DISTRIBUTED DATABASE MANAGEMENT
(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) The part of a database management system which ensures that the data remains in aconsistent state is
(a)authorization and integrity manager (b)buffer manager
(c)transaction manager (d)file manager

(ii) The _____ is a join algorithm used for query optimization
(a) Block nested Loop Join (b) Quick Join
(c) Left Join (d) Heapify Join

(iii) A set of simple predicates P is minimal if
(a) All its predicates are minimal (b) All its predicates are relevant
(c) All its predicates are disjoint (d) Its predicates are duplicate

(iv) Relation R=(A,B,C,D) with AB as primary key. Choose one FD such that R should be in 1NF but not in 2NF.
(a) AB -> C (b)AB -> D
(c) A -> D (d)AB-> CD

(v) 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

(vi) A set of simple predicates P1 consists of 5 predicates, therefore number of minterm predicates are
(a) 25 + 1 (b) 2
(c) 28 (d) 25

(vii) In a distributed database the levels of distribution transparencies are
(a) Fragmentation (b) Location
(c) Local Mapping (d) All

(viii) 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

(ix) In distributed database what does DTM signify?
(a) Distributed Transition maintenance
(b) Distributed transaction manager
(c) Distributed Terminal manager
(d) None

(x) Completeness, Disjointness, and _____are the correctness rules of fragmentation
(a) Reconstruction (b) Cohesion
(c) Uniqueness (d) Conversion

Group- B

2. (a) A relation R (A, B , C , D , E,F) with attributes has the following given set of FD's

A→B, B→A,
A→C, (A,D) → F
C→E

i) Find out the candidate keys from the set of given FD's.
ii) Find the canonical cover of F.
iii) Find out in which normal form the relation is? Explain.
iv) Convert the relation into its higher normal form such that dependency is preserved and lossless decomposition occurs. Explain.[(CO6) (Apply/IOCQ)]

(b) What is entity integrity constraint? Explain with example.
[(CO6) (Understand/LOCQ)]
8 +4 =12

3. (a) Given a relation Employee{<u>EmpID</u>, EmpName, Dept, Salary, Course , DateCompleted} where the primary key is {EmpID, Course}, Dept is the department where employee belongs and DateCompleted is the date when an employee completed a particular training in a course. Every other field is self-explanatory. Some employees have taken more than one course in their training.
The FDs are given below: -
FD1: EmpID, Course →EmpName, Salary, Dept, DateCompleted
FD2: EmpID→EmpName, Salary, Dept

(i) Is this relation in 2NF? Explain.
(ii) What nature of dependency exists? What problems can occur?
(iii) How will this be fixed? [(CO6) (Apply/IOCQ)]

(b) (i) What is transitive dependency? Explain with an example.
(ii) What problems occur? Explain with proper examples.
(iii) How it is corrected? [(CO6) (Understand/LOCQ)]

(2+2+2) + (2+3+1)=12

Group - C

4. (a) Consider a relation Course with the following attributes: Course-id, Course-name, Course-fee, Location. There are three locations where the courses are conducted: Bankura, Birbhum, and Burdwan. There exists only one application that accesses the tuples of the course table using its location basis.
(i) Design the set of simple predicates.
(ii) Also design the minterm predicates from those.
(iii) Find out the valid minterm predicates.
(iv) Justify completeness property with respect to the fragments created.
(v) A new application arrives which accesses the tuples with respect to Course-fee. All tuples with Course-fee less than 7000 are accessed. Redesign the set of simple predicates, so that completeness is achieved.
[(CO2) (Apply/IOCQ)]

(b) Explain the difference between horizontal and derived horizontal fragmentation Schema. [(CO1) (Understand/LOCQ)]

(c) What are the levels of Distributed Transparency? Explain with diagram.
[(CO1) (Understand/LOCQ)]

5 + 2 + 5 = 12

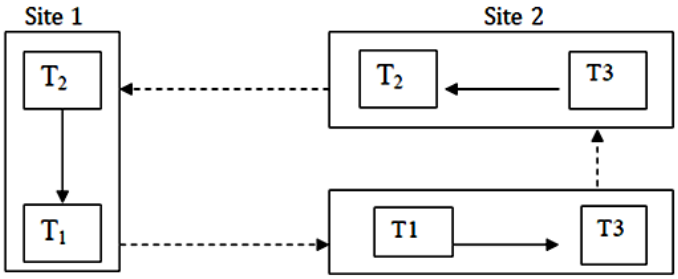
5. (a) A distributed database system, having 3 sites is designed, where 3 fragments R1, R2, and R3 exists. Now, consider the following two allocations of fragments:
1. R1 at site 1 R2 at site2 and R3 at site 3
2. R1 and R2 at site 1 and R2 and R3 at site 3
The following applications are issued with same frequency.
A1 issued at site 1, reads 6 records of R1, updates 2 records of R1 and 7 records of R2
A2 issued at site 3 reads 4 records of R3 and 8 records of R2
A3 issued at site 2 reads 15 records of R2
If we take locality of references as objective, then which solution is the best.
[(CO2) (Apply/IOCQ)]

(b) What is distributed database? What is the advantage of distributed database over centralized? [(CO1) (Understand/LOCQ)]

7 + (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, CO5) (Apply/IOCQ)]

(b) Consider a distributed DBMS with two sites running transactions T1 and T2 with following data objects: A,B,C,D
Show whether the following execution sequences running at site 1 and site2 are conflict serializable or not. If the answer is affirmative, determine all possible total orders of transactions. If answer is negative, prove that there is no total order possible.
Execution 1:
Site 1: R1(A) R2(A) W1(A) W2(B)
Site 2: R1(C) R2(C) W2(C) W1(D)
Execution 2:
Site1: R1(B) R2(A) W2(A) W1(C)
Site2: W1(C) R1(D) R2(D) W1(D). [(CO3,CO5) (Evaluate/HOCQ)]

6 + 6 = 12

7. (a) Explain the algorithm of 2PC protocol in distributed environment for both Coordinator and Participants. [(CO3,CO5) (Understand/LOCQ)]

(b) Explain the different scenarios of node/site failures of both coordinator and participants in 2PC environment. [(CO3,CO5) (Analyse/IOCQ)]

6 + 6 = 12

Group - E

8. (a) Explain the hash join algorithm. [(CO4) (Remember/LOCQ)]

(b) Joining tables is a costly operation. To optimize the cost joining algorithms are applied. Consider, two relations Employers and Department. Employer has 33000tuples and requires 211 blocks to store these tuples in the disk file. Department has 16000 records, and blocks required is 90.
In order to join the two tables, if block nested loop join algorithm is used then calculate the number of block access required for worst case and also in the best case. [(CO4) (Apply/IOCQ)]

6 + 6 = 12

9. (a) Consider relationsStudent(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 born before 27th April, 2019. Project only the roll number, course name and location-district of the students. Exclude any courses, whose start-date is before 21st Aug, 2022.
Draw the initial query tree, and thereafter step by step transform the query tree to more efficient query tree. [(CO4) (Apply/IOCQ)]

(b) With the help of a diagram state the different steps of processing a high-level query. [(CO4) (Understand/LOCQ)]

6 + 6 = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	41.67%	52.08%	6.25%

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

Department & Section	Submission Link
IT	https://classroom.google.com/c/NDA1NjI0NDM4NDQ5/a/NDYzNjc3NDg0NDMy/details