

图书基本信息

书名：<<数据仓库与数据挖掘原理.工具及应用>>

13位ISBN编号：9787508363103

10位ISBN编号：7508363108

出版时间：2007-12

出版时间：中国电力

作者：潘华

页数：263

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>

## 前言

为贯彻落实教育部《关于进一步加强高等学校本科教学工作的若干意见》和《教育部关于以就业为导向深化高等职业教育改革的若干意见》的精神，加强教材建设，确保教材质量，中国电力教育协会组织制订了普通高等教育“十一五”教材规划。

该规划强调适应不同层次、不同类型院校，满足学科发展和人才培养的需求，坚持专业基础课教材与教学急需的专业教材并重、新编与修订相结合。

本书为新编教材。

自20世纪90年代以来，信息技术在我国电力系统的应用得到了前所未有的发展，各级电力企业纷纷建立各种各样的信息系统，如办公自动化（OA）、生产管理系统、设备管理系统、燃料管理系统、电力市场和营销系统、电力调度系统、送电和配电地理信息系统、呼叫中心（Call Center）等。

然而，这些信息系统往往是根据某个企业，甚至是某个部门自身需求而设计的，信息的采集、加工和存储大多着眼于本企业或本部门的信息，忽视了相互之间信息沟通和共享的要求。

这样建立起来的信息系统虽然覆盖了各方面的信息，但同时也形成了一个信息孤岛，使得原本可以相互沟通和共享的信息被一道道“篱笆”分隔开来。

2002年电力体制改革之后，电力企业解除管制的商业环境以及更加多变的电力市场，使得信息和知识成为电力公司最有价值的资源，而上述情况使得电力企业信息化最终不能构造有效的知识管理系统，信息传递困难，难以提供企业级的决策分析支持。

目前的问题主要表现为以下几项。

（1）异构性强，信息集成度差。

电力企业各应用系统在数据建模、软硬件平台、应用系统平台和开发工具等方面都存在着显著的差异，从而导致彼此数据交换困难，使得各个应用系统在信息上成为相对孤立的“自动化孤岛”，不易与其他系统交换数据或在企业范围内实现集成。

（2）数据冗余和多信息源问题。

由于建设时期的不同以及当时技术水平的限制，造成了过量的数据冗余和多信息源等问题，使得数据资源访问困难，难以进行有效的决策分析。

（3）缺乏企业级的决策支持系统。

电力企业各应用系统信息共享困难，管理系统难以跨应用系统实施生产业务流程管理，不能构造有效的知识管理系统，难以提供管理层和决策层的综合分析和辅助决策支持。

数据仓库和数据挖掘技术可以很好地解决以上问题。

这种技术自20世纪90年代初开始在美国等国家流行，并在20世纪90年代中期传入我国，现在已经逐渐在我国推广应用，特别是在金融、电信、制造、零售等企业，发挥着越来越重要的作用。

相比而言，由于体制、观念、技术、人才等方面的原因，数据仓库与数据挖掘技术在电力行业的应用尚处于起步阶段。

但是可以预测，随着电力体制改革和行业信息化的进一步深入，数据仓库和数据挖掘技术将会在电力行业有很大的应用。

本书编者一直从事数据仓库、数据挖掘方面的研究与开发，所参与设计和开发的项目涉及金融、保险、电力等多个领域。

近年来在上海电力学院也开设了相关课程。

本书是在此基础上编写而成的。

## 内容概要

本书为普通高等教育“十一五”规划教材，是电力企业信息化系列教材之一。

本书全面深入介绍了数据仓库、联机分析处理（OLAP）和数据挖掘的基本概念、工具及实际应用。全书分成三篇，数据仓库与数据挖掘原理篇的主要内容包括数据仓库的基本概念和结构、创建过程、联机分析处理、数据挖掘的基本概念和方法等；数据仓库与数据挖掘工具篇介绍几个现在市场上主流的数据仓库和数据挖掘工具，包括ETL工具Data Stage、商务智能工具 Congos和数据挖掘工具SAS；数据仓库与数据挖掘应用篇以某电力公司为例介绍一个数据仓库应用系统的建设过程，包括系统需求分析、系统架构设计、数据模型设计、数据库规划、ETL开发等。

本书可作为计算机、信息管理与信息系统等相关专业的学生学习数据仓库、OLAP及数据挖掘技术的实用教程，也可供从事数据仓库、数据挖掘研究、设计、开发等工作的科研人员和工程人员参考。

