



普通高等教育“十三五”计算机类规划教材

数据库 原理及应用

◎ 胡孔法 主编

Shujuku
Yuanli Ji Yingyong

第2版



机械工业出版社
CHINA MACHINE PRESS

普通高等教育“十三五”计算机类规划教材

数据库原理及应用

第2版

主 编 胡孔法

副主编 余侃侃

参 编 张福安 胡晨骏 朱新山



机械工业出版社

本书将数据库基本原理、方法和应用技术相结合,以关系数据库系统为核心,完整地介绍了数据库系统的基本概念及基本原理,并对 Microsoft SQL Server 等主流数据库管理系统, Visio 和 PowerDesigner、ASP.NET 和 ADO.NET 等数据库设计与软件开发工具进行了讲解,还对数据仓库与数据挖掘、XML 数据库等新技术进行了介绍。主要内容包括数据库系统概述、数据模型、数据库系统的体系结构、关系数据库方法、关系数据库的结构化查询语言、关系模式的规范化理论、数据库设计、数据库保护、Microsoft SQL Server 2012、ASP.NET 和 ADO.NET 数据库开发技术、数据仓库与数据挖掘和 XML 数据库等。

本书可作为高等院校计算机类专业、信息管理与信息系统以及相关专业的本科生教材,还可作为广大软件设计与开发工程技术人员及信息领域的科技人员的参考书。通过本书的学习,读者应能熟练地使用现有的数据库管理系统和软件设计与开发工具,进行数据库结构的设计和数据库应用系统开发。

为方便教师教学,本书配有免费教学课件,欢迎选用本书作为教材的教师登录 www.cmpedu.com 下载或发邮件到 llm7785@sina.com 索取。

图书在版编目(CIP)数据

数据库原理及应用/胡孔法主编. —2版. —北京:机械工业出版社, 2015.8

普通高等教育“十三五”计算机类规划教材
ISBN 978-7-111-51129-8

I. ①数… II. ①胡… III. ①关系数据库系统—高等学校—教材
IV. ①TP311.138

中国版本图书馆 CIP 数据核字(2015)第 184648 号

机械工业出版社(北京市百万庄大街 22 号 邮政编码 100037)
策划编辑:刘丽敏 责任编辑:刘丽敏 吴晋瑜
责任校对:刘怡丹 封面设计:张静
责任印制:康朝琦
北京圣夫亚美印刷有限公司印刷
2015 年 9 月第 2 版第 1 次印刷
184mm×260mm·18 印张·489 千字
标准书号:ISBN 978-7-111-51129-8
定价:38.00 元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

电话服务

网络服务

服务咨询热线:010-88379833

机工官网:www.cmpbook.com

读者购书热线:010-88379649

机工官博:weibo.com/cmp1952

教育服务网:www.cmpedu.com

封面无防伪标均为盗版

金书网:www.golden-book.com

前 言

20世纪60年代末,数据库技术是作为数据处理中的一门新技术发展起来的。时至今日,数据库技术已形成了较为完整的理论体系,是计算机软件领域的一个重要分支。

随着数据库系统的推广,计算机应用已深入人类社会的各个领域,如当前的管理信息系统(MIS)、企业资源规划(ERP)、计算机集成制造系统(CIMS)、地理信息系统(GIS)、决策支持系统(DDS)等都是数据库技术为基础的。此外,我国实施的国家信息化、“金”字工程、数字城市等都是数据库为基础的大型计算机系统。目前,数据库的建设规模和性能、数据库信息量的大小和使用水平已成为衡量一个国家信息化程度的重要标志。我国高等院校从20世纪80年代开始就把数据库原理及应用作为计算机类专业的主要课程之一。目前,数据库原理及应用课程是各大高等院校计算机类专业、信息管理与信息系统以及相关专业的一个重要专业基础课程。

本书很好地将数据库基本原理、方法和应用技术相结合,以关系数据库系统为核心,在全面解析数据库系统的基本概念及基本原理的基础上,对Microsoft SQL Server等主流数据库管理系统、Visio和PowerDesigner、VS.net等数据库设计与软件开发工具进行讲解,并对数据仓库与数据挖掘、XML数据库等新技术进行了介绍,旨在让高等院校计算机类专业、信息管理与信息系统以及相关专业的本科生以及广大软件设计与开发工程技术人员在学习本书后,能熟练使用现有的数据库管理系统和软件设计与开发工具,进行数据库结构的设计和数据库应用系统开发。

