



普通高等教育“十二五”规划教材

数据库系统原理及应用



主 编 李晓峰 李 东
副主编 翟 霞 吕志峰
主 审 黄国兴



中国水利水电出版社
www.waterpub.com.cn

普通高等教育“十二五”规划教材

数据库系统原理及应用

主 编 李晓峰 李 东

副主编 翟 霞 吕志峰

主 审 黄国兴



中国水利水电出版社
www.waterpub.com.cn

内 容 提 要

本书从不同层次、角度,由浅入深地介绍数据库系统的原理和应用,吸取新思想、新技术,详细介绍数据库系统的基本概念、关系模型的运算理论、SQL语言和SQL Server 2008、规范化设计理论、数据库设计、数据保护等基本知识,并对目前正在发展的高级数据库技术作了介绍,强调内容的特色化与创新性。

本书本着突出理论知识的应用和实践能力的培养,基础理论以必需、够用为度,专业教学加强针对性和实用性等原则,共分三部分:基础篇、应用篇和高级篇。“基础篇”内容涉及数据库的基本概念、操作、模型;“应用篇”结合SQL Server 2008和计算机实验签到系统实例介绍数据库管理和高级应用技术,达到理论与实践紧密结合;“高级篇”内容包括一些前沿技术,如“分布式数据库”、“面向对象数据库”。

本书内容循序渐进,深入浅出,概念清晰,条理性强,每一章都给出了大量的实例,并且每章后面附有大量习题,从各种不同角度帮助读者了解和掌握所学的知识点,完成全部知识体系的衔接。

本书可作为大学计算机及相关专业的本科教材,也可供从事计算机软件工作的科技人员、工程技术人员以及其他相关人员参阅。

本书提供免费电子教案,读者可以从中国水利水电出版社网站和万水书苑免费下载,网址为: <http://www.waterpub.com.cn/softdown/>和 <http://www.wsbookshow.com>。

图书在版编目(CIP)数据

数据库系统原理及应用 / 李晓峰, 李东主编. — 北京: 中国水利水电出版社, 2011.2
普通高等教育“十二五”规划教材
ISBN 978-7-5084-8294-1

I. ①数… II. ①李… ②李… 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(2011)第001840号

策划编辑: 石永峰 责任编辑: 张玉玲 封面设计: 李 佳

书 名	普通高等教育“十二五”规划教材 数据库系统原理及应用
作 者	主 编 李晓峰 李 东 副主编 翟 霞 吕志峰 主 审 黄国兴
出版发行	中国水利水电出版社 (北京市海淀区玉渊潭南路1号D座 100038) 网址: www.waterpub.com.cn E-mail: mchannel@263.net (万水) sales@waterpub.com.cn 电话: (010) 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	北京市天竺颖华印刷厂
规 格	184mm×260mm 16开本 23印张 624千字
版 次	2011年2月第1版 2011年2月第1次印刷
印 数	0001—3000册
定 价	42.00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

前 言

数据库技术是计算机学科中比较活跃的一个分支，是高校计算机专业教学中一门重要的核心课程，已经成为计算机信息系统与应用系统的核心技术。随着数据库应用技术的发展，越来越多的软件开发人员希望了解并掌握数据库的开发和管理方法。数据库的应用为高效、精确地处理各种数据创造了条件。数据库的建设规模、数据库信息量的大小和使用频率，已成为衡量一个国家信息化程度的重要标志。

本书在内容编排上，由理论到实践、从技术基础到具体开发应用，将数据库原理知识与实际数据库开发工具结合在一起，旨在培养学习者提高综合实践与创新能力。本书根据数据库发展的过程与特点，从不同角度出發，凝煉出数据库发展的三条线索：数据模型经历了从层次型、网状型、关系型、对象关系型、面向对象型到网络型的演变过程；数据库查询语言经历了交互式、嵌入式、对象关系式、面向对象式等演变过程；数据库设计从实体联系图、对象联系图的发展过程，将全书内容提携起来，使得复杂内容条理化、清晰化，易于学生掌握。

本书突出理论知识的应用和实践能力的培养，由浅入深，并突出应用型人才培养要求，对实用性强的内容有所侧重，每一章都附有相应的习题，便于读者加深理解和巩固提高。

本书是作者多年从事数据库教学的经验和感受的总结，特点是覆盖的知识比较全面，既包括数据库理论，又包括数据库应用技术，并提供大量的实例分析，为读者理解相关知识点，提高实际能力提供了方便。

本书第1、3、6章由李晓峰、李东编写，第2和7章由翟霞、黄磊编写，第4和5章由吕志峰、孙海龙编写，第8章由张振蕊、李兰竹编写，第9章由姚路、杨丽丽编写，第10章由孙鹤、冯占伟编写，全书由李晓峰统稿。

值此本书付梓之际，首先向华东师范大学黄国兴教授表示感谢，他为审阅本书付出了艰辛的劳动并提出了很多宝贵意见，其次还要感谢哈尔滨工业大学李东教授参与了编写工作，最后感谢中国水利水电出版有关领导和编辑的大力支持和帮助。

