



# 关系模型和关系代数



PART 01  
关系模型的基本  
概念



PART 02  
关系数据结构



PART 03  
关系操作



PART 04  
关系的完整性



PART 05  
关系数据模型  
的优缺点



PART 06  
关系代数



PART 07  
关系演算



- 掌握关系数据模型的结构、操作与约束
- 理解关系性质，掌握主键、外键等基本术语
- 理解并掌握运用关系数据模型的实体完整性、参照完整性
- 掌握关系代数运算方法



## ● 关系实例

关系实例，是由命名的若干列和行组成的**表格**。关系中的**列**称为**属性**，**行**称为**元组**；关系实例中元组的**数目**称为**基数**（Cardinality）。

**“医生”关系的示例**

属性						
Dno	Dname	Dsex	Dage	Ddeptno	Ttype	Snumber
140	郝亦柯	男	28	102	医师	1800
21	刘伟	男	43	103	副主任医师	2800
368	罗晓	女	27	102	主治医师	2000
73	邓英超	女	43	201	主任医师	3200
82	杨勋	男	36	101	副主任医师	2800

元组



## ● 关系模式

- (1) **关系名**，关系名在数据库中必须唯一
- (2) 关系中的**属性名**以及相关联的**域名**
- (3) **完整性约束**，数据的相关形式规则

一般表示为：

**关系名 (属性1, 属性2, ..., 属性n)**

关系名

属性和对应域

**Doctor**

Dno	Dname	Dsex	Dage	Ddeptno	Ttype	Snumber
140	郝亦柯	男	28	102	医师	1800
21	刘伟	男	43	103	副主任医师	2800
368	罗晓	女	27	102	主治医师	2000
73	邓英超	女	43	201	主任医师	3200
82	杨勋	男	36	101	副主任医师	2800



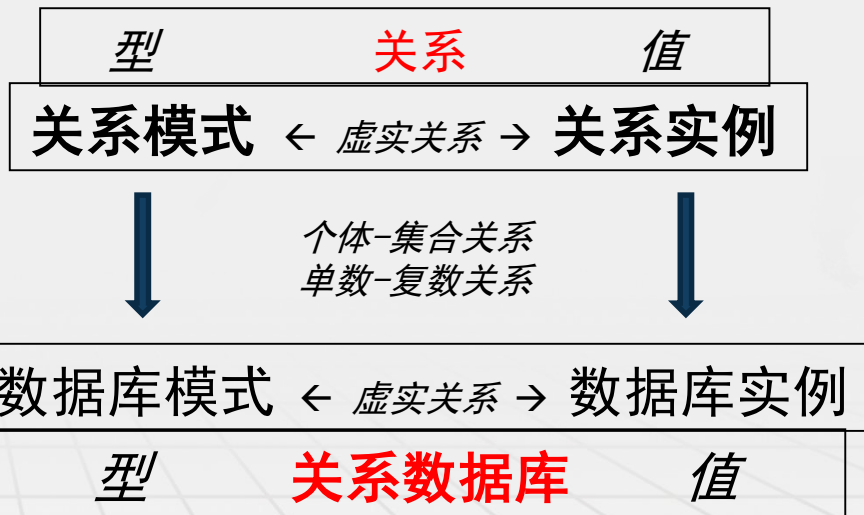
# 1. 关系模型

## ● 关系数据库

关系数据库 = 关系实例的集合 = “**关系实例**  $s$ ”

关系数据库 = 关系模式的集合 = “**关系模式**  $s$ ”

关系数据库 = “**实体+联系**  $s$ ”





## ● 关系

关系是一种规范化的二维表，应满足如下3个条件：

- 每个关系都有一个唯一的\*\*关系名\*\*；
- 关系表中的每一列都是不可再分的基本属性，各属性不能重名；
- 表中的行列次序不重要，可以交换行、列的前后顺序；

**“医生”关系的示例**

**DOCTOR**

Dno	Dname	Dsex	Dage	Ddeptno	Ttype	Snumber
140	郝亦柯	男	28	102	医师	1800
21	刘伟	男	43	103	副主任医师	2800
368	罗晓	女	27	102	主治医师	2000
73	邓英超	女	43	201	主任医师	3200
82	杨勋	男	36	101	副主任医师	2800



## 患者与医生关系的实例(基本表)

就诊信息表

DGno	Pno	Dno	Symptom	Diagnosis	DGtime	Rfee
1645	481	140	呼吸道感染	伤风感冒	2017-7-21 01:12:01	3
2170	201	21	皮肤和软组织感染	细菌感染	2017-7-22 10:10:03	5
3265	161	82	胃溃疡	螺杆菌感染	2017-7-23 10:59:42	5
3308	181	82	消化不良	胃病	2017-7-23 11:11:34	5
3523	501	73	心力衰竭	高血压	2017-7-23 02:01:05	7
7816	421	368	肾盂结石	肾结石	2018-1-8 05:17:03	3