本书共分12章。第1章数据库系统概述,介绍数据库系统基本概念,并讨论了数据管理和数据库系统的发展过程。第2章数据模型,着重介绍E-R数据模型、层次数据模型、网状数据模型、关系数据模型、面向对象数据模型等的基本概念和设计方法。第3章数据库系统的体系结构,主要介绍数据库系统的三级模式结构,DBS、DBMS、DBMS的系统结构等内容。第4章关系数据库方法,介绍关系数据库的基本概念、代数运算、关系演算、关系查询优化等内容。第5章关系数据库的结构化查询语言,主要介绍SQL概述、SQL的数据定义语言、SQL数据更新、SQL数据查询、SQL聚集函数、SQL中的视图等内容。第6章关系模式的规范化理论,主要介绍函数依赖、范式和模式设计方法等关系模式规范化理论。第7章主要介绍数据库设计、概念设计、数据库的物理设计、PowerDesigner辅助设计工具等内容。第8章主要介绍数据库完整性、数据库安全性、数据库恢复技术和并发控制等数据库保护措施。第9章主要介绍关系数据库系统Microsoft SQL Server的基本知识,并以Microsoft SQL Server 2012为背景,介绍了数据库系统设计方法和SQL Server的高级应用技术。第10章主要介绍ASP.NET和ADO.NET基础知识、ASP.NET连接数据库方法、ADO.NET读取和操作数据库数据等基于.NET的数据库开发技术。第11章主要介绍数据仓库和数据挖掘的基本概念、数据仓库设计方法与实现过程、数据挖掘的主要技术和过程。第12章主要介绍XML文档、XML模式、XPath查询语言、XQuery查询语言等知识,并给出了XML在数据管理方面的应用实例。

本书第1~6章、第11章由胡孔法编写,第7章由胡晨骏编写,第8章由张福安编写,第9章和第10章由余侃侃编写,第12章由朱新山编写。



作者结合自己在读硕士和博士期间所从事的数据库及信息系统方面的研究成果以及十余年的软件开发与设计经验和多年的教学经验，完成了本书的编写。

本书可作为计算机及相关专业的本科生教材，也可供研究生、广大软件设计和开发人员参考。书中如有不足之处，敬请广大读者指正。

作者

目 录

前言

第1章 数据库系统概述	1	2.6 面向对象数据模型	24
1.1 数据库技术的产生与发展	1	2.6.1 对象和对象标识符	24
1.1.1 数据管理的发展	1	2.6.2 属性和方法	24
1.1.2 数据和数据管理技术	2	2.6.3 封装和消息传递	25
1.1.3 数据库系统的三个发展阶段	3	2.6.4 类和实例	25
1.2 数据库系统的基本概念	4	2.6.5 类层次结构和继承	25
1.2.1 数据库	4	2.6.6 持久性和版本	26
1.2.2 数据库系统	5	2.6.7 多态、重载、重定义与动态联编	26
1.3 数据库系统的特点	6	2.6.8 面向对象数据模型与关系数据 模型的比较	27
第2章 数据模型	8	第3章 数据库系统的体系结构	28
2.1 数据模型概述	8	3.1 数据库的体系结构	28
2.1.1 数据模型的基本组成	8	3.1.1 三级模式结构	28
2.1.2 数据模型的发展	9	3.1.2 两级映像和两级数据独立性	29
2.2 E-R 数据模型	9	3.2 数据库系统	30
2.2.1 基本概念	9	3.2.1 DBS 的组成	30
2.2.2 E-R 图	11	3.2.2 DBS 的全局结构	32
2.2.3 扩充 E-R 数据模型	12	3.2.3 DBS 结构的分类	34
2.3 层次数据模型	13	3.3 数据库管理系统	36
2.3.1 基本概念和结构	13	3.3.1 DBMS 的工作模式	36
2.3.2 数据操作	15	3.3.2 DBMS 的主要功能	36
2.3.3 数据约束	16	3.4 DBMS 的系统结构	38
2.3.4 层次数据模型的优缺点	17	3.4.1 DBMS 的进程结构和多线索结构	38
2.4 网状数据模型	17	3.4.2 DBMS 的组成	38
2.4.1 基本概念和结构	17	3.4.3 DBMS 的层次结构	40
2.4.2 数据操作	19	3.5 几种典型结构的 DBMS	41
2.4.3 数据约束	20	3.5.1 C/S 结构的 DBMS	41
2.4.4 网状数据模型的优缺点	20	3.5.2 基于 Web 的数据库系统	43
2.5 关系数据模型	20	第4章 关系数据库方法	47
2.5.1 基本概念	20	4.1 关系数据库的基本概念	47
2.5.2 关系数据模型的数据结构	21	4.1.1 关系的形式化定义	47
2.5.3 数据操作	22	4.1.2 关系模式、关系子模式和存储 模式	49
2.5.4 数据约束	22		
2.5.5 关系数据模型的优缺点	23		



