

数据库系统概论

SHUJUKU XITONG GAILUN

主编 / 刘 庆 韩蓓蓓

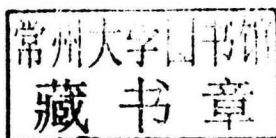


电子科技大学出版社

数据库系统概论

SHUJUKU XITONG GAILUN

主编 / 刘 庆 韩蓓蓓



电子科技大学出版社

图书在版编目(CIP)数据

数据库系统概论 / 刘庆, 韩蓓蓓主编. —成都:

电子科技大学出版社, 2015.9

ISBN 978-7-5647-3280-6

I. ①数… II. ①刘… ②韩… III. ①数据库系统—
高等学校—教材 IV. ①TP311.13

中国版本图书馆 CIP 数据核字 (2015) 第 220363 号

数据库系统概论

刘 庆 韩蓓蓓 主编

出 版: 电子科技大学出版社

地 址: 成都市一环路东一段 159 号电子信息产业大厦 (邮编 610051)

策划编辑: 辜守义

责任编辑: 辜守义

主 页: www.uestcp.com.cn

电子邮箱: uestcp@uestcp.com.cn

发 行: 新华书店经销

印 刷: 四川永先数码印刷有限公司

成品尺寸: 185mm×260mm 印张 12 字数 290 千字

版 次: 2015 年 9 月第 1 版

印 次: 2015 年 9 月第 1 次印刷

书 号: ISBN 978-7-5647-3280-6

定 价: 30.00 元

版权所有★侵权必究

- ◆ 本社发行部电话: 028-83202463; 本社邮购电话: 028-83201495。
- ◆ 本书如有缺页、破损、装订错误, 请寄回印刷厂调换。

前 言

数据库技术作为计算机软件领域的一个重要分支，是计算机科学技术中发展最快的领域之一，也是应用最广的技术之一。数据库技术发展至今已成为以计算机为中心的信息系统与应用系统的核心技术和重要基础。

1964 年，美国通用电器公司的 Bachman 等人成功开发了世界上第一个 DBMS (Data-Base Management System) ——IDS 系统，标志着人们对数据的管理进入了数据库系统阶段。与文件系统相比，数据库系统实现了数据整体结构化。在文件系统中，虽然存在记录内的结构性，但整体上数据是无结构的，即不同文件之间的记录是没有联系的。但在数据库系统中，不仅存在记录内部的联系，还描述了数据之间的联系，实现了数据整体结构化。这是数据库系统与文件系统的本质差别。同时，数据库系统使数据面向整个应用系统，降低了数据的冗余度，实现了数据的共享。

人类在 21 世纪进入以知识经济为主要基础的信息社会，而数据库正是信息社会信息资源管理与开发利用的基础。

随着信息产业的发展，数据库在社会中发挥了越来越重要的作用。可以说，几乎所有的信息系统都依赖于数据库。通俗地讲，数据库是存放数据的仓库，而这个仓库是存放在计算机存储设备上，且是按照一定格式存放的。按照数据库理论的定义，数据库是长期存储在计算机内的、有组织的、可共享的数据集合。在今天，电子商务、电子政务等都得到了迅速的发展，并由此产生了大量的数据。为了能够高效、准确地分析这些数据，人们使用了数据库。

众所周知，“3C”即计算机、通信和信息内容 (Computer, Communication and Contents) 已成为信息技术的核心，而信息内容则主要存放在数据库中，书中共十章，分别讲述了数据库的基本内容，以及设计的过程和优化发展阶段等。因此，数据库的建设规模和使用水平便成为衡量一个国家信息化程度的重要标志。

本书在编写时之所以把内容新颖作为考虑的一个重要因素，是希望在数据库技术发展很快的情况下，可作为教材在 10 年之内都能基本适用，有利于教师熟悉教材及实验内容并能融会贯通，从而达到最好的教学效果。

本书的主要特点是：内容新颖，系统全面；突出重点，注重总结；概念清晰，分析深入；例题丰富，实用性强；叙述深入浅出，语言流畅生动。

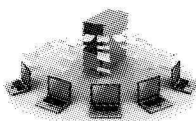
由于水平有限，书中难免有不妥之处，敬请读者批评改正。

编者

2015 年 6 月

目 录

第1章 绪 论	1
第1节 数据库概念	1
第2节 数据库系统概述	3
第3节 数据管理技术的演变	13
第4节 数据模型	16
第5节 数据库体系结构	21
第2章 关系数据库	25
第1节 关系数据结构及形式化定义	25
第2节 关系模型	30
第3节 关系的完整性	39
第4节 关系代数	41
第5节 关系演算	48
第3章 关系数据库标准语言 SQL	51
第1节 SQL 概述	51
第2节 数据定义	54
第3节 数据查询	57
第4节 数据更新	69
第5节 数据控制	72
第6节 视图	75
第4章 数据库设计	83
第1节 数据库设计概述	83
第2节 需求分析	87
第3节 概念设计	90
第4节 逻辑结构设计	94
第5节 物理结构设计	98
第6节 数据库的实施、运行和维护	102
第5章 数据库安全性	104
第1节 数据库安全性概述	104
第2节 数据库安全控制技术	107
第3节 视图机制	111
第4节 审计	111
第5节 数据加密	113



