第2章 数据模型

北京理工大学 计算机学院 张文耀

zhwenyao@bit.edu.cn



- 数据模型(Data Model)
 - 信息领域采用的模型
 - 将现实世界的各种事物以及事物之间的联系,表示为数据 以及数据之间的联系
 - 是对现实世界数据特征的抽象和模拟
 - 用来描述数据、组织数据和操作数据
 - 是数据库系统的核心和基础



概念数据模型(概念模型)

数据模型

基本数据模型 (数据模型)

层次数据模型

网状数据模型

关系数据模型

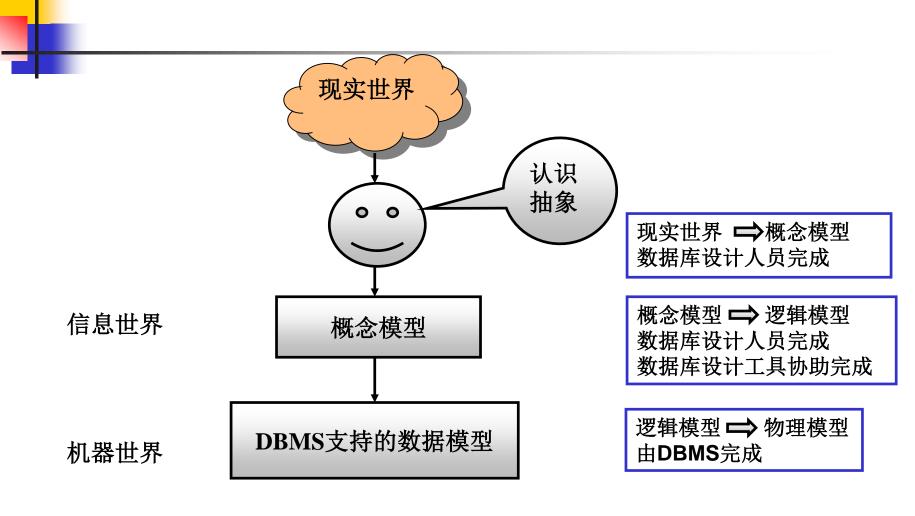
面向对象模型

逻辑数据模型

数据结构 数据操作 数据完整性约束

物理数据模型

- 4
 - 在数据库中用数据模型来抽象、表示和处理现实 世界中的数据和信息。
 - 基本过程---两步抽象
 - 把现实世界中的客观对象抽象为概念模型;
 - 把概念模型转换为某一DBMS支持的数据模型。



概念模型是现实世界到机器世界的一个中间层次。

主要内容

介绍具体的数据模型

- E-R概念模型
- 层次数据模型
- 网状数据模型
- 关系数据模型
- 面向对象数据模型

2.1 E-R概念模型

- 概念模型,也称信息模型
 - 是按用户的观点对数据和信息建模
 - 是现实世界到信息世界的抽象表示
- ■信息

客观世界中存在的事物在人们头脑中的反映,人们把这种 反映用文字、图形等形式记录下来,经过命名、整理、分 类就形成了信息。

- 概念模型的用途
 - 用于信息世界的建模
 - 用于数据库设计,是数据库设计的有力工具
 - 是现实世界到机器世界的一个中间层次
 - 数据库设计人员和用户之间进行交流的语言



- 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。
- ■最常用的概念模型
 - 实体一联系模型

(Entity-Relationship data model, E-R模型)

- 实体
- 属性
- 实体之间的联系

E-R模型的基本概念

- 实体(Entity)
 - 表示客观存在的可以相互区别的事物。
 - 可以是具体的对象,如一个学生、一本书、一辆汽车。
 - 也可以是抽象的概念或联系,如一堂课、一次比赛等。
 - 实体用特征来描述,如学生可以用姓名、性别、年龄来表征。
 - 实体是指现实世界存在的个体,具有同类特征的个体的 集合形成一个实体集(Entity Set)。
 - 实体集用实体名和表征实体的多个特征的集合来描述, 其结果称为实体的"型"(Entity Type)。
 - 实体集中的每个实体称为实体型的值或实例。



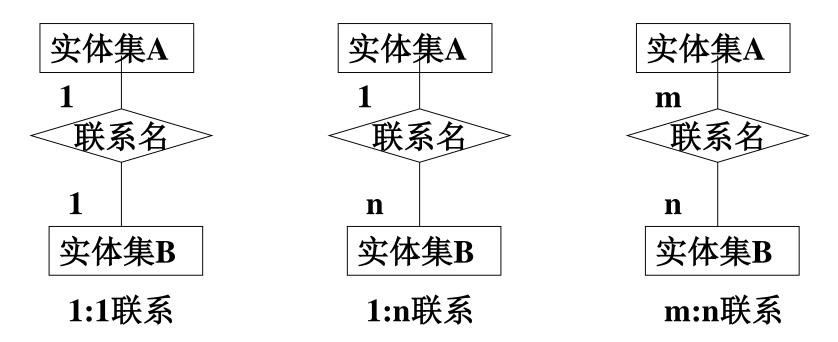
- 属性 (Attribute)
 - 实体所具有的某一特征称为属性;
 - 一个实体可以由若干个属性来刻画;如:学生有学号、姓名、性别、年龄、系等方面的属性。
 - 属性有"类型"和"值"之分
 - ■"类型"为属性名,如姓名、性别、年龄是属性的型;
 - "值"为属性的具体内容,如(张三,男, 20);
 - 属性的取值范围称为属性的域(domain)。
 - 实体标识符:一个或一组用来区分实体集中每个实体的属性。例如:学号可以作为学生实体的标识符。 简称:键(码)(Key)



- 实体间的联系(Relationship)
 - 在现实世界中
 - 事物内部的联系
 - 事物之间的联系
 - 在信息世界(概念模型)中
 - 实体内部的联系 是指组成实体的各属性之间的联系。
 - 实体之间的联系 是指不同实体集之间的联系。
 - 二元联系: 两个实体集之间的联系
 - 多元联系
 - 一元联系



