

数据库原理

The Theory of Database System

第一章 绪论



中国矿业大学计算机学院



中国矿业大学数据库原理精品课程

第一章 绪论

1.1 数据库、数据库管理系统和数据库系统

1.2 数据库系统的产生与发展

1.3 数据模型

1.4 数据库系统结构

1.5 数据库管理系统

1.6 数据库应用系统常见的几种结构



1.3 逻辑数据模型

1.3.1 逻辑数据模型的组成要素

1.3.2 常用逻辑数据模型



1.3.1 数据模型的组成要素

- 数据结构
- 数据操作
- 数据的约束条件



1. 数据结构

- 什么是数据结构
 - 对象类型的集合
- 两类对象
 - 与数据类型、内容、性质有关的对象
 - 与数据之间联系有关的对象
- 数据结构是对系统静态特性的描述



2. 数据操作

➤ 数据操作

- 对数据库中各种对象（型）的实例（值）允许执行的**操作**及有关的**操作规则**。

➤ 数据操作的类型

- 检索
- 更新（包括插入、删除、修改）



数据操作(续)

➤ 数据模型对操作的定义

- 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. 数据的约束条件

➤ 数据的约束条件

- 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储存规则，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。



数据的约束条件(续)

➤ 数据模型对约束条件的定义

- 反映和规定本数据模型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。
- 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。



1.3.2 常用数据模型

➤ 层次模型

➤ 网状模型

➤ 关系模型



一、层次模型

1. 层次数据模型的数据结构
2. 层次数据模型的数据操纵
3. 层次数据模型的与完整性约束
4. 层次数据模型的存储结构
5. 层次数据模型的优缺点
6. 典型的层次数据库系统



1. 层次数据模型的数据结构

➤ 层次模型

满足下面两个条件的基本层次联系的集合为层次模型。

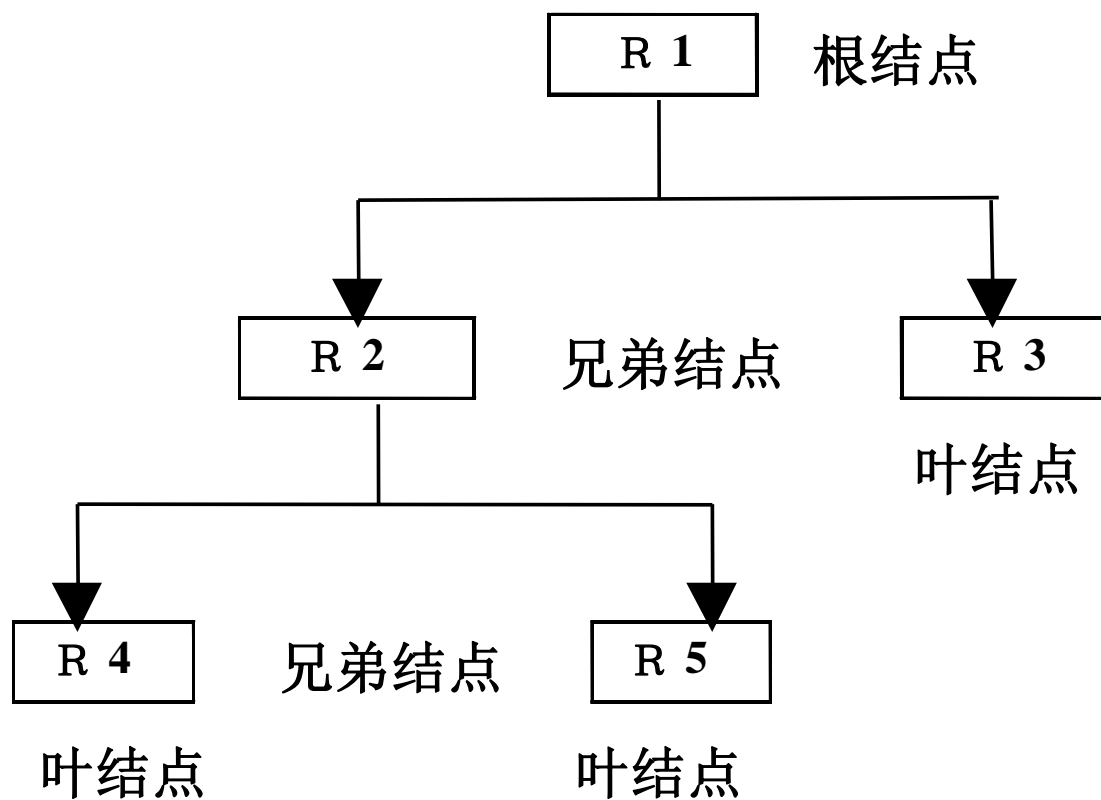
1. 有且只有一个结点没有双亲结点，这个结点称为根结点
2. 根以外的其它结点有且只有一个双亲结点

