



SHUJUKU JISHU
SHIYONG
JIAOCHENG

数据库技术 实用教程

徐洁磐 周嵩 / 编著



东南大学出版社
SOUTHEAST UNIVERSITY PRESS

数据库技术实用教程

徐洁磐 周 嵩 主 编

东南大学出版社
· 南京 ·

内 容 简 介

本书全面介绍了计算机软件应用、软件设计和数据库的主要基础知识和内容。介绍了计算机软件的历史与现状、常用数据结构及其算法实现、计算机操作系统、计算机软件工程、面向对象方法学、计算机的数据库系统。

本书编写力求反映应用型本科的要求和理工类专业的教学特点,内容力求由浅入深、循序渐进、通俗易懂,基本概念和基本知识准确清晰,计算机软件的说明简明扼要,尽量避免繁琐的数学推导,重点放在计算机软件应用和软件设计,注重将计算机软件应用、软件设计和数据库知识有机地结合起来,并且特别注意以形象直观的形式来配合文字表述,重点突出,以帮助读者掌握计算机软件应用和软件开发的主要内容。

本书适于不同层次的读者选用,既可用于高等学校理工类本科教材,也适用于各类工程技术人员参考、阅读。

图书在版编目(CIP)数据

计算机软件基础/鲍玉军等编. —南京:东南大学出版社, 2008. 9

(新世纪计算机课程系列精品教材)

ISBN 978-7-5641-1223-3

I. 计… II. 鲍… III. 软件—高等学校—教材
IV. TP31

中国版本图书馆 CIP 数据核字(2008)第 103389 号

计算机软件基础

出版发行 东南大学出版社
出版人 江 汉
社 址 南京市四牌楼 2 号
邮 编 210096
经 销 全国各地新华书店
印 刷 盐城印刷总厂有限责任公司
开 本 787mm×1092mm 1/16
印 张 16.75
字 数 415 千
版 次 2008 年 9 月第 1 版
印 次 2008 年 9 月第 1 次印刷
印 数 1—4000 册
书 号 ISBN 978-7-5641-1223-3/TP·199
定 价 30.00 元

(凡因印装质量问题,请与我社读者服务部联系。电话:025-83792328)

前 言

数据库技术与数据库应用是目前信息技术发展的重要内容,它的发展也带动了数据库的教学与培训。目前数据库的教学与培训除正规计算机专业开设“数据库”课程外,还涉及到与计算机相关的大量专业,如电子商务、企业管理、信息管理等专业本科正规“数据库”课程以及相关的培训类“数据库”课程,而这些数据库课程大都需要的是以实用性为主的内容。

根据这种分析,本书以计算机相关专业本科“数据库”课程教材为目标,以实用性内容为主。

本书实用性体现在如下几个方面:

1. 介绍数据库技术的基本概念、原理与内容。
2. 介绍数据库技术的基本操作——SQL,并以 SQL Server 2000 为蓝本作介绍。
3. 介绍数据库技术的开发技术,包括数据库设计、开发平台及数据库编程等内容。
4. 介绍数据库技术的应用,重点介绍数据库在电子商务、客户关系管理、企业资源规划以及数据挖掘中的应用。

本书是一本特色显著的教材,主要表现在如下几个方面:

1. 实用性 本书突出实用性,四个篇章涵盖了数据库实用的所有方面,学完本书后学生应能掌握数据库基本技术并能操作、使用与开发数据库应用系统。

2. 先进性 本书既有传统的数据库内容,更主要的是大量吸收了现代的数据库技术与新方法,同时也摒弃了过时落后与不实用的内容,并按新的思想重新构造新的结构体系。使得本书具有内容新颖、结构合理的特色。

3. 理论与实际相结合 本书既包括实际应用内容也包括一定的基本原理与理论,而两者是不可缺一的。要解决此问题关键是需两者有机结合,而目前教材中普遍存在理论与实际脱节的倾向。在本书中我们突出“以理论引导实际”同时又“以实际支撑理论”,两者相辅相成、无缝结合,构成了一个完整的体系。

