

数据库系统原理与设计

第2章 关系模型与关系代数

目 录



关系模型



关系代数

2.1 关系模型

- 系统而严格地提出关系模型的是美国IBM公司的E.F.Codd
 - 1970年提出关系数据模型
E.F.Codd, “A Relational Model of Data for Large Shared Data Banks”, 《Communication of the ACM》,1970
- 关系数据库系统是支持关系数据模型的数据库系统。
关系数据库管理系统是当今的主流数据库管理系统。
- 关系模型由**关系数据结构**、**关系完整性约束**和**关系操作集合**三部分组成。

2.1.1 关系数据结构

- 关系
- 关系模式
- 码
- 关系数据库模式

2.1.1 关系数据结构

■ 关系

- 关系模型的数据结构非常简单，它就是二维表，亦称为关系。
- 关系数据库是表的集合，即关系的集合。
- 表是一个实体集，一行就是一个实体，它由共同表示一个实体的有相关联的若干属性的值所构成。
- 由于一个表是这种有相关联的值的集合(即行的集合)，而表这个概念和数学上的关系概念密切相关，因此称为关系模型。
- 关系模型中，现实世界的实体以及实体间的各种联系都是用关系来表示。

2.1.1 关系数据结构

举例：

给定两个域：

学生的姓名集合： $D_1 = \{ \text{‘李小勇’}, \text{‘刘方晨’}, \text{‘王红敏’} \}$

课程的名称集合： $D_2 = \{ \text{‘数据库系统概论’}, \text{‘操作系统’} \}$

则 D_1, D_2 的笛卡尔积为：

$$\begin{aligned} D_1 \times D_2 = \{ & (\text{‘李小勇’}, \text{‘数据库系统概论’}), \\ & (\text{‘李小勇’}, \text{‘操作系统’}), \\ & (\text{‘刘方晨’}, \text{‘数据库系统概论’}), \\ & (\text{‘刘方晨’}, \text{‘操作系统’}), \\ & (\text{‘王红敏’}, \text{‘数据库系统概论’}), \\ & (\text{‘王红敏’}, \text{‘操作系统’}) \} \end{aligned}$$

2.1.1 关系数据结构

- 定义2.1 域是一组具有相同数据类型的值的集合
- 空值(null)是所有可能的域的一个取值, 表明值未知或值不存在。
 - 对于学位的取值域, 某员工的学位为空值null, 表示不知道该员工所获得的学位, 或该员工没有获得学位;
 - 对于成绩的取值域, 某学生的成绩为空值null, 表示不知道该学生的成绩, 或该学生没有成绩(如没有参加考试就没有获得成绩)。
- 关系的最基本要求:
 - 关系中的每个属性的域必须是原子的, 即域中的每个值都是不可再分的一个完整单元。
 - 关系中的每个元组都是可区分的, 即存在唯一标识不同元组的属性(集)——码。

2.1.1 关系数据结构

■ 关系的形式化定义（定义2.2）

- $D_1 \times D_2 \times \dots \times D_n$ 的子集称为在域 D_1, D_2, \dots, D_n 上的关系，表示为：

$$r(D_1, D_2, \dots, D_n)$$

其中， r 表示关系的名字， n 是关系的目或度(degree)。当 $n=1$ 时，该关系称为单元关系；当 $n=2$ 时，称为二元关系。

- 关系是笛卡尔积的有限子集，所以关系也是一个二维表，表的每行对应于关系的一个元组，表的每列对应于关系的一个域。由于域可以相同，为了区别就必须给每列起一个名字，称为属性(attribute)。 n 目关系共有 n 个属性。

2.1.1 关系数据结构

■ 关系模式

- 对于一个二维表，有表头部分和表体部分：
 - 表头部分定义了该表的结构，即定义了该表由哪些列构成(假设由 n 列构成)，每个列的名字和取值范围等；
 - 表体部分就是所有数据行(元组)的集合，每一个数据行都是由表头部分规定的 n 列有关联的取值的集合构成。
- 对应于关系数据库，表的每一个数据行对应于关系的一个元组，表体对应于关系，关系是元组的集合，关系是值的概念；
- 表头部分对应于关系模式，关系模式是型的概念，它定义了元组集合的结构，即定义了一个元组由哪些属性构成(假设由 n 个属性构成)，每个属性的名字和来自的域等。

2.1.1 关系数据结构

■ 关系模式

- 关系的描述称为**关系模式**(relation schema)。

- 形式化地表示为： $r(U, D, DOM, F)$

r 为关系名， U 为组成该关系的属性名的集合， D 为属性集 U 中所有属性所来自的域的集合， DOM 为属性向域的映像集合， F 为属性间**数据的依赖关系**集合(即体现一个元组的各属性取值之间的“**关联**”性)。

- 关系模式通常被简记为： $r(U)$ 或 $r(A_1, A_2, \dots, A_n)$

r 为关系名， U 为属性名的集合 $\{A_1, A_2, \dots, A_n\}$

2.1.1 关系数据结构

■ 码

● **超码**：对于关系 r 的一个或多个属性的集合 A ，如果

举例：Student关系

studentNo	studentName	sex	birthday	speciality
0701001	李小勇	男	1990-12-21	计算机
0701008	王 红	男	1992-04-26	计算机
0802002	刘方晨	女	1990-11-11	信息系统
0802005	王红敏	女	1990-10-01	信息系统
0703045	王 红	男	1992-04-26	会计学
0703010	李宏冰	女	1992-03-09	会计学

超码？：studentNo, {studentNo, studentName}、{studentNo, sex}、speciality、{studentNo, studentName, speciality}、sex？

候选码：studentNo 和 {studentName, speciality} —— 与语义有关

主码：studentNo

学生文件Student的记录结构

学号	姓名	性别	出生日期	所学专业	家庭住址	联系电话
----	----	----	------	------	------	------

2.1.1

课程文件Course的记录结构

课程号	课程名称	学时	学分	教材名称
-----	------	----	----	------

学生成绩文件Score的记录结构

学号	课程号	学期	成绩
----	-----	----	----

图1-3 学生、课程、学生成绩文件结构

■ 码

- **外码**：设 F 是关系 r 的码， s 是关系 s 的主码。如果 F 的取值范围对应于关系 s 中主码 K_s 的取值范围的子集)，则称 F 是关系 r 参照关系 s 的**外码**(foreign key)，简称 F 是关系 r 的**外码**。

➤ 称关系 r 为**参照关系**，关系 s 为**被参照关系**或**目标关系**。

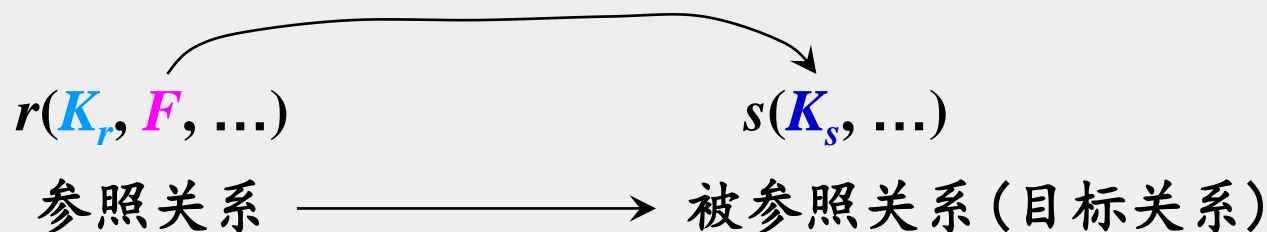


图2-3 外码参照图

2.1.1 关系数据结构

■ 关系数据库模式

举例：学生成绩管理数据库ScoreDB的模式

Class(classNo: char(6), **className**: varchar(30), **institute**: varchar(30),
grade: smallint, **classNum**: tinyint)

Student(studentNo: char(7), **studentName**: varchar(20), **sex**: char(2), **birthday**:
datetime, **native**: varchar(20), **nation**: varchar(30), *classNo*: char(6))

Course(courseNo: char(5), **courseName**: varchar(30), **creditHour**: numeric,
courseHour: int, *priorCourse*: char(5))

Score(*studentNo*: char(7), *courseNo*: char(5), term: char(5), **score**: numeric)

说明：带下划线的属性(集)为**主码**，**斜体**属性为**外码**。

基本**数据类型**请参见3.8.2节。

小结

- 主要讲述了**关系数据结构**的概念，如关系，关系模式，码，和关系数据库的模式集等

关系完整性约束条件

2.1.2 关系完整性约束条件

- 实体完整性
- 参照完整性
- 自定义完整性

2.1.2 关系完整性约束条件

■ 实体完整性

- 若属性集A是关系r的主码，则A不能取空值null。

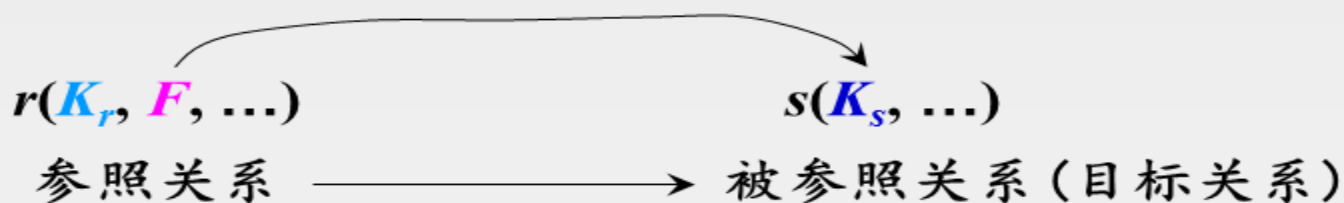
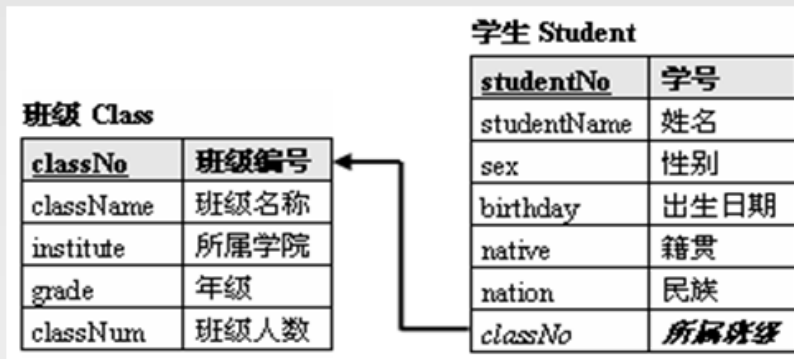
➤例如，关系Student，由于studentNo是关系Student的主码，因此它在任何时候的取值都不能为空值null，但其他属性如sex、birthday、speciality等都可以取空值，表示当时该属性的值未知或不存在。

