

高等教育“十三五”规划教材



数据库技术与应用教程 (SQL Server 2014)

主 编 蒋丽影 李建东
副主编 邓长春 林 琳

SQL Server 2014

中国矿业大学出版社

China University of Mining and Technology Press

”规划教材

数据库技术与应用教程 (SQL Server 2014)

主 编 蒋丽影 李建东
副主编 邓长春 林 琳



中国矿业大学出版社

图书在版编目(CIP)数据

数据库技术与应用教程(SQL Server 2014) / 蒋丽影, 李建东主编. —徐州: 中国矿业大学出版社, 2017. 8
ISBN 978 -7 -5646 -3565 -7

I. ①数… II. ①蒋…②李… 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128573号

书 名 数据库技术与应用教程(SQL Server 2014)
主 编 蒋丽影 李建东
责任编辑 仓小金
出版发行 中国矿业大学出版社有限责任公司
(江苏省徐州市解放南路 邮编 221008)
营销热线 (0516)83885307 83884995
出版服务 (0516)83885767 83884920
网 址 <http://www.cumtp.com> E-mail: cumtpvip@cumtp.com
印 刷 徐州中矿大印发科技有限公司
开 本 787×1092 1/16 印张 15.25 字数 380 千字
版次印次 2017 年8月第 1 版 2017 年8月第 1 次印刷
定 价 30.00元

(图书出现印装质量问题, 本社负责调换)

前 言

在互联网日益发达的今天,Internet 使数据库技术、知识、技能的重要性得到了充分的发挥,数据库应用逐渐涉及社会生活的各个方面。数据库技术是现代信息科学与技术的重要组成部分,是计算机数据处理与信息管理的核心。数据库技术具有强大的事务处理功能和数据分析能力,可有效减少数据存储冗余、实现数据共享、保障数据安全以及高效地检索数据和处理数据。

SQL Server 数据库是跨平台且功能完备的网络数据库管理系统,它提供了完整的数据库创建、开发和管理功能。因其功能强大、操作简便,得到了广泛应用。此外,伴随因特网的广泛应用,SQL Server 在网络开发、网络系统集成、网络应用中也发挥着重要的作用。

SQL Server 数据库是计算机及理工类相关专业的必修课程。当前学习 SQL Server 数据库程序设计知识、掌握数据库开发应用的关键技能,已经成为网站及网络信息系统从业工作的先决和必要条件。

本书在参考了全日制高等学校本科数据库教学大纲的基础之上,同时为了顺应大学教学改革的需要,结合作者多年从事数据库类相关课程的教学体会和科研实践编写而成。本书在编写过程中,以目前较为先进的 SQL Server 2014 数据库管理系统为依托,以一个小型的数据库应用系统的整个开发流程为主线,由浅入深、贯穿始终地讲解了数据库程序设计学习及应用的基本过程和规律。本书以解析的观点、从应用的角度、站在开发与实现的立场来进行讨论。突出“实例与理论的紧密结合”,循序渐进地进行知识要点讲解,通俗易懂地讨论了关系数据库管理系统的功能、结构、设计理论和实现方法以及组织和开发过程。

本书由蒋丽影和李建东主编,邓长春、林琳为副主编,蒋丽影负责统稿。分工如下:蒋丽影编写前言、第 4 章、第 5 章;邓长春编写第 2 章、第 3 章、第 7 章;李建东编写第 8 章、第 9 章、第 10 章、第 11 章及教学课件的制作;林琳编写第 6 章;贤继红编写第 1 章。

在本书编写过程中,参阅了国内外有关 SQL Server 2014 数据库设计应用的最新书刊及相关网站资料,并得到业界专家教授的具体指导,在此一并致谢。为方便教学,本书配有电子课件,读者可以从 <http://www.cumtp.com/download?cat=resource-materials> 免费下载。

因编者水平所限,书中难免存在疏漏和不足,恳请专家、同行和读者予以批评指正。

编 者

2017 年 4 月

目 录

第 1 章 数据库系统概述	1
1.1 数据库基础知识	1
1.2 数据模型	4
1.3 关系数据库	8
1.4 数据库设计	13
1.5 初识 SQL Server 2014	17
习题	25
第 2 章 数据库的创建与管理	27
2.1 SQL Server 数据库	27
2.2 数据库的创建	29
2.3 数据库的管理	35
习题	46
第 3 章 数据表的创建与管理	48
3.1 表结构与数据类型	48
3.2 表的创建与管理	54
3.3 数据完整性	59
习题	70
第 4 章 数据表的基本操作	74
4.1 数据操纵	74
4.2 数据查询	79
习题	102
第 5 章 T-SQL 语言	105
5.1 T-SQL 语言概述	105
5.2 T-SQL 语法要素	106
5.3 常用系统函数	112
5.4 流程控制语句	117
5.5 游标	122
5.6 事务	129
习题	130

