

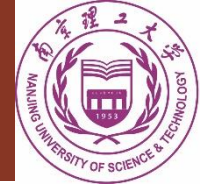
第10讲 怎样管理和利用数据(I)

用数据说话、用数据决策、用数据创新已形成社会的一种常态和共识，数据已被视为知识的来源、被认为是一种财富。

学习数据库相关知识，体验基于数据库的数据运用方法，理解基本的数据化思维方式

学生登记表

学号	姓名	性别	出生年月	入学日期	家庭住址
98110101	张三	男	1980.10	1998.09	黑龙江省哈尔滨市
98110102	张四	女	1980.04	1998.09	吉林省长春市
98110103	张五	男	1981.02	1998.09	黑龙江省齐齐哈尔市
98110201	王三	男	1980.06	1998.09	辽宁省沈阳市
98110202	王四	男	1979.01	1998.09	山东省青岛市
98110203	王武	女	1981.06	1998.09	河南省郑州市



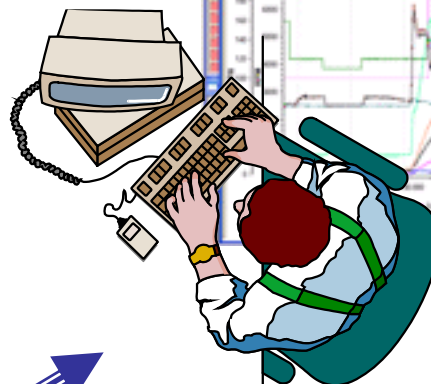
数据为什么要管理？

1. 数据为什么要管理

3/48



1.1 信息社会的工作方式？



信息社会：业务工作 + 计算机支持

- 网络/Internet
- 数据库

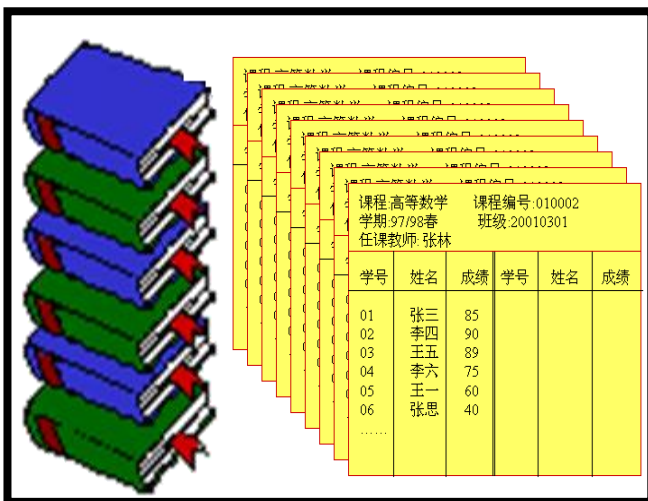
Everything Over DB



传统社会：业务工作

1.2 数据为什么要管理？

数据与数据库



数据

形成“库”，实现“积累”
应用“库”，实现积累的效益

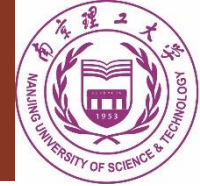
“库”的管理与控制

- 纸面数据 vs. 电子数据
- 单一数据文件 vs. 数据库
- 数据产生的分散化 vs. 数据应用的共享化
- 小规模数据 vs. 大规模数据

1. 数据为什么要管理

1.3 数据自有黄金屋？

5/48



大数据价值发现

2008年全球产生的数据量为0.49ZB(2^{50} MB)

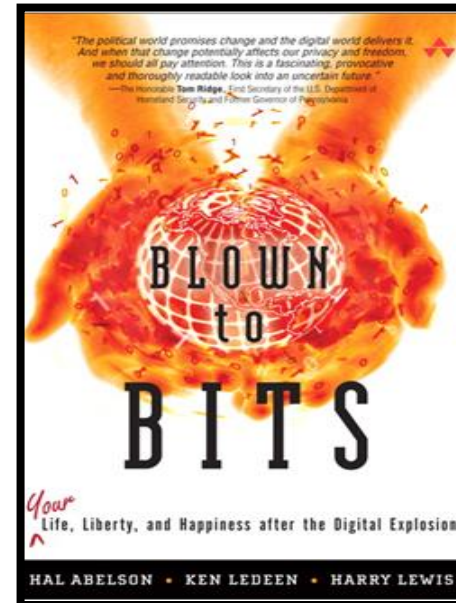
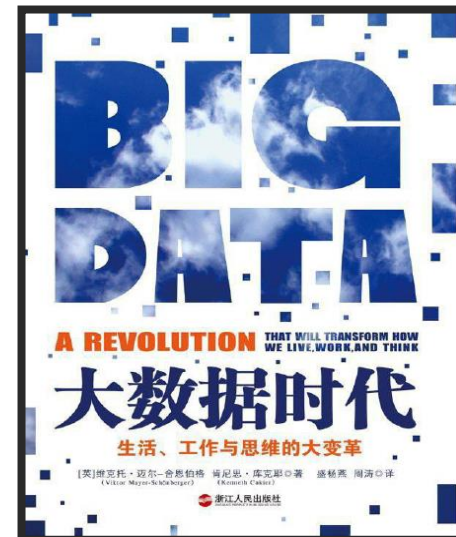
2009年的数据量为0.8ZB

2010年增长为1.2ZB

2011年的数量更是高达1.82ZB

2012年为止，人类所有印刷材料的数据量是200PB

预计到2020年，全世界的数据规模将达今天的44倍。



一个例子。大家乘坐飞机时都希望买到更便宜的机票，可能都相信“购买机票，越早预订越便宜”，果真然否？2003年Farecast公司创始人奥伦·埃齐奥尼(Oren Etzioni)提前几个月在网上订了一张机票，在飞机上与邻座若干乘客交谈时，他发现尽管很多人机票比他买的更晚，但票价却比他的便宜得多。出了什么问题？是航空公司或者网站有意“欺诈”，还是常识“购买机票，越早预订越便宜”？

Farecast: 飞机票价格预测

怎样判断机票价格是否合理呢？又该何时降价、什么原因降价，只有航空公司才知道。埃齐奥尼要做的就是仅仅依靠数据“特洛伊木马”来破解这个谜。他发现，如果一小时内会上涨还是下降，如果一小时内会下降还是上涨，机票的明智选择。反过来说，机票。

购票时机与机票价格的关系？

他开发了一个价格预测工具，它通过在一个航班网站上收集来的41天内价格波动产生12000条数据，并确定票价与提前购买天数的关系。它不知道是哪些因素导致了机票价格的波动，还是所谓的周六晚上不出门，它都不知道。只知通过利用其他航班的数据来预测未来机票价格的走势以及增减幅度，能帮助消费者抓住最佳时机。

只求关系，不求因果

为了建立一个行业机票预订数据库。有了这个数据库，系统就能在商业航空产业中，每一条航线上每一架飞机内的每一个座位一周内的综合票价记录中得出。如今，Farecast已经拥有惊人的约2000亿条飞行数据。Farecast系统预测的准确度，到2012年为止，Farecast票价预测平均每张机票可节省50美元。这项技术也可以延伸到其他领域，如宾馆预订、二手车购买等。只要这些领域内的产品差异不大，同时存在大幅度的价格差和大量可运用的数据，就都可以应用这项技术。

