

第四章 数据仓库与联机分析处理

笔记本： 数据挖掘：概念与技术

创建时间： 2017/12/20 15:42

更新时间： 2017/12/26 11:30

作者： Passero

4.1 数据仓库：基本概念

什么是数据仓库

宽泛的讲，数据仓库是一种数据库，它与单位的操作数据库分别维护。数据仓库系统允许将各种应用系统集成在一起，为统一的历史数据分析提供坚实的平台，为信息处理提供支持。

“数据仓库是一个面向主题的，集成的，时变的，非易失的数据集合，支持管理者的决策过程”

数据仓库主要特征：面向主题的，集成的，时变的，非易失的

- 面向主题的（subject-oriented）：数据仓库围绕一些重要主题，如顾客、供应商、产品和服务组织。数据仓库关注决策者的数据建模与分析，而不是单位的日常操作和事务处理。因此，数据仓库通常排除对于决策无用的数据，提供特定主题的简明视图。
 - 集成的（integrated）：通常，构造数据仓库是将多个异构数据源，如关系数据库、一般文件和联机事务处理记录集成在一起。使用数据清理和数据集成技术，确保命名约定、编码结构、属性度量等的一致性。
 - 时变的（time-variant）：数据存储从历史的角度（例如，过去 5 ~ 10 年）提供信息。数据仓库中的关键结构都隐式或显式地包含时间元素。
 - 非易失的（nonvolatile）：数据仓库总是物理地分离存放数据，这些数据源于操作环境下的应用数据。由于这种分离，数据仓库不需要事务处理、恢复和并发控制机制。通常，它只需要两种数据访问操作：数据的初始化装入和数据访问。
-

操作数据库系统与数据仓库的区别

为什么需要分离的数据仓库

数据仓库：一种多层体系结构

数据仓库模型：企业仓库、数据集市和虚拟仓库

数据提取、变换和装入

元数据库

4.2 数据仓库建模：数据立方体与OLAP

数据立方体：一种多维数据模型

星形、雪花形和事实星座：多维数据模型的模式

维：概念分层的作用

度量的分类和计算

典型的OLAP操作

查询多维数据库的星网查询模型

4.3数据仓库的设计与使用

数据仓库的设计的商务分析框架

数据仓库的设计过程

数据仓库用于信息处理

从联机分析处理到多维数据挖掘

4.4数据仓库的发现

数据立方体的有效计算：概述

索引|OLAP数据：位图索引和连接索引|

OLAP查询的有效处理

OLAP服务器结构：ROLAP、MOLAP、HOLAP的比较

4.5数据泛化：面向属性的归纳

数据特征的面向属性的归纳

面向属性归纳的有效实现

类比较的面向属性归纳
