



工业和信息化人才培养规划教材
Industry And Information Technology Training Planning Materials

Technical And Vocational Education

高职高专计算机系列

数据库系统 及应用

Database System and Its
Application

赵立江 王欣 © 主编

本书贴近实际需要，在内容取舍上既考虑数据库及其管理系统的基本概念、基本原理介绍的系统性，又兼顾到数据库应用系统分析和设计的案例教学的实用性。



人民邮电出版社
POSTS & TELECOM PRESS



工业和信息化人才培养规划教材
Industry And Information Technology Training Planning Materials

Technical And Vocational Education
高职高专计算机系列

数据库系统 及应用

Database System and Its
Application

赵立江 王欣 © 主编

本书贴近实际需要，在内容取舍上既考虑数据库及其管理系统的基本概念、基本原理介绍的系统性，又兼顾到数据库应用系统分析和设计的案例教学的实用性。



人民邮电出版社
POSTS & TELECOM PRESS



图书在版编目 (CIP) 数据

数据库系统及应用 / 赵立江, 王欣主编. — 北京 : 人民邮电出版社, 2013. 9

工业和信息化人才培养规划教材. 高职高专计算机系列

ISBN 978-7-115-32199-2

I. ①数… II. ①赵… ②王… III. ①数据库系统—
高等职业教育—教材 IV. ①TP311.13

中国版本图书馆CIP数据核字(2013)第175968号

内 容 提 要

全书系统地讲述数据库系统基本概念、关系模型、SQL 语言、数据库设计与规范化理论、数据库控制以及数据库新技术等内容。每章后均附有习题, 全书后还附有实验指导。书中所有的案例、习题、实验都提供完整的、已上机验证过的程序文档供下载, 同时为教师提供 PPT 电子课件。

本书语言简洁、概念清楚、重点突出, 适合作为本科院校、高职高专院校、成人教育学院“数据库原理与应用”课程的教材, 也适合作为数据库原理自学教材和参考书。对于非计算机专业的学生, 如果希望学到关键、实用的数据库技术, 也可采用本书作为教材。

-
- ◆ 主 编 赵立江 王 欣
责任编辑 王 平
责任印制 杨林杰
 - ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
邮编 100061 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京中新伟业印刷有限公司印刷
 - ◆ 开本: 787×1092 1/16
印张: 13.5 2013 年 9 月第 1 版
字数: 329 千字 2013 年 9 月北京第 1 次印刷
-

定价: 29.80 元

读者服务热线: (010)67170985 印装质量热线: (010)67129223

反盗版热线: (010)67171154

前 言

数据库技术是 20 世纪 60 年代后期发展起来的一门计算机数据管理技术，伴随着数据管理技术的不断发展，计算机的应用领域已渗透到人类社会生活的各个方面，数据库技术也因此成为计算机软件领域的一个重要分支，已经形成较为完整的理论体系。在专业教学中，数据库技术课程是计算机专业与信息管理等学科中一门重要的专业基础课。

作为计算机及相关专业的学生，数据库技术是一门入门课程，由于本书的两位作者均具有多年的各类高等学校的教学经验，所以本书贴近学生，在内容取舍上既考虑数据库及其管理系统的基本概念、基本原理介绍的系统性，又兼顾到数据库应用系统分析和设计的案例教学的实用性。为方便师生使用，每章后均附一定的习题，在全书后，还附带了实验指导。

本书章节安排如下：第 1 章绪论，第 2 章介绍关系模型和关系运算理论，第 3 章 SQL 关系数据库语言，第 4 章关系系统的查询及优化，第 5 章和第 6 章关系规范化设计理论和数据库设计，第 7 章介绍系统实现技术，第 8 章、第 9 章和第 10 章介绍对象数据库系统和分布式数据库系统以及数据库新技术。

本书的主要特点是理论与实践结合紧密，既强调理论介绍的系统性和完整性，又注重教学效果的实用性。在系统实现和应用中以当今较为流行的 SQL Server 作为介绍和分析对象。

