

数据库

原理与技术

周志逵 编著

吴鹤龄 审校

DATABASE

DATABASE

DATABASE

科学出版社

数据库原理与技术

周志逵 编著

吴鹤龄 审校

科学出版社

1 9 9 4

(京)新登字 092 号

内 容 简 介

本书全面地叙述数据库系统的基本概念、原理和技术。全书共十章：数据库系统概论；网状数据库；层次数据库；关系数据库；典型关系数据库系统；关系数据设计理论；数据库保护；数据库设计；分布式数据库系统；面向对象数据库。本书内容取材新颖，重点突出，尽量反映国内外最新研究成果，并注重理论联系实际，叙述简炼易懂，除第十章外，其余各章均配有习题。

本书可作为高等院校计算机、经济管理、信息管理等有关专业数据库课程的教材，亦可供从事计算机研究、开发和应用的科研人员、工程技术人员和其他科技工作者学习和参考。

数 据 库 原 理 与 技 术

周志遼 编著

吴鹤龄 审校

责任编辑 徐津津

科学出版社出版

北京东黄城根北街 16 号

邮政编码：100717

中国人民解放军第一二〇二工厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1994 年 9 月 第 一 版 开本：787×1092 1/16

1994 年 9 月 第 一 次 印 刷 印张：17张

印数：1—4 500 字数：400 000

ISBN 7-03-004337-5/TP·392

定价：12.50 元

前 言

数据库是 60 年代末发展起来的一门新技术,它的出现使数据处理进入了一个崭新的时代。它能把大量的数据按照一定的结构存储起来,在数据库管理系统的集中管理下,实现数据共享。由于数据库具有数据结构化,数据独立性高、冗余度小,数据共享,便于使用,易于扩充等特点,被广泛地应用于各种管理信息系统。在当今信息化的社会里,数据库已成为管理和利用信息资源不可缺少的工具。数据库还广泛地应用于办公自动化,计算机辅助设计,人工智能等领域。可以毫不夸张地说,计算机的广泛应用和普及,与数据库应用是不无关系的。

数据库的广泛应用,给计算机应用开辟了广阔的天地,受到了人们的极大关注,学习和应用数据库,研究数据库技术成了广大科技工作者和管理人员的共同需要。在高等学校,数据库也早已成为计算机专业大学生的必修课之一,非计算机专业的学生也要求学习和应用数据库来解决本专业的问题。本书就是为此而编写的。

本书是作者在多年从事数据库教学和科研的基础上编写出来的。书中较全面地讲述了数据库的基本概念、原理和技术,介绍了网状型,层次型,关系型三种数据库系统。特别是对现在应用很广的关系数据库系统作了重点介绍,较深入地讨论了关系数据库的原理和技术,并详细介绍了广为流行的三种关系数据库管理系统:SQL/DS、ORACLE、INGRES,力求从理论到实际系统使读者对数据库有一个较全面的了解。书中用少量篇幅介绍了网状和层次数据库。目前,这两种数据库的应用逐渐减少,但了解它们有助于理解数据库的基本概念和进一步了解关系数据库的特点。书中还介绍了数据库技术的一些发展新方向:分布式数据库和面向对象的数据库。分布式数据库是 80 年代发展起来的,它的出现和发展与网络技术有关。目前,分布式数据库技术已趋于成熟,应用也很广。面向对象的数据库是数据库领域中正在研究的新课题,还很不成熟,但它是数据库重要发展方向之一。

全书共分十章。第一章数据库系统概述,介绍数据库管理技术的发展、数据模型、数据库系统结构等一般概念;第二章网状数据库,第三章层次数据库,通过分别介绍 DBTG 和 IMS 系统,叙述网状、层次模型的基本概念、存储和存取方法、数据操纵语言等。第四、五、六章介绍关系数据库。第四章介绍关系模型的基本概念,关系代数和关系演算,关系系统的查询优化;第五章关系数据库系统,介绍三个广为流行的关系数据库系统 SQL/DS、ORACLE、INGRES,主要讨论关系数据库的标准语言 SQL 以及三个关系数据库系统的不同结构、特点;第六章介绍关系数据库设计理论;第七章数据库保护,内容包括数据的安全性、完整性、并发控制和数据恢复;第八章介绍数据库设计的基本技术和方法;第九章介绍分布式数据库的基本概念和特点;第十章简单介绍面向对象数据库的基本概念,旨在使读

