



# **Advances in Data Base Management**

## **VOLUME 2**

Edited by

**E. A. UNGER**

Kansas State University  
Manhattan

**P. S. FISHER**

Kansas State University  
Manhattan

and

**JACOB SLONIM**

GEAC Computer International Inc  
Markham, Ontario

*A Wiley Heyden Publication*

**JOHN WILEY & SONS**

Chichester • New York • Brisbane • Toronto • Singapore

Copyright © 1984 by Wiley-Heyden Ltd.

All rights reserved.

No part of this book may be reproduced by any means, nor transmitted, nor translated into a machine language without the written permission of the publisher.

***Library of Congress Catalog Card No.: 81-640184***

***British Library Cataloguing in Publication Data:***

Advances in data base management.—Vol. 2.—(Wiley  
Heyden advances in EDP management)

1. Data base management—Periodicals

001.64'42      OA76.9.D3

ISBN 0 471 26231 5

Typeset by Pintail Studios Ltd,  
Duck Island Lane, Ringwood, Hampshire.  
Printed by Pitman Press, Bath, Avon

**Advances in Data  
Base Management**

**VOLUME 2**

**WILEY-HEYDEN ADVANCES LIBRARY IN EDP MANAGEMENT**

Edited by Thomas A. Rullo

Other Advances Series in the Library

**ADVANCES IN DATA PROCESSING MANAGEMENT Volume 1**

**ADVANCES IN COMPUTER PROGRAMMING MANAGEMENT  
Volume 1**

**ADVANCES IN DISTRIBUTED PROCESSING MANAGEMENT  
Volumes 1 and 2**

**ADVANCES IN DATA COMMUNICATIONS MANAGEMENT  
Volumes 1 and 2**

**ADVANCES IN COMPUTER SECURITY MANAGEMENT  
Volumes 1 and 2**

**ADVANCES IN DATA BASE MANAGEMENT Volume 1**

## LIST OF CONTRIBUTORS

- DANIEL APPLETON, Appleton Company, Inc., P.O. Box 838, Manhattan Beach, California 90266, U.S.A. (p. 119).
- SUDHIR K. ARORA, Ph.D., Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada L8S 4L7. (p. 23).
- DONALD R. DEUTSCH, Ph.D., Manager, Advanced Data Base Management Systems Development, General Electric, Information Services Co., 3200 West End, Suite 201, Nashville, Tennessee 37203, U.S.A. (p. 1).
- JESSE M. DRAPER, Center for Programming Science and Technology Institute for Computer Sciences and Technology, National Bureau of Standards, Washington, D.C. 20234, U.S.A. (p. 1).
- DAVID W. EMBLEY, Ph.D. Department of Computer Science, Brigham Young University, Provo, Utah 84602, U.S.A. (p. 197).
- DENNIS W. FIFE, Ph.D., Systems and Applied Sciences Corporation, 6811 Kenilworth Avenue, Riverdale, Maryland, 20737, U.S.A. (p. 101).
- RICHARD A. FOX, Lt. Col., U.S. Army, West Point Military Academy, West Point, New York, U.S.A. (p. 147).
- W. TERRY HARDGRAVE, Ph.D., Department of Decision Sciences, School of Business Administration, George Mason University, 4400 University Drive, Fairfax, Virginia 22030, U.S.A. (p. 67).
- ALAN R. HEVNER, Ph.D., College of Business and Management, University of Maryland, College Park, Maryland 20742, U.S.A. (p. 225).
- J. PENTON HOLLIST, Major, U.S. Army, 9302 Ivanhoe Road, Fort Washington, Maryland 20744, U.S.A. (p. 43).
- WON KIM, Ph.D., IBM Research Laboratory, K55-282 5600 Coltle Road, San Jose, California 95193, U.S.A. (p. 171).

