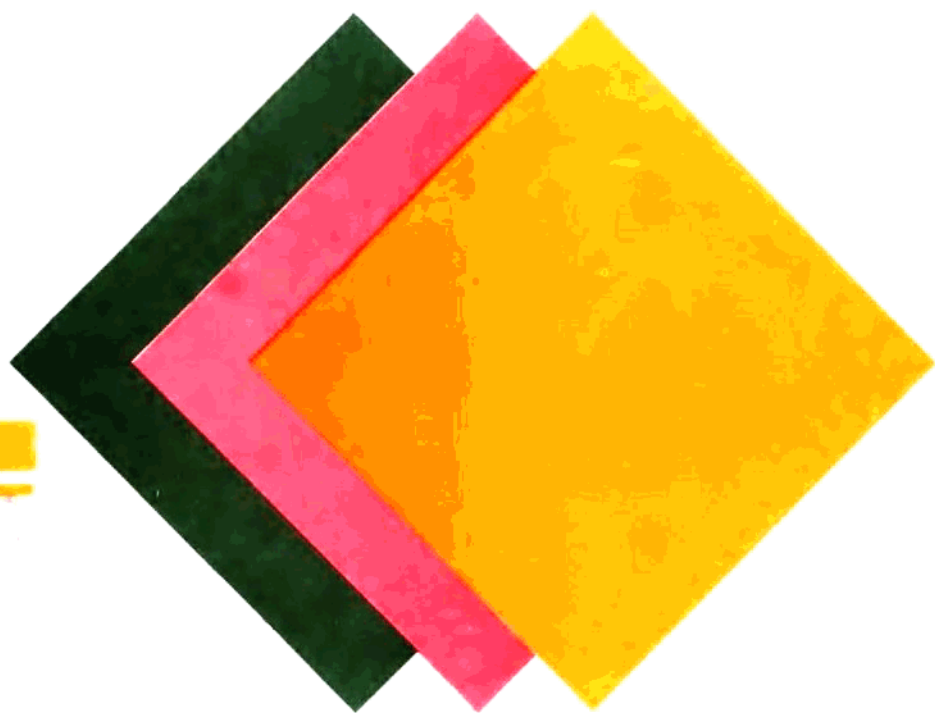


计算机系列教材

数据库系统概论

(第二版)

刘云生 卢正鼎 卢炎生



华中理工大学出版社

内 容 简 介

本书在1992年第一版的基础之上进行了很大的调整、修改与充实。全书分为三篇。第一篇系统、完整地叙述了数据库系统的基本概念、基本原理、基本方法与技术。主要内容有:数据库的发展、系统组成与结构(第一章),数据库的物理组织方法(第二章),主要的数据库模型(第三章),以DBMS为核心的数据库管理软件(第四章),三种主要的数据库技术与方法——层次、网状、关系系统(第五~七章),数据库的完整性、一致性、安全性与可靠性(第八章),数据库的设计方法与技术(第九章),数据库的运行维护与管理(第十章)。第二篇从实例与实验环境角度,系统地介绍了流行的FoxPro数据库系统及其使用(第十一~十五章)。第三篇给出了教学与学习的指导和实验。

本书可作为高等学校计算机各专业或非计算机专业数据库课程的教材,能适合多种层次(本科或专科)、多种培养目标的人员(如数据库的一般用户或程序设计员、数据库与信息系统的开发人员、数据库管理员等)的学习需要,亦可供计算机应用研究工作者、工程技术人员和其他有关人员参考。

《计算机系列教材》再版说明

我社出版的《计算机系列教材》已有 12 种相继问世。这 12 种教材分别是：《逻辑设计》、《PASCAL 语言程序设计》、《IBM-PC 宏汇编语言程序设计》、《计算机组成原理》、《数据结构》、《计算机操作系统》、《微型计算机及其应用》、《数据库系统概论》、《计算机及其使用基础》、《计算机的操作·应用·维护》、《计算方法》、《FOXBASE 程序设计教程》等。这批教材自 1990 年 12 月先后出版以来，多数已重印了七八次，深受广大师生、计算机方面自学成才者的欢迎，并获有关专家的好评。其中，《IBM-PC 宏汇编语言程序设计》荣获全国电子专业类优秀教材一等奖。

湖北省高等教育自学考试委员会已指定这批教材为湖北省高等教育自学考试计算机及应用专业的专业课教材，其中，《计算机及其使用基础》被指定为现行开考各专业的公共基础课教材。

这批教材都是约请具有丰富教学经验的各课程骨干教师编写的。每种教材都是按照计算机专业该课题教学大纲进行编撰，在内容的剪裁、文字的表述、例题和习题的选择上都力求遵循理论联系实际的原则，面向应用，重视实践，便于自学。每种教材都有“学习指导与实验”部分，其内容包括该课程大纲、学习辅导、习题解答和实验，目的是为了更方便读者自学、复习和加强实践环节的指导。

