iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 1 **Issue:** III **Month of publication:** October 2013

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

Fundamentals & Concepts of Distributed Database

Prof. Pooja Singh

Information Technology Department,
Vadodara Institute of Engineering,
Vadodara Halol Toll Road, Vadodara, India
Pooja1042006@yahoo.com

Abstract: *This paper is prepared with the objective to present an introduction to distributed databases through its two main parts: The first part, is about the fundamentals of distributed databases (DDBS). The Points discussed in this part is related to the motivations of DDBS, architecture, design, features & function, Pros & Cons, etc. In the second part, we include some of the research that has been done in this specific area of DDBS. The topics of this research include, query optimization, distribution optimization, fragmentation, optimization, and join optimization on the net. Some of the examples and results are been provided to demonstrate the topics we are presenting.*

Keywords: *Distributed databases functions and fundamentals, New topics covered: query optimization, distribution optimization, fragmentation optimization and join optimization on the Internet.*

INTRODUCTION: The idea is that distributed database has become an important area of information processing for both organizational and technological reasons: They eliminate many shortcomings of centralized database, & fit more naturally in decentralized structures of many organizations. In centralized database, all the information whether it is important or not are stored on the same database which when required must sometimes not possible to access/use due to some unsustainable circumstances. If due to some unfavorable scenarios if the database goes down, there is major loss with no information on hand. If the user wants to access the information irrespective of the location but majorly focus on content of information required.

Definition: A **distributed database** is a collection of data which belong to the same system but are spread over the sites of computer network. They mainly emphasize on two equal aspects of the distributed database:

1. ***Distribution***:- The fact that the data are not resident of same site or processor that we can distinguish a distributed database from single, centralized database.
2. ***Logical Correlation:-*** It is fact that data have some properties which tie them together so that it is easily distinguishable whether it is from local database or different sites of computer.

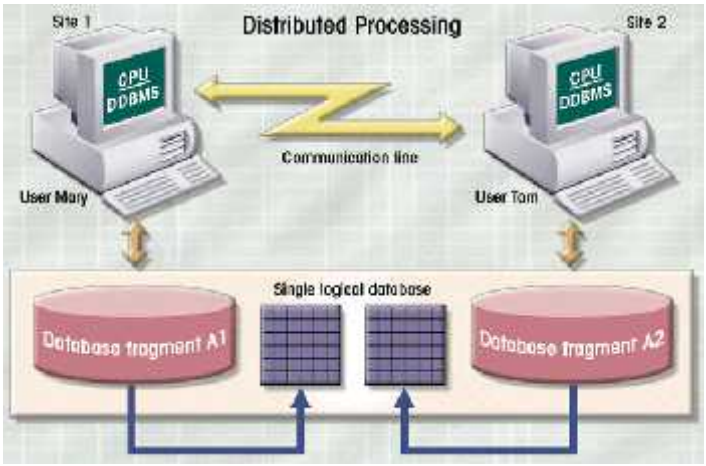
A distributed database (DDB) is a collection of multiple, logically interrelated databases distributed over a computer network. A distributed database management system (distributed DBMS) is the software system that permits the management of the distributed database and makes the distribution transparent to the users The term distributed database system (DDBS) is typically used to refer to the combination of DDB and the distributed DBMS. Distributed DBMSs are similar to distributed file systems (*see* Distributed File Systems) in that both facilitate access to distributed data. However, there are important differences in structure and functionality.

These characterize a distributed database system:

1. Distributed file systems simply allow users to access files that are located on machines other than their own. These files have no explicit structure (i.e., they are flat) and the relationships among data in different files (if there are any) are not managed by the system and are the users responsibility.

2. A distributed file system provides a simple interface to users which allows them to open, read/write (records or bytes), and close files.

3. A distributed DBMS provides *transparent* access to data, while in a distributed file system the user has to know (to some extent) the location of the data. A DDB may be partitioned (called *fragmentation*) and replicated in addition to being distributed across multiple sites. All of this is not visible to the users.



