

21 世纪计算机应用技术系列规划教材

数据库原理及应用

范剑波 编著

人民邮电出版社

图书在版编目 (CIP) 数据

数据库原理及应用 / 范剑波编著. —北京: 人民邮电出版社, 2006.1

(21 世纪计算机应用技术系列规划教材)

ISBN 7-115-13918-0

I. 数... II. 范... 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 (2005) 第 114822 号

内 容 提 要

本书详细地介绍了数据库的基本知识, 阐述了关系数据库系统的理论、方法和设计技术, 讲解了 SQL Server 2000 中 T-SQL, 给出了基于 C/S 和 B/S 结构的数据库应用系统设计的实例, 在最后还是介绍了数据库技术的新进展。具体内容包括: 数据库系统概述、实体-联系数据模型、关系数据模型、关系数据库结构化查询语言、关系数据库的模式设计、关系数据库设计、数据库保护、数据库应用系统设计实例、面向对象数据库系统和分布式数据库系统。

本书每章后附有习题, 供学生平时练习之用; 本书最后附有课程考试模拟试题及参考答案、课程实验指南和课程设计大纲, 供教师和学生使用。本书配有电子教案, 若有需要请从人民邮电出版社网站下载。

本书为普通高等学校计算机及相关专业本科教材, 也可供高职高专计算机相关专业学生选用, 并可供从事计算机软件工作的科技人员和工程技术人员参考。

21 世纪计算机应用技术系列规划教材

数据库原理及应用

-
- ◆ 编 著 范剑波
责任编辑 邹文波
执行编辑 韩学义
 - ◆ 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>
北京艺辉印刷有限公司印刷
新华书店总店北京发行所经销
 - ◆ 开本: 787×1092 1/16
印张: 18.25
字数: 434 千字 2006 年 1 月第 1 版
印数: 1-3 000 册 2006 年 1 月北京第 1 次印刷

ISBN 7-115-13918-0/TP · 4905

定价: 25.00 元

读者服务热线: (010)67170985 印装质量热线: (010)67129223



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。

应用型本科教材的编写应该注重与工程应用相结合，注重与能力培养相联系，注重与培养目标相一致。在这样的思想指导下，我们对十多年数据库课程教学与科研工作的实践进行了总结，对教育部计算机重点示范性专业教学改革的工作进行了探索和实践，形成了计算机类专业“数据库原理及应用”课程的教学模式：由实践（可视化程序设计）→ 理论（数据库原理及应用）→ 再实践（数据库课程设计，网络数据库技术及应用）→ 再提高（毕业设计，参与教师的科研），取得了明显的效果。经过近几年的实践，学生的数据库设计能力及实际编程动手能力均有了较大的提高。

本书具有以下特点。（1）集理论、应用和系统于一体。理论的目标是最终能进行关系数据库的设计；应用的目标是能按软件工程的规范和数据库设计的步骤来进行数据库信息管理系统的设计；系统的目标是通过使用 SQL Server 2000 和 B/S、C/S 结构来进行网络数据库的开发和应用。（2）将 SQL Server 2000 的内容结合到数据库原理中去讲，避免了部分内容的重复，也避免了在原理部分讲解时无实际系统作为参照时的空对空现象。（3）增加了数据库应用系统设计实例的内容，使学生能运用关系数据库的设计理论和软件工程的规范来设计和开发一个基于 C/S 或 B/S 结构的数据库应用系统。（4）增加了面向对象数据库系统和分布式数据库系统的内容，展示了数据库系统的新进展和广阔的发展前景和应用趋势。

本书主要内容共包括 10 章，各章主要内容如下。

第 1 章（数据库系统概述）主要介绍文件技术和数据库技术、数据描述和数据模型、数据库管理系统、数据库系统结构和数据库技术的研究，这些内容是学习后面几章的基础。本章重点要求掌握数据库技术的特点、数据模型的概念、数据库管理系统的评价和选择、数据库系统的三级模式结构等内容。

