

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

analytical processing (OLAP) cubes noted where appropriate. We presume you have basic knowledge of relational database concepts such as tables, rows, keys, and joins. Given we will be discussing dimensional models in a nondenominational manner, we won't dive into specific physical design and tuning guidance for any given database management systems.

Chapter Preview

The book is organized around a series of business vignettes or case studies. We believe developing the design techniques by example is an extremely effective approach because it allows us to share very tangible guidance and the benefits of real world experience. Although not intended to be full-scale application or industry solutions, these examples serve as a framework to discuss the patterns that emerge in dimensional modeling. In our experience, it is often easier to grasp the main elements of a design technique by stepping away from the all-too-familiar complexities of one's own business. Readers of the earlier editions have responded very favorably to this approach.

Be forewarned that we deviate from the case study approach in Chapter 2: Kimball Dimensional Modeling Techniques Overview. Given the broad industry acceptance of the dimensional modeling techniques invented by the Kimball Group, we have consolidated the official listing of our techniques, along with concise descriptions and pointers to more detailed coverage and illustrations of these techniques in subsequent chapters. Although not intended to be read from start to finish like the other chapters, we feel this technique-centric chapter is a useful reference and can even serve as a professional checklist for DW/BI designers.

With the exception of Chapter 2, the other chapters of this book build on one another. We start with basic concepts and introduce more advanced content as the book unfolds. The chapters should be read in order by every reader. For example, it might be difficult to comprehend Chapter 16: Insurance, unless you have read the preceding chapters on retailing, procurement, order management, and customer relationship management.

Those of you who have read the last edition may be tempted to skip the first few chapters. Although some of the early fact and dimension grounding may be familiar turf, we don't want you to sprint too far ahead. You'll miss out on updates to fundamental concepts if you skip ahead too quickly.

NOTE This book is laced with tips (like this note), key concept listings, and chapter pointers to make it more useful and easily referenced in the future.

Chapter 1: Data Warehousing, Business Intelligence, and Dimensional Modeling Primer

The book begins with a primer on data warehousing, business intelligence, and dimensional modeling. We explore the components of the overall DW/BI architecture and establish the core vocabulary used during the remainder of the book. Some of the myths and misconceptions about dimensional modeling are dispelled.

Chapter 2: Kimball Dimensional Modeling Techniques Overview

This chapter describes more than 75 dimensional modeling techniques and patterns. This official listing of the Kimball techniques includes forward pointers to subsequent chapters where the techniques are brought to life in case study vignettes.

Chapter 3: Retail Sales

Retailing is the classic example used to illustrate dimensional modeling. We start with the classic because it is one that we all understand. Hopefully, you won't need to think very hard about the industry because we want you to focus on core dimensional modeling concepts instead. We begin by discussing the four-step process for designing dimensional models. We explore dimension tables in depth, including the date dimension that will be reused repeatedly throughout the book. We also discuss degenerate dimensions, snowflaking, and surrogate keys. Even if you're not a retailer, this chapter is required reading because it is chock full of fundamentals.

Chapter 4: Inventory

We remain within the retail industry for the second case study but turn your attention to another business process. This chapter introduces the enterprise data warehouse bus architecture and the bus matrix with conformed dimensions. These concepts are critical to anyone looking to construct a DW/BI architecture that is integrated and extensible. We also compare the three fundamental types of fact tables: transaction, periodic snapshot, and accumulating snapshot.

Chapter 5: Procurement

This chapter reinforces the importance of looking at your organization's value chain as you plot your DW/BI environment. We also explore a series of basic and advanced techniques for handling slowly changing dimension attributes; we've built on the long-standing foundation of type 1 (overwrite), type 2 (add a row), and type 3 (add a column) as we introduce readers to type 0 and types 4 through 7.

Chapter 6: Order Management

In this case study, we look at the business processes that are often the first to be implemented in DW/BI systems as they supply core business performance metrics—what are we selling to which customers at what price? We discuss dimensions that play multiple roles within a schema. We also explore the common challenges modelers face when dealing with order management information, such as header/line item considerations, multiple currencies or units of measure, and junk dimensions with miscellaneous transaction indicators.

Chapter 7: Accounting

We discuss the modeling of general ledger information for the data warehouse in this chapter. We describe the appropriate handling of year-to-date facts and multiple fiscal calendars, as well as consolidated fact tables that combine data from multiple business processes. We also provide detailed guidance on dimension attribute hierarchies, from simple denormalized fixed depth hierarchies to bridge tables for navigating more complex ragged, variable depth hierarchies.

Chapter 8: Customer Relationship Management

Numerous DW/BI systems have been built on the premise that you need to better understand and service your customers. This chapter discusses the customer dimension, including address standardization and bridge tables for multivalued dimension attributes. We also describe complex customer behavior modeling patterns, as well as the consolidation of customer data from multiple sources.

