

数据库系统原理

教程：数据库系统概论（第5版）

结合：CMU 15-445/645 INTRO TO DATABASE SYSTEMS

华中科技大学 计算机学院

左琼

1.2 数据模型

1.2.1 两大类数据模型

1.2.2 概念模型

1.2.3 数据模型三要素

1.2.4 最常用的数据模型

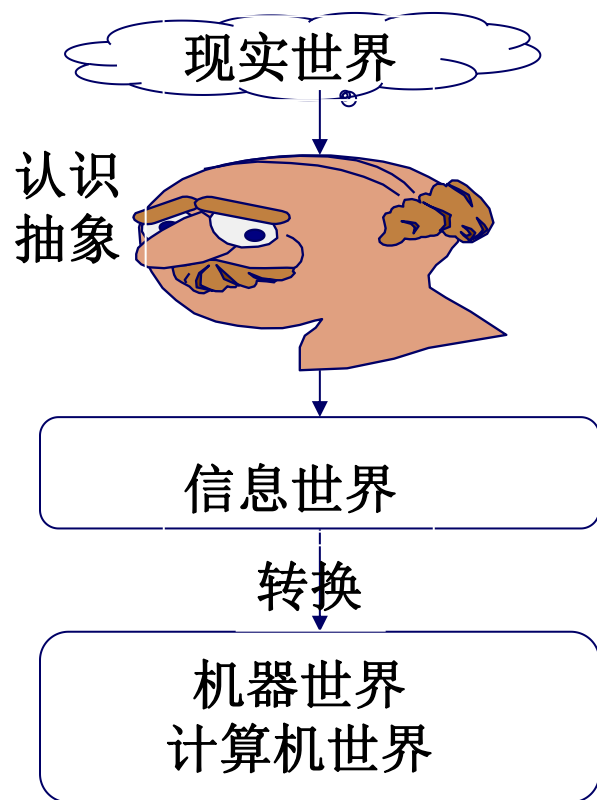
1.2.5 层次模型

1.2.6 网状模型

1.2.7 关系模型

1.2 数据模型 (Data Model)

- **数据模型**：是一种描述数据、数据联系、数据语义及一致性约束的抽象工具，对现实世界数据特征的描述。

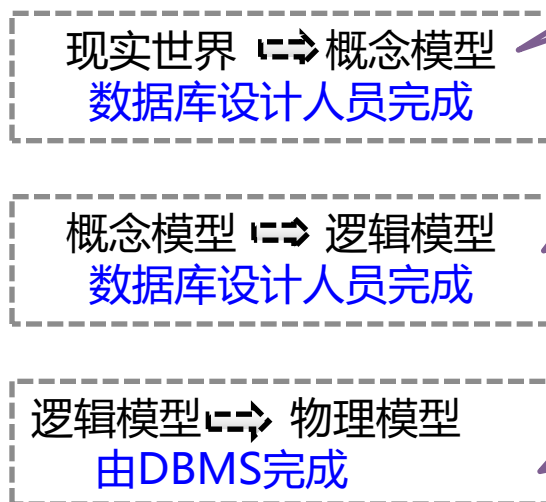
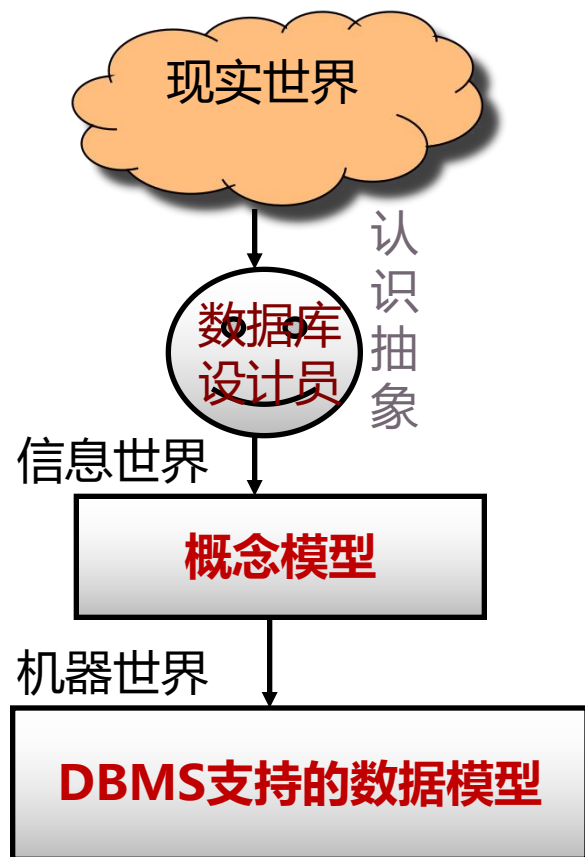


数据模型应满足3方面要求：

- 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中用于提供信息表示和操作手段的形式构架。即：在数据库中用数据模型这个工具来**抽象、表示和处理**现实世界中的数据 and 信息。

1.2.1 两类数据模型



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，用于**数据库设计**。

逻辑模型 主要包括：网状 / 层次 / 关系 / 面向对象 / 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。

图 现实世界中客观对象的抽象过程

1.2.2 概念模型

- **概念模型**：按用户的观点对数据建模。
 - **独立于**特定DBMS的现实世界的抽象模型。
 - 数据库设计的有力工具；用户与DB设计人员进行交流的语言。
- **特点**：
 - 1) 较强语义表达能力；
 - 2) 便于直接表示应用语义；
 - 3) 简单、清晰，易于理解。
- **实体-联系(E-R)模型**：
 - P.P.S.Chen于1976年提出：用E-R图来描述现实世界的概念模型：世界由一组称作**实体**的基本对象及这些对象间的**联系**组成。
 - 是一种**语义模型**，语义由模型力图来表达数据。

1. 信息世界中的基本概念

1. **实体(entity)**——客观存在可相互区别的事物、事件和概念。(静态、动态、物质、精神、联系等)。注意：不仅可以是具体的人、事、物（如：一个学生 / 书 / 车），或抽象的概念和联系（一堂课、一次比赛）。
2. **属性(Attribute)**——实体具有的特征。
例：Student (学号, 姓名, 性别, 出生日期, ...)
3. **码 / 键(KEY)** ——**唯一标识**一个实体集中任何实体值又**不含多余属性**的属性集。（至少一个属性；至多n个属性）
例：学生 (KEY): 学号
选课 (KEY): (学号, 课程号)
例：学生： (学号, 姓名) **KEY?**