患者信息表

Pno	Pname	Pid	Pino	Pmno	Psex	Pbd	Padd
161	刘景	142201198702130061	1201676	6781121941	男	1987-2-13	新华路光源街
181	陈禄	142201196608190213	1204001	5461021938	男	1966-8-19	城建路茂源巷
201	曾华	142201197803110234	0800920	1231111932	男	1978-3-11	新建路柳巷
421	傅伟相	142202199109230221	0700235	4901021947	男	1991-9-23	高新区西源大道
481	张珍	142201199206200321	1200432	3451121953	女	1992-6-20	西湖区南街
501	李秀	142203198803300432	0692015	3341111936	女	1988-3-30	泰山大道北路

医生信息表

Dno	Dname	Dsex	Dage	Ddeptno	Tno
140	郝亦柯	男	28	101	01
21	刘伟	男	43	104	01
368	罗晓	女	27	103	04
73	邓英超	女	43	105	33
82	杨勋	男	36	104	35



## ● 元组

表中的一行（即一个记录），表示一个实体。关系是由元组组成的。

## ● 属性

二维表中的每一列在关系中称为属性。每个属性都有一个属性名，属性值则是各元组属性的取值。



Pno	Pname	Pid	Pino	Pmno	Psex	Pbd	Padd
161	刘景	142201198702130061	1201676	6781121941	男	1987-2-13	新华路光源街
181	陈禄	142201196608190213	1204001	5461021938	男	1966-8-19	城建路茂源巷
201	曾华	142201197803110234	0800920	1231111932	男	1978-3-11	新建路柳巷
421	傅伟相	142202199109230221	0700235	4901021947	男	1991-9-23	高新区西源大道
481	张珍	142201199206200321	1200432	3451121953	女	1992-6-20	西湖区南街
501	李秀	142203198803300432	0692015	3341111936	女	1988-3-30	泰山大道北路

## ● 域

属性的取值范围称为域。同一属性只能在相同域中取值。





# 1. 关系模型



## ● 度

属性域的个数称为关系的目或度。

## ● 分量

元组中的一个属性值。



Pno	Pname	Pid	Pino	Pmno	Psex	Pbd	Padd
161	刘景	142201198702130061	1201676	6781121941	男	1987-2-13	新华路光源街
181	陈禄	142201196608190213	1204001	5461021938	男	1966-8-19	城建路茂源巷
201	曾华	142201197803110234	0800920	1231111932	男	1978-3-11	新建路柳巷
421	傅伟相	142202199109230221	0700235	4901021947	男	1991-9-23	高新区西源大道
481	张珍	142201199206200321	1200432	3451121953	女	1992-6-20	西湖区南街
501	李秀	142203198803300432	0692015	3341111936	女	1988-3-30	泰山大道北路

## ● 键

关系中能**唯一区分**不同元组的**属性或属性组**，称为关系的一个**键**（Key），或者称为**关键字或码**。需要强调的是，关键字的属性值必须唯一，并且不能取“空值”。



### ● 候选键

凡在关系中能够唯一区分确定不同元组的属性或属性组，均称为候选键（Candidate Key）。

候选键的特征主要包括两点：唯一性和最小性。

### ● 主属性 & 非主属性

包括在候选键中的属性成为主属性（Prime Attribute），不包括在任何候选键中的属性称为非主属性。

### ● 全码

在最简单的情况下，候选键只包含一个属性。在最极端的情况下，关系模式的所有属性都是这个关系模式的候选键，称为全码（All Key）。



# 1. 关系模型

## ● 主键

当一个关系中有多个候选键的时候，则从中选定一个作为关系的主键（Primary Key），关系中主键是唯一的。原则上每个关系中都必定有且仅有一个主键。

主键

Pno	Pname	Pid	Pino	Pmno	Psex	Pbd	Padd
161	刘景	142201198702130061	1201676	6781121941	男	1987-2-13	新华路光源街
181	陈禄	142201196608190213	1204001	5461021938	男	1966-8-19	城建路茂源巷
201	曾华	142201197803110234	0800920	1231111932	男	1978-3-11	新建路柳巷
421	傅伟相	142202199109230221	0700235	4901021947	男	1991-9-23	高新区西源大道
481	张珍	142201199206200321	1200432	3451121953	女	1992-6-20	西湖区南街
501	李秀	142203198803300432	0692015	3341111936	女	1988-3-30	泰山大道北路



# 1. 关系模型

## ● 外键

F是关系R中某个属性或属性组, **而非**该关系的键, 但**F却是**另一个关系S的主键, 则称F为**关系R的外键** (Foreign Key)。其中, 关系R称为**参照关系** (Referencing Relation), 关系S称为**被参照关系** (Referenced Relation)。

S. 医生信息表

Dno	Dname	Dsex	Dage	Ddeptno	Tno
140	郝亦柯	男	28	101	01
21	刘伟	男	43	102	02
368	罗晓	女	27	103	04
73	邓英超	女	43	105	33
82	杨勋	男	36	104	35

主键

被参照关系

R. 就诊信息表

DGno	Pno	Dno	Symptom	Diagnosis	DGtime	Rfee
1645	481	140	呼吸道感染	伤风感冒	2007-7-21 14:12:01	3
2170	201	21	皮肤和软组织感染	细菌感染		
3265	161	82	胃溃疡	螺杆菌感染	2007-7-23 10:59:42	5
3308	81	82	消化不良	胃病	2007-7-23 11:11:34	5
3523	1	73	心力衰竭	高血压	2007-7-23 02:01:05	7
7816	421	368	肾盂结石	肾结石	2008-1-8 05:17:03	3

