

第 2 章 关系数据库

【学习目标】

- 掌握关系的定义及性质、候选码、主码、外码等基本概念。
- 掌握关系数据库三种完整性规则，即实体完整性、参照完整性和用户定义完整性的内容和意义。
- 掌握关系代数运算的使用方法。

2.1 关系模型

关系模型就是用二维表格结构来表示实体及实体之间联系的模型。关系模型是各个关系的框架的集合，即关系模型是一些表格的格式，其中包括关系名、属性名和关键字等。

关系数据库是表的集合，即关系的集合，关系就是二维表。表中一行代表的是若干值之间的关联，即表的一行是由有关联的若干值构成的。一般来说，一个表是一个实体集，一行就是一个实体，即一个元组，它由共同表示一个实体的有关联的若干属性的值所构成；一列就是一个属性。

表 2-1 所示的是学生信息的关系模型；表 2-2 所示的是课程信息的关系模型；表 2-3 所示的是学生选修课程的关系模型。

表 2-1 学生表

学号	姓名	性别	年龄
201501	张娟	女	18
201502	李强	男	21
201503	刘莉	女	19
201504	王华	男	19

表 2-2 课程表

课程号	课程名	课时
C1	计算机基础	48
C2	数据库原理及应用	64
C3	大学英语	56

表 2-3 选修表

学号	课程号	分数
201501	C1	80
201501	C2	90
201501	C3	70
201502	C1	60
201502	C3	85
201503	C1	76
201503	C2	88

从以上列表中，我们可以看出 3 个表之间有联系。学生关系和选修关系有公共的属性“学号”，表明这两个关系有联系；而课程关系和选修关系有公共的属性“课程号”，则表明这两个关系也有联系；学生关系和课程关系之间通过选修关系存在间接联系。至于元组之间的联系，则与具体的数据有关，只有在公共属性上具有相同属性值的元组之间才有联系。

由以上内容可以看出，在一个关系中可以存放以下两类信息：

- (1) 描述实体本身的信息。
- (2) 描述实体（关系）之间的联系的信息。

在层次模型和网状模型中，把有联系的实体（元组）用指针链接起来，实体之间的联系是通过指针来实现的。而关系模型采用不同的思想，不仅用二维表来表示实体，同时用二维表表示实体与实体之间的联系，这就是关系模型的本质所在。

因此，在建立关系模型时，只要把所有的实体及其属性用表来表示，同时把实体之间的联系也用表来表示，就可以得到一个关系模型。

2.2 关系的形式化定义

关系模型建立在集合代数的基础之上，我们可以从集合论的角度给出关系的形式化的定义。

2.2.1 域 (Domain)

定义：具有相同数据类型的值的集合称为一个域，又称为值域（用 D 表示）。如自然数集是一个域，记为 D （自然数）；整数集是一个域，记为 D （整数）；奇数集是一个域，记为 D （奇数）。域中所包含的值的个数称为域的基数（用 m 表示）。在关系中就是用域来表示属性的取值范围。例如：

$D1 = \{\text{张娟, 李强, 刘莉, 王华}\}$, $m=4$;

$D2 = \{\text{男, 女}\}$, $m=2$;

其中 $D1, D2$ 为域名，分别表示学生关系中姓名、性别的集合。

2.2.2 笛卡尔积 (Cartesian Product)

给定一组域 $D1, D2, \dots, Dn$ （这些域中可以有相同的）， $D1, D2, \dots, Dn$ 的笛卡尔积定义为：

$$D1 \times D2 \times \cdots \times Dn = \{(d1, d2, \cdots, dn) | di \in Di, i=1, 2, \cdots, n\}$$

