

<<数据仓库与数据分析教程>>

图书基本信息

书名：<<数据仓库与数据分析教程>>

13位ISBN编号：9787040341300

10位ISBN编号：7040341301

出版时间：2012-8

出版时间：王珊、李翠平 高等教育出版社 (2012-08出版)

作者：王珊,李翠平

页数：222

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>

<<数据仓库与数据分析教程>>

内容概要

数据仓库技术和数据分析技术是信息领域的核心技术之一，是基于海量数据的决策支持系统体系化环境的核心。

《普通高等教育“十一五”国家级规划教材：数据仓库与数据分析教程》详尽地介绍了数据仓库和数据分析技术的基本概念和基本原理，建立数据仓库和进行数据分析的方法和过程。

全书分为数据仓库技术篇、联机分析处理技术篇、数据挖掘技术篇三部分，共10章。

附录中介绍了一些典型的数据仓库产品和工具。

《普通高等教育“十一五”国家级规划教材：数据仓库与数据分析教程》可以作为高等学校计算机专业、信息管理专业以及其他相关专业本科生和研究生的教材和参考书，也可以作为企事业单位信息管理部门及相关行业从事数据库和数据仓库的研究与开发人员、数据分析人员和管理人员的参考资料。

<<数据仓库与数据分析教程>>

书籍目录

第一篇数据仓库技术 第一章从数据库到数据仓库 1.1数据仓库产生的原因 1.1.1操作型数据处理 1.1.2分析型数据处理 1.1.3两种数据处理模式的差别 1.1.4数据库系统的局限性 1.2数据仓库的基本概念 1.2.1主题与面向主题 1.2.2数据仓库的其他三个特征 1.2.3数据仓库的功能 1.3数据仓库的体系结构 1.3.1体系结构 1.3.2数据集市 小结 习题 第二章操作数据存储 2.1什么是ODS 2.1.1ODS的定义及特点 2.1.2ODS的功能和实现机制 2.2DB ~ ODS ~ DW体系结构 2.2.1ODS与DW 2.2.2DB ~ ODS ~ DW三层体系结构 小结 习题 第三章数据仓库中的数据及组织 3.1数据仓库中的数据组织 3.2数据仓库中数据的追加 3.3数据仓库中的元数据 3.3.1元数据的定义 3.3.2元数据的分类 3.3.3元数据管理的标准化 小结 习题 第二篇联机分析处理技术 第四章概述及模型 4.1OLAP技术概述 4.1.1OLAP的起源 4.1.2OLAP的定义 4.1.3OLAP与OLTP的区别 4.1.4OLAP核心技术 4.2多维数据模型 4.2.1基本概念 4.2.2星形、雪片和事实群模型 4.3多维分析操作 4.3.1多维分析基础：聚集 4.3.2常用多维分析操作 4.3.3其他多维分析操作 4.3.4聚集的一些限制 4.3.5水平层次结构和非水平层次结构 4.4多维查询语言 4.4.1MDX简介 4.4.2MDX对象模型 4.5多维数据展示 4.5.1三维数据展示 4.5.2高维数据展示 小结 习题 第五章数据方体的存储、预计算和缩减 5.1数据方体的存储 5.1.1MOLAP 5.1.2ROLAP 5.1.3MOLAP和ROLAP实现机制的比较 5.2数据方体的预计算 5.2.1预计算的相关概念 5.2.2数据方体格结构 5.2.3数据方体格存储方法 5.3完整数据方体的预计算方法 5.3.1流水线算法 5.3.2BUC算法 5.4部分数据方体的预计算方法 5.4.1BPUS算法 5.4.2PBS算法 5.5数据方体的缩减技术 5.5.1Drawf数据方体 5.5.2Condensed数据方体 5.5.3Quotient数据方体 小结 习题 第六章数据方体的索引、查询和维护 6.1数据方体的索引技术 6.1.1树索引 6.1.2位图索引 6.2数据方体的查询处理和优化技术 6.2.1子查询划分技术 6.2.2子查询处理及优化技术 6.3数据方体的维护技术 小结 习题 第三篇数据挖掘技术 第七章数据挖掘概述 7.1数据挖掘简介 7.1.1数据挖掘的特点 7.1.2数据挖掘与KDD 7.1.3数据挖掘与OLAP 7.1.4数据挖掘与数据仓库 7.1.5数据挖掘的分类 7.1.6数据挖掘的应用 7.2数据挖掘算法的组件化思想 7.2.1模型或模式结构 7.2.2数据挖掘的任务 7.2.3评分函数 7.2.4搜索和优化方法 7.2.5数据管理策略 7.2.6组件化思想的应用 小结 习题 第八章频繁模式挖掘 8.1频繁项集和关联规则 8.1.1问题描述 8.1.2关联规则分类 8.1.3关联规则挖掘的经典算法Apriori 8.1.4关联规则挖掘的重要算法FP—Growth 8.1.5其他关联规则挖掘方法 8.1.6关联规则的兴趣度 8.2序列模式挖掘 8.2.1问题描述 8.2.2GSP算法 8.2.3PrefixSpan算法 8.3频繁子图挖掘 8.3.1问题描述 8.3.2基于Apriori的宽度优先算法 8.3.3基于FP—Growth的深度优先搜索算法 小结 习题 第九章预测建模：分类和回归 9.1预测建模简介 9.1.1预测的模型结构 9.1.2用于预测的评分函数 9.1.3用于预测的搜索和优化策略 9.2决策树分类 9.2.1建树阶段 9.2.2剪枝阶段 9.2.3分类规则的生成 9.2.4可扩展性问题 9.2.5其他问题 9.3贝叶斯分类 9.3.1基本概念 9.3.2朴素贝叶斯分类 9.4支持向量机分类 9.4.1线性可分时的二元分类问题 9.4.2线性不可分时的二元分类问题 9.4.3多元分类问题 9.4.4可扩展性问题 9.5人工神经网络分类 9.5.1神经网络的组成 9.5.2神经网络的分类方法 小结 习题 第十章描述建模：聚类 10.1聚类分析简介 10.1.1对象间的相似性 10.1.2其他相似性度量 10.2聚类方法概述 10.2.1基于划分的聚类方法 10.2.2基于密度的聚类方法 10.2.3基于层次的聚类方法 10.2.4.基于模型的聚类方法 10.2.5基于方格的聚类方法 小结 习题 附录产品与工具 附录AIBM数据仓库解决方案 附录BOracle数据仓库解决方案 附录CMicrosoftSQLServer2005数据仓库解决方案 附录DSybase数据仓库解决方案 附录EGroup1Sagem介绍 附录FInformatica介绍 参考文献

