Computer Science
and Applied Mathematics

# DATA BASE ORGANIZATION FOR DATA MANAGEMENT

Sakti P. Ghosh

# Data Base Organization for Data Management

SAKTI P. GHOSH

*IBM Research Laboratory*
*San Jose, California*



ACADEMIC PRESS New York San Francisco London 1977
*A Subsidiary of Harcourt Brace Jovanovich, Publishers*

## Computer Science and Applied Mathematics

A SERIES OF MONOGRAPHS AND TEXTBOOKS

Editor

**Werner Rheinboldt**

*University of Maryland*

HANS P. KÜNZI, H. G. TZSCHACH, and C. A. ZEHNDER. Numerical Methods of Mathematical Optimization: With ALGOL and FORTRAN Programs, Corrected and Augmented Edition

AZRIEL ROSENFELD. Picture Processing by Computer

JAMES ORTEGA AND WERNER RHEINBOLDT. Iterative Solution of Nonlinear Equations in Several Variables

AZARIA PAZ. Introduction to Probabilistic Automata

DAVID YOUNG. Iterative Solution of Large Linear Systems

ANN YASUHARA. Recursive Function Theory and Logic

JAMES M. ORTEGA. Numerical Analysis: A Second Course

G. W. STEWART. Introduction to Matrix Computations

CHIN-LIANG CHANG AND RICHARD CHAR-TUNG LEE. Symbolic Logic and Mechanical Theorem Proving

C. C. GOTLIEB AND A. BORODIN. Social Issues in Computing

ERWIN ENGELER. Introduction to the Theory of Computation

F. W. J. OLVER. Asymptotics and Special Functions

DIONYSIOS C. TSICHRITZIS AND PHILIP A. BERNSTEIN. Operating Systems

ROBERT R. KORFHAGE. Discrete Computational Structures

PHILIP J. DAVIS AND PHILIP RABINOWITZ. Methods of Numerical Integration

A. T. BERZTISS. Data Structures: Theory and Practice, Second Edition

N. CHRISTOPHIDES. Graph Theory: An Algorithmic Approach

ALBERT NIJENHUIS AND HERBERT S. WILF. Combinatorial Algorithms

AZRIEL ROSENFELD AND AVINASH C. KAK. Digital Picture Processing

SAKTI P. GHOSH. Data Base Organization for Data Management

DIONYSIOS C. TSICHRITZIS AND FREDERICK H. LOCHOVSKY. Data Base Management Systems

# Contents

## 3 Searching on One Field



## 4 Key to Address Transformation



## 5 Algebraic Filing Schemes



## 6 Consecutive Retrieval Property

## 7 Organization on Drum Storage



## 8 Access Path Retrieval

# Preface

In the past few years the area of data management has become an extremely important one in computer science. Large volumes of nonnumerical data are stored in large data banks and are processed by complex queries from time to time as the need arises. The techniques that are used for numerical information processing cannot be easily applied to nonnumerical information because of the complex logical structures inherent in them. Most of the techniques for handling nonnumerical information processing have been developed by practitioners, and very few formal descriptions are available in written form. Researchers associated with prestigious universities and large computer manufacturers have been doing basic research to understand these techniques at a fundamental level. Their works are scattered in many journals, technical reports, proceedings of meetings, etc. This presents some very difficult problems for an instructor putting together a course in fundamentals of data management organization. The problem is more difficult for students. I have lost count of the number of times I have been asked by a student for reprints and references and by teachers to give seminars and lectures in their classes. There is no textbook on the market dealing with the theories of data base organization and usable at a graduate level. There is no book on the market covering such basic concepts as theories of data description models, logical structures of queries, combinational query sets, quadratic residue transformations, balanced filing schemes, the consecutive retrieval property, and organization on drums. These topics can be found in journal articles, which are on a level that good researchers can understand. They are not written for students. In view of the great importance of this area of knowledge in computer science, I felt that this vacuum should not be allowed to remain any more.

The textbooks that have been written on data management are directed toward undergraduate students at elementary or senior level and concentrate on either algorithmic aspects or system aspects of data management. In this book I have tried to put the basic theories and techniques of data management in the foreground, ignoring the systems details of information management. Chapters 1–4 are directed toward senior-level undergraduate students and Chapters 5–8 are directed toward graduate students.

The eight chapters of the book cover the following subjects: (1) data structure, (2) queries and query languages, (3) searching on one field, (4) key to address transformation, (5) algebraic filing schemes, (6) consecutive retrieval property, (7) organization on drum storage, and (8) access path retrieval.

Chapter 1 covers some primitive concepts of data description, basic mathematical ideas relevant to the materials covered in the book, and some data description models, namely, entity, relational, and graph theoretic models.

Chapter 2 covers various types of queries, their parametric representation, simple queries, complex queries, logical structured queries, queries based on a relational algebra set representation query language, etc.

Chapter 3 contains techniques of searching on one field; namely, serial search, sorting, binary tree search, index sequential search, and hierarchical access methods.

Chapter 4 deals with techniques used in key to address transformation, the overflow problems, and comparison of different key to address transformation methods.

Chapter 5 contains filing schemes for answering queries based on multiple values of multiple attributes. In this chapter properties of finite geometries are extensively used for construction of filing schemes.

Chapter 6 deals with file organizations that need no redundant storage and at the same time have minimum access time.

Chapter 7 discusses techniques of organizing records on drum storage; namely, binary search on drums, consecutive storage on drums, scheduling on drums, etc.

Chapter 8 deals with types of access paths used for retrieval of data and definition of logical structures. It contains linear lists, logical access paths, search path algorithms, etc.