

# 第 1 章 数据库基础理论



随着计算机技术的发展和推广,数据库技术在管理信息系统或计算机应用技术中,已发展成为一个重要的学科分支。数据库技术是研究数据库结构、存储、设计和使用的—门软件科学,是进行数据管理和处理的技术。在信息社会里,信息已成为各行各业的重要财富资源,数据库技术已经成为信息基础设施的核心技术和重要基础。数据库技术作为数据管理的最有效的手段,极大地促进了计算机应用的发展。

本章通过数据的发展史、数据库的基本概念、数据模型、关系数据库、数据库系统的设计步骤等,让大家对数据库有个基本的了解,为后面各章的学习打下坚实的基础。



- 了解数据库的基本概念
- 掌握关系数据库的基本概念
- 掌握关系运算的类别
- 了解数据库设计的原则及步骤



- 数据、信息与数据处理
- 数据管理技术的发展
- 数据库的应用
- 数据库系统
- 数据模型
- 关系数据库
- 关系运算
- 数据库设计

## 1.1 数据库概述

数据库自20世纪60年代末产生至今,已经成为计算机科学的一个重要分支,目前大多数的计算机系统都离不开数据库技术,数据库技术被广泛地应用于各个领域。从字面的角度来看,“库”是储存东西的地方,“数据库”可以简单地理解为储存数据的地方。在了解数据库系统基本概念之前,先从数据管理技术的产生和发展过程来认识数据是如何进行处理的。从数据处理的演变过程,就不难看出数据库技术的历史地位和发展前景。

### 1.1.1 数据、信息与数据处理

#### 1. 信息

信息(information)是现实世界事物的存在方式或运动状态的反映。泛指通过各种方式传播、可被感受的声音、文字、图像、符号等表示的某一特定事物的消息、情报或知识。它具有可感知、可存储、可加工、可传递和可再生等特点。

#### 2. 数据

数据(data)是指存储在某一媒体上能识别的物理符号,是反映客观事物特性的记录。数据的概念包括两个方面:一是所反映的事物特性的数据内容;二是存储在某一种媒体上的数据形式。数据形式可以是多种多样的,如某人的出生日期是“1992年8月10日”,也可表示为“92/8/10”,其含义没有改变。

信息和数据二者之间是不可分离但又有一定区别的。一方面,并非任何数据都能表示信息,信息只是消化了的数据;另一方面,信息是更基本的直接反映现实的概念,而数据则是信息的具体表现,所以信息不随承载它的物理设备的改变而改变。数据具有任意性,用不同的数据可以表示同一信息,但在某些不需要严格区分的场合,可以把二者不加区分地使用。例如,可以说信息处理和信息管理,也可以说数据处理和数据管理。

#### 3. 数据处理

数据处理(Data Procession, DP)又称为信息处理。实际上就是利用计算机对各种类型的数据进行处理,并将其转换成信息的过程。它包括对数据的采集、整理、存储、分类、排序、检索、维护、加工、统计和传输等一系列的操作过程。数据处理的目的是从大量的、原始的数据中获得所需的资料并提取有用的数据成分,作为行为和决策的依据。

### 1.1.2 数据管理技术的发展

自从计算机应用于数据处理领域以来,就面临着如何管理大量复杂数据的问题。时至今日,随着计算机软硬件技术与数据管理手段的不断发展,数据处理过程发生了划时代的变革,数据管理技术已经大致经历了5个发展阶段。

## 1. 人工管理阶段

在 20 世纪 50 年代中期之前,计算机主要用于科学计算,当时的硬件和软件系统都很原始,没有专门管理数据的软件,也没有磁盘之类的可以随机访问、直接存取的外部存储设备,数据处理方式基本是批处理。数据和应用程序之间的关系如图 1-1 所示。

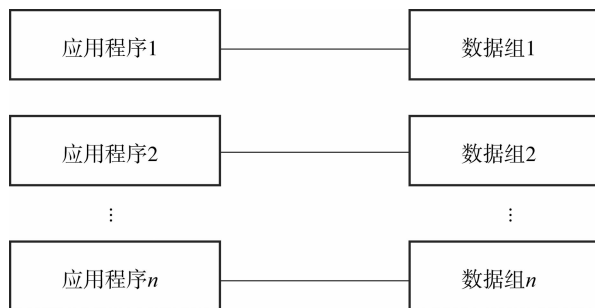


图 1-1 人工管理数据与程序的关系

这个时期,数据管理的特点如下。

①数据不保存。因为计算机主要用于科学计算,不要求保存数据。每次计算机先将程序和数据输入主存,计算结束后,将结果输出,计算机不保存程序和数据。

②编写程序时要确定数据的物理存储。程序员编写应用程序时,还要安排数据的物理存储。程序和数据混为一体,一旦数据的物理存储改变,必须要重新编程,程序员的工作量大,繁琐,程序难以维护。