为了适应计算机科学技术飞速发展的新形势，为了满足计算机专业教学的需要和培养新型计算机专业人才的需要，必须更新教材的内容，把已经陈旧、过时的内容删去，增添新的知识。为此，我社约请每种教材的作者们对原教材进行精心修订。这次修订，重点是内容增删，同时根据初版以来作者的教学实践和有关专家提出的改进意见，作出修改。

湖北省高等教育自学考试委员会办公室、华中理工大学成人教育学院有关领导对这批教材的初版和再版给予了大力支持，华中理工大学计算机科学工程系和计算机软件教研室、计算机及应用教研室有关领导和老师们对这批教材的初版和再版给予热情而具体的帮助，在此表示衷心的感谢。

相信这批再版计算机系列教材将会继续受普通高校、成人高等教育自学考试、函授大学、夜大学、广播电视大学、职工大学等计算机类有关专业师生的欢迎。这批系列教材也将继续是广大工程技术人员在职较系统地自学计算机知识颇好的读物。

华中理工大学出版社

1996 年 5 月

第二版前言

本书第一版自 1992 年出版以来,历时 5 年,印刷 8 次,发行近 5 万册,受到广大读者的支持与厚爱,在此表示衷心感谢。

数据库及与之不可分割的各类信息系统的应用在日益普及,数据库技术本身在快速发展,希望学习和掌握数据库技术的人员的应用水平与要求也不断提高。因此,本书再版时从整体结构、内容、组织、语言叙述等方面对第一版进行了很大的增删、调整、更新与修改,但原来的编写宗旨与基本风格仍保持未变。

全书分为三篇。第一篇数据库系统基础,第一至十章,涵盖了原第一至十章,其中第一章增加了数据库的现代发展,主要介绍分布式数据库、面向对象数据库和客户机/服务器数据库的基本概念。第三章增加了当前时兴的面向对象数据模型的基本知识。第四章对第一版中有关 SQL 和 ORACLE 的介绍,尤其是关于 SQL 标准部分作了较大的精简。在原第四章和第七章有关部分内容重新组织的基础上,再增加客户机/服务器数据库系统结构,加强了 ORACLE 开发工具与环境的内容,组成第七章。原第八章调整为第九章。新增加第八章数据库保护,介绍事务的概念和事务的管理,包括完整性、一致性、安全性、可靠性(或恢复)。原第九、十章经精简压缩后,重新写成第十章数据库运行与管理,着重叙述数据库的运行环境及运行性能的监视与维护,数据库管理员对数据库及其开发的管理。其余各章也都进行了适当增删与修改,充实了内容,加强了基本理论及实践应用。

第二篇 FoxPro 及其程序设计是新撰写的部分,是以一个实例和实验环境来介绍的,这满足了新大纲的要求。同时也使相关读者能学习或了解当前较为流行的 FoxPro 数据库系统及其使用情况,以满足他们的需要。

第三篇学习指导与实验,保持原有框架,根据前面各章内容的变动与修改,进行了相应修订,以作为教学或自学的指导与参考。

本书能适合多种层次、多种培养目标的人员学习需要,学习时只需作适当取舍即可。例如,本科生学习整个第一篇,参考第二篇(将它作为实验环境)。专科生则可着重学习第一篇的一至七章、九章,其中有的章节还可根据情况适当从简、从略;熟悉第二篇,了解第八章。作为一般用户或应用程序员,可重点学习第一篇中的一、三至七、九章,了解第二篇。若是数据库与信息系统的设计开发或管理人员,则应学习整个第一篇,尤其是,八、九和十章较之其他人员来说更是重点,阅读第二篇以作实验。第三篇对各类人员都可作参考、指导。

本书由刘云生主编,其中:第一篇的第一、五、八、十章由刘云生编写;第二、三、四章由卢正鼎编写;第六、七、九章由卢炎生编写。第二篇的第十一、十二章由刘云生编写;第十三章由卢炎生编写;第十四、十五章由卢正鼎编写;全篇由卢炎生审定。第三篇按第一版的分工分别由三人相应修订。全书由刘云生统稿、审定。

