

普.通.高.等.学.校
计算机教育“十二五”规划教材

数据库 原理与应用

(第3版)

DATABASE CONCEPT AND
APPLICATION
(3rd edition)

赵杰 杨丽丽 陈雷 ◆ 编著



人民邮电出版社
POSTS & TELECOM PRESS

普.通.高.等.学.校
计算机教育“十二五”规划教材

数据库 原理与应用

(第3版)

*DATABASE CONCEPT AND
APPLICATION
(3rd edition)*

赵杰 杨丽丽 陈雷 ◆ 编著

人 民 邮 电 出 版 社
北 京

图书在版编目 (C I P) 数据

数据库原理与应用 / 赵杰, 杨丽丽, 陈雷编著. —
3版. — 北京: 人民邮电出版社, 2013. 2
普通高等学校计算机教育“十二五”规划教材
ISBN 978-7-115-29835-5

I. ①数… II. ①赵… ②杨… ③陈… 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(2012)第287128号

内 容 提 要

本书是在第2版的基础上, 根据教育部计算机等级考试 Access 科目考试大纲要求, 补充了应用部分的内容, 修订了相应的习题, 同时完善了教学案例, 使其更加符合教学要求。

本书全面、系统地介绍了数据库系统的基本概念、数学模型和关系理论, 通过可视化的实际操作, 讲解了数据库的建立、查询及标准 SQL 查询语言, 通过“课堂教学质量评价”这一广大师生熟悉和关心的功能需求入手, 实现了复杂数据库的设计、建立、查询和报表输出。

本书注重理论和实践的统一。每一章节列举了大量的实例, 各章后面都附有习题, 主要包括选择题、填空题、判断对错题、简答题、应用题和综合题等, 从各种不同的侧面进一步帮助读者了解和掌握知识点。

本书可作为大学计算机及相关专业的本、专科教材, 也可供从事软件开发工作的科技工作者和信息管理人员参考, 同时适合备考计算机等级考试 Access 科目的考生参考。

普通高等学校计算机教育“十二五”规划教材

数据库原理与应用 (第3版)

-
- ◆ 编 著 赵 杰 杨丽丽 陈 雷
责任编辑 滑 玉
 - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街14号
邮编 100061 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京隆昌伟业印刷有限公司印刷
 - ◆ 开本: 787×1092 1/16
印张: 17.75 2013年2月第3版
字数: 464千字 2013年2月北京第1次印刷

ISBN 978-7-115-29835-5

定价: 36.00 元

读者服务热线: (010)67170985 印装质量热线: (010)67129223
反盗版热线: (010)67171154

第3版序

2012年北京初秋的早晨，天气凉爽，阳光娇媚，经过一夏天的暑热，心情甚是舒畅。回想当年，应研究生导师殷光复老师之招，回到阔别5年的母校——中国农业大学（东校区），开始从事计算机科学与技术的研究和教学工作，特别是开始研究数据库原理与应用教学以来，成全了我十数年的事业发展，同时也验证了一句古话：“书中自有颜如玉，书中自有黄金屋”。

1996年的夏天，我接触了一个电力行业的报表项目，丰富的报表展现需求和复杂的业务数据逻辑关系，让我充分认识到，作为一名大学毕业生，掌握数据库原理，对工作是多么的重要。在相关领导的支持和帮助下，我面向全校研究生和计算机（辅修）专业开设了一门新课——数据库原理与应用。当时，我顶着很大的压力，摒弃了dBASE3和Fox Pro，我深信Access数据库，将来会有极大的发展空间。后来的发展，也证实了我的预见。

在教学研究的过程中，不断地接触到很多实际项目。特别是一些“硬骨头”，也就是别人做不下去的项目。经过研究和分析，绝大多数的问题，都是在初期数据结构设计的时候，违反了数据库设计的原则，在软件工程学中总结的“积累效应”、“传递效应”和“放大效应”的作用下，一发而不可收拾，最终陷入了“泥潭”。尽管相当多的软件项目投入了巨大的人力和财力，但仅仅因为最初的数据库设计的缺陷，不得不推翻重来，教训深刻。因此，“数据库原理与应用”这门课程，也就越来越受到各行各业的广泛重视。

不仅仅是软件开发人员需要掌握数据库原理，软件的使用者同样需要了解和掌握数据库原理。我在进行数据库原理与应用的教学和研究过程中，带了几名非计算机专业的本科学生。起初几乎所有的学生都对找工作发愁，当他们在我的鼓励下，认真严格地接受了数据库设计和开发的正规训练后，都出乎他们意料地找到了满意的工作。如今他们都成了公司的骨干，购房安家、娶妻生子，过上了安居乐业的生活，有的甚至做到了集团副总的职位。

如今，《数据库原理与应用》的教材，在书店里可以说是琳琅满目，sybase、DB2、Oracle、SQL Server、MySQL、Access等，比比皆是。但是最适合教学的，应用也最广泛的，Access还是数一数二的。恐怕也正是如此，每年参加教育部考试中心组织的计算机等级考试Access科目的同学逐年增加，2012年竟然达到了20万之多，这也完全出乎我的意料，也促使我决心完成本书第3版的修订。

为了配合教育部计算机等级考试大纲的要求，在完善了原理部分的同时，第3版教材特别增加了应用的内容，相应地增加和调整了习题。为了配合Office 2010版的推出，将书中的部分截图也相应地进行了调整，特别是将贯穿于本书的教学案例《课堂教学质量评价系统》进行了兼容性改造，使我们的教学案例涵盖了所有不同的Access版本。同时，我们推出了相应的教学网站，<http://Access.qiandao.org>。本书作者和编辑诚挚地希望广大师生，共同献计献策，不断传承，将这个网站办成优秀的学习和研究平台。

