Assignment 4

Indexing Techniques

Submission Date: Tuesday 02-Nov-2017 (Start of Class)

Note:

- Plagiarism will result in zero credit in all assignments
- Read the assignment statement carefully
- If you have any confusion , try posting it on Piazza
- You can also make suitable assumptions
- Mention any assumptions before solving the question

Consider the following table and statistics which are part of a leaning management system:

Student (RollNo, Name, DegreeID, BatchID, DeptID, GPA .....);

Block Size	64 KB
Available Memory	100 Blocks
Rows	1,000,000
Row Width	512 bytes
Index Row Width	32 bytes

Assume batch 2014 students are 15%, CS department students are 60%, students having GPA>2.5 are 55% and students having GPA>3.7 are 3%.

Question:

Find the I/O cost for the two given queries for all the indexes specified:

Query 1: Students of the batch 2014 who are from CS dept. and have GPA > 2.5
Query 2: Students of the batch 2014 who are from CS dept. and have GPA > 3.7

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:

1) **FULL TABLE SCAN**
   a. base table is scanned once, so I/O cost is equal to the number of blocks i.e. (1000000*512)/(64*1024) = 7813 (Or Blocking factor = block size / record size = (64*1024)/512=128 , Number of blocks = Number of rows / blocking factor =7813)

   b. base table is scanned once, so I/O cost is equal to the number of blocks i.e. 7813
2) **SINGLE INDEXING**
   a. Choose highest selectivity column i.e. batch
      0.15* 1000000 = 10,000 i.e. 150,000 students are of 2014 batch
      As 150,000 >7813so we have to read all the blocks of base table
      Bfr_index = (64*1024)/32 = 2048
      Index access cost is 150,000 /Bfr_index = 150,000 /2048 = 74
      Total cost = index access cost + base tabe access cost
      Total cost =74+ 7813 = 7887
   b. Choose the highest selectivity column i.e. gpa
      0.03*1000000 = 30000 i.e. 30000 students have gpa > 3.7
      as 30000 > 7813 so we have to read all the blocks of base table
      Bfr_index = (64*1024)/32 = 2048
      Index access cost is 30000/Bfr_index = 30000/2048 = 15
      Total cost = index access cost + base table access cost
      Total cost = 15 +7813= 7828
3) **COMBING MULTIPLE INDEXES**
   a. Batch 2014 students = 15% = 0.15 * 100,0000 = 150000 rows
      CS dept students = 60% = 0.6 * 1000000 = 600000 rows
      GPA>2.5 students = 55% = 0.55 * 1000000 = 550,000 rows
      Index on each of the above columns, we get index costs of 150000/2048 = 74; 600000/2048 = 293; 550,000 /2048 = 269 on batch, dept, and gpa indexes
      Combine selectivity = 0.15*0.6*0.55*1000000 = 49500 > 7813 blocks => so we read all the 7813 blocks of base table
      Total cost = index access cost + base table access cost
      Total cost = 74 + 293 + 269 +7813 = 8449
   b. Batch 2014 students = 15% = 0.15 * 1000000 = 150000 rows
      CS dept students = 60% = 0.6 * 1000000 = 600000 rows
      GPA>3.7 students = 3% = 0.03 * 1000000 = 30000 rows
      Index on each of the above columns, we get index costs of 150000/2048 = 74, 600000/2048 = 293, 30000/2048 = 15 on batch, dept, and gpa indexes
      Combine selectivity = 0.15*0.6*0.03*1000000 = 2700 <7813 blocks => so we read only 2700 blocks of base table
      Total cost = index access cost + base table access cost
      Total cost = 74 + 293 + 15 + 2700 = 3082

4) **DYNAMIC BITMAP INDEX**
   a. Same as Combining Multiple Index cost
   b. Same as Combining Multiple Index cost

5) **STATIC BITMAP INDEX**

   a. Static bitmap size = 1000000/(64*1024*8) = 2 blocks for each value indexed.
      Batch 2014 -> 2 blocks
      Degree CS -> 2 blocks
      Gpa > 2.5 -> 2 blocks
      0.15*0.6*0.55*1000000 = 49500 > 7813 blocks => so we read all the 7813 blocks of base table
      Total cost = index access cost + base table access cost
      Total cost = 2 + 2 + 2 + 7813 = 7819
   b. Static bitmap size = 1000000/(64*1024*8) = 2 blocks for each value indexed.
      Batch 2014 -> 2 blocks
      Degree CS -> 2 blocks
      Gpa >3.7 -> 2 blocks
      0.15*0.6*0.03*1000000 = 2700< 7813 blocks => so we read only 7813 blocks of base table
      Total cost = index access cost + base table access cost
      Total cost = 2 + 2 + 2 + 2700 = 2706

6) **COMPOSITE INDEX**

   a. Assume Size of composite index = 32 bytes
      Order of composite index = batch, degree, gpa
      0.15*0.6*0.55*1000000 = 49500 rows > 7813 so we read all blocks of base table
      Bfr_index = (64*1024)/32 = 2048
      So 49500 RIDs will be stored in 49500/2048 = 25 blocks (index cost)
      Total cost = index access cost + base table access cost
      Total cost = 25 + 7813 = 7838
   b. Assume Size of composite index = 32 bytes
      Order of composite index = batch, degree, gpa
      0.15*0.6*0.03*1000000 = 2700 rows < 7813 so we read only 2700 blocks of base table
      Bfr_index = 2048
      So 2700 RIDs will be stored in 2700 /2048 = 2 blocks (index cost)
      Total cost = index access cost + base table access cost
      Total cost = 2 + 2700= 2702

7) **CLUSTERED INDEX**

   a. Suppose cluster index on batch column
      Batch 2014 students = 15% = 0.15 * 1000000 = 150000 rows
      CS dept students = 60%
      GPA>2.5 students = 55%
      These 150000 rows of students of batch 2014 are stored consecutively
      Bfr = 64*1024/512 = 128 rows per block
      So 150000 rows are in 150000/128 = 1172 blocks
      Bfr_index = (64* 1024)/32 = 2048
      Index access cost = 150000/2048 = 74
      Total cost = index access cost + base table access cost
      Total cost = 74+ 1172 = 1246
   b. Suppose cluster index on gpa column
      Batch 2014 students = 15%
      CS dept students = 60%
      GPA>3.7 students = 3% = 0.03 * 1000000 = 30000 rows
      These 30000 rows of students having gpa > 3.7 are stored consecutively

Bfr = 64*1024/512 = 128 rows per block
So 30000 rows are in 30000/128 = 235 blocks
Bfr_index = 64*1024/32 = 2048
Index access cost = 30000/2048 = 15
Total cost = index access cost + base table access cost
Total cost = 15 + 235 = 250