

# 第1章 数据库基础

计科2201zzy友情分享😎😊

## 数据库系统概述

### 基础概念

- 数据：
  - 收集->保存->抽取有用部分
  - 数据和它的语义密不可分。
  - 数据是数据库中存储的基本对象。
- 数据库：
  - 数据库是长期储存在计算机内、有组织的、可共享的大量数据集合。
  - 数据库中的数据按一定的数据模型组织、描述和储存，具有较小的冗余度、较高的数据独立性和易扩展性，并可为各种用户共享。
- 数据库管理系统：位于用户和操作系统之间的一层数据管理软件
  - 在数据库建立、运用和维护时对数据库进行统一控制，以保证数据的完整性和安全性，并在多用户同时使用数据库时进行并发控制，在发生故障后对数据库进行恢复。
  - 数据库管理系统的主要功能：
    - 提供数据定义与数据操作功能
      - 通过数据定义语言(Data Definition Language)与数据操纵语言(Data Manipulation Language)
      - 用增删改查四个字可以概括
    - 数据组织、存储和管理功能
      - 包括数据字典、用户数据、数据的存储路径，同时也确定文件结构、存取方式等。
    - 数据库的运行管理(事务管理和运行管理)功能
      - 保证数据的安全性、完整性；
      - 确保多用户对数据的并发使用
      - 发生故障后的系统恢复
    - 数据库的建立和维护功能
      - 数据的批量装载
      - 数据库转储与恢复
      - 介质故障恢复
      - 数据库的重组织
      - 性能监视与分析 等

- 数据库系统：在计算机系统中引入数据库后的系统构成。通常将数据库系统称为数据库。
  - 即数据库系统包含了以上几个概念。
  - 由数据库、数据库管理系统及其开发工具、应用程序、用户和数据库管理员构成。
  - 存储、管理、处理和维持数据的系统
- 数据库管理员：负责全面管理和控制数据库系统，有以下职责：
  - 决定数据库中的信息内容和结构
  - 决定数据库的存储结构和存取策略
  - 定义数据的安全性要求和完整性约束条件
  - 监控数据库的使用和运行

## 数据管理技术

### 数据管理

- 对数据进行分类、组织、编码、存储、检索和维护，是数据处理的中心问题

### 发展过程

- 人工管理->文件系统->数据库系统阶段

表 1.1 数据管理三个阶段的比较

		人工管理阶段	文件系统阶段	数据库系统阶段
背景	应用背景	科学计算	科学计算、数据管理	大规模数据管理
	硬件背景	无直接存取存储设备	磁盘、磁鼓	大容量磁盘、磁盘阵列
	软件背景	没有操作系统	有文件系统	有数据库管理系统
	处理方式	批处理	联机实时处理、批处理	联机实时处理、分布处理、批处理
特点	数据的管理者	用户（程序员）	文件系统	数据库管理系统
	数据面向的对象	某一应用程序	某一应用	现实世界（一个部门、企业、跨国组织等）
	数据的共享程度	无共享，冗余度极大	共享性差，冗余度大	共享性高，冗余度小
	数据的独立性	不独立，完全依赖于程序	独立性差	具有高度的物理独立性和一定的逻辑独立性
	数据的结构化	无结构	记录内有结构、整体无结构	整体结构化，用数据模型描述
	数据控制能力	应用程序自己控制	应用程序自己控制	由数据库管理系统提供数据安全性、完整性、并发控制和恢复能力

- 人工管理(40年代中-50年代中):
  - 冗余，无共享、不一致、(数据和代码)不独立
- 文件系统(50-60年代): 和上一结构无本质差别
- 数据库系统(60年代末以来)

