

2. 差 (Difference)

- R 和 S
 - 具有相同的目 n
 - 相应的属性取自同一个域
- $R - S$
 - 仍为 n 目关系，由属于 R 而不属于 S 的所有元组组成

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



差运算

R

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

S

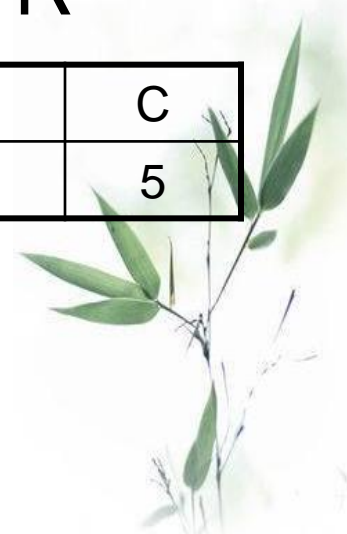
A	B	C
3	4	5
7	2	3

R—S

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

S—R

A	B	C
3	4	5

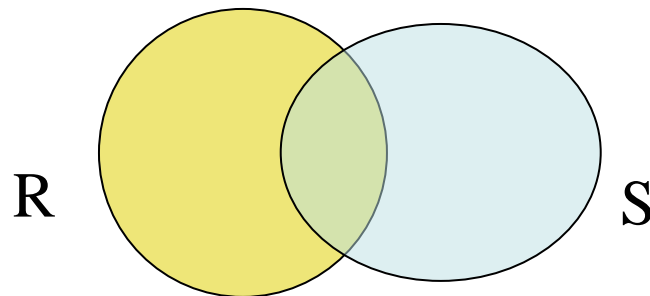


3. 交 (Intersection)

- 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

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

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



交运算

R

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

S

A	B	C
3	4	5
7	2	3

$R \cap S$

A	B	C
7	2	3



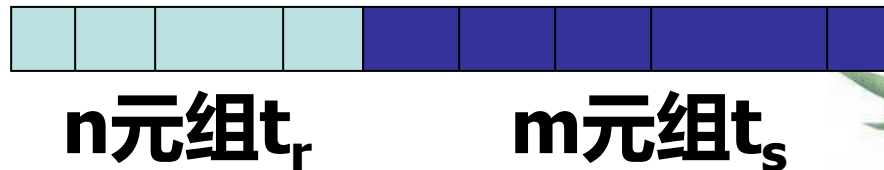
表示记号

- $\widehat{t_r t_s}$

- 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。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

$(n+m)$ 元组 $\widehat{t_r t_s}$



4. 笛卡尔积 (Cartesian Product)

- 严格地讲应该是广义的笛卡尔积 (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
 - $R \times S = \{ \overbrace{t_r t_s} \mid t_r \in R \wedge t_s \in S \}$



笛卡尔积(续)

R		
A	B	C
a_1	b_1	c_1
a_1	b_2	c_2
a_2	b_2	c_1

S		
A	B	C
a_1	b_2	c_2
a_1	b_3	c_2
a_2	b_2	c_1

$R \times S$					
$R.A$	$R.B$	$R.C$	$S.A$	$S.B$	$S.C$
a_1	b_1	c_1	a_1	b_2	c_2
a_1	b_1	c_1	a_1	b_3	c_2
a_1	b_1	c_1	a_2	b_2	c_1
a_1	b_2	c_2	a_1	b_2	c_2
a_1	b_2	c_2	a_1	b_3	c_2
a_1	b_2	c_2	a_2	b_2	c_1
a_2	b_2	c_1	a_1	b_2	c_2
a_2	b_2	c_1	a_1	b_3	c_2
a_2	b_2	c_1	a_2	b_2	c_1

Relations r , s :

A	B
-----	-----

α	1
β	2

r

C	D	E
-----	-----	-----

α	10	a
β	10	a
β	20	b
γ	10	b

s

$r \times s$:

A	B	C	D	E
-----	-----	-----	-----	-----

α	1	α	10	a
α	1	β	19	a
α	1	β	20	b
α	1	γ	10	b
β	2	α	10	a
β	2	β	10	a
β	2	β	20	b
β	2	γ	10	b



2.4 关系代数

- 概述
- 传统的集合运算
- 专门的关系运算



专门的关系运算(续)

- 选择
- 投影
- 连接
- 除



➤图为学生-课程数据库中的student关系、Course关系、SC关系

SC:

Sno	Cno	Grade
200215121	1	92
200215121	2	85
200215121	3	88
200215122	2	90
200215122	3	80

Student:

Sno	Sname	Ssex	Sage	Sdept
200215121	李勇	男	20	CS
200215122	刘晨	女	19	IS
200215123	王敏	女	18	MA
200215125	张立	男	19	IS

Course:

Cno	Cname	Cpno	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

1.选择(Selection)

- 问题：如何查询信息系(IS)的全体学生信息

选择满足条件：Sdept='IS'的元组

学号 Sno	姓名 Sname	性别 Ssex	年龄 Sage	所在系 Sdept
200215121	李勇	男	20	CS
200215122	刘晨	女	19	IS
200215123	王敏	女	18	MA
200215125	张立	男	19	IS

选择(续)

- 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

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

- F : 选择条件, 是一个逻辑表达式, 基本形式为:

$X_1 \theta Y_1$

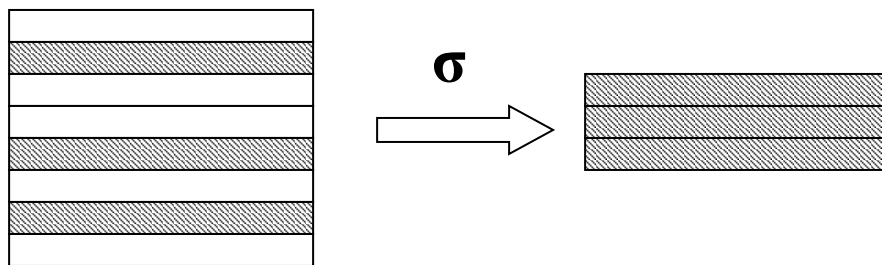
比较运算符, 如 $>$, \geq ,
 $<$, \leq , $=$, $<>$



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



• [例1] 查询信息系

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

或 $\sigma_5='IS' (Student)$

选

学号 Sno	姓名 Sname	性别 Ssex	年龄 Sage	所在系 Sdept
200215121	李勇	男	20	CS
200215122	刘晨	女	19	IS
200215123	王敏	女	18	MA
200215125	张立	男	19	IS

□ 结果:

属性名可以用属性序号代替

Sno	Sname	Ssex	Sage	Sdept
95002	刘晨	女	19	IS
95004	张立	男	19	IS



• [例2] 查询年龄小

$\sigma_{\text{Sage} < 20}(\text{Student})$
或 $\sigma_{4 < 20}(\text{Student})$

学号 Sno	姓名 Sname	性别 Ssex	年龄 Sage	所在系 Sdept
200215121	李勇	男	20	CS
200215122	刘晨	女	19	IS
200215123	王敏	女	18	MA
200215125	张立	男	19	IS

□ 结果:

Sno	Sname	Ssex	Sage	Sdept
95002	刘晨	女	19	IS
95003	王敏	女	18	MA
95004	张立	男	19	IS



• [例] 查询年龄小于

$\sigma_{Sage < 20 \wedge Ssex = '女'}$

或 $\sigma_{4 < 20 \wedge 3 = '女'} (Student)$

学号 Sno	姓名 Sname	性别 Ssex	年龄 Sage	所在系 Sdept
200215121	李勇	男	20	CS
200215122	刘晨	女	19	IS
200215123	王敏	女	18	MA
200215125	张立	男	19	IS

□ 结果:

Sno	Sname	Ssex	Sage	Sdept
95002	刘晨	女	19	IS
95003	王敏	女	18	MA

——选择是在关系R中选择满足给定条件的诸元组

2.投影 (Projection)

- **问题：**如何查询全部学生的姓名及其所在系信息

显示学生关系中的姓名和系属性列

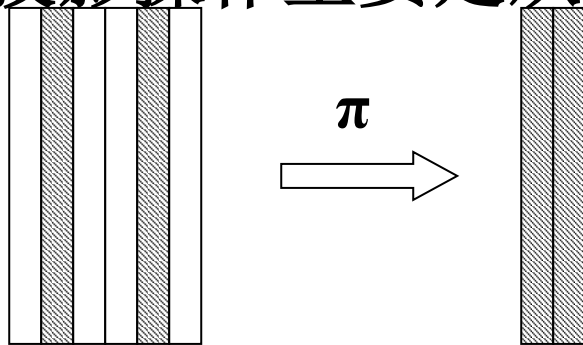
学号 Sno	姓名 Sname	性别 Ssex	年龄 Sage	所在系 Sdept
200215121	李勇	男	20	CS
200215122	刘晨	女	19	IS
200215123	王敏	女	18	MA
200215125	张立	男	19	IS

投影(续)

- 投影运算符的含义：从 R 中选择出若干属性列组成新的关系

$$\pi_A(R) = \{ t[A] \mid t \in R \} \text{ ---其中 } A \text{ 为 } R \text{ 中的属性列}$$

