



高等学校计算机规划教材

数据库原理与应用

——基础·开发技术·实践

■ 刘玉宝 主编 ■ 徐大伟 戴银飞 副主编

电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

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

高等学校计算机规划教材

数据库原理与应用

——基础·开发技术·实践

刘玉宝 主编

徐大伟 戴银飞 副主编

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书详细介绍了数据库原理、方法及其应用开发技术。介绍了数据库系统基础，关系数据库理论，关系数据库标准语言 SQL，SQL Server 2005 应用基础，SQL 高级应用，数据库设计，数据库保护技术，数据库访问技术，C 语言数据库应用程序开发技术，C#和 ADO.NET 数据库应用程序开发技术，Java 数据库应用程序开发技术，以及数据库新技术等内容。同时介绍了使用 C、C#和 Java 三种语言开发数据库应用程序的基本方法和技术，使具有不同语言基础的读者有选择性地学习，同时 C 语言开发数据库程序的技术还为嵌入式系统相关专业读者开发嵌入式软件打下良好的基础。附录提供本课程的实验指导、课程设计指导等实践环节。书中配有实例、习题，有利于教师教学和学生自学。为方便教师教学，本书配有教学课件和书中实例源代码。

本书可作为普通高等学校计算机及信息专业的本、专科生的教材，也可作为高职高专院校在校生的教材，同时也适合从事数据库应用程序开发人员参考之用。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容
版权所有，侵权必究

图书在版编目(CIP)数据

数据库原理与应用：基础·开发技术·实践/刘玉宝主编. —北京：电子工业出版社，2010.9
高等学校计算机规划教材
ISBN 978-7-121-11658-2

I. ①数… II. ①刘… 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(2010)第 162751 号

策划编辑：史鹏举

责任编辑：史鹏举

特约编辑：王 崧

印 刷：

北京市李史山胶印厂

装 订：

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本：787×1092 1/16 印张：21 字数：605 千字

印 次：2010 年 9 月第 1 次印刷

定 价：33.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010) 88254888。

质量投诉请发邮件至 zltz@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：(010) 88258888。

前 言

数据库技术产生于 20 世纪 60 年代末。经过 40 多年的迅猛发展，已经形成了完整的理论与技术体系，并已成为计算机科学与技术中的一个重要分支。随着信息技术的迅猛发展，数据库技术已经成为国家信息基础设施和信息化社会中的最重要的支撑技术之一。

伴随着数据库技术在国民经济、科技和文化等各个领域的广泛应用，数据库及其设计技术已经受到了各行各业人员的普遍关注。数据库课程已成为高等院校计算机科学与技术、信息工程、管理工程、信息管理与信息系统等专业的核心专业课程，并有越来越多的专业及相关技术人员提出了对数据库知识的需求。

本书共分为 12 章。

第 1 章数据库系统基础，主要介绍数据库的基本概念，数据库技术的发展，数据库系统的组成与结构，数据模型的概念与分类。

第 2 章关系数据库理论，主要介绍关系模型，关系数据结构及形式化定义，关系代数及关系规范化理论等。

第 3 章关系数据库标准语言 SQL，主要介绍了 SQL 语言的功能与特点，表的基本操作，基本查询，视图操作，子查询，组合查询及数据的插入、修改、删除等。

第 4 章 SQL Server 2005 应用基础，主要介绍 SQL Server 2005 的基本知识，安装方法及常见问题，数据库、基本表及索引的管理。

第 5 章 SQL 高级应用，主要介绍 Transact_SQL 语言程序设计基础，存储过程和触发器的基本应用。

第 6 章数据库设计，主要介绍数据库系统设计的内容和特点，数据库设计的步骤，需求分析，概念结构设计，逻辑结构设计，物理结构设计及数据库的实施和维护等。

第 7 章数据库保护技术，主要介绍事务机制，数据库安全性，数据库完整性，数据库的恢复，并发控制等。

第 8 章数据库访问技术，主要介绍 ODBC 工作原理及使用方法，ADO 模型的层次结构，使用 ADO 技术访问数据库的方法，ADO.Net 的体系结构的组成及工作原理，JDBC 数据库访问技术。

第 9 章 C 语言数据库应用程序开发技术，主要介绍 C 语言嵌入式 SQL 程序开发环境搭建，嵌入式 SQL 语句中使用的 C 变量，数据库的连接、查询和更新，SQL 通信区，游标的使用，SQLDA。

