

21 世纪高等教育面向新工科软件工程系列规划教材

DATABASE SYSTEM PRINCIPLE
DESIGN AND PROGRAMMING (MOOC Edition)

数据库系统

原理、设计与编程

〔MOOC 版〕

陆鑫 张凤荔 陈安龙 / 编著



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS

21 世纪高等教育面向新工科软件工程系列规划教材

DATABASE SYSTEM PRINCIPLE
DESIGN AND PROGRAMMING (MOOC Edition)

数据库系统

原理、设计与编程

〔MOOC 版〕

陆鑫 张凤荔 陈安龙 / 编著

人民邮电出版社
北京

图书在版编目(CIP)数据

数据库系统：原理、设计与编程：MOOC版 / 陆鑫，张凤荔，陈安龙编著. — 北京：人民邮电出版社，2019.3

21世纪高等教育面向新工科软件工程系列规划教材
ISBN 978-7-115-50274-2

I. ①数… II. ①陆… ②张… ③陈… 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(2018)第277210号

内 容 提 要

本书以先进的开源对象——关系数据库 PostgreSQL 和典型的分布式非关系数据库 NoSQL 为背景，介绍数据库系统的原理、设计与编程技术。全书共 7 章，内容包括数据库系统基础、数据库关系模型、数据库操作语言 SQL、数据库设计与实现、数据库管理、数据库应用编程、NoSQL 数据库技术。针对关系数据库系统，本书重点介绍关系数据模型原理、数据库操作语言 SQL、数据库服务端编程、数据库管理技术；针对非结构化数据管理，本书介绍 NoSQL 数据库原理及其基本技术，如列存储数据库、键值对数据库、文档数据库、图形数据库等。同时，本书还针对数据库应用系统开发，介绍数据库设计方法、数据库应用 Java 编程方法。

本书取材新颖、内容实用、案例丰富，同时注重数据库工程实践应用。为支持面向新工科人才培养的翻转课堂教学，本书在每节均给出课程线上学习视频和课堂讨论问题。

本书既可作为高等学校计算机专业、软件工程专业数据库课程的教材，也可作为相关开发人员学习数据库知识与技术的参考书。

◆ 编 著 陆 鑫 张凤荔 陈安龙

责任编辑 邹文波

责任印制 彭志环

◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路 11 号

邮编 100164 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

山东百润本色印刷有限公司印刷

◆ 开本：787×1092 1/16

印张：21.25

2019 年 3 月第 1 版

字数：558 千字

2019 年 3 月山东第 1 次印刷

定价：59.80 元

读者服务热线：(010)81055256 印装质量热线：(010)81055316

反盗版热线：(010)81055315

21 世纪高等教育面向新工科软件工程系列规划教材

编委会名单

主 任：周世杰

副主任：汤 羽 蔡竞业 戴思俊

委 员：（按照拼音字母顺序排列）

傅 翀 管 庆 江春华 匡 平 蓝 天

李 凡 廖 勇 刘 博 刘 玓 陆 鑫

罗绪成 吴 劲 张 翔 邹文波

前 言

数据库技术是信息化应用的核心技术，任何信息系统都离不开数据库系统对数据的组织、存储和管理。掌握了数据库系统原理与技术就拥有了进入 IT 行业的钥匙。

为满足工程教育课程教学的需要，编者根据新工科人才培养的要求，遵循厚实专业基础，注重工程实践能力培养，反映产业先进技术的总体思路，编写了本书。编者在本书内容的组织、项目案例的设计、实践练习题的设计等方面突出工程教育的特点，注重内容对学生的工程师核心潜质能力（专业技能、工程实践能力、创新设计能力）的培养，解决传统教材理论知识与实际工程应用脱节、工程案例偏少等问题，为学生掌握数据库领域的专业知识、提升专业技能提供了丰富的学习素材。通过本书的学习，读者可以理解数据库系统的原理，掌握数据库系统的设计方法与开发技术，初步具备数据库应用系统的开发能力。

为取得更好的教学与学习效果，本书提供了 MOOC 教学资源，包括教学视频、教学 PPT、单元练习题等。此外，附录还提供了与本书配套的实验方案。

全书内容共 7 章。前 3 章介绍数据库系统基础、数据库关系模型、数据库操作语言 SQL 等数据库原理知识及基本技术。第 4 章介绍数据库设计与实现，包括 E-R 模型、数据库建模设计、数据库规范化设计、数据库设计模型的 SQL 实现等内容，并详细介绍了主流数据库设计工具 PowerDesigner 的具体应用。第 5 章对数据库管理技术进行介绍，主要包括事务管理、并发控制、安全管理、备份与恢复等内容。第 6 章结合 Java Web 介绍数据库应用编程技术。第 7 章介绍主流的 NoSQL 数据库技术及其应用方法，包括列存储数据库、键值对数据库、文档数据库、图形数据库等内容。

