

21世纪高等学校计算机规划教材

21st Century University Planned Textbooks of Computer Science

数据库基础 与应用 (第2版)

Database Foundation
and Applications (2nd Edition)

王珊 李盛恩 编著

- 突出应用, 内容丰富
- 重视理论, 难易得当
- 关注发展, 技术新颖



名家系列

 人民邮电出版社
POSTS & TELECOM PRESS

21世纪高等学校计算机规划教材

21st Century University Planned Textbooks of Computer Science

数据库基础 与应用 (第2版)

Database Foundation
and Applications (2nd Edition)

王珊 李盛恩 编著



名家系列

人民邮电出版社

北京

图书在版编目 (CIP) 数据

数据库基础与应用 / 王珊, 李盛恩编著. —2版. —北京:
人民邮电出版社, 2009.6
21世纪高等学校计算机规划教材
ISBN 978-7-115-20508-7

I. 数… II. ①王…②李… 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 (2009) 第050120号

内 容 提 要

本书侧重于数据库系统的应用, 重点介绍了开发关系数据库系统必备的基本知识和基本方法, 包括数据库系统的基本概念、基本技术及数据库应用开发技术, 数据仓库和联机分析新技术及新应用等。全书内容丰富, 系统性强, 知识体系新颖, 理论与实践结合, 具有先进性和实用性。

本书可作为高等学校理工科计算机专业数据库课程的教材, 也可供相关工程技术人员参考使用。

21 世纪高等学校计算机规划教材 数据库基础与应用 (第 2 版)

-
- ◆ 编 著 王 珊 李盛恩
责任编辑 滑 玉
执行编辑 武恩玉
 - ◆ 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 电子函件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京隆昌伟业印刷有限公司印刷
 - ◆ 开本: 787×1092 1/16
印张: 18.5
字数: 482 千字
印数: 28 501—31 500 册

2009 年 6 月第 2 版

2009 年 6 月北京第 1 次印刷

ISBN 978-7-115-20508-7/TP

定价: 30.00 元

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

反盗版热线: (010)67171154

出版者的话

计算机科学与技术日新月异的发展,对我国高校计算机人才的培养提出了更高的要求。许多高校主动研究和调整学科内部结构、人才培养目标,提高学科水平和教学质量,精炼教学内容,拓宽专业基础,优化课程结构,改进教学方法,逐步形成了“基础课程精深,专业课程宽新”的良性格局。作为大学计算机教材建设的生力军,人民邮电出版社始终坚持服务高校教学、致力教育资源建设的出版理念,在总结前期教材建设成功经验的同时,深入调研和分析课程体系,并充分结合我国高校计算机教育现状和改革成果,推出“推介名师好书,共享教育资源”的教材建设项目,出版了“21 世纪高等学校计算机规划教材”名家系列。

本套教材的突出特点如下:

(1) 作者权威 本套教材的作者均为国内计算机学科中的学术泰斗或高校教学一线的教学名师,他们有着深厚的科研功底和丰富的教学经验。可以说,这套教材汇聚了众师之精华,充分显示了这套教材的格调和品位。无论是刚入杏坛的年轻教师,还是象牙塔内的莘莘学子,细细品读其中的章节文字,定会受益匪浅。

(2) 定位准确 本套教材是为普通高等院校的学生量身定做的精品教材。具体体现在:一是本套教材的作者长期从事一线科研和教学工作,对高校教学有着深刻而独到的见解;二是本套教材在选题策划阶段便多次召开调研会,对普通高校的教学需求和教材建设情况进行充分摸底,从而保证教材在内容组织和结构安排上更加贴近实际教学;三是组织有关作者到较为典型的普通高等院校讲授课程教学方法,深入了解教师的教学需求,充分把握学生的理解能力,以教材内容引导授课教师严格按照科学方法实施教学。

(3) 教材内容与时俱进 本套教材在充分吸收国内外最新计算机教学理念和教育体系的同时,更加注重基础理论、基本知识和基本技能的培养,集思想性、科学性、启发性、先进性和适应性于一身。

(4) 一纲多本,合理配套 根据不同的教学法,同一门课程可以有多本不同的教材,教材内容各具特色,实现教材系列资源配套。

总之,本套教材中的每一本精品教材都切实体现了各位教学名师的教学水平,充分折射出名师的教学思想,淋漓尽致地表达着名师的教学风格。我们相信,这套教材的出版发行一定能够启发年轻教师们真正领悟教学精髓,教会学生科学地掌握计算机专业的基本理论和知识,并通过实践深化对理论的理解,学以致用。

我们相信,这套教材的策划和出版,无论在形式上还是在内容上都能够显著提高我国高校计算机专业教材的整体水平,为培养符合时代发展要求的具有较强国际竞争力的高素质创新型计算机人才,为我国普通高等教育的计算机教材建设工作做出新的贡献。欢迎各位老师和读者给我们的工作提出宝贵意见。

前 言

数据库技术是对数据进行存储、管理、处理和维护的最先进、最常用的技术。随着计算机技术的飞速发展和计算机系统在各行各业的广泛应用，数据库技术的发展尤为迅速，已成为计算机信息系统和应用的核心技术和重要基础。

有关数据库系统的理论和技术是计算机科学技术教育中必不可少的部分。但是，不同的学校对数据库课程的要求是不一样的。本书针对培养应用型人才的要求，从开发一个数据库应用系统以及使用数据库系统的角度讲解数据库系统的基本概念、基本方法和基本技术。