第 2 章（实体-联系数据模型）主要介绍数据库建模概述、E-R 模型（E-R 图）、设计原则、子类和继承以及约束的建模。本章重点要求掌握数据库建模的主要内容、E-R 图建模的方法和原则、E-R 图中类与子类的层次关系和子类的继承性、几种常用的约束建模方法等。另外，要深入理解键码、实体完整性和引用完整性这三个基本概念。

第3章（关系数据模型）系统讲解关系数据库的重要概念，包括关系的定义、关系模型的三要素（数据结构、关系操作和关系的完整性）和关系的完备性。要求重点掌握关系代数的运算，学会关系演算的表达方法，掌握关系数据库查询优化的内容。

第4章（关系数据库结构化查询语言）主要介绍 T-SQL 的数据定义、数据操纵和数据控制功能，介绍视图、存储过程和触发器，介绍批处理、脚本、局部变量、全局变量的概念，重点介绍流程控制语句和游标的使用方法。本章重点要求掌握 T-SQL 的查询和更新功能，灵活掌握视图的作用和用法，掌握流程控制语句和游标的使用方法。

第5章（关系数据库的模式设计）关系数据库的逻辑设计主要是设计关系模式，而深入理解函数依赖和键码的概念则是设计和分解关系模式的基础，要求初步掌握计算属性的闭包。模式设计是本章的重点，要求了解数据冗余和更新异常产生的根源；理解关系模式规范化的途径；准确理解第一范式、第二范式、第三范式和 BC 范式的含义、联系与区别；深入理解模式分解的原则；熟练掌握模式分解的方法，能正确而熟练地将一个关系模式分解成属于第三范式或 BC 范式的模式。了解多值依赖和第四范式的概念，掌握把关系模式分解成属于第四范式的模式的方法。

第6章（关系数据库设计）要求了解数据库设计的方法、特点和步骤，掌握需求分析、概念结构、逻辑结构、物理结构、数据库实施、数据库运行和维护六个阶段设计的目标、方法和应注意的事项，重点掌握概念结构和逻辑结构设计这两个阶段。

第7章（数据库保护）要求了解数据库保护的四种措施：数据库的恢复、并发控制、安全性和完整性，要求重点领会这些措施的基本概念、应用方法和常用技术。

第8章（数据库应用系统设计实例）要求读者了解应用系统开发模型，掌握数据库应用系统的设计技术。了解网络数据库的基本概念，掌握基于 C/S 和 B/S 结构应用系统的开发技术。

第9章（面向对象数据库系统）主要介绍了面向对象数据模型的基本概念、对象定义语言（ODL）、子类、从 ODL 设计到关系设计和对象查询语言（OQL）等内容，要求读者重点掌握面向对象数据模型的基本概念，学会用对象定义语言（ODL）来描述类和子类，掌握将 ODL 设计转换为关系设计的方法，了解对象查询语言。

第10章（分布式数据库系统）要求读者掌握分布式数据库系统的定义、特点和优缺点，了解分布式数据库系统的体系结构、查询处理和优化等内容。

本书作为教材使用时，建议总学时为 68 学时，其中理论要求 52 学时，实验要求 16 学时，带*章节根据需要可以选择或不选。各章理论学时分配如下：

章	学时数	章	学时数
第1章	6	第6章	6
第2章	3	第7章	4
第3章	6	第8章	4
第4章	8	第9章	6
第5章	6	第10章*	3

“数据库原理及应用”课程获 2003 年度浙江省高等学校精品建设课程。
本书配有电子教案，若有需要请从人民邮电出版社网站下载。
限于作者水平，书中难免存在错误和不妥之处，殷切期望广大读者给予指正。