本书可作为高等学校计算机、软件工程等专业数据库课程的教材，建议授课学时为 48 小时，实验学时为 16 小时。

本书由陆鑫、陈安龙、张凤荔老师编写。其中，陆鑫编写了第 1、2、3、4、5 章及附录，并负责全书统稿；陈安龙编写了第 6 章；张凤荔编写了第 7 章。在本书编写过程中，编者得到电子科技大学教务处、信息与软件工程学院有关领导和老师的支持，在此表示诚挚感谢。

由于时间仓促，书中难免存在不妥之处，请读者原谅，并提出宝贵意见。

编者

2019 年 1 月于成都

目 录

第1章 数据库系统基础.....1

1.1 数据库及其系统的概念.....1	
1.1.1 数据库的定义.....1	
1.1.2 数据模型.....2	
1.1.3 数据库系统的组成.....4	
1.2 数据库技术的发展.....7	
1.2.1 数据管理技术的演化.....7	
1.2.2 数据库技术的发展阶段.....8	
1.2.3 数据库领域的新技术.....9	
1.3 数据库应用系统.....12	
1.3.1 数据库应用系统的类型.....12	
1.3.2 数据库应用系统的结构.....13	
1.3.3 数据库应用系统的生命周期.....15	
1.4 典型的数据库管理系统.....17	
1.4.1 Microsoft SQL Server.....17	
1.4.2 Oracle DataBase.....18	
1.4.3 MySQL.....18	
1.4.4 PostgreSQL.....19	
1.5 PostgreSQL 对象-关系数据库系统 软件.....19	
1.5.1 PostgreSQL 软件的获得.....19	
1.5.2 PostgreSQL 软件的功能程序.....20	
1.5.3 PostgreSQL 数据库的管理工具.....21	
1.5.4 PostgreSQL 数据库对象.....23	
习题.....24	

第2章 数据库关系模型.....25

2.1 关系及其相关概念.....25	
2.1.1 关系的通俗定义.....25	
2.1.2 关系的数学定义.....26	
2.1.3 关系模式表示.....28	
2.1.4 关系键的定义.....28	
2.2 关系模型的原理.....29	

2.2.1 关系模型的组成.....29	
2.2.2 关系模型的操作.....30	
2.2.3 数据完整性约束.....37	
2.3 PostgreSQL 数据库关系操作实践.....40	
2.3.1 项目案例——选课管理系统.....40	
2.3.2 关系数据库的创建.....42	
2.3.3 关系表的创建.....43	
2.3.4 实体完整性定义.....44	
2.3.5 参照完整性定义.....46	
2.3.6 用户自定义完整性.....48	
习题.....49	

第3章 数据库操作语言 SQL.....53

3.1 SQL 概述.....53	
3.1.1 SQL 的发展.....53	
3.1.2 SQL 的特点.....54	
3.1.3 SQL 的类型.....54	
3.1.4 SQL 的数据类型.....55	
3.2 数据定义 SQL 语句.....58	
3.2.1 数据库的定义.....58	
3.2.2 数据库表对象的定义.....60	
3.2.3 数据表索引对象的定义.....65	
3.3 数据操纵 SQL 语句.....67	
3.3.1 数据插入 SQL 语句.....67	
3.3.2 数据更新 SQL 语句.....68	
3.3.3 数据删除 SQL 语句.....69	
3.4 数据查询 SQL 语句.....70	
3.4.1 查询语句基本结构.....70	
3.4.2 从单表读取指定列.....71	
3.4.3 从单表读取指定行.....72	
3.4.4 从单表读取指定行和列.....73	
3.4.5 WHERE 子句条件.....73	
3.4.6 查询结果排序.....75	
3.4.7 内置函数的使用.....76	