限于编者的知识与水平,书中不妥或谬误之处恐怕难免,恳请同行及读者指正。

编者

1997 年 7 月

于武昌喻家山华工园

目 录

第一篇 数据库系统原理

第一章 数据库系统概念	(1)
1.1 数据管理技术的进展	(1)
1.1.1 数据管理概述	(1)
1.1.2 传统的文件	(2)
1.1.3 文件管理系统	(4)
1.1.4 数据库方法	(4)
1.2 数据与联系	(8)
1.2.1 数据与信息	(9)
1.2.2 三个领域	(9)
1.2.3 数据项间的联系	(12)
1.2.4 记录间的联系	(12)
1.3 数据库系统的组成	(13)
1.3.1 数据库	(14)
1.3.2 数据库系统软件	(14)
1.3.3 数据库系统用户	(15)
1.3.4 数据库系统硬件	(16)
1.4 数据的分层视图	(16)
1.5 数据库与信息系统的联系	(18)
1.6 数据库的新发展	(19)
1.6.1 分布式数据库系统	(19)
1.6.2 面向对象数据库	(20)
1.6.3 客户机/服务器数据库系统结构	(21)
习题一	(22)
第二章 数据库存储结构	(23)
2.1 数据的外存组织	(23)
2.1.1 主要的外存储器——磁盘	(23)
2.1.2 物理块	(23)
2.1.3 逻辑记录的组块与解块	(24)
2.1.4 物理块的编址	(25)
2.2 基本文件组织	(26)
2.2.1 顺序文件组织	(26)
2.2.2 随机文件组织	(27)
2.2.3 链表组织	(28)
2.2.4 树状结构	(30)
2.3 索引结构	(30)
2.3.1 稀疏索引与稠密索引	(30)
2.3.2 B-树	(31)
2.3.3 B ⁺ 树	(35)

习题二	(37)
第三章 数据模型	(38)
3.1 数据模型概述	(38)
3.2 层次数据模型	(39)
3.2.1 模型结构	(39)
3.2.2 转换为层次树结构的方法	(40)
3.3 网状数据模型	(41)
3.3.1 模型结构	(41)
3.3.2 DBTG 网状模型	(42)
3.3.3 转换为 DBTG 网状模型的方法	(43)
3.4 关系数据模型	(45)
3.4.1 模型结构	(45)
3.4.2 对关系模型的几点说明	(46)
3.5 E-R 模型	(48)
3.5.1 语义数据模型概述	(48)
3.5.2 E-R 模型结构	(48)
3.5.3 E-R 模型的转换	(50)
3.6 面向对象数据模型	(50)
3.6.1 面向对象数据模型概述	(51)
3.6.2 面向对象数据模型的一些基本概念	(51)
习题三	(53)
第四章 数据库管理软件	(54)
4.1 数据库管理软件的主要目标	(54)
4.2 数据库管理系统	(56)
4.2.1 DBMS 功能	(56)
4.2.2 DBMS 组成	(57)
4.2.3 数据字典系统	(57)
4.3 SQL 标准简介	(58)
4.4 ORACLE 系统	(59)
4.4.1 ORACLE 系统概述	(59)
4.4.2 ORACLE RDBMS 与 ORACLE 产品	(61)
4.4.3 ORACLE 数据库的管理	(63)
4.4.4 ORACLE 应用开发工具	(66)
4.5 客户机/服务器数据库及其开发工具	(70)
4.5.1 客户机/服务器结构	(70)
4.5.2 SYBASE 的 C/S 结构	(72)
4.5.3 一些客户机端的应用设计开发工具	(73)
习题四	(75)
第五章 层次数据库系统	(77)
5.1 IMS 的系统结构	(77)
5.2 IMS 的数据结构	(78)

5.2.1	物理数据库	(78)
5.2.2	物理数据库的描述	(80)
5.2.3	层次序列	(81)
5.3	IMS 的逻辑数据库	(81)
5.3.1	逻辑数据库概念	(81)
5.3.2	逻辑数据库定义	(82)
5.3.3	逻辑数据库的处理	(84)
5.4	IMS 的数据存储结构	(84)
5.4.1	概述	(84)
5.4.2	HSAM 存储结构	(85)
5.4.3	HISAM 存储结构	(86)
5.4.4	HD 指针结构	(87)
5.4.5	HDAM 和 HIDAM 存储结构	(88)
5.5	IMS 的数据操作	(89)
5.5.1	DL/I 操作	(89)
5.5.2	数据库检索	(90)
5.5.3	数据库维护	(91)
习题五		(92)
第六章 网状数据库系统		(94)
6.1	DBTG 数据结构	(94)
6.1.1	数据对象	(94)
6.1.2	系	(95)
6.2	DBTG 数据描述	(97)
6.2.1	系的实现方式	(97)
6.2.2	系序	(99)
6.2.3	系成员籍	(100)
6.2.4	记录码	(101)
6.2.5	数据库码	(101)
6.2.6	当前值	(102)
6.2.7	保密及并发控制	(102)
6.2.8	记录定位方式	(103)
6.2.9	系值选择	(104)
6.3	DBTG 模式数据描述语言	(104)
6.3.1	DDL 结构	(104)
6.3.2	模式条目	(105)
6.3.3	域条目	(105)
6.3.4	记录条目	(105)
6.3.5	系条目	(106)
6.3.6	模式示例	(107)
6.4	DBTG 子模式数据描述语言	(108)
6.4.1	SDDL 构成	(109)

