

第 2 章 关系数据库

本章学习目标

本章主要讲述了关系模型的数据结构、数据操纵和完整性约束以及关系系统的定义和分类。通过本章的学习，读者应该掌握以下内容：

- 关系模型的数据结构
- 并、交、差和笛卡尔积 4 种传统的集合运算
- 选择、投影、连接和除 4 种专门的关系运算
- 关系的实体完整性规则和参照完整性规则
- 关系系统的定义和分类

2.1 关系模型的基本概念

系统而严格地提出关系模型的是美国 IBM 公司的 E.F.Codd 于 1970 年提出的关系数据模型——“A Relational Model of Data for Large Shared Data Banks”（《Communication of the ACM》），1970 年之后提出了关系代数和关系演算的概念，1972 年提出了关系的第一、第二、第三范式。

关系数据库应用数学方法来处理数据库中的数据，20 世纪 80 年代后，关系数据库系统成为最重要、最流行的数据库系统。典型的商用系统有：

- Oracle
- SQL Server
- Access
- Sybase
- Informix
- DB2

关系模型建立在集合代数的基础上，关系数据结构的基本概念：关系、关系模式、关系数据库。

2.1.1 关系

1. 域

定义 2.1：域是一组具有相同数据类型的值的集合。

例如：学生集合={李勇,刘晨,王敏,...}

教师集合={张清玫,刘逸,...}

班级集合={计算机班级,信管班级}

2. 笛卡尔积 (Cartesian Product)

定义 2.2: 给定一组域 $D_1, D_2, \dots, D_n$, 则 $D_1, D_2, \dots, D_n$ 的笛卡尔积为:

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

所有域的所有取值的一个组合, 不能重复。

【例 2-1】给出三个域:

教师集合={张清玫,刘逸}

班级集合={计算机班级,信管班级}

学生集合={李勇,刘晨,王敏}

则 D_1, D_2, D_3 的笛卡尔积为:

$D_1 \times D_2 \times D_3 = \{(张清玫, 计算机班级, 李勇), (张清玫, 计算机班级, 刘晨),$
 $(张清玫, 计算机班级, 王敏), (张清玫, 信管班级, 李勇),$
 $(张清玫, 信管班级, 刘晨), (张清玫, 信管班级, 王敏),$
 $(刘逸, 计算机班级, 李勇), (刘逸, 计算机班级, 刘晨),$
 $(刘逸, 计算机班级, 王敏), (刘逸, 信管班级, 李勇),$
 $(刘逸, 信管班级, 刘晨), (刘逸, 信管班级, 王敏) \}$

定义 2.3: 基数 (Cardinal Number)。若 $D_i (i=1, 2, \dots, n)$ 为有限集, 其基数为 $m_i (i=1, 2, \dots, n)$, 则 $D_1 \times D_2 \times \dots \times D_n$ 的基数 M 为:

$$M = \prod_{i=1}^n m_i$$

在上例中, 基数: $2 \times 2 \times 3 = 12$, 即 $D_1 \times D_2 \times D_3$ 共有 $2 \times 2 \times 3 = 12$ 个元组。

笛卡尔积的表示方法: 笛卡尔积可表示为一个二维表。表中的每行对应一个元组, 表中的每列对应一个域。在上例中, 12 个元组可列成一张二维表, 如表 2-1 所示。

表 2-1 D_1, D_2, D_3 的笛卡尔积

教师	班级	学生
张清玫	计算机班级	李勇
张清玫	计算机班级	刘晨
张清玫	计算机班级	王敏
张清玫	信管班级	李勇
张清玫	信管班级	刘晨
张清玫	信管班级	王敏
刘逸	计算机班级	李勇
刘逸	计算机班级	刘晨
刘逸	计算机班级	王敏
刘逸	信管班级	李勇
刘逸	信管班级	刘晨
刘逸	信管班级	王敏