4.1.3	关系模型的完整性规则	51	5.6.5	视图的作用	101
4.1.4	关系数据库模式	52	第6章 关系模式的规范化理论		102
4.2	关系代数	53	6.1	关系模式设计中的问题	102
4.2.1	关系代数的五个基本操作	53	6.2	函数依赖	103
4.2.2	关系代数的组合操作	57	6.2.1	函数依赖定义	103
4.2.3	关系代数表达式应用举例	61	6.2.2	关键字和超关键字	104
4.3	关系演算	62	6.3	数据依赖的公理系统	105
4.3.1	元组关系演算	62	6.3.1	函数依赖的逻辑蕴涵	105
4.3.2	域关系演算	63	6.3.2	Armstrong 公理系统	105
4.4	关系查询优化	64	6.3.3	函数依赖集的等价和覆盖	109
4.4.1	查询优化的一般策略	64	6.4	关系模式的分解及其问题	110
4.4.2	关系代数优化	65	6.4.1	分解的无损连接性	111
4.4.3	基于存取路径的规则优化	69	6.4.2	保持函数依赖性	116
第5章 关系数据库的结构化查询语言		76	6.5	关系模式的规范化	118
5.1	SQL 概述	76	6.5.1	范式	118
5.2	SQL 的数据定义语言	77	6.5.2	模式分解的算法	120
5.2.1	数据类型	77	6.6	多值函数依赖与4NF	124
5.2.2	数据库模式的定义	78	6.6.1	BCNF 关系模式存在的问题	124
5.2.3	基本表的定义	79	6.6.2	多值函数依赖	124
5.2.4	基本表的修改和删除	82	6.6.3	4NF	125
5.2.5	索引的建立和删除	83	6.7	连接依赖和5NF	126
5.3	SQL 数据更新	84	第7章 数据库设计		127
5.3.1	元组插入	84	7.1	数据库设计概述	127
5.3.2	元组删除	85	7.1.1	数据库系统生存期	127
5.3.3	元组修改	85	7.1.2	数据库设计方法	128
5.4	SQL 数据查询	86	7.1.3	数据库设计的基本过程	129
5.4.1	SQL 查询语句的格式	86	7.2	规划	130
5.4.2	简单查询	87	7.3	需求分析	131
5.4.3	复杂查询	89	7.3.1	需求描述与分析	131
5.5	SQL 聚集函数	96	7.3.2	需求分析阶段的输入和输出	132
5.5.1	聚集函数的运算符	96	7.3.3	需求分析的步骤	132
5.5.2	数据分组	96	7.3.4	数据字典	136
5.5.3	数据排序	97	7.4	概念设计	138
5.6	SQL 中的视图	98	7.4.1	概念设计的必要性	138
5.6.1	视图定义	98	7.4.2	概念设计的主要步骤	139
5.6.2	视图查询	99	7.4.3	采用 E-R 模型方法的数据库概念设计	140
5.6.3	视图更新	100	7.5	数据库逻辑结构设计及优化	148
5.6.4	视图删除	100	7.5.1	逻辑设计环境	148

7.5.2 E-R 模型向关系模型的转换	148	9.1.3 Microsoft SQL Server 2012 的工具 介绍	178
7.5.3 用关系规范化理论对关系数据 模型进行优化	151	9.2 数据库的创建、修改和删除	181
7.6 数据库的物理设计	153	9.2.1 创建 SQL Server 数据库	181
7.7 数据库的实现	153	9.2.2 分离和附加 SQL Server 数据库	183
7.8 数据库的运行与维护	153	9.2.3 删除数据库	186
7.9 PowerDesigner 辅助设计工具	154	9.3 表和主键的创建	186
7.9.1 绘制 E-R 图	154	9.3.1 表的创建	186
7.9.2 后台生成 SQL 数据库及 数据表	157	9.3.2 主键的创建	186
第 8 章 数据库保护	161	9.3.3 用 SQL 语句方式创建表	188
8.1 事务	161	9.4 数据的插入、修改、删除和 查询	189
8.1.1 事务的定义	161	9.4.1 数据的插入	189
8.1.2 事务的 ACID 准则	161	9.4.2 数据的修改	191
8.2 数据库完整性	162	9.4.3 数据的删除	191
8.2.1 完整性子系统和完整性规则	162	9.4.4 数据的查询	191
8.2.2 SQL 中的完整性约束	163	9.5 数据库的备份和恢复	192
8.2.3 SQL3 触发器	164	9.5.1 数据库的备份	192
8.3 数据库安全性	165	9.5.2 数据库的恢复	194
8.3.1 数据库安全性级别	165	9.6 高级应用技术	197
8.3.2 数据访问权限	165	9.6.1 存储过程	197
8.3.3 SQL 中的安全性机制	166	9.6.2 触发器	201
8.4 数据库恢复技术	167	第 10 章 ASP.NET 和 ADO.NET 数据库 开发技术	206
8.4.1 恢复的定义、原则和实现方法	167	10.1 ASP.NET 基础知识	206
8.4.2 故障恢复方法	167	10.1.1 Visual Studio .NET 的特色	206
8.4.3 具有检查点的恢复技术	168	10.1.2 .NET Framework	207
8.5 并发控制	169	10.1.3 ASP.NET 开发环境介绍	208
8.5.1 并发所引起的问题	169	10.2 ADO.NET 介绍	208
8.5.2 封锁	170	10.2.1 ADO.NET 与数据管理简介	208
8.5.3 活锁和死锁	172	10.2.2 ADO.NET 命名空间	209
8.5.4 死锁的防止、检测和处理	173	10.3 连接数据库	209
8.5.5 并发调度的可串行化	174	10.3.1 连接 Microsoft SQL Server 数据库	209
8.5.6 两段封锁协议	175	10.3.2 连接到 Microsoft Access 数据库	214
第 9 章 Microsoft SQL Server 2012	177	10.4 读取和操作数据	215
9.1 SQL Server 简介	177	10.5 数据集	218
9.1.1 Microsoft SQL Server 2012 的 特点	177	10.5.1 使用 DataAdapter 类	219
9.1.2 Microsoft SQL Server 2012 环境 介绍	177		



