

第一章 绪论

1.1 数据库系统概述

1.2 数据模型

1.3 数据库系统结构

1.4 数据库系统的组成

1.5 小结

每一年级共有7个班，分为3个专业方向，每个班有若干学生，入学时，都会分配个学号，可以选修的课程(高等数学、英语、计算机基础、数据库...)，有很多教师，每个教师可以上某些课程，每个学生要选若干门课,.....



管理者： 学生情况、教师情况、 ...



学生： 选修课程、查成绩、 ...



教师： 查成绩、提交成绩, ...

现实世界是非常复杂的，必须予以抽象。

1.2 数据模型

- 数据模型是对现实世界数据特征的抽象。
- 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
- 数据模型是数据库系统的核心和基础

1.2 数据模型

1.2.1 两类数据模型

1.2.2 概念模型

1.2.3 数据模型的组成要素

1.2.4 常用的数据模型

1.2.5 层次模型

1.2.6 网状模型

1.2.7 关系模型

数据模型(续)

数据模型分成两个不同的层次

(1) **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，用于数据库设计。

(2) **逻辑模型和物理模型**

- 逻辑模型主要包括网状模型、层次模型、关系模型、面向对象数据模型、对象关系数据模型、半结构化数据模型等。按**计算机系统的观点**对数据建模，用于**DBMS实现**。
- 物理模型是对数据最底层的抽象，描述数据在系统内部的表示方式和存取方法，在磁盘或磁带上的存储方式和存取方法。是**面向计算机系统的**。

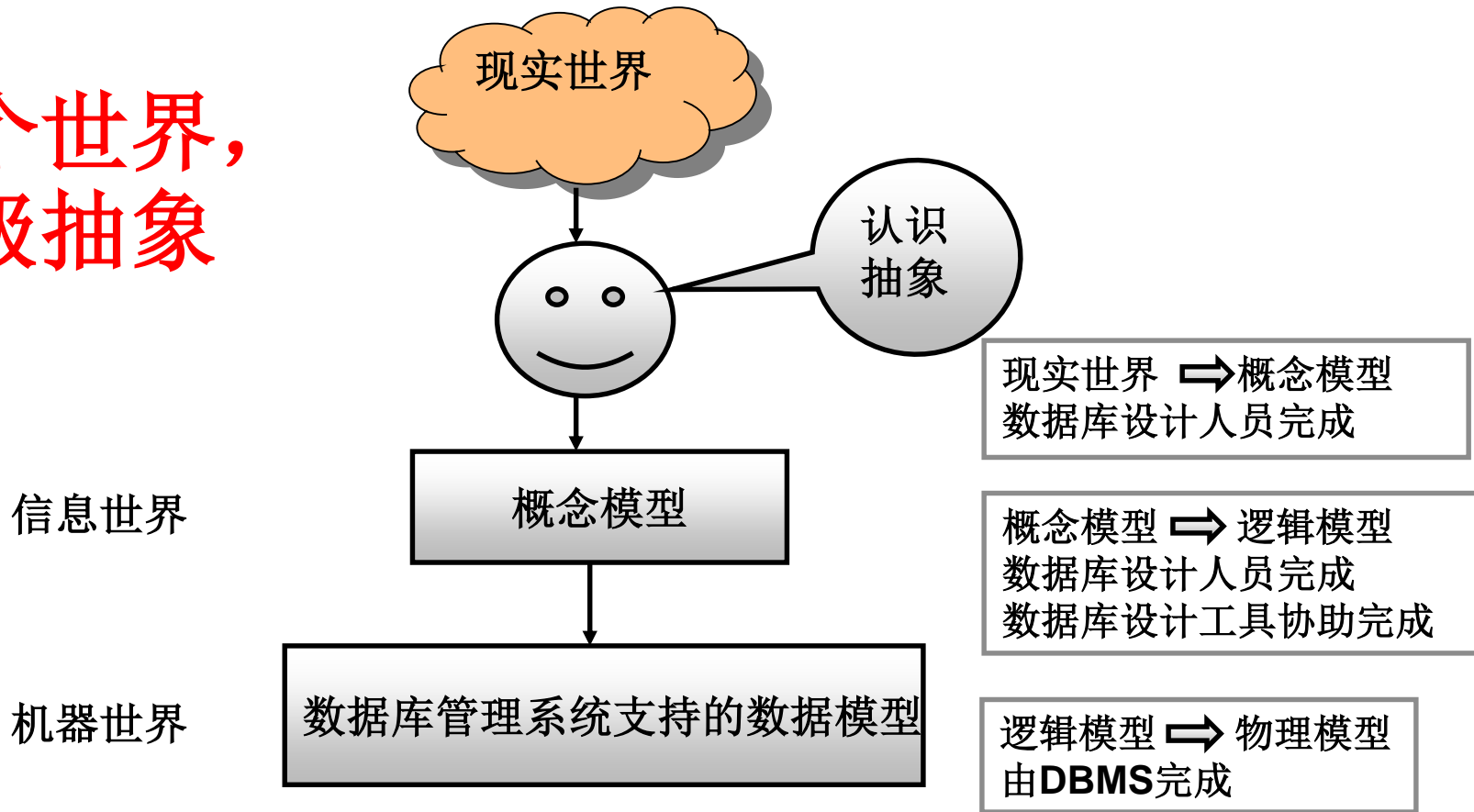
数据模型(续)

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(将现实世界抽象为信息世界);
- 把概念模型转换为某一DBMS支持的数据模型(将信息世界转换为机器世界)。

两类数据模型（续）

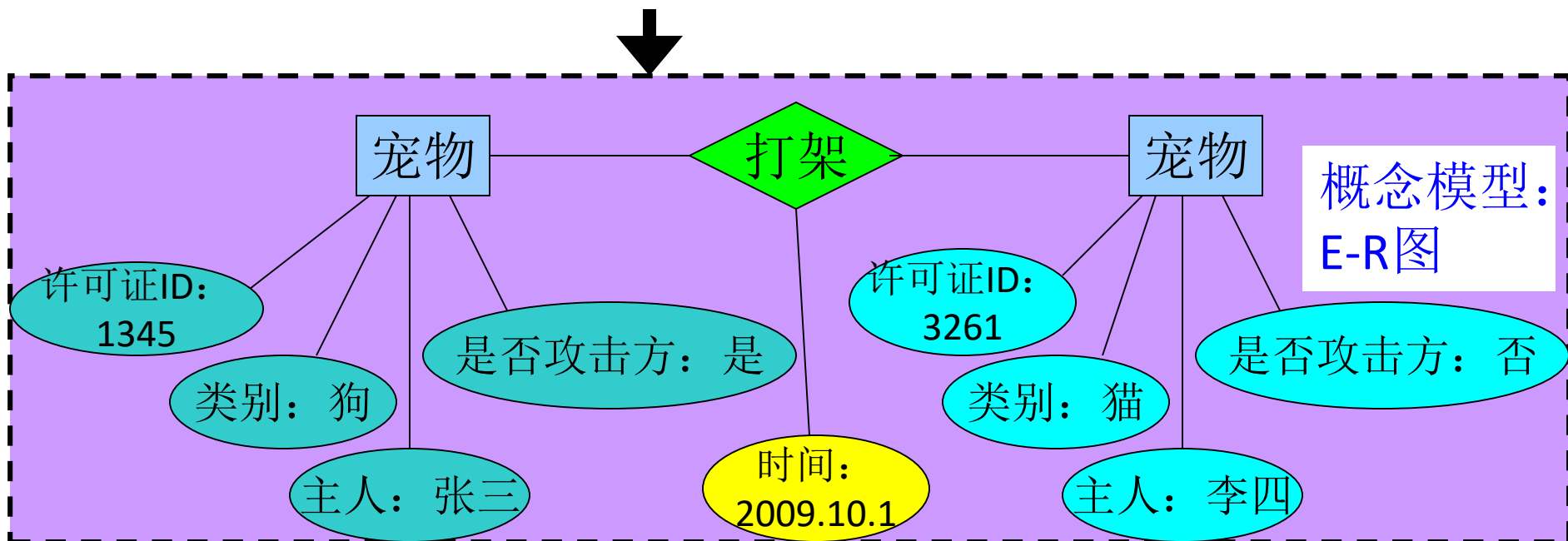
三个世界， 两级抽象



现实世界中客观对象的抽象过程

例如 **现实世界**的事件：
某年某月某日，在某小区，张三的狗咬了李四的猫

小区宠物
调查



小区宠物信息表

许可证 ID	类别	主人
1345	狗	张三
3261	猫	李四
...

