



全国高等教育自学考试指定教材 计算机及应用专业(独立本科段)

数据库系统原理

附：数据库系统原理自学考试大纲

课程代码
4735
[2007年版]

组编 / 全国高等教育自学考试指导委员会
主编 / 丁宝康

本教材附赠网络学习卡

经济科学出版社

全国高等教育自学考试指定教材
计算机及应用专业(独立本科段)

数据库系统原理

(附:数据库系统原理自学考试大纲)

(2007年版)

全国高等教育自学考试指导委员会 组编

丁宝康 主编

丁宝康 杜 贇 乔 健 编著

经济科学出版社

图书在版编目 (CIP) 数据

数据库系统原理: 附数据库系统原理自学考试大纲 /
丁宝康主编. —北京: 经济科学出版社, 2007. 3
全国高等教育自学考试指定教材. 计算机及应用专业
(专科)
ISBN 978 - 7 - 5058 - 6122 - 0

I. 数… II. 丁… III. 数据库系统 - 高等教育 - 自学考试 - 教材 IV. TP311.13

中国版本图书馆 CIP 数据核字 (2007) 第 020971 号

责任编辑: 沙超英
责任校对: 杨海
版式设计: 代小卫
技术编辑: 邱天

数据库系统原理

附: 数据库系统原理自学考试大纲
(2007 年版)

全国高等教育自学考试指导委员会 组编
丁宝康 主编

经济科学出版社出版

社址: 北京市海淀区阜成路甲 28 号 邮编: 100036

网址: www.esp.com.cn

电子邮件: esp@esp.com.cn

北京友谊印刷有限公司印刷

北京友谊印刷有限公司装订

787 × 1092 16 开 19.75 印张 470000 字

2007 年 2 月第一版 2007 年 3 月第 1 次印刷

印数: 000001 - 010100 册

ISBN 978 - 7 - 5058 - 6122 - 0/F·5383 定价: 30.00 元

(图书出现印装问题, 请与当地教材供应部门调换)

(版权所有 翻印必究)

组编前言

21世纪是一个变幻莫测的世纪，是一个催人奋进的时代，科学技术飞速发展，知识更替日新月异。希望、困惑、机遇、挑战，随时随地都有可能出现在每一个社会成员的生活之中。抓住机遇、寻求发展、迎接挑战、适应变化的制胜法宝就是学习——依靠自己学习、终生学习。

作为我国高等教育组成部分的自学考试，其职责就是在高等教育这个水平上倡导自学、鼓励自学、帮助自学、推动自学，为每一位自学者铺就成才之路。组织编写供读者学习的教材就是履行这个职责的重要环节。毫无疑问，这种教材应当适合自学，应当有利于学习者掌握、了解新知识、新信息，有利于学习者增强创新意识，培养实践能力，形成自学能力，也有利于学习者学以致用，解决实际工作中所遇到的问题。具有如此特点的书，我们虽然沿用了“教材”这个概念，但它与那种仅供教师讲、学生听，教师不讲、学生不懂，以“教”为中心的教科书相比，已经在内容安排、编写体例、行文风格等方面都大不相同了。希望读者对此有所了解，以便从一开始就树立起依靠自己学习的坚定信念，不断探索适合自己的学习方法，充分利用已有的知识基础和实际工作经验，最大限度地发挥自己的潜能，以达到学习的目标。

欢迎读者提出意见和建议。

祝每一位读者自学成功！

全国高等教育自学考试指导委员会

2007年1月

内 容 简 介

《数据库系统原理》是全国高等教育自学考试指导委员会组编的计算机及应用专业（独立本科段）的教材。

本书详细介绍了数据库系统的基本原理、方法和应用技术。内容包括：数据库系统基本概念、数据库设计和 ER 模型、关系模式设计理论、关系运算、SQL 语言、数据库管理、SQL Server 2000 简介及应用、PowerBuilder 9.0 简介及应用、数据库技术的发展。

