



高等学校精品规划教材

数据库技术系列

主 编 程学先 程传慧

参 编 曾 玲 郑秋华 陈永辉 杨晓艳

数据库原理与技术 (第二版)



中国水利水电出版社
www.waterpub.com.cn

21 世纪高等学校精品规划教材

数据库原理与技术

(第二版)

主 编 程学先 程传慧

参 编 曾 玲 郑秋华 陈永辉 杨晓艳



中国水利水电出版社
www.waterpub.com.cn

内 容 提 要

本书在第一版的基础上加强了 SQL Server 与 Oracle 两种数据库管理系统软件的教学内容,突出在网络环境中数据库的应用研究。全书系统、全面地叙述了数据库系统的基本概念、基本原理和基本方法。内容包括数据库概念、数据模型、存储结构、关系数据理论和关系数据库的基本概念、SQL 语言、数据库应用系统设计、SQL Server 与 Oracle 数据库的基础知识与开发方法。还介绍了 C/S 与 B/S 模式、数据库保护、数据仓库、数据挖掘、分布式数据库和面向对象数据库等知识。

本书可作为高等院校应用型本科在校学生学习数据库理论与技术的教材,也可供研究生和从事计算机工作的科技工作者参考。

为方便自学,并迅速掌握数据库应用系统的开发方法,将随书公开发布教学幻灯片、部件库最小系统(局域网版)、数据库试题库。读者可以到中国水利水电出版社网站免费下载,网址: <http://www.waterpub.com.cn/softdown/>。

图书在版编目(CIP)数据

数据库原理与技术 / 程学先, 程传慧主编. —2 版. —北京: 中国水利水电出版社, 2009
21 世纪高等学校精品规划教材
ISBN 978-7-5084-6-6362-9

I. 数… II. ①程…②程… 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 (2009) 第 040528 号

策划编辑: 杨庆川 加工编辑: 李 冰 责任编辑: 李 炎 封面设计: 李 佳

书 名	21 世纪高等学校精品规划教材 数据库原理与技术 (第二版)
作 者	主 编 程学先 程传慧 参 编 曾 玲 郑秋华 陈永辉 杨晓艳
出版 发行	中国水利水电出版社 (北京市三里河路 6 号 100044) 网址: www.waterpub.com.cn E-mail: mchannel@263.net (万水) sales@waterpub.com.cn 电话: (010) 63202266 (总机)、68367658 (营销中心)、82562819 (万水)
经 售	全国各地新华书店和相关出版物销售网点
排 版	北京万水电子信息有限公司
印 刷	北京蓝空印刷厂
规 格	184mm×260mm 16 开本 18.25 印张 443 千字
版 次	2001 年 10 月第 1 版 2009 年 3 月第 2 版 2009 年 3 月第 7 次印刷
印 数	26001—30000 册
定 价	30.00 元

凡购买我社图书, 如有缺页、倒页、脱页的, 本社营销中心负责调换
版权所有·侵权必究

第二版前言

数据库是设计与建立管理信息系统的主要支撑，而管理信息系统是计算机应用最主要的内容之一。学习数据库的目的，除了学习其思想、方法之外，还要掌握它在管理信息系统中应用的理论与方法。要学好数据库，也必须与管理信息系统的建设密切联系，由管理信息系统的需求分析其对数据库技术的要求；联系管理信息系统的设计讨论应用数据库的方法；从管理信息系统的构成看数据库的组成与结构；根据管理信息系统的发展研究数据库的进步等。孤立地讲述数据库的概念、方法与技术会大大减少这门课程的价值，会使其理论变得枯燥无味与难以理解，会出现理论与实践相脱离的弊病。

本书第一版自问世以来，得到了广大读者的肯定，不少读者还提出了许多有益的建议，为新版的编写打下了良好基础，我们深表感谢。第一版教材基于 VFP 组织教学，使用了整章介绍 VFP 的界面及句法语法知识。目的是借 VFP 的简单、易学但又是一种基本具备关系数据库特征的数据库系统等特性帮助组织教学，将数据库的学习和实际应用系统设计的理论与技术比较紧密地结合起来，加深读者对数据库概念与方法的理解，了解数据库的应用，使学有所目的、学有所用。应当说，VFP 在我国数据库历史中是具有重要地位的，在相当一段时间内，它及其前身是大众主要使用的数据库，为数据库技术的启蒙、传播、推广、应用发挥了极大作用。一直到现在，许多工作人员仍能使用甚至利用 VFP 设计简单的应用系统，具有相当的数据库方面的知识能力，从一个侧面证明了当时教学内容与教学方法的成功。在将来，VFP 应当还会在数据库的大家族里占有一席之地。数据管理从机器语言到文件、到数据库、到关系数据库，还将发展成许多种新的或专门的数据库。VFP 是其中一种特殊的不完全的数据库。它结构简单、使用方便、汉化程度高，将开发语言嵌入到数据库之中，使其易学、易用、易开发，在单机桌面系统建设中具有明显优势。如果说类似于 Excel 这样的基于文件系统的数据处理工具还在、也还会继续广泛使用的话，VFP 毕竟是较文件系统高一级的系统工具，由于其相对其他大型数据库的学习门槛低、成本低、基本无环境要求、在我国有深厚的技术沉淀等特点，也一定还会有广阔的应用空间与市场。