不要相信经验，一切以数据说话



什么是数据库与数据库系统？

2.1 数据库？

数据库：相关有关联关系的数据的集合

- 一个表聚集了具有相同结构类型的若干个对象
- 一行数据反映了某一对象的相关内容
- 一列数据具有相同的数据类型
- 表与表间也存在着相互关联

学生登记表

学号	姓名	性别	出生年月	入学日期	家庭住址
98110101	张三	男	1980.10	1998.09	黑龙江省哈尔滨市
98110102	张四	女	1980.04	1998.09	吉林省长春市
98110103	张五	男	1981.02	1998.09	黑龙江省齐齐哈尔市
98110201	王三	男	1980.06	1998.09	辽宁省沈阳市
98110202	王四	男	1979.01	1998.09	山东省青岛市
98110203	王武	女	1981.06	1998.09	河南省郑州市

数据库//Database

相互有关联关系
的数据的集合

学生成绩单

班级	课程	教师	学期	学号	姓名	成绩
981101	数据库	李四	98秋	98110101	张三	100
981101	数据库	李四	98秋	98110102	张四	90
981101	数据库	李四	98秋	98110103	张五	80
981101	计算机	李五	98秋	98110101	张三	89
981101	计算机	李五	98秋	98110102	张四	98
981101	计算机	李五	98秋	98110103	张五	72
981102	数据库	李四	99秋	98110201	王三	30
981102	数据库	李四	99秋	98110202	王四	90
981102	数据库	李四	99秋	98110203	王武	78

2. 什么是数据库与数据库系统

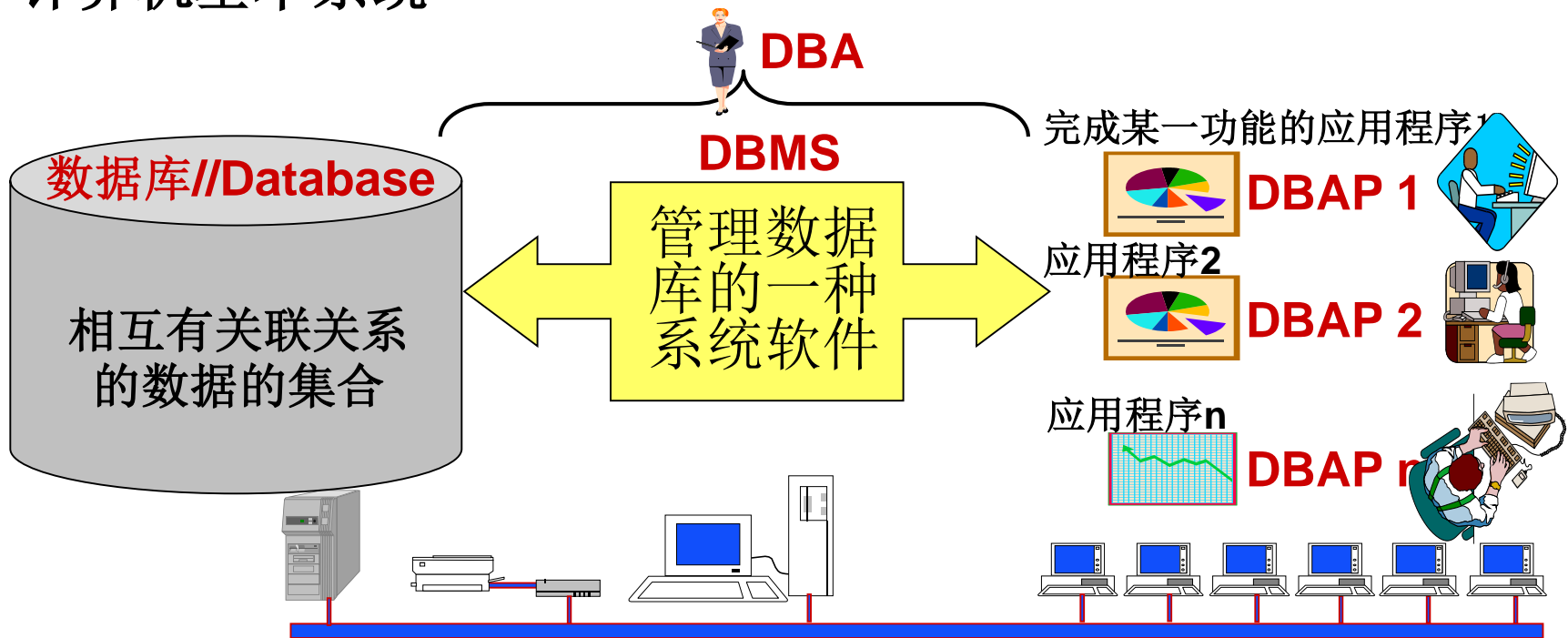
8/48



2.2 数据库系统的几个构成部分?

数据库系统(工作环境)

