



“十二五”普通高等教育本科国家级规划教材

“高等学校本科计算机类专业应用型人才培养研究”项目规划教材

数据库系统及应用 (第4版)

Database System and Application
(Fourth Edition)

崔巍 编著

高等教育出版社



“十二五”普通高等教育本科国家级规划教材

“高等学校本科计算机类专业应用型人才培养研究”项目规划教材

数据库系统及应用 (第4版)

Database System and Application
(Fourth Edition)

崔巍 编著

高等教育出版社·北京

内容简介

本书第2版和第3版被评为北京市高等教育精品教材,第3版同时被列为“十二五”普通高等教育本科国家级规划教材。

第4版保持前版教材的特色,在结构、内容上都做了调整,使其更完善。全书共12章,主要包括数据库概论、概念数据模型、关系数据库基础、关系数据理论、逻辑数据模型和物理数据模型、数据定义、操作与完整性约束、SQL查询、数据库编程基础、数据库安全、事务管理与并发控制、数据库存储管理和数据恢复,以及数据库应用和研究的新领域(面向对象数据库、数据仓库与数据分析、分布式数据库与云计算、大数据与NoSQL等)。

本书可作为高等院校计算机、信息管理与信息系统等相关专业的数据库课程教材,也可供从事计算机软件及数据库应用、管理和开发的科技人员、工程技术人员以及其他有关人员阅读参考。

图书在版编目(CIP)数据

数据库系统及应用 / 崔巍编著. —4版. —北京:
高等教育出版社, 2017.12
ISBN 978-7-04-048771-8

I. ①数… II. ①崔… 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(2017)第258743号

Shujuku Xitong ji Yingyong

策划编辑 倪文慧
责任校对 刘丽娴

责任编辑 倪文慧
责任印制 田甜

版式设计 徐艳妮

插图绘制 杜晓丹

出版发行 高等教育出版社
社址 北京市西城区德外大街4号
邮政编码 100120
印刷 北京宏伟双华印刷有限公司
开本 787mm×1092mm 1/16
印张 22.75
字数 470千字
购书热线 010-58581118
咨询电话 400-810-0598

网 址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.hepmall.com.cn>
<http://www.hepmall.com>
<http://www.hepmall.cn>
版 次 1999年6月第1版
2017年12月第4版
印 次 2017年12月第1次印刷
定 价 43.00元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换

版权所有 侵权必究

物料号 48771-00

数字课程资源使用说明

与本书配套的数字课程资源包括微视频、电子教案、有关程序模板等，具体使用方法如下：

1. 电脑访问 <http://abook.hep.com.cn/1865070>，或手机扫描二维码、下载并安装 Abook 应用。

2. 注册并登录，进入“我的课程”。

3. 输入教材封底防伪标签上的数字课程账号（20 位密码，刮开涂层可见），或通过 Abook 应用扫描封底数字课程账号二维码，完成课程绑定。

4. 单击“进入课程”按钮，开始本数字课程的学习。

课程绑定后一年为数字课程使用有效期。受硬件限制，部分内容无法在手机端显示，请按提示通过计算机访问学习。

如有使用问题，请发邮件至 abook@hep.com.cn。



扫描二维码

下载 Abook 应用

前 言

本书第1版于1999年出版,第2版于2003年出版,第3版于2012年出版。第2版和第3版被评为北京市高等教育精品教材,第3版还被列为“十二五”普通高等教育本科国家级规划教材。

本次修订仍以在夯实基础的前提下加强实践的思想为指导,对传统教学内容的安排进行了调整和优化,希望在结构上更科学,内容上更丰富,教学上更好用,没有最好,努力做到更好。

