



2026年春季学期

数据库系统概论

An Introduction to Database Systems

第二章 关系数据库

中国科学技术大学
人工智能与数据科学学院

黄振亚, huangzhy@ustc.edu.cn



2.1 关系数据结构

26

2.1.1 关系

2.1.2 关系模式

2.1.3 关系数据库

2.1.4 关系模型的存储结构



2.1.2 关系模式

27

1. 什么是关系模式
2. 定义关系模式
3. 关系模式与关系



1. 什么是关系模式

28

关系数据库中

- 关系模式（Relation Schema）是**型**
- 关系是**值**
- 关系模式是对关系的描述
 - 元组集合的结构
 - 属性构成
 - 属性来自的域
 - 属性与域之间的映象关系
 - 完整性约束条件
 - 属性间数据的依赖关系集合
 - 如，取值范围，相互关联



2. 定义关系模式

29

定义2.4 关系模式可以形式化地表示为：

$R (U, D, DOM, F)$

R 关系名

U 组成该关系的属性名集合

D 属性组 U 中属性所来自的域

DOM 属性向域的映象集合

F 属性间的数据依赖关系集合

第6章 关系数据理论



定义关系模式 (

SUPERVISOR	SPECIALITY	POSTGRADUATE
张清玫	计算机专业	李勇
张清玫	计算机专业	刘晨
刘逸	信息专业	王敏

例: SAP模式:

SAP(SUPERVISOR, SPECIALITY, POSTGRADUATE)

□ 定义域

□ D1: PERSON。 D2: SPECIALITY

□ 定义DOM

□ 取不同的属性名，并在模式中定义属性向域的映象，即说明它们分别出自哪个域:

DOM (SUPERVISOR)

= DOM (POSTGRADUATE)

= PERSON



定义关系模式 (续)

关系模式通常可以简记为

$R(U)$ 或 $R(A_1, A_2, \dots, A_n)$

- R : 关系名
- A_1, A_2, \dots, A_n : 属性名

注: 域名及属性向域的映象常常直接说明为属性的类型、长度

```
CREATE TABLE Student 关系R  
  (Sno CHAR(9) PRIMARY KEY,  
   Sname CHAR(20) UNIQUE,  
   Ssex CHAR(2)  
   Sage INT)  
  
  属性名      DOM
```



3. 关系模式与关系

32

- 关系模式
 - 对关系的描述
 - 静态的、稳定的
- 关系
 - 关系模式在**某一时刻**的状态或内容
 - 动态的、随时间不断变化的
- 注：关系模式和关系往往统称为关系

通过上下文加以区别



2.1 关系数据结构

33

2.1.1 关系

2.1.2 关系模式

2.1.3 关系数据库

2.1.4 关系模型的存储结构



2.1.3 关系数据库

34

- 关系数据库
 - 在一个给定的应用领域中，所有关系的集合构成一个关系数据库
- 关系数据库的型与值



2.1.3 关系数据库

35

- 关系数据库的型: 关系数据库模式
对关系数据库的描述。
- 关系数据库模式包括
 - 若干域的定义
 - 在这些域上定义的若干关系模式
- 关系数据库的值: 关系模式在某一时刻对应的关系的集合，简称为关系数据库



2.1.3 关系数据库（示例）

- 学生表: Student (Sno, Sname, Ssex, Sage, Sdept)
- 课程表: Course (Cno, Cname, Cpno, Ccredit)
- 学生选课表: SC (Sno, Cno, Grade, Semester)

Sno	Sname	Ssex	Sage	Sdept
0001	李勇	男	20	CS
0002	刘晨	女	19	CS

Cno	Cname	Cpno	Ccredit
1	数据库	3	3.5
2	数学		4
3	数据结构		4

Sno	Cno	Grade	Semester	Sno	Cno	Grade	Semester
0001	1	92	20192	0002	2	90	20192
0001	2	85	20201	0002	3	80	20201
0001	3	88	20202				



2.1 关系数据结构

37

2.1.1 关系

2.1.2 关系模式

2.1.3 关系数据库

2.1.4 关系模型的存储结构



2.1.4 关系数据库的存储结构

38

- 关系数据库的逻辑结构表示：“表”
- 关系数据库的**物理结构**
 - 一个表对应一个操作系统文件，物理结构组织交给操作系统完成；
 - 从操作系统那里申请若干个大的文件，自己划分文件空间，组织表、索引等结构，并进行存储管理



第二章 关系数据库

39

2.1 关系模型概述

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



2.2.1 基本关系操作

40

- 常用的关系操作
 - 查询：选择、投影、连接、除、并、交、差
 - 数据更新：插入、删除、修改
- 查询的表达能力是其中最主要的部分
 - 5种基本操作：选择、投影、并、差、笛卡尔积
-



2.2.2 关系数据库语言的分类

41

- 关系代数语言 $\pi_{Sno} (\sigma_{Cno='2'} (SC))$
 - 用对关系的运算来表达查询要求。代表：ISBL
- 关系演算语言：用谓词来表达查询要求
 - 元组关系演算语言
 - 谓词变元的基本对象是元组变量。代表：APLHA, QUEL
 - 域关系演算语言
 - 谓词变元的基本对象是域变量。代表：QBE
- 具有关系代数和关系演算双重特点的语言
 - 代表：SQL (Structured Query Language)
- 关系数据语言 (SQL)
 - 查询、数据定义语言、数据操纵语言、数据控制语言



2.2.2 关系数据库语言

42

- 关系数据库语言的特点
 - 集合操作方式
 - 操作的对象和结果都是集合，一次一集合的方式
 - 语言具有完备的表达能力
 - 非过程化的操作语言，是描述性语言
 - “要什么”，而非“怎么要”
 - 不必请求建立特殊存取路径，存取路径的选择由DBMS完成
 - 例如，顺序查询，or 索引查询
 - 系统自动查询优化