小区宠物冲突记录表

事件 ID	时间	攻击方	被攻击方
0001	2009.10.1	1345	3261
0002

(DBMS支持)
数据模型：
关系模型

1.2 数据模型

1.2.1 两类数据模型

1.2.2 概念模型

1.2.3 数据模型的组成要素

1.2.4 常用的数据模型

1.2.5 层次模型

1.2.6 网状模型

1.2.7 关系模型

1.2.2 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
- 简单、清晰、易于用户理解

1.2.2 概念模型

- 1) 信息世界中的基本概念
- 2) 概念模型的一种表示方法：实体-联系方法

1) 信息世界中的基本概念

(1) 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。 某位学生

(2) 属性 (Attribute)

实体所具有的某一特性称为属性。

一个实体可以由若干个属性来刻画。

(学号、姓名、性别、出生年月、籍贯、所在系别、入学时间)

(3) 码 (Key)

唯一标识实体的属性集称为码。

1) 信息世界中的基本概念（续）

（4）实体型（Entity Type）

用实体名及其属性名集合来抽象和刻画同类实体称为实体型

学生（学号、姓名、性别、出生年月、籍贯、所在系别、入学时间）

（5）实体集（Entity Set）

全体学生

同一类型实体的集合称为实体集

1) 信息世界中的基本概念（续）

（6）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。
- 实体内部的联系通常是指组成实体的各属性之间的联系
- 实体之间的联系通常是指不同实体集之间的联系
- 实体之间的联系有一对一、一对多和多对多等多种类型

学院 (前面的例子)

实体名	属性	键/码	实体
班级	班号, 班级名称	班号	(S021, 1班)
学生	学号, 姓名, 性别	学号	(2008001, 张力, 男)
课程	课程代号, 课程名称	课程代号	(C005, 数据库)
教师	教师编号, 姓名	教师编号	(S9527, 王珊)

联系: **学生**在一个**班级**, **学生**选修**课程**, **教师**讲授**课程**,

实体集: **班级** { (S021, 1班), (S022, 2班), (S023, 3班) }
课程 { (C001, 高等数学), (C002, 大学英语) (C005, 数据库) }

2) 实体-联系方法

实体-联系方法（Entity-Relationship Approach）

- 用E-R图来描述现实世界的概念模型
- E-R方法也称为E-R模型

7.3.2 E-R模型

1. 实体之间的联系

(1) 两个实体型之间的联系：

① 一对一联系 ($1:1$)

② 一对多联系 ($1:n$)

③ 多对多联系 ($m:n$)

E-R模型（续）

①一对一联系（1：1）

- 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（也可以没有）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。
- 例如，学校里一个班级只有一个正班长，而一个班长只在一个班中任职，则班级与班长之间具有一对一联系。

E-R模型（续）

②一对多联系（1：n）

- 如果对于实体集A中的每一个实体，实体集B中有n个实体（ $n \geq 0$ ）与之联系，反之，对于实体集B中的每一个实体，实体集A中至多只有一个实体与之联系，则称实体集A与实体集B有一对多联系，记为1：n。
- 例如，一个班级中有若干名学生，而每个学生只在一个班级中学习，则班级与学生之间具有一对多联系。

E-R模型（续）

③多对多联系（ $m:n$ ）

- 如果对于实体集A中的每一个实体，实体集B中有 n 个实体（ $n \geq 0$ ）与之联系，反之，对于实体集B中的每一个实体，实体集A中也有 m 个实体（ $m \geq 0$ ）与之联系，则称实体集A与实体集B具有多对多联系，记为 $m:n$ 。
- 例如，一门课程同时有若干个学生选修，而一个学生可以同时选修多门课程，则课程与学生之间具有多对多联系。

E-R模型（续）

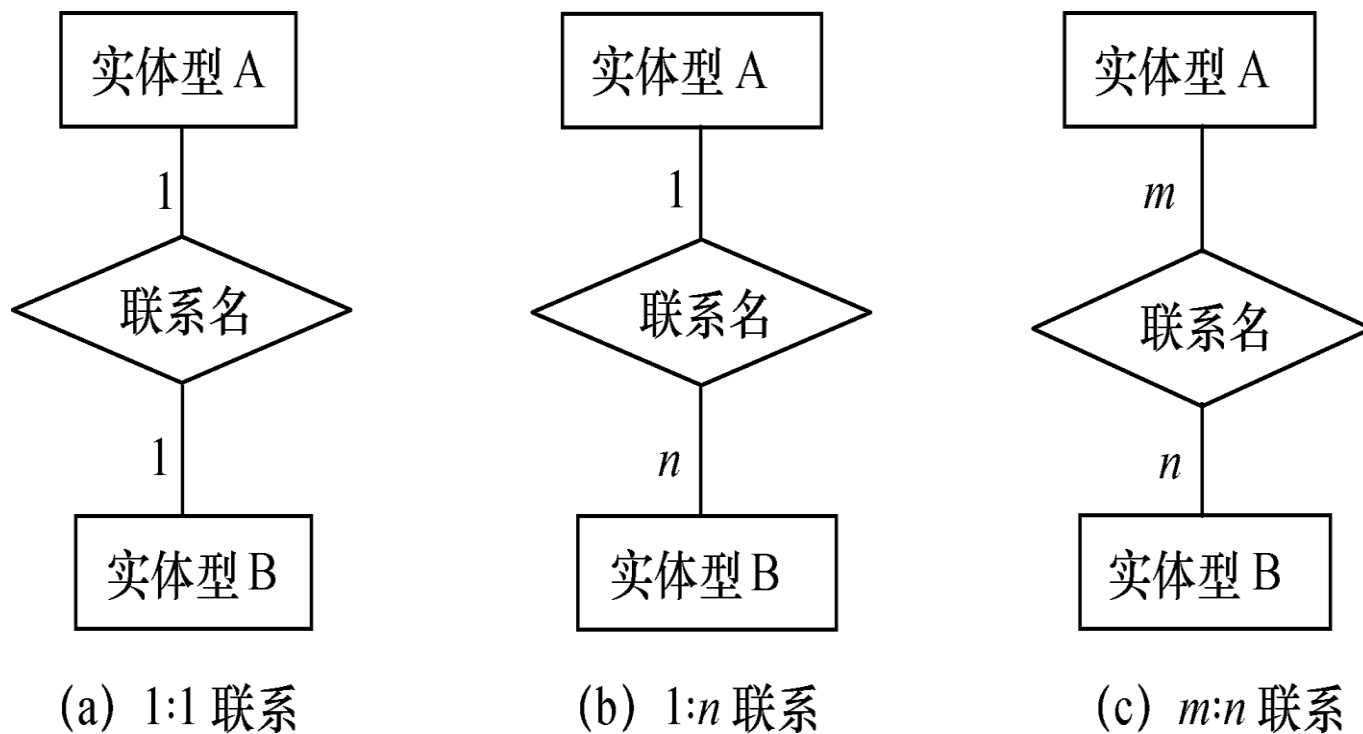


图7.6 两个实体型之间的三类联系

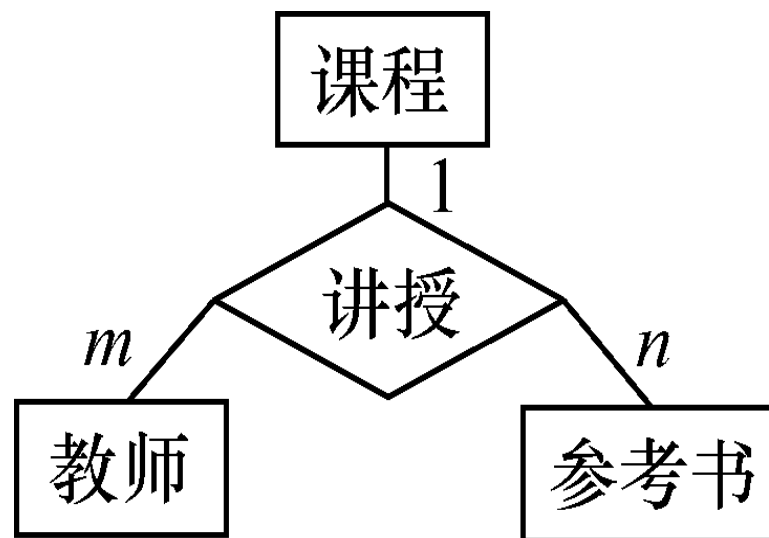
E-R模型（续）

（2）两个以上的实体型之间的联系

- 一般地，两个以上的实体型之间也存在着一对一、一对多、多对多联系。
- 对于课程、教师与参考书3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，而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。

