

数据挖掘与数据仓库_3

1、数据库与数据仓库对比

数据库	数据仓库
细节的	综合或提炼的
在存取时准确的	代表过去的数据
可更新的	不更新
一次操作数据量小	一次操作数据量大
面向应用	面向分析
支持管理	支持决策
用于事务处理	主要用于分析数据

2、DW（数据仓库）、ODS（操作型数据存储）、DM（数据集市）属性

DW（数据仓库）（侧重于计算能力）	ODS（操作型数据存储）（能够实时报表分析）
A、面向主题的	A、面向主题的
B、集成的	B、集成的
C、时变的	C、当前的
D、非易失的/稳定的	D、易失的

DM（数据集市）（Data Mart）

独立的数据集市

从属的数据集市

A、面向部门的数据仓库

B、为满足用户特定需求而创建的数据仓库

C、数据仓库的子集

附：

- Data Lake（数据湖）**（侧重于数据存储能力）
- A、保存原始的、细节的数据
 - B、结构化数据、非结构化数据均存在（多模态）

发展趋势：

一体化：比如湖仓一体（Lakehouse）：融合了数据湖的存储能力和数据仓库的计算能力

多模态：既能处理关系表数据，也能处理时序数据、图数据

云原生：数据库迁移到云上（提升数据库存储和计算的拓展性）

3、元数据：用来描述数据的数据。

业务元数据

技术元数据

4、数据仓库模型和数据库模型对比

数据仓库模型

A、概念模型：分析主题（边界、主题、维度、类别、指标、事实）

B、逻辑模型：星型模型、雪花模型、星座模型

C、物理模型：

用户



系统

数据库模型

A、概念模型：ER（实体关系模型）

B、逻辑模型：关系模型、网状模型、层次模型

C、物理模型：

5、数据仓库模型

销售分析的概念模型

信息包：销售分析

日期	销售地点	销售产品	年龄组别	性别
年（10）	国家（15）	产品类（6）	年龄组（8）	性别组（2）
季度（40）	区域（45）	产品组（48）		
月（120）	城市（280）	产品（240）		
	区（880）			
	商店（2000）			
度和事实： 预测销售量、实际销售量、预测误差				

销售分析的逻辑模型（星型模型）

星型模型（Star Schema）

