

普通高等教育“十三五”规划教材 (软件工程专业)

数据库原理及应用

—— SQL Server 2012

主 编 赖 玲 李祥琴 胡 秀 王娅纷 沈成涛

- ◎ 扎实基础，由浅入深地组织内容。
- ◎ 结合实践，实践侧重数据库应用。
- ◎ 巩固练习，配备实例、习题和配套教材。

SQL Server
2012



中国水利水电出版社
www.waterpub.com.cn

普通高等教育“ (软件工程专业)

数据库原理及应用

——SQL Server 2012

主 编 赖 玲 李祥琴 胡 秀 王娅纷 沈成涛



中国水利水电出版社
www.waterpub.com.cn

·北京·

内 容 提 要

本书全面介绍了数据库的原理及应用,全书共13章,分成两部分,第一部分侧重数据库理论知识,包括数据库的基本概念、数据模型、关系数据库、关系代数、关系规范化、SQL、事务和锁、数据库设计和数据库技术的新发展;第二部分侧重数据库应用,以Microsoft SQL Server 2012为平台,详细介绍了数据库的管理及应用,包括索引、T-SQL程序设计、存储过程、触发器、游标、数据库的安全管理、数据的备份与恢复、SQL Server开发工具。每章以丰富的实例进行讲解,并配备了大量课后习题。此外,本书还有配套教材《数据库原理及应用上机指导与习题解答——SQL Server 2012》。

本书可作为高等学校计算机专业“数据库”课程的教材,也可作为其他相关专业“数据库”课程的教材,还可作为从事数据库开发和应用的有关人员的参考书。

图书在版编目(CIP)数据

数据库原理及应用:SQL Server 2012 / 赖玲等主编.
—北京:中国水利水电出版社,2017.5
普通高等教育“十三五”规划教材. 软件工程专业
ISBN 978-7-5170-5279-1

I. ①数… II. ①赖… 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(2017)第065432号

策划编辑:杨庆川 责任编辑:李炎 加工编辑:郭继琼 封面设计:李佳

书 名	普通高等教育“十三五”规划教材(软件工程专业) 数据库原理及应用——SQL Server 2012 SHUJUKU YUANLI JI YINGYONG——SQL Server 2012
作 者	主 编 赖 玲 李祥琴 胡 秀 王娅纷 沈成涛
出版发行	中国水利水电出版社 (北京市海淀区玉渊潭南路1号D座 100038) 网址: www.waterpub.com.cn E-mail: mchannel@263.net(万水) sales@waterpub.com.cn
经 售	电话:(010) 68367658(营销中心)、82562819(万水) 全国各地新华书店和相关出版物销售网点
排 版	北京万水电子信息有限公司
印 刷	三河市鑫金马印装有限公司
规 格	184mm×260mm 16开本 19.25印张 474千字
版 次	2017年5月第1版 2017年5月第1次印刷
印 数	0001—3000册
定 价	39.00元

凡购买我社图书,如有缺页、倒页、脱页的,本社营销中心负责调换
版权所有·侵权必究

前 言

数据库技术是信息处理的基础，其应用范围广，几乎涵盖了信息技术的各个领域。SQL Server 是微软的核心数据库平台。如今，Microsoft SQL Server 已经从一个只能支持小型部门任务的产品成长为能够处理部署于世界各地的任务的超大型数据库平台。近年来，其不断发布的新版本已涵盖越来越广泛且强大的功能与组件，从而使其在本领域绝大多数竞争对手的角逐中脱颖而出。目前 SQL Server 已经是市场上最流行的大中型关系数据库管理系统。

为了适应市场的需求，我国高校的许多专业都开设了“数据库原理及应用”课程。开设数据库课程的目的是使学生在掌握数据库的基本原理、方法和技术的基础上，能根据应用需求灵活设计适合的数据库，并能结合现有的数据库管理系统进行数据库的管理及数据库应用系统的开发。根据教育部高等学校计算机基础课程教学指导委员会《高等学校计算机基础核心课程教学实施方案》的要求，本书以 SQL Server 2012 为平台，结合作者多年来教学与应用开发的实践经验，将完整的数据库原理及应用知识体系，按照理论和实例相结合的模式，由浅入深地组织和安排内容。通过对本书的学习，读者无论对 SQL Server 数据库应用开发，还是对数据库管理都会有新的认识和提高。

本书共分为 13 章，主要内容如下：

第 1 章 数据库系统概述。从数据管理技术的发展开始介绍，进而讲解了数据库技术的相关概念，还介绍了数据模型。

第 2 章 关系数据库。首先介绍了关系模型、关系的形式化定义、关系的键和关系完整性，然后重点介绍了关系代数。

第 3 章 关系数据库理论。首先提出了关系规范化问题，接着介绍了函数依赖、范式，然后讲解了关系模式的规范化。

第 4 章 关系数据库标准语言 SQL。从 SQL 的数据定义语言 DDL、数据操作语言 DML 和数据查询语言 DQL 三个方面，对 SQL 进行了详细的介绍，最后介绍了视图，包括视图的创建、修改、删除和使用。

