

Distributed Database Problem areas and Approaches

D.S.Hiremath¹, Dr.S.B.Kishor²

¹(hiremath.dayanand@rediffmail.com, Research Scholar Gondwana University, Gadchiroli,)

²(s.b.kishor.spc@gmail.com, Head of Department of Computer Science, S.P.College, Chandrapur

Abstract: The main purpose of this paper is to study and identify the Distributed Database Issues and approaches of Distributed database. Distributed Database is Emerging technology to stored the database in several location or sites, this maintain the Reliability and Availability of the data. In the paper we study the various problems areas in distributed database such concurrency control, distributed database design, transaction management problem etc. Distributed database is a collection of multiple, logically interrelated database distributed over a Computer network. Distributed database would allow to end user to create and stored data anywhere in the network where database is situated, while storing and accessing data from distributed database through computer network there are various problems occurs such deadlock ,concurrency and data allocation using fragmentation and replication to manage these problems it is necessary to design the distributed database carefully manner. The main objective of Distributed database stored the data where it is required.

Keywords– Availability, Computer Network, Distributed Database, Repliation, Reliability

I. INTRODUCTION

Distributed databases is defined as databases located at different machines at the Same or different locations that looks like one centralized database to the end user. Thus, instead of having one centralized database bear the entire load, it is shared by a collection of machines/computers. It is actually a set of server machines working in synchronization to cater the needs to multiple users. These machines in a distributed system are connected to each other either through wireless connection or through various communication media that serve data transfer at high rate. The machines don't have a shared memory nor do they share a clock. The processors in the distributed system may vary from microcomputers to work station to mini computers to computers used in day to day life. The distributed database can be shown as:

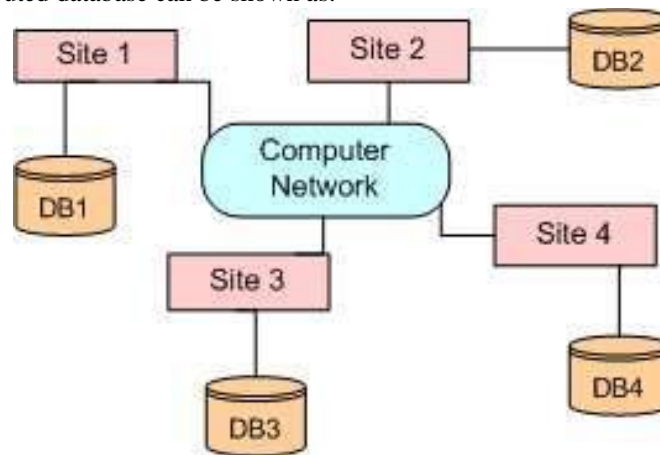


Fig.1: Distributed Database

Distributed database is very useful now due to promising feature to manage a database, Data is most important property of any organization to secure and proper management of organizational data Distributed database is helpful.

II. ADVANTAGES OF DISTRIBUTED DATABASE

The Advantages of Distributed Database is as follows:-

- a) **Availability:** - The data is replicated at multiple sites. In case the local server is unavailable due to some reason, the data can be retrieved from the other available server.

- b) **Robustness:-** The entire system becomes more robust as multiple servers are involved in handling data. Thus, failure of one system doesn't lead to failure of entire system.
- c) **Sharing:** Data at multiple sites is shared by users at different sites
- d) **Performance:** It leads to improved performance as many machines are involved, the load is distributed. The database is divided into database fragments, thus local queries can be resolved by local databases rather than all queries being targeted to one centralized database. Thus, the query Processing time is reduced and performance is increased.
- e) **Ease of growth:** To add more clients to such a system is quite easy as Overloading is never an issue.
- f) Management of distributed data with different levels of transparency, Hardware, Operating System, Network and Location Independence. It provides Continuous operation.
No more reliance on the central site.

