Course Code : CST 408-2 **ITSJ/RW – 17 / 1400**

Eighth Semester B. E. (Computer Science and Engineering) Examination

Elective - III

DISTRIBUTED AND PARALLEL DATABASES

Time : 3 Hours] [Max. Marks : 60

Instructions to Candidates :—

(1) All questions carry marks as indicated.

(2) Assume suitable data and illustrate answers with neat sketches wherever necessary.

1. (a) Compare the following parallel database architectures with their pros and cons :

Shared Disk and Shared Nothing
Also explain their working. 5 (CO 1)

(b) Draw and discuss the following three scenarios with respect to distributed databases :

(1) A distributed database on a geographically dispersed network

(2) A distributed database on a local network

(3) A multiprocessor system.

Categorize the above three scenarios into distributed and non distributed database with reason. 5 (CO 1)

2. Solve any **Two** :—

(a) Consider the following global, fragmentation, and allocation schemata :

Global Schema : STUDENT (NUMBER, NAME, DEPT)
Fragmentation Schema : STUDENT 1 = $SL_{\mathbf{DEPT}}$ = "EE" STUDENT
STUDENT2 = $SL_{\mathbf{DEPT}}$ = "CS" STUDENT
Allocation Schema : STUDENT1 at sites 1, 2
STUDENT2 at sites 3, 4

Consider an application which requires the student number from the terminal and outputs the name and department. If this application is repeated for many possible values of the student number. Write the application :

(i) Accessing the database for each student number given at the terminal

(ii) Accessing the database after having collected several inputs from the terminal. 5 (CO 2)

(b) Illustrate the formula for measure of costs and benefits for horizontal and vertical fragmentation. 5 (CO 2)

(c) Assume a relation EMP as given below :

EMP(Empno, Ename, Job, Sal, Dept)

Furthermore assume that there are two applications which accessing the above table. One application retrieves information about employees who earn more than 5000 and another application manages data about employees who are "Clerk" (Job).

(1) Infer the simple predicates for the above applications.

(2) How many valid min term predicates we can derive from the above problem ? List them.

(3) Assume that there are three departments out of which "Finance" is accesses more frequently by one more application. What would be the number of minterm predicates ?

(4) What happens when a set P of simple predicates is complete but not minimal ? Give an example for the above applications.

(5) If by mistake one of the valid min term fragments is missed, what will be the effect on reconstruction of global relation EMP ? 5 (CO 2)

3. (a) Analyze the behavior of 2 phase commit protocol in case of site failures. 5 (CO 2)

(b) Draw and discuss the reference model for distributed concurrency control. 5 (CO 2)

4. Solve any **Two** :—

(a) Consider the join R JN $_{A = B}$ S; assume that R and S are at different sites and disregard the cost of collecting the result of the join. Let C0 = 0 and C1 = 1. The following profiles are given :

Size(R) = 25; card(R) = 50; val(A[R]) = 50; size(A) = 3
Size(S) = 10; card(S) = 50; val(B[S]) = 50; size(B) = 3
RSJ $_{A=B}$ S has selectivity ρ = 0.2
SSJ $_{B=A}$ R has selectivity ρ = 0.8

Determine the transmission cost of performing the join at the site R using semi join reduction. 5 (CO 3)

(b) Provide the equivalence transformation for the relational algebra specifying with example the values for validity indicator and sufficient and necessary conditions.
Consider following properties :

(i) Commutativity of unary and binary operations.

(ii) Associativity of binary operations.

(iii) Idempotence of unary operations. 5 (CO 3)

(c) Design a query which requires supplier number of suppliers that have issued a supply order in the South area of the country and having quantity of orders more than 500. Apply the canonical expression for the same query. Apply all the possible criteria to simplify the query. 5 (CO 3)

5. (a) Consider the two FUND_TRANSFER transactions Ti and Tj given in book. Assume that initially RTM(x) = 25, WTM(x) = 25, RTM(y) = 30 and WTM(y) = 30. Determine the possible executions and transaction restarts in the following cases :

(i) TS(Ti) = 35; TS(Tj) = 40

(ii) TS(Ti) = 20; TS(Tj) = 40

(iii) TS(Ti) = 40; TS(Tj) = 35

(iv) TS(Ti) = 45; TS(TJ) = 30

(v) TS(Ti) = 35; TS(Tj) = 45 5 (CO 2)

(b) What is distributed deadlock ? Explain the following in the context of distributed deadlock prevention :-

(1) Non-preemptive method.

(2) Preemptive method. 5 (CO 2)

6. (a) Enlist and explain distributed database mining challenges. 5 (CO 4)

(b) Provide the details of any Two access tools for data warehousing. Comment on improvement of performance by the application of parallelization. 5 (CO 4)