

高等学校计算机专业规划教材

数据库应用、设计与实现

党德鹏 编著

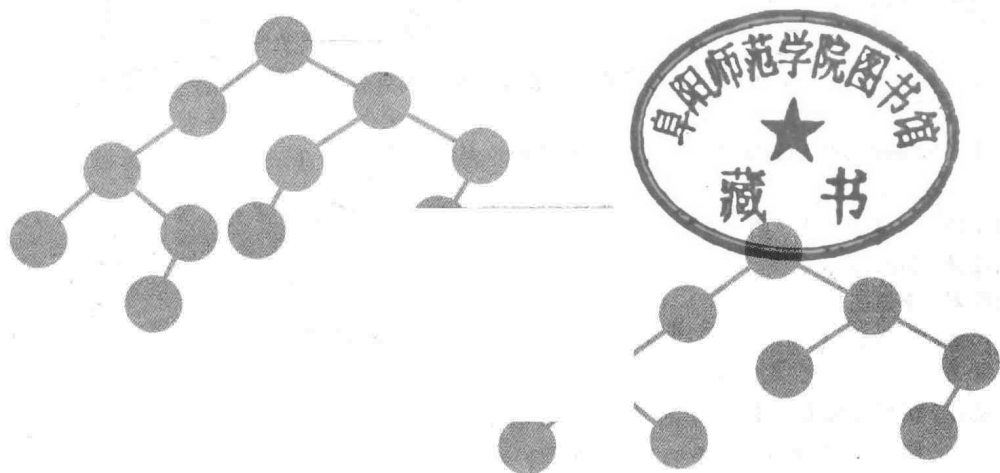
清华大学出版社



高等学校计算机专业规划教材

数据库应用、设计与实现

党德鹏 编著



清华大学出版社
北京

内 容 简 介

近几年数据管理技术发展迅猛,从大数据管理和数据库应用、设计和实现的角度,梳理数据库系统原理课程的知识点体系,调整教材结构,丰富教材内容,乃大数据时代数据库教学的尝试和实践。本书正是基于此目标编写的,全书内容主要包括5大部分,第一部分是基本概念和基础知识,第二部分是SQL语言及其应用,第三部分是数据库设计,第四部分是DBMS实现技术,第五部分是大数据新技术简介。

本书可以作为高等学校计算机、软件工程、信息管理和信息系统等相关专业数据库课程教材,也可供从事数据库系统、信息系统、Web系统、“互联网+”平台系统研究、开发与应用的工程技术人员、科技工作者以及其他相关人员参考阅读。

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

数据库应用、设计与实现/党德鹏编著. —北京:清华大学出版社,2017

(高等学校计算机专业规划教材)

ISBN 978-7-302-46403-7

I. ①数… II. ①党… III. ①关系数据库系统—高等教育—教材 IV. ①TP311.138

中国版本图书馆CIP数据核字(2017)第017909号

责任编辑:龙启铭 薛 阳

封面设计:何凤霞

责任校对:李建庄

责任印制:王静怡

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦A座

邮 编:100084

社总机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课件下载: <http://www.tup.com.cn>, 010-62795954

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:185mm×260mm

印 张:15

字 数:367千字

版 次:2017年3月第1版

印 次:2017年3月第1次印刷

印 数:1~2000

定 价:35.00元

产品编号:072586-01



前言

当前已经是大数据时代,关系数据库一统天下的局面已不可能重现。随着近几年数据管理技术的迅猛发展,传统数据库课程面临一系列挑战,并亟待改革。一方面,在空气中无处不渗透着大数据气息的今天,数据库课程不能没有大数据技术,这是技术发展的呼唤,也是提高学生学习兴趣、提高教学质量、加强人才技术素质的迫切要求。另一方面,传统数据库关键技术和方法仍然是数据库课程必不可少的核心内容,这些技术和方法如今依然广泛应用于管理着各行各业宝贵信息资源的系统,对改进部门管理、提高企业效益、提升人民生活水平均产生着实实在在的效果,而且也是大数据技术的基础和依托,大数据技术则是数据库技术的进一步延伸和发展。在有限教学学时内,融合传统数据库关键技术以及大数据最新进展,乃大数据时代数据库教学的尝试和实践。

