

高职高专计算机系列教材

# 数据库原理与应用

## ——灾备建设案例

高春玲摇张文学摇白桂梅摇等编著

中国铁道出版社

北京·月

## 内 容 简 介

本书介绍数据库基本原理和数据库管理系统软件 ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 具体内容包括数据库基本概念、关系数据库理论、数据库设计、数据库保护、~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 数据库和表的操作、视图与查询、~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 面向对象程序设计、表单和菜单设计、报表和标签设计、应用程序集成和发布。本书从实用性和先进性出发,内容组织合理、通俗易懂。

本书可作为高职高专计算机专业“数据库原理与应用”课程的教材,也可作为各类成人教育计算机专业的教材或作为从事数据库应用开发技术人员的参考资料。

摇摇未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。

摇摇版权所有,翻版必究。

摇摇图书在版编目(CIP)数据

数据库原理与应用——~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 高春玲等编著 北京:电子工业出版社, ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 高职高专计算机系列教材  
~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~

I 数...摇 II 高...摇 III 援关系数据库 原数据库管理系统, ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 原高等学校 原教材材摇  
IV 援 ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~

中国版本图书馆 CIP 数据核字( ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ ) 第 ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 号

丛 书 名: 高职高专计算机系列教材

书摇摇名: 数据库原理与应用——~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~

编 著 者: 高春玲摇张文学摇白桂梅摇等

责任编辑: 张孟玮

特约编辑: 吴浩源

排版制作: 电子工业出版社计算机排版室

印 刷 者: 北京东光印刷厂

装 订 者: 三河市新伟装订厂

出版发行: 电子工业出版社 ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 编辑部  
北京市海淀区万寿路 ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 信箱摇邮编 ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~

经摇摇销: 各地新华书店

开摇摇本: ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 摇印张: ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 摇字数: ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 摇

版摇摇次: ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 年 苑月第 ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 版摇 ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 年 苑月第 ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 次印刷

书摇摇号: ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 摇

~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 摇

裁 ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 摇

印摇摇数: ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 摇册摇摇定价: ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 摇元

凡购买电子工业出版社的图书,如有缺页、倒页、脱页、所附磁盘或光盘有问题者,请向购买书店调换;

若书店售缺,请与本社发行部联系调换。电话摇 ~~灾~~ ~~害~~ ~~与~~ ~~操~~ ~~作~~ ~~原~~ ~~理~~ ~~与~~ ~~应~~ ~~用~~ 摇

## 前 摇 摇 言

本书是中国计算机学会高职高专教育学会《高职高专计算机系列教材》之一。符合教育部最新制定的高职高专计算机专业“数据库原理与应用”课程的教学要求,可以作为高职高专计算机及相关专业“数据库原理与应用”课程的教材。

本书编写的出发点是将数据库基本原理和具体数据库管理系统的介绍合二为一。使读者通过学习该教材,既掌握数据库应用系统开发所需的数据库理论知识,又能使用一个实际的数据库管理系统进行小型数据库应用系统的开发。

本书的编写以“理论必需够用,侧重实际应用”为原则。在内容选择上突出实用性,主要介绍实际数据库应用开发所必要的知识。在数据库基本原理部分,重点介绍关系数据库理论和数据库设计。在实际数据库管理系统部分,介绍 ~~数据库管理系统~~ ~~数据库管理系统~~ 的操作及其应用程序设计。~~数据库管理系统~~ ~~数据库管理系统~~ 是目前应用最广泛的微机数据库管理系统及应用开发平台,在微机数据库管理系统中具有代表性。

本书的编写在内容组织和安排上注重合理性,以实际应用系统的开发过程为主线,循序渐进地安排各个章节。在介绍 ~~数据库管理系统~~ ~~数据库管理系统~~ 的操作和应用中,以一个应用实例贯穿始终,理论和实例结合,易于理解和掌握。随着计算机行业的发展,许多术语均有变动,在本书编写中我们注意使用规范的术语和简明的文字,使本书内容通俗易懂。但是,窗口图中的用词由于历史的原因,且不易修改,容易与规范用语发生不一致现象,请读者鉴谅。

本书由高春玲任主编,张文学、白桂梅任副主编。第 1 章猿源章由白桂梅编写,第 2 章猿源章由高春玲编写,第 3 章猿愿怨章由张文学编写,第 4 章猿章由王琼编写,第 5 章猿章由张新颜编写,第 6 章猿章由马军编写。朱乃立担任本书主审。

在本书编写中,朱乃立老师提出了许多宝贵的指导意见,电子工业出版社也给予大力的支持和帮助,在此表示衷心感谢。

由于编者水平有限,加之时间仓促,书中难免存在缺点和错误,敬请读者和专家批评指正。

编著者

圆园园年 源月

# 第 1 章 数据库基本概念

## 1.1 数据库管理技术的发展

数据处理是对各种形式的数据进行收集、组织、加工、存储、传播等工作。日常工作事务中包含大量的数据，人们只抽取那些与所解决的问题有关联的数据，从中加工整理出反映实际问题情况的数据。数据管理是指对数据的组织、存储、检索和维护等工作。数据管理是数据处理的基本环节。早期的数据处理是手工进行的，属于手工数据处理。计算机的发展，特别是高效率存储设备的出现，借助于计算机对数据进行处理，使得数据处理的规模和范围以及处理速度都有了很大的提高，这就是电子数据处理。