主键

外键

参照关系



- 关系操作的特点

- 关系操作语言操作一体化

数据定义，查询，更新，控制一体化。

可以作为宿主语言嵌入到主语言中，又可以作为独立的语言交互使用。

- 操作方式是一次一集合（set-at-a-time）方式

操作的对象和结果都是集合。

- 关系操作语言是高度非过程化的语言

具有强大的表达能力，用户使用关系语言时，只需要指出作什么，而不需要指出怎么做，数据存取路径的选择，数据操作方法的选择和优化由DBMS自动完成。



### ● 关系操作的查询语言

早期的关系操作能力通常用代数方式或逻辑方式来表示，分别称为**关系代数**和**关系演算**。

关系代数是通过对关系的运算来表达查询要求的。

关系演算是用谓词来表达查询要求的。

关系演算可按谓词变元的基本对象是元组变量还是域变量分为**元组关系演算**和**域关系演算**。

关系代数、元组关系演算和域关系演算三种语言在表达能力上是完全等价的，均是抽象的查询语言。





- 关系操作采用集合操作方式，即操作的对象和结果都是集合  
结构化查询语言（Structured Query Language, SQL）。SQL不仅具有丰富的查询功能，而且具有数据定义和数据控制功能，是集数据查询语言（Data Query Language, DQL）、数据定义语言（Data Definition Language, DDL）、数据操作语言（Data Manipulate Language, DML）和数据控制语言（Data Control Language, DCL）于一体的关系数据语言。



## ● 关系操作语言种类

- 关系代数语言，如ISBL
- 关系演算语言：包括元组关系演算语言（如APLHA，QUEL）和域关系演算语言（如QBE）
- 基于映象的语言，它具有关系代数和关系演算双重特点，如SQL。

SQL不仅具有丰富的查询功能，而且具有数据定义和数据控制功能，是集

**数据查询语言（Data Query Language, DQL）、**

**数据定义语言（Data Definition Language, DDL）、**

**数据操作语言（Data Manipulate Language, DML）和**

**数据控制语言（Data Control Language, DCL）**

于一体的关系数据语言。





### ● 关系数据库的约束分类

- (1) **数据模型中固有的约束**。例如，要求关系中不能有重复元组的约束就是一个固有约束。
- (2) 可以在数据模型的模式中**直接表述**的约束，通常用DDL加以指定。关系模型的完整性约束就是可以在模式中直接表述的约束。
- (3) 不能在数据模型的模式中直接表述的约束，因此必须由**应用程序表示和执行**。
- (4) **数据依赖**，包括函数依赖和多值依赖。



- 域 (Domain) 是一组具有相同数据类型的值的集合

域中所包含的值的个数称为域的基数 (用 $m$ 表示)

- 笛卡儿积 (Cartesian Product) 是给定一组域 $D_1, D_2, \dots, D_n$ , 这些域中可以有相同的域。

笛卡儿积 $D_1 \times D_2 \times \dots \times D_n$ 的子集叫做在域 $D_1, D_2, \dots, D_n$ 上的 $n$ 元关系 (Relation)。

用 $R(D_1, D_2, \dots, D_n)$ 表示,  $R$ 是关系名,  $n$ 是关系的目或度。

关系中的每个元素是关系中的元组, 通常用 $t$ 表示。当 $n = 1$ 时, 称该关系为单元或一元关系, 当 $n = 2$ 时, 称该关系为二元关系。

- 关系可以有三种类型: 基本关系 (通常又称为基本表)、查询表和视图表。

**基本表**是实际存在的表, 它是实际存储数据的逻辑表示。**查询表**是查询结果对应的表。**视图表**是由基本表或其他视图导出的表。



### ● 关系的性质

(1) 每一个列的分量必须来自同一个域，必须是同一类型的数据。

科室信息表

姓 名	职 称	科 室
刘伟	主治医师	门诊部
罗晓	主治医师	消化内科
杨勋	主任医师	门诊部

域 — 科室（消化内科，门诊部，急诊外科，内分泌科，肿瘤科，口腔科）



(2) 不同列的取值可来自同一个域

(3) 列的顺序可以任意交换(交换时, 连同属性名一起交换)

该表的两个属性“职业”“兼职”来自同一个域

姓 名	职 业	兼 职
张三	教师	工人
李四	工人	教师
王五	工人	辅导员



该表的两个属性列的顺序可以任意交换

姓 名	兼 职	职 业
张三	工人	教师
李四	教师	工人
王五	辅导员	工人



- (4) 关系中的元组的顺序可任意交换
- (5) 关系中不允许出现相同的元组
- (6) 关系中每一分量必须是不可分割的数据项。

姓 名	住 址	
	省 市	行 政 区
李四	四川省成都市	金牛区
王五	四川省成都市	高新区

该表中，列“住址”不符合关系的性质要求



### ● 关系模式

关系模式 (Relation Schema) 是型，关系是值，

关系模式是一个五元组，形式化定义为： $R(U, D, DOM, F)$

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

R: 关系名，

U: 组成该关系的属性名集合，一般为  $(A_1, A_2, \dots, A_n)$ ；

D: 属性组U中属性所来自的域；

DOM: 属性向域的映象集合；

F: 属性间数据的依赖关系集合。

一般表示为：关系名（属性1，属性2，…，属性n），即

$R(A_1, A_2, \dots, A_n)$  or  $R(U)$



## 2. 关系模型的数据结构



关系数据库模式 S 包含

关系模式的集合  $S = \{R_1, R_2, \dots, R_m\}$  和 完整性约束的集合 IC。

下表显示了一个关系数据库模式，

记作  $HIS = \{\text{Dept}, \text{Doctor}, \text{Patient}, \text{Diagnosis}\}$ ，（带下划线的表示主键）

