

(Time: 2½ Hours)

[Total Marks: 75]

- N. B.:
- (1) All questions are compulsory.
 - (2) Make suitable assumptions wherever necessary and state the assumptions made.
 - (3) Answers to the same question must be written together.
 - (4) Numbers to the right indicate marks.
 - (5) Draw neat labelled diagrams wherever necessary.
 - (6) Use of Non-programmable calculators is allowed.

1. Attempt any three of the following:

- a. Suppose you want to build a video site similar to YouTube and keep data in file-processing system. Discuss the relevance of each of the following points to the storage of actual video data, and to metadata about the video, such as title, the user who uploaded it, tags, and which users viewed it.
 - i. Data redundancy and inconsistency
 - ii. Difficulty in accessing data
 - iii. Data isolation
 - iv. Integrity problems
 - v. Atomicity problems
 - vi. Concurrent system anomalies
 - vii. Security problems
- b. State the advantages and disadvantages of the following data models: Hierarchical, Network, Relational, Entity Relationship, Object Oriented and NoSQL. State if the models support data and structural independence.
- c. State and explain the twelve Codd's rules for relational databases.
- d. What is Unified modelling language? What are its parts? Show the ER diagram notations and equivalent notations in UML.
- e. Construct an E-R diagram for a car insurance company whose customers own one or more cars each. Each car has associated with it zero to any number of recorded accidents. Each insurance policy covers one or more cars, and has one or more premium payments associated with it. Each payment is for a particular period of time, and has an associated due date, and the date when the payment was received.
- f.
 - i. Design an E-R diagram for keeping track of the exploits of your favourite sports team. You should store the matches played, the scores in each match, the players in each match, and individual player statistics for each match. Summary statistics should be modelled as derived attributes.
 - ii. Consider an E-R diagram in which the same entity set appears several times, with its attributes repeated in more than one occurrence. Why is allowing this redundancy a bad practice that one should avoid?

2. Attempt any three of the following:

- a. The natural outer-join operations extend the natural-join operation so that tuples from the participating relations are not lost in the result of the join. Describe how the theta join operation can be extended so that tuples from the left, right, or both relations are not lost from the result of a theta join.

[TURN OVER]

Q.P. Code: 20943

- b. Given the following relational schemas: $R = (A, B, C)$ $S = (D, E, F)$
 Suppose the relations $r(R)$ and $s(S)$ are defined. Write the expressions in tuple relational calculus equivalent to each of the following:

i. $\Pi_A(r)$ ii. $\sigma_{B=17}(r)$ iii. $r \times s$ iv. $\Pi_{A,F}(\sigma_{C=D}(r \times s))$

- c. Consider the relational database below, where primary keys are underlined.

employee (person name, street, city)

works (person name, company name, salary)

company (company name, city)

manages (person name, manager name)

Give an expression in tuple relational calculus for each of the following queries:

- i. Find all employees who work directly for "Jones."
 ii. Find all cities of residence of all employees who work directly for "Jones."
 iii. Find the name of the manager of the manager of "Jones."
 iv. Find those employees who earn more than all employees living in the city "Mumbai."
 d. What is normalization? What is its objective? Give a distinguishing characteristic of 1NF, 2NF, 3NF, 4NF and BCNF.

- e. i. Using the INVOICE table structure shown in table below, write the relational schema, draw its dependency diagram and identify all dependencies (including all partial and transitive dependencies). You can assume that the table does not contain repeating groups and that any invoice number may reference more than one product. (Hint: This table uses a composite primary key.)

Attribute Name	Sample Value	Sample Value	Sample Value	Sample Value	Sample Value
INV_NUM	211347	211347	211347	211348	211349
PROD_NUM	AA-E3422QW	QD-300932X	RU-995748G	AA-E3422QW	GH-778345P
SALE_DATE	15-Jan-2016	15-Jan-2016	15-Jan-2016	15-Jan-2016	16-Jan-2016
PROD_LABEL	Rotary sander	0.25-in. drill bit	Band saw	Rotary sander	Power drill
VEND_CODE	211	211	309	211	157
VEND_NAME	NeverFail, Inc.	NeverFail, Inc.	BeGood, Inc.	NeverFail, Inc.	ToughGo, Inc.
QUANT_SOLD	1	8	1	2	1
PROD_PRICE	₹4995	₹345	₹3999	₹4995	₹8775

- ii. Using the initial dependency diagram drawn in question i, remove all partial dependencies, draw the new dependency diagrams, and identify the normal forms for each table structure you created.
 iii. Using the table structures you created in question ii, remove all transitive dependencies and draw the new dependency diagrams. Also identify the normal forms for each table structure you created.
 f. Explain the phases of database design.

3. Attempt any three of the following:

- a. What are constraints? What are the different types of constraints? Explain.
 b. What is a view? What are its advantages?
 c. State the rules for performing DML operations on a view.
 d. Explain GROUP BY and ORDER BY clauses with examples.
 e. What are NULL values? Explain.

[TURN OVER]

- f. What are joins? What are different types of joins? Explain.
4. Attempt **any three** of the following:
- a. The lost update anomaly is said to occur if a transaction T_j reads a data item, then another transaction T_k writes the data item (possibly based on a previous read), after which T_j writes the data item. The update performed by T_k has been lost, since the update done by T_j ignored the value written by T_k .
- Give an example of a schedule showing the lost update anomaly.
 - Give an example schedule to show that the lost update anomaly is possible with the read committed isolation level.
 - Explain why the lost update anomaly is not possible with the repeatable read isolation level.
- b. State and explain the ACID properties of transactions.
- c. Consider a database for a bank where the database system uses snapshot isolation. Describe a particular scenario in which a nonserializable execution occurs that would present a problem for the bank.
- Consider a database for an airline where the database system uses snapshot isolation. Describe a particular scenario in which a nonserializable execution occurs, but the airline may be willing to accept it in order to gain better overall performance.
- d. Show that the two-phase locking protocol ensures conflict serializability, and that transactions can be serialized according to their lock points.
- e. Consider the following two transactions:
- T_{34} :
 read(A);
 read(B);
 if $A = 0$ then $B := B + 1$;
 write(B).
- T_{35} :
 read(B);
 read(A);
 if $B = 0$ then $A := A + 1$;
 write(A).
- Add lock and unlock instructions to transactions T_{34} and T_{35} , so that they observe the two-phase locking protocol. Can the execution of these transactions result in a deadlock?
- f. Explain the different ways to handle deadlocks.

5. Attempt **any three** of the following:
- What are triggers? What are different types of triggers? How are they created? Give the syntax and examples of the same.
 - What are packages? What are the components of packages? How are packages developed? Explain with syntax and example.
 - What are functions? What are procedures? How do they differ from each other? What are the benefits of stored procedures and functions?
 - What is a cursor? Explain implicit and explicit cursors. How are explicit cursors controlled?
 - What are hierarchical queries? Explain the syntax of hierarchical queries.
 - What are composite data types? Explain the PL/SQL records. How is a PL/SQL record created?