10.5.2 使用 DataTables、DataColumns 和 DataRows	219	11.4.3 数据挖掘的主要技术	253
10.6 DataReader 类	221	11.4.4 数据挖掘的过程	256
10.7 DataGrid 和 GridView 控件	223	11.4.5 DM 与 OLAP	257
10.8 DataList 控件	226	第 12 章 XML 数据库	258
第 11 章 数据仓库与数据挖掘	238	12.1 引言	258
11.1 概述	238	12.2 XML 文档	259
11.2 数据仓库的概念	239	12.3 XML 文档类型定义	262
11.2.1 数据仓库的定义	239	12.3.1 元素定义	263
11.2.2 数据仓库的结构	241	12.3.2 属性定义	264
11.3 数据仓库的设计与实现	246	12.4 XML 模式	265
11.3.1 数据仓库的设计原则	246	12.5 XPath 查询语言	269
11.3.2 数据仓库的三级数据模型	247	12.6 XQuery 查询语言	270
11.3.3 数据仓库的设计步骤	248	12.7 XML 应用	271
11.3.4 数据仓库的实现结构	251	12.7.1 基于 XML 的数据交换与异构 数据集成	271
11.4 数据挖掘	251	12.7.2 XML 索引与查询处理	272
11.4.1 数据挖掘定义	251	12.7.3 XML 文档聚类	275
11.4.2 数据挖掘技术的应用 研究现状	252	12.7.4 XML 流处理	276
		参考文献	278

第 1 章 数据库系统概述

20 世纪 60 年代末，数据库技术是作为数据处理中的一门新技术发展起来的。时至今日，数据库技术已是计算机软件领域的一个重要分支，形成了较为完整的理论体系和实用技术。本章先介绍数据库技术的发展过程，然后介绍数据库中的基本概念，以使读者对数据库的概貌有所了解。

1.1 数据库技术的产生与发展

1.1.1 数据管理的发展

数据库系统的萌芽出现于 20 世纪 60 年代。当时，计算机开始广泛应用于数据管理，对数据的共享提出了越来越高的要求。传统的文件系统已经不能满足人们的需要，能够统一管理和共享数据的数据库管理系统（Database Management System, DBMS）应运而生。数据模型是数据库系统的核心和基础，各种 DBMS 软件均基于某种数据模型。所以，通常也按照数据模型的特点将传统数据库系统分成网状数据库、层次数据库和关系数据库。

世界上第一个 DBMS 应当是美国通用电气公司 Bachman 等人在 1964 年开发成功的 IDS（Integrated Data Store 集成数据存储）。IDS 奠定了网状数据库的基础，并在当时得到了广泛应用。IDS 具有数据模式和日志的特征，但它只能在 GE 主机上运行，并且数据库只有一个文件，数据库所有的表必须通过手工编码来生成。

最著名最典型的层次数据库系统是 IBM 公司在 20 世纪 60 年代末开发的 IMS（Information Management System），它是一种适合其主机的层次数据库。这是 IBM 公司研制的最早的大型数据库系统程序产品。

网状数据库和层次数据库已经很好地解决了数据的集中和共享问题，但是在数据独立性和抽象级别上仍有很大欠缺。用户在对这两种数据库进行存取时，仍然需要明确数据的存储结构，指出存取路径。而后来出现的关系数据库较好地解决了这些问题。

1970 年，IBM 的研究员 E. F. Codd 博士在《Communication of the ACM》上发表了一篇名为“A Relational Model of Data for Large Shared Data Banks”的论文，提出了关系模型的概念，奠定了关系模型的理论基础。尽管之前在 1968 年 Childs 已经提出了面向集合的模型，然而这篇论文被普遍认为是数据库系统历史上具有划时代意义的里程碑。后来 Codd 又陆续发表多篇文章，论述了范式理论和衡量关系系统的 12 条标准，用数学理论奠定了关系数据库的基础。关系模型有着严格的数学基础，抽象级别比较高，而且简单清晰，便于理解和使用。但是当时也有人认为关系模型是理想化的数据模型，用来实现 DBMS 是不现实的，尤其担心关系数据库的性能难以接受，更有人视其为当时正在进行中的网状数据库规范化工作的严重威胁。为了促进对问题的理解，1974 年，美国计算机协会（Association for Computing Machinery, ACM）牵头组织了一次研讨会，开展了一场分别以 Codd 和 Bachman 为首的支持和反对关系数据库两派之间的辩论。这次著名的辩论推动了关系数据库的发展，不少工业和学术单位投入到关系 DBMS 原型的研制，使其最终成为现代数据库产品的主流。在众多的关系 DBMS 原型中，功能最全面、技术上最有代表性的是 IBM 公司的 System R 和加州大学伯克利分校的 INGRES。这两个原型系统全面地提供了比较成熟

