

高等职业教育

计算机类专业 规划教材

INFORMATION TECHNOLOGY

网络数据库SQL Server 2005教程

平震宇 主 编
臧武军 李朝晖 副主编

8SQ



中国电力出版社

<http://jc.cepp.com.cn>

高等职业教育

计算机类专业 规划教材

INFORMATION TECHNOLOGY

网络数据库SQL Server 2005教程

主 编 平震宇

副主编 臧武军 李朝晖

主 审 郭力子



中国电力出版社

<http://jc.cepp.com.cn>

内 容 提 要

本书为高等职业教育计算机类专业规划教材。本书共分 14 章, 分别介绍了数据库基础知识、数据库创建与维护、表的创建与数据完整性、数据查询、索引、视图、存储过程、触发器、安全性、数据备份与恢复及 ADO 数据访问技术等内容。本书内容循序渐进, 任务丰富生动, 以任务驱动引导知识点的学习, 具有较强的趣味性和可操作性。

本书可作高职高专软件技术、计算机网络技术、电子商务、经济管理等专业的网络数据库程序设计课程的教材, 也可供相关工程技术人员参考。

图书在版编目 (CIP) 数据

网络数据库 SQL Server 2005 教程 / 平震宇主编. —北京:
中国电力出版社, 2010

高等职业教育计算机类专业规划教材

ISBN 978-7-5083-9856-3

I. ①网… II. ①平… III. ①关系数据库—数据库管理系统,
SQL Server 2005—高等学校: 技术学校—教材 IV. ①TP311.138

中国版本图书馆 CIP 数据核字 (2009) 第 225085 号

中国电力出版社出版、发行

(北京三里河路 6 号 100044 <http://jc.cepp.com.cn>)

汇鑫印务有限公司印刷

各地新华书店经售

*

2010 年 1 月第一版 2010 年 1 月北京第一次印刷
787 毫米×1092 毫米 16 开本 14.75 印张 358 千字
印数 0001—3000 册 定价 24.00 元

敬告读者

本书封面贴有防伪标签, 加热后中心图案消失
本书如有印装质量问题, 我社发行部负责退换

版 权 专 有 翻 印 必 究

前言

数据库技术是计算机领域中最重要技术之一,是计算机软件学科中的一个独立分支。SQL Server 是基于客户/服务器体系结构的关系数据库管理系统,在电子商务、数据仓库等应用中起着重要的作用,可以为企业的数据库管理提供强大的支持。《网络数据库 SQL Server》课程的显著特点是实践性强,本教材以“工作过程导向”和“项目式学习”为主线,每个模块单元都是围绕典型项目展开。本书以提高学生的技能水平为目的,根据高职高专教育和教学特点,结合教学改革和应用实践修订而成的,可作为高职高专计算机软件、计算机网络、电子商务、经济管理等专业的网络数据库程序设计课程的教材。

本教材根据实际工作过程,分解学习领域为主题学习单元;将学习领域中的能力目标和学习内容分解为主题学习单元,如项目、任务、案例等,根据长期的项目开发、教学研究与实践所积累的认识与经验,针对高职学生的认知特点,不过分追求知识、技能的“完整性”和“系统性”,从实际应用的需求出发,提炼、整合最基本、最核心的内容;使用项目实例带动知识点,即把知识点的讲授贯穿在实际应用项目的分析、实践之中,指导学生掌握面对真实的数据库应用系统时的发现问题、分析问题、解决问题的能力。

本教材把数据库课程与网络应用及编程类课程结合起来,使学生能够掌握开发数据库应用软件的相关技术。包含的知识模块有:数据库基础理论、数据库管理系统软件、SQL 语言、数据库编程接口、数据库应用程序开发实例。

全书共 14 章,全面介绍了 SQL Server 数据库系统的管理、使用与开发。第 1 章数据库系统的概念与规划;第 2 章 SQL Server 概述;第 3 章创建数据库及表;第 4 章数据库完整性;第 5 章查询与视图;第 6 章索引;第 7 章 Transact-SQL 及其程序设计;第 8 章实现存储过程;第 9 章实现触发器;第 10 章游标与事务;第 11 章 SQL Server 的安全性;第 12 章备份与还原;第 13 章 ADO.NET 2.0 介绍;第 14 章数据库应用系统开发实例。本课程建议参考教学学时数为 64~80 学时。其中,授课学时为 38 学时,实训学时为 26~42 学时。本书由郭力子主审。

在本书编写、修订、成稿的过程中,除封面署名作者外,还得到了编者所在学院和系部的领导、同事及朋友的帮助和大力支持,在此向他们的辛勤劳动表示衷心的感谢!