6.4.2 子模式示例	(110)
6.5 DBTG 数据操纵语言	(111)
6.5.1 控制语句	(111)
6.5.2 检索语句	(113)
6.5.3 更新语句	(117)
6.5.4 应用程序例	(119)
习题六	(121)
第七章 关系数据库系统	(123)
7.1 基本概念	(123)
7.1.1 关系模式	(123)
7.1.2 关系数据库	(123)
7.1.3 视图	(124)
7.2 关系数据库系统的数据描述	(123)
7.3 关系数据库系统的数据操作	(124)
7.3.1 关系代数	(124)
7.3.2 关系演算	(128)
7.3.3 SQL 语言	(130)
习题七	(140)
第八章 数据库保护	(142)
8.1 完整性	(142)
8.1.1 完整性限制	(142)
8.1.2 完整性检验	(142)
8.1.3 完整性检验的时机控制	(143)
8.2 安全性	(143)
8.2.1 用户识别	(144)
8.2.2 权限审定	(144)
8.3 并发控制	(145)
8.3.1 基本概念	(145)
8.3.2 三种不一致性	(146)
8.3.3 封锁	(147)
8.3.4 活锁与死锁	(148)
8.3.5 可串行性	(149)
8.3.6 两段封锁法	(150)
8.4 恢复	(151)
8.4.1 故障的类型	(151)
8.4.2 数据库的后备	(152)
8.4.3 日记	(152)
8.4.4 恢复过程	(153)
习题八	(154)
第九章 数据库设计	(156)
9.1 数据库设计步骤	(156)

9.2	需求分析	(158)
9.2.1	确定数据库范围	(158)
9.2.2	应用过程分析	(158)
9.2.3	数据的收集与分析	(160)
9.2.4	文档整理	(161)
9.3	概念设计	(162)
9.3.1	局部信息结构设计	(162)
9.3.2	全局信息结构设计	(164)
9.4	实现设计	(167)
9.4.1	实现设计的步骤	(167)
9.4.2	模型转换	(168)
9.5	物理设计	(168)
9.5.1	建立索引	(168)
9.5.2	建立聚集	(169)
9.6	实施与维护	(169)
9.6.1	数据库建立	(169)
9.6.2	应用程序设计	(170)
9.6.3	数据库试运行	(170)
9.6.4	运行与维护	(170)
9.7	关系规范化	(171)
9.7.1	函数依赖	(171)
9.7.2	关键字	(172)
9.7.3	关系规范化过程	(172)
9.7.4	关系分解的正确性	(175)
9.7.5	关系规范化理论的应用	(176)
习题九		(176)
第十章 数据库系统的运行与管理		(177)
10.1	数据库运行环境	(177)
10.1.1	运行环境的构成	(177)
10.1.2	运行环境的类型	(178)
10.1.3	运行环境的建立	(178)
10.2	数据库运行维护	(182)
10.2.1	数据库的运行	(182)
10.2.2	数据库性能监视与分析	(184)
10.2.3	数据库的重组与重构	(184)
10.3	数据库的管理	(187)
10.3.1	引言	(187)
10.3.2	数据库生命周期	(187)
10.3.3	数据库管理的职责	(188)
10.3.4	数据库管理员选择	(191)
10.4	数据库的开发管理	(191)

习题十	(193)
-----------	-------

第二篇 FoxPro 及其程序设计

第十一章 FoxPro 概述	(194)
11.1 FoxPro 系列简介	(194)
11.2 FoxPro 的技术指标	(195)
11.3 FoxPro 使用入门	(196)
11.3.1 运行环境	(196)
11.3.2 安装	(197)
11.3.3 启动与退出	(197)
11.3.4 用户接口	(197)
11.3.5 联机帮助	(199)
11.4 FoxPro 应用基础	(199)
11.4.1 数据及其类型	(199)
11.4.2 函数	(200)
11.4.3 表达式	(202)
11.4.4 文件	(202)
11.4.5 命令的使用	(204)
第十二章 数据库的建立与维护	(206)
12.1 数据库结构的建立	(206)
12.2 数据库结构的维护	(206)
12.2.1 数据库结构的显示	(207)
12.2.2 数据库结构的修改	(207)
12.2.3 数据库结构的拷贝	(207)
12.3 数据库的数据输入	(208)
12.3.1 初始数据录入	(208)
12.3.2 追加记录	(208)
12.3.3 备注字段的数据输入	(208)
12.3.4 其他数据输入方法	(209)
12.4 数据库的基本数据操作	(210)
12.4.1 数据库显示	(210)
12.4.2 记录的定位	(210)
12.4.3 数据库的查询	(210)
12.4.4 数据库的维护	(212)
12.4.5 数据库的浏览	(214)
12.5 数据库文件管理操作	(216)
12.5.1 打开与关闭	(216)
12.5.2 基本的文件操作	(217)
12.5.3 文件管理器	(218)
12.6 数据库的管理操作	(219)
12.6.1 排序(Sort)	(219)
12.6.2 索引(Index)	(220)

