

关于本课程

第1章 绪论

本章要点

本章从数据库基本概念与知识出发，依次介绍了数据库系统的特点、数据模型的三要素及其常见数据模型、数据库系统的内部体系结构等重要概念与知识。本章的另一重点是围绕数据库管理系统（DataBase Management System, DBMS）介绍其功能、组成与操作，还介绍了数据库技术的研究点及其发展变化情况。

本章目录

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 数据库技术的研究领域及其发展*

1.6 小结

习题

1.1 数据库系统概述

1.1.1 数据，数据库，数据库管理系统，数据库系统

1.1.2 数据管理技术的产生和发展

1.1.3 数据库系统的特点

1.1.1 数据，数据库，数据库管理系统，数据库系统

- 数据
- 数据库
- 数据库管理系统
- 数据库系统

数据 (Data)

✓ 定义

数据是用来记录信息的可识别的符号，是信息的具体表现形式。

✓ 表现形式

数据是数据库中存储的基本对象。按广义的理解来说，数据的种类有很多，如文字、图像、声音、等等都可以转化为计算机可以识别的标识，可以以数字化后的二进制形式存入计算机。

在日常生活中人们可以对事务直接用自然语言描述。在计算机中存储和处理时，就要抽出对这些事物感兴趣的特征组成一个记录来描述。

与信息的联系

数据解释的含义称为数据的语义(即信息)，数据与其语义是不可分的。可以这样认为：数据是信息的符号表示或载体，信息则是数据的内涵，是对数据的语义解释。

数据库（DataBase，简称DB）

✓ 数据库

从字面意思来说就是存放数据的仓库。具体而言就是长期存放在计算机内的有组织的可共享的数据集合，可供多用户共享，数据库中的数据按一定的数据模型组织、描述和储存，具有尽可能小的冗余度和较高的数据独立性和易扩张性。

✓ 数据库具有两个比较突出的特点

①把在特定的环境中与某应用程序相关的数据及其联系集中在一块并按照一定的结构形式进行存储，即集成性。

②数据库中的数据能被多个应用程序的用户所使用，即共享性。

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

数据库管理系统是数据库系统的核心组成部分，是对数据进行管理的大型系统软件，用户在数据库系统中的一些操作都是由数据库管理系统来实现的。功能：

✓ 数据定义

DBMS提供数据定义语言 (Data Definition Language,简称DDL)，用户通过它可以方便地对数据库中的数据对象（包括表、视图、索引、存储过程等）进行定义。定义相关的数据库系统的结构和有关的约束条件。

✓ 数据操纵

DBMS提供数据操纵语言(Data Manipulation Language,简称DML)，通过DML操纵数据实现对数据库的一些基本操作，如查询，插入，删除和修改等等。其中，国际标准数据库操作语言——SQL语言就是DML的一种。

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

✓ 数据库的运行管理

这一功能是数据库管理的核心所在。DBMS通过数据库在建立、运用和维护时统一管理和控制，以保证数据安全、正确、有效的正常运行。DBMS主要通过数据的安全性控制、完整性控制、多用户应用环境的并发性控制和数据库数据的系统备份与恢复四个方面来实现对数据库的统一控制功能的。

数据库的建立和维护功能

数据库的建立和维护功能包括数据库初始数据的输入、转换功能、数据库的转储、恢复功能、重组织功能和性能监视、分析功能等。

数据库系统 (DataBase System, 简称DBS)

数据库系统是指在计算机系统中引入数据库后的系统构成，主要有数据库（及相关硬件）、数据库管理系统（及开发工具）（及相关软件）、应用系统、数据库管理员和用户这几部分。其中，在数据库的建立、使用和维护的过程要有专门的人员来完成，这些人被称为数据库管理员(DataBase Administrator,简称DBA)。

数据库系统可以用 [图1.1](#) 表示。

数据库系统在整个计算机系统中的地位如 [图1.2](#) 所示。

图 1.1 数据库系统

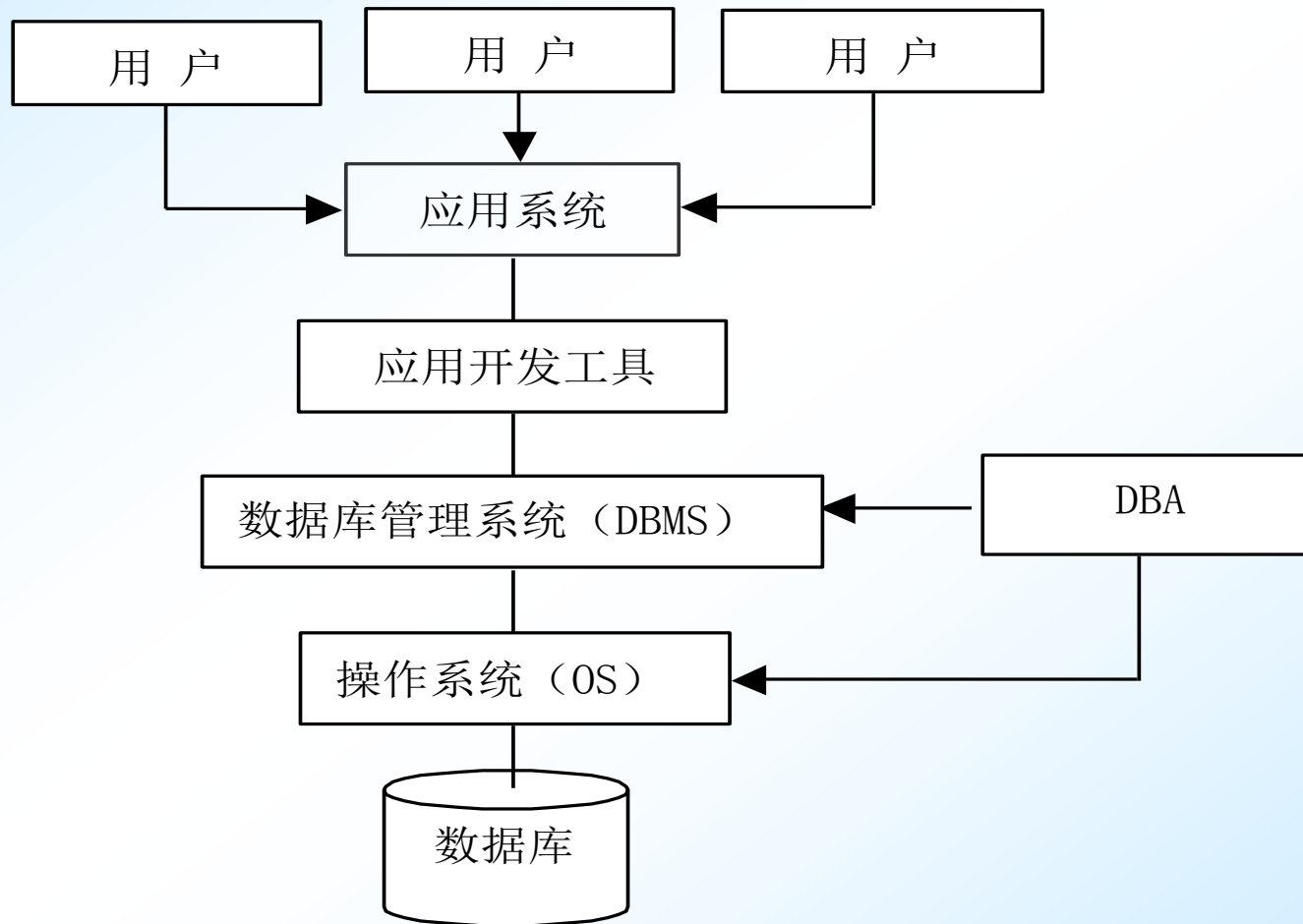
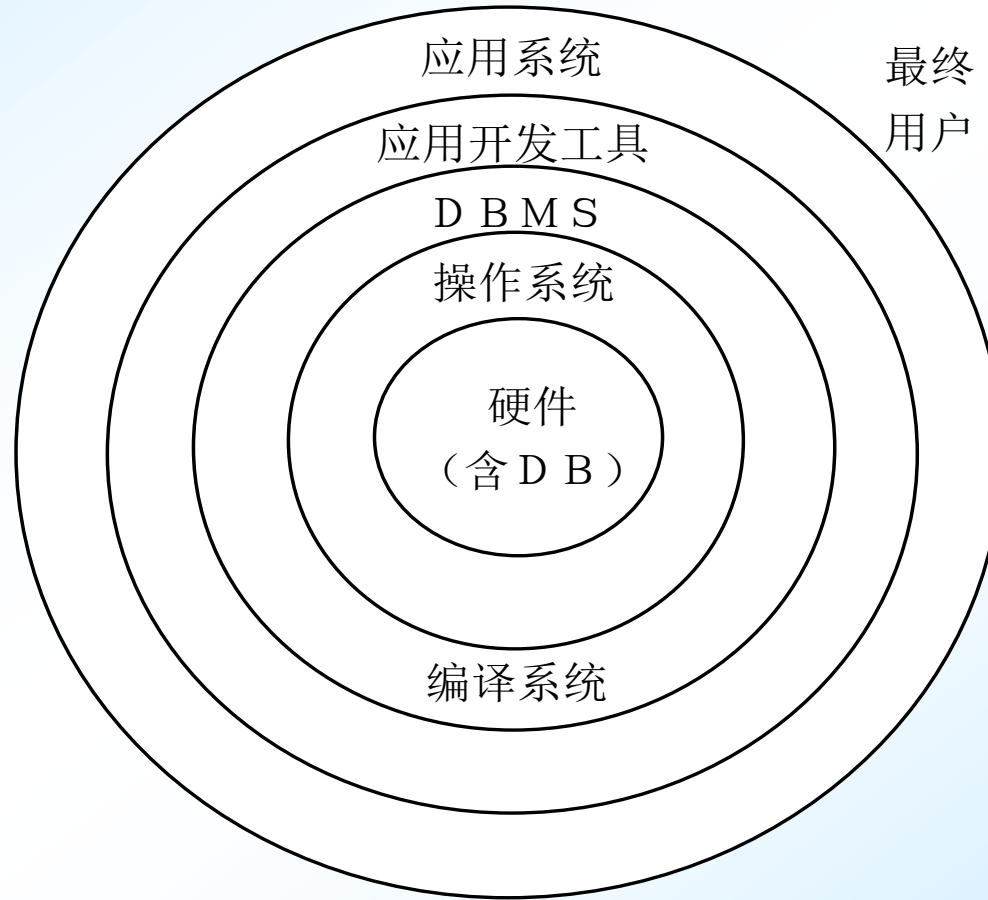


图 1.2 数据库在计算机系统中的地位



1.1.2 数据管理技术的产生和发展

谈数据管理技术，先要讲到数据处理，所谓数据处理是指对各种数据进行收集、存储、加工和传播的一系列活动的总和。数据管理则是数据处理的中心问题，为此，数据管理是指对数据进行分类、组织、编码、存储、检索和维护的管理活动总称。就用计算机来管理数据而言，数据管理是指数据在计算机内的一系列活动的总和。

数据管理技术已经历了人工管理、文件系统及数据库系统三个发展阶段。这三个阶段的特点及其比较如[表1.1](#)所示。

人工管理阶段

文件系统阶段

数据库系统阶段

表1.1 数据管理三个阶段的比较

		人工管理阶段	文件系统阶段	数据库系统阶段
背 景	应用背景	科学计算	科学计算、管理	大规模管理
	硬件背景	无直接存取存储设备	磁盘、磁鼓	大容量磁盘
	软件背景	没有操作系统	有文件系统	有数据库管理系统
	处理方式	批处理	联机实时处理、批处理	联机实时处理、分布处理、批处理
特 点	数据的管理者	用户(程序员)	文件系统	数据库管理系统
	数据面向的对象	某一应用程序	某一应用	现实世界
	数据的共享程度	无共享，冗余度极大	共享性差，冗余度大	共享性高，冗余度小
特 点	数据的独立性	不独立，完全依赖于程序	独立性差	具有高度的物理独立性和一定的逻辑独立性
	数据的结构化	无结构	记录内有结构、整体无结构	整体结构化，用数据模型描述
	数据控制能力	应用程序自己控制	应用程序自己控制	由数据库管理系统提供数据安全性、完整性、并发控制和恢复能力

人工管理阶段

20世纪50年代中期以前，计算机主要用于科学计算。硬件设施方面：外存只有纸带、卡片、磁带，没有磁盘等直接存取设备；软件方面：没有操作系统和管理数据的软件；数据处理方式是批处理。

人工管理数据具有以下几个特点：

- **数据不保存**

- ✓ **应用程序管理数据**

- ✓ **数据不共享**

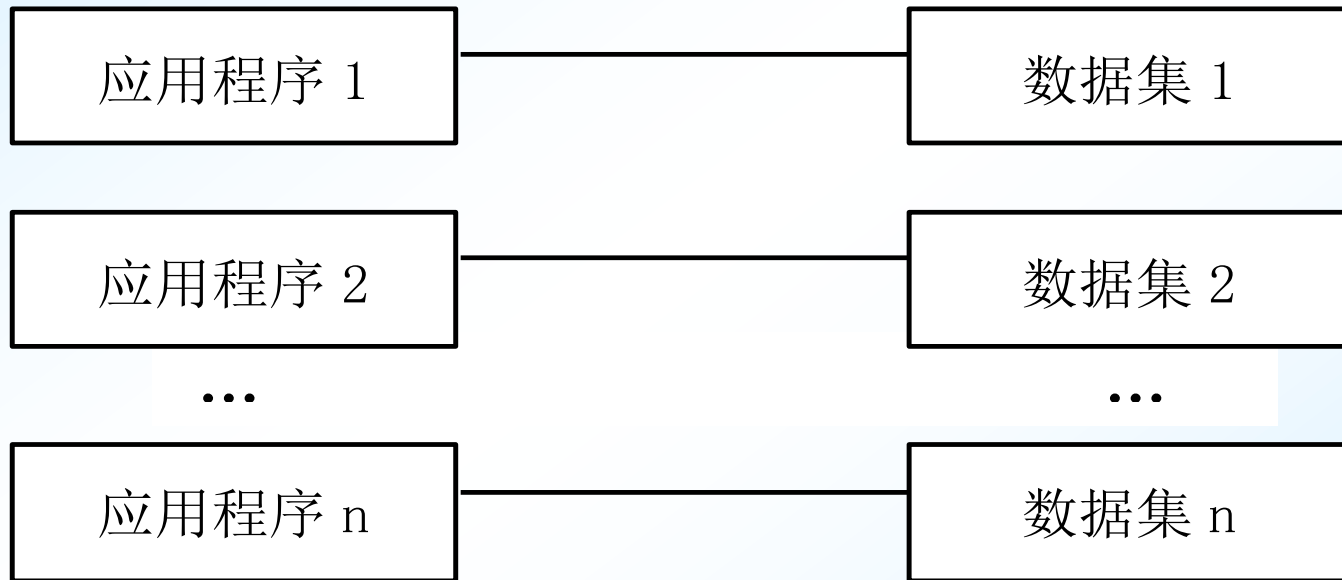
数据是面向应用的，一组数据只能对应一个程序。如果多个应用程序涉及某些相同的数据，则由于必须各自进行定义，无法进行数据的参照，因此程序间有大量的冗余数据。

数据不具有独立性

数据的独立性包括了数据的逻辑独立性和数据的物理独立性。当数据的逻辑结构或物理结构发生变化时，必须对应用程序做相应的修改。

在人工管理阶段，程序与数据之间的对应关系可用[图1.3](#)表示。

图1.3 人工管理阶段应用程序与数据之间的对应关系



文件系统阶段

20世纪50年代后期到60年代中期，这时计算机已大量用于数据的管理。硬件方面：有了磁盘、磁鼓等直接存取存储设备；软件方面：操作系统中已经有了专门的管理软件，一般称为文件系统；处理方式有批处理、联机实时处理。

特点如下：

数据长期保存

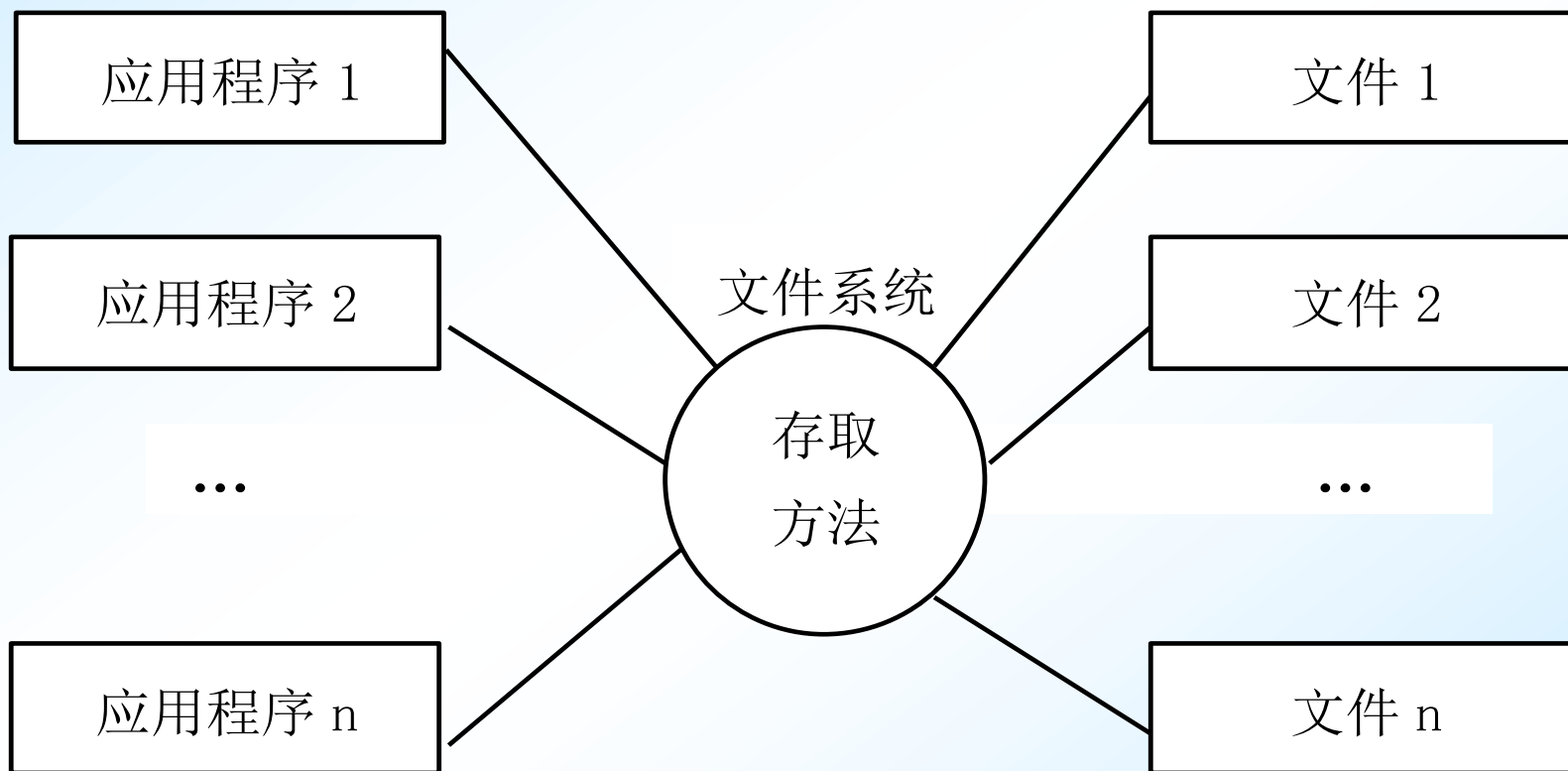
文件系统管理数据

数据共享性差，冗余度大

数据独立性差

在文件系统阶段，程序与数据之间的关系如[图1.4](#)所示。

图1.4 文件系统阶段应用程序与数据之间的对应关系

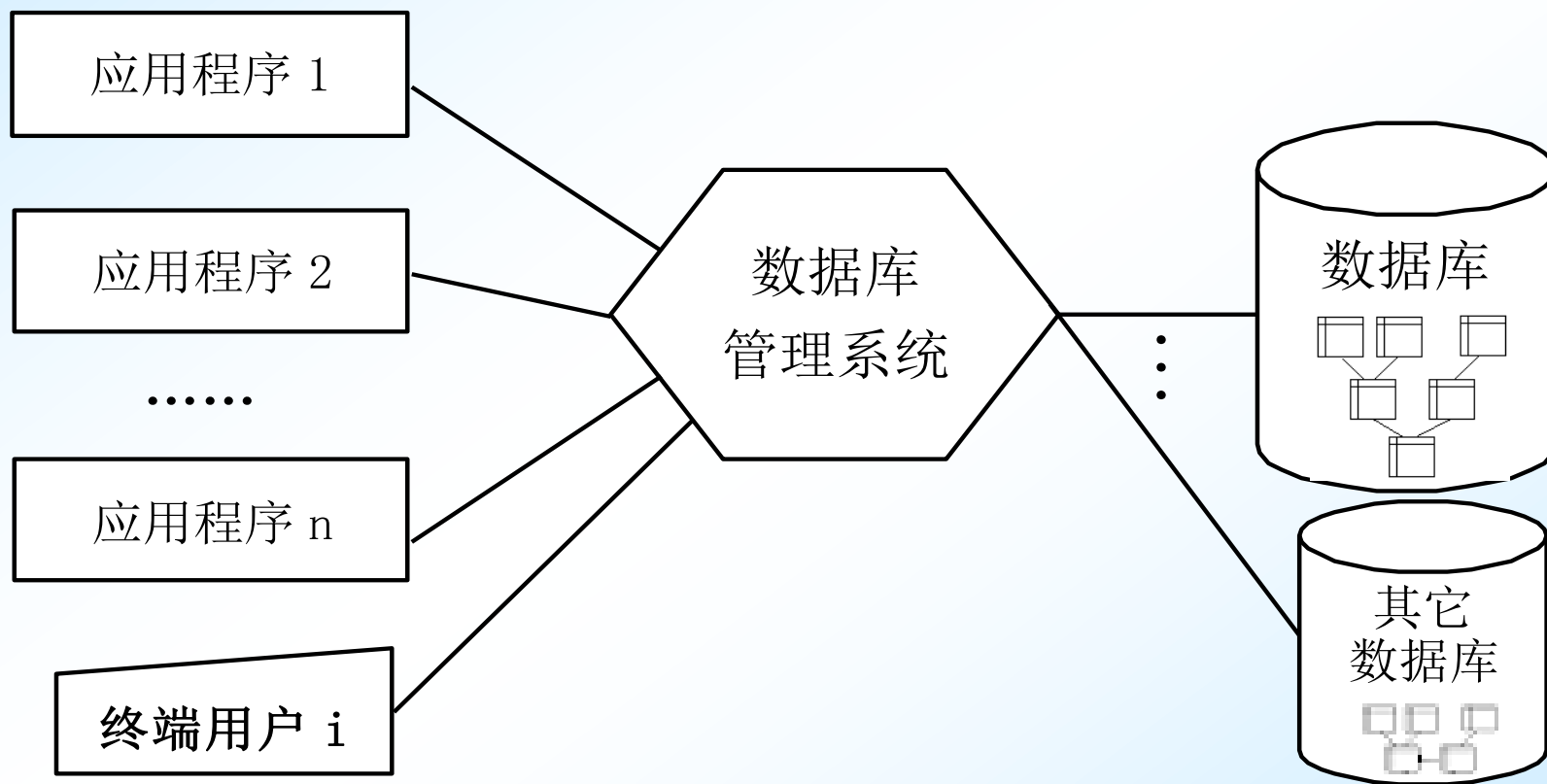


数据库系统阶段

20世纪60年代后期以来，计算机用于管理的规模更为庞大，数据量急剧增长，硬件已有大容量磁盘，硬件价格下降；软件则价格上升，使得编制、维护软件及应用程序成本相对增加；处理方式上，联机实时处理要求更多，分布处理也在考虑之中。介于这种情况，文件系统的数据库管理满足不了应用的需求，为解决共享数据的需求，随之从文件系统中分离出了专门软件系统——数据库管理系统，用来统一管理数据。

