

可下载教学资料

<http://www.tup.tsinghua.edu.cn>



高等学校教材
信息管理与信息系统

数据库系统 应用教程

王成 杨铭 王世波 等 编著

清华大学出版社



高等学校教材
信息管理与信息系统

数据库系统 应用教程

王成 杨铭 王世波 等 编著

清华大学出版社
北京

内 容 简 介

本书系统阐述了数据库的基础理论、基本技术和方法。全书共 7 章, 主要内容包括数据库系统概述、关系数据库基本理论、关系数据库标准语言 SQL、关系数据库规范化理论、数据库设计、数据保护以及数据库技术发展。本书结构完整、内容精练、实用性强。本书在阐述数据库基本理论的同时, 围绕基本理论介绍了 SQL Server 2000 的相关知识。此外, 本书还提供配套的实验教程, 以方便实验课程的开展。

本书可作为高等学校非计算机专业数据库课程的教材, 也可作为从事数据库系统研究和开发人员的参考书。

本书封面贴有清华大学出版社防伪标签, 无标签者不得销售。

版权所有, 侵权必究。侵权举报电话: 010-62782989 13701121933

图书在版编目(CIP)数据

数据库系统应用教程 / 王成等编著. —北京: 清华大学出版社, 2008.8
(高等学校教材·信息管理与信息系统)

ISBN 978-7-302-17972-6

I. 数… II. 王… 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 (2008) 第 093669 号

责任编辑: 郑寅堃 李 晔

责任校对: 白 蕾

责任印制: 杨 艳

出版发行: 清华大学出版社

地 址: 北京清华大学学研大厦 A 座

<http://www.tup.com.cn>

邮 编: 100084

社 总 机: 010-62770175

邮 购: 010-62786544

投稿与读者服务: 010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者: 山东新华印刷厂临沂厂

经 销: 全国新华书店

开 本: 185×260 印 张: 15.5 字 数: 372 千字

版 次: 2008 年 8 月第 1 版 印 次: 2008 年 8 月第 1 次印刷

印 数: 1~5000

定 价: 23.00 元

本书如存在文字不清、漏印、缺页、倒页、脱页等印装质量问题, 请与清华大学出版社出版部联系调换。联系电话: (010)62770177 转 3103 产品编号: 026989-01

编审委员会成员

(按地区排序)

清华大学	周立柱	教授
	覃 征	教授
	王建民	教授
	刘 强	副教授
	冯建华	副教授
北京大学	杨冬青	教授
	陈 钟	教授
	陈立军	副教授
北京航空航天大学	马殿富	教授
	吴超英	副教授
	姚淑珍	教授
中国人民大学	王 珊	教授
	孟小峰	教授
	陈 红	教授
北京师范大学	周明全	教授
北京交通大学	阮秋琦	教授
北京信息工程学院	孟庆昌	教授
北京科技大学	杨炳儒	教授
石油大学	陈 明	教授
天津大学	艾德才	教授
复旦大学	吴立德	教授
	吴百锋	教授
	杨卫东	副教授
华东理工大学	邵志清	教授
华东师范大学	杨宗源	教授
	应吉康	教授
东华大学	乐嘉锦	教授
上海第二工业大学	蒋川群	教授
浙江大学	吴朝晖	教授
	李善平	教授
南京大学	骆 斌	教授
南京航空航天大学	秦小麟	教授
南京理工大学	张功萱	教授

南京邮电学院	朱秀昌	教授
苏州大学	龚声蓉	教授
江苏大学	宋余庆	教授
武汉大学	何炎祥	教授
华中科技大学	刘乐善	教授
中南财经政法大学	刘腾红	教授
华中师范大学	王林平	副教授
	魏开平	副教授
	叶俊民	教授
国防科技大学	赵克佳	教授
	肖 依	副教授
中南大学	陈松乔	教授
	刘卫国	教授
湖南大学	林亚平	教授
	邹北骥	教授
西安交通大学	沈钧毅	教授
	齐 勇	教授
长安大学	巨永峰	教授
西安石油学院	方 明	教授
西安邮电学院	陈莉君	教授
哈尔滨工业大学	郭茂祖	教授
吉林大学	徐一平	教授
	毕 强	教授
长春工程学院	沙胜贤	教授
山东大学	孟祥旭	教授
	郝兴伟	教授
山东科技大学	郑永果	教授
中山大学	潘小轰	教授
厦门大学	冯少荣	教授
福州大学	林世平	副教授
云南大学	刘惟一	教授
重庆邮电学院	王国胤	教授
西南交通大学	杨 燕	副教授

