



普通高等学校计算机科学与技术应用型规划教材

数据库原理应用教程

主 编 俞俊甫
副主编 王钟庄
参 编 邹 璇 刘 敏



北京邮电大学出版社
www.buptpress.com

普通高等学校计算机科学与技术应用型规划教材

数据库原理应用教程

主 编 俞俊甫
副主编 王钟庄
参 编 邹 璇 刘 敏

北京邮电大学出版社
·北京·

内 容 简 介

本教材是根据普通本科院校的计算机教材要求,为突出重点、加强实践、学以致用为目的编写而成,主要是为全国大学二本、三本及大专学校有关计算机专业学生学习“数据库原理”这门课程服务。

本书共分9章,第1章数据库概论;第2章数据库系统结构;第3章关系运算;第4章标准查询语言SQL;第5章关系数据库的规范化设计;第6章E-R模型的设计方法;第7章关系数据库的设计方法;第8章数据库管理;第9章SQL Server 2005数据库管理系统介绍。

本教材的编写工作完全由多年从事二本、三本数据库原理教学的老师承担。每章后附有较多的练习题,配有上机指导书,并提供习题答案和上课电子教案。教材中较大幅度加强了对学生实践知识的培养,编写的教材理论联系实际,由浅入深、通俗易懂,便于自学。它适用于各大专院校有关计算机专业的教学用书。

图书在版编目(CIP)数据

数据库原理应用教程/俞俊甫主编.--北京:北京邮电大学出版社,2010.7

ISBN 978-7-5635-2309-2

I. ①数… II. ①俞… III. ①数据库系统—高等学校—教材 IV. ①TP311.13

中国版本图书馆CIP数据核字(2010)第132967号

书 名:数据库原理应用教程

主 编:俞俊甫

责任编辑:王丹丹 赵静

出版发行:北京邮电大学出版社

社 址:北京市海淀区西土城路10号(邮编:100876)

发 行 部:电话:010-62282185 传真:010-62283578

E-mail: publish@bupt.edu.cn

经 销:各地新华书店

印 刷:北京忠信诚胶印厂

开 本:787 mm×1 092 mm 1/16

印 张:16.25

字 数:401千字

版 次:2010年7月第1版 2010年7月第1次印刷

ISBN 978-7-5635-2309-2

定 价:29.00元

· 如有印装质量问题,请与北京邮电大学出版社发行部联系 ·

前 言

“数据库原理”这门课程是计算机有关专业的主干课程,对于这门课程的教学历来是重理论轻实践,学生(特别是大学二本和三本的学生)学完后,总觉得没学到什么知识,对于书中的理论知识也是似懂非懂,更不知道如何把书本理论应用到实际的程序设计中。针对这种情况,编写本教材时加强了实践教学内容,特别增强了数据库实际操作(本教材是应用 SQL Server 2005)知识和利用嵌入式 SQL 语言开发数据库应用系统的知识。在讲述理论知识时尽量多讲实例,避免枯燥无味的纯理论教学。在学习本课程前,应学习过“Visual Basic 程序设计”、“C 语言程序设计”、“Java 程序设计”三门课程中的一门。学完本课程后马上安排两周的工程训练,要求学生在这两周的工程训练中学会利用本课程知识,自己设计一个小型数据库应用系统。本教材编写了配套的上机实验教材,配备的后续课程是:大型数据库系统。

本书共分 9 章,第 1 章介绍数据库概论知识;第 2 章介绍数据库系统结构知识;第 3 章介绍关系型数据库中关系运算的知识;第 4 章介绍关系型数据库中的标准查询语言 SQL,其中增加了如何利用嵌入式 SQL 语言设计数据库应用系统的知识;第 5 章介绍关系数据库的规范化设计的知识;第 6 章介绍在数据库概念设计中使用的 E-R 模型的设计方法知识;第 7 章介绍关系数据库的设计方法的知识;第 8 章介绍数据库安全管理方面的知识;第 9 章介绍 SQL Server 2005 数据库管理系统的知识。