- 如果主码是由若干个属性的集合构成，则要求构成主码的每一个属性的值都不能取空值。例如，学生成绩管理数据库ScoreDB中的关系Score，它的主码是{studentNo, courseNo, term}，因此这3个属性都不能取空值。

关系完整性约束条件

■ 参照完整性

- 若属性(或属性集) F 是关系 r 的外码，它与关系 s 的主码 K_s 相对应，则对于关系 r 中的每一个元组在属性 F 上的取值要么为空值null，要么等于关系 s 中某个元组的主码值。



classno	classname	...	classnum
1	计算机1		29
2	计算机2		32
3	计算机3		31

studentno	studentname	...	classno
10153111	梁泽铎		1
10160960	施智罡		1
10161706	薛盛丰		2
10161733	李金明		
10161764	张诚		2
10161765	张琮昊		2
10161766	邓积懋		1

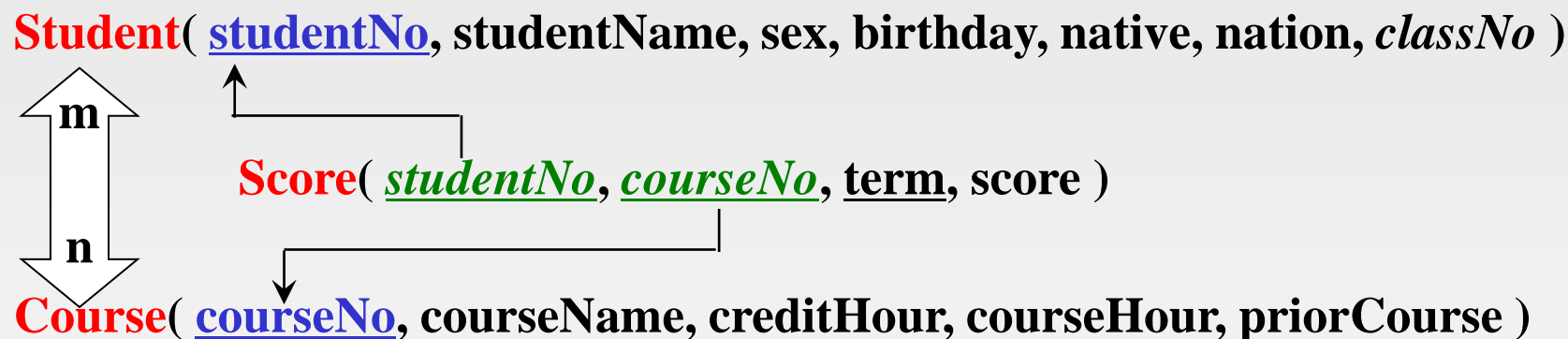


图2-5 实现“多对多”联系的联系关系及外码

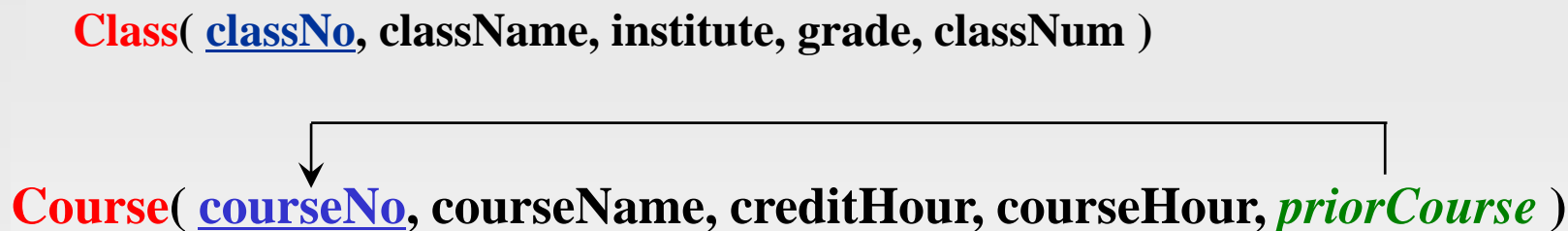


图2-6 实现关系内部多对一联系的外码

图2-4 实现“多对一”联系的外码

学生成绩管理数据库**ScoreDB**中，假设一门课程可能存在先修课程，且关系**Course**中的**priorCourse**属性用来存放先修课程的课程编号。属性**priorCourse**是课程关系**Course**参照课程关系**Course**的外码。

关系完整性约束条件

■ 数据库模式导航图

- 一个含有主码和外码依赖的数据库模式可以通过**模式导航图**来表示。

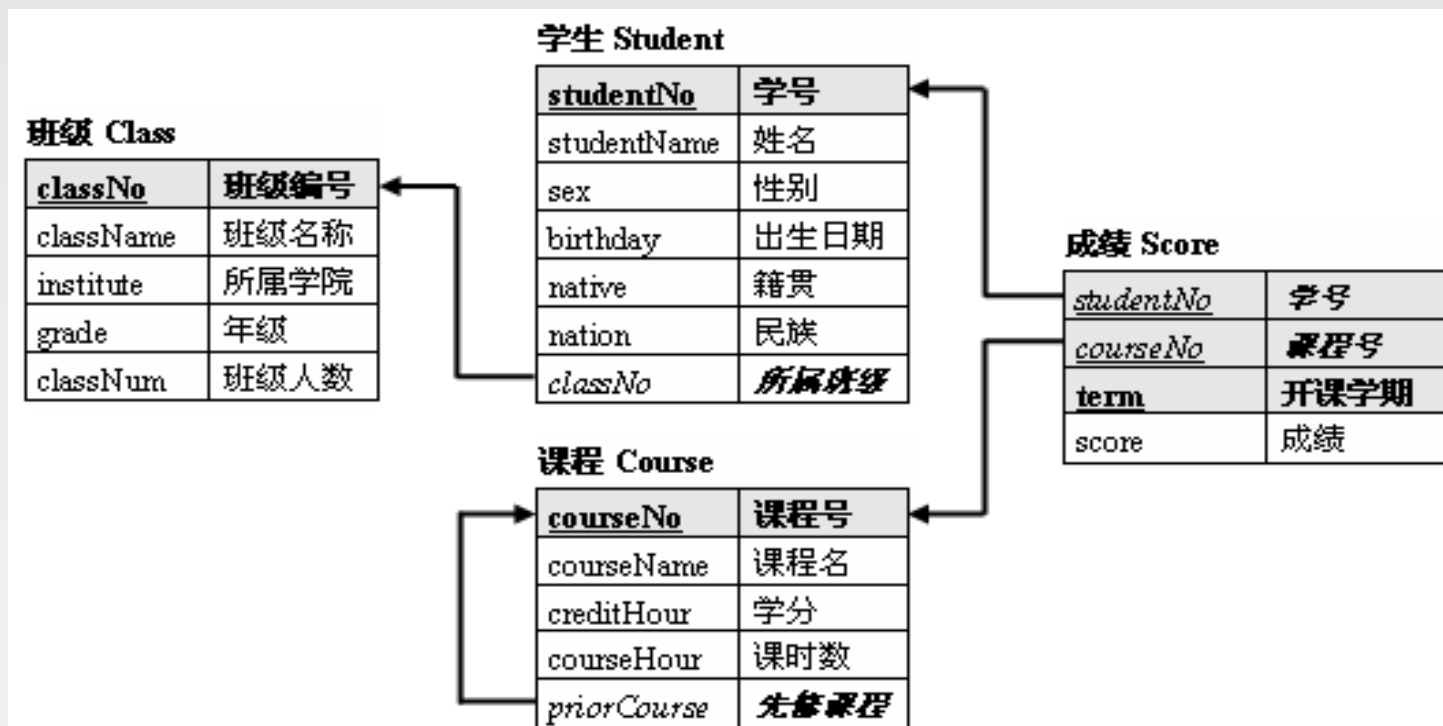


图 2-8 成绩管理数据库 ScoreDB 的模式导航图

关系完整性约束条件

■ 数据库模式导航图

- 一个含有主码和外码依赖的数据库模式可以通过**模式导航图**来表示。

- 关系Student与关系Class之间存在多对一的“归属”联系(一个班由多个学生组成，一个学生只能归属于某个班)，通过外码classNo实现该联系。
- 关系Course与关系Student之间存在多对多的“选修”联系。
- 关系Score的主码是{studentNo, courseNo, term}，显然同一个学生在同一个学期不允许修读同一门课程多次。
- 关系Course的外码priorCourse参照本关系的主码courseNo。

