Answer any three out of the following six questions

1. Answer the following three parts.

   a. Define the concepts of entity integrity and referential integrity. Illustrate your definitions using an example database schema having a STUDENT relation schema and a DEPARTMENT relation schema.

   [11 marks]

   b. Give an incremental algorithm for testing referential integrity of a database d having two relations r and s over relation schemas, R and S, respectively, where X $\subseteq$ schema(R) is a foreign key referencing the primary key Y $\subseteq$ schema(S).

   (Hint: You should assume that the database d satisfies referential integrity prior to the insertion of a new tuple t into r.)

   [11 marks]

   c. Give an efficient (polynomial time) algorithm for finding one key for a relation schema R with respect to a set of functional dependencies F over R.

   [11 marks]

   [Total 33 marks]

2. Answer the following three parts.

   a. Answer the following two parts.

      i. Express the natural join, $r \bowtie s$, of the relations r over R, with schema(R) = $\{$A, B$\}$, and s over S, with schema(S) = $\{$B, C$\}$, using only the relational algebra operators: renaming, projection, selection and Cartesian product.

      [6 marks]

      ii. Show how the Boolean combinations of simple selection formulae, with AND, OR and NOT, can be expressed using the set-theoretic relational algebra operators.

      [5 marks]

      [Total 11 marks]

   b. It is often claimed that the join operator is the bottleneck in relational database query processing. Suggest how queries involving joins may be optimised.

   [11 marks]

   c. Assume that we have a relational data model at the conceptual database level. Argue for and against adding a new internal database level between the conceptual and physical levels, whose database schema is either hierarchical or nested relational.

   [11 marks]

   [Total 33 marks]

3. Answer the following three parts.

   a. Discuss the pros and cons of having a computationally complete database language as a user query language. (Such a language may be similar to Oracle's PL/SQL language.)

   [11 marks]

   b. Answer the following two parts.

      i. Discuss the pros and cons of allowing, general, n-ary relationships as opposed to allowing only binary relationships, in the Entity-Relationship model.

      [6 marks]

      ii. In the Entity-Relationship model we have the separate concepts of entity and relationship, although it is is possible to view a relationship as a special kind of entity. Argue for and against such a view.

      [5 marks]

      [Total 11 marks]

   c. Discuss the view that the Entity-Relationship model supports the fundamental concepts of an object-oriented data model.

   [11 marks]

   [Total 33 marks]

4. a. Give an account of the database recovery arrangements normally provided by a database management system making reference to each of the following terms: transaction, transaction log, before and after images, checkpoints.

[12 marks]

b. Explain the best way that recovery from each of the following situations can be handled :

(i) loss of the communication link while a user is entering a transaction

(ii) damage of a disk pack on which part of the database is stored making it unreadable

(iii) a power failure to the computer room

(iv) noticing that some payment amount was incorrectly credited to a customer saccount record two weeks ago.

[12 marks]

c. Using a clearly annotated example explain how timestamping can be used to prevent loss of data integrity caused by interference between transactions in a multi-user database environment.

[9 marks]

[Total 33 marks]

5. a. Using examples for illustration, show how data redundancy and data anomalies can be overcome by normalising data structures successively through first, second and third normal form. In each case identify an anomaly which is overcome by the further normalisation.

[18 marks]

b. Define the term multi-valued dependency. Give an example of a data structure which contains a multi-valued dependency and show how the data anomalies it can give rise to can be overcome by further normalisation.

[8 marks]

c. Briefly outline the reasons why fully normalised data structures might not form the basis of the relations eventually implemented in a database system.

[7 marks]

[Total 33 marks]

6. a. Discuss the part played by data volume and usage analysis during the process of database design.

[13 marks]

b. Answer each of the following questions concerning file organisation and access methods:

(i) What is the main advantage of a balanced tree index over other types of indexes?

(ii) How does page-splitting in a balanced tree come about? Use an example to show how a page split keeps such a structure balanced.

(iii) The use of hashing for file organisation and access may result in fewer accesses for random record retrieval than would be required using a B+ - tree. Why, therefore, are B+ -trees the most common method for accessing database files?

(iv) Why would you expect to use more indexes for a database primarily used for data retrieval than for one where transactions to perform updating predominate?

(v) What are the benefits to be gained from clustered indexes?

[$4 \times 5 = 20$ marks]

[Total 33 marks]

END OF PAPER