③数据面向程序。每个程序都有属于自己的一组数据,程序与数据相互结合成为一体,互相依赖。各程序之间的数据不能共享,因此数据就会重复存储(冗余度大)。

## 2. 文件系统阶段

20 世纪 50 年代后期到 60 年代,计算机软、硬件技术有了飞速发展。在硬件方面,出现了能存储大量数据的磁鼓、磁盘;在软件方面,出现了高级语言和操作系统,或者说,操作系统提供了文件管理的功能。

文件系统是操作系统的高层部分。用户和应用程序通过文件系统,对文件中的数据进行处理和加工。此时,程序与数据有了一定的独立性,有了程序文件和数据文件之分,如图 1-2 所示。

和人工管理相比,文件系统的优点如下。

①数据以文件的形式长期保存。由于计算机大量用于数据处理,数据需要长期保留在外存上反复处置,即经常对其进行查询、修改、插入和删除等操作。因此,在文件系统中,按一定的规则将数据组织为一个文件,存放在外存储器中长期保存。

②尽管数据的物理结构与逻辑结构有了区别,但过于简单。程序员只需用文件名与数据打交道,不必关心数据的物理位置,可由文件系统提供的读写方法去读/写数据。

③文件形式多样化。为了方便数据的存储和查找,人们研究了许多文件类型,如索引文件、链接文件、顺序文件和倒排文件等。数据的存取基本上是以记录为单位的。

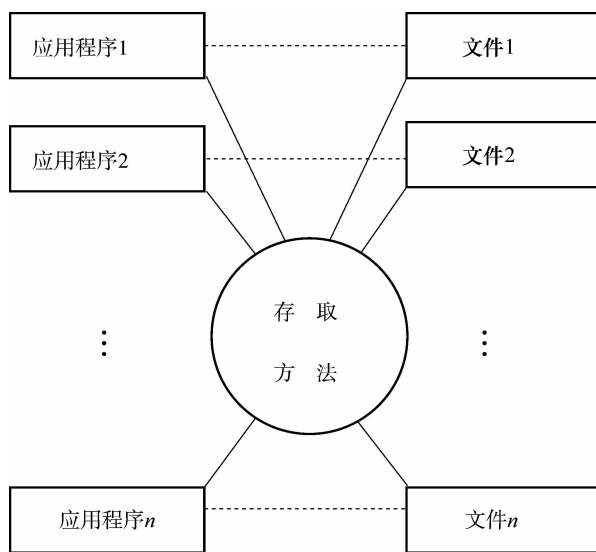


图 1-2 文件系统中程序和数据的关系

④程序与数据之间有一定的独立性。应用程序通过文件系统对数据文件中的数据进行存取和加工。因此,处理数据时,程序不必过多地考虑数据的物理存储的细节。文件系统充当应用程序和数据之间的一种接口,这样可使应用程序和数据都具有一定的独立性。这样,程序员可以集中精力于算法,而不必过多地考虑物理细节。并且,数据在存储上的改变不一定反映在程序上,这可以大大节省维护程序的工作量。

尽管文件系统有上述优点,但是,这些数据在数据文件中只是简单地存放,文件中的数据没有结构,文件之间并没有有机的联系,仍不能表示复杂的数据结构;数据的存放仍依赖于应用程序的使用方法,基本上是一个数据文件对应于一个或几个应用程序;数据面向应用,独立性较差,仍然出现数据重复存储,冗余度大,一致性差(同一数据在不同文件中的值不一样)等问题。

### 3. 数据库系统阶段

20 世纪 60 年代以来,计算机用于管理的规模越来越大,应用越来越广泛,硬件有了大容量磁盘,出现越来越多的系统软件,并且出现了统一管理数据的专门软件系统——数据库管理系统。

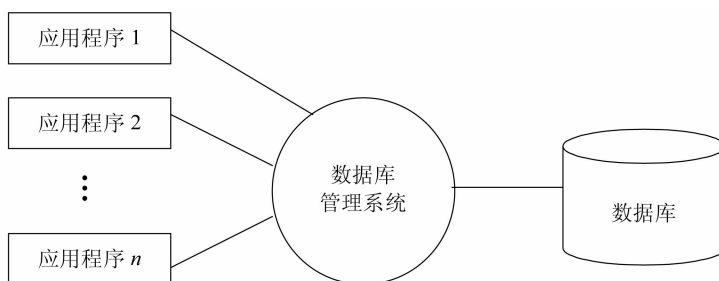
在计算机软件体系中,数据库管理系统建立在操作系统之上,程序员可以用它设计具体的数据库。从层次上看,数据库管理系统和数据库之间的关系如图 1-3 所示。

所以,数据库管理系统和数据库是两个不同层次的概念。数据库管理系统十分复杂,比较著名的有 Oracle、SQL Server、Sybase、Informix、DB2 等产品。程序员和一般的软件公司需要使用这些产品开发面向具体应用的数据库,如图书管理数据库、超市管理数据库等,这些数据库由 DBMS 管理。