1. 信息世界中的基本概念

4. **实体型 (entity type)** ——具有相同特征和性质的实体与属性命名序列。

实体值 (entity value) ——实体型的具体实例。

例: **Student**

学号	姓名	性别	年龄
----	----	----	----

2001	余得水	男	20
------	-----	---	----

值集/域(domain)——属性的取值范围。

例: 学号的域为6位整数, 姓名的域为字符串集合,

年龄的域为小于40的整数, 性别的域为 (男, 女) 。

简单属性 / 复合属性 / 单值属性 / 多值属性 / NULL属性 / 派生属性

1. 信息世界中的基本概念

5. **实体集 (Entity Set)** —— 具有相同类型及相同性质 (或属性) 的实体集合。
实体集可以相交。



1. 信息世界中的基本概念

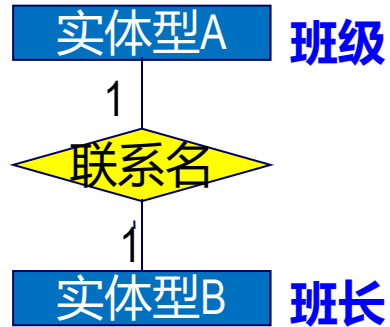
6. **联系 (relationship)** ——现实世界中事物内部以及事物之间的联系在信息世界中被抽象为实体内部的联系和实体之间的联系。实体间的联系可分为三类：

- ① 一对一联系 (1:1)
- ② 一对多联系 (1:m)
- ③ 多对多联系 (m:n)

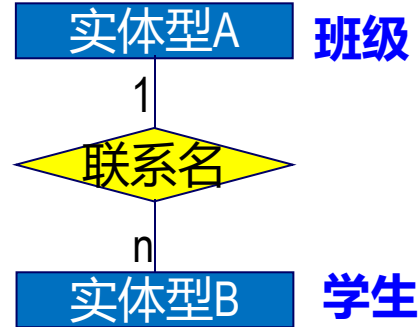
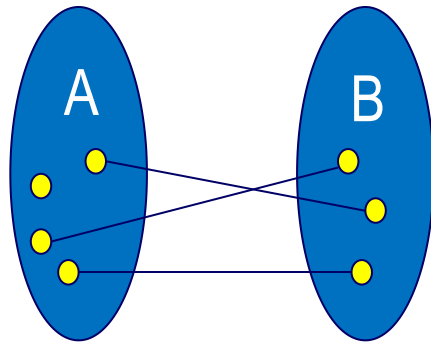
两个实体模型之间的联系——二元联系集的映射基数

1: 1 (一对一联系)

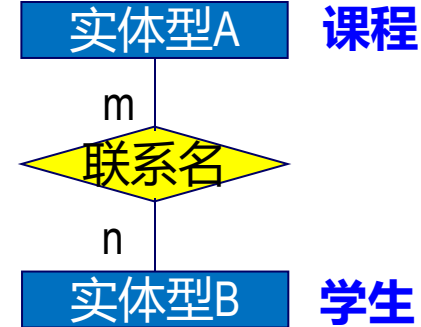
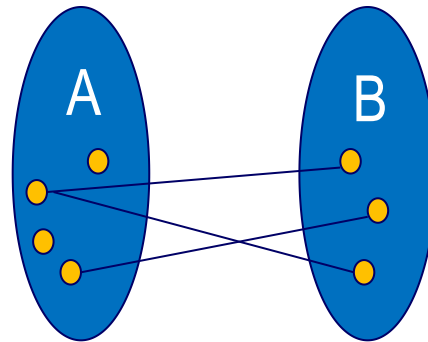
定义：设有实体集A、B，若其中任何一个实体集中每一实体至多与另一实体集中的一个实体有联系，则称A、B间存在一对一联系。



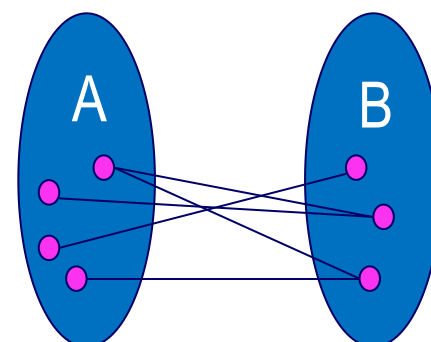
(a) 1: 1的联系



(b) 1: n 的联系



(c) m : n 的联系



m:n联系

定义：设有实体集A、B，若其中任何一个实体集中的每一个实体均与另一个实体集中的n个实体 ($n \geq 0$) 有联系，则称A、B间存在多对多联系。

1: m(一对多联系)

定义：设有实体集A、B，若A中的每一个实体，与B中的n个实体 ($n \geq 0$) 有联系，反之，对于B中的每一个实体，至多与A中的一个实体有联系，则称A、B间存在一对多联系。

1. 信息世界中的基本概念

课堂练习:

顾客可以到多个商场购物，商场可以有很多顾客来购物，则商场与顾客间的联系是_____。

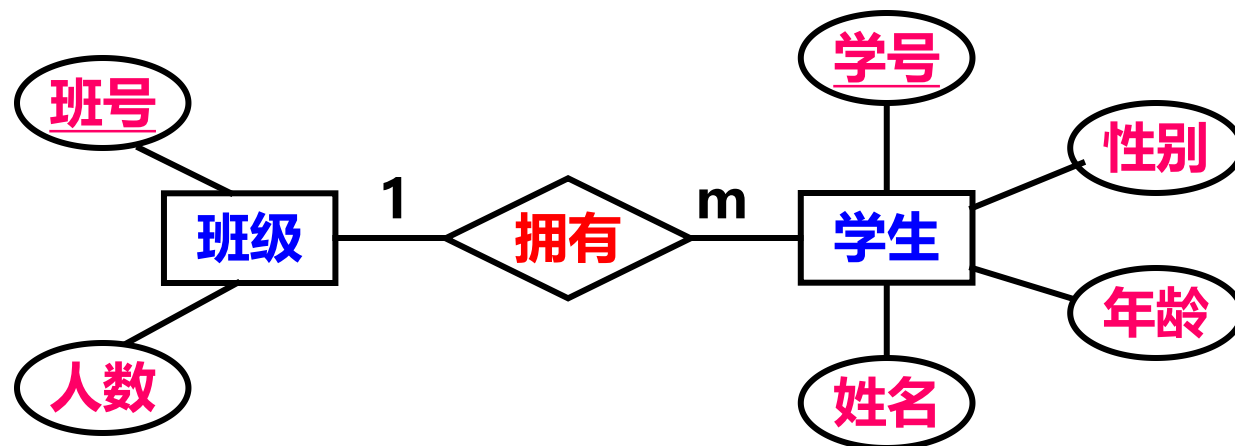
- A. 1: 1
- B. 1: n
- C. ☒ m: n
- D. m: 1