12.6.3	视图(View)	(221)
12.6.4	数据库的统计	(222)
12.6.5	数据库的筛选	(223)
12.6.6	数据库间的关联	(223)
第十三章	FoxPro 程序设计基础	(226)
13.1	FoxPro 程序的建立、运行和结束	(226)
13.1.1	程序的建立	(226)
13.1.2	程序的运行	(227)
13.1.3	程序的结束	(227)
13.2	内存变量的创建与使用	(228)
13.2.1	内存变量说明	(228)
13.2.2	内存变量建立	(229)
13.2.3	内存变量显示	(229)
13.2.4	内存变量释放	(230)
13.2.5	内存变量存储	(230)
13.2.6	内存变量恢复	(230)
13.3	数组	(231)
13.3.1	数组建立	(231)
13.3.2	数组的赋值、显示与释放	(232)
13.3.3	数组下标	(232)
13.3.4	数组操作函数	(233)
13.3.5	数组与数据库文件的数据交换	(235)
13.4	数据的输入与输出	(237)
13.4.1	非格式化输入	(238)
13.4.2	非格式化输出	(239)
13.4.3	格式化输入	(239)
13.4.4	格式化输出	(241)
13.5	FoxPro 程序结构	(241)
13.5.1	分支结构	(242)
13.5.2	循环结构	(244)
13.5.3	过程	(247)
13.5.4	自定义函数	(250)
第十四章	FoxPro 程序设计技巧	(253)
14.1	窗口设计	(253)
14.1.1	定义窗口	(253)
14.1.2	启动窗口	(254)
14.1.3	关闭窗口	(255)
14.1.4	其他窗口命令与函数	(255)
14.2	菜单设计	(256)
14.2.1	基本菜单建立命令概述	(256)
14.2.2	菜单的其他操作	(260)

14.3	报表设计简述	(261)
第十五章	局域网络环境下的 FoxPro	(262)
15.1	网络站点配置文件设置	(262)
15.2	并发控制	(263)
15.2.1	数据库文件共享与独占的设置	(263)
15.2.2	数据记录和文件的封锁和解锁	(263)
15.3	FoxPro 网络多用户命令与函数	(265)
15.4	FoxPro 的网络错误信息	(266)
第三篇 学习指导与实验		
第一部分	课程大纲	(267)
第二部分	学习辅导	(271)
第一章	数据库系统概念	(271)
第二章	数据库存储结构	(272)
第三章	数据模型	(273)
第四章	数据库管理软件	(274)
第五章	层次数据库系统	(275)
第六章	网状数据库系统	(276)
第七章	关系数据库系统	(277)
第八章	数据库保护	(278)
第九章	数据库设计	(279)
第十章	数据库的运行与管理	(281)
第三部分	习题解答	(282)
习题一		(282)
习题二		(282)
习题三		(284)
习题四		(285)
习题五		(285)
习题六		(287)
习题七		(289)
习题八		(292)
习题九		(293)
习题十		(296)
第四部分	实验	(297)
第一章	实验例子	(297)
第二章	实验内容	(298)
第三章	实验结果举例	(300)

第一篇

数据库系统原理

第一章 数据库系统概念

数据库技术的应用越来越广泛,几乎深入到各个领域、各个部门,因此人们普遍要求掌握数据库系统的基本原理。本章首先介绍数据库技术的发展,有关数据的基本概念和特征,然后介绍数据库系统的主要构成及其与信息系统的关系,最后简要地介绍数据库在几个方面发展的新概念。

1.1 数据管理技术的进展

社会的一切组织、人类思想和行动的各个领域都离不开信息。在各种组织和个人的活动中都需要决策,而决策正确与否主要依赖于信息是否准确、及时、完善和真实。数据是信息的载体,是表示信息的物理符号(可以是数字、字母或其他形式的符号),信息是数据处理过程的结果。

1.1.1 数据管理概述

与生产制造过程中将原材料变成成品一样,数据处理是将原始数据转换成信息的过程(图1.1)。依其使用技术的不同与设备的演变历史,数据处理的方式可以分为人工式(人工数据处理阶段,1800年以前)、机械辅助式(机械辅助阶段,1800~1890年)、机电穿孔卡片式(机电阶段,1890~1946年)、电子计算机式(电子阶段,1946年以后)。

