Name : .....

Roll No. : .....

Invigilator's Signature : .....

CS/M.Tech (CSE)/SEM-2/PGCSE-202/2013

2013
ADVANCED AND DISTRIBUTED DBMS

Time Allotted : 3 Hours *Full Marks* : 70

The figures in the margin indicate full marks.
Candidates are required to give their answers in their own words as far as practicable.

Answer any *five* of the following : 5 × 14 = 70

1. What is distributed database ? Explain with a neat diagram. What are the main advantages and disadvantages of distributed databases ? Differentiate between parallel and distributed databases. What are the desired properties of distributed databases ? What are the various types of distributed databases ?
2. What do you mean by architecture of a distributed database system ? What are different types of architectures ? Discuss each of them with neat sketch. What are homogeneous and heterogeneous DDBSs ? Explain in detail with an example.
3. What is a fragment of a relation ? What are the main types of data fragments ? Why is fragmentation a useful concept in distributed database design ? What are horizontal data fragmentation, vertical data fragmentation, mixed data fragmentation ? Explain with examples.

4. consider the following relation :

   EMPLOYEE (EMP, NAME, ADDRESS, SKILL, PROJ-ID)

   EQUIPMENT (EQP-ID, EQP-TYPE, PROJECT)

   Suppose that EMPLOYEE relation is horizontally fragmented by PROJ-ID and each fragment is stored locally at its corresponding project site. Assume that the EQUIPMENT relation is stored in its entirely at the Tokyo location. Describe a good strategy for processing each of the following queries :

   a) Find the join of relations EMPLOYEE and EQUIPMENT.
   b) Get all employees for projects using EQP-TYPE as "Welding machine".
   c) Get all machines being used at the Mumbai Project.
   d) Find all employees of the project using equipment number 110.

5. Write short notes on any *seven* of the following :

   a) Distributed Database
   b) Data Fragmentation
   c) Data Allocation
   d) Data Replication
   e) Two-phase Commit
   f) Three-phase Commit
   g) Timestamping
   h) Distributed Locking
   i) Semi-JOIN
   j) Distributed Deadlock.

6. How do we achieve concurrency control in a distributed database system ? What should be the characteristics of a good concurrency control mechanism ? How do we achieve recovery control in a distributed database system ? What is distributed locking ? What are its advantages and disadvantages ? Explain the function of two-phase and three-phase commit protocols used in recovery control of distributed database system.

7. Consider a relation that is fragmented horizontally by PLANT-NO and given as EMPLOYEE (NAME, ADDRESS, SALARY, PLANT-NO).

   Assume that each fragment has two replicas; one stored at the Bangalore site and one stored locally at the plant site of Jamshedpur. Describe a good processing strategy for the following queries entered at the Singapore site :

   a) Find all employees at the Jamshedpur plant.
   b) Find the average salary of all employees.
   c) Find the highest-paid employee at each of the plant site namely Thailand, Mumbai, New Delhi and Chennai.
   d) Find the lowest-paid employee in the company.