3. 关系

笛卡尔积是没有实际语意的，只有它的子集（关系）才有实际意义。

定义 2.4： 关系。 $D_1 \times D_2 \times \dots \times D_n$ 的子集叫作在域 $D_1, D_2, \dots, D_n$ 上的关系，表示为：

$$R(D_1, D_2, \dots, D_n)$$

其中：R 为关系名；n 为关系的目或度（Degree）。

4. 关系的性质

(1) 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。

(2) 不同的列可出自同一个域，称其中的每一列为一个属性，不同的属性要给予不同的属性名。

(3) 列的顺序无所谓，即列的次序可以任意交换。

(4) 任意两个元组不能完全相同。

(5) 行的顺序无所谓，即行的次序可以任意交换。

(6) 分量必须取原子值，即每一个分量都必须是不可再分的数据项。这是规范条件中最基本的一条。

2.1.2 关系数据结构

在用户看来，一个关系模型的逻辑结构是一张二维表，它由行和列组成。例如，表 2-2 中的学生记录就是一个关系模型，它涉及下列概念。

表 2-2 学生关系

学号	姓名	性别	年龄	所在系
000101	王萧	男	17	经济系
000207	李云虎	男	18	机械系
010302	郭敏	女	18	信息系
010408	高红	女	20	土木系
⋮	⋮	⋮	⋮	⋮
020309	王睿	男	19	信息系
020506	路旭青	女	21	管理系

- 关系：一个关系对应一张二维表，表 2-2 中的这张学生记录表就是一个关系。
- 元组：图中的一行称为一个元组，若表 2-2 有 20 行，就有 20 个元组。
- 属性：图中的一列称为一个属性，表 2-2 有 5 列，对应 5 个属性：学号、姓名、性别、年龄和所在系。
- 码：表中的某个属性（组），它可以唯一确定一个元组，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。如表 2-2 中的学号列，可以作为该学生关系的码来唯一标识一个学生的信息。
- 域：属性的取值范围。如表 2-2 中学生年龄的域应是（16~28），性别的域是（男,女），系别的域是一个学校所有系名的集合。

- 分量：元组中的一个属性值。
- 关系模式：关系模式（Relation Schema）是对关系的描述（表的数据结构）。

关系模式通常可以简记为：

$$R(U) \text{ 或 } R(A_1, A_2, \dots, A_n)$$

其中：R 为关系名； $A_1, A_2, \dots, A_n$ 为属性名。

注意：域名及属性常常直接说明为：属性的类型、长度。

如表 2-2 的学生关系可描述为：学生(学号,姓名,性别,年龄,所在系)。

关系模式：对关系的描述是静态的、稳定的。

关系：关系模式在某一时刻的状态或内容（具体的表的值），是动态的、随时间不断变化的。

关系模式和关系往往统称为关系。

2.1.3 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。

关系模型中有三类完整性约束：实体完整性、参照完整性、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，应该由关系系统自动支持。

1. 实体完整性（Entity Integrity）

若属性 A 是基本关系 R 的主属性，则属性 A 不能取空值。

例：学生(学号,姓名,...)，学号属性为主码，则属性学号不能取空值。

关系模型必须遵守实体完整性规则的原因：

（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或多对多联系。

（2）现实世界中的实体和实体间的联系都是可区分的，即它们具有某种唯一性标识。

（3）相应地，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。空值就是“不知道”或“无意义”的值。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。

注意：实体完整性规则规定基本关系的所有主属性都不能取空值。

例如：课程(课程号,课程名)，“课程号”为主码，则属性不能取空值。

2. 参照完整性

关系间的引用：在关系模型中实体及实体间的联系都是用关系来描述的，因此可能存在着关系与关系间的引用。

例如：学生实体、班级实体以及班级与学生间的一对多联系：

学生(学号,姓名,性别,出生日期,班级编号)

班级(班级编号,班级名称)，如图 2-1 所示。

学号	姓名	性别	出生日期	班级编号
9601001	岳艳玲	女	1977-08-21 00:...	9601
9601002	罗军	男	1975-11-05 00:...	9601
9601003	张英	女	1977-09-07 00:...	9601
9601004	王静波	男	1976-02-03 00:...	9601
9601005	蔡尧	男	1974-06-23 00:...	9601
9601006	高峰	男	1973-10-26 00:...	9601
9601007	孙琴	女	1976-07-30 00:...	9601
9601008	罗军	男	1975-04-16 00:...	9601
9601009	李阳	男	1977-05-10 00:...	9601
9601010	王海鸥	男	1975-09-29 00:...	9601

班级编号	班级名称
9601	96计算机
9602	96财会
9701	97计算机
9702	97财会

图 2-1 学生与班级之间的关系

定义 2.5: 外码 (Foreign Key)。设 F 是基本关系 R 的一个或一组属性, 但不是关系 R 的码。如果 F 与基本关系 S 的主码 K_s 相对应, 则称 F 是基本关系 R 的外码。

