

<<数据仓库设计>>

图书基本信息

书名：<<数据仓库设计>>

13位ISBN编号：9787302230748

10位ISBN编号：7302230749

出版时间：2010-8

出版时间：清华大学出版社

作者：（意）戈尔法雪利 等著

页数：393

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

## <<数据仓库设计>>

### 前言

数据仓库设计是一个重要的课题，是商业信息的核心，而信息是现代组织的核心，小管组织是一家企业、一家非盈利性组织还是一个政府机构，信息都具有无比重要的价值。

就像生物体需要使用信息一样，组织需要使用信息来调节内部操作以及适应外部变化。

人员组织使用有关内部运营状况的信息来管理成本及正确地分配资源，使用有关外部世界的信息来管理客户和供应商，以及应耐竞争和市场变化。

能够很好地利用信息的组织往往能够更加稳定地发展，而不能有效利用信息的组织则在苦苦挣扎。

现在，信息比以往任何时候都更加重要，生成和存储的信息量浩如烟海。

我们面临着庞大的信息量。

我们居住的模拟世界正在逐渐数字化，而在数字化的世界中，每个事件都可以被已求、归类并存储起来，供以后分析使用。

每笔商业交易、每次网站访问、每幅查看的图片、每个打出的电话和花费的每一美元都被数字化并记录下来。

在过去，购买就是发生在商店售货员和顾客之间的“模拟交易”，而现在，相同的交易通过互联网或商店的销售点终端系统进行，并将被立即记录下来。

## <<数据仓库设计>>

### 内容概要

《数据仓库设计：现代原理与方法》内容详实，图文并茂，介绍最前沿的数据仓库设计技术，指导您构建安全可靠的决策支持基础结构。

它阐述了遵循成熟可靠的软件工程原理的实用设计方法，讨论如何得心应手地构建富有表达力的概念模式、将概念模式转换为关系模式以及设计最新ETL过程，还讲述如何集成异构数据源、实现星型和雪花模式、管理动态和不规则层次结构以及如何通过实体化和拆分视图来优化性能。

## 作者简介

作者：（意大利）戈尔夫雪利（Matteo Golfarelli）（意大利）Stefano Rizzi 译者：战晓苏 吴云浩 皮人杰 Matteo Golfarelli 是意大利博洛尼亚大学计算机科学与技术学院副教授，讲授信息系统、数据库和数据挖掘课程。

Matteo 从2008年开始担任 Busiless Intelligence Systems 会议的联合主席，并任国际杂志 Data Mining and Management 的编委。

Stefano Rizzi 是意大利博洛尼亚大学计算机科学与技术学院教授，讲授高级信息系统和软件工程课程。他已在国际期刊和会议文献上发表了近100篇关于信息系统、移动机器人系统和模式识别的论文。