本书由赵立江和王欣主编，全书最后由赵立江统稿和审定。

由于编者水平有限，加之时间仓促，书中难免存在纰漏和错误之处，殷切希望广大读者不吝指教。

编 者

2013 年 2 月 25 日

目 录

第1章 绪论	1
1.1 数据库与数据库系统引言	1
1.1.1 数据、信息与知识	2
1.1.2 数据处理与数据管理	2
1.1.3 数据库技术的基本概念	3
1.2 数据管理技术的演变	4
1.2.1 人工管理阶段(20世纪40年代中期—50年代中期)	4
1.2.2 文件系统阶段(20世纪50年代后期—60年代中后期)	4
1.2.3 数据库技术阶段	5
1.3 数据描述与数据模型	6
1.3.1 三级世界的划分	6
1.3.2 逻辑数据与物理数据	7
1.3.3 数据之间的联系	8
1.4 数据模型	9
1.4.1 数据模型的定义和类型	9
1.4.2 概念模型	9
1.4.3 层次模型、网状模型	10
1.4.4 关系模型	12
1.4.5 面向对象模型	13
1.5 数据库系统结构	13
1.5.1 数据库三级模式结构	14
1.5.2 数据库二级映象功能与数据独立性	15
1.5.3 数据库系统组成	15
本章小结	17
习题一	17
第2章 关系模型和关系运算理论	19
2.1 关系模型概述	19
2.1.1 基本术语	20
2.1.2 关系数据结构及形式化定义	21
2.1.3 关系模型的3类完整性规则	22

2.2 关系模式与关系数据库	23
2.3 关系代数	23
2.3.1 关系查询语言和关系运算	23
2.3.2 传统的集合运算	24
2.3.3 专门的关系运算	25
2.4 关系演算	28
2.4.1 元组关系演算	28
2.4.2 域关系演算	29
2.4.3 关系运算的安全约束	29
本章小结	30
习题二	30
第3章 关系数据库标准语言 SQL	32
3.1 SQL 概述	32
3.1.1 SQL 的产生与发展	32
3.1.2 SQL 数据库的体系结构	33
3.1.3 SQL 的特点	34
3.2 数据定义	36
3.2.1 基本表的定义、删除与修改	36
3.2.2 索引的建立与删除	38
3.3 SQL 数据查询	38
3.3.1 SELECT 查询语句	38
3.3.2 单表查询	40
3.3.3 连接查询	45
3.3.4 嵌套查询	47
3.3.5 集合查询	52
3.4 数据更新语句	53
3.4.1 插入数据	53
3.4.2 删除数据	54
3.4.3 修改数据	55
3.5 视图	56
3.5.1 视图的定义	56
3.5.2 视图的查询	58
3.5.3 视图的更新	59
3.5.4 视图的作用	60
3.6 数据控制	61
3.7 嵌入式 SQL	62
3.7.1 嵌入式 SQL 的一般形式	62
3.7.2 嵌入式 SQL 的工作原理	63



3.7.3 动态 SQL 简介	67
3.8 Transact-SQL 简介	68
本章小结	69
习题三	69
第 4 章 关系系统及查询优化	73
4.1 关系系统的定义及分类	73
4.2 关系系统中的查询优化	75
4.2.1 一个优化实例	75
4.2.2 关系代数表达式中的查询优化	77
4.2.3 查询优化的算法	79
本章小结	80
习题四	80
第 5 章 关系规范化设计理论	83
5.1 数据库设计的基本问题	83
5.2 规范化理论	85
5.2.1 函数依赖	85
5.2.2 码	86
5.2.3 范式	87
5.2.4 2NF	88
5.2.5 3NF	89
5.2.6 BCNF	90
5.2.7 多值依赖	90
5.2.8 4NF	92
5.3 数据依赖的公理系统	93
5.3.1 Armstrong 公理系统	93
5.3.2 函数依赖的闭包 F^+ 及属性的闭包 X_F^+	94
5.3.3 函数依赖集的等价和最小依赖集	95
5.4 关系模式的分解	98
5.4.1 无损连接的模式分解	99
5.4.2 保持函数依赖分解	100
5.4.3 模式分解的算法	100
本章小结	101
习题五	102
第 6 章 数据库设计	104
6.1 数据库设计概述	104
6.1.1 数据库设计方法简述	105
6.1.2 数据库设计的基本步骤	106

