

21世纪高等学校计算机规划教材

21st Century University Planned Textbooks of Computer Science

Access数据库基础 与应用教程

DataBase System and Access Applications

巫张英 编著

- 结构合理、由浅入深
- 图文并茂、实例丰富
- 内容覆盖等级考试二级



高校系列



人民邮电出版社
POSTS & TELECOM PRESS

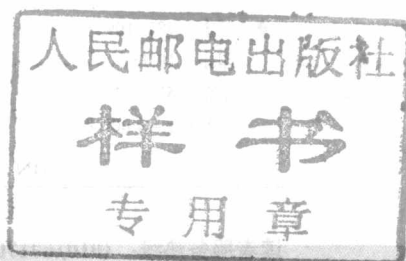
21世纪高等学校计算机规划教材

21st Century University Planned Textbooks of Computer Science

Access数据库基础 与应用教程

DataBase System and Access Applications

巫张英 编著



高校系列

人民邮电出版社

北京

图书在版编目(CIP)数据

Access数据库基础与应用教程 / 巫张英编著. — 北京: 人民邮电出版社, 2009. 11

21世纪高等学校计算机规划教材
ISBN 978-7-115-21463-8

I. ①A… II. ①巫… III. ①关系数据库—数据库管理系统, Access—高等学校—教材 IV. ①TP311.138

中国版本图书馆CIP数据核字(2009)第177032号

内 容 提 要

本书全面介绍了使用 Access 2003 创建数据库、创建数据库的各种对象(包括表、查询、窗体、报表、页、宏和模块)、VBA 编程、DAO 编程和 ADO 编程的方法。全书内容覆盖了全国计算机等级考试二级 Access 数据库程序设计考试大纲(部分二级公共基础知识除外)的基本内容。

本书可作为各大专院校所有专业的有关数据库应用基础课程的参考书,也可作为参加“全国计算机等级考试二级 Access 数据库程序设计”科目考试的考生的参考书,还可作为各层次 Access 用户的自学参考用书。

21 世纪高等学校计算机规划教材

Access 数据库基础与应用教程

◆ 编 著 巫张英

责任编辑 刘 博

◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号

邮编 100061 电子函件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

北京昌平百善印刷厂印刷

◆ 开本: 787×1092 1/16

印张: 19.75

字数: 512 千字

印数: 1—3 000 册

2009 年 11 月第 1 版

2009 年 11 月北京第 1 次印刷

ISBN 978-7-115-21463-8

定价: 34.00 元

读者服务热线: (010)67170985 印装质量热线: (010)67129223

反盗版热线: (010)67171154

现今社会对人才的基本要求之一就是应用计算机的能力。在高等学校,培养学生应用计算机的能力,主要是通过计算机课程的体制改革,即计算机教学分层、分类规划与实施;密切联系实际,恰当体现与各专业其他课程配合;教学必须以市场需求为导向,目的是培养高素质创新型人才。

人民邮电出版社经过对教学改革新形势充分的调查研究,依据目前比较成熟的教学大纲,组织国内优秀的有丰富教学经验的教师编写一套体现教学改革最新形势的“高校系列计算机教材”。在本套教材的出版过程中,我社多次召开教材研讨会,广泛听取了一线教师的意见,也邀请众多专家对大纲和书稿做了认真的审读与研讨。本套教材具有以下特点。

1. 覆盖面广,突出教改特色

本套教材主要面向普通高等学校(包括计算机专业和非计算机专业),是在经过大量充分的调研基础上开发的计算机系列教材,涉及计算机教育领域中的所有课程(包括专业核心骨干课程与选修课程),适应了目前经济、社会对计算机教育的新要求、新动向,尤其适合于各专业计算机教学改革的特点特色。

2. 注重整体性、系统性

针对各专业的特点,同一门课程规划了组织结构与内容不同的几本教材,以适应不同教学需求,即分别满足不同层次计算机专业与非计算机专业(如工、理、管、文等)的课程安排。同时本套教材注重整体性的策划,在教材内容的选择上避免重叠与交叉,内容系统完善。学校可根据教学计划从中选择教材的各种组合,使其适合本校的教学特点。

3. 掌握基础知识,侧重培养应用能力

目前社会对人才的需要更侧重于其应用能力。培养应用能力,须具备计算机基础理论、良好的综合素质和实践能力。理论知识作为基础必须掌握,本套教材通过实践教学与实例教学培养解决实际问题的能力和知识综合运用能力。

4. 教学经验丰富的作者队伍

高等学校在计算机教学和教材改革上已经做了大量的工作,很多教师在计算机教育与科研方面积累了相当多的宝贵经验。本套教材均由有丰富教学经验的教师编写,并将这些宝贵经验渗透到教材中,使教材独具特色。

