第二章 关系数据库

目录

1 关系数据结构及形式化定义

1.1 关系

- **域**：域是一组具有相同数据类型的值的集合
- **笛卡尔积**：给定一组域 $D_1, D_2, \cdots, D_n$，允许其中某些域是相同的，$D_1, D_2, \cdots, D_n$ 的笛卡尔积为

$$D_1 \times D_2 \times \cdots \times D_n = \{(d_1, d_2, \cdots, d_n) | d_i \in D_i, \, i = 1, 2, \cdots, n\}$$

  - 其中，每一个元素 $(d_1, d_2, \cdots, d_n)$ 叫作一个 n **元组**，或简称元组
  - 元素中的每一个值 d_i 叫做一个**分量**
  - 一个域允许的不同取值个数称为这个域的**基数**
  - 若 $D_i (i = 1, 2, \cdots, n)$ 为有限集，其基数为 $m_i (i = 1, 2, \cdots, n)$ ，则 $D_1 \times D_2 \times \cdots \times D_n$ 的基数 $M = \prod_{i=1}^{n} m_i$
- **关系**：$D_1 \times D_2 \times \cdots \times D_n$ 的子集叫做在域名 $D_1 \times D_2 \times \cdots \times D_n$ 上的关系，表示为 $R(D_1 \times D_2 \times \cdots \times D_n)$
  - R 表示关系的名字
  - n 是关系的**目**或者**度**
- 若关系中的某一属性组的值能唯一地标识一个元组，则称该属性组为**候选码**
  - 若一个关系有多个候选码，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**
  - 在简单的情况下，候选码只包含一个属性；在最极端的情况下，关系模式的所有属性组是这个关系模式的候选码，称为**全码**
- 关系可以有三种类型：**基本关系**（基本表或基表）、**查询表**和**视图表**
  - 基本关系是实际存在的表，是实际存储数据的逻辑表示
  - 查询表是查询结果对应的表
  - 视图表是由基本表或其他视图表导出的表，是虚表，不对应实际存储的数据
- 基本关系的性质：
  - 列是同质的，即每一列中的分量是同一类型的数据，来自同一个域
  - 不同的列可出自同一个域，称其中的每一列称为一个属性，不同的属性要给予不同的属性名
  - 列的顺序无所谓
  - 任意两个元组的候选码不能相同
  - 行的顺序无所谓
  - 分量必须取原子值

1.2 关系模式

- 关系模式是对关系的描述，关系模式是型，关系是值
- 关系的描述称为**关系模式**。可以形式化地表示为 $R(U, D, DOM, F)$
  - R 为关系名
  - U 为组成该关系的属性名集合
  - D 为 U 中属性所来自的域
  - DOM 为属性向域的映像集合
  - F 为属性间数据的依赖关系集合

1.3 关系数据库

- 在一个给定的应用领域中，所有关系的集合构成一个关系数据库
- 关系数据库的型，也称关系数据库模式，是对关系数据库的描述
- 关系数据库的值，是这些关系模式在某一时刻对应的关系的集合，通常就称为关系数据库

2 关系的完整性

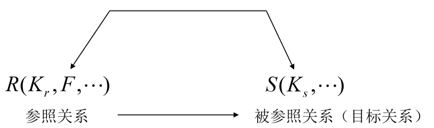
- 实体完整性和参照完整性
  - 关系模型必须满足的完整性约束条件称为关系的两个不变性，应该由关系系统自动支持
- 用户定义的完整性
- 应用领域需要遵循的约束条件，体现了具体领域中的语义约束

2.1 实体完整性

- **实体完整性规则：**若属性 A 是基本关系 R 的主属性，则属性 A 不能取空值。空值就是“不知道”或“不存在”或“无意义”的值
- 对于实体完整性规则的说明：
  1. 实体完整性规则是针对基本关系而言的。一个基本表通常对应现实世界的一个实体集
  2. 现实世界中的实体是可区分的，即它们具有某种唯一性标识
  3. 关系模型中以主码作为唯一性标识
  4. 主码中的属性即主属性不能取空值。主属性取空值，就说明存在某个不可标识的实体，即存在不可区分的实体，这与第 2 点相矛盾，因此这个规则称为实体完整性

2.2 参照完整性

- 在关系模型中实体及实体间的联系都是用关系来描述的，自然存在着关系与关系间的引用
- 设 F 是基本关系 R 的一个或一组属性，但不是关系 R 的码，K_s 是基本关系 S 的主码。如果 F 与 K_s 相对应，则称 F 是 R 的**外码**，并称基本关系 R 为**参照关系**，基本关系 S 为**被参照关系**或**目标关系**
  - 其中关系 R 和 S 不一定是不同的关系
  - 目标关系 S 的主码 K_s 和参照关系的外码 F 必须定义在同一个（或一组）域上
  - 外码并不一定要与相应的主码同名
    - 当外码与相应的主码属于不同关系时，往往取相同的名字，以便于识别