第 6 章 视图与索引	134
6.1 视图	134
6.2 索引	144
习题	154
第 7 章 存储过程与用户自定义函数	156
7.1 存储过程	156
7.2 用户自定义函数	162
习题	170
第 8 章 触发器	172
8.1 触发器概述	172
8.2 创建与使用触发器	173
8.3 修改触发器	177
8.4 禁用或启用触发器	179
8.5 删除触发器	179
习题	180
第 9 章 SQL Server 2014 的安全性	182
9.1 SQL Server 2014 的安全机制	182
9.2 SQL Server 2014 的验证模式	183
9.3 登录帐户管理	185
9.4 数据库用户管理	189
9.5 权限管理	190
9.6 角色管理	197
习题	201
第 10 章 数据库的备份、恢复与数据的导入、导出	202
10.1 备份与恢复	202
10.2 数据的导入与导出	212
习题	222
第 11 章 SQL Server 2014 综合应用实例	223
11.1 教学管理系统的需求分析	223
11.2 教学管理系统数据库设计	225
11.3 数据库的实现	229
参考文献	238

第1章 数据库系统概述

当前,人类早已迈进了信息化时代,信息作为一种战略资源,其占有和利用水平成为衡量一个国家、地区、组织或企业综合实力的一项重要标志。而在信息化社会中,人类的知识也以惊人的速度增长,如何有效地组织和利用这样庞大的知识以及如何收集、存储、加工和管理维护这个信息化社会海量的信息成为非常棘手的问题,其解决之道就是数据库技术。从20世纪60年代中期开始,计算机的应用由科学研究逐渐扩展到社会各领域,经过几十年的发展,已形成较为完整的理论体系和实用技术。本章在回顾数据库技术发展过程的基础之上,介绍与数据库技术相关的基本概念和数据库系统知识,最后介绍 SQL Server 2014 这一关系数据库管理系统的基本结构、主要功能与新特性。

1.1 数据库基础知识

数据库技术是计算机技术的重要分支,是计算机数据处理与信息管理的核心,具有强大的数据分析与处理能力。下面首先介绍数据库的基本概念,这些概念将贯穿数据处理的整个过程。

1.1.1 基本概念

1.1.1.1 数据、信息与数据处理

数据(Data)是记录客观事实的符号,它是数据库存储的基本对象。这里的“符号”不仅指数字、字母、文字和其他符号,而且还包括图形、图像、声音等。例如,“辽宁”,“87”都是数据,“87”表示某人某门课程的成绩或是某人的体重等信息;“辽宁”表示某人的籍贯。

信息是经过加工后的数据,它会对接收者的行为和决策产生影响,具有现实的或潜在的价值,如各门课程的平均分85分可以作为评定奖学金的依据。数据是信息的载体,是信息的表现形式。信息是对数据语义的解释,是经过加工处理后的有用数据。

数据处理是对数据的加工与整理,包括对数据的采集、整理、分类、存储、检索、维护、传输等操作。

数据、信息、数据处理三者之间的关系如图1-1所示。

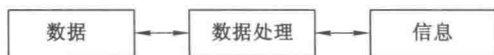


图 1-1 数据、信息、数据处理三者之间的关系

1.1.1.2 数据库

数据库(DataBase,DB)是长期存储在计算机内,有组织的、大量的、可共享的数据的集合。数据库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,并可被各种用户共享。

简单地讲,数据库数据具有永久存储、有组织和可共享 3 个基本特点。

1.1.1.3 数据库管理系统

数据库管理系统(DataBase Management System,DBMS)是位于用户与操作系统之间的数据管理软件,它为用户或应用程序提供操作数据库的接口,包括数据库的建立、使用与维护等。目前常见的大中型数据库管理系统有甲骨文公司的 Oracle、IBM 公司的 DB2、微软公司的 SQL Server、Sybase 公司的 Sybase 等,小型的数据库管理系统有微软公司的 Access、Visual FoxPro 等。