5. 配套资源完善

所有教材均配有PPT电子教案,部分教材配有实践教程、题库、教师手册、学习指南、习题解答、程序源代码、演示软件、素材、图书出版后要更新的内容等,以方便教与学。

我社致力于优秀教材的出版,恳请大家在使用的过程中,将发现的问题与提出的意见反馈给我们,以便再版时修改。

人民邮电出版社

前言

随着计算机广泛应用于各个领域,数据库技术的应用也得到快速的发展,现在许多单位的业务开展都离不开数据库,如银行业务、各种交通部门售票业务、企业的电子商务、政府的电子政务、教务管理、财务管理等都离不开数据库。从当前信息时代特征来看,不管是理科、文科、医科、工科,还是其他学科的本科生、专科生,学习数据库应用技术知识都很有必要。有些大学已对各专业学生增开了有关数据库应用技术的课程,甚至有些大学还把“数据库基础与应用”课程作为全校非计算机专业学生的公共课。

根据编者从1990年以来曾经成功开发出 Oracle、FoxPro、Access、Sybase 等多种类型的数据库应用系统的经验和教授数据库有关课程的教学体会,觉得现在的大学生至少要掌握一种数据库的应用技术。而 Access 2003 是美国微软公司推出的,基于关系数据模型的数据库管理系统,它作为 Office 2003 的一部分,具有与 Word 2003、Excel 2003 和 PowerPoint 2003 等相同的操作界面和使用环境,功能强,使用方便,深受广大用户的喜爱,故以 Access 2003 数据库管理系统为基础来学习数据库应用知识可行。目前,国内高校(包括本科和专科)的计算机基础实验室大都还是使用 Windows XP 和 Office 2003,许多学校的有关数据库基础的课程也是以 Access 数据库为平台来讲授和组织上机实验,教育部考试中心每年两次组织“全国计算机等级考试二级 Access 数据库程序设计”科目的考试。综合上述学习数据库应用知识的必要性,而编写了本书。

本书是先编写出讲义,使用该讲义进行了一个学期的教学实践,在此基础上,再修订成本书。本书结构合理、由浅入深、图文并茂,以丰富的实例展示了使用 Access 2003 创建数据库及其各种数据库对象的基本知识和技巧,可操作性强。本书内容覆盖了全国计算机等级考试二级 Access 数据库程序设计考试大纲(部分二级公共基础知识除外)的基本内容。

为了帮助教师使用本教材进行教学,编者准备了这本书的教学辅导课件,包括各章的电子教案(ppt 文档)、书中实例数据库等,可从人民邮电出版社教学资源与服务网(www.ptpedu.com.cn)上下载。

在编写本书的过程中,林卓然教授给予了大力支持和帮助,薛春香老师认真看完本书讲义的初稿并提出了许多宝贵意见,何丁海老师也对本书提出了宝贵的建议,在此表示衷心感谢!

