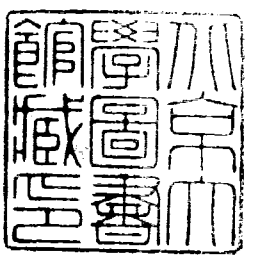
高等学校21世纪教材

数据库基础与应用

王　珊　李盛恩　编著



人民邮电出版社

需要全本请访问 www.ertongbook.com

图书在版编目(CIP)数据

数据库基础与应用 / 王珊，李盛恩编著. —北京：人民邮电出版社，2002.8
ISBN 7-115-10318-6

I. 数… II. ①王…②李… 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(2002)第 045218 号

内 容 提 要

本书主要讲解数据库系统的基本概念和基本技术，数据库设计的理论和方法步骤，数据库应用开发技术，数据仓库和联机分析新技术和新应用。本书为高等学校理工科计算机课程中数据库的教材，也可供数据库应用部门、从事数据库系统设计和开发部门的工程技术人员学习参考。

高等学校 21 世纪教材
数据库基础与应用

◆ 编　　著　王　珊　李盛恩
责任编辑　潘春燕

◆ 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　电子函件　315@ptpress.com.cn
网址　http://www. ptpress.com.cn
读者热线　010-67180876
北京汉魂图文设计有限公司制作
北京隆昌伟业印刷有限公司印刷
新华书店总店北京发行所经销

◆ 开本：787×1092　1/16
印张：16.75
字数：393 千字　　2002 年 8 月第 1 版
印数：1-6 000 册　　2002 年 8 月北京第 1 次印刷

ISBN 7-115-10318-6/TP・2883

定价：22.00 元

本书如有印装质量问题，请与本社联系　电话：(010)67129223

丛书前言

当今世界，科学技术突飞猛进，知识经济已见端倪，国际竞争日趋激烈。教育在综合国力的形成中处于基础地位，国力的强弱将越来越取决于劳动者的素质，取决于各类人才的质量和数量，这对于培养和造就我国 21 世纪的一代新人提出了更加迫切的要求。21 世纪初，我国高等教育呈快速发展的势头。教材是体现教学内容和教学方法的知识载体，是进行教学的基本工具，也是深化教育教学改革、全面推进素质教育、培养创新人才的重要保证。因此，高等教育教材建设必须有一个与之相适应的快速发展。

随着计算机软硬件的不断升级换代，计算机教学内容也随之更新，尤其随着教育部“高等教育面向 21 世纪教育内容与课程体系改革”计划的实施，对教材也提出了新的要求。为此我们聘请了国内高校计算机教学方面知名的专家教授，精心策划编写了这套“高等学校 21 世纪教材”。

为真正实施精品战略，组织编写好这套教材，我们在国内高校做了系统、详细的调查，对教育部制订的教育计划做了认真的研究，还对国内外已出版的教材做了理性的分析，确立了依托国家教育计划、传播先进教学理念、为培养符合社会需要的高素质创新型人才服务的宗旨。

在本套教材的策划过程中，我们多次组织了由专家及高校一线教师参加的研讨会，对现有比较出色的教材的特点及优点进行了分析，博采众长，力求实现教材权威性与实用性的完美结合。

本套教材有如下特点：

1．考虑到全国普通高等院校学生的知识、能力、素质的特点和实际教学情况，在编写教材时把重点放在基本理论、基础知识、基本技能与方法上。

2．紧密结合当前技术的新发展，在阐述理论知识的同时侧重实用性。

3．力求在概念和原理的讲述上严格、准确、精练，理论适中，实例丰富，写作风格上深入浅出，图文并茂，便于学生学习。

4．为适应当前高校课程种类多、课时数要压缩的教学特点，教材不仅篇幅有很大的压缩，而且均配有电子教案，以满足现代教学新特点的需要，做到易教易学。

5．所选作者均是国内有丰富教学实践经验的知名专家、教授，所编教材具有较高的权威性。

教育的改革将不会停止，教材也将会不断推陈出新。目前本套教材即将推出，将接受广大教学第一线教师的检验。

由于我们的水平和经验有限，这批教材在编审、出版工作中还存在不少缺点和不足，希望使用本套教材的学校师生和广大读者提出批评和建议，以便改进我们的工作，使教材质量不断提高。

