

数据库应用基础

电子信息应用基础知识丛书

吴保国 孙国光 刘瑞

朱鹏举 审
编写

清华大学出版社

<http://www.tup.tsinghua.edu.cn>

(京)新登字 158 号

内 容 简 介

35.150/28-20

本书较全面地介绍了数据库系统的基本概念、基本原理和基本技术。内容力求深入浅出,通俗易懂,它将告诉读者:数据库是什么,能做什么事,以及如何通过掌握的数据库理论设计自己的数据库应用系统。

本书可作为各行各业的干部、职工等学习和了解数据库系统的入门读物,亦可供数据库管理系统的人员参考。

图书在版编目(CIP)数据

数据库应用基础/吴保国等编著. - 北京:清华大学出版社,1998.12

(电子信息应用基础知识丛书)

ISBN 7-302-03234-3

I.数… II.吴… III.数据库管理系统-基础知识 IV.TP311.13

中国版本图书馆 CIP 数据核字(98)第 35770 号

出版者:清华大学出版社(北京清华大学校内,邮编:100084)

<http://www.tup.tsinghua.edu.cn>

印刷者:清华大学印刷厂印刷

发行者:新华书店总店北京科技发行所

开 本:787×1092 1/16 印张:6.5 字数:130 千字

版 次:1998 年 12 月第 1 版 1998 年 12 月第 1 次印刷

书 号:ISBN 7-302-03234-3/TP·1731

印 数:0001~8000

定 价:10.00 元

序 言

当今信息化的热潮席卷全球,在发达国家已受到公众和政府的普遍关注,在发展中国家也已引起政府的高度重视。信息化同工业化一样,是人类社会生产力发展的新标志,信息化将改变人们的工作、学习和生活方式。

信息化是一个发展过程,在这个过程中,要利用现代信息技术改造传统工业,实现信息资源普遍共享,推动经济和社会的优质发展。

为了推进信息化事业,首先要普及信息技术知识,让人民大众懂得和能够应用电子信息技术知识。为此,电子部信息中心与清华大学出版社共同策划编辑出版一套普及宣传电子信息技术及其应用的丛书,以满足广大读者的要求。丛书从电子信息技术及其应用的方方面面,用形象易懂的语言,用非专业人员的思维逻辑,用通俗易懂的比喻来描述和表达电子信息技术的深奥知识,介绍其在各个方面的广泛应用。

今天人类发展和进步到了信息化时代,掌握电子信息技术并应用电子信息技术改造客观世界和主观世界,推动国民经济以及人们生活的各个领域的信息化,是我们每一个公民义不容辞的责任。

让《电子信息应用基础知识丛书》,枝繁叶茂、五彩缤纷,能受到广大读者的喜爱。

陈 正 清

1997年3月10日

电子信息应用基础知识丛书

编 委 会

主 编 陈正清

副主编 朱鹏举 徐培忠

编 委 吴克忠 侯炳辉

李思三 王海燕

王 永 帅志清

前 言

数据库技术是 60 年代后期开始发展起来的一项信息管理的最新技术,是现代计算机系统的一个重要组成部分,是人们有效地存储、共享和处理数据的工具。在信息“爆炸”的今天,数据库技术的发展极为迅速,已广泛应用于各行各业,成为人们的得力助手。现代的管理信息系统几乎都是以数据库技术作为核心的。

随着计算机科学的迅速发展,计算机在我国的普遍推广应用,数据库“热”正在悄然兴起,越来越多的人希望了解、学习和使用数据库。本书是一本关于数据库的入门性读物,它所介绍的内容包括数据库是什么,能做什么事,数据库的一般原理,数据库产品及其特点,怎样设计一个实用的数据库系统等,帮助读者对数据库系统有一个较全面的认识。

本书分为三部分,第一部分介绍基础知识(第 1 章~第 3 章);第二部分介绍数据库的基础理论、数据库安全性保护和数据库的一般设计开发方法(第 4 章~第 6 章);第三部分介绍数据库系统的发展、国内数据库管理系统流行产品和开发工具(第 7 章)。

由于篇幅所限,书中所述内容较为简略,对于初级用户,如遇到难于理解之处,可查阅其他相关书籍。本书的读者对象是各行各业的干部、职工以及非计算机专业的大中专学生。凡希望知识更新,希望在工作中应用计算机做管理的人也可从中受益。

