

# Review of fragmentation and replication in Distributed Database

**Yogita Yashveer Raghav,**

Assistant professor, K.R Mangalam University, SohnaRoad, Gurgaon

**Praveen Kumar,**

M.D.University, Rohtak

**Abstract:** *This paper reviews the basic concepts of Distributed database in distributed database system. Distributed database is the collection of the databases stored at different sites in the interrelated manner over the network. A distributed database is the combination of Distributed database and database management system. This paper provides differentiation between the architecture of central Database and distributed database and various aspects like fragmentation, replication and various problems that can be faced into distributed database system. Data Fragmentation is another way to store data in distributed manner.*

*Replication of data states, that the same copy of data has been maintained in different sites so that if one data sites fails, the same data can be recovered by other sites and execution of the query can be done by taking the data from other sites. Fragmentation is a design technique to divide the Database into multiple sites such that the combination of the partitions provides the original database without any loss of data. The benefits of the fragmentation are that the whole load is not on only one server, and data is scattered by some techniques which will be discussed further. By doing so, this reduces the amount of irrelevant data accessed by the applications of the database, thus reducing the disk accesses, which increases the performance of system.*

**Keywords:** *Distributed Database, Deadlock, transaction, Fragmentation, Replication, and Transparency.*

## **1 Introduction:**

**Information plays a vital role in** today's life. Everyone needs the information at a very high speed with secure and reliable connection. Distributed database is the solution. A distributed database is a database in which storage devices are not all attached to a common processing unit such as the CPU. It is stored at multiple computers. A distributed database system consists of loosely coupled sites which does not share physical component. In distributed database, users at any location can access information from anywhere in the network. The main goal of a distributed database system is to control the management of a distributed database in such a way that it seems to user as data is not scattered at various sites but stored at centralized database.

The purpose of this paper is to present an Introduction to distributed databases, fragmentation, replication, transparency which are becoming very popular now days. Today's business requirements are rapidly changing and requiring the distributed database. Distributed database systems provide an improvement on Communication and data processing because of its data distribution at multiple sites, this makes accessing of data faster but recoverable also, and it provides local control of data for users.

This appearance of centralized database can be achieved with the help of several different kinds of transparencies like: Transaction transparency, performance transparency, location transparency,

Fragmentation transparency, schema change transparency, and local DBMS transparency. The most vital issue in distributed database system is concurrency control. Concurrency control is the method of synchronizing simultaneous access to a database in a multiuser database management system. There are different techniques that Deliver concurrency control. Some of the techniques are: Time stamping, two phase locking and Multiversion time Concurrency control is the process of coordinating concurrent access to a database in multi user fashion.

## 2 Why distributed databases?

1. Today's Business environment requirement
2. Interconnection of existing databases
3. Incremental growth
4. Reduced communication overhead
5. Performance improvement
6. Reliability and availability
7. Batter control of user on local database
8. Batter Data recovery

Distributed Database system consists of two types of sites in a network: query sites and data sites. Query sites are used to query information and are not associated with and local Database other hand data sites are related with database, which is local database.

### 2.1 Types of Distributed Database System:

**Homogeneous Distributed Database System** – In this data is distributed but all servers run on the same database Management system software.

**Heterogeneous distributed database system:** In this different site runs under the control of different DBMS.

### 2.2 Advantages of Distributed database

1. Data is distributed in data sites, so the data that is on high demand can be accessed by the nearby query sites.
2. In case of failure of any data site, query can be executed by taking data from other data sites, as data is replicated at various sites to face these types of crucial conditions.
3. Overall network traffic reduced.

4. It can operate on different platforms. Thus, processor independence is one of the advantages.
5. It is flexible to expand the distributed system over the network since it provides platform independence.
6. Queries are local, so high performance

### 2.3 Disadvantage of Distributed database

1. In Distributed System query sites requests the data from the data sites, so it is mandatory to manage the data on the data sites. For managing the data we need to manage the resources, it involves transaction management, concurrency control, security, backup and recovery as well as query optimization. These tasks need to be implemented on all the data sites to synchronize the data transfer, which becomes difficult in case of huge database.

2. Storage requirement can increase over time, leading to complex data management in the query sites.
3. For using DDBMS, one need to get trained properly, thereby increasing the cost of training.
4. Distributed System is not efficient if there is heavy Interaction between sites.
5. Security problem as data are distributed.

### Architecture of Centralized Database System

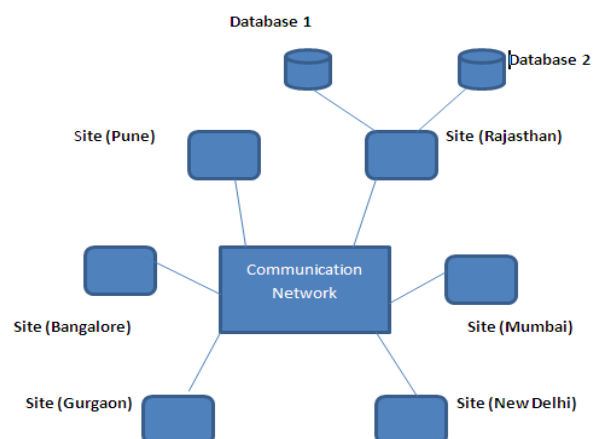


Figure2: Centralized Database System Architecture

Centralized system is the system where the database is stored locally. It can be single user system and multi user system both.

