Distributed Database Unit-3 for final year students of CSE and IT

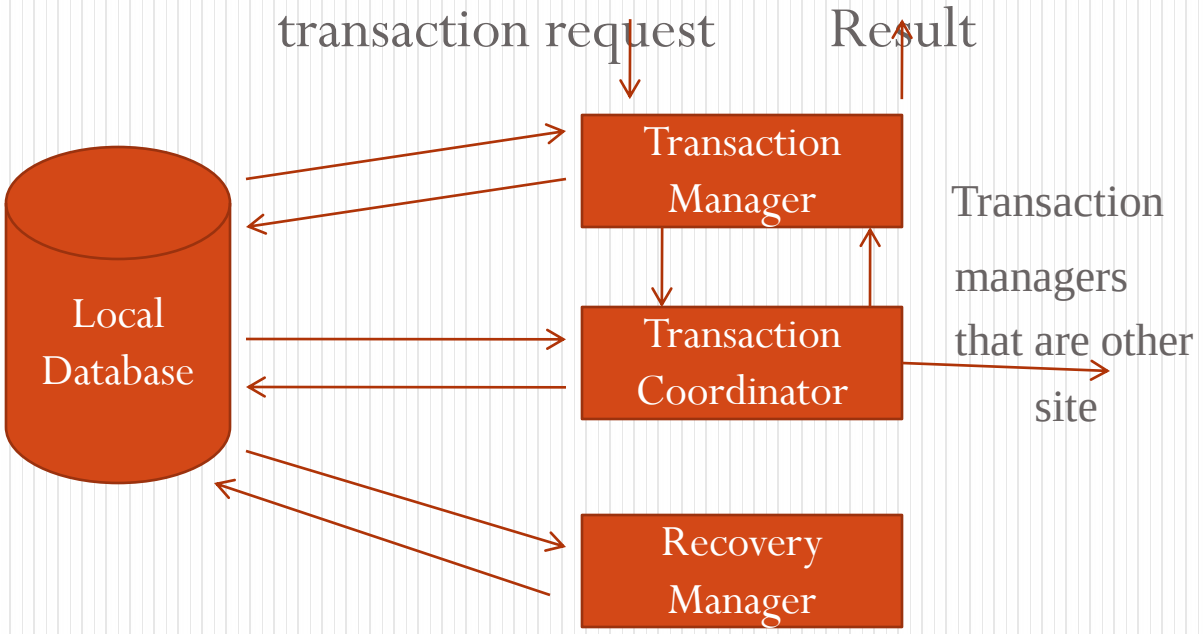
Developed and Presented By:

Dileep Kumar Yadav
Assistant Professor
CSE Dept.
V.B.S P.U,Jaunpur
Mb. No.8726943272
Email-dileep1482@gmail.com

Distributed Transaction Management

- A distributed transaction is a data base transaction in which two or more network hosts are involved, usually hosts are provided transactional resources.
- While the transaction manager is responsible for creating and managing a global transaction that encompasses all operations against such resources.