- 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



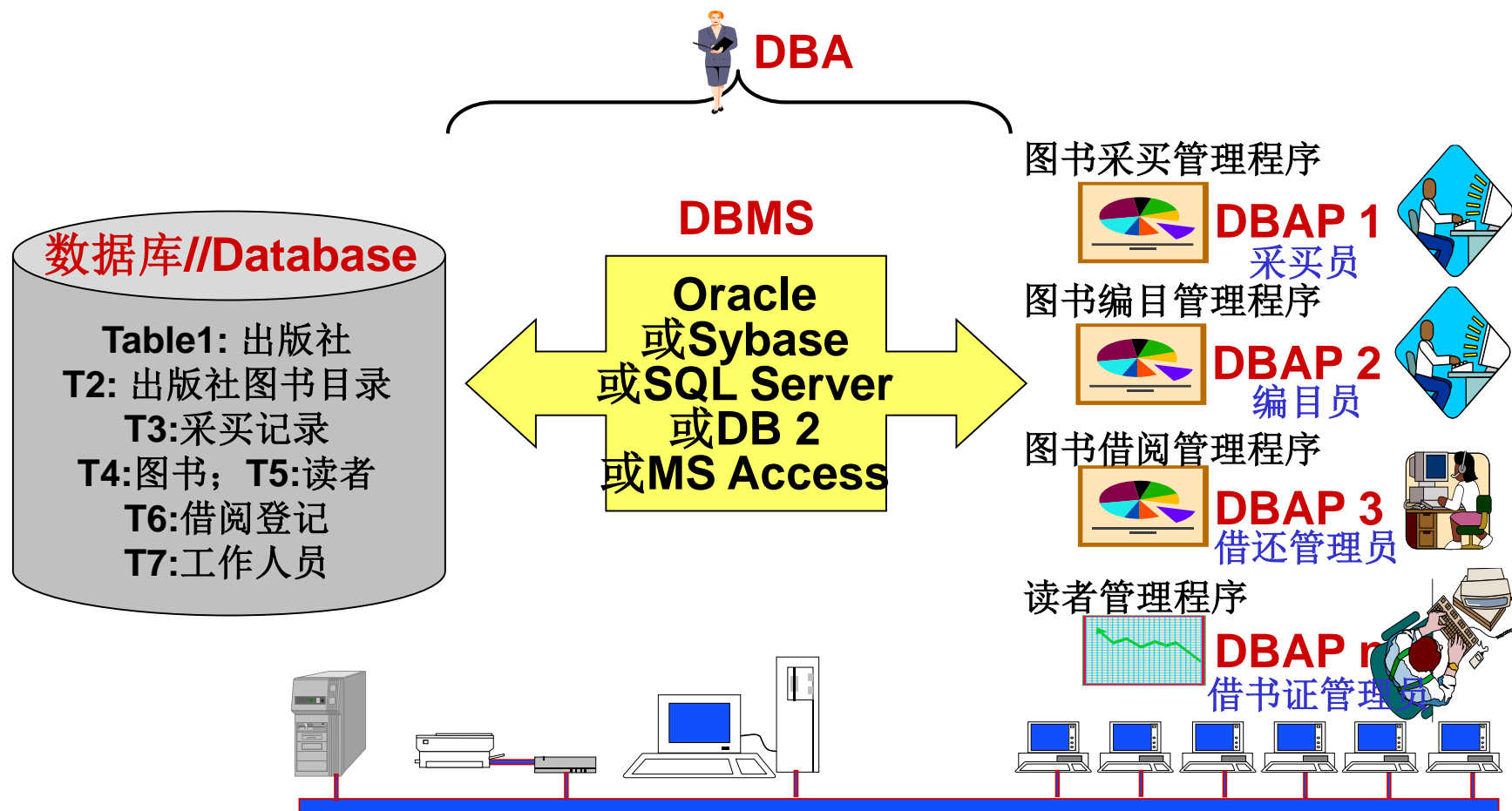
2. 什么是数据库与数据库系统

9/48



2.2 数据库系统的几个构成部分？

数据库系统(工作环境)示例：图书管理数据库系统

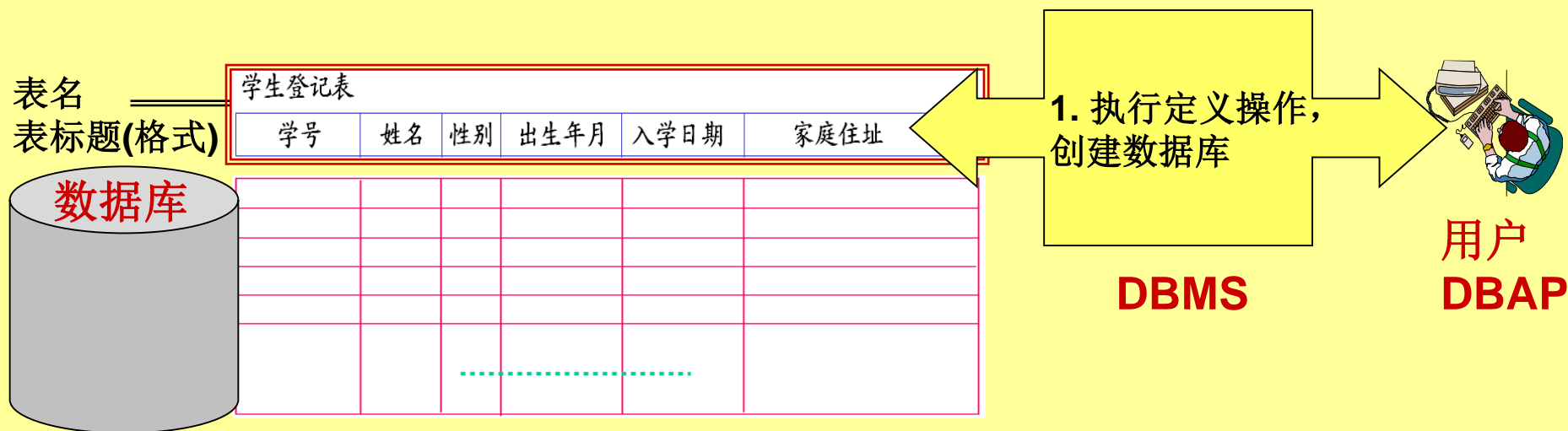




数据库管理系统的基本功能

➤数据库定义: 定义数据库中数据表的名称、标题(内含的属性名称及对该属性的值的值的要求)等。

- ❑ **DBMS提供一套数据定义语言(DDL:Data Definition Language)给用户**
- ❑ **用户使用DDL描述其所要建立表的格式**
- ❑ **DBMS依照用户的定义，创建数据库及其中的Table**

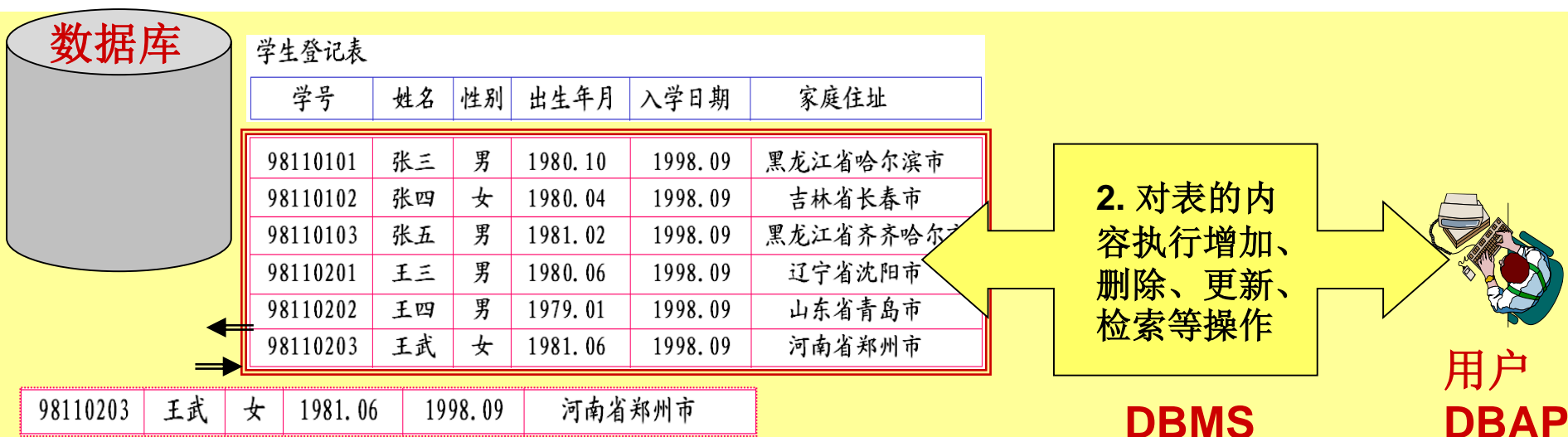


3.2 什么是数据库操纵?