6.2	需求分析	108
6.2.1	需求分析方法	108
6.2.2	数据流图	109
6.2.3	数据字典	110
6.3	概念结构设计	112
6.3.1	概念结构设计方法	113
6.3.2	视图设计	114
6.3.3	视图的集成	116
6.3.4	EE-R 模型	119
6.4	逻辑结构设计	121
6.4.1	E-R 图向关系模型的转化	121
6.4.2	EE-R 图向关系模型的转化	122
6.4.3	数据模型的优化	124
6.4.4	用户模式的设计	125
6.4.5	逻辑结构的生成和图示	125
6.5	物理结构设计	126
6.5.1	索引设计	127
6.5.2	聚簇存取方法的选择	128
6.5.3	确定数据库的存储结构	128
6.5.4	物理结构设计的评价	129
6.6	数据库的实施	129
6.6.1	定义数据库结构	130
6.6.2	数据装载	130
6.6.3	编制与调试应用程序	130
6.6.4	数据库试运行	130
6.7	数据库的运行和维护	131
	本章小结	132
	习题六	132
第 7 章	系统实现技术	134
7.1	事务	134
7.1.1	事务的基本概念	134
7.1.2	事务的 ACID 性质	135
7.2	数据库的并发控制	136
7.2.1	并发控制概述	136
7.2.2	封锁与封锁协议	138
7.2.3	活锁与死锁	140
7.2.4	可串行性调度与两段锁协议	141
7.2.5	封锁的粒度	143



7.2.6	封锁的粒度	144
7.2.7	SQL Server 的并发控制技术	145
7.3	数据库的恢复	146
7.3.1	故障类型	146
7.3.2	恢复的实现技术	147
7.3.3	恢复的策略	149
7.3.4	具有检查点的恢复技术	150
7.3.5	数据库镜像	152
7.3.6	SQL Server 的恢复技术	152
7.4	数据库的完整性	153
7.4.1	完整性约束条件	153
7.4.2	完整性控制	155
7.4.3	完整性约束与触发器	155
7.4.4	SQL Server 的完整性技术	155
7.5	数据库的安全性	156
7.5.1	计算机安全性概论	156
7.5.2	数据库安全性控制	157
7.5.3	统计数据库的安全性	158
7.5.4	SQL Server 的安全性技术	158
	本章小结	159
	习题七	160
第 8 章	对象数据库系统	161
8.1	面向对象数据模型	161
8.1.1	数据模型的发展	161
8.1.2	面向对象数据模型的基本概念	162
8.1.3	面向对象数据模型	163
8.2	面向对象数据库语言	164
8.2.1	SQL-3 语言	164
8.2.2	OQL/ODL 语言	167
8.2.3	ORDB 和 OODB 两种面向对象语言的比较	170
	本章小结	170
	习题八	170
第 9 章	分布式数据库系统	172
9.1	分布式数据库概述	172
9.1.1	分布式数据库系统的定义	172
9.1.2	分布式数据库系统的特点	173
9.1.3	分布式数据库系统的分类	173
9.2	分布式数据库体系结构	174



9.2.1 分布式数据库系统的模式结构	174
9.2.2 分布式数据库管理系统	175
9.2.3 数据分片	175
9.2.4 分布透明性	175
9.3 查询和优化	176
9.3.1 查询的简化	176
9.3.2 连接查询的优化	176
9.4 分布事务管理	178
9.4.1 分布式事务的恢复	178
9.4.2 并发控制	179
本章小结	179
习题九	179
第 10 章 数据库新技术	181
10.1 数据库的新应用	181
10.1.1 数据模型及其新特征	181
10.1.2 现代数据库应用领域	183
10.2 数据库新技术的研究和发展	190
10.2.1 数据库技术的发展阶段	190
10.2.2 数据库新技术的研究和发展	191
本章小结	193
习题十	194
实验指导	195
实验一 认识和熟悉 DBMS 系统 (2 学时)	195
实验二 交互式 SQL 语言 (4 学时)	198
实验三 SQL 的查询优化 (2 学时)	199
实验四 宿主语言操纵数据库 (2 学时)	199
实验五 编写简单的数据库应用系统 (10 学时)	204
参考文献	206

