

本讲内容

- ✿ 数据库发展
- ✿ 数据管理技术
- ✿ 关系数据库体系结构
- ✿ 数据模型
- ✿ 关系代数运算



数据库技术的发展

✿ 数据库技术的诞生以 20 世纪 60 年代 IBM 推出的数据库管理产品 IMS 为标志。



三代数据库系统

✿ 第一代的层次、网状数据库系统

层次模型：节点 网状模型

✿ 第二代数据库系统是关系数据库系统（二维表）

✿ 第三代数据库系统是 80 年代开始的面向对象的数据库系统，也称新一代数据库系统，其发展方向包括：

- ✿ 一是建立纯粹的面向对象数据库管理系统，主要是支持持久对象和实现数据共享
- ✿ 第二种实现途径是在传统的关系数据库基础上增加面向对象的特性。



数据管理技术的发展

✿ 人工管理阶段

✿ 文件系统阶段

✿ 数据库系统阶段



问与答

✿使用数据库系统进行数据管理有哪些优点？

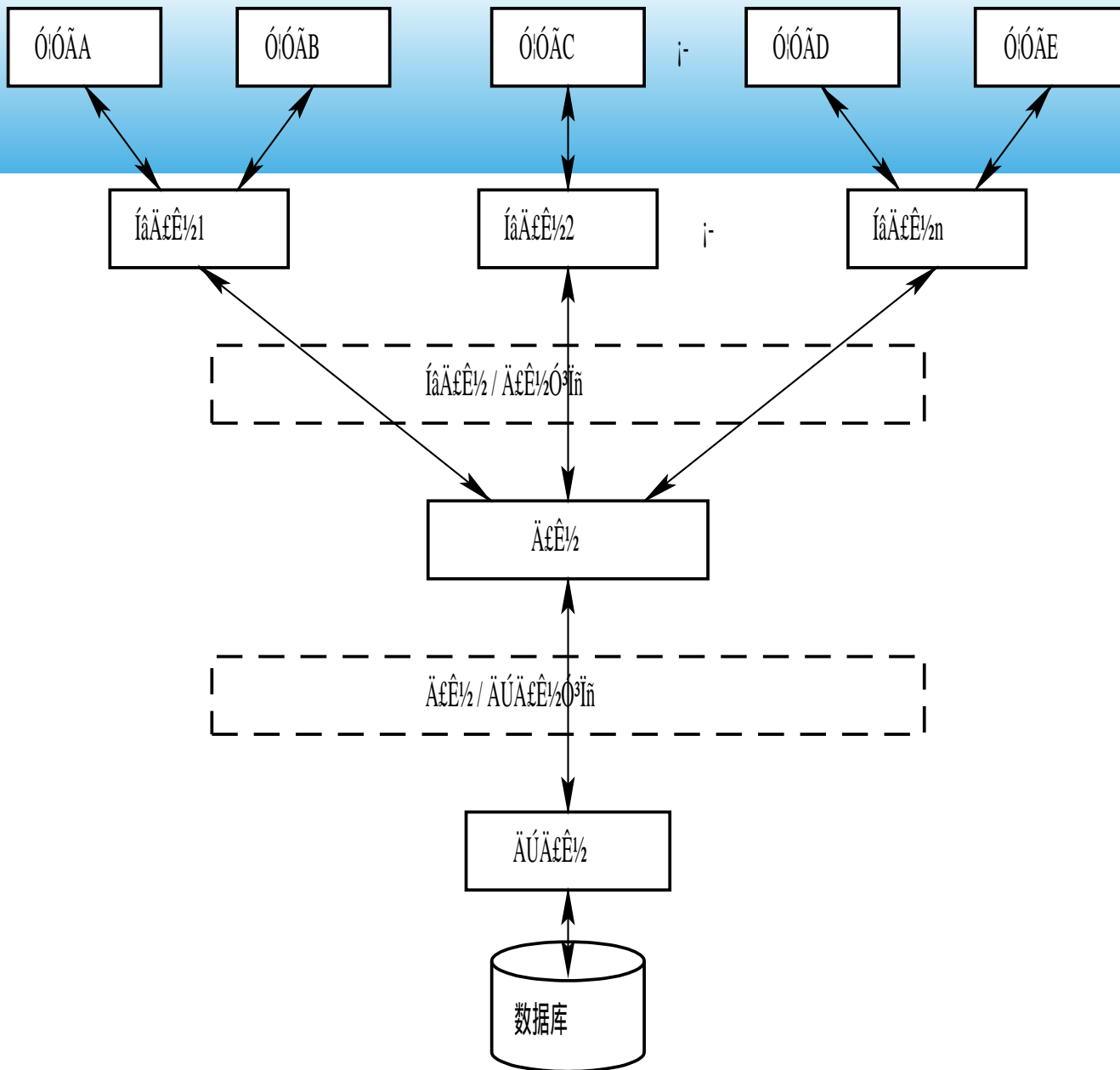
✿答：数据独立性，共享性和完整性



三级模式结构和两级映射

- ✿ 模式 (Schema) 是数据库中全体数据的逻辑结构和特征的描述，是 DBA 所看到的数据库，使用 DBMS 提供的模式数据定义语言 (模式 DDL) 来严格定义。