其中，集合中的每一个元素 $(d1, d2, \cdots, dn)$ 称为一个 n 元组，简称为元组，元素中的每个值 di 称为该元素的一个分量。

【例 2-1】设 $D1 = \{\text{学士, 硕士, 博士}\}$, $D2 = \{\text{王慧, 李宏}\}$, 则笛卡尔积

$D1 \times D2 = \{(\text{学士, 王慧}), (\text{学士, 李宏}), (\text{硕士, 王慧}), (\text{硕士, 李宏}), (\text{博士, 王慧}), (\text{博士, 李宏})\}$

【例 2-2】设 $D1 = \{\text{计算机软件专业, 信息科学技术专业}\}$, $D2 = \{\text{张珊, 李海, 王宏}\}$, $D3 = \{\text{男, 女}\}$, 则笛卡尔积 $D1 \times D2 \times D3$ 为:

计算机软件专业	×	张珊	×	男	=	计算机软件专业	张珊	男
信息科学技术专业		李海		女		计算机软件专业	张珊	女
		王宏				计算机软件专业	李海	男
						计算机软件专业	李海	女
						计算机软件专业	王宏	男
						计算机软件专业	王宏	女
						信息科学技术专业	张珊	男
						信息科学技术专业	张珊	女
						信息科学技术专业	李海	男
						信息科学技术专业	李海	女
						信息科学技术专业	王宏	男
						信息科学技术专业	王宏	女

笛卡尔积的任意一行数据就是一个元组，它的第一个分量来自 $D1$ ，第二个分量来自 $D2$ ，第三个分量来自 $D3$ 。笛卡尔积就是所有这样的元组的集合。

根据笛卡尔积的概念，可以给出一个关系的形式化定义：

设 $D1, D2, \cdots, Dn$ 是 n 个属性域，笛卡尔积 $D1 \times D2 \times \cdots \times Dn$ 的任意一个子集称为定义在属性域 $D1, D2, \cdots, Dn$ 上的一个 n 元关系。

2.2.3 关系的基本性质

(1) 每一分量必须是不可分的最小数据项，即每个属性都是不可再分解的，或者说所有属性都是原子项。

(2) 每一列中的分量是同类型的数据，来自同一个值域。

(3) 不同的列可以出自同一个值域，每一列都称为一个属性，每个属性要给予不同的属性名。

(4) 列的顺序是无关紧要的，即列的次序可以任意交换，但一定是整体交换，属性名和属性值必须作为整列同时交换。

(5) 行的顺序是无关紧要的，即行的次序可以任意交换。

(6) 元组不可以重复，即在一个关系中任意两个元组不能完全一样。

2.2.4 关系模式

关系模式是对关系的描述，它包括关系名、组成该关系的属性名、值域名、模式的主关键字、属性间数据的依赖关系。关系模式通常可以简记为： $R(U)$ 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。例如教师关系模式：教师（教师号，姓名，性别，年龄，职称，系部）。

2.3 关系的键

2.3.1 候选关键字与主关键字

候选关键字（候选码、候选键）：能唯一标识关系中元组的一个属性或属性组。

候选关键字的性质：

唯一性。任何两个候选关键字值都是不相同的。

最小性。组成候选关键字的属性组中，任一属性都不能从该属性组中删掉，否则会破坏唯一性的性质。

注意：在一个关系中，候选关键字可能有多个。

主码（主关键字、主键）：如果一个关系中有多个候选码，可以从中选择一个作为查询、插入或删除元组的操作变量，被选用的候选码称为主码。

2.3.2 主属性与非主属性

主属性：包括在候选码中的各个属性。

非主属性：不包含在任何候选码中的属性。

全码：所有属性的组合是关系的唯一候选码。

2.3.3 外关键字

定义：如果关系 R_2 的一个或一组属性 X 不是 R_2 的主码，而是另一个关系 R_1 的主码，则该属性或属性组 X 称为关系 R_2 的外关键字或外码，并称关系 R_2 为参照关系，关系 R_1 为被参照关系。