则课程与教师、参考书之间的联系是一对多的，如图7.7(a)所示。

E-R模型（续）



(a)

图7.7 三个实体型之间的联系示例

E-R模型（续）

（3）单个实体型内的联系

- 同一个实体集内的各实体之间也可以存在一对一、一对多、多对多的联系。
- 例如，职工实体型内部具有领导与被领导的联系，即某一职工（干部）“领导”若干名职工，而一个职工仅被另外一个职工直接领导，因此这是一对多的联系，如图7.8所示。

E-R模型（续）

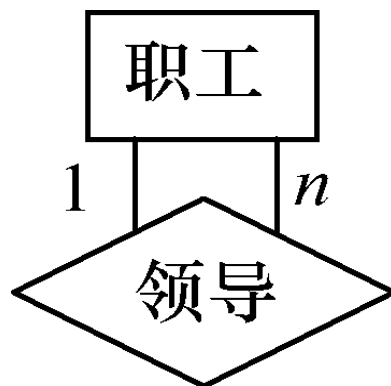


图7.8 单个实体型内的一对多联系示例

■联系的度：参与联系的实体型的数目

- 2个实体型之间的联系度为2，也称为二元联系；
- 3个实体型之间的联系度为3，称为三元联系；
- N个实体型之间的联系度为N，也称为N元联系


E-R模型（续）

2. E-R图

E-R图提供了表示实体型、属性和联系的方法：

实体型

- 用矩形表示，矩形框内写明实体名。



学生

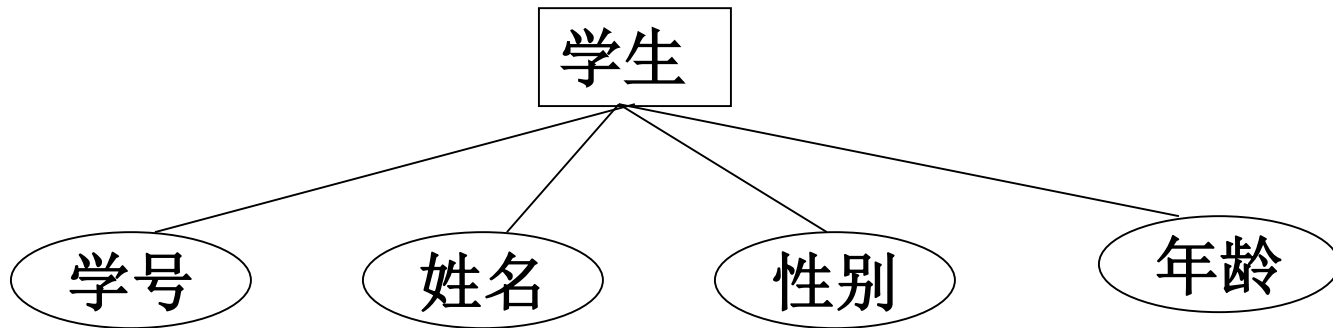


教师

E-R模型（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



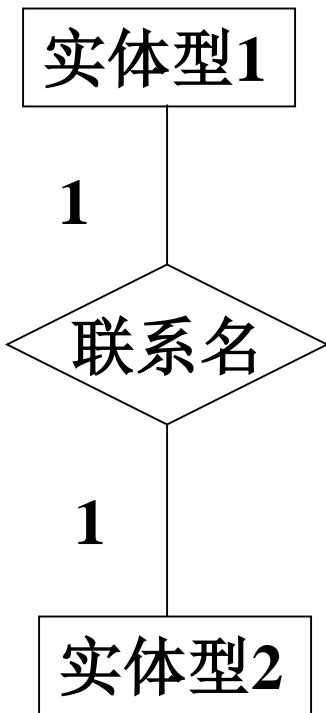
E-R模型（续）

联系

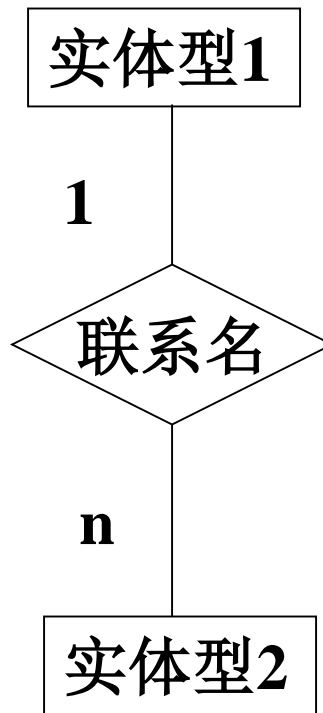
- **联系本身**：用菱形表示，菱形框内写明联系名，并用无向边分别与有关实体连接起来，同时无向边旁标上联系类型（1:1、1:n或m:n）
- **联系的属性**：联系本身也是一种实体型，也可以有属性。如果一个联系具有属性，则这些属性也要用无向边与该联系连接起来