- ✿ 外模式也称子模式 (SubSchema) 或用户模式，它是数据库用户能够看见和使用的局部数据的逻辑结构和特征的描述
- ✿ 内模式又称存储模式，一个数据库只有一个内模式。它描述了数据的物理结构和存储方式，是数据库内部的表示方法
- ✿ 逻辑独立性与物理独立性。
 - ✿ 模式与外模式之间的独立性称为逻辑独立性
 - ✿ 模式与物理存储结构之间的独立性称为物理独立性





问与答

✿ 数据库三级模式划分的优点是什么？两级映射有什么作用？

✿ 答：数据库的三级模式将用户、逻辑数据库与物理数据库进行划分，使彼此之间的相互干扰减到最小。这使得底层的修改和变化尽量不影响到上层。

✿ 两级映像是指外模式 / 概念模式映像和概念模式 / 内模式映像，这样就带来了逻辑独立性和物理独立性。

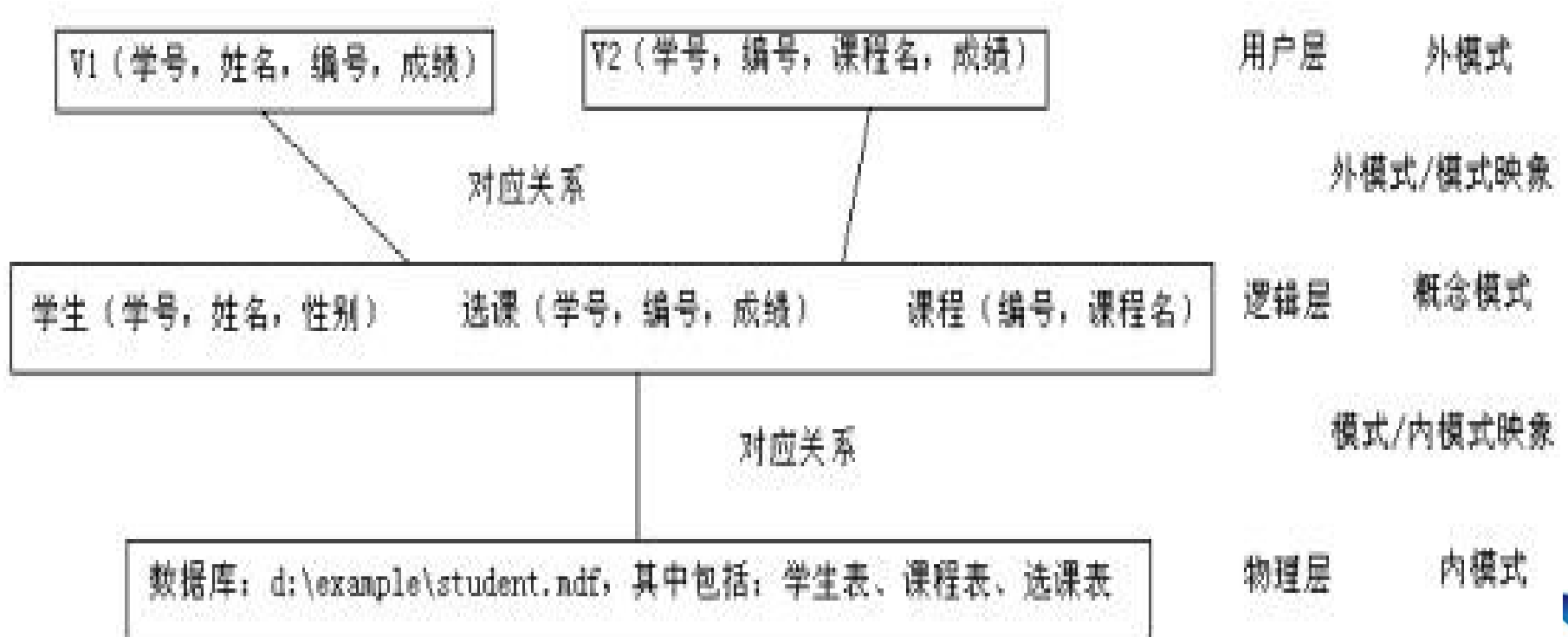
✿ 物理独立性是指内模式的修改只需对概念模式 / 内模式映像做相应修改，而不需修改概念模式，逻辑独立性是指概念模式的修改只需对外模式 / 概念模式映像做相应修改，而不需修改外模式



