# 主要内容

- 2.1 关系数据结构及形式化定义
- 2.2 关系操作
- 2.3 关系的完整性
- 2.4 关系代数

# 2.4 关系代数

- 2.4.1 传统的集合运算 并、差、交、笛卡尔积 — 行
- 2.4.2 专门的关系运算 选择、投影、连接、除 — 行和列

| 运 算 符 | | 含 义 |
|---|---|---|
| 集合运算符 | ∪ | 并 |
| | - | 差 |
| | ∩ | 交 |
| | × | 笛卡尔积 |
| 专门的关系运算符 | $\sigma$ | 选择 |
| | $\pi$ | 投影 |
| | $\bowtie$ | 连接 |
| | ÷ | 除 |

# 并（union）

- $R \cup S$
  - 仍为$n$目关系，由属于$R$或属于$S$的元组组成

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

**R**

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

**S**

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

**RUS**

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

# 差（except）

- $R - S$
  - 仍为$n$目关系，由属于$R$而不属于$S$的所有元组组成

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

**R**

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

**S**

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

**R-S**

| A | B | C |
|---|---|---|
| a1 | b1 | c1 |

# 交（intersection）

- $R \cap S$
  - 仍为$n$目关系，由既属于$R$又属于$S$的元组组成

$$R \cap S = \{ t \mid t \in R \wedge t \in S \} \quad R \cap S = R - (R - S)$$

**R**

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

**S**

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

**R ∩ S**

| A | B | C |
|---|---|---|
| a1 | b2 | c2 |
| a2 | b2 | c1 |

# 笛卡尔积（cartesian product）

- $R \times S$
  - 列：（$n+m$）列元组的集合
    - 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

$$R \times S = \{t_r\, t_s \mid t_r \in R \wedge t_s \in S\}$$

R

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

S

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

$R \times S$

| R.A | R.B | R.C | S.A | S.B | S.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 |

- R， $t \in R$， $t[A_i]$

  设关系模式为$R(A_1, A_2, \ldots, A_n)$

  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]，A

若$A=\{A_{i1}, A_{i2}, \ldots, A_{ik}\}$，其中$A_{i1}$，$A_{i2}$，…，$A_{ik}$是$A_1$，$A_2$，…，$A_n$中的一部分，则$A$称为属性列或属性组。

$t[A]=(t[A_{i1}], t[A_{i2}], \ldots, t[A_{ik}])$表示元组$t$在属性列$A$上诸分量的集合。

$\bar{A}$则表示$\{A_1, A_2, \ldots, A_n\}$中去掉$\{A_{i1}, A_{i2}, \ldots, A_{ik}\}$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（Images Set）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

# 单击此处编辑母版标题样式

| $R$ | |
|---|---|
| $x_1$ | $Z_1$ |
| $x_1$ | $Z_2$ |
| $x_1$ | $Z_3$ |
| $x_2$ | $Z_2$ |
| $x_2$ | $Z_3$ |
| $x_3$ | $Z_1$ |
| $x_3$ | $Z_3$ |

- $x_1$在$R$中的象集

  $Z_{x1}$ ={$Z_1$，  $Z_2$，  $Z_3$}，

- $x_2$在$R$中的象集

  $Z_{x2}$ ={$Z_2$，  $Z_3$}，

- $x_3$在$R$中的象集

  $Z_{x3}$={$Z_1$，  $Z_3$}

学生-课程数据库:

学生关系Student、课程关系Course和选修关系SC

Student

| 学号<br>Sno | 姓名<br>Sname | 性别<br>Ssex | 年龄<br>Sage | 所在系<br>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 |