第 10 章 C#和 ADO.NET 数据库应用程序开发技术，主要介绍数据提供程序的选择，数据提供程序的选择，SqlConnection 的使用，OleDbConnection 的使用，OracleConnection 的使用，数据的获取，DataReader 的使用，DataSet 和 DataAdapter 的使用等。

第 11 章 Java 数据库应用程序开发技术，主要介绍 JDBC API 介绍，SQL 和 Java 数据类型的映射关系，Java 数据库操作的基本步骤，使用 JDBC 实现对数据库的操作，JDBC 连接其他类型的数据库。

第 12 章数据库新技术，主要介绍分布式数据库的概念、特点和体系结构，面向对象数据库的理论和实现方法，数据仓库技术及数据挖掘技术等。本书内容覆盖了关系数据库系统的原理、设计和应用技术。

本书主要特点:

(1) 以关系数据库系统为核心。在系统论述数据库基本知识的基础上,着重讨论了关系数据库的原理与实现,其中对关系数据模型、关系数据库体系结构、关系规范化理论等都有较详细、系统的说明。

(2) 教材对传统数据库的内容进行了精简,如对层次数据库、网状数据库仅对其模型做了简要介绍,删除了一些与操作系统联系较密切的存储理论等。

(3) 为了反映当前数据库领域的新技术、新水平和新趋势。本教材介绍了分布式数据库系统、面向对象数据库、数据仓库和数据挖掘等内容,力求反映当前数据库技术的发展。

(4) 注重理论联系实际,加强数据库应用开发技术的介绍。教材在数据库语言(SQL)等数据库应用技术方面进行了较为全面的论述,并结合一些实例较详细地讲解了数据库设计方法,为读者进行数据库应用程序的开发提供了较扎实的基础。结合 SQL Server 2005 的具体的数据库管理系统,讲解了数据库一些管理技术的应用,使读者在学习理论的同时有了具体的应用,也为读者维护管理大中型数据库系统打下基础。本教材还介绍了 C、C#和 Java 语言开发数据库应用程序的基本技术,为读者在选择开发语言上提供了更广阔的空间。

(5) 在内容选取、章节安排、难易程度、例子选取等方面充分考虑到理论教学和实践教学的需要,力求使教材概念准确、清晰,重点明确,内容广泛,便于取舍,每章均配有习题便于教学。

全书内容丰富,结构合理,实用性强。其中第 7、8、9 章的内容可以根据实际情况进行取舍,针对读者现有的语言基础选择相应的章节学习。

本书配有电子课件、完整的开发案例源代码、习题参考答案等教学资源,需要者可从华信教育资源网 <http://www.hxedu.com.cn> 免费注册下载。

本书由刘玉宝担任主编,徐大伟、戴银飞担任副主编,参加编写的人员还有赵耀红、杨丽萍、祝海英、王薇。第 1、3 章由戴银飞编写,第 2、6 章由赵耀红编写,第 4、7 章由杨丽萍编写,第 5、8、12 章由徐大伟编写,第 9 章、附录 B 由刘玉宝编写,第 10、11 章由祝海英编写,附录 A 由王薇编写,最后由刘玉宝统一定稿。

在本书编写的过程中得到了单位领导和同仁的热情帮助和支持,在此表示衷心的感谢!

本书的编写参考了广大同行专家的著作和成果,在此对他们表示衷心的感谢!