由于编者水平有限,书中错误之处难免,敬请广大读者指正。

编 者

2009年7月

目 录

第 1 章 数据库基础概述.....1

1.1 数据管理发展概况.....1

1.1.1 数据及数据处理.....1

1.1.2 人工管理.....1

1.1.3 文件系统.....2

1.1.4 数据库系统.....2

1.1.5 分布式数据库系统.....2

1.1.6 面向对象数据库系统.....3

1.1.7 数据仓库.....3

1.1.8 数据挖掘.....3

1.2 数据库系统简述.....3

1.2.1 数据库.....3

1.2.2 数据库管理系统.....4

1.2.3 数据库系统.....4

1.2.4 数据库应用系统.....5

1.2.5 数据库系统的三级模式及二级映射结构.....5

1.3 数据模型.....6

1.3.1 概念数据模型——E-R 模型.....7

1.3.2 逻辑数据模型.....8

1.3.3 物理数据模型.....10

1.4 关系数据库.....10

1.4.1 关系模型的基本术语.....10

1.4.2 关系的性质.....11

1.4.3 关系完整性约束.....11

1.4.4 关系规范化.....12

1.5 关系运算.....12

1.5.1 传统的集合运算.....12

1.5.2 专门的关系运算.....14

1.6 数据库设计简述.....16

1.6.1 数据库应用开发.....16

1.6.2 数据库需求分析阶段.....16

1.6.3 数据库的概念设计阶段.....17

1.6.4 数据库的逻辑设计阶段.....17

1.6.5 数据库的物理设计阶段.....17

习题.....17

第 2 章 Access 2003 数据库设计与创建.....20

2.1 Access 2003 的启动与退出.....20

2.1.1 启动 Access 2003.....20

2.1.2 Access 2003 窗口简介.....20

2.1.3 退出 Access 2003.....21

2.2 Access 数据库的设计.....21

2.2.1 Access 数据库设计步骤.....21

2.2.2 “学生管理系统”数据库设计实例.....22

2.3 Access 数据库的创建.....25

2.3.1 创建空数据库.....25

2.3.2 使用模板创建数据库.....26

2.3.3 Access 数据库对象简介.....29

2.4 数据库的打开与关闭.....32

2.4.1 打开 Access 2003 数据库.....32

2.4.2 关闭 Access 2003 数据库.....33

习题.....33

第 3 章 表.....34

3.1 表结构设计概述.....34

3.1.1 字段的命名规定.....35

3.1.2 字段的数据类型.....35

3.1.3 学生管理系统数据库的表结构设计实例.....37

3.2 创建表.....38

3.2.1 创建表的方法	38	4.1.1 查询的类型	72
3.2.2 使用设计器创建表	39	4.1.2 创建查询的方法	74
3.2.3 使用向导创建表	40	4.2 创建选择查询	74
3.2.4 通过输入数据创建表	43	4.2.1 使用向导创建查询	74
3.2.5 修改表的结构	44	4.2.2 在设计视图中创建查询	77
3.2.6 字段属性的设置	46	4.3 查询的运行和修改	79
3.2.7 设置和取消表的主键	48	4.3.1 运行查询的基本方法	79
3.3 表之间的关系	49	4.3.2 修改查询设计	80
3.3.1 表之间关系类型的确定	49	4.4 查询条件的设置	82
3.3.2 建立表之间的关系	49	4.4.1 表达式	82
3.3.3 “编辑关系”对话框中的关系选项	51	4.4.2 比较运算符	82
3.3.4 修改表之间的关系	52	4.4.3 逻辑运算符	83
3.3.5 删除表之间的关系	52	4.4.4 特殊运算符	83
3.3.6 子表	52	4.4.5 常用字符串函数	84
3.4 表的基本操作	53	4.4.6 常用日期时间函数	85
3.4.1 打开和关闭表	53	4.4.7 设置查询的组合条件	85
3.4.2 在表中添加记录	53	4.5 设置查询的计算	87
3.4.3 在表中修改记录	54	4.5.1 设置查询的总计计算	88
3.4.4 在表中删除记录	54	4.5.2 设置查询的分组总计计算	89
3.4.5 表中记录排序	54	4.5.3 设置查询的自定义计算	91
3.4.6 表中记录筛选	56	4.6 交叉表查询	92
3.4.7 设置表的外观	58	4.6.1 使用向导创建交叉表查询	92
3.4.8 查找表中的数据	62	4.6.2 使用设计视图创建交叉表查询	93
3.4.9 替换表中的数据	64	4.7 参数查询	95
3.4.10 表的重命名	64	4.7.1 在设计视图中创建单个参数的 查询	95
3.4.11 删除表	64	4.7.2 在设计视图中创建多个参数的 查询	96
3.4.12 复制表	65	4.8 操作查询	97
3.5 导入表、导出表与链接表	65	4.8.1 生成表查询	98
3.5.1 导入表	65	4.8.2 追加查询	99
3.5.2 导出表	68	4.8.3 更新查询	101
3.5.3 链接表	68	4.8.4 删除查询	102
习题	70	4.9 SQL 查询	103
第4章 查询	72	4.9.1 SELECT 语句	104
4.1 查询概述	72		

4.9.2 INSERT 语句	108
4.9.3 UPDATE 语句	109
4.9.4 DELETE 语句	110
4.9.5 SQL 特定查询	111
习题	113

第 5 章 窗体

5.1 窗体概述	116
5.1.1 窗体的组成	116
5.1.2 窗体的视图类型	117
5.1.3 窗体的类型	118
5.1.4 主/子窗体	119
5.1.5 控件	119
5.1.6 工具箱	120
5.1.7 窗体和控件的属性	121
5.1.8 窗体和控件的事件与事件过程	121
5.2 创建窗体	121
5.2.1 创建窗体的方法	122
5.2.2 使用自动创建窗体创建窗体	122
5.2.3 使用“窗体向导”创建一个表的窗体	125
5.2.4 使用“窗体向导”创建基于两个表的主/子窗体	127
5.2.5 使用“数据透视表向导”创建窗体	129
5.2.6 使用“图表向导”创建窗体	131
5.2.7 使用“设计视图”创建窗体	133
5.3 在设计视图中进行自定义窗体设计	135
5.3.1 使用“设计视图”创建一个只读的主/子窗体	135
5.3.2 使用“设计视图”创建一个输入窗体	142
5.3.3 使用“设计视图”创建一个含选项卡的控制窗体	145
5.3.4 在窗体“设计视图”中使用“工具箱”创建“选项组”控件	149

习题	151
----------	-----

第 6 章 报表

6.1 报表概述	153
6.1.1 报表的组成	153
6.1.2 报表的视图类型	155
6.1.3 报表的类型	155
6.1.4 创建报表的方法	155
6.1.5 修改报表的设计	156
6.2 创建报表	156
6.2.1 自动创建纵栏式报表	157
6.2.2 自动创建表格式报表	157
6.2.3 使用“报表向导”创建报表	158
6.2.4 使用图表向导创建报表	160
6.2.5 使用标签向导创建报表	162
6.2.6 使用设计视图创建报表	164
6.3 报表设计的一些技巧	169
6.3.1 报表记录的排序	169
6.3.2 报表记录的分组	170
6.3.3 添加计算控件实现计算	173
6.3.4 两级分组统计及百分比计算	174
6.3.5 多列报表	176
6.3.6 子报表	177
6.3.7 报表快照	181
习题	181

