

Query Processing in a Distributed Environment

by

Han Ying Chao

B.S., Soochow University, Taipei, Taiwan R.O.C. 1980

---

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Computer Science

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1982

Approved by:



Major Professor

SPEC  
COLL  
LD  
2668  
R4  
1982  
C436  
C.2

A11202 337323

**Acknowledgements**

The writer is indebted to her major advisor, Dr. Paul Fisher for his knowledge and perspectives in the development of this research.

Special gratitude goes to my parents, Mr. and Mrs. Chin Chi Chao, for their encouragement throughout my graduate studies. They instilled in me the value of education.

**THIS BOOK  
CONTAINS  
NUMEROUS PAGES  
WITH DIAGRAMS  
THAT ARE CROOKED  
COMPARED TO THE  
REST OF THE  
INFORMATION ON  
THE PAGE.**

**THIS IS AS  
RECEIVED FROM  
CUSTOMER.**

## Query Processing in a Distributed Environment

## Table of Contents

Chapter 1	Introduction
Chapter 2	Definition of Data Bases
Section 2.1A	Hierarchical
Section 2.1B	Network
Section 2.1C	Relational
Section 2.2	Definition of Homogeneous DDB
Section 2.3	Definition of Heterogeneous DDB
Chapter 3	Query language
Section 3.1	Relational Algebra
Section 3.2	Tuple Relational Calculus
Section 3.3	Domain Relational Calculus
Section 3.4	ISBL
Section 3.5	SQUARE and SEQUEL
Section 3.6	QUEL
Section 3.7	QBE
Chapter 4	Query Processing in Homogeneous DDB's
Section 4.1	Multiple Machines in Homogeneous DDB Environment
Section 4.2	Network Decomposition Algorithm
Section 4.3	The Optimization Problem
Section 4.4	Summary
Chapter 5	Query Processing in Heterogeneous DDB's
Section 5.1	Query Processing Mechanism
Section 5.2	Alternative Solutions
Section 5.2.1	Multi-level Integration
Section 5.2.2	Partial Integration
Section 5.3	Summary
Chapter 6	Conclusions

## Chapter 1 Introduction

Distributed computing systems are a new type of system. They emerged from the necessity to provide a vast amount of data to satisfy the requirements of geographically separate end users. Therefore, the need for data bases containing required information is becoming more and more a part of our society. Several approaches concerned with data access by means of a query language have been proposed. To achieve the goal of flexibility and efficiency when accessing data in a distributed environment, query processing needs to be studied carefully.

First, the diverse categories of abstract query languages are discussed as they relate to the relational data model. The form and structure of each language are overviewed. Subsequently, several current query languages are discussed with respect to the type of the abstract query language used.

The purpose of this report is to focus on the procedure for analyzing, decomposing, transmitting, and synthesizing a query. The object is to try to obtain the optimal solution of query processing under the conditions of minimum network traffic and minimum response time. Some methodologies dealing with query processing in both homogeneous and heterogeneous environments are discussed. These include the network decomposition algorithm, query processing mechanisms for heterogeneous DDB's, multi-level

integration, and partial integration. Although the technique for the latter two methods is not practical to implement, it demonstrates future tendencies.

## Chapter 2 Definition of Data Bases

In order to obtain an overall concept of diverse data bases, the hierarchical, network, and relational data bases will be discussed respectively.

### Section 2-1.A Hierarchical

A data base with data structure that a parent record type may have one or more child record types satisfies the restrictions (1) a child record type may not have more than one parent record type, and (2) an M:N relationship of instances between two record types is not allowed.

The relationships between records (or segments as termed in hierarchical data base) are not explicitly expressed as those sets of network data model. The hierarchy of segments are ranked by means of the parent-child relationships. The root segment type is the one with the highest hierarchy. All other segment types, called dependent segment types or child record types, are arranged in the tree structure in 'physical' parent-child relationships. Currently, IBM's Information Management System (IMS) is one of the most commercially used systems based on hierarchical data model.

### Section 2-1.B Network

A hierarchical data base can be viewed as a special case of network data base. Here, the network data base can