编 者 的 话

随着21世纪的到来，计算机技术的发展越来越快，应用越来越广泛。为了适应这一发展趋势，高等学校各个专业都增设了不同层次的计算机课程，包括数据库方面的课程。

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。它已成为计算机信息系统与应用系统的核心技术和重要基础。数据库技术已经成为人们存储数据、管理信息的最先进最常用的技术。

但是，不同院系、不同专业对数据库课程的要求是不一样的。我和萨师煊教授编写的《数据库系统概论》(高等教育出版社出版)主要是针对计算机科学与技术本科专业的要求编写的。而本书是供高等学校理工科专业使用的数据库教材。

本书从开发一个数据库应用系统以及使用数据库系统的角度来讲解数据库系统的基本概念和方法技术。

全书分为4个部分，共12章。

第1部分（第1～4章）是数据库系统的基本知识和关系数据库系统的基本知识。内容包括：数据管理的进展和数据库系统构成的一般概念；关系模型和关系代数；关系数据库的SQL语言；数据库的事务管理、恢复、并发控制、完整性和安全性等控制保护技术。

第2部分（第5～8章）讲解数据库应用系统开发中数据库设计的方法和技术。内容包括数据库设计的特点和基本问题；概念模型与E-R方法；关系数据库设计理论和数据库逻辑设计；数据库设计的具体步骤。

第3部分（第9、10章）主要讲解在网络环境下数据库应用系统开发中使用的ODBC技术，数据库系统的体系结构，重点是客户/服务器结构的介绍。

第4部分（第11、12章）是数据库技术的新应用—数据仓库技术和OLAP技术。

本书第1到第4章由李盛恩副教授编写，第5到第12章由王珊教授编写。全书由王珊教授修改定稿。

在编写本书的过程中，如何把握理工科专业学生学习数据库系统课程的要求，教材内容应该如何选取和组织，对我们来说是一个新的课题。带着这个问题我们一边思考一边编写这本教材。在教材具体内容的选取上尽力注意先进性和实用性。重点介绍数据库系统的基本概念和基本技术，适当增加数据库新技术方面的内容，如数据仓库技术和OLAP、网络环境下的数据库技术。在写作上力求讲解原理深入浅出，从具体实例到抽象概念的提升。但由于学识浅陋，存在的问题一定不少，希望读者提出批评意见。更希望学术同仁不吝赐教。

王 珊

2002年初春 于中国人民大学

目　录

第 1 章　概述 …… 1

1.1　数据处理 …… 1
1.2　数据管理技术的发展 …… 1
1.2.1　人工管理阶段 …… 2
1.2.2　文件系统阶段 …… 2
1.2.3　数据库技术阶段 …… 4
1.2.4　数据库系统的优点 …… 4
1.3　数据库系统的组成 …… 6
1.3.1　数据库 …… 6
1.3.2　数据库管理系统 …… 7
1.3.3　数据库应用 …… 10
1.3.4　数据库系统的分类 …… 11
1.4　数据模型 …… 12
习题 …… 13
本章参考文献 …… 13

第 2 章　关系模型 …… 14

2.1　关系模型的基本概念 …… 14
2.1.1　二维表 …… 14
2.1.2　关系 …… 15
2.1.3　关系模型 …… 16
2.1.4　关系模型的优点 …… 18
2.2　关系代数 …… 18
2.2.1　传统的集合运算 …… 18
2.2.2　专门的关系运算 …… 20
2.2.3　用关系代数表示关系操作 …… 22
2.3　查询优化 …… 24
2.3.1　一个实例 …… 24
2.3.2　优化策略 …… 25
2.3.3　关系代数等价变换规则 …… 26
2.3.4　查询优化的优点 …… 28
习题 …… 28
本章参考文献 …… 30

