

关系数据库

基本概念

1. 关系及其性质

- **域**：属性的取值单位
- **笛卡尔集合**
- **关系**： $D_1 \times D_2 \times \dots \times D_n$ 的有意义的子集称为在域 D_1, D_2, \dots, D_n 上的关系，记为 $R(D_1, D_2, \dots, D_n)$

2. 关系的常用术语

- **候选码**：唯一标识一个元组，且不含多余属性
- **主码**：候选码中其中一个
- **外码**：非本关系的候选码，但是是其他关系的候选码
- **全码**：整个属性的集合是 R 的候选码

关系模型的特点

- 结构简单 表达能力强
- 语言一体化
- 非过程化操作
- 坚实的数学基础
- 操作效率低

关系模型的组成

- **关系数据结构**
- **关系数据操作**
 - 数据查询
 - 数据增加
 - 数据删除
 - 数据修改
- **关系完整性约束**
 - 实体完整性 【主码不能为空】
 - 参照完整性 【主码与外码的引用规则 定义在同一组】
 - 用户定义的完整性 【取值范围 非空限制】

关系数据语言

1. 关系代数语言 【集合操作】

- 集合运算（并，交，差；广义笛卡尔积）
- 关系运算（投影，选择，连接，除运算）
- 扩充的关系运算

2. 关系演算语言 【谓词演算】

3. 结构化查询语言

查询优化

技术分类

- **规则优化** 启发式规则【例如 先选择 后投影】
- **物理优化** 物理组织和访问路径 【例如 索引优化】
- **代价估算优化** 【多个策略中选择代价最小的】

一般策略

一般先代数优化，后物理优化

基本原则：尽量减少查询过程中的产生的结果

1. 先做选择运算 投影运算
2. 笛卡尔集合与选择 =》连接运算
3. 投影运算与选择运算同时进行
4. 让投影运算和其前后的其他运算同时进行
5. 在执行连接前对关系适当的预处理
6. 寻找公共子表达式