本书第 1 章~第 6 章由吴保国编写,第 7 章由孙国光、刘瑞、吴保国编写。孙国光在全书的编写过程中做了大量的策划、组织、资料收集等具体工作,韩琳琳、何平洪做了资料整理等具体工作。

本书引用了一些参考书的内容,在此对各位作者表示感谢。

我们在编写中尽可能引进新观点,力求简单、通俗、易懂,但由于水平有限,必有许多不足之处,诚请读者批评指正。

作 者

1998 年 5 月

目 录

第 1 章 数据库概述	1
1.1 信息、数据与数据处理	1
1.1.1 信息与数据	1
1.1.2 数据处理	2
1.2 数据管理的三个阶段	2
1.2.1 人工管理阶段	2
1.2.2 文件系统阶段	3
1.2.3 数据库系统阶段	4
1.3 数据库是什么	4
1.4 数据库的特点	6
1.5 数据模型	7
1.5.1 数据之间的联系	7
1.5.2 实体模型	9
1.5.3 数据模型	10
第 2 章 数据库系统	14
2.1 数据库系统的组成	14
2.2 数据库管理系统	16
2.2.1 数据库管理系统的抽象层次	16
2.2.2 DBMS 的主要功能	19
2.2.3 DBMS 的组成	20
2.2.4 DBMS 对数据的存取	21
第 3 章 关系数据库	22
3.1 关系模型的基本概念	22
3.1.1 关系模型的基本术语	22
3.1.2 关系的性质	23
3.1.3 关系的键	24
3.1.4 关系模型的完整性	25
3.1.5 关系模式与关系数据库	25
3.2 关系数据语言	27
3.2.1 关系数据语言的分类	27
3.2.2 关系代数	28

3.3	SQL 语言	30
3.3.1	SQL 语言概述	30
3.3.2	SQL 数据定义	31
3.3.3	SQL 数据查询	31
3.3.4	SQL 数据操纵	32
3.3.5	SQL 数据控制语言	33
3.4	QBE 语言	33
3.4.1	QBE 语言概述	33
3.4.2	QBE 语言的查询操作	33
3.4.3	FoxPro 系统的 QBE	34
3.5	xBASE 语言与 FoxPro	34
3.5.1	xBASE 语言概述	34
3.5.2	FoxPro 语言的特点和语法规则	35
3.5.3	FoxPro 数据定义	35
3.5.4	FoxPro 数据查询	36
3.5.5	FoxPro 数据操纵	37
3.6	关系数据语言的特点	38
第 4 章	关系数据库设计理论	40
4.1	数据库设计中的问题	40
4.1.1	数据库设计概述	40
4.1.2	问题的提出	40
4.1.3	解决办法	42
4.2	函数依赖	43
4.3	关系模式的规范化	44
4.3.1	范式	44
4.3.2	规范化和模式分解	44
第 5 章	数据库安全保护	48
5.1	数据库故障	48
5.2	数据库的恢复	48
5.2.1	恢复的基本原则和实现方法	49
5.2.2	实现数据库恢复的各系统功能	49
5.3	数据库的安全性控制	50
5.3.1	操作系统层进行标识证实	51
5.3.2	DBMS 存取控制	52
5.3.3	密码存储	53
5.4	数据库完整性控制和完整性约束分类	53

5.4.1	数据库完整性控制	53
5.4.2	完整性约束的分类	53
5.5	数据库并发控制	54
5.5.1	封锁	55
5.5.2	死锁及消除的方法	56
第 6 章	数据库应用系统设计	57
6.1	数据库应用系统设计开发综述	57
6.1.1	数据库应用系统的中心问题和开发方法	57
6.1.2	开发策略和开发条件	58
6.1.3	数据库应用系统的设计原则和准则	59
6.2	数据字典	59
6.3	软件工程分步设计法	60
6.3.1	软件工程分步设计法的开发过程	60
6.3.2	系统规划需求分析	61
6.3.3	概念设计	64
6.3.4	逻辑设计	65
6.3.5	建立数据库	66
6.3.6	应用程序设计、编写和调试	67
6.3.7	数据库的运行和维护	67
6.4	快速原型法	68
6.4.1	快速原型法的思想	68
6.4.2	快速原型法的生命周期	68
6.4.3	快速原型开发方法的特点	69
6.5	直接设计法	70
第 7 章	数据库技术的发展和国内流行的 DBMS	71
7.1	数据库技术发展历程	71
7.2	新一代数据库技术的研究和发展	72
7.3	微机关系数据库的发展	73
7.4	关系数据库管理系统及其产品	74
7.4.1	Oracle	74
7.4.2	DB2	76
7.4.3	Sybase	76
7.4.4	Informix	78
7.4.5	Ingres	80
7.4.6	SQL Server	81
7.4.7	xBASE 类数据库	81