全书分为四部分，共 13 章。

第一部分（第 1~5 章）介绍了数据库系统的基本知识和基本使用方法。内容包括：数据库系统的基本概念，关系数据模型和关系代数，关系数据库的 SQL 语言，查询处理初步，事务的基本概念和事务管理的相关技术。

第二部分（第 6、7 章）主要讲解在网络环境下开发数据库应用系统所要使用的嵌入式 SQL 技术、ODBC 接口、JDBC 接口、存储过程、触发器的基本概念和使用方法。

第三部分（第 8、9 章）简单介绍了数据库设计的基本过程，着重介绍了实体联系模型，关系规范化理论。

第四部分（第 10~13 章）介绍了数据库的新技术。内容包括对象关系数据库、XML 数据库、数据仓库和联机分析技术。

本书第 6、9、12 章和 13 章由王珊教授编写，其余各章由李盛恩教授编写。全书由王珊教授修改定稿。

本书侧重于数据库系统的应用，重点介绍了开发关系数据库系统必备的基本知识和基本方法。由于数据库技术的快速发展，出现了很多新技术，如对象关系数据库、XML、数据仓库和联机分析，在很多实际工作中要用到这些技术，本书对此作了较详细的介绍。同时，书中也介绍了基本的关系数据库理论。

限于作者水平，书中疏漏和错误难免，欢迎批评指正。

王 珊
中国人民大学

目 录

第 1 章 概述.....1

1.1 数据库的基本概念.....1

1.1.1 数据.....1

1.1.2 数据库.....1

1.1.3 数据库管理系统.....3

1.1.4 数据库系统.....5

1.2 数据模型.....6

1.2.1 数据模型的三要素.....6

1.2.2 3 种数据模型.....7

1.3 数据库系统的三级模式结构.....8

1.4 数据库系统的特点.....9

1.5 数据库系统的分类.....11

1.6 数据库管理系统的演变.....12

小结.....12

习题.....13

第 2 章 关系模型.....14

2.1 关系模型概述.....14

2.1.1 关系模型的数据结构.....14

2.1.2 关系模型的数据操作.....17

2.1.3 关系模型的完整性约束.....18

2.2 关系代数.....19

2.2.1 传统的集合运算.....20

2.2.2 专门的关系运算.....21

2.3 事例数据库.....23

小结.....25

习题.....25

第 3 章 关系数据库标准语言 SQL.....28

3.1 SQL 概述.....28

3.1.1 SQL 的产生和发展.....28

3.1.2 SQL 的组成.....28

3.1.3 SQL 的特点.....29

3.2 数据查询.....30

3.2.1 单表查询.....31

3.2.2 多表查询.....41

3.2.3 集合操作.....47

3.2.4 子查询.....48

3.3 数据定义.....56

3.3.1 表的定义.....57

3.3.2 索引的定义.....61

3.3.3 视图的定义.....63

3.4 数据更新.....68

3.5 存取控制.....70

3.6 空值的处理.....74

小结.....75

习题.....76

第 4 章 查询处理及优化.....78

4.1 查询处理的步骤.....78

4.2 查询处理算法.....78

4.2.1 外部排序.....79

4.2.2 集合操作算法.....80

4.2.3 选择操作算法.....82

4.2.4 连接操作算法.....83

4.3 查询优化.....84

4.3.1 概述.....84

4.3.2 一个实例.....85

4.3.3 查询优化的一般准则.....87

4.3.4 关系代数等价变换规则.....87

4.3.5 关系代数表达式的优化算法.....88

4.3.6 优化的一般步骤.....89

小结.....90

习题.....91

第 5 章 事务管理.....92

5.1 事务.....92

5.1.1 事务的特性.....92

5.1.2 定义事务的 SQL 语句.....93

5.2 恢复技术.....94

5.2.1 故障种类.....94

5.2.2 应对措施	95
5.2.3 恢复过程	97
5.3 并发控制	98
5.3.1 并发带来的问题	99
5.3.2 封锁技术	101
5.3.3 隔离级别	103
小结	104
习题	104

第6章 客户机/服务器数据库环境 106

6.1 客户机/服务器的一般概念	106
6.2 数据库应用系统结构的演变	107
6.3 两层与多层客户机/服务器结构	110
小结	113
习题	114

第7章 在应用中使用 SQL 115

7.1 嵌入式 SQL	115
7.1.1 嵌入式 SQL 的一般形式	115
7.1.2 嵌入式 SQL 语句与宿主语言之间的通信	116
7.1.3 查询结果为单个记录的 SELECT 语句	116
7.1.4 游标	117
7.1.5 动态 SQL 简介	120
7.1.6 实例	122
7.2 存储过程	124
7.2.1 SQL/PSM	124
7.2.2 PL/SQL	128
7.2.3 Transact-SQL	136
7.3 ODBC 简介	147
7.3.1 ODBC 原理概述	147
7.3.2 ODBC 驱动程序分类	150
7.3.3 ODBC 的工作流程及相关函数	152
7.3.4 实例	157
7.4 JDBC 简介	159
7.4.1 JDBC 原理概述	159
7.4.2 JDBC 的工作流程	160
7.4.3 实例	164

