Course Code : CST 321/CST 308 **CXDW/RW – 18 / 5505**

Sixth Semester B. E. (Computer Science and Engineering) Examination

DATABASE MANAGEMENT SYSTEM

Time : 3 Hours] [Max. Marks : 60

Instructions to Candidates :—

(1) Attempt **all** questions.
(2) All questions carry marks as indicated against them.
(3) Assume suitable data and illustrate answers neat sketches wherever necessary.
(4) Mobile phones and / or electronic gadgets are prohibited in the examination hall.
(5) Use of non-programmable electronic calculator is permitted.

1. Attempt Either **(a)** or **(b)**. **Part (c) is Compulsory.**

(a) Discuss the categorization of database users. State the functions of a DBA. 5 (CO 1)

(b) State and discuss different DBMS interfaces clearly indicating the targeted users and user needs addressed. 5 (CO 1)

(c) Consider the following schema and write the relational algebra and SQL equivalent of the mentioned queries.

Parts (partID, partName, Color)
Catalog (supID, partID, Cost)
Suppliers (supID, supName, Address)

(i) Find the supplier IDs of suppliers who supply some red or green part.

(ii) Find the supplier IDs of suppliers who supply some red part or stays at Gulmohar Ridge Road.

(iii) Find names of suppliers who supply every part. 5 (CO 1,3)

2. Attempt Either **(a)** or **(b)**. **Part (c) is Compulsory.**

(a) Consider the following relation for published books :

Book (Title, Author, BookType, Price, Affiliation, Publisher)

Suppose the following dependencies exist :

Title $\longrightarrow$ Publisher, BookType
BookType $\longrightarrow$ Price
Author $\longrightarrow$ Affiliation

What normal form is the relation in ? Explain your answer. Normalize until it cannot be decomposed into relations further. 5 (CO 2)

(b) Elaborate on minimal set of functional dependencies.
Let the set of functional dependencies, FX = {A $\longrightarrow$ B, AB $\longrightarrow$ C, D $\longrightarrow$ AC, D $\longrightarrow$ E) and GX = {A $\longrightarrow$ BC, D $\longrightarrow$ AE). Are FX and GX equivalent ? 5 (CO 2)

(c) For the following relational schema specification, write SQL code to create the database applying appropriate constraints. [due credit will be given to named constraints]

Student (stuRoll, lastName, firstName, EnrolID, Email, Advisor)
Staff (staffID, Name, Branch, Designation, joiningDate)
Dept (deptName, Branch, studentIntake, yearEstd)

Following design considerations apply :

(i) The columns lastName, firstName, Email, Name, Designation, deptName contain alphanumeric data. The columns stuRoll, Advisor, staffID, studentIntake and yearEstd are numeric. EnrolID and Branch columns stores alphabetic data.

(ii) The columns firstName, lastName, Name, joiningDate, Designation, studentIntake, yearEstd and deptName will always store some data value.

(iii) EnrolID and Email are unique columns.

(iv) Designation column can only assume values from {Professor, Associate, Assitant}. Branch column can only assume values from {CS, IT, EN}. The studentIntake can be among {30, 60,, 90} and yearEstd should be after 2005.

(v) Advisor column establishes referential integrity in Staff table on primary key column.

(vi) Other PK and FK constraints (implicit) are evident from the schema. 5 (CO 3)

3. Attempt any **Two** of the following :

(a) What are the differences among primary, secondary and clustering indexes ? Which of the indexes are dense, and which are not ? Why can we have at most one primary or clustering index on a file, but several secondary indexes ? 5 (CO 3,4)

(b) A PARTS file with Part# as the key field includes records with the following Part# values : 23, 65, 37, 59, 18, 10, 74, 15, 20, 24, 28, 50, 75, 8, 49. Suppose that the search field values are inserted in the given order in a B-tree of order p = 4 and pleaf = 3; show how the tree will expand and what the final tree will look like. 5 (CO 3,4)

(c) Explain the distinction between closed and open hashing. Discuss the relative merits of each technique in database applications. Why is a hash structure not the best choice for a search key on which range queries are likely ? 5 (CO 3,4)

4. Attempt any **Two** of the following :—

(a) A relation of 40 gigabytes with 4 kilobyte blocks has to be sorted using a memory size of 40 megabytes. Suppose the cost of seek is 5 milliseconds, while the disk transfer rate is 40 megabytes per second.

(i) Find the cost of sorting the relation, in seconds, with $b_b = 1$, and with $b_b = 100$.

(ii) In each cases, how many merge passes are required ? 5 (CO 4)

(b) Discuss the reasons for converting SQL queries into relational algebra queries before optimization is done. Discuss the difference between pipelining and materialization. 5 (CO 4)

(c) For the relation schemas –

Instructor (ID, Name, deptName, Salary)
Teacher (ID, courseID, secID, Semester, Year)
Course (courseID, courseTitle, deptName, Credits)

Give relational-algebra expression for the query "Find the names of all instructors in the Music department together with the course title of all the courses that the instructors taught in 2009."
Apply appropriate transformations to convert the initial query tree into the optimized one. Show the transformation steps. 5 (CO 5)

5. Attempt any **Two** of the following :—

(a) During its execution, a transaction passes through several states, until it finally commits or aborts. Elaborate with neat sketch, the all possible sequences of states through which a transaction may pass. Explain why each state transition may occur. 5 (CO 5)

(b) What is a cascadeless schedule ? Why is cascadelessness of schedules desirable? Are there any circumstances under which it would be desirable to allow non-cascadeless schedules ? Explain your answer. 5 (CO 5)

(c) Describe the isolation levels specified by the SQL standard.

(i) Give example of a schedule showing the lost update anomaly.

(ii) Give an example schedule to show that the lost update anomaly is possible with the read committed isolation level. 5 (CO 5)

6. Attempt any **Two** of the following :—

(a) Outline the drawbacks of the no-steal and force buffer management policies. Explain why logical undo logging is used widely, whereas logical redo logging (other than physiological redo logging) is rarely used. 5 (CO 5)

(b) Explain why log records for transactions on the undo-list must be processed in reverse order, whereas redo is performed in a forward direction. 5 (CO 5)

(c) Suppose the deferred modification technique is used in a database.

(i) Is the old-value part of an update log record required any more ? Why or why not ?

(ii) If old values are not stored in update log records, transaction undo is clearly not feasible. How would the redo-phase of recovery have to be modified as a result ? 5 (CO 5)