Distributed Database Management Systems for Information Management and Access

N Geetha

Abstract

Libraries play an important role in the academic world by providing access to world-class information resources and services. The explosive growth of Information Technology has paved way to unprecedented phenomenon of access to information. The evolving developments of the recent decades in information access and management have resulted in a situation where information is stored and managed by a large variety of systems around the world. The growing requirements for information support to the globalised socio-economic forces makes efficient and effective availability of distributed information and correlation of relevant data to the user need. Databases and database systems have become essential components of everyday life in our modern society. In our daily lives, most of us would engage in one activity or another that involves some interaction with a database. This paper discusses the technology used to access and manage the distributed information.

Key words : Database, Database Management System, Distributed Database Management System, Data Access, Client/Server Systems, Client/Server Architecture.

0. Introduction

Distributed database technology is one of the most important developments of the past decades. The maturation of database management systems (DBMS) technology has coincided with significant developments in distributed computing results the emergence of Distributed Database Management Systems (DDBMS). Distributed database is a logical database that is physically divided among computers at several sites on a network. A distributed database management system is a database management system capable of supporting and manipulating distributed databases. The distributed database management systems have started to become the dominant management tools for highly intensive applications.

The basic motivations for distributed databases are improved performance, increased availability, shareability, expandability and access facility. Computers in a network communicate through messages. Accessing the data using messages over network is slower than the accessing data on disk. In general to access the data rapidly in a distributed databases we must attempt to minimize the numbers of messages. Distributed database management system manages applications based on data access from different sources at multiple locations. As institutions implement networked information strategies which call for sharing and licensing access to information resources in the networked environment. Access management have emerged as major issues, which threaten to impede progress.

1. Information Access

Information access to the west is much quicker and less dangerous than those days. Today's information highway system had been available to the prospectors of the gold rush. All the data in the enterprise are useless unless they are accessible to the relevant users and processes that can become into knowledge.

The distributed database system allows databases to be spread over a network of multiple processors and supports access to those databases by any application running on the network. It also allows distribution of data and applications for faster response time and distributed control and provides communication services across processor nodes.

2. Distributed Database Management Systems (DDBMS)-Definition

A distributed database is defined as a collection of simple, logically interrelated databases distributed over a computer network. A distributed database management system is then defined as the software system permits the management of the distributed databases and makes this distribution transparency to the users. Distributed database system is to referred as a combination of distributed databases and distributed database management system. The two important terms in these definitions are “logically interrelated databases distributed over a computer network”.

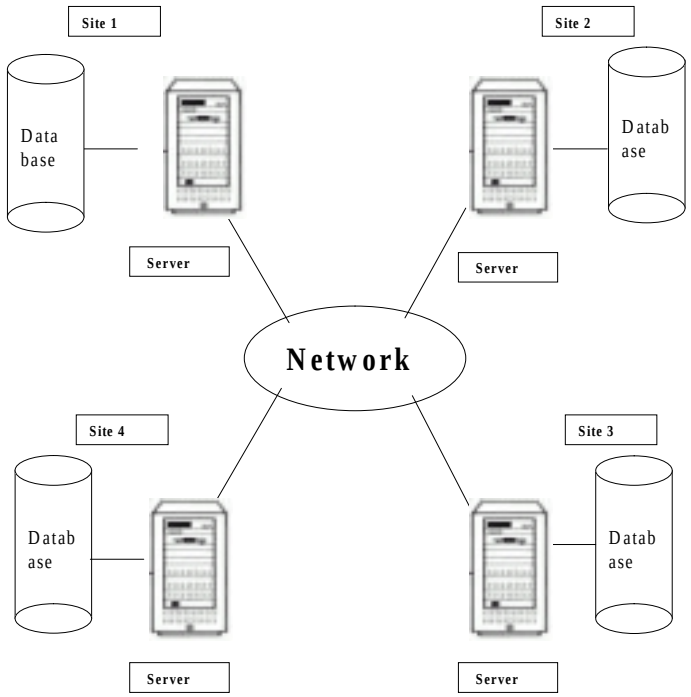
2.1 Characteristics of Distributed Database Management Systems

A Distributed database management system can be Homogeneous (same local DBMS at each site) or Heterogeneous (different local DBMS). Heterogeneous DBMSs are more complex and more difficult to manage. The DDBMS share the characteristics of location transparency, replication transparency, and fragmentation transparency.

