

21 世纪高等院校计算机系列规划教材

Computer

□ 熊才权 主编

数据库原理及应用



华中科技大学出版社
<http://www.hustp.com>

TP311.13/329

2008

21 世纪高等院校计算机系列规划教材

数据库原理及应用

主 编 熊才权

副主编 李志辉 曾 玲 江 南

康瑞华 何 花

华中科技大学出版社

中国·武汉

图书在版编目(CIP)数据

数据库原理及应用/熊才权 主编. —武汉:华中科技大学出版社,2008年2月
ISBN 978-7-5609-4361-9

I. 数… II. 熊… III. 数据库系统 IV. TP311.13

中国版本图书馆 CIP 数据核字(2007)第 196908 号

数据库原理及应用

熊才权 主编

责任编辑:张毅

封面设计:刘卉

责任校对:代晓莺

责任监印:周治超

出版发行:华中科技大学出版社(中国·武汉)

武昌喻家山 邮编:430074 电话:(027)87557437

录排:龙文排版工作室

印刷:华中科技大学印刷厂

开本:787mm×1092mm 1/16

印张:22.75

字数:563 000

版次:2008年2月第1版

印次:2008年2月第1次印刷

定价:37.50元

ISBN 978-7-5609-4361-9/TP·646

(本书若有印装质量问题,请向出版社发行部调换)

前 言

数据库是一门在各行各业均有广泛应用的技术,是计算机科学与技术的重要分支,是高校计算机科学与技术专业的必修课之一。数据库技术的出现极大地促进了计算机应用的发展,目前基于数据库技术的计算机应用已成为计算机应用的主流。数据库技术本身也在不断地发展和完善,关系数据库已取代了早期的层次数据库与网状数据库,成为主流数据库,而新一代数据库也在不断地开发。

为了适应教学与科研需要,我们在多年从事数据库系统教学与科研的基础上编写了这本书。全书以关系数据库为重点,全面系统地介绍了数据库的基本概念和基本原理,并结合实际应用,详细介绍了数据库设计的基本方法,取材上力图反映当前数据库技术的发展水平和发展趋势。

本书共 10 章:第 1 章介绍数据管理技术的发展历史,数据、数据库和数据模型的基本概念,以及数据库系统的基本原理;第 2 章讲解关系数据库的基本概念,包括关系模型的数据结构、完整性约束以及关系操作,介绍了用代数方式或逻辑方式来表达的关系语言、关系代数、元组关系演算和域关系演算;第 3 章结合一个数据库实例,详细地介绍了数据库语言 SQL 语法结构及其用法,并结合 SQL 语言进一步讲述关系数据库中的基本表、视图等基本概念;第 4 章讲述关系数据理论,内容包括函数依赖及 Armstrong 公理系统、关系规范化及模式分解;第 5 章以数据库概念结构设计和逻辑结构设计为重点,讲述了数据库设计的基本步骤和基本方法;第 6 章详细讨论了事务的概念,以及事务概念在数据库恢复和并发控制中的应用,介绍了数据库恢复和并发控制的一般原理和方法;第 7 章介绍安全性和完整性等数据库保护技术;第 8 章从数据库管理系统的背景出发,通过与文件系统的对比,介绍了数据库管理系统的目标、数据库管理系统的目标和基本功能;第 9 章介绍数据库编程技术,以及如何利用编程工具和语言访问、连接以及操纵后台数据库;第 10 章介绍数据仓库与联机分析处理技术,为数据库的进一步应用打下基础。

为了帮助读者迅速掌握当前流行商用数据库管理系统的使用,附录 A 提供了简明 SQL Server 2000 教程,分三节讲解了商业数据库管理系统 SQL Server 2000 的基本组成、基本功能和使用方法。为了配合数据库原理课程的实验教学,附录 B 提供了数据库实验教程,其中有 9 个验证型实验,1 个综合设计型实验。通过实验课教学可以进一步加深读者对数据库系统基本概念的理解,使其学会数据库应用系统的基本设计与开发方法。书中含有丰富的例题与习题,便于教学与自学。

