

新编考研辅导丛书
(2003版)

Database System

数据库系统原理辅导

王亚平 编著

核心内容阐述
典型例题分析
常见错误点评
全真考题解答

西安电子科技大学出版社

<http://www.xduph.com>



新编考研辅导丛书(2003 版)

数据库系统原理辅导

王亚平 编著

西安电子科技大学出版社

2003

内 容 简 介

本书为西安电子科技大学出版社规划出版的“考研辅导丛书”之一。其内容包括数据库系统原理的基本概念,要点以及典型题剖析,模拟题及解答,还有近年来研究生入学考试题等。本书的目的是让学生真正了解和掌握该课程基本概念及解题思路与方法,进一步提高分析问题和解决问题的能力与技巧。

本书共9章,每章通过大量的典型题解析、习题(选择题、填空题、简答题、综合题)解答,揭示了数据库系统原理课程学习过程中实际问题的求解方法。全书编写力求做到精益求精,真正激励读者思维。

本书既可以作为报考计算机专业等学科的硕士研究生的复习用书,也可作为在校本科学生学习或教师讲授“数据库系统原理”课程的参考书。

新编考研辅导丛书(2003版)

数据库系统原理辅导

王亚平 编著

策 划 李惠萍

责任编辑 阎彬

出版发行 西安电子科技大学出版社(西安市太白南路2号)

电 话 (029)8227828 邮 编 710071

http://www.xduph.com E-mail: xdupfxb@pub.xaonline.com

经 销 新华书店

印 刷 西安文化彩印厂

版 次 2003年1月第1版 2003年1月第1次印刷

开 本 787毫米×960毫米 1/16 印张 17.375

字 数 309千字

印 数 1~6000册

定 价 21.00元

ISBN 7-5606-1041-2/TP·0514

XDUP 1312A01-1

如有印装问题可调换

前 言

“数据库系统原理”是计算机应用、计算机软件及相关专业的主干课程，也是硕士研究生入学考试课程。本书是为了配合课程的学习，帮助读者抓住重点，高效率地复习，适应考题类型，在应试中正常发挥水平而专门编写的。

本书适用的对象是：本科学生、报考计算机专业硕士研究生的考生、数据库系统概论的自学者和担任“数据库系统概论”课程的教师。

全书共分9章，主要内容如下：

第1章：数据库系统概论。本章主要介绍了数据库系统中的基本概念，数据模型及数据模型的分类，概念模型及概念模型的表示方法，数据库系统的结构及其组成等内容。通过例题及习题进一步加深对这些内容的理解。

第2章：关系数据库。本章通过例题及习题帮助读者理解关系数据库的基本概念，掌握如何用关系代数表达式、元组演算表达式以及域演算表达式来表达实际查询问题，以及如何将上述三类表达式互相转换的方法。

第3章：关系数据库标准语言 SQL。本章主要介绍了 SQL 的功能特点和体系结构，如何利用 SQL 定义、修改、删除表，如何定义、删除视图和索引，如何利用 SQL 的 SELECT 完成数据的查询，如何利用 SQL 的 UPDATE、INSERT 完成数据的修改、插入操作，嵌入式 SQL 的使用方法，以及给定一个关系代数表达式，如何用 SQL 语句描述等内容。

第4章：关系系统及查询优化。本章介绍了查询优化的基本概念、目的、实现方法等内容。

第5章：关系数据库理论。本章是非常重要而且也是比较难理解的章节，特别是对函数依赖及 Armstrong 公理系统的理解。为什么要对模式进行分解？如何分解？如何判断关系模式达到几范式？如何求属性的闭包及如何求最小函数依赖集？如何判断分解后的关系模式既是无损连接又保持函数依赖？对于这一系列问题，本章和自测题中都有很多典型的实例及求解的方法。