1.1.1.4 数据库应用系统

数据库应用系统(DataBase Application System,DBAS)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应用软件系统。如以数据库为基础的面向内部业务与管理的学生管理系统、图书管理系统、人事管理系统、工资管理系统等信息管理系统,以及面向外部提供信息服务的电子政务系统、电子商务系统等开放式信息系统。

1.1.1.5 数据库系统

数据库系统(DataBase System,DBS)是指在计算机系统中引入数据库后的系统,一般由数据库、数据库管理系统(及其开发工具)、应用系统、数据库管理员和用户构成。应当指出的是,数据库的建立、使用和维护等工作只靠一个 DBMS 是远远不够的,还要有专门的人员来操作和维护,这些人被称为数据库管理员(DataBase Administrator,DBA)。在一般不引起混淆的情况下常常把数据库系统简称为数据库,数据库系统的组成如图 1-2 所示。

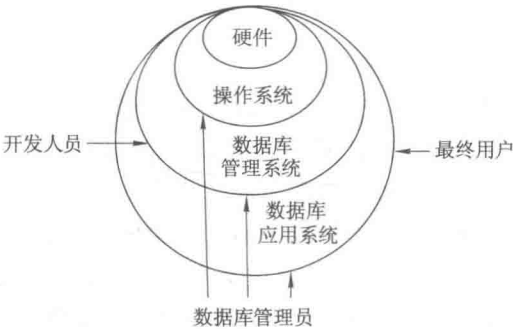


图 1-2 数据库系统的组成

1.1.2 数据库管理技术的发展

数据库技术是 20 世纪 60 年代末出现的以计算机技术为基础的数据处理技术,数据处理的核心问题是数据管理。数据管理指的是对数据进行组织、编码、分类、存储、检索与维护等操作。数据管理经历了人工管理、文件管理和数据库管理三个阶段。

1.1.2.1 人工管理阶段

在20世纪50年代中期以前,计算机主要用于科学计算。在硬件方面,外存只有纸带、卡片、磁带,没有磁盘;在软件方面,还没有操作系统,没有管理数据的软件。在人工管理阶段,由于数据量少,加上计算机硬件的限制,数据处理中的数据不需要也不允许长期保存。解决某一问题时将数据输入,用完就删除。除此之外,程序员还需要设计数据的存储结构、存取方法和输入输出方法等,这不仅使程序员负担加重,而且程序也严重依赖于数据。数据存储方式的改变必然要导致程序的修改,即数据也不具有独立性。另外,即使两个应用程序都涉及某些共同的数据,也必须各自定义,无法共享,程序之间有大量的冗余数据。人工管理阶段的示意图如图1-3所示。

1.1.2.2 文件管理阶段

从20世纪50年代后期到60年代中期,计算机的应用已扩展到数据处理领域。这期间,在硬件方面,已经有了磁盘、磁鼓等直接存取的存储设备;在软件方面,操作系统中的文件系统专门用于管理数据,文件系统不仅有批处理的处理方式,而且能够实现联机实时处理。在这一管理方式下,数据以数据文件的形式长久地保存,通过对数据文件的存取来实现对数据的查询和操纵等操作。文件系统把数据按其内容、结构和用途组织成若干个独立的数据文件,实现了“按文件名访问,按记录进行存取”的数据管理技术。文件一般为某一用户或用户组所有,仅供指定的其他用户共享。目前,如C语言等仍采用这种数据管理方式。虽然文件系统比人工管理有了很大的进步,但是仍然存在诸如数据独立性差、冗余不可避免、不支持并发访问等不足。文件系统阶段的示意图如图1-4所示。