产生信息的数据来源于两个方面:内部和外部。内源由组织内部的对象,如个人、部门等组成。外源就是环境源,它由处于组织之外的对象组成。它们都给组织决策提供事实,即数据。数据的处理就是指对数据的收集、分类、整理、计算、精炼、储存、提取、传递等。

每个组织单位都保存了各种各样大量的数据,如企业中有关各种生产计划、生产组织与调度、工艺技术、设备与工具、物料供应、经营销售的数据;学校中关于学生和教职员的档案、开设的课程、学生成绩等方面的数据;医院中的病历、药物管理、诊断与治疗记录、财务结算的数据

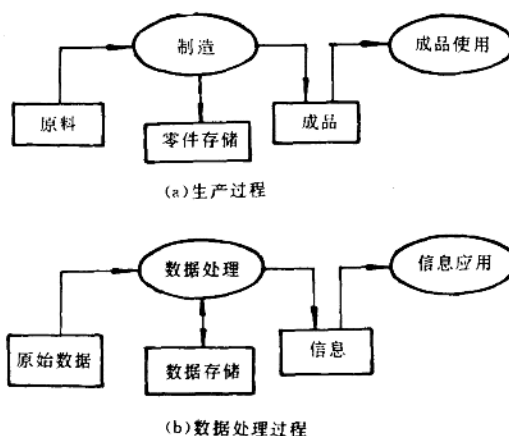


图 1.1 变换过程

等等。与其他资源,如人、财、物一样,数据也是组织的一种珍贵资源,甚至比其他资源更珍贵。虽然各种资源的作用不同,但它们都有一些共同特征:①花费了代价;②对组织有价值;③为整个组织统一管理与共享。数据也完全具有这些特征。事实上,从某种意义上讲,数据更为重要,因为组织对其他各种资源的有效管理与合理使用都依赖于有关数据的有效性,任何不将数据作为一种资源并给予有效管理的组织,在如何管理人、财、物等资源方面都将陷入窘境。

对数据进行管理可以分为两个方面:一是侧重于组织业务的管理,负责制订并执行整个组织中关于数据的定义、组织、保护与有效使用的政策、过程、实践和计划;二是侧重于技术,负责实现数据作为一种资源的集中控制管理。数据管理的目标就是要在适当的时刻以适当的形式给适当的人提供适当的数据。为此,必须进行数据的收集、组织、控制、存储、选取、维护,这也就是数据管理的任务。

要处理数据,首先必须以便于处理的某种方式收集数据,并将记录在纸介文件上的数据转换成计算机可处理的形式。将收集的数据进行适当的构造,则称为数据组织。数据的组织分为逻辑组织和物理组织两种:数据的逻辑组织是用户(或应用程序)所使用的数据结构形式;其物理组织是数据在物理存储设备上的结构形式,两者之间可以相对独立。为了备用,需要将数据归类存入文件。显然,应该是其价值超过存储代价的数据,才值得存储。为了向用户提供信息,存储的数据要能够方便地被选择提取,这叫作检索。为了保护数据的正确性与安全性等,必须建立一些相应的规则和执行这些规则的过程来控制数据的存与取,以实现前面所述数据管理的目标。保管的数据必然是要较长期地多次使用的数据,但随着时间的推移,组织单位的内外环境会发生变化,因而存储的数据在量和结构方面都可能变化,所以数据管理必须能对其进行维修。下面我们简述几种电子数据处理技术,它们的排序则描绘了计算机数据处理的发展里程,也反映了数据库系统的演变过程。

1.1.2 传统的文件

在早期的数据管理系统中,数据处理过程如图 1.2 所示,即注意力集中于处理功能,主要面向批处理,而数据起着辅助作用。这样的数据处理系统是面向单个数据处理应用的,因为尽管系统中可以包含多个处理功能,但开发者往往是一次一个地开发应用程序来响应用户的单

个应用要求。进行这种系统开发时,首先说明功能,然后导出数据需求,因此,每一应用只有它自己的被设计成满足其特定需要的输入文件和输出文件,没有一个总体的布局、计划或模型来指导系统的开发或应用的增长。系统所提供的管理功能也仅仅是一些简单的 I/O 操作。

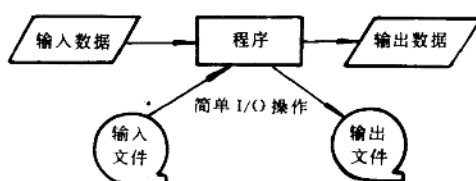


图 1.2 传统的数据处理

当应用增加时,必然引起两种情况发生。第一,由于有的数据与多种应用相关,故新文件中的许多乃至全部数据可能在对应于其他应用的已有文件中已存在,然而为了满足新的应用要求,这些已有的文件又不得不重新构造。第二,因为每一应用都是独立地开发的,故对于同一信息常常可能会使用不同的表示;反之,相同的数据表示亦可能代表不同的信息内容。

