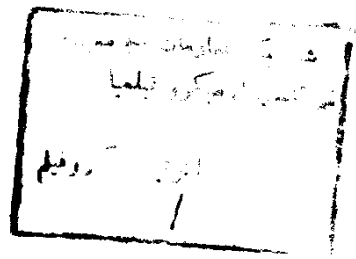
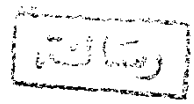
Ain Shams University
Faculty of Engineering
Department of Computers and System Engineering



Distributed Database Systems Performance Analysis

by
Eng. Mohamed Abdel Kader Salem



A Thesis

Submitted in fulfillment for the
Requirements of the degree of Master in Science
in Computer Engineering

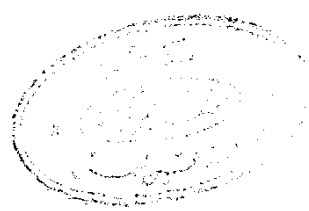
621.392
م. A

Supervised by

Prof. Dr. Osman. A. Badr
Professor of Computers, Ain Shams University

Dr. El Sayed A. S.El Sakka
Computers and Information Systems Consultant

Cairo 1993

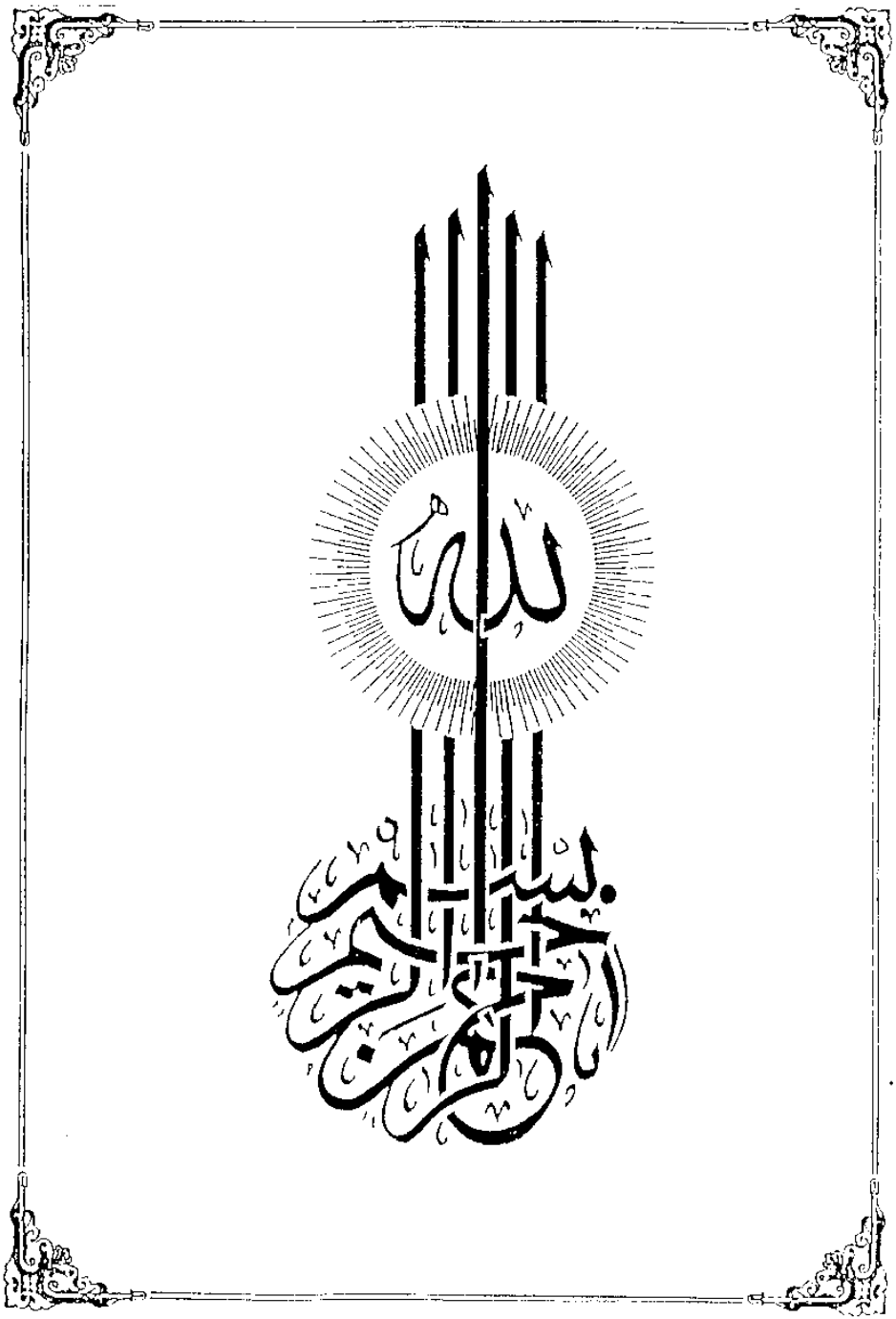


Central Library - Ain Shams University

TO ITS REAL OWNER
THIS THESIS IS DEDICATED

TO NEVINE

Central Library - Ain Shams University



Central Library - Ain Shams University

Examiners Committee

The thesis on
Distributed Database Systems - Performance Analysis

By
Eng. **Mohamed Abdel Kader Salem**

Approved By

Professor Dr. **M. A. R. Ghonaimy** : M. A. R. Ghonaimy
Ain Shams University

Professor Dr. **A. Fahmy** : A. Fahmy
Cairo University

Professor Dr. **O. A. Badr** : O. A. Badr
Ain Shams University (Supervisor)