但是，VFP 毕竟只是基于单机而设计的，安全性、数据共享能力差，其语言简单且不严密，与目前普遍使用的大、中型数据库存在许多不同。由于数据库课程学时数普遍紧张，为突出数据库理论与技术的教学，提高教学层次，本书再版时删除了涉及 VFP 语法、句法等有关章节的内容。为了保证教学质量，求取好的教学效果，本书再版时加强了原书 SQL Server 与 Oracle 两种数据库管理系统软件的教学内容，突出在网络环境中数据库的应用研究，作为全书应用部分的主线。希望加强基于 SQL Server、Oracle 的实验教学，开展如涉及数据库与数据表的设计与管理、SQL 语句及内嵌语言的应用、数据完整性与安全性保护、备份与恢复等方面的实验。这些内容既扩展了数据库的基本理论与技术，也使数据库理论教学落实到实践之中。为了延续上一版风格，重视教学工具的使用，我们将在中国水利水电出版社网站上推出部件库最小系统（局域网版），利用部件库最小系统组织实验教学。

部件库最小系统是在本书第一版所配光盘中框架性软件（查询部件与涉及维护表单公用

件)的基础上发展起来的。目前公开发表的版本仍然基于 VFP。但将之用于数据库教学,可以不要求学习 VFP,可以不懂 VFP 的语法与句法,只要求安装 VFP 6.0 系统软件,除了可以在 SQL Server 2000 或其他具有 ODBC 接口的数据库上操作之外,所有实验与应用都可以脱离 VFP 界面进行。可以利用该系统在几分钟到几十分钟的时间里搭建一个局域网上的功能比较齐全的管理系统,这个系统将能拥有多种数据的录入与维护程序、有能足够满足需要的查询程序、多种数据处理程序、多种数据导入或导出的通信程序、多种打印与图形输出程序。可以让学生联系具体应用系统需求实地认识数据冗余、共享、数据独立性、多类数据完整性及数据完整性保护、关键字、视图、数据安全、SQL 语言及其应用、数据表结构及其对系统设计的影响、代码表与派生数据及其处理等基本概念、基本理论与基本方法。可以让学生结合数据库设计实地开发应用系统,通过操作去分析数据库设计的正确性与合理性,更深入切实地掌握数据库设计的理论与方法,从而大大提高数据库的学习质量与动手能力。

书中部分章、节前加有“*”号,可作为选讲或选学内容。

软部件是应用系统中由原子构件组成的能独立完成一项工作的具有自适应与即插即用特性的通用程序模块。随着部件与构件技术的发展,在管理信息系统建设时,参与原始代码设计的人员将减少,大部分开发人员的主要工作将集中到数据库设计、应用系统结构研究、系统扩展与维护等工作上来,我们希望数据库课程的教学内容与方法能与这一发展过程相适应。我们目前的研究还处于早期阶段,缺点与错误在所难免,希望广大读者多提宝贵意见。

为配合教学,我们还将在中国水利水电出版社网站上公开发布教学幻灯片、数据库试题库系统等。

本书由程学先、程传慧主编,参加编写的有曾玲、郑秋华、陈永辉、杨晓艳等。程传慧承担第 1、2 章、曾玲承担第 3、4 章、陈永辉承担第 5、6 章、郑秋华承担第 7、8 章、杨晓艳承担第 9 章的编写及相应章节教学幻灯片设计与试题设计,程学先、程传慧负责全书统稿与修改,程学先、程传慧、王莉、程传庆等承担部件库最小系统等软件程序设计。林姗、刘伟、胡显波、赵岚、江南、肖模艳、龚晓明、王富强、陈义、史函、刘玲玲、熊晓菁、童亚拉、周金松、祝苏薇、王嘉、黎柳柳、苏艳、蒋慧婷、陈莉、谌章恒、高霞、张军、赵普、钱涛、张俊、李珺等参加了本书编审与软件设计等工作,特此表示感谢。

