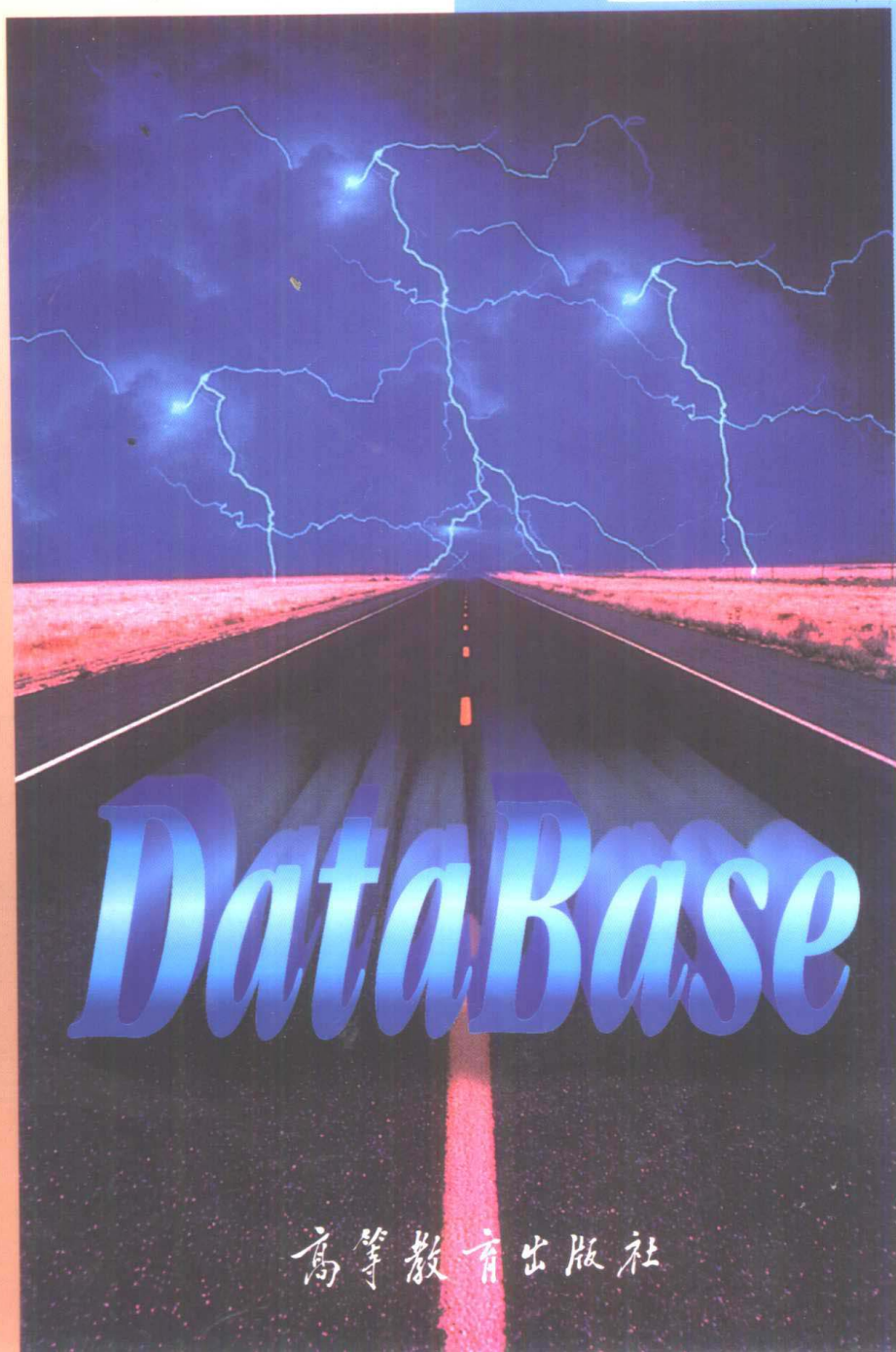


高等学校教材

数据库系统 及应用

崔巍 编著



1.13
15

高等教育出版社

1195645

高等学校教材

数据库系统及应用

崔巍 编著

高等教育出版社

内 容 提 要

本书全面介绍了数据库系统的原理、方法和发展趋势。本书首先介绍数据库系统的基础,内容包括数据库的基本概念、信息结构、数据库的体系结构、关系数据库及其理论等;接着介绍了 Client/Server 体系结构数据库的概念、管理和使用,内容包括 Client/Server 数据库的基础、如何管理和使用数据库服务器、如何进行客户端的开发、数据库保护以及分布式数据库等内容;最后介绍了数据库设计和数据库的新领域等,内容包括数据库的规划和设计、数据仓库、因特网上的 Web 数据库和面向对象数据库等。

本书以全新的取材奉献给读者,既可以作为高等院校计算机和信息管理等有关专业的数据库课程教材,也可供从事计算机软件以及数据库管理、使用和开发的科技人员、工程技术人员以及其他有关人员阅读参考。

图书在版编目(CIP)数据

数据库系统及应用/崔巍编著. —北京:高等教育出版社,1999(2000 重印)

