



普通高等教育“十三五”规划教材

数据库原理

(第四版)

张红娟 傅婷婷 编著
郭盈发 主审

SHUJUKU YUANLI



西安电子科技大学出版社
<http://www.xduph.com>

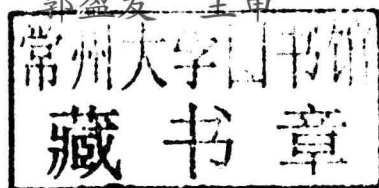
□ 普通高等教育“十三五”规划教材

数据库原理

(第四版)

张红娟 傅婷婷 编著

郭盈发 主审



西安电子科技大学出版社

内 容 简 介

本书系统地介绍了关系数据库的基本概念、基本原理和设计技术等,内容包括数据库系统概论、SQL 语言、关系数据模型及其运算基础、数据库管理、数据库设计等。本书根据数据库发展的趋势,在介绍数据库新技术时,重点介绍了面向对象和对象-关系数据库。

全书共 11 章,内容丰富,结构合理,通俗易懂。各章都安排有大量的例题,章末附有习题,书末还附有 SQL 语言的上机实验指导。

本书可作为大学计算机专业及其他专业的教材,也可作为科技人员学习数据库的自学教材和应用参考书。

图书在版编目(CIP)数据

数据库原理 / 张红娟, 傅婷婷编著. —4 版. —西安: 西安电子科技大学出版社, 2016.7

普通高等教育“十三五”规划教材

ISBN 978-7-5606-4156-0

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

中国版本图书馆 CIP 数据核字(2016)第 148201 号

策 划 马乐惠

责任编辑 南 景 马乐惠

出版发行 西安电子科技大学出版社(西安市太白南路 2 号)

电 话 (029)88242885 88201467 邮 编 710071

网 址 www.xduph.com 电子邮箱 xdupfb001@163.com

经 销 新华书店

印刷单位 陕西华沐印刷科技有限责任公司

版 次 2016 年 7 月第 4 版 2016 年 7 月第 13 次印刷

开 本 787 毫米×1092 毫米 1/16 印 张 19

字 数 447 千字

印 数 55 001~58 000 册

定 价 33.00 元

ISBN 978 - 7 - 5606 - 4156 - 0/TP

XDUP 4448004-13

如有印装问题可调换

本社图书封面为激光防伪覆膜, 谨防盗版。

第四版前言

本书自出版以来多次修订,受到了读者和同行的认可,综合各方面的意见,并结合数据库应用的最新发展,作者从去年开始就着手进行第四版的修订。

本版在组织结构上与第三版保持了一致。由于数据库技术发展迅速,此次修订将全书基于 SQL Server 2000 版本的内容修订为基于 SQL Server 2014 版本,主要涉及第 3 章 SQL 语言初步的部分内容,以及附录 A 上机实验指导的全部内容。随着近几年大数据和云计算的应用, NoSQL 数据库得到了飞速发展,所以在本版中增加了 NoSQL 的介绍。同时,本版在第 2 章、第 3 章、第 7 章中均新增了大量的实例,并增加了第 8 章习题中的设计实例。

本书的内容如下:

第 1 章首先介绍贯穿全书的示例:职工项目管理系统,引出了数据库基本概念以及数据库系统的构成、体系结构。

第 2 章是建立数据模型,介绍概念模型和关系数据模型中的概念,以及两者间的转换,最后介绍了 NoSQL 的概念和产品种类。

第 3、4、5 章集中介绍 SQL 语言及其编程,第 3 章介绍 SQL 的基本知识,SQL 的 DDL 语句和 DML 语句的使用,以及视图的概念及使用;第 4 章介绍完整性和安全性的概念及它们在 SQL 中的实现,触发器的概念和编程;第 5 章介绍数据库编程相关的技术,包括嵌入式 SQL、数据库接口 ODBC 和 JDBC、存储过程和用户自定义函数。本书中 SQL 的语句句法,都遵循 Microsoft SQL Server 的语法。

第 6 章介绍关系运算和查询优化,主要介绍关系代数的基本操作,并添加了包的运算和关系代数的扩展操作作为选修内容。本书对除法的介绍方法是比较独特的,既突出了除法的本质,又降低了除法的难度,希望这种方法能帮助读者更快地理解除法的本质。

第 7 章从关系的规范化作用和函数依赖,引出了关系模式的分解和范式的定义。

第 8 章介绍数据库设计的一般过程,从需求分析到概念结构设计、逻辑结构设计、物理结构设计、数据库实施和维护,最后介绍了 UML 设计方法和使用 PowerDesigner 工具实现数据库设计的方法,作为教师的选讲内容。

第 9 章介绍事务的概念,以及并发控制和数据库系统的故障及恢复技术。

第 10、11 章分别介绍了面向对象数据库技术和数据库新技术。

此次修订工作由张红娟完成，傅婷婷提供了安装 SQL Server 2014 的素材。郭盈发老师虽然年事已高，但仍一直关心本书的修订工作，在此向他表示深深的谢意。还要感谢金洁老师、赵丽娜老师以及数据库实验室的同学们给出了很好的修改意见。