数据库应用系统则是以数据库为基础和核心的计算机应用系统,是利用 DBMS 提供的编程语言或其他独立的编程语言开发的。



在数据库管理系统的支持下,数据与程序的关系如图 1-4 所示。



#### 4. 分布式数据库系统阶段

20 世纪 70 年代以前,数据库多数是集中式的,随着计算机网络技术的发展,使数据库从集中式发展到了分布式。分布式数据库把数据库分散存储在网络的多个结点上,彼此用通信线路连接。

网络技术的发展为数据库提供了分布运行环境,从主机——终端体系结构发展到客户/服务器系统结构。分布式数据库系统既可以把全局数据模式按数据来源和用途,合理分布在系统的多个结点上,使大部分数据可以就地存取,而用户觉察不到的分布,即物理上分布、逻辑上集中的分布式数据库结构;又可将多个集中式数据库系统通过网络连接起来,各结点上的计算机可以利用网络通信功能访问其他结点上的数据资源,即物理上、逻辑上分布的分布式数据库结构。

#### 5. 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程序设计相结合的产物。面向对象数据库系统是面向对象方法在数据库系统中的实现和应用,它既是一个面向对象的系统,又是一个数据库系统。面向对象数据库吸收了面向对象程序设计方法的核心概念和基本思想,采用面向对象的观点来描述现实世界实体(对象)的逻辑组织、对象之间的限制和联系等。

### 1.1.3 数据库的应用

数据库的应用领域十分广泛。传统数据库中很大一部分用于商务领域,现代数据库有很多新的应用领域。

#### 1. 多媒体数据库

数据库技术与多媒体技术相结合产生了多媒体数据库系统,这类数据库主要存储与多媒体相关的数据,如声音、图形、视频等数据。多媒体数据最大的特点是数据连续,数据量比较大,存储需要的空间较大。

#### 2. 移动数据库

数据库技术与移动通信技术相结合产生了移动数据库系统。移动数据库是在移动计算机系统上发展起来的,如3G手机的实时图像处理等。这类数据库最大的特点是通过无线网络传输。移动数据库可以随时随地获取和访问数据,为一些商务应用带来很大的方便,有利于处理紧急情况。

#### 3. 空间数据库

空间数据库是随着地理信息系统GIS的开发和应用发展起来的数据库新技术,它主要包括地理信息数据库(GIS)和计算机辅助设计(CAD)数据库。地理信息数据库一般存储与地图相关的信息数据。计算机辅助设计数据库一般存储设计信息的空间数据库,如机械、集成电路、电子设备设计图等。

#### 4. 信息检索系统

信息检索就是根据用户输入的信息,从数据库中查找相关的文档和信息,并把查找的信息反馈给用户。信息检索领域和数据库是同步发展的,它是一种典型的联机文档系统或者联机图书目录。

#### 5. 专家决策系统

专家决策系统也是数据库应用的一部分。越积越多的数据可以联机获取,企业可以根据联机获取的数据做出正确的决策。人工智能的发展,使得专家决策系统的应用更加广泛。

#### 6. 知识库系统

数据库技术与人工智能技术相结合产生了知识库系统。人工智能是研究机器智能和智能机器的高科技学科,它需要大量的演绎和推理规则的支持,这就为数据库提供了用武之地。它通过将知识抽象化,条理化,利用数据库技术建立知识库,从而使数据库智能化。

### 1.1.4 数据库系统

数据库是用户应用计算机系统求解一类问题的泛称。数据库系统实际上是一个应用系统,它是在计算机硬、软件系统支持下,由数据库、数据库管理系统、数据库应用系统和人员构成的数据处理系统。其中数据库管理系统是数据库系统的核心组成部分。

## 1. 基本概念

### (1) 数据库

数据库(DataBase, DB)是存储在计算机存储设备上,结构化的相关数据的集合。它不仅存放数据,而且还存放数据之间的联系。数据库中的数据是以文件的形式存储在存储介质上的,它是数据库系统操作的对象和结果。

### (2) 数据库管理系统

数据库管理系统(DataBase Management System, DBMS)是位于用户与操作系统之间的帮助用户建立、使用和管理数据库的数据管理软件。用户使用的各种数据库命令以及应用程序的执行,都要通过数据库管理系统来统一管理和控制。数据库管理系统还承担着数据库的维护工作,按照数据库管理员所规定的要求,保证数据库的安全性和完整性。数据库管理系统通常有 4 个方面的主要功能:数据定义功能、数据操纵功能、数据控制功能和数据通信功能。

### (3) 数据库系统

数据库系统(DataBase System, DBS)是引入数据库技术后的计算机系统。数据库系统不但能够实现有组织地、动态地存储大量相关的数据,而且为数据处理和信息资源共享提供了便利条件。数据库系统主要由 5 部分组成:计算机硬件系统、数据库、数据库管理系统及相关软件、数据库管理员和用户。