学生(学号,姓名,性别,出生日期,班级编号)

班级(班级编号,班级名)

基本关系 R (学生)的外码是班级编号, 称为参照关系 (Referencing Relation)。

基本关系 S (班级)的主码是班级编号, 称为被参照关系 (Referenced Relation) 或目标关系 (Target Relation)。

说明: 关系 R 和 S 是不同的关系, 被参照关系 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, 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。

定义 2.6: 参照完整性。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, 则对于 R 中每个元组在 F 上的值必须等于 S 中某个元组的主码值。

学生(学号,姓名,性别,出生日期,班级编号)

班级(班级编号,班级名)

学生关系中每个元组的“班级编号”属性只取非空值, 这时该值必须是班级关系中某个元组的“班级编号”值。班级编号是班级关系中的主属性, 按照实体完整性和参照完整性规则, 学生关系中班级编号只能取相应被参照关系中已经存在的主码值。

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，以使用统一的系统的方法处理它们，而不应由应用程序承担这一功能。

例如：课程(课程号,课程名,学分)
“课程号”属性必须取唯一值，非主属性“课程名”也不能取空值，“学分”属性只能取值{1,2,3,4}。

2.2 关系代数

关系代数是一种抽象的查询语言，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：关系；运算结果：关系；运算符：四类（见表 2-3）。

表 2-3 关系代数运算符

运算符		含义	运算符		含义
集合运算符	∪	并	比较运算符	>	大于
	-	差		≥	大于等于
	∩	交		<	小于
	×	广义笛卡尔积		≤	小于等于
				=	等于
				≠	不等于
专门的关系运算符	σ	选择	逻辑运算符	¬	非
	Π	投影		∧	与
	⋈	连接		∨	或

2.2.1 传统的集合运算

传统的集合运算是二目运算，包括并、交、差和广义笛卡尔积四种运算。设关系 R 和关系 S 具有相同的目 n（即两个关系都具有 n 个属性），且相应的属性取自同一个域，表示记号：

- (1) $R, t \in R, t[A_i]$ 。
- 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。

- (2) $\widehat{t_r t_s}$ 。
- R 为 n 目关系，S 为 m 目关系。 $t_r \in R, t_s \in S$ ， $t_r t_s$ 称为元组的连接。它是一个 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。

则四种运算定义如下：

1. 并

关系 R 与关系 S 的并由属于 R 或属于 S 的元组组成，其结果关系仍为 n 目关系。记作 $R \cup S$ 。R 和 S 具有相同的目 n（即两个关系都有 n 个属性），相应的属性取自同一个域。

$$R \cup S = \{t | t \in R \vee t \in S\}$$

2. 交

关系 R 与关系 S 的交由既属于 R 又属于 S 的元组组成，其结果关系仍为 n 目关系。记作 $R \cap S$ 。

$$R \cap S = \{t | t \in R \wedge t \in S\}$$

3. 差

关系 R 与关系 S 的差由属于 R 而不属于 S 的所有元组组成。其结果关系仍为 n 目关系。记作 $R - S$ 。

$$R - S = \{t | t \in R \wedge t \notin S\}$$

4. 广义笛卡尔积

R 为 n 目关系， k_1 个元组， S 为 m 目关系， k_2 个元组，则：

$R \times S$ 列：(n+m)列的元组的集合，元组的前 n 列是关系 R 的一个元组，后 m 列是关系 S 的一个元组。

行： $k_1 \times k_2$ 个元组。

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

【例 2-2】有关系 R 、 S ，则 $R \cup S$ 、 $R \cap S$ 、 $R - S$ 、 $R \times S$ 的结果分别为表 2-4 中的 (c)、(d)、(e)、(f) 所示。

表 2-4 传统的集合运算

R			S		
A	B	C	A	B	C
a1	b1	c1	a1	b2	c2
a1	b2	c2	a1	b3	c2
a2	b2	c1	a2	b2	c1

(a) (b)

R ∪ S		
A	B	C
a1	b1	c1
a1	b2	c2
a1	b3	c2
a2	b2	c1

(c)

R ∩ S		
A	B	C
a1	b2	c2
a2	b2	c1

(d)

R - S		
A	B	C
a1	b3	c2

(e)

