

第二章 关系数据库

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

2.5 *关系演算

2.6 小结

2.4 关系代数

- 关系代数是一种抽象的查询语言，它用对关系的运算来表达查询
- 关系代数
 - 运算对象是关系
 - 运算结果亦为关系
 - 关系代数的运算符有两类：集合运算符和专门的关系运算符
- 传统的集合运算是从关系的“水平”方向即行的角度进行
- 专门的关系运算不仅涉及行而且涉及列

2.4 关系代数（续）

表2.4 -2.5 关系代数与条件表达式运算符

运算符		含义	运算符		含义
集合运算符	\cup	并	比较运算符	$>$	大于
	$-$	差		\geq	大于等于
	\cap	交		$<$	小于
	\times	广义笛卡尔积		\leq	小于等于
				$=$	等于
				\neq 或 $<>$	不等于

2.4 关系代数（续）

表2.4-2.5 关系代数与条件表达式运算符

运算符		含义	运算符		含义
专门的 关系运 算符	σ	选择	逻辑 运算符	\neg	非
	π	投影		\wedge	与
	\div	除		\vee	或
	\bowtie	连接			

2.4 关系代数

2.4.1 传统的集合运算

2.4.2 专门的关系运算

(1) 并 (Union)

● 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

● $R \cup S$

- 仍为 n 目关系, 由属于 R 或属于 S 的元组组成

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

(1) 并 (续)

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

$R \cup S$

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

(2) 差 (Difference)

● R 和 S

- 具有相同的目 n
- 相应的属性取自同一个域

● $R - S$

并不要求 S 是 R 的子集

- 仍为 n 目关系，由属于 R 而不属于 S 的所有元组组成

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

(2) 差 (续)

R

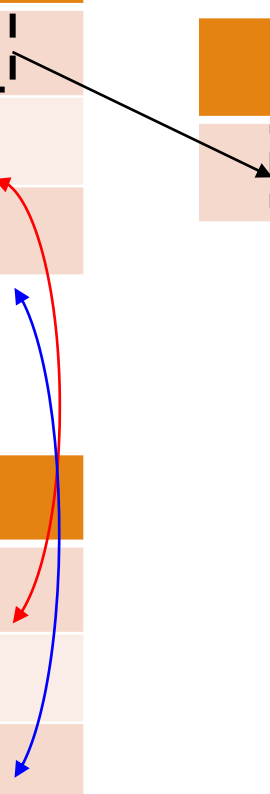
A	B	C
a1	b1	c1
a1	b2	c2
a2	b2	c1

S

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

R-S

A	B	C
a1	b1	c1



(3) 交 (Intersection)

- R 和 S

- 具有相同的目 n
- 相应的属性取自同一个域

- $R \cap S$

- 仍为 n 目关系，由既属于 R 又属于 S 的元组组成

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

$$R \cap S = R - (R - S)$$

(3) 交 (Intersection) (续)

R

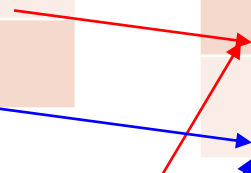
A	B	C
a1	b1	c1
a1	b2	c2
a2	b2	c1

S


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

$R \cap S$

A	B	C
a1	b2	c2
a2	b2	c1



(4) 笛卡儿积 (Cartesian Product)

- 严格地讲应该是广义的笛卡儿积 (Extended Cartesian Product)
- R : n 目关系, k_1 个元组
- S : m 目关系, k_2 个元组
- $R \times S$
 - 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
 - ✓ $R \times S = \{ \widehat{t_r t_s} \mid t_r \in R \wedge t_s \in S \}$ 

(4) 笛卡儿积 (续)

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

2.4 关系代数

2.4.1 传统的集合运算

2.4.2 专门的关系运算

2.4.2 专门的关系运算

先引入几个记号

(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

专门的关系运算（续）

(2) A , $t[A]$, \bar{A}

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

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

\bar{A} 则表示 $\{A_1, A_2, \dots, A_n\}$ 中去掉 $\{A_{i1}, A_{i2}, \dots, A_{ik}\}$ 后剩余的属性组。