第6章：数据库设计。本章重点介绍了数据库设计各个阶段的步骤和任务，通过实例特别详细介绍了逻辑结构设计阶段如何将概念模型（E-R 图）转换为关系模型的方法。

第7章：数据库安全。本章重点介绍了安全性的基本概念，事务的基本概念，

事务的特性，故障的类型及恢复方法，并发控制带来的新问题，封锁及封锁协议，并发调度的可串行性，两段锁协议，完整性约束条件，完整性控制等内容。

第 8 章：面向对象数据系统。本章重点介绍了面向对象数据库系统的基本概念，对象联系图以及对象关系数据库（ORDB）等内容。

第 9 章：分布式数据库系统。本章重点介绍了分布式数据库系统的基本概念，分布式数据库系统的体系结构，查询优化，事务管理等内容。

数据库技术发展很快，需要学习的内容也不断变化，书中难免有需要改进和更新的地方，我们衷心地希望本书能够得到广大读者的支持和帮助，但愿我们的书给您送去一份希望！

编 者

2002 年 10 月

于西安电子科技大学

目 录

第一部分 数据库系统原理的基本概念、要点及典型题剖析

第 1 章 数据库系统概论	2
1.1 基本概念及复习要点	2
1.2 典型例题分析	14
1.3 自测题	17
第 2 章 关系数据库	23
2.1 基本概念及复习要点	23
2.2 典型例题分析	34
2.3 自测题	44
第 3 章 关系数据库标准语言 SQL	49
3.1 基本概念及复习要点	49
3.2 典型例题分析	60
3.3 自测题	68
第 4 章 关系系统及其查询优化	72
4.1 基本概念及复习要点	72
4.2 典型例题分析	76
4.3 自测题	82
第 5 章 关系数据理论	86
5.1 基本概念及复习要点	86
5.2 典型例题分析	106
5.3 自测题	114
第 6 章 数据库设计	122
6.1 基本概念及复习要点	122
6.2 典型例题分析	127
6.3 自测题	130

第 7 章 数据库保护	135
7.1 基本概念及复习要点	135
7.2 典型例题分析	144
7.3 自测题	149
第 8 章 面向对象的数据库系统	157
8.1 基本概念及复习要点	157
8.2 典型例题分析	161
8.3 自测题	162
第 9 章 分布式数据库系统	165
9.1 基本概念及复习要点	165
9.2 典型例题分析	169
9.3 自测题	171

第二部分 模 拟 试 题

模拟试题 1.....	174
模拟试题 2.....	175
模拟试题 3.....	176
模拟试题 4.....	178

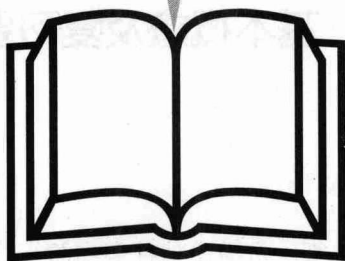
第三部分 历年攻读计算机专业硕士研究生入学考试试题 (西安电子科技大学)

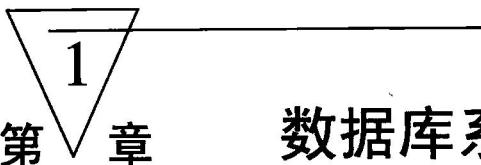
1997 年试题.....	182
1998 年试题.....	183
1999 年试题.....	185
2000 年试题.....	186
2001 年试题.....	188
2002 年试题.....	190
附录 自测题及模拟试题答案.....	192
第 1 章自测题答案	192
第 2 章自测题答案	197
第 3 章自测题答案	201

第4章自测题答案	208
第5章自测题答案	212
第6章自测题答案	225
第7章自测题答案	232
第8章自测题答案	242
第9章自测题答案	243
模拟试题1 解答	246
模拟试题2 解答	253
模拟试题3 解答	256
模拟试题4 解答	263
参考文献	269

第一部分

数据库系统原理的基本概念、要点及典型题剖析





第 1 章

数据库系统概论

内 容 提 要

数据库系统(DBS)本质上是一个用计算机存储记录的系统。数据库管理系统(DBMS)是介于用户与操作系统之间的一种数据管理软件,其基本目标是提供一个可以方便地、有效地存取数据库信息的环境。数据库本身被看作为一种电子文件柜,它是收集计算机数据文件的仓库或容器,系统用户可以对这些文件执行一系列的操作。

设计数据库系统的目的是为了管理大量信息。对数据的管理涉及到信息存储结构的定义,信息操作机制的提供,安全性保证,以及多用户对数据的共享问题。

本章重点介绍了如下内容:

- ◇ DB、DBMS、DBS 含义及它们之间的关系;
- ◇ 三个重要的数据模型;
- ◇ 如何用 E-R 模型描述现实世界的问题;
- ◇ 数据库系统的结构与组成;
- ◇ 典型例题分析及自测题。

1.1 基本概念及复习要点

1.1.1 基本概念

1. 数据(Data)

数据是数据库存储的基本对象,是描述某种事物的符号记录。这里的数据并非一般意义上的数字,它包含有很多种类,如:数字、文字、图形、图像、声音、语言等。

2. 数据库(DataBase, 简称 DB)