■ 两个实体集之间的三类联系





- 一对一的联系(1:1)
 - 实体集A中的一个实体至多与实体集B中的一个实体相对应,反之亦然,则称实体集A与实体集B为一对一的联系。
 - 记作1:1。
 - 示例: 班级与班长之间的联系
 - 一个班级只有一个正班长
 - 一个班长只在一个班中任职



- 一对多的联系(**1:n**)
 - 实体集A中的一个实体与实体集B中的多个实体相对应, 反之,实体集B中的一个实体至多与实体集A中的一个 实体相对应。
 - 记作1:n。
 - 示例: 班级与学生之间的联系
 - 一个班级中有若干名学生,
 - 每个学生只在一个班级中学习

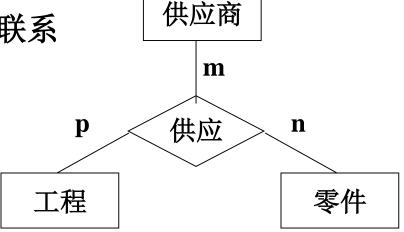


- 多对多的联系(m:n)
 - 实体集A中的一个实体与实体集B中的多个实体相对应; 反之,实体集B中的一个实体与实体集A中的多个实体 相对应。
 - 记作(m:n)。
 - 示例: 课程与学生之间的联系
 - 一门课程同时有若干个学生选修
 - 一个学生可以同时选修多门课程

- - 多个实体集之间也可以存在联系,称多元联系。
 - 若规定:
 - 一个供应商可供应多种零件给多个工程;
 - 一个工程可由多个供应商供应多种零件;
 - 一种零件可由多个供应商供应给多个工程。
 - 供应商、零件和工程间存在着多对多的联系,

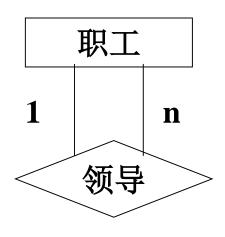
表示为m: n: p。

■ 不同于多个实体两两间的联系





- 实体集内部不同实体间的联系
 - 一对多的联系
 - 职工实体集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
 - 一对一的联系
 - 夫妻关系
 - 多对多的联系
 - ■朋友关系



同一实体型内部的 1:n联系

E-R数据模型

- 概念模型的表示方法
 - E-R模型,也称为E-R方法
 - UML类图(软件工程、面向对象)
 -
- E-R模型 用E-R图表示现实世界中实体与实体间联系的模型
- E-R图
 - 实体
 - 联系
 - 属性

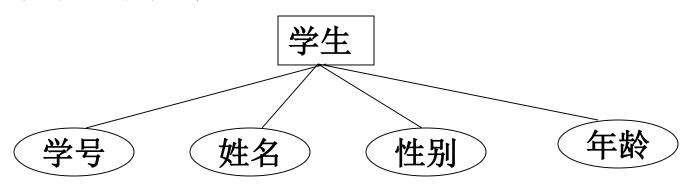


- 实体
 - 实体型用矩形表示,矩形框内写明实体名。

学生

教师

- 属性
 - 用椭圆形表示,用属性名标注,并用无向边将其与相应 的实体连接起来





■联系

■ 联系本身

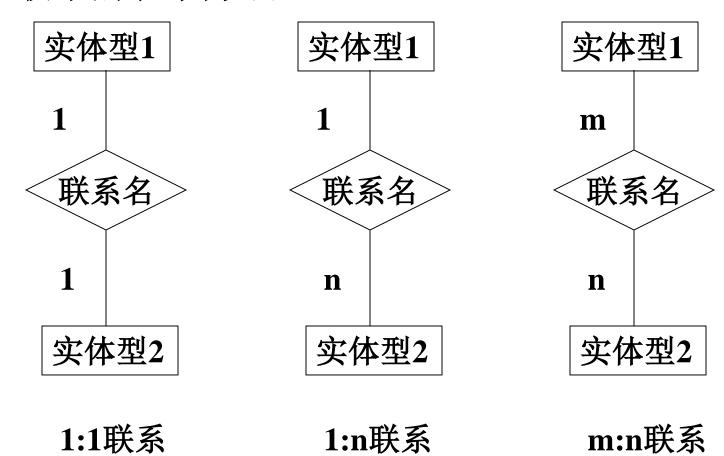
用菱形表示,菱形框内写明联系名,并用无向边分别与有关实体连接起来,同时在无向边旁标上联系的类型 (1:1、1:n或m:n)。

■ 联系的属性

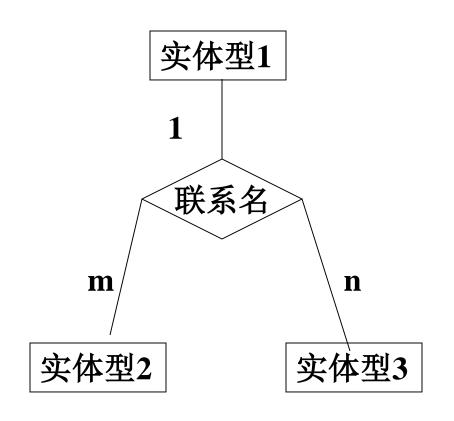
联系本身也是一种<mark>实体型</mark>,可以有<mark>属性</mark>。如果一个联系具有属性,则这些属性也要用无向边与该联系连接起来。

