



21st CENTURY
规划教材

面向21世纪高等院校计算机系列规划教材

COMPUTER COURSES FOR UNDERGRADUATE EDUCATION

数据库原理与应用

陆慧娟 主编



科学出版社

www.sciencep.com



面向21世纪高等院校计算机系列规划教材
COMPUTER COURSES FOR UNDERGRADUATE EDUCATION

数据库原理与应用

陆慧娟 主编

吴达胜 刘建平 黄长城 副主编

科学出版社

北京

内 容 简 介

本书以关系数据库系统为核心,全面介绍了数据库系统的基本原理、技术实现及其应用。全书共12章,内容包括数据库系统概论、关系模型、结构化查询语言、关系数据库设计理论、索引和哈希技术、数据库文件存储技术、数据库安全性与完整性、数据库设计、数据库设计的典型案例、SQL Server 数据库基础、数据库应用系统开发技术,最后介绍了几种新型的数据库技术。本书每章后均附有习题。书中所涉及例子均以学生学习过程为主线,具有系统性和一致性,其程序代码已经过严格调试。

本书可作为普通高等院校计算机及相关专业的数据库课程教材,也可作为自学计算机技术的参考书。

图书在版编目(CIP)数据

数据库原理与应用/陆慧娟主编.—北京:科学出版社,2006

(面向21世纪高等院校计算机系列规划教材)

ISBN 7-03-016734-1

I.数… II.陆… 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159600号

责任编辑:吕建忠 陈砾川 / 责任校对:刘彦妮

责任印制:吕春珉 / 封面设计:飞天创意

科学出版社出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

北京彩色印务有限公司印刷

科学出版社发行 各地新华书店经销

*

2006年1月第 一 版 开本:787×1092 1/16

2006年1月第一次印刷 印张:25

印数:1—3 000 字数:574 000

定价:33.00元

(如有印装质量问题,我社负责调换<路通>)

销售部电话 010-62136131 编辑部电话 010-62138978-8001 (H101)

前 言

近年来,数据库技术发展迅速,已广泛地应用于我们生产与生活的各个领域,成为计算机科学的重要分支。可以说,计算机应用几乎都与数据库有关。

本书全面介绍了数据库系统的基本原理、技术实现和基本应用,编写中力求内容全面、概念清晰、语言流畅、图文并茂、理论与实际相结合,相关章节还结合 SQL Server 2000 数据库系统进行介绍,并尽力反映数据库领域的最新研究成果,比如 ADO、ADO.NET、Visual Studio.NET 等技术。

本教材的教学大约需要 51~68 学时,其中理论讲授安排 39~57 学时,实验安排 12~18 学时。第 1~5 章及第 7~9 章为必修内容;第 6 章、第 12 章为选修内容。

本书可作为普通高等院校、高等职业技术学校、成人高等院校计算机类及相关专业的教材,也可作为非计算机专业学生的选修课或辅修课教材,还可供计算机应用人员及工程技术人员参考。

本书第 1 及第 5~7 章由陆慧娟编写,第 2 章由张建美老师编写,第 3、4 章由刘建平老师编写,第 8 章由李慧老师和黄长城老师合写,第 9 章和第 11 章由吴达胜老师编写,第 10 章由黄长城老师编写,第 12 章由关伟老师编写。全书由陆慧娟统稿。

在本书的编写过程中,几位编者得到了所在学校的领导和同事的支持,特别是得到了中国计量学院计算机系师生的帮助,在此对他们表示衷心的感谢。另外,编者参考了国内外同行的研究成果和相关资料。在此,谨向本书参考文献中列出的作者表示感谢!