### 1.1.1 电子数据处理技术发展的三个阶段

随着计算机软件和硬件的发展，电子数据处理技术的发展经历了以下 3 个阶段。

#### 1. 程序管理方式阶段

电子数据管理技术发展的第一阶段（大约在 1960 年以前）是程序管理方式阶段。用户在编制程序时就要确定自己对数据的组织、管理方式。因此，不仅要在程序中表现处理算法，还要表现对数据的管理原则；不仅考虑数据的逻辑定义，还要考虑数据的物理特征。数据包含在程序中。这一阶段的背景是，计算机主要用于科学计算；硬件方面的状况是外存只有磁带、卡片、纸带，没有磁盘等直接存取的存储设备；软件系统中没有操作系统，没有管理数据的软件。该阶段数据处理的特点是：

- (1) 数据不保存；
- (2) 数据不能独立，它是程序的组成部分，即数据和程序结合为一个不可分割的整体；程序员对数据的存储结构、存取方法以及输入输出的格式拥有绝对的控制权，数据的修改即意味着程序的修改；
- (3) 数据是面向应用的，不同应用的数据之间是相互独立的、彼此无关的，即使两个不同应用涉及到相同数据，也必须各自定义。数据不仅高度冗余，而且不能共享。

#### 2. 文件系统阶段

这一阶段的背景是计算机不仅用于科学计算，还用于简单的数据处理；外存储器有了磁盘、磁鼓等直接存取的存储设备，软件系统中有了初级的操作系统，有了专门管理数据的软件——文件系统。

它的特点是：

- (1) 数据出现在文件中，即所谓的数据文件；
- (2) 在存储上，数据文件独立于应用程序(文件)，但逻辑上它仅供该应用程序使用；
- (3) 数据冗余大，因为数据文件在逻辑上它仍然是应用程序的一部分。

### 3. 数据库系统阶段

利用数据库系统方式实现数据处理，是在文件系统的基础上发展起来的最先进技术。数据不仅存储上独立于应用程序，而且逻辑上也独立于应用程序，因此数据能为各应用程序所共享。

数据库系统的目标是实现数据独立，对数据的操作简单化，解决数据冗余问题，实现数据共享并且解决由于数据共享而带来的数据安全性、完整性及并发控制等一系列问题。为实现这一目标，数据库的运行必须有一个软件来控制，这个软件就是数据库管理系统 DBMS (Data Base Management System)。

#### 1.1.2 数据库技术的发展

数据库技术萌芽于 20 世纪 60 年代中期，到 20 世纪 60 年代末 70 年代初出现了三个事件，标志着数据库技术日益成熟，并有了坚实的理论基础。

(1) 1969 年 IBM 公司研制、开发的数据库管理系统的商品化软件 IMS(Information Management System)，IMS 的数据模型是层次结构的。

(2) 美国数据语言协商会 CODASYL(Conference On Data System Language)下属的数据库任务组 DBTG(Data Base Task Group)对数据库方法进行系统的研究、讨论。于 20 世纪 60 年代末 70 年代初提出了若干报告，成为 DBTG 报告。DBTG 报告确定并且建立了数据库系统的许多概念、方法和技术。DBTG 所提议的方法是基于网状结构的，它是数据库网状模型的基础和典型代表。

(3) 1970 年 IBM 公司 San Jose 研究实验室的研究员 E.F.Codd 发表了题为“大型共享数据库数据的关系模型”论文，提出了数据库系统的关系模型，开创了数据库关系方法和关系数据库理论的研究，为关系数据库技术奠定了理论基础。由于 E.F.Codd 的杰出工作，他于 1981 年获得了 ACM 图灵奖。

20 世纪 70 年代，数据库技术有了很大发展。

首先，数据库方法，特别是 DBTG 方法和思想应用于各种计算机系统，出现许多商品化数据库系统。它们大都基于网状模型和层次模型。

其次，这些商用系统的运行使数据库系统日益广泛地应用到企业管理、交通运输、情报检索、军事指挥、政府管理和辅助决策等各个方面，深入到人类生产和生活的各个领域。数据库技术成为实现和优化信息系统的基本技术。

第三，关系方法的理论研究和软件系统的研制取得了很大的成果。IBM 公司 San Jose 研究实验室在 IBM 370 系列机上研究关系数据库系统 System R 获得成功(1974 年~1979 年)。1981 年 IBM 公司又宣布了具有 System R 全部特征的新的数据库软件产品 SQL/DS 问世。与此同时，美国加州大学柏克利分校也研制 INGRES 关系数据库实验系统。并紧接着出现了商用 INGRES 系统，使关系方法从实验室走向社会。因此，在计算机领域中，有人把 20 世纪 70 年代称为数据库时代是不无道理的。20 世纪 80 年代，几乎所有的新开发的数据库系统均是关系数据库系统，微型机的关系数据库系统也越来越丰富，性能越来越好，功能越来越强，它的应用遍及各个领域。