者了解一些数据库的发展新方向。

本书在编写过程中注意取材合理,尽量反映国内外最新研究成果,力求做到理论联系实际,概念清晰,叙述简炼,通俗易懂,以便于自学。

北京理工大学的吴鹤龄教授详细审阅了书稿,提出了许多宝贵的意见和建议,并在本书的编写过程中给予了很多支持和帮助,同时也得到了江涛教授和陈朔鹰老师的大力支持和帮助,在此谨向他们表示衷心的感谢。

由于作者水平有限,书中错误和不足之处肯请读者批评指正。

作 者

1994年4月于北京理工大学

目 录

第一章 数据库系统概论

1.1 引言	(1)
1.1.1 数据管理技术的发展	(1)
1.1.2 什么是数据库	(3)
1.1.3 数据库系统的主要特征	(4)
1.2 数据模型	(5)
1.2.1 数据模型	(5)
1.2.2 实体间的联系	(6)
1.2.3 层次模型	(7)
1.2.4 网状模型	(8)
1.2.5 关系模型	(9)
1.3 数据库系统结构	(10)
1.3.1 数据库系统的三级模式结构	(10)
1.3.2 数据库管理系统的组成	(12)
1.3.3 数据库系统的工作过程	(14)
1.3.4 数据库系统的不同视图	(15)
习题	(16)

第二章 网状数据库

2.1 DBTG 系统概述	(17)
2.2 DBTG 中的几个重要概念	(17)
2.2.1 DBTG 中的系	(17)
2.2.2 数据库码	(21)
2.3 模式 DDL	(22)
2.3.1 模式 DDL	(22)
2.3.2 示例	(29)
2.4 子模式 DDL	(31)
2.5 数据操纵语言 DML	(33)
2.5.1 DML 与主语言的接口	(34)
2.5.2 程序运行环境	(34)
2.5.3 DML 命令简介	(36)
2.5.4 检索命令 FIND	(39)

2.5.5 程序实例	(41)
2.6 小结	(43)
习题	(43)

第三章 层次数据库

3.1 IMS 概述	(45)
3.1.1 IMS 的系统结构	(45)
3.1.2 IMS 中的层次结构	(46)
3.1.3 IMS 的存取方法	(47)
3.2 IMS 的数据模式	(48)
3.2.1 物理数据库(PDB)	(48)
3.2.2 逻辑数据库(LDB)	(50)
3.3 IMS 的子模式	(54)
3.4 IMS 的数据操纵语言	(55)
3.4.1 DL/1 与主语言的接口	(55)
3.4.2 DL/1 命令	(56)
3.5 IMS 的存储结构	(57)
3.6 小结	(61)
习题	(61)

第四章 关系数据库

4.1 关系模型的基本概念	(63)
4.1.1 关系的定义	(63)
4.1.2 键	(64)
4.1.3 关系模型	(65)
4.2 关系代数	(66)
4.2.1 传统的集合运算	(67)
4.2.2 专门的关系运算	(68)
4.2.3 举例	(70)
4.3 关系演算	(71)
4.3.1 元组关系演算	(72)
4.3.2 域关系演算	(73)
4.3.3 关系运算的安全限制和三种关系运算的等价性	(74)
4.4 几种关系查询语言	(77)
4.4.1 ISBL—“纯”关系代数语言	(77)
4.4.2 元组关系演算语言—ALPHA	(79)
4.4.3 域关系演算语言—QBE	(81)
4.5 查询优化	(84)
4.5.1 查询优化概述	(84)

4.5.2 查询优化的一般策略	(85)
4.5.3 关系代数表达式的优化算法	(85)
4.6 关系管理系统及其评价	(88)
习题	(91)

