Physical Database Design

The Morgan Kaufmann Series in Data Management Systems

Series Editor: Jim Gray, Microsoft Research

Data Preparation for Data Mining Using SAS
Mamdouh Refaat

Querying XML: XQuery, XPath, and SQL/XML in Context
Jim Melton and Stephen Buxton

Data Mining: Concepts and Techniques, Second Edition
Jiawei Han and Micheline Kamber

Database Modeling and Design: Logical Design, Fourth Edition
Toby J. Teorey, Sam S. Lightstone and Thomas P. Nadeau

Foundations of Multidimensional and Metric Data Structures
Hanan Samet

Joe Celko's SQL for Smarties: Advanced SQL Programming, Third Edition
Joe Celko

Moving Objects Databases
Ralf Hartmut Güting and Markus Schneider

Joe Celko's SQL Programming Style
Joe Celko

Data Mining, Second Edition: Concepts and Techniques
Ian Witten and Eibe Frank

Fuzzy Modeling and Genetic Algorithms for Data Mining and Exploration
Earl Cox

Data Modeling Essentials, Third Edition
Graeme C. Simsion and Graham C. Witt

Location-Based Services
Jochen Schiller and Agnès Voisard

Database Modeling with Microsft® Visio for Enterprise Architects
Terry Halpin, Ken Evans, Patrick Hallock, Bill Maclean

Designing Data-Intensive Web Applications
Stephano Ceri, Piero Fraternali, Aldo Bongio, Marco Brambilla, Sara Comai, and Maristella Matera

Mining the Web: Discovering Knowledge from Hypertext Data
Soumen Chakrabarti

Advanced SQL: 1999—Understanding Object-Relational and Other Advanced Features
Jim Melton

Database Tuning: Principles, Experiments, and Troubleshooting Techniques
Dennis Shasha and Philippe Bonnet

SQL:1999—Understanding Relational Language Components
Jim Melton and Alan R. Simon

Information Visualization in Data Mining and Knowledge Discovery
Edited by Usama Fayyad, Georges G. Grinstein, and Andreas Wierse

Transactional Information Systems: Theory, Algorithms, and Practice of Concurrency Control and Recovery
Gerhard Weikum and Gottfried Vossen

Spatial Databases: With Application to GIS
Philippe Rigaux, Michel Scholl, and Agnes Voisard

Information Modeling and Relational Databases: From Conceptual Analysis to Logical Design
Terry Halpin

Component Database Systems
Edited by Klaus R. Dittrich and Andreas Geppert

Managing Reference Data in Enterprise Databases: Binding Corporate Data to the Wider World
Malcolm Chisholm

Understanding SQL and Java Together: A Guide to SQLJ, JDBC, and Related Technologies
Jim Melton and Andrew Eisenberg

Database: Principles, Programming, and Performance, Second Edition
Patrick and Elizabeth O'Neil

The Object Data Standard: ODMG 3.0
Edited by R. G. G. Cattell and Douglas K. Barry

Data on the Web: From Relations to Semistructured Data and XML
Serge Abiteboul, Peter Buneman, and Dan Suciu

Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations
Ian Witten and Eibe Frank

Joe Celko's SQL for Smarties: Advanced SQL Programming, Second Edition
Joe Celko

Joe Celko's Data and Databases: Concepts in Practice
Joe Celko

Developing Time-Oriented Database Applications in SQL
Richard T. Snodgrass

Web Farming for the Data Warehouse
Richard D. Hackathorn

Management of Heterogeneous and Autonomous Database Systems
Edited by Ahmed Elmagarmid, Marek Rusinkiewicz, and Amit Sheth

Object-Relational DBMSs: Tracking the Next Great Wave, Second Edition
Michael Stonebraker and Paul Brown, with Dorothy Moore

A Complete Guide to DB2 Universal Database
Don Chamberlin

Universal Database Management: A Guide to Object/ Relational Technology
Cynthia Maro Saracco

Readings in Database Systems, Third Edition
Edited by Michael Stonebraker and Joseph M. Hellerstein

Understanding SQL's Stored Procedures: A Complete Guide to SQL/PSM
Jim Melton

Principles of Multimedia Database Systems
V. S. Subrahmanian

Principles of Database Query Processing for Advanced Applications
Clement T. Yu and Weiyi Meng

Advanced Database Systems
Carlo Zaniolo, Stefano Ceri, Christos Faloutsos, Richard T. Snodgrass, V. S. Subrahmanian, and Roberto Zicari

