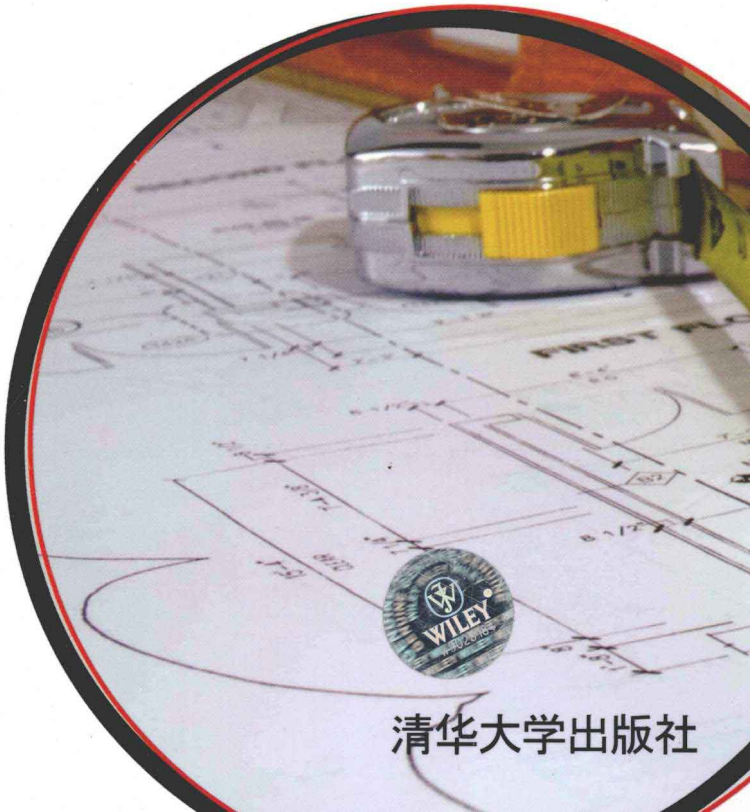
WILEY

The Microsoft Data Warehouse Toolkit: With SQL Server 2008 R2 and the Microsoft Business Intelligence Toolset, Second Edition

Microsoft 数据仓库工具箱(第2版)

——使用SQL Server 2008 R2和Microsoft BI工具集



[美] Joy Mundy
Warren Thornthwaite 著
Ralph Kimball
包 战 孔祥亮 译

清华大学出版社

Microsoft 数据仓库工具箱(第 2 版)

——使用 SQL Server 2008 R2 和 Microsoft BI 工具集

Joy Mundy
[美] Warren Thornthwaite 著
Ralph Kimball
包 战 孔祥亮 译

清华大学出版社
北 京

Joy Mundy, Warren Thornthwaite, Ralph Kimball
The Microsoft Data Warehouse Toolkit: With SQL Server 2008 R2 and the Microsoft Business Intelligence Toolset, Second Edition
EISBN：978-0-470-64038-8

图书在版编目(CIP)数据

Microsoft 数据仓库工具箱(第 2 版)：使用 SQL Server 2008 R2 和 Microsoft BI 工具集/ (美) 蒙迪(Mundy,J.)，(美) 桑思韦特(Thornthwaite,W.)，(美) 金博尔(Kimball,R.) 著；包战，孔祥亮 译.
—北京：清华大学出版社，2012.5

书名原文：The Microsoft Data Warehouse Toolkit: With SQL Server 2008 R2 and the Microsoft Business Intelligence Toolset, Second Edition
ISBN 978-7-302-28336-2

Ⅰ. ①M…　Ⅱ. ①蒙…　②桑…　③金…　④包…　⑤孔…　Ⅲ. ①关系数据库—数据库管理系统，SQL Server 2008　②数据库系统—软件工具　Ⅳ. ①TP311.138　②TP311.56

中国版本图书馆 CIP 数据核字(2012)第 044926 号

责任编辑：王　军　张立浩
装帧设计：牛艳敏
责任校对：成凤进
责任印制：张雪娇

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：三河市金元印装有限公司
经　　销：全国新华书店
开　　本：185mm×260mm　　**印　　张**：30.25　　**字　　数**：736 千字
版　　次：2012 年 5 月第 2 版　　**印　　次**：2012 年 5 月第1 次印刷
印　　数：1～3000
定　　价：78.00 元

产品编号：042869-01

作 者 简 介

Joy Mundy 在斯坦福大学、WebTV 和 Microsoft SQL Server 产品研发小组中一直关注 DW/BI 系统。Joy 在塔夫茨大学获得经济学学士学位，然后在斯坦福大学获得工程经济系统硕士学位。

Warren Thornthwaite 自 1980 年起就开始了 DW/BI 生涯。在离开 Metaphor 咨询公司后，他为斯坦福大学和 WebTV 工作。Warren 从密西根州立大学获得传媒学的学士学位，从宾夕法尼亚州的沃顿商学院获得决策学的 MBA 学位。