## 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 |

SC

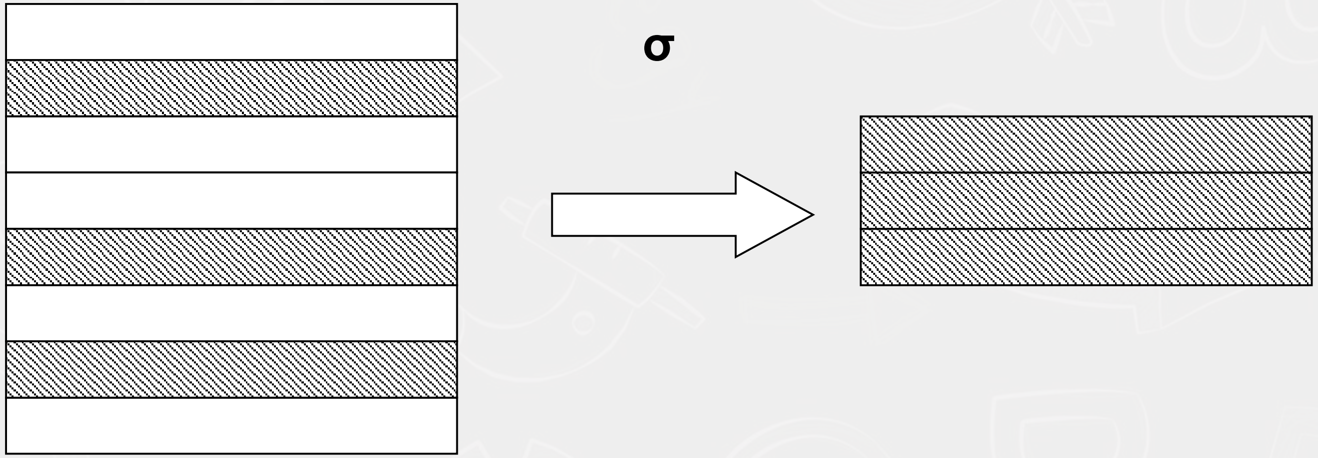
| 学号<br>Sno | 课程号<br>Cno | 成绩<br>Grade |
|---|---|---|
| 201215121 | 1 | 92 |
| 201215121 | 2 | 85 |
| 201215121 | 3 | 88 |
| 201215122 | 2 | 90 |
| 201215122 | 3 | 80 |

# 选择（selection）

- 在关系$R$中选择满足给定条件的诸元组

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

- F：选择条件，是一个逻辑表达式，取值为“真”或“假”
  - 基本形式为：$X_1\theta Y_1$
  - θ表示比较运算符，它可以是＞，≥，＜，≤，＝或<>



- 查询信息系（IS系）全体学生。

$$\sigma_{Sdept\ =\ 'IS'}(Student)$$

| Sno | Sname | Ssex | Sage | Sdept |
|---|---|---|---|---|
| 201215125 | 张立 | 男 | 19 | IS |

- 查询年龄小于20岁的学生。

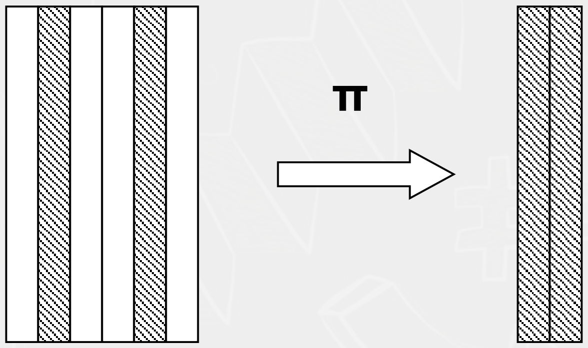
$$\sigma_{Sage < 20}(Student)$$

| Sno | Sname | Ssex | Sage | Sdept |
| --- | --- | --- | --- | --- |
| 201215122 | 刘晨 | 女 | 19 | IS |
| 201215123 | 王敏 | 女 | 18 | MA |
| 201215125 | 张立 | 男 | 19 | IS |