Principles of Transaction Processing
Philip A. Bernstein and Eric Newcomer

Using the New DB2: IBMs Object-Relational Database System
Don Chamberlin

Distributed Algorithms
Nancy A. Lynch

Active Database Systems: Triggers and Rules For Advanced Database Processing
Edited by Jennifer Widom and Stefano Ceri

Migrating Legacy Systems: Gateways, Interfaces, & the Incremental Approach
Michael L. Brodie and Michael Stonebraker

Atomic Transactions
Nancy Lynch, Michael Merritt, William Weihl, and Alan Fekete

Query Processing for Advanced Database Systems
Edited by Johann Christoph Freytag, David Maier, and Gottfried Vossen

Transaction Processing: Concepts and Techniques
Jim Gray and Andreas Reuter

Building an Object-Oriented Database System: The Story of O_2
Edited by François Bancilhon, Claude Delobel, and Paris Kanellakis

Database Transaction Models for Advanced Applications
Edited by Ahmed K. Elmagarmid

A Guide to Developing Client/Server SQL Applications
Setrag Khoshafian, Arvola Chan, Anna Wong, and Harry K. T. Wong

The Benchmark Handbook for Database and Transaction Processing Systems, Second Edition
Edited by Jim Gray

Camelot and Avalon: A Distributed Transaction Facility
Edited by Jeffrey L. Eppinger, Lily B. Mummert, and Alfred Z. Spector

Readings in Object-Oriented Database Systems
Edited by Stanley B. Zdonik and David Maier

Physical Database Design

The Database Professional's Guide to Exploiting Indexes, Views, Storage, and More

Sam Lightstone
Toby Teorey
Tom Nadeau

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO
Morgan Kaufmann Publishers is an imprint of Elsevier

MORGAN KAUFMANN PUBLISHERS

Publisher	Diane D. Cerra
Publishing Services Manager	George Morrison
Project Manager	Marilyn E. Rash
Assistant Editor	Asma Palmeiro
Cover Image	Nordic Photos
Composition:	Multiscience Press, Inc.
Interior Printer	Sheridan Books
Cover Printer	Phoenix Color Corp.

Morgan Kaufmann Publishers is an imprint of Elsevier.
500 Sansome Street, Suite 400, San Francisco, CA 94111

This book is printed on acid-free paper.

Library of Congress Cataloging-in-Publication Data
Lightstone, Sam.
Physical database design : the database professional's guide to exploiting indexes, views, storage, and more / Sam Lightstone, Toby Teorey, and Tom Nadeau.
p. cm. -- (The Morgan Kaufmann series in database management systems)
Includes bibliographical references and index.
ISBN-13: 978-0-12-369389-1 (alk. paper)
ISBN-10: 0-12-369389-6 (alk. paper)
1. Database design. I. Teorey, Toby J. II. Nadeau, Tom, 1958– III. Title.
QA76.9.D26L54 2007
005.74--dc22 2006102899

For information on all Morgan Kaufmann publications, visit our Web site at www.mkp.com or www.books.elsevier.com

Printed in the United States of America
07 08 09 10 11 10 9 8 7 6 5 4 3 2 1

Contents

Preface

Since the development of the relational model by E. F. Codd at IBM in 1970, relational databases have become the de facto standard for managing and querying structured data. The rise of the Internet, online transaction processing, online banking, and the ability to connect heterogeneous systems have all contributed to the massive growth in data volumes over the past 15 years. Terabyte-sized databases have become commonplace. Concurrent with this data growth have come dramatic increases in CPU performance spurred by Moore's Law, and improvements in disk technology that have brought about a dramatic increase in data density for disk storage. Modern databases frequently need to support thousands if not tens of thousands of concurrent users. The performance and maintainability of database systems depends dramatically on their physical design.

A wealth of technologies has been developed by leading database vendors allowing for a fabulous range of physical design features and capabilities. Modern databases can now be sliced, diced, shuffled, and spun in a magnificent set of ways, both in memory and on disk. Until now, however, not much has been written on the topic of physical database design. While it is true that there have been white papers and articles about individual features and even individual products, relatively little has been written on the subject as a whole. Even less has been written to commiserate with database designers over the practical difficulties that the complexity of "creeping featurism" has imposed on the industry. This is all the more reason why a text on physical database design is urgently needed.