由于作者水平有限,书中难免有不当之处,敬请广大读者和专家批评指正。

咨询、意见和建议,可反馈至本书责任编辑邮箱: shipj@phei.com.cn。

编者

2010年6月

目 录

第 1 章 数据库系统概述	(1)
1.1 数据库系统的基本概念	(1)
1.1.1 数据管理与数据处理	(1)
1.1.2 数据库	(2)
1.1.3 数据库管理系统	(2)
1.1.4 数据库系统	(2)
1.2 数据描述与数据模型	(4)
1.2.1 数据的 3 种范畴	(4)
1.2.2 信息世界中所涉及的基本概念	(4)
1.2.3 机器世界中所涉及的基本概念	(5)
1.2.4 实体间的联系	(5)
1.3 概念数据模型与 E-R 表示方法	(6)
1.3.1 数据模型	(6)
1.3.2 概念数据模型	(7)
1.3.3 概念数据模型的 E-R 表示方法	(8)
1.3.4 概念数据模型实例	(8)
1.4 传统数据模型概述	(10)
1.4.1 层次模型	(10)
1.4.2 网状模型	(11)
1.4.3 关系模型	(13)
1.5 数据库系统结构	(13)
1.5.1 数据库系统模式的概念	(13)
1.5.2 数据库系统的三级模式结构	(14)
1.5.3 数据独立性	(14)
1.6 数据库管理系统的组成	(15)
1.7 数据库系统的组成	(16)
1.8 小结	(16)
习题 1	(17)
第 2 章 关系数据库理论	(21)
2.1 关系模型	(21)
2.1.1 关系模型的基本概念	(21)
2.1.2 关系模式	(23)
2.2 关系模型的完整性	(23)
2.3 关系代数	(24)
2.3.1 关系代数概述	(24)
2.3.2 关系代数的基本操作	(24)
2.3.3 关系代数的其他操作	(27)

2.4	关系数据库规范化理论	(28)
2.4.1	关系规范化理论概述	(28)
2.4.2	函数依赖	(30)
2.4.3	关系的规范化	(32)
2.4.4	多值依赖与第四范式	(34)
2.4.5	各种范式之间的关系	(35)
2.5	小结	(36)
	习题 2	(36)
第 3 章	关系数据库语言 SQL	(41)
3.1	SQL 语言的功能与特点	(41)
3.2	表的基本操作	(41)
3.2.1	定义表	(41)
3.2.2	修改表	(44)
3.2.3	删除基本表	(45)
3.3	SQL 的数据查询	(45)
3.3.1	单表查询	(46)
3.3.2	多表查询	(52)
3.4	SQL 的视图操作	(54)
3.4.1	定义视图	(54)
3.4.2	创建视图	(54)
3.4.3	使用视图	(55)
3.4.4	删除视图	(56)
3.5	子查询	(56)
3.6	组合查询	(58)
3.7	数据的插入、修改与删除	(59)
3.7.1	插入数据	(59)
3.7.2	修改数据	(60)
3.7.3	删除数据	(60)
3.8	小结	(61)
	习题 3	(61)
第 4 章	SQL Server 2005 应用基础	(66)
4.1	SQL Server 2005 系统概述	(66)
4.2	SQL Server 2005 版本说明	(67)
4.3	SQL Server 2005 Express Edition 简介	(68)
4.4	SQL Server 2005 安装与配置	(70)
4.4.1	安装时考虑的关键点	(70)
4.4.2	SQL Server 2005 Express Edition 安装	(71)
4.4.3	SQL Server 2005 组件	(77)
4.5	常见故障分析	(80)
4.6	数据库的创建	(80)
4.6.1	操作系统文件	(80)
4.6.2	数据库文件组	(81)

4.6.3	使用数据文件和文件组的建议	(81)
4.6.4	创建数据库	(81)
4.6.5	修改数据库	(84)
4.6.6	删除数据库	(85)
4.7	基本表的定义	(85)
4.7.1	创建基本表	(85)
4.7.2	修改基本表	(89)
4.7.3	删除基本表	(91)
4.8	索引的建立和删除	(91)
4.8.1	索引的概念	(91)
4.8.2	索引的类型	(91)
4.8.3	建立索引	(92)
4.8.4	删除索引	(92)
4.9	小结	(92)
	习题 4	(93)
第 5 章	SQL 高级应用	(94)
5.1	Transact_SQL	(94)
5.1.1	Transact_SQL 简介	(94)
5.1.2	Transact_SQL 语法格式	(94)
5.1.3	Transact_SQL 系统元素	(96)
5.2	Transact_SQL 程序流程控制	(99)
5.2.1	IF...ELSE 语句	(99)
5.2.2	BEGIN...END 语句	(100)
5.2.3	GOTO 语句	(101)
5.2.4	WHILE、BREAK、CONTINUE 语句	(101)
5.2.5	WAITFOR 语句	(101)
5.2.6	RETURN 语句	(102)
5.2.7	CASE 表达式	(102)
5.3	存储过程	(103)
5.3.1	存储过程的概念	(103)
5.3.2	存储过程的创建与执行	(104)
5.3.3	存储过程与参数	(108)
5.4	触发器	(110)
5.4.1	触发器的概念与工作原理	(110)
5.4.2	创建触发器	(111)
5.4.3	管理触发器	(114)
5.4.4	触发器的用途	(115)
5.5	小结	(115)
	习题 5	(115)
第 6 章	数据库设计	(116)
6.1	数据库设计的内容与特点	(116)
6.2	数据库设计方法	(116)

