



普通高校“十二五”规划重点教材

数据库基础及应用

主编 樊重俊 刘 臣 杨坚争



立信会计出版社

LIXIN ACCOUNTING PUBLISHING HOUSE

普通高校“十二五”规划重点教材

数据库基础及应用

Shujuku Jichu Ji Yingyong

主编 樊重俊 刘 臣 杨坚争



立信会计出版社

LIXIN ACCOUNTING PUBLISHING HOUSE

图书在版编目(CIP)数据

数据库基础及应用 / 樊重俊, 刘臣, 杨坚争编著.

—上海: 立信会计出版社, 2015. 1

普通高校“十二五”规划重点教材

ISBN 978-7-5429-4378-1

I. ①数… II. ①樊…②刘…③杨… 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(2014)第 270568 号

责任编辑 黄成良

封面设计 周崇文

数据库基础及应用

出版发行 立信会计出版社

地 址 上海市中山西路 2230 号 邮政编码 200235

电 话 (021)64411389 传 真 (021)64411325

网 址 www.lixinaph.com 电子邮箱 lxaph@sh163.net

网上书店 www.shlx.net 电 话 (021)64411071

经 销 各地新华书店

印 刷 常熟市梅李印刷有限公司

开 本 787 毫米×1092 毫米 1/16

印 张 20 插 页 1

字 数 480 千字

版 次 2015 年 1 月第 1 版

印 次 2015 年 1 月第 1 次

书 号 ISBN 978-7-5429-4378-1/F

定 价 38.00 元

如有印订差错,请与本社联系调换

编委会名单

主 任	杨坚争	上海理工大学管理学院副院长,教授,博士,博导
副主任	窦瀚修	立信会计出版社社长
	龚炳铮	华北计算机研究所教授级高级工程师
编 委	时启亮	上海理工大学管理学院教授
	蔡建平	上海理工大学外语学院副院长,教授
	万以娴	上海权亚智博律师事务所高级合伙人,博士
	劳帼龄	上海财经大学信息管理与工程学院副教授,博士
	张宝明	上海理工大学管理学院副教授
	魏 忠	上海海事大学经济管理学院副教授

前 言

经过近 50 年的发展,数据库系统和数据库技术已经成为计算机信息系统的重要基础和核心技术,数据库技术、计算机操作系统和互联网通信均已成为 IT 基础设施的重要组成部分,几乎每个行业都涉及数据库系统管理、存储和处理数据的应用与实践,比如各大银行、证券公司、通信公司、行政管理等等。数据库课程已经成为国内各大高校的核心课程之一,不仅是计算机学院,也是管理学院部分专业(信息管理与信息系统、金融学、国际贸易等)的核心课程,更是许多其他专业的选修课程。作为高等院校的学生,具备基础的数据库系统知识,掌握一定程度的数据库操纵技术,对现在的学习和将来的工作,都是必不可少的。

SQL Server 作为一种大型数据库,是当前应用最广泛的数据库系统之一。本书以 SQL Server 为背景,将实例贯穿其中,由浅入深地介绍数据库课程,学习理解数据库系统基础知识以及使用数据库技术必须掌握的常用方法,本着以应用为核心,以基础与操作为支撑,注重理论与实际相结合,力图让读者在有限的时间之内,对该课程的主要知识和技术有一个比较清晰完整的理解和把握,能进行数据库操作,而且能从事数据库领域的实际工作。

本书将基础知识与实验环节的讲解融为一体,除了详尽的讲解之外,每章还包括了循序渐进的实验方法说明和精心安排的习题。全书共有 11 章,包括以下几部分内容:

第 1 章主要介绍数据库的基本知识,包含数据库概念与发展、数据模型、数据库系统结构和语言等内容。第 2 章主要介绍关系数据库的基础知识,包含关系、关系模式、关系数据模型和关系代数等内容。第 3 章主要介绍关系数据库的标准语言 SQL,包含 SQL 数据定义、数据操纵和数据查询等内容。第 4 章主要介绍数据库的管理与保护,包含数据库安全性、数据库完整性、事务、并发控制、备份与恢复等内容。第 5 章主要介绍关系数据理论,包含函数依赖、公理系统、规范化等内容。第 6 章主要介绍数据库设计,包含数据库设计概念、设计步骤与实施、运行与维护等内容。第 7 章主要介绍数据库编程操作数据库,包含编程基础、游标、存储过程、触发器和嵌入式 SQL 编程等内容。第 8 章和第 9 章主要介绍数据库技术发展趋势和研究领域,包含分布式数据库、数据仓库与商务智能等内容。第 10 章主要介绍 SQL Server 2008 管理系统,包含 SQL Server 数据库的发展、数据库系统简介以及 SQL Server 2008 安装过程、相关组件、管理工具的功能与使用等内容。第 11 章主要介绍大数据的发展、非关系型数据库的应用实例以及 Hadoop 内容简要。