数据库是指长期储存在计算机内的, 有组织的, 可共享的数据集合。数据库中的数据按一定的数学模型组织、描述和储存, 具有较小的冗余度, 较高的数据独立性和易扩展性, 并可被各种用户共享。

特别需要说明的是, 讲到数据库时必须区分数据库模式和数据库实例。数据库模式是数据库的逻辑设计, 而数据库实例是给定时刻数据库中数据的一个快照。

3. 数据库管理系统(DataBase Management System, 简称 DBMS)

DBMS 是解决如何科学地组织和储存数据, 如何高效地获取和维护数据的一个介于用户和操作系统之间的系统软件。其主要功能有以下几方面:

- (1) 数据定义功能;
- (2) 数据操纵功能;
- (3) 数据库的运行管理;
- (4) 数据库的建立与维护。

4. 数据库系统(DataBase System, 简称 DBS)

数据库系统由数据库、数据库管理系统及其开发工具、应用系统、数据库管理员和用户构成。需要说明的是, 数据库的建立、使用和维护等工作仅靠 DBMS 远远不够, 还要由专门的人员来完成, 这些人被称为数据库管理员(DataBase Administrator, 简称 DBA)。

5. 数据的独立性

数据库中数据的独立性是指数据与程序独立, 将数据的定义从程序中分离出去, 由 DBMS 负责数据的存储, 从而简化应用程序的编制。数据的独立性是由 DBMS 的二级映射功能来保证的。数据的独立性包括数据的物理独立性和数据的逻辑独立性。

数据的物理独立性是指用户的应用程序与存储在磁盘上的数据库中的数据是相互独立的, 即存储在磁盘上的数据由 DBMS 统一管理, 应用程序处理的只是数据的逻辑结构, 这样即使数据的物理结构改变了, 应用程序也不用改变。

数据的逻辑独立性是指用户的应用程序与数据库的逻辑结构是相互独立的。数据的逻辑结构发生变化后, 用户程序也可以不修改。

1.1.2 模型

模型是现实世界特征的模拟和抽象。模型分为两类: 概念模型和数据模型。

(1) 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第一层抽象。概念模型强调其语义表达功能, 易于被用户理解, 是用户和数据库设计人员交流的语言, 主要用于数据库设计。

(2) 数据模型: 是按计算机系统的观点对数据建模, 是现实世界数据特征的抽象, 用于 DBMS 的实现。它主要包括层次模型(Hierarchical Model)、网状模型(Network Model)、关系模型(Relation Model)和面向对象模型(Object Oriented Model), 其中前两种为非关系模型。

一、概念模型

我们知道, 计算机只能处理数据, 所以首先要解决的问题是按用户的观点对数据和信息建模, 然后按计算机系统的观点对数据建模。换句话说, 就是要解决现实世界的问题如何表达为信息世界的问题, 以及信息世界的问题如何在具体的机器世界表达的问题。图 1-1 所示的是现实世界客观对象的抽象过程。

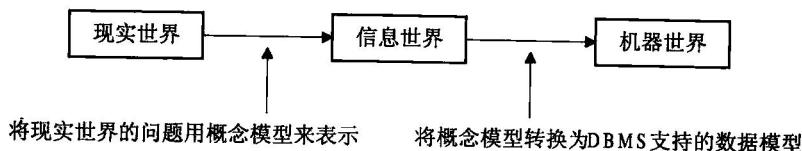


图 1-1 现实世界客观对象的抽象过程

1. 基本概念

实体: 表示现实(客观)世界中存在的并能区别的事物(可以是人或物)。

属性: 实体某方面的特性。

码: 能唯一区分实体的属性集。

域: 属性的取值范围, 例如, 姓名的域为字符串的集合, 性别的域为男、女等。

实体型: 具有相同属性的实体必然具有共同的特征和性质, 将实体名及属性名的集合抽象并刻画同类实体称为实体的类型, 如表示学生学习情况的属性的集合学生(S#, SN, SAG, SG, SD, C#……), 就是一个实体型。

实体集: 具有共同特性的实体的集合。

联系: 描述实体集之间的依赖关系。联系的类型: 1:1; 1:m; m:n。

信息世界和机器世界术语的对应关系见表 1-1。

表 1-1 术语的对应关系

信息世界	机器世界
实体	记录
属性	字段(数据项)
实体集	文件
实体标识符	码

2. E-R 方法

概念模型中最常用的方法为实体—联系方法, 简称 E-R 方法。该方法直接从现实世界中抽象出实体和实体间的联系, 然后用 E-R 图来表示数据模型。在 E-R 图中实体用方框□表示; 联系用菱形◇表示, 并且用边将其与有关的实体连接起来, 并在边上标上联系的类型; 属性用椭圆○表示, 并且用边将其与相应的实体连接起来。