数据库系统阶段应用程序与数据之间的对应关系可用图1.5表示。

图1.5 数据库系统阶段应用程序与数据之间的对应关系



结论

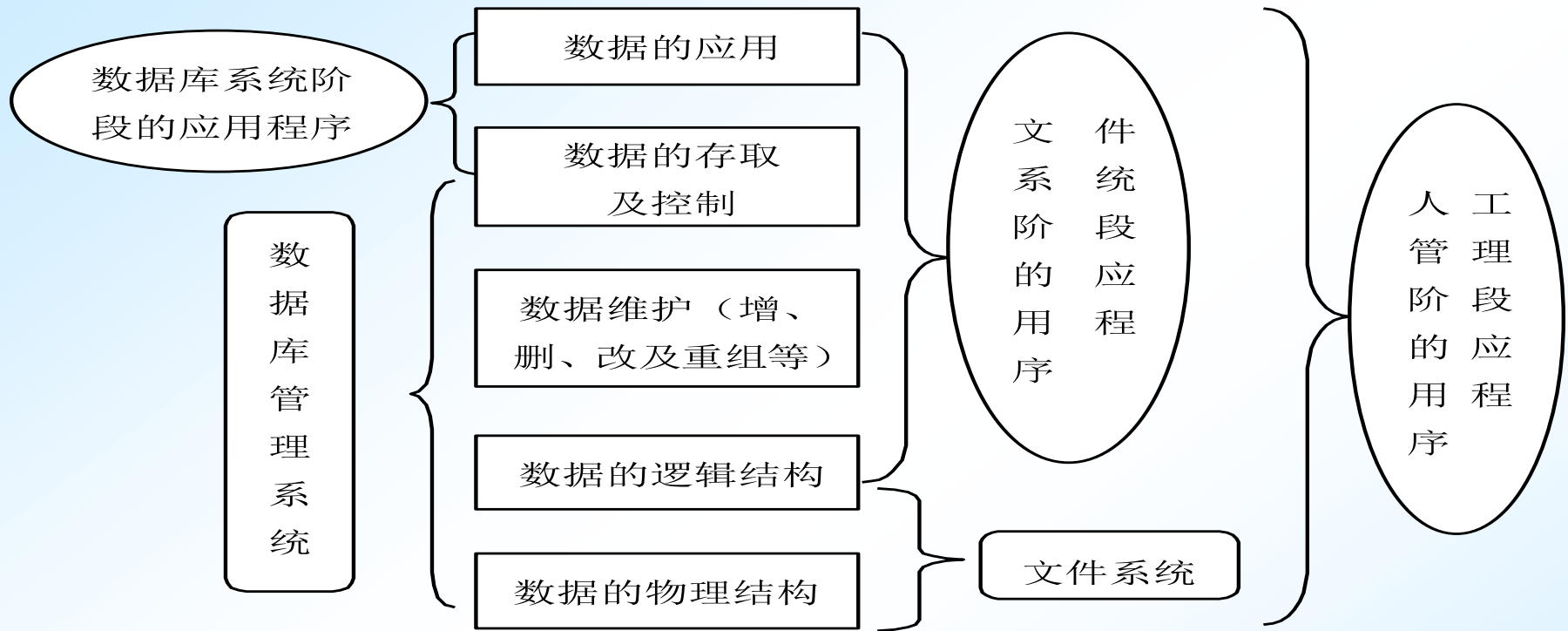


图1.6 三个阶段应用程序与数据管理的工作任务划分示意图

综上所述，如图1.6所示，随着数据管理技术的不断发展，应用程序不断从底层的、低级的、物理的数据管理工作中解脱出来，能独立的、较高逻辑级别地轻松处理数据库数据。从而能极大地提高了应用软件的生产力。目前关系数据库系统已逐渐淘汰了网状数据库和层次数据库，成为当今最流行的商用数据库系统。

1.1.3 数据库系统的特点

主要特点：

- 数据结构化
- 数据的共享性高，冗余度低，易扩充
- 数据独立性高
- 数据由DBMS统一管理和控制

结论

数据结构化

数据结构化是数据库系统与文件系统的根本区别。

在文件系统中，相互独立的文件的记录内部是有结构的。传统文件的最简单形式是等长同格式的记录集合。例如：一个教师人事记录文件。

- ◆ 每个记录都有如图1.7的记录格式：
- ◆ 图1.7 教师记录格式示例
- ◆ 图1.8 主记录——详细记录格式示例
- ◆ 图1.9 教师王名记录示例

数据结构化

教师人事记录

教师号	姓名	性别	年龄	政治面貌	籍贯	家庭出身	职称	所在系	家庭成员	奖惩情况
-----	----	----	----	------	----	------	----	-----	------	------

图1.7 教师记录格式示例

教师人事记录

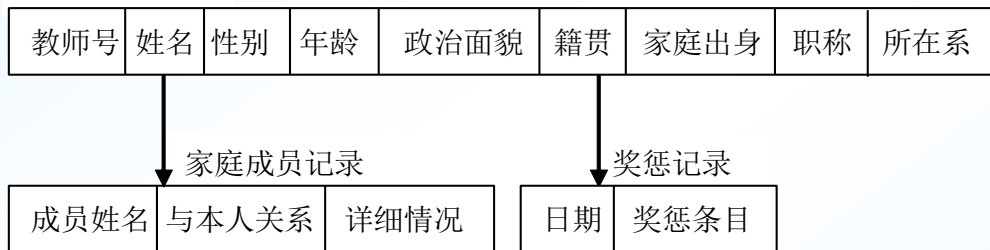


图1.8 主记录——详细记录格式示例

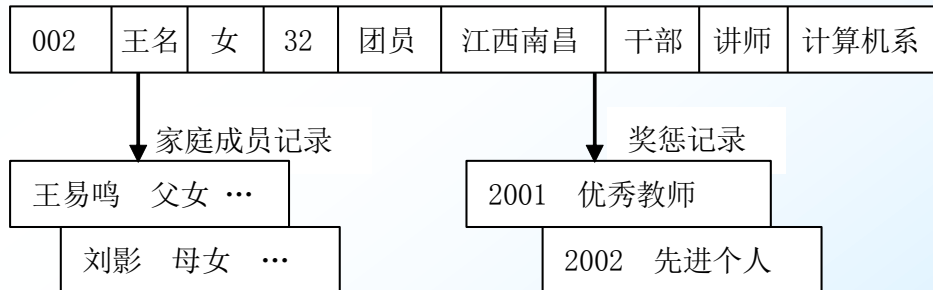


图1.9 教师王名记录示例

数据结构化

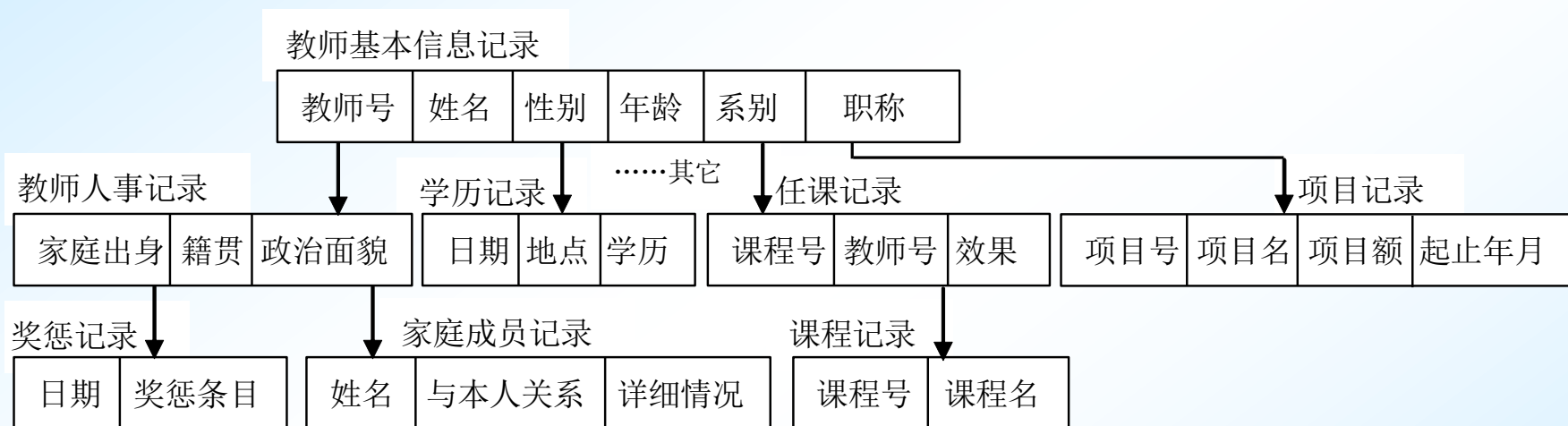


图1.10 教师数据组织

数据结构化

- 这样就可以节省许多存储空间，灵活性也相对提高。
- 但这样建立的文件还有局限性，因为这种结构上的灵活性只是针对一个应用而言。
- 而图1.10 教师数据组织，这种数据组织方式为各部分的管理提供了必要的记录，使数据结构化了。这就要求在描述数据时不仅要描述数据本身，还要描述数据之间的联系。
- 在文件系统中，尽管其记录内已经有了某些结构，但记录之间没有联系。

数据结构化

- 数据库系统实现整体数据的结构化，是数据库的主要特征之一，也是数据库系统与文件系统的本质区别。
- 在数据库系统中，数据不再针对某一应用，而是面向全组织，是整体结构化的。不仅数据是结构化的，而且存取数据的方式也是很灵活的，可以存取数据库中的某一个数据项（或字段）、一组数据项、一个记录或是一组记录。而在文件系统中，数据的最小单位是记录（一次一记录的读写），粒度不能细到数据项。

数据的共享性高，冗余度低，易扩充

- ✓ 数据库系统从整体角度看待和描述数据，数据不再面向某个应用而是面向整个系统，因此数据可以被多个用户、多个应用共享使用。数据共享可以大大的减少数据冗余，节约存储空间。数据共享还能够避免数据之间的不相容性与不一致性。
- ✓ 所谓数据的不一致性是指同一数据不同拷贝的值不一样。采用人工管理或文件系统管理时，由于数据被重复存储，当不同的应用使用和修改不同的拷贝时就容易造成数据的不一致。在数据库中数据共享，减少了由于数据冗余造成的不一致现象。

数据的共享性高，冗余度低，易扩充

- ✓ 由于数据面向整个系统，是有结构的数据，不仅可以被多个应用共享使用，而且容易增加新的应用，这就使得数据系统弹性大，易于扩充，可以适应各种用户的要求。可以取整体数据的各种子集用于不同的应用系统，当应用需求改变或增加时，只要重新选取不同的子集或加上一部分数据便可以满足新的需求。

数据独立性高

- ✓ 数据独立性包括了数据的物理独立性和数据的逻辑独立性两方面。
- ✓ 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数据是相互独立的。也就是说，数据在磁盘上的数据库中怎样存储是由DBMS管理的，用户程序不需要了解，应用程序要处理的只是数据的逻辑结构，这样当数据的物理存储改变时，应用程序不用改变。
- ✓ 逻辑独立性是指用户的应用程序与数据库的整体逻辑结构是相互独立的，也就是说，数据的逻辑结构改变了，用户程序也可以不需修改的。

数据独立性高

- ✓ 数据独立性是由DBMS的三级模式结构与二级映象功能来保证的，将在后面介绍。
- ✓ 数据与程序的独立，把数据的定义从程序中分离出去，加上数据的存取又由DBMS负责，从而简化了应用程序的编制，大大减少了应用程序的维护和修改。

数据由DBMS统一管理和控制

- 数据库是长期在计算机内有组织的大量的可共享的数据集合,数据库的共享是并发的共享,即多个用户可以同时存取数据库中的数据甚至可以同时存取数据库中的同一块数据。
- DBMS提供以下几方面的数据控制功能:
 - (1) 数据的安全性控制
 - (2) 数据的完整性约束
 - (3) 并发控制
 - (4) 数据库恢复

结 论

- 目前，数据库已经成为现代信息系统的不可分离的重要组成部分。
- 数据库系统的出现使信息系统从以加工数据的程序为中心转向围绕共享的数据库为中心的新阶段。这样既便于数据的集中管理，又有利于应用程序的研制和维护，提高了数据的利用率和相容性，提高了决策的可靠性。
- 数据库技术是计算机领域中发展最快的技术之一，数据库技术的发展是沿着数据模型的主线展开的。

1.2 数据模型

模型是现实世界特征的模拟和抽象。数据模型也是一种模型，它能实现对现实世界数据特征的抽象。

数据模型应满足三方面的要求：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。

根据模型应用的不同目的，分为两类。

第一类模型是概念模型，也称信息模型。一般应具有以下能力：

- (1) 具有对现实世界的抽象与表达能力
- (2) 完整、精确的语义表达力
- (3) 易于理解和修改。
- (4) 易于向DBMS所支持的数据模型转换

1.2 数据模型

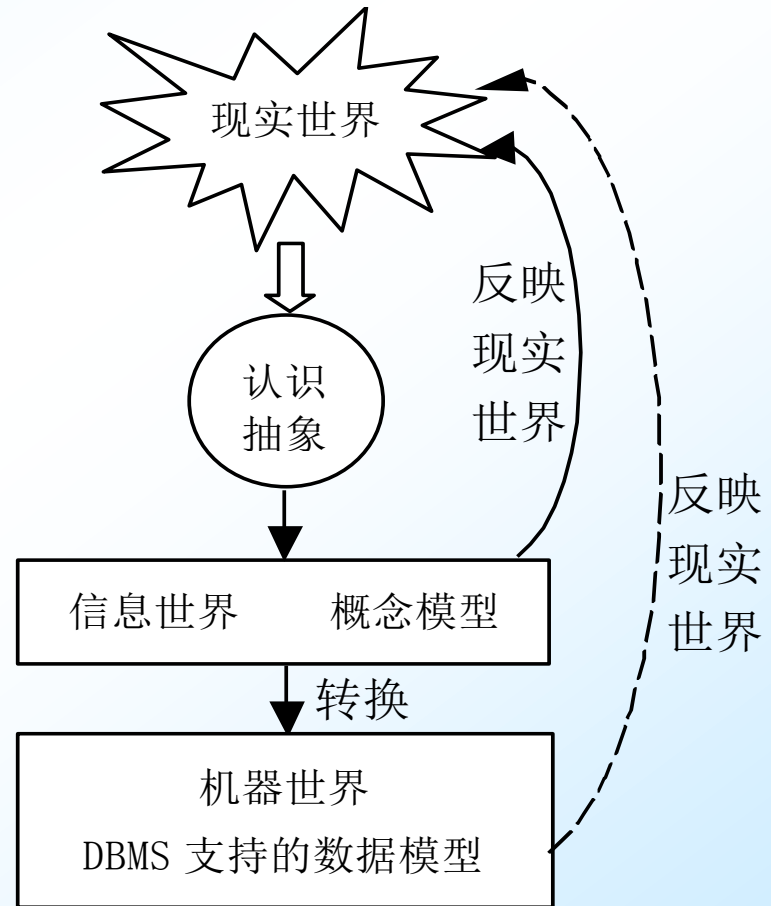
另一类模型是数据模型，包括层次模型、网状模型、关系模型等，它是按计算机系统对数据建模，主要用于DBMS的实现。

数据模型是数据库系统的核心和基础，各种机器上实现的DBMS软件都是基于某种数据模型的。

1.2 数据模型

- 为了把现实世界中的具体事物抽象、组织为某一DBMS支持的数据模型，人们常常首先将现实世界抽象为信息世界，然后将信息世界转换（或数据化）为机器世界。这一过程如图1.11所示。

- 图1.11 现实世界中客观对象的抽象过程 =====>



1.2.1 数据模型的组成要素

- 数据模型描述了系统的三个方面：静态特性、动态特性和完整性约束条件。一般由数据结构、数据操作和完整性约束三部分组成，是严格定义的一组概念的集合。

1.2.1 数据模型的组成要素

一、数据结构

数据结构用于描述系统的静态特性，是所研究的对象类型的集合。

二、数据操作

数据操作用于描述系统的动态特性，是指对数据库中各种对象及对象的实例允许执行的操作的集合。

三、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。数据模型应该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。数据模型还应该提供自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。

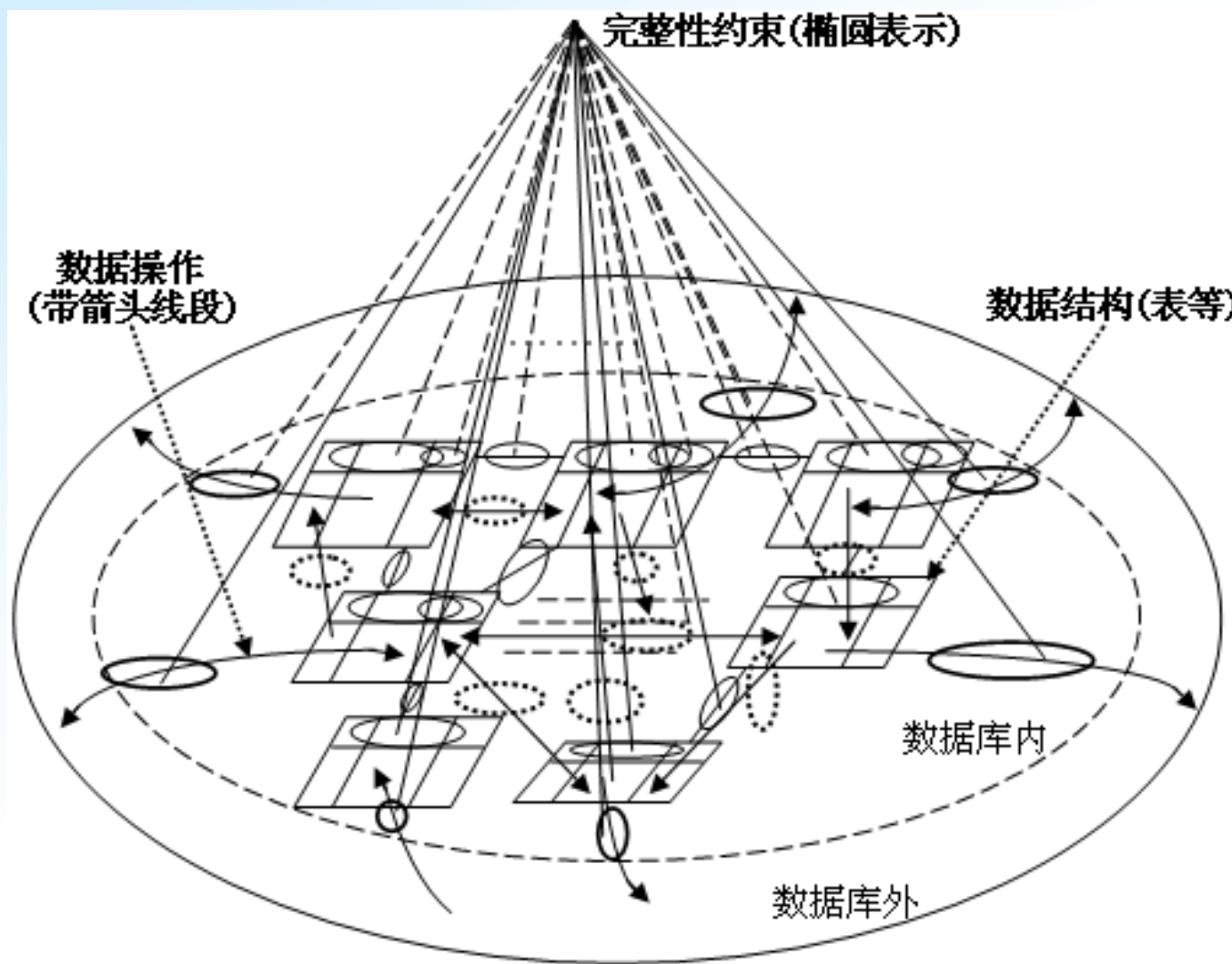


图1.12 数据模型的三要素示意图

- 要说明的是图1.12是简单化、逻辑示意的图，数据模型的三要素在数据库中都是严格定义的一组概念的集合。在关系数据库可以简单理解为：数据结构是表结构定义及其它数据库对象定义的命令集；数据操作是数据库管理系统提供的数据库操作（操作命令、语法规则、参数说明等）命令集；数据完整性约束是各关系表约束的定义及操作约束规则等的集合。