全书由樊重俊、刘臣、杨坚争统稿。刘臣、张青磊撰写第 1 章;申瑞娜、孟栋、郭强撰写第 2 章;袁光辉、杨云鹏、马淑娇撰写第 3 章;何蒙蒙、林叮云、樊重俊撰写第 4 章;杨云鹏、刘臣撰写第 5 章;申瑞娜、秦江涛、马淑娇撰写第 6 章;何蒙蒙、刘臣、张青磊撰写第 7 章;杨云鹏、申瑞娜、樊重俊撰写第 8 章;袁光辉、张惠珍撰写第 9 章;杨云鹏、霍良安撰写第 10 章;刘臣、

霍良安、杨云鹏撰写第 11 章。最后,由樊重俊、刘臣、马淑娇、杨云鹏对全书校对。本书作为高等院校数据库课程的教材以 32~48(包括上机时间)学时为宜。本书配有习题参考答案和教学用 PPT,任课教师可发 E-mail 至 Chenggen765@163.com 联系索取。

数据库技术仍在不断发展变化,本书编写过程中,力求跟踪数据库学科最新的技术水平和发展方向,引入新的技术和方法。但由于篇幅、时间等限制,书中疏漏之处在所难免,殷切希望同行专家和读者批评指正。

本书是上海理工大学核心课程(数据库原理)建设的一个组成部分,本书获得了沪江基金研究基地专项“电子商务智库”(D14008)和立信会计出版社的大力支持,在此表示诚挚的谢意。

编者

2014 年 11 月

目 录

第 1 章 绪论	1
1.1 概述	1
1.2 数据描述与数据模型	10
1.3 概念模型	13
1.4 逻辑模型	17
1.5 数据库系统结构	23
1.6 数据库系统语言类型	25
1.7 小结	27
习题	27
第 2 章 关系数据模型	29
2.1 关系数据库基本概念	29
2.2 关系操作	34
2.3 关系的完整性	35
2.4 关系代数	37
2.5 小结	43
习题	43
第 3 章 结构化查询语言(SQL)	47
3.1 SQL 概述	47
3.2 数据的定义	51
3.3 数据的操纵	61
3.4 数据的查询	65
3.5 视图和索引	85
3.6 小结	90
习题	90
第 4 章 数据库管理与保护	92
4.1 数据库的安全性	92
4.2 数据库的完整性	98
4.3 数据库的并发控制	103
4.4 数据库的备份与恢复	112
4.5 小结	114

习题	114
第 5 章 关系数据库的规范化	116
5.1 规范化的重要性	116
5.2 数据依赖	118
5.3 数据库的规范化	125
5.4 更高级别范式	130
5.5 小结	135
习题	136
第 6 章 数据库设计	138
6.1 数据库设计的步骤	138
6.2 数据库需求分析	141
6.3 数据库结构设计	145
6.4 数据库实施	158
6.5 数据库运行与维护	160
6.6 小结	161
习题	161
第 7 章 数据库编程	162
7.1 数据库编程概述	162
7.2 T-SQL 编程基础	163
7.3 游标	172
7.4 存储过程	177
7.5 触发器	183
7.6 嵌入式 SQL 编程	189
7.7 小结	194
习题	195
第 8 章 分布式数据库系统概论	196
8.1 分布式数据库系统的发展	196
8.2 分布式数据库系统的概述	198
8.3 分布式数据库系统的结构	203
8.4 分布式数据库管理系统	209
8.5 分布式数据库系统的主要技术及应用	212
8.6 小结	214
习题	214

