An Overview of Distributed Database System's Advantages and its Problem Areas

Shital Jadhav[1], Ashwini Gawande[2]

Department of Computer Sceience, Sinhagd College of Science, Pune, Maharashtra, India

ABSTRACT: A distributed database (DDB) is a collection of multiple, logically interrelated databases distributed over a computer network. [2].The goal of distributed database system is to achieve data integration and data distribution transparency. Data integration means overall completeness, accuracy & consistency of data. Distribution transparency means that the DDBMS hides all the added complexities of distribution, allowing users to think that they are working with a single centralized system.. DDBMS is a Software system that permits the management of the distributed database and makes the distribution transparent to users. It is adapted from Bell & Grim son in 1992. This paper presents an overview of Distributed Database System along with their advantages and various problem areas. This paper also provides various aspects like replication, fragmentation.

KEYWORDS: Database, Deadlock, Distributed Database Management System, Fragmentation, Replication, Concurrency Control

I. INTRODUCTION

In today's world of universal dependence on information systems, all sorts of people need access to companies' databases. In addition to a company's own employees, these include the company's customers, potential customers, suppliers, and vendors of all types. It is possible for a company to have all of its databases concentrated at one mainframe computer site with worldwide access to this site provided by telecommunications networks, including the Internet. Although the management of such a centralized system and its databases can be controlled in a well-contained manner and this can be advantageous, it poses some problems as well. For example, if the single site goes down, then everyone is blocked from accessing the databases until the site comes back up again. Also the communications costs from the many far PCs and terminals to the central site can be expensive. One solution to such problems, and an alternative design to the centralized database concept, is known as distributed database. The idea is that instead of having one, centralized database, we are going to spread the data out among the cities on the distributed network, each of which has its own computer and data storage facilities. [3]. A Database is systematically organized or structured repository of indexed information that allows easy retrieval, updating, analysis, and output of data. Each database may involve different database management systems and different architectures that distribute the execution of transactions [1].A DDBMS is not a "collection of files" that can be stored at each node of a computer network. A multiprocessor system based DBMS (parallel database system) is not a DDBMS.A DDBMS is not a system wherein data resides only at one node

II. ADVANTAGES AND DISADVANTAGES

 The distribution of data and applications has promising advantages. Although they may not be fully satisfied by the time, these advantages are to be considered as objectives to be achieved.

Local Autonomy: Since data is distributed, a group of users that commonly share such data can have it placed at the site where they work, and thus have local control. By this way, users have some degree of freedom as accesses can be made independently from the global users.

Improved Performance: Data retrieved by a transaction may be stored at a number of sites, making it possible to execute the transaction in parallel. Besides, using several resources in parallel can significantly improve performance.

Improved Reliability/Availability: If data is replicated so that it exists at more than one site, a crash of one of the sites, or the failure of a communication line making some of these sites inaccessible, does not necessarily make the data impossible to reach. Furthermore, system crashes or communication failures do not cause total system not operable and distributed DBMS can still provide limited service.

Economics: If the data is geographically distributed and the applications are related to these data, it may be much more economical, in terms of communication costs, to partition the application and do the processing at each site. On the other hand, the cost of having smaller computing powers at each site is much more less than the cost of having an equivalent power of a single mainframe.

Expandability: Expansion can be easily achieved by adding processing and storage power to the existing network. It may not be possible to have a linear improvement in power but significant changes are still possible.

Shareability: If the information is not distributed, it is usually impossible to share data and resources. A distributed database makes this sharing feasible.

Security: Security can be easily controlled in a central location with the DBMS enforcing the rules. However, in distributed database system, network is involved which it has its own security requirements and security control becomes very complicated. [4]

Disadvantages
Complexity: A distributed database is more complicated to setup and maintain as compared to central database system.
Security: There are many remote entry points to the system compared to central system leading to security threats.
Data Integrity: In distributed system it is very difficult to make sure that data and indexes are not corrupted.
In distributed database systems, data need to be carefully placed to make the system as efficient as possible.

III. PROBLEM AREAS IN DDBMS

Distributed database design: To minimize the cost of data storage and transaction processing and how to the fragment distribute the database and application program. The two fundamental design issues are fragmentation, the separation of the database into partitions called fragments, and distribution, the optimum distribution of fragments.

Query Processing: The problem is how to decide on a strategy for executing every query over the network in the most cost effective way. Query processing deals with designing algorithms that analyze queries and convert them into a series of data manipulation operations.

Distributed Concurrency Control: It indicates to maintain the integrity of database i.e. when multiple transaction are executing concurrently on a particular site.

Directory Management: It indicates how to manage meta-data (Data about Data) for entire database.

Deadlock Management: It is more complex to acquire or prevent and detect deadlock situation in distributed environment. The competition among users for access to a set of resources (data, in this case) can result in a deadlock

Reliability : To ensure the consistency of the database as well as to detect failures and recover from them. The implication for DDBSs is that when a failure occurs and various sites become either inoperable or inaccessible, the databases at the operational sites remain consistent and up to date.

Operating System Support: Problem is traditional OS does not provide support necessary for all DDBMS operations.

IV. CONCLUSION

Design and implementation of DDBMS involves many challenges and choices. This paper has attempted to presents different aspects of DDBMS. Distribution of data has its own advantages and disadvantages. This paper presents a complete review on distributed databases. It is clear from the study that distribution of data involves the problem of deadlock. We need to find out the methods to data distribution and accessing which leads to minimization of deadlock and thus resulting in proper utilization of resources.

REFERENCES

[1] Www.Businessdictionary.Com/Definition/Database.Html
[2] M.T. Özsu And P. Valduriez. Principles Of Distributed Database Systems.   Englewood Cliffs, Nj: Prentice-Hall, 1991
[3] Distributed Databases Fundamentals And Research   By Haroun Rababaah
[4] An Overview Of Distributed Database Management By Ayse Yasemin Seydim
[5] An Overview Of Distributed Database Management System By B Johnsirani And M   Natarajan