



第二章 关系数据库

主要内容

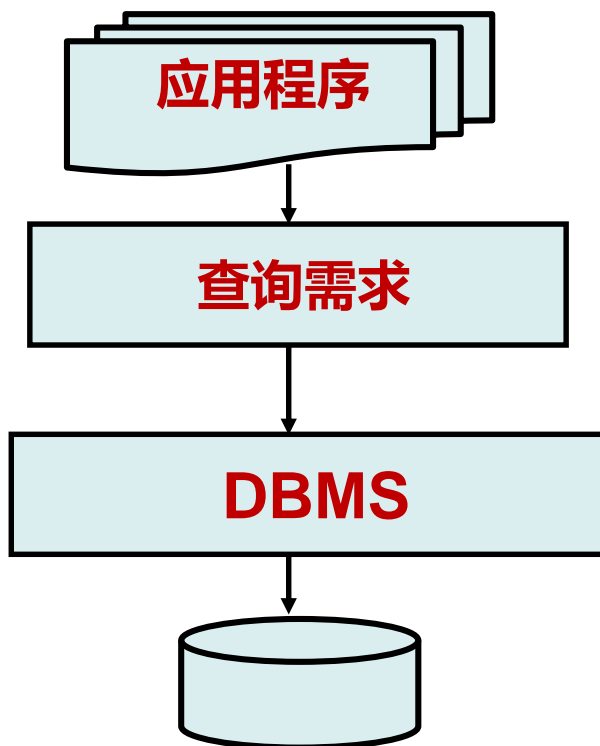
- 如何用离散数学中的代数系统来定义关系数据库？
(2.1、2.2、2.3内容)
- 如何定义代数运算，用数学语言描述查询需求？
(2.4、2.5内容分别介绍了代数语言和谓词语言)



关系代数系统的引入动机

问题：为什么需要代数语言？

关系代数产生动机



如何描述查询需求？

查询需求描述自动化
查询需求描述自然化

- 关系数据库是表集
- 定义完备的表操作
- 用表的运算表达式描述查询需求
- 系统解决表达式与查询语言的转化、执行

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



2.1 关系数据结构及形式化定义

- **动机** 能否将一个关系数据库用一个集合代数系统来表示？
- **作用** 应用对DB的查询、处理请求，能够转换成相应的关系操作表达式，再设计DBMS能够自动执行表达式。
- **思考** 关系代数语言虽然用户不友好，但形成一套理论，可以指导数据库研究
用户友好语言可以自动转换为关系代数语言，
例如 SQL



2.1 关系数据结构及形式化定义

2.1.1 关系

2.1.2 关系模式

2.1.3 关系数据库



2.1.1 关系

- 单一的数据结构----关系

现实世界的实体以及实体间的各种联系均用关系来表示

- 逻辑结构----二维表

从用户角度，关系模型中数据的逻辑结构是一张二维表

- 建立在集合代数的基础上



关系(续)

1. 域 (Domain)
2. 笛卡尔积 (Cartesian Product)
3. 关系 (Relation)



1. 域 (Domain)

- 域是一组具有相同数据类型的值的集合。例:

- 整数

- 实数

- 介于某个取值范围的整数

- 长度指定长度的字符串集合

- {‘男’ , ‘女’ }

-



2. 笛卡尔积 (Cartesian Product)

- 笛卡尔积

给定一组域 D_1, D_2, \dots, D_n , 这些域中可以有相同的。

D_1, D_2, \dots, D_n 的笛卡尔积为:

$$D_1 \times D_2 \times \dots \times D_n =$$

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

- 所有域的所有取值的一个组合
- 不能重复



例如，有三个域：

D1=导师集合 SUPERVISOR=张清玫，刘逸

D2=专业集合 SPECIALITY=计算机专业，信息专业

D3= 研究生集合 POSTGRADUATE=李勇，刘晨，王敏

则D1， D2， D3的笛卡尔积为

**D1XD2XD3={ (张清玫， 计算机专业， 李勇)， (张清玫， 计算机专业， 刘晨)，
(张清玫， 计算机专业， 王敏)， (张清玫， 信息专业， 李勇)，
(张清玫， 信息专业， 刘晨)， (张清玫， 信息专业， 王敏)，
(刘逸， 计算机专业， 李勇)， (刘逸， 计算机专业， 刘晨)，
(刘逸， 计算机专业， 王敏)， (刘逸， 信息专业， 李勇)，
(刘逸， 信息专业， 刘晨)， (刘逸， 信息专业， 王敏)}**




表 2.1 D_1 , D_2 , D_3 的笛卡尔积

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



笛卡尔积（续）

- 元组（Tuple）

- 笛卡尔积中每一个元素 (d_1, d_2, \dots, d_n) 叫作一个 n 元组（ n -tuple）或简称元组(Tuple)
- (张清玫, 计算机专业, 李勇)、(张清玫, 计算机专业, 刘晨)等都是元组