第1章

绪论

本章要点

数据库技术是在 20 世纪 60 年代后期发展起来的计算机数据管理技术,它作为数据管理的一种有效手段,极大地促进了计算机应用技术的发展,数据库技术也因此成为计算机软件领域的一个重要分支。在学科教育中,数据库课程是计算机科学与技术专业以及信息管理专业等相关学科中一门重要的专业基础课。

本章通过介绍数据库技术产生和发展的历史介绍,应达到如下目标要求。

- (1) 掌握数据库系统的基本概念、数据库技术发展和演变过程及数据库的基本知识。
- (2) 掌握和理解数据模型,了解不同数据模型的结构特点及区别,掌握概念模型的概念以及主要建模方法 E-R 方法、关系模型的定义及相关概念,掌握数据库三级模式结构、两级映象及数据独立性的含义。
- (3) 了解数据库管理系统的组成与功能,数据库系统中各部分的作用和相互关系。

1.1 数据库与数据库系统引言

技术的快速发展和数据获取手段的多样化,使得人类进入信息社会后所拥有的数据和信息量呈一种迅猛增长的态势。据有关资料统计,到 20 世纪末,全球人类拥有的信息量总和已经达到了令人难以置信的程度和规模。面对极度膨胀的数据和信息量,人们用“信息爆炸”、“数据过剩”(Data Glut) 等来形容激增的数据量。但如果没有有效的数据管理工具,不能迅速从中发现与之相关的信息,那么随着数据信息量的增长,对人类而言,就不再是财富,而只能成为一场灾难。因此,为了能有效地管理和使用如此大量的数据,及时地从中抽取有用的信息,数据管理和数据库技术应运而生了。



1.1.1 数据、信息与知识

数据(Data)是数据库中存储和管理的基本对象,是描述事物属性的一种符号记录。数据可以分为两大类:一类是能参与数值运算的数值型数据,如表示工资、成绩、身高等的的数据;另一类则是不能参与数值运算的非数值型数据,如文字、图形、图像、声音等数据。数据也因此具有多样性、可构造性、持久性/挥发性、私有性/共享性、小量/海量等特性。

但需要注意的是,数据的形式并不能完全表达其内容,必须经过相应的语义解释,因此数据与其语义信息是密不可分的。

如果说数据是记录信息的符号载体的话,那么信息则是数据语义的内涵表示,数据与信息之间有着密不可分固有的联系。数据与信息的关系,犹如原料与成品的关系。未经过加工的数据只是一种原始材料,是一种符号记录,它只有经过抽取加工之后,具有了新的事实知识之后才能成为信息。此时的信息,仍然以数据形式表示,因而信息同样具有可表征性、可传播性、可处理性、可用性、可增值性、可替代性等特征。

需要注意的是,数据与信息这两个概念,在不严格区分的场合下,可以混用,如将“信息处理”、“信息管理”说成是“数据处理”、“数据管理”等,这并不影响对它们在概念上的不同理解。

知识也是一种信息,但它是经过人脑加工和理解之后的数据和信息,可以说这是信息的一种更高级形态。就像拥有了数据并不能简单地等同于拥有了信息一样,知识的获得并非是简单的信息复制,而是一种缓慢、复杂的,人脑通过思维重新组合、系统化信息的过程。

1.1.2 数据处理与数据管理

1. 数据处理

数据处理(Data Processing)是对各种类型的数据进行采集、存储、检索、加工、变换和传输的过程。其基本目的是从大量的、可能是杂乱无章的、难以理解的原始数据中抽取并推导出对于某些特定的人们来说是有价值、有意义的的数据,也就是信息。