#### ① 计算机硬件系统。

任何一个计算机系统都需要有中央处理器、存储器和输入输出设备等硬件。一个数据库系统需要有足够快的处理器来处理这些数据,以便快速响应用户的数据处理和数据检索请求,同时还需要有足够容量的内存与外存来存储大量的数据。对于分布式数据库系统,还需要有网络通信设备的支持。

#### ② 数据库。

在一个数据库系统中,可以根据需要创建多个数据库,并且数据库中的数据通常可以被多个用户所共享。

#### ③ 数据库管理系统及相关软件。

数据库管理系统是整个数据库系统的核心,它可以对数据库进行集中统一的管理。除了数据库管理系统之外,一个数据库系统必须还有其他相关软件的支持,如操作系统、编译系统、应用软件开发工具等。

#### ④ 数据库管理员。

数据库管理员是对整个数据库系统进行全面维护和管理的专门人员。

#### ⑤ 用户。

也称最终用户,他们可以通过应用系统的用户接口使用数据库。

### (4) 数据库应用系统

数据库应用系统(DataBase Application System, DBAS)是利用数据库系统资源开发的面向某一类实际应用的应用软件。例如,学生成绩管理系统、人事工资管理系统、产品销售管

理系统等。

## 2. 数据库系统的特点

数据库系统的出现是计算机数据处理技术的重大进步,它主要有以下特点。

### (1) 数据共享性好

数据共享是数据库系统最重要的特点。数据库中的数据能够被多个用户、多个应用程序所共享。数据共享可以大大减少数据冗余,节约存储空间,给数据应用带来很大的灵活性。

### (2) 数据独立性强

在数据库系统中,应用程序与数据是相互独立的,因此,对数据结构的修改不会对应用程序产生影响或者不会有大的影响,而对应用程序的修改也不会对数据产生影响或者不会有大的影响。

### (3) 数据结构化

数据库中的数据是以一定的逻辑结构存放的,这种结构是由数据库管理系统所支持的数据模型决定的。数据库系统不仅可以表示事物内部各数据项之间的联系,而且还可以表示事物和事物之间的联系。只有按一定结构组织和存放的数据,才便于对它们实现有效的管理。

### (4) 统一的数据控制功能

由于多个用户可以同时使用同一个数据库,因此必须提供必要的数据安全保护措施,包括安全性控制措施、完整性控制措施、并发操作控制措施等。

特别说明的是,一般说法中的“数据库”,可能指数据库本身,也可能指数据库管理系统,也可能指数据库系统,因此用户必须加以区别。

## 3. 数据库管理系统的功能

数据库管理系统的功能可以概括为以下4个方面。

### (1) 数据库的定义

按照用户的要求定义数据库的结构,并在计算机内部真正建立装有数据的数据库,包括逻辑结构的定义、物理结构(即存储结构)的定义和保密规定等。

数据库管理系统提供数据定义语言或者操作命令以便对数据库进行具体的描述,同时系统包含 DDL 的编译或解释程序。

### (2) 数据库的操作

数据库的操作是指接受、分析和执行用户对数据库中数据的存取要求,主要包括检索、插入、删除和更新等操作。

数据库管理系统提供语言(或者命令)对数据库中的数据进行追加、插入、修改、删除和检索等操作,称为数据操纵语言(Data Manipulation Language, DML)。在不同的数据库管理系统中,语言的语法格式也不相同,根据其实现方法可分为两种类型:一类数据操纵语言可以独立使用、不依赖于任何其他程序设计语言,称为自含型或自主型数据操纵语言;另一类是需嵌入到某一程序设计语言中使用,例如嵌入到 FORTRAN、Basic、C 等程序设计语言中,



这些程序设计语言称为宿主型操纵语言。在使用高级语言编写应用程序时,如果需要调用数据库中的数据,则需要用宿主型数据操纵语言的语句来操纵数据。同时 DBMS 还包含 DML 的编译或解释程序。

### (3) 数据库的运行控制

数据库中的数据必须采用一定的控制手段加以保护以免损坏,数据库管理系统主要提供以下 4 方面的数据控制功能,即数据的完整性控制、数据库的并发操作控制、数据的安全性控制和数据库的恢复。

此外,还要对数据存取、数据库的初始装入、数据库转存、数据库回复、数据的内部维护和通信等进行控制。

### (4) 数据字典

数据字典(Data Dictionary, DD)提供了对数据库中的数据进行管理的规程,对数据库的使用和操作可通过查阅数据字典来完成。

## 4. DBMS 的三级模式

为了提高数据库的逻辑独立性和物理独立性,美国国家标准局(ANSI)计算机与信息处理委员会(代号为 X3)以及标准规划和要求委员会(SPARC)分别于 1975 年、1978 年提出了标准化建议,将数据库模式定义为三级:外模式、内模式和概念模式。