1.2.2 概念模型

- 概念模型是现实世界到机器世界的一个中间层次。
- 信息世界中的基本概念
- 概念模型的表示
- 层次模型概述

信息世界中的基本概念

主要有：

- 实体：实体是指客观存在并可以相互区别的事物。
- 属性：属性是指实体所具有的某一特性。
- 码：码是指唯一标识实体的属性集。
- 域：域是指属性的取值范围，具有相同的数据类型的数据集合。
- 实体型：具有相同属性的实体必然具有共同的特征和性质。
- 实体集：实体集是指同型实体的集合。
- 联系：在现实世界中，事物内部以及事物之间是有关联的。

两个实体型之间的联系有以下三种：

一对一联系；一对多联系；多对多联系。

- 两个实体型之间的三类联系可以用如[图1.13](#)、[1.14](#)所示来表示。

图1.13 两个实体型之间的三类联系示意图

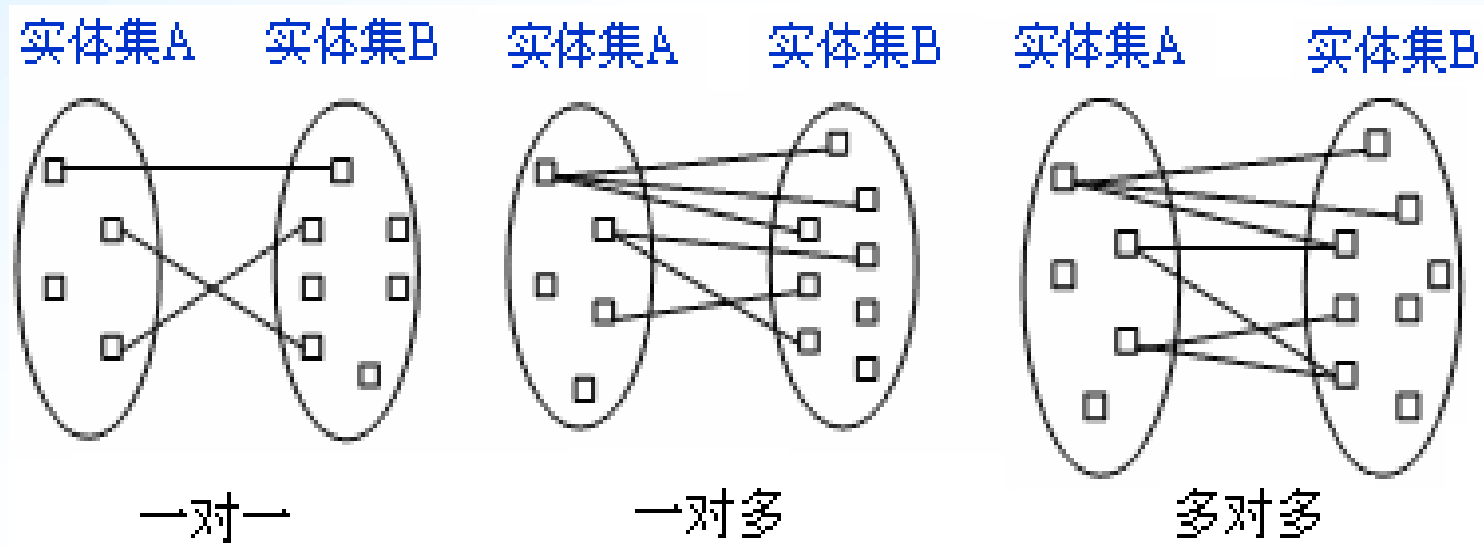
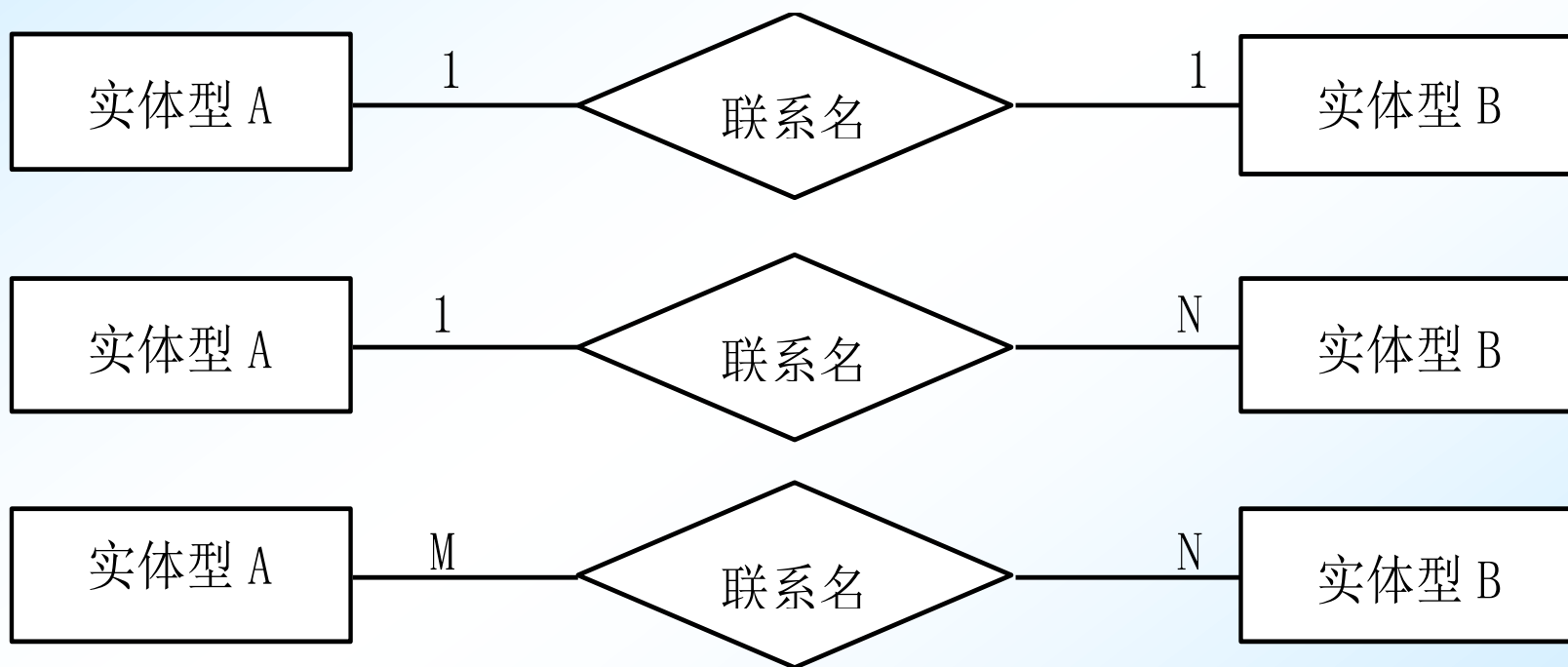


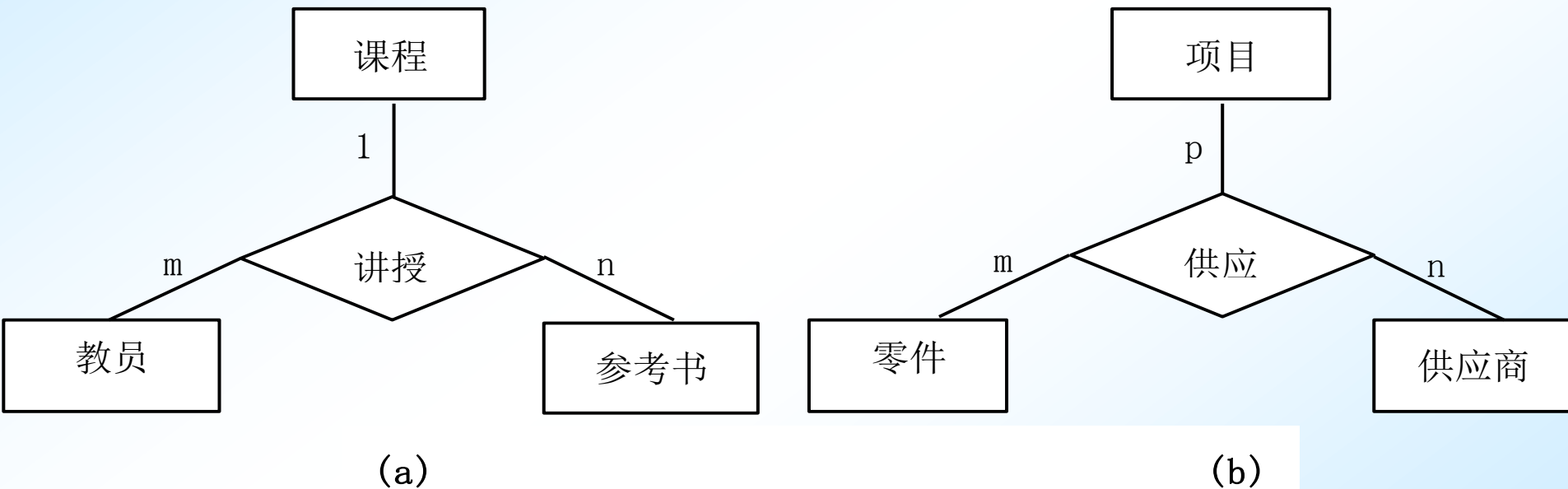
图1.14 两个实体型之间的三类联系



信息世界中的基本概念

- 单个或多个实体型之间也有类似于两个实体型之间的三种联系类型。
- 例如，对于教师、课程与参考书三个实体型，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、参考书三者间的联系是一对多的，如图1.15 (a)所示。
- 要注意的是三个实体型之间多对多联系与三个实体型两两之间的多对多联系（共有三个）的语义及E-R图是不同的。

图1.15 两个实体型之间的三类联系



信息世界中的基本概念


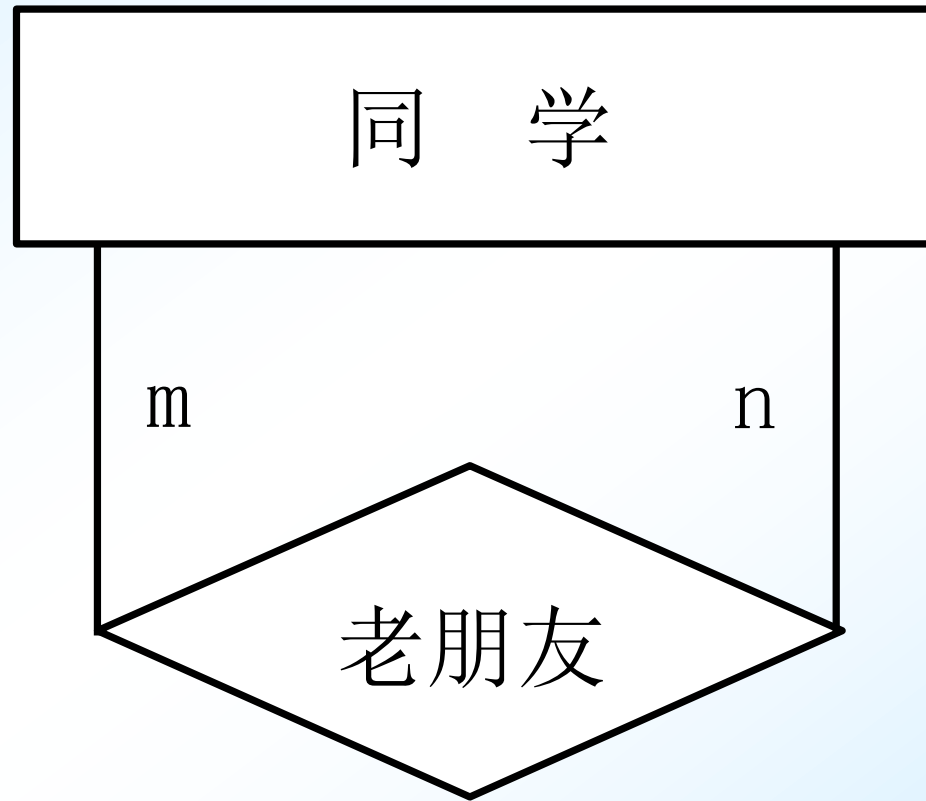
- 同一个实体型对应的实体集内的各实体之间也可以存在一对一、一对多、多对多的联系的（可以把一个实体集逻辑上看成两个与原来一样的实体集来理解）。例如：同学实体集内部同学与同学之间老朋友的关系可能是多对多的（如图1.16所示），这是因为每位同学的老朋友往往有多位。

图1.16 一个实体型实体之间的多对多联系

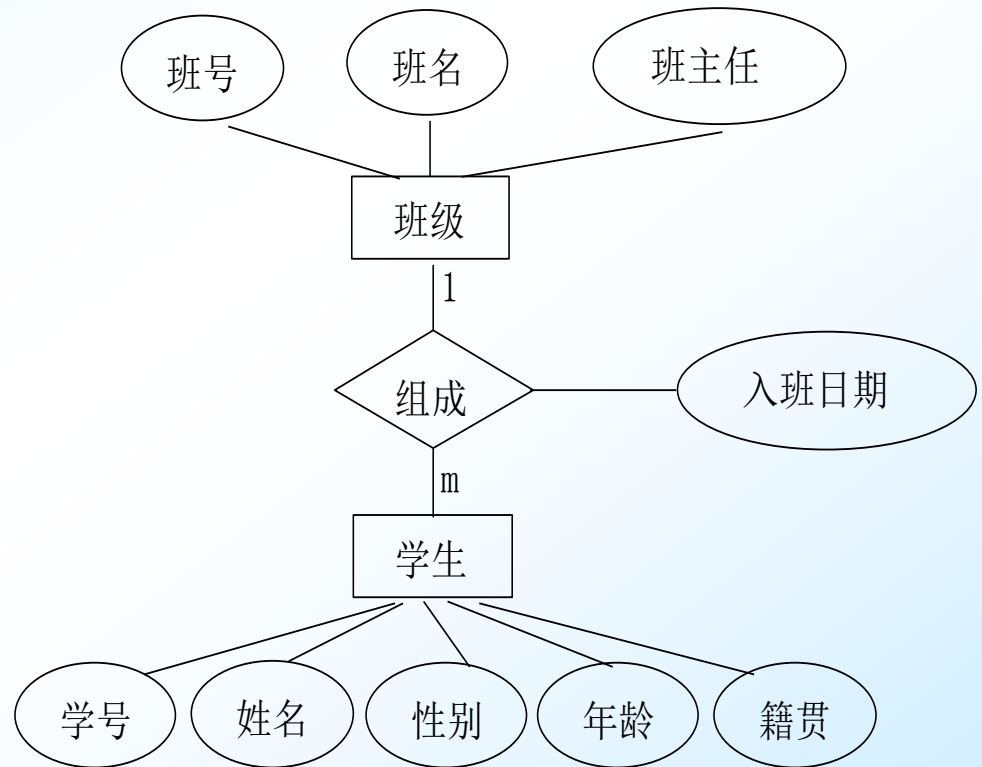


概念模型的表示

- 最常用的是实体—联系方法。该方法用E-R图来描述现实世界的概念模型。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。
- 属性：用椭圆表示，椭圆形内写明属性名。并用无向边将其与相应的实体连接起来。
- 联系：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）。

概念模型的表示

- 如右图1.17所示就是一个班级、学生的概念模型（用E-R图表示），班级实体型与学生实体型之间很显然是一对多关系。应用时应针对某实际情况，试着设计反映实际内容的实体及实体联系的E-R图。



E-R模型的变换

- E-R模型在数据库概念设计过程中根据需要可进行变换，包括实体类型、联系类型和属性的分裂、合并和增删等，以满足概念模型的设计、优化等的需要。

E-R模型的变换

- 垂直分裂，如图1.18所示。

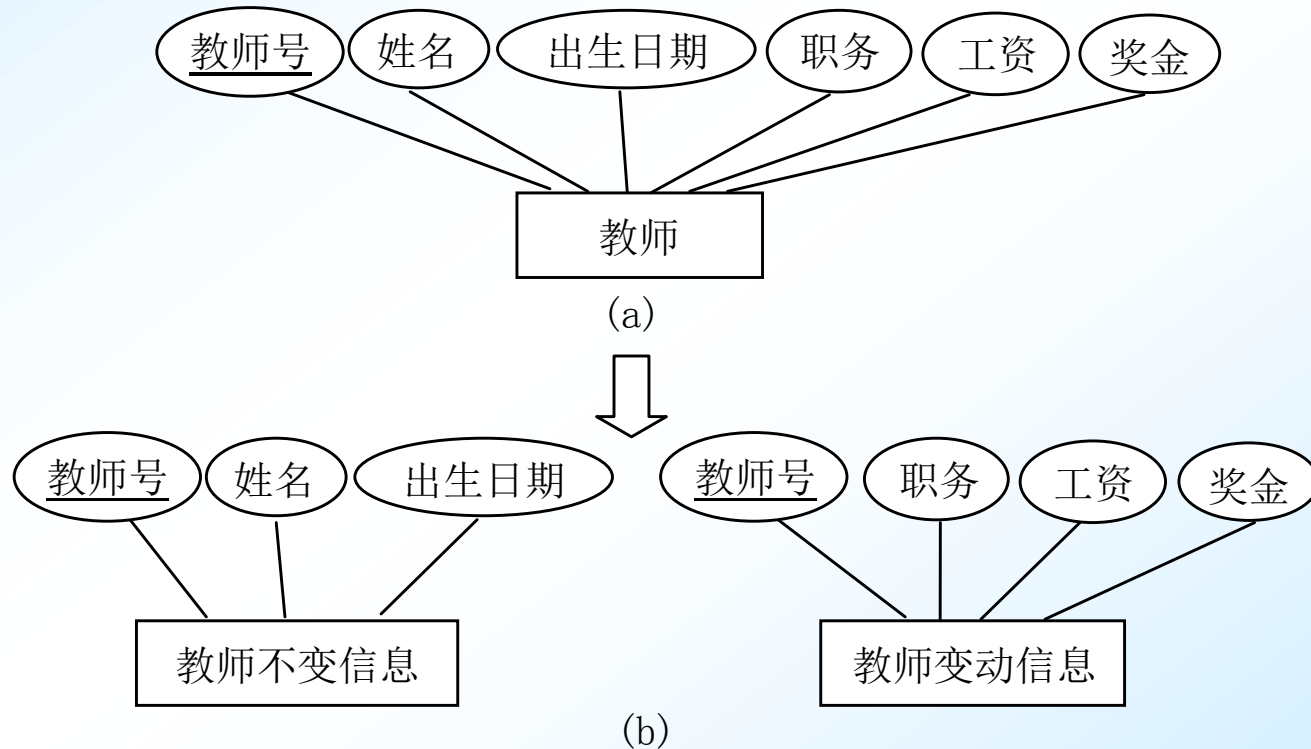


图1.18 实体类型的垂直分裂

E-R模型的变换

- 联系类型也可分裂。如教师与课程间的“担任”教学任务的联系，可分裂为“主讲”和“辅导”两个新的联系类型，如图1.19所示

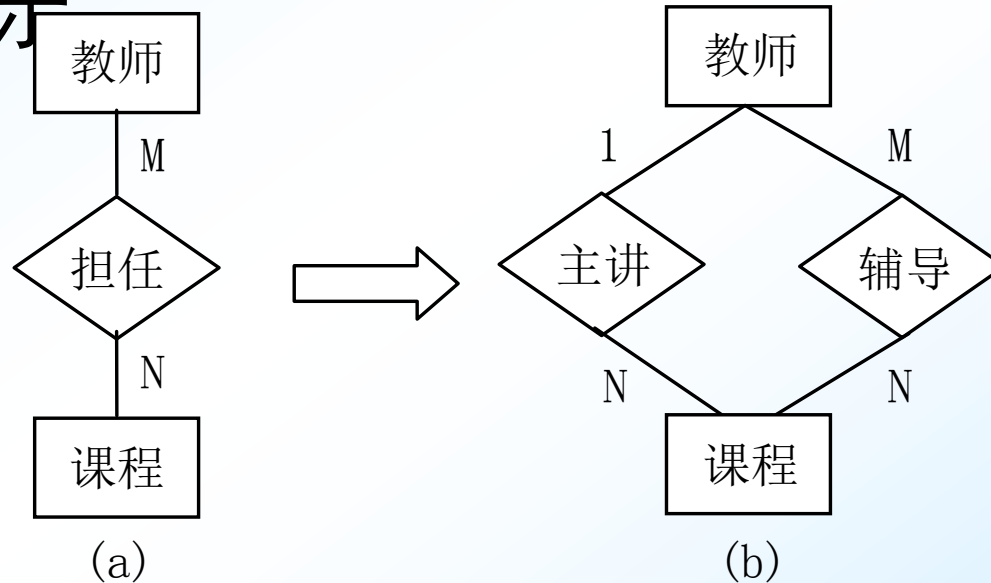


图1.19 联系类型的分裂

1.2.3 基本E-R模型的扩展*

- 基本ER（实体-联系）模型是对现实世界的一种抽象，它的主要成份是实体、联系和属性，以实体、属性和联系三个抽象概念为基础的E-R数据模型是基本E-R数据模型。
- 但是在现实世界中还有一些特殊的语义，需要扩展E-R模型（Extended E-R data model，简称ERR data model）的概念才能更好地模拟与抽象现实世界。