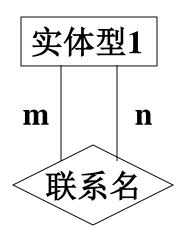


■ 联系的表示方法







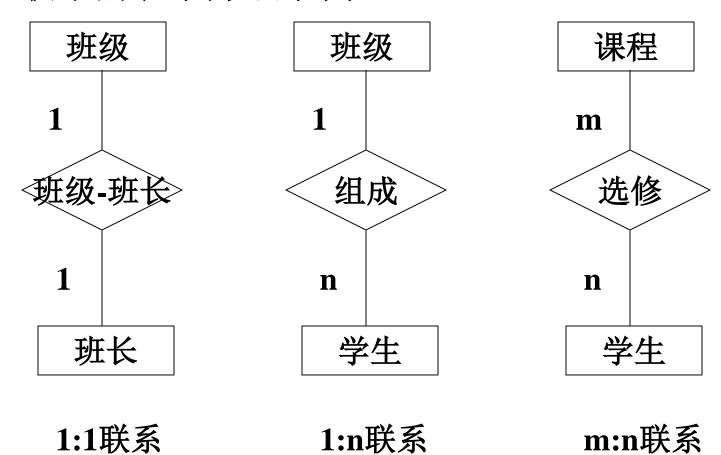


同一实体型内 部的m:n联系

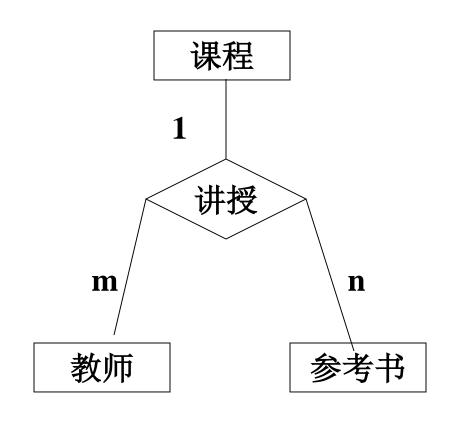
多个实体型间的1:n联系

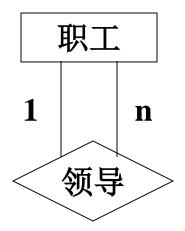


■ 联系的表示方法示例





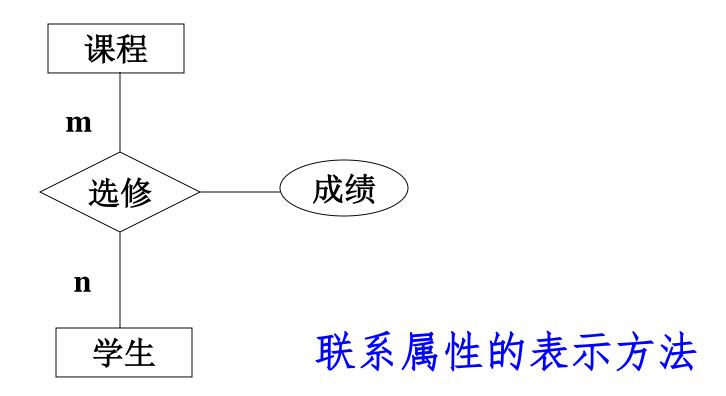




同一实体型内 部的1:n联系

多个实体型间的1:n联系

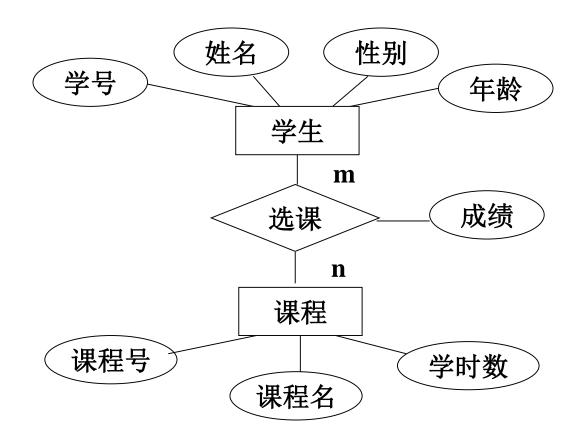




4

· E-R图实例

—学生选课的E-R图





用椭圆表示实 体的属性 用无向边把实体与其属性连接起来

姓名

学号

系别

课程号

课程名

学分

学生

m

选修

n 湯程

用矩形表 示实体

用菱形表示实 体间的联系

成绩

实体与联系用线段 连接并注明类型

联系的属性

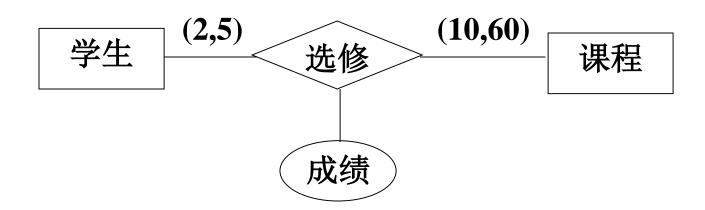
E-R图表示方法小结

4

- E-R模型中实体间联系的语义
 - 1. 基数比约束 如在二元联系中有**1:1**,**1:n**,**m:n**的联系
 - 参与约束 根据实体集中的实体是否全部参与联系来描述实体参与联系的约束
 - 一个实体集中的所有实体都参与联系称为完全参与, 否则称为部分参与。
 - 3. 实体的参与度 实体参与联系的最小和最大次数,称实体的参与度
 - 可以表示基数比约束和参与约束



- 联系的语义示例
 - 如果规定一个学生最少选修2门课,最多选修5门课,则学生在选课联系中的参与度是(2,5)
 - 规定一门课至少要有10个学生选修,至多有60个学生 选修,则课程在选课联系中的参与度是(10,60)





■ 弱实体

- E-R模型中有一类特殊的实体,这种实体的存在是依赖于其他实体而存在的,称这类实体为弱实体。
- 相对于弱实体,它所依赖的实体称为强实体。
- 弱实体不能脱离依赖实体而单独存在,全部参与联系。
- 弱实体本身不一定具有标识属性。
- 在E-R模型中,弱实体用双框矩形表示; 为了表示全部参与,与菱形框间用双线连接。