第 5 章 索引。对索引和全文索引进行了详细的介绍，包括索引的创建、修改、删除和全文索引的启用、创建等。

第 6 章 T-SQL 程序设计。从 T-SQL 语言基础开始，逐步介绍了数据类型、变量和运算符，然后重点详细介绍了流程控制语句，最后介绍了系统内置函数和用户自定义函数。

第 7 章 存储过程、触发器和游标。分别详细介绍了存储过程的概念、创建、执行、修改、查看和删除；触发器的概念、创建、修改、查看和删除；游标的概念、使用、删除等。

第 8 章 事务和锁。首先介绍了事务的相关概念，进而提出并介绍了并发控制，最后介绍了锁的相关知识。

第 9 章 数据库的安全管理。首先介绍了 SQL Server 的安全机制，然后从各个方面分别详细介绍了安全机制的实现。

第 10 章 数据的备份与恢复。首先介绍了数据库备份的概念、备份设备及备份操作，然

后介绍了数据恢复的策略及操作。

第 11 章 数据库设计。结合软件工程的思想对数据库设计的各个步骤进行了详细的介绍。

第 12 章 数据库技术的新发展。介绍了数据库技术的发展趋势及数据库发展的新技术。

第 13 章 SQL Server 开发工具。分别介绍了 SQL Server 常用的代理服务、集成服务、报表服务和分析服务。

本书由赖玲、李祥琴、胡秀、王娅纷、沈成涛共同主编，胡波、李俊梅、张牧、吴际林也参加了本书的编写与校对工作。全书由赖玲统稿。第 1 章由王娅纷编写；第 2 章由沈成涛编写；第 4 章、第 5 章、第 9 章、第 10 章、第 11 章、第 13 章由赖玲编写；第 6 章、第 7 章、第 8 章由李祥琴编写；第 3 章、第 12 章由胡秀编写。本书在编写过程中得到了荆楚理工学院计算机工程学院田原院长和任正云等专家的指导，学院的领导也对本书的出版付出了大量的心血，在此一并表示衷心的感谢。

由于编者水平所限，书中难免存在疏漏之处，恳请广大读者批评指正。

编 者

2017 年 2 月

目 录

前言

第1章 数据库系统概述	1	习题	23
1.1 数据管理技术的发展	1	第3章 关系数据库理论	25
1.2 数据库技术概述	3	3.1 规范化问题的提出	25
1.3 数据模型	5	3.1.1 规范化理论的主要内容	25
1.3.1 数据模型概念	5	3.1.2 关系模式存在的问题	25
1.3.2 实体的描述	6	3.1.3 解决问题的方法	27
1.3.3 联系	6	3.2 函数依赖	28
1.3.4 实体—联系模型	7	3.2.1 函数依赖的定义	28
1.3.5 层次模型	9	3.2.2 有关函数依赖的说明	29
1.3.6 网状模型	9	3.2.3 函数依赖的基本性质	30
1.3.7 关系模型	10	3.2.4 平凡函数依赖与非平凡函数依赖	30
习题	10	3.2.5 完全依赖与部分依赖	31
第2章 关系数据库	12	3.2.6 传递依赖	31
2.1 关系模型	12	3.2.7 属性的封闭集	31
2.2 关系的形式化定义	13	3.3 范式	32
2.2.1 域 (Domain)	13	3.3.1 第一范式 (1NF)	32
2.2.2 笛卡尔积 (Cartesian Product)	13	3.3.2 第二范式 (2NF)	33
2.2.3 关系的基本性质	14	3.3.3 第三范式 (3NF)	35
2.2.4 关系模式	15	3.3.4 BC 范式 (BCNF)	36
2.3 关系的键	15	3.4 关系模式的规范化	37
2.3.1 候选关键字与主关键字	15	3.4.1 关系模式规范化的目的	38
2.3.2 主属性与非主属性	15	3.4.2 关系模式规范化的基本思想	38
2.3.3 外关键字	15	3.4.3 关系模式规范化的原则	38
2.4 关系完整性	16	3.4.4 关系模式规范化的步骤	39
2.4.1 实体完整性	16	3.4.5 分解的方法	39
2.4.2 参照完整性	16	习题	41
2.4.3 用户定义完整性	17	第4章 关系数据库标准语言 SQL	42
2.5 关系代数	17	4.1 SQL 简介	42
2.5.1 关系代数的定义、分类及运算符	17	4.2 数据定义语言 (DDL)	43
2.5.2 传统的集合运算	18	4.2.1 定义数据库	43
2.5.3 专门的关系运算	19	4.2.2 定义数据库表	49