A single user centralized system is the system where the operating system installed supports the single user environment.

A multi user centralized system is the system where the operating system installed supports the multi user environment.

### 3 Architecture of Distributed Database System

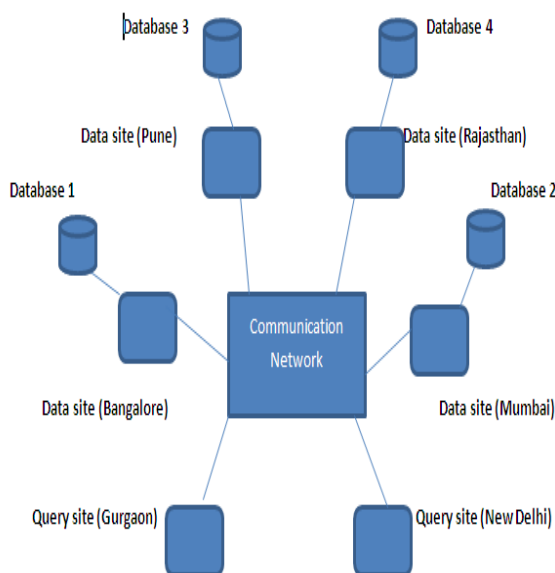


Figure1: Distributed Database System Architecture

#### 3.1 Design of Distributed Database Management System

A distributed database is database whose storage device are not all attached to a common processing unit since database are stored at multiple computer. A distributed database resides on network server on the internet.

#### 3.2 Fragmentation

Data can be stored at various sites in distributed manner. In terms of DBMS, data fragmentation means dividing the Relation  $R$  into many parts, say  $R_1, R_2, \dots, R_n$ , any loss of data, it means it means when we decomposed the relation into many Relations the

functional dependencies should be preserved and data should be lossless. The information and data that we were gaining before decomposition should also gain same after decomposition. Decomposed relations stored at various locations on various data sites in distributes database system

#### 3.3 Advantages of fragmentation

Before we discuss fragmentation in detail, we list four reasons for fragmenting a relation

##### Maintained Confidentiality

In general we never share with other users the whole relations, only that part of the relation that he is concern for is to be shared by the use of the virtual table such as view. So it seems appropriate to work with decomposed relations as the unit of distribution. That maintained the Confidentiality of the relation.

##### Efficiency

Since Data is available closest to the site of usage so it is easy to access the data faster because data is locally available..

##### Batter query optimization

Because data is scattered at various sites, so local query optimization techniques are sufficient for most of the queries, because data is available locally.

##### Parallelism

With fragments as the unit of distribution, a transaction can be divided into several sub queries that operate on fragments. This should increase the degree of concurrency, or parallelism, in the system, thereby allowing transactions that can do so safely to execute in parallel

#### 3.4 Disadvantages of fragmentation

Fragmentation has two primary disadvantages, which we have mentioned previously:

##### Performance

Sometimes query demands data from various sites then we need to gather the data from various sites, which need complex which affect the performance of the system.

### Integrity

Integrity control may be more difficult if data and functional dependencies are fragmented and located at different sites.

### 3.5 Types of Fragmentation

Fragmentation can be of three types:

#### Horizontal Fragmentation

#### Vertical Fragmentation

#### Hybrid Fragmentation

#### Horizontal Fragmentation

It divides the relation into tuples or rows. It divides table horizontally by selecting the relevant rows and these fragments can be assigned to different sites in the distributed system. It allows a class to be partitioned into disjoint instances. In horizontal fragmentation table remains same, only rows gets split. Horizontal fragmentation is defined as selection operation,  $\sigma_{Student(Attributes)}$ .

**Consider Relation:** Relation R is: Student Table S (Roll No., Name, Address, and Contact No. , Fee Balance).

Roll No.	Name	Address	Contact No.	Fee balance
1	Rahul	Gurgaon	9344553456	1000
2	Abhinav	Delhi	7834345563	4000
3	Arman	Noida	3568877576	5000
4	Vishvas	Gurgaon	7835645421	3000
5	Ravi	Rohtak	6788765476	4000

The Student Table S has five fields, as Roll No, Name, Address, Contact No., Fee balance of Student. We are fragmenting this Student Table by

using Horizontal Fragmentation below as we are taking rows based fragmentation on some given condition, Condition is:

**Student Table (S1):** Fragment with Fee Balance Less than 3000.

**Student Table (S2):** Fragment with Fee Balance Greater than 3000.

Student Table(S1)

Roll No.	Name	Address	Contact No.	Fee balance
1	Rahul	Gurgaon	9344553456	1000
2	Vishvas	Gurgaon	7835645421	3000

Student Table (S2)

Roll No.	Name	Address	Contact No.	Fee balance
1	Abhinav	Delhi	7834345563	4000
2	Arman	Noida	3568877576	5000
3	Ravi	Rohtak	6788765476	4000

Whenever we need to access the data from both the sites then we need to reconstruct the relation R from the complete horizontal fragmentation then we use Union set theoretic operation on the Student Table (S1) and Student (S2) which are located at different sites.