限于编者水平,书中存在错误和不妥之处,殷切期望广大读者给予批评指正。

编者

2009 年 10 月于无锡

目 录

前言

第 1 章 数据库系统的概念与规划	1
1.1 数据库系统的基本概念	1
1.2 数据模型	2
1.3 关系数据模型	5
1.4 项目实例 1——人力资源管理系统数据库设计	9
小结	15
习题	15
第 2 章 SQL Server 概述	16
2.1 SQL Server 简介	16
2.2 SQL Server 2005 的新特性	16
2.3 SQL Server 常用工具及其组件	18
2.4 项目实例 2——安装 SQL Server 2005	22
小结	28
习题	28
第 3 章 创建数据库及表	29
3.1 SQL Server 数据库概念	29
3.2 数据库操作	31
3.3 操作数据库表	37
3.4 维护数据表	46
3.5 项目实例 3——人力资源管理数据库操作	49
3.6 独立实践项目	52
小结	52
习题	53
第 4 章 数据库完整性	54
4.1 数据完整性的基础知识	54
4.2 约束	55
4.3 默认值	62

241	4.4	规则	64
241	4.5	项目实例 4——实现数据完整性	67
241	4.6	独立实践项目	75
041		小结	77
021		习题	77
	第 5 章	查询与视图	78
221	5.1	使用 SELECT 语句查询数据	78
221	5.2	数据分组和汇总	83
021	5.3	多表连接与子查询	86
021	5.4	视图的基本知识	90
001	5.5	创建视图	92
101	5.6	独立实践项目	97
101		小结	99
201		习题	99
	第 6 章	索引	100
001	6.1	索引基本知识	100
301	6.2	创建索引	104
031	6.3	操作索引	109
051	6.4	联机索引	112
121	6.5	查询优化器	115
121	6.6	项目实例 5——创建索引	117
281		小结	119
281		习题	119
	第 7 章	Transact-SQL 及其程序设计	120
101	7.1	Transact-SQL 语言基础	120
001	7.2	常量与变量	122
002	7.3	流程控制语句	124
002	7.4	系统函数	127
102	7.5	用户自定义函数	129
102	7.6	项目实例 6——Transact-SQL 程序设计	131
002		小结	133
112		习题	134
	第 8 章	实现存储过程	135
012	8.1	存储过程的基本概念	135
012	8.2	实现存储过程	137
012	8.3	存储过程中参数的使用	140
022	8.4	项目实例 7——实现存储过程	142
822	8.5	独立实践项目	143
822		小结	144
022		习题	144

第 9 章 实现触发器	145
9.1 触发器介绍	145
9.2 创建 DML 触发器	147
9.3 创建 DDL 触发器	149
9.4 查看、修改和删除触发器	150
9.5 项目实例 8——实现触发器	152
小结	155
习题	155
第 10 章 游标与事务	156
10.1 游标	156
10.2 事务	160
10.3 项目实例 9——实现游标和事务	163
小结	164
习题	165
第 11 章 SQL Server 的安全性	166
11.1 安全性基础知识	166
11.2 管理服务器的安全性	168
11.3 数据库的安全性	173
11.4 数据库对象的安全性	179
小结	181
习题	181
第 12 章 备份与还原	182
12.1 备份与还原简介	182
12.2 备份数据库	184
12.3 还原数据库	191
12.4 自动备份维护计划	196
小结	200
习题	200
第 13 章 ADO.NET 2.0 介绍	201
13.1 ADO.NET 概述	201
13.2 项目实例 10——生成 ADO.NET 应用程序	206
小结	213
习题	213
第 14 章 数据库应用系统开发实例	214
14.1 数据库应用系统开发流程	214
14.2 项目实例 11——进销存管理系统开发实例	216
14.3 独立实践项目	226
小结	228
习题	228
参考文献	229

第 1 章

数据库系统的概念与规划

本章要点:

- ☒ 数据库系统的基本概念
- ☒ 概念模型
- ☒ 关系数据库模型

1.1 数据库系统的基本概念

计算机数据管理,就是指利用计算机对数据进行分类、组织、编码、存储、检索和维护,是数据处理的中心问题。计算机数据管理大致经历了 3 个阶段:人工管理阶段(从 20 世纪 40 年代中期到 20 世纪 50 年代中期),文件系统阶段(从 20 世纪 50 年代末期到 20 世纪 60 年代中期),数据库系统阶段(从 20 世纪 60 年代末期到现在)。

在数据库系统阶段,数据管理具有如下特点:

(1) 数据的共享程度高。数据的高度共享可以有效地降低数据的冗余度,节省存储空间,避免数据间的不一致性,还有利于系统扩充。

(2) 数据的独立性高。在数据库系统阶段,数据管理具有高度的物理独立性和一定的逻辑独立性。物理独立性是指用户的应用程序与存储在磁盘上的数据库中的数据是相互独立的,数据的物理存储改变了,应用程序不用改变。逻辑独立性是指用户的应用程序与数据库的逻辑结构是相互独立的,数据的逻辑结构改变了,用户程序也可以不变。

(3) 数据结构化。在文件系统中,单个文件是有结构的,但整个系统在整体上又缺乏结构性。数据库系统中同一数据库中的文件是有联系的,在整体上也服从一定的结构形式。

(4) 数据控制能力强。利用数据库管理系统进行统一的数据管理和数据控制,具体提供了数据的安全性保护、完整性检查、并发访问控制、数据恢复能力。

数据、数据库、数据库管理系统、数据库系统是数据库技术中密切相关的 4 个基本概念。

1.1.1 数据

数据是对客观事物的描述,是数据库中存储的基本对象。数据的概念在数据处理领域中比科学计算领域中大大地扩展了,如数字、文本、图形、图像、声音等多种可被计算机存储的信息都可以作为数据。在计算机中,为了存储和处理这些客观事物,需要对事物的主要特征进行抽象,形成由若干符号组成的一条条的记录来描述事物。例如,在图书管理系统中,人们通常关心的是图书的书名、作者、出版社、出版日期等信息,可以这样描述图书:

(数据库程序设计-SQL Server 2000 数据库程序设计, 微软公司, 人民邮电出版社, 2004)

上述数据中包含的信息是: 人民邮电出版社在 2004 出版的由微软公司编写的《数据库程序设计-SQL Server 2000 数据库程序设计》一书。从上面的解释中可以看出, 数据的形式还不能完全表达其内容, 需要结合语义来解释。如果仅有记录符号, 而缺乏相应的语义环境, 数据中包含的信息也就不能充分被人理解, 从而达不到信息交流的目的。

1.1.2 数据库

数据库 (DataBase) 是长期储存在计算机内、有组织的、可共享的大量数据集合, 是存放数据的仓库, 这个仓库是在计算机存储设备上, 如硬盘。人们收集并抽取出一个应用所需要的大量数据之后, 应将其保存起来以供进一步加工处理, 进一步抽取有用的信息, 这就形成了数据库。数据必须按照一定的格式存放, 以利于以后应用。可以认为数据库是长期存储在计算机内、与应用彼此独立的、以一定的组织方式存储在一起的、彼此相互关联的、具有较少冗余的、能被多个用户共享的数据的集合。

1.1.3 数据库管理系统

数据库是依靠数据库管理系统 (DataBase Management System, DBMS) 来实现管理数据的功能的。数据库管理系统是位于用户与操作系统之间的数据管理软件, 为用户提供科学地组织和存储数据、高效地获取和维护数据的功能。它提供应用程序与数据库的接口, 使用户能够方便、快速地建立、维护、查询、存取和处理数据库中的信息。它的主要功能如下:

(1) 数据定义功能。DBMS 提供数据定义语言 (Data Definition Language), 用户可以方便地定义数据库中的各种数据对象, 如数据库、表等。

(2) 数据操纵功能。DBMS 同时提供数据操纵语言 (Data Manipulation Language), 用户可以操纵数据库中的数据, 实现对数据库的基本操作, 如从数据库中查询符合条件的数据、向数据库中插入数据、从数据库中删除和修改数据等。

(3) 数据库的运行管理。数据库在建立、运行和维护时由 DBMS 进行统一管理、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。包括数据库数据批量载入、数据库的转储、介质故障的恢复、数据库的重组、性能监视与分析等。这些功能通常是由一些实用程序完成的。

1.1.4 数据库系统

数据库系统 (DataBase System) 是指在计算机系统中引入数据库后的系统构成。数据库系统一般由数据库、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, 其层次关系如图 1-1 所示。

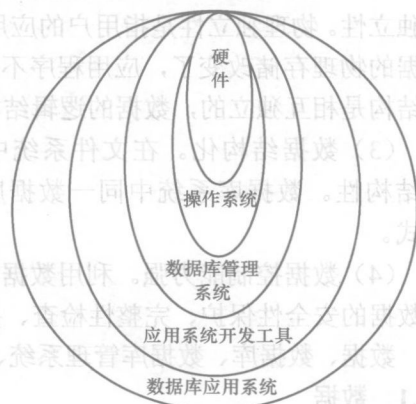


图 1-1 数据库系统的层次结构

1.2 数据模型

模型是人们对现实世界特征的模拟和抽象。要将现实世界中的信息转变为机器能识别的

数据形式, 需要使用数据模型来完成这个模拟和抽象的过程。数据模型应满足三方面要求: 能比较真实地模拟现实世界; 容易为人所理解; 便于在计算机上实现。

数据模型按照抽象的层次不同又可分为概念模型与数据模型。

1.2.1 概念模型的基本概念

概念模型是对信息世界建模, 应该能够方便、准确地表示出现实世界中的常用概念。

1. 概念模型

概念模型反映现实世界。概念模型中的概念主要有以下几种:

(1) 实体 (Entity)。客观存在并可相互区别的事物称为实体。从具体的人、物、事件到抽象的状态与联系都可以用实体来抽象地描述。例如, 在图书借阅中, 读者、图书及读者借书、还书等都可看做是实体。

(2) 属性 (Attribute)。实体所具有的某一特性称为属性。一个属性描述实体的某一方面性质, 一个实体正是通过若干个属性来刻画的。如一个读者由学号、姓名、性别、年龄、地址、电话等属性来描述。实体究竟具有哪些属性则取决于人们对实体的哪些特性感兴趣。

(3) 码 (Key)。能够唯一标识实体的属性集称为实体的码。如学号是读者实体中能够唯一确定是哪位读者的属性, 一个学校里不可能有两个学生具有相同的学号。

(4) 域 (Domain)。属性的取值范围称为该属性的域。如读者的年龄只能在 $0 \sim 150$ 之间。

(5) 联系 (Relationship)。现实世界中事物内部以及事物之间的联系在信息世界中反映为实体内部的联系和实体之间的联系。例如, 图书实体与读者实体之间发生着借书或还书联系。

2. 实体的联系

实体的联系可分为: 一对一联系 (1:1)、一对多联系 (1:M) 和多对多联系 (M:N), 如图 1-2 所示。

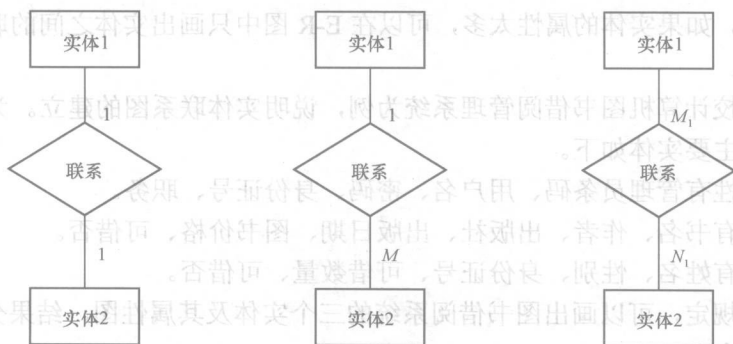


图 1-2 实体间的联系

(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, 班级与班长之间的联系: 一个班级只有一个正班长, 一个班长只在一个班中任职。

(2) 一对多联系 (1:M)。如果对于实体集 A 中的每一个实体, 实体集 B 中有 M 个实体 ($M \geq 0$) 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 有一对多联系, 记为 1:M。例如, 班级与学生之间的联系:

一个班级中有若干名学生，每个学生只在一个班级中学习。

(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 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 中也有 M 个实体 ($M \geq 0$) 与之联系，则称实体集 A 与实体 B 具有多对多联系，记为 $M:N$ 。例如，读者与图书之间的联系：一个读者可以借多本不同的图书，一本图书也可以在不同的时候被多个读者所借阅。

联系还可以在多个实体间发生。例如，在借书联系中就涉及读者、图书、管理员三个实体，它们的联系如图 1-3 所示。



图 1-3 多个实体间发生联系

1.2.2 概念模型的图形化描述

概念模型的表示方法很多，其中最常用的是实体-联系方法，也叫 E-R 方法，由 E-R 图来表示。E-R 图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联系刻画出来，为客观事物建立概念模型。在 E-R 图中规定如下：

(1) 用长方形表示实体，在框内写上实体名。

(2) 用椭圆形表示实体的属性，并用线段把实体与其属性连接起来。

(3) 用菱形表示实体之间的联系，菱形内写上联系名。用线段把菱形与有关的实体连接起来，在线段旁标上联系的类型。若实体之间的联系也有属性，则把属性和菱形也用线段连接起来。