➤ 层次模型中的几个术语

- 根结点，双亲结点，兄弟结点，叶结点



层次数据模型的数据结构(续)



层次数据模型的数据结构(续)

➤表示方法

实体型：用记录类型描述。

每个结点表示一个记录类型。

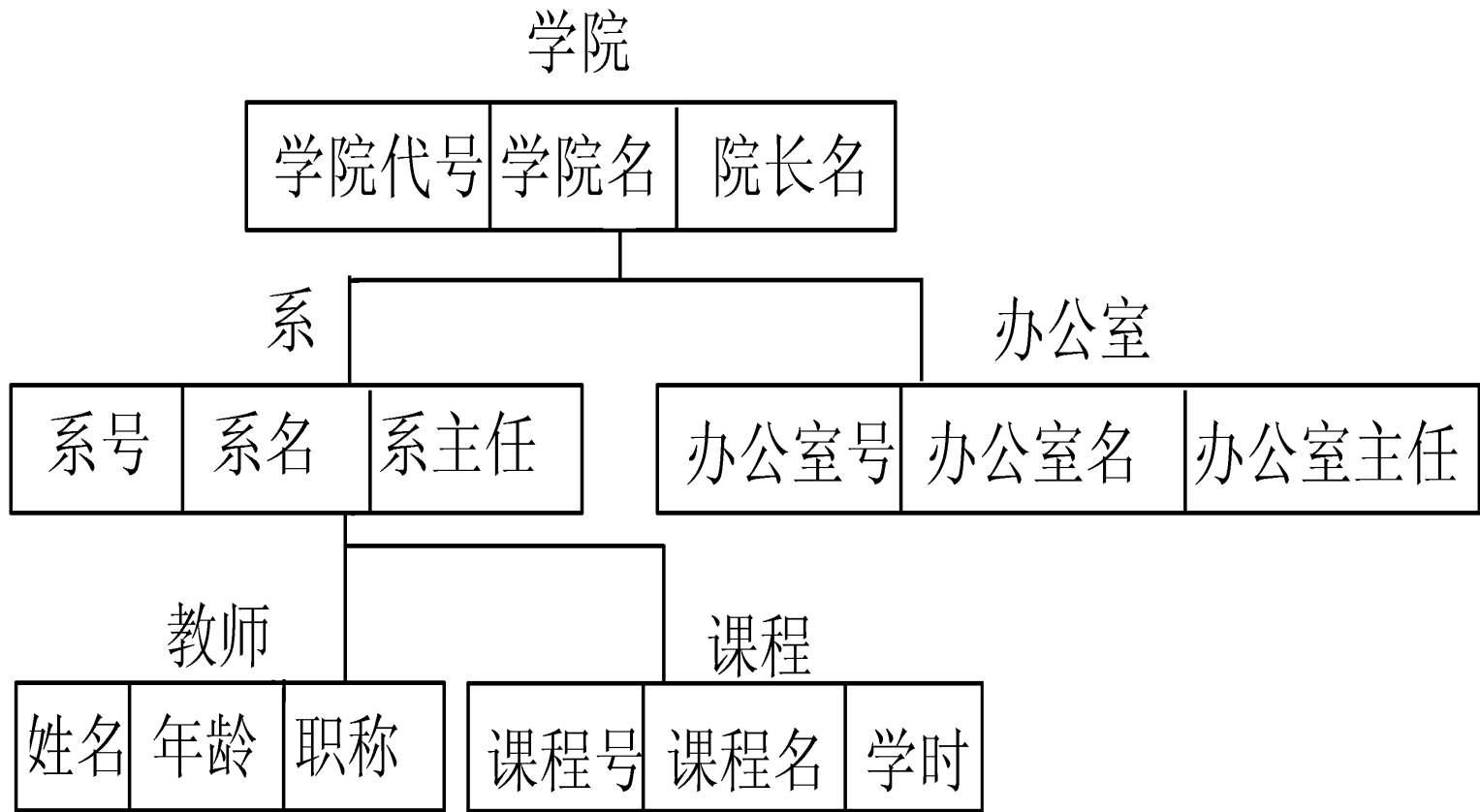
属性：用字段描述。每个记录类型可包含若干个字段。

联系：用结点之间的连线表示记录（类）型之间的

一对多的联系



实例：学院数据模型



层次数据模型的数据结构(续)

➤ 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标识字段，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)

➤ 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

- 方法

 - 将多对多联系**分解**成一对多联系

- 分解方法

 - 冗余结点法
 - 虚拟结点法



2. 层次模型的数据操纵

- 查询
- 插入
- 删除
- 更新



3. 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
- 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. 层次数据模型的存储结构