Ralph Kimball 是 Kimball Group 的创立者，自从 20 世纪 80 年代中期开始，他就是 DW/BI 行业中维度方法的思想领袖，培训了 10 000 多名 IT 专业人员。他曾在 Metaphor 工作过，创立了 Red Brick Systems，之后在 Xerox 的 Palo Alto Research Center(PARC)工作时还与其他人一起创立了星型工作站。Ralph 获得了斯坦福大学的电气工程博士学位。

致　谢

首先，我们要感谢阅读 Kimball Group 工具集图书的读者、参加培训的学员、在咨询项目中雇用我们的客户。我们总是能从这些人身上学到知识，他们对我们的思维和商业智能行业有深远的影响。

没有 SQL Server 产品研发小组中许多人的帮助，本书就不可能面世。Dave Wickert 审阅了 PowerPivot 和 SharePoint 相关章节，提供了许多宝贵的内容改进建议。Bryan Smith 审阅了 Integration Services 和 Analysis Services 相关章节，这些章节因为获得了 Bryan 的帮助而变得更好。Carolyn Chau 审阅了 Reporting Services 相关章节，Eric Hanson 审阅了关系数据库相关章节，Pej Javaheri 审阅了 SharePoint 相关章节，Raman Iyer 审阅了数据挖掘相关章节，在此对他们表示诚挚的感谢。

SQL Server 小组的其他成员在审阅本书的 SQL Server 2005 版时也提供了重要的帮助，很惭愧直到本书的第 2 版才感谢他们：Bill Baker、Stuart Ozer、Grant Dickinson、Donald Farmer、Siva Harinath、Jamie MacLennan、John Miller、Ashvini Sharma、Stephen Quinn 和 Rob Zare。

Kimball Group 的同事也非常重要。我们在编写本书时，他们一直在鼓励我们，并审阅书稿，帮助我们给书稿润色。当然，Ralph Kimball 对本书的影响最大，他不仅参与了本书的编写，加入了他在商业智能领域方面的思考，还直接帮助我们改进了本书的整体结构和流程。

Sara Sheaer、Ginny Munroe 和 Bob Elliott 是 Wiley 出版社的编辑，他们也提供了很大的帮助和支持，很高兴能与他们合作。

感谢和我们一起生活的人，感谢他们总是适时出现在我们身边，感谢他们给我们提供了必要的时间，并提醒我们休息一会儿。Tony Navarrete 和 Elizabeth Wright，本书没有你们是不可能面世的。

序

在本书第 1 版出版以来的 5 年中，Microsoft 在构建数据仓储和商业智能工具集方面取得了长足的进步。工作在这个领域的人员会高兴地看到 Microsoft 一直承诺提供有用的、高质量的专业工具。在这 5 年中，Warren 和 Joy 为数十位客户提供咨询服务，教授了大量的课程，回答了数以百计的问题，还在吃午饭时讨论模式，并斟酌了 Microsoft 的 DW/BI 工具集中的每个模块。本书第 2 版保留了独特的观点，并根据精确的评估给出工具能完成的所有任务。本书介绍的是“该做什么”，而不是“如何做”！

Ralph Kimball

前　言

本书描述如何使用 Microsoft SQL Server 产品集设计并构建一个成功的商业智能系统及其底层的数据仓库数据库。

0.1　数据仓库和商业智能系统

数据仓储和商业智能的作用在于，为业务人员提供制定操作性和战略性业务决策所需的信息和工具。我们将详细剖析这方面的问题，以便您能切实了解要采纳的决策的性质和规模。

首先，顾客是指公司的业务人员。但是，并非所有的业务用户都同样重要——您应对制定战略性业务决策的人更感兴趣。一个非常好的业务决策可以给许多公司带来数百万美元的收益。您主要的顾客是公司的主管、经理以及分析师，因此，数据仓库和商业智能(DW/BI)系统影响深远、意义重大。

战略性也意味着重要性。这些决策可以决定公司的成败。因此，DW/BI 系统是一个高风险的尝试。当做出了某个战略性的决策时，总是有人成功，有人失败。因此，DW/BI 系统也是高度政治化的尝试。

DW/BI 系统正日益支持着操作性决策，特别是在决策制定人员需要从多个数据源中查询历史数据或集成数据的地方。许多“分析型应用程序”都支持这一操作。不管制定的决策是战略性的还是操作性的，DW/BI 小组都需要提供必要的信息来制定这些决策。

任何给定的决策都需要一个独特的信息子集，但这个子集一般不能预先确定。需要构建一个信息基础结构，以集成公司内部和外部的数据，然后清理、排列和重构这些数据，使它们尽可能灵活、有效。尽管大部分事务系统模块处理一种类型的数据，例如收到的账单、订单或账目，但 DW/BI 系统最终必须将它们集成在一起。因此，DW/BI 系统要求进行技术上很复杂的数据收集和管理。

