DISTRIBUTED DATABASES
(CSEN 4241)

Time Allotted : 3 hrs **Full Marks : 70**

Figures out of the right margin indicate full marks.

Candidates are required to answer Group A and any 5 (five) from Group B to E, taking at least one from each group.

Candidates are required to give answer in their own words as far as practicable.

Group – A
(Multiple Choice Type Questions)

1. Choose the correct alternative for the following: **10 × 1 = 10**

(i) Preservation of functional dependency is ensured by which of the correctness rule of the fragmentation?
(a) Disjointness (b) Reconstruction
(c) Completeness (d) All of (a), (b) & (c).

(ii) When some of the columns of a relation are at different sites, which of the following does it refer to?
(a) Data Replication (b) Horizontal fragmentation
(c) Vertical fragmentation (d) Both (a) and (c).

(iii) In order to represent run-time simplification in operator trees, the operator that is used is
(a) Join (b) Union
(c) Cut (d) Intersection

(iv) A physical image of global relation R at site j constitutes
(a) All the fragments of one or more relations allocated at site j
(b) All the fragments of relation R while R is located at site j
(c) All the fragments corresponding to R and located at the same site j
(d) None of the above.

(v) Two-phase commitment protocol is used for
(a) concurrency control (b) integrity control
(c) recovery (d) redundancy

(vi) Join graph is used in
(a) primary horizontal fragmentation (b) vertical fragmentation
(c) derived fragmentation (d) all of (a), (b) & (c).

(vii) Parallelism in which the relations are partitioned on multiple disks to reduce the retrieval time of relations from disk is called as
(a) I/O parallelism (b) Inter-operation parallelism
(c) Intra-query parallelism (d) Inter-query parallelism.

(viii) Which of the following failures are typical to distributed database systems?
(a) Failure of a site (b) Loss of messages
(c) Network Partition (d) All of (a), (b) & (c).

(ix) The type of mapping defined in the allocation schema is
(a) One-to-many (b) One-to-one
(c) Many-to-many (d) Many-to-one.

(x) Commit and rollback are related to _____.
(a) data integrity (b) data consistency
(c) data sharing (d) data security

Group - B

2. (a) Compare the role of centralized control in traditional and distributed databases. *[(CO1)(Analyze/IOCQ)]*
(b) What are the parameters that characterize the basic functions of a computer network? *[(CO1)(Remember/LOCQ)]*
(c) Describe briefly the functions of the Transport layer of the ISO/OSI reference architecture. *[(CO1)(Remember/LOCQ)]*

4 + 6 + 2 = 12

3. Consider the following global schema:

FACULTY (emp_id, eName, dept)

(i) Assume FACULTY has a fragmentation by Dept and 'CSE', 'ECE' and 'IT' are the only possible values of Dept. Define the fragmentation.
(ii) Assume the allocation of the fragments as follows:
Allocation schema: $FACULTY_1$ at sites 1, 2
$FACULTY_2$ at sites 3
$FACULTY_3$ at sites 3, 4
Write an application that receives the emp_id from the user and outputs the name and dept at levels 1 and 2 of transparency.
(iii) Write an application that transfers the faculty having emp_id=2120 from department 'ECE' to 'CSE', at levels 1, 2 and 3 of transparency.

[(CO2)(Design/HOCQ)]

[2 + (2 × 2) + (3 × 2)] = 12

Group - C

4. (a) What is horizontal and vertical fragmentation? What are the types of horizontal fragmentation. *[(CO2)(Remember/LOCQ)]*
(b) Perform horizontal Fragmentation for student relation as given below. Also give the correctness criteria for it.

Students (studentrollno., Student Name, Course Name, Course Name, Course fees, year) *[(CO2)(Analyze/LOCQ)]*

(2 + 2) + 8 = 12

5. (a) Describe briefly the two methods used for determining redundant allocation of fragments. What do you understand by Distributed Join? *[(CO3)(Understand/LOCQ)]*

   (b) Write the equivalent expressions in relational algebra:

   (i) Select AVG(QTY)
   From SUPPLY
   Where PNUM = "P001"

   (ii) Select PNUM, SNUM, SUM(QTY)
   From SUPPLY
   Group by PNUM, SNUM

   (iii) Select PNUM, SNUM, SUM(QTY)
   From SUPPLY
   Group by PNUM, SNUM
   Having SUM(QTY) > 100. *[(CO2)(Understand/IOCQ)]*

(3 + 2) + (2 + 2 + 3) = 12

Group - D

6. (a) Consider relation EMP(ENo, EName, Title). Let EMP be fragmented as follows:

   $EMP1 = \sigma_{ENo"E3"}(EMP)$
   $EMP2 = \sigma_{"E3"<ENo\leq"E6"}(EMP)$
   $EMP3 = \sigma_{ENo>"E6"}(EMP)$

   Query: "Find all employees whose employee number is E5"
   Write the SQL and draw the operator graphs for the generic query as well as for the reduced query. *[(CO3)(Understand/IOCQ)]*

   (b) Explain in brief the three schemes used to detect conflicts during lock-based concurrency control in Distributed Database. *[(CO4)(Remember/LOCQ)]*

(2 + 2 + 2) + 6 = 12

7. (a) Briefly mention the steps of transforming the centralized 2PC protocol in a distributed commit protocol. *[(CO4)(Understand/LOCQ)]*

   (b) What is hierarchical deadlock detector? *[(CO4)(Understand/LOCQ)]*

   (c) What is the difference between centralized and distributed deadlock detection? *[(CO4)(Remember/LOCQ)]*

4 + 4 + 4 = 12

Group - E

8. (a) What are Catalogs used for in Distributed database? How are the allocated in the distributed database? *[(CO2)(Understand/LOCQ)]*

   (b) What do you understand by parallel databases? What are their basic differences with a distributed database? *[(CO1)(Understand/LOCQ)]*

(3 + 3) + (3 + 3) = 12

9. Consider the following fragments:
   P1 : $\sigma_{\text{Branch\_no"B003" } \wedge \text{ type = "House"}}$ (PROPERTY)
   P2 : $\sigma_{\text{Branch\_no"B003" } \wedge \text{ type = "Flat"}}$ (PROPERTY)
   P3 : $\sigma_{\text{Branch\_no!="B003"}}$ (PROPERTY)
   B1 : $\sigma_{\text{Branch\_no="B003"}}$ (BRANCH)
   B2 : $\sigma_{\text{Branch\_no!="B003"}}$ (BRANCH)
   Optimize the following query:
   Select *
   From BRANCH b, PROPERTY p
   Where b.BranchNo. = p.BranchNo.
   And p.type = "Flat". *[(CO5)(Analyze/HOCQ)]* **12**

Cognition Level	*LOCQ*	*IOCQ*	*HOCQ*
Percentage distribution	*57.29*	*17.71*	*25*

Course Outcome (CO):

After the completion of the course students will be able to
CO1. Understand the basic concepts of database, communication network, and distributed database.
CO2. Identify the concepts of creating and maintaining databases in a distributed environment.
CO3. Learn to design a distributed database using horizontal and vertical fragmentation.
CO4. Learn to manage distributed transactions and concurrency control.
CO5. Design all types of distributed queries using query optimization techniques.

**LOCQ: Lower Order Cognitive Question; IOCQ: Intermediate Order Cognitive Question; HOCQ: Higher Order Cognitive Question.*