

数据库系统概论



刘云生 卢正鼎 卢炎生

数据库系统 概论

华中理工大学出版社

3143

/S/1

数据库系统概论

刘云生 卢正鼎 卢炎生

华中理工大学出版社

内 容 简 介

本书较系统、全面而又简明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：数据库系统组成结构(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)，数据库设计方法与技术(第八章)，数据库的运行、维护与管理(第九~十章)。每章后都附有习题。为了便于自学和加强实践，在“学习指导与实验”中给出了课程大纲、学习辅导与习题解答，还安排了以一个应用例子贯串全过程的四个数据库开发实验。

本书可作为高等院校计算机各专业及有关非计算机专业数据库课程的教材，亦可供计算机应用研究工作者、工程技术人员及其他有关人员参考。

数据库系统概论

刘云生 卢正鼎 卢炎生

责任编辑 吴凤萍

*

华中理工大学出版社出版发行

(武昌喻家山)

新华书店湖北发行所经销

华中理工大学出版社沔阳印刷厂印刷

*

开本：787×1092 1/16 印数：19 字数：435 000

1992年9月第1版 1994年3月第3次印刷

印数：8 001—13 000

ISBN7—5609—0709—1/TP·71

定价：11.20元

(鄂)新登字第10号

《计算机系列教材》序

计算机是20世纪最伟大的发明之一。在迎接世界新技术革命的挑战和加快我国四个现代化建设的进程中,大力推广应用计算机具有很现实的和深远的意义。目前,计算机特别是微型计算机的应用已广泛渗透到我国社会和国民经济的各个领域。要使我们计算机的开发和应用进一步向深度和广度发展,以便为我国的社会主义建设带来更大的经济效益和社会效益,其中一个很重要、很迫切的任务就是,要迅速培养一大批主要从事计算机应用工作的专业人才。为此,我校已设置了“计算机及应用”专业。

为满足计算机专业教学工作的需要,我们特邀请和组织华中理工大学计算机系具有丰富教学经验的有关课程的一批骨干教师编写了这套计算机系列教材。该系列教材包括《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》和《数据库系统概论》等8种,并均由华中理工大学出版社出版。

该系列教材是参照计算机专业教学大纲,按照面向应用、重视实践、便于自学的原则和以广泛使用的IBM-PC微机作为教学机进行编写的。为了使该系列教材成为符合上述原则的统一整体,在教材的编写过程中,对于各书的内容、取材、界面、衔接及格式等问题,都反复地进行了研讨和协商。同时,为了便于读者自学,各书中均编写了“学习指导”方面的内容。因此,该系列教材不仅可供普通高校计算机类有关专业使用,而且也可供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使用。同时,该系列教材也是广大工程技术人员和其他在职人员较系统地自学计算机知识与技术颇为适宜的读物。

华中理工大学成人教育学院胡润清副研究员、郑青林助理研究员等领导同志具体担负了该系列教材的规划、组织工作。

该系列教材在编、审和出版过程中,华中理工大学计算机系李崇阁副研究员、冯著明副教授等领导同志以及有关课程的教师做了大量的工作,华中理工大学出版社给予了大力的支持与帮助,在此一并表示感谢。

限于水平与经验,该系列教材肯定会有许多缺点和不足,诚望有关专家和广大读者积极提出批评与建议,共同为不断提高该系列教材的质量而努力。

华中理工大学成人教育学院

1992年4月

前 言

本书是参照计算机专业教学大纲编写的。

数据是社会组织的重要资源,必须对它进行科学的管理与使用,这在当今信息社会尤为重要。数据库系统就是一门研究有关计算机数据管理理论、方法与技术的现代学科,它发展迅速,其应用正在日益广泛地深入到科学与社会的各个领域和部门,因而它越来越普遍地受到人们的关注与重视。

我们在多年从事数据库系统教学与科研的基础之上编写了这本教材。本书作为计算机专业系列教材之一,在编写时力求在内容上既考虑到系统性和完整性,又照顾到实际需要,在阐述上既突出基本概念与理论、基本原理与方法、基本技术与工具,又体现重视应用、加强实践、便于自学的原则。