HIS

关系数据库模式  
的模式图

Patient

<u>Pno</u>	Pname	Pid	Pino	Pmno	Psex	Pbd	Padd
------------	-------	-----	------	------	------	-----	------

Doctor

<u>Dno</u>	Dname	Dsex	Dage	Ddeptno	Tno
------------	-------	------	------	---------	-----

Diagnosis

<u>Dgno</u>	Pno	Dno	Symptom	Diagnosis	DGtime	Rfee
-------------	-----	-----	---------	-----------	--------	------

Dept

<u>DeptNo</u>	DeptName	ParentDeptNo	Manager
---------------	----------	--------------	---------



## 2. 关系模型的数据结构



### HIS数据库模式一个可能的数据库

Doctor

Dno	Dname	Dsex	Dage	Ddeptno	Tno
140	郝亦柯	男	28	101	01
21	刘伟	男	43	104	01
368	罗晓	女	27	103	04
73	邓英超	女	43	105	33
82	杨勋	男	36	104	35

Patient

Pno	Pname	Pid	Pino	Pmno	Psex	Pbd	Padd
161	刘景	142201198702130061	1201676	6781121941	男	1987-2-13	新华路光源街
181	陈禄	142201196608190213	1204001	5461021938	男	1966-8-19	城建路茂源巷
201	曾华	142201197803110234	0800920	1231111932	男	1978-3-11	新建路柳巷
421	傅伟相	142202199109230221	0700235	4901021947	男	1991-9-23	高新区西源大道
481	张珍	142201199206200321	1200432	3451121953	女	1992-6-20	西湖区南街
501	李秀	142203198803300432	0692015	3341111936	女	1988-3-30	泰山大道北路

Diagnosis

Dgno	Pno	Dno	Symptom	Diagnosis	DGtime	Rfee
1645	481	140	呼吸道感染	伤风感冒	2017-7-21 01:12:01	3
2170	201	21	皮肤和软组织感染	细菌感染	2017-7-22 10:10:03	5
3265	161	82	胃溃疡	螺杆菌感染	2017-7-23 10:59:42	5
3308	181	82	消化不良	胃病	2017-7-23 11:11:34	5
3523	501	73	心力衰竭	高血压	2017-7-23 02:01:05	7
7816	421	368	肾盂结石	肾结石	2018-1-8 05:17:03	3

Dept

DeptNo	DeptName	ParentDeptNo	Manager
00	XX医院		
10	门诊部	00	
101	消化内科	10	82
102	急诊内科	10	368
103	门内三诊室	10	21
20	社区医疗部	00	
201	家庭病床病区	20	73





## ● 基本的关系操作

关系模型中常用的关系操作包括两大部分：

**查询**操作和

**更新**操作（具体分为插入、删除、修改操作）

查询是关系操作中最重要的部分。查询操作可分为：**选择、连接、投影、并、差、除、交、笛卡尔积**等。

其中，**选择、投影、并、差、笛卡尔积**是5种基本操作。



## 4. 关系的完整性



### ● 完整性约束

关系模型中，有三类完整性约束，即**实体完整性**、**参照完整性**和**用户定义的完整性**。

其中**实体完整性**和**参照完整性**是关系模型中必须满足的完整性约束条件，被称作关系的两个不变性。

任何关系数据库系统都应该支持这两类完整性。

除此之外，不同的关系数据库系统由于应用环境的不同，往往还需要一些特殊的约束条件，这就是**用户自定义完整性**。



### ● 实体完整性

实体完整性要求表中的所有行都有唯一的标识符，称为主键。

实体完整性规则规定：基本关系的**所有主关键字对应的主属性都不能取空值，且取值唯一。**

**在关系模型中，主关键字作为唯一的标识，且不能为空值。**



### ● 参照完整性

参照完整性规则：若属性（或属性组）F是基本关系R的外键，它与基本关系S的主键相对应（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中某个元组的主键值。