为简洁起见，如果实体的属性太多，可以在 E-R 图中只画出实体之间的联系，而将实体及属性单独画出。

下面以某学校计算机图书借阅管理系统为例，说明实体联系图的建立。为简单起见，假设该系统涉及的主要实体如下。

管理员：属性有管理员条码、用户名、密码、身份证号、职务。

图书：属性有书名、作者、出版社、出版日期、图书价格、可借否。

读者：属性有姓名、性别、身份证号、可借数量、可借否。

根据上述的规定，可以画出图书借阅系统的三个实体及其属性图，结果分别如图 1-4、图 1-5 和图 1-6 所示。

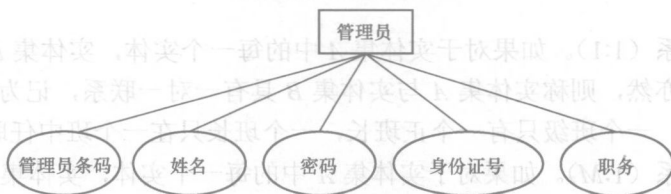


图 1-4 管理员实体及其属性

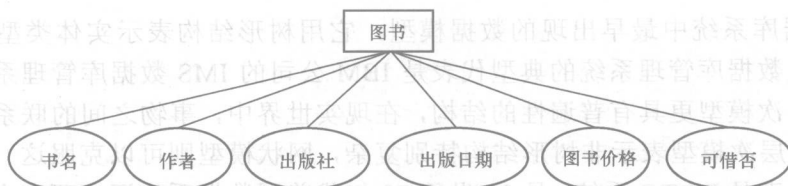


图 1-5 图书实体及其属性

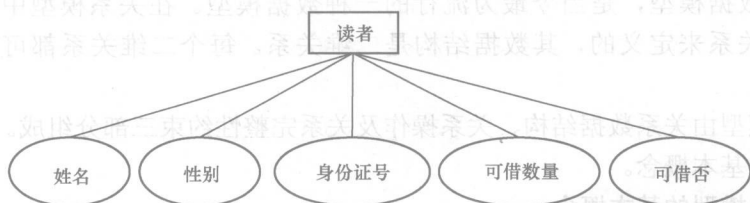


图 1-6 读者实体及其属性

假设系统实体间的联系描述如下：

管理员对图书的管理，包括对图书的增加、删除、修改等操作，涉及管理员与图书之间的联系；一个管理员可以管理多本图书，一本图书也可以由多个管理员进行管理；它们之间是多对多的联系；联系的属性除了管理员和图书的条码之外，还包括管理的时间。

管理员对读者的管理，包括对读者的增加、删除、修改等操作，涉及管理员与读者之间的联系；一个管理员可以管理多个读者，一个读者也可以由多个管理员进行管理；它们之间是多对多的关系；联系的属性除了管理员和读者的条码之外，还包括管理的时间。

读者通过管理员进行图书的借阅与还书，该联系涉及读者、图书及管理员。一本图书只能被一个读者借阅，一个读者可借多本图书，它们之间是一对多的关系。一个管理员可完成对多个读者、多本图书的借还，也是一对多的关系。联系属性包括借阅图书的读者 ID、图书 ID、借书管理员 ID、借阅时间、还书管理员 ID、还书时间。

图书借阅系统的实体联系 E-R 图如图 1-7 所示。

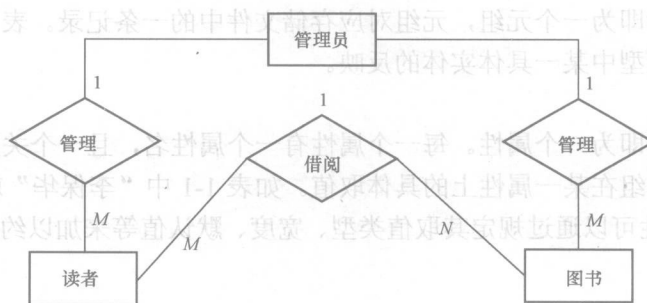


图 1-7 图书管理系统 E-R 图

1.3 关系数据模型

数据库基本的数据模型有三种：层次数据模型、网络数据模型和关系数据模型。层

次模型是数据库系统中最早出现的数据模型，它用树形结构表示实体类型及实体间联系。层次模型数据库管理系统的典型代表是 IBM 公司的 IMS 数据库管理系统。网状模型是一种比层次模型更具有普遍性的结构，在现实世界中，事物之间的联系更多是非层次关系的，用层次模型表示非树形结构特别复杂，网状模型则可以克服这一弊病。网状模型的典型代表是 DBTG 系统，是 20 世纪 70 年代美国数据系统语言研究会 CODASYL 下属的数据库任务组 DBTG 提出的方案。关系模型自诞生以来，经过不断的发展，已经成为最成熟的数据模型，是当今最为流行的一种数据模型。在关系模型中，实体间的联系是通过二维关系来定义的，其数据结构是二维关系。每个二维关系都可以用一张二维表来表示。

