

# 软件工程专业导论

## ——数据库篇

哈尔滨工业大学（威海）软件学院  
朱东杰

# 第五章 计算机数据管理—数据库

## 5.1 数据库系统的基本概念

数据库系统: DB、DBMS、DBA、DBAP、Computer System

数据库管理系统: 数据库定义、数据库操纵、数据库控制、数据库维护

数据库语言: DDL、DML、DCL

数据库控制: 安全性、完整性、并发性

## 5.2 关系模型与关系数据库

关系的通俗解释

关系模型:

一个结构: 关系

七个操作: 并、差、积、选择、投影、交、连接(自然连接)

三条规则: 实体完整性、参照完整性和用户定义完整性

# 为什么需要数据库?

## 信息社会以信息为基础



——传统纸张上记录的各种信息使用起来非常不方便... ..

- 查询/统计消耗时间
- 人工计算不能很细致
- .....

## 例如：学生成绩管理(1)

课程:计算机基础 学期:98/99春 任课教师: 王敏			课程编号:510001 班级:985101		
学号	姓名	成绩	学号	姓名	成绩
01	张三	85			
02	李四	90			
03	王五	89			
04	李六	75			
05	王一	60			
06	张思	40			
.....					

—— 一个班一门课程有一张成绩单

## 例如：学生成绩管理(2)

课程:数据库

学期:98/99春

任课教师: 刘东

课程编号:510002

班级:985101

学号	姓名	成绩	学号	姓名	成绩
01	张三	85			
02	李四	90			
03	王五	89			
04	李六	75			
05	王一	60			
06	张思	40			
.....					

一个班的所有课程的成绩单

## 例如：学生成绩管理(3)

课程:高等数学 课程编号:010002  
学期:97/98春 班级:974101  
任课教师: 张林

学号	姓名	成绩	学号	姓名	成绩
01	张三	85			
02	李四	90			
03	王五	89			
04	李六	75			
05	王一	60			
06	张思	40			
.....					

} 所有班所有课程的成绩单

# 一个学校成绩管理信息量有多大?

任课教师: 王敏 课程编号: 510001 课程: 计算机基础 学期: 98/99 班级: 985101		任课教师: 王刚 课程编号: 510002 课程: 物理 学期: 98/99春 班级: 985101		任课教师: 高扬 课程编号: 510003 课程: 化学 学期: 98/99春	
学号	姓	学号	姓	学号	姓
01	张三	01	张三	01	张三
02	李四	02	李四	02	李四
03	王五	03	王五	03	王五
04	李六	04	李六	04	李六
05	王一	05	王一	05	王一
06	张思	06	张思	06	张思
.....	.....	.....	.....	.....	.....
06	李六	06	李六	06	李六
.....	.....	.....	.....	.....	.....

- 本科生7个学期有课  $A_1=7$
- 每学期平均有 8 门课  $A_2=8$
- 每个班平均有25人  $A_3=25$
- 每学年平均有70个班  $A_4=70$
- 学校同时有4个学年  $A_5 = 4$

$$\begin{aligned}
 \text{基本信息量} &= A_5 \times A_4 \times A_3 \times A_2 \times A_1 \\
 &= 4 \times 70 \times 25 \times 8 \times 7 = 392000
 \end{aligned}$$

# 完成如下工作需多少人, 多少时间?

任课教师: 王敏 课程编号: 510001 课程: 计算机基础 学期: 98/99 班级: 98510		任课教师: 王刚 课程编号: 510002 课程: 物理 学期: 98/99春 班级: 985101		任课教师: 高扬 课程编号: 510003 课程: 化学 学期: 98/99春	
学号	姓	学号	姓	成绩	
01	张	01	张三	85	
02	李	02	李四	90	
03	王	03	王五	89	
任课教师: 李敏 课程编号: 510001 课程: 计算机基础 学期: 98/99春 班级: 985102		学号	姓名	成绩	
01	张三	01	张三	85	
02	李四	02	李四	90	
03	王五	03	王五	89	
04	李六	04	李六	75	
05	王一	05	王一	60	
06	张思	06	张思	40	
.....		.....			
		06	李六	75	
			王一	60	
			张思	40	
		.....			

- 对一个班一门课:
  1. 计算平均成绩
  2. 统计各分数段的人数
- 对一个班所有门课:
  1. 统计每门课平均成绩
  2. 统计每门课各分数段人数
  3. 统计所有课各分数段人数
- 对所有班所有门课:
  1. 统计各班各门课平均成绩
  2. 统计各人平均成绩
  3. 统计各课程平均成绩
  4. 统计各分数段人数

... ..