## 1.2 数据库系统的组成

数据库系统是由数据库、数据库系统的系统软件（数据库管理系统 DBMS）、支持数据库运行的软、硬件环境、应用程序和数据库管理员等组成的。

### 1. 数据库

数据库是一组相互联系的若干文件的集合，其中最基本的是包含用户数据的文件（通常称为主文件）。用户所需的数据，按逻辑分类存储于数据库文件中，文件之间的联系是由它们之间的逻辑关系决定的，这种联系也要存储于数据库文件中。

### 2. 数据库管理系统 DBMS

在数据库系统中，DBMS 是专门用于数据管理的软件。通常，DBMS 提供数据库定义和数据装入功能，提供数据操纵（包括检索与数据存取操作）、数据控制（包括安全性、完整性和并发控制）和数据库维护（包括数据库整理、修改与重定义等）等功能。

DBMS 提供应用程序与数据库的接口，允许用户逻辑地访问数据库中的数据，负责逻辑数据与物理地址之间的映射，是控制和管理数据库运行的工具。

### 3. 支持数据库运行的软、硬件环境

每种数据库管理系统（DBMS）都有它自己要求的软、硬件环境。一般对硬件需要说明所需的基本配置以及所建议的配置（较高配置），对软件则要说明其适应于哪些底层软件，与哪些软件兼容等。

### 4. 应用程序

数据库是多用户共享的，不同用户有不同业务，他们所需要的数据，已由数据库设计者精心规划在数据库中。但是如何使用数据库中的数据是用户的事，需要自编应用程序处理自己的业务，其操作范围仅是数据库的一个子集（由用户所需的那部分数据组成）。

### 5. 数据库管理员 DBA(Data Base Administrator)和用户

顾名思义，数据库管理员 DBA 是管理、维护数据库系统的人员，起着联络数据库系统与用户的作用；用户则是使用数据库系统的人员。大型数据库系统，一般配备专职 DBA，微型机数据库系统，DBA 一般由用户自己承担。

数据库是整个数据库系统的数据资源，它们要由专门的机构来管理，DBA 就是这一机构中的人员，DBA 的具体职责包括如下 5 个方面。

#### (1) 决定数据库的内容与结构

为了决定数据库的内容与结构，DBA 必须参与整个数据库的设计过程，从应用出发，提取实际问题所涉及的实体及其相互联系（实际问题所涉及的全部数据即数据库的内容），用模式 DDL（Data Description Language——数据描述语言）给出数据库模式的描述（数据库结构的描述）。这一过程就是 DBA 与用户交流的过程，了解、分析用户的所需，从而决定数据库的内容与结构。

## (2) 决定数据库的存储结构和存取策略

DBA 要从整体上分析各个用户对数据的需求,与数据库设计人员一道共同决定数据库的存储结构和存取策略,使得对数据的存储空间效率、存取时间效率都能达到一个较好的水准。

## (3) 实施数据库系统的保护

数据库中数据被多个用户共享的同时,必须考虑数据库系统的安全性问题。不同用户对不同的数据应具有不同的操作权限;为使数据的正确性与相容性得到一定的保障,对数据库系统中的数据应当设置完整性约束条件。DBA 负责实施用户权限、完整性约束条件等的设置。

## (4) 监督和控制数据库的使用和运行

DBA 负责监督数据库的运行,尽量满足用户在使用数据库时提出的要求,解决运行过程中出现的问题。一旦数据库系统出现问题,DBA 能够分析问题所在;出现严重问题或数据库遭到破坏时,需要根据定期转储的数据、日志文件恢复系统。当然,要做到这些,DBA 必须要有过硬的业务能力。

## (5) 改进与重组数据库系统

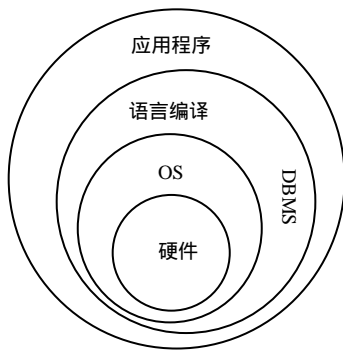


图 1.1 软件层次关系

DBA 监督数据库的运行,用户对数据库中数据以及文件的不断插入、修改与删除,必然导致数据库系统的使用效率下降,这就要求 DBA 阶段性地对数据库系统进行重组。记录数据库系统的运行情况,分析、统计其性能,必要时对系统进行改进。