7.4.8 其他数据库系统	83
7.5 新一代数据库应用系统开发工具	84
7.5.1 PowerBuilder	86
7.5.2 Delphi	87
7.6 客户/服务器结构的数据库系统	88
7.6.1 客户/服务器各部分的任务	88
7.6.2 客户/服务器结构的特点	88
7.7 分布式数据库系统	89
7.7.1 分布式数据库系统概述	89
7.7.2 分布式数据库的特点	89
参考文献	91

第 1 章 数据库概述

随着计算机技术的蓬勃发展,计算机应用从科学计算、过程控制进入数据处理,使计算机从少数科学家手中的珍品成为人们日常工作中处理数据的得力助手和有力工具。当今世界,在计算机的三大主要应用(科学计算、过程控制和数据处理)领域中,数据处理迅速上升为计算机应用的主要方面。在本章中,我们从数据处理开始介绍数据处理的发展、数据模型和数据库以及所涉及的基本概念,作为后面各章学习的基础。

1.1 信息、数据与数据处理

在信息社会中,信息是一种资源,其重要性可以与物质和能量相提并论。对企业来说,各种必须的信息是其赖以生存和发展的命根子。按信息论的观点,任何社会实践活动都可以抽象为人流、物流、财流、能源流和信息流这五种流运动,其中起主导和支配作用的是信息流,它调节和控制着其他各流的数量、方向、速度和目标,从而使社会实践活动更具有目的性和规律性。信息是维持生产活动、经济活动和社会活动必不可少的资源,因此,信息是有价值的,信息成为构成客观世界的三大要素(信息、能源和材料)之一。人们为了获取有价值的信息用于决策,就需要对数据进行处理和管理。

人们把用计算机对数据进行处理的应用系统称为计算机信息系统。信息系统是“一个由人、计算机等组成的能进行信息的收集、传递、储存、加工、维护、分析、计划、控制、决策和使用的系统”。信息系统的核心是数据库。

1.1.1 信息与数据

在数据处理这一领域中,我们常常遇到的基本概念是数据(data)和信息(information),既然信息与数据是两个不同的术语,它们之间必定存在一定的差别。

1. 信息

信息是现实世界事物的存在方式或运动状态的反映。具体地说,信息是一种已经被加工为特定形式的数据,这种数据形式对接收者来说是有意义的,而且对当前和将来的决策具有明显的或实际的价值。

信息有如下一些重要特征:

- ① 信息传递需要物质载体,信息的获取和传递要消耗能量。
- ② 信息是可以感知的。
- ③ 信息是可以存储、压缩、加工、传递、共享、扩散、再生和增值的。

2. 数据

数据是将现实世界中的各种信息记录下的、可以识别的符号,是信息的载体,信息的具体表现形式。可用多种不同形式的数据来表示同样的信息,信息不随它的数据形式不

同而改变。数据的表现形式多种多样,不仅有我们熟知的数字和文字,还可以有图形、图像、声音等形式。

3. 数据与信息的关联

数据与信息是密切相关联的。信息是向人们(或机器)提供关于现实世界有关事物的知识;数据则是载荷信息的物理符号,二者是不可分离而又有一定区别的概念。例如,“今年大学招生录取线为 450 分,张红的高考成绩为 500 分。”这段文字(数据)提供了张红可能被某一个大学录取的信息。但是,在一些不很严格的场合下,对它们不做严格的区分,当作同义词互换使用。譬如,信息处理与数据处理、信息资源与数据资源、信息采集与数据采集等。