# 人与计算机的差异?

任课教师: 王敏 课程编号:510001 课程:计算机基础 学期:98/99 班级:98510		任课教师: 王刚 课程编号:510002 课程:物理 学期:98/99春 班级:985101		任课教师: 高扬 课程编号:510003 课程:化学 学期:98/99春	
学号	姓	学号	姓	学号	姓
01	张三	01	张三	01	张三
02	李四	02	李四	02	李四
03	王五	03	王五	03	王五
任课教师: 李敏 课程编号:510001 课程:计算机基础 学期:98/99春 班级:985102		任课教师: 王敏 课程编号:510001 课程:计算机基础 学期:97/98春 班级:975101		成绩	
学号	姓名	学号	姓名	成绩	
01	张三	01	张三	85	
02	李四	02	李四	90	
03	王五	03	王五	89	
04	李六	04	李六	75	
05	王一	05	王一	60	
06	张思	06	张思	40	
.....		.....			
		06	李六	75	
			王一	60	
			张思	40	
		.....			

人工进行统计:

1. 容易出现差错
2. 枯燥, 乏味
3. 统计分析周期长, 所以不能分析很细

计算机进行统计:

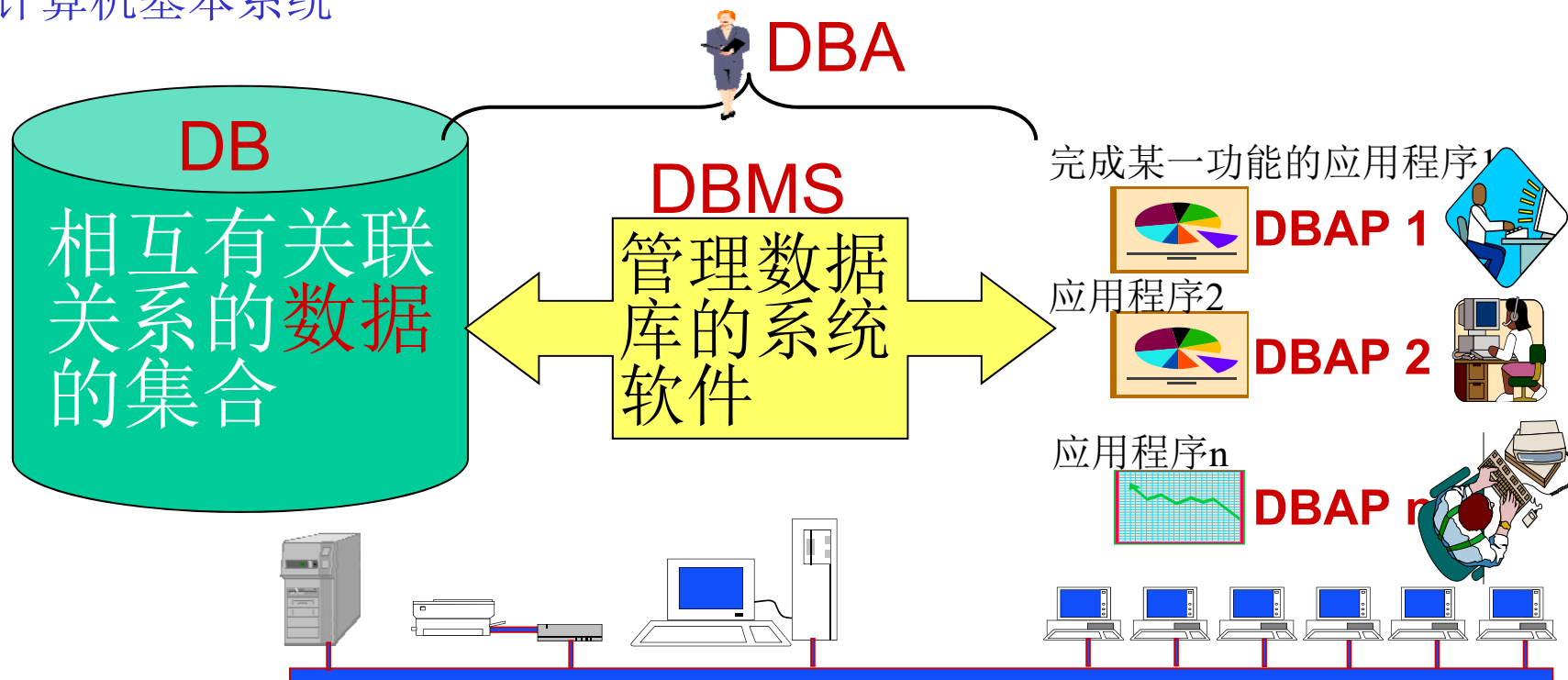
1. 克服人工查询的弱点
2. 使分析工作更为细致

- 计算机应去做人力所不及的工作  
人应更多地去创造性工作!
- 科学化管理就是以详实的数据为依据进行管理, 有了详实的数据, 就有行动的措施.