本书内容丰富，概念阐述细致清楚。书中编写了丰富的例题和习题。本书是自学考试有关专业的教材，也可作为从事信息领域工作的科技人员的参考书。

编者的话

《数据库系统原理》是根据全国高等教育自学考试计算机及应用专业（独立本科段）课程考试大纲组织编写的自学考试教材。

数据库技术产生于20世纪60年代末。经过40年的发展，数据库技术已形成完整的理论体系和一大批实用系统。数据库管理已经从一种专门的计算机应用发展成现代计算环境的一个核心部分。数据库应用领域也已从数据处理、信息管理、事务处理扩大到计算机辅助设计、人工智能、决策支持和网络应用等新的应用领域。数据库系统的推广使用使得计算机应用迅速渗透到国民经济各个部门和社会的每一个角落，并改变着人们的工作方式和生活方式。

数据库技术作为计算机软件领域的一个重要分支，是计算机科学技术中发展最快的领域之一，也是应用最广泛的技术之一，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。人们充分认识到，数据库是信息化社会中信息资源管理与开发利用的重要基础。对于一个国家，数据库的建设规模、使用水平已成为衡量该国信息化程度的重要标志。因此，数据库是计算机技术中一门重要的课程。

本书的主要目的是使读者掌握数据库技术的基本原理、方法和应用技术，能有效地使用现有的数据库管理系统和软件开发工具，掌握数据库结构的设计和数据库应用系统的开发原理。

全书共分9章，具体内容如下：

第1章介绍数据库系统的基本概念，包括数据描述、数据模型、全局结构等内容。

第2章介绍数据库应用系统的设计过程，重点介绍ER模型以及ER模型到关系模型的转换规则。

第3章介绍关系模式的设计理论，包括函数依赖、模式分解特性、范式和模式设计方法等问题。

第4章介绍关系模型的运算理论，包括关系代数、关系演算和查询优化等内容。

第5章介绍关系数据库标准语言SQL的全貌，包括数据定义、查询、更新，以及嵌入式SQL、存储过程等内容。

第6章介绍数据库系统的管理技术，包括事务、恢复、并发控制、完整性和安全性等内容。

第7章介绍SQL Server 2000的开发环境，主要介绍企业管理器的使用。

第8章介绍PowerBuilder 9.0的集成开发环境，通过“学生选课系统”的开发来介绍数据库应用系统的开发过程。

第9章介绍面向对象的概念建模和开放数据库互连技术等内容。

本课程具有较强的理论性、实用性和可操作性。理论性体现在第3章的模式设计理论和第4章的关系运算理论。实用性体现在第2章的ER模型的设计和第5章的SQL语言的使用上。可操作性体现在本课程已形成较为完整的系列习题、思考题和设计题,使学生在数据库领域有回旋的余地。在每一道例题或定理的结尾处,加了符号“□”,以示与正文的区别。

本书第1~5章由丁宝康执笔,第6、8章由杜赞执笔,第7、9章由乔健执笔,参与编写的还有唐慧君、杨卫稼、陈长洪、薛剑虹、许建军等老师。全书由丁宝康任主编。

上海交通大学黄上腾教授、华东理工大学杨明福教授和上海第二工业大学何守才教授审阅了全稿,并提出了许多宝贵的意见,在此向他们致以衷心的感谢。

复旦大学施伯乐教授对全书的结构和取材进行了指导,上海交通大学陈敏逊教授、华中科技大学卢炎生教授和杭州电子工业大学章坚武教授等为本书的组稿和出版进行了大量的工作,在此一并向他们表示感谢。

限于水平,书中欠妥之处,敬请广大读者和专家批评指正。对本书的意见请通过邮箱 dn@citiz.net 反馈给我们,谢谢。

丁宝康

2006年12月

目 录

数据库系统原理

| | |
|------------------------------|--------|
| 第 1 章 数据库系统基本概念 | (1) |
| 1.1 数据管理技术的发展 | (1) |
| 1.1.1 人工管理阶段 | (1) |
| 1.1.2 文件系统阶段 | (2) |
| 1.1.3 数据库阶段 | (3) |
| 1.1.4 高级数据库阶段 | (6) |
| 1.2 数据描述 | (6) |
| 1.2.1 概念设计中的数据描述 | (7) |
| 1.2.2 逻辑设计中的数据描述 | (7) |
| 1.2.3 物理设计中的数据描述 | (8) |
| 1.2.4 数据联系的描述 | (9) |
| 1.3 数据抽象的级别 | (11) |
| 1.3.1 数据抽象的过程 | (11) |
| 1.3.2 概念模型 | (12) |
| 1.3.3 逻辑模型 | (13) |
| 1.3.4 外部模型 | (17) |
| 1.3.5 内部模型 | (18) |
| 1.3.6 三层模式和两级映像 | (18) |
| 1.3.7 高度的数据独立性 | (19) |
| 1.4 数据库管理系统 (DBMS) | (20) |
| 1.4.1 DBMS 的工作模式 | (20) |
| 1.4.2 DBMS 的主要功能 | (21) |
| 1.5 数据库系统 (DBS) | (22) |
| 1.5.1 DBS 的组成 | (22) |
| 1.5.2 DBS 的全局结构 | (24) |
| 1.5.3 DBS 的效益 | (26) |