1.1.2 数据处理

数据处理也称为信息处理。数据处理实际上就是利用计算机对各种形式的数据进行处理。它包括:数据采集、整理、编码和输入,有效地把数据组织到计算机中,由计算机对数据进行一系列储存、加工、计算、分类、检索、传输、输出等操作过程。数据处理的目的是从大量的、原始的数据中抽取和推导出对人们有价值的信息以作为行动和决策的依据。例如,一个企业,需要对其收集的大量的有关市场产品销售的信息(数据)进行存储、加工、计算,生成市场销售情况图表,从而获得哪种型号的产品最受欢迎的信息,以指导生产计划。

在数据处理的一系列活动中,数据收集、分类、组织、编码、存储、检索、传输和维护等操作作为基本操作,我们将这些基本操作环节称为数据管理,而加工、计算、输出等操作是千变万化的,不同业务有不同的处理,数据管理技术是解决上述基本环节的,而其他操作由应用程序实现。

数据处理的中心问题是数据管理。

数据库技术所研究的问题就是如何科学地进行数据管理。数据库技术是数据管理的最新技术。数据库系统是当代计算机系统的重要组成部分。

随着计算机的软硬件技术的发展,数据管理技术的发展大致经历了从人工管理、文件系统和数据库系统三个阶段。

1.2 数据管理的三个阶段

1.2.1 人工管理阶段

50 年代初期,当计算机一出现,人们就试图使用计算机来处理这些数据。在这一阶段,计算机除硬件外,没有任何软件可供数据处理使用,因而计算机主要用于科学计算。而对数据管理时,设计人员除考虑应用程序、数据的逻辑定义和组织外,还必须考虑数据在存储设备内的存储方式和地址。这个阶段的数据管理有如下特点:

1. 数据不保存

由于没有软件系统对数据进行管理和计算机硬件的限制,每次计算前,先将程序和数

据输入。计算结束后,将结果输出,计算机不保存程序和数据。

2. 数据面向程序

每个程序都有属于自己的一组数据,当数据在存储上稍有改变,就要修改程序。程序与数据相互结合成为一体,互相依赖,独立性差。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)。

数据与程序之间的关系如图 1.1 所示。

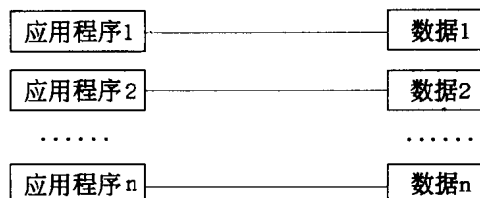


图 1.1 数据与程序之间的关系

1.2.2 文件系统阶段

在 50 年代后期至 60 年代中期,计算机外存已有了磁鼓、磁盘等存储设备,软件有了操作系统。人们在操作系统的支持下,设计开发了一种专门管理数据的计算机软件,称之为文件系统。这时,计算机不仅是用于科学计算,而且还大量用于数据处理。由于有了文件系统,这个阶段的数据管理有如下特点:

1. 数据与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理,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. 文件形式多样化

为了便于数据的存储和查找,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存储数据。数据的存取基本上是以记录为单位。(注:字段的有序集合称为记录,例如一个学生的记录是由学生编号、姓名、年龄、性别等字段组成。)

3. 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,管理数据时,不必过多地考虑数据物理存储的细节,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,数据在存储设备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,数据的存放仍依赖于应用程序的使用方法,基本上是一个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,不同的应用程序很难共享同一数据文件。因此出现数据重复存储,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,造成应用程序编制繁琐,数据的正确性、安全性、保密性等得不到保证。文件系统中程序与数据之间的关系如图 1.2 所示。

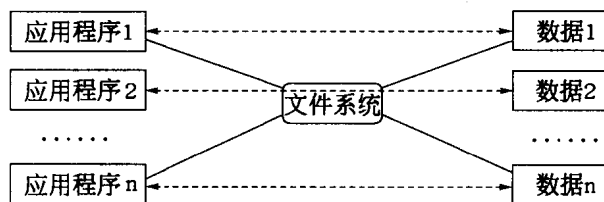


图 1.2 文件系统中程序与数据之间的关系

1.2.3 数据库系统阶段

60 年代末期开始,随着计算机技术的发展,数据管理规模的扩大,数据量急剧增加,为了提高数据处理的效率,满足数据共享的要求,人们对文件系统进行了扩充,研制了一种新的、先进的数据处理方法,即数据库系统。数据库系统克服了以前所有处理方式的缺点,试图提供一种完美的、更高级的数据处理方式。它的指导思想是对所有的数据实行统一的、集中的、独立的管理,使数据存储独立于使用数据的程序,实现数据共享。