编者的联系方式分别如下：赵杰（106732718@QQ.com）、杨丽丽（llyang@cau.edu.cn）、陈雷（chen528@cau.edu.cn），感谢读者长期以来对本书及作者的厚爱，欢迎广大读者与我们联系交流，你们的意见和建议是我们成长的源动力。

编 者

2012年11月

目 录

第 1 章 数据库系统概述1	
1.1 引言.....1	
1.1.1 数据库是计算机技术发展的产物.....1	
1.1.2 数据库是计算机应用的基础.....1	
1.1.3 Access 数据库是中小型数据库的最佳选择.....1	
1.1.4 Access 数据库系统是企业级开发工具.....2	
1.2 数据库技术的发展.....2	
1.2.1 数据库发展阶段的划分.....2	
1.2.2 人工管理阶段.....2	
1.2.3 文件系统阶段.....3	
1.2.4 数据库系统阶段.....3	
1.2.5 高级数据库阶段.....6	
1.2.6 数据库应用的体系结构.....9	
1.3 数据模型.....9	
1.3.1 数据模型.....10	
1.3.2 概念模型.....10	
1.3.3 概念模型的表示方法-实体联系模型.....11	
1.3.4 几种常见的逻辑模型.....14	
1.4 数据库系统结构.....17	
1.4.1 数据库系统的三级模式结构.....17	
1.4.2 数据库系统的组成.....18	
1.5 数据库管理系统的功能及工作过程.....19	
1.5.1 数据库管理系统的主要功能.....19	
1.5.2 数据库系统的工作过程.....20	
1.5.3 数据库系统的不同视图.....21	
1.6 小结.....22	
习题.....22	
第 2 章 关系数据库数学模型27	
2.1 关系模型的基本概念.....27	
2.1.1 关系的数学定义.....27	
2.1.2 关系模型.....29	
2.2 EER 模型到关系模型的转换.....30	
2.2.1 实体类型的转换.....30	
2.2.2 二元关系的转换.....31	

2.2.3 实体内部之间联系的转换.....32	
2.2.4 三元关系的转换.....33	
2.2.5 子类型的转换.....33	
2.3 关系代数.....34	
2.3.1 基于传统集合理论的关系运算.....34	
2.3.2 专门的关系运算.....35	
2.4 关系演算.....39	
2.4.1 元组关系演算.....40	
2.4.2 域关系演算.....42	
2.5 小结.....43	
习题.....43	
第 3 章 关系数据库设计理论51	
3.1 问题的提出.....51	
3.2 函数依赖.....52	
3.2.1 数据依赖.....52	
3.2.2 函数依赖.....52	
3.2.3 函数依赖的逻辑蕴涵.....53	
3.2.4 键.....53	
3.3 关系模式的规范化.....54	
3.3.1 第一范式 (1NF).....55	
3.3.2 第二范式 (2NF).....56	
3.3.3 第三范式 (3NF).....57	
3.3.4 BCNF 范式.....57	
3.3.5 多值函数依赖.....58	
3.3.6 第四范式 (4NF).....61	
3.4 函数依赖的公理系统.....62	
3.4.1 Armstrong 公理.....62	
3.4.2 公理的正确性.....62	
3.4.3 公理的推论.....63	
3.5 模式分解.....63	
3.5.1 无损连接.....63	
3.5.2 保持函数依赖的分解.....65	
3.5.3 3NF 无损连接和保持函数依赖的分解算法.....66	
3.6 闭包及其计算*.....66	
3.7 函数依赖集的等价和覆盖*.....68	
3.8 公理的完备性*.....70	

3.9 小结	71	5.4 创建参数查询	110
习题	71	5.5 创建动作查询	112
第4章 Access 数据库设计	75	5.5.1 生成表查询	112
4.1 数据库设计的总体思路	75	5.5.2 删除查询	113
4.2 建立数据库	76	5.5.3 追加查询	114
4.3 数据表的建立	78	5.5.4 更新查询	115
4.4 使用数据表向导创建表	78	5.6 保存查询	117
4.5 通过输入数据建立新表	80	5.7 小结	117
4.6 使用表的“设计”视图	81	上机题	117
4.6.1 数据表设计窗体	83	习题	118
4.6.2 数据表窗体中的工具栏	83	第6章 使用高级查询——SQL	
4.6.3 建立和命名字段	83	语言	123
4.6.4 指定字段的数据类型	83	6.1 SQL 的数据定义	123
4.6.5 字段说明	85	6.1.1 基表	123
4.6.6 字段属性的设置	85	6.1.2 索引	124
4.6.7 定义主关键字	85	6.2 SQL 的数据操纵	125
4.6.8 为需要的字段建立索引	86	6.2.1 查询	125
4.7 更改数据表的结构	86	6.2.2 插入	130
4.7.1 移动字段的位置	86	6.2.3 删除	131
4.7.2 添加新字段	87	6.2.4 更新	131
4.7.3 删除字段	87	6.3 视图	132
4.7.4 更改字段名称	87	6.3.1 视图的建立和撤销	132
4.7.5 更改字段大小	87	6.3.2 视图的操作	133
4.7.6 更改字段的数据类型	88	6.3.3 视图的优点	133
4.8 保存完成的数据表	89	6.4 小结	134
4.9 数据表关联	89	上机题	134
4.9.1 如何定义表间的关系	89	习题	135
4.9.2 编辑已有的关系	91	第7章 窗体设计	141
4.9.3 删除关系	91	7.1 认识数据输入的窗体	141
4.9.4 查看已有的关系	91	7.1.1 窗体的结构	141
4.9.5 设置参照完整性	92	7.1.2 窗体的种类	142
4.9.6 联系类型	92	7.2 窗体的用途	143
4.10 小结	92	7.3 建立窗体	143
上机题	93	7.4 窗体窗口介绍	144
习题	94	7.4.1 窗体的工具栏	144
第5章 数据查询	104	7.4.2 窗体弹出式菜单	144
5.1 认识查询	104	7.4.3 “设计视图”窗体中的工具箱	145
5.1.1 查询的目的	104	7.5 窗体属性、控件属性及节的属性	146
5.1.2 查询的种类	104	7.6 将图片或其他 OLE 对象添加到	
5.1.3 查询准则	105	窗体中	147
5.2 创建选择查询	106	7.7 定制窗体的外观	147
5.3 创建交叉表查询	108		

