

**Broadview**  
WWW.BROADVIEW.COM.CN

数据仓库与数据挖掘  
技术应用丛书

IBM® Data Warehousing  
With IBM Business Intelligence Tools

**IBM 数据仓库  
及 IBM 商务智能工具**

[美] Michael L. Gonzales 著

吴刚 董志国 等译

 **WILEY**



电子工业出版社  
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

<http://www.phei.com.cn>

数据仓库与数据挖掘技术应用丛书

# IBM 数据仓库 及 IBM 商务智能工具

---

IBM<sup>®</sup> Data Warehousing  
With IBM Business Intelligence Tools

[美] Michael L. Gonzales 著

吴刚 董志国 等译

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

## 内 容 简 介

本书主要介绍了商务智能和数据仓库的一系列问题,如数据体系结构、技术体系结构、OLAP、数据挖掘、空间分析,以及数据抽取、转换和装载(ETL),更集中关注的是如何利用IBM的一系列应用广泛的成熟技术解决这些问题,例如,SMP和MPP技术体系结构、DB2通用数据库、DB2 OLAP Server数据管理技术、智能挖掘器、空间扩展器等。本书写作方向明确,主题清晰,兼顾了深度与广度,在彻底地剖析了商业智能之后,对使用IBM的关于建立、维护和挖掘数据仓库领域中的丰富产品族提出了颇具实用性的建议。

本书可供在公司中决策数据仓库所需工具和技术的决策者学习参考,也适于数据架构师、项目计划人或发起人阅读。

**John Wiley & Sons, Inc.**

© 2002 by Michael L. Gonzales

All rights reserved. Authorized translation from the English language edition published by John Wiley & Sons, Inc.

本书简体中文专有翻译出版版权由John Wiley & Sons Inc. 授予电子工业出版社,未经许可,不得以任何方式复制或抄袭本书的任何部分。

版权贸易合同登记号 图字: 01-2003-0739

### 图书在版编目(CIP)数据

IBM 数据仓库及 IBM 商务智能工具/(美)冈萨雷斯(Gonzales,M.L.)著;吴刚等译. —北京:电子工业出版社,2004.7

(数据仓库与数据挖掘技术应用丛书)

书名原文: IBM® Data Warehousing With IBM Business Intelligence Tools

ISBN 7-121-00016-4

I. I… II. ①冈… ②吴… III. ①数据库系统 ②电子商务 IV. ①TP311.13 ②F713.36

中国版本图书馆CIP数据核字(2004)第058443号

责任编辑:孙学瑛

印刷:北京智力达印刷有限公司

出版发行:电子工业出版社

北京市海淀区万寿路173信箱 邮编 100036

经 销:各地新华书店

开 本:787×980 1/16 印张:40 字数:593千字

印 次:2004年7月第1次印刷

印 数:4000册 定价:88.00元

凡购买电子工业出版社的图书,如有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系。联系电话:(010)68279077。质量投诉请发邮件至 zltz@phei.com.cn,盗版侵权举报请发邮件至 dbqq@phei.com.cn。

# 出版说明

如果没有对海量数据进行科学分析的能力，沃尔玛的老板再精明，也绝对想不到“啤酒与尿布”这两个风马牛不相及的东西之间还有着千丝万缕的联系。而将它们放在一起，竟然增加了啤酒销量，可见数据分析的巨大威力。

信息系统数年中收集了海量数据，且数据还正以指数级增长，企业迫切地需要高效、精确、科学地分析数据，以找出其背后的寓意，进而了解企业的经营状况和外部环境，做出科学的决断，在现代激烈的竞争中胜出。所以，如何将数据点石成金，更是摆在我们面前很现实也很诱人的一个问题。

现在，很多人已经意识到数据中潜在的大量商机，并踏踏实实地进行着从数据中沙里淘金的工作。特别是在信息化的大潮中，上至政府，下到企业，从银行到电信，再到网站、超市，人们都希望用数据分析这根魔杖赢得先机。与此同时，人们也在期盼着相关书籍，以便工作中学习参考。在广泛征询专家和用户的基础上，秉着选题全面、内容经典、译者严谨的原则，我们适时地推出了这套《数据仓库与数据挖掘技术应用丛书》，以飨读者。本丛书有如下几本：

