

WILEY



大数据应用与技术丛书

The Kimball Group Reader: Relentlessly Practical Tools for
Data Warehousing and Business Intelligence, Second Edition

数据仓库与商业 智能宝典(第2版)

成功设计、部署和维护DW/BI系统

Ralph Kimball

Margy Ross

[美] Bob Becker

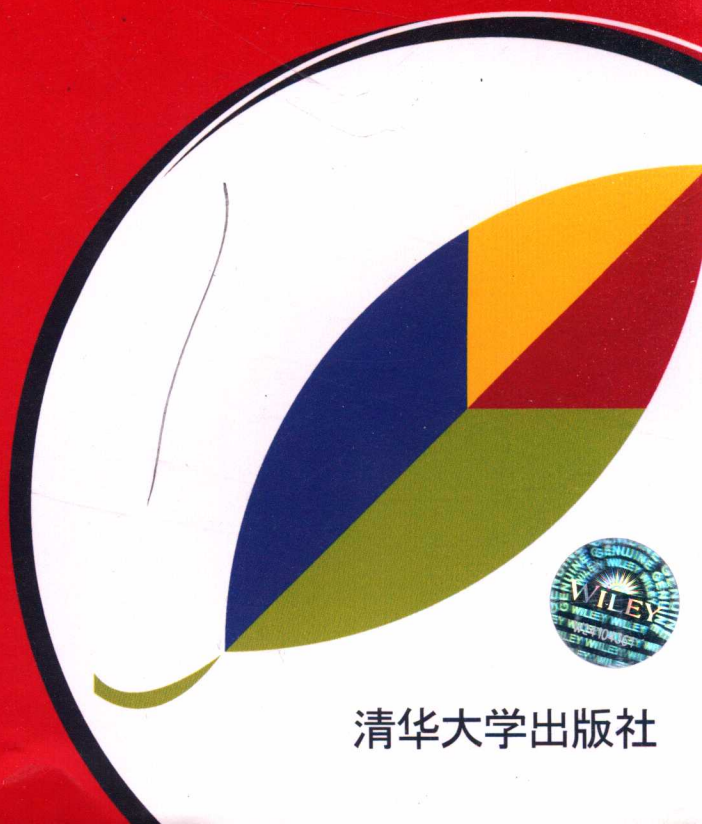
Joy Mundy

Warren Thornthwaite

蒲 成

著

译



清华大学出版社

大数据应用与技术丛书

数据仓库与 商业智能宝典

(第2版)

成功设计、部署和维护 DW/BI 系统

Ralph Kimball
Margy Ross
[美] Bob Becker
Joy Mundy
Warren Thornthwaite
蒲 成 译



C00521500
清华大学出版社
北 京

mm005x200mm : 本
2011年10月 : 第
1-3200 : 册
00.211 : 价
元

10-300050 : 印

Ralph Kimball, Margy Ross, Bob Becker, Joy Mundy, Warren Thornthwaite
The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence,
Second Edition

EISBN: 978-1-119-21631-5

Copyright © 2016 by Ralph Kimball and Margy Ross

All Rights Reserved. This translation published under license.

Trademarks: Wiley and the Wiley logo are trademarks or registered trademarks of John Wiley & Sons, Inc., and/or its affiliates, in the United States and other countries, and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc., is not associated with any product or vendor mentioned in this book.

北京市版权局著作权合同登记号 图字: 01-2016-8091

Copies of this book sold without a Wiley sticker on the cover are unauthorized and illegal.

本书封面贴有 Wiley 公司防伪标签, 无标签者不得销售。

版权所有, 侵权必究。侵权举报电话: 010-62782989 13701121933

图书在版编目(CIP)数据

数据仓库与商业智能宝典: 成功设计、部署和维护DW/BI系统: 第2版 / (美)拉尔夫·金博尔(Ralph Kimball) 等著; 蒲成 译. —北京: 清华大学出版社, 2017

(大数据应用与技术丛书)

书名原文: The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence,
Second Edition

ISBN 978-7-302-47579-8

I. ①数… II. ①拉… ②蒲… III. ①数据库系统 IV. ①TP311.13

中国版本图书馆 CIP 数据核字(2017)第 150840 号

责任编辑: 王 军 于 平

装帧设计: 孔祥峰

责任校对: 成凤进

责任印制: 王静怡

出版发行: 清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址: 北京清华大学学研大厦 A 座 邮 编: 100084

社 总 机: 010-62770175

邮 购: 010-62786544

投稿与读者服务: 010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者: 北京鑫海金澳胶印有限公司

经 销: 全国新华书店

开 本: 185mm×260mm 印 张: 44.25 字 数: 1161 千字

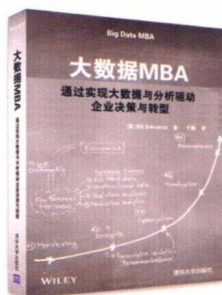
版 次: 2017 年 8 月第 1 版 印 次: 2017 年 8 月第 1 次印刷