(1) 两个不同实体集之间存在 $1:n$ (一对多) 和 $m:n$ (多对多) 联系, 例如部门 **DEPT** 和职工 **EMP** 两个实体集, 若一个职工只能属于一个部门, 那么这两个实体集之间应是一个 $1:n$ 的联系, 用 E-R 图表示如图 1-2(a) 所示。又如, 工程项目 **PROJ** 和职工 **EMP** 两个实体集, 若一个职工可以参加多个项目, 一个项目可以由多个职工参加, 那么这两个实体集之间应是一个 $m:n$ 的联系, 用 E-R 图表示如图 1-2(b) 所示。

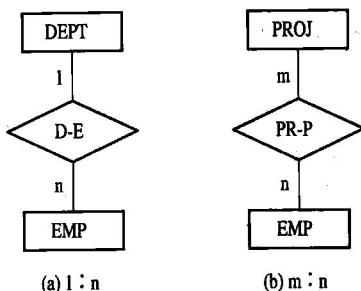


图 1-2 两个不同实体集之间的 $1:n$ 和 $m:n$ 联系

(2) 两个以上不同实体集之间的 $1:n:m$ 、 $r:n:m$ 联系的表示方法如图 1-3 所示。其中, 图 1-3(a) 表示供应商 **Supp**、项目 **Proj** 和零件 **Part** 之间的 $r:n:m$ 的联系; 图 1-3(b) 表示病房、病人和医生之间 $1:n:m$ 的联系。

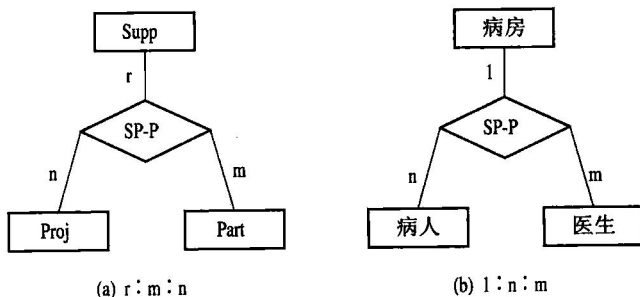


图 1-3 三个不同实体集之间的 $r:n:m$ 和 $1:n:m$ 联系

(3) 同一实体集内的二元联系, 如图 1-4 所示。从图中可见, 职工实体集中的领导与被领导的联系是 $1:n$ 的。但是, 职工实体集中的婚姻联系是 $1:1$ 的。

需要说明的是, E-R 模型是软件工程设计中的一个重要方法, 因为它接近于人的思维方式, 容易理解并且与计算机无关, 所以用户容易接受。但是, E-R 模型只能说明实体间的语义联系, 还不能进一步地说明详细的数据结构。一般遇到实际问题时, 应先设计一个 E-R 模型, 然后再把它转换成计算机能接受的数据模型。

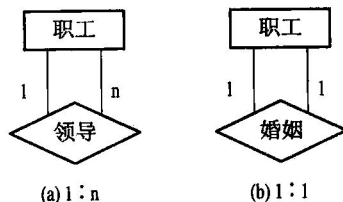


图 1-4 同一实体集内的 $1:n$ 和 $1:1$ 联系

二、数据模型

数据库管理技术进入数据库阶段的标志是三种数据模型(层次模型、网状模型、关系模

型)的推出。组成数据模型的三要素是数据结构、数据操作、数据的约束条件(指一组完整性约束条件的集合)。

(1) 数据结构是所研究的对象类型的集合,是对系统静态特性的描述。

(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,包括操作及操作规则。操作指检索、插入、删除、修改;操作规则是按照优先级别来操作。

数据操作与数据结构的区别是数据操作是对数据库动态特性的描述,而数据结构是对数据库静态特性的描述。

(3) 数据的约束条件是完整性规则的集合。例如,对于具体的应用,数据必须遵循特定的语义约束条件,如在某个人事档案管理数据库中,规定人的年龄必须大于 18 岁且小于 150 岁;每个职工可担任一个工种等,这都可以通过建立数据的约束条件来实现。

一、层次模型

定义: 满足以下两个条件的基本层次联系的集合称为层次模型。

(1) 有且仅有一个结点无双亲,该结点称之为根结点;