需要全本请访问 www.ertongbook.com

第 3 章　关系数据库标准语言 SQL …… 31

3.1　SQL 概述 …… 31
3.1.1　SQL 的产生与发展 …… 31
3.1.2　SQL 的基本概念 …… 31
3.1.3　SQL 的组成 …… 32
3.1.4　SQL 的特点 …… 32
3.2　学生—课程数据库 …… 33
3.3　数据定义 …… 35
3.3.1　定义、修改与删除表 …… 35
3.3.2　建立与删除索引 …… 38
3.4　数据更新（一） …… 39
3.4.1　插入操作 …… 39
3.4.2　修改操作 …… 40
3.4.3　删除操作 …… 41
3.5　数据查询 …… 41
3.5.1　单表查询 …… 42
3.5.2　连接查询 …… 50
3.5.3　SQL2 中的连接查询表达方法 …… 53
3.5.4　嵌套查询 …… 54
3.5.5　集合操作 …… 64
3.5.6　SELECT 语句的一般格式 …… 65
3.6　数据更新（二） …… 66
3.6.1　插入操作 …… 67
3.6.2　修改操作 …… 67
3.6.3　删除操作 …… 67
3.7　视图 …… 67
3.7.1　定义视图 …… 68
3.7.2　查询视图 …… 70
3.7.3　更新视图 …… 71
3.7.4　视图的作用 …… 73
3.8　数据控制之一：安全性控制 …… 74
3.8.1　授权 …… 75
3.8.2　收回权限 …… 76
3.9　数据控制之二：完整性控制 …… 77
3.9.1　实体完整性 …… 77
3.9.2　参照完整性 …… 78
3.9.3　属性值限制 …… 79
3.9.4　元组级限制 …… 79

3.9.5　完整性修改 …… 80
3.10　空值的处理 …… 81
3.11　嵌入式 SQL …… 82
3.11.1　嵌入式 SQL 的一般形式 …… 83
3.11.2　嵌入式 SQL 语句与主语言之间的通信 …… 83
3.11.3　不用游标的 SQL 语句 …… 86
3.11.4　使用游标的 SQL 语句 …… 89
3.11.5　动态 SQL 简介 …… 94
3.12　PL/SQL 简介 …… 95
3.12.1　PL/SQL 的块结构 …… 95
3.12.2　变量常量的定义 …… 96
3.12.3　控制结构 …… 96
3.12.4　异常处理 …… 99
3.12.5　存储过程和函数 …… 102
3.12.6　游标 …… 105
3.13　数据库系统的三级模式 …… 106
3.13.1　模式 …… 107
3.13.2　外模式 …… 107
3.13.3　内模式 …… 108
习题 …… 108
本章参考文献 …… 110

第 4 章　数据保护 …… 111

4.1　事务的概念 …… 111
4.1.1　实例 …… 111
4.1.2　事务的特性 …… 112
4.1.3　SQL 中的事务控制 …… 113
4.2　恢复 …… 113
4.2.1　故障的种类 …… 113
4.2.2　转储操作 …… 115
4.2.3　日志文件 …… 115
4.2.4　恢复过程 …… 116
4.3　并发控制 …… 117
4.3.1　串行调度和可串行化调度 …… 118
4.3.2　冲突可串行化调度 …… 121
4.3.3　封锁并发控制方法 …… 122
4.4　完整性和安全性 …… 126
4.4.1　完整性 …… 126
4.4.2　安全性 …… 130

习题 …… 137
本章参考文献 …… 137

第 5 章　数据库设计概述 …… 139

5.1　数据库设计的特点 …… 139
5.1.1　“三分技术，七分管理，十二分基础数据”是数据库建设的基本规律 …… 139
5.1.2　结构（数据）设计应该和行为（处理）设计密切结合 …… 140
5.2　数据库设计方法 …… 141
5.3　数据库的生命周期 …… 142
5.4　数据库设计步骤 …… 142
习题 …… 144
本章参考文献 …… 144

第 6 章　概念模型与 E-R 方法 …… 146

6.1　概念模型 …… 146
6.2　概念模型的主要概念 …… 147
6.3　实体型之间的联系 …… 147
6.3.1　两个实体型之间的联系 …… 147
6.3.2　两个以上的实体型之间的联系 …… 148
6.4　实体-联系方法 …… 149
6.4.1　E-R 图的表示方法 …… 149
6.4.2　实体与属性的划分 …… 152
习题 …… 153
本章参考文献 …… 154

