Richard T. Watson

DATA MANAGEMENT

Databases and Organizations

Data Management:

Databases and Organizations

Second edition

Richard T. Watson
Department of Management
The University of Georgia

John Wiley & Sons, Inc.
New York • Chichester • Brisbane • Toronto • Singapore

Acquisitions Editor	Beth L. Golub
Marketing Manager	Carlisle Paulson
Senior Production Editor	Kelly Tavares
Cover Designer	Madelyn Lesure
Cover Image	The Pierpont Morgan Library/Art Resource, NY

This book was set in 10/12 Garamond by Richard and Ned Watson and printed and bound by Donnelley/Crawfordsville. The cover was printed by Lehigh Press.

This book was printed on acid-free paper.

Library of Congress Cataloging -in-Publication Data

Watson, Richard T., 1948-
Data Management : Databases and Organizations / Richard T. Watson.—2nd ed.
p. cm.
Includes index.
ISBN 0-471-18074-2 (cloth : alk. paper)
1. Database management. I.Title
QA76.9.D3W375 1999 98-39266
005.74 — dc21 CIP

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To Clare

Preface

This is not your traditional database textbook. It differs in three fundamental ways.

First, it is deeper than most database books in its coverage of data modeling and SQL. The market seeks graduates who have these fundamental skills. Time and again, students who have completed my data management class have told me how these skills have been invaluable in their first job. The intention is to place great emphasis on the core skills of data management. The consequence is that there is a better match between the skills students develop and market needs. This means that students find this text highly relevant.

Second, the treatment of data modeling and SQL is intertwined 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 a 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.

Third, the book is broader than most database books. Databases are one component of an expansive organizational memory. Information systems professionals need to develop a wide perspective of data management if they are to comprehend fully the organizational role of information technology. Thus the book includes coverage of groupware (e.g., Lotus Notes), imaging systems, and the Web.

In essence, the book is deeper where it matters, data modeling and SQL, and broader to give students a managerial outlook.

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 chang-

ing needs, continue to operate under adverse conditions, and are protected from unauthorized access.

An information delivery system has two components: data and processes. This book focuses on data, which is customarily thought of as a database. I deliberately set out to extend this horizon, however, by including all forms of organizational data stores, 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 section devoted to data management technologies such as groupware and imaging systems.

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 in a data management text.

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 in-depth coverage of 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.

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.

Although 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, client/server models, distributed database, and the 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 groupware, imaging systems, organizational intelligence technologies (data warehousing, OLAP, and data mining), and the Web.

The final section examines the management of organizational data stores. 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 data in today's organization.

Supplements

Accompanying this book are an instructors' manual[1] and an extensive Web site[2] that provides:

- overhead slides in PowerPoint format;
- all relational tables in the book in electronic format;
- answers to many of the exercises;
- additional exercises;
- revisions;
- links to useful Web sites.

Acknowledgments

The support of Beth Golub at John Wiley & Sons was much appreciated. I thank my son, Ned, for help with the typesetting and my wife, Clare, for indexing the book.

1. Instructors should contact Wiley to gain access to the instructors' manual.
2. www.negia.net/~rwatson/

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:

Michael Barrett	University of Alberta, Canada
John E. Boggess	Purdue University
John Bradley	East Carolina University
Traci Carte	University of Oklahoma
Patrick Michael Doran	Hawaii Pacific University
Mohammad Dadashzadeh	Wichita State University
George Federman	Santa Barbara City College
Colin Freeman	University of New South Wales, Australia
Lisa Friedrichsen	Keller Graduate School of Management
Mary Gebelt	Florida Atlantic University
Barbara Haley	University of Virginia
Herman P. Hoplin	Syracuse University
Mark Hwang	Central Michigan University
Lakshmi Iyer	University of Dayton
David Kemp	University of Melbourne, Australia
Someswar Kesh	Central Missouri State University
Constance A. Knapp	Pace University
Chris F. Kemerer	University of Pittsburgh
Shaoyi Liao	City University of Hong Kong, Hong Kong
Thomas Lucy-Bouler	Auburn University
Ronald Maier	University of Regensburg, Germany
Maggie McClintock	Mississippi University for Women
Scott McIntyre	University of Colorado
Anthony F. Norcio	University of Maryland
Bruce Rollier	University of Baltimore
Thomas E. Sandman	California State University—Sacramento
Avanti Sethi	Wichita State University
Liao Shaoyi	City University of Hong Kong, Hong Kong
Ramesh Subramanian	University of Alaska, Anchorage
Ramesh Venkataraman	Indiana University
Linda Volonino	Canisius College
Charles J. Wertz	Buffalo State 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, and contributed extensively to two chapters. His extremely thorough review of the first edition added considerable value. I am very grateful for his many contributions to this project.

