

高等院校信息技术规划教材

实用数据库教程

(第2版)

赵池龙 编著



清华大学出版社

九江学院图书馆



1822248

1524575

高等院校信息技术规划教材

实用数据库教程

(第2版)

TP311.13/21401

赵池龙 编著

不外借

九江学院图书馆
藏书章

清华大学出版社
北京

内 容 简 介

本书主要介绍数据库的基本原理、主要操作、编程方法、设计模式、设计工具、设计案例,并且系统地提出了数据库规范化的“四个原子化”理论,以及七个数据库设计模式方法论。

本书适合于各类理工科大学计算机相关专业的数据库原理与设计课程,也适合于 IT 企业的软件工程师自学之用。作为大学教材,教学内容应涵盖全部章节(非重点院校可省略打星号的章节),教学计划是 4 学分 72 学时。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

实用数据库教程 / 赵池龙编著. —2 版. —北京:清华大学出版社, 2012.1
(高等院校信息技术规划教材)

ISBN 978-7-302-27337-0

I. ①实… II. ①赵… 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(2011)第 237442 号

责任编辑:袁勤勇 顾 冰

责任校对:白 蕾

责任印制:何 芊

出版发行:清华大学出版社

地 址:北京清华大学学研大厦 A 座

<http://www.tup.com.cn>

邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62795954, jsjje@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:北京鑫海金澳胶印有限公司

经 销:全国新华书店

开 本:185×260

印 张:17

字 数:389 千字

版 次:2012 年 1 月第 2 版

印 次:2012 年 1 月第 1 次印刷

印 数:1~3000

定 价:29.50 元

产品编号:044725-01

前言

foreword

实用数据库教程(第2版)是一本具有自主创新、特色鲜明、自成体系的大学教材,是作者在IT企业多年数据库设计经验与在高校多年数据库教学经验的积累与总结。为了克服第1版的某些缺陷,该版对第1篇进行了重新构思与编写,并且增加了数据仓库简明原理1章;对第2篇进行了充实与加强,使其更加丰满;对第3篇进行了彻底更新,使其更加实用。该版围绕数据库的“基本原理、主要操作、编程方法、设计模式、设计工具、设计案例”等核心内容,进行深入论述,使其精益求精。从实用数据库教程的角度上看,该版内容上耳目一新,理论上深入浅出,实践上通俗易懂,其目的是使高校数据库原理与设计课程及IT企业实际需求,实现无缝连接,变为平滑接轨,帮助读者成为数据库高手。

本书第1篇是“数据库基本原理”篇。数据库基本原理是研究关系模型原理、结构化查询原理、事务处理与并发控制原理、数据库规范化原理、数据库编程原理、数据库管理原理、数据仓库原理的科学。这些原理不仅对数据库管理系统的设计与实现很有必要,而且对数据库规范化设计与数据库编程大有好处。

本书第2篇是“数据库设计模式”篇。该篇不但提出了数据库规范化的四个原子化理论(属性原子化、实体原子化、主键原子化、关系原子化),而且还提出了七个数据库设计模式(主从模式、弱实体插足模式、强实体插足模式、列变行模式、星系模式、主题模式、全局模式)。由于四个原子化理论揭示了数据库规范化的内在规律,又有七个数据库设计模式的强力支持,才使得数据库规范化与数据集成化的道路,由崎岖山间小道变为宽敞平原大道。

本书第3篇是“数据库设计实践”篇。该篇包括数据库设计工具和数据库设计案例分析。PowerDesigner是一种数据库设计图形语言,重点是设计实体-联系图(E-R图)。该篇具体分析了三个真实的数据库设计案例。

