Faculty of Science

A Model for Open Gate Algorithm vs Heterogeneous Distributed Database Issues

A thesis Submitted to

The Faculty of Science
Ain Shams University

In partial fulfillment of the requirement for the degree of doctor of philosophy (PH.D.) in computer science

By

Abdullah Faisal Atmaz Al-Sebai

Supervised By

Prof. E. Fayed F. M. Ghaleb
Department of Mathematics
Faculty of Science
Ain Shams University

Dr. Naglaa M. Reda Taher
Department of Mathematics
Faculty of Science
Ain Shams University

Asc. Prof. Hawaf A. H. Telb
Department of Information Technology
Faculty of Computers and Information
Helwan University

Cairo Egypt

2010

Acknowledgement

I would like to thank all people, who have helped me to complete my thesis, I am Grateful their assistance.

I am deeply obliged to my supervisors A. Prof. Fayed F. M. Ghaleb and Dr. Naglaa M. Reda for their great help, and their precious time which they have given me.

A great grateful to my family (all of them); they have always encouraged me, I indebted to all; my father, mother, brothers, and sisters.

From my heart big thanks for my friends for their help and encourage.

At the first and the last, a great thanks to my wife who gives me all of her time and support.

Summary

Distributed database systems have become an essential and vital component to trading information and international exchange. Heterogeneous systems spread currently because it allows dealing with various forms of data sources. These systems address the problem of integrating heterogeneous data sources in several ways; the most famous systems use middleware approach. Although many middleware systems have been produced, most of them don't achieve distributed database advantages. Also they face some problems that have bad effects on their performance.

This thesis proposes a new model of a middleware system special for heterogeneous distributed databases. The objective of this system is to achieve some advantages like scalability, autonomy, reliability and high performance.

Also a new wrapper is designed to overcome the main weaknesses in former middleware systems such as bottleneck, and high network communication cost.

The proposed wrapper has been implemented, and tested for many of the various queries for assessing the performance. Also a comparison between the new wrapper's processing time with the time for one of the known systems have been made. Better results and clear improvement in performance have been achieved.

The thesis is divided into five chapters and two appendices:

Chapter 1: This chapter presents fundamentals of distributed database systems, their basic concepts, advantages and problems, with their different types and architectures. It also gives an overview of the basic technical issues of distributed databases design.

Chapter 2: This chapter reviews the basic concepts of data integration systems, their definitions, characteristics, classification, and various approaches. It also surveys the most popular systems that relay in its design on middleware systems.

Chapter 3: In this chapter a detailed study is devoted to a proposed wrapper (IWRAP) for dealing with queries coming from the middleware system. This study includes the proposed wrapper architecture, services and components (*the controller, the schema integrator, and the query translator*). An algorithm with its flow chart for each component is presented. This chapter, also, surveys other wrappers. Finally it presents experimental results of measuring the proposed wrapper's performance compared with another wrapper.

Chapter 4: This chapter proposes a model for a new middleware system (Open-Gate) which is specially designed for heterogeneous distributed databases using the proposed wrapper. It contains a study of the system's environment, architecture, storages (*the query queue, and the data cache*), and components (*The manager, the global integrator, the query processor, and the data collector*). It introduces algorithms and flow charts for all storages and components. It also focuses on the main characteristics of the proposed system such as autonomy, scalability, reliability and high performance.

Chapter 5: Conclusion and future work.

Appendix A: It gives a brief study of some other middleware systems.

Appendix B: It presents the source code for the proposed wrapper.

Table of Contents

List of Figures

List of tables