由于作者水平有限和数据库技术发展迅猛，书中不足之处犹存，恳请读者批评指正。

作者邮箱：hjzhang@hdu.edu.cn。

编 者

于杭州电子科技大学

2016 年 4 月

第一版前言

数据库技术产生至今不到四十年，但它却已浸透到计算机应用的各个方面。数据库技术在理论上不断得到创新，在应用上则遍及人类社会的各个角落。在学校里，数据库技术不但早已成为计算机专业的重要课程，也已成为非计算机专业的必修课或选修课。学生怀着浓厚的兴趣，要求学习数据库的有关课程。

数据库技术高速发展的根本原因，在于它的应用性。应用上不断提出的迫切要求，促使数据库技术飞速发展。数据库的所有理论都是与应用紧密联系的，从应用中提出，得到提高，又立即返回到应用中去，指导应用不断向高深和普及两方面发展。在数据库学科中，没有永久不变的理论。理论完全依赖于应用的条件，应用条件发生变化，理论也随即跟着变化。因此，数据库是计算机中一个最活跃的分支，是一个发展速度最快的分支。

本书在编排上，特别强调摆正理论与应用的联系；在叙述上，则强调严谨与通俗紧密结合。全书共分十章。数据库概论中，介绍了数据库特别是关系数据库的基本知识、基本术语。接着，介绍 SQL 语言的基本内容，使学生能立即上机实践。既可使学生获得初步应用能力，又可从应用中发现问题的，从而提高学习兴趣。关系数据库模型及其运算基础的内容，从理论上给学生打下基础。接下来的三章：建立数据模型、关系数据库规范化理论和数据库设计，都是为了提高学生数据库设计能力的。而数据库保护一章则讲述了并发控制、恢复技术和安全控制。最后两章：面向对象和对象-关系数据库、数据库新技术，则介绍了数据库的发展方向，特别重点介绍了面向对象和对象-关系数据库。

本书还给出了 SQL 的上机实验指导，以便学生循序渐进地提高实际应用能力。

本书由郭盈发、张红娟共同编著而成。张红娟执笔了第三、五、六、七、九、十章，并设计了全书的习题和上机实验指导。郭盈发为主编，并执笔了其余内容。

由于水平有限，本书难免会存在缺点和错误，恳请读者批评指正。

编 者

于杭州电子工业学院计算机分院

2001 年 10 月

目 录

第 1 章 数据库系统概论.....1	2.2.8 子类和 Is-a 层次联系.....26
1.1 数据库系统的应用.....1	2.3 E-R 模型的设计方法.....27
1.1.1 数据库系统的应用领域.....1	2.4 数据模型.....32
1.1.2 数据库应用系统示例.....3	2.5 关系模型.....32
1.2 数据管理技术的发展.....4	2.5.1 关系模型数据结构及基本概念.....33
1.2.1 人工管理阶段.....5	2.5.2 关系模型的数据操作.....35
1.2.2 文件系统阶段.....5	2.5.3 关系的完整性约束.....35
1.2.3 数据库系统阶段.....6	2.6 E-R 模型向关系模型的转化.....36
1.3 数据库系统的构成.....8	*2.7 历史上有影响的数据模型.....39
1.3.1 数据库系统.....8	2.7.1 层次模型.....39
1.3.2 数据库管理系统.....9	2.7.2 网状模型.....40
1.3.3 数据库系统人员的构成.....10	2.8 数据模型与数据库系统的发展.....41
1.4 数据库系统的结构.....11	2.9 小结.....44
1.4.1 数据的三级模式结构.....11	习题 2.....45
1.4.2 数据独立性.....13	
1.4.3 SQL 中的三级模式结构.....14	第 3 章 SQL 语言初步.....48
1.4.4 面对用户的数据库系统体系结构.....15	3.1 SQL 简介.....48
1.4.5 用户访问数据库的过程.....16	3.2 基本的数据定义.....50
1.5 数据库系统研究概述.....17	3.2.1 创建基本表——CREATE TABLE.....50
1.6 小结.....18	3.2.2 表结构的修改——ALTER TABLE.....53
习题 1.....19	3.2.3 删除基本表——DROP TABLE.....54
	3.2.4 创建索引——CREATE INDEX.....54
第 2 章 建立数据模型.....20	3.2.5 删除索引——DROP INDEX.....55
2.1 现实世界的数字化过程.....20	3.3 基本的数据操作.....56
2.2 概念模型.....21	3.3.1 表中增加元组——INSERT.....56
2.2.1 概念模型的基本概念.....21	3.3.2 修改表中数据——UPDATE.....56
2.2.2 E-R 图的基本表示方法.....22	3.3.3 删除元组——DELETE.....57
2.2.3 联系.....23	3.3.4 更新操作与数据库的一致性.....57
2.2.4 多元联系.....24	3.4 数据查询——SELECT.....58
2.2.5 联系的属性.....25	3.4.1 无条件单关系查询.....58
2.2.6 自身联系.....25	3.4.2 带条件单关系查询.....61
2.2.7 弱实体集.....26	3.4.3 分组查询.....63