第 9 章 数据仓库与商务智能	215
9.1 商务智能的需求	215
9.2 商务智能简介	216
9.3 商务智能应用及模型	221
9.4 数据仓库	224
9.5 数据仓库案例	229
9.6 小结	234
习题	234
第 10 章 SQL Server 2008 概述	235
10.1 SQL Server 数据库的发展历程	235
10.2 SQL Server 2008 的特性简介	238
10.3 SQL Server 2008 的安装过程	243
10.4 SQL Server 2008 的管理工具	264
10.5 小结	273
习题	274
第 11 章 大数据和非关系型数据库	275
11.1 大数据的产生、发展与挑战	275
11.2 非关系型数据库(NoSQL)	280
11.3 常见的非关系数据库实例	290
11.4 Hadoop 简介	302
11.5 小结	309
习题	309
参考文献	310

第 1 章

绪 论

随着信息技术的不断发展,信息管理水平的不断提高,对信息资源的掌握就成为各个行业部门取得竞争优势的重要战略财富和资源。建立一个高效的信息系统也就成为其生存和发展的重要基础条件。而数据库技术是信息系统的核心技术之一,是一种计算机辅助管理数据的方法。它研究如何组织和存储数据,即数据库的结构、存储、设计、管理以及应用的基本理论和实现方法;并利用这些理论来实现对数据库中的数据进行处理、分析和理解的技术。因此,数据库技术是研究、管理和应用数据库的一门软件科学。

数据库技术产生于 20 世纪 60 年代末 70 年代初,其主要目的是有效地管理和存取大量的数据资源。近年来,数据库技术和计算机网络技术的发展相互渗透,相互促进,明显加快了经济信息化和社会信息化进程,广泛地应用在事务处理、情报检索、人工智能、专家系统、计算机辅助设计等领域。所以,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。

本章主要介绍了数据库的基本概念、数据库技术的发展、数据模型及数据库新技术等内容。通过本章的学习,可以充分了解数据库的相关概念,了解数据库技术的发展历程及数据库技术在现实中的作用,以便为以后章节的学习打下坚实的基础。

1.1 概 述

在了解数据库模型、结构、语言之前,我们首先了解一下数据库常用的一些术语和基本概念,然后对数据库技术的发展和应用领域进行简单的介绍。

1.1.1 基本概念

1. 数据(Data)

数据用于描述现实世界中各种具体事物或抽象概念,可存储具有明确意义的信息符号,包括数字、文字、图形、音频、视频等。数据处理是指对各种形式的数据进行收集、存储、加工和传播的一系列活动的总和。在日常生活中当提到数据时,人们的第一反应是数字。例如,在“张三 25 岁”这个描述中,人们一般只把“25”当做数据,而“张三”、“岁”都不当作数据。实际上,这个整体或其每一个组成部分都是数据,因为它们也是在表达和传递具体的信息,都是存储在计算机中的符号。

所以在现代计算机系统中,凡是描述事物或概念的符号记录,能为计算机所接受和处理的各种字符、数字、图形、图像、音频及视频等,都可称为数据。数据可以分为数值型数据和

非数值型数据。数值型数据是可以用于数学计算的数字,如一个企业的员工编号、人数、年龄及工资等;非数值型数据范围非常广泛,如文字、图形、音频、视频等都是非数值型数据。数值型数据和非数值型数据都可以被转化成计算机能够处理的形式并存储在数据库中。

数据有“型”和“值”之分。数据的型是指数据的结构类型或表现形式,是数据的内部结构和对外联系的方式;数据的值是指数据的具体取值。数据的表现形式不能完全表达其内容,还需要在不同的背景下进行语义的解释。例如,25 是一个数据,它可以是一个人的年龄,也可以是一个班级中学生的人数,还可以是某件物品的价格。数据的语义是指对数据含义的解释和说明,同一个数据在不同的背景下可以有不同的语义。因此,数据与语义是不可分割的。

在计算机中,为了存储和处理现实世界中的事物,需要根据应用场景抽取出能够清楚表达其本质的一些特征组成一条记录来描述。例如,日常生活中人们一般用自然语言来描述某企业的一位员工的信息:张小青,女,1986 年 9 月出生,复旦大学,本科,2013 年 9 月入职。在计算机中通常这样记录这名员工的信息:

(张小青,女,198609,复旦大学,本科,201309)