6.3	数据库设计步骤	(117)
6.4	数据库规划	(118)
6.5	需求分析	(118)
6.5.1	需求分析的任务	(118)
6.5.2	需求分析的方法	(119)
6.5.3	需求分析的步骤	(119)
6.6	概念结构设计	(122)
6.6.1	设计各局部应用的 E-R 模型	(122)
6.6.2	全局 E-R 模型的设计	(122)
6.7	逻辑结构设计	(124)
6.7.1	逻辑结构设计的步骤	(124)
6.7.2	E-R 图向关系模型的转换	(124)
6.7.3	逻辑模式的优化	(125)
6.7.4	外模式的设计	(126)
6.8	物理结构设计	(127)
6.8.1	数据库物理结构设计的内容与方法	(127)
6.8.2	关系模式存取方法的选择	(127)
6.8.3	系统存储结构的确定	(129)
6.9	数据库的实施	(130)
6.9.1	数据库试运行	(131)
6.9.2	数据库的运行与维护	(131)
6.10	数据库应用的结构和开发环境	(132)
6.10.1	数据库应用模型	(132)
6.10.2	数据库应用开发环境 ODBC	(134)
6.11	小结	(137)
	习题 6	(137)
第 7 章	数据库保护技术	(140)
7.1	事务机制	(140)
7.1.1	事务的概念与特性	(140)
7.1.2	事务的提交与回退	(141)
7.2	数据库安全性	(142)
7.2.1	对数据库安全的威胁	(142)
7.2.2	数据库安全性控制	(143)
7.2.3	视图机制	(144)
7.2.4	数据加密	(144)
7.2.5	SQL Server 2005 的安全性	(144)
7.3	数据库完整性	(147)
7.3.1	数据库完整性概述	(147)
7.3.2	完整性控制	(148)
7.3.3	数据完整性的实现	(150)
7.4	数据库恢复	(153)
7.4.1	数据库的故障分类	(153)
7.4.2	数据库故障的基本恢复方式	(154)

7.4.3	恢复策略	(156)
7.4.4	具有检查点的恢复技术	(158)
7.4.5	SQL Server 2005 备份与还原	(159)
7.5	并发控制	(163)
7.5.1	数据库并发操作带来的数据不一致性问题	(163)
7.5.2	封锁技术	(165)
7.5.3	锁协议	(165)
7.5.4	封锁带来的问题——活锁与死锁	(166)
7.5.5	并发调度的可串行性	(168)
7.5.6	SQL Server 的并发控制	(170)
7.6	小结	(171)
	习题 7	(172)
第 8 章	数据库访问技术	(176)
8.1	ODBC 的使用	(176)
8.1.1	ODBC 概述	(176)
8.1.2	ODBC 数据源的配置	(177)
8.2	ADO 的使用	(179)
8.2.1	ADO 概述	(179)
8.2.2	使用 ADO 技术访问数据库举例	(181)
8.3	ADO.NET 简介	(184)
8.3.1	ADO.NET 技术的设计目标	(184)
8.3.2	ADO.NET 的体系结构	(184)
8.3.3	ADO.NET 数据对象	(190)
8.4	JDBC 技术	(191)
8.4.1	JDBC 概述	(191)
8.4.2	JDBC 驱动程序	(192)
8.4.3	JDBC 常用类	(193)
8.5	小结	(194)
	习题 8	(194)
第 9 章	C 语言数据库应用程序开发	(195)
9.1	嵌入式 SQL 语句	(195)
9.1.1	在 C 语言程序中嵌入 SQL 语句的程序开发环境的搭建	(196)
9.1.2	第一个在 C 语言程序中嵌入 SQL 语句的程序	(197)
9.2	静态 SQL 语句	(199)
9.2.1	声明嵌入式 SQL 语句中使用的 C 变量	(199)
9.2.2	连接数据库	(201)
9.2.3	数据的查询与更新	(202)
9.2.4	SQL 通信区	(203)
9.3	动态 SQL 语句	(205)
9.3.1	动态修改	(205)
9.3.2	动态游标	(206)
9.3.3	SQLDA	(208)