例子

- 例:

关系 R (学号, 姓名, 性别, 年龄, 所在系)

- 若 $A = \{\text{学号}, \text{所在系}\}$

$\overline{A} = \{\text{姓名}, \text{性别}, \text{年龄}\}$



专门的关系运算（续）

(3) $\widehat{t_r t_s}$

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。

例：元组的连接

学生 R (学号, 姓名, 性别, 专业号, 年龄)

专业 S (专业号, 专业名)

学号	姓名	性别	专业号	年龄	专业号	专业名
801	张三	女	01	19	01	信息
802	李四	男	01	20	02	数学
803	王五	男	01	20	03	计算机
804	赵六	女	02	20		
805	钱七	男	02	19		

804	赵六	女	02	20	01	信息
-----	----	---	----	----	----	----

t_r t_s



专门的关系运算（续）

（4）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

专门的关系运算（续）

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

象集举例

- x_1 在 R 中的象集

$$Z_{x_1} = \{Z_1, Z_2, Z_3\},$$

- x_2 在 R 中的象集

$$Z_{x_2} = \{Z_2, Z_3\},$$

- x_3 在 R 中的象集

$$Z_{x_3} = \{Z_1, Z_3\}$$

专门的关系运算（续）

1) 选择

2) 投影

3) 连接

4) 除运算

专门的关系运算（续）

学生-课程数据库:

学生关系Student、课程关系Course和选修关系SC

P52

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)

例2.4

例2.5

例2.6

例2.7

专门的关系运算（续）

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)

专门的关系运算（续）

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)

例2.10

1) 选择 (Selection)

- 选择又称为限制 (Restriction)

- 选择运算符的含义

- 在关系 R 中选择满足给定条件的诸元组

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

- 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 $>$, \geq , $<$, \leq , $=$ 或 \neq

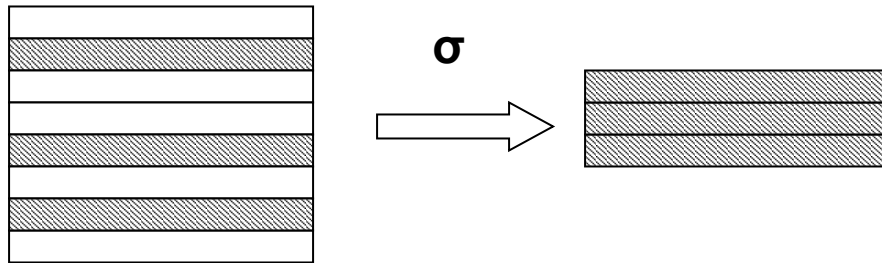
- ✓ 例子—— F : $X_1 \theta Y_1 \wedge X_2 \theta Y_2 \vee (\neg X_3 \theta Y_3)$:

选择年龄大于15岁的男性同学和年龄不等于17岁的女性同学

$(\text{age} > 15 \wedge \text{Ssex} = \text{'男'}) \vee (\text{age} \neq 17 \wedge \text{Ssex} = \text{'女'})$

1) 选择（续）

- 选择运算是从关系 R 中选取使逻辑表达式 F 为真的元组，是从行的角度进行的运算



1) 选择 (续)

[例2.4-1] 查询信息系 (IS系) 全体学生。

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

结果:

Sno	Sname	Ssex	Sage	Sdept
20121512 5	张立	男	19	IS

[例2.4-2] 查询信息系 (IS系) 且性别为男性的全体学生。

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

结果:

Sno	Sname	Ssex	Sage	Sdept
20121512 5	张立	男	19	IS



1) 选择 (续)