作 者
2005 年 8 月于宁波

目 录



第 1 章 数据库系统概述	1
1.1 文件技术与数据库技术	1
1.1.1 文件技术的特点与局限性	1
1.1.2 数据库技术的产生与发展	3
1.1.3 数据库技术的特点	4
1.1.4 数据库基本概念	6
1.2 数据描述和数据模型	7
1.2.1 数据描述的领域	7
1.2.2 物理存储介质层次及数据描述	8
1.2.3 数据联系的描述	10
1.2.4 数据模型的概念	11
1.2.5 概念数据模型及实例	12
1.2.6 结构数据模型及实例	14
1.3 数据库管理系统	17
1.3.1 数据库管理系统的基本功能	17
1.3.2 DBMS 的组成	18
1.3.3 DBMS 的评价和选择	20
1.4 数据库系统结构	22
1.4.1 数据库系统的三级模式结构	22
1.4.2 数据与程序的独立性	23
1.4.3 数据库系统的组成	24
1.5 数据库技术的研究	26
小结	26
习题	26
第 2 章 实体-联系数据模型	29
2.1 数据库建模概述	29

2.2 E-R 模型 (E-R 图)	30
2.2.1 E-R 图的主要组成部分	30
2.2.2 E-R 图联系中的角色	31
2.2.3 E-R 图联系的多向性	31
2.3 设计原则	32
2.3.1 真实性	32
2.3.2 避免冗余	32
2.3.3 简单性	33
2.3.4 合理选择元素类型	33
2.4 子类和继承	34
2.4.1 E-R 图中的子类	34
2.4.2 E-R 图中的继承	34
2.5 约束的建模	35
2.5.1 E-R 图中的键码	35
2.5.2 单值约束	36
2.5.3 引用完整性	36
2.5.4 其他类型的约束	37
小结	37
习题	38
第 3 章 关系数据模型	40
3.1 关系模型的基本概念	40
3.1.1 关系的通俗解释	40
3.1.2 关系的数学定义	41
3.1.3 关系模型	42
3.1.4 关系数据库管理系统	44
3.2 关系代数	44
3.2.1 传统的集合运算	45
3.2.2 专门的关系运算	46
3.2.3 关系代数表达式	47
3.3 关系演算	49
3.3.1 元组关系演算	49
3.3.2 域关系演算	52
3.4 关系数据库查询的优化	53
3.4.1 查询优化问题的提出	53
3.4.2 关系代数的等价变换	56
3.4.3 查询优化的一般策略	56
3.4.4 查询优化的步骤	57
小结	59

习题	59
第 4 章 关系数据库的结构化查询语言	62
4.1 SQL 概述	62
4.1.1 标准 SQL 的特点	62
4.1.2 关系数据库的三级模式结构	63
4.1.3 SQL Server 中的 T-SQL 对标准 SQL 的扩充	64
4.2 T-SQL 的数据定义	64
4.2.1 SQL Server 中的数据库	64
4.2.2 SQL Server 中的数据库表和索引	65
4.3 T-SQL 数据查询功能	69
4.3.1 数据基本查询	69
4.3.2 数据分组查询	70
4.3.3 多表连接查询	71
4.3.4 数据的子查询	72
4.3.5 附加子句	73
4.4 T-SQL 数据更新功能	74
4.4.1 数据插入	74
4.4.2 数据修改	75
4.4.3 数据删除	76
4.5 T-SQL 的视图、存储过程和触发器	76
4.5.1 T-SQL 视图的建立、查询和删除	76
4.5.2 T-SQL 存储过程的建立、执行和删除	78
4.5.3 T-SQL 触发器的建立和删除	79
4.6 T-SQL 数据控制	80
4.6.1 语句权力的授予与收回	80
4.6.2 对象权力的授予与收回	81
4.7 批处理、脚本和变量	82
4.7.1 批处理和脚本	82
4.7.2 局部变量和全局变量	83
4.8 流程控制语句	85
4.8.1 程序注释语句	85
4.8.2 BEGIN...END 语句块	85
4.8.3 IF...ELSE 语句	86
4.8.4 WHILE 语句	87
4.8.5 BREAK 和 CONTINUE 语句	87
4.8.6 WAITFOR 语句	88
4.8.7 GOTO 语句	89
4.8.8 RETURN 语句	89

