COMP 5 – 5 (RC)

T.E. (Comp.) (Semester – V) (RC) Examination, Nov./Dec. 2016
DATABASE MANAGEMENT SYSTEMS

Duration : 3 Hours | Total Marks : 100

Instructions: *1) Answer **any five** questions by selecting atleast **one** question from **each** Module.*

*2) **Assume** necessary data, **wherever** required.*

MODULE – I

1. a) Assume we have the following application that models soccer teams, the games they play and the players in each team. In the design, we want to capture the following : **10**

We have a set of teams, each team has an ID (unique identifier), name, main stadium and to which city this team belongs.

- Each team has many players and each player belongs to one team. Each player has a number (unique identifier), name, Date of Birth, start year and shirt number that he uses.
- Teams play matches, in each match there is a host team and a guest team. The match takes place in the stadium of the host team.
- For each match we need to keep track of the following :
  - o The date on which the game is played
  - o The final result of the match
  - o The players participated in the match. For each player, how many goals he scored, whether or not he took yellow card and whether or not he took red card.
  - o During the match, one player may substitute another player. We want to capture this substitution and the time at which it took place.
- Each match has exactly three referees. For each referee we have an ID (unique identifier), name, Date of Birth, years of experience. One referee is the main referee and the other two are assistant referee.

Design an ER diagram to capture the above requirements. State any assumptions you have that affects your design. Make sure cardinalities and primary keys are clear.

P.T.O.

b) What do you mean by data model ? Explain any two data models. **4**

c) What do you mean by consistency constraints ? Explain the various types of constraints. **6**

2. a) Explain the various data abstraction levels. What is physical data independence ? **6**

b) Explain the following constraints on generalization : **8**

i) Completeness constraint

ii) Disjoint v/s overlapping constraint

iii) Condition defined v/s user defined constraint

c) What is a storage manager in DBMS ? Explain the various components of storage manager. **6**

MODULE – II

3. a) What do you mean by a functional dependency ? Do all the tuples in the relation should satisfy these dependencies ? Justify. **7**

Consider a relation R(A, B, C, D) with functional dependency $AB \rightarrow C$ and $CD \rightarrow E$. Suppose there are atmost 4 different values for each of A, B and D. What is the maximum number of different values for E ? Explain. What is the key for relation R.

b) Consider the following relations : **10**

Doctor(SSN,FirstName, LastName, Specialty, YearsOfExperience,PhoneNum)

Patient(SSN, FirstName, LastName, Address, DOB, PrimaryDoctor_SSN)

Medicine(TradeName, UnitPrice, GenericFlag)

Prescription(Id, Date, Doctor_SSN, Patient_SSN)

Prescription_Medicine(Prescription Id, TradeName, NumOfUnits)

Answer the following queries in Relational Algebra.

i) List the first and last name of doctors who are not primary doctors to any patient.

ii) For medicines written in more than 24 prescriptions, report the trade name and the total number of units prescribed.

iii) List the SSN of patients who have 'Aspirin' and 'Vitamin' trade names in one prescription.

iv) List the SSN of distinct patients who have 'Aspirin' prescribed to them by doctor named 'James Smith'.

v) List the first and last name of patients who have no prescriptions written by doctors other than their primary doctors.

c) Using the same schema as in 3 b) answer the following query in Tuple relational calculus. **3**

List the first and last name of patients who has primary doctor named 'James Smith'.

4. a) Consider the following relations : **10**

Employee (emp_no, name, skill, pay_rate)

Position(posting_no, skill)

Duty_allocation (posting_no, emp_no,day,shift)

Write the following queries in SQL.

i) Retrieve employees whose rate of pay is more than or equal to the rate of pay of employee 'XYZ'.

ii) Find the employees with the lowest pay rate.

iii) Find the name of all employees who are assigned to all positions that require a Chef's skill.

iv) Get a list of employees who are not assigned a duty.

v) Get the names of the employees working on at least two dates.

b) Let relation R(A,B,C,D) satisfy the following functional dependencies : 5

$A \rightarrow B, B \rightarrow C, C \rightarrow A$

Call this set S1. A different set S2 of functional dependencies is equivalent to S1 if exactly the same FDs follow from S1 and S2. Which of the following sets of FDs is equivalent to the set above ?

i) $C \rightarrow B, B \rightarrow A, A \rightarrow C$

ii) $A \rightarrow BC, C \rightarrow AB$

iii) $B \rightarrow A, B \rightarrow C, C \rightarrow B$

iv) $A \rightarrow B, B \rightarrow A, C \rightarrow A$.

c) Consider the following set of functional dependencies : 5

$F = \{A \rightarrow BCE, AB \rightarrow DE, BI \rightarrow J\}$. Compute the canonical cover for F.