注意：实体间的联系类型反映的是**实体集**间的普遍联系。

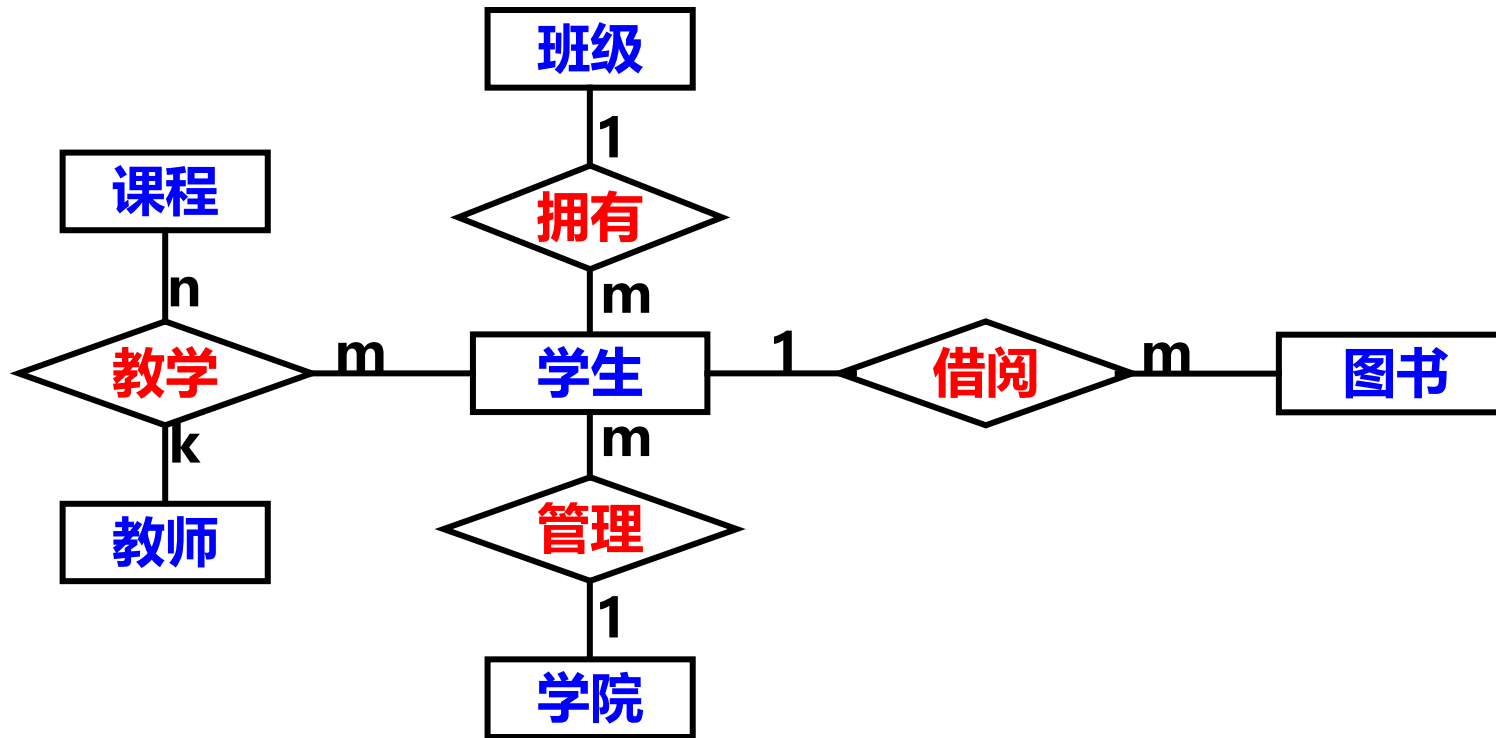
实体-联系 (Entity-Relationship) 模型

采用画E-R图的形式直接表示实体及实体间联系的方法：

- (1) **矩形**表示**实体型**，框内标明实体名；
- (2) **椭圆**表示**属性**，用无向边与其相应实体连接；
- (3) **菱形**表示**联系**，标明联系名，用无向边与相关实体连接；
- (4) **无向边**上标明**联系的类型** ($1:1$, $1:m$, $m:n$)；
- (5) 可据需要任意展开 (E-R图复杂时可略去属性)。



实体-联系 (Entity-Relationship) 模型



E-R图是数据库设计人员与用户进行沟通、交流的工具。
但是，DBMS很难直接支持E-R模型。

1.2.3 数据模型的组成要素

- 数据模型是按计算机的观点对现实世界进行抽象、表示事物及事物间联系的模型。
- 数据模型通常由**数据结构**、**数据操作**和**数据的约束条件**三个要素组成。

1. **数据结构**：描述数据库的组成对象及对象之间的关系。

- 数据结构用于描述系统的**静态特性**。
- 数据结构是所研究的对象类型的集合，它是刻画一个数据模型性质最重要的方面。
- 在数据库系统中，人们通常按照其数据结构的类型来命名数据模型。如：**层次模型、网状模型和关系模型的数据结构分别为树型结构、网状结构和关系结构**

1.2.3 数据模型的组成要素

2. **数据操作**：是指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。
- 数据操作用于描述系统的**动态特性**。
 - 数据操作是对数据库中各种数据操作的集合，包括操作及相应的操作规则。如数据的检索、插入、删除和修改等。
 - 数据模型必须定义这些操作的确切含义、操作规则以及实现操作的语言。