数据库管理系统DBMS

➤ 数据库操纵: 向数据库的**Table**中增加/删除/更新数据及对数据进行查询、检索、统计等

- ❑ DBMS提供一套数据操纵语言(**DML: Data Manipulation Language**)给用户
- ❑ 用户使用**DML**描述其所要进行的增、删、改、查等操作
- ❑ DBMS依照用户的操作描述, 实际执行这些操作

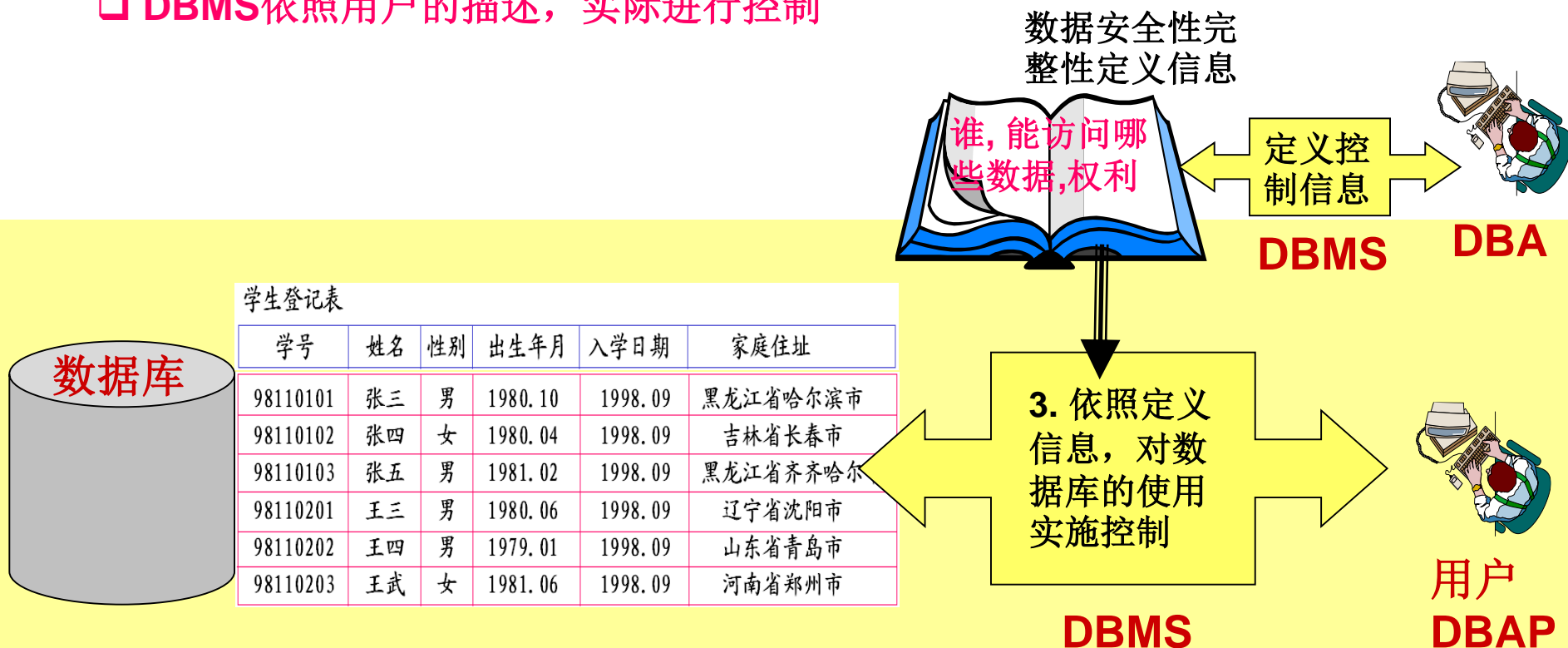


3.3 什么是数据库控制?

数据库管理系统DBMS

➤ 数据库控制: 控制数据库中数据的使用——哪些用户可以使用, 哪些不可以

- ❑ DBMS提供一套数据控制语言(DCL:Data Control Language)给用户
- ❑ 用户使用DCL描述其对数据库所要实施的控制
- ❑ DBMS依照用户的描述, 实际进行控制



3.4 数据库维护要做什么？

数据库管理系统DBMS

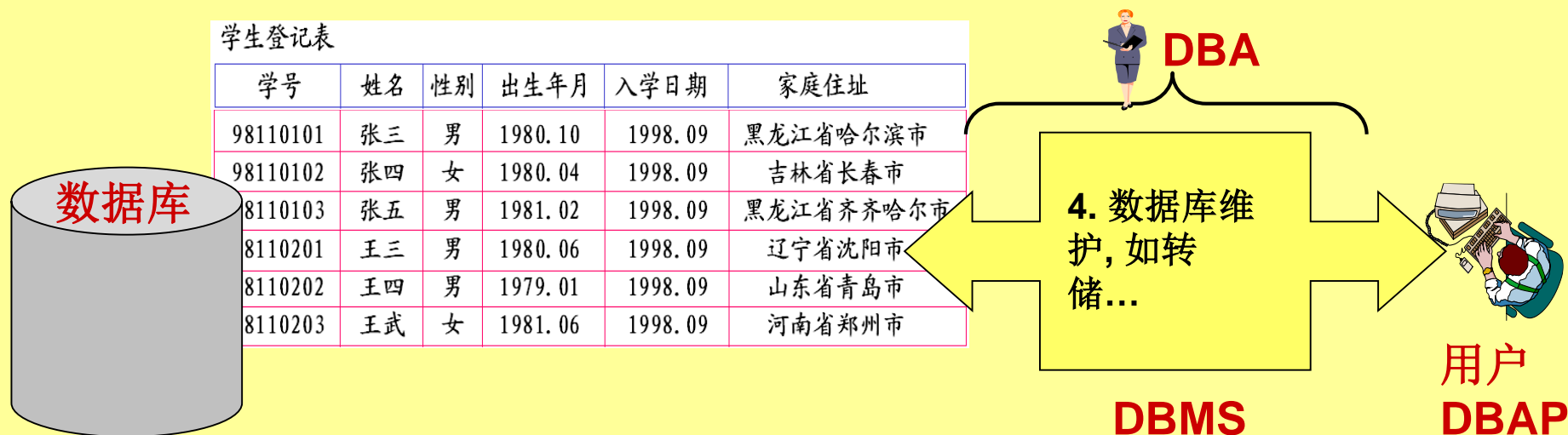
➤ 数据库维护: 转储/恢复/重组/性能监测/分析...

❑ DBMS提供一系列程序(实用程序/例程序)给用户

❑ 在这些程序中提供了对数据库维护的各种功能

❑ 用户使用这些程序进行各种数据库维护操作

➤ 数据库维护的实用程序，一般都是由数据库管理员(DBA)来使用和掌握的



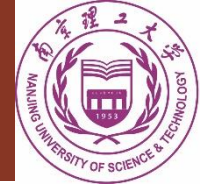


3.5 DBMS还有什么功能?

数据库管理系统DBMS

➤ **DBMS**为完成**DB**管理，在后台运行着一系列程序...

- ❑ 数据库物理存储
- ❑ 数据库查询执行及查询优化
- ❑ 并发控制
- ❑ 故障恢复
- ❑ 安全性控制
- ❑ 完整性控制
- ❑ 数据字典管理
- ❑ 应用程序接口(API)
- ❑



关系模型：什么是关系



4.1 什么是数据模型？

数据模型 一组严格定义的概念的集合

- ❶ **数据结构** 有哪些格式的数据
- ❷ **数据操作** 对这些格式的数据都可能有哪些操作
- ❸ **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

现实世界  信息世界  数据世界

概念数据模型(简称概念模型)

数据库三大经典的数据模型

关系模型

层次模型

网状模型

4. 关系模型：什么是关系

18/48



4.2 你理解关于关系的一些术语的含义吗？

数据库的关系模型起源于规范化“表(Table)”的处理

Table: 以按行按列形式组织及展现的数据

(关系)模式

表名

表标题(格式)

表内容(值)

表/关系

列/字段/属性/数据项(column/field/attribute/data item)

列名

行/元组/记录 (row / tuple / record)

列值

班级	课程	教师	学期	学号	姓名	成绩
981101	数据库	李四	98秋	98110101	张三	100
981101	数据库	李四	98秋	98110102	张四	90
981101	数据库	李四	98秋	98110103	张五	80
981101	计算机	李五	98秋	98110101	张三	89
981101	计算机	李五	98秋	98110102	张四	98
981101	计算机	李五	98秋	98110103	张五	72
981102	数据库	李四	99秋	98110201	王三	30
981102	数据库	李四	99秋	98110202	王四	90
981102	数据库	李四	99秋	98110203	王武	78