本书依然特别强调数据库课程的实践性,建议以SQL Server 2008为教学和实践环境,书中涉及的实践环节和示例均可以在SQL Server 2008下运行。

本书按顺序可以分为以下4部分。

第1部分为数据库基础和数据库设计(第1~5章)。第1章介绍数据库的基本概念以及数据管理技术的发展过程,概述数据模型、三层结构、数据库系统等内容。第2章介绍概念数据模型,从理解现实世界的数据库管理开始去理解数据模型设计和数据库系统,讨论概念数据模型设计中的普遍问题,并利用一个案例完成从局部E-R模型到全局E-R模型的设计。第3章介绍关系数据库基础,包括关系数据模型、关系完整性约束、关系代数和关系数据库系统等,简单介绍SQL Server和关系数据库标准语言SQL。第4章介绍关系数据理论,包括函数依赖、公理系统、规范化和模式分解等内容。第5章介绍逻辑数据模型设计和物理数据模型设计,包括数据库设计的一般过程、E-R模型到关系模型的转换、关系规范化理论的应用,以及反规范化的概念等。在第2章和第5章的数据模型设计中使用数据库建模工具PowerDesigner。

第2部分围绕SQL深入学习数据定义、数据操作、数据查询和数据库编程等内容(第6~8章)。第6章详细介绍SQL的表定义和完整性约束定义功能,通过操作功能体验数据完整性约束的作用。第7章通过大量的实例分别从简单查询、连接查询、汇总查询和嵌套查询等方面详细介绍SQL的查询功能,还介绍了需要查询支持的数据操作功能以及视图的定义、作用和应用等。第8章介绍游标、存储过程、触发器等基本数据库编程技术,还介绍了动态SQL的基本内容。

第3部分为数据库系统和管理(第9~11章)。第9章介绍数据库的安全问题,包括用户管理和权限管理等内容,还介绍了数据加密等其他与数据库安全相关的问题。第10章介绍数据库的事务管理与并发控制。第11章介绍数据库的存储管理和存储优化,以及

数据库恢复等内容。通过这一部分内容的学习与实践,希望学生不仅能够掌握数据库的基本概念和技术,同时还具备管理数据库的基本实践技能。

第4部分为数据库新技术(第12章)。按顺序介绍了面向对象数据库、数据仓库与数据分析、分布式数据库与云计算、大数据与NoSQL,以及主动数据库、知识库、模糊数据库、并行数据库和多媒体数据库等基本内容。

本书可按48~64学时使用,学时大致分配如下:第1部分数据库基础和数据库设计约12~16学时,第2部分SQL及其编程约18~24学时,第3部分数据库系统和管理约12~14学时,第4部分数据库新技术约6~10学时。

为便于读者自学,本书为重难点配有视频讲解。读者可扫描二维码观看。此外,本书还为使用教材的教师提供了电子教案、习题解答和有关的程序模板,如需要可与作者联系(E-mail: cuiw@bistu.edu.cn),或访问本书配套的数字课程网站下载。本书相应课程的北京市精品课程网站为<http://dbcourse.bistu.edu.cn>。

在教材修订的过程中,笔者广泛听取了兄弟院校一些教师的意见以及一些学生的建议,力求克服前3版教材中的不足,在科学性、完备性、实用性和教材的可用性方面都有一定程度的提高。但由于水平有限,书中仍然会有许多不足之处,还望同行和专家批评指正。

北京信息科技大学信息管理学院数据库课程教学团队的王晓波、车蕾、卢益清、王晓敏、王磊等老师在教材使用和修订的过程中都提出过很好的建议。在此向他们一并致以衷心的感谢。

