Course Code : CST 418 – 2 / CST 408 – 2   KOLP/RW – 19 /9624

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)   For the following parallel database architecture, describe their topology and functionality :—

               (1) Shared Disk Architecture   (2) NUMA Architecture   5(CO1)

    (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(CO1)

2.   (a)   Consider the following relation schema and the set of applications that are frequently accessing the relation :

        DEPARTMENT (Deptno, DeptName, School, Phone)

        Application1 : Find the department details of 'DPS' schools
        Application2 : Find the phone of "Digital Communication" department

        Find the simple predicates and min – term predicates. Fragment the given table horizontally according to the requirement and check for the correctness of the fragment with the help of completeness, reconstruction and disjointness condition.   5(CO2)

(b)   Derive the measure of cost and benefit in redundant and non redundant cases in horizontal fragmentation.   5(CO2)

3. Solve any **Two** :—

    (a)   Consider a data item x. Let $RTM(x) = 40$ and $WTM(x) = 35$. Let the pair $\{Ri(x), TS\}$ $\{Wi(x), TS\}$ denote a read (write) request of transaction Ti on the item x with Timestamp TS. Indicate the behavior of the conservative timestamp method with the following sequence of requests :

    $\{R1(x), 22\}, \{R2(x), 26\}, \{W3(x), 20\}, \{R4(x), 28\}, \{R5(x), 32\}, \{W6(x), 31\}$

    Also explain the conservative timestamp method.   5(CO2)

    (b)   Compare the concurrency based locking in centralized databases and distributed databases.   5(CO2)

    (c)   Consider the following diagram of two transactions at two sites :
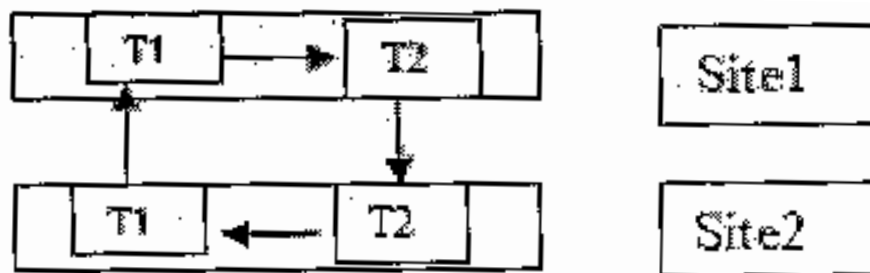
    Write the procedure for distributed deadlock detection.   5(CO2)

4. Solve any **Two** :—

    (a)   List all criteria's needed for query simplification. Show the application of each criterion on query tree with proper examples.   5(CO3)

    (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(CO3)

(c) 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) = 50 ; card (R) = 100 ; val (A [R]) = 50 ; size (A) = 3
Size (S) = 5 ; card (R) = 50 ; val (B [S]) = 25 ; size (B) = 3

Calculate the transmission cost and delay of :

(a) Performing the join at the site R without semijoin

(b) Performing the join at the site S without semijoin

(c) Performing the join at the site R with semijoin

(d) Performing the join at the site S with semijoin

where D1 corresponds to 100 bits per second.   5(CO3)

5. (a) List the reasons for failure in distributed systems and the basic fault tolerance approaches and techniques to cope with them.   5(CO2)

    (b) The use of replication and caching in distributed database systems improves performance and reliability, but it requires attention to consistency of data. Mention the types of strict replica control protocols. Elaborate any one protocol.   5(CO2)

6. (a) Provide the details of any two access tools for data warehousing. Comment on improvement of performance by the application of parallelization.   5(CO4)

    (b) Identify various data mining techniques and try to create a correlation between these techniques and centralized algorithms. Mention the important features of these algorithms.   5(CO4)