4.3 数据操作语言 (DML)	53	6.2 数据类型	117
4.3.1 插入数据	53	6.2.1 系统提供的数据类型	118
4.3.2 更新数据	54	6.2.2 自定义数据类型	120
4.3.3 删除数据	55	6.3 变量和运算符	121
4.4 数据查询语言 (DQL)	55	6.3.1 变量	121
4.4.1 SELECT 语句的基本语法格式	55	6.3.2 运算符	122
4.4.2 简单查询	58	6.4 流程控制语句	123
4.4.3 汇总查询	70	6.4.1 BEGIN...END 语句	123
4.4.4 关联表查询	75	6.4.2 IF...ELSE 语句	124
4.4.5 连接查询	77	6.4.3 IF [NOT] EXISTS 语句	124
4.4.6 子查询	83	6.4.4 CASE 语句	124
4.5 视图	89	6.4.5 WHILE 语句	126
4.5.1 视图概述	89	6.4.6 其他流程控制语句	126
4.5.2 创建视图	90	6.5 函数	127
4.5.3 使用视图	92	6.5.1 系统内置函数	127
4.5.4 修改视图	92	6.5.2 自定义函数	134
4.5.5 删除视图	93	习题	143
习题	93	第 7 章 存储过程、触发器和游标	146
第 5 章 索引	97	7.1 存储过程	146
5.1 索引概述	97	7.1.1 存储过程的概念	146
5.2 索引的类型	98	7.1.2 存储过程的优点	147
5.3 创建索引	99	7.1.3 存储过程的分类	147
5.4 修改索引	104	7.1.4 创建存储过程	148
5.5 删除索引	105	7.1.5 执行存储过程	151
5.6 全文索引	105	7.1.6 修改存储过程	155
5.6.1 开启 SQL Full-text 服务	106	7.1.7 查看存储过程	156
5.6.2 启用全文索引	106	7.1.8 删除存储过程	157
5.6.3 创建全文目录	107	7.2 触发器	158
5.6.4 创建全文索引	108	7.2.1 触发器的定义	158
5.6.5 添加到全文索引	112	7.2.2 触发器的作用	158
习题	113	7.2.3 触发器的类型	159
第 6 章 T-SQL 程序设计	115	7.2.4 触发器的工作原理	160
6.1 T-SQL 基础	115	7.2.5 创建触发器	161
6.1.1 标识符	115	7.2.6 修改触发器	168
6.1.2 批处理	116	7.2.7 查看触发器	169
6.1.3 脚本	116	7.2.8 禁用、启用和删除触发器	170
6.1.4 注释	117	7.3 游标	172

7.3.1 游标的概念	172	第 10 章 数据的备份与恢复	220
7.3.2 游标的分类	172	10.1 数据的备份	220
7.3.3 游标的使用	173	10.1.1 数据库备份的概念	220
7.3.4 游标变量	177	10.1.2 备份设备	221
7.3.5 利用游标修改或删除数据	178	10.1.3 备份数据库	223
习题	180	10.2 数据的恢复	225
第 8 章 事务和锁	183	10.2.1 恢复策略	225
8.1 事务	183	10.2.2 恢复数据库	225
8.1.1 事务的概念	183	习题	227
8.1.2 事务的性质	183	第 11 章 数据库设计	230
8.1.3 事务的模式	184	11.1 需求分析	230
8.1.4 事务控制	186	11.1.1 需求分析的任务	230
8.1.5 分布式事务	188	11.1.2 需求分析的方法	231
8.1.6 事务隔离级别	189	11.2 概念结构设计	232
8.2 并发控制	191	11.2.1 概念模型的特点	232
8.2.1 串行执行与并发执行	191	11.2.2 概念结构设计的方法与步骤	233
8.2.2 并发导致的问题	191	11.3 逻辑结构设计	237
8.3 锁	192	11.3.1 E-R 图向关系模型的转换	237
8.3.1 锁定粒度	193	11.3.2 数据模型的优化	238
8.3.2 锁模式	194	11.4 物理结构设计	239
8.3.3 锁协议	195	11.4.1 确定数据库的存取方法	239
8.3.4 活锁与死锁	196	11.4.2 确定数据库的存储结构	239
习题	197	11.4.3 确定系统存储参数的配置	240
第 9 章 数据库的安全管理	200	11.5 数据库的实施	240
9.1 SQL Server 的安全机制	200	11.6 数据库的运行和维护	241
9.2 服务器级的安全性	200	习题	241
9.2.1 SQL Server 的身份验证模式	201	第 12 章 数据库技术的新发展	244
9.2.2 配置身份验证模式	202	12.1 影响数据库技术发展的因素	244
9.2.3 SQL Server 登录账户	202	12.2 面向对象的数据技术	245
9.2.4 服务器角色	205	12.3 分布式数据库	246
9.3 数据库级的安全性	208	12.3.1 分布式数据库系统简介	246
9.3.1 数据库用户	208	12.3.2 分布式数据库的特点	246
9.3.2 数据库角色	211	12.3.3 分布式数据库与集中式数据库 相比的优缺点	247
9.4 数据库对象级的安全性	213	12.4 多媒体数据库技术	248
9.4.1 权限类型	214	12.5 数据仓库	249
9.4.2 管理权限	215	12.6 数据挖掘技术	250
习题	218		