本书共分十章并附有“学习指导与实验”。第一章概括性地叙述了数据库系统的基本概念、基本原理,使读者对什么是数据库系统、为什么要有数据库系统有一个基本认识。第二章扼要地介绍了在实际的数据库中数据的组织方法与技术,使读者了解数据库系统如何存贮大量的数据,怎样使之易于检索与维护。第三章介绍了目前最流行的三种主要数据模型——层次、网状和关系模型以及在数据库设计中作为描述现实世界的有力工具的 E-R 模型。第四章介绍了以 DBMS 为核心的数据库管理软件,着重叙述了 DBMS 的功能及主要组成部件,并作为实例较具体地介绍了具有代表性的关系数据库语言 SQL 和关系数据库管理软件系统 ORACLE 的组成结构与使用方法。第五~七章分别叙述了层次、网状和关系数据库管理技术与方法。第八章介绍数据库设计方法学,包括设计的过程、步骤、各步的主要任务和使用的技术与工具。第九、十章叙述数据库的运行、维护与管理。“学习指导与实验”包括了教学大纲、学习辅导、习题解答和四个实验,四个实验以同一个例子串成一个完整的数据库开发实践过程。这样有利于自学和加强实践。

本书由刘云生副教授主编,其中:第一、五、九、十章由刘云生编写;第二、三、四章由卢正鼎编写;第六、七、八章由卢炎生编写;“学习辅导与实验”由三人共同编写,主要由卢炎生、刘云生执笔。全书由刘云生统稿。

在本书的编写过程中得到了华中理工大学成人教育学院、计算机系与软件教研室、华中理工大学出版社等各方面的领导与同志们的大力支持与帮助,谨此一并表示衷心感谢。

限于编者水平,书中不妥或错误之处必定难免,敬请读者指教。

编 者

1992年3月

于武昌喻家山华工园

目 录

第一章 数据库系统概念	(1)
1.1 数据处理技术	(1)
1.1.1 数据处理概述.....	(1)
1.1.2 传统的文件.....	(2)
1.1.3 文件管理系统.....	(3)
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.....	(8)
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	(13)
1.3 数据库系统的组成	(14)
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)
习 题 一	(19)
第二章 数据库存贮结构	(20)
2.1 数据的外存组织	(20)
2.1.1 主要的外存贮器——磁盘	(20)
2.1.2 物理块	(20)
2.1.3 逻辑记录的组块与解块	(21)
2.1.4 物理块的编址	(22)
2.2 基本文件组织.....	(23)
2.2.1 顺序文件组织	(23)
2.2.2 随机文件组织	(24)
2.2.3 链表组织	(25)
2.2.4 树状结构	(27)
2.3 索引结构	(27)
2.3.1 稀疏索引与稠密索引	(27)
2.3.2 B ₊ 树	(29)
2.3.3 B ⁺ 树	(32)
习 题 二	(34)
第三章 数据模型	(35)

3.1	数据模型概述	(35)
3.2	层次数据模型	(36)
3.2.1	模型结构	(36)
3.2.2	到层次树结构的转换	(37)
3.3	网状数据模型	(38)
3.3.1	模型结构	(39)
3.3.2	DBTG 网状模型	(39)
3.3.3	到 DBTG 网状模型的转换	(40)
3.4	关系数据模型	(42)
3.4.1	模型结构	(42)
3.4.2	对关系模型的几点说明	(44)
3.5	E-R 模型	(45)
3.5.1	语义数据模型概述	(45)
3.5.2	E-R 模型结构	(46)
3.5.3	E-R 模型的转换	(47)
	习 题 三	(48)
	第四章 数据库管理软件	(49)
4.1	数据库管理软件的主要目标	(49)
4.2	数据库管理系统 DBMS	(51)
4.2.1	DBMS 功能	(51)
4.2.2	DBMS 组成	(52)
4.2.3	数据字典系统	(53)
4.2.4	一些 DBMS 产品简介	(54)
4.3	SQL 标准	(55)
4.3.1	SQL 标准的结构	(56)
4.3.2	公用成分	(57)
4.3.3	模式定义语言	(64)
4.3.4	数据操纵语言	(65)
4.3.5	嵌入 SQL 的主程序	(68)
4.4	ORACLE 系统	(70)
4.4.1	ORACLE 系统概述	(70)
4.4.2	ORACLE RDBMS 构成与 ORACLE 产品	(71)
4.4.3	ORACLE 数据库的管理	(73)
4.4.4	ORACLE 应用开发工具	(78)
	习 题 四	(86)
	第五章 层次数据库系统	(88)
5.1	IMS 的系统结构	(88)
5.2	IMS 的数据结构	(89)
5.2.1	物理数据库	(89)
5.2.2	物理数据库的描述	(90)
5.2.3	层次序列	(92)
5.3	IMS 的逻辑数据库	(92)