第 7 章　关系数据库设计理论 …… 155

7.1　数据依赖对关系模式的影响 …… 155
7.2　函数依赖 …… 157
7.2.1　函数依赖 …… 157
7.2.2　码 …… 157
7.3　范式 …… 158
7.3.1　第 1 范式（1NF） …… 158
7.3.2　第 2 范式（2NF） …… 160
7.3.3　第 3 范式（3NF） …… 161
7.3.4　BC 范式（BCNF） …… 162
7.4　多值依赖与第 4 范式（4NF） …… 163
7.4.1　多值依赖 …… 164
7.4.2　第 4 范式（4NF） …… 165
7.5　关系模式的规范化 …… 166

7.6 数据依赖的公理系统 …… 167
习题 …… 170
本章参考文献 …… 171

第 8 章 数据库设计步骤 …… 173

8.1 需求分析 …… 173
8.1.1 需求分析的任务和过程 …… 173
8.1.2 数据流图（Data Flow Diagram，简称 DFD） …… 175
8.1.3 数据字典 …… 178
8.1.4 需求分析注意点 …… 179
8.2 概念结构设计 …… 180
8.2.1 概念结构设计的方法与步骤 …… 180
8.2.2 局部视图设计 …… 181
8.2.3 视图的集成 …… 185
8.3 逻辑结构设计 …… 189
8.3.1 E-R 图向关系模型的转换 …… 189
8.3.2 数据模型的优化 …… 190
8.3.3 设计用户子模式 …… 191
8.4 数据库的物理设计 …… 192
8.4.1 数据库物理设计的内容 …… 192
8.4.2 选择关系存取方法 …… 192
8.4.3 确定数据库的存储结构 …… 194
8.4.4 评价物理结构 …… 197
8.5 数据库的实施 …… 197
8.5.1 数据的载入和应用程序的调试 …… 197
8.5.2 数据库的试运行 …… 198
8.6 数据库的运行和维护 …… 198
习题 …… 199
本章参考文献 …… 200

第 9 章 开放数据库互连（ODBC） …… 201

9.1 数据库互连概述 …… 201
9.2 ODBC 的工作原理概述 …… 202
9.3 使用 ODBC 的系统结构 …… 203
9.3.1 应用程序 …… 204
9.3.2 驱动程序管理器 …… 204
9.3.3 数据库驱动程序 …… 204
9.3.4 ODBC 数据源管理 …… 205
9.4 ODBC 驱动程序的分类 …… 205

9.4.1 API 一致性级别 …… 205
9.4.2 SQL 语法一致性级别 …… 206
9.4.3 驱动程序类型 …… 206
9.5 ODBC 的工作流程 …… 208
9.5.1 建立和释放 ODBC 环境 …… 209
9.5.2 建立和释放 ODBC 连接 …… 209
9.5.3 连接和断开数据源 …… 210
9.5.4 分配和释放语句句柄 …… 211
9.5.5 执行 SQL 语句 …… 213
9.5.6 结果集的处理 …… 214
习题 …… 218
本章参考文献 …… 218

第 10 章 客户/服务器结构的数据库系统 …… 219

10.1 数据库系统体系结构概述 …… 219
10.1.1 主/从式结构的数据库系统 …… 219
10.1.2 分布式结构的数据库系统 …… 220
10.1.3 客户/服务器结构的数据库系统 …… 220
10.2 客户/服务器一般概念 …… 222
10.2.1 客户/服务器的工作模式 …… 222
10.2.2 客户/服务器的主要技术特征 …… 222
10.2.3 客户/服务器结构的组成 …… 222
10.2.4 客户/服务器中的服务器类型 …… 223
10.3 客户/服务器结构的数据库系统 …… 223
10.3.1 客户/服务器数据库系统的功能划分 …… 224
10.3.2 客户/服务器数据库系统实例 …… 224
10.3.3 客户/服务器数据库系统优点 …… 225
10.4 两层与三（多）层结构 …… 226
10.4.1 两层客户/服务器结构的局限性 …… 226
10.4.2 三（多）层体系结构 …… 227
习题 …… 229
本章参考文献 …… 229