## 4. 关系的完整性

### ● 参照完整性

请问哪些数据可以成功添加到医生信息表中？

不可

医生信息表

医生编号	姓名	科室编号	职称
140	郝亦柯	102	医师
121	刘伟	103	副主任医师

科室信息表

科室编号	科室名称	科室负责人	科室位置
102	门诊部	张三	H1-1
103	消化内科	李四	H1-2
104	急诊外科	王五	H1-3

‘医生信息表’ 添加的三条记录

医生编号	姓名	科室编号	职称
001	冯如意	109	医师
002	李伟	NULL	副主任医师
003	黄海兴	103	副主任医师

可以

可以



### ● 用户定义完整性

用户定义完整性则是根据应用环境的要求和实际的需要，对某一具体应用所涉及的数据提出约束性条件。

例如：某个属性必须取唯一值，规定某个属性的取值范围等。

**性别** 属性：男/女，

(百分制) **成绩** 属性：0和100分之间，

**婚姻状态** 属性：未婚/已婚/离婚，



### ● 关系数据模型的优点

- (1) 关系数据模型与非关系数据模型不同，它是建立在严格的数学概念基础上的。
- (2) 无论实体还是实体之间的联系都用关系来表示。对数据的检索结果也是关系（即表），概念单一，其数据结构简单、清晰。
- (3) 关系数据模型的存取路径对用户透明，从而具有更高的数据独立性，更好的安全保密性，简化了程序员的工作和数据库开发建立的工作。
- (4) 关系数据模型具有丰富的完整性，如实体完整性、参照完整性和用户定义的完整性，大大降低了数据的冗余和数据不一致的概率。



### ● 关系数据模型的缺点

- (1) 对“现实世界”实体的表达能力弱。
- (2) 由于存取路径对用户透明，查询效率往往不如非关系数据模型。
- (3) 关系数据模型只有一些固定的操作集，如面向集合和记录的操作，操作是在SQL规格说明中提供的。
- (4) 能很好地支持业务规则，很多商业化系统不能完全支持实体和参照完整性、域等业务规则，所以需要将它们内置到应用程序中。





### ● 关系代数概述

关系代数是一个由各种运算组成的系统，

**运算对象、运算符、运算结果**是运算的三大要素。

关系代数用到的运算符包括四类：

**集合运算符、专门的关系运算符、比较运算符和逻辑运算符。**



### 关系代数运算符

运 算 符		含 义	运 算 符		含 义
集合运算符	$\cup$	并 差 交 笛卡尔积	比较运算符	$>$	大于
	$-$			$\geq$	大于等于
关系运算符 专门的	$\cap$	选 择 投 影 连 接 除 更 名	逻辑运算符	$<$	小于
	$\times$			$\leq$	小于等于
				$=$	等于
				$\neq$	不等于
	$\sigma$			$\neg$	非
	$\pi$			$\wedge$	与
	$\bowtie$			$\vee$	或
	$\div$				
	$\rho$				



### ● 关系代数概述

关系代数中的操作可分为三类。

- 传统的集合运算：并、差、交、笛卡尔积；
- 专门的关系运算：投影、选择、连接、除；
- 扩展的关系运算：广义投影、聚集函数和分组、递归闭包。



### ● 传统集合运算 — 并 交 差

设关系R和S具有相同的模式，t是元组变量， $t \in R$ 表示t是R的一个元组。则并、差、交运算定义如下：

#### (1) 并运算

R和S的并是由属于R或属于S的元组构成的集合，记为： $R \cup S = \{t \mid t \in R \vee t \in S\}$

#### (2) 差运算

R和S的差是由属于R但不属于S的元组构成的集合，记为： $R - S = \{t \mid t \in R \wedge t \notin S\}$



### ● 传统集合运算 — 并 交 差

#### (3) 交运算

R和S的交是由既属于R又属于S的元组构成的集合，其结果关系仍为n目关系，记为：

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

式中，“ $\cap$ ”为交运算符。关系的交可以用差来表示，即：

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



## 6. 关系代数



【例】给定以下三个关系 R1和 R2、R3，求 (1)  $R1 \cup R2$  (2)  $R1 - R2$  (3)  $R1 \cap R2$

表 R1

患者编号	患者姓名	患者性别	患者年龄	社会保险号
206001	金荣	男	24	500230241
206002	丁冬	男	21	301236542
206003	唐雯	女	50	250413692

表 R2

患者编号	患者姓名	患者性别	患者年龄	社会保险号
206004	李华林	男	65	111425255
206003	唐雯	女	50	250413692

表 R3

患者编号	电话类型	电话号码
206001	家庭电话	85639456
206003	手机	1301525xxxx



## 6. 关系代数



表 R1 U R2

患者编号	患者姓名	患者性别	患者年龄	社会保险号
206001	金荣	男	24	500230241
206002	丁冬	男	21	301236542
206003	唐雯	女	50	250413692
206004	李华林	男	65	111425255



表 R1-R2

患者编号	患者姓名	患者性别	患者年龄	社会保险号
206001	金荣	男	24	500230241
206002	丁冬	男	21	301236542

表  $R1 \cap R2$

患者编号	患者姓名	患者性别	患者年龄	社会保险号
206003	唐雯	女	50	250413692





### ● 传统集合运算 — 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) 列的元组的集合。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R有k1个元组，S有k2个元组，则关系R和关系S的笛卡尔积有k1xk2个元组。

记作： $R \times S = \{\widehat{t_r t_s} \mid t_r \in R \wedge t_s \in S\}$



【例】给定以下两个关系 R1、R3，求  $R1 \times R3$

表 R1

患者编号	患者姓名	患者性别	患者年龄	社会保险号
206001	金荣	男	24	500230241
206002	丁冬	男	21	301236542
206003	唐雯	女	50	250413692