4.8.9 CASE 语句	90
4.8.10 PRINT 和 RAISERROR 语句	91
4.9 游标	93
4.9.1 声明游标	93
4.9.2 打开游标	94
4.9.3 关闭游标	95
4.9.4 释放游标	95
4.9.5 使用游标取数	95
4.9.6 利用游标修改数据	97
小结	99
习题	99
第 5 章 关系数据库的模式设计	101
5.1 关系模式的存储异常和数据依赖	101
5.2 函数依赖的概念	103
5.2.1 函数依赖的定义	103
5.2.2 完全函数依赖和部分函数依赖	103
5.2.3 传递函数依赖	104
5.2.4 关系中的键码	104
5.3 函数依赖的规则	105
5.3.1 三个推理规则	105
5.3.2 闭包的计算	106
5.4 关系的规范化	107
5.4.1 第一范式	108
5.4.2 第二范式	108
5.4.3 第三范式	109
5.4.4 BCNF 范式	110
5.4.5 多值依赖和第四范式	111
5.5 模式分解的优劣	114
5.5.1 模式分解的等价性	114
5.5.2 模式分解的规则和方法	116
5.6 关系模式规范化小结	119
小结	119
习题	119
第 6 章 关系数据库设计	121
6.1 数据库设计概述	121
6.1.1 数据库设计的方法	122
6.1.2 数据库设计的特点	122

6.1.3 数据库设计的步骤	122
6.2 需求分析	123
6.2.1 需求分析的任务和方法	124
6.2.2 数据字典	124
6.2.3 数据流图	125
6.3 概念结构设计	126
6.3.1 概念结构设计的方法与步骤	126
6.3.2 数据抽象与局部视图设计	127
6.3.3 视图的集成	131
6.4 逻辑结构设计	134
6.4.1 E-R 图向关系模型的转换	134
6.4.2 关系模型的优化	135
6.5 数据库的物理设计	136
6.5.1 数据库物理设计的要求与内容	136
6.5.2 关系模式存取方法的选择	137
6.5.3 数据库存储结构的确定	138
6.5.4 物理结构的评价	139
6.6 数据库的实施、运行和维护	139
6.6.1 数据库的实施	139
6.6.2 数据库的运行和维护	140
小结	141
习题	141
第 7 章 数据库保护	143
7.1 数据库的恢复	143
7.1.1 事务的概念和性质	143
7.1.2 故障的种类和恢复的基本原则	146
7.1.3 日志文件优先原则	147
7.1.4 SQL Server 的恢复技术	147
7.2 数据库的并发控制	148
7.2.1 并发操作存在的问题	148
7.2.2 并发调度的可串行性	149
7.2.3 并发控制的封锁机制	150
7.2.4 SQL Server 系统的并发控制	153
7.3 数据库的安全性	154
7.3.1 数据库的安全性概念	154
7.3.2 数据库的安全性控制	154
7.3.3 SQL Server 系统的安全性	155
7.4 数据库的完整性	157

7.4.1 完整性约束条件	157
7.4.2 数据库的完整性控制	158
7.4.3 SQL Server 系统的完整性	160
小结	163
习题	163
第 8 章 数据库应用系统设计实例	165
8.1 应用系统开发模型	165
8.1.1 瀑布模型	165
8.1.2 原型模型	168
8.2 宾馆信息管理系统的设计	168
8.2.1 系统的需求分析	169
8.2.2 系统的概念结构设计	170
8.2.3 系统的逻辑结构设计	170
8.2.4 系统的物理结构设计	171
8.3 基于 C/S 结构的宾馆信息管理系统实现	172
8.3.1 客户机/服务器应用系统概述	172
8.3.2 创建工程项目	175
8.3.3 创建系统主窗体	175
8.3.4 创建公用模块	175
8.3.5 系统管理模块的创建	181
8.3.6 客房标准管理模块的创建	182
8.3.7 客房信息管理模块的创建	189
8.3.8 订房信息管理模块的创建	193
8.3.9 结算信息管理模块的创建	195
8.4 图书网上销售系统的设计	196
8.4.1 网上购物流程设计	196
8.4.2 网上购物功能设计	196
8.4.3 网上购物数据库设计	197
8.5 基于 B/S 结构的图书网上销售系统的实现	198
8.5.1 浏览器/服务器应用系统概述	198
8.5.2 网上书店首页的创建	204
8.5.3 畅销图书网页的创建	205
8.5.4 购物袋网页的创建	207
8.5.5 结账区网页的创建	209
8.5.6 订单确认网页的创建	214
小结	217
习题	218