[例2.5] 查询年龄小于20岁的学生。

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

结果：

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



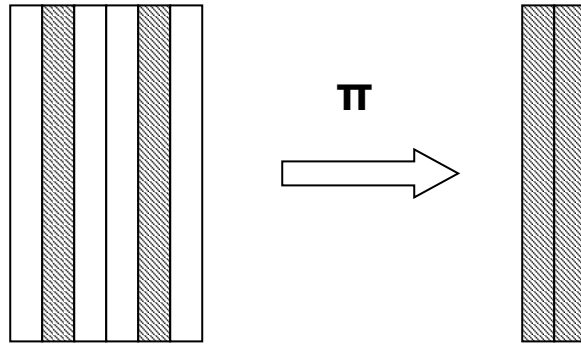
2) 投影 (Projection)

- 从 R 中选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）

2) 投影（续）

[例2.6] 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

$\Pi_{\text{Sname}, \text{Sdept}}(\text{Student})$

结果：

Sname	Sdept
李勇	CS
刘晨	CS
王敏	MA
张立	IS

如果原Student表中有两位同学都叫李勇，但学号不一样，且都来自CS系，请问该题的结果还是原答案么？



2) 投影（续）

[例2.7] 查询学生关系Student中都有哪些系。

$\Pi_{\text{Sdept}}(\text{Student})$

结果：

Sdept
CS
IS
MA



3) 连接 (Join)

- 连接也称为 θ 连接

- 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

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

- ✓ A 和 B : 分别为 R 和 S 上列数相等且可比的属性组

- ✓ θ : 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 A 属性组 (R 关系) 上的值与 B 属性组 (S 关系) 上的值满足比较关系 θ 的元组

3) 连接（续）

- 两类常用连接运算

- 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：

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

3) 连接（续）

➤ 自然连接（Natural join）

● 定义：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

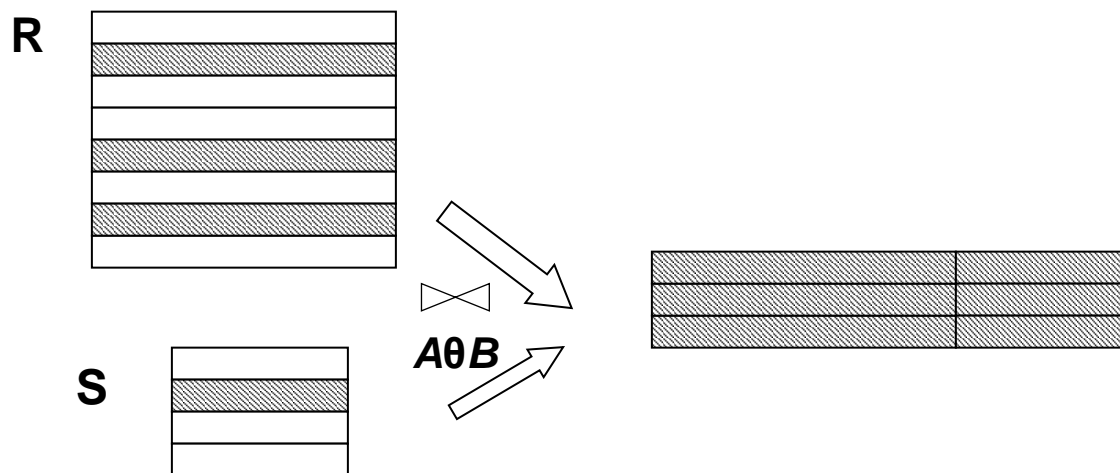
● 自然连接的含义

R 和 S 具有相同的属性组 B ， U 为 R 和 S 的全体属性集合

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

3) 连接 (续)

- 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。

3) 连接 (续)

[例2.8] 关系 R 和关系 S 如下所示:

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
b5	2

3) 连接（续）

一般连接 $R \bowtie_{C < E} S$ 的结果如下：

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

3) 连接（续）

等值连接 $R \bowtie S$ 的结果如下：

$$R.B=S.B$$

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) 连接 (续)

自然连接 $R \bowtie S$ 的结果如下:

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

3) 连接（续）

- 悬浮元组（Dangling tuple）

➤ 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。

3) 连接（续）

● 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

3) 连接（续）

下图是例2.8中关系***R***和关系***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

3) 连接 (续)