- 分量（Component）

- 笛卡尔积元素 (d_1, d_2, \dots, d_n) 中的每一个值 d_i 叫作一个分量
- 张清玫、计算机专业、李勇、刘晨等都是分量



笛卡尔积（续）

- 基数（Cardinal number）
 - 若 D_i ($i=1, 2, \dots, n$) 为有限集，其基数为 m_i ($i=1, 2, \dots, n$)，则 $D_1 \times D_2 \times \dots \times D_n$ 的基数 M 为：

$$M = \prod_{i=1}^n m_i$$



例

- $D_1 = \{\text{Zhang, Li, Wang}\}$, $D_2 = \{21, 22\}$, $D_3 = \{M, F\}$

- $D_1 \times D_2 \times D_3$ 的笛卡尔积为:

D1	D2	D3
Zhang	21	M
Zhang	21	F
Zhang	22	M
Zhang	22	F
Li	21	M
Li	21	F
Li	22	M
Li	22	F
Wang	21	M
Wang	21	F
Wang	22	M
Wang	22	F

3. 关系 (Relation)

1) 关系

$D_1 \times D_2 \times \dots \times D_n$ 的子集叫作在域 D_1, D_2, \dots, D_n 上的关系, 表示为

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

- R : 关系名
- n : 关系的目或度 (Degree)



例. 已知三个域

D1=导师集合S={张清玫,刘逸}

D2=专业集合SP={计算机,信息}

D3=学生集合P={李勇,刘晨,王敏}

计算思考

从笛卡尔集中取出一个子集，选择有意义的结果组成关系

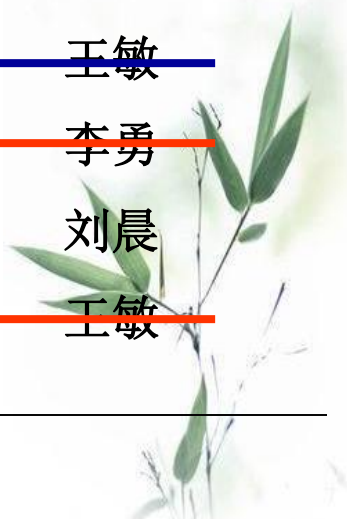
R=(导师，专业，研究生)

•一个研究生只能有一个专业，如李勇和王敏是计算机专业、刘晨是信息专业

现在导师与研究生是什么关系？

限定一个学生只能有一个导师，如张是计算机导师，刘是信息专业导师

导师S	专业SP	研究生P
张清玫	计算机	李勇
张清玫	计算机	刘晨
张清玫	计算机	王敏
张清玫	信息	李勇
张清玫	信息	刘晨
张清玫	信息	王敏
刘逸	计算机	李勇
刘逸	计算机	刘晨
刘逸	计算机	王敏
刘逸	信息	李勇
刘逸	信息	刘晨
刘逸	信息	王敏



关系（续）

关系的表示

关系也是一个二维表，表的每行对应一个元组，表的每列对应一个域

表 2.2 SAP 关系

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



关系（续）

2) 元组

关系中的每个元素是关系中的元组，通常用 t 表示。

3) 属性

- 关系中不同列可以对应相同的域
- 为了加以区分，必须对每列起一个名字，称为属性（Attribute）
- n 目关系必有 n 个属性



关系（续）

4) 码

- 候选码

设R是一个关系模式，K是R的子集。若对于R中任意两个不同的元组 $t_1 \neq t_2$, 有 $t_1[K] \neq t_2[K]$, 且K是这样的最小集，则称K是R的候选码

- 候选码可能有多个；
- 唯一确定元组；
- 最小化；

- 主码 若一个关系有多个候选码，则选定其中一个为主码
- 全码 关系模式的所有属性组是这个关系模式的候选码
- 主属性 候选码的诸属性称为主属性，不包含在任何侯选码中的属性称为非主属性或非码属性



候选码

候选码

外码

主码

学号	身份证号	姓名	性别	系别
0101	4201111985	张	男	CS
0102	36020211986	李	女	CS
0203	3671111985	赵	男	MA

系号	系名
CS	计算机
EN	英语
MA	数学系

主码

主码

学号	课号	成绩
0101	CS145	88
0101	CS148	90
0102	CS180	87
0203	CS145	78

课号	课名
CS145	数据库
CS148	操作系统
CS180	数据结构

全码关系

- 学校中某一门课程由多个教师讲授，他们使用相同的一套参考书。每个教员可以讲授多门课程，每种参考书可以供多门课程使用。

Teaching

课程C	教员T	参考书B
物理	李勇	普通物理学
物理	李勇	光学原理
物理	李勇	物理习题集
物理	王军	普通物理学
物理	王军	光学原理
物理	王军	物理习题集
数学	李勇	数学分析
数学	李勇	微分方程
数学	李勇	高等代数
数学	张平	数学分析
数学	张平	微分方程
数学	张平	高等代数
...