Stefano 是 Encyclopedia of Database Systems 杂志的数据仓库设计编辑。

## &lt;&lt;数据仓库设计&gt;&gt;

## 书籍目录

- 第1章 数据仓库简介 1.1 决策支持系统 1.2 数据仓库 1.3 数据仓库的体系结构 1.3.1 单层体系结构 1.3.2 两层体系结构 1.3.3 三层体系结构 1.3.4 另一种体系结构类别 1.4 数据准备和ETL 1.4.1 提取 1.4.2 清洗 1.4.3 转换 1.4.4 加载 1.5 多维模型 1.5.1 限制 1.5.2 聚合 1.6 元数据 1.7 访问数据仓库 1.7.1 报表 1.7.2 OLAP 1.7.3 仪表盘 1.8 ROLAP、MOLAP和HOLAP 1.9 其他问题 1.9.1 质量 1.9.2 安全 1.9.3 进化第2章 数据仓库系统的生命周期 2.1 风险因素 2.2 自上而下与自下而上 2.2.1 商业维度生命周期 2.2.2 快速数据仓库方法 2.3 数据集市设计阶段 2.3.1 数据源的分析 和协调 2.3.2 需求分析 2.3.3 概念设计 2.3.4 工作负荷细化和概念模式的验证 2.3.5 逻辑设计 2.3.6 物理设计 2.3.7 数据准备设计 2.4 系统方法架构 2.4.1 场景1: 数据驱动的方法 2.4.2 场景2: 需求驱动的方法 2.4.3 场景3: 混合方法 2.5 测试数据集市第3章 数据源的分析与协调 3.1 检查和规范化模式 3.2 集成问题 3.2.1 不同视角 3.2.2 等效建模构造 3.2.3 不兼容的规范 3.2.4 共有概念 3.2.5 相互关联的概念 3.3 集成阶段 3.3.1 预集成 3.3.2 比较模式 3.3.3 对齐模式 3.3.4 合并和重构模式 3.4 定义映射第4章 用户需求分析 4.1 采访 4.2 基于词汇表的需求分析 4.2.1 事实 4.2.2 预备性工作负荷 4.3 面向目标的需求分析 4.3.1 Tropos简介 4.3.2 组织建模 4.3.3 决策建模 4.4 其他要求第5章 概念建模 5.1 维度事实模型: 基本概念 5.2 高级建模 5.2.1 描述性属性 5.2.2 跨维度属性 5.2.3 聚合 5.2.4 共享层次结构 5.2.5 多弧线 5.2.6 可选弧线 5.2.7 不完整层次结构 5.2.8 递归层次结构 5.2.9 可加性 5.3 事件和聚合 5.3.1 聚合可加性度量 5.3.2 聚合不可加度量 5.3.3 使用聚合和跨维度属性聚合 5.3.4 使用可选弧线或者多弧线聚合 5.3.5 空事实模式聚合 5.3.6 使用维度间的函数依赖进行聚合 5.3.7 沿着不完整或者递归层次结构聚合 5.4 时间 5.4.1 事务模式与快照模式 5.4.2 迟更新 5.4.3 动态层次结构 5.5 重叠事实模式 5.6 正式化维度事实模式 5.6.1 元模型 5.6.2 内涵特性 5.6.3 外延特性第6章 概念设计 6.1 基于实体-关系模式的设计 6.1.1 定义事实 6.1.2 构建属性树 6.1.3 修剪和移植属性树 6.1.4 一对一关系 6.1.5 定义维度 6.1.6 时间维度 6.1.7 定义度量 6.1.8 生成事实模式 6.2 基于关系模式的设计 6.2.1 定义事实 6.2.2 构建属性树 6.2.3 其他阶段 6.3 基于XML模式的设计 6.3.1 建立XML关联模型 6.3.2 预备阶段 6.3.3 选择事实并构建属性树 6.4 混合方法设计 6.4.1 映射需求 6.4.2 构建事实模式 6.4.3 细化 6.5 需求驱动的方法设计第7章 工作负荷和数据卷 7.1 工作负荷 7.1.1 维度表达式和对事实模式的查询 7.1.2 横向钻取查询 7.1.3 复合查询 7.1.4 嵌套GPSJ查询 7.1.5 验证概念模式中的工作负荷 7.1.6 工作负荷和用户 7.2 数据卷第8章 逻辑建模 8.1 MOLAP和HOLAP系统 8.2 ROLAP系统 8.2.1 星型模式 8.2.2 雪花模式 8.3 视图 8.4 时间场景 8.4.1 动态层次结构: 类型1 8.4.2 动态层次结构: 类型2 8.4.3 动态层次结构: 类型3 8.4.4 动态层次结构: 完整数据记录 8.4.5 删除元组第9章 逻辑设计 9.1 事实模式到星型模式 9.1.1 描述性属性 9.1.2 跨维度属性 9.1.3 共享层次结构 9.1.4 多弧线 9.1.5 可选选弧线 9.1.6 不完整层次结构 9.1.7 递归层次结构 9.1.8 退化维度 9.1.9 可加性问题 9.1.10 使用雪花模式 9.2 视图实体化 9.2.1 使用视图来回答查询 9.2.2 问题公式化 9.2.3 实体化算法 9.3 视图碎片化 9.3.1 垂直视图碎片化 9.3.2 水平视图碎片化第10章 数据准备设计 10.1 填充协调数据库 10.1.1 提取数据 10.1.2 转换数据 10.1.3 加载数据 10.2 清洗数据 10.2.1 基于字典的技术 10.2.2 近似合并 10.2.3 即席技术 10.3 填充维度表 10.3.1 确定要加载的数据 10.3.2 替换键 10.4 填充事实表 10.5 填充实体化视图第11章 数据仓库的索引 11.1 B+树索引 11.2 位图索引 11.2.1 位图索引与B+树 11.2.2 高级位图索引 11.3 投影索引 11.4 联接和星型索引 11.5 空间索引 11.6 联接算法 11.6.1 嵌套循环 11.6.2 排序一合并 11.6.3 哈希联接第12章 物理设计 12.1 优化器 12.1.1 基于规则的优化器 12.1.2 基于开销的优化器 12.1.3 直方图 12.2 选择索引 12.2.1 索引维度表 12.2.2