5.3.1	逻辑数据库概念	(92)
5.3.2	逻辑数据库定义	(93)
5.3.3	逻辑数据库的处理	(95)
5.4	IMS 的数据存贮结构	(95)
5.4.1	概述	(95)
5.4.2	HSAM 存贮结构	(96)
5.4.3	HISAM 存贮结构	(97)
5.4.4	HD 指针结构	(98)
5.4.5	HDAM 和 HIDAM 存贮结构	(98)
5.5	IMS 的数据操作	(101)
5.5.1	DL/I 操作	(101)
5.5.2	数据库检索	(101)
5.5.3	数据库维护	(103)
习题五	(104)
第六章	网状数据库系统	(105)
6.1	DBTG 数据结构	(105)
6.1.1	数据对象	(105)
6.1.2	系	(107)
6.2	DBTG 数据描述	(109)
6.2.1	系的实现方式	(109)
6.2.2	系序	(110)
6.2.3	系成员籍	(112)
6.2.4	记录码	(113)
6.2.5	数据库码	(113)
6.2.6	当前值	(114)
6.2.7	保密及并发控制	(115)
6.2.8	记录定位方式	(116)
6.2.9	系值选择	(117)
6.3	DBTG 模式数据描述语言	(118)
6.3.1	DDL 结构	(118)
6.3.2	模式条目	(118)
6.3.3	域条目	(119)
6.3.4	记录条目	(119)
6.3.5	系条目	(120)
6.3.6	模式示例	(120)
6.4	DBTG 子模式数据描述语言	(122)
6.4.1	SDDL 构成	(122)
6.4.2	子模式示例	(124)
6.5	DBTG 数据操纵语言	(124)
6.5.1	控制语句	(124)
6.5.2	检索语句	(126)
6.5.3	更新语句	(131)
6.5.4	应用程序例	(135)

习题六	(136)
第七章 关系数据库系统	(138)
7.1 基本概念	(138)
7.1.1 关系模式	(138)
7.1.2 关系数据库	(138)
7.1.3 视图	(138)
7.2 关系数据库系统的数据描述	(139)
7.2.1 关系数据库模式定义	(139)
7.2.2 视图定义	(140)
7.2.3 索引定义	(140)
7.3 关系数据库系统的数据操作	(141)
7.3.1 关系代数	(141)
7.3.2 关系演算语言	(146)
7.3.3 SQL 语言	(148)
习题七	(159)
第八章 数据库设计	(161)
8.1 数据库设计步骤	(161)
8.2 需求分析	(162)
8.2.1 确定数据库范围	(163)
8.2.2 应用过程分析	(163)
8.2.3 数据的收集与分析	(164)
8.2.4 文档整理	(166)
8.3 概念设计	(167)
8.3.1 自顶向下方法	(167)
8.3.2 关系的规范化	(178)
8.4 实现设计	(184)
8.4.1 实现设计的步骤	(184)
8.4.2 模型转换	(185)
8.5 物理设计	(189)
8.5.1 关系数据库物理设计	(189)
8.5.2 网状数据库物理设计	(190)
8.5.3 层次数据库物理设计	(191)
习题八	(192)
第九章 数据库系统运行	(194)
9.1 数据库运行环境	(194)
9.1.1 运行环境的构成	(194)
9.1.2 运行环境的类型	(195)
9.1.3 运行环境的建立	(195)
9.2 应用程序的运行	(199)
9.3 数据库系统用户接口	(201)
9.3.1 用户特征	(201)
9.3.2 接口特性	(202)