Dr. **S. A. S. El Sakka** : Dr - Sayed
(Supervisor)

Central Library - Ain Shams University

Statement

This dissertation is submitted to Ain Shams university for the degree of Master of science in computer Engineering.
The work included in this thesis was carried out by the author in the department of computers and system engineering, Faculty of Engineering, Ain Shams university, from 9/1990 : 9/1993.
No part of this thesis has been submitted for a degree or a qualification at any other university or institute.

Date 30-12-1993
Signature M. Sal
Name Mohamed Abdel Kader Salem

Central Library - Ain Shams University

ACKNOWLEDGMENT

I am greatly indebted to ***Professor Dr. Osman Badr*** for all his support, help and generous contributions.

and

To my intelligent colleague and dear friend ***Dr. El Sayed El Sakka,*** I sincerely dedicate my great thanks for his academic support, unlimited moral and continuous guidance and tremendous encouragement which made this work come true..

I would like to thank as well ***Professor Dr. Adeeb El Ghonaimy*** and ***Professor Dr. Aly Fahmy***, for honoring me by participating and enriching the discussion of this thesis.

Central Library - Ain Shams University

Distributed Database Systems Performance Analysis

Abstract

Performance analysis is considered one of the most important operations which is carried out on any new technologies, because it gives the best results with minimum cost and time frame.

In the field of Distributed Database Systems, and as a result of the huge and continuing progress in this field, and because of its speedy, accurate and low cost methods, ***Performance Analysis*** has become may be the leading method in dealing with those systems problems .

Using Simulation for performance analysis gives deep insight in the Distributed Database systems as well as accurate results.

Central Library - Ain Shams University

Thesis Construction

In this study different areas have been surveyed and analyzed, so many programs were written starting from the very simple analytic level up to the main model which was made simple enough to reflect the essential goal of the study, comprehensive enough to cover all the environments of the real-life system and most accurate to achieve the best results available. The study consists from three chapters and a conclusion with the necessary figures, graphs, tables and appendices.

the background study of this work includes Distributed Database systems concepts, structure as well as simulation and analytic modeling techniques including Queuing Network Modeling and Simulation Modeling.

A survey for Performance Analysis in Distributed Database Systems have been carried out, using Queuing Network Modeling a simple primitive analytic models have been constructed, implemented and used for verification of the main simulation model.

A simulation model to study Query Processing in Distributed Database Systems has been constructed. An application package was analyzed, designed and implemented using C programming language and Object Oriented techniques in C++ to the solve the model. A deep detailed study of the effect of database file organization is carried out, the results of this study is included in the model parameters.

The simulation model successfully ran, followed by gathering and

Central Library - Ain Shams University

analyzing the results, the conclusion and recommendation were given.
The following is a brief description of the thesis chapters.

Chapter #1 is a highly tutorial one and this contribution was meant to be so for two reasons, the first came from the research thesis objective to study the distributed database systems from all aspects, the second was to give the interested reader in that topic a short concentrated logical summery of that area.

This chapter covers the Distributed Database Systems definition, concepts, structure, advantages which lies in the fact that it mirrors the real life system, disadvantages and always complexity was the main problem of the that sort of systems. Following that the chapter calls for the fundamental principles of a distributed database system like, local autonomy, location transparency, replication transparency, distributed query processing and some more rules which are presented in this chapter. The chapter moves after that to the main problems of the distributed database systems like query processing, recovery control, concurrency control, update propagation and catalog management together with a very brief suggested solutions of some of these problems.

Chapter #2 focuses the light on Performance Analysis generally and in Distributed Database specifically.

It starts by discussing the techniques of Performance Evaluation namely Measurement, Analytic Modeling and Simulation. The chapter goes through those techniques in brief, touches some sort of comparison between these techniques and branches after that into two main sections, Queuing Network Modeling and Simulation Modeling which had been discussed in details in

Central Library - Ain Shams University

addition to some Analytic and Simulation models that had been presented and discussed to complete the idea of that sort of performance evaluation.

A definition of Queuing Network Modeling is presented followed by the reason why it was decided to work with queuing network modeling rather than Queuing theory. The next part describes how to construct a queuing network model going through its definition then parameterization and evaluation of that model to determine the behavior through out the alternatives and specifying to what extent the model behavior is the same manner as the real system. Following to that the section calls for queuing network basic definitions, basic quantities and basic laws. A case study was considered for the reason of construction and validation the main simulation model. The case study was discussed from the very basic concepts of open and closed models, single and multiple classes and centers. All the relevant relations and laws were overviewed before starting discussing a multiple classes multiple centers open queuing network model which was written in C programming language. Inputs were proposed output were analyzed, all the necessary flow charts, output graphs and tables were presented.

