THE IMPACT OF DISTRIBUTED DATABASES IN E-LEARNING SYSTEMS

IACOB NICOLETA – MAGDALENA, Ph.D. candidate
Faculty of mathematics - computer science / department of computer science, University of Pitesti

DEFTA COSTINELA – LUMINIȚA University Asist. Ph.D. candidate
Faculty of Tourist and Commercial Management, „Dimitrie Cantemir” Christian University of Bucharest

Abstract

This paper presents the advantages of e-Learning education and focus on the importance of data replication in the design of a distributed database system. In addition, we analyze the methods of data replication, for improving performance and increasing data availability. The purpose of this paper is to present the advantages of using distributed databases in e-learning systems by universities with geographically distributed offices.

Keywords: e-Learning, distributed databases, fragmentation, replication, allocation, methods, strategies

1. Introduction

The distributed systems are attractive for the users with various, interactive activities that require large processing capabilities, resource sharing and high availability. The distributed nodes of information systems allow replication of data between data centers, ensuring maximum availability of resources.

Academic activities are targeted overall towards socio-economic realities and various changes in contemporary society. In this context, the universities provide training for future generations in an adapted environment for society to which they belong.

The evolution of information technology generated also a tremendous evolution of database systems and in this context distributed database technology has changed the centralized point of view by offering major advantages.

A distributed database (DDB) is a collection of shared data, logically interrelated, which are geographically distributed in a computer network (Iacob, 2010).

---

1 This work was partially supported by the strategic grant POSDRU/88/1.5/S/52826, Project ID52826 (2009), co-financed by the European Social Fund – Investing in People, within the Sectoral Operational Programme Human Resources Development 2007-2013.
A distributed database management system (DDBMS) is a set of programs that allows the management of distributed database and makes the distribution transparent to the users. The objective of transparency is to make a distributed system appear similar to a centralized system. This is usually called the fundamental principle of DDBMS.

A distributed database system is a database system which is fragmented or replicated on the various configurations of hardware and software, located usually at different geographical sites within an organisation (Beynon-Davies, 2004).

The evolution of modern information technologies and the Internet have changed traditional education and training methods resulting in reorganization and transformation of those in order to provide opportunity to develop student’s level of professional training, irrespective of age, sex or geographical area. Using new technologies, any student can communicate with specialists in their area of interest or other colleagues for exchanging information, knowledge and experience.

E-learning is the interaction between teaching and learning, information technology and communications, covering a wide spectrum of activities from computer aided education to education conducted entirely online.

E-learning incorporates traditional or modern learning methods and techniques and use computers to deliver them to the students - multimedia processing and asynchronous or synchronous communication. E-learning courses include both content (that is, information) and instructional methods (that is, techniques) that help people learn the content (Clark & Mayer, 2008).

Asynchronous e-learning can be defined as a form of computer based training, in which the teacher (trainer) and the course participants do not use educational resources at the same time, hence the name asynchronous e-learning.

Synchronous e-learning can be defined as a form of computer based training, in which teacher (trainer) and the course participants use educational resources at the same time hence the name synchronous e-learning. The students can log into the system and can interact with the teacher and other students that are in the same location or different locations (as far as the teacher allows it) in real time.

2. Distributed database

Distributed database design: The methodology used for the logical design of a centralized database applies to the design of the distributed one as well. However, for a distributed database three additional factors have to be considered: data fragmentation, data replication and data allocation.

2.1. Data fragmentation

Before we decide how to distribute the data we must determine the logical units of distribution. The database may be broken up into logical units called fragments which will be stored at different sites. The simplest logical units are the tables themselves.

Fragmentation aims to improve:
- Reliability;
- Performance;
- Communication costs;
- Security.

There are two types of fragmentation: horizontal fragmentation and vertical fragmentation.

- **Horizontal fragmentation:** A horizontal fragment of a table is a subset of rows in it. So horizontal fragmentation divides a table "horizontally" by selecting the relevant rows and these fragments can be assigned to different sides in the distributed system.
- **Vertical fragmentation:** a vertical fragment of a table keeps only certain attributes of it. It divides a table vertically by columns. It is necessary to include the primary key of the table in each vertical fragment so that the full table can be reconstructed if needed.
- **Mixed fragmentation:** is achieved by applying successive the operations for horizontal and vertical fragmentation. Such, mixed fragments can be obtained by applying horizontal fragmentation to a vertical fragment, respectively by applying a vertical fragment to a horizontal fragment. In practice it is recommended to have at most two levels of fragmentation.

2.2. Data replication

A copy of each fragment can be maintained at several sites. Replication is the operation of storing portions from a database, as copies, on
multiple nodes on a network. If a user updates a local copy then DDBMS automatically updates all copies of that data. The fragmentation and the replication can be combined: A relationship can be partitioned into several pieces and can have multiple replicas of each fragment (Silberschatz, Korth & Sudarshan, 2010).

The purposes of replication are multiple:
- **System availability.** Even when some sites are down, data may be accessible from other sites.
- **Performance.** Replication enables us to locate the data closer to their access points, that contributes to a reduction in response time.
- **Scalability.**
- **Application requirements.** Finally, replication may be dictated by the applications, which may wish to maintain multiple data copies as part of their operational specifications (Ozsu & Valduriez, 2011).