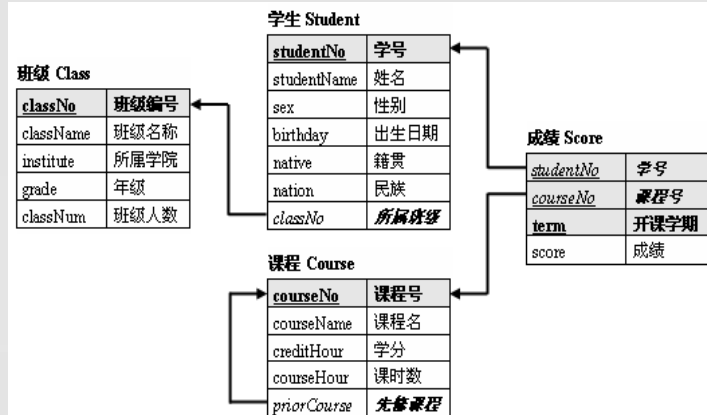


图 2-8 成绩管理数据库 ScoreDB 的模式导航图

关系完整性约束条件

■ 数据库模式导航图

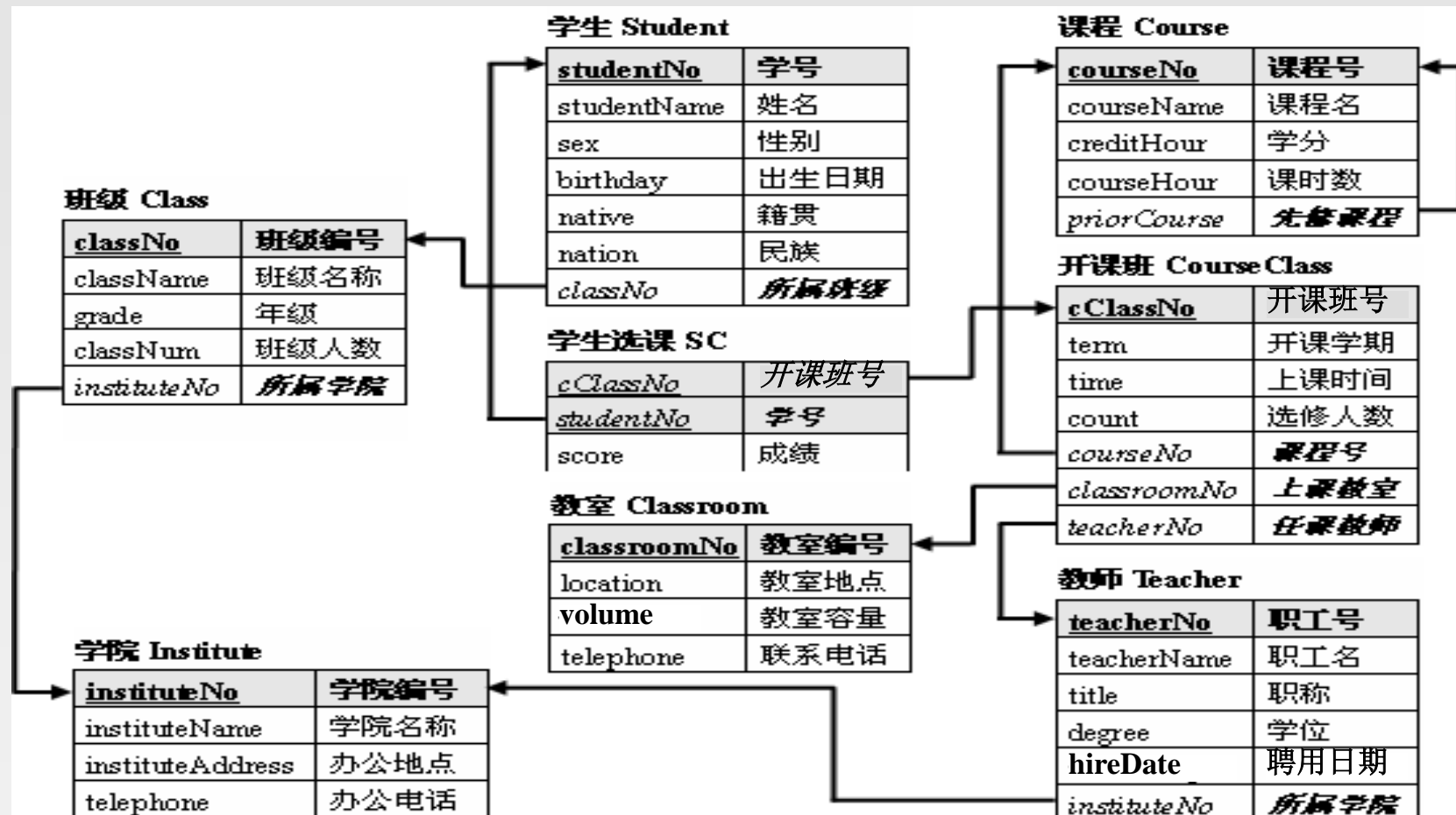


图 2-9 学生选课数据库 SCDB 的模式导航图

关系完整性约束条件

数据库模式导航图

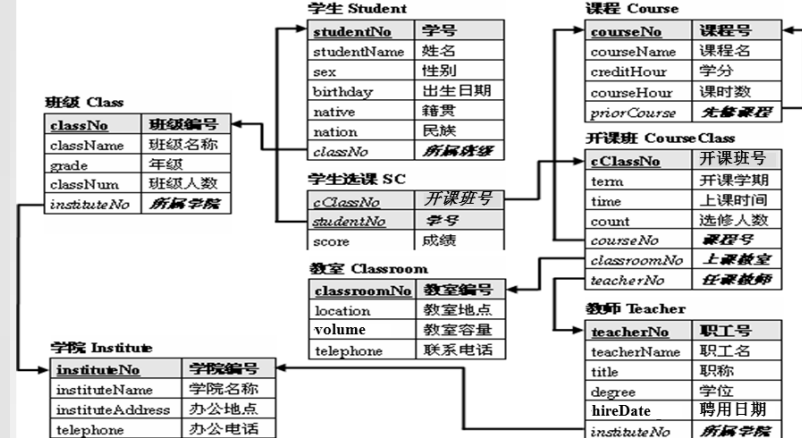


图 2-9 学生选课数据库 SCDB 的模式导航图

● 一个含有主码和外码依赖的数据库模式可以通过**模式导航图**来表示

- 关系 **Student** 与关系 **Class** 之间、关系 **Class** 与关系 **Institute** 之间、关系 **Teacher** 与关系 **Institute** 之间都存在多对一的“归属”联系。
- 关系 **Course** 的外码 **priorCourse** 参照本关系的主码 **courseNo**。
- 关系 **CourseClass** 与关系 **Course** 之间存在多对一的“开课”联系(每个学期一门课程可能开设多个教学班，一个教学班只讲授一门课程)。
- 关系 **CourseClass** 与关系 **Classroom** 之间存在多对一的“上课”联系
- 关系 **CourseClass** 与关系 **Teacher** 之间存在多对一的“授课”联系
- 关系 **CourseClass** 与关系 **Student** 之间存在多对多的“选课”联系。
- 关系 **SC** 的主码是 {**cClassNo**, **studentNo**}，显然同一个学生不允许选修同一个教学班的课程多次。

关系完整性约束条件

■ 用户自定义完整性

- 任何关系数据库管理系统都应该支持**实体完整性**和**参照完整性**。
- **用户定义的完整性**就是针对某一具体应用要求来定义的约束条件，它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，
 - 限制关系中某些属性的取值要符合业务语义要求。
 - 限制关系中某些属性的取值之间需要满足一定的逻辑关系。
 - 限制关系中某属性集上的取值必须唯一。

小结

- 讲述了关系模型中关系完整性约束条件。分为三种约束。

■ 2019.9.4

关系操作

■ 关系操作

- 关系操作的特点是**集合操作方式**，即操作的对象和结果都是集合。这种操作方式也称为**一次一个集合的方式**。相应地，非关系数据模型的数据操作方式则为**一次一个记录的方式**。
- 关系模型中的关系操作有**查询操作**和**更新操作**(插入、删除和修改)两大类。
- 查询操作是关系操作中最主要的部分。查询操作又可以分为**选择(select)**、**投影(project)**、**连接(join)**、**除(divide)**、**并(union)**、**交(intersection)**、**差(except)**、**笛卡尔积**等。

<操作符> <工作空间名> (<目标表>)[:<操作条件>]

内存空间 (W)

GET
PUT
HOLD
UPDATE
DELETE
DROP

(1) 查询所有被选修的课程号码

GET W (SC.Cno)

(2) 查询所有学生的数据

GET W (Student)

关系操作

■ 关系操作能力可用两种方式来表示
——代数方式和逻辑方式。

- 关系代数是用代数方式表达的关系查询语言。
- 关系演算是用逻辑方式表达的关系查询语言。
- 对于关系代数、关系演算均是抽象的查询语言，在表达能力上是完全等价的。

目 录

2.1

关系模型

2.2

关系代数

- ✓ 传统的集合运算
- ✓ 专门的关系运算
- ✓ 举例说明

关系代数

■ 关系代数

- 关系代数是通通过关系代数运算构成的表达式来表达查询。
- 基本的关系代数运算有选择、投影、集合并、集合差、笛卡尔积和更名等。
- 关系代数运算是以一个或两个关系作为输入(即运算对象)产生一个新的关系作为结果。
- 切记关系代数运算也是集合运算，下面分为传统集合运算和专门的关系运算

关系代数

r	id	age	addr
	1	18	haisi 99

s	id	nianling	dizhi
	3	19	meilong130

■ 传统的集合运算

- 前提假设：关系 r 和关系 s 具有相同的 n 个属性，且相应的属性取自同一个域，即两个关系的模式或结构相同。

t 是元组变量， $t \in r$ 表示 t 是 r 的一个元组。

● 并运算

关系 r 与关系 s 的并记作： $r \cup s = \{ t \mid t \in r \vee t \in s \}$

其结果关系仍为 n 目关系，由属于 r 或属于 s 的所有元组组成。

并运算

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

RUS

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

关系代数

■ 传统的集合运算

● 差

关系 r 与关系 s 的差记作： $r - s = \{ t \mid t \in r \wedge t \notin s \}$

其结果关系仍为 n 目关系，由属于 r 而不属于 s 的所有元组组成。

● 交

关系 r 与关系 s 的交记作： $r \cap s = \{ t \mid t \in r \wedge t \in s \}$

其结果关系仍为 n 目关系，由既属于 r 又属于 s 的所有元组组成。关系的交可以通过差来表达，即 $r \cap s = r - (r - s)$ 。

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

关系代数

R(3目)			S (2目)	
id	age	addr	id	nianling
1	18	haisi 99	3	19
...