全书计划 64 学时,其中实验 18 学时,授课教师可以根据实际情况和教学的具体要求,进行适当的取舍。

本书由熊才权、李志辉、曾玲、江南、康瑞华、何花、王海燕等编写,全书由熊才权负责统稿。

由于水平有限,书中难免存在不足或错误之处,恳请专家和读者批评指正。

编 者

2007-8-21

目 录

(041)		1.2
(151)		2.2
(061)		3.2
(821)		7.2
第 1 章 数据库系统概述		(1)
1.1 数据与数据管理		(1)
1.2 数据库系统基本概念		(6)
1.3 数据库系统的模式结构		(11)
1.4 数据模型		(14)
1.5 小结		(27)
习题 1		(28)
第 2 章 关系数据库		(30)
2.1 关系模型概述		(30)
2.2 关系的完整性		(34)
2.3 关系代数		(36)
2.4 关系演算		(45)
2.5 查询优化		(48)
2.6 小结		(52)
习题 2		(52)
第 3 章 SQL 语言		(54)
3.1 SQL 概述		(54)
3.2 数据定义		(58)
3.3 SQL 数据查询功能		(62)
3.4 视图操作		(88)
3.5 数据操作		(93)
3.6 数据控制		(97)
3.7 嵌入式 SQL		(100)
3.8 小结		(103)
习题 3		(104)
第 4 章 关系数据理论		(107)
4.1 问题的提出		(107)
4.2 规范化		(109)
4.3 数据依赖的公理系统		(117)
4.4 模式的分解		(123)
4.5 小结		(128)
习题 4		(128)
第 5 章 数据库设计		(131)
5.1 数据库设计概述		(131)
5.2 需求分析		(135)
5.3 概念结构设计		(139)

5.4	逻辑结构设计阶段	(149)
5.5	物理结构设计阶段	(154)
5.6	数据库的实施	(156)
5.7	数据库的运行和维护	(158)
5.8	基于 3NF 的泛关系数据库设计方法	(159)
5.9	PowerDesigner 辅助设计	(159)
5.10	小结	(165)
习题 5		(166)
第 6 章	事务处理技术	(167)
6.1	事务	(167)
6.2	数据库恢复	(169)
6.3	并发控制	(177)
6.4	小结	(189)
习题 6		(190)
第 7 章	安全性和完整性	(193)
7.1	安全性	(193)
7.2	SQL Server 的安全性	(195)
7.3	完整性	(200)
7.4	SQL Server 完整性控制	(202)
7.5	小结	(205)
习题 7		(205)
第 8 章	数据库管理系统	(207)
8.1	DBMS 简介	(207)
8.2	DBMS 系统结构	(210)
8.3	DBMS 磁盘和文件管理	(212)
8.4	索引技术	(216)
8.5	查询处理	(221)
8.6	数据库安全和认证	(226)
8.7	小结	(227)
习题 8		(227)
第 9 章	数据库编程	(230)
9.1	嵌入式编程	(230)
9.2	游标	(235)
9.3	存储过程与触发器	(237)
9.4	面向数据库的中间件	(251)
9.5	小结	(265)
习题 9		(265)
第 10 章	数据仓库技术	(267)
10.1	数据仓库技术产生的背景	(267)
10.2	数据仓库概述	(268)

10.3 联机分析处理.....	(276)
10.4 SQL Server 2000 中数据仓库工具.....	(288)
10.5 小结.....	(289)
习题 10.....	(290)
附录 A 简明 SQL Server 2000 教程.....	(292)
A.1 SQL Server 2000 常用工具介绍.....	(292)
A.2 SQL Server 2000 数据库管理.....	(295)
A.3 SQL Server 2000 表、索引和视图.....	(302)
附录 B 实验指导.....	(311)
实验一 数据库的创建、修改与删除.....	(311)
实验二 基本表的创建、修改与删除.....	(315)
实验三 数据更新操作.....	(320)
实验四 SQL 数据查询与索引.....	(323)
实验五 视图定义与管理.....	(327)
实验六 数据库的安全性控制.....	(330)
实验七 数据库的完整性控制.....	(334)
实验八 数据库的备份与恢复.....	(340)
实验九 数据库应用开发基础.....	(346)
参考文献.....	(353)

第1章 数据库系统概述

当今社会是一个信息化的社会,信息已经成为社会上各行各业的重要资源。数据是信息的载体,数据库是互相关联的数据集合。数据库能利用计算机保存和管理大量复杂的数据,快速而有效地为多个不同的用户和应用程序提供数据,帮助人们有效利用数据资源。以数据处理为研究对象的数据库技术正迅速发展,并得到广泛应用。

数据库的应用来自于已发展了数十年的数据处理技术,这些数据处理技术蕴藏在被称为数据库管理系统的专业化软件中。引入数据库后的计算机系统称为数据库系统,它由数据库、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等组成,其中数据库是系统的核心和基础。本章介绍数据、数据库和数据模型的基本概念以及数据库系统的基本原理。

1.1 数据与数据管理

数据库是计算机信息管理的基础,其研究对象是数据。因此,在介绍数据库技术之前,有必要了解数据与信息的基本概念和数据管理技术的发展历程。

1.1.1 数据与信息

提到数据,人们往往会想到信息。数据中隐含着信息,但是数据并不是信息本身。数据是对事实和概念的描述,它的最初表示形式是符号,不同的事实或概念用不同的符号表示,人们通过对符号的辨识获取不同的数据。描述事物的符号可以是数字,也可以是文字、图形、图像、声音、语言等,它们通过数字化存入计算机中。

对一个事物的描述,往往需要多个符号,不同的符号代表事物的不同特征。例如,在人事档案管理中,如果人们对一个员工最感兴趣的是员工的姓名、性别、家庭住址、进入公司时间、所属部门,那么就可以这样描述:

(谭林,男,南区6#,2001年12月,研发部)