1.2.3 数据模型的组成要素

3. 数据的约束条件

- 数据的约束条件是一组完整性规则的集合。
- 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。
- 数据模型还应该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。

例，在学生数据库中，学生的年龄不得超过40岁。

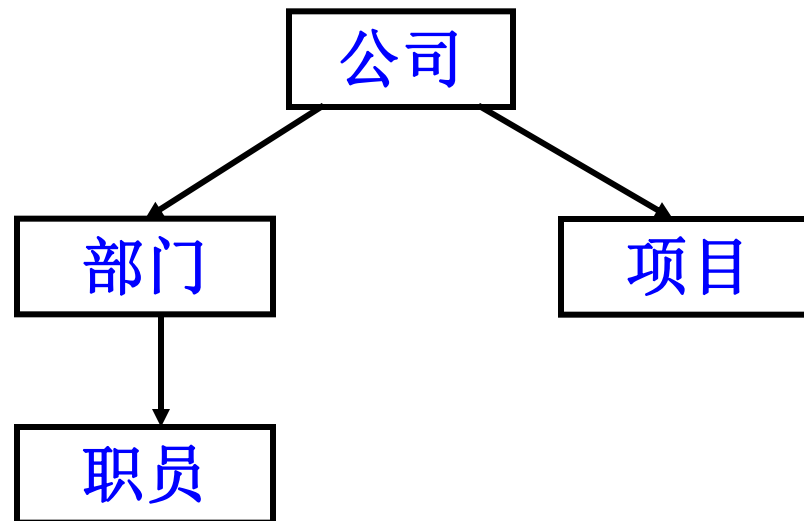
1.2.4 常用的数据模型

- 层次模型(Hierarchical Model)
- 网状模型(Network Model)
- 关系模型(Relational Model)
- 面向对象模型(Object Oriented Model)
- 对象关系模型(Object Relational Model)
- 半结构化模型 (semistructure Model)

- 数据模型主要用于DBMS的实现。
- 一种DBMS通常只支持一种数据模型。数据模型不同，对应的DBMS差别很大，因此，DBMS的类型也通常依据数据模型的不同来划分。

1.2.5 层次模型

- 用**树型结构**表示实体及实体间联系的数据模型。
- 表示方法：
 - ① 一个结点表示一个实体；
 - ② 结点间连线表示实体间联系；
 - ③ 联系中表示1的实体在上层，表示n的结点在下层。



层次数据模型的数据结构(续)

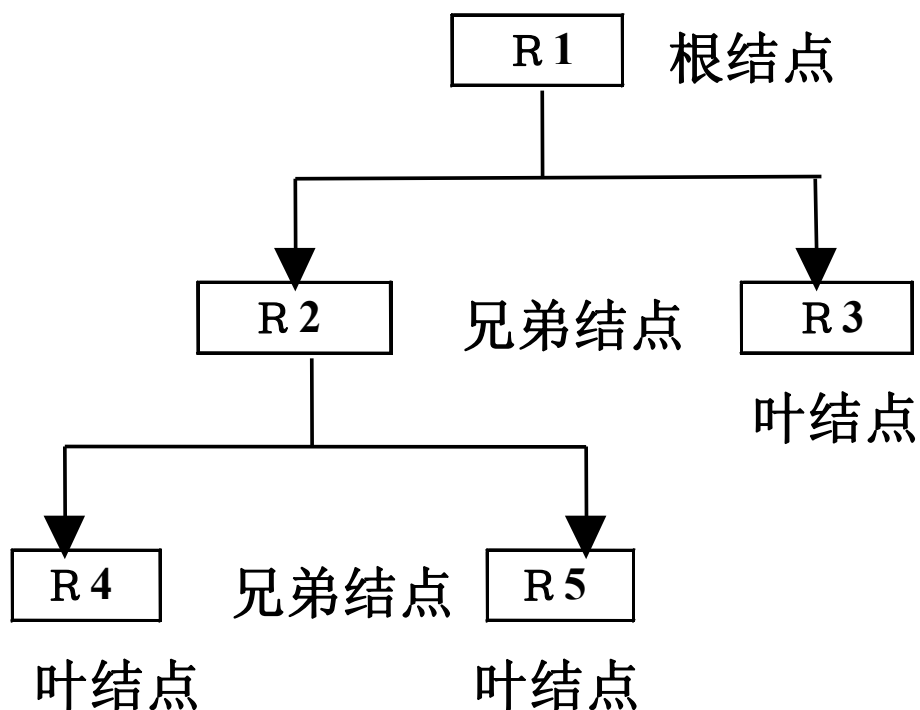


图 一个层次模型的示例

□ 层次模型的特点:

- 结点的双亲是**唯一**的：根节点**没**双亲，其他结点只有一个双亲。
- 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

层次数据模型的数据结构(续)