■ 弱实体示例

在学生管理信息中,学生实体与家长实体之间存在着 "所属"关系,家长实体是不能脱离学生实体而独立存 在的,成为弱实体。





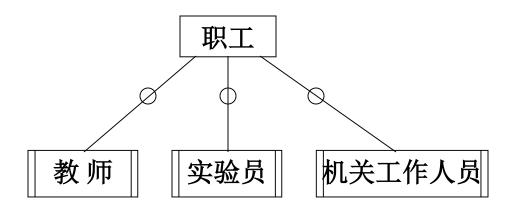
■ 子类实体

- 在扩展E-R数据模型中增加了子类和超类的概念,使E-R数据模型具有了更多的语义。
- 将实体集根据个体的不同特性分为多个子集,由此产生 了子类实体。(子类和超类)
- 子类是根据个体的不同特征进行特殊化的结果; 超类是对不同实体集的共同特征的概括。
- 子类可以继承超类的所有属性,也可以继承超类的联系。
- 子类可以有自己的联系。
- 子类的表示方法,见示例。



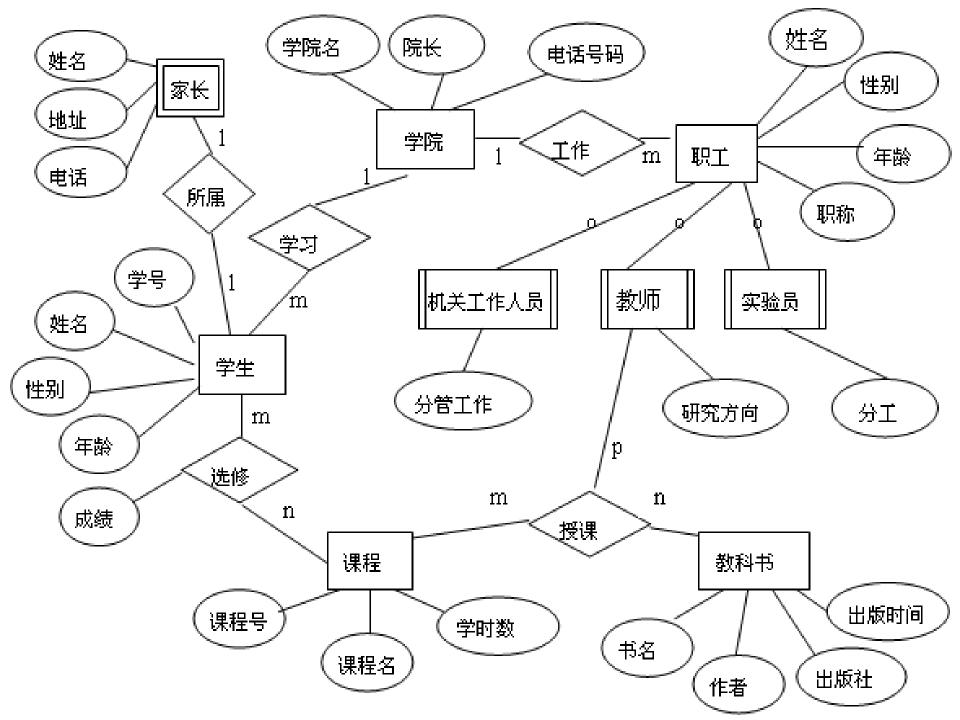
■ 子类实体示例

学校的职工,按照不同工作的特点可分为教师、实验人员和机关工作人员等。他们除具有共同的特性如姓名、 年龄、性别外,还有各自不同的特性



E-R模型示例

- 教务管理(参见电子版书稿)
 - ■学院
 - 职工
 - 学生
 - 家长
 - 课程
 - 教科书
 - 教师
 - 实验员
 - 机关工作人员





■ 教务管理实体间的联系

• • • • • •

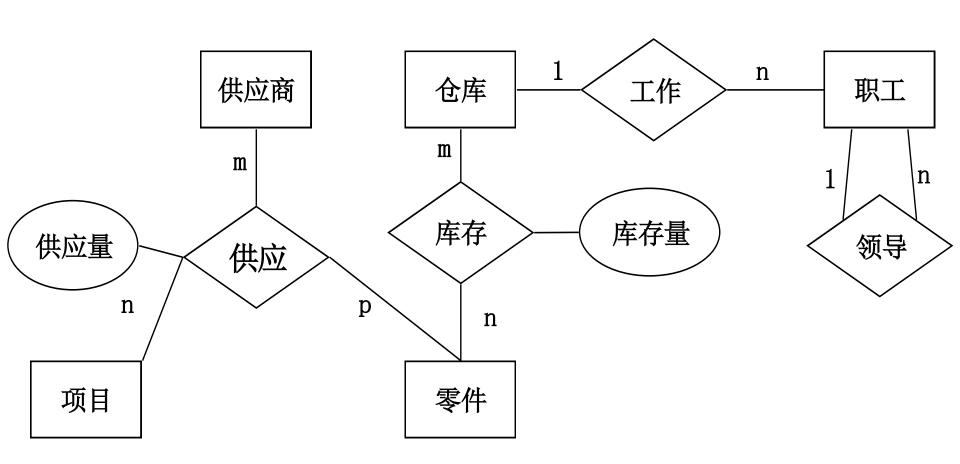
一个学院有若干名职工,一个职工仅在一个学院工作,学院与职工间是 1: m 的联系; 一个学院有若干名学生,一名学生仅在一个学院学习,学院与学生间是 1: m 的联系; 一个学生可以选修多门课,一门课可供若干学生选修,学生与课程间是 m: n 的联系; 一个学生有一位家长联系,一位家长对应一个学生,学生与家长间是 1:1 的联系; 一个教师可以讲授多门课,在讲授某门课时同时确定所用教材; 一门课可以有多个授课教师,不同教师可以用不同的教材; 一种教材可用于多门课且被不同的教师选用。教师、课程、教科书之间存在着 p: m: n 的联系。