数据处理离不开软件的支持,数据处理软件包括用以书写处理程序的各种程序设计语言及其编译程序,管理数据的文件系统和数据库系统,以及各种数据处理方法的应用软件包。为了保证数据安全可靠,还有一整套数据安全保密的技术。

根据处理设备的结构方式、工作方式,以及数据的时间空间分布方式的不同,数据处理有不同的方式。不同的处理方式要求不同的硬件和软件支持。每种处理方式都有自己的特点,应当根据应用问题的实际环境选择合适的处理方式。

数据处理主要有4种分类方式。

(1) 根据处理设备的结构方式区分,有联机处理方式和脱机处理方式。

(2) 根据数据处理时间的分配方式区分,有批处理方式、分时处理方式和实时处理方式。

(3) 根据数据处理空间的分布方式区分,有集中式处理方式和分布处理方式。

(4) 根据计算机中央处理器的工作方式区分,有单道作业处理方式、多道作业处理方式和交互式处理方式。



2. 数据管理

数据处理的核心问题是数据管理，计算机对数据的管理是指对数据的组织、分类、编码、存储、检索和维护提供操作手段。在应用需求的推动下，在计算机硬件、软件发展的基础上，计算机数据管理经历了人工管理、文件系统、数据库系统等几个阶段。

数据库技术就是研究如何科学正确地组织、存储数据，如何高效地获取和处理数据，这是在文件管理系统上发展起来的一种理想数据管理技术。

1.1.3 数据库技术的基本概念

1. 数据库

数据库 (Data Base, DB) 是长期存储在计算机存储设备上，有组织、结构化、可共享的相关数据集合。数据库的数据按照一定的数据模型组织、描述和存储，具有较小的冗余度，较高的数据独立性和易扩展性，并且能为各种用户共享。

数据库中的数据面向多种应用，可以被多个用户、多个应用程序共享。例如，某个企业、组织或行业所涉及的全部数据的汇集。

数据库的结构是独立于使用数据的程序的，而对于数据库的数据增删、修改、检索等操作是由系统软件进行统一的控制的。

总之，数据库是实行统一管理的相关数据的集合，并且可以为各类用户所共享，具有较小的数据冗余和较高的数据独立性等特点。

2. 数据库应用系统

数据库应用系统 (Data Base Application System, DBAS) 是指系统开发人员利用数据库系统资源开发出来的，面向某一类实际应用的应用软件系统。例如，财务管理系统，人事管理系统等。

3. 数据库管理系统

数据库管理系统 (Data Base Management System, DBMS) 是位于用户与操作系统之间的一层数据管理软件。为了让多种应用程序并发地使用数据库中具有最小冗余度的共享数据，必须是数据与程序既有较高的独立性。这需要一个软件系统对数据实行专门管理，进行统一控制，方便用户对数据库进行的各种操作并保证数据库中数据始终处于正确和安全的状态。

数据库在建立、运用和维护过程中就由数据库管理系统 DBMS 统一管理、统一控制。数据库管理系统使用户能方便地定义数据和操纵数据，并保证数据的安全性、完整性、提供多用户对数据的并发访问以及进行故障后的系统恢复。

4. 数据库系统

数据库系统 (Data Base System, DBS) 是指引进数据库技术后的计算机系统。它可以有组织地、动态地存储大量相关数据，提供数据处理和信息资源共享的便利手段。

数据库系统由 5 部分组成：硬件系统、数据库集合、数据库管理系统及相关软件、数据库管理员和用户。图 1-1 所示为这几个层次之间的关系。

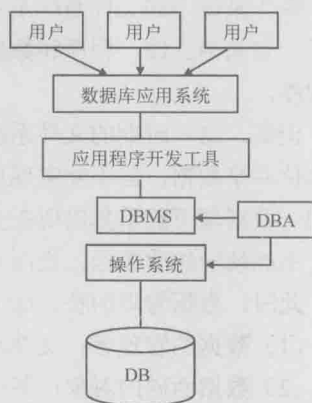


图 1-1 数据库系统示意图

1.2 数据管理技术的演变

1.2.1 人工管理阶段（20世纪40年代中期—50年代中期）

在20世纪40年代中期—50年代中期这一阶段中，计算机的主要应用目的是完成科学计算，在硬件中的外存储器没有磁盘一类可以随机访问、直接存取的设备，软件上没有专门的管理数据的软件，数据由计算或处理数据的程序自行携带，所以数据管理任务由人工完成。