图 1.1 是数据库管理系统与硬件及其他软件的层次关系,外层依赖于内层,内层支持外层。

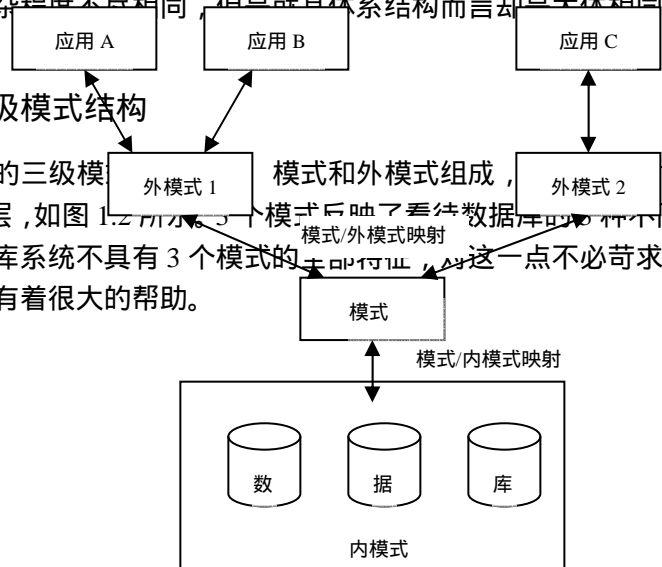
综上所述,数据库文件包含用户所需的数据,是存储在磁鼓、磁盘或其他存储介质上的数据集合;每个用户只享用其中一部分数据,不同用户所使用的数据可以重叠,并且同一片数据可以为多个用户共享;应用程序以数据库文件为背景进行检索、修改、插入或删除等操作;DBMS 向用户提供对数据的存储组织、数据操作的方便界面,它还提供有保护数据库的功能。本书所讲述的 Visual FoxPro 即是一种关系型数据库管理系统(DBMS),它实现了操作界面的可视化,提供的界面更是友好、直观。

## 1.3 数据库体系结构

为了有效地组织、管理数据,人们为数据库设计了一个严谨的体系结构。目前,尽管数据库的种类和复杂程度不尽相同,但其体系结构而言却是大体相同的,包括 3 个模式和 2 个映射。

### 1.3.1 三级模式结构

数据库系统的三级模式结构由外模式、模式和内模式组成,如图 1.3 所示。外模式是用户视图,模式反映了看待数据库的不同数据观点。当然,微型机上的数据库系统不具有 3 个模式的结构,这一点不必苛求,重要的是它对深刻理解数据库系统有着很大的帮助。



## 1. 模式

模式是数据库中全部数据的一个逻辑表示。它既不同于内模式，比内模式抽象，不涉及数据的物理存储细节；也不同于外模式，与具体的应用程序及使用的高级程序设计语言无关。模式是数据库数据在逻辑级上的视图，它是一种对数据库组织的全局逻辑观点，反映了数据的总体观。

数据库系统提供数据模式描述语言（模式 DDL——Data Description Language）来定义数据的模式，数据模式描述语言又称为数据定义语言，模式是对数据库结构的一种描述，它是装填数据的框架。

## 2. 内模式

内模式又称为存储模式，具体描述数据在外部存储器上如何组织存储。它涉及数据库文件中数据的存储方式，可用顺序表形式存储，也可以用哈希（hash）表方式，还可以是线性链表或树型结构。内模式由内模式描述语言（内模式 DDL）来定义。内模式反映数据库的存储观。

## 3. 外模式

外模式是个别用户的数据视图，与某一应用有关的数据的逻辑表示。外模式是模式的一个子集，所以又称外模式是子模式，它包含模式中允许特定用户使用的部分数据。外模式由外模式描述语言（外模式 DDL）来描述，外模式 DDL 和用户选用的程序设计语言具有相容的语法。外模式反映了数据库的用户观。

三个模式中，模式是内模式的逻辑表示，内模式是模式的物理实现，外模式则是模式的部分抽取。总体观和存储观只有一个，而用户观可能有多个，有一个应用就有一个用户观。

### 1.3.2 三级模式之间的映射

三个模式间有着两种映射：模式/外模式映射、模式/内模式映射。模式/外模式这种映射定义某个外模式和模式的对应关系，当模式发生改变时，只要改变其映射，就可以使外模式保持不变（这是 DBA 的责任）。另一映射模式/内模式，定义数据的逻辑结构与存储结构之间的对应关系，数据的存储结构发生变化时，改变模式/内模式映射，就能使模式不变。正是通过这两种映射，换来用户使用数据库的方便，最终把用户对数据库的逻辑操作转换成对数据库的物理操作。用户不必关心数据库全局，更不必关心物理数据库，它们之间的映射是由 DBMS 实现的。用户看到的只是其外模式。

## 1.4 数据库管理系统

数据库管理系统是数据库系统的系统软件，能够实现对数据库的建立、使用和维护。对于各种数据库命令和程序中各条语句，数据库管理系统都将其转换成对系统存储文件的操作，把用户对数据的一次访问，从外模式带到模式，再导向内模式，DBMS 实现数据库三级模式之间的转换。数据库管理系统的主要功能包括以下 3 个方面。

## 1.4.1 数据定义和操纵功能

### 1. 数据定义功能

DBMS 一般提供数据定义（描述）语言（DDL）定义外模式、模式和内模式。各种模式翻译程序把用 DDL 书写的各种源模式翻译为相应的内部表示，分别称为目标外模式、目标模式和目标内模式，如图 1.3 所示。这些目标模式是对数据库的描述，而不是数据本身。它们是刻画数据库的框架，被保存在数据字典（亦称系统目录）中，是 DBMS 存取和管理数据的依据。具体 DBMS 包括哪些 DDL，差别很大，如关系式 DBMS 没有内模式 DDL，还有的 DBMS 不一定将三种 DDL 分开，而是提供一二种 DDL 来完成三种模式的描述。