在本书编写过程中，作者参考了大量的相关书籍和资料，在此对这些参考文献的作者表示感谢。

由于编者水平有限，书中难免存在错误及不当之处，敬请广大专家和读者批评指正，以便及时修订和补充。作者邮箱是 mberse@126.com。

编 者

2010年11月

目 录

前言

第一篇 基础篇

第 1 章 数据库系统概论 1	2.2.2 关系模型的数据操作..... 29
1.1 数据库系统基本概念..... 1	2.2.3 关系完整性约束..... 30
1.1.1 信息、数据与数据处理..... 1	2.3 关系代数..... 31
1.1.2 数据库..... 2	2.3.1 关系代数的传统集合运算..... 32
1.1.3 数据库管理系统与管理信息系统..... 3	2.3.2 关系代数专门的关系运算..... 33
1.1.4 数据库系统..... 4	2.3.3 关系代数表达式应用举例..... 37
1.2 数据库技术的产生与发展..... 6	2.4 关系演算..... 37
1.2.1 数据管理技术的发展..... 6	2.4.1 元组关系演算..... 37
1.2.2 数据库系统的特点..... 9	2.4.2 域关系演算..... 39
1.3 数据模型..... 10	2.5 关系系统的查询优化技术..... 40
1.3.1 信息的三个世界..... 10	2.5.1 查询优化问题的提出..... 41
1.3.2 模型与数据模型..... 12	2.5.2 查询优化的一般策略..... 43
1.3.3 数据模型的基本组成..... 12	2.5.3 关系代数表达式的等价变换规则..... 43
1.3.4 数据模型的分类..... 13	2.5.4 关系表达式的优化算法..... 44
1.4 数据库系统的体系结构..... 15	本章小结..... 47
1.4.1 数据库三级模式结构..... 16	习题二..... 48
1.4.2 数据库应用系统的体系结构..... 18	第 3 章 关系数据库设计理论 50
1.5 数据库系统研究领域..... 20	3.1 关系模式设计问题..... 50
1.5.1 数据库技术、方法、理论..... 20	3.1.1 数据冗余与操作异常..... 50
1.5.2 DBMS 产品及工具..... 20	3.1.2 冗余产生原因分析和问题解决思路..... 51
1.5.3 数据库技术与相关技术的融合..... 21	3.2 关系模式的规范化..... 52
本章小结..... 21	3.2.1 函数依赖..... 52
习题一..... 22	3.2.2 码的形式定义..... 54
第 2 章 关系数据库 23	3.2.3 关系模式基础范式..... 55
2.1 关系数据库的基本概念..... 23	3.2.4 多值依赖与第四范式..... 59
2.1.1 关系的数学定义..... 23	3.3 数据依赖的公理系统..... 62
2.1.2 关系的性质..... 26	3.3.1 Armstrong 公理..... 62
2.1.3 关系模式与关系数据库..... 27	3.3.2 属性集的闭包及其算法..... 63
2.2 关系数据库的数据模型..... 28	3.3.3 候选码的求解理论和算法..... 65
2.2.1 关系模型结构..... 28	3.3.4 函数依赖集的等价与覆盖..... 66

3.3.5 最小函数依赖集	66	3.5.1 规范化的优缺点	75
3.4 关系模式的分解	68	3.5.2 反规范化设计	75
3.4.1 模式分解的等价性	68	3.5.3 规范化设计与反规范化设计	76
3.4.2 无损连接分解	69	3.5.4 规范化示例	76
3.4.3 保持依赖性分解	71	本章小结	78
3.4.4 关系模式分解算法	73	习题三	79
3.5 规范化理论的应用	75		

第二篇 应用篇

第4章 SQL Server 2008 和关系数据库

标准语言 SQL	81
4.1 SQL Server 2008 概述	81
4.1.1 SQL Server 2008 的发展历史	81
4.1.2 SQL Server 2008 简介	82
4.1.3 SQL Server 2008 的版本	83
4.1.4 SQL Server 2008 的安装与配置	86
4.2 SQL Server 2008 的管理工具	97
4.2.1 SQL Server 2008 管理工具一览	97
4.2.2 SQL Server 2008 Management Studio	98
4.2.3 SQL Server 2008 配置管理器	100
4.3 SQL Server 2008 的体系结构	101
4.3.1 逻辑结构	101
4.3.2 物理结构	102
4.4 关系数据库语言 SQL	104
4.4.1 SQL 的发展史	104
4.4.2 SQL 的功能特点	105
4.4.3 SQL 的基本组成	105
4.5 Transact-SQL 的数据定义语言	107
4.5.1 数据库管理与使用	107
4.5.2 表的管理与使用	118
4.5.3 视图的管理与使用	126
4.5.4 索引的管理与使用	132
4.5.5 同义词的管理和使用	134
4.6 Transact-SQL 的数据查询语言	135
4.6.1 SELECT 的基本结构	136
4.6.2 简单查询	136
4.6.3 连接查询	144
4.6.4 子查询	147

4.6.5 集合查询	149
4.6.6 递归合并查询	150
4.7 Transact-SQL 的数据操纵语言	152
4.7.1 数据插入	152
4.7.2 数据修改	153
4.7.3 数据删除	154
4.8 Transact-SQL 的数据控制语言	155
4.8.1 数据控制的方法	155
4.8.2 权限的授予与收回	155
4.9 嵌入式 SQL	157
4.9.1 嵌入式 SQL 的实现方式	158
4.9.2 嵌入式 SQL 语言的分类	158
4.9.3 嵌入式 SQL 和宿主语言之间的通信	159
4.9.4 嵌入式 SQL 的使用规定	160
4.9.5 嵌入式 SQL 的使用技术	160
4.9.6 嵌入式 SQL 应用实例	163
本章小结	164
习题四	164

第5章 SQL Server 2008 高级应用

5.1 Transact-SQL 程序设计	167
5.1.1 Transact-SQL 程序结构	167
5.1.2 Transact-SQL 程序的批处理	168
5.1.3 系统数据类型	168
5.1.4 常量与变量	171
5.1.5 注释符与运算符	173
5.1.6 标准(库)函数	174
5.1.7 流程控制语句	175
5.1.8 常用命令	179
5.2 存储过程	180