3.4.4 排序查询	64	5.2 数据库编程方法	107
3.4.5 多关系连接查询	65	5.3 嵌入式 SQL 的使用	108
3.4.6 嵌套查询	68	5.3.1 嵌入式 SQL 的一般形式	109
3.4.7 多个 SELECT 语句的集合操作	73	5.3.2 嵌入式 SQL 与宿主语言之间的 信息传递	109
3.5 含有子查询的数据更新	73	5.3.3 游标	112
3.5.1 INSERT 与子查询的结合	74	5.3.4 通过游标的更新和删除	114
3.5.2 UPDATE 与子查询的结合	74	5.3.5 动态 SQL 介绍	116
3.5.3 DELETE 与子查询的结合	75	5.4 数据库接口技术	117
3.6 视图	75	5.4.1 开放数据库互连——ODBC	117
3.6.1 定义视图——CREATE VIEW	75	5.4.2 ODBC 实例分析	119
3.6.2 删除视图——DROP VIEW	77	5.4.3 Java 数据库连接——JDBC	122
3.6.3 视图的查询	77	5.5 SQL 中的存储过程	123
3.6.4 视图的更新	78	5.5.1 存储过程的定义	124
3.6.5 视图的作用	80	5.5.2 存储过程的执行	126
3.7 小结	80	5.6 SQL 中的用户定义函数	127
习 题 3	81	5.6.1 函数的定义	127
第 4 章 完整性和安全性	85	5.6.2 函数的调用和执行	130
4.1 完整性约束的 SQL 定义	85	5.6.3 函数与存储过程	131
4.1.1 实体完整性约束和主码	85	5.7 小结	131
4.1.2 参照完整性约束和外部码	86	习题 5	132
4.1.3 用户自定义完整性约束	88	第 6 章 关系数据模型及其运算基础	134
4.1.4 约束的更新	90	6.1 关系模型的基本概念	134
4.2 SQL 中的触发器	90	6.2 关系模式	136
4.2.1 触发器的组成和类型	91	6.3 关系代数	136
4.2.2 创建触发器	91	6.3.1 基于传统集合运算的关系运算	137
4.2.3 触发器创建实例	92	6.3.2 投影(Projection)	138
4.3 数据库安全	94	6.3.3 选择(Selection)	139
4.3.1 用户标识和鉴别	94	6.3.4 连接(Join)	139
4.3.2 访问控制	95	6.3.5 更名(Rename)	142
4.3.3 强制存取控制方法	99	6.3.6 除(Division)	142
4.3.4 视图和查询修改	100	6.3.7 综合运算举例	144
4.3.5 跟踪审计	100	*6.4 包	145
4.3.6 数据加密	101	*6.5 扩展关系代数	146
4.4 小结	101	6.5.1 消除重复	146
习题 4	102	6.5.2 聚集运算和分组运算	146
第 5 章 数据库编程	104	6.5.3 排序运算	147
5.1 客户机/服务器体系结构	104	6.5.4 广义投影运算	148

*6.6 关系演算	149	8.1 数据库设计简介	189
6.6.1 元组关系演算	149	8.1.1 数据库设计的一般策略	189
6.6.2 域关系演算	150	8.1.2 数据库设计的步骤	189
6.7 查询优化	151	8.1.3 数据库设计的主流方法	191
6.7.1 查询处理概述	152	8.2 需求分析	192
6.7.2 关系代数等价变换规则	153	8.2.1 需求调查	192
6.7.3 查询优化的一般策略	154	8.2.2 需求分析的方法	193
6.7.4 关系代数表达式的优化算法	155	8.3 概念结构设计	196
*6.8 关系系统	157	8.4 逻辑结构设计	197
6.8.1 全关系系统的基本准则	157	8.4.1 模式评价	198
6.8.2 关系系统的定义	158	8.4.2 逻辑模式的修正	198
6.8.3 关系系统的分类	159	8.4.3 设计用户外模式	199
6.9 小结	160	8.5 物理结构设计	200
习题 6	161	8.5.1 存储记录结构设计	200
第 7 章 关系数据库规范化理论	163	8.5.2 存储记录布局	201
7.1 关系规范化的作用	163	8.5.3 存取方法的设计	201
7.2 函数依赖	165	8.6 数据库的实施和维护	202
7.2.1 属性间的联系	165	8.6.1 数据库的实施	203
7.2.2 函数依赖	166	8.6.2 数据库的维护	205
7.2.3 码的定义	167	*8.7 UML 方法规范数据库设计	206
7.3 关系模式的规范化	168	8.7.1 UML 用于数据库设计	206
7.3.1 非规范化的关系	168	8.7.2 各种 UML 图	207
7.3.2 第一范式(1NF)	169	*8.8 使用 PowerDesigner 工具设计数据库	211
7.3.3 第二范式(2NF)	170	8.8.1 PowerDesigner 简介	212
7.3.4 第三范式(3NF)	171	8.8.2 概念数据模型	214
7.3.5 改进的 3NF——BCNF	171	8.8.3 物理数据模型	215
*7.4 多值依赖和第四范式	173	8.9 小结	216
7.4.1 多值依赖(Multivalued Dependency)	173	习题 8	217
7.4.2 第四范式(4NF)	176	第 9 章 事务管理	220
7.5 关系的规范化程度	176	9.1 事务概述	220
*7.6 函数依赖公理系统	177	9.1.1 事务的概念	220
7.6.1 闭包及其计算	178	9.1.2 事务的特性	222
7.6.2 最小函数依赖集	180	9.2 并发控制技术	224
7.6.3 关系模式的分解	181	9.2.1 并发可能出现的问题	224
7.7 小结	184	9.2.2 封锁(Locking)	226
习题 7	185	9.2.3 三级封锁协议	226
第 8 章 数据库设计	188	9.2.4 加锁请求的选择策略和活锁	227
		9.2.5 死锁	228