- 从$R$中选择出若干属性列组成新的关系

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

$A$ ：$R$中的属性列




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）

- 查询学生的姓名和所在系。

即求Student关系上学生姓名和所在系两个属性上的投影

$\pi_{Sname,Sdept}(Student)$

| Sname | Sdept |
|---|---|
| 李勇 | CS |
| 刘晨 | CS |
| 王敏 | MA |
| 张立 | IS |

- 查询学生关系Student中都有哪些系。

$$\pi_{Sdept}(Student)$$

| Sdept |
|---|
| CS |
| IS |
| MA |

# 连接（join）

- 连接也称为θ连接
  - 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

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

*A*和*B*：分别为*R*和*S*上度数相等且可比的属性组

θ：比较运算符

等值连接
equijoin

自然连接
natural join

- 等值连接（equijoin）
  - 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：

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

- 自然连接（Natural join）

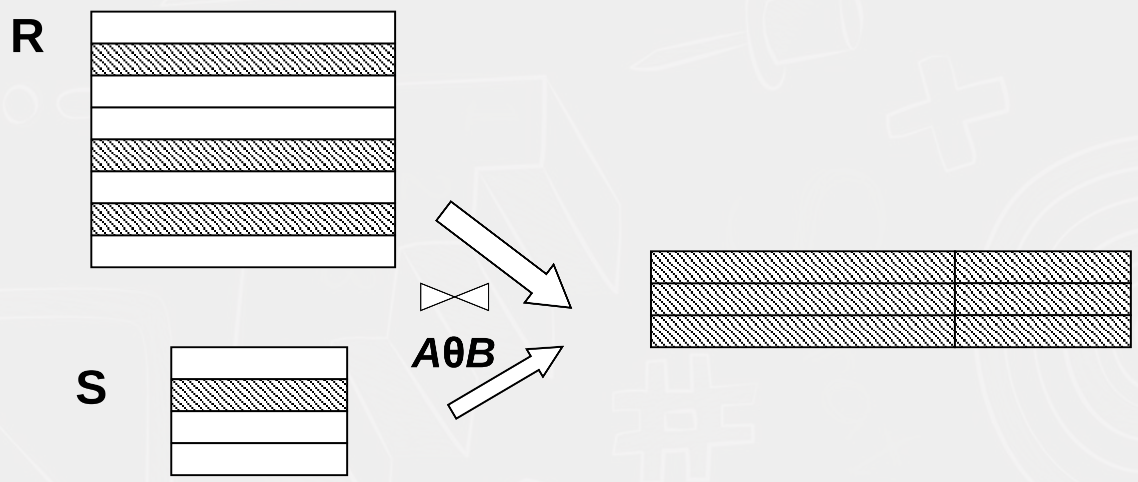
  自然连接是一种特殊的等值连接

  - 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$B$

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

- 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。

**R**

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

**S**

| B | E |
|---|---|
| b1 | 3 |
| b2 | 7 |
| b3 | 10 |
| b3 | 2 |
| b2 | 2 |

$$R \underset{C<E}{\bowtie} S$$

| 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 |

**R**

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

**S**

| B | E |
|---|---|
| b1 | 3 |
| b2 | 7 |
| b3 | 10 |
| b3 | 2 |
| b2 | 2 |

$$R \underset{R.B=S.B}{\bowtie} S$$

| 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 |

**R**

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

**S**

| B | E |
|---|---|
| b1 | 3 |
| b2 | 7 |
| b3 | 10 |
| b3 | 2 |
| b2 | 2 |

$R \bowtie S$

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

- 外连接（Outer Join）
  - 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
  - 左外连接(LEFT OUTER JOIN或LEFT JOIN)
    - 只保留左边关系$R$中的悬浮元组
  - 右外连接(RIGHT OUTER JOIN或RIGHT JOIN)
    - 只保留右边关系$S$中的悬浮元组

R

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

S

| B | E |
|---|---|
| b1 | 3 |
| b2 | 7 |
| b3 | 10 |
| b3 | 2 |
| b2 | 2 |

*R*和*S*的外连接

| A | B | C | E |
|---|---|---|---|
| a1 | b1 | 5 | 3 |
| a1 | b2 | 6 | 7 |
| a2 | b3 | 8 | 10 |
| a2 | b3 | 8 | 2 |
| a2 | b4 | 12 | NULL |
| NULL | b5 | NULL | 2 |

R

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

S

| B | E |
| --- | --- |
| b1 | 3 |
| b2 | 7 |
| b3 | 10 |
| b3 | 2 |
| b2 | 2 |

*R*和*S*的左外连接

| A | B | C | E |
| --- | --- | --- | --- |
| a1 | b1 | 5 | 3 |
| a1 | b2 | 6 | 7 |
| a2 | b3 | 8 | 10 |
| a2 | b3 | 8 | 2 |
| a2 | b4 | 12 | NULL |

R

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

S

| B | E |
|---|---|
| b1 | 3 |
| b2 | 7 |
| b3 | 10 |
| b3 | 2 |
| b2 | 2 |

*R*和*S*的右外连接

| A | B | C | E |
|---|---|---|---|
| a1 | b1 | 5 | 3 |
| a1 | b2 | 6 | 7 |
| a2 | b3 | 8 | 10 |
| a2 | b3 | 8 | 2 |
| NULL | b5 | NULL | 2 |