本教材由从事多年“数据库原理”教学一线老师组织编写,有的老师从事过大学一本、二本和三本学生该门课程的教学,有较强的实践教学经验。本教材由南昌大学科技学院俞俊甫教授为主编,王钟庄副教授为副主编,邹璇、刘敏讲师为编委。另外张忻、范晰、罗少彬、邓伦丹、吴赞婷、赵金萍、罗丹、熊婷、汪伟、梅毅、江东明、卢钢、兰长明、汪文俊、张剑等老师做了大量的辅助工作,并提出了许多宝贵意见。虽然编写这本教材编者们花费了很多时间,但由于水平有限,难免会产生一些缺点和错误,敬请读者批评指正,以便再版时改正。

在本书编写过程中,受到北京邮电大学出版社、各大学的各级领导、各计算机系全体教职员工的全力支持,对此我们全体编写人员对这些单位的领导和有关同志表示衷心感谢!

编 者

目 录

第 1 章 数据库概论	1
1.1 数据管理技术的发展阶段与数据库技术概念	1
1.1.1 数据管理技术的 3 个发展阶段	1
1.1.2 数据库系统阶段产生的 3 大新技术	4
1.1.3 数据库系统的主要特点	4
1.1.4 数据库技术中的几个主要名称	5
1.2 数据库技术的新发展	7
1.2.1 数据库技术研究的主要领域	7
1.2.2 数据库技术的新发展	8
本章小结	13
习题	14
第 2 章 数据库系统结构	16
2.1 数据模型.....	16
2.1.1 概念设计中的数据描述.....	16
2.1.2 数据模型的定义和组成结构数据模型的 3 要素.....	17
2.1.3 实体联系模型简述.....	19
2.1.4 结构数据模型.....	21
2.2 数据库系统结构.....	25
2.2.1 数据库系统的三级模式结构.....	26
2.2.2 数据库的二级映像功能与数据独立性.....	27
2.2.3 数据库管理系统的工作过程.....	28
本章小结	28
习题	29
第 3 章 关系运算	31
3.1 关系数据模型.....	31
3.1.1 关系数据模型的定义.....	31
3.1.2 关键码和表之间的联系.....	34
3.1.3 关系模式.....	35

3.1.4	关系模型的完整性规则	36
3.2	关系运算	37
3.2.1	关系查询语言和关系运算	37
3.2.2	关系代数运算符的分类	39
3.2.3	传统的集合运算	39
3.2.4	专门的关系运算	42
3.2.5	关系代数表达式应用举例	46
3.2.6	扩充的关系代数操作	48
3.3	关系代数表达式的查询优化	50
	本章小结	56
	习题	56
第4章	标准查询语言 SQL	59
4.1	SQL 语言概述及其数据定义	59
4.1.1	SQL 语言的基本概念及其特点	59
4.1.2	SQL 语言的数据定义	61
4.1.3	SQL 对索引的创建与删除	64
4.2	SQL 语言的数据查询	65
4.2.1	SELECT 命令的格式及其含义	65
4.2.2	单表查询	66
4.2.3	多表间连接和合并查询	73
4.2.4	嵌套查询	75
4.2.5	保存查询结果及分步查询	78
4.3	SQL 语言的数据更新与视图	79
4.3.1	插入数据	79
4.3.2	修改数据	80
4.3.3	删除数据	82
4.3.4	视图创建、删除与更新	82
4.3.5	SQL 数据控制	84
4.4	嵌入式 SQL	84
4.4.1	嵌入式 SQL 语言介绍	84
4.4.2	VB 平台下的嵌入式 SQL 语言	86
4.4.3	C# 平台下的嵌入式 SQL 语言	95
4.4.4	Java 平台下的嵌入式 SQL 语言	103
	本章小结	110
	习题	110