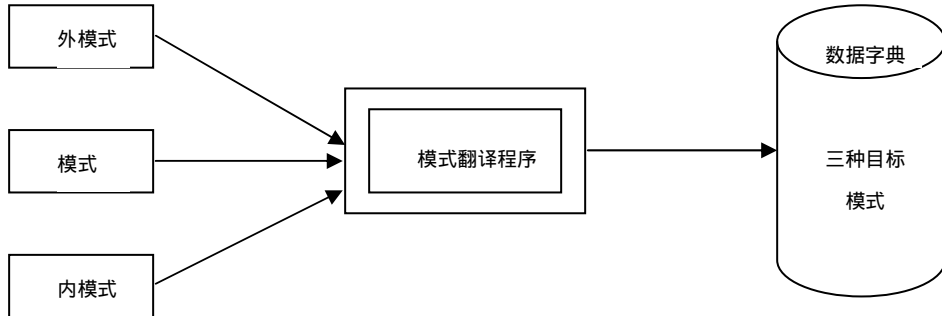


图 1.3 模式转换为目标模式

## 2. 数据操纵功能

DBMS 提供数据操纵语言 (Data Manipulation Language 简称 DML) 实现对数据库的基本操作: 检索、插入、修改和删除等各种操作命令。DML 有两类: 一类是嵌入主语言中的, 如嵌入 Cobol, FORTRAN, C 等高级语言中, 这类 DML 语言本身不能独立使用, 因此称为宿主型语言; 另一类是交互式命令语言, 它们语法简单, 可以独立使用, 所以称为自含语言。

DBMS 控制并执行 DML 语句, 完成对数据库的操作。对于交互式命令语言, DBMS 通常用解释执行的方法。对于宿主型语言, DBMS 一般提供以下两种编译方法。

(1) 预编译方法: DBMS 提供预处理程序, 对源程序进行扫描, 识别 DML 语句并把它们转换成主语言调用语句, 以便原来的编译程序接受和执行它们, 处理过程如图 1.4 所示。

(2) 增强编译方法: 扩充主语言编译程序, 对其进行补充, 使它能够编译数据库 DML。

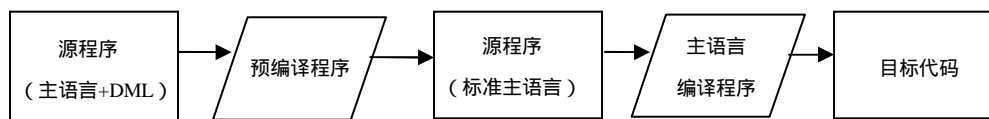


图 1.4 预编译方法

### 1.4.2 数据库运行管理

这是 DBMS 运行时的核心部分, 包括并发控制、存取控制 (即存取权限的检查)、完整性约束条件的检查和执行、数据库内部的维护 (如索引、数据字典的自动维护) 等等。所有数据库的操作都要在这些控制程序的统一管理下进行, 以保证运行的正确执行, 保证数据库的正确有效。

DBMS 的功能随不同系统而有所差异, 通常大型系统功能较强、较全, 小型系统功能较弱。例如, 目前微机上的许多 DBMS 就没有存取控制功能, 而网络环境下的数据库管理系统具有存取控制以及并发控制功能。

### 1.4.3 用户与数据库接口

用户可以通过两种方式使用数据库: 交互方式、应用程序方式。

交互方式直观、使用简单, 通常是借助于 DBMS 的查询语言个别地对数据库中的数据进行操作。应用程序方式则是用户依据外模式编写程序模块, 以实现数据库中的数据各种复杂的操作。

## 习 题 一

1. 何谓数据处理? 电子数据处理经历了哪三个阶段? 文件系统阶段与数据库系统阶段有什么不同?
2. 何谓数据的独立性?
3. 数据库系统由哪几部分组成? 哪部分属于数据库系统的系统软件?
4. DBA 的职责是什么?
5. 试画出数据库体系结构图。

6. 数据库体系结构的三个模式反映了哪三种数据观？两层映射各完成什么模式间的转换？这种映射转换给用户带来什么好处？
7. 用户可以通过哪几种方式使用数据库？数据库各种数据操作命令是否能够以交互方式使用，是否能够以应用程序方式使用？程序控制语句是否能够以交互方式使用？
8. 数据库管理系统（DBMS）包含哪些功能？
9. 外模式、模式、内模式分别对应哪种 DDL？何处存放三个目标模式？
10. 何谓宿主语言？自含语言？
11. 宿主语言有哪两种编译方式？它们的区别是什么？

# 第2章 数据模型

概念模型是信息世界中信息的描述。数据模型是计算机世界中对数据的描述。数据模型有网状、层次、关系模型三种类型。本章首先讨论概念模型与数据模型，然后再讨论怎样从概念模型导出数据模型。