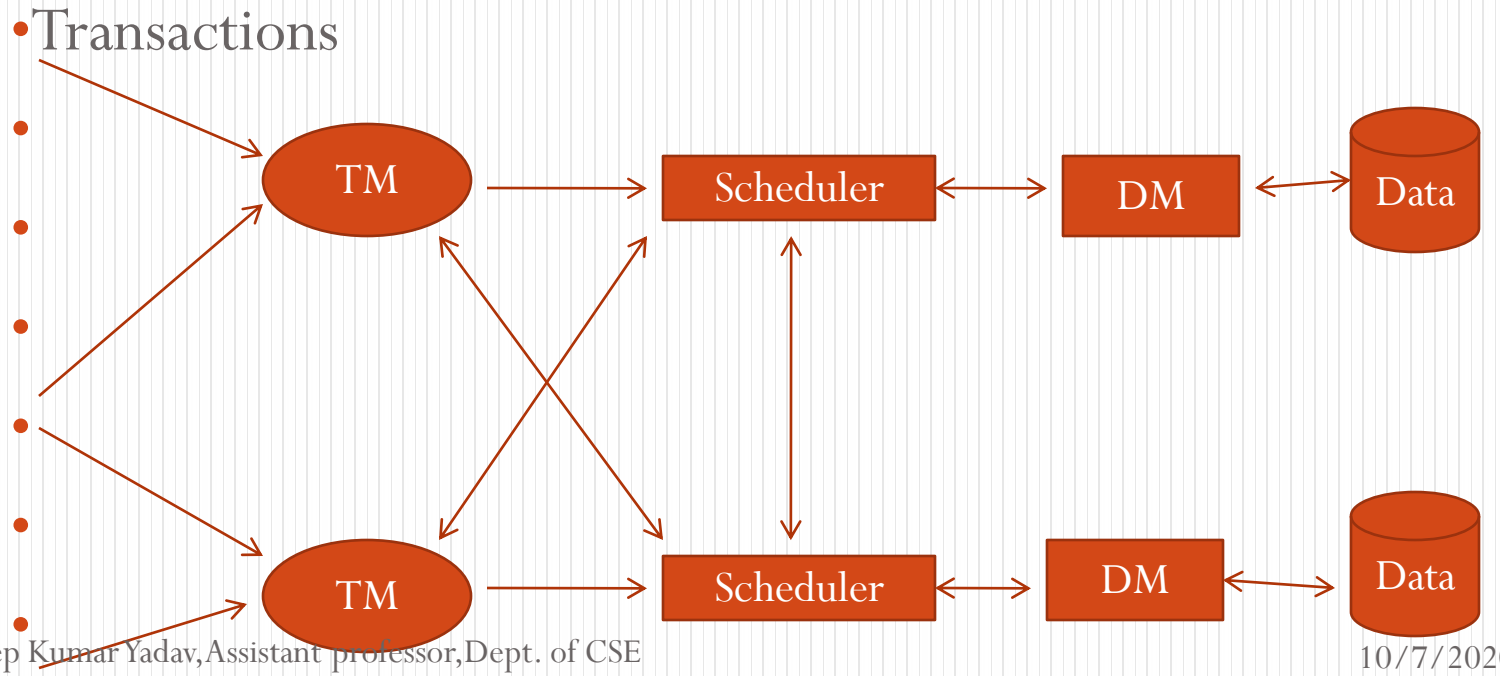
Cont…


transaction request
Result
Transaction Manager
Transaction managers that are other site
Local Database
Transaction Coordinator
Recovery Manager

Cont…

- There are two types of transaction that we need to consider the local transaction which access and update data in only on local database.
- The global transaction which access and update in several local database.
- Each site have its own local transaction managers whose function is to ensure that ACID properties of those transactions.

Cont…



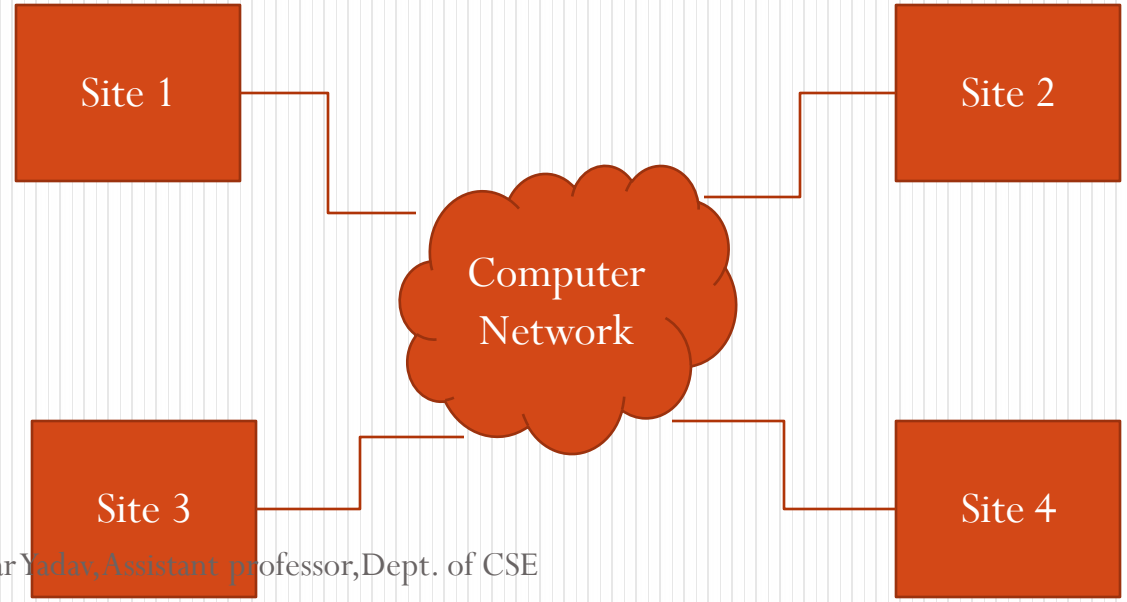
Data Distribution

- A distributed DBMS manages the distributed database in a manner so that it appears as one single database to users.
- A distributed database is a collection of multiple interconnected databases, which are spread physically across various locations that communicate via a computer network.

Cont…

- A distributed database is a database in which data is stored across different physical locations. It may be stored in multiple computers located in the same physical location (e.g. a data centre);
- or maybe dispersed over a network of interconnected computers. Unlike parallel systems, in which the processors are tightly coupled and constitute a single database system, a distributed database system consists of loosely coupled sites that share no physical components.