数据库系统概论

第6节 数据库安全性实例——Oracle 系统	115
第6章 数据库完整性	119
第1节 实体完整性	119
第2节 参照完整性	121
第3节 完整性的约束命名	124
第4节 触发器	125
第7章 并发控制	129
第1节 并发控制概述	129
第2节 封锁	130
第3节 活锁和死锁	133
第4节 并发调度和可串行性	136
第5节 两段锁协议	137
第6节 封锁的粒度	139
第8章 数据库恢复技术	141
第1节 数据库故障	141
第2节 恢复的实现技术	143
第3节 恢复策略	146
第4节 具有检查点的恢复技术	147
第5节 数据库镜像	149
第9章 查询处理和查询优化	150
第1节 关系数据库系统的查询处理	150
第2节 表达式计算	154
第3节 查询优化	157
第4节 物理优化	171
第5节 查询计划的执行	174
第10章 数据库技术的发展阶段	175
第1节 数据库技术的发展阶段	175
第2节 数据库新技术的研究和发展	176
第3节 大数据的应用	179
第4节 内存数据库的关键技术	186
第5节 数据仓库技术	200
第6节 联机分析处理技术	202
第7节 数据挖掘技术	204
参考文献	207

第1章 绪 论

第1节 数据库概念

数据库是一个很复杂的系统，涉及面很广，难以用简练的语言准确地概括其全部特征。因此，先从简单分析入手，逐步认识什么是数据库。

顾名思义，数据库就是存储数据的“仓库”。但它和普通的仓库是有所不同的。首先，数据不是存放在容器或空间中，是存放在计算机的外存储器中（如磁盘），并且是有组织地存放的。数据的管理和利用是通过计算机的数据管理软件——数据库管理系统来完成的。因此，我们讲的数据库，不单是指存有数据的计算机外存，也是指存放在外存上的数据集合以及管理它们的计算机软件的总和。数据库系统的组成可以用图 1-1 表示。

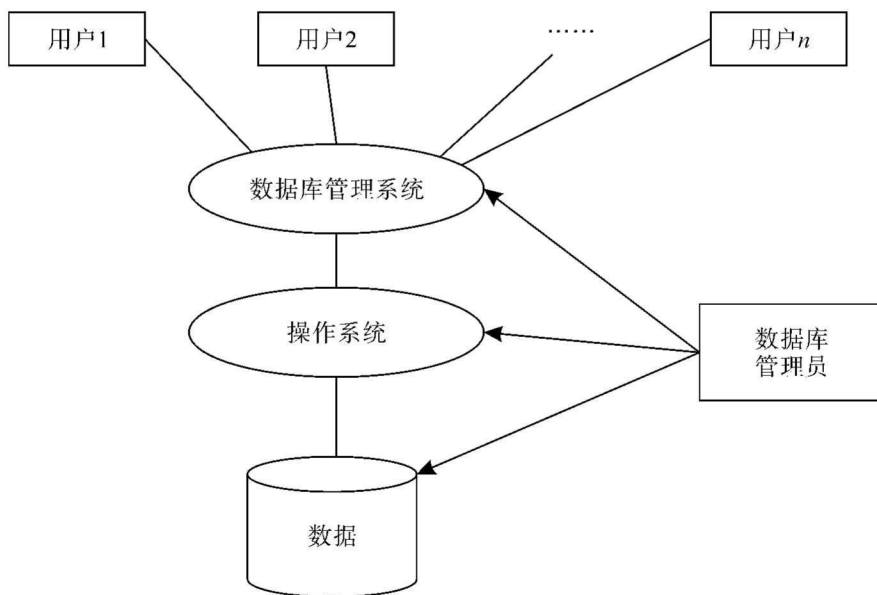
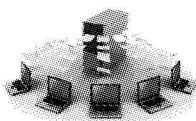


图 1-1 数据库系统

通常一个数据库系统包含下列内容：

（1）有一个结构化的相关数据的集合。这一集合称为数据库（DataBase 或 DB）。在这个数据集合中没有有害的或不必要的冗余，而且能够为多种应用提供服务。它独立于应用程序而存在。这种结构化的数据集合就是数据库本身，是数据库系统的核心和管理对象。

（2）有一个负责数据库管理和维护的软件系统。这一系统称为数据库管理系统（DataBase Management System 或 DBMS）。用户一般不直接加工或使用数据库中的数据，而