第二章 关系数据库

43

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



2.3 关系的完整性

44

2.3.1 关系的三类完整性约束

2.3.2 实体完整性

2.3.3 参照完整性

2.3.4 用户定义的完整性



2.3.1 关系的三类完整性约束

45

- 现实世界的要求
- 实体完整性和参照完整性
 - 关系模型**必须满足**的完整性约束条件
 - 称为关系的两个**不变性**，应该由关系系统自动支持
- 用户定义的完整性
 - 应用领域需要遵循的约束条件，体现了具体领域中的语义约束（只需定义）
- **任何关系在任何时刻**都满足这些语义约束



2.3 关系的完整性

46

2.3.1 关系的三类完整性约束

2.3.2 实体完整性

2.3.3 参照完整性

2.3.4 用户定义的完整性



2.3.2 实体完整性

47

规则2.1 实体完整性规则 (Entity Integrity)

- 若属性A 是基本关系R 的主属性，则属性A不能取空值NULL
 - 例：SAP (SUPERVISOR, SPECIALITY, POSTGRADUATE)
POSTGRADUATE: 主码 (假设研究生不会重名)
不能取空值
 - 例：选修 (Sno, Cno, Grade, Semester)
Sno和Cno: 两个属性均不能取空值
- 空值就是“不知道”或“不存在”或“无意义”的值。非0
 - 例如，Grade与空值的关系



实体完整性

48

实体完整性规则的说明

- (1) 实体完整性规则是针对**基本关系**而言的。一个基本表通常对应现实世界的一个实体集。
- (2) 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。
- (3) 关系模型中以主码作为**唯一性标识**。
- (4) **主码中的属性，即主属性**不能取空值。

主属性取空值，就说明存在某个不可标识的实体，即存在不可区分的实体，这与第（2）点相矛盾，

这个规则称为**实体完整性**



2.3关系的完整性

49

2.3.1关系的三类完整性约束

2.3.2 实体完整性

2.3.3 参照完整性

2.3.4 用户定义的完整性



2.3.3 参照完整性

50

1. 关系间的引用（参照）
2. 外码（Foreign Key）
3. 参照完整性规则



1. 关系间的引用

- 在关系模型中，实体及实体间的联系都是用关系来描述的，因此可能存在着**关系与关系间的引用**。

例2.1 学生实体、专业实体

学生 (学号, 姓名, 性别, **专业号**, 年龄)

主码

专业 (专业号, 专业名)

主码

- ✓ 学生关系引用了专业关系的主码“专业号”。
- ✓ 学生关系中的“专业号”值必须是确实存在的专业的专业号，即专业关系中有该专业的记录



关系间的引用(续)

52

例2.2 学生、课程、学生与课程之间的多对多联系

学生（学号，姓名，性别，年龄，所在系）

课程（课程号，课程名，先修课，学分）

学生选修（学号，课程号，成绩，开课学期）



关系间的引用(续)

53

Sno	Sname	Ssex	Sage	Sdept
0001	李勇	男	20	CS
0002	刘晨	女	19	CS

Cno	Cname	Cpno	Ccredit
1	数据库	3	3.5
2	数学		4
3	数据结构		4

Sno	Cno	Grade	Semester
0001	1	92	20192
0001	2	85	20201
0001	3	88	20202
0002	2	90	20192
0002	3	80	20201



关系间的引用(续)

例2.3 学生实体及其内部的一对多联系

学生 (学号, 姓名, 性别, 专业号, 年龄, 班长)

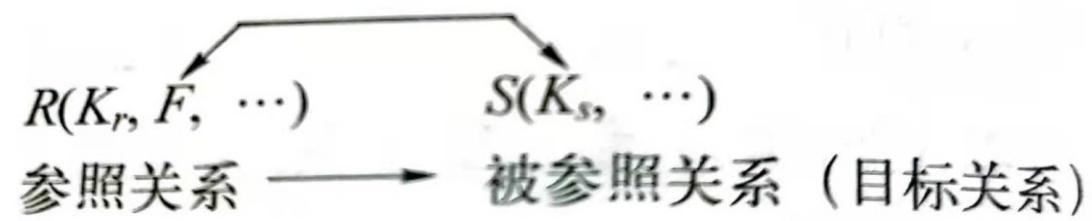
学号	姓名	性别	专业号	年龄	班长
801	张三	女	01	19	802
802	李四	男	01	20	
803	王五	男	01	20	802
804	赵六	女	02	20	805
805	钱七	男	02	19	

- ✓ “学号” 是主码，“班长” 是外码，它引用了本关系的“学号”
- ✓ “班长” 必须是确实存在的学生的学号



2. 外码 (Foreign Key)

- 定义2.5 设 F 是基本关系 R 的一个或一组属性，但不是关系 R 的码。如果 F 与基本关系 S 的主码 K_s 相对应，则称 F 是基本关系 R 的外码 (Foreign Key)
- 基本关系 R 称为参照关系 (Referencing Relation)
- 基本关系 S 称为被参照关系 (Referenced Relation) 或目标关系 (Target Relation)

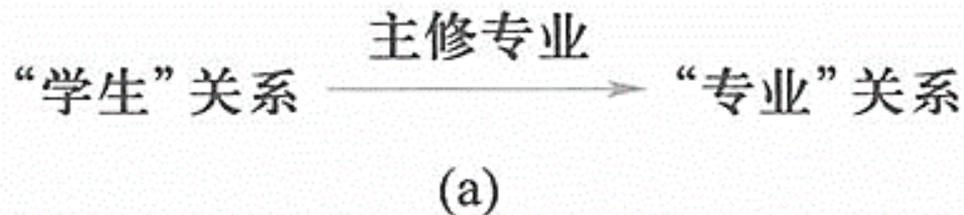




外码(续)

