



数据库原理与应用

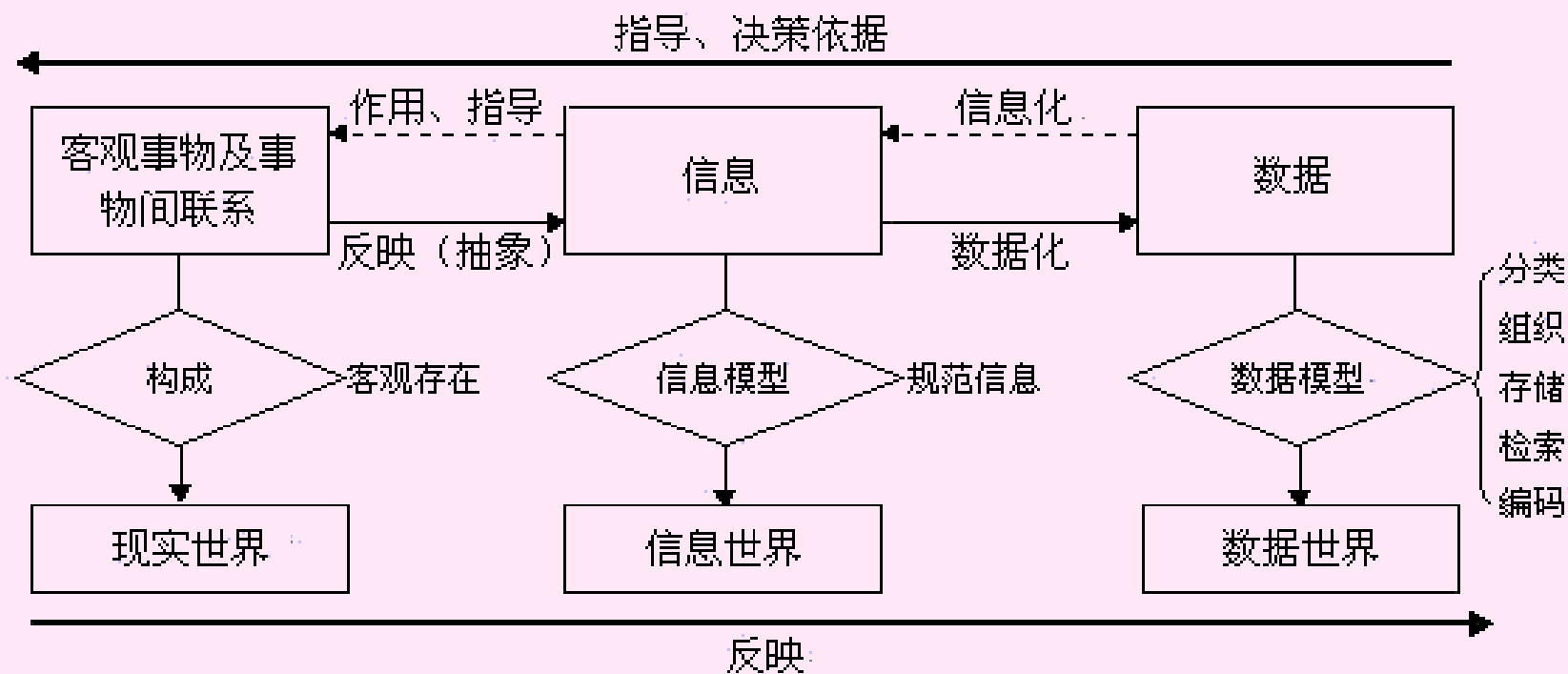
苏州大学计算机科学与技术学院





知识点7: 数据模型介绍

信息的三个世界



现实世界



■ 现实世界

- 指存在于人脑之外的客观世界，泛指客观存在的事物及其相互间的联系。
- 一个实际存在并可以识别的事物称为个体，每个个体都有自己的特征，用以区别其他个体

信息世界



- **现实世界中的事物反映到人们的头脑里，经过认识、选择、命名、分类等综合分析而形成了概念和认识，这就是信息，即进入了信息世界**
 - **在信息世界中，每一个被认识的个体称为实体，个体的特征在头脑里形成的知识称为属性。**
 - **一个实体是由它所有的属性来表示的。**



- 信息要能在计算机中表示出来，则需要进一步的转换。
- 在信息世界里，有些信息可以直接用数字表示，有些需要由符号、文字等来表示；但在机器世界里，所有的信息只能用二进制数来表示。
- 一切信息进入计算机时，必须是数据化的。可以说数据是信息的具体表现形式。

客观对象的抽象过程



■ 客观对象的抽象过程---两步抽象

- 现实世界中的客观对象抽象为概念模型；

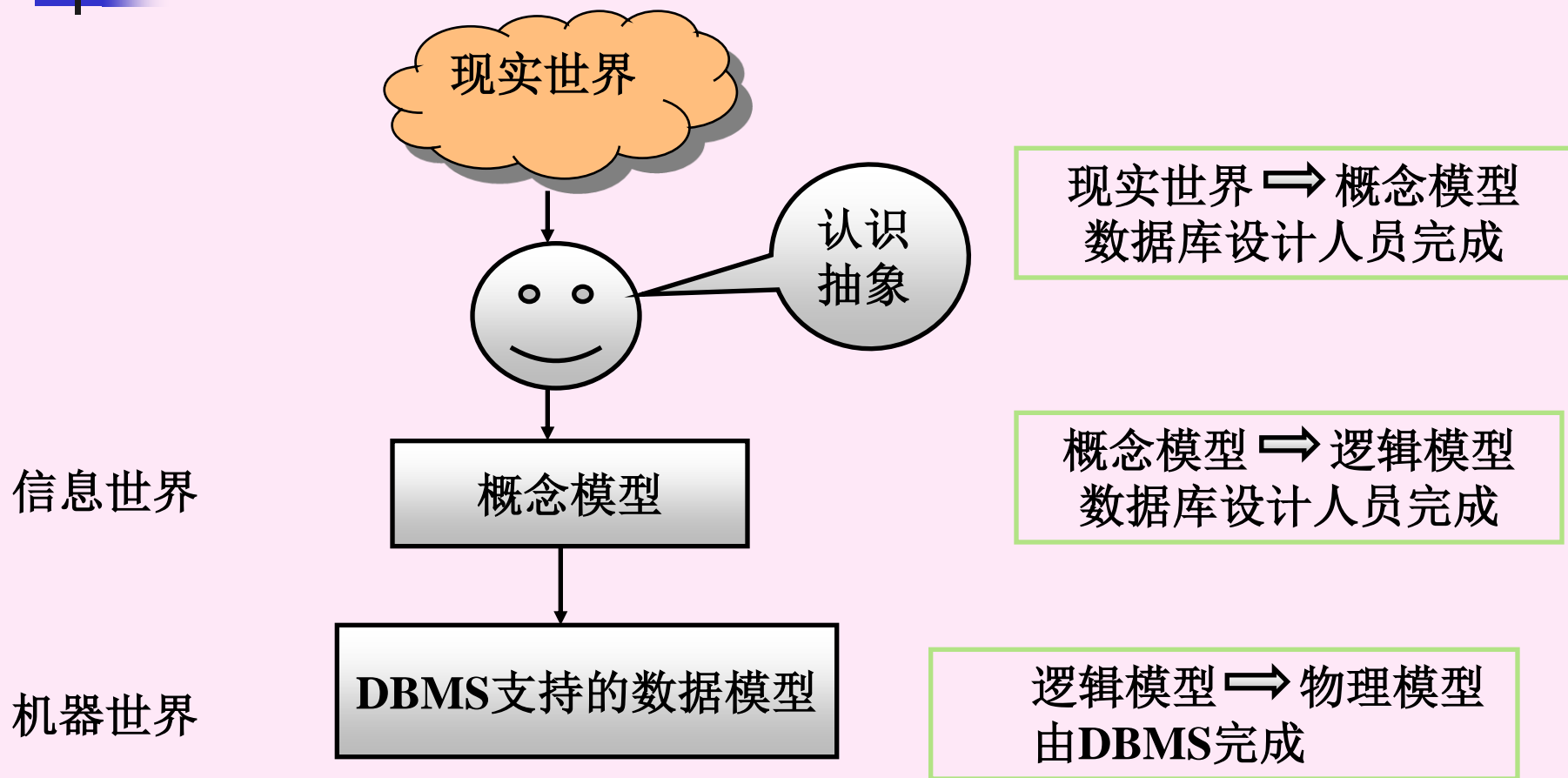
概念模型 也称信息模型，它是按用户的观点来对数据和信息建模。

- 把概念模型转换为某一DBMS支持的数据模型。

数据模型 主要包括网状模型、层次模型、关系模型等，它是按计算机系统的观点对数据建模。



客观对象的抽象过程（续）



现实世界中客观对象的抽象过程

数据模型



- 在数据库中用数据模型这个工具来抽象、表示和处理现实世界的信息 and 数据。
- 通俗地讲数据模型就是现实世界的模拟。
- 数据模型应满足三方面要求
 - 能比较真实地模拟现实世界
 - 容易为人所理解
 - 便于在计算机上实现



