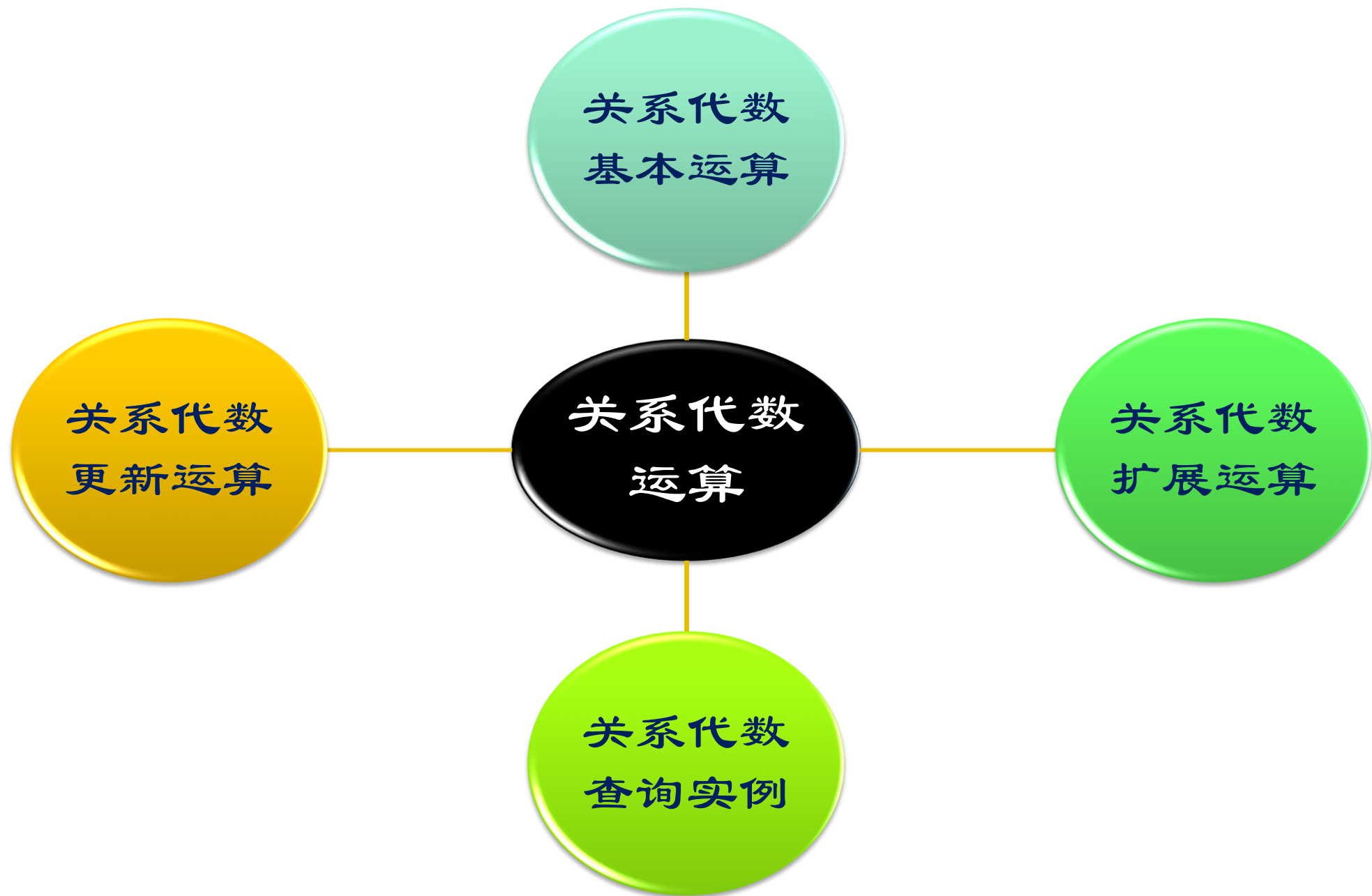
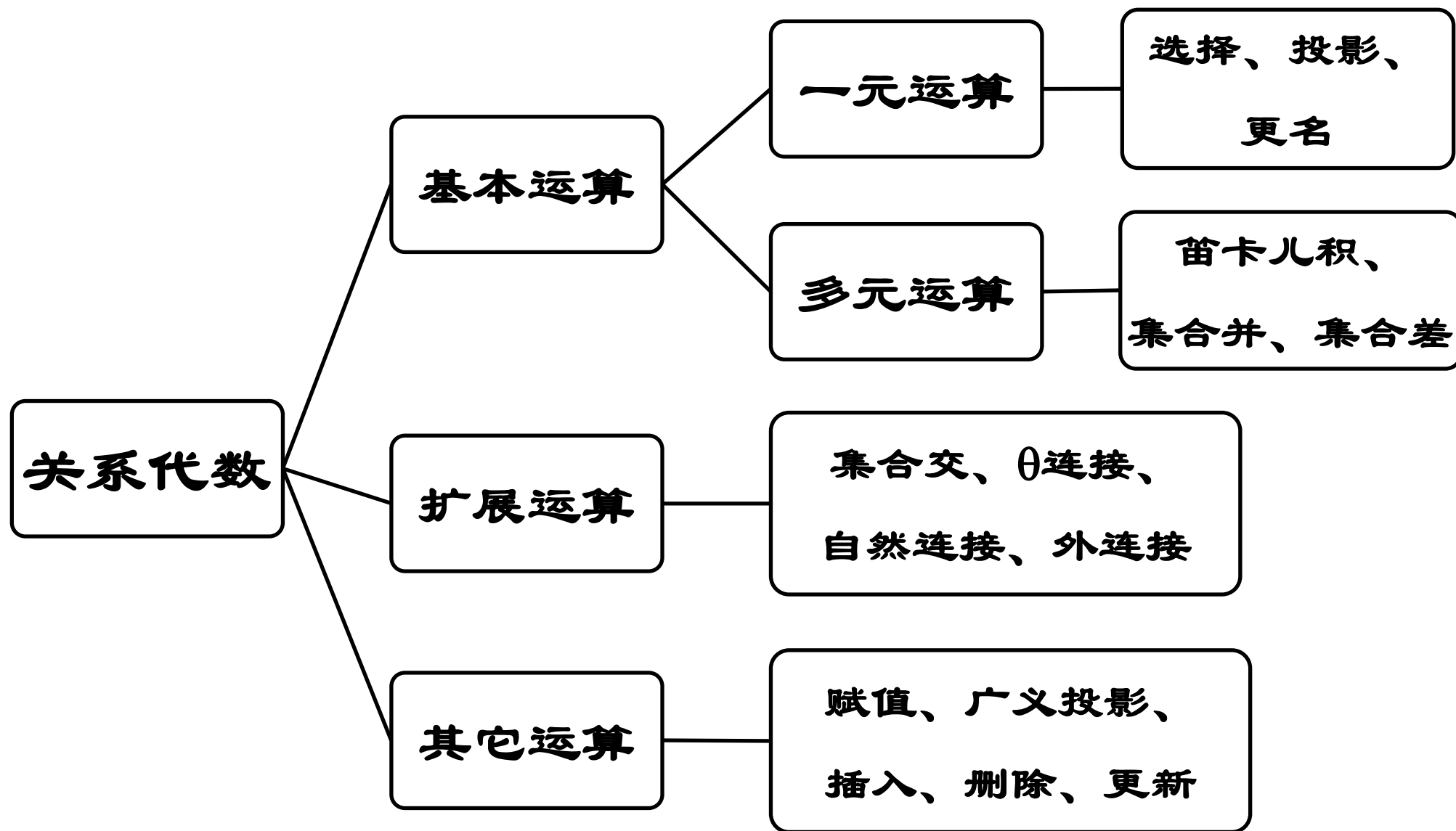