■ 传统的集合运算

● 笛卡尔积

- 两个分别为 n 目和 m 目的关系 r 和 s 的笛卡尔积是一个 $n+m$ 目元组的集合。
- 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
- 若关系 r 有 k_r 个元组，关系 s 有 k_s 个元组，则关系 r 和 s 的笛卡尔积有 $k_r \times k_s$ 个元组。记作：

$$r \times s = \{ t_r \cdot t_s \mid t_r \in r \wedge t_s \in s \}$$

笛卡尔积

- 两个分别为 n 目和 m 目的关系 r 和 s 的笛卡尔积是一个 $n+m$ 目元组的集合。
- 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
- 若关系 r 有 k_r 个元组，关系 s 有 k_s 个元组，则关系 r 和 s 的笛卡尔积有 $k_r \times k_s$ 个元组。记作：

$$r \times s = \{ t_r \cdot t_s \mid t_r \in r \wedge t_s \in s \}$$

r

A	B
α	1
β	2

s

C	D	E
α	10	a
β	10	a
β	20	b
γ	10	b

笛卡尔积运算

$r \times s$

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

举例说明

成绩管理数据库ScoreDB的实例数据 (P48)

Class关系

ClassNo	ClassName	instiute	grade	ClassNum
AC0703	会计学07(3)班	会计学院	2007	46
CS0701	计算机07(1)班	信息学院	2007	48
IS0802	信息系统08(2)班	信息学院	2008	43

Student关系

StudentNo	StudentName	sex	birthday	native	nation	classNo
0701001	李小勇	男	1990-12-21	南昌	汉族	CS0701
0701008	王 红	男	1992-04-26	上海	汉族	CS0701
0703010	李宏冰	女	1992-03-09	太原	蒙古族	AC0703
0703045	王 红	男	1992-04-26	北京	汉族	AC0703
0802002	刘方晨	女	1990-11-11	南昌	傣族	IS0802
0802005	王红敏	女	1990-10-01	上海	蒙古族	IS0802

Course关系

CourseNo	CourseName	creditHour	courseHour	priorCours
AC001	基础会计	48	3	null
CN028	大学语文	48	3	nul
CS012	操作系统	80	5	null
CS015	数据库系统	64	4	CS012

Score关系

StudentNo	courseNo	term	score
0701001	CN028	07081	85
0701001	CS012	07082	88
0701001	CS015	08091	92
0701008	AC001	07081	76
0701008	CN028	07081	86
0701008	CS012	07082	93
0701008	CS015	08091	96
0703010	AC001	07081	92
0703010	CN028	07081	83
0703010	CS012	07082	73
0703045	AC001	07081	52
0703045	AC001	08091	94

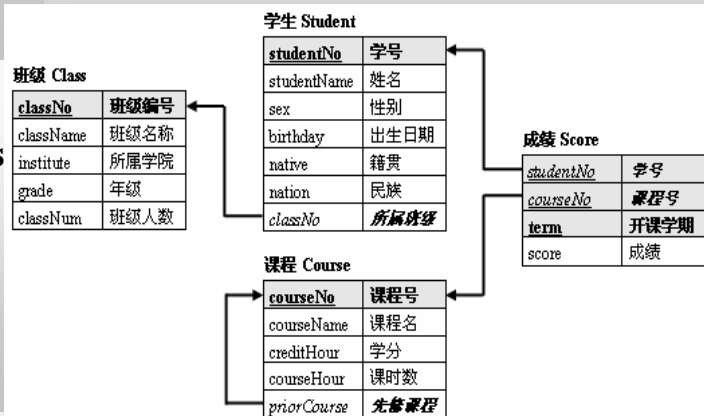


图 2-8 成绩管理数据库 ScoreDB 的模式导航图

关系代数

■ 传统的集合运算 什么用？

- 对于数据库ScoreDB，笛卡尔积Class × Course的结果关系为：

关系 Class × Course

classNo	className	institute	grade	classNum	courseNo	courseName	creditHour	courseHour	priorCourse
AC0703	会计学 07(3)班	会计学院	2007	46	AC001	基础会计	48	3	null
AC0703	会计学 07(3)班	会计学院	2007	46	CN028	大学语文	48	3	null
AC0703	会计学 07(3)班	会计学院	2007	46	CS012	操作系统	80	5	null
AC0703	会计学 07(3)班	会计学院	2007	46	CS015	数据库系统	64	4	CS012
CS0701	计算机 07(1)班	信息学院	2007	48	AC001	基础会计	48	3	null
CS0701	计算机 07(1)班	信息学院	2007	48	CN028	大学语文	48	3	null
CS0701	计算机 07(1)班	信息学院	2007	48	CS012	操作系统	80	5	null
CS0701	计算机 07(1)班	信息学院	2007	48	CS015	数据库系统	64	4	CS012
IS0802	信息系统 08(2)	信息学院	2008	43	AC001	基础会计	48	3	null
IS0802	信息系统 08(2)	信息学院	2008	43	CN028	大学语文	48	3	null
IS0802	信息系统 08(2)	信息学院	2008	43	CS012	操作系统	80	5	null
IS0802	信息系统 08(2)	信息学院	2008	43	CS015	数据库系统	64	4	CS012

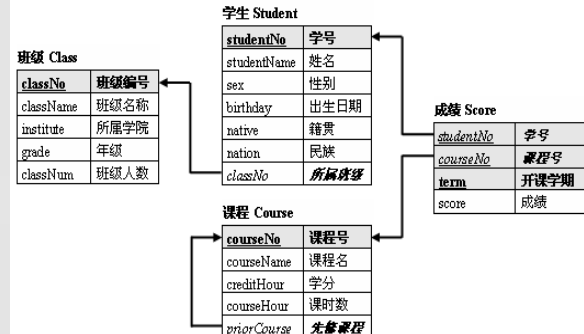


图 2-8 成绩管理数据库 ScoreDB 的模式导航图

RXS

什么用？

R

姓名	课程	成绩
张军	物理	93
王红	数学	86
张军	数学	89

S

姓名	课程	成绩
张军	物理	93
王红	数学	86
张军	数学	89

R.a	R.b	R.c	S.d	S.e	S.f
姓名	课程	成绩	姓名	课程	成绩
张军	物理	93	张军	物理	93
张军	物理	93	王红	数学	86
张军	物理	93	张军	数学	89
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
张军	数学	89	张军	物理	93
张军	数学	89	王红	数学	86
张军	数学	89	张军	数学	89

关系代数

■ 专门的关系运算

● 选择

选择操作是在关系 r 中查找满足给定谓词(即选择条件)的所有元组，记作：

$$\sigma_P(r) = \{ t \mid t \in r \wedge P(t) \}$$

P 表示谓词(即选择条件)，它是一个逻辑表达式($X \text{ op } Y$ ，
可以产生复杂的运算关系表达式，逻辑表达式等)，取值
为“真”或“假”。

关系代数

专门的关系运算

选择

◆ 例如，在数据库S

$\sigma_{\text{grade}=2007}(\text{Class})$

ClassNo	ClassName
AC0703	会计学08(3)班
CS0701	计算机07(1)班

Class关系

ClassNo	ClassName	instiute	grade	ClassNum
AC0703	会计学07(3)班	会计学院	2007	46
CS0701	计算机07(1)班	信息学院	2007	48
IS0802	信息系统08(2)班	信息学院	2008	43

Student关系

StudentNo	StudentName	sex	birthday	native	nation	classNo
0701001	李小勇	男	1990-12-21	南昌	汉族	CS0701
0701008	王 红	男	1992-04-26	上海	汉族	CS0701
0703010	李宏冰	女	1992-03-09	太原	蒙古族	AC0703
0703045	王 红	男	1992-04-26	北京	汉族	AC0703
0802002	刘方晨	女	1990-11-11	南昌	傣族	IS0802
0802005	王红敏	女	1990-10-01	上海	蒙古族	IS0802

◆ 例如，在数据库ScoreDB中，查找所有太原出生的女学生情况？

$\sigma_{\text{native}='太原' \wedge \text{sex}='女'}(\text{Student})$

StudentNo	StudentName	sex	birthday	native	nation	classNo
0703010	李宏冰	女	1992-03-09	太原	蒙古族	AC0703

关系代数

■ 专门的关系运算

● 投影

关系是一个二维表，对它的操作可以从水平(行)的角度进行，即选择操作；也可以从纵向(列)的角度进行，即投影操作。

关系 r 上的投影是从 r 中选择出若干属性列组成新的关系。
记作：

$$\Pi_A(r) = \{ t[A] \mid t \in r \}$$

A 为关系 r 的属性集合。

Student关系

关系代数

■ 专门的关系运算

● 投影

- ◆ 例如，在数据库ScoreDB中，查找所有学生的姓名和民族

$\Pi_{\text{studentName}, \text{nation}}(\text{Student})$

- ◆ 例如，在数据库ScoreDB中，查找所有“蒙古族”学生的姓名和籍贯

$\Pi_{\text{studentName}, \text{native}}(\sigma_{\text{nation}=\text{'蒙古族'}}(\text{Student}))$

StudentNo	StudentName	sex	birthday	native	nation	classNo
0701001	李小勇	男	1990-12-21	南昌	汉族	CS0701
0701008	王红	男	1992-04-26	上海	汉族	CS0701
0703010	李宏冰	女	1992-03-09	太原	蒙古族	AC0703
0703045	王红	男	1992-04-26	北京	汉族	AC0703
0802002	刘方晨	女	1990-11-11	南昌	傣族	IS0802
0802005	王红敏	女	1990-10-01	上海	蒙古族	IS0802

StudentName	nation
李小勇	汉族
王红	汉族
王宏冰	蒙古族
刘方晨	傣族
王红敏	蒙古族

StudentName	native
王宏冰	太原
王红敏	上海

关系代数

■ 专门的关系运算

● 连接

◆ 连接也称为 θ 连接。记为 $A \text{ op } B$ ，其中 A 、 B 分别为关系 r 和 s 中的度数相等且可比的连接属性集， op 为比较运算符。 θ 连接是从两个关系的笛卡尔积中选取连接属性间满足谓词 θ 的所有元组。记作：

$$r \bowtie_\theta s = \{ t_r \cdot t_s \mid t_r \in r \wedge t_s \in s \wedge (r.A \text{ op } s.B) \}$$

◆ θ 连接运算就是从关系 r 和 s 的笛卡尔积 $r \times s$ 中，选取 r 关系在 A 属性集上的值与 s 关系在 B 属性集上的值满足连接谓词 θ 的所有元组，即

$$r \bowtie_\theta s = \sigma_\theta(r \times s)$$

关系代数

■ 专门的关系运算

● 连接

- ◆ 连接运算中有两种最常用、最重要的连接，一种是**等值连接(equijoin)**，另一种是**自然连接(natural join)**。 θ 为等值比较谓词的连接运算称为**等值连接**。
- ◆ **自然连接**是一种**特殊的等值连接**，它要求两个参与连接的关系具有**公共的属性集**，并在这个公共属性集上进行**等值连接**；同时，还要求将连接结果中的**重复属性列**去掉，即在公共属性集中的列**只保留一次**。