这里的员工记录就是一个数据。其中的姓名、性别、家庭住址、进入公司时间、所属部门等称为数据项,它们本身也是数据。

一般来说,数据库系统中的数据可以有两种类型。一种是作业层数据,如产品数量、财务收支等。这类数据会引起数据库的频繁操作,它反映了现实世界中的日常活动,这些活动是必须的、重复的、可以预见和计划的。另一种是管理控制层数据,这种数据主要用来做统计、分析、预测等,它们主要来自于计划和决策部门,是通过对作业层数据进行分析和处理得到的数据。

信息是对数据的理解或解释,是通过对数据进行处理、加工、提炼而得到的能为人所理解和交流的知识。在现实世界中,人们一般用自然语言表示和交流信息。而在计算机中,为了表示信息,必须对数据赋予一定的含义,数据的含义称为数据的语义。例如,对上面的员工数据,了解其含义的人会得到如下信息:谭林是某公司的男性员工,住在南区6号,2001

年 12 月进入公司,在研发部工作。不了解其语义的人则无法理解其含义,也获取不了信息。可见在信息系统中,数据与其语义是密不可分的。一个信息系统一般有多个数据,数据与数据之间有复杂的联系,要从数据中获取信息,不仅要了解数据的语义,还要了解数据之间的联系,这也是信息系统所要解决的问题。

信息与数据是相互关联的,数据是信息的载体,而信息是数据的内涵。

1.1.2 数据管理技术的发展

每个组织都保存了大量各种各样的数据。例如,企业有关于生产计划、生产调度、生产工具、物质供应、经营销售等方面的数据;学校有关于学生和教职工档案、教学计划、学生成绩等方面的数据;医院有关于病历、药品、病房、财务等方面的数据。数据是一个组织的重要资源,有时甚至比其他资源更珍贵,因此必须对组织的各种数据实行有效管理。所谓数据管理,是指对数据进行收集、整理、存储、检索、加工和传递等一系列活动的总和。数据管理的最终目的是从数据中获取有用的信息,以便服务于组织的管理工作。从原始数据中获取信息,或将原始数据转换成信息的过程就是数据处理。数据处理是数据管理的中心工作,也是信息系统的根本任务。

可以说人类社会自从有了组织,或者说自从有了管理工作,就面临数据管理的任务,数据管理贯穿于一个组织的管理工作的全过程,没有数据管理的管理工作是不存在的。我们平时所说的用事实说话、用数字说明问题,指的就是这个意思。根据数据管理工具和管理技术的发展历史,数据处理的方式大致可以分为人工式(1800 年以前)、机械辅助式(1800—1890 年)、机电穿孔卡片式(1890—1946 年)和电子计算机式(1946 年以后)几个阶段。利用计算机进行数据管理的历史虽然不长,但发展迅速,尤其是数据库技术应用以来,计算机处理数据的能力和范围大为提高。计算机数据管理技术经历了人工管理、文件系统和数据库系统三个阶段。

1. 人工管理阶段

20 世纪 50 年代中期以前,计算机数据管理的能力很差,这一阶段称为人工管理阶段。那时计算机没有磁盘等直接存储设备,没有操作系统,没有管理数据的软件,数据处理方式是批处理,数据管理的任务主要由应用程序员自己承担,系统所提供的数据管理功能仅仅是一些简单的 I/O 操作,如图 1.1 所示。

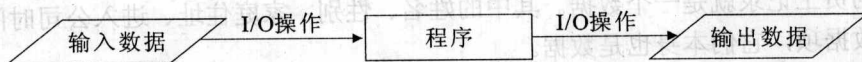


图 1.1 程序中数据的输入与输出

在人工管理阶段,不同的应用程序处理不同的数据,数据与程序之间是一一对应的关系,如图 1.2 所示。

人工管理数据存在以下缺点。

(1) 数据不保存。

当时的计算机主要用于科学计算,一般不需要将数据长期保存,只是在计算某一具体题目时将数据输入,运行完后得到输出结果,输入、输出和中间数据都不保存。这与信息系统



图 1.2 人工管理阶段数据与程序之间的对应关系

中对数据的管理思想是完全不一样的，因为在信息系统中，数据作为一个重要资源，不仅要对其进行加工处理，还要予以保存，以便以后检索和修改。

(2) 数据不共享。

在人工管理阶段，数据是面向应用的。一组数据只对应一个应用程序，当多个应用程序要用到相同数据时，必须各自定义，不能共享。

(3) 数据冗余度大。

由于数据不能共享，必然会出现相同数据的多个副本，不同的副本对应不同的应用程序。这样会导致数据冗余度大。

(4) 数据缺乏独立性。

数据与程序是紧密结合在一起的，数据的逻辑结构、物理结构、存取方式都由程序规定，当数据的逻辑结构、物理结构、存取方式发生变化时，必须对应用程序作相应修改，因此程序员的负担很重。