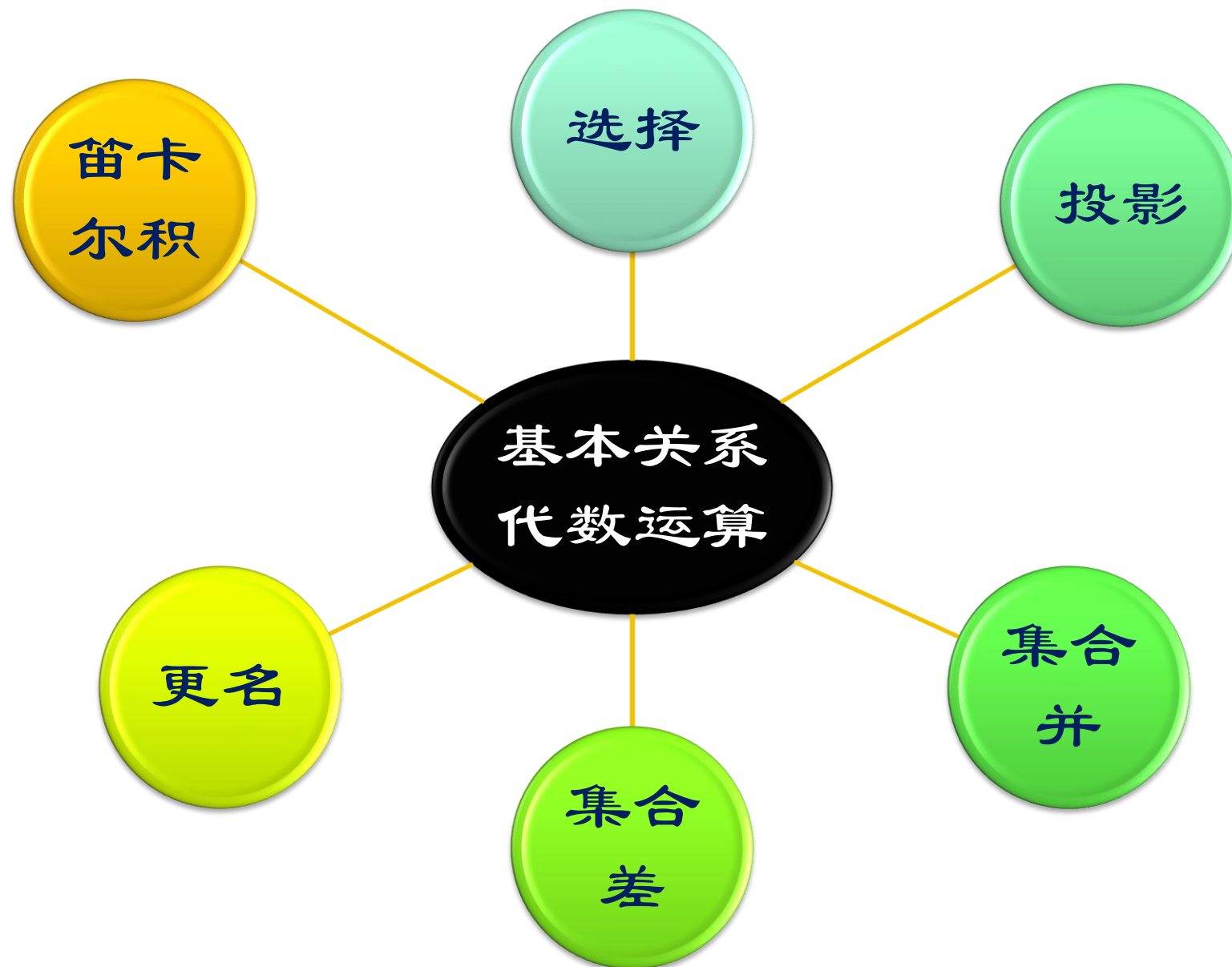
第 3 章

关系模型



关系代数运算汇览





选择运算

在关系中选择满足给定条件的元组（行角度）

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

F 是选择的条件， $\forall t \in R, F(t)$ 要么为真，要么为假

F 由逻辑运算符连接算术表达式而成

逻辑表达式： \wedge, \vee, \neg

算术表达式： $X \theta Y$

X, Y 是属性名、常量、或简单函数

θ 是比较算符， $\theta \in \{>, \geq, <, \leq, =, \neq\}$

选择运算计算例子

R

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

$\sigma_{A < 5}(R)$

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

$\sigma_{A < 5 \wedge C = 7}(R)$

A	B	C
3	6	7
2	5	7

用选择运算表达查询

找年龄不小于20的男学生

$$\sigma_{age \geq 20 \wedge sex = 'M'}(S)$$

$$\sigma_{age \geq 20}(\sigma_{sex = 'M'}(S))$$

$$\sigma_{sex = 'M'}(\sigma_{age \geq 20}(S))$$

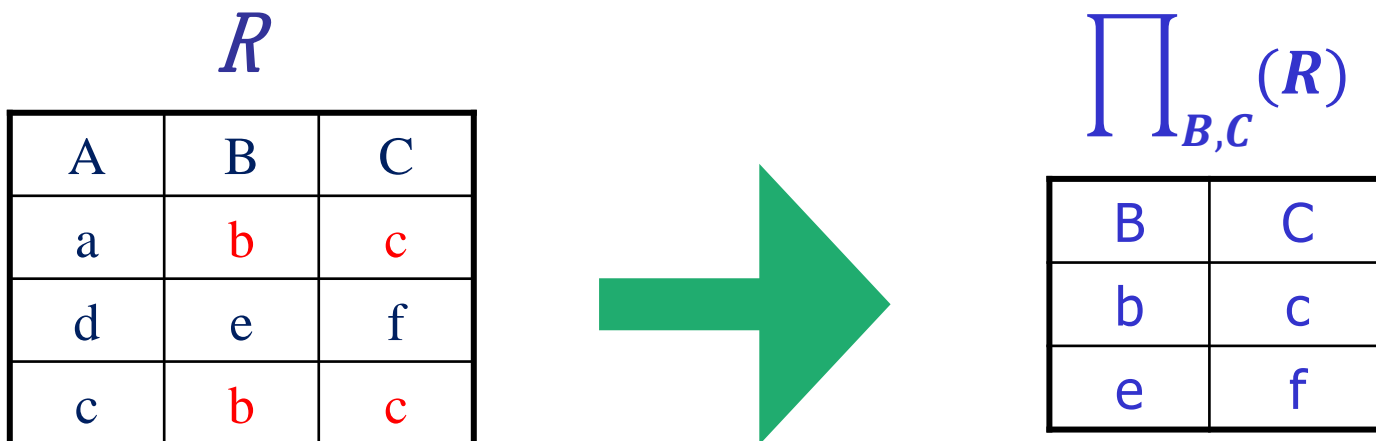
哪种执行方式更为高效，主要取决于数据分布和索引配置

投影运算

从关系中取若干列组成新的关系（从列的角度）

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

注意：投影的结果中要去掉相同的行



投影查询示例

给出所有学生的姓名和年龄

$$\Pi_{sno, age}(S)$$

找001号学生所选修的课程号

$$\Pi_{cno}(\sigma_{sno=001}(SC))$$

更名

将关系R更名为S

$$\rho_S(R)$$

将计算表达式E更名为关系S

$$\rho_{S(A_1, A_2, \dots, A_n)}(E)$$

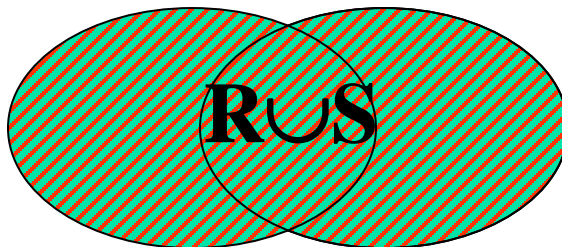
更名运算的必要性：

- 将更名运算施加到关系上，得到具有不同名字的同一关系
- 当同一关系多次参与同一运算时需要更名

并运算

所有至少出现在两个关系中之一的元组集合

$$R \cup S = \{ r \mid r \in R \vee r \in S \}$$



关系R和S进行并运算的前提是它们必须是相容的

- 关系R和S必须是同元的，其属性数目必须相同
- 对 $\forall i$ ，R的第i个属性和S的第i个属性的域必须相同

并运算计算例子

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

RUS

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

并运算查询示例

求选修了001号或002号课程的学生号

方案1:

$$\Pi_{sno}(\sigma_{cno=001 \vee cno=002}(SC))$$

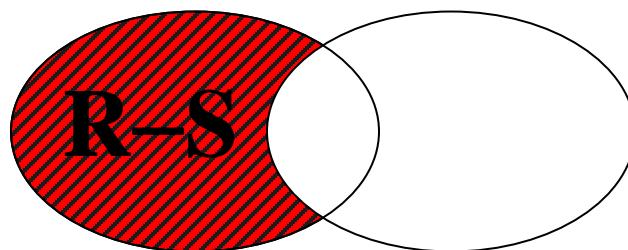
方案2:

$$\Pi_{sno}(\sigma_{cno=001}(SC)) \cup \Pi_{sno}(\sigma_{cno=002}(SC))$$

差运算

所有出现在一个关系而不在另一关系中的元组集合

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



R 和 S 必须是相容的

差运算计算例子

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

差运算查询示例

求选修了001号但未选修002号课程的学生号

方案1:

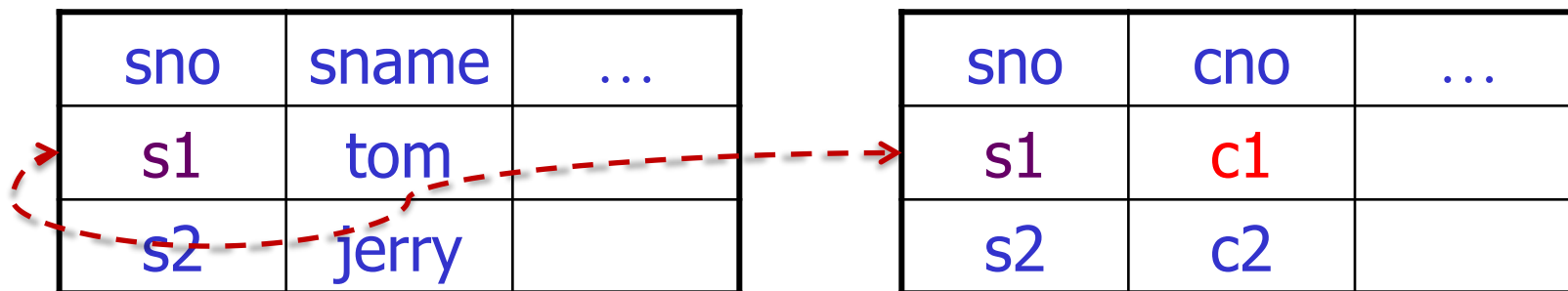
$$\Pi_{sno}(\sigma_{cno=001}(SC)) - \Pi_{sno}(\sigma_{cno=002}(SC))$$

方案2:

$$\Pi_{sno}(\sigma_{cno=001 \wedge cno \neq 002}(SC))$$

笛卡尔积运算：如何跨表（多行比较）查询？

求选修c1课程的学生姓名



行级思维： 逐一扫描SC中的行，过滤出cno=c1的行，获得其sno，再根据得到的sno到S中查找对应的行

什么是集合级思维？运算的对象是集合

如何把多表数据融合到一个表中，以便施加选择和投影运算？

笛卡尔积运算定义

元组的连串 (Concatenation)

$r = (r_1, \dots, r_n)$, $s = (s_1, \dots, s_m)$, r 与 s 的连串定义为:

$$\widehat{rs} = (r_1, \dots, r_n, s_1, \dots, s_m)$$

关系的笛卡尔积:

$$R \times S = \{ \widehat{rs} \mid r \in R \wedge s \in S \}$$

- $R \times S$ 的度为 R 与 S 的度之和
- $R \times S$ 的元组个数为 R 和 S 的元组个数的乘积

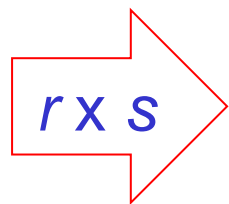
笛卡尔积运算示例

r

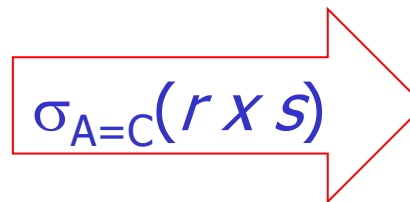
A	B
α	1
β	2

s

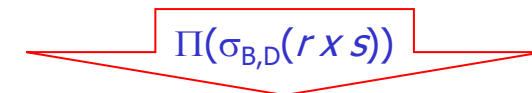
C	D	E
α	10	a
β	10	a
β	20	b
γ	10	b



A	B	C	D	E
α	1	α	10	a
α	1	β	10	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



A	B	C	D	E
α	1	α	10	a
β	2	β	20	a
β	2	β	20	b



B	D
1	10
2	20

包打天下：选择、投影、笛卡尔积

笛卡尔积运算查询示例

sno	sname	...
s1	tom	
s2	jerry	

sno	cno	...
s1	c1	
s2	c2	

求选修c1课程的学生姓名

$S \times SC$

S.sno	sname	...	SC.sno	cno	...
s1	tom		s1	c1	
s1	tom		s2	c2	
s2	jerry		s1	c1	
s2	jerry		s2	c2	

$\sigma_{S.sno=SC.sno \wedge cno=c1}$

sname
tom

Π_{sname}

S.sno	sname	...	SC.sno	cno	...
s1	tom		s1	c1	

$\Pi_{sname}(\sigma_{S.sno=SC.sno \wedge cno=c1}(S \times SC))$

关系自身的笛卡尔积运算

求数学成绩比王红同学高的学生

R

$R \times \rho_S(R)$



姓名	课程	成绩
张军	物理	93
王红	数学	86
张军	数学	89

R.姓名	R.课程	R.成绩	S.姓名	S.课程	S.成绩
王红	数学	86	张军	物理	93
王红	数学	86	王红	数学	86
王红	数学	86	张军	数学	89

$\Pi_{S.姓名} \left(\sigma_{R.姓名=王红 \wedge R.课程=数学 \wedge S.课程=数学 \wedge R.成绩 < S.成绩} (R \times \rho_S(R)) \right)$

