Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

A Constituent Institution of Manipal University

V SEMESTER B.TECH. (MECHATRONICS ENGINEERING)

END SEMESTER EXAMINATIONS, NOV 2019

SUBJECT: DATABASE MANAGEMENT SYSTEMS [MTE 4011]

(27/11/2019)

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:

- Answer **ALL** questions.
- Data not provided may be suitably assumed

		Marks	CO
1A.	Cite an example of how specific clustering methods may be integrated, i.e., when one clustering algorithm is used as a pre-preprocessing step for another. In addition provide reasoning on why the integration of two methods may lead to improved clustering quality and efficiency.	**4**	**CO4**
1B.	Association rule mining often generates a large number of rules. Discuss one effective method that can be used to reduce the number of rules generated while still preserving most of the interesting rules.	**3**	**CO5**
1C.	Suppose that frequent item sets are saved for a large transaction database, *DB*. Discuss how to efficiently mine the (global) association rules under the same minimum support threshold if a set of new transactions, denoted as ΔDB, is (incrementally) added in.	**3**	**CO5**
2A.	Construct a decision-tree classifier with binary splits at each node, using tuples in relation r (A, B,C) shown below as training data; attribute C denotes the class. Show the final tree, and with each node show the best split for each attribute along with its information gain value. (1, 2, a), (2, 1, a), (2, 5, b), (3, 3, b), (3, 6, b), (4, 5, b), (5, 5, c), (6, 3, b), (6, 7, c)	**3**	**CO5**
2B.	The following Table 2B an example of customer purchase transaction data set. Calculate the support and confidence of the following association rule. Infer whether the items in the association rule are independent of each other or have negative or positive impacts on each other. {10} → {50,70}	**3**	**CO5**

Table 2B

CID	TID	Date	Items Purchased
1	1	01/01/2001	10, 20
1	2	01/02/2001	10, 30, 50, 70
1	3	01/03/2001	10, 20, 30, 40

		2	4	01/03/2001	20, 30		
		2	5	01/04/2001	20, 40, 70		
		3	6	01/04/2001	10, 30, 60, 70		
		3	7	01/05/2001	10, 50, 70		
		4	8	01/05/2001	10, 20, 30		
		4	9	01/06/2001	20, 40, 60		
		5	10	01/11/2001	10, 20, 30, 60		
	Note: CID = Customer ID and TID = Transactions ID						
2C.	Let the following relational schemas be given: $R = (A, B, C)\,; S = (D, E, F)$ Let relations $r(R)$ and $s(S)$ be given. Develop an expression in the tuple relational calculus that is equivalent to each of the following: a) $\Pi_A(r)$ b) $\sigma_{B=17}(r)$ c) $\Pi_{A,F}(\sigma_{C=D}(r \times s)$					**4**	**CO2**
3A.	Consider the relational schema of Figure 3A. *employee* (*person_name*, *street*, *city*) *works* (*person_name*, *company_name*, *salary*) *company* (*company_name*, *city*) *manages* (*person_name*, *manager_name*) **Figure 3A** Develop an expression in the relational algebra to express the following query; a) List the names of all employees who do not work for "First Bank Corporation". b) List the names of all employees managed by Mark Zuckerberg.					**6**	**CO2**
3B.	Consider following Table 3B. **Table 3B** *ID* | *name* | *salary* | *dept_name* | *building* | *budget* 22222 Einstein 95000 Physics Watson 70000 12121 Wu 90000 Finance Painter 120000 32343 El Said 60000 History Painter 50000 45565 Katz 75000 Comp. Sci. Taylor 100000 98345 Kim 80000 Elec. Eng. Taylor 85000 76766 Crick 72000 Biology Watson 90000 10101 Srinivasan 65000 Comp. Sci. Taylor 100000 58583 Califieri 62000 History Painter 50000 83821 Brandt 92000 Comp. Sci. Taylor 100000 15151 Mozart 40000 Music Packard 80000 33456 Gold 87000 Physics Watson 70000 76543 Singh 80000 Finance Painter 120000 Explain two problems caused by the design given in the table. Suggest a normalization technique to resolve the anomaly in the table.					**4**	**CO1**
4A.	Explain the distinctions among the terms primary key, candidate key and super key.					**3**	**CO3**
4B.	Design a database for an airline company. The database must keep track of customers and their reservations, flights and their status, seat assignments on individual flights, and the schedule and routing of future flights. Your design should					**7**	**CO3**

	include an E-R diagram, a set of relational schemas, and a list of constraints, including primary-key and foreign-key constraints.		
5A.	Compute the closure of the following set of functional dependencies for the relational schema $r(A, B, C, D, E)$. $A \rightarrow BC$ $CD \rightarrow E$ $B \rightarrow D$ $E \rightarrow A$ List the candidate keys for R.	**3**	**CO3**
5B.	The definition of a schedule assumes that operations can be totally ordered by time. Consider a database system that runs on a system with multiple processors, where it is not always possible to establish an exact ordering between operations that executed on different processors. However, operations on a data item can be totally ordered. Does the above situation cause any problem for the definition of conflict serializability? Explain your answer.	**3**	**CO3**
5C.	Consider the bank database of Figure 5C where the primary keys are underlined and the following query holds. *branch(branch_name, branch_city, assets)* *customer (customer_name, customer_street, customer_city)* *loan (loan_number, branch_name, amount)* *borrower (customer_name, loan_number)* *account (account_number, branch_name, balance)* *depositor (customer_name, account_number)* **Figure 5C** **select** *T.branch_name* **from** *branch T, branch S* **where** *T.assest > S. assests* and *S. branch_city = 'Brooklyn'* Develop an efficient relational algebra expression that is equivalent to this query. Justify your choice.	**4**	**CO3**