7.5 触发器	168
小结	173
习题	174

第8章 实体联系模型 176

8.1 基本的实体联系模型	176
8.1.1 基本概念	177
8.1.2 常见问题及解决方法	181
8.2 扩充的实体联系模型	183
8.2.1 IsA 联系	183
8.2.2 part-of 联系	184
8.3 从 E-R 图到关系数据库模式	185
8.3.1 实体型的表示方法	185
8.3.2 联系型的表示方法	187
8.3.3 IsA 联系的表示方法	188
小结	189
习题	189

第9章 关系规范化理论 191

9.1 数据依赖对关系模式的影响	191
9.2 函数依赖	192
9.2.1 函数依赖的基本概念	192
9.2.2 码	193
9.3 范式	193
9.3.1 第1范式(1NF)	194
9.3.2 第2范式(2NF)	195
9.3.3 第3范式(3NF)	196
9.3.4 BC 范式(BCNF)	197
9.4 多值依赖与第4范式(4NF)	198
9.4.1 多值依赖	198
9.4.2 第4范式(4NF)	200
9.5 关系模式的规范化	200
9.6 数据依赖的公理系统	201
9.7 模式分解	204
9.7.1 模式分解的3个定义	204
9.7.2 分解的无损连接性和保持函数依赖性	206
9.7.3 模式分解的算法	207
小结	210
习题	211

第 10 章 对象关系数据库.....212

10.1 关系模型的不足.....212

10.2 面向对象数据模型.....214

10.2.1 面向对象的基本概念.....214

10.2.2 面向对象数据模型的核心.....216

10.3 对象关系数据模型.....219

10.3.1 类型系统.....220

10.3.2 对象关系.....223

10.3.3 子表和超表.....224

10.3.4 查询和更新.....224

小结.....226

习题.....226

第 11 章 XML 数据库.....228

11.1 XML 简介.....228

11.1.1 XML 的特点.....228

11.1.2 XML 的应用.....229

11.1.3 XML 的相关标准.....229

11.1.4 XML 的存储.....230

11.2 XML 文档.....231

11.3 DTD-XML 模式定义语言.....233

11.4 XML Schema-XML 模式定义语言.....234

11.5 XPath 查询语言.....237

11.5.1 数据模型.....237

11.5.2 路径表达式.....239

11.5.3 XPath 函数.....241

11.6 XQuery 查询语言.....242

11.6.1 FLWOR 表达式.....242

11.6.2 连接.....243

11.6.3 嵌套查询.....245

11.6.4 排序.....245

11.7 XML 应用程序接口.....245

11.7.1 SAX.....246

11.7.2 DOM.....248

11.8 SQL/XML 标准.....252

11.8.1 发布 XML 文档.....252

11.8.2 存储和查询 XML 文档.....260

小结.....264

习题.....265

第 12 章 数据仓库技术.....266

12.1 从数据库到数据仓库.....266

12.2 数据仓库的基本概念.....268

12.2.1 什么是数据仓库.....268

12.2.2 主题与面向主题.....268

12.2.3 数据仓库的数据是集成的.....270

12.2.4 数据仓库的数据是不可更新的.....271

12.2.5 数据仓库数据是随时间不断

变化的.....271

12.3 数据仓库中的数据组织.....271

12.4 数据仓库系统的体系结构.....273

12.4.1 数据仓库的后台工具.....273

12.4.2 数据仓库服务器和 OLAP

服务器.....274

12.4.3 前台工具.....274

12.5 企业的体系化数据环境.....275

12.5.1 数据环境的层次.....275

12.5.2 数据集市.....275

12.6 创建数据仓库.....276

小结.....277

习题.....277

第 13 章 联机分析处理
(OLAP) 技术.....278

13.1 什么是 OLAP.....278

13.2 多维数据模型.....278

13.2.1 多维数据模型的基本概念.....278

13.2.2 多维分析的基本操作.....281

13.3 OLAP 的实现.....283

13.3.1 MOLAP 结构.....283

13.3.2 ROLAP 结构.....284

小结.....285

习题.....285

参考文献.....286

第 1 章

概 述

计算机最早被用于科学计算，目前已经深入到了我们生活的各个方面。计算机应用可大致分为三大类：科学计算、数据处理和过程控制。其中，数据处理占了很大的比重（约 70%），而数据管理又是数据处理的一个重要方面，数据库技术是数据管理的最新技术，是计算机科学的重要分支，是信息技术的基石。

本章介绍数据库系统的基本概念等内容，是后续章节的准备和基础。

1.1 数据库的基本概念

在学习本书之前，本节将首先介绍一些数据库最常用的术语和基本概念。

1.1.1 数据

数据是数据库中存储的基本对象。数据的种类很多，文本、图形、图像、音频、视频等都是数据。数据实际上是描述事物的符号记录。在日常生活中，人们可以直接用自然语言来描述事物。例如，可以这样描述一位同学的基本情况：王红，女，1985 年 5 月出生，山东省济南市人。

而在计算机中，为了存储和处理这些事物，就要抽象出对这些事物感兴趣的特征并组成一个记录来描述。例如：

（王红，女，198505，山东省济南市）

1.1.2 数据库

1. 什么是数据库

关于数据库的定义有很多，一般认为数据库是长期存储在计算机内有组织的、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、存储，具有较小的冗余度、较高的数据独立性和易扩展性，并可为各种用户共享。