9.2.6 并发调度的可串行性	229	11.1.2 分布式数据库系统的特点	253
9.2.7 两段封锁协议	230	11.1.3 分布式数据库系统的模式结构	254
9.2.8 多粒度封锁	231	11.2 并行数据库	255
9.2.9 意向锁	232	11.2.1 并行数据库系统的体系结构	256
9.3 数据库恢复技术	233	11.2.2 并行处理技术	257
9.3.1 故障的种类	233	11.3 工程数据库	258
9.3.2 故障恢复的手段	234	11.3.1 工程数据库的特点	258
9.3.3 故障恢复的方法	235	11.3.2 工程数据库的系统结构	259
9.4 小结	237	11.3.3 工程数据库的数据模型	260
习题 9	238	11.3.4 版本和版本管理	261
第 10 章 面向对象和对象-关系数据库	240	11.4 数据仓库	262
10.1 面向对象数据模型	240	11.4.1 数据仓库的定义与特征	262
10.1.1 对象	240	11.4.2 数据仓库系统	263
10.1.2 类和实例	241	11.4.3 数据仓库的数据库模式	264
10.1.3 类的继承	241	11.4.4 数据仓库的分析工具	266
10.2 面向对象数据库建模	242	11.5 小结	267
10.2.1 ODL 的类说明	242	习题 11	267
10.2.2 ODL 中属性的说明	242	附录 A 上机实验指导	268
10.2.3 ODL 中的联系	243	上机实验一 安装和了解 SQL Server 2014	268
10.2.4 ODL 中类的继承	244	上机实验二 创建 SQL Server 数据库和表	274
10.2.5 ODL 中方法的说明	245	上机实验三 基本表的建立和修改	279
10.2.6 E-R 模型向面向对象数据模型的 转换	246	上机实验四 SELECT 语句基本格式的使用	280
10.3 对象-关系数据库	246	上机实验五 SELECT 语句高级格式和完整 格式的使用	280
10.3.1 基本数据类型的扩充	247	上机实验六 SQL 的存储操作	281
10.3.2 支持复杂对象	248	上机实验七 视图的建立及操作	282
10.3.3 支持继承	249	上机实验八 完整性约束的实现	282
10.3.4 强大、通用的规则系统	250	*上机实验九 安全性的实现	283
10.4 小结	250	*上机实验十 创建存储过程和用户定义 函数	284
习题 10	251	*附录 B PowerDesigner 入门实验	285
第 11 章 数据库新技术	252	参考文献	294
11.1 分布式数据库	252		
11.1.1 分布式数据库系统的定义	253		

第 1 章 数据库系统概论

数据库技术是使用计算机管理数据的一门新技术，是计算机学科的重要分支。使用数据库对数据进行管理是计算机应用的一个重要而广阔的领域。

数据管理是数据处理的中心问题。数据处理是指对各种形式的数据进行收集、储存、加工和传播的一系列活动的总和。数据处理的目的是有两个：一是借助计算机科学地保存和管理大量复杂的数据，以便人们能方便而充分地利用这些宝贵的信息资源；二是从大量的原始数据中抽取、推导出对人们有价值的信息。数据管理指的是对数据的分类、组织、编码、储存、检索和维护。

数据库技术所研究的是如何科学地组织和储存数据，如何高效地处理数据以获取其内在信息。数据库系统是当代计算机系统的重要组成部分。

本章通过数据库系统的各个具体应用领域初步介绍了数据库系统，并给出了一个贯穿全书的数据库应用示例：职工项目管理系统。其中，回顾了数据管理技术发展的三个阶段，介绍了数据库系统的各个组成部分，讨论了数据库管理系统 DBMS 的三层模式体系结构和数据独立性，并从用户的角度考察一个 DBMS 所做的工作。最后，通过数据库系统研究的三个方面对本书的各章节内容作一整体介绍。