56

- [例2.1] : 学生 (学号, 姓名, 性别, 主修专业, 年龄)
 专业 (专业名, 专业号)
- 学生关系的“主修专业”与专业关系的主码“专业名”相对应
 - “主修专业”属性是学生关系的外码
 - 专业关系是被参照关系, 学生关系为参照关系





外码(续)

57

- [例2] : 学生 (学号, 姓名, 性别, 年龄, 所在系)
课程 (课程号, 课程名, 先修课, 学分)
学生选修 (学号, 课程号, 成绩, 开课学期)

选修关系的“学号”与学生关系的主码“学号”相对应

选修关系的“课程号”与课程关系的主码“课程号”相对应

- “学号”和“课程号”是选修关系的外码
- 学生关系和课程关系均为被参照关系
- 选修关系为参照关系



(b)



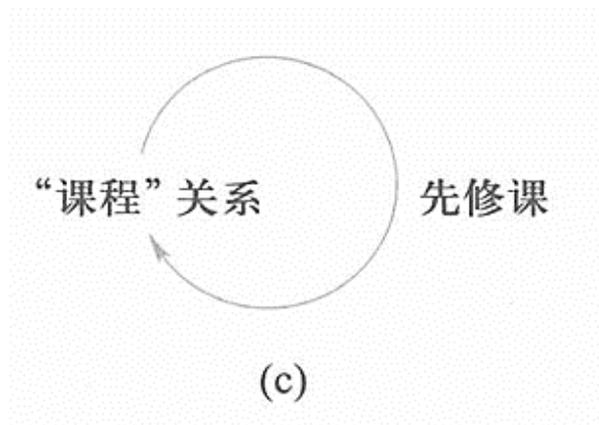
外码(续)

58

- [例3]：课程（课程号，课程名，先修课，学分）

“课程名”与本身的主码“课程号”相对应

- “先修课”是外码
- 课程关系既是参照关系也是被参照关系





外码(小结)

59

- 关系 R 和 S 不一定是不同的关系
- 目标关系 S 的主码 K_s 和参照关系的外码 F 必须定义在同一个（或一组）域上
- 外码并不一定要与相应的主码同名
 - 可以改名，但含义相同，但
 - 当外码与相应的主码属于不同关系时，往往取相同的名字，以便于识别



3. 参照完整性规则

60

规则2.2 参照完整性规则

若属性（或属性组） F 是基本关系 R 的外码，它与基本关系 S 的主码 K_s 相对应（基本关系 R 和 S 不一定是不同的关系），则对于 R 中每个元组在 F 上的值必须为：

- 或者取空值（ F 的每个属性值均为空值）
- 或者等于 S 中某个元组的主码值



参照完整性规则(续)

61

[例2.1]: **学生** (学号, 姓名, 性别, 专业号, 年龄)

专业 (专业名, 专业号)

学生关系中每个元组的“专业号”属性只取两类值:

- (1) **空值**, 表示尚未给该学生分配专业
- (2) **非空值**, 这时该值必须是专业关系中某个元组的“专业名”值, 表示该学生不可能分配一个不存在的专业



参照完整性规则(续)

62

(例2.2) : 学生 (学号, 姓名, 性别, 年龄, 所在系)
课程 (课程号, 课程名, 先修课, 学分)
学生选修 (学号, 课程号, 成绩, 开课学期)

“学号”和“课程号”可能的取值：

- (1) 选修关系中的主属性，**不能取空值**
- (2) 只能取相应被参照关系中已经存在的主码值



参照完整性规则(续)

63

例2.3) :

课程 (课程号, 课程名, 先修课, 学分)

“先修课”属性值可以取两类值:

- (1) 空值, 表示该课程无先修课
- (2) 非空值, 该值必须是本课程关系中某个元组的课程号值



完整性规则示例

64

建立一个“课程”表Course

```
CREATE TABLE Course
```

```
(Cno CHAR(4) PRIMARY KEY,
```

```
Cname CHAR(40),
```

```
Cpno, CHAR(4),
```

```
Ccredit INT,
```

```
FOREIGN KEY Cpno REFERENCE Course(Cno)
```

```
)
```



关系的完整性(续)

65

2.3.1 关系的三类完整性约束

2.3.2 实体完整性

2.3.3 参照完整性

2.3.4 用户定义的完整性



2.3.4 用户定义的完整性

66

- 用户定义完整性：针对某一具体关系数据库的约束条件，反映某一具体应用所涉及的数据必须满足的语义要求
 - 例如，课程关系中“课程名”可以要求“不能取空值”
- 关系模型应提供定义和检验这类完整性的机制，以使用统一的方法处理它们，而不要由应用程序承担这一功能
 - 例如，插入一条记录，需要检查其中的学号是否等于学生关系中的某个学号。若等于，则插入记录。否则，则拒绝



用户定义的完整性(续)

67

例:

课程(课程号, 课程名, 学分)

- “课程号” 属性必须取唯一值
- 非主属性 “课程名” 也不能取空值
- “学分” 属性只能取值{1, 2, 3, 4}



第二章 关系数据库

71

2.1 关系模型概述

2.2 关系数据结构

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

72

- 概述
- 传统的集合运算
- 专门的关系运算



概述

73

□ 关系代数概述

□ 抽象的查询语言，用对关系的运算来表达查询

□ 运算的三要素

■ 运算对象：关系

■ 运算结果：关系

■ （关系代数）运算符：集合运算符、专门的关系运算符

□ 传统的集合运算

■ 将关系看成元组的集合，“水平”方向，行的角度

□ 专门的关系运算

■ 涉及行，也涉及列

■ 还包含：比较运算符和逻辑运算符



概述

表2.4 关系代数运算符

运算符		含义
集合运算符	\cup	并
	$-$	差
	\cap	交
	\times	笛卡尔积
专门的关系运算符	σ	选择
	π	投影
	\bowtie	连接
	\div	除