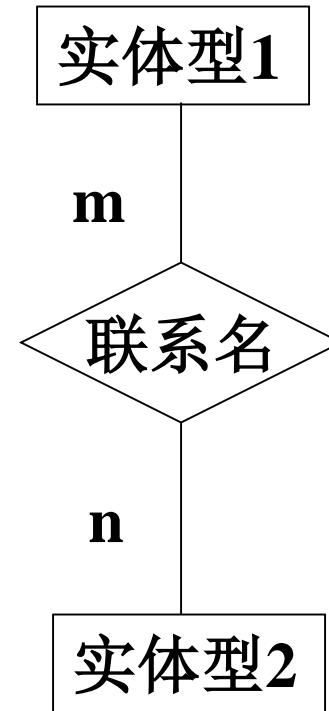
联系的表示方法



1:1联系



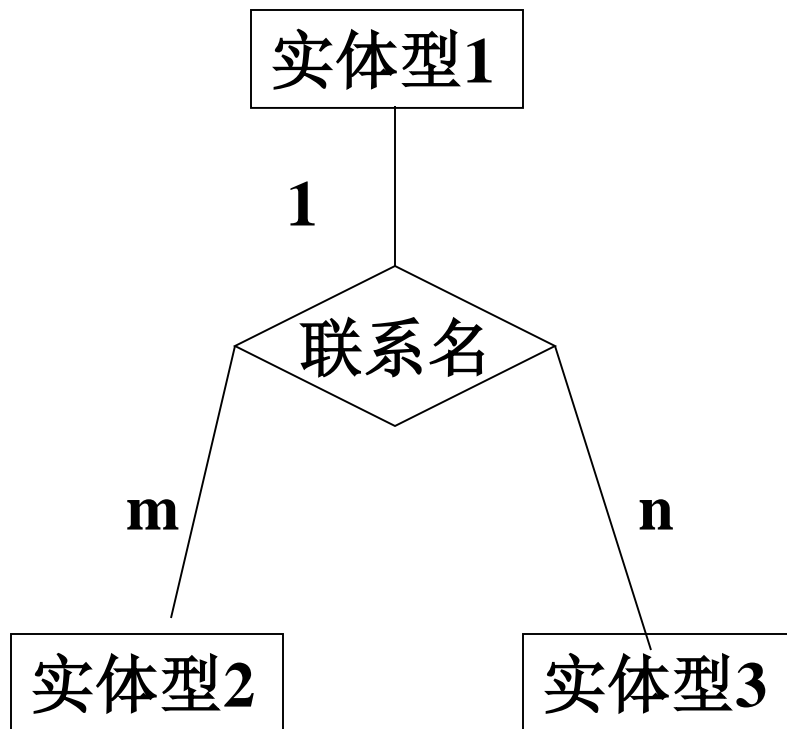
1:n联系



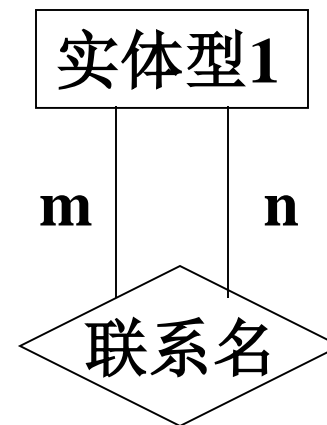
m:n联系



联系的表示方法



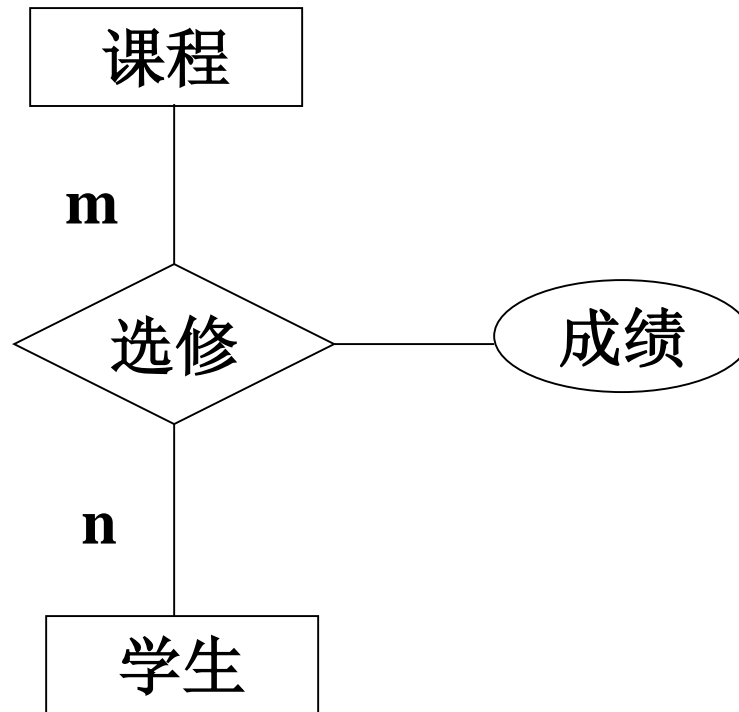
多个实体型间的1:n联系



同一实体型内部的m:n联系



联系属性的表示方法



E-R图实例

利用**E-R**图来表达某工厂物资管理概念模型（P218）

第一步：请分析，工厂的物资管理涉及到哪些实体？

第二步：这些实体分别都有哪些属性？

人？物？事？

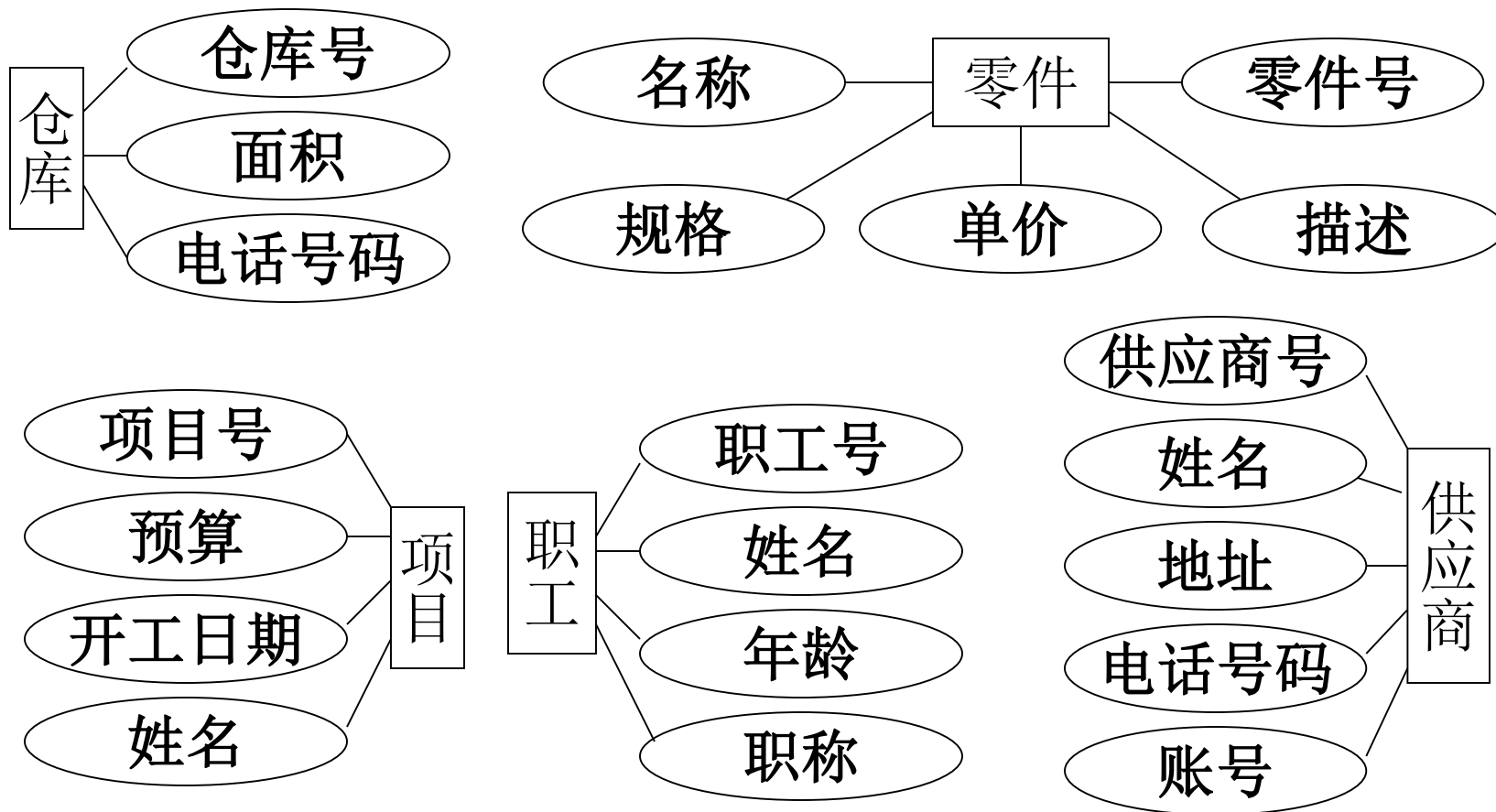
如何标识、什么特征、如何刻画？

仓库
零件
供应商
项目
职工

- 仓库号、面积、电话号码
- 零件号、名称、规格、单价、描述
- 供应商号、姓名、地址、电话号码、账号
- 项目号、预算、开工日期
- 职工号、姓名、年龄、职称