9.4	C 语言数据库应用程序开发实例	(209)
9.4.1	需求说明	(209)
9.4.2	数据库结构设计	(209)
9.4.3	数据库行为设计	(210)
9.4.4	系统实现	(210)
	习题 9	(215)
第 10 章	C#和 ADO.NET 数据库应用程序开发	(216)
10.1	数据库的连接	(216)
10.1.1	SqlConnection 的使用	(217)
10.1.2	OleDbConnection 的使用	(218)
10.1.3	OdbcConnection 的使用	(219)
10.1.4	OracleConnection 的使用	(220)
10.2	数据的获取	(223)
10.2.1	创建 Command 对象	(223)
10.2.2	执行命令	(223)
10.2.3	参数化查询	(227)
10.2.4	执行存储过程	(229)
10.3	DataReader 的使用	(231)
10.3.1	DataReader 简介	(231)
10.3.2	使用 DataReader 读取数据	(232)
10.3.3	在 DataReader 中使用多个结果集	(234)
10.4	DataSet 和 DataAdapter 的使用	(236)
10.4.1	DataSet 简介	(236)
10.4.2	DataAdapter 简介	(236)
10.4.3	利用 DataSet 和 DataAdapter 访问数据	(237)
10.4.4	类型和无类型 DataSet	(239)
10.5	性能	(240)
10.6	C#数据库应用程序开发实例	(240)
10.6.1	需求说明	(241)
10.6.2	数据库结构设计	(241)
10.6.3	数据库行为设计	(242)
10.6.4	系统实现	(242)
	习题 10	(244)
第 11 章	Java 数据库应用程序开发	(245)
11.1	JDBC API 简介	(245)
11.2	SQL 和 Java 之间的映射关系	(246)
11.3	JDBC 编程	(247)
11.3.1	数据库操作基本步骤	(248)
11.3.2	JDBC 数据库操作实现	(249)
11.4	连接其他类型数据库	(267)
11.4.1	连接 Oracle 数据库	(267)
11.4.2	连接 MySQL 数据库	(268)

11.4.3	连接 SQL Server 数据库	(269)
11.4.4	JDBC:ODBC 连接 Access 数据库	(271)
11.5	Java 数据库应用程序开发实例	(272)
11.5.1	需求说明	(272)
11.5.2	数据库结构设计	(273)
11.5.3	数据库行为设计	(273)
11.5.4	系统实现	(274)
习题 11	(279)
第 12 章	数据库新技术	(282)
12.1	面向对象数据库系统	(282)
12.1.1	对象-关系数据库	(284)
12.1.2	面向对象数据库与传统数据库的比较	(284)
12.2	分布式数据库系统	(286)
12.2.1	分布式数据库系统概述	(286)
12.2.2	分布式数据库系统的设计	(288)
12.2.3	分布式数据库系统的安全技术	(289)
12.2.4	分布式数据库系统的发展前景与应用趋势	(289)
12.3	数据仓库	(291)
12.3.1	数据仓库的定义与特点	(291)
12.3.2	数据仓库的种类	(292)
12.3.3	数据集市	(292)
12.3.4	数据仓库中的几个数据	(293)
12.3.5	数据仓库模型	(294)
12.4	数据挖掘	(295)
12.4.1	数据挖掘的定义	(295)
12.4.2	数据挖掘的分类	(295)
12.4.3	数据挖掘的数据来源	(295)
12.4.4	数据挖掘的体系结构	(296)
12.4.5	数据挖掘的步骤	(296)
12.4.6	数据挖掘的功能	(296)
12.4.7	数据挖掘的常用技术	(296)
12.5	数据库技术的新应用	(297)
12.5.1	数据模型研究	(297)
12.5.2	与新技术结合的研究	(297)
12.5.3	与应用领域结合的研究	(300)
12.6	小结	(302)
习题 12	(302)
附录 A	实验部分	(303)
附录 B	课程设计指导书	(317)