为进行更详实的数据管理-----就需要**数据库**


# 数据库系统构成?

- 数据库(DB): Database
- 数据库管理系统(DBMS): Database Management System
- 数据库应用(DBAP): DataBase APplication
- 数据库管理员(DBA): DataBase Administrator
- 计算机基本系统



# 数据库?

DB是由表构成的



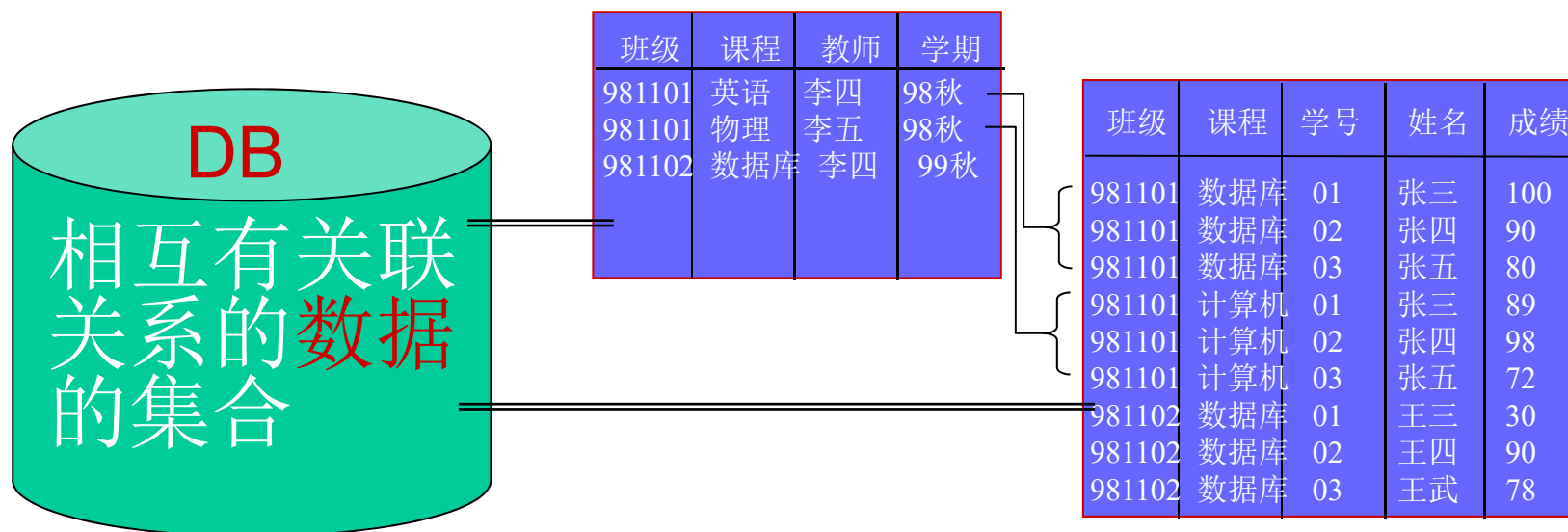
班级	课程	教师	学期	学号	姓名	成绩
981101	数据库	李四	98秋	01	张三	100
981101	数据库	李四	98秋	02	张四	90
981101	数据库	李四	98秋	03	张五	80
981101	计算机	李五	98秋	01	张三	89
981101	计算机	李五	98秋	02	张四	98
981101	计算机	李五	98秋	03	张五	72
981102	数据库	李四	99秋	01	王三	30
981102	数据库	李四	99秋	02	王四	90
981102	数据库	李四	99秋	03	王武	78

== 数据格式

== 数据

# 数据库?

DB中的表是有关联的



# 数据库管理系统 (DBMS) 功能?

- 定义数据库：定义数据库中数据的格式
- 数据库操纵：向数据库中增/删/改数据及查询数据
- 数据库控制：控制数据的使用, 保证数据的正确
- 数据库维护：转储/恢复/重组/性能监测/分析...