表 R3

患者编号	电话类型	电话号码
206001	家庭电话	85639456
206003	手机	1301525xxxx



表 R1XR3

患者编号	患者姓名	患者性别	患者年龄	社会保险号	患者编号	电话类型	电话号码
206001	金荣	男	24	500230241	206001	家庭电话	85639456
206001	金荣	男	24	500230241	206003	手机	1301525xxxx
206002	丁冬	男	21	301236542	206001	家庭电话	85639456
206002	丁冬	男	21	301236542	206003	手机	1301525xxxx
206003	唐雯	女	50	250413692	206001	家庭电话	85639456
206003	唐雯	女	50	250413692	206003	手机	1301525xxxx



### ● 专门关系运算 — 选择

选择即从关系中找出满足给定条件的所有元组。

使用“选择”运算符，可以提取关系水平方向的切片。选择运算符用“ $\sigma_b$ ”表示

这里b给出了用来提取关系水平切片的布尔谓词。

选择运算记为： $\sigma_F(R) = \{t | t \in R \wedge F(t) == 'True'\}$

F：是一个条件，由以下三部分组成：运算对象、算术比较符、逻辑运算符。

t：满足给定条件的元组。

R：操作对象



## 6. 关系代数

【例】在下面所示患者信息表R中，利用选择运算把30岁以下男患者找出来。

操作对象

表 患者信息表R

条件

患者编号	患者姓名	患者性别	患者年龄	社会保险号
206001	金荣	男	24	500230241
206002	丁冬	男	21	301236542
206003	唐雯	女	50	250413692
206004	李华林	男	65	111425255
206005	文娟	女	45	789256342



$$\sigma_{(\text{患者性别}='男') \wedge (\text{患者年龄}<'30')}(R)$$

R: 代表操作对象: 患者信息表R

F: 代表条件:  $(\text{患者性别}='男') \wedge (\text{患者年龄}<'30')$

结果如下图

表  $\sigma_F(R)$

患者编号	患者姓名	患者性别	患者年龄	社会保险号
206001	金荣	男	24	500230241
206002	丁冬	男	21	301236542



### ● 专门关系运算 — 投影

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： $\Pi_A(R)$

A为R中的属性列。投影操作是从列的角度进行的运算。

表 患者信息表R

患者编号	患者姓名	患者性别	患者年龄	社会保险号
206001	金荣	男	24	500230241
206002	丁冬	男	21	301236542
206003	唐雯	女	50	250413692
206004	李华林	男	65	111425255
206005	文娟	女	45	789256342



【例】在患者信息表R中，利用投影运算得到患者的姓名和社会保险号，投影结果如下表所示。

$$\Pi_{\text{患者姓名}, \text{社会保险号}}(R)$$

表 将R关系在患者姓名和社会保险号属性上投影

患 者 姓 名	社会 保 险 号
金 荣	500230241
丁 冬	301236542
唐 雯	250413692
李华林	111425255
文 娟	789256342





### ● 专门关系运算 — 连接

连接也称为 $\theta$ 连接。连接运算从R和S的笛卡尔积 $R \times S$ 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的元组。

连接运算中有两种最为重要也最为常见的连接，一种是自然连接，另一种是条件连接。

### ● 自然连接

自然连接要求两个关系中进行比较的分量必须是相同的属性组，并且要在结果中将重复的属性去掉。它的形式定义为：

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



### ● 自然连接

自然连接步骤：

- 计算  $R \bowtie S$ ;
- 选择同时出现在R和S中属性相等元组;
- 去掉重复属性。

连接运算一般是从行的角度进行的操作，但自然连接是同时从行和列的角度进行的操作。  
如果两个关系没有公共属性，自然连接就是笛卡尔积。



## 6. 关系代数

### 专门关系运算-连接



#### ● 自然连接

【例】定义以下两个关系R1和R2，求  $R1 \bowtie R2$

表 R1

A 医生姓名	B 医生性别	C 工资
郝亦柯	男	2500
刘伟	男	2850
邓英超	女	3000

表 R2

D 职称	E 工资
医师	2500
副主任医师	2850
主任医师	3000

表  $R1 \bowtie R2$

A 医生姓名	B 医生性别	C 工资	D 职称
郝亦柯	男	2500	医师
刘伟	男	2850	副主任医师
邓英超	女	3000	主任医师



### ● 条件连接

假定 $\theta$ 为算术运算符。则关系R的第i个属性（第i列）和关系S的第j个属性（第j列）的连接定义为：

$$R \bowtie_{[i]\theta[j]} S = \sigma_{[i]\theta[j]}(R \times S)$$

当 $\theta$ 为“=”时的连接称为等值连接。



## 6. 关系代数



### ● 条件连接

【例】如图所示的关系R1和R2，求  $R1 \bowtie_{C=E} R2$

表 R1

A 医生姓名	B 医生性别	C 工资
郝亦柯	男	2500
刘伟	男	2850
邓英超	女	3000

表 R2

D 职称	E 工资
医师	2500
副主任医师	2850
主任医师	3000

表  $R1 \bowtie_{C=E} R2$

A 医生姓名	B 医生性别	C 工资	D 职称	E 工资
郝亦柯	男	2500	医师	2500
刘伟	男	2850	副主任医师	2850
邓英超	女	3000	主任医师	3000



### ● 专门关系运算 — 除

设关系模式为 $R(A_1, A_2, \dots, A_n)$ 。它的一个关系设为 $R$ ， $t[A_i]$ 则表示元组 $t$ 中相应于属性 $A_i$ 的一个**分量**。