关系数据模型由关系数据结构、关系操作及关系完整性约束三部分组成。下面先介绍关系数据模型中的基本概念。

1.3.1 关系数据模型的基本概念

下面以表 1-1 描述的图书信息为例，说明关系数据模型的基本概念。

1. 关系

一个关系对应通常说的一张二维表，每个关系都有一个关系名。关系是对概念模型中某类实体的反映。例如，表 1-1 的关系名为“读者信息表”，是对多个读者信息的反映。在计算机里，一个关系可以存储为一个文件。

表 1-1 读者信息表

学 号	姓 名	性 别	身 份 证 号
1000001	张三力	男	510612197005061763
1000002	李保华	女	512632198008121763
1000003	王力刚	男	510216198205221760
1000004	陈大伟	男	51041919850802176X

2. 元组

二维表中的一行即为一个元组，元组对应存储文件中的一条记录。表 1-1 中包括 4 个元组。元组是对概念模型中某一具体实体的反映。

3. 属性

二维表中的一列即为一个属性。每一个属性有一个属性名，且一个关系中的属性不能重复。属性值是某一元组在某一属性上的具体取值。如表 1-1 中“李保华”就是第 2 个元组的“姓名”属性值。属性可以通过规定其取值类型、宽度、默认值等来加以约束。属性对应存储文件中的一个字段。

4. 域

属性的取值范围即为域。例如，读者信息中性别的取值范围只能是男和女。

5. 关键字/码

若关系中的某一个属性或属性组的值唯一决定其他所有属性，即元组的值，这个属性或属性组称为该关系的关键字。在如表 1-1 所示的学生关系中，学号是关键字，所有学生的学号是唯一的。

6. 候选键/候选码 如果一个关系中有多个属性或属性组都能用来标识该关系的元组,那么这些属性或属性组都称为关系的候选键。在如表 1-1 所示的学生关系中,身份证号同样是唯一的,也可以作为关键字。

7. 主键/主码 在一个关系的多个候选关键字中指定其中一个作为该关系的关键字,它就被称为主键。

8. 关系模式 关系模式是对关系的信息结构及语义限制的形式化描述,用关系名和包含的属性名集合来表示:

关系名(属性 1, 属性 2, ..., 属性 n)

例如:

读者信息表的关系模式为:读者(学号, 姓名, 性别, 身份证号);

图书信息表的关系模式为:图书(编号, 书名, 作者, 出版社, 出版日期);

借书信息表的关系模式为:借书信息(读者学号, 图书编号, 借书日期);

1.3.2 关系数据库的基础理论

关系数据库建立在关系数据模型的基础之上,具有严格的数学理论基础。关系数据库的理论包括关系运算、关系的完整性规则及关系规范化理论。

1. 关系运算

关系数据库对数据的操作除了包括集合代数的并、差等运算之外,更定义了一组专门的关系运算:连接、选择和投影。关系运算的特点是运算的对象和结果都是表。

选择是单目运算,其运算对象是一个表。该运算按给定的条件,从表中选出满足条件的行形成一个新表,作为运算结果。投影也是单目运算,该运算从表中选出指定的属性值组成一个新表。表的选择和投影运算分别从行和列两个方向上分割一个表。连接是把两个表中的行按照给定的条件进行拼接而形成一个新表。两个表连接最常用的条件是两个表的某些列值相等,这样的连接称为等值连接。数据库应用中最常用的是自然连接。进行自然连接运算要求两个表有共同的属性。自然连接运算的结果表是在参与操作两个表的共同属性上进行等值连接后,再去除重复属性后所得的新表。自然连接运算记为 $R \bowtie S$,其中 R 、 S 是参与运算的两个表。

2. 关系的完整性规则

关系的完整性规则也称为关系的约束条件,它是对关系的一些限制和规定,通过这些限制保证数据库的数据合理、正确和一致。关系的完整性规则包括域完整性规则、实体完整性规则和参照完整性规则三个方面。

(1) 域完整性规则。由用户根据实际情况,对数据库中数据的内容所做的规定称为域完整性规则。通过这些规则限制数据库只接受符合约束条件的数据值,不接受违反约束条件的数据,从而保证数据库中的数据合理可靠。例如,表中的性别数据只能是男和女;对价格数据也应该有一定的限制,不能是任意值。