高等学校教材

ISBN 7-04-006950-4

I. 数… II. 崔… III. 数据库系统-高等学校-教材 IV. TP311.13

中国版本图书馆 CIP 数据核字(1999)第 10899 号

数据库系统及应用

崔巍 编著

出版发行 高等教育出版社

社 址 北京市东城区沙滩后街 55 号

邮政编码 100009

电 话 010-64054588

传 真 010-64014048

网 址 <http://www.hep.edu.cn>

经 销 新华书店北京发行所

印 刷 北京地质印刷厂

开 本 787×1092 1/16

版 次 1999 年 6 月第 1 版

印 张 19.5

印 次 2000 年 4 月第 2 次印刷

字 数 470 000

定 价 15.80 元

凡购买高等教育出版社图书,如有缺页、倒页、脱页等质量问题,请在所购图书销售部门联系调换。

版权所有 侵权必究

前 言

数据库技术始于 20 世纪 60 年代,经历了最初的基于文件的初级系统、20 世纪 60~70 年代流行的层次系统和网状系统,而现在广泛使用的是关系数据库系统。数据库应用也从简单的事务管理扩展到各个应用领域,如用于工程设计的工程数据库、用于因特网的 Web 数据库、用于决策支持的数据仓库技术、用于多媒体技术的多媒体数据库等,但应用最广泛的还是在基于事务管理的各类信息系统领域。数据库的体系结构也从最初的集中式数据库变化为基于客户/服务器机制的分布式数据库。随着面向对象技术的发展,关系对象数据库系统正在逐步完善和投入使用。而随着时代的进步和发展,数据库的应用领域会越来越广泛,数据库技术也将是所有信息技术和信息产业的基础。

本书分为四个部分。第一部分是数据库基础;第二部分利用 SQL Server 介绍基于客户/服务器机制的数据库管理系统、数据库保护技术和分布式数据库;第三部分介绍数据库系统的开发、数据库设计等;第四部分概要介绍数据库技术的最新研究领域和应用领域及其发展。

第一部分分为四章(第一章至第四章)。第一章主要介绍什么是数据库及与之相关的一些基本概念;第二章介绍数据模型和三层模式数据库结构,包括实体-联系方法、概念数据模型、传统的三大数据模型、数据库的三层模式和数据库管理系统等内容;第三章介绍关系数据库、关系代数以及关系数据库的标准数据语言 SQL,并用大量的实例介绍了 SQL 的使用方法;第四章介绍关系数据库理论,包括函数依赖、公理系统、规范化与模式分解等内容。

第二部分分为五章(第五章至第九章)。在这一部分我们选择 Microsoft SQL Server 作为数据库管理系统的实例来介绍数据库管理系统的功能、管理和使用方法,其目的不是为了教会读者如何使用 SQL Server,也不是为了单纯介绍 SQL Server,而是通过 SQL Server 了解客户/服务器机制的数据库体系结构,了解一个完备的数据库管理系统应该具备的功能,了解应该如何管理和使用数据库管理系统。通过这一部分的学习,可以为日后管理和使用这类数据库管理系统打下一个良好的基础。第五章介绍客户/服务器数据库系统的基础知识,包括客户/服务器数据库的基本概念、SQL Server 的基本概念以及 SQL Server 的数据语言 Transact-SQL 等;第六章介绍如何充分利用数据库服务器的功能,包括 SQL Server 数据库的管理、数据完整性、存储过程及触发器的应用等;第七章介绍数据库保护技术,包括并发控制、安全控制、转储和恢复及镜像等;第八章介绍客户端的开发方法,包括 ODBC 的概念, Visual FoxPro、PowerBuilder 和 Visual Basic 等一些常用的客户端开发工具,并介绍如何利用这些工具进行开发等;第九章介绍分布式数据库和 SQL Server 的数据复制等内容。

第三部分即第十章,介绍一些常用数据库系统的开发方法、数据库设计以及规划客户/服务器数据库应用系统的方法和要考虑的一些问题。

第四部分即第十一章,介绍数据库的一些最新研究领域和应用领域及其发展,包括数据仓库技术、因特网上的 Web 数据库、面向对象数据库、主动数据库、并行数据库、工程数据库和知识库等。