Replication is a process that consists in making and distributing copies of data between different places, at remote or mobile users, by using Internet. In addition, allows changes to be propagated consistently to the relevant copies. Distribution of these replicas must ensure that the data processing is done locally, in order to maximize efficiency and reduce communication costs. Furthermore, replication provides more reliability, minimizes the chance of total data loss, and greatly improves disaster recovery (Rahimi & Haug, 2010).

2.2.1. Replication methods

The methods that can be used for replication (Table 1) are:
- **not replicated** data: DDBMS allocates space for some data on a single node from a computer network. The characteristics of this method are: minimal redundancy, data access concurrency is maximum, update time is small and retrieval time is great.
- **partitioned:** each fragment resides at only one site.
- **partially replicated data:** DDBMS allocates for a part of data a single copy on a computer, and for another part of data multiple copies on multiple computers on the network. The characteristics of this method are: redundancy is increased, concurrency access to data is decreased, update and retrieval time is medium.
- **fully replicated** data: DDBMS allocates for the entire database multiple copies on different computers on the network. The characteristics of this method are: redundancy is high, concurrency access to data is minimal, the time update is great and retrieval time is small.

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods of data replication</td>
</tr>
<tr>
<td>Data consistency</td>
</tr>
<tr>
<td>Storage costs</td>
</tr>
<tr>
<td>Reliability and availability</td>
</tr>
<tr>
<td>Update speed</td>
</tr>
<tr>
<td>Communication costs</td>
</tr>
<tr>
<td>Redundancy</td>
</tr>
<tr>
<td>Concurrency access to data</td>
</tr>
<tr>
<td>Update time</td>
</tr>
<tr>
<td>Retrieval time</td>
</tr>
</tbody>
</table>

The replication of a fragment \( R_i \) at \( j \) station, formula (1), is justified if the fragment allocation cost is less than the cost of remote accessing the applications \( k \) from \( j \) station to other copies of the fragment found on other stations \( j' \) (\( j' \neq j \)):

\[
C_{ij} < \sum_{k} d_{kij}
\]

2.3. Data allocation

Each fragment has to be allocated to one or more sites, where it will be stored.
There are three strategies regarding the allocation of data:

- **Centralized**: Consists of a single DB and DBMS stored at one site with users distributed across the network.
- **Fragmented (or Partitioned)**: The database is partitioned into disjoint fragments, with each fragment assigned to one site (no replication).
- **Complete replication**: A complete copy of the database is maintained at each site (no fragmentation). Here, storage costs and communication costs for updates are most expensive.
- **Selective replication**: A combination of fragmentation and replication.

3. Distributed queries

A user interacts with the database by providing read and write requests in the database. These read and write operations are grouped into transactions. In a computer network it is very important not to allow the unauthorized persons to gain access to the information sent between computers (Defta, 2010).

The global users working with DDB would work as with a centralized database:

- make requests (global transactions);
- requests are assessed, decomposed and delivered to distributed execution supervisor;
- the supervisor sends the requests to nodes in which there are local databases that will be queried;
- queried databases send replies to the supervisor.

3.1. Distributed query optimization

The problem of optimizing execution of distributed queries must be addressed in order to obtain a system response time as low as possible in terms of minimizing the total cost of implementation.

The total cost of a query execution in a distributed system, formula (2), includes:

- CCPU – the cost of CPU processing for the execution of data operations in main memory,
- CI/O - the costs of accessing (input/output) physical data on the respective support,
- communication costs between nodes associated with sending and receiving messages between nodes (CMSG) and the transfer of data between nodes (CTR).

\[
(2) \text{Total cost} = \text{CCPU} \times \text{no instructions} + \text{CI/O} \times \text{no I/O} + \text{CMSG} \times \text{no messages} + \text{CTR} \times \text{data quantity}
\]

The response time of the distributed system, formula (3), is calculated from the time the query starts until it receives response from the system.

\[
(3) \text{Response time} = \text{TCPU} \times \text{no instructions} + \text{TI/O} \times \text{no I/O} + \text{TMSG} \times \text{no messages} + \text{TTR} \times \text{data quantity}
\]

The response time minimization is achieved by increasing the parallelism degree of distributed query execution. This does not necessarily imply that the total cost will be minimized. Contrary, the total cost may increase when trying to increase the degree of parallelism of the execution and transmission. On the other hand, minimizing the total cost involves better use of resources in the detriment of response time which will increase. In practice, it is desirable to do a compromise between the two objectives.

4. Conclusion

Distributed databases have appeared as a necessity, because they improve availability and reliability of data and also high performance in data processing by allowing parallel processing of queries, and reduce processing costs.

The use of distributed databases in e-learning systems has the goal to improve access to information and also rapid data collection.

The aim is the creation of an educational network based on e-learning tools which allows greater flexibility in the training of persons in terms of efficiency in accordance with national standards.

The main objective of improving the e-Learning technology consists in implementation of educational applications that will offer greater flexibility, better reliability, a new set of features and also increased communications security over Internet.
References