通过第1篇的学习,读者应该知道:元组是属性的集合,关系是

元组的集合,数据库是关系的集合,属性、元组、关系、数据库的操作方式都是集合操作;通过第2篇的学习,读者应该知道:属性、元组、关系、数据库这四者之间实现规范化的充分必要条件,是满足四个原子化理论,而数据库的七个设计模式,则是实现数据库规范化的保证措施;通过第3篇的学习,读者应该知道,只要学会 PowerDesigner 对话框中工具模板 Palette 的用法,设计 E-R 图将是很简单的事情。读者还应该知道:数据库的需求分析、概念设计、物理设计、编程实现都要以 E-R 图为中心;通过全书的学习,读者应了解关系数据库的精髓,就是一张二维表加上四个原子化理论。

学习的最好方法是理论联系实践,比如在网上下载一个免费的数据库管理系统(SQL Server 等),针对某个信息系统,设计出该系统的 E-R 图,建立几个基本表,向基本表加载记录,初步建立一个数据库,在这个数据库上进行查询和更新操作,还可以在基本表上建立索引。这样你就对数据库有了感性认识并产生兴趣,而兴趣是最好的动力。

因为重点院校的师生更应该掌握数据库设计的四个原子化理论与七个设计模式方法论,所以本书适合于各类理工科大学计算机相关专业各类学生的数据库原理与设计课程,也适合于 IT 企业的软件工程师自学之用。作为大学教材,教学内容应涵盖全部章节(第10章可适当提前讲授,非重点院校可省略打星号的章节),教学计划的4学分72学时。

在第2版的编写过程中,得到了张松、蔡勇、杨林、吕威、陈红顺、周鹏、李冬、邹文蓬、程庆威、李大林、张安波的支持与帮助,作者在此深表感谢。

赵池龙

2011年10月

目录

Contents

第 1 篇 数据库基本原理

第 1 章 数据库系统概论	3
1.1 数据库系统的基本概念	3
1.2 数据模型与数据库系统	6
1.3 数据库系统的特点	11
1.4 数据库系统的组成	12
1.5 数据库系统的三级模式结构	13
习题 1	15
第 2 章 关系数据库原理	17
2.1 关系数据库基本概念	17
2.1.1 关系的名词解释	17
2.1.2 关系的数据结构	22
2.1.3 关系的数据操作	25
2.2 关系的完整性约束	26
2.2.1 实体完整性约束	26
2.2.2 参照完整性约束	27
2.2.3 用户定义完整性约束	28
2.3 关系代数	29
2.3.1 传统的集合运算	30
2.3.2 专门的关系运算	32
习题 2	36
第 3 章 数据库 SQL 语言原理	38
3.1 SQL 语言特点	38

3.2	数据库定义语句	40
3.2.1	经典示例数据库	40
3.2.2	模式定义语句	43
3.2.3	表定义语句	43
3.2.4	索引定义语句	47
3.3	数据查询语句	48
3.3.1	单表查询语句	49
3.3.2	连接查询语句	52
3.3.3	嵌套查询语句	54
3.3.4	集合查询语句	57
3.3.5	函数与表达式语句	57
3.3.6	分组与筛选语句	59
3.4	查询优化技术	60
3.5	数据更新语句	61
3.5.1	插入数据语句	61
3.5.2	修改数据语句	62
3.5.3	删除数据语句	62
3.6	视图语句	63
3.6.1	视图概论	63
3.6.2	定义视图	64
3.6.3	操作视图	65
3.6.4	删除视图	66
3.7	数据控制语句	66
3.7.1	授予权限语句	66
3.7.2	收回权限语句	68
习题 3		69

第4章 关系数据库规范化原理

4.1	关系规范化的作用	72
4.1.1	问题的提出	72
4.1.2	解决的方法	74
4.1.3	规范化理论的出现	76
4.2	函数依赖	77
4.2.1	关系模式的简化表示	77
4.2.2	函数依赖的基本概念	77
4.2.3	码的函数依赖表示	79
4.2.4	函数依赖的推理规则	79
4.3	关系模式的规范化理论	80