在教材的编写过程中，笔者尽可能引入新的技术和方法，力求能反映当前的技术水平和未来的发展方向。但由于水平有限，定有许多不足之处，望同行和专家批评指正。

编 者

1999 年 2 月

目 录

第一章 绪论1	
1.1 什么是数据库.....1	
1.1.1 相互关联的数据的集合.....1	
1.1.2 用综合的方法组织数据.....1	
1.1.3 低冗余与数据共享.....1	
1.1.4 数据具有较高的独立性.....2	
1.1.5 保证数据的安全、可靠.....2	
1.1.6 最大限度地保证数据的正确性.....2	
1.1.7 数据可以并发使用并能同时 保证数据的一致性.....2	
1.2 数据库管理系统.....3	
1.3 数据库系统和数据库管理员.....3	
1.3.1 数据库系统.....3	
1.3.2 数据库管理和数据库管理员.....4	
1.4 数据库的过去、现在和未来.....5	
1.4.1 人工管理阶段.....5	
1.4.2 文件系统阶段.....5	
1.4.3 数据库系统阶段及其发展过程.....7	
1.4.4 数据库技术的广泛应用领域.....8	
习题.....9	
第二章 数据模型和三层模式数据库10	
2.1 信息结构与 E-R 方法.....10	
2.1.1 数据的三种范畴.....10	
2.1.2 实体-联系方法.....11	
2.2 概念数据模型.....14	
2.2.1 概念数据模型中的基本联系.....14	
2.2.2 理解和转换多对多联系.....17	
2.2.3 扩展 Bachman 图.....18	
2.2.4 连接陷阱.....19	
2.2.5 概念数据模型——实例研究.....20	
2.3 传统的三大数据模型.....23	
2.3.1 层次数据模型.....23	
2.3.2 网络数据模型.....25	
2.3.3 关系数据模型.....28	
2.4 数据独立性与三层结构.....29	
2.4.1 存储数据独立性.....29	
2.4.2 概念数据独立性.....31	
2.5 数据库管理系统的结构.....32	
2.5.1 数据库的三层模式结构.....32	
2.5.2 模式说明实例.....33	
2.5.3 数据库管理系统的总体结构.....37	
习题.....39	
第三章 关系数据库及其操作语言40	
3.1 关系模型的基本概念.....40	
3.1.1 笛卡儿积与关系定义.....40	
3.1.2 关系的一些基本概念.....41	
3.2 关系模式.....42	
3.2.1 关系概念模式.....42	
3.2.2 关系存储模式.....43	
3.2.3 关系外部模式.....44	
3.3 关系代数.....44	
3.3.1 传统的集合运算.....44	
3.3.2 专门的关系运算.....45	
3.4 关系数据库标准数据语言——SQL.....50	
3.4.1 SQL 的数据定义功能.....52	
3.4.2 SQL 的数据查询功能.....56	
3.4.3 视图.....74	
3.4.4 SQL 的数据操作功能.....77	
3.4.5 SQL 的数据控制功能.....79	
3.4.6 SQL 的宿主使用.....81	
习题.....85	
第四章 关系数据理论87	
4.1 基本概念.....87	
4.1.1 函数依赖.....87	
4.1.2 术语和符号.....88	
4.1.3 为什么要讨论函数依赖.....88	
4.1.4 模式分解.....89	
4.2 函数依赖的公理系统.....89	
4.2.1 Armstrong 公理的内容及正确性.....89	
4.2.2 Armstrong 公理的推论.....90	

4.2.3 逻辑蕴涵和闭包.....	91	6.1 SQL Server 的数据库管理.....	150
4.2.4 公理的完备性.....	92	6.1.1 数据库设备的管理.....	150
4.2.5 闭包的计算.....	93	6.1.2 数据库管理.....	154
4.2.6 函数依赖集的等价和最小化.....	94	6.1.3 使用段优化存储.....	159
4.3 规范化.....	97	6.2 数据完整性.....	162
4.3.1 第一范式(1NF).....	97	6.2.1 实体完整性.....	162
4.3.2 第二范式(2NF).....	98	6.2.2 域完整性.....	163
4.3.3 第三范式(3NF).....	98	6.2.3 参照完整性.....	164
4.3.4 BC 范式(BCNF).....	99	6.2.4 用户定义完整性.....	165
4.3.5 多值依赖与第四范式(4NF).....	101	6.3 存储过程.....	165
4.3.6 规范化小结.....	102	6.3.1 基本概念.....	165
4.4 模式分解.....	103	6.3.2 创建和执行存储过程.....	166
4.4.1 模式分解的准则.....	103	6.3.3 存储过程的返回值和状态信息.....	167
4.4.2 3NF无损连接和保持函数依赖算法.....	105	6.3.4 要用好存储过程.....	168
4.4.3 使分解后的关系模式数最少.....	106	6.4 触发器及其用途.....	168
习题.....	107	6.4.1 基本概念.....	168
第五章 客户/服务器数据库基础.....	109	6.4.2 插入视图和删除视图.....	169
5.1 客户/服务器体系结构.....	109	6.4.3 删除类触发器.....	170
5.1.1 客户/服务器和其他应用结构.....	109	6.4.4 插入类触发器.....	170
5.1.2 文件服务器与客户/服务器的 数据库操作.....	112	6.4.5 更新类触发器.....	170
5.1.3 客户/服务器结构的特点.....	113	习题.....	171
5.1.4 三层客户/服务器模型.....	114	第七章 数据库保护.....	173
5.2 Microsoft SQL Server 基础.....	116	7.1 数据一致性和并发控制.....	173
5.2.1 SQL Server 功能概述.....	116	7.1.1 事务.....	173
5.2.2 SQL Server 的存储结构.....	120	7.1.2 干扰问题.....	174
5.2.3 SQL Server 安装说明.....	122	7.1.3 封锁.....	175
5.3 Transact-SQL 简介.....	123	7.1.4 死锁.....	177
5.3.1 数据库管理的命令.....	123	7.2 安全性控制.....	179
5.3.2 数据库对象管理命令.....	126	7.2.1 安全性概述.....	179
5.3.3 变量、数据类型与表达式.....	128	7.2.2 SQL Server 的安全模式.....	180
5.3.4 数据操作命令.....	134	7.2.3 创建用户和用户组.....	181
5.3.5 数据管理和数据控制命令.....	139	7.2.4 权限管理.....	182
5.3.6 函数.....	142	7.3 备份与恢复.....	186
5.3.7 流程控制和输入输出命令.....	146	7.3.1 故障与备份类型.....	186
5.3.8 系统提供的存储过程.....	147	7.3.2 日志的概念.....	188
5.3.9 数据字典.....	147	7.3.3 备份或转储.....	188
习题.....	149	7.3.4 恢复.....	194
第六章 充分利用数据库服务器的特点.....	150	7.4 镜像.....	196
		7.4.1 SQL Server 的镜像配置选择.....	196