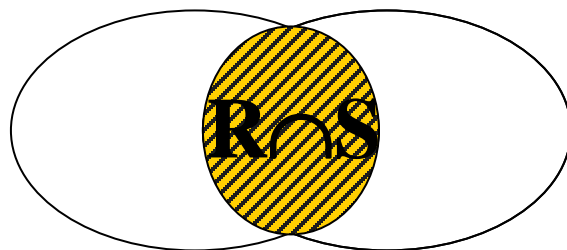




交运算

所有同时出现在两个关系中的元组集合

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



交运算可以通过差运算来重写

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

交运算计算例子

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

交运算查询示例

求同时选修了001号和002号课程的学生号

方案1:

$$\Pi_{sno}(\sigma_{cno=001}(SC)) \cap \Pi_{sno}(\sigma_{cno=002}(SC))$$

方案2:

$$\Pi_{sno}(\sigma_{cno=001 \wedge cno=002}(SC))$$

θ 连接

从两个关系的广义笛卡儿积中选取给定属性间满足一定条件的元组

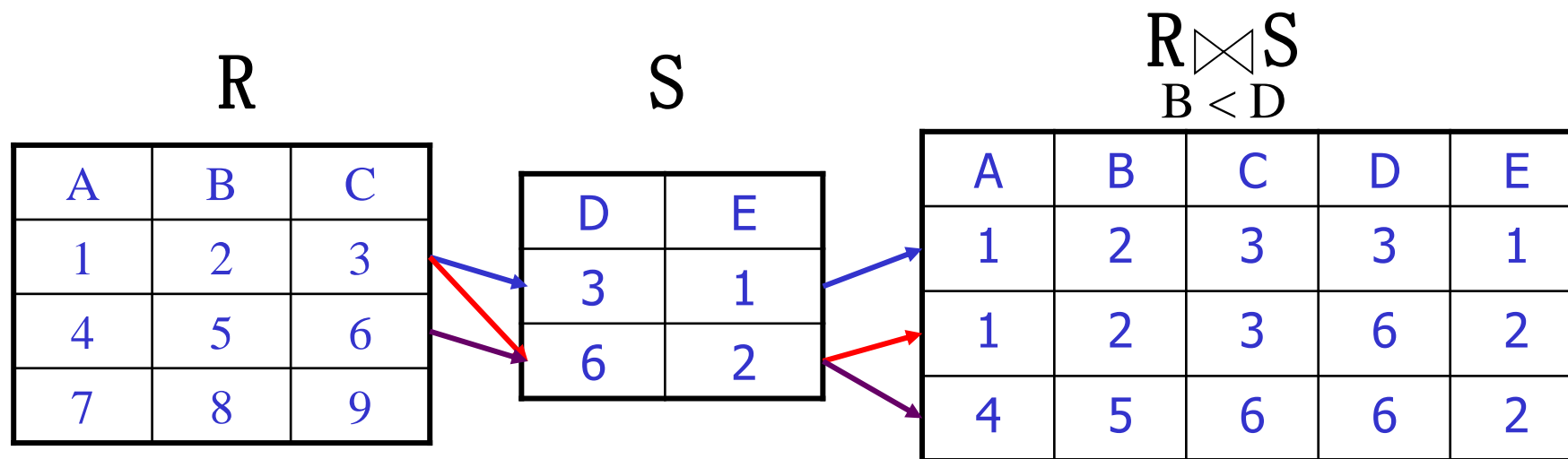
$$R \bowtie_{A\theta B} S = \{ \widehat{rs} \mid r \in R \wedge s \in S \wedge r[A] \theta s[B] \}$$

$$R \bowtie_{A\theta B} S = \sigma_{r[A] \theta s[B]}(R \times S)$$

A, B为R和S上度数相等且可比的属性列

θ 为算术比较符，为等号时称为等值连接

θ连接



求数学成绩比王红同学高的学生

$$\Pi_{S.姓名} \left(\left(\sigma_{\text{课程=数学} \wedge \text{姓名=王红}}(R) \right) \bowtie_{R.成绩 < S.成绩} \left(\sigma_{\text{课程=数学}}(S) \right) \right)$$

自然连接

从两个关系的广义笛卡儿积中选取在相同属性列B上取值相等的元组，并去掉重复的列

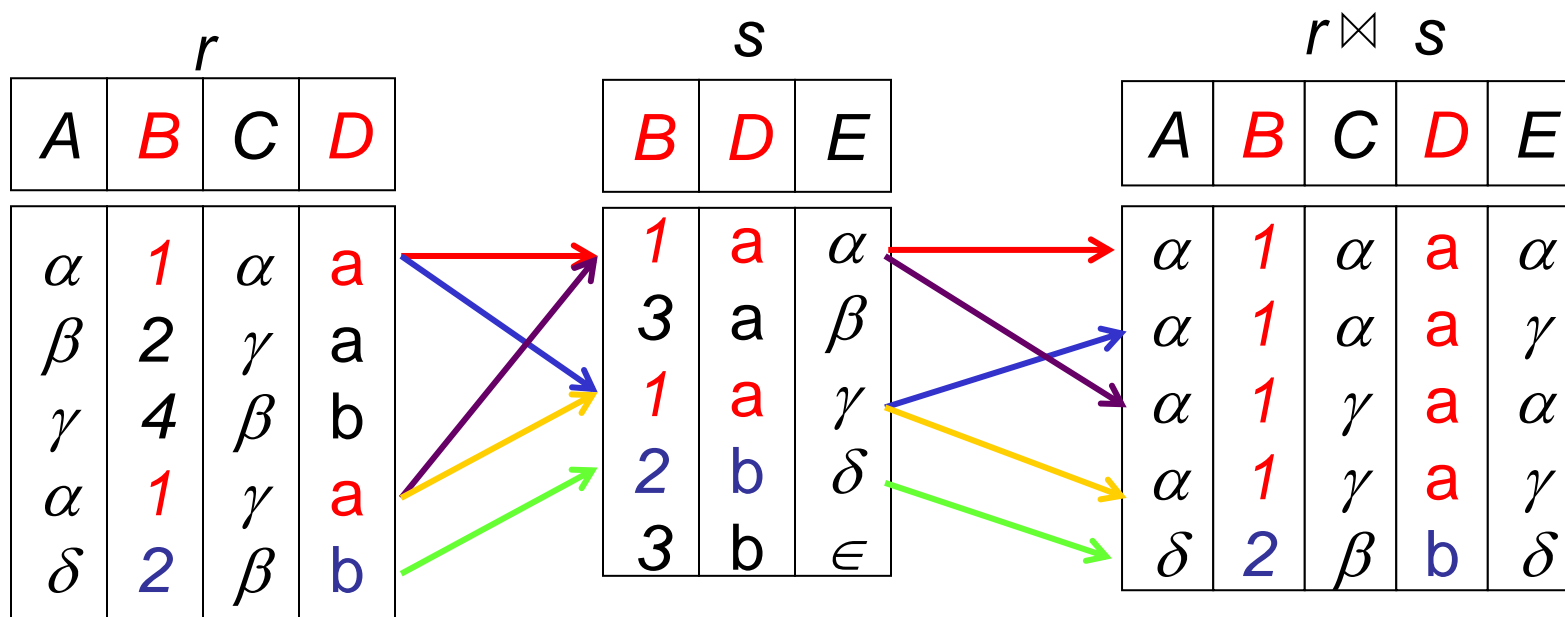
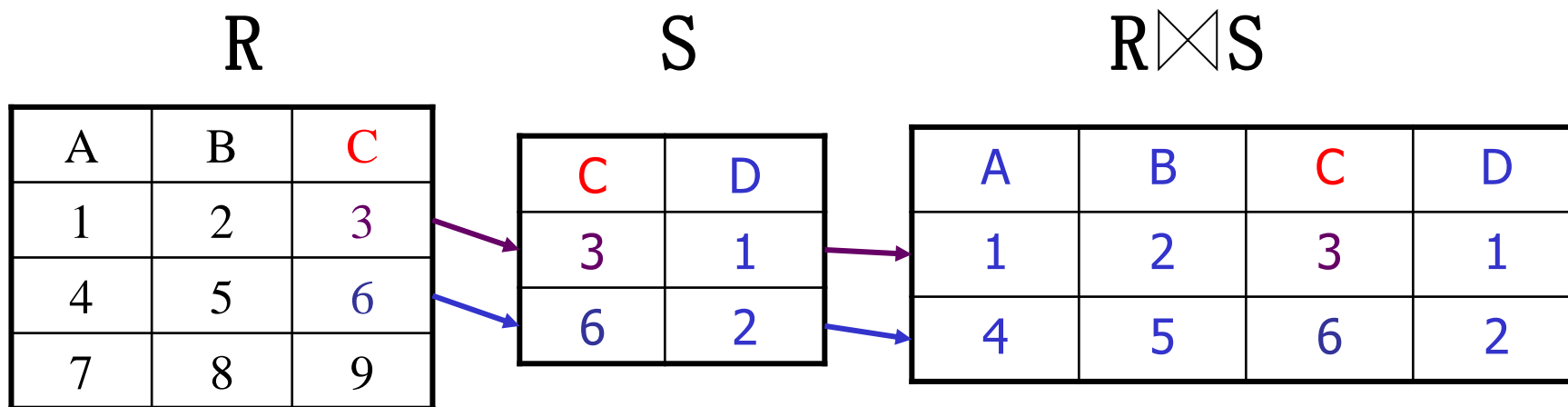
$$R \bowtie S = \{ \widehat{rs}[\bar{B}] \mid r \in R \wedge s \in S \wedge r[B] = S[B] \}$$