表 2-4 传统的集合运算 (续表)

R×S					
A	B	C	A	B	C
a1	b1	c1	a1	b2	c2
a1	b1	c1	a1	b3	c2
a1	b1	c1	a2	b2	c1
a1	b2	c2	a1	b2	c2
a1	b2	c2	a1	b3	c2
a1	b2	c2	a2	b2	c1
a2	b2	c1	a1	b2	c2
a2	b2	c1	a1	b3	c2
a2	b2	c1	a2	b2	c1

(f)

2.2.2 专门的关系运算

专门的关系运算包括选择、投影、连接等。

1. 选择

选择 (Selection) 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 为选择条件, 是一个逻辑表达式, 基本形式为:

$$[\neg (X_1 \theta Y_1)] [\phi [\neg (X_2 \theta Y_2)]] \dots$$

θ 为比较运算符 ($>$, $\geq$, $<$, $\leq$, $=$ 或 $<>$)。

X_1 、 Y_1 等: 属性名、常量、简单函数; 属性名也可以用它的序号来代替。

ϕ 为逻辑运算符 ($\wedge$ 或 $\vee$), $[\]$ 表示任选项。

因此, 选择运算实际上是从关系 R 中选取使逻辑表达式 F 为真的元组。这是从行的角度进行的运算。

设有一个学生—课程关系数据库, 包括学生关系 S、课程关系 C 和选修关系 SC, 如表 2-5 所示。下面的例子将对这三个关系进行运算。

表 2-5 学生—课程关系数据库

S					C		
学号 S#	姓名 SN	性别 SS	年龄 SA	所在系 SD	课程号 C#	课程名 CN	学分 CC
000101	李晨	男	18	信息系	1	数学	6
000102	王博	女	19	数学系	2	英语	4
010101	刘思思	女	18	信息系	3	计算机	4
010102	王国美	女	20	物理系	4	制图	3
020101	范伟	男	19	数学系			

(a)

(b)

表 2-5 学生—课程关系数据库（续表）

SC		
学号 S#	课程号 C#	成绩 G
000101	1	90
000101	2	87
000101	3	72
010101	1	85
010101	2	42
020101	3	70

(c)

【例 2-3】查询数学系学生的信息。

 $\sigma_{SD='数学系'}(S)$

或

 $\sigma_{5='数学系'}(S)$

结果如表 2-6 所示。

表 2-6 查询数学系学生的信息

学号 S#	姓名 SN	性别 SS	年龄 SA	所在系 SD
000102	王博	女	19	数学系
020101	范伟	男	19	数学系

【例 2-4】查询年龄<20 的学生的信息。

 $\sigma_{SA<20}(S)$

或

 $\sigma_{4<20}(S)$

结果如表 2-7 所示。

表 2-7 查询年龄<20 的学生的信息

学号 S#	姓名 SN	性别 SS	年龄 SA	所在系 SD
000101	李晨	男	18	信息系
000102	王博	女	19	数学系
010101	刘思思	女	18	信息系
020101	范伟	男	19	数学系

2. 投影

关系 R 上的投影（Projection）是从 R 中选择出若干属性列组成新的关系。记作：

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

其中：A 为 R 中的属性列。

投影操作是从列的角度进行的运算。投影之后不仅取消了原关系中的某些列，而且还能取消某些元组，因为取消了某些属性列后，就可能出现重复行，应取消这些完全相同的行。

【例 2-5】查询学生的学号和姓名。

 $\Pi_{S\#,SN}(S)$

或

$$\Pi_{1,2}(S)$$

结果如表 2-8 所示。

表 2-8 查询学生的学号和姓名

学号 S#	姓名 SN
000101	李晨
000102	王博
010101	刘思思
010102	王国美
020101	范伟

【例 2-6】查询学生的所在系，即查询学生关系 S 在所在系属性上的投影。

$$\Pi_{SD}(S)$$

或

$$\Pi_5(S)$$

结果如表 2-9 所示。

表 2-9 查询结果

所在系 SD
信息系
数学系
物理系

3. 连接

(1) 连接 (Join) 也称为 θ 连接。

(2) 连接运算的含义。

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：

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

其中：A 和 B 分别为 R 和 S 上度数相等且可比的属性组；

θ 为比较运算符，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 属性组上值满足比较关系的元组。

(3) 两类常用的连接运算。

1) 等值连接 (Equal join)， θ 为 “=” 的连接运算称为等值连接。

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：

$$R \bowtie_{A=B} S = \{\widehat{t_r t_s} \mid t_r \in R \wedge t_s \in S \wedge t_r[A] = t_s[B]\}$$