问与答

✿ 请举一个例子来具体说明模式、外模式与内模式？

✿ 答：参见下图，说明（略）



现实世界、信息世界、数据世界

- ✿现实世界：存在于人们头脑之外的客观世界
- ✿信息世界：是现实世界在人们头脑中的反映
- ✿数据世界：信息世界的信息的数据表示



三种主要的数据库模型

✿ 网状模型

- * 每个节点可以有一个以上的节点无双亲。
- * 至少有一个节点有多于一个的双亲。

✿ 层次模型

- * 有且仅有一个节点无双亲，这个节点称为根节点。
- * 其他节点有且仅有一个双亲。

✿ 关系模型

- * 关系模型的基本组成是关系





广州大学华软软件学院

South China Institute of Software Engineering.GU

关系代数



主要内容

- ✿ 基于传统集合理论的关系运算
- ✿ 关系代数特有的关系运算
- ✿ 4 种非基本关系运算
- ✿ 关系代数运算在查询中的应用



一、笛卡儿积的数学定义

✿ 1、笛卡儿积的数学定义

✿ **定义 2.1** 设有属性 A_1, A_2, \dots, A_n 分别在值域 D_1, D_2, \dots, D_n 中取值，则这些值域的笛卡儿积定义为：

✿ $D_1 \times D_2 \times \dots \times D_n = \{(d_1, d_2, \dots, d_n) | d_j \in D_j, j=1, 2, \dots, n\}$

✿ 其中，每个元素 (d_1, d_2, \dots, d_n) 称为**元组**。

✿ 元组中的第 j 个值 d_j 称为元组的第 j 个**分量**。

✿ 若 $D_j (j=1, 2, \dots, n)$ 为有限集，且其基数为 $m_j (j=1, 2, \dots, n)$ ，则
$$m = \prod_{j=1}^n m_j$$

笛卡儿积 $D_1 \times D_2 \times \dots \times D_n$ 的基数为

注：笛卡儿积的基数即为笛卡儿积定义的元组集合中元组的个数。



一、笛卡儿积的数学定义

✿ 例 1：设 $D_1=\{1,2,3\}$ ，基数为 3；

✿ $D_2=\{a,b\}$ ，基数为 2； 则有：

✿ $D_1 \times D_2 = \{(1,a), (1,b), (2,a), (2,b), (3,a), (3,b)\}$

✿ 基数为 $3 \times 2 = 6$ 。



一、笛卡儿积的数学定义

✿ 例 2：设 $D_1 = \{ \text{李兵, 王芳} \}$, $D_2 = \{ \text{男, 女} \}$,

✿ $D_3 = \{ \text{北京, 上海} \}$ 。

✿ $D_1 \times D_2 \times D_3 = \{ (\text{李兵, 男, 北京}), (\text{李兵, 男, 上海}),$

✿ $(\text{李兵, 女, 北京}), (\text{李兵, 女, 上海}),$

✿ $(\text{王芳, 男, 北京}), (\text{王芳, 男, 上海}),$

✿ $(\text{王芳, 女, 北京}), (\text{王芳, 女, 上海}) \}$

✿ 基数为 $2 \times 2 \times 2 = 8$ 。



一、笛卡儿积的数学定义

✿可见，笛卡儿积实际上是一个二维表，结果共有8个元组。



姓名 (D_1)	性别 (D_2)	籍贯 (D_3)
李兵	男	北京
李兵	男	上海
李兵	女	北京
李兵	女	上海
王芳	男	北京
王芳	男	上海
王芳	女	北京
王芳	女	上海



