

**DISTRIBUTED DATABASES**  
**(CSEN 3231)**

**Time Allotted: 2½ hrs**

**Full Marks: 60**

*Figures out of the right margin indicate full marks.*

*Candidates are required to answer Group A and any 4 (four) from Group B to E, taking one from each group.*

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

**Group - A**

1. Answer any twelve:

**12 × 1 = 12**

*Choose the correct alternative for the following*

- (i) A distributed database has which of the following advantages over a centralized database?  
(a) Software complexity (b) Slow Response  
(c) Modular growth (d) Software cost.
- (ii) A transaction manager is which of the following?  
(a) Maintains Sql Queries (b) Maintain Schema  
(c) Maintains a log of transactions (d) Maintains database.
- (iii) Some of the columns of a relation are at different sites is which of the following?  
(a) Data Replication (b) Horizontal Partitioning  
(c) Horizontal and Vertical Partitioning (d) Vertical Partitioning.
- (iv) Storing a separate copy of the database at multiple locations is which of the following?  
(a) Horizontal and Vertical Partitioning (b) Data Replication  
(c) Horizontal Partitioning (d) Vertical Partitioning.
- (v) Which of the following is a disadvantage of replication?  
(a) Each transaction may proceed without coordination across the network  
(b) Each site must have the same storage capacity  
(c) If the database fails at one site, a copy can be located at another site  
(d) Reduced network traffic.
- (vi) A distributed transaction can be \_\_\_\_\_ if queries are issued at one or more nodes  
(a) fully read-write (b) partially read-write  
(c) partially read-only (d) fully read-only
- (vii) What is the mechanism which guarantees that all database servers participating in a distributed transaction either all commit or all roll back the statements in the transaction?  
(a) Commit transaction (b) Commit  
(c) Three-phase commit (d) Two-phase commit.

- (viii) Which of the following is true concerning a global transaction?  
 (a) The required data are at one local site and the distributed DBMS routes requests as necessary  
 (b) The required data are located in at least one nonlocal site and the distributed DBMS routes requests as necessary  
 (c) The required data are at one local site and the distributed DBMS passes the request to only the local DBMS  
 (d) The required data are located in at least one nonlocal site and the distributed DBMS passes the request to only the local DBMS.
- (ix) Global Wait-for graph is used for \_\_\_\_\_ in Distributed database.  
 (a) Handling concurrency control (b) Handling deadlock  
 (c) Handling failures (d) None of the above
- (x) The procedure of choosing a suitable query out of all the queries is classified as  
 (a) Query Optimization (b) Parser Optimization  
 (c) Processing optimization (d) All of these.

*Fill in the blanks with the correct word*

- (xi) The \_\_\_\_\_ layer of ISO/OSI model provides services to the user.  
 (xii) Storing a separate copy of the database at multiple locations is \_\_\_\_\_.  
 (xiii) Database profiles for DDBs are kept in \_\_\_\_\_.  
 (xiv) In \_\_\_\_\_ phase new lock(s) are acquired.  
 (xv) The \_\_\_\_\_ is responsible for ensuring correct execution in the presence of failures.

### Group - B

2. (a) What do you understand by a Distributed database? What are its two important aspects? [[CO1](Understand/LOCQ)]  
 (b) What were the main motivations behind the development of distributed databases? [[CO1](Understand/LOCQ)]  
**(2 + 2) + 8 = 12**
3. (a) Consider the following:  
 Global schema : Guest (G\_id, name, block\_id, room\_no)  
 Fragmentation schema :  
 $F_1 = \sigma_{\text{block\_id}=\text{'East'}}(\text{Guest})$   
 $F_2 = \sigma_{\text{block\_id}=\text{'West'}}(\text{Guest})$   
 Allocation schema: F1 at Site 2 and F2 at Site 1 and 3.  
 Write an application that accepts G\_id from the user and outputs the guest name at Level 1, 2 and 3 of transparency? [[CO2](Design/HOCQ)]
- (b) Write an application that shifts the guest having G\_id=100 from block\_id = 'West' to 'East', at levels 1 and 3 of transparency. [[CO2](Design/HOCQ)]  
**(2 + 2 + 3) + (2 + 3) = 12**