数据格式

班级	课程	教师	学期	学号	姓名	成绩
----	----	----	----	----	----	----

数据

981101	数据库	李四	98秋	01	张三	100
981101	数据库	李四	98秋	02	张四	90
981101	数据库	李四	98秋	03	张五	80
981101	计算机	李五	98秋	01	张三	89
981101	计算机	李五	98秋	02	张四	98
981101	计算机	李五	98秋	03	张五	72
981102	数据库	李四	99秋	01	王三	30
981102	数据库	李四	99秋	02	王四	90
981102	数据库	李四	99秋	03	王武	78

981102 数据库 李四 99秋 02 王四 90



谁能访问数据  
更新数据有什么要求



检查用户合法性  
检查数据合法性

DBMS



用户

# 数据库语言?

- 数据定义语言 (DDL: Data Definition Language)  
----DBMS提供给用户, 以使用户定义数据格式
- 数据操纵语言 (DML: Data Manipulation Language)  
----DBMS提供给用户, 以使用户对数据进行操作
- 数据控制语言 (DCL: Data Control Language)  
----DBMS提供给用户, 以使用户对数据进行控制

数据格式

数据

班级	课程	教师	学期	学号	姓名	成绩
981101	数据库	李四	98秋	01	张三	100
981101	数据库	李四	98秋	02	张四	90
981101	数据库	李四	98秋	03	张五	80
981101	计算机	李五	98秋	01	张三	89
981101	计算机	李五	98秋	02	张四	98
981101	计算机	李五	98秋	03	张五	72
981102	数据库	李四	99秋	01	王三	30
981102	数据库	李四	99秋	02	王四	90
981102	数据库	李四	99秋	03	王武	78

981102 数据库 李四 99秋 02 王四 90

数据安全性完整性定义



谁能访问数据  
更新数据有什么要求

检查用户合法性  
检查数据合法性

DBMS

用户

# 数据库语言与数据模型?

- 数据定义语言 (DDL) ? 定义哪些内容?
- 数据操纵语言 (DML) ? 有哪些操作? 如何操作?

