

普通高中课程标准实验教科书

信息技术

选修 4

数据管理技术

高一高二年级用



上海科技教育出版社

编写人员名单

主 编：应吉康

主要编写人员(按姓氏笔画)：

杨福兴 吴 萍 胡 杨 洪如蕙

蔡福民

上海科技教育出版社欢迎广大师生来电来函指出教材的差错和不足,提出宝贵意见。

上海科技教育出版社地址:上海市柳州路218号。

邮政编码:200235。

联系电话:021-64367970转202分机。

传真电话:021-64702835。

写给学生的话

亲爱的同学们：

曾几何时,我们的耳畔时常响起大数据、人工智能、云平台等新概念,我们的生活中经常出现 Web、5G、AR、VR、MR、App 等新名词。你可曾留意过,我们生活中发生的许多新鲜事都与这些新概念、新名词有联系,这些新概念、新名词都出自正在蓬勃发展的现代信息技术。

从打开《数据管理技术》课本起,你将再一次加入到探索信息技术奥秘的行列中,让我们携手共度这一段美好的时光,去领略现代信息技术的无限风光。

如果没有数据库,今天的因特网将不会如此“绚丽多彩”;如果没有数据库,现代企业管理系统将毫无意义;如果没有数据库,上述新事物也都将失去魅力……

当你循着课本中“任务”的足迹——

从身边正在运用计算机“做事”的地方,探寻到数据库的“身影”;

从需要应用计算机的地方,调查到“需要计算机帮什么忙”;

找到可以让计算机做的“事”;

把怎么做这些“事”交代给了计算机;

……

你一定会发现,“不知不觉”之中你也设计了一个数据库。同时你一定能感受到在“多姿多彩”的计算机应用、网络应用背后,数据库这一发明的巧妙极致。

学习数据管理技术能使你发现,如此神通广大的技术原来并不那么神秘,并非高不可攀。

学习数据管理技术能使你发现,有了需要就会有办法,办法是人想出来的;需要不断提出,办法不断产生;想出了与别人不同的办法,就是创造。创造是永无止境的。

学习数据管理技术能使你发现,多具备一些现代信息技术能力的愿望可以实现。

……

或许,明天你也能成为信息技术领域某一项技术的发明者。

为使你在“做”中学数据管理技术有更大的收获,课本里安排了这样一些栏目:



情景 每章开头都呈现了一些背景情况,表明需要解决的真实问题的来源,提出值得思考的问题。



任务 以任务串的形式,呈现每一章中每一节的主体内容。在这一实践园地里,你与同伴将共同合作经历这样的过程:探索解决问题的方法,产生主动学习相应技术的欲望,尝试选择合适的手段和方法解决问题、完成任务……



提示板 在你解决问题、完成任务的过程中伴你同行,适时地在技术上向你提个醒或适当地为你释疑解惑。



学习指引 为你解决问题、完成任务提供技术知识的支持,为你形成技术能力和能力的迁移提供具体的帮助。



体验活动 为你提供不同技术程度的实践机会。你可以根据自己的需要作出选择,使自己的技术本领得到发挥。



后续活动 设在每章之后,为你与同伴综合运用相应技术合作解决问题,提供了实践机会。



评价 是你关注自己学习过程的向导,有助于你感受探索的乐趣,总结学习技术、运用技术解决问题的经验,体会成功的喜悦。



参考资料 设在每章最后,旨在拓展相关内容,反映数据管理技术的最新发展动态。

在这些栏目的帮助下,相信你一定能很快掌握数据管理技术的基本思想和方法,为以后进一步学习打下坚实的基础。

编者

目录



第 1 章 走进数据库	1
第一节 了解数据管理技术	2
任务一 调查图书馆的图书管理技术	2
任务二 了解哪些部门或行业应用计算机管理	5
学习指引	6
■ 信息和数据	6
■ 数据处理和数据管理	7
■ 数据管理技术	7
■ 数据管理技术的发展	8
第二节 了解数据库技术及应用	9
任务一 浏览网上书店	10
任务二 了解数据库的资源类型和特点	13
学习指引	15
■ 数据库	15
■ 数据库管理数据的基本特征	16
■ 数据库技术的发展	17
第三节 初识关系数据库	19
任务一 了解图书馆数据库中的基本数据	19
任务二 解读实例中的表	22
任务三 了解数据库管理系统的作用	22
任务四 了解数据库应用系统	24
学习指引	26
■ 数据库的数据	26
■ 实体、属性和关系	26
■ 数据库的数据组织	27
■ 表与关系数据库	27
■ 数据库管理系统	28
■ 数据库应用系统	28
■ 数据库系统的组成	28
后续活动	30
评价	31
参考资料	31
第 2 章 建立数据库	35
第一节 设计旅游信息数据库	36
任务一 旅行社旅游业务需求分析	36
任务二 旅游信息数据库的建模	39
任务三 定义旅游信息数据库表及表间联系	41

学习指引	43
■ 数据模型	43
■ 数据库设计的一般步骤	44
第二节 数据库设计的规范化	49
任务一 分析表设计的合理性	49
任务二 优化表	50
学习指引	52
■ 规范化理论概要	52
■ 关系数据库规范化总结	54
第三节 创建旅游信息数据库	55
任务一 选用数据库管理系统	55
任务二 建立空数据库	55
任务三 建立表	56
任务四 建立表间关系	59
任务五 向表中输入数据	61
任务六 导入导出数据	61
学习指引	63
■ 数据库管理系统 Access	63
■ 数据类型	64
■ 字段属性	66
后续活动	67
评价	68
参考资料	68
第 3 章 使用数据库	71
第一节 筛选与查询	72
任务一 使用筛选检索旅游线路	72
任务二 在单一表中查询三日游线路及其行程、交通	74
任务三 通过多个表查询各旅游线路的发团日期和回程日期	76
任务四 使用参数查询线路信息	77
学习指引	79
■ 排序	79
■ 筛选	79
■ 查询	79
第二节 统计与报表	81
任务一 统计各线路旅游团的报名人数和缴费情况	81
任务二 制作各旅游团缴费名单及缴费汇总报表	82
任务三 制作旅游线路介绍报表	88
学习指引	90
■ 数据统计中常用的方法及函数	90
■ 报表的制作	90

第三节 SQL 语言	93
任务一 报名工作分析	94
任务二 用 SELECT 语句查询信息	94
任务三 用 INSERT 语句追加记录	96
任务四 用 UPDATE 语句修改游客报名标志	98
任务五 用 DELETE 语句删除退团游客记录	99
学习指引	100
■ SQL 语言常用语句	100
■ SQL 语言中常用的操作符	102
后续活动	103
评价	104
参考资料	104
第 4 章 建立简单的应用系统	107
第一节 设计旅游信息数据库应用系统	108
任务一 确定旅游信息数据库应用系统的功能	108
任务二 确定旅游信息数据库应用系统的表、查询、报表	110
任务三 设计旅游信息数据库应用系统的窗体	110
学习指引	111
■ 数据库应用系统的设计	111
第二节 建立旅游信息数据库应用系统	113
任务一 准备数据	113
任务二 制作数据查询窗体	114
任务三 制作游客数据维护窗体	115
任务四 建立应用系统的切换面板	118
学习指引	124
■ 窗体的组成	124
■ 窗体的类型	125
■ 窗体的创建	125
后续活动	126
评价	127
附 录 部分名词术语中英文对照	128