7.4.2 建立镜像.....	196	9.5.1 用于配置复制服务器的存储过程.....	257
7.4.3 取消镜像.....	198	9.5.2 用于出版管理的存储过程.....	257
7.4.4 重镜像.....	198	9.5.3 用于订阅管理的存储过程.....	258
习题.....	199	9.5.4 用于复制操作的存储过程.....	258
第八章 开放的客户端开发.....	200	9.5.5 用于复制事务管理的存储过程.....	258
8.1 开放式客户体系结构与 ODBC.....	200	9.6 建立复制.....	258
8.1.1 ODBC 的基本概念.....	200	9.6.1 复制前的准备工作.....	259
8.1.2 ODBC 的构成.....	201	9.6.2 配置复制服务器.....	259
8.1.3 ODBC 接口的使用.....	202	9.6.3 出版.....	260
8.2 Visual FoxPro 作为 SQL Server 的客户端.....	207	9.6.4 订阅.....	260
8.2.1 连接到 SQL Server.....	207	习题.....	260
8.2.2 对 SQL Server 进行操作.....	209	第十章 数据库系统的开发	
8.2.3 Visual FoxPro 中其他 ODBC 函数.....	210	与数据库设计.....	262
8.3 PowerBuilder 介绍.....	215	10.1 概述.....	262
8.3.1 PowerBuilder 的数据库接口.....	215	10.2 结构化生命周期方法.....	262
8.3.2 PowerBuilder 应用的构成与		10.3 快速原型方法.....	264
开发方法.....	215	10.4 DADM 方法.....	265
8.3.3 对象、属性与事件.....	217	10.5 面向对象方法.....	267
8.3.4 PowerBuilder 的工具.....	218	10.5.1 面向对象的基本概念.....	267
8.3.5 与数据库的连接.....	220	10.5.2 面向对象的分析和设计方法.....	269
8.3.6 一个简单实例.....	222	10.6 可能会失败.....	270
8.4 用 Visual Basic 开发客户/服务器应用.....	233	10.6.1 认识问题.....	270
8.4.1 远程数据对象.....	233	10.6.2 方法问题.....	270
8.4.2 RDO 与 SQL Server 的连接.....	235	10.6.3 工具问题.....	270
8.4.3 创建和执行 SQL 查询.....	237	10.6.4 管理问题.....	271
8.4.4 插入、更新和删除记录.....	241	10.7 数据库设计.....	271
习题.....	246	10.7.1 概念模型设计.....	271
第九章 分布式数据库与分布数据处理.....	247	10.7.2 逻辑数据库设计与规范化	
9.1 什么是分布式数据库.....	247	理论的应用.....	273
9.2 分布式数据库的数据分布方式.....	249	10.7.3 物理数据库设计.....	274
9.2.1 数据的分布方式.....	249	10.8 客户/服务器应用规划综述.....	275
9.2.2 数据分片.....	250	10.8.1 可以量化的需求分析.....	275
9.3 SQL Server 的分布数据管理概述.....	250	10.8.2 性能需求.....	277
9.4 SQL Server 的复制概念.....	251	10.8.3 并发需求.....	278
9.4.1 SQL Server 的复制.....	251	10.8.4 数据分布需求.....	278
9.4.2 复制技术中使用的术语和基本概念.....	252	10.8.5 恢复需求.....	278
9.4.3 SQL Server 的复制模型.....	253	10.8.6 安全问题.....	279
9.4.4 复制处理.....	255	10.8.7 系统需求.....	279
9.5 复制存储过程.....	257	习题.....	280

第十一章 数据库应用和研究的

新领域	281
11.1 数据仓库.....	281
11.1.1 什么是数据仓库.....	281
11.1.2 数据仓库的结构.....	282
11.1.3 数据仓库系统.....	283
11.1.4 建立数据仓库.....	284
11.1.5 实现数据仓库的数据库环境.....	284
11.1.6 数据仓库与决策支持.....	285
11.1.7 软件厂商的数据仓库解决方案.....	285
11.2 因特网上的 Web 数据库.....	286
11.2.1 因特网及其所提供的服务.....	287
11.2.2 Web 与 HTML.....	288

11.2.3 Web 数据库应用的一般 结构和方法.....	290
11.2.4 各种 Web 数据库应用方案概述.....	291
11.2.5 Sybase 的 Web-SQL 方案.....	292
11.3 数据库的其他应用和研究领域.....	295
11.3.1 面向对象数据库.....	295
11.3.2 并行数据库.....	297
11.3.3 主动数据库.....	297
11.3.4 工程数据库.....	298
11.3.5 知识库.....	299
习题.....	301
参考文献	302

第一章 绪 论

数据库是数据管理的工具。数据管理经历了从手工管理阶段、文件管理阶段到数据库管理阶段的变迁。数据库技术自产生以来便为广大用户所接受，并获得了广泛的应用。无论是对学生，还是对广大计算机用户，数据库都已不再陌生。

1.1 什么是数据库

数据库，顾名思义就是存放数据的仓库，这种想当然的理解是不准确的。数据库对应的英文单词是 DataBase，如果直译则是数据基地；而数据仓库则另有其词——DataWarehouse。所以数据库和数据仓库不是同义词，数据仓库是在数据库技术的基础上发展起来的又一新的应用领域（数据仓库将在第十一章介绍）。