<<数据仓库与数据分析教程>>

章节摘录

版权页：插图：2.数据存储 如前所述，MOLAP以多维数组为主要存储结构，聚集计算通过数组下标的直接偏移进行。

ROLAP以传统的关系数据库系统为基础，以关系表为主要存储结构。

在数据的存储容量上，由于关系数据库的技术较为成熟，因此ROLAP占优势，并且可以支持的维数也较MOLAP多。

但值得注意的是，限制MOLAP数据量的不是维数，而是数据单元数。

如果MOLAP能在数据单元的存储管理上有进一步的提高，再辅之以高效的稀疏处理能力，其数据量也可以达到很大。

另一点需要注意的是，尽管ROLAP的数据容量大，但为了提高分析响应速度，须构造大量的中间表（即预综合，以避免重复连接），因此数据冗余度也大。

3.数据存取 在前面已经提到，由于ROLAP是用关系表来模拟多维数据，因此其存取较MOLAP复杂。首先用户的分析请求由ROLAP服务器转为SQL请求，然后交由RDBMS处理，处理结果经多维处理后返回给用户，而且SQL并非可以处理所有的多维分析和计算工作，有些时候只能依赖附加的应用程序来完成。

而MOLAP可以利用多维查询语言（如MDX）或以其他方式直接将用户查询转为MOLAP可以处理的形式，基本不借助附加程序。

4.适应性 可以从以下几个不同方面对MOLAP和ROLAP的适应性进行比较。

（1）适应分析维数动态变化 由于MOLAP的预综合度相当高（85%以上），因而增加一维则数据方体的规模会迅猛增长。

而ROLAP的预综合度相当灵活，大多根据用户需要进行，一般在85%以下，增加一维意味着增加一些维表及与用户分析相关的综合表，还有事实表中的相应内容，相对来说比较容易。

（2）适应数据变化 同样，因为MOLAP的预综合度高，因此当数据或计算频繁变化时，其重新计算量相当大，甚至需要重新构建多维数据库。

相比而言，ROLAP的预综合度低，适应数据变化的范围大。

（3）适应海量数据 由于RDBMS已有20多年的历史，其技术上比较成熟，加之近年来并行处理技术的发展和运用，ROLAP在适应海量数据上的能力强于MOLAP。

（4）适应软硬件的能力 理由同上，ROLAP在软硬件上的适应力明显强于MOLAP。

尽管ROLAP在适应性方面明显强于MOLAP，但这种差距是历史造成的。

可以预见，随着时间的推移，MOLAP的技术会不断成熟，像并行处理等RDBMS上用到的技术也会逐渐用到MOLAP上来。

<<数据仓库与数据分析教程>>

编辑推荐

《普通高等教育"十一五"国家级规划教材:数据仓库与数据分析教程》可以作为高等学校计算机专业、信息管理专业以及其他相关专业本科生和研究生的教材和参考书,也可以作为企事业单位信息管理部门及相关行业从事数据库和数据仓库的研究与开发人员、数据分析人员和管理人员的参考资料。

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>