员工的姓名、性别、出生年月、毕业院校、学历、入职时间等特征被抽取出来,形成一条记录。这样,孤立的、没有意义的数据就被组织成具有特定语义的形式。数据的语义就隐含在这样的结构之中。同时,记录中的每个数据还受到数据类型和取值范围的约束,有些只能填充数字型数据,有些更是约束了其范围,这就是数据的值域,保证了数据的有效性。例如,姓名只能用字符型数据填充,性别只能取值为汉字字符“男”或“女”。记录是计算机中最为常见的表示和存储数据的方法。

2. 信息(Information)

信息通常定义为经过加工处理后,有用的,对人的行为产生影响的数据,是反映客观现实世界的知识。同一个信息可以用不同的数据形式来表示。例如,当我们在表达一棵树时,我们可以用汉字“树”表示,也可以用图画来表示;对于同一条新闻,我们可以通过报纸、电台、电视、网络等多种方式获得。所以,信息也可以这样来定义:指通过各种方式传播的可以被感知关于某事物的消息、情报和知识,信息需要依赖于数字、文字、图像、声频、视频等不同的形式或符号作为载体。

信息依赖于数据但与数据不同,它是一种抽象的事物,是数据经过数据处理过程转变而成的。数据处理过程包括对数据的收集、存储、加工、分类、检索、传播等一系列操作。通过数据处理过程,再经过人的分析、归纳、演绎、推理等,形成有价值、有意义的信息。

总之,数据和信息是相似但有所区别的两个不同的概念。数据是信息的载体,信息是数据的内涵。数据只是一种符号,本身是没有意义的,而信息是有意义的知识。但没有数据,信息也将成为无源之水、无本之木而不复存在。因此,数据、信息,以及计算机技术之间的关系可用图 1-1 表示。

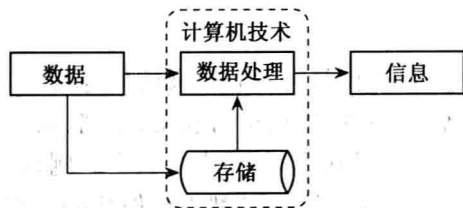


图 1-1 信息 = 数据 + 数据处理

3. 数据库(Data Base, DB)

数据库是长期储存在计算机内的、有组织的、可共享的大量数据的集合。数据库具有如

下的特点:

- (1) 数据库中的数据按一定的数据模型组织、描述和存储;
- (2) 具有较小的冗余度、较高的独立性和易扩张性;
- (3) 数据库可被各种用户共享。

总的来说,数据库是具有逻辑关系和确定意义的数据集合。数据库是针对明确的应用目标而设计、建立和加载的;每个数据库都具有一组用户,并为这些用户的应用需求服务。一个数据库反映了客观事物的某些方面,而且需要与客观事物的状态始终保持一致。

4. 数据库管理系统(Database Management System, DBMS)

数据库管理系统是对数据库进行管理的系统软件,它的职能是有效地组织和存储数据,接受和完成用户提出的各种数据访问请求。我们经常用的数据库管理系统有 SQL Server、Oracle、DB2、MySQL 等。

(1) 数据库管理系统的功能

数据库管理系统是介于用户与操作系统之间的一套用于对数据库进行管理的系统软件,主要的目标就是使数据成为方便各种用户使用的资源,并提高数据的安全性、完整性和可用性。所以 DBMS 一般都具有如下功能:

① 数据定义。数据库管理系统提供数据定义语言(Data Definition Language,简称 DDL)。利用 DDL 可以方便地对数据库中存储的数据进行定义。例如,对数据库、表、字段、视图和索引进行定义、创建和修改等。

② 数据操纵。为了对数据库中的数据进行查询、插入、修改和删除等操作,数据库管理系统还需要提供相应的工具,即数据操纵语言(Data Manipulation Language,简称 DML)。

③ 数据控制。数据控制语言(Data Control Language,简称 DCL)用于实现数据库运行控制的功能,包括并发控制(即处理多个用户同时使用某些数据时可能产生的问题)、安全性检查、完整性约束条件的检查和执行,以及数据库的内部维护(如索引的自动维护)等。

④ 数据组织、存储和管理。DBMS 要分类、组织、存储和管理各种数据,包括数据字典、用户数据、数据的存取路径等,此外还需要确定这些数据以何种文件结构和存取方式组织在存储器上,如何实现数据之间的联系等,从而提高存取效率。