数据库技术发展到今天已经是一门成熟的技术，但却没有一个被普遍接受的、严格的定义。数据库是相互关联的数据的集合，这是大家公认的数据库的基本特征之一。下面一段话概括了数据库应该具备的一些特征，也可以把它作为数据库的定义：

数据库是相互关联的数据的集合，它用综合的方法组织数据，具有较小的数据冗余，可供多个用户共享，具有较高的数据独立性，具有安全控制机制，能够保证数据的安全、可靠，允许并发地使用数据库，能有效、及时地处理数据，并能保证数据的一致性和完整性。

1.1.1 相互关联的数据的集合

数据库中的数据不是孤立的，数据与数据之间是相互关联的。也就是说，在数据库中不仅要能够表示数据本身，还要能够表示数据与数据之间的联系。

比如在学籍管理中，有学生和课程两类数据，在数据库中除了要存放这两类数据之外，还要存放哪些学生选修了哪些课程或哪些课程由哪些学生选修这样的信息，这就反映了学生数据和课程数据之间的联系。

1.1.2 用综合的方法组织数据

数据库能够根据不同的需要按不同的方法组织数据，比如可以有顺序组织方法、索引组织方法，现在很多数据库还支持聚集（Cluster）组织方法等。

1.1.3 低冗余与数据共享

由于在数据库技术之前，数据文件都是独立的，所以任何数据文件都必须含有满足某一应用的全部数据。比如，人事部门有一个职工文件，教育部门也有一个职工文件，人事部门的职工文件的记录格式是：

职工基本情况	有关人事管理的数据
--------	-----------

教育部门的职工文件的记录格式是：

职工基本情况	有关教育培训的数据
--------	-----------

这样在两个部门的职工文件中都有“职工基本情况”的数据，也就是说这一部分数据是重复存储的，如果还有第三、第四个部门也有类似的职工文件，那么重复存储所造成的空间浪费是很大的。在数据库中，可以共享类似“职工基本情况”这样的共用数据，从而降低数据的冗余度。

1.1.4 数据具有较高的独立性

数据独立性是指数据的组织和存储方法与应用程序互不依赖、彼此独立的特性。在数据库技术之前，数据文件的组织方式和应用程序是密切相关的，当改变数据结构时，相应的应用程序也必须随之修改，这样就大大增加了应用程序的开发代价和维护代价。而数据库技术却可以使数据的组织和存储方法与应用程序互不依赖，从而大大降低应用程序的开发代价和维护代价。

1.1.5 保证数据的安全、可靠

数据库技术要能够保证数据库中的数据是安全、可靠的。数据库要有一套安全机制，以便可以有效地防止数据库中的数据被非法使用或非法修改；数据库还要有一套完整的备份和恢复机制，以便保证当数据遭到破坏时（软件或硬件故障引起的），能立刻将数据完全恢复，从而保证系统能够连续、可靠地运行。

1.1.6 最大限度地保证数据的正确性

保证数据正确的特性在数据库中称之为数据完整性。在数据库中可以通过建立一些约束条件保证数据库中的数据是正确的。比如，某学生的年龄是 20 岁，当误输入为 2 岁或 200 岁时，数据库能够主动拒绝这类错误。

1.1.7 数据可以并发使用并能同时保证数据的一致性

数据库中的数据是共享的，并且允许多个用户同时使用相同的数据，这就要求数据库能够协调一致，保证各个用户之间对数据的操作不发生矛盾和冲突，即在多个用户同时使用数据库时，能够保证数据的一致性和正确性。

以上概括介绍了数据库的主要特性或特征，在后续的章节中对它们会作出更详细的解释，将会回答一些为什么……？什么是……？如何做……？等问题。

1.2 数据库管理系统

上一节提到的数据库的各种功能和特性，并不是数据库中的数据固有的，它是靠管理或支持数据库的系统软件——数据库管理系统（DataBase Management System，简称 DBMS）提供的。一个完备的数据库管理系统应该具备上一节提到的各种功能，它的任务就是对数据资源进行管理，并且使之能为多个用户共享，同时还能保证数据的安全性、可靠性、完整性、一致性，还要保证数据的高度独立性。

具体来说，一个数据库管理系统应该具备如下功能：

1. 数据库定义功能：可以定义数据库的结构和数据库的存储结构，可以定义数据库中数据之间的联系，可以定义数据的完整性约束条件和保证完整性的触发机制等；
2. 数据库操纵功能：可以完成对数据库中数据的操纵，可以装入、删除、修改数据，可以重新组织数据库的存储结构，可以完成数据库的备份和恢复等操作；
3. 数据库查询功能：可以以各种方式提供灵活的查询功能，使用户可以方便地使用数据库中的数据；
4. 数据库控制功能：可以完成对数据库的安全性控制、完整性控制、多用户环境下的并发控制等各方面的控制；
5. 数据库通信功能：在分布式数据库或提供网络操作功能的数据库中还必须提供数据库的通信功能。