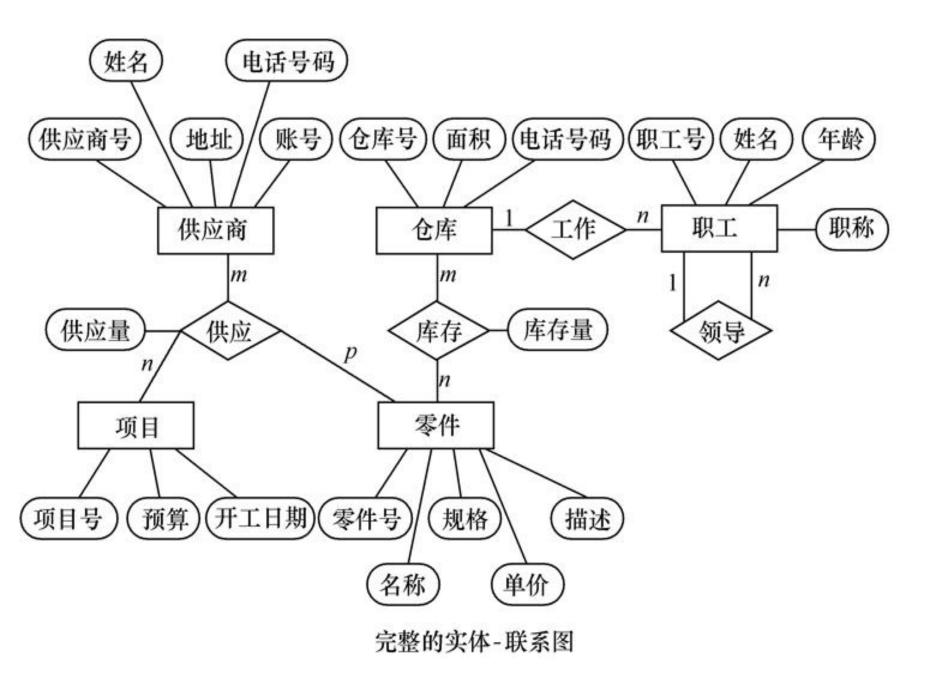


■物资管理

- 基本实体: 仓库、零件、供应商、项目、职工
- 基本联系:
 - 一个仓库可以存放多种零件,一种零件可以存放在 多个仓库中;
 - 一个仓库有多个职工当仓库保管员,一个职工只能 在一个仓库工作;
 - 职工之间具有领导与被领导的关系,即:仓库主任 领导若干保管员;
 - 一个供应商可以供给若干项目多种零件,每个项目可以使用不同供应商供应的零件,每种零件可有不同供应商供给。







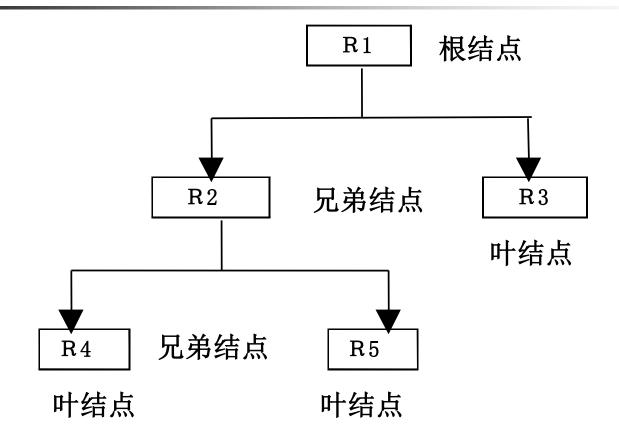


- E-R模型被广泛地用于数据库概念模型的设计。
- 在E-R图中仅表示现实世界中的信息结构及信息之间的关系,不涉及任何信息在计算机中的表示。
- 只要用户的需求不变,E-R模型是稳定的。
- 运用E-R模型,可以很方便地将其转换为具体的 DBMS所支持的数据模型。

2.2 层次数据模型

- 层次数据模型是数据库系统中最早出现的数据模型, 典型代表是1968年IBM推出的IMS
- 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。
- 层次模型能很好地模拟自然的层次关系。
- 层次模型采用树型结构表示各类实体以及实体间的联系。
 - 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。
 - 子女结点、兄弟结点、叶结点





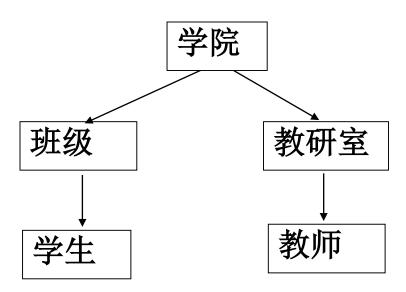
层次模型的树型结构

- 4
 - 层次模型的表示方法
 - 实体: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(field)。
 - 联系:用结点之间的连线表示记录(类)型之间的一对多的联系(包括一对一的联系)。
 - 层次模型的基本数据结构
 - 记录 存取数据的基本单位,由若干字段组成。
 - 由记录组成的层次结构

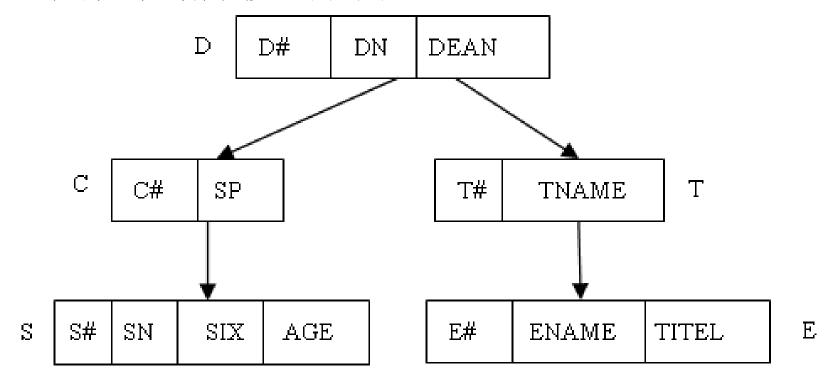
- 4
 - 若干记录型按照层次结构组织成一个层次数据库模式。
 - 一个层次数据库模式可以有多个实例,每个实例 是一棵值树。
 - 一个层次数据库模式的所有实例组成一个层次数据库。