编者

2008 年 12 月

第一版前言

数据库技术是一门发展迅速，在科学与社会各个领域广泛应用的技术，是计算机科学的重要分支，是建立各类信息系统的基础，目前其应用已从信息检索、一般管理扩大到科技计算、计算机辅助设计、人工智能等各个领域。

我们在多年从事数据库系统教学与科研的基础上编写了这本教材，力求由浅入深，由具体到抽象，理论与实践紧密相结合，既系统完整又突出应用，使读者能了解什么是数据库及数据库要达到的目标，掌握数据库的设计方法并了解实际数据库应用系统的组成及利用 Visual FoxPro 关系数据库系统设计与实现的方法。

本书内容包括数据库原理，数据库应用系统设计技术，并介绍了可视化与面向对象技术与理论，符合《IEEE-CS》与 ACM 计算机教程联合工作组推出的《CC2001》教程关于计算机教学大纲的要求。

本书开始即对比了应用 C 语言进行数据处理的程序编写方法和 VFP 环境下数据存储与应用程序编写方法的不同，导出数据库的基本概念，并且说明了数据库理论中一些至关重要的概念。通过对不同数据逻辑结构与物理结构的介绍，让读者更切实地体会要完成数据库系统的目标需要解决的矛盾。

在书中将关系代数与 SQL 查询语言作对照性的介绍，使理论与实际高度结合，既方便理解又方便应用。

在介绍面向对象技术时，介绍了知识是如何发展的，以前是怎样做的，现在怎样做，启发读者去思考科学今后将如何发展，生活会向我们提出什么要求、什么课题，希望启发读者进行创新式的思维。

在数据库应用系统设计中，我们以真实的系统为例，尽量要求内容的全面与完整。在介绍 VFP 程序设计时，从最基本、最简单的程序开始，以方便读者入门，在此基础上又讲述了有实际应用价值的复杂结构及其应用。目的是要求读者不仅仅停留在语法、概念、原理和方法上，而且要清楚在实际系统中如何使用这些理论知识。我们希望读者能设计出有实际意义的数据库应用系统。

随书所附光盘既是范例又具有实用价值。将其中的某些模块与对象拷贝到你的系统中，可让学完本书的初学者能在几个小时内用 VFP 设计出简单的数据库应用系统。但我们还是希望读者真正能了解每个模块的结构，理解它们的设计目标、设计方法与设计技巧，对其融会贯通，使自己能够熟练掌握数据库应用系统的设计，真正成为一名设计大师。

本书还介绍了 SQL Server 和 Oracle 的相关知识，希望借之提出并介绍数据库系统中一些重要的理论与技术。

本书可作为本、专科及在职职工学习数据库理论与技术的教材，也可供研究生学习参考。

本书由程学先教授任主编，由林姗、张明武、徐晓任副主编，参加编写的还有王莉、曾玲、程传庆、谭再峰等。本书所附光盘程序由程学先、林姗、程传慧设计。PowerPoint 演示文稿由曾玲设计制作。刘伟、曾立平、程传慧、雷飞宇、陈永辉、李江平、江南等参加了前期部分工作。

编者

2001年6月

目 录

第二版前言

第一版前言

第 1 章 数据库基础知识	1
本章学习目标	1
1.1 数据处理	1
1.1.1 利用文件系统进行数据处理.....	1
1.1.2 从实例看数据库的数据处理技术.....	4
1.2 数据库技术概述	9
1.3 数据库的数据结构及存储结构.....	14
1.3.1 链表式数据结构	14
1.3.2 关系数据库结构概述	15
1.4 索引文件组织	18
1.4.1 索引文件	18
1.4.2 非关键字索引文件	19
1.4.3 B+树索引结构	21
本章小结	22
习题一	22
第 2 章 数据库设计中的数据模型	23
本章学习目标	23
2.1 数据模型	23
2.1.1 数据模型概念	23
2.1.2 数据之间的联系	24
2.1.3 实体—联系模型	25
2.2 关系数据模型	28
2.2.1 关系数据模型的概念	28
2.2.2 关系数据模型的设计	29
2.3 面向对象数据模型	30
2.3.1 UML 定义的类型图.....	31
2.3.2 利用 Rose 建模操作	35
2.3.3 从建模到建库与建表的自动化操作.....	36
本章小结	38
习题二	38
第 3 章 关系数据库	39
本章学习目标	39