5.2.1 存储过程的概念	181	6.3.3 数据抽象与局部概念模式设计	220
5.2.2 存储过程的优点	181	6.3.4 全局概念模式设计	222
5.2.3 存储过程的分类	181	6.4 逻辑设计	224
5.2.4 存储过程的建立	182	6.4.1 逻辑设计的步骤	224
5.2.5 执行存储过程	184	6.4.2 E-R 模型向关系模型的转换	224
5.2.6 修改存储过程	185	6.4.3 模型的优化	228
5.2.7 删除存储过程	186	6.4.4 设计用户外模式	230
5.3 触发器	187	6.4.5 E-R 模型实例分析	230
5.3.1 触发器的概念	187	6.5 物理设计	233
5.3.2 触发器的作用	187	6.5.1 物理设计的要求和内容	233
5.3.3 触发器的种类	188	6.5.2 存取方法	234
5.3.4 DML 触发器的分类	188	6.5.3 存储结构	235
5.3.5 DML 触发器的工作原理	188	6.5.4 性能评价	236
5.3.6 设计 DML 触发器的注意事项 及技巧	189	6.6 数据库的实施	236
5.3.7 创建 DML 触发器	191	6.6.1 数据库结构建立	236
5.3.8 创建 Instead Of 触发器	194	6.6.2 数据载入	236
5.3.9 查看 DML 触发器	195	6.6.3 编写、调试应用程序	237
5.3.10 修改 DML 触发器	196	6.6.4 数据库试运行	237
5.3.11 删除 DML 触发器	197	6.6.5 整理文档	237
5.3.12 DDL 触发器	197	6.7 数据库的运行维护	238
5.3.13 存储过程和触发器的区别	200	6.8 数据库应用系统设计实例	239
5.4 用户自定义函数	200	6.8.1 系统需求分析	239
5.4.1 创建用户自定义函数	201	6.8.2 系统概念设计	242
5.4.2 修改和删除用户自定义函数	204	6.8.3 系统的逻辑设计	243
本章小结	204	6.8.4 系统物理设计	247
习题五	204	本章小结	251
第 6 章 数据库设计	207	习题六	252
6.1 数据库设计概述	207	第 7 章 数据库安全性与完整性	254
6.1.1 数据库设计的任务、特点和方法	207	7.1 数据安全与保护概述	254
6.1.2 数据库设计过程	209	7.2 数据库的安全性	255
6.2 需求分析	211	7.2.1 对数据库安全的威胁	255
6.2.1 需求分析的任务	211	7.2.2 数据库安全性级别	255
6.2.2 需求分析的步骤	211	7.2.3 数据库安全性控制	256
6.2.3 数据流图	213	7.2.4 SQL Server 的安全性控制	259
6.2.4 数据字典	213	7.3 数据库的完整性	275
6.3 概念设计	215	7.3.1 完整性约束条件	276
6.3.1 E-R 模型	216	7.3.2 完整性控制	278
6.3.2 概念设计的方法和步骤	220	7.3.3 SQL Server 完整性的实现	280
		本章小结	286

习题七	286
第 8 章 数据库事务管理	288
8.1 事务与事务管理	288
8.1.1 事务的概念及 ACID 特性	288
8.1.2 事务的基本操作与活动状态	290
8.2 并发控制	291
8.2.1 事务的并发执行	291
8.2.2 并发操作引起的不一致问题	291
8.2.3 封锁与封锁协议	293
8.2.4 活锁与死锁	295

8.2.5 并发调度的可串行性	297
8.2.6 两段锁协议	299
8.2.7 SQL Server 并发控制的实现	301
8.3 数据库恢复技术	305
8.3.1 数据库故障分类	305
8.3.2 数据库恢复实现技术	306
8.3.3 SQL Server 数据恢复机制的实现	310
本章小结	322
习题八	322

第三篇 高级篇

第 9 章 分布式数据库	324
9.1 分布式数据库简介	324
9.1.1 分布式数据库的特点	324
9.1.2 分布式数据库的优缺点	325
9.1.3 分布式数据库的分类	326
9.2 分布式数据库的体系结构	326
9.3 分布式数据库管理系统	329
9.4 分布式数据存储	330
9.4.1 数据分片	330
9.4.2 数据分配	332
9.5 分布式数据库的组成	333
9.6 分布式查询	333
9.7 分布式数据库的发展及应用趋势	335
本章小结	336
习题九	337
第 10 章 对象数据库	338
10.1 面向对象数据库	338
10.1.1 面向对象数据模型	339
10.1.2 面向对象数据库管理系统	339
10.1.3 面向对象数据库系统	340
10.1.4 面向对象技术与数据库技术相结合的	

多种方法	340
10.1.5 面向对象数据库的不足	341
10.2 对象关系数据库	341
10.2.1 关系模型扩充——后关系模型	342
10.2.2 后关系模型扩充——对象关系 数据模型	346
10.3 对象关系数据库管理系统	348
10.4 对象关系数据库的定义语言	350
10.4.1 对象关系数据库模型的定义	350
10.4.2 数据类型定义	350
10.4.3 继承性定义	352
10.4.4 引用类型定义	353
10.5 对象关系数据库的数据查询	356
10.5.1 嵌套与元组变量设置	356
10.5.2 路径表达式	356
10.5.3 复合值创建和查询	357
10.5.4 面向对象数据库与对象关系 数据库的比较	358
本章小结	358
习题十	359
参考文献	360

第一篇 基础篇

第 1 章 数据库系统概论

数据库技术主要研究如何存储、使用、设计和管理数据。它自 20 世纪 60 年代产生至今已有 40 多年的历史，造就了 C.W.Bachman、E.F.Codd 和 James Gray 三位图灵奖获得者，带动了一个巨大的软件产业。数据库技术一直是最活跃、发展速度最快、应用最广的 IT 技术之一。

