

高等学校计算机系列教材

数据库系统概论

——习题、实验与考试辅导

史嘉权 史红星 李 博 编著

清华大学出版社



高等学校计算机系列教材

数据库系统概论

——习题、实验与考试辅导

史嘉权 史红星 李 博 编著

清华大学出版社

北京

内 容 简 介

本书是为高等院校《数据库系统概论》教材而编写的配套辅助教材。本书的主要特色是:内容新颖,在以关系数据库为基础的前提下,引进面向对象的数据库技术,以微软公司新推出的关系数据库管理系统 SQL Server 2000 为实验环境;与主教材紧密结合,习题、实验与主教材三位一体,互为补充;题型多,题量大,覆盖范围宽;要求明确、重点突出;有易有难,对部分难题给出不同解法,并有深入分析;为综合实验提供了有参考价值的实验报告;便于灵活而科学的组卷。

主要内容包括:数据库系统要点;习题及其解答;实验环境与基础知识;教学实验内容及综合实验报告示例;考试大纲;组卷方案,模拟试题及参考答案。

本书可作为高等院校数据库课程的辅助教材。对于从事数据库研制、开发和应用的有关人员,本书也是一本很有价值的参考书。

版权所有,翻印必究。举报电话:010-62782989 13501256678 13801310933

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。

本书防伪标签采用特殊防伪技术,用户可通过在图案表面涂抹清水,图案消失,水干后图案复现;或将表面膜揭下,放在白纸上用彩笔涂抹,图案在白纸上再现的方法识别真伪。

图书在版编目(CIP)数据

数据库系统概论——习题、实验与考试辅导 / 史嘉权, 史红星, 李博编著. —北京:清华大学出版社, 2006.2

(高等学校计算机系列教材)

ISBN 7-302-11937-6

I. 数… II. ①史… ②史… ③李… III. 数据库系统—高等学校—教学参考资料 IV. TP311.13

中国版本图书馆 CIP 数据核字(2005)第 114089 号

出 版 者: 清华大学出版社

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

社 总 机: 010-62770175

地 址: 北京清华大学学研大厦

邮 编: 100084

客户服务: 010-62776969

责任编辑: 马瑛珺

印 装 者: 北京嘉实印刷有限公司

发 行 者: 新华书店总店北京发行所

开 本: 185×260 印张: 13 字数: 304 千字

版 次: 2006 年 2 月第 1 版 2006 年 2 月第 1 次印刷

书 号: ISBN 7-302-11937-6/TP·7740

印 数: 1~5000

定 价: 18.00 元

出版说明

清华大学出版社推出的这套《高等学校计算机系列教材》是《清华大学计算机系列教材》的姊妹篇。

《清华大学计算机系列教材》出版以来,多次获得国家和部级奖项。我们经常收到一些师生热情洋溢的来信,强烈感受到他们对新的知识与教育模式的渴求,同时也感受到广大师生对清华大学计算学科教学工作的关注和信任。

随着高等教育规模的持续扩大和高等教育改革的不断深入,不同院校对于计算学科的教学工作提出了新的要求,突出体现在:理论课时的压缩、实践能力的要求提高,以及学科教育与行业需求的不断结合。根据这些发展趋势,清华大学一批学术水平高、教学经验丰富的教授总结了他们几十年的教学和科研经验,有针对性地编写了《高等学校计算机系列教材》。这套教材的特点体现在:

1. 课程内容在《清华大学计算机系列教材》的基础上,进行了适时的修订更新,并且明确了教学基本要求,区分应该熟练掌握和只需一般了解的内容。

2. 强调加强基础理论教育,重视学生实践能力的培养。课程内容为进一步的实践教学既提供了基础知识,又留出了足够的时间。

另外,本套丛书同时出版了相关辅导用书,并为教师免费提供电子课件,便于师生的教学使用。

清华大学计算学科坚持推行具有启发性的、富于创造性的教学工作,为国家源源不断地培养出一批又一批优秀人才。从《清华大学计算机系列教材》中就可以体会到这些艰辛的探索历程,希望作为姊妹篇的《高等学校计算机系列教材》也能得到师生的认可。