由于时间仓促,书中如有不当之处,恳请读者批评指正,并欢迎读者将意见发至电子邮件地址 hjlu8@cjlue.edu.cn 反馈给编者。

陆慧娟

2005 年 12 月于中国计量学院

目 录

第 1 章 数据库系统概论	1
1.1 引言	2
1.1.1 信息与数据	2
1.1.2 数据处理	2
1.2 数据管理技术的演变	3
1.2.1 人工管理阶段	3
1.2.2 文件系统阶段	3
1.2.3 数据库阶段	4
1.2.4 高级数据库阶段	4
1.3 数据库系统的组成与结构	6
1.3.1 数据库系统的组成	6
1.3.2 数据库系统的结构	8
1.4 数据库管理系统	12
1.4.1 DBMS 的主要功能	12
1.4.2 DBMS 的组成	13
1.4.3 DBMS 的数据存取过程	15
1.5 数据模型	15
1.5.1 数据模型的组成要素	15
1.5.2 数据之间的联系	16
1.5.3 数据模型的分类	18
1.6 数据库系统的发展	22
小结	24
习题	24
第 2 章 关系模型	26
2.1 关系模型	27
2.1.1 关系数据结构	27
2.1.2 关系操作集合	27
2.1.3 关系的三类完整性约束	28
2.2 关系的形式化定义	28
2.2.1 域	28
2.2.2 笛卡儿积	28
2.2.3 关系	29
2.3 关系的性质	30
2.4 关系的码	31
2.4.1 候选码与主码	31

2.4.2 主属性与非码属性	32
2.4.3 外部键码	32
2.4.4 关系模型的完整性	32
2.5 关系模式与关系数据库	34
2.5.1 关系模式	34
2.5.2 关系数据库	35
2.6 关系代数	36
2.6.1 关系代数的分类及其运算符	36
2.6.2 传统的集合运算	36
2.6.3 专门的关系运算	38
2.7 关系演算	41
2.7.1 元组关系演算语言	42
2.7.2 域关系演算语言	44
2.8 关系代数表达式的优化	45
2.8.1 关系代数表达式的等价变换规则	45
2.8.2 关系代数表达式的优化算法	47
2.8.3 关系代数表达式的优化策略	47
小结	48
习题	48
第3章 结构化查询语言 SQL	53
3.1 SQL 的基本概念	54
3.1.1 基本操作模式、数据类型和空值	55
3.1.2 表达式与函数	59
3.1.3 SQL 命令和保留字	67
3.2 SQL 的数据定义功能	70
3.2.1 基本表的定义、修改和删除	70
3.2.2 视图	71
3.2.3 索引	73
3.3 SQL 的数据操纵功能	73
3.3.1 SQL 查询语句	73
3.3.2 SQL 更新语句	86
3.3.3 视图的查询、更新	89
3.3.4 关系操作符	91
3.4 SQL 的数据控制功能	97
小结	100
习题	100
第4章 关系数据库理论	106
4.1 问题的提出	107
4.1.1 冗余量大	108

4.1.2 插入异常	108
4.1.3 删除异常	108
4.1.4 潜在的不一致性	108
4.2 函数依赖性	109
4.2.1 函数依赖	109
4.2.2 函数依赖的蕴涵性	110
4.2.3 关键字	111
4.3 关于函数依赖性的公理系统	111
4.3.1 阿姆斯特朗公理	111
4.3.2 阿姆斯特朗公理的完备性	112
4.3.3 闭包的计算	114
4.3.4 函数依赖集的等价、覆盖和最小集	115
4.4 关系模式的分解	117
4.4.1 无损连接分解	117
4.4.2 保持依赖的分解	121
4.5 关系模式的规范化	122
4.5.1 第一范式和第二范式	122
4.5.2 第三范式	123
4.5.3 BCNF 范式	125
4.5.4 BCNF 范式的无损连接分解	126
4.5.5 第三范式的保持依赖分解	127
4.5.6 第三范式的无损连接且保持依赖的分解	129
4.5.7 多值依赖和 4NF	130
小结	132
习题	132
第 5 章 数据库文件存储技术	135
5.1 文件组织	136
5.1.1 文件组织方式	136
5.1.2 顺序文件组织	140
5.1.3 聚集文件组织	142
5.2 索引技术	143
5.2.1 基本概念	143
5.2.2 线性索引	144
5.2.3 B 树索引文件	144
5.2.4 B ⁺ 树索引文件	146
5.3 哈希技术	151
5.3.1 基本概念	151
5.3.2 哈希索引	155