Richard T. Watson

Brief Table of Contents

Table of Contents

Section 1

The Managerial Perspective

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

Organizations are accumulating vast volumes of data because of the implementation of technology (e.g., bar codes and scanners) that makes it easier and cheaper for them to collect data. The world's data are estimated to be doubling every 20 months, and many large companies now routinely manage terabytes (10^{12} bytes) of data. Data management has become a key function for many organizations.

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 data management systems, or organizational 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 data management and its many facets. The second chapter discusses the relationship between information and organizational goals. Again, a very broad outlook is adopted in order to provide a sweeping perspective on the relationship of information 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 model[1] 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.

---

1. Newell, A., and H.A. Simon. 1972. *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

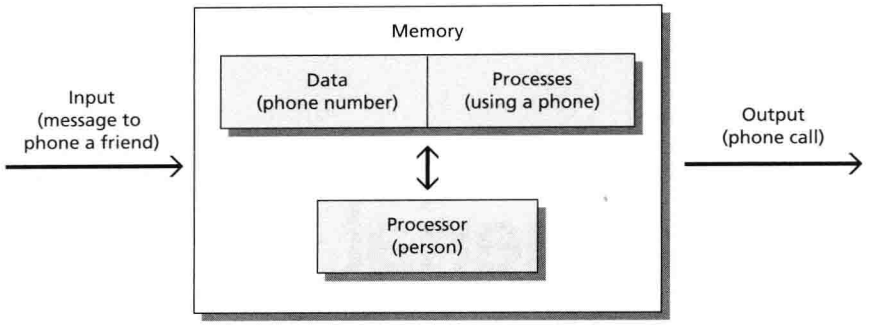

Figure S1-1. The Newell-Simon model of human information processing

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 memory and 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 ingrained in our memory we never think about retrieving them, we just do it automatically.

Human information processing systems can get overloaded easily. Our memory is limited, and our ability to process data is restricted; thus we use a variety of external tools to extend and augment our capacities. A telephone book is an example of external data memory. A recipe, a description of the process for preparing food, is an example of external process memory. Calculators and computer s are examples of external processors we use to augment our limited processing capacity.

Database skills in high demand

Database professionals are in high demand because of the increasing use of client/server technology and massive growth in corporate Web site complexity. Professionals with skills in Oracle, Sybase, and Informix are in highest need. Because of their central role in the development of information systems, database professionals must be able to work well with both IT staff and clients.

Average database specialist salaries in 1997 were $85,000 in New York and $69,000 in San Francisco. Database managers in the same cities averaged $102,000 and $83,000 respectively. Salaries are expected to grow as the shortage is not likely to disappear soon.

Adapted from Steen, M. 1997. Database skills highly sought. *Infoworld*, Nov 24, 1997, 99.

The original model of human information processing can be extended to include external memory, for storing data and processes, and external processors, for executing processes (see Figure S1-2).