概述

表2.4 关系代数运算符（续）

运算符		含义	运算符		含义
比较运算符	$>$	大于	逻辑运算符	\neg	非 与 或
	\geq	大于等于			
	$<$	小于			
	\leq	小于等于			
	$=$	等于			
	\neq	不等于			



2.4 关系代数

76

- 概述
- 传统的集合运算
- 专门的关系运算



1. 并 (Union)

77

□ 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
 - 例如, 学生和课程Union

□ $R \cup S$

- 仍为 n 目关系, 由属于 R 或属于 S 的元组组成

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



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

RUS

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



2. 差 (Difference)

79

- 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(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)

81

- 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 (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. 笛卡尔积 (Cartesian Product)

83

- 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: $(n+m)$ 目关系, $(n+m)$ 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
a1	b1	c1
a1	b2	c2
a2	b2	c1

S

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

R × 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

86

- 概述
- 传统的集合运算
- 专门的关系运算



2.4.2 专门的关系运算

87

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

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



专门的关系运算(续)

88

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

若 $A = \{A_{i_1}, A_{i_2}, \dots, A_{i_k}\}$, 其中 $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ 是 A_1, A_2, \dots, A_n 中的一部分, 则 A 称为属性列或属性组。

$t[A] = (t[A_{i_1}], t[A_{i_2}], \dots, t[A_{i_k}])$ 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 $\{A_{i_1}, A_{i_2}, \dots, A_{i_k}\}$ 后剩余的属性组。(R关系中)



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(Concatenation)

或元组的串接。

$\widehat{t_r t_s}$ 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。

(注意: 存在相同的属性(来自相同的域))



专门的关系运算(续)

90

(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：

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

这里 $t[X]$ 表示元组 t 中相应于属性 X 的一个分量

一个理解：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

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

- x_1 在 R 中的象集

$$Z_{x1} = \{Z1, Z2, Z3\},$$

- x_2 在 R 中的象集

$$Z_{x2} = \{Z2, Z3\},$$

- x_3 在 R 中的象集

$$Z_{x3} = \{Z1, Z3\}$$



专门的关系运算(续)

92

- 选择 (select)
- 投影 (project)
- 连接 (join)
- 除 (division)



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**

学号 Sno	姓名 Sname	性别 Ssex	年龄 Sage	所在系 Sdept
20180001	李勇	男	20	CS
20180002	刘晨	女	19	CS
20180003	王敏	女	18	MA
20180004	张立	男	19	IS
20180005	陈新奇	男	19	DS

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	Python语言	6	4
8	数据分析及 实践	6	3

学号 Sno	课程号 Cno	成绩 Grade	开课学期 Semester
20180001	1	92	20192
20180001	2	85	20201
20180001	3	88	20202
20180002	2	90	20192
20180002	3	80	20201

SC



1. 选择 (Selection)

- 1) 选择又称为限制 (Restriction)
- 2) 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: $X_1 \theta Y_1$

□ $X_1 \theta Y_1$

- θ 表示 比较运算符
- X_1, Y_1 为属性名、或常量
- 注意: 运算符的优先级

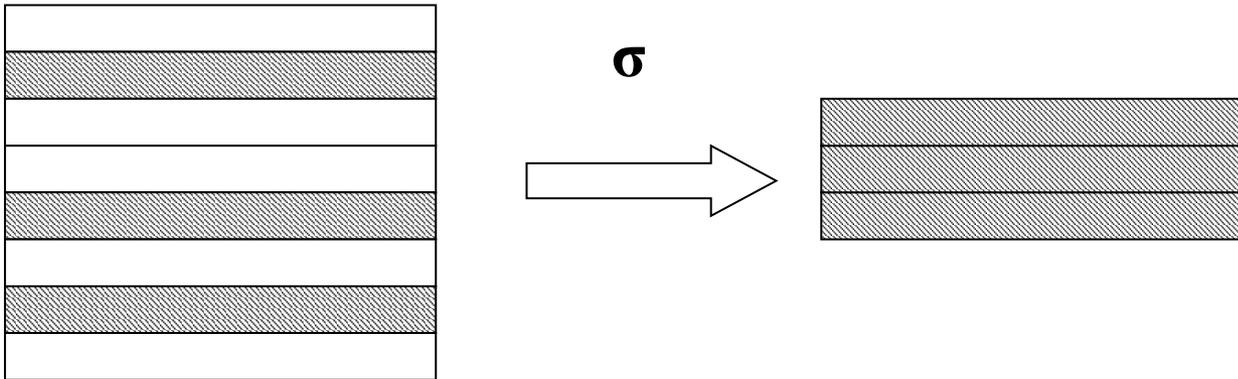
运算符		含义	运算符		含义
比较运算符	>	大于	逻辑运算符	\neg	非
	\geq	大于等于		\wedge	与
	<	小于		\vee	或
	\leq	小于等于			
	=	等于			
	<	不等于			
	>				



选择（续）

95

- 3) 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





选择 (续)

96

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 Sno	姓名 Sname	性别 Ssex	年龄 Sage	所在系 Sdept
20180001	李勇	男	20	CS
20180002	刘晨	女	19	CS
20180003	王敏	女	18	MA
20180004	张立	男	19	IS
20180005	陈新奇	男	19	DS



选择（续）

97

[例2.4] 查询信息系（IS系）全体学生

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

结果:

Sno	Sname	Ssex	Sage	Sdept
20180004	张立	男	19	IS



选择 (续)