本教材按顺序可以分为5大部分。第一部分是基本概念和基础知识,包括第1章绪论和第2章关系模型,都是为后面内容打基础。其中,关于关系代数部分,也曾把关系代数放到后面和查询优化一起讲,试了一届发现把关系代数放第2章SQL语言前面对学生整体把握SQL帮助明显,所以此后放在第2章。总的来说,1、2两章是全书其他部分的基础。第二部分主要分3、4、5三章讲SQL语言及其应用。第三部分是第6章和第7章讲数据库设计。第6章主要是ER设计以及转换为关系,所以也涉及了关系设计或者说目标也是关系设计,但方法主要是从实体及联系的角度来做的;第7章讲关系设计,主要是从数据依赖角度来做的,而数据依赖实质上是属性及其联系。6、7两章的目标是一致的,都是讲关系设计,只是方法不同。第四部分是8、9、10三章,主要讲DBMS共性的技术。第五部分是大数据新技术简介。

本教材主要特色包括:

(1) 站在大数据管理的角度讲述数据库设计和实现的新思想,在数据库设计和实现的讲述中融入大数据思维。

(2) 提出了数据库管理系统的目标:简单、安全、高效地共享数据,并以此为线索贯穿全书内容。

(3) 以PostgreSQL为平台,以网络考试系统为案例,实施案例驱动的教学模式。通过案例分析,解析传统数据管理和大数据管理中数据库技术的基本思想和特点,融合理论与实践,贯通技术思想与职业理念。

(4) 从大数据管理的角度讲述数据管理技术发展趋势,并对大数据技术进行简介。

(5) 在课程内容安排上,主要是先讲语言,让学生通过上机使用数据库,对数据库有直观了解,进而再讲数据库设计,最后讲实现,由浅到深,由表及里,便于理解。

(6) 研磨了与其他课程如信息安全、操作系统、数据结构、组成原理等的关系和衔接。

本教材是北京师范大学首批本科大规模在线开放课程建设项目“数据库系统原理”课程指定教材。

感谢编辑龙启铭老师细致入微的辛勤工作和为本教材顺利出版付出的心血!感谢参考文献中列出和未能列出的老师们及其数据库教材,正是受这些精品的指引一步步踏入数据库领域!感谢历届博士生、硕士生担任助教期间对教材及实验的积极探讨!感谢历届本科生参与的校对工作!感谢专家、老师、同事、朋友们的有益建议和帮助!特别感谢刘莹(中科院)、张笑然(中移动)、黄仕航(秒针)、王楠(中国银行)、姚颖婷(网易)、王俐之(中国银行)、周鹏霞(中国农业银行)、甘锐琦(北京大学)、陶燕飞(中国人寿)、张波(山西大学)、姜雪(北京科技大学)、张莹(中国银行)、阮慧(IBM)、王洪杰(中办)、张永妹(淘宝)、王俊杰(中国建设银行)、徐娟(中航信)、叶璐婷、胡华晓、胡新、王兴建、徐俏、旷洁燕、方真、邹蓉、孟真、王心欣、余文慧、陈闯霞、古丽斯坦·阿卜杜克然木、罗福莉、赵帅帅、张宇、徐冲冲等等。恕不能一一列出。大家的无私付出使得教材不断完善多受裨益!