(5) 数据的不统一性。

在一个组织中，要想将数据作为一种资源共享，必须对数据的命名、格式、存取方式等标准进行统一规定。但是在人工管理阶段，数据与程序紧密结合，不同应用程序会对同一数据做不同的定义，因而往往会出现“同名异物”和“同物异名”的现象。

2. 文件系统阶段

20 世纪 50 年代到 60 年代中期，计算机数据管理技术进入文件系统阶段。这时已经有了磁盘、磁鼓等直接存储的硬件设备，也有了专门的数据管理软件，一般称为文件系统；处理方式上不仅有了批处理，而且还有了能够实现联机实时处理的处理方式。

在文件系统阶段，文件系统把数据组织成相互独立的数据文件，数据可以长期保存在存储设备上，应用程序利用“按文件名访问，按记录进行存取”的管理技术，可以对文件中的数据进行修改、插入和删除等操作。文件系统实现了一定的数据独立性，它将数据的逻辑结构与物理结构分离，应用“存取方式”实现逻辑结构与物理结构之间的映射，如图 1.3 所示。

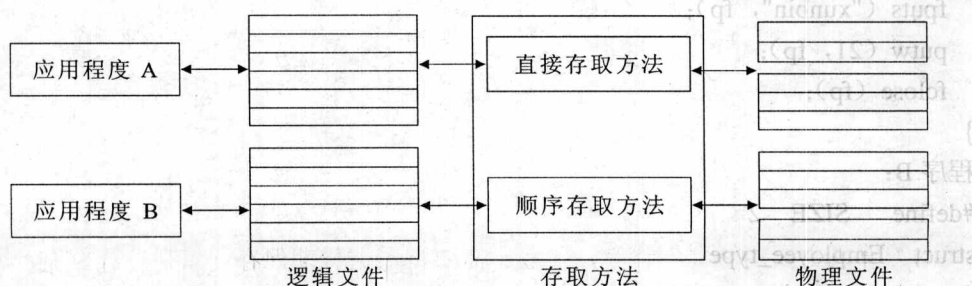


图 1.3 文件系统的数据处理

在这种方式下,应用程序只涉及数据的逻辑结构,当数据的物理结构改变时,不会导致应用程序的修改,这就是数据的物理独立性。数据的物理独立性使应用程序脱离了数据的物理结构,也使其适应性得以提高。同时,程序员在编写程序和对程序进行维护时不必过多考虑数据的物理存储细节,只需将精力集中在算法上,因而生产效率得以大大提高。另外,数据的物理独立性使数据共享成为可能,多个应用程序可以对同一文件进行操作。

与早期的人工管理阶段相比,文件系统已经有了很大的进步。但文件系统管理数据还存在以下不足。

(1) 数据的共享性较差。

文件系统提供了数据的物理独立性,实现了一定的数据共享,但它只能实现文件级共享而不能在记录或数据项级实现数据共享。

(2) 数据的冗余度较大。

在文件系统中,文件的逻辑结构是根据它的应用而设计的,数据的逻辑结构与应用程序之间相互依赖。即使不同应用程序具有部分相同的数据(记录或数据项),也必须构造各自的文件,这样就存在大量的数据冗余,浪费大量的存储空间。

(3) 数据存在不一致性。

数据的冗余度大与数据的不一致性是密切相关的。同一数据在多个地方同时存放,同一数据在不同存放地的值可能不相同,这会降低信息的价值,有时甚至会造成重大损失。

(4) 数据的独立性较差。

文件系统只实现了数据的物理独立,而没有实现数据的逻辑独立。数据的逻辑结构对应一个特定的应用,当应用发生变化时,数据的逻辑结构也要发生改变,当数据的逻辑结构发生变化时,程序也要做相应的修改。因此,文件系统的数据与程序之间缺乏逻辑独立性。

