

DATA MANAGEMENT

AN ORGANIZATIONAL PERSPECTIVE



Richard T. Watson

Data Management

An Organizational Perspective

Richard T. Watson



John Wiley & Sons, Inc.

New York • Chichester • Brisbane • Toronto • Singapore

Acquisitions Editor	Beth L. Golub
Assistant Marketing Manager	Leslie Hines
Senior Production Editor	John Rouselle
Senior Designer	Kevin Murphy
Assistant Manufacturing Manager	Mark Cirillo
Photo Researcher	Hilary Newman
Illustration Freelance Coordinator	Jaime Perea
Illustrators	Wellington Studios

This book was set in 10/12 Minion by Dignotype and printed and bound by Donnelley/Crawfordsville.
The cover was printed by Phoenix Color.

Recognizing the importance of preserving what has been written, it is a policy of John Wiley & Sons, Inc. to have books of enduring value published in the United States printed on acid-free paper, and we exert our best efforts to that end.

Copyright © 1996, by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

Reproduction or translation of any part of this work beyond that permitted by Sections 107 and 108 of the 1976 United States Copyright Act without the permission of the copyright owner is unlawful. Requests for permission or further information should be addressed to the Permissions Department, John Wiley & Sons, Inc.

ISBN: 0-471-30534-0

Printed in the United States of America

10 9 8 7 6 5 4 3 2

Preface

Information is a key resource for modern organizations. It is a critical input to managerial tasks. Because managers need high quality information to manage change in a turbulent, global environment, many organizations have established systems for storing and retrieving data, the raw material of information. These storage and retrieval systems are an organization's memory. The organization relies on them, just as individuals rely on their personal memory, to continue as a going concern.

The central concern of information systems management is to design, build, and maintain information delivery systems. Information systems management needs to discover its organization's information requirements so that it can design systems to serve these needs. It must merge a system's design and information technology to build an application that provides the organization with data in a timely manner, appropriate format, and at a convenient location. Furthermore, it must manage applications so they evolve to meet changing needs, continue to operate under adverse conditions, and are protected from unauthorized access.

An information delivery system has two components: organizational memory and processes. This book focuses on organizational memory, which is customarily thought of as a database. I deliberately set out to extend this horizon, however, by including all forms of organizational memory, because I believe students need to understand the role of data management that is aligned with current practice. In my view, data management is the design and maintenance of computer-based organizational memory. Thus, you will find a complete section devoted to organizational memory technologies such as hypertext, groupware, and imaging systems.

This book is written for students learning to design and manage organizational memory. Typically, these students are majoring in information systems and taking a required database course. Although this book covers the material found in a typical database text, it is broader in a number of ways. First, it takes the viewpoint that databases are one component of an expansive organizational memory. Information systems professionals need to develop this wider perspective if they are to comprehend the organizational role of information technology. Second, because managers use organizational memory to initiate change, there is considerable discussion of how, why, and when managers use organizational memory.

The decision to start the book with a managerial perspective arises from the belief that successful information systems practice is based on matching managerial needs, social system constraints, and technical opportunities. I want readers to appreciate the "big picture" before they become immersed in the intricacies of data modeling and SQL. In line with this perspective, business stories are used to support and enhance the text. Many of these vignettes serve double duty because they also alert students to current economic trends such as the globalization of business and the growth of the service sector. To provide an international flavor, I selected organizational stories from a variety of nations. The broad, international, managerial approach is one of several innovative pedagogical features.

The first chapter introduces the case study, *The Expeditioner*, which is used in most subsequent chapters to introduce the key themes discussed. Often it sets the scene for the ensuing

material by presenting a common business problem. I hope the case study also injects a little humor.

The second section of the book provides depth in data modeling and SQL. Data modeling is the foundation of database quality. A solid grounding in data modeling principles and extensive practice are necessary for successful database design. In addition, this book exposes students to the full power of SQL.

The book intertwines the treatment of data modeling and SQL because my database teaching experience indicates that students more readily understand the intent of data modeling when they grasp the long-term goal—querying a well-designed relational database. The double helix, upward, intertwined, spiraling of data modeling and SQL is another unique pedagogical feature. Classroom testing indicates it is a superior method of teaching compared to handling data modeling and SQL separately. Students quickly understand the reason for data modeling and appreciate why it is a valuable skill. Also, rapid exposure to SQL means students gain hands-on experience that much sooner.

I intend this book to be a long-term investment for students. There are useful reference sections for data modeling and SQL. The data modeling section details the standard structures and their relational mappings. The SQL section contains an extensive list of queries that serves as a basis for developing other SQL queries. The purpose of these sections is to facilitate “pattern matching.” For example, a student with an SQL query that is similar to a previous problem can rapidly search the SQL reference section to find the closest match. The student can then use the model answer as a guide to formulating the SQL query for the problem at hand. These reference sections are another unique teaching feature that will serve students well during the course and in their subsequent careers.