3.4.8 查询结果分组处理	80	4.4.4 逆规范化处理	134
3.4.9 使用子查询处理多表	81	4.5 数据库设计模型的 SQL 实现	136
3.4.10 使用连接查询多表	82	4.5.1 确定数据库设计的实现方式	136
3.4.11 SQL JOIN...ON 连接	83	4.5.2 设计模型转换为 SQL 脚本程序	137
3.5 数据控制 SQL 语句	86	4.6 基于 PowerDesigner 的数据库设计建模实践	138
3.5.1 GRANT 语句	86	4.6.1 项目案例——图书借阅管理系统	138
3.5.2 REVOKE 语句	87	4.6.2 系统概念数据模型设计	138
3.5.3 DENY 语句	87	4.6.3 系统逻辑数据模型设计	142
3.6 视图 SQL 语句	87	4.6.4 系统物理数据模型设计	144
3.6.1 视图的概念	87	4.6.5 PostgreSQL 数据库实现	146
3.6.2 视图的创建与删除	88	习题	154
3.6.3 视图的使用	89		
3.7 PostgreSQL 数据库 SQL 实践	92	第 5 章 数据库管理	156
3.7.1 项目案例——工程项目管理系统	92	5.1 数据库管理概述	156
3.7.2 数据库的创建	93	5.1.1 数据库管理的目标与内容	156
3.7.3 数据库表的定义	94	5.1.2 数据库管理工具	157
3.7.4 数据的维护操作	95	5.1.3 DBMS 管理功能	157
3.7.5 多表的关联查询	98	5.1.4 DBMS 结构	158
3.7.6 视图的应用	100	5.2 事务管理	159
习题	102	5.2.1 事务的概念	159
第 4 章 数据库设计与实现	105	5.2.2 事务的特性	161
4.1 数据库设计概述	105	5.2.3 事务的并发执行	161
4.1.1 数据库设计方案	105	5.2.4 事务 SQL 语句	162
4.1.2 数据库设计过程与策略	107	5.3 并发控制	165
4.1.3 数据库建模设计工具	108	5.3.1 并发控制问题	165
4.2 E-R 模型	109	5.3.2 并发事务调度	168
4.2.1 模型基本元素	109	5.3.3 数据库锁机制	170
4.2.2 实体联系类型	111	5.3.4 基于锁的并发控制协议	171
4.2.3 强弱实体	114	5.3.5 两阶段锁定协议	174
4.2.4 标识符依赖实体	114	5.3.6 并发事务死锁解决	175
4.2.5 E-R 模型图	115	5.3.7 事务隔离级别	176
4.3 数据库建模设计	116	5.4 安全管理	176
4.3.1 概念数据模型设计	116	5.4.1 数据库系统安全概述	177
4.3.2 逻辑数据模型设计	119	5.4.2 数据库系统安全模型	177
4.3.3 物理数据模型设计	121	5.4.3 用户管理	179
4.4 数据库规范化设计	128	5.4.4 权限管理	182
4.4.1 非规范化关系表的问题	129	5.4.5 角色管理	184
4.4.2 函数依赖理论	130	5.5 备份与恢复	187
4.4.3 规范化设计范式	132	5.5.1 数据库备份	187

5.5.2 PostgreSQL 数据库的备份方法	189
5.5.3 数据库恢复	194
5.5.4 PostgreSQL 数据库的恢复方法	195
5.6 PostgreSQL 数据库管理项目实践	198
5.6.1 项目案例——成绩管理系统	198
5.6.2 数据库角色管理	199
5.6.3 数据库权限管理	200
5.6.4 数据库用户管理	201
5.6.5 数据库备份与恢复管理	202
习题	204

第 6 章 数据库应用编程 206

6.1 数据库连接技术	206
6.1.1 ODBC 技术	206
6.1.2 JDBC 技术	209
6.2 Java Web 数据库编程	215
6.2.1 Java Web 简介	215
6.2.2 Java Web 开发运行环境	215
6.2.3 Servlet 技术	216
6.2.4 JSP 技术	217
6.2.5 JavaBean 技术	227
6.2.6 MyBatis 访问数据库技术	227
6.2.7 MyBatis 数据库编程	229
6.3 存储过程编程	235
6.3.1 存储过程的基础知识	236
6.3.2 存储过程的优点	244
6.3.3 存储过程的缺点	244
6.4 触发器编程	245
6.4.1 触发器的语法结构	245
6.4.2 触发器的编程技术	248
6.4.3 事件触发器	250
6.4.4 触发器的应用	252
6.4.5 使用触发器的优点	253
6.5 游标编程	253
6.5.1 游标的基础知识	253
6.5.2 游标的应用编程	255
6.6 嵌入式 SQL 编程	257
6.6.1 嵌入式 SQL 的处理过程	257
6.6.2 嵌入式 SQL 的基本语法	257
6.6.3 嵌入式 SQL 的通信方式	258

6.7 数据库应用编程项目实践	260
6.7.1 项目案例——课程管理系统	260
6.7.2 功能模块设计	262
6.7.3 模块的编码实现	263
习题	272

第 7 章 NoSQL 数据库技术 274

7.1 NoSQL 数据库概述	274
7.1.1 关系数据库的局限	275
7.1.2 NoSQL 理论基础	276
7.1.3 NoSQL 基本概念	277
7.2 列存储数据库	280
7.2.1 列存储的概念	280
7.2.2 HBase 数据库的概念	281
7.2.3 HBase 数据库的存储结构	284
7.2.4 HBase 数据库的系统架构与组成	286
7.2.5 HBase 数据库的应用场景	288
7.3 键值对数据库	288
7.3.1 键值对存储的概念	289
7.3.2 Redis 数据库的基本知识	289
7.3.3 Redis 数据库的结构	290
7.3.4 Redis 数据库的相关操作	294
7.3.5 Redis 数据库的体系结构	295
7.3.6 Redis 数据库的应用场景	296
7.4 文档数据库	297
7.4.1 文档存储的概念	297
7.4.2 MongoDB 数据库的基本概念	297
7.4.3 MongoDB 数据库的管理	299
7.4.4 MongoDB 数据库的集群架构	300
7.4.5 MongoDB 数据库的应用场景	304
7.5 图形数据库	304
7.5.1 图形数据模型	306
7.5.2 Neo4j 图形数据库	309
7.5.3 Neo4j 图形数据库的存储结构	311
7.5.4 Neo4j 数据库的集群结构	313
7.5.5 Neo4j 数据库的查询语言 CQL	314
7.6 NoSQL 数据库项目实践	316
7.6.1 项目案例——成绩管理系统	316
7.6.2 设计 HBase 数据库表	317