- 数据仓库基础
- OLAP 解决方案：多维信息系统的构建技术
- 数据仓库工具箱：维度建模的完全指南（第二版）
- 数据仓库生命周期工具箱：设计、开发和部署数据仓库的专家方法
- 数据仓库及其在电信领域中的应用
- 疑难数据仓库专家解决方案
- IBM 数据仓库及 IBM 商务智能工具
- 可视化数据挖掘：数据可视化和挖掘的技术和工具
- 点击流数据仓库
- Web 数据挖掘：将客户数据转化为客户价值
- 企业信息工厂
- 机器学习与数据挖掘：方法和应用

本丛书既包括商业智能（BI）的基础——数据仓库（DW），也包括数据仓库上的两类不同目的的数据增值操作——联机分析处理（OLAP）和数据挖掘（DM）；既覆盖基础理论，如数据仓库基础，又提供不同领域的解决方案，



如数据仓库在电信、银行、保险等领域的应用。

本丛书来自国外数据库领域一些著名作者的畅销书，以及国内第一线实施者的精心总结。如一直位居 AMAZON 畅销书榜的数据仓库领域的畅销书作家 Ralph Kimball 的《数据仓库工具箱：维度建模的完全指南（第二版）》、《数据仓库生命周期工具箱：设计、开发和部署数据仓库的专家方法》，数据仓库之父 William H.Inmon 的《企业信息工厂》（Corporate Information Factory）等。

丛书的译者均来自工作在该领域一线的人员，既有该领域的理论和实践经验，又具备中英文翻译的功底。且多位译者先前均已读过原著，所以，自感翻译的过程不再是枯燥，而是情趣盎然，乐在其中。

出版高品位、高品质的图书是博文视点的努力目标。希望您对我们的工作多提宝贵意见。您的意见是我们创造精品的动力源泉。

如果您希望将您的工作经验感悟等总结成书，我们将为您提供一流的服务，共创精品图书。

我们的联系方式如下：

地址：北京复兴路 47 号天行建商务大厦 604

邮编：100036

电话：010-51922832，68216158

传真：010-51922823

E-mail: [jsj@phei.com.cn](mailto:jsj@phei.com.cn); [zsh@phei.com.cn](mailto:zsh@phei.com.cn)

博文视点资讯有限公司

2003 年 10 月

# 译者序

21 世纪是知识爆炸的时代，我们每个人在处理日常工作和学习事务时都面对着一种同样的过程：从采集知识、筛选知识、使用知识到存储和管理知识。在纷繁复杂的知识面前，我们一边感叹着弗朗西斯·培根的名言：**Knowledge is power**（知识就是力量），一边却在知识的海洋中迷失了方向。

《IBM 数据仓库及 IBM 商务智能工具》正是一本能够引导我们在 IBM 的产品和技术中获取、使用和管理知识的上乘之作。

本书比较系统地阐述了商务智能（BI）和数据仓库问题，以及如何通过 IBM 的技术来解决这些问题。它包含的内容只和数据仓库有关，涉及到 IBM 产品时，不具体介绍细节，比如不介绍 SQL 的语法，因为市面上各式各样的 DB2 SQL 的书籍唾手可得。只有那些专门描述与商务智能或数据仓库相关的问题才会在本书中有所提及。不仅如此，本书也不具体展开其所研究技术的各个方面，比如，不讨论 DB2 V8 的所有特征和功能，因为你可以发现成打的介绍数据库引擎的一般功能的书。此书将重点放在了阐述数据存储的架构、分析和数据管理方面。所以，读者在本书中看到的将是 IBM 产品中那些影响 BI 和数据仓库的技术。这使本书避免了一些无病呻吟的噪声。作者的目的是为了让学生在阅读后可以收到立竿见影的效果。

承担本书翻译工作的有南京理工大学的吴刚、王永明、郑立青、赵昱、鞠飞、罗力恒和苏州科技学院的董志国。

限于译者水平，译文中定有许多不妥之处，敬请读者批评指正。

# 致 谢

我谨向 Gary Robinson 致以特别的感谢，感谢他所付出的劳动及对我们的指导和帮助。假如没有他的帮助，我们不可能找得到写这本书所必要的确认过的资料。

## 对本书做出贡献的人

**Nagraj Alur** 是坐落于圣何塞 (San Jose) 的 IBM 国际技术支持组织中的一位项目领导者。他有 28 年以上的关于数据库管理系统 (DBMS) 的经验。他曾经是一位程序员、系统分析师、项目领导者、咨询师和研究员。他的专业领域涵盖了数据库管理系统、数据仓库、分布式系统管理、数据库性能及客户-服务器和 Internet 计算。