改革开放以来，特别是党的十五大以来，我国教育事业取得了举世瞩目的辉煌成就，高等教育实现了历史性的跨越，已由精英教育阶段进入国际公认的大众化教育阶段。在质量不断提高的基础上，高等教育规模取得如此快速的发展，创造了世界教育发展史上的奇迹。当前，教育工作既面临着千载难逢的良好机遇，同时也面临着前所未有的严峻挑战。社会不断增长的高等教育需求同教育供给特别是优质教育供给不足的矛盾，是现阶段教育发展面临的基本矛盾。

教育部一直十分重视高等教育质量工作。2001年8月，教育部下发了《关于加强高等学校本科教学工作，提高教学质量的若干意见》，提出了十二条加强本科教学工作提高教学质量的措施和意见。2003年6月和2004年2月，教育部分别下发了《关于启动高等学校教学质量与教学改革工程精品课程建设工作的通知》和《教育部实施精品课程建设提高高校教学质量和人才培养质量》文件，指出“高等学校教学质量和教学改革工程”是教育部正在制定的《2003—2007年教育振兴行动计划》的重要组成部分，精品课程建设是“质量工程”的重要内容之一。教育部计划用五年时间（2003—2007年）建设1500门国家级精品课程，利用现代化的教育信息技术手段将精品课程的相关内容上网并免费开放，以实现优质教学资源共享，提高高等学校教学质量和人才培养质量。

为了深入贯彻落实教育部《关于加强高等学校本科教学工作，提高教学质量的若干意见》精神，紧密配合教育部已经启动的“高等学校教学质量与教学改革工程精品课程建设工作”，在有关专家、教授的倡议和有关部门的大力支持下，我们组织并成立了“清华大学出版社教材编审委员会”（以下简称“编委会”），旨在配合教育部制定精品课程教材的出版规划，讨论并实施精品课程教材的编写与出版工作。“编委会”成员皆来自全国各类高等学校教学与科研第一线的骨干教师，其中许多教师为各校相关院系主管教学的院长或系主任。

按照教育部的要求，“编委会”一致认为，精品课程的建设工作从开始就要坚持高标准、严要求，处于一个比较高的起点上；精品课程教材应该能够反映各高校教学改革与课程建设的需要，要有特色风格、有创新性（新体系、新内容、新手段、新思路，教材的内容体系有较高的科学创新、技术创新和理念创新的含量）、先进性（对原有的学科体系有实质性的改革和发展、顺应并符合新世纪教学发展的规律、代表并引领课程发展的趋势和方向）、示范性（教材所体现的课程体系具有较广泛的辐射性和示范性）

和一定的前瞻性。教材由个人申报或各校推荐（通过所在高校的“编委会”成员推荐），经“编委会”认真评审，最后由清华大学出版社审定出版。

目前，针对计算机类和电子信息类相关专业成立了两个“编委会”，即“清华大学出版社计算机教材编审委员会”和“清华大学出版社电子信息教材编审委员会”。首批推出的特色精品教材包括：

（1）高等学校教材·计算机应用——高等学校各类专业，特别是非计算机专业的计算机应用类教材。

（2）高等学校教材·计算机科学与技术——高等学校计算机相关专业的教材。

（3）高等学校教材·电子信息——高等学校电子信息相关专业的教材。

（4）高等学校教材·软件工程——高等学校软件工程相关专业的教材。

（5）高等学校教材·信息管理与信息系统。

（6）高等学校教材·财经管理与计算机应用。

清华大学出版社经过 20 多年的努力，在教材尤其是计算机和电子信息类专业教材出版方面树立了权威品牌，为我国的高等教育事业做出了重要贡献。清华版教材形成了技术准确、内容严谨的独特风格，这种风格将延续并反映在特色精品教材的建设中。

清华大学出版社教材编审委员会
E-mail: dingl@tup.tsinghua.edu.cn

数据库技术是计算机科学的重要分支,也是计算机领域中应用最广泛、发展最迅速的技术之一。当今,信息资源已成为社会的重要财富和资源。建立一个行之有效的信息系统已成为企业或组织生存和发展的重要条件。作为信息系统核心和基础的数据库技术由此得到了越来越广泛的应用,从小型事务处理系统到大型信息系统,从联机事务处理到联机分析处理,从传统的数据管理到空间数据库、工程数据库等特定应用领域等,数据库的应用几乎遍及社会的各个领域。对于一个国家来说,数据库的建设规模、数据库信息量的大小和使用频度已成为衡量这个国家信息化程度的重要标志。