第五章 典型关系数据库系统

5.1 SQL/DS 的系统结构	(92)
5.1.1 SQL/DS 的数据结构	(92)
5.1.2 SQL/DS 的系统组成	(93)
5.1.3 SQL/DS 的操作环境和操作方式	(95)
5.1.4 SQL/DS 的物理存储结构	(97)
5.2 SQL 数据子语言	(98)
5.2.1 SQL 的数据定义	(99)
5.2.2 SQL 的数据操纵	(101)
5.2.3 视图	(107)
5.2.4 SQL 的授权控制	(110)
5.2.5 SQL/DS 的数据字典	(113)
5.3 嵌入式 SQL	(114)
5.3.1 SQL 与主语言	(115)
5.3.2 嵌入式 SQL	(116)
5.3.3 应用举例	(119)
5.3.4 SQL 的动态语句	(122)
5.4 SQL/DS 的交互接口—ISQL	(124)
5.5 小结	(127)
5.6 ORACLE 关系数据库系统	(128)
5.6.1 ORACLE 关系数据库系统组成及特点	(128)
5.6.2 ORACLE 的 SQL * PLUS	(131)
5.6.3 ORACLE 的应用开发工具 SQL * FORMS	(137)
5.6.4 报表生成器 SQL * Report	(140)
5.6.5 电子表格软件 SQL * Calc	(143)
5.6.6 图形生成工具 SQL * Graph	(146)
5.7 INGRES 关系数据库系统	(149)
5.7.1 INGRES 的组成及特点	(149)
5.7.2 INGRES 的 QUEL 语言	(151)
5.7.3 INGRES 中的查询优化方法	(155)
习题	(158)

第六章 关系数据库设计理论

6.1 关系模式的存储异常	(159)
---------------	-------

6.2	函数依赖	(160)
6.3	函数依赖的蕴涵性	(161)
6.4	函数依赖公理	(162)
6.4.1	Armstrong 公理	(162)
6.4.2	公理的完备性	(163)
6.4.3	闭包的计算	(164)
6.4.4	函数依赖集的等价和覆盖	(166)
6.5	关系模式的分解	(168)
6.5.1	无损连接分解	(169)
6.5.2	分解的依赖保持性	(171)
6.6	关系模式的规范化	(172)
6.6.1	第一范式	(172)
6.6.2	第二范式	(173)
6.6.3	第三范式	(173)
6.6.4	BCNF	(174)
6.7	多值依赖和第四范式	(175)
6.7.1	多值依赖	(175)
6.7.2	第四范式	(176)
6.8	连接依赖和第五范式	(177)
6.8.1	连接依赖	(177)
6.8.2	第五范式	(177)
	习题	(179)

第七章 数据库保护

7.1	安全性	(180)
7.1.1	用户标识与确认	(180)
7.1.2	存取控制	(181)
7.1.3	数据加密	(182)
7.1.4	几种系统采用的安全性措施	(182)
7.2	完整性	(185)
7.2.1	完整性规则	(185)
7.2.2	几个系统的完整性措施	(187)
7.3	并行性	(189)
7.3.1	共享系统中的数据一致性问题	(189)
7.3.2	封锁	(190)
7.3.3	可串行性	(192)
7.3.4	时标法	(194)
7.3.5	实际系统中的并发控制方法	(195)
7.4	数据库恢复	(197)

7.4.1 故障恢复的基本手段	(197)
7.4.2 故障恢复的方法	(198)
7.4.3 实际系统中的数据库恢复方法	(200)
7.4.4 计算机病毒	(201)
习题	(202)

第八章 数据库设计

8.1 数据库设计概述	(203)
8.1.1 数据库设计的内容和要求	(203)
8.1.2 数据库设计过程	(204)
8.2 需求分析	(206)
8.3 概念设计	(209)
8.3.1 E—R 模型	(209)
8.3.2 设计局部概念模式	(211)
8.3.3 设计全局概念模式	(213)
8.4 逻辑设计	(218)
8.4.1 导出数据模型	(219)
8.4.2 模型评价	(222)
8.4.3 模型优化	(224)
8.5 物理设计	(225)
8.5.1 数据库物理设计的内容	(225)
8.5.2 SQL/DS 的物理数据库设计	(227)
8.5.3 小结	(229)
8.6 数据库的建立和维护	(230)
8.6.1 建立数据库	(230)
8.6.2 数据库的重组织和重构造	(230)
8.7 数据库设计的辅助工具	(232)
8.8 小结	(234)
习题	(235)

