

<<数据仓库与数据挖掘原理及应用>>

图书基本信息

书名：<<数据仓库与数据挖掘原理及应用>>

13位ISBN编号：9787030254009

10位ISBN编号：7030254007

出版时间：2009-9

出版时间：科学出版社

作者：王丽珍 等编著

页数：296

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>

## <<数据仓库与数据挖掘原理及应用>>

### 前言

进入信息社会以来，信息技术经历了从计算机主机的信息集中处理方式到个人计算机（PC）的信息分布处理形式的转变；从单一的计算机操作系统到计算机互网络操作的改变：从客户机 / 服务器（Client / Server）计算体系到多层体系结构计算模式的转变；从单一数据库到大型数据仓库和从局域网到Internet全球网的改变。

现代信息技术的发展和现代科学技术的进步，使人类迈入新的时期——信息化时代。

信息处理技术的发展使得各类数据、信息急剧增长，给数据的传输、存储带来了许多新的问题，特别是由于各类不同事务产生大量不同类型的数据，这些数据分别被许多各个时期建立的应用系统所使用。

人们希望能够看到所有数据和信息的综合情况，而这些数据和信息有许多不能被统一描述，不能被现有应用系统综合使用。

针对这一问题，人们设想专门为业务的统计分析建立一个数据中心，它的数据来自联机的事务处理系统、异构的外部数据源、脱机的历史业务数据等，这个数据中心就是数据仓库。

数据仓库技术的应运而生，成为信息技术领域非常热门的话题之一。

数据仓库技术的提出建立了一种体系化的数据存储环境，将分析决策所需要的大量数据从传统的操作环境中分离出来，使分散、不一致的操作数据转换成集成、统一的信息。

企业内不同单位、不同角色的成员都可以在此单一的环境之下，通过运用其中的数据与信息，发现全新的视野和新的问题，产生用于决策的新分析方法。

作为决策支持系统的重要组成部分，数据仓库为决策支持系统提供了分析决策所需的数据：OLAP的产生，则进一步增强了决策支持系统快速、一致和交互性的分析能力，它利用存储在数据仓库中的数据完成各种分析操作，并以直观易懂的形式将分析结果展现给决策分析人员；而数据挖掘是从大量数据中提取或“挖掘”知识，从而实现从“数据—信息—知识”的过程，为企业的管理阶层提供各种层次的决策支持。

## <<数据仓库与数据挖掘原理及应用>>

### 内容概要

本书全面深入地介绍了数据仓库、联机分析处理和数据挖掘的基本概念、基本方法和应用技术。全书分成三篇：数据仓库与OLAP篇的主要内容包括：数据仓库的基本概念、体系结构、模型设计、开发方法、ETL、元数据和数据集市，OLAP的基本概念、基本操作、数据模型和OLAP的实现及准则；数据挖掘与空间数据挖掘篇的主要内容包括关联分析方法、聚类分析技术、分类与预测方法、异常检测算法以及空间数据挖掘技术等；工具与实例篇介绍了数据挖掘工具及可视化、Cognos公司的BI主要产品和企业数据仓库系统构建。

本书可作为高等院校计算机软件与应用、信息科学等专业的学生学习数据仓库、OLAP及数据挖掘技术的实用教程或参考书，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、工程人员参考。

# <<数据仓库与数据挖掘原理及应用>>

## 书籍目录

第一篇 数据仓库与OLAP 第1章 数据仓库基本概念 1.1 从数据库到数据仓库 1.1.1 蜘蛛网问题 1.1.2 事务处理和分析处理数据环境的分离 1.2 什么是数据仓库 1.2.1 面向主题 1.2.2 集成 1.2.3 稳定性 1.2.4 随时间而变化 1.3 数据仓库与传统数据库的比较 1.3.1 两个系统的主要区别 1.3.2 两个系统的查询支持不同 1.3.3 两个系统数据组织模式示例比较 1.4 数据仓库的系统结构 1.4.1 三层数据仓库结构 1.4.2 数据仓库中的关键名词 1.5 数据仓库的数据组织 1.5.1 数据仓库的数据组织结构 1.5.2 数据粒度与数据分割 1.5.3 数据仓库的数据组织形式 1.5.4 数据仓库的数据追加和清理 1.6 小结 习题 第2章 数据仓库中的ETL和元数据 2.1 ETL 2.1.1 ETL概念 2.1.2 ETL作用 2.1.3 ETL工具 2.2 元数据 2.2.1 什么是元数据 2.2.2 元数据的标准化 2.2.3 数据仓库中的元数据管理 2.2.4 在数据仓库项目中使用元数据的建议 2.3 外部数据 2.3.1 外部数据和非结构化数据 2.3.2 元数据和外部数据 2.3.3 外部数据的存储 2.3.4 外部数据的管理 2.4 小结 习题 第3章 数据仓库模型设计及数据仓库建立 3.1 数据仓库的概念模型设计 3.1.1 E.R模型 3.1.2 面向对象的分析方法 3.2 数据仓库的逻辑模型设计 3.2.1 分析主题, 确定当前要装载的主题 3.2.2 确定数据粒度的选择 3.2.3 确定数据分割策略 3.2.4 增加导出字段 3.2.5 定义关系模式 3.2.6 定义记录系统 3.3 数据仓库的物理模型设计 3.3.1 索引策略 3.3.2 数据存储策略 3.4 数据仓库的建立过程 3.4.1 需求分析 3.4.2 数据路线 3.4.3 技术路线 3.4.4 应用路线 3.4.5 数据仓库部署 3.4.6 运行维护 3.5 提高数据仓库性能 3.6 小结 习题 第4章 联机分析处理 4.1 OLAP概念 4.1.1 什么是OLAP 4.1.2 OLAP的相关基本概念 4.1.3 OLAP和OLTP的区别 4.1.4 OLAP和数据仓库的区别 .....第二篇 数据挖掘与空间数据挖掘第三篇 工具与实例主要参考文献

章节摘录

第一篇 数据仓库与OLAP 第1章 数据仓库基本概念 计算机技术的迅速发展使得处理数据成为可能，这就推动了数据库技术的极大发展，但是面对不断增加如潮水般的数据，人们不再满足于数据库的查询功能，提出了深层次问题：能不能从数据中提取蕴藏于其中的知识为决策服务。就数据库技术而言已经显得无能为力了，这就急需有新的方法和技术来处理这些海量般的数据。在这种情况下，数据库逐步发展到了数据仓库。

世界上最早的数据仓库是NCR公司为全美、也是全世界最大的连锁超市集团Wal—Mart在1981年建立的，而最早将数据仓库提升到理论高度进行分析并提出数据仓库这个概念的则是著名学者W.H.Inrnon，他对数据仓库所下的定义是：数据仓库是一个面向主题的、集成的、稳定的、随时间变化的数据的集合，用于支持管理决策过程。

由此可见，数据仓库是一个综合的解决方案，主要用来帮助企业有关主管部门和业务人员做出更符合业务发展规律的决策。

1.1 从数据库到数据仓库 传统数据库以及联机事务处理(on-line transaction processing, OLTP)在日常的管理事务处理中获得了巨大的成功，但是对管理人员的决策分析要求却无法实现。因为，管理人员常常希望能够通过对组织中的大量数据进行分析，了解业务的发展趋势。而传统数据库只保留了当前的业务处理信息，缺乏决策分析所需要的大量的历史信息。为满足管理人员的决策分析需要，就需要在数据库的基础上产生适应决策分析的数据环境——数据仓库(data warehouse)。

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>