Instructions: Answer any 3 questions. All questions carry equal marks

PTO

CT3532 Database Systems 2

Question 1   (25 marks)

a) Explain the term *minimal cover set*. Given a set of functional dependencies F, outline an algorithm to generate a minimal set of functional dependencies. Illustrate the operation of your algorithm on the answer with the following set **F**.

$$F = \{B \rightarrow CE, D \rightarrow E, BC \rightarrow D, C \rightarrow E\} \qquad (8)$$

b) Given the following relation, R, and functional dependencies, F:

$$R = \{A, B, C, D, E, F, G, H, I, J\}$$

$$F = \{\{A,B,C\} \rightarrow \{D, E, F, G\}, \{A,C\} \rightarrow \{J\},$$
$$\{B\} \rightarrow \{H\}, \{H\} \rightarrow \{I\}, \{J\} \rightarrow \{C\}\}$$

   i)   Outline, in your words, an approach to identifying a suitable key; illustrate that approach on relation R.
   ii) Draw a dependency graph for the set F.
   iii) Decompose the relation to BCNF.

   (11)

c) Explain, with the use of a suitable example, when de-normalisation may be appropriate. With respect to your example, outline the associated advantages and disadvantages.

   (6)

Question 2   (25 marks)

a) Explain the structure of a B+ tree and outline an algorithm for the insertion of items into a B+ tree. Illustrate your algorithm by showing how a B+tree would develop given the following values to insert. You may assume each node in the tree can hold two search values.

$$34, 25, 16, 30, 15, 19, 7, 18$$

(10)

b) Linear hashing and dynamic hashing are two approaches to hash values to a dynamically changing file. Briefly outline either approach and illustrate the approach using the following record key values. You may assume each block can hold two records.

$$34, 25, 16, 15, 19, 9, 30, 24$$ (10)

c) Suggest an indexing strategy to jointly index a number of attributes *a1, a2, .. an,* where expect queries involving selection criteria on a number of these attributes.   (5)

Question 3   (25 marks)

a) Explain what is meant by the *lost update problem* and show with an example schedule of transactions how it may arise.   (5)

b) Explain what is meant by *two phase locking* and show how the schedule in a) would proceed under a two phase locking protocol.   (6)

c) Prove that two phase locking guarantees conflict-serializability.   (8)

d) Outline a general approach to database recovery. Your answer should explain the following concepts: system log, commit point, checkpoint.   (6)

Question 4   (25 marks)

a) Outline an efficient approach to implementing sorting of data to satisfy, for example, the ORDER BY command. You may assume that the quantity of data to sort is too large to fit in main memory. Illustrate your approach with a small example. Comment on the efficiency of the approach.   (8)

b) Parallel architectures have been used to increase the efficiency of many standard operations used in database management systems. Describe how range partitioning operates and discuss any advantages or limitations that exist with this approach.   (9)

c) Explain, with an example, a means to sort the tuples of a relation where the relations has been partitioned across multiple partitions in a parallel database.   (8)