第 11 章 数据仓库技术 …… 230

11.1 概述 …… 230
11.2 从数据库到数据仓库 …… 230
11.3 数据仓库的基本概念 …… 232
11.3.1 什么是数据仓库 …… 232
11.3.2 主题与面向主题 …… 233

11.3.3　数据仓库的数据是集成的 …… 235
11.3.4　数据仓库的数据是不可更新的 …… 235
11.3.5　数据仓库的数据是随时间不断变化的 …… 235
11.4　数据仓库的数据组织 …… 236
11.5　数据仓库系统的体系结构 …… 237
11.5.1　数据仓库的后台工具 …… 238
11.5.2　数据仓库服务器和 OLAP 服务器 …… 238
11.5.3　前台工具 …… 239
11.6　企业的体系化数据环境 …… 239
11.6.1　数据环境的层次 …… 240
11.6.2　数据集市（Data Mart） …… 240
11.7　创建数据仓库 …… 241
习题 …… 241
本章参考文献 …… 242

第 12 章　联机分析处理（OLAP）技术 …… 243

12.1　什么是 OLAP …… 243
12.2　多维数据模型 …… 243
12.2.1　基本概念 …… 243
12.2.2　多维分析的基本操作 …… 246
12.3　OLAP 的实现 …… 248
12.3.1　MOLAP 结构 …… 248
12.3.2　ROLAP 结构 …… 250
习题 …… 251
本章参考文献 …… 251

需要全本请访问 www.ertongbook.com

第1章 概 述

计算机最早用于科学计算，目前计算机已经深入到了我们生活的各个方面。计算机应用大致分为三大类：科学计算、数据处理和过程控制。其中，数据处理占了很大比重（约70%），而数据管理又是数据处理的一个重要方面，数据库技术是数据管理的最新技术，是计算机科学的重要分支，是信息技术的基石。

本章介绍数据库系统的基本概念，包括数据管理技术的发展过程、数据库系统的组成部分和数据模型的基本概念。本章是后面各章节的准备和基础。

1.1 数据处理

数据是描述事物的符号记录。数字是一种常见的数据表现形式，其他表现形式有文字、图形、图像和声音等。例如，183 是一个数据，如果它是一个人的身高，我们会得到这是一个很高的人的信息，如果它代表一个学生三门考试课的总成绩，我们会得出这个同学的成绩很一般的结论。可见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。数据的解释是对数据含义的解释，数据的含义称为数据的语义，数据与其语义是不可分的。另外数据之间是有联系的，是有结构的。例如，关于一个学生的数据是需要组织在一起的，如后面表 1.1 中把学生的学生编号、姓名和性别等组织在一起。

数据处理是指从某些已知的数据出发，推导加工出一些新的数据，这些新的数据又表示了新的信息。例如，从学生的基本信息中可以推导出学生的性别构成、年龄构成和地区构成，分析不同专业这些构成的差异。在具体操作中，涉及到数据收集、管理、加工和输出等过程。在数据处理中，通常数据的计算比较简单，而数据的管理比较复杂。数据管理是指数据的收集、整理、组织、存储和检索等操作，这部分操作是数据处理业务的基本环节，是任何数据处理业务中必不可少的共有部分，因此有必要研究数据管理的技术，对数据处理提供有利的支持。

1.2 数据管理技术的发展

数据管理技术的发展经历了人工、文件系统和数据库技术等 3 个阶段。我们通过一个学校的教务部门对学生、课程和成绩的管理来阐述各阶段的特点。

在没有使用计算机时，教务部门的工作人员将有关学生的信息抄写在一张张卡片上，为了方便查找，将同一个班、同一个年级、同一个系的学生的卡片存放在相邻的地方，并对不同的班、年级和系做上标签。每门课程的信息也是抄写在卡片上，将同一个专业的卡片放在