7.7.1 文本	147	8.6.4 在报表上绘制线条	169
7.7.2 设置窗体属性	147	8.6.5 在报表上绘制矩形	170
7.7.3 在窗体中添加当前日期和时间	148	8.7 制作图表	170
7.7.4 设置控件属性	148	8.8 创建子报表	171
7.8 使用计算性表达式	149	8.8.1 子报表的定义和作用	171
7.9 在窗体中使用宏	150	8.8.2 在已有报表中创建子报表	172
7.10 设置值	150	8.8.3 将某个已有报表添加到其他 已有报表来创建子报表	173
7.10.1 改变当前的控件焦点、 页号和记录号	151	8.8.4 链接主报表和子报表	173
7.10.2 查找记录	151	8.8.5 更改子报表的布局	174
7.10.3 显示提示信息	151	8.9 创建多列报表	174
7.10.4 打印窗体	151	8.10 设计复杂的报表	175
7.10.5 利用宏来引入和导出数据	151	8.10.1 报表属性	175
7.10.6 设计菜单	152	8.10.2 节属性	176
7.11 创建与使用主/子窗体	152	8.10.3 使用选项组显示选项	176
7.11.1 同时创建主窗体和子窗体	153	8.10.4 给报表添加分页符	176
7.11.2 创建子窗体并将其添加到 已有的窗体中	154	8.11 打印预览报表	177
7.12 小结	156	8.12 小结	177
上机题	156	上机题	177
习题	157	习题	177
第 8 章 建立和打印报表	162	第 9 章 Web 页	180
8.1 认识报表的用途	162	9.1 概述	180
8.2 报表的结构组成	162	9.2 使用向导创建数据访问页	180
8.3 报表设计区	163	9.3 使用设计器创建和修改数据访问页	183
8.3.1 报表页眉	163	9.3.1 使用设计器创建数据访问页	183
8.3.2 页面页眉	164	9.3.2 使用“设计”视图修改数据 访问页	186
8.3.3 群组页眉	164	9.4 小结	187
8.3.4 主体	164	上机题	187
8.3.5 群组页脚	165	习题	187
8.4 报表的分类	165	第 10 章 宏操作	189
8.4.1 群组/合计报表	165	10.1 宏的概念	189
8.4.2 纵栏式报表	166	10.2 宏的分类	190
8.4.3 邮件标签	166	10.3 宏窗体中的工具栏	191
8.5 认识报表窗体	166	10.4 宏动作	192
8.5.1 预览窗体中的工具栏	167	10.5 创建宏	193
8.5.2 报表设计窗体中工具栏和 工具箱	167	10.6 在宏中设置操作参数的提示	194
8.6 设计报表	167	10.7 创建宏组	195
8.6.1 用预定义格式来设置报表 的格式	167	10.8 宏的条件表达式	195
8.6.2 添加分页符和页码	168	10.9 执行宏	196
8.6.3 在报表上使用节	169	10.10 调试宏	196

10.11 宏的触发	198	11.12.1 运算符	219
10.12 小结	198	11.12.2 表达式和优先级	220
上机题	198	11.13 程序控制流程	221
习题	200	11.13.1 以 GoTo 转移程序控制	221
第 11 章 面向对象程序设计		11.13.2 条件语句	221
语言——VBA	202	11.13.3 循环	223
11.1 VBA 概述	202	11.14 处理执行时的错误	225
11.1.1 为什么要使用 VBA	202	11.15 VBA 的基本操作	226
11.1.2 VBA 简介	202	11.15.1 打开和关闭窗体和报表对象	226
11.1.3 VBA 与 xBase、PAL 和 Visual Basic 的比较	202	11.15.2 信息输入和消息输出	228
11.1.4 如何使用 Visual Basic 程序代码	203	11.15.3 数据验证	230
11.2 模块的定义和使用	203	11.15.4 计时事件 Timer 和动画处理	230
11.2.1 模块的概念	203	11.15.5 鼠标和键盘事件处理	231
11.2.2 模块的定义和使用	206	11.15.6 文件操作	232
11.3 VBA 程序设计概念和书写原则	207	11.16 简单数据库编程	238
11.3.1 集合和对象	207	11.16.1 数据库编程技术简介	238
11.3.2 属性和方法	207	11.16.2 数据库编程技术分析	240
11.3.3 事件和事件过程	208	11.16.3 Docmd.RunSQL 方法运用	248
11.3.4 程序语句书写原则	208	11.17 小结	250
11.4 Visual Basic 的数据类型与数据库对象	208	上机题	250
11.5 变量	210	习题	251
11.5.1 隐含型变量	210	第 12 章 复杂数据库设计	261
11.5.2 显式的变量	210	12.1 数据库设计概述	261
11.5.3 变量的作用域与生命周期	210	12.1.1 数据库设计的内容和要求	261
11.5.4 用户定义的数据类型	211	12.1.2 数据库设计过程	263
11.5.5 Visual Basic 数组	211	12.2 需求分析	265
11.6 将数据库对象命名为 VBA 程序代码中的变量	212	12.2.1 确认用户需求, 确定设计范围	265
11.7 变量命名的法则	213	12.2.2 收集和分析需求数据	265
11.8 符号常量	213	12.2.3 需求说明书	266
11.9 Access 系统定义的常量	214	12.3 概念设计	266
11.10 Access 固有常量	214	12.3.1 设计局部概念模式	267
11.11 Access 常用标准函数	214	12.3.2 设计全局概念模式	268
11.11.1 算术函数	214	12.4 逻辑设计	270
11.11.2 字符串函数	215	12.5 物理设计	271
11.11.3 日期/时间函数	216	12.6 数据库的建立和维护	272
11.11.4 类型转换函数	219	12.6.1 建立数据库	272
11.12 运算符和表达式	219	12.6.2 数据库的重组织和重构造	272
		12.7 小结	273
		习题	274

