MANIPAL INSTITUTE OF TECHNOLOGY
(Constituent Institute of Manipal University)
MANIPAL-576104

FIRST SEMESTER M.Tech.(CSE) END SEMESTER EXAMINATION – DEC– 2014
ADVANCED CONCEPTS IN DATABASE MANAGEMENT SYSTEMS (CSE 501)
02-12-2014

TIME : 3 HOURS   MAX.MARKS : 50

Instruction to Candidates
- Answer **any five** full questions.

1A. Design an ER diagram for keeping track of information about votes taken in the U.S. House of Representatives during the current two-year congressional session. The database needs to keep track of each U. S. STATE's Name(e.g., Texas, New York, California) and include the Region of the state(whose domain is {Northeast, Midwest, Southeast, Southwest, West}). Each CONRESSPERSON in the House of Representatives is described by his or her Name, plus the District represented, the StartDate when the congressperson was first elected, and the political Party to which he or she belongs (whose domain is {Republican, Democrat, Independent, Other}). The database keeps track of each BILL(i.e., proposed law), including the BillName, the DateOfVote on the bill, whether the bill PassedOrFailed (whose domain is {Yes, No}), and the Sponsor (the congressperson(s) who sponsored – that is, proposed – the bill). The database keeps track of how each congressperson voted on each bill (domain of vote attribute is {Yes, No, Abstain, Absent}). State clearly any assumptions you make. Also map the above E-R diagram to the relational schema.

1B. Consider the following database schema:
  Book(*ISBN*, title, Category, Price, PName)
  Publisher(*PName*, BuildingName, Road, City)
  Author(*Aid*, AName, HouseNo, Road, City)
  Writes(*Aid, ISBN*, Position)
  Write the following queries in SQL:
  i. Find the name of authors who has written more than one book
  ii. Find the name of authors who has written maximum number of books
  iii. Find the name of Publishers who have published books in all categories

1C. Explain the basic steps in query processing   (4+4+2)

2A. Compute the query cost of external sort merge for following two cases with a block size of one record.
  i. No. of Blocks = 6, Buffer size = 3 blocks ii. No. of Blocks = 15, Buffer size = 3 blocks

2B. Projection operation distributes over the theta join operation. Explain with the appropriate equivalence rules.

2C. What is the size estimate of join operation, $R \bowtie S$. Illustrate with an example.

2D. Explain incremental view maintenance for join, selection and projection operations

                                                                            (3+2+2+3)

3A. Consider the following nested relation:

title	author_array	publisher	keyword_set
		(name, branch)	
Compilers	[Smith, Jones]	(McGraw-Hill, NewYork)	{parsing, analysis}
Networks	[Jones, Frick]	(Oxford, London)	{Internet, Web}

    Map it to corresponding schema in   i. relational DB and   ii. Object relational DB

3B. Explain the different options available to map specialization to relational schema. What is the drawback of each of these options?

3C. Consider a object relational database schema with a relation 'Travel' whose attributes are as shown below, with types specified for multivalued attributes.
Travel = (TripID, CustomerSet multiset(Customer), Vechile(*VehicleID*, Type, Year), PlacesVisited multiset(Place), Driver(DriverID, Name, Age), StartDate, Amount)
Customer = (name, gender, Profession, Age), Place(City, State, Country)
 a. Define the above schema in SQL:2003, with appropriate types for each attribute.
 b. Using the above schema, write the following queries in SQL:2003
    i.   Find the names of all customers who travelled in trip with TripID 100.
    ii.   Find the trip(s) with max. no. of customers
    iii.   List all the places visited by John M.   (3+2+(2+3))

4A. Explain the different partitioning techniques used in parallel databases. How they facilitate scan, point and range queries?

4B. Why replication and fragmentation is required to store the data in distributed databases? What are its drawbacks?

4C. Explain the different variants of distributed lock managers.

4D. Give Bully algorithm used to elect a new coordinator in case of failures   (3+2+2+3)

5A. Illustrate different multidimensional schema used to model a data warehouse.

5B. What is a cross tab? Explain how it can be i. mapped to a relation? ii. extended to handle hierarchies? and iii. used to apply slicing and dicing?

5C. Consider the following DTD:

```
<!DOCTYPE bank-2[
        <!ELEMENT account (branch, balance)>
        <!ATTLIST account
          account_number ID       # REQUIRED
          owners            IDREFS # REQUIRED>
         <!ELEMENT customer(customer_name, customer_street,
                                  customer_city)>
        <!ATTLIST customer
            customer_id     ID       # REQUIRED
            accounts        IDREFS # REQUIRED>
    … declarations for branch, balance, customer_name,
             customer_street and customer_city
   ]>
```

  i. Give an XML document satisfying the DTD.
  ii. Write the following queries in XQuery:
    a. List the details of the account(s) held by Mr. Amith
    b. List the no. of accounts held by each customer located in Manipal
    c. Convert the flat structure of bank data to a nested structure   (2+4+4)

6A. What is R Tree? How it is used in indexing spatial databases? How the search operation handled in it?

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

6C. Give MapReduce implementation for the following relational algebra operations:
    i. Selection   ii. Natural Join   (3+4+3)

****