DISTRIBUTED DATABASE MANAGEMENT SYSTEMS
(INFO 3132)

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) Which of the following statements are true with respect to Fragmentation Design in a distributed database?
Statement1: Simple predicate set should be minimal
Statement 2: Simple predicate set should be complete
Statement 3: Simple predicate set should be disjoint
Statement 4: Simple predicate set should be closed
(a) Only statement 1 (b) Only Statement 2
(c) Both Statement 1 and 2 (d) All.

(ii) In 2PC protocol of distributed transaction, preparing for commit and waiting for AAM or READY message happens in
(a) Phase I (b) Phase II (c) Phase III (d) Phase IV.

(iii) 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.

(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 atmost one fragment
Statement 3: A tuple should be allocated to all the fragments
(a) Only statement 1 (b) Only Statement 2
(c) Both Statement 1 and 2 (d) Only Statement 3.

(v) Consider a Distributed Database Design, where a global relation R has been horizontally fragmented into R1, R2, R3. If another relation Z is to be fragmented based on R1 then the type of fragmentation that will be performed is
(a) Vertical Fragmentation (b) Derived Horizontal Fragmentation
(c) Derived Vertical Fragmentation (d) Hybrid Fragmentation.

(vi) The _____ is not a join algorithm applied for optimization of queries.
(a) Left Loop Join (b) Block nested loop Join
(c) Hash join (d) Nested loop join

(vii) The design of creating horizontal fragments based on the already designed horizontal fragments of another relation is
(a) Vertical Fragmentation (b) Derived Horizontal Fragmentation
(c) Derived Vertical Fragmentation (d) Hybrid Fragmentation.

(viii) In a Distributed Database, the benefit of allocating a fragment at a particular site is calculated using Bij= $\sum$ 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.

(ix) Which of the following statements are true with respect to query optimization?
Statement 1: Query optimizer needs to show the order of operations for query execution, which is not possible in query tree
Statement 2: Query optimizer needs to show the order of operations for query execution, which is not possible in query graphs.
(a) Only Statement 1 (b) Only Statement 2
(c) Both Statement 1 and 2 (d) None is true.

(x) In best fit approach of allocation design, the fragments are allocated to sites keeping _____.
(a) locality of reference (b) locality of allocation
(c) locality of fragmentation (d) locality of regions

Group- B

2. (a) A relation R with attributes X,Y,Z have the following FDs holding on it:
XY→Z
Z→Y
(i) Find out the candidate keys.
(ii) Find out the highest normal form in which this relation exists.
(iii) Is it possible to normalize R to BCNF along with dependency preservation? Critically comment on it. [(CO6)(Apply/IOCQ)]

(b) A relation R(A,B,C,D,E,F) with attributes has the following given set of FD's. Find the canonical cover of F.
F= {A→B,B→A, A→C,(A,D)→F,C→E}. [(CO6)(Apply/IOCQ)]
(2 + 3 + 3) + 4 = 12

3. (a) Considering an online electric bill payment site. Considering the bill amount of Rs. 2000, Write a PL/SQL code to check the balance of e-wallet. If it is greater than 2000 update the e-wallet accordingly. Otherwise, display "Insufficient Balance in wallet". [(CO4)(Apply/IOCQ)]

(b) State the features of distributed database that makes it different from centralized database. [(CO1)(Understand/LOCQ)]
8 + 4 = 12