崔巍

2017年10月

目 录

第 1 章 数据库概论.....1

1.1 数据管理及其发展过程.....1

1.1.1 数据管理.....1

1.1.2 人工管理阶段.....1

1.1.3 文件系统阶段.....2

1.1.4 数据库系统阶段及其发展过程.....4

1.1.5 数据管理和数据库技术的 持续发展.....5

1.2 数据库系统的特点.....6

1.3 数据模型初步.....8

1.3.1 概念数据模型.....8

1.3.2 传统的三大数据模型.....9

1.4 数据独立性与三层结构.....13

1.4.1 存储数据独立性.....13

1.4.2 概念数据独立性.....15

1.5 三层模式结构和数据库管理系统.....16

1.5.1 数据库管理系统的基本功能.....16

1.5.2 数据库的三层模式结构.....17

1.5.3 模式说明实例.....18

1.5.4 数据库管理系统的基本框架.....22

1.6 数据库系统.....23

1.6.1 数据库系统的组成.....23

1.6.2 数据库管理和数据库管理员.....23

1.6.3 数据库应用系统的体系结构.....24

本章小结.....25

习题与思考题.....26

第 2 章 概念数据模型.....27

2.1 概念数据模型综述.....27

2.2 实体-联系方法.....28

2.2.1 实体.....28

2.2.2 实体之间的联系.....29

2.3 数据库建模工具.....31

2.3.1 SAP PowerDesigner 简介.....31

2.3.2 使用 PowerDesigner 建立概念 数据模型的基本方法.....32

2.4 深入讨论联系的几个问题.....36

2.4.1 为什么需要讨论联系.....36

2.4.2 理解和转换多对多联系.....40

2.4.3 PowerDesigner 中的多对多联系 转换为一对多联系.....41

2.4.4 连接陷阱.....43

2.5 概念数据模型设计实例.....44

2.5.1 问题描述.....44

2.5.2 库存业务局部概念模型.....44

2.5.3 订购业务局部概念模型.....45

2.5.4 将局部 E-R 模型合并为 全局 E-R 模型.....46

本章小结.....49

习题与思考题.....50

实验 1 概念数据模型设计.....50

第 3 章 关系数据库基础.....52

3.1 关系数据库系统概述.....52

3.1.1 关系数据库的发展.....52

3.1.2 关系数据库管理系统简介.....53

3.1.3 关系数据库标准语言 SQL 简介.....54

3.1.4 关系数据库的三层模式结构.....56

3.1.5 SQL Server 的数据库存储结构	56	4.3.1 第一范式	89
3.1.6 SQL Server 的用户数据库	58	4.3.2 第二范式	90
3.2 关系数据模型	58	4.3.3 第三范式	91
3.2.1 关系数据模型的 3 个要素	58	4.3.4 BC 范式	92
3.2.2 关系的形式定义	60	4.3.5 多值依赖与第四范式	93
3.2.3 关系的基本性质	61	4.3.6 规范化小结	95
3.2.4 关系模型的数据结构和基本术语	62	4.4 模式分解	97
3.3 关系模型的完整性约束	64	4.4.1 模式分解的准则	97
3.3.1 实体完整性约束	64	4.4.2 3NF 无损连接和保持函数依赖算法	99
3.3.2 参照完整性约束	65	4.4.3 使分解后的关系模式数最少	100
3.3.3 用户定义完整性约束	66	本章小结	101
3.3.4 完整性约束的作用	67	习题与思考题	101
3.4 关系代数	68	第 5 章 逻辑数据模型和物理数据模型	103
3.4.1 基本概念和几个符号	68	5.1 数据库设计的概念和方法	103
3.4.2 传统的集合运算	70	5.1.1 数据库设计的概念	103
3.4.3 专门的关系运算	71	5.1.2 数据库设计的一般步骤	104
3.4.4 基本运算及变换	76	5.2 逻辑数据模型设计	106
本章小结	77	5.2.1 逻辑数据模型设计的主要内容	106
习题与思考题	77	5.2.2 把 E-R 模型转换为关系数据模型	107
第 4 章 关系数据理论	79	5.2.3 规范化理论的应用	108
4.1 基本概念	79	5.2.4 反规范化	109
4.1.1 函数依赖	79	5.2.5 设计视图	109
4.1.2 术语和符号	80	5.3 物理数据模型设计	109
4.1.3 为什么要讨论函数依赖	80	5.3.1 由逻辑数据模型生成物理数据模型	110
4.1.4 模式分解	81	5.3.2 物理数据库设计	111
4.2 函数依赖的推理规则	82	5.3.3 建立数据库	112
4.2.1 函数依赖的推理规则及正确性	82	本章小结	115
4.2.2 Armstrong 公理的推论及正确性	83	习题与思考题	115
4.2.3 逻辑蕴涵和闭包	83	实验 2 数据库设计	115
4.2.4 公理的完备性	84		
4.2.5 属性集闭包的计算	86		
4.2.6 函数依赖集的等价和最小化	87		
4.3 规范化	89		