数据库是一个组织机构（如企业、机关、银行、学校等）赖以生存的数据集合。例如，在一个学校中，有关学生数据、教师数据、课程数据、教学计划数据、教室及宿舍数据等就构成了学校数据库。客户数据、客户存取款记录和账户余额、往来账目等构成了银行数据库。

组织机构使用数据库开展日常工作，将部分工作自动化。例如，银行的存取款业务、铁路和民航的订票业务。我们拨打一次电话，通信公司的交换机除了在通话双方建立一个连接外，还把主叫号码、被叫号码、通话的起始时间等记录到公司的数据库中，月底根据这些通话记录以及客

户的其他信息自动生成账单。

数据库使用操作系统的若干个文件存储数据，也有一些数据库使用磁盘的一个或若干个分区存放数据。

文件中的数据由操作系统的文件系统进行管理，文件系统屏蔽了数据在磁盘或磁带等存储介质存储的细节。文件系统向用户提供了一个访问接口，这个接口一般包括文件名、fopen、fread、fwrite、fseek 和 fclose 函数调用，使用这个接口可以方便地向文件中写入数据以及从文件中读取数据。

数据库中的数据由数据库管理系统（DataBase Management System, DBMS）进行统一管理，DBMS 屏蔽了数据在数据库中的存放细节。DBMS 向用户提供了一个简单易学的语言，用户使用这个语言操纵数据库中的数据。DBMS 利用文件系统提供的接口读写数据库中的数据。

文件系统不“关心”文件中存放了什么样的数据，以及这些数据之间存在何种联系。从文件系统的视角看上去，文件是无结构的，文件只是一个字节流，因此，我们经常把文件叫做流式文件。实际上文件中的数据是有结构的，数据的结构需要程序员通过编写程序来建立和维护。

数据库中的数据一定是有结构的，很难想象把一个组织机构赖以生存的数据像家中储藏室的物品那样杂乱无章地堆放。无论数据库中数据的结构如何复杂，这些都被 DBMS 屏蔽了，我们看到的是一个简单的数据逻辑结构。

数据库中的数据除了存放组织机构日常业务要用到的数据外，还要存放保证数据库管理系统运行所要用到的数据，称为系统数据，例如，数据库管理系统中的用户信息、用户权限信息、各种统计信息等。

2. 数据库应用

数据库的应用十分广泛，在数据库管理系统的支持下可以开发面向各种领域的应用。各行各业都有其独特的应用。例如在银行业、保险业、证券业、电力、制造业等都有自己的数据库应用。

数据库应用可以分为两大类：联机事务处理（On-Line Transaction Processing, OLTP）和联机分析处理（On-Line Analytical Processing, OLAP）。联机事务处理解决了组织机构业务自动化问题，而联机分析处理帮助管理层更好地分析组织机构的运转情况，辅助领导进行正确决策。

操作系统通过调度执行进程完成用户对计算机的操纵。DBMS 则是执行事务达到用户自动进行业务处理的目的。一个数据库应用由若干个事务组成，而事务由一系列的对数据库的查询操作和更新操作构成，这些操作是一个整体，不能分割，即要么所有的操作都顺利完成，要么一个操作也不要做，绝不能只完成了部分操作，而还有一些操作没有完成。例如，我们要订购一张火车票，售票员运行一个订票事务，事务更新数据库中车票的余额，并打印火车票给客户。DBMS 每秒能处理的事务的数量叫做事务吞吐率，事务吞吐率是衡量 DBMS 的一个重要指标。为了完成一些繁忙的应用，DBMS 要具有很高的事务吞吐率。例如，一个门户网站在 1s 内会有成千上万甚至更多的并发访问用户，这就要求网站所用的数据库能同时处理很多的事务。

联机分析处理是决策支持的一种常用技术。例如，一个销售经理发现某种商品的销售量逐月下滑，他可以使用 OLAP 分析该商品在不同地区不同商店的一个时间段的销售情况，同其他同类商品的销售情况进行比对，最终发现问题并采取措施。

组织机构在多年使用数据库的过程中积累了大量的历史数据，这是一笔宝贵的财富。例如，一个大型连锁超市的数据库中保存了过去多年中客户的购物记录，通过对这些海量的购物记录进行关联分析后发现，在傍晚，有很多客户会同时购买尿布和啤酒。如果是同时购买面包和牛奶，我们不会感到惊讶，但是同时购买尿布和啤酒则让人觉得不可思议。经过调查后得知，很多年轻

的父亲在购物时会给孩子购买尿布而给自己购买啤酒。获得这个经验后，超市就将摆放尿布和啤酒的货架安排在相邻的位置，提高了两种商品的销售额。这是应用数据挖掘（Data Mining）技术的一个典型案例。数据挖掘也是决策支持的一种常用方法。

1.1.3 数据库管理系统

数据库管理系统（DBMS）是一类重要的系统软件，由一组程序构成，其主要功能是完成对数据库中数据的定义、数据操纵，提供给用户一个简明的应用接口，实现事务处理等。

1. 基本功能

（1）数据定义功能

数据库管理系统提供数据定义语言（Data Definition Language, DDL），用户通过它可以方便地对数据库中的数据对象进行定义。

（2）数据操纵功能

数据库管理系统提供数据操纵语言（Data Manipulation Language, DML），用户可以使用 DML 操纵数据，实现对数据库的基本操作——查询、插入、删除和修改。