1.2.3 基本E-R模型的扩展*

- 1、属性的分类
- 2、联系的再理解
- 3、实体的角色
- 4、依赖联系与弱实体
- 5、特殊化和普遍化
- 6、聚集
- 7、范畴

1、属性的分类

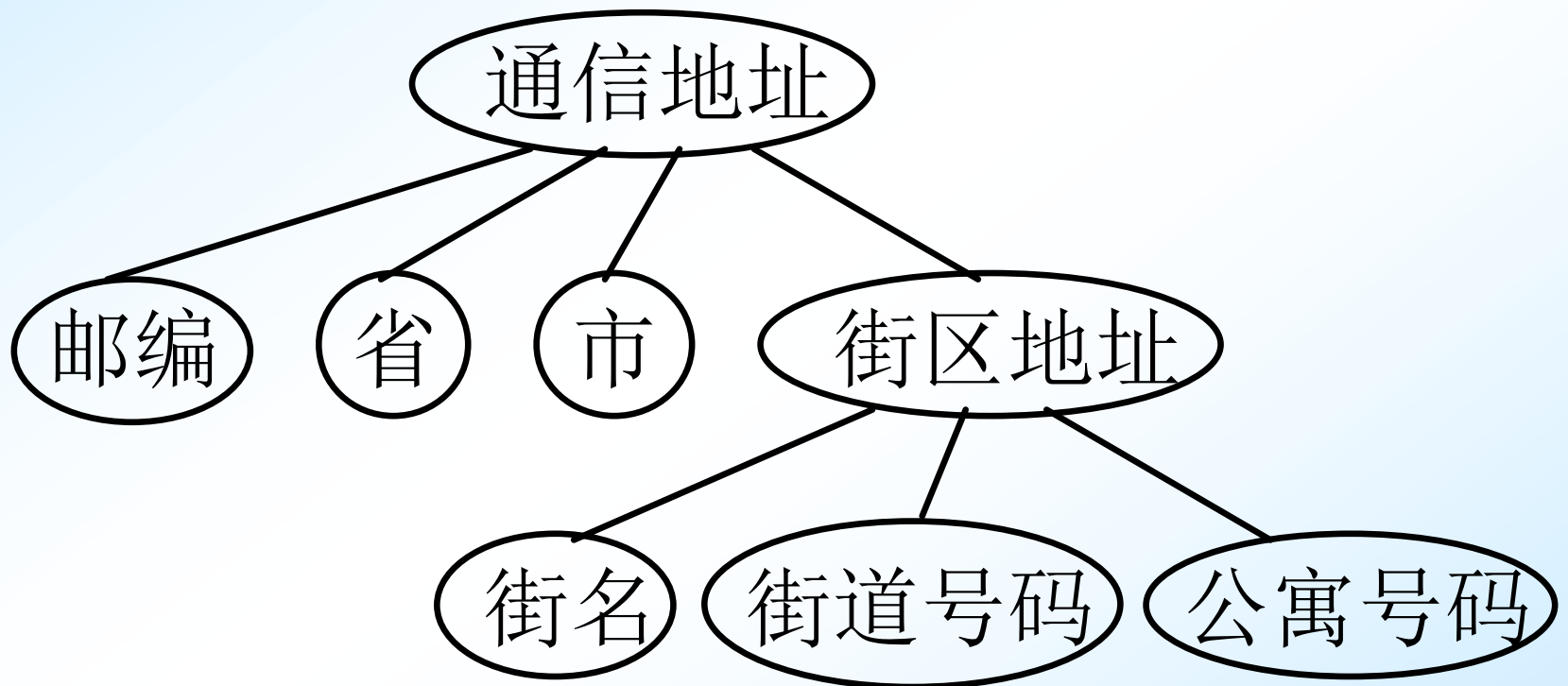
扩展E-R图中的属性，有这几种分类：简单属性与复合属性、单值属性与多值属性、NULL属性、派生属性与基本属性等。

- ✓ 简单属性与复合属性
- ✓ 单值属性与多值属性
- ✓ NULL属性
- ✓ 派生属性与基本属性

1、属性的分类——简单属性与复合属性

在E-R模型中，属性的取值范围称为域，属性可以是单域的简单属性，即不可再分的属性，如学号、年龄、性别；也可以是多域的组合属性。复合属性由简单属性和其他复合属性组成。复合属性中允许包含其他复合属性，这意味着属性可以是一个层次结构。复合属性可用一个树形结构表示，图1.20表示通信地址这一属性的层次结构。

图1.20 通信地址属性



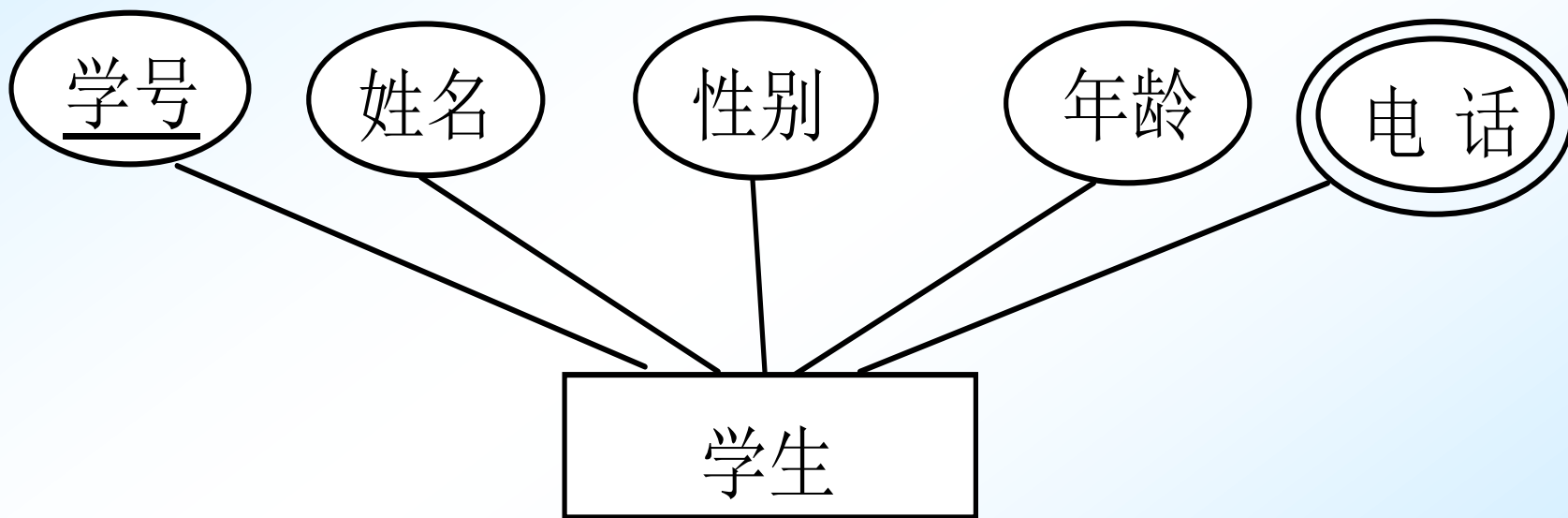
1、属性的分类——单值属性与多值属性

在E-R模型中，属性的取值可以是单值的（single-valued），是指同一实体的属性只能取一个值；也可以是多值的（multi-valued），是指同一实体的某个属性可能取多个值。

例如某个学生的学号、姓名、性别、年龄等属性上的取值是唯一的，它们是单值属性；而学生的“联系电话”应该是多值属性，因为如今一位学生往往有多个联系电话的。

多值属性用双椭圆形表示。如图1.21为学生实体E-R图。

图1.21 含多值属性的学生实体E-R图



1、属性的分类——NULL属性

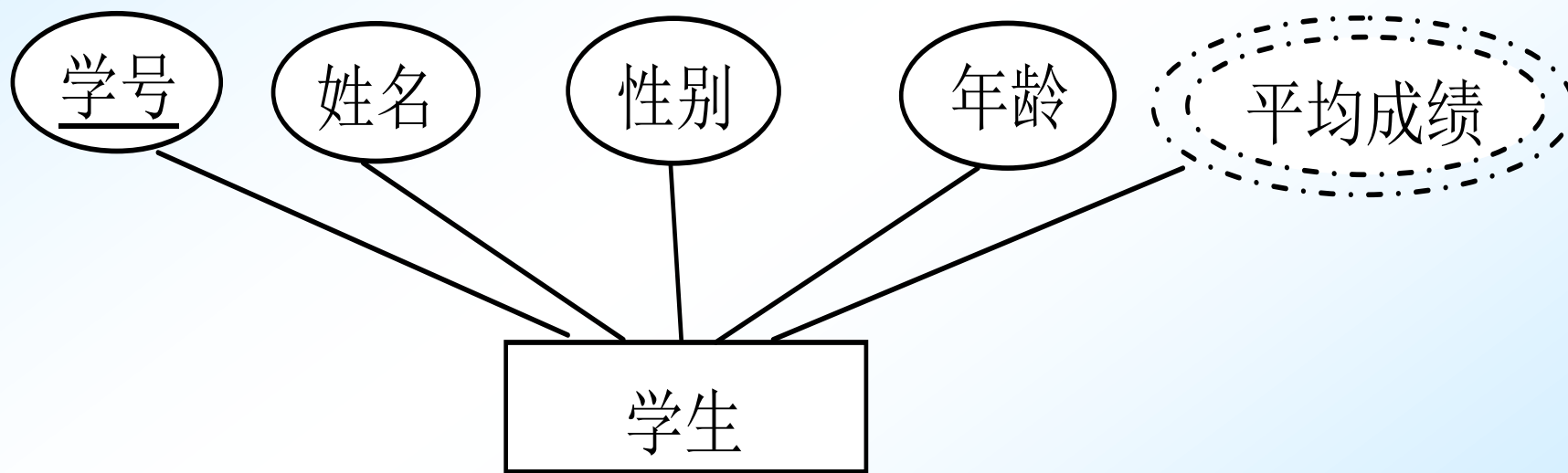
当某个属性对某个实体不适用、属性值未知或属性值暂时未定时，也可用空缺符Null来表示。Null表示“无意义”或“不确定”。如通讯录（姓名、email、电话、BP），若某人没有email地址，则在email属性上取值为Null。要注意的是作为码的属性，它的取值不能为Null的。

1、属性的分类——派生属性与基本属性

可以从其他相关的属性或实体派生出来的属性称为派生属性，相反不能由其它属性或实体派生出来的属性为基本属性或存储属性。

如学生（学号、姓名、平均成绩），选课（学号、课程号、成绩），则平均成绩可由学生所选课程的总成绩除以课程总数来计算得到，则称“平均成绩”为派生属性，而选课实体型中的“成绩”为基本属性。

图1.22 含派生属性的学生实体E-R图



2、联系的再理解

- 实体型间的联系，具体来说多个实体间的相互关联。联系集是同类联系的集合。
- 定义 1-x 联系集是 $n(n \geq 1)$ 个实体集上的数学关系，这些实体集不必互异。如果 E_1, E_2, \dots, E_n 为 n 个实体集，那么联系集 R 是
- $\{(e_1, e_2, \dots, e_n) \mid e_1 \in E_1, e_2 \in E_2, \dots, e_n \in E_n\}$ 的一个子集，而 (e_1, e_2, \dots, e_n) 是一个联系。
- 关于联系的内容有联系的元数、联系的类型和联系的实体参与度等三方面。

2、联系的再理解——联系的元数

- 定义 一个联系涉及到的实体集个数，称为该联系的元数或度数（Degree）。通常，同一个实体集内部实体之间的联系，称为一元联系，也称为递归联系；两个不同实体集实体之间的联系，称为二元联系；三个不同实体集实体之间的联系，称为三元联系。以此类推。

2、联系的再理解——联系的类型

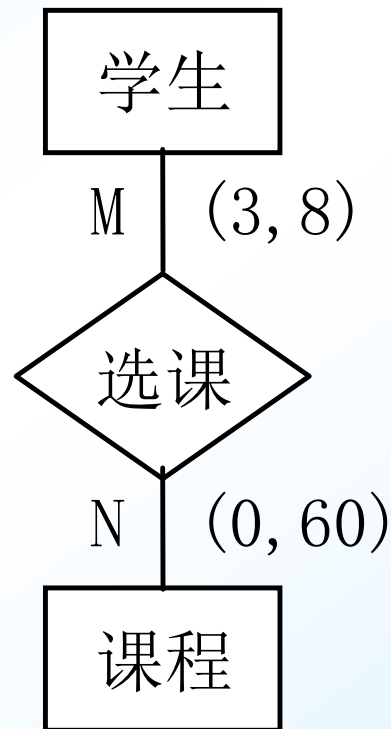
- 定义 联系涉及到的实体集之间实体对应的方式，称为联系的类型。这里“实体的对应方式”是指实体集E1中一个实体与实体集E2中至多有一个还是多个实体有联系，同时相反看实体集E2中一个实体与实体集E1中至多有一个还是多个实体有联系（注意E1与E2可以是指相同的实体集）。
- 二元联系的类型有：一对一（1：1），一对多（1：N）或多对一（N：1），多对多（M：N）等三种。在E-R数据模型中，这种联系类型还可推广到多元联系。如三元联系可分为：M：N：P，1：N：P，1：1：P，1：1：1等种类。这种实体间联系的约束也称为基数比约束（cardinality ratio constraint）。

2、联系的再理解——联系的实体参与度

- 定义 设二元联系有两个实体集E1和E2，E1中每个实体与E2中有联系实体的数目的最小值min和最大值max，称为E1实体的参与度，用（min,max）形式表示。
- 例如：学校里规定每学期学生至少选修3门课程，最多选修8门课程；每门课程至多有60人选修，最少可以没人选修。这样，学生实体的参与度是（3，8），课程实体的参与度是（0，60），如图1.23所示。

2、联系的再理解——联系的实体参与度

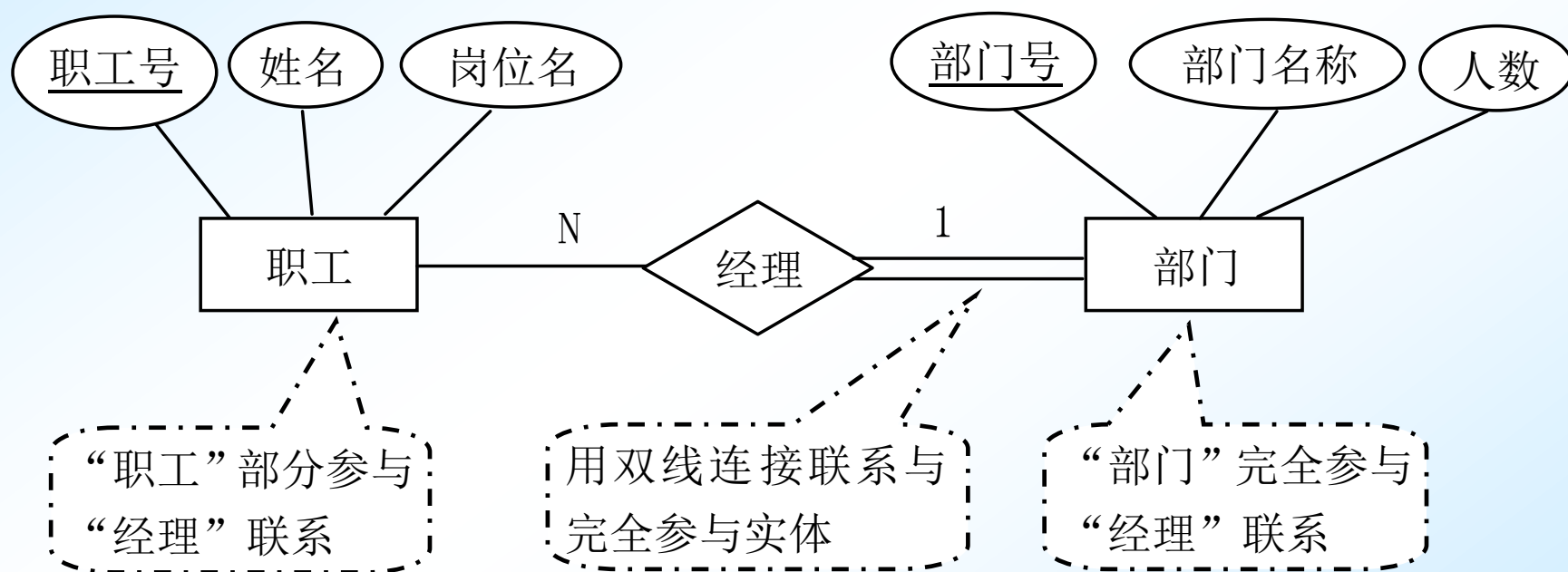
图1.23 联系的类型及实体的参与度



2、联系的再理解——联系的实体参与度

- 在E-R图表示时，如果无向边为双线，则表示该实体完全参与；如果无向边画为单线，则表示该实体部分参与。如图1.24所示。
- 实体参与联系的方式也是重要的语义约束，称为参与约束（participation constraint）。
- 基数比约束和参与约束构成联系的语义约束，有时合称为结构约束（structural constraint）。这些概念很容易推广到多元联系的（具体略），这样对实体参与联系的程度有定性与定量的概念了。

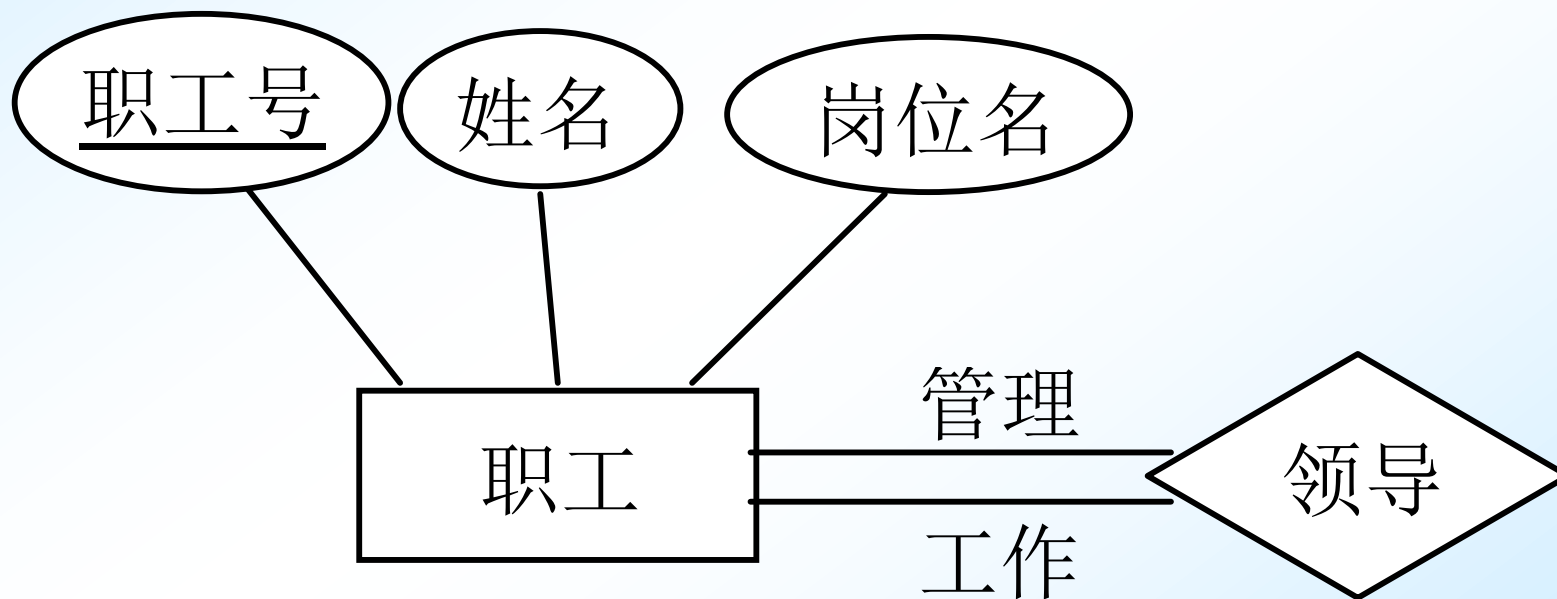
图1.24 部分参与与完全参与的表示



3、实体的角色

实体在联系中的作用称为实体的角色。当同一个实体集不止一次参与一个联系集时，为区分各实体的参与联系的方式，需要显式指明其角色。如学生与学生间的班长关系，职工与职工之间的领导关系(如图1.25)，课程之间的先修关系等。当需要显式区分角色时，在连接菱形和矩形的线上加上说明性标注以区别不同的角色。

图1.25 “职工” 实体在 “领导” 联系中的不同角色



4、依赖联系与弱实体

- (1) 依赖联系

在现实世界中，有些实体对另一些实体有很强的依赖关系，即一个实体的存在必须以另一实体的存在为前提。比如，一个职工可能有多个社会关系，社会关系是多值属性，为了消除冗余，设计两个实体：职工与社会关系。在职工与社会关系中，社会关系的信息以职工信息的存在为前提，因此社会关系的存在以职工的存在为前提。这说明了职工与社会关系是一种依赖联系。

4、依赖联系与弱实体

- (1) 依赖联系

如果实体A的存在依赖于实体B的存在，则称A依赖联系于B，B称作支配实体，A称作从属实体。如果B被删除，则A也要被删除。例如考虑分期付款的例子，对每一个“贷款”实体，有若干个“还款”实体与之关联，“还款”实体存在依赖于“贷款”实体。

完全参与与依赖联系的关系：设A与B间存在联系R，若A的存在依赖于B，则A必然是完全参与联系R。

4、依赖联系与弱实体

- (2) 弱实体

一个实体对于另一些实体具有很强的依赖联系，而且该实体的主码部分或全部从其父实体中获得，称该实体为弱实体（即从属实体），另一实体称为强实体（即支配实体）。如在人事管理系统中，职工子女的信息就是以职工的存在为前提的，子女实体是弱实体，职工实体是强实体，子女与职工的联系是一种强依赖联系。

4、依赖联系与弱实体

- 在E-R图中用双线框的矩形表示弱实体，与弱实体联系的联系用双线菱形框表示，从联系集用双线（全部参与）连接弱实体集，用箭头（一对多联系）指向强实体集。弱实体集的分辨符用下划虚线标明。如图1.26所示。