1.3 数据库系统和数据库管理员

数据库管理员负责数据库的全面管理工作，数据库系统则包括了数据库、数据库管理系统、数据库管理员等内容。

1.3.1 数据库系统

前面介绍了数据库的各种特征和管理数据库的系统软件 DBMS，那么什么是数据库系统呢？

简单地说，数据库系统就是基于数据库的计算机应用系统。这样一个系统除了包括：

- 以数据为主体的数据库；
- 管理数据库的系统软件 DBMS；

此外还要包括：

- 支持数据库系统的计算机硬件环境和操作系统环境；
- 管理和使用数据库系统的人，特别是负责设计、维护数据库的技术人员——数据库管理员；
- 方便使用和管理系统的各种技术说明书和使用说明书。

至此我们可以看出，数据库、数据库管理系统和数据库系统三个不同的概念。数据库

强调的是数据，数据库管理系统是系统软件，而数据库系统强调的是系统。

1.3.2 数据库管理和数据库管理员

数据库的使用会改变企事业单位的管理方式，最简单的原因就是很多部门或用户把他们的数据集中放在了数据库中，这自然会带来很多好处，如：

1. 由于消灭了数据重复和数据不一致性，数据将更加可靠、实用；
 2. 由于数据独立性，减少了程序的维护代价；
 3. 为数据的特定查询请求提供了快速响应；
- 等等。

然而，如果认为仅仅靠对称之为数据库或数据库管理系统的特定技术“黑箱”进行投资就可以得到这些好处，那将是错误的。因为要把众多部门或用户的数据放在同一个数据库中，那就必须要考虑这些数据会不会产生冲突？会不会有越权使用数据的情况发生？重要的数据会不会丢失？诸如此类的问题是用户非常关心的。为了解决这些问题就要有一个数据库管理的部门，这个部门将负责和数据库管理有关的所有工作，也就是说，负责数据库的管理。

从事数据库管理工作的人员称之为数据库管理员（Database Administrator，简称为 DBA），DBA 有大量的工作要做，既有技术方面的工作，又有管理方面的工作。DBA 要和各种人员打交道，包括普通应用人员、计算机专业人员、企事业管理人员等，DBA 要参加数据库开发和使用的全部工作，总的来说 DBA 的工作可以概括如下：

1. 首先在数据库规划阶段要参与选择和评价与数据库有关的计算机软件和硬件，要与数据库用户共同确定数据库系统的目标和数据库应用需求，要确定数据库的开发计划；
2. 在数据库设计阶段要负责数据库标准的制定和共用数据字典的研制，要负责各级数据库模式的设计，负责数据库安全、可靠方面的设计；
3. 在数据库运行阶段首先要负责对用户进行数据库方面的培训；负责数据库的转储和恢复；负责对数据库中的数据进行维护；负责监视数据库的性能，并调整、改善数据库的性能，提高系统的效率；继续负责数据库安全系统的管理；在运行过程中发现问题、解决问题。

由此可见，数据库管理员的工作是十分繁重和十分重要的，要求作为 DBA 的人员除了要掌握一定的数据处理、数据库技术之外，还应有处理好人际关系的优良素质和能力。在一个企事业中，特别是对一个规模较大的数据库，不能指望一两个人能够完成数据库管理工作，所以 DBA 通常是指数据库管理部门。

数据库管理员这一职位是非常重要、非常关键的，任何一个数据库系统如果没有数据库管理员或相应的部门执行 DBA 的职责，数据处理自动化就难以成功，数据库就会失去统一的管理和控制，从而造成数据库的混乱。所以，在开发数据库系统时，一开始就应设置数据库管理员的职位或相应的机构，要明确 DBA 的责任，也要保证 DBA 的权限。DBA 一定是那些懂得和掌握数据库全局工作、并作为设计和管理数据库的核心人员，在数据库系统的开发过程中，DBA 应该发挥极其重要的作用。

1.4 数据库的过去、现在和未来

数据库的核心任务是数据管理，它包括数据的分类、组织、编码、存储、检索和维护等。但并不是一开始就有数据库技术，在计算机诞生的初期，计算机主要用于科学计算，虽然此时同样有数据管理的问题，但这时的数据管理是以人工的方式进行的，后来发展到文件系统，再后来才是数据库。也就是说，数据管理经历了人工管理阶段、文件系统阶段和数据库系统阶段。

1.4.1 人工管理阶段

人工管理阶段是指计算机诞生的初期（20 世纪 50 年代中期以前）。这个时期的计算机技术，从硬件看还没有磁盘这样的可直接存取的存储设备，从软件看没有操作系统，更没有管理数据的软件。所以这个时期数据管理的特点是：

1. 数据不保存。因为计算机主要用于科学计算，一般也不需要长期保存数据，只是在完成某一个计算或课题时才将数据输入，然后不仅原始数据不保存，计算结果也不保存。
2. 还没有文件的概念。这个时期的数据组织必须由每个程序的程序员自行组织和安排。
3. 一组数据对应一个程序。每组数据只对应一个应用，即使两个程序用到相同的数据，也必须各自定义、各自组织，数据无法共享、无法相互利用和互相参照。因此，程序和程序之间有大量的数据重复。
4. 没有形成完整的数据管理的概念。由于以上几个特点及没有对数据进行管理的软件系统，所以这个时期的每个程序都要包括数据存取方法、输入/输出方法和数据组织方法等。因为程序是直接面向存储结构的，所以存储结构的任何一点修改，都会导致程序的修改，程序与数据不具有独立性。