第 9 章 面向对象数据库系统	220
9.1 面向对象数据模型的基本概念	220
9.1.1 OO 模型的核心概念	220
9.1.2 对象与对象标识	221
9.1.3 类与类层次	223
9.1.4 继承	224
9.1.5 对象的嵌套	225
9.2 对象定义语言 ODL	225
9.2.1 面向对象的设计	226
9.2.2 类的声明	226
9.2.3 ODL 中的属性	226
9.2.4 ODL 中的联系和反向联系	227
9.2.5 联系的三种类型	229
9.3 子类	231
9.3.1 ODL 中的子类	231
9.3.2 ODL 中的多重继承	231
9.4 从 ODL 设计到关系设计	233
9.4.1 从 ODL 属性到关系属性	233
9.4.2 类中的非原子属性	233
9.4.3 单值联系的表示	234
9.4.4 多值联系的表示	235
9.4.5 联系和反向联系的表示	235
9.4.6 ODL 子类的表示	236
9.5 对象查询语言 OQL	236
9.5.1 OQL 概述	236
9.5.2 OQL 表达式	240
9.5.3 OQL 及其宿主语言编程	245
小结	246
习题	247
第 10 章 分布式数据库系统	248
10.1 分布式数据库系统概述	248
10.1.1 分布式数据库系统的定义	248
10.1.2 分布式数据库系统的特点	249
10.1.3 分布式数据库系统的优缺点	249
10.2 分布式数据库系统的体系结构	250
10.2.1 分布式数据库系统的模式结构	250
10.2.2 数据分片和分布透明性	251

10.2.3 分布式数据库管理系统	253
10.2.4 分布式数据库系统存在的问题	253
10.3 查询处理和优化	254
10.3.1 一个实例	254
10.3.2 查询处理和优化要解决的问题	255
10.3.3 查询优化的目标	256
10.4 分布式数据库系统的发展前景	256
小结	257
习题	257
附录 A “数据库原理及应用” 考试模拟试题及参考答案	259
附录 B SQL Server 2000 实验	268
实验 1 使用 SQL Server 的管理工具和实用程序	268
实验 2 T-SQL 的数据定义	268
实验 3 T-SQL 的数据查询	269
实验 4 T-SQL 的数据更新	269
实验 5 T-SQL 的数据完整性和数据控制	269
实验 6 T-SQL 的视图和存储过程	270
实验 7 T-SQL 的流程控制和游标	271
实验 8 数据库综合设计练习	272
附录 C “数据库原理及应用” 课程设计大纲	273
参考文献	278

第 1 章

数据库系统概述



电子计算机在其问世的前 10 年，主要用于解决科学研究和工程设计中的数值计算问题，这个时期它的技术与应用只被少数科学家所拥有。随着计算机的不断发展，计算机的应用已经从科学研究部门扩展到了各企业和行政部门。因此，在计算机的三大主要应用领域（即科学计算、数据处理和过程控制）中，数据处理迅速上升为计算机应用的主要方面。数据库技术就是作为数据处理的一门新技术而发展起来的。

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。其目的之一是从大量的、原始的数据中抽取、推导出对人们有价值的信息，以作为行动和决策的依据；目的之二是为了借助计算机科学地保存和管理复杂的大量的数据，以便人们能方便而充分地利用这些宝贵的信息资源。

数据库技术所研究的问题是如何科学地组织和储存数据，如何高效地获取和处理数据。数据库技术作为数据管理的最新技术，目前已广泛应用于各个领域，数据库系统已成为当今计算机系统的重要组成部分。

本章主要介绍文件技术和数据库技术、数据描述和数据模型、数据库管理系统、数据库系统结构以及数据库技术的研究，这些内容是学习后面几章的基础。本章重点要求掌握数据库技术的特点、数据模型的概念、DBMS 的评价和选择以及数据库系统的三级模式结构等内容。