的关系 DBMS 技术, 为研制商品化的关系 DBMS 提供了技术上的准备。IBM 公司在 System R 的基础上先后推出了 SQL/DS 和 DB2 两个商品化的关系 DBMS。关系型数据库系统以关系代数为坚实的理论基础, 经过几十年的发展和实际应用, 其技术日趋成熟和完善。其代表产品有 Oracle、Sybase、Informix、MS SQL Server 等。

20 世纪 80 年代是关系数据库的全盛时代。随着计算机的广泛应用, 新的应用又提出了新的要求。人们开始发现关系数据库的诸多限制和不足, 这又推动了数据库技术的新一轮研究: 一是改造和扩充关系数据库, 以适应新的应用需求; 二是改用新的数据模型 (例如, 面向对象数据模型等), 研制新型数据库。

数据管理与数据处理一样, 都是计算机系统最基本的支撑技术。尽管计算机科学技术经历了飞速的发展, 但数据管理的这一地位没有变化。可以预见, 数据管理将作为计算机科学技术的一个重要分支一直发展下去, 因为社会愈信息化, 对数据管理的要求也愈高。推动这门学科发展的动力是计算机应用, 发展这门学科的基础是新的硬、软件技术。数据库是现阶段数据管理的主要形式。

1.1.2 数据和数据管理技术

1. 数据

为了了解世界、研究世界和交流情况, 人们需要描述事物。人们常常通过抽象那些感兴趣的事物的特征或属性来对事物进行描述。

例如, 一个大学生可用如下的记录来描述: <李定, 091231142, 男, 1991, 江苏, 计算机系, 2009>。这表示李定是位大学生, 学号是 091231142, 1991 年出生, 江苏人, 2009 年考入计算机系。这种描述事物的符号记录称为数据。

(1) 数据的含义

数据是载荷信息的媒体, 是对现实世界中客观事物的符号表示。它既可以是数值数据, 也可以是非数值数据, 如声音、图像等。它是能输入计算机, 并能为其处理的符号序列。

(2) 数据与信息的关系与区别

数据是数据库系统研究和处理的对象。数据与信息是分不开的, 它们既有联系又有区别。

所谓数据, 通常指用符号记录下来的、可以识别的信息。

信息与数据之间存在着固有的联系, 数据是信息的符号表示 (或称为载体); 信息则是数据的内涵, 是对数据语义的解释。

2. 数据管理技术

数据管理是指数据的收集、整理、分类、组织、存储、查询、检索、维护等操作, 这部分操作是数据处理业务的基本环节, 而且是任何数据处理业务中必不可少的共有部分。

数据处理是指从某些已知的数据出发, 推导加工出一些新的数据, 而这些新的数据又表示了新的信息。

数据处理是与数据管理相联系的, 数据管理技术的优劣将直接影响数据处理的效率。

3. 数据管理技术的发展

计算机数据处理的速度和规模是人工方式或机械方式所无可比拟的, 随着数据处理量的增长, 数据管理技术应运而生。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人工管理阶段、文件系统阶段和数据库系统阶段三个发展阶段。

(1) 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, 数据管理处于人工管理阶段, 数据处

理的方式基本上是批处理。

当时计算机主要用于科学计算，数据量小、结构简单，如高阶方程、曲线拟合等。外存主要为顺序存取设备，如磁带、卡片、纸带，没有磁盘等直接存取设备。没有操作系统，没有数据管理软件，用户用机器指令编码，通过纸带机输入程序和数据，程序运行完毕后，由用户取走纸带和运算结果，再让下一用户上机操作。人工管理阶段的主要特点是没有专用的软件对数据进行管理，由应用程序管理数据；数据面向应用程序，即一组数据对应一个应用程序，数据不能共享；应用程序完全依赖于数据，数据与程序没有独立性；数据不保存在计算机内，数据一般也不需要长期保存。

(2) 文件系统阶段

20世纪50年代后期至60年代中期，数据管理进入文件系统阶段。这个阶段是将数据组织成若干个相互独立的文件，用户通过操作系统对文件进行打开、读写、关闭等操作。

这时计算机不仅用于科学计算，还用于信息管理。外部存储器已有磁盘、磁鼓等直接存取设备。专门管理数据的软件（即文件系统）已经出现，用于文件存储空间的管理，目录管理，文件读写管理，文件保护，向用户提供操作接口等。

文件系统阶段的主要特点是数据以“文件”形式可长期保存在外部存储器的磁盘上；数据的逻辑结构与物理结构有了区别，但比较简单；文件组织已多样化，有索引文件、链接文件和直接存取文件等；数据不再属于某个特定的程序，可以重复使用，即数据面向应用；对数据的操作以记录为单位。

随着数据管理规模的扩大，数据量急剧增加，文件系统显露出数据存在大量冗余（Redundancy），造成了数据不一致（Inconsistency），数据之间的联系也比较弱（Poor Data Relationship）等缺陷。

(3) 数据库系统阶段

20世纪60年代后期以来，为了克服文件系统的弊端，适应日益增长的数据处理的需求，人们开始探索新的数据管理方法和工具。在这一时期，存储技术取得了重要进展，大容量和快速存取的磁盘相继投入市场，为新型数据管理技术的研制提供了良好的物质基础。