Original Student is constructed via:

Student  $\leftarrow$  Student (S1) U Student (S2)

#### Vertical Fragmentation:

Vertical fragmentation technique divides the relation into attributes called columns. In vertical fragmentation it is necessary to include primary key of table in each vertical fragmentation. If any time we need to construct the original table, then it is possible with the help of primary key. In vertical fragmentation one table is split into two or more table. The main objective of vertical fragmentation is to partition relation into a set of smaller relations so that many of the applications will run on only one fragment. Vertical fragmentation of relation S produces S1, S2 each of which contains a subset of E's attributes. Vertical fragmentation is defined using the projection operation of the relational algebra.

II (E1, E2) (R)

**Student Table (S1)**

Roll No.	Name	Address
1	Rahul	Gurgaon
2	Vishvas	Gurgaon

**Student Table (S2)**

Roll No.	Contact No.	Fee balance
1	9344553456	1000
2	7835645421	3000

Condition:

**Student Table (S1):** Fragment with Name and Address

**Student Table (E2):** Fragment with Contact No. , Fee balance

**Student Table (E1) Student Table (E2)**

Student ← Student (S1) ⊗ Student (S2)

### Hybrid Fragmentation:

Hybrid fragmentation can be achieved by performing horizontal and vertical partition together. Mixed fragmentation is group of rows and columns in relation. The original relation in hybrid fragmentation is reconstructed by performing UNION and FULL OUTER JOIN.

**3.6 Data Replication:** In replication,[11] multiple copies of the data are stored at different data sites, and different processes work with different copies..The problem in managing replicated data is to maintain the consistency of the data. The Advantages of replication is to increased availability, increased reliability, improved response time, reduced network traffic, improved system throughput and better scalability. Replicated data are Subject to the mutual consistency rule. The mutual consistency rule requires that all copies of data fragments be identical. Therefore, to maintain data consistency among the replicas, the DDBMS must

ensure that a database update is performed at all sites where replicas exist.

Three replication scenarios exist [11]:

1. A fully replicated database stores multiple copies of each database fragment at multiple sites.
2. A partially replicated database stores multiple copies of some database fragments at multiple sites.
3. An replicated database stores each database fragment at a single site. Therefore, there are no database fragments.

### REFERENCES

- [1] Prof. Vinayak Sinde, Preeti A. Aware, *concurrency control in distributed Database systems, International Journal for Research in Engineering Application & Management (IJREAM) ISSN : 2494-9150 Vol-01, Issue 10, JAN 2016.*
- [2] Gupta V.K., Sheetlani Jitendra, Gupta Dhiraj and Shukla Brahma Datta, *Concurrency control and Security issues in Distributed Database system*, Vol. 1(2), 70-73, August (2012)
- [3] Arun Kumar Yadav & Ajay Agarwal, *An Approach for Concurrency Control in Distributed Database System*, Vol.1, No. 1, pp. 137-141, January-June (2010).
- [4] Navathe Elmasri, *Database Concepts*, Pearson Education, V edition (2008).
- [5] Fundamentals of DBMS, Lakhanpal Publisher, III edition (2008).
- [6] Swati Gupta, Kuntal Saroha, Bhawna, *Fundamental Research of Distributed Database*, IJCSMS International Journal of Computer Science and Management Studies, Vol.11, Issue 02, Aug 2011.
- [7] Distributed Database [http://wps.pearsoned.co.uk/wps/media/objects/10977/11240737/Web%20chapters/Chapter%2012\\_WEB.pdf](http://wps.pearsoned.co.uk/wps/media/objects/10977/11240737/Web%20chapters/Chapter%2012_WEB.pdf)
- [8] D. G. Shin, and K. B. Irani, "Fragmenting relations horizontally using a knowledge based approach," *IEEE Transactions on Software Engineering (TSE)*, Vol. 17, No. 9, pp. 872–883, 1991.
- [9] E. S. Abuelyaman, "An optimized scheme for vertical partitioning of a distributed database," *Int. Journal of Computer Science & Network Security*, Vol. 8, No.1, 2008.
- [10] Gupta Dhiraj and Gupta V.K., *Approaches for Deadlock Detection and Deadlock Prevention for*



Distributed, Res. J. Recent Sci., 1(ISC-2011), 422-425 (2012)  
[11] *Manoj Kumar Sah 1, Vinod Kumar 2, Ashish Tiwari 3, Security and Concurrency Control in Distributed Database System, International Journal of scientific research and management(IJSRM), Volume 2, Issue 11 Pages 1839-1845 ,2014*