Ain Shams University
Faculty of Computer and Information Sciences
Information Systems Department

Multidatabase Query Optimization

THESIS

Submitted to the Information Systems department, Faculty of Computer and Information Sciences, Ain Shams University in partial fulfillment of the requirements for the degree of Master of Computer and Information sciences

BY

AMIRA REZK ABDOU REZK

B.Sc. Degree in Computer and Information sciences (2000)
Demonstrator, Information System Dept., Faculty of Computer Science and Information System, Mansoura University.

SUPERVISED **BY**

Prof. Dr. **MOHAMED ELSHARKAWY**
Faculty of Computer and Information Sciences
Ain Shams University.

Asst. Prof. Dr. **HESHAM ARAFAT**
Faculty of Engineering
Mansoura University.

Cairo – 2006

To my parent …

Who taught me the value of learning.

Acknowledgment

Thanks are due to ALLAH for getting this work done.

I am very grateful for the encouragement of Professor Esam Khalifa the faculty dean and his support to finish this work.

I wish to express my gratitude to my supervisor Professor Mostafa M. Syaim for his support during the preparation of this work.

I would like to thank my supervisor Professor M. A. Sharkawy for his helpful, guidance and valuable comments on my work.

My sincere thank to my supervisor Assistant Professor H. A. Ali for his cooperative, guidance, and encouragement during the preparation of this work.

Many thanks to my friend Eng. Mervat M. Fahmy for her support and advices.

I am thankful for every one urges me to go forward.

CONTENTES

ABSTRACT

Multidatabase Query Optimization, By: Amira Rezk Abdou
Submitted to the Information Systems department, Faculty of Computer and Information Sciences, Ain Shams University

The researchers have an interest in the Multidatabase system (MDBS) as a new trend to integrate the pre-existing database systems, as information resources, and provide the global users with a global uniform view. However, the heterogeneity and autonomous of such systems make it difficult to integrate their data.

The most critical issue in Multidatabase system (MDBS) is query optimization, and the most important problem associated to global query optimization in a MDBS is that some required local information about local database components (DBCs) will not be available at the global level due to local autonomy.

The main goal of this work is to solve the problem associated to Query optimization in MDBS. In this thesis an overview of the MDBS characteristics and different issues are introduced. The most previous work that interested in the Multidatabase query optimization are studied carefully. Such study and analysis depict a lot of problems that must be solved to improve the performance of the system such (The lack of information about the local DB in MDBS, Integrate data from heterogeneous sources, Perform large join queries efficiency). The main concern of this work is solving the first one.

The main contribution of this thesis is solving the query optimization problems in the Multidatabase system. To achieve this goal a new framework is proposed to establish a global protocol. This framework can be applied to any autonomy database systems which have the ability to integrate together. It deals with Federated Database System as special case from the Multidatabase system where there is no global catalog.

The proposed framework aims at enhancing the performance of the query optimization process through four designed issues:

- Overcoming the lack of local information.
- Improving the decomposition process.
- Increasing the response of the system.
- Reducing the data transmission in the system.

Validation and evaluation of the efficiency of the proposed framework is done via illustrated experiments of a case study and simulation experimental result. The experiment results indicate that, the proposed framework achieves its goal and objectives to enhance the performance of the query optimization process along the four design issues as follow:

- Collect the local data from its site using the suggested driver overcomes the lack of local information.
- Distributing the decomposition process among the local database component enhances it and avoids wastage of the decomposition effort. The average of eliminating the wastage of the decomposition is around 50% of the submitted Query.
- The suggestion of executing the query partially, i.e. the framework provides the global user with an answer for his/her query even if there are sub-queries that will not be executed, increases the ratio of return result for the user around 45%.
- Applying the proposed routing technique that depends on a set of messages between the two layers reduces the data transmission in the system. The percent of data exchange reducing is increase when the no of join operation which join tables from different DBCs increases.

The simulation profess that the proposed framework achieves its goal to improve the query optimization process and enhance the performance of the Multidatabase system.