### (1) 外模式

数据库的外模式是最接近用户的层次,相应的数据库层次称为“外部层”“外视图”或“用户视图”。它是用户看到的数据库的内容。

外模式还可称为子模式。因为一个用户往往只能用到数据库的一部分,看到的只是一部分数据库,是允许用户使用的那部分数据的逻辑结构,所以它表示的是数据库的局部逻辑。

### (2) 内模式

内模式是数据库的物理结构,相应的数据库层次称为“内部层”“存储视图”或“内部视图”,是机器或操作系统“看到”的数据库结构。内模式要定义数据的物理存储策略,以及各种管理细节。

### (3) 概念模式

外模式是面向单个或部分用户的数据库局部逻辑结构,概念模式则是面向所有用户的数据库全局逻辑结构。因此,与概念模式对应的数据库层次也称为“概念层”“全局视图”或“概念视图”,它是数据库中所有信息的抽象表示。

概念模式也可简称为“模式”。

### (4) 三级模式之间的关系

数据库系统的三级模式结构如图 1-5 所示。

在三级模式结构之间,存在着两级映射。应用程序根据外模式进行操作,通过外模式到概念模式的映射,与概念模式联系起来;概念模式又通过概念模式到内模式的映射,与内模式联系起来。

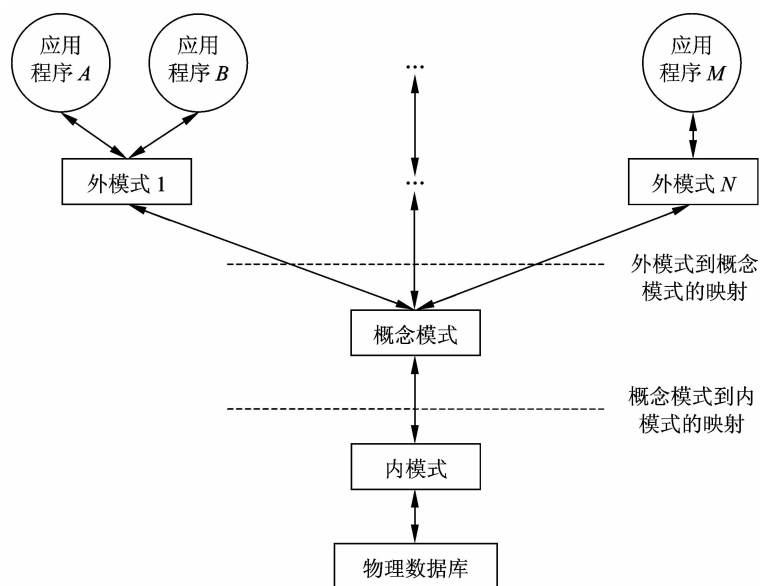


图 1-5 数据库系统的三级模式结构

将数据库系统结构分成 3 个层次,对于提高数据独立性具有重要意义。如果数据库的物理存储结构发生变化,可以相应地改变概念模式到内模式之间的映射,从而使概念模式保持不变,这就保证了数据的物理独立性;如果数据库的概念模式发生变化,也可以相应地改变外模式到概念模式之间的映射,从而使外模式保持不变,用户的应用程序也就不必改变,这就保证了数据的逻辑独立性。

所以,数据库管理系统 DBMS 的中心任务,就是管理数据库系统的三级模式,并完成它们之间的模式转换,把用户对数据库的操作从用户级转换到物理级,从外模式转换到内模式,完成具体执行。

## 1.2 数据模型

模型是对现实世界事物特征的模拟和抽象。由于计算机不可能直接处理现实世界中的具体事物,人们必须事先把具体事物转换成计算机能够处理的数据,这种具有联系性的相关数据总是按照一定的组织关系排列,从而构成一定的结构,对这种结构的描述就是数据模型。

所以说,数据模型就是一种模型,它是现实世界数据特征的抽象。

### 1.2.1 实体间的联系

#### 1. 三个“世界”

现实世界的事物无穷无尽,形态千差万别,将现实世界的事物及其联系转化为计算机及

数据库所允许的形式的过程,一般要经过 3 个阶段,或者说,要通过三个“世界”。

首先,分析现实世界中的事物及其联系;其次,将现实世界中的客观事物抽象为信息世界中的实体;最后,再将实体转化为 DBMS 支持的数据世界中的数据,如图 1-6 所示。

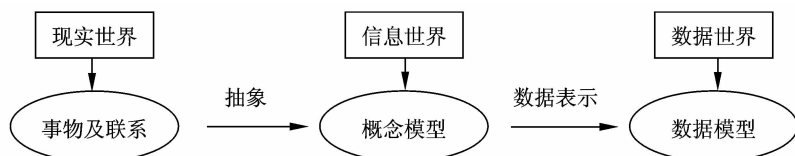


图 1-6 三个“世界”之间的关系

其中,面向客观世界、面向用户的数据模型称为概念数据模型,简称概念模型。它是现实世界的直接抽象,强调其表达能力和易理解性,要求用户和设计者都能理解。另一个是面向 DBMS 的、用以刻画实体在数据库中的存储形式及实体之间的联系的模式,称为逻辑数据模型,简称数据模型。