（3）数据库的运行管理

数据库在建立、运行和维护时由数据库管理系统统一管理、统一控制，以保证数据的安全性、完整性、多用户对数据的并发使用及发生故障后的系统恢复。

（4）数据库的建立和维护功能

它包括数据库初始数据的输入、转换功能，数据库的转储、恢复功能，数据库的重组、重构功能和性能监视、分析功能等。这些功能通常是由一些实用程序完成的。

2. 组成模块

作为一个庞大的系统软件，数据库管理系统由众多程序模块组成，它们分别实现数据库管理系统复杂而繁多的功能。数据库管理系统由两大部分组成：查询处理器和存储管理器。查询处理器包含数据定义语言（DDL）编译器、数据操作语言（DML）编译器、嵌入型 DML 的预编译器及查询优化等核心程序。存储管理器包含授权和安全性控制、完整性检查管理器、事务管理器、存储管理器和缓冲区管理器。下面介绍各模块所完成的功能。

（1）数据定义方面的程序模块

数据定义的程序模块主要包括以下方面。

- 数据库逻辑结构的定义模块，在关系数据管理系统中就是创建数据库、创建表、创建视图、创建索引等定义模块。

- 安全性定义（如授权定义）及处理模块。

- 完整性定义（如主码、外码、其他完整性约束定义）及处理模块。

这些 DDL 程序模块接收相应的定义，进行语法、语义检查，把它们翻译为内部格式存储在数据字典中。创建数据库的模块还根据定义建立数据库的框架（即形成一个空库），等待装入数据。

（2）数据操纵方面的程序模块

数据操纵的程序模块主要包括查询处理程序模块、数据更新（增加、删除、修改）程序模块、交互式查询程序模块和嵌入式查询程序模块。

这些程序模块对用户的数据操纵请求进行语法分析和语义检查，生成某种内部表示（通常是

语法树)。对于查询语句,要由查询优化器(模块)进行优化,如根据一定的等价变换规则把语法树转换成标准(优化)形式;对于语法树中的每一个操作,根据存取路径、数据的存储分布、数据的聚簇等信息来选择具体的执行算法;最后生成查询计划(生成代码),提交查询执行模块执行,完成对数据库的存取操作。

(3) 数据库运行管理方面的程序模块

数据库运行管理的程序模块主要有系统初启程序,负责初始化数据库管理系统,建立数据库管理系统的系统缓冲区、系统工作区,打开数据字典等,此外,还有安全性控制、完整性检查、并发控制、事务管理、运行日志管理等程序模块,在数据库运行过程中监视对数据库的所有操作,控制管理数据库资源,处理多用户的并发操作等,它们一方面保证用户事务的正常运行及其原子性,一方面保证数据库的安全性和完整性。

(4) 数据库组织、存储和管理方面的程序模块

数据库组织、存储和管理的程序模块有文件读写与维护程序、存取路径(如索引)管理和维护程序、缓冲区管理程序(包括缓冲区读、写、淘汰模块)等。这些程序负责维护数据库的数据和存取路径,提供有效的存取方法。

(5) 数据库建立、维护和其他方面的程序模块

数据库建立、维护和其他的程序模块有数据库初始装入程序、转储程序、恢复程序、数据库重构程序、数据转换程序、通信程序等。

数据库管理系统的这些组成模块互相联系、互相依赖,共同完成数据库管理系统的复杂功能。这些模块之间的联系也不是平面的、散乱无章的,正如下面要讲的,它们具有一定的层次联系。

3. 层次结构

和操作系统一样,可以将数据库管理系统划分成若干层次。清晰、合理的层次结构不仅可以帮助我们更清楚地认识数据库管理系统,更重要的是有助于数据库管理系统的设计和维护。许多数据库管理系统实际上就是分层实现的。

例如,IBM公司最早研制的著名的数据库管理系统——System R,其核心分为两层,即底层的关系存储系统(RSS)和上层的系数据系统(RDS)。RDS本质上是一个语言翻译和执行层,完成包括语法检查与分析、优化、代码生成、视图实现、合法性检查等功能。RSS则是一个存取方法层,其功能包括存储空间、设备管理,索引、存取路径管理,并发控制,运行日志管理和恢复等。

图1-1给出了一个数据库管理系统的层次结构示例。这个层次结构是根据处理对象的不同,按照最高级到最低级的次序来划分的,具有普遍性。图中包括了与数据库管理系统密切相关的应用层和操作系统。

最上层是应用层,位于数据库管理系统核心之外。它处理的对象是各种各样的数据库应用,可以用开发工具开发或者用宿主语言编写。应用程序要利用数据库管理系统提供的接口来完成事务处理和查询处理。

第2层是语言翻译处理层。它处理的对象是数据库语言,如SQL。其功能是对数据库语言的各类语句进行语法分析、视图转换、授权检查、完整性检查、查询优化等。通过对下层基本模块的调用,生成可执行代码,运行这些代码即可完成数据库语句的功能要求。该层向上提供的接

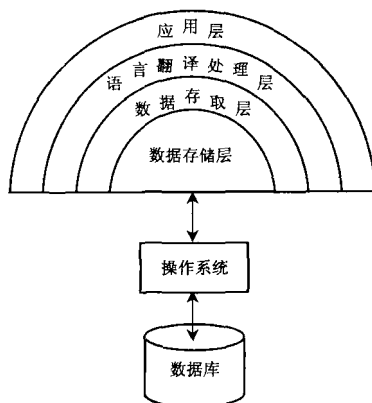


图 1-1 数据库管理系统的层次结构

口是关系、视图，它们是元组的集合。

第3层是数据存取层。该层处理的对象是单个元组。完成扫描（如表扫描）、排序、查找、插入、修改、删除、封锁等基本操作，完成存取路径维护、并发控制、事务管理、安全控制等工作。该层向上提供的接口是单记录操作。