二、关系的数学定义

✿ 2、关系的数学定义

✿ **定义 2.2** 笛卡儿积 $D_1 \times D_2 \times \dots \times D_n$ 的任一子集 R 是 D_1, D_2, \dots, D_n 上的关系。

$n=1$ ，一元关系
 $n=2$ ，二元关系
度为 n 时， n 元关系

✿ 值域集合 D_1, D_2, \dots, D_n 是关系中元组的取值范围，称为关系的域（Domain）， n 称为关系的目或度（Degree）。

姓名	性别	籍贯
李兵	男	北京
王芳	女	上海

有意义！

姓名	性别	籍贯
李兵	男	上海
王芳	女	北京



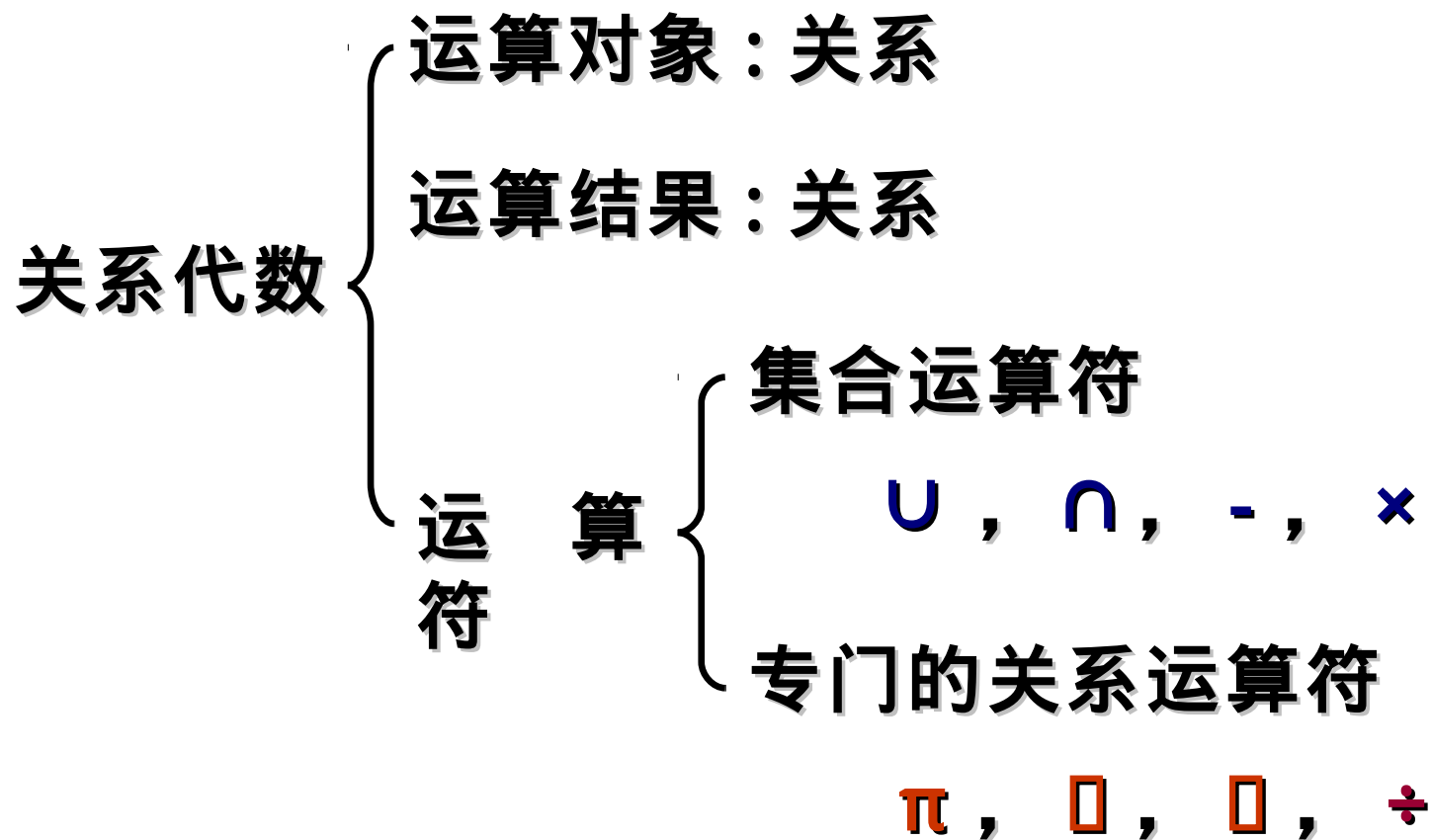
三、关系的性质

关系是一种规范化了的二维表。

✿ 3、关系的性质

- ✿ (1) 关系中的每个属性值都是不可再分的数据单位，即关系表中不能再有子表；
- ✿ (2) 关系中任意两行不能完全相同，即关系中不允许出现相同的元组；
- ✿ (3) 关系是一个元组的集合，所以关系中元组间的顺序可以任意；
- ✿ (4) 关系中的属性是无序的，使用时一般按习惯排列各列的顺序；
- ✿ (5) 每一个关系都有一个主键，唯一地标识它的各个元组。



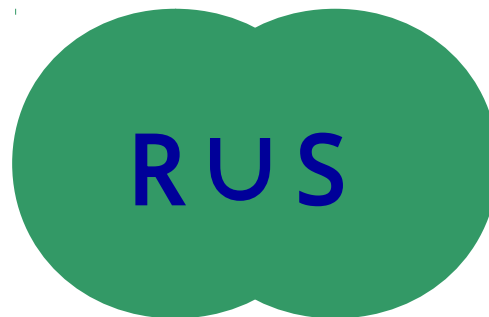
