



石家莊鐵道大學
SHIJIAZHUANG TIEDAO UNIVERSITY

在线开放课程

管理信息系统技术基础

数据仓库与数据 挖掘

主讲：王兴鹏

目录

- 数据仓库
- 数据挖掘



在线开放课程

一、数据仓库

1. 数据仓库的定义

- 数据仓库（Data Warehouse）是支持管理决策过程的、面向主题的、集成的、随时间变化的、但信息本身是相对稳定的数据集合。
- DW的最终目标是把企业范围内的所有数据集成起来中，让用户能运行查询、产生报告、执行分析。

一、数据仓库

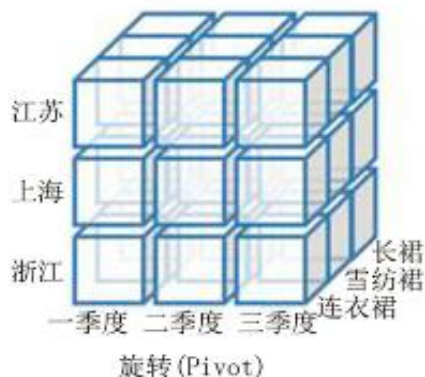
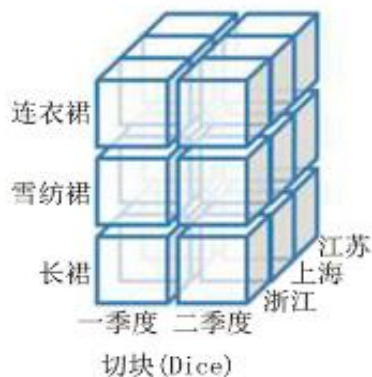
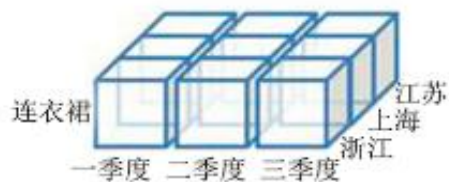
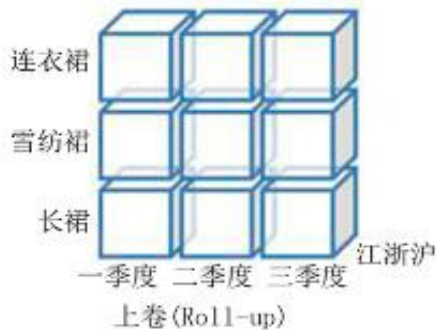
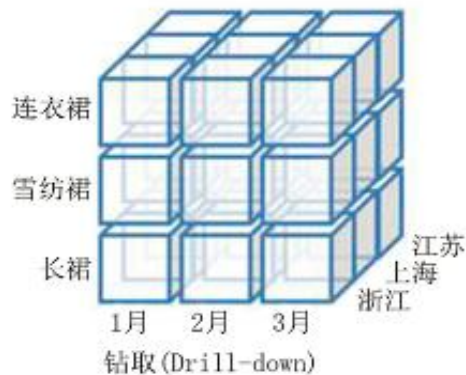


在线开放课程

2. 数据仓库的特征：

- 多源数据集成
- 多维数据
- 支持决策而不是事务处理 (OLAP vs. OLTP)

多维数据的例子



一、数据仓库

3. 数据仓库和操作数据库的比较

对比内容	数 据 库	数 据 仓 库
数据内容	当前值	历史的、存档的、归纳的、计算的数据
数据目标	面向业务操作程序，重复处理	面向主题域，分析应用
数据特性	动态变化，按字段更新	静态、不能直接更新，只能定时添加、刷新
使用频率\访问量	高，少量记录	低，海量记录
响应时间	以秒为单位	时间长
主要用户	业务人员和管理人员	高层决策人员

一、数据仓库

4. 数据集市 (Data Marts)