泊禾科技出版

第1章

走进数据库

走进现代化的图书馆，只需在计算机上进行简单操作，就能在汗牛充栋的图书中找到想要的图书；来到机票销售网点，只需说出目的地和启程日期，售票员轻轻击打键盘，就能获悉尚有空座的若干航班班次及起飞时间；在超市购物，只需在结账处扫入物品的条形码，瞬间就能获得购物清单……生活中这些情景比比皆是，我们不禁会纳闷：图书馆是如何使用计算机管理图书的？飞机的座位是怎样确保不被重复出售的？超市中物品的名称和价格是从哪里查到的……

在这些场景的背后，支撑着一种先进的数据管理技术——数据库技术，它在现代社会的各个领域中正发挥着重要的作用。

那么，究竟什么是数据库，它在我们的生活中扮演着怎样的角色？它是如何一步步发展起来的？它还将会有一些新的发展？

学习目标

- ★ 了解数据管理的基本知识。
- ★ 知道数据库技术的产生历史和发展趋势。
- ★ 了解数据库在网络和多媒体方面的应用。
- ★ 掌握关系数据库中的库、表、字段、记录等概念。
- ★ 了解关系数据库中数据的组织方式。
- ★ 了解“关系”的含义。

第一节 了解数据管理技术

在我们身边,时时刻刻能看到计算机的身影。学生学习,使用计算机查阅资料;教师教学,使用计算机演示多媒体课件;工程设计人员运用计算机设计产品、绘制图纸;商场用计算机结账;仓库用计算机管理;人们休闲在家也时不时地在计算机的显示屏前欣赏电影……这许许多多的计算机应用得到了一项强有力的技术的支持——数据管理技术。

让我们去调查一下身边的计算机应用,了解数据管理技术是如何与我们的现代生活休戚相关、息息相通的。

任务一 调查图书馆的图书管理技术

提**示****板**

由于条件限制,不同规模的图书馆可能会采取不同的图书管理技术。小组内最好分别选取采用不同管理技术的图书馆作调查。

1. 以小组为单位,分别选取不同规模的图书馆进行调查,并填写调查表。

图书馆调查表

调查组成员 _____

一、图书馆基本情况

1. 我们选择调查的图书馆属于:

☐ 中、小学图书馆 ☐ 高校图书馆 ☐ 县、市图书馆
☐ 省图书馆 ☐ 其他 _____

2. 图书馆馆藏图书数量 _____ 册;读者人数约为 _____ 人。

3. 图书的排放方法:

☐ 简单地按类排放 ☐ 按照索书号排放,位置唯一
☐ 其他 _____

4. 图书馆是否使用计算机管理?

☐ 是,开始使用的时间为 _____ 年 ☐ 否

二、手工管理图书馆的借、还书管理办法调查

(供调查不使用计算机管理的图书馆使用)

1. 读者查找图书的办法:

☐ 直接开架查找
☐ 通过卡片式目录检索出图书的索书号
☐ 其他 _____

2. 图书馆提供哪些卡片式目录供检索:

☐ 书名目录 ☐ 作者目录 ☐ 分类目录

提**示****板**

索书号又称排架号,它指定了该本书在书库相应书架中的具体位置。

提**示****板**

图书馆的目录一般为卡片式。每张目录卡片记载一种图书的相关信息。新书入库时均须编制相应的目录卡片。

下页续

接上页

☐ 其他_____

3. 卡片式目录的每张卡片上记载着图书的哪些信息?

4. 读者借阅图书的办理方法:

☐ 闭架借书——通过卡片式目录查找到图书的索书号,凭索书单和借书证向管理员办手续。

☐ 开架借书——自行到开架书库查找到图书后,凭图书和借书证向管理员办手续。

☐ 其他_____

5. 管理人员如何为读者办理借书手续?

☐ 抽出每本书后所附的书袋卡,填写_____等信息;填写读者的借书卡或借书证(一式两份),具体填写_____等信息。将其中一份借书卡归还读者,另一份借书卡和书袋卡一起放置在馆内。

☐ 仅在借书登记簿上登记书名、读者姓名、借书日期等。

☐ 其他_____

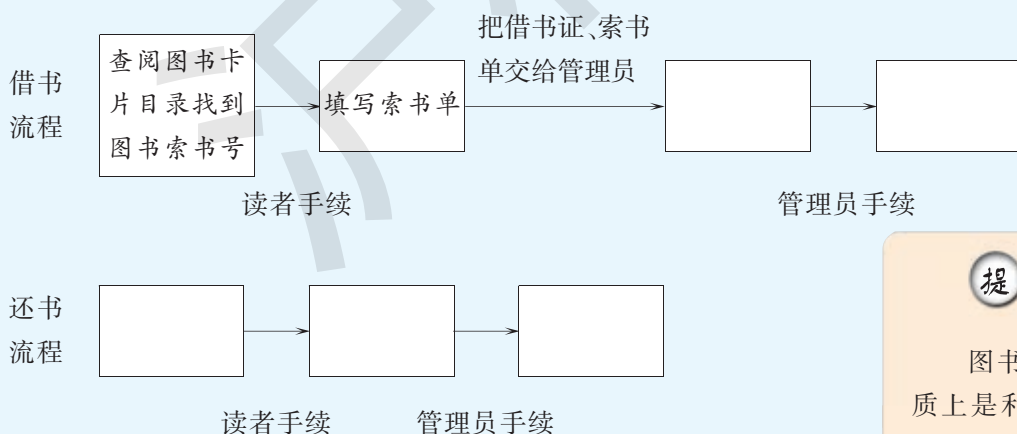
6. 管理人员如何为读者办理还书手续?

☐ 凭读者借书卡上的_____信息找出馆内对应的另一份书袋卡及归还图书的借书卡;在书袋卡上填写_____等信息,将书袋卡放入图书的书袋;填写读者的借书卡,具体填写_____等信息。

☐ 在借书登记簿上找出该书的借出记录,填写还书日期。

☐ 其他_____

7. 画出手工管理图书馆的借、还书程序的流程图。



三、使用计算机管理图书馆的借、还书管理办法调查 (供调查使用计算机管理的图书馆用)

1. 图书馆使用的计算机是:

☐ 单机 ☐ 多台连网的计算机

提

示

板

书袋卡插在图书封底的书袋中,唯一代表着该书。卡上记录该书历次借出及归还信息。借书卡则唯一代表着读者,卡上记录该读者的历次借书情况。

提

示

板

图书馆的计算机管理,本质上是利用计算机技术对图书馆的业务信息(如书目信息、读者信息等)进行管理。我们将这种运用计算机管理数据的技术称为**数据管理技术**。参见学习指引。

下页续

接上页

提

示

板

借书证或图书上的条形码,与读者、图书建立起了一一对应关系。借助于条形码扫描仪扫入电脑,可以调出该读者或图书的信息。

2. 用计算机查询馆藏图书,有哪些检索方法?