第 7 章 数据访问页

7.1 数据访问页概述	183
7.1.1 数据访问页的视图类型	183
7.1.2 数据访问页的组成	183
7.1.3 创建数据访问页的方法	184
7.1.4 修改数据访问页的设计	185
7.2 在当前数据库中创建数据访问页	186
7.2.1 自动创建数据访问页	186
7.2.2 使用向导创建数据访问页	187

7.2.3 使用设计视图创建数据访问页	189	9.1.1 模块类型	216
7.2.4 使用现有网页创建数据访问页	192	9.1.2 模块的组成	217
7.3 在不打开数据库的情况下创建数据访问页	192	9.2 VBA 程序设计概述	218
7.3.1 创建独立的数据访问页	192	9.2.1 对象和对象名	218
7.3.2 修改独立的数据访问页	194	9.2.2 对象的属性	219
7.4 设计和编辑数据访问页	195	9.2.3 对象的方法	221
7.4.1 页设计的工具箱	195	9.2.4 对象的事件	221
7.4.2 创建标签	195	9.2.5 事件过程	222
7.4.3 创建命令按钮	196	9.3 VBA 编程环境——VBE 窗口	224
7.4.4 创建滚动文字控件	196	9.3.1 进入 VBE 编程环境	224
7.4.5 应用主题	198	9.3.2 VBE 窗口	225
7.4.6 设置背景	199	9.3.3 VBE 环境中编写 VBA 代码	227
7.4.7 在数据访问页中添加电子表格	200	9.4 VBA 编程基础	228
7.5 在 IE 浏览器中查看数据访问页	201	9.4.1 VBA 的标准数据类型	228
习题	202	9.4.2 变量	230
第 8 章 宏	204	9.4.3 变量的作用域与生命周期	231
8.1 宏概述	204	9.4.4 数组	231
8.1.1 宏设计窗口	204	9.4.5 用户自定义的数据类型	233
8.1.2 常用的宏操作简介	205	9.4.6 数据库对象变量	234
8.2 宏的创建	207	9.4.7 常量	235
8.2.1 创建操作序列宏	207	9.4.8 运算符	236
8.2.2 创建条件操作宏	207	9.4.9 表达式和运算符的优先级	238
8.2.3 创建宏组	209	9.4.10 常用标准函数	239
8.3 运行宏和调试宏	211	9.5 VBA 程序语句	245
8.3.1 宏的运行	211	9.5.1 VBA 程序语句编写规则	245
8.3.2 宏的调试	213	9.5.2 VBA 注释语句	246
8.4 宏与 VB	213	9.5.3 VBA 声明语句	246
8.4.1 宏与 Visual Basic	213	9.5.4 VBA 赋值语句	246
8.4.2 宏向 Visual Basic 程序代码转换	214	9.5.5 输入/输出的语句	247
习题	214	9.5.6 选择结构	249
第 9 章 模块与 VBA 程序设计	216	9.5.7 循环结构	256
9.1 模块概述	216	9.5.8 标号和 GoTo 语句	262
		9.5.9 过程的声明和调用及参数传递	262
		9.6 VBA 中的常用操作方法	265
		9.6.1 打开和关闭窗口	265

9.6.2 打开和关闭报表	266	第 1 章习题	292
9.6.3 VBA 编程验证数据	267	第 2 章习题	292
9.6.4 计时事件 Timer	268	第 3 章习题	292
9.6.5 几个特殊函数的使用	269	第 4 章习题	292
9.7 VBA 程序调试和错误处理	271	第 5 章习题	293
9.7.1 错误处理	271	第 6 章习题	293
9.7.2 程序的调试	273	第 7 章习题	293
习题	276	第 8 章习题	293
第 10 章 VBA 的数据库编程	279	第 9 章习题	293
10.1 数据库访问接口	279	第 10 章习题	294
10.2 数据访问对象	279	附录 2 上机实验	295
10.2.1 DAO 模型结构	280	附录 2.1 上机实验 1	295
10.2.2 DAO 对象变量的声明和赋值	281	附录 2.2 上机实验 2	295
10.2.3 DAO 对象的部分属性和方法	282	附录 2.3 上机实验 3	296
10.2.4 利用 DAO 访问数据库	283	附录 2.4 上机实验 4	298
10.3 ActiveX 数据对象	286	附录 2.5 上机实验 5	299
10.3.1 ADO 模型结构	286	附录 2.6 上机实验 6	300
10.3.2 ADO 对象变量的声明	287	附录 2.7 上机实验 7	300
10.3.3 ADO 对象的部分属性和方法	287	附录 2.8 上机实验 8	301
10.3.4 利用 ADO 访问数据库	289	附录 2.9 上机实验 9	301
习题	290	附录 2.10 上机实验 10	303
附录 1 习题参考答案	292	参考文献	304