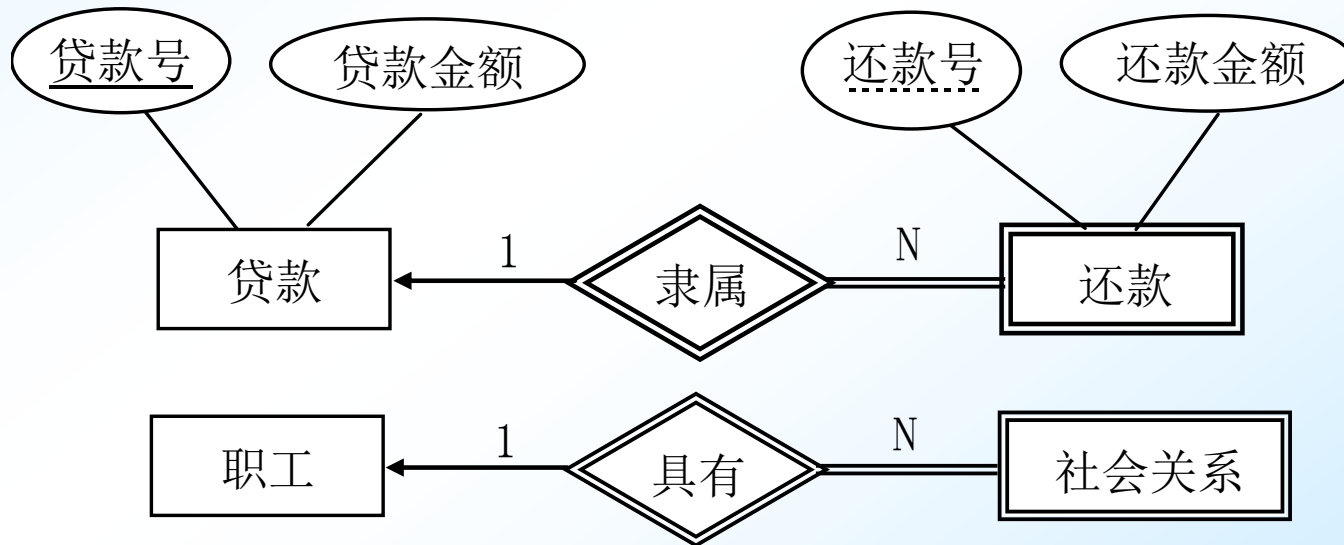


图1.26 弱实体的表示方法

5、特殊化和普遍化

- 我们知道一个实体集是具有某些共性的实体的集合。这些实体一方面具有共性，另一方面还具有各自的特殊性。往往一个实体集可以按照某个特征区分为几个实体集，例如学生这个实体集可以分为研究生、本科生、大专生等子集。如果需要的话，还可以把研究生这个实体集再分为博士生、硕士生等子集。这个从普遍到特殊的过程，叫做特殊化（specialization）。与特殊化相反的过程则叫普遍化，即把几个具有某些共性的实体集概括成一个更普遍的实体集。例如，把研究生、本科生、大专生三个实体集概括为学生实体集，还可以把学生、教师、职工这些实体集概括为“人”这个实体集。
- 从一般到特殊，从特殊到一般，本来就是人们认识世界常用的方法。因而在E-R数据模型中引入特殊化和普遍化这两个概念，对模拟现实世界是有用的。

5、特殊化和普遍化

- 超实体集与子实体集间的多层继承关系，能形成实体集间的层次结构(Hierarchy)关系，子实体集一般只继承一个超实体集。若子实体集参与继承到多个超实体集，则形成所谓的格结构(Lattice)。如在职进修生子实体集同时继承超实体集教职工与超实体集学生的情况所示。因此，在数据模型中，引入子实体集，可以很方便地描述实体集中部分实体的特殊属性和联系。

5、特殊化和普遍化

- 特殊化和普遍化可用扩展的E-R图表示，图1.27是一个例子。

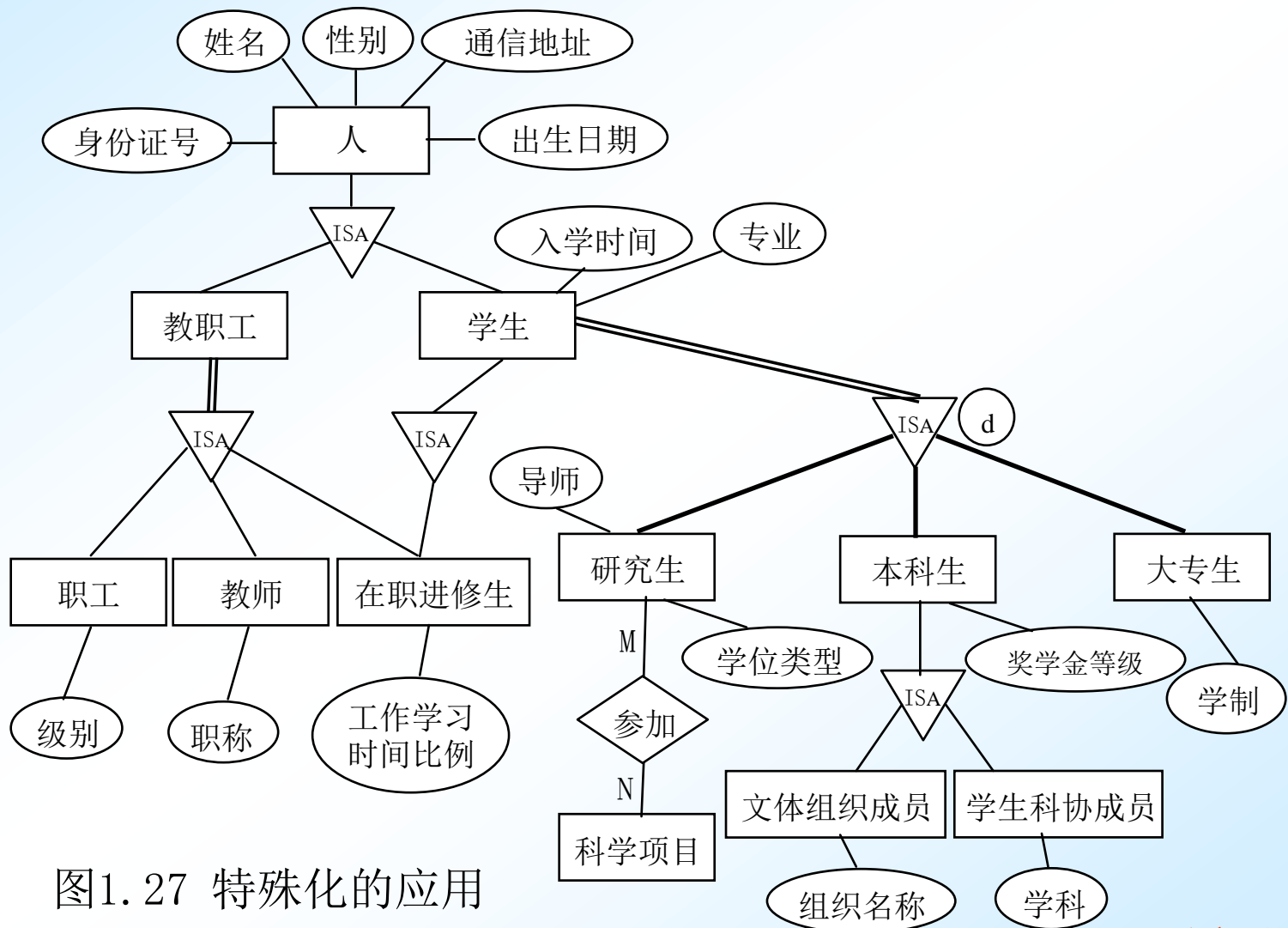


图1.27 特殊化的应用

6、聚集

- 联系之间存在重叠，如何表达联系之间的联系？在基本E-R数据模型中，只有实体才能参与联系，不允许联系参与联系。在扩展E-R数据模型中，可以把联系看成由参与组合而成的新的实体，其属性为参与联系的实体的属性和联系的属性的并。这种新的实体称为参与联系的实体的聚集（aggregation）。有了聚集这个抽象概念，联系也可以参与联系。图1.28是应用聚集的例子。
- 聚集是一种抽象，通过它联系被作为高层实体集。例如设实体集A与实体集B有联系R，则实体A、B以及联系R整体可被看成是另一高层实体集C。提出与使用聚集的概念能带来这些好处：能一定程度上来消除冗余；能将联系作为抽象实体看待，增加了分析问题的新思路；这样能允许联系之间存在联系，有了新的表达能力。

6、聚集

- 实例：职工参加项目，并在此过程中可能使用机器，则其E-R图的设计有两种方案可选，如图1.28（联系表为表1.2所示）、图1.29（联系表为表1.3、表1.4所示）所示，其中图1.29是应用聚集的一个例子。读者可分析这两种方案的利弊。

图1.28 三个实体间的联系的E-R图

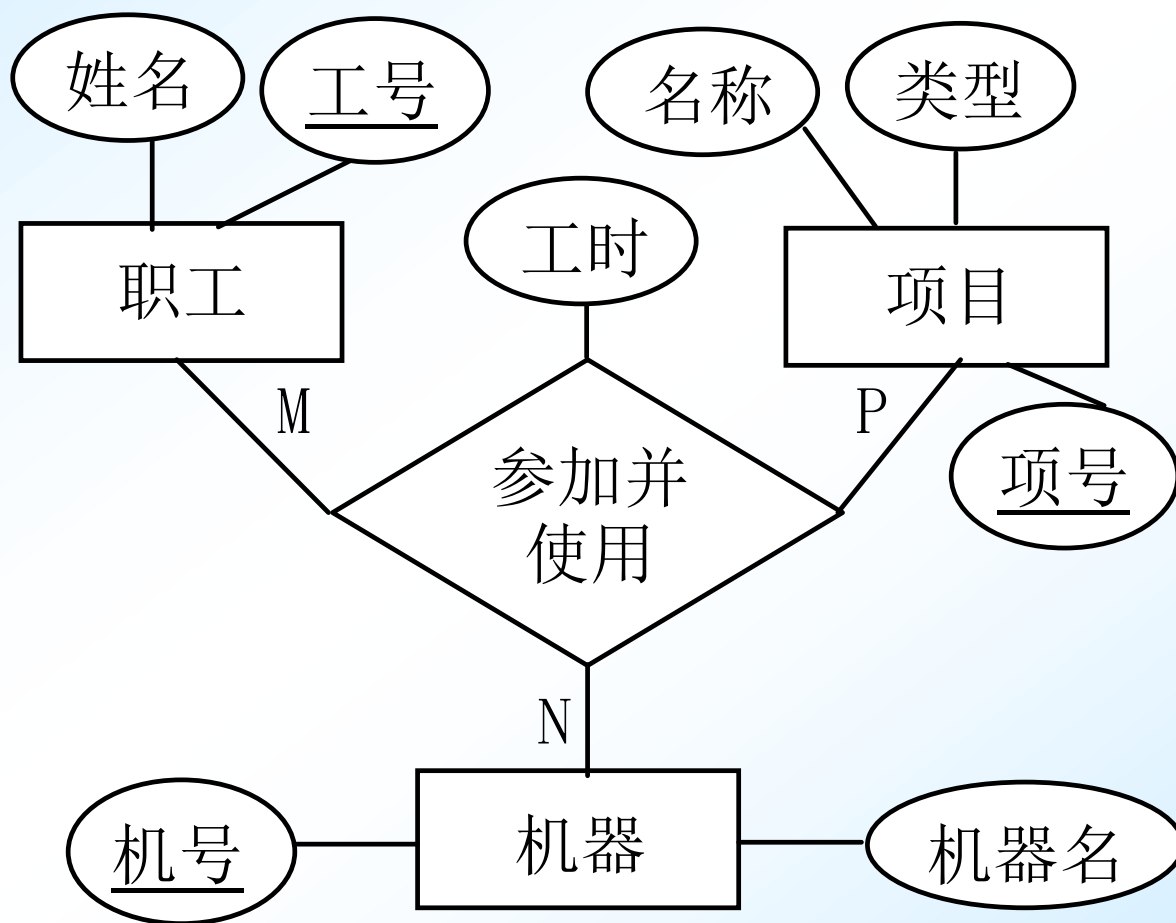
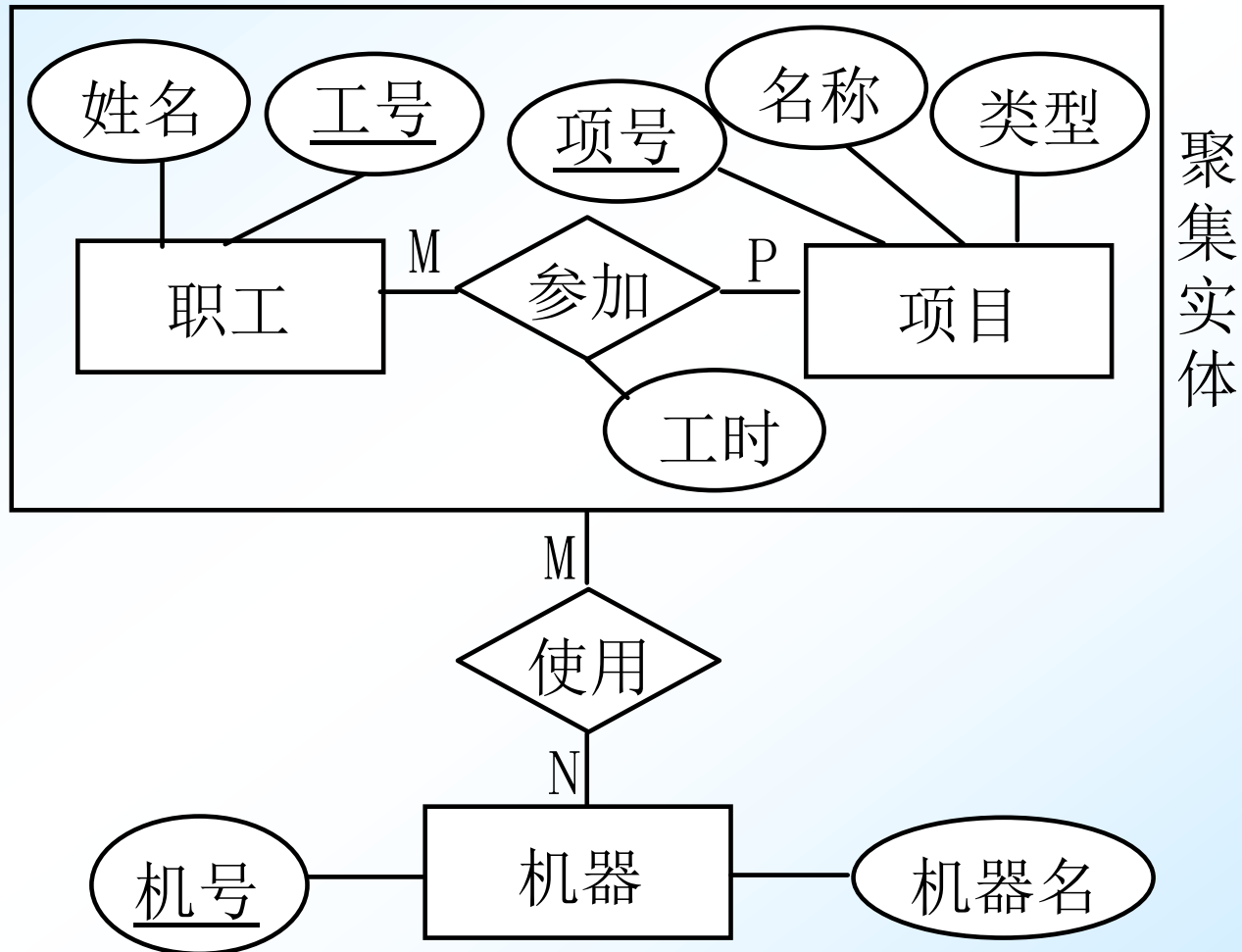


图1.29 应用聚集的E-R图



7、范畴

- 在模拟现实世界时，有时要用到由不同类型的实体组成的实体集，例如车主这个实体集的成员可能是单位，也可能是个人。这种由不同类型实体组成的实体集不同于前面所定义的实体集，为了区分起见，特名之为范畴（category）。设 E_1, E_2, \dots, E_n 是 n 个不同类型的实体集，则范畴 T 课定义为：

$$T (E_1 \cup E_2 \cup \dots \cup E_n)$$

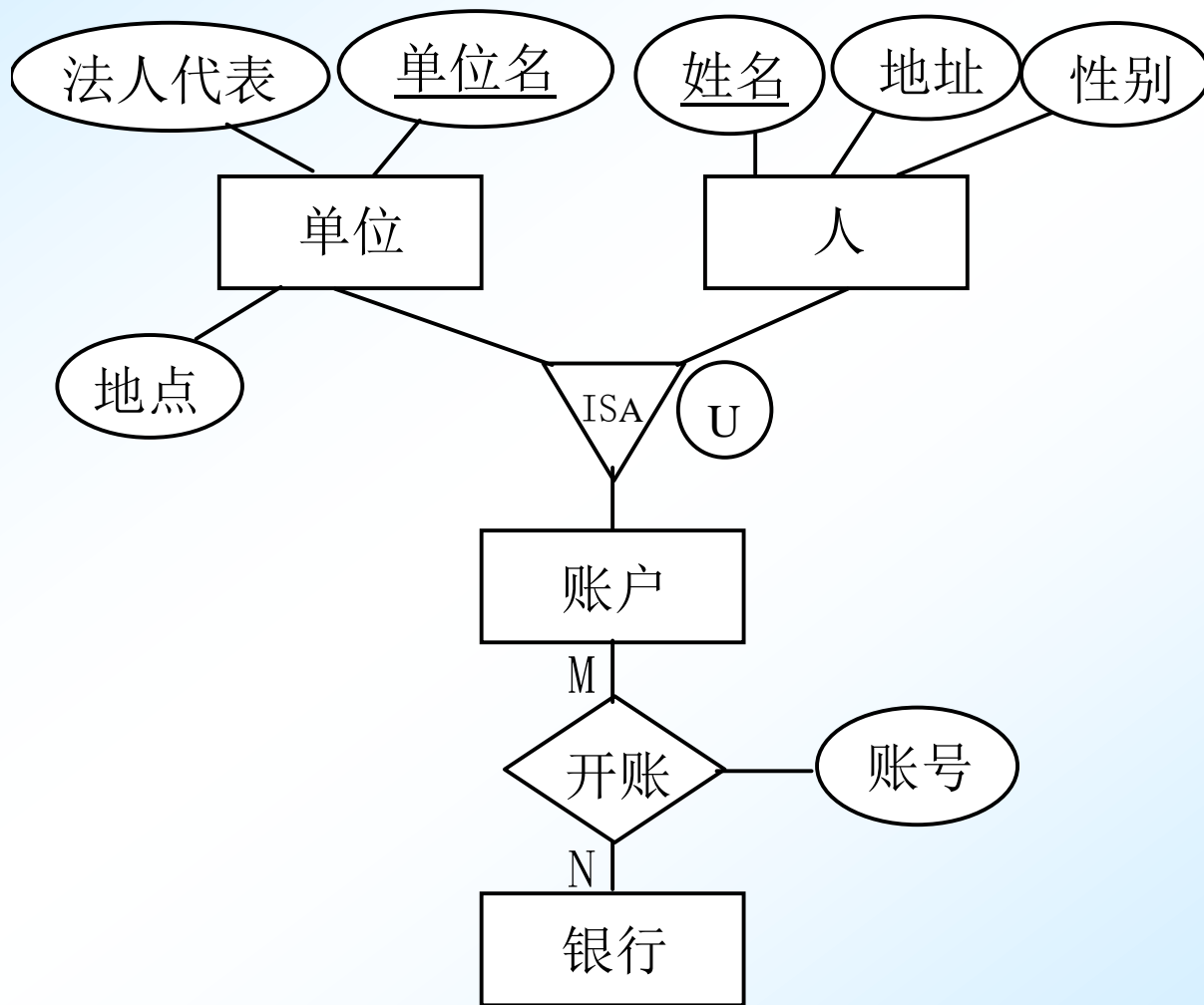
7、范畴

- E1,E2,...En也称为T的超实体集。图1.30是应用范畴一个例子，其中账户是个范畴，可以是单位，也可以是个人。用旁含“U”的圆圈的“ISA”三角形标记来表示并操作。范畴也继承其超实体集的属性，但与子实体集的继承规则不一样。子实体集继承所有超实体集的属性，例如图1-27的在职进修生继承教职工和学生的所有属性。

7、范畴

- 而范畴的继承是有选择性的，例如在图1.30中，如果账户是单位，则继承单位的属性；如果账户是个人，则继承人的属性。这叫选择性继承。范畴与具有多超实体集的子实体集在形式上有些相似，但意义完全不同。范畴是超实体集并的子集，而子实体集是超实体集交的子集。例如，图1.26中在职进修生既是教职工成员，又是学生的成员，而在图1.30中，账户只是一个单位或者一个个人。

图1.30 应用范畴的E-R图



1.2.4 层次模型

- 在数据库领域中，有四种最常用的数据模型，它们是：被称为非关系模型的层次模型、网状模型、关系模型和面向对象模型。本章简要介绍层次模型、网状模型、关系模型。
 - 层次模型的数据结构
 - 多对多联系在层次模型中的表示
 - 层次模型的数据操纵与约束条件
 - 层次模型的存储结构
 - 层次模型的优缺点