印 数: 1~3500

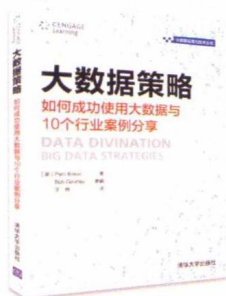
定 价: 118.00 元

产品编号: 072066-01

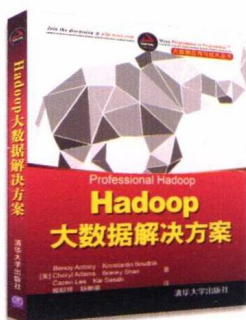
大数据应用与技术丛书



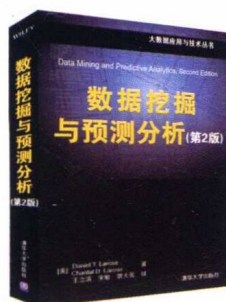
大数据是第四生产要素，数据的价值日益不可估量。越来越多的政府、企业、机构开始意识到数据正在成为最重要的资产，数据分析能力正在成为组织的核心竞争力。本书作者是大数据行业领军企业易安信的首席技术官。通过引入“大数据商业模式成熟度指数”概念，分析具体企业的大数据运营案例，指导企业像“数据科学家一样思考”，化繁为简，教会企业如何利用大数据和数据科学构建合理的关键商业计划，获得竞争优势，挖掘隐藏在大数据中的巨大财富。



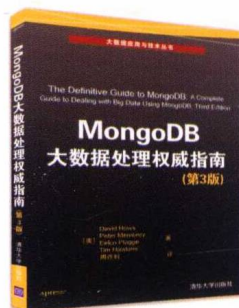
大数据正在改变我们的世界。互联网发展以及移动通信市场和相关技术的迅速扩展已创建大量数据，包括结构化数据和非结构化数据。数据可用性和数据应用对商业和更广泛的社会领域带来巨大影响。有效使用大数据有助于公司更准确地对重要信息进行分析，最终提高运营效率，降低成本，加快创新，增加收入。本书详细介绍大数据策略的规划和执行，配以10个不同行业中不计其数的现实案例予以阐述。你将了解大数据的概念以及如何运用大数据。每章都会解答关键问题，并列出了需要掌握的技能，以确保大数据项目成功。



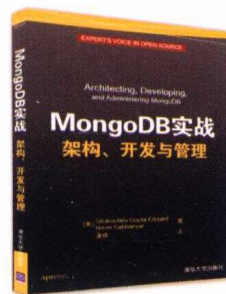
本书单独涵盖了各个组件，最后用实际项目将它们联系起来并构建示例应用。跳过数据库开发基础，本书直奔主题，帮助有经验的开发者更上一层楼，并开始在真实场景中使用Hadoop。



《数据挖掘与预测分析(第2版)》提供了从数据准备到探索性数据分析、数据建模及模型评估等整个数据分析过程的内容。《数据挖掘与预测分析(第2版)》不仅提供了理解软件底层算法的“白盒”方法，而且提供了能够使读者利用现实世界数据集开展数据挖掘与预测分析的应用方法。

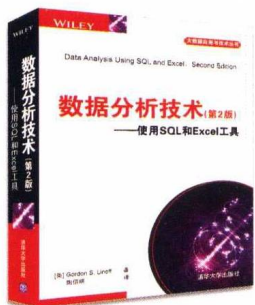


本书针对MongoDB 3做了精细更新，呈现MongoDB的所有最新特性，涵盖2.2版引入的聚集框架、2.4版引入的哈希索引以及3.2版本的WiredTiger，还新纳入Node.js和Python。MongoDB是最流行的“大数据”NoSQL数据库技术，目前仍在蓬勃发展。

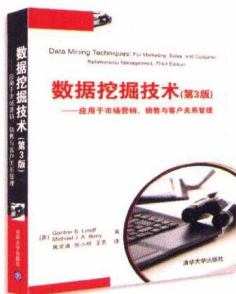


本书涵盖了数据模型、底层架构、使用Mongo Shell编码、管理MongoDB平台以及其他主题。本书还提供了使用MongoDB平台进行架构、开发与部署应用程序的清晰指导与实践示例。数据库开发人员、架构师和管理员将在本书中找到涵盖MongoDB平台所有知识点的有用信息，以及如何将它用于实践。

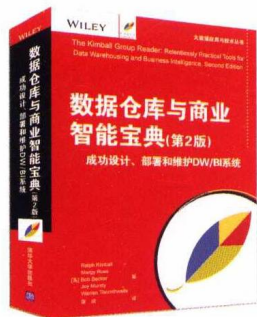
大数据应用与技术丛书



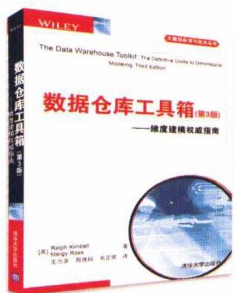
SQL是数据查询的基本语言，Excel是数据分析和展示的最常见工具。两者结合，可以组成一个强大且易于理解的业务数据分析工具。很多类重要的数据分析并不需要复杂且昂贵的数据挖掘工具。答案就在你的电脑桌上。