以这种方法管理数据,除了由于同一数据有多个副本而产生大量冗余,浪费大量的存储空间外,更严重的是难以维护数据的一致性。例如,当一个数据变更时,无法自动将这种变更由一个文件传播到另一文件,而必须分别逐个地由各用户去修改。这样就很难保证其一致性,而且久而久之,不同文件中的相同数据之间的差别会越来越大,使得同一数据在不同文件中完全不同。数据表示的这种不一致性极大地阻碍着信息的交流与共享。

在这种文件系统中,文件与记录的描述安置在每一个应用程序内,且直接给出它们的物理结构,即实质上并无逻辑结构与物理结构之分,因而每一程序“自我拥有”其数据文件,程序逻辑与数据的格式和描述彼此紧密地相关联,毫无各自的独立性。

这种传统的文件处理系统针对各个应用来设计各自的文件,十分简单,处理效率高,数据管理相当容易。但它有许多固有的缺点,主要是:

(1)无控冗余 如上所述,在这种系统中各个应用有各自的文件,其中包含了大量冗余数据。这样既浪费了宝贵的存储空间,又对同一数据更新时,需要输入多次来逐个修改各重复版本的值;还要花时间来处理和改正所产生的不一致性。

(2)不一致数据 这是引起计算机应用出错的普遍根源之一,它们将导致不一致的资料和报表,并会削弱用户对信息完整性的信任度。

(3)数据缺乏独立性 在这种系统中,文件和记录及数据项的描述、存取方法以及物理的 I/O 语句都嵌入在单个的应用程序中,这样程序和数据文件任何一方的有关变动都必定引起对方的修改。系统对新的或变更的信息所要求的响应灵活性很差,任何变动都要求重构造新的文件,修改或设计新的应用程序。这导致了大量的程序维护(修改现有程序的过程称为程序维护),降低了程序员的生产效率,因为,重构造文件与描述数据、重写 I/O 语句等对每一应用程序都要重复进行,且这些均是构成系统开发工作的主要部分。

(4)数据难以共享 各应用程序拥有它自己的私有数据文件,彼此之间没有什么共享数据的机会。这又促成了上述三个缺点。

(5)无法施加统一标准 在一个组织中,必定有关于数据命名、格式、存取限制等各种标准。然而,在传统的文件处理系统中,很难施加这些标准,因而往往带来“同名异物”和“同物异名”的情况,这种不统一性阻碍了数据的独立和可共享性。

1.1.3 文件管理系统

上面叙述了早期的文件处理系统存在许多缺点,在那里文件固定地存储在顺序介质上,所以只有顺序文件这一种组织形式。到60年代中期,直接存取存储设备的出现使人们能研制专门的文件管理系统,如图1.3所示。它将数据的逻辑结构和物理结构分离,由“存取方法”实现逻辑结构与物理结构之间的映射。应用程序只涉及数据的逻辑结构,系统决定数据的物理结构,两者之间可以有差别。这样,当物理结构改变时,不会导致应用程序的修改,这叫数据的物理独立性。数据的物理独立性使应用程序脱离数据的物理结构,使其适用性得以提高。同时,应用程序员不必关心数据的物理存储细节,因而其生产效率也得到提高。

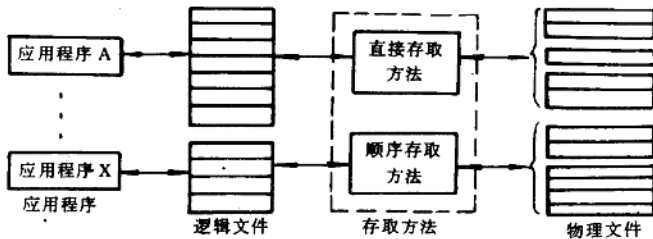


图 1.3 文件系统的数据处理

文件系统提供了多种文件组织形式,如不仅有顺序文件组织,还可以有直接文件组织等。与这些文件组织形式相对应,系统提供了多种存取策略,如顺序存取、随机存取、相对地址存取等,应用程序员可以根据应用的需要进行选择。此时,系统不再仅提供简单的 I/O 操作命令,还能进行记录操作及一些数据组织、管理控制、维护等操作,如索引、排序、文件复制、存储设备或格式的转换、备份以及两个文件的比较等。