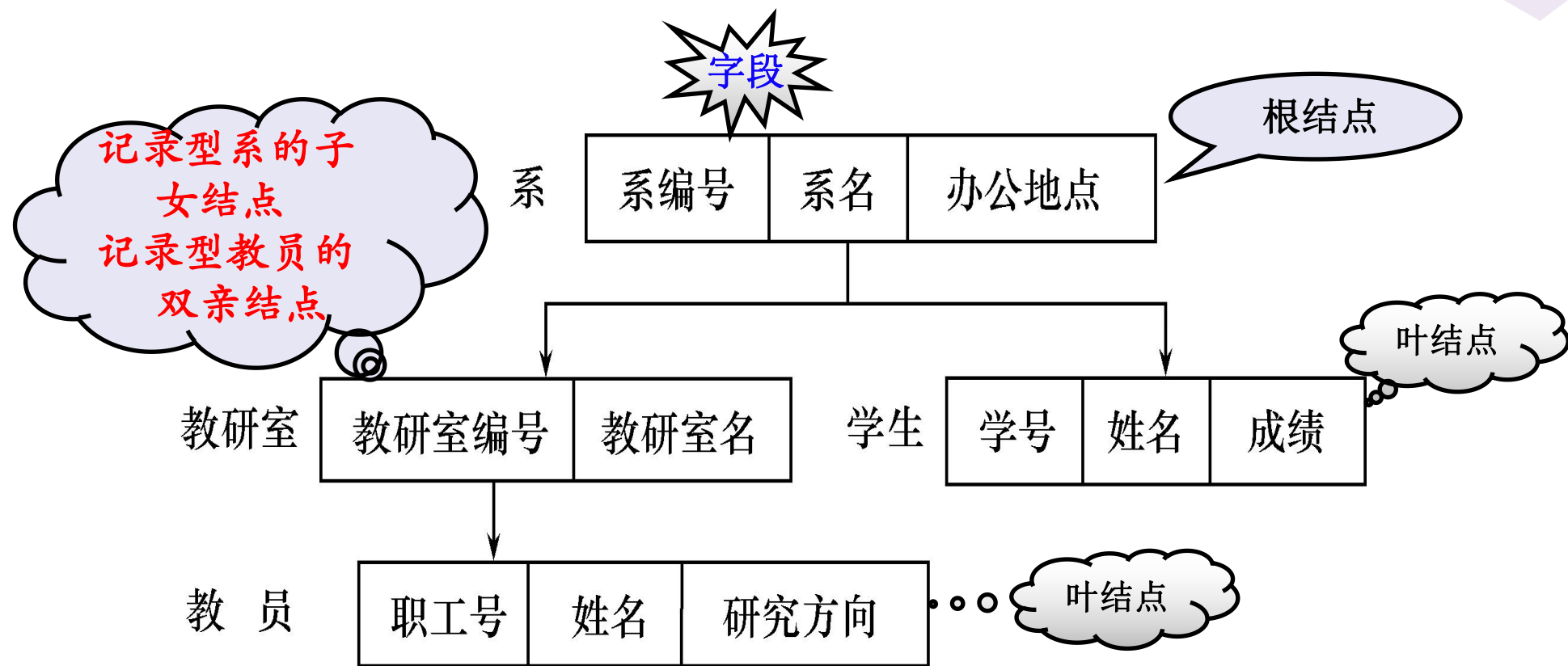


图 教员学生层次数据库模型

层次数据模型的数据结构(续)

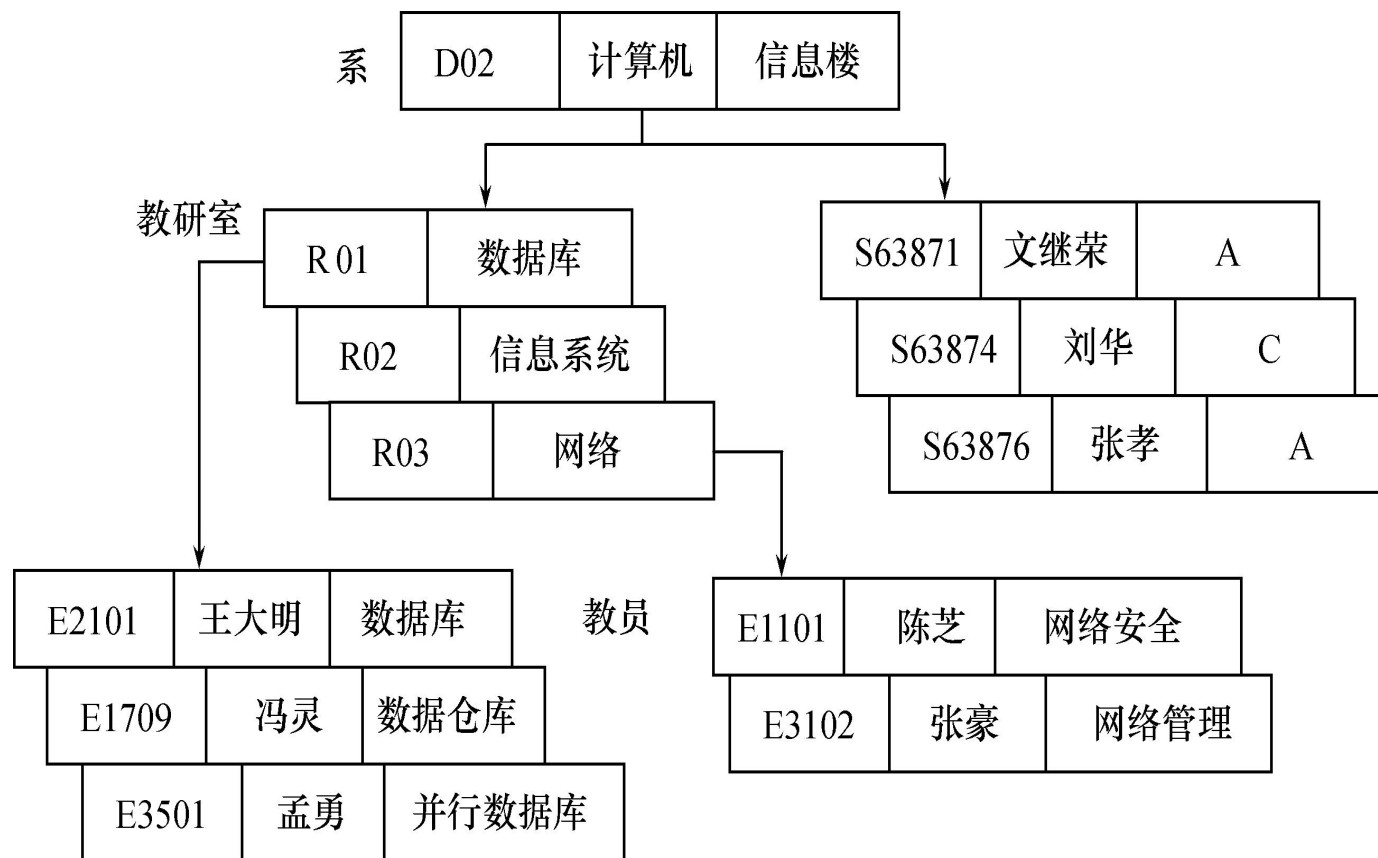


图 教员学生层次数据库的一个值

层次数据模型的存储结构:

1. 邻接法

按照层次树前序遍历的顺序把所有记录值依次邻接存放，即通过物理空间的位置相邻来实现层次顺序。

2. 子女-兄弟链接法

每个记录设两类指针，分别指向最左边的子女（每个记录型对应一个）和最近的兄弟。

3. 层次序列链接法

按树的前序穿越顺序链接各记录值。

1.2.5 层次模型(续)

2. 操作:增、删、查、改

3. 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。

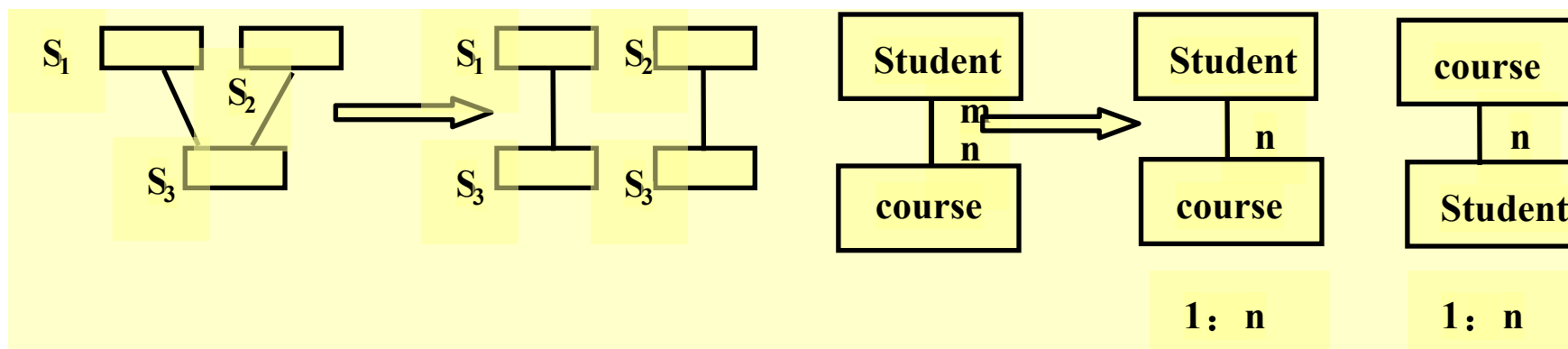
4. 优点

- 简单易用 (几条操作命令) ;
- 自然表示1: M联系;
- 速度较快。

1.2.5 层次模型(续)

缺点:

1) 不能直接表示 $m:n$ 联系 (须引进冗余结点)



2) 插入, 删除操作限制多

3) 在层次模型中具有一定的存取路径, 查子女须经过双亲 (从上到下, 从左到右: 它仅允许自顶向下的单向查询, 随机存取效率低)

1.2.6 网状模型

- 现实世界中事物之间的联系更多的是非层次关系的，用层次模型表示这种关系很不直观，出现网状模型；
- 20世纪70年代，数据系统语言研究会CODASYL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，成为了网状模型的代表。
- 网状模型取消了层次模型的两个限制：两个或两个以上的结点都可以有多个双亲结点，则此时有向树变成了有向图，该有向图描述了网状模型。

1.2.6 网状模型

□ 用网状结构表示实体及实体间联系的数据模型。

□ 表示方法：

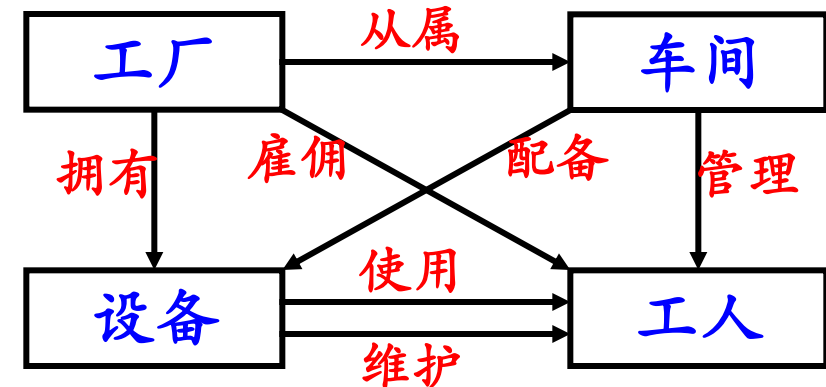
- ① 一个结点表示一个实体；
- ② 有向连线表示实体间联系；
- ③ 两个结点间的联系不唯一，因此联系必须命名。