目前很多高等院校都开设了数据库课程,并将其作为一门基础必修课。了解和掌握有关数据库的基础知识并具备一定的实践能力,已经不仅仅是针对计算机专业学生所提出的要求。本书主要是为高等院校非计算机专业学生学习数据库课程而编写的,是在作者多年的数据库课程教学 and 实际数据库系统开发工作基础之上完成的,它简洁而又精练地介绍了数据库的基础理论知识,同时围绕基本理论,介绍了 SQL Server 2000 的相关知识。为了配合数据库课程的实验教学,同时编写了相应的数据库实验教程。

本书的内容有两条主线:一条主线是数据库的基础理论知识,如关系数据库理论、关系规范化理论、数据库设计理论等;另一条主线是数据库实际应用产品,本书重点介绍了 SQL Server 2000,这部分内容附在相关理论之后,主要讲述在 SQL Server 2000 中如何实现上述基本理论以及相关的基本操作,这样可以帮助学生理论联系实践,便于消化理解基本理论。这两条主线相辅相成、相互渗透,方便学生学习。

本书包括 7 章,主要内容如下:

第 1 章主要介绍数据库的基础知识,包括数据管理技术的产生和发展、数据库的基本概念、数据模型的分类以及 SQL Server 2000 概述。

第 2 章主要介绍关系数据库的基本理论,包括关系的数据结构、关系完整性和关系操作的概念,在关系操作中主要介绍关系代数。

第 3 章主要介绍关系数据库标准语言 SQL,包括 SQL 的基本概念、SQL 数据定义、数据查询、数据更新、视图和数据控制等命令,最后重点介绍 SQL Server 2000 中 T-SQL 语言。

第 4 章主要介绍关系规范化理论,包括数据依赖、范式、关系模式规范化以及函数依赖公理。

第 5 章主要介绍数据库设计理论,包括数据库设计的原则和方法、数据库设计的步骤,以及如何利用 PowerDesigner 进行数据建模。

第 6 章主要介绍数据保护,包括数据的安全性、完整性、并发控制、数据恢复以及数据库复制与数据库镜像。

第 7 章主要介绍数据库技术的发展,包括数据库技术发展的 3 个阶段和数据库的新技术。

本书由王成老师主编,杨铭老师、王世波老师任副主编。其中王成老师编写了第 1 章和第 7 章,杨铭老师编写了第 2 章和第 3 章,吴占坤老师编写了第 4 章,王世波老师编写了第 5 章,王铁老师编写了第 6 章,最后由王成老师进行了统稿。

由于时间比较仓促,加之作者水平有限,如有不当之处,恳请广大读者批评指正。

编 者

2008 年 6 月

第 1 章 数据库系统概述	1
1.1 数据管理技术	1
1.1.1 数据管理技术的产生和发展	1
1.1.2 数据库系统的基本概念	4
1.1.3 数据库技术的发展及研究领域	5
1.2 数据模型	7
1.2.1 数据模型的组成要素	7
1.2.2 数据模型分类	8
1.2.3 概念数据模型及表示方法	8
1.2.4 主要的逻辑数据模型	12
1.3 数据库系统结构	16
1.3.1 数据库系统模式的概念	16
1.3.2 数据库系统的三级模式结构	16
1.3.3 数据库的二级映像功能与数据独立性	18
1.4 SQL Server 2000 概述	19
1.4.1 SQL Server 2000 的版本及特点	19
1.4.2 SQL Server 2000 运行环境	20
1.4.3 SQL Server 2000 服务器的启动与注册	22
1.4.4 SQL Server 2000 的数据库	26
1.5 小结	27
1.6 习题	27
第 2 章 关系数据库	28
2.1 关系模型概述	28
2.1.1 关系模型的数据结构	28
2.1.2 关系操作	29
2.1.3 完整性约束	29
2.2 关系数据结构	30

