Acknowledgment

First and foremost, thanks to Allah, who gave me the strength and the support to accomplish this work.

I am deeply, indebted to Professor Dr. Ahmed Sharaf Eldin, the vice dean of faculty of computers and information at Helwan University for suggesting the topic of the research and for his valuable guidance, support, help, and encouragement during the work.

I would like to express my great and deep gratitude to Dr. Doaa Saad for her help and for the huge amount of effort she spent in guiding my research.

Finally, with all my appreciation and love that no words can express, I dedicate this Master's to my dear parents who always provided me with love, supplications, blessing, and care. They are behind any success in my life. I am also thankful to my wonderful sister: Ola for her understanding and supporting me, my lovely brothers: Ashraf, Sherief, Taha and Youssef, my valuable friends: Amira Fathy, Sherine Awad, Hanan Fahmy, Hanan Saeed, Hala Aboelhamael and Ranya Elsayed.

Abstract

Quality of service in distributed query processing

Query optimization implies creating different query execution plans, then selecting the best plan such that the execution cost function is a minimal. In distributed database systems, the query optimization cost function includes the communication cost of transmitting data across the network, in addition to the CPU cost and I/O cost. Two different goals for distributed query optimization are often considered, one is to minimize the total cost and the other is to minimize the response time for each query. These two goals are usually in conflict. However, since operations can be executed in parallel at different sites, it is possible to optimize a distributed query using both criteria, with one as the primary goal and the other as a secondary goal.

One of the most important issues that may affect query processing in distributed database is the used network links. Consequently, selecting an optimized plan is one of the most important query processing problems. Distributed query processing is usually done assuming a cost function associated with each communication link in the network connecting different databases. Quality of service (QoS) is an

important issue that should be considered when transmitting data over a network link.

In fact, what we propose in this thesis is to add QoS factors to the traditional cost function used in optimizing query processing. We propose adding the network delay, error rate, loss probability and bandwidth to the distributed query optimization cost function to contribute in selecting the most optimal plan.

We propose a multiple objective optimization model that considers the cost function and the QoS factors as a goal programming (GP) problem, where minimizing the cost function and handling (either to minimize or to maximize) the QoS factors are its goals. The GP approach consists of a multi-objective function and a set of constraints. In addition, there is a non-negativity requirement which means that all variables must be non-negative. The model is solved using the available software (WINQSB). The results we obtained showed that the best execution plan may be changed if the QoS factors are taken into consideration.