本教材可按 36~54 学时使用。为了教师教学的方便,本教材配有电子教案、习题解答以及程序代码,所有程序在 Java EE 和 PostgreSQL 9.1 上实际运行通过。针对实践教学,书末附有实验指导书。

在教材编写过程中,尽可能引入新技术,力求反映技术发展趋势,但由于水平有限,定有许多不足之处,还望老师、同行和专家批评指正。

编 者

2017 年 2 月



目录

第 1 章 绪论 /1

1.1	什么是数据库系统	1
1.1.1	数据库	1
1.1.2	DBMS	2
1.1.3	DBS	2
1.2	为什么需要数据库系统	3
1.2.1	DBS 前的困境	3
1.2.2	DBS 的吸引力	4
1.3	数据抽象	5
1.3.1	四层抽象	5
1.3.2	数据抽象的表达	6
1.3.3	三层模式和两级映射	7
1.4	DBMS	8
1.4.1	数据定义语言	10
1.4.2	数据操作语言	10
1.4.3	数据保护语言	10
1.4.4	查询处理	10
1.4.5	存储管理	11
1.4.6	保护管理	11
1.4.7	物理数据结构	11
1.4.8	立足点	12
1.5	DBS	12
1.5.1	硬件	12
1.5.2	软件	13
1.5.3	用户	13
1.5.4	工作过程	14
1.5.5	在网络上	15
1.6	数据库技术发展趋势	16
	习题	17

第 2 章 关系模型 /18

2.1 关系结构与约束	18
2.1.1 关系与表	18
2.1.2 关系键	20
2.1.3 约束	21
2.2 关系操作	21
2.2.1 基本关系代数运算	21
2.2.2 附加关系代数运算	27
2.2.3 扩展关系代数运算	30
习题	33

第 3 章 PostgreSQL 数据定义与操作 /34

3.1 SQL 与 PostgreSQL	34
3.1.1 SQL 发展史	34
3.1.2 PostgreSQL	35
3.1.3 数据库语言组成	35
3.1.4 数据库语言特点	36
3.1.5 考试系统数据库	36
3.2 数据定义	38
3.2.1 SQL 的基本数据类型	38
3.2.2 表的创建、修改和撤销	39
3.3 投影与广义投影	42
3.4 选择	43
3.5 集合操作	47
3.6 联接查询	48
3.6.1 笛卡儿积	49
3.6.2 内联接	49
3.6.3 外联接	50
3.6.4 按属性联接	51
3.7 更名	51
3.8 聚集查询	52
3.8.1 基本聚集	52
3.8.2 分组	53
3.8.3 排名	54
3.8.4 分窗	55
3.9 基本查询语句的一般形式	56
3.10 嵌套查询	57



3.10.1	子查询作为表	57
3.10.2	子查询作为集合	58
3.10.3	子查询作为标量	60
3.10.4	关系除	62
3.11	递归查询	64
3.12	数据修改	65
3.12.1	数据插入	65
3.12.2	数据删除	66
3.12.3	数据更新	67
	习题	68

第4章 PostgreSQL 应用 /70

4.1	应用体系结构	70
4.1.1	C/S 结构	70
4.1.2	B/S 结构	71
4.2	嵌入式 pgSQL	73
4.2.1	pgSQL 的宿主使用	73
4.2.2	嵌入式 pgSQL 的使用技术	74
4.2.3	动态 SQL 语句	76
4.3	JDBC 编程	79
4.3.1	JDBC 基础	79
4.3.2	JDBC 程序	80
4.3.3	预备语句	81
4.3.4	元数据	82
4.3.5	Java 应用连接访问数据库实例	82
4.3.6	Java 小应用连接访问数据库实例	83
4.3.7	JSP 连接访问数据库实例	85
4.3.8	Servlet 连接访问数据库实例	86
4.4	函数	92
4.4.1	PL/pgSQL 的块结构	92
4.4.2	变量常量的定义和赋值	92
4.4.3	控制结构	93
4.4.4	存储函数	95
	习题	96