- ☐ 按书名检索 ☐ 按作者检索
☐ 按关键词检索
☐ 按书名、作者、关键词等条件综合检索
☐ 其他_____

3. 用计算机能查出图书的哪些书目信息?

- ☐ 书名 ☐ 作者 ☐ 出版信息
☐ 页数 ☐ 价格 ☐ 国际标准图书号
☐ 馆藏信息(在馆,已借出等) ☐ 其他_____

4. 读者除能用计算机查找书目信息外,还能查询哪些信息?

- ☐ 查询图书去向、何时应归还等信息
☐ 查询读者本人历次借阅图书的信息
☐ 查询读者本人过期未还图书的催还信息
☐ 新书查询
☐ 其他_____

5. 管理员如何为读者办理借书手续?

- ☐ 在计算机借书系统中输入读者的借书证号码,调出读者的信息,再输入读者所借图书的书号或书名等。
☐ 用条形码扫描仪扫描借书证上的条形码,在计算机借书系统中调出读者的信息,再逐一扫描所借图书上的条形码,利用条形码将图书信息存放在该读者当日的借书信息中。
☐ 其他_____

6. 管理人员如何为读者办理还书手续?

- ☐ 在计算机还书系统中输入所还图书的书号或书名,调出该书的出借信息,将该书的出借状态改为已归还。
☐ 用条形码扫描仪扫描所还图书上的条形码,调出该书的出借信息,自动将该书的出借状态改为已归还。
☐ 其他_____

7. 画出使用计算机管理图书馆的借、还书程序的流程图。

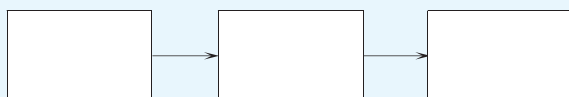
借书
流程



读者手续

管理员手续

还书
流程



读者手续

管理员手续

2. 分析比较手工管理图书馆和应用计算机管理图书馆的管理方式。

管理方法 图书馆	手工管理图书馆	应用计算机管理图书馆
管理内容		
借书卡	传统借书卡 卡上信息：	
检索	卡片式目录 速度：	关键词查询 速度：
索书单填写内容		
查询图书是否在库		
入库图书附加信息	书内附书袋卡 卡上信息：	条形码 码上信息：
查询图书是否外借超期		

3. 交流：在应用计算机管理的图书馆中，哪些方面的工作是由计算机帮助或替代完成的。

任务二 了解哪些部门或行业应用计算机管理

1. 校内调查。以小组为单位，到学校的教导处、总务处、财务室等部门采访有关老师，了解以下信息，并把调查结果填入表中。

提

示

板

用计算机进行管理一般有三种层次的技术，它们被称作人工管理、文件系统管理和数据库管理。参见学习指引。

调查项目 部门	是否应用计算机管理	计算机管理的主要业务	使用时间	使用技术
教导处				
总务处				
财务室				

2. 校外调查。以小组为单位，调查与人们生活有密切关系的部门和行业，哪里应用了计算机进行管理，主要管理什么业务。把调查结果填入表中。

调查项目 调查对象	应用计算机管理领域	计算机管理的主要业务
银行	柜面、ATM机	
超市	收款处(POS机)	
火车售票处		
飞机售票处		



讨论

应用计算机管理的业务有何共同特点?



学习指引

计算机管理已渗透到人们生活的方方面面。在广泛的计算机管理应用中,可以发现,这些应用计算机管理的业务都有着共同的特点。例如,在计算机管理的图书馆中,读者每办理一次借书手续,计算机中就记录了一条该读者借阅图书的信息;用超市的 POS 机结账,每扫一次商品的条形码,就记录了一个商品售出的数据,刷新一次记录;在飞机售票处,旅客能查询详尽的航班班次……这些业务处理的过程都是在计算机中记录、保存和查询数据的过程。我们将这些工作称为数据管理工作。数据管理工作是目前计算机应用最广泛的领域之一。

■ 信息和数据

信息(information)是目前许多学科广泛使用的概念。在不同领域,其含义有所不同。在数据管理领域,一般把信息理解为关于现实世界事物存在方式或运动状态的反映的组合。例如,我们使用的粉笔,它的颜色有白、红、绿等各色,形状呈圆台形,高 80 毫米,上底半径 5



图 1-1

毫米,下底半径 4 毫米,材料是石膏粉,这些都是关于粉笔的信息,都是粉笔存在状态的反映,从不同角度反映或刻画了粉笔这个事物。

数据(data)通常指用符号记录下来的可加以鉴别的信息。例如,为了描述粉笔的信息,可以用一组数据“白色、圆台形、高 80 毫米……”来表述,因为“白色、圆台形、高 80 毫米……”已被赋予了特定的语义,所以它们就有了传递信息的功能。而“白色、圆台形”又被称作文本型数据,“80”被称作数字型数据。

信息和数据之间有着固有的联系:数据是信息的符号表示或称为载体,信息是数据的内涵,是对数据的语义解释。

今天,除了人的思维活动,大量信息处理活动都采用将信息表示为数据,然后对这类符号进行加工这一方式。因此,在大多数场合下,将不再区分信息和数据这两个概念。

■ 数据处理和数据管理

围绕着数据所做的工作,如数据的收集、组织、整理、加工、存储和传播等均称作**数据处理**(data processing)。数据处理有三大类工作——数据传播、数据加工、数据管理。

数据传播,是指在空间或时间上以各种形式传播信息,而不改变数据的结构、性质和内容。数据传播会使更多的人得到并理解信息,从而使信息的作用充分发挥出来。

数据加工,有时也称作信息处理,就是将原始数据进行加工(包括对数据进行交换、抽取和运算)产生新数据的过程。通过数据加工会得到更有用的数据,以指导或控制人的行为或事物的变化趋势。例如,根据上述粉笔的尺寸数据和石膏浇注定额标准可以推算出制作粉笔所需的石膏数量和制作费用。

数据管理(data management)是数据处理的核心和基础,它能在需要的时候为各种应用和数据处理提供数据。数据管理工作包括三项内容:(1)组织和保存数据,即将收集到的数据合理地分类组织,将其存储在物理载体上,使数据能被长期保存。(2)数据维护,即根据需要随时进行插入新数据、修改原数据和删除失效数据的操作。(3)供数据查询和数据统计,以便快速地得到需要的正确数据,以满足各种使用要求。

事实上,生活中涉及数据管理的工作非常普遍。各行各业都有许多人在从事着各种行政管理工作。例如,学校的教导处要管理学生学籍,学校的财务室要管理来往的账务;各企业的人事部门要管理员工的档案,行政部门要管理单位的财产。这些管人、管财、管物或管事的工作就是数据管理工作。习惯上,人、财、物和事统称为事务(transaction)。在事务管理中,事务以数据的形式被记录和保存。例如,在财务管理中,财务部门通过对各种账目的记账、对账或者查账等来实现对财务数据的管理。目前许多数据管理工作开始利用计算机进行,因此,数据管理也成为计算机应用的一个重要分支。