自然连接与等值连接的不同

自然连接中相等的分量必须是相同的属性组，并且要在结果中去掉重复的属性，而等值连接则不必

自然连接为什么非常有用？

自然连接的计算例子



自然连接表达查询的例子

求001号学生所在系的名称

$$\Pi_{dname}(\sigma_{sno=001}(S \bowtie dept))$$

$$\Pi_{dname}(\sigma_{sno=001}(S) \bowtie dept)$$

关系R(A, B), S(A, C), R与S中元组个数分别为10, 15, 试填写下表

条件	表达式	最小元组数	最大元组数
无任何条件	$R \bowtie S$	0	150
	$\Pi_A(R) \cup \Pi_A(S)$	1	25
A是R的主码	$R \bowtie S$	0	15
	$\Pi_A(R) \cup \Pi_A(S)$	10	25
A是R的主码	$R \bowtie S$	15	15
A是S的外码	$\Pi_A(R) \cup \Pi_A(S)$	10	10

自然连接的问题

列出**所有**老师的有关信息，包括姓名、工资、所教授的课程

$\Pi_{pno,pname,sal,cno,cname}(PROF \bowtie PC \bowtie C)$

pno	pname	sal
P01	赵明	800
P02	钱广	700
P03	孙立	600
P04	李三	500



cno	pno
C01	P01
C02	P02
C02	P04



cno	cname
C01	物理
C02	数学
C03	化学

==

pno	pname	sal	cno	cname
P01	赵明	800	C01	物理
P02	钱广	700	C02	数学
P04	李三	500	C02	数学

问题：有关P03号职工的姓名和工资信息没有显示出来

如何由关系表生成报表？

教工号	教工名	年龄	工资	论文	教学	项目	服务
S1	√	√	√	√			√
S2	√	√	√	√		√	
S3	√	√	√		√	√	
S4	√	√	√	√		√	
S5	√	√	√		√	√	√

涉及到多个表之间的连接
全能榜？

外连接

为避免自然连接时因失配而发生的信息丢失，可以假定往参与连接的一方表中附加一个取值全为空值的行，它和参与连接的另一方表中的任何一个未匹配上的元组都能匹配，称之为外连接

外连接 = 自然连接 + 未匹配元组(悬挂元组)

外连接的形式：左外连接、右外连接、全外连接

❏ 左外连接 = 自然连接 + 左侧表中未匹配元组

❏ 右外连接 = 自然连接 + 右侧表中未匹配元组

❏ 全外连接 = 自然连接 + 两侧表中未匹配元组

外连接的执行顺序

输出所有老师的信息

pno	pname	sal
P01	赵明	800
P02	钱广	700
P03	孙立	600
P04	李三	500



cno	pno
C01	P01
C02	P02
C02	P04



cno	cname
C01	物理
C02	数学
C03	化学

==

pno	pname	sal	cno	cname
P01	赵明	800	C01	物理
P02	钱广	700	C02	数学
P04	李三	500	C02	数学
P03	孙立	600	null	null

外连接的结合律成立否？

外连接结合律不成立的反例

R1

A	B
1	2

R2

B	C
2	3

R3

A	C
4	5

$(R1 \bowtie R2) \bowtie R3$

A	B	C
1	2	3
4	null	5

$R1 \bowtie (R2 \bowtie R3)$

A	B	C
1	2	null
null	2	3
4	null	5

半连接

R

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

S

B	C	D
b	c	d
b	c	e
e	b	a
a	d	b

猜想一下什么是
反半连接？

$R \bowtie S$

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

$S \bowtie R$

B	C	D
b	c	d
b	c	e
a	d	b

R anti join S

A	B	C
b	b	f

半连接的应用场景：减少分布式查询中的通讯量

场地1

emp(eno, ename, dno)
10,000行

场地2

dept(dno, dname, mgr)
10行

在场地2查询每个
部门的经理姓名

$\Pi_{ename}(emp \bowtie_{emp.eno=dept.mgr} dept)$

方案1

场地1

emp(10,000行)

场地2

场地1

1. dept(10行)

方案2

2. emp \bowtie dept(10行)

场地2

外部并

$R(AB)$ outer union $S(BC)$

在B上相同的行合并为一行

- 一般用于**信息集成**，将子实体集合并为父实体集
- **实体识别**是判别两个子实体是否对应同一个实体

R

A	B
a	b
c	d

S

B	C
b	c
a	d

R outer union S

A	B	C
a	b	c
c	d	null
null	a	d

A	B	C
a1	b1	c1
a2	b2	c2

B	C	D
b1	c3	d1

R outer union S

A	B	R.C	S.C	D
a1	b1	c1	c3	d1
a2	b2	c2	null	null

象集 (Image Set)

关系 $R(X, Z)$, X, Z 是属性组, x 是 X 上的取值

定义 x 在 R 中的象集为:

$$Z_x = \{ t[Z] \mid t \in R \wedge t[X] = x \}$$

象集:

选择 + 投影

从 R 中选出在 X 上取值为 x 的元组, 只留 Z 属性

X	Z
姓名	课程
张军	物理
王红	数学
张军	数学

$x = \text{张军}, Z_x$

课程
数学
物理

象集的含义:

张军同学所选
修的全部课程

问题：如何求得选修了全部课程的学生？

思路一：判断每个学生的课程象集是否包含了整个课程集合

SC

姓名	课程名
张军	物理
王红	数学
张军	数学

C

课程名
数学
物理

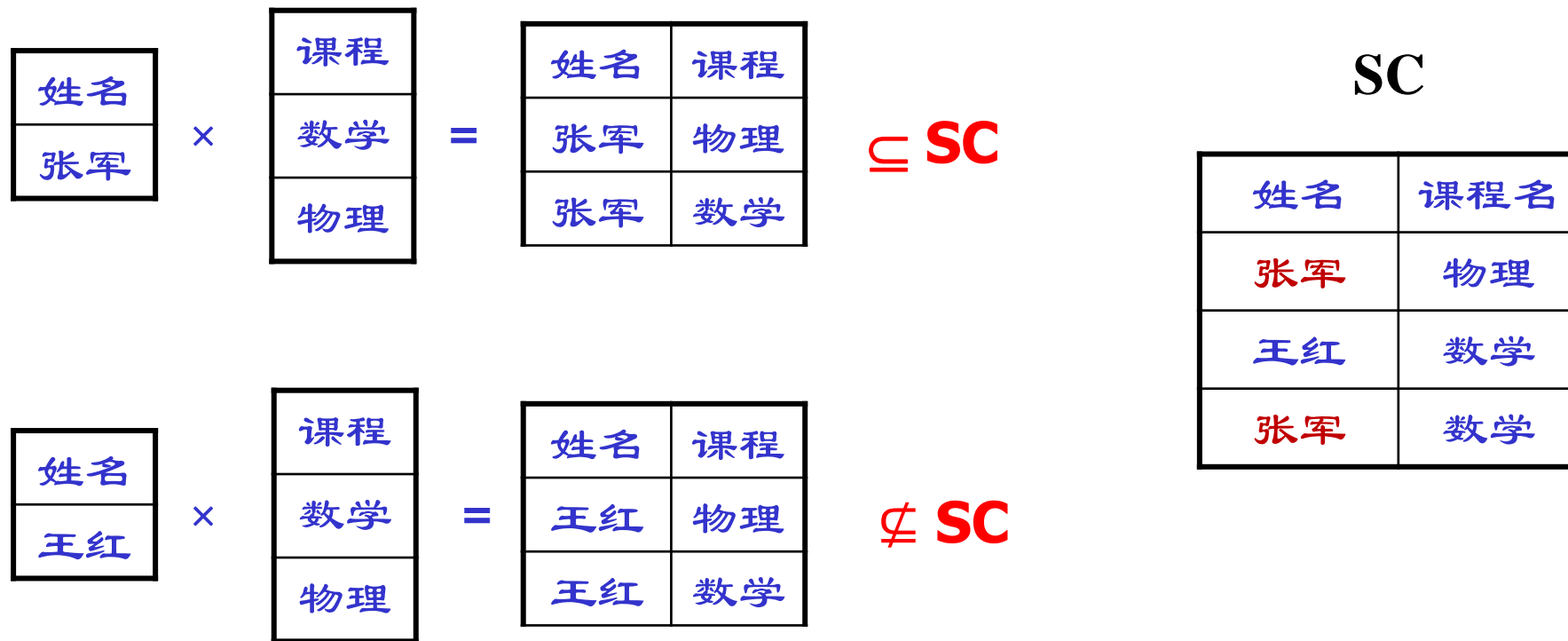
课程_{张军} = { 物理, 数学 } \supseteq C

课程_{王红} = { 数学 } $\not\supseteq$ C

$$\{ u \mid r \in SC \wedge u = r[\text{姓名}] \wedge \text{课程名}_u \supseteq C \}$$

如何求得选修了全部课程的学生？

思路二：判断学生与课程集合构成的笛卡尔积是否完全包含在选课集合中



$$\{u | u \in \Pi_{\text{姓名}}(SC) \wedge \forall v \in C(\widetilde{uv} \in SC)\}$$

除法的本能：所有、全部、任意

除法的定义表达式

$$R(X, Y) \div S(Y) = \{x | r \in R \wedge x = r[X] \wedge Y_x \supseteq S\}$$

$$R(X, Y) \div S(Y) = \{u | u \in \Pi_X(R) \wedge \forall v \in S(\tilde{u}v \in R)\}$$