12.7 基于移动 Ad Hoc 无线网络的数据库技术.....	250	13.1.7 维护计划.....	276
12.8 嵌入式数据库技术.....	251	13.2 SQL Server Integration Services.....	279
习题.....	251	13.2.1 使用导入/导出向导转换数据.....	280
第 13 章 SQL Server 开发工具	253	13.2.2 SSIS 设计器.....	283
13.1 SQL Server 代理服务.....	253	13.3 SQL Server Reporting Services.....	286
13.1.1 SQL Server 代理简介.....	253	13.3.1 报表服务器项目向导.....	286
13.1.2 启用 SQL Server 代理.....	254	13.3.2 报表设计器.....	290
13.1.3 配置数据库作业.....	256	13.3.3 报表发布.....	292
13.1.4 数据库邮件.....	262	13.4 SQL Server Analysis Services.....	293
13.1.5 配置操作员.....	269	习题.....	297
13.1.6 配置警报.....	270	参考文献	299

第 1 章 数据库系统概述

【学习目标】

- 了解数据管理技术的发展过程。
- 理解数据、数据库、数据库管理系统的概念。
- 掌握建立 E-R 概念模型的基本方法。
- 了解层次模型、网状模型、关系模型。

当今时代是信息技术飞速发展的时代。所谓信息，是以数据为载体的客观世界实际存在的事物、事件或概念在人们头脑中的反映，也可以理解为被赋予特殊含义的符号、文字或者数据等在人们头脑中的反映。信息系统是以计算机为核心，以数据库为基础，对信息进行收集、组织、存储、加工、传播、管理和使用的系统。数据库能借助计算机保存和管理大量复杂的数据，快速而有效地为多个不同的用户和各种应用程序提供需要的数据，以便人们能更方便、更充分地利用这些宝贵的信息资源。

1.1 数据管理技术的发展

数据是指存储在某一种媒体上能够识别的物理符号。数据的概念在数据处理领域中已经大大拓宽了。数据不仅具有数字、字母、文字和其他特殊字符组成的文本形式，而且还具有图形、图像、动画、影像、声音等多媒体形式。

数据处理是指将数据转换成信息的过程。从数据处理的角度而言，信息是一种被加工成特定形式的数据，这种数据形式对于数据接受者来说是有意义的。通过数据处理可以获得信息，通过分析和筛选信息可以产生决策。例如，一个人的“出生日期”是有生以来不可改变的基本特征之一，属于原始数据，而“年龄”则是通过现年与出生日期相减的简单计算而得到的二次数据。根据某人的年龄、性别、职称等有关信息和离退休年龄的规定，可以判断此人何时应当办理离退休手续。

数据处理的中心问题是数据管理。数据管理是指数据的收集、整理、组织、存储、查询、维护和传输等各种操作，是数据处理的基本环节，是任何数据处理任务必有的共性部分。因此应当开发出既通用又方便好用的软件，把数据有效地管理起来，以便最大限度地减轻程序员的负担。

计算机在数据管理方面经历了由低级到高级的发展过程。它随着计算机硬件、软件技术的发展和计算机应用范围的扩大而不断发展。多年来，数据管理经历了人工管理、文件系统、数据库系统、分布式数据库系统和面向对象数据库系统等几个阶段。

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。存储设备只有卡片、纸带、磁带，

没有像磁盘这样的可以随机访问、直接存取的外部存储设备。软件方面，没有专门管理数据的软件，无操作系统与高级语言，数据由计算或处理它的程序自行携带。数据处理的方式基本上是批处理。数据管理的任务，包括存储结构、存取方式、输入/输出方式等完全由程序设计人员自行负责。

这一时期的数据管理的特点是：数据与程序不具有独立性，一组数据只对应一组程序；数据不长期保存，程序运行结束后就退出计算机系统，一个程序中的数据无法被其他程序利用，因此程序与程序之间存在大量的重复数据，称为数据冗余。

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，还用于信息管理。随着数据量的增加，数据的存储、检索和维护问题成为亟待解决的问题，这使数据结构和数据管理技术迅速发展起来。此时，外部存储设备已有磁盘、磁鼓等直接存取存储设备。软件领域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部存储器的数据管理软件。数据处理方式有批处理，也有联机实时处理。

在文件系统阶段，程序与数据有了一定的独立性，程序和数据分开存储，有了程序文件和数据文件的区别，数据文件长期保存在外存储器上多次存取。在文件系统的支持下，程序只需用文件名访问数据文件，程序员可以集中精力在数据处理的算法上，而不必关心记录在存储器上的地址和内、外存储器交换数据的过程。

文件系统阶段是数据管理技术发展中的一个重要阶段。在这一阶段中，得到充分发展的数据结构和算法丰富了计算机科学，为数据管理技术的进一步发展打下了基础，现在仍是计算机软件科学的重要基础。但是，随着数据管理规模的扩大，数据量的急剧增加，文件系统显露出了缺陷。

文件系统中的数据文件是为了满足特定业务领域或某部门的专门需要而设计的，服务于某一特定应用程序，数据和程序相互依赖，同一数据项可能重复出现在多个文件中，导致数据冗余度大。这不仅浪费存储空间，增加更新开销，更严重的是，由于不能统一修改，容易造成数据的不一致性。