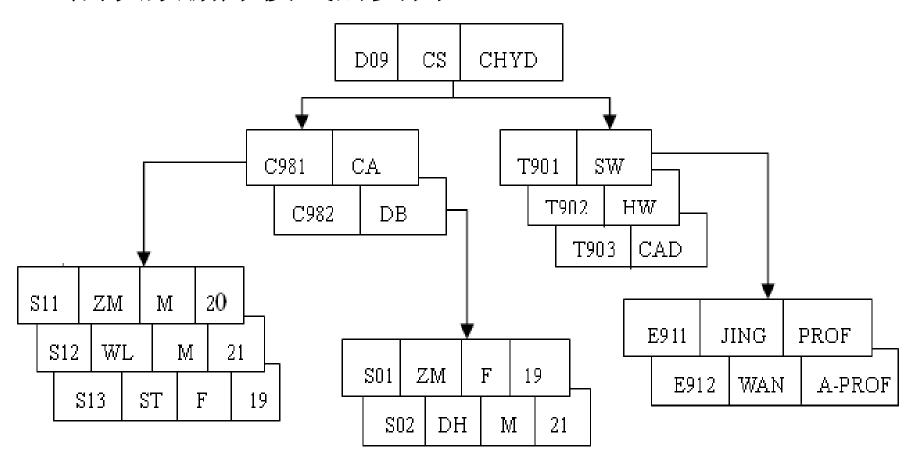
■ 层次模型示例



■ 层次数据库模式示例



■ 层次数据库模式的实例





- 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
 - 每个记录类型定义一个排序字段,也称为码字段
 - 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操纵
 - 查询
 - 插入
 - 删除
 - 更新



- 层次模型的完整性约束
 - 无相应的双亲结点值就不能插入子女结点值
 - 删除双亲结点值时相应子女结点值也被同时删除



- 层次模型的存储结构
 - 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
 - 链接法 用指针来反映数据之间的层次联系

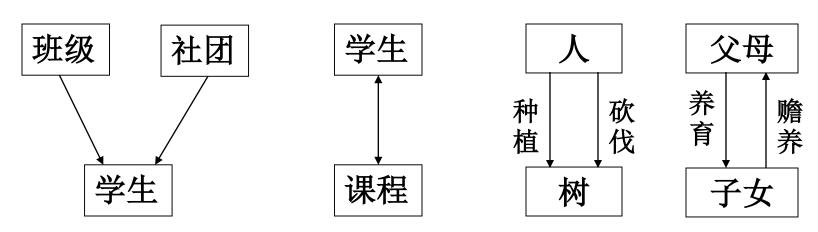
- 4
 - 多对多联系在层次模型中的表示 用层次模型间接表示多对多联系,即将多对多联系分解成 一对多联系。
 - 冗余结点法存在数据的不一致问题
 - 虚拟结点法引入虚拟指针,增加系统开销



- 层次模型的优点
 - 数据结构比较简单,清晰、自然、直观,容易理解
 - 层次数据库的查询效率高
 - 提供了良好的完整性支持
- 层次模型的缺点
 - 多对多联系表示不自然
 - 对插入和删除操作的限制多
 - 查询子女结点必须通过双亲结点
 - 层次命令趋于程序化

2.3 网状数据模型

- 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;网状模型可以清晰的表示这种非层次关系。
- 在网状模型中,结点间的联系可以是任意的:
 - 一个结点可以有多个双亲结点:
 - 允许多个结点可以无双亲结点。





- 网状模型的典型代表
 - DBTG系统,亦称CODASYL系统
 - 20世纪70年代由DBTG(数据库任务组)提出的一个系统方案
 - 非实际系统,C. W. Bachman,1973的年图灵奖
 - 奠定了数据库系统的基本概念、方法和技术
- 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

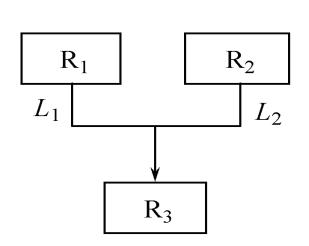
- - 网状模型的表示方法(与层次模型相同)
 - 实体:用记录类型描述, 每个结点表示一个记录类型。
 - 属性:用字段描述, 每个记录类型可包含若干个字段。
 - 联系:用结点之间的连线表示记录(类)型之间的一对多的联系(包括一对一的联系)。

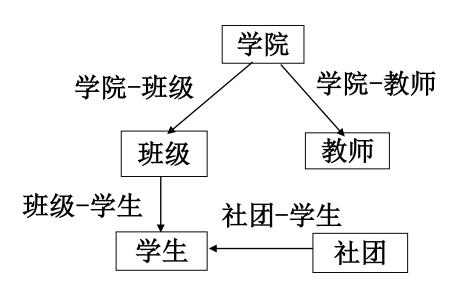


- 网状模型(与层次模型的区别)
 - 允许多个结点没有双亲结点
 - 允许结点有多个双亲结点
 - 允许两个结点之间有多种联系(复合联系)
 - 联系可以不唯一
 - 要命名每个联系,并指出双亲记录和子女记录
 - 网状模型可以更直接地描述现实世界,描述结果是一个有向图。
 - 层次模型实际上是网状模型的一个特例