第 1 章

数据库系统概述

【本章提要】

本章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、数据库系统的功能和工作过程等，以使大家对数据库系统有一个概括的了解。准备参加程序员水平等级考试的读者，应特别注意本章介绍的基本概念。

1.1 引言

【提要】本节主要介绍数据库技术的重要性及 Access 数据库系统的优越之处。

1.1.1 数据库是计算机技术发展的产物

数据库技术是计算机科学技术的一个重要分支。从 20 世纪 50 年代中期开始，计算机应用由科学研究部门扩展到企业和行政部门，数据处理很快上升为计算机应用的一个重要方面。自 1968 年第一个商品化的数据管理系统（Information Management System, 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、办公自动化和计算机辅助设计等的主要软件工具之一。

1.1.2 数据库是计算机应用的基础

数据库技术研究如何科学地组织数据和存储数据，如何高效地检索数据和处理数据，以及如何既减少数据冗余，又能保障数据安全，实现数据共享。在计算机应用的领域中，管理信息系统方面的应用占 90% 以上，而数据库技术又是管理信息系统的基础。因此，可以说数据库技术是计算机的重要应用之一。

1.1.3 Access 数据库是中小型数据库的最佳选择

Access 数据库系统是在 Windows 环境下开发的一种全新的关系型数据库系统。它具有大型数据库的一些基本功能，支持事务处理功能，具有 Transaction、Commit、Rollback、Withdraw 等指令。Access 数据库系统支持数据库加密，具有用户组和多用户管理功能，可以设置用户组或用户的密码和权限。Access 数据库系统支持数据压缩、备份和恢复功能，能够保证数据的安全性。Access 数据库系统还具备级联修改和级联删除功能，能够严格保证数据的一致性。

1.1.4 Access 数据库系统是企业级开发工具

目前，世界上有许多软件开发公司以 Access 数据库系统为主要开发工具之一。Access 不仅是数据库管理系统，而且还是一个功能强大的开发工具。它提供了丰富和完善的可视化开发手段，引入了 VBA（Visual Basic for Application）面向对象的编程技术，可以设计出友好的用户界面。在 Access 数据库管理系统上开发应用程序，开发者可以直接将 Access 系统的界面改造成应用程序的用户界面，只需花费很小的代价，就能得到功能完善的应用软件。

1.2 数据库技术的发展

【提要】本节主要介绍数据库发展的四个阶段和各类数据库的基本特征。

1.2.1 数据库发展阶段的划分

数据处理的首要问题是数据管理。数据管理是指如何分类、组织、存储、检索及维护数据。自 1946 年世界上第一台计算机诞生以来，随着计算机硬件和软件的发展，数据管理技术不断更新、完善，数据库的发展经历了如下四个阶段：人工管理阶段、文件系统阶段、数据库系统阶段和高级数据库阶段。

1.2.2 人工管理阶段

1. 人工管理阶段的年代及特征

从 1946 年计算机诞生至 20 世纪 50 年代中期，计算机主要用于科学计算。计算机除硬件设备外没有任何软件可用，使用的外存只有磁带、卡片和纸带，没有磁盘等直接存取的设备。软件中只有汇编语言，没有操作系统，对数据的处理，完全由人工进行管理。

2. 人工数据管理的特点

(1) 数据不保存。一组数据对应于一个应用程序，应用程序与其处理的数据结合成一个整体。有时也把数据与应用程序分开，但这只是形式上的分开，数据的传输和使用完全取决于应用程序。在进行计算时，系统将应用程序与数据一起装入，用完后就将它们撤销，释放被占用的数据空间与程序空间。

(2) 没有软件对数据进行管理。应用程序的设计者不仅要考虑数据的逻辑结构，还要考虑存储结构、存取方法以及输入输出方式等。如果存储结构发生变化，程序中的取数子程序也要发生变化，数据与程序不具有独立性。

(3) 没有文件概念。数据的组织方法由应用程序开发人员自行设计和安排。

(4) 数据面向应用。一组数据对应一个程序。即使两个应用程序使用相同的数据，也必须各自定义数据的存储和存取方式，不能共享相同的数据定义，因此程序与程序之间可能会有大量的重复数据。

3. 人工数据管理的模型

人工数据管理的模型如图 1-1 所示。

图 1-1 (a) 说明数据和程序是一体的，即数据置于程序内部；图 1-1 (b) 说明数据和程序是一一对应的，即一组数据只能用于一个程序。

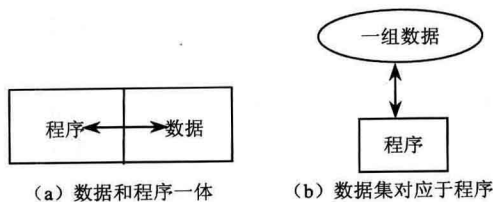


图 1-1 数据的人工管理模型

1.2.3 文件系统阶段

1. 文件系统阶段的年代及特征

20 世纪 50 年代后期至 60 年代中期，计算机不仅用于科学计算，而且还大量用于管理。计算机的硬件中有了磁盘和磁鼓等直接存储设备；计算机软件中有了高级语言和操作系统。