除法的计算表达式

$$R(X, Y) \div S(Y) = \Pi_X(R) - \Pi_X(\Pi_X(R) \times \Pi_Y(S) - R)$$

如何求得没有选修全部课程的学生？

所有学生

姓名
张军
王红

所有课程

课程
数学
物理

×

=

所有学生选修全部课程

姓名	课程
张军	物理
王红	数学
张军	数学
王红	物理

没有选修全部课程的学生

没有发生的选课记录

姓名	课程
王红	物理

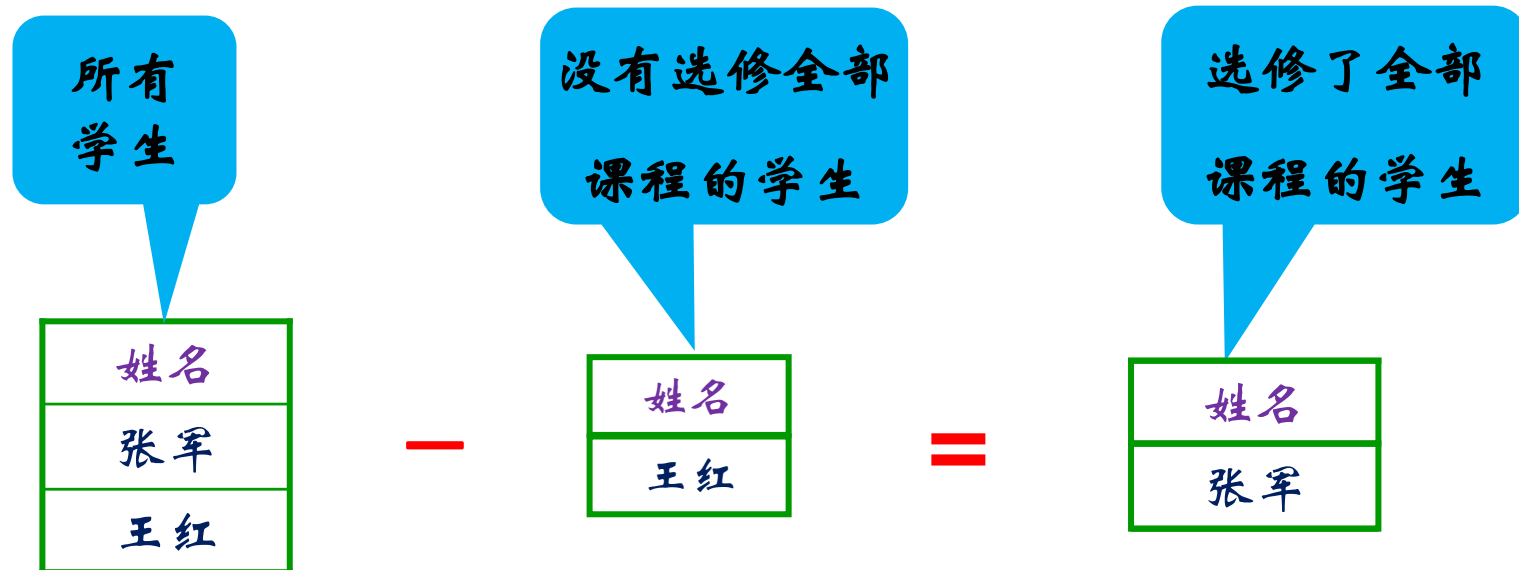
=

-

实际的学生选修课程

姓名	课程
张军	物理
王红	数学
张军	数学

如何求得选修全部课程的学生？



体会减法操作在区分两类不同性质子集时的作用

除运算示例

$SC(sno, cno, grade)$

求选修了所有课程的学生

方案1: $\Pi_{sno, cno}(SC) \div \Pi_{cno}(C)$

方案2: $\Pi_{sno}(SC \div \Pi_{cno}(C))$



对除运算中的被除关系要正确投影

姓名	课程
张军	物理
王红	数学
张军	数学
王红	物理

÷

课程
数学
物理

=

姓名
张军
王红

选修了全部课程的学生

姓名	课程	成绩
张军	物理	93
王红	数学	86
张军	数学	93
王红	物理	92

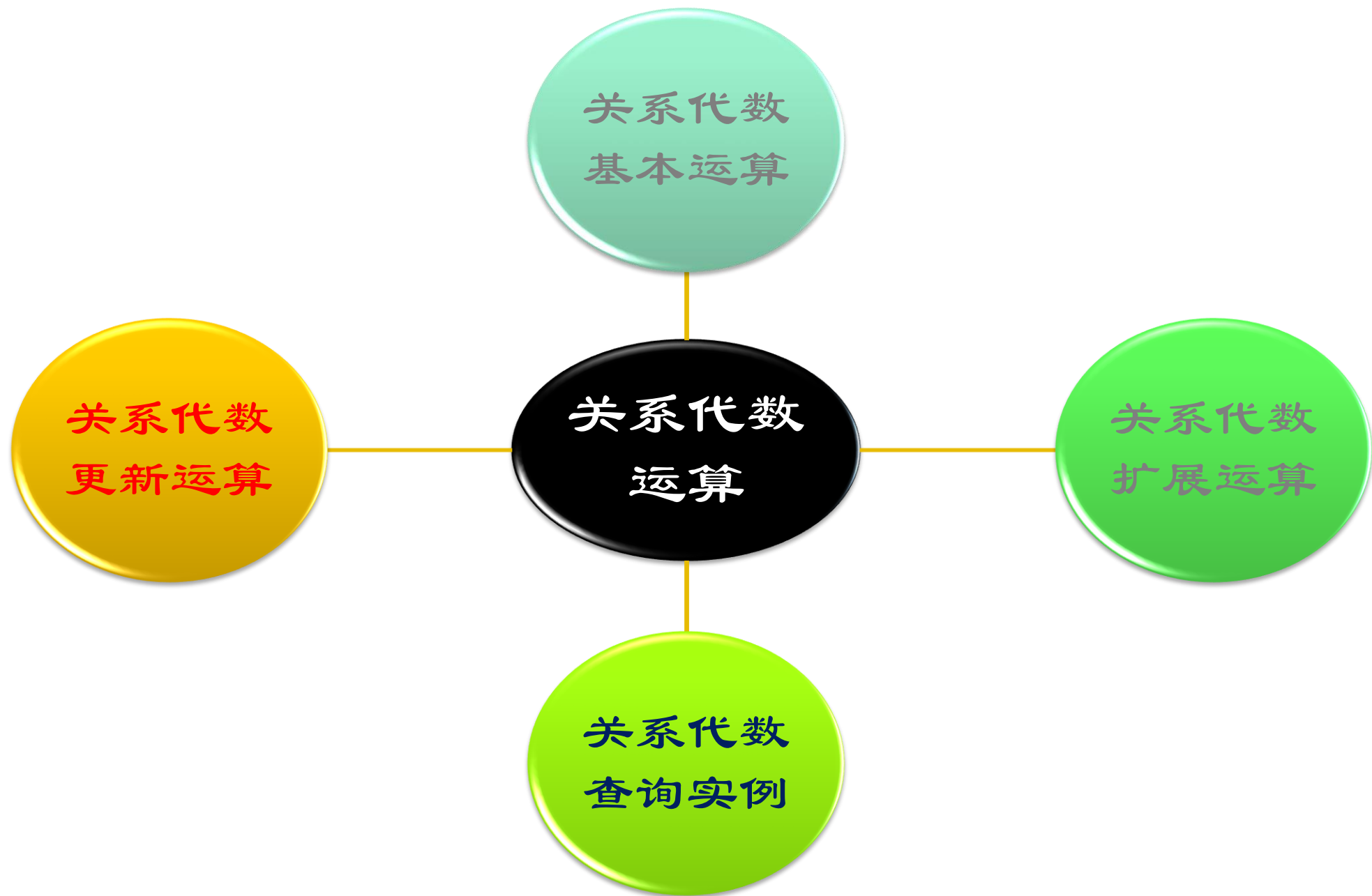
÷

课程
数学
物理

=

姓名	成绩
张军	93

选修了全部课程并且成绩都相同的学生



赋值运算

- 为使查询表达简单、清晰，可以将一个复杂关系代数表达式分成几个部分
- 每一部分都赋予一个临时关系变量
- 该变量可被看作关系而在后续表达式中使用

临时关系变量 \leftarrow 关系代数表达式

赋值运算

$$R \div S = \Pi_X(R) - \Pi_X(\Pi_X(R) \times \Pi_Y(S) - R)$$



$$temp1 \leftarrow \Pi_X(R)$$

$$temp2 \leftarrow \Pi_X(temp1 \times \Pi_Y(S) - R)$$

$$result \leftarrow temp1 - temp2$$

广义投影

在投影列表中使用算术表达式来对投影进行扩展

$$\prod_{F_1, F_2, \dots, F_n}(E)$$

F_1, F_2, \dots, F_n 是算术表达式

求教工应缴纳的所得税

$$\prod_{pno, sal * 0.05}(PROF)$$

$$\rho_{pno, income-tax}(\prod_{pno, sal * 0.05}(PROF))$$

数据库修改：删除

将满足条件的元组从关系中删除

$$R \leftarrow R - E$$

是对永久关系的赋值运算

数据库修改：删除

删除001号老师所担任的课程

$$PC \leftarrow PC - \sigma_{pno=001}(PC)$$

删除没有选课的学生

$$S \leftarrow S - (\Pi_{sno}(S) - \Pi_{sno}(SC)) \bowtie S$$

数据库修改：插入

插入一个指定的元组，或者插入一个查询结果

$$R \leftarrow R \cup E$$

新加入一个老师

$$PC \leftarrow PC \cup \{ (p07, \text{周正}, 3000, d08) \}$$

数据库修改：插入

加入计算机系学生选修“数据结构”的信息

$$SC \leftarrow SC \cup (\Pi_{sno} (S \bowtie \sigma_{dname='计算机系', (DEPT)}) \\ \times \Pi_{cno} (\sigma_{name='数据结构', (C)}) \\ \times \{null\})$$

