

<<数据仓库与数据挖掘的原理及应用>>

图书基本信息

书名：<<数据仓库与数据挖掘的原理及应用>>

13位ISBN编号：9787040230147

10位ISBN编号：7040230143

出版时间：2008-2

出版范围：高等教育

作者：李志刚

页数：340

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

随着计算方法和信息技术的不断发展，大量数据的产生和收集导致信息爆炸。

现代社会的竞争趋势要求对这些数据进行实时的和深层次的分析。

虽然现在已经出现更强大的存储系统和检索系统，但是使用者发现在分析所拥有的信息方面变得越来越困难。

数据仓库提供了容纳大量信息的场所，但它只有和数据挖掘技术相结合才能最终解决用户的困惑，使用户能够从繁杂的数据中找出真正有价值的信息和知识。

数据仓库可以加强企业对信息的管理能力，数据挖掘可以改善企业的经营状况，使企业的决策制定过程更加科学化和快速，为企业带来巨大的收益，增强企业的竞争优势。

数据仓库和数据挖掘是20世纪90年代中期兴起的决策支持新技术，它们是基于大规模数据库的决策支持系统的核心。

数据仓库是区别于数据库的一种新型数据存储形式，它是面向主题的、集成的、不可更新的、随时间不断变化的数据集，用以支持经营管理中的决策制定。

数据挖掘是从数据库中发现知识的核心技术，它从大量的数据中提取隐含的、人所未知的、可信而有效的知识。

数据挖掘能够对数据进行再分析，以期获得更加深入的了解。

它具有预测功能，可以通过已有数据预测发展趋势。

数据仓库与数据挖掘技术相结合，与现代的管理决策方法相结合，就能够使数据仓库在组织机构的经营管理决策中发挥巨大的作用。

我国数据挖掘技术的研究始于20世纪90年代，经过十几年的发展，这一领域目前正处于蓬勃发展时期。

但是由于数据仓库、数据挖掘技术都是数据处理与分析领域出现的新技术，大部分人把目光投向基于这两项技术的基础理论的研究，特别是具体技术和算法的实现，而忽略对数据仓库、数据挖掘理论与实践相结合的应用研究。

因此，笔者在结合科研项目的基础上，对数据仓库、数据挖掘技术进行较系统的研究，并将研究成果应用于财经、证券投资等领域；同时笔者在研究过程中不断地学习，既对原有的理论和实践进行总结，又不断地将所学到的知识运用到实践中去，丰富了原有的理论。

本书详细阐述数据仓库与数据挖掘的基本原理，系统而全面地介绍数据仓库与数据挖掘的概念、作用、算法以及应用领域、相关学科和发展趋势，并着重讨论数据仓库和数据挖掘在企业管理中的应用及构建策略。

基于SQL Senrer 2005介绍数据仓库与数据挖掘工具的操作和应用，并结合具体实例，阐述企业数据仓库和数据挖掘的实施过程。

<<数据仓库与数据挖掘的原理及应用>>

内容概要

《数据仓库与数据挖掘的原理及应用》详细阐述数据仓库与数据挖掘的基本原理，系统而全面地介绍数据仓库与数据挖掘的概念、作用、算法以及应用领域、相关学科和发展趋势，并着重讨论数据仓库和数据挖掘在企业管理中的应用及构建策略。

基于SQL Server 2005介绍数据仓库与数据挖掘工具的操作和应用，并结合具体实例，阐述企业数据仓库和数据挖掘的实施过程。

最后，以证券行业为对象提供一个数据挖掘的开发实例。

《数据仓库与数据挖掘的原理及应用》的指导思想是在系统阐述基本知识和基本理论的基础上，强调实际应用能力的培养，充分体现数据仓库和数据挖掘技术作为解决实际问题的方法和工具的特点。

《数据仓库与数据挖掘的原理及应用》既可以作为信息系统、电子商务、管理科学与工程、计算机应用、软件工程等专业的本科高年级和研究生教材，又可以作为从事竞争情报、信息管理、知识管理、战略管理和软科学的研究人员的参考资料。