由于计算机管理的数据量大，关系复杂，共享性要求强，外存出现了大容量磁盘和光盘。软件价格上升，硬件价格下降，编制和维护软件及应用程序成本相对增加，其中维护的成本更高，急待降低，这些也推动了数据库技术的发展。

数据管理技术进入数据库系统阶段的标志是20世纪60年代末的三个里程碑：1968年，美国IBM公司推出层次模型的IMS系统；1969年，DBTG报告于美国数据系统语言会议（Conference on Data Systems Languages, CODASYL）上发布，总结了当时各式各样的数据库，提出了网状模型；1970年，美国IBM公司的E. F. Codd连续发表论文，提出关系模型，奠定了关系数据库的理论基础。

数据库系统阶段的主要特点是采用数据模型表示复杂的数据结构；有较高的数据独立性；为用户提供了方便的用户接口；提供了数据库的并发控制、数据库的恢复、数据的完整性和数据安全性四方面的数据控制功能；增加了系统的灵活性和安全性。

1.1.3 数据库系统的三个发展阶段

1. 第一代数据库系统

20世纪70年代，以层次型数据库和网状型数据库为代表的第一代数据库系统得到广泛应用。它们基本实现了数据管理中的“集中控制与数据共享”这一目标。



2. 第二代数据库系统

20 世纪 80 年代出现了以关系型数据库为代表的第二代数据库系统。

由于关系数据库具有坚实的理论基础、结构简单、操作方便，且在查询效率和数据库性能等方面取得了突破性的进展，因此在这一时期关系数据库得到了迅猛发展。关系数据库已逐渐成为数据库发展的主流，一批性能不断改善、版本不断更新的商品化关系数据库软件相继投入使用，如 Oracle、Sybase、Informix、Ingres 等关系数据库系统已广泛用于大型信息管理系统。关系数据库系统得到了以下方面的发展：

1) 关系数据库系统的发展促进了数据库系统的小型化。

20 世纪 80 年代推出了一批适于微机环境下运行的关系数据库系统，美国 Ashton-Tate 先后在 1981 年推出微机版关系数据库软件 dBASE II，1984 年升级到 dBASE III，1986 年又推出 dBASE III Plus。1987 年美国 FOX SOFTWARE 公司推出了 FoxBase + (V2.0)，之后又推出 Foxpro 等软件。Oracle 公司在 1986 年也研制出 Professional Oracle 微机版数据库软件。

2) 随着数据库技术的发展和计算机网络的广泛应用，分布式数据库系统在 20 世纪 80 年代得到了很大的发展。

如 1986 年 Oracle 公司和 Ingres 公司先后推出 SQL * Star 和 Ingres/Star 的开放型分布式关系数据库系统。

3) 随着计算机的广泛应用，特别是一些新的应用领域不断提出新的应用要求（如图形图像处理、GIS 空间数据管理、XML 数据管理），关系型数据库、层次型数据库、网状型数据库等传统数据库都表现出不同程度的局限性。因此，人们在 20 世纪 80 年代后期又提出了研制新一代数据库的设想。

3. 第三代数据库系统

20 世纪 80 年代末 90 年代初，新一代数据库技术的研究和开发已成为数据库领域学术界和工业界的研究热点。

例如，人们要求对图形、图像、声音等多媒体数据、时态数据、空间数据、知识信息以及各种复杂对象等非常规数据进行有效处理，为了适应应用需求，人们提出了许多新概念、新思想和新方法以及一些新的数据模型和新数据库管理系统的体系结构，如多媒体数据库、时态数据库、空间数据库、面向对象数据库、分布式数据库、并行数据库系统、数据仓库、移动数据库、XML 数据管理技术等。

1.2 数据库系统的基本概念

1.2.1 数据库

1. 数据库 (Database, DB)

DB 是长期存储在计算机内、有组织的、统一管理的相关数据的集合。DB 能为各种用户共享，具有较小冗余度、数据间联系紧密而又有较高的数据独立性等特点。

2. 数据库管理系统 (Database Management System, DBMS)

DBMS 是位于用户与操作系统 (OS) 之间的一层数据管理软件，它为用户或应用程序提供访问 DB 的方法，包括 DB 的建立、查询、更新及各种数据控制。DBMS 总是基于某种数据模型，可以分为层次型、网状型、关系型和面向对象型等。

1.2.2 数据库系统

数据库系统的简单结构如图 1.1 所示。

图 1.1 中的数据库是数据的汇集，它们经一定的组织形式存储在磁盘介质中。数据库管理系统（DBMS）是管理数据库的软件，它实现数据库系统的各种功能。应用程序是指以数据库为基础的各种应用程序，它必须通过 DBMS 访问数据库。数据库管理员是负责进行数据库规划、设计、协调、维护和管理等工作的人员。由这些应用程序、数据库管理系统、数据库和数据库管理员构成了数据库系统。