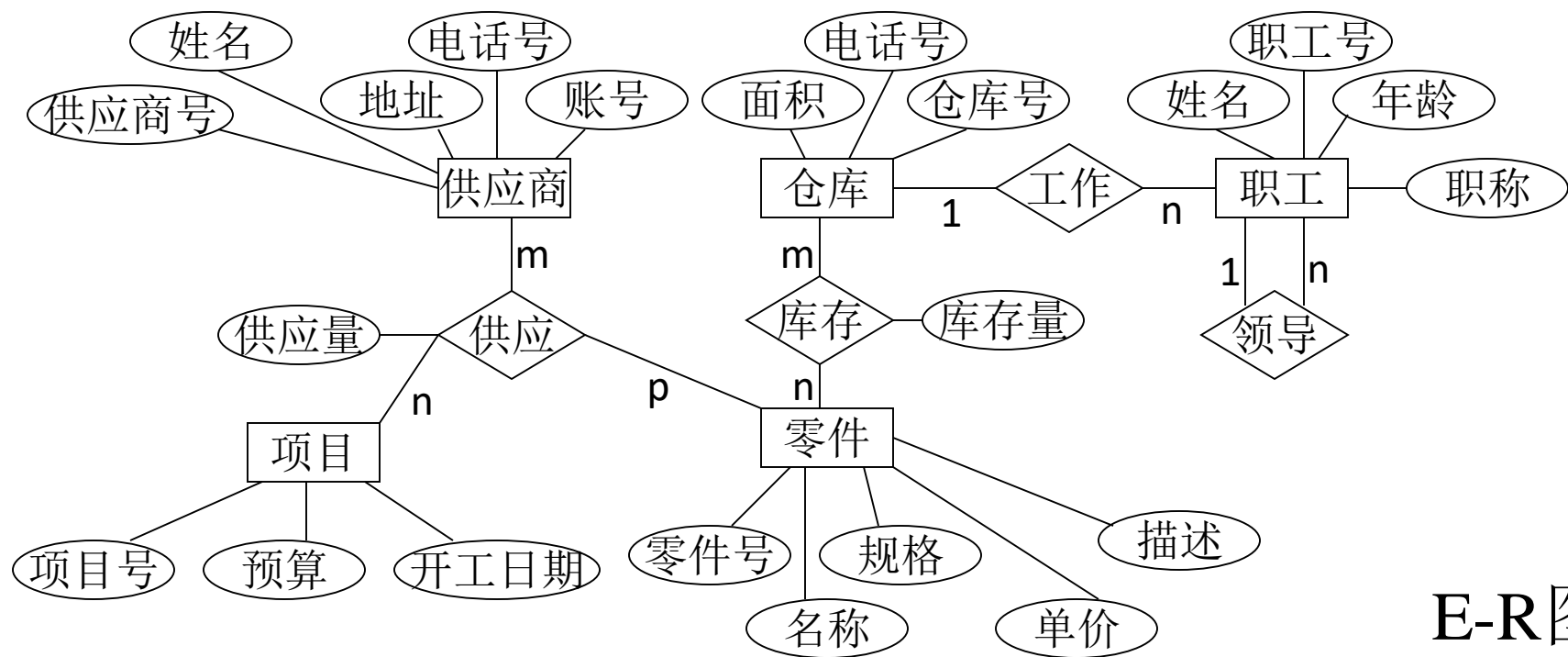
E-R图实例

实体及其属性



E-R图实例

第三步：这些实体间存在何种联系？



E-R图

1.2 数据模型

1.2.1 两类数据模型

1.2.2 概念模型

1.2.3 数据模型的组成要素

1.2.4 常用的数据模型

1.2.5 层次模型

1.2.6 网状模型

1.2.7 关系模型

1.2.3 数据模型的组成要素

- 1) 数据结构
- 2) 数据操作
- 3) 数据的完整性约束条件

1) 数据结构

数据模型的数据结构

- 描述数据库的组成对象，以及对象之间的联系

描述的内容

- 与对象的类型、内容、性质有关（实体）
- 与数据之间联系有关（联系）

数据结构是对系统静态特性的描述

2) 数据操作

数据操作是对数据库中各种对象（**型**）的实例（**值**）允许执行的**操作的集合**，包括操作及有关的操作规则

数据操作的类型

- 查询
- 更新（包括插入、删除、修改）

2) 数据操作（续）

数据模型必须对操作有如下定义

- 操作的确切含义
- 操作符号
- 操作规则（如优先级）
- 实现操作的语言

数据操作是对系统动态特性的描述

3) 数据的完整性约束条件

- 数据的完整性约束条件是一组完整性规则的集合
- 完整性规则：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（一致性）

3) 数据的完整性约束条件（续）

- 数据模型反映和规定必须遵守的基本的通用的完整性约束条件。

例如在关系模型中，任何关系必须满足实体完整性和参照完整性两个条件。

- 数据模型提供定义完整性约束条件的机制，以反映具体应用所涉及的数据必须遵守的特定的语义约束条件。

（性别：男或女） （夫妻关系为1对1联系）

1.2 数据模型

1.2.1 两类数据模型

1.2.2 概念模型

1.2.3 数据模型的组成要素


1.2.4 常用的数据模型

1.2.5 层次模型

1.2.6 网状模型

1.2.7 关系模型

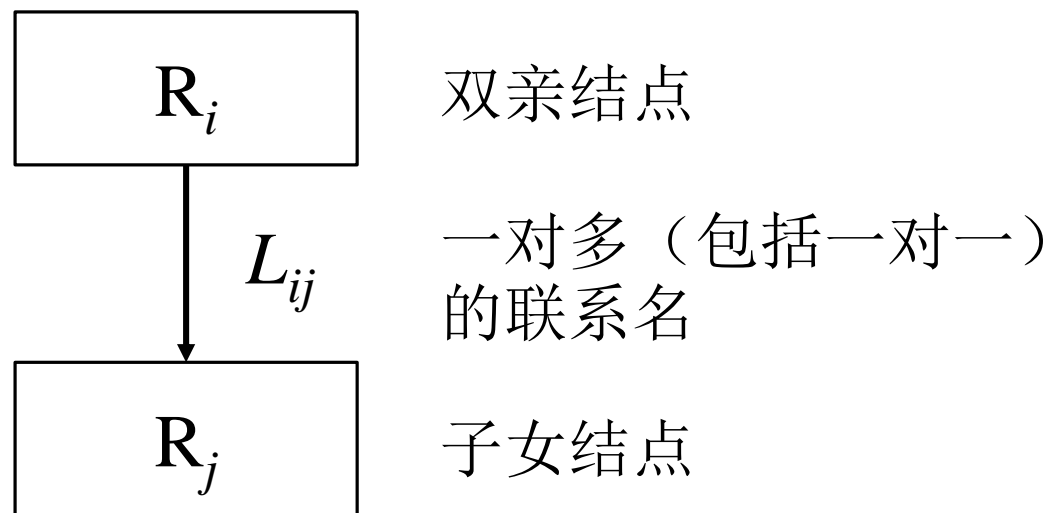
1.2.4 常用的数据模型

- 层次模型 (Hierarchical Model)
 - 网状模型 (Network Model)
 - 关系模型 (Relational Model)
 - 面向对象数据模型 (Object Oriented Data Model)
 - 对象关系数据模型 (Object Relational Data Model)
 - 半结构化数据模型 (Semistruture Data Model)
- 
- A diagram consisting of a right-facing curly bracket grouping the first two items of the list, '层次模型 (Hierarchical Model)' and '网状模型 (Network Model)'. To the right of the bracket is a rectangular box containing the text '格式化模型' (Formatted Model).

1.2.4 常用的数据模型（续）

格式化模型中基本层次联系：

两个记录及它们之间的一对多（包括一对一）的联系



1.2 数据模型

1.2.1 两类数据模型

1.2.2 概念模型

1.2.3 数据模型的组成要素

1.2.4 常用的数据模型

1.2.5 层次模型

1.2.6 网状模型

1.2.7 关系模型

1.2.5 层次模型

- 层次模型是数据库系统中最早出现的数据模型
- 层次数据库系统的典型代表是IBM公司的IMS（Information Management System）数据库管理系统
- 层次模型用树形结构来表示各类实体以及实体间的联系