We've designed this new book with a broad audience in mind, with both students of database systems and industrial database professionals clearly within its scope. In it

we introduce the major concepts in physical database design, including indexes (B+, hash, bitmap), materialized views (deferred and immediate), range partitioning, hash partitioning, shared-nothing design, multidimensional clustering, server topologies, data distribution, underlying physical subsystems (NUMA, SMP, MPP, SAN, NAS, RAID devices), and much more. In keeping with our goal of writing a book that had appeal to students and database professionals alike, we have tried to concentrate the focus on practical issues and real-world solutions.

In every market segment and in every usage of relational database systems there seems to be nowhere that the problems of physical database design are not a critical concern: from online transaction processing (OLTP), to data mining (DM), to multidimensional online analytical processing (MOLAP), to enterprise resource planning (ERP), to management resource planning (MRP), and in both in-house enterprise systems designed and managed by teams of database administrators (DBAs) and in deployed independent software vendor applications (ISVAs). We hope that the focus on physical database design, usage examples, product-specific syntax, and best practice, will make this book a very useful addition to the database literature.

Organization

An overview of physical database design and where it fits into the database life cycle appears in Chapter 1. Chapter 2 presents the fundamentals of B+tree indexing, the most popular indexing method used in the database industry today. Both simple indexing and composite indexing variations are described, and simple performance measures are used to help compare the different approaches. Chapter 3 is devoted to the basics of query optimization and query execution plan selection from the viewpoint of what a database professional needs to know as background for database design.

Chapters 4 through 8 discuss the individual important design decisions needed for physical database design. Chapter 4 goes into the details about how index selection is done, and what alternative indexing strategies one has to choose from for both selection and join operations. Chapter 5 describes how one goes about choosing materialized views for individual relational databases as well as setting up star schemas for collections of databases in data warehouses. The tradeoffs involved in materialized view selection are illustrated with numerical examples. Chapter 6 explains how to do shared-nothing partitioning to divide and conquer large and computationally complex database problems. The relationship between shared-nothing partitioning, materialized view replication, and indexing is presented.

Chapter 7 is devoted to range partitioning, dividing a large table into multiple smaller tables that hold a specific range of data, and the special indexing problems that need to be addressed. Chapter 8 discusses the benefits of clustering data in general, and how powerful this technique can be when extended to multidimensional data. This

allows a system to cluster along multiple dimensions at the same time without duplicating data.

Chapter 9 discusses the problem of integrating the many physical design decisions by exploring how each decision affects the others, and leads the designer into ways to optimize the design over these many components. Chapter 10 looks carefully at methods of counting and sampling data that help improve the individual techniques of index design, materialized view selection, clustering, and partitioning. Chapter 11 goes more thoroughly into query execution plan selection by discussing tools that allow users to look at the query execution plans and observe whether database decisions on design choices, such as index selection and materialized views, are likely to be useful.

Chapter 12 contains a detailed description of how many of the important physical design decisions are automated by the major relational databases—DB2, SQL Server, and Oracle. It discusses how to use these tools to design efficient databases more quickly. Chapter 13 brings the database designer in touch with the many system issues they need to understand: multiprocessor servers, disk systems, network topologies, disaster recovery techniques, and memory management.

Chapter 14 discusses how physical design is needed to support data warehouses and the OLAP techniques for efficient retrieval of information from them. Chapter 15 defines what is meant by denormalization and illustrates the tradeoffs between degree of normalization and database performance. Finally, Chapter 16 looks at the basics of distributed data allocation strategies including the tradeoffs between the fast query response times due to data replication and the time cost of updates of multiple copies of data.

Appendix A briefly describes a simple computational performance model used to evaluate and compare different physical design strategies on individual databases. The model is used to clarify the tradeoff analysis and design decisions used in physical design methods in several chapters. Appendix B includes a comparison of two commercially available disaster-recovery technologies—IBM's High Availability Disaster Recovery and Oracle's Data Guard.

Each chapter has a tips and insights section for the database professional that gives the reader a useful summary of the design highlights of each chapter. This is followed by a literature summary for further investigation of selected topics on physical design by the reader.

Usage Examples

One of the major differences between logical and physical design is that with physical design the underlying features and physical attributes of the database server (its software and its hardware) begin to matter much more. While logical design can be performed in the abstract, somewhat independent of the products and components that will be used to materialize the design, the same cannot be said for physical design. For this reason we

have made a deliberate effort to include examples in this second book of the major database server products in database server products about physical database design. In this set we include DB2 for zOS v8.1, DB2 9 (Linux, Unix, and Windows), Oracle 10g, SQL Server 2005, Informix Dataserver, and NCR Teradata. We believe that this covers the vast majority of industrial databases in use today. Some popular databases are conspicuously absent, such as MySQL and Sybase, which were excluded simply to constrain the authoring effort.