## 2. 实体的相关概念

由于数据模型是用来描述现实世界中的事物与事物之间联系的,所以,在介绍数据模型之前,先来了解一下实体和实体之间的联系。

### (1) 实体(Entity)

客观存在并相互区别的事物称为实体,实体可以是实际的事物,也可以是抽象的事物。如职工、学生、教师、课程就是具体的实体,而学生的选课、教师的授课、产品的订货、一次比赛等也是实体,它们是抽象实体。

### (2) 实体集

性质相同的同类实体的集合称为实体集,如一群学生、一个公司的职工、一批书籍等。

### (3) 属性(Attribute)

描述实体的特性称为属性。一个实体可以由若干个属性来刻画,例如,学生实体可以由学号、姓名、性别、年龄、院系等属性描述;一场比赛实体可以由比赛名、比赛时间、比赛地点、参赛者等属性描述。

### (4) 实体型

所谓实体型就是对实体的型的描述,通常使用实体名和属性名的集合,如学生实体,其型的描述为:学生(学号,姓名,性别,出生年份,系,入学时间)是一个实体型。

### (5) 实体值

所谓实体值就是指实体的具体实例,它是属性值的结合,如学生王劲的实体值是:98101,王劲,男,1986,计算机,1998。

## 3. 实体间的联系及其类型

实体之间的对应关系称为联系,它反映了现实世界事物之间的相互关联。联系分为两种,一种是实体内部各属性之间的联系,如具有相同职称的有很多人,但一个职工当前只有一种职称;另一种是实体之间的联系,如一位运动员可以参加若干项比赛,同一项比赛有若干个运动员参加。

实体间的联系类型是指一个实体类型中的每一个实体与另一个实体类型中多个实体存在联系。实体之间的联系可以归纳为一对一的联系、一对多的联系和多对多的联系3种类型。

### (1) 一对一联系

如果对于实体集A中的每一个实体,实体集B中至少有一个实体与之联系,反之亦然,则称实体集A与实体集B具有一对一联系。记为1:1。例如,一个班只有一个正班长,而班长只在一个班任职,则班级与班长之间有一对一联系。但必须注意,一对一联系不一定是——对应的联系,如图1-7所示。

### (2) 一对多联系

如果对于实体集A中的每一个实体,实体集B中有 $n(n \geq 0)$ 个实体与之联系,反之,对于实体集B中的每一个实体,实体集A中至多只有一个实体与之联系,则称实体集A与实体集B有一对多联系。记为1:n。例如,一个班级有多个学生,而一个学生只能在一个班级中,所以实体集班级与学生是一对多的联系。此外还有公司与员工,院系等都是——一对多联系,如图1-8所示。

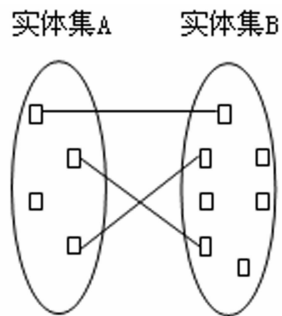


图 1-7 一对一联系

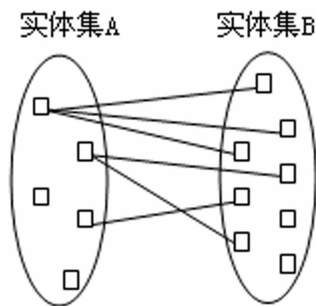


图 1-8 一对多联系

### (3) 多对多联系

如果对于实体集A中的每一个实体,实体集B中有 $n(n \geq 0)$ 个实体与之联系,反之,对于实体集B中的每一个实体,实体集A中也有 $m(m \geq 0)$ 个实体与之联系,则称实体集A与实体集B具有多对多联系。记为 $m:n$ 。例如,一个教师有多个学生,一个学生有多个教师,所以实体集教师与学生是多对多的联系。此外还有学生与课程,图书与借书人等都是多对多联系,如图1-9所示。

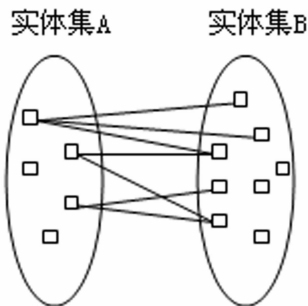


图 1-9 多对多联系

#### 4. 实体 - 联系模型

了解实体、实体的属性及实体之间的联系之后,就可以着手建立实体 - 联系模型了。

该模型是用来对实体、实体的属性及实体之间的联系进行描述的。实体 - 联系图(简称 E - R 图)就是一种描述实体 - 联系模型的常用工具。

E - R 图的图例说明如下。

- ①实体:用矩形表示,矩形框内写明实体名。
- ②实体的属性:用椭圆形表示,并用无向边将其与相应的实体连接起来。
- ③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、m: n)。