数据库技术的发展经历了三代的演变：第一代数据库以网状模型和层次模型为特征；第二代数据库以关系模型为基础，在 20 世纪 70 年代末到 80 年代得到广泛应用；从 20 世纪 90 年代开始，随着网络和多媒体技术的迅速发展，数据库的应用也得到了更为广泛的拓展，进入了一个新的发展时期，融合了多种技术，面向对象是其主要特征。

在过去 20 年里，数据库技术的应用有了巨大的增长，几乎每个行业都要用数据库来存储、操纵、检索数据。如今，数据库和数据库系统已经成为现实生活中不可缺少的一部分。我们每天都会或多或少地、不知不觉地和数据库发生某些联系。我们可能会去邮局订阅一本杂志；可能会去预订旅馆房间或机票；可能会为找一本书而去检索一个由计算机管理的图书馆目录；也可能会去银行取款，所有这些活动都会涉及对数据库的使用。在商业、医疗保健、教育、政府组织、图书馆、军事、工业控制等领域都有数据库的应用，它已成为信息管理、电子商务、网络服务等应用系统的核心技术和重要基础。

本章主要介绍数据库中有关的基本概念、数据模型及数据库系统的体系结构、数据库的发展状况、数据库系统研究领域等基本知识，为以后各章的学习打下基础。

1.1 数据库系统基本概念

1.1.1 信息、数据与数据处理

信息和数据是数据处理中的两个基本概念，它们有着不同的含义。

1. 数据

(1) 数据的定义。数据 (Data) 是描述事物的符号记录，也是数据库中存储、用户操纵的基本对象。数据不仅是数值，还可以是文字、图形、图像、声音、视频等。数据是信息的符号表示。例如描述新生信息，可用一组数据“2010 (年)，哈尔滨工业大学华德学院计算机系学生，22 (个班)，920 (人)”。这些符号被赋予了特定的语义，具体描述了信息，因此它们就具有了传递信息的功能。

(2) 数据的表现形式。数据的表现形式是多种多样的，主要有数字、文字、声音、图形和图像等形式。数据的形式有时还不能完全表达其内容，需要经过解释。因而数据与数据的解释是不可分的，数据的解释是指对数据含义的说明，数据的含义定义为数据的语义。

2. 信息

(1) 信息的定义。信息 (Information) 是指数据经过加工处理后所获取的有用知识, 是以某种数据形式表现的。信息与原始数据同样重要。原始资料需要加工处理才能成为有用的信息。例如气象局每天从各地气象站收集到大量有关气象的图形或文字记录表示各地气象信息的数据。当气象局对这些记录数据进行综合处理、分析、判断, 作出气象预报时, 这些处理过的数据又表示了新的信息。信息对决策或行动是有价值的。例如, 人们可以根据气象预报安排生产和生活。

(2) 信息的特征。信息有以下 3 个重要特征:

1) 信息源于物质和能量。信息的传递需要物质载体, 信息的获取和传递需要消耗能量, 信息不可能脱离物质而存在, 其载体可以是广播、报纸和杂志等。

2) 信息是可以感知的。人类对客观事物的感知, 可以通过感觉器官, 也可以通过各种仪器仪表来获得, 不同的信息源于不同的感知方法。例如, 我们从广播上获得的信息是通过听觉器官来感知的, 而从报纸上获得的信息则是通过视觉来感知的。

3) 信息是可存储、加工、传递和再生的。人类可以用大脑存储少量的信息, 用文字长久存储信息, 而计算机的发展进一步扩大了信息存储的范围。人类可以对所收集的信息进行取舍整理以及通过各种手段和方法进行传递和再生。

(3) 数据与信息的关系。一方面, 数据和信息是两个相互联系的概念: 数据是信息的具体表现形式, 信息是数据有意义的表现。另一方面, 数据和信息又是相互区别的, 某一具体的信息和表示它的数据之间的对应关系又会因环境而变化, 并非所有的数据都能表示信息; 而同一信息可以有不同的符号表示, 同一数据也可以有不同的解释。因此, 信息是人们消化理解了的数据, 是抽象的, 它不随数据设备所决定的数据形式而改变, 而数据的表现方式及存在方式是客观具体的。

3. 数据处理

数据是人类活动的重要资源。数据处理是对数据进行收集、管理、加工、传播等一系列工作。其目的是从大量的原始数据中抽取和推导出有价值的信息, 以进行各种应用。

目前在计算机的各类应用中, 用于数据处理的约占 80%。而数据处理中的数据管理是指对数据的组织、存储、检索、维护等工作。由于数据管理技术的优劣直接影响数据处理的效率, 因此它是数据处理的核心。数据库技术正是瞄准这一目标而发展起来的专门技术。我们可以简单地用下式表示信息、数据与数据处理的关系:

$$\text{信息} = \text{数据} + \text{数据处理}$$

可以将数据形象化地比喻为原料, 是输入; 信息就像产品, 是输出; 而数据处理是把原料变成产品的过程。从这个角度出发, “数据处理”的真正含义应该是为了产生信息而处理数据。

1.1.2 数据库

“数据库”这个名词起源于 20 世纪中叶, 当时美军为作战指挥需要建立起了一个高级军事情报基地, 把收集到的各种情报存储在计算机中, 并称之为“数据库”。起初人们只是简单地将数据库看做是一个电子文件柜, 即将它当成是一个存储数据文件的仓库或容器。后来随着数据库技术的产生, 人们引申并沿用了该名词, 给“数据库”这个名词赋予了更深层的含义。

那么, 数据库到底是什么呢? 可以归纳为: 数据库 (DataBase) 是按一定结构组织并长期存储在计算机内的、有组织、可共享的大量数据的有机集合。

1.1.3 数据库管理系统与管理信息系统

1. 数据库管理系统

数据库管理系统 (DataBase Management System, DBMS) 是管理和维护数据库的系统软件, 是数据库和用户之间的一个接口, 其主要作用是在数据库建立、运行和维护时对数据库进行统一管理和控制。