Keywords: Multidatabase, Federated database,
Query optimization, Heterogeneity, Autonomy.

LIST OF FIGURES

LIST OF TABLES

1

INTRODUCTION

Chapter1

INTRODUCTION

1.1 Motivation

Quick review for the large modern enterprise finds a large number of data sources that contain a huge amount of data. Nowadays, there is a new trend to combine the information from these various systems. Therefore, the enterprise can realize the full value of the data they contain. Nevertheless, these systems are usually not designed to interoperate with each other in a uniform manner [1, 2, 3, 4]. In this way, they are isolated data sources, only used for local purpose. The heterogeneity and autonomous of these systems make it difficult to integrate their data. Throughout the 1980s, the database market matured and companies attempted to standardize on a single database vendor. However, the reality of doing business generally made such a strategy unrealistic [57]. Therefore, there is a growing need for tools to maximize the reusability and interoperability of these arbitrary computing systems. This tool is the Multidatabase management systems (MDBMS). MDBMS is built on the top of the existing database system, to integrate data from them and provide the user with a global uniform view.

This trend debuted commercially in the 1990s under various names. Papers of that time often referred to such products as next-generation gateways, data access middleware and Multidatabase. Then a new terminology Federated database system (FDBS) appears [58].

Although Multidatabase system has the common database systems characteristics, it has its own characteristics that make it more complexes than the other types of database system (DBS), and imposes certain constraints that must be taken into consideration. MDBS is an integration of a set of database system components (DBCs), each one of these DBCs is located in different location (Distributed), built using various database management system, different hardware platform, divers schema, or data model. (Heterogeneous), and work stand-alone, i.e. each DBC is isolated from the other (fully Autonomy). However, when these DBCs decide to integrate together, they give up part of their autonomy. So MDBS can be characterized as Distributed, Heterogeneous and semi-Autonomy [5].

1.2 Research Goals

The most critical issue in Multidatabase system (MDBS) is query optimization. It can be considered the backbone of any successful database system. Therefore, there is a growing need for query optimization algorithm that can effectively deal with Multidatabase system. There are many obstacles and challenges are implied within query optimization in MDBS. These obstacles rise from the natural of query optimization itself, or from the characteristics of the MDBS.

The major challenge of the global query optimization in a Multidatabase system is that some required local information about local database components (DBCs) might not be available at the global level due to local autonomy.

In this thesis, the introduced solution is oriented to the federated database system, the special case of the multidatabase system, where there is no global catalog. The goals and research activities are outlined as follows:

1- Surveying the Multidatabase system environment, its characteristics, and issues.

2- Surveying the query optimization problems in the database system and its challenges in the multidatabase system environment.
3- Introducing a framework to manage the relation between the FDBMS and the DBCs, and state a set of rules for any database systems decided to integrate together.
4- Providing a solution for the lack of local information based on the introduced framework.

1.3 Publications

- Mostafa Syaim, H. A. Ali, Amira Rezk, "A New Framework For Query Optimization in Multidatabase System Environment". Proceedings of the 14th International Conference on Computer Theory and Applications, Alex, 2004

** Mostafa Syaim, H. A. Ali, Amira Rezk, "A New Framework For Query Optimization in Multidatabase System Environment", MJCSIS Headquarters Journals Staff, Volume 1, Number 0, Jan 2005

- Mostafa Syaim, H. A. Ali, Amira Rezk, Solving Query Optimization Problems in Multidatabase System. Proceedings of 3rd International Conference on Computer Science, Software Engineering, e-Business and Applications, Cairo, 2004

- Mostafa Syaim, H. A. Ali, Amira Rezk, Analysis Of Multidatabase System Framework To Solve Query Optimization Problems. Proceedings of 2nd International Conference on Intelligent Computing & Information Systems, Cairo, 2005.

1.4 Thesis Outline

The work in this thesis is organized as follow:

Chapter 2 introduces some concepts within the Multidatabase system environment contain an answer for the