在该阶段，数据与程序不具有独立性，一组数据对应一组程序。数据不长期保存，一个程序中的数据无法被其他程序利用，程序与程序间存在大量的重复数据，称为数据冗余。此时的数据管理有如下特点。

(1) 数据不保存。由于没有文件的概念，数据的对应管理者只能是应用程序，对于数据保存的需求尚不迫切。一旦程序结束，数据也随之撤销。

(2) 应用程序管理数据。由于没有专用的软件对数据进行管理，应用程序不但定义数据的逻辑结构，而且还要设计物理结构，造成程序员负担很重。

(3) 数据无共享、冗余度极大。由于没有文件的概念，并且没有相同的数据结构，出现了当多个应用程序涉及存取相同的数据时，由于对数据的管理只能各自为政，造成了数据大量冗余的现象。

(4) 数据不具有独立性。数据是完全依赖于应用程序的，不具备独立性。人工管理阶段的特点如图1-2所示。

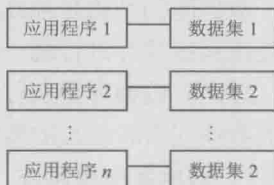


图 1-2 人工管理阶段的特点

1.2.2 文件系统阶段（20世纪50年代后期—60年代中后期）

到20世纪50年代后期至60年代中后期，大量的数据存储、检索和维护成为当时紧迫的需求，可直接存取的磁盘成为联机的主要外存，软件上已经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储器的数据管理软件。在文件系统阶段，程序与数据有了一定的独立性，程序和数据分开，数据文件和程序之间的独立性增强，存储设备的改变不影响程序。

但是，这一时期的文件系统的文件主要是服务于某一特定的应用程序，用户允许以文件为单位共享数据，但不能实现以记录和数据项为单位共享，因此数据和程序还存在相互依赖，而且同一数据项可能重复出现在多个文件中，数据冗余量大，浪费空间，增加更新开销。由于冗余多，不能统一修改数据，进而可能出现数据的不一致性。

此时，数据管理的特点如下。

(1) 数据的管理者：文件系统，数据可长期保存。

(2) 数据面向的对象：某一应用程序。

(3) 数据的共享程度：共享性差、冗余度大。

(4) 数据的结构化：记录内有结构，整体无结构。

(5) 数据的独立性差: 数据的逻辑结构改变必须修改应用程序。

(6) 数据控制能力: 应用程序自己控制。

文件管理阶段的特点如图 1-3 所示。

文件系统阶段是数据管理技术发展中的一个重要阶段, 这一阶段中得到充分发展的各种数据结构和算法丰富了计算机科学, 并为数据管理技术的发展奠定了基础。

但随着数据量的激增, 此时的文件系统日益暴露出以下 3 个缺陷。

(1) 数据的冗余度大。文件之间缺乏联系, 造成相同数据在不同地方(文件)的重复存储, 浪费存储空间。

(2) 数据之间出现不一致性。数据冗余的另一个后果就是造成数据间的不一致性, 出现无用甚至是矛盾的数据。

(3) 数据之间的联系弱。文件之间的相互独立, 造成了数据孤立, 相互访问困难以及并发访问异常, 安全性脆弱等问题。

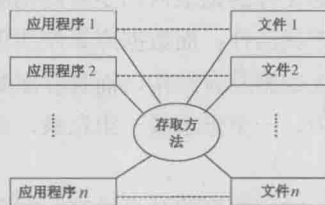


图 1-3 文件管理阶段的特点

1.2.3 数据库技术阶段

到了 20 世纪 60 年代后期, 计算机的应用大为普及, 数据规模急剧增长, 对数据共享的需求日益增强, 同时也为了克服文件系统存在的问题, 人们开始研究和发展数据库技术。