Cont…



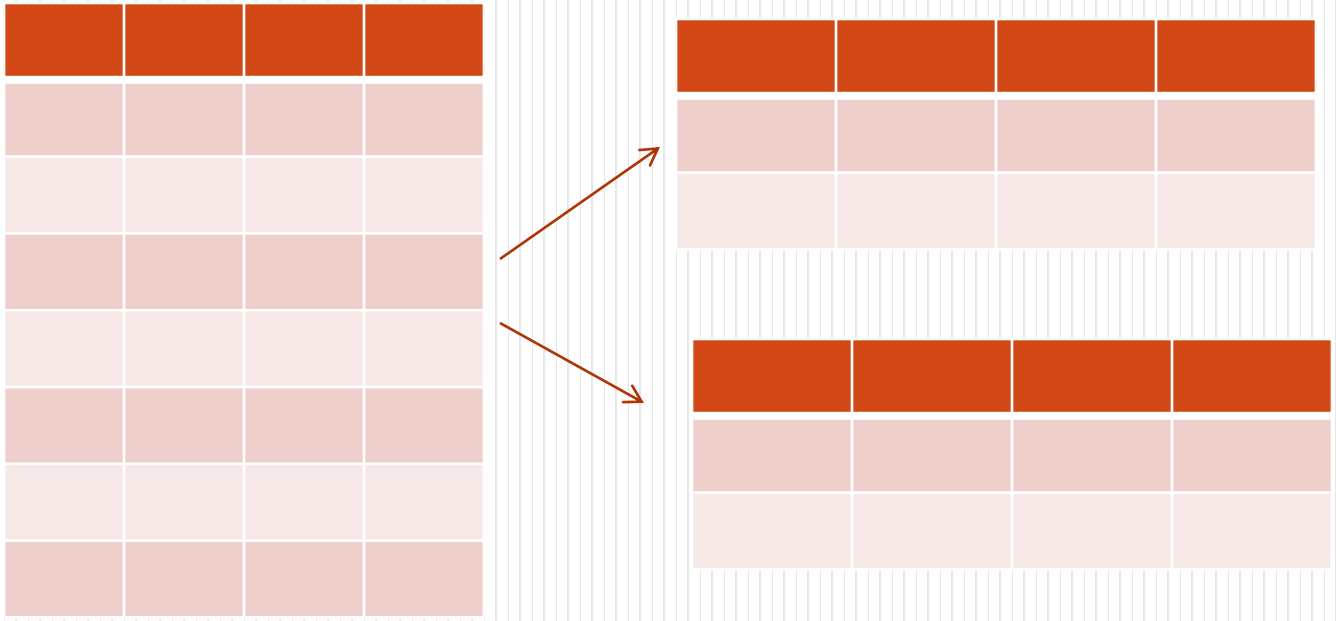
Fragmentation

- The process of dividing the database into a smaller multiple parts is called as fragmentation.
- These fragments may be stored at different locations.
- The data fragmentation process should be carried out in such a way that the reconstruction of original database from the fragments is possible.