以上 3 种联系如图 1-10 所示。

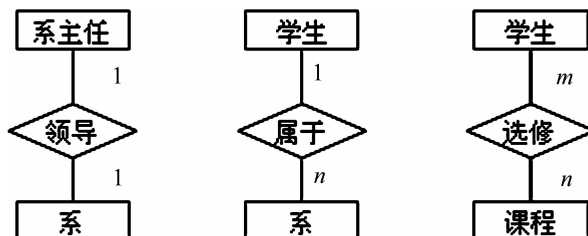


图 1-10 实体之间的联系 E - R 图

例如,考查学生和考试成绩所组成的系统。学生有学号、姓名等属性,某学生和他取得的各门课成绩之间是一对多联系,联系方式是考试,而成绩实体又包括学号(唯一标识是哪个学生的成绩)、科目、成绩等属性。可以用图 1-11 表示该系统的概念模型。

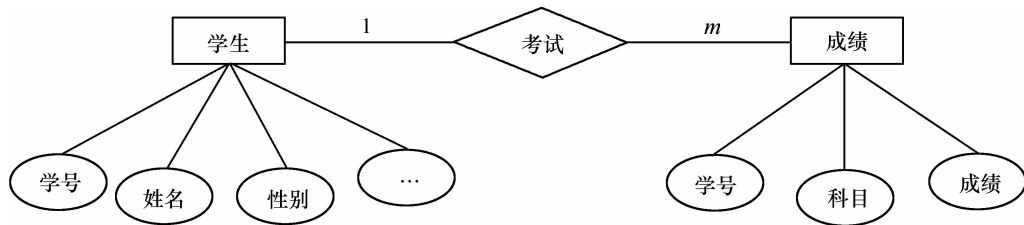


图 1-11 学生与考试成绩实体的 E - R 图

### 1.2.2 数据模型

从理论上讲,数据模型是指反映事物与事物之间联系的数据组织结构和形式。任何一个数据库管理系统都是基于某种数据模型的。数据库管理系统常用的数据模型有多种,但完全成熟并常用的只有层次数据模型、网状数据模型和关系数据模型 3 种数据模型。与之相对应,数据库也分为 3 种基本类型:层次数据库、网状数据库和关系数据库。

#### 1. 层次数据模型(Hierarchical Model)

层次模型是用树形结构来表示实体与实体之间的联系。在这种模型中,记录类型为结点,由根结点、父结点和子结点构成。层次模型像一棵倒置的树,根结点在上,层次最高,

子结点在下,逐层排列。其主要特征是有且只有一个无双亲的根结点;根结点以外的子结点,向上仅有一个父结点,向下可以有若干个子结点。

层次模型表示的是一个父结点对应于多个子结点,而一个子结点只能对应于一个父结点的一对多的联系,它不能表示较复杂的数据结构,但却简单、直观、处理方便、算法规范,如图 1-12 所示。

支持层次数据模型的 DBMS 称为层次数据库管理系统,在这种系统中建立的数据库是层次数据库。层次模型可以方便地表示一对一联系和一对多联系,但不能用它直接表示多对多联系。

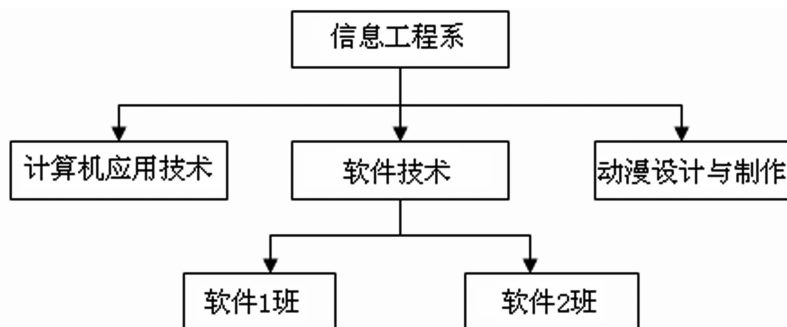


图 1-12 层次模型

## 2. 网状数据模型 (Network Model)

网状模型是用网状结构表示实体与实体之间的联系。在这种模型中,记录类型为结点,由结点与结点之间的相互关联构成,网状模型是层次模型的扩展,表示多个从属关系的层次结构,呈现一种交叉关系的网络结构。其主要特征是允许有一个以上的结点无双亲结点,至少有一个结点有多于一个的双亲结点。