第4层是数据存储层。该层处理的对象是数据页和系统缓冲区。执行文件的逻辑打开、关闭，读写数据页，完成缓冲区管理、内外存交换、外存的数据管理等功能。

操作系统是数据库管理系统的基础。它处理的对象是数据文件的物理块。执行物理文件的读写操作，保证数据库管理系统对数据逻辑上的读写真实地映射到物理文件上。操作系统提供的存取原语和基本的存取方法通常作为与数据库管理系统存储层连接的接口。

以上所述的数据库管理系统层次结构划分的思想具有普遍性。当然，具体系统在划分细节上会是多种多样的，这可以根据数据库管理系统实现的环境和系统的规模灵活地处理。

1.1.4 数据库系统

数据库系统就是基于数据库的计算机应用系统，由4部分组成：数据库（Data Base）、数据库管理系统（DataBase Management System）、应用程序和用户，这4部分之间的关系如图1-2所示。

数据库是数据的汇集，它以一定的组织形式保存在存储介质上；数据库管理系统是管理数据库的系统软件，它可以实现数据库系统的各种功能；应用程序是指以数据库管理系统和数据库中的数据为基础的程序。

数据库系统还包括用户，一般将用户分为应用程序开发人员、系统管理员和最终用户。

（1）应用程序开发人员

这里的应用程序开发人员包括系统分析员、数据库设计人员和应用程序员。

系统分析员使用软件工程的方法对业务流程进行分析，提出应用系统的需求分析和规范说明，与用户及数据库管理员一起确定系统的硬件、软件配置，并参与数据库管理系统的概要设计。

数据库设计人员根据分析阶段产生的数据流图确定数据库中数据的组织。数据库设计人员必须参加用户需求调查和系统分析，然后进行数据库设计。在很多情况下，数据库设计人员就由数据库管理员担任。

应用程序员根据详细设计说明书负责设计和编写应用系统的程序模块，并进行调试和安装。

（2）数据库管理员（DataBase Administrator, DBA）

数据库管理员管理数据库和数据库管理系统的日常运行，具体职责如下。

- 决定数据库上要存储的数据及数据结构。数据库上要存放哪些数据，DBA要参与决策。因此DBA必须参加数据库设计的全过程，并与用户、应用程序员、系统分析员密切合作、共同协商，搞好数据库设计。

- 决定数据库的存储结构和存取策略。DBA要综合各用户的应用要求，与数据库设计人员共同决定数据的存储结构和存取策略，以求获得较高的存取效率和存储空间利用率。

- 保证数据的安全性和完整性。DBA负责确定各个用户对数据库的存取权限、数据的保密级别和完整性约束条件。

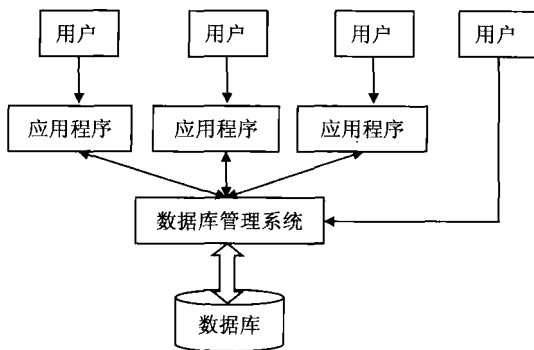


图1-2 数据库系统的组成

● 监控数据库的使用和运行。DBA 要监视数据库管理系统的运行情况，及时处理运行过程中出现的问题。例如，系统发生各种故障时，数据库会因此遭到不同程度的破坏，DBA 必须在最短时间内将数据库恢复到正确状态，并尽可能不影响或少影响计算机系统其他部分的正常运行。为此，DBA 要定义和实施适当的后备和恢复策略，如周期性的转储数据、维护日志文件等。

● 数据库的改进和重组重构。DBA 还负责在系统运行期间监视系统的空间利用率、处理效率等性能指标，对运行情况进行记录、统计分析，依靠工作实践并根据实际应用环境不断改进数据库设计。不少数据库产品都提供了对数据库运行状况进行监视和分析的实用程序，DBA 可以使用这些实用程序完成这项工作。

另外，在数据运行过程中，大量数据不断被插入、删除、修改，时间一长，会影响系统的性能。因此，DBA 要定期对数据库进行重组，以提高系统的性能。当用户的需求增加和改变时，DBA 还要对数据库进行较大的改造，包括修改部分设计，即数据库的重构。

（3）用户

这里用户是指最终用户（End User）。最终用户通过应用程序的用户接口使用数据库。常用的用户接口方式有浏览器、菜单驱动、表格操作、图形显示、报表书写等。

最终用户可以分为如下 3 类：

● 偶然用户。这类用户不经常访问数据库，但每次访问数据库时往往需要不同的数据库信息，这类用户一般是企业或组织机构的高中级管理人员。

● 简单用户。数据库的多数最终用户都是简单用户。其主要工作是查询和更新数据库，一般都是通过由应用程序员精心设计并具有友好界面的应用程序存取数据库。银行的职员、航空公司的机票预订工作人员、旅馆总台服务员等都属于这类用户。