书籍目录

前言 电力企业信息化系列教材序 第一篇 数据仓库与数据挖掘原理篇 第一章 数据仓库概述 第一节 数据仓库的产生 第二节 数据仓库的相关概念 第三节 数据仓库与OLTP的比较 第四节 数据仓库的发展历程 第二章 数据仓库的基本结构 第一节 数据仓库的参考架构 第二节 数据仓库的数据存储和数据模型 第三节 数据加载模块 第四节 数据分析展现模块 第五节 元数据管理模块 第六节 数据仓库门户管理模块 第七节 数据仓库监控和日常管理 第三章 数据仓库的构建 第一节 数据仓库设计开发过程 第二节 数据仓库模型设计 第三节 数据加载设计 第四节 应用及门户系统建设 第五节 元数据管理系统设计 第四章 联机分析处理 第一节 OLAP概述 第二节 OLAP基本操作 第三节 OLAP体系结构和分类 第四节 基于多维数据库的OLAP 第五节 基于关系数据库的OLAP 第六节 OLAP的评价标准 第七节 OLAP的前端展现 第五章 数据挖掘技术 第一节 数据仓库与数据挖掘 第二节 数据挖掘概述 第三节 数据挖掘的决策支持及其方法 第二篇 数据仓库与数据挖掘工具篇 第六章 ETL工具——Data Stage 第一节 Data Stage概述 第二节 创建一个Data Stage工程 第三节 Data Stage作业的开发 第四节 创建BASIC：表达式 第七章 商务智能工具——Cognos 第一节 Cognos概述 第二节 Framework Manager建模过程 第三节 使用Report Studio开发固定式报表 第四节 使用Powerplay开发OLAP报表 第八章 数据挖掘工具——SAS 第一节 SAS系统工作环境 第二节 SAS程序结构 第三节 SAS数据挖掘实例 第三篇 数据仓库与数据挖掘应用篇 第九章 数据仓库与数据挖掘在电力行业应用概述 第一节 电力行业信息化建设概况 第二节 数据仓库与数据挖掘在电力行业应用 第十章 某省电力营销数据仓库应用系统建设 第一节 系统需求分析 第二节 系统架构设计 第三节 数据模型设计 第四节 数据库规划 第五节 ETL开发 第六节 系统实现参考文献

## 章节摘录

数据仓库的开发应用规划是开发数据仓库的首要任务。

只有制定了正确的数据仓库规划，才能组织主要力量有序地实现数据仓库的开发应用。

在数据仓库的规划中一般要经历这样几个步骤：选择实现策略、确定数据仓库的开发目标和实现范围、选择数据仓库的体系结构、建立商业和项目规划预算。

当数据仓库规划完成后，就需要编制数据仓库的规划说明书，说明数据仓库与企业战略的关系，以及与企业急需处理的、范围相对有限的开发机会，重点支持的职能部门和今后数据仓库开发工作的建议，实际使用的方案和开发预算，以作为数据仓库实际开发的依据。

1. 选择数据仓库的实现策略数据仓库的开发策略主要有自顶向下、自底向上以及这两种策略的联合使用。

自顶向下策略在实际应用中比较困难，因为数据仓库是一种决策支持功能，这种功能在企业战略的应用范围中常常是很难确定的，因为数据仓库的应用机会往往超出了企业当前的实际业务范围。

而且，在开发前就确定目标会在实现了预定的目标后不再追求新的应用，使数据仓库丧失更有战略意义的应用。

但是，由于该策略在开发前就可以给出数据仓库的实现范围，能够清楚地向决策者和企业描述系统的收益情况和实现目标，因此是一种有效的数据仓库开发策略。

该方法的使用需要开发人员具有丰富的自顶向下开发系统的经验，企业决策层和管理人员完全知道数据仓库使用的预定目标并了解数据仓库能够在哪些决策中发挥作用。

自底向上策略一般从某一数据仓库的原型开始，选择一些特定的为企业管理人员所熟知的管理决策问题作为数据仓库开发的对象，在此基础上进行数据仓库的开发。

因此，该策略常常用于一个数据集市或一个部门的数据仓库的开发。

该策略的优点在于企业能以较小的投入获得较高的数据仓库应用效益，在开发过程中，人员投入较少，也容易获得成效。

当然，如果某个项目的开发失败可能会造成企业整个数据仓库系统开发的推迟。

该策略一般用于企业希望对数据仓库的技术进行评价，以确定该技术的应用方式、地点和时间，或希望了解实现和运行数据仓库所需要的各种费用，或在数据仓库的应用目标并不是很明确以及数据仓库对决策过程影响不是很明确时采用。

在自顶向下的开发策略中可以采用结构化或面向对象方法。

按照数据仓库的规划、需求确定、系统分析、系统设计、系统集成、系统测试和系统试运行的阶段完成数据仓库的开发。

而在自底向上的开发中，则可以采用螺旋式的原型开发方法，使用户可以根据新的需求对试运行的系统进行修改。

螺旋式的原型开发方法要求在较短时间内快速生成可以不断增加功能的数据仓库。

螺旋式的原型开发方法适用于这样一些场合：企业的市场动向和需求无法预测；市场的时机是实现产品的重要组成部分；不断地改进对于企业的市场调节是必需的；持久的竞争优势来自连续不断的改进；系统的改进基于用户在使用过程中的不断发现。

自顶向下和自顶向上策略的联合使用具有两种策略的优点，既能够快速地完成数据仓库的开发与应用，又可以建立具有长远价值的数据仓库方案，但是在实际使用中难以操作，通常需要能够建立、应用和维护企业模型、数据模型和技术结构的、具有丰富经验的开发分析人员，能够熟练地从具体（如业务系统中的元数据）转移到抽象（只基于业务性质而不是基于实现系统技术的逻辑模型）。

企业需要拥有由最终用户和信息系统人员组成的有经验的开发小组，能够清楚地指出数据仓库在企业战略决策中的应用。

编辑推荐

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>