

<<数据仓库原理与实践>>

图书基本信息

书名：<<数据仓库原理与实践>>

13位ISBN编号：9787115100443

10位ISBN编号：7115100446

出版时间：2003-1-1

出版单位：人民邮电出版社

作者：林宇

页数：360

字数：563

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>

<<数据仓库原理与实践>>

内容概要

本书比较全面系统地介绍了数据仓库（Data Warehouse）、联机分析处理（OLAP）、数据挖掘（Data Mining）等3个层次的基本概念、原理和应用技术。

全书分成4篇，基本原理篇和设计建模篇的内容主要包括：数据仓库的基本概念、体系结构、创建过程、建模设计、项目规划，OLAP的基本概念、ROLAP和MOLAP的实现原理、OLAP模型设计，数据挖掘的基本概念、基本过程、常见模型和算法。

产品介绍篇介绍了现有数据仓库厂商产品工具的基本情况，并对产品选择进行了一些分析。

应用实践篇结合电信领域的实例，介绍了数据仓库项目在设计和实施中的关键问题。

本书的编写以理论联系实际为原则，内容系统全面，对于从事数据仓库研究、设计、开发等工作的人员具有宝贵的参考价值，对于需要了解数据仓库技术的系统集成人员、系统分析师、系统设计师也具有一定的参考价值。