9.4	处理接口方式	(203)
9.5	数据库性能监视与分析	(205)
9.6	数据库的重组与重构	(206)
	习 题 九	(208)
第十章	数据库的管理	(210)
10.1	引言	(210)
10.2	数据的集成管理	(210)
10.3	数据库管理功能	(212)
10.3.1	数据库生命周期	(212)
10.3.2	数据库管理的职责	(213)
10.3.3	数据库管理接口	(215)
10.4	数据库规划	(217)
10.4.1	数据库规划的重要性	(217)
10.4.2	数据库的战略目标	(218)
10.4.3	应用环境描述	(219)
10.4.4	规划的过程与任务	(220)
10.5	数据库设计过程	(221)
10.6	数据库管理的组织	(222)
10.6.1	数据库管理员选择	(222)
10.6.2	数据库管理职能的地位	(223)
10.6.3	数据库管理组织结构	(224)
10.6.4	分散组织中的 DBA 结构	(225)
10.7	代价与效益分析	(226)
10.7.1	效益的标识与定量表示	(226)
10.7.2	代价因子	(227)
10.7.3	代价效益比分析	(228)
	习 题 十	(228)
	学习指导与实验	(229)

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。本章将介绍数据库系统的基本概念,包括数据库技术的发展,有关数据的基本概念与特征,数据库系统的主要构成及其与信息系统的关系。

1.1 数据处理技术

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,而决策是否正确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 数据处理概述

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)。

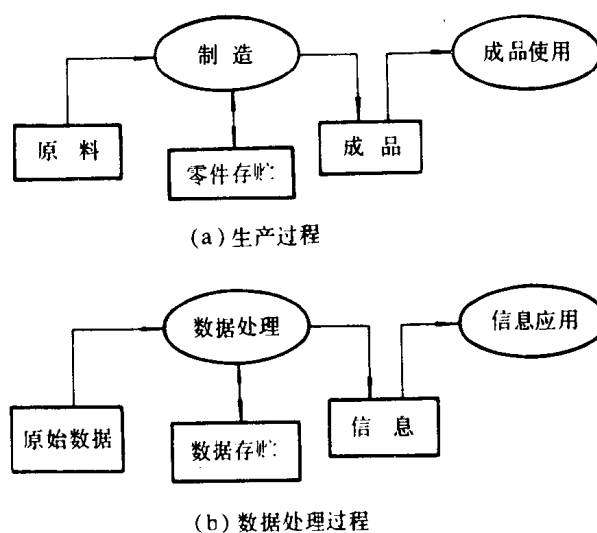


图 1.1 变换过程

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、贮存、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、学生成绩记录等方面的数据；医院中的病历、药物管理、诊断与治疗记录、财务结算数据等等。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妥当的时刻以妥当的形式给妥当的人提供妥当的数据。为此，必须进行数据的收集、组织、控制、存贮、选取、维护。

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，收集的数据需要进行适当的构造，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贮设备上的结构形式，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贮代价的数据，才值得存贮。为了向用户提供信息，存贮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，因而存贮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

在早期的文件处理系统中，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。系统所提供的数据管理功能也仅仅是一些简单的 I/O 操作。

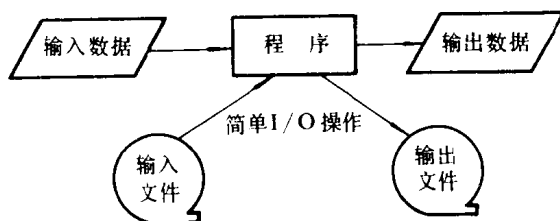


图 1.2 传统的数据处理

当应用增加时，必然引起两种情况发生。首先，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，这些已有的文件也不得不重新构造。其次，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,浪费大量的存贮空间外,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,一是浪费了宝贵的存贮空间;二是同一数据的更新要输入多次来逐个修改各重复版本的值;三是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

上一节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贮在顺序介质上,所以也只有顺序文件这一种组织形式。到 60 年代中期,直接存取存贮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贮细节,因而其生产效率也得到提高。