■ 数据管理技术

总结本节广泛开展的调查活动,不难发现,我们始终关心着这样一件事:“有没有应用计算机”,做着“什么工作”。这里的“什么工作”,绝大部分是事务方面的业务,实质上就是数据管理工作。而“有没有应用计算机”关注的则是:在数据管理工作中采用了什么技术和手段。

最早的数据管理采用的是手工管理方式。例如,在不使用计算机的图书馆,就是通过手工登记的方式,记录图书编号、借书证号、图书借还日期等来实现图书管理的;财务部门通过手工记账、算账和保管账的方法实现对财务的管理。计算机的发展为科学地进行数据管理提供了先进的技术和手段。当前,可以毫不夸张地说,几乎所有涉及计算机应用的领域中,数据管理方面的应用担当着相当重要的角色。据统计,计算机的主要应用(如科学计算、数据处理和过程控制等)中,数据处理所占比重约为 70%。现今,“数据管理技术”(data management technology)一词已经泛指应用了计算机的数据管理技术。

应用了计算机,数据管理的速度与规模相对于手工方式有了很大的变化。随着计算机软硬件技术的不断发展和数据处理量的不断增长,数据管理技术本身也在不断地发展着。

■ 数据管理技术的发展

1. 人工管理阶段

20 世纪 50 年代初期,计算机刚刚出现,主要用于科学计算。当时,计算机的外存储器是磁带、卡片和纸带。软件只有汇编语言,没有数据管理方面的软件。人们使用计算机进行数据处理时,程序的编写十分烦琐。数据的编排、存取等全部由程序员完成。每个程序都必须有属于自己的一组数据,一旦需要少许改变数据的存储,就要修改整个程序。计算机也不保存程序和数据。

2. 文件系统阶段

计算机中进行数据处理,首先需把大量的信息以数据形式存放在存储器中,存储器的容量和速率直接关系到数据管理技术的发展。20 世纪 50 年代后期,能存储相当于 10000 张卡片信息的磁盘被开发出来了。磁盘、磁鼓等节约空间和时间、方便可靠的存储设备代替了磁带、卡片和纸带。随着操作系统的诞生,人们在操作系统的支持下,开发了一种专门管理数据的软件,称之为文件系统(file systems)。高级编程语言的出现,使程序员能够容易地编写程序。

文件系统阶段是数据管理技术发展中的一个重要阶段。今天,一些单位和个人仍在使用文件系统进行管理。文件系统管理数据的特点是:数据按某种结构以一条条记录的形式存放在文件中。程序通过文件名和数据打交道,不关心数据在存储器中的存放位置。程序通过操作系统的文件系统对文件进行读或写,从而完成对数据的查询、修改和增删。例如,某学校的总务处这样来管理学生的名单:将学生的信息按照“学号 姓名 性别 年龄 家庭地址”顺序,一条条输入,存放在文件名为“班级”的文本文件中。高一(1)班的学生名单存放在名为“101.TXT”的文件中,打开该文件,有“1 李刚 男 17 鼓楼路 13 号 2 王明 女 18 解放路 3 号 102 室 3 刘洪 男 17 ……”格式的数据。在用 Basic 语言编写的应用程序中,教师只要输入班级就能查询或打印该班的学生名单等。

但是,我们也发现用文件系统管理数据有不少缺点。例如,如果学校想加强和学生家长之间的联系,要在原数据文件中添加家庭电话信息,原数据文件数据顺序将改为“1 李刚 男 17 鼓楼路 13 号 6582178 2 王明 女 18 解放路 3 号 102 室 5826356 3 刘洪 男 17……”但由于数据的编排改变了,整个应用程序将全部不能使用,只能重新编写。另外,若图书馆需要管理学生的借书信息,但图书馆的应用程序无法直接利用总务处这些已有的数据文件,只能重新建立新的数据文件。文件系统管理致使同样的学号、姓名等数据在多个文件中重复存储。不仅如此,由于重复存储,还可能造成数据的不一致。比如,若班主任发现高一(1)班学号为 2 的学生名叫王敏,但入学报名时老师误写成了王明,班主任请总务处的老师在学生名单中作了修改,但班主任忘了通知图书馆,结果王敏拿到的借书证上的名字仍是“王明”。

由于文件系统管理数据存在以上种种缺陷,已经不能满足很多应用的需求,为了提高效率,人们在文件系统的基础上发展了一种新技术——数据库技术。

3. 数据库阶段

20 世纪 60 年代后期,计算机应用于管理的规模越来越大,数据量急剧增长。大容量和快速存取的硬盘也进入市场。此时,以科学地组织和存储数据,高效地获取和处理数据为目的的数据库技术产生了。

数据库技术更加关注数据本身的管理,它将所有数据保存在数据库(database)中,以对

数据实行统一的、集中的、独立的管理。上例中的学校,若采用数据库技术,整个学校只需建立一个统一的学生信息数据库,即将图书馆、总务处、教务处等各部门需要的数据都保存在该数据库中,各个部门通过不同的应用程序访问数据库,使用自己所需的数据。这样,这些部门都需要的有关学生的一些基本数据如学号、姓名、性别等,无需重复存放在多个文件中,减少了数据的重复存储。

数据库技术作为数据管理技术中的一项重要技术自产生开始,一直飞速发展。今天,全世界有成千上万个数据库系统在运行着,数据库技术已广泛应用于工农业生产、商业、行政、科学研究、工程技术和国防军事的各个部门。

4. 大数据时代下的数据管理

21 世纪初期以来,随着互联网和物联网技术的发展,不仅数据量在不断增大,而且数据形式也越来越多样化,海量的结构化、非结构化和半结构化数据不断产生。一方面,传统的关系数据库难以存储海量的数据;另一方面,关系数据库管理非结构化和半结构化数据显得力不从心,而非结构化和半结构化数据所占比例目前已接近 80%。针对非结构化和半结构化数据的管理,提倡运用非关系模型来存储数据的 NoSQL 数据管理技术得到了急速发展。

为了应对数据量和用户服务请求的爆炸性增长,一些互联网公司设计并开发了分布式数据库系统。分布式数据库将数据分散地存储在不同的存储单元上,不仅提升了传统数据库的数据存储能力,而且改善了用户体验。

面对海量的、形式多样的数据,企业往往需要同时部署关系型和非关系型数据库系统,这大大增加了数据管理的成本。为降低数据管理成本,一些云服务提供商设计并开发了云端数据库。云端数据库不仅可以免去企业自建数据库的硬件投入和运营维护成本,而且具有部署快和可扩展性好等优点。



体验活动

组成小组,向自己的亲朋好友或邻居所在单位调查,了解数据管理技术在政府部门、事业单位或企业的应用情况。

第二节 了解数据库技术及应用

浏览网上书店已成为当今许多人的时尚之举。确实,上网购书不乏是一种省时、省力的购书方法。目前,能在网上为读者提供购书服务的网上书店不胜枚举。要在一个网上书店呈现几十万种图书品种,又要为顾客选购提供最便捷的服务,这些功能是怎么实现的呢?数据库技术功不可没。庞大的图书目录、不断增添的上架新书、看不见却登记在册的顾客群、每时每刻可能出现的订单……数据库技术将所有这些都管理得井井有条。