| | |
|--------------------------------|-------------|
| 小结 | (26) |
| 习题1 | (27) |
| 第2章 数据库设计和 ER 模型 | (29) |
| 2.1 数据库系统生存期 | (29) |
| 2.1.1 规划阶段 | (30) |
| 2.1.2 需求分析阶段 | (30) |
| 2.1.3 概念设计阶段 | (31) |
| 2.1.4 逻辑设计阶段 | (32) |
| 2.1.5 物理设计阶段 | (33) |
| 2.1.6 数据库的实现 | (33) |
| 2.1.7 数据库的运行与维护 | (35) |
| 2.2 ER 模型的基本概念 | (36) |
| 2.2.1 ER 模型的基本元素 | (36) |
| 2.2.2 属性的分类 | (38) |
| 2.2.3 联系的设计 | (40) |
| 2.2.4 ER 模型的操作 | (45) |
| 2.2.5 采用 ER 模型的数据库概念设计步骤 | (46) |
| 2.3 关系模型的基本概念 | (52) |
| 2.3.1 关系模型的基本术语 | (52) |
| 2.3.2 关系的定义和性质 | (53) |
| 2.3.3 三类完整性规则 | (53) |
| 2.4 ER 模型到关系模型的转换 | (55) |
| 2.4.1 ER 图转换成关系模式集的算法 | (55) |
| 2.4.2 采用 ER 模型的逻辑设计步骤 | (57) |
| 2.5 ER 模型实例分析 | (58) |
| 2.5.1 库存管理信息系统的 ER 模型及转换 | (58) |
| 2.5.2 人事管理信息系统的 ER 模型 | (59) |
| 2.5.3 住院管理信息系统的 ER 模型 | (60) |
| 2.5.4 公司车队信息系统的 ER 模型 | (61) |
| 2.6 增强的 ER 模型 | (62) |
| 2.6.1 弱实体与强实体 | (62) |
| 2.6.2 子类实体与超类实体 | (63) |
| 小结 | (65) |
| 习题2 | (65) |
| 第3章 关系模式设计理论 | (68) |
| 3.1 关系模式的设计准则 | (68) |
| 3.1.1 关系模式的冗余和异常问题 | (68) |
| 3.1.2 关系模式的非形式化设计准则 | (70) |

| | | |
|-------|--------------------|--------|
| 3.2 | 函数依赖 | (70) |
| 3.2.1 | 函数依赖的定义 | (71) |
| 3.2.2 | FD 的逻辑蕴涵 | (72) |
| 3.2.3 | FD 的推理规则 | (72) |
| 3.2.4 | FD 和关键码的联系 | (74) |
| 3.2.5 | 属性集的闭包 | (74) |
| 3.2.6 | FD 集的最小依赖集 | (75) |
| 3.3 | 关系模式的分解特性 | (76) |
| 3.3.1 | 关系模式的分解 | (76) |
| 3.3.2 | 无损分解 | (76) |
| 3.3.3 | 模式分解的优缺点 | (79) |
| 3.3.4 | 无损分解的测试方法 | (79) |
| 3.3.5 | 保持 FD 的分解 | (80) |
| 3.3.6 | 模式分解与模式等价问题 | (81) |
| 3.4 | 范式 | (82) |
| 3.4.1 | 第一范式 (1NF) | (82) |
| 3.4.2 | 第二范式 (2NF) | (82) |
| 3.4.3 | 第三范式 (3NF) | (83) |
| 3.4.4 | BCNF | (85) |
| 3.4.5 | 分解成 BCNF 模式集的分解算法 | (86) |
| 3.4.6 | 分解成 3NF 模式集的合成算法 | (86) |
| 3.4.7 | 模式设计方法小结 | (87) |
| 3.5 | 多值依赖和第四范式 | (87) |
| 3.5.1 | 多值依赖 | (87) |
| 3.5.2 | 关于 FD 和 MVD 的推理规则集 | (89) |
| 3.5.3 | 第四范式 (4NF) | (90) |
| | 小结 | (90) |
| | 习题 3 | (91) |