- **参照完整性规则：**若属性（或属性组）F 是基本关系 R 的外码它与基本关系 S 的主码 K_s 相对应（基本关系 R 和 S 不一定是不同的关系），则对于 R 中每个元组在 F 上的值必须为：
  - 或者取空值（F 的每个属性值均为空值）
  - 或者等于 S 中某个元组的主码值

2.3 用户定义的完整性

- 用户定义的完整性是针对某一具体关系数据库的约束条件，反映某一具体应用所涉及的数据必须满足的语义要求
- 关系模型应提供定义和检验这类完整性的机制，以便用统一的系统的方法处理它们，而不需由应用程序承担这一功能

3 关系代数

3.1 传统的集合运算

- 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：$R \times S = \{\widehat{t_r t_s}|t_r \in R \wedge t_s \in S\}$

3.2 专门的关系运算

首先引入几个记号：

- R，$t \in R$，$t[A_i]$：
  - 设关系模式为 $R(A_1, A_2, \cdots, A_n)$
  - 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
- A，$t[A]$，$\overline{A}$
  - 若 $A = \{A_{i1}, A_{i_2}, \cdots, A_{ik}\}$，其中 $A_{i1}, A_{i_2}, \cdots, A_{ik}$ 是 $A_1, A_2, \cdots, A_n$ 中的一部分，则 A 称为属性列或属性组
  - $t[A] = (t[A_{i1}], t[A_{i2}], \cdots, t[A_{ik}])$ 表示元组 t 在属性列 A 上诸分量的集合
  - $\overline{A}$ 则表示 $\{A_1, A_2, \cdots, A_n\}$ 中去掉 $\{A_{i1}, A_{i_2}, \cdots, A_{ik}\}$ 后剩余的属性组
- $\widehat{t_r t_s}$：
  - R 为 n 目关系，S 为 m 目关系
  - $t_r \in R, t_s \in S, \widehat{t_r t_s}$ 称为元组的连接
  - $\widehat{t_r t_s}$ 是一个 $n+m$ 列的元组，前 n 个分量为 R 中的一个 n 元组，后 m 个分量为 S 中的一个 m 元组
- 象集 Z_X
  - 给定一个关系 $R(X, Z)$，X 和 Z 为属性组
  - 当 $t[X] = x$ 时，x 在 R 中的象集为：$Z_X = \{t[Z]|t \in R, t[X] = x\}$
  - 它表示 R 中属性组 X 上值为 x 的诸元组在 Z 上分量的集合

下例中的学生—课程数据库如下：

Student

学号 Sno	姓名 Sname	性别 Ssex	年龄 Sage	所在系 Sdept
201215121	李勇	男	20	CS
201215122	刘晨	女	19	CS
201215123	王敏	女	18	MA
201215125	张立	男	19	IS

(a)

Course

课程号 Cno	课程名 Cname	先行课 Cpno	学分 Ccredit
1	数据库	5	4
2	数学		2
3	信息系统	1	4
4	操作系统	6	3
5	数据结构	7	4
6	数据处理		2
7	PASCAL语言	6	4

(b)

SC

学号 Sno	课程号 Cno	成绩 Grade
201215121	1	92
201215121	2	85
201215121	3	88
201215122	2	90
201215122	3	80

(c)

3.2.1 选择

- 选择又称为限制
- 选择是在关系 R 中选择满足给定条件的诸元组，记作

$$\sigma_F(R) = \{t|t \in R \wedge F(t) = \text{“真”}\}$$

  其中 F 表示选择条件，它是一个逻辑表达式，取逻辑值“真”或“假”
  - 逻辑表达式 F 的基本形式为 $X_1 \theta Y_1$，其中 θ 表示比较运算符。在基本的选择条件上可以进一步进行逻辑运算
- 选择运算是从行角度进行的运算

例：查询信息系（IS 系）全体学生：$\sigma_{\text{Sdept='IS'(Student)}}$，结果如下：

Sno	Same	Sex	Sage	Slept
201215125	张立	男	19	IS

3.2.2 投影

- 投影是从 R 中选择出若干属性列组成新的关系，记作

$$\Pi_A(R) = \{t[A]|t \in R\}$$

- 投影操作主要是从列的角度进行运算
- 投影之后不仅取消了原关系中的某些列，而且还可能取消某些元组（避免重复行）

例：查询学生关系 Student 中都有哪些系，即查询关系 Student 上所在系属性上的投影：$\Pi_{\text{Sdept}}(\text{Student})$，结果如下：

Sdept
CS
IS
MA

3.2.3 连接

- 连接也称为 θ 连接，是从两个关系的笛卡尔积中选取属性间满足一定条件的元组，记作

$$R \bowtie S = \{\widehat{t_r t_s} | t_r \in R \wedge t_s \in S \wedge t_r[A]\theta t_s[B]\}$$