- 投影操作主要是从列的角度进行运算



■但投影之后可能取消某些元组(避免重复行)



投影(续)

- [例3] 查询学生的姓名和所在系

即求Student关系上Sname和Sdept两个属性上的投影

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

或 $\pi_{2, 5}(\text{Student})$

属性名
可以用
属性序
号代替

□ 结果:

Sname	Sdept
李勇	CS
刘晨	IS
王敏	MA
张立	IS

学号 Sno	姓名 Sname	性别 Ssex	年龄 Sage	所在系 Sdept
200215121	李勇	男	20	CS
200215122	刘晨	女	19	IS
200215123	王敏	女	18	MA
200215125	张立	男	19	IS

• [例4] 查询学生关系

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

学号 Sno	姓名 Sname	性别 Ssex	年龄 Sage	所在系 Sdept
200215121	李勇	男	20	CS
200215122	刘晨	女	19	IS
200215123	王敏	女	18	MA
200215125	张立	男	19	IS

□ 结果:

Sdept
CS
IS
MA

投影之后取消了原关系中的某些元组
(避免重复行)



➤图为学生-课程数据库中的student关系、Course关系、SC关系

SC:

Sno	Cno	Grade
200215121	1	92
200215121	2	85
200215121	3	88
200215122	2	90
200215122	3	80

Student:

Sno	Sname	Ssex	Sage	Sdept
200215121	李勇	男	20	CS
200215122	刘晨	女	19	IS
200215123	王敏	女	18	MA
200215125	张立	男	19	IS

Course:

Cno	Cname	Cpno	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

选择投影运算复合使用示例

- 复合运用投影、选择、笛卡尔积运算，可以从任意n张表中截取满足条件的子表

例 查询选修了2号课程的学生学号。

$$\pi_{Sno} (\sigma_{Cno = '2'} (SC)) = \{ 200215121, 200215122 \}$$

例 查询选过“数据库”课程学生学号。

$$\pi_{Sno} (\sigma_{sc.cno=course.cno \wedge Cname = '数据库'} (SC \times Course))$$

解：需查询的数据要根据“选课”信息以及“课程”

信息来完成，因此，查询涉及SC和COURSE两张表

step1: 先将“选课”与“课程”表合并为一张表

SC × Course，结果为所有选过任意课程的信息

step2: 再从上一步结果中，对于任意一次选课，所选课程

存在sc.cno=course.cno，且选的课程是“数据库”即Cname=‘数据库’的元组



笛卡尔连接运算存在的问题

- 问题示例 观察上例中表达式 $SC \times Course$ ，其结果中包含所有的选课与所有课程的组合，而实际中，一个学生不可能选所有课程，这样结果元组中许多是无意义的。上例对200215121学生只有第一个元组有意义

SNO	CNO	GRADE	CNO	CNAME	CPNO	CREDIT
200215121	1	92	1	数据库	5	4
200215121	1	92	2	数学	null	2
200215121	1	92	3	信息系统	1	4
200215121	1	92	4	操作系统	6	3
200215121	1	92	5	数据结构	7	4
200215121	1	92	6	数据处理	null	2
200215121	1	92	7	C 语言	6	4
...
200215121	2	92	1	数据库	5	4

连接运算

- **问题分析** 笛卡尔乘运算为了保证数学上的**完整性**，将两张表的‘**所有**’内容合并，若该两张表之间存在关联关系，则合并后结果中将那些**无关联**关系的元组也合并了。

数学上的“完整”在使用中不便，甚至造成“信息丢失”

- **问题解决**
 - 1) 通过在 σ_P 运算中确定适当的P，去除那些无关联的元组。如 `sc.cno=course.cno`
 - 2) 定义扩展的笛卡尔乘运算，在**合并时去除**无关联元组



连接 (续)

- 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 上度数相等且可比的属性组
- ◆ θ : 比较运算符



连接(续)

- 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 θ 的元组

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

■ 连接后，结果关系的关系模式由 R 和 S 的所有属性组成， R 的属性在前， S 的属性在后



连接(续)

- [例] 计算 SC  Course
SC. Cno < Course. Cno

关系SC

Sno	Cno	Grade
001	1	92
001	3	88
002	2	90
003	6	80

关系Course

Cno	Cname	Cpno	Ccredit
1	操作系统	6	3
3	数据结构	7	4
6	数据处理		2

----从SC × Course中选取SC关系中课程号小于Course关系中课程号值的元组



连接(续)