第 4 章 关系运算 (95)

| | | |
|-------|---------------|---------|
| 4.1 | 关系代数 | (95) |
| 4.1.1 | 关系代数的五个基本操作 | (95) |
| 4.1.2 | 关系代数的四个组合操作 | (97) |
| 4.1.3 | 关系代数运算的应用实例 | (100) |
| 4.1.4 | 关系代数的两个扩充操作 | (101) |
| 4.2 | 关系演算 | (102) |
| 4.2.1 | 元组关系演算 | (102) |
| 4.2.2 | 域关系演算 | (106) |
| 4.2.3 | 关系运算的安全约束和等价性 | (107) |
| 4.3 | 关系代数表达式的优化 | (108) |

| | | |
|------------------|-----------------------|--------------|
| 4.3.1 | 关系代数表达式的优化问题 | (108) |
| 4.3.2 | 关系代数表达式的启发式优化算法 | (109) |
| 小结 | | (112) |
| 习题4 | | (113) |
| 第5章 SQL语言 | | (115) |
| 5.1 | SQL简介 | (115) |
| 5.1.1 | SQL发展史 | (115) |
| 5.1.2 | SQL数据库的体系结构 | (116) |
| 5.1.3 | SQL的组成 | (116) |
| 5.1.4 | SQL的特点 | (117) |
| 5.2 | SQL的数据定义 | (117) |
| 5.2.1 | SQL模式的创建和撤销 | (117) |
| 5.2.2 | SQL的基本数据类型 | (118) |
| 5.2.3 | 基本表的创建和撤销 | (119) |
| 5.2.4 | 索引的创建和撤销 | (121) |
| 5.3 | SQL的数据查询 | (122) |
| 5.3.1 | SELECT查询语句的基本结构 | (122) |
| 5.3.2 | SELECT语句完整的结构 | (126) |
| 5.3.3 | 数据查询中的限制和规定 | (128) |
| 5.3.4 | 条件表达式中的比较操作 | (129) |
| 5.3.5 | 嵌套查询的改进写法 | (133) |
| 5.3.6 | 基本表的连接操作 | (134) |
| 5.4 | 数据更新 | (135) |
| 5.4.1 | 数据插入 | (135) |
| 5.4.2 | 数据删除 | (136) |
| 5.4.3 | 数据修改 | (137) |
| 5.5 | 视图 | (138) |
| 5.5.1 | 视图的创建和撤销 | (138) |
| 5.5.2 | 对视图的操作 | (139) |
| 5.6 | 嵌入式SQL | (139) |
| 5.6.1 | 嵌入式SQL的实现方式 | (140) |
| 5.6.2 | 嵌入式SQL的使用规定 | (141) |
| 5.6.3 | 嵌入式SQL的使用技术 | (142) |
| 5.6.4 | 动态SQL语句 | (145) |
| 5.7 | 存储过程与SQL/PSM | (146) |
| 5.7.1 | 数据库存储过程与函数 | (146) |
| 5.7.2 | SQL/PSM | (148) |
| 小结 | | (149) |
| 习题5 | | (149) |