【例 2-3】假设职工管理数据库中有两个关系模式：

部门（部门号，部门名称，负责人）

职工（职工号，姓名，性别，年龄，部门号）

“部门号”是部门关系的主码，而“部门号”并不是职工关系的主码，所以“部门号”是职工关系的外码。

如表 2-3 所示的选修关系中，“学号”属性与学生关系的主码“学号”相对应，“课程号”属性与课程关系的主码“课程号”相对应。因此，“学号”和“课程号”属性是选修关系的外关键字。学生关系和课程关系为被参照关系，选修关系为参照关系。

由外关键字的定义可知，被参照关系的主码和参照关系的外码必须定义在同一个域上。如将选修关系中“学号”属性与学生关系的主码“学号”定义在同一个域上，“课程号”属性

与课程关系的主码“课程号”定义在同一个域上。

2.4 关系完整性

完整性是数据模型的一个非常重要的方面，为了维护数据库中数据与现实世界的一致性，对关系数据库的插入、删除和更新操作必须有一定的约束条件，关系数据库从多个方面来保证数据的完整性。

关系模型的三类完整性，即实体完整性、参照完整性和用户定义完整性。其中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，被称为是关系的两个不变性，应该由关系系统自动生成。

2.4.1 实体完整性

实体完整性是指主码的值不能为空值或部分为空值。

实体完整性规则保证关系中的每个元组都是可识别和唯一的。

关系模型中的一个关系对应一个实体集，一个元组对应一个实体。例如，一条课程记录对应着一门课程，课程关系对应着课程的集合。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。与此相对应，关系模型中以主码来唯一识别元组。例如，课程关系中的属性“课程号”可以唯一标识一个元组，也可以唯一标识课程实体。如果主码中的值为空或部分为空，则不符合主码的定义条件，不能唯一标识元组及其相对应的实体。这就说明存在不可区分的实体，从而与现实世界中的实体是可以区分的事实相矛盾。因此主码的值不能为空或部分为空。例如课程关系中的主码“课程号”不能为空，选修关系中的主码“学号+课程号”不能部分为空，即“学号”和“课程号”两个属性都不能为空。

2.4.2 参照完整性

如果关系 R2 的外码 X 与关系 R1 的主码相符，则 X 的每个值或者等于 R1 中主码的某一个值，或者取空值。

在【例 2-3】中，部门关系中的属性“部门号”是职工关系的外码。如表 2-4 所示，职工关系中某个职工（如 T1 或 T2）“部门号”的取值，必须在参照的部门（如表 2-5 所示）中主关系键“部门号”的值中能够找到，否则表示把该职工分配到一个不存在的部门中，显然不符合语义。如果某个职工（如 T16）“部门号”取空值，则表示该职工尚未分配到任何一个部门；否则，他只能取部门关系中某个元组的部门号值。

表 2-4 职工表

职工号	姓名	性别	年龄	部门号
T1	张丽	女	32	X1
T2	赵勇	男	35	X2
.....
T16	李刚	男	23	

表 2-5 部门表

部门号	部门名称	负责人
X1	人事部	周强
X2	财务部	王蕾
X3	采购部	李华

在表 2-3 中，如果按照参照完整性规则，选修关系中的外部关系键“学号”和“课程号”可以取空值或者取被参照关系中已经存在的值。但由于“学号”和“课程号”是选修关系中的主属性，根据实体完整性规则，两个属性都不能为空。所以选修关系中外部关系键“学号”和“课程号”中只能取被参照关系（表 2-1、表 2-2）中已经存在的值。

实体完整性和参照完整性是关系模型必须满足的完整约束条件，被称作关系的两个不变性。任何关系数据库系统都应该支持这两类完整性。除此之外，不同的关系数据库由于应用环境不同，往往还需要一些特殊的约束条件，这就是用户定义完整性。