第 5 章 关系数据库的规范化设计	113
5.1 关系模式的设计问题	113
5.1.1 概述	113
5.1.2 关系模式存在的问题	113
5.2 规范化理论	116
5.2.1 函数依赖	116
5.2.2 码	118
5.2.3 范式	119
5.3 Armstrong 公理系统	125
5.3.1 Armstrong 公理系统推理规则	125
5.3.2 属性集的闭包	125
5.3.3 最小函数依赖集	126
5.3.4 规范化模式设计的 3 条原则	127
本章小结.....	129
习题.....	129
第 6 章 E-R 模型的设计方法	132
6.1 E-R 模型的基本元素	132
6.2 属性的基本分类	134
6.2.1 属性类别分类	134
6.2.2 属性取值特点分类	135
6.3 联系的设计方法	136
6.4 E-R 模型的扩充	140
6.5 E-R 模型实例介绍	141
本章小结.....	144
习题.....	145
第 7 章 关系数据库的设计方法	147
7.1 数据库设计概述	147
7.1.1 数据库设计的内容	147
7.1.2 数据库设计的特点	148
7.1.3 数据库设计的方法简述	148
7.1.4 数据库设计的步骤	149
7.1.5 数据库应用系统生命周期	152
7.2 数据库系统的需求分析	153
7.2.1 需求分析的任务	153

7.2.2	需求分析的方法	153
7.2.3	数据字典	154
7.2.4	数据流图	155
7.3	概念结构设计	156
7.3.1	概念结构设计的特点	156
7.3.2	概念结构设计的方法与步骤	156
7.3.3	数据抽象与局部视图设计	158
7.3.4	全局 E-R 模型设计	160
7.4	数据库系统的逻辑设计	163
7.4.1	逻辑结构设计任务和步骤	164
7.4.2	E-R 图向关系模型的转换	164
7.4.3	数据模型的优化	165
7.5	数据库的物理设计	166
7.5.1	确定物理结构	166
7.5.2	确定数据库的储存结构	167
7.5.3	评价物理结构	167
7.6	数据库的实现	167
7.6.1	建立实际数据库结构	167
7.6.2	数据的载入和应用程序的调试	168
7.6.3	编制与调试应用程序	168
7.6.4	数据库的试运行	168
7.7	数据系统的运行和维护	169
	本章小结	170
	习题	170
第 8 章	数据库管理	173
8.1	数据库中事务的概念	173
8.2	数据库的恢复	174
8.2.1	事务的故障	174
8.2.2	数据库恢复的基本原理及实现技术	175
8.2.3	故障恢复策略	177
8.3	数据库的并发控制	178
8.3.1	并发控制概述	178
8.3.2	封锁及其解决问题的办法	179
8.4	数据库的完整性	183
8.4.1	数据库完整性约束条件的分类	183
8.4.2	数据库完整性控制	184

8.5 数据库的安全性	185
本章小结	187
习题	188
第 9 章 SQL Server 2005 数据库管理系统介绍	190
9.1 SQL Server 2005 概述	190
9.1.1 SQL Server 2005	190
9.1.2 SQL Server 2005 的版本	191
9.1.3 SQL Server 2005 的主要组件与实用程序	192
9.2 数据库的创建及其管理	198
9.2.1 系统数据库	198
9.2.2 使用 SQL Server Management Studio 创建数据库	199
9.2.3 使用 SQL Server Management Studio 查看数据库信息	201
9.2.4 使用 SQL Server Management Studio 修改数据库	203
9.2.5 使用 SQL Server Management Studio 删除数据库	206
9.3 数据表的创建及其管理	207
9.3.1 SQL Server 数据类型	207
9.3.2 使用 SQL Server Management Studio 创建数据表	208
9.3.3 使用 SQL Server Management Studio 查看数据表	210
9.3.4 使用 SQL Server Management Studio 修改数据表	211
9.3.5 使用 SQL Server Management Studio 删除数据表	214
9.4 数据的添加、修改、删除和查询	214
9.4.1 使用 SQL Server Management Studio 添加数据	214
9.4.2 使用 SQL Server Management Studio 修改数据	214
9.4.3 使用 SQL Server Management Studio 删除数据	215
9.4.4 使用 SQL Server Management Studio 查询数据	215
9.5 数据完整性	217
9.5.1 SQL Server 2005 完整性概述	217
9.5.2 主键约束	218
9.5.3 规则	219
9.5.4 默认	222
9.5.5 数据库关系图	225
9.6 视图	227
9.6.1 视图设计窗口	227
9.6.2 使用 SQL Server Management Studio 创建视图	228
9.6.3 使用 SQL Server Management Studio 查看及修改视图	230
9.6.4 使用 SQL Server Management Studio 删除视图	231