## Group - C

4. (a) Describe briefly the two methods used for determining redundant allocation of fragments. [[CO4](Understand/LOCQ)]
- (b) What do you understand by Distributed join? How is it represented using a graph? [[CO4](Remember/LOCQ)]
- (c) Consider the following global relation  
*Emp(eno, ename, city, sal)*  
 Write a suitable parametric query to find the name of employees who earns a salary greater than Rs. 40000 and work in the city \$X. [[CO5](Apply/IOCQ)]  
**4 + 4 + 4 = 12**
5. (a) Consider the following global relations:  
*Emp(eno, ename, city, sal)*  
*Assign(eno, pno, duration)*  
*Proj(pno, pname, budget)*
- (i) Write an expression using relational algebra to find the salaries (sal) of those employees working on a “DDBMS” project other than Mr. John.
- (ii) Draw the corresponding operator tree for the above query.
- (iii) Is the operator tree optimized? If not, transform it into the optimized tree, with proper explanations. [[CO5](Implement/IOCQ)]
- (b) What do you understand by Replication transparency? [[CO3](Understand/LOCQ)]  
**(2 + 3 + 4) + 3 = 12**

## Group - D

6. (a) Given is the following algebraic expression. Draw the equivalent operator tree and apply simplification rules, if applicable, to get an optimized tree.  
 $(\mathbf{SL}_{DEPTNUM=10} \mathbf{DEPT} \mathbf{N} \mathbf{J} \mathbf{N} (\mathbf{SL}_{PNUM="P1"} \mathbf{SUPPLY} \mathbf{D} \mathbf{F} \mathbf{SL}_{PNUM="P2"} \mathbf{SUPPLY}))$   
 $\mathbf{UN} (\mathbf{SL}_{DEPTNUM=10} \mathbf{DEPT} \mathbf{N} \mathbf{J} \mathbf{N} \mathbf{SL}_{PNUM="P1"} \mathbf{SUPPLY})$   
 Where the global schema is as follows:  
 DEPT = (DEPTNUM, NAME, AREA, MGRNUM)  
 SUPPLY = (SNUM, PNUM, DEPTNUM, QUAN). [[CO3](Design/HOCQ)]
- (b) What are parametric queries? Why are temporaries used in parametric queries? Explain your answer with suitable example(s). [[CO4](Design/HOCQ)]  
**6 + (3 + 3) = 12**
7. (a) What are the problems that need to be addressed during query optimization? [[CO3](Design/HOCQ)]
- (b) Consider a given relation *Student* as follows:
- | Attribute | Size | No. Of distinct values |
|-----------|------|------------------------|
| Roll      | 5    | 120                    |
| Name      | 12   | 100                    |
| Dept      | 3    | 12                     |
| Phn       | 10   | 120                    |
- There are 120 tuples in the relation *Student* and Roll is the primary key.  
 (i) Draw the database profile for *Student*.

(ii) Suppose there are two horizontal fragments of *Student*, namely *Student1* (allocated at site 1) and *Student2* (allocated at site 3).

If *Student1* has 50 tuples and the number of distinct values of the attributes in the fragments is given as follows, draw the database profiles for the fragments.

Attributes	<i>Student1</i>	<i>Student2</i>
Name	42	58
Dept	8	10
Phn	50	70

[[CO3](Design/HOCQ)]

**3 + (3 + 6) = 12**

### Group - E

8. (a) Explain clearly how the 2PC protocol will behave in the following kind of failures:

(i) The coordinator fails after writing a 'global-commit' or 'global-abort' record but before writing the 'complete' record in the log.

(ii) A participant fails before writing the 'ready' record in the log.

[[CO4](Analyze/IOCQ)]

(b) Describe briefly the 'Write-lock-all, read-locks-one' locking and 'Primary Copy' locking schemes used for concurrency control in DDB.

[[CO4](Remember/LOCQ)]

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

9. (a) Describe in brief the contents of distributed database catalogs. [[CO4](Learn/IOCQ)]

(b) What are the two ways in which authorization rules are allocated and enforced in distributed database systems?

[[CO4](Learn/IOCQ)]

**6 + 6 = 12**

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution			

#### Course Outcome (CO):

After the completion of the course students will be able to

CSEN3231.1. Understand the basic concepts of database, communication network, and distributed database.

CSEN3231.2. Identify the concepts of creating and maintaining databases in a distributed environment Understand and use the various core module

CSEN3231.3. Learn to design a distributed database using horizontal and vertical fragmentation.

CSEN3231.4. Learn to manage distributed transactions and concurrency control.

CSEN3231.5. 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.