DISTRIBUTED DATABASES
(CSEN 4241)

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) Relation R=(P,B,C,D) with PB as primary key. Choose one FD such that R should be in 1NF but not in 2NF
(a) PB→ P (b) PB→D (c) P →D (d) P→P

(ii) The fragmentation where a base table is required to divide tuples is
(a) Derived Horizontal Fragmentation (b) Derived Vertical Fragmentation
(c) Dedicated Horizontal Fragmentation (d) Dedicated Vertical Fragmentation

(iii) 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.

(iv) Commit starts
(a) when coordinator requests all subordinates to "get ready"
(b) when coordinator acknowledges that they are all agreed
(c) both (a) and (b)
(d) none of these.

(v) The number of tuple of each fragment Ri of a global relation R is called
(a) Cardinality (b) Size (c) Value (d) Domain.

(vi) Depending on the situation, each site in the distributed database system can act as
(a) A client (b) A server (c) Both (a) and (b) (d) None of the above.

(vii) Let us assume that in 2PC protocol a transaction coordinator failed after a decision is taken (to abort/commit) and shared among the participating sites. What should the coordinator do during restart (recovery)?
(a) Abort the transaction in any case
(b) Commit the transaction in any case
(c) Commit/abort only if received all acknowledgements from participating sites
(d) Cannot be decided by the coordinator on what to be done.

(viii) In a Distributed Database, the benefit of allocating a fragment at a particular site is calculated using Bij= Σ fkj * nki, considering k as the application index, j as the site index and i as the fragment index. The component nki represents
(a) Only Update by jth application
(b) Only Read by jth application
(c) Both Read and Update by jth application
(d) None of these.

(ix) A relation R(X, Y, Z), holds F= { XY→Z, Z→Y}. The candidate keys will be
(a) {XY} only (b) {XY} and {XZ}
(c) {XY}, {XZ} and {YZ} (d) X only.

(x) A Semi-join is which of the following?
(a) Only the joining attributes are sent from one site to another and then all of the rows are returned
(b) All of the attributes are sent from one site to another and then only the required rows are returned
(c) Only the joining attributes are sent from one site to another and then only the required rows are returned
(d) All of the attributes are sent from one site to another and then only the required rows are returned.

Fill in the blanks with the correct word

(xi) There are _____ layers in the ISO/OSI network reference model.

(xii) _____ parallelism enables the execution of multiple queries generated by concurrent transactions.

(xiii) Parallelism in which the relations are partitioned on multiple disks to reduce the retrieval time of relations from disk is called as _____.

(xiv) Two-phase commitment protocol is used for _____.

(xv) Minterm predicate is a _____ of simple predicates.

Group - B

2. (a) Consider a given relation EMP(<u>empno</u>, 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)(Remember/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)(Remember/LOCQ)]*

(2 + 2) + 4 + 4 = 12

Group - C

4. (a) What are the basic objectives of designing the data distribution in DDB? *[(CO2)(Understand/LOCQ)]*

(b) Consider a given relation *Product (p_id, brand, quantity, cost)*. Write the equivalent expressions in relational algebra:

(i) Select sum (cost) from Product where quantity >5;

(ii) Select brand, count (*) from Product group by brand;

(iii) Select brand, sum (cost) from Product group by brand having sum(cost) >= 1000; *[(CO2)(Apply/IOCQ)]*

5 + (2 + 2 + 3) = 12

5. (a) What are the levels of distribution transparency? Explain with diagram. *[(CO3)(Analyse/IOCQ)]*

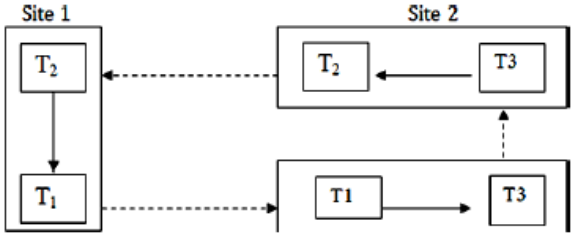
(b) Describe briefly the two methods used for determining redundant allocation of fragments. What do you understand by Distributed Join? *[(CO3)(Remember/LOCQ)]*

(c) What are the correctness rules of fragmentation? *[(CO3)(Apply/IOCQ)]*

4 + (3 + 2) + 3 = 12

Group - D

6. (a) Consider the following distributed wait-for-graph (DWFG):



Illustrate each step of detecting deadlock using distributed Deadlock Detection Algorithm for the above graph. *[(CO3)(Analyse/HOCQ)]*

(b) Explain the steps of both coordinator and participant with respect to two phase commitment protocol, along with the 2PC algorithm. *[(CO4)(Remember/LOCQ)]*

6 + 6 = 12

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

(i) The coordinator fails after writing a '*prepare*' record in the log but before writing a '*global_commit*' or '*global_abort*' record in the log.

(ii) An answer message (READY or ABORT) from a participant is lost.

[(CO4)(Analyze/IOCQ)]

(b) Explain the stepwise transformation of the centralized 2PC protocol into a distributed commit protocol. How many messages are required for the distributed protocol with N participants? *[(CO4)(Remember/LOCQ)]*

6 + (4 + 2) = 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*. *[(CO2)(Design/IOCQ)]*

(b) How is Transmission requirements of a distributed query evaluated in terms of cost and delay? *[(CO4)(Understand/LOCQ)]*

(c) What do you understand by fragment reducers in query optimization? *[(CO5)(Remember/LOCQ)]*

4 + (3 + 3) + 2 = 12

9. (a) What are the two ways in which authorization rules are allocated and enforced in distributed database systems? *[(CO5)(Understand/IOCQ)]*

(b) Consider relations:
Student(roll, name, birthdate, stream, cid),
Course(cid, cname, start-date, location-id),
Location(location-id, location-state, location-district).
Write the algebra to find the courses undertaken by students whose course location is in the state Bihar. Project only the roll number, course name, start-date, and location-districts of the students. Exclude any courses whose start-date is before 21st Aug, 2022, and also exclude students whose stream is S1.
Draw the initial query tree, and thereafter step by step transform the query tree to more efficient query tree. Show and discuss each and every step undertaken to transform your initial query tree. *[(CO5)(Analyse/HOCQ)]*

4 + 8 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	45.83	29.17	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.*