| | |
|--|-------|
| 第 6 章 数据库管理 | (152) |
| 6.1 事务 | (152) |
| 6.1.1 事务的定义 | (152) |
| 6.1.2 事务的 ACID 性质 | (153) |
| 6.2 数据库的恢复 | (154) |
| 6.2.1 典型的恢复策略 | (154) |
| 6.2.2 故障类型和恢复方法 | (155) |
| 6.2.3 检查点技术 | (156) |
| 6.2.4 SQL 对事务的支持 | (157) |
| 6.3 数据库的并发控制 | (157) |
| 6.3.1 并发操作带来的三个问题 | (157) |
| 6.3.2 封锁技术 | (160) |
| 6.3.3 并发操作的调度 | (163) |
| 6.3.4 SQL 对并发处理的支持 | (163) |
| 6.4 数据库的完整性 | (164) |
| 6.4.1 完整性子系统 | (164) |
| 6.4.2 SQL 中的完整性约束 | (165) |
| 6.4.3 SQL3 的触发器 | (169) |
| 6.5 数据库的安全性 | (172) |
| 6.5.1 安全性问题 | (172) |
| 6.5.2 SQL 中的安全性机制 | (173) |
| 6.5.3 常用的安全性措施 | (175) |
| 小结 | (177) |
| 习题 6 | (178) |
| 第 7 章 SQL Server 2000 简介及应用 | (180) |
| 7.1 SQL Server 2000 概述 | (180) |
| 7.1.1 SQL Server 2000 的结构和版本 | (180) |
| 7.1.2 SQL Server 2000 的环境介绍 | (182) |
| 7.1.3 工具介绍 | (183) |
| 7.2 企业管理器 | (185) |
| 7.2.1 数据库的操作 | (188) |
| 7.2.2 表的操作 | (189) |
| 7.2.3 索引 | (191) |
| 7.2.4 视图 | (192) |
| 7.2.5 存储过程 | (193) |
| 7.2.6 触发器 | (196) |
| 7.2.7 用户自定义函数 | (199) |
| 7.2.8 备份恢复与导入导出 | (202) |

| | |
|---|--------------|
| 7.2.9 安全管理 | (206) |
| 7.3 T-SQL 语言 | (209) |
| 7.3.1 T-SQL 批处理 | (209) |
| 7.3.2 变量 | (210) |
| 7.3.3 流程控制 | (211) |
| 7.3.4 临时表和表变量 | (213) |
| 小结 | (214) |
| 习题 7 | (214) |
| 第 8 章 PowerBuilder 9.0 简介及应用 | (216) |
| 8.1 PowerBuilder 9.0 集成开发环境 | (216) |
| 8.1.1 PB 的特点 | (216) |
| 8.1.2 开发空间 | (217) |
| 8.1.3 PB 9.0 的启动 | (217) |
| 8.1.4 系统树、剪贴板和输出窗口 | (217) |
| 8.1.5 PB 9.0 的工具栏 | (219) |
| 8.1.6 PB 9.0 的主要画板 | (220) |
| 8.2 “学生选课系统”的开发过程 | (221) |
| 8.2.1 “学生选课系统”概貌 | (221) |
| 8.2.2 建立数据库 newdb, 并自动连接数据库 | (226) |
| 8.2.3 在数据库中建立三张表及一个视图 | (228) |
| 8.2.4 新建工作空间、目标和应用对象 | (232) |
| 8.2.5 在应用中建五个数据窗口 | (235) |
| 8.2.6 在应用中建六个窗口 | (237) |
| 8.2.7 编写脚本 | (240) |
| 8.2.8 运行应用程序 | (249) |
| 8.3 PB 9.0 与数据库的连接 | (249) |
| 8.3.1 建立 ODBC 数据源 | (249) |
| 8.3.2 建立数据库描述文件 | (253) |
| 8.3.3 连接数据源 | (255) |
| 小结 | (256) |
| 习题 8 | (256) |
| 第 9 章 数据库技术的发展 | (257) |
| 9.1 面向对象的概念建模 | (257) |
| 9.1.1 面向对象的数据类型系统 | (257) |
| 9.1.2 对象联系图 | (259) |
| 9.1.3 UML 类图 | (260) |
| 9.2 开放数据库互连 (ODBC) | (267) |
| 9.2.1 ODBC 概念 | (267) |

| | |
|----------------------------|-------|
| 9.2.2 ODBC 的体系结构 | (268) |
| 9.2.3 SQL/CLI | (270) |
| 9.2.4 典型的数据库应用系统开发工具 | (273) |
| 小结 | (274) |
| 习题9 | (274) |
| 参考文献 | (276) |

数据库系统原理自学考试大纲

| | |
|----------------------------------|-------|
| 出版前言 | (279) |
| 一、课程性质与设置目的 | (281) |
| 二、课程内容与考核目标 | (283) |
| 第1章 数据库系统基本概念 | (283) |
| 第2章 数据库设计和 ER 模型 | (284) |
| 第3章 关系模式设计理论 | (286) |
| 第4章 关系运算 | (287) |
| 第5章 SQL 语言 | (288) |
| 第6章 数据库管理 | (290) |
| 第7章 SQL Server 2000 简介及应用 | (291) |
| 第8章 PowerBuilder 9.0 简介及应用 | (292) |
| 第9章 数据库技术的发展 | (293) |
| 实践环节 | (293) |
| 三、关于大纲的说明与考核实施的要求 | (295) |
| 附录 题型举例 | (298) |
| 后记 | (300) |