数据库系统概论

必须通过数据库管理系统。DBMS 的主要功能是维持数据库系统的正常活动,接受并响应用户对数据库的一切访问要求,包括建立和删除数据文件,检索、统计、修改和组织数据库中的数据以及为用户提供对数据库的维护手段等。通过使用 DBMS,用户可逻辑、抽象地处理数据,不必关心数据在计算机中的具体存储方式以及计算机处理数据的过程细节。这样,把一切处理数据的具体而繁杂的工作交给 DBMS 去完成。就好像操作系统的出现,解放了用户,使用户不必关心数据的实际存放和读取,而只需给出文件名和路径一样。

(3) 有供数据库及数据库管理系统运行的运行环境。运行环境主要是指计算机硬件系统以及基础的软件支持系统(如:操作系统等)。

(4) 有一个(或一组)负责整个数据系统的建立、维护和协调工作的专门人员。这就是数据库管理员(DataBase Administrator 或 DBA)。在一个安全性较高的大型的数据库管理系统中,如金融部门等,必须有专门的数据库管理员,从事监视应用程序、维护硬件设备、定时备份等工作。他们也是一个数据库管理系统中不可缺少的重要部分。

(5) 有若干个用户。这里的用户可以是使用该数据库的终端用户或者应用程序。数据库系统作为计算机的一个分支,与计算机硬件及其他基础软件和系统软件有密切的关系。它几乎涉及硬件及软件的所有知识,是许多重要软件技术的综合应用。例如,数据结构、操作系统、编译技术、程序设计等知识在数据库中都将用到。所以数据库系统是一门综合性的计算机技术,也是一门很有意义很有趣味的学科。要更好地研究掌握它,必须了解计算机的各个方面,以便加深理解这些知识的内在联系,并将它们统一起来。

数据库技术之所以能够如此快速地发展,受到计算机科学界普遍的重视,成为引人注目的一门重要学科,是因为它具有如下的特点。

(1) 采用数据模型表示复杂的数据结构。在文件系统中,尽管其记录内部已有了某些结构,但记录之间没有联系。而采用数据模型表示复杂的数据结构,实现整体数据的结构化,是数据库的主要特征之一,也是数据库系统与文件系统的本质区别。

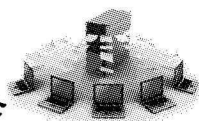
在数据库系统中,数据不再针对某一应用,而是面向全组织,具有整体的结构化。不仅数据是结构化的,而且其存取数据的方式也很灵活,可以存取数据库中的某一个数据项、一组数据项、一个记录或一组记录。而在文件系统中,数据的最小存取单位是记录,粒度不能细到数据项。

(2) 数据的共享性高,冗余度低,易扩充。数据库系统从整体角度看待和描述数据,数据不再面向某个应用而是面向整个系统,因此数据可以被多个用户、多个应用共享使用。数据共享可以大大减少数据冗余,节约存储空间。数据共享还能够避免数据之间的不相容性与不一致性。

所谓数据的不一致性是指同一数据不同拷贝(副本)的值不一样。采用人工管理或文件系统管理时,由于数据被重复存储,当不同的应用程序使用和修改不同的拷贝时就很容易造成数据的不一致。在数据库中数据共享,减少了由于数据冗余造成的一致现象。

由于数据面向整个系统,是具有结构化的数据,因此不仅可以被多个应用共享使用,也使用户更容易增加新的应用。这就使得数据库系统弹性大,易于扩充,可以适应各种用户的要求。例如,可以取整体数据的各种子集用于不同的应用系统。当应用需求改变或增加时,只要重新选取不同的子集或加上一部分数据便可以满足新的需求。

(3) 数据具有较高的独立性。数据独立性是数据库领域一个常用术语,包括数据的物



理独立性和数据的逻辑独立性。

物理独立性是指用户的应用程序与存储在磁盘上的数据库中数据是相互独立的。也就是说，数据在磁盘上的数据库中怎样存储是由 DBMS 管理的，用户程序不需要了解，应用程序要处理的只是数据的逻辑结构。这样即使数据的物理存储改变时，应用程序也不用改变。

逻辑独立性是指用户的应用程序与数据库的逻辑结构是相互独立的。也就是说，即使数据的逻辑结构改变了，应用程序也可以不变。

数据独立性是由 DBMS 的二级映射功能来保证的。

数据与程序独立，把数据的定义从程序中分离出去，加上数据的存取是由 DBMS 负责，从而简化了应用程序的编制，大大减少了应用程序的维护和修改。

(4) 数据由 DBMS 统一管理和控制。数据库的共享是并发的共享，即多个用户可以同时存取数据库中的数据甚至可以同时存取数据库中同一个数据。

数据库管理阶段应用程序与数据之间的对应关系如图 1-1 所示。

在这里需要强调的是数据库系统的核心是 DBMS。一个数据库系统的功能和性能在很大程度上取决于 DBMS 的功能和性能。

数据库系统的出现使信息系统从以加工数据的程序为中心转向围绕共享的数据库为中心的新阶段。这样既便于数据的集中管理，又有利于应用程序的研制和维护，提高了数据的利用率和相容性，也提高了决策的可靠性。

