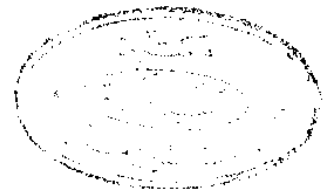
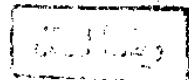


Ain Shams University  
Faculty of Engineering

**Computer-Aided Design  
for  
Distributed Database Systems**

**By  
Hoda Korashy Mohamed**

626-38195  
11 K  
A Thesis  
Submitted in fulfillment of the  
requirements of the Degree of PH.D.  
in Electrical Engineering  
(Electronics and Computer)



Supervised By

Prof. Dr. M. Adeeb R. Ghonaimy

Prof. Dr. Osman A. Badr

350

**Cairo  
(1992)**



**Examiners Committee**

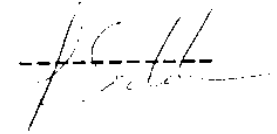
**Name , Title and Affiliation**

**Signature**

1- Prof. Dr. R.A. Amer  
Prof. at King Abdeel Aziz  
University, Saudi Arabia.

-----

2- Prof. Dr. F.H. Saleh  
Prof. at Cairo University,  
Egypt .



3- Prof. Dr. M.A.R. Ghonaimy  
Prof. at Ain Shams University,  
Egypt.



Date: 10 / 3 / 1992

### **STATEMENT**

This dissertation is submitted to **Ain Shams University** for the degree of **PH.D.** in **Electrical Engineering (Electronics and computers)**.

The work included in this thesis , was carried out by the author in the Department of **Faculty of Engineer**, **Ain Shams University** , from **2/1986** to **3/1992** .

No part of this thesis has been submitted for a degree or a qualification at any other university or institution .

Date : 3 /1992

Name : Hoda Korashy Mohamed

Signature : *Hoda Korashy*

## **ACKNOWLEDGMENT**

I would like to express my limitless gratitude to **Prof. Dr. M.A.Ghonaimy** for his invaluable advices, supervision and guidance through the research.

I am also grateful to **Prof. Dr. O.A.Badr** for his patient guidance and continuing encouragement.

I would like also to thank all my colleagues, especially Eng. Mona Fahmy, for their help and encouragement.

Finally, I would like to acknowledge the encouragement and support shown by my husband and my mother.

## ABSTRACT

A comparative study for the design of distributed database has been done. The design of distributed database, may be divided into four phases, each has its own techniques and tools.

A proposed system is applied to design the distributed database system. The proposed system (**DBDSG**) is a semiautomated methodology which covers all the design steps. It is divided into two parts.

**DBDSG-1** system is a menu driven system to facilitate the interface with designers. Data are represented using *Extended-Entity Relationship model [EERM]*. View integration is applied if more than one view exists. The system transforms the EERM into a relational data model which is the data model required for the next phase of the design. Analysis of the important transactions is applied, to develop all possible partitioning of objects and the access path for each transaction .

In **DBDSG-2** system, a model is proposed for the nonreplicated allocation of data over the sites of the computer network. A modified model is presented to modify some of the constraints in the original model. Zero/one implicit enumeration algorithm is used to solve the allocation problem. For replicated allocation of data, a heuristic postoptimization of the optimal solution without replication is applied. An analysis of the results of the allocation model proposed with the effect of different parameters is also discussed . The parameters which affect the allocation model are the cost parameter such as transmission cost and the access cost of the computers at each site. The load of transactions at each site has a pronounced effect in the allocation problem.

## **TABLE OF CONTENTS**

### **1 INTRODUCTION**

1.1 What is meant by Distributed Database? . . . . .	1
1.2 Features of Distributed versus Centralized Database . . .	2
1.3 Distributed Database Design . . . . .	3
1.4 Computer-aided design tools . . . . .	5
1.5 Organization of the thesis . . . . .	7

### **2 DATABASE DESIGN**

2.1 Database Design Criteria . . . . .	9
2.2 Database design inputs and outputs . . . . .	10
2.2.1 Inputs to database design . . . . .	11
2.2.2 Outputs from database design . . . . .	12
2.3 Database Design Phases . . . . .	12
2.4 Requirement Formulation and Analysis . . . . .	13
2.4.1 Capturing Requirements to Database Design . . .	13
2.4.2 Requirements definition and specification . . .	15
2.4.3 Requirement Validation . . . . .	20
2.5 Conceptual Design . . . . .	21
2.5.1 View Modeling . . . . .	21
2.5.2 View Integration . . . . .	33



2.6 Logical Design . . . . .	42
2.6.1 The translation process . . . . .	43
2.6.2 Overview of systems implemented to support logical design . . . . .	45
2.7 Physical Design . . . . .	46

### **3 Data Distribution Design**

3.1 Objectives of the design of data distribution . . . . .	49
3.2 Parameters that affect the design of Data Distribution . . . . .	50
3.2.1 Type of distributed Database Management System (DDBMS) . . . . .	50
3.2.2 Database Schema . . . . .	51
3.2.3 Types of Computer Networks . . . . .	51
3.2.4 Transactions (Application Requirements) . . . . .	52
3.2.5 Computers connected at each site . . . . .	53
3.3 Fragmentation Design . . . . .	54
3.3.1 What is meant by fragmentation . . . . .	54
3.3.2 Fragmentation Types . . . . .	54
3.3.3 Horizontal Fragmentation . . . . .	55
3.3.4 Vertical Fragmentation . . . . .	61
3.4 Allocation Problem . . . . .	64
3.4.1 Allocation model approaches . . . . .	64
3.4.2 General Criteria for Fragment Allocation . . . . .	65
3.4.3 Overview of the allocation problems . . . . .	66