3.1	基本概念	39
3.2	函数依赖	40
3.2.1	函数依赖概念	40
3.2.2	部分函数依赖	41
3.2.3	完全函数依赖	41
3.2.4	传递函数依赖	41
3.3	候选关键字与主属性	42
3.3.1	候选关键字	42
3.3.2	主属性	43
3.4	关系规范化	43
3.4.1	问题的提出	43
3.4.2	范式	45
3.4.3	关系分解的正确性	50
	本章小结	50
	习题三	51
第4章	关系数据库实现及使用	52
	本章学习目标	52
4.1	关系代数	52
4.1.1	传统的集合运算	53
4.1.2	专门的关系运算	55
*4.2	关系演算	58
4.3	SQL 语言概貌	59
4.4	SQL 数据定义功能	60
4.4.1	基本表的定义和修改	60
4.4.2	索引的建立和删除	61
4.5	SQL 数据查询语句	61
4.5.1	语句格式	61
4.5.2	对单一表查询语句	63
4.5.3	对两个以上表的连接查询	64
4.5.4	嵌套查询	64
4.5.5	关系除法	65
4.6	SQL 查询语句的加强	66
4.7	视图	68
4.8	SQL 数据更新语句	69
4.8.1	修改 (UPDATE) 语句	69
4.8.2	删除 (DELETE) 语句	69
4.8.3	插入 (INSERT) 语句	69
4.8.4	视图的更新	70
4.9	SQL 数据控制功能	70

4.10	嵌入式 SQL	71
4.11	查询优化	72
	本章小结	74
	习题四	74
第 5 章	数据库的保护	77
	本章学习目标	77
5.1	数据库的完整性实施方法	77
5.2	数据库的安全性实施方法	78
5.2.1	定义视图	78
5.2.2	访问权限控制	78
5.3	事务处理	80
5.3.1	事务的基本概念	80
5.3.2	事务处理过程分析	81
5.3.3	SQL 的事务管理	81
5.4	并发控制	82
5.4.1	并发处理产生的三种不一致性	82
5.4.2	封锁	84
5.5	数据库的备份与恢复	86
5.5.1	故障的类型	86
5.5.2	事务日志	87
5.5.3	恢复	87
5.5.4	数据的转储	88
	本章小结	88
	习题五	89
第 6 章	SQL Server 数据库系统	91
	本章学习目标	91
6.1	SQL Server 管理工具	91
6.1.1	服务管理器	91
6.1.2	建立数据库、表、索引的操作	92
6.1.3	建立视图的操作	98
6.1.4	数据完整性保护	101
6.1.5	备份与恢复	102
6.1.6	数据库安全性管理	105
6.1.7	查询分析器	112
6.2	Transact-SQL 语言简介	114
6.2.1	数据定义语言 (DDL)	114
6.2.2	数据操纵语言 (DML)	118
6.2.3	数据控制语言 (DCL)	119
6.2.4	其他语言元素	121

6.3	SQL Server 中的存储过程.....	126
6.3.1	存储过程的概念	127
6.3.2	存储过程的优点	127
6.3.3	使用 Transact-SQL 命令创建存储过程	128
6.3.4	使用企业管理器创建存储过程.....	130
6.3.5	重新命名存储过程	131
6.3.6	删除存储过程	132
6.3.7	执行存储过程	132
6.3.8	系统存储过程	132
6.4	SQL Server 中的触发器.....	133
6.4.1	触发器的概念及作用	133
6.4.2	触发器的种类	134
6.4.3	创建触发器	134
6.4.4	触发器的原理	137
6.4.5	INSTEAD OF 触发器	138
6.4.6	触发器的应用	139
6.4.7	管理触发器	141
*6.5	SQL Server 中的数据导入和导出.....	143
6.5.1	使用 Transact-SQL 进行数据导入、导出	144
6.5.2	使用命令行 bcp 导入、导出数据.....	144
6.5.3	使用数据转换服务 (DTS) 导入、导出数据	145
6.5.4	如何选择具体的数据导入、导出方法.....	146
*6.6	SQL Server 应用系统开发环境.....	147
6.6.1	SQL Server 应用系统的两种系统结构.....	147
6.6.2	ODBC	147
6.6.3	ADO.....	152
6.6.4	JDBC.....	153
	本章小结	157
	习题六	158
第 7 章	数据库应用系统设计	160
	本章学习目标	160
7.1	概述	160
7.2	数据库结构设计	161
7.2.1	数据库结构设计步骤	161
7.2.2	需求分析	161
7.2.3	概念结构设计	166
7.2.4	逻辑结构设计	169
7.2.5	数据库物理设计	170
7.3	应用程序结构设计	172