第5章 PostgreSQL 数据保护 /98

5.1	数据保护	98
5.2	视图	99



5.2.1	视图的创建和撤销	99
5.2.2	对视图的操作	100
5.3	访问控制	101
5.3.1	授予权限	101
5.3.2	收回权限	103
5.4	完整性约束	103
5.4.1	约束含义	103
5.4.2	声明及检验	104
5.5	触发器	109
5.5.1	定义触发器	109
5.5.2	激活触发器	112
5.5.3	删除触发器	112
5.6	事务	112
	习题	114
第6章 数据库设计: 实体-联系方法 /115		
6.1	数据库设计过程	115
6.2	E-R 模型的基本元素	116
6.3	E-R 图	118
6.4	属性的分类	119
6.5	联系的设计	120
6.6	弱实体集	123
6.7	特殊化与一般化	124
6.8	将 E-R 模式变换为关系模式	125
	习题	128
第7章 数据库设计: 属性-联系方法 /130		
7.1	问题的提出	130
7.2	数据依赖	132
7.2.1	函数依赖的定义	132
7.2.2	函数依赖的逻辑蕴涵	133
7.2.3	函数依赖的推理规则	133
7.2.4	属性集的闭包	135
7.2.5	函数依赖集的最小依赖集	136
7.2.6	多值依赖	137
7.3	模式分解	137
7.3.1	无损联接分解	139
7.3.2	分解无损联接检验	139



7.3.3	保持函数依赖的分解	143
7.4	范式	145
7.4.1	第一范式(1NF)	145
7.4.2	第二范式(2NF)	146
7.4.3	第三范式(3NF)	146
7.4.4	BC 范式(BCNF)	147
7.4.5	第四范式	147
7.5	规范化	148
7.6	大数据与反规范化	149
	习题	149

第 8 章 存储和存取 /152

8.1	存储器件	152
8.2	磁盘	152
8.3	DBMS 文件管理	154
8.4	数据库文件组织	154
8.4.1	行存储	154
8.4.2	列存储	158
8.5	文件中元组组织	158
8.6	索引	160
8.6.1	稠密索引和稀疏索引	161
8.6.2	多级索引	162
8.6.3	B+树索引	163
8.6.4	哈希方法	165
8.7	数据字典的存储	167
	习题	167

第 9 章 查询处理与优化 /169

9.1	查询处理过程及查询优化问题	169
9.2	关系代数表达式的等价变换与优化	171
9.2.1	关系代数表达式等价变换规则	171
9.2.2	关系代数等价变换的启发式规则	174
9.3	实现关系运算的算法与优化	174
9.3.1	选择运算的算法与优化	174
9.3.2	联接运算的算法与优化	175
9.4	表达式的求值方法与优化	177
9.4.1	实体化	178
9.4.2	流水线	178



9.5 基于代价的定量优化	178
习题	179

第 10 章 事务处理 /181

10.1 事务概念	181
10.1.1 如果没有事务	181
10.1.2 事务及其特性	182
10.2 并发执行和调度	185
10.2.1 并发执行	185
10.2.2 可串行化	187
10.3 并发控制	189
10.3.1 锁	189
10.3.2 两阶段封锁	190
10.3.3 死锁	191
10.4 故障恢复	191
10.4.1 恢复准备	191
10.4.2 恢复处理	194
10.5 小结	196
习题	196

第 11 章 大数据技术 /198

11.1 大数据特征	198
11.2 大数据关键技术	198
11.3 分布式文件系统	199
11.3.1 计算机集群	199
11.3.2 分布式文件系统	200
11.4 NoSQL 数据模型	200
11.4.1 键值存储	200
11.4.2 列族存储	201
11.4.3 文档存储	201
11.4.4 图存储	203
11.5 大数据计算	203
11.5.1 批处理	204
11.5.2 流式计算	205
11.6 大数据应用	206
11.6.1 基于内容推荐	206
11.6.2 协同过滤推荐	207
11.6.3 小结	209