Teaching具有唯一候选码(C, T, B)，即全码



关系（续）

5) 三类关系

基本关系（基本表或基表table）

实际存在的表，是实际存储数据的逻辑表示

查询表

查询结果对应的表

视图表view

由基本表或其他视图表导出的表，是虚表，不对应实际存储的数据



关系（续）

6)基本关系的性质

- ① 列是同质的,即每一列中的分量是同一类型的数据, 来自同一个域
- ② 不同的列可出自同一个域
 - 其中的每一列称为一个属性
 - 不同的属性要给予不同的属性名
- ③ 列的顺序无所谓,, 列的次序可以任意交换
- ④ 任意两个元组的候选码不能相同
- ⑤ 行的顺序无所谓, 行的次序可以任意交换
- ⑥ 分量必须取原子值, 即每一个分量都必须是不可分的数据项。



关系性质1——同质的列

学号	姓名	性别	年龄	曾用名
9901	张三	男	20	张狗子
9902	李四	女	18	李朋
王五	9903	1	19	王麻子
9904	赵六	2	1981/2/28	赵薇



关系性质2—不同的属性名

姓名2
曾用名

学号	姓名	性别	年龄	姓名
9901	张三	男	20	张狗子
9902	李四	女	18	李朋
9903	王五	男	19	王麻子
9904	赵六	女	20	赵薇



关系性质3—属性无序

学号	姓名	性别	年龄	曾用名
9901	张三	男	20	张狗子
9902	李四	女	18	李朋
9903	王五	男	19	王麻子
9904	赵六	女	20	赵薇

学号	姓名	曾用名	性别	年龄
9901	张三	张狗子	男	20
9902	李四	李朋	女	18
9903	王五	王麻子	男	19
9904	赵六	赵薇	女	20

关系性质4—元组不重复

学号	姓名	性别	年龄	姓名
9901	张三	男	20	张狗子
9902	李四	女	18	李朋
9903	王五	男	19	王麻子
9904	赵六	女	20	赵薇
9901	张三	男	20	张狗子

重复的
元组!

×



关系性质5—元组无序

学号	姓名	性别	年龄
9901	张三	男	20
9902	李四	女	18
9903	王五	男	19
9904	赵六	女	20

学号	姓名	性别	年龄
9904	赵六	女	20
9901	张三	男	20
9903	王五	男	19
9902	李四	女	18



关系性质6—分量是原子

父	母	孩子	
		大	小
李男	丁女	李一	李二
王男	肖女	王一	

父	母	孩子
李男	丁女	李一 李二
王男	肖女	王一

非规范化关系

父	母	大孩	小孩
李男	丁女	李一	李二
王男	肖女	王一	

父	母	孩子
李男 李男 王男	丁女 丁女 肖女	李一 李二 王一

规范化关系



在许多实际关系数据库产品中，基本表并不完全具有这六条性质。

举例

- 列的顺序无所谓，次序可以任意交换
 - 遵循这一性质的数据库产品(如**ORACLE**)，增加新属性时，永远是插至最后一列
 - 但也有许多关系数据库产品没有遵循这一性质，例如**FoxPro**仍然区分了属性顺序
- 任意两个元组不能完全相同，由笛卡尔积的性质决定
 - 但许多关系数据库产品没有遵循这一性质，例如**Oracle**，**FoxPro**等都允许关系表中存在两个完全相同的元组，除非用户特别定义了相应的约束条件



在许多实际关系数据库产品中，基本表并不完全具有这六条性质。

举例

- 行的顺序无所谓，行的次序可以任意交换
 - 遵循这一性质的数据库产品(如**ORACLE**)，插入一个元组时永远插至最后一行
 - 但也有许多关系数据库产品没有遵循这一性质，例如**FoxPro**仍然区分了元组的顺序



2.1 关系数据结构

2.1.1 关系

2.1.2 关系模式

2.1.3 关系数据库



2.1.2 关系模式

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

- 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
 - 元组语义以及完整性约束条件
 - 属性间的数据依赖关系集合



2. 定义关系模式

关系模式可以形式化表示：

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

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



例：Student关系模式的定义

– Student(U,D,dom,F)

- $U=\{\text{sno}, \text{name}, \text{age}\}$

- $D=\{\text{CHAR}, \text{INT}\}$

- $\text{Dom}=\{\text{dom}(\text{sno})=\text{dom}(\text{name})=\text{CHAR}, \text{dom}(\text{age})=\text{INT}\}$

- $F=\{\text{sno} \rightarrow \text{name}, \text{sno} \rightarrow \text{age}\}$

- 关系模式通常简写为 $R(U)$ ，或 $R(A_1, A_2, \dots, A_n)$



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



3. 关系模式与关系

- 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
 - 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

2.1.1 关系

2.1.2 关系模式

2.1.3 关系数据库



2.1.3 关系数据库

- 关系数据库
 - 在一个给定的应用领域中，所有实体及实体间联系的关系的集合构成一个关系数据库。



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



2.2.1基本关系操作

- 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
 - 选择、投影、并、差、笛卡尔积是5种基本操作
- 关系操作的特点
 - 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



2.2.2 关系数据库语言的分类

- 关系代数语言
 - 用对关系的运算来表达查询要求

- 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

域关系演算语言

- 谓词变元的基本对象是域变量



关系数据库语言的分类