需要指出的是，本章内容均为基础知识，对全面正确认识数据库系统的特征及功能至关重要。但同时，本章讨论的内容范围较广，有些抽象，而且涉及到大量新的概念与术语。当然，在以后章节中将对相关的概念和术语进行详细介绍。

1.1 数据库系统的应用

1.1.1 数据库系统的应用领域

在生活中，我们经常使用数据库，只是应用系统的用户界面隐藏了访问数据库的细节，以至于多数人没有意识到数据库已成为生活中的基本组成部分。例如，当你从银行的 ATM 取款机上提取现金或查询账户余额时，这些信息就是从银行的数据库系统中取出来的；当你使用某个购物网站进行网上购物时，你的订单也就保存在该网站的数据库中；当你在线查询、下载或是播放音乐时，其实访问的就是存储在某个数据库中的视频或音频数据。

数据库的应用领域非常广泛，无论是家庭、企业、政府部门，都需要使用数据库来存储数据信息。传统数据库中的很大一部分用于商务领域，如证券机构、银行、企业的销售部门、医院、公司以及国家政府部门、国防军工领域、科技发展领域等。以下是数据库系

统一一些具有代表性的应用。

(1) 航空售票系统:是最先使用分布式数据库的行业之一,通过网络上分散的结点来访问中央数据库系统。这个系统主要用来存储和管理航班信息(如航班号、飞机型号、机组号、起飞地、目的地、起飞时间、到达时间、飞行状态等)、机票信息(如票价、折扣、是否有票等)和座位预定信息(如座位分配、座位确认、餐饮选择等)。

(2) 银行账务管理系统:用于存储客户的信息、账户、存贷款信息以及银行的交易记录等,同时也管理记录信用卡消费的情况和产生每月清单等。

(3) 学校学籍管理系统:用于存储学生姓名、课程名、所选课程及成绩等信息。

(4) 电子政务系统:将政府机构的经济管理、市场监管、社会管理和公共服务这四大职能电子化、网络化。

(5) 销售管理系统:用于管理客户、订货、库存、往来账款、产品信息、销售人员,以及收集与处理市场有关信息。

(6) 人力资源管理系统:包括人事日常事务、薪酬、招聘、培训、考核以及人力资源的管理。

(7) 企业资源规划(Enterprise Resource Planning, 简称 ERP):用来合并一个企业中的各种功能领域,包括产品生产、销售、分发、市场、财务和人力资源等。

(8) 电子商务系统:利用 Web 技术来传输和处理商业信息,在 Internet 上进行商务活动,其主要功能包括基于互联网的广告、订货、付款、客户服务和货物递交等销售、售前和售后服务,以及市场调查分析、财务核计及生产安排等多项利用 Internet 开发的商业活动。

数据库技术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。随着信息时代的发展,当今数据库系统是一个大家族,数据模型丰富多样,新技术层出不穷,应用领域日益广泛。以下是当前数据库的一些新的应用领域。

(1) 多媒体数据库。多媒体数据库主要存储与多媒体相关的数据,如声音、图像和视频等数据。多媒体数据库最大的特点是数据连续,而且数据量比较大,存储需要的空间较大。

(2) 移动数据库。移动数据库是在移动计算机系统上发展起来的,如笔记本电脑、掌上计算机等。该数据库最大的特点是通过无线数字通信网络传输数据。移动数据库可以随时随地获取和访问数据,为商务应用和一些紧急情况带来了很大的便利。

(3) 空间数据库。空间数据库目前发展比较迅速。它主要包括地理信息数据库(又称为地理信息系统,即 Geographic Information System,简称 GIS)和计算机辅助设计(Computer-Aided Design,简称 CAD)数据库。其中地理信息数据库一般存储与地图相关的信息数据;计算机辅助设计数据库一般存储设计信息的空间数据,如机械、集成电路以及电子设备设计图等。

(4) 网格数据库。网格数据库是把整个网络整合成一个虚拟的、巨大的超级计算环境,实现计算资源、存储资源、数据资源、信息资源、知识资源、专家资源等的全面共享。其目的是解决多机构虚拟组织中的资源共享和协同工作问题。

(5) 传感器数据库。传感器网络由大量的低成本设备组成,用来测量诸如目标位置、环境温度和湿度等数据。每个设备都是一个数据源,将会提供重要的数据,这就产生了新的数据管理需求,同时也要求传感器数据库系统解决好传感器设备特有的移动性、分散性、动态性和传感器资源的有限性等特点所带来的许多新问题。

(6) 信息检索系统。信息检索就是根据用户输入的信息，从数据库中查找相关的文档或信息，并把查找的信息反馈给用户。信息检索领域和数据库是同步发展的，它是一种典型的联机文档管理系统或者联机图书目录。