- L. J. MacRAE, NDX Corporation, P.O. Box 7028, Station A, Toronto, Ontario, Canada M5W1X7. (p. 253).
- FRED J. MARYANSKI, Ph.D., Digital Equipment Corporation, Corporate Research Group, 9 Highfield Drive, Lancaster, Maryland 01523, U.S.A. (p. 253).
- RICHARD A. McBRIDE, Ph.D., Department of Computer Science, Kansas State University, Manhattan, Kansas, 66502, U.S.A. (p. 253).
- EDGAR H. SIBLEY, Ph.D., The Alpha Omega Group, Inc., World Building 8121 Georgia Avenue, Silver Spring, Maryland, 20910, U.S.A. (p. 81).
- JACOB SLONIM, Ph.D., GEAC Computer International Inc., 350 Steelcase Road West, Markham, Ontario, Canada L3R 1B3. (p. 253).
- KENNETH C. SMITH, Ph.D., Chairman, Departments of Electrical Engineering and Computer Science, University of Toronto, Toronto, Ontario, Canada, M5S 1A4. (p. 23).
- E. A. UNGER, Ph.D., Department of Computer Science, Kansas State University, Manhattan, Kansas 66506, U.S.A. (p. 147).

## **PREFACE TO THE WILEY-HEYDEN ADVANCES LIBRARY IN EDP MANAGEMENT**

Information, to be of value, must be timely, accurate, and accessible. By addressing each of these criteria, the *Wiley-Heyden Advances Library in EDP Management* has proved to be a unique and valuable source of information for managers.

The timeliness of the material is maintained by issuing additional volumes within each series. These volumes build upon the base of previous ones to provide new insights and add topics of current interest. Because we continue to generate new ideas our readers receive the most up-to-date information.

The second attribute of useful information is accuracy. To create the *Wiley-Heyden Advances Library in EDP Management*, we assign a separate editor to each volume as an expert in the field, that editor being responsible for the overall content of the volume. Each chapter is in turn developed by an expert in a specific area and then submitted to the volume editor for review. Wherever necessary, additional technical review support is obtained. Because of this detailed editorial and review process, a high degree of accuracy can be assured.

The third attribute that makes information of value is accessibility. This feature is the least commonly discussed and the most difficult of the three to accomplish. In fact the lack of easy accessibility to information was the main reason for developing the *Wiley-Heyden Advances Library in EDP Management*. Our primary role as developers and managers of this base of material is to provide our users with the information they want in a form in which they can readily use it. This task is accomplished through the selection of relevant topic areas, maintenance of a consistent level of content, and organization of material in line with its intended use.

Although we have achieved our timeliness, accuracy, and accessibility objectives, we strive to improve the library through the added dimension of communication with our end users. Thus we welcome any suggestions for improvement of our current series or expansion into new areas.

THOMAS A. RULLO  
*Executive Editor*



## **PREFACE TO VOLUME 2**

The concept of a data base management system is roughly a decade old. During this period of time, many organizations have moved from the concepts of file management systems to the concept of a data management system for their information processing needs. Managers in this phase of information automation are concerned with establishing a uniform corporate information policy covering data, software, hardware, and personnel. This volume was designed to address the issues managers face as the evolution of information management progresses. The articles are tutorial in nature and selection of the authors was based upon their breadth and depth knowledge in an area of concern.

The first chapter by Donald Deutsch and Jesse Draper of the Center for Programming Science of the National Bureau of Standards presents each of the three major data structuring models. This tutorial provides a framework for the reader as the management issues of standardization, selection, tools and efficiency are discussed. The authors include fundamental definitions pertinent to each model, describe the data definition language for establishing the data base under that model, and give the reader a sound overview of the user interface languages.

S. K. Arora and K. C. Smith were selected to write a tutorial on data base machines because of their experience in the design of such devices. The von Neumann machine was designed to accomplish numerical computing 'number crunching,' and it does this well, but is a misfit when used to accomplish the operations involved in information processing. The second chapter provides the reader with an understanding and definitional basis for the area of data base machines. It includes a survey of memory technologies and machine architectures and concludes with a prediction of what the future holds in this area.

The third chapter discusses the information needs of an organization, the data which can support these needs, and the flow of those data through the organization. The author discusses the design of a generalized information system and the use of three tools—software requirements engineering methodology (SREM), input/output requirements language (IORL), and information systems design optimization system (ISDOS)—to accomplish that task.