2. 文件系统阶段数据管理的特点

文件系统阶段数据管理有以下四个特点。

- (1) 数据可长期保存在磁盘上。用户可使用程序经常对文件进行查询、修改、插入或删除等操作。
- (2) 文件系统提供数据与程序之间的存取方法。文件管理系统是应用程序与数据文件之间的一个接口。应用程序通过文件管理系统建立和存储文件；反之，应用程序要存取文件中的数据，必须通过文件管理系统来实现。用户不必关心数据的物理位置，程序与数据之间有了一定的独立性。
- (3) 文件的形式多样化。因为有了直接存取设备，所以可建立索引文件、链接文件和直接存取文件等。对文件的记录可顺序访问、随机访问。文件之间是相互独立的，文件与文件之间的联系要用程序来实现。
- (4) 数据的存取以记录为单位。

3. 文件系统的模型

文件系统的模型如图 1-2 所示。通过文件管理系统，程序和数据文件之间可以组合，即一个程序可以使用多个数据文件，多个程序也可以共享同一个数据文件。

4. 文件系统的缺陷

文件管理系统的使用，使得应用程序按规定的组织方式建立文件并按规定的存取方法使用文件，不必过多地考虑数据物理存储方面的问题。但是文件管理下的数据仍然是无结构的信息集合，它可以反映现实世界中客观存在的事物，但不能反映出各事物之间客观存在的本质联系。文件系统有三大缺陷。

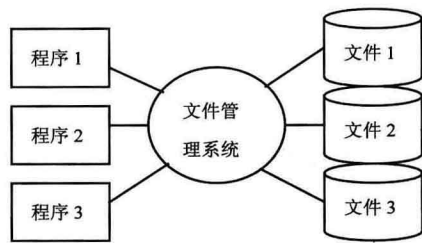


图 1-2 文件系统的模型

(1) 数据冗余。因为文件之间缺乏联系，可能有同样的数据在多个文件中重复存储。

(2) 不一致性。由于数据冗余，在对数据进行修改时，若不小心，同样的数据在不同的文件中可能不一样。

(3) 数据联系弱。这是文件之间缺乏联系造成的。

1.2.4 数据库系统阶段

1. 数据库系统阶段的年代和特征

从 20 世纪 60 年代后期开始，存储技术取得很大发展，有了大容量的磁盘。计算机用于管理

的规模更加庞大，数据量急剧增长，为了提高效率，人们着手开发和研制更加有效的数据管理模式，提出了数据库的概念。

美国 IBM 公司 1968 年研制成功的数据库管理系统（Information Management System, IMS）标志着数据管理技术进入了数据库系统阶段。IMS 为层次模型数据库。在 1969 年美国数据系统语言协会（Conference On Data System Language, CODASYL）公布了数据库工作组（Data Base Task Group, DBTG）报告，对研制开发网状数据库系统起了重大推动作用。从 1970 年起，IBM 公司的 E.F.Codd 连续发表论文，又奠定了关系数据库的理论基础。

从 20 世纪 70 年代以来数据库技术发展很快，得到了广泛的应用，已成为计算机科学技术的一个重要分支。

2. 数据库系统的特点

数据库系统与文件系统相比，克服了文件系统的缺陷。用数据库技术管理数据，主要有以下的特点。

（1）数据库中的数据是结构化的。在文件系统中，从整体上来看，数据是无结构的，即不同文件中的记录型之间没有联系，它仅关心数据项之间的联系。数据库系统不仅考虑数据项之间的联系，还要考虑记录型之间的联系，这种联系是通过存储路径来实现的。例如在学生选课情况的管理中，一个学生可以选修多门课，一门课可被多个学生选修。可用三种记录型（学生的基本情况、课程的基本情况以及选课的基本情况）来进行这种管理，如图 1-3 所示。

在查询“张三的学习成绩及学分”时，如果用文件系统实现学生选课的管理，程序员要编程，从 3 个文件中查找出所需的信息；如果用数据库系统管理学生选课，可通过存取路径来实现。利用存取路径，从一个记录型走到另一个记录型。事实上，学生记录、课程记录与选课记录有着密切的联系，存取路径表示了这种联系。这是数据库系统与文件系统的根本区别。

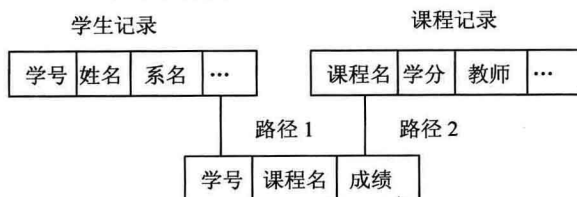


图 1-3 学生选课管理中的数据联系

（2）数据库中的数据是面向系统的，不是面向某个具体应用的，减少了数据冗余，实现了数据共享。数据库中的数据共享情况如图 1-4 所示。

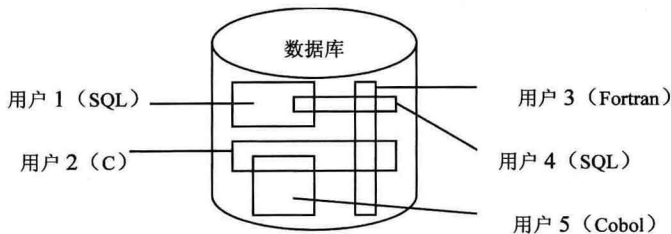


图 1-4 数据共享

（3）数据库系统比文件系统有较高的数据独立性。数据库系统的结构分为 3 级：用户（应用程序或终端用户）数据的逻辑结构、整体数据的逻辑结构（用户数据逻辑结构的最小并集）和数据的物理结构。当整体数据的逻辑结构或数据的物理结构发生变化时，应用不变。数据的独立性是通过数据库系统在数据的物理结构与整体结构的逻辑结构、整体数据的逻辑结构与用户的数据