## —— 数据模型

严格定义的概念的集合

① 数据结构

有哪些格式的数据

② 数据操作

对这些格式的数据都可能有哪些操作

③ 完整性约束

为保证操作后和操作过程中产生的数据仍符合规定所必须遵守的约束条件

## —— 三种计算机数据模型

关系模型

层次模型

网状模型

# 第5章 计算机数据管理——数据库

1. 数据库系统的基本概念

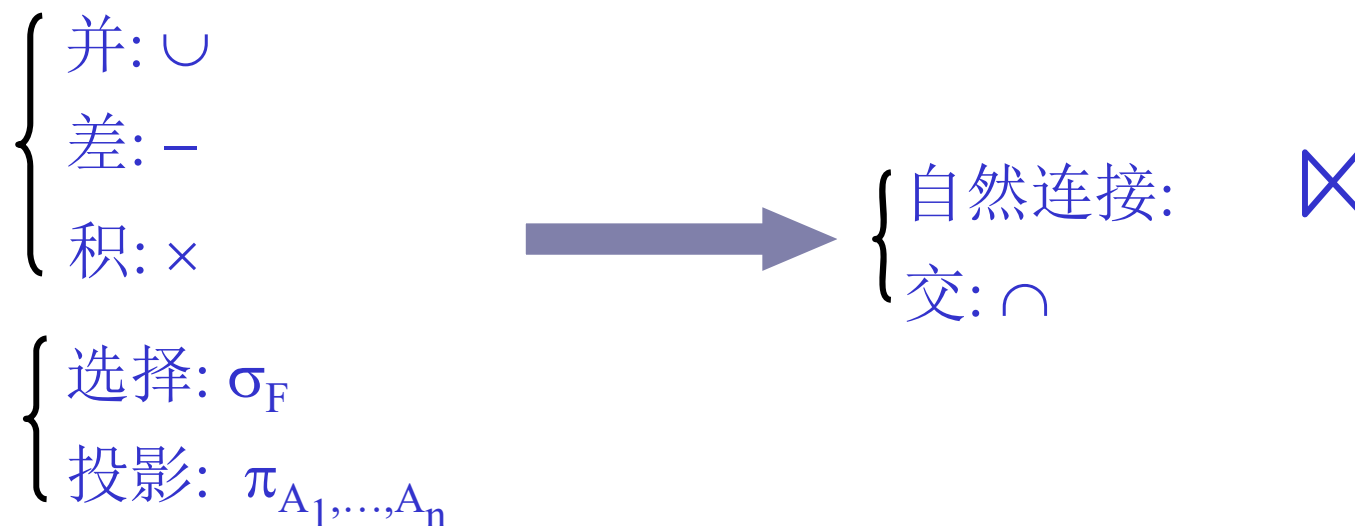
2. 关系模型与关系数据库

3. 关系数据库标准语言——SQL



# 关系模型

- 数据结构      关系
- 数据操作      关系操作/关系代数



- 完整性约束
  - 实体完整性
  - 参照完整性
  - 用户定义的完整性

# 关系模型与关系数据库

- 关系数据库是基于关系模型的数据库
- 关系模型是从表格抽象出来的, 是对传统表格及其操作的严格定义

班级	课程	学号	姓名	成绩
981101	数据库	01	张三	100
981101	数据库	02	张四	90
981101	数据库	03	张五	80
981101	计算机	01	张三	89
981101	计算机	02	张四	98
981101	计算机	03	张五	72
981102	数据库	01	王三	30
981102	数据库	02	王四	90
981102	数据库	03	王武	78

# 关系的通俗解释

表(Table)/关系(Relation)

关系模式

学生成绩单(班级,课程,学号,姓名,成绩)

学生成绩单

班级	课程	学号	姓名	成绩
981101	数据库	01	张三	100
981101	数据库	02	张四	90
981101	数据库	03	张五	80
981101	计算机	01	张三	89
981101	计算机	02	张四	98
981101	计算机	03	张五	72
981102	数据库	01	王三	30
981102	数据库	02	王四	90
981102	数据库	03	王武	78

码(Key)/关键字

某一列或某几列的数据  
能惟一区分每一行

行(Row)/

元组(Tuple)/

记录(Record)

值域(Domain)

列数据的取值范围

列(Column)/字段(Field)/

属性(Attribute)/数据项

列有名称, 有值/数据

# 关系的数学定义

---

定义1 域(Domain): 是值的集合

定义2 笛卡尔积(Cartesian Product)

$$D_1 \times D_2 \times \dots \times D_n = \{(d_1, d_2, \dots, d_n) \mid d_i \in D_i, i = 1, 2, \dots, n\}$$

定义3 关系(Relation)

$D_1 \times D_2 \times \dots \times D_n$  的子集为关系, 记为  $R(D_1, D_2, \dots, D_n)$

# 关系的数学定义的说明

怎样把一张表格定义清楚呢?

班级	课程	学号	姓名	成绩
981101	数据库	01	张三	100
981101	数据库	02	张四	90
981101	数据库	03	张五	80
981101	计算机	01	张三	89
981101	计算机	02	张四	98
981101	计算机	03	张五	72
981102	数据库	01	王三	30
981102	数据库	02	王四	90
981102	数据库	03	王武	78

## 4. 指出关系中的元组

关系中元组是有意义的组合  
----笛卡尔积的子集

## 3. 指出所有可能的元组

值域中值的所有可能的组合  
----笛卡尔积

## 2. 值域(Domain)

说清楚每一列数据可能的取值

## 1. 指出有多少列

# 笛卡尔积与关系的例子

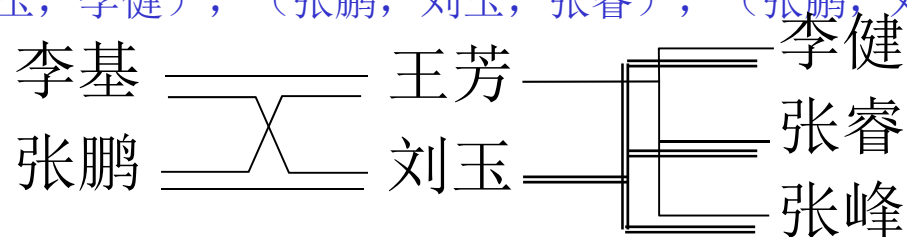
$D_1$  = 男人集合 (MAN) = {李基, 张鹏}

$D_2$  = 女人集合 (WOMAN) = {王芳, 刘玉}

$D_3$  = 儿童集合 (CHILD) = {李健, 张睿, 张峰}

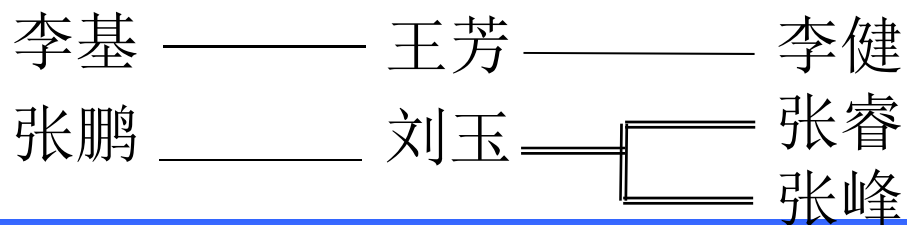
$D_1 \times D_2 \times D_3 = \{ \text{所有组合} \} =$