## 2.1 实体及其联系

借助于计算机解决现实世界的问题，首先是对问题的直观印象：要解决的问题包含哪些事物，这些事物都具有什么性质。然后人们从中提取关于解决问题所必需的事物，以及解决问题所涉及的事物特性——这就是进入信息世界的活动，即通过描述、分类等抽象加工处理来自现实世界的信息。现实世界的事物及其性质对应信息世界的实体及其属性，并进一步研究与事物间联系相对应的实体间联系。

### 2.1.1 实体及其属性

#### 1. 实体

现实世界任何可被认识、区分的事物叫实体。实体可以是人或物，可以是实际的对象，也可以是抽象的概念，比如一个人，一个学生，一门课，一项设计，一个规划等等。

#### 2. 属性

实体所具有的特性叫做属性，一个实体可用若干属性来刻画。例如，学生实体可以由学号、姓名、年龄、性别、系、年级等属性组成，“99032，王大为，19，男，电子系，三年级”这些属性值组合起来描述一个学生。

对于一个对象，它究竟具有哪些属性？这完全取决于考虑问题的角度。当规定了不同属性时，便成为两个不同实体。例如一个人，在人口统计中，关心他的姓名、年龄、性别、文化程度、婚姻状况等属性；而在工资处理中，关心他的工资、奖金、水电费、房租等方面的属性。选择什么属性要根据处理问题的需要而确定。

#### 3. 实体和属性的型与值

实体和属性有型与值之分。型是结构，具有相同属性的实体，称它们具有相同的实体型。对于实体型，用实体名及其属性的集合来描述。例如，工人（姓名，年龄，性别，工种，工龄，技术级别，文化程度）是一个实体型。但是不同的个体有着不同的属性内容（即属性值），如：工人（李明明，23，女，车工，4，2，高中）就是一个具体的实体值。一个实体型对应若干该型的值，这些不同值的实体构成该实体型的实体集。实体型、实体值经常被笼统地说成实体，至于究竟是指型还是值，可以通过上下文叙述明了其含义。

类似地，属性也有型与值之分，如“姓名”是属性的型，李明明是属性值，如表 2.1

所示。

表 2.1 实体和属性的型与值

学号	姓名	年龄	性别	系	年级
99032	王大为	19	男	电子系	2
99097	李明明	21	女	机械	2
98035	陈惠	19	女	机械	3
00159	赵示杰	18	男	外语	1

#### 4. 关键字

在实体属性中，用于区别实体集合中不同个体的某个属性或某几个属性（最少属性）的组合，称为关键字。表 2.1 所示实体集合中，能做关键字的属性是学号，因为一旦关键字有一取值，便惟一地标识了集合中的某一实体。若无重名，姓名也可以做关键字；而年龄、性别等属性都没有这种特征，不能做关键字。所以，关键字是用来标识实体的。

#### 5. 信息转化为数据

实体、属性是信息世界的术语，它与计算机世界的数据库模型术语有着对应关系，可用图 2.1 形象地表示。

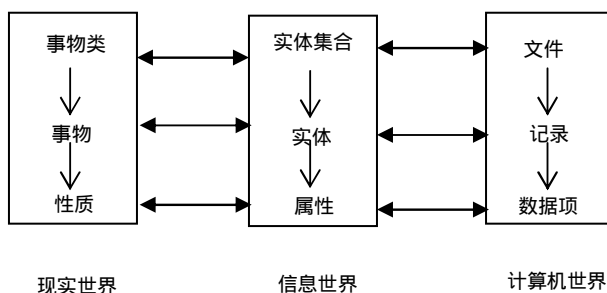


图 2.1 三个世界术语对照

### 2.1.2 实体间的联系

现实世界的事物是有联系的，这种联系必然要在信息世界中反映。一般存在两类联系：一是实体之间的联系，另一是实体内部的联系，如组成实体的属性之间的联系。

#### 1. 实体间的联系方式

##### (1) 一对一联系(1:1)

若两个不同型实体集合中，任一方的一个实体只与另一方的一个实体相对应，则这种联系称为一对一联系，记作 1:1。

如生产车间与车间主任两实体间的联系，每个生产车间有一个车间主任，反过来，每个

车间主任又对应一个生产车间。

(2) 一对多联系(1:n)

若两个不同型实体集合中，一方一个实体对应另一方若干实体；反之，另一方一个实体只对应本方一个实体，则这种联系称为一对多联系，记作 1:n。

如车间主任与本车间工人间的联系，一个车间主任对应若干工人，反过来，每个工人只对应一个车间主任。

(3) 多对多联系(m:n)