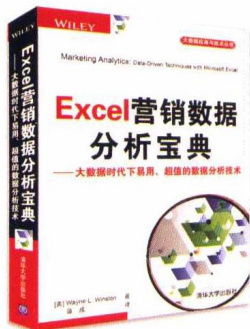
自本书第1版问世以来，数据挖掘已经日益成为现代商业不可缺少的工具。在这个最新版本中，作者对每个章节都进行了大量更新和修订，并添加了几个新章节。本书保留了早期版本的重点，指导市场分析师、业务经理和数据挖掘专家利用数据挖掘方法和技术来解决重要的商业问题。每个技术主题都通过案例研究和源自作者经验的真实案例进行说明，每章都包含针对从业者的宝贵提示。



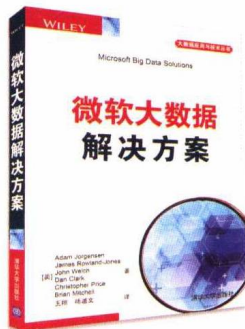
数据仓库和商业智能自诞生之日起就是以提供企业运营所需的数据和决策依据为目的，在过去的传统行业中得到了大力推广和应用，为传统行业的企业信息化发展提供了基础。本书提供了应用Kimball方法论来构建DW/BI系统的清晰指导与实践示例。书中的文章描绘了从最初的业务需求直到最后部署的DW/BI系统开发的整个生命周期。本书可以作为DW/BI系统开发过程中将出现的数以百计问题和情况的深度参考。



随着The Data Warehouse Toolkit(1996)第1版的出版发行，Ralph Kimball为整个行业引入了维度建模技术。从此，维度建模成为一种被广泛接受的表达数据仓库和商业智能(DW/BI)系统中数据的方法。该经典书籍被认为是维度建模技术、模式和最佳实践的权威资源。本书汇集了迄今为止最全面的维度建模技术。本书采用新的思路和最佳实践对上一版本进行了全面修订，给出了设计维度模型的全面指南，既适合数据仓库新手，也适合经验丰富的专业人员。



大多数企业都充斥着数据。为使数据有助于业务开展，需要一个简单的、具有成本效益的工具——最好是已经有所了解的一个工具。Excel就是这样工具。本书中的每个示例都具有分步指导、可下载的Excel文件以及大量截图。要提高自己的市场营销分析技术，这本指南和Excel是必不可少的。



Microsoft强大的大数据平台Windows Azure HDInsight和Hortonworks Data Platform for Windows改变了企业处理、存储和管理企业数据的方式。Microsoft的大数据解决方案套件被设计用于与公司现有的数据基础设施、SQL Server和Hadoop等产品进行无缝集成，使客户不必中断工作流或关键流程即可实施这些方案。如果需要设计和实施全面部署的大数据解决方案的详尽规划蓝图，本书无疑是最佳选择。

译者序

数据仓库为用户提供了用于决策支持的当前和历史数据，这些数据在传统的操作型数据库中很难或无法得到。数据仓库技术是为了有效地把操作型数据集成到统一的环境中以提供决策型数据访问的各种技术和模块的总称。所做的一切都是为了让用户更快、更方便地查询所需要的信息并且提供决策支持。而商业智能是一个广泛意义上的概念，它包括一系列技术，数据仓库是其重要的基石之一，另外一个基石技术就是 ETL(extract, transform, and load; 提取、转换和加载)。

数据仓库和商业智能自诞生之日起就是以提供企业运营所需的数据和决策依据为目的的，在过去的传统行业中得到了大力推广和应用，为传统行业的企业信息化发展提供了基础。不过随着互联网技术、物联网技术以及移动互联网技术和人工智能技术的兴起，数据仓库和商业智能的含义正不断扩大，越来越多的新兴概念和技术被纳入其中，以便能够满足现代互联网企业的运营需求。

Ralph Kimball 和 Kimball Group 的成员都是数据仓库和商业智能领域的权威专家。Ralph 作为一名最早的数据仓库架构师，他长期以来一直坚信数据仓库的设计必须遵循易于理解和快速反应的准则。他创立的维度建模方法论(或者称为 Kimball 方法论)已经成为决策支持领域的金科玉律。

本书的内容都来自过去 20 多年间 Kimball Group 根据自身丰富的行业实践经验所总结的指导性文章，涵盖了数据仓库和商业智能所涉及的方方面面的主题。同时，依据当前的技术和市场现实，加入了新的概念和技术实现方面的内容，使这些极具历史意义和实践指导价值的文章焕发出新的生机。目前，Kimball Group 的所有成员都已退休，但其思想方法以及对 DW/BI(data warehouse/business intelligence)领域所做的贡献将被人们所铭记。相信行业中的后来者会继承其衣钵并且让 DW/BI 领域得到持续发展。

本书提供了应用 Kimball 方法论构建 DW/BI 系统的清晰指导与实践示例。书中的文章描绘出从最初的业务需求收集直到最后部署的 DW/BI 系统开发的整个生命周期。这一系列文章可以作为 DW/BI 系统开发过程中出现的数以百计的问题和情况的深度参考。希望读者在阅读过程中重点思考 Kimball 方法论的原理，结合自己的环境进行适应性调整，而不是原样照搬书中描述的技术和解决方案。毕竟，对于 DW/BI 系统来说，并不存在放之四海而皆准的准绳。这也是本书作者多次强调的。