第 1 章 数据库系统基本概念

今天，“数据库”（Database）这个名词已是家喻户晓，成为人们日常生活中不可缺少的一部分。例如，我们可能会去银行取款，可能需要预订飞机票，可能网上购物，等等。数据库的应用领域已遍及现代社会的每一个角落。

本章介绍数据库系统的基本概念，以使读者对数据库概貌有个基本的了解。1.1 节介绍数据管理技术的发展过程，1.2 节介绍数据库设计中各阶段的数据描述，1.3 节介绍数据库设计中各阶段的数据抽象（即数据模型），1.4 节介绍数据库管理系统的工作模式和主要功能，1.5 节介绍数据库系统的组成和全局结构。

1.1 数据管理技术的发展

使用计算机以后，数据处理的速度和规模是手工方式或机械方式无可比拟的，随着数据处理量的增长，产生了数据管理技术。数据管理技术的发展，与计算机硬件（主要是外部存储器）、系统软件及计算机应用的范围有着密切的联系。数据管理技术的发展经历了人工管理、文件系统、数据库和高级数据库阶段。下面分别介绍。

1.1.1 人工管理阶段

在这一阶段（20 世纪 50 年代中期以前），计算机主要用于科学计算，其他工作还没有展开。外部存储器只有磁带、卡片和纸带等，还没有磁盘等直接存取存储设备。软件只有汇编语言，尚无数据管理方面的软件。数据处理的方式基本上是批处理。这个时期的数据管理有下列特点：

（1）数据不保存在计算机内。计算机主要用于计算，一般不需要长期保存数据。在进行某一课题计算时，将原始数据随程序一起输入内存，运算处理后将结果数据输出。随着计算任务的完成，用户作业退出计算机系统，数据空间随着程序空间一起被释放。

（2）没有专用的软件对数据进行管理。每个应用程序都要包括存储结构、存取方法、输入输出方式等内容。程序中的存取子程序随着存储结构的改变而改变，因而数据与程序不具有独立性。存储结构改变时，应用程序必须改变。此时，由于程序直接面向存储结构，因此数据的逻辑结构与物理结构没有区别。

（3）只有程序（Program）的概念，没有文件（File）的概念。数据的组织方式必须由程序员自行设计与安排。

（4）数据面向程序。即一组数据对应一个程序。

1.1.2 文件系统阶段

在这一阶段（20世纪50年代后期至60年代后期），计算机不仅用于科学计算，还用于信息管理。随着数据量的增加，数据的存储、检索和维护问题成为紧迫的需要，数据结构和数据管理技术迅速发展起来。此时，外部存储器已有磁盘、磁鼓等直接存取存储设备。1956年生产的第一台磁盘，其容量仅为5MB，60年代时已达到数十个MB。软件领域出现了高级语言和操作系统。操作系统中的文件系统是专门管理外存的数据管理软件。数据处理的方式有批处理，也有联机实时处理。

这一阶段的数据管理有以下特点：

(1) 数据以“文件”形式可长期保存在外部存储器的磁盘上。由于计算机的应用转向信息管理，因此对文件要进行大量的查询、修改和插入等操作。

(2) 数据的逻辑结构与物理结构有了区别，但比较简单。程序与数据之间具有“设备独立性”，即程序只需用文件名就可与数据打交道，不必关心数据的物理位置。由操作系统的文件系统提供存取方法（读/写）。

(3) 文件组织已多样化。有索引文件、链接文件和直接存取文件等。但文件之间相互独立、缺乏联系。数据之间的联系要通过程序去构造。

(4) 数据不再属于某个特定的程序，可以重复使用，即数据面向应用。但是文件结构的设计仍然是基于特定的用途，程序基于特定的物理结构和存取方法，因此程序与数据结构之间的依赖关系并未根本改变。

在文件系统阶段，由于具有设备独立性，因此当改变存储设备时，不必改变应用程序。但这只是初级的数据管理，还未能彻底体现用户观点下的数据逻辑结构独立于数据在外存的物理结构要求。在数据的物理结构修改时，仍然需要修改用户的应用程序，即应用程序具有“程序—数据依赖”性。有关物理表示的知识和访问技术将直接体现在应用程序的代码中。