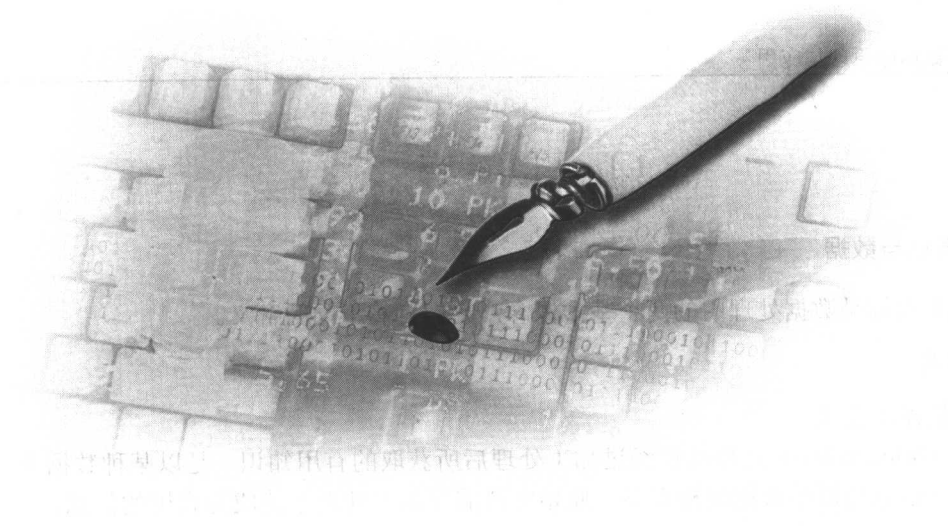
小结	156
习题	156
第6章 安全性与完整性	158
6.1 数据库的安全性	159
6.1.1 数据库安全性的含义	159
6.1.2 安全性控制的一般方法	159
6.2 完整性控制	163
6.2.1 数据库完整性的含义	163
6.2.2 完整性规则的组成	163
6.2.3 完整性约束条件的分类	165
6.3 并发控制与封锁	166
6.3.1 数据库并发性的含义	166
6.3.2 事务	166
6.3.3 并发控制的数据不一致性与加锁	168
6.4 数据库的恢复	174
6.4.1 数据库恢复的含义	174
6.4.2 恢复的原理及其实现技术	174
6.4.3 数据库的故障和恢复策略	177
小结	179
习题	180
第7章 数据库设计	182
7.1 数据库设计概述	183
7.1.1 数据库设计的任务、内容和特点	183
7.1.2 数据库设计方法简述	184
7.1.3 数据库设计的步骤	186
7.2 系统需求分析	188
7.2.1 系统需求分析的任务	189
7.2.2 系统需求分析的方法	190
7.3 概念结构设计	193
7.3.1 概念模型的特点	193
7.3.2 概念结构设计的方法与步骤	193
7.3.3 数据抽象与局部 E-R 模型设计	194
7.3.4 视图的集成	197
7.4 逻辑结构设计	201
7.4.1 逻辑结构设计任务和步骤	201
7.4.2 E-R 图向数据模型的转换	201
7.4.3 数据模型的优化	202
7.4.4 设计用户子模式	204

7.5 数据库物理设计	205
7.5.1 选择存取方法	205
7.5.2 确定数据库的存储结构	206
7.5.3 评价物理结构	206
7.6 数据库实施	207
7.6.1 建立实际数据库结构	207
7.6.2 装入数据	207
7.6.3 应用程序编码与调试	207
7.6.4 数据库试运行	208
7.6.5 整理文档	208
7.7 数据库运行和维护	208
7.7.1 维护数据库的安全性及完整性	209
7.7.2 监测并改善数据库性能	209
7.7.3 重新组织和构造数据库	209
小结	210
习题	210
第8章 数据库设计的典型案例	213
8.1 案例的系统需求简介	214
8.1.1 总体需求描述	214
8.1.2 用户总体业务构造	214
8.1.3 其他要求	214
8.1.4 系统功能设想	214
8.1.5 业务流程分析	215
8.2 需求描述	216
8.2.1 数据流图	216
8.2.2 数据字典	219
8.3 概念设计	220
8.3.1 实体	221
8.3.2 系统局部 E-R 图	221
8.3.3 系统全局 E-R 图	223
8.4 逻辑设计	224
8.4.1 E-R 图到关系模式的转换	224
8.4.2 关系模式的规范及调整	224
8.4.3 各个数据表的表结构设计	225
8.5 数据库的物理设计	227
8.5.1 存储介质的选择	227
8.5.2 数据库“学生选课”的存储结构设计	228
8.5.3 各个数据表(视图)的建立	229