第1章

数据库基础概述

数据库及其应用是计算机科学中一个重要的分支。从最初简单的人工管理数据的方式到当前各种先进的数据库系统,数据管理技术发生了翻天覆地的变化。特别是数据库技术的应用非常快、非常广泛,以至在计算机的应用领域中的许多地方都与数据库应用有关。现在,许多单位的正常业务开展都离不开数据库,如银行业务、证券市场业务、飞机订票业务、火车订票业务、超市业务、电子商务等都离不开数据库。如果支持这些业务的数据库出现故障,那么,有关的业务将立即无法开展。

1.1 数据管理发展概况

自从世界上第一台电子数字计算机(简称计算机)诞生以来,数据管理经历了从较为低级的人工管理到先进的数据库、数据仓库、数据挖掘的演变。

1.1.1 数据及数据处理

数据是存储在某一媒体(如计算机)上能够识别的物理符号。也可以说,数据是描述事物的符号记录,如“黄山”、“95”。数据不仅可以包括数字、字母、文字和其他特殊字符组成的文本形式,而且还可以包括图像、图形、影像、声音、动画等多媒体形式,它们经过数字化后可以存入计算机。

数据处理是把数据加工处理成为信息的过程。

信息是数据根据需要进行加工处理后得到的结果。

信息对于数据接收者来说是有意义的。例如,“黄山”、“95”只是单纯的数据,没有具体意义,而“黄山同学本学期的英语期末考试成绩为95分”就是一条有意义的信息。此外,“旅游景点黄山的门票费是每人95元人民币”也是一条有意义的信息。

1.1.2 人工管理

在计算机诞生初期,计算机主要用于科学计算。由于受到当时硬件和软件技术的限制,外存储器只有纸带、卡片和磁带,而没有磁盘等可以直接进行存取的存储设备。在软件方面,没有操作系统和高级语言,没有数据管理软件。数据处理的方式是批处理。因此,在这种条件下决定了当时的数据管理只能依赖手工进行。那时,计算机系统不提供对用户数据的管理功能。用户编制程序时,必须全面考虑好相关的数据,包括数据的定义、存储结构、存取方法等。程序和数据是

一个不可分割的整体。数据无独立性，数据脱离了程序就无任何存在的价值。数据不能共享。所有程序的数据均不单独保存。

1.1.3 文件系统

在 20 世纪 50 年代后期到 60 年代中期，计算机不仅用于科学计算，还大量用于信息处理。随着数据量的增加，数据的存储、检索和维护问题成为紧迫的需要，数据结构和数据管理技术也迅速发展起来。此时，外部存储器已有磁盘、磁鼓等直接存取的存储设备。软件领域出现了操作系统、高级语言等系统软件。操作系统中的文件系统是专门管理外存的数据管理软件，文件是操作系统管理的重要资源之一。数据处理方式有批处理，也有联机实时处理。

数据可以“文件”形式长期保存在外部存储器的磁盘上。由于计算机的应用转向信息管理，因此对文件要进行大量的查询、修改、插入等操作。

对数据的操作以记录为单位。这是由于文件中只存储数据，不存储文件记录的结构描述信息。文件的建立、存取、查询、插入、删除、修改等所有操作，都要用程序来实现。

随着数据管理规模的扩大，数据量急剧增加，文件系统显露出数据冗余、数据联系弱等缺陷。

1.1.4 数据库系统

数据库系统阶段开始于 20 世纪 60 年代末。随着计算机应用的日益广泛，数据管理的规模也越来越大，需要处理的数据量急剧增长。同时硬件技术的发展，出现了大容量的磁盘。在这种情况下促使人们去研究更加有效的数据管理手段，从而催生了数据库技术，使数据管理进入了数据库系统阶段。应用程序与数据库的关系通过数据库管理系统 (DBMS) 来实现，如图 1-1 所示。

数据库系统克服了文件系统的缺陷，提供了对数据更高级、更有效的管理。数据库系统采用数据模型表示复杂的数据结构。数据模型不仅描述数据本身的特征，还要描述数据之间的联系，这种联系通过存取路径实现。通过所有存取路径表示自然的数据联系是数据库与传统文件的根本区别。这样，数据不再面向特定的某个或多个应用，而是面向整个应用系统。数据冗余明显减少，实现了数据共享。

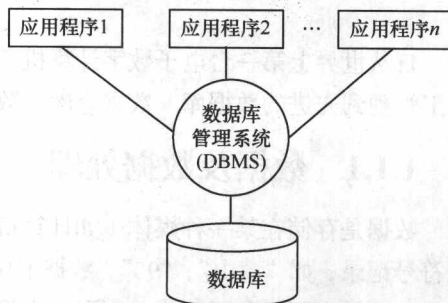


图 1-1 应用程序与数据库的关系

1.1.5 分布式数据库系统

分布式数据库是数据库技术与网络技术相结合的产物。随着传统的数据库技术日趋成熟，计算机网络技术的飞速发展和应用范围的扩充，数据库应用已经普遍建立于计算机网络之上。这时集中式数据库系统表现出它的不足：数据按实际需要已在网络上分布存储，再采用集中式处理，势必造成通信开销大；应用程序集中在一台计算机上运行，一旦该计算机发生故障，则整个系统受到影响，可靠性不高；集中式处理引起系统的规模和配置都不够灵活，系统的可扩充性差。在这种形势下，集中式数据库的“集中计算”向“分布计算”发展。