第九章 分布式数据库系统

9.1 分布式数据库系统概述	(236)
9.1.1 分布式数据库系统的产生与发展	(236)
9.1.2 分布式数据库系统的定义	(237)
9.1.3 分布式数据库与集中式数据库的比较	(238)
9.1.4 分布式数据库管理系统	(239)
9.1.5 分布式数据库系统的优点	(240)
9.2 分布式数据库系统的体系结构	(241)
9.2.1 分布式数据库系统的参考体系结构	(241)

9.2.2 数据分段	(242)
9.2.3 分布透明性	(244)
9.3 分布式查询处理	(246)
9.4 更新传播	(248)
9.5 并发控制	(249)
9.5.1 基于封锁的并发控制	(249)
9.5.2 基于时标的并发控制	(251)
9.6 分布式事务处理的恢复	(252)
9.7 ORACLE 分布式数据库系统	(253)
习题	(256)

第十章 面向对象数据库

10.1 面向对象数据库的兴起	(257)
10.2 面向对象的程序设计语言	(259)
10.3 面向对象数据库系统	(260)
10.3.1 面向对象数据库的核心概念	(260)
10.3.2 面向对象数据库系统的特点	(265)
10.3.3 面向对象的数据库管理系统	(265)
10.4 面向对象数据库的主要研究内容	(268)
10.5 与面向对象数据库系统的相关领域	(271)

第一章 数据库系统概述

数据库技术是计算机科学技术的一个重要分支。自1968年第一个商品化的数据库管理系统IMS问世以来,数据库技术得到了迅速发展。随着计算机应用的不断深入,数据库的重要性日益被人们所认识,它已成为信息管理、办公自动化、计算机辅助设计等方面的主要手段。学习和应用数据库已成为广大科技人员和管理人员的共同需要。

这一章,我们将介绍数据库系统以及所涉及的一些基本概念,包括数据模型、数据库的系统结构、数据库系统组成和功能,使大家对数据库有一个概括的了解。下面,我们先回顾一下数据处理的发展过程。

1.1 引言

1.1.1 数据管理技术的发展

电子计算机诞生于1946年,在一段时间里,计算机主要用于科学和工程技术领域的数值计算。但随着社会生产力和文明的不断发展,信息在人类社会的活动中起着越来越重要的作用,人们经常需要收集、加工和处理大量的信息数据。50年代初期,人们开始用计算机进行数据处理。30多年来,数据管理技术随着计算机软、硬件的发展在不断的不断发展,它大致经历了如下三个阶段:

- (1)人工管理阶段(50年代中期以前)
- (2)文件系统阶段(50年代后期到60年代中期)
- (3)数据库系统阶段(从60年代后期开始)

1. 人工管理阶段

这一阶段,计算机除硬设备外没有任何软件可用。在应用程序中除了要规定数据的逻辑结构外,还要考虑数据在计算机中如何存储和组织,并为数据分配空间、决定存取方法。应用程序完全依赖于数据。数据结构的改变,存储与存取方法的变化,都会使应用程序跟着改变。这一阶段,应用程序与数据是一一对应的,如图1-1所示。如果几个应用要用到同一数据,这些数据也要重复存储,数据的冗余度很大。这一阶段,数据的管理还是手工性质的、分散的,处理效率很低。

2. 文件系统阶段

50年代末,随着数据处理的日渐增多,出现了专用于数据处理的软件——文件管理系统。文件的应用,是计算机数据处理的重大进展。数据文件可以按名引用,应用程序通过文件管理系统与数据文件发生联系(如图1-2所示),数据的物理结构与逻辑结构间有了简单变换。这样,在应用程序中,可以不必过多地考虑数据的物理存储细节,简化了程序员的数据管理工作。同时,一个应用程序可以和几个数据文件发生联系,增加了数据处理的灵活性。但是,数据仍然是分散的,是面向应用的,用文件系统处理数据存在着以下几个问题。