**Steve Benner** 目前负责 ESRI 公司战略事务。过去 13 年，他在地理信息系统 (GIS) 行业从事过各种不同的工作，他曾经在 TDWI 教授过 GIS 和数据仓库方面的课程，并就 GIS 和 SAP 的集成为《SAP 技术周刊》杂志撰文。

**Ron Fryer** 一直从事 IBM 数据管理工作。作为一个数据库建模和数据库管理人员，他在设计、组建决策支持环境上拥有 20 年以上的工作经验，这其中包括超过 10 年的数据仓库经历。他曾经为世界上最大的几个数据仓库工作过。Ron 发表了大量的关于数据库设计及数据库管理系统结构的文章。他也是“Understanding Database Management Systems, Second Edition” (Rob Mattison, McGraw-Hill, 1998) 一书的作者之一。

**Jacques Labrie** 从 1984 年起至今，曾经是 IBM 多个产品的研发团队的领导者和关键研发人员，他同时也是 IBM DB2 Data Warehouse Center 和数据仓库管理器的架构师。Jacques 有超过 15 年的领导和管理数据管理产品的经历，其中包括 ETL 工具，如 IBM 的数据抽取产品；基于工作站的元数据管理工具，如 IBM 数据指南和信息分类管理器；数据仓库管理工具，如 IBM 可视化数据仓库和 DB2 Data Warehouse Center。Jacques 在位于圣何塞的加利福尼亚州立大学获得了数学学士学位。

**Gregor Meyer** 从 1997 年起为 IBM 工作，他当时加入了位于德国的 DB2 智能挖掘器的开发团队。目前，他在 IBM 的位于圣何塞的硅谷实验室工作，

主要负责数据挖掘集成和其他 DB2 商务智能技术的工作。Gregor 曾经在德国的布伦瑞克和斯图加特学习计算机科学，他在德国的哈根大学获得了博士学位。

**Wendell B. Mitchell** 目前是 Focus 集团公司资深的数据架构师。他在众多 TDWI 会议上提出数据挖掘、ETL、商务智能和 OLAP 等方面的实验指导。Wendell 在密歇根州的西密歇根大学分别获得了数学和计算机科学的学士学位。

**Roger D. Roles** 目前是信息分类元数据管理应用的架构师。他是一位软件开发的老手，具有 27 年的开发经历：在 Fortran 语言环境下从事计算机辅助设计和应用开发，用 C 语言及汇编语言开发 UNIX 操作系统内核。自 1993 年起为 IBM 工作的这一段时间里，他在不同部门从事过微内核、文件系统和应用开发。在最近的 6 年中，作为一个团队领导者和核心开发人员，他从事在 Java 环境下开发商务智能应用的工作。

**Richard Sawa** 自从 1998 年起一直在 Hyperion Solutions 供职。他目前在俄亥俄州的哥伦比亚担任 IBM 数据管理部门的 Hyperion Solutions 的技术开发经理。他是 IBM 红皮书“DB2 OLAP Server Theory and Practice” (2001 年 4 月) 的主要写作者，先前作为一个独立的顾问，Sawa 拥有 10 年左右的相关决策支持和 OLAP 技术的经验。

**William Sterling** 从 1992 年起就一直从事 OLAP 工作，当时他在 Arbor Software (ESSBASE 的诞生地)。他的专长是调节 OLAP 数据库，重点放在商业系统建模、量化分析和设计上。1999 年，他加入 IBM，成为 IBM 全球商务智能分析团队的一名技术人员。

**Phong Truong** 是 IBM DB2 Data Warehouse Center 和数据仓库管理器中主要的数据仓库服务器的开发人员，并且他也是 Trillium、MQSeries 及 OLE DB 集成项目组的领导者。他具有 13 年以上的广泛的研发经验及 DB2 UDB 组件的客户服务经验。他在加拿大的阿尔伯达的卡尔加里大学获得了理科学士学位。