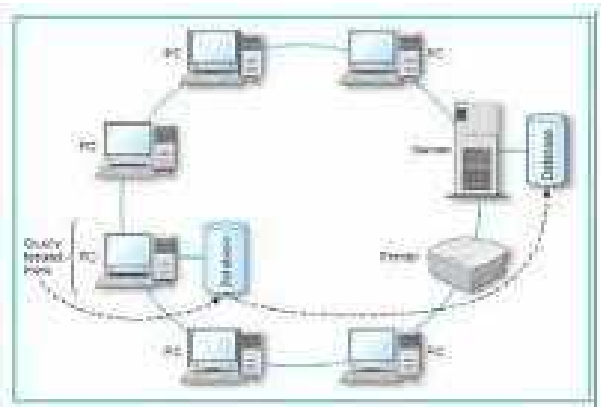
HARDWARE ASPECT:

Two-tier client/server:

Another issue involving the data on a LAN is the fact that some databases can be stored on a client PC's own hard drive while other databases that the client might access are stored on the LAN's server. This is also known as a two-tier approach, (Figure 2). Software has been developed that makes the location of the data transparent to the user at the client. In this mode of operation, the user issues a query at the client, and the software first checks to see if the required data is on the PC's own hard drive. If it is, the data is retrieved from it, and that is the end of the story. If it is not there, then the software automatically looks for it on the server.

Three-tier approach:

In another use of the term three-tier approach, the three tiers are the client PCs, servers known as application servers, and other servers known as database servers, In this arrangement, local screen and keyboard interaction is still handled by the clients, but they can now request a variety of applications to be performed at and by the application servers. The application servers, in turn, rely on the database servers and their databases to supply the data needed by the applications. Though certainly well beyond the scope of LANs, an example of this kind of arrangement is the World Wide Web on the Internet. The local processing on the clients is servers the data input and data display capabilities of browsers such as Netscape's Communicator and Microsoft's Internet Explorer. The application servers are the computers at company Web sites that conduct. the companies' business with the "visitors" working through their browsers. The company application servers in turn rely on the companies' database servers to provide the necessary data to complete the transactions.



Two-tier client/server
Three-tier approach

Distributed Database:

Replication:

REPLICATION is the process of copying and maintaining database objects in multiple databases that make up a distributed database system. Changes applied at one site are captured and stored locally before being forwarded and applied at each of the remote locations. Replication provides user with fast, local access to shared data, and protects availability of applications because alternate data access options exist. Even if one site becomes unavailable, users can continue to query or even update the remaining locations.

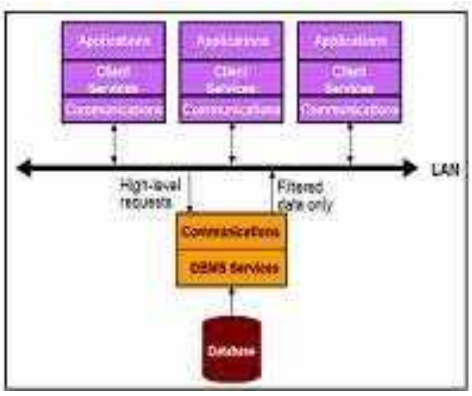
SOFTWARE ASPECT:

In a typical DDBS, three levels of software modules are defined:

1. The server software: responsible for local data management at site.
2. The client software: responsible for most of the distribution functions; DDBMS catalog, processes all requests that require more than one site. Other functions for the client include: consistency of replicated data,

atomicity of global transactions.

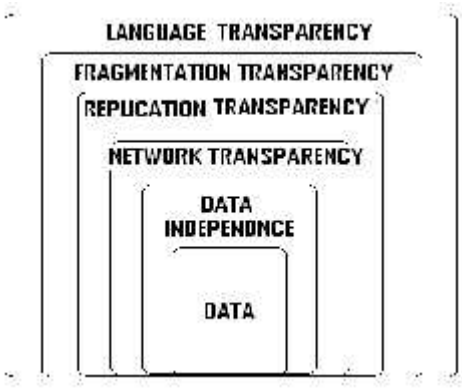
3. The communications software: provides the communication primitives, used by the client/server to exchange data and commands.