数据冗余度大。由于数据是面向应用程序的,一个数据文件只为某一用户的特定用途

服务,其它应用所需要的相同数据只能重复存储,造成了大量的数据冗余。

例如,某单位的人事、教学、科研部门,都需要处理职工的有关信息。

人事处要处理的信息包括:部门、职工号、姓名、性别、年龄、籍贯、文化程度、职务、工资、社会关系、个人简历。

教务处要处理的信息包括:部门、职工号、姓名、性别、年龄、职务、学历、专业、教学历史。

科研处要处理的信息包括:部门、职工号、姓名、性别、年龄、职务、学历、科研历史。

以上三个应用需要将各自所处理的数据分别进行存储。这样,相同的数据,如部门、职工号、姓名、性别等要重复存储多次。数据冗余不仅浪费了存储空间,降低了存储器的利用率,而且带来潜在的不一致性。

在文件系统中,数据的更新是每个应用程序各自进行的,相同数据间没有一致的保护措施。如某人的职务发生了变化,人事处及时作了更新,但教务处和科研处可能还没有修改,发生了数据的不一致性。由于数据的使用价值在很大程度上依赖于其可靠性,这种不一致性是不可忽视的。

在文件系统中,应用程序与数据结构互相依赖。数据文件是按具体应用要求建立的,而应用程序的编制直接依赖于数据的存储格式、存取方法。一旦由于存储介质发生变化或数据结构改变时,相应的应用程序也得进行修改,增加了程序的编制和维护的工作量。这种应用程序与数据的存储、存取方式密切相关的情况,称为“数据依赖”。

另外,文件系统缺乏对数据进行统一的控制方法,数据的完整性、安全性等无法得到保证,完全要应用程序自己管理,增加了程序的复杂性。

文件系统的这些缺点,促使人们开始研究新的数据管理技术。

3. 数据库系统阶段

文件系统处理数据存在着诸多不足,而计算机用于管理,规模越来越大,数据量急剧增长。这一时期,磁盘技术取得了重要进展,大容量和快速存取的磁盘开始进入市场,给数据库系统的研究提供了良好的物质基础。

1968年美国IBM公司研制了世界上第一个数据库管理系统IMS(Information Management System);1969年美国CODASYL委员会(Conference On Data System Language)的DBTG(Data Base Task Group)小组公布了它的研究成果DBTG报告;1970年IBM公

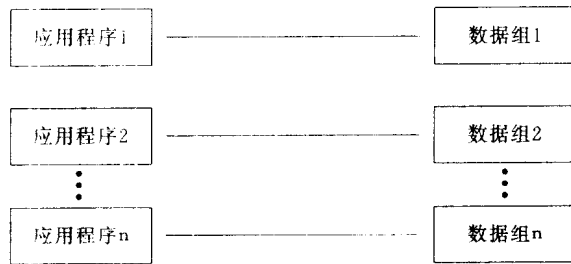


图1 - 1 手工管理阶段数据与程序间的关系

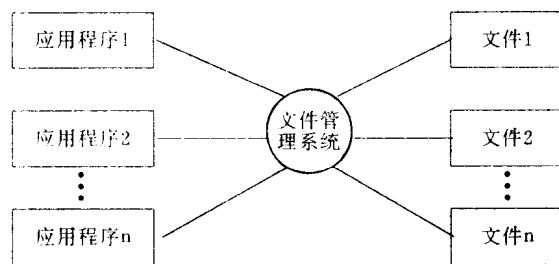


图1 - 2 文件系统阶段数据程序间的关系

司的研究员 E. F. Codd 发表了题为“大型共享数据库数据的关系模型”等一系列关系数据库论文。这三大事件标志着数据处理已进入了数据库技术的新时代。

数据库技术解决了冗余和数据依赖问题,实现了数据共享,数据库中的数据以一定方式组织起来,由一个软件共同管理,作为应用程序与数据的接口。这一阶段数据与程序的关系如图 1-3 所示。

70 年代,数据库技术得到了迅速发展。在这期间,世界上研制出了许多数据库系统,较著名的有 IMS、IDMS、ADABAS、DMS1100、TOTAL、SYSTEM R 等。70 年代末,数据库技术逐渐成熟。进入 80 年代,关系数据库由于其结构简单、操作方便、并

