Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY

MANIPAL

A Constituent Institution of Manipal University

V SEMESTER B.TECH. (MECHATRONICS ENGINEERING)

END SEMESTER EXAMINATIONS, NOV 2018

SUBJECT: DATABASE MANAGEMENT SYSTEMS [MTE 4011]

REVISED CREDIT SYSTEM
(27/11/2017)

Time: 3 Hours MAX. MARKS: 50

Instructions to Candidates:

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

1A.	Describe at least 3 tables that might be used to store information for a social networking system such as Facebook.	**4**
1B.	**i.** The following relational schema belongs to which normal form? Convert it into the next normal form. Professor_id → Professor_name Student_rollno→Student_name (Professor_id, Student_rollno→Subject_marks **ii.** With a suitable example differentiate between the Third normal form (3NF) and Boyce Codd normal form (BCNF).	**6**
2A.	Consider the following relations: Student (ID, name, dept_name, tot_cred) Advisor (Student_id, Instructor_id) What is the result of first performing the cross product of the Student and Advisor relations and then performing a selection operation on the result with the predicate Student_id = ID. Represent this using the notations of Relational algebra. Also show the output with some example data.	**2**
2B.	Consider the E-R diagram in the figure below for Q 2B., which models an online bookstore. **i.** List the entity sets and their primary keys. **ii.** Suppose the bookstore adds Blu-ray discs and downloadable video to its collection. The same item may be present in one or both formats, with differing prices. Extend the E-R diagram to model this addition, ignoring the effect on shopping baskets. **iii.** Now extend the E-R diagram, using generalization, to model the case where a shopping basket may contain any combination of books, Blu-ray discs, or downloadable video.	**4**
2C.	Design a database for an automobile company to provide to its dealers to assist them in maintaining customer records and dealer inventory and to assist sales staff in ordering cars. Each vehicle is identified by a vehicle identification number (VIN). Each individual vehicle is a particular model of a particular brand offered by the company (e.g., the XF is a model of the car brand Jaguar of Tata Motors). Each model can be offered with a variety of options, but an individual car may have only some (or none) of the available options. The database needs to store information about models, brands, and options, as well as information about individual dealers, customers, and cars. Your design should include an E-R diagram, a set of relational schemas, and a list of constraints, including primary-key and foreign-key constraints.	**4**