图(b)是例2.8中关系**R**和关系**S**的左外连接,图(c)是右外连接

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

图(b) 左外连接

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

图(c) 右外链接

4) 除 (Division)

要求：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。

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

Y_x : x 在 R 中的象集, $x = t_r[X]$

R 中，给定 x ，所有可能出现的 y 的集合

S 中，所有可能出现的 y 的集合

4) 除 (续)

[例2.9]

R

<i>A</i>	<i>B</i>	<i>C</i>
<i>a</i> ₁	<i>b</i> ₁	<i>c</i> ₂
<i>a</i> ₂	<i>b</i> ₃	<i>c</i> ₇
<i>a</i> ₃	<i>b</i> ₄	<i>c</i> ₆
<i>a</i> ₁	<i>b</i> ₂	<i>c</i> ₃
<i>a</i> ₄	<i>b</i> ₆	<i>c</i> ₆
<i>a</i> ₂	<i>b</i> ₂	<i>c</i> ₃
<i>a</i> ₁	<i>b</i> ₂	<i>c</i> ₁

<i>B</i>	<i>C</i>	<i>D</i>
<i>b</i> ₁	<i>c</i> ₂	<i>d</i> ₁
<i>b</i> ₂	<i>c</i> ₁	<i>d</i> ₁
<i>b</i> ₂	<i>c</i> ₃	<i>d</i> ₂

S

$R(X, Y)$ 和 $S(Y, Z)$
在本例中什么是 Y ?

$$R \div S = ?$$

分析:

在关系 R 中, A 可以取四个值 $\{a_1, a_2, a_3, a_4\}$

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

$\{(b_1, c_2), (b_2, c_1), (b_2, c_3)\}$

只有 a_1 的象集包含了 S 在 (B, C) 属性组上的投影

所以 $R \div S = \{a_1\}$

思考: 1. 如果 a_1 的象集为

$\{(b_1, c_2), (b_2, c_3), (b_2, c_1), (b_4, c_6)\}$,
该题结果变化么? 还是 a_1 么?

2. S 中可不可以没有属性 Z ?

除的另一个例子

A	B	C	D	E
a1	b1	5	d1	3
a1	b1	5	d2	7
a1	b2	5	d3	10
a2	b3	5	d3	2
a4	b5	6	d2	7
a4	b5	6	d3	10
a4	b5	6	d3	2
a6	b6	6	d5	2

÷

什么是Y?

C	D	E	F
6	d2	7	fa
6	d3	10	fb
6	d3	2	fc

除的另一个例子(续)

(A, B)



(C, D, E)



A	B	C	D	E
a1	b1	5	d1	3
a1	b1	5	d2	7
a1	b2	5	d3	10
a2	b3	5	d3	2
a4	b5	6	d2	7
a4	b5	6	d3	10
a4	b5	6	d3	2
a6	b6	6	d5	2

(a1,b1)的象集



C	D	E
5	d1	3
5	d2	7

(a1,b2)的象集



C	D	E
5	d3	10

(a4,b5)的象集



C	D	E
6	d2	7
6	d3	10
6	d3	2

(a2, b3)、
(a6, b6)的象集?



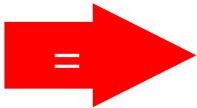
R 上分量值 X 的象集 Y_x 包含 S 在 Y 上的投影



A	B	C	D	E
a1	b1	5	d1	3
a1	b1	5	d2	7
a1	b2	5	d3	10
a2	b3	5	d3	2
a4	b5	6	d2	7
a4	b5	6	d3	10
a4	b5	6	d3	2
a6	b6	6	d5	2