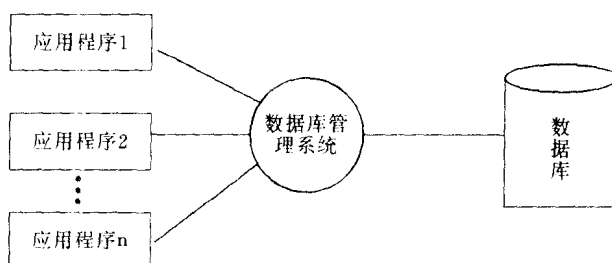


图1 - 3 数据库阶段数据与程序的关系

有扎实的理论基础,而得到长足的发展,有关关系数据库的研究和产品日益增多,关系数据库逐渐成为数据库发展的主流。同时,各种微机关系数据库系统不断出现,并得到了广泛应用。数据库技术与计算机网络相结合,导致了一个崭新的领域——分布式数据库,它是八十年代数据库技术的重大发展。分布式数据库系统的出现,使不同地点、不同用户共享数据库资源成为可能。短短几年里,已有不少分布式数据库系统面世,如 SDD-1、INGRES、POREL、R* 等,它是数据库的发展方向之一。

近年来,随着计算机科学技术的发展,数据库技术与人工智能、软件工程、自然语言理解、CAD、CAM 等相结合,不断扩展形成新的方向,如对知识库、面向对象的数据库的研究等。数据库技术已成为计算机领域里最活跃的一个分支,向人们展示着广阔的前景。

1.1.2 什么是数据库

文件系统中数据是分散的,应用程序对应着各自的数据文件,而在数据库系统中,数据被集中进行管理,就象货物仓库中的物资一样,用户需要什么数据就去库中提取。因此,有人形象地把这样的系统称为“数据库”(Data Base)。

那么,什么是数据库?数据库是一个复杂的系统,给它下一个确切的定义是困难的,目前还没有一个公认的、统一的定义。

但具体地,对一个特定数据库来说,它是集中、统一地保存、管理着某一单位或某一领域内所有有用信息的系统,这个系统根据数据间的自然联系结构而成,数据较少冗余,且具有较高的数据独立性,能为多种应用服务。

从以上叙述我们可以看出,数据库系统是管理数据且为不同应用服务的工具。它所管理的数据是大量的有关某一方面的信息,需要较长时间的保存;它所提供的数据应该是正确的、可靠的、高效率的。

这样一个数据库系统至少应该包含以下三个部分:

(1)数据库。一个结构化的相关数据的集合,包括数据本身和数据间的联系。它独立于应用程序而存在,是数据库系统的核心和管理对象。

(2)物理存储的。这是保存数据的硬件介质,一般为磁盘、磁带等大容量的存储器。

(3)数据库软件。负责对数据库管理和维护的软件。具有对数据定义、描述、操作、维护等功能,接受并完成用户程序及终端命令对数据库的不同请求,并负责保护数据免受各种干扰和破坏。数据库软件的核心是数据库管理系统(Data Base Management System—DBMS)。