一起并做上标签。每个学期末将同一个班的各门课的成绩单收集起来存放在档案中。

当查找一个学生的信息时，如果知道他所在的系和班级，按照建好的标签可以很快找到该学生的卡片；如果只知道他的姓名，那只有在所有学生的卡片中一个一个地查找，需要花费很多的时间。当计算一个学生某个学期的平均成绩时，首先在档案中找到该学生所在班级这个学期的所有成绩单，从中找出该学生各门课程的成绩，再计算平均成绩。统计某一门课的成绩分布时也只能进行手工计算。

1.2.1 人工管理阶段

在使用计算机的初期（20 世纪 50 年代中期），计算机主要用于科学计算，其它工作尚未开展。外部存储器只有纸带、卡片、磁带，没有磁盘等直接存取的存储设备；软件只有汇编语言，没有操作系统，没有管理数据的软件；数据处理方式是批处理。该阶段的特点是：

（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）由应用程序管理数据

数据需要由应用程序自己管理，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，由于必须各自定义，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，这就进一步加重了程序员的负担。

1.2.2 文件系统阶段

计算机的操作系统实现了文件系统，用户可以把相关数据组织成一个文件存放在计算机中，在需要时只要提供文件名，计算机就能从文件系统中找出用户所要的文件，把文件中存储的数据提供给用户进行处理。

为了改变手工管理阶段查找，计算工作量大，需要花费很长时间的被动局面，教务部门编写了一套学籍管理软件。将学生卡片、课程卡片和学生成绩单中的内容分别存放到了文件 student、course 和 score 中，文件的格式和部分内容见表 1.1～1.3。

表 1.1　　student 表

学生编号	姓　名	性　别	年　龄	系	家庭住址	联系电话
20000121	王林	男	19	计算机	北京	63458967
20000122	张大民	男	18	计算机	烟台	2678090
20000123	顾芳	女	19	计算机	天津	89561234

表 1.2　　course 表

课程编号	名　称	学 时 数	教材名称
J013	数据库原理	80	数据库系统概论
J028	C 语言程序设计	60	C 语言教程
J032	人工智能	64	人工智能原理

表 1.3　　score 表

学生编号	课程编号	学　期	成　绩
20000121	J013	7	86
20000122	J028	2	91
20000123	J032	6	79

对每个文件编写了一组程序用于数据维护，包括增加一条记录、删除一条记录、修改一条记录和查询一条记录。在此基础上根据实际工作的需要编写一些查询和报表打印程序。例如，根据学生姓名、学生编号查找学生的信息，统计某学期某个学生的平均成绩，统计某门课的平均成绩等。

软件投入使用后，工作人员的效率大大提高。例如，学期末将各门课的考试成绩输入计算机以后，可以很快计算出学生的平均成绩，打印出需要补考的学生的名单。但经过一段时间的使用后，工作人员发现有时必须修改程序和文件结构才能适应工作的需要。例如，学校领导让统计新入学的学生中有多少是来自山东的，因为学籍管理软件没有实现这个查询，必须编写一段程序来实现这个功能。又如当需要往 student 文件中增加“血型”一项时，因为这涉及到改变文件的结构，需要若干步骤才能完成：第一步，建立一个新文件 student_new，其结构是在 student 的结构中加入血型这一项。第二步，编写一个程序将文件 student 中的数据转存到 student_new 中。第三步，删除文件 student。第四步，将文件 student_new 重命名为 student。这项工作到此并没有结束，因为文件中保存的是一组数据，不保存数据结构，而数据结构是在程序中实现的，文件 student 的结构写到了所有使用它的程序中，必须一一修改这些程序以适应新的文件结构，否则程序运行就会出错。

可以看到用文件系统管理数据具有如下特点：

（1）数据可以长期保存

数据可以组织成文件长期保存在计算机中反复使用。

（2）由文件系统管理数据

文件系统把数据组织成内部有结构的记录，实现“按文件名访问，按记录进行存取”的管理技术。

文件系统使应用程序与数据之间有了初步的独立性，程序员不必过多地考虑数据存储的物理细节，数据在存储上的改变不一定反映在程序上，大大节省了维护程序的工作量。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，一个文件基本上对应于一个应用程序。当不同的应用程序使用部分相同的数据时，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，文件的逻辑结构对该应用程序来说是优化的，因此要想对现有的数据再增加一些新的应用会很困难，系统不容易扩充。一旦数据的逻辑结构改变，必须修改应用程序，修改文件结构的定义。因此数据与程序之间仍缺乏独立性。