第 6 章 数据定义、操作与

完整性约束 117

6.1 SQL 的架构和定义 117

6.1.1 什么是架构 117

6.1.2 定义架构 117

6.2 SQL 的表定义和完整性定义功能 119

6.2.1 定义表及其完整性约束 119

6.2.2 修改表结构 125

6.3 SQL 数据操作与完整性

约束的作用 127

6.3.1 插入操作 127

6.3.2 删除操作 133

6.3.3 更新操作 134

本章小结 135

习题与思考题 136

实验 3 建立表和定义完整性约束 136

实验 4 数据操作及体验完整性约束 138

第 7 章 SQL 查询 140

7.1 SQL 的数据查询命令 140

7.2 简单查询 141

7.2.1 简单无条件查询 141

7.2.2 简单条件查询 142

7.2.3 使用[NOT] BETWEEN... AND...的查询 143

7.2.4 字符串匹配查询 144

7.2.5 空值查询 145

7.2.6 使用 IN 表达式的查询 146

7.2.7 ALL 和 DISTINCT 短语的 作用 147

7.2.8 存储查询结果 147

7.2.9 查询结果的排序 148

7.2.10 TOP 短语的作用 149

7.2.11 集合运算 151

7.3 连接查询 152

7.3.1 连接查询的语法格式 152

7.3.2 一般连接 153

7.3.3 多个表的连接 153

7.3.4 别名和自连接查询 154

7.3.5 外连接查询 156

7.3.6 广义笛卡儿积 159

7.4 分组及汇总查询 160

7.4.1 聚合函数与汇总查询 160

7.4.2 一般汇总查询 161

7.4.3 带明细的汇总查询 162

7.4.4 使用 GROUP BY 的分组 汇总查询 164

7.4.5 使用 COMPUTE BY 的分组 汇总查询 166

7.4.6 使用 COMPUTE BY 和 COMPUTE 的汇总查询 169

7.5 嵌套查询 169

7.5.1 普通嵌套查询 169

7.5.2 使用量词的嵌套查询 172

7.5.3 内、外层互相关嵌套查询 173

7.5.4 使用 EXISTS 的嵌套查询 174

7.6 需要查询支持的数据操作 175

7.6.1 插入操作 175

7.6.2 更新操作 176

7.6.3 删除操作 176

7.7 视图及其操作 177

7.7.1 视图的建立和使用 177

7.7.2 视图的修改和删除 180

7.7.3 视图的作用 181

7.7.4 在设计数据库时设计视图 181

本章小结 182

习题与思考题 183

实验 5 数据查询 185

实验 6 视图的应用 187

第 8 章 数据库编程基础	189	8.5.2 动态定义功能.....	217
8.1 Transact-SQL 介绍.....	189	8.5.3 动态操作功能.....	217
8.1.1 数据类型与变量说明.....	189	8.5.4 动态查询功能.....	217
8.1.2 运算符与表达式.....	189	本章小结.....	219
8.1.3 函数.....	190	习题与思考题.....	219
8.1.4 全局变量.....	192	实验 7 游标的应用.....	220
8.1.5 程序语句.....	192	实验 8 存储过程及应用.....	221
8.2 游标与 SQL 的宿主使用.....	194	实验 9 触发器及应用.....	222
8.2.1 SQL 宿主使用面临的问题.....	194	第 9 章 数据库安全	224
8.2.2 嵌入识别与预编译.....	195	9.1 安全性概述.....	224
8.2.3 数据通信区与主变量.....	196	9.1.1 安全性措施的层次.....	224
8.2.4 游标.....	197	9.1.2 数据库管理系统的安全功能.....	225
8.2.5 游标应用举例.....	198	9.1.3 自主存取控制与强制存取控制.....	226
8.2.6 利用游标进行删除和 更新操作.....	201	9.1.4 数据库管理系统的 身份验证模式.....	227
8.3 存储过程.....	202	9.2 用户管理.....	229
8.3.1 基本概念.....	203	9.2.1 登录用户和数据库用户.....	229
8.3.2 创建和执行存储过程.....	204	9.2.2 登录用户管理.....	229
8.3.3 存储过程的修改和删除.....	205	9.2.3 数据库用户的管理.....	232
8.3.4 存储过程应用举例.....	205	9.2.4 数据库用户的分类.....	234
8.3.5 获得有关存储过程的信息.....	209	9.3 角色及其管理.....	234
8.3.6 在数据库设计阶段设计 存储过程.....	209	9.3.1 基本概念.....	235
8.3.7 用户定义函数.....	210	9.3.2 public 角色.....	235
8.4 触发器及其用途.....	210	9.3.3 角色管理.....	236
8.4.1 基本概念.....	210	9.3.4 系统预定义角色.....	237
8.4.2 建立触发器.....	211	9.4 权限管理.....	239
8.4.3 deleted 表和 inserted 表.....	213	9.4.1 概述.....	239
8.4.4 触发器应用举例.....	213	9.4.2 对象权限管理.....	240
8.4.5 查看导致触发器触发的事件.....	215	9.4.3 架构权限管理.....	246
8.4.6 在数据库设计阶段设计 触发器.....	216	9.4.4 语句权限管理.....	248
8.5 动态 SQL.....	216	9.4.5 禁止权限.....	250
8.5.1 动态 SQL 语句的划分.....	216	9.4.6 查询授权情况.....	251
		9.4.7 所有权链接.....	251
		9.4.8 角色与存取控制.....	252