笛卡尔积运算

r

<i>A</i>	<i>B</i>
α	1
β	2

s

<i>C</i>	<i>D</i>	<i>E</i>
α	10	<i>a</i>
β	10	<i>a</i>
β	20	<i>b</i>
γ	10	<i>b</i>

r X *s*

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
α	1	α	10	<i>a</i>
α	1	β	10	<i>a</i>
α	1	β	20	<i>b</i>
α	1	γ	10	<i>b</i>
β	2	α	10	<i>a</i>
β	2	β	10	<i>a</i>
β	2	β	20	<i>b</i>
β	2	γ	10	<i>b</i>

广义笛卡尔积运算

■ $\sigma_{A=C}(r \times s)$

■ $r \times s$

■ $\sigma_{A=C}(r \times s)$

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	10	a
β	2	β	20	b

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

Student关系

StudentNo	StudentName	sex	birthday	native	nation	classNo
0701001	李小勇	男	1990-12-21	南昌	汉族	CS0701
0701008	王红	男	1992-04-26	上海	汉族	CS0701
0703010	李宏冰	女	1992-03-09	太原	蒙古族	AC0703
0703045	王红	男	1992-04-26	北京	汉族	AC0703
0802002	刘方晨	女	1990-11-11	南昌	傣族	IS0802
0802005	王红敏	女	1990-10-01	上海	蒙古族	IS0802

Class关系

ClassNo	ClassName	instiute	grade	ClassNum
AC0703	会计学07(3)班	会计学院	2007	46
CS0701	计算机07(1)班	信息学院	2007	48
IS0802	信息系统08(2)班	信息学院	2008	43

- 例如，在数据库ScoreDB中，查找所有2008级的“蒙古族”学生的姓名

StudentNo	StudentName	sex	birthday	native	nation	classNo	ClassNo	ClassName	instiute	grade	ClassNum
0701001	李小勇	男	1990-12-21	南昌	汉族	CS0701	CS0701	计算机07(1)班	信息学院	2007	48
0701008	王红	男	1992-04-26	上海	汉族	CS0701	CS0701	计算机07(1)班	信息学院	2007	48
0703010	李宏冰	女	1992-03-09	太原	蒙古族	AC0703	AC0703	会计学07(3)班	会计学院	2007	46
0703045	王红	男	1992-04-26	北京	汉族	AC0703	AC0703	会计学07(3)班	会计学院	2007	46
0802002	刘方晨	女	1990-11-11	南昌	傣族	IS0802	IS0802	信息系统08(2)班	信息学院	2008	43
0802005	王红敏	女	1990-10-01	上海	蒙古族	IS0802	IS0802	信息系统08(2)班	信息学院	2008	43

- ② 根据模式导航图可知，关系Student与关系Class可通过外码classNo关联起来，这种外码引用关系可通过自然连接表示

$$\text{Student} \bowtie \text{Class} = \sigma_{\text{Student.classNo}=\text{Class.classNo}}(\text{Student} \times \text{Class})$$

- ③ 最后的查询可表达为：

$$\Pi_{\text{studentName}}(\sigma_{\text{nation}='蒙古族'}(\text{Student}) \bowtie \sigma_{\text{grade}=2008}(\text{Class}))$$

StudentName

王红敏

Student关系

StudentNo	StudentName	sex	birthday	native	nation	classNo
0701001	李小勇	男	1990-12-21	南昌	汉族	CS0701
0701008	王红	男	1992-04-26	上海	汉族	CS0701
0703010	李宏冰	女	1992-03-09	太原	蒙古族	AC0703
0703045	王红	男	1992-04-26	北京	汉族	AC0703
0802002	刘方晨	女	1990-11-11	南昌	傣族	IS0802
0802005	王红敏	女	1990-10-01	上海	蒙古族	IS0802

Class关系

ClassNo	ClassName	instiute	grade	ClassNum
AC0703	会计学07(3)班	会计学院	2007	46
CS0701	计算机07(1)班	信息学院	2007	48
IS0802	信息系统08(2)班	信息学院	2008	43

$$\Pi_{\text{studentName}}(\sigma_{\text{nation}='蒙古族'}(\text{Student}) \bowtie \sigma_{\text{grade}=2008}(\text{Class}))$$

$$= \Pi_{\text{studentName}}(\sigma_{\text{Student.classNo}=\text{Class.classNo}}(\sigma_{\text{nation}='蒙古族'}(\text{Student}) \times \sigma_{\text{grade}=2008}(\text{Class})))$$

$$= \Pi_{\text{studentName}}(\sigma_{\text{Student.classNo}=\text{Class.classNo}}(\sigma_{\text{nation}='蒙古族' \wedge \text{grade}=2008}(\text{Student} \times \text{Class})))$$

$$= \Pi_{\text{studentName}}(\sigma_{\text{nation}='蒙古族' \wedge \text{grade}=2008 \wedge \text{Student.classNo}=\text{Class.classNo}}(\text{Student} \times \text{Class}))$$

$$= \Pi_{\text{studentName}}(\sigma_{\text{nation}='蒙古族' \wedge \text{grade}=2008}(\sigma_{\text{Student.classNo}=\text{Class.classNo}}(\text{Student} \times \text{Class})))$$

$$= \Pi_{\text{studentName}}(\sigma_{\text{nation}='蒙古族' \wedge \text{grade}=2008}(\text{Student} \bowtie \text{Class}))$$

更名运算

■ 定义

- 给一个关系表达式赋予名字

$$\rho_x (E)$$

返回表达式 E 的结果，并把名字 x 赋给 E

$$\rho_x (A_1, A_2, \dots, A_n) (E)$$

返回表达式 E 的结果，并把名字 x 赋给 E ，同时将各属性更名为 A_1, A_2, \dots, A_n

- 关系被看作一个最小的关系代数表达式，可以将更名运算施加到关系上，得到具有不同名字的同—关系。这在同一关系多次参与同一运算时很有帮助

更名运算

■ 示例

- 求数学成绩比王红同学高的学生

姓名	课程	成绩	姓名	课程	成绩
张军	物理	93	张军	物理	93
张军	物理	93	王红	数学	86
张军	物理	93	张军	数学	89
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
张军	数学	89	张军	物理	93
张军	数学	89	王红	数学	86
张军	数学	89	张军	数学	89

$$(R \times \rho_S(R))$$

R

姓名	课程	成绩
张军	物理	93
王红	数学	86
张军	数学	89

S

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

关系代数

- 例如，在数据库ScoreDB中，查找课程号为“AC001”课程的考试中比学号为“0703045”的学生考得更好的所有学生的姓名和成绩。

分析：①

成绩管理数据库ScoreDB的实例数据

成绩元组

Class关系

②

ClassNo	ClassName	instiute	grade	ClassNum
AC0703	会计学07(3)班	会计学院	2007	46
CS0701	计算机07(1)班	信息学院	2007	48
IS0802	信息系统08(2)班	信息学院	2008	43

③

Student关系

④

StudentNo	StudentName	sex	birthday	native	nation	classNo
0701001	李小勇	男	1990-12-21	南昌	汉族	CS0701
0701008	王红	男	1992-04-26	上海	汉族	CS0701
0703010	李宏冰	女	1992-03-09	太原	蒙古族	AC0703
0703045	王红	男	1992-04-26	北京	汉族	AC0703
0802002	刘方晨	女	1990-11-11	南昌	傣族	IS0802
0802005	王红敏	女	1990-10-01	上海	蒙古族	IS0802

Course关系

=

=

CourseNo	CourseName	creditHour	courseHour	priorCourse
AC001	基础会计	48	3	null
CN028	大学语文	48	3	nul
CS012	操作系统	80	5	null
CS015	数据库系统	64	4	CS012

Score关系

StudentNo	courseNo	term	score
0701001	CN028	07081	85
0701001	CS012	07082	88
0701001	CS015	08091	92
0701008	AC001	07081	76
0701008	CN028	07081	86
0701008	CS012	07082	93
0701008	CS015	08091	96
0703010	AC001	07081	92
0703010	CN028	07081	83
0703010	CS012	07082	73
0703045	AC001	07081	52
0703045	AC001	08091	94
0703045	CN028	07081	80
0703045	CS015	08091	82
0802002	AC001	08091	98
0802002	CN028	08091	72
0802005	CS015	09101	85
0802005	AC001	09101	88
0802005	CS012	08092	90
0802005	CS015	09101	87

关系)) AS r1

关系

实例数据 (P58)

Score关系			
StudentNo	courseNo	term	score
0701001	CN028	07081	85
0701001	CS012	07082	88
0701001	CS015	08091	92
0701008	AC001	07081	76
0701008	CN028	07081	86
0701008	CS012	07082	93
0701008	CS015	08091	96
0703010	AC001	07081	92
0703010	CN028	07081	83
0703010	CS012	07082	73
0703045	AC001	07081	52
0703045	AC001	08091	94
0703045	CN028	07081	80
0703045	CS015	08091	82
0802002	AC001	08091	98
0802002	CN028	08091	72
0802005	CS015	09101	85
0802005	AC001	09101	88
0802005	CS012	08092	90
0802005	CS015	09101	87

CS012	操作系统	80	5	null
CS015	数据库系统	64	4	CS012

关系代数

- 例如，在数据库ScoreDB中，查找课程号为“AC001”课程的考试
中比学号为“0703045”的学生考得更好的所有学生的姓名和成绩

分析：①找出学号为“0703045”的学生在课程号为“AC001”的课程中的成绩元组
(结果关系记为 r_1)，可表达为： $(\sigma_{\text{studentNo}='0703045' \wedge \text{courseNo}='AC001'}(\text{Score})) \text{ AS } r_1$

②找出选修了课程号为“AC001”课程的所有学生的成绩元组(结果关系
记为 r_2)，其查询可表达为： $(\sigma_{\text{courseNo}='AC001'}(\text{Score})) \text{ AS } r_2$

③将关系 r_1 与关系 r_2 进行 θ 连接(结果关系记为 r_3)，其查询可表达为：
 $r_1 \bowtie_{r_1.\text{score} < r_2.\text{score}} r_2 = \sigma_{r_1.\text{score} < r_2.\text{score}} (r_1 \times r_2)$

④将关系 r_3 与学生关系Student按外码 studentNo 进行自然连接，并对连接
结果在属性 studentName 和 $r_2.\text{score}$ 上进行投影，其查询可表达为：