9.7 数据库备份与还原	232
9.7.1 使用命令语句备份和还原数据库	232
9.7.2 使用 SQL Server Management Studio 备份数据库	233
9.7.3 使用 SQL Server Management Studio 还原数据库	236
9.7.4 使用 SQL Server Management Studio 导入/导出数据表	238
本章小结	245
习题	245
参考文献	248

第 1 章 数据库概论



数据库技术是计算机科学中重要的也是发展最快的分支之一,是信息系统的核心和基础,它的出现极大地促进了计算机应用向各行各业的渗透。数据库的建设规模、数据库信息量的大小和使用频度已成为衡量一个国家信息化程度的重要标志。学好“数据库原理”这门课程是掌握数据库技术的必修课程。

数据库概论知识是了解和掌握数据库原理及其应用的基础,只有理解这些概论知识,才能开始进一步学习数据库原理的理论和实践知识。本章主要介绍数据库管理技术发展的三个过程,介绍数据库技术中数据、数据库、数据库管理系统和数据库系统的基本概念,同时阐明数据库系统的主要特点和数据库技术发展方向。

1.1 数据管理技术的发展阶段与数据库技术概念

数据处理实际应用中数据的日益改进引发了数据库技术不断发展,数据处理需要存储大量数据。在计算机中,存储数据的硬件是存储器。在 20 世纪 50 年代末,存储器容量只能存储 5~10 MB 的信息容量,到 60 年代末才达到近 100 MB 的容量,初步具有了存放大量数据信息的条件,这时候数据管理技术才真正开始发展,所以说数据管理技术的发展是与计算机中外部存储器的发展密切有关。从 70 年代开始,存储器的存储容量得到飞速发展,到现在为止,几百兆字节的硬盘随处都可以买到,数据库中存储数据的困难完全解决,数据管理新技术也不断出现。下面介绍数据管理技术的 3 个发展阶段和目前使用的数据库技术的基本概念。

1.1.1 数据管理技术的 3 个发展阶段

从 20 世纪 50 年代开始至 70 年代初,数据管理技术的发展经历了人工管理、文件管理和数据库管理 3 个阶段。

1. 人工管理阶段

在 20 世纪 50 年代之前,计算机的应用主要是在科学计算方面。在硬件设施方面,外存只有纸带、卡片、磁带,没有磁盘等直接存取设备;在软件方面,没有操作系统和管理数据的软件;数据处理方式是批处理。在这种情况下,数据管理的基本方法只能是采用人工管理。

人工管理数据具有以下几个特点:

(1) 数据不能长期保存在计算机内。在当时的科学计算中,计算所需的原始数据连同程序一起输入内存,运算后输出结果,随着计算任务的完成和程序作业退出计算机系统,数

据空间连同程序空间一起被释放。

(2) 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,因此程序间数据不能共享,有大量的冗余数据。

(3) 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,必须对应用程序做出相应的修改。当时只有程序(Program)的概念,没有文件(File)的概念。

在人工管理阶段,程序与数据之间的对应关系如图 1.1 所示。

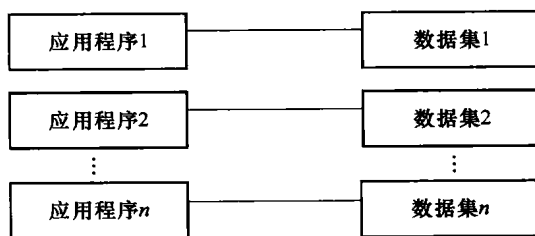


图 1.1 数据处理的人工管理阶段

2. 文件管理阶段

20 世纪 50 年代末至 60 年代末,当时计算机不仅用于数据计算,还用于信息管理,这时需要对数据进行增加、修改、删除等操作,还要求把数据保存下来,并根据需要进行维护。此时硬件水平已有较大的提高,出现了磁盘、磁鼓等存储设备;软件水平上已有成熟的高级语言,并开始使用操作系统,数据处理方式已应用联机实时处理和批处理方式。