{ (李基, 王芳, 李健), (李基, 王芳, 张睿), (李基, 王芳, 张峰),  
 (李基, 刘玉, 李健), (李基, 刘玉, 张睿), (李基, 刘玉, 张峰),  
 (张鹏, 王芳, 李健), (张鹏, 王芳, 张睿), (张鹏, 王芳, 张峰),  
 (张鹏, 刘玉, 李健), (张鹏, 刘玉, 张睿), (张鹏, 刘玉, 张峰) }



家庭关系表(丈夫,妻子,孩子) = { 有意义的组合 } =

{ (李基, 王芳, 李健), (张鹏, 刘玉, 张睿), (张鹏, 刘玉, 张峰) }

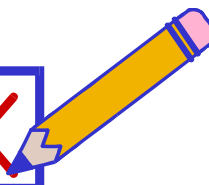
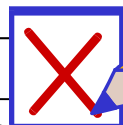


# 关系的性质

- 列是同质的 (Homogeneous)，即每一列中的分量是同一类型数据，来自同一个域
- 不同的列可出自同一个域，每一列称为属性，要给予不同的属性名
- 列的顺序可以任意交换，行的顺序也可以任意交换
- 任意两个元组不能完全相同
- 每一分量必须是不可分的数据项

班级	课程	学号	姓名	成绩
981101	数据库	01	张三	100
981101	数据库	02	张四	90
981101	数据库	03	张五	80
981101	计算机	01	张三	89
981101	计算机	02	张四	98
981101	计算机	03	张五	72
981102	数据库	01	王三	30
981102	数据库	02	王四	90
981102	数据库	03	王武	78

丈夫	妻子	孩子	
		第一个	第二个
李基	王芳	李健	张峰
张鹏	刘玉	张睿	



# 关系代数之操作一

## 并: $\cup$

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

R=9811班学生

学号	姓名	年龄	住址
981101	李四	22	3010
981102	李五	22	3010
981103	李三	21	3011
981104	李一	22	3011
981105	李六	22	3011

S=9812班学生

学号	姓名	年龄	住址
981201	王四	22	3010
981202	王五	22	3010
981203	王三	21	3011
981204	王一	19	3011
981205	王六	21	3010
981206	王二	19	3012
981207	王七	19	3012

$R \cup S$  = 9811和9812班学生

学号	姓名	年龄	住址
981101	李四	22	3010
981102	李五	22	3010
981103	李三	21	3011
981104	李一	22	3011
981105	李六	22	3011
981201	王四	22	3010
981202	王五	22	3010
981203	王三	21	3011
981204	王一	19	3011
981205	王六	21	3010
981206	王二	19	3012
981207	王七	19	3012

R和S合并



# 关系代数之操作二

差: -

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

R=9811班学生

学号	姓名	年龄	住址
981101	李四	22	3010
981102	李五	22	3010
981103	李三	21	3011
981104	李一	22	3011
981105	李六	22	3011

R - S = 是9811班但不是校运动队的学生

学号	姓名	年龄	住址
981102	李五	22	3010
981103	李三	21	3011
981105	李六	22	3011

S=校各种运动队的学生

学号	姓名	年龄	住址
981101	李四	22	3010
981104	李一	22	3011
981204	王一	19	3011
981205	王六	21	3010
981207	王七	19	3012

从R中把属于S的拿掉

# 关系代数之操作三

## 积: $\times$

$$R \times S = \{ (a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_m) \mid (a_1, a_2, \dots, a_n) \in R \wedge (b_1, b_2, \dots, b_m) \in S \}$$

R=9811班学生

学号	姓名	年龄	住址
981101	李四	22	3010
981103	李三	21	3011
981105	李六	22	3011

$R \times S = 9811班及其可能选课$

学号	姓名	年龄	住址	课程号	课程名	教师	学时
981101	李四	22	3010	C1	计算机	教1	52
981101	李四	22	3010	C2	物理	教2	36
981101	李四	22	3010	C3	高数	教5	40
981103	李三	21	3011	C1	计算机	教1	52
981103	李三	21	3011	C2	物理	教2	36
981103	李三	21	3011	C3	高数	教5	40
981105	李六	22	3011	C1	计算机	教1	52
981105	李六	22	3011	C2	物理	教2	36
981105	李六	22	3011	C3	高数	教5	40