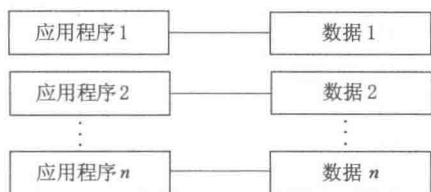


图 1-3 人工管理阶段

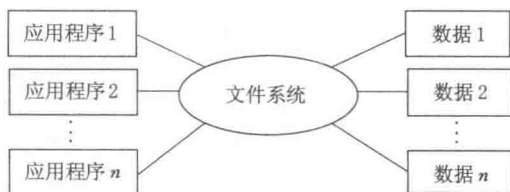


图 1-4 文件系统阶段

1.1.2.3 数据库管理阶段

数据库管理阶段是从20世纪60年代后期开始至今。这一时期,计算机性能得到很大提高,有了大容量磁盘,并且磁盘价格也急剧下降,能联机处理更多信息。人们开发出了一种新的软件系统——数据库管理系统,用户通过数据库管理系统来使用数据库中的数据。在此阶段,数据不再面向某个应用程序,而是面向整个系统,具有整体的结构性,数据与应用程序间相互独立,数据彼此联系,共享性高,冗余度小,保证了数据的一致性、完整性与安全性。数据库管理阶段的示意图如图1-5所示。

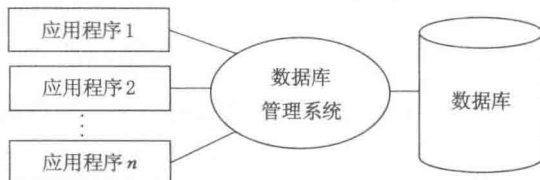


图 1-5 数据库管理阶段

20 世纪 60 年代诞生的数据库技术标志数据管理技术产生了质的飞跃。随着计算机技术与网络通信技术的发展,数据库系统结构由主机/终端的集中式结构发展到网络环境的分布式结构,Internet 环境下的浏览器/服务器结构与移动环境下的动态结构,产生了分布式数据库系统、多媒体数据库系统、面向对象数据库系统、专家数据库系统、空间数据库系统等,以满足不同应用的需求,适应不同的应用环境。

1.2 数据模型

计算机不能直接处理现实世界中的客观事物,所以人们必须事先将客观事物进行抽象、组织成计算机最终能处理的某一数据库管理系统支持的数据模型(Data Model)。

1.2.1 数据处理的 3 个阶段

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中的具体表示的逐级抽象过程,这就需要进行两级抽象,即首先把现实世界转换为概念世界,然后将概念世界转换为某一个数据库管理系统所支持的数据模型,即现实世界——概念世界——数据世界 3 个阶段。有时也将概念世界称为信息世界,将数据世界称为机器世界。其抽象过程如图 1-6 所示。

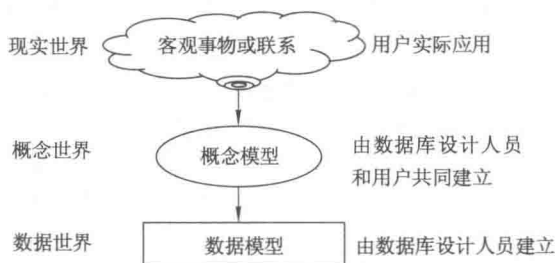


图 1-6 现实世界到数据世界的抽象过程

数据模型是现实世界中数据的抽象,它表现为一些相关数据组织的集合。在实施数据处理的不同阶段,需要使用不同的数据抽象,包括概念模型、逻辑模型和物理模型。

1.2.1.1 概念模型

概念模型也称为信息模型,是对现实世界的认识和抽象描述,按用户的观点对数据和信息进行建模,不考虑在计算机和数据库管理系统上的具体实现,所以被称为概念模型。概念模型是对客观事物及其联系的一种抽象描述,它的表示方法很多,目前较常用的是美籍华人陈平山(Peter Chen)于 1976 提出的实体—联系模型(Entity-Relationship Model, E-R 模型)。E-R 模型是现实世界到数据世界的一个中间层,它表示实体及实体间的联系,涉及的基本术语如下。

(1) 实体(Entity)

客观存在、可以相互区别的事物称为实体。实体可以是具体的对象,也可以是抽象的对象,如一个学生,一门课程,一个部门,一个比赛项目等都是实体。

(2) 属性(Attribute)

描述实体的特征称为实体的属性。如学生实体的属性有学号、姓名、性别、出生日期、简况等,课程实体的属性有课程号、课程名称、学分等。

(3) 实体型(Entity Type)

实体名与实体属性的集合表示一种实体类型,称为实体型。如学生实体的实体型表示为学生(学号,姓名,性别,出生日期,简况)。

(4) 实体集(Entity Set)

同型实体的集合称为实体集。如所有学生构成学生实体集,所有课程构成课程实体集。

(5) 联系(Relationship)