1968—1970 年发生的三大历史事件标志着数据库技术的成熟。

- IBM 公司 1968 年成功研制层次数据管理系统 (IMS)。
- 美国 CODASYL (Conference On Data System Language, 数据系统语言协商会) 1971 年公布的 DBTG 报告提出网络数据库系统 CODASYL 系统 (或 DBTG 系统)。
- IBM 公司 E.F.Codd 1970 发表一系列论文, 奠定了关系数据库系统 (RDBMS) 理论基础。

从文件系统发展到数据库管理系统是数据管理发展史上的一个重大变革, 它将过去在文件系统中的以应用程序为核心, 数据服从应用程序的数据管理模式改变为以数据库设计为核心、应用程序退居次位的数据管理模式。

数据库技术的主要目的是有效地管理和存取大量数据资源, 包括提高数据的共享性, 使多个用户能够同时访问数据库中的数据, 减小数据的冗余度, 以提高数据的一致性和完整性; 提供数据与应用程序的独立性, 从而减少应用程序的开发和维护代价。

为数据库的建立、使用和维护而配置的软件称为数据库管理系统 (DBMS)。数据库管理系统

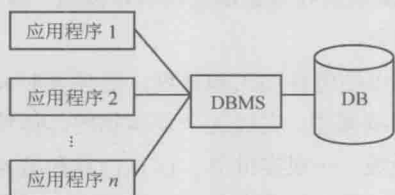


图 1-4 数据库管理阶段的特点

利用了操作系统提供的输入/输出控制和文件访问功能, 因此它需要在操作系统的支持下运行。

从文件系统发展到数据库管理系统是数据管理发展的一个重大转折, 它将过去在文件系统中以程序设计为核心, 数据服从程序设计的数据库管理模式改变为以数据库设计为核心、应用程序退居次位的数据管理模式。这一阶段数据管理的特点如图 1-4 所示。

(1) 数据实现了整体结构化。这是数据库系统与文件系统的根本区别。文件系统中, 相互独

立的文件的记录内部是有结构的，最简单的形式是等长记录，但这种面向某一具体应用的结构，缺乏灵活性；而数据库系统实现了整体的结构化，数据不再是面向某一应用，而是面向全组织，不仅数据具有结构，而且存取数据的方式非常灵活，可以存取数据库中的某一个数据项、一组数据项、一个记录或一组记录，而在文件系统中，数据的最小存取单位是记录，粒度不能细到数据项。

(2) 数据的共享性高，冗余度低，易扩充。数据库系统从整体角度看待和描述数据，数据不再面向某个应用而是面向整个系统，数据可以被多个用户、多个应用共享使用。数据共享可以大大减少数据冗余，节约存储空间，避免数据之间的不相容性与不一致性，使数据库系统易于扩充，可适应各种用户的需求。

(3) 数据的独立性程度高。数据的独立性包括数据的物理独立性和逻辑独立性。前者是指用户的应用程序与存储在磁盘上的数据库中数据是相互独立的，当数据的物理存储改变了，应用程序不用改变；后者是指用户的应用程序与数据库的逻辑结构是相互独立的，数据的逻辑结构改变了，用户程序也可以不变。

(4) 数据实现了 DBMS 统一管理和控制。DBMS 提供以下控制功能。

① 数据的安全性 (Security): 保护数据以防止不合法的使用造成的数据的泄密和破坏，使每个用户只能按指定方式操纵数据。

② 数据的完整性 (Integrity): 将数据控制在有效的范围内，或保证数据之间满足一定的关系。

③ 并发 (Concurrency) 控制: 对多用户的并发操作加以控制，防止相互干扰而得到错误的结果。

④ 数据库恢复 (Recovery): 将数据库从错误状态恢复到某一已知的正确状态。

1.3 数据描述与数据模型

显然，数据应该是整个数据库系统的基础，实际上在任何一个数据库中都存储了两种类型的数据，第一类是实际的数据，第二类是描述数据语义的数据，即关于数据库结构、组织形式的元数据 (Meta Data)。早期的系统是用数据定义语言来刻画模式，用这种方式描述的企业的数据需求，层次太低，用户不便理解，因此需要一种更高层次的描述模式的方法，即数据模型。那么，数据模型的概念是怎样提出来的呢？