4.3.1	第一范式	81
4.3.2	第二范式	83
4.3.3	第三范式	84
4.3.4	BCNF 范式	85
4.4	多值依赖与第四范式	86
4.4.1	问题的引入	86
4.4.2	多值依赖基本概念	87
4.4.3	第四范式	88
* 4.4.4	保持函数依赖介绍	89
4.4.5	六个范式的发展历程	91
4.5	关系模式的规范化方法	92
4.5.1	关系模式规范化步骤	92
4.5.2	对关系模式规范化的范式理论进行反思	93
习题 4		94
第 5 章	数据库编程与管理	95
5.1	事务的基本概念	95
5.1.1	事务的特性	95
5.1.2	事务的类型	97
5.2	并发控制	98
5.2.1	锁的概述	99
5.2.2	死锁及处理	101
5.3	存储过程编程	101
5.4	触发器编程	103
* 5.5	嵌入式 SQL 编程	106
* 5.6	JDBC 编程	107
5.6.1	加载驱动程序	108
5.6.2	建立连接	108
5.6.3	使用语句	109
5.6.4	结果集	110
5.6.5	关于 ODBC 编程	113
5.7	数据库管理	113
5.7.1	数据库故障	114
5.7.2	数据库备份	114
5.7.3	数据库恢复	115
习题 5		117



第6章 数据仓库简明原理	118
6.1 数据仓库启蒙	118
6.2 数据仓库定义	119
6.3 数据仓库原理	120
* 6.3.1 维表	121
* 6.3.2 事实表	124
6.3.3 数据仓库体系结构	128
6.4 数据仓库应用案例分析	131
习题6	133

第2篇 数据库设计模式

第7章 四个原子化理论	137
7.1 设计模式基本概念	138
7.2 数据库中的四种表	140
7.3 原始单据与实体之间的联系	143
7.4 原始 E-R 图与现代 E-R 图	144
7.5 数据库设计的内容与步骤	146
7.6 四个原子化理论的提出	148
7.7 四个原子化的实现方法	150
7.8 四个原子化理论的定理与推论	152
7.9 四个原子化是一种软件工程方法论	154
7.10 四个原子化理论与范式理论比较	155
习题7	158
第8章 数据库的微观设计模式	160
8.1 微观设计模式1(主从模式)	160
8.2 微观设计模式2(弱实体插足模式)	163
8.3 微观设计模式3(强实体插足模式)	171
8.4 微观设计模式4(列变行模式)	174
8.5 微观设计模式的综合练习	178
习题8	181
第9章 数据库的宏观设计模式	182
9.1 宏观设计模式1(星系模式)	182

9.2 宏观设计模式 2(主题模式)	188
9.3 宏观设计模式 3(全局模式)	195
9.4 四个原子化理论与七个设计模式方法论小结	200
习题 9	203

第 3 篇 数据库设计实践

第 10 章 数据库设计工具	207
10.1 安装与启动	208
10.1.1 PowerDesigner 的安装	208
10.1.2 PowerDesigner 的启动	210
* 10.2 数据库业务模型设计	211
10.2.1 建立 BPM	211
10.2.2 创建起点	212
10.2.3 定义处理过程	213
10.2.4 定义资源	214
10.2.5 定义终点	214
10.2.6 定义流程	214
10.3 概念数据库模型设计	215
10.3.1 创建概念数据模型	215
10.3.2 创建实体	217
10.3.3 建立实体之间的联系	218
10.3.4 定义域	220
10.3.5 定义业务规则	222
10.3.6 定义 CDM 属性	223
10.4 物理数据库模型设计	224
10.4.1 创建物理数据模型	224
10.4.2 创建表	226
10.4.3 创建列	227
10.4.4 创建索引	227
10.4.5 创建视图	229
10.4.6 创建触发器	230
10.4.7 创建存储过程和函数	232
10.4.8 通过 CDM 生成 PDM	233
* 10.5 面向对象模型设计	236
10.5.1 创建 OOM	236
10.5.2 设计 Use Case 图	236