Terry Hardgrave of the Institute for Computer Science and Technology (ICST)

within the National Bureau of Standards was asked to write a tutorial describing current efforts of that agency in the area of standardizing data base systems. He presents the concept of positional notation and provides several examples illustrating the use of such a notation.

The second section of the book which is addressed to providing a survey of current standardization efforts concludes with an article by Edgar Sibley of the College of Business at the University of Maryland. He approaches this issue by providing a survey of efforts in all those areas impacting data management systems, including the ISO standard for protocols between systems, the operating system, the information resource dictionary standardization (IRDS) effort, and the DBMS and user language standardization efforts.

The third section of this volume is devoted to a discussion of tools for use by managers and others in the selection, design, and use of data management systems. Dennis Fife of the National Bureau of Standards provides an excellent introduction into the concept of the data dictionary. He also surveys a few available dictionary packages, providing the reader with a list of features against which to benchmark available packages. The chapter concludes with a prediction of the future of data dictionaries.

Automated information resource management is the topic of the seventh chapter. The author provides a list to delineate the dimensions involved in data management. This guide is used to illustrate the features of some available systems. A general introduction is provided to information resource management software, systems development life cycles, and the integration process which occurs when one implements the logical structure with a system.

The selection of a DBMS for a single site is the topic of Chapter 8. A two-phase model is designed for use by middle managers who may have little computing technology knowledge. The tool selects those factors which are important in the selection process, poses questions for vendors to answer, defines a framework in which the vendor responses must fit and then quantifies the results. The first phase is designed to be a quick elimination scheme to cull the available systems down to reasonable candidates. The second phase is designed to carefully discriminate between reasonable systems.

Won Kim of the IBM Research Laboratory in San Jose presents an optimization scheme for queries in a relational data base system in Chapter 9. A query language and a programming language based upon ordinary forms is suggested in Chapter 10. This system is potentially useful in application areas where forms are common, e.g. most business applications.

The author of the eleventh chapter, Alan Hevner, provides a survey of recent work in the distribution of data and data retrieval functions when a data base is resident on a distributed system. He includes the minimization of communication and file storage costs for most of the current models of data distribution.

The author of the final chapter focuses on generalized distributed information management architectures. It integrates much of the information in this volume and analyzes for the reader the various functions involved in such a system. The

issue of concurrency in DBMS is discussed and the benefits to be gained from it are illustrated through a simulation analysis of eight different data bases that span the types of most data bases found in use today.

The editors wish to express their appreciation to the various contributors for the quality of their insight and their willingness to impart concise, comprehensive information. Special thanks are also extended to Russell William Hultgren for his indefatigable efforts in compiling and combing the manuscript.

# CONTENTS

List of Contributors . . . . .	xi
Preface to the Wiley–Heyden Advances Library in EDP Management . . . . .	xiii
Preface to Volume 2 . . . . .	xv

## CHAPTER 1: Data Models: Keys to Understanding Data Base Management Systems—

<b>Donald R. Deutsch and Jesse M. Draper</b>		<b>1</b>
Definitions . . . . .		<b>1</b>
<i>Definition of Data Model</i>		
<i>Sample Data Base</i>		
<i>Data Model Descriptions</i>		
The Relational Model . . . . .		<b>3</b>
<i>Structures</i>		
<i>Data Definition Language</i>		
<i>Operations</i>		
<i>User Interfaces</i>		
The Network Model . . . . .		<b>9</b>
<i>Structures</i>		
<i>Data Definition Language</i>		
<i>Operations</i>		
<i>User Interfaces</i>		
The Hierarchical Model . . . . .		<b>16</b>
<i>Structures</i>		
<i>Data Definition Language</i>		
<i>Operations</i>		
<i>User Interfaces</i>		
Using Data Model Concepts . . . . .		<b>20</b>
<i>DBMS Selection</i>		
<i>DBMS Standards Development</i>		
Conclusion . . . . .		<b>21</b>