III. Disadvantages

- a) **Complexity:-**DBAs may have to do extra work to ensure that the distributed nature of the system is transparent. Extra work must also be done to maintain multiple disparate systems, instead of one big one. Extra database design work must also be done to account for the disconnected nature of the database for example, joins become prohibitively expensive when performed across multiple systems.
- b) **Economics :-** increased complexity and a more extensive infrastructure means extra labour costs
- c) **Security:** Remote database fragments must be secured, and they are not centralized so the remote sites must be secured as well. The infrastructure must also be secured (for example, by encrypting the network links between remote sites.
- d) **Difficult to maintain integrity :-** but in a distributed database, enforcing integrity over a network may require too much of the network's resources to be feasible
- e) **Inexperience :-** distributed databases are difficult to work with, and in such a young field there is not much readily available experience in "proper" practice
- f) **Lack of standards:-** there are no tools or methodologies yet to help users convert a centralized DBMS into a distributed DBMS
- g) **Database design more complex :-** In addition to traditional database design challenges, the design of a distributed database has to consider fragmentation of data, allocation of fragments to specific sites and data replication
- h) Additional software is required
- i) Operating system should support distributed environment

IV. Problems Areas of Distributed Database

Following are the Problems Areas of Distributed database. :-

- 1) **Distributed Concurrency Control:** - Distributed Concurrency Control specifies that synchronization of access to the distributed database such that the integrity of the database is maintained. To maintain Concurrency in distributed database different locking techniques should be used which is based on mutual exclusion of access to data. Time stamping algorithm also used where transactions are executed in some order [1].
- 2) **Distributed Deadlock Management :-** In distributed database several users are request for resources from the database if the resources are available at that time , then database grant the resources to that user if not available the user has to wait until the resources are released by other user. Sometimes the users are not released the resources are blocked by some other user. This situation is known as Deadlock. Distributed Deadlock is managed using the different algorithm and techniques such as avoidance and detection algorithm.
- 3) **Replication Control:** - Replication is a technique that only applies to distributed systems. A database is said to be replicated if the entire database or a portion of it (a table, some tables, one or more fragments, etc.) is copied and the copies are stored at different sites. The issue with having more than one copy of a database is maintaining the mutual consistency of the copies—ensuring that all copies have identical schema and data content [2].
- 4) **Operating Environment:** - To Implement Distributed Database Environment a Specific Operating System is requirement as per Organizational needs. Operating system plays an important role for managing the distributed database. Some time Operating system is not supported for Distributed database.
- 5) **Transparent Management:** - Transparent management of Data is one of the major problem areas in Distributed database. In Distributed database data is situated in multiple locations and number

of users are used that database. To maintain the integrity of database transparent management of data is important.

- 6) **Security and privacy:** - How to apply the security policies to the interdependent system is a great issue in distributed system. Since distributed systems deal with sensitive data and information so the system must have a strong security and privacy measurement. Protection of distributed system assets, including base resources, storage, communications and user-interface I/O as well as higher-level composites of these resources, like processes, files, messages, display windows and more complex objects, are important issues in distributed system
- 7) **Resource management:** - In distributed systems, objects consisting of resources are located on different places. Routing is an issue at the network layer of the distributed system and at the application layer. Resource management in a distributed system will interact with its eterogeneous Nature.