知识点8: 数据模型的组成要素



数据模型的组成要素



- **数据结构**
- **数据操作**
- **数据的完整性约束条件**

数据结构



■ 什么是数据结构

- 数据库的组成对象以及对象之间的联系

■ 两类对象

- 与数据类型、内容、性质有关的对象（对数据的描述）
- 与数据之间联系有关的对象（对数据间联系的描述）

■ 数据结构是对系统静态特性的描述

■ 数据结构是刻画一个数据模型性质最重要的方面，因此在数据库系统中通常按照数据结构的类型来命名数据模型

- 采用层次型、网状型和关系型数据结构的数据模型分别被称为层次模型、网状模型和关系模型

数据操作



■ 数据操作

- 对数据库中各种对象的实例允许执行的**操作**及有关的操作规则。

■ 数据操作的类型

- 检索和更新（包括插入、删除、修改）
- 数据操作是对系统动态特性的描述。
- 数据模型要给出这些操作确切的含义、操作规则和实现操作的语言。

数据的约束条件



■ 数据的完整性约束条件

■ 一组完整性规则的集合。

- 层次模型中的每个结点有且仅有一个双亲结点；关系模型中的实体完整性和参照完整性

- 完整性规则是给定的数据模型中数据及其联系所具有的制约和依存规则，用以限定符合数据模型的数据库状态以及状态的变化，以保证数据的正确、有效、相容。

- 数据约束还应包括能够反映某一应用所涉及的数据必须遵守的特定的语义约束条件



知识点9: 数据模型的发展

传统数据模型



■ 传统数据模型

- 20世纪60年代后期发展起来的层次、网状和关系数据模型
- 传统数据模型是在文件系统基础上发展起来的，它们都在记录的基础上定义了各自数据的基本结构、操作和完整性约束及不同类型记录间的联系

■ 缺点

- 过于面向机器实现、模拟现实世界能力不足、语义贫乏等

非传统数据模型



- **E-R模型（20世纪70年代后期）**
 - 提供了丰富的语义和直接模拟现实世界的能力，且具有直观、自然、易于用户理解等优点
 - E-R数据模型与传统数据模型的主要区别在于，它不是面向机器实现，而是面向现实世界的
- **20世纪80年代以来又相继推出面向对象数据模型、面向对象关系数据模型等新的数据模型**



知识点**10**: 概念模型介绍

概念模型



■ 概念模型的用途

- 概念模型用于信息世界的建模
- 是现实世界到机器世界的一个中间层次
- 是数据库设计的有力工具

■ 数据库设计人员和用户之间进行交流的语言

- 对概念模型的基本要求
- 较强的语义表达能力
- 能够方便、直接地表达应用中的各种语义知识
- 简单、清晰、易于用户理解



ER模型



- 数据库设计中广泛使用的概念模型当属E-R数据模型
- E-R数据模型（Entity-Relationship data model），即实体联系数据模型，是P.P.S. Chen于1976年提出的一种语义数据模型。

ER模型的优点



- E-R数据模型描述现实世界，不必考虑信息的存储结构、存取路径、存取效率以及如何在计算机中实现。
- 它与传统数据模型相比，更便于直接描述现实世界。
- 具有直观、自然、语义丰富、易于向传统数据模型转换等优点。



知识点11: ER模型基本概念

ER模型介绍



- 设计E-R数据模型的目标是有效地和自然地模拟现实世界
- E-R数据模型只应包含那些对描述现实世界有普遍意义的抽象概念
- 三要素：
 - 实体
 - 联系
 - 属性

实体



■ 实体 (Entity)

- 客观存在并可相互区别的事物称为实体。
- 可以是具体的人、事、物或抽象的概念。如一个学生，一本书，学生的一次选课

■ 实体型

- 具有相同属性的实体必然具有共同的特征和性质。
- 用实体名及其属性名集合来抽象和刻画同类实体，例如，学生（学号，姓名，性别，出生年份，系，入学时间）就是一个实体型。

实体集



■ 实体集 (Entity Sets)

- 具有相同性质的一类实体的集合
- 实体集中各个实体是借助实体标识符（称为关键字）加以区别的。
- 例如，可以定义学校的“教师”为一实体集，而学校中每个教师都是该实体集的成员。



知识点**12**: 联系

联系



■ 联系 (Relationship)

- 现实世界中事物内部以及事物之间的联系在信息世界中反映为实体型内部的联系和实体型之间的联系

联系

两个实体型间

三个或三个以上实体型间

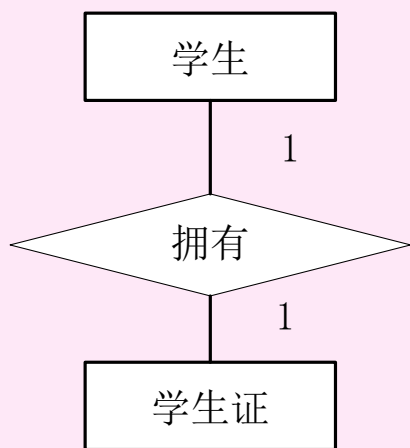
同一个实体型内

一对一联系 (1:1)
一对多联系 (1:n)
多对多联系 (m:n)

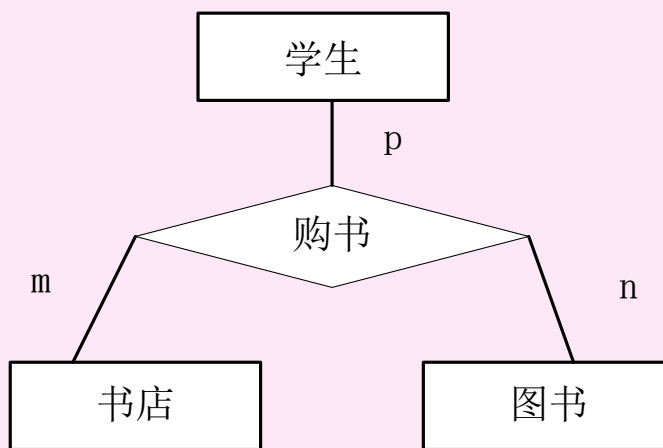
n元联系



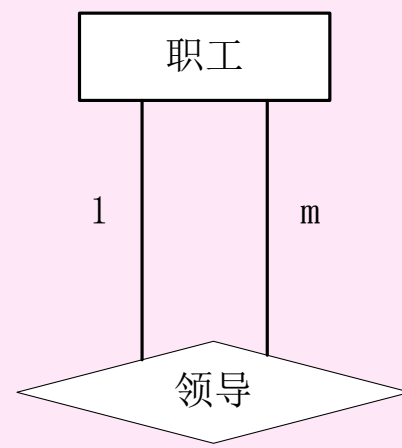
- 如果参与联系的实体集的数目为 n ，则称这种联系数为 n 元联系



二元联系



多元联系



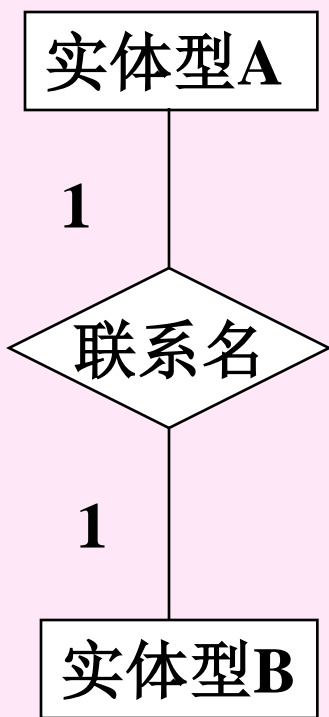
自反联系

二元联系

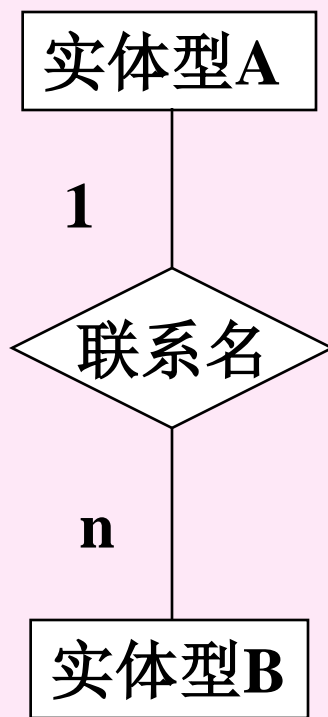


- 只有两个实体集参与的联系称为二元联系，它是现实世界中大量存在的联系
- 二元联系可进一步细分为以下三种联系：
 - 一对一 (1: 1) 联系
 - 一对多 (1: n) 的联系
 - 多对多 (m: n) 联系