现实世界的客观事物之间是有联系的,即很多实体之间是有联系的。例如,学生和课程之间存在选课联系;教师和学生之间存在讲授联系。实体之间的联系是错综复杂的,有两个实体之间的联系,称为二元联系;也有多个实体之间的联系,称为多元联系。本书只对二元联系进行描述。

二元联系主要有以下3种情况:

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中最多有一个实体与之联系,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。例如,在学校中一个班级只有一个正班长,而一个班长只在一个班中任职,则班级与班长之间的联系就是一对一联系。

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 $N(N \geq 0)$ 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中最多有一个实体与之联系,则称实体集A与实体集B有一对多联系,记为1:N。例如,一个班级中有多名学生,而每个学生只能属于一个班级,则班级与学生之间的联系就是一对多联系。

③ 多对多联系(M:N)。如果对于实体集A中的每一个实体,实体集B中有 $N(N \geq 0)$ 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中也有 $M(M \geq 0)$ 个实体与之联系,则称实体集A与实体集B具有多对多联系,记为M:N。例如,一门课程可以同时有多个学生选修,而一个学生又可以同时选修多门课程,所以课程与学生之间的联系就是多对多联系。

实际上,一对一联系是一对多联系的特例,而一对多联系又是多对多联系的特例。

(6) E-R图

E-R图是用一种直观的图形方式建立现实世界中实体与联系模型的工具,也是进行数据库设计的一种基本工具。在E-R图中,用矩形表示现实世界中的实体,用椭圆形表示实体的属性,用菱形表示实体间的联系。实体名、属性和联系名分别写在相应的图形框内,并用线段将各框连接起来。用E-R图表示两个实体之间的3种联系如图1-7所示。

概念模型反映了实体之间的联系,是独立于具体的数据库管理系统所支持的数据模型,是各种数据模型的共同基础。

1.2.1.2 逻辑模型

逻辑模型是对应于数据世界的模型,是数据库中实体及其联系的抽象描述,数据库系统的逻辑模型不同,相应的数据库管理系统也不同。在数据库系统中,传统的逻辑模型有层次模型、网状模型和关系模型3种,非传统的逻辑模型有面向对象模型(Object-Oriented mod-

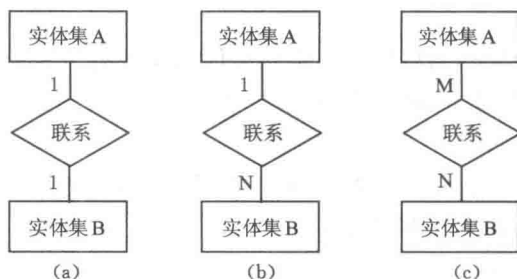


图 1-7 两个实体之间的联系

(a) 1:1 联系;(b) 1:N 联系;(c) M:N 联系

el,OO)。

客观事物是信息之源,是设计、建立数据库的基础,也是使用数据库的目的。概念模型和逻辑模型是对客观事物及其相互关系的两种描述,实现了数据处理 3 个阶段的对应转换。

逻辑模型中的数据描述如下:

(1) 字段(field)

标记实体属性的命名单位称为字段或数据项。由于它是可以命名的最小信息单位,所以又被称为数据元素或数据项。字段的命名往往和属性名相同,例如学生有学号,姓名,性别,出生日期等字段。

(2) 记录(record)

字段的有序集合称为记录。通常用一个记录描述一个实体,所以记录又可以定义为能完整地描述一个实体的字段集。例如一个学生记录由有序的字段集组成,即(学号,姓名,性别,出生日期……)。

(3) 文件(file)

同一类记录的集合称为文件,文件是用来描述实体集的。例如所有的学生记录组成了一个学生文件。

(4) 关键码(key)

能唯一标识文件中每个记录的字段或字段集,称为记录的关键码,简称为键。

1.2.1.3 物理模型

物理模型用于描述数据在物理存储介质上的组织结构,与具体的数据库管理系统、操作系统和计算机硬件有关。

从概念模型到逻辑模型的转换是由数据库设计人员完成的,从逻辑模型到物理模型的转换是由数据库管理系统完成的。因此,一般人员不必考虑物理实现的细节。

1.2.2 常见的数据模型

数据模型即上面所述逻辑模型,它是数据库系统中的一个关键内容。数据模型不同,相应的数据库管理系统就完全不同,任何一个数据库管理系统都是基于某种数据模型的。数据库管理系统常用的数据模型主要有层次模型、网状模型和关系模型。