(7) 分布式信息检索数据库。分布式信息检索数据库是随着 Internet 的发展而产生的数据库。它一般用于因特网及远距离计算机网络系统中。特别是随着电子商务的发展，这类数据库发展更加迅猛。许多网络用户(如个人、公司或企业等)在自己的计算机中存储信息，同时希望通过网络使用发送电子邮件、文件传输、远程登录方式和别人共享这些信息。分布式信息检索便满足了这一要求。

(8) 专家决策系统。专家决策系统也是数据库应用的一部分。由于越来越多的数据可以联机获取，特别是企业通过这些数据可以对企业的发展作出更好的决策，以使企业更好地运行，因此人工智能的发展，使得专家决策系统的应用更加广泛。

1.1.2 数据库应用系统示例

图 1-1 给出了一个有关职工项目管理系统数据库的简单示例。

Employee						
Eno	Ename	Sex	Age	Is_Marry	Title	Dno
1002	胡一民	男	38	1	工程师	01
1004	王爱民	男	60	1	高工	03
1005	张小华	女	50	1	工程师	02

Item					
Ino	Iname	Start_date	End_date	Outlay	Check_date
200801	硬盘伺服系统的改进	3/1/2008	2/28/2009	10.0	2/10/2009
200802	巨磁阻磁头研究	6/1/2008	5/30/2009	6.5	5/20/2009
201001	磁流体轴承改进	4/1/2009	2/1/2010	4.8	3/18/2010

Item_Emp		
Ino	Eno	IENo
200801	1004	1
200801	1016	2
200802	1002	1

Salary						
Eno	Basepay	Service	Price	Rest	Insure	Fund
1002	685.00	1300.00	85.00	488.40	18.80	630.50
1004	728.34	3500.00	85.00	580.00	21.00	800.50
1005	685.00	2500.00	85.00	512.00	18.80	700.50

Department			
Dno	Name	Phone	Manager
01	技术科	88090121	1002
02	设计所	88090122	1010
03	车间	88090123	1004

图 1-1 职工项目管理系统数据库

应用该数据库，可以查询到全公司职工和职工工资信息、公司正在进行或者已经完成的项目信息，以及职工参加项目的情况，并对这些相关信息进行维护。图 1-1 是该应用系统所涉及的部分数据信息。

图 1-1 所示数据库由 5 个表组成，每个表都存储了同一记录结构的数据。**Employee** 表存储了每个职工的相关信息，包括职工的职工号、姓名、性别、年龄、婚否、职称和所在部门号；**Item** 表存储了有关项目的基本信息，包括项目编号、项目名称、起始日期、终止日期、项目经费和验收日期；**Item_Emp** 表存储了有关职工参与项目的信息，包括项目编号、参加该项目的职工编号、职工在该项目中的排名；**Salary** 表存储了有关职工的薪酬信息，包括职工号、基本工资、津贴、物价补贴、养老保险、医疗保险和公积金；**Department** 表存储了有关部门的基本信息，包括部门号、部门名称、联系电话和部门经理。

要定义这个数据库，必须指定每个表的记录结构，并指定各个数据元素的数据类型。为了构建这个职工项目管理系统数据库，必须把每个职工、项目、职工参与项目情况等信息都以记录的形式存储在适当的表中。这里需要注意的是，不同表中的记录可能是相关的。比如，**Employee** 表中的部门号 **Dno** 与 **Department** 表中所存储的部门号 **Dno** 有关。

该职工项目管理系统数据库构建好以后，接下来，就要考虑如何对数据进行存储操作以及读取操作。比如，需要做的查询操作有：查询所有参加了 200801 项目的职工号、姓名和职称；查询职工胡一民参加的所有项目情况。又如，可能要执行的数据更新操作有：将参加 200802 项目的职工的津贴值都增加 200；删除所有张小华参与项目的相关信息等。这些都是职工项目管理系统的任务。

这个示例系统将贯穿本书的大多数章节。要实现这样一个管理系统，首先需要从需求分析阶段开始为该系统设计一个新的数据库。在本书的第 2 章着重介绍了概念设计阶段的成果——实体-联系(Entity-Relationship, 简称 ER)模型，接着实现逻辑设计，将 ER 模型转换成用商业数据库管理系统 DBMS 实现的数据模型(即关系模型)来表达，并利用第 7 章所讲解的规范化理论改进该数据库设计，最后对该关系模型进行物理设计，为存储和访问数据库提供进一步的详细规范说明。整个数据库设计过程将在第 8 章中借助于某种具体的计算机辅助软件工程(Computer-Aided Software Engineering, 简称 CASE)工具来整体实现。在第 3、4、5 章中，通过具体的数据库应用程序设计实现整个系统所预设的各个功能，包括从数据存储、数据查询到一些特定的业务逻辑功能的实现，以及保证数据的正确有效和安全性。

1.2 数据管理技术的发展

数据库技术是应数据管理任务的需要而产生的。所谓数据管理是指对数据进行分类、组织、编码、存储、检索和维护，是数据处理的中心问题。而数据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。