本章小结	175
习题七	176
*第 8 章 Oracle 数据库系统	177
本章学习目标	177
8.1 Oracle 数据库概述	177
8.1.1 Oracle 10g 的特点	177
8.1.2 Oracle 10g 的安装和基本管理	178
8.2 Oracle 10g 的体系结构	185
8.2.1 概述	185
8.2.2 物理存储结构	185
8.2.3 逻辑、存储结构	186
8.2.4 内存结构	187
8.2.5 数据字典	188
8.3 Oracle 10g 的安全管理	189
8.3.1 用户管理和权限管理	189
8.3.2 角色管理	190
8.3.3 配置文件 PROFILE 管理	191
8.4 Oracle 10g 的空间管理	191
8.4.1 表空间	191
8.4.2 段	193
8.4.3 表	193
8.4.4 索引	196
8.4.5 视图	196
8.5 Oracle 10g PL/SQL 编程	197
8.5.1 PL/SQL 简介	197
8.5.2 PL/SQL 语言基础	197
8.6 Oracle 数据查询	210
8.7 Oracle 系统管理	212
本章小结	217
习题八	217
*第 9 章 其他数据处理技术概述	219
本章学习目标	219
9.1 数据仓库	219
9.1.1 数据仓库的概念	220
9.1.2 数据仓库的架构	221
9.1.3 数据收集	222
9.1.4 数据集市	224
9.1.5 数据仓库系统开发	225
9.1.6 联机分析 (OLAP) 技术概述	231

9.1.7	SQL Server 2000 中的数据仓库组件.....	233
9.2	数据挖掘	233
9.2.1	数据挖掘技术概述	234
9.2.2	数据挖掘的定义	235
9.2.3	数据挖掘的过程模型及常用技术.....	235
9.2.4	目前数据挖掘的主要应用.....	236
9.3	面向对象的数据库技术.....	237
9.3.1	面向对象数据模型	237
9.3.2	面向对象数据库系统中的程序设计语言	239
9.3.3	一个持久化程序设计语言实现面向对象数据库的实例	240
9.4	分布式数据库	246
9.4.1	分布式数据库系统概述	246
9.4.2	分布式数据存储	247
9.4.3	分布式数据的查询处理	248
9.4.4	分布式数据库系统中的事务处理.....	248
9.4.5	数据对象的命名方式与目录表的管理.....	252
9.4.6	更新传播	253
9.5	XML	253
9.5.1	HTML 概念	254
9.5.2	XML 文档设计.....	256
	本章小结	259
	习题九	259
	附录	261
	参考文献	279

第 1 章 数据库基础知识

本章学习目标

本章联系数据库管理系统的发展历程介绍数据库管理系统的基本概念,包括用文件管理系统进行数据管理的缺点,数据库管理系统的优点,数据库管理系统的组成,数据库的数据结构和存储结构等。通过本章学习,读者应该掌握以下内容:

- 数据库管理系统与传统的文件管理系统的主要区别与各自的特点
- 数据库管理系统(DBMS)的组成与功能
- 数据库的数据结构和存储结构
- 索引的概念,索引文件类型
- 索引文件结构与使用
- B+树结构及应用

当今时代是信息技术飞速发展的时代。所谓信息,是以数据为载体的客观世界实际存在的事物、事件或概念在人们头脑中的反映。信息系统是以计算机为核心,以数据库为基础,对信息进行收集、组织、存储、加工、传播、管理和使用的系统。数据库能借助计算机保存和管理大量复杂的数据,快速而有效地为多个不同的用户和各种应用程序提供需要的数据,以便人们能更方便更充分地利用这些宝贵的信息资源。

数据管理是指数据的收集、整理、组织、存储、查询、维护和传送等各种操作,是数据处理的基本环节,是任何数据处理任务必有的共性部分。因此应当开发出既通用又方便好用的软件,把数据有效地管理起来,以便最大限度地减轻程序员的负担。数据库技术正是针对这一目标逐渐完善的一门计算机软件技术。它所研究的问题就是如何科学地组织和存储数据,如何高效地获取和处理数据,如何更广泛、更安全地共享数据。

1.1 数据处理

1.1.1 利用文件系统进行数据处理

根据计算机软件和硬件的发展,数据管理技术的发展大体上分为三个阶段:人工管理阶段、文件系统阶段和数据库阶段。

1. 人工管理阶段(20世纪50年代中期以前)

20世纪50年代中期以前,计算机主要用于科学计算。当时尚无操作系统与高级语言,软件采用机器语言编写。在科学计算公式中用到一些数据,但数据量很小,一般不对它们进行添加、修改、删除等维护操作,也不要求检索,因此无论从软件开发环境,还是从应用需求上都