习题.....	209
---------	-----

附录 实验 /210

实验一 Access 数据库	210
实验二 PostgreSQL 基础和安装	210
实验三 数据库的基本操作	211
实验四 数据表的基本操作	212
实验五 数据备份与还原	214
实验六 简单数据查询	216
实验七 高级数据查询	217
实验八 Java 连接数据库	218
实验九 简单的 PL/pgSQL 程序	219
实验十 索引和视图	220
实验十一 PostgreSQL 权限管理	221
实验十二 触发器	222
实验十三 性能优化	222
实验十四 事务与并发控制	223
实验十五 PowerDesigner	224
实验十六 综合应用	224
实验评分标准	225
实验报告要求	225
实验报告模板	226

参考文献 /227

1.1 什么是数据库系统

1.1.1 数据库

包含关于某单位、机构、部门,或是某领域、业务、主题,或是某对象(如某个人)的信息的互相关联的数据集合称作数据库(Database)。

如今,数据库已应用到生产生活的方方面面。

如高等学校管理信息系统数据库。它包含学生、课程、学生选课数据,教师教学数量和评价数据,科研项目和论文数据;也可以包括课程习题集、答案集以及试卷和考生答卷细节、得分等数据。

如应急平台中心数据库。它包含支持突发公共事件(如地震、洪水、干旱、泥石流、动车追尾、生产安全等)的应急综合管理、监测防控、预测预警、智能方案、指挥调度、应急保障、应急评估等业务的人口、经济、资源、地理等背景数据,重大自然灾害事件时间、地点、强度、受灾人畜等数据,历史自然灾害损失、救援、恢复重建等数据,还有预案数据等。

如社交网络微信、微博、博客和腾讯 QQ 中的数据库,都保存了用户基本注册信息数据、相互发送信息数据、评论数据、转载数据、微信支付相关的数据等;电子商务平台京东、淘宝和亚马逊等的数据库,都保存了商家及其供销商品数据和客户及其购买、订单处理等数据。技术本质上,社交网络、电子商务等互联网+应用就是大量用户通过网络交互访问平台数据库。

如云平台监测管理系统依赖数据库保存整个云平台的实时运行状态、各个应用基本信息和应用状态,服务及其基本信息以及服务状态,平台设备即所有虚拟机运行状态,应用生命周期内的日志信息。系统能以折线图、柱状图等形式展示平台运行参数指标的实时监控数据,如 CPU 使用率、应用负载、应用响应时间等,当平台监测到数据异常时便会预警,提示人工干预。

如电信数据库保存客户通话、短信及网络流量等数据,以及账户余额及月账单支付数据;银行数据库保存客户账户、存取款、贷还款数据;航空、铁路、公路运输数据库保存运输班次、购票数据;企业信息数据库中保存员工工资、工作量、津贴、所得税,配件购买、存放,零件逐环节加工情况和产品销售数据。

数据库是包含了关于某方面信息的互相关联的数据集合。这里,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。现代计算机系统广泛应用于各种领域,存储和处理的对象也非常广泛,表示这些对象的数据也越来越复杂多样了,除了数值,还可以是文本、图形、图像、音频、视频等。

1.1.2 DBMS

包含互相关联的数据集合的数据库是广义的,可以是各种形式,例如可以是打印或写在纸上的;或者是保存在 txt 文件、Word 文件、Excel 或者其他文件中的;或用 PostgreSQL、Access、Oracle 等软件系统管理起来的。

本课程所讲的数据库是特指用专门通用软件管理,长期存储在计算机内、有组织、可共享的大量数据的集合,这些数据库构成了现代企业、大学、国家、政府机构、公司单位等的信息化基础,帮助其大大提高管理效率和水平。数据库可以科学地组织和存储数据、高效地获取和维护数据,从而提供一个可以方便、安全地存取信息的环境,将这类专门的、通用的软件称为数据库管理系统(Database Management System, DBMS)。DBMS 是位于用户与操作系统之间的一层软件,它是一个大型的复杂的系统软件。领先的著名数据库管理系统有 PostgreSQL、IBM DB2、Oracle、Microsoft SQL Server 等。