各个网上书店如何为读者展现图书风貌?利用数据库技术为读者提供了哪些服务?让我们去领略一下网上书店的风采,感受数据库技术在其中散发的魅力。

任务一 浏览网上书店

1. 查找国内的网上书店及其网址。

找到的网上书店	网址

2. 访问几家网上书店,了解各家网上书店可供书目数。

网上书店	了解可供书目的方法	可供品种数

提示板

不少网上书店都为读者准备了几十万种图书。这样的图书量,若置于一般书店,需要多大的店堂场地?网上书店的优越性可见一斑。那么,网上书店的图书信息是如何存放与管理的?参见学习指引。

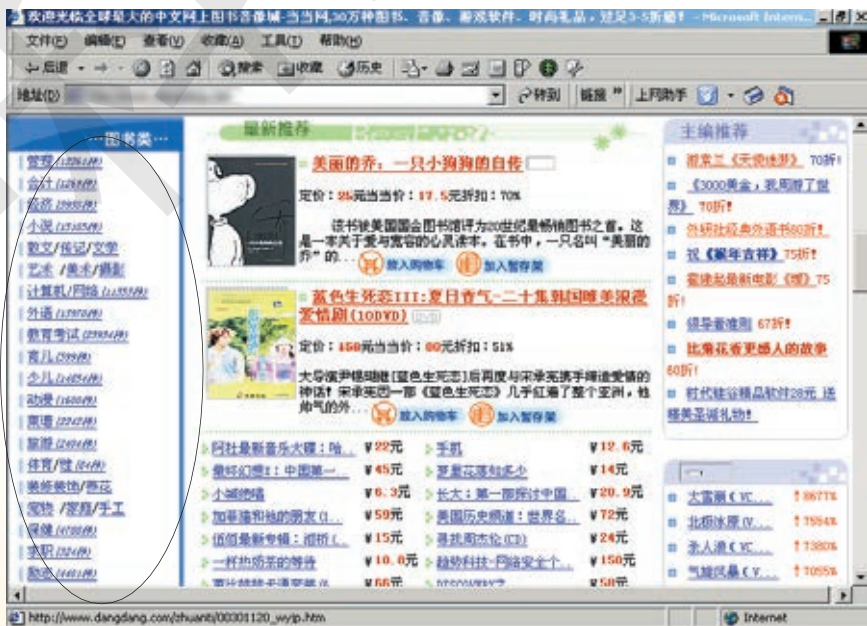


图 1-2

3. 试从网上书店查找一本书,了解各家网上书店为读者提供了哪些查询方法?在查找过程中,还发现哪些了解图书信息的途径?

网上书店	提供的查询方法	提供的了解图书信息的途径



提示板

要快速地对众多的图书信息进行检索，如果没有非常有序的存储方式，是难以想像的。

网页上所显示的这些图书信息都来自后台的数据库。参见学习指引。

图 1-3

4. 试从几家网上书店查找同一本书,了解各网上书店是如何提供图书信息的。比较介绍方法的相同处与差异。

网上书店				
图书信息				
书名				
价格				
封面图片				
浏览次数				

提 示 板

除了书名、内容简介、出版社等常见的文字信息和价格等数字信息外,大多数网上书店还提供图片——图书封面。这些可以存放图形、图像、动画、音频等多媒体数据的数据数据库被称为**多媒体数据库**。参见学习指引。

书名

哈利·波特与火焰杯

浏览次数

该产品已经被浏览6352次

内容介绍

6月1日对孩子们而言无疑是一个重大的日子,而2001年的6月1日,更是非同凡响,因为这一天,孩子们翘首以待多日的儿童故事书《哈利·波特》第四集面市了!哈利注定永远都不可能平平常常,即使拿魔法界的标准来衡量。黑魔的阴影始终挥之不去,种种暗藏杀机的神秘事件将他一步步推向了伏地魔的魔爪。他渴望在百年不遇的三强争霸赛中战胜自我,完成三个惊险艰巨的魔法项目,谁知整个竞赛竟是一个天大的阴谋……

提 示 板

顾客填写信息,提交成功后,网站将在后台数据库中插入这些数据。用户在网上每增加、修改或删除一次自己的订单,网站就需要对数据库中的数据进行一次更新。

封面图片

价格

出版社

出版日期



图 1-4

5. 选择其中一个网站,了解在网站上进行用户注册以及购书的方法。



图 1-5



讨论

在网络上购书有哪些优点和不足？

任务二 了解数据库的资源类型和特点

1. 在网络上查找更多的数据库资源,分析各数据库的类型特点。

找到的数据库及其创建机构	数据库的专业领域和资源类型

提示板

因特网上有大量专供检索信息的数据库。这些数据库大多由信息服务商或大型信息机构创建维护。这类网络数据库已成为人们检索信息的主要工具。参见学习指引。



图 1-6

2. 尝试使用网上数据库资源。
(1) 组成小组,确定有兴趣的专题。

提示板

网络上的数据库内容丰富,涉及各种不同的专业领域。此外,数据库所涉及的资源类型也多种多样,有报刊论文、书目文摘、图片、专利文献、多媒体数据等。参见学习指引。

专题方向提示

教育热点问题(如,素质教育、中学生的校园生活、中学生心理健康教育等);

学科学习资料收集;

动物保护方面;

气象信息方面;

文献检索方面;

.....

(2) 从相关的网上数据库中寻找资料,完成专题介绍报告。报告中注明资料来源。

资料寻找示例一

专题——现阶段的师生关系

① 访问新华网,在网站新闻的全文检索窗口,输入“师生关系”关键词,能查找出该网站上发表的所有有关“师生关系”的新闻全文。

② 进入CNKI中小学数字图书馆上的中学期刊报纸全文数据库,可以了解在期刊、报纸上有哪些已经发表过的有关“师生关系”的文章。通过在标题或全文关键词检索窗口中输入“师生关系”,能检索出来源于国内大部分中学期刊、报纸上的相关文章。由于该数据库必须付费使用,免费用户只能浏览、下载本年度以前的题录、摘要。付费用户则能下载全文。

资料寻找示例二

专题——中国的猴子保护

① 想了解属于中国一级保护动物的猴有哪些,可以访问科学数据库中心网站上中国动物主题数据库中的中国濒危和保护动物数据库。在检索窗口中输入中文名“猴”,可以检索出属国家保护动物的猴的物种列表。点击表中的“蜂猴”,还可查看数据库中有关蜂猴的详细信息,如图片、英文名、国家保护等级、濒危原因等。

② 想了解国内出版了哪些有关“保护猴子”的图书,可以进入中国国家图书馆网站,匿名登陆图书馆的书目数据库。进入书目检索页面,选择中文普通图书库。用关键词“猴AND保护”进行检索,可以查询出中国国家图书馆馆藏中文图书中有关“保护猴子”的图书书目信息。

(3) 交流:

① 专题内容介绍;

② 在专题介绍中,用到了哪些数据库中的资源? 相关数据库提供了哪些信息检索的方法?