第1章 数据库系统概述

本章从数据库管理技术的产生和发展引出数据库的概念，围绕着数据库系统介绍有关名词术语。通过本章学习，将了解以下内容：

- 📖 数据库系统的基本概念
- 📖 数据模型
- 📖 概念数据模型及表示
- 📖 传统数据模型
- 📖 数据库系统结构
- 📖 数据库管理系统的组成
- 📖 数据库系统的组成

数据库是计算机科学的重要分支。当今，信息资源已成为各个部门的重要财富和资源。建立一个满足各级部门信息处理要求的、行之有效的信息系统已成为一个企业或组织生存和发展的重要条件。目前，基于数据库技术的计算机应用已成为计算机应用的主流。它使计算机应用渗透到各部门。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。

1.1 数据库系统的基本概念

1.1.1 数据管理与数据处理

数据(Data)是描述事物的符号记录，是数据库中存储的基本对象。数据在大多数人头脑中的第一个反应就是数字。其实数字只是一种简单的数据，是对数据传统、狭义的理解。从广义上理解，数据的种类很多，文字、图形、图像、声音、语言、学生的档案记录、货物的运输情况等都是数据，即

数据 = 量化特征描述 + 非量化特征描述

例如，对天气预报中的温度的高低可以进行量化表示，而“刮风”、“下雨”等特征则需要用文字或图形符号进行描述，它们都是数据，只是数据类型不同而已。自然界的任何事物都可以通过记录的形式进行描述。

(1) 人：(王一，男，21，1988，吉林)

(2) 学生：(王一，男，21，1988，吉林，计算机系，计算机应用专业)

数据形式本身并不能完全表达其内容，需要经过数据语义解释。数据与其语义是不可分的。例如：

(1) (王一，78) 可以赋予它一定的语义，它表示王一的期末考试平均成绩为78分。如果不了解其语义，则无法对其进行解释，甚至解释为王一的年龄为78。

(2) 99: 8179, 7954, 521 舅舅：不要吃酒，吃酒误事，我爱你

- | | |
|--|------|
| (3) $1 \times 1 = 1$ | 一成不变 |
| (4) $1000^2 = 100 \times 100 \times 100$ | 千方百计 |
| (5) $7/8$ | 七上八下 |
| (6) $7 \div 2$ | 不三不四 |

1.1.2 数据库

数据库(Database, DB)是存放数据的仓库。只不过这个仓库位于计算机存储设备上,而且数据是按一定的格式存放的。数据是描述自然界事物特征的符号,而且能够被计算机处理。对数据进行存储的目的是为了从大量的数据中发现有价值的信息,这些有价值的信息就是“信息”。

数据库是长期存储在计算机内的、有组织的、可共享的数据集合。数据库中的数据按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,并为各种用户所共享,数据库本身不是独立存在的,它是组成数据库系统的一部分,在实际应用中,人们面对的是数据库系统(Database System, DBS)。

1.1.3 数据库管理系统

数据库管理系统(Database Management System, DBMS)是一个系统软件,是数据库系统的一个重要组成部分,位于用户与操作系统之间。它的任务是科学地组织和存储数据,高效地获取和维护数据。DBMS负责对数据库的建立、运用和维护进行统一管理和控制,使用户能方便地定义数据和操纵数据,并能够保证数据的安全性、完整性,在多个用户同时使用数据库时进行并发控制,在发生故障后对系统进行恢复。它的主要功能有如下几个:

- (1) 数据定义。
- (2) 数据操纵。
- (3) 数据库运行管理。
- (4) 数据组织、存储和管理。
- (5) 数据库建立和维护。
- (6) 数据通信接口。

1.1.4 数据库系统

数据库系统是指在计算机系统中引入数据库后的系统构成,一般由数据库、数据库管理系统(及开发工具)、应用系统、数据库管理员和用户构成。其中数据库管理员(Database Administrator, DBA)是负责数据库的建立、使用和维护等工作的专门人员。

1. 数据库的基本特征

数据库是相互关联的数据的集合。数据库中的数据不是孤立的,数据和数据之间是相互关联的,也就是说,在数据库中不仅要能够表示数据本身,还要能够表示数据与数据之间的关系。

数据库有以下几个基本特征。

- (1) 数据库具有较高的数据独立性。
- (2) 数据库用综合的方法组织数据,保证尽可能高的访问效率。
- (3) 数据库具有较小的数据冗余,可供多个用户共享。