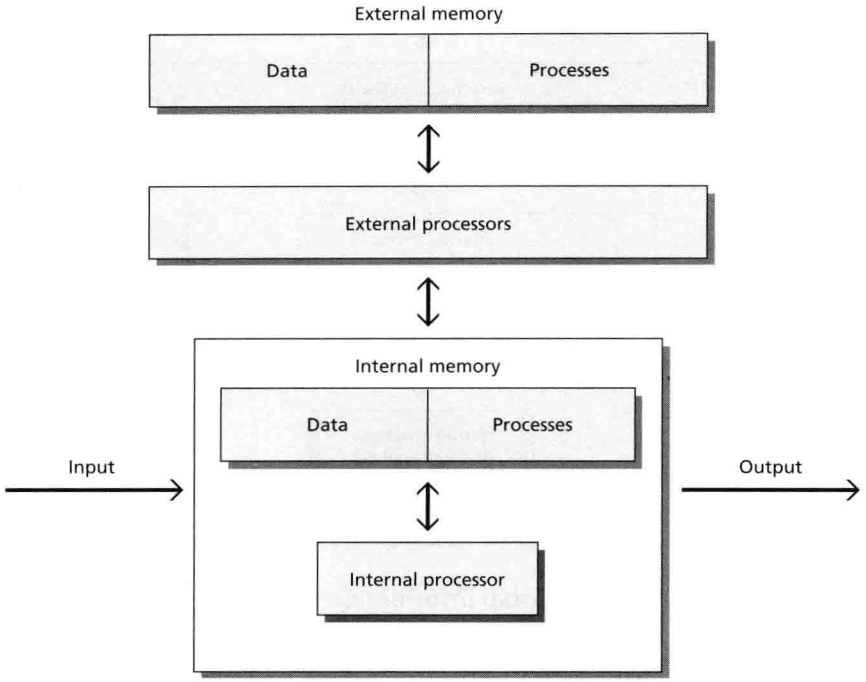

Figure S1-2. An augmented human information processing model

This model of augmented human information processing translates directly to an organizational setting. Organizations collect inputs from the environment—market research, customer complaints, and competitor actions. They process these data and produce outputs—sales campaigns, new products, price changes, and so on. Figure S1-3 gives an example of how an organization might process data. As a result of some market research (input) a marketing analyst (an internal processor) retrieves sales data (data) and does a sale forecast (process). The analyst also requests a marketing consultant (an external processor) to analyze (process) some demographic data (data) before deciding to launch a new promotion (output).

An organization's memory comes in a variety of forms, as you will see in Chapter 1. This memory also can be divided into data and processes. The data part may contain information about customers. The process portion may store details of how to handle a customer order. Organizations use a variety of processors to handle data, including people and computers. Organizations also rely on external sources to extend their information-processing

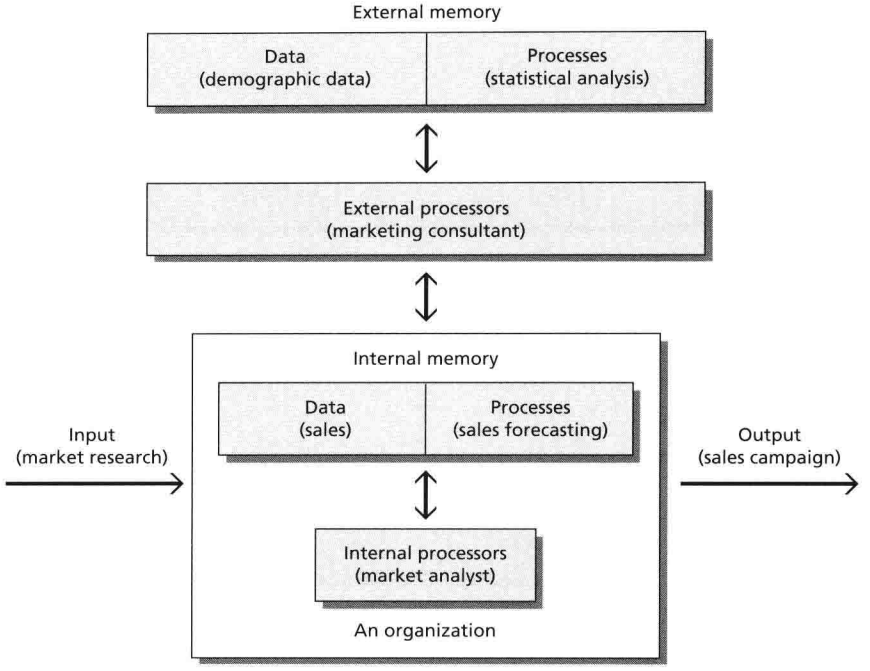

Figure S1-3. An organizational information processing model

capacity. For example, a business may use a specialist credit agency to check a customer's creditworthiness, or an engineering firm may buy time on a university's supercomputer for structural analysis of a bridge. Viewed this way, the augmented human information processing model becomes an organizational information processing system.

This book focuses on the data side of organizational memory. While it is primarily concerned with data stored within the organization, there is also coverage of data in external memory. The process side of organizational memory is typically covered in a systems analysis and design course.