A Review on Fragmentation Techniques in Distributed Database

Akashkumar Patel\textsuperscript{1}, Rakshitkumar Hirapara\textsuperscript{2}, Vivekkumar Dhamecha\textsuperscript{3}
\textsuperscript{1}UG Student, G. H. Patel College of Engineering & Technology, patelakash317@yahoo.com
\textsuperscript{2}UG Student, G. H. Patel College of Engineering & Technology, rakshithirapara@gmail.com
\textsuperscript{3}UG Student, G. H. Patel College of Engineering & Technology, vivekdhamecha@rediffmail.com

Abstract—The distributed database systems are developed for balancing the load and scattering the data over different sites on an organization. So in order to distribute the database on different sites of an organization, fragmentation methods are used. There are several fragmentation methods reviewed in this article.

Keywords—Fragmentation, Distributed Database System, Horizontal fragmentation algorithm, Vertical fragmentation algorithm, Hybrid fragmentation.

I. INTRODUCTION

Data fragmentation allows you to break a single object into two or more segments or fragments. The object might be a user’s database, a system database, or a table. Each fragment can be stored at any site over a computer network. Information about data fragmentation is stored in the distributed data catalog (DDC), from which it is accessed by the TP to process user requests.

Distributed processing is an effective way to improve reliability and performance of a database system. Distribution of data is a collection of fragmentation, allocation and replication processes. The purpose of this work is to present an introduction to

Distributed Databases which are becoming very popular now days with the description of distributed database environment, fragmentation and horizontal fragmentation technique. Horizontal fragmentation has an important impact in improving the applications performance that is strongly affected by distributed databases design phase

II. BACKGROUND THEORY

2.1 Distributed Database System

A distributed database (DDB) is a collection of data that logically belongs to the same system but is spread over the sites of a computer network. It is not necessary that database system have to be geographically distributed. The sites of the distributed database can have the same network address and may be in the same room but the communication between them is done over a network instead of shared memory. As communication technology, hardware, software protocols advances rapidly and prices of network equipment’s falls every day, developing distributed database systems become more and more feasible. Design of efficient distributed database is one of the major research problems in database & information technology areas

A distributed database management system (DDBMS) is then defined as the software system that permits the management of the DDB and makes the distribution transparent to the users. Distributed
database system (DDBS) is the integration of DDB and DDBMS. This integration is achieved through the merging the database and networking technologies together. Or it can be described as, a system that runs on a collection of machines that do not have shared memory, yet looks to the user like a single machine.

Assumptions regarding the system that underlie these definitions are:

1. Data is stored at a number of sites. Each site is assumed to logically consist of a single processor. Even if some sites are multiprocessor machines, the distributed DBMS is not concerned with the storage and management of data on this parallel machine. [3]

2. The processors at these sites are interconnected by a computer network rather than a multiprocessor configuration.

3. To form a DDB, distributed data should be logically related, where the relationship is defined according to some structural formalism, and access to data should be at a high level via a common interface. The typical formalism that is used for establishing the logical relationship is the relational model. [3]

4. The system has the full functionality of a DBMS. Distributed processing on database management systems (DBMS) is an efficient way of improving performance of applications that manipulate large volumes of data. This may be accomplished by removing irrelevant data accessed during the execution of queries and by reducing the data exchange among sites, which are the two main goals of the design of distributed databases. Primary concern of distributed database system design is to making fragmentation of the relations in case of relational database or classes in case of object oriented databases, allocation and replication of the fragments in different sites of the distributed system, and local optimization in each site. [3]

III. FRAGMENTATION

Fragmentation is a design technique to divide a single relation or class of a database into two or more Partitions such that the combination of the partitions provides the original database without any loss of information. This reduces the amount of irrelevant data accessed by the applications of the database,
thus reducing the number of disk accesses. Fragmentation can be of any type: horizontal, vertical and hybrid/mixed. [5]

![Figure 3 Data](image1)

![Figure 4 E-R Diagram](image2)

- Fragmentation aims to improve:
  - Reliability
  - Performance
  - Balanced storage capacity and costs
  - Communication costs
  - Security

- The following information is used to decide fragmentation:
  - Quantitative information: frequency of queries, site, where query is run, selectivity of the queries, etc.
  - Qualitative information: types of access of data, read/write, etc.
3.1 Vertical Fragmentation

Vertical fragmentation splits a single relation $R$ into sub-relations that are projections of relation $R$ with respect to subset of attributes. These relations are in grouping with attributes and frequently accessed by queries. Projection built the vertical fragments \(^1\). By joining the fragments the original relation is reconstructed. \(^5\)

Explanation with example:
Vertical fragmentation of $PROJ$ relation
- $PROJ_1$: information about project budgets
- $PROJ_2$: information about project names and locations

![Figure 5 Vertical Fragmentation](image)

3.2 Horizontal fragmentation

Horizontal fragmentation, divides a single relation $R$ into subsets of rows using query predicates. It reduces query processing costs by selecting the horizontal fragments that are built and the original relation is reconstructed by union of the fragments. In short, each fragment represents the equivalent of a SELECT statement, with the WHERE clause on a single attribute. \(^5\)

Explanation with example:
Horizontal fragmentation of $PROJ$ relation
- $PROJ_1$: projects with budgets less than 200,000
- $PROJ_2$: projects with budgets greater than or equal to 200,000
Mixed fragmentation (hybrid fragmentation)

The Mixed/Hybrid fragmentation is Combination of horizontal and vertical fragmentations. This type is most complex one, because both types are used in which horizontal and vertical fragmentation of the DB application [1]. The original relation is obtained back by join or union operations. [5]

IV CONCLUSION

It is important to manage an appropriate methodology for data fragmentation in order to utilize the resources and thus it is must to select an accurate and efficient fragmentation methodology to enrich the power of distributed database system.
REFERENCES