S=课程表

课程号	课程名	教师	学时
C1	计算机	教1	52
C2	物理	教2	36
C3	高数	教5	40

所有组合

# 关系代数之操作四

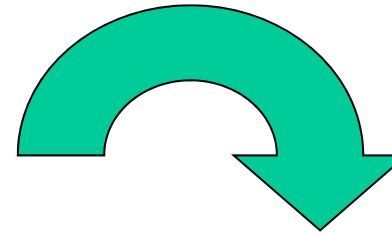
## 选择: $\sigma_F$

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

R=9812班学生

学号	姓名	年龄	住址
981201	王四	22	3010
981202	王五	22	3010
981203	王三	21	3011
981204	王一	19	3011
981205	王六	21	3010
981206	王二	19	3012
981207	王七	19	3012

把满足条件F的行选出来/只要某些行



$\sigma_{\text{年龄} < 22}(R)$

学号	姓名	年龄	住址
981203	王三	21	3011
981204	王一	19	3011
981205	王六	21	3010
981206	王二	19	3012
981207	王七	19	3012

$\sigma_{\text{年龄} > 19 \text{ OR } \text{住址} = 3012}(R)$

学号	姓名	年龄	住址
981201	王四	22	3010
981202	王五	22	3010
981203	王三	21	3011
981205	王六	21	3010
981206	王二	19	3012
981207	王七	19	3012

$\sigma_{\text{年龄} < 22 \text{ AND } \text{住址} = 3011}(R)$

学号	姓名	年龄	住址
981203	王三	21	3011
981204	王一	19	3011

# 关系代数之操作五

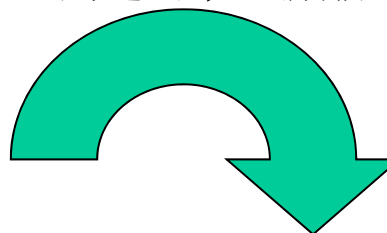
投影:  $\pi_{A_1, \dots, A_n}$

$$\pi_{A_1, \dots, A_n}(R) = \{ t[A_1, \dots, A_n] \mid t \in R \}$$

R=9812班学生

学号	姓名	年龄	住址
981201	王四	22	3010
981202	王五	22	3010
981203	王三	21	3011
981204	王一	19	3011
981205	王六	21	3010
981206	王二	19	3012
981207	王七	19	3012

把某些列选出来重新排列/只要某些列



$\pi_{\text{姓名, 学号}}(R)$

姓名	学号		
王四	981201		
王五	981202		
王三	981203		
王一	981204		
王六	981205		
王二	981206		
王七	981207		

$\pi_{\text{学号, 住址}}(R)$

学号	住址		
981201	3010		
981202	3010		
981203	3011		
981204	3011		
981205	3010		
981206	3012		
981207	3012		

$\pi_{\text{学号, 年龄, 住址}}(R)$

学号	年龄	住址	
981201	22	3010	
981202	22	3010	
981203	21	3011	
981204	19	3011	
981205	21	3010	
981206	19	3012	
981207	19	3012	

# 关系代数之操作六

交:  $\cap$

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

R=9811班学生

学号	姓名	年龄	住址
981101	李四	22	3010
981102	李五	22	3010
981103	李三	21	3011
981104	李一	22	3011
981105	李六	22	3011

$R \cap S$  = 是9811班同时是校运动队的学生

学号	姓名	年龄	住址
981101	李四	22	3010
981104	李一	22	3011

S=校各种运动队的学生

学号	姓名	年龄	住址
981101	李四	22	3010
981104	李一	22	3011
981204	王一	19	3011
981205	王六	21	3010
981207	王七	19	3012

# 关系代数之操作七

## 自然连接: $\bowtie$

$$R \bowtie S = \{ \langle t_r', t_s' \rangle \mid t_r \in R \wedge t_s \in S \wedge t_r[B] = t_s[B] \}$$

说明:  $\langle t_r', t_s' \rangle$  为串接  $t_r, t_s$  两个元组所组成的新元组后去掉一组重复属性B后形成的新元组

R=9811班学生

学号	姓名	年龄	住址
981101	李四	22	3010
981103	李三	21	3011
981105	李六	22	3011

S=选课表

学号	课程号	课程名	
981101	C1	计算机	教1
981105	C2	物理	教2
981101	C3	高数	教5
981101	C4	化学	教5
981103	C2	物理	教2