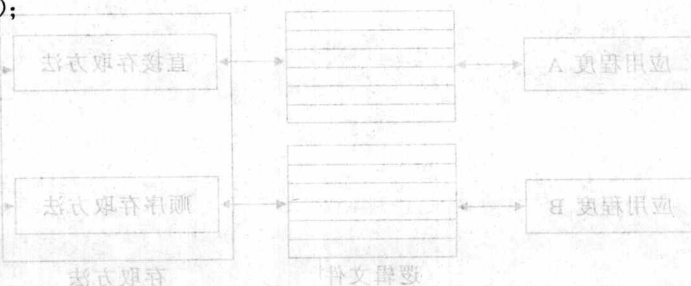
为了说明上面的问题,现在来看如下一段 C 语言程序。

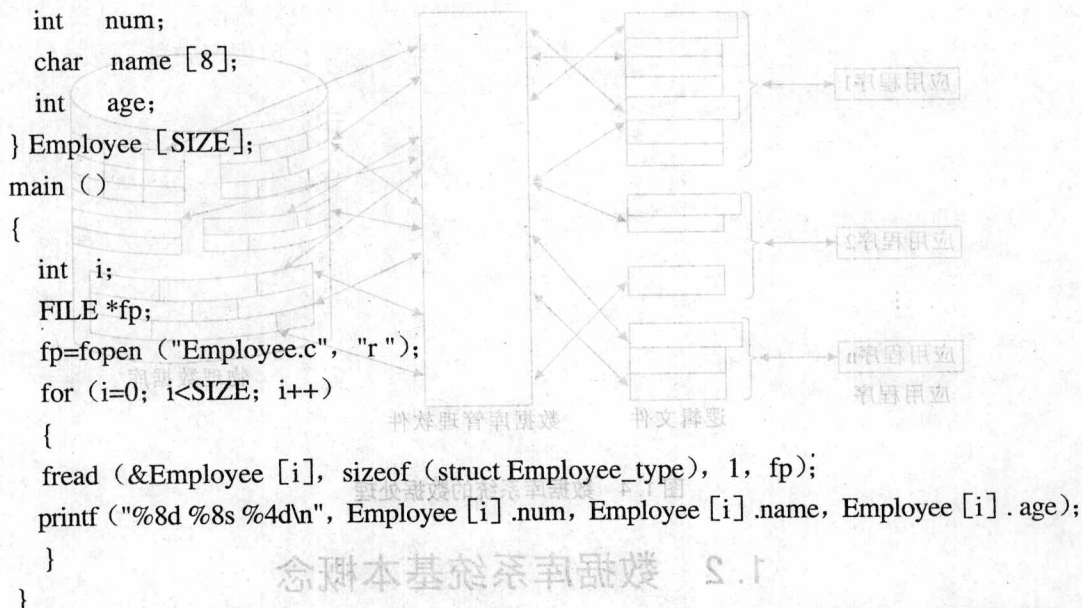
```
程序 A:
#include "stdio.h"
main ( )
```

```
{
FILE *fp;
fp = fopen ("Employee.c", "w");
putw (200005, fp);
fputs ("tanlin", fp);
putw (20, fp);
putw (200009, fp);
fputs ("xunbin", fp);
putw (21, fp);
fclose (fp);
}
```

程序 B:

```
#define SIZE 2
struct Employee_type
{
```





这是对文件进行操作的一段程序，文件中保存员工记录。程序 A 将两个员工的员工号、姓名、年龄输入到一个文件中，程序 B 从文件中读取记录，存入变量中。程序 B 中结构型 Employee 的设计必须与文件的逻辑结构一致，即程序设计与文件逻辑结构是紧密相关的。

可以看到，程序 B 对文件操作时必须知道文件的逻辑结构，或者说，若文件的逻辑结构发生了变化，就必须修改程序。而在关系数据库系统中，只需用一个 SQL（结构化查询语言）语句即可完成记录的插入和读取，只要不违反数据库完整性和安全性约束，写 SQL 语句完全无需考虑数据文件的逻辑结构，而数据库完整性和安全性控制是由数据库管理系统自动实现的。

3. 数据库系统阶段

20 世纪 60 年代后期以来，计算机数据管理技术开始进入数据库系统阶段。这时计算机技术发展迅速，硬件方面有了大容量磁盘，硬件价格下降；软件方面出现了包括操作系统在内的大量的系统软件；在处理方式上，联机实时处理要求增多，并开始提出和考虑分布式处理方法。数据库系统阶段开始的标志是产生了一种称为数据库管理系统的专门用于数据管理的软件。

数据库系统的产生是企业海量数据处理需求的必然结果。随着计算机用于管理的规模越来越大，应用越来越广泛，数据量急剧增大，企业对数据管理技术的要求也越来越高。首先，企业要求数据作为企业组织的公共资源而集中管理控制，为企业的各种用户共享，因此，应大量地消去冗余数据，节省存储空间。其次，当数据变更时，能减少对多个数据副本的多次变更操作，从而可大大节省计算时间，更为重要的是，不会因遗漏某些副本的变更而使系统给出一些不一致的数据。再次，要求数据具有更高的独立性，不但要具有物理独立性，而且还要具有逻辑独立性，即当数据逻辑结构改变时，不影响用户的应用程序，从而降低应用程序开发和维护的成本。所有这些，是文件系统所不能满足的，而数据库管理系统却可以做到。

与文件系统不同，数据库系统是面向数据的而不是面向程序的，数据是系统的中心，各处理功能处于外围，它们通过数据管理软件从数据库中获取所需数据和存储处理结果。按数据库的方法，数据处理的过程如图 1.4 所示。