8.5.4 选择存取方法.....	230
8.5.5 数据库服务器性能优化.....	231
小结.....	231
习题.....	232
第9章 SQL Server 数据库基础	233
9.1 客户/服务器体系结构.....	234
9.2 SQL Server 2000 应用环境概述.....	234
9.2.1 SQL Server 2000 软件架构.....	235
9.2.2 SQL Server 数据库组成.....	236
9.2.3 SQL Server 工具.....	243
9.3 Transact-SQL 程序设计.....	246
9.3.1 Transact-SQL 的数据定义功能及实例.....	247
9.3.2 Transact-SQL 的数据操纵功能及实例.....	249
9.3.3 注释符与运算符.....	253
9.3.4 变量.....	254
9.3.5 流程控制语句.....	254
9.3.6 常用命令.....	258
9.3.7 函数.....	260
9.4 存储过程.....	260
9.4.1 存储过程的类型.....	260
9.4.2 创建存储过程.....	261
9.4.3 执行存储过程.....	262
9.4.4 存储过程实例.....	263
9.5 触发器.....	265
9.5.1 触发器结构.....	265
9.5.2 创建触发器.....	266
9.5.3 触发器实例.....	269
小结.....	271
习题.....	271
第10章 数据库应用系统开发技术	273
10.1 概述.....	274
10.2 C/S 体系结构.....	275
10.2.1 传统的二层 C/S 结构.....	276
10.2.2 二层 C/S 结构的问题.....	276
10.2.3 “瘦客户机/胖服务器”模式.....	277
10.2.4 三层及多层的 C/S 结构.....	278
10.2.5 N 层体系结构.....	279
10.3 B/S 结构.....	280
10.4 开放式数据库开发.....	281

10.4.1 ODBC.....	282
10.4.2 基于 ODBC 接口的编程.....	284
10.4.3 数据库操作对象 ADO	286
10.5 ADO.NET 技术	290
10.5.1 ADO.NET 的设计目标	291
10.5.2 ADO.NET 模型构造	291
10.5.3 .NET Framework 数据提供程序.....	292
10.5.4 DataSet.....	297
10.5.5 基于 Visual Studio.NET 环境应用程序示例.....	306
小结.....	314
习题.....	314
第 11 章 Web 数据库及其开发技术	316
11.1 Web 数据库概述.....	317
11.2 ASP.NET 概述.....	317
11.2.1 ASP.NET 初识	317
11.2.2 ASP.NET 的编程模型	318
11.2.3 安装和配置 IIS 5.0	319
11.2.4 安装.NET Framework	319
11.2.5 选择开发环境.....	319
11.3 Web 数据库应用实例	320
11.3.1 DataGrid 控件介绍	320
11.3.2 DataGrid 控件实现数据库记录的分页显示	323
11.3.3 DataGrid 控件对记录排序	326
11.3.4 DataGrid 控件定制列元素	328
小结.....	340
习题.....	340
第 12 章 数据库新技术	341
12.1 概述.....	342
12.2 面向对象的数据库技术.....	342
12.2.1 面向对象的模型的核心概念	342
12.2.2 面向对象的数据库系统的特点	344
12.2.3 面向对象的数据库语言	345
12.2.4 面向对象的数据库研究内容	350
12.3 分布式数据库技术	352
12.3.1 分布式数据库系统的概念	352
12.3.2 分布式数据库系统的体系结构	353
12.3.3 分布式数据库系统的设计	354
12.3.4 分布式查询.....	357

12.3.5 分布式并发控制	359
12.4 数据仓库技术	360
12.4.1 数据仓库的概念	360
12.4.2 数据仓库系统的体系结构	362
12.4.3 数据仓库中的数据	367
12.4.4 数据仓库的数据模型	367
12.4.5 数据仓库的设计和开发	368
12.4.6 数据仓库的解决方案及工具介绍	374
12.4.7 数据仓库的主要技术	377
12.5 其他新型的数据库系统	379
12.5.1 知识库	379
12.5.2 模糊数据库与演绎数据库	380
12.5.3 主动数据库	380
12.5.4 移动数据库	384
12.5.5 统计数据库	385
小结	386
习题	386
参考文献	387