## **4 The Proposed Model (DBDSG)**

4.1 DBDSG-1 : Designing global schema . . . . .	76
4.2 View modeling . . . . .	77
4.2.1 Why Extended Entity-Relationship Model (EERM) ?	77
4.2.2 Entity-Relationship model (ERM) . . . . .	78
4.2.3 Extended Entity-Relationship Model (EERM) . . . .	80
4.2.4 EER modeling of requirements . . . . .	83
4.2.5 User-defined Constraints . . . . .	86
4.2.6 View modeling in DBDSG-1 System . . . . .	87
4.3 View Integration . . . . .	99
4.3.1 Preintegration phase . . . . .	100
4.3.2 Comparison and Conforming phases . . . . .	102
4.3.3 Merging and Restructuring Phase . . . . .	106
4.4 Logical Design . . . . .	108
4.4.1 Step 1 : Transformation of the EERM to relations . . . . .	108
4.4.2 Step 2 : Normalization of Relations . . . . .	111
4.5 Transaction modeling . . . . .	119

## **5 Data Allocation Model (DBDSG-2)**

5.1 Nonreplicated model (the original model) . . . . .	124
5.1.1 Inputs to the model . . . . .	124
5.1.2 Assumption and Simplification . . . . .	139
5.1.3 Objective function . . . . .	139

5.1.4 Constraints . . . . .	142
5.1.5 Coefficient Evaluation . . . . .	143
5.1.6 Algorithm . . . . .	151
5.2 The Modified Model . . . . .	159
5.2.1 Constraints . . . . .	160
5.2.2 Objective function . . . . .	163
5.3 Design of a partially replicated distribution of the database . . . . .	165
5.3.1 Assumption . . . . .	165
5.3.2 Evaluation of Processing Costs of a Transaction	169
5.3.3 Algorithm . . . . .	172

## **6 RESULTS AND ANALYSIS OF THE ALLOCATION MODEL**

6.1 Examples Specifications . . . . .	181
6.2 Nonreplicated Model . . . . .	187
6.2.1 Cost Parameters . . . . .	187
6.2.2 Load of Transactions . . . . .	209
6.3. The effect of the implicit enumeration algorithm . . .	211
6.2. The Replication model . . . . .	220

## **7 CONCLUSION AND FURTHER WORK**

7.1 CONCLUSION . . . . .	228
7.1.1 Strength of DBDSG system . . . . .	229
7.1.2 Shortcoming of DBDSG system . . . . .	230

7.2 Further Work . . . . .	231
7.2.1 Expert Systems . . . . .	231
7.2.2 Object-Oriented Databases . . . . .	232

APPENDIX A : An Example Applied to the Proposed system  
(DBDSG)

REFERENCES

## CHAPTER 1

### INTRODUCTION

<i>1.1 What is meant by Distributed Database? . . . . .</i>	<i>1</i>
<i>1.2 Features of Distributed versus Centralized Database . . .</i>	<i>2</i>
<i>1.3 Distributed Database Design . . . . .</i>	<i>3</i>
<i>1.4 Computer-aided design tools . . . . .</i>	<i>5</i>
<i>1.5 Organization of the thesis . . . . .</i>	<i>7</i>

## INTRODUCTION

### 1.1 What is meant by Distributed Database?

A typical definition of a distributed database is :

*" A distributed database is a collection of data which are distributed over different computers forming a network. Each site of the network has autonomous processing capability and can perform local applications. Each site also participates in the execution of at least one global application, which requires accessing data at several sites using a communication subsystem" [CER184].*

The definition emphasizes two important aspects :

#### 1- Distribution

Distribution means the fact that the data are not resident at the same site. That will distinguish a distributed database from a single centralized database.

#### 2- Logical correlation

The fact that the data have some properties which tie them together is one of the characteristics of distributed databases , which are different from a set of local databases residing at different sites of a computer network.

## 1.2 Features of Distributed versus Centralized Database

### 1- Centralized Control

In a **centralized database**, database administrator **DBA** is responsible for the safety of data. In a **distributed database**, there is a global database administrator who has the central responsibility of the whole database, and local database administrators who have the responsibility of their respective local databases. But **local DBA** may have a high degree of autonomy so the **global DBA** is completely missing and the intersite coordination is performed by the local administrators.

### 2- Data independence

**Data independence** means that the actual organization of data is transparent to the users. In **centralized databases**, this was achieved by conceptual schema. In **distributed databases**, we have **distribution transparency** which means that users see the databases as if they were not distributed.

### 3- Reduction of redundancy

In **centralized databases**, redundancy is reduced as far as possible by using data sharing that means allowing several applications to access the same files and records. In **distributed databases**, there are several reasons for considering data redundancy as a desired feature :

- a- To increase site locality of applications.
- b- The availability of the system can be increased.

### 4- Complex physical structures and efficient access

In **centralized databases**, complex physical structures are a major aspect for efficient access to the data. In **distributed**