(2) DBMS 的组成及工作流程

DBMS 由数据和元组、存储管理器、查询处理器、事务管理器、模式更新、查询和更新等部分组成,如图 1-2 所示。

① 数据和元组。数据是数据库管理系统的管理对象;元组是描述数据的数据,是有关数据结构的信息。例如,在关系数据库管理系统中,数据是由用户添加到基本表中的数据,元组是

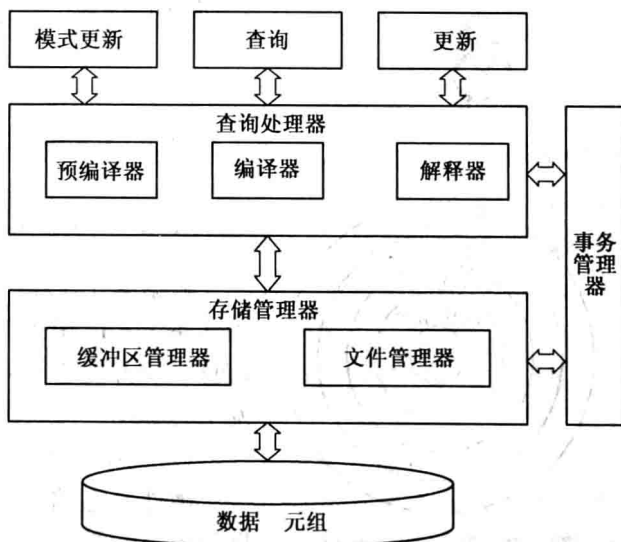


图 1-2 DBMS 的主要组成

描述有基本表名、列名、数据类型等数据库对象的数据。

② 存储管理器。存储管理器根据获得的请求,从数据存储器中获得信息或修改数据存储器中的信息。存储管理器由缓冲区管理器和文件管理器组成。

③ 查询处理器。查询处理器负责处理数据查询、数据更新和模式更新等工作。查询处理器包括:预编译器、编译器、解释器。

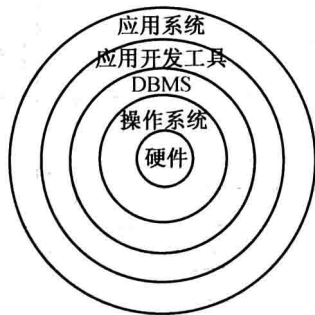
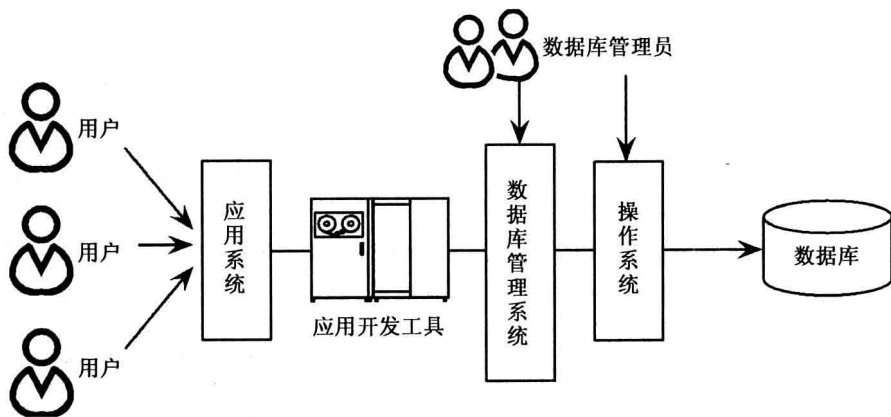
④ 事务管理器。负责系统的完整性工作,必须确保同时运行的查询语句不相互影响。

综上所述,数据库管理系统接受应用程序的数据请求和处理请求,然后将用户的数据请求转换成机器代码实现对数据库的操作,并接受对数据库操作从而得到查询结果,同时对查询结果进行处理,最后将结果返回给用户。

5. 数据库系统 (Data Base System, 简称 DBS)

数据库系统是指在计算机系统中引入数据库后的系统。它可以实现有组织地、动态地存储大量数据,并提供数据处理和信息资源共享等功能。一般由数据库、数据库管理系统、其开发工具、应用系统、人员(包括数据库管理员、软件开发员及用户)构成。数据库的建立、使用和维护等工作仅仅依靠数据库管理系统还远远不够,最终还需要专业人员来管理和操作,称为数据库管理员(Data Base Administrator, 简称 DBA)。一般情况下,常常把数据库系统简称为数据库。