$\pi_{\text{studentName}, r_2.\text{score}} (r_3 \bowtie \text{Student})$
 $= \pi_{\text{studentName}, r_2.\text{score}} (\sigma_{r_2.\text{studentNo}=\text{Student.studentNo}} (r_3 \times \text{Student}))$
 $= \pi_{\text{studentName}, r_2.\text{score}} (\sigma_{r_2.\text{studentNo}=\text{Student.studentNo}} (\sigma_{r_1.\text{score} < r_2.\text{score}} (r_1 \times r_2)) \times \text{Student}))$

studentNo courseNo term score

0703045 AC001 07081 52

0703045 AC001 08091 94

(a) 关系 r_1

r1.studentNo r1.courseNo r1.term r1.score r2.studentNo r2.courseNo r2.term r2.score

0703045 AC001 07081 52

0703045 AC001 07081 52

0703045 AC001 07081 52

0703045 AC001 07081 52

0703045 AC001 07081 52

0703045 AC001 08091 94

(c) 关系 r_3

关系代数

studentNo courseNo term

0701008 AC001 07081

0703010 AC001 07081

0703045 AC001 07081

0703045 AC001 08091

0802002 AC001 08091

0802005 AC001 09101

(b) 关系 r_2

Score关系

StudentNo courseNo term score

0701001 CN028 07081 85

0701001 CS012 07082 88

0701001 CS015 08091 92

0701008 AC001 07081 76

0701008 CN028 07081 86

0701008 CS012 07082 93

0701008 CS015 08091 96

0703010 AC001 07081 92

0703010 CN028 07081 83

0703010 CS012 07082 73

0703045 AC001 07081 52

0703045 AC001 08091 94

0703045 CN028 07081 80

0703045 CS015 08091 82

0802002 AC001 08091 98

0802002 CN028 08091 72

0802002 CS015 09101 85

0802005 AC001 09101 88

0802005 CS012 08092 90

0802005 CS015 09101 87

图2-19 例2.16的 θ 连接的计算过程

外连接(outer join)

- 例：列出老师的有关信息，包括姓名、工资、所教授的课程

$\Pi_{P\#,PN,SAL,C\#,CN}((PROF) \bowtie PC \bowtie C)$

DEPT(D#, DN, DEAN)

S(S#, SN, SEX, AGE, D#)

C(C#, CN, PC#, CREDIT)

SC(S#, C#, SCORE)

PROF(P#, PN, D#, SAL)

PC(P#, C#)

C#	P#
C01	P01
C02	P02
C02	P04



C#	CN
C01	物理
C02	数学
C03	化学

P#	PN	SAL	C#	CN
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，C03？

外连接

■ 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 + 失配的元组

- 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 + 左侧表中失配的元组
 - ⋈⌋ 右外连接 = 自然连接 + 右侧表中失配的元组
 - ⌋⋈⌋ 全外连接 = 自然连接 + 两侧表中失配的元组

外连接

P#	PN	SAL
P01	赵明	800
P02	钱广	700
P03	孙立	600
P04	李三	500



C#	P#
C01	P01
C02	P02
C02	P04



C#	CN
C01	物理
C02	数学
C03	化学

==

P#	PN	SAL	C#	CN
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

所有老师的信息

外连接

P#	PN	SAL
P01	赵明	800
P02	钱广	700
P03	孙立	600
P04	李三	500



C#	P#
C01	P01
C02	P02
C02	P04



C#	CN
C01	物理
C02	数学
C03	化学

==

P#	PN	SAL	C#	CN
P01	赵明	800	C01	物理
P02	钱广	700	C02	数学
P04	李三	500	C02	数学
null	null	null	C03	化学

所有课程的信息

外连接

P#	PN	SAL
P01	赵明	800
P02	钱广	700
P03	孙立	600
P04	李三	500



C#	P#
C01	P01
C02	P02
C02	P04



C#	CN
C01	物理
C02	数学
C03	化学

==

P#	PN	SAL	C#	CN
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
null	null	null	C03	化学

所有老师和课程的信息

小结

- θ 连接?
- 什么是外连接?

关系代数查询实例

- 求未选修01号课程的学生号

方案1: $\Pi_{S\#}(S) - \Pi_{S\#}(\sigma_{C\# = '01'}(SC))$

方案2: $\Pi_{S\#}(\sigma_{C\# \neq '01'}(SC))$

哪一个正确?

DEPT(D# , DN , DEAN)

S(S# , SN , SEX , AGE , D#)

C(C# , CN , PC# , CREDIT)

SC(S# , C# , SCORE)

PROF(P# , PN , D# , SAL)

PC(P# , C#)

S#	SN	AGE	S#	C#	G
s1	s1	01	90
s2	s2	01	95
s3	s1	02	96

求仅选修了01号课程的学生号

关系代数查询实例

- 求仅选修了01号课程的学生号

选修01号课程的学生 - 仅选01号课程之外的学生

$$= \Pi_{S\#}(\sigma_{C\# = '01'}(SC)) - \Pi_{S\#}(SC - \sigma_{C\# = '01'}(SC))$$

DEPT(D# , DN , DEAN)

S(S# , SN , SEX , AGE , D#)

C(C# , CN , PC# , CREDIT)

SC(S# , C# , SCORE)

PROF(P# , PN, D# , SAL)

PC(P# , C#)

	SCORE
	96
	90

=

S#	C#	SCORE
003	02	88
001	03	92

=

S#
002

关系代数

■ 专门的关系运算

● 除运算

◆例如，需要查找修读过信息学院开设的所有课程的学生学号，如何表达查询？

分析：①查找出修读过信息学院课程的所有学生

$$r1 = \pi_{\text{studentNo}, \text{courseNo}}(\sigma_{\text{courseNo LIKE 'CS\%'}}(\text{Score}))$$

②找出信息学院开设的所有课程

$$r2 = \pi_{\text{courseNo}}(\sigma_{\text{courseNo LIKE 'CS\%'}}(\text{Course}))$$

StudentNo courseNo

0701001 CS012
0701001 CS015
0701008 CS012
0701008 CS015
0703010 CS012
0703045 CS015
0802002 CS015
0802005 CS012
0802005 CS015

CourseNo

CS012
CS015

Course换成
Score? 可以吗?

Class关系				
ClassNo	ClassName	institute	grade	ClassNum
AC0703	会计学07(3)班	会计学院	2007	46
CS0701	计算机07(1)班	信息学院	2007	48
IS0802	信息系统08(2)班	信息学院	2008	43

Student关系				
StudentNo	StudentName	sex	birthday	native nation classNo
0701001	李小勇	男	1990-12-21	南昌 汉族 CS0701
0701008	王 红	男	1992-04-26	上海 汉族 CS0701
0703010	李宏冰	女	1992-03-09	太原 蒙古族 AC0703
0703045	王 红	男	1992-04-26	北京 汉族 AC0703
0802002	刘方晨	女	1990-11-11	南昌 傣族 IS0802
0802005	王红敏	女	1990-10-01	上海 蒙古族 IS0802

Course关系				
CourseNo	CourseName	creditHour	courseHour	priorCourse
AC001	基础会计	48	3	null
CN028	大学语文	48	3	null
CS012	操作系统	80	5	null
CS015	数据库系统	64	4	CS012

除运算

■ 象集 (Image Set)

- 关系 $R(X, Y)$, X, Y 是属性组, x 是 X 上的取值, 定义 x 在 R 中的象集为

$$Y_x = \{ t[Y] \mid t \in R \wedge t[X] = x \}$$

从 R 中选出在 X 上取值为 x 的元组, 去掉 X 上的分量, 只留 Y 上的分量

X	Y
姓名	课程
张军	物理
王红	数学
张军	数学

$x = \text{张军}$

Y_x
课程
数学
物理


张军同学所选修的全部课程

关系代数

■ 专门的关系运算

◆ 设关系 $r(R)$ 和 $s(S)$ ，属性集 S 是 R 的子集，即 $S \subseteq R$ ，则关系 $r \div s$ 是关系 r 中满足下列条件的元组在属性集 $R-S$ 上的投影： $\forall t_r \in r$ ，记 $x = t_r[R-S]$ ，则关系 r 中属性集 $R-S$ 的取值 x 的象集 S_x 包含关系 s 。记作

$$r \div s = \{ t_r[R-S] \mid t_r \in r \wedge s \subseteq S_x \}$$

StudentNo	courseNo		CourseNo
0701001	CS012		CS012
0701001	CS015		CS015
0701008	CS012		
0701008	CS015		
0703010	CS012		
0703045	CS015		
0802002	CS015		
0802005	CS012		
0802005	CS015		

关系代数

■ 专门的关系运算

- ③ 比较图2-20(a)和(b)，修读过信息学院开设的所有课程的学生就是关系 r_1 中满足“**courseNo**列包含关系 r_2 的所有行”的那些学生。

studentNo	courseNo
0701001	CS012
0701001	CS015
0701008	CS012
0701008	CS015
0703010	CS012
0703045	CS015
0802002	CS015
0802005	CS012
0802005	CS015

(a) 关系 r_1

象集 courseNo '0701001'

象集 courseNo '0701008'

象集 courseNo '0802005'

courseNo
CS012
CS015

(b) 关系 r_2

studentNo
0701001
0701008
0802005

(c) 关系 $r_1 \div r_2$

studentNo	courseNo
0701001	CN028
0701001	CS012
0701001	CS015
0701008	AC001
0701008	CN028
0701008	CS012
0701008	CS015
0703010	AC001
0703010	CN028
0703010	CS012
0703045	AC001
0703045	CN028
0703045	CS015
0802002	AC001
0802002	CN028
0802002	CS015
0802005	AC001
0802005	CS012
0802005	CS015

(a) $\Pi_{\text{studentNo, courseNo}}(\text{Score})$ 的结果

关系代数

关系数据库ScoreDB 中

$(\text{Score})) \div$

$\text{courseName} = \text{'基础会计'} \vee \text{courseName} = \text{'操作系统'} (\text{Course}))$

“基础会计”又选修了“操作系统”课
查询结果如图2-21(c)所示。其中，

$\text{courseNo}(\text{Score})$ 和除数

$\text{courseName} = \text{'基础会计'} \vee \text{courseName} = \text{'操作系统'} (\text{Course}))$ 的