➤ 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

用指引元来反映数据之间的层次联系

- 子女—兄弟链接法
- 层次序列链接法



5. 层次模型的优缺点

➤ 优点

- 层次数据模型简单，对具有一对多的层次关系的部门描述自然、直观，容易理解
- 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
- 查询子女结点必须通过双亲结点
- 层次命令趋于程序化



6. 典型的层次数据库系统

IMS数据库管理系统

- 第一个大型商用**DBMS**
- **1968**年推出
- **IBM**公司研制



1.3.2 常用数据模型

➤ 层次模型

➤ 网状模型

➤ 关系模型



二、网状模型

1. 网状数据模型的数据结构
2. 网状数据模型的数据操纵
3. 网状数据模型的完整性约束
4. 网状数据模型的存储结构
5. 网状数据模型的优缺点
6. 典型的网状数据库系统



1. 网状数据模型的数据结构

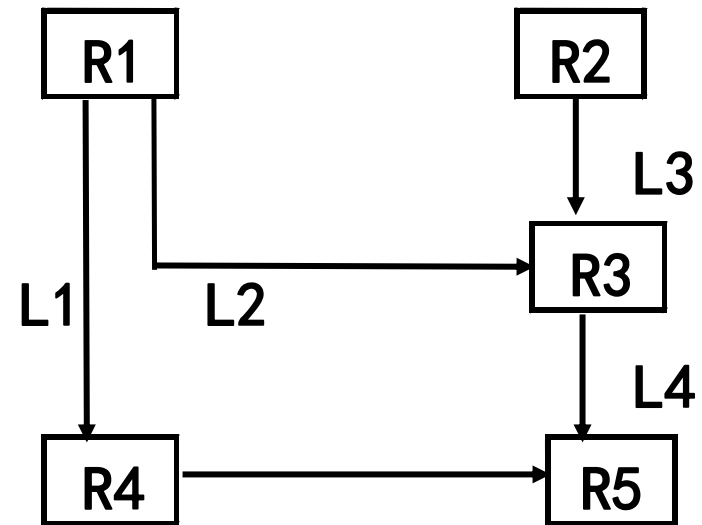
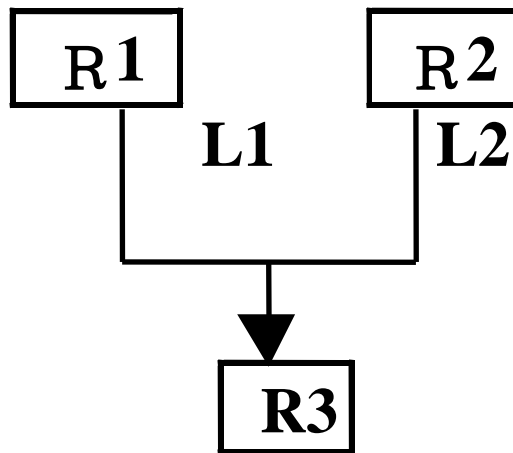
➤ 网状模型