1.1 文件技术与数据库技术

1.1.1 文件技术的特点与局限性

在 20 世纪 50 年代到 60 年代之间，计算机不仅用于科学计算，而且已经大量用于数据处理，此时计算机外存储器有了磁鼓、磁盘等直接存取的存储设备，为计算机进行事务管理奠定了硬件基础。与此同时，数据结构设计和数据管理技术研究的软件技术也得到了迅速的发展，从而出现了专门的数据管理的软件，这就是所谓的“文件技术”阶段。

在文件技术阶段，数据管理的主要特点可概括为如下几点。

(1) 外存储器成为计算机系统不可缺少的组成部分，用户可以随时通过程序对文件进行查询、修改和增删等处理。

(2) 文件组织形式日益多样化, 索引文件、直接存取文件、链接文件相继出现, 既可以满足批处理应用的需要, 又能有效地实现记录的随机存取。

(3) 出现了专用数据管理系统的软件, 即文件系统, 它能对驻留在外存储器上的数据文件实施统一管理。这种专用数据管理软件构成了操作系统的一个重要组成部分。由于应用程序不再需要了解数据在存储介质上的实际地址, 因而大大减少了程序设计的工作量。

(4) 数据不再仅仅属于某个特定的程序, 而可以重复使用。

文件技术的上述特点使得这项技术在 20 世纪 60 年代中得到了充分的发展, 把计算机应用推向了一个新的高潮。但由于文件结构的设计仍然是基于特定用途的, 程序仍然是基于文件的特定物理结构和存取方法编制的, 因此, 数据结构与程序之间的依赖关系并未根本改变, 因而限制了它的进一步发展。

20 世纪 60 年代中期以后, 计算机在数据处理领域的应用迅速发展, 由个别部门的应用逐步发展成多个部门的普遍应用, 由简单孤立的单项应用发展为彼此相关的复杂应用, 从而使管理的规模更加庞大, 数据量急剧增长, 共享性也更增强了。这就带来了数据管理上的一些新问题, 现举一例加以说明。

例 1 某学校的学生处、教务处和卫生院均要使用计算机对学生的有关信息进行管理, 但各自处理的内容不同, 如用文件系统实现, 可按如下方式进行组织。

学生处要处理的信息包括: 学号、姓名、系名、年级、专业、年龄、性别、籍贯、政治面目、家庭住址、个人简历、社会关系……为此, 学生处的应用程序员必须定义一个文件 F1, 该文件结构中的记录应包括上述几个数据项。

教务处要处理的信息包括: 学号、姓名、系名、年级、专业、课名、成绩、学分、……显然, 教务处的应用程序员需定义一个文件 F2, 该文件结构中的每一记录包括以上几个数据项。

卫生院要记录和处理学生的有关健康情况信息时, 其创建的文件 F3 应包括下列数据项: 学号、姓名、系名、年级、专业、年龄、性别、身高、体重、健康状况、……

这样当上述三个部门共用某台计算机时, 该计算机的外存中要同时保存 F1、F2、F3 三种文件, 可这三种文件中均有学生的学号、姓名、系名、年级和专业等信息, 因此重复的数据项达到了 1/3 以上, 这就是数据冗余。数据冗余将会产生以下问题。

(1) 数据冗余不仅浪费存储空间, 更严重的是带来潜在的 inconsistency。由于数据存在多个副本, 所以当发生数据更新时, 就很可能发生某些副本被修改而另一些副本被遗漏的情况, 从而使数据发生不一致, 影响数据的正确性和可靠性。比如, 某学生因故需从数学系转到计算机系, 当学生处得到该信息后, 将该生所属的系名改为计算机系, 因而 F1 文件中保存了正确的信息。但若教务处和卫生院没有得到此信息, 或者没有及时更改 F2 和 F3 文件, 这就造成了数据的不一致性。由于数据的使用价值很大程度上依赖其可靠性, 所以, 这种不一致的后果是不可忽视的。可以想象, 当这种情况发生在军事、航天、金融等行业部门时, 其后果是非常严重的。