Table中描述了一批相互有关联关系的数据==>关系

用数学严格地定义Table

怎样把一张表格定义清楚呢？

家庭

丈夫	妻子	子女
李基	王方	李键
张鹏	刘玉	张睿
张鹏	刘玉	张峰

2. 值域(Domain)

说清楚每一列数据可能的取值

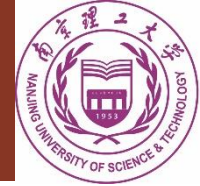
1. 指出有多少列

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

元组是值的一个组合；值域中值的所有可能的组合----笛卡尔积



4.3 如何用数学来定义关系呢？

用数学严格地定义Table

➤首先定义“列”的取值范围“域(Domain)”

➤域(Domain)

□一组值的集合，这组值具有相同的数据类型

□如整数的集合、字符串的集合、全体学生的集合

□再如, 由8位数字组成的数字串的集合，由0到100组成的整数集合

□集合中元素的个数称为域的基数(Cardinality)

家庭

丈夫	妻子	子女
李基	王方	李健
张鹏	刘玉	张睿
张鹏	刘玉	张峰

$D_3 = \text{儿童集合(CHILD)} = \{\text{李健, 张睿, 张峰}\}$

$D_2 = \text{女人集合(WOMAN)} = \{\text{王芳, 刘玉}\}$

$D_1 = \text{男人集合(MAN)} = \{\text{李基, 张鹏}\}$

4.3 如何用数学来定义关系呢？

用数学严格地定义Table

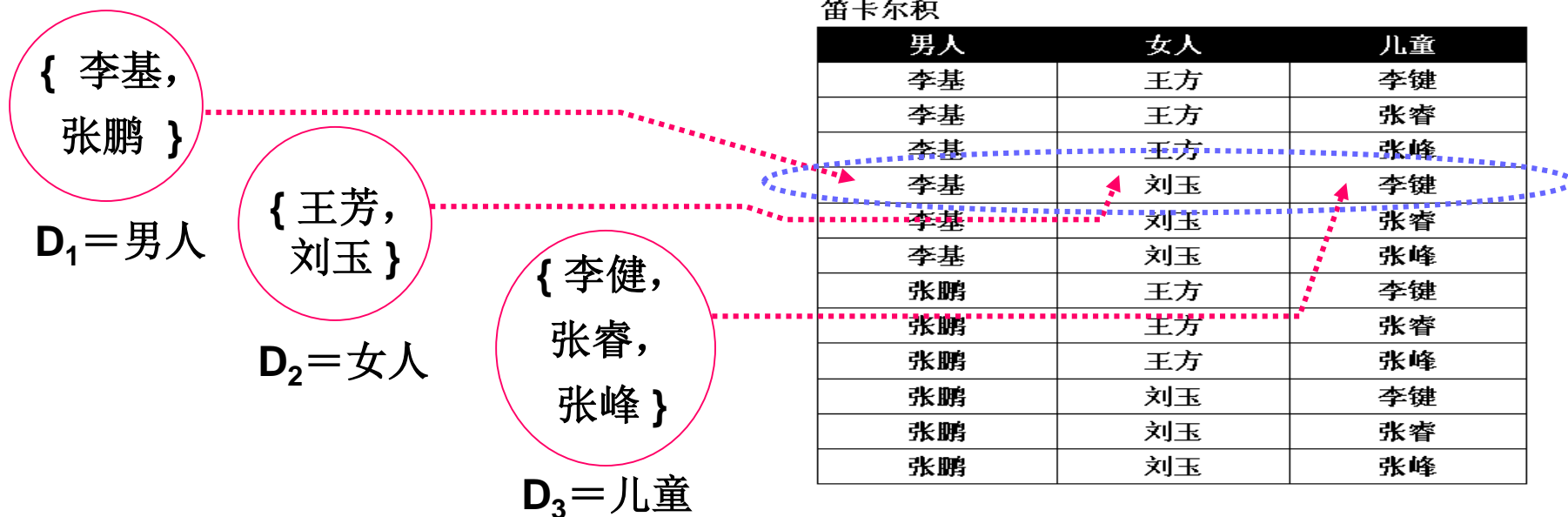
➤再定义“元组”及所有可能组合成的元组：笛卡尔积

➤笛卡尔积(Cartesian Product)

□一组域 D_1, D_2, \dots, D_n 的笛卡尔积为:

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

□笛卡尔积的每个元素 (d_1, d_2, \dots, d_n) 称作一个n-元组 (n-tuple)



4.3 如何用数学来定义关系呢？

用数学严格地定义Table

➤由于笛卡尔积中的所有元组并不都是有意义的，因此...

➤关系(Relation)

□一组域 D_1, D_2, \dots, D_n 的笛卡尔积的子集:

□笛卡尔积中具有某一方面意义的那些元组被称作一个关系(Relation)

□由于关系的不同列可能来自同一个域，为区分，需要为每一列起一个名字，该名字即为属性名。不同列名的列值可以来自相同域。

例如：家庭(丈夫:男人, 妻子:女人, 子女:儿童)或家庭(丈夫, 妻子, 子女)

笛卡尔积

男人	女人	儿童
李基	王方	李键
李基	王方	张睿
李基	王方	张峰
李基	刘玉	李键
李基	刘玉	张睿
李基	刘玉	张峰
张鹏	王方	李键
张鹏	王方	张睿
张鹏	王方	张峰
张鹏	刘玉	李键
张鹏	刘玉	张睿
张鹏	刘玉	张峰



列名(属性名)

丈夫	妻子	子女
李基	王方	李键
张鹏	刘玉	张睿
张鹏	刘玉	张峰

列值：来自域

4. 关系模型：什么是关系

23/48



4.4 关系有什么性质？

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

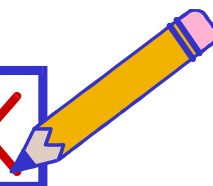
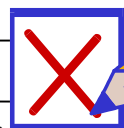
任意两个元组不能完全相同

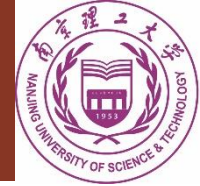
每一分量必须是不可分的数据项

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

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	





4.5 关系中的候选键？

候选码(Candidate Key)/候选键

□关系中的一个属性组，其值能唯一标识一个元组，若从该属性组中去掉任何一个属性，它就不具有这一性质了，这样的属性组称作候选码。

学生(S#, Sname, Sage, Sclass)

课程(C#, Cname, Credit, T#)

4.6 关系中的外键？

外码(Foreign Key)/外键

□关系R中的一个属性组，它不是R的候选码，但它与另一个关系S的候选码相对应，则称这个属性组为R的外码或外键。

□外码是两个关系(数据表)的连接纽带

合同					
主码	合同号	合同名称	合同签定人	客户号	外码
	HT0001	购煤合同	张三	CUST01	
	HT0002	销售机床合同	李四	CUST01	
	HT0003	购钢材合同	张五	CUST02	

两个关系可以靠外码联接起来

主码	客户			
	客户号	客户名称	客户地址	联系人
	CUST01	依兰煤矿	哈尔滨市	王三
	CUST02	长春电机厂	长春市	赵六
	CUST03	鞍钢集团	鞍山市	钱七