逻辑结构之间提供的映像实现的。

例如在图 1-3 中, 根据需把课程记录中的字段“学分”移出, 加到选课记录中, 即课程记录中减少一个字段, 选课记录中增加一个字段。原来的应用不变, 仍然可用。

(4) 数据库系统为用户提供了方便的接口。用户可以用数据库系统提供的查询语言和交互式命令操纵数据库。用户也可以用高级语言(如 C、Fortran、Cobol 等)编写程序来操纵数据库, 拓宽了数据库的应用范围。

3. 数据库系统的控制功能

(1) 数据的完整性

保证数据库存储数据的正确性。例如, 预订同一班飞机的旅客不能超过飞机的定员数; 订购货物中, 订货日期不能大于发货日期。使用数据库系统提供的存取方法, 设计一些完整性规则, 对数据值之间的联系进行校验, 可以保证数据库中数据的正确性。

(2) 数据的安全性

并非每个应用都可以存取数据库中的全部数据。例如, 建立一个人事档案的数据库, 只有那些需要了解工资情况并且有一定权限的工作人员才能存取这些数据。数据的安全性是保护数据库不被非法使用, 防止数据的丢失和被盗。

(3) 并发控制

当多个用户同时存取、修改数据库中的数据时, 可能会发生相互干扰, 使数据库中的数据完整性受到破坏, 而导致数据的不一致性。数据库的并发控制防止了这种现象的发生, 提高了数据库的利用率。

(4) 数据库的恢复

任何系统都不可能永远正确无误地工作, 数据库系统也是如此。在运行过程中, 会出现硬件或软件的故障。数据库系统具有恢复能力, 能把数据库恢复到最近某个时刻的正确状态。

4. 数据库的定义

综合上面的叙述, 可为数据库下一个定义: 数据库是与应用彼此独立的以一定的组织方式存储在一起的彼此相互关联、具有较少冗余的能被多个用户共享的数据集合。

数据库技术的发展使数据管理上了新台阶, 几乎所有的信息管理系统都以数据库为核心, 数据库系统在计算机领域中的应用越来越广泛, 数据库系统本身也越来越完善。目前, 数据库系统已深入到人类生活的各个领域, 从企业管理、银行业务管理到情报检索、档案管理、普查、统计都离不开数据库管理。随着计算机应用的发展, 数据库系统也在不断更新、发展和完善。

5. 数据库系统的数据管理模型

数据库中数据的最小存取单位是数据项(文件系统的最小存取单位是记录)。应用(应用程序和终端用户)和数据库的联系如图 1-5 所示。其中, 数据库管理系统(DBMS)是一个软件系统, 它能够操纵数据库中的数据, 对数据库进行统一控制。

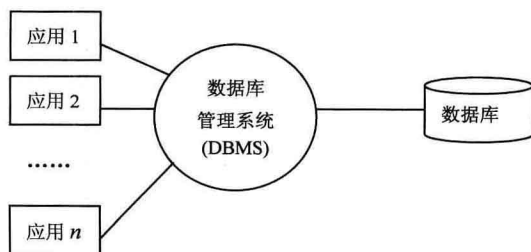


图 1-5 应用与数据库的联系

1.2.5 高级数据库阶段

1. 高级数据库阶段的年代及类别

20 世纪 70 年代中期以来,随着计算机技术的不断发展,出现了分布式数据库、面向对象数据库和智能型知识数据库等,通常被称为高级数据库技术。

计算机领域中其他新兴技术的发展对数据库技术产生了重大影响。传统的数据库技术和其他计算机技术的互相结合,建立和实现了一系列新型数据库系统,数据库技术与网络通信技术、并行计算技术、面向对象程序设计技术、人工智能技术等互相渗透,互相结合,成为当前数据库技术发展的主要特征,涌现出各种新型的数据库系统。例如,数据库技术与分布式处理技术相结合,形成了分布式数据库系统;数据库技术与并行处理技术相结合,形成了并行数据库系统;数据库技术与面向对象技术相结合,形成了面向对象数据库系统;数据库技术与多媒体技术相结合,形成了多媒体数据库系统;数据库技术与人工智能技术相结合,形成了知识库系统和主动数据库系统;数据库技术与模糊技术相结合,形成了模糊数据库系统等。

2. 并行数据库系统的特点

并行数据库系统 (Parallel Database System) 是在并行机上运行的具有并行处理能力的数据库系统。并行数据库系统是数据库技术与并行计算技术相结合的产物。并行计算技术利用多处理机并行处理产生的规模效益来提高系统的整体性能,为数据库系统提供了一个良好的硬件平台。并行数据库技术包括了对数据库的分区管理和并行查询。它通过将一个数据库任务分割成多个子任务的方法由多个处理机协同完成这个任务,从而极大地提高了事务处理能力,并且通过数据分区可以实现数据的并行 I/O 操作。DBMS 进程结构的最新发展为数据库的并行处理奠定了基础。多线程技术和虚拟服务器技术是并行数据库技术实现中采用的重要技术。一个理想的并行数据库系统应能充分利用硬件平台的并行性,采用多进程多线程的数据库结构,提供不同粒度 (Granularity) 的并行性、不同用户事务间的并行性、同一事务内不同查询间的并行性、同一查询内不同操作间的并行性和同一操作内的并行性。