DBMS 的主要功能包括以下几个方面:

(1) 数据定义功能。DBMS 提供数据定义语言 DDL (Data Definition Language) 来定义数据库的三级模式和两级映像, 定义数据完整性和保密限制等约束。

(2) 数据操纵功能。DBMS 提供数据操纵语言 DML (Data Manipulation Language) 来实现对数据库的操作, 如查询、插入、修改和删除。DML 有两类: 一类是嵌入在宿主语言中, 如嵌入在 C、Java、Delphi、PowerBuilder 等高级语言中, 这类 DML 称为宿主型 DML; 另一类是可以独立交互使用的 DML, 称为自主型或内含型 DML, 常用的有 Transact-SQL、SQL Plus 等。

(3) 数据保护功能。数据库中的数据是信息社会的战略资源, 因此对数据库的保护是至关重要的。DBMS 对数据的保护主要包括 4 个方面: 数据安全性控制、数据完整性控制、数据并发控制和数据库的恢复。

1) 数据安全性控制。数据安全性控制是对数据库的一种保护措施。它的作用是防止未被授权的用户破坏或存取数据库中的数据。用户首先必须向 DBMS 标识自己, 在系统确定该用户有权对指定的数据进行存取时才能存取数据。防止未被授权的用户蓄意或无意地修改数据是很重要的, 否则会导致数据完整性的破坏, 从而使企事业单位蒙受巨大损失。

2) 数据完整性控制。数据完整性控制是 DBMS 对数据库提供保护的另一个重要方面。完整性是数据的准确性和一致性的测度。当数据加入到数据库中时, 对数据的一致性和合法性进行检验将会提高数据的完整性程度。完整性控制的目的是保证进入数据库中的数据的语义的正确性和有效性, 防止操作对数据的修改会违反其语义。因此, DBMS 允许对数据库中的各类数据定义若干语义完整性约束, 由 DBMS 强制执行。

3) 并发控制。DBMS 一般允许多个用户并发地访问数据库, 即数据共享。但是多个用户同时对数据库进行访问可能会破坏数据的正确性, 例如存储错误的数据库, 或者读取不正确的数据库, 即所谓的“脏数据”。因此 DBMS 中必须具有并发控制机制, 解决多用户下的并发冲突。

4) 恢复功能。恢复功能是保护数据库的又一个重要方面。数据库在运行中可能会出现各种故障, 如停电、软硬件错误等, 导致数据库的损坏或不一致。DBMS 必须把处于故障中的数据库恢复到以前的某个正确状态, 保持数据库的一致性。DBMS 的其他保护功能还有系统缓冲区管理以及数据存储的某些自适应调节机制。

(4) 数据库维护功能。DBMS 提供一系列的实用程序来完成包括数据库的初始数据的装入、转化功能, 数据库的存储、恢复功能, 数据库的重新组织功能和性能监视、分析功能等。

(5) 数据字典。数据字典 (Data Dictionary, DD) 是对数据库结构的描述, 不仅存放着对实际数据库三级模式的定义, 是数据库系统中各种描述信息和控制信息的集合, 而且还存放数据库运行时的系统信息, 如记录个数和访问次数等。数据字典是数据库管理的有力工具。数据字典是数据库系统的一部分, 但用户通常不能直接访问它, 只有 DBMS 才能对它进行访问

说明: 从操作系统角度: DBMS 是使用者, 它建立在操作系统的基础之上, 需要操作系统提供底层服务, 如创建进程、读写磁盘文件、CPU 和内存管理等。

从数据库角度：DBMS 是管理者，是数据库系统的核心，是为数据库的建立、使用和维护而配置的系统软件，负责对数据库进行统一的管理和控制。

从用户角度：DBMS 是工具或桥梁，是位于操作系统与用户之间的一层数据管理软件。用户发出的或应用程序中的各种操作数据库的命令，都要通过 DBMS 来执行。

常用的 DBMS 可分为文件型和服务器型两种。文件型 DBMS 有 FoxPro、Prodax、Access 等，服务器型 DBMS 有 Oracle、SQL Server、Informix、Sybase、DB2 等。

2. 管理信息系统

管理信息系统 (Management Information System, MIS) 是计算机应用领域的一个重要分支。MIS 帮助人们完成原来需要手工处理的复杂工作，它不仅能明显地提高工作效率，降低劳动强度，而且能提高信息管理的质量或水平。因而，MIS 不是在模拟手工劳动，它要更合理地组织数据和更科学地管理数据，为控制事务发展提供控制信息，为预测事务变化的状态提供事务发展趋势信息和变化规律的信息。MIS 的应用非常广泛，它可以用于事务管理、计算机辅助设计、计算机图形及人工智能等系统中，即所有数据量大、数据成分复杂的地方都可以使用 MIS 进行其数据管理的工作。

MIS 以数据库技术为基础，它的核心是数据库。MIS 的数据存放在数据库中，数据库技术为 MIS 提供了数据管理的手段，DBMS 为 MIS 提供了系统设计的方法、工具和环境。学习数据库及 DBMS 的基础理论和设计方法，目的就是要掌握数据库系统的设计、管理和应用，以便能够胜任 MIS 的设计、开发和应用工作。

1.1.4 数据库系统

数据库系统 (DataBase System, DBS) 是实现有组织地、动态地存储大量关联数据，方便多用户访问的计算机软件、硬件和数据资源组成的系统。

在一般计算机系统中引入数据库技术后即形成数据库系统，故可以简单地说数据库系统是具有管理数据库功能的计算机系统，即可简化表示为：一个数据库系统通常由数据库、硬件、软件、用户四部分组成。数据库系统如图 1.1 所示。