□ 特征：

- ① 可以有多个结点无父结点；
- ② 至少有一个结点有多个父结点。

□ 缺点：

- ① 模型复杂；
- ② 数据独立性不高。



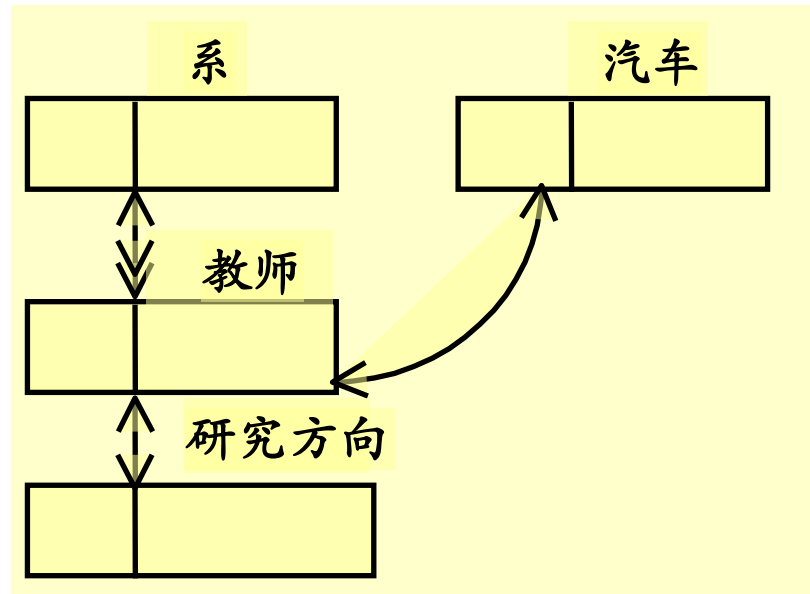
网状数据模型的数据结构

- 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**

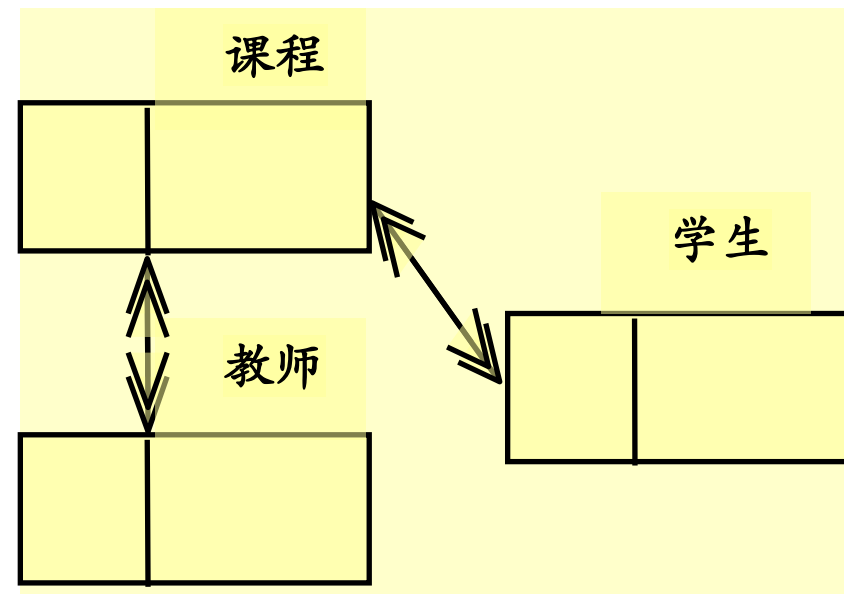
1.2.6 网状模型

1. 数据结构

a: 简单网

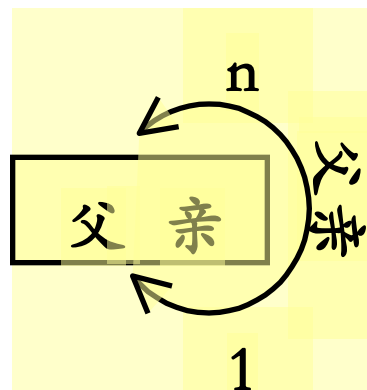


b: 复杂网



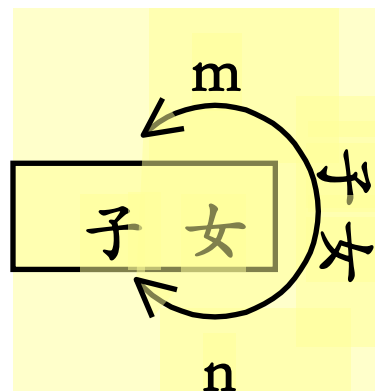
1.2.6 网状模型(续)

c: 简单环网



- 一个父亲有多个已为人父的儿子;
- 已为人父的儿子只有一个父亲。

d: 复杂环网

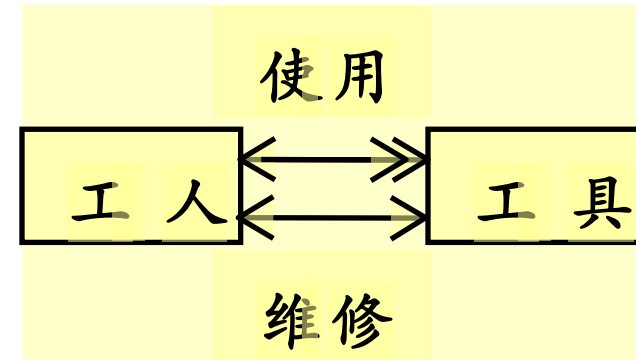


- 每个子女可多个子女;
- 每个为人子女者又可有多多个子女。

12.6 网状模型(续)

□ e: 多种联系

- 1) 结点表示实体，称为**记录类型**；
- 2) 结点内含数据项，表示属性；
- 3) **有向连线**表示实体间联系；
- 4) **属性可嵌套**。
- 5) 特征
 - (1) 可多个结点无双亲结点；
 - (2) 子女结点可多个双亲结点；
 - (3) 两记录间可多种联系。