(2) 在传统文件技术阶段, 文件是为某一特定应用服务的, 应用程序和文件是一对一的 (如图 1.1 所示), 这就造成了应用程序与数据结构过分地互相依赖, 而且系统很难扩充。一旦逻辑结构改变, 就必须修改应用程序和文件结构的定义; 反之, 应用程序的改变, 也会影响文件的数据结构的改变。比如, 学生处在管理学生信息时发现, 只登记学生年龄是不够的, 还应记录学生的出生年月, 因此需要在 F1 的文件中加一个数据项, 这就必须修改 F1 的文件

结构，同时还要修改涉及 FI 文件的应用程序；另一方面，若用户原先是用 Pascal 语言编制的访问 FI 文件的应用程序，现想改用 C 语言，那么，FI 文件的数据结构也必须随之改变。由此可见在文件技术阶段，扩充已有的系统是很困难的，因此，用户迫切希望应用程序能够不受数据结构变化的影响。

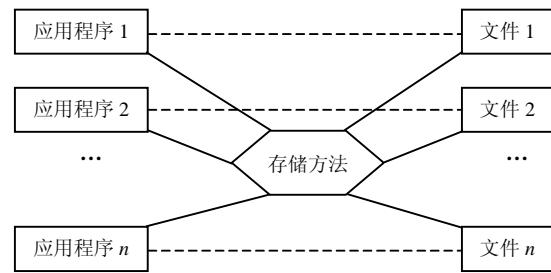


图 1.1 应用程序与文件之间一对一关系

(3) 文件系统缺乏对数据操作进行控制的方法，对于数据的安全性、保密性和正确性等方面的控制，完全要用户程序自己负责，这使得应用程序的编制相当繁琐。

综上所述，传统的文件技术有许多缺点，不能满足人们的要求，因此迫切需要新的数据管理技术来实现对数据的共享，实现数据与程序的独立性，并提供安全性和完整性。也就是说，在操作系统之上必须有一个软件系统——数据库管理系统(DataBase Management System, DBMS)，在数据库的建立、运用和维护时，对数据库进行统一的控制，这就是数据库技术。

1.1.2 数据库技术的产生与发展

数据库技术从 20 世纪 60 年代中期产生到现在也不过 40 年左右的时间，但其发展速度很快、使用范围很广。20 世纪 60 年代末出现了第一代数据库——网状数据库和层次数据库，20 世纪 70 年代出现了第二代数据库——关系数据库。目前关系数据库已逐渐淘汰了网状数据库和层次数据库，成为当今最为流行的商用数据库系统。而 20 世纪 80 年代出现的以面向对象为主要特征的数据库系统又向关系数据库系统提出了挑战。以下几个事件标志着数据库技术日益成熟的过程。

(1) 1969 年 IBM 公司研制开发了 DBMS 的商品化软件 IMS (Information Management System)，它是层次模型数据库系统的典型代表。

(2) 美国数据系统语言协商会 (The Conference On Data Systems Languages) 下属的数据库任务组 (Data Base Task Group, DBTG) 对数据库方法进行了系统的研究、讨论。于 20 世纪 60 年代末、70 年代初提出了若干报告，称为 DBTG 报告。DBTG 报告确定并建立了数据库系统的许多概念、方法和技术，它是网状模型数据库系统的典型代表。

(3) 1970 年 IBM 公司 San Jose 研究室的 E.F.Codd 在美国计算机学会会刊 “Communication of the ACM” 上发表了题为“大型共享数据库数据的关系模型”(A Relational Model of Data for Shared Data Banks) 的著名论文，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，为关系数据库技术奠定了理论基础。E.F.Codd 本人作为关系数据库的创始人和奠基人，获得了 1981 年 ACM 图灵奖。