若两个不同型实体集合中，两实体集中任一实体，都与另一实体集合中一个或多个实体相对应，则这种联系称为多对多联系，记作 m:n。

如教师与学生之间的联系，一位教师对应若干学生，一个学生也对应若干上课的教师。

图 2.2 表示了实体间的以上三种联系。

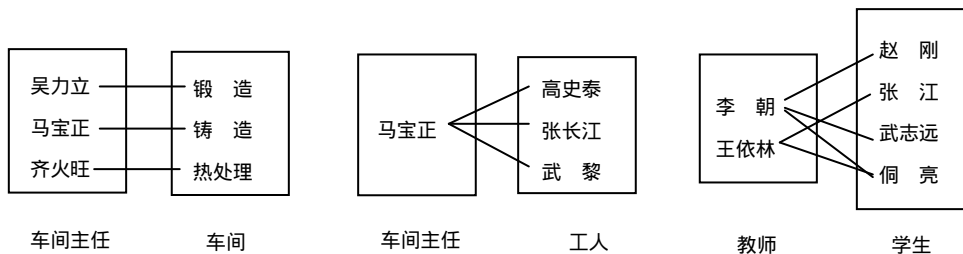


图 2.2 实体间的联系

2. 同一实体集内实体间的联系

同一实体集内实体之间也可以有某种联系。参见图 2.5，考察职工与领导的联系，领导是职工实体集中的一员，某一职工领导若干其他职工，一个职工仅被另一职工所领导，所以该联系属于实体集内的 1:n 联系。再如零部件之间的装配联系，一个零部件可由其他多个零（部）件组成，而该零部件又可以组装成其他多个（零）部件，所以该联系属于实体集内的 m:n 联系。

2.1.3 实体及实体联系的表示

所谓建立概念模型，就是用模型表示实体及实体间的联系，最常用的表示方法之一就是 E-R 方法，即用 E-R 图（Entity-Relation diagram）表示实体及实体间的联系。E-R 图的绘制，就是建立概念模型的过程。

(1) 用长方形表示实体型，在框内填上实体名。

(2) 用带圆弧的长方形表示实体的属性，实体框与属性框之间用直线连接。通常在 E-R 图中，表示单个实体时，除要标明实体名外，还要标明实体所具有的各个属性；而在表示多个实体之间的联系时，每个实体只需标明其实体名即可。

(3) 用菱形表示实体间的联系，框内填上联系名，用直线把菱形与相关实体相连接，在直线旁标明联系的类型，即标明联系是 1:1 还是 1:n 或者是 m:n。

根据上述约定，图 2.3 是工人这一实体的 E-R 图表示，图 2.2 所示不同类型的实体联系。其对应的 E-R 图如 2.4 所示。图 2.5 是同一实体集内实体间联系的 E-R 图。

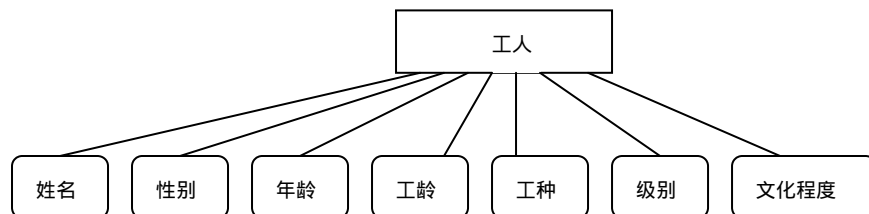


图 2.3 实体及其属性

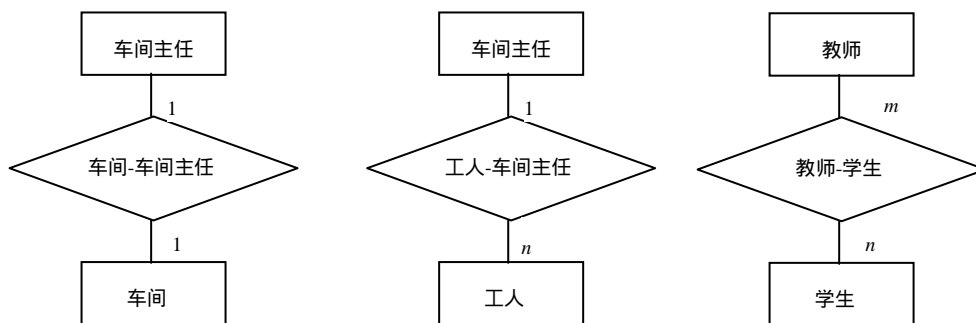


图 2.4 实体间联系的表示

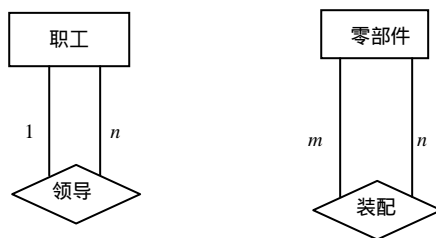


图 2.5 实体集内的 1:n、m:n 联系

## 2.2 数据模型与数据库类型

概念模型 (E-R 图) 是对信息世界的模型表示, 而数据模型是对计算机世界的模型表示。数据库类型是由采用的数据模型的类型决定的。

### 2.2.1 数据模型

数据库的数据结构形式叫数据模型, 它是对数据库中数据如何组织的一种模型化表示。显然, 如果这个模型只能表示存储些什么信息, 那么它是简单的, 是文件系统早已解决的问题。问题的关键在于数据库不仅要表示存储些什么信息, 更重要的是要以一定的数据结构方式表示各种数据间的联系。