在此要特别感谢清华大学出版社的编辑们，在本书翻译过程中他们提供了颇有助益的帮助，没有其热情付出，本书将难以付梓。

本书全部章节由蒲成翻译，参与翻译的还有何东武、李鹏、李文强、林超、刘洋洋、茆永锋、潘丽臣、王滨、陈世佳、申成龙、王佳、赵栋、潘勇、负书谦、杨达辉、赵永兰、郑斌、杨晔。

由于译者水平有限,难免会出现一些错误或翻译不准确的地方,欢迎读者指出并勘正,译者将不胜感激。

译者

作者简介

Ralph Kimball 创立了 Kimball Group。自 20 世纪 80 年代中期开始，他就一直是 DW/BI 行业关于维度化方法的思想领袖，并且已经培训了超过 20 000 名 IT 专家。在任职于 Metaphor 和创立 Red Brick Systems 之前，Ralph 在施乐帕克研究中心(Xerox PARC)参与创建了 Star 工作站。Ralph 拥有斯坦福大学电子工程专业的博士学位。

Margy Ross 是 Kimball Group 和 Decision Works Consulting 的董事长。她从 1982 年开始就专注于数据仓库和商业智能。截止现在，Margy 已经为数百个客户提供过咨询服务，并且向数万人讲解过 DW/BI 的最佳实践。在任职于 Metaphor 和联合创办 DecisionWorks Consulting 之前，她毕业于美国西北大学，并且获得了工业工程专业的学士学位。

本书来源和目标

本书的主要作者正在交付培训课程以便为企业管理提供知识更新。该课程是面向管理者的。书中的文章提供了 DW/BI 系统开发所需的基本知识，从最初的需求分析到实施和运营。我们期望，这一系列文章会作为 DW/BI 系统开发过程中出现的难以言喻的经验教训具有深度的权威性。

这些文章的经验涵盖了对于高级技术管理者的关注。在所有情况下，这些文章的基本原理是提供咨询。在过去的 20 年中，这些文章在 Kimball Group 网站上被广泛地分享，并被我们认为是有益的。通过组织这些文章并以系统的方式表达它们可以确保保持一贯性，并为本书增加了重要价值。

内容预览

本书的内容组织对于《数据仓库与数据集市》(第 1 卷)和第 2 卷(第 2 卷)的读者是有益的。

致 谢

首先，我们要感谢 *Kimball Design Tip* 的 33 000 个订阅者，以及无数访问过 Kimball Group 网站以研读我们档案资料的人。这本书将重新修订过的 *Design Tip* 和文章放在了一起，我们希望这是一种非常有用的形式。

如果没有我们业务合作伙伴的协助，《数据仓库与商业智能宝典(第 2 版) 成功设计、部署和维护 DW/BI 系统》一书就不可能成型。Kimball Group 成员 Bob Becker、Joy Mundy 和 Warren Thornthwaite 编写过许多有价值的文章和 *Design Tip*，这些都收录在本书中。感谢 Julie Kimball 极具见解的评论。还要感谢前 Kimball Group 成员 Bill Schmarzo 对于分析应用程序的贡献。

感谢我们的客户和学生，他们与我们一起拥抱、实践并且验证了 Kimball 方法论。我们都从对方身上学到了很多！

Jim Minatel，我们在 Wiley 出版社的执行编辑、项目编辑 Tom Dinse 以及 Wiley 团队的其他人都用其技能、鼓励和热情为这个项目提供了支持。与他们共事非常愉快。

我们要感谢我们的家庭成员，谢谢你们在过去 20 年中支持我们写作这些 *Design Tip* 和文章。Julie Kimball 和 Scott Ross：如果没有你们，我们无法完成本书！当然，还要感谢我们的孩子，Sara Kimball Smith、Brian Kimball 和 Katie Ross，他们在这 20 年中已经长大成人了！

前言

Kimball Group 的文章和 *Design Tip* 归档文件一直都是我们网站(www.kimballgroup.com)上浏览量最大的。回顾 20 年前 Ralph 最初于 1995 年发表的 *DBMS* 杂志文章, 这些归档文件探究了超过 250 个主题, 有时比我们的书籍或课程探究的程度还要深。

在《数据仓库与商业智能宝典(第 2 版)成功设计、部署和维护 DW/BI 系统》中, 我们以连贯的方式组织了所有这些文章。不过本书并不仅仅是过去的杂志文章和 *Design Tip* 一字不变的集合。我们已经精简了多余的内容, 确保所有文章都以一致词汇来编写, 并且更新了许多图片。本书中的文章都进行了重新编辑和改进。

经过慎重的讨论之后, 我们决定更新整本书中的时间基准以及内容, 以便提供 2015 的视角, 而不是将旧日期或者过时的概念留在这些文章中。因此, 2007 年所写的文章可能会在一个示例中使用 2015 年这一时间! 当出现涉及过去多少年的文章时, 我们已经将这些时间基准更新为相对于 2015 年而言。例如, 如果一篇 2005 年的文章最初描述“在过去 5 年之间”, 那么该文章现在就会描述为“在过去 15 年之间”。在提及关于我们多年来的经验、售出的图书量、编写的文章或者教过的学生时, 这些也都被更新为 2015 年的描述。最后, 我们偶尔会将像“调制解调器”这样过时技术的引用变更为更加现代的技术, 尤其是“互联网”。我们相信, 这些变更不会造成误导或引起混淆, 相反, 它们会让本书的阅读体验变得更好。

