

# 1 Introduction

- 数据

- 对现实世界中客观事物的符号表示
- 可以是数值数据, 也可以是非数值数据, 如声音、图像等
- 计算机中数据
  - 能输入计算机, 并能为其处理的符号序列
- 信息与数据的联系
  - 数据是信息的符号表示, 或称载体
  - 信息是数据的内涵, 是数据的语义解释
  - 数据是符号化的信息
  - 信息是语义化的数据
- 数据抽象: 通过多个数据抽象级别隐藏数据结构的复杂性, 以向用户表示数据库中的数据。

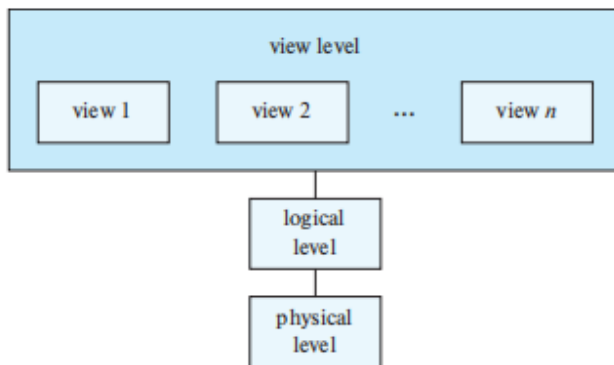


Figure 1.2 The three levels of data abstraction.

- 物理层: 描述数据实际上是怎样存储的
- 逻辑层: 描述存储什么样的数据以及数据之间存在什么关系。
  - **物理数据独立**: 在逻辑层的用户不需要意识到物理层的复杂性
- 视图层: 只提供一部分信息, 交互简单
- 数据模型: 一组用于描述数据、数据关系、数据语义和一致性约束的概念工具。
  - 概念数据模型: 按用户的观点来对数据和信息建模 (E-R model)
    - E-R model: 世界是由一组称作实体的基本对象和这些对象之间的联系构成的
      - 实体
      - 属性
      - 域: 属性的取值范围
  - 逻辑数据模型: 从计算机实现的观点来对数据建模 (层次模型、网状模型、关系模型、面向对象模型)
    - 三要素: 数据结构、数据操作、数据的约束条件
    - 关系模式 (Ted Codd提出): 用二维表格表示

- 物理数据模型
- 数据库语言
  - **数据定义语言** (Data Definition Language)
    - 定义数据库模式
    - DDL 编译器生成一组存储在数据字典中
      - **数据字典**包含**元数据**（即关于数据的数据，metadata），特殊的表，只能由数据库系统本身访问修改
      - 数据库架构
      - 完整性约束
      - 授权
    - create table 那些
  - **数据操纵语言** (Data Manipulation Language)
    - 让用户访问或操纵按照某种适当的数据模型组织起来的数据。
    - 两种
      - 过程化DML：要求指定需要什么数据以及如何获取
      - 声明式DML：只指定需要什么数据
    - DML中涉及信息检索的部分称为查询语言（query language）
    - SQL 查询语言是非过程化的。它以几个表作为输入，总是仅返回一个表(select那些)
- 数据库设计
  - 逻辑设计——数据库模式，“好的”关系模式
  - 业务决策——属性
  - 计算机科学决策：关系模式、属性如何分布
  - 物理设计：物理布局
- 实例和模式
  - 类似于编程语言中的类型和变量
  - 逻辑模式——数据库的整体逻辑结构，类似于程序中变量的类型信息
  - 物理模式——数据库的整体物理结构，类似于变量的值
- 事务管理
  - 事务是在数据库应用程序中执行单个逻辑功能的操作集合
  - 事务管理组件确保数据库在系统故障（例如电源故障和操作系统崩溃）和事务失败的情况下保持一致（正确）状态。
  - 并发控制管理器控制并发事务之间的交互，以保证数据库的一致性。