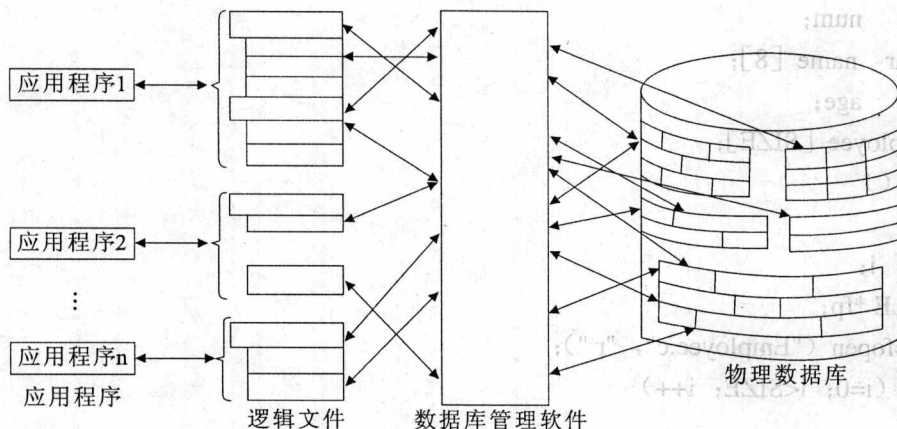


图 1.4 数据库系统的数据处理

1.2 数据库系统基本概念

随着数据管理技术的不断发展和计算机应用的普及，数据库已经成为很多人熟悉的概念和术语，但是不同的人对数据库的理解并不相同。在系统介绍数据库原理之前，有必要先介绍数据库、数据库管理系统、数据库系统等几个常用的基本概念。

1.2.1 数据库

简单地说，数据库是存在一定联系的数据的集合，它可以人工地建立、维护和使用，也可以通过计算机建立、维护和使用。当然，本书关心的是后者，即计算机化的数据库。因此，我们定义数据库（database，简称 DB）为，长期存储在计算机内的、相互联系的数据集合，它按一定的数据模型组织、描述和存储，具有较小的冗余度、较高的数据独立性和易扩展性，并可为各种用户共享。数据库一般都通过应用程序或数据库管理系统来建立、维护和使用。

数据库具有以下特点。

- （1）数据库是具有逻辑关系和确定含义的数据集合。逻辑上无关的数据集合不能称为数据库。
- （2）数据库是针对明确的应用目标而设计、建立和加载的，并为这些用户的应用服务。
- （3）数据库是对一个现实世界（如一个单位或组织）的映像，现实世界的某些改变必须及时地反映到该数据库中来。

1.2.2 数据库管理系统

数据库管理系统（database management system，简称 DBMS）是一个位于用户与操作系统之间的数据管理软件。DBMS 的目标是为用户提供一个能方便、快速、有效地建立、维护、检索、存取和处理数据库中的数据的环境。DBMS 能够对数据库进行有效的管理，包括存储管理、安全性管理、完整性管理等，其主要功能包括以下几个方面。

(1) 持久存储数据。

DBMS 支持对独立于应用程序的超大数据量（吉字节或更多）数据长期存储，其数据独立性优于文件系统，并能防止对数据的意外和非授权的访问，且在数据库查询和更新时支持对数据的有效存取。

(2) 数据定义功能。

DBMS 允许用户使用专门的数据定义语言（data definition language，简称 DDL）对数据库中的数据对象进行定义，如定义或删除模式、索引、视图等，并能保证数据库完整性。

(3) 数据操纵功能。

DBMS 提供合适的查询语言（query language）或数据操纵语言（data manipulation language，简称 DML），用户使用 DML 可以实现对数据库的基本操作，如查询、插入、删除和修改数据等。

(4) 事务管理。

DBMS 支持对数据的并发存取，即可以同时有很多不同的进程（称为“事务”）对数据访问，为了避免存取错误数据，DBMS 必须提供一种机制保证事务正确执行。

(5) 数据库的运行管理。

数据库在建立、运用和维护时由 DBMS 统一管理、统一控制，以保证数据的安全性、完整性和多用户对数据库使用的并发控制及发生故障后的系统恢复等。

(6) 数据库维护功能。

它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，数据库的重组织功能和性能监视、分析功能等。

DBMS 是数据库系统的一个重要组成部分。DBMS 核心技术的研究和实现是三十余年来数据库领域所取得的主要成就。我国对 DBMS 的研制时间不长，但发展迅速，目前已有国产 DBMS 产品走向商业应用。

1.2.3 数据库用户和管理员

数据库系统的一个主要目标是从数据库中检索数据和往数据库中存储新数据。使用数据库的人员可分为数据库用户和数据库管理员两大类。

1. 数据库用户

根据用户与系统交互方式和使用目标不同，数据库用户分为偶然用户、初级用户、高级用户、应用程序员等几类。

(1) 偶然用户。这类用户不经常访问数据库。他们每次访问数据库时需求比较单一，一般通过事先设置好的窗口与数据库进行交互。例如，一个用户想通过互联网查询其银行账户上的余额。这个用户会访问一个用来输入他的账号和密码的窗口。位于服务器上的一个应用程序就用账户的号码取出账户的余额，并将这个信息返回给用户。此类用户的典型用户界面是窗口，用户只需要填写窗口相应项就可以使用数据库。对于企业，偶然用户还有企业的高级管理人员。

(2) 简单用户。大多数数据库用户都是简单用户。他们的主要工作是查询和更新数据库，一般通过事先设计好的应用系统与数据库进行交互。例如，银行出纳员需要将账户 A 的 100