具有一定规模的网上书店不但含有几十万种图书的信息,还有众多注册用户的信息和每日不断产生的订单信息……需要管理的数据不计其数。如此庞大的数据必须通过精心地组织和管理才可能为众多的用户有效地使用。网上书店之所以能为众多用户提供方便快捷的服务,不得不提及其后台的有力技术支持——数据库。在当今数据管理技术领域,数据库是最有效的数据管理的技术和手段。

■ 数据库

数据库是一个长期存储在计算机内、有组织的、可共享的、统一管理的数据集合。它是一个按数据结构来存储和管理数据的计算机软件系统。

数据库的概念实际上包含两层意思:

① 数据库这一数据集合是一个实体。它好比是一个能够合理保管数据的“库”,用户在该“库”中存放要管理的事务数据,“数据”和“库”两个概念结合成为“数据库”。

② 数据库是数据管理的新方法和技术,它能够更合理地组织数据、更方便地维护数据、更严密地控制数据和更有效地利用数据。

数据库中的数据具有以下两个特性:

① 数据的整体性。数据库是一个单位或一个应用领域的数据处理系统,它存储的是属于企业和事业部门、团体和个人的有关数据的集合。数据库中的数据是从全局观点出发建立的,它按一定的数据结构组织、描述和存储,其结构基于数据间的自然联系,从而可提供一切必要的存取路径,且数据不再只针对某一应用,而是面向全局或整体,这样,也使数据具有最小的冗余度(重复程度)。

② 数据共享性。由于数据库中的数据是为众多用户共享其信息而建立的,不同的用户可以按各自的用法使用数据库中的数据。多个用户可以同时共享数据库中的数据资源,即不同的用户可以同时访问数据库中的同一个数据。

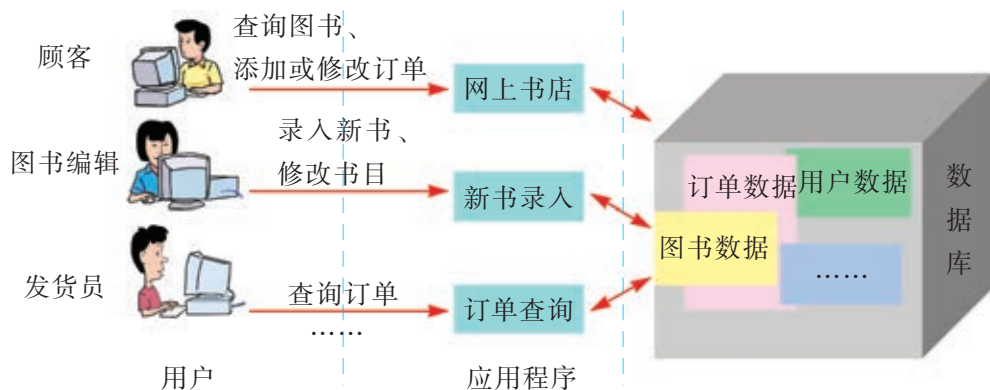


图 1-7

以网上书店为例,其后台数据库存储着大量图书、用户、订单等信息数据。从上网查询图书信息的顾客角度看,这些数据能为众多顾客同时查询图书信息服务,实现了一定程度的数据共享。但网上书店后台数据库的数据共享,绝非仅止于此。该数据还为更多的网上书店的业务所用。通过不同的应用程序,书店发货员可使用数据库查询每日的订单信息,并据此发货;网站的图书编辑及时补充最新上架图书信息以更新数据库……顾客、发货员、图书编辑等不同的用户所存与取的数据,都来自于同一个包含图书、客户、订单等业务数据的数据库。

正因为此,在设计数据库之初,数据库中要存放哪些数据,如何存放,必须考虑整个书店业务特点,考虑面向图书采购员、发货员、图书编辑、业务经理等多方面的需要,从而减少数据的冗余(重复),实现共享。

■ 数据库管理数据的基本特征

1. 数据一致性

由于数据共享,多个用户同时访问同一数据时,数据因相互干扰可能会产生不一致。以民航订票问题为例,若有两个终端(两个售票点)同时操作购买去北京的机票,而此时数据库中仅剩一张去北京的该航班机票。甲终端首先操作并预订了机票,但当该终端正准备修改还未来得及修改数据库之际,乙终端先前一步抢入,并发现数据库内仍有 1 张去北京的机票,因此,也预订了该票,这样就会造成一张机票两人预订的错误结果。为此,数据库采取并发性控制来避免并发程序之间的相互干扰,杜绝提供给用户不正确的数据。此外,如果数据库中仍然存在冗余(某种程度上完全消除冗余是不实际的),同一数据出现在两个地方,那么当其中一个更新时,数据库也保证另一处数据同时更新,亦即数据的一致性(coherency)。

2. 数据独立性

在数据库中,系统提供一种功能,以确保应用程序与数据的结构和存储方式无关,即程序与数据具有较高的独立性(independence)。这样,当为了提高效率或增加新的数据,需要改变数据的结构时,不会影响到应用程序。

3. 数据的完整性与安全性

数据完整性(integrity)是指确保数据库中的数据是正确、有效的。由于数据库中的数据是共享的,若不控制数据的完整性,有可能会因一个用户错误地更新数据库生成错误数据,而殃及其他无辜的用户。因此,在数据库中往往会采取设计一些完整性规则用以确保数据值

的正确。例如,顾客在某网上书店注册时,如果输入不包括“@”的错误的电子邮箱地址,会出现“您输入的 E-mail 地址有误”的窗口,这实际上就是对数据的检验,以保证其完整性。

数据的安全性(security)是指保证数据的安全和机密,防止数据丢失或被窃取。数据库的安全性控制可防止未经允许的用户存取数据。

综上所述,数据库是数据的集合,是为特定组织的多种应用服务的。数据结构独立于使用数据的程序,数据之间相互关联并且冗余度尽可能小。数据库对数据的增添、删除、修改和检索,都有统一、严格的控制,从而确保数据的安全、正确与一致。

■ 数据库技术的发展

作为一种先进的数据管理技术,数据库自产生之日起就在不断地发展。在数据库得到广泛应用的今天,针对不同的应用环境和用户需求,出现了许多新领域、新成果和新技术。

1. 多媒体数据库

网上书店网页上显示的图书信息,不仅有常见的文字数据和数字数据,还包括了文本、图像等形式的数据,这些数据都来自书店后台的数据库。我们把文字、数字形式记录的数据称作传统数据,把图像、视频等形式记录的数据称作多媒体数据。

早期的数据库无法存放图片、声音这种多媒体数据,但现实生活中人们对此又有很迫切的需要。可以想像,如果网页上缺少封面图片,只有书名、作者等枯燥的文字介绍,这样的网上书店对顾客有多少吸引力?随着数据库技术的发展,能存放图形、图像、声音、动画等的多媒体数据库(multimedia database)出现了。将多媒体引入数据库后,数据库变得内容丰富、功能强大而更具实用价值。