(2) 实体完整性规则。关系数据模型对应的是现实世界中的数据实体,而关键字是

实体唯一性的表现。没有关键字就没有实体,所有关键字不能是空值。可见,在任何关系的任何一个元组中,主关键字又称主键的值不能为空值。这条规则称为实体完整性规则。

(3) 参照完整性规则。参照完整性规则也称为引用完整性规则,这条规则是对外部关键字的规定,要求外部关键字的取值必须是存在的,即不允许在一个关系中引用另一个关系里不存在的元组。

3. 关系规范化理论

在关系数据库中,对于同一个系统,选用不同关系模式集合作为数据库模式,其性能优劣是大不相同的,某些数据库模式常常带来存储异常,这是不利于实际应用的。为了区分数据库模式的优劣,关系规范化理论中将数据库模式的不同等级称为范式。通常将关系分为 5 种范式。满足最低条件的称为第一范式 1NF。1NF 是关系模式应满足的最起码的条件。

在介绍各个范式之前,先定义一个重要概念——函数依赖。通俗地讲,函数依赖是指一个关系模式中的一个属性集和另一个属性集之间的对应关系。函数依赖的定义如下。

关系模式 R 上的函数依赖,是行如 $X \rightarrow Y$ 的一个约束,其中 X 和 Y 是 R 中的属性集。如果 R 中任意两个元组在 X 上属性值相等,则它们在 Y 上属性值也一定相等,称 X 函数确定 Y ,或 Y 函数依赖于 X 。

(1) 第一范式 1NF。

要求组成关系的所有属性都是原子属性,即属性不可分性。设 R 是一个关系模式,如果 R 中的每个属性都是不可分解的,则称 R 是第一范式。即实体中的某个属性不能有多值或者不能重复的属性。如果出现重复的属性,就可能需要定义一个新的实体,新的实体由重复的属性构成,新实体与原实体之间为一对多关系。在第一范式中,表的每一行只包含一个实体的信息。例如对于图书信息表,不能将图书的出版社、出版日期放在一列中显示;每一本图书的信息在整个表中只占一行,一行也只能存储一本图书的信息。简而言之,第一范式就是无重复的列。

(2) 第二范式 2NF。

要求关系在满足第一范式的前提下,除去所有不完全依赖于主键的属性,即要求关系的所有非主属性都完全依赖于主键,则称 R 是第二范式。第二范式要求数据库表中的每个实体或行必须可以被唯一地区分。为实现区分就需要在表上加一标识列。例如,读者信息表中的读者编号列,就是为每个读者分配一个唯一的编号,从而区分不同的读者。这个唯一属性就是主键或主码。第二范式要求实体的属性完全依赖于主关键字。所谓完全依赖是指不能存在仅依赖主关键字的一部分属性,如果存在这种依赖,那么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。

存储部分的函数的依赖关系是造成插入异常的原因。例如,在关系模式(读者编号,姓名,图书编号,书名,借书日期,系名,系主任)(前提是该读者为在校学生)中,主关键字是读者编号和图书编号,对于非主属性姓名、系名来说,它们只部分依赖于读者编号,而与图书编号无关。可见该关系模式存在着数据冗余、更新异常、插入异常和删除异常等存储问

题。第二范式消除了部分函数依赖,也就消除了存储异常。解决办法就是将上述关系模式分解出若干个第二范式的关系模式。分解结果如下:

读者关系模式 READER (读者编号, 姓名, 系名, 系主任)

借书关系模式 BORROW (读者编号, 图书编号, 借书日期)

此时,在读者关系模式中,系名、系主任等非主属性完全依赖于读者编号这一主属性;借书关系模式中,借书日期也完全依赖于读者编号、图书编号这两个主属性。

简而言之,第二范式就是非主属性完全依赖于主属性。

(3) 第三范式 3NF。

要求关系在满足第二范式的前提下,且没有一个非主属性传递依赖于主属性,则称 R 是第三范式。

传递函数依赖关系是造成删除异常的原因。如读者关系模式 READER (读者编号, 姓名, 系名, 系主任) 存在着学号决定着系名,系名决定系主任,即系主任这一非主属性传递依赖于学号,存在着删除异常的问题。解决办法是进一步将读者关系模式分解为两个独立的第三范式的关系模式,从而消除传递依赖。分解结果如下:

读者关系模式 (读者编号, 姓名, 系名)

系关系模式 (系名, 系主任)

可以看出,读者关系模式和系关系模式都单独描述了一个实体,不存在传递依赖,符合第三范式。

第三范式基本解决了异常问题,但还不能彻底解决数据冗余问题。更完善的范式是 Boyce-Codd 范式。

(4) Boyce-Codd 范式 (BCNF 范式)。

如果关系模式 R 中的所有决定因素都是码,则称为 BCNF 范式。BCNF 范式消除了关系中冗余的码,所有非主属性对每个码完全函数依赖,所有主属性对每个不包含它的码完全函数依赖,没有任何属性完全函数依赖于非码的任何一组属性。