关系模型：关系运算

5.1 什么是关系运算？

什么是关系运算？

学生登记表					
学号	姓名	性别	出生年月	入学日期	家庭住址
98110101	张三	男	1980.10	1998.09	黑龙江省哈尔滨市
98110102	学生成绩单				
98110103					
学号	姓名	性别	出生年月	入学日期	家庭住址
98110201	张三	男	1980.10	1998.09	黑龙江省哈尔滨市
98110202	张四	男	1981.11	1999.01	黑龙江省哈尔滨市
98110203	张五	男	1982.12	2000.03	黑龙江省哈尔滨市

有哪些运算？

并： $R \cup S$ 差： $R - S$ 积： $R \times S$ 选择： $\sigma(R)$ 投影： $\pi(R)$ 连接： $R \bowtie S$ 交： $R \cap S$

5.2 什么情况用并运算呢？

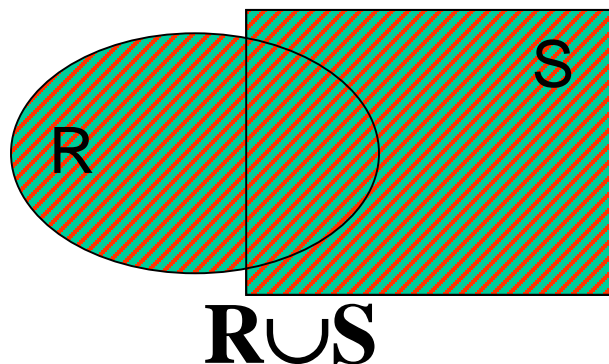
并(Union)

□ **定义**： 设关系**R**和关系**S**是并相容的(即：属性数目相同，其对应的域也相同)， 则关系**R**与关系**S**的并运算结果也是一个关系， 记作：

$R \cup S$, 它由**或者出现在关系R中， 或者出现在S中的元组**构成

□ **数学描述**： **$R \cup S = \{ t \mid t \in R \vee t \in S \}$** ， 其中**t**是元组

□ **$R \cup S$ 与 $S \cup R$** 运算的结果是同一个关系



5.2 什么情况用并运算呢？

并(Union)

R		
A1	A2	A3
a	b	c
a	d	g
f	b	e

S		
B1	B2	B3
a	b	c
a	b	e
a	d	g
h	d	g

RUS		
C1	C2	C3
a	b	c
a	d	g
f	b	e
a	b	e
h	d	g

R(参加体育队的学生)

S#	Sname	Ssex	Sage	D#	Sclass
98030101	张三	男	20	03	980301
98030102	张四	女	20	03	980301
98030103	张五	男	19	03	980301
98040201	王三	男	20	04	980402
98040202	王四	男	21	04	980402
98040203	王五	女	19	04	980402

S(参加文艺队的学生)

S#	Sname	Ssex	Sage	D#	Sclass
98020101	孙三	女	18	02	980201
98020102	孙四	男	20	02	980201
98020103	孙五	女	19	02	980201
98030101	张三	男	20	03	980301
98030102	张四	女	20	03	980301
98030103	张五	男	19	03	980301

RUS(或者参加体育队或者文艺队的学生)

S#	Sname	Ssex	Sage	D#	Sclass
98030101	张三	男	20	03	980301
98030102	张四	女	20	03	980301
98030103	张五	男	19	03	980301
98040201	王三	男	20	04	980402
98040202	王四	男	21	04	980402
98040203	王五	女	19	04	980402
98020101	孙三	女	18	02	980201
98020102	孙四	男	20	02	980201
98020103	孙五	女	19	02	980201

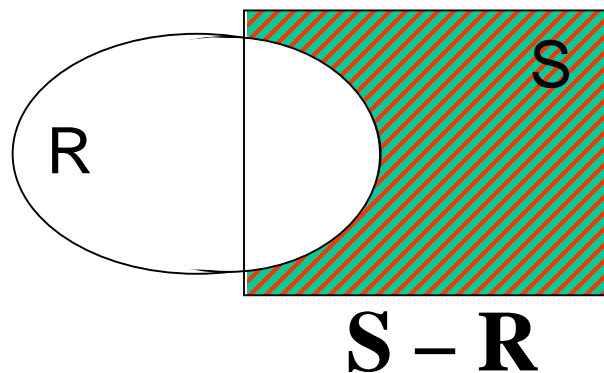
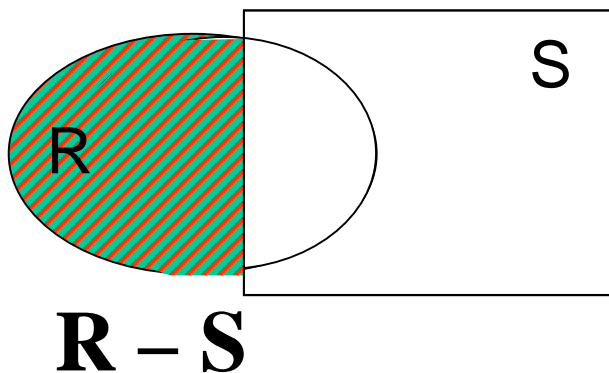
5.3 什么情况用差运算呢？

差(Difference)

□ **定义**：假设关系**R** 和关系**S**是并相容的，则关系**R** 与关系**S** 的差运算结果也是一个关系，记作： $R - S$ ，它由**出现在关系R中但不出现在关系S中的元组**构成

□ **数学描述**： $R - S = \{ t \mid t \in R \wedge t \notin S \}$ ，其中**t**是元组

□ **注意**： $R - S$ 与 $S - R$ 是不同的



5.3 什么情况用差运算呢？

差(Difference)

R		
A1	A2	A3
a	b	c
a	d	g
f	b	e

S		
B1	B2	B3
a	b	c
a	b	e
a	d	g
h	d	g

R - S		
D1	D2	D3
f	b	e

S - R		
E1	E2	E3
a	b	e
h	d	g

R(参加体育队的学生)

S#	Sname	Ssex	Sage	D#	Sclass
98030101	张三	男	20	03	980301
98030102	张四	女	20	03	980301
98030103	张五	男	19	03	980301
98040201	王三	男	20	04	980402
98040202	王四	男	21	04	980402
98040203	王五	女	19	04	980402

R-S(参加体育队而未参加文艺队的学生)

S#	Sname	Ssex	Sage	D#	Sclass
98040201	王三	男	20	04	980402
98040202	王四	男	21	04	980402
98040203	王五	女	19	04	980402

S(参加文艺队的学生)

S#	Sname	Ssex	Sage	D#	Sclass
98020101	孙三	女	18	02	980201
98020102	孙四	男	20	02	980201
98020103	孙五	女	19	02	980201
98030101	张三	男	20	03	980301
98030102	张四	女	20	03	980301
98030103	张五	男	19	03	980301

S-R(参加文艺队而未参加体育队的学生)

S#	Sname	Ssex	Sage	D#	Sclass
98020101	孙三	女	18	02	980201
98020102	孙四	男	20	02	980201
98020103	孙五	女	19	02	980201

5.4 什么情况用交运算呢？

交(Intersection)