将它们与程序紧密结合在一起，对这些数据不需要共享，而且当时也不允许共享。

2. 文件系统阶段（20 世纪 50 年代后期至 60 年代）

20 世纪 50 年代后期到 60 年代中期，随着计算机科学的发展，这时硬件上已有了磁盘、磁鼓等直接存取存储设备；软件方面出现了操作系统及高级语言。操作系统中有专门的文件管理软件，称为文件系统，处理方法上不仅有了文件批处理，而且能够联机实时处理。计算机应用也从单纯科学计算、控制，扩大到电子数据处理系统，包括电算系统（例如工资系统与成本会计计算）、统计系统（例如国民经济数据统计系统）、数据更新系统（例如飞机预订订票系统）等，数据量大大增加，且同一组数据往往要求用于不同的计算和统计之中，以供不同客户查询，存在修改变动了的数据、根据用户的需要添加新的数据、删除一些过时无用的数据等维护操作。为便于对数据进行维护，也方便不同用户查询的需求，人们利用文件系统将数据从程序中分离出来形成专门的数据文件。例如将两件商品（Commodity）的有关数据：商品名称（WareName）、规格（Specification）、单价（Unitprice）、说明（Illuminate）等输入到一个文件中的 C 语言程序如例 1.1 所示。

【例 1.1】使用 C 语言将 Commodity 的有关数据：WareName、Specification、Unitprice、Illuminate 等数据输入到一个文件中。

```
# include "stdio.h"
main()
{ FILE *fp;
  fp = fopen("Commodity.c","w");
  fputs("Silverware ",fp);
  fputs ("Wwmottle      ",fp);
  putw(40,fp);
  fputs("11111111      ",fp);
  fputs("Chinaware   ",fp);
  fputs("Popularware   ",fp);
  putw(100,fp);
  fputs("222222222222      ",fp);
  fclose(fp);
}
```

如要显示文件中的数据，可使用例 1.2 所示程序。

【例 1.2】显示文件 Commodity.c 中的数据。

```
#include "stdio.h"
#define SIZE 2
struct Commodity _type
{ char WareName [12];
  char Specification [16];
  int Unitprice;
  char Illuminate [20];
}stud[SIZE];
main()
{ int i;
  FILE *fp;
  fp=fopen("Commodity.c","r");
  for(i=0;i<SIZE;i++)
    {fread(&stud[i], sizeof(struct Commodity_type),1,fp);
```

```
printf("%28s%20s %8d %200s %\n", stud[i].WareName, stud[i].Specification,  
stud[i].Unitprice, stud[i].Illuminate);  
}  
}
```

如果要按一定条件显示一定范围的数据，上述程序需略作修改：对每一组数据逐一判别是否满足要求，再按预定范围显示。如果要修改文件中的数据，要首先使用类似于例 1.2 的程序，以读的方式打开文件，将文件中数据读入到变量中，修改变量中数据的值，再以类似于例 1.1 的程序以写的方式打开文件，将修改后的数据输回到文件中。数据处理全过程如图 1.1 所示。

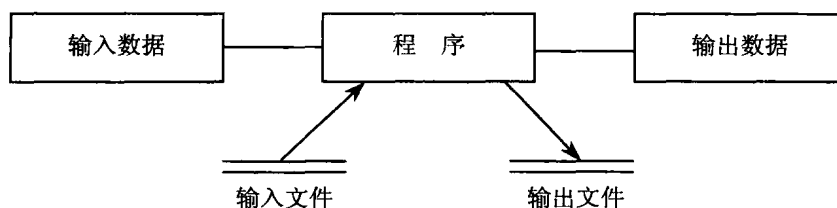


图 1.1 传统数据处理模式

使用数据文件实现了数据与程序相分离，分别采用两个文件各自存放数据与程序，我们将之称为实现了数据的物理独立。

利用文件系统解决程序和文件数据的存取操作，程序员只需关心文件的逻辑结构，无须关心如何转为物理存储，从而使程序设计变得简单，还可专门对数据文件进行管理，单独对数据进行操作或维护。不同的程序只要以读的方式打开数据文件，并按数据存放的格式将数据取出便可使用这些数据，在一定的范围内可以做到数据为不同程序所共享。数据可长期保存，大大方便了用户的使用。在这一阶段，文件越来越多样化和结构化，如出现了方便查询与直接存取的索引文件，方便多条件查询的倒排表文件，将内容联系起来的链接文件、线性表文件等各种可适应不同应用需要的文件。

使用数据文件的缺点是如果要使用数据和维护数据，必须知道数据存放的格式，即要知道数据存取的逻辑结构。例如在前述例 1.2 中，必须知道共存放了两件商品的记录，且每个商品的记录包括 4 个数据：WareName、Specification、Unitprice、Illuminate。其中第一、二、四 3 个数据是字符串类型数据，宽度为 12、16 与 20；第三个数据是整型数据。数据个数、数据类型与宽度必须与数据文件中的数据一一对应，否则，程序所读出的数据将会出错。其原因显然是因为数据结构和程序语句紧密相关，或者说数据和程序之间缺少数据逻辑独立性。

