Composite Index:**

Let's assume that size of the composite index is given as:
Composite index size = 16 bytes

Also consider that order of composite index is: Gender, Department and Salary.
Then,
Combined Selectivity = Sc = 0.55 ×(0.12 ×(0.60 ×1280000) ) = 50688 rows

Now since Qualifying rows > Base table blocks
i.e. 50688 > 6400
Hence, we have to read all the blocks of base table

Blocking Factor for composite index= *bfri* = B/ *Ri* = 51200 / 16 = 3200
Index access cost = *Sc* / Composite index size = 50688 /3200= 16 blocks

Total I/O cost = Base table access cost + Index access cost
Total I/O cost = 6400 + 16 = 6416 blocks

**7. Clustered Index:**

Let's assume that clustered index is on Gender attribute. Its Selectivity is given as:
Qualifying rows = S2 =153600
Now since it is clustered index, these rows will be co-located in the memory.
Number of Blocks for of Base Table = *S*2 /bfr = 153600/ 200 = 768 blocks
Also, Index table access cost = S2 / bfri
= 153600/3200 = 48 blocks

Total I/O cost = Base table access cost + Index table access cost
Total I/O cost = 768 + 48 = 816 blocks