1.2.5 层次模型

- 1) 层次模型的数据结构
- 2) 层次模型的数据操纵与完整性约束
- 3) 层次模型的优缺点

1) 层次模型的数据结构

- 满足下面两个条件的**基本层次联系的集合**为**层次模型**

- ✓ **有且只有一个**结点没有双亲结点，这个结点称为**根结点**

- ✓ 根以外的其它结点有且只有一个双亲结点

- 层次模型中的几个术语

- ✓ 根结点，双亲结点，兄弟结点，叶结点

1) 层次模型的数据结构（续）

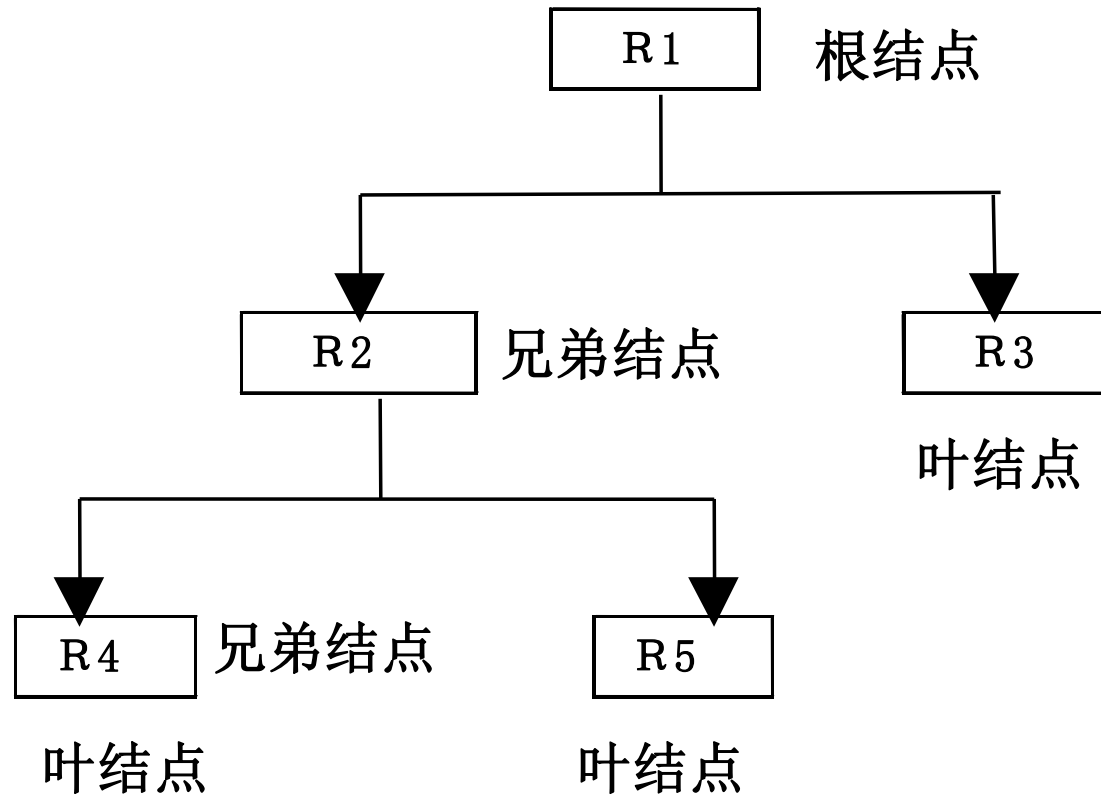


图1.9 一个层次模型的示例

1) 层次模型的数据结构（续）

层次模型的特点：

- 结点的双亲是唯一的
- 只能直接处理一对多的实体联系
- 每个记录类型可以定义一个排序字段，也称为码字段
- 任何记录值只有按其路径查看时，才能显出它的全部意义
- 没有一个子女记录值能够脱离双亲记录值而独立存在