÷

C	D	E	F
6	d2	7	fa
6	d3	10	fb
6	d3	2	fc

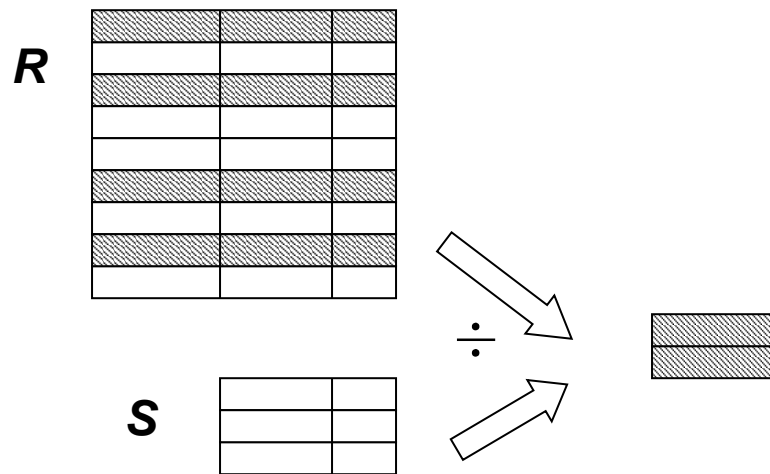


(a4,b5)

A	B
a4	b5

除运算（续）

- 除操作是同时从行和列角度进行运算



综合举例

以学生-课程数据库为例

[例2.10] 查询至少选修1号课程和3号课程的学生号码。

首先建立一个临时关系 K :

Cno
1
3

然后求: $\Pi_{Sno, Cno}(SC) \div K$



综合举例（续）

[例2.10]续

$\Pi_{\text{Sno}, \text{Cno}}(\text{SC})$

201215121象集{1, 2, 3}

201215122象集{2, 3}

$K = \{1, 3\}$

于是：

$\Pi_{\text{Sno}, \text{Cno}}(\text{SC}) \div K = \{201215121\}$

Sno	Cno
201215121	1
201215121	2
201215121	3
201215122	2
201215122	3

综合举例（续）

[例2.11] 查询选修了2号课程的学生学号。P52 图2.4

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

[例2.12] 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

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

$$\text{或 } \Pi_{Sname}(\sigma_{Cpno='5'}(Course \bowtie SC \bowtie Student))$$

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

[例2.13] 查询选修了全部课程的学生号码和姓名。

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

课堂练习

- 查询选修了全部直接先行课为6号课程的学生姓名

$$\Pi_{\text{Sname}}((\Pi_{\text{Sno}, \text{Cno}}(\text{SC}) \div \Pi_{\text{Cno}}(\sigma_{\text{Cpno}=6}(\text{Course}))) \bowtie \text{Student})$$

小结

● 关系代数运算

➤ 关系代数运算

✓ 并、差、交、笛卡尔积、投影、选择、连接、除

➤ 基本运算

✓ 并、差、笛卡尔积、投影、选择

➤ 交、连接、除

✓ 可以用5种基本运算来表达

✓ 引进它们并不增加语言的能力，但可以简化表达

小结（续）

- 关系代数表达式

- 关系代数运算经有限次复合后形成的式子

- 典型关系代数语言

- ISBL（Information System Base Language）

- ✓ 由IBM United Kingdom研究中心研制

- ✓ 用于PRTV（Peterlee Relational Test Vehicle）实验系统

第二章 关系数据库

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

2.5 *关系演算

2.6 小结

2.6 小结

- 关系数据库系统是目前使用最广泛的数据库系统
- 关系数据库系统与非关系数据库系统的区别：
 - 关系系统只有“表”这一种数据结构
 - 非关系数据库系统还有其他数据结构，以及对这些数据结构的操作

小结（续）

● 关系数据结构

➤ 关系

✓ 域

✓ 笛卡尔积

✓ 关系

- 关系，属性，元组
- 候选码，主码，主属性
- 基本关系的性质

➤ 关系模式

➤ 关系数据库

➤ 关系模型的存储结构

小结（续）

● 关系操作

➤ 查询

✓ 选择、投影、连接、除、并、交、差

➤ 数据更新

✓ 插入、删除、修改

小结（续）

- 关系的完整性约束

- 实体完整性

- 参照完整性

- ✓ 外码

- 用户定义的完整性

小结（续）

- 关系数据语言

- 关系代数语言

- 关系演算语言

- ✓ 元组关系演算语言 ALPHA

- ✓ 域关系演算语言 QBE