10.5.3 通过 PDM 转换生成 OOM	239
第 11 章 数据库设计案例分析	241
11.1 网上投票系统数据库案例分析	241
11.1.1 投票系统需求分析	241
11.1.2 数据库概念设计	243
11.1.3 数据库物理设计	243
11.1.4 数据库表设计	244
* 11.2 网上论坛系统数据库案例分析	245
11.2.1 MySQL 数据库管理系统	245
11.2.2 论坛管理系统数据库脚本	246
11.2.3 概念数据模型	249
11.2.4 数据库表的详细清单	250
11.2.5 数据库连接实现	252
11.3 学生管理信息系统数据库案例分析	253
11.3.1 概念数据模型 CDM 设计	253
11.3.2 物理数据模型 PDM 设计	255
11.3.3 用 SQL Server 2000 创建数据库	255
11.3.4 用 PowerBuilder 创建表	257
11.3.5 用 ADO 控件连接数据库	257
参考文献	259

第 1 篇

数据库基本原理

第1篇的特点是将CASE工具设计的现代E-R图引入到数据库原理的各章之中。通过该篇的学习,读者将会知道:元组是属性的集合,关系是元组的集合,数据库是关系的集合,属性、元组、关系、数据库的操作方式都是集合操作,这些操作通过数据库编程加以实现,而关系代数与范式理论,则是这种数据库操作的理论基础。读者还会知道,维表与事实表是数据仓库的基本表,表中的数据来源于数据库,这充分说明了数据仓库与数据库的紧密程度。

数据库系统概论

本章以深入浅出的道理与通俗易懂的语言,论述数据库系统的基本概念,数据库系统的发展历史,数据库系统的组成和特点,以及数据库系统的三级模式与两层映像结构。通过本章的学习,读者应在宏观上对数据库系统有一个大体的印象。

1.1 数据库系统的基本概念

要弄清什么是数据库系统,就要首先弄清什么是数据?什么是数据库?什么是数据库管理系统?只有将数据、数据库、数据库管理系统这三个基本概念弄清后,才能最终懂得什么是数据库系统这个基本概念。

1. 数据

在计算机中,数据(Data)是数据库存储的基本对象。而在大多数人头脑中,对数据的第一个反应就是数字。为了弄清什么是数据,我们给数据下一个广泛的、科学的、精确的、现代的定义,然后再对数据进行解释。

【定义 1-1】 计算机中描述事物的符号称为数据。

这里所指的事物,包括事物本身和事物的各种状态。

这里所指的符号,可以是数字,也可以是文字、图形、图像、声音、语言等。

通过这个定义可以知道,尽管“数据”有多种多样的表现形式,但是它们都可以经过“数字化”后,存入计算机中。

在计算机系统中,数据的概念是广义的。早期的计算机系统主要用于科学计算,处理的数据都是数值型数据,如整数、实数、浮点数等。现代的计算机用于各种领域,存储和处理的对象十分广泛,表示这些对象的数据也越来越复杂了。

由于数据的表现形式还不能完全表达数据的内容,所以对数据还需要进行解释。数据和关于数据的解释是不可分的,请看下例。

【例 1-1】 99 是一个数据,它可以是某个同学某门课程的成绩,也可以是某个人的体重,还可以是计算机系 2011 级的学生人数,甚至可以是某男生送给某女生的 99 朵玫瑰。

对数据的解释是指对数据含义的说明,数据的含义称为数据的语义,数据与其语义是不可分的。



【例 1-2】 在日常生活中,人们可以直接用自然语言(如汉语或英语)来描述事物。例如,可以用自然语言这样来描述某校计算机系一位同学的基本情况:张本悟同学,男,1982年11月12日生,广东省广州人,2000年入学。而在计算机中,常常这样来描述事物:(张本悟,男,19821112,广东省广州市,计算机系,2000),即把学生的姓名、性别、出生年月日、出生地、所在院系、入学时间等组织在一起,组成一条记录。