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、存贮设备

或格式的转换、备份以及两个文件的比较等。

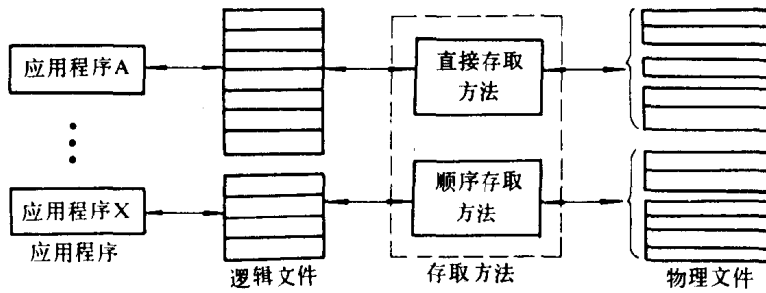


图 1.3 文件系统的数据处理

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贮结构分离,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,为企业的各种用户所普遍共享,从而应大量地消去数据冗余,节省存贮空间。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,因此导致了数据库技术的发展。

数据库方法代表了不同的数据处理观点,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贮处理结果。

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;同一数据多次重复,冗余量极大;根本无统一的标准与控制,甚至无统一的数据理解与定义;同一数据可能有多个不同的值;也难以实施安全保密控制。采用数据库方法则可以克服这些缺点,且