目前，数据库已经成为现代信息系统不可分离的重要组成部分。具有数百万甚至数十亿字节信息的数据库已经普遍存在于科学技术、工业、农业、商业、服务业和政府部门的信息系统。20 世纪 80 年代后不仅在大型机上，大多数微机上也配置了 DBMS，使数据库技术得到更加广泛的应用。

第 2 节 数据库系统概述

在系统地介绍数据库的基本概念之前，这里首先介绍一些数据库最常用的术语和基本概念。

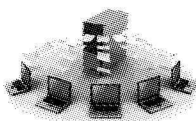
1.2.1 数据库的 4 个基本概念

数据、数据库、数据库管理系统和数据库系统是与数据库技术密切相关的 4 个基本概念。

1. 数据(Data)

数据是数据库中存储的基本对象。数据在大多数人头脑中的第一个反应就是数字，如 93、1 000、99.5、-330.86、¥6 880、\$726 等。其实数字只是最简单的一种数据，是数据的一种传统和狭义的理解。广义的理解认为数据的种类很多，文本(Text)、图形(Graph)、图像(Image)、音频(Audio)、视频(Video)、学生的档案记录、货物的运输情况等，这些都是数据。

可以对数据做如下定义：描述事物的符号记录称为数据。描述事物的符号可以是数



数据库系统概论

字,也可以是文字、图形、图像、音频、视频等,数据有多种表现形式,它们都可以经过数字化后存入计算机。

在现代计算机系统中数据的概念是广义的。早期的计算机系统主要用于科学计算,处理的数据是数值型数据,如整数、实数、浮点数等。现在计算机存储和处理的对象十分广泛,表示这些对象的数据也随之变得越来越复杂。

数据的表现形式还不能完全表达其内容,需要经过解释,数据和关于数据的解释是不可分的。例如,93 是一个数据,可以是一个同学某门课的成绩,也可以是某个人的体重,还可以是计算机系 2013 级的学生人数。数据的解释是指对数据含义的说明,数据的含义称为数据的语义,数据与其语义是不可分的。

在日常生活中,人们可以直接用自然语言(如汉语)来描述事物。例如,可以这样来描述某校计算机系一位同学的基本情况:李明同学,男,1995 年 5 月生,江苏省南京市人,2013 年入学。在计算机中常常这样来描述:

(李明,男,199505,江苏省南京市,计算机系,2013)

即把学生的姓名、性别、出生年月、出生地、所在院系、入学时间等组织在一起,构成一个记录。这里的学生记录就是描述学生的数据。这样的数据是有结构的。记录是计算机中表示和存储数据的一种格式或一种方法。

2. 数据库(DataBase, DB)

数据库,顾名思义,是存放数据的仓库。只不过这个仓库是在计算机存储设备上,而且数据是按一定的格式存放的。

人们收集并抽取出一个应用所需要的大量数据之后,应将其保存起来,以供进一步加工处理,进一步抽取有用信息。在科学技术飞速发展的今天,人们的视野越来越广,数据量急剧增加。过去人们把数据存放在文件柜里,现在人们借助计算机和数据库技术科学地保存和管理大量复杂的数据,以便能方便而充分地利用这些宝贵的信息资源。

严格地讲,数据库是长期储存在计算机内、有组织的、可共享的大量数据的集合。数据库中的数据按一定的数据模型组织、描述和储存,具有较小的冗余度(redundancy)、较高的数据独立性(data independency)和易扩展性(scalability),并可为各种用户共享。

概括地讲,数据库数据具有永久存储、有组织和可共享三个基本特点。

3. 数据库管理系统(DataBase Management System, DBMS)

了解了数据和数据库的概念,下一个问题就是如何科学地组织和存储数据,如何高效地获取和维护数据。完成这个任务的是一个系统软件——数据库管理系统。

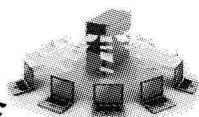
数据库管理系统是位于用户与操作系统之间的一层数据管理软件。数据库管理系统和操作系统一样是计算机的基础软件,也是一个大型复杂的软件系统。它的主要功能包括以下几个方面:

(1) 数据定义功能

数据库管理系统提供数据定义语言(Data Definition Language, DDL),用户通过它可以方便地对数据库中的数据对象的组成与结构进行定义。

(2) 数据组织、存储和管理

数据库管理系统要分类组织、存储和管理各种数据,包括数据字典、用户数据、数据



的存取路径等。要确定以何种文件结构和存取方式在存储级上组织这些数据,如何实现数据之间的联系。数据组织和存储的基本目标是提高存储空间利用率和方便存取,提供多种存取方法(如索引查找、hash 查找、顺序查找等)来提高存取效率。