这里的学生记录就是描述学生的数据,这样的数据是有结构的。“记录”这个概念很重要,它是计算机中表示和存储数据的一种格式或一种方法。

通过对数据的定义与解释,我们不难发现,计算机中的数据,从宏观上可以分为两类数据:组织数据的数据和被组织的数据,前者称为元数据,后者称为基础数据。例如,(姓名、性别、出生年月日、出生地、所在院系、入学时间)就是元数据,由元数据构成了记录的“型”,而记录的“值”可以是(张山,男,19891116,广东省广州市,计算机系,2010),也可以是(李斯,女,19891217,广东省珠海市,软件系,2010),还可以是(王武,男,19900222,广东省东莞市,电子系,2010),记录的这些“值”就是基础数据。

记录的“型”与记录的“值”、“元数据”与“基础数据”,这四个概念太重要了,今后我们还要多次提到与解释,因为它对于学好数据库原理、数据库设计、数据库应用与数据仓库,都很有帮助。

2. 数据库

顾名思义,数据库(Data Base,DB)是存放数据的仓库或容器。只不过这个仓库或容器是建立在计算机存储设备上,而且数据是以一定的格式存放的。

人们在收集并抽取出一个应用所需要的大量数据之后,并将其保存起来,以供进一步加工处理。在科学技术飞速发展的今天,人们的视野越来越广阔,数据量也越来越庞大。人们过去把数据存放在文件柜里,现在借助计算机技术和数据库技术,把数据存放在数据库里。由文件柜发展到数据库,标志着信息时代的来临。

严格地讲,数据库是指长期储存在计算机内、有组织、可共享的大量数据的集合。数据库中的数据,是按一定的数据模型来组织、描述和储存的,这些数据具有较小的冗余度、较高的独立性和较强的扩展性,只有这样,才能有利于各种用户共享这些宝贵的信息资源。下面,我们给数据库下一个科学的定义。

【定义 1-2】 数据库是存储在一起的相关数据的集合,这个集合是结构化的、无不必要冗余的、长久存储的,并且能为多种应用共享与服务。

由此可见,数据库中的数据,具有永久存储、科学组织和共享这三项基本特性。

3. 数据库管理系统

了解数据和数据库的基本概念后,下一个问题就是如何科学地组织和存储数据库中的数据,以及如何高效地获取和维护数据,完成这个艰巨任务的软件,就是数据库管理系统(DataBase Management System,DBMS),它是数据库系统中最重要的概念。

DBMS是位于用户与操作系统之间的一层软件,它不是一个应用软件,而是一个大

型复杂的系统软件。DBMS 的用途,包括科学地组织和存储数据、高效地获取和维护数据,它的主要功能包括以下几个方面。

(1) 数据定义功能。它提供数据定义语言(Data Definition Language, DDL),用户通过 DDL,可以方便地对数据库中的数据对象进行定义。

(2) 数据组织、存储和管理功能。DBMS 要分类组织、存储和管理各种数据,包括数据字典管理、用户数据管理、数据存取路径管理,还要确定以何种文件结构或何种存取方式将数据存储到存储器上,以及如何实现这些数据之间的联系,其目的是为了提高存储空间的利用率,并加快查询速度(如索引查找、Hash 查找、顺序查找等)。

(3) 数据操纵功能。DBMS 还提供数据操纵语言(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,如数据的查询、插入、删除和更新等操作。

(4) 数据库的事务管理和运行管理功能。数据库在建立、运用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性,以及发生故障后的系统恢复性。

(5) 数据库的建立和维护功能。数据库原始数据的输入和转换功能,数据库的转储和恢复功能,数据库的重组功能、分析功能、监视性能等。这些功能通常是由一些实用程序或管理工具来完成。

(6) 其他功能。例如 DBMS 与网络中其他软件系统的通信功能,一个 DBMS 与另一个 DBMS 或文件系统的数据库转换功能,异构数据库之间的互访和互操作功能等。