V. Distributed Database Architecture Strategies:

a) Top Down Approach:-

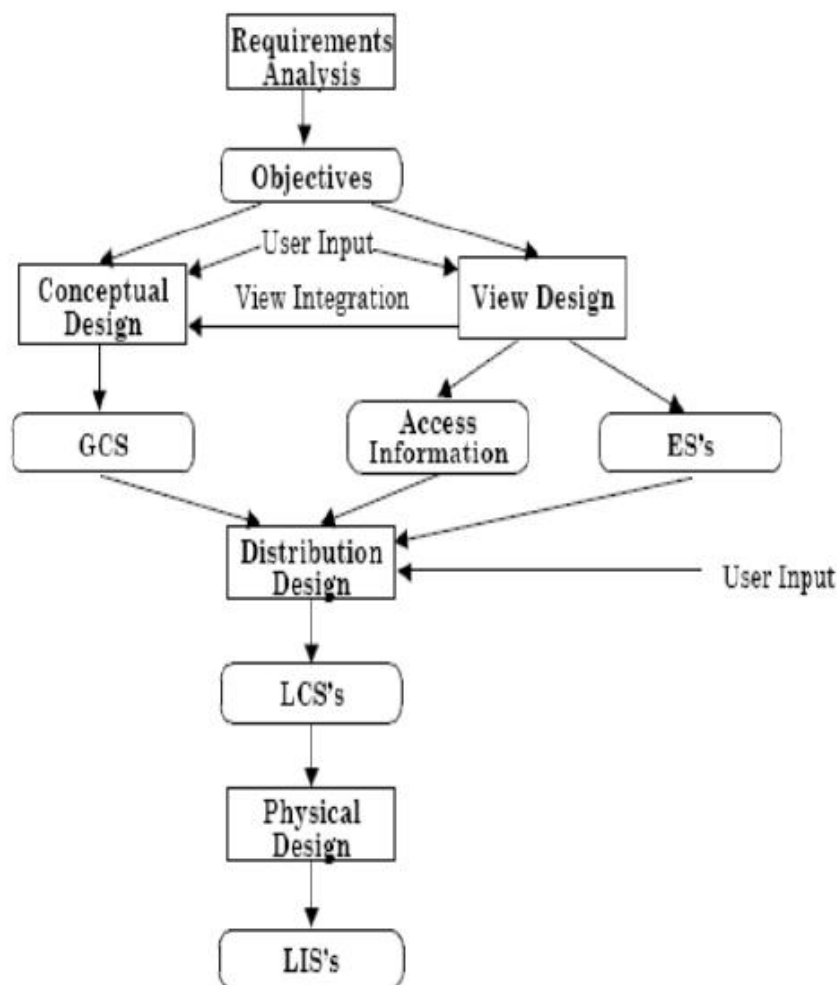


Fig. 2 Top down Approach

Top Down approach is mostly used when we have implement Distributed database from beginning. The process starts from a requirement analysis phase including analyzing of the company situation, defining problems and constraints, defining objectives, and designing scope and boundaries. The next two activities are conceptual design and view design. Focus on the data requirements, the conceptual design deals with entity relationship modeling and normalization [6].

The view design activity deals with defining the interfaces for end users. The Conceptual design is the process by which determine entity types and relationships among these entities. Entity analysis is used

to determining the entities and their attributes and the relationships among them. Functional analysis is concerned with determining the fundamental functions with which the modeled organization is involved. View Integration activity is used to determine conceptual model should support not only the existing applications but also future applications [5]

b) Bottom up Approach: - This type of Approach is used when Distributed Database is already exist and we have to add another database in existing environment. In this environment we can enhance additional features in to the existing database.

The important aspect of design straties is determine how to integrate multiple database system together

VI. CONCLUSION:-

Distributed database is one of emerging technology research field and business organization. In this Paper we discuss various problem areas and advantages and disadavantages of Distributed database. The problem areas mentioned in the paper are very useful while implementing distributed database so that concurrency, deadlock, security and privacy is easily managed. In this paper we also study architecture of distributed database for implementing Distributed database environment in business environment. In short difficulties encountered while implementing Distributed database must be carefully handle aginst its advantages over Centralized database system.

References:

- [1]. Andrew S Tanenbaum, "Distributed systems principles and paradigms", PHI
- [2]. Ceri, S Pelagatti, G, "Distributed Databases: Principles and Systems"
- [3]. DBMS – Complete Practical Approach by Sharad Maheshwari, Ruchin Jain
- [4]. George Coulouris, "Distributed System", ean Dollimore Publication.
- [5]. M Tamer Ozsu, "Principle of Distributed Database System" PHI
- [6]. <http://www.ijeset.com/media/11N5-IJES0202505>
- [7]. Saeed K.Rahimi and Frank S.Haug, "A Distributed Database Management System a practical approach", Wiley Publication.
- [8]. Stefano Ceri, "Distributed database principals systems by", TMH
- [9]. William Buchanan, "Distributed systems and networks", TMH
- [10]. Wesley W. Chu, "Distributed Systems: Vol. II: Distributed Data Base Systems",