# 数据库系统的特点

- 数据结构化
  - 数据库系统实现整体数据的结构化(和文件系统的本质区别)
  - 结构用数据模型描述，可以变长。
  - 数据的最小存取单位是数据项。
- 高共享性：低冗余、易于保持一致、易于扩充
- 数据独立性高：包括数据的物理独立性和逻辑独立性
  - 物理独立性：应用程序和数据的物理存储方式之间独立，数据由DBMS隔离管理
  - 逻辑独立性：应用程序与数据库的逻辑结构之间独立，数据的逻辑结构改变了，应用程序不用改变
- 具有数据控制功能
  - 数据的安全性保护：防止非法使用造成的泄密与破坏
  - 数据的完整性检查：数据的正确性、有效性与相容性
  - 并发控制：协调多用户并发操作
  - 数据库恢复

## 数据模型

- 一种模型，是对现实世界数据特征的抽象。也就是说数据模型是用来描述数据、组织数据和对数据进行操作的。
- 数据模型是数据库系统的核心和基础。

## 两个层次

- 概念模型(信息模型)：按用户的观点对数据和信息建模；现实世界→信息世界
- 逻辑模型&物理模型：
  - 逻辑模型主要包括层次模型、网状模型、关系模型等。按计算机系统的观点对数据建模。
  - 物理模型是对数据最底层的抽象。

## 概念模型：从现实世界到信息世界

- 实体(entity)：客观存在且可相互区别的事物
- 属性(attribute)：实体所具有的某一特性。一个实体可具有多个属性。
- 码(key)：唯一标识实体的属性
- 域(domain)：属性的取值范围
- 实体型(entity type)：
  - 用实体名及其属性名集合来抽象和刻画
  - 类比编程中“类”的概念
- 联系：一对一、一对多、多对多
  - 同一实体集内各实体联系

- 多个不同实体之间联系

## 概念模型的表示方法

- 实体-联系方法(E-R方法)
  - 详见教材P215
  - 用E-R图来描述现实世界的概念模型
  - 实体用矩形表示，属性用椭圆框表示
  - 联系本身用菱形表示，框内写明联系名，用无向边连接有关实体，并在旁边标上联系的类型
  - 联系也可以具有属性。

## 逻辑模型 - 关系模型：从信息世界到机器世界

逻辑模型不止关系模型，但只需要掌握关系模型即可。

### 逻辑结构：一张二维表

- **关系(relation)**：一个关系对应一张表
- **元组(tuple)**：表中的一行
- **属性(attribute)**：表中的一列即为一个属性，
- **码(key)**：表中的某个属性组，可以唯一确定一个元组
- **域(domain)**：一组具有相同数据类型的值的集合。属性的取值范围来自某个域，如性别的域是（男，女）
- 分量：元组中的一个属性值
  - 注意：每一个分量必须是不可分的数据项，即不允许表中有表

## 实体模型的表示方法

- 实体型：直接用关系(表)表示
- 属性：用属性名表示
- 一对一联系：隐含在实体对应的关系中
- 一对多联系：隐含在实体对应的关系中
- 多对多联系：直接用实体表示

例：学生、课程之间的多对多联系  
学生（学号，姓名，性别，班级，年级）  
课程（课程号，课程名，学分）  
选修（学号，课程号，成绩）

## 关系模型的完整性约束

- 实体完整性

- 参照完整性
- 用户定义的完整性

## 数据模型的组成要素

- 数据结构：描述数据库的组成对象以及对象之间的联系
- 数据操作：（对型和值）增删改查
- 约束条件：一组完整性规则的集合

## 数据库系统的结构

- 模式是数据库中全体数据的逻辑结构和特征的描述。
- 数据库系统的三级模式结构：外模式、模式和内模式
  - 模式：所有用户的公共数据视图，是数据库中的逻辑结构和特征的描述。一个数据库只有一个模式。
  - 外模式：数据库用户能够看见和使用的、局部的数据逻辑结构和特征的描述。是数据库用户的数据视图，是与某一应用有关的数据的逻辑表示
  - 内模式：存储模式，是对数据物理结构和存储方式的描述，是数据在数据库内部的组织方式。
- 两层映像
  - 外模式/模式映像：确保模式改变时，外模式&应用程序保持不变——数据的逻辑独立性
  - 模式/内模式映像：确保数据库存储结构（内模式）改变时，模式可以保持不变——数据的物理独立性