另外,数据库系统需要有专门的人员负责数据库的设计、建立、执行和维护,这些人员被称为数据库管理员(Data Base Administrator—DBA)。

一个基本数据库系统组成及其间的关系如图 1-4 所示。

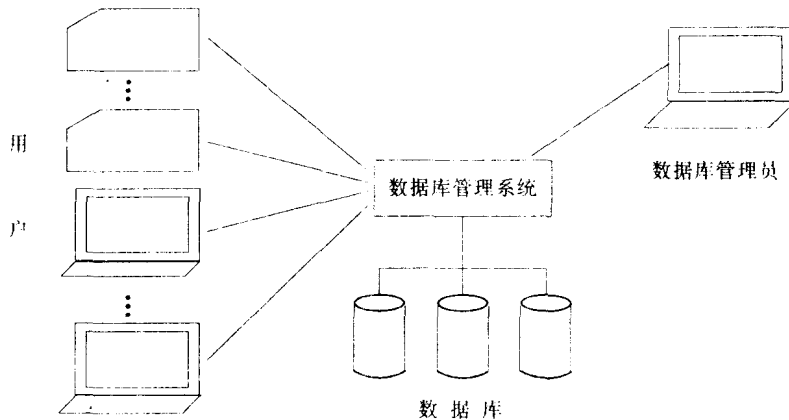


图1 - 4 数据库系统

数据库系统的运行需要有一定的硬件支持,除大容量的磁盘直接存储以外,对内存有一定的要求。在数据库系统运行期间,内存中除操作系统外还有数据库管理系统和一定数量的缓冲区。数据库管理系统一般为几百 K 字节,通常有一部分是常驻内存的。此外,与数值计算不同,数据处理需要大量的缓冲区,以减少内外存交换次数,提高数据存取效率。早期的数据库管理系统 IMS 占内存 120K 字节,缓冲区占 50K 字节;IDMS 占内存 50K 字节,缓冲区占 20K 字节;现在微机上使用较多的 FoxBASE 中,正常运行至少要 360K 的可用内存;而 ORACLE 除需要 640K 字节内存外,还需要约 900K 的扩展内存空间。

数据库系统的运行速度,除了与 CPU 本身的速度有关外,主要还与输入/出操作所占时间有关。

1.1.3 数据库系统的主要特征

数据库系统管理数据,较文件系统有许多不同。它的主要特征如下:

(1)数据共享。实现数据共享是数据库发展的一个主要原因,也是数据库系统的一个重要特征。数据库中的数据可以供多个用户所使用,每个用户只与库中的一部分数据打交道;用户数据可以是重叠的,在同一时刻不同的用户可以同时存取数据而互不影响,大大地提高了数据的利用率。数据共享还表现在对存储的数据开展一些新应用,即可为将要使用数据库的新用户服务。

(2)数据独立性。在文件系统中,应用程序不但与数据文件相互对应,而且与数据的存储和存取方式密切相关。在数据库系统中,应用程序不再同物理存储器上具体文件相对应,每个用户所使用的数据有其自身的逻辑结构。在数据库方式下,数据独立性表现在二

个方面:物理独立性和逻辑独立性。

1)物理独立性。指数据库物理结构的改变,如数据的存储格式和组织方法的改变,外存设备的更换等,不影响数据的逻辑结构,因而不会使应用程序随物理结构的改变而进行修改,数据的物理存储是独立于应用程序的。

2)逻辑独立性。指数据库中的逻辑数据结构因增加新应用或某些应用发生变化而需要对数据重新定义,或数据间的联系发生改变时,其它应用程序无需修改,即其它用户感觉不出数据逻辑结构的更改。

数据独立性是通过数据库系统所提供的二级映象来实现的。一级是数据的存储结构与逻辑结构之间的映象关系,二级是数据的全局逻辑结构到应用所涉及的局部逻辑结构之间的映象。

数据独立性给数据库的使用、调整、优化和进一步扩充带来了方便,提高了数据库应用系统的稳定性,减轻了程序员的负担。

(3)减少数据冗余。在数据库系统管理下的数据,不再是面向应用,而是面向系统。数据集中管理,统一进行组织、定义和存储,避免了不必要的冗余,因而也避免了数据的不一致性。

在实际系统中,为提高数据的存取效率,少量数据的冗余还是存在的,但这些数据的冗余将受到控制,系统负责对冗余数据的检查、维护工作。

(4)数据的结构化。文件系统中的数据从整体上看是无结构的,数据文件之间不存在任何联系。而数据库系统中的数据是相互关联的,这种联系不仅表现在记录内部,更重要的是记录类型之间的相互联系。整个数据库是以一定的形式结构而成的。用户可以通过不同的路径存取数据,以满足用户的不同需要。

(5)统一的数据保护功能。多个用户共享数据资源,需要解决数据的安全性、一致性及并发控制问题。为使数据安全、可靠,系统对用户使用数据有严格检查,对非法用户将拒绝进入数据库。专用数据可规定密码或存取权限,有权用户才能对数据进行操作。