最后，需要给业务决策制定人员提供使用这些数据的工具。在这样的情形下，工具并不仅仅意味着软件，还意味着业务用户需要了解哪些信息是可用的，查找需要的子集，并将数据结构化，以阐明潜在的业务动态。因此，工具意味着培训、文档、支持，以及即席查询工具、报表和分析型应用程序。

DW/BI 系统：

- 意义重大，影响深远。
- 高风险。
- 高度战略性。
- 需要专业技术，收集和管理复杂的数据。
- 需要频繁地访问、培训和支持用户。

创建和管理 DW/BI 系统是一项极具挑战性的任务，希望您以自己掌握的全部知识来接受这一任务。以我们的经验来看，如果您事先有所了解，那么将更容易应对这些挑战。

我们并不想使您气馁，而是想在您跳进深水之前警告您。使数据仓库具有挑战性的所有因素也是该项目很有趣且令人兴奋的原因。

0.1.1 Kimball Group

尽管构建和管理成功的 DW/BI 系统是具有挑战性的，但也有一些增加成功可能性的方法，即 Kimball Group 提供的方法。我们已经在 DW/BI 领域工作了超过 25 年。本书的作者也是 Kimball Group 的成员，一直从事着数据仓储和商业智能系统的工作，是该系统的厂商、顾问、实现者和用户。我们的格言是“Practical techniques—proven results(通过实践技术来证明结果)”，我们共同的动力是指明构建和管理成功的 DW/BI 系统的最佳方式。我们也是热心的老师，衷心希望您获得成功，避免我们和其他人犯过的错误。

0.1.2 本书目标

数据仓储和商业智能至少自 20 世纪 70 年代以来就有相同的形式，且有着很长的技术生命周期。1995 年，我们的主要作者成立了第一家顾问公司，其中一位作者认为数据仓储不会继续发展，这个浪潮已经开始回落，而我们幸运地获得了更多的项目，不必再去找工作。多年后，数据仓储和商业智能依然很强大。

随着 DW/BI 行业的成熟，它逐渐被单源提供商垄断——对不愿冒风险的公司来说这是一种安全的选择。DW/BI 技术涵盖了所有方面，包括深奥的源系统知识、用户界面设计以及具有最佳实践的 BI 应用程序。数据库销售商的最佳定位是提供端到端解决方案。自从推出 SQL Server 2000、尤其是 SQL Server 2005 以来，Microsoft 就以诱人的价格将自己变成一个可行的、单源数据仓库系统提供商。

本书是《数据仓库工具集——面向 Microsoft 和 SQL Server 2005 商业智能工具集》的修订版。除了添加新功能(如 PowerPivot 和 Master Data Services)之外，新版本还把以前的建议替换为我们近年来用 Microsoft 工具建立 DW/BI 系统的所有心得。本书基于 SQL Server 2008 R2 版，但大多数建议对 SQL Server 2008 也有效。适用于 SQL Server 2008 R2 的技术或建议会在文中标识出来。

0.1.3 本书读者对象

本书覆盖了 DW/BI 系统的整个生命周期，因而可以给 DW/BI 团队的每个成员提供有

用的指导，项目经理、业务分析师、数据建模人员、ETL 开发人员、DBA，BI 应用程序开发人员甚至业务用户都可以从本书中受益。我们相信本书对从事 Microsoft SQL Server DW/BI 项目的任何人都非常有价值。

本书的主要读者是在 Microsoft SQL Server 平台上启动项目的新 DW/BI 团队，读者不需要有构建 DW/BI 系统的经验，但基本了解 Microsoft 世界：操作系统、基础架构组件以及资源，对关系数据库(表、列和简单 SQL)有基本的认识，大致熟悉 SQL Server 关系数据库，但这并不是必备条件。全书将提供其他书和资源的许多参考。

第二个读者群是有 Kimball Method DW/BI 使用经验但却首次接触 Microsoft SQL Server 工具集的读者。对于曾经阅读过其他 *Toolkit* 书籍、实践过我们的方法的读者，我们将指出需要复习哪些章节，不过再次阅读这些材料并没有坏处。

不管您的背景如何，只要从新项目开始，就将从本书中受益匪浅。尽管本书提供使用现有的数据仓库的建议，但在理想情况下，已有的数据仓库或数据集市不会令人满意，至少在新系统部署后，已有的系统不会令人满意。

0.2 Kimball 生命周期

在深入一个项目后，我们都会感到恐慌，因为需要完成的工作量要比一开始想象的多得多。在许多 BI/DW 项目中，您一开始都以为只需要移动一些数据到新计算机中，清理数据，然后建立一些报表。这听起来很不错——只需要 6 周，最多两个月就可以完成。您走入森林后不久就发现，里面比自己想象得更暗、树木更密。实际上，您甚至看不到道路。