一个并行数据库系统应该实现如下目标。

(1) 高性能:并行数据库系统通过将数据库管理技术与并行处理技术有机结合,发挥多处理机结构的优势,从而提供比相应的大型机系统要高得多的性能价格比和可用性。

(2) 高可用性:并行数据库系统可通过数据复制来增强数据库的可用性。

(3) 可扩充性:数据库系统的可扩充性指系统通过增加处理和存储能力而平滑地扩展性能的能力。

在国内,并行数据库系统的研究刚起步不久,有中国人民大学开发的基于曙光天演系列并行计算机的 PBASE/3 系统,另外联想集团正在投入巨资进行并行数据库服务器系统的开发和推广。

3. 分布式数据库系统的特点

分布式数据库 (Distributed Database System) 是由一组数据组成的,这组数据分布在计算机网络的不同计算机上,网络中的每个节点具有独立处理的能力 (称为场地自治),可以执行局部应用。同时,每个节点也能通过网络通信子系统执行全局应用。一个分布式数据库系统由一个逻辑数据库组成,这个逻辑数据库的数据存储在一个或多个节点的物理数据库上,通过两阶段提交 (2PC) 协议来提供透明的数据访问和事务管理。分布式数据库系统在系统结构上的真正含义是指物理上分布、逻辑上集中的分布式数据库结构。数据在物理上分布后,由系统统一管理,使用户不感到数据的分布。用户看到的似乎不是一个分布式数据库,而是一个数据模式为全局数据模式

的集中式数据库。

分布式数据库系统具有如下特点。

(1) 数据独立性：在分布式数据库系统中，数据独立性这不仅包括数据的逻辑独立性和物理独立性，还包括数据分布独立性，也称为分布透明性。

(2) 集中与自治相结合的控制结构：各局部的 DBMS 可以独立地管理局部数据库，具有自治的功能；同时，系统又设有集中控制机制，协调各局部 DBMS 的工作，执行全局应用。

(3) 适当增加数据冗余度：在不同的场地存储同一数据的多个副本，这样可以提高系统的可靠性和可用性，同时也能提高系统性能。

(4) 全局的一致性、可串行性和可恢复性。

比较常见的分布式数据库有 Sybase 数据库和 Oracle 数据库。分布式数据库系统现阶段主要用在银行、证券和政府部门的系统中。

4. 面向对象数据库系统

面向对象数据库系统 (Object Oriented Data Base System, OODBS) 是数据库技术与面向对象程序设计方法相结合的产物。面向对象数据库系统支持面向对象数据模型 (以下简称 OO 模型)。即面向对象数据库系统是一个持久的、可共享的对象库的存储和管理者，而一个对象库是由一个 OO 模型所定义的对象集合体。

对象-关系数据库系统就是将关系数据库系统与面向对象数据库系统两方面的特征相结合。对象-关系数据库系统除了具有原来关系数据库的各种特点外，还应该提供以下特点。

(1) 扩充数据类型，例如可以定义数组、向量、矩阵、集合等数据类型以及这些数据类型上的操作。

(2) 支持复杂对象，即由多种基本数据类型或用户自定义的数据类型构成的对象。

(3) 支持继承的概念。

(4) 提供通用的规则系统，大大增强对象-关系数据库的功能，使之具有主动数据库和知识库的特性。

近 10 年来，面向对象数据库系统一直是数据库学术界和工业界研究的热点之一。自从 1987 年以来，已陆续有多个 OODB 产品投入市场，其中某些产品已经占有一定的市场份额。总的说来，当前世界 OODB 市场只占整个数据库产品市场的很小一部分。作为数据库产品 OODB 还是不够成熟的，原因是缺乏某些数据库基本特性，例如完全非过程化的查询语言、视图、授权动态模式变化、参数化的性能调整等，这些都是数据库用户已经熟知的，因而希望提供的。此外，关系数据库产品还提供触发器、元数据管理、数据完整性约束等，而目前大多数的 OODB 产品不提供这样的支持。目前 OODB 产品的应用还很不普遍，主要应用在一些特殊的行业或特殊的应用领域中。

5. 数据仓库的特点

随着市场竞争的加剧和信息社会需求的发展，从大量数据中提取 (检索、查询等) 制订市场策略的信息就显得越来越重要了。这种需求既要求联机服务，又涉及大量用于决策的数据，而传统的数据库系统已无法满足这种需求。其具体体现在 3 个方面：(1) 历史数据量很大；(2) 辅助决策信息涉及许多部门的数据，而不同系统的数据难以集成；(3) 由于关系数据库访问数据的能力不足，它对海量数据的访问性能明显下降。

随着 C/S 技术的成熟和并行数据库的发展，信息处理技术的发展趋势是从大量的事务型数据库中抽取数据，并将其清理、转换为新的存储格式，即为了决策目标把数据聚合在一种特殊的格式中。随着这项技术的发展和完善，这种支持决策的、特殊的数据存储即被称为数据仓库 (Data

Warehouse, DW)。

数据仓库(Data Warehouse)是一个面向主题的(Subject Oriented)、集成的(Integrated)、相对稳定的(Non-Volatile)、反映历史变化(Time Variant)的数据集合,用于支持管理决策。这个定义中的数据是以下内容。

- (1) 面向主题的:因为仓库是围绕大的企业主题(如顾客、产品、销售量)而组织的。
- (2) 集成的:来自于不同数据源的面向应用的数据集成在数据仓库中。
- (3) 时变的:数据仓库的数据只在某些时间点或时间区间上是精确的、有效的。
- (4) 非易失的:数据仓库的数据不能被实时修改,只能由系统定期地进行刷新。刷新时将新数据补充进数据仓库,而不是用新数据代替旧的。