2) 自然连接 (Natural join)。

自然连接是一种特殊的等值连接，两个关系中进行比较的分量必须是相同的属性组，在结果中把重复的属性列去掉。R 和 S 具有相同的属性组 B，自然连接为：

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

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。

综上所述，我们可以看出等值连接与自然连接的区别：

(1) 等值连接中不要求相等属性值的属性名相同，而自然连接要求相等属性值的属性名必须相同，即两关系只有在同名属性才能进行自然连接。如前面例题中：R 中的 C 列和 S 中的 D 列可进行等值连接，但因为属性名不同，不能进行自然连接。

(2) 等值连接不将重复属性去掉，而自然连接去掉重复属性，也可以说，自然连接是去掉重复列的等值连接。如前面例题中：R 中的 B 列和 S 中的 B 列进行等值连接时，结果有两个重复的属性列 B，而进行自然连接时，结果只有一个属性列 B。

【例 2-7】查询选修了 2 号课程的学生的学号。

$\Pi_{S\#}(\sigma_{C\#=2}(SC))$

【例 2-8】查询选修了 3 号课程的学生的姓名。

$\Pi_{SN}(\sigma_{C\#=3}(SC \bowtie S))$

【例 2-9】查询选修了数学课的学生的姓名和成绩。

$\Pi_{SN,G}(\sigma_{CN=\text{数学}}(C \bowtie SC \bowtie S))$

本章小结

本章主要讲述了关系模型的数据结构、数据操作和完整性约束以及关系系统的定义和分类。通过本章的学习，读者应该理解关系模型的数据结构和关系的两种完整性规则；掌握选择、投影、连接和除四种专门的关系运算；能运用关系代数、关系演算进行简单操作；并了解关系系统的定义和分类。

习题 2

一、填空

1. 关系数据模型中，实体及实体间的联系都用_____来表示。在数据库的物理组织中，它以_____形式存储。
2. 常用的关系操作有两类：传统的集合操作，如并、交、差、_____。专门的关系操作，如_____、_____、_____和除等。
3. 关系数据库的完整性约束包括_____、_____和_____三类。

二、操作题

有如下的四个关系：

S（供应商）：

SNO（供应商号）	SNAME（供应商姓名）	CITY（供应商所在城市）
S1	精益	天津
S2	万胜	北京
S3	东方	北京
S4	丰泰隆	上海
S5	康健	南京

P (零件):

PNO (零件号)	PNAME (零件名称)	COLOR (零件颜色)	WEIGHT (零件重量)
P1	螺母	红	12
P2	螺栓	绿	17
P3	螺丝刀	蓝	14
P4	螺丝刀	红	14
P5	凸轮	蓝	40

J (项目):

JNO (项目号)	JNAME (项目名称)	CITY (项目所在城市)
J1	三建	北京
J2	一汽	长春
J3	弹簧厂	天津
J4	造船厂	天津
J5	机车厂	唐山
J6	无线电厂	常州

SPJ (供应情况):

SNO (供应商号)	PNO (零件号)	JNO (项目号)	QTY (供应数量)
S1	P1	J1	200
S1	P1	J3	100
S1	P1	J4	700
S1	P2	J2	100
S2	P3	J1	400
S2	P3	J2	200
S2	P3	J4	500
S2	P3	J5	400
S2	P5	J1	400
S2	P5	J2	100
S3	P1	J1	200
S3	P3	J1	200
S4	P5	J1	100
S5	P6	J2	200
S5	P6	J4	500

试用关系代数完成下列操作:

1. 求供应商供应的商品的零件号。

2. 求供应商 S5 供应的商品的零件号。
3. 求供应工程 J1 零件的供应商号。
4. 求供应工程 J1 零件 P1 的供应商号。
5. 求供应工程 J1 红色零件的供应商号。

三、简答题

1. 关系模型的完整性规则有哪几类？在关系模型的参照完整性规则中，外部码属性的值是否可以为空？什么情况下才可以为空？
2. 关系系统可以分为哪几类？各类关系系统的定义是什么？