满足下面两个条件的基本层次联系的集合为网状模型。

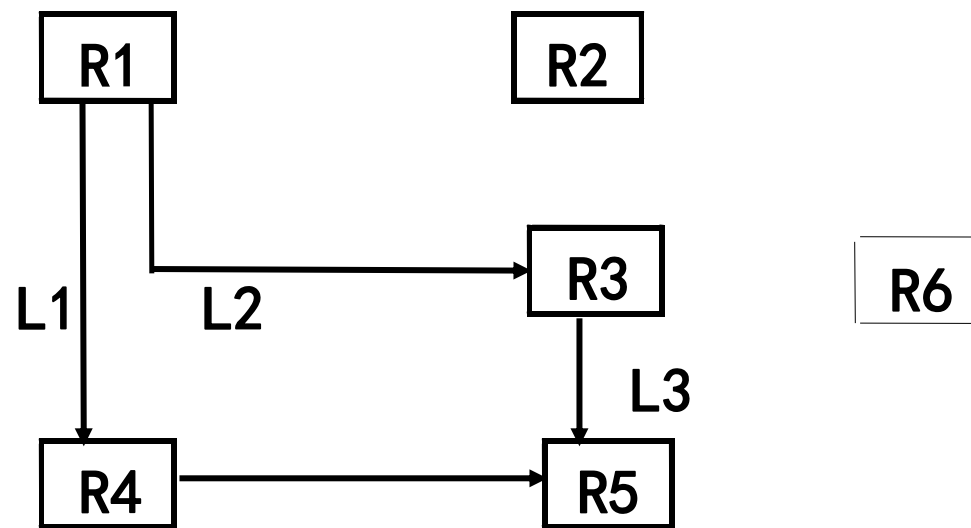
1. 允许一个以上的结点无双亲；
2. 一个结点可以有多个的双亲。



网状数据模型的数据结构



网状数据模型的数据结构(续)



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）型之间的一对多的父子联系。



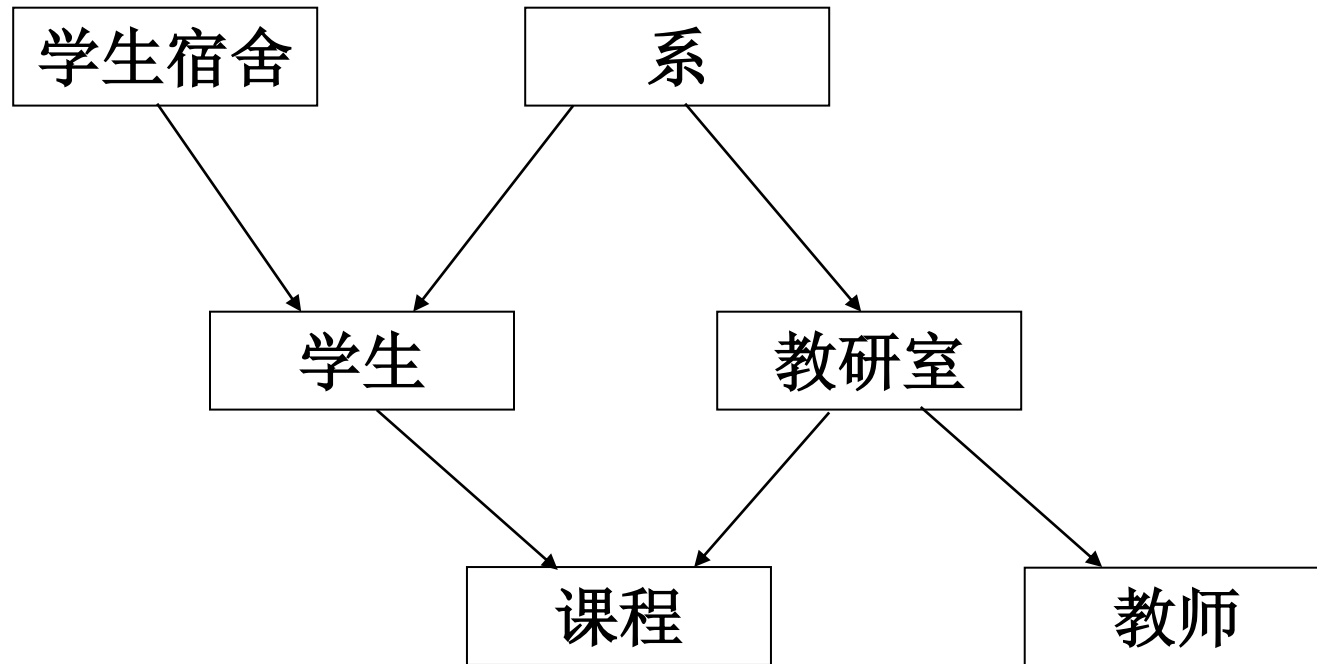
网状数据模型的数据结构(续)