本书读者对象和目标

本书的主要读者是正在交付数据仓库以便为商业智能提供支持的分析师、设计者、建模者或者管理者。本书中的文章描绘了 DW/BI 系统开发的整个生命周期, 从最初的业务需求收集直到最后的部署。我们相信, 这一系列文章会充当 DW/BI 系统开发过程中出现的数以百计的问题和情况具有深度的极佳参考。

这些文章的范围涵盖了对于高度技术性重点的关注, 在所有情况下, 这些文章的基调都力求提供咨询。在过去的 20 年中, 这些文章在 Kimball Group 网站上每天都会被访问数千次, 因此我们确信它们是有用的。通过组织这些归档文件以及系统地编辑这些文章以便确保其一致性和相关性, 为本书增加了重要价值。

内容预览

本书的内容组织对于《数据仓库生命周期工具箱(第二版)》(Wiley 出版社于 2008 年出版)

的读者来说看起来会有些熟悉,因为我们是根据对应于数据仓库/商业智能(DW/BI)实现的主要里程碑的主题来组织这些文章的。鉴于“Kimball”一词差不多就是维度建模的代名词,所以不足为奇的是,本书的大部分内容都会专注于该主题。

- **第1章:读本概览。**我们会以 Ralph 几年前为 *DM Review* 杂志所写的一系列文章作为本书的开始。这一系列文章以连贯内聚的方式简洁地封装了 Kimball 方法,因此这些文章为本书提供了绝佳的概述,类似于克利夫笔记。
- **第2章:深入研究之前。**Ralph 文章的长期读者会发现,该章充满了对于旧日的记忆,因为这些文章中的许多内容都具有重要历史意义。有些让人惊讶的是,尽管这些文章中的大多数都写于 20 世纪 90 年代,但其内容仍然具有相关性。
- **第3章:项目/程序规划。**在了解了概述和历史的经验教训之后,第 3 章会继续推进从而启动 DW/BI 程序和项目。我们会思考项目团队和主办干系人的职责,然后深入研究 Kimball 生命周期方法。
- **第4章:需求定义。**要在缺乏业务需求的情况下取得 DW/BI 的成功是很难的。该章将为有效获得业务需求提供具体的建议。它强调了围绕业务过程来组织需求调研结果的重要性,并且提供了就恰当的后续步骤达成组织共识的策略建议。
- **第5章:数据架构。**切实理解了业务需求之后,我们会将注意力转向数据(第 11 章同样也会继续关注数据)。该章首先会论证维度建模的正确性。然后会描述企业数据仓库总线架构、探讨敏捷开发方法以便支持数据仓库,为必不可少的集成和管理工作提供合理化机制,然后将 Kimball 架构与企业信息工厂的中枢辐射模型作对比。
- **第6章:维度建模基础。**该章将介绍维度建模的基础知识,首先探讨事实与维度的区别,以及在数据仓库中向下钻取、横向钻取和处理时间的核心活动。我们还会探究大家熟悉的关于维度模型的传言。
- **第7章:维度建模任务和职责。**第 6 章涵盖了围绕维度建模的根本性“内容和原因”,专注于“如何、谁和何时”。第 7 章描述了维度建模过程和任务,以便组织起一支有效的团队,不管是全新开始还是重新考虑已有模型。
- **第8章:事实表核心概念。**第 8 章的主题可以被描述为“仅仅着眼于事实”。我们首先会探讨事实表的粒度性和三种基本类型,然后将我们的注意力转向事实表键和退化维度。该章将以常用的事实表模式集锦作为结尾,其中包括空值、文本和稀疏填充的指标,以及非常类似于维度属性的事实。
- **第9章:维度表核心概念。**在第 9 章中我们会将注意力转向维度表,从对代理键和无处不在的时间(或日期)维度的探讨开始。然后将探究角色扮演、杂项和因果性维度模式,随后将探讨对于渐变维度的全面处理,其中包括四种新的高级维度类型。
- **第10章:更多的维度模式和注意事项。**第 10 章将用更丰富的维度表范围来补充前一章的内容。我们会描述星型模式和外支架,以及一节关于桥接的被大幅更新过的内容,以便应对多值维度属性和不规则的多变层次结构。我们会探讨顾客维度中经常会遇到的细微差别以及国际化问题。该章会以一系列案例研究作为结束,其中涵盖了保险、航程和网络、人力资源、金融、电子商务、文本搜索以及零售。我们鼓励每个人都仔细研读这些描述,因为这些模式和推荐做法超出了行业或应用程序边界。
- **第11章:后台 ETL 和数据质量。**在第 11 章中,我们要将话题切换到设计目标维度模型以便填充它。预先警告:这是篇幅很长的一章,正如根据该主题你可以预见到的一

样。本书在该章中提供了大量的新素材。我们首先会描述提取、转换和加载(ETL)数据所需的 34 个子系统, 以及使用商业化 ETL 工具的优缺点。基于此, 我们会深入研究数据质量的注意事项, 为构建事实表和维度表提供具体指导, 并且探讨实时 ETL 的影响。