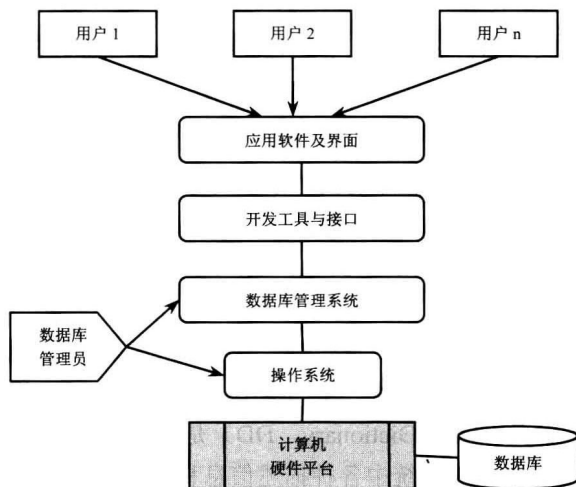


图 1.1 数据库系统图示

1. 数据库

数据库 (DataBase, DB) 是指相互关联的数据集合。一般定义为：长期存储在计算机内的、

有组织的、可共享的数据集合。

数据库由以下两大部分组成：

- 用户数据库。
- 系统数据库，即数据字典。

数据字典是一组关于数据的数据，也叫元数据（meta-data）。数据库系统运行时要涉及许多对象，如数据结构、数据类型、文件、记录、数据项、用户和程序等均要反映在数据字典中。通过数据字典，数据处理人员能够有效地控制和管理用户数据库。

数据库有以下几个特点：

- (1) 数据结构化。在数据库系统中，数据不再像文件系统中的数据那样从属于特定的应用，而是面向全局的复杂的数据结构，数据的结构化是数据库系统区别于文件系统的根本特征。
- (2) 数据共享。数据库系统中的数据可供多个用户、多种语言和多个应用程序共享，这是数据库技术的基本特征。数据共享大大减少了数据冗余和不一致性，大大提高了数据的利用率和工作效率。
- (3) 数据独立性。数据的独立性包括数据的物理独立性和逻辑独立性。用户的应用程序与存储在磁盘上的数据库中的数据是相互独立的，这就是数据的物理独立性；同时用户的应用程序与数据的逻辑结构是相互独立的，这就是数据的逻辑独立性。它不会因一方的改变而改变，这大大减少了应用程序设计和数据库维护的工作量。

2. 硬件

计算机系统的硬件包括中央处理器、内存、外存、输入/输出设备等。在数据库系统中特别要关注内存、外存、I/O 存取设备、可支持的节点数和性能稳定性指标，现在还要考虑支持联网的能力和必要的后备存储器等因素。此外，还要求系统有较高的通道能力，以提高数据的传输速度。

3. 软件

数据库系统的软件主要包括操作系统（OS）、数据库管理系统（DBMS）、各种宿主语言和应用开发支撑软件。DBMS 是在操作系统的文件系统基础上发展起来的，在操作系统的支持下工作，是数据库系统的核心软件。为了开发应用系统，需要各种宿主语言，这些语言大部分属于第三代语言（3GL）范畴，例如 COBOL、C、PL/I 等；有些是面向对象的程序设计语言，例如 C++、Java 等。应用开发支撑软件是为应用开发人员提供高效率的、多功能的交互式程序设计系统，一般属于第四代语言（4GL）范畴，包括报表生成器、表格系统、图形系统，以及具有数据库访问、I/O 功能和数据字典功能的软件开发系统等。当前比较流行的应用开发工具主要有 PowerBuilder、Delphi、Visual Basic 等。

4. 用户

管理、开发和使用数据库系统的用户主要有普通用户、应用程序员和数据库管理员。数据库系统中不同人员涉及不同的数据抽象级别，具有不同的数据视图。

(1) 普通用户。普通用户主要指应用程序的最终用户，也称终端用户。它们通过应用程序的用户接口使用数据库，目前常用的接口方式有菜单驱动、表格操作、图形显示、报表生成等，这些接口使得用户的操作简单易学，适合非计算机专业人员的使用。

(2) 应用程序员。应用程序员负责设计和调试数据库系统的应用程序。他们通常使用 4GL 开发工具编写数据库应用程序，供各种用户使用。

(3) 系统分析员和数据库设计员。系统分析员是数据库系统设计中的高级人员，主要负责数据库系统建设的前期工作，包括应用系统的需求分析、规范说明和数据库系统的总体设计等。数据库设计人员参与用户需求调查、应用系统的需求分析后，主要负责数据库的设计，包括各级模式的

设计、确定数据库中的系统数据等。

(4) 数据库管理员 (DataBase Administrator, DBA)。DBA 是 DBS 中的一个重要角色, 主要负责设计、建立、管理和维护数据库, 协调各用户对数据库的要求等。因此, DBA 要熟悉、掌握程序语言和系统软件 (如 OS、DBMS 等), 充分了解各种用户的需求, 了解各应用部门的所有业务工作, 具有系统分析员和运筹学专家的知识。所以, DBA 通常是信息技术方面的专业人员, 负责全局控制。DBA 不一定是一个人, 它往往是一个工作小组, 具有如下职责:

1) 决定数据库中的信息内容和结构。数据库中要存放哪些信息, DBA 要参与决策。因此 DBA 必须参加数据库设计全过程, 并与用户、应用程序员、系统分析员密切合作, 共同协商, 搞好数据库设计。

2) 决定数据库的存储结构和存取策略。DBA 要综合各用户的应用要求, 和数据库设计人员共同决定数据的存储结构和存储策略, 以求获得较高的存取效率和存储空间利用率。

3) 定义数据的安全性和完整性。DBA 的重要职责是保证数据库的安全性和完整性, 因此 DBA 负责确定各个用户对数据库的存取权限、数据的保密级别和完整性约束条件。