(3) 数据操纵功能

数据库管理系统还提供数据操纵语言(Data Manipulation Language, DML),用户可以使用它操纵数据,实现对数据库的基本操作,如查询、插入、删除和修改等。

(4) 数据库的事务管理和运行管理

数据库在建立、运用和维护时由数据库管理系统统一管理和控制,以保证事务的正确运行,保证数据的安全性、完整性、多用户对数据的并发使用及发生故障后的系统恢复。

(5) 数据库的建立和维护功能

数据库的建立和维护功能包括:数据库初始数据的输入、转换功能,数据库的转储、恢复功能,数据库的重组功能和性能监视、分析功能等。这些功能通常是由一些实用程序或管理工具完成的。

(6) 其他功能

其他功能包括:数据库管理系统与网络中其他软件系统的通信功能;一个数据库管理系统与另一个数据库管理系统或文件系统的数据库转换功能;异构数据库之间的互访和互操作功能等。

4. 数据库系统(DataBase System, DBS)

数据库系统是由数据库、数据库管理系统(及其应用开发工具)、应用程序和数据库管理员(DataBase Administrator, DBA)组成的存储、管理、处理和维护数据的系统。应当指出的是,数据库的建立、使用和维护等工作只靠一个数据库管理系统远远不够,还要有专门的人员来完成,这些人被称为数据库管理员。

数据库系统可以用图 1-2 表示。其中数据库提供数据的存储功能,数据库管理系统提供数据的组织、存取、管理和维护等基础功能,数据库应用系统根据应用需求使用数据库,数据库管理员负责全面管理数据库系统。图 1-3 是引入数据库后计算机系统的层次结构。

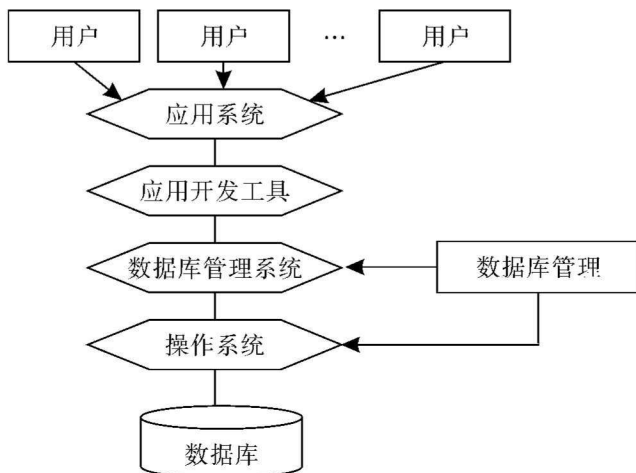


图 1-2 数据库系统

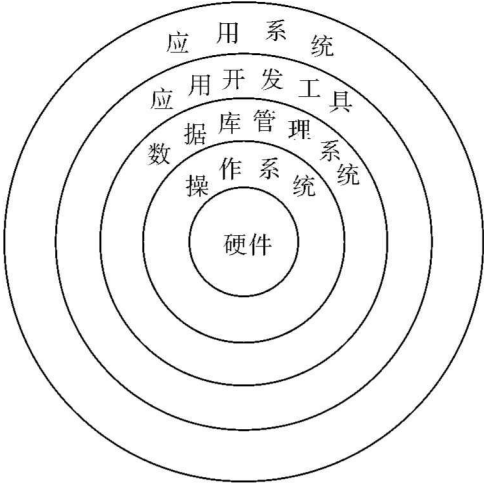
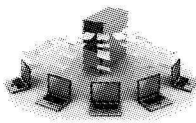


图 1-3 引入数据库后计算机系统的层次结构

在一般不引起混淆的情况下，人们常常把数据库系统简称为数据库。

1.2.2 数据管理系统的产生和发展

数据库技术是应数据管理任务的需要而产生的。数据管理是指对数据进行分类、组织、编码、存储、检索和维护，它是数据处理的中心问题。而数据的处理是指对各种数据进行收集、存储、加工和传播的一系列活动的总和。

在应用需求的推动下，在计算机硬件、软件发展的基础上，数据管理技术经历了人工管理、文件系统、数据库系统三个阶段。这三个阶段的特点及其比较如表 1-1 所示。

1. 人工管理阶段

20 世纪 50 年代中期以前，计算机主要用于科学计算。当时的硬件状况是，外存只有纸带、卡片、磁带，没有磁盘等直接存取的存储设备；软件状况是，没有操作系统，没有管理数据的专门软件；数据处理方式是批处理。人工管理数据具有如下特点：

表 1-1 数据管理 3 个阶段的比较

		人工管理阶段	文件系统阶段	数据库系统阶段
背景	应用背景	科学计算	科学计算、数据管理	大规模数据管理
	硬件背景	无直接存取存储设备	磁盘、磁鼓	大容量磁盘、磁盘阵列
	软件背景	没有操作系统	有文件系统	有数据库管理系统
	处理方式	批处理	联机实时处理、批处理	联机实时处理、分布处理、批处理