1.3 数据库是什么

数据库一词的英文写法为 database 或 data base,简称 DB。data 的中文含义是数据,base 是基地,database 意指供给数据的基地即为数据库。顾名思义,数据库是计算机中存放数据的“仓库”。仓库建立在外存储媒介中,如磁盘、光盘等。

在我们的日常生活中,有各种各样的库,例如书库、粮库、汽车配件库等等。这些库都具有这样几个特点:仓库由一到多个库房组成,可以存放现实世界中一定的东西,这些东西有条理、有组织、合理地存放在仓库中某个库房的货位或货架上,且这些东西由仓库管理人员进行管理,有一套管理体制,即系统。人们到仓库去取货,必须通过仓库管理员。

数据库系统实质上就是一个记录保存系统。它可以将各种数据很方便地保存到数据库中,数据库中的数据由一个称为数据库管理系统的软件进行管理,对数据库的访问必须通过数据库管理系统。

例如,在计算机上有一个学生情况数据库 STUDENT,它帮助我们记录学生信息。现在如果我们想查询有关管理信息系学生的学号、姓名、性别、年龄情况,可输入下面的命令:

```
SELECT 学号、姓名、性别、年龄 FROM STUDENT WHERE 系名 = "管理信息系"
```

于是,在计算机屏幕上就会立即显示所需要的学生库中管理信息系全部学生的信息,如图 1.3 所示。

学 号	姓 名	性 别	系 名	年 龄
970001	李 一	男	管理信息系	16
970002	王 二	男	管理信息系	17
970003	张小丽	女	管理信息系	15
970004	赵 芳	女	管理信息系	17
...				...

图 1.3 显示查询结果

当记录的数据发生变化时,我们就要对数据库中的数据进行更新。例如,每过一年,学生年龄增长一岁,要在学生数据库 STUDENT 中反映出来,就要更新数据库中所有的记录,命令如下:

```
UPDATE STUDENT SET 年龄 = 年龄 + 1
```

我们可以将数据库中没有必要保留的记录删除掉。例如,学号为“970002”的学生已退学,用命令将该学生的记录从库文件 STUDENT 中删除掉,命令如下:

```
DELETE FROM STUDENT WHERE 学号 = "930002"
```

我们可以向数据库中插入 (INSERT) 新记录。例如,向库中增加一个新学生“周丽同学”,命令如下:

```
INSERT INTO STUDENT ( 学号, 姓名, 性别, 系名, 宿舍, 年龄, 家庭住址);
VALUES ("970040", "周丽", "女", "计算机", "学 2202", 16, "北京海淀")
```

在上述例子中,我们对数据库进行了四种基本操作:检索、更新、删除和插入。这四种操作通常称为数据库操作,它们是最常用的操作。

一个数据库系统的主要功能之一就是允许数据操作,但是数据库系统绝非仅此而已,还需要有许多的附加功能(例如屏幕格式定义、菜单定义、打印输出控制等等)和数据定义(即建库功能)功能。上述这些功能,说准确一点,是支持一种数据库语言,用户可以用这种语言编写程序去处理数据。本章的例子采用的是 Visual FoxPro 3.0 数据库管理系统提供的数据库标准语言 SQL 来表达的。

每一种数据库系统都有它自己支持的语言(命令)来实现对数据的操作,这有点类似于地球上不同的国家和民族使用各自的语言来表达各自的思想。但是,语言不同给不同的民族之间的交流带来了困难。数据库的语言同样如此,因而国际标准化组织(ISO)选定了 SQL 语言作为数据库的标准语言。有关 SQL 语言和其他数据库语言我们在后面的章节将做专门介绍。

从上述例子中我们可以看出,数据库看起来就像一张大二维表,对数据库的各种操作就像在表上按条件插入几行或删除几行,或从表中选取几行和几列组成新的表。我们还可看出,数据库中对数据的存取不一定以记录为单位(文件系统对数据的存取是以记录为单位的),可以是字段。事实上,今天我们用的数据库就是一个或几个彼此相关的表的集合,数据记录就是以表中行的形式存放在库中的。