分布式数据库系统有两种：一种是物理上分布的，但逻辑上却是集中的。另一种在物理上和逻辑上都是分布的，也就是所谓联邦式分布数据库系统。

1.1.6 面向对象数据库系统

将面向对象技术与数据库技术结合产生出面向对象的数据库系统,这是数据库应用发展的迫切需要,也是面向对象技术和数据库技术发展的必然结果。

面向对象的数据库系统必须支持面向对象的数据模型,具有面向对象的特性。一个面向对象的数据模型是用面向对象的观点来描述现实世界实体(对象)的逻辑组织、对象之间的限制和联系等的模型。

另外,将面向对象技术应用到数据库应用开发工具中,使数据库应用开发工具能够支持面向对象的开发方法并提供相应的开发手段,这对于提高应用开发效率,增强应用系统界面的友好性、系统的可伸缩性、可扩充性等具有重要的意义。

1.1.7 数据仓库

随着客户机/服务器技术的成熟和并行数据库的发展,信息处理技术的发展趋势是从大量的事务型数据库中抽取数据,并将其清理、转换为新的存储格式,即为决策目标把数据聚合在一种特殊的格式中。随着此过程的发展和完善,这种支持决策的、特殊的数据存储即被称为数据仓库(Data Warehouse)。数据仓库领域的著名学者 W.H.Inmon 对数据仓库的定义是:数据仓库是支持管理决策过程的、面向主题的、集成的、稳定的、随时间变化的数据集合。

1.1.8 数据挖掘

数据挖掘(Data Mining)又称数据库中的知识发现(Knowledge Discovery in DataBase),它是一个从数据库中获取有效的、新颖的、潜在有用的、最终可理解的知识的复杂过程。简单来说,数据挖掘就是从大量数据中提取或“挖掘”知识。

数据挖掘和数据仓库的协同工作,一方面,可以迎合和简化数据挖掘过程中的重要步骤,提高数据挖掘的效率和能力,确保数据挖掘中数据来源的广泛性和完整性。另一方面,数据挖掘技术已经成为数据仓库应用中极为重要和相对独立的方面和工具。

1.2 数据库系统简述

数据库系统是指引入数据库技术后的计算机系统。数据库系统实际上是一个集合体,一般由硬件系统、数据库、数据库管理系统及其相关的软件、数据库管理员和用户组成。数据库系统的内部体系结构是三级模式和二级映射结构。

1.2.1 数据库

数据库(DataBase, DB)是长期存储在计算机内、有组织的、可共享的、统一管理的相关数据的集合。数据库中的数据按一定的数据模型进行组织、描述和存储,具有较小的冗余度、较高数据独立性和易扩展性。

数据库中不仅包括描述事物的数据本身,而且包括相关事物之间的关系。

数据库中的数据不只是面向某一种特定的应用,而是可以面向多种应用,可以被多个用户、多个应用程序所共享。

例如,某一企业的数据库,可以被该企业下属的各个部门的有关管理人员共享使用,而且可供各个管理人员运行的不同应用程序共享使用。当然,为保障数据库的安全,对于使用数据库的用户应有相应权限的限制。

1.2.2 数据库管理系统

数据库管理系统(DataBase Management System, DBMS)是数据库系统的核心软件,其主要任务是支持用户对数据库的基本操作,对数据库的建立、运行和维护进行统一管理和控制。注意:用户不能直接接触数据库,而只能通过 DBMS 来操作数据库。

DBMS 的基本功能如下。

1. 数据定义功能

DBMS 提供了数据定义语言(Data Description Language, DDL)供用户定义数据库的结构、数据之间的联系等。具体来说,DDL 供用户定义数据库的外模式、模式、内模式、各级模式之间的映射以及有关的约束条件等。

2. 数据操纵功能

DBMS 提供了数据操纵语言(Data Manipulation Language, DML)来完成用户对数据库提出的各种操作要求,以实现数据库的插入、修改、删除、检索等基本操作。DML 分为宿主型 DML 和自主型 DML 两种类型。宿主型 DML 本身不能独立使用,必须嵌入到主语言中,如嵌入 C、Cobol 等高级语言中。自主型 DML 又称为自含型 DML,它是交互式命令语言,可以独立使用。

3. 数据库运行控制功能

DBMS 提供了数据控制语言(Data Control Language, DCL)来实现对数据库进行并发控制、安全性检查、完整性约束条件的检查等功能。它们在数据库运行过程中监视对数据库的各种操作,控制管理数据库资源,处理多用户的并发操作等。