➤ 特点

- 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
- 任何记录值只有按其路径查看时，才能显出它的全部意义



网状数据模型的数据结构(续)



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)

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



2. 网状模型的数据操纵

- 查询
- 插入
- 删除
- 更新



3. 网状数据模型的完整性约束

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
- 双亲结点与子女结点之间是一对多联系
- 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



4. 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



5. 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



6. 典型的网状数据库系统

- **DBTG系统，亦称CODASYL系统**
 - 由**DBTG**提出的一个系统方案
 - 奠定了数据库系统的基本概念、方法和技术
 - **70年代推出**
- **实际系统**
 - **Cullinet Software Inc.公司的 IDMS**
 - **Univac公司的 DMS1100**
 - **Honeywell公司的IDS/2**
 - **HP公司的IMAGE**



1.3.2 常用数据模型

- 层次模型
- 网状模型
- 关系模型



三、关系模型

1. 关系数据模型的数据结构
2. 关系数据模型的操纵
3. 关系数据模型的完整性约束
4. 关系数据模型的存储结构
5. 关系数据模型的优缺点
6. 典型的关系数据库系统



关系模型

- 最重要的一种数据模型。也是目前主要采用的数据模型
- 1970年由美国IBM公司San Jose研究室的研究员E.F.Codd提出
- 本课程的重点



1. 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
95004	王小明	19	女	社会学	95
95006	黄大鹏	20	男	商品学	95
95008	张文斌	18	女	法律学	95
...



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**)

表中的某个最小属性组，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

对关系中信息内容结构的描述

$R(U, D, \text{dom}, I, F)$



学生关系模式

- 关系的名称(**R**): 学生
- 包含的属性(**U**): {学号, 姓名, 性别, 出生年份, 所在院系.....}
- 每个属性的取值范围(**D**): {000000000~999999999, (男, 女),}
- 属性和取值范围的映射(**dom**):
 - 学号 \leftrightarrow 000000000~999999999
 - 性别 \leftrightarrow (男, 女)
- 完整性约束条件(**I**)
- 属性间的函数依赖(**F**)



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。
- **一对一联系**：可以隐含在实体对应的关系中。
- **一对多联系**：可以隐含在实体对应的关系中。
- **多对多联系**：直接用关系表示。



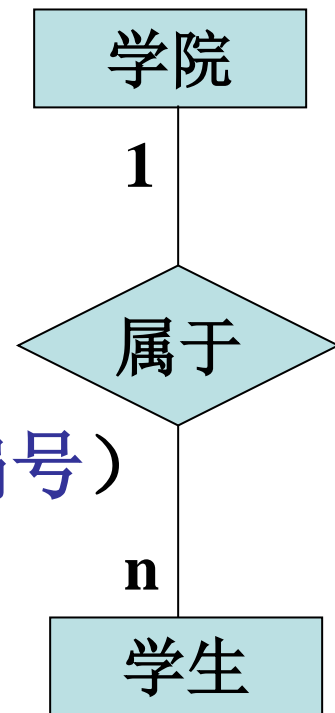
关系数据模型的数据结构(续)