在实际系统中，由于数据值及逻辑结构都可能不断变化，如果每次变化都要求所有应用程序作相应修改，其工作量之大实在令人无法承受。而且，对这些数据文件要么控制不让打开，一旦能够打开，每个程序都可取用全部数据，所有人使用数据的权限都相同，完全无安全性可言。由于数据文件中除了数据不再有其他信息，也就无法对数据作统一的控制和管理。基于以上原因，这样的系统中数据共享就只能局限在一定范围内。同一数据常需在多个地方同时存放，数据冗余度大且无法有效控制，一方面浪费存储空间，降低运行效率；另一方面降低系统可靠性、正确性，降低系统价值，对系统进行修改、维护都麻烦。

同一数据在多个地方同时存放，那么同一数据在不同存放地的值可能不相同，称为数据不一致，这将会降低信息价值，甚至造成重大损失。

从数据处理需求来看，计算机辅助管理的内容逐步向纵深发展，应用面越来越广，部门的壁垒开始打破，数据不再只是用于计算、统计，而作为信息的载体被存储，成为人们宝贵的信息资源，用于检索、统计、预测及决策。要求对人工管理过程广泛予以模拟；要求收集并保存大量的数据；要求不断对历史上收集的数据进行筛选、分析和提炼；要求为决策提供大量的数据，并产生不同决策方案以供决策者参考；要求随时将各类信息向各方面发布等。这就要求系统效率更高，实现更大范围的数据共享。

数据共享是指同一数据能为同一用户或不同用户的相同或不同程序所使用。

目前广泛使用的 Excel 的一大弱点是数据共享难度大。在管理信息系统中，信息代码是保证管理规范化、标准化，方便统计操作的基本工具，但在 Excel 中很难进行多数据表的相关联操作，无法直接进行代码变换。在管理信息系统中可以十分容易地根据应用的每一需要设计专门的处理程序，用户只需要用鼠标做极少的点击就能操作，学习与使用都十分方便，而使用 Excel 就很难做到这点，稍微复杂一些的操作，难度就迅速加大。管理信息系统强调确保数据的安全与准确，而对于 Excel 表，要么不能打开，要么一旦打开所有内容就都可读、可修改。学习与使用 Excel 确实很简单，不需要预先建立数据结构，这是优点，但也是其缺点的根源。

3. 数据库系统阶段（20 世纪 60 年代后期）

20 世纪 60 年代后期以来，计算机用于管理的规模更为庞大，应用也越来越广泛，数据量急剧增长，同时多种应用、多种语言互相交叉，共享数据的要求越来越强烈。这时硬件也飞快发展，大容量磁盘出现，硬件价格下跌，软件价格上升，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，以文件系统作为管理手段已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，就出现了数据库技术，出现了统一管理数据的专门软件系统：数据库管理系统。数据库系统使前述例 1.1、例 1.2 的工作变得十分简单，也解决了文件系统中存在的许多问题。

1.1.2 从实例看数据库的数据处理技术

当前关系数据库管理系统的产品主要有 Oracle、DB2、Informix、Sybase ASE、SQL Server、MySQL、达梦、Access、Visual FoxPro 等，Oracle 是大型数据库，SQL Server 是可以应用于网上的小型数据库，Access、Visual FoxPro 是桌面式数据库，达梦是我国自行开发的国产数据库管理系统等。

如果在 SQL Server 中要完成前述任务，首先要建立数据存储结构。可先建立数据库，例如建立名为 Waremanage 的数据库，当打开该数据库后，可使用下述命令建立空表结构：

```
CREATE TABLE Commodity(WareName char(12),Specification char(16),  
Unitprice int,Illuminate text);
```

如要存入两件商品数据，可使用下述命令：

```
INSERT INTO Commodity VALUES("Silverware", "Wwmottle",40, "11111111111111");  
INSERT INTO Commodity VALUES("Chinaware", "Popularware",100, "22222222222222222222");
```

与 C 语言程序不同，在输入商品名称、规格、说明等数据时，数据值的宽度不到预定的字符宽度时，无须加填充格；在数据维护过程中对数据格式要求已经不是那么严格了。

例如要显示表 Commodity 中的内容，可使用下述命令：


```
SELECT * FROM Commodity
```

如要求按一定条件显示一定范围内的内容,例如只显示单价为 40 的商品的名称、规格与单价,可使用下述命令:

```
SELECT WareName,Specification, Unitprice FROM Commodity WHERE Unitprice =40
```