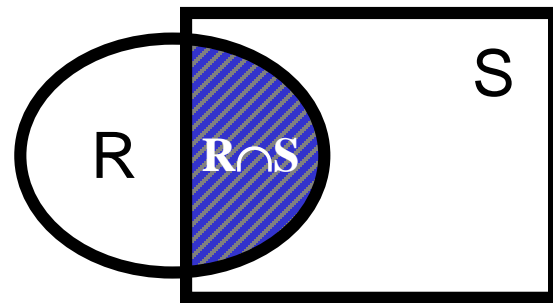
□ **定义**：假设关系**R**和关系**S**是并相容的，则关系**R**与关系**S**的交运算结果也是一个关系，记作： $R \cap S$ ，它由**同时出现在**关系**R**和关系**S**中的元组构成

□ **数学描述**： $R \cap S = \{ t \mid t \in R \wedge t \in S \}$ ，其中**t**是元组

□ $R \cap S$ 和 $S \cap R$ 运算的结果是同一个关系

□ 交运算可以通过差运算来实现：

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



5.4 什么情况用交运算呢？

交(Intersection)

R			S			$R \cap S$		
A1	A2	A3	B1	B2	B3	F1	F2	F3
a	b	c	a	b	c	a	b	c
a	d	g	a	b	e	a	d	g
f	b	e	a	d	g			
			h	d	g			

R(参加体育队的学生)

S#	Sname	Ssex	Sage	D#	Sclass
98030101	张三	男	20	03	980301
98030102	张四	女	20	03	980301
98030103	张五	男	19	03	980301
98040201	王三	男	20	04	980402
98040202	王四	男	21	04	980402
98040203	王五	女	19	04	980402

 $R \cap S$ (既参加体育队又参加文艺队的学生)

S#	Sname	Ssex	Sage	D#	Sclass
98030101	张三	男	20	03	980301
98030102	张四	女	20	03	980301
98030103	张五	男	19	03	980301

S(参加文艺队的学生)

S#	Sname	Ssex	Sage	D#	Sclass
98020101	孙三	女	18	02	980201
98020102	孙四	男	20	02	980201
98020103	孙五	女	19	02	980201
98030101	张三	男	20	03	980301
98030102	张四	女	20	03	980301
98030103	张五	男	19	03	980301



5.5 什么情况用笛卡尔积运算呢？

广义笛卡尔积 (Cartesian Product)

□ **定义**：关系 $R(<a_1, a_2, \dots, a_n>)$ 与关系 $S(<b_1, b_2, \dots, b_m>)$ 的广义笛卡尔积(简称广义积)运算结果也是一个关系，记作： $R \times S$ ，它由关系 R 中的元组与关系 S 的元组进行所有可能的拼接(或串接)构成。

□ **数学描述**： $R \times S = \{ <a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_m> \mid <a_1, a_2, \dots, a_n> \in R \wedge <b_1, b_2, \dots, b_m> \in S \}$

□ 笛卡尔积可将两个表串接起来作为一个表进行操作

学生表

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

课程表

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

(所有学生)的(所有课程)

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	教师 1	52
981101	李四	22	3010	C2	物理	教师 2	36
981101	李四	22	3010	C3	高数	教师 5	40
981103	李三	21	3011	C1	计算机	教师 1	52
981103	李三	21	3011	C2	物理	教师 2	36
981103	李三	21	3011	C3	高数	教师 5	40
981105	李六	22	3011	C1	计算机	教师 1	52
981105	李六	22	3011	C2	物理	教师 2	36
981105	李六	22	3011	C3	高数	教师 5	40

5.5 什么情况用笛卡尔积运算呢？

广义笛卡尔积 (Cartesian Product)

R		
A1	A2	A3
a	b	c
a	d	g
f	b	e

S		
B1	B2	B3
a	b	c
a	b	e
a	d	g
h	d	g

R × S					
A1	A2	A3	B1	B2	B3
a	b	c	a	b	c
a	b	c	a	b	e
a	b	c	a	d	g
a	b	c	h	d	g
a	d	g	a	b	c
a	d	g	a	b	e
a	d	g	a	d	g
a	d	g	h	d	g
f	b	e	a	b	c
f	b	e	a	b	e
f	b	e	a	d	g
f	b	e	h	d	g

Relations r, s :

A	B
α	1
β	2

C	D	E
α	10	a
β	10	a
β	20	b
γ	10	b

$r \times s$:

A	B	C	D	E
α	1	α	10	a
α	1	β	19	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

[illegible]

5. 关系运算

5.6 什么情况用选择运算呢？

37/48



选择(Selection)

R		
A1	A2	A3
a	a	10
a	d	-4
f	b	5

$\sigma_{A3>0}(R)$		
A1	A2	A3
a	a	10
f	b	5

$\sigma_{A2="a" \vee A2="b"}(R)$		
A1	A2	A3
a	a	10
f	b	5

$\sigma_{A3>0 \wedge A1=A2}(R)$		
A1	A2	A3
a	a	10

R(学生表)

S#	Sname	Ssex	Sage	D#	Sclass
98030101	张三	男	20	03	980301
98030102	张四	女	21	03	980301
98030103	张五	男	19	03	980301
98040201	王三	男	18	04	980402
98040202	王四	男	21	04	980402
98050104	孙六	女	19	05	980501

查询所有年龄小于20同学的信息

$\sigma_{Sage<20}(R)$

S#	Sname	Ssex	Sage	D#	Sclass
98030103	张五	男	19	03	980301
98040201	王三	男	18	04	980402
98050104	孙六	女	19	05	980501

查询所有3系或5系的同学信息

$\sigma_{D#="03" \vee D#="05"}(R)$

S#	Sname	Ssex	Sage	D#	Sclass
98030101	张三	男	20	03	980301
98030102	张四	女	21	03	980301
98030103	张五	男	19	03	980301
98050104	孙六	女	19	05	980501

□ **定义**：给定一个关系R, 投影运算结果也是一个关系，记作 $\Pi_A(R)$ ，它从关系R中选出属性包含在A中的列构成

✓ 设 $R(A_1, A_2, \dots, A_n)$

✓ $\{A_{i1}, A_{i2}, \dots, A_{ik}\} \subseteq \{A_1, A_2, \dots, A_n\}$

✓ $t[A_i]$ 表示元组 t 中相应于属性 A_i 的分量

✓ 投影运算可以对原关系的列在投影后重新排列

R		
A1	A2	A3

5.7 什么情况用投影运算呢？

投影(Projection)

R		
A1	A2	A3
a	b	c
a	d	g
f	b	e

$\Pi_{A3}(R)$
A3
c
g
e

$\Pi_{A3, A1}(R)$	
A3	A1
c	a
g	a
e	f

R(学生表)

S#	Sname	Ssex	Sage	D#	Sclass
98030101	张三	男	20	03	980301
98030102	张四	女	21	03	980301
98030103	张五	男	19	03	980301
98040201	王三	男	18	04	980402
98040202	王四	男	21	04	980402
98050104	孙六	女	19	05	980501

 $\Pi_{Sname, Sage}(R)$

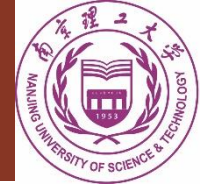
查询所有
学生的姓名
和年龄

Sname	Sage
张三	20
张四	21
张五	19
王三	18
王四	21
孙六	19

 $\Pi_{Sname, D\#}(R)$

查询所有
学生的姓名
及其所在
的系

Sname	D#
张三	03
张四	03
张五	03
王三	04
王四	04
孙六	05



5.8 什么情况用连接运算呢？

 θ -连接(θ -Join)