正是这些方面显示了它的优越性。下面分别叙述数据库方法的主要优点。

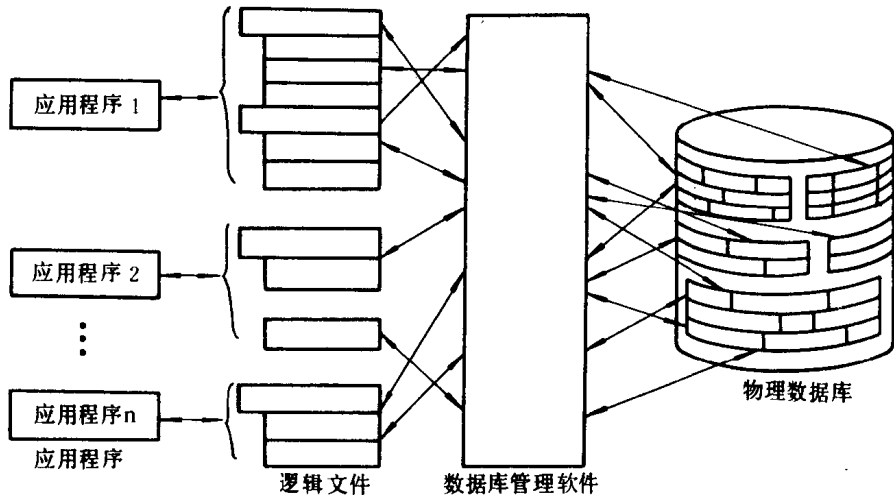


图 1.4 数据库系统的数据处理视图

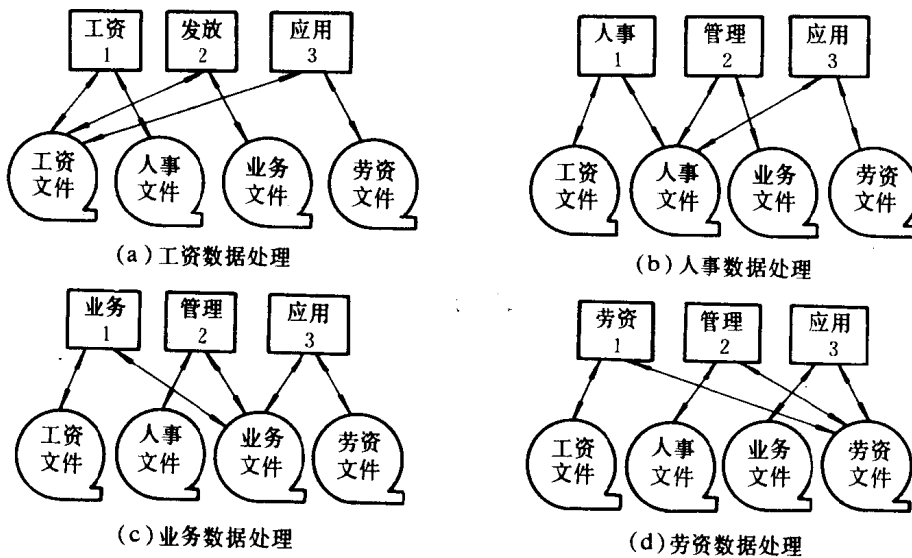


图 1.5 各自独立的数据处理

1. 数据集成

数据的集成是数据库管理系统的主要目的。通过数据集成来统一计划与协调遍及各相关应用领域的信息资源,这样可使数据得到最大程度的共享,而冗余最少。图 1.6 是图 1.5 所示系统的分离数据管理形式,它虽然实现了数据共享,但在这个蜘蛛网式的错综复杂的系统中,数据冗余量还是很大。而且,修改或扩充系统的任何一部分都极其困难,花费的代价极高,其原因在于存贮数据的高度重叠或冗余,以及一个应用到另一个应用存在复杂的转换。图 1.7 表明了一个集成的人事工资信息系统,其中的人事工资数据库由所有相关应用共享。一个全面的人事工资数据库易于容纳改变和接受新增加的数据,因而系统在修改或开发新的应用时,也可共享数据库。因为,在数据库中通过相联数据间定义的逻辑联系,数据被组织成统一的逻辑结构(这些工作由数据库管理软件实现),与数据的物理组织与定位分离,而应用的修改与增加只与