(4) 数据库具有安全控制机制,能够保证数据的安全、可靠。

(5) 数据允许多用户共享,数据库能有效、及时地处理数据,并能保证数据的一致性和完整性。

2. 数据管理技术的发展

如同其他科学技术的发展一样,数据管理技术也有一个发展的历程,大体上经历了3个阶段。

(1) 人工管理阶段(20世纪50年代中期以前)。这一阶段计算机主要用于科学计算。硬件中的外存只有卡片、纸带、磁带,没有磁盘等直接存取设备。软件只有汇编语言,没有操作系统和管理数据的软件。数据处理的方式基本上是批处理。

人工管理数据具有以下特点。

- ① 数据不保存。
- ② 应用程序管理数据。
- ③ 数据不共享。
- ④ 数据不具有独立性。

(2) 文件系统阶段(20世纪50年代后期至60年代中后期)。计算机不仅用于科学计算,而且还逐渐扩大到非计算领域,如用于管理。硬件方面:已经有磁盘、磁鼓等直接存取存储设备,磁盘已经成为联机应用的主要存储设备。软件方面:有了操作系统和高级语言,而且有了专门的数据管理软件,也就是文件管理系统(或操作系统的文件管理部分),处理方式不仅有了文件批处理,而且能够进行联机实时处理。

文件系统管理数据的优点有如下3个。

- ① 数据可以长期保存。
- ② 有专门的软件即文件系统用于管理数据。
- ③ 文件的形式多样化。

文件系统管理数据的缺点也有如下3个。

- ① 数据共享性差,冗余度大。
- ② 数据独立性差。
- ③ 数据联系弱。文件与文件之间是独立的,文件之间的联系必须通过程序来构造,可见,文件是一个不具有弹性的、无结构的数据集合,不能反映现实世界事务之间的内在联系。

文件管理系统示例如图1-1所示。

(3) 数据库系统阶段(20世纪60年代后期以来)。20世纪60年代后期,在硬件方面出现了大容量的磁盘,价格下降,在软件方面出现了数据库管理系统。在数据库系统阶段使用数据库技术来管理数据。它克服了文件系统的不足,并增加了许多新功能。在这一阶段,数据由数据库管理系统统一控制,数据不再面向某个应用而是面向整个系统,因此数据可以被多个用户、多个应用共享。

数据库系统的特点如下。

- ① 数据结构化,这是数据库与文件系统的根本区别。
- ② 由DBMS提供统一的管理控制功能(安全性、完整性、并发控制、数据库恢复)。
- ③ 数据的共享性好。
- ④ 数据的独立性高。
- ⑤ 可控数据冗余度低。

数据库管理系统示例如图1-2所示。

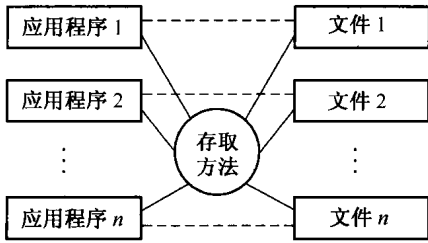


图 1-1 文件管理系统示例

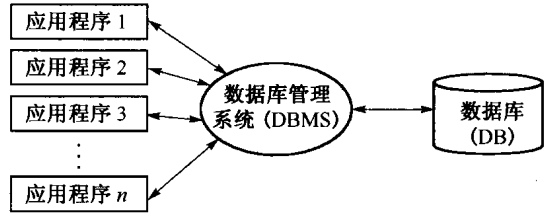


图 1-2 数据库管理系统示例

1.2 数据描述与数据模型

1.2.1 数据的 3 种范畴

数据不能直接从现实世界存放到数据库中，它需要经过人们的认识、理解、整理、规范和加工，也就是说，数据从现实世界进入到数据库中经历了 3 个层次，即现实世界、信息世界和机器世界，称为数据的 3 种范畴。

1. 现实世界

现实世界也称为客观世界。人们头脑之外的客观事物及其相互联系就在这个世界中。现实世界中所有客观存在的事物及其相互之间的联系只是处理对象最原始的表现形式。

2. 信息世界

信息世界又称为观念世界，是现实世界在人们头脑中的反映，或者说，在信息世界中所存在的信息是现实世界中的客观事物在人们头脑中的反映，并经过一定的选择、命名和分类。

3. 机器世界

当信息进入计算机后，即进入机器世界范畴或存储世界范畴。其中机器世界也称为数据世界。

由于计算机只能处理数据化的信息，所以对信息世界中的信息必须进行数据化。信息经过加工、编码后即进入数据世界，由计算机来进行处理。因此，数据世界中的对象是数据。现实世界中的客观事物及其联系在数据世界中是用数据模型来描述的。