多媒体数据库的应用领域相当广泛。例如,贸易系统的电子购物应用中,人们不仅可以从商务信息库中迅速查询到有关商品的文字信息,而且可以同时看到该商品的照片和显示其性能的录像片断;医疗会诊系统中,医生调阅病人的病例,除了可以看到文字记载外,还可以查看有关化验结果以及 X 光照片等;工程设计及建筑设计系统应用中,设计人员可以从各种图形库、图像库、照片库中迅速获得所需要的图符、图像及照片作为设计的模块;教育系统中,从多媒体电子教材库中选择的用于组成现场教学的信息,可以包括文字、图表、图像、视频及声音等各种多媒体信息。

多媒体数据库技术对社会生活、生产的影响越来越大。现在,数字图书馆、数据仓库、科学可视化、远程教育等都体现了多媒体数据库的应用价值。

2. 网络数据库

读者通过因特网访问网上书店,查询图书或提交订单。网站提供的方便、可视化的用户接口,使读者顺利完成了对后台数据库的操作。网上书店可以说是数据库技术和网络技术完美结合的一个典范。

数据库技术目前是计算机处理与存储数据的最有效、最成功的技术,而计算机网络的特点则是资源共享,这种“数据+资源共享”的新技术就是当前广泛应用的网络数据库(web database),也叫 Web 数据库。

网络数据库是以后台数据库为基础,加上一定的前台程序,通过浏览器完成数据存储、查询等操作的系统。简单地说,一个网络数据库就是用户利用浏览器作为输入接口,输入所

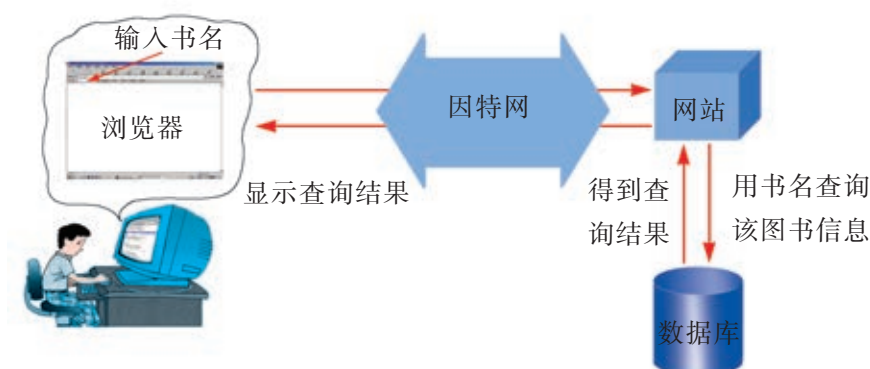


图 1-8

需要的数据,浏览器将这些数据传送给网站,而网站再对这些数据进行处理,如将数据存入数据库,或者对数据库进行查询操作等,最后网站将操作结果传回给浏览器,通过浏览器将结果告知用户。

除网上书店这类大量用于企业网站、记载公司的商业活动的网络数据库外,网络上还有一类应用很广泛的专供人们检索信息的数据库,如存储大量原始信息全文的期刊全文数据库、专利全文数据库、百科全书全文数据库等,又如存储目录、题录、文摘等书目线索的各种图书馆目录数据库、文摘数据库等,以及存储生产与经营活动信息的机构名录数据库、存储人物信息的人物传记数据库、存储基金信息的基金数据库等。这类网络数据库一部分是商业性网络数据库,只有付费获得授权才能使用,但也有相当数量的免费数据库,只要连入因特网就能使用。国内目前比较著名的网络数据库有《中国科学引文索引》、《中国社会科学引文索引》、《全国报刊索引数据库》、国图数字图书馆等。

3. 其他

数据仓库(data warehouse, DW)是当前数据库发展的一个重要方面。目前的数据库应用多是联机事务处理(online transaction processing, OLTP)。联机事务处理主要是指对数据的操作型处理,也即普通的查询和修改等日常操作,如在网上书店查书、购书,在 ATM 机上取款、查余额等。在联机事务处理中,用户对数据的存取操作频繁,数据库要强调的是数据更新处理性能和系统的可靠性,注重响应时间、数据的安全性和完整性。数据库经长期应用,往往会积累一大批历史数据。对这些数据进行分析,以给决策人员提供决策支持,就是联机分析处理(online analytical processing, OLAP)。银行对顾客信用的评估等就是联机分析处理一例。传统数据库系统擅长于企业的日常事务处理工作,而难于实现对数据的分析处理。针对这一问题,研究和解决从数据库中获取信息的数据仓库开始发展。数据仓库是把不同数据库,不同存储介质,不同业务系统上的数据按某一应用主题集成起来的数据集合。

当前同样蓬勃崛起的有数据挖掘技术。数据挖掘(data mining, DM)是指从大量数据中发现有用的信息和知识。数据挖掘采用了统计学、决策树等方法分析数据内在关联,预测某些趋势和行为。

数据挖掘在企业经营中十分有用。它根据企业目标,对大量的数据进行挖掘,发现内在蕴含的规律,预测企业的发展趋势,获取企业的经营信息,从而使决策者能基于信息丰富的

数据作出重要的判断和决策,而非仅依赖于直觉。数据挖掘这一沙里淘金的技术的采用,有可能使那些在计算机应用中长期积累起来的看似无用的历史数据,成为企业的财富。

数据处理技术一直在随着实际应用的需要不断地发展,一代代的数据库产品也在不断地改变着我们的生活。由于篇幅所限,我们无法一一罗列讲解。未来的数据库世界将往何处发展,我们也仅仅只能预测而不能断定,但是,有一点是毋庸置疑的,那就是任何一项技术的发展都是与实际应用的需求紧密联系在一起,数据库技术也不能例外。



体验活动

1. 以小组为单位,确定一个研究专题,如不同地区青少年体质和营养现状,城乡同一年龄段人群的体质与营养状况比较……通过上网或其他途径查找相关资料,写出专题报告。

要求:

- (1) 至少查找到两个以上的数据库,并能充分利用数据库提供的数据。
- (2) 比较各数据库对完成该研究的作用。

2. 登陆新华社多媒体数据库,了解多媒体数据库的应用。

要求:

- (1) 确定一个新闻主题,检索出相关的图片、视频新闻。
- (2) 通过上网或其他途径,了解其他的多媒体数据库,分析其应用方法和应用价值。

第三节 初识关系数据库

随着数据库技术不断成熟,数据库在各行各业的管理中也发挥着越来越重要的作用。那么数据库是一个怎样的“库”?数据库中的数据又是如何组织的?数据间存在什么关系?人们又是如何从数据库中获取数据的?本节将从分析图书馆数据库入手,通过解读数据库中的数据及其关系,了解数据库应用系统和数据库管理系统在实际应用中的作用,探寻数据库技术是如何为读者和管理员提供各种服务的。

任务一 了解图书馆数据库中的基本数据

1. 组成小组,分别前往应用了数据库技术的图书馆,采访有关人员,了解他们在借、还书过程中最关心什么事情;了解他们使用的数据库中有哪些基本数据。

调查记录单

_____ 学校
调查者: _____;
调查的图书馆是 _____ (学校、县、大学、省) 图书馆;
采访的人员有 _____ (读者、管理员、业务主管、馆长、其他); 他们各自最关心的事情是 _____, _____, _____;
图书馆的工作人员每天的主要工作是 _____;
该图书馆的数据库中有哪些基本数据? (调查结果请另附表)

2. 交流各小组调查获得的结果和图书馆数据库中的基本数据表。



讨论

各图书馆中, 最了解该数据库基本数据的老师具有什么身份?