While I set out to cast data management in a new light, I have not ignored the traditional core of a database course. Section 3 presents database architectures and their implementation. Coverage includes data storage technologies, data and file structures, and the relational, hierarchical, network, and object-oriented models. Naturally, this section reflects a managerial perspective and discusses the trade-offs for the various options facing the data manager.

In keeping with the organizational memory theme introduced in Chapter 1, Section 4 covers other information technologies including the Internet, information services, hypertext, groupware, and imaging systems.

The final section examines the management of organizational memory. The outstanding features of this section are the rigorous treatment of data integrity and data administration.

A student completing this text will:

- have a broad, managerial perspective of an organization’s need for a memory;
- be able to design and create a relational database;
- be able to formulate complex SQL queries;
- have a sound understanding of database architectures and their managerial implications;
- be familiar with the full range of information technologies available for organizational memory;
- understand the fundamentals of data administration;
- know about data management developments and their organizational implications.

My purpose is to create a data management text that is innovative, relevant, and lively. I trust that you will enjoy reading this book and learn a great deal about managing organizational memory in today's organization.

SUPPLEMENTS

Accompanying this book are several aids for the instructor:

- an instructor's manual;
- overhead slides in Power Point 4.0 format;
- all relational tables in the book in electronic format;
- a World Wide Web site to provide additional tracking material.

Instructors should contact their Wiley representative to gain access to these aids.

ACKNOWLEDGMENTS

The support of Beth Golub and David Kear at John Wiley & Sons was much appreciated. They were always cooperative, supportive, and willing to try something new.

I wish to recognize the important contribution of my copyeditor Lorena Akioka. As well as assuring me that someone appreciated my sense of humor, she made the text cleaner and more consistent. I thank John Rouselle, production editor, for his friendly advice and willingness to consider changes to the format of the final work.

I would like to thank my reviewers for their many excellent suggestions and ideas for improving the quality of the content and presentation of this book:

Mohammad Dadashzadeh	Wichita State University
George Federman	Santa Barbara City College
Thomas Lucy-Bouler	Auburn University
Scott McIntyre	University of Colorado
Anthony F. Norcio	University of Maryland
Thomas E. Sandman	California State University—Sacramento
Avanti Sethi	Wichita State University
Linda Volonino	Canisius College

I thank Tore Ørvik of Agder College, Norway for his major contribution to the chapter on object-oriented database. His experience and knowledge of the object-oriented approach were most valuable.

My mate and colleague, Bob Bostrom of The University of Georgia, provided many insights and suggestions, as well as writing two chapters. His extremely thorough review of the manuscript added considerable value. I am very grateful for his many contributions to this project.

Richard T. Watson

Brief Table of Contents

Preface	vii
Section 1 The Managerial Perspective	1
Chapter 1 Organizational Memory	5
Chapter 2 Information	35
Section 2 Data Modeling and SQL	59
Chapter 3 The Single Entity	61
Chapter 4 The One-to-Many Relationship	89
Chapter 5 The Many-to-Many Relationship	108
Chapter 6 One-to-One and Recursive Relationships	124
Chapter 7 Data Modeling	142
Chapter 8 Basic Structures and the Relational Model	185
Chapter 9 SQL	198
Chapter 10 The SQL Playbook	235
Section 3 Database Architectures and Implementations	271
Chapter 11 Data Structure and Storage	275
Chapter 12 Data Storage and Data Processing Architectures	307
Chapter 13 The Relational Model	327
Chapter 14 Hierarchical and Network Models	344
Chapter 15 Object-Oriented Database Management Systems	365
Section 4 Organizational Memory Technologies	399
Chapter 16 Groupware, Hypertext, and Imaging	401
Chapter 17 Internet and Information Services	435

Section 5	Managing Organizational Memory	469
Chapter 18	Data Integrity	470
Chapter 19	Data Administration	503
Glossary		539
Photo Credits		554
Index		555

Contents

Preface	vii
----------------	------------

Section 1 The Managerial Perspective	1
--	----------

Chapter 1 Organizational Memory	5
---	----------

Learning objectives	5
Introduction	5
Individual memory systems	6
Organizational memory	9
Components of organizational memory	15
Problems with organizational memory systems	26
Data, information, and knowledge	28
The challenge	30
Summary	31
Key terms and concepts	32
References and additional readings	32
Exercises	32

Chapter 2 Information	35
---------------------------------	-----------

Learning objectives	35
Introduction	35
A historical perspective	36
Information characteristics	37
Information and organizational change	40
Information and managerial work	47
Information delivery systems	49
Summary	55
Key terms and concepts	56
References and additional readings	56
Exercises	57
Case questions	58