4. 适合教学 本书内容丰富、结构合理、书写通俗,非常适合于教学。本书每章配有大量应用性习题,并附有复习提纲,为方便学生实验在本书中还附有 10 个实验内容。

本书配有电子教案与课件供教师免费使用。本书还将配有相关辅导教材,不久即将出版。

值本书付梓之际,作者特别感谢山东大学董继润教授、南京大学费翔林教授为本书所付出的艰辛劳动及提出的许多宝贵意见。同时也感谢南京大学计算机软件新技术国家重点实验室及南京大学计算机科学与技术系的多位老师的支持与帮助。

作者水平有限,书中错误、缺点在所难免,恳切希望读者批评指正。

编著者

南京大学计算机软件新技术国家重点实验室

南京大学计算机科学与技术系

2008. 3 于南京

目 录

第一篇 基础篇

1 基本概念	(2)
1.1 数据库技术概述	(2)
1.2 有关数据的基本概念介绍	(5)
1.2.1 数据的基本性质	(5)
1.2.2 数据在计算机中的位置与作用	(6)
1.3 有关数据管理的内容与概念介绍	(7)
1.3.1 数据管理的变迁	(7)
1.3.2 数据管理中的几个基本概念	(8)
1.3.3 数据管理中数据库内部结构体系	(10)
1.3.4 数据管理特点	(11)
1.3.5 数据管理工具	(12)
1.4 有关数据处理的内容与概念介绍	(13)
1.4.1 数据库应用系统	(13)
1.4.2 数据处理环境	(14)
1.4.3 数据处理应用开发	(15)
1.4.4 数据处理应用领域	(15)
2 数据模型	(18)
2.1 数据模型的基本概念	(18)
2.2 数据模型的四个世界	(19)
2.3 概念世界与概念模型	(19)
2.4 信息世界与逻辑模型	(24)
2.4.1 概述	(24)
2.4.2 关系模型	(24)
2.5 计算机世界与物理模型	(28)
2.5.1 数据库的物理存储介质	(29)
2.5.2 磁盘存储器及其操作	(29)
2.5.3 文件系统	(30)
2.5.4 数据库物理结构	(30)

3 关系模型的数学理论——关系代数	(34)
3.1 关系、代数与关系代数	(34)
3.1.1 关系	(34)
3.1.2 代数	(35)
3.1.3 关系代数	(35)
3.2 关系模型中的关系代数.....	(36)
3.2.1 关系与二维表	(36)
3.2.2 关系运算与表的操作	(36)
3.3 关系代数在关系模型中的应用.....	(40)
4 关系模型数据库管理系统	(43)
4.1 关系数据库管理系统概述.....	(43)
4.2 关系数据库管理系统基本内容组成.....	(43)
4.2.1 数据定义功能	(44)
4.2.2 数据操纵功能	(45)
4.2.3 数据控制功能	(46)
4.2.4 数据交换功能	(53)
4.2.5 数据服务及数据字典	(59)
4.2.6 关系数据库管理系统的扩展功能	(59)
4.3 关系数据库管理系统标准语言 SQL	(62)
4.3.1 SQL 概貌	(62)
4.3.2 SQL 的功能	(63)
5. 关系模型数据库管理系统 SQL Server 2000 介绍	(67)
5.1 SQL Server 2000 系统概貌	(67)
5.1.1 SQL Server 2000 的应用环境与结构模式	(67)
5.1.2 SQL Server 2000 的版本	(67)
5.1.3 SQL Server 2000 的运行平台	(67)
5.1.4 一个典型的系统架构	(68)
5.1.5 SQL Server 2000 组成	(69)
5.1.6 SQL 语言	(72)
5.2 SQL Server 2000 的特色	(72)
5.3 SQL Server 2000 的基本功能	(73)

