

数据库概念

第二章 关系模型与关系代数

关系模型的几个基本概念：**关系模型的数据结构、关系操作和关系完整性约束**

关系数据结构

关系数据结构为：

- **关系**
 - 关系模型的数据结构非常简单，它就是二维表，亦称为**关系**
 - **关系数据库**就是表的集合，即**关系的集合**
- **关系模式**
 - 表的**表头部分**对应于关系模式，关系模式是型的概念
- **码**
 - 超码：属性集A能唯一地标识关系r中的一个元组
 - 候选码：属性集A是关系r的超码，且属性集A的任意真子集都不能成为关系r的超码
候选码是最小的超码
 - 主码：若一个关系有多个候选码，则可以选定其中一个候选码作为该关系的主码
 - 外码：设F是关系r的一个属性（集）， K_s 是关系s的主码。如果F与 K_s 相对应，则称F是关系r**参照**关系s的外码
- **关系数据库模式**
 - 关系数据库也有值和型之分，关系数据库的**型**就是**关系数据库模式**；值就是关系数据库实例

关系操作

关系模型中的关系操作有**查询操作**和**关系操作**两大类

5个基本关系操作：**选择、投影、集合并、集合差和笛卡尔积**

关系完整性约束

- 实体完整性
 - 若属性集A是关系r的主码，则A不能取null
- 参照完整性
 - 若关系r的外码参照关系s的主码，则对于关系r中的每一个元组在属性F上的取值，要么取null，要么等于关系s中某个元组的主码值
- 用户自定义完整性

第四章 数据库建模

数据库设计过程

- 需求分析
- 概念设计
 - 设计E-R模型
- 逻辑设计
 - 将E-R模型转化为关系数据库模式
- 模式求精
 - 减少数据冗余、消除更新、插入与删除异常
- 物理设计
 - 包括数据库文件组织格式、内部存储结构、建立索引、表的聚集
- 应用和安全设计

E-R模型基本概念及表示

E-R模型采用实体集、属性集和属性三个基本概念来分别描述事物、联系和特征

- **实体和实体集**
 - 实体是客观世界中可区别于其他事物的“事物”和对象
 - 实体集是指具有相同类型即相同性质（或属性）的实体集合
- **属性**
 - 属性是实体集中每个实体都具有的特征描述
 - 可分为：简单属性和复合属性、单值属性和多值属性（双椭圆表示）、派生属性（虚线椭圆表示）
- **联系与联系集**
 - 联系是指多个实体间的相互关联
 - 多联系：实体之间有多种不同的联系
 - 实体的角色：实体在联系中的作用叫做实体的角色
 - 联系集的度：**参与联系集的实体集数目**称为联系集的度

E-R模型转化方法

从E-R模型转化为关系模式：

- **强实体集转化方法**：将实体集的每个属性对应为关系模式的属性，实体集的码作为关系的码
- **弱实体集转化方法**：弱实体集对应的关系模式属性由弱实体集本身的描述属性（**部分码**）加上所依赖的**标识实体集的主码**属性组成
- **属性集转化方法**：一个联系集映射而成的关系模式的属性，由来自参与联系的所有实体集的主码属性和联系本身的描述属性组成。
 - 生成模式的主码：
 - 一对一联系集：主码可选择任何一个参与实体集的主码
 - 一对多(多对一)联系集：主码由“多”的一方的实体集的主码组成
 - 多对多联系集：主码由所有参与实体集的主码的**并集**组成
 - 特别地，当为一对多和一对一联系集时可以不转化为单独的关系模式
 - **一对多，多的一方加少的外码和联系的属性**

- 一对一，将某一方的主码属性及联系集属性的描述增加到另一方实体集所转化的关系模式中去

- **复合属性及多值属性转化方法：**

- 对于复合属性，对每个子属性创建一个单独的属性，而不是复合属性自身创建一个单独的属性。例如**地址应该转化为省、城市、街道**
- 对于多值属性，则必须为其**创建一个新的关系模式**，其属性为多值属性所在的实体集（或联系集）的**主码属性和该多值属性对应的属性组成，主码为全部属性**

- PhoneNumber(studentNo, pNumber)

- **类层次转化方法：**父类实体集和子类实体集分别转化为单独的模式。其中，父类实体集对应的关系模式属性为父类实体集的属性，而各个子类实体集对应的模式由该子类的特殊属性和父类实体集的主码属性组成。**子类的主码与父类实体集的主码相同**

- Student(**stuNo**, name, sex, birthday)
- Undergraduate(**stuNo**, interest)
- Graduate(**stuNo**, direction)

第五章 模式求精

对于给定的关系R (A1,A2,.....,An) 和其上的函数依赖集F，可将其属性分为如下四类：

- **L类：**仅出现在F的函数依赖左侧的属性
- **R类：**仅出现在F的函数依赖右侧的属性
- **N类：**在F的函数依赖左右两侧均未出现的属性
- **LR类：**在F的函数依赖左右两侧均出现的属性

码=超键：能够唯一标识一条记录的属性或属性集

候选键=候选码：能够唯一标识一条记录的**最小**属性集

主键=主码：某个能够唯一标识一条记录的最小属性集（是从候选码里人为挑选的一条）

主属性：包含在任一候选码中的属性称主属性。简单来说，主属性是候选码所有属性的并集

非主属性：不包含在候选码中的属性称为非主属性。非主属性是相对于主属性来定义的

基于函数依赖理论，关系模式可以分成：

- 第一范式（1NF）：所有属性都是原子的
- 第二范式（2NF）：不存在**非主属性对候选码的部分依赖**
- 第三范式（3NF）：不存在**非主属性对候选码的传递依赖**
- BC范式（BCNF）：排除了任何属性（包括主属性和非主属性）对候选码的部分依赖以及传递依赖，并且排除了**主属性之间的传递依赖**