(续表)

		人工管理阶段	文件系统阶段	数据库系统阶段
特点	数据的管理者	用户(程序员)	文件系统	数据库管理系统
	数据面向的对象	某一应用程序	某一应用	现实世界(一个部门、企业、跨国组织等)
	数据的共享程度	无共享、冗余度极大	共享极差,冗余度大	共享性高,冗余度小
	数据的独立性	不独立,完全依赖于程序	独立性差	具有高度的物理独立性和一定的逻辑独立性
	数据的结构化	无结构	记录内有结构、整体无结构	整体结构化,用数据模型描述
	数据控制能力	应用程序自己控制	应用程序自己控制	由数据库管理系统提供数据安全性、完整性、并发控制和恢复能力

(1) 数据不保存

由于当时计算机主要用于科学计算,一般不需要将数据长期保存,只是在计算某一课题时将数据输入,用完就撤走。不仅对用户数据如此处置,对系统软件有时也是这样。

(2) 应用程序管理数据

数据需要由应用程序自己设计、说明(定义)和管理,没有相应的软件系统负责数据的管理工作。应用程序中不仅要规定数据的逻辑结构,而且要设计物理结构,包括存储结构、存取方法、输入方式等。因此程序员负担很重。

(3) 数据不共享

数据是面向应用程序的,一组数据只能对应一个程序。当多个应用程序涉及某些相同的数据时必须各自定义,无法互相利用、互相参照,因此程序与程序之间有大量的冗余数据。

(4) 数据不具有独立性

数据的逻辑结构或物理结构发生变化后,必须对应用程序做相应的修改,数据完全依赖于应用程序,称之为数据缺乏独立性,这就加重了程序员的负担。

在人工管理阶段,应用程序与数据之间的一一对应关系可用图 1-4 表示。

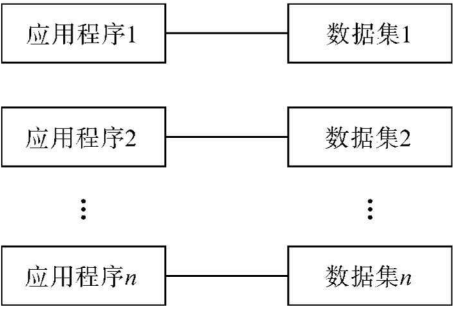
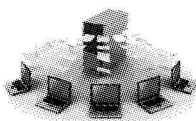


图 1-4 人工管理阶段应用程序与数据之间的一一对应关系

2. 文件系统阶段

20 世纪 50 年代后期到 60 年代中期,这时硬件方面已有了磁盘、磁鼓等直接存取存储



数据库系统概论

设备；软件方面，操作系统中已经有了专门的数据管理软件，一般称为文件系统；处理方式上不仅有了批处理，而且能够联机实时处理。

用文件系统管理数据具有如下特点。

(1) 数据可以长期保存

由于计算机大量用于数据处理,数据需要长期保留在外存上反复进行查询、修改、插入和删除等操作。

(2) 由文件系统管理数据

由专门的软件即文件系统进行数据管理,文件系统把数据组织成相互独立的数据文件,利用“按文件名访问,按记录进行存取”的管理技术,提供了对文件进行打开与关闭、对记录读取和写入等存取方式。文件系统实现了记录内的结构性。但是,文件系统仍存在以下缺点:

①数据共享性差,冗余度大。在文件系统中,一个(或一组)文件基本上对应于一个应用程序,即文件仍然是面向应用的。当不同的应用程序具有部分相同的数据时,也必须建立各自的文件,而不能共享相同的数据,因此数据的冗余度大,浪费存储空间。同时由于相同数据的重复存储、各自管理,容易造成数据的不一致性,给数据的修改和维护带来了困难。

②数据独立性差。文件系统中的文件是为某一特定应用服务的,文件的逻辑结构是针对具体的应用来设计和优化的,因此要想对文件中的数据再增加一些新的应用会很困难。而且当数据的逻辑结构改变时,应用程序中文件结构的定义必须修改,应用程序中对数据的使用也要改变,因此数据依赖于应用程序,缺乏独立性。可见,文件系统仍然是一个不具有弹性的无整体结构的数据集合,即文件之间是孤立的,不能反映现实世界事物之间的内在联系。

文件系统阶段应用程序与数据之间的关系如图 1-5 所示。

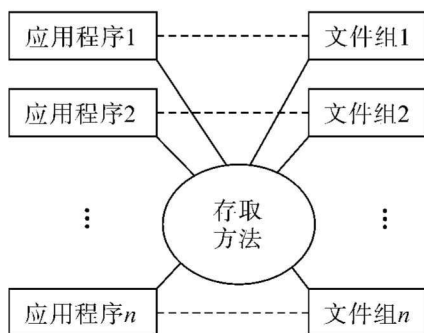
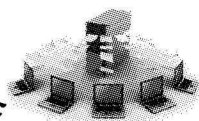