元转入账户 B 时,启动银行转账系统,调用一个转账程序。该程序要求出纳员输入转账金额、转出的账户以及转入的账户。简单用户一般都不直接使用 DBMS,而是通过程序员精心设计并具有友好界面的应用程序存取数据库。银行的职员、航空公司的机票预订工作人员、旅馆总台服务员等都属于这类用户。

(3) 高级用户。这类用户不通过应用程序与数据库进行交互,而是用数据库查询语言来表达他们的要求,有时还使用联机分析处理(OLAP)和数据挖掘(DM)来发现数据库中的其他模式。高级用户包括工程师、科学家、经济学家、科学技术工作者等具有较高科学技术背景的人员。

(4) 系统分析员:系统分析员负责分析数据库用户特别是简单用户的需求,确定用户所需要的数据,给出适应这些用户需求的数据库模式定义、文件结构设计、存取方式选择等。系统分析员一般与数据库管理员合作工作。

(5) 应用程序员。应用程序员是编写供多数人使用数据库的应用程序的计算机专业人员。应用程序员可以选择多种工具来开发满足用户要求的用户界面。大多数主要的商业数据库系统都提供了快速应用开发工具。应用程序员必须十分熟悉整个 DBMS,以完成他们的任务。

2. 数据库管理员

在任何—个组织机构中,如果有很多人共享相同资源,则需要有一个特殊的人员来监督和管理这个共享资源。在数据库系统环境下,共享资源有两类:第一类是数据库,第二类是 DBMS 软件和相关软件。这些资源的监督管理由数据库管理员(DBA)完成。数据库管理员可以由一个人担任,也可以由一组人担任。数据库管理员的职责如下。

(1) 模式定义。数据库管理员用 DBMS 中的数据定义语言约定的一系列定义来创建最初的数据库模式。模式的概念将在下节介绍。

(2) 数据存储结构和存储方式定义。

(3) 模式和存储结构的修改。由数据库管理员对模式和存储结构进行修改,以反映组织的需求变化,或为提高性能选择不同的存储结构。

(4) 数据访问授权。通过授权管理,数据库管理员可以规定不同的用户各自访问数据库中不同数据,从而保障数据库安全。授权信息保存在一个特殊的系统结构中,一旦系统中有访问数据的要求,数据库系统就去查阅这些信息。

(5) 日常维护。日常维护是数据管理员经常性的工作,主要维护活动有:定期将数据库备份在磁带或磁盘上或在远程服务器上,以防止灾难发生时数据库丢失;确保运转时所需的空余磁盘空间,并且在需要时升级磁盘空间;监视数据库的运行,确保数据库性能不因一些用户提交了需花费时间较多的任务而下降很多。

1.2.4 数据库系统的组成

数据库系统(database system,简称 DBS)是指在计算机系统中引入数据库后的系统,一般由数据库、DBMS(及其开发工具)、应用系统、数据库用户和管理员构成。其中数据库是系统的核心,DBMS(及其开发工具)和数据库管理员是系统的基础,应用系统和用户是系统服务的对象。在不引起混淆的情况下数据库系统常常简称为数据库。