Chapter 9: Human Resources Management

This chapter explores several unique aspects of human resources dimensional models, including the situation in which a dimension table begins to behave like a fact table. We discuss packaged analytic solutions, the handling of recursive management hierarchies, and survey questionnaires. Several techniques for handling multivalued skill keyword attributes are compared.

Chapter 10: Financial Services

The banking case study explores the concept of supertype and subtype schemas for heterogeneous products in which each line of business has unique descriptive attributes and performance metrics. Obviously, the need to handle heterogeneous products is not unique to financial services. We also discuss the complicated relationships among accounts, customers, and households.

Chapter 11: Telecommunications

This chapter is structured somewhat differently to encourage you to think critically when performing a dimensional model design review. We start with a dimensional design that looks plausible at first glance. Can you find the problems? In addition, we explore the idiosyncrasies of geographic location dimensions.

Chapter 12: Transportation

In this case study we look at related fact tables at different levels of granularity while pointing out the unique characteristics of fact tables describing segments in a journey or network. We take a closer look at date and time dimensions, covering country-specific calendars and synchronization across multiple time zones.

Chapter 13: Education

We look at several factless fact tables in this chapter. In addition, we explore accumulating snapshot fact tables to handle the student application and research grant proposal pipelines. This chapter gives you an appreciation for the diversity of business processes in an educational institution.

Chapter 14: Healthcare

Some of the most complex models that we have ever worked with are from the healthcare industry. This chapter illustrates the handling of such complexities, including the use of a bridge table to model the multiple diagnoses and providers associated with patient treatment events.

Chapter 15: Electronic Commerce

This chapter focuses on the nuances of clickstream web data, including its unique dimensionality. We also introduce the step dimension that's used to better understand any process that consists of sequential steps.

Chapter 16: Insurance

The final case study reinforces many of the patterns we discussed earlier in the book in a single set of interrelated schemas. It can be viewed as a pulling-it-all-together chapter because the modeling techniques are layered on top of one another.

Chapter 17: Kimball Lifecycle Overview

Now that you are comfortable designing dimensional models, we provide a high-level overview of the activities encountered during the life of a typical DW/BI project. This chapter is a lightning tour of *The Data Warehouse Lifecycle Toolkit, Second Edition* (Wiley, 2008) that we coauthored with Bob Becker, Joy Mundy, and Warren Thornthwaite.

Chapter 18: Dimensional Modeling Process and Tasks

This chapter outlines specific recommendations for tackling the dimensional modeling tasks within the Kimball Lifecycle. The first 16 chapters of this book cover dimensional modeling techniques and design patterns; this chapter describes responsibilities, how-tos, and deliverables for the dimensional modeling design activity.

Chapter 19: ETL Subsystems and Techniques

The extract, transformation, and load system consumes a disproportionate share of the time and effort required to build a DW/BI environment. Careful consideration of best practices has revealed 34 subsystems found in almost every dimensional data warehouse back room. This chapter starts with the requirements and constraints that must be considered before designing the ETL system and then describes the 34 extraction, cleaning, conforming, delivery, and management subsystems.

Chapter 20: ETL System Design and Development Process and Tasks

This chapter delves into specific, tactical dos and don'ts surrounding the ETL design and development activities. It is required reading for anyone tasked with ETL responsibilities.

Chapter 21: Big Data Analytics

We focus on the popular topic of big data in the final chapter. Our perspective is that big data is a natural extension of your DW/BI responsibilities. We begin with an overview of several architectural alternatives, including MapReduce and

Hadoop, and describe how these alternatives can coexist with your current DW/BI architecture. We then explore the management, architecture, data modeling, and data governance best practices for big data.

Website Resources

The Kimball Group's website is loaded with complementary dimensional modeling content and resources:

- Register for *Kimball Design Tips* to receive practical guidance about dimensional modeling and DW/BI topics.
- Access the archive of more than 300 *Design Tips* and articles.
- Learn about public and onsite Kimball University classes for quality, vendor-independent education consistent with our experiences and writings.
- Learn about the Kimball Group's consulting services to leverage our decades of DW/BI expertise.
- Pose questions to other dimensionally aware participants on the Kimball Forum.

Summary

The goal of this book is to communicate the official dimensional design and development techniques based on the authors' more than 60 years of experience and hard won lessons in real business environments. DW/BI systems must be driven from the needs of business users, and therefore are designed and presented from a simple dimensional perspective. We are confident you will be one giant step closer to DW/BI success if you buy into this premise.

Now that you know where you are headed, it is time to dive into the details. We'll begin with a primer on DW/BI and dimensional modeling in Chapter 1 to ensure that everyone is on the same page regarding key terminology and architectural concepts.