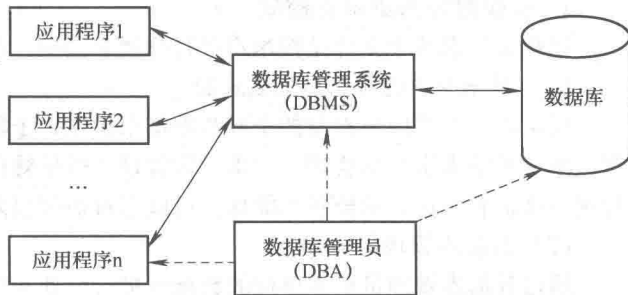


图 1.1 数据库系统的简单结构

从图 1.1 可以看出，DBMS 是数据库系统的核心，一般是由厂家提供的通用软件，如 Microsoft 公司提供的 Microsoft SQL Server 2012、Oracle 公司提供 Oracle 12C 等。DBMS 的功能因产品而异，但现代 DBMS 一般必须具备以下基本功能：

1. 提供高级的用户接口

用户通过 DBMS 看到的是与数据物理存储细节无关的数据逻辑形式，通过 SQL 等非过程数据库语言来对数据库进行操纵。为了适应不同用户的需要，DBMS 通常提供各种接口。

2. 查询处理和优化

这里查询是泛指用户对数据库所提出的访问请求，包括数据修改、数据定义等。由于用户用非过程数据库语言来查询数据库，查询处理过程的拟定和优化都由 DBMS 来完成。查询处理和优化是 DBMS 的基本任务，它对 DBMS 的性能影响颇大。

3. 数据目录管理

由于数据库中保存了持久和共享的数据，数据的定义也长期保存在数据库，这就构成了数据目录。数据目录不但包含数据的逻辑属性，而且包含数据的存储结构定义、数据访问和管理所必需的信息。DBMS 在完成各种功能时都离不开数据目录，数据目录管理是 DBMS 的基本功能。

4. 并发控制

现代 DBMS 一般允许许多用户并发访问数据库，这就不可避免地会产生冲突，例如，对同一数据，一个用户要读，另一个用户要写；若是并发执行，则可能产生不正确的结果。为了解决冲突，DBMS 中需要有并发控制机制。

5. 恢复功能

这要求数据库在发生故障时，甚至遭到破坏时，也要能够恢复到数据一致状态。

6. 完整性约束功能

由于数据库中的数据是持久和共享的，保证其正确性是至关重要的。在 DBMS 中不但要对数据进行语法检查，而且要对数据进行语义检查。数据在语义上的约束称为完整性约束。

例如，每个学生的学号应该是唯一的，年龄不能是负数或大于 100 等，都是完整性约束。

7. 访问控制

DBMS 应有控制用户访问数据库中数据的访问权限控制。例如，对于学生成绩这一数据，允许所有教务人员读，但只允许成绩登录人员写，只允许学生查看自己的课程成绩等。

1.3 数据库系统的特点

1. 实现数据的集中化控制

以往文件系统中文件是按用户进行分散管理的，而数据库系统则能集中控制和管理数据。

(1) 数据库中的数据是集成式的

通常将一个部门所涉及的全部数据都组织在一个数据库中。例如，我们通常利用数据库系统将一个学校的人事档案管理、学生学籍管理、教学管理等各种相关的数据集中在一个数据库中进行统一维护和管理，学校各职能部门可以通过访问权限从数据库中提取相应的所需数据。

(2) 数据库管理员 DBA

通过数据库管理员对本单位的数据库进行管理和维护。

2. 数据的冗余度小，易扩充

通过数据库系统可以将相同的数据在数据库中只存储一次，并为不同的应用共享。这比文件系统的冗余度要小得多。

3. 采用一定的数据模型实现数据结构化

数据库中的数据是按照一定的数据模型来组织、描述和存储的，称为数据结构化。在数据库系统中常用的数据模型有层次模型、网状模型和关系模型等。通常每一个数据库系统只选用一种数据模型，根据选用的数据模型不同，数据库分为层次型数据库、网状型数据库和关系型数据库。

4. 避免了数据的不一致性

数据的不一致性是指数据的不相容性和矛盾性，即数据违反了数据完整性约束条件，使同一数据在数据库内重复出现不同的值。例如，职工工资在工资单上和人事档案中有不同的值，这就是数据的不一致性，主要是由于存在数据冗余所造成的。

在数据库系统中，由于冗余度的大幅降低，避免了大量的数据不一致产生。同时，数据库系统提供了对数据进行控制和检查的功能，从而使得数据在更新的同时也能更新数据的所有副本，以保证数据的一致性。

5. 实现数据共享

在数据库系统中，一组数据集合为多种应用程序和多个用户共同使用。实现数据共享主要体现在当前所有用户（批处理用户、终端用户）可以同时使用数据库；数据库既能满足用户的需求，又能适应新的应用需求，即增加新应用需求且不会影响原有的应用；具有多种用户接口，可以提供与 PL/SQL、C/C++、COBOL、VB 等数据库接口。

6. 提供数据库保护

数据库系统提供了安全性控制、完整性控制、并发控制、故障的检测与恢复等一系列数据控制功能，以确保数据库中的数据是安全可靠的。

7. 数据独立性

数据独立性是指数据库中的数据独立于应用程序，即数据的逻辑结构、存储结构与存取方式的改变不影响应用程序。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，如图 1.2 所示。