**Paul Wilms** 已经在 IBM 从事了 20 年以上的分布式数据库和商务智能工作。他撰写及与别人合作撰写了几篇有关 IBM 的 R\* 及 Starburst 研究项目的论文。在最近的 10 年中，他为 IBM 商务智能及 ETL 工具的客户的技术支持及咨询工作。Paul 在美国及海外的许多国际会议上做过讲座。他在法国格勒诺布尔的国家理工学院获得了计算机科学的博士学位。

**Cheung-Yuk Wu** 目前是 IBM DB2 Data Warehouse Center 和数据仓库管理器的架构师。她有 15 年以上的相关数据库工具开发经验，包括运行于 Windows



或 UNIX 平台之上的 DB2、Oracle、Sybase、Microsoft SQL Server 和 Informix 数据库。她也开发了一些产品，如为 DB2 开发 Tivoli，为 UNIX 开发 IBM Data Hub 和 QMF。此外，她还是 IBM 圣何塞制造数据中心的 DB2、CICS 和 IMS 的数据库管理员。她于加利福尼亚州立理工大学获得计算机科学学士学位。

**Chi Yeung** 是 IBM Data Warehouse Center 和数据仓库管理器主要的 GUI 开发人员。目前，他是多个数据仓库 GUI 组件的团队领导者，这些 GUI 组件包括数据仓库源、进库/出库/发布、用户团体、代理机构，以及复制步骤。他有超过 13 年的在 IBM 产品上的可扩展 GUI 和面向对象程序设计和开发的经验，这些产品包括智能挖掘器、内容管理器、Lotus Approach 和 QMF 的集成及可视化工具等。他在康奈尔大学获得了理学学士学位，在斯坦福大学获得了理学硕士学位，在加州大学伯克利分校获得了商业管理硕士学位。

**Calisto Zuzarte** 是 IBM Toronto 实验室 DB2 Query Rewrite 开发组的资深技术管理人员。他的特长在于主键的查询回写，以及那些影响数据库中的复杂查询性能的基于成本的优化组件。

**Vijay Bommireddipal** 是 IBM DB2 Data Warehouse Center 和数据仓库管理器开发团队的一名成员，一直从事数据仓库进库/出库的应用(包括 tag 及 CWM 格式)、数据仓库案例，以及数据仓库元数据交换的 ISV 工具包方面的工作。他于 2000 年 7 月加入 IBM，那时他已获得位于达特茅斯的麻萨诸塞大学的电子和计算机工程的硕士学位。

# 导 言

系统设计师、项目策划者和发起人总是要面对各种技术、一些相互冲突的方法，以及剧烈竞争的商业问题的。这些问题结合在一起，引发了许多商务智能（BI）和数据仓库（DW）所面临的挑战。现在，您需要问自己的问题是：为了要处理手头的的一个商务要求，我是否拥有做出关于正确地使用什么样的技术和方法的这一决策所需的信息？

我们当然可以把这些技术划分为一些大类，比如，数据获取软件、数据管理软件、数据访问软件，甚至是硬件。然而，这些分类通常会误导决策者，让他们认为选择是如此的简单，而事实却是，每一类下面或许就有浩如烟海的技术，而每种技术产品又有各式各样让人犯晕的特征和功能。这种种选择问题只有当您试图在决策过程中引入方法概念时才会被激化。

由技术和方法所导致的无数选择使得决策者们看起来好像是一只被汽车前灯照射下的手足无措的小鹿。他们因下面这一系列的问题而苦恼：

- 是需要建立多个相互依赖的数据集市，还是让它们彼此独立？
  - 为什么这样建立？
  - 它们之间有什么区别？
- 数据仓库环境应该是集中式还是分布式？
  - 每种情况下，分别需要什么样的硬件技术？
  - 什么是 SMP（对称多进程）、MPP 及聚簇（clustering）？为什么这些技术会影响到数据仓库的工作效果？
  - 这种结构会如何影响数据仓库的原子层及所有相关的数据集市？
- 应该如何给公司各用户团体提供空间数据？
  - 应该建立星状结构还是立方体结构？
  - 它们有什么区别？
  - 为什么择其一而弃其他，或者说它们之间甚至是互斥的吗？
  - 什么是 MOLAP（Multidimensional Online Analytical Processing，多维在线分析处理）、ROLAP（Relational OLAP）以及 HOLAP（Hybrid OLAP）？它如何影响结构？如何影响用户团体？