不同用户同时使用数据库,可能会相互干扰而出现数据的不一致,数据库系统具有并发控制功能,以保证数据的正确性。

此外,系统还提供其它的数据保护措施,如数据的有效性检查、故障恢复等来保证数据的正确性。

1.2 数据模型

为了了解世界、研究世界,相互交流情况,人们需要描述各种事物。但客观存在的事物是复杂的,因此,人们把表示事物的主要特征抽象地用一种形式化的描述反映出来,以简化问题,便于处理,模型方法就是这种抽象的一种表示。如为了研究地球概貌,人们设计出了地球仪,地球仪上表示了山川、河流、湖泊、海洋、国家、城市等的地理分布位置,而与之无关的其它特征则被略去。

信息领域中采用的模型通常称为数据模型。这一节我们将讨论基本数据模型的三个组成部分,实体间的联系和数据库中常用的三种数据模型。

1.2.1 数据模型

在数据库中,数据模型是现实世界中的各种事物及其间的联系用数据及数据间的联

系来表示的一种方法。目前使用的数据模型大体可以分为两种类型：一种是独立于计算机之外的，如实体—联系模型，语义数据模型等，这类模型不涉及信息在计算机中如何表示，而是用来描述某一特定范围内人们所关心的信息结构，这种模型常称为概念模型或信息模型；另一类模型是直接面向计算机的，它们以记录为单位构造数据模型，数据库中常用的层次、网状和关系模型，就属于这类模型，我们称它为基本数据模型。

数据模型是数据特征的抽象，它不是描述个别数据，而是描述数据的共性。一般来说，一个数据库的数据模型包含以下三个组成部分：

- (1) 一组目标型的集合；
- (2) 一组算子；
- (3) 一组通用的完整性规则。

目标型是构造数据库的基本数据结构类型，它包括应用所涉及的对象和对象具有的特征，对象间的联系，它是对数据静态特性的描述。如网状模型中的数据项、记录、系型，关系模型中的域、关系等。

算子是一组对目标型的任何实例执行的数据操作。通常对数据库的操作有检索、插入、删除、修改，这些操作反映了数据的动态特性，因为现实世界中的实体及实体间的联系是在不断变化的，数据模型应能反映出这种变化。

完整性规则是对数据静态和动态特性的限定，它定义相容的数据库状态的集合及可允许的状态改变。如在关系模型中，规定任一记录都必须有一个确定的关键字值来标识；又如在进行数据操作时，不能破坏数据间存在着的联系，因此要规定插入、删除、修改规则等。现实世界中的实体是按一定方式相互制约、相互依存的，完整性规则应反映数据模型中数据间的这种制约和依存关系。

从以上叙述可知，一个基本数据模型实际上给出了在计算机系统上进行描述和动态模拟现实世界信息结构及其变化的一种抽象方法。数据模型不同，描述和实现方法亦不相同，相应的支持软件——数据库管理系统也就不同。

一个数据模型，严格地讲应该由上述三部分组成，但在讨论数据模型时，通常只涉及构造数据库的基本数据结构，即一组目标型的集合，他表示了应用所涉及的对象和对象间的联系，这是区别数据模型最本质的部分。下面在讨论具体数据模型时也将忽略算子和完整性规则，这二部分在以后的章节再讨论。

1.2.2 实体间的联系

数据模型是对现实世界的抽象描述。在组织数据模型时，人们首先将现实世界中存在的客观事物用某种信息结构表示出来，然后再转化为用计算机能表示的数据形式。

所谓信息，是客观世界中存在的事物在人们头脑中的反映，人们把这种反映用文字、图形等形式记录下来，经过命名、整理、分类就形成了信息。

在信息领域中，用名词“实体”表示客观存在的事物，如一个人、一本书、一次活动，都是实体；属性表示实体所具有的特征，如书名、作者、出版单位等都是图书实体具有的属性，一个实体是由它的若干属性来表示的；具有相同特征的同类实体，形成实体的类型，简称实体型。可用实体名及若干属性名的集合来描述；同类实体的集合称为实体集。

数据模型反映了现实世界中事物间的各种联系，即实体间的联系。究竟有哪些联系呢？联系有二种：一种是实体内部的联系，即实体中属性间的联系；一种是实体与实体之间