2.2.1	关系	30
2.2.2	关系模式	33
2.2.3	关系数据库	34
2.3	关系的完整性	34
2.3.1	实体完整性	34
2.3.2	参照完整性	35
2.3.3	用户定义的完整性	36
2.4	关系代数	36
2.4.1	传统的集合运算	37
2.4.2	专门的关系运算	40
2.5	小结	45
2.6	习题	45
第3章	关系数据库标准语言 SQL	46
3.1	SQL 概述	46
3.2	数据定义	49
3.2.1	基本表的定义、修改与删除	49
3.2.2	索引的建立与删除	52
3.3	数据查询	53
3.3.1	单表查询	54
3.3.2	连接查询	62
3.3.3	嵌套查询	65
3.3.4	集合查询	73
3.4	数据更新	74
3.4.1	插入数据	74
3.4.2	修改数据	75
3.4.3	删除数据	76
3.5	视图	77
3.5.1	定义视图	77
3.5.2	查询视图	79
3.5.3	更新视图	80
3.5.4	视图的作用	82
3.6	数据控制	83
3.6.1	授权	84
3.6.2	收回权限	85
3.7	Transact-SQL 语言	86
3.7.1	Transact-SQL 语言基础	86
3.7.2	Transact-SQL 数据查询	101
3.7.3	存储过程和触发器	107

3.8	小结	119
3.9	习题	119
第4章	关系数据库规范化理论	121
4.1	数据依赖	121
4.1.1	函数依赖	121
4.1.2	函数依赖对关系模式的影响	125
4.2	范式与关系模式规范化	126
4.2.1	第一范式	127
4.2.2	第二范式	129
4.2.3	第三范式	131
4.2.4	Boyce-Codd 范式	132
4.2.5	多值依赖与第四范式	132
4.3	关系模式规范化	134
4.3.1	关系模式分解的步骤	135
4.3.2	关系模式分解的等价标准	135
4.4	函数依赖公理	138
4.5	小结	141
4.6	习题	141
第5章	数据库设计	142
5.1	数据库设计概述	142
5.1.1	数据库设计的原则	142
5.1.2	数据库设计方法	143
5.2	数据库设计的步骤	144
5.2.1	需求分析阶段	145
5.2.2	概念结构设计阶段	150
5.2.3	逻辑结构设计阶段	153
5.2.4	物理结构设计阶段	156
5.2.5	数据库实施阶段	157
5.2.6	运行和维护阶段	161
5.3	PowerDesigner 数据建模	162
5.3.1	PowerDesigner 概述	162
5.3.2	PowerDesigner 功能介绍	162
5.3.3	PowerDesigner 数据建模实例	164
5.4	小结	171
5.5	习题	172
第6章	数据保护	173
6.1	安全性	173

6.1.1	数据库安全控制的一般方法	173
6.1.2	SQL Server 中的安全性控制	176
6.2	完整性	186
6.2.1	完整性约束条件	187
6.2.2	完整性控制	189
6.2.3	SQL Server 中的完整性控制	191
6.3	并发控制	197
6.3.1	并发控制概述	197
6.3.2	并发操作的调度	199
6.3.3	封锁	201
6.3.4	SQL Server 中的并发控制	202
6.4	数据恢复	203
6.4.1	故障的种类及恢复	203
6.4.2	恢复的实现技术	205
6.4.3	SQL Server 中的数据库备份与恢复	206
6.5	数据库复制与数据库镜像	210
6.5.1	数据库复制	210
6.5.2	数据库镜像	212
6.6	小结	213
6.7	习题	213
第 7 章	数据库技术发展	215
7.1	数据库技术发展的 3 个阶段	215
7.1.1	第一代数据库系统	215
7.1.2	第二代数据库系统	216
7.1.3	第三代数据库系统	217
7.2	数据库新技术	218
7.2.1	数据模型的发展	218
7.2.2	数据库技术与其他相关技术结合	219
7.2.3	面向应用领域的数据库新技术	226
7.3	小结	230
7.4	思考题	230
参考文献		231

数据库系统概述

日常工作生活中，在人们的周围会有各种各样的数据库系统在运行。当人们进行股票交易、银行取款、订购车票、查询资料等活动时都需要与数据库打交道。数据库系统已成为人们提高工作效率和管理水平的重要手段，也是企业提高竞争力的有力武器。

数据库是数据管理的最新技术，是计算机科学的重要分支。在信息时代，信息资源已经成为各行各业的重要财富和资源，针对各行业或组织设计的信息系统已经成为其发展的重要基础条件。数据库技术是信息系统的核心和基础，因而得到了快速发展和越来越广泛的应用。数据库技术主要研究如何科学地组织和存储数据、高效地获取和处理数据，可以为各种用户提供及时的、准确的、相关的信息，满足这些用户的各种不同的需要。本章主要介绍数据管理技术发展过程、数据库系统基本概念、数据模型以及关系型数据库管理系统 SQL Server 2000 的基本知识。