2.4.3 用户定义完整性

用户定义完整性是针对某一具体关系数据库的约束条件，它反映某一具体应用所涉及的数据必须满足的语义要求。在用户定义完整性中最常见的是限定属性的取值范围，即对值域的约束，所以在用户定义完整性中最常见的是域完整性约束。如学生关系中性别只能为“男”或“女”，选修关系中成绩不能为负数；某些数据的输入格式要有一些限制等。关系模型应该提供定义和验证这类完整性的机制，以便采取统一的、系统的方法处理它们，而不要由应用程序承担这一功能。

2.5 关系代数

2.5.1 关系代数的定义、分类及运算符

关系代数是一种抽象的查询语言，是关系数据操纵语言的一种传统表达方式，用对关系的运算来表达查询，作为研究关系数据语言的数学工具。关系代数的运算对象是关系，运算结果亦为关系。

关系代数用到的运算符有四类：集合运算符、专门的关系运算符、比较运算符和逻辑运算符，如表 2-6 所示。

关系代数的运算按运算符的不同，主要分为传统的集合运算和专门的关系运算两类。

(1) 传统的集合运算将关系看成元组的集合，其运算是从关系的“水平”方向即行的角度来进行的。它包括并、交、差和笛卡尔积。

(2) 专门的关系运算不仅涉及行而且涉及列。它包括选择、投影、连接和除法。

比较运算符和逻辑运算符是用来辅助专门的关系运算符进行操作的。

表 2-6 关系代数运算符

运算符		含义	运算符		含义
集合运算符	∪	并 交	比较运算符	>	大于
	∩			≥	大于等于
	-	差		<	小于
	×	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
	÷	除			

2.5.2 传统的集合运算

传统的集合运算是二目运算，包括四种运算：并、交、差、广义笛卡尔积。

设关系 R 和关系 S 具有相同的目 n （即两个关系都有 n 个属性），且相应的属性取自同一个域，则可以定义并、交、差、广义笛卡尔积如下：

1. 并 (Union)

设关系 R 和 S 具有相同的模式， R 和 S 的并是由属于 R 或属于 S 的元组构成的集合，记为 $R \cup S$ 。形式定义如下：

$$R \cup S = \{t \mid t \in R \vee t \in S\}, \text{ } t \text{ 是元组变量, } R \text{ 和 } S \text{ 的元数相同。}$$

2. 交 (Intersection)

关系 R 和 S 的交是由属于 R 又属于 S 的元组构成的集合，记为 $R \cap S$ ，这里要求 R 和 S 定义在相同的模式上。形式定义如下：

$$R \cap S = \{t \mid t \in R \wedge t \in S\}, \text{ } t \text{ 是元组变量, } R \text{ 和 } S \text{ 的元数相同。}$$

3. 差 (Difference)

设关系 R 和 S 具有相同的模式， R 和 S 的差是由属于 R 但不属于 S 的元组构成的集合，记为 $R - S$ 。形式定义如下：

$$R - S = \{t \mid t \in R \wedge t \notin S\}, \text{ } t \text{ 是元组变量, } R \text{ 和 } S \text{ 的元数相同。}$$

由于 $R \cap S = R - (R - S)$ ，或 $R \cap S = S - (S - R)$ ，因此交操作不是一个独立的操作。

4. 广义笛卡尔积 (Extended Cartesian Product)

两个分别为 n 元和 m 元的关系 R 和 S 的广义笛卡尔积是一个 $(n + m)$ 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。若 R 有 i 个元组， S 有 j 个元组，则关系 R 和关系 S 的广义笛卡尔积有 $i \times j$ 个元组。

【例 2-4】 有关系 R 和 S ，分别求 $R - S$ ， $R \cup S$ ， $R \cap S$ ， $R \times S$ 。

A	B	C
a1	b1	c1
a1	b2	c2
a2	b2	c1

A	B	C
a1	b2	c2
a1	b3	c2
a2	b2	c1

解:

A	B	C
a1	b1	c1

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

A	B	C
a1	b2	c2
a2	b2	c1

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

2.5.3 专门的关系运算

专门的关系运算包括选择、投影、连接、除等。其中，选择、投影为一元操作，连接、除为二元操作。

1. 选择

选择运算是从指定的关系中选择满足条件的某些元组形成一个新的关系。

选择运算表示为:

$$\sigma_F(R) = \{t | t \in R \wedge F(t) = \text{'真'}\}$$

其中， σ 是选择运算符， R 是关系名， t 是元组， F 表示选择条件，它是一个逻辑表达式，取逻辑值“真”或“假”。

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。

假设教学管理数据库中有三个关系模式：

学生（学号、姓名、性别、年龄）

课程（课程号、课程名、课时）

选修（学号、课程号、分数）

【例 2-5】查询所有的男生。

$\sigma_{\text{性别}=\text{男}}(\text{学生})$

【例 2-6】查询年龄小于 20 岁的学生。

$\sigma_{\text{年龄}<20}(\text{学生})$

【例 2-7】查询年龄小于 20 岁的男生。

$\sigma_{\text{性别}=\text{男} \wedge \text{年龄}<20}(\text{学生})$

2. 投影

关系 R 上的投影是从 R 中选出若干属性列组成新的关系。从关系中消除某些属性，就可能出现重复行，应取消这些完全相同的行。

投影运算表示为：

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

【例 2-8】查询学生的姓名和性别。

$\Pi_{\text{姓名}, \text{性别}}(\text{学生})$

【例 2-9】查询女学生的姓名和年龄。

$\Pi_{\text{姓名}, \text{年龄}}(\sigma_{\text{性别}=\text{女}}(\text{学生}))$

3. 连接

连接运算是两个表之间的运算，是从两个关系的笛卡尔积中选择满足条件的元组，组成新的关系。将满足两个表之间运算关系的记录连接成一条记录，所有这样的记录构成新的表（连接运算的结果）。连接运算也称为 θ 运算。

连接运算一般表示为：

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

其中， \bowtie 是连接运算符， θ 为算术比较运算符，也称 θ 连接。

$X\theta Y$ 为连接条件，其中：

θ 为 “=” 时，称为等值连接；

θ 为 “<” 时，称为小于连接；

θ 为 “>” 时，称为大于连接。

连接运算中最常用的连接有两个：一个是等值连接，另一个是自然连接。

自然连接是一种特殊的连接，在等值连接的情况下，当连接属性 A 与 B 具有相同属性组时，在结果中要去掉相同的属性列。也就是说，若关系 R 和 S 具有相同的属性组 B ，则自然连接可记为：

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

一般的连接运算是从行的角度进行运算的，但自然连接还需要去掉相同的列，所以它是从

行和列的角度进行运算的。

综上所述，自然连接与等值连接的区别如下：

(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，即两关系只有同名属性才能进行自然连接。

(2) 等值连接不将重复属性去掉，而自然连接要去掉重复属性，也就是说自然连接是去掉重复列的等值连接。

【例 2-10】下面有两个关系 R 和 S。

A	B	C
a1	b1	5
a1	b2	6
a2	b3	8
a2	b4	12

B	E
b1	3
b2	7
b3	10
b3	2
b5	2

- (1) 查询关系 R 中属性 C 小于关系 S 中属性 E 的连接。
- (2) 查询关系 R 中属性 B 与关系 S 中属性 B 相等的等值连接。
- (3) 查询关系 R 中属性 B 与关系 S 中属性 B 相等的自然连接。

解：

(1)

A	R.B	C	S.B	E
a1	b1	5	b2	7
a1	b1	5	b3	10
a1	b2	6	b2	7
a1	b2	6	b3	10
a2	b3	8	b3	10

(2)

A	R.B	C	S.B	E
a1	b1	5	b1	3
a1	b2	6	b2	7
a2	b3	8	b3	10
a2	b3	8	b3	2

(3)