3A.	Consider the insurance database of the figure for Q 3A., where the primary keys are underlined. Construct the following SQL queries for this relational database. **i.** Find the total number of people who owned cars that were involved in accidents in 2009. **ii.** Add a new accident to the database; assume any values for required attributes. **iii.** Delete the Toyota belonging to "John Smith".	**3**
3B.	Suppose that we have a relation *marks* (ID, *score*) and we wish to assign grades to students based on the score as follows: grade *F* if *score* < 40, grade *C* if 40 ≤ *score* < 60, grade *B* if 60 ≤ *score* < 80, and grade *A* if 80 ≤ *score*. Write SQL queries to do the following: **i.** Display the grade for each student, based on the *marks* relation. **ii.** Find the number of students with each grade.	**3**
3C.	Consider the SQL query: **select distinct** *p*.*a*1 **from** *p*, *r*1, *r*2 **where** *p*.*a*1 = *r*1.*a*1 **or** *p*.*a*1 = *r*2.*a*1 Under what conditions does the preceding query select values of *p*.*a*1 that are either in *r*1 or in *r*2? Examine carefully the cases where one of *r*1 or *r*2 may be empty.	**4**
4A.	Give an example of a serializable schedule with two transactions such that the order in which the transactions commit is different from the serialization order.	**4**
4B.	The following contingency table summarizes supermarket transaction data, where *hot dogs* refers to the transactions containing hot dogs, $\overline{hot\ dogs}$ refers to the transactions that do not contain hot dogs, *hamburgers* refers to the transactions containing hamburgers, and $\overline{hamburgers}$ refers to the transactions that do not contain hamburgers. (see table below) **i.** Suppose that the association rule $hotdogs \rightarrow hamburgers$ is mined. Given a minimum support threshold of 25% and a minimum confidence threshold of 50%, is this association rule strong? **ii.** Based on the given data, is the purchase of *hot dogs* independent of the purchase of *hamburgers*? If not, what kind of *correlation* relationship exists between the two?	**4**
4C.	Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8): **i.** Compute the *Euclidean distance* between the two objects. **ii.** Compute the *Manhattan distance* between the two objects.	**2**
5A.	Suppose that the data mining task is to cluster the following eight points (with (*x*, *y*) representing location) into three clusters: *A*1(2, 10), *A*2(2, 5), *A*3(8, 4), *B*1(5, 8), *B*2(7, 5), *B*3(6, 4), *C*1(1, 2), *C*2(4, 9): The distance function is Euclidean distance. Suppose initially we assign *A*1, *B*1, and *C*1 as the center of each cluster, respectively. Use the *k-means* algorithm to show *only* **i.** The three cluster centers after the first round execution **ii.** The final three clusters	**5**
5B.	A research article entitled, "Application of improved K-medoids algorithm in charging station planning for mobile robot" authored by Q. Yuan *et. al.* has been published in 2017 at the 7th IEEE International Conference on Electronics Information and Emergency Communication (ICEIEC) held at Macau China. The abstract of the article is as stated below. *In order to solve the problem of battery replacement when the battery is consumed to a certain extent by mobile robot working in outdoor environment, the clustering analysis is introduced to solve the problem of charging station planning for mobile robot. In this paper, combined with motion energy model, the criterion that similarity between two nodes is measured according to the distance in the traditional K-medoids is adjusted.* ***Aiming at the limitation of random selection of the initial clustering centers, an improved K-medoids clustering algorithm is formed****, finally obtaining a reasonable number and location of the charging stations, achieving a good balance between less motion energy consumption of mobile robot returning the charging station and the economic benefits of charging station construction. The experiment shows that*	**5**

Table for 4B:

	hot dogs	$\overline{hot\ dogs}$	Σ_{row}
hamburgers	2,000	500	2,500
$\overline{hamburgers}$	1,000	1,500	2,500
Σ_{col}	3,000	2,000	5,000

	application of improved K-medoids algorithm in charging station planning for mobile robot is effective and practical. With a sample dataset, show as to how choosing a random centroid in the K-medoids algorithm, could be a limitation. (**Hint:** Justify your answer by using the changing value of the Absolute Error Criterion when shifting the centroid of a cluster from one point to another.)	

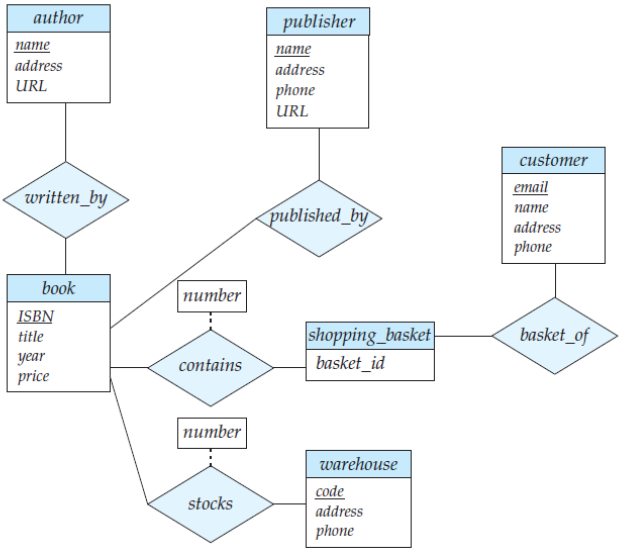

Figure for Q 2B

person (*driver_id*, *name*, *address*)
car (*license*, *model*, *year*)
accident (*report_number*, *date*, *location*)
owns (*driver_id*, *license*)
participated (*report_number*, *license*, *driver_id*, *damage_amount*)

Figure for Q 3A