9.5 数据加密	252	11.1.1 SQL Server 数据库的存储结构	285
9.5.1 概述	253	11.1.2 调整数据库	286
9.5.2 加密的一般步骤及示例	253	11.1.3 文件组	287
9.6 其他安全问题	255	11.1.4 分区	290
9.6.1 审计	255	11.1.5 索引	292
9.6.2 统计数据库	256	11.2 备份与恢复	296
9.6.3 用户定义的安全性措施	256	11.2.1 数据库恢复概述	296
本章小结	257	11.2.2 故障类型	297
习题与思考题	257	11.2.3 备份类型	298
实验 10 用户管理和权限管理	258	11.2.4 日志的概念	299
第 10 章 事务管理与并发控制	261	11.2.5 恢复模型	299
10.1 事务管理	261	11.2.6 备份和恢复策略	300
10.1.1 事务的概念	261	11.2.7 备份操作	301
10.1.2 事务的性质	263	11.2.8 恢复或还原	304
10.1.3 SQL 对事务的支持	264	本章小结	309
10.1.4 简单事务管理	264	习题与思考题	310
10.1.5 事务保存点	265	实验 12 数据恢复	310
10.1.6 隐含事务与自动提交	266	第 12 章 数据库应用和研究的	
10.1.7 使用 TRY...CATCH 控制		新领域	312
事务	267	12.1 概述	312
10.2 并发控制	269	12.2 面向对象数据库	313
10.2.1 干扰问题	269	12.2.1 新的数据库应用和新的	
10.2.2 可串行性	271	数据类型	313
10.2.3 封锁	272	12.2.2 面向对象数据库研究中的	
10.2.4 死锁	277	几个特点	314
10.2.5 活死锁	279	12.2.3 RDBMS、ORDBMS 和	
10.2.6 隔离级别	279	OODBMS 的比较	316
10.2.7 封锁与隔离级别	280	12.3 数据仓库与数据分析	317
本章小结	282	12.3.1 什么是数据仓库	317
习题与思考题	282	12.3.2 数据分析与数据挖掘	321
实验 11 并发事务控制	283	12.4 分布式数据库与云计算	324
第 11 章 数据库存储管理与		12.4.1 分布式数据库的基本概念和	
数据恢复	285	特点	324
11.1 数据库存储管理与存储优化	285	12.4.2 分布式数据库的数据分布方式	326

12.4.3 SQL Server 的分布数据管理概述	327	12.6.1 主动数据库	344
12.4.4 SQL Server 的复制实现技术	329	12.6.2 知识库	345
12.4.5 云计算与云数据管理	333	12.6.3 模糊数据库	346
12.5 大数据与 NoSQL	335	12.6.4 并行数据库	347
12.5.1 什么是大数据	335	12.6.5 多媒体数据库	348
12.5.2 NoSQL	336	本章小结	348
12.6 其他研究方向和领域	344	习题与思考题	349
		参考文献	351