- **第 12 章: 技术架构注意事项。**直到第 12 章, 我们才开始探讨围绕技术架构的问题, 首先会介绍面向服务架构(server oriented architecture, SOA)、主数据管理(master data management, MDM)以及打包分析。关于大数据的新一节内容刊载了 Ralph 编写的两份重要白皮书。该章的最后几节内容会专注于展示服务器, 其中包括聚合导航和在线分析处理(online analytical processing, OLAP)、用户界面设计、元数据、基础设施和安全性的作用。
- **第 13 章: 前台商业智能应用程序。**在第 13 章中, 我们开始介绍 DW/BI 系统的前台, 其中业务用户会与数据进行交互。我们描述了一个典型业务分析的生命周期, 从历史绩效的审查开始, 但并不会止步于此。之后我们会将注意力转向标准化 BI 报告, 然后深入探究数据挖掘和预测式分析。该章会以探究用于业务分析的 SQL 限制作为结束。
- **第 14 章: 维护和发展的注意事项。**在倒数第 2 章中, 我们为成功部署 DW/BI 系统以及保持其健康以便持续取得成功提供了建议。
- **第 15 章: 最后的思考。**该章总结了来自每个 Kimball Group 原则的关于数据仓库和商业智能的最终观点。这些见解涵盖了我们已经获得的最重要的来之不易的经验教训, 以及所能预见到的一些未来数据仓库可能具有的特性。

导航标识

鉴于《数据仓库与商业智能宝典(第 2 版) 成功设计、部署和维护 DW/BI 系统》中文章的广度和深度, 我们非常审慎地找出了 20 多篇文章作为“Kimball 经典”, 因为它们涵盖了非常有效的概念, 我们和行业中的许多人在过去 20 年中反复地引用了这些文章。这些经典的文章使用如下特殊图标作为区分:



我们期望大多数人以某种随机的顺序阅读这些文章, 而不是从前到后地阅读本书。因此, 我们特别强调此读本的索引, 因为我们期望许多读者会通过搜索特定技术或建模情形的索引来深入进行探究。

术语说明

引以为荣的是, Ralph 确立的词汇表如此经久不衰并且被广泛采用, 包括维度、事实、渐变维度、代理键、事实表粒度、非事实型事实表以及退化维度在内的 Kimball “特征词”, 在整个行业中已经持续使用 20 多年了。不过尽管我们充满了善意, 但一些术语自其被引入以来已经

发生了变化。我们已经以追溯的方式使用被广泛采用的当前术语替换了老的术语。

- 人造键现在被称为代理键。
- 数据集市已经被替换成业务过程维度模型、业务过程主题领域或者主题领域，这取决于上下文。
- 数据暂存现在被称为提取、转换和加载。
- 终端用户应用程序已经被商业智能应用程序所替代。
- 帮助表现在被称为桥接表。

由于大多数人都不会从头到尾一页不漏地阅读本书，因此我们需要事先介绍一些常用的缩写词：

- DW/BI 是端到端数据仓库/商业智能系统的英文缩写。这个缩写对于简约性来说很有用，不过它也明确地将数据仓库和商业智能链接为一个共存体。最后，它反映出从数据仓库本身作为终点到商业智能(BI)的重心转换推动我们所做的一切事情。毕竟，数据仓库是所有形式 BI 的平台。
- 本书中的许多图片都包含 DD(degenerate dimension)、FK(foreign key)和 PK(primary key)缩写，它们分别代表退化维度、外键和主键。
- ETL 的意思是提取、转换和加载，这是获取数据并且让数据准备好暴露给 BI 工具的标准范式。
- ER(entity-relationship)指的是实体关系。我们会在探讨第三范式(3NF)或者与维度数据模型相反的数据模型时频繁使用 ER。
- OLAP 代表在线分析处理，通常用于将在多维数据库或多维数据集中捕获的维度模型与被称为星型模式的关系型 DBMS 中的维度模型区分开来。这些关系型星型模式有时也被称为 ROLAP。
- SCD(slowly changing dimension)是渐变维度的缩写，指的是所确立的用于处理维度属性变更的技术。

目 录

第 1 章 读本概览	1
1.1 抑制住立即开始编码的冲动	1
1.2 设置边界	3
1.3 数据争夺	5
1.4 流言终结者	7
1.5 划分数据世界	9
1.6 集成式企业数据仓库的 必要步骤	10
1.6.1 集成式 EDW 会交付什么	11
1.6.2 集成的终极试金石	11
1.6.3 组织挑战	12
1.6.4 一致化维度和事实	12
1.6.5 使用总线矩阵与管理层交流	12
1.6.6 管理集成式 EDW 的主干	13
1.6.7 维度管理器	14
1.6.8 事实提供者	15
1.6.9 配置商业智能(BI)工具	16
1.6.10 连带责任	17
1.7 钻取以寻求原因	17
1.8 渐变维度	19
1.8.1 渐变维度的三种原生类型	20
1.8.2 高级渐变维度	22
1.9 通过维度评价 BI 工具	22
1.10 事实表	24
1.10.1 忠实于粒度	24
1.10.2 从最低的可能粒度进行 构建	25
1.10.3 三类事实表	25
1.11 开发利用事实表	26
1.11.1 前端: 聚合导航	26
1.11.2 前端: 钻取不同的粒度	26