(5) 对数据的操作以记录为单位。这是由于文件中只存储数据，不存储文件记录的结构描述信息。文件的建立、存取、查询、插入、删除、修改等所有操作，都要用程序来实现。

文件系统阶段是数据管理技术发展中的一个重要阶段。在这一阶段中。得到充分发展的数据结构和算法丰富了计算机科学，为数据管理技术的进一步发展打下了基础，现在仍是计算机软件科学的重要基础。

但是，随着数据管理规模的扩大，数据量急剧增加，文件系统显露出三个缺陷：

(1) 数据冗余（Redundancy）。由于文件之间缺乏联系，造成每个应用程序都有对应的文件，有可能同样的数据在多个文件中重复存储；

(2) 数据不一致（Inconsistency）。这往往是由数据冗余造成的，在进行更新操作时，稍不谨慎，就可能使同样的数据在不同的文件中显示的结果不一样；

(3) 数据联系弱（Poor Data Relationship）。这是由于文件之间相互独立，缺乏联系造成的。

由于这些原因，促使人们研究新的数据管理技术，在20世纪60年代末产生了数据库

技术。

【例 1.1】 某单位添置了一台计算机，各部门纷纷在计算机中建立了文件。譬如建立了职工档案文件、职工工资文件和职工保健文件。每一职工的电话号码在三个文件中重复出现。这就是“数据冗余”；如果某职工的电话号码要修改，就要修改三个文件中的数据，否则会引起同一数据在三个文件中不一样；产生上述问题的原因是三个文件中数据没有联系。如图 1.1 (a) 所示。

如果在职工档案文件中存放电话号码值，而在另外文件中不存放电话号码值，而存放档案文件中电话号码值的位置（即“指针”），如图 1.1 (b) 所示。这样就能消除文件系统中的三个缺陷。此时电话号码不重复存储，只存储在档案文件中，而其他两个文件中的值为指针；修改时只需修改档案文件中的电话号码，而其他两个文件的值为指针值不必修改，这样就不会产生不一致现象；三个文件中的数据通过指针，加强了联系。这种存储结构就进入了数据库方式。 □

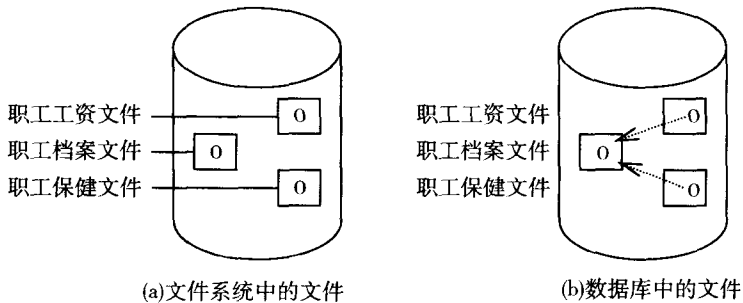


图 1.1 文件结构的对照

1.1.3 数据库阶段

随着数据管理规模一再扩大，数据量急剧增长。为了提高效率，人们开始时只是对文件系统加以扩充，研制成倒排文件系统，但这并不能解决问题。恰好在 20 世纪 60 年代末，磁盘技术取得重要进展，具有数百兆字节容量和快速存取的磁盘陆续进入市场，成本也不高，这就为数据库技术的产生提供了良好的物质条件。

20 世纪 60 年代中期出现的系统（Database 或 Databank）还不能真正地称为数据库系统。数据管理技术进入数据库阶段的标志是 20 世纪 60 年代末的三件大事：

- (1) 1968 年美国 IBM 公司推出层次模型的 IMS（Information Management System）系统；
- (2) 1969 年美国 CODASYL（Conference On Data System Language）组织发布了 DBTG（Data Base Task Group）报告。总结了当时各式各样的数据库，提出网状模型，尔后于 1971 年 4 月正式通过；
- (3) 1970 年美国 IBM 公司的 E. F. Codd 连续发表论文，提出关系模型，奠定了关系数据库的理论基础。

20 世纪 70 年代以来，数据库技术得到迅速发展，开发出了许多产品，并投入运行。数据库系统克服了文件系统的缺陷，提供了对数据更高级、更有效的管理。概括起来，数据库