第1章 数据库概论

数据库是现代社会数据管理的工具。数据管理由来已久，数据管理也无处不在，因此数据库已经成为当今社会的重要基础设施。

本章概括介绍与数据库相关的基础知识，使读者了解数据管理及其发展过程、数据库系统的特点、数据模型基础、数据库的三层体系结构、数据库管理系统的功能以及数据库系统的组成等。

1.1 数据管理及其发展过程

早在计算机诞生之前就有数据管理，太遥远的事情这里不再追溯，本书只关心使用计算机进行数据管理的发展历程和管理方法及技术。

1.1.1 数据管理

为了使用计算机进行数据管理和数据处理，必须首先对现实世界中的数据进行认识、分类、组织、归纳、编码，接着采取有效技术和手段将数据存储到计算机中，然后才能对数据进行查询和操作。但并不是一开始就有数据库技术。

在计算机诞生的初期，计算机主要用于科学计算，虽然此时同样有数据管理的问题，但这时的数据管理是以人工的方式进行的，后来发展到文件系统，再后来才是数据库。也就是说，数据管理经历了人工管理阶段、文件系统阶段和数据库系统阶段。使用计算机进行数据管理也经历了一个由简单到复杂的变化过程。

“数据库”作为一个名词已经家喻户晓，现代人的生活已经离不开数据库。在大学校园人人有校园一卡通，一卡通的信息存储在数据库中，到餐厅就餐需要数据库的服务，到图书馆借书需要数据库的服务；走出校园，到银行存取款需要数据库的服务，到商场购物需要数据库的服务，购买火车票、飞机票同样需要数据库的服务。数据库的应用已经遍布人类生活的方方面面。

1.1.2 人工管理阶段

在计算机诞生的初期（20 世纪 50 年代中期以前）计算机主要用于科学计算，如果说是数据处理，也是对数据的计算处理。

这个时期的计算机技术，从硬件看还没有磁盘这样的可直接存取的存储设备，从软件看没有操作系统，更没有管理数据的软件。所有为程序提供完成科学计算和数据处理的数据必须人工来完成，因此这个阶段就称为人工管理阶段。这个时期数据管理的特点有以下几方面。

① 数据不保存。因为还没有磁盘这样的可直接存取的存储设备，当时用于科学计算的数据也没有考虑要长期保存，只是在完成某一个计算或课题时才将数据输入，不仅原始数据不保存，计算结果也不保存。

② 还没有文件的概念。这个时期的数据必须由每个程序的程序员自行组织和安排。

③ 一组数据对应一个程序。每组数据只对应一个应用，即便两个程序用到相同的数据，也必须各自定义、各自组织，数据无法共享、无法相互利用和互相参照。因此，程序和程序之间有大量的数据重复。

④ 没有形成完整的数据管理的概念。

由于以上几个特点及没有对数据进行管理的软件系统，所以这个时期的每个程序都要包括数据存取方法、输入/输出方法和数据组织方法等。因为程序是直接面向存储结构的，所以存储结构的任何一点修改都会导致程序的修改，程序与数据之间不具有独立性，即程序的编写严重依赖于数据的存储结构。

人工管理阶段的特点可用图 1-1 表示。



图 1-1 数据的人工管理阶段

1.1.3 文件系统阶段

文件系统阶段是指 20 世纪 50 年代后期到 20 世纪 60 年代中期这一阶段。从那时起，计算机不仅大量用于科学计算，也开始大量用于信息管理。像磁盘这样的直接存取存储设备已经出现，在软件方面有了操作系统和高级语言，也有了专门用于数据管理的软件——文件系统（或操作系统的文件管理部分）。这个阶段的数据管理有以下一些特点。