文件系统存在的问题阻碍了数据处理技术的发展，不能满足日益增长的信息需求，这正是数据库技术产生的原动力。

3. 数据库系统阶段

20 世纪 60 年代后期以来，计算机用于管理的规模更为庞大，应用也越来越广泛，数据量急剧增加，同时多种应用、多种语言相交叉，共享数据的要求越来越强烈。这时硬件也飞快发展，大容量磁盘出现，硬件价格下跌，软件价格上升，为编制和维护系统软件及应用程序所需的成本相对增加，在处理方式上，联机实时处理要求更多，并开始提出和考虑分布处理。在这种背景下，以文件系统作为管理手段已不能满足应用的需求，于是为解决多用户、多应用共享数据的需求，使数据为尽可能多的应用服务，就出现了数据库技术，出现了统一管理数据的专门软件系统：数据库管理系统。

数据管理进入数据库系统阶段的标准是 20 世纪 60 年代末的 3 件大事：IMS 系统、DBTG 报告和 E.F.Codd 的文章。

(1) IMS 系统

IBM 公司研制的 IMS（信息管理系统）是一个典型的层次数据库系统。1969 年成功研制

了 IMS/I, 在 IBM360/370 机上投入运行, 1969 年 9 月投入市场, 后又于 1974 年推出 IMS/VS (虚拟存储信息管理系统, Visual System) 版本, 在操作系统 OS/VS 支持下运行。

IMS 原先是 IBM 公司为满足阿波罗计划的数据库要求而与北美洛氏 (Rockwell) 公司一起开发的。这是一个庞大的、花费资源的和有点不灵巧的系统, 但它是数据库系统中的第一个商用产品, 20 世纪 70 年代在商业、金融系统中得到了广泛的应用。我国国家计委和许多银行也曾先后采用过该系统。

(2) DBTG 报告

CODASYL 是美国数据系统语言协会 (Conference On Data System Language) 的缩写。该组织是由用户和厂商自发组织的团体, 成立于 1959 年。CODASYL 组织在 1967 年成立了一个 DBTG (Database Task Group), 专门研究数据库语言。1969 年 DBTG 提出一份报告, 即著名的“DBTG 报告”。在 1971 年 4 月, 这份报告获得通过, 它对数据库和数据操作的环境建立了标准的规范。

根据 DBTG 报告实现的系统一般称为 DBTG 系统 (或 CODASYL 系统), 它是一种网状数据库系统。现有的网状数据库系统不少均采用 DBTG 方案。DBTG 系统在 20 世纪 70 年代初期到中期得到了广泛的卓有成效的应用。

(3) E.F.Codd 的文章

第一次提出关系模型的文章是 E.F.Codd 于 1970 年在美国计算机学会通信杂志 (CACM) 上发表的《A Relation Model of Data for Large Shared Data Banks》一文。尽管 40 多年过去了, 但它仍值得再次阅读。关系数据库的许多概念都是这篇文章思想的继承和发展。这篇文章奠定了关系数据库的理论基础, 使关系数据库从一开始就建立在集合论和谓词演算的基础上。

关系模型极其简单, 因此, 它完全能为任何数据库系统提供统一的结构。交给用户用来设计数据库的逻辑结构只有一种——二维表, 用户不必涉及链接、树、图、索引等方面的复杂事情。

由于关系数据库的语言属于非过程语言, 在当时的条件下, 效率偏低, 因此, 到 20 世纪 70 年代还处于实验阶段。但随着硬件性能的改善和系统性能的提高, 在 20 世纪 80 年代, 关系数据库产品逐步投入市场, 并逐步取代层次、网状数据库产品, 成为主流产品。目前成功的产品有 SQL Server、Oracle、Sybase 和 DB2 等。

1.2 数据库技术概述

1. 数据库技术的术语

(1) 数据库

数据库 (Database, 简称 DB) 指有组织、动态存储的、结构化的、相互关联的数据的集合。它不仅包括描述事物的数据本身, 而且还包括相关事物之间的联系。

数据库中的数据具有较小的冗余和较高的数据独立性, 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, 其数据结构独立于使用数据的程序, 对于数据的增加、删除、修改和检索由系统软件进行统一的控制。

(2) 数据库管理系统

数据库能够被建立, 并能够在其中存储数据, 进而被用户或应用程序进行访问, 这些功能要依靠一种专门的软件才能够实现, 这种软件就是数据库管理系统 (Database Management

System, DBMS)。

数据库管理系统一般需要安装在一种操作系统之上,专门用来管理数据库的软件系统,位于用户与操作系统之间,对数据库的建立、运行和维护进行统一管理、统一控制。它一方面要向用户提供接口,使用户能方便地定义和操纵数据;另一方面还要能够保证数据的安全性、完整性以及多个用户对数据的并发访问与出现故障后的数据库恢复等。

人们平时经常提到的 Access、VFP、Oracle、SQL Server、Sybase、MySQL 等都是数据库管理系统。

(3) 数据库应用系统

数据库应用系统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软件系统。例如,以数据库为基础的财务管理系统、人事档案管理系统、图书管理系统等,从实现技术角度而言,它们就是以数据库为基础和核心的计算机应用系统。