避免这种恐慌以及是随之而来的灾难的最好方式是在开始之前就知道要去哪里，路标和方向会指出您必须访问什么地方，前面有哪些危险区域，从而有助于引领您安全地穿过不熟悉的区域。本书就是 Microsoft SQL Server DW/BI 系统项目的路标。它遵循 *The Data Warehouse Lifecycle Toolkit*，第 2 版(Wiley，2008)一书描述的 Kimball 生命周期中的基本流程。本书汇集了我们的经验，详细描述生命周期的步骤、任务和依赖关系。生命周期是一个基于 4 个主要原则的迭代方法：

- **专注于业务**：识别业务需求和及其相关的价值，以便建立牢靠的业务关系，使您的业务感觉更敏锐，提高咨询技能。
- **构建信息基础架构**：设计一个集成的、易使用的、高性能的信息基础架构，旨在满足在企业中找出的各种业务需求。
- **进行有意义的增量发布**：以增量方式构建数据仓库，用 6 到 12 个月的时间来发布。使用清晰可辨的业务价值来决定增量的执行次序。
- **发布整个解决方案**：提供给业务用户传递价值所需的所有元素。这意味着，一个稳固的、设计合理的、经过高质量测试的、可访问的数据仓库仅仅是一个开始。还必须发布即席查询工具、报表设计应用程序及高级分析表、培训、支持、网站和文档。

本书使用 Kimball 生命周期构建 DW/BI 系统，以帮助您遵循这 4 个原则，这 4 个原则已加入到生命周期中。理解 Kimball 生命周期的秘密是：它是基于业务的，采用维度方法

设计展现给最终用户的数据模型，而且它是一个真实的生命周期。

0.2.1 生命周期跟踪和任务区域

DW/BI 系统是复杂的实体，构建这种系统的方法必须有助于简化复杂性。图 0-1 展现了 Kimball 生命周期。13 个方框显示了构建成功的数据仓库的主要任务区域，以及这些任务之间的主要依赖关系。

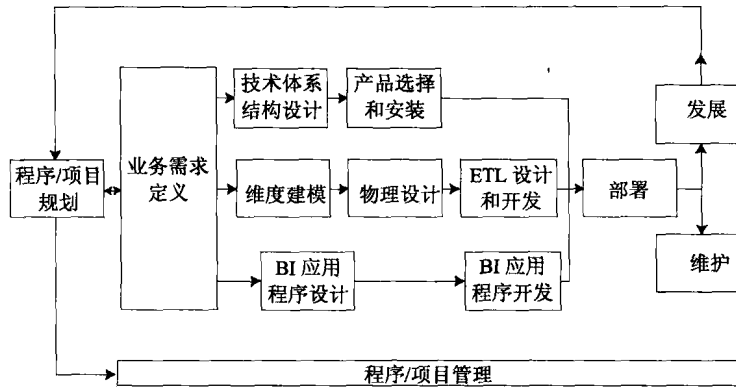


图 0-1 业务维度生命周期

在生命周期这一级可以进行多方观察，首先注意业务需求定义方框的中心角色。业务需求提供了其后的 3 个任务的基础，它们也影响着项目计划，因此箭头也指回项目规划方框。最终，我们经常要根据对业务需求和优先权的深入理解来修改项目规划。

其次，生命周期中间的 3 个任务主要考虑 3 个不同的领域。

- 顶部的任务是关于技术的。这些任务主要确定需要什么功能，并计划使用哪些 Microsoft 技术，以及如何安装和配置它们。
- 中间的任务是关于数据的。数据任务将设计和实例化维度模型，然后开发 ETL(提取、转换和加载)系统来填充它。可以将数据任务视作“构建数据仓库数据库”，但数据仓库直到生命周期的其余任务完成后才会成功。
- 底部的任务是关于商业智能应用程序的。这些任务将为业务用户设计和开发 BI 应用程序。

当部署系统时，这些任务会合并起来，这时要非常小心，因为系统仅有一次机会可以获得初次的好印象。尽管图 0-1 把维护放在部署之后，但在设计系统时就必须有维护它的能力和工具。项目增长阶段有一个箭头指向开始的项目规划，这个简单的箭头有重要的意义。生命周期的增量方法是发布业务价值的一个基本元素。

生命周期的下方是项目管理方框，这里最重要的是需要一个能与高管沟通的领导。在理想情况下，团队领导可以同技术人员和业务人员进行有效沟通，包括公司最高级的行政人员。

0.2.2 关键术语和 Microsoft 工具集

商业智能行业充斥着没有正确使用的或者矛盾的术语。该行业中一些长期存在的争论源自于对术语的误解，就像哲学上的差异一样。记住这一点，即使我们不能解决所有的历史争论，也将尽量保持清晰和一致。这里介绍一些重要的术语。