数据库系统的组成如图 1.5 所示。数据库系统在整个计算机系统的位置如图 1.6 所示。

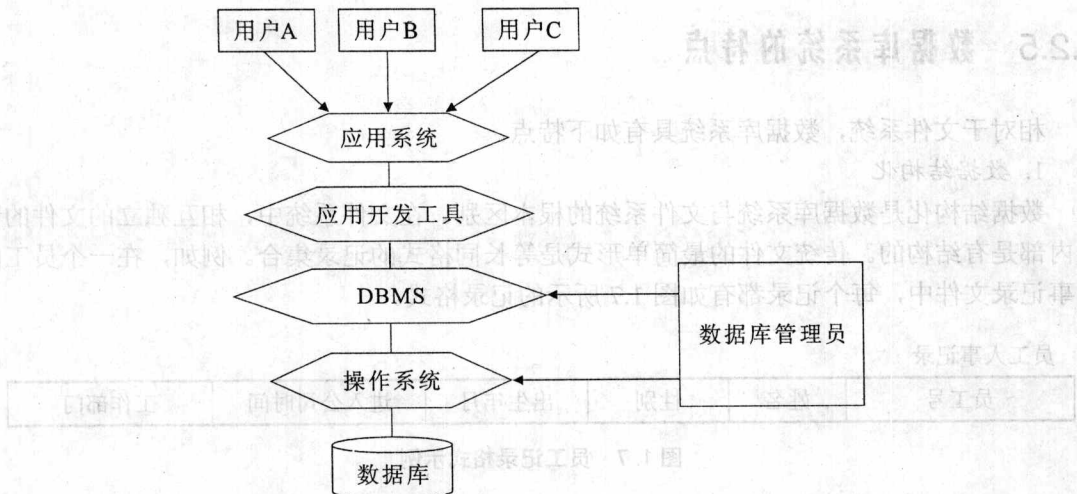


图 1.5 数据库系统的组成

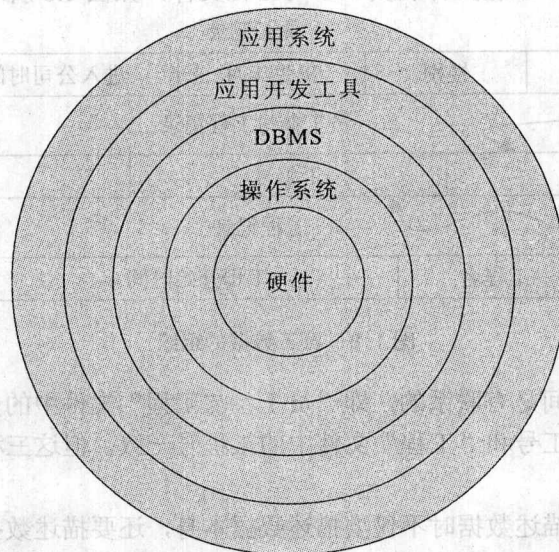


图 1.6 数据库系统在整个计算机系统的位置

数据库系统对硬件的要求除了一般计算机系统对硬件的要求外，还要求有足够大的内存，以便存放操作系统、DBMS 的核心模块、数据缓冲区和应用程序；有容量足够大的磁盘等直接存取设备，以便存放数据和作数据备份；有较高的数据传输速率。

数据库系统的软件主要有 DBMS、支持 DBMS 运行的操作系统、具有与数据库接口的高级程序设计语言及其编译程序、以 DBMS 为核心的应用开发工具和为特定应用开发的数据库应用系统。

数据库用户和数据管理员是数据库系统的重要组成部分，他们的作用是开发、管理和使用数据库系统。不同人员涉及不同的数据抽象级别，对应不同的数据视图。

1.2.5 数据库系统的特点

相对于文件系统，数据库系统具有如下特点。

1. 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，在一个员工的人事记录文件中，每个记录都有如图 1.7 所示的记录格式。

员工人事记录

员工号	姓名	性别	出生年月	进入公司时间	工作部门
-----	----	----	------	--------	------

图 1.7 员工记录格式示例

一个文件只能面向一个应用，而一个管理信息系统则涉及许多应用。在数据库系统中，不仅要考虑某个应用的数据结构，还要考虑整个组织的数据结构。例如，在一个企业管理信息系统中不仅要考虑员工管理，还要考虑工程管理、销售管理等，就工程管理来说，其涉及三个文件，即员工文件、工程文件和员工参与工程文件，如图 1.8 所示。

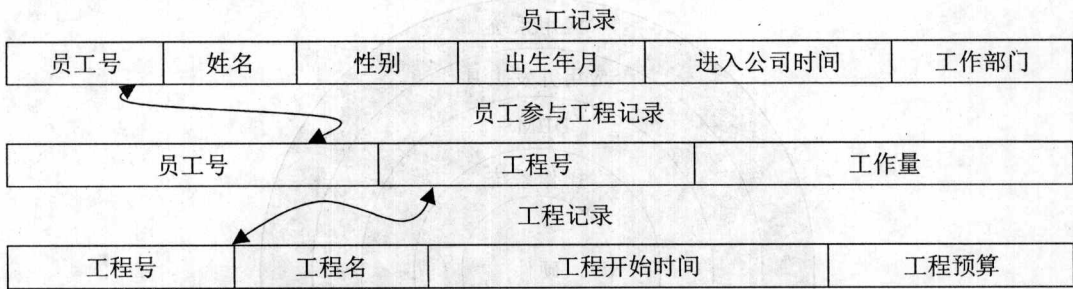


图 1.8 员工数据的组织

这三种记录相互之间是有联系的，即“员工参与工程”文件中的员工号和工程号应分别与“员工”文件中的员工号和“工程”文件中的工程号一致。但这三种记录在文件系统中又是相互独立的。

数据结构化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，尽管其记录内部有了某些结构，但记录之间没有联系，一个文件往往只针对某一特定应用，文件之间是相互独立的；数据的最小存储单位是记录，不能细到数据项。在数据库系统中，存在多个数据文件，这些数据文件之间是相互联系的，数据不再只针对某一特定应用，而是面向全组织，具有整体的结构化特点。在某一特定应用中，所用到的是结构化数据中的一个子集。数据库系统存取数据的方式也很灵活，可以存取数据库中的某一数据项或一组数据项、一个记录或一组记录。

2. 数据共享度高

数据只有实现共享才能发挥更大作用，实现数据共享是数据管理的目标。在人工管理阶段，数据无共享可言；在文件系统阶段，数据只能实现文件级共享，而不能实现系统级共享。在数据库系统中，一个数据可以为多个不同的用户共同使用，即各个用户可以为了不同的目的来存取相同的数据。在数据库系统中，还可以实现数据并发共享，即多个不同的用户可以