自 E.F.Codd 提出关系模型数据库方法以后，1974 年 IBM 公司就提出了一种基于关系方法实现对数据库存取的 SQL (Structured Query Language) 语言，并经大约五年的努力，成功地研制了一个实现 SQL 语言的关系数据库系统原型 System R，该系统在 IBM 370 上运行。之后，IBM 又将 SQL 语言引入到 DB2 (IBM DataBase2) 中，配置在 MVS 上运行，并于 1983 年推出 DB2 产品，表明了关系模型数据库系统终于被人们所接受，并得到了最终的肯定。目前，有影响的关系模型数据库系统有许多，值得一提的 Oracle 系统采用 SQL 语言，用 C 语

言开发,使 Oracle 不仅使用方便、速度快且移植性好。1979 年开始推出 Oracle 第一版,它是由核心和一组支持软件组成的软件包。现在已推出第八版,增加了与各种语言、决策支持工具以及网络通信软件的接口,形成了开放式体系结构,是当前惟一可以通用于大、中、小和微型机的产品。关系数据库系统的出现,促进了数据库的小型化和普及化,使得在微型机上配置数据库系统成为可能。较流行的 Oracle、UNIFY、SQL Server 等较完善的关系数据库管理系统也都有其微型机版本。

(4) 从 20 世纪 80 年代以来,数据库技术在商业领域的巨大成功刺激了其他领域对数据库技术需求的迅速增长,这些新的领域为数据库应用开辟了新的天地。另一方面在应用中提出的一些新的数据管理的需求也直接推动了数据库技术的研究与发展,尤其是面向对象数据库系统(Object Oriented DataBase System)的研究与发展。1990 年美国高级 DBMS 功能委员会发表了“第三代数据库系统宣言”的文章,提出了第三代 DBMS 应具有的三个基本特征,即

- ① 第三代数据库系统应支持数据管理、对象管理和知识管理;
- ② 第三代数据库系统必须保持或继承第二代数据库系统的技术;
- ③ 第三代数据库系统必须对其他系统开放。

数据库技术与网络通信技术、人工智能技术、面向对象程序设计技术、并行计算技术等互相渗透、互相结合,成为当前数据库技术发展的主要特征。

我国数据库技术的发展大致始于 20 世纪 70 年代中、后期,其标志是 1977 年的第一次数据库技术研讨会。自 1982 年起,几乎每年由该学会下设的数据库学组召开一次全国性的学术交流会,会议起到了交流我国数据库应用与有关科研成果、指导我国数据库学术发展的作用。经过十多年的努力,我国数据库技术水平已从初期的学习、理解、试探阶段发展到消化、改造和创新阶段。

1.1.3 数据库技术的特点

与文件系统相比,数据库技术提供了对数据的更高级、更有效的管理,它有如下主要特点。

(1) 数据共享性。数据共享的意义是多种应用、多种语言互相覆盖的共享数据集合。在传统的文件组织中,每个数据文件是特定的应用所私有的,如图 1.1 所示。而数据库是从整体观点来看待和描述数据的,数据不再是面向某一应用,而是面向整个系统,如图 1.2 所示。这可以大大减少数据冗余度,既节约存储空间,减少存取时间,又可避免数据之间的不兼容性和不一致性。前面所举文件系统例子,若从数据库角度进行组织,则如图 1.2 所示。

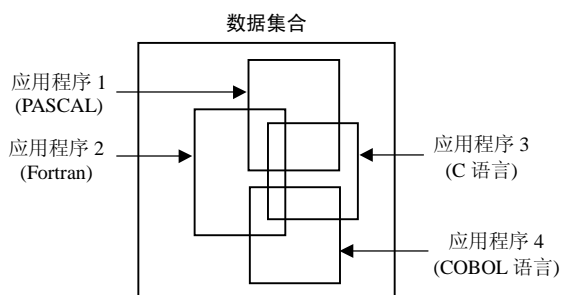


图 1.2 应用程序共享数据集合

由图 1.2 可看出,数据不再分属于各个应用程序,而是集中存储在一起。对一个特定的组织而言,除了由于需要保密而强加的某些限制以外,这些数据是为该组织的各个下属部门和各个方面的应用所共享的,这样就大大提高了数据的使用价值。在图 1.3 中,学生记录是为教务处、学生处和卫生院所共享的,若某个学生需要转系,则只要修改学生