Moving to section II, it starts by defining simulation modeling, then it goes for the types of simulation then describing the mechanics of constructing and executing a simulation model whether it is an event or time driven. Being one of the most important phases of simulation; gathering observation, was discussed from the point of view of methods of gathering these observations, namely Replication, subintervals and cycles method.

Dealing with the amount of observations gathered needed some rules of sampling, and that was the reason for over viewing sampling theory which had

Central Library - Ain Shams University

been used as a tool for dealing with the gathered data. Many simulation models were designed and implemented for the possible combinations, starting from the simple open single class single server, up to multiple classes multiple centers models so a comprehensive case study one could be studied. A case study was considered, its results were analyzed for the purpose of determining some values of the inputs that was used in designing the main simulation model.

Chapter #3 is dedicated for the simulation model. It details the model in a progressive manner that starts from the model construction, representation then it proceeds to the model objectives, following to that, it calls for the model characterization and description.

Being one of the basic concepts of the model, deep analysis had to be done starting from the very beginning and that was the reason of detailing the customer arrival/departure routine with its two major events, arrival and departure. After what could be called base description, the query itself follows, and related items are descriped going through paths overview, model assumptions, environments, inputs, centers description, databases and outputs.

The model design primarily presented by the model flowchart, path optimization comes next with the relevant basic file system organization, different performance relationships, service time calculations for each path for the used file organizations in the model.

It is worthwhile mentioning in this point that many models (analytic and simulation) had been designed and implemented for the purpose of reaching the best available form used for the principle model verification & validation as well

Central Library - Ain Shams University

as using the most realistic data to get at the end the highest accurate results. Model solution starts from this point, presenting the results with all the relevant graphs for detailed description of the obtained data. Model validation uses two techniques, which by applying these techniques the model will prove its true representation of the real life system and then by using the statistical rules of confidence test the model will prove with no doubt that it has a high degree of accuracy.

Chapter #4, is the conclusion chapter, it contains the final analysis and conclusions as well as the recommendation of the study.

Central Library - Ain Shams University

TABLE OF CONTENTS

Chapter 1 Distributed Database Systems 1

1.1 Introduction 2

1.2 Database System 2

1.3 Distributed Database systems 3

1.4 Advantages and Disadvantages of distributed database 5

1.5 Fundamental Principles 7

1.5.1 Local Autonomy 9

1.5.2 Noncentralized Operation 9

1.5.3 Location Transparency 10

1.5.4 Fragmentation Transparency 10

1.5.5 Replication Transparency 10

1.5.6 Distributed Query Processing 11

1.5.7 Distributed Transaction Management 12

1.5.8 Hardware Independence 12

1.5.9 Operation System Independence 12

1.5.10 Network Independence 13

1.5.11 DBMS Independence 13

1.6 Problems of Distributed Database Systems 13

1.6.1 Query Processing 13

1.6.2 Recovery Control 14

1.6.3 Concurrency Control 15

1.6.4 Update Propagation 16

1.6.5 Catalog Management 17

Central Library - Ain Shams University

Chapter 2 Performance Analysis 19

2.1 Introduction 20

2.2 Techniques of Performance Evaluation 20

2.1.1 Measurement 20

2.1.2 Analytic Modeling 21

2.1.3 Simulation Modeling 21

Section I Queuing Network Modeling 22

2.I.1 Introduction 22

2.I.2 Queuing Network Modeling 23

2.I.3 Queuing Network Basic definitions 24

2.I.4 Queuing Network Basic Laws 26

2.I.5 Case Studies 28

Section II Simulation Modeling 44

2.II.1 Introduction 44

2.II.2 Types of Simulation 45

2.II.3 Mechanics of Constructing and Executing Simulation Models 45

2.II.4 Procedures of Gathering Observations 47

2.II.4.1 Replication Method 47

2.II.4.2 Subintervals Method 48

2.4.II.3 Cycles Method 48

2.II.5 Theoretical Background 49

2.II.5.1 Sampling Theory 49

2.II.5.2 Random Samples & Random Numbers 50

2.II.5.3 Confidence Interval Estimates of Population Parameters 51

Central Library - Ain Shams University

2.II.6 Case Study 52

Chapter 3 Simulation Model 60

3.1 Introduction 60

3.2 Model Construction 61

3.3 Model Objectives 62

3.4 Model Characterization 65

3.4.1 Customer Arrival/Departure Routine 66

3.4.2 The Query 70

3.5 Model Assumptions 72

3.5.1 Flow Chart 75

3.6 Path Optimization 78

3.6.I Basic System Organization 79

3.6.1 Performance of Direct File 80

3.6.2 Performance of Indexed File 81

3.6.3 Performance of Sequential File 82

3.6.4 Paths Algorithms 83

3.6.5 Equations used derivation 84

3.6.5.1 Direct file 84

3.6.5.2 Indexed file 85

3.6.5.3 Sequential file 85

3.6.5.4 Service time calculation 86

3.6.6 Algorithm Description 87

3.7 Model Solution 87

Central Library - Ain Shams University