● 复杂用户。复杂用户包括工程师、科学家、经济学家、科学技术工作者等具有较高科学技术背景的人员。这类用户一般都比较熟悉数据库管理系统的各种功能，能够直接使用数据库语言访问数据库，甚至能够基于数据库管理系统的 API 编制自己的应用程序。

数据库技术是计算机领域中发展最快的技术之一。数据库技术的发展是沿着数据模型的主线展开的。

1.2 数据模型

模型的概念我们并不陌生。数据模型（Data Model）也是一种模型，它是对现实世界数据特征的抽象。也就是说，数据模型是用来描述数据、组织数据和对数据进行操作的。通俗地讲，数据模型就是现实世界的模型。

现有的数据库系统均是基于某种数据模型的。数据模型是数据库系统的核心和基础。因此，了解数据模型的基本概念是学习数据库的基础。

数据库中的数据是按照一定规律存放的，这样便于管理和使用。数据模型用于描述数据库中数据的结构和性质、描述数据之间的联系以及施加在数据或数据联系上的一些限制。

1.2.1 数据模型的三要素

一般地讲，数据模型是对现实世界数据特征的抽象。也就是说，数据模型是用来描述数据、组织数据和对数据进行操作的。因此数据模型通常由数据结构、数据操作和完整性约束三部分组

成。数据模型描述了系统的静态特性、动态特性和完整性约束条件。

1. 数据结构

数据结构是所研究的对象类型的集合。这些对象是数据库的组成成分，它们包括两类：一类是与数据类型、内容、性质有关的对象，例如网状模型中的数据项、记录及关系模型中的域、属性、关系等；一类是与数据之间联系有关的对象，例如网状模型中的系型（Set Type）。

数据结构是刻画一个数据模型性质最重要的方面。因此在数据库管理系统中，人们通常按照其数据结构的类型来命名数据模型。例如，层次结构、网状结构和关系结构的数据模型分别命名为层次模型、网状模型和关系模型。

数据结构是对系统静态特性的描述。

2. 数据操作

数据操作是指对数据库中各种对象（指数据的型）的实例（指数据的值）允许执行的操作的集合，包括操作及有关的操作规则。数据库主要有检索和更新（包括插入、删除、修改）两大类操作。数据模型必须定义这些操作的确切含义、操作符号、操作规则（如优先级）以及实现操作的语言。数据操作是对系统动态特性的描述。

3. 数据的约束条件

数据的约束条件是一组完整性规则的集合。完整性规则是给定的数据模型中数据及其联系应满足的制约和依存规则，用以限定符合数据模型的数据库状态以及状态的变化，以保证数据的正确、有效、相容。

数据模型应该反映和规定本数据模型必须遵守的基本的通用的完整性约束条件。例如，在关系模型中，任何关系必须满足实体完整性和参照完整性两个条件（第2章将详细讨论这两个完整性约束条件）。

此外，数据模型还应该提供定义完整性约束条件的机制，以反映具体应用所涉及的数据必须遵守的特定的语义约束条件。

1.2.2 3种数据模型

如同在建筑设计和施工的不同阶段需要不同的图纸一样，在实施数据库应用中也需要使用不同的数据模型：概念模型（也称信息模型）、逻辑模型和物理模型。

概念模型独立于计算机系统，它完全不涉及信息在计算机系统上的表示，只是用来描述某个特定组织所关心的信息结构，是按用户的观点来对数据和信息建模，是对企业主要数据对象的基本表示和概括性描述，主要用于数据库设计。这类模型强调其语义表达能力，概念应该简单、清晰，易于用户理解，是数据库设计人员和用户之间进行交流的工具。著名的实体—联系模型就是概念模型中的代表，将在第8章中介绍。

逻辑模型是直接面向数据库的逻辑结构的，通常有一组严格定义的无二义性的语法和语义的数据库语言，人们可以用这种语言来定义、操纵数据库中的数据。应用程序员根据逻辑模型编程，每种数据库管理系统都支持一种逻辑模型。目前，数据库领域中最常用的逻辑模型有层次模型（Hierarchical Model）、网状模型（Network Model）、关系模型（Relational Model）、面向对象模型（Object Oriented Model）、对象关系模型（Object Relational Model）等，其中层次模型和网状模型

统称为非关系模型。非关系模型的数据库系统在 20 世纪 70 ~ 80 年代初非常流行，但是现已逐渐被关系模型的数据库系统所取代。由于早期开发的应用系统都是基于层次数据库或网状数据库系统的，因此目前仍有不少层次数据库或网状数据库系统在继续使用。