下面在定义每个术语时也强调了关联的 Microsoft 技术，其中大部分是 SQL Server 的成员。

- 数据仓库是商业智能的平台。在 Kimball 方法中，数据仓库包含了从原始数据提取到用户见到的软件和应用程序等所有内容。其他作者坚持认为，数据仓库仅仅是远离最终用户的、放在屋中的、集中式的、高度规范化的数据存储。 我们不同意这个观点。为了减少混淆，本书始终使用短语“数据仓库/商业智能系统(DW/BI 系统)”来表示整个端到端系统。当专门讨论原子级的用户可查询数据存储时，就称为数据仓库数据库。
- 业务过程维度模型是建模数据的特定准则，也是规范化建模的一个替代品。维度模型包含了与规范化模型一样的信息，但以对称的方式打包数据，其设计目标是用户可理解性、商业智能查询性能和对变化的适应性。规范化模型有时称为第三范式模型，用于支持大数据量的单行插入和更新，以定义事务系统，但一般在可理解性、快速和适应变化方面较差。术语“业务过程维度模型”既指支持业务过程的逻辑维度模型，又指数据库中相应的物理表。换句话说，维度模型既是逻辑的，又是物理的。
- 关系型数据库是存储、管理和查询数据的一般技术。SQL Server 数据库引擎是 Microsoft 的关系数据库引擎。业务过程维度模型可以存储在关系数据库中。支持事务处理的规范化的数据模型也可以存储在关系数据库中。
- 联机分析处理(OLAP)数据库是存储、管理和查询数据的一种技术，专门用于支持商业智能的使用。SQL Server Analysis Services 是 Microsoft 的 OLAP 数据库引擎。业务过程维度模型可以存储在 OLAP 数据库中，但事务数据库不能，除非首先将它转换成明确的维度形式。
- ETL 系统是一个过程集合，可以清理、转换、合并、重复数据删除、日常处理、存档、一致化、结构化用于数据仓库的数据。这些术语在本书中都有描述。早期的 ETL 系统使用 SQL 和其他脚本来构建。尽管一些小型 ETL 系统仍然这么做，但是更重要的大型 ETL 系统使用专门的 ETL 工具。而且，几乎每个 DW/BI 系统都使用 SQL Server Integration Services 等 ETL 工具，因为收益很大，而增加的成本很低或者没有成本。
- 商业智能(BI)应用程序是一些预定义的应用程序，它们可以查询、分析和展现信息，以满足业务需求。有许多复杂性各异的 BI 应用程序，从一系列预定义的静态报表，到直接影响事务系统和公司日常运作的分析型应用程序，应有尽有。使用 SQL Server Reporting Services 可以构建报表设计应用程序，使用各种 Microsoft 和第三方技术可以构建复杂的分析型应用程序。

- 数据挖掘模型是一个统计模型，经常用于根据过去行为的数据来预测未来的行为，或者标识出数据中紧密相关的子集(称为群集)。数据挖掘是一个术语，意指用于不同目标的统计技术或算法的松散(经常改变的)集合，主要包括群集、决策树、神经网络和预测。Analysis Services 数据挖掘是数据挖掘工具的一个示例。
- 即席查询由用户即时创建。维度建模方法被广泛认为是支持即席查询的最好技术，因为简单的数据库结构易于理解。Microsoft Office(特别是 Excel 透视表和 PowerPivot)是市场上最流行的即席查询工具，使用 Reporting Services Report Builder 可以执行即席查询和报表定义。尽管如此，许多系统仍为其重要用户增加了第三方的即席查询工具，作为对 Excel 和 Report Builder 的补充。
- 此外，数据仓库/商业智能(DW/BI)系统包括：源系统提取、ETL、既是关系型又是 OLAP 的维度数据库、BI 应用程序以及即席查询工具。DW/BI 系统也包括了管理工具和实践、面向用户的文档和培训、安全系统，以及本书讨论的所有其他成员。

0.2.3 角色和职责

DW/BI 系统在生命周期中需要许多不同的角色和技能，它们来自业务和技术领域。本节将介绍创建 DW/BI 系统所涉及的主要角色。角色和人之间很少是一对一关系。与我们合作的团队小到只有一人，大到有 40 人(听说有更大的)，大部分 DW/BI 团队在 3～10 个全职成员之间，并根据需要增加其他人。