(2) 根以外的其它结点有且仅有一个双亲。

1) 层次模型的表示

在层次模型中,分别用记录和链接表示数据和数据间的联系。其中,每一个结点表示一个记录类型,记录(类型)之间的联系用结点之间的连线(有向边)表示。例如:学校中的“系教课”模型如图 1-5 所示(层次模型及实例)。

2) 多对多联系在层次模型中的表示方法

层次模型只能表示一对多(包括一对一)的联系,但为了表示多对多的联系,可采用如下两种方法。

方法 1: 采用冗余结点法。两个实体的多对多的联系转换为两个一对多的联系。

该方法的优点是结点清晰,允许结点改变存储位置。缺点是需要额外的存储空间,有潜在的数据不一致性。

方法 2: 采用虚拟结点分解法。将冗余结点转换为虚拟结点。虚拟结点是一个指引元,指向所代替的结点。

该方法的优点是减少对存储空间的浪费,避免数据不一致性。缺点是改变存储位置可能引起虚拟结点中指针的修改。

例如:学生与课程,学生可以选多门课程,一门课程可以由多个学生选取,所以学生与课程之间是一个多对多的联系,如图 1-6(a)所示。采用冗余结点法将其转换为两个一对多的联系,如图 1-6(b)所示。采用虚拟结点分解法,将冗余结点转换为虚拟结点,如图 1-6(c)

所示。

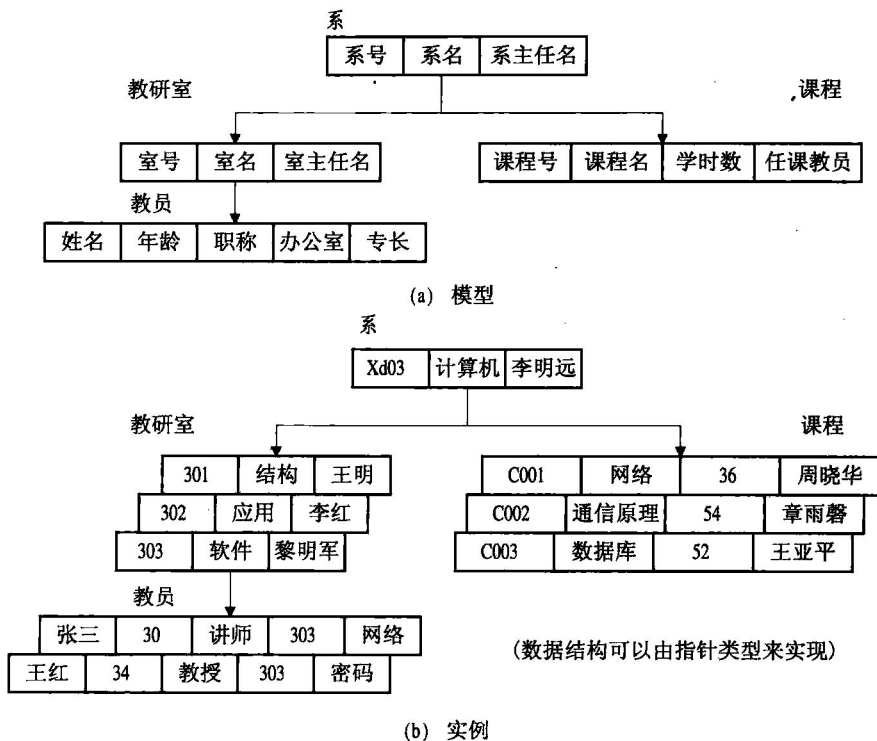


图 1-5 “系教课”层次模型及实例

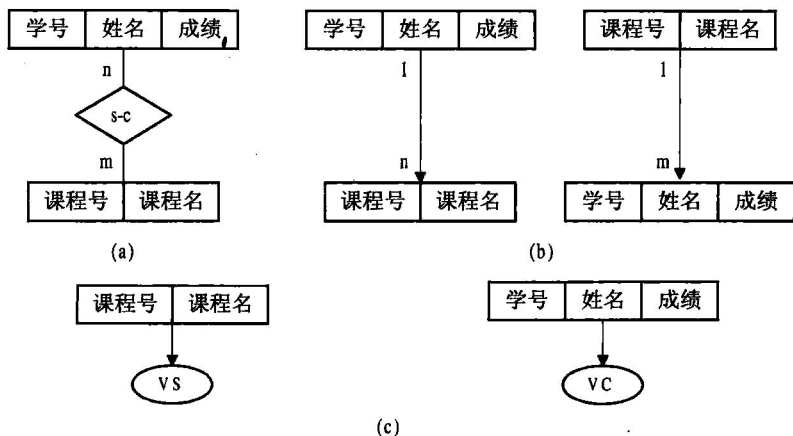


图 1-6 多对多联系在层次模型中的表示

3) 层次模型的存储结构

层次模型的存储结构采用两种方法:邻接法和链接法。

(1) 邻接法:按照层次树前序穿越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来体现(或隐含)层次顺序。