数据化后的信息称为数据，所以说数据是信息的符号表示。

1.2.2 信息世界中所涉及的基本概念

(1) 实体(Entity)。实体是客观存在的事物在人们头脑中的反映，或者说，客观存在并可相互区别的客观事物或抽象事件称为实体。实体可以指人，如一名教师、一名护士等；也可以指物，如一把椅子、仓库、一个杯子等。实体不仅可以指实际的事物，还可以指抽象的事物，如一次访问、一次郊游、订货、演出、足球赛等；甚至还可以指事物与事物之间的联系，如“学生选课记录”和“教师任课记录”等。

(2) 属性(Attribute)。在观念世界中，属性是一个很重要的概念。所谓属性是指实体所具有的某一方面的特性。一个实体可由若干属性来刻画。例如，教师的属性有姓名、年龄、性别、职称等。

属性所取的具体值称为属性值。例如,某一教师的姓名为李辉,这是教师属性“姓名”的取值;该教师的年龄为45,这是教师属性“年龄”的取值,等等。

(3) 域(Domain)。一个属性可能取的所有属性值的范围称为该属性的域。例如,教师属性“性别”的域为男、女;教师属性“职称”的域为助教、讲师、副教授、教授等。

由此可见,每个属性都是一个变量,属性值就是变量所取的值,而域则是变量的变化范围。因此,属性是表征实体的最基本的信息。

(4) 码(Key)。唯一标识实体的属性集称为码。例如,学号是学生实体的码。

(5) 实体型(Entity Type)。具有相同属性的实体必然具有共同的特性和性质。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,教师(姓名,年龄,性别,职称)就是一个实体型。

(6) 实体集(Entity Set)。同一类型实体的集合。例如,某一学校中的教师具有相同的属性,他们就构成了实体集“教师”。

在信息世界中,一般就用上述这些概念来描述各种客观事物及其相互的区别与联系。

1.2.3 机器世界中所涉及的基本概念

与信息世界中的基本概念对应,在数据世界中也涉及一些相关的基本概念。

(1) 数据项(字段, Field): 对应于信息世界中的属性。例如,在实体型“教师”中的各个属性中,姓名、性别、年龄、职称等就是数据项。

(2) 记录(Record): 每个实体所对应的数据。例如,对应于某一教师的各项属性值李辉、45、男、副教授等就构成一条记录。

(3) 记录型(Record Type): 对应于信息世界中的实体型。

(4) 文件(File): 对应于信息世界中的实体集。

(5) 关键字(Key)。能够唯一标识一个记录的字段集。

1.2.4 实体间的联系

在现实世界中,事物内部及事物之间是有联系的,这些联系在信息世界中反映为实体(型)内部的联系和实体(型)之间的联系。实体内部的联系通常是指组成实体的各属性之间的联系。实体之间的联系通常是指不同实体集之间的联系。

两个实体型之间的联系可以分为以下3类。

1. 一对一联系(1:1)

如果对于实体集 A 中的每一个实体,实体集 B 中至多有一个(也可以没有)实体与之联系,反之亦然,则称实体集 A 与实体集 B 具有一对一联系,记为 $1:1$,用图 1-3 表示。

例如,实体集“学院”与实体集“院长”之间的联系就是 $1:1$ 的联系。因为一个院长只领导一个学院,而且一个学院也只有一个院长。再如学校里,实体集“班级”与实体集“班长”之间也具有 $1:1$ 联系,一个班级只有一个班长,而一个班长只在一个班中任职。

2. 一对多联系(1:n)

如果对于实体集 A 中的每一个实体,实体集 B 中有 n 个($n \geq 0$) 实体与之联系,反之,对于实体集 B 中的每一个实体,实体集 A 中至多有一个实体与之联系,则称实体集 A 与实体集 B 具有一对多联系,记为 $1:n$,用图 1-4 表示。

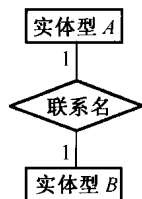


图 1-3 1:1 联系