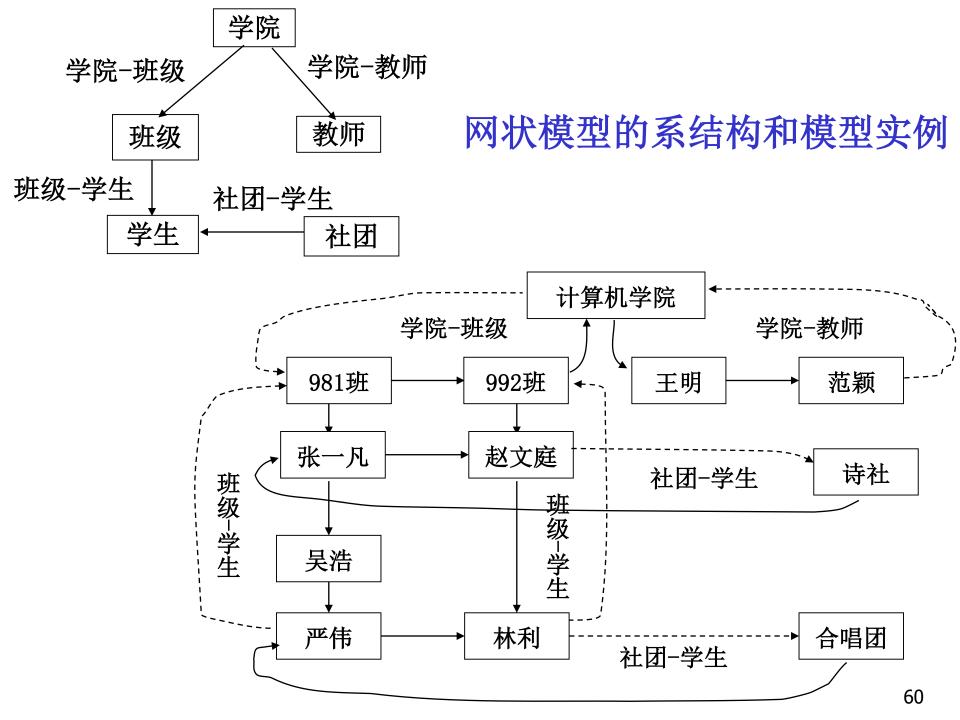
■ 网状模型示例



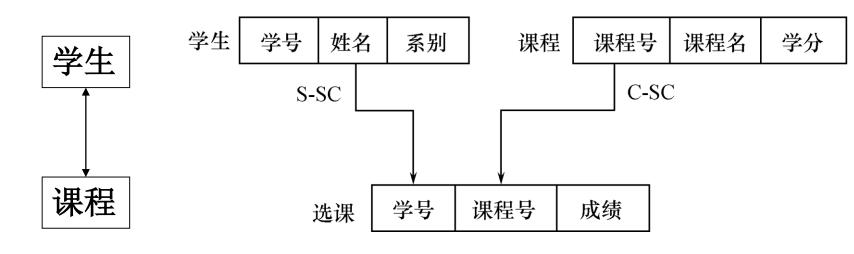




- 网状模型中联系的表示方法
 - 联系比较复杂;
 - 引入 "系(Set)"的概念来表示联系;
 - 系可以看成一个二级树,由一个父结点和多个子女结点 组成,表示上下层之间的**1:m**的联系。

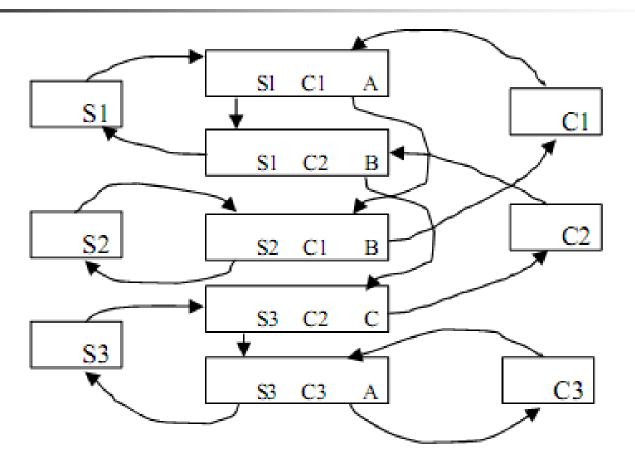


- 4
 - 多对多联系在网状模型中的表示
 - 间接表示多对多联系
 - 方法: 将多对多联系直接分解成一对多联系



增加联结记录,将m:n转换为1:m的联系





学生/选课/课程的网状数据模型实例



- 网状模型的数据操纵与完整性约束
 - 数据操纵与层次模型相同;
 - 完整性约束比层次模型宽松;
 - 允许插入尚未确定双亲结点值的子女结点值;
 - 允许只删除双亲结点值;
 - 具体的网状数据库系统都对数据操纵加了一些限制。



■ 网状模型的优点

- 没有层次结构的限制,能够更好的模拟事物之间各种复杂的联系;
- 一般采用指针实现数据间的联系,存取效率较高。

■ 网状模型的缺点

- 结构比较复杂;随着应用的扩大,数据库结构将越来越复杂,不容易掌握;
- 数据定义和数据操纵语言比较复杂,用户不容易使用;
- 访问数据时必须自行选择存取路径,复杂的导航机制, 增加了程序编写的负担。

2.4 关系数据模型

- 目前数据库系统中应用最普遍的数据模型
- 关系模型的历史
 - 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。
 - 他在"大型共享数据银行的数据关系模型"(A Relation Model of Data for Large Shared Data Banks) 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。他为此获得了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。
- 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、DB2、SQLServer等都是关系数据库系统。



关系模型的数据结构

在关系模型中,基本数据结构被限制为二维表,一张二维表称为一个关系。

学号	姓名	出生年月	性别	入学年份	班 级
2006901	张伟	1988.01	男	2006	200602
2007912	王刚	1989.03	男	2007	200705
•••	•••				

学生关系

关系模型的基本概念

关系(Relation)

- 一张二维表,由多个行和列组成。
- 一个关系可以用来描述一个实体集。
- 关系是数学上集合论中的一个概念,关系模型是以关系 为基础发展起来的。

属性(Attribute)

- 一个关系有多个列,每一列称为关系的一个属性
- 如: 学生关系中的学号、姓名、出生年月等

域(Domain)

