Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY, MANIPAL 576104

(Constituent College of Manipal University)

SECOND SEMESTER M.Tech (Network Engg.) DEGREE END SEMESTER EXAMINATION, MAY– 2016
SUBJECT: ELECTIVE-II ADVANCED DATABASE MANAGEMENT SYSTEMS – ICT 542
(REVISED CREDIT SYSTEM)

TIME: 3 HOURS **14/05/2016** **MAX. MARKS: 50**

Instructions to candidates

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

1A. Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70. Give the five-number summary of the data.

1B. List the ACID properties. Explain the usefulness of each.

1C. How do you place the attribute if there exists 1-1, 1-many and many-many relationship between the entity sets E1 and E2?

(5+3+2)

2A. Use pincer search algorithm and discover all frequent itemsets for the following Customer Basket data given in Table Q.2A considering minimum support of 20%.

Table Q.2A

TID	Products
1	Burger, Coke, Juice
2	Juice, Potato, Chips
3	Coke, Burger
4	Juice, Groundnuts
5	Coke, Groundnuts

2B. Determine whether buying coffee and milk together is positively correlated or negatively correlated for the data given in Table Q.2B.

Table Q.2B

	coffee	$\overline{\text{coffee}}$	$\sum_{row}$
milk	4000	3500	7500
$\overline{\text{milk}}$	2000	500	2500
$\sum_{col}$	6000	4000	10000

2C. How can a weak Entity set be made into strong entity set ? Explain with an example.

(5+3+2)

3A. Compute the canonical cover for the relation R=(A,B,C) with the set of functional dependencies F= {A →BC , B →C, A→B, AB →C}

3B. Suppose that a data warehouse consists of the four dimensions, date, spectator, location, and game, and the two measures, count and charge, where charge is the fare that a spectator pays when watching a game on a given date. Spectators may be students, adults, or seniors, with each category having its own charge rate.

(a) Draw a lattice of cuboid for the above mentioned dimension.
(b) Starting with the base cuboid [date, spectator, location, game], what specific OLAP operations should one perform in order to list the total charge paid by student spectators at Bangalore in 2015 for cricket?

3C. Write the BCNF decomposition algorithm and explain how it is stronger than 3NF?

(5+3+2)

4A. Consider the following set of data objects (2,6), (3,4), (3,8), (4,7), (6,2), (6,4), (7,3), (7,4),(8,5),(7,6). Use K-medoid algorithm and Manhattan distance measure to discover two clusters by considering (3,4) and (7,4) as cluster medoids. Check whether the replacement of (i) (3,4) by (2,6) (ii) (7,4) by (8,5) on the initial clusters formed is a good replacement or not.

4B. Write the 3NF decomposition algorithm and hence decompose R=(A, B, C, D, E) into 3NF relations with respect to the following set of functional dependencies, F ={ A →(B, C, D, E), (B, C)→(A , D, E), D→E}

4C. Define downward closure property of frequent itemsets. Give an example.

(5+3+2)

5A. The Table in Q.5A consists of training data. Using the algorithm for decision tree induction, construct a Decision tree based on this data. Classify the records by the "status" attribute and write the rules that can be generated from the obtained decision tree.

Table Q.5A

Department	Age	Marital Status	Salary Class	Status
Sales	Young	Male	High	Senior
Sales	Senior	Female	Low	Senior
Systems	Senior	Male	Low	Junior
Marketing	Young	Female	Average	Junior
Secretary	Middle_Aged	Male	Average	Senior
Sales	Young	Male	High	Junior
Marketing	Middle_Aged	Female	High	Senior
Secretary	Senior	Female	Average	Junior
Systems	Middle_Aged	Male	Average	Senior
Marketing	Senior	Female	Low	Junior
Sales	Middle_Aged	Female	Low	Junior
Systems	Middle_Aged	Female	High	Senior
Secretary	Young	Male	High	Senior

5B. Check whether the decomposition R= (R1, R2) is dependency preserving, where R1 = (A,B,C) , R2 = (A, D, E) and F ={ A→BC, CD→E, B→D,E→A}.

5C. With a neat diagram explain all the steps in Knowledge Discovery in Databases

(5+3+2)

6A. Write the Dynamic Item set Counting algorithm for finding frequent item sets.

6B. Discuss the four cases that need to be considered for conflict serializability with the help of an example.

6C. Distinguish among the terms primary key, candidate key and superkey.

(5+3+2)

*****