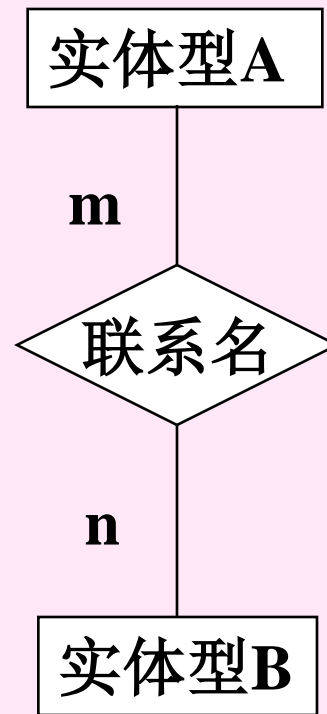
两个实体型间的联系



1:1联系



1:n联系

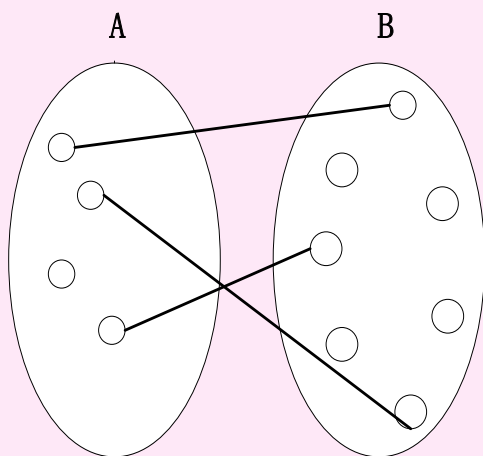


m:n联系

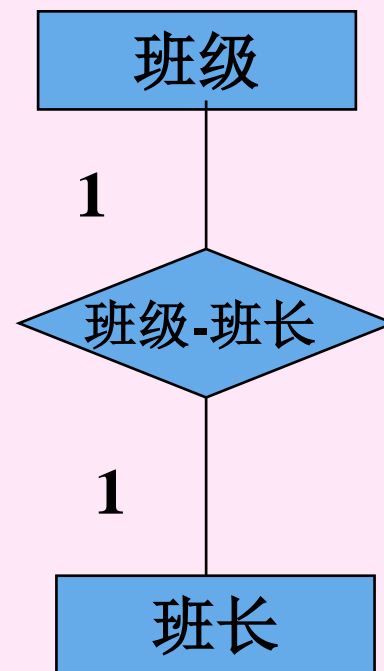
一对一联系



- 如果对于实体集A中的每一个实体，实体集B中至多有一个实体与之联系，反之亦然，则称实体集A与实体集B具有一对一联系。记为1:1。



(a) 一对一联系

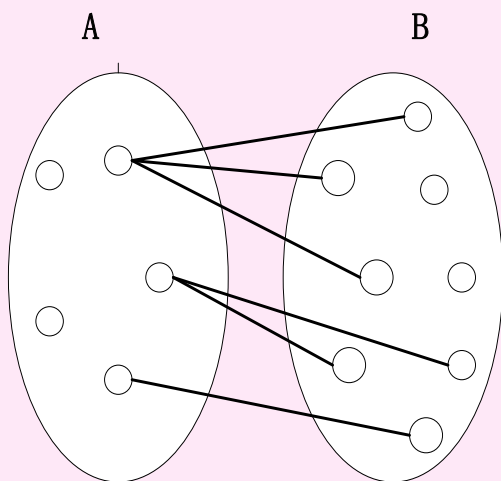


1:1联系

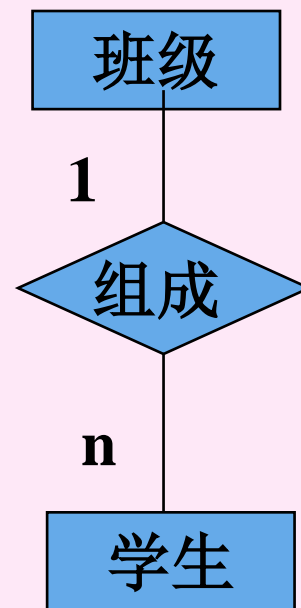
一对多 (1: n) 的联系



- 设有两个实体集A和B，若A中每个实体与B中任意个实体（包括零个）发生关联，但B中每个实体至多和A中一个实体有联系、则称A和B是一对多的联系，记为1: n。



(b) 一对多联系

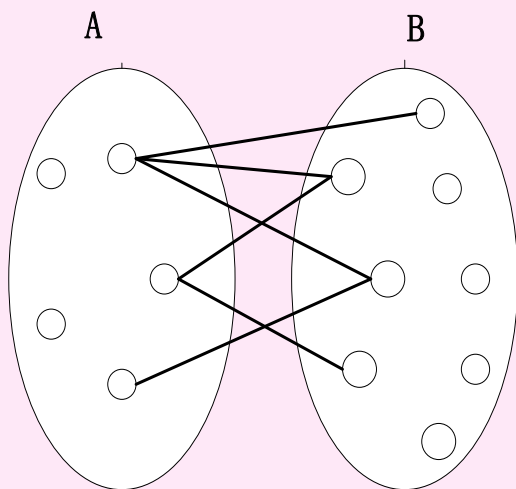


1:n联系

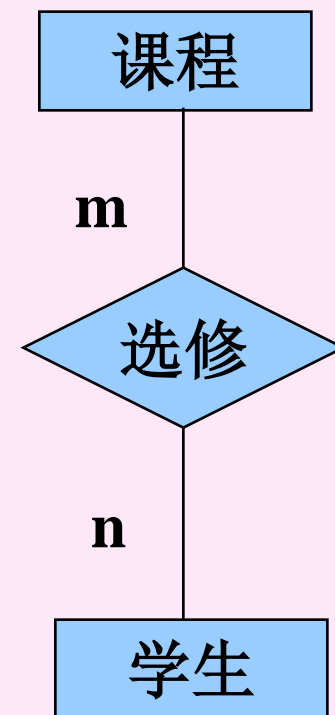
多对多 (m: n)联系



- 若两个实体集A和B，每个实体集中的每一个实体都和另一个实体集中任意多个（包括0个）实体有联系，则称A和B是多对多的联系，记为m: n。



(c) 多对多联系



m:n联系

多个实体型间的联系



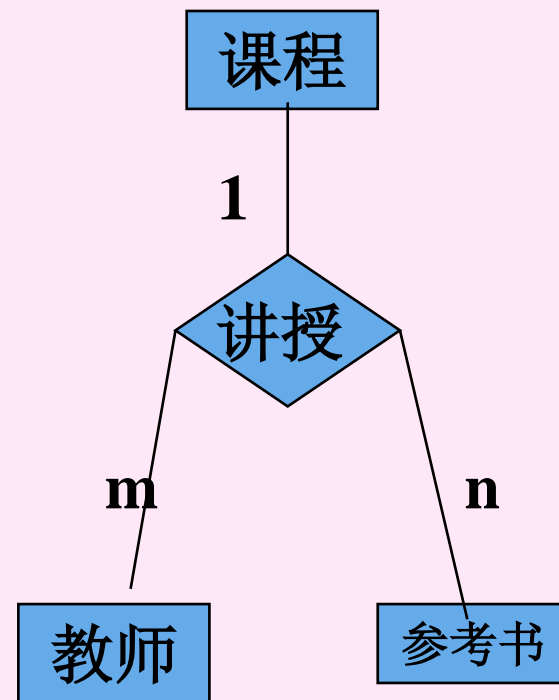
- 两个以上不同实体型之间也可能存在各种联系，以3个不同实体型A、B、C为例，它们之间的典型联系有1:m:n和m:n:p联系。
- 对于1:m:n联系，表示A和B之间是1:m(一对多)联系，B和C之间m:n(多对多)联系，A和C之间是1:n(一对多)联系。

多个实体型间的联系(续)



实例

- 课程、教师与参考书三个实体型
- 如果一门课程可以有若干个教师讲授，使用若干本参考书，每一个教师只讲授一门课程，每一本参考书只供一门课程使用课程与教师、参考书之间的联系是一对多的



两个以上实体型间1:n联系

同一实体型内部的联系



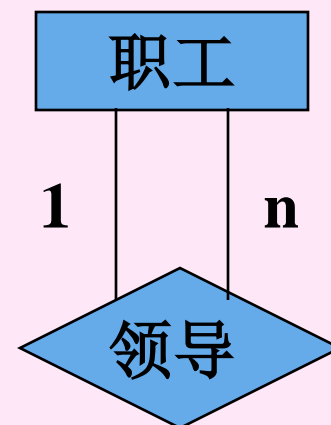
实例

职工实体型内部具有领导与被领导的联系

某一职工（干部）“领导”若干名职工；

一个职工仅被另外一个职工直接领导

这是一对多的联系



单个实体型内部1:n联系



知识点**13**: 属性

属性 (Attribute)