Advantages of Client/Server architecture include: More efficient division of labor, horizontal and vertical scaling of resources, better price/performance on client machines, ability to use familiar tools on client machines, client access to remote data (via standards), full DBMS functionality provided to client workstations, and overall better system price/performance

Disadvantages of Client/Server architecture include: server forms bottleneck, server forms single point of failure, and database scaling is difficult.

It is preferable for a DDBMS to have the property of distribution transparency where the user's can issue global queries without knowing or worrying about the global distribution in the DDBS.



Fragmentation: Breaking up the database into logical units called fragments and assigned for storage at various sites.

Types of Fragmentation :

Horizontal: partitions a relation along its tuples,
Vertical: partitions a relation along its attributes,Mixed/hybrid: a co mbination of horizontal and vertical fragmentation.

Data replication: The process of storing fragments in more than one site.

Data Allocation: The process of assigning a particular fragment to a particular site in a distributed system.

The information concerning the data fragmentation, allocation and replication is stored in a global directory.

Correctness Rules of Fragmentation:

- **Completeness:-** Decomposition of relation R into several fragments R1, R2,….., Rn is complete if each data item in base relation R can also be found in some fragmented relation Ri.
- **Reconstruction:-** If relation R is decomposed into fragments R1,R2…..Rn, then there should be some relational operator existing that can reconstruct relation R from its fragments.

  i.e., R = R1∇……..∇Rn

  ∗ Union to combine horizontal fragments.

  ∗ Join to combine vertical fragments
- **Disjointness: -** If relation R is decomposed into fragments R1, R2… Rn and data item di appears in fragment Rj, then di should not appear in any other fragment Rk. K=j.

  (exception: primary key attribute for vertical fragmentation)∗ For horizontal fragmentation, data item is tuple.

∗ For vertical fragmentation, data item is an attribute.

Centralized vs. Distributed database:-

In centralized database, the data or information stored is on centrally located server or computer. Thus, data to be stored on server requires lot of memory and bandwidth with the network connectivity. Data which when fetched requires performing the smooth operation which can be done by maintaining proper bandwidth and network connection.

In distributed database, the data is stored at various locations depending on usage. The data is stored on several sites depending on operation and accessibility of particular data. The data is although stored at different sites but user feels as is working on single homogeneous computer with more re resources is working on single homogeneous computer with more resources depending on operation and accessibility of particular data. The data is although stored at different sites but user feels as is working on single homogeneous computer with more resources.

Distributed systems have been built with the objective of attaining the following:

- Transparency
- Openness
- Reliability
- Performance
- Scalability.

Types of Distributed System:-

There are many different types of distributed systems.

1. Grid Computing
2. Cluster Computing
3. Distributed Database.

GRID COMPUTING:-Grid is defined as parallel and distributed system that is capable of selecting, sharing and aggregating resources dynamically at runtime based on their availability, capability, performance, and cost meeting the users' Quality of Service (QoS) requirements. Drug discovery, economic forecasting, seismic analysis, and back office data processing for e-commerce are a few of the tasks that are commonly solved using grid computing.

- **CLUSTER COMPUTING:-**

A set of computers that are grouped together in such a manner that they form a single resource pool is called a cluster.Any task that has been assigned to the cluster would run on all the computers in the cluster in a parallel fashion by breaking the whole task into smaller self contained tasks.Cluster Computing increases the computation power of the organizations.