# 除（division）

- 给定关系R (X， Y) 和S (Y， Z)，其中$X$，$Y$，$Z$为属性组。

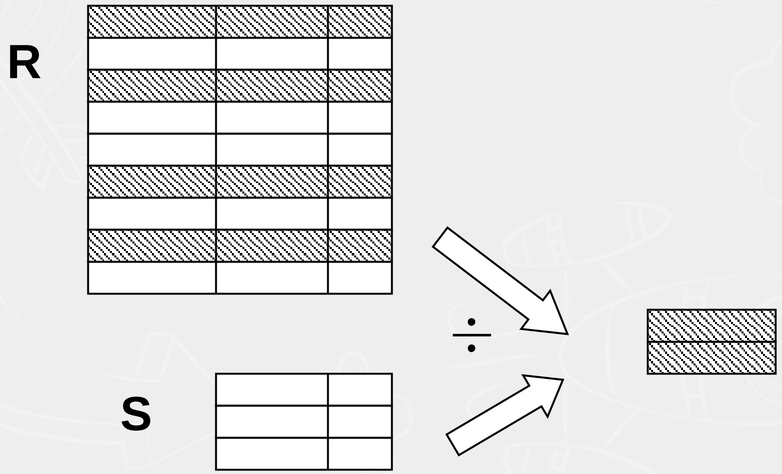
R与S的除运算得到一个新的关系P(X)，

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：

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

$Y_x$：$x$在$R$中的象集，$x = t_r[X]$



**R**

| A | B | C |
|---|---|---|
| a1 | b1 | c2 |
| a2 | b3 | c7 |
| a3 | b4 | c6 |
| a1 | b2 | c3 |
| a4 | b6 | c6 |
| a2 | b2 | c3 |
| a1 | b2 | c1 |

**S**

| B | C | D |
|---|---|---|
| b1 | c2 | d1 |
| b2 | c1 | d1 |
| b2 | c3 | d2 |

**R÷S**

| A |
|---|
| a1 |

- 在关系R中，A可取四个值{a1，a2，a3，a4}
  - $a_1$的象集为 $\{(b_1，c_2)，(b_2，c_3)，(b_2，c_1)\}$
  - $a_2$的象集为 $\{(b_3，c_7)，(b_2，c_3)\}$
  - $a_3$的象集为 $\{(b_4，c_6)\}$
  - $a_4$的象集为 $\{(b_6，c_6)\}$
- $S$在$(B，C)$上的投影为

  $\{(b1，c2)，(b2，c1)，(b2，c3)\}$
- 只有$a_1$的象集包含了$S$在$(B，C)$属性组上的投影

  所以: $R \div S = \{a_1\}$

- 查询至少选修1号课程和3号课程的学生号码
  - 建立一个临时关系K：

| Cno |
| --- |
| 1 |
| 3 |

  - 再求：$\pi_{Sno,Cno}(SC) \div K$

$\pi_{Sno,Cno}(SC)$

| Sno | Cno |
| --- | --- |
| 201215121 | 1 |
| 201215121 | 2 |
| 201215121 | 3 |
| 201215122 | 2 |
| 201215122 | 3 |

201215121象集{1，2，3}

201215122象集{2，3}

K={1，3}

于是：$\pi_{Sno,Cno}(SC) \div K=\{201215121\}$

- 查询选修了2号课程的学生的学号

$$\pi_{Sno}(\sigma_{Cno='2'}(SC))=\{201215121,201215122\}$$

- 查询至少选修了一门其直接先行课为5号课程的学生姓名

$$\pi_{Sname}(\sigma_{Cpno='5'}(Course \bowtie SC \bowtie \pi_{Sno,Sname}(Student))$$

$$\pi_{Sname}(\pi_{Sno}(\sigma_{Cpno='5'}(Course) \bowtie SC) \bowtie \pi_{Sno,Sname}(Student))$$

- 查询选修了全部课程的学生号码和姓名

$$\pi_{Sno,Cno}(SC) \div \pi_{Cno}(Course) \bowtie \pi_{Sno,Sname}(Student)$$