以文件形式处理数据有如下特点:

(1) 数据可以以“文件”形式长期保存在外部存储器的磁盘上。由于计算机除了科学计算外,还要进行信息管理,因此可以对文件进行查询和各种更新操作。

(2) 数据的逻辑结构与物理结构有区别,但比较简单。与数据打交道,不需要知道数据的物理位置,只要知道文件名及其位置就可以。

(3) 文件组织已有多多样化。有索引文件、链接文件和直接存取文件,但这些文件之间相互独立,没有联系,获取它们之间的数据要依靠程序去解决。

(4) 数据不再属于某个特定的程序,可以重复使用,即数据面向应用。一般来说,设计文件时总是根据某一要求设计的,所以总是与程序密切相关,缺乏真正的独立性。

(5) 对数据的操作以记录为单位。在设计文件时,完全是为了存储数据和如何方便地让程序能获得数据,所以不会去记录数据的结构,所有数据的更新全靠编写程序来解决。

虽然文件中的数据没有记录结构,但文件系统管理阶段是数据管理技术发展阶段中的重要一步。在这一阶段中,为了能在文件中快速得到数据,程序中的结构算法得到了很大的发展,为数据管理技术的发展打下了坚实的基础。直到现在,许多高级语言中还在使用这种方法获取数据。

文件系统阶段中设计的文件相互独立,本质上存在许多缺陷,主要是数据冗余(Redundancy)大、各文件中数据的不一致性(Inconsistency)不可避免、数据之间联系弱(Poor data relationship)。

例如,某一单位要管理职工的档案、工资和医保,要建立3个文件:职工工资文件、职工档案文件和职工医保文件。在每一个文件中,必须要有职工号、职工姓名、性别、年龄、单位、级别等内容,否则很难正确管理。从现在的观点来看,这样做存在着明显的结构上的问题,形成重复输入,有增加大量内存、浪费资源、增加成本等缺点。

在20世纪60年代中期,在文件系统阶段的后期,数据管理规模一再扩大,数据量急剧增加,为了提高系统性能,人们开始对文件系统加以扩充,研制成倒排文件系统(Inverted File)。倒排文件是索引文件的推广,对每个字段都提供单独的索引。这些文件很适合于信息检索系统。缺点是要占用许多内存(而当时内存昂贵)。倒排文件系统的出现虽然是属于文件系统阶段,但它大大提高了对当时数据查询和数据处理的速度。

在文件系统阶段,程序与数据之间的对应关系如图1.2所示。

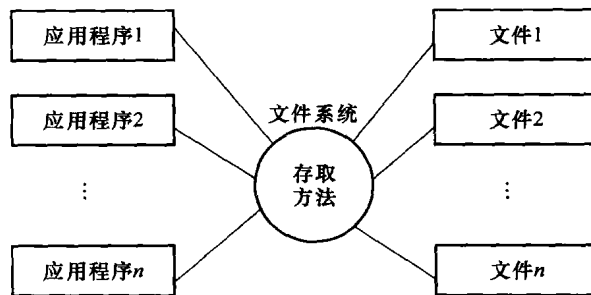


图 1.2 数据处理的文件系统阶段

3. 数据库管理阶段

20世纪60年代末以来,计算机用于数据处理的发展已成为大规模管理,数据量大大增加;硬件方面出现了大容量磁盘(能生产上百兆字节容量的硬盘),而价格却大大降低;软件价格越来越高,使得编写程序和维护软件的成本增加;在处理方式上,联机实时处理方式越来越多,分布式处理也已经开始实施。原先文件系统处理的方式已不能满足要求,为了解决数据共享的要求,能够用来统一管理数据的数据库管理系统随即产生。

