Manipal Institute of Technology, Manipal

(A Constituent Institute of Manipal University)

SECOND SEMESTER M.TECH (CSIS)

END SEMESTER EXAMINATIONS, MAY 2016

SUBJECT: ADVANCED CONCEPTS IN DATABASE MANAGEMENT SYSTEMS
[CSE 501]   REVISED CREDIT SYSTEM

Time: 3 Hours   DATE: 05-05-2016   MAX. MARKS: 50

Instructions to Candidates:

❖ Answer **ANY FIVE FULL** questions.
❖ Missing data, if any, may be suitably assumed.

1A. Consider the following set of requirements for a UNIVERSITY database that is used to keep track of students' transcripts.

   a. The university keeps track of each student's name, student number, Social Security number, current address and phone number, permanent address and phone number, birth date, sex, class (freshman, sophomore, ..., graduate), major department, minor department (if any), and degree program (B.A., B.S., ..., Ph.D.). Some user applications need to refer to the city, state, and ZIP Code of the student's permanent address and to the student's last name. Both Social Security number and student number have unique values for each student.

   b. Each department is described by a name, department code, office number, office phone number, and college. Both name and code have unique values for each department.

   c. Each course has a course name, description, course number, number of semester hours, level, and offering department. The value of the course number is unique for each course.

   d. Each section has an instructor, semester, year, course, and section number. The section number distinguishes sections of the same course that are taught during the same semester/year; its values are 1, 2, 3, ..., up to the number of sections taught during each semester.

   e. A grade report has a student, section, letter grade, and numeric grade (0, 1, 2, 3, or 4).

   Design an ER schema for this application, and draw an ER diagram for the schema. Specify key attributes of each entity type, and structural constraints on each relationship type. Note any unspecified requirements, and make appropriate assumptions to make the specification complete.   4M

1B. Consider the following university placements schema. Write SQL queries for the following:
   COMPANY (cname, clocation)
   STUDENT (srollno, sname, sdegree, slocation)
   INTERVIEW (cname, srollno, idate)
   OFFER (cname, srollno, osalary)
   i. Find the name of the students who have been interviewed but don't have any offer.
   ii. Find the name of all the students along with the name of the company if they have been offered a job by a company located in student's location.
   iii. Find the name of the companies that has interviewed every student in the college
   iv. Find the names of the student whose total offer salary is greater than 10 lakhs.   4M

1C. Explain the basic steps in query processing .   2M

2A. Let relations $r1(A, B, C)$ and $r2(C, D, E)$ have the following properties: $r1$ has 20,000 tuples, $r2$

has 45,000 tuples, 25 tuples of *r*1 fit on one block, and 30 tuples of *r*2 fit on one block. Estimate the number of block transfers and seeks required, using each of the following join strategies for *r*1 $\bowtie$ *r*2 : a. Nested-loop join.   b. Block nested-loop join.   c. Merge join.   3M

2B. Consider the relations *r*1(*A*, *B*,*C*), *r*2(*C*, *D*, *E*), and *r*3(*E*, *F*), with primary keys *A*, *C*, and *E*, respectively. Assume that *r*1 has 1000 tuples, *r*2 has 1500 tuples, and *r*3 has 750 tuples. Estimate the size of *r*1 $\bowtie$ *r*2 $\bowtie$ *r*3, and give an efficient strategy for computing the join.   4M

2C. Explain incremental view maintenance for join, selection and projection operations.   3M

3A. Consider a database schema with a relation Customer whose attributes are as shown below, with types specified for multivalued attributes.
Customer = (Cid, Cname, DOB, Address(Street, City, PinCode), AccountSet multiset (AcctNo, Balance, InterestRate), LoanSet multiset(LNo, Amount, Type, Payment(No. of months, EMI))).
   a. Define the above schema Customer in SQL, with appropriate types for each attribute.
   b. Using the above schema, write the following queries in SQL.
      i. List AccNo. of all the accounts held by Mr. Ramesh
      ii. Find the name of customers who has one lakh or above in any of the account
      iii. Find the name of customers whose total balance is greater than total loan amount.   4M

3B. Explain fragment and replicate join in case of intra-operation parallelism.   3M

3C. Explain the different partitioning techniques used in parallel databases. How they facilitate scan, point and range queries?   3M

4A. Give the DTD for an XML representation of the following nested-relational schema: 3M
   *Emp = (ename, ChildrenSet **setof**(Children), SkillsSet **setof**(Skills))*
   Children = (name, Birthday)
   Birthday = (day, month, year)
   *Skills = (type, ExamsSet **setof**(Exams))*
   Exams = (year, city)

4B. Explain the storage of XML data in relational databases with following methods:
   i. string representation   ii. Tree representation   iii. Map to relations   3M

4C. Explain ranking with three variations.   **i.**   rank ( ) over (order by …. desc)
   ii.   rank ( ) over (partition by .....order by…. desc)   **iii.**   ntile()   4M

5A. Explain 2 PC protocol in case of distributed transactions.   4M

5B. Explain an election algorithm to elect a new coordinator in case of failure.   3M

5C. Explain with diagram Star and Snowflake multidimensional schema used to model a data warehouse.   3M

6A. Briefly discuss the different types of NOSQL Data Models :
   i. Key-Value   **ii.**Document-based   **iii.** Column-based   3M

6B. Give details of MapReduce execution. How it cope with node failures?   3M

6C. How matrix multiplication can be done using natural join followed by grouping and aggregation? Give MapReduce implementation for the same.   4M
****