Location transparency is the characteristic that states that users do not need to be aware of the location of data in a distributed database. Replications let users at different sites use and update copies of a database and then share their updates with other users. Replication transparency refers to the characteristic that a DDBMS should update various copies of data behind the scenes; users should be unaware of the steps. A Distributed Database Management Systems supports data fragmentation is the DDBMS can divide and manage a logical object, such as records in a table, among the various locations under its control. If users are unaware of fragmentation, the DDBMS has fragmentation transparency.

2.2 Functions of a DDBMS

- Application interface
- Validation to analyze data requests
- Transformation to determine request’s components
- Query-optimization to find the best access strategy
- Mapping to determine the data location
- I/O interface to read or write data
- Formatting to prepare the data for presentation
- Security to provide data privacy
- Backup and recovery
- DB administration
- Concurrency control
- Transaction management



Distributed Database Management System (Diagram)

2.3 Rules for Distributed Databases

C.J.Date has formulated 12 rules for Distributed databases. The rules are as follows:

1. Local autonomy - No site should depend on another site to perform its functions.
2. No reliance on a central site - A DDBMS should not need to rely on one site more than any other site.
3. Continuous operation - Performing any function should not shut down the entire distributed database.
4. Location transparency - Users should feel as if the entire database is stored at their location.
5. Fragmentation transparency - Users should feel as if they are using a single central database.
6. Replication transparency - Users should not be aware of any data replication.
7. Distributed query processing - A DDBMS must process queries as rapidly as possible even though the data is distributed.
8. Distributed transaction management - A DDBMS must effectively manage transaction updates at multiple sites.
9. Hardware independence - A DDBMS must be able to run on different types of hardware.
10. Operating system independence - A DDBMS must be able to run on different operating systems.
11. Network independence - A DDBMS must be able to run on different types of networks.
12. DBMS independence - A DDBMS must be heterogeneous.

2.4 Advantages of Distributed Database Management Systems

- 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 application 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.
- Expandibility : 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. On the other hand, distribution of the database can cause several problems.

2.5 Disadvantages of Distributed Database Management Systems

- Lack of Experience : Some special solutions or prototype systems have not been tested in actual operating environments. More theoretical work is done compared to actual implementations.
- Complexity : Distributed DBMS problems are more complex than centralized DBMS problems.
- Cost : Additional hardware for communication mechanisms are needed as well as additional and more complex software may be necessary to solve the technical problems. The trade-off between increased profitability due to more efficient and timely use of information and due to new dataprocessing sites, increased personnel costs has to be analyzed carefully.
- Distribution of Control : The distribution creates problems of synchronization and coordination as the degree to which individual DBMSs can operate independently.
- 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 securityrequirements and security control becomes very complicated.
- Difficulty of Change : All users have to use their legacy data implemented in previous generation systems and it is impossible to rewrite all applications at once. A distributed DBMS should support a graceful transition into a future architecture by allowing old applications for obsolete databases to survive with new applications written in current generation DBMSs.

In spite of the problems and its complexity, the users that will mostly benefit from the distributed DBMSs.