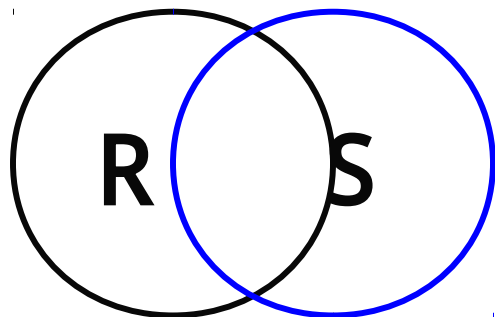


一、基于传统集合理论的关系运算

✿1、并

① 前提：关系 R 和 S 具有相同的模式

② 定义： $R \cup S \equiv \{ t | t \in R \vee t \in S \}$



③ 运算过程：

将关系 R 和 S 的元组放在一起，然后消去重复的元组。



一、基于传统集合理论的关系运算

✿ 1、并

④ 示例：求 $R1 \cup R2$

A	B	C
---	---	---

a	b	c
---	---	---

a	d	e
---	---	---

f	d	c
---	---	---

关系 R1

A	B	C
---	---	---

a	d	e
---	---	---

f	d	c
---	---	---

f	d	e
---	---	---

关系 R2

$R1 \cup R2$

A	B	C
---	---	---

a	b	c
---	---	---

a	d	e
---	---	---

f	d	c
---	---	---

f	d	e
---	---	---

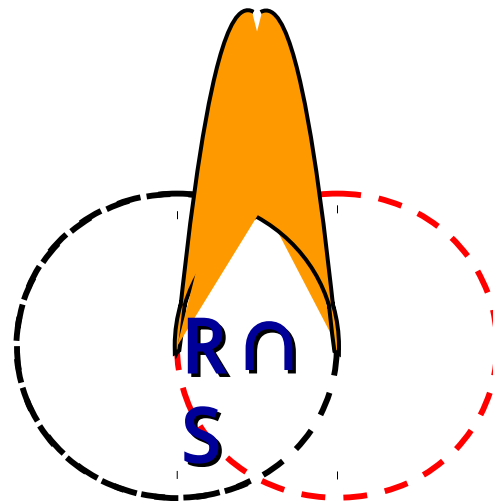
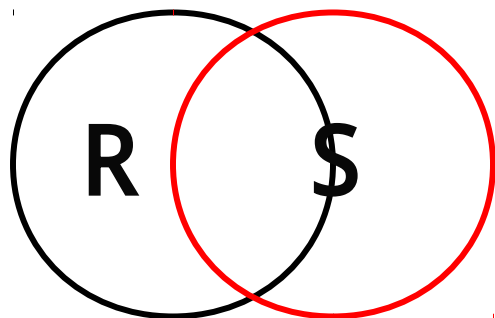
⑤ **适用于**：找出所有出现在两个关系之一的或同时出现在两个关系中的元组。