1.1.3 DBS

数据库系统是指面向数据管理应用、在计算机系统中引入数据库管理系统之后的整个系统,它一般由硬件系统、操作系统、数据库、数据库管理系统(及其开发工具)、应用系统、常规用户和数据库管理员(Database Administrator, DBA)等构成,即包括数据库分析、设计、实施、运行和维护涉及的硬件、软件和人员。

如今,数据库系统已然成为几乎所有企业不可或缺的重要部分,是人们日常生活中最普遍(直接或间接地)使用的技术之一。无论有没有意识到,现代社会中人们每天都在使用数据库系统,例如查看或打印课程成绩单就是对学校教务数据库系统的访问;每次通电话双方的号码及通话时间和时长都会被通信公司通话数据库系统自动保存,查看或打印电话费单也是对通信公司通话数据库系统的访问;通过银行的出纳员或 ATM 机存取款、打印银行卡明细等都是对银行业务数据库系统的访问;预订火车票/飞机机票就是对铁路票务数据库/航空票务数据库系统的访问。如今,越来越多地通过 Web 界面访问大量的在线服务和信息数据库系统。当访问当当在线书店,浏览数据库教材或足球鞋时,其实就是正在访问存储在当当数据库系统中的数据。当你往购物车里添加了商品,或下了订单,都是由当当数据库系统保存相关数据;在选购的过程中,当当会根据选购者之前购买过的商品、正在选购浏览的商品以及之前在当当购买过类似商品的客户购买历史,预测当前选购者可能会购买的商品并给予推荐,这都是当当数据库系统提供的功能。

数据库系统与 Web 技术的结合,既带给 Web 交互性,又带来随时随地访问数据库系统的便捷性,大大推动了数据库系统和 Web 技术的普及,以及各行各业的信息化水平。远不止此,数据库系统与 Web 技术这两种技术的联姻孵化出当今最时髦的术语“云计算/大数据”所蕴含的新计算时代。Google、Amazon、Facebook 等这些根植于网上数据管理的 Web 公司,既积累了巨大量的数据(2015 年,Google 每天处理的数据量达 8.5TB),同时也建立了由许多分布于世界各地的拥有成千上万服务器聚集的数据中心组成的计算平

台。立足于技术更好地服务于人,诸如 Google、Facebook、YouTube、Yahoo、Twitter 和 Amazon,无一例外,它们的业务主要以“Web+数据库”的方式运作,同时它们的计算平台本身的运行管理也依赖“Web+数据库”方式以提高管理自动化。它们所不断积累的“大数据”,既源于数据库,又对数据库技术提出了新挑战,推动数据管理技术不断发展。本教程的最后部分将专门介绍大数据相关关键技术。

1.2 为什么需要数据库系统

如今,专门生产通用数据管理软件——数据库管理系统(DBMS)的 Oracle 公司已经成为世界上最大的软件公司之一;数据库管理系统及其衍生品是微软、IBM 等许多世界顶尖级公司产品链中极其重要的组成部分;Google、Baidu、Amazon 是数据管理技术的先锋,数据库系统是其企业的“心脑血管”。不仅如此,数据库系统几乎成为所有软件不可或缺的核心部件。数据库系统之所以如此重要并广泛地应用于社会生活的方方面面,是因为数据库系统以数据库管理系统为核心,支持安全、方便地共享数据,或者说是安全、简单、高效地共享数据。