## &lt;&lt;数据仓库设计&gt;&gt;

索引事实表 12.3 其他物理设计元素 12.3.1 将数据库划分为表空间 12.3.2 分配数据文件  
12.3.3 磁盘块大小第13章 数据仓库项目文档 13.1 数据仓库层 13.1.1 数据仓库模式  
13.1.2 部署模式 13.2 数据集市层 13.2.1 总线矩阵和重叠矩阵 13.2.2 操作模式  
13.2.3 数据准备模式 13.2.4 域术语表 13.2.5 工作负荷和用户 13.2.6 逻辑模式和物理  
模式 13.2.7 测试文档 13.3 事实层 13.3.1 事实模式 13.3.2 属性和度量术语表 13.4  
系统方法指导原则第14章 案例研究 14.1 应用领域 14.2 计划Transport数据仓库 14.3 销售数  
据集市 14.3.1 数据源分析和协调 14.3.2 用户需求分析 14.3.3 概念设计 14.3.4 逻辑  
设计 14.3.5 数据准备设计 14.3.6 物理设计 14.4 营销数据集市第15章 超越数据仓库范畴  
的商业智能 15.1 商业智能简介 15.2 数据挖掘 15.2.1 关联规则 15.2.2 群集化 15.2.3  
分类器和决策树 15.2.4 时间序列 15.3 假设分析 15.3.1 归纳技术 15.3.2 演绎技术  
15.3.3 系统方法注意事项 15.4 商业绩效管理术语表参考文献

## &lt;&lt;数据仓库设计&gt;&gt;

## 章节摘录

插图：数据仓库层是本书重点讨论的部分。

我们在这里引入一个数据仓库的关键字：多维。

您需要熟悉这里使用的概念和术语才能理解本书中讲解的信息，尤其是关于概念和逻辑建模及设计的信息。

在过去几年中，多维数据库已经引起了很大的研究和市场兴趣，因为对于许多决策制定支持应用程序（比如数据仓库系统），它们起着奠基性的作用。

多维模型用作数据仓库数据表示范式的根本原因是它易于使用并且很直观，即使对于IT新手也是如此。

多维模型的成功也与采用多维模型作为可视化范式的生产工具（比如电子表格）的广泛应用有关。

可能有效地了解多维模型的最佳起点是定义这种模型最适用的查询的类型。

1.7节提供了关于典型的决策查询的更多细节，比如下面的查询（Jarkeeta1.2000）：“去年记录的每个州和每个产品类别的总收入额为多少？”

” “过去5年中PC厂商的份额与季度收入之间的关系是什么？”

” “哪些订单可以实现收入的最大化？”

” “两种治疗方法中哪一种会造成平均住院时间的下降？”

” “包含少于10件物品的配送带来的利润与包含多于10件物品的配送带来的利润之间的关系是什么？”

”

## <<数据仓库设计>>

### 编辑推荐

《数据仓库设计:现代原理与方法》：使用数据和需求驱动的方法创建协调数据库来完善数据集市架构采集并明确表达终端用户要求使用维度事实模型构建概念数据集市模式评估数据集市卷和负载使用高级逻辑建模技术提升性能提取、转换、清洗及加载数据源的数据使用高级索引技术优化查询执行计划编写完整的数据仓库项目文档了解创新商业智能技术



版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>