一、基于传统集合理论的关系运算

✿ 2、交

- ① 前提：关系 R 和 S 具有相同的模式
- ② 定义： $R \cap S \equiv \{t | t \in R \wedge t \in S\}$



③ 运算过程：

找出同时存在于关系 R 和 S 中的所有相同的元组

。



一、基于传统集合理论的关系运算

✿ 2、交

④ 示例： 求 $R1 \cap R2$

A	B	C
a	b	c
a	d	e
f	d	c

关系 R1

A	B	C
a	d	e
f	d	c
f	d	e

关系 R2

A	B	C
a	d	e
f	d	c

$R1 \cap R2$

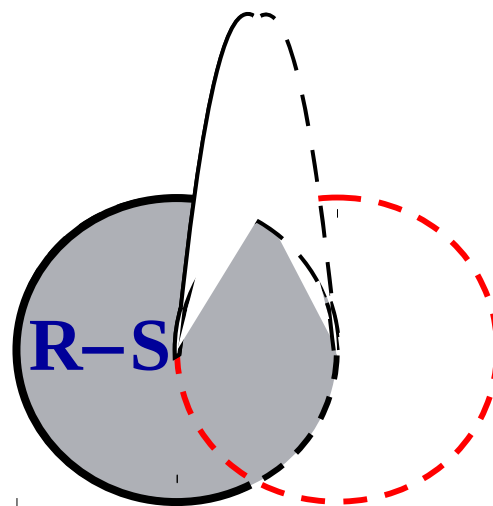
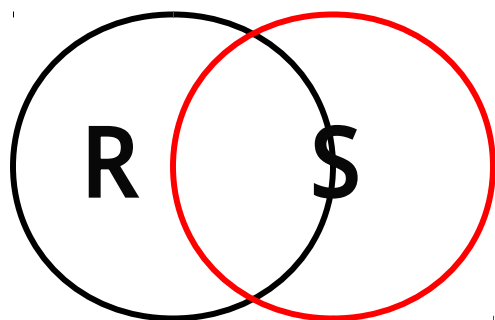
⑤ **适用于：** 需要找出所有同时出现在两个关系中的元组时。



一、基于传统集合理论的关系运算

✿ 3、差

- ① 前提：关系 R 和 S 具有相同的模式
- ② 定义： $R - S \equiv \{ t | t \in R \wedge t \notin S \}$



③ 运算过程：

从关系 R 的元组中去除它与关系 S 相同的那些元组



一、基于传统集合理论的关系运算

✿ 3、差

④ 示例： 求 $R1 - R2$

A	B	C
a	b	c
a	d	e
f	d	c

关系 R1

A	B	C
a	d	e
f	d	c
f	d	e

关系 R2

A	B	C
a	b	c

$R1 - R2$

⑤ **适用于：**找出在一个关系中而不在另一个关系中的那些元组。



一、基于传统集合理论的关系运算

✿ 4、广义笛卡儿积

✿ ①前提：关系 R 和 S 的目数分别为 r 和 s。

✿ ②定义：

✿
$$R \times S \equiv \{ t \mid t = \langle t^r, t^s \rangle \wedge t^r \in R \wedge t^s \in S \}$$

③运算过程：

用 R 的第 i (i=1, 2, ..., m) 个元组与 S 的全部元组结合成 n 个元组，所以 $R \times S$ 有 $m \times n$ 个元组。



一、基于传统集合理论的关系运算

✿ 4、广义笛卡儿积

④ 示例： 求 $R1 \times R3$

A	B	C
a	b	c
a	d	e
f	d	c

关系 R1

D	E
g	h
i	j

关系 R3

$R1 \times R3$

A	B	C	D	E
a	b	c	g	h
a	b	c	i	j
a	d	e	g	h
a	d	e	i	j
f	d	c	g	h
f	d	c	i	j



一、基于传统集合理论的关系运算

✿ 4、广义笛卡儿积

✿ ⑤ 命名机制

✿ 关系名. 属性名



A B C

a b c

a d e

f d c

关系 R1

C D

g h

i j

关系 R3

$R1 \times R3$

A B $R_1.C$ $R_3.C$ D

a b c g h

a b c i j

a d e g h

a d e i j

f d c g h

f d c i j

⑥ 适用于：

将任意两个关系的信息无条件组合在一起



二、关系代数特有的关系运算

✿ 1、投影

① 定义

设关系 R 为 r 目关系，其元组变量为 $t^r = (t_1, t_2, \dots, t_r)$ ，关系 R 在其属性分量 $A_{j_1}, A_{j_2}, \dots, A_{j_k} (k \leq r; j_1, j_2, \dots, j_k \text{ 为 } 1 \text{ 到 } r \text{ 之间互不相同的整数})$ 上的投影是一个 k 目关系

$$\pi_{j_1, j_2, \dots, j_k}(R) = \{ t \mid t = \underline{(t_{j_1}, t_{j_2}, \dots, t_{j_k})} \wedge \underline{(A_{j_1}, A_{j_2}, \dots, A_{j_k})} \in R \}$$



