Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

(A constituent unit of MAHE, Manipal)

I SEMESTER M.TECH. (SOFTWARE ENGINEERING)

END SEMESTER EXAMINATIONS, NOV 2018

SUBJECT: ADVANCED DATABASE MANAGEMENT SYSTEMS [ICT 5124]

REVISED CREDIT SYSTEM
(27/11/2018)

Time: 3 Hours

MAX. MARKS: 50

Instructions to Candidates:

- Answer **ALL** the questions.
- Missing data if any, may be suitably assumed.

1A. 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}. **5**

1B. 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.

i) Draw a boxplot of the data and also find outliers

ii) What is the mode of the data? **3**

1C. Consider a schema for an unnested relation book = (title, author, pub_name, pub_branch, keyword), where author is of type array, publisher = { pub_name and pub_branch} and keyword is a multiset. Write a query to nest book relation on keyword and array. **2**

2A. 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 (7,6) on the initial clusters formed is a good replacement or not. **5**

2B. Consider the 2 X 2 contingency table given in Table Q.2B, summarizing observed count and the total transactions with respect to Gender and Eye Color.

Table Q.2B

		Eye Color		
		Black	*Gray*	*∑row*
Gender	*Female*	2000	1000	3000
	Male	1000	1500	2500
	∑col	3000	2500	5500

Use $\chi 2$ test to check the dependency of Gender and *Eye Color* for degree of freedom n=1, significance level 0.001and $\chi 2$ in the statistical table is 10.828. **3**

2C. With a neat diagram explain two types of parallel database architectures. **2**

3A. Write and explain the Pincer-Search algorithm for finding frequent itemsets. **5**

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.

i) Draw a lattice of cuboid for the above mentioned dimension.

ii) 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 2017 for cricket? 3

3C. Write an XQuery statement using FLWOR syntax to find all accounts with balance > 400. 2

4A. The Table Q.4A consists of training data. Find the best splitting attribute to construct decision tree using information gain considering the "status" attribute as the class label.

Table Q.4A

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

5

4B. Define and explain data warehouse. What are the major distinguishing features between OLTP and OLAP? 3

4C. List out the databases on which data mining can be performed. 2

5A. Find frequent itemsets for the database given in Table Q.5A by using the Dynamic itemset counting algorithm, assuming stop number as 3 and support count as 3. Clearly show all the steps of the algorithm.

Table Q.5A

TID	ITEMS
1	MILK, BREAD, EGGS
2	BREAD, SUGAR
3	BREAD, CEREAL
4	MILK, BREAD, SUGAR
5	MILK, CEREAL
6	BREAD, CEREAL
7	MILK, CEREAL
8	MILK, BREAD, CEREAL, EGGS
9	MILK, BREAD, CEREAL

5

5B. Explain all partitioning techniques to partition a relation. 3

5C. List and explain any four requirements of clustering in Data Mining. 2