上一节介绍了信息世界中的实体以及实体之间的联系。对于实体、实体间的联系反映到计算机世界中就是数据模型。数据库按模式、外模式、内模式三个层次进行组织。模式是源于数据模型的, 它是用 DBMS 所提供的 DDL 对数据模型的描述。数据模型可分为网状模型、

层次模型与关系模型，这三种模型就是在“数据结构”课程中熟知的：图、树和表三种数据结构。

## 2.2.2 数据库类型

数据库的类型依据数据模型的类型分为：

- 层次模型
- 网状模型
- 关系模型

相应的数据库类型为：

- 层次数据库
- 网状数据库
- 关系数据库

前二种数据模型，统称非关系模型。非关系模型用“有向图”表示，然后直接从有向图出发定义数据库；关系模型用“二维表”表示。关系模型的结构简单，使用普遍，是本书的重点。

### 1. 层次模型

层次模型中实体及其联系表现为树结构。树的基本特点是：

- 有且仅有一个结点无双亲，该结点称根；
- 其他结点有且仅有一个双亲结点。

在层次模型中，同一双亲的子女结点称为兄弟结点，无子女的结点称为叶子结点。图 2.6 中，N1 是根，N2、N3、N4 互为兄弟，是 N1 的子女结点；同样，N5、N6 互为兄弟，是 N3 的子女结点；N2、N4、N5、N6 是叶子结点。其中每个结点代表一个实体型，各实体型由上而下是 1:n 联系。

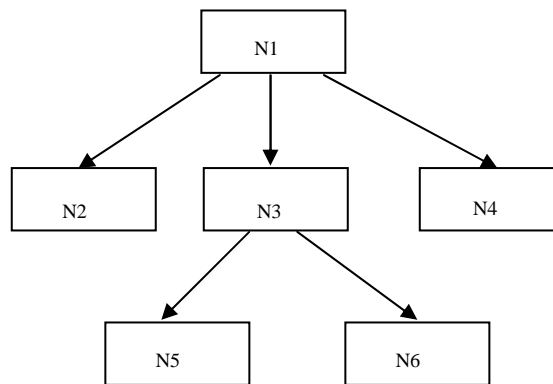


图 2.6 层次模型

在层次数据库中查找记录，必须指定查找记录的存取路径。路径即是从根开始到查找记录所在结点沿途经过的结点序列。

### 2. 网状模型

网状模型的结点间可以任意发生联系，能够表示各种复杂的联系。图 2.7 是网状模型的几个例子。在网状模型中：

- (1) 有一个以上结点无双亲；
- (2) 至少有一个结点有多于一个的双亲。

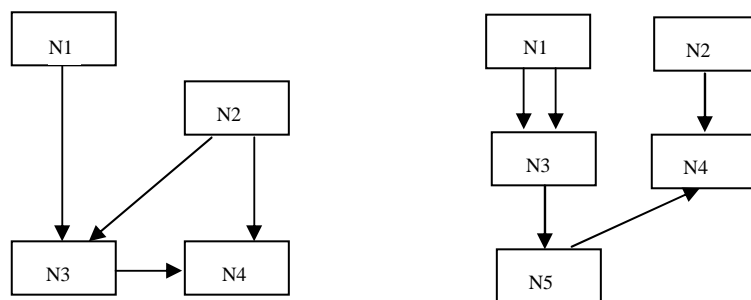


图 2.7 网状模型

网状模型反映的是  $m:n$  联系。对于网状模型数据库，它只能支持基本的  $1:n$  联系，对于  $m:n$  联系要进行演变，将其演变成两个  $1:n$  联系，这样才能被网状 DBMS 所接受。

关于网状模型，有以下 3 点需要特别注意：

- (1) 能反映现实世界各种复杂联系，对  $m:n$  联系经演变后能够支持；
- (2) 由于网络中一个结点多于一个双亲，因而使子女与双亲间的联系不惟一，因此对不同联系要分别命名加以区分（每个命名的联系，就是由  $m:n$  联系转换得到的  $1:n$  联系）；
- (3) 查找记录时，可以根据联系名从任一结点找到所需记录。

### 3. 关系模型

#### (1) 关系与关系模型

关系是数学名词，有严格的数学定义，第3章将深入讨论关系理论。现在暂且可把关系理解为二维表。所谓关系模型就是指用若干关系表示实体及其联系。有关关系模型的术语如下：

- 关系——一个关系对应一张二维表；
- 元组——表中的一行称为一个元组；
- 属性——表中的一列称为一个属性，给每一列起一个名称，该名称就是属性名；
- 域——属性的取值范围；
- 关键字——能够唯一地标识一个元组的属性或最小属性组；
- 分量——元组中的一个属性值；
- 关系模式——对关系的描述，用关系名(属性名 1, 属性名 2, ..., 属性名  $n$ )来表示。

#### (2) 用关系模式表示实体与实体间的联系

无论实体还是实体间的联系，在关系模型中都用关系模式来表示。

实体对应关系，实体名用关系名表示，实体的属性对应关系的属性。为方便起见，通常就取关系名为实体名，关系的属性名为实体的属性名。

假定有“教师”、“学生”两个实体，可用以下两个关系模式表示：