人工管理阶段的特点可用图 1.1 表示。

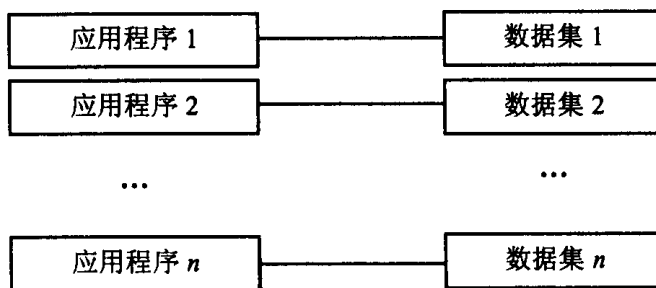


图 1.1

1.4.2 文件系统阶段

文件系统阶段是指 20 世纪 50 年代后期到 60 年代中期这一阶段。从那时起，计算机不仅大量用于科学计算，也开始大量用于信息管理。像磁盘这样的直接存取存储设备也已经出

现,在软件方面也有了操作系统和高级语言,也有了专门用于数据管理的软件——即文件系统(或操作系统的文件管理部分)。这个阶段的数据管理有以下一些特点:

1. 数据可以长期保存在磁盘上,也可以反复使用,即可以经常对文件进行查询、修改、插入和删除等操作。
2. 操作系统提供了文件管理功能和访问文件的存取方法,程序和数据之间有了数据存取的接口,程序开始通过文件名和数据打交道,可以不再关心数据的物理存放位置。因此,这时也有了数据的物理结构和数据的逻辑结构的区别。程序和数据之间有了一定的独立性。
3. 文件的形式已经多样化。由于有了磁盘这样的直接存取存储设备,文件也就不再局限于顺序文件,也有了索引文件、链表文件等。因而,对文件的访问可以是顺序访问,也可以是直接访问。但文件之间是独立的,它们之间的联系要通过程序去构造,文件的共享性也还比较差。
4. 有了存储文件以后,数据就不再仅仅属于某个特定的程序,而可以由多个程序反复使用。但文件结构的设计仍然是基于特定的用途,程序仍然是基于特定的物理结构和存取方法编制的。因此,数据的存储结构和程序之间的依赖关系并未根本改变。
5. 数据的存取基本上以记录为单位。

图 1.2 示意了文件阶段的特点及程序和数据之间的关系。

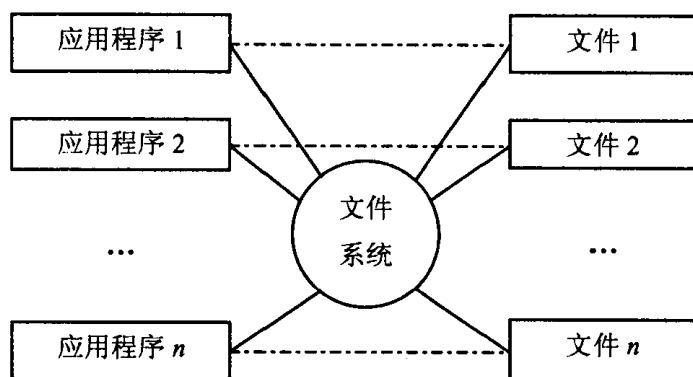


图 1.2

虽然文件系统比人工管理有了长足的进步,但是文件系统所能提供的存取方法和数据管理仅仅是初级水平。无论如何,文件系统阶段是数据管理技术发展中的一个重要阶段,在这个阶段中得到充分发展的各种数据结构和算法等,都大大丰富了计算机科学,今天的数据库技术也正是在文件系统的基础上发展起来的。

现在,站在数据库的观点上反观文件系统,可以看出文件系统有以下几方面明显的缺陷:

1. 数据冗余大。这是因为每个文件都是为特定的用途设计的,因此就会造成同样的数据在多个文件中重复存储。
2. 数据不一致性。这往往是由数据冗余造成的,在进行更新时,稍不谨慎就会造成同一数据在不同文件中的不一致。
3. 程序和数据之间的独立性差。应用程序依赖于文件的存储结构,使得若修改文件的存储结构则必须修改程序。

4. 数据联系弱。文件与文件之间是独立的，文件之间的联系必须通过程序来构造。因此，文件系统是一个不具有弹性的、无结构的数据集合，不能反映现实世界事物之间的联系。

1.4.3 数据库系统阶段及其发展过程

数据库系统阶段从 20 世纪 60 年代后期开始，数据库技术的诞生既有计算机技术的发展做依托，又有数据管理的需求做动力。数据库的数据不再是面向某个应用或某个程序，而是面向整个企业（组织）或整个应用的，图 1.3 示意了这种特点。

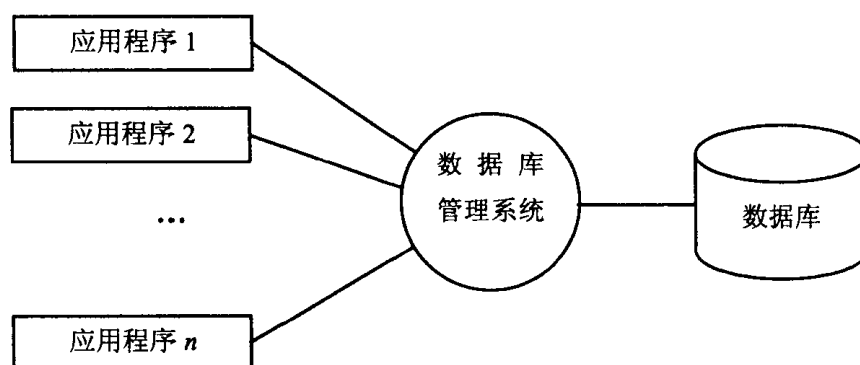


图 1.3

关于数据库的特点及优点在本章的第一节已经做了比较详细的叙述，这里不再重复。

数据库技术的诞生可以以 20 世纪 60 年代末和 70 年代初的三个事件做标志，它们是：

1. 1968 年研制成功、1969 年形成产品的美国 IBM 公司的数据库管理系统 IMS (Information Management System) 的问世，该系统支持的是层次数据模型。