网状数据模型的存储结构

- 关键

- 实现记录之间的联系

- 常用方法

- 单向链接
- 双向链接
- 环状链接
- 向首链接

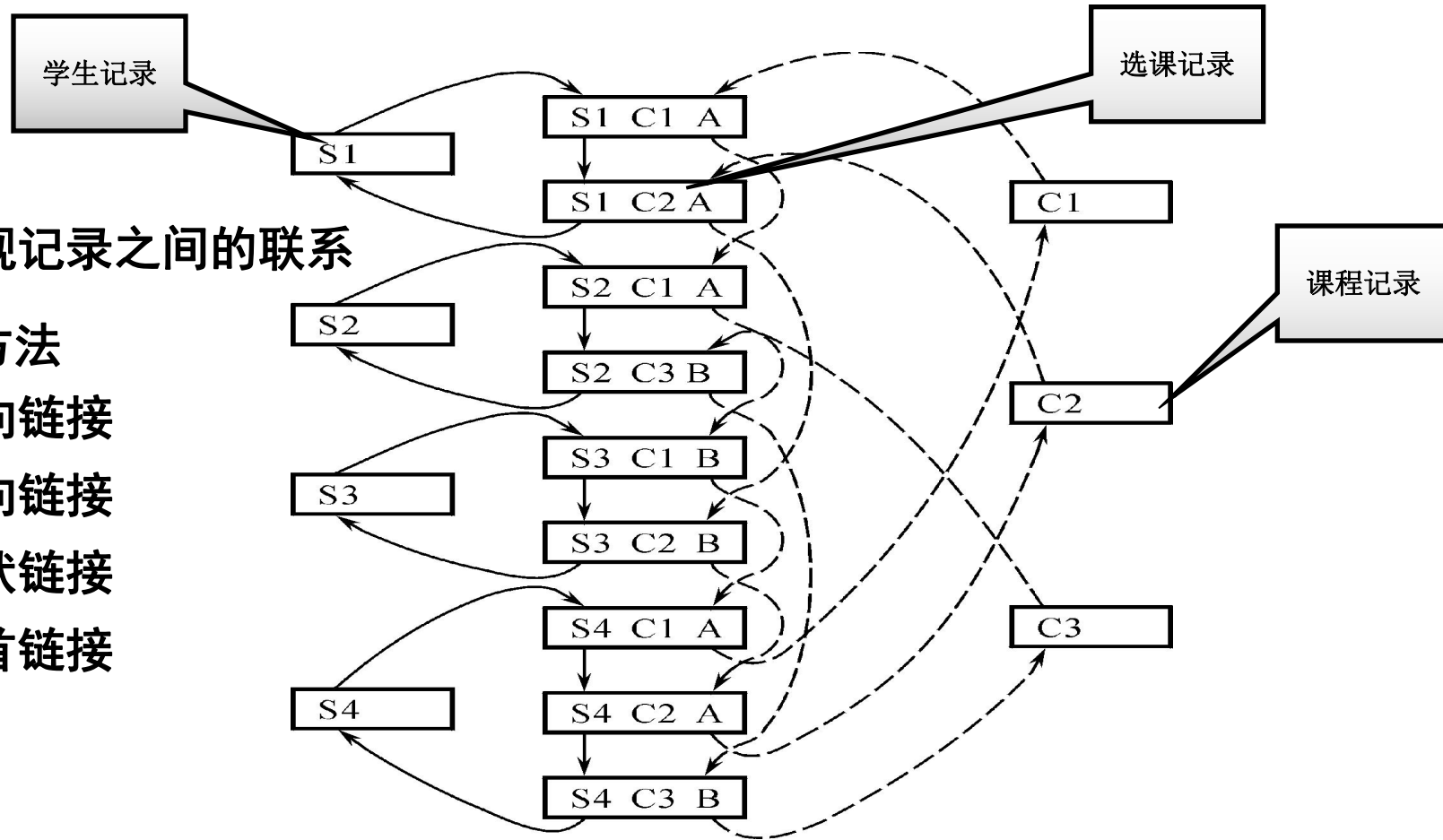


图 学生/选课/课程的网状数据库实例

1.2.6 网状模型(续)

2. 操作：增、删、查、改

3. 约束

1) 插入不受限制； 2) 删去双亲，子女不受影响。

4. 优点

1) 直接表示的m：n联系； 2) 存取效率高。

5. 缺点

1) 结构复杂；

2) DDL、DML复杂；

3) 一次存取一个记录值；

4) 应用程序与数据结构相互依赖，应用程序在访问数据时必须选择适当的存取路径。

1.2.7 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。
 - 他发表论文 “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。为此他获得了1981年的图灵奖。
 - 此后许多人把研究方向转到关系方法上，陆续出现了关系数据库系统。
 - 1977年IBM公司研制的关系数据库的代表系统System R开始运行，其后又进行了不断的改进和扩充，出现了基于System R的数据库系统SQL/DB。
- 关系数据库已成为目前应用最广泛的数据库系统。

1.2.7 关系模型

- 关系数据模型是以集合论中的关系(relation)概念为基础发展起来的数据模型。
- 在关系模型中，无论是实体还是实体间的联系，均由单一的结构类型即关系(表)来表示。
- 比E-R模型的抽象层次更低。

1. 关系模型的数据结构

- 从用户的角度来看，关系模型是用二维表格表示实体及其间联系的数据模型。
- 一个关系就是没有重复行和重复列的二维表。



关系名：学生登记表

关系模式：对关系的描述

元组中的一个属性值为元组的一个分量

关系数据模型的数据结构（续）

表 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 一个工资表(表中有表)实例

举例

- 表1、表2、表3 给出三个不同的关系，其中表1 和 表2是基本关系，用于表示现实世界的基本实体。
- 而表3则是“联系”关系，用于表示哪些学生选修哪些课程的联系，这种**联系**是通过在“选课”关系中引入“学生”关系中的“学号”和“课程”关系中的“课程号”属性的数据建立的。

表1：学生

学号	姓名	性别
----	----	----

表2：课程

课程号	课程名	学时
-----	-----	----

表3：选课

学号	课程名	成绩
----	-----	----

关系数据模型结构简单、理论基础严密、数据独立性高、支持非过程化语言、一次操作可存取多个元组，并且可直接表示多对多联系。
主要不足是查询效率低。

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**，即若干元组的集合：
 - 查询、插入、删除、更新
- **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”。
- 关系的完整性约束条件：
 - **实体完整性**
 - **参照完整性**
 - **用户定义的完整性**

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透明
 - 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

■ 缺点:

- 存取路径对用户透明导致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。

数据库系统的发展

- 按照数据模型的进展情况，数据库系统的发展可划分为三代：
- 第一代：层次数据库系统和网状数据库系统
 - 主要支持层次和网状数据模型
- 第二代：关系数据库系统
 - 支持关系数据模型，该模型有严格的理论基础，概念简单、清晰，易于用户理解和使用。因此一经提出便迅速发展，成为实力性最强的产品。
- 第三代：面向对象数据库系统
 - 基于扩展的关系数据模型或面向对象数据模型的一代数据库系统。
 - 特点：支持包括数据、对象和知识的管理，在保持和继承第二代技术的基础上引进新技术(如OO)，对其他系统开放，具有良好的可移植性、可连结性、可扩充性、互操作性。

1.2 小结

- 使用计算机来管理信息，首要解决如何将现实世界恰当地抽象、表示为计算机中的数据集合的问题。 解决该问题要用到的工具就是——**数据模型**。
- **二级抽象**的过程是先由现实世界到信息世界，再到机器世界。
- 从现实世界映射到信息世界的模型称为**概念模型**，以**实体-联系模型**（E-R模型）为主要代表，它将现实世界抽象为**E-R图**。
- 从信息世界映射到机器世界的模型称为**数据模型**，根据其采用的数据结构的不同，经历了由层次模型、网状模型到关系模型、面向对象模型、对象关系模型的过程。目前主流DBMS支持的数据模型为关系模型。

课堂练习

- 层次型、网状型、关系型数据库划分原则是?
 - A. 记录长度
 - B. 文件大小
 - C. 联系的复杂程度
 - D. 数据之间的联系方式

课堂练习

???

□ 什么模型比层次模型、网状模型更利于与用户交流?

- A. 关系模型
- B. 概念模型
- C. 面向对象模型

