Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

A Constituent Institution of Manipal University

I SEMESTER M.TECH. (SOFTWARE ENGINEERING)

END SEMESTER EXAMINATIONS, NOV/DEC 2016

SUBJECT: **ELECTIVE-II ADVANCED DATABASE MANAGEMENT SYSTEMS**

[ICT 5124]

REVISED CREDIT SYSTEM
(01/12/2016)

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:

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

1A. Construct Frequent Pattern Tree for the following database and compute the frequent items considering a minimum support count of 3, for the following set of transactions {T1={1,5,6,8},T2={2,4,8},T3={4,5,7},T4={2,3},T5={5,6,7},T6={2,3,4},T7={2,6,7,9},T8={5},T9={8},T10={3,5,7},T11={3,5,7},T12={5,6,8},T13={2,4,6,7},T14={1,3,5,7}, T15={2,3,9}} **5**

1B. Write the necessary pseudo code for Candidate generation and pruning step of Apriori algorithm. **3**

1C. Consider a schema for an un nested 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 {(1,6), (3,4), (3,8), (4,7), (5,2), (6,4), (7,3), 7,4),(8,5),(7,6)}. Use K-medoids 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 (4,7) (ii) (7,4) by (6,4) on the initial clusters formed is a good replacement or not. **5**

2B. The research data 736, 767, 576, 793, 779, 761, 568, 768, 491, 883, 890, 681 describe the life span (in days) of mice. Transform the value 761 using

i. Min-Max normalization on to a range [0.0, 1.0].
ii. Z-score normalization.
iii. Decimal scaling. **3**

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

3A. Consider the relation R (A, B, C, D, E) with the functional dependencies F = {{A → BC}, {CD → E}, {B → D}, {E → A}}. Find all the candidate keys of R and also check whether R is in BCNF? If not, then decompose it into BCNF. **5**

3B. Consider the 2 X 2 contingency table given in Table Q.3B, summarizing observed count and the total transactions with respect to type of drinks and snacks preferred by students of an Engineering college.

Table Q.3B

		Drinks		
		Milk	*Coffee*	*∑row*
Snacks	*Pizza*	3000	1000	4000
	Burger	1000	1000	2000
	∑col	4000	2000	6000

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

3C. Write the advantages and disadvantage of data replication. **2**

4A. Construct a decision tree for the dataset given in Table Q.4A, considering the class label "defaulted borrower".

Table Q.4A

TID	Home Owner	Marital Status	Annual Income	Defaulted Borrower
1	Yes	Single	101K – 220K	No
2	No	Married	90K – 100K	No
3	No	Single	60K – 89K	No
4	Yes	Married	101K – 220K	No
5	No	Divorced	90K – 100K	Yes
6	No	Married	60K – 89K	No
7	Yes	Divorced	101K – 220K	No
8	No	Single	60K – 89K	Yes
9	No	Married	60K – 89K	No
10	No	Single	90K – 100K	Yes

5

4B. 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}. **3**

4C. Distinguish between Intraoperation parallelism and Interoperation parallelism. **2**

5A. During the execution of Pincer-search Algorithm, it is found that when k=1(First scan of the database) L_1 ={ {1} , {2} , {3} , {4}, {5}} , S_1 ={} and MFCS = {1,2 3,4,5} , MFS={}
When k=2 (Second scan of the database) L_2= { {1, 2}, {1,3}, {1,5}, {2,3}, {2,4}, {2,5} } S_2 = { {1,4}, {3,4},{3,5}, {4,5} } . Find the MFCS and MFS when k=2. **5**

5B. Explain how fragment and replicate join work. **3**

5C. How do you provide a document type definition for a bank data consisting of account, customer and depositor? **2**