下面,给 DBMS 下一个科学的定义。

【定义 1-3】 DBMS 是一种用于建立、操纵、使用和维护数据库的大型系统软件。

由此可见,数据库管理系统的作用,是对数据库进行统一的管理和控制,以保证数据库的安全性和完整性。用户通过它访问数据库中的数据,数据库管理员通过它进行数据库的维护。DBMS 还提供多种功能,例如,可同时供多个应用程序或多个用户使用;可以使用不同的方法、在不同地点、串行或并发去建立、修改、查询或维护数据库,从而使用户能方便地定义和操纵数据。

我们知道,导弹、火箭、飞机、汽车、火车、轮船都有引擎,即发动机。数据库系统也有引擎,这个引擎或发动机就是 DBMS。因此,有时又将 DBMS 称为数据库引擎,即数据库发动机。形象地说,其所以叫数据库引擎或数据库发动机,就是它不但能跟随着数据库用户实时地运转,而且在运转中永不熄火、永不疲倦,如同游戏引擎能实时地、全心全意地跟踪游戏玩家运转一样。

人们利用数据库引擎可控制访问权限并快速处理事务,从而满足企业内大多数需要处理大量数据的应用程序的要求,这包括创建用于存储数据的表和用于查看、管理和保护数据安全的数据对象。

4. 数据库系统

数据库系统(DataBase Systems, DBS)是一个复杂而庞大的系统,它包括许多人员、硬件和软件,如图 1-1 表示。

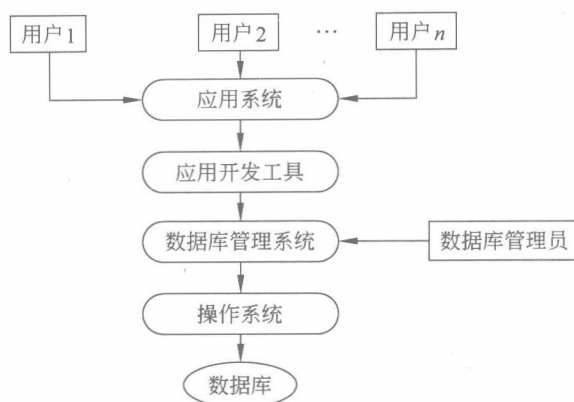


图 1-1 数据库系统

由图 1-1 可知,数据库系统是指在计算机系统中引入数据库之后的系统,它一般由网络硬件系统、网络操作系统、数据库、数据库管理系统(及其开发工具)、应用系统、用户和数据库管理员构成。

应当指出的是,数据库的建立、使用和维护等日常工作,只靠几个程序设计人员是远远不够的,还必须有专职的数据库管理员(Data Base Administrator, DBA),由他专门进行数据库的管理。

下面,给数据库系统下一个科学的定义。

【定义 1-4】 数据库系统是由数据库及其管理数据库的软硬件设施所组成的系统。

由此可见,数据库系统是为了适应数据处理的需要,而发展起来的一种较为理想的数据处理系统。它是一个为应用系统提供数据的存储、处理和维护的系统,是存储介质、处理对象和管理系统的集合体。

由图 1-1 还可以知道,DBMS 是 DBS 的一个重要组成部分,数据库(DB)不仅需要 DBMS 的支持,又是 DBS 的一部分,它们三者之间的依存关系,如下面的公式所示:

$$\text{DBS} \supset \text{DB} \supset \text{DBMS}$$

另外,在不引起混淆的情况下,有时在 IT 行业内部,人们习惯上常常把数据库系 DBS 统简称为 DB。

1.2 数据模型与数据库系统

什么是数据模型(Data Model)呢?通俗地说,数据模型是站在计算机的角度,用模型的方式来描述数据、组织数据、处理数据的一种思想或一种方法。因为计算机不能直接组织与处理现实世界中的具体事物,所以人们必须发挥主观能动性,事先将客观事物转换为某种模型,或者说用某种模型来模拟客观事物,然后再将这些模型交给计算机去组织与处理。