20 世纪 80 年代以来，面向对象的方法和技术的流行促进了数据库中面向对象模型的研究和发展。面向对象的数据模型将在第 10 章中介绍。

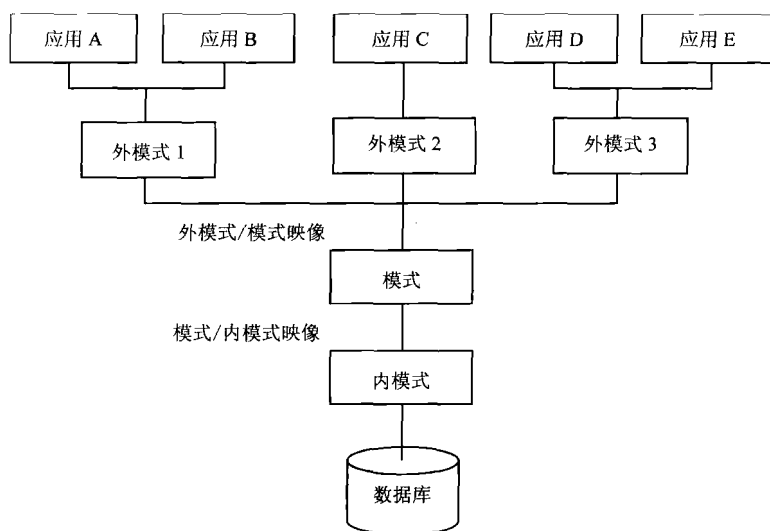
物理模型是对数据最底层的抽象，它描述数据在磁盘或磁带上的存储方式和存取方法，是面向计算机系统的。物理模型的具体实现是数据库管理系统的任务，在使用支持关系模型的数据库管理系统时，用户不必考虑物理级的细节。

从概念模型到逻辑模型的转换由数据库设计人员完成，从逻辑模型到物理模型的转换则由数据库管理系统完成。

一般人员掌握了逻辑模型就可以很方便地使用数据库。

1.3 数据库系统的三级模式结构

数据库系统通常采用三级模式结构，即外模式、模式和内模式，如图 1-3 所示。这是数据库系统内部的系统结构。



（1）模式

模式又被称为逻辑模式，是对数据库中全部数据的逻辑结构和特性的描述，是数据库所有用户的公共数据视图。

模式不仅定义数据的逻辑结构，如记录名及数据项的名字、类型、长度等，而且要定义与数据有关的安全性、完整性要求，此外，还要定义数据记录内部的结构，以及数据项之间的联系，进一步表示不同记录之间的联系。

一般数据库系统提供模式描述语言严格地表示这些内容。用 DDL 写出的一个数据库的定义的全部语句称为一个数据库的模式。模式是对数据库结构的一种描述，而不是数据库本身。在关

系数据库中对表的定义,以及对安全性、完整性的定义构成了数据库模式。

(2) 外模式

外模式又称为用户模式或子模式,通常是模式的子集,是数据库系统中每个用户看到和使用的的数据视图,即是与某一应用有关的数据的逻辑表示。

不同的用户因为需求不同,对待数据的方式可以不同,对数据的保密要求、使用的程序设计语言都可以不同,因此他们的用户模式描述是不同的。即使对于模式中的同一数据,在用户模式中的结构、类型、长度、保密级别均可以不同。

数据库系统提供外模式描述语言(外模式 DDL)描述用户数据视图。用外模式 DDL 写出的一个用户数据视图的逻辑定义的全部语句叫做此用户的外模式。外模式与用户使用的编程语言具有相容的语法。

在关系数据库中,用户可以对基本表或定义在基本表的视图上操作,对于某个用户来说,他对自己使用的表和视图的定义构成了他的用户模式。

(3) 内模式

内模式是数据库所有数据的内部表示或者说是底层的描述。内模式用来定义数据的存储方式和物理结构。例如,是按 B⁺树结构存储还是 Hash 方法存储,是否压缩存储,是否建立索引,是否加密,如何进行存储管理等。

系统也要提供内模式描述语言(内模式 DDL)来定义和描述内模式。现有的关系数据库产品中一般给出一系列的 DBA 实用程序来定义内模式,内模式的定义和修改是 DBA 的责任。

数据库系统的三级模式是对数据的 3 个抽象级别。它把数据的具体组织留给 DBMS 去做,用户只要抽象地逻辑地处理数据,而不用关心这些数据如何在计算机中表示和存储,大大减轻了用户使用计算机的负担。为了实现 3 个抽象层次的联系和转换,数据库系统在这三级模式中提供了以下两个层次的映像。

- 外模式/模式映像。
- 模式/内模式映像。

模式/内模式映像定义了数据的逻辑结构和存储结构的对应关系。例如,说明逻辑记录和字段在内部如何表示,当存储结构变了,模式/内模式的映像也必须做出相应的修改使模式不变。例如,在关系系统中某关系原来是以堆文件方式存储,现在按 B⁺树方式存储。DBA 做了文件存储方式的转换,对原关系建了 B⁺树索引,但关系名仍不变,关系的其他所有定义也没有变,即模式没有变化,保证系统中数据具有较高的物理独立性。

外模式/模式映像定义了某一外模式和模式之间的对应关系。这些映像定义通常包含在外模式中,当模式改变时,外模式/模式的映像要做相应的改变,以保证外模式不变。例如,在关系系统中,用户的外模式由一系列视图组成,若基本表的结构发生变化,如将一个表垂直分成两个表,这两个表的自然连接构成了原来的表,只要修改视图的定义,用户通过应用程序看到的视图并没有变化,应用程序不用修改,使得模式发生变化,外模式不变,应用程序不变,数据具有逻辑独立性。

1.4 数据库系统的特点

数据库系统是在文件系统的基础上发展起来的,数据库系统与文件系统相比有很多优点。

(1) 数据结构化

实现整体数据的结构化,是数据库的主要特征之一,也是数据库系统与文件系统的本质区别。