<<数据仓库原理与实践>>

书籍目录

第一篇 基本原理篇

第1章 数据仓库概述 2

1.1 数据库到数据仓库的演变 2

1.1.1 蜘蛛网问题 2

1.1.2 操作型系统和分析型系统的分离 5

1.2 商业智能系统的功能和构成 7

1.2.1 商业智能系统的数据处理循环 7

1.2.2 决策支持系统的总体功能 7

1.3 仓库的应用前景 9

第2章 数据仓库的基本原理 12

2.1 数据仓库的体系结构 12

2.1.1 数据仓库的体系结构 12

2.1.2 数据仓库中的关键名词 13

2.2 数据仓库的特点 16

2.2.1 数据库的功能和特征 16

2.2.2 数据仓库的功能和特征 17

2.2.3 面向主题 17

2.2.4 数据的集成性 19

2.2.5 数据的稳定性 21

2.2.6 数据随时间变化的特点 22

2.3 数据仓库的数据组织 23

2.3.1 数据仓库的数据组织结构 23

2.3.2 数据颗粒度 25

2.3.3 数据的分割 29

2.3.4 数据仓库的数据组织形式 30

2.3.5 数据仓库的数据追加技术 32

2.3.6 清理数据仓库的数据 36

2.4 数据仓库建设的两条技术路线 36

2.5 操作数据存储ODS 38

2.5.1 ODS的概念 38

2.5.2 ODS的应用 39

2.5.3 DB - ODS - DW的3层体系结构 41

2.5.4 ODS/DW、ODS/DB间的比较 43

2.6 外部数据和非结构数据 44

2.6.1 外部数据的特征 44

2.6.2 为什么将外部数据放在数据仓库 45

2.6.3 对外部数据进行管理的元数据 46

2.6.4 外部数据/非结构化数据的存储 47

2.6.5 外部数据的使用 49

第3章 OLAP的基本原理 50

3.1 OLAP的基本概念 50

3.1.1 OLAP的基本概念 50

3.1.2 OLAP的基本分析动作 53

3.1.3 OLAP的展现方式 57

3.1.4 OLAP和OLTP 59

<<数据仓库原理与实践>>

3.1.5 OLAP的体系结构和分类	60
3.2 以多维数据库为基础的OLAP服务器	61
3.2.1 多维数据库 (Multi-Dimensional Database)	61
3.2.2 MDDB产品实例	64
3.2.3 MOLAP产品的结构	66
3.3 基于关系型数据库的OLAP (ROLAP)	66
3.3.1 维表	67
3.3.2 事实表	67
3.3.3 星型结构	69
3.3.4 ROLAP和MOLAP的比较	73
3.3.5 HOLAP (Hybrid OLAP)	76
3.4 OLAP的特征和衡量	76
3.4.1 OLAP的12准则	76
3.4.2 OLAP的简洁准则	79
3.5 OLAP的前端展现方式	80
3.5.1 OLAP的C/S方式	80
3.5.2 OLAP的Web	80
3.5.3 瘦客户机方式	81
3.5.4 OLAP的局限性	82
第4章 数据挖掘基础	83
4.1 数据挖掘的概念	83
4.1.1 什么是数据挖掘	83
4.1.2 数据挖掘的形式化定义	84
4.1.3 数据挖掘的基本步骤	85
4.1.4 数据挖掘方法的分类	85
4.2 描述性挖掘分析	87
4.2.1 关联规则	87
4.2.2 序列模型分析	92
4.2.3 聚类分析 (Clustering)	93
4.3 预测类的挖掘算法	98
4.3.1 分类问题	99
4.3.2 回归问题	102
4.3.3 时间序列问题	102
4.3.4 神经网络	105
4.3.5 决策树分析	110
4.4 数据挖掘的体系结构	112
4.5 数据挖掘系统的应用实例	113
4.5.1 从用户的行为模式来自动地生成动态链接	113
4.5.2 用数据挖掘的方法来解决网络拥挤的问题	119
4.6 数据挖掘和相关系统的联系	120
4.6.1 数据挖掘和人工智能、统计学的关系	120
4.6.2 数据挖掘和数据仓库的关系	121
4.6.3 数据挖掘和OLAP的区别和联系	122
4.6.4 数据挖掘不是万能的	124
第二篇 设计建模篇	
第5章 企业模型设计	126
5.1 数据仓库设计和数据库设计的区别	126

<<数据仓库原理与实践>>

5.2 企业模型	129
5.2.1 什么是企业模型	129
5.2.2 面向对象的分析方法	133
5.2.3 使用面向对象的方法建立企业模型	139
5.2.4 面向对象方法和ER模型的对比	144
5.3 企业模型到数据库模型的映射	145
5.3.1 限定集成的范围	145
5.3.2 映射到关系模型	146
5.3.3 对比映射结果和现有系统	148
5.4 将企业模型映射到数据仓库概念模型	149
第6章 数据仓库的模型设计	152
6.1 数据仓库的逻辑模型设计	152
6.1.1 系统数据量估算	152
6.1.2 数据颗粒度的选择	153
6.1.3 表的分割	157
6.1.4 增加时间字段	157
6.1.5 去除纯操作型数据	157
6.1.6 合理的表划分	158
6.1.7 定义关系模式	161
6.1.8 增加导出字段	161
6.1.9 记录系统的定义	162
6.2 数据仓库的物理模型设计	163
6.2.1 确定数据的存储结构	163
6.2.2 索引策略	166
6.2.3 数据存储策略	173
6.2.4 存储分配优化	176
6.3 数据装载接口设计	176
6.4 并行优化	177
6.4.1 数据仓库中并行优化的必要性和可能性	177
6.4.2 并行体系结构	179
第7章 OLAP建模方法	183
7.1 维表	183
7.1.1 维表的特征	183
7.1.2 维的变化	184
7.1.3 维表的共享	187
7.1.4 雪花型结构处理多对多关系	189
7.1.5 层次信息和分类信息的位置	191
7.1.6 非分析数据的分离	194
7.1.7 典型的维层次	195
7.2 事实表	198
7.2.1 事实	198
7.2.2 事实表的特征	199
7.2.3 数据的粒度	199
7.2.4 聚合操作和聚合表	200
7.2.5 没有度量变量的事实表	201
7.2.6 通用数据和专用数据事实表	202
7.3 OLAP的并行优化	204