显然,采用数据库系统管理软件后,有以下优点:

(1) 程序极为简单,在 C 语言中需要大量语句才能实现的功能,目前仅需一两句即可完成。对数据的每一种维护往往也只需要一两行语句即可实现。

(2) 程序中不需要了解数据的数量和顺序,往往只需要知道准备操作的那部分数据的名字。

(3) 可直接对数据的某一部分分量进行操作,而无须知道全面的数据结构及其他分量的个数、名字和数据类型。

(4) 只要初始定义的数据结构所需的那一部分分量的名字和数据类型不发生变化,程序将无须随数据结构改变而修改。

这些优点为用户和程序员均带来方便,也使数据广泛共享真正成为可能。

数据库系统管理软件是如何实现上述功能的呢?对于上面的例子,在 SQL Server 2000 中利用 SQL Server 管理工具中企业管理器建立数据库 waremanage 之后会产生:主数据文件 waremanage_Data.MDF 与日志文件 waremanage_Data.LDF 两个文件。如果在其中用上述 CREATE 命令建立数据表“Commodity”之后,我们查看 waremanage_Data.MDF 中的内容,可以看到,其中包含了如下内容(如图 1.2 所示)。

```

33BA0 30 00 2E 00 78 2E 50 15-AF 02 AF 00 1A 00 00 01 0...x.P.....
33BB0 01 00 04 00 00 00 00-00 00 00 00 00 00 00 .....$.....
33BC0 00 00 01 00 FF FF 24 D0-00 00 20 00 20 00 15 00 .....$.....
33BD0 00 00 02 01 00 47 00 77-00 61 00 72 00 75 00 6E .....G.w.a.r.u.n
33BE0 00 61 00 6D 00 65 00 00-30 00 2E 00 78 2E 50 15 .a.m.e..0...x.P.
33BF0 AF 02 AF 00 14 00 00 00-02 00 1E 00 00 00 00 .....$.....
33C00 00 00 00 00 00 00 00-00 00 02 00 FE FF 24 D0 .....$.....
33C10 00 00 20 00 20 00 15 00-00 00 02 01 00 49 00 73 .. .I.s
33C20 00 65 00 63 00 66 00 63-00 61 00 74 00 6F 00 6E .e.c.f.c.a.t.o.n
33C30 00 51 9F 19 30 00 2E 00-78 2E 50 15 38 00 38 00 .Q..0...x.P.8.8.
33C40 04 00 0A 00 03 00 32 00-00 00 00 00 00 00 00 .....2.....
33C50 00 00 00 00 00 00 03 00-FD FF 00 00 00 00 20 00 .....$.....
33C60 20 00 15 00 00 00 0A 01-00 49 00 75 00 6E 00 69 .....I.u.n.i
33C70 00 74 00 70 00 72 00 69-00 63 00 65 00 00 00 00 .t.p.r.i.c.e....
33C80 30 00 2E 00 78 2E 50 15-23 00 23 00 10 00 00 00 0...x.P.#.#.....
33C90 04 00 FF FF 00 00 00 00-00 00 00 00 00 00 00 .....$.....
33CA0 00 00 04 00 FC FF 24 D0-00 00 20 00 20 00 15 00 .....$.....
33CB0 00 00 02 01 00 4B 00 69-00 6C 00 6C 00 75 00 6D .....K.i.l.l.i.u.m
33CC0 00 69 00 6E 00 61 00 74-00 65 00 00 30 00 2E 00 .i.n.a.t.e..0...
33CD0 B1 52 44 16 AF 02 AF 00-1A 00 00 00 01 00 04 00 .RD.....

```

图 1.2 数据表文件 MDF 中关于表结构定义的内容的示意

每个字段的特性用一段数据描述,用 30 00 表示对一个字段属性描述的开始;第 4 字节起的 4 个字节(例如 78 2E 50 15 代表 15502E78)是关于该字段所属数据表的描述;第 8 字节起的 3 个字节(例如 AF 02 AF)描述该字段的数据类型;第 12 字节存放字段宽度,用十六进制数字表示(例如 1A 表示宽度为 26);第 16 字节存放该字段在数据表字段中的顺序号;第 51 字节说明该字节的数据是以二进制还是以 ASCII 码形式存放,02 表示以 ASCII 码形式存放,0A 表示以二进制数存放;第 55 字节起开始存放字段名称。

在存放数据时,数据库将数据分为不同类型,不同类型的数据存放格式不同、将来允许进行的处理也不相同。数据在计算机内存放有二进制与 ASCII 码两种形式,前者适于存放如