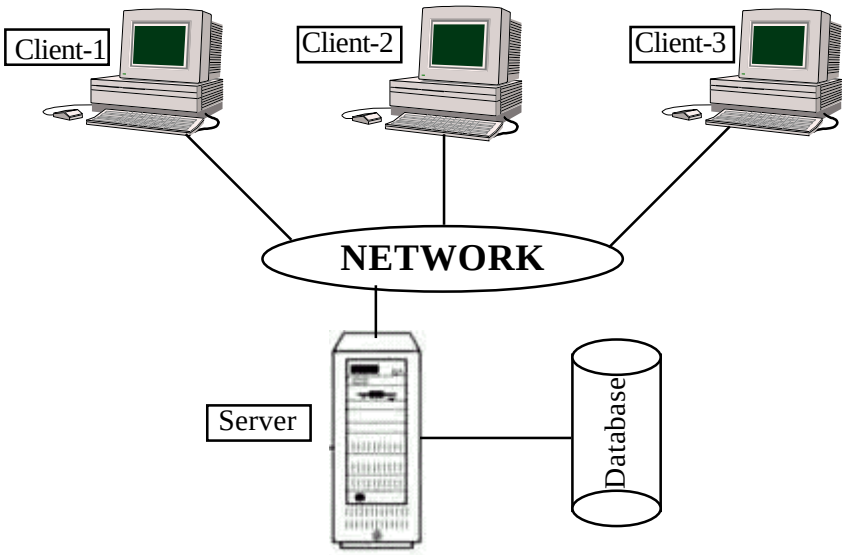
2.6 An Example for Distributed Database Management System

2.6.1 Client/Server Systems

The term client/server was first used in the 1980s in reference to personal computers (PCs) on a network. The actual client/server model started gaining acceptance in the late 1980s. The client/server software architecture is a versatile, message-based and modular infrastructure that is intended to improve usability, flexibility, interoperability and scalability. A client is defined as a requester of services and a server is defined as the provider of services. A single machine can be both a client and a server depending on the software configuration. In a network environment, a file server stores files required by users on the network. When users need data from a file, the entire file is sent.

In client/server architecture, the server is a computer providing data to the clients, which are the computers that are connected to a network and that people use to access data stored on the server. The DBMS runs on the server and a client sends a request for specific data to the server. Only the necessary data and not the entire file or files are sent. Client/server architecture may be either two-tier or three-tier. In a two-tier architecture, the server performs database functions and the clients perform the presentation (user interface) functions. Either the server or the clients may perform business functions. The term fat client refers to an arrangement where the clients perform the business functions. If the business functions reside on the server, each client is called a thin client. In a three-tier architecture, the clients perform the presentation functions, a database server performs the database functions, and separate computers, called application servers, perform the business functions and act as interface between clients and database server. The advantages to using a client/server system instead of a file server are: lower network traffic; improved processing distribution; thinner clients; greater processing transparency; increased network, hardware, and software transparency; improved security; decreased costs; and increased scalability. Triggers, which are actions that occur automatically in response to associated database operations, provide additional integrity support.

2.6.2 Client/Server Architecture



3. Conclusion

Advances in information technology have changed the entire gamut of Library and Information service today. It is time to develop new means of information access and resource sharing with the help of fast growing computer and telecommunication technologies. The libraries serve three roles in learning. They serve practical role in sharing expensive sources both physical and human resource and cultural role in preserving and organizing facts and ideas. Third the libraries serve social and intellectual role by bringing together people and ideas. The Libraries also serve as centers of interdisciplinary places shared by learners from all disciplines.

Digital libraries extend such interdisciplinary approach by making diverse information resources available beyond physical space shared by a group of learners. However explosion of data and increased cost of information, force the libraries to work together, and technological advancements acts as platform for accessing and sharing the diversified information. The Library and Information Science profession is preparing for adapting itself to the new environment created by information technology. High performance networks have become key ingredients of competitive scientific research. Research projects especially interdisciplinary ones often require high-speed access to massive data archive and analysis facilities.

4. References

1. Principles of Database systems by M.T.Ozsu, P.Valduriez.
2. The Push/Pull Effect - Can Distributed, Database Technology Meet The Challenges of New, Applications?Ó Database Programming & Design by M. T. Ozsu.
3. http://www.engr.smu.edu
4. http://www.edtech.uis.edu
5. Proceedings of National Seminar on Information Access, Management and Exchange in the technological age, May 2002.
6. Distributed database systems Bell D

About Author

Mrs. N. Geetha is Assistant Librarian at Anna University, Chennai-600 025, India. She holds Masters in Zoology, LIS and undergone training in British Council Library, Chennai.
E-mail : geethkowsalya@hotmail.com, aulib@annauniv.edu