- 如何加强、完善和补充那些导入数据仓库的数据以支持商务智能？
  - 怎样将从第三方（如 Dunn & Bradstreet）获得的数据和通过如地理编码这样的技术获得数据融合？
  - 什么是空间分析？它如何为组织建立信息内容？
  - 什么是数据挖掘？用户团体如何从中获利？

这本书帮助您回答在 IBM 技术领域中的诸如此类的问题，而此技术领域本身就包罗万象。IBM 提供了一系列广泛的成熟技术，以支持企业级的商务智能环境和数据仓库议案，从 SMP 和 MPP 技术体系结构到 DB2 通用数据库，从 DB2 OLAP 服务器数据管理技术到智能挖掘机和空间扩展器，IBM 系列产品正是创建商务智能环境和建立企业数据仓库所需的技术支架。

本书将注意力集中于 BI 和数据仓库问题，以及如何用 IBM 技术来解决这些问题，数据架构、技术架构、OLAP、数据挖掘、空间分析，以及抽取、转换、装载（ETL）代表了本书所涵盖的一些中心议题。

我们认为，既然题目是数据仓库，包含的内容也应该只和数据仓库有关。因此，您不会在本书中看到完整的 SQL 语法，对所有对其感兴趣的人来讲，市面上各式各样的 DB2 SQL 的书籍唾手可得。只有那些专门解决与商务智能或数据仓库有关问题的 SQL 才会在本书中有所提及。

不仅如此，本书也不会涵盖其所研究的技术的各个方面。比如，我们不会讨论 DB2 V8 的所有特征和功能，因为您可以发现成打的介绍数据库引擎的一般功能的书。本书的重点放在每项技术中和商务智能及数据仓库相关的方面。

所以，您将在本书中看到是 IBM 产品中那些只涉及商务智能和数据仓库的技术。例如，本书的第 5 部分标题为“OLAP 和 IBM”，分为三章：第 11 章集中讨论 DB2 OLAP Server，第 12 章讨论 DB2 Data Warehouse Center 中支持 DB2 OLAP Server 的方面，第 13 章则是关于 DB2 V8 中的 OLAP 函数。

之所以将重点这样分配，理由很简单：它避免了一些无病呻吟的噪声，只提供那些对我们的商务智能和数据仓库工作极为关键的问题，以及相关的且紧密联系的内容。仅此而已。我们的目的就是为了让您的阅读可以收到立竿见影的效果。

## ▲ 本书的组织结构

本书共分 16 章，分成如下的 6 大部分。

**第 1 部分：** 商务智能和数据仓库的基础知识      这部分的重点是建立一

种通用的语言描述，以及介绍商务智能和数据仓库的基本概念。如果您是这个领域的新手，请务必通读前面的这些章节。如果您是经验丰富的“仓库管理员”，则可以跳到下一部分。这部分的章节分布如下。

- 第 1 章：BI 组织概述
- 第 2 章：商务智能基础
- 第 3 章：设计数据仓库的迭代过程

**第 2 部分：商务智能架构** 这是至关重要的一部分，因为它覆盖了仓库中的两个架构领域：数据架构和技术架构。这对刚刚开始从事数据仓库工作的一些人来说，是必读章节，对经验丰富的人员来说，为了保证他们可以跟踪 IBM 在这些中心结构方面的最新科技成果，这部分仍然值得翻阅。这部分只包括两章。

- 第 4 章：设计数据架构
- 第 5 章：技术架构和数据管理基础

**第 3 部分：数据管理** 尽管 DB2 V8 的特征和功能很广泛，但是我们只提供了 DB2 V8 中和商务智能及数据仓库有关的部分。这部分包括两章，都是关于 DB2 的。

- 第 6 章：DB2 BI 基础
- 第 7 章：DB2 物化查询表

**第 4 部分：数据仓库管理** 我们将在这里展示帮助管理数据仓库的 IBM 技术，这部分包括三章，主要涵盖 IBM DB2 Data Warehouse Center 的内容。

- 第 8 章：利用 IBM DB2 Data Warehouse Center 管理数据仓库
- 第 9 章：在 IBM DB2 Data Warehouse Center 中进行数据变换
- 第 10 章：元数据和 IBM DB2 仓库管理器