Group - C

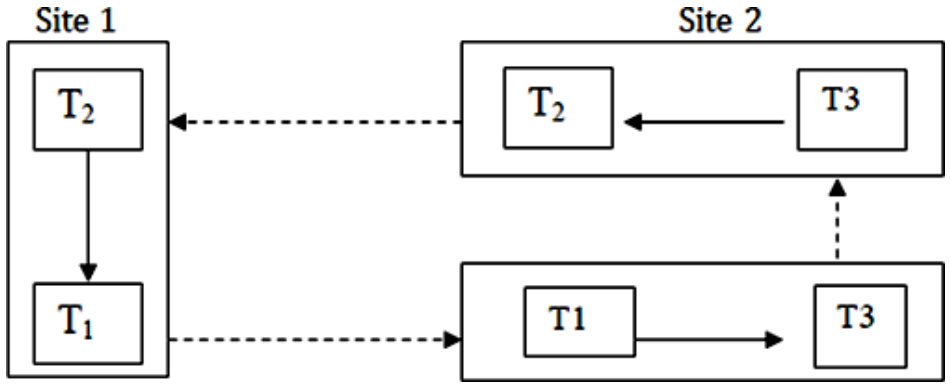
4. (a) Consider a Distributed Database of a Real estate company, where a Global relation Land has the following attributes: Land-Plot-No, Land-Mouza, Land-Area, Land-Price. There are three Mouza where the land exists: Choushal, Basudevpur, and Bedhanpur. There exists only one application that accesses the tuples of the Land relation using its mouza in formation.
   (i) Design the set of simple predicates.
   (ii) Also design the minterm predicates from those.
   (iii) Find out the valid minterm predicates.
   A new application arives which accesses the tuples with respect to Land-Price. All tuples with Land-Price less than 700000/katha are accessed. Redesign the set of simple predicates, so that completeness is achieved. [(CO2)(Apply/IOCQ)]
   (b) Explain vertical horizontal fragmentation. [(CO1)(Understand/LOCQ)]
   (6 + 3) + 3 = 12

5. (a) Consider a bank database distributed at sites Delhi, Mumbai, and Kolkata, having a global relation F that has been fragmented into F1, F2, and F3. Three applications A1, A2, and A3 issued at same frequency at different sites. A1 is issued at Delhi, reads 9 records of F1 and 3 records of F2. A2 is issued at Kolkata reads 7 records of F3 and 5 records of F2. A3 is issued at Mumbai reads 10 records of F2. If we take locality of references as objective, then determine the best allocation of fragments. [(CO2)(Evaluate/HOCQ)]
   (b) What are the levels of distribution transparency? Explain with diagram. [(CO1)(Analyse/IOCQ)]
   8 + 4 = 12

Group - D

6. (a) Explain the state transition diagram of both coordinator and participant with respect to two phase commitment protocol, along with the 2PC algorithm. [(CO3)(Understand/LOCQ)]
   (b) 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,CO5)(Analyse/IOCQ)]
   6 + 6 = 12

7. (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,CO5)(Analyze/IOCQ)]



   (b) Explain the algorithm for 3PC protocol in distributed environment for both Coordinator and Participants. [(CO3,CO5)(Understand/LOCQ)]
   6 + 6 = 12

Group - E

8. (a) Consider, a movie ticket booking system, where two of the relations are **Movies_Info** and **Showtime_Info**.
Movies_Info has 20,000 tuples and needs 241 blocks to store the tuples of the relation in the disk file. **Showtime_Info** has 40000 records, and needs 482 blocks to store the tuples in the disk file.
There is an application which requires information from both the tables, and there by will have to perform join operation. Determine both in the worst and best case the number of block access required if block nested loop join algorithm is applied.
[(CO4)(Evaluate/HOCQ)]

(b) Write the pseudo code along with explanation for any one join algorithm used for efficiently joining two objects. [(CO4)(Alalyse/IOCQ)]

7 + 5 = 12

9. (a) 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. [(CO4)(Create/HOCQ)]

(b) Explain with example how Hash Join optimizes the join operation.
[(CO4)(Analyse/IOCQ)]

8 + 4 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	19.8	56.25	23.95

Course Outcome (CO):

After the completion of the course students will be able to
INFO3132.1: Understand the need of distributed database and various architectures of DDBMS
INFO3132.2: Apply various distribution fragmentation techniques given a problem
INFO3132.3: Apply the different transaction recovery techniques.
INFO3132.4: Analyze and apply query optimization algorithms
INFO3132.5: Compare various approaches to concurrency control in Centralized and Distributed database
INFO3132.6: Design a normalized centralized database, and can convert into distributed database with respect to a given problem

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