单个 DW/BI 团队常常同时承担开发和操作任务，不同于大部分技术项目团队，这与 DW/BI 项目开发周期的高度迭代相关。下面的角色与设计和开发活动相关：

- DW/BI 经理负责项目的总体领导和方向把握。DW/BI 经理必须能够与高级业务和 IT 管理人员进行有效的通信，并能够和团队一起工作，以规划 DW/BI 系统的总体体系结构。
- 项目经理负责系统开发过程中项目任务和活动的日常管理。
- 业务项目领导者是业务领域的成员，并和项目经理紧密合作。
- 业务系统分析师或业务分析师负责领导业务需求定义活动，并且经常参与业务过程维度模型的开发。业务系统分析师需要能够在业务和技术之间架起桥梁。
- 数据建模人员负责执行详细的数据分析，包括数据剖析和开发详细的维度模型。
- 系统架构师设计 DW/BI 系统的各个组件，包括 ETL 系统、安全系统、审核系统和维护系统。
- 开发数据库管理员(DBA)创建关系型数据仓库数据库，并且负责总体的物理设计，包括磁盘布局、分区和初始的索引计划。
- OLAP 数据库设计人员创建 OLAP 数据库。
- ETL 系统开发人员创建 Integration Services 程序包、脚本及其他元素，把数据从源数据库移动到数据仓库中。
- 测试领导建立测试环境，编写自动执行测试的脚本；在测试日志数据库上开发并发布报表；进入业务用户群体，获得用户输入，以进行数据质量测试；在系统进

入生产阶段后，不间断地管理自动测试数据质量的过程；给用户群体发布数据质量报表。

- DW/BI 管理工具开发人员负责编写持续管理 DW/BI 系统所需的定制工具。这些工具包括输入元数据的简单 UI、执行系统备份和恢复的脚本或 Integration Services 程序包，以及维护维度体系结构的简单 UI。
- BI 应用程序开发人员负责构建 BI 应用程序，包括标准报表和业务需要的高级分析型应用程序，他们也负责开发 BI 门户中的定制组件，以及把数据挖掘模型集成到业务操作中。

当团队进入部署和操作系统的阶段时，其他大部分角色在 DW/BI 项目开发周期的后期起到一定的作用，其中几个角色是属于严格操作型的。

- 数据干事负责保证数据仓库中的数据是正确的。数据干事一般最好由业务用户群体中的人员担任，因为他对数据有深入的理解，并能很好地判断数据的准确性。
- 安全经理规定业务用户需要的新用户访问角色，以及添加用户到现有的角色中，安全经理也决定 DW/BI 系统的 ETL 后台中的安全过程。
- 关系数据库管理员(DBA)负责管理关系数据仓库数据库的性能和操作。
- OLAP DBA 负责管理 OLAP 数据仓库数据库的性能和操作。
- 协调经理负责保证 DW/BI 的政策和操作遵循企业的规章制度和常规的法令，如隐私权、HIPAA 和 Sarbanes-Oxley。协调经理、安全经理和内部审核人员要紧密合作。
- 元数据经理决定收集哪些元数据、放在哪里以及如何将它们发布到业务领域。如第 15 章所述，元数据一般不进行管理，除非有专门的人负责。
- 数据挖掘分析师对业务很熟悉，常常有一定的统计学背景。数据挖掘分析师开发数据挖掘模型，并和 BI 应用程序开发人员一起设计使用数据挖掘模型的操作型应用程序。
- BI 门户内容经理管理 BI 门户。他决定门户的内容、布局和更新。
- DW/BI 培训人员创建和发布 DW/BI 系统的培训材料。
- DW/BI 团队的用户支持人员必须能够帮助业务用户，特别是即席查询访问。企业提供的帮助除了连接问题之外，并不能提供专业技术的帮助。

0.3 本书内容

本书分为 4 个部分：

- 需求、现实情况和体系结构。
- 建立和填充数据库。
- 商业智能应用程序的开发。
- DW/BI 系统的部署和管理。

第 I 部分：需求、现实情况和体系结构

第 I 部分为本书的其余部分奠定基础。大部分人渴望了解 Microsoft 工具集。虽然这对

体验和了解这个技术是很好的，但对项目来说却是死神之吻。停下来，离开键盘，思考您要做的事情。

第 1 章：定义业务需求

本书首先概述 Kimball 生命周期，接着深入讨论最重要的步骤——收集业务需求，然后简要介绍本书使用的 Adventure Works Cycles 案例研究的业务需求。第 1 章引用图 0-1 中的业务需求定义方框。

非常熟悉 Kimball 方法的读者可以跳过第 I 部分，但必须学习案例研究。

第 2 章：业务过程维度模型设计

该章简要介绍如何开发维度模型，还解释本书使用的术语和概念，理解这些内容是很重要的。该章引用图 0-1 中的维度建模方框。

熟悉 Kimball 方法的读者可以略过该章的大部分内容，只阅读该章最后的 Adventure Works 案例研究。

第 3 章：工具集

体系结构和产品选择任务对 Microsoft DW/BI 系统是很简单的。该章详细讨论如何使用以及在何处使用 SQL Server 的不同组件与其他 Microsoft 产品，以及在系统中的何处使用第三方软件。该章概述图 0-1 中的技术体系结构设计、产品选择和安装方框。