二、关系代数特有的关系运算

1、投影

② 运算过程：

首先按照 j_1, j_2, \dots, j_k 的顺序，从关系 R 中取出列序号为 j_1, j_2, \dots, j_k 的 k 列，然后除去结果中的重复元组，构成一个以 j_1, j_2, \dots, j_k 为顺序，以 $A_{j_1}, A_{j_2}, \dots, A_{j_k}$ 为属性的 k 目关系

- 投影是从列的角度进行的运算
- 投影的下标可是列序号，也可是列属性名



二、关系代数特有的关系运算

✿ 1、投影 列的角度进行计算

③ 示例： $\pi_{1,2} (R2)$

A	B	C
a	d	e
f	d	c
f	d	e

A	B
a	d
f	d
f	d

A	B
a	d
f	d

关系 R2

④ **适用于**：可以从某一关系中选出若干属性列构成新的关系，通常用于查询结果的输出



二、关系代数特有的关系运算

✿ 2、选择 从行的角度进行计算

① 定义： $\sigma_f(R) \equiv \{t \mid t \in R \wedge F(t)=\text{true}\}$

② 运算过程：

根据公式 F 对关系 R 作水平分割，从中挑选出满足公式 F 的那些元组组成新关系。

● 设 F 是一个命题公式

< 属性名 > < 比较操作符 > < 常量值 >

或 < 属性名 > < 比较操作符 > < 属性名 >

● 选择是从行的角度进行的运算，产生的关系具有与 R 相同的属性，但其元组数量总是小于或等于 R 中的元组数量



二、关系代数特有的关系运算

✿ 2、选择

③ 示例

关系 R1	A	B	C	$\sigma_{B='d' \wedge C='e'} (R1)$		
	A	B	C	A	B	C
	a	b	c			
	a	d	e	a	d	e
	f	d	c			

④ **适用于**：可以在某一关系中选择满足给定条件的诸元组构成新的关系



二、关系代数特有的关系运算

练习

A	B	C
a	b	c
a	d	e
f	d	c

C	A
c	a
e	a
c	f

关系
R1

$$\textcircled{1} \pi_{3, 1}(R1)$$



二、关系代数特有的关系运算

练习

A	B	C
a	d	e
f	d	c
f	d	e

关系
R2

A	B	C
a	d	e
f	d	e

A	B	C
f	d	c

$$\textcircled{2} \sigma_{2>3}(R2)$$



二、关系代数特有的关系运算

✿ 3、商

1) 定义:

设关系 R 和 S 的目数分别为 r 和 s ，且 $r > s$ ， $s \neq \varnothing$ ，则关系 R 关于 S 的商是一个由 $r-s$ 目元组组成的集合，且如果 $t^{r-s} \in \pi_{1,2,\dots,r-s}(R)$ ，则 t^{r-s} 与 S 中的每一个元组 u^s 组成的新元组 $\langle t^{r-s}, u^s \rangle$ 必在关系 R 中。

关系 R 关于 S 的商记为 $R \div S$ ：

$$R \div S = \{t \mid t = (t_1^r, t_2^r, \dots, t_{r-s}^r) \wedge \text{“ 如果 } t^{r-s} \in \pi_{1,2,\dots,r-s}(R), \text{ 则对于所有的 } u^s \in S, \text{ 成立 } \langle t^{r-s}, u^s \rangle \in R \text{”}\}$$



二、关系代数特有的关系运算

✿ 3、商

2) 运算过程

- ① 计算 $\pi_{1,2,\dots,r-s}(R)$;
- ② 对于 $\pi_{1,2,\dots,r-s}(R)$ 中每一个元组 t^{r-s} 和所有 $u^s \in S$, 如果均有 $\langle t^{r-s}, u^s \rangle \in R$ 成立, 则 t^{r-s} 属于结果关系 $R \div S$ 中的元组; 否则, 不属于 $R \div S$ 中的元组。