清华大学出版社

2005 年 9 月

前 言

数据库技术作为计算机软件领域的一个重要分支,是计算机科学技术发展最快、应用最广的领域之一。数据库技术已经成为以计算机为中心的信息技术与应用系统的核心技术和重要基础。

人类在 21 世纪将进入以知识经济为主要基础的信息社会,而数据库正是信息社会中信息资源管理和开发利用的基础。因此,数据库的建设规模和使用水平已成为衡量一个国家信息化程度的重要标志。

综上所述,在计算机专业的教学中,数据库课程的地位和作用不言而喻。

本书作为与主教材《数据库系统概论》配套的辅助教材,主要特色如下:

(1) 内容新颖:在关系数据库的基础上引进面向对象数据库的对象定义语言 ODL 和对象查询语言 OQL。习题及其解答均能反映数据库技术的最新发展。实验环境采用微软公司新推出的关系数据库管理系统 SQL Server 2000,这是目前最流行的数据库管理系统之一。

(2) 与主教材紧密结合:把习题、实验与主教材作为学好数据库课程的有机组成部分,三位一体,互为补充。习题和实验是深入理解和熟练掌握数据库课程的两个重要环节。

(3) 题型多,题量大,覆盖范围宽:题型有判断题、单项选择题、多项选择题、填空题、简答题和综合题(包括编程题和应用题),有助于启发读者从不同的角度思考问题。共有 508 题,覆盖了全书的主要内容,有助于全面复习和检查所学的知识。

(4) 要求明确,重点突出:把相关知识点分为 3 个层次:熟悉或熟练掌握;比较熟悉或基本掌握;了解或初步掌握。对要求熟悉或熟练掌握的知识点,题型较多,难度也有适度加深。

(5) 难度上拉开距离:题目难度上有易、中、难之分,有助于检查对知识理解和掌握的深度,也有助于组卷时参考,从而在成绩上拉开档次。

(6) 给出不同解法,并有深入分析:部分难度较大的习题给出了两种解法,有助于开阔思路,灵活掌握所学的知识,也有助于教师对试卷的评阅;对个别难度较大的习题,解答中给出了详细的分析,有助于深入、准确地理解其中的难点,从中学习解题的思路和方法。

(7) 对综合实验提供了有参考价值的实验报告:既防止作为标准答案而限制读者的创造性思维,也防止初学者由于缺少经验而走过多的弯路。

(8) 组卷既有灵活性也有科学性:题型多,题量大,并按主教材的章节出题,为组卷提供了方便;每题都注明要求层次和难度等级,为组卷提供了参考性依据;按书中给出的两种组卷方案,题库中各种题型的数量与试题的数量至少达到 10 : 1,从而可以相信考试的成绩能真实地反映学生对数据库课程掌握的情况。

本书由如下 6 部分组成:

(1) 数据库系统概论要点:按主教材分 10 章,系统、全面而又有重点地总结、归纳各章的主要内容。

(2) 数据库系统习题及其解答:按主教材分 9 章,并按题型给出习题及其解答。部分习题给出两种解答;个别难题给出解题分析。

(3) 数据库系统实验环境与基础知识:对微软公司的 SQL Server 2000 以及与教学实验相关的基础知识做了较为系统的介绍。

(4) 数据库系统教学实验:给出 3 个基础实验题和 3 个综合实验题,并给出 3 个综合实验报告示例。

(5) 数据库系统考试大纲:按主教材分 10 章,给出考试的知识点和考试要求。

(6) 数据库系统模拟试题及参考答案:先给出两种组卷方案,再按组卷方案给出两种模拟试题及其参考答案。

总之,本书会帮助读者在理解主教材内容的基础上,在分析问题、解决问题、提高动手能力等方面有质的飞跃。

在本书的编写和书稿的审阅过程中,清华大学计算机系戴梅萼教授提出不少宝贵的建议和意见,在此表示衷心的感谢。

由于水平有限,难免有不妥之处,敬请读者批评指正。

作 者

2005 年 9 月于清华园

目 录

第一部分 数据库系统概论要点	1
第 1 章 数据库系统概述	1
第 2 章 数据库建模	2
第 3 章 关系模型和关系运算	3
第 4 章 数据库语言 SQL	5
第 5 章 查询优化与并发控制	6
第 6 章 关系数据库设计理论	7
第 7 章 数据库设计	8
第 8 章 SQL 系统环境	9
第 9 章 面向对象查询语言	9
第 10 章 数据库技术发展动态	10
第二部分 数据库系统习题及其解答	11
第 1 章 数据库系统概述	12
1.1 填空题	12
1.2 单项选择题	13
1.3 多项选择题	15
1.4 判断题	16
1.5 简答题	16
第 2 章 数据库建模	18
2.1 填空题	18
2.2 单项选择题	19
2.3 多项选择题	22
2.4 判断题	25
2.5 简答题	26
2.6 综合题	27
第 3 章 关系模型和关系运算	32
3.1 填空题	32
3.2 单项选择题	33
3.3 多项选择题	36
3.4 判断题	38
3.5 简答题	39
3.6 综合题	40

第 4 章	数据库语言 SQL	48
4.1	填空题	48
4.2	单项选择题	49
4.3	多项选择题	51
4.4	判断题	52
4.5	简答题	53
4.6	综合题	53
第 5 章	查询优化与并发控制	69
5.1	填空题	69
5.2	单项选择题	69
5.3	多项选择题	71
5.4	判断题	73
5.5	简答题	74
5.6	综合题	76
第 6 章	关系数据库设计理论	82
6.1	填空题	82
6.2	单项选择题	82
6.3	多项选择题	85
6.4	判断题	87
6.5	简答题	89
6.6	综合题	91
第 7 章	数据库设计	94
7.1	填空题	94
7.2	单项选择题	94
7.3	多项选择题	96
7.4	判断题	98
7.5	简答题	99
第 8 章	SQL 系统环境	101
8.1	填空题	101
8.2	单项选择题	101
8.3	多项选择题	103
8.4	判断题	104
8.5	简答题	106
8.6	综合题	107
第 9 章	面向对象查询语言	112
9.1	判断题	112
9.2	单项选择题	113
9.3	多项选择题	115

9.4	填空题	116
9.5	简答题	117
9.6	综合题	117
第三部分 数据库系统实验环境与基础知识		121
第1章	SQL Server 2000 简介	121
第2章	SQL Server 2000 的安装	123
2.1	SQL Server 2000 的硬件和软件安装要求	123
2.2	安装过程简介	125
第3章	查询分析程序(查询分析器)	128
第4章	数据库操作	130
4.1	建立数据库	130
4.2	更改数据库	131
4.3	删除数据库	131
第5章	标准 SQL 表操作	131
5.1	建表	132
5.2	删除表	133
5.3	更改表	134
5.4	索引	134
第6章	数据操作语句	136
6.1	插入、修改和删除语句	136
6.2	查询语句	138
第7章	T-SQL 语言概述	142
7.1	Transact-SQL 语法规则	142
7.2	T-SQL 中主要流程控制语句	143
7.3	数据类型	145
7.4	常用符号	146
7.5	常用函数	147
7.6	建立存储过程	149
第四部分 数据库系统教学实验		151
第1章	基础实验题	151
第2章	综合实验题	151
2.1	实验一:零件交易中心管理系统	151
2.2	实验二:图书管理系统	152
2.3	实验三:民航订票管理系统	152
第3章	综合实验报告示例	152
3.1	零件交易中心管理系统实验报告(参考答案)	152
3.2	图书管理系统实验报告(参考答案)	163