1) 层次模型的数据结构（续）

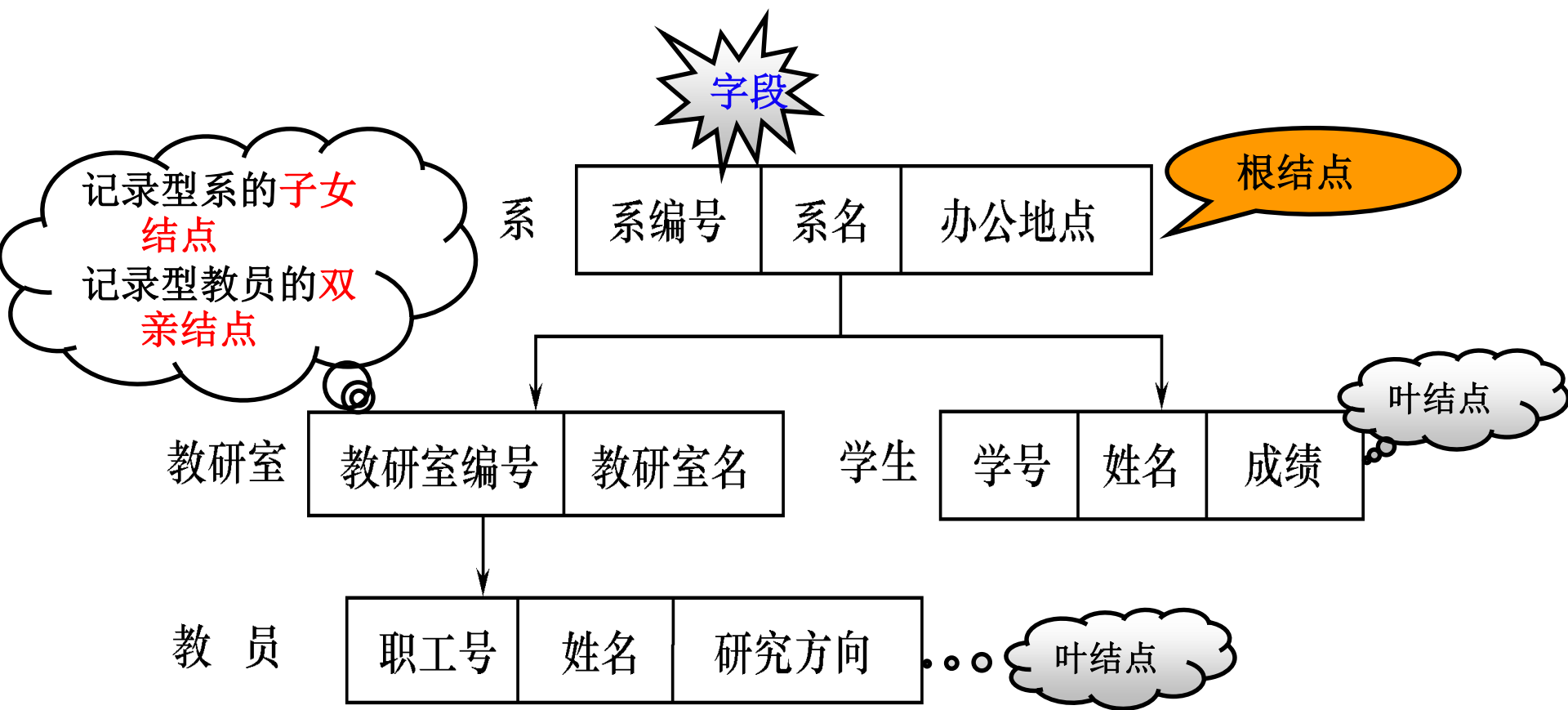


图1.10 教员学生层次数据库模型

1) 层次模型的数据结构（续）

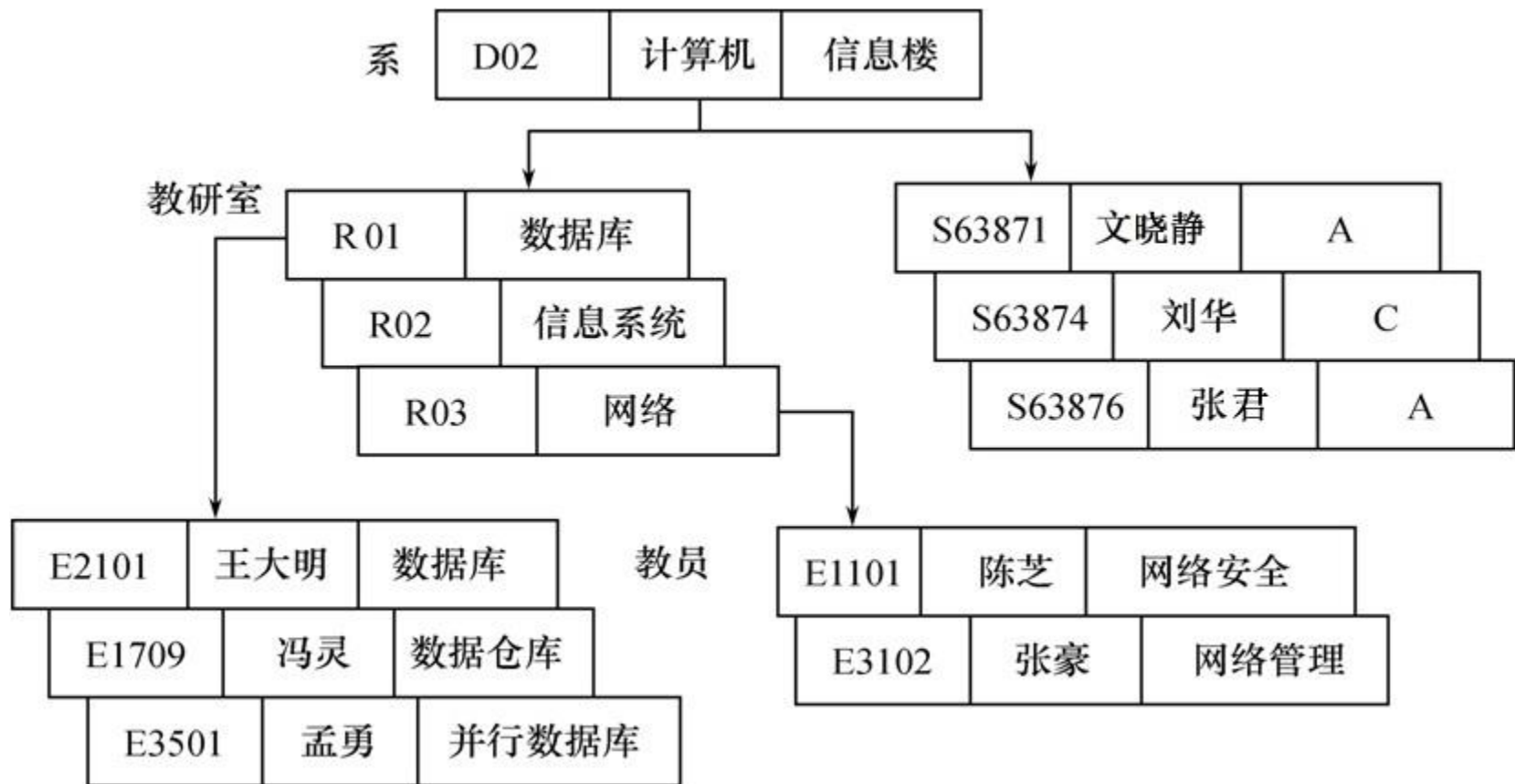


图1.11 教员学生层次数据库的一个值

2) 层次模型的数据操纵与完整性约束

层次模型的数据操纵

- 查询
- 插入
- 删除
- 更新

2) 层次模型的数据操纵与完整性约束

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
(新调来一个教师, 还没确定划归哪个教研室)
- 如果删除双亲结点值, 则相应的子女结点值也被同时删除
(删除教研室信息, 则教研室所有教师信息被删除)
- 更新操作时, 应更新所有相应记录, 以保证数据的一致性

3) 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

- 结点之间的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
- 层次命令趋于程序化

1.2 数据模型

1.2.1 两类数据模型

1.2.2 概念模型

1.2.3 数据模型的组成要素

1.2.4 常用的数据模型

1.2.5 层次模型

1.2.6 网状模型

1.2.7 关系模型

1.2.6 网状模型

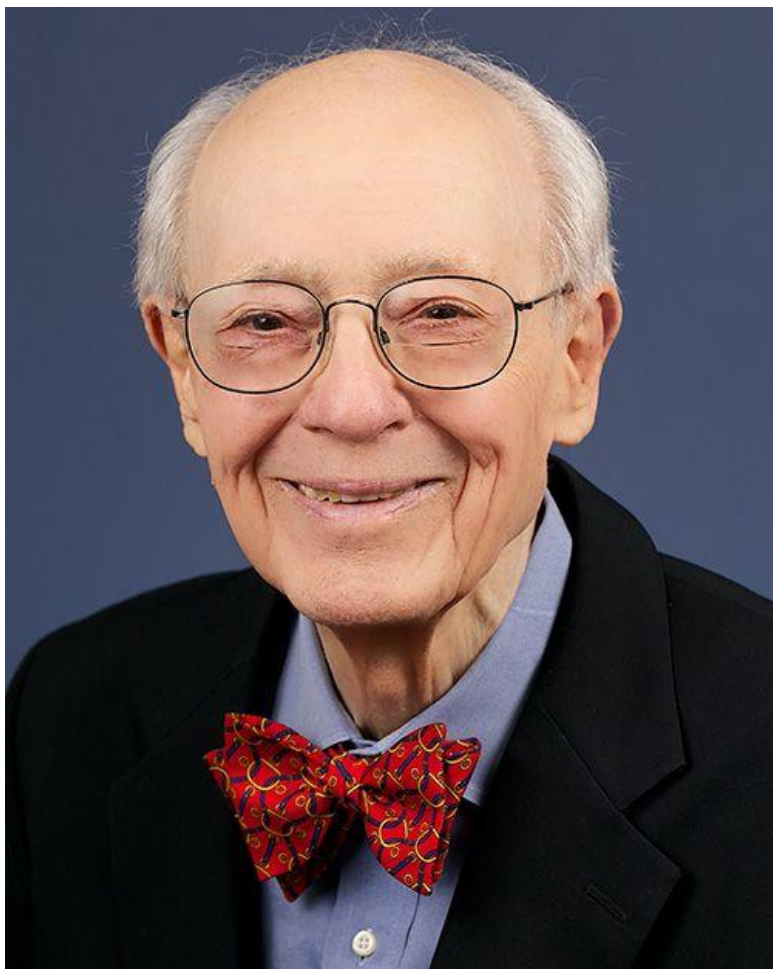
网状数据库系统采用网状模型作为数据的组织方式

典型代表是DBTG系统：

- 亦称CODASYL系统
- 20世纪70年代由DBTG提出的一个系统方案

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



Charles.W.Bachman
1924-2017

- 1960年为通用电气制造了世界上第一个网状数据库系统IDS
- 积极推动与促成了数据库标准的制定：DBTG报告
- 由于他在在数据库技术的产生、发展与推广应用方面都发挥了巨大的作用
1973获图灵奖

1.2.6 网状模型

- 1) 网状模型的数据结构
- 2) 网状模型的数据操纵与完整性约束
- 3) 网状模型的优缺点

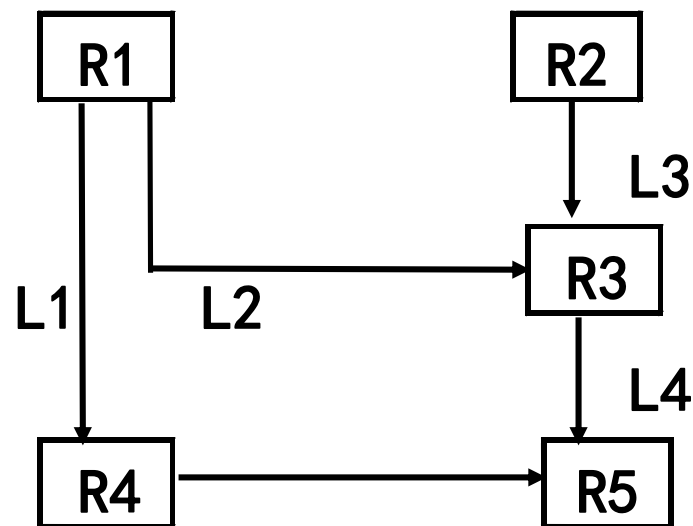
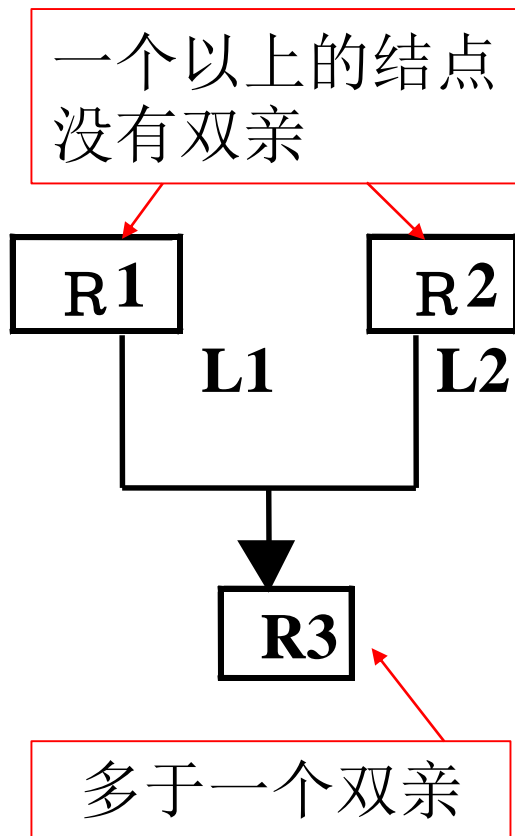
1) 网状模型的数据结构