例1:

学生、学院、学生与学院之间的一对多联系

学生 (学号, 姓名, 年龄, 性别, 年级, 学院编号)

学院 (学院编号, 学院名称, 办公地点)



1:n联系



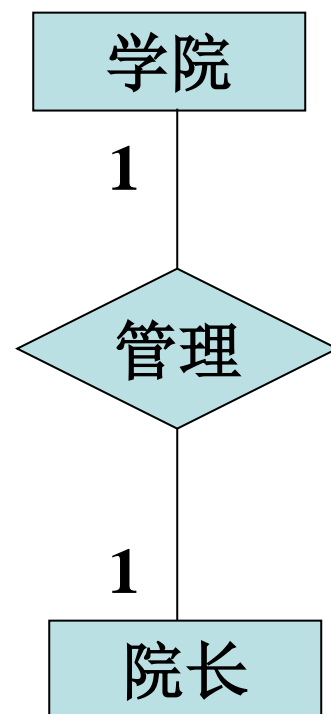
关系数据模型的数据结构(续)

例2:

学院、院长、学院与院长间的一对一联系

学院（学院编号，学院名称，办公地点）

院长（职工号，姓名，年龄，职务，职称）



1:1联系



关系数据模型的数据结构(续)

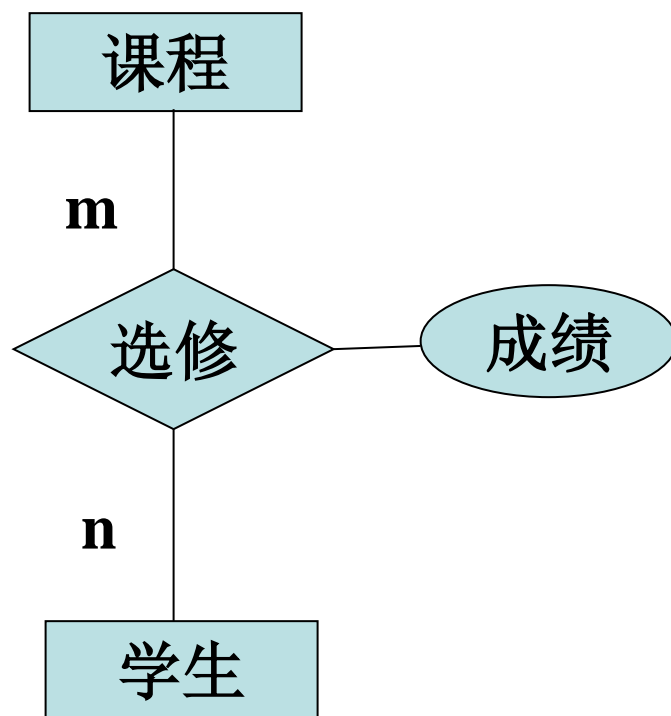
例3:

学生、课程、学生与课程之间的多对多联系

学生（学号，姓名，年龄，……）

课程（课程号，课程名，学分）

选修（学号，课程号，成绩）



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。

职工号	姓名	职称	工资			扣除		实发
			基本	工龄	职务	房租	水电	
0802110	陈平	讲师	1620	126	220	360	105	1501
.



2.关系模型的数据操纵

- 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”。



3. 关系模型的完整性约束

- 实体完整性
- 参照完整性
- 用户定义的完整性



4. 关系数据模型的存储结构

- 表以文件形式存储
- 有的DBMS一个表对应一个操作系统文件
- 有的DBMS自己设计文件结构



5. 关系模型的优缺点

➤ 优点

- 建立在严格的数学概念的基础上
- 概念单一。数据结构简单、清晰，用户易懂易用
 - 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



关系模型的优缺点（续）

➤ 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，增加了开发数据库管理系统的难度



6. 典型的关系数据库系统

- **SQL Server**
- **Oracle**
- **MySQL**
- **SQL Lite**
- **SYBASE**
- **DB/2**