1.11.3 前端: 将约束暴露给不同的 业务过程	26
1.11.4 后端: 事实表代理键	27
第 2 章 深入研究之前	29
2.1 Ralph Kimball 和施乐帕克 研究中心(Xerox PARC)	29
2.2 数据库市场分化	31
2.3 提出超市概念(Kimball 经典)	33
2.3.1 危机规划	33
2.3.2 具有架构的数据集市	34
2.3.3 一致化维度的重要性	34
2.3.4 设计一致化维度	35
2.3.5 做出承诺	36
2.3.6 允许的一致化维度变体	36
2.3.7 建立标准事实定义	36
2.3.8 粒度的重要性	37
2.3.9 更高级别的数据集市	38
2.3.10 解决烟囱问题	38
2.3.11 不需要一致化维度的情形	38
2.3.12 清晰视角	39
2.4 数据仓库的全新需求	39
2.5 应对全新需求	42
2.5.1 数据集市和维度建模	42
2.5.2 将数据集市插入数据仓库 总线架构中	44
2.6 挑起事端	46
2.7 设计约束和不可避免的现实	49
2.7.1 设计约束	49
2.7.2 不可避免的现实	50
2.7.3 摆脱困境	51

2.8 两个强有力的观点	52	3.9 包含 DW/BI 范围蔓延并且 避免范围冒用	77
2.8.1 分离系统	52	3.10 IT 过程对于 DW/BI 项目 是否有益	79
2.8.2 对称的星型结构和多维 数据集	53	3.10.1 规范	79
2.8.3 巨大的回报	54	3.10.2 命名规范	79
2.8.4 我们已经取得了什么成果	54	3.10.3 教条主义	80
2.9 数据仓库就餐体验 (Kimball 经典)	54	3.11 有效主办者的行为	80
2.9.1 厨房	55	3.11.1 为成功做准备	81
2.9.2 就餐区	56	3.11.2 抵制阻力最小的路径	81
2.10 用于更艰难问题的更简单 方法	57	3.11.3 团结周边可用资源	82
2.10.1 增量集成	57	3.11.4 耐心是一种美德	82
2.10.2 递增的数据质量	58	3.11.5 保持对目标的专注	83
2.11 扩展数据仓库的边界	58	3.12 从终端用户开始计算的总体 拥有成本(Kimball 经典)	83
第3章 项目/程序规划	61	3.12.1 不好的决策也是成本	83
3.1 专家边界	61	3.12.2 仔细查看这些成本	84
3.2 工程师的观点	63	3.13 简要概括 Kimball 生命周期	87
3.2.1 数据仓库使命	64	3.13.1 程序/项目规划和管理	88
3.2.2 设计驱动	65	3.13.2 业务需求	88
3.2.3 设计约束	65	3.13.3 技术轨迹	88
3.2.4 工程师的回应	66	3.13.4 数据轨迹	88
3.3 当心异议消除者	66	3.13.5 商业智能轨迹	89
3.4 中央团队要做些什么	69	3.13.6 部署、维护和发展	89
3.4.1 定义和发布共享维度	70	3.14 挺身而出	89
3.4.2 提供跨部门的应用程序	71	3.15 持相反意见的架构师	90
3.4.3 定义一致化数据仓库安全性 架构	72	3.16 在应用最佳实践时慎重思考	93
3.5 避免隔离的 DW 和 BI 团队	72	3.16.1 采取一种企业方法	93
3.6 BI 和数据仓库专家可用的、 更好的业务技能	73	3.16.2 拥抱商业智能	93
3.6.1 建立对业务的理解	73	3.16.3 设计维度模式	93
3.6.2 建立人际交往能力	73	3.16.4 将一致化维度用于集成	94
3.6.3 掌握公开演讲技巧	74	3.16.5 仔细规划 ETL 架构	94
3.6.4 掌握书面沟通技巧	74	3.17 低风险企业数据仓库的 八个准则	95
3.6.5 实践决定一切	75	3.17.1 做正确的事情	95
3.7 有风险的项目资源就是 有风险的业务	75	3.17.2 赋予业务用户控制权	96
3.8 无法实现分析	76	3.17.3 渐进式处理	96
		3.17.4 从轻量级、专注的治理 开始	96