第二篇 操作篇

6 SQL 数据定义语句	(79)
6.1 数据定义功能介绍.....	(79)

6.2	SQL 的数据定义语句	(79)
6.2.1	SQL 基本数据类型	(80)
6.2.2	SQL 的模式定义语句	(80)
6.2.3	SQL 的表定义语句	(81)
6.2.4	SQL 的索引定义语句	(82)
6.3	SQL Server 2000 中的数据定义语句	(82)
6.3.1	创建与删除数据模式	(82)
6.3.2	表	(85)
7	SQL 数据操作语句	(87)
7.1	SQL 的查询语句	(87)
7.2	SQL 的更新语句	(93)
7.3	SQL 的统计、计算及分类	(94)
7.4	SQL 中的视图语句	(95)
7.5	SQL Server 2000 中数据操纵及视图功能	(96)
8	SQL 数据控制语句	(101)
8.1	SQL 的数据控制功能	(101)
8.2	SQL 的安全性控制语句	(101)
8.2.1	SQL 对数据库安全的支持	(101)
8.2.2	SQL Server 2000 中的数据安全性操作	(103)
8.3	SQL 的完整性控制语句	(107)
8.3.1	SQL 完整性控制语句	(107)
8.3.2	SQL Server 2000 中的数据完整性语句	(109)
8.4	SQL 的事务语句	(111)
8.4.1	有关事务的 SQL 语句	(111)
8.4.2	SQL Server 2000 中的事务语句	(111)
8.5	SQL 的故障恢复操作	(112)
8.5.1	故障恢复三大功能	(112)
8.5.2	SQL Server 2000 中的故障恢复操作	(112)
9	SQL 数据交换及扩展语句	(115)
9.1	数据交换中的 SQL 语句	(115)
9.1.1	会话管理语句	(115)
9.1.2	连接管理语句	(115)
9.1.3	游标管理语句	(116)
9.1.4	诊断管理语句	(117)
9.1.5	动态 SQL 管理	(117)
9.2	SQL 扩展操作	(118)
9.2.1	人机交互方式	(118)

9.2.2	自含式方式及 T-SQL	(130)
9.2.3	调用层接口方式及 ODBC 接口	(136)
9.2.4	Web 方式	(143)

第三篇 开发篇

10	数据库设计	(150)
10.1	数据库设计概述	(150)
10.2	数据库设计的需求分析	(151)
10.2.1	需求调查	(151)
10.2.2	需求分析	(151)
10.2.3	需求分析说明书	(152)
10.3	数据库的概念设计	(152)
10.3.1	数据库概念设计概述	(152)
10.3.2	数据库概念设计的过程	(153)
10.3.3	概念设计说明书	(155)
10.4	数据库的逻辑设计	(156)
10.4.1	数据库逻辑设计基本方法	(156)
10.4.2	关系视图设计	(158)
10.4.3	逻辑设计说明书	(158)
10.5	数据库的物理设计	(158)
10.5.1	存取方法设计	(159)
10.5.2	存贮结构设计	(159)
10.5.3	物理设计说明书	(160)
10.6	一个数据库设计实例	(160)
10.6.1	需求分析	(160)
10.6.2	概念设计	(162)
10.6.3	逻辑设计	(163)
10.6.4	物理设计	(164)
11	数据库应用系统开发	(167)
11.1	数据库应用系统概述	(167)
11.1.1	数据库应用系统组成	(167)
11.1.2	数据库应用系统的开发	(167)
11.2	数据库应用系统的平台介绍	(168)
11.2.1	网络结构平台	(168)
11.2.2	中间件	(169)
11.3	数据库应用系统中的数据层	(172)
11.3.1	数据库应用系统中的数据层的组成	(172)

11.3.2	数据库应用系统中数据层的开发	(173)
11.4	数据库应用系统中的应用层	(173)
11.4.1	数据库应用系统中的应用层组成	(173)
11.4.2	数据库应用系统中应用层的开发	(174)
11.5	数据库应用系统中的界面层	(174)
11.6	典型的数据库应用系统组成介绍	(174)
12	数据库应用编程	(178)
12.1	自含式方式中的编程	(178)
12.2	调用层接口的编程	(183)
12.2.1	ODBC 编程步骤	(184)
12.2.2	连接阶段	(188)
12.2.3	执行 SQL 语句	(192)
12.2.4	结束阶段	(199)
12.2.5	ODBC 编程示例	(200)

第四篇 应用篇

13	数据库在电子商务领域中的应用	(208)
13.1	电子商务简介	(208)
13.2	电子商务发展历程	(209)
13.3	电子商务的特点与优势	(209)
13.4	电子商务应用系统的构成	(210)
13.5	电子商务应用系统的开发	(211)
14	数据库在客户关系管理中的应用	(214)
14.1	客户关系管理介绍	(214)
14.2	CRM 内容的确定	(215)
14.3	CRM 应用系统的构成	(216)
14.4	CRM 应用系统的开发	(218)
15	数据库在企业资源规划中的应用	(220)
15.1	企业资源规划介绍	(220)
15.2	企业资源规划发展史	(220)
15.3	企业资源规划的基本内容	(221)
15.4	ERP 应用系统构成	(222)
15.5	ERP 应用系统的开发	(224)
16	数据库在分析领域中的应用	(226)
16.1	决策支持系统	(226)