数据库系统的构成如图 1-3 所示。



1.1.2 数据库在计算机系统中的地位

数据库在计算机系统中的地位如图 1-4 所示,它是介于操作系统和应用系统之间的软件层,其中应用系统由专门的信息系统开发工具开发,其数据存取与管理功能也由应用开发工具实现。

(1) 硬件平台

硬件是存储数据库和运行 DBMS 的物质基础。由于数据量大且频繁进行存取等访问操作,因此需要较高的硬件配置:

① 要有足够大的内存,用于存放操作系统、DBMS 的核心模块、数据缓冲区、应用程序和数据库表等;

② 要有足够大的外存,用于存放数据库和数据备份(如磁盘或磁盘阵列数据库、光盘、磁带用作数据备份)。

③ 要有较高的通信能力,保证较高数据传送率。

(2) 软件

软件是指数据库系统中被计算机使用的程序的集合。需要以下几类软件来实现数据库系统的全部功能:

① DBMS 是数据库系统的核心,用于数据库的建立、使用和维护。前面曾经提到常见的 DBMS 的软件有 SQL Server、Oracle、ERP、DB2、MySQL 等,每种 DBMS 软件都有着自己的特色与优势,用户需要根据自己的需求进行选择;

② 支持 DBMS 运行的操作系统,各种 DBMS 软件几乎都提供了对主流操作系统的支持,如 Linux、Unix、Windows 等;DBMS 运行在操作系统之上,接受操作系统的控制和调度;

③ 具有与数据库接口的高级语言及开发工具,常见的程序设计语言如 C、C++、Java、VB、C#、Python 等都对各种 DBMS 提供支持,比较流行的数据库系统开发工具及平台有 JSP、ASP、PHP 等;

④ 特定应用环境开发的数据库应用系统。终端用户借助应用程序通过 DBMS 访问数据库。

(3) 人员

开发、管理和使用数据库系统的人员主要包括:系统分析员和数据库设计人员、应用程序开发人员、数据库管理员和最终用户。

① 系统分析员负责应用系统的需求分析和规范说明,与用户及数据库管理员协商以确定系统的硬软件配置,并参与数据库系统的概要设计;

② 数据库设计人员负责数据库中数据的确定,数据库中各模式的设计。数据库设计人员必须参加用户需求调查和系统分析,然后进行数据库设计。在很多情况下数据库设计人员由数据库管理员担任;

③ 应用程序开发人员负责设计和编写应用系统的程序模块,并进行调试和安装;

④ 数据库管理员负责全面管理和控制数据库系统,其主要职责包括:决定数据库中的信息内容和结构;决定数据库的存储结构和存取策略;定义数据的安全性要求和完整性约束条件;监控数据库的使用和运行(例如,周期性转储数据库、数据文件、日志文件、系统故障恢复、介质故障恢复、监视审计文件);数据库的改进和重组,性能监控和调优,定期对数据库进行重组织,以提高系统的性能;需求增加和改变时,数据库则需要重构。

⑤ 最终用户通过应用系统的用户接口使用数据库,浏览器是最常使用的用户接口。

1.1.3 数据管理技术的产生与发展

数据库管理技术伴随着计算机的产生而产生,与计算机的发展密不可分。在计算机诞生之前,由各种人类活动所产生的数据很简单而且数量有限,因此对数据管理没有太高的要求。随着计算机的发展,需要存储并处理的数据量大增,结构也越来越复杂,

于是数据管理技术应运而生。数据管理技术是指对数据进行分类、组织、编码、存储、检索和维护的计算机应用技术。

随着计算机硬件、软件的发展,在应用需求的推动下数据管理技术也不断发展。根据数据管理技术的时间、主要应用领域、依赖的计算机技术,以及应用的特征和数据特征,可以粗略地将其划分为人工管理、文件系统、数据库系统三个阶段。每阶段都是对前一个阶段数据管理的完善,数据管理技术向数据存储冗余小、数据独立性不断增强、数据操作和转换更加方便和简单的方向发展(如表 1-1 所示)。