在数据库管理系统阶段,程序与数据之间的对应关系如图1.3所示。

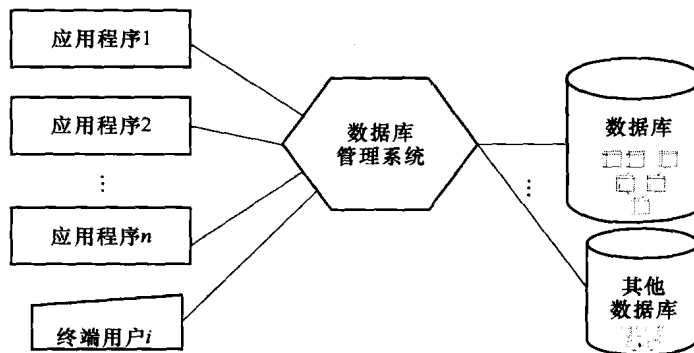


图 1.3 数据库管理系统对数据的处理方式

从图1.3可以看出,所有数据都放在数据库内,是公用的,也就是说共享的,所有应用程序都可以通过数据库管理系统来调用数据。

1.1.2 数据库系统阶段产生的 3 大新技术

1. 层次数据库的发明

由于利用计算机来进行数据计算技术的迅速发展,许多有实力的公司都愿意花钱和时间来发展这方面的技术。1968年,当时具有很强实力的美国 IBM 公司与当时一家制造火箭的公司为阿波罗登月火箭处理数据,研制出 IMS(Information Management System)信息管理系统,这是一个典型的层次数据库系统,当时命名为 IMS/1,安装在 IBM360/370 计算机上,1969年9月正式投入运行,取得了很大的成功,为当时数据库的发展做出了重要贡献。

这一系统在 1974 年推出 IMS/VS(Virtual System)版本,在 20 世纪 70 年代的商业和金融业得到广泛的应用。当时我国国家计委和一些银行都使用过这一系统。

2. 网状数据库的发明

1969 年美国数据系统语言协会的数据库研究小组(专门研究数据库语言)DBTG(Date Base Task Group)提出了网状数据库系统的报告,这份报告对数据库和数据库操作的环境建立了统一的规范,一般被称为 DBTG 系统(或 CODASYL 系统)。在这种系统下研制成功的 IDMS、IDS II、DMS100、TOTAL、IMAGE 等数据库管理系统,在 20 世纪 70 年代至 80 年代中期得到了广泛的应用,为数据库管理技术的发展做出了很大的贡献。

3. 关系数据库的发明

1970 年,IBM 公司的一位科学家 E. F. Codd 首先在美国计算机学会《通信》杂志上发表关系型数据库方面的论文,名称为“A Relation Model of Data for Large Shared Data Banks”。现代数据库的许多概念都是从此篇文章中的思想继承和发展来的,它奠定了关系数据库的理论基础,这一基础是建立在集合论和谓词演算的基础上。由于关系数据库模型简单,为任何一种关系数据库管理系统提供一种统一的结构,这种结构就是二维表,用户在使用时不必去考虑数据链接的方法和存放数据结构等复杂的问题。

由于关系数据库模型的建立是在集合论和谓词演算的基础上,所以形成的数据库语言是一种非过程性语言,所谓非过程性语言是指编写程序时,只要告诉做什么,不必告诉怎样做的编程语言,它比结构化过程语言(既要告诉做什么,还要告诉怎样做)方便很多。关系模型在开始使用时,因为当时计算机硬件条件差,运行速度慢,效率稍低,但到了 20 世纪 80 年代初,计算机硬件条件得到改善,很快关系模型产品占领市场,逐步取代层次模型和网状模型产品,目前使用最广泛的关系数据库产品有:Visual FoxPro、SQL Server、Sybase、Oracle、Informix、DB2 等。

1.1.3 数据库系统的主要特点

到现在为止管理数据都在使用数据库管理系统,淘汰了人工管理数据和文件管理数据,这是因为数据库管理系统有其独特的特点,这些特点主要表现在以下几个方面。

1. 数据结构化

数据库系统采用数据模型表示复杂的数据结构,这是数据库与文件系统的根本区别。在文件系统中,相互独立的文件记录内部是有结构的,最简单的形式是等长记录,这种结构是面向某一具体应用的,缺乏灵活性;而数据库系统实现了整体的结构化,数据不再面向某一应用,而是面向全组织,不仅数据具有结构,而且存取数据的方式非常灵活,可以存取数据