图2-21(a)和(b)所示。

studentNo
0701008
0703010
0802005

(c) 查询(1)的结果

courseNo
AC001
CS012

(b) $\Pi_{\text{courseNo}}(\sigma_{\text{courseName} = \text{'基础会计'} \vee \text{courseName} = \text{'操作系统'}}(\text{Course}))$ 的结果

studentNo	courseNo
0701001	CN028
0701001	CS012
0701001	CS015
0701008	AC001
0701008	CN028
0701008	CS012
0701008	CS015
0703010	AC001
0703010	CN028
0703010	CS012
0703045	AC001
0703045	CN028
0703045	CS015
0802002	AC001
0802002	CN028
0802002	CS015
0802005	AC001
0802005	CS012
0802005	CS015

关系代数

关系数据库 **Score**

studentNo	studentName
0701008	王红

(e) 查询(2)的结果

ne $((\Pi_{\text{studentNo, courseNo}}(\text{Score})) \div$
 $(\Pi_{\text{courseNo}}(\text{Course}))) \bowtie \text{Student})$

所有课程的学生学号和姓名，查询结果
 其中除数 $\Pi_{\text{courseNo}}(\text{Course})$ 的查询结果如

courseNo
AC001
CN028
CS012
CS015

(d) $\Pi_{\text{courseNo}}(\text{Course})$ 的结果

(a) $\Pi_{\text{studentNo, courseNo}}(\text{Score})$ 的结果

关系代数

关系数据库 **Scor**

courseNo	courseName
CN028	大学语文
CS015	数据库系统

(g) 查询(3)的结果

$((\Pi_{\text{courseNo}, \text{studentNo}}(\text{Score})) \div$

$(\Pi_{\text{studentNo}}(\sigma_{\text{sex}='男'}(\text{Student})))) \bowtie \text{Course})$

男同学选修过的课程的课程号和课程名，
1(g)所示。其中除数

$\Pi_{\text{studentNo}}(\sigma_{\text{sex}='男'}(\text{Student}))$

2-21(f)所示。

studentNo	courseNo
0701008	AC001
0703010	AC001
0703045	AC001
0802002	AC001
0802005	AC001
0701001	CN028
0701008	CN028
0703010	CN028
0703045	CN028
0802002	CN028
0701001	CS012
0701008	CS012
0703010	CS012
0802005	CS012
0701001	CS015
0701008	CS015
0703045	CS015
0802002	CS015
0802005	CS015

(a) $\Pi_{\text{studentNo}, \text{courseNo}}(\text{Score})$ 的结果

studentNo
0701001
0701008
0703045

(f) $\Pi_{\text{studentNo}}(\sigma_{\text{sex}='男'}(\text{Student}))$ 的结果

studentNo	courseNo
0701001	CN028
0701001	CS012
0701001	CS015
0701008	AC001
0701008	CN028
0701008	CS012
0701008	CS015
0703010	AC001
0703010	CN028
0703010	CS012
0703045	AC001
0703045	CN028
0703045	CS015
0802002	AC001
0802002	CN028
0802002	CS015
0802005	AC001
0802005	CS012
0802005	CS015

关系代数

关系数据库 Score

studentNo	studentName
0802002	刘方晨

(h) 查询(4)的结果

$$\text{select } ((\prod_{\text{studentNo, courseNo}}(\text{Score})) \div (\prod_{\text{studentNo}}(\sigma_{\text{sex}='男'}(\text{Student})))) \bowtie (\sigma_{\text{sex}='女'}(\text{Student}))$$

courseNo
CN028
CS015

$(\prod_{\text{courseNo, studentNo}}(\text{Score})) \div (\prod_{\text{studentNo}}(\sigma_{\text{sex}='男'}(\text{Student})))$ 的结果 (根据第(3)小题的结果)

studentNo	
0701001	(男)
0701008	(男)
0703045	(男)
0802002	(女)

最后除法的结果

(a) $\prod_{\text{studentNo, courseNo}}(\text{Score})$ 的结果

除运算

■ 示例（CS,C,S表名）

● 求同时选修了01和02号课和
方案1:

$$\Pi_{S\#, C\#}(SC) \div \sigma_{C\# = '01' \vee C\# = '02'}(C)$$

方案2:

$$\Pi_{S\#}(SC \div \sigma_{C\# = '01' \vee C\# = '02'}(C))$$

哪一个正确？

■ 专门的关系运算

◆ 设关系 $r(R)$ 和 $s(S)$ ，属性集 S 是 R 的子集，即 $S \subseteq R$ ，则关系 $r \div s$ 是关系 r 中满足下列条件的元组在属性集 $R-S$ 上的投影： $\forall t_r \in r$ ，记 $x = t_r[R-S]$ ，则关系 r 中属性集 $R-S$ 的取值 x 的象集 S_x 包含关系 s 。记作

$$r \div s = \{ t_r[R-S] \mid t_r \in r \wedge s \subseteq S_x \}$$

StudentNo	courseNo	CourseNo
0701001	CS012	CS012
0701001	CS015	CS015
0701008	CS012	CS012
0701008	CS015	CS015
0703010	CS012	CS012
0703045	CS015	CS015
0802002	CS015	CS015
0802005	CS012	CS012
0802005	CS015	CS015

■ 示例

● 求同时选修了01和02号课程的学生号

方案1:

$\Pi_{S\#, C\#}(SC) \div \sigma_{C\# = '01' \vee C\# = '02'}(C)$

方案2:

$\Pi_{S\#}(SC \div \sigma_{C\# = '01' \vee C\# = '02'}(C))$

哪一个正确?

除运算

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

选修了全部课程并且成绩都相同的学生

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

关系代数查询综合举例

- 给定一个查询需求，构造其关系代数表达式的步骤
 - 明确该查询涉及到哪些属性；
 - 明确该查询涉及到哪些关系；
 - 根据数据库模式导航图，通过多对一联系(或一对多联系)把所有涉及的关系连接起来，每一个多对一联系(或一对多联系)都可以表示为外码属性的自然连接。

关系代数查询综合举例

- ◆ 例如，ScoreDB数据库中，查找“蒙古族”学生所修各门课程的情况，要求输出学生姓名、课程名和成绩。

分析：

- 该查询共涉及4个属性，分别是民族nation、姓名studentName、课程名courseName和成绩score，其中，nation属性用于选择条件nation='蒙古族'。

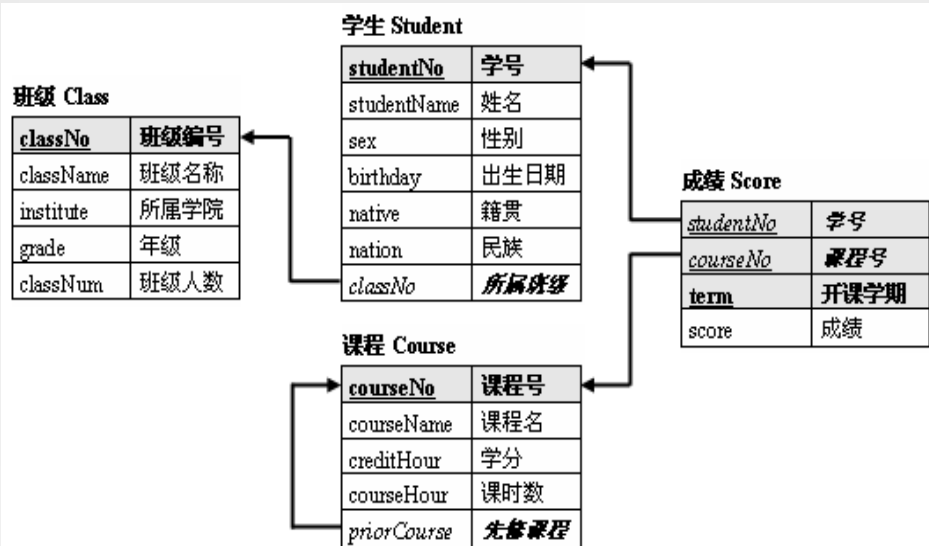


图 2-8 成绩管理数据库 ScoreDB 的模式导航图

成绩管理数据库 ScoreDB 的实例数据和成绩

ClassNo	ClassName	institute	grade	ClassNum
AC0703	会计学07(3)班	会计学院	2007	46
CS0701	计算机07(1)班	信息学院	2007	48
IS0802	信息系统08(2)班	信息学院	2008	43

StudentNo	StudentName	sex	birthday	native	nation	classNo
0701001	李小勇	男	1990-12-21	南昌	汉族	CS0701
0701008	王 红	男	1992-04-26	上海	汉族	CS0701
0703010	李宏冰	女	1992-03-09	太原	蒙古族	AC0703
0703045	王 红	男	1992-04-26	北京	汉族	AC0703
0802002	刘方晨	女	1990-11-11	南昌	傣族	IS0802
0802005	王红敏	女	1990-10-01	上海	蒙古族	IS0802

CourseNo	CourseName	creditHour	courseHour	priorCourse
AC001	基础会计	48	3	null
CN028	大学语文	48	3	null
CS012	操作系统	80	5	null
CS015	数据库系统	64	4	CS012

StudentNo	courseNo	term	score
0701001	CN028	07081	85
0701001	CS012	07082	88
0701001	CS015	08091	92
0701008	AC001	07081	76
0701008	CN028	07081	86
0701008	CS012	07082	93
0701008	CS015	08091	96
0703010	AC001	07081	92
0703010	CN028	07081	83
0703010	CS012	07082	73
0703045	AC001	07081	52
0703045	AC001	08091	94
0703045	CN028	07081	80
0703045	CS015	08091	82
0802002	AC001	08091	98
0802002	CN028	08091	72
0802002	CS015	09101	85
0802005	AC001	09101	88
0802005	CS012	08092	90
0802005	CS015	09101	87

关系代数查询综合举例

- ◆ 例如，ScoreDB数据库中，查找2007级的“南昌”籍同学修读了哪些课程，要求输出学生姓名、课程名。

分析：

- 该查询共涉及4个属性，分别是年级grade、籍贯native、姓名studentName和课程名courseName，其中年级grade和籍贯native用于选择条件。
- 共涉及3个关系，分别是班级关系Class、学生关系Student和课程关系Course。
- 学生关系Student与班级关系Class