(4) 数据库管理员和用户

数据库管理员(Database Administrator, DBA)是指管理和维护数据库的专业人员,其主要职责为:规划、设计数据库结构;对数据库中的数据安全性、完整性、并发控制及数据备份与恢复等进行管理和维护;监视数据库的运行,不断调整和优化内部结构,使系统保持最佳性能。

数据库系统中还有一类人员——用户,用户是数据库应用系统的使用人员,是数据库的最终用户,用户通过数据库应用系统提供的功能菜单、表格及图形用户界面等实现对数据的查询及管理工作。

(5) 数据库系统

数据库系统(Database System, DBS)是指引入数据库技术后的计算机系统,通常由数据库、数据库管理系统、数据库开发工具、应用系统、数据库管理员和用户构成,其中数据库管理系统是数据库系统的核心组成部分,如图 1-1 所示。

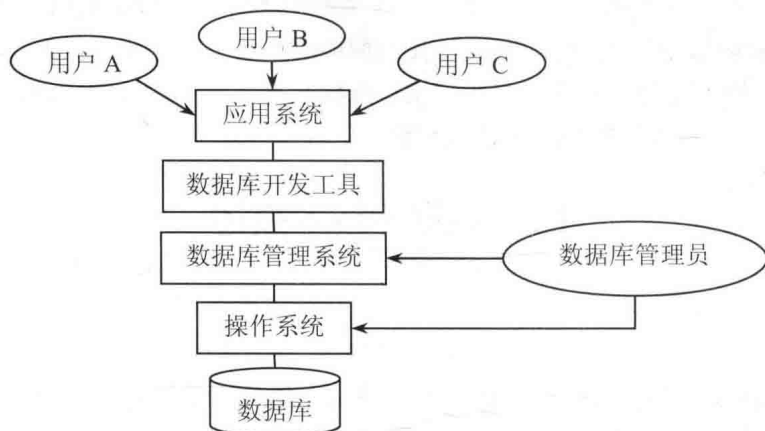


图 1-1 数据库系统组成

2. 数据库系统的特点

数据库系统克服了文件系统的缺陷,提供了对数据更高级、更有效的管理。概括起来,数据库系统具有以下特点。

(1) 采用特定的数据模型

数据库中的数据是有结构的,这种结构由数据库管理系统所支持的数据模型表现出来。

数据模型不仅描述数据本身的特征,还要描述数据之间的联系,从而反映出现实世界事物之间的联系。这样,数据不再面向特定的某个或多个应用,而是面向整个应用系统。数据冗余明显减少,实现了数据共享。任何数据库管理系统都支持一种抽象的数据模型。

(2) 具有较高的数据独立性

数据的结构分成用户的局部逻辑结构、数据库的整体逻辑结构和物理结构三级。用户(应用程序或终端用户)的数据和外存中的数据之间的转换由数据库管理系统实现。

数据独立性是指应用程序与数据库的数据结构之间相互独立。在改变物理结构时,能尽量不影响整体逻辑结构、用户的逻辑结构以及应用程序,这样就认为数据库达到了物理数据独立。在改变整体逻辑结构时,能尽量不影响用户的逻辑结构以及应用程序,这样就认为数据库达到了逻辑数据独立。

在数据库系统中,数据库管理系统提供映像功能,实现了应用程序对数据的整体逻辑结构、物理结构之间较高的独立性。用户只以简单的逻辑结构来操作数据,无需考虑数据在存储器上的物理位置与结构。

(3) 具有统一的数据控制功能

数据库系统提供了以下四个方面的数据控制功能。

数据库的并发控制:对程序的并发操作加以控制,防止数据库被破坏,杜绝提供给用户不正确的数据。

数据库的恢复:在数据库被破坏或数据不可靠时,系统有能力把数据库恢复到最近某个正确状态。

数据完整性:保证数据库中数据始终是正确的。

数据安全性:保证数据的安全,防止数据丢失或被窃取、破坏。

目前世界上已有数以百万计的数据库系统在运行,其应用已深入到人类社会生活的各个领域,从企业管理、银行业务、资源分配、经济预测一直到信息检索、档案管理、普查、统计等。现在几乎各行各业都普遍建立了以数据库技术为基础的信息系统。

1.3 数据模型

1.3.1 数据模型概念

数据库是某个企业、组织或部门所涉及的数据的存储库,它存放所有的数据并且反映数据彼此之间的联系。我们设计数据库系统时,一般先用图或表的形式抽象地反映数据彼此之间的关系,称为建立数据模型。常用的数据模型一般可分为两类,一是语义数据模型,如实体-联系模型(E-R模型)、面向对象模型等;二是经典数据模型,如层次模型、网状模型和关系模型等。第一类模型强调语义表达能力,建模容易、方便,概念简单、清晰,易于用户理解,是现实世界到信息世界的第一层抽象,是用户和数据库设计人员之间进行交流的语言。第二类模型用于机器世界,一般和实际数据库对应,例如层次模型、网状模型、关系模型分别和层次数据库、网状数据库、关系数据库对应,可在机器上实现。这类模型有更严格的形式和定义,常需加上一些限制或规定。我们设计数据库系统通常利用第一类模型作初步设计,之后用一定方法转换成第二类模型,再进一步设计全系统的数据库结构。数据模型包括数据结构、数据操