4) 监控数据库的使用和运行。DBA 还有一个重要职责是监测数据库系统的运行情况, 及时处理运行过程中出现的问题。比如系统发生各种故障时, 数据库会因此遭到不同程度的破坏, DBA 必须在最短时间内将数据库恢复到正确状态, 并尽可能不影响或少影响计算机系统其他部分的正常运行。为此, DBA 要定义和实施适当的后备和恢复策略, 如周期性的转储数据、维护日志等。

5) 数据库的改进和重组重构。DBA 还负责在系统运行期间监视系统的空间利用率、处理效率性能指标, 对运行情况进行记录、统计分析, 依靠工作实践并根据实际应用环境不断改进数据库设计。不少数据库产品都提供了对数据库运行状况进行监视和分析的工具, DBA 可以使用这些软件完成这项工作。

另外, 在数据运行过程中, 大量数据不断插入、删除、修改, 时间一长, 会影响系统的性能。因此, DBA 要定期对数据库进行重新组织, 以提高系统的性能。当用户的需求增加和改变时, DBA 还要对数据库进行较大的改造, 包括修改部分设计, 即数据库的重新构造。

DBA 不仅要有较高的技术水平和较深的资历, 还应具有了解和阐明管理要求的能力。特别是对于大型数据库系统, DBA 极为重要。而常见的系统往往只有一个用户, 没有必要设置专职的 DBA。DBA 由应用程序员等兼任。

1.2 数据库技术的产生与发展

1.2.1 数据管理技术的发展

数据库技术是应数据管理任务的需求而产生的。数据管理技术是随着计算机技术的发展而完善的。数据管理是研究如何对数据分类、组织、编码、存储、检索和维护的一门技术, 是数据处理的中心问题。数据管理经历了人工管理、文件系统管理、数据库系统和高级数据库阶段。

1. 人工管理阶段

在人工管理阶段 (20 世纪 50 年代中期以前), 计算机主要用于科学计算, 对于数据保存的需求尚不迫切, 数据的管理是靠人工进行的, 计算机不保存数据, 也没有专用的数据管理软件, 只有程序的概念, 没有文件的概念, 一组数据对应一个应用程序, 如图 1.2 所示, 数据存在着大量的重复存储现象。

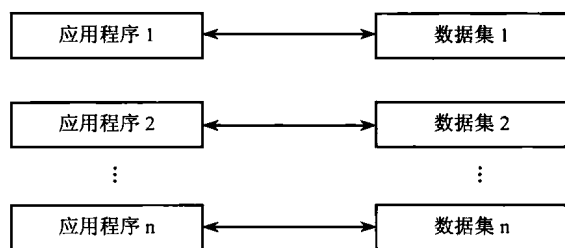


图 1.2 人工管理阶段程序与数据的关系

2. 文件系统阶段

在文件系统阶段（20 世纪 50 年代后期至 60 年代中期），计算机开始应用于信息管理。硬件方面出现了可以直接存取的外部存储设备，软件方面有了操作系统中专门管理数据的文件系统。数据的管理是以独立的文件形式存放，并可按记录存取。在文件系统阶段，一个应用程序可以处理多个数据文件，文件系统在程序与数据间起到接口的作用，使程序和数据有了一定的独立性，如图 1.3 所示。这使得程序员可以把精力集中于算法，不必过多考虑物理细节，因此在这个时期各种数据结构和算法得到了充分的发展，大大丰富了计算机科学。但文件系统的致命缺陷使各种数据文件之间缺乏有机的联系，数据和程序之间缺乏独立性，不能有效地共享相同的数据，从而造成了数据的冗余和不一致，给数据的修改和维护带来了困难。

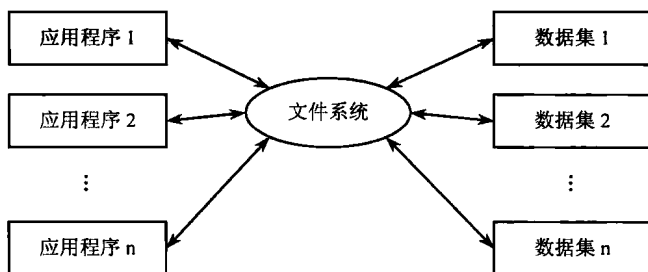


图 1.3 文件系统阶段程序与数据之间的关系

3. 数据库阶段

随着计算机技术的迅速发展和广泛应用，磁盘技术取得了重要进展，数据管理中数据量也急剧增长，对数据共享和数据管理就提出了更高的要求。此时文件系统已经不能满足应用的需求，数据库技术也就应运而生。

数据管理技术在 20 世纪 60 年代末的发展经历了以下几件大事：

(1) 1969 年 IBM 公司研制、开发了数据库管理系统商品化软件 IMS (Information Management System)，IMS 的数据模型是层次结构的。

(2) 美国数据库系统语言协会 CODASYL (Conference On Data System Language) 下属的数据库任务组 DBTG (Data Base Task Group) 对数据库方法进行了系统的讨论、研究，提出了若干报告，最终形成 OBTG 报告。OBTG 报告确定并且建立了数据库系统的许多概念、方法和技术。

(3) 1970 年 IBM 公司 SanJoe 研究实验室的研究员 E.F.Codd 发表了著名的“大型共享系统的关系数据库的关系模型”论文，为关系数据库技术奠定了理论基础。