即使非常熟悉 SQL Server 2005 的读者也应该阅读该章，因为它包含了 SQL Server 2008 R2 的新功能，其中一些功能有显著的区别。

第 4 章：系统设置

第 4 章关注图 0-1 中的产品选择和安装方框，描述如何安装和配置 SQL Server 2008 R2 的不同组件，还讨论系统的缩放和配置，选择在多台服务器上发布 DW/BI 系统的方法和原因。

第 II 部分：建立和填充数据库

本书的第 II 部分介绍有效地构建和填充数据仓库数据库的必要步骤。大部分 Microsoft DW/BI 系统都用关系数据库和 Analysis Services 数据库实现了维度数据仓库。

第 5 章：创建关系数据仓库

第 5 章讨论如何为关系数据仓库创建数据库结构。我们还没有移动数据，但离之很近了。该章首先讨论 Kimball 逻辑设计和物理数据模型之间的微小差异，包括初始索引计划、键结构和存储决策等问题。

关系数据仓库的一个重要决策是是否对事实数据分区。如该章所述，分区有许多优点，是大型数据仓库的必备品。

第 6 章：主数据的管理

主数据是为企业集中管理的引用数据。SQL Server 2008 R2 新增的 Master Data Services 为建立主数据管理系统提供了一个工具集。第 6 章描述主数据管理和 Master Data Services

工具，接着讨论使用这个新技术改进数据仓库的一些快捷方法。随着时间的推移，一些公司可能把维度管理从在 SQL Server Integration Services 中实现的传统 ETL 系统转向通过 Master Data Services 进行更主动的数据管理。

第 7 章：设计和开发 ETL 系统

终于开始移动数据了。该章讨论 ETL 系统的基本设计，首先介绍 SQL Server Integration Services(SSIS)，然后说明如何使用 SSIS 建立 ETL 系统的 34 个子系统。34 个子系统分为 4 组：数据提取、数据清理和一致化、数据显示和系统管理。这些子系统都在 SSIS 中讨论。

第 8 章：核心 Analysis Services OLAP 数据库

我们建议 Microsoft DW/BI 系统使用 Analysis Services 作为主要的数据库，供用户查询。关系数据库和 ETL 过程设计得越满足业务需求，设计 Analysis Services 数据库也就越简单。Analysis Services 向导很容易使用，对于小系统，不必考虑高级设置。然而，如果有大量的数据或者许多用户，就需要对 OLAP 引擎有深入的了解。该章的许多部分主要介绍在企业内实现 Analysis Services 的更多内容。

Analysis Services 包含 3 个主要功能：核心 OLAP 数据库、数据挖掘平台和 PowerPivot 用户驱动的 Excel 分析。数据挖掘参见第 13 章，PowerPivot 参见第 11 章。

第 9 章：实时商业智能的设计需求

第 9 章的主题是实时商业智能，讨论如何将实时数据(一般定义为每天多次刷新的数据)引入 DW/BI 系统中。SQL Server 包含了许多支持实时商业智能的功能，该章将讨论如何使用这些功能，以及在实现实时 BI 时面临的不可避免的妥协。

第III部分：商业智能应用程序的开发

本书的第III部分介绍给业务用户提供数据的必要步骤。BI 应用程序是完整 DW/BI 系统的一个主要组件。对于大多数业务用户而言，BI 应用程序与数据仓库同义。BI 应用程序包括简单的静态报表；复杂的数据挖掘应用程序；用户驱动的、使用 PowerPivot 和 Excel 的即席分析；为商业智能系统提供一个入口点的 BI 门户。

第 10 章：在 Reporting Services 中构建 BI 应用程序

该章提供理解可用的 BI 应用程序所需的基本信息。首先大致介绍 BI 应用程序，再在 Kimball 生命周期的环境下论述 BI 应用程序的开发过程。该章的剩余部分深入探讨 Reporting Services，作为创建和发布标准报表的平台。

第 11 章：PowerPivot 和 Excel

PowerPivot 的核心组件是用于 Excel 2010 的一个内存数据库插件，它允许 Excel 用户以内存速度处理数百万行数据。业务用户可以在 PowerPivot 数据库中连接多个离散数据源中的数据，创建复杂的计算和度量。

第 11 章首先介绍 Excel，作为一个分析和报表设计工具。之后专门讨论 PowerPivot，首先简要描述 PowerPivot 及其产品体系结构，之后建立一个示例。最后在 SharePoint 环境下简要论述 PowerPivot 及其在整个 DW/BI 系统中的作用。

第 12 章：BI 门户和 SharePoint

BI 门户是大多数业务群体获取信息的主要起点。它需要结构化，以允许人们在日益增长的报表设计和分析中查找需要的信息。SharePoint 是 Microsoft 的门户平台产品。