作和数据的约束条件三部分内容。

(1) 数据结构

数据结构描述的是数据库数据的组成、特征及数据相互间的联系。在数据库系统中通常按数据结构的类型来命名数据模型,如层次结构、网状结构和关系结构的模型分别命名为层次模型、网状模型和关系模型。

(2) 数据操作

数据操作指对数据库中各种对象的实例允许执行的操作的集合,包括操作及有关的操作规则。数据库的操作主要有检索和维护(包括录入、删除、修改)两大类。数据模型要定义这些操作的确切含义、操作符号、操作规则及实现操作的语言。数据结构是对系统静态特征的描述。数据操作是对系统动态特征的描述。

(3) 数据的约束条件

数据的约束条件指数据完整性规则的集合,它是给定数据模型中数据及其联系所具有的制约和依存规则,用以限定符合数据模型的数据库状态及其变化,以保证数据的完整性。

数据模型这三方面内容完整地描述了数据与数据之间的联系,数据结构是其中的首要内容。

1.3.2 实体的描述

数据库需要根据应用系统中数据的性质和内在联系,按照管理的要求来设计和组织。人们把客观存在的事物以数据的形式存储到计算机中,经历了对现实生活中事物特性的认识、概念化到计算机数据库里的具体表示的逐级抽象过程。

现实世界存在各种事物,事物与事物之间存在联系。这种联系是客观存在的,由事物本身的性质所决定。例如,图书馆中有图书和作者,读者借阅图书;学校的教学系统中有教师、学生、课程,教师为学生授课,学生选修课程并取得成绩等。如果管理的对象较多,或者比较特殊,事物之间的联系就可能较为复杂。

(1) 实体

客观存在并且可以相互区别的事物称为实体。实体可以是实际的事物,也可以是抽象的事件。例如,职工、图书等属于实际事物;订货、借阅图书、比赛等活动则是比较抽象的事件。

(2) 属性

描述实体的特征称为属性。例如,职工实体用职工号,姓名,性别,出生日期,职称等若干个属性来描述。学生实体用学号,姓名,性别,出生日期,班级等多个属性来描述。

(3) 实体集和实体型

属性值的集合表示一个具体的实体,而属性的集合表示一种实体的类型,称为实体型。同类型的实体的集合称为实体集。

例如,在职工实体集中,(0508005,张晓明,男,1978/1/1,副教授)表示一个具体的职工。在学生实体集中,(20150101001,李丽,女,1998/4/5,15 计算机科学与技术1班)则表示一个具体的学生。

1.3.3 联系

实体之间的对应关系称为联系,它反映现实世界事物之间的相互关联。例如,一位读者可以借阅若干本图书;同一本书可以相继被几个读者借阅。

实体间联系的种类是指一个实体集中可能出现的每一个实体与另一个实体集中多少个具体实体存在联系。两个实体间的联系主要归结为以下三种类型。

1. 一对一联系 (1:1)

若对于实体集 A 中的每一个实体, 实体集 B 中至多只有一个实体与之联系; 反之, 对于实体集 B 中的每一个实体, 实体集 A 中也至多只有一个实体与之联系。这称为实体集 A 与实体集 B 之间具有一对一联系, 记为 1:1。例如, 考虑公司和总经理两个实体集, 如果一个公司只有一个总经理, 一个总经理不能同时在其他公司兼任总经理。在这种情况下公司和总经理之间存在一对一联系。

2. 一对多联系 (1:N)

若对于实体集 A 中的每一个实体, 实体集 B 中可有 n 个实体 ($n \geq 0$) 与之联系; 反之, 对于实体集 B 中的每一个实体, 实体集 A 中至多只有一个实体与之联系, 则称实体集 A 与实体集 B 有一对多的联系, 记为 1:N。例如, 考虑部门和职工两个实体集, 一个部门有多名职工, 而一名职工只在一个部门就职, 部门与职工之间则存在一对多联系; 考虑学生和班级两个实体集, 一个学生只能属于一个班级, 而一个班级有很多个学生, 班级与学生之间则存在一对多联系。

一对多联系是最普遍的联系。也可以把一对一的联系看作一对多联系的一个特殊情况。

3. 多对多联系 (M:N)

若对于实体集 A 中的每一个实体, 实体集 B 中可有 n 个实体 ($n \geq 0$) 与之联系。反之, 对于实体集 B 中的每一个实体, 实体集 A 中可有 m 个实体 ($m \geq 0$) 与之联系, 则称实体集 A 与实体集 B 之间有多对多联系, 记为 M:N。

例如, 考虑学生和课程两个实体集, 一个学生可以选修多门课程, 一门课程由多个学生选修。因此, 学生和课程之间存在多对多联系。图书与读者之间也是多对多联系, 因为一位读者可以借阅若干本图书, 同一本书可以相继被多个读者借阅。