- 属性的取值范围。
- 如:学号的域是7位字符数字的集合,学生姓名是汉字字符串的集合等

68



元组(tuple)

- 关系是元组的集合;
- 一个元组对应实体集中的一个个体;
- 一个元组由若干个分量组成,每个分量对应一个属性值。
- 在学生关系中,一个元组对应一个学生实体,学生实体是由**6**个分量组成的。

■ 键(key)【码】

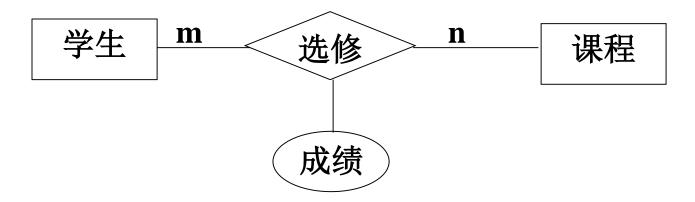
- 键是由一个或多个属性组成的,能够惟一标识一个元组。
- 一个关系中可能有多组属性都能够起到标识元组的作用。 因而,一个关系中可能有多个键,选择其中的一个作为 主键,其余为候选键。

- 4
 - 关系模式(relation schema)
 - 对关系结构的描述称为关系模式。
 - 关系模式可用如下形式表示:
 - 关系名(属性名1,属性名2,...,属性名n)。
 - 如学生关系可表示为:学生(学号,姓名,出生年月,性别,入学年份, 班级)
 - 关系数据库模式 一组关系模式的集合
 - 关系数据库 一组关系模式所对应的关系的集合

关系模型的表示方法

- 关系模型中基本的数据结构是单一的关系
- 实体和实体间的联系都用关系表示
 - 实体(型): 直接用关系(表)表示
 - 属性: 用属性名表示
 - 实体之间的联系
 - 一对一联系: 隐含在实体对应的关系中
 - 一对多联系: 隐含在实体对应的关系中
 - 多对多联系: 直接用关系表示

- 4
 - 关系模型示例
 - 学生、课程、学生与课程之间的多对多联系



学生(<u>学号</u>,姓名,出生年月,性别,入学年份,班级) 课程(课程号,课程名,学分,任课教师)

选修(学号,课程号,成绩,备注)



学号	姓名	出生年月	性别	入学年份	班 级
2006901	张伟	1988.01	男	2006	200602
2007912	王刚	1989.03	男	2007	200705
	•••	•••		•••	

学号	课程号	成绩	备注
2006901	CS001	计算机	重修
2007912	CS002	计算机	正常
•••	•••	•••	•••



- 关系必须是规范化的,满足一定的规范条件;
- 最基本的规范条件是: 关系的每一个分量必须是一个不可分的数据项。
- 下表不符合关系模型要求

职工号姓名	州夕	名即称	工资			扣除		实发
	姓石		基本工资	岗位津贴	业绩津贴	三险	个人所得税	
86051	陈平	讲师	1305	1200	1850	160	112	4083
•	•	•	•	•	•	•	•	•



- 为了维护数据库中的数据与现实世界的一致性, 需要对数据施加一定的约束条件
 - 实体完整性
 - ■参照完整性
 - ■用户自定义完整性

关系模型的数据操纵

- 关系模型中,对关系中的数据可进行查询、插入、 删除和修改操作;
- 在关系数据库系统中,对数据的全部操作都可以归结为对关系的运算;
 - 对关系可以进行多种运算,运算结果形成一个新关系,操作对象和操作结果都是关系。
 - 关系运算分为: 关系代数和关系演算。
- 数据的存取对用户是透明的,
 - ——用户只要指出"做什么",不必详细说明"怎么做"。

关系模型的优点

- 数据结构简单
- 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
 - 数据的检索结果也是关系
- 存取路径对用户透明
 - 具有更高的数据独立性,更好的安全性
 - 简化了编程工程和数据库开发建立的工作



- 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查询优化, 增加了开发数据库管理系统的难度;
- 语义比较贫乏,不能明确表示实体间的联系,难以 描述复杂对象。

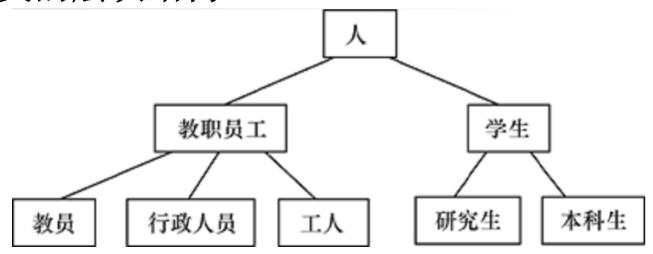
2.5 面向对象数据模型

- 20世界80年代初提出来的
- 源于面向对象程序设计
- 数据库技术与面向对象程序设计方法学相结合的产物
- 数据结构不是记录,而是具有复杂结构的对象
- 有助于解决传统数据模型的不足
- 许多课题有待进一步研究



- 面向对象数据库系统支持的核心概念
 - 对象
 - 类
 - ■消息
 - ■方法
 - ■封装
 - 继承

类的层次结构



- 教员、行政人员、工人中只有本身的特殊属性和方法
- 同时它们又继承教职员工类和人的所有属性和方法
- 逻辑上它们具有人、教职员工和本身的所有属性和方法

类的其他结构

本章小结

- 数据模型:
 - 信息世界的概念模型
 - 表示信息结构的模型,不涉及信息在计算机中的表示
 - 常用的概念模型有实体-联系模型,即E-R模型。
 - E-R模型可用于数据库设计的概念建模
 - 数据库系统支持的数据模型
 - 层次数据模型
 - 网状数据模型
 - 关系数据模型
 - ■面向对象数据模型

•

思考题

- 数据模型的螺旋式发展
- 教材习题

课外阅读:数据库界的四位图灵奖得主