1.2.3 数据库技术阶段

随着各部门计算机应用水平的提高，学校提出了共享教务处保存的有关学生数据的要求，学校购置了一个关系数据库管理系统（Relational DataBase Management System，RDBMS），对全校各部门的数据进行统一组织，将教务处和学生工作处保存的学生数据进行了合并，各系的有关工作人员都可以访问这些数据。

在系统中建立三个关系：student、course、score，这时不用自己再编写程序来实现了，而是向 RDBMS 提交一条命令：

```
CREATE TABLE STUDENT（Sno       CHAR（8）NOT NULL UNIQUE，
                      Sname     CHAR（10），
                      Ssex      CHAR（2），
                      Sage      INT，
                      Sdept     CHAR（20），
                      Saddress  CHAR（40），
                      Stelephone CHAR（14））；
```

这条命令在数据库中建立了一个关系 student，用来保存学生的信息，更重要的是将 student 的结构也保存到数据库的数据字典中。向关系中增加一个元组（记录）、删除一个元组、修改一个元组用 DBMS 提供的语句 INSERT、DELETE 和 UPDATE 来完成，这些命令如何操作磁盘上的数据是由 RDBMS 来完成的，程序员不用编写专门的程序，节省了程序员大量的时间和精力。当需要往关系 student 增加血型一项时，可以用一条命令来实现：

```
ALTER TABLE STUDENT ADD BloodType CHAR（2）；
```

因为关系 student 的数据和结构都由 RDBMS 管理，在仅向关系 student 中添加一个属性的情况下，既不需要编写转数据的程序也不用修改那些使用了关系 student 的程序。

1.2.4 数据库系统的优点

从上面的例子中，我们可以初步体会到使用数据库系统给人们带来的方便。与人工管理和文件系统相比，数据库系统的优点主要有以下几个方面：

1．数据结构化

数据结构化是数据库和文件系统之间的根本区别。在文件系统中每个文件是一个“流式”文件，文件中存储的数据是若干个字节的数据，可以将这些字节流解释成由一个一个记录构成的，而每个记录是有结构的。例如，文件 student 的一个记录是由学生编号、姓名、性别、

年龄、系、家庭住址和联系电话组成的。由于每个文件对应一个应用，这种结构只适用于这个特定的应用，对于其他应用并不适用。例如，教务处编写的有关程序，使用教务处的 student 文件；而学生处编写的有关程序使用学生处的 student 文件，两个 student 文件的结构可以不一样，也不能互相公用。

在文件系统中任何两个文件的记录之间是不存在任何联系的。例如，文件 student 的记录和文件 score 的记录是毫无关系的，但实际上是有关系的，如文件 student 的第一条记录和 score 的第一条记录是有联系的，前者表示学生王林的基本信息，后者表示王林数据库原理课的考试成绩，这种联系只能写在程序中。在关系数据库中，两个关系的记录之间的联系是可以用参照完整性（见 2.1.3 节）来表述的。如果向 score 中增加一个学生的考试成绩，但是这个学生并没有出现在 student 关系中，RDBMS 保证这样的增加操作是无效的，是不能完成的，从而保证了数据的正确性。在文件系统中要做到这一点，必须由程序员编写一段代码来实现。

在文件系统中最小的存取单位是记录，而在数据库中既可以存取一个元组（记录）或一组元组，又可以存取一组数据项或一个数据项。

2．数据的共享性提高、冗余度低、易扩充

在文件系统中文件是面向一个特定应用的，造成了不同的应用之间不能共享数据，其后果是数据的冗余存放，即同一个数据存放在多个文件中。例如，学生信息既存放在教务处的 student 文件中，又存储在学生处的 student 文件中。冗余存放会带来数据不一致性。例如，一个学生搬家后，电话号码变了，教务处将保存的电话号码修改为新的电话号码，但是学生处并没有修改该学生的电话号码，这样同一个学生就有了两个电话号码，不符合实际情况。冗余存放也浪费了存储空间。