3.3 民航订票管理系统实验报告(参考答案)	170
第五部分 数据库系统考试大纲	179
第1章 数据库系统概述	179
第2章 数据库建模	179
第3章 关系模型和关系运算	180
第4章 数据库语言 SQL	180
第5章 查询优化和并发控制	181
第6章 关系数据库设计理论	181
第7章 数据库设计	182
第8章 SQL 系统环境	183
第9章 面向对象查询语言	183
第10章 数据库技术发展动态	184
第六部分 数据库系统模拟试题及参考答案	185
第1章 组卷方案	185
1.1 《数据库系统》组卷方案一	185
1.2 《数据库系统》组卷方案二	185
第2章 模拟试题一及参考答案	186
2.1 模拟试题一	186
2.2 参考答案	187
第3章 模拟试题二及参考答案	190
3.1 模拟试题二	190
3.2 参考答案	194
参考文献	196

第一部分 数据库系统概论要点

第 1 章 数据库系统概述

随着计算机软硬件的发展,数据管理技术不断地完善,数据管理技术经历了 3 个阶段:人工管理阶段、文件系统阶段和数据库系统阶段。其中重点是数据库系统阶段。

数据库系统的特点有:数据结构化;数据共享性好;数据独立性好;数据存取粒度小;数据库管理系统对数据进行统一的管理和控制;为用户提供了友好的接口。应该通过理解数据库系统的这些主要特点,对数据库系统的全貌有个初步的了解,能够正确地理解和数据库有关的基本术语。下面列出一些主要的术语及其含义。

(1) **数据:**凡是计算机中用来描述事物的记录,可统称为数据。它包括数字、文字、图形、图像、声音等。

(2) **数据模型:**数据模型是一种对客观事物抽象化的表现形式,它要真实地反映现实世界,易于理解,便于实现。

(3) **数据库:**数据库是按照一定的数据模型组织的、长期储存在计算机内、可为多个用户共享的数据的聚集。也可以说,数据库是由数据库管理系统统一管理和控制的数据的聚集。

(4) **数据库管理系统:**专门用于建立和管理数据库的一套软件,介于应用程序和操作系统之间。

(5) **数据库系统:**包括和数据库有关的数据库、DBMS、应用程序以及数据库管理员和用户等。

关系数据库系统是当前数据库系统的主流,也是本书的重点。在主教材第 1 章,要理解关系数据库中的一些基本概念:关系、关系模型、关系数据库系统等。所谓关系,就是一张表。表的各列以属性开始,属性是列的入口;数据以“关系”的形式,也就是以二维表的形式来表示,其数据模型就是所谓的关系模型;关系数据库系统就是以关系模型为基础的数据库系统。关系模型有严格的数学基础,而且简单清晰,便于理解和使用。

数据库管理系统的组成为 3 个部分。

(1) **查询处理程序:**它不仅负责查询,也负责发出更新数据或模式的请求。

(2) **存储管理程序:**它的功能是从数据库中获得想要查询的数据,并按照更新请求更新相应的信息。

(3) **事务管理程序:**它的作用是保证多个事务并发执行。为此,它要与查询处理程序和存储管理程序互相配合。

第 2 章 数据库建模

主教材第 2 章要求掌握数据库建模的两种基本方法：对象定义语言(ODL)和实体-联系模型(E-R 图)。会用这两种方法建立简单的数据库模型。

ODL(对象定义语言)是用面向对象的术语来说明数据库结构的一种推荐的标准语言,其主要用途是书写面向对象数据库的设计,进而将其直接转换成面向对象数据库管理系统的说明。