SC  Course
SC.Cno < Course.Cno

- 计算步骤

(1) 计算笛卡尔积 SC × Course, 结果为

Sno	SC.Cno	Grade	Course.Cno	Cname	Cpno	Ccredit
001	1	92	1	操作系统	6	3
001	1	92	3	数据结构	7	4
001	1	92	6	数据处理		2
001	3	88	1	操作系统	6	3
001	3	88	3	数据结构	7	4
001	3	88	6	数据处理		2
002	2	90	1	操作系统	6	3
002	2	90	3	数据结构	7	4
002	2	90	6	数据处理		2
003	6	80	1	操作系统	6	3
003	6	80	3	数据结构	7	4
003	6	80	6	数据处理		2

连接(续)

SC  Course
SC.Cno < Course.Cno

(2) 从SC × Course中选择满足SC.Cno < Course.Cno的元组

Sno	SC.Cno	Grade	Course.Cno	Cname	Cpno	Ccredit
001	1	92	1	操作系统	6	3
001	1	92	3	数据结构	7	4
001	1	92	6	数据处理		2
001	6	88	1	操作系统	6	3
001	3	88	3	数据结构	7	4
001	3	88	6	数据处理		2
002	2	90	1	操作系统	6	3
002	2	90	3	数据结构	7	4
002	2	90	6	数据处理		2
003	6	80	1	操作系统	6	3
003	6	80	3	数据结构	7	4
003	6	80	6	数据处理		2

连接(续)

SC  Course
SC. Cno < Course. Cno

(3) 结果为

对关系SC和Course进行连接的结果

Sno	SC. Cno	Grade	Course. Cno	Cname	Cpno	Ccredit
001	1	92	3	数据结构	7	4
001	1	92	6	数据处理		2
001	3	88	6	数据处理		2
002	2	90	3	数据结构	7	4
002	2	90	6	数据处理		2



连接(续)----例

计算 $R \bowtie_{1>1} S$

R

S

A	B	D
2	5	8
5	4	1
4	5	8
3	4	9

B	C	D
3	5	8
4	4	1
4	1	8
6	4	1

$R \bowtie_{1>1} S$

A	R. B	R. D	S. B	C	S. D
5	4	1	3	5	8
5	4	1	4	4	1
5	4	1	4	1	8
4	5	8	3	5	8

- ❖ 注意连接运算的执行情况!!
- ❖ 满足条件的元组 ----匹配的元组(保留)
- ❖ 不满足条件的元组 ----不匹配的元组(舍去)



连接(续)

- 常用的两类连接运算
 - 等值连接(equijoin)
 - 自然连接(Natural join)



连接(续)----等值连接

- 等值连接

- θ 为“=”的连接运算
- 含义：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

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



等值连接(续)--例

计算 $R \bowtie_{1=1} S$

R

A	B	D
2	5	8
7	4	1
4	5	8
3	4	9

S

B	C	D
3	5	8
4	4	1
4	1	8
6	4	1

$R \bowtie_{1=1} S$

A	R. B	R. D	S. B	C	S. D
4	5	8	4	4	1
4	5	8	4	1	8
3	4	9	3	5	8

1. $R \times S$

2. $\sigma_{R.A=S.B}(R \times S)$



连接(续)----自然连接

❖ 自然连接：一种特殊的等值连接

■ 两个关系中进行比较的分量须是相同的属性组

■ 在结果中把重复的属性列去掉

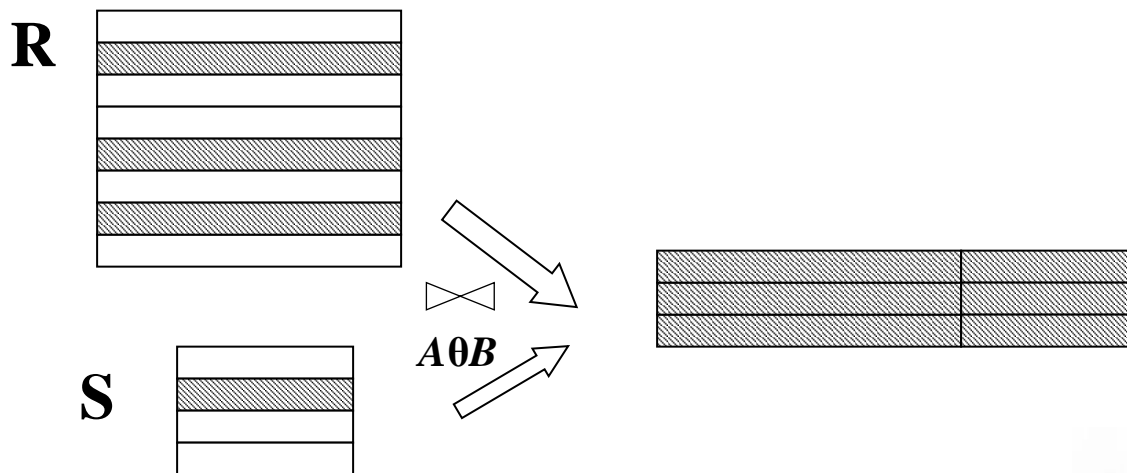
➤ 含义： R 和 S 具有相同的属性组 B

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



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。



自然连接(续)