Literature Summaries and Bibliography

Following the style of the our earlier text on logical database design, *Database Modeling and Design: Logical Design, Fourth Edition*, each chapter concludes with a literature summary. These summaries include the major papers and references for the material covered in the chapter, specifically in two forms:

- Seminal papers that represent the original breakthrough thinking for the physical database design concepts discussed in the chapter.
- Major papers on the latest research and breakthrough thinking.

In addition to the chapter-centric literature summaries, a larger more comprehensive bibliography is included at the back of this book.

Feedback and Errata

If you have comments, we would like to hear from you. In particular, it's very valuable for us to get feedback on both changes that would improve the book as well as errors in the current content. To make this possible we've created an e-mail address to dialogue with our readers: please write to us at db-design@rogers.com.

> Has everyone noticed that all the letters of the word *database* are typed with the left hand? Now the layout of the QWERTY typewriter keyboard was designed among other things to facilitate the even use of both hands. It follows, therefore, that among other things, writing about databases is not only unnatural, but a lot harder than it appears.
>
> —*Anonymous*

While this quip may appeal to the authors who had to personally suffer through left-hand-only typing of the word *database* several hundred times in the authoring of this book,[1] if you substitute the words "writing about databases" with "designing data-

[1] Confession of a bad typist: I use my right hand for the t and b. This is an unorthodox but necessary variation for people who need to type the word "database" dozens of times per day.

bases," the statement rings even more powerfully true for the worldwide community of talented database designers.

Acknowledgments

As with any text of this breadth, there are many people aside from the authors who contribute to the reviewing, editing, and publishing that make the final text what it is. We'd like to pay special thanks to the following people from a range of companies and consulting firms who contributed to the book: Sanjay Agarwal, Eric Alton, Hermann Baer, Kevin Beck, Surajit Chaudhuri, Kitman Cheung, Leslie Cranston, Yuri Deigin, Chris Eaton, Scott Fadden, Lee Goddard, Peter Haas, Scott Hayes, Lilian Hobbs, John Hornibrook, Martin Hubel, John Kennedy, Eileen Lin, Guy Lohman, Wenbin Ma, Roman Melnyk, Mughees Minhas, Vivek Narasayya, Jack Raitto, Haider Rizvi, Peter Shum, Danny Zilio and Calisto Zuzarte. Thank you to Linda Peterson and Rebekah Smith for their help with manuscript preparation.

We also would like to thank the reviewers of this book who provided a number of extremely valuable insights. Their in-depth reviews and new directions helped us produce a much better text. Thank you to Mike Blaha, Philippe Bonnet, Philipe Carino, and Patrick O'Neil. Thank you as well to the concept reviewers Bob Muller, Dorian Pyle, James Bean, Jim Gray, and Michael Blaha.

We would like to thank our wives and children for their support and for allowing us the time to work on this project, often into the wee hours of the morning.

To the community of students and database designers worldwide, we salute you. Your job is far more challenging and complex than most people realize. Each of the possible design attributes in a modern relational database system is very complex in its own right. Tackling all of them, as real database designers must, is a remarkable challenge that by all accounts ought to be impossible for mortal human beings. In fact, optimal database design can be shown mathematically to truly be impossible for any moderately involved system. In one analysis we found that the possible design choices for an average database far exceeded the current estimates of the number of atoms in the universe (10^{81}) by several orders of magnitude! And yet, despite the massive complexity and sophistication of modern database systems, you have managed to study them, master them, and continue to design them. The world's data is literally in your hands. We hope this book will be a valuable tool for you. By helping you, the students and designers of database systems, we hope this book will also lead in a small incremental but important way to improvements in the world's data management infrastructure.

> Engineering is a great profession. There is the satisfaction of watching a figment of the imagination emerge through the aid of science to a plan on paper. Then it moves to realization in stone or metal or energy. Then it brings homes to men or women. Then

it elevates the standard of living and adds to the comforts of life. This is the engineer's high privilege.

—Herbert Hoover (1874–1964)

The most likely way for the world to be destroyed, most experts agree, is by accident. That's where we come in; we're computer professionals. We cause accidents.

—Nathaniel Borenstein (1957–)