在进行 ODL 类的设计时,有 3 种特性需要描述:属性、联系、方法。

在 ODL 中,形式最简单的类的说明包括:关键字 interface(接口)、类的名字、用花括号括起来的类的特性表(特性包括属性、联系和方法)。

要说明一个类,最简单的形式是:

```
interface<名字>{  
    <特性表>  
};
```

两个类之间的联系可分为 3 种不同的类型:多对多、多对一或一对多以及一对一。

E-R 图方法建模也有 3 个主要的部分:实体集、属性、联系。在 E-R 图中,不仅实体集可以有属性,实体集之间的联系也可以有属性。在 E-R 图中,可以用箭头的有无来区别 3 种不同的联系。假如一个联系是从实体集 A 到实体集 B 的多对一的联系,就画一个指向 B 的箭头;假如是 A 到 B 的一对一联系,就画两个箭头,分别指向 A 和 B。

主教材第 2 章中一个很重要的部分就是了解数据库设计的基本原则,并且在数据库设计中能够实践这些原则。这些原则是:真实性、避免冗余、简单性、合理选择元素类型。

类和子类的层次关系反映了现实世界的层次结构。在 ODL 中,定义子类的一般方法是:在类 A 说明的类名 A 之后加上冒号和另一个类 B 的名字,就可以定义类 A 是类 B 的子类。ODL 中的子类和类都与 E-R 模型中的实体集类似。所以在 E-R 图中,假定类 A 是类 B 的子类,类 A 对应于 E-R 图中的实体集 A,类 B 对应于 E-R 图中的实体集 B,为了表示出 A 和 B 之间的关系,用一种称作“属于”(“isa”)的特殊联系将实体集 A 和 B 相连。任何只和子类 A 有关的属性和联系都连到实体集 A 的方框上,而与类 A 和 B 都有有关的属性和联系则连到实体集 B 的方框上。还应该了解子类的继承性。

约束建模是数据库建模的重要组成部分。要深入理解键码和引用完整性这两个基本概念。键码是在类的范围内惟一标识一个对象(在 ODL 中),或者在实体集的范围内惟一标识一个实体(在 E-R 图中)的属性或属性集。一个实体的某个属性(集)值只能引用另一实体确实存在的键码属性(集)值,称为引用完整性。

第3章 关系模型和关系运算

关系模型是关系数据库的基础,因此理解关系模型的基本概念就成为学习关系数据库的基础。下面列出关系模型中的一些概念。

- ① **属性**: 属性就是关系标题栏中各列的名字。
- ② **模式**: 关系的名称和关系的属性集称为关系的模式。
- ③ **元组**: 除了关系的标题栏以外,其他各行统称元组。
- ④ **域**: 与关系的每个属性相关的特定的基本类型称为域。
- ⑤ **关系实例**: 给定关系中元组的集合称为该关系的实例。

在用对象定义语言或实体-联系模型对数据库建模之后,面临的问题就是如何将其转换成所需要的关系模型,这是要求掌握的一个重点内容。

(1) 从 ODL 设计到关系设计

ODL 类转换为关系分为属性和联系两方面的转换。对于原子类型的属性,类的每个属性对应于关系的一个属性。对于非原子属性,若为结构,则把结构中的每个元素作为关系的一个属性;若为集合,则按元素的个数把相关的一个元组扩展为多个元组;若为数组,则按元素的个数既可扩展为多个元组,也可扩展为多个属性。

ODL 中联系的转换: 若为单值联系,则把相关类中构成键码的属性(集)作为关系的附加属性(集)。若为多值联系,如为集合类型,那么,首先把相关类的键码属性(集)作为关系的附加属性(集);其次为相关对象集合的每个元素建立一个关系元组。对于联系与反向联系,常用的方法是将其独立出来作为一个新的关系。

ODL 子类转换成关系模式: 每个子类都对应于一个关系;这个关系用相应子类的所有特性(包括从超类继承下来的全部特性)来表示。

(2) 从 E-R 图到关系设计

实体集与联系转换为关系: 实体集可直接转换为关系,实体集的每个属性都对应于关系中的一个属性。E-R 图中的联系转换为关系时,其属性由两部分组成——与该联系有关的每个实体集的键码属性(集);该联系本身的属性。

多向联系转换为关系也非常类似。不管一个联系 R 涉及几个实体集,在将 R 转换为关系的时候,只要让 R 的属性包括与其相关的所有实体集的键码属性(集)和它本身的属性即可。