7.6.3 安装 HBase 数据库	318
7.6.4 使用 HBase Shell	318
7.6.5 操作 HBase 数据库	320
习题	324

附录 PostgreSQL 数据库系统

实验项目 326

实验 1 图书销售管理系统数据库设计	326
实验 2 图书销售管理系统数据库的创建与 数据访问操作	327

实验 3 图书销售管理系统数据库后端 编程	328
实验 4 图书销售管理系统数据库安全 管理	329
实验 5 图书销售管理系统数据库应用 JSP 访问编程	330

参考文献 332

第 1 章

数据库系统基础

信息技术离不开数据库系统对数据的管理。数据库技术是实现信息化应用与服务的核心技术之一，它涉及数据组织与存储、数据存取模型、数据操作语言、数据架构与模型设计、数据管理与系统处理、应用数据访问编程等技术主题。本章将介绍数据库及其系统的概念，数据库技术的发展、数据库应用系统、典型的数据库管理系统、PostgreSQL 对象-关系数据库系统软件。

本章学习目标如下。

- (1) 了解数据库的基本概念、数据库的作用、数据库文件的特点。
- (2) 理解数据模型与数据库所使用的模型、数据库系统的组成等基本知识。
- (3) 了解数据库系统的技术发展过程、技术特点、技术趋势。
- (4) 了解典型数据库管理系统的技术特点、数据库应用适用场合。
- (5) 理解数据库应用系统的类型、结构，以及生命周期。

1.1 数据库及其系统的概念

在信息时代，无论是信息系统，还是互联网服务，它们都需要对各类数据进行存储、管理、分析等。这些数据处理都离不开数据库及其数据库管理系统软件的支持。什么是数据库？数据库如何组织、存储数据？数据库管理系统如何创建、访问和管理数据库？这些都是学习者需要了解与掌握的数据库领域知识。



1.1.1 数据库的定义

任何信息系统的技术实现，均需要使用具有特定数据模型的系统容器组织与存储数据，同时还需要相应系统软件支持应用程序对系统容器中的数据进行共享操作。在计算机领域中，这类组织与存储数据的系统容器被称为“数据库”。例如，电子商务系统将商品信息、销售价格信息、销售服务信息等业务数据分别写入由多个相关数据表构成的数据库中进行数据存储。当客户访问电子商务系统时，系统就立刻将每个商品的信息从数据库中提取出来，并呈现在电子商务网站页面中。若客户需要购买商品，可以在电子商务系统中填写订单信息，并支付货款，以完成一次客户与商家的线上交易活动。在电子商务系统中，所有业务实现都依赖数据库的支持。同样，各个机构的办公管理系统、财务管理系统、人力资源管理系统、

薪酬管理系统等业务信息系统都需要数据库来实现数据管理。因此，数据库是信息系统最重要的组成部分。

一些学者分别对数据库（DataBase）给出了更专业的定义。

定义1：简单来说，数据库是一种电子化的文件柜，用户可以对文件柜中的数据进行新增、检索、更新、删除等操作。

定义2：数据库是以一定方式存储在一起的、能为多个用户所共享的数据集合。

定义3：数据库是依照某种数据模型组织起来，并存放在存储器中的数据集合。这种数据集合具有如下特点：尽可能不重复存储数据，以优化方式为用户存取数据提供服务，其数据结构独立于使用它的应用程序，其数据的增、删、改和检索由统一软件进行管理和控制。

综合上述，我们可以将数据库理解为一种依照特定数据模型组织、存储和管理数据的文件集合。在信息系统中，数据库的基本作用是组织与存储系统数据，并为系统软件从中存取数据提供支持。与文件系统中普通数据文件有明显不同，数据库文件具有如下特点。

- 数据一般不重复存放。
- 可支持多个应用程序并发访问。
- 数据结构独立于使用它的应用程序。
- 对数据增、删、改、查操作均由数据库系统软件进行管理和控制。

1.1.2 数据模型

从数据库的定义可知，数据库使用了特定的数据模型来组织与存储数据，那么数据模型是什么，数据库一般使用哪种数据模型组织与存储数据，这些都是深入了解数据库原理的基本问题。