1.2.1 DBS 前的困境

在数据库系统出现以前,数据管理自动化是通过操作系统文件来实现的。例如某高校想要保存该校所有老师、学生和考试的信息,就将它们存放在操作系统文件(例如二进制文件)中。对这些信息的访问,需设计相应的应用程序对文件进行操作,可能包括:教师基础数据文件,以及打开教师数据文件并从中对某个或某些教师具体数据的查询、更新,新教师数据插入或过期教师数据删除的应用程序;学生基础数据文件,以及打开学生数据文件并从中对某个或某些学生具体数据的查询、更新,新学生数据插入或过期学生数据删除的应用程序;学生考试数据文件,以及打开学生考试数据文件并从中对某个或某些学生的某次考试具体数据的查询、更新,新报考数据插入或过期考试数据删除的应用程序。这些应用程序由程序员根据当时确定的需求编写,之后当需求发生变化时,就需要修改相应的应用程序;当有新的需求出现时,就需要加入相应的新应用程序。随着时间的推移,越来越多的数据文件和应用程序就会加入到系统中。

像这样使用文件系统来对数据进行管理,使得数据相对孤立、冗余,导致数据共享访问和安全保护都很困难。

使用文件系统管理数据,各个文件是分散的、独立的、专用的、私有的,相同的信息可能在不同地方的不同文件中重复存储;文件系统主要负责建立文件、读写文件,通常仅提供整个文件层面的共享和保护,而这相对数据项共享和安全保护的需求来说是非常粗糙的。例如,某个学生的联系方式和身份证号码既出现在教务信息文件中,又出现在学生工作信息文件,以及宿舍管理信息文件中。这种冗余除了导致存储开销增大外,还可能导致数据不一致性,即同一数据的不同副本不一致。例如,某个学生联系方式的更改可能在学生工作信息文件元组中得到反映而在系统的其他地方(如教务信息文件)却没有及时更新。

使用文件系统管理数据,如果系统使用时间跨度较长,通常系统不同功能部分的文件和程序会在不同时间由不同程序员创建。不同程序可能采用不同的程序设计语言编写,数据分散在不同文件中,这些文件又可能具有不同的结构,编写应用程序来共享不同文件中的数据是很困难的。例如,学生信息文件是.txt文件,而教务信息文件是二进制文件,即使功能相近的程序也不能互相访问,跨文件数据访问需仔细考虑文件格式和文件内部结构。

再例如,假设系统中原15位身份证号须升级为18位,这既需要编写多个专门应用程序对多个数据文件中的每一个身份证号一一进行修改,同时也不得不对已经存在的每一个应用程序仔细找出每一受影响之处并逐个修改。也就是说,应用的小小变化都可能导致雪崩式的数据文件和应用程序修改或编写工作量。而实际情况总是在不断地发展变化中,造成无休止编写大量功能相近的新应用程序的困境。

1.2.2 DBS 的吸引力

不仅考试系统、教学系统,还有各行各业的业务系统,如工矿企业生产管理系统、客户管理系统、销售系统、电子商务系统等,都涉及大量数据,需要有效组织存储数据,并以数据项插入、删除、更新和查询等为基础,安全高效地实现各自关键功能。抽取所有类似应用需求的共性,实现为通用数据库管理系统,极大地简化了相应应用的开发。数据库系统的核心数据库管理系统从应用需求数据管理的共性特征出发,能有效支持安全、方便的数据管理,已经极大地、并将继续推动各行各业的信息化发展。

数据库系统的核心数据库管理系统集中管理数据,并提供数据的一个多层抽象,既使得系统易于扩展以适应应用需求的变化,又允许用户以一种非常接近自然语言的简单方式高效地访问数据,并能(包括大量并发访问、故障情况下)有效保护数据安全。

数据库管理系统能自动防止明显不正确的数据进入数据库。例如,大学某门课实行0~100分的百分制,使用数据库管理系统时只需对此进行简单声明,如果出现0~100以外的分数,数据库管理系统就会自动报错。而对于文件系统,开发者需要仔细检查处理应用程序中所有可能涉及分数输入或修改的代码。