**第 5 部分：OLAP 和 IBM** 这部分侧重于 IBM 技术的 OLAP 主题。包括三章，每章介绍一种不同的技术，分别是 DB2 OLAP Server、DB2 V8，以及 IBM DB2 Data Warehouse Center。

- 第 11 章：多维数据和 DB2 OLAP Server
- 第 12 章：OLAP 和 IBM DB2 Data Warehouse Center
- 第 13 章：DB2 OLAP 函数

**第 6 部分：高级分析** 最后，本书介绍真正可以丰富数据仓库，将数据转换成信息内容的 IBM 技术。这里展示的是数据挖掘和空间分析的技术方法。包括以下三章。

- 第 14 章：带智能挖掘器的数据挖掘

- 第 15 章：DB2——增强的 BI 特征和函数
- 第 16 章：向数据仓库中添加空间数据

只要您对以上讨论的这些技术在什么时候，以及应该怎样合适地切入 BI 组织的总体结构体系有全局的认识，那么这 6 个部分均可以独立阅读。

## 本书适合的读者群

两类读者可以从本书中获益：决策者和实现者。如果您是公司数据仓库所需工具和技术的决策者，并且采用（或正在考虑采用）IBM 的技术，那么本书很适合您，您将从中获得对这项技术中包含的主要议题的一个清晰认识。再者，如果您是数据体系结构架构师、项目计划人或者发起人，并因此可以影响决策过程，您同样应该阅读本书，它将帮助您获得 IBM 技术的相关信息，以及如何应用这项技术中的某些特定特征和功能满足您的 BI 或数据仓库所需。

此外，如果您负责 IBM 技术在环境中的实现，那么这本书正适合您的需求。它没有喋喋不休的废话，直接将您带到支持您的 BI 和数据仓库项目的特征和功能面前。您不必费心温习那些无关的语法，以及对您的 BI 项目而言无关紧要的特征。

## 网站内容

站点([www.wiley.com/compbooks/gonzales](http://www.wiley.com/compbooks/gonzales))提供了最新技术信息和参考资料、本书中提到的产品的软件更新版本，以及其他 BI 技术的链接。除了 IBM 的产品，我们也计划提供一系列补充 IBM BI 环境的合作伙伴的解决方案。

## 小结

BI 和数据仓库环境需要不断的监视和适当的调节，以保证满足公司的需要。技术日新月异，每天醒来，您都会发现一些技术特征的改进，软件公司对一些软件的升级，或者一个产品新版本的发布。这意味着，如果您负责为公司选择和实现合适的技术，那么您将必须时刻注意这些变化，并因此承受巨大的压力。我们希望本书能提供您所需的、特定的、相关的信息，以使您能紧跟 BI 发展的脚步。

# 对 IBM 数据仓库一书的赞评

“这本书兼顾了深度与广度，它以一种超越平常的方式涵盖了商务智能领域。它不仅描述了各种不同的 IBM 产品，如 IBM DB2、IBM 智能挖掘器，以及 IBM DB2 OLAP，而且还全面地剖析了对数据挖掘体系结构、分析及数据管理。”

Wayne Eckerson  
数据仓库研究所主任

“如今，各个企业都面临着一种‘数据泛滥’的境地，而这些数据是关于顾客、原料供应商、合作伙伴、雇员和竞争对手的。为了生存和发展下去，企业越来越需要信息管理的解决方案。Michael Gonzales 的书对于 IBM 公司的商务智能软件介绍得非常好，而这些软件可以帮助企业更快、更好地做出决策。在全面地介绍了 IBM 数据仓库、OLAP、数据挖掘器和空间分析能力的同时，Michael Gonzales 清楚地指出了在信息竞争日益激烈的时代中，在企业组织和数据的体系结构上取得成功所必不可少的支柱。”

Jeff Jones  
IBM 数据管理方案的高级程序经理

“IBM 公司在提供完备和易用的数据仓库、数据分析和数据管理技术方面处于业界领先。这本书阐述了对于每本涉及数据仓库的书最为重要的专业要点：在彻底地领悟了商业智能基础之后，伴随着的是使用 IBM 关于建立、维护和挖掘数据仓库领域中的丰富产品族的颇具实用性的建议。”