二、关系代数特有的关系运算

3) 示例例 2.3 已知关系 R 和 S，求 $R \div S$

r-s									
A	B	C	D	C	D	A	B	A	B
a	b	c	d	c	d	a	b	a	b
a	b	e	f	e	f	b	c	e	d
b	c	e	f			e	d		
e	d	c	d						
e	d	e	f						
a	b	d	e						

关系 R

关系 S

$\pi_{A,B}(R)$

$R \div S$

- ① 计算 $\pi_{1,2,\dots,r-s}(R)$;
- ② 验证 $\langle t^{r-s}, u^s \rangle \in R$ 成立。



二、关系代数特有的关系运算

4、联接

- 1) 设关系 R 和 S 的目数分别为 r 和 s , θ 是算术比较运算符, 则联接运算定义为:

$$R \bowtie_{j\theta k} S = \{t \mid t = (t^r, t^s) \wedge \underline{t^r} \in R \wedge \underline{t^s} \in S \wedge \underline{t_j^r \theta t_k^s}\}$$

2) 运算过程

将 R 的每个元组的第 j 个分量与 S 的每个元组的第 k 个分量做 θ 比较运算, 当满足比较条件时, 就把 S 的该分量所在元组接在 R 的相应元组的右边构成一个新关系的元组; 当不满足比较条件时, 继续下一次比较, 直到关系 R 和 S 中的元组都比较完为止。



二、关系代数特有的关系运算

4、联接

3) 说明

$$\textcircled{1} R \underset{F}{\bowtie}, F = F_1 \wedge F_2 \wedge \dots \wedge F_m$$

② 等值联接: θ 为 “=”



二、关系代数特有的关系运算

4) 示例 2.4 已知关系 R 和 S，求 $R \bowtie_{B < D} S$ 或 $R \bowtie_{2 < 1} S$

	A	B	C
t^r	1	2	3
	4	5	6
	7	8	9

关系 R

	D	E
t^s	3	1
	6	2

关系 S

A	B	C	D	E
1	2	3	3	1
1	2	3	6	2
4	5	6	6	2

5) 适用于

有选择条件的多个关系的数据组合



二、关系代数特有的关系运算

✿ 5、自然联接

- 1) 设关系 R 和 S 的目数分别为 r 和 s , 且关系 R 和 S 的属性中有部分相同属性 A_1, A_2, \dots, A_k , 则自然联接定义为:

$$R \bowtie S = \{t \mid t = (\underline{t^r}, \overline{t^s}) \wedge \underline{t^r} \in R \wedge \overline{t^s} \in S \wedge \underline{R.A_1 = S.A_1} \wedge \dots \wedge \underline{R.A_k = S.A_k}\}$$

2) 运算过程

将 R 的每个元组的 A_1, A_2, \dots, A_k 列的值和关系 S 的每个元组的 A_1, A_2, \dots, A_k 列的值进行比较, 比较条件是 $R.A_1 = S.A_1 \wedge R.A_2 = S.A_2 \wedge \dots \wedge R.A_k = S.A_k$, 当条件都满足时, 就从关系 S 中正在比较的元组中去掉被比较的 k 个分量后, 把剩余的分量依原序接在关系 R 的元组的右边构成新关系的一个元组; 当至少有一个条件不满足时, 就继续下一次比



二、关系代数特有的关系运算

3) 示例 2.5 已知关系 R 和 S，求 $R \bowtie S$

R				S							
	A	B	C		B	C	D	A	B	C	D
a		b	c	b	b	c	d	a	b	c	d
d		b	c	b	b	c	e	a	b	c	e
b		b	f	a	a	d	b	d	b	c	d
c		a	d					d	b	c	e
								c	a	d	b



二、关系代数特有的关系运算

▲ 5、自然联接

✿ 4) 适用于：自然联接在关系数据库理论和实践中处于中心地位，它的使用非常方便。

✿ 5) 与等值联接的区别：

- 当两个关系 R 和 S 有相同属性时，自然联接与等值联接都是判断在相同属性上的值是否相等。但结果关系中，自然联接的公共属性只出现一次，而等值联接的公共属性则要重复出现；
- 当关系 R 和 S 无公共属性时， R 与 S 的自然联接即为 R 与 S 的广义笛卡儿积。



小结