相同属性值  
相等的行连接

$R \bowtie S$

学号	姓名	年龄	住址	课程号	课程名		
981101	李四	22	3010	C1	计算机		
981101	李四	22	3010	C4	化学		
981101	李四	22	3010	C3	高数		
981103	李三	21	3011	C2	物理		
981105	李六	22	3011	C2	物理		

# 关系代数各种操作之组合

$$\pi_{\text{姓名,课程名}}(\sigma_{\text{课程号}='c2'}(R \bowtie S))$$

R=9811班学生

学号	姓名	年龄	住址
981101	李四	22	3010
981103	李三	21	3011
981105	李六	22	3011

S=选课表

学号	课程号	课程名	
981101	C1	计算机	教1
981105	C2	物理	教2
981101	C3	高数	教5
981101	C4	化学	教5
981103	C2	物理	教2

学号	姓名	年龄	住址	课程号	课程名		
981101	李四	22	3010	C1	计算机		
981101	李四	22	3010	C4	化学		
981101	李四	22	3010	C3	高数		
981103	李三	21	3011	C2	物理		
981105	李六	22	3011	C2	物理		

$$\sigma_{\text{课程号}='c2'}(R \bowtie S)$$

学号	姓名	年龄	住址	课程号	课程名		
981103	李三	21	3011	C2	物理		
981105	李六	22	3011	C2	物理		

$$\pi_{\text{姓名,课程名}}(\sigma_{\text{课程号}='c2'}(R \bowtie S))$$

姓名	课程名
李三	物理
李六	物理

# 码(关键字)和外码

R=9811班学生

学号	姓名	年龄	住址
981101	李四	22	3010
981103	李三	21	3011
981105	李六	22	3011

哪一列数据能够惟一区分每一行/或说没有重复, 哪一列便称为码

S=选课表

学号	课程号	课程名	
981101	C1	计算机	教1
981105	C2	物理	教2
981101	C3	高数	教5
981101	C4	化学	教5
981103	C2	物理	教2

在S中“学号”不是码, 但“学号”是另一关系R的码, 则“学号”称S的外码

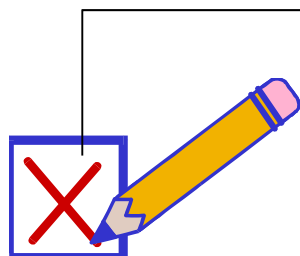


# 完整性约束一 实体完整性

码不能为空

R=9812班学生

学号	姓名	年龄	住址
981201	王四	22	3010
981202	王五	22	3010
981203	王三	21	3011
	王一	19	3011
981205	王六	21	3010
981206	王二	19	3012
981207	王七	19	3012



# 完整性约束二

## 参照完整性

S=选课表

外码 — 学号

学号	课程号	课程名	
981101	C1	计算机	教1
981105	C2	物理	教2
981101	C3	高数	教5
981101	C4	化学	教5
981103	C2	物理	教2

R=9811班学生

码 — 学号

学号	姓名	年龄	住址
981101	李四	22	3010
981103	李三	21	3011
981105	李六	22	3011

外码或者为空, 或者只能取另一关系中码的值

- 外码为空: 此值尚不能确定是哪一个
- 否则: 作为码的关系(R)中没有值, 则该外码的值也不存在.

# 完整性约束三

## 用户定义的完整性

用户可定义某些列的值必须满足的约束条件

R=9811班学生

学号	姓名	年龄	住址
981101	李四	22	3010
981103	李三	21	3011
981105	李六	22	3011

年龄 > 16    AND    年龄 < 30

学号只能是数字构成的

... ..

# 本章小结

---

1. 数据库系统的基本概念
2. 关系模型与关系数据库

# 本章内容概述

---

(1)数据库，数据库管理系统，数据库管理员

(2)关系，关系的七种操作

- 关系
- 并、交、差、笛卡儿积、选择、投影、连接、自然连接

# 本章的基本知识点

---

- 三组概念:

数据库、数据库管理系统、数据库管理员

- 一个构成两个作用

数据库系统的构成，数据库的作用，数据库管理员的作用

- 一个过程

数据库的操作过程(定义结构、增删改数据、查询统计)

- 一组定义七个操作一套语言

关系的定义；并/交/差/选择/投影/连接/自然连接；SQL语言

- 一个应用技能:

应用SQL语言进行数据查询统计分析的技能