随着计算机硬件、软件的发展，计算机数据管理方法是不断发展的，至今经历了三个阶段：人工管理阶段、文件系统阶段和数据库系统阶段。

1.2.1 人工管理阶段

20 世纪 50 年代中期以前, 计算机本身的水平较低。在硬件方面, 计算机的运算速度低、内存容量小, 外存还没有磁盘等直接存取的存储设备; 在软件方面, 还没有操作系统, 没有管理数据的软件。这一阶段的计算机主要用于科学计算。这个阶段数据管理的特点是:

- (1) 数据不保存。需要时把数据输入, 用完就撤走。数据不保存在计算机中。
- (2) 没有管理数据的软件系统。应用程序不仅要管理数据的逻辑结构, 还要设计其物理结构、存取方法、输入/输出方法等。当存储改变时, 程序中存取数据的子程序就需随之改变, 即数据和程序不具有独立性。
- (3) 基本上没有文件概念。数据的组织方式必须由程序员自行设计。
- (4) 数据是面向应用的。一组数据只对应于一个应用程序。即使两个应用程序都涉及到了某些相同数据, 也必须各自定义, 无法相互利用, 不仅在程序之间有大量重复数据, 还易产生数据的不一致性(相同数据在不同程序中出现的值不同)。

在人工管理阶段, 上述数据与程序关系的特点如图 1-2 所示。

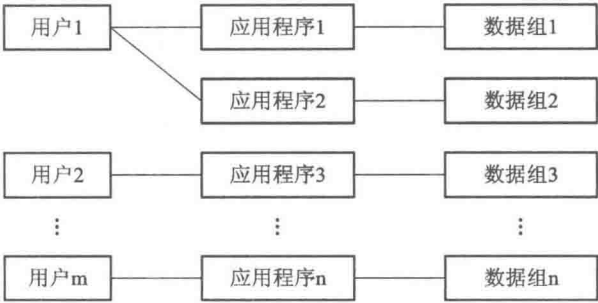


图 1-2 人工管理阶段数据与程序的关系

1.2.2 文件系统阶段

从 20 世纪 50 年代后期到 60 年代中期, 计算机的硬件、软件都有了很大发展: 有了磁盘、磁鼓等直接存取的存储设备; 有了操作系统, 其中有专门管理数据的文件系统; 在处理方式上, 不仅有了文件批处理, 而且能够联机实时处理。此时, 计算机不仅用于科学计算, 还大量用于管理。文件系统阶段有以下几个特点:

- (1) 数据存放在数据文件中, 数据文件可长期保存在外存, 需要时, 可方便地把数据文件调入内存。因此, 可经常方便地对文件进行检索、修改、插入和删除等操作。
- (2) 数据文件由记录组成(记录由数据组成)。数据的存取基本上以记录为单位, 按文件名访问, 按记录进行存取。数据的逻辑结构保存在使用数据的应用程序中。应用程序取得记录后, 首先根据数据的逻辑结构把记录分解成有含义的数据, 然后才使用这些数据; 应用程序要存储数据时, 首先要把数据组成记录, 才能进一步把记录存入数据文件。
- (3) 程序和数据有了一定的独立性。由于有了管理数据的软件——文件管理系统, 文件的逻辑结构(记录的逻辑结构)与存储结构由系统进行了转换。文件在存储上的改变不一定反

映到程序上，既可大大节省维护程序的工作量，程序员也可不必过多地考虑物理细节，可把精力集中在算法上。

(4) 文件多样化。由于有了直接存取存储设备，也就有了索引文件、链接文件、直接存取文件等。

在文件系统阶段，用户虽有了一定的方便，但系统仍有很多缺点，这主要有：

(1) 数据冗余度大。由于数据的基本存取单位是记录，因此，程序员之间很难明白他人数据文件的逻辑结构。虽然在理论上，一个用户可通过文件管理系统访问很多数据文件，但实际上，一个数据文件只能对应于同一个程序员的一个或几个程序，不能共享，数据仍然是面向应用的。随着应用的增加，数据文件也同步增加，且数据很多都是相同的。数据冗余度大，不仅浪费存储空间，而且数据的修改和维护也较困难，容易造成数据的不一致性(同一数据在多处出现，但值不同)。

例如，在项目管理中，可能保存有每个职工参与项目的信息文件，统计职工参与项目的情况或者新增职工参与项目的信息都可能是其中的某一个应用，但财务处可能会关注职工的薪酬以及每个项目的经费等情况。尽管这两个应用都可能会涉及到职工的相关数据，但是每个用户却因为彼此需要的数据并不能从对方所管理的文件中得到，所以只能各自维护各自不同的文件，以及操作这些文件的程序。

(2) 数据和程序缺乏独立性。文件处理过程中，如果某一用户要完成一个特定的软件应用，那么必须将该应用中所需要的文件的定义和实现当作该应用编程的一部分，即文件是为特定程序服务的，改变数据的逻辑结构就必须修改程序。应用程序若有改变，就可能影响文件中数据的数据结构，因此，数据和程序缺乏独立性。