Section 2 Data Modeling and SQL	59
---	-----------

Chapter 3 The Single Entity	61
---------------------------------------	-----------

Learning objectives	61
<i>The Expeditioner</i>	61
Introduction	61
Modeling a single entity database	62

Creating a single table database	63	
Querying a single table database	68	
Summary	85	
Key terms and concepts	85	
Exercises	85	
Chapter 4 The One-to-Many Relationship		89
Learning objectives	89	
<i>The Expeditioner</i>	89	
Relations	89	
Creating a database with a one-to-many relationship	91	
Querying a two-table database	95	
Summary	104	
Key terms and concepts	105	
Exercises	105	
Chapter 5 The Many-to-Many Relationship		108
Learning objectives	108	
<i>The Expeditioner</i>	108	
The many-to-many relationship	109	
Creating a relational database with a many-to-many relationship	111	
Querying a many-to-many relationship	113	
Summary	121	
Key terms and concepts	121	
Exercises	121	
Chapter 6 One-to-One and Recursive Relationships		124
Learning objectives	124	
<i>The Expeditioner</i>	124	
Modeling the one-to-one relationship	125	
Mapping a one-to-one relationship	126	
Mapping a recursive relationship	126	
Querying a one-to-one relationship	128	
Querying a recursive relationship	128	
<i>The Expeditioner</i>	132	
Modeling a one-to-one recursive relationship	132	
Mapping a one-to-one recursive relationship	133	
Querying a one-to-one recursive relationship	133	
<i>The Expeditioner</i>	135	
Modeling a many-to-many recursive relationship	135	
Mapping a many-to-many recursive relationship	136	
Querying a many-to-many recursive relationship	137	
Summary	139	
Key terms and concepts	139	
Exercises	139	

Chapter 7	Data Modeling	142
Introduction	142	
The building blocks	142	
Data model quality	146	
Quality improvement	147	
Entity types	159	
Data modeling hints	162	
Normalization	168	
Other data modeling methods	179	
Summary	180	
Key terms and concepts	181	
References and additional readings	181	
Exercises	181	
 Chapter 8	 Basic Structures and the Relational Model	 185
One entity	185	
Two entities	189	
Relationship descriptors as identifiers	191	
Exercises	196	
 Chapter 9	 SQL	 198
Learning objectives	198	
Introduction	198	
Data definition	199	
Data manipulation	211	
Nulls—much ado about missing information	223	
Security	223	
Synonyms	226	
The catalog	226	
Embedded SQL	227	
Summary	231	
Key terms and concepts	232	
References and additional readings	233	
Exercises	233	
 Chapter 10	 The SQL Playbook	 235
1. A slow full toss	236	
2. Skinning a cat	236	
3. Another full toss	238	
4. Subtracting from all	238	
5. Dividing	239	
6. At least some number	240	
7. A friendly IN for an SQL traveler	240	

8.	Joining a table with itself	240	
9.	A combination of subtract from all and a self-join	241	
10.	Self-join with GROUP BY	241	
11.	A self-join with two matching conditions	241	
12.	Averaging with grouping	242	
13.	Self-join with a GROUP BY and HAVING	242	
14.	An IN with GROUP BY and COUNT	242	
15.	A self-join with some conditions	243	
16.	Making comparisons	243	
17.	An IN with GROUP BY and SUM	243	
18.	A double divide!	244	
19.	A slam dunk	245	
20.	A 6-inch putt for a birdie	245	
21.	Making the count	245	
22.	Minus and divide	245	
23.	Division with copies	246	
24.	A difficult pairing	247	
25.	Two divides and an intersection	247	
26.	A divide with a matching condition	248	
27.	Restricted divide	248	
28.	A NOT IN variation on divide	249	
29.	All and only	249	
30.	Divide with an extra condition	250	
31.	At least some count	250	
32.	Double divide with a restriction	250	
33.	Triple divide with an intersection	251	
34.	An easy one count	252	
35.	The only one	252	
36.	At least some number	253	
37.	A three-table join	253	
38.	Using NOT IN like NOT EXISTS	253	
39.	Minus after grouping	254	
40.	Something to all	254	
41.	Intersection (AND)	254	
42.	Union (OR)	255	
43.	Intersection/union	255	
44.	Averaging with a condition	256	
45.	Averaging with grouping	256	
46.	Averaging with a join, condition, and grouping	256	
47.	Averaging with multiple joins	256	
48.	Complex counting	257	
49.	Summing with joins and conditions	257	
50.	Summing with joins, conditions, and grouping	257	
51.	Advanced summing	258	
52.	Comparing to the average with a join	258	