网状模型在概念上、结构上都比较复杂,实现的算法也难以规范化,但这种数据模型可以表示较复杂的数据结构,如图 1-13 所示。

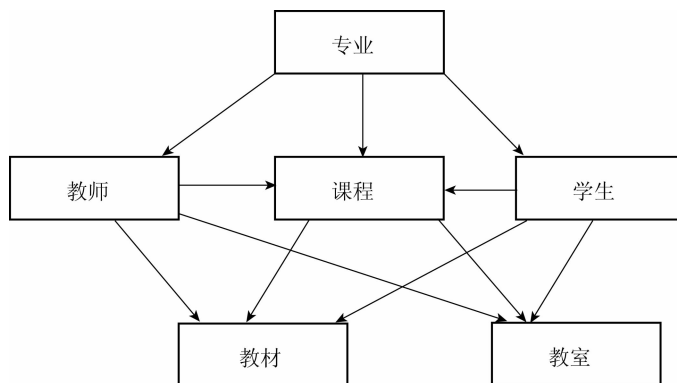


图 1-13 网状模型

支持网状数据模型的 DBMS 称为网状数据库管理系统,在这种系统中建立的数据库是

网状数据库。网状结构可以直接表示多对多联系,这也是网状模型的主要优点,当然在一些已经实现的网状模型 DBMS 中,对这一点做了限制。

### 3. 关系数据模型 (Relational Model)

二维表在数据库理论中称为关系,用关系(即二维表)表示的数据模型称为关系数据模型。

在关系数据模型中,操作的对象和结果都是二维表。其主要特征是关系中每一数据项不可再分,是最基本的单位;每一竖列是同属性的,列数根据需要而设,且各列的顺序是任意的;每一行由一个事物的诸多属性构成,行数根据需要而定,且各行的顺序是任意的。由于浏览和设计二维表十分方便,所以二维表的关系数据模型比层次数据模型和网状数据模型更具优势,是目前应用最广的数据库模型。表示一组有关学生的信息,就是一种关系数据模型,如表 1-1 所示。

表 1-1 关系模型

学号	姓名	性别	班级	照片	团员
98101	王劲	男	计算机 98	Gen	F
98102	王浩天	男	计算机 98	Gen	T
98301	何芳	女	软件 98	Gen	F
99101	章达	男	计算机 99	Gen	F
99301	徐静	女	软件 99	Gen	T

支持关系数据模型的 DBMS 称为关系数据库管理系统,例如,Access 就是一种广泛使用的关系型数据库管理系统。

## 1.3 关系数据库

由于关系数据模型比层次数据模型和网状数据模型更具优势,自 20 世纪 80 年代以来,新推出的数据库管理系统几乎都支持关系数据模型,早期许多采用层次和网状模型的产品也加上了关系数据接口。

### 1.3.1 关系的基本概念及特征

#### 1. 关系的基本概念

关系数据库是基于关系模型的。有关的一些基本术语如下。

##### (1) 关系

一个关系就是一张二维表,通常将一个没有重复行、重复列的二维表看成一个关系,每个关系都有一个关系名。对关系的描述称为关系模式,一个关系模式对应一个关系的结构,

其一般格式为:关系名(属性名1,属性名2,⋯,属性名n)

例如:学生(学号,姓名,性别,班级,团员否,相片)

### (2)元组

表中的行称为元组,每一行是一个元组,对应数据表中的一个具体记录,元组的各分量分别对应于关系的各个属性。例如,表1-1中姓名为“王劲”所在行上的所有数据就是一个元组或一条记录。

### (3)属性

二维表中的每一列在关系中称为属性或字段,每个属性都有一个属性名,属性值则是各个元组的属性的取值。例如,表1-1中的第2列就是一个属性,而“姓名”则是它的属性名,“徐静”则为第5元组姓名属性的属性值。

### (4)域

表中属性的取值范围称为域,即表中字段的取值范围。例如,表1-1中的第3列性别的取值范围是“男”或“女”。

### (5)关键字

能唯一确定一个元组的单个属性或多个属性的组合称为关键字。例如,表1-1中的“学号”可以作为主关键字,因为每名学生的学号不允许相同。而“姓名”、“性别”则不能作为主关键字,因为学生中可能出现重名或相同出生日期。

### (6)外部关键字

如果表中的一个属性不是本表的关键字,而是另外一个表的关键字,这个属性就称为外关键字。

例如:学生(学号,姓名,性别,出生日期,专业编号)课程(课程编号,课程名称,学分,课时)学生选课(学号,课程编号,成绩)其中,对于学生选课表而言,学号或课程编号都是外部关键字。

## 2. 关系的特点

在关系数据库中,每一个关系都必须满足一定的条件,即关系必须规范化。一个规范化的关系必须具备以下几个特点。

①关系中的每一个属性都必须是不可分解的,是最基本的数据单元。即数据表中不能再包含表。

②一个关系中不允许有相同的属性名。即在定义表结构时,一张表中不能出现重复的字段名。

③关系中不允许出现相同的元组。即数据表中任意两行不能完全相同。否则不仅会增加数据量,造成数据的“冗余”(重复存储)。

④关系中同一列的数据类型必须相同,即同一属性的数据具有同质性。指数据表中任一字段的取值范围应属于同一个域。

⑤关系中行、列的次序任意。即数据表中元组和字段的顺序无关紧要。任意交换两行或两列的位置并不影响数据的实际含义。