- **DISTRIBUTED DATABASE:-**

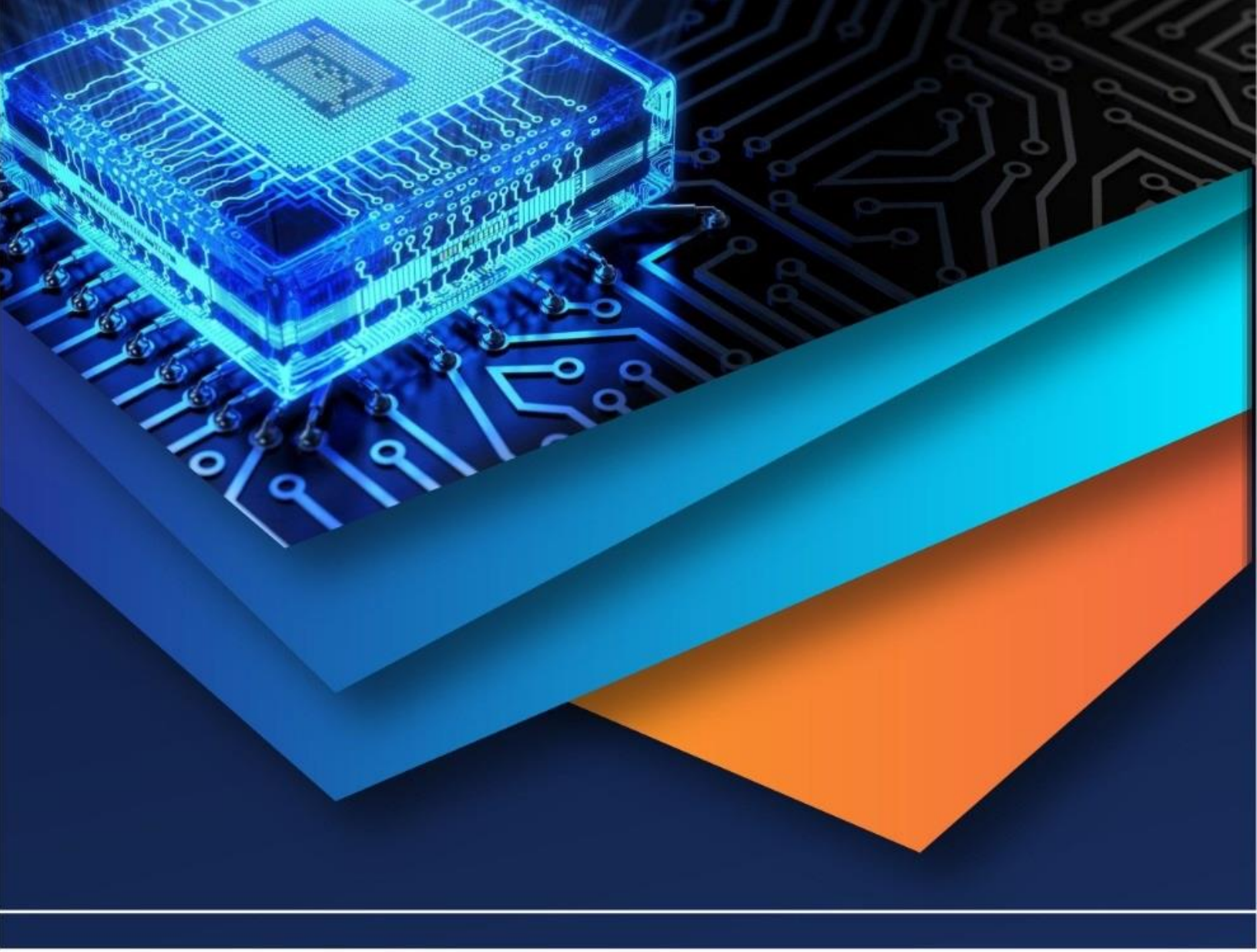
Distributed database system is a collection of independent database systems distributed across multiple computers that collaboratively store data in such a manner that a user can access data from anywhere as if it has been stored locally irrespective of where the data is actually stored.

CONCLUSION:

Through this paper, we want to attract readers towards features and performance factors of distributed database system. We also mentioned. The software architecture being used for the distributed database .We also described Fragmentation, and replication aspect also in order to make readers completely aware about the topic being described here. Besides having a fruitful side of DDBs.

BIBLIOGRAPHY:-

1. Principles of Distributed Database Systems. By M. Tamer Ozsu and Patrick Valduriez.
2. Distributed Database Principles & Systems By Stefano Ceri.
3. Distributed Database System By Chanda Ray.
4. Haroun Rababaah, Distributed Database Fundamentals & Research, Spring 2005.,
5. Swati Gupta1, Kuntal Saroha2, Bhawna, IJCSMS, Vol. 11, Issue 02, Aug 2011.
6. Mohamed Firdhous, IJCIS, Vol. 2, No. 2, 2011



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)