由此可见,用数据库对数据管理比之传统的手工对数据管理优越得多,主要表现在:紧凑、迅速、简洁和及时。

由于数据库是一个很复杂的系统,涉及面很广,因而许多专家从不同角度给数据库下过定义。总的来说,数据库是存储在一起的相关数据集合,数据被结构化。这些数据去掉了有害的或不必要的冗余,为多种应用服务,数据的存储独立于使用它的程序;对数据库插入新数据,修改和检索原有的数据,均可按一种公用的可控制方式进行。

数据库系统 DBS(Data Base System)则是一个由计算机硬件、操作系统、数据库管理系统、其他相关的软件(如编译系统等)、数据库和用户构成的系统。

1.4 数据库的特点

1. 数据共享

这是数据库系统区别于文件系统的最大特点之一,也是数据库系统技术先进性的重要体现。共享是指多用户、多种应用、多种语言互相覆盖地共享数据集合,所有用户可同时存取数据库中的数据。

2. 面向全组织的数据结构化

在数据库中,数据不再像文件系统那样从属于特定的应用,而是按照某种数据模型组织成为一个结构化的整体。它不仅描述了数据本身的特性,而且也描述了数据与数据之间的种种联系,这使数据库具备复杂的结构。

数据结构化,有利于实现数据共享。比如,一个学校,可以把学校所有的各个应用(人事、学籍、科研、财务、后勤等)的数据组织到一个数据库中,并且结构化。数据实现集中、统一的存储与管理,各种应用存取各自相关的数据子集,满足各种应用要求,实现数据共享。

3. 数据独立性

文件系统中,应用程序严重依赖于数据文件,如果把应用程序使用的磁带顺序文件改成为磁盘索引文件,则必须对应用程序进行修改。而数据库技术的重要特征就是数据独立于应用程序而存在,数据与程序相互独立,互不依赖,不因一方的改变而改变另一方,这大大简化了应用程序的设计与维护的工作量。

数据库是面向整个系统的,以最优的方式服务于一个或多个应用程序(用户),实现数据共享。

4. 可控数据冗余度

数据共享、结构化和数据独立性的优点可使数据存储不必重复,不仅可以节省存储空间,而且从根本上保证了数据的一致性,这也是有别文件系统的重要特征。

从理论上讲,数据存储完全不必重复,即冗余度为零,但有时为了提高检索速度,常有意安排若干冗余,这种冗余由用户控制,称为可控冗余度。可控冗余要求任何一个冗余的改变都能自动地对其余冗余加以改变,这个过程叫做传播更新。

5. 统一数据控制功能

数据库是系统中各用户的共享资源,因而计算机的共享一般是并发的,即多个用户同时使用数据库。因此,数据库管理系统必须提供以下四个方面的数据控制功能,保证整个系统的正常运转。

(1) 数据安全性控制

数据的安全性是指采取一定安全保密措施确保数据库中的数据不被非法用户存取而造成数据的泄密和破坏。

(2) 数据完整性控制

数据的完整性指数据的正确性、有效性与相容性。系统要提供必要的功能,保证数据库中的数据在输入、修改过程中始终符合原来的定义和规定。

(3) 并发控制

当多个用户并发进程同时存取,修改数据库中数据时,可能会发生互相干扰而得到错误结果,并使数据库完整性遭到破坏,因此必须对多用户的并发操作加以控制和协调。

(4) 数据恢复

当系统发生故障造成数据或当对数据库数据的操作发生错误时,系统能进行应急处理,把数据库恢复到正确状态。

数据库系统中应用程序与数据库之间的关系如图 1.4 所示。

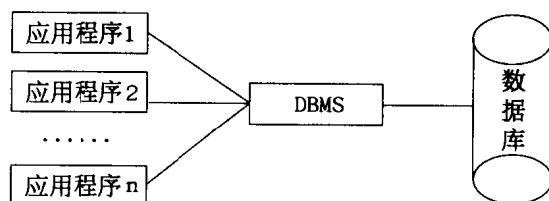


图 1.4 数据库系统中应用程序与数据库之间的关系

1.5 数据模型

前面我们已经提到,数据库中的数据是结构化的,即数据库要考虑如何去组织数据,考虑如何表示数据及其数据之间的联系,并将其合理地存放到计算机中。数据模型是指用图解方法说明数据库的数据结构,即描述数据及其数据之间联系的结构形式。