1.2.2.1 层次模型

用树形结构表示实体及其之间联系的模型称为层次数据模型。这样的树由结点和连线组成,结点表示实体集,连线表示两实体之间的联系,树形结构只能表示一对多联系。通常将表示“一”的实体放在上方,称为父结点;表示“多”的实体放在下方,称为子结点。树的最高位置只有一个结点,称为根结点。根结点以外的其他结点都有一个父结点与它相连,同时可能有一个或多个子结点与它相连。没有子结点的结点称为叶,它处于分支的末端,树形层次模型有两个特点:

① 树的最高结点,即根结点只有一个。

② 根结点以外的其他结点都与一个且只与一个父结点相连,如图 1-8 所示。

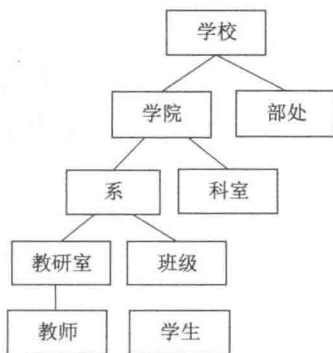


图 1-8 树形层次模型

采用层次模型的数据库是最早出现的,它的典型代表是 IBM 公司的 IMS(Information Management System)系统。该系统于 1968 年问世,是世界上第一个数据库管理系统。

1.2.2.2 网状模型

用网状结构来表示实体及实体之间联系的模型称为网状数据模型,它是对层次模型的发展,能够更直接地描述现实世界的多对多联系。网状模型克服了层次模型的两个限制,即可以有任意个结点无父结点,同时允许一个结点有多个父结点。网状模型的示例如图 1-9 所示。

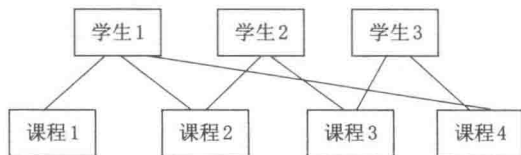


图 1-9 网状模型

网状模型更适合于表达客观事物复杂的隶属关系和横向联系,但其结构也更加复杂,记录的存取也往往不是唯一的。从网状模型与层次模型的相互关系看,后者仅是前者的一种特殊情况。

网状模型中最有代表性的是美国数据系统语言协会的下属机构“数据库任务组”于 1969 年提出的 DBTG 报告中给出的模型。

1.2.2.3 关系模型

用二维表表示实体以及实体之间联系的模型称为关系模型,它是目前应用最广泛的一种数据类型。支持关系模型的数据库管理系统称为关系数据库管理系统。现在几乎所有流行的数据库管理系统都是关系数据库系统,如 Oracle、Sybase、SQL Server、Informix 等。关系模型中采用规范化的二维表表示实体及实体间的联系,关系模型的操作对象与结果都是二维表。关系模型结构简单、概念单一,插入、修改、删除操作方便,但查询效率较低。关系模型示例如图 1-10 所示。



图 1-10 关系模型

1.3 关系数据库

关系数据库是目前应用最为广泛的主流数据库,由于它以数学方法为基础管理并处理数据库中的数据,所以关系数据库与其他数据库相比具有比较突出的优点。20 世纪 80 年代以来,数据库厂商新推出的数据库管理系统产品主要以关系数据库为主,非关系数据库产品也大多添加了关系接口。正是关系数据库的出现和发展,促进了数据库应用领域的扩大和深入,因此研究和学习关系数据库的理论、技术和应用十分重要。

1.3.1 关系术语

(1) 关系

一个关系就是一张规范化的二维表,每个关系都有一个关系名,即表名。

(2) 元组

二维表中的行称为元组或记录,即一个实体的各个属性值的集合。元组表现为一个二维表中的数据,一个关系有多条记录。

(3) 属性

二维表中的每一列在关系中称为属性或字段,每一列的标题称为属性名或字段名,列的值称为属性值或字段值。

(4) 关系模式

关系模式是对关系的描述,由关系名与组成该关系的所有属性名构成,表示为关系名(属性名 1,属性名 2,……,属性名 n),在不引起混淆的情况下,也常称关系名为关系模式或关系,例如关系模式 $R(A, B, C)$,也称为关系 R 或关系模式 R。关系模式表现为一个二维表的结构。