书籍目录

第1章 数据仓库概述本章主要内容1.1 从数据库到数据仓库1.1.1 决策支持技术与数据库技术的发展1.1.2 数据仓库与数据库的区别1.2 数据仓库的概念与特点1.2.1 数据仓库概念1.2.2 面向主题1.2.3 数据的集成性1.2.4 数据的非易失性1.2.5 数据因时而变的特点1.3 数据仓库中的关键概念1.3.1 外部数据源1.3.2 数据抽取1.3.3 数据清洗1.3.4 数据转换1.3.5 数据加载1.3.6 元数据1.3.7 数据集市1.3.8 数据粒度1.4 数据仓库的数据组织1.4.1 数据仓库的数据组织结构1.4.2 数据粒度与数据分割1.4.3 数据仓库的数据组织形式1.4.4 数据仓库的数据追加与清理1.5 数据仓库与数据集市的关系1.5.1 数据集市的类型1.5.2 数据集市与数据仓库的区别1.5.3 数据集市的特点1.6 数据仓库体系结构1.6.1 数据仓库系统的层次结构1.6.2 数据仓库的构造模式1.7 操作数据存储ODS1.7.1 操作数据存储ODS的概念1.7.2 操作数据存储ODS的应用1.7.3 DB-ODS-DW三层体系结构1.7.4 ODS / DW、ODS / DB之比较习题一第2章 联机分析处理本章主要内容2.1 联机分析处理的概念2.1.1 OLAP的定义2.1.2 OLAP的相关基本概念2.1.3 OLAP与OLTP的关系及比较2.1.4 OLAP准则2.2 OLAP多维数据分析2.2.1 OLAP基本分析动作2.2.2 广义OLAP功能2.2.3 多维数据分析实例2.3 OLAP数据组织2.3.1 多维数据组织2.3.2 关系数据组织2.3.3 两种数据组织的比较2.3.4 HOLAP2.4 OLAP的体系结构与展现方式2.4.1 OLAP体系结构2.4.2 OLAP前端展现方式2.4.3 OLAP结果的展现方法2.5 OLAP工具及评价2.5.1 OracleOLAP工具2.5.2 OLAP服务器和工具的评价指标2.5.3 OLAP的局限性习题二第3章 数据仓库设计本章主要内容3.1 仓库中数据模型概述3.1.1 数据模型的概念3.1.2 数据仓库模型的构建原则3.1.3 企业数据模型3.2 概念模型设计3.2.1 企业模型的建立3.2.2 数据模型的规范化3.2.3 常见的概念模型3.3 逻辑模型设计3.3.1 概念模型到逻辑模型的转换3.3.2 数据表的规范化与分割3.3.3 维度表的设计3.3.4 事实表的设计3.3.5 数据集市的设计3.4 物理模型设计3.4.1 定义数据存储结构3.4.2 索引策略3.4.3 存储分配优化3.4.4 数据加载设计3.4.5 物理模型的设计对数据仓库性能的影响3.5 元数据模型3.5.1 元数据的类型3.5.2 元数据的作用3.5.3 元数据的收集与维护3.5.4 元数据的使用3.6 粒度模型3.6.1 粒度的划分3.6.2 粒度级别的确定习题三第4章 数据仓库的规划与开发本章主要内容4.1 数据仓库的投资分析4.1.1 建立数据仓库的必要性4.1.2 数据仓库的投资回报分析与风险分析4.2 数据仓库的开发方法4.2.1 瀑布式开发4.2.2 螺旋式开发4.3 数据仓库的建立过程4.3.1 数据进入数据仓库的过程与建立数据仓库的步骤4.3.2 需求分析4.3.3 数据路线4.3.4 技术路线4.3.5 应用路线4.3.6 数据仓库部署4.3.7 运行维护4.4 数据仓库的维护4.4.1 数据周期4.4.2 参照完整性4.4.3 数据环境信息4.4.4 数据备份与恢复4.5 提高数据仓库性能4.5.1 提高I/O性能4.5.2 缩小查询范围4.5.3 采取并行优化技术4.5.4 选择适当的初始化参数4.6 数据仓库的安全性4.6.1 安全类型4.6.2 安全方法4.7 分布式数据仓库4.7.1 分布式数据仓库的优点4.7.2 分布式数据仓库的模型建立与数据划分4.7.3 分布式数据仓库的建设策略4.7.4 分布式数据仓库的技术缺陷习题四第5章 数据仓库的工具本章主要内容5.1 数据仓库工具的选择5.1.1 数据仓库工具的组成5.1.2 数据仓库工具应具备的主要功能5.1.3 数据仓库的发展趋势5.1.4 选取数据仓库工具的方法5.2 常用数据仓库产品简介5.2.1 Oracle9i5.2.2 NCR TeraData第6章 数据挖掘概述第7章 数据挖掘的算法第8章 数据挖掘新技术第9章 数据挖掘的工具及其应用第10章 数据仓库与数据挖掘的综合应用第11章 基于数据挖掘的上市公司财务危机预警应用实例参考文献