(2) 链接法:用指引元来反映数据间的层次联系。

4) 层次模型的优缺点

(1) 层次模型的优点:

- ✧ 层次数据模型比较简单;
- ✧ 对实体间的联系固定;
- ✧ 提供了良好的完整性支持。

(2) 层次模型的缺点:

✧ 只能表示 1:n 的联系,尽管有许多辅助手段可以实现 m:n 的联系,如冗余结点(易产生数据的不一致性)和虚拟结点(改变存储位置可能会引起虚拟结点指针的变化);

- ✧ 插入或删除操作的限制比较多;
- ✧ 查询子女结点必须通过双亲结点;
- ✧ 由于结构严谨,层次命令趋于程序化。

二、网状模型

定义:满足以下两个条件的基本层次联系的集合称为网状模型。

- (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。

网状模型(也称 DBTG 模型)是一个比层次模型更具普遍性的数据结构,是层次模型的一个特例。网状模型可以更直接地描述现实世界,因为它去掉了层次模型的两个限制,允许多个结点没有双亲,允许结点有多个双亲,并且允许两个结点之间有多种联系(称之为复合联系)。

网状模型中的每个结点都表示一个记录类型(实体),每个记录类型都可以包含若干字段(实体的属性),结点间的连线表示记录类型之间一对多的联系。

1) 层次模型和网状模型的主要区别

层次模型和网状模型的主要区别如下:

- (1) 网状模型中子女结点与双亲结点的联系不唯一,因此需要为每个联系命名。
- (2) 网状模型允许复合联系,即两个结点之间有两种以上的联系,如图 1-7(a)所示。
- (3) 网状模型不能表示记录之间的多对多的联系,为此需要引入联结记录来表示多对多

的联系。

例如：学生与课程之间的 E-R 图不能直接用网状模型表示，因为一个学生可以选若干门课，而一门课可以被多个学生选。为此，引入选课联结记录，如图 1-7(b)所示。这样，学生与选课之间的 S-SC 是一对多的联系，课程与选课之间的 C-SC 也是一对多的联系。

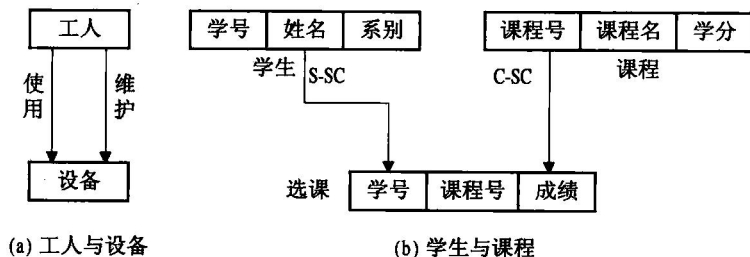


图 1-7 网状模型举例

2) 网状数据模型的操纵与完整性约束

网状数据模型没有层次模型那样严格的完整性约束条件，但 DBTG 在模式 DDL 中提供了定义 DBTG 数据库完整性的若干概念和语句，主要有：

- (1) 支持记录码的概念，码是能唯一标识记录的数据项的集合。
- (2) 保证一个联系中双亲记录和子女记录之间是一对多的联系。
- (3) 可以支持双亲记录和子女记录之间的某些约束条件。例如在图 1-7(b)中，当插入一条选课记录时，只有数据库中存在该学生及该课程，系统才认为是合法的操作。

3) 网状模型的优缺点

(1) 网状模型的优点主要有：

- ✧ 能更为直接地描述现实世界；
- ✧ 具有良好的性能，存取效率高。

(2) 网状模型的缺点主要有：

- ✧ 结构比较复杂，不利于最终用户使用；
- ✧ 其 DDL、DML 语言复杂，用户不易使用。

三、关系模型

1) 基本概念

关系模型最大特色是描述的一致性，对象(如实体集)和联系均用二维表格表示。例如：反映教学情况的数据，可用关系模型表示为下列的二维表格：