库中的某一个数据项、一组数据项、一个记录或一组记录,而在文件系统中,数据的最小存取单位是记录,粒度不能细到数据项。

2. 有较高的数据独立性

数据的独立性包括数据的物理独立性和逻辑独立性。前者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,当数据的物理存储改变了,应用程序不用改变;后者是指用户的应用程序与数据库的逻辑结构是相互独立的,数据的逻辑结构改变,在一定条件下用户程序也可以不变。这一点在后面数据库系统结构中会作比较详细的说明。

3. 数据库系统为用户提供了方便的用户接口

用户可以使用查询语言或终端命令操作数据库,也可以用程序方式(如使用各种高级语言 C++、C#、Visual Basic 等编写的程序调用)操作数据库,解决了数据共享的问题,而且降低了数据的冗余度并易于扩充。

4. 数据库系统提供 5 方面的数据控制功能

(1) 数据库的并发控制(Concurrency)。对多用户同时需要应用数据库的并发操作加以控制和协调,防止相互干扰而得到错误的结果。

(2) 数据库恢复(Recovery)。在某一特殊情况下,数据库被破坏或数据不可靠时,系统有能力将数据库从错误状态恢复到某一已知的正确状态。

(3) 数据库的完整性(Integrity)。在数据库运行过程中,系统能将数据控制在有效的范围内,或保证数据之间能满足一定的关系。

(4) 数据的安全性(Security)。保护数据,防止不合法用户的使用造成数据的泄密和破坏,使每个用户只能按指定方式操纵数据。

(5) 增加了系统的灵活性。对数据的操作可以是集合形式操作,也可以以记录为单位操作,还可以以单个字段或多个字段为单位操作。

数据处理技术发展数据库系统管理阶段是信息处理领域的一个重大变化,是从不同程序调用数据到可共享的数据库为中心的新阶段,使数据能集中管理和维护,增加了数据的利用率和相容性,数据使用的可靠性得到了极大的提高。

目前世界上已有百万计的数据库在运行,几乎深入到各个领域。我国在 20 世纪 90 年代开始,在各行各业装备了以数据库为基础的大型计算机系统,这些系统分布在工业、农业、科研、学校、商贸、金融等领域。例如在邮电、银行、电力、交通、气象、旅游、股票交易、期货交易、情报、公安、网络等行业的应用尤为突出,人们都能感受到,现在几乎在各行各业都普遍建立了以数据库为核心的信息系统。

1.1.4 数据库技术中的几个主要名称

数据、数据库、数据库管理系统和数据库系统是学习数据库技术一开始就该明确的 4 个基本概念。

1. 数据(Data)

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。

数据是数据库中存储的基本对象,是描述事物的符号记录,如数字、文字、图形、图像、声音等都是数据,按一定的方法转化为计算机能识别的二进制符号进行处理,处理后需要显示的数据,再按一定的方法恢复成人们能看懂或理解的数据。

如果要了解某一个人的姓名、年龄和性别的信息,可以用一些具体的数据值来表达,例如:周一鸣,36,男。

数据与信息概念是不完全相同的。我们可以这样解释:数据是信息的符号表示形式,或者说数据是信息的载体,信息则是数据的内涵,是对数据的语义解释。例如老王每月工资3 000元,数据“3 000”表示特定数目,说明是以“元”为单位的老王月工资信息。

2. 数据库(DataBase, DB)

数据库是存放数据的仓库。具体而言,数据库是长期储存在计算机内、有组织的、可共享的大量数据集合,意思是说数据库中的数据是按一定数据结构组织、描述和存储,具有尽可能小的冗余度和较高的数据独立性和易于扩展性,可供许多用户共用。所以说数据库具有集成性(把要使用的数据及其联系集中在一起,并按一定的结构形式进行存储)和共享性两个突出的特点。例如,能反映学生一般情况的数据库,可以把学生的所在学校、学号、姓名、年龄、性别、进校日期、所在系别、专业和班级的数据存放在一起。

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

数据库管理系统是数据库的核心组成部分,是对数据库中数据进行管理的大型系统软件,是位于用户与操作系统之间的一层数据管理软件,它为用户或应用程序提供了访问数据库的方法,包括数据库的建立、查询、更新及各种数据控制。具体来说包括以下4个主要功能:

(1) 数据定义。DBMS提供数据定义语言(Data Definition Language, 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。

(2) 数据操纵。DBMS提供数据操纵语言(Data Manipulation Language, 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。其中,国际标准数据库操作语言——SQL语言就是DML的一种(它同时还包含DDL、DCL语言功能)。

(3) 数据库的运行管理。这一功能是数据库管理系统的核心所在。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、有效地正常运行。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4个方面来实现对数据库的统一控制功能的(这一内容会在第7章中详细介绍)。

(4) 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、性能监视和分析功能等。这些功能均可以使用DBMS中的一些专用命令来解决。

4. 数据库系统(DataBase System, DBS)

数据库系统是实有组织地、动态地存储大量关联数据、方便多个用户访问的硬件、软件和数据资源组成的系统,即它是采用数据库技术的计算机系统。它主要由以下4部分组成:

(1) 数据库(DB)。

(2) 硬件。这一部分包括中央处理器、内存、外存、输入/输出设备等硬件设备,在DBS中特别要关注内存、外存、I/O存取速度、可支持终端数和性能稳定性等指标和考虑连网能力。此外要求系统有较高的通道能力,以提高数据的传输速度。

(3) 软件。这一部分包括DBMS、操作系统(OS)、各种开发数据库的高级语言和各种应用开发支撑软件程序。

(4) 数据库管理员(DataBase Administrator, DBA)。DBA 是控制数据整体结构的一组人员,负责 DBS 的正常运行,承担创建、监控和维护数据库结构的责任。DBA 要熟悉使用数据库单位全部数据的性质和用途、对所有用户的需求有充分的了解、对系统的性能非常熟悉,兼有系统分析员的知识。

数据库系统结构如图 1.4 所示。

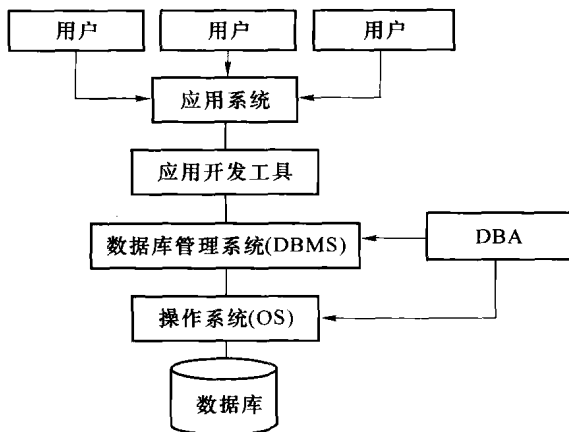


图 1.4 数据库系统的结构

1.2 数据库技术的新发展

虽然目前的数据库技术已经比较完美,但随着社会信息量持续增加,信息处理技术不断发展,数据库技术也在不断发展。每个学员都要及时跟踪和学习数据库技术的最新进展,以便能掌握它的最新技术为自己的工作服务。本节主要简单介绍数据库研究领域和数据库发展两方面的知识。

1.2.1 数据库技术研究的主要领域

数据库技术研究领域研究的问题十分广泛,也十分复杂,综合起来,可以归结为以下 3 个方面。

1. 数据库理论研究领域

数据库理论研究领域是在数据库研究领域中最难突破的领域,近 20 多年来,这一方面的研究从来没有中断过,这一领域中的研究主要集中在关系数据库的规范化理论、关系模型的关系运算、新型数据库模型建立与应用等方面。20 世纪 80 年代后,随着计算机其他领域技术的不断发展,与数据库技术相互渗透,在应用方面产生了许多新的理论研究方向,如数据库的逻辑演绎和知识推理、数据库中的知识发现、并行数据库产生与应用、分布式数据库的产生与应用、多媒体数据库系统、数据仓库系统、工程数据库系统、模糊数据库系统等。

2. 数据库系统软件的研制

数据库系统软件主要是指 DBMS 软件及其配套的检测和维护软件,还包含以 DBMS 为核心的一组相互联系的软件系统(如工具软件和中间件软件等),研制的最终目的是提高系