表 1-1 数据管理技术的发展

	人工管理阶段	文件系统阶段	数据库系统阶段
时 间	20 世纪 40~50 年代	20 世纪 50~60 年代	20 世纪 60 年代~现在
应用领域	科学计算	科学计算、数据处理	大规模数据管理
计算机技术	没有操作系统和文件管理技术,数据存储纸带、卡片或磁带之中	出现了批处理操作系统和文件系统,数据存储磁盘、磁鼓之中	利用 DBMS 管理大规模数据集,大容量、分布式存储技术
应用特征	数据与应用程序不能分离,数据不能共享,冗余量极大	数据以文件形式存储,可以由多个程序应用,但独立性很差	数据独立于应用程序,共享程序很高,冗余度很低
数据特征	数据无结构,不能存储在计算机中,程序运行结束即退出	数据结构化程度很低,以文件形式存储	数据完全结构化,用数据模型来描述、组织和存储

1. 人工管理阶段(20 世纪 50 年代中期以前)

人工管理阶段是计算机数据管理的初级阶段,当时计算机主要用于科学计算。在计算机出现之前,人们利用纸张记录数据,利用计算工具(算盘、计算尺等)进行数据计算。早期的计算机也是用于数值计算,数据存储纸带、卡片、磁带等之上。那个时期还没有出现操作系统,数据的组织和管理完全依靠手工完成,因此称作“人工管理阶段”。计算机仅由极少数专业人员编制针对性极强的程序来使用,应用程序和数据集之间是一对一的关系(如图 1-5 所示)。

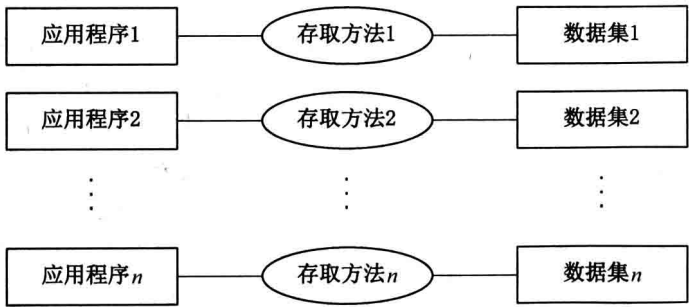


图 1-5 人工管理阶段应数据使用方式

人工管理阶段,数据管理技术有如下特点:

① 数据不能在计算机中保存。当时计算机主要用于科学计算,一般不需要长期保存数据,利用卡片、纸带,将数据输入,运算后输出结果,完成数据处理过程。

② 数据需要由应用程序自己管理,没有相应的软件系统负责数据的管理工作。应用程序中不仅要规定数据的逻辑结构,而且要设计物理结构,包括存储结构、存取方法、输入方式等。

③ 数据不能在应用程序之间共享。数据是面向应用的,当多个应用程序涉及某些相同的数据时,必须各自定义,且无法互相利用、互相参照,因此程序与程序之间有大量的冗余数据。

④ 数据不具有独立性。数据的逻辑结构或物理结构发生变化后,必须对应用程序做相应的修改。

2. 文件系统阶段(20世纪50年代末至60年代中期)

20世纪50年代后期到60年代中期,随着计算机技术的发展,在软件方面出现了批处理操作系统及早期的文件系统,硬件方面出现了磁鼓、磁盘等数据存储设备。数据以相互独立的文件形式存储,不同的应用程序可以利用文件系统以统一的方式存取数据(如图1-6所示)。尽管在文件内部可以实现一定程度的结构化,但该工作由应用程序自己完成,因此数据在不同的应用程序之间共享还很困难。

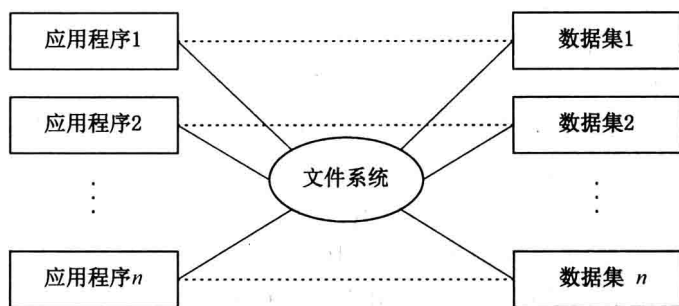


图 1-6 文件系统阶段的数据使用方式

文件系统管理阶段的数据管理技术有如下特点:

① 数据在计算机存储设备中保存。磁鼓、磁带等存储设备能够提供大容量、快速的数据存取,而且能够长期保存。

② 由文件系统提供统一的数据存取方法。文件系统把数据组织成相互独立的数据文件,使用“按文件名访问,按记录进行存取”的管理技术,可以对文件进行创建、修改和删除等操作。

③ 数据共享程度低。尽管文件系统能够提供统一的文件管理方法,但是应用程序还需要自己根据需要定义数据结构,因此,从数据管理的角度来看仅具有内部结构而整体可以说是无结构的。数据在应用程序之间的共享难度很大,数据冗余度依旧较高,安全性、完整性得不到可靠的保证。

④ 数据独立性差。虽然应用程序与数据之间具有独立性,但应用程序仍然依赖于文件的逻辑结构,应用程序的语言环境的变化也将引起文件数据结构的改变。文件系统不容易扩充;数据与程序之间仍缺乏独立性;是一个无弹性的无结构的数据集合。

文件系统阶段是数据管理技术发展中的一个里程碑式的阶段。但是,随着人们对数据需求的增加以及计算机技术的不断发展,文件系统仍然不能满足人们对于数据库管理技术的要求。

3. 数据库系统阶段(20 世纪 60 年代末以来)

20 世纪 60 年代末以来,人们逐渐认识到计算机在信息存储与管理上的巨大潜力,计算机被应用到各种领域。这直接导致数据量急剧增加,数据类型及结构越来越复杂,迫切要求能够实现高共享、低冗余,高效快捷的数据管理技术。计算机软硬件的发展,存储技术的提高为数据库系统的出现提供了可能性。数据库理论也不断发展,层次模型、网状模型、关系模型(见本章 1.2 节)等为商业数据库奠定了坚实的理论基础。1968 年 IBM 推出了世界上第一个数据库管理系统 IMS(Information Management System),它是基于层次模型的系统。在 1969 年美国数据系统语言协会(CODASYL)的数据库任务组(DBTG)发表了网状数据模型的 DBTG 报告。1970 年美国 IBM 公司的高级研究员 E. F. Codd 连续发表论文,提出了关系数据模型及相关概念,奠定了关系数据库的理论基础。于是,数据管理技术全面进入数据库系统阶段。数据库系统的出现标志着数据管理技术的一个质的飞跃,相对于文件系统有着明显的区别和巨大的优势。

(1) 数据高度结构化

数据库系统实现整体数据的结构化,是数据库的主要特征之一,也是数据库与文件系统的根本区别。在文件系统中,相互独立的文件可以由应用程序实现内部的结构化,但不同文件之间是没有联系的。在数据库中,全部数据集由 DBMS 统一进行管理,实现了不同层次的、整体的结构化。例如,在关系数据库中记录与记录之间、表与表之间、记录与表之间都存在着联系,满足各种完整性约束。

数据的高度结构化使得数据存取的灵活性也大大提高。在文件系统中,数据访问的最小粒度是行(对应着记录)。而在数据库系统中,数据可以以各种粒度存取,最小的粒度是数据项,甚至可以方便地存取数据库中互不相邻的数据项或记录。

(2) 数据高度共享,冗余度低,易扩充

在采用人工管理或文件系统管理时期时,由于数据被重复存储,当不同的应用使用和修改不同的拷贝时就很容易造成数据的不一致。数据库系统不再面向某个应用,而是面向整个系统,从整体角度看待和描述数据,在降低冗余的同时也保证了数据的一致性。因此,数据可以被多个用户、多个应用共享使用。

由于数据面向整个系统,是有结构的数据,不仅可以被多个应用共享使用,而且容易增加新的应用。这就使得数据库系统弹性大,易于扩充,可以适应各种用户的要求。可以取整体数据的各种子集用于不同的应用系统,当应用需求改变或增加时,只要重新选取不同的子集或加上一部分数据便可以满足新的需求。

(3) 数据独立性高

数据库系统使得应用程序从数据存取和管理的繁重工作中解脱出来,可以专注于具体的业务流程。通俗地说,数据独立性是指应用程序和数据结构之间的依赖程度,高度的数据独立性使得当应用程序修改时数据尽可能不发生变动,反之亦然。在数据库系统中,数据独立性包括数据的物理独立性和数据的逻辑独立性。

物理独立性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。数据