实例:

以下是某市第三中学的同学调查了本校图书馆的记录。

调查记录单

某市第三中学

调查者: 张蕙蕴、周娇娇、杨晴青

我们调查的是本校图书馆;

我们采访了在图书馆借书的学生和图书馆的老师;

借书的学生关心自己想借的书和借了几本书;

图书馆老师关心这几方面的事: 师生员工、书和借还图书情况;

图书馆里有三位老师。三位老师都告诉我们, 我校图书馆使用了数据库技术管理图书, 其中有一位顾老师专门负责录入新购图书的有关数据。他较详细地告诉了我们图书馆数据库的基本数据内容(见附表)。

从调查中得知, 图书馆计算机一旦发生故障, 由我校计算机中心的老师负责检修和维护。

提

示

板

如果用数据库的术语来描述, 那么图书馆老师关心的“这几方面的事”称为**实体**, 即图书馆管理员关心的实体是: 读者、书、借还情况。参见学习指引。

调查记录单附单

图书馆专门录入新书数据的顾老师告诉我们,我校图书馆数据库的基本数据是这样三个表:

读者表

借书证号	姓名	性别	所在班级或部门	电子邮箱	读者身份
0001	顾莉	女	教导处	lgu@163.net	教师
0002	李晓明	男	高二(1)	xml@163.net	学生
0003	王芳	女	高一(2)	fwang@163.net	学生
0004	张晓凯	男	高二(2)	xkzh@163.net	学生

书目表

分类号	书名	出版社	第一作者	出版日期	价格
wx1011	骆驼祥子	人民文学出版社	老舍	2004-1	¥22.00
wx1018	二马	人民文学出版社	老舍	2001-1	¥14.00
wx1023	四世同堂	人民文学出版社	老舍	2001-1	¥49.00
wx1045	茶馆	人民文学出版社	老舍	2003-5	¥6.00
wx2006	红楼梦	人民文学出版社	曹雪芹	2001-1	¥39.00
wx2016	三国演义	人民文学出版社	罗贯中	2001-1	¥28.00

借还信息表

借书编号	分类号	借书证号	出借日期	还书标志
00001	wx1011	0001	2003-1-4	<input checked="" type="checkbox"/>
00002	wx2016	0004	2003-1-6	<input checked="" type="checkbox"/>
00003	wx1018	0003	2003-1-5	<input checked="" type="checkbox"/>
00004	wx1023	0001	2003-1-4	<input checked="" type="checkbox"/>
00005	wx1023	0003	2003-10-30	<input type="checkbox"/>
00006	wx1045	0004	2003-9-28	
00007	wx1045	0002	2003-12-4	
00008	wx2006	0001	2003-2-4	
00009	wx2016	0002	2003-10-4	

因为每个表中按规律记了很多信息,顾老师让我们在每个表中摘了几行下来。

顾老师还告诉我们,图书馆管理员老师不仅关心我校有多少师生员工会来借书,关心图书馆有多少藏书,关心每本书的借还情况;还关心哪个读者借了哪些书,出借的书被谁借去了等信息。

提 示 板

管理员顾老师告诉学生的三个基本数据表,分别是三个实体(读者、书目、借还信息)的具体信息。

提 示 板

管理员顾老师告诉学生还关心的这些信息,是实体和实体之间关系的信息。参见学习指引。

提

示

板

一张表(table)是一个实体的数据的集合。三个实体的具体信息都以表的形式表示。

表的每一行称为一个**记录**(record);记录中的每一项称为一个**数据项**(item)。每一列称为一个**字段**(field);表头中的每一项称为**字段名**。

任务二 解读实例中的表

1. 分析:从实例中的三张表,分别可以获得什么信息?

归纳出这三张表的表名、字段名和记录数,并从表中找出可唯一找到读者、书和借还信息的字段名。

从读者表中可以获得的信息:_____;

从书目表中可以获得的信息:_____;

从借还信息表中可以获得的信息:_____。

提

示

板

表中能唯一确定某个记录的字段称为**主关键字**。

表名	字段名	记录数	主关键字

提

示

板

某位读者借了哪几本书?书名是什么?需要从三张表中寻得。全校有几千名师生,为每个读者都查找一次,工作就相当繁杂。使用**数据库管理系统**这样的工具就可以帮助管理员迅速完成这一工作。参见学习指引。

2. 试从表中查找信息:“王芳同学借了哪几本书,它们的书名是什么?”并记录寻找过程。

从_____表中查到_____;

再从_____表中查到_____;

最后从_____表中查到_____。

3. 交流,在查找信息过程中从一张表找到另一张表是通过哪些字段找到联系的。请举例说明。

提

示

板

能够建立起两表间的联系的字段称为**关联关键字**。

任务三 了解数据库管理系统的作用

1. 启动数据库管理系统 Microsoft Access,打开配套光盘上的数据库文件“图书馆.mdb”。

(1) 在“对象”栏中选择“表”,查看数据库中建立了几张表,各表中有哪些数据。

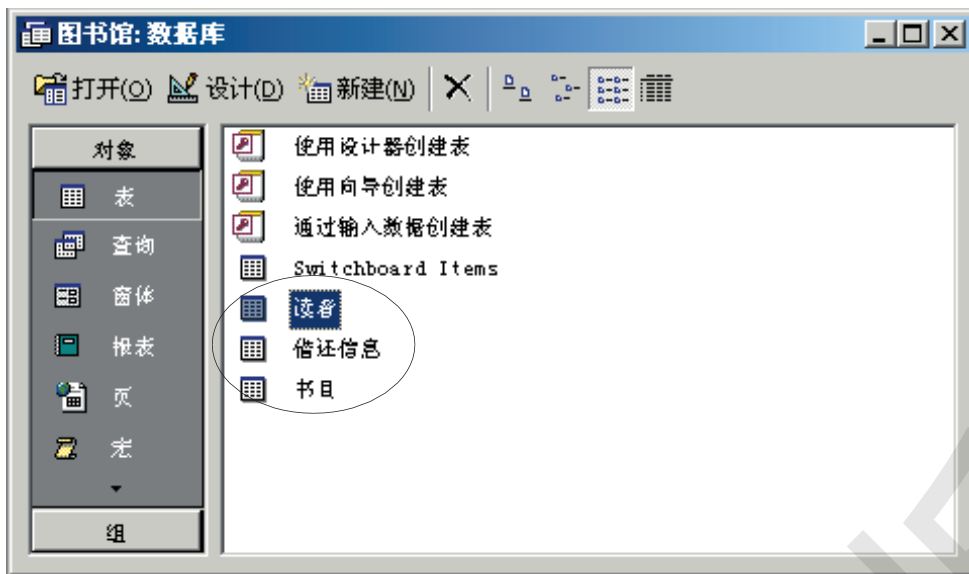


图 1-9

(2) 在“对象”栏中选择“查询”,了解图书馆数据库管理系统提供了哪些查询功能。

(3