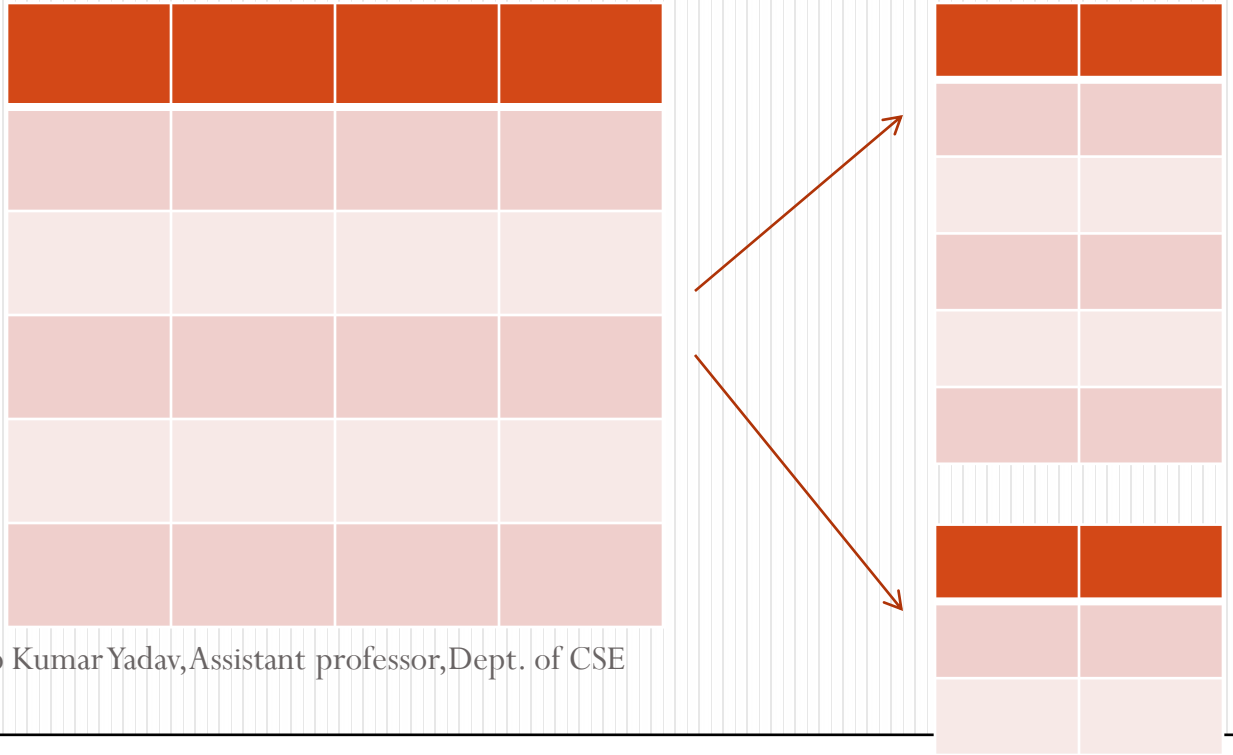
Types of Fragmentation

- Horizontal Fragmentation
- Vertical Fragmentation
- Hybrid Fragmentation

Horizontal Fragmentation



Vertical Fragmentation



Advantages of Fragmentation

- Since data is stored close to the site of usage, efficiency of the database system is increased.
- Local query optimization techniques are sufficient for most queries since data is locally available.
- Since irrelevant data is not available at the sites, security and privacy of the database system can be maintained.

Disvantages of Fragmentation

- When data from different fragments are required, the access speeds may be very high.
- In case of recursive fragmentations, the job of reconstruction will need expensive techniques.
- Lack of back-up copies of data in different sites may render the database ineffective in case of failure of a site.

Replication

- Data replication is the process of storing separate copies of the database at two or more sites. It is a popular fault tolerance technique of distributed databases.

Advantages of Replication

- **Reliability** – In case of failure of any site, the database system continues to work since a copy is available at another site.
- **Reduction in Network Load** – Since local copies of data are available, query processing can be done with reduced network usage, particularly during prime hours. Data updating can be done at non-prime hours.

Cont…

- **Quicker Response** − Availability of local copies of data ensures quick query processing and consequently quick response time.
- **Simpler Transactions** − Transactions require less number of joins of tables located at different sites and minimal coordination across the network. Thus, they become simpler in nature.

Disvantages of Replication

- **Increased Storage Requirements:** Maintaining multiple copies of data is associated with increased storage costs. The storage space required is in multiples of the storage required for a centralized system.
- **Increased Cost and Complexity of Data Updating** : Each time a data item is updated, the update needs to be reflected in all the copies of the data at the different sites. This requires complex synchronization techniques and protocols.

Cont..

- **Undesirable Application i.e. Database coupling** : If complex update mechanisms are not used, removing data inconsistency requires complex co-ordination at application level. This results in undesirable application database coupling.

Distributed Commit

- A commit is the updating of a record in a database. In the context of a database transaction a commit refers to the saving of data permanently after a set of tentative changes.
- A commit ends a transaction within a relational database and follows all other users to see the changes.

Cont…

- Distributed database use commit protocols to ensure atomicity in case of distributed transactions and a wide variety of commit protocols have been proposed.
- When it receives the message that all the sites are ready to commit, it starts to commit.
- In a distributed system either all sites commit or none of them does.

Long Duration Transaction

- It is represented as a nested transaction with atomic database operation i.e. read/write at the lowest level.
- If a transaction fails only active short duration transaction abort. Active long duration transaction resume once, any short duration transactions have recovered.
- Long duration transaction that avoid locks on non-local resources use compensation to handle failures, atomic transactions and use a coordinator to complete or abort the transaction.

Properties of Long duration Transaction

- It support partial roll back.
- It support partial update to design database.
- In crash state data should be restored even for yet to be committed data, so user work is not lost.
- Fast response time is essential so user time is not wasted.

Moss Concurrency Control

- It is the management of concurrent transaction execution. DBMS or DD implement concurrency control techniques to ensure serializability and isolation of transaction in order to guard the consistency and integrity of the database.
- Moss concurrency control is the process of managing simultaneously operations queries like updates, inserts, and deletes on the database without having them interfere with one an other.

Reference

- Database management system by Abraham Silberschatz and Henry F. Korth, Tata McGraw-hill publications.