1.3.4 实体—联系模型

实体—联系模型 (E-R 模型、E-R 图) 是 P.P.Chen 于 1976 年提出的一种概念模型, 用 E-R 图来描述一个系统中的数据及其之间的关系。在 E-R 图中, 用长方形表示实体集, 在长方形框内写上实体名。用菱形表示实体间联系, 菱形框内写上联系名。用无向边把菱形和有关实体相连接, 在无向边旁标上联系的类型, 如 1 或 M 或 N。用椭圆表示实体或联系的属性, 以无向边将椭圆与一个相应实体相连接。

实体—联系模型是抽象描述现实世界的有力工具。它通过画图将实体以及实体间的联系刻画出来, 为客观事物建立概念模型。下面以某学校计算机系的教学管理系统为例, 说明实体—联系模型的建立方法。

第一步: 确定现实系统可能包含的实体。

为了简单起见, 假设教学管理系统所涉及的实体有教师、学生、课程。

第二步: 确定每个实体的属性, 特别要注明每个实体的键。

本例教学管理系统的实体包含的属性和键如下。

- (1) 对教师实体, 属性有职工号、姓名、性别、年龄、职称和专业, 其中职工号是键。
- (2) 对学生实体, 属性有学号、姓名、性别、年龄、籍贯和专业, 其中学号是键。

(3) 对课程实体, 属性有课程号、课程名、学时数、学分和教材, 其中课程号是键。

第三步: 确定实体之间可能有的联系, 并结合实际情况给每个联系命名。

本例教学管理系统的实体之间存在如下联系。

(1) 一位教师可以讲授多门课程, 一门课程可以被多位教师讲授, 这里将教师与课程之间的联系命名为讲授。

(2) 一位学生可以选修多门课程, 一门课程可以被多位学生选修。这里将学生与课程之间的联系命名为选修。

(3) 在某个时间和地点, 一位教师可指导多位学生, 但每位学生在某个时间和地点只能被一位教师指导。这里把教师和学生之间的联系命名为指导。

在对联系命名时, 一般用动词, 当用动词连接两边的实体时, 通常能表达一个符合逻辑的比较完整的意思。例如, 用动词“讲授”为教师与课程的联系命名, 并且教师“讲授”课程是一个符合逻辑的完整句子。这也是判断实体之间是否有联系和对联系命名是否恰当简单标准。

第四步: 确定每个联系的种类和可能有的属性。有时, 为了更好地刻画联系的某些特性, 需要对联系指定属性。

根据教学管理系统的实体间联系情况, 可以确定教师和课程之间的讲授联系是 M:N 联系; 学生和课程之间的选修联系是 M:N 联系, 为了更好地刻画选修的结果, 为选修联系指定成绩属性; 教师和学生之间的指导联系是 1:N 联系, 为了更好地刻画指导的环境因素, 为指导联系指定时间和地点属性。

第五步: 画 E-R 图, 建立概念模型, 完成现实世界到信息世界的第一次抽象。

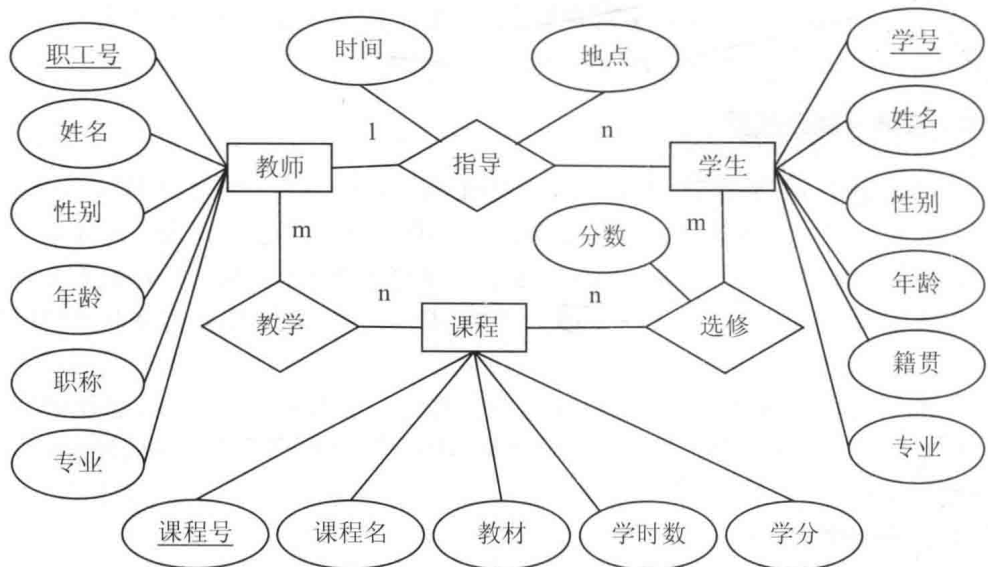


图 1-2 教学管理系统 E-R 图

在建立实体—联系模型时, 应注意以下几个问题:

- (1) 实体—联系模型要全面正确地刻画客观事物, 各类命名要清楚明了, 易于理解。
- (2) 实体中键的选择应注意确保唯一性, 即作为键的属性确实应该是那些能够唯一识别