(5) 域

属性的取值范围称为该属性的域。如性别的取值范围为“男”或“女”,成绩的取值范围为 0~100。

(6) 键

在关系数据库中,键(key)也称为码或关键字,它通常由一个或几个属性组成,能唯一地表示一个元组。

① 超键(super key)。在一个关系中,能唯一标识元组的属性或属性组称为关系的

超键。

② 候选键(candidate key)。如果一个属性组能唯一标识元组,且不含有多余的属性,那么这个属性组称为关系的候选键或称候选关键字。

③ 主键(primary key,PK)。若一个关系中有多个候选键,则选择其中的一个为关系的主键又称主关键字。通常用主键实现关系定义中“表中任意两行(元组)不能相同”的约束。包含在任何一个候选键中的属性称为主属性,不包含在任何一个键中的属性称为非主属性或非键属性。

例如,对于关系模式学生(学号,姓名,性别,出生日期),考虑一个班级的情况,假设没有重名,则学号,姓名,(姓名,性别),(姓名,性别,出生日期)等都能唯一标识学生记录,所以都是超键;而在这些超键当中,学号、姓名都能唯一标识学生记录且没有多余属性,所以都是候选键。关系的候选键可以有多个,但不能同时使用,一次只能使用其中的一个。如考虑输入查询的方便性,可以选择学号为主键。

④ 外键(foreign key,FK)。若一个关系中的属性或属性组合是另一个关系的主键或候选键时,称该属性或属性组合为当前关系的外键又称外关键字。通过外键可实现两个表的联系。

1.3.2 关系特点

在关系模型中,关系必须具有以下特点:

- ① 关系必须规范化,每个属性必须是不可再细分的数据项,即表中不能再包含表。
- ② 每列具有相同的数据类型、相同的域。
- ③ 每一列的标题不能相同,即属性名不能重复。
- ④ 任意两行的内容不能完全相同,即元组不能重复。
- ⑤ 在一个关系中,元组的次序无关紧要。
- ⑥ 在一个关系中,列的次序也无关紧要,即可以任意交换两行、两列的次序。

这里特别提醒读者注意:关系、元组及属性等是数学领域中的术语;二维表、行、列等均为日常用语;而数据文件、数据记录、数据项等则是计算机领域中的术语。这些术语是相互对应的,只不过因应用领域有所不同而称呼不同,其含义是完全相同的。

1.3.3 关系模式的规范化

关系模式是否具备以上的关系特点就可以了呢?设有选课关系模式,选课(学号,课程编号,姓名,性别,班级,班主任,课程名称,学分,成绩),由于成绩由学号与课程编号所决定,则该关系模式的主键为“学号,课程编号”。该关系模式的具体数据如表 1-1 所示。

表 1-1 选课表

学号	课程编号	姓名	性别	班级	班主任	课程名称	学分	成绩
1001	C101	王一	女	151 班	张平	大学英语	4	90
1001	C103	王一	女	151 班	张平	马克思理论	2	89
1002	C102	李男	男	152 班	刘辉	高等数学	5	85
1002	C101	李男	男	152 班	刘辉	大学英语	4	78
1002	C104	李男	男	152 班	刘辉	C 语言程序设计	4	83

从表 1-1 中的数据可见,该关系存在以下问题。

① 数据冗余。如果一个学生选修了多门课程,这个学生的信息(学号,姓名,性别……)就会重复多次。同样,如果一门课程有多人选修,则课程信息(课程编号,课程名称,学分)也将重复多次。

② 插入异常。由于主关键字(学号,课程编号)的值不能为空,当添加一个没有选课的学生信息时就会引起插入异常。同样,当添加一门无人选修的新课时也会出现相同的问题。

③ 更新异常。由于存在数据冗余,当更新信息时,需要将所有重复的信息同时更新,如更新学号为“1001”的学生姓名,当有一个元组没有更新时,便会造成数据不一致的现象。

④ 删除异常。当要删除学生信息时,可能造成课程信息被彻底删除。如表 1-1 中,删除“王一”的信息时,“马克思理论”和“大学英语”的课程信息被彻底删除了,引起删除异常。

由此可见,选课关系模式并不是一个合理、有效的关系模式。关系模式需要在满足关系特点的基础上做进一步的规范化处理。规范化是指按照统一的标准对关系进行优化,以提高关系的质量,为构造一个高效的数据库应用系统打下基础。