16.2	数据仓库的基本原理	(227)
16.2.1	概论	(227)
16.2.2	数据仓库特点	(228)
16.3	数据仓库的基本结构	(230)
16.3.1	数据源	(230)
16.3.2	数据抽取	(230)
16.3.3	数据仓库管理层	(231)
16.3.4	数据集市层	(231)
16.4	联机分析处理 OLAP	(232)
16.4.1	OLTP 与 OLAP	(232)
16.4.2	OLAP 的基本概念	(232)
16.4.3	OLAP 的基本数据模式——星形与雪花模式	(233)
16.4.4	OLAP 的多维数据结构——数据立方体及超立方体	(234)
16.4.5	OLAP 多维结构的物理存储	(237)
16.4.6	OLAP 的分析操作	(237)
16.5	数据挖掘	(238)
16.5.1	数据挖掘的方法	(238)
16.5.2	数据挖掘的步骤	(239)
16.6	DSS 中的建模与展示	(240)
16.7	DSS 整体结构	(240)
附录 “数据库课程”实验指导		(242)
一	实验计划与要求	(242)
二	实验 1 实验准备	(242)
三	实验 2 数据模式定义	(243)
四	实验 3 数据查询	(246)
五	实验 4 数据更新及视图	(247)
六	实验 5 数据库安全保护与备份、恢复	(248)
七	实验 6 数据库设计	(249)
八	实验 7 C/S 结构方式的数据库应用系统开发	(250)
九	实验 8 B/S 结构方式的数据库应用系统开发	(250)
十	实验总结	(251)
参考文献		(252)

第一篇 基础篇

数据库技术是计算机学科中的一门独立技术,其主要内容包括基础知识、基本操作、系统开发及应用等多个部分。

数据库技术的基础知识部分是此门学科的基石,它给出了该学科的抽象、全局的内容并对整门学科起指导性作用。

在本篇中我们给出数据库技术的一般性基础知识以及关系数据库管理系统中的基础知识,其具体内容如下:

1. 数据库技术中的一般性基础知识。

在本篇中数据库技术的一般性基础知识由两章组成,它们是第1章与第2章。其中第1章介绍数据库技术中数据、数据管理及数据处理等三个基本部分的概念、内容及相互间关系。第2章介绍数据管理中的核心内容——数据模型。

2. 关系数据库管理系统的基础知识。

数据库管理系统中目前最流行的是关系数据库管理系统,因此本篇的第二部分介绍关系数据库管理系统的基础知识,它由第3、第4及第5章共三章组成,其中第3章介绍关系数据库模型基本理论——关系代数,第4章介绍关系数据库管理系统基本内容组成及其标准语言,而第5章则是第4章的一个具体体现,即用第4章原理对 SQL Server 2000 作分析与介绍。

上述内容可用下面的图1表示之。

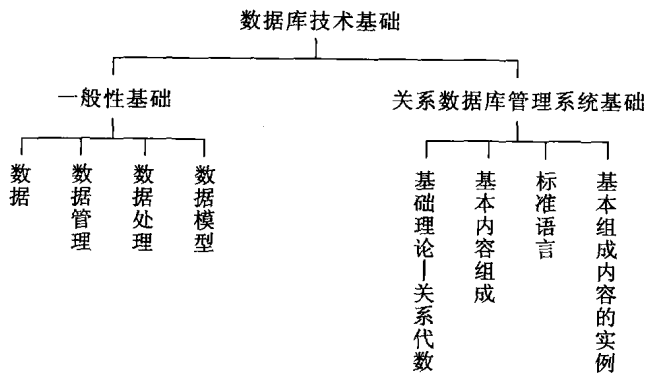


图1 数据技术基础内容分类图

本章介绍数据库技术与探讨的对象、内容与应用以及相关的基本概念。本章是全书的总纲,学习后使读者对数据库技术有一个全面、完整的认识。

1.1 数据库技术概述

本书是介绍数据库技术原理与应用的一本基础性教材,因此首先须对数据库技术有一个明确的了解与认识。

关于数据库技术可用下面的三句话概括之。

- 数据库技术的研究对象是数据;
- 数据库技术的研究内容是数据管理;
- 数据库技术的应用领域是数据处理。

下面我们将以这三个内容为主轴对数据库技术作介绍。

1) 数据

数据(data)是客观世界(包括自然界与人类社会)事物和现象的抽象表示,在本书中我们讨论的是计算机中的数据。它在计算机中表示为二进符号串,它是计算机处理与加工的对象。下面所列的形式都是数据。