网状模型

满足下面两个条件的**基本层次联系**的集合：

- 允许一个以上的结点无双亲；
- 一个结点可以有多个的双亲。

1) 网状模型的数据结构（续）



1) 网状模型的数据结构（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（实体）。




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型（实体）之间的一对多的父子联系。

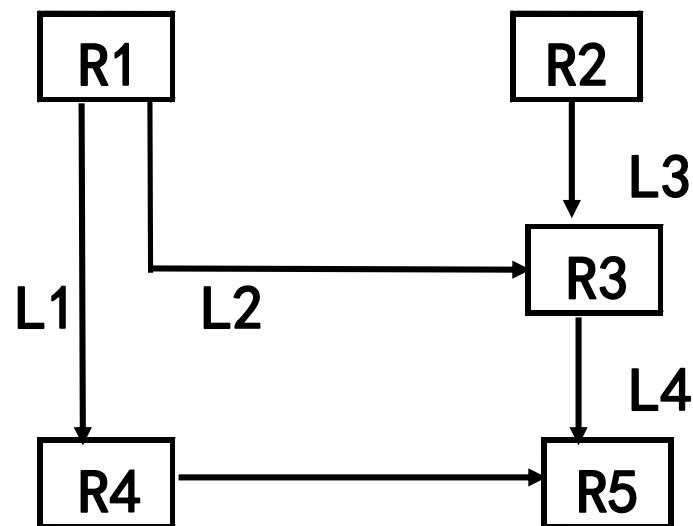
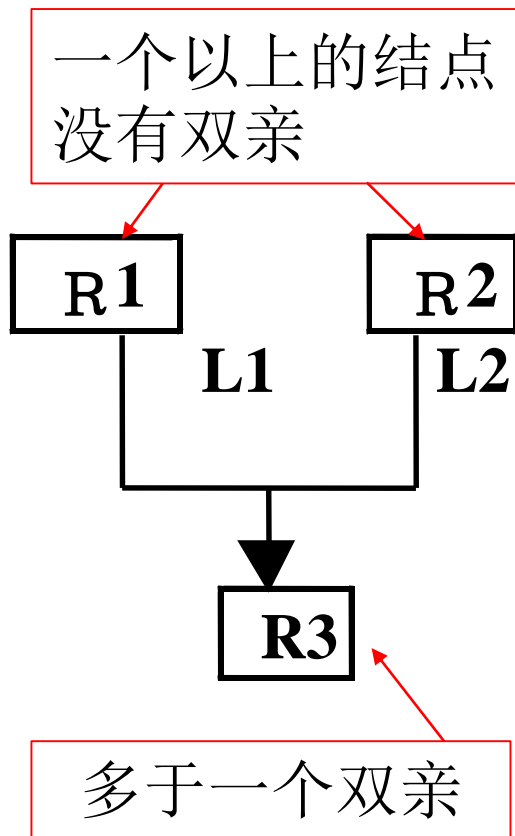
1) 网状数据模型的数据结构（续）

网状模型与层次模型的区别

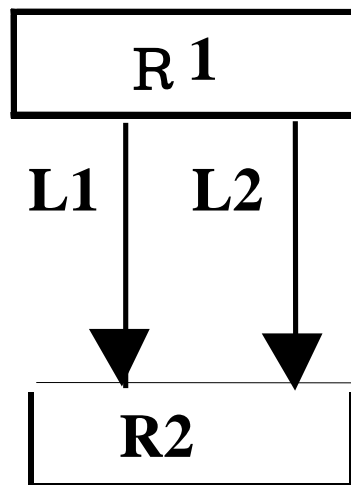
- 网状模型允许多个结点没有双亲结点 
- 网状模型允许结点有多个双亲结点 
- 网状模型允许两个结点之间有多种联系（复合联系） 
- 网状模型可以更直接地去描述现实世界
- 层次模型实际上是网状模型的一个特例



1) 网状模型的数据结构（续）



1) 网状数据模型的数据结构（续）



1) 网状数据模型的数据结构（续）

多对多联系在网状模型中的表示

- 用网状模型间接表示多对多联系

- 方法：

将多对多联系直接分解成一对多联系

1) 网状数据模型的数据结构（续）

例如：一个学生可以选修若干门课程，某一课程可以被多个学生选修，**学生与课程之间是多对多联系**

● 引进一个学生选课的联系记录，由3个数据项组成

✓ 学号

✓ 课程号

✓ 成绩

=》表示某个学生选修某一门课程及其成绩

1) 网状数据模型的数据结构（续）

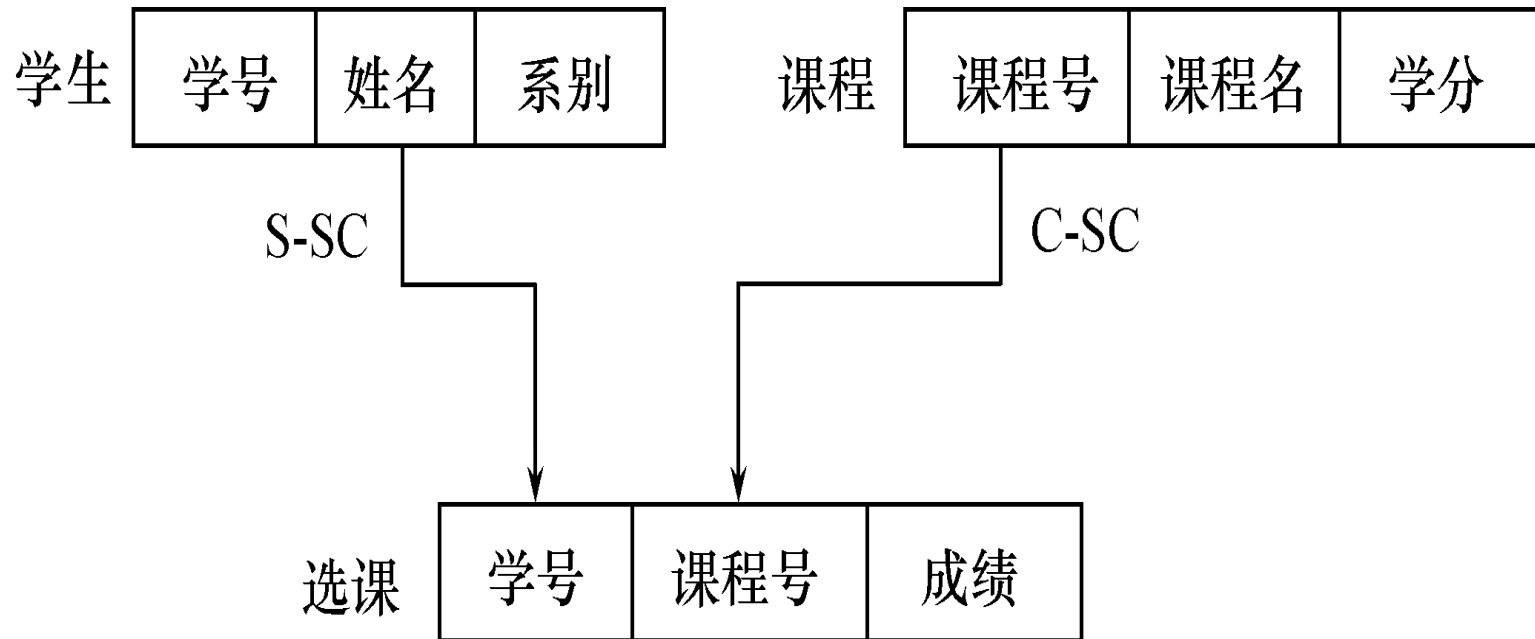


图1.13 学生/选课/课程的网状数据模型

2) 网状模型的操纵与完整性约束

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
（双亲记录存在才能插入子女记录，双亲记录删除时也连同删除子女记录）

3) 网状模型的优缺点

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
- 记录之间联系是通过存取路径实现的，用户必须了解系统结构的细节