2. 美国数据系统语言协会 CODASYL (Conference On DAta SYstem Language) 下属的数据库任务组 DBTG (DataBase Task Group) 对数据库方法进行了系统的研究，在 20 世纪 60 年代末和 70 年代初发表了若干个报告（称为 DBTG 报告），该报告建立了数据库技术的很多概念、方法和技术。DBTG 所提议的方法是基于网状数据模型的。

3. 从 1970 年起，IBM 的研究员 E.F.Codd 发表了一系列的论文，提出了数据库的关系模型，开创了数据库关系方法和关系数据理论的研究，为关系数据库的发展和理论研究奠定了基础。

可以说，20 世纪 70 年代是以 IMS 为代表的层次数据库和以 DBTG 为代表的网状数据库的鼎盛时期。这些数据库都有很高的效率，特别是 DBTG 报告包含了很多完备的概念和技术，这大大推动了计算机在信息管理领域的应用，也使许多商品化的数据库开始出现，数据库技术日益广泛地应用到企业管理、交通运输、情报检索、军事指挥、政府管理和辅助决策等各个方面。层次数据库和网状数据库是面向专业人员的，要求使用人员要有较高的技术水平和专业水平，使用起来比较困难。

20 世纪 70 年代也是关系数据库的萌动时期。与层次数据库和网状数据库相比，研究关系数据库的出发点之一就是简单、易用，这样就要尽可能地减少人为的操作，让计算机自动

地完成更多的工作。受当时计算机硬件和软件技术的约束，关系数据库经历了比较长时间的研究过程。尽管在 20 世纪 60 年代末和 70 年代初就提出了关系数据库的概念，但是关系数据库真正得到广泛的应用，则是进入 20 世纪 80 年代以后的事情。

在国内，20 世纪 70 年代就开始使用数据库的单位主要集中在国家部委、国防军事、气象预报、石油勘探等一些特殊的行业和部门。而数据库技术真正得到广泛推广和使用，可以说是从 20 世纪 80 年代初的 dBASE II 开始的，尽管 dBASE II、甚至其后的 xBASE 系列都不能称为是一个完备的数据库管理系统，但是它们支持基本的关系数据模型，使用起来非常方便，一般也能满足中、小规模的信息管理的应用。现在使用的各种数据库系统均是基于关系数据模型的，层次数据库和网状数据库由于其自身的缺点均已被关系数据库取代。

从 20 世纪 80 年代算起，又经过了近 20 年，关系数据库系统的技术也在不断完善和提高，有关数据库的新的研究课题不断取得进展，如分布式数据库已经成为实用技术，基于关系模型、支持面向对象方法的关系对象模型数据库也已经问世等。数据库技术还会朝着支持更大规模、更快速度、更广泛的应用等方向发展。

1.4.4 数据库技术的广泛应用领域

数据库技术的最初应用领域主要是信息管理领域，如政府部门、工商企业、图书情报、交通运输、银行金融、科研教育等各行各业的信息管理和信息处理。事实上，只要有大量的数据要管理、需要有大量数据支持的工作，都可以使用数据库，下面介绍几个比较有代表性的应用领域：

1. 因特网上的 Web 数据库。因特网即国际互连网，它用一种统一的网络协议可以将全世界的计算机连成网络，最初它只能提供事先制作好的静态信息。后来，人们为了从因特网上得到动态的、实时的信息，则将数据库技术引入因特网，从而有了 Web 数据库，Web 数据库是因特网上的数据库的一种称呼。

2. 多媒体数据库。数据库不仅可以存储和管理文字和报表数据，还可以存储和管理如声音、图片、动画等各种媒体的数据，这种多媒体数据库可以支持广泛的应用。

3. 工程数据库。主要用于管理工程数据，如计算机辅助设计（CAD）和计算机辅助制造（CAM）过程中使用的数据，以及设计的图纸、工艺流程等数据的存储和管理。

4. 辅助决策支持。早期的模型库、方法库和现在广泛讨论的数据仓库技术都是用于辅助决策支持的。

5. 人工智能领域的知识库。人工智能是从 20 世纪 60 年代开始发展起来的研究机器智能和智能机器的高科技学科，它需要大量的演绎和推理规则的支持，这无疑又为数据库提供了一个用武之地。通过将人的知识抽象化、条理化，利用数据库技术建立知识库，从而使数据库智能化。

以上只是概括介绍了一些数据库的应用领域，实际上还远远不止这些，这些技术也不只是用于单一领域，通过这些技术结合可以有更广阔的领域，或者说一些应用领域需要综合数据库技术的支持。在本书的最后一章（第十一章）将对这些领域做进一步的介绍。