

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 can use which of the following strategies?
 - (a) Totally centralized at one location and accessed by many sites
 - (b) Partially or totally replicated across sites
 - (c) Partitioned into segments at different sites
 - (d) All of the above
- (ii) Which network topology requires a central controller or hub?
 - (a) Star
 - (b) Mesh
 - (c) Ring
 - (d) Bus
- (iii) “Users view each data object as logically unique.” Which property of DDBMS supports this given statement?
 - (a) Fragmentation Transparency
 - (b) Replication Transparency
 - (c) Location Transparency
 - (d) Allocation Transparency
- (iv) Which of the statements with respect to horizontal fragmentation is true in a distributed database?
 - Statement 1: A tuple should be allocated to atleast one fragment
 - Statement 2: A tuple should be allocated to atleast one fragment
 - Statement 3: A tuple should be allocated to all the fragments
 - (a) Only statement 1
 - (b) Only statement 2
 - (c) Both Statements 1 and 2
 - (d) Only statement 3
- (v) 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
- (vi) 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 the above

- (vii) 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
- (viii) Distributed join means
 (a) Relations are first collected and then joined
 (b) Relations are first joined and then collected
 (c) Fragments are first collected and then joined
 (d) Fragments are first joined and then collected
- (ix) In some cases of query processing, the operations in a single query are executed at the same time or in random order. What type of parallelism is this?
 (a) Intra-operation parallelism (b) Pipelined parallelism
 (c) Independent parallelism (d) Inter-operation parallelism
- (x) 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 corresponding to R and located at the same site j
 (c) All the fragments of relation R while R is located at site j
 (d) None of the above.

Fill in the blanks with the correct word

- (xi) _____ operation is used to reconstruct the global relation from horizontal fragments.
- (xii) _____ property will check whether all the operation of a transaction are completed or none.
- (xiii) The relational algebraic command used to extract some records from a relation is _____.
- (xiv) Join graph is used in _____ fragmentation.
- (xv) During the shrinking phase of Two Phase Locking the 'locks' are _____.

Group - B

2. (a) Consider a given relation EMP(empno, name, desig, salary, branchno). The EMP table is fragmented as follows:

Fragment No	Fragmentation	Location
F ₁	$\Pi_{empno, desig, salary} (EMP)$	Site 5
F ₂	$\Pi_{empno, name, branchno} (EMP)$	
F ₂₁	$\sigma_{branchno='B003'} (F_2)$	Site 2
F ₂₂	$\sigma_{branchno='B005'} (F_2)$	Site 3
F ₂₃	$\sigma_{branchno='B007'} (F_2)$	Site 5

- (i) What type of fragmentation is done on EMP?
 (ii) Draw the corresponding fragmentation tree.
 (iii) Write the queries to find the names of the employees for a branch number (branchno) given by the user, at levels 1, 2 and 3 of transparency.

[[CO3](Design/HOCQ)]

- (b) What do you understand by physical image of a global relation?
 [[CO2](Remember/LOCQ)]
(1 + 2 + 7) + 2 = 12
3. (a) What do you understand by a Distributed database? What are its two important aspects?
 [[CO1](Understand/LOCQ)]
 (b) State the features of distributed database that makes it different from centralized database.
 [[CO1](Remember/LOCQ)]
 (c) Describe briefly the functions of the Transport layer of the ISO/OSI reference architecture.
 [[CO1](Understand/IOCQ)]
(2 + 2) + 4 + 4 = 12

Group - C

4. (a) Draw the equivalent operator tree and apply simplification rules, if applicable, to get an optimized tree for the following query:
 Find the scores of those students assigned to the “DDBMS” course except the student named “XYZ”. The global schema is as follows:
Student(roll, sname, city, score)
Assgn(roll, cno, duration)
Course(cno, cname, credit)
 Assume that the *Student* table is horizontally fragmented into students with $roll < 100$ and $roll \geq 100$.
 [[CO5](Design/HOCQ)]
 (b) What do you understand by temporaries? Why are they used in parametric queries?
 [[CO5](Remember/LOCQ)]
8 + 4 = 12
5. (a) What are the levels of distribution transparency? Explain with diagram.
 [[CO3](Analyse/HOCQ)]
 (b) Describe briefly the two methods used for determining redundant allocation of fragments. What do you understand by Distributed Join?
 [[CO4](Remember/LOCQ)]
 (c) What are the correctness rules of fragmentation?
 [[CO2](Apply/IOCQ)]
4 + (3 + 2) + 3 = 12

Group - D

6. (a) What is False deadlock? Explain with an appropriate example.
 [[CO4](Remember/LOCQ)]
 (b) What is the difference between centralized and distributed deadlock detection?
 [[CO4](Contrast/IOCQ)]
 (c) What is a hierarchical deadlock detector?
 [[CO4](Remember/LOCQ)]
4 + 5 + 3 = 12
7. (a) A distributed database consisting of two sites executing transactions T1 and T2. The associated data objects are P, Q, R, S. Critically comment whether the execution sequence is conflict serializable or not.

Site1: R1(Q)	R2(P)	W1(P)	W2(Q)
Site2: R1(S)	R2(R)	W2(R)	W1(S).

 [[CO3](Analyse/HOCQ)]

- (b) What is a network failure with respect to distributed transaction? [[C04](Remember/LOCQ)]
- (c) Why are temporaries used in parametric queries? Explain your answer with suitable example(s). [[C02](Apply/IOCQ)]
- 6 + 3 + 3 = 12**

Group - E

8. (a) Consider a given relation *Employee* as follows:

Attribute	Size	No. of distinct values
EMP_No	5	200
eName	25	150
Department	12	100

There are 200 tuples in the relation *Employee* and EMP_no is the primary key. Draw the database profile for *Employee*. [[C02](Design/IOCQ)]

- (b) How is Transmission requirements of a distributed query evaluated in terms of cost and delay? [[C04](Understand/LOCQ)]
- (c) What do you understand by fragment reducers in query optimization? [[C05](Remember/LOCQ)]
- 4 + (3 + 3) + 2 = 12**

9. (a) What are the uses of Catalogs in DDMBS? [[C05](Remember/LOCQ)]
- (b) What are the contents of a Catalog? [[C05](Remember/LOCQ)]
- (c) What are the different alternatives of Catalog management? [[C05](Understand/LOCQ)]
- 3 + 4 + 5 = 12**

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	51.04	19.79	29.17