<<数据仓库原理与实践>>

7.3.1 B-TREE索引、位图索引和Bit-Wise索引	204
7.3.2 星型查询优化	204
7.3.3 预连接技术	207
7.4 多维数据库	207
第8章 数据仓库规划和开发方法	210
8.1 数据仓库的投资分析	210
8.1.1 数据仓库的应用目标	210
8.1.2 建设数据仓库的必要性	211
8.1.3 数据仓库的投资回报分析	211
8.1.4 技术选择分析	212
8.1.5 IDC的统计结果	213
8.2 数据仓库的开发方法	213
8.2.1 瀑布式开发	213
8.2.2 螺旋式开发	214
8.3 数据仓库主题的选择和阶段规划	216
8.3.1 阶段规划的原则	216
8.3.2 维护阶段	216
8.4 现有数据库系统的改造问题	216
8.5 数据仓库和数据库系统的相互作用	217
8.5.1 相互促进的过程	217
8.5.2 解决“蜘蛛网”问题	218
8.5.3 数据仓库的“间接使用”	218
8.6 分布式数据仓库	219
8.6.1 采用分布式数据仓库的原因	219
8.6.2 分布式下的模型建立和数据划分	221
8.6.3 分布式数据仓库的建设策略	224
8.6.4 分布式数据仓库技术的缺点	227
8.7 需要避免的错误	228
第9章 数据挖掘的实施过程	231
9.1 数据挖掘过程模型5A	231
9.2 数据挖掘过程模型CRISP-DM	233
9.3 数据挖掘过程中的相关问题	235
9.3.1 定义商业问题	235
9.3.2 建立数据挖掘库	236
9.3.3 分析数据/选择变量	241
9.3.4 模型训练方法	245
9.3.5 数据挖掘模型的评价方法	248
9.3.6 数据仓库的实施和维护	250
9.3.7 模型实例	250
第10章 数据仓库的建立和维护	252
10.1 数据仓库建立的过程	252
10.1.1 建立企业模型	252
10.1.2 阶段规划/主题选取	253
10.1.3 技术准备工作	253
10.1.4 逻辑设计	254
10.1.5 物理设计	255
10.1.6 数据载入接口设计	255

<<数据仓库原理与实践>>

10.1.7 装载一个主题的数据和数据校验	256
10.1.8 OLAP模型设计和应用开发	256
10.1.9 数据准备程序设计	257
10.1.10 数据挖掘模型设计	257
10.1.11 界面系统设计	258
10.1.12 装载其他主题数据	258
10.1.13 同客户交流	259
10.1.14 重新开始循环	260
10.2 数据仓库的维护工作	260
10.2.1 数据周期	260
10.2.2 参照完整性	261
10.2.3 数据环境信息	262
第三篇 产品介绍篇	
第11章 数据仓库产品的介绍	266
11.1 数据仓库工具	266
11.2 INFORMIX数据仓库产品简介	268
11.2.1 INFORMIX数据仓库解决方案	268
11.2.2 数据抽取、转换和装载	269
11.2.3 数据存储	270
11.2.4 数据访问/呈现	274
11.3 SAS产品简介	278
第12章 数据仓库产品的选择	283
12.1 数据仓库产品应具备的关键技术	283
12.2 各数据仓库厂商产品的比较	285
12.3 数据仓库工具的选择	286
12.4 数据仓库工具的互通问题	287
第四篇 应用实践篇	
第13章 项目的需求和目标分析	292
13.1 电信领域建立数据仓库的常见主题	292
13.2 电信领域常见的分析问题	295
13.2.1 客户群体划分	295
13.2.2 客户流失分析	296
13.2.3 客户欺诈分析	296
13.2.4 网络规划优化	297
13.2.5 网管中的分析问题	298
13.3 项目规划	301
13.4 需求分析的形成	301
13.4.1 任务说明书	301
13.4.2 需求说明书	302
第14章 系统结构和模型设计	307
14.1 系统结构设计	307
14.1.1 数据量的估算	307
14.1.2 系统硬件结构/软件结构选择	307
14.2 数据仓库模型的设计	310
14.2.1 可利用的数据	310
14.2.2 粒度的确定	311
14.3 OLAP模型设计	314

<<数据仓库原理与实践>>

14.3.1 项目涉及的维度分析	315
14.3.2 各个主题的维度设计	317
第15章 系统装载、数据挖掘和界面设计	324
15.1 数据装载/数据综合模块设计	324
15.2 OLAP模型生成程序	329
15.3 数据挖掘宽表设计和生成	335
15.3.1 确定同目标变量相关的数据	335
15.3.2 创建新变量	336
15.3.3 准备训练集合与验证集合	342
15.3.4 确定分析的次序	343
15.3.5 变量选择	343
15.3.6 模型的维护和完善	344
15.4 创建多维数据库模块设计	345
第16章 界面设计和项目总结	348
16.1 界面展现设计	348
16.1.1 三层体系结构	348
16.1.2 按照内容对界面进行规划	349
16.2 系统的工作流程总结	351
16.3 系统的数据流程总结	353
16.4 系统的模块组成	355
附录 常用名词表	356

<<数据仓库原理与实践>>

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>