<b>CHAPTER 2:</b>	Data Base Machines—	
	<b>Sudhir K. Arora and Kenneth C. Smith</b>	23
	Data Base Management Systems .....	23
	Data Base Machines .....	25
	Current Architectures .....	26
	<i>RAP</i>	
	<i>DIRECT</i>	
	<i>DBC</i>	
	Future Architectures .....	31
	<i>WCRC</i>	
	<i>GDBMS</i>	
	Performance .....	37
	Conclusion .....	38
 <b>CHAPTER 3:</b>	What is Information?—	
	<b>J. Penton Hollist</b>	43
	Information and Organizations .....	44
	Information System Design .....	49
	A Generalized Information Design System .....	50
	Input/Output Requirements Language (IORL) .....	51
	Software Requirements Engineering Methodology (SREM) ....	57
	Information Systems Design Optimization System (ISDOS) ....	61
	Summary .....	64
 <b>CHAPTER 4:</b>	Positional Set Notation—	
	<b>W. Terry Hardgrave</b>	67
	Motivation .....	67
	Notation .....	68
	Primitive Mathematical Concepts .....	69
	<i>Sets</i>	
	<i>Pairs, n-Tuples, and Binary Relations</i>	
	<i>Labeled Tuples</i>	
	Positional Set Notation .....	71
	<i>The Syntax of Positional Set Notation</i>	
	<i>Sets, Sequences, and Tuples in PSN</i>	
	<i>Mathematical Description of PSN</i>	
	<i>PSN Definition Using Extended Set Theory</i>	
	The Relational Model .....	75
	A Sample Data Base .....	76
	<i>Domains</i>	
	<i>The EMP Relation</i>	

<i>The SALES Relation</i>	
<i>The SUPPLY Relation</i>	
<i>The TYPE Relation</i>	
Conclusion .....	77
 <b>CHAPTER 5:</b> Standardization and Distributed Processing— <b>Edgar H. Sibley</b>	 81
Operating Systems Control Language Standardization Efforts ..	88
The Information Resource Dictionary Standardization (IRDS) Effort .....	90
The DBMS Standardization Effort .....	96
 <b>CHAPTER 6:</b> The Dictionary Becomes a Tool for System Management— <b>Dennis W. Fife</b>	 101
Basic Features of Dictionary Software .....	102
<i>Entity Types and Attributes</i>	
<i>Relationships and Schema</i>	
<i>Processing</i>	
Management Uses for Dictionary Capabilities .....	107
<i>Maximizing Benefits from Available Resources</i>	
<i>Integrating Requirements to Improve Design</i>	
<i>Controlling Software Development Costs</i>	
Dictionary State of the Art .....	111
Typical Application of the System Dictionary .....	112
<i>Technical Support of IRDS Application</i>	
<i>Application Policy</i>	
<i>Standards and User Responsibilities</i>	
Future Trends .....	115
<i>Conceptual Schema Definition</i>	
<i>System and Software Modeling Languages</i>	
<i>National IRDS Standardization</i>	
 <b>CHAPTER 7:</b> Automating Management of the Information Resource— <b>Daniel Appleton</b>	 119
What are Information Resources? .....	121
IRRIASSPA .....	124
Viewing IRM Software .....	125
<i>Data Dictionary Systems (DDSs)</i>	
<i>Data Definition Languages (DDLs)</i>	
<i>Data Manipulation Languages (DMLs)</i>	
<i>Data Base Query Languages (DQLs)</i>	