Thomas W. Rosamilia  
IBM 数据管理 (DB2) 全球开发副总裁



# 目 录

## 第 1 部分 商务智能和数据仓库的基础知识

<b>第 1 章 BI 组织概述</b> .....	3
1.1 BI 组织架构概述 .....	4
1.2 提供信息内容 .....	8
1.2.1 信息内容的计划 .....	9
1.2.2 信息内容的设计 .....	12
1.2.3 信息内容的实现 .....	13
1.3 证明 BI 工作的价值 .....	16
1.3.1 项目与已知的商务需求相结合 .....	16
1.3.2 衡量 ROI .....	16
1.3.3 让数据仓库首次迭代最大限度地发挥作用 .....	19
1.4 IBM 和 BI 组织 .....	20
1.4.1 无缝集成 .....	21
1.4.2 简化数据传递系统 .....	23
1.4.3 零滞后 .....	25
1.5 小结 .....	25
<b>第 2 章 商务智能基础</b> .....	27
2.1 BI 组成及技术 .....	28
2.1.1 BI 组成 .....	28
2.2 数据仓库的组成 .....	32
2.2.1 数据抽取、转换和装载 .....	34
2.2.2 数据管理 .....	37
2.2.3 数据访问 .....	37
2.2.4 元数据 .....	38
2.3 分析化的用户要求 .....	39
2.3.1 报告和查询 .....	40
2.3.2 在线分析处理 .....	40
2.3.3 数据统计 .....	43

2.3.4	数据挖掘	43
2.4	维度技术和 BI	44
2.4.1	OLAP 服务器	45
2.4.2	定义空间谱	47
2.5	触点	48
2.6	零滞后和数据仓库环境	49
2.7	闭环学习	49
2.8	历史完整性	50
2.9	小结	53
<b>第 3 章</b>	<b>设计数据仓库的迭代过程</b>	<b>54</b>
3.1	设计任何的迭代	55
3.1.1	建立你自己的 BI 计划	57
3.2	计划初期迭代	63
3.2.1	设定数据仓库以符合公司战略	64
3.2.2	实施一个准备就绪的评估	65
3.2.3	源计划	66
3.3	用 DIF 矩阵选择方案	69
3.3.1	确定正确的方法	70
3.3.2	应用 DIF 矩阵	70
3.3.3	功能紊乱	74
3.3.4	影响力	75
3.3.5	可行性	75
3.3.6	DIF 矩阵的结果	75
3.4	计划随后的迭代	78
3.4.1	确定范围	78
3.4.2	确定战略性的商务询问	79
3.5	完成项目的方法	80
3.5.1	BI 骇客方法	80
3.5.2	Inmon 方法	81
3.5.3	商务空间的生命周期方法	81
3.5.4	螺旋形方法	82
3.5.5	IBM 方法	90
3.5.6	选择合适的方法	93
3.6	小结	93

## 第 2 部分 商务智能架构

<b>第 4 章 设计数据架构</b> .....	97
4.1 选择正确的架构 .....	99
4.1.1 原子层选择 .....	101
4.1.2 数据集市 .....	106
4.1.3 有依赖性的数据集市的原子层 .....	107
4.1.4 独立的数据集市 .....	109
4.1.5 数据传输架构 .....	110
4.2 预测传送量 .....	114
4.3 架构建模 .....	116
4.3.1 商务逻辑模型 .....	117
4.3.2 原子层模型 .....	118
4.3.3 数据集市的建模 .....	120
4.3.4 原子和星型数据的比较 .....	124
4.4 可操作的数据的存储 .....	125
4.5 数据架构策略 .....	126
4.6 小结 .....	129
<b>第 5 章 技术架构和数据管理基础</b> .....	131
5.1 决定架构的技术依据 .....	133
5.1.1 集中的数据仓库 .....	133
5.1.2 分布式的数据仓库 .....	137
5.1.3 并行和仓库 .....	139
5.1.4 分区数据存储 .....	142
5.2 数据管理的技术基础 .....	143
5.2.1 DB2 和原子层 .....	143
5.2.2 DB2 和星型模式 .....	148
5.3 DB2 技术架构的本质 .....	150
5.3.1 SMP、MPP 和簇 .....	150
5.3.2 共享资源与无共享 .....	152
5.3.3 静态和动态并行 .....	154
5.3.4 目录分区 .....	155
5.3.5 高可用性 .....	155
5.4 尺寸需求 .....	162
5.5 小结 .....	164