章节摘录

插图：在决策过程中经常用到外部数据，这些数据通常也是非结构化的。

在事务处理系统中，由于未能对外部数据进行统一管理，用到这些数据的DSS应用必须对其自行集成。

（3）数据动态集成问题由于每次分析处理时都进行数据集成的开销太大，某些应用仅在开始时对所需数据进行集成，以后就一直以这部分数据作为分析处理的基础，不再与数据源发生联系，这种方式的集成是静态集成。

但是，数据源中的数据如果在数据集成后发生变化，这些改变未能反映给决策者，这将导致决策者使用过时的数据。

对于决策者而言，虽然并不要求实时准确地知道系统内的任何数据变化，但也不希望所分析的是很久以前的数据。

因此，集成数据必须以一定的周期（如24小时）进行刷新，我们称其为动态集成。

显然，事务处理系统并不具备动态集成的能力。

（4）历史数据问题事务处理通常只需要当前数据，在数据库中一般也只存储短期数据，且不同数据的保存期限也不一样。

即使有一些历史数据保存下来了，也被束之高阁，未能得到充分利用。

但对于决策分析而言，历史数据是相当重要的，许多分析方法必须以大量的历史数据为依托。

不对历史数据进行详细分析，是难以把握企业的发展趋势的。

（5）数据的综合问题在事务处理系统中积累了大量的细节数据，一般而言，DSS并不对这些细节数据进行分析。

这主要存在两个原因，一是细节数据的数据量太大，会严重影响分析数据的效率；二是过多的细节数据不利于分析人员将注意力集中在有用的信息上。

因此，在进行分析之前，往往需要对细节数据进行不同程度的综合。

而事务处理系统不具备这种综合能力，根据规范化理论，这种综合数据的过程还往往因为会产生数据冗余而被加以限制。

市场商业经营模式的改变以及市场竞争日趋激烈都要求捕获和分析事务级的业务数据。

建立在事务处理环境上的分析系统无法达到这一要求。

要提高分析数据和做出决策的效率和有效性，分析型处理及其数据必须与操作型处理及其数据相分离，必须把分析型数据从事务处理环境中提取出来，按照DSS处理的需要重新组织，建立单独的分析处理环境，数据仓库正是为了构建这种新的分析处理环境而出现的一种数据存储和组织技术。

表1.2 是以数据库为代表的事务处理环境与以数据仓库为代表的分析处理环境的对比。

编辑推荐

《数据仓库与数据挖掘的原理及应用》的目的旨在向读者系统阐述数据仓库与数据挖掘的基本原理、方法和实用工具，介绍国内外的最新研究成果。

全书共有11章，第1章介绍数据仓库的基本概念和知识；第2章介绍联机分析处理的基本理论；第3章介绍数据仓库的设计思想、方法和技巧；第4章介绍数据仓库的规划与开发；第5章介绍各种数据仓库工具的基本功能及SQL Server 2005数据仓库工具的应用；第6章介绍数据挖掘的概念和相关知识；第7章介绍数据挖掘的算法；第8章介绍文本挖掘、Web挖掘等数据挖掘新技术；第9章介绍数据挖掘的工具及其应用；第10章介绍数据仓库与数据挖掘的综合应用；第11章介绍基于数据挖掘的上市公司财务危机预警应用实例，使读者能结合具体应用进行上机操作，消化和理解所学的知识。

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>