数据仓库是一个决策支撑环境,它从不同的数据源得到数据、组织数据、使用数据,有效地支持企业决策。一般来说,数据仓库的结构包括数据源、装载管理器、数据仓库管理器、查询管理器、详细数据、汇总数据、归档/备份数据、元数据和终端用户访问工具几大部分。

6. 多媒体数据库的特点

所谓多媒体数据库是指数据库中的信息不仅涉及各种数字、字符等格式化的表达形式,而且还包括多媒体的非格式化的表达形式,数据管理要涉及各种复杂对象的处理。多媒体是指多种媒体,如数字、文本、图形、图像和声音的有机集成,而不是简单的组合。其中数字、字符等称为格式化数据,文本、图形、图像、声音、视像等称为非格式化数据,非格式化数据具有数据量大、处理复杂等特点。多媒体数据库实现对格式化和非格式化的多媒体数据的存储、管理和查询,其主要特征如下。

(1) 能够表示多种媒体的数据。非格式化数据表示起来比较复杂,需要根据多媒体系统的特点来决定表示方法。如果感兴趣的是它的内部结构且主要是根据其内部特定成份来检索,则可将它按一定算法映射成包含它所有子部分的一张结构表,然后用格式化的表结构来表示它。如果感兴趣的是它本身的内容整体,要检索的也是它的整体,则可以用源数据文件来表示它,文件由文件名来标记和检索。

(2) 能够协调处理各种媒体数据。正确识别各种媒体数据之间在空间或时间上的关联。例如,关于乐器的多媒体数据包括乐器特性的描述和乐器的照片,利用该乐器演奏某段音乐的声音等,这些不同媒体数据之间存在着自然的关联,比如多媒体对象在表达时必须保证时间上的同步特性。

(3) 提供更强的适合非格式化数据查询的搜索功能。

(4) 多媒体数据库提供特种事务处理与版本管理能力。

7. 模糊数据库的特点

模糊数据库(Fuzzy Database)指能够处理模糊数据的数据库。一般的数据库都是以精确的数据工具为基础的,不能表示模糊不清的事情。随着模糊数学理论体系的建立,人们可以用数量来描述模糊事件并能进行模糊运算和模糊查询。这样就可以把不完全性、不确定性、模糊性引入数据库系统中,从而形成模糊数据库。目前,模糊数据库研究主要有两方面,首先是如何在数据库中存放模糊数据,其次是定义各种运算建立模糊数据上的函数。

8. 知识库系统的特点

人们对数据进行分析找出其中关系并形成信息,然后对信息进行再加工,获得更有用的信息,即知识。人工智能的发展,要求计算机不仅能够管理数据,还能管理知识。管理知识可用知识库系统实现。

知识库是一门新的学科,它研究知识表示、结构、存储和获取等技术。知识库是专家系统、

知识处理系统的重要组成部分。知识库系统把人工智能的知识获取技术和机器学习的理论引入到数据库系统中,通过抽取隐含在数据库实体间的逻辑蕴涵关系和隐含在应用中的数据操纵之间的因果联系,形式化地描述数据库中的实体联系。在知识库系统中可以把语义知识自动提供给推理机,从已有的事实知识推出新的事实知识。

1.2.6 数据库应用的体系结构

在软件体系架构设计中,分层式结构是最常见也是最重要的一种结构。基于数据库和网络应用的数据库应用系统实现模式有多种,可以采用传统的客户机/服务器(C/S)架构,也可以采用目前流行的基于 WEB 的方式。

所谓客户机/服务器(C/S)模式,即数据库(比如试题库)内容放在远程的服务器上,在客户机上安装相应软件;C/S 结构在技术上很成熟,但该结构的程序往往只局限在小型的局域网内部,不利于扩展。该结构的每台客户机都需要安装相应的客户端程序。采用该结构,系统的安装和维护工作比较繁重。同时,由于应用程序直接安装在客户机,客户机直接和数据库服务器交换数据,系统的安全性也受到一定的影响。

基于 Web 的方式其实是一种特殊的客户/服务器方式,在这种方式中,客户端是各种各样的浏览器。为了区别传统的 C/S 模式,通常称为浏览器/服务器(B/S)模式。B/S 采用三层体系结构,即包括数据库系统、应用服务器、客户浏览器三部分。由于采用了互联网的相关技术,B/S 结构的系统开放性好,易维护和扩展。客户浏览器只跟 Web 服务器交换数据,数据安全性比较高。当然,B/S 结构在网络安全方面也有弱点。在 C/S 结构中,应用程序是在客户机上运行的独立程序,如果这台计算机安全的话,那么应用程序就是安全的。而在 B/S 结构中,众多的客户浏览器访问的是同一个 Web 服务器,Web 服务器会成为攻击的对象。

结合 C/S 和 B/S 结构的特点,还有基于 B/S 的三层体系结构,即 Browser/Web 服务器/数据库服务器的三层模式,是在客户端与数据库之间加入了一个中间层。三层体系结构的数据库应用系统将业务规则、数据访问、合法性校验等工作放到了中间层进行处理。通常情况下,客户端不直接与数据库进行交互,而是通过与中间层通讯建立连接,再经由中间层与数据库进行交互,保证其具有开放性和可扩充性,同时提高了数据库的安全性,如图 1-6 所示。

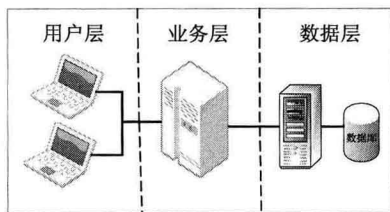


图 1-6 三层模式架构图

1.3 数据模型

【提要】人们把表示事物的主要特征抽象地用一种形式化的描述表示出来,模型方法就是这种