3.17.5 构建一个简单、通用的平台.....	97	5.2 一个维度建模宣言(Kimball 经典).....	122
3.17.6 使用一致化维度来集成.....	97	5.2.1 什么是 3NF 标准化建模.....	122
3.17.7 每次都有一些过滤来管理质量.....	97	5.2.2 什么是 DM.....	124
3.17.8 自始至终使用代理键.....	97	5.2.3 DM 与 3NF 的对比.....	125
第 4 章 需求定义	99	5.2.4 DM 的优势.....	126
4.1 将 Alan Alda 的访问技巧用于揭示业务需求(Kimball 经典).....	99	5.2.5 对 DM 的误解.....	127
4.1.1 保持好奇心, 但不要自作聪明.....	100	5.2.6 捍卫 DM.....	128
4.1.2 要口语式对话.....	100	5.3 没有百分百的保证	128
4.1.3 倾听并且期望被改变.....	101	5.3.1 3NF 建模是否会处理业务规则吗.....	129
4.2 业务需求收集的更多注意事项.....	102	5.3.2 早期维度建模.....	130
4.3 平衡需求与现实(Kimball 经典).....	104	5.4 分而治之	131
4.4 在收集业务需求时克服障碍.....	105	5.4.1 是否需要通用标签.....	132
4.5 令人吃惊的数据剖析价值.....	106	5.4.2 业务过程主题领域并不是部门化的.....	132
4.6 专注于业务过程, 而非业务部门.....	108	5.4.3 一致化维度和事实.....	132
4.7 识别业务过程.....	109	5.4.4 数据仓库总线架构.....	133
4.8 业务过程全面揭秘.....	110	5.4.5 是否仅仅为了高度分布式系统.....	133
4.9 战略业务举措和业务过程之间的关系.....	111	5.4.6 净收益.....	134
4.10 自下而上属于用词不当.....	112	5.5 矩阵(Kimball 经典)	134
4.10.1 专注于企业, 而非部门.....	112	5.5.1 邀请主题领域小组参加一致化会议.....	136
4.10.2 起草企业数据仓库总线矩阵.....	112	5.5.2 与老板进行沟通.....	136
4.10.3 进行优先级排序以便得到一份有序的结论.....	113	5.5.3 二级主题领域.....	136
4.10.4 绘制企业路线图.....	115	5.6 再次探讨矩阵(Kimball 经典)	137
4.11 (超越数据建模的)维度化思维.....	115	5.6.1 用于引用数据的矩阵列.....	137
4.12 使用维度模型验证业务需求.....	116	5.6.2 数据管理.....	137
第 5 章 数据架构	119	5.6.3 以过程为中心的行.....	138
5.1 ER 建模是否对 DSS 有害(Kimball 经典).....	119	5.6.4 关联列和行.....	138
		5.6.5 常见的矩阵不幸事件.....	139
		5.6.6 矩阵扩展.....	139
		5.7 向下钻取到详细的总线矩阵中	140
		5.8 关于敏捷方法论	142
		5.9 敏捷企业数据仓库是不是一个矛盾混合体	143
		5.10 采用敏捷方法? 先要从总线矩阵开始	144

5.11 作为敏捷数据仓库基础的	
一致化维度	145
5.12 为现实中的人而集成	146
5.12.1 定义集成	146
5.12.2 集成标签	147
5.12.3 集成测量	147
5.12.4 维度管理者的职责	148
5.12.5 事实提供者的职责	148
5.13 为企业维度构建即时可用的	
资源	149
5.14 数据管理基础知识: 质量和	
一致性的第一步	150
5.14.1 为何管理是必要的	150
5.14.2 管理职责	151
5.14.3 管理的正确举措	152
5.14.4 沟通工具和技术	152
5.14.5 如何开始	152
5.15 要不要集中化	153
5.15.1 闪光的未必都是金子	153
5.15.2 不要畏惧伟大	154
5.15.3 结果好意味着一切都好	155
5.16 观点差异(Kimball 经典)	155
5.16.1 共同之处	155
5.16.2 Kimball 总线架构	156
5.16.3 企业信息工厂	157
5.16.4 根本性差异	158
5.16.5 混合方法怎么样	159
5.16.6 成功标准	159
5.17 庸人自扰	160
5.18 不要用一个标准化 EDW	
支持商业智能	160
5.19 使用维度展示区域补充	
3NF EDW	162

第6章 维度建模基础 165

6.1 事实表和维度表	165
6.1.1 测量和上下文	165
6.1.2 维度键	166
6.1.3 把两个建模方法关联起来	167
6.1.4 声明粒度	167

6.1.5 可累加事实	167
6.1.6 退化维度	168
6.2 向下、向上和横向钻取	168
6.2.1 向下钻取	168
6.2.2 向上钻取	170
6.2.3 横向钻取	170
6.3 数据仓库的灵魂之第一部分:	
向下钻取	171
6.4 数据仓库的灵魂之第二部分:	
横向钻取	173
6.4.1 实现横向钻取	174
6.4.2 令人惊讶的神奇之处	175
6.5 数据仓库的灵魂之第三部分:	
时间处理	175
6.5.1 时间的有效性	176
6.5.2 正确关联	176
6.5.3 自然粒度	177
6.5.4 是否兑现了承诺	178
6.6 优雅修改已有的事实和	
维度表	178
6.7 Kimball 关于维度建模的十项	
必要规则(Kimball 经典)	179
6.8 不该做的事情	181
6.9 危险的先入为主的想法	183
6.10 虚言和事实	185
6.10.1 并非所有的维度模型都是	
同等创建的	185
6.10.2 专注于测量过程, 而	
非部门报告	185
6.10.3 从原子详情开始, 而	
非汇总数据	186
6.10.4 目标是集成, 而非标准化	187

第7章 维度建模任务和职责 189

7.1 让用户安然入睡	189
7.2 用于设计维度模型的实践	
步骤	194
7.2.1 参与其中	195
7.2.2 深究细节	195
7.2.3 审核结果	196