(1) 数据的逻辑独立性

数据的逻辑独立性是指数据库整体逻辑结构的改变（如修改数据定义、增加新的数据类型、改变数据间的联系等）不需要修改应用程序。

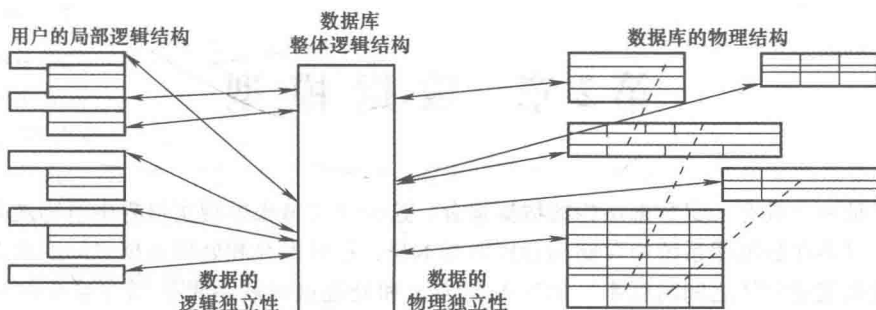


图 1.2 数据独立性示意

(2) 数据的物理独立性

数据的物理独立性是指数据的物理结构（存储结构、存取方式等）的改变（如存储设备的更换、物理存储格式和存取方式的改变等）不影响数据库的逻辑结构，因而不会引起应用程序的变化。

8. 数据由 DBMS 统一管理和控制

数据库中的多用户并发共享数据须由 DBMS 进行统一管理和控制。

(1) 安全性保护

安全性保护是指保护数据以防因不合法的使用而造成数据的泄密和破坏。

(2) 完整性保护

完整性保护是指将数据控制在有效范围内（如年龄在 1~200）或使数据之间满足一定的关系（外键值为引用实体的关键字值或为空值），以保证数据的正确性、有效性和相容性。

(3) 并发控制

当多个用户进程同时对数据库进行并发存取时，DBMS 必须对多用户的并发操作加以控制和协调。

(4) 数据库恢复

当数据库系统发生故障时，DBMS 必须具备将数据库从错误状态恢复到某个正确状态的功能。

第2章 数据模型

数据库是一个具有一定数据结构的数据集合，这个结构是根据现实世界中事物之间的联系来确定的。不仅要在数据库系统中存储和管理数据本身，还要保存和处理数据之间的联系。这种数据之间的联系就是实体之间的联系。研究如何表示和处理这种联系是数据库系统的一个核心问题，用以表示实体以及实体之间联系的数据库的数据结构称为数据模型。本章将着重介绍一下概念模型、层次模型、网状模型、关系模型、面向对象模型等数据库系统的数据模型的基本概念和设计方法，为后面的数据库设计打下基础。

2.1 数据模型概述

数据模型 (Data Model) 是对现实世界数据特征的抽象，是用来描述数据的一组概念和定义。

为了把现实世界的具体事物抽象、组织为某一 DBMS 支持的数据模型，通常先把现实世界中的客观对象抽象为概念模型，然后把概念模型转换为某一 DBMS 支持的数据模型，这一过程如图 2.1 所示。

数据模型按不同的应用层次可划分为以下两类：

1. 概念数据模型

概念数据模型 (又称概念模型) 是一种面向客观世界、面向用户、独立于计算机系统的数据模型，完全不涉及信息在计算机中的表示，只是用来描述某个特定组织所关心的信息结构。概念模型是按用户的观点对数据建模，是用户和数据设计人员之间进行交流的工具，主要用于数据库的设计。例如 E-R 模型、扩充 E-R 模型均属于这类模型。

2. 逻辑数据模型

逻辑数据模型 (又称数据模型) 是一种直接面向数据库系统的模型，主要用于 DBMS 的实现。例如，层次模型、网状模型、关系模型，它们均属于这类模型。这类模型有严格的形式化定义，以便在计算机系统中实现。

2.1.1 数据模型的基本组成

数据模型是现实世界中的事物及其之间联系的一种抽象表示，是一种形式化描述数据、数据间联系以及有关语义约束规则的方法。通常，一个数据库的数据模型由数据结构、数据操作和数据约束条件三个部分组成。

1. 数据结构

数据结构是指对实体类型和现实间联系的表达。它是数据模型最基本的组织部分，规定了数据模型的静态特性。在数据库系统中，通常按照数据结构的类型来命名数据模型，例如，采用层次型数据结构、网状型数据结构、关系型数据结构的数据模型分别称为层次模型、网状模型和关系模型。

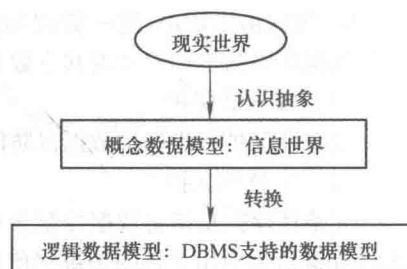


图 2.1 客观事物的抽象过程