4. 数据库维护功能

DBMS 还提供了一些实用程序,以对已经建立好的数据库进行维护,包括数据库的转储与恢复、数据库的重组与重构、数据库性能的监视与分析等。

5. 数据库通信功能

DBMS 还提供了与通信有关的实用程序,以实现网络环境下的数据通信功能。

1.2.3 数据库系统

1. 数据库系统的组成

数据库系统(DataBase System, DBS)是指引入数据库技术后的计算机系统。数据库系统实际上是一个集合体,通常包括如下 5 部分。

(1) 数据库(DB)。

(2) 数据库管理系统(DBMS)及其相关的软件。

(3) 计算机硬件系统。

(4) 数据库管理员(DataBase Administrator, DBA),全面负责建立、维护、管理和控制数据库系统。

(5) 用户。

2. 数据库系统的特点

数据库系统的基本特点如下。

(1) 数据低冗余、共享性高。

(2) 数据独立性高。数据的独立性包括逻辑独立性和物理独立性。

数据的逻辑独立性是指当数据的总体逻辑结构改变时,数据的局部逻辑结构不变,由于应用程序是依据数据的局部逻辑结构编写的,所以应用程序不必修改,从而保证了数据与程序间的逻辑独立性。例如,在原有的某些记录类型中增加新的数据项,而应用程序不必修改,就体现了数据的逻辑独立性。

数据的物理独立性是指当数据的存储结构改变时,数据的逻辑结构不变,从而应用程序也不必改变。例如,改变存储设备(如换了一个磁盘来存储该数据库),而应用程序不必修改,就体现了数据的物理独立性。

(3) 有统一的数据控制功能。数据控制功能通常包括数据的安全性控制、数据的完整性控制、并发控制等。

1.2.4 数据库应用系统

数据库应用系统是指系统开发人员利用数据库系统的资源,为某一类实际应用的用户使用数据库而开发的软件系统,如银行业务管理系统、仓库管理系统、财务管理系统、飞机售票管理系统、教务管理系统等。

1.2.5 数据库系统的三级模式及二级映射结构

根据美国国家标准化协会和标准计划与需求委员会提出的建议,数据库系统的内部体系结构是三级模式和二级映射结构。

三级模式分别是外模式、概念模式和内模式。

二级映射分别是外模式到概念模式的映射和概念模式到内模式的映射。

三级模式反映了数据库的3种不同的层面(见图1-2)。

以外模式为框架所组成的数据库称作用户级数据库,体现了数据库操作的用户层。

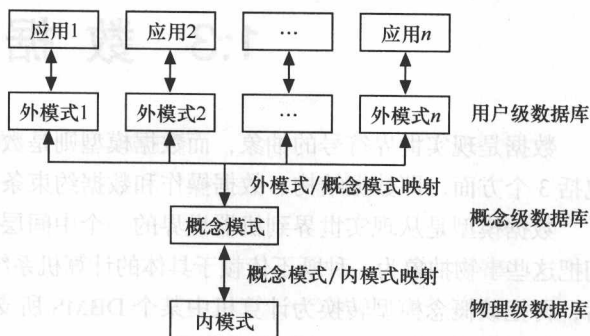


图 1-2 数据库系统的三级模式结构

以概念模式为框架所组成的数据库称作概念级数据库,体现了数据库操作的接口层。

以内模式为框架所组成的数据库称作物理级数据库,体现了数据库操作的存储层。

1. 数据库系统的三级模式

(1) 外模式。外模式也称子模式或用户模式。它是数据库用户(包括应用程序员和最终用户)所见到和使用的局部数据逻辑结构的描述,是数据库用户的数据视图,是与某一应用有关的数据的逻辑表示。一个概念模式可以有若干个外模式,每个用户只关心与他有关的外模式,这样不仅可以屏蔽大量无关信息而且有利于数据库中数据的保密和保护。对外模式的描述,DBMS一般都提供有相应的外模式数据定义语言(外模式DDL)来定义外模式。

(2) 概念模式。概念模式也称为模式。它是数据库中全局数据逻辑结构的描述,是所有用户的公共数据视图。定义模式时不仅要定义数据的逻辑结构(如数据记录由哪些数据项构成,数据项的名字、类型等),而且要定义与数据有关的安全性、完整性约束要求,定义这些数据之

间的联系等。对概念模式的描述, DBMS 一般都提供有相应的模式数据定义语言(模式 DDL)来定义模式。

(3) 内模式。内模式也称存储模式或物理模式,它是数据库物理存储结构和物理存储方法的描述,是数据在存储介质上的保存方式。例如,数据的存储方式是顺序存储、还是按照 B 树结构存储等。内模式对一般用户是透明的,但它的设计直接影响数据库的性能。对内模式的描述, DBMS 一般都提供有相应的内模式数据定义语言(内模式 DDL)来定义内模式。一个数据库只有一个内模式。

2. 数据库系统的二级映射

数据库系统的三级模式之间的联系是通过二级映射来实现的,当然实际的映射转换工作是由 DBMS 来完成的。

(1) 外模式到概念模式的映射。外模式到概念模式的映射(即外模式/概念模式映射)定义了外模式与概念模式之间的对应关系。外模式是用户的局部模式,而概念模式是全局模式。当概念模式改变时,由数据库管理员对各个外模式/概念模式映射作相应改变,可以使外模式保持不变,从而应用程序不必修改,保证了数据的逻辑独立性。