98

[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
20180002	刘晨	女	19	CS
20180003	王敏	女	18	MA
20180004	张立	男	19	IS
20180005	陈新奇	男	19	DS



2. 投影 (Projection)

99

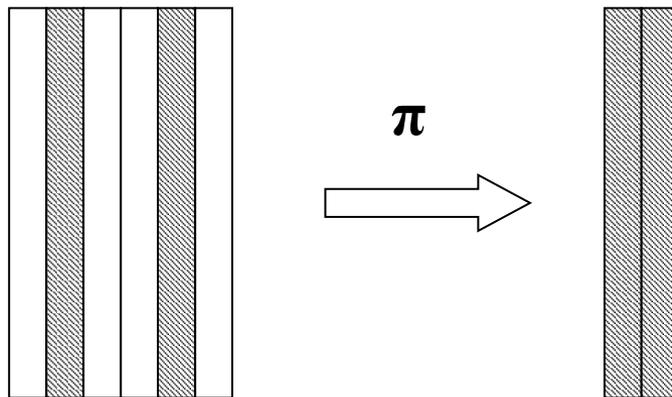
□ 1) 投影运算符的含义

- 从 R 中选择出若干属性列组成新的关系

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

A : R 中的属性列

- 投影操作主要是从列的角度进行运算



- 注意：但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,元组数可能减少）



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

Student

学号 Sno	姓名 Sname	性别 Ssex	年龄 Sage	所在系 Sdept
20180001	李勇	男	20	CS
20180002	刘晨	女	19	CS
20180003	王敏	女	18	MA
20180004	张立	男	19	IS
20180005	陈新奇	男	19	DS

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	Python 语言	6	4
8	数据分析 及实践	6	3

SC

学号 Sno	课程号 Cno	成绩 Grade	开课学期 Semester
20180001	1	92	20192
20180001	2	85	20201
20180001	3	88	20202
20180002	2	90	20192
20180002	3	80	20201



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

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

结果:

Sname	Sdept
李勇	CS
刘晨	CS
王敏	MA
张立	IS
陈新奇	DS



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
DS



3. 连接 (Join)

103

- 1) 连接也称为 θ 连接
- 2) 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

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

- A 和 B : 分别为 R 和 S 上**列数相等且可比**的属性组
- θ : 比较运算符
- 一种理解: 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 的元组作为结果元组



连接(续)

104

□ 3) 两类常用连接运算

□ 1. 等值连接 (equijoin)

➤ 什么是等值连接

θ为“=”的连接运算称为等值连接

➤ 等值连接的含义

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，即等值连接为：

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



连接(续)

105

□ 2. 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

■ 自然连接的含义

R 和 S 具有相同的属性组 B , Z 是 R 和 S 的全体属性集合

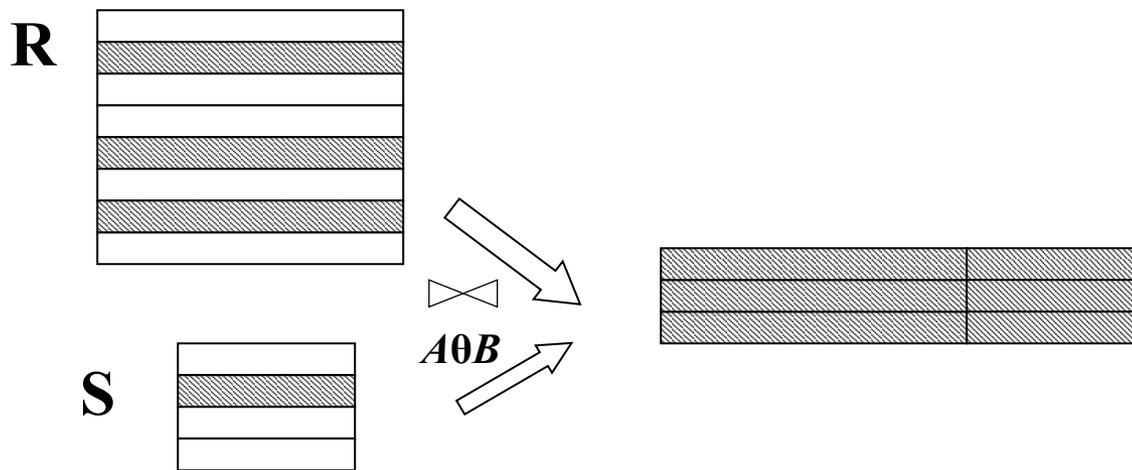
$$R \bowtie S = \{ \widehat{t_r t_s} [Z-B] \mid t_r \in R \wedge t_s \in S \wedge t_r[B] = t_s[B] \}$$

注意：自然连接与等值连接的区别



连接(续)

- 4) 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。



连接(续)

□ [例5]关系R和关系S 如下所示:

R

<i>A</i>	<i>B</i>	<i>C</i>
a_1	b_1	5
a_1	b_2	6
a_2	b_3	8
a_2	b_4	12

S

<i>B</i>	<i>E</i>
b_1	3
b_2	7
b_3	10
b_3	2
b_5	2



连接(续)

一般连接 $R \bowtie_{C < E} S$ 的结果如下:

R			S	
A	B	C	B	E
a ₁	b ₁	5	b ₁	3
a ₁	b ₂	6	b ₂	7
a ₂	b ₃	8	b ₃	10
a ₂	b ₄	12	b ₃	2
			b ₅	2

$R \bowtie_{C < E} S$

A	R.B	C	S.B	E
a ₁	b ₁	5	b ₂	7
a ₁	b ₁	5	b ₃	10
a ₁	b ₂	6	b ₂	7
a ₁	b ₂	6	b ₃	10
a ₂	b ₃	8	b ₃	10