数据的逻辑结构发生关系。

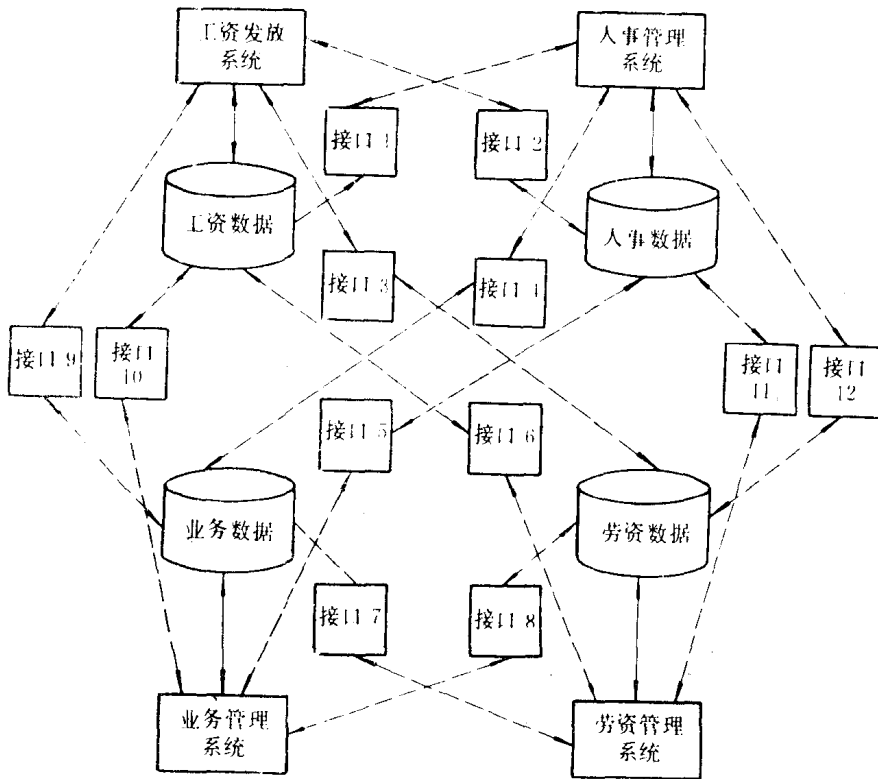


图 1.6 分离的数据管理

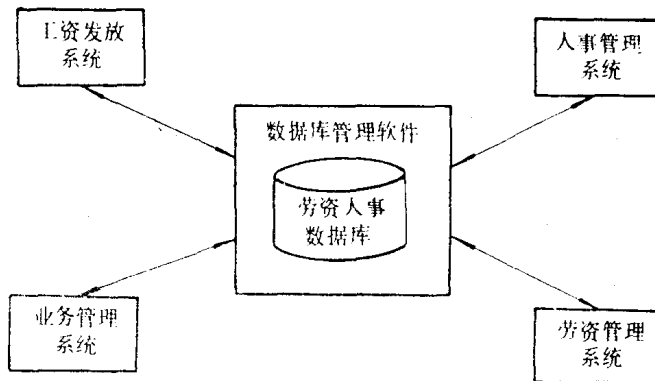


图 1.7 集成的数据管理

2. 数据共享

数据共享乃指在数据库中,一个数据可以为多个不同的用户共同使用,即各个用户可以为不同的目的来存取相同的数据。这种共享实际上是数据库集成的产物。例如在上述人事工资系统的例子中,关于职工记录中的数据“姓名”、“性别”、“部门”、“工资”等可以为人事部门、劳资部门、工资发放部门以及业务档案管理部门的各个用户共享。由数据库集成而产生的另一结果是任何给定的用户通常只与整个数据库的某一子集相关,而且不同用户的相关子集在许

多方面可以重叠。换句话说,不同的用户可以从各种不同的角度来看待数据库,即一个数据库有多种不同的用户视图。这些用户视图简化了数据的共享,因为它们可给每一用户提供执行其业务职能所要求的数据的准确视图,使用户无需知道数据库的全部复杂组成。

共享不只是指同一数据可以为多个不同用户存取,还包含了并发共享,即多个不同用户同时存取同一数据的可能性。此外,不仅为现有的应用(用户)共享,还可开发新的应用来针对数据库中同样的数据进行操作。换句话说,现有数据库中的数据可能满足将来新应用的需要而无需建立任何新的存贮数据文件。当前大多数数据库系统允许多个用户并发地共享一个数据库,尽管可能会有某些限制。

3. 数据冗余最少

在非数据库系统中,每个应用拥有它各自的文件,这常常带来大量的数据冗余,如上述工资发放应用、人事应用、劳资应用和业务档案应用,每一个都可能拥有一个包含职工信息(如职工号、姓名、性别、职称、工资等)的文件。对于数据库方法,如前所述,这些分立而有冗余的数据文件被集成为单一的逻辑结构,而且每一数据项值可以理想地只存贮一次。

我们并没有认为所有的冗余都可以或应该消除,有时,由于应用业务或技术上的原因,如数据合法性检验、数据存取效率等方面的需要,同一数据可能在数据库中保持多个副本。但是,在数据库系统中,冗余是受控的。系统知道冗余,保留必要的冗余也是系统预定的。

4. 数据一致性

通过消除或控制数据冗余,可以在一定范围内避免数据的不一致性。例如,假定张三的工资“135元”这个数据存贮在数据库的两个不同记录中,则当张三工资变动而要更新他的工资额时,若无控制,且只更新一个记录,则会引起同一数据的两个副本的不一致性,此时称数据库是不一致的。

显然,引起不一致性的根源是数据冗余。若一个数据在数据库中只存贮一次,则根本不可能发生不一致性。然而,如上段所述,冗余在数据库中是难免的,但它是受控的,所以当发生更新时,数据库系统本身可以通过更新所有其他副本来自动保证数据的一致性。遗憾的是当前许多数据库管理系统都不管数据冗余,因而也不支持数据一致性的自动保证。如 dBASE- III 就是一个典型的例子。

5. 实施统一标准

数据库对组织的数据进行集中管理控制,但数据库必须由人实现和进行维护管理,所以—个数据库系统必须包括一个称为数据库管理的组织机构(DBA)。数据库管理的职能包括管理和技术两个方面(见第九章)。在管理上负责制定并实施有关组织数据的命名、组织结构、格式及有效使用等方面的统一标准和控制过程。统一标准的数据有利于共享与彼此交换,有利于数据定义的重叠或冲突问题的解决以及今后的变更。

6. 统一安全、保密和完整性控制

DBA 机构对数据库有完全的管辖权且负责建立对数据的加入、检索、修改、删除权限及有效性的检验过程,可以对数据库中各种数据的每一类型的操作建立不同的检验过程。这种集中控制和标准过程较之分散数据文件的系统加强了对数据库的保护,使数据的定义或结构与数据之间的使用发生冲突的可能性最小。在检验控制方面,数据库比传统的文件危险性更大,因为它牵涉的用户更多。

7. 数据独立

数据说明与使用数据的程序分离称为数据独立。换句话说,就是数据或应用程序的修改不