- 数字:如某企业采购钢材 40 吨,计金额人民币 206 万元中数字“40”、“206”均为数据。
- 文字:如李白的诗句:“床前明月光,疑是地上霜,抬头望明月,低头思故乡。”是数据。
- 数学公式:如公式“ $y=ax+b$ ”为数据。
- 图象:如中央电视台天气预报中的“云图”为图象数据。
- 图形:如“动漫、电脑游戏”中的图形为数据。
- 音频:如电脑中所播放的贝多芬的“英雄交响曲”、柴可夫斯基的“天鹅湖”均是数据。
- 视频:如电视台播放的数字电视等均为数据。

目前,充塞于社会上的诸如:“数码时代”中的“数码”、“信息化社会”中的“信息”以及“数字化地球”、“数字化城市”、“数字化工厂”中的“数字”等都是数据的不同表现形式。因此,数字表示最具广泛性,使用与传播最具普遍性。

此外数据还具有很多实用的特性。

(1) 数据是一种重要的信息资源。

当今社会人们拥有巨大的物质资源与财富,此外,人们还可拥有大量的数据,它们构成了巨大的信息资源与财富,它们可以为人类社会与国民经济建设服务。

(2) 数据可以创造财富、创造文明。

利用物质资源可以创造财富、创造文明。同样,利用信息资源也可以创造财富、创造文明。大量的数据可以通过归纳、整理与分析而获得创造性的规则从而可以为人类服务,创造财富与

创造文明。

2) 数据管理

人类社会有着巨大的数据资源,为使它们有效的使用,必须对它们进行集中、统一的管理,这种管理称为数据管理(data management)。目前,数据管理方式有三种,它们是:

- 文件管理:是一种较为原始、初级的管理,它一般属操作系统范畴。
- 数据库管理:是一种严格、规范的管理,是一门独立的技术。
- Web 数据管理:是在 Web 环境下的一种数据管理方式,是一种相对开放、自由的管理方式。

在本书中我们主要讨论目前最为常用的数据库管理方式。在下面讨论中,如不特别指明数据管理即意指数据库管理。

数据管理是数据库技术研究的核心,它的内容有如下几个方面:

(1) 数据组织

为便于数据管理必须对数据进行有序与有机的组织,使其能存储在一个统一的组织结构下,这是数据管理的首要工作。

(2) 数据定位与查找

在浩如烟海的数据中如何查找到所需的数据是数据管理的重要任务。这种查找的难度可形容为“大海捞针”,而查找的关键是数据的定位,亦即是找到数据的位置。只有定位后数据查找才成为可能。因此数据定位与查找是数据管理的一项艰巨任务。此外,它还包括对数据的修改、删除与增添等工作。

(3) 数据的保护

数据是一种资源,其中大量的是不可再生资源,因此须对它作保护以防止丢失与破坏。数据保护一般包括以下几个部分:

① 数据语法与语义正确性保护:数据是受一定语法、语义约束的,如职工年龄一般在 18~60 岁之间,职工工资一般在 1 000~8 000 元之间等。又如职工的工资与其工龄、职务均有一定语义关联。任何违反约束的数据必为不正确数据。因此,必须保护其语法、语义的正确性。

② 数据访问正确性保护:数据是共享的,而共享是受限的。过分的共享会产生安全上的弊病,如职工工资,职工自身只有读权限而无写与改权限。因此数据访问权限是受限的,而正确访问权限是受到保护的。

③ 数据动态正确性保护:在多个数据访问并发执行时相互间会产生干扰从而造成数据的不正确,因此要防止此种现象产生,这称为数据动态正确性保护。

④ 最后,是数据动态正确性保护的另一种现象,即在执行数据操作时受外界破坏而产生故障的防止。

(4) 数据交换

为方便使用数据,必须为不同应用环境的用户提供不同使用数据的方式,它称为数据交换。

(5) 数据服务与元数据

为方便使用数据,在数据管理中还提供大量的服务功能称数据服务(data service)。数据服务是目前计算机领域中的一大潮流,其目的是为用户提供更多的方便与个性化需求。数据服务一般包括两种,它们分别是操作服务与信息服务,其中操作服务主要为用户提供多种操作上的方便,而信息服务则为用户使用数据库提供信息,特别是数据结构信息、数据控制信息。这种信息是有关数据的数据,因此又称元数据(metadata)。元数据是一种特殊的数据服务,由