53. Comparing to the average with a product	259
54. Averaging with multiple grouping	259
55. More than the average with grouping	259
56. The simplest average	260
57. Difference from the average	260
58. Averaging with multiple joins, multiple grouping, and a condition	260
59. Detailed averaging	261
60. Counting pairs	261
61. No Booleans	261
Summary	262
Key terms and concepts	262
Exercises	262
Appendix—Data for Tables	265

Section 3 Database Architectures and Implementations **271**

Chapter 11 Data Structure and Storage **275**

Learning objectives	275
<i>The Expeditioner</i>	275
Introduction	276
Data structures	276
Data storage devices	290
Conclusion	302
Summary	303
Key terms and concepts	304
Exercises	305

Chapter 12 Data Storage and Data Processing Architectures **307**

Learning objectives	307
<i>The Expeditioner</i>	307
Introduction	307
Client/server	315
Distributed database	319
A hybrid architecture	320
Summary	324
Key terms and concepts	325
References and additional readings	325
Exercises	325

Chapter 13 The Relational Model **327**

Learning objectives	327
Background	327
Data structures	328

Integrity rules	331
Manipulation—relational algebra	331
A primitive set of relational operations	337
A fully relational database	337
Summary	341
Key terms and concepts	342
References and additional readings	343
Exercises	343

Chapter 14 Hierarchical and Network Models 344

Learning objectives	344
<i>The Expeditioner</i>	344
Introduction	345
The hierarchical model	346
The network data model and CODASYL/DBTG	355
Data extraction	361
Summary	361
Key terms and concepts	362
Exercises	362

Chapter 15 Object-Oriented Database Management Systems 365

Learning objectives	365
Introduction	365
Historical development	366
Key OO concepts	367
Why OO?	370
Objects and information system modeling	373
<i>The Expeditioner</i>	376
Persistent objects	390
Object-oriented database management systems	390
The future of ODBMS	394
Summary	394
Key terms and concepts	395
References and additional readings	396
Exercises	396

Section 4 Organizational Memory Technologies 399

Chapter 16 Groupware, Hypertext, and Imaging 401

Learning objectives	401
Introduction	401
<i>The Expeditioner</i>	401
Groupware	402
<i>The Expeditioner</i>	418

Hypertext	418
<i>The Expeditioner</i>	424
Imaging	425
Summary	431
Key terms and concepts	432
References and additional readings	433
Exercises	433

Chapter 17 Internet and Information Services 435

Learning objectives	435
Introduction	435
The Internet	436
Information services	455
Summary	466
Key terms and concepts	466
References and additional readings	467
Exercises	467

Section 5 Managing Organizational Memory 469

Chapter 18 Data Integrity 470

Learning objectives	470
<i>The Expeditioner</i>	470
Introduction	470
Single-user and multi-user database systems	474
Protecting existence	475
Maintaining quality	483
Ensuring confidentiality	490
Summary	499
Key terms and concepts	500
References and additional readings	501
Exercises	501

Chapter 19 Data Administration 503

Learning objectives	503
<i>The Expeditioner</i>	503
Introduction	504
Management of the database environment	505
Data administration	508
Database Management System (DBMS)	527
Organizing data administration	530
Summary	536
Key terms and concepts	536

XX Contents

References and additional readings	536
Exercises	537
Glossary	539
Photo Credits	554
Index	555

The Managerial Perspective

People only see what they are prepared to see.
Emerson, *Journals*, 1863

The first section prepares you to see the role of data and information in an organization. The managerial perspective on data management concentrates on why organizations design and maintain memories. Chapter 1 examines this topic by detailing the components of organizational memory and then discussing some of its common problems. The intention is to make you aware of the scope of organizational memory and its many facets.

The second chapter discusses the use of memory in supporting organizational goals. Again, a very broad outlook is adopted in order to provide a sweeping perspective on the relationship of organizational memory to organizational change.

At this point, we want to give you some “maps” for understanding the terrain you will explore. Since the territory is possibly very new, these maps initially may be hard to read and so you may need to read them several times before you understand the terrain you are about to enter.

The first map (see Figure S1–1) is based on the Newell-Simon model¹ of the human information processing system, which shows that humans receive input, process it, and produce output. The processing is done by a processor, which is linked to a memory divided into data and processes. The processor retrieves both data and processes from memory.

To understand this model, consider a person receiving a message to telephone a close friend. The message is input to the human information processing system. The person retrieves the friend’s telephone number from the data portion of his memory. He or she also retrieves the process, or instructions, for making a telephone call (e.g., pick up hand piece, press numbers, and so on). The person then makes the phone call, the processing of the input message. The phone call is the output. Sometimes these processes are so well

¹A. Newell and H. A. Simon, *Human Problem Solving* (Englewood Cliffs, NJ: Prentice-Hall, 1972).