这个阶段数据库与应用程序的关系可由图 1.4 表示。

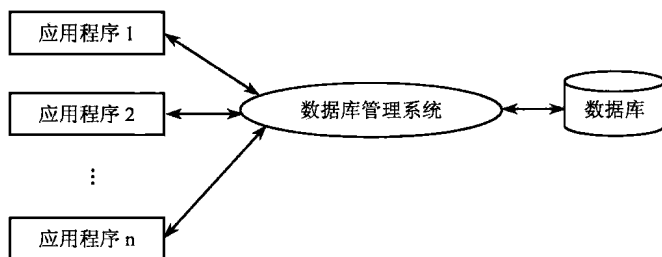


图 1.4 应用程序与数据库之间的关系

4. 高级数据库阶段

20 世纪 70 年代开始，数据库技术又有了很大的发展，表现如下：

(1) 数据库方法，特别是 OBTG 方法和思想应用于各种计算机系统，出现了许多商品化数据库系统，它们大都是基于网状模型和层次模型的。

(2) 商用数据库系统的运行使数据库技术日益广泛地应用到企业管理、事务处理、交通运输、信息检索、军事指挥、政府管理和辅助决策等各个方面，深入到生产、生活的各个领域。数据库技术成为实现和优化信息系统的基本技术。

(3) 关系方法的理论研究和软件系统的研制取得了很大的成果。

这一阶段的主要标志是 20 世纪 80 年代出现的分布式数据库系统、90 年代出现的面向对象数据库系统和各种新型数据库系统。

1) 分布式数据库系统。分布式数据库系统 (Distributed DataBase System, DDBS) 是在集中式数据库基础上发展起来的，是数据库技术与计算机网络技术、分布处理技术相结合的产物。分布式数据库系统的主要特点如下：

- 数据是分布的。
- 数据是逻辑相关的。
- 节点具有自治性。

2) 面向对象数据库系统。面向对象数据库系统 (Object-Oriented DataBase System, OODBS) 是将面向对象的模型、方法和机制与先进的数据库技术有机结合而形成的新型数据库系统。它从关系模型中脱离出来，强调在数据库框架中发展类型、数据抽象、继承和持久性；它的基本设计思想是：一方面把面向对象语言向数据库方向扩展，使应用程序能够存取并处理对象；另一方面扩展数据库系统，使其具有面向对象的特征，提供一种综合的语义数据建模概念集，以便对复杂应用中的实体和联系建模。

3) 多媒体数据库系统。多媒体数据库系统 (Multi-Media DataBase System, MDBS) 是将数据库技术与多媒体技术相结合的产物。从实际应用的角度考虑，多媒体数据库管理系统 (MDBMS) 应具有如下基本功能：

- 能够有效地表示多媒体数据，对不同媒体类型的数据，如文本、图形、图像、声音等能够按应用的不同采用不同的表示方法。
- 能够处理各种媒体数据，正确识别和表现各种媒体数据的特征、各种媒体间的空间或时间的关联。
- 能够像对其他格式化数据一样对多媒体数据进行操作。

4) 数据仓库。数据仓库可以提供对企业数据方便地访问和具有强大分析能力的工具，从企业数据中获得有价值的信息，发掘企业的竞争优势，提高企业的运营效率和指导企业决策。数据仓库

作为决策支持系统（Decision Support System, DSS）的有效解决方案，涉及三方面的技术内容：数据仓库技术、联机分析处理（On-Line Analysis Processing, OLAP）技术和数据挖掘（Data Mining, DM）技术。

各阶段有自己的背景及特点，数据管理技术也在发展中不断完善。随着社会信息化的不断发展，对数据管理的要求也会越来越高。计算机不断更新的软件、硬件技术是这门学科发展的基础，数据库技术的发展是推动这门学科发展的动力。

人工管理阶段、文件系统阶段、数据库系统阶段的比较如表 1.1 所示。

表 1.1 数据管理 3 个阶段的比较

数据管理的 3 个阶段		人工管理阶段	文件系统阶段	数据库系统阶段
比较项目	应用背景	科学计算	科学计算、管理	大规模管理
	硬件背景	无直接存取存储设备	磁盘、磁鼓	大容量磁盘
	软件背景	没有操作系统	有文件系统	有数据库管理系统
特点	处理方式	批处理	联机实时处理、批处理	联机实时处理、分布处理、批处理
	数据的管理者	人	文件系统	数据库管理系统
	数据面向的对象	某一应用程序	某一应用程序	现实世界
	数据的共享程度	无共享、冗余度极大	共享性差、冗余度大	共享性高、冗余度小
	数据的独立性	不独立，完全依赖于应用程序	独立性差	具有高度的物理独立性和一定的逻辑独立性
	数据的结构化	无结构	记录内有结构，整体无结构	整体结构化，用数据模型描述
	数据的控制能力	应用程序自己控制	应用程序自己控制	由数据库管理系统提供数据安全性、完整性、并发控制和恢复能力

1.2.2 数据库系统的特点

与传统的文件系统相比，数据库系统呈现出许多潜在的优点，主要表现在如下几个方面：

（1）程序——数据独立性。

在数据库系统中，数据描述（元数据）从使用这些数据的应用程序中分离出来，这种分离称为数据独立性。数据独立性一般分为数据的逻辑独立性和数据的物理独立性。数据逻辑独立性是指数据库总体逻辑结构的改变，如修改数据定义、增加新的数据类型、数据间联系的改变等不需要修改应用程序。数据物理独立性是指用户的应用程序与存储在磁盘上的数据库中的数据是相互独立的。也就是说，数据在磁盘上的数据库中怎样存储是由 DBMS 管理的，用户程序不需要去了解。当数据的物理结构（存取结构、存取方式）发生改变时，不影响数据的逻辑结构，因而应用程序并不需要改变。

数据独立的特性使得数据的定义可以从程序中分离出来，同时数据的存取由数据库管理系统 DBMS 负责完成，从而很大程度上减少了程序的编制和维护。

（2）提高了数据的共享性。

数据共享是指数据库中的一组数据集合为多个用户、多个应用共享使用。数据共享可以减少数据冗余，节约存储空间，同时还能够避免数据的不相容性和不一致性。数据共享是促使数据库技术