对于已经建成的数据库,人们最关心的是查询其中的数据,在学习具体的查询语言之前,先学 3 种抽象的查询语言,以便为以后的学习打下坚实的基础。其中关系代数部分用到的运算如下:

(1) 关系的 3 种集合运算

对于集合 R 和 S,3 种集合运算定义如下:

$R \cup S$ R 和 S 的并,它是 R 中的元素和 S 中的元素共同组成的集合。

$R \cap S$ R 和 S 的交,它是既出现在 R 中又出现在 S 中的元素组成的集合。

$R - S$ R 和 S 的差,它是只在 R 中出现,不在 S 中出现的元素组成的集合。

要想对两个关系 R 和 S 进行上述运算, R 和 S 必须满足如下条件: R 和 S 的模式具有相同的属性集;在对 R 和 S 进行集合运算之前,要对 R 和 S 的属性列进行排序,保证两个关系的属性顺序相同。

(2) 投影

投影运算符是 π , 该运算符作用于关系 R 将产生一个新关系 S, S 只具有 R 的某几个属性列。投影运算的一般表达式如下:

$$S = \pi_{A_1, A_2, \dots, A_n}(R)$$

(3) 选择

选择运算符是 σ , 该运算符作用于关系 R 也将产生一个新关系 S, S 的元组集合是 R 的一个满足某条件 C 的子集。选择运算的一般表达式为:

$$S = \sigma_C(R)$$

(4) 笛卡儿积

两个关系 R 和 S 的笛卡儿积记作 $R \times S$, 它是一个新关系, 它的关系模式是 R 和 S 的模式 的并集, 假如 R 和 S 有同名的属性, 例如 A, 则至少要为其中一个属性重新命名, 通常用 R.A 和 S.A 来区分来自 R 的属性 A 和来自 S 的属性 A。 $R \times S$ 的元组是 R 的一个元组和 S 的一个元组串联而成的长元组。 $R \times S$ 是把 R 和 S 的元组以所有可能的方式组合起来, 因此, $R \times S$ 拥有的元组数量应该是 R 的元组数与 S 的元组数的乘积。

(5) 自然连接

两个关系 R 和 S 的自然连接, 记作 $R \bowtie S$, 得到一个新关系: 它的关系模式是 R 和 S 模式的并集。 $R \bowtie S$ 所拥有的元组是这样产生的: 假设 $A_1, A_2, \dots, A_n$ 是 R 和 S 模式中的公共属性, 那么如果 R 的元组 r 和 S 的元组 s 在这些属性上取值都相同, r 和 s 组合而成的元组就归入 $R \bowtie S$ 中。

(6) θ 连接

两个关系 R 和 S 基于条件 C 的 θ 连接用

$$R \bowtie_C S$$

表示, 它是这样得到的: 先作 R 和 S 的笛卡儿积, 然后从 $R \times S$ 的元组中选择满足条件 C 的元组集合。显然, R 与 S 的 θ 连接其关系模式应该与 $R \times S$ 相同, 即为 R 和 S 模式的并集。

(7) 改名

如下运算

$$\rho_{S(A_1, A_2, \dots, A_n)}(R)$$

用来把关系 R 改名为关系 S, 同时把关系 S 的属性从左至右依次命名为 $A_1, A_2, \dots, A_n$ 。假如只想改变关系名, 不想改变关系模式中的属性名, 那么用如下形式:

$$\rho_S(R)$$

即可达到目的。经过改名运算所得到的关系 S 具有和关系 R 完全相同的元组。

主教材第 3 章要求熟练掌握关系代数语言的使用, 并掌握关系演算语言和关系逻辑语言的使用。

第 4 章 数据库语言 SQL

用结构化查询语言 SQL 对数据库进行查询,是本书的重要内容,也是主教材第 4 章的重点。下面是应该熟练掌握的内容。

(1) 简单查询

SELECT *

FROM 关系名

WHERE 选择条件

(2) 别名

SELECT 属性的原名 AS 别名

(3) 模糊匹配

s LIKE p 和 s NOT LIKE p

其中,s 是普通字符串,p 是模式,表示包含‘%’和‘_’两种任选的具有特殊含义的字符的字符串。p 中的普通字符只和完全相同的字符匹配,而‘%’能和由任意字符组成的任意长度的字符序列匹配,‘_’能和任意一个字符匹配。SQL 的习惯用法是在字符串中用两个连续的单引号来表示一个真正的、单独的单引号,而在%和_之前加上转义符“\”表示这里的%和_是一个普通的字符,而不是为了匹配之用,连续两个\就表示一个真正的反斜线\。

(4) 排序

ORDER BY <属性表>

默认的顺序为升序。也可以在属性的后面加上关键字 DESC 来实现降序输出。SQL 还提供了另一个关键字 ASC 来明确地指定升序输出,不过这个关键字其实没有必要存在,因为默认的顺序就是升序。

(5) 分组

GROUP BY 分组属性列表

如果只希望查询满足一定条件的分组情况,可以使用关键字 HAVING 来选择具有给定条件的分组。

(6) 聚合运算符

- ① **SUM** 求某列中所有值的和;
- ② **AVG** 求某列中所有值的平均值;
- ③ **MIN** 求某列中的最小值;
- ④ **MAX** 求某列中的最大值;
- ⑤ **COUNT** 求某列中值的个数。

对关系中某一列的值可以用关键字 SUM、AVG(平均值)、MIN、MAX 或 COUNT 加以汇总,这 5 种处理统称为聚合。可在聚合之前用关键字 GROUP BY 对元组进行分组,从而得到各分组的汇总数据。还可在分组的基础上用关键字 HAVING 给出选择条件,从而取出具有给定条件的分组。

(7) 连接与笛卡儿积

在 FROM 子句中列出涉及到的每个关系,这样 SELECT 和 WHERE 子句就可以引用 FROM 子句中列出的任何关系的属性了。

(8) 嵌套查询

把 SELECT-FROM-WHERE 查询作为子查询用于另一个查询的 WHERE 子句中,称为嵌套查询。可把运算符 EXISTS、IN、ALL 和 ANY 用于子查询的结果关系来表示布尔值条件;若对表达式求反,可在适当位置加上 NOT。如果能确信子查询的结果为单值(仅一个元组、一个分量),就可以在外层查询的 WHERE 子句中直接用比较运算符。

(9) SQL 的关系模型

SQL 实际上把关系看作是元组的包(Bag)(即允许有完全相同的元组),而不是元组的集合(Set)(不允许有完全相同的元组)。如果要求结果的关系为集合,则需另加关键字 DISTINCT 以删除重复的元组。

(10) 集合查询

SQL 使用关键字 UNION、INTERSECT 和 EXCEPT 分别代表并、交和差。这类关键字用在两个查询之间(将每个查询结果作为一个运算对象),查询必须用圆括号括起来。

当两个查询生成具有相同属性集的关系时,可用关键字 UNION、INTERSECT 和 EXCEPT 进行关系代数中的集合运算:并、交和差,从而实现对多个关系的查询。用上述 3 种运算符将自动删除重复元组,若想保留重复元组,则需在运算符后加上关键字 ALL。

(11) 数据更新

SQL 中的数据更新可利用关键字 INSERT(插入新元组)、DELETE(删除已有元组)或 UPDATE(修改已有元组中的部分分量)来实现。更新语句既可处理单一元组,也可处理一批元组(因格式或给定条件而异)。

(12) 数据定义

对数据库模式中的各个元素进行说明称为数据定义。SQL 的数据定义包括定义和撤销模式(Schema)、表(Table,又称基本表,也就是关系)、视图(View)、索引(Index)等内容。其中,最基本的语句是 CREATE TABLE,用于定义关系(SQL 中称为表)模式,指定其属性和类型,规定各种约束。此外,还可用 ALTER 语句更改模式,比如,在关系模式中增加或删除属性;或者用 DROP 语句完全撤销关系或视图。