数据库管理系统能在出现故障时自动保护数据。例如学校通过银行划扣学费,银行欲把学生账户A中的5000元转入学校账户B。假设在程序的执行过程中发生了系统故障,很可能A账户上减去的5000元还没来得及存入B账户,这就导致5000元无端消失。使用数据库管理系统时只需对此进行简单的事务性声明,数据库管理系统就能自动保证这里的加和减两个操作要么都发生,要么都不发生。而这对于文件系统几乎无法实现。

数据库管理系统能有效协调大量并发访问以自动对数据进行保护。例如,选课期间通常有万余名学生在短时间内密集同时访问选课数据库,其中,并发的数据库访问相互干涉,可能导致数据出错。艺术鉴赏课今有30个上课名额,假如选课程序如此操作:读取空缺名额数,对其减1,然后将结果写回,当有多个学生几乎同时进行选课操作,即多个选课程序并发执行,可能他们同时读到的空缺额都是2,并分别写回1到数据库,这样上课名额还剩1个。而实际上这两种结果都是错的,正确的值应该是0个。当数据可能被多个不同的应用程序同时访问时,数据库管理系统能自动地协调,而对于文件系统则会因数

据量之大导致信号量枯竭从而无法实现。

数据库管理系统能简单地声明并自动实施安全策略,并非数据库系统的所有用户都可以访问所有数据。例如在大学信息系统中,每个主讲教师只能看到自己主讲课程的学生信息并可以登入成绩,每个学院的教务老师能看到自己学院全体学生的课程和成绩,教务处教务老师能看到全校学生的课程和成绩,对已经登录入库的成绩只有教务处老师在校定后有权修改等。对于数据库管理系统只需简单声明,就能很好地自动实施检查保护。相反,对于文件系统,由于应用程序总是即兴加入到系统中来,在不同应用程序中提供这种保护很复杂。

总之,数据库系统的核心数据库管理系统支持安全、方便地共享数据,以及对需求的变化做出更快的反应。本课程从数据库系统的应用、设计和实现等方面来讲述数据库系统建设的关键技术方法和核心思想。

1.3 数据抽象

1.3.1 四层抽象

一个实际可用的系统必须能高效地管理数据,这就需要使用一定的结构来表示数据。一个数据库系统涉及的用户多种多样,不同用户具有不同的岗位职责,各用户没有必要全面了解数据库中用来表示数据的复杂数据结构;不同用户观察、认识和理解数据的范围、角度和方法不同,涉及不同的数据抽象,即不同用户从不同角度来观察数据库中的数据。数据抽象的相互关系如图 1-1 所示。

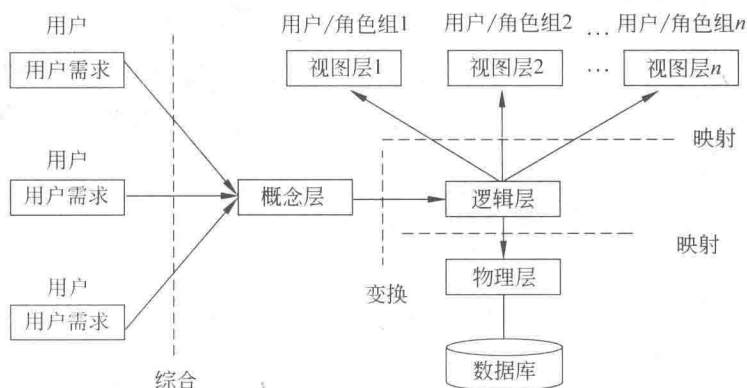


图 1-1 数据抽象的层次

(1) 概念层。对现实世界事物状态选择、加工、组织,形成对全部用户数据需求的认识。

(2) 物理层。“物理”意指具体的硬件设备,如存储器、磁带、磁盘等。物理层是最低层次的抽象,即数据实际上是怎样在辅助存储设备上组织的。

(3) 逻辑层。由 DBMS 提供,通过便于人理解的相对简单的结构来描述数据库中存储什么数据及这些数据之间存在什么关系,一个大型数据库通常有很多用户,逻辑层描述