1. 数据模型

数据模型是一种描述事物对象数据特征及其结构的形式化表示，通常由数据结构、数据操作、数据约束3个部分组成。

（1）数据结构用于描述事物对象的静态特征，其中包括事物对象的属性数据、数据类型、数据组织方式等。数据结构是数据模型的基础，数据操作和约束都是基于该数据结构进行的。

（2）数据操作用于描述事物对象的动态特征，即对事物对象的属性数据可进行的数据操作，如插入、更新、删除和查询等。在数据模型中，我们通常还需定义数据操作语言，如操作语句、操作规则及操作结果表示。

（3）数据约束用于描述事物对象的数据之间语义的联系，以及数据取值范围等规则，从而确保数据的完整性、一致性和有效性。

例如，在许多高级编程语言中，数据文件是一种典型的数据模型实例，其数据结构由若干数据记录行组成，数据记录之间彼此独立，每行数据记录又由若干数据项组成。在数据文件的编程访问中，我们可以移动指针定位来记录位置，然后对该记录数据进行读/写或删除处理。

2. 数据库所使用的数据模型

传统数据库先后使用的数据模型主要有层次数据模型、网状数据模型、关系数据模型。这3种模型是按其数据结构命名的。它们之间的根本区别在于数据之间的联系方式不同，即数据记录之间的联系方式不同：层次数据模型以“树结构”方式表示数据记录之间的联系；网状数据模型以“图结构”方式来表示数据记录之间的联系；关系数据模型用“关系二维表”

方式来表示数据记录之间的联系。

(1) 层次数据模型

层次数据模型是数据库系统最早使用的一种数据模型，其数据结构是一棵包含多个数据结点的“有向树”。根结点在最上端，层次最高；子结点在下层；最低层为叶结点。每个数据结点存储一个数据记录，数据结点之间通过链接指针进行联系。当需要访问层次数据模型的数据时，需要使用树结点遍历方法进行数据记录的操作。例如，高校教务系统的层次数据模型如图 1-1 所示。

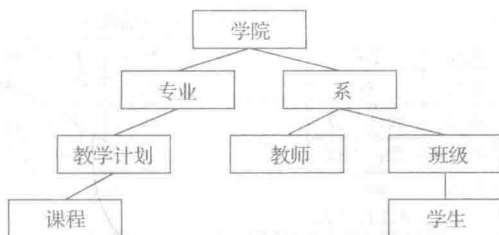


图 1-1 高校教务系统的层次数据模型

层次数据模型的特征：该模型将数据结点组织成多叉树关系的数据结构，程序采用关键字检索来遍历访问各个数据结点。其优点：数据结构层次清晰，使用指针可遍历访问各个数据结点；数据结点的更新和扩展容易实现；关键字检索查询效率高。其缺点：系统数据结构局限于层次结构，缺乏灵活性；相同信息的数据结点可能会多次存储，数据冗余大；层次数据模型不适合于具有拓扑空间的数据组织。

层次数据模型的数据库产品出现于 20 世纪 60 年代末，最具代表性的数据库产品是 IBM 公司推出的 IMS 层次模型数据库系统。

(2) 网状数据模型

网状数据模型以网络图表示数据结点之间的联系。该网状结构中每一个数据结点代表一个数据记录，结点之间的联系也使用链接指针来实现，但网状数据模型可以表示数据结点之间的多个从属关系，同时也可以表示数据结点之间的横向关系。网状数据模型扩展了层次数据模型的数据关系，其数据处理更方便。例如，高校教务系统的网状数据模型如图 1-2 所示。

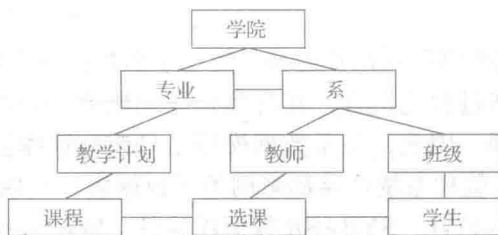


图 1-2 高校教务系统的网状数据模型

该模型采用链接指针来确定数据结点之间的联系，可支持具有多对多联系的数据结点组织方式。网状数据模型允许各个结点有多于一个父结点，也可以有一个以上的结点没有父结点。其优点：能明确而方便地表示数据间的复杂关系，数据冗余小。其缺点：结构较复杂，增加了数据查询和数据定位的困难；需要存储数据间联系的指针，使得数据存储量增大；数据更新不方便，除了对数据进行更新外，还必须修改关联指针。

网状数据模型的数据库产品出现于 20 世纪 70 年代，使用网状数据模型的典型数据库系

统产品有 Cullinet 软件公司的 IDMS、Honeywell 公司的 IDSII、Univac 公司的 DMS1100、HP 公司的 IMAGE 等。

（3）关系数据模型

关系数据模型（简称关系模型）是以关系代数理论为基础，通过二维表结构来表示数据记录之间联系的数据模型。关系数据模型的每个二维表应具有关系特征，它们又被称为关系表。在关系数据模型中，多个二维表可通过相同属性列的一致性约束进行数据关联。例如，课程目录系统的关系数据模型如图 1-3 所示。