(13) 视图

由已有关系定义的、并不实际存在而只是逻辑上的关系称为视图。视图的定义以子查询为基础,这正体现了视图是由已有的关系(或视图)构造而成。对视图可以像对关系一样进行查询,只是具体的执行过程有所不同。

第 5 章 查询优化与并发控制

对于给定的查询选择代价最小的操作序列,使查询过程既省时间,又省空间,具有较高的效率,这就是所谓的查询优化。对于关系数据库系统,用户只需提出“做什么”,而由

系统解决“怎么做”的问题。具体来说,是数据库管理系统中的查询处理程序自动实现查询优化。查询优化是数据库系统实现范畴的一个重要问题。在学会用 SQL 语言对数据库进行查询的基础上,了解数据库系统如何对查询进行优化,有助于在认识上实现从感性到理性的飞跃。

查询优化既有逻辑方法的优化(如关系代数表达式的优化),也有物理方法的优化(如存取路径和存取方法的优化)。由于磁盘的读写比 CPU 的处理在速度上慢得多,因此查询优化的关键就是减少对磁盘的访问。具体策略主要包括:一元选择(只针对一个关系、只涉及一个属性的选择)首先做;投影、选择同时做;乘积、选择合并做(把笛卡儿积与随后以选择形式出现的连接条件一起做连接运算);索引、排序预先做。深入理解查询优化的策略;掌握用关系代数等价变换规则对关系代数查询表达式进行优化的方法;学会按查询优化的步骤对查询进行优化。

安排事务执行的次序称为调度;利用分时的方法同时处理多个事务,则称为事务的并发调度,也就是并发操作的调度。数据库并发操作可能带来数据不一致的问题,主要有 3 种:丢失修改、读“脏”数据和不可重复读。并发控制是数据库系统实现范畴的又一个重要问题。封锁是实现并发控制的主要技术。封锁分排它锁(X 锁)和共享锁(S 锁)两种类型。对 X 锁或 S 锁何时申请(加锁)、何时释放(解锁)均有约定的规则,称之为封锁协议。共有三级封锁协议可分别达到系统一致性的不同级别,依次解决不丢失修改、不读“脏”数据、可重复读等问题。把事务的执行过程分成申请封锁(加锁)阶段和释放封锁(解锁)阶段,这种规则称为两段锁协议。两段锁协议是保证并发调度可串行性的封锁协议。也就是说,凡遵守两段锁协议的任何调度,都是可串行化的调度。

第 6 章 关系数据库设计理论

关系数据库的逻辑设计主要是设计关系模式,而深入理解函数依赖和键码的概念则是设计和分解关系模式的基础。

函数依赖:如果关系 R 中的两个元组在某个特定的属性集 A 上一致,则它们在某个其他特定属性 B 上也一致,就称 B 与 A 之间的关系为函数依赖。具体地说,就是 B 函数依赖于 A,或者 A 函数决定 B。

键码:关系 R 中能函数决定该关系所有属性的最小属性集称为关系 R 的键码。也就是说,键码的任何真子集都不能函数决定该关系的所有属性。

属性的封闭集:对于给定的函数依赖集 S,属性集 A 函数决定的属性的集合称为属性集 A 在依赖集 S 下的封闭集。可利用各种属性组合的封闭集确定关系的键码。要学会计算封闭集。

下面是几个范式的概念。

第二范式:若关系模式 R 属于第一范式,且每个非主属性都完全函数依赖于键码,则 R 属于第二范式。要使关系模式属于第二范式,就要消除非主属性对键码的部分依赖。

第三范式:若关系模式 R 属于第一范式,且每个非主属性都不传递依赖于键码,则 R 属于第三范式。因为部分依赖是传递依赖的特例,所以要使关系模式 R 属于第三范式,