WILEY

KIMBALL GROUP

The Data Warehouse Toolkit

Third Edition

The Definitive Guide to Dimensional Modeling

Ralph Kimball
Margy Ross

The Data Warehouse Toolkit

The Data Warehouse Toolkit

The Definitive Guide to Dimensional Modeling

Third Edition

Ralph Kimball
Margy Ross

WILEY

The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition

Published by
John Wiley & Sons, Inc.
10475 Crosspoint Boulevard
Indianapolis, IN 46256
www.wiley.com

Published by John Wiley & Sons, Inc., Indianapolis, Indiana

Published simultaneously in Canada

ISBN: 978-1-118-53080-1
ISBN: 978-1-118-53077-1 (ebk)
ISBN: 978-1-118-73228-1 (ebk)
ISBN: 978-1-118-73219-9 (ebk)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Library of Congress Control Number: 2013936841

About the Authors

Ralph Kimball founded the Kimball Group. Since the mid-1980s, he has been the data warehouse and business intelligence industry's thought leader on the dimensional approach. He has educated tens of thousands of IT professionals. The Toolkit books written by Ralph and his colleagues have been the industry's best sellers since 1996. Prior to working at Metaphor and founding Red Brick Systems, Ralph coinvented the Star workstation, the first commercial product with windows, icons, and a mouse, at Xerox's Palo Alto Research Center (PARC). Ralph has a PhD in electrical engineering from Stanford University.

Margy Ross is president of the Kimball Group. She has focused exclusively on data warehousing and business intelligence since 1982 with an emphasis on business requirements and dimensional modeling. Like Ralph, Margy has taught the dimensional best practices to thousands of students; she also coauthored five Toolkit books with Ralph. Margy previously worked at Metaphor and cofounded DecisionWorks Consulting. She graduated with a BS in industrial engineering from Northwestern University.

Credits

Executive Editor
Robert Elliott

Project Editor
Maureen Spears

Senior Production Editor
Kathleen Wisor

Copy Editor
Apostrophe Editing Services

Editorial Manager
Mary Beth Wakefield

Freelancer Editorial Manager
Rosemarie Graham

Associate Director of Marketing
David Mayhew

Marketing Manager
Ashley Zurcher

Business Manager
Amy Knies

Production Manager
Tim Tate

Vice President and Executive Group Publisher
Richard Swadley

Vice President and Executive Publisher
Neil Edde

Associate Publisher
Jim Minatel

Project Coordinator, Cover
Katie Crocker

Proofreader
Word One, New York

Indexer
Johnna VanHoose Dinse

Cover Image
iStockphoto.com / teekid

Cover Designer
Ryan Sneed

Acknowledgments

First, thanks to the hundreds of thousands who have read our Toolkit books, attended our courses, and engaged us in consulting projects. We have learned as much from you as we have taught. Collectively, you have had a profoundly positive impact on the data warehousing and business intelligence industry. Congratulations!

Our Kimball Group colleagues, Bob Becker, Joy Mundy, and Warren Thornthwaite, have worked with us to apply the techniques described in this book literally thousands of times, over nearly 30 years of working together. Every technique in this book has been thoroughly vetted by practice in the real world. We appreciate their input and feedback on this book—and more important, the years we have shared as business partners, along with Julie Kimball.

Bob Elliott, our executive editor at John Wiley & Sons, project editor Maureen Spears, and the rest of the Wiley team have supported this project with skill and enthusiasm. As always, it has been a pleasure to work with them.

To our families, thank you for your unconditional support throughout our careers. Spouses Julie Kimball and Scott Ross and children Sara Hayden Smith, Brian Kimball, and Katie Ross all contributed in countless ways to this book.

Contents

Introduction

The data warehousing and business intelligence (DW/BI) industry certainly has matured since Ralph Kimball published the first edition of *The Data Warehouse Toolkit* (Wiley) in 1996. Although large corporate early adopters paved the way, DW/BI has since been embraced by organizations of all sizes. The industry has built thousands of DW/BI systems. The volume of data continues to grow as warehouses are populated with increasingly atomic data and updated with greater frequency. Over the course of our careers, we have seen databases grow from megabytes to gigabytes to terabytes to petabytes, yet the basic challenge of DW/BI systems has remained remarkably constant. Our job is to marshal an organization's data and bring it to business users for their decision making. Collectively, you've delivered on this objective; business professionals everywhere are making better decisions and generating payback on their DW/BI investments.

Since the first edition of *The Data Warehouse Toolkit* was published, dimensional modeling has been broadly accepted as the dominant technique for DW/BI presentation. Practitioners and pundits alike have recognized that the presentation of data must be grounded in simplicity if it is to stand any chance of success. Simplicity is the fundamental key that allows users to easily understand databases and software to efficiently navigate databases. In many ways, dimensional modeling amounts to holding the fort against assaults on simplicity. By consistently returning to a business-driven perspective and by refusing to compromise on the goals of user understandability and query performance, you establish a coherent design that serves the organization's analytic needs. This dimensionally modeled framework becomes the *platform for BI*. Based on our experience and the overwhelming feedback from numerous practitioners from companies like your own, we believe that dimensional modeling is absolutely critical to a successful DW/BI initiative.

Dimensional modeling also has emerged as the leading architecture for building integrated DW/BI systems. When you use the conformed dimensions and conformed facts of a set of dimensional models, you have a practical and predictable framework for incrementally building complex DW/BI systems that are inherently distributed.

For all that has changed in our industry, the core dimensional modeling techniques that Ralph Kimball published 17 years ago have withstood the test of time. Concepts such as conformed dimensions, slowly changing dimensions, heterogeneous products, factless fact tables, and the enterprise data warehouse bus matrix

continue to be discussed in design workshops around the globe. The original concepts have been embellished and enhanced by new and complementary techniques. We decided to publish this third edition of Kimball's seminal work because we felt that it would be useful to summarize our collective dimensional modeling experience under a single cover. We have each focused exclusively on decision support, data warehousing, and business intelligence for more than three decades. We want to share the dimensional modeling patterns that have emerged repeatedly during the course of our careers. This book is loaded with specific, practical design recommendations based on real-world scenarios.

The goal of this book is to provide a one-stop shop for dimensional modeling techniques. True to its title, it is a toolkit of dimensional design principles and techniques. We address the needs of those just starting in dimensional DW/BI and we describe advanced concepts for those of you who have been at this a while. We believe that this book stands alone in its depth of coverage on the topic of dimensional modeling. It's the definitive guide.

Intended Audience

This book is intended for data warehouse and business intelligence designers, implementers, and managers. In addition, business analysts and data stewards who are active participants in a DW/BI initiative will find the content useful.

Even if you're not directly responsible for the dimensional model, we believe it is important for all members of a project team to be comfortable with dimensional modeling concepts. The dimensional model has an impact on most aspects of a DW/BI implementation, beginning with the translation of business requirements, through the extract, transformation and load (ETL) processes, and finally, to the unveiling of a data warehouse through business intelligence applications. Due to the broad implications, you need to be conversant in dimensional modeling regardless of whether you are responsible primarily for project management, business analysis, data architecture, database design, ETL, BI applications, or education and support. We've written this book so it is accessible to a broad audience.

For those of you who have read the earlier editions of this book, some of the familiar case studies will reappear in this edition; however, they have been updated significantly and fleshed out with richer content, including sample enterprise data warehouse bus matrices for nearly every case study. We have developed vignettes for new subject areas, including big data analytics.

The content in this book is somewhat technical. We primarily discuss dimensional modeling in the context of a relational database with nuances for online