连接(续)

等值连接 $R \bowtie_{R.B=S.B} S$ 的结果如下:

R			S	
A	B	C	B	E
a ₁	b ₁	5	b ₁	3
a ₁	b ₂	6	b ₂	7
a ₂	b ₃	8	b ₃	10
a ₂	b ₄	12	b ₃	2
			b ₅	2

A	R.B	C	S.B	E
a ₁	b ₁	5	b ₁	3
a ₁	b ₂	6	b ₂	7
a ₂	b ₃	8	b ₃	10
a ₂	b ₃	8	b ₃	2



连接(续)

自然连接 $R \bowtie S$ 的结果如下:

R			S	
A	B	C	B	E
a_1	b_1	5	b_1	3
a_1	b_2	6	b_2	7
a_2	b_3	8	b_3	10
a_2	b_4	12	b_3	2
			b_5	2

A	B	C	E
a_1	b_1	5	3
a_1	b_2	6	7
a_2	b_3	8	10
a_2	b_3	8	2

悬浮元组 (dangling tuple)



连接(续)

111

- 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。



连接(续)

112

- 外连接
 - 如果把舍弃的元组也保存在结果关系中，而在其他属性上填空值(Null)，这种连接就叫做**外连接**(OUTER JOIN)。记作：P51页
- 左外连接
 - 如果只把**左边关系R**中要舍弃的元组保留就叫做**左外连接**(LEFT OUTER JOIN或LEFT JOIN)
- 右外连接
 - 如果只把**右边关系S**中要舍弃的元组保留就叫做**右外连接**(RIGHT OUTER JOIN或RIGHT JOIN)。



连接(续)

下图是例5中关系R和关系S的外连接

R			S	
A	B	C	B	E
a ₁	b ₁	5	b ₁	3
a ₁	b ₂	6	b ₂	7
a ₂	b ₃	8	b ₃	10
a ₂	b ₃	8	b ₃	2
a ₂	b ₄	12	b ₅	2

A	B	C	E
a ₁	b ₁	5	3
a ₁	b ₂	6	7
a ₂	b ₃	8	10
a ₂	b ₃	8	2
a ₂	b ₄	12	NULL
NULL	b ₅	NULL	2

(a) 外连接



连接(续)

图(b)是例5中关系R和关系S的左外连接,图(c)是右外连接

R			S	
A	B	C	B	E
a ₁	b ₁	5	b ₁	3
a ₁	b ₂	6	b ₂	7
a ₂	b ₃	8	b ₃	10
a ₂	b ₃	8	b ₃	2
a ₂	b ₄	12	b ₅	2

A	B	C	E
a ₁	b ₁	5	3
a ₁	b ₂	6	7
a ₂	b ₃	8	10
a ₂	b ₃	8	2
a ₂	b ₄	12	NULL

(b) 左外连接

A	B	C	E
a ₁	b ₁	5	3
a ₁	b ₂	6	7
a ₂	b ₃	8	10
a ₂	b ₃	8	2
NULL	b ₅	NULL	2

(c) 右外连接



4. 除 (Division)

115

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。

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, $x = t_r[X]$

一种理解: $P(X)$ 包含所有在 R 但不在 S 中的属性 X 及其值的元组,

而且 $P(X)$ 的元组与 S 的元组的所有组合都在 R 中

R 中的元组, 这些元组在 R 中的象集 Y_x 必须完全覆盖与 S 在 Y 上的投影



除(续)

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

Y_x : x 在 R 中的象集, $x = t_r[X]$

[例2.10] 设关系 R 、 S 分别为下图的(a)和(b), $R \div S$ 的结果为图(c)

R		
A	B	C
a_1	b_1	c_2
a_2	b_3	c_7
a_3	b_4	c_6
a_1	b_2	c_3
a_4	b_6	c_6
a_2	b_2	c_3
a_1	b_2	c_1

(a)

S		
B	C	D
b_1	c_2	d_1
b_2	c_1	d_1
b_2	c_3	d_2

(b)

$R \div S$
A
a_1

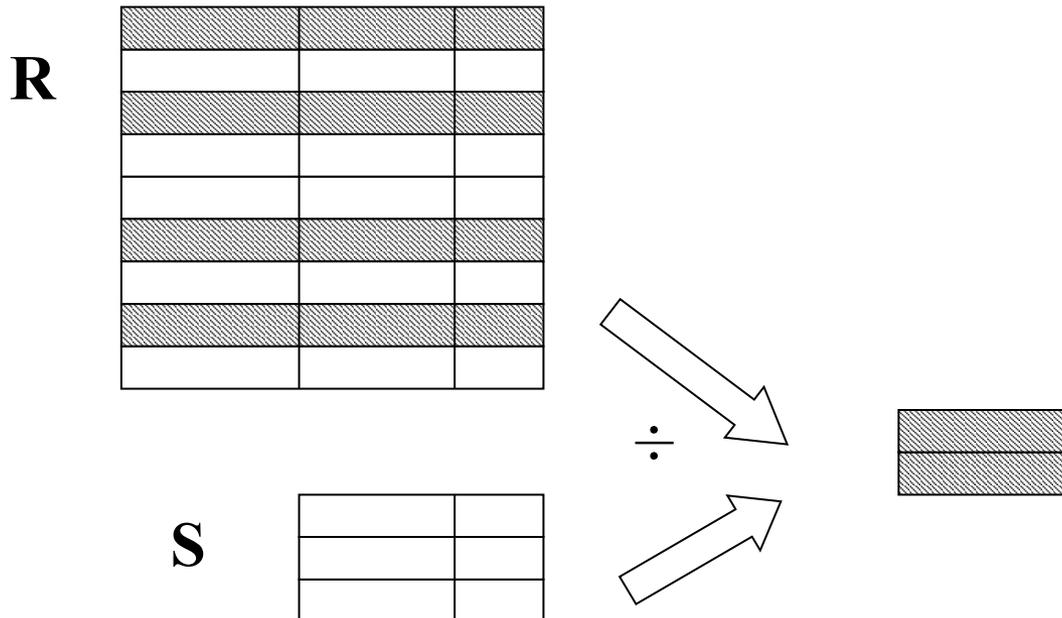
(c)