图 1-3 课程目录系统的关系数据模型

在图 1-3 所示的关系数据模型实例中，“教师信息表”“课程信息表”“开课目录表”均为具有关系特征的二维表。每个关系表分别存放各自主题的数据，表之间通过具有相同列属性的数据值进行约束关联。其中，“开课目录表”的“工号”属性列与“教师信息表”的“工号”属性列的数据值必须保持一致。同样，“开课目录表”的“课程号”属性列的数据值也要求与“课程信息表”的“课程号”属性列的数据值匹配一致。因而，这些关系表之间通过相同属性列建立了约束关联。

关系数据模型的优点：数据结构简单、数据操作灵活；支持关系与集合运算操作；支持广泛使用的结构化查询语言（SQL）；容易实现与应用程序的数据独立性。其缺点：局限于结构化数据组织与存储；支持的数据类型较简单；难以支持非结构化数据和复杂数据处理。

关系数据模型存取数据路径对用户隐蔽，使程序和数据具有高度的独立性。关系数据模型的数据操作语言为非过程化语言，具有集合处理能力，并能实现数据对象的定义、操纵、控制等一体化处理。因此，关系数据模型是目前数据库使用最广泛的数据模型，几乎所有的结构化数据库管理系统产品都采用关系数据模型实现数据库，如 Oracle 数据库软件、IBM DB2 数据库软件、SYBASE 数据库软件、SQL Server 数据库软件、MySQL 数据库软件等。

1.1.3 数据库系统的组成

数据库系统（DataBase Systems）是一种基于数据库进行数据管理与信息服务的软件系统。当数据库系统在应用领域实现数据存储、数据处理、数据检索、数据分析等功能时，数据库系统又被称为数据库应用系统。数据库系统由用户、数据库应用程序、数据库管理系统和数据库 4 个部分所组成，如图 1-4 所示。

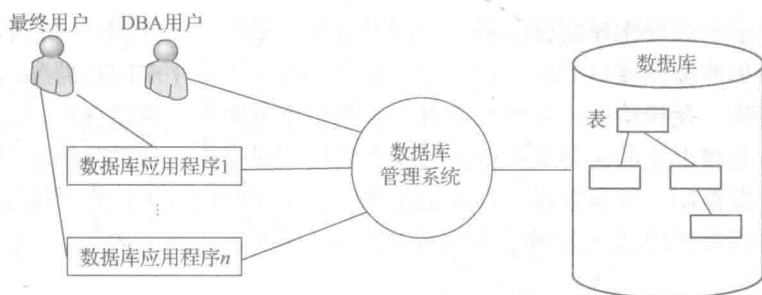


图 1-4 数据库系统组成

1. 用户

在数据库系统中，用户可分为最终用户和 DBA（DataBase Administrator）用户两类。最终用户通过操作数据库应用程序处理业务，并利用程序存取数据库信息。当然，数据库应用程序不能直接存取数据库信息，必须基于数据库管理系统（DataBase Management System, DBMS）提供的接口和环境才能访问数据库。DBA 用户是一种专门进行数据库管理与运行维护的系统用户，该用户通过使用 DBMS 软件提供的管理工具对数据库进行创建、管理和维护，从而为数据库系统的正常运行提供支持与保障。

2. 数据库应用程序

数据库应用程序是一种在 DBMS 支持下对数据库进行访问和处理的应用程序。它们以窗口或页面等表单形式来读取、更新、查询或统计数据库信息，从而实现业务数据处理与信息服务。数据库应用程序需要使用 DBMS 提供的标准接口（如 ODBC、JDBC 等）驱动程序连接数据库。在程序设计语言（如 Java、C++、C#、VB、PB 等）编程中，我们需要使用数据库访问接口实现对数据库的操作。

3. 数据库管理系统

数据库管理系统是一类用于创建、操纵和管理数据库的系统软件。数据库管理系统与操作系统一样都属于系统平台软件。数据库管理系统一般具有如下功能：①创建数据库、数据库表及其他对象；②读/写、更新、删除数据库表数据；③维护数据库结构；④执行数据访问规则；⑤提供数据库并发访问控制和安全控制；⑥执行数据库备份和恢复。

