CAT 1

Answer all the questions

a) **Explain the different ways in which distributed databases are applied in business organizations (5 Marks)**

I. Geographic distribution: When an organization operates across different geographic locations, a distributed database can be used to store and manage data locally while ensuring data consistency across all locations.

II. Replication: Distributed databases can be used to replicate data across multiple sites to improve data availability and fault tolerance. This ensures that data can be accessed even if one or more sites fail.

III. Load balancing: Distributed databases can be used to distribute the workload across multiple sites to improve performance and scalability.

IV. Cost savings: Distributed databases can help organizations save on hardware, software, and infrastructure costs by utilizing existing resources more efficiently.

V. Security: Distributed databases can improve data security by replicating data across multiple sites and implementing access controls to ensure that data is only accessed by authorized users.

b) **Describe fragmentation correctness rules in distributed databases (4Marks)**

I. Completeness rule: Every data item in the original database should be present in at least one fragment of the distributed database.

II. Reconstruction rule: The original database should be able to be reconstructed from the fragments of the distributed database.

III. Disjointness rule: No data item should be stored in more than one fragment of the distributed database.

c) **Discuss the following data distribution design alternatives in DBMS**

(4Marks)

I. Horizontal partitioning: This involves dividing a table into multiple smaller tables based on rows. Each smaller table contains a subset of the original table's rows, and all tables have the same schema.

II. Vertical partitioning: This involves dividing a table into multiple smaller tables based on

columns. Each smaller table contains a subset of the original table's columns, and all tables have different schemas.

III. Hybrid partitioning: This involves combining horizontal and vertical partitioning to achieve the benefits of both.

d)**Discuss three aspects of distribution transparency in DDBMS (3 Marks**

I. Location transparency: This ensures that users and applications can access data without knowing the physical location of the data.
II. Fragmentation transparency: This ensures that users and applications can access data without knowing how the data is fragmented across different sites.
III. Replication transparency: This ensures that users and applications can access data without knowing how the data is replicated across different sites.

e)**Discuss some of anomalies that need to be dealt with by the scheduler in controlling concurrency (3 Marks)**

I. Lost update: This occurs when two or more transactions update the same data item at the same time, and one of the updates is lost.
II. Dirty read: This occurs when one transaction reads data that has been modified by another transaction that has not been committed.
III. Phantom read: This occurs when a transaction reads data that has been modified by another transaction that has been committed.

f**)Explain the two ways for the Local Recovery Management (LRM) to deal with update/write operations (2 Marks)**

I. Deferred update: In this approach, the changes made by a transaction are recorded in a transaction log, but the actual changes are not made to the database until the transaction is committed.
II. Immediate update: In this approach, the changes made by a transaction are immediately written to the database, and the transaction is committed.

g)In optimistic concurrency control, a transaction's life cycle is divided into three phases, discuss these phases (3 Marks)

I. Read phase: In this phase, the transaction reads data without acquiring any locks. The transaction checks for conflicts with other concurrent transactions and may choose to abort if conflicts are detected.
II. Validation phase: In this phase, the transaction validates that the data it read during the read phase is still valid. If the data is still valid, the transaction proceeds to the commit phase. Otherwise, the transaction is aborted.
III. Commit phase: In this phase, the transaction commits the changes it made to the database. If conflicts were detected during the validation phase, the transaction is aborted instead of committed.

h)In a distributed system, data may be fragmented, and each fragment can have a number of replicas to increase data availability and reliability. Discuss the objectives that must be considered while designing the fragmentation and allocation of these

fragments to different sites in a distributed system

(6Marks)

I. Minimize data communication: The fragmentation and allocation design should aim to minimize the amount of data that needs to be transferred between sites. This can be achieved by grouping related data items together and allocating them to the same site.
II. Balance load: The fragmentation and allocation design should aim to balance the load across different sites. This can be achieved by evenly distributing the fragments across sites based on their processing capabilities.
III. Ensure data availability: The fragmentation and allocation design should ensure that each fragment has at least one replica on a different site to ensure data availability in case of site failures.
IV. Optimize performance: The fragmentation and allocation design should aim to optimize query performance by ensuring that frequently accessed data is stored on the same site.
V. Ensure data consistency: The fragmentation and allocation design should ensure that data consistency is maintained across all sites. This can be achieved by using appropriate consistency protocols and ensuring that updates are propagated to all replicas in a timely manner.
VI. Minimize cost: The fragmentation and allocation design should aim to minimize the cost of data storage and communication by utilizing available resources efficiently. This can be achieved by considering the cost of communication and storage when deciding on the allocation of fragments to different sites.