- ❖ 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 $R \div S = \{a_1\}$



除(续)

- 2) 除操作是同时从行和列角度进行运算





5. 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

Student

学号 Sno	姓名 Sname	性别 Ssex	年龄 Sage	所在系 Sdept
20180001	李勇	男	20	CS
20180002	刘晨	女	19	CS
20180003	王敏	女	18	MA
20180004	张立	男	19	IS
20180005	陈新奇	男	19	DS

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	Python语言	6	4
8	数据分析及实践	6	3

SC

学号 Sno	课程号 Cno	成绩 Grade	开课学期 Semester
20180001	1	92	20192
20180001	2	85	20201
20180001	3	88	20202
20180002	2	90	20192
20180002	3	80	20201



综合举例(续)

[例 2.11] 查询选修了2号课程的学生们的学号。

$$\pi_{Sno} (\sigma_{Cno='2'} (SC))$$
$$= \{ 20180001, 20180002 \}$$

SC

学号 Sno	课程号 Cno	成绩 Grade	开课学期 Semester
20180001	1	92	20192
20180001	2	85	20201
20180001	3	88	20202
20180002	2	90	20192
20180002	3	80	20201

请写出其他表达式？



综合举例(续)

122

[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

Student

学号 Sno	姓名 Sname	性别 Ssex	年龄 Sage	所在系 Sdept
20180001	李勇	男	20	CS
20180002	刘晨	女	19	CS
20180003	王敏	女	18	MA
20180004	张立	男	19	IS
20180005	陈新奇	男	19	DS

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	Python 语言	6	4
8	数据分析 及实践	6	3

SC

学号 Sno	课程号 Cno	成绩 Grade	开课学期 Semester
20180001	1	92	20192
20180001	2	85	20201
20180001	3	88	20202
20180002	2	90	20192
20180002	3	80	20201



综合举例(续)

[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

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

Step1: $\sigma_{Cpno='5'}(Course)$

课程号 Cno	课程名 Cname	先行课 Cpno	学分 Ccredit
1	数据库	5	4

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	Python 语言	6	4
8	数据分析 及实践	6	3



综合举例(续)

[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

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

Step2: $\sigma_{Cpno='5'}(Course) \bowtie SC$

课程号 Cno	课程名 Cname	先行课 Cpno	学分 Ccredit
1	数据库	5	4



SC

学号 Sno	课程号 Cno	成绩 Grade	开课学期 Semester
20180001	1	92	20192
20180001	2	85	20201
20180001	3	88	20202
20180002	2	90	20192
20180002	3	80	20201

学号 Sno	课程号 Cno	课程名 Cname	先行课 Cpno	学分 Ccredit	成绩 Grade	开课学期 Semester
20180001	1	数据库	5	4	92	20192



综合举例(续)

[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

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

Step3: $\sigma_{Cpno='5'}(Course) \bowtie SC \bowtie \pi_{Sno, Sname}(Student)$

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



学号 Sno	姓名 Sname
20180001	李勇
20180002	刘晨
20180003	王敏
20180004	张立
20180005	陈新奇

学号 Sno	姓名 Sname	课程号 Cno	课程名 Cname	先行课 Cpno	学分 Ccredit	成绩 Grade	开课学期 Semester
20180001	李勇	1	数据库	5	4	92	20192



综合举例(续)

[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

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

Step4: $\pi_{Sname}(\sigma_{Cpno='5'}(Course) \bowtie SC \bowtie \pi_{Sno, Sname}(Student))$

学号 Sno	姓名 Sname	课程号 Cno	课程名 Cname	先行课 Cpno	学分 Ccredit	成绩 Grade	开课学期 Semester
20180001	李勇	1	数据库	5	4	92	20192

姓名
Sname
李勇



综合举例(续)

127

[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

方式1: $\pi_{\text{Sname}}(\sigma_{\text{Cpno}='5'}(\text{Course}) \bowtie \text{SC} \bowtie \pi_{\text{Sno}, \text{Sname}}(\text{Student}))$

方式2: $\pi_{\text{Sname}}(\pi_{\text{Sno}}(\sigma_{\text{Cpno}='5'}(\text{Course}) \bowtie \text{SC}) \bowtie \pi_{\text{Sno}, \text{Sname}}(\text{Student}))$

方式3: $\pi_{\text{Sname}}(\sigma_{\text{Cpno}='5'}(\text{Course} \bowtie \text{SC} \bowtie \text{Student}))$



综合举例(续)

[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

方法2: $\pi_{Sname} (\pi_{Sno} (\sigma_{Cpno='5'} (Course) \bowtie SC) \bowtie \pi_{Sno, Sname} (Student))$

Step1: $\sigma_{Cpno='5'} (Course)$

课程号 Cno	课程名 Cname	先行课 Cpno	学分 Ccredit
1	数据库	5	4

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	Python语言	6	4
8	数据分析及实践	6	3



综合举例(续)

[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

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

Step2: $\sigma_{Cpno='5'} (\text{Course}) \bowtie \text{SC}$

课程号 Cno	课程名 Cname	先行课 Cpno	学分 Ccredit
1	数据库	5	4

SC

学号 Sno	课程号 Cno	成绩 Grade	开课学期 Semester
20180001	1	92	20192
20180001	2	85	20201
20180001	3	88	20202
20180002	2	90	20192
20180002	3	80	20201