MODULE – III

5. a) Find a BCNF decomposition of the relation schema : 10

SHIPPING (ship,capacity, shipDate,cargo,value) with the following set of functional dependencies :

$ship \rightarrow capacity$

$shipDate \rightarrow Cargo$

$cargo,capacity \rightarrow Value$

Specify the key of the relation. Also mention whether the decomposition is lossless and dependency preserving or not.

b) Consider the relations r1(A, B, C), r2 (C, D, E) and r3 (E, F), with primary keys A, C and E, respectively. Assume that r1 has 1000 tuples, r2 has 1500 tuples and r3 has 750 tuples. Estimate the size of $r1 \bowtie r2 \bowtie r3$ and give an efficient strategy for computing the join. 6

c) Explain the fourth normal form. 4

6. a) Consider the following relational schema : **10**

Project(pno, pname, plocation)

Works_on(empno,pno)

Employee (empno, ename, salary)

Write a query to retrieve the name of all employees working on 'research' project at 'Noida' and earning salary over 50,000/-

Draw the query graph and the initial query tree for the above query and then show (each step) how the query can be optimised. Provide the final optimised query.

b) Consider two relations R1(A, B, C) and R2(A, D). A is the primary key for R2 and R1. A is a foreign key for R1. R1 and R2 are stored as an unordered file (blocks are assumed to be 100% full); R1 contains 200000 tuples that are stored in 500 blocks and R2 has 50000 tuples that are stored in 1000 blocks. Moreover, assume that only a small buffer size of 8 blocks is available. Based on these assumptions answer the following questions (indicate, if you assume in your computations that the output relation is written back to disk or not) : **10**

i) How many tuples will the output relation R = R1 ⋈ R2 contain ? Explain.

ii) What is the cost for implementing the natural join using the block-nested loop join ?

iii) Now assume that either a hashed index on A of R1 or a hashed index on A of R2 is available (assume that there are no overflow pages). Compute the cost for using the index nested loops join using the index for R1. A and for using the index nested loops join for R2.A.

iv) Is it possible to apply the hash-join in this case (explain your answer!) ?

v) Which of the 4 proposed implementations (above) is the best ?

MODULE – IV

7. a) Consider the following classes of schedules: conflict-serializable, view-serializable, recoverable, avoids-cascading-aborts and strict. Explain your answer. For the following schedule, state which of the preceding classes it belongs to. If you cannot decide whether the schedule belongs in a certain class based on the listed actions, explain briefly. **10**

The actions are listed in the order they are scheduled and prefixed with the transaction name.

T1:R(X), T1:R(Y), T2:R(Z), T1:W(X) T2:R(X) T2:W(Y) T2: commit,

T3:R(A), T1:R(Z), T1: Commit, T3:R(Z) T3:W(A), T3:Commit

b) Consider the following actions taken by transaction T1 on database objects A and B: **6**

R(A), W(A), R(B), W(B)

i) Give an example of another transaction T2 that, if run concurrently to transaction T1 without some form of concurrency control, could interface with T1.

ii) Explain how the use of Strict 2PL would prevent interference between the two transactions.

iii) Strict 2PL is used in many database systems. Give two reasons for its popularity.

c) Bring out the differences between the basic two-phase locking and strict two phase locking using a appropriate example. **4**

8. a) For each of the following locking protocols, assuming that every transaction follows that locking protocol, state which of these desirable properties are ensured : serializability conflict-serializability, recoverability, avoidance of cascading aborts. Justify your answer. **8**

i) Always obtain an exclusive lock before writing; hold exclusive locks until end-of-transaction. No shared locks are ever obtained.

ii) In addition to (i), obtain a shared lock before reading; shared locks can be released at any time.

iii) As in (ii) and in addition, locking is two-phase.

iv) As in (ii) and in addition, all locks held until end-of-transaction.

b) Suppose that a DBMS recognizes increment, which increments an integer valued object by 1 and decrement as actions, in addition to reads and writes. A transaction that increments an object need not know the value of the object; increment and decrement are versions of blind writes. In addition to shared and exclusive locks, two special locks are supported: An object must be locked in I mode before incrementing it and locked in D mode before decrementing it. An I lock is compatible with another I or D lock on the same object, but not with S and X locks. Illustrate how the use of I and D locks can increase concurrency. (Show a schedule allowed by Strict 2PL that only uses S and X locks. Explain how the use of I and D locks can allow more actions to be interleaved, while continuing to follow Strict 2PL.) **6**

c) Show that there are schedules that are possible under the two phase locking protocol, but are not possible under the timestamp protocol and vice versa. **6**