层次模型的数据结构

在数据库中，对满足以下两个条件的数据模型称为层次模型。

(1) 有且仅有一个节点无双亲，这个节点称为“根节点”。

(2) 其他节点有且仅有一个双亲。

一个层次模型在理论上可以包含任意有限个记录型和字段，但任何实际的系统都会因为存储容量或实现复杂度而限制层次模型中包含的记录型个数和字段的个数。

若用图来表示，层次模型是一棵倒立的树。节点层次（Level）从根开始定义，根为第一层，根的子节点称为第二层，根称为其子女的双亲，同一双亲的子节点称为兄弟。

层次模型的数据结构

图1.31 给出了一个系的层次模型。

层次模型对具有一对多的层次关系的描述非常自然、直观、容易理解，这是层次数据库的突出优点。

图1.32 是图1.31的具体化，成为一个教师-学生层次数据库。

图1.33 是图1.32数据模型的一个值。

图1.31 一个层次模型的示例

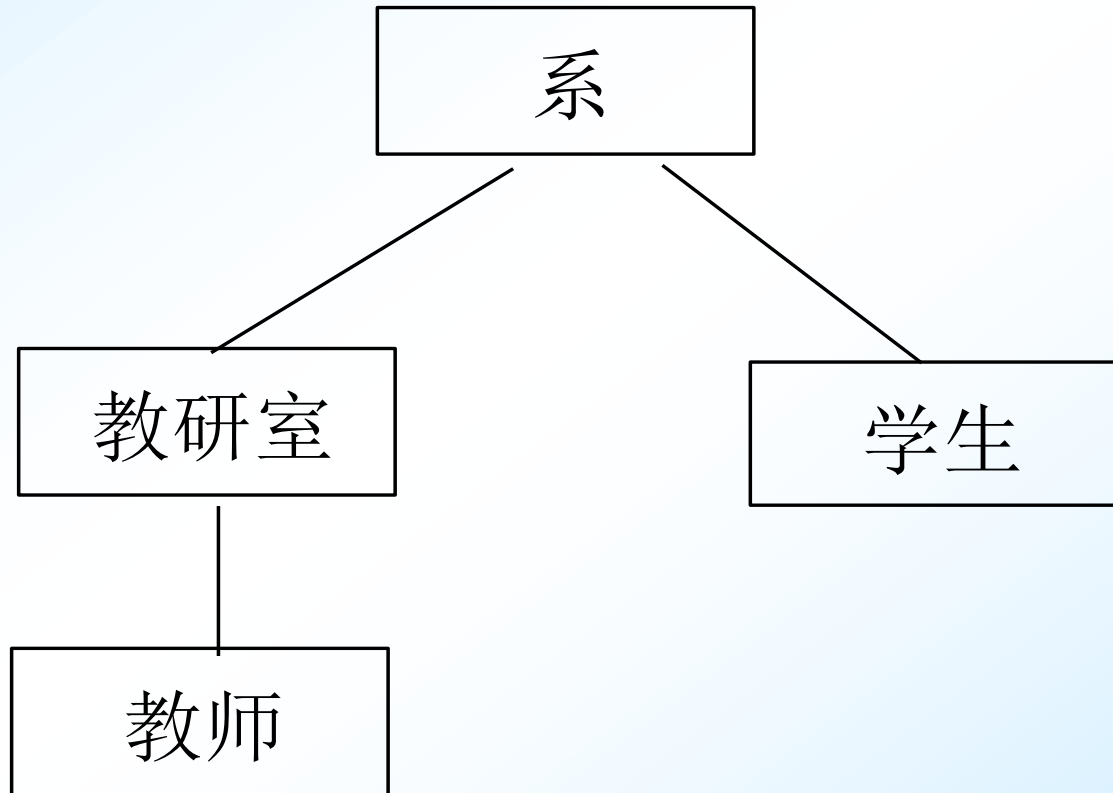


图1.32 教师-学生数据库模型

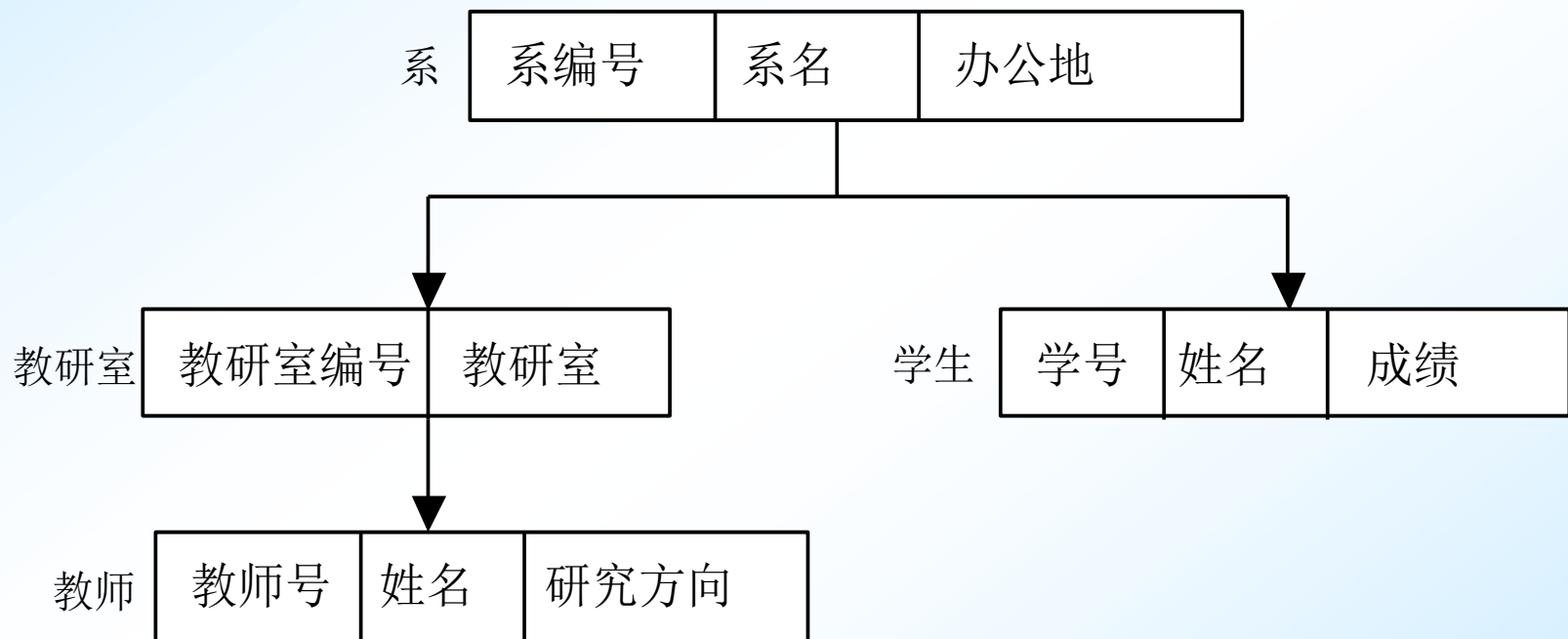
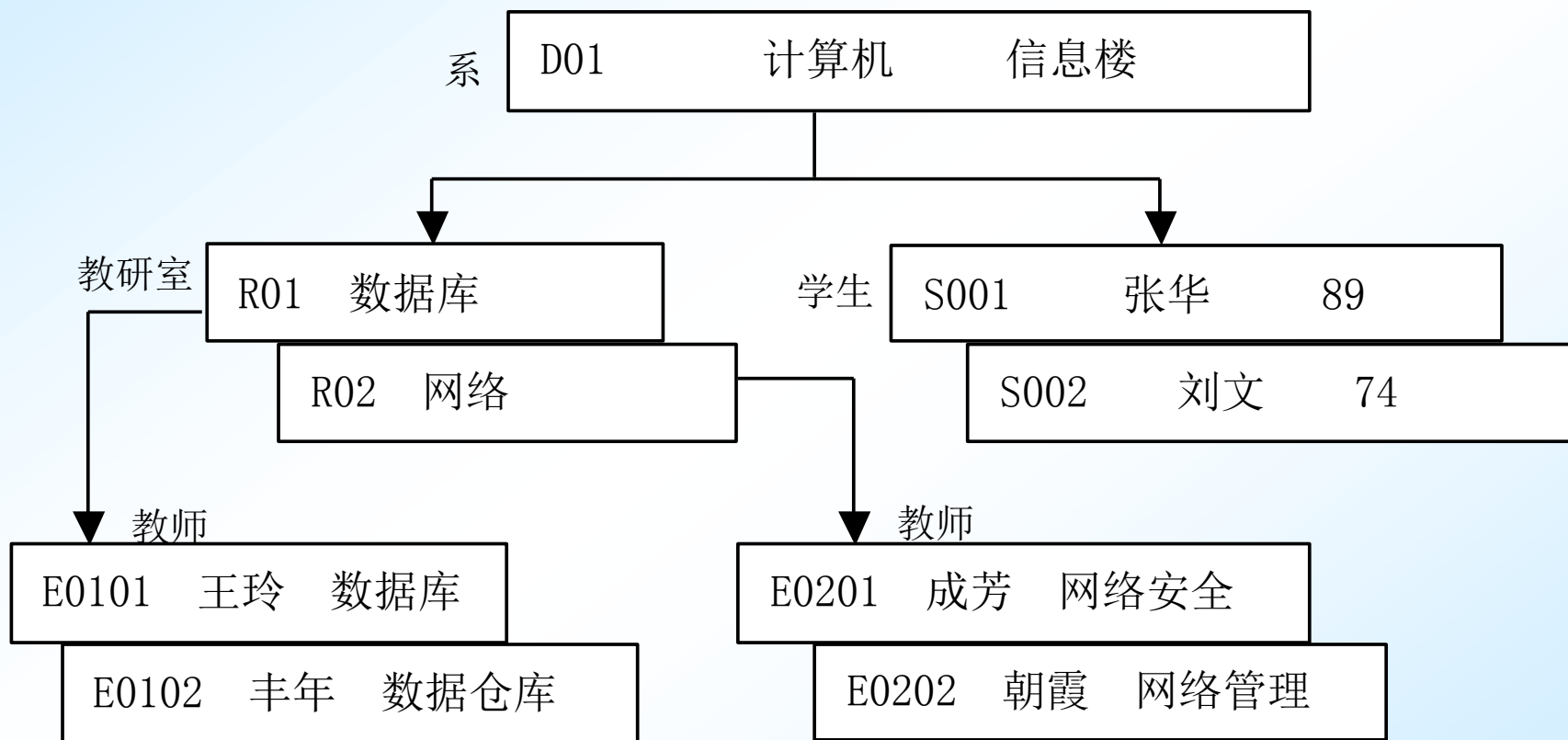


图1.33 教师-学生数据库的一个值



层次模型的存储结构

层次数据库中不仅要存储数据本身，还要存储数据之间的层次联系。

方法：

(1) 邻接法

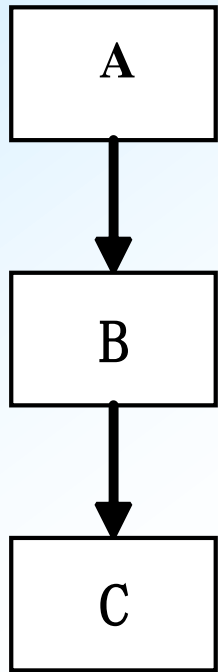
按照层次树前序的顺序把所有记录值依次邻接存放，即通过物理空间的位置相邻来体现层次顺序。例如对于图1.34 (a) 的数据库，按邻接法存放图1.34(b)中以记录A1为首的层次记录实例集，则应如图1.35所示存放。

层次模型的存储结构

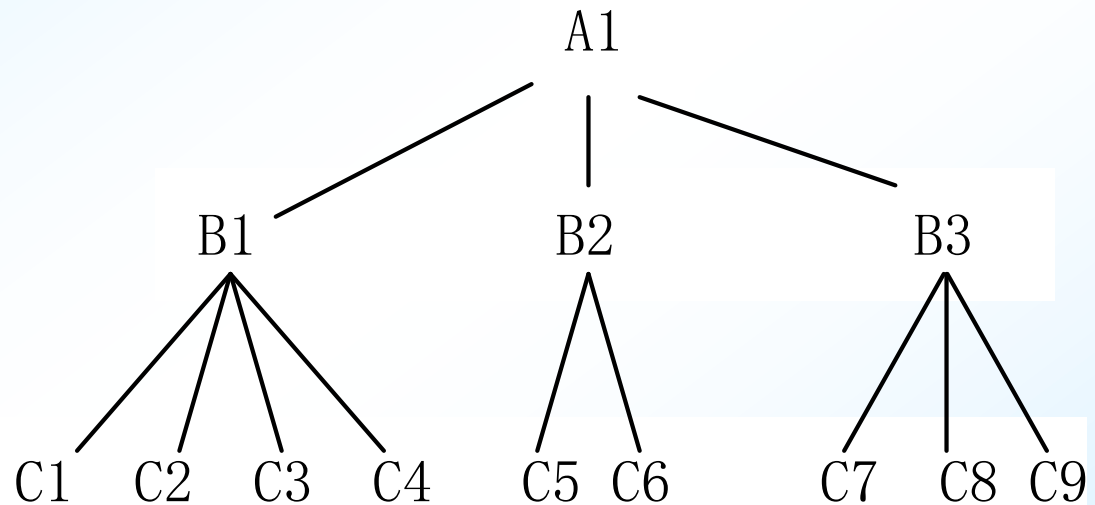
(2) 链接法

用指引元来反映数据之间的层次联系。则如图1.36所示，其中图1.36 (a) 每个记录设两类指引元，分别指向最左边的子女和最近的兄弟，这种链接方法称为子女-兄弟链接法；图1.36 (b) 按树的前序顺序链接各记录值，这种链接方法称为层次序列链接法。

图1.34 层次数据库及其实例



(a)

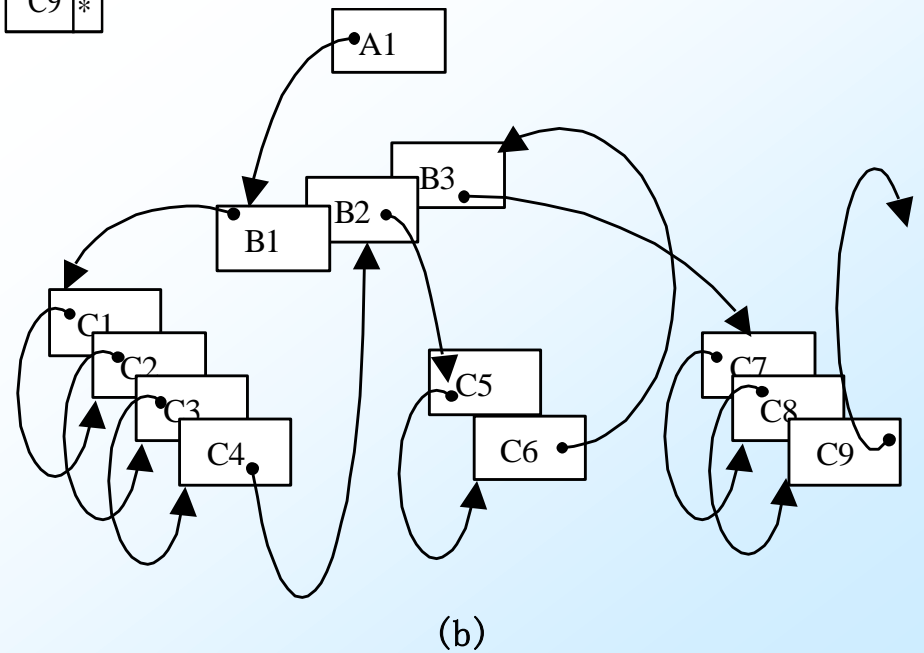
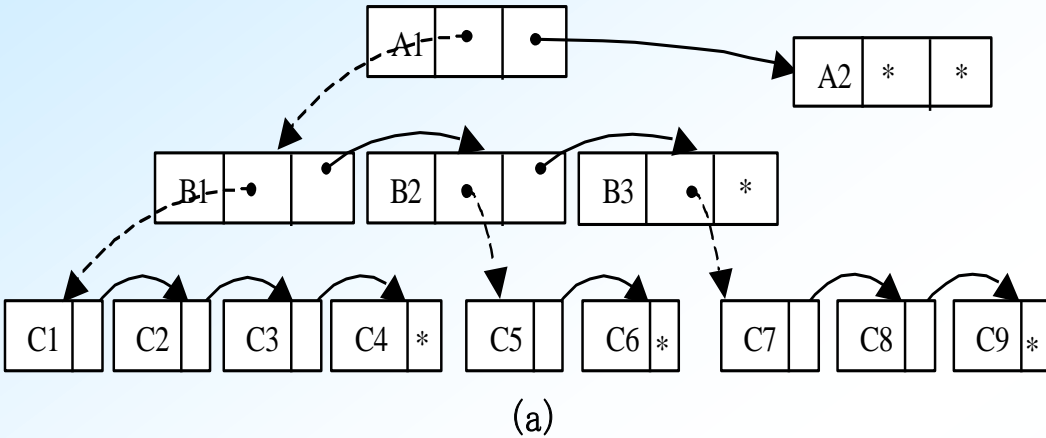


(b)

图1.35 邻接法

A1	B1	C1	C2	C3	C4	B2	C5	C6	B3	C7	C8	C9
----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----

图1.36 链接法



层次模型的优缺点

- 优点：

- (1) 层次模型本身比较简单
- (2) 实体间联系是固定的
- (3) 良好的完整性支持

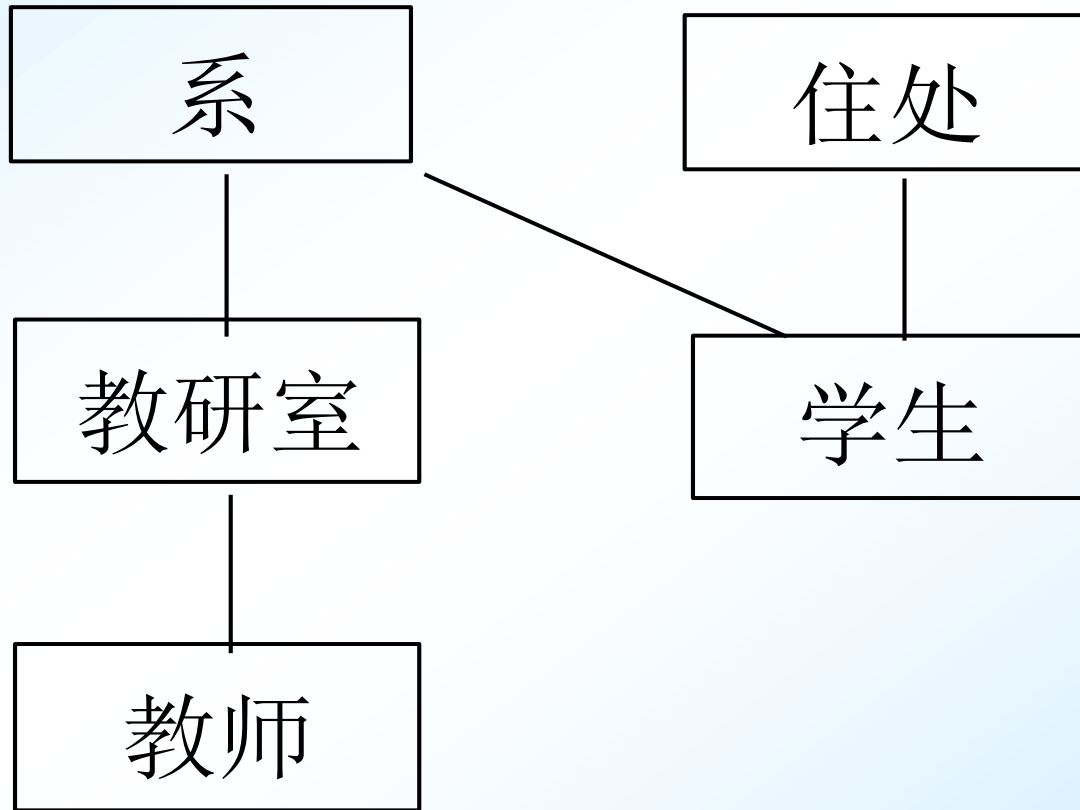
- 缺点：

- (1) 层次模型表示非层次性的方法很笨拙
- (2) 对插入和删除操作的限制太多
- (3) 查询子女结点必须通过双亲结点，缺乏快速定位机制。
- (4) 由于结构严密，层次命令趋于程序化。

1.2.5 网状模型

- 网状数据模型的典型代表是DBTG系统，也称CODASYL系统，它是20世纪70年代数据系统语言研究会CODASYL（Conference On Data Systems Language）下属的数据库任务组（Data Base Task Group，简称DBTG）提出的一个系统方案。
- 图1.37给出了一个抽象的简单的网状模型。

图1.37 简单的网状模型



一、网状模型的数据结构

- 在数据库中，把满足以下两个条件的基本层次联系集合称为网状模型：

- (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。

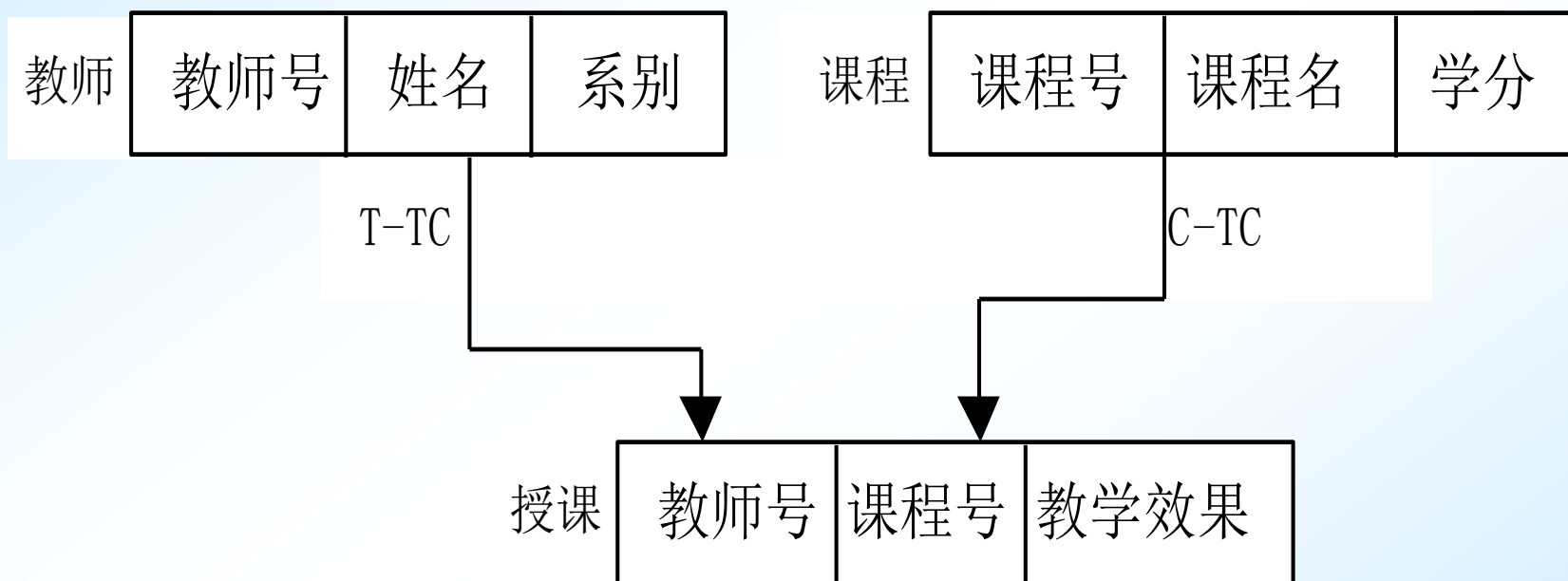
它去掉了层次模型的两个限制，此外它还允许两个结点之间有多种联系。因此网状模型可以更直接地去描述现实世界。

以教师授课为例，看看网状数据库模式是怎样组织数据的。

一、网状模型的数据结构

- 按照常规语义，一个教师可以讲授若干门课程，一门课程可以由多个教师讲授，因此教师与课程之间是多对多联系。这里引进一个教师授课的连结记录，它由两个数据项组成，即教师号、课程号，表示某个教师讲授一门课程。
- [图1.38](#)为教师授课数据库的网状数据库模式。