学号 Sno	课程号 Cno	课程名 Cname	先行课 Cpno	学分 Ccredit	成绩 Grade	开课学期 Semester
20180001	1	数据库	5	4	92	20192



综合举例(续)

[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

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

Step3: $\pi_{Sno} (\sigma_{Cpno='5'} (Course) \bowtie SC)$

学号 Sno
20180001



学号 Sno	课程号 Cno	课程名 Cname	先行课 Cpno	学分 Ccredit	成绩 Grade	开课学期 Semester
20180001	1	数据库	5	4	92	20192



综合举例(续)

[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

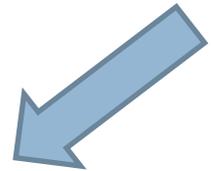
$$\pi_{Sname} (\pi_{Sno} (\sigma_{Cpno='5'} (\text{Course}) \bowtie \text{SC}) \bowtie \pi_{Sno, Sname} (\text{Student}))$$

Step4: $\pi_{Sno} (\sigma_{Cpno='5'} (\text{Course}) \bowtie \text{SC}) \bowtie \pi_{Sno, Sname} (\text{Student})$

学号 Sno
20180001



学号 Sno	姓名 Sname
20180001	李勇



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

学号 Sno	姓名 Sname
20180001	李勇
20180002	刘晨
20180003	王敏
20180004	张立
20180005	陈新奇



综合举例(续)

[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

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

Step5: $\pi_{Sname} (\pi_{Sno} (\sigma_{Cpno='5'} (\text{Course}) \bowtie \text{SC}) \bowtie \pi_{Sno, Sname} (\text{Student}))$

学号 Sno	姓名 Sname
20180001	李勇



姓名 Sname
李勇



5. 综合举例

133

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)

K=

Cno
1
3

□ [例2.10] $\pi_{Sno,Cno}(SC)$

20180001 象集{1, 2, 3}

20180002 象集{2, 3}

K={1, 3}

于是:

$\pi_{Sno,Cno}(SC) \div K = \{20180001\}$

Sno	Cno
20180001	1
20180001	2
20180001	3
20180002	2
20180002	3



综合举例(续)

135

[例2.13] 查询选修了全部课程的学生号码和姓名。

- 问题思考：
1. 找到全部课程（Course表）
 2. 找到选修了全部课程的学生学号（SC表）；
 3. 找到相应的学生姓名（Student表）



专门的关系运算(续)

[例2.13] 查询选修了全部课程的学生号码和姓名。

Student

学号 Sno	姓名 Sname	性别 Ssex	年龄 Sage	所在系 Sdept
20180001	李勇	男	20	CS
20180002	刘晨	女	19	CS
20180003	王敏	女	18	MA
20180004	张立	男	19	IS
20180005	陈新奇	男	19	DS

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	Python 语言	6	4
8	数据分析 及实践	6	3

SC

学号 Sno	课程号 Cno	成绩 Grade	开课学期 Semester
20180001	1	92	20192
20180001	2	85	20201
20180001	3	88	20202
20180002	2	90	20192
20180002	3	80	20201



综合举例(续)

137

[例2.13] 查询选修了全部课程的学生号码和姓名。

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



综合举例(续)

138

[例2.13] 查询选修了全部课程的学生号码和姓名。

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

Step1: $\pi_{Cno} (Course)$

Course

课程号 Cno	课程名 Cname	先行课 Cpno	学分 Ccredit	课程号 Cno
1	数据库	5	4	1
2	数学		2	2
3	信息系统	1	4	3
4	操作系统	6	3	4
5	数据结构	7	4	5
6	数据处理		2	6
7	Python语言	6	4	7
8	数据分析及 实践	6	3	8



综合举例(续)

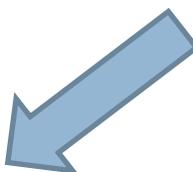
[例2.13] 查询选修了全部课程的学生号码和姓名。

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

Step2: $\pi_{Sno, Cno} (SC)$

修改后的SC

学号 Sno	课程号 Cno
20180001	1
20180001	2
20180001	3
20180001	4
20180001	5
20180001	6
20180001	7
20180001	8
20180002	2
20180002	3



学号 Sno	课程号 Cno	成绩 Grade	开课学期 Semester
20180001	1	92	20192
20180001	2	85	20201
20180001	3	88	20202
20180001	4	82	20202
20180001	5	87	20192
20180001	6	88	20202
20180001	7	91	20202
20180001	8	90	20202
20180002	2	90	20192
20180002	3	80	20201



综合举例(续)

[例2.13] 查询选修了全部课程的学生号码和姓名。

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

Step3: $\pi_{Sno, Cno} (SC) \div \pi_{Cno} (Course)$

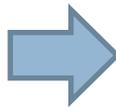
$\pi_{Sno, Cno} (SC)$

学号 Sno	课程号 Cno
20180001	1
20180001	2
20180001	3
20180001	4
20180001	5
20180001	6
20180001	7
20180001	8
20180002	2
20180002	3



$\pi_{Cno} (Course)$

课程号 Cno
1
2
3
4
5
6
7
8



学号 Sno
20180001



综合举例(续)

[例2.13] 查询选修了全部课程的学生号码和姓名。

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

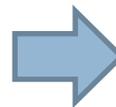
Step4: $\pi_{Sno, Cno} (SC) \div \pi_{Cno} (Course) \bowtie \pi_{Sno, Sname} (Student)$

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

学号 Sno
20180001



学号 Sno	姓名 Sname
20180001	李勇
20180002	刘晨
20180003	王敏
20180004	张立
20180005	陈新奇



学号 Sno	姓名 Sname
20180001	李勇