- 是一种更小、更集中的数据仓库，它是面向特定应用对象的数据仓库。
- 为公司分析提供了一种相对廉价的解决方案，又称为部门级的数据仓库。



一、数据仓库

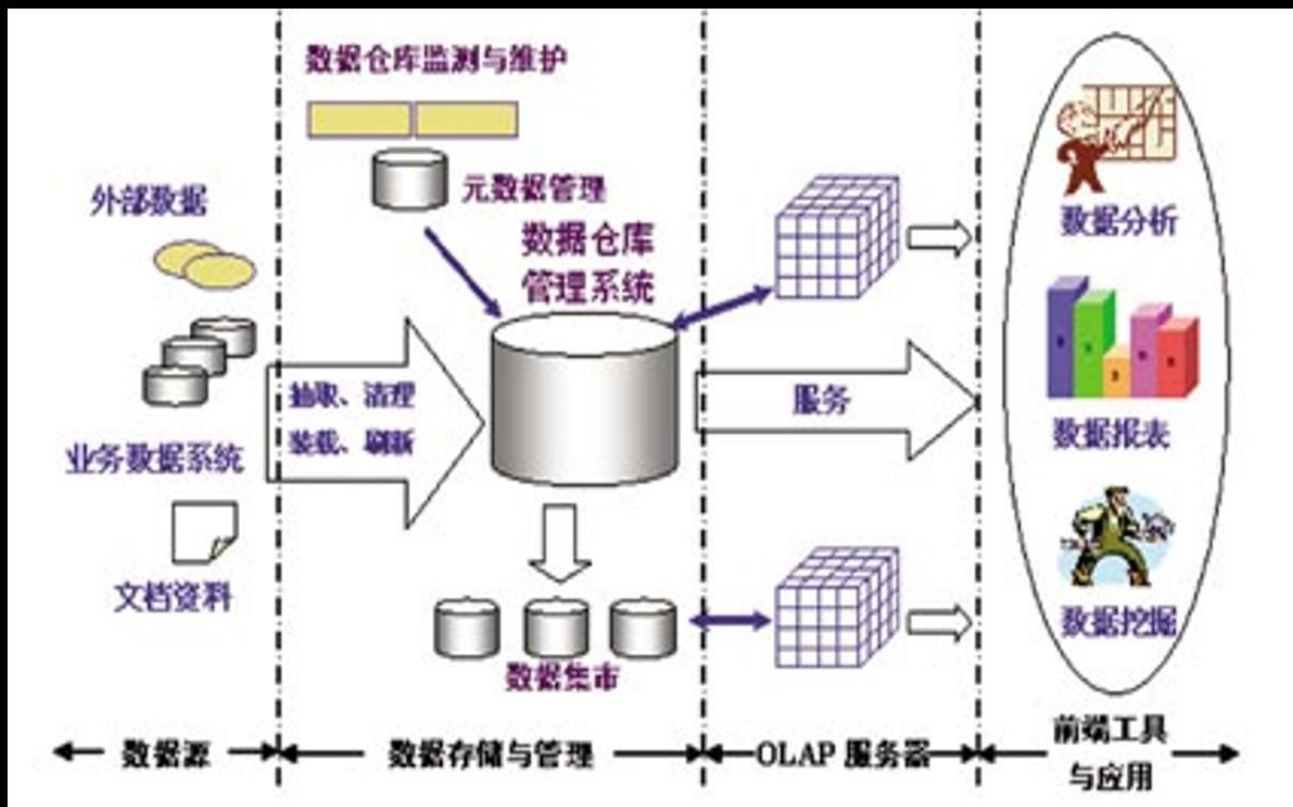


在线开放课程

4. 数据集市的特点

- (1) 规模小，组织灵活；
- (2) 能快速实现；
- (3) 价格相对低廉；
- (3) 集成的工具集。

数据仓库的体系结构



二、数据挖掘

1. 数据挖掘的定义

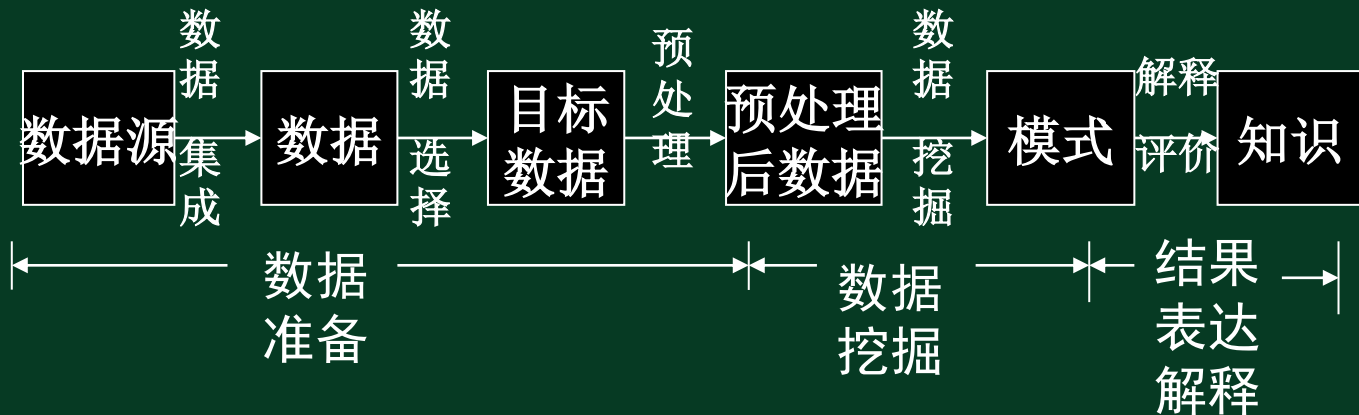
- “啤酒和纸尿裤的故事”
- 数据挖掘是从大量的、不完全的、模糊的、随机的数据中抽取有效的、新颖的和潜在有用的知识的过程。

二、数据挖掘

2. 数据挖掘的过程

DM的过程一般由三个主要的阶段组成：

数据准备、挖掘操作、结果表达和解释。



二、数据挖掘



在线开放课程

3. 常用的数据挖掘技术

- 关联分析
- 分类分析
- 聚类分析
- 时间序列分析

二、数据挖掘

3. 数据挖掘的应用

- 在商业经营方面
- 在金融保险方面
- 在客户关系管理方面
- 在移动通信方面
- 其他：公共管理、政府、教育等

小结

- 数据仓库的概念
- 数据挖掘的概念



在线开放课程