关系模式的规范化可以分为几个等级,每一个等级称为一个范式。如第一范式(1NF)、第二范式(2NF)、第三范式(3NF)……每一范式比前一范式的要求更为严格,即范式之间存在 $1NF \subseteq 2NF \subseteq 3NF \dots$ 的关系,通常满足第三范式即可。

1.3.3.1 第一范式(First Normal Form,1NF)

第一范式是最基本的要求,即关系模式的所有属性都是不可再分的数据项。如果关系模式 R 的所有属性都是不可再分的,则称 R 满足第一范式,记做 $R \in 1NF$ 。满足第一范式的关系称为规范化关系;否则称为非规范化关系。非规范化关系示例如表 1-2 所示。

表 1-2 非规范化关系

用户名	地址			电话
	省	市	县	

1.3.3.2 第二范式(Second Normal Form,2NF)

如果一个关系模式 R 满足第一范式,且每个非主属性完全函数依赖于主关键字,则称 R 满足第二范式,记做 $R \in 2NF$ 。

第二范式要求实体的非主属性完全依赖于主关键字。完全依赖是指不能存在仅依赖主关键字一部分的属性,如果存在,这个属性和主关键字的这一部分应该分离出来形成一个新的实体,新实体与原实体之间是一对多的关系。

例如,在选课(学号,课程编号,姓名,性别,班级,班主任,课程名称,学分,成绩)关系模式中,成绩属性完全依赖于主关键字(学号,课程编号),姓名、性别、班级、班主任属性依赖于主关键字中的学号,即存在部分依赖;课程名称、学分属性依赖于主关键字中的课程编号,也

存在部分依赖。所以选课模式不满足第二范式,可分解如下:

学生(学号,姓名,性别,班级,班主任)

课程(课程编号,课程名称,学分)

选课(学号,课程编号,成绩)

以上3个关系模式均满足第二范式,但学生关系模式仍存在数据冗余,如一个班级有多名学生时,班主任的信息就会重复多次。

1.3.3.3 第三范式(Third Normal Form, 3NF)

如果关系模式 R 满足第二范式,且每个非主属性都不传递函数依赖于 R 的主关键字,则称 R 满足第三范式,记做 $R \in 3NF$ 。

第三范式要求实体的非主属性不传递依赖于主关键字。传递依赖指的是如果存在“ $A \rightarrow B \rightarrow C$ ”的决定关系,则 C 传递依赖于 A 。

例如,在学生(学号,姓名,性别,班级,班主任)关系模式中,存在学号 $\rightarrow$ 班级 $\rightarrow$ 班主任的决定关系,所以学生关系模式不满足第三范式,可分解如下:

学生(学号,姓名,性别,班级)

班级(班级,班主任)

将关系模式分解到第三范式,可以在相当程度上减轻数据冗余。但在实际设计中,完全消除冗余是很难做到的,有时为了提高数据检索等处理效率,也允许存在适当的冗余。

1.3.4 关系运算

在对关系数据库进行访问,希望找到所需要的数据时,就要对关系进行运算。关系运算有两类:一类是传统的集合运算,另一类是专门的关系运算。关系运算的操作对象是关系,结果也是关系。

1.3.4.1 传统的集合运算

传统的集合运算包括并($\cup$)、交($\cap$)、差($-$)、笛卡儿积($\times$)4种。其中并、交、差运算要求参加运算的两个关系具有相同的模式,即具有相同的结构。

(1) 并

两个相同结构关系的并是由这两个关系的元组组成的集合。

(2) 交

设有两个关系 R 和 S , R 交 S 的结果是由既属于 R 又属于 S 的记录组成的集合。即交运算的结果是 R 和 S 中共同的记录。

(3) 差

设有两个关系 R 和 S , R 差 S 的结果是由属于 R 但不属于 S 的记录组成的集合。即差运算的结果是从 R 中去掉 S 中也有的记录。

例如 R 和 S 这两个关系中分别存放选修大学英语与高等数学课程的学生信息,则 R 与 S 并、交、差的集合运算如图 1-11 所示。

(4) 笛卡儿积

这里的笛卡儿积严格地讲应该是广义的笛卡儿积,因为这里的笛卡儿积的元素是元组。

如果关系 R_1 有 m 列,关系 R_2 有 n 列,则 R_1 与 R_2 的广义笛卡儿积记做 $R_1 \times R_2$,是一个