① 数据可以长期保存在磁盘上，也可以反复使用，即可以经常对文件进行查询、修改、插入和删除等操作。

② 操作系统提供了文件管理功能和访问文件的存取方法, 程序和数据之间有了数据存取的接口, 程序开始通过文件名和数据打交道, 可以不再关心数据的物理存放位置。因此, 这时也有了数据的物理结构和逻辑结构的区别。程序和数据之间有了一定的独立性。

③ 文件的形式已经多样化。由于有了磁盘这样的直接存取存储设备, 文件也就不再局限于顺序文件, 有了索引文件、链表文件等。因此, 对文件的访问既可以是顺序访问, 也可以是直接访问。但文件之间是独立的, 它们之间的联系要通过程序去构造, 文件的共享性也还比较差。

④ 有了存储文件以后, 数据就不再仅仅属于某个特定的程序, 而可以由多个程序反复使用。但文件结构的设计仍然是基于特定的用途, 程序仍然是基于特定的物理结构和存取方法编制的。因此, 数据的存储结构和程序之间的依赖关系并未根本改变。

⑤ 数据的存取基本上以记录为单位。

图 1-2 示意了数据的文件管理阶段的特点及程序和数据之间的关系。

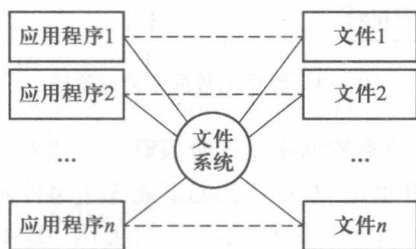


图 1-2 数据的文件管理阶段

虽然文件系统比人工管理有了长足的进步, 但是文件系统所能提供的存取方法和数据管理仅仅是初级水平。无论如何, 文件系统阶段是数据管理技术发展中的一个重要阶段, 在这个阶段中得到充分发展的各种数据结构和算法等都大大丰富了计算机科学, 今天的数据库技术也正是在文件系统的基础上发展起来的。

文件系统还是有以下几方面明显的缺陷。

① 数据冗余大。这是因为每个文件都是为特定的用途设计的, 因此就会造成同样的数据在多个文件中重复存储。

② 数据不一致性。这往往是由数据冗余造成的, 在进行更新时, 稍不谨慎就会造成同一数据在不同文件中的不一致。

③ 程序和数据之间的独立性差。应用程序依赖于文件的存储结构, 使得若修改文件的存储结构则必须修改程序。

④ 数据联系弱。文件与文件之间是独立的, 文件之间的联系必须通过程序来构造。

因此, 文件系统是一个不具有弹性的、无结构的数据集合, 不能反映现实世界事物之间的联系。

1.1.4 数据库系统阶段及其发展过程

数据库系统阶段从 20 世纪 60 年代后期开始, 数据库技术的诞生既有计算机技术的发展做依托, 又有数据管理的需求做动力。数据库的数据不再是面向某个应用或某个程序, 而是面向整个企业(组织)或整个应用, 它克服了上述人工管理阶段和文件系统阶段的缺陷, 图 1-3 示意了这种特点, 1.2 节将详细阐述数据库系统的特点。

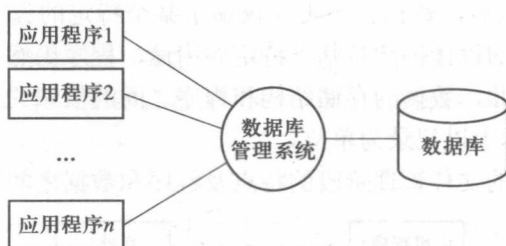


图 1-3 数据的数据库管理阶段

数据管理由文件系统阶段到数据库系统阶段有一个质的变化。数据库技术的诞生可以以 20 世纪 60 年代末和 20 世纪 70 年代初的下述 3 个事件做标志。

① 1968 年研制成功、1969 年形成产品的美国 IBM 公司的数据库管理系统 (Information Management System, IMS) 的问世。该系统使用层次结构来表示数据和数据之间的关联或联系, 因此也说该系统支持的是层次数据模型。