1.1 数据管理技术

1.1.1 数据管理技术的产生和发展

数据管理技术是应数据管理任务的需要而产生的，数据管理是指对数据进行收集、组织、编码、存储、检索和维护等活动。自计算机问世以来，数据管理技术经历了手工管理、文件系统、数据库系统 3 个阶段。

1. 手工管理阶段

20 世纪 50 年代中期以前，计算机主要用于科学计算。硬件存储设备主要有磁带、卡片机、纸带机等，还没有磁盘等直接存取的存储设备。软件也处于初级阶段，没有操作系统和管理数据的工具。数据处理方式是批处理。数据的组织和管理完全靠程序员手工完成，因此称为“手工管理阶段”。这个阶段数据的管理效率很低，其特点是：

- (1) 数据不保存。该时期的计算机主要用于科学计算，一般不需要将数据长期保存，只是在计算某一课题时将数据输入，用完后不保存原始数据，也不保存计算结果。
- (2) 应用程序管理数据。数据要由应用程序自己管理，没有相应的软件系统负责数据

的管理工作。所以，程序员不但要规定数据的逻辑结构，而且还要设计物理结构。

(3) 数据不共享。数据是面向应用的，一组数据只能对应一个程序。当多个应用程序涉及某些相同的数据时，必须各自定义，无法相互利用、参照，因此，程序与程序之间有大量的冗余数据。

(4) 数据不具有独立性。数据的逻辑结构或物理结构发生变化后，必须对应用程序做相应的修改，这更增加了程序员的负担。

手工管理阶段应用程序与数据之间的对应关系如图 1-1 所示。

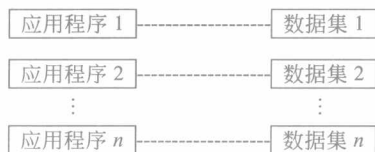


图 1-1 手工管理阶段

2. 文件系统阶段

20 世纪 50 年代后期到 60 年代中期，计算机得到广泛应用。这时在硬件方面已经有了磁盘、磁鼓等直接存取的存储设备；在软件方面，操作系统中已经有了专门的数据管理软件，一般称为文件系统；在处理方式上不但能进行批处理，而且能够实现联机实时处理。用文件管理数据具有如下特点：

(1) 数据可以长期保存。由于在进行数据处理时大量使用计算机，所以，数据需要长期保留在外存上，以供查询、更新等操作。

(2) 由文件系统管理数据。应用程序与数据库之间有了一定的独立性，程序员可以不必过多地考虑物理细节，将精力集中于算法。数据在存储上的改变不一定反映在程序上，节省了维护程序的工作量。

(3) 数据共享性差，冗余度大。在文件系统中，一个文件对应一个应用程序，文件是面向应用的。当不同的应用程序具有部分相同的数据时，也必须建立各自的文件，而不能共享相同的数据，因此数据冗余度大，浪费存储空间；同时可能造成数据的不一致性，给数据维护带来困难。

(4) 数据独立性差。文件系统上的文件是为某个特定应用服务的，文件的逻辑结构对该应用是最优的，因此，想对现有的数据增加一些新的应用是很困难的，系统扩充性不好。数据的逻辑结构要变化就必须修改应用程序。数据和应用程序之间缺乏独立性。

文件系统阶段应用程序与数据之间的对应关系如图 1-2 所示。

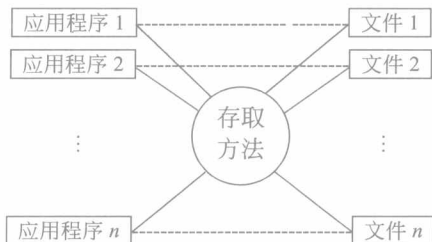


图 1-2 文件系统阶段

3. 数据库系统阶段

20 世纪 60 年代后期以来，计算机用于管理的规模更为庞大，应用越来越广泛，数据量急剧增长，同时多种应用、多种语言互相覆盖地共享数据集合的要求越来越强烈。硬件

已有了大容量的磁盘，硬件价格下降，软件价格上升；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。在这样的背景下，以文件系统作为数据管理手段已经不能满足应用的需求，为解决多用户、多应用共享数据的需求，出现了统一管理数据的专门软件系统——数据库管理系统。