于它的重要性在本书中将单独命名与单独介绍。

(6) 数据扩展

在不同数据环境中不同数据交换方式,为此需建立不同数据接口,这些接口的建立扩展了数据管理功能。这种功能可以方便数据处理使用数据,同时也为数据处理开拓应用提供支撑。

上述六种管理功能前五种为基本的管理功能而后一种则是扩展的管理功能。而在基本管理功能中前三个则是数据管理的核心功能,它们构成如图 1.1 所示的结构图。

这六种数据管理功能可分两个层次进行管理,其中一个层次是低层次管理,它负责数据管理中简单、常规的管理,它由系统软件——数据库管理系统进行管理;另一个层次是高层次管理,它由人——数据库管理员进行管理,它负责数据管理中复杂、智能性管理。

数据管理是数据库技术的主要研究内容。

3) 数据处理

将客观世界中的现象与事物抽象成计算机中的数据后,我们对客观世界现象与事物的研究即可转化为对计算机中数据的处理,它可称为“数据处理”(data process)。

数据处理是一种计算机的应用,它以批量数据多种方式处理为其特色,主要从事数据的加工、转换、分类、统计、计算、存取、传递、采集、发布等工作。

数据处理是一种新的处理问题的方法,它可以将对客观世界的研究借助于计算机中的数据处理而实现。正因为如此,目前世界上多门学科和多种应用均可以以数据处理为其基本方法与工具,其处理流程如图 1.2 所示。

在图中数据处理流程分成为四个阶段,它们分别是:

(1) 第一阶段——数据形成阶段:

此阶段的工作是将客观世界的现象与事物抽象成计算机中的数据。这是数据处理的首要关键的一步,一般而言,客观世界中某领域的现象与事物若能抽象成数据,此时对该领域的研究即可转化成为计算机中的数据处理,并正式进入数据处理流程。

(2) 第二阶段——数据管理阶段:

此阶段的工作是将所形成的数据按一定结构形式组织,对其作保护,并提供访问该数据的多种方式与接口,它为数据处理使用数据提供便利。

(3) 第三阶段——数据处理阶段:

此阶段的工作是数据处理,对数据作加工、转换等处理工作,在处理结束后以数据形式给出处理结果。因此我们说,数据处理是由数据到数据的一种转换工作。

数据处理一般由程序完成。

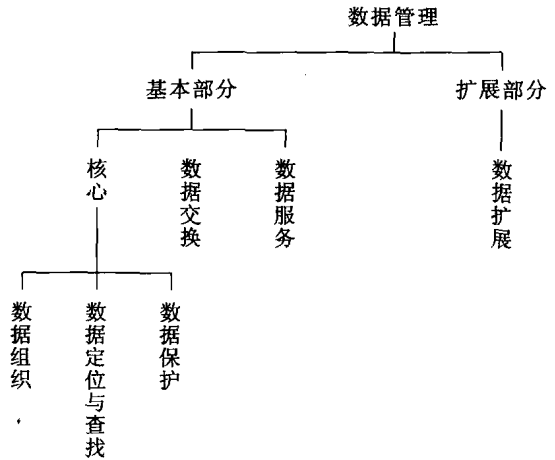


图 1.1 数据管理组织结构图

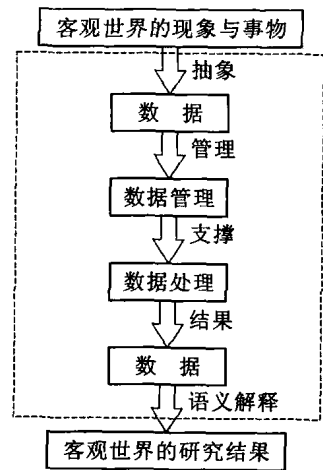


图 1.2 数据处理流程示意图

(4) 第四阶段——语义解释阶段:

此阶段的工作是将数据处理结果所形成的数据还原成客观世界中的现象与事物,从而结束整个处理流程。

整个流程看来,从客观世界中某领域的现象与事物出发,经计算机中的数据处理最终获得客观世界中的一些新的结果(包括新的现象与事物)。这就是数据处理作为工具参与解决客观世界各领域、学科的研究过程中的作用与方法。

1.2 有关数据的基本概念介绍

1.2.1 数据的基本性质

前面已经讲到,数据是计算机中的基本概念,它是计算机中处理、加工的对象与目标,其具体表示为存储于计算机内的二进符号串。数据一般有下面几个性质:

- 数据的抽象性

从形式看,数据是一种抽象的符号串,一般来讲它不含有语义,如数据:“15”可以是“15岁”、“15元钱”,也可以是“15公斤”、“15本”等等。因此在必要时须对它作出语义解释,此时它才有真实的意义。

- 数据的可构造性

从结构看,数据分为结构化数据(structured data)、半结构化数据(semi-structured data)与非结构化数据(non-structured data)。所谓非结构化数据即表示符号串是不规则结构形式;所谓半结构化数据即表示符号串呈半规则结构形式,如文件中的流式文件,如互联网中的 Web 结构等均属非结构化及半结构化形式。而在软件中的数据大多是有结构的,它们称结构化数据。首先,结构化数据有型(type)与值(value)之分,数据的型给出了数据表示的类型如整型、实型、字符型等,而数据的值给出了符合给定型的数值。随着应用需求的扩大,数据的型有了进一步的扩大,它包括了将多种相关数据以一定结构方式组合构成特定的数据框架,称为数据结构(data structure),具有统一结构形式和特性的具体描述可称为数据模式(data schema)。

- 数据的持久性

从存储时间看,数据一般分为两部分,其中一部分与程序仅有短时间的交互关系,随着程序的结束而消亡,它们称为临时性数据或挥发性(transient)数据。这类数据一般存放于计算机内存中;而另一部分数据则对系统起着持久的作用,它们称为持久性(persistent)数据,这类数据一般存放于计算机中的次级存储器内(如磁盘)。

- 数据的共享性

从其使用对象看,数据可分为私有性与共享性两种。为特定应用(程序)服务的数据称私有(private)数据,而为多个应用(程序)服务的数据则称为共享(share)数据。

- 数据的海量性

从其存储数量看,数据可分为小量、大量及海量三种。数据的量是衡量与区别数据的重要标志,这主要是由于数据“量”的变化可能会引起数据“质”的变化。数据量由小变大后,数据就需要进行管理,需要保护与控制。目前数据以海量数据为多见,因此一般数据均需管理、保护与控制。

随着技术的进步与应用的扩大,数据的特性都在发生变化,这些变化主要表现为:

- 数据的量由小量到大量进而到海量;
- 数据的组织由非结构化到结构化;
- 数据的服务范围由私有到共享;
- 数据的存储周期由挥发到持久。

数据的这些变化使得现代数据具有海量的、结构化的、持久的和共享的特点,本书如不作特别说明,所提数据即具此四种特性。

1.2.2 数据在计算机中的位置与作用

我们知道,计算机是由硬件与软件两部分组成,其中软件是运行实体而硬件则是运行平台。软件好像是一台运作的机器,将原料送入机器进行加工后变成成品。软件由程序与数据两部分组成,其中程序给出了运行的加工过程而数据则给出了运行的原料与成品。它们的关系可见图 1.3,而数据在计算机中的位置,则可见图 1.4。

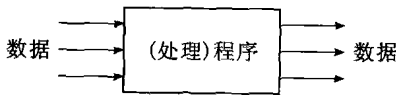


图 1.3 数据处理结构原理图

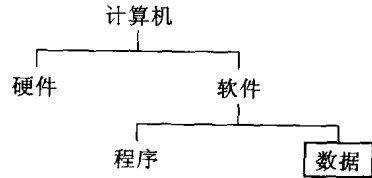


图 1.4 数据在计算机中位置示意图

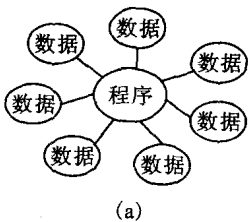
在软件中数据(主要指其结构)是其最稳定部分,而程序则是可变部分,因此数据称为软件中的不动点(fixed point),它在软件中起着基础性的作用。

