INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR

End-Spring Semester Examination 2022-23

Date of Examination: _____ Session: (FN/AN) _____ Duration: **3 hours** Full Marks: **50**

Subject No.: **MA40004/MA60308/MA60050** Subject: **FILE ORGANIZATION AND DATABASE SYSTEMS**

Department/Center/School: **Mathematics**

Specific charts, graph paper, log book, etc., required: **NO**

Special Instructions (if any): Answer **ALL EIGHT (8)** Questions. This question paper consists of **3 pages.**

---

1. Consider the following BANK database.

Branch (branch_name, branch_city, asset)
Customer (customer_name, customer_street, customer_city)
Borrow (loan_no, branch_name, customer_name, amount)
Deposit (branch_name, customer_name, account_no, balance)

Express the following queries in **SQL** and **QUEL**:

(I) Find all customers who have loan and account at 'Kharagpur' branch.
(II) Find all account numbers whose balance is more than the average balance of that branch.
(III) Find all customers who have account at 'Kharagpur' branch but no loan from 'Kharagpur' branch.

(3M+3M+3M)

2. Which of the file organizations would you choose for a file where the most frequent operations are as follows? Justify your answers.

(I) Search for records based on a range of field values.
(II) Perform inserts and scans, where the order of records does not matter.
(III) Search for a record based on a particular field value.

(3M)

3.

(I) Outline the essential features of a heap file organization. For what type of operations heap file is most suitable? Justify your answer.
(II) Distinguish between dense index and sparse index. When is it preferable to use a dense index rather than a sparse index? Justify your answer.
(III) If a file uses hashed file organization, is it possible to perform range queries on a search key value? If yes, justify your answer.

(2M+3M+2M)

4. Construct a B+-tree for the following set of key values:
(2, 3, 5, 7, 11, 17, 19, 23, 29, 31)

(I) Assume that the tree is initially empty and values are added in ascending order. Construct B+ trees for the cases where the number of pointers that will fit in one node is as follows: (a) Four, (b) Six, (c) Eight.

(P.T.O)

(II) Also, for all 3 B+ trees, show the steps involved in this query: Find records with a search-key value between 7 and 17, inclusive.

(3M+2M)

5.

(I) Suppose that we are using extendable hashing on a file that contains records with the search-key values: {2, 3, 5, 7, 11, 17, 19, 23, 29, 31}
Show the extendable hash structure for this file if the hash function is h(x) = x mod 8 and buckets can hold three records.

(II) Show how this extendable hash structure changes as the result of each of the following steps: (a) Delete 31, (b) Insert 1, (c) Insert 15.

(III) Consider the following COMPANY database.
EMPLOYEE (SSN, Name, Bdate Address, Sex, Salary, SuperSSN, DNo)
DEPARTMENT (DNo, DName, MgrSSN, MgrStartDate)
DLOCATION (DNo,DLoc)
PROJECT (PNo, PName, PLocation, DNo)
WORKS_ON (SSN, PNo, Hours)
DEPENDENT (SSN, Dependent_name, Sex, Bdate, Relationship)
Write the SQL command and draw the initial operator graph for the following query. Then apply heuristic rules to transform the queries into a more efficient form. Show all the intermediate steps.
Query: Find the names of employees born after 1987 who work on a project named 'IndicNLP'.

(3M+2M+3M)

6.

(I) Consider the following schedules S_1 and S_2 for three transactions T_1, T_2, and T_3.
S_1: $R_1(X)$, $R_3(Y)$, $R_3(X)$, $R_2(Y)$, $R_2(Z)$, $W_3(Y)$, $W_2(Z)$, $R_1(Z)$, $W_1(X)$, $W_1(Z)$
S_2: $R_1(X)$, $R_3(Y)$, $R_2(Y)$, $R_3(X)$, $R_1(Z)$, $R_2(Z)$, $W_3(Y)$, $W_1(X)$, $W_2(Z)$, $W_1(Z)$

a) Design the precedence graphs for the schedules S_1 and S_2.

b) Is both S_1 and S_2 conflict serializable? Explain why or why not.

(II) T1: Read(A); Read(B); if A = 0 then B := B + 1; Write(B).
T2: Read(B); Read(A); if B = 0 then A := A + 1; Write(A).
Add lock and unlock instructions to transactions T1 and T2, so that they observe the two-phase locking protocol. Can the execution of these transactions result in a deadlock? Justify.

(4M+3M)

7.

(I) In the process of recovery from a system failure, discuss the importance and structure of log table. Write short note on (a) Transaction-consistent check point, (b) Action-consistent checkpoint, and (c) Transaction- oriented check point.

(II) Discuss the potential advantages of distributed databases. Discuss different kinds of data fragmentation and the utility of fragmentation in the design and query processing of distributed database. Why data replication is useful in Distributed Databases Management System?

(3M+3M)

(P.T.O)

8.

Write Apriori algorithm to generate frequent item sets from a set of transactions consisting of items.

Consider the following transactions for the set of items

T-ID	Items purchased
T1	Bread, Coke, Beer
T2	Coke, Diaper
T3	Coke, Milk
T4	Bread, Coke, Diaper
T5	Bread, Milk
T6	Coke, Milk
T7	Bread, Coke, Milk, Beer
T8	Bread, Milk
T9	Bread, Coke, Milk

Use Apriori algorithm to find the frequent Item sets. Hence generate the corresponding association rules. Use $\frac{2}{9}$ as minimum support value and $\frac{7}{9}$ as minimum confidence value.

How many candidate item sets and possible rules can be generated with the database of n items?

(2M+2M+1M)

----------X----------