A	B	C	E
a1	b1	5	3
a1	b2	6	7
a2	b3	8	10
a2	b3	8	2

4. 除法

除法运算是二目运算，设有关系 $R(X,Y)$ 与关系 $S(Y,Z)$ ，其中 X 、 Y 、 Z 为属性集合， R 中的 Y 与 S 中的 Y 可以有不同属性名，但对应属性必须出自相同的域。关系 R 除以关系 S 所得的商是一个新关系 $T(X)$ ， T 是 R 中满足下列条件的元组在 X 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：

$$R \div S = \{t_r[x] \mid t_r \in R \wedge \Pi_y(S) \subseteq Y_x\}$$

其中 Y_x 为 X 在 R 中的象集， $X = [X]$ 。

【例 2-11】已知关系 R 和 S ，求 $R \div S$ 。

R

A	B	C
a1	b2	c3
a1	b2	c4
a2	b2	c3
a2	b2	c5

S

C
c3
c4

解：

$R \div S$

A	B
a1	b2

【例 2-12】查询选修了全部课程的学生的学号。

$\Pi_{\text{学号, 课程号}}(\text{选修}) \div \Pi_{\text{课程号}}(\text{课程})$

学号
201501

习 题

一、选择题

1. 关系数据库管理系统应能实现的专门的关系运算包括 ()。
A. 排序、索引、统计 B. 选择、投影、连接
C. 关联、更新、排序 D. 显示、打印、制表
2. 关系模型中，一个关键字是 ()。
A. 可以由多个任意属性组成
B. 至多由一个属性组成
C. 可以由一个或多个其值能唯一标识该关系模式中任何元组的属性组成
D. 以上都不是
3. 在一个关系中如果有这样一个属性存在，它的值能唯一地标识关系中的每一个元组，称这个属性为 ()。
A. 关键字 B. 数据项 C. 主属性 D. 主属性值
4. 在关系代数的传统集合运算中，假设有关系 R 和关系 S，运算结果为 W。如果 W 中的元组属于 R，或者属于 S，则 W 为 () 运算的结果。如果 W 中的元组属于 R 而不属于 S，则 W 为 () 运算的结果。如果 W 中的元组既属于 R 又属于 S，则 W 为 () 运算的结果。
A. 笛卡尔积 B. 并 C. 差 D. 交
5. 在关系代数的专门关系运算中，从表中取出满足条件的属性的操作称为 ()；从表中选出满足条件的元组的操作称为 ()；将两个关系中具有共同属性值的元组连接到一起构成新表的操作称为 ()。
A. 选择 B. 投影 C. 连接 D. 扫描
6. 自然连接是构成新关系的有效方法。一般情况下，当对关系 R 和 S 使用自然连接时，要求 R 和 S 含有一个或多个共有的 ()。
A. 元组 B. 行 C. 记录 D. 属性
7. 关系模式的任何属性 ()。
A. 不可再分 B. 可再分
C. 命名在该关系模式中可以不唯一 D. 以上都不是
8. 在关系代数中，五种基本运算为 ()。
A. 并、差、选择、投影、自然连接 B. 并、差、交、选择、投影
C. 并、差、选择、投影、笛卡尔积 D. 并、差、交、选择、乘积
9. 在一个关系模式中，() 是用于唯一标识一条记录的表关键字。
A. 主关键字 B. 外关键字 C. 公共关键字 D. 候选关键字

二、计算题

1. 设有如下所示的关系 R 和 S，计算：

- (1) $R1=R-S$ (2) $R2=R \cup S$
 (3) $R3=R \cap S$ (4) $R4=R \times S$

A	B	C
a	b	c
b	a	f
c	b	d

A	B	C
b	a	f
d	a	d

2. 设有如下所示的关系 R 和 S, 计算:

- (1) $R1=R \bowtie S$
 (2) $R2=R_{B<D} \bowtie S$
 (3) $R3=\sigma_{B=D}(R \times S)$

A	B	C
3	6	7
4	5	7
7	2	3
4	4	3

C	D	E
3	4	5
7	2	3