MANIPAL INSTITUTE OF TECHNOLOGY
(Constituent Institute of Manipal University)
MANIPAL-576104
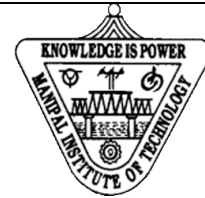
VI SEMESTER B.TECH. (COMPUTER SCIENCE AND ENGINEERING) DEGREE
MAKE-UP EXAMINATION JUNE-JULY 2016
SUBJECT: ADVANCED DATABASE SYSTEMS (CSE 322) [Program Elective]
DATE: 04-07-2016

TIME: 3 HOURS   MAX.MARKS: 50

Instructions to Candidates
- **Note:** Answer any **FIVE** full questions.

1.A. Explain any six selection operations using File Scans and Indices.   3M

1.B. Explain the following:
      i) Pipelined evaluation.   ii) Materialized evaluation.   4M

1.C. Consider the relations $r1$ (A, B, C), $r2$(C, D, E), and $r3$ (E, F). Assume that there are no primary keys, except the entire schema. Let $V(C, r1)$ be 900, $V(C, r2)$ be 1100, $V(E, r2)$ be 50, and $V(E, r3)$ be 100. 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.   3M

2.A. Explain any four lock modes of multiple granularity with the help of compatibility matrix.   3M

2.B. Show by example that there are schedules possible under the tree protocol that are not possible under the two-phase locking protocol, and vice versa.   4M

2.C   Suppose the deferred modification technique is used in a database.
      i). Is the old-value part of an update log record required anymore? 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?
      iii). What problem would arise with the above technique, if transactions perform a large number of updates?   3M

3.A. Explain the Undo and Redo operations in transaction recovery algorithm.   3M

3.B. Write a recursive algorithm of construction of a decision tree.   3M

3.C   Suppose that the data mining task is to cluster the following eight points (with ($x; y$) representing location) into two clusters.
      $A1.(2,10)$, $A2(2,5)$ $A3(8,4)$, $B1(5,8)$,$B2(7,5)$,$B3(6,4)$

      The distance function is Euclidean distance. Suppose initially we assign $A1$ and $B2$ as the center of each cluster, respectively. Use the *k-means* algorithm to show the following:

(a) The two cluster centers after the first round of execution and
(b) The final two clusters.   4M

4.A. Explain the working of web crawlers in indexing the web.   3M

4.B. What are the different types of queries in an IR system? Describe each with an example.   3M

4.C. A car-rental company maintains a database for all vehicles in its current fleet. For all vehicles, it
      includes the vehicle identification number, license number, manufacturer, model, date of
      purchase, and color. Special data are included for certain types of vehicles:
       Trucks: cargo capacity.
       Sports cars: horsepower, renter age requirement.
       Vans: number of passengers.
       Off-road vehicles: ground clearance, drivetrain.
      i) Construct an SQL schema definition for this database. Use inheritance where appropriate.
      ii) Find the vehicle identification number of sports car with highest horsepower and no renter
      requirement.
      iii)Find the manufacturers of the Trucks with higher cargo capacity than the average of all the
         Trucks.   4M

5.A. List and Explain the features of object persistence in Java programs.   3M

5.B. Consider a relation that is fragmented horizontally by *plant number*:   4M
      employee (*name, address, salary, plant number*)
      Assume that each fragment has two replicas: one stored at the New York
      site and one stored locally at the plant site. Describe a good processing
      strategy for the following queries entered at the San Jose site.
      a. Find all employees at the Boca plant.
      b. Find the average salary of all employees.
      c. Find the highest-paid employee at each of the following sites: Toronto, Edmonton, Vancouver,
      Montreal.
      d. Find the second highest paid employee in the company.

5.C. 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)

6.A. Explain the structure of XML data with an example.   3M

6.B. Explain the mobile computing architecture with a neat diagram.   4M

6.C. List and explain any six characteristics of biological data.   3M