在数据库系统中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一致性。

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，而且可以增加新的应用，这就使得数据库系统弹性大，易于扩充，可以适用各种用户的要求。我们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者加上一部分数据便可以满足新的需求。这就是系统弹性大的含义。

3．数据独立性高

数据独立性是数据库领域中的一个常用术语，包括数据的物理独立性和数据的逻辑独立性，数据独立性是数据库系统的一个重要概念。

物理独立性是指应用程序和数据库中数据的存放位置和存储格式是相互独立的。当数据库中的数据被转移到存取速度更快的磁盘或者 DBMS 采用性能更好的存储格式时，不用修改应用程序。在文件系统中如果将一个文件从一种类型（如顺序文件）改为另一种类型（如索引顺序文件），从一个目录转移到另一个目录或者从一个磁盘转移到另外一个磁盘，都必须修改程序，因为文件的具体位置、存储格式是写到程序中的。

逻辑独立性是指应用程序和数据库的逻辑结构是相互独立的，逻辑结构改变了，用户的程序可以不修改。例如，关系 student 的逻辑结构是每个元组由 Sno、Sname、Ssex、Sage、Sdept、Saddress 和 Stelephone 属性组成，当增加了一个血型属性 BloodType，关系 student 的逻辑结构发生了改变，可是不用修改应用程序，因为数据的结构是保存在数据库中而不是写

在程序中。但是如果删除了 Sage 属性，就必须修改那些存取了 Sage 属性的应用程序。

4. 数据由 DBMS 统一管理和控制

数据库提供了数据共享使得多个用户可以同时存取数据库中的数据，为了在多用户环境下保证数据的安全性和一致性，DBMS 还提供了以下的数据控制功能：

- 数据的安全性保护

数据的安全性是指保护数据以防止不合法的使用造成数据泄密和破坏。例如，数据库管理员根据工作的需要给每个用户一定的权限，使得用户只能访问职权范围内的数据。

- 数据的完整性检查

数据的完整性是指数据的正确性、有效性和相关性。关系数据库管理系统对实体完整性、参照完整性以及用户定义的其他限制进行检查，对违反完整性的操作加以限制和处理。

- 并发控制

当多个用户同时存取数据库中的数据时可能发生相互干扰使得用户得到错误的结果或数据库的完整性受到破坏，并发控制机制使得多个用户同时对数据库的操作结果等同于多个用户顺序操作数据库的结果。

- 数据库恢复

计算机系统的硬件故障、软件故障、操作员的失误以及故意的破坏都可能影响数据库中数据的正确性，甚至造成数据库部分或全部数据的丢失。DBMS 通过数据备份以及日志技术在发生故障时将数据库恢复到某个已知的正确状态。

1.3 数据库系统的组成

数据库系统由三部分组成：数据库（DB）、数据库管理系统（DBMS）和数据库应用系统，这三部分之间的关系如图 1.1 所示。

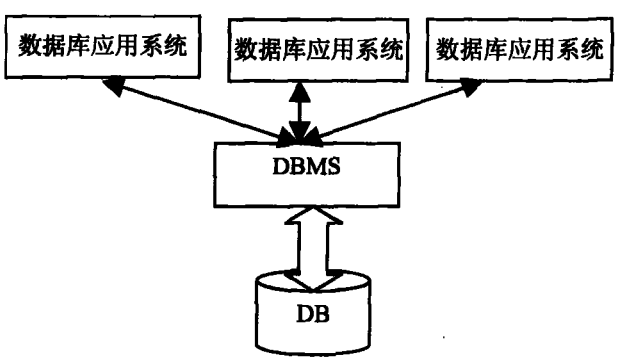


图 1.1 数据库系统的组成

1.3.1 数据库

数据库（DB）是用来存储数据的。数据库中有两大类数据，一类是用户数据，如关系 student 中每个学生的信息。另一类是系统数据，如关系 student 的结构（即每个学生的信息

需要全本请访问 www.ertongbook.com