第 1 章 数据库系统概论



本章要点

- ◇ 信息、数据、数据处理、数据库、数据库系统与数据库管理系统
- ◇ 数据库技术的产生、发展及其未来
- ◇ 数据库系统的组成与结构
- ◇ 数据库管理系统
- ◇ 数据模型
- ◇ 数据模型与数据库系统的发展



本章学习目标

- ◇ 理解信息、数据和数据处理的基本概念，区分数据、数据库、数据库系统和数据库管理系统
- ◇ 了解数据库的四个发展阶段、各阶段的主要特点
- ◇ 掌握数据库系统的组成与结构
- ◇ 掌握数据库管理系统的主要功能、组成及数据存取过程
- ◇ 了解几种主要数据模型的特点

1.1 引言

1.1.1 信息与数据

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. 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。

(2) 信息的特征

信息有以下三个重要特征。

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。

2. 数据

(1) 数据的定义

数据(data)是用来记录信息的可识别的符号。也有人把数据定义为关于人、地点、行为或事件的事实。

(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。

3. 数据与信息的联系

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。

1.1.2 数据处理

数据处理又称信息处理，是将数据转换成信息的过程，包括对数据的收集、存储、

加工、检索和传输等一系列活动，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。

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。

1.2 数据管理技术的演变

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、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以下四个阶段：人工管理阶段、文件系统阶段、数据库阶段和高级数据库阶段。

1.2.1 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 1.1 所示，数据存在着大量的重复存储现象。

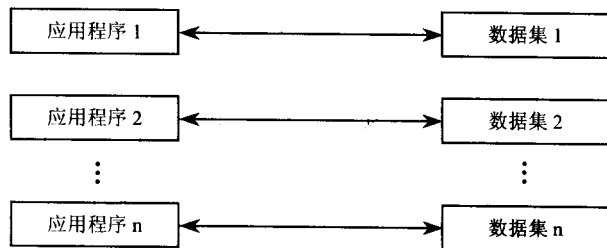


图 1.1 人工管理阶段程序与数据的关系

1.2.2 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数据文件形式存放，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 1.2 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

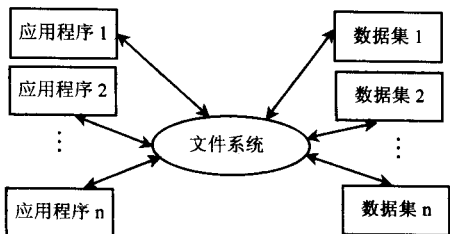


图 1.2 文件系统阶段程序与数据之间的关系

这个时期各种数据结构和算法得到了充分发展，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。

1.2.3 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地讨论、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 San Jose 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 1.3 表示。

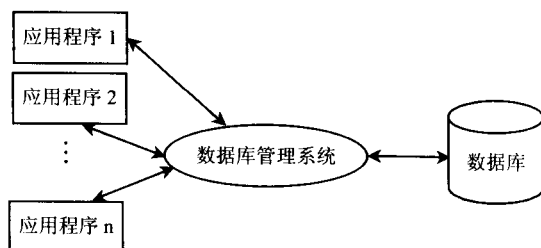


图 1.3 应用程序与数据库之间的关系

1.2.4 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。

- 1) 数据是分布的。
- 2) 数据是逻辑相关的。
- 3) 结点具有自治性。

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地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的发展类型、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) 是数据库技术与多媒体技术相结合的产物。其主要特点是:

- 1) 数据量大。
- 2) 结构复杂。
- 3) 时序性。
- 4) 数据传输的连续性。

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:

- 1) 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。
- 2) 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。
- 3) 能够像对其他格式化数据一样对多媒体数据进行操作。
- 4) 具有开放功能, 提供多媒体数据库的应用程序接口。

4. 数据仓库

数据仓库可以提供对企业数据方便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。

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。