这样，文件系统仍然是一个不具有弹性的无结构的数据集合。文件之间是孤立的，不能反映现实世界中事物之间的内在联系。在文件系统阶段，数据与程序的关系如图 1-3 所示。

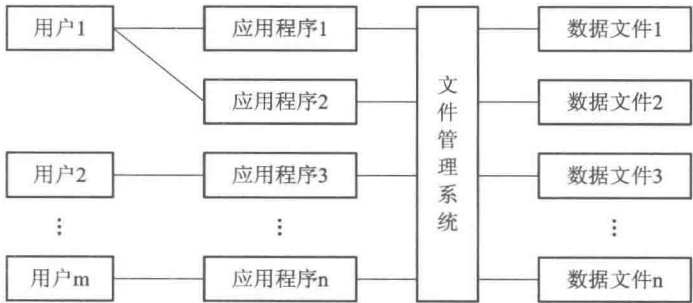


图 1-3 文件系统阶段数据与程序的关系

1.2.3 数据库系统阶段

20 世纪 60 年代后期，计算机硬件、软件有了更进一步的发展。计算机的运算速度快，内存容量大，并有了大容量磁盘。随着管理数据规模的增大，文件系统的缺点越来越令人难以忍受。人们迫切盼望能有数据冗余度小、可共享数据的系统。

文件系统中，数据为什么难以共享呢？原因就在于：数据的逻辑结构不在数据文件中(在

对应的应用程序中)。在文件系统中,采用“按文件名访问,按记录存取”的数据管理技术,一般用户虽然都可以访问文件、数据,但却不知道这些数据的含义,当然就不能共享数据了。以如下一条记录为例:

1005 张三 7006151800 设计所

一般用户很难理解其数据的逻辑结构,甚至连数据的创建者在经过一段时间后重新看到此记录时,也会感到不可理解了。但如果数据的显示形式为

职工号	姓名	出生日期	工资	部门
1005	张三	700615	1800	设计所

则数据的逻辑结构就比较清楚,共享自然就容易了。数据的逻辑结构包含:数据的含义(名称)、类型、数据本身的约束条件以及数据之间的关系(约束条件)等。

定义数据时,能简单又明确地定义数据的逻辑结构;存储数据时,能同时把数据的逻辑结构也存入数据文件,就成了改进文件系统的关键,从而出现了数据管理的新方式——数据库系统(DataBase System,简称 DBS)。在数据库系统中,数据以一种全新的方式——数据库方式存储;管理数据库生成、修改的是一种新的管理数据的软件——数据库管理系统(DataBase Management System,简称 DBMS)。

数据库(DataBase,简称 DB)是存储在计算机系统内的有结构的数据集合,是相关数据的集合,数据由数据库管理系统统一管理和维护。在此,数据(Data)指的是可记录的客观事实,并且有隐含的含义。例如,考虑某班级的全体同学(甚至是全校的学生)的姓名、出生日期、电话号码和家庭地址,把这些数据保存在一个带索引的通讯簿上,或者使用 Access、Excel 等软件保存到某个电子文档中,这就是带隐含含义的相关数据的集合,就是一个数据库。

在数据库中,数据与数据的含义(数据名称及说明)同时存储。数据的最小存取单位是构成记录的、有名称的、有含义的最小数据单位——数据项。定义数据库时,必须定义数据项的逻辑结构。使用数据库时,以数据项名存储数据、更新数据以及查询和使用数据。

所以,在数据库中,不仅包含数据本身,还包含了数据结构和约束的完整性定义或者描述。这些定义存储在数据库管理系统的目录(Catalog)中,称为数据库的元数据(Meta-Data)(也称数据字典),元数据描述了数据库的结构。任何合法用户都可在数据库管理系统的元数据的帮助下,利用数据项名方便地访问数据库中的数据及它们的逻辑定义,并使用这些数据(即使所用数据是其他人建立的,也不会增加使用的难度),亦即数据可高度共享。

由于数据可高度共享,因此,在数据库中,数据不再以各个应用程序各自的要求来分别存储,而是把整个系统所有的数据,根据它们之间固有的关系,分门别类地加以存储。也就是说,数据库中的数据不再是互相独立、毫无联系的,不再有有害的、不必要的冗余了。数据与应用程序互相独立,数据可为所有合法用户共享。

可见,数据库系统区别于传统文件处理系统的最重要特征就是引入了数据库这个概念,以及产生了专门用来实现和维护数据库而建立的通用软件——数据库管理系统。也就是说,数据库是存储在计算机系统内的有结构的数据的集合,数据是由数据库管理系统管理的。

数据库系统的出现,极大地推动了计算机数据管理事业,也极大地推动了计算机事业本身。可以毫不夸张地说,目前,计算机的任何应用都离不开数据库。数据库系统的水平已经成为国家实力的标志之一。