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：

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



### ● 专门关系运算 — 除

给定关系 $R(X, Y)$ 和 $S(Y, Z)$ ，其中 $X, Y, Z$ 为属性组。 $R$ 中的 $Y$ 与 $S$ 中的 $Y$ 可以有不同的属性名，但必须出自相同的域集， $Z$ 可以为空。 $R$ 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。记为：

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

$Y_x$ 是 $x$ 在 $R$ 中的象集， $x = t_r[X]$



### ● 专门关系运算 — 除

【例】给定两个关系R1、R2，求 $R1 \div R2$ 。

表 R1

pid	did
206001	3501
206002	3501
206003	3502
206001	3502
206002	3502
206001	3503

表 R2

did
3501
3502

表  $R1 \div R2$

pid
206001
206002





### ● 专门关系运算—除

解析：

在关系R1中，Pid可以取3个值 {206001, 206002, 206003}。其中：

206001的象集为 {3501, 3502, 3503}

206002的象集为 {3501, 3502}

206003的象集为 {3502}

R2在Did上的投影为 {3501, 3502}

于是可以得出：206001和206002的象集包含了R2在did上的投影，即

$$R1 \div R2 = \{206001, 206002\}$$



### ● 专门关系运算 — 更名

对给定关系代数表达式 $E$ ，表达式 $\sigma_x(E)$ ，返回表达式 $E$ 的结果，并把名字 $x$ 赋给了它。

更名运算的另一形式如下。假设关系代数表达式 $E$ 是 $n$ 元的，则表达式

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

返回表达式 $E$ 的结果，并赋给它名字 $x$ ，同时将各属性更名为 $A_1, A_2, \dots, A_n$ 。

更名运算一般运用于以下情况：给属性重新命名；给表达式结果重新命名；关系的自身连接（即单个关系与自己进行连接）中，将一个关系命名为多个名称便于区分等。



### ● 扩展关系运算—广义投影

广义投影运算通过允许在投影列表中使用算术函数来对投影进行扩展。广义投影运算形式为：

$$\pi F_1, F_2, \dots, F_n (R)$$

这里 $F_1, F_2, \dots, F_n$ 是关系 $R$ 属性上的函数，并可能含有常量。

**【例】**在关系药品(药品编号, 药品名称, 价格, 包装单位, 药品类型)中, 药品类型为西药的药品的价格增加1%, 则可用下式表示:

$$\pi_{\text{药品编号, 药品名称, 价格} \times 1.01, \text{包装单位, 药品类型}} (\sigma_{\text{药品类型} = \text{'西药'}} (\text{药品}))$$



#### ● 扩展关系运算—聚集函数和分组

**聚集函数输入值的一个集合，将单一值作为结果返回。**

常用的聚集函数：最大值max，最小值min，平均值avg，总和值sum，计数值count等。

使用聚集函数的集合中，一个值可以出现多次，值出现的顺序是无关紧要的，这样的集合称为**多重集**。

集合是多重集的特例，其中每个值都只出现一次。有时，我们只想对重复的值计算一次，可以使用distinct去除重复值。



- 扩展关系运算—聚集函数和分组

【例】统计医生有几种职称，这里每个职称应只计算一次，查询表达式为：

$$\Gamma_{\text{count-distinct}(\text{职称})}(\text{医生})$$

其中，符号 $\Gamma$ 是字母G的手写体，关系代数运算符 $\Gamma$ 表示聚集将被应用。



【例】我们希望分别找出每级职称对应的医生人数而不是所有的医生人数，查询表达式如下所示：

$$(\text{职称}) \Gamma_{\text{count}(\text{医生编号})}(\text{医生})$$

在这个表达式中  $\Gamma$  左侧的下标职称属性表明输入关系医生按照职称的值进行分组， $\Gamma$  右侧下标的表达式  $\text{count}(\text{医生编号})$  表明对每组元组（每级职称）计数。



### ● 扩展关系运算—递归闭包

假定 $k$ 是关系 $R$ 的一个关键字； $f$ 是 $R$ 中的一个外关键字，且后向引用到 $R$ ；

$S = \sigma_c(R)$   $X_0 = S$  是包含于 $R$ 的一个初始元组集。对于 $i = 0, 1, 2, \dots$ ，考虑下面的

操作序列：  $X_i = X_{i-1} \cup \pi_R \sigma_{b_i}(\rho(X_{i-1}) \times R), i > 0, b_i = (X_{i-1}.k = f)$

$R$ 中 $S$ 的递归闭包用  $\sum_{k=f}(R, S) = \min\{i \mid X_i = X_{i+1}\}$

$R$ 的子集 $S$ 是 $R$ 中一些元组的集合，它被用做种子，且 $S$ 的属性名与 $R$ 相同。 $X_0$ 的取值为种子 $S$ ，以后每一个随后的 $X_{i+1}$ 都给它的前趋 $X_i$ 添加一些 $R$ 中的元组。

因为 $R$ 中的元组个数是有限的，所以必定存在某个 $i$ 使得 $X_{i+1}$ 不能向 $X_i$ 中添加更多的元组，即：  $X_i = X_{i+1} = X_{i+2} \dots$ 。这个 $X_i$ 的稳定值便是 $R$ 中 $S$ 的递归闭包。