图 1-5 文件系统阶段应用程序与数据之间的对应关系

3. 数据库系统阶段

20 世纪 60 年代后期以来,计算机管理的对象规模越来越大,应用范围越来越广泛,数据量急剧增长,同时多种应用、多种语言互相覆盖地共享数据集合的要求越来越强烈。

这时硬件已有大容量磁盘,硬件价格下降;软件则价格上升,为编制和维护系统软件及应用程序所需的成本相对增加;在处理方式上,联机实时处理要求更多,并开始提出和考虑分布处理。在这种背景下,以文件系统作为数据管理手段已经不能满足应用的需求,于是为解决多用户、多应用共享数据的需求,使数据为尽可能多的应用服务,数据库技术



便应运而生，出现了统一管理数据的专门软件系统——数据库管理系统。

用数据库系统来管理数据比文件系统具有明显的优点，从文件系统到数据库系统标志着数据管理技术的飞跃。

1.2.3 数据库系统的特点

下面首先通过一个简单的例子——学生学籍管理系统来比较文件系统和数据库系统的差异，从而阐述数据库系统的特点。

【例 1-1】设一个学生的信息包括学号、姓名、性别、年龄、专业和奖励。部分学生的情况，如表 1-2 所示。假设该学籍管理系统具有录入学生信息、根据学号可以找到一个学生信息等功能。

(1) 采用文件系统实现学籍管理。首先是如何存储学生数据。计算机操作系统实现了文件系统，可以把每个学生都具备的信息采用定长记录方式存放在一个“学生基本信息”文件中。有关奖励的数据有的学生较多，有的学生没有，因此采用变长记录方式把它存放在另外一个“奖励”文件中。“学生基本信息”文件的记录包括学号、姓名、性别、年龄、专业、位置和长度等字段，如表 1-3 所示。其中位置和长度描述的是另一个奖励文件(表 1-4 所示)中记录的开始位置和长度。文件系统管理的文件是流式文件，或者说只是一些字节流。因此，表 1-3 和表 1-4 是为了方便程序员理解的结构。程序员必须把它们在应用程序中加以说明和描述。

表 1-2 学生情况

学号	姓名	性别	年龄	专业	奖励
20100001	史玉明	女	20	计算机	2011 校奖学金, 2012 国家奖学金
20100100	李明虎	男	21	机械	2012 校优秀学生
...	...	...	...	...	...

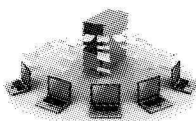
表 1-3 “学生基本信息”文件的结构和内容

学号	姓名	性别	年龄	专业	位置	长度
20100001	史玉明	女	20	计算机	0	30
20100100	李明虎	男	21	机械	30	15
20100234	张翔	男	21	化工	45	0
...	...	...	...	...	...	...

表 1-4 “奖励”文件的结构和内容

奖励
2011 校奖学金, 2012 国家奖学金
2012 校优秀学生

确定了学生数据的存储方式后，需要编写程序来实现数据的录入功能和查询功能。录入功能的基本过程包括从键盘读入学生信息，把基本信息写到“学生基本信息”文件中，把奖励情况写到“奖励”文件中。特别要注意的是，为了能正确地表达“学生基本信息”



数据库系统概论

文件中一条记录和“奖励”文件中记录的对应关系，在程序中要把奖励情况在文件中的开始位置和长度再写回“学生基本信息”文件中。

查询功能采用顺序查找方法。首先从“学生基本信息”文件中读入第1条记录，然后比较学号字段的值是否和要查找的学号相同。如果相同，读出该学生的信息，并根据位置字段和长度字段的值到“奖励”文件中读出该学生的奖励信息，查找过程结束；如果不相同，则从“学生基本信息”文件中读入下一条记录，直到找到该学号的记录，或者读到文件末尾也没有该学号的学生为止。

(2) 采用数据库系统实现学籍管理

首先在数据库中建立两张表:STUDENT 存放学生的基本信息, AWARD 存放学生的奖励情况。这可以用数据库管理系统的两条 CREATE 语句来实现:

```
CREATE TABLE STUDENT(  
    Sno      CHAR(8),  
    Sname    CHAR(10),  
    Ssex     CHAR(2),  
    Sage     SMALLINT,  
    Major    CHAR(20));
```

```
CREATE TABLE AWARD(  
    Sno      CHAR(8),  
    Details  VARCHAR(2000));
```

建立好表结构后,向数据库管理系统提交下面的两条插入命令就可以把学生的基本信息和奖励情况保存到 STUDENT 和 AWARD 表中,完成录入功能。

```
INSERT INTO STUDENT (Sno, Sname, Ssex, Sage, Major)  
VALUES('20100001', '史玉明', '女', 20, '计算机') /*插入学生的基本信息*/  
INSERT INTO AWARD (Sno, Details)  
VALUES('20100001', '2011 校奖学金, 2012 国家奖学金') /*插入学生获得的奖励*/
```