- 连接运算从 R 和 S 的广义笛卡尔积 $R \times S$ 中选取 R 关系在 A 属性组上的值与 S 关系在 B 属性组上的值满足比较关系 θ 的元组
- 连接运算中最为重要且最为常用的连接为等值连接和自然连接
  - θ 为“=”的连接运算称为等值连接
    - 是从关系 R 与 S 的广义笛卡尔积中选取 A, B 属性值相等的那些元组，即等值连接为：

$$R \underset{A=B}{\bowtie} S = \{\widehat{t_r t_s} | t_r \in R \wedge t_s \in S \wedge t_r[A] = t_s[B]\}$$

  - 自然连接是一种特殊的等值连接
    - 要求两个关系中进行比较的分量必须是相同的属性组
    - 并且在结果中把重复的属性列去掉
    - 若 R 和 S 具有相同的属性组 B，U 为 R 和 S 的全体属性集合，则自然连接可记作

$$R \bowtie S = \{\widehat{t_r t_s}[U - B] | t_r \in R \wedge t_s \in S \wedge t_r[B] = t_s[B]\}$$

  - 一般的连接操作是从行的角度进行运算
  - 自然连接还需要取消重复列，所以是同时从行和列的角度进行运算。

例：设下图 (a) 和 (b) 分别为关系 R 和关系 S，图 (c) 为非等值连接 $R \underset{C<E}{\bowtie} S$ 的结果，图 (d) 为等值连接 $R \underset{R.B=S.B}{\bowtie} S$ 的结果，图 (e) 为自然连接 $R \bowtie S$ 的结果

R

A	B	C
a_1	b_1	5
a_1	b_2	6
a_2	b_3	8
a_2	b_4	12

(a)关系R

S

B	E
b_1	3
b_2	7
b_3	10
b_3	2
b_5	2

(b)关系S

A	$R.B$	C	$S.B$	E
a_1	b_1	5	b_2	7
a_1	b_1	5	b_3	10
a_1	b_2	6	b_2	7
a_1	b_2	6	b_3	10
a_2	b_3	8	b_3	10

(c)非等值连接

A	$R.B$	C	$S.B$	E
a_1	b_1	5	b_1	3
a_1	b_2	6	b_2	7
a_2	b_3	8	b_3	10
a_2	b_3	8	b_3	2

(d)等值连接

A	B	C	E
a_1	b_1	5	3
a_1	b_2	6	7
a_2	b_3	8	10
a_2	b_3	8	2

(e)自然连接

- 两个关系 R 和 S 在做自然连接时，关系 R 中某些元组有可能在 S 中不存在公共属性上值相等的元组，从而造成 R 中这些元组在操作时被舍弃了，这些被舍弃的元组称为**悬浮元组**
- 如果把悬浮元组也保存在结果关系中，而在其他属性上填空值（Null），就叫做**外连接**，记作 $R ⟗ S$
  - 如果只保留左边关系 R 中的悬浮元组，则称为**左外连接**，记作 $R ⟕ S$
  - 如果只保留右边关系 S 中的悬浮元组，则称为**右外连接**，记作 $R ⟖ S$

例：下图 (a) 是上例中关系 R 和关系 S 的外连接，图 (b) 是左外连接，图 (c) 是右外连接

A	B	C	E
a_1	b_1	5	3
a_1	b_2	6	7
a_2	b_3	8	10
a_2	b_3	8	2
a_2	b_4	12	NULL
NULL	b_5	NULL	2

(a)外连接

A	B	C	E
a_1	b_1	5	3
a_1	b_2	6	7
a_2	b_3	8	10
a_2	b_3	8	2
a_2	b_4	12	NULL

(b)左外连接

A	B	C	E
a_1	b_1	5	3
a_1	b_2	6	7
a_2	b_3	8	10
a_2	b_3	8	2
NULL	b_5	NULL	2

(c)右外连接

3.2.4 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，且 T 的元组与 S 的元组的所有组合都在 R 中

用**象集**定义除运算：

- 给定关系 $R(X,Y)$ 和 $S(Y,Z)$，其中 X,Y,Z 为属性组
- R 中的 Y 与 S 中的 Y 可以有不同的属性名，但必须出自相同的域集
- R 与 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：

$$R \div S = \{t_r[X] | t_r \in R \wedge \Pi_Y(S) \subset Y_X\}$$

  其中 Y_x 为 x 在 R 中的象集，$x = t_r[X]$
- 除是同时从行和列的角度进行运算

例：设关系 R, S 分别为下图中的 (a) 和 (b) ，$R \div S$ 的结果如图 (c)

R

A	B	C
a_1	b_1	c_2
a_2	b_3	c_7
a_3	b_4	c_6
a_1	b_2	c_3
a_4	b_6	c_6
a_2	b_2	c_3
a_1	b_2	c_1

(a)

S

B	C	D
b_1	c_2	d_1
b_2	c_1	d_1
b_2	c_3	d_2

(b)

$R \div S$

A
a_1

(c)