

第2章 关系数据库

计科2201zzy友情分享😁😊

关系数据结构

- 关系模型建立在集合代数的基础上。以下给出关系数据结构的形式化定义。

关系

- 域：域是一组具有相同数据类型的值的集合。
- 关系： $D_1 \times D_2 \times \dots \times D_n$ 的子集叫做在域 D_1, D_2, \dots, D_n 上的关系，表示为 $R(D_1, D_2, \dots, D_n)$
- 候选码：若关系中的某一属性组能唯一地标识一个元组（这样的属性组称为超码），而其子集不能，则称该属性组为候选码。
 - 主码：若一个关系中有多个候选码，则选定其中一个为主码。
 - 候选码的诸属性称为主属性。不包含在任何候选码中的属性称为非主属性或非码属性。

关系模式

- 关系模式是对关系的描述，是型；关系是值。
 - 关系模式是静态的、稳定的。
 - 关系是关系模式在某一时刻的状态或内容，是动态的、随时间变化的。
- 关系模式可形式化地表示为： $R(U, D, DOM, F)$ ；可以简记为 $R(A_1, A_2, \dots, A_n)$
 - R ：关系名
 - A_1, A_2, \dots, A_n ：属性名

关系数据库

- 关系数据库也有型和值的区分
 - 型：关系数据库模式，包含若干个关系模式
 - 值：这些关系模式在某一时刻对应的关系的集合

关系操作

- 增删改查
- 特点：集合操作方式——操作的对象和结果都是集合

关系数据语言

- 关系代数语言：

- 用对关系的运算来表达查询要求
- 关系演算语言：用谓词来表达查询要求
- 具有双重特点的语言：SQL
 - SQL（结构化查询语言），集查询、数据定义语言、数据操纵语言和数据控制语言（DCL）于一体的关系数据语言。

关系的完整性

- 关系模型的完整性规则是对关系的某种约束条件。包括实体完整性、参照完整性和用户定义的完整性
- 实体完整性和参照完整性是关系模型必须满足的完整性约束条件，被称作是关系的两个不变性

实体完整性

- 若属性A是基本关系R的主属性，则属性A不能取空值。
 - 详细说明

参照完整性

外码的概念

- 现有关系R和S，F是关系R的一个或一组属性，但不是它的码。
 - 如果F和S的主码 K_S 相对应，则称F是R的外码。
 - 称基本关系R为参照关系；基本关系S为被参照关系或目标关系。谁的主码被用，谁就是被参照的。
 - 关系R和S不一定是不同的关系。如学生表 学生(学号, 姓名, 性别, 班长) 中的一个属性“ta的班长”（填对应学生学号）。

关系代数

- 关系代数用对关系的运算来表达查询。
- 关系代数运算的三个要素
 - 运算对象&运算结果：都是关系
 - 运算符：见下表

集合运算符	含义		关系运算符	含义
\cup	并		σ	选择
$-$	差		Π	投影
\cap	交		\bowtie	连接
\times	笛卡尔积		\div	除

传统的集合运算

- 二目运算，包括并、差、交、笛卡尔积四种。
- **记号引入**：设关系 R 和关系 S 具有相同的目 n （即具有 n 个属性），且相应的属性取自同一个域。 t 是元组变量，表示任意一个元组。以下是对四种运算的形式化阐述。
 - 并： $R \cup S = \{t | t \in R \vee t \in S\}$
 - 差： $R - S = \{t | t \in R \wedge t \notin S\}$
 - 交： $R \cap S = \{t | t \in R \wedge t \in S\}$
 - 笛卡尔积： $R \times S = \{\widehat{t_r t_s} | t_r \in R \wedge t_s \in S\}$
 - $\widehat{t_r t_s}$ 称为元组的连接。它是一个 $n + n$ 列的元组。

专门的关系运算

- 专门的关系运算包括选择、投影、连接、除运算等。
- **记号引入**：设关系模式为 $R(A_1, A_2, \dots, A_n)$ ，它的一个关系设为 R 。 $t \in R$ 表示 t 是 R 的一个元组。 $t[A_i]$ 则表示元组 t 中相应于属性 A_i 的一个分量。

选择

- 在关系 R 中选择满足给定条件的诸元组。即选取符合条件的行。
- 记作： $\sigma_F(R) = \{t | t \in R \wedge F(t) \text{ 为真} \}$
 - 其中 F 表示选择条件，是一个逻辑表达式。用比较运算符和逻辑运算符进行连接。
- 以StudentCourse数据库为例
 - 查询信息系全体学生： $\sigma_{Sdept='IS'}(Student)$
 - 查询年龄小于20岁的学生： $\sigma_{Sage < 20}(Student)$

投影

- 关系 R 上的投影：指从 R 中选择出若干属性列组成新的关系。投影操作是从列的角度进行的运算。
- 记作： $\Pi_A(R) = \{t[A] | t \in R\}$
 - 其中 A 为 R 中的属性列。
- 同样以SC数据库为例
 - 查询学生的姓名与所在系： $\Pi_{Sname, Sdept}(Student)$

连接

- 连接也称为 θ 连接，指从两个关系的笛卡儿积中选取属性间满足一定条件的元组。
- 记作： $R \bowtie_{A\theta B} S = \{\widehat{t_r t_s} | t_r \in R \wedge t_s \in S \wedge t_r[A] \theta t_s[B]\}$
 - 其中 θ 是比较运算符。连接运算从 $R \times S$ 中选取对应属性列满足这一比较关系的元组。
- 等值连接： θ 为 "=" 的连接运算称为等值连接。

- 自然连接：一种特殊的等值连接。
 - 要求两关系中比较的分量同名，并在结果中去掉重复的属性列。（同时从行和列的角度进行的运算）。
 - 此时 $A\theta B$ 可以省略
- 外连接：把悬浮元组也保留在结果关系中的自然连接。其他属性上填入空值。
 - 悬浮元组：一个关系中的有些元组，在另一关系中并不存在公共属性相等的元组。

除运算

- 设关系 R 除以关系 S 的结果为关系 T ，则：
 - T 包含所有在 R 但不在 S 中的属性及其值
 - T 的元组与 S 的元组的所有组合都在 R 中
- **记号引入**：给定关系 $R(X, Y)$ 和 $S(Y, Z)$ ，其中 X, Y, Z 为属性组。
- R 和 S 的除运算得到一个新的关系 $P(X)$ ， P 是 R 中满足下列条件的元组在 X 属性列上的投影：
 - 元组在 X 上分量值 x 的象集 Y_x 包含 S 在 Y 上投影的集合。
- 记作： $R \div S = \{t_r[X] | t_r \in R \wedge \Pi_Y(S) \subseteq Y_x\}$
 - 其中 Y_x 是 x 在 R 中的象集， $x = t_r[X]$ 。
- 形象化理解by zzy：
 - 找到R关系和S关系的共有属性列。取出R中的所有元组，每个元组都按照列纵向掰成A和B两个元组，其中A元组只包含共有列，B元组包含其它的列。则A元组就是B的象集中的一个元素。
 - 上述的所有的A元组，有哪些也包含在S关系中？记下它们。
 - 如果有一个B元组，它对应的所有A元组都在S里面，那么它就是结果元组之一。