从软件结构来看，数据库管理系统由操作界面层、语言翻译处理层、数据存取层、数据存储层组成，其层次结构如图 1-5 所示。

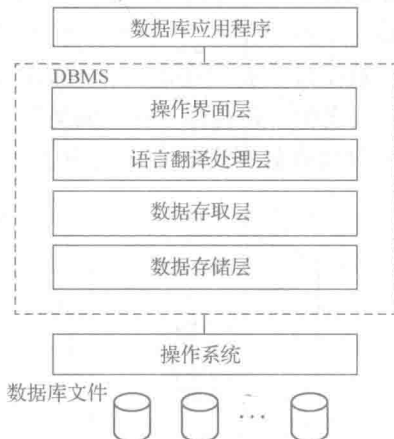


图 1-5 数据库管理系统的层次结构

数据库管理系统的操作界面层由若干管理工具和应用程序 API 组成，它们分别为用户和应用程序访问数据库提供接口界面。语言翻译处理层对应用程序中的数据库操作语句进行语法分析、视图转换、授权检查、完整性检查、查询优化等处理。数据存取层处理的对象是数据表中的记录，它将上层的集合关系操作转换为数据记录操作，如对数据记录进行存取、维护存取路径、并发控制、事务管理、日志记录等。数据存储层基于操作系统提供的系统调用对数据库文件进行数据块读/写操作，并完成数据页、系统缓冲区、内外存交换、外存数据文件等系统操作管理。

目前，有较多的软件厂商提供了功能强大的数据库管理系统产品，如 Microsoft 公司提供的 SQL Server 数据库软件、Access 数据库软件，Oracle 公司提供的 Oracle Database 数据库软件，SAP Sybase 公司提供的 Sybase ASE 数据库软件、Sybase Anywhere 数据库软件，IBM 公司提供的 DB2 数据库软件。此外，不少开源组织机构也发布了 DBMS 软件，如 MySQL、PostgreSQL、MongoDB、InterBase 等。

4. 数据库

在数据库系统中，数据库是存放系统各类数据的容器。该容器按照一定的数据模型组织与存储数据。目前，在数据库系统应用中，使用最多的数据库软件产品是关系数据库软件。这类数据库采用关系数据模型实现数据组织与存储。例如，使用关系数据库进行成绩管理，该数据库由 COURSE 关系表、GRADE 关系表和 STUDENT 关系表组成，各关系表之间的联系如图 1-6 所示。

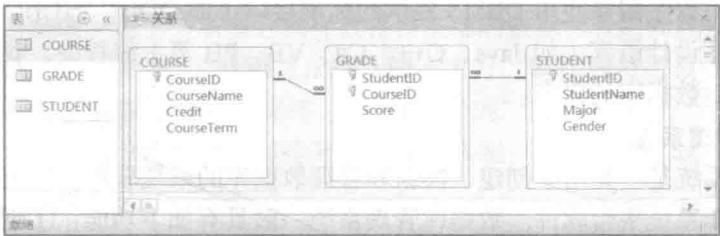


图 1-6 成绩管理关系数据库

在图 1-6 所示的成绩管理关系数据库示例中，该数据库的 COURSE 表、GRADE 表和 STUDENT 表均为关系表，它们之间通过公共列 CourseID、StudentID 建立了表之间的参照关联。因此，我们可以在成绩表（GRADE）中，针对某个学生（StudentID）参加某课程（CourseID）考试，给出具体课程的成绩分数（Score）值。

在数据库中，除了存放用户数据外，我们还需要存放描述数据库结构的元数据。例如，在关系数据库中，各个关系表的表名称、列名称、列数据类型、数据约束规则等都是元数据，这些描述数据库结构的数据需要存放在数据库的系统表中。图 1-7 给出了关系数据库存储的主要数据信息类型。



图 1-7 关系数据库存储的主要数据信息类型

在关系数据库中,数据库元数据、索引数据、运行数据等存放在系统表中,用户数据则只能存放在用户表中。

课堂讨论——本节重点与难点知识问题

1. 如何区分数据库、数据库管理系统、数据库应用系统、数据库系统?
2. 在数据库中,数据模型有何作用?它有哪些组成部分?
3. 在数据库技术发展阶段中,先后采用过哪些数据模型?各有什么特点?
4. 举例说明哪些典型数据库软件产品或开源系统采用关系数据模型?
5. 数据库系统有哪几个组成部分?各个部分有何作用?
6. 什么是元数据?它是如何产生的,存储在哪里?

1.2 数据库技术的发展

数据库技术是一种利用计算机组织、存储和管理数据的软件技术。它涉及研究数据库的结构、存储、设计、管理、应用的基本理论和实现方法,并利用这些理论方法对数据库中的数据进行存取、计算、统计及分析等操作。

扫码预习



1.2.1 数据管理技术的演化

在20世纪60年代末,计算机技术快速发展并被广泛应用,大量用户提出对数据资源进行存储管理和数据存取处理的需求,由此产生了利用计算机进行数据管理的原始数据库技术。在当时,数据库技术主要研究如何存储、使用和管理数据。随着计算机技术的发展,数据库技术与计算机相关技术的发展相互渗透与相互促进,现已成为当今计算机软件领域发展最迅速、应用最广泛的技术。数据库技术不仅应用于数据资源存取管理,还应用到信息检索、数据仓库、数据挖掘、商业智能、大数据处理等领域。在利用计算机进行数据管理的技术发展历程中,数据管理经历了人工数据管理、文件系统管理、数据库管理3个阶段。