- 自然连接运算的步骤
 - 计算笛卡尔积 $R \times S$
 - 从 $R \times S$ 中选择那些公共属性A的数值相同 ($\sigma_{R.A=S.A}$) 的元组
 - 去掉S.A(或R.A)，将留下来的R.A(或S.A)改为A，即得所要的结果



连接(续)

- 自然连接例：计算SC \bowtie Course
- 步骤：
 - (1) 计算SC \times Course

Sno	SC. Cno	Grade	Course. Cno	Cname	Cpno	Ccredit
001	1	92	1	操作系统	6	3
001	1	92	3	数据结构	7	4
001	1	92	6	数据处理		2
001	6	88	1	操作系统	6	3
001	3	88	3	数据结构	7	4
001	3	88	6	数据处理		2
002	2	90	1	操作系统	6	3
002	2	90	3	数据结构	7	4
002	2	90	6	数据处理		2
003	6	80	1	操作系统	6	3
003	6	80	3	数据结构	7	4
003	6	80	6	数据处理		2



自然连接例(续)

(2) 选择 $\sigma_{SC.Cno=Course.Cno}$ ($SC \times Course$)

Sno	SC. Cno	Grade	Course. Cno	Cname	Cpno	Ccredit
001	1	92	1	操作系统	6	3
001	3	88	3	数据结构	7	4
003	6	80	6	数据处理		2

自然连接例(续)

(3) 删除重复列SC.Cno，并将留下来的Course.Cno改为Cno，得到结果

进行SC ⋈ Course运算的结果

Sno	Grade	Cno	Cname	Cpno	Ccredit
001	92	1	操作系统	6	3
001	88	3	数据结构	7	4
003	80	6	数据处理		2



自然连接与等值连接

- 自然连接的两个关系要求有公共的属性组B，等值连接则不要求；
- 自然连接中等值的条件一定是公共属性组的值相等 ($R.B = S.B$)，而等值连接不一定；
- 自然连接中等值的条件隐含，不显式地写出来，而等值连接要写出来；
- 自然连接要在结果中去掉重复的一个属性组B，而等值连接则不。



自然连接运算----例

R和S有公共属性B, D

R

A	B	C	D
α	1	α	a
β	2	γ	a
γ	4	β	b
α	1	γ	a
δ	2	β	b

S

B	D	E
1	a	α
3	a	β
1	a	γ
2	b	δ
3	b	ϵ

结果中只有一个B, D

R \bowtie **S**

A	B	C	D	E
α	1	α	a	α
α	1	α	a	γ
α	1	γ	a	α
α	1	γ	a	γ
δ	2	β	b	δ



自然连接与笛卡尔积的区别

- 1) 两者的语义上都有合并两张表的作用

笛卡尔乘是数学意义上的所有可能组合的乘积，而自然连接则是将两张表中有关联关系的元组合并

例如 $SC \times Course$ 是所有可能的选课及所有课程信息

$SC \bowtie Course$ 则是所有的选课及所选课程的信息

- 2) 自然连接有选择 σ_p 的语义

例如 将两个表 SC 、 $Course$ 做自然连接 \bowtie ，可以理解为在 SC 中选择那些选过 “...课程” 的学生

示例 找出选过学分为4的学生号

$\sigma_{credit=4}(Course)$ // 找出学分为4的课程

$SC \bowtie \sigma_{credit=4}(Course)$ // 选出选过4分的选课

$\pi_{Sno} (SC \bowtie \sigma_{credit=4}(Course))$ // 选出选过4分的学号



假设 R_1 、 R_2 和 R_3 是三个关系

R_1

A	B	C
a ₁	b ₁	55
a ₂	b ₂	45
a ₄	b ₂	35

R_2

A	B	C
a ₁	b ₂	55
a ₂	b ₂	45
a ₃	b ₁	35
a ₄	b ₂	35

R_3

A	D
a ₁	101
a ₂	101
a ₂	102
a ₃	102

$R_1 \bowtie R_2$

$R_1.C < R_2.C$

$R_1 \bowtie R_2$

$R_1.C = R_2.C$

$R_1 \bowtie R_3$