图1.38 为教师授课数据库的网状数据库模式。



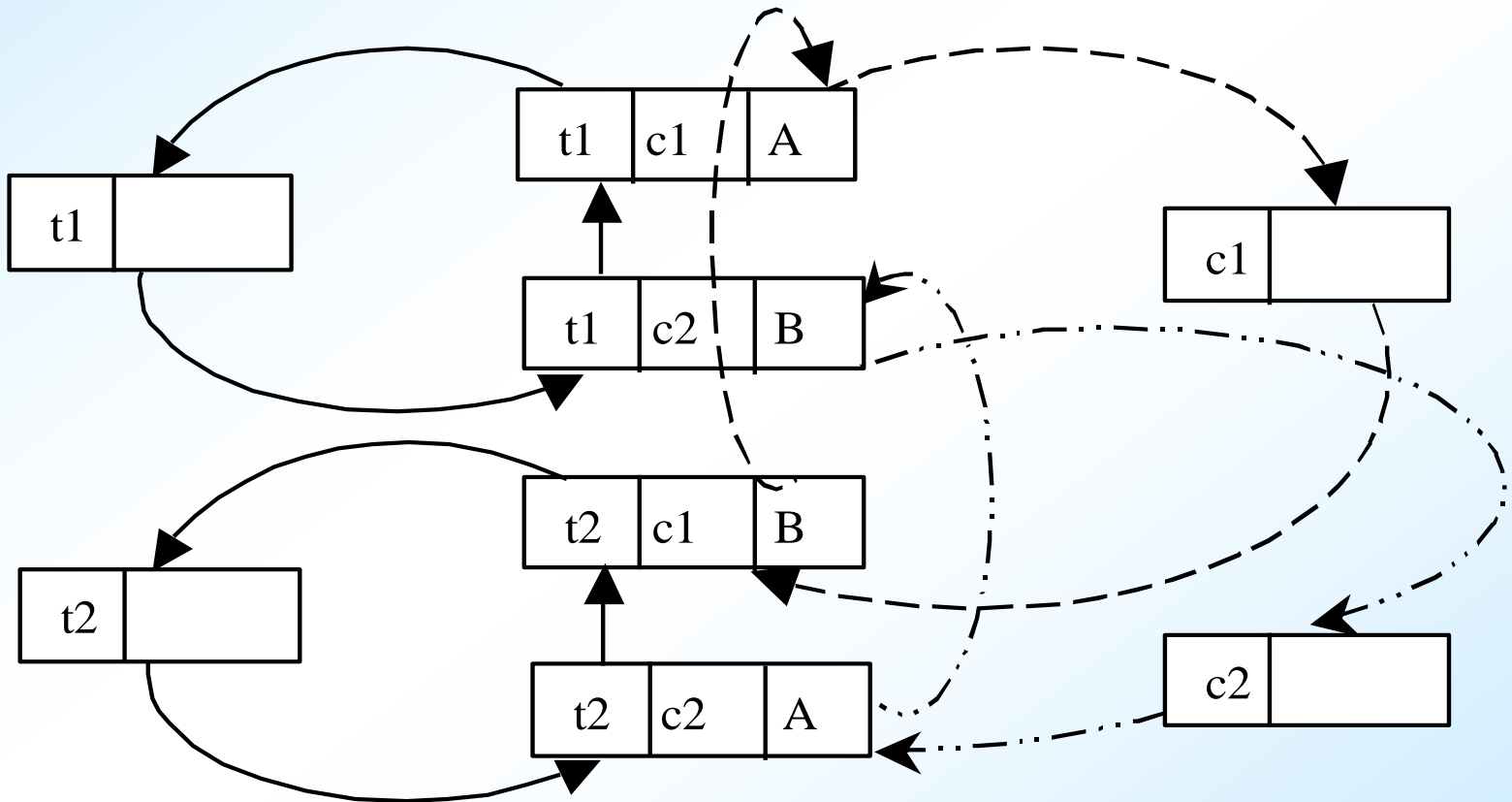
二、网状模型的数据操作与完整性约束

- 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，码即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某些约束条件。

三、网状模型的存储结构

- 常用的方法是链接法，包括单向链接、双向链接、环状链接、向首链接等，此外其他实现方法，如指引元阵列法、二进制阵列法、索引法等，依具体系统不同而不同。
- 教师任课数据库中，记录之间的联系用单向环状链接法实现，[如图1.39](#)所示。

图1.39 教师/授课/课程的网状数据库实例



四、网状模型的优缺点

- 优点主要有：
 - (1) 能够更为直接地描述现实世界
 - (2) 具有良好的性能，存取效率较高。
- 缺点主要有：
 - (1) 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。
 - (2) 其DDL、DML语言复杂，用户不容易使用。

1.2.6 关系模型

- 关系模型是目前最重要的一种模型。美国IBM公司的研究员E. F. Codd于1970年发表题为“大型共享系统的关系数据库的关系模型”的论文，文中首次提出了数据库系统的关系模型。
- 数据库领域当前的研究工作都是以关系方法为基础的。
- 数据模型的三个组成要素，主要体现如下：
 - 一、关系模型的数据结构
 - 二、关系模型的数据操纵与约束条件
 - 三、关系模型的存储结构

关系模型的数据结构

- 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。每一行称为一个元组，每一列称为一个属性（或字段）
- 通过 [图1.40](#) 所示的教师登记表，介绍关系模型中的相关的术语
- 关系：一个关系对应一张二维表，[图1.40](#) 表示的就是一张教师登记表
- 元组：二维表中的一行称为一个元组
- 属性：二维表中的一列称为一个属性，对应每一个属性的名字称为属性名
- 主码：二维表中的某个属性或是属性组可以唯一确定一个元组，则称为主码，也称为关系键

关系模型的数据结构

- 域：属性的取值范围称为域
- 分量：元组中的一个属性值
- 关系模式：表现为关系名和属性集的集合，是对关系的具体描述。一般表示为：
- 关系名（属性1，属性2，...，属性N）

关系模型要求关系必须是规范化的，即要求关系必须满足一定规范条件，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。

例如：[图1.41](#) 中出产日期是可分的数据项，出产日期有可以分为年，月，日。因此，图1.41的表就不符合关系模型要求。

图1.40 关系模型的数据结构及术语

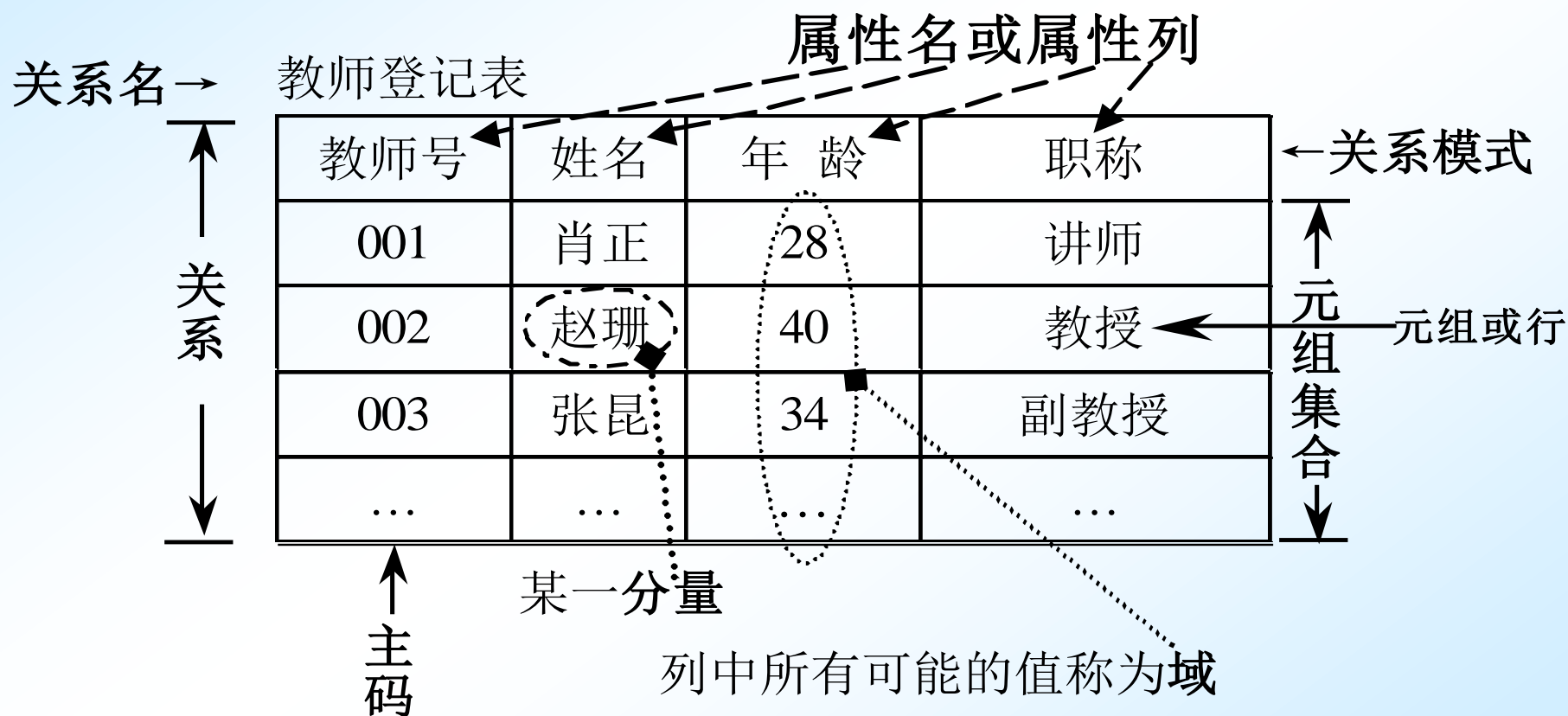


图1.41 表中有表的示例

产品号	产品名	型号	出产日期		
			年	月	日
032456	风扇	A134	2004	05	12
...

二、关系模型的数据操纵与约束条件

- 关系模型的操作主要包括查询、插入、删除和修改数据。这些操作必须满足关系的完整性约束条件，即实体完整性、参照完整性和用户定义完整性。
- 在非关系模型中，操作对象是单个记录，而关系模型中的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。关系模型把对数据的存取路径向用户隐蔽起来，从而大大地提高了数据的独立性。

三、关系模型的存储结构

在关系数据模型中，实体及实体间的联系都用表来表示。在数据库的物理组织中，表以文件形式存储，每一个表通常对应一种文件结构，也有多个表对应一种文件结构。

四、关系模型的优缺点

关系模型优点：

- (1) 有较强的数学理论根据。
- (2) 数据结构简单、清晰，用户易懂易用。
- (3) 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。

关系模型具有查询效率不如非关系模型效率高缺点。为了提高性能，必须对用户的查询进行优化，增加了开发数据库管理系统的负担。

1.2.7 面向对象模型*

- 面向对象数据模型（Object-Oriented data model,简称O-O data model）就是一种可扩展的数据模型。它又称对象数据模型（Object data model），以面向对象数据模型为基础的DBMS称为O-O DBMS或对象数据库管理系统（ODBMS）。
- 面向对象数据模型提出于20世纪70年代末，80年代初。它吸收了语义数据模型和知识表示模型的一些基本概念，同时又借鉴了面向对象程序设计语言和抽象数据类型的一些思想。面向对象数据模型及其数据库系统依然在不断发展成熟中，其相关的概念与术语仍不统一。

1.2.7 面向对象模型*

- 虽然关系模型比层次模型、网状模型简单灵活，但它还不能很好地表达现实世界中存在的许多复杂的数据结构，譬如CAD数据、图形数据、嵌套递归的数据等，它们就需要如面向对象模型这样的新模型来表达。
- 面向对象模型中，基本的概念是对象、类及其实例、类的层次结构和继承、对象的标识等。

1.2.7 面向对象模型*

- (1) 对象 (Object)
- 对象是现实世界中实体的模型化，与记录概念相仿，但远比记录复杂。每个对象有一个唯一的标识符，把状态 (State) 和行为(havior)封装 (Encapsulate) 在一起。其中，对象的状态是该对象属性值的集合，对象的行为是在对象状态上操作的方法集。

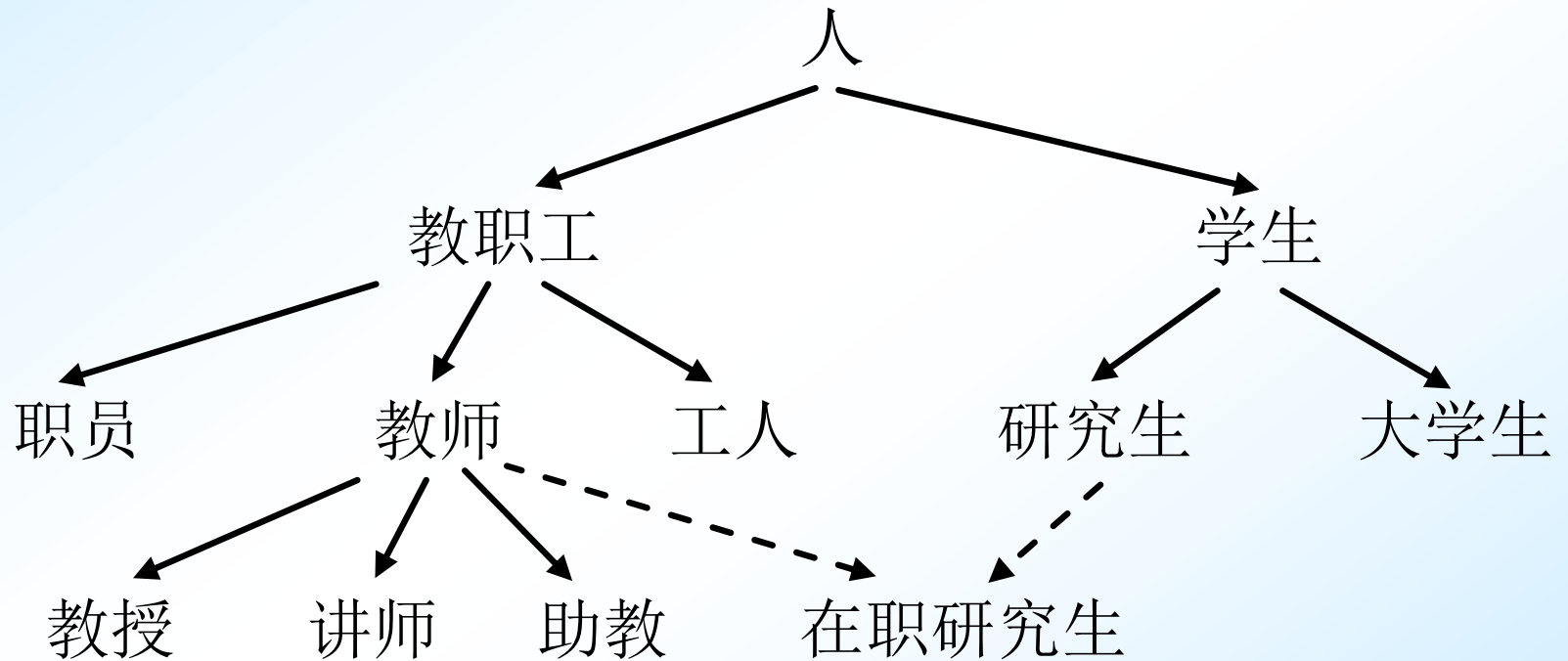
1.2.7 面向对象模型*

- (2) 类 (Class) 和实例 (Instance)
- 一个数据库一般包含大量的对象。如果每个对象都附有属性和方法的定义说明, 则会有大量的重复。为了解决这个问题, 同时也为了概念上的清晰, 常常把类似的对象归并为类。类中每个对象称为实例。同一类的对象具有共同的属性和方法, 这些属性和方法可以在类中统一说明, 而不必在类的每个实例中重复。消息传送到对象后, 可以在其所属的类中找到相应的方法和属性说明。同一类中的对象的属性虽然是一样的, 但这些属性所取的值会因各个实例而各不同。因此, 属性又称为实例变量。有些变量的值在全类中是共同的, 这些变量称为类变量。

1.2.7 面向对象模型*

- (3) 类层次结构和继承
- 类的子集也可以定义为类，称为这个类的子类，而该类称为子类的超类（superclass）。子类还可以再分为子类，如此可以形成一个层次结构。图2-35是一个类层次结构的例子，一个子类可有多个超类，有直接的，也有间接地。
- 一个对象既属于它的类，也属于它的所有超类。为了在概念上区分起见，在O-O数据模型中，对象与类之间的关系有时用不同的名词。对象只能是它所属类中最特殊化的那个子类的实例，但可以是它的所有超类的成员。

图1.42 类层次结构



1.2.7 面向对象模型*

- (4) 对象的标识
- 在O-O数据模型中，每个对象都有一个在系统内唯一的和不变的标识符，称为对象标识符（Object Identifier，简称OID）。OID一般由系统产生，用户不得修改。两个对象即使属性值和方法都一样，若OID不同，则仍被认为是两个相等而不同的对象。
- 面向对象数据模型已经用做O-O DBMS的数据模型。由于其语义丰富，表达比较自然，它也适合作为数据库概念设计的数据模型（即概念模型）。随着面向对象程序设计的广发应用和数据库新应用的不断涌现。面向对象数据模型可望在计算机科学技术领域中得到普遍的认可。

表1.5 数据模型比较表

比较项	层次模型	网状模型	关系模型	面向对象模型
创始	1968年IBM公司的IMS系统	1969年CODASYL的DBTG报告（1971年通过）	1970年E. F. Codd提出关系模型	20世纪80年代
典型产品	IMS	IDS/ II， IMAGE/3000， IDMS等	Oracle, Sybase, DB2, SQL Server等	ONTOS DB
盛行时期	20世纪70年代	20世纪70年代到80年代中期	20世纪80年代至今	20世纪90年代至今
数据结构	复杂（树形结构），要加树形限制	复杂（有向图结构），结构上无需严格限制	简单（二维表），无需严格限制	复杂（嵌套，递归），无需严格限制
数据联系	通过指针连接记录型，联系单一	通过指针连接记录型，联系多样，较复杂	通过联系表（含外码），联系多样	通过对象标识
查询语言	过程式，一次一记录。查询方式单一（双亲到子女）	过程式，一次一记录。查询方式多样	非过程式，一次一集合。查询方式多样	面向对象语言
实现难易	在计算机中实现较方便	在计算机中实现较困难	在计算机中实现较方便	在计算机中实现有一定难度
数学理论基础	树（研究不规范，不透彻）	无向图（研究不规范，不透彻）	关系理论（关系代数、关系演算），研究深入、透彻、	连通有向图（研究还不透彻）

表1.6 现实世界、信息世界、机器世界/关系数据库间术语对照表

现实世界	信息世界	机器世界/关系数据库
事物	实体	记录/元组（或行）
若干同类事物	实体集	记录集（即文件）/元组集（即关系）
若干特征刻画的事物	实体型	记录型/二维表框架（即关系模式）
事物的特征	属性	字段（或数据项）/属性（或列）
事物之间的关联	实体型（或实体）之间的联系	记录型之间的联系/联系表（外码）
事物某特征的所有可能值	域	字段类型/域
事物某特征的一个具体值	一个属性值	字段值/分量
可区分同类事物的特征或若干特征	码	关键字段/关系键（或主码）

1.3 数据库系统结构

- 可以有多种不同的层次或不同的角度来考查数据库系统的结构。从数据库管理系统内部系统结构看，数据库系统通常采用三级模式结构。从数据库外部的体系结构看，数据库系统的结构分为集中式结构、分布式结构、客户/服务器和并行结构等。

1.3.1 数据库系统的三级模式结构

1.3.2 数据库的二级映象功能与数据独立性

1.3.3 数据库管理系统的工作过程

数据库系统的三级模式结构

- 数据库系统的三级模式结构是指外模式、模式和内模式。如图1.44
- 一、外模式 (External Schema)
- 外模式也称子模式 (SubSchema) 或用户模式，是三级模式的最外层，它是数据库用户能够看到和使用的局部数据的逻辑结构和特征的描述。
- 把普通用户看到和使用的数据库内容称为视图。视图集也称为用户级数据库，它对应于外模式。外模式通常是模式的子集。一个数据库可以有多个外模式。同一外模式也可以为某一用户的多个应用系统所用，但一个应用程序只能使用一个外模式。

数据库系统的三级模式结构

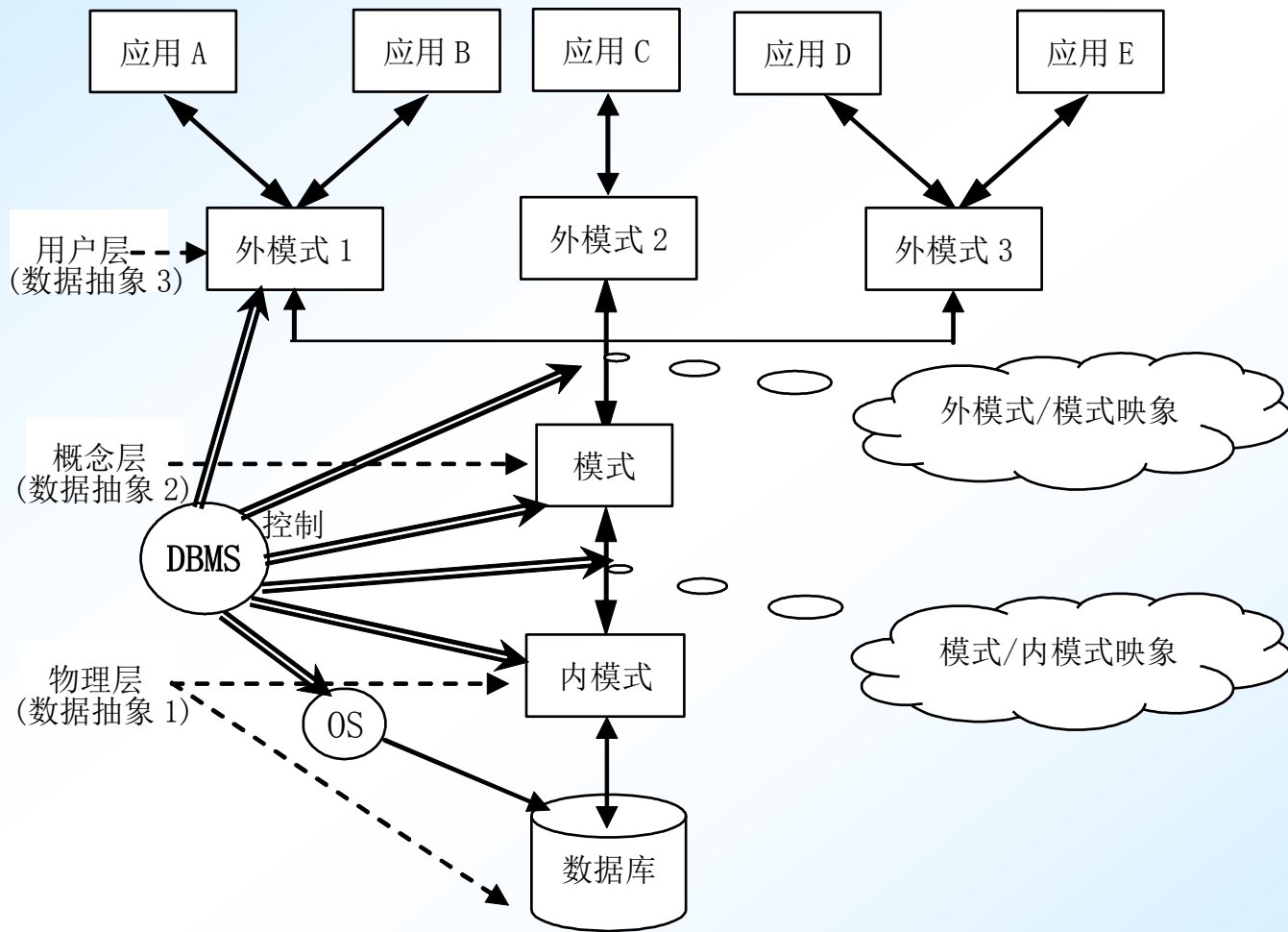


图1.44 数据库系统的三级模式结构

数据库系统的三级模式结构

● 二、模式 (Schema)