(2) 概念模式到内模式的映射。概念模式到内模式的映射(即概念模式/内模式映射)定义了数据全局逻辑结构与物理存储结构之间的对应关系。当数据库的存储结构改变时(如换了另一个磁盘来存储该数据库),由数据库管理员对概念模式/内模式映射作相应改变,可以使概念模式保持不变,从而保证了数据的物理独立性。

1.3 数据模型

数据是现实世界符号的抽象,而数据模型则是数据特征的抽象。数据模型描述所描述的内容包括 3 个方面,即数据结构、数据操作和数据约束条件。

数据模型是从现实世界到机器世界的一个中间层次。现实世界的事物反映到人的大脑中,人们把这些事物抽象为一种既不依赖于具体的计算机系统又不依赖于具体的 DBMS 的概念模型,然后,再把该概念模型转换为计算机中某个 DBMS 所支持的数据模型。

数据模型包括如下 3 类。

1. 概念数据模型

概念数据模型是面向数据库用户的现实世界的数据模型,也称概念模型。概念模型主要用来描述现实世界的概念化结构。它使数据库的设计人员在设计的初始阶段,摆脱计算机系统及 DBMS 的具体技术问题,集中精力分析数据以及数据之间的联系等。概念模型与具体的计算机平台无关,与具体的 DBMS 无关。

2. 逻辑数据模型

逻辑数据模型也称数据模型。逻辑数据模型主要用来描述数据库中数据的表示方法和数据库结构的实现方法。它是计算机实际支持的数据模型,是与具体的 DBMS 有关的数据模型。它包括层次数据模型、网状数据模型、关系数据模型、面向对象数据模型等。

3. 物理数据模型

物理数据模型也称物理模型,它是一种面向计算机物理表示的模型。物理数据模型给出了数据模型在计算机上物理结构的表示,它是描述数据在储存介质上的组织结构的数据模型。

1.3.1 概念数据模型——E-R 模型

概念数据模型是一种面向客观世界、面向用户的模型。概念数据模型是按用户的观点,从概念上描述客观世界复杂事物的结构以及事物之间的联系,而不管事物和联系如何在数据库中存储。概念数据模型与具体的 DBMS 无关,与具体的计算机平台无关。概念数据模型是整个数据模型的基础。概念数据模型的设计方法有多种,其中的 E-R (实体-联系) 方法是最广泛使用的方法。在此,仅介绍使用 E-R 方法设计概念数据模型——E-R 模型。

1. 概念模型中的基本概念

(1) 实体: 客观存在并可相互区别的事物称为实体。实体可以是具体的人、事、物,也可以是抽象的概念或联系。例如,一名教师、一门课、一本书、一次作业、一次考试等。

(2) 属性: 描述实体的特性称为属性。一个实体可以由若干个属性来刻画,如一个学生实体有学号、姓名、性别、出生日期等方面的属性。属性有属性名和属性值,属性的具体取值称为属性值。例如,对某一学生的“性别”属性取值“女”,其中“性别”为属性名,“女”为属性值。

(3) 关键字: 能够唯一标识实体的属性或属性的组合称为关键字。如学生的学号可以作为学生实体的关键字,但学生的姓名有可能有重名,因此不能作为学生实体的关键字。

(4) 域: 属性的取值范围称为该属性的域。例如,学号的域为 8 个数字字符串集合,性别的域为“男”和“女”。

(5) 实体型: 属性的集合表示一个实体的类型,称为实体型。例如,学生(学号,姓名,性别,出生日期)就是一个实体型。

属性值的集合表示一个实体。例如,属性值的集合(02091001, 李楠, 女, 1986-01-12)就是代表一名具体的学生。

(6) 实体集: 同类型的实体的集合称为实体集。例如,对于“学生”实体来说,全体学生就是一个实体集。

2. 两个实体之间的联系

现实世界中事物内部以及事物之间是有联系的,在概念模型中反映为实体内部的联系和实体之间的联系。实体内部的联系通常是指组成实体的各属性之间的联系,而实体之间的联系通常是指不同实体集之间的联系。两个实体之间的联系可分为如下 3 种类型。

(1) 一对一联系 (1:1): 实体集 A 中的一个实体至多与实体集 B 中的一个实体相对应,反之亦然,则称实体集 A 与实体集 B 之间为一对一的联系,记作 1:1。例如,一个学校只有一位校长,一位校长只能管理一个学校。

(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 中有多个实体与之对应,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对应,则称实体集 A 与实体集 B 之间为一对多联系,记为 1:n。例如,学校的一个系有多个专业,而一个专业只属于一个系。

(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 中有多个实体与之对应,反之,对于实体集 B 中的每一个实体,实体集 A 中也有多个实体与之对应,则称实体集 A 与实体集 B 之间为多对多联系,记为 m:n。例如,一名学生可以选修多门课程,一门课程可以被多名学生选修。

3. E-R 方法

E-R 方法(即实体-联系方法)是最广泛使用的概念数据模型设计方法,该方法用 E-R 图来描