查询功能可以用一条查询语句实现:

```
SELECT A.Sno, Sname, Ssex, Sage, Major, Details /*查询学号为 20100001 学生的信息*/  
FROM STUDENT A LEFT JOIN AWARD B ON A.Sno=B.Sno  
WHERE A.Sno = '20100001'
```

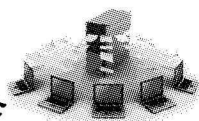
可以看出,无论采用文件系统还是数据库系统都可以实现本例中学籍管理的功能。但是,使用文件系统时,程序员要关注记录的结构和不同文件中记录之间的联系,使用文件系统提供的 fopen(打开)、fread(读)、fwrite(写)、fseek(移动读写位置)、fclose(关闭)等操作来编程,工作量大、编程复杂,且开发速度慢;而数据库系统提供了功能强大的操作,如查询操作只需要写一条语句就可以实现,程序员的开发效率大大提高。

与人工管理和文件系统相比,数据库系统的特点主要有以下几个方面。

1. 数据结构化

数据库系统实现整体数据的结构化,这是数据库的主要特征之一,也是数据库系统与文件系统的本质区别。

在文件系统中,文件中的记录内部具有结构,但是记录的结构和记录之间的联系被固



化在程序中,需要由程序员加以维护。这种工作模式既加重了程序员的负担,又不利于结构的变动。

所谓“整体”结构化是指数据库中的数据不再仅仅针对某一个应用,而是面向整个组织或企业;不仅数据内部是结构化的,而且整体也是结构化的,数据之间是具有联系的。也就是说,不仅要考虑某个应用的数据结构,还要考虑整个组织的数据结构。例如,一个学校的信息系统中不仅要考虑教务处的课程管理、学生选课管理、成绩管理,还要考虑学生处的学生学籍管理,同时还要考虑研究生院的研究生管理、人事处的教员人事管理、科研处的科研管理等。因此,学校信息系统中的学生数据就要面向各个处室的应用而不仅仅是教务处的一个学生选课应用。可以参照图1-6为该学校的信息系统组织其中的学生数据。

学生基本记录

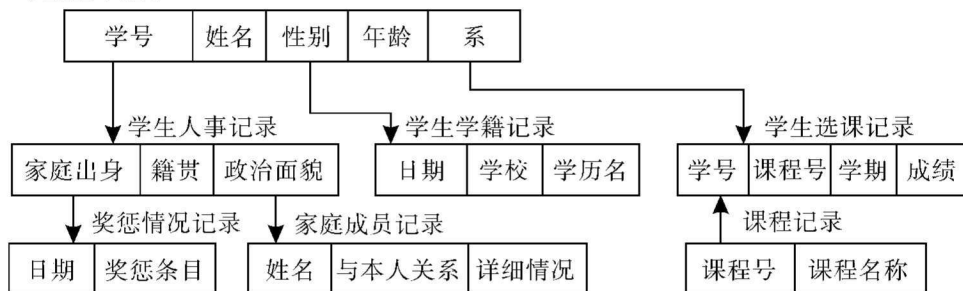


图1-6 某学校信息系统中的学生数据

图1-6的数据组织方式为各部门的应用提供了必要的记录,使整体数据结构化了。这就要求在描述数据时不仅要描述数据本身,还要描述数据之间的联系。在数据库系统中,记录的结构和记录之间的联系由数据库管理系统维护,从而减轻了程序员的工作量,提高了工作效率。

在数据库系统中,不仅数据是整体结构化的,而且存取数据的方式也很灵活,可以存取数据库中的某一个或一组数据项、一个或一组记录。而在文件系统中,数据的存取单位是记录,粒度不能细到数据项。

2. 数据的共享性高、冗余性低且易扩充

数据库系统从整体角度看待和描述数据,数据不再面向某个应用而是面向整个系统,因此数据可以被多个用户、多个应用共享使用。数据共享可以大大减少数据冗余,节约存储空间。数据共享还能够避免数据之间的不相容性与不一致性。

所谓数据的不一致性是指同一数据不同副本的值不一样。采用人工管理或文件系统管理时,由于数据被重复存储,当不同的应用使用和修改不同的副本时就很容易造成数据的不一致。在数据库中数据共享,减少了由于数据冗余造成的不一致现象。

由于数据面向整个系统,是有结构的数据,不仅可以被多个应用共享使用,而且容易增加新的应用,这就使得数据库系统弹性大,易于扩充,可以适应各种用户的要求。可以选取整体数据的各种子集用于不同的应用系统,当应用需求改变或增加时,只要重新选取不同的子集或加上一部分数据,便可以满足新的需求。

3. 数据独立性高

数据独立性是借助数据库管理数据的一个显著优点,它已成为数据库领域中一个常用