模式又称概念模式，也称逻辑模式，是数据库中全体数据的逻辑结构和特征的描述，是所有用户的公共数据视图，是数据视图的全部。它是数据库系统模式结构的中间层。

概念模式实际上是数据库数据在逻辑级上的视图。一个数据库只有一个模式。定义模式时不仅要定义数据的逻辑结构，而且要定义数据之间的联系、定义与数据有关的安全性、完整性要求。

数据库系统的三级模式结构

- **三、内模式 (Internal Schema)**

内模式也称为存储模式，一个数据库只有一个内模式。它是数据物理结构和存储方式的描述，是数据在数据库内部的表示方式。

数据库系统三级模式结构概念比较，请参见表1.7

比较	外模式	模式	内模式
定义	也称子模式或用户模式，还称用户级模式	也称逻辑模式，还称概念级模式	也称存储模式，还称物理级模式
	是数据库用户能够看见和使用的局部数据的逻辑结构和特征的描述	是数据库中全体数据的逻辑结构和特征的描述，它包括：数据的逻辑结构、数据之间的联系和与数据有关的安全性、完整性要求。	它是数据物理结构和存储方式的描述。
特点1	是各个具体用户所看到的数据视图，是用户与DB的接口。	是所有用户的公共数据视图。一般只有DBA能看到全部	数据在数据库内部的表示方式。
特点2	可以有多个外模式	只有一个模式	只有一个内模式

数据库系统三级模式结构概念比较，请参见表1.7

比较	外模式	模式	内模式
特点3	针对不同用户，有不同的外模式描述。每个用户只能看见和访问所对应的外模式中的数据，数据库中其余数据是不可见的。所以外模式是保证数据库安全性的一个有力措施。	数据库模式以某一种数据模型（层状、网状、关系）为基础，统一综合地考虑所有用户的需求，并将这些需求有机地结合成一个逻辑整体。	以前由DBA定义，现基本由DBMS定义。
特点4	面向应用程序或最终用户。	由DBA定义与管理。	由DBA定义或由DBMS预先设置。
DDL	DBMS提供三种模式的描述语言（DDL）来严格定义三种模式。如：子模式DDL、模式DDL和内模式DDL。子模式DDL和用户选用的程序设计语言具有相容的语法。如：Cobol子模式DDL。关系数据库三种模式的描述语言统一于SQL语言中。		

数据库的二级映象功能与数据独立性

- 为了能够在内部实现这三个抽象层次的联系和转换，数据库管理系统在这三级模式之间提供了两层映象：外模式/模式映象，模式/内模式映象。
- 这两层映象保证了数据库系统的数据能够具有较高的逻辑独立性和物理独立性。
- **一、外模式/模式映象**
- 模式描述的是数据的全局逻辑结构，外模式描述的是数据的局部逻辑结构。数据库系统都有一个外模式/模式映象，它定义了该外模式与模式之间的对应关系。
- 当模式改变时，由数据库管理员对各个外模式/模式映象作相应改变，可以使外模式保持不变。
- 应用程序是依据数据的外模式编写的，保证了数据与程序的逻辑独立性，简称为数据逻辑独立性。

数据库的二级映象功能与数据独立性

● 二、模式/内模式映象

- 数据库中只有一个模式，也只有一个内模式，所以模式/内模式映象是唯一的，它定义了数据库全局逻辑结构与存储结构之间的对应关系。当数据库的存储结构改变了，由数据库管理员对模式/内模式映象作相应改变，可以使模式保持不变，从而应用程序也不必改变。保证了数据与程序的物理独立性，简称为数据物理独立性。
- 在数据库的三级模式结构中，数据库模式即全局逻辑结构是数据库的中心与关键，它独立于数据库的其他层次。

数据库的二级映象功能与数据独立性

- 数据库的内模式依赖于它的全局逻辑结构，但独立于数据库的用户视图即外模式，也独立于具体的存储设备。
- 数据库的外模式面向具体的应用程序，它定义在逻辑模式之上，但独立于内模式和存储设备。
- 数据库的二级映象保证了数据库外模式的稳定性，从而根本上保证了应用程序的稳定性，使得数据库系统具有较高的数据与程序的独立性。数据库的三级模式与二级映象使得数据的定义和描述可以从应用程序中分离出去。

1.3.3 数据库管理系统的工作过程

- 数据库管理系统控制的数据操作过程基于数据库系统的三级模式结构与二级映象功能，总体操作过程能从其读或写一个用户记录的过程大体反映出来。
- 下面就以应用程序从数据库中读取一个用户记录的过程（如图1.45所示）来说明。
- 按照步骤解释运行过程如下：
 - （1）应用程序A向DBMS发出从数据库中读用户数据记录的命令；
 - （2）DBMS对该命令进行语法检查、语义检查，并调用应用程序A对应的子模式，检查A的存取权限，决定是否执行该命令。如果拒绝执行，则转（10）向用户返回错误信息

1.3.3 数据库管理系统的工作过程

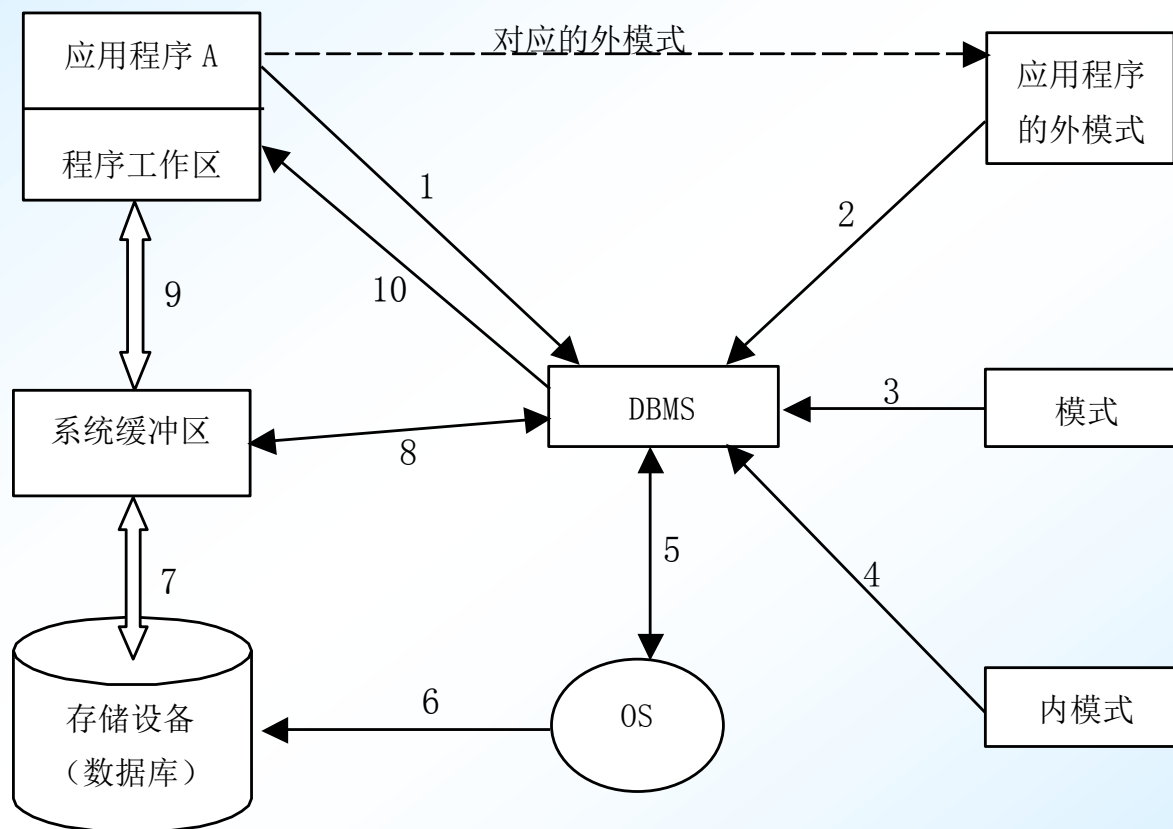


图1.45 DBMS读取用户记录的过程示意图

1.3.3 数据库管理系统的工作过程

- (3) 在决定执行该命令后，DBMS调用模式，依据子模式/模式映象的定义，确定应读入模式中的哪些记录；
- (4) DBMS调用内模式，依据模式/内模式映象的定义，决定应从哪个文件、用什么存取方式、读入哪个或哪些物理记录；
- (5) DBMS向操作系统发出执行读取所需物理记录的命令；

1.3.3 数据库管理系统的工作过程

- (6) 操作系统执行从物理文件中读数据的有关操作；
- (7) 操作系统将数据从数据库的存储区送至系统缓冲区；
- (8) DBMS依据内模式/模式（模式/内模式映象的反方向看待，并不是另一种新映象，模式/子模式映象也是类似情况）、模式/子模式映象的定义，导出应用程序A所要读取的记录格式；

1.3.3 数据库管理系统的工作过程

- (9) DBMS将数据记录从系统缓冲区传送到应用程序A的用户工作区；
- (10) DBMS向应用程序A返回命令执行情况的状态信息。
- 至此，DBMS就完成了读用户数据记录的过程。DBMS向数据库写一个用户数据记录的过程经历的环节类似于读，只是过程基本是相反而已。由DBMS控制的大量用户数据的存取操作，可以理解为就是由许许多多这样的读或写的基本过程组合完成的。

1.4 数据库系统的组成

数据库系统是指计算机系统中引入数据库后的整个人机系统。为此，数据库系统应由计算机硬件、数据库、计算机软件及各类人员组成。

一、硬件平台

二、数据库

数据库是存放数据的地方，是存储在计算机内有组织的大量可共享的数据集合，具有尽可能少的冗余和较高的数据独立性，有完整的自我保护能力和数据恢复能力。

1.4 数据库系统的组成

三、软件

数据库系统的软件主要包括：

- (1) 支持DBMS运行的操作系统；
- (2) DBMS：DBMS可以通过操作系统对数据库的数据进行存取、管理和维护；
- (3) 具有与数据库接口的高级语言及其编译系统；
- (4) 以DBMS为核心的应用开发工具，为特定应用环境开发的数据库应用系统。

1.4 数据库系统的组成

四、用户

进行管理和维护数据库系统的人员——数据库管理员；

具体职责包括：

- (1) 决定数据库中的信息内容和结构
- (2) 决定数据库的存储结构和存取策略
- (3) 定义数据的安全性要求和完整性约束条件
- (4) 监控数据库的使用和运行
- (5) 数据库的改进和重组重构

1.4 数据库系统的组成

数据库应用系统分析设计的人员——系统分析员和数据库设计人员；

- ◆具体开发数据库系统的人员——数据库应用程序员；
- ◆使用数据库系统的人员——最终用户。

1.5 数据库技术的研究领域及其发展*

1.5.1 数据库技术的研究领域

数据库技术的研究领域十分广泛，概括地而言包括以下三方面：

- 1、DBMS系统软件的研制
- 2、数据库应用系统设计与开发的研制
- 3、数据库理论的研究

1.5.2 数据库技术的发展

数据库技术产生于20世纪60年代中期，由于其在商业领域的成功应用，在20世纪80年代后，得到迅速推广，新的应用对数据库技术在数据存储和管理方面提出了更高的要求，从而进一步推动了数据库技术的发展。

1.5 数据库技术的研究领域及其发展

- 1、数据模型的发展和三代数据库系统
- 数据模型是数据库系统的核心和基础，数据模型的发展带动着数据库系统不断更新换代。
- 数据模型的发展可以分为三个阶段：
 - 第一阶段为格式化数据模型，包括层次数据模型和网状数据模型
 - 第二阶段为关系数据模型
 - 第三阶段则是以面向对象数据模型为代表的非传统数据模型

1.5 数据库技术的研究领域及其发展

- 据上述的数据模型三个发展阶段，数据库系统也可以相应地划分为三代：

第一代数据库系统为层次与网状数据库系统

第二代数据库系统为关系数据库系统

新一代数据库系统（即第三代）的发展呈现百花齐放的局面

1.5 数据库技术的研究领域及其发展

● 2、数据库技术与其他相关技术的结合

- 将数据库技术与其他相关技术相结合，是当代数据库技术发展的主要特征之一
- (1) 面向对象数据库系统
- 面向对象数据库系统是数据库技术与面向对象技术相结合的产物。

1.5 数据库技术的研究领域及其发展

- (2) 分布式数据库系统
- 分布式数据库系统是数据库技术与计算机网络技术相结合的产物。
- (3) 并行数据库系统
- 并行数据库系统就是在并行计算机上运行的具有并行处理能力的数据库系统，它是数据库技术与并行计算机技术相结合的产物。

1.5 数据库技术的研究领域及其发展

- (4) 多媒体数据库系统
- 多媒体数据库系统是数据库技术与多媒体技术相结合的产物。
- (5) 主动数据库系统
- 主动数据库系统是数据库技术与人工智能技术相结合的产物。
- (6) 模糊数据库系统
- 模糊数据库系统是数据库技术与模糊技术相结合的产物。

1.5 数据库技术的研究领域及其发展

- 3、数据库技术的新应用
- 数据库技术在不同领域中的应用，也导致了一些新型数据库系统的出现，这些应用领域往往无法直接使用传统数据库系统来管理和处理其中的数据对象。
 - (1) 数据仓库系统
 - (2) 工程数据库系统

1.6 小结

- 本章概述了数据库的基本概念，介绍了数据管理技术发展三个阶段及各自的优缺点，说明了数据库系统的优点。
- 数据模型是数据库系统的核心和基础。本章介绍了组成数据模型的三要素及其内涵、概念模型和三种主要的数据库模型。
- 概念模型也称信息模型，用于信息世界的建模，E-R模型是这类模型的典型代表，E-R方法简单、清晰，应用十分广泛。数据模型包括非关系模型（层次模型和网状模型）和关系模型。本章简要地讲解了层次模型、网状模型和关系模型，而关系模型将在后续章节中作更详细的介绍。

1.6 小结

- 数据库系统的结构包括三级模式和两层映象。数据库系统三级模式和两层映象的系统结构保证了数据库系统中能够具有较高的逻辑独立性和物理独立性。数据库系统不仅是一个计算机系统，而且是一个人-机系统，人的作用特别是DBA的作用最为重要。

习题

一、选择题

- 1、 () 是位于用户与操作系统之间的一层数据管理软件。数据库在建立、使用和维护时由其统一管理、统一控制。
A. DBMS B. DB C. DBS D. DBA
- 2、文字、图形、图像、声音、学生的档案记录、货物的运输情况等，这些都是 ()
A. DATA B. DBS C. DB D. 其他
- 3、目前 () 数据库系统已逐渐淘汰了网状数据库和层次数据库，成为当今最为流行的商用数据库系统。
A. 关系 B. 面向对象 C. 分布 D. 对象-关系
- 4、 () 是刻画一个数据模型性质最重要的方面。因此在数据库系统中，人们通常按它的类型来命名数据模型。
A. 数据结构 B. 数据操纵 C. 完整性约束 D. 数据联系
- 5、 () 属于信息世界的模型，实际上是现实世界到机器世界的一个中间层次。
A. 数据模型 B. 概念模型 C. 非关系模型 D. 关系模型



习题

- 6、当数据库的()改变了，由数据库管理员对()映像作相应改变，可以使()保持不变，从而保证了数据的物理独立性。
- (1)模式 (2) 存储结构 (3)外模式/模式 (4)用户模式 (5)模式/内模式
- A. (1)和(3)和(4) B. (1)和(5)和(3)
C. (2)和(5)和(1) D. (1)和(2)和(4)
- 7、数据库的三级体系结构即子模式、模式与内模式是对()的三个抽象级别。
- A. 信息世界 B. 数据库系统
C. 数据 D. 数据库管理系统
- 8、英文缩写DBA代表()
- A. 数据库管理员 B. 数据库管理系统
C. 数据定义语言 D. 数据操纵语言
- 9、模式和内模式()。
- A. 只能各有一个 B. 最多只能有一个
C. 至少两个 D. 可以有多个



习题

10、在数据库中存储的是()。

- A. 数据 B. 信息 C. 数据和数据之间的联系 D. 数据模型的定义

二、填空题

- 1、数据库就是长期储存在计算机内_____、_____的数据集合。
- 2、数据管理技术已经历了人工管理阶段、_____和_____三个发展阶段。
- 3、数据模型通常都是由_____、_____和_____三个要素组成。
- 4、数据库系统的主要特点：_____、数据冗余度小、具有较高的数据程序独立性、具有统一的数据控制功能等。
- 5、用二维表结构表示实体以及实体间联系的数据模型称为_____数据模型。
- 6、在数据库的三级模式体系结构中，外模式与模式之间的映象，实现了数据库的_____独立性。
- 7、数据库系统是以_____为中心的系统。
- 8、E-R图表示的概念模型比_____更一般、更抽象、更接近现实世界。

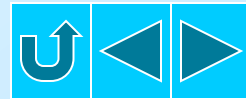


习题

- 9、外模式，亦称为子模式或用户模式，是_____能够看到和使用的局部数据的逻辑结构和特征的描述。
- 10、数据库系统的软件主要包括支持_____运行的操作系统以及_____本身。

三、简答题

- 1、简述计算机数据管理技术发展的三个阶段。
- 2、常用的三种数据模型的数据结构各有什么特点？
- 3、试述数据库系统的特点。
- 4、试述数据模型的概念、数据模型的作用和数据模型的三要素。
- 5、试述概念模型的作用。
- 6、定义并理解概念模型中的以下术语：
实体、实体型、实体集、属性、码、实体联系图（E-R图）、三种联系类型
- 7、学校有若干个系，每个系有若干班级和教研室，每个教研室有若干教师，每个教师只教一门课，每门课可由多个教师教；每个班有若干学生，每个学生选修若干课程，每门课程可由若干学生选修。请用E-R图画出该学校的概念模型，注明联系类型。



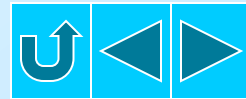
习题

- 8、每种工厂生产的产品由不同的零件组成，有的零件可用于不同的产品。这些零件由不同的原材料制成，不同的零件所用的材料可以相同。一个仓库存放多种产品，一种产品存放在一个仓库中。零件按所属的不同产品分别放在仓库中，原材料按照类别放在若干仓库中（不跨仓库存放）。请用E-R图画出此关于产品，零件，材料，仓库的概念模型，注明联系类型。
- 9、分别给出一个层次、网状和关系模型的实例。
- 10、试述层次、网状和关系数据库的优缺点。
- 11、定义并理解关系模型中的以下术语：
关系、元组、属性、主码、域、分量、关系模式
- 12、数据库系统的三级模式结构是什么？为什么要采用这样的结构？
- 13、数据独立性包括哪两个方面，含义分别是什么？
- 14、数据库管理系统有哪些主要功能？
- 15、数据库管理系统通常由哪几部分组成？



习题

- 16、设学生含有学号、姓名、性别、系别、选修课程、平均成绩（是经计算得到的）等信息，若把选修课程、平均成绩也作为学生的属性，请用E-R图表示学生信息。
- 17、请用扩展E-R图来表示客户的相关信息。客户含有姓名（由姓和名两部分组成）、客户号、电话（客户一般有多个电话）、出生日期、年龄、地址（含省、市、街道名、街道号、楼道号、房号与邮政编码）。
- 18、请用扩展E-R图中的实体角色的表示方法来分别表示：学生与学生间的班长关系与课程之间的先修关系等。
- 19、请用扩展E-R图的表示方法来表示公司生产特有产品的关系，其中公司含有公司名、地址、联系电话等信息，产品含有产品名与价格等信息。



习题

- 20、请用扩展E-R图的表示方法来表示如下信息：人有姓名、所在城市等信息；人可分成客户（含购买金额）和雇员（有雇员号、薪水信息等）两类；雇员可再分为办公人员（有累计办公天数信息）、车间职员（有累计产生产品数量信息）、销售人员（有累计销售数量与销售金额信息）等三类。
- 21、请用扩展E-R图的聚集表示方法来表示如下信息：制造商有名称与地址信息、批发商有名称与地址信息、产品有产品名与价格信息。制造商与批发商联营产品的销售有个具体的时间，共同联营的产品才共同负责发送。
- 22、设要为医院的核心主题建立概念模型E-R图，包括病人、医生及病人的病历。其中医生有姓名与专长信息；病人有姓名、身份证号、保险号等信息；病历有门诊日期、诊断病症、治疗方法及治疗结果等信息。病人找医生治疗有个具体的时间。