<i>Data Base Utilities</i>	
<i>Data Communications (DC) Facilities</i>	
The Systems Development Life Cycle .....	130
What is an Isomorphism? .....	131
Phase 3 System Structures and Information Structures .....	132
Logical Information Structures .....	135
Integrating System Structures and Information Structures .....	137
A Further Look at the Integration Paradox .....	137
Closing the Loop .....	139
Phase 3 Life Cycles .....	140
Forecasting .....	145
 <b>CHAPTER 8:</b> A DBMS Selection Model for Managers—	
<b>Richard W. Fox and E. A. Unger</b>	147
The Tool .....	148
The Filter .....	149
<i>Hardware</i>	
<i>Staff Experience</i>	
<i>Capital Expenditure</i>	
<i>Vendor</i>	
The Discriminator .....	155
<i>User Requirements</i>	
<i>Requirements Consolidation Form</i>	
<i>Vendor Capability Form</i>	
<i>Features Evaluation Form</i>	
<i>System Comparison Form</i>	
Conclusion .....	168
 <b>CHAPTER 9:</b> Query Optimization for Relational Data Base Systems—	
<b>Won Kim</b>	171
Optimization of a Single Query (Part 1) .....	172
<i>Operation and Cost of the Nested-Block Method</i>	
<i>Nested-Block Method of Computing the Product of</i>	
<i>Relations</i>	
<i>Nested-Block Method of Computing Joins of Relations</i>	
<i>Nested-Block Method of Computing the Division of</i>	
<i>Relations</i>	
Optimization of a Single Query (Part 2)	179
<i>Hashing</i>	
<i>Sorting</i>	
Optimization of Nested Queries .....	182
<i>Nesting of Query Blocks</i>	

<i>Processing Nested Queries</i>	
Optimization of Multiple Queries . . . . .	188
<i>Basic Strategy</i>	
<i>Rules for Preprocessing Queries</i>	
<i>Grouping and Sequencing Queries</i>	
Summary and Conclusion . . . . .	192

<b>CHAPTER 10:</b> A Forms-Based Programming System— <b>David W. Embley</b>	197
A Programming System Based on Forms . . . . .	197
<i>Related Research</i>	
Relationship of Forms to Programming Languages and Data Bases . . . . .	199
<i>Forms</i>	
<i>Relationship between Forms and Programming Languages</i>	
<i>Relationships between Forms and Data Bases</i>	
Computational Relationship Formulation . . . . .	203
<i>Data Frames</i>	
<i>An Approach to Computation Formulation</i>	
Data Base Retrieval . . . . .	208
<i>The Forms Data Base Model</i>	
<i>Data Dependencies</i>	
<i>Form Retrieval Requests</i>	
Data Base Update . . . . .	216
<i>A Forms Data Base Editor</i>	
<i>Editing User-Defined Forms</i>	
<i>The View Update Problem</i>	
Form Flow and Data Flow Among Forms . . . . .	219
Conclusion . . . . .	222

<b>CHAPTER 11:</b> Data Allocation and Retrieval in Distributed Systems— <b>Alan R. Hevner</b>	225
Distributed Data Allocation . . . . .	226
<i>Early Work on Distributed Data Allocation</i>	
<i>Distributed Data and Program Allocation</i>	
<i>Distributed Data and Hardware Allocation</i>	
<i>Summary</i>	
Distributed Data Retrieval . . . . .	236
<i>Single File Retrieval</i>	
<i>Distributed Query Optimization</i>	
Combined Research on Distributed Data Allocation and Retrieval . . . . .	248



<b>CHAPTER 12:</b>	A Quantitative Analysis of Information Processing in Centralized and Distributed Architecture— <b>Jacob Slonim, Fred J. Maryanski, L. J. MacRae, E. A. Unger, and Richard A. McBride</b>	253
	Generalized Distributed Information Management Architecture .	254
	A Throughput Model for Concurrent Processing . . . . .	257
	<i>Description of Compared Cases</i>	
	<i>Access Times and Levels of Concurrency: Query Complexity 4</i>	
	<i>Cost and Levels of Concurrency: Query Complexity 4</i>	
	<i>Summary of Experiments</i>	
<b>Appendix A:</b>	Cardenas' Performance Equations and Parameters	274
	<i>Physical Storage Parameters</i>	
	<i>Average List Length for a Query</i>	
	<i>Number of Data Blocks Accessed by a Single Query</i>	
<b>Appendix B:</b>	Equation for Access Time and Cost	277
	<i>Access Time</i>	
	<i>Cost</i>	
<b>Appendix C:</b>	Data Base Characteristics	282
<b>Index</b>		287