第 12 章的第一部分讨论 BI 门户，包括设计原则和一个简单的示例。第二部分在较高的层面上讨论 SharePoint，作为一个 BI 门户平台，并论述使用 SharePoint 和一组 BI 相关功能的过程，包括 Reporting Services 和 PowerPivot for SharePoint。

第 13 章：数据挖掘的加入

数据挖掘也许是 BI 工具箱中最强大、最不被理解的工具。该章定义数据挖掘，提供如何使用它的示例。我们讨论的是 Microsoft 的数据挖掘技术，也包括包含在 SQL Server Analysis Services 中的算法。该章提供如何构造数据挖掘模型以及如何将挖掘结果合并到系统中的实践指导。为了使这个理论的讨论更加具体，该章还包含两个案例研究。

第Ⅳ部分：DW/BI 系统的部署和管理

本书的第Ⅳ部分包含如何部署和操作 DW/BI 系统的信息，这是整本书中最令人兴奋的部分。

第 14 章：设计和实施安全保护

该章通过鼓励开发一个信息的公开访问策略来开始对 DW/BI 系统安全的讨论。敏感数据当然需要保护，但大部分通过验证的用户都应能访问数据仓库的大部分内容。

即使有了公开访问策略，也必须保护一些数据。该章描述如何在 SQL Server 的不同组件中控制访问，包括 Reporting Services、关系数据库和 Analysis Services。该章也讨论数据仓库后备开发中不同的安全问题。

安全的讨论与图 0-1 中部署方框和维护方框关联最紧密。

第 15 章：元数据规划

许多人都讨论过元数据，但全面而成功地实现元数据的例子很少。我们当然希望在 SQL Server 中有集成的元数据服务，这样就可以简单地描述它，但事实并非如此。该章花费大量篇幅详细讨论了我们认为最重要的元数据，接着描述了维护和发布该信息的步骤。

元数据与图 0-1 中的部署方框和维护方框关联。

第 16 章：部署

部署 DW/BI 系统包含两个主要的任务集。首先需要部署系统。这个任务主要包括测试数据、过程、性能以及部署脚本本身。部署脚本应该包括一个小册子，一步步地指导您如何部署系统变化。

另一个主要的部署活动集专注于业务用户而不是技术。需要开发、发布培训和文档材料，将第 12 章描述的 BI 门户放在一起，然后开发一个业务用户支持计划，因为这些业务用户不可避免地会存在一些问题。

第 17 章：运行与维护

当业务用户开始使用数据仓库来回答常见问题时，就开始依赖它。如果用户认为数据

仓库不可靠，就将使用原来获取信息的方式。这种依赖是一种信任，必须尽其所能保持这种信任。您需要监控数据加载和用户查询的使用率和性能，追踪系统资源，并确保不会用光磁盘空间。简而言之，要像产品系统一样维护数据仓库，必须时刻注意加载到数据仓库的数据质量。一旦业务用户对数据的精确性失去了信任，这种信任几乎是不可能重建的。

第 18 章：目前的需要及未来的展望

第 18 章回顾 DW/BI 项目的主要阶段，并强调最大的风险也就是项目的成功所在。本章最后还介绍 Microsoft BI 工具集即将具有的特征和功能。

0.4 补充信息

本书包含了使用 SQL Server 2008 R2 成功构建和部署一个基本的 DW/BI 系统所需要的大部分信息。为了使本书较小，以便放进背包中，本书并没有包含可以在 SQL Server 联机丛书中很容易找到的工具指令，而是在合适的地方提供帮助寻找相关材料的搜索主题。在许多地方，都建议您在完全理解某些技术材料之前阅读 SQL Server 自带的指南。

本书并没有讲授数据仓库的基础知识。我们在 *Kimball Data Warehouse Toolkit* 系列的其他书籍中总结了许多概念和技术，而不是将所有细节包含进来，并根据需要提供这些书籍关键部分的参考。表 0-1 显示了 *Toolkit* 系列的核心书籍、其主要内容以及面向的读者。

这些书体现了 Kimball Group 关于数据仓库和商业智能的集体智慧，建议您将这些书添加到团队的库中。

本书包含一些技巧、关键概念、侧边栏和章节指针，使本书更有用，更便于参考。它们采用如下格式：

参考：
一些参考指南，用于查找可以补充到 DW/BI 库中的其他材料。这包括 SQL Server 联机丛书、其他书籍和文章以及在线参考材料。

资源：
提供对其他 Kimball Group 材料的参考，例如 *Toolkit* 书籍、文章或设计提示。

注意：
提供了当前讨论的主题的一些额外信息，以及澄清材料的解释或细节。

警告：
帮助您避免花费时间、数据或脑力的潜在危险。

下载：
可以下载的资源。