软件发展至今,程序与数据间不同关系形成了目前流行的两种结构方式:

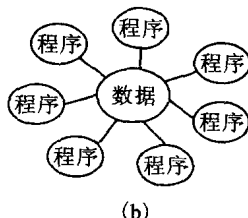
(1) 以程序为中心的结构:在此种软件结构中以程序为中心以数据为辅助,即每个程序有若干个数据为其支撑,它们构成了如图 1.5(a)所示结构。

(2) 以数据为中心的结构:在此种软件结构中以数据为中心以程序为辅助,即以一个数据集合为中心,围绕它有若干个程序对数据作处理,它们构成了如图 1.5(b)所示结构。

在目前,大多数软件结构采用以数据为中心的结构。



(a)



(b)

图 1.5 软件的两种结构图

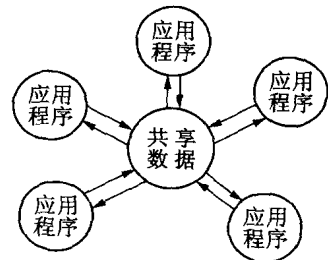


图 1.6 以数据为主体的软件系统

在过去,软件是以程序为中心,而数据则以私有形式从属于程序。在此种系统中,数据是分散、凌乱的,它造成了数据管理的混乱,如数据冗余大、一致性差、结构复杂等多种弊病,但经过若干年的发展,数据在软件中的地位和作用发生了本质的变化。在软件中它已占主体地位,而程序则已退居附属地位,它们构成了以数据为中心的结构。在此种结构中,需要对数据作集

中、统一的管理,并使其为多个应用程序共享,它们构成了如图 1.6 所示的结构图,这种结构方式为数据库系统的产生与发展奠定了基础。

1.3 有关数据管理的内容与概念介绍

数据管理是数据库技术探讨与研究的主要内容。在本节中主要介绍数据管理中的基本内容与基本概念。

1.3.1 数据管理的变迁

数据管理是数据库技术的核心,在数据库发展的 50 余年历史中,它经历了多个不同时段与阶段,它们是:

1) 人工管理阶段

20 世纪 40 年代自计算机出现至 50 年代这段时间中,由于当时计算机结构简单,应用面狭窄且存储单元少,对计算机内的数据管理非常简单,它们由应用程序编制人员各自管理自身的数据,此阶段称人工管理阶段。

2) 文件管理阶段

文件系统是数据库管理系统发展的初级阶段,它出现于 20 世纪 50 年代,此时计算机中已出现有磁鼓、磁盘等大规模存储设备,计算机应用也逐步拓宽,此时计算机内的数据已开始有专门的软件管理,这就是文件系统(file system)。

文件系统能对数据进行初步的组织,并能对数据作简单查找及更新操作,但是文件对数据的保护能力差,同时由于当时应用环境简单因此接口能力差。由于文件系统的数据库管理能力简单,因此它只能附属于操作系统而不能成为独立部分。目前一般将其看成是数据库系统的雏形,而不是真正的数据库管理系统。

3) 数据库管理阶段

自 20 世纪 60 年代起,数据管理进入了数据库管理系统阶段。由于计算机规模日渐庞大,应用日趋广泛,计算机存储设备已出现有大容量磁盘与磁盘组,且数据量已跃至海量,文件系统已无法满足新的数据管理要求,因此数据管理职能由附属于操作系统的文件系统而脱离成独立的数据管理机构,它即是数据库管理系统。

在数据库管理系统阶段,因不同的数据结构组织而分成为三个时代,它们是:

(1) 第一代数据库管理——层次与网状数据库管理时代

20 世纪 60 年代以后所出现的数据库管理系统是层次数据库与网状数据库,它们具有真正的数据库管理系统特色,但是它们脱胎于文件系统,受文件的物理影响大,因此给数据库使用带来诸多不便。

(2) 第二代数据库管理——关系数据库管理时代

关系数据库管理系统出现于 20 世纪 70 年代,在 20 世纪 80 年代得到了蓬勃的发展并逐步取代前两种系统。关系数据库管理系统结构简单、使用方便、逻辑性强、物理性少,因此 20 世纪 80 年代以后一直占据数据库领域的主导地位。关系数据库管理系统起源于商业应用,它适合于事务处理领域并在该领域内发挥主要作用。