□ **定义**：给定关系R和关系S, R与S的 θ 连接运算结果也是一个关系，记作 $R \bowtie_{A \theta B} S$ ，它由关系R和关系S的笛卡尔积中，选取R中属性A与S中属性B之间满足 θ 条件的元组构成。

□ **数学描述**：
$$R \bowtie_{A \theta B} S = \sigma_{t[A] \theta s[B]} (R \times S)$$

✓ 设 $R(A_1, A_2, \dots, A_n)$, $A \in \{A_1, A_2, \dots, A_n\}$

✓ $S(B_1, B_2, \dots, B_m)$, $B \in \{B_1, B_2, \dots, B_m\}$

✓ t是关系R中的元组，s是关系S中的元组

✓ 属性A和属性B具有可比性

✓ θ 是比较运算符, $\theta \in \{>, \geq, <, \leq, =, \neq\}$

□ 在实际应用中， θ -连接操作经常与投影、选择操作一起使用



5.8 什么情况用连接运算呢？

自然连接(Natural-Join)

□ **定义**：给定关系R和关系S, R与S的自然连接运算结果也是一个关系，记作 $R \bowtie S$ 它由关系R和关系S的笛卡尔积中选取相同属性组B上值相等的元组所构成。

□ **数学描述**：
$$R \bowtie S = \sigma_{t[B] = s[B]} (R \times S)$$

- ✓ 自然连接是一种特殊的连接运算
- ✓ 要求关系R和关系S必须有相同的属性组B(如R,S共有一个属性 B_1 ,则B是 B_1 , 如R, S共有一组属性 B_1, B_2, \dots, B_n , 则B是这些共有的所有属性)
- ✓ R, S属性相同, 值必须相等才能连接, 即

$R.B_1 = S.B_1$ and $R.B_2 = S.B_2 \dots$ and $R.B_n = S.B_n$ 才能连接

- ✓ 要在结果中去掉重复的属性列(因结果中 $R.B_i$ 始终是等于 $S.B_i$ 所以可只保留一列即可)

5.8 什么情况用连接运算呢？

 θ -连接 vs. 连接 vs. 笛卡尔积

R	
A	B
a	1
b	2

S	
B	C
1	x
1	y
3	z

$R \times S$			
A	B	B	C
a	1	1	x
a	1	1	y
a	1	3	z
b	2	1	x
b	2	1	y
b	2	3	z

$R \bowtie S$		
A	B	C
a	1	x
a	1	y

R	
A	B
a	1
b	2

S	
H	C
1	x
1	y
3	z

$R \times S$			
A	B	H	C
a	1	1	x
a	1	1	y
a	1	3	z
b	2	1	x
b	2	1	y
b	2	3	z

$R \bowtie_{B \leq H} S$			
A	B	H	C
a	1	1	x
a	1	1	y
a	1	3	z
b	2	3	z



6. 应用关系运算进行数据库的查询

应用关系运算进行数据库的查询

6.应用关系运算进行数据库的查询

44/48



6.1 利用关系运算进行查询？

查询表达式

组合各种运算

Student					
S#	Sname	Ssex	Sage	D#	Sclass
98030101	张三	男	20	03	980301
98030102	张四	女	21	03	980301
98030103	张五	男	19	03	980301
98040201	王三	男	18	04	980402
98040202	王四	男	21	04	980402
98050104	孙六	女	19	05	980501

SC		
S#	C#	Score
98030101	001	92.0
98030101	002	85.0
98030101	003	88.0
98040202	002	90.5
98040202	003	80.0
98040202	001	55.0
98050104	003	56.0
98030102	001	54.0
98030102	002	85.0
98030102	003	48.0

- 查询学习课程号为002的学生学号和成绩

$\pi_{S\#, Score}(\sigma_{C\#="002"}(SC))$

- 查询学习课程号为001的学生学号、姓名

$\pi_{S\#, Sname}(\sigma_{C\#="001"}(Student \bowtie SC))$

Course				
C#	Cname	Chours	Credit	T#
001	数据库	40	6	001
003	数据结构	40	6	003
004	编译原理	40	6	001
005	C 语言	30	4.5	003
002	高等数学	80	12	004

- 查询学习课程名称为数据结构的学生学号、姓名和这门课程的成绩

$Student \bowtie SC \bowtie Course$

$\sigma_{Cname="数据结构"}(Student \bowtie SC \bowtie Course)$

$\pi_{S\#, Sname, Score}(\sigma_{Cname="数据结构"}(Student \bowtie SC \bowtie Course))$

6.应用关系运算进行数据库的查询

45/48



6.1 利用关系运算进行查询？

查询表达式

注意连接与积的差别

Student					
S#	Sname	Ssex	Sage	D#	Sclass
98030101	张三	男	20	03	980301
98030102	张四	女	21	03	980301
98030103	张五	男	19	03	980301
98040201	王三	男	18	04	980402
98040202	王四	男	21	04	980402
98050104	孙六	女	19	05	980501

SC		
S#	C#	Score
98030101	001	92.0
98030101	002	85.0
98030101	003	88.0
98040202	002	90.5
98040202	003	80.0
98040202	001	55.0
98050104	003	56.0
98030102	001	54.0
98030102	002	85.0
98030102	003	48.0

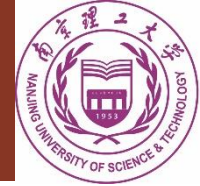
➤ 查询学习课程号为001的学生学号、姓名

$\pi_{S\#,Sname}(\sigma_{C\#="001"}(Student \bowtie SC))$

Course				
C#	Cname	Chours	Credit	T#
001	数据库	40	6	001
003	数据结构	40	6	003
004	编译原理	40	6	001
005	C 语言	30	4.5	003
002	高等数学	80	12	004

$\pi_{S\#,Sname}(\sigma_{C\#="001" \wedge Student.S\# = SC.S\#}(Student \times SC))$

连接条件



关系运算式

$\Pi_{\text{列名}, \dots, \text{列名}} (\sigma_{\text{检索条件}} (\text{表名1} \times \text{表名2} \times \dots))$

□数据库语言SQL

Select 列名 [[, 列名] ...]

From 表名1 [[, 表名2], ...]

[**Where** 检索条件] ;

语义：将**From**后面的所有表串接起来，检索出满足“检索条件”的元组，并按给定的列名及顺序进行投影显示。

$\Pi_{S\#, Sname, Score} (\sigma_{Cname=\text{“数据结构”} \wedge Student.S\#=SC.S\# \wedge Course.C\#=SC.C\#} (Student \times SC \times Course))$

Select S#, Sname, Score

From Student, SC, Course

Where Cname= ‘数据结构’ **and** Student.S#=SC.S# **and** Course.C#=SC.C#;



- 数据中蕴含着价值，需要聚集与管理
- 数据库是若干相互有关联关系的数据的集合
- 数据库系统包括数据库、数据库管理系统、数据库管理员、数据库应用程序、计算机与网络基本系统构成
- 数据库管理系统包括数据库定义、数据库操纵、数据库控制、数据库维护等功能，它还能够进行数据库物理存储、查询实现与查询优化、安全性控制、完整性控制、故障恢复等。
- 结构化数据库的核心是关系模型，基本数据结构为关系，有并、差、笛卡尔积、选择、投影、连接等运算

第10讲 怎样管理和利用数据(I)

Questions & Discussion?