1.3.1 三级世界的划分

关于模型大家并不陌生，现实生活中有车模、航模、建筑模型等形象模型和模拟模型。显然，有了模型后就可以对所研究的事物进行相应特征的模拟和抽象。

为了能够让计算机对现实世界中的具体事物进行描述，以便进行加工和处理，必须要对具体事物进行抽象，即要将其表示成某一个 DBMS 所能支持的数据模型。完成这个从事物的具体特性到计算机中的对应表示的过程，这中间实际上要经历 3 个领域——现实世界、信息世界和机器世界的变换。本节首先介绍 3 个世界中的数据描述，再介绍逻辑数据与物理数据，最后介绍数据之间的联系。

1. 现实世界

现实世界也就是我们所处的客观世界，通常由各种各样的事物组成。由于计算机不能直接处理现实世界中的具体事物，必须用一定的数据模型来刻画和描述。人们常常将现实世界通过概念模型抽象为信息世界，这是第一层抽象（见图 1-5），然后再通过数据模型将信息世界转换为机器世界。

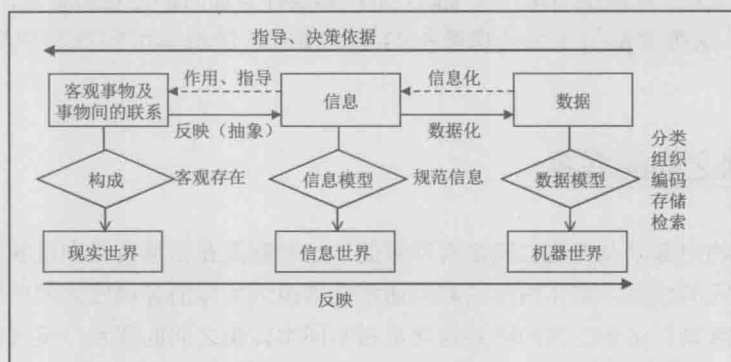


图 1-5 3 个世界之间的联系及对应关系

2. 信息世界

信息世界是现实世界在人头脑中的反映。客观的事物及其性质被抽象成实体以及属性。信息世界所涉及的概念如下。

(1) 实体 (Entity)。客观存在，可相互区别的事物被称为实体。

实体可以是具体的人、事、物，也可以是抽象的概念或联系。例如，一个学生、一门课程、学生的一次选课、老师与课程的讲授关系等都是实体。

(2) 属性 (Attribute)。实体有若干个特性，所具有的某一特性称为属性。属性的取值范围称作其值域。属性有型与值的区分，如姓名、年龄等是属性型，而李明、18 岁等则是属性值。

(3) 码 (Key)。能唯一标识实体的属性或最小属性集称为码。例如，学号是学生实体的码。

(4) 实体型 (Entity Type)。具有相同属性的实体必然具有共同的特征和性质。用实体名及其属性名集合来抽象和刻画同类实体，称为实体型。例如，学生（学号，姓名，性别，出生年份，系，入学时间）就是一个实体型。

(5) 实体集 (Entity Set)。同型实体的集合称为实体集。例如，全体学生就是一个实体集。

3. 机器世界

信息世界中的信息在机器世界中只能以数据形式存储，常常用如下术语进行描述。

(1) 字段 (Field)：标记实体属性的最小单位，又称为数据项。字段的命名往往和属性名相同，如学生有学号、姓名、性别、年龄等字段。

(2) 记录 (Record)：字段的有序集合称为记录。

(3) 现实世界中的实体及其联系，在机器世界中用数据模型表示。

1.3.2 逻辑数据与物理数据

在数据库中按照数据模型对数据进行组织和描述。数据的描述有两种形式：逻辑数据和物理