1. 人工数据管理阶段

在20世纪50年代以前,计算机主要用于科学计算。计算机外部存储只有纸带、卡片、磁带等设备,没有直接存取设备。计算机软件只是一些操作控制程序,还没有操作系统及数据管理软件。计算机可处理的数据量小,数据无结构,数据依赖于特定的应用程序,缺乏独立性。在涉及数据处理的计算机程序中,程序员必须在代码中进行数据管理。因此,当时的数据管理存在很大局限性,难以满足应用数据管理的要求。

2. 文件系统管理阶段

在20世纪50年代后期到60年代中期,新的计算机外部存储设备如磁鼓、磁盘等出现,它们可以用来长久存储程序与数据,并支持直接数据块存取。在这个时期,计算机软件也得到快速发展,出现了控制计算机软硬件运行的操作系统软件。在操作系统中,可使用数据文件方式来组织、存储数据,并采用文件系统工具管理各个独立的数据文件。文件系统可以按照数据文件的名称对其进行访问,既可对数据文件中的数据记录进行存取,也可对数据记录进行更新、插入和删除。文件系统实现了数据在记录内的结构化,即在数据文件的各个记录

内，数据项组成是一致的。但是从整体看数据文件，数据记录之间是无结构的，不能处理数据记录之间的关联性。

在这个阶段，用户可以使用高级语言程序对数据文件进行数据记录的存取，解决了人工数据管理的限制，可以满足应用的基本数据管理要求。但在文件管理数据方式中，存在如下不足：①编写应用程序管理数据的过程较烦琐。②数据文件对应用程序存在依赖，难以实现独立。③不支持多用户对数据文件并发访问。④不能实现数据文件的安全控制。⑤难以解决不同数据文件间的数据冗余。⑥在文件中，数据记录之间缺少联系，难以满足用户对数据的关联访问需求。

3. 数据库管理阶段

在 20 世纪 60 年代末期，计算机软硬件技术得到较大发展：计算机处理能力得到较大提高，并且大容量磁盘设备开始出现；计算机软件也出现专门管理数据的系统软件——数据库管理系统。这些技术都为实现大规模计算机数据管理提供了支持。在这个阶段，用户可使用数据库管理系统来实现应用系统的数据管理。应用程序连接数据库后，用户可使用数据库操作语言对其表中数据进行操作。所有对数据库的操作都由数据库管理系统自动去完成，应用程序不需要考虑数据库文件的物理操作和系统控制。数据库管理与文件数据管理相比较，具有如下优点：①应用程序与数据相互独立，避免了应用程序对数据的依赖性。②应用程序访问数据库使用标准语言操作，编程访问简单。③数据组织结构化、共享性高、冗余小。④提供数据的安全访问机制，并保证数据的完整性、一致性、正确性。

因此，数据库技术成为当今计算机数据管理的基本技术。虽然数据库技术从 20 世纪 60 年代末期到现在经历了几十年发展，其技术也发生了许多变化，但数据库组织与管理数据的基本思想是一致的，这说明数据库技术管理数据的生命力是长久的。

1.2.2 数据库技术的发展阶段

数据模型是数据库技术的核心基础，数据模型的发展演变可以作为数据库技术发展阶段的主要标志。按照数据模型的发展演变过程，数据库技术从出现到如今半个多世纪中，主要经历了 3 代：第一代是以层次数据模型和网状数据模型为特征的数据库技术，第二代是以关系数据模型为特征的数据库技术，第三代是以面向对象数据模型为主要特征的数据库技术。目前数据库技术进入到第四代，数据库技术与计算机网络技术、人工智能技术、并行计算技术、多媒体技术、云计算技术、大数据技术等相互结合与相互促进，衍生出大量数据库新技术，其典型特征是采用非结构化的数据模型处理大数据。

1. 第一代数据库技术

第一代数据库技术出现于 20 世纪 60 年代末人们研制的层次模型数据库系统和网状模型数据库系统。层次模型数据库系统的典型代表是 1968 年 IBM 公司研制出的世界上第一个数据库管理系统 IMS（Information Management System）。该数据库系统最早运行在 IBM 360/370 计算机上。经过多年技术改进后，该系统至今还在 IBM 部分大型主机中使用。网状模型数据库系统的典型代表是 1964 年通用电器公司研制的网状数据模型的数据库管理系统 IDS（Integrated Data System）。IDS 奠定了网状数据库技术基础，并在当时得到了广泛的发行和应用。

20 世纪 70 年代初，美国数据库系统语言协会 CODASYL（Conference On Data System Language）下属的数据库任务组 DBTG（DataBase Task Group）对数据库技术方法进行了系