1.4 项目实例 1——人力资源管理系统数据库设计

随着市场竞争的日趋激烈,人才已经成为实现企业自身战略目标的一个非常关键的因素。巧虎公司为了人力资源系统更加有效的运行,管理层要求开发一套新的人力资源管理系统以适应公司要求。人力资源部门负责的工作是管理职位和薪水,招聘雇用以及培训员工,提升和分配人员以及处理退休事项等。

1.4.1 项目任务单

(1) 仔细阅读人力资源关系系统需求。

(2) 根据需求进行数据库概念结构设计,完成 E-R 图的绘制。

(3) 根据 E-R 图进行数据库逻辑结构设计,完成数据表结构的设计。

(4) 对照需求检查并修改最后的设计结果。

1.4.2 人力资源管理系统需求

一个标准的人力资源管理系统应该包括雇员信息管理,人事档案管理,考勤管理,职位

管理,薪水福利管理,招聘管理,合同管理等几大功能。除此之外系统还应包括信息系统必须具备的通用功能,如系统管理、权限设置、数据备份与恢复等。其中每个功能都由若干相关联的子功能模块组成。在本教材中主要关注雇员信息管理,职位管理,薪水福利管理,招聘管理四个子系统。需要完成功能主要有:

(1) 雇员信息管理用于输入、修改和查询员工的信息,常用的信息包括雇员编码、姓名、候选人编码、地址、城市、国家、邮政编码、国家编码、电话、资历、出生日期、性别、现任职位、团体名称、邮箱、部门编码、地区、照片、加入日期、注册日期、社会保险号码、管理编码等。

(2) 职位管理可以查看新增、修改、删除职务。可以执行设立职务、撤销职务等业务操作。内容包括职位代码,职位描述,预算人数,当前人数等信息。

(3) 招聘管理用于管理和跟踪企业招聘新员工时的全过程。系统包括以下几个部分:“招聘申请登记”用于管理企业下属各部门的招聘申请信息,内容包括申请编号、部门编号、拟招聘人数、工作内容、招聘日期、截止时间等;“外部求职者人员登记”用于管理社会应聘人员的信息,内容包括应聘者编号、姓名、地址、城市、状况、邮编、手机、职位号、申请时间、雇员代码、报纸编号、合同招聘号、工作公平号、校园招聘号、前工号、资历、工作经历、生日、性别、学校号、以前的月收入、照片、电子邮箱、状态、测试时间、测试成绩、面试日期、面试考官、面试评论、评价等。“内部求职人员登记”用于管理企业内部岗位的招聘工作,内容包括内部求职者编码、雇员编码、内部职位公告编码、职位信息、申请日期、测验时间、测验成绩、面试日期、面试者、面试评语、等级、状态。“报纸信息”用于管理刊登招聘广告的信息,内容包括报纸编码、报纸名称、地区、报纸类型、联系人、地址、城市、省、邮编、国家代码、传真、电话。“内部岗位公告”用于刊登公司内部岗位招聘信息,内容包括岗位信息、空缺额、地区、发布日期、截止日期等信息。

(4) 薪水福利管理用于管理企业员工的薪资和福利,本系统简化了薪水福利管理子系统,只关心雇员的月薪、年薪、奖金。

1.4.3 数据库概念结构设计

人力资源管理信息系统可以划分的实体有:雇员信息实体、职位信息实体、招聘申请信息实体、外部求职者人员信息实体,内部求职人员信息实体,报纸信息实体等,用 E-R 图描述实体,如图 1-8 所示。

在概念模型的图形化描述中介绍了 E-R 图绘制的传统方法,在工作中是使用 PowerDesigner 和 Visio 等实体关系图建模工具来绘制 E-R 图。依赖于所选择的用来建立数据模型的软件包,建模工具可能会根据模型生成 SQL 命令或直接修改数据库模式,这种功能带来了极大的便利。一些建模工具的功能适合于大量的数据库类型,例如 PostgreSQL、MySQL、Oracle、DB2 等。对于简单的数据库修改,改动操作可以从建模工具通过 ODBC 直接完成。数据库改动还允许以增量方式进行。当第一次使用建模工具时,可以查看建模工具生成的 SQL,看看自己是否可以信任和认可建模工具对数据模型的解释。一段时间之后,就会熟悉建模工具对各种关系和表细节的解释。同学们可以首先学习使用 Visio 工具来绘制 E-R 图,本教材中不介绍 Visio 工具的使用方法。