数据库修改：更新

利用广义投影改变元组的某些属性上的值

$$R \leftarrow \prod_{F_1, F_2, \dots, F_n}(R)$$

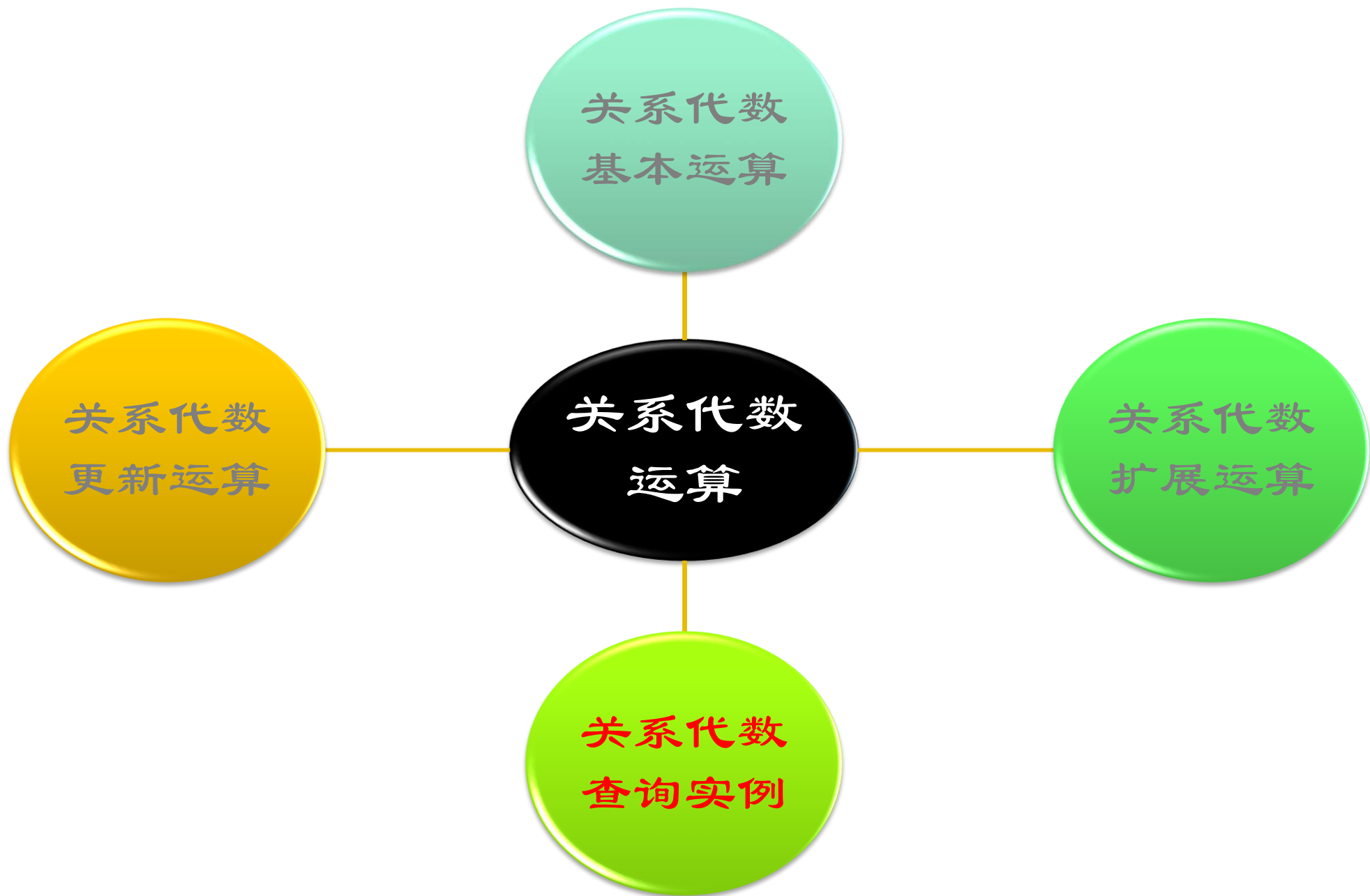
给每位老师上调10%的工资

$$PROF \leftarrow \prod_{pno, pname, sal \leftarrow sal * 1.1, dno}(PROF)$$

数据库修改：更新

对工资超过3000的老师征收5%所得税

$$PROF \leftarrow \Pi_{pno, pname, sal \leftarrow sal * 0.95, dno} (\sigma_{sal > 3000}(PROF)) \\ \cup (\sigma_{sal \leq 3000}(PROF))$$



思为双飞燕 衔泥巢君屋

求没有选修c1号课程的学生



所有学生 - 选修了c1号课程学生

$$\Pi_{sno}(SC) - \Pi_{sno}(\sigma_{cno=c1}(SC))$$

明月皎夜光 促织鸣东壁

求没有选修c1号课程的学生

写成 $\Pi_{sno}(\sigma_{cno \neq c1}(SC))$ 对不对?

仅选c1号课程
之外的其他学生

sno	cno
s1	c1
s1	c2
s2	c1
s3	c2

$\sigma_{cno \neq c1}(SC)$

sno	cno
s1	c2
s3	c2

Π_{sno}

sno
s1
s3

胡马依北风 越鸟巢南枝

求仅选修了c1号课程的学生号

选修c1号课程的学生 - 仅选c1号课程之外的学生

$$\Pi_{sno}(\sigma_{cno=c1}(SC)) - \Pi_{sno}(\sigma_{cno \neq c1}(SC))$$

青青陵上柏，磊磊涧中石

如何甄别开只选C1的同学和

选修C1同时又选修其他课程的同学？

sno	cno
s1	c1
s1	c2
s2	c1
s3	c2

—

sno	cno
s1	c1
s2	c1

=

sno	cno
s1	c2
s3	c2

$$\Pi_{sno}(\sigma_{cno=c1}(SC)) - \Pi_{sno}(SC - \sigma_{cno=c1}(SC))$$

青青河畔草 郁郁园中柳

求选修c1课程比s1学生的该门课程成绩高的学生

sno	cno	G
s1	c1	g1
s2	c1	g2

SC × SC


(s1, c1, g1, s2, c1, g2)

且 $g1 < g2$

$\Pi_{S.sno}(\sigma_{R.sno=s1 \wedge R.cno=c1 \wedge R.G < S.G}(\rho_R(SC) \times \rho_S(SC)))$

西北有高楼 上与浮云齐

求每门课程的先修课的先修课



cno	pcno
c1	null
c2	c1
c3	c2

$C \times C$

$(c3, c2, c2, c1)$

$$\Pi_{C.cno, R.pcno}(\sigma_{C.pcno=R.cno}(C \times \rho_R(C)))$$

递归查询:

找到所有的祖辈先修课?

关系代数不具有生成传递闭包

的能力, 不满足关系完备性

涉江采芙蓉 兰泽多芳草

求选修了至少两门课的学生

sno	cno	G
s1	c1	g1
s1	c2	g2

$SC \times SC$

$(s1, c1, g1, s1, c2, g2), c1 \neq c2$

$$\Pi_{R.sno}(\sigma_{R.sno=S.sno \wedge R.cno \neq S.cno}(\rho_R(SC) \times \rho_S(SC)))$$

求选修了至少N门课的学生

求只选修了1门课的学生

迢迢牵牛星 皎皎河汉女

求最低的成绩

sno	cno	G
s1	c1	60
s1	c2	70
s2	c1	80
s3	c2	90

$$\sigma_{R.G < S.G}(\rho_R(SC) \times \rho_S(SC))$$

R.sno	R.cno	R.G	S.sno	S.cno	S.G
s1	c1	60	s1	c2	70
s1	c1	60	s2	c1	80
s1	c1	60	s3	c2	90
s1	c2	70	s2	c1	80
s1	c2	70	s3	c2	90
s2	c1	80	s3	c2	90

$$\Pi_G(SC) - \Pi_{S.G}(\sigma_{R.G < S.G}(\rho_R(SC) \times \rho_S(SC)))$$

生年不满百 常怀千岁忧

求选修课程中包含了 所有 S01号学生所选修课程的学生号

$$\Pi_{sno,cno}(SC)$$
$$\div$$
$$\Pi_{cno}(\sigma_{sno=s01}(SC))$$


冉冉孤生竹 结根泰山阿

stock(sno, date, price)

找出一直上涨的股票

持续上涨10天的股票😊, 持续上涨的股票😞

上涨过的股票: if $d1 > d2$ then $p1 > p2$

$$\prod_{R.sno} (\sigma_{R.sno=S.sno \wedge R.date>S.date \wedge R.price>S.price} (\rho_R(stock) \times \rho_S(stock)))$$

持续上涨的股票 = 所有股票 - 下跌过的股票

庭中有奇树 绿叶发华滋

guanxi(source, destination)

source	destination
p1	p2
p1	p3
p2	p1

- 互相认识的人，互不认识的人…
- 认识所有人的人，被所有人认识的人…
- 认识的传递性：间接认识
- 社会网络分析：度分布、聚类系数、介度、紧度

舍弟江南歿 家兄塞北亡

source	destination
p1	p2
p2	p1

目标：
消除对称关系

source	destination
p1	p2

$\rho_R(guanxi) \times \rho_S(guanxi)$

任意列都包含p1和p2

R.s	R.d	S.s	S.d
p1	p2	p1	p2
p1	p2	p2	p1
p2	p1	p1	p2
p2	p1	p2	p1

如何筛选？

$\sigma_{R.d=S.s \wedge R.s<S.s}$

$guanxi -$

S.s	S.d
p2	p1

$\Pi_{S.s,S.d}$

R.s	R.d	S.s	S.d
p1	p2	p2	p1

$guanxi - \Pi_{S.s,S.d}(\sigma_{R.d=S.s \wedge R.s=S.d \wedge R.s<S.s}(\rho_R(guanxi) \times \rho_S(guanxi)))$