与早期的文件处理系统相比,文件(管理)系统已有很大的改进。它提供了物理数据独立性,使应用程序与数据的具体物理存储结构分离,并通过数据的抽取、排序、合并等可以为应用提供新的文件,从而使数据共享成为可能。但在文件系统中,不能实现数据的普通共享,只能实现文件级而不能在记录或数据项级实现数据的共享。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。当不同应用程序使用的数据(记录或数据项)大部分相同(而只有少部分不同)时,还必须构造各自的文件,这样仍然还存在大量的数据冗余。

1.1.4 数据库方法

随着计算机技术与工业的迅速发展,计算机日益广泛地应用于企业管理,这对计算机数据管理提出了更高的要求。首先,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,因此应大量地消去数据冗余,节省存储空间。其次,当数据变更时,能节省对多个数据副本的多次变更操作,从而可大大缩小计算机时间开销,且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,还要求数据具有更高的独立性,不但具有物理独立性,而且具有逻辑独立性,即当数据逻辑结构改变时,不影响那些不要求这种改变的用户的程序,从而节省应用程序开发和维护的代价。所有这些,用文件系统的数据库管理方法都不能满足,这导致了数据库技术的发展。

数据库方法代表了不同的数据管理观点,它将数据视为一种与人、财、物同等重要,甚至更

重要的组织资源,所以要像其他资源一样来统一管理、控制、共享使用。数据库的概念起源于“共享”数据资源、“离释”数据的个人控制,并将数据管理作为公共职能权力,进行合作与协调的数据资源的统一控制与维护。

数据库与文件不同,一个基本的差别在于它们的使用形式不同。文件一般限制于一个或少数几个用户,且只有一种为应用程序(通常是少数)共享的文件逻辑视图;而数据库将各种数据集成在一起,且使各种用户能以不同的数据库逻辑视图共享数据库。按数据库方法,数据处理视图如图 1.4 所示。它是面向数据而不是面向程序的,在这里,数据库处于中心地位,各处理功能处于外围,它们都通过数据库管理软件自数据库中获取所需数据和存储处理结果。

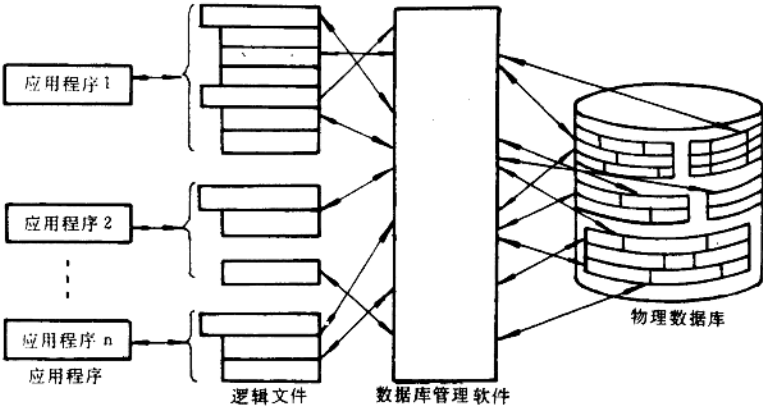


图 1.4 数据库系统的数据处理视图

为说明数据库方法的主要优点,下面首先举一个关于劳资人事信息管理的例子。它涉及到人事、劳资、工资发放、业务技术等方面,如图 1.5 所示。按传统的方法,它们完全是孤立的,各自单独处理,在每一部门的系统中,各个应用也是独立的。因此,根本没有什么数据共享;同一

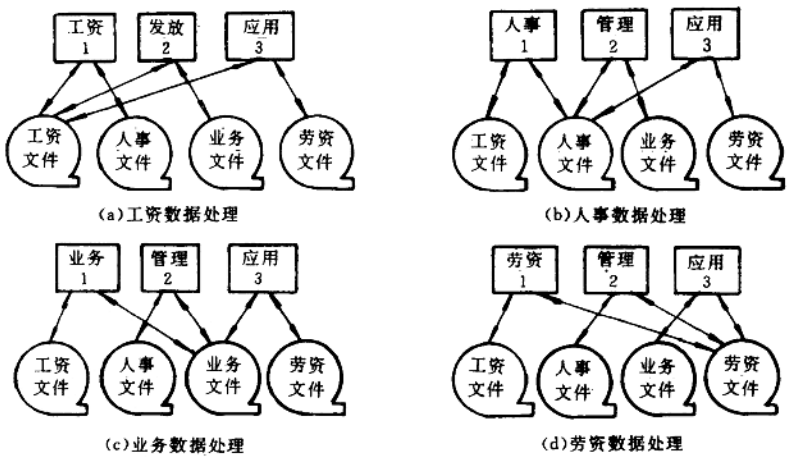


图 1.5 各自独立的数据处理