

Monsoon Mid-Semester Examination, Session 2023-24

Examination & Semester: M.Tech. (Computer Science & Engineering) I Semester
 Subject: Advanced DBMS (CSC502)

Time: 2 Hours
 Max. Marks: 32

Instructions:

- (a) All questions are compulsory.
- (b) Attempt the questions in serial order.
- (c) Make a suitable assumption in case of any missing data.

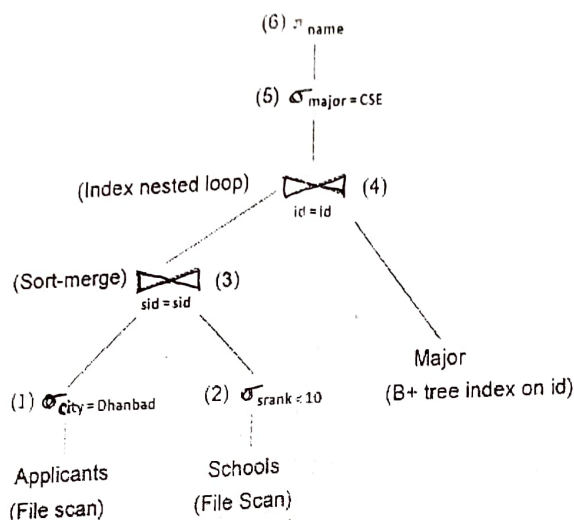
Q. No.	Question	Mar
1. (a)	Give an expression in the Relational Algebra (using Basic Operators only) for each request on schema: <i>employee</i> (<u>person_name</u> , street, city) <i>works</i> (<u>person_name</u> , company_name, salary) <i>company</i> (<u>company_name</u> , city) <i>manages</i> (<u>person_name</u> , manager_name) (i) Modify the database so that Amit now lives in Dhanbad city. (ii) Give all employees of Facebook a 10 percent salary raise. (iii) Find the names of all employees in this database who live in the same city as the company for which they work. (iv) Give all managers in this database a 10 percent salary raise. (v) Assume the companies may be located in several cities. Find all companies located in every city in which Amazon is located.	5
(b)	Given a relation R(A, B, C) and Functional Dependency set $FD = \{ A \rightarrow B, B \rightarrow C, \text{ and } C \rightarrow A \}$, determine given R is in which normal form?	3
2. (a)	Show the steps to remove extraneous attributes from the functional dependency. Find the canonical cover of the given F. $F = \{ A \rightarrow BC, B \rightarrow CE, A \rightarrow E, AC \rightarrow H, D \rightarrow B \}$	5
(b)	Suppose we have a relation R(a, b, c, d, e) and there are at least 1000 distinct values for each of the attributes. Consider each of the following query workloads, independently of each other. If it is possible to speed it up significantly by adding up to two additional indexes to relation R, specify for each index (1) which attribute or set of attributes form the search key of the index, (2) if the index should be clustered or unclustered, (3) if the index should be a hash-based index or a B+-tree. You may add at most two new indexes. If adding a new index would not make a significant difference, you should say so. Give a brief justification for your answers. (i) 100,000 queries have the form: select * from R where b<? 10,000 queries have the form: select * from R where c=? (ii) 100,000 queries have the form: select * from R where b<? and c=? 10,000 queries have the form: select * from R where d=? 1,000 queries have the form: select * from R where a=?	3
3. (a)	Consider the following SQL query that finds all applicant who want to major in CSE, live in Dhanbad, and go to a school ranked better than 10 (i.e., rank<10). <pre> SELECT A.name FROM Applicants A, Schools S, Major M WHERE A.sid=S.sid AND A.id=M.id AND A.city='Dhanbad' AND S.rank<10 AND M.major='CSE' </pre> Details of relations used are: Applicants(id, name, city, sid) : Cardinality: 2,000 and Number of pages or blocks: 100	5

Schools(sid, sname, srnk) : Cardinality: 100 and Number of pages or blocks: 10
 Major(id, major) : Cardinality: 3,000 and Number of pages or blocks: 200

And assuming:

- Each school has a unique rank number (srnk value) between 1 and 100.
- There are 20 different cities.
- Applicants.sid is a foreign key that references Schools.sid.
- Major.id is a foreign key that references Students.id.
- There is an unclustered, secondary B+ tree index on Major.id and all index pages are in memory.

What is the cost (in terms of number of page I/Os) of the query plan below? Explain the process of calculation in detail.



- (b) Consider an institute and the relation *Registered*(*Stu_Admn_no*, *Course_No*, *Session*, *Year*, *Units*, *Grade*) contains the grades for the courses completed by students during the last 20 years. For simplicity, assume that there are 25,000 students enrolled each session, and that each student takes four courses per session, and that there are four sessions each year. Then we get a total of 8,000,000 records. If 10,000 new students enter institute every year, we can assume that in *Registered* there are 200,000 different students, each identified by its *Stu_Admn_no*. On the average, a student took 40 different courses. The file blocks hold 2048 bytes and each *Registered* tuple requires 50 bytes. The table has a primary index on *Stu_Admn_no*. If the *Stu_Admn_no* index is a B+ tree with order $n = 101$, how many levels does the B+ tree use, in the worst case. Justify your answer.

3

4. (a) Consider join processing using symmetric fragment and replicate with range partitioning. How can you optimize the evaluation if the join condition is of the form $|r.A - s.B| \leq k$, where k is a small constant. Here, $|x|$ denotes the absolute value of x . A join with such a join condition is called a band join.
- (b) Differentiate between pipelined and independent parallelism. Set up a pipeline architecture to compute three joins in parallel.

5

3