### ● 元组关系演算

元组关系演算中的查询表达式为：

$$\{t | \Phi(t)\}$$

该表达式的含义是使谓词 $\Phi$ 为真的元组 $t$ 的集合。其中 $t$ 为元组变量，表示一个定长的元组。谓词演算公式 $\Phi(t)$ 是由原子公式和运算符组成的公式。

原子公式有以下三类：

(1)  $R(t)$ 。 $R$ 是关系名， $t$ 是元组变量。

(2)  $t[X] \theta u[Y]$ 。 $t$ 和 $u$ 是元组变量， $\theta$ 是算术比较运算符。 $t[X]$ 和 $u[Y]$ 分别表示 $t$ 的 $X$ 分量和 $u$ 的 $Y$ 分量。 $t[X] \theta u[Y]$ 表示 $t$ 的 $x$ 分量与 $u$ 的 $Y$ 分量满足比较关系 $\theta$ 。

(3)  $t[X] \theta C$ 。这里 $C$ 是常量。

例如：查询医院年龄大于40的医生信息： $\{t | \text{医生}(t) \wedge t[\text{年龄}] > 40\}$





### ● 元组关系演算

公式的递归定义如下。

- (1) 每个原子公式是一个公式。
- (2) 设 $\Phi_1$ 和 $\Phi_2$ 是公式. 则 $\neg\Phi_1$ ,  $\Phi_1 \wedge \Phi_2$ ,  $\Phi_1 \vee \Phi_2$ 也都是公式。
- (3) 若 $\Phi$ 是公式,  $s$ 是元组变量, 则 $\exists s(\Phi)$ 和 $\forall s(\Phi)$ 也是公式。 $\exists$ 为存在量词,  $\forall$ 是全称量词。 $\exists s(\Phi)$ 表示的命题是: “存在一个元组 $s$ 使得公式 $\Phi$ 为真”。 $\forall s(\Phi)$ 表示的命题是: “对于所有元组 $s$ 都使得公式 $\Phi$ 为真”。如果元组变量不被 $\exists$ 和 $\forall$ 修饰, 则称为自由变量。



### ● 元组关系演算

在元组演算公式中，运算符的**优先级**有下面四项。

- (1) 算术比较运算符 $=$ 、 $\neq$ 、 $<$ 、 $\leq$ 、 $>$ 、 $\geq$ 最高。
- (2) 存在量词 $\exists$ ，全称量词 $\forall$ 。
- (3) 逻辑运算符最低， $\neg$ 、 $\wedge$ 、 $\vee$ 的先后次序进行。
- (4) 如果有括号，则括号的优先级别最高。



## ● 等价表示

(1) 并  $R \cup S = \{t | R(t) \vee S(t)\}$ 。

(2) 差  $R - S = \{t | R(t) \wedge \neg S(t)\}$ 。

(3) 笛卡儿积 
$$R \times S = \{t^{(n+m)} \mid (\exists u^n)(\exists v^m)(R(u) \wedge S(v) \wedge t[1] = u[1] \wedge t[2] = u[2] \wedge \cdots \wedge t[n] = u[n] \wedge t[n+1] = v[1] \wedge t[n+2] = v[2] \wedge \cdots t[n+m] = v[m])\}$$

式中， $t^{(n+m)}$ 表示 $t$ 是 $(n + m)$ 元组， $u^n$ 表示 $u$ 是 $n$ 元组， $v^m$ 表示 $v$ 是 $m$ 元组。

(4) 选择  $\sigma_F(R) = \{t | R(t) \wedge P\}$ ，式中， $P$ 为布尔函数。

(5) 投影  $\Pi_x(R) = \{t^k \mid (\exists u)(R(u) \wedge t[1] = u[i_1] \wedge t[2] = u[i_2] \wedge \cdots t[k] = u[i_k])\}$

式中， $x = i_1, i_2, \cdots, i_k$ 。

$\{t | NOT 医生((t))\}$

【例】查询所有在门诊部工作的医生的基本信息。

$\{t | (\exists t) 医生(t) \wedge (\exists s)(组织机构(s) \wedge s[部门编号] = t[部门编号] \wedge s[部门名称] = '门诊部')\}$



### ● 域关系演算语言QBE

通过**例子查询**（Query By Example, QBE）是一种域关系运算的关系语言，同时也指使用此语言的关系数据库管理系统

QBE的特点：

- （1）QBE是交互式语言
- （2）QBE是表格语言
- （3）QBE是基于例子的查询语言



## 6. 关系演算



【例】以学生选课数据库为例：  
学生(学号, 姓名, 性别, 年龄, 系)  
课程(课程号, 课程名)  
选课(学号, 课程号, 成绩)  
找出数学系的所有学生姓名

T是示例  
元素

学生	学号	姓名	性别	年龄	系
		P.T			数学系

学生	学号	姓名	性别	年龄	系
		P.张三			数学系

学生	学号	姓名	性别	年龄	系
	P.1001	P.张三	P.男	P.20	P.数学系

- (1) 用户提出请求。
- (2) 屏幕显示空白表格
- (3) 用户在最左边一栏输入关系名“学生”
- (4) 系统显示该关系的属性名
- (5) 用户构造查询条件
- (6) 屏幕显示查询结果