② 美国数据系统语言协会 (Conference On DAta SYstem Language, CODASYL) 下属的数据库任务组 (DataBase Task Group, DBTG) 对数据库方法进行了系统的研究, 在 20 世纪 60 年代末和 20 世纪 70 年代初发表了若干个报告, 称为 DBTG 报告, 该报告建立了数据库技术的很多概念、方法和技术。DBTG 所提议的方法可以表示任意数据之间的联系, 突破了层次结构的限制, 因此将其称为网状数据模型。

③ 从 1970 年起, IBM 的研究员 E.F.Codd 发表了一系列论文, 提出了数据库的关系模型 (即数据之间的联系用数学上的“关系”来表示), 开创了数据库关系方法和关系数据库理论的研究, 为关系数据库的发展和理论研究奠定了基础。

可以说, 20 世纪 70 年代是以 IMS 为代表的层次数据库和以 DBTG 为代表的网状数据库的鼎盛时期。这些数据库都有很高的效率, 特别是 DBTG 报告包含了很多完备的概念和技术, 这大大推动了计算机在信息管理领域的应用, 也使许多商品化的数据库开始出现, 数据库技术日益广泛地应用到企业管理、交通运输、情报检索、军事指挥、政府

管理和辅助决策等各领域。层次数据库和网状数据库是面向专业人员的, 要求使用人员有较高的技术水平和专业水平, 使用起来比较困难。

层次数据库和网状数据库也可以看作是第一代数据库系统。

20 世纪 70 年代也是关系数据库的萌动时期。与层次数据库和网状数据库相比, 研究关系数据库的出发点之一就是简单、易用, 这样就要尽可能地减少人为的操作, 让计算机自动完成更多的工作。受当时计算机硬件和软件技术的约束, 关系数据库经历了比较长时间的研究过程。尽管在 20 世纪 60 年代末到 20 世纪 70 年代初就提出了关系数据库的概念, 但是关系数据库真正得到广泛应用则是进入 20 世纪 80 年代以后的事情。

关系数据库可以看作是第二代数据库系统。

具体的数据模型和特点将在 1.3 节做简单介绍。第 3 章将对关系数据模型做详细介绍。

在国内, 20 世纪 70 年代就开始使用数据库的主要是国家各部委、国防军事、气象预报、石油勘探等一些特殊的部门和行业。而数据库技术真正得到广泛的推广和使用, 可以说是从 20 世纪 80 年代初的 dBASE II 开始的。尽管 dBASE II, 甚至其后的 xBASE 系列都不能称为是一个完备的数据库管理系统, 但是它们支持基本的关系数据模型, 使用起来非常方便, 一般也能满足中、小规模信息管理的应用。这些现在看起来不起眼的小软件, 当时为数据库和计算机应用的普及做出了极其重要的贡献。

1.1.5 数据管理和数据库技术的持续发展

可以说自 20 世纪 80 年代以来, 关系数据库系统一直占据主流地位, 层次数据库和网状数据库由于自身的缺点很快被关系数据库所取代。从 20 世纪 80 年代算起, 又经过了将近 40 年, 关系数据库系统的技术也在不断完善和提高, 有关数据管理和数据库的新的研究课题不断取得进展, 如分布式数据库已成为实用技术, 基于关系模型、支持面向对象方法的关系对象模型数据库也已问世, 如开放数据库技术和互联网上的 Web 数据库、支持移动应用的移动数据库等, 同时随着大数据时代的到来, 数据管理的范畴已经不局限于结构化数据管理, 关系数据库不再一枝独秀, NoSQL 技术应运而生 (这里暂且把 NoSQL 泛称为非关系数据库, 具体内容在 12.5 节介绍)。

现在的数据库不仅可以管理一般数据, 还可以管理对象数据, 可以进行知识管理等, 相应的技术在持续、交叉发展, 很难再划分第 n 代数据库, 所以人们也把这些应用了新技术的数据库称作新一代数据库系统, 把基于这些新技术的数据库应用称作高级数据库系统阶段。

数据库是数据管理的基础, 但是随着互联网的普及, 人们“制造”数据的速度越来越快、能力越来越强, 数据的形式也多种多样, “大数据”这个名词也已经耳熟能详。随着计算机技术的发展和各种应用的普及, 数据库及数据管理技术还会朝着支持更大规模、