1.5.1 数据之间的联系

1. 三个世界的划分

“三个世界”即现实世界、信息世界和计算机世界。存在于人们头脑之外的客观世界称为现实世界;信息世界是现实世界在人们头脑中的反映,人们把它用文字和符号记载下来;信息世界中的信息在计算机世界是以数据形式存储的。

在现实世界中,各种事物或事件千差万别且都是彼此关联的,任何事物或事件都不是独立存在的。数据库是模拟现实世界中某应用环境(一个企业、单位或部门)所涉及的信息(各种图、表、单据等)集合,是一个统一的、集中的数据管理机构。这个集合或者包含了信息的一部分,或者包含了信息的全部。这种模拟是通过数据模型来进行的。一个好的数据模型必须能够反映现实世界中各种事物及其复杂的联系。现实世界中各种事物及其

复杂的联系是不能直接输入到计算机的数据库中的,因此,在实际的数据处理中,我们首先将现实世界的事物及其联系抽象为信息世界的信息模型,然后再将信息模型抽象转换为计算机世界的数据库模型。所以说,数据库模型是现实世界的两级抽象的结果。

在图 1.5 中,我们给出了数据处理的抽象和转换过程。从图示中可以看出,在数据处理中,数据加工经历了现实世界(Real World)、信息世界(Information World)和计算机世界(Computer World)三个不同的世界,经历了两级抽象和转换。

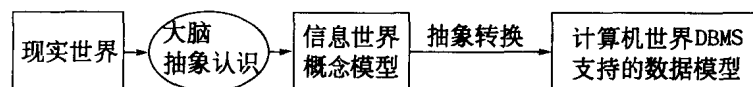


图 1.5 三个世界的关系

2. 第一级抽象

现实世界是存在于人们头脑之外的客观世界,现实世界的事物反映到人们头脑中来,经过人们头脑的认识、选择、命名、分类之后进入信息世界。信息世界是现实世界在人们头脑中的反映和抽象。现实世界的事物在信息世界中被抽象为“实体”。

实体可定义为任何可以区别的“事物”。可以指人,也可以指物,可以指实际的东西,也可以指概念性的东西,还可指事物与事物之间的联系。例如,一个学生、一辆吉普车、一个学校、一次演出、一次订货等都是实体(即实体个体)的例子。性质相同的同类实体的集合称为“实体集”,也叫实体整体。例如,所有的学生、所有的吉普车、所有的学校、所有的演出、所有的课程、所有的零件都可称为实体集。

实体的某一特征称为“属性(Attribute)”。例如,学生实体有姓名、年龄、性别……等方面的属性。属性有“型”和“值”的区分,例如属性名:姓名、年龄、性别等是属性的型,而其值是具体的内容,如李一、18、男分别是姓名、年龄和性别的值。可见,若干个属性值所组成的集合可表征一个实体,若干个属性型所组成的集合可表征一个实体的类型,简称“实体型”。同类型的实体集合组成“实体集”。

现实世界事物之间的联系也同样要抽象和反映到信息世界来。现实世界事物之间的联系是复杂的,在信息世界中将被抽象为实体型内部之间的联系(即属性间的联系)和各种实体型之间的联系(也称实体之间的联系)。反映实体型集合及其联系的结构形式可称为实体模型。实体模型也就是信息模型,它是现实世界事物及其联系的抽象表示。实体型之间的联系有如下三种类型:

(1) 一对一联系(1:1)

实体型 A 中的一个实体至多与实体型 B 中的一个实体相对应(相联系),反之亦然,则称实体型 A 与实体型 B 的联系为一对一的联系。比如公司与经理、班长与班、病人与床位、旅客与车票之间都是一对一联系,如图 1.6(a)所示。

(2) 一对多联系(1:m)

实体型 A 中的一个实体与实体型 B 中的多个实体相对应,反之,实体型 B 中的一个实体至多与实体型 A 中的一个实体相对应,则称实体型 A 与实体型 B 的联系为一对多的联系(其逆是多对一),如图 1.6(b)所示。比如,父亲对子女、学校对系、班级对学生、公司