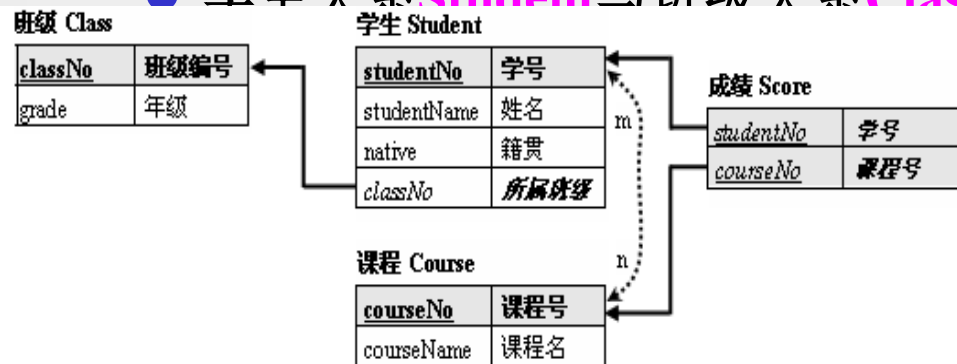


图 2-22 例 2.20 的模式导航图

Class关系

ClassNo	ClassName	institute	grade	ClassNum
AC0703	会计学07(3)班	会计学院	2007	46
CS0701	计算机07(1)班	信息学院	2007	48
IS0802	信息系统08(2)班	信息学院	2008	43

Student关系

StudentNo	StudentName	sex	birthday	native	nation	classNo
0701001	李小勇	男	1990-12-21	南昌	汉族	CS0701
0701008	王红	男	1992-04-26	上海	汉族	CS0701
0703010	李宏冰	女	1992-03-09	太原	蒙古族	AC0703
0703045	王红	男	1992-04-26	北京	汉族	AC0703
0802002	刘方晨	女	1990-11-11	南昌	傣族	IS0802
0802005	王红敏	女	1990-10-01	上海	蒙古族	IS0802

Course关系

CourseNo	CourseName	creditHour	courseHour	priorCourse
AC001	基础会计	48	3	null
CN028	大学语文	48	3	null
CS012	操作系统	80	5	null
CS015	数据库系统	64	4	CS012

Score关系

StudentNo	courseNo	term	score
0701001	CN028	07081	85
0701001	CS012	07082	88
0701001	CS015	08091	92
0701008	AC001	07081	76
0701008	CN028	07081	86
0701008	CS012	07082	93
0701008	CS015	08091	96
0703010	AC001	07081	92
0703010	CN028	07081	83
0703010	CS012	07082	73
0703045	AC001	07081	52
0703045	AC001	07081	94
0703045	CN028	07081	80
0703045	CS015	08091	82
0802002	AC001	08091	98
0802002	CN028	08091	72
0802002	CS015	09101	85
0802005	AC001	09101	88
0802005	CS012	08092	90
0802005	CS015	09101	87

综合运算-小结

- 给定一个查询需求，构造其关系代数表达式的步骤
 - 明确该查询涉及到哪些属性；
 - 明确该查询涉及到哪些关系；
 - 根据数据库模式导航图，通过多对一联系(或一对多联系)把所有涉及的关系连接起来，每一个多对一联系(或一对多联系)都可以表示为外码属性的自然连接。

扩展的关系代数运算

定义

- 求一组值的统计信息，返回单一值
- 使用聚集的集合可以是多重集，即一个值可多次。如果想去除重复值，可以用连接符‘distinct’附加在聚集函数名后，如sum-distinct
- sum: 求和
求数量

sum_{ClassNum}(Class)

求0802005号学生的总成绩

sum_{score}(σ_{Studentno = '0802005'}(Score))

Score关系				
StudentNo	courseNo	term	score	
0701001	CN028	07081	85	
0701001	CS012	07082	88	
0701001	CS015	08091	92	
0701008	AC001	07081	76	
0701008	CN028	07081	86	
0701008	CS012	07082	93	
0701008	CS015	08091	96	
0703010	AC001	07081	92	
0703010	CN028	07081	83	
0703010	CS012	07082	73	
0703045	AC001	07081	52	
0703045	AC001	08091	94	
0703045	CN028	07081	80	
0703045	CS015	08091	82	
0802002	AC001	08091	98	
0802002	CN028	08091	72	
0802002	CS015	09101	85	
0802005	AC001	09101	88	
0802005	CS012	08092	90	
0802005	CS015	09101	87	

Class关系

ClassNo	ClassName	instiute	grade	ClassNum
AC0703	会计学07(3)班	会计学院	2007	46
CS0701	计算机07(1)班	信息学院	2007	48
IS0802	信息系统08(2)班	信息学院	2008	43

聚集函数

Score关系

StudentNo	courseNo	term	score
0701001	CN028	07081	85
0701001	CS012	07082	88
0701001	CS015	08091	92
0701008	AC001	07081	76
0701008	CN028	07081	86
0701008	CS012	07082	93
0701008	CS015	08091	96
0703010	AC001	07081	92
0703010	CN028	07081	83
0703010	CS012	07082	73
0703045	AC001	07081	52
0703045	AC001	08091	94
0703045	CN028	07081	80
0703045	CS015	08091	82
0802002	AC001	08091	98
0802002	CN028	08091	72
0802002	CS015	09101	85
0802005	AC001	09101	88
0802005	CS012	08092	90
0802005	CS015	09101	87

- avg: 求平均

求0701001号同学选修课程的平均成绩。

$\text{avg}_{\text{score}}(\sigma_{\text{StudentNo\#} = '0701001'}(\text{Score}))$

- count: 计数

求0701001号同学选修的课程数。

$\text{count-distinct}_{\text{courseNo}}(\sigma_{\text{StudentNo\#} = '0701001'}(\text{Score}))$

Score关系			
StudentNo	courseNo	term	score
0701001	CN028	07081	85
0701001	CS012	07082	88
0701001	CS015	08091	92
0701008	AC001	07081	76
0701008	CN028	07081	86
0701008	CS012	07082	93
0701008	CS015	08091	96
0703010	AC001	07081	92
0703010	CN028	07081	83
0703010	CS012	07082	73
0703045	AC001	07081	52
0703045	AC001	08091	94
0703045	CN028	07081	80
0703045	CS015	08091	82
0802002	AC001	08091	98
0802002	CN028	08091	72
0802002	CS015	09101	85
0802005	AC001	09101	88
0802005	CS012	08092	90
0802005	CS015	09101	87

聚集函数

● **max**: 求最大值

min: 求最小值

求学生选修数学的最高成绩

max_{score}($\sigma_{\text{courseName} = \text{'数学'}}$ ((**score**) \bowtie **course**))

■ 分组

● 将一个元组集合分为若干个组，在每个分组上使用聚集函数。

属性下标 **G** 聚集函数 属性下标 (关系)

按此属性上的
值对关系分组

对此属性在每个分
组上运用聚集函数

聚集函数

● 分组运算G 的一般形式

G_1, G_2, \dots, G_n **G** $F_1A_1, F_2A_2, \dots, F_mA_m$ (E)

G_i 是用于分组的属性, F_i 是聚集函数, A_i 是属性

G 将E分为若干组, 满足:

- 1) 同一组中所有元组在 G_1, G_2, \dots, G_n 上的值相同。
- 2) 不同组中元组在 G_1, G_2, \dots, G_n 上的值不同。

■ 示例

求每位学生的总成绩和平均成绩

StudentNo **G** sum_{score} , avg_{score} (Score)

Score关系			
StudentNo	courseNo	term	score
0701001	CN028	07081	85
0701001	CS012	07082	88
0701001	CS015	08091	92
0701008	AC001	07081	76
0701008	CN028	07081	86
0701008	CS012	07082	93
0701008	CS015	08091	96
0703010	AC001	07081	92
0703010	CN028	07081	83
0703010	CS012	07082	73
0703045	AC001	07081	52
0703045	AC001	08091	94
0703045	CN028	07081	80
0703045	CS015	08091	82
0802002	AC001	08091	98
0802002	CN028	08091	72
0802002	CS015	09101	85
0802005	AC001	09101	88
0802005	CS012	08092	90
0802005	CS015	09101	87

数据库修改

■ 删除(deletion)

- 将满足条件的元组从关系中删除

$$r \leftarrow r - E$$

关系

- 是对永久关系的赋值运算
- 例：

关系代数表达式

➤ 删除001号学生

➤ $Student \leftarrow Student - \sigma_{P\#studentNo = '001'}(Student)$

数据库修改

■ 插入(insertion)

- 插入一个指定的元组，或者插入一个查询结果

$$r \leftarrow r \cup E$$

- 示例：新加入一个学生
- $Student \leftarrow Student \cup \{ (\text{“002”}, \text{“周正”}, \text{“男”}, 1995-12-12, \text{“上海”}, \text{“汉”}) \}$

数据库修改

■ 更新(updating)

- 利用 **广义投影** 改变元组的某些属性上的值

$$r \leftarrow \Pi_{F_1, F_2, \dots, F_n}(r)$$

- 示例：给每位老师上调10%的工资

$$\text{score} \leftarrow \Pi_{\text{studentno}, \text{courseno}, \text{term}, \text{score} \leftarrow \text{score} * 1.1, (\text{score})}$$

总结

- 讲述了各类运算等，连接，除运算，聚集函数，一些操作等。

本章结束！

请同学们对本章内容进行
复习、总结！！！！

- 四、应用题（共12分，每小题3分）
 - (1) 商品表（商品编号，商品名称，单位，成本价，商品类别，库存数量）
Item (code, Iname, unit, cost, type, amount)
 - (2) 商品销售主表（销售单号，销售日期，客户编号）
SaleHead (saleOrder, saleDate, CustomerId)
 - (3) 商品销售明细表（销售单号，商品编码，单价，销售数量）
SaleDetail (saleOrder, code, price, qty)
 - (4) 客户表（客户编号，客户名称，电话，客户地址）
Customer (CustomerId, Cname, TeleNo, Address)
- 用关系代数完成如下查询：
 - (1) 查询销售给“瑞嘉家具”的所有数量在100以上的商品的销售日期和销售数量。
 - (2) 查询从没有销售过的商品编号，商品名称和库存数量。
 - (3) 查询亏本（成本价大于单价）销售商品的商品名称、单价和客户名称。
 - (4) 查询各类“冰箱”成本价在2000元以上的商品编号、商品名称和库存量。