从文件系统到数据库系统，是数据管理技术的一个飞跃。数据库系统的主要特点如下：

(1) 数据结构化。数据结构化是数据库与文件系统的根本区别。在文件系统阶段，只考虑了同一文件记录内部数据项之间的联系，而不同文件的记录之间是没有联系的，也就是说，从整体上看数据是无结构的。在数据库中，实现了整体数据的结构化，把文件系统简单的记录结构变成了记录和记录之间的联系所构成的结构化数据。在描述数据的时候，不仅要描述数据本身，还要描述数据之间的联系。

(2) 数据的共享性好，冗余度低。数据的共享程度直接关系到数据的冗余度。文件系统中，一个文件基本上对应一个应用程序，文件是面向应用的。不能共享相同的数据，因此冗余度大。数据库中的数据考虑所有用户的数据需求，是面向整个系统组织的，而不是面向某个具体应用的，减少了数据的冗余。

(3) 数据独立性好。数据独立性是指数据库中的数据与应用程序之间不存在依赖关系，而是相互独立的，包括数据的物理独立性和数据的逻辑独立性。物理独立性是指用户的应用程序与存储在硬盘上的数据库中数据是相互独立的；逻辑独立性是指用户的应用程序与数据库的逻辑结构是相互独立的，也就是说数据的逻辑结构改变了，用户程序可以不变。数据独立性是由数据库管理系统的二级映像功能保证的。

(4) 数据由数据库管理系统统一管理和控制。数据库的共享是并发的共享，即多个用户可以同时存取数据库中的数据甚至可以同时存取数据库中的同一个数据，这要求数据不仅要由数据库管理系统进行统一的管理，同时还要进行统一的控制。具体的控制功能包括：数据的安全性保护、数据的完整性保护、数据的并发控制、数据库的恢复。

数据库系统阶段应用程序与数据之间的关系如图 1-3 所示。

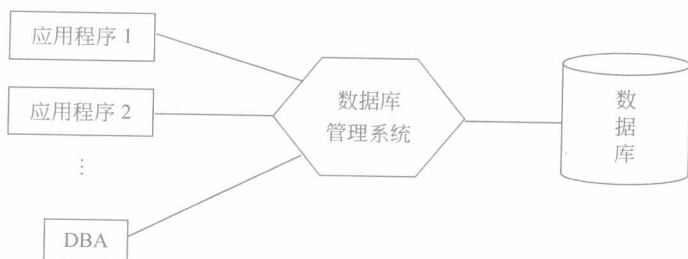


图 1-3 数据库系统阶段

数据管理技术 3 个阶段的比较如表 1-1 所示。

表 1-1 数据管理技术阶段比较

要素 \ 阶段	手工管理阶段	文件系统阶段	数据库系统阶段
时间	20 世纪 50 年代中期	20 世纪 50 年代后期～ 20 世纪 60 年代中期	20 世纪 60 年代后期至今

续表

要素\阶段	手工管理阶段	文件系统阶段	数据库系统阶段
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	有操作系统（文件系统）	有 DBMS
处理方式	批处理	批处理、联机实时处理	批处理、联机实时处理、分布处理
数据保存方式	数据不保存	以文件的形式长期保存，但无结构	以数据库形式保存，有结构
数据管理	考虑安排数据的物理存储位置	与数据文件名打交道	对所有数据实行统一、集中、独立的管理
数据与程序	数据面向程序	数据与程序脱离	数据与程序脱离 实现数据的共享
数据的管理者	人	文件系统	DBMS
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	共享性差	共享性高
数据的冗余度	冗余度极大	冗余度大	冗余度小
数据的独立性	不独立，完全依赖于程序	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 DBMS 提供数据的安全性、完整性、并发控制和恢复能力

1.1.2 数据库系统的基本概念

1. 数据（Data）

数据是数据库中存储的基本对象，它有多种表现形式，大多数人头脑中的第一反应是数据就是数字，其实数字只是最简单的一种数据，数据还包括文字、图形、图像、声音、语言等，这些表现形式可以经过数字化后存入计算机。

可以对数据定义如下：数据是指描述事物的符号记录，这些符号可以是文字、图形、声音、图像等。

数据的含义称为数据的语义，数据与其语义是不可分的。例如，学生档案表中有一个记录的描述如下：

（王一，男，1985-7-2，黑龙江，管理科学与工程系）

这个记录就是数据。对于了解其含义的人会得到这样的信息：姓名是王一，男，黑龙江人，1985 年 7 月 2 日出生，在管理科学与工程系读书；不了解其语义的人则无法理解其含义。可见，数据的形式还不能完全表达其内容，需要经过解释。所以，数据和关于数据的解释是不可分的。