- 实体或联系所具有的特征称之为属性。
- 通常一个实体可以由多个属性来描述，即实体可用属性集表示。
 - 例如学生实体可用学号、姓名、性别、年龄、系、籍贯等属性来描述。(990001, 张立, 20, 男, 计算机) 这些属性值的集合表示了一个学生实体。
- 联系也可以用属性来描述使其语义更加丰富。
 - 学生实体集和课程实体集间存在 $m:n$ 的“选课”联系，这种联系可以有“成绩”、“选修时间”等属性

属性的说明



- 一个实体可以有若干个属性，但在数据库设计中通常只选择那些数据管理需要的属性，而不是全部属性。
- 不能再细分的属性称为原子属性，如性别、民族等。
- 属性有型与值的区别。如学生实体中的学号、姓名等属性名是属性型，而“0000001”、“刘强”等具体数据称为属性值。
- 每个属性值都有一定的变化范围，通常称属性取值的变化范围为属性值的域(domain)。例如，性别属性域是男、女，年龄属性域是1-200岁。
- 能惟一标识实体集中某一实体的属性或属性组称为实体集的关键字 (Key)。



知识点14: ER图

E-R图



- 用E-R图来描述现实世界的概念模型
- 实体型
 - 用矩形表示，矩形框内写明实体型名。

学生

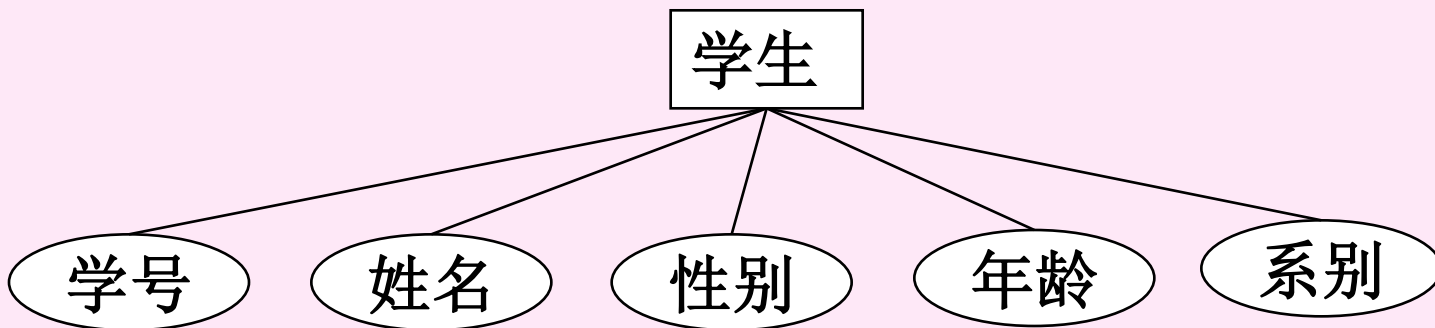
教师

E-R图(续)



■ 属性

- 用椭圆形表示，并用无向边将其与相应的实体型连接起来。



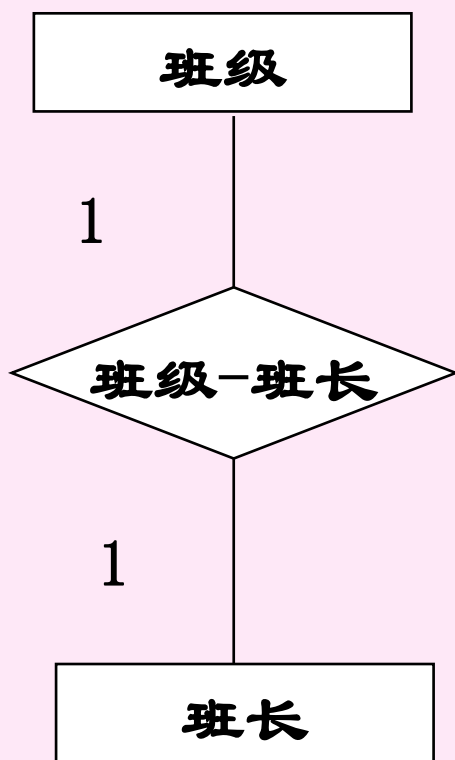
E-R图(续)



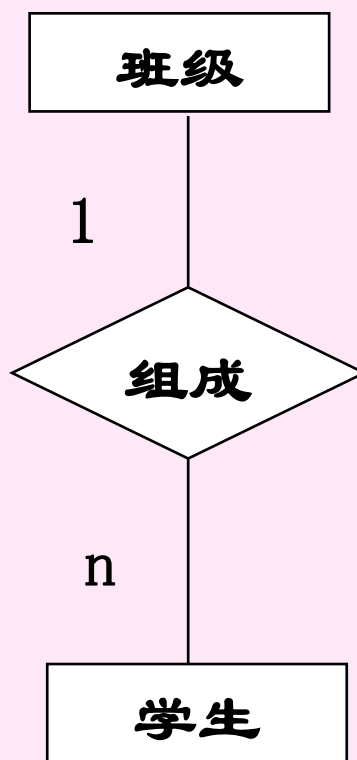
■ 联系

- **联系本身：** 用菱形表示，菱形框内写明联系名，并用无向边分别与有关实体型连接起来，同时也在无向边旁标上联系的类型（1:1、1:n或m:n）
- **联系的属性：** 联系本身也是一种实体型，也可以有属性。如果一个联系具有属性，则这些属性也要用无向边与该联系连接起来。

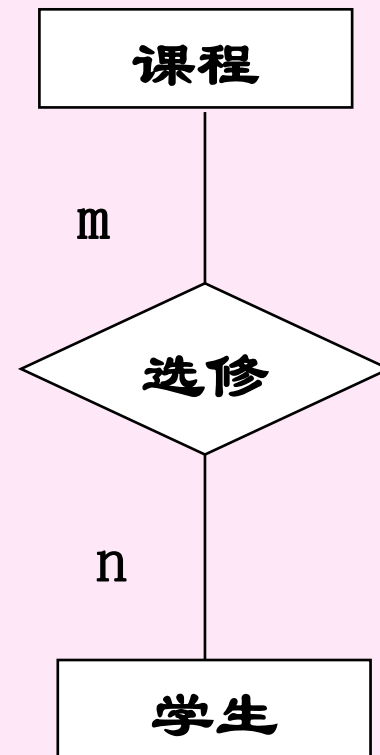
联系的表示方法示例



1:1联系

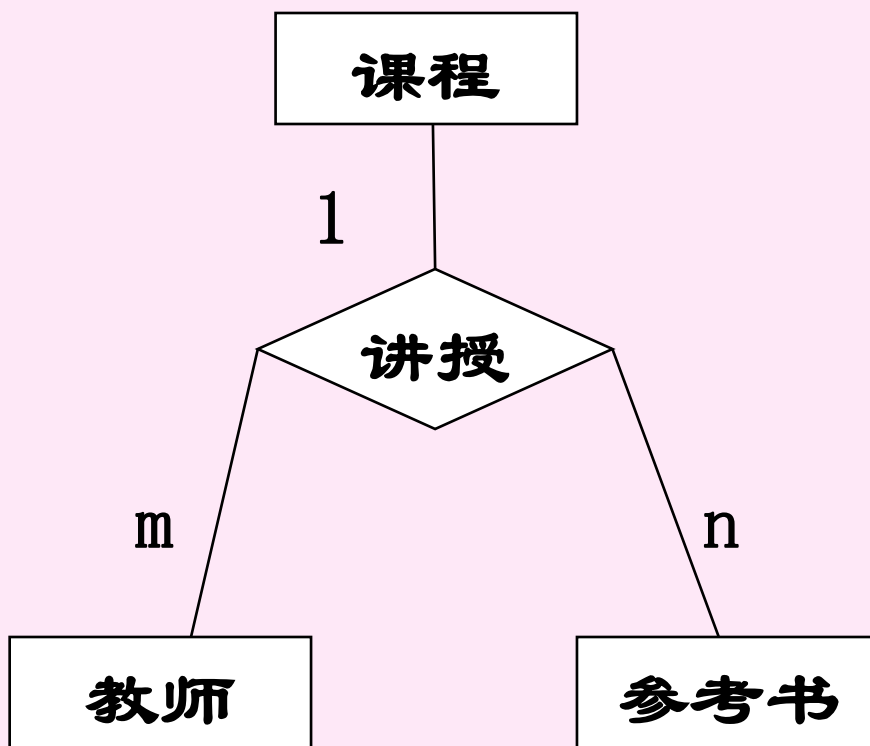


1:n联系

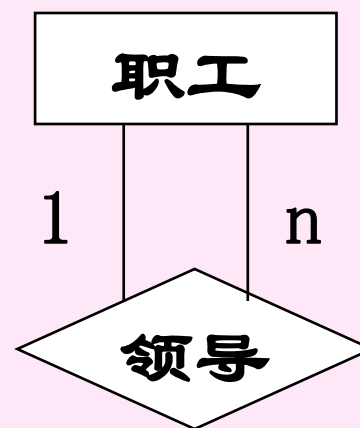


m:n联系

联系的表示方法示例（续）

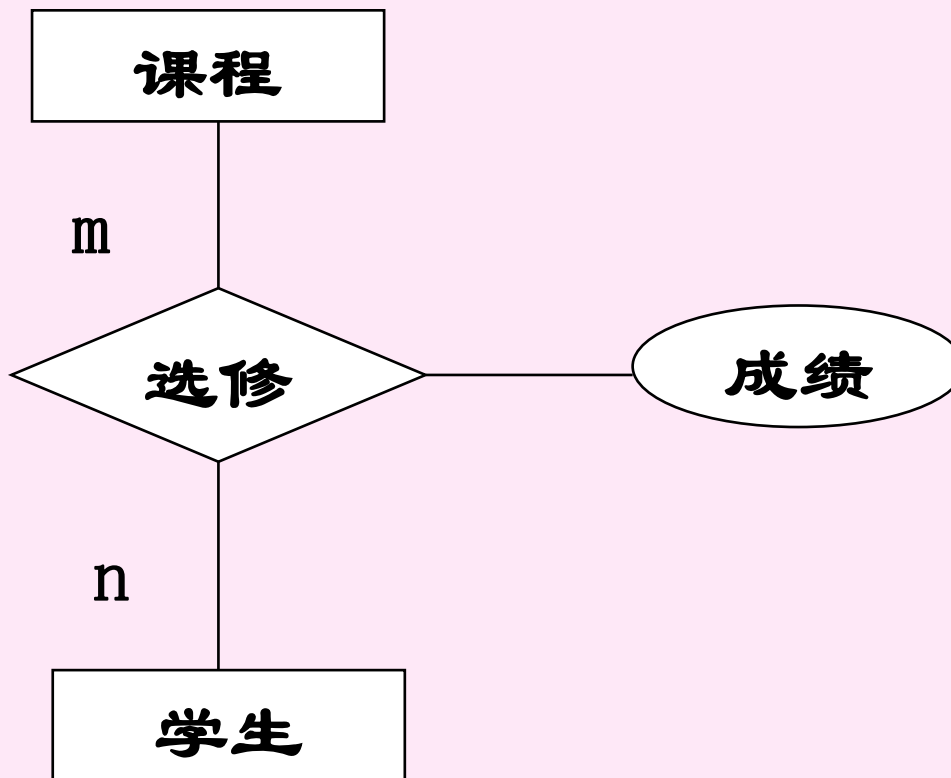


多个实体型间的1:n联系



同一实体型内部的1:n联系

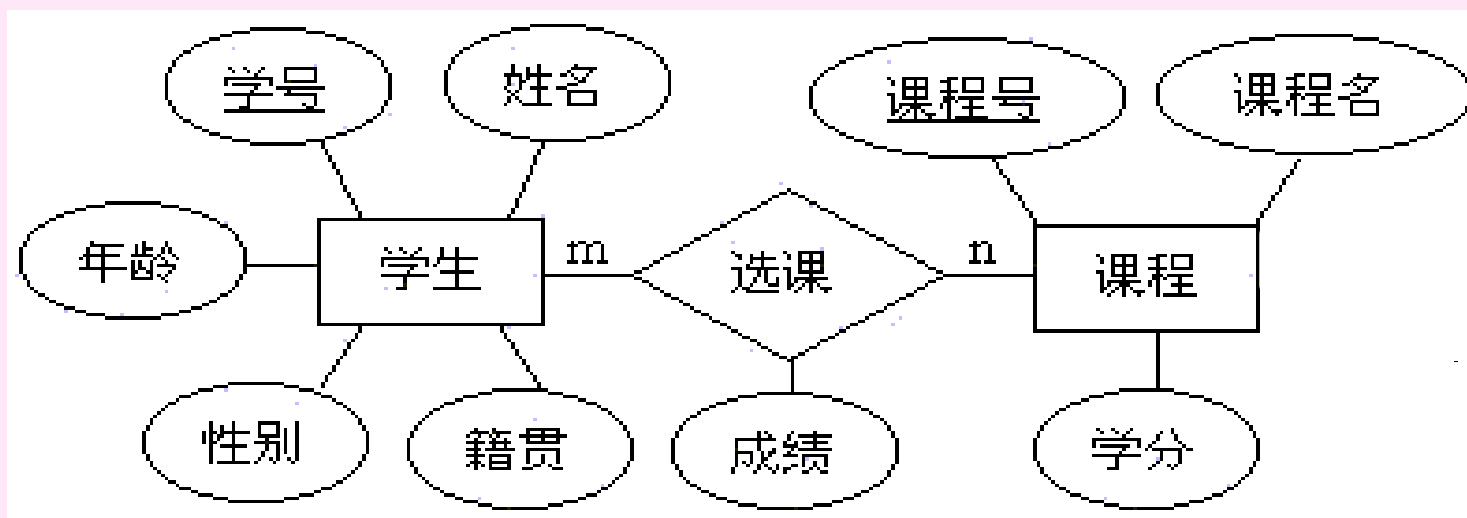
联系属性的表示方法



ER图练习



学生实体集中有学号、姓名、性别和年龄、籍贯等属性，课程实体集中有课程号、课程名和学分等属性；学生与课程之间联系名为选课，联系属性为成绩，用**ER**工具画出教学管理的概念模型，标示出实体标示符。





在简单的教务管理系统中，包括四个实体，分别为：

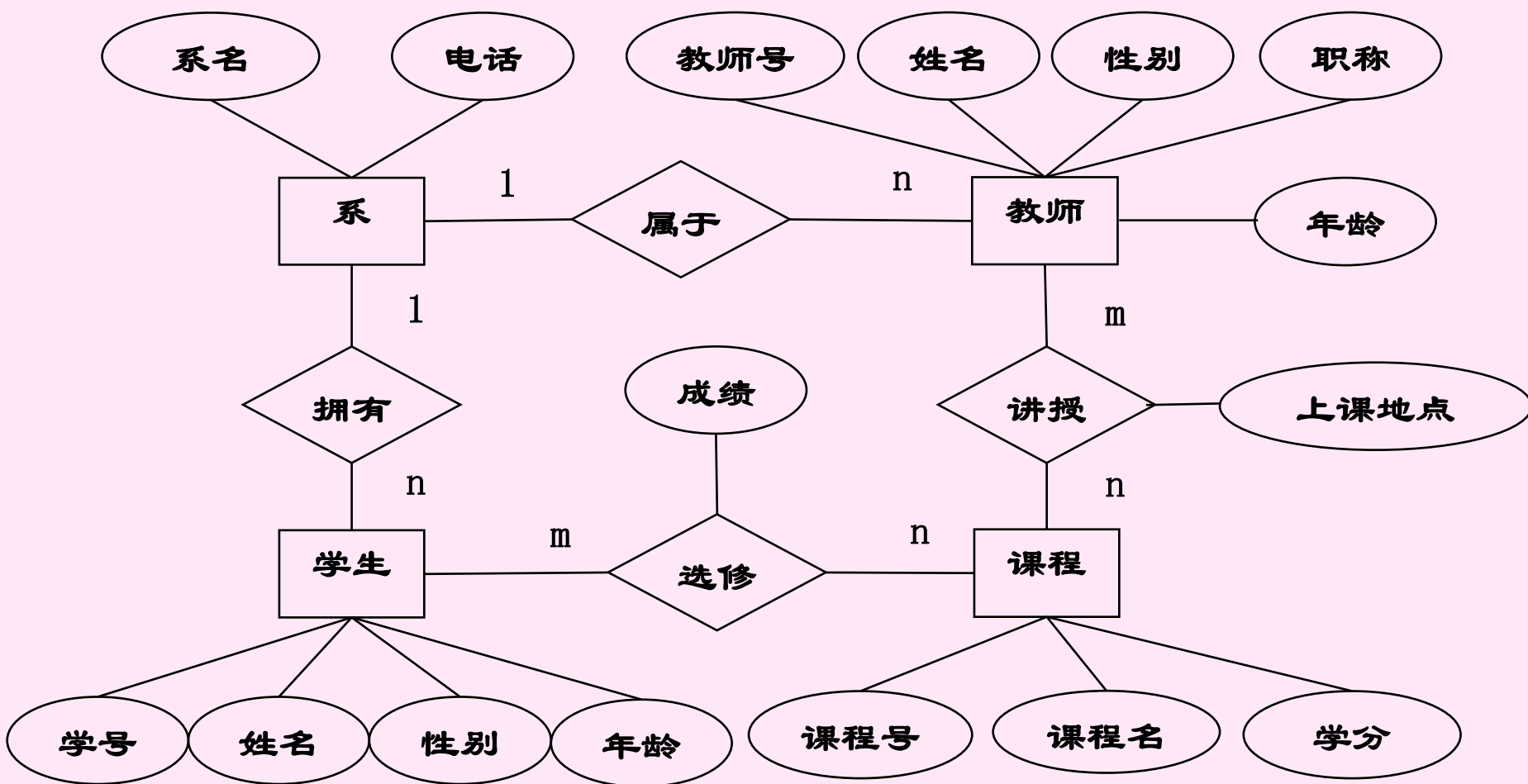
- 系：系名，电话
- 教师：教师号，姓名，性别，职称，年龄
- 学生：学号，姓名，性别，年龄
- 课程：课程号，课程名，学分

且存在如下语义约束：

- 一个系可拥有多个教师，一个教师只能属于一个系。
- 一个系可拥有多个学生，一个学生只能属于一个系。
- 一个学生可选修多门课程，一门课程可为多个学生选修，每一个学生选修每门课程都有一个成绩。
- 一个教师可讲授多门课程，一门课程可为多个教师讲授。

画出E-R图。

简单教务管理系统的E-R图



实体-联系模型的局限性



- **实体-联系模型是一种静态信息模型，只能反映当前状态，不能反映实体的变化过程。**
- **目前数据库多是采用实体-联系模型。**



知识点**15**: 层次模型

层次模型



- 层次模型是数据库系统中最早出现的数据模型，采用层次模型的数据库的典型代表是**IBM**公司的IMS (Information Management System) 数据库管理系统。
- 现实世界中，许多实体之间的联系都表现出一种很自然的层次关系，如家族关系，行政机构等。

层次数据模型的数据结构



■ 层次模型的数据结构：“有向树”

实体型：用**记录类型**描述。

每个结点表示一个记录类型。

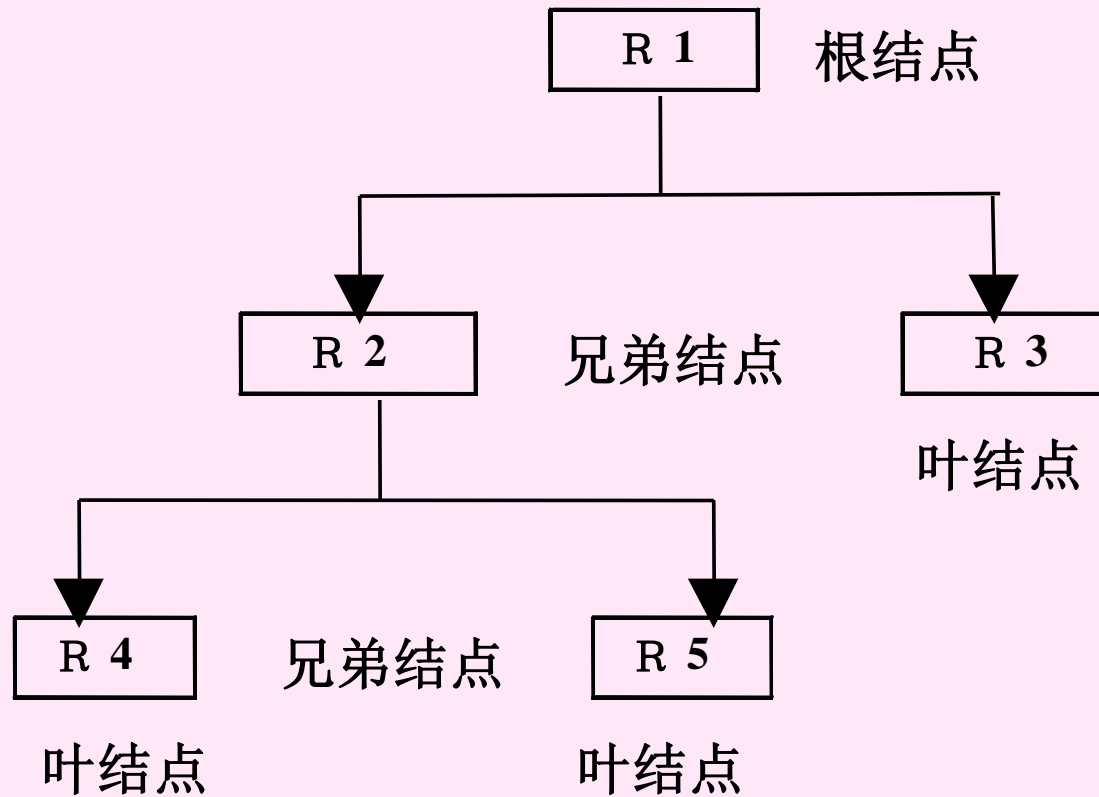
属性：用**字段**描述。每个记录类型可包含若干个字段。

联系：用**结点之间的连线（有向边）**表示记录类型之间的一对多的联系

■ 层次模型中的几个术语

■ 根结点，双亲结点，兄弟结点，叶结点

层次数据模型的数据结构（续）



层次数据模型的数据结构（续）



层次模型的特征：

- (1) 有且仅有一个结点没有双亲，该结点就是根结点；
- (2) 根结点以外的其他结点有且仅有一个双亲结点，这就使得层次数据库系统只能直接处理**一对多**的实体关系；



层次数据模型的数据结构(续)

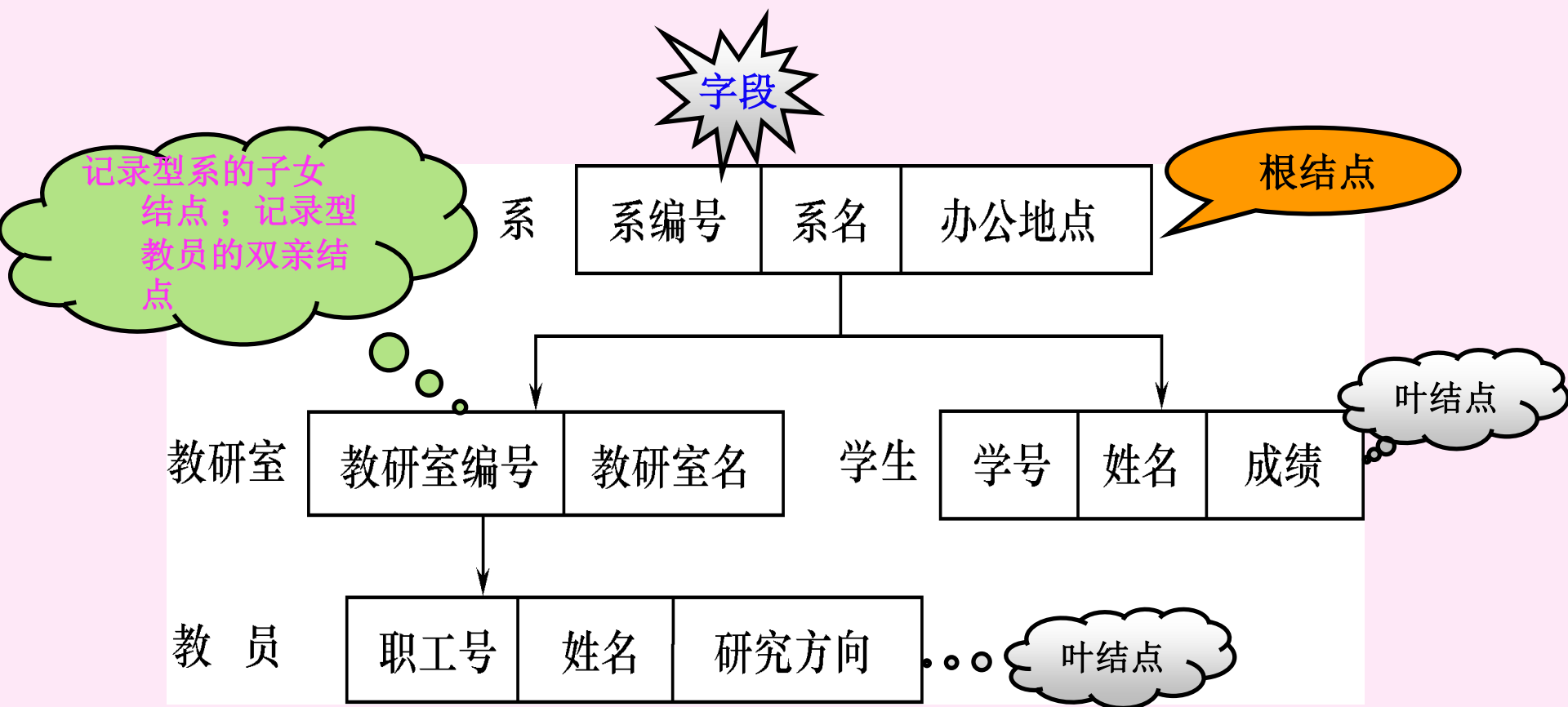


图1.17 教员学生层次数据库模型



层次数据模型的数据结构(续)

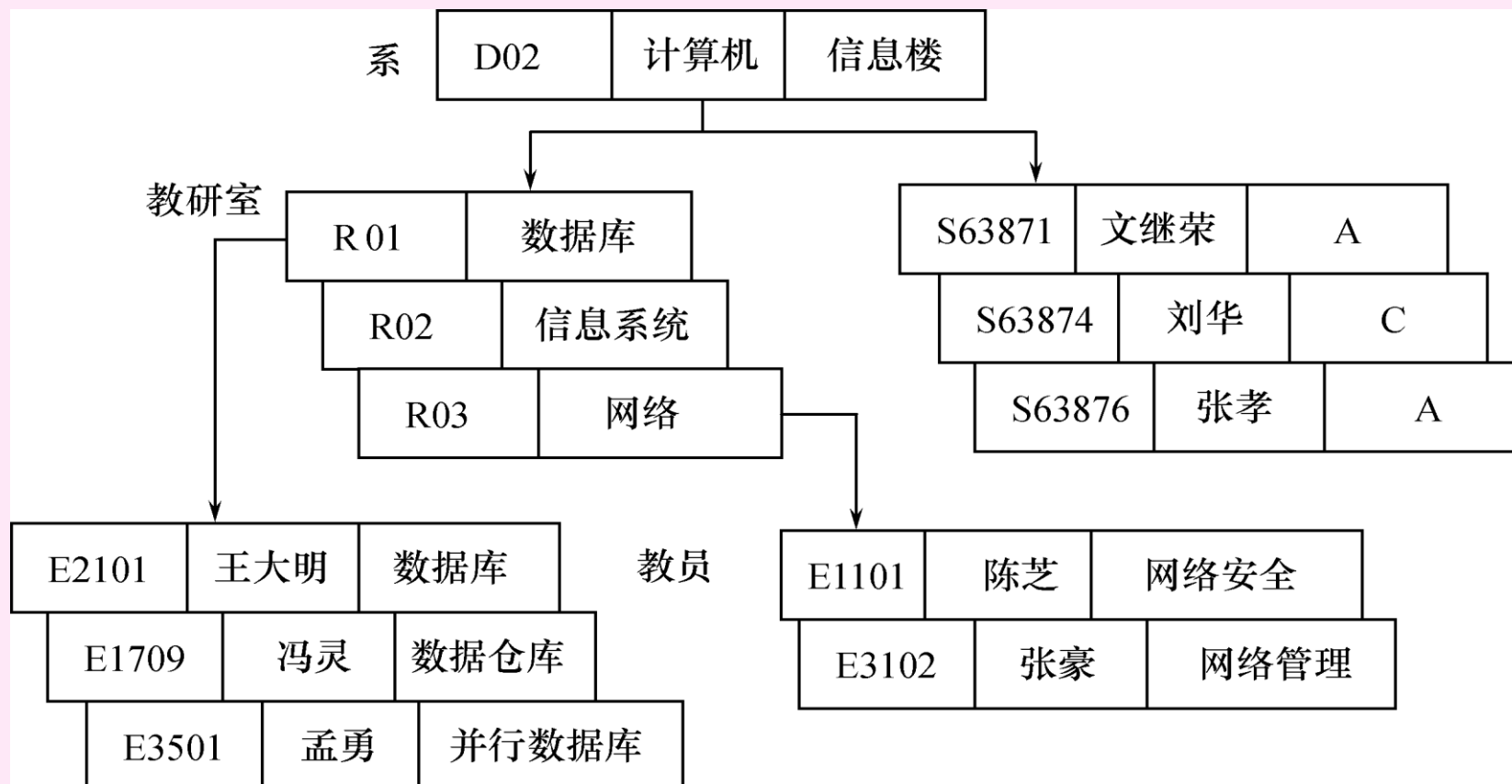


图1.18 教员学生层次数据库的一个值

层次模型的数据操纵与完整性约束（续）



❖ 层次模型的数据操作：

- ❖ 查询、添加、修改、删除

❖ 层次模型的完整性约束条件

- 无相应的双亲结点值就不能插入子女结点值
- 如果删除双亲结点值，则相应的子女结点值也被同时删除
- 更新操作时，应更新所有相应记录，以保证数据的一致性



层次模型的优缺点

❖ 优点

- 层次模型的数据结构比较简单清晰
- 查询效率高，性能优于关系模型，不低于网状模型
- 层次数据模型提供了良好的完整性支持

❖ 缺点

- 多对多联系表示不自然
- 对插入和删除操作的限制多，应用程序的编写比较复杂
- 查询子女结点必须通过双亲结点
- 由于结构严密，层次命令趋于程序化



知识点**16**: 网状模型

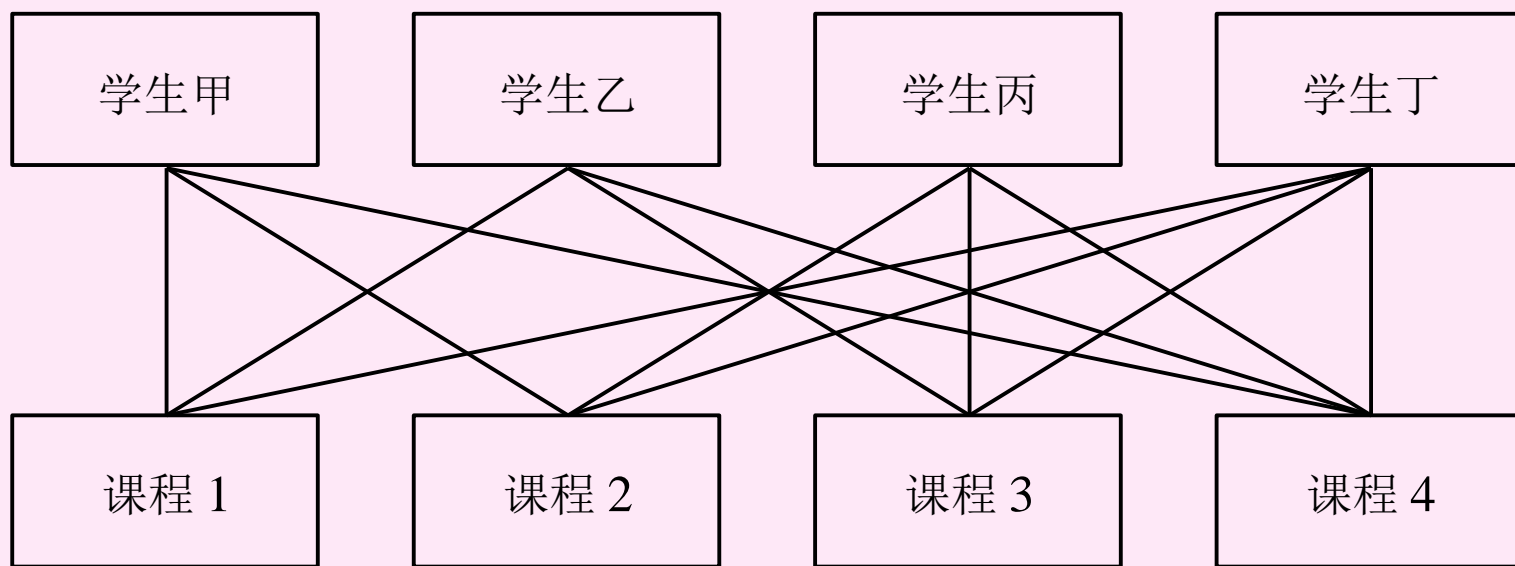
网状模型



- 20世纪70年代，数据系统语言研究会CODASYL（Conference On Data System Language）下属的数据库任务组DBTG（Data Base Task Group）提出了一个系统方案，**DBTG系统**，也称CODASYL系统，成为了网状数据模型的代表。
- 实际系统
 - Cullinet Software 公司的 IDMS
 - Univac公司的 DMS1100
 - Honeywell公司的IDS/2
 - HP公司的IMAGE



实 例



1.网状数据模型的数据结构



■ 网状模型

满足下面两个条件的基本层次联系的集合为网状模型。

1. 允许一个以上的结点无双亲；
2. 一个结点可以有多个的双亲。

网状数据模型的数据结构（续）



■ 表示方法（与层次数据模型相同）

实体型：用记录类型描述。

每个结点表示一个记录类型。

属性：用字段描述。

每个记录类型可包含若干个字段。

联系：用结点之间的连线表示记录（类）类型之间的一对多的父子联系。

网状数据模型的数据结构（续）



❖ 网状模型与层次模型的区别

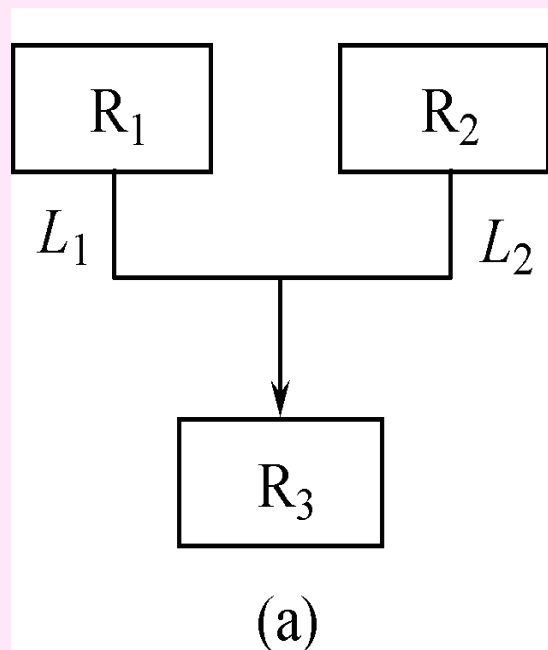
- 网状模型允许多个结点没有双亲结点
- 网状模型允许结点有多个双亲结点
- 网状模型允许两个结点之间有多种联系（复合联系）
- 网状模型可以更直接地去描述现实世界
- 层次模型实际上是网状模型的一个特例



网状数据模型的数据结构（续）

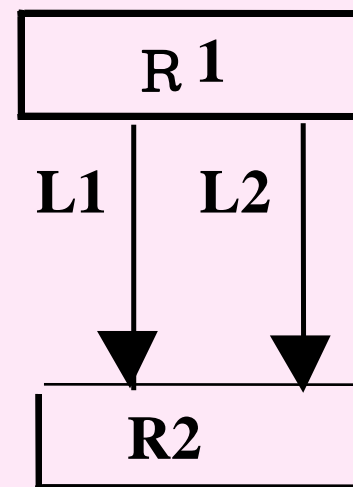
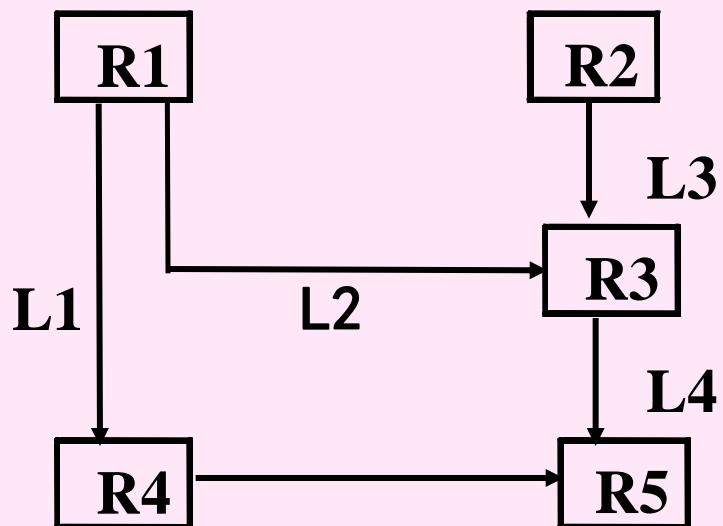
- 网状模型中子女结点与双亲结点的联系可以不唯一
- 要为每个联系命名，并指出与该联系有关的双亲记录和子女记录

R1与R3之间的
联系L1

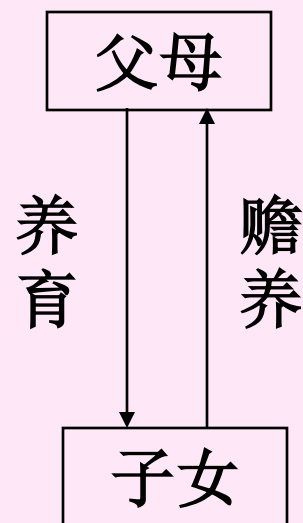
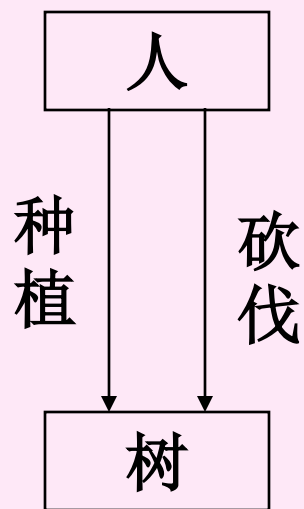


R2与R3之间的
联系L2

网状数据模型的数据结构



网状数据模型的数据结构（续）





网状数据模型的操纵与完整性约束（续）

- ❖ 网状数据库系统（如**DBTG**）对数据操纵加了一些限制，提供了一定的完整性约束
 - 码：唯一标识记录的数据项的集合
 - 一个联系中双亲记录与子女记录之间是一对多联系
 - 支持双亲记录和子女记录之间某些约束条件



网状数据模型的优缺点

❖ 优点

- 能够更为直接地描述现实世界，如一个结点可以有多个双亲
- 具有良好的性能，存取效率较高

❖ 缺点

- 结构比较复杂，而且随着应用环境的扩大，数据库的结构就变得越来越复杂，不利于最终用户掌握
- **DDL、DML**语言复杂，用户不容易使用



知识点**17**: 关系模型

关系模型



- 1970年美国IBM公司的研究员E.F.Codd首次提出了数据库系统的关系模型。
 - 他发表了题为“大型共享银行数据的关系模型” (A Relation Model of Data for Large Shared Data Banks), 在文中解释了关系模型, 定义了某些关系代数运算, 研究了数据的函数相关性, 定义了关系的第三范式, 从而开创了数据库的关系方法和数据规范化理论的研究, 为关系数据库技术奠定了理论基础。
 - 他因此获得了1981年的图灵奖。

关系模型（续）



- 20世纪80年代以来，计算机厂商新推出的数据库管理系统几乎都支持关系模型，非关系系统的产品也都加上了关系接口。
- 关系数据库已成为目前应用最广泛的数据库系统，如现在广泛使用的小型数据库系统Foxpro、Access，大型数据库系统Oracle、Informix、Sybase、SQL Server等都是关系数据库系统。



关系数据模型的数据结构

- ❖ 在用户观点下，关系模型中数据的逻辑结构是一张二维表，它由行和列组成。

学生登记表

学 号	姓 名	年 龄	性 别	系 名	年 级
2005004	王小明	19	女	社会学	2005
2005006	黄大鹏	20	男	商品学	2005
2005008	张文斌	18	女	法律	2005
...

属性

元组

关系模型的基本概念



■ 关系 (Relation)

一个关系对应通常说的一张表。

■ 元组 (Tuple)

表中的一行即为一个元组。

■ 属性 (Attribute)

表中的一列即为一个属性，给每一个属性起一个名称即属性名。

关系模型的基本概念



■ 码 (Key)

表中的某个属性 (组), 它可以唯一确定一个元组。

■ 域 (Domain)

属性的取值范围。

■ 分量

元组中的一个属性值。

■ 关系模式

对关系的描述

关系名 (属性1, 属性2, ..., 属性n)

学生 (学号, 姓名, 年龄, 性别, 系, 年级)

关系数据模型的数据结构(续)



■ 实体及实体间的联系的表示方法

- **实体型**：直接用关系（表）表示。
- **属性**：用属性名表示。
- **联系**：用键表示。

■ 在层次和网状模型中，联系用指针来实现，在关系模型中，记录之间的联系通过**键**来实现

■ 关系模型的操作语言是**SQL**



关系数据模型的数据结构（续）

❖ 关系必须是规范化的，满足一定的规范条件

最基本的规范条件：关系的每一个分量必须是一个不可分的数据项，
不允许表中还有表

图1.27中工资和扣除是可分的数据项，不符合关系模型要求

职工号	姓名	职 称	工 资			扣 除		实 发
			基 本	津 贴	职 务	房 租	水 电	
86051	陈 平	讲 师	1305	1200	50	160	112	2283

图1.27 一个工资表(表中有表)实例

关系数据模型的数据结构（续）



表1.2 术语对比

关系术语	一般表格的术语
关系名	表名
关系模式	表头（表格的描述）
关系	（一张）二维表
元组	记录或行
属性	列
属性名	列名
属性值	列值
分量	一条记录中的一个列值
非规范关系	表中有表（大表中嵌有小表）

关系数据模型的操纵与完整性约束



- ❖ 数据操作是集合操作，操作对象和操作结果都是关系
 - 查询 /插入 /删除 /更新
- ❖ 数据操作是集合操作，操作对象和操作结果都是关系，即若干元组的集合
- ❖ 存取路径对用户隐蔽，用户只要指出“干什么”，不必详细说明“怎么干”



关系数据模型的优缺点

❖ 优点

- 建立在严格的数学概念的基础上
- 概念单一
 - 实体和各类联系都用关系来表示
 - 对数据的检索结果也是关系
- 关系模型的存取路径对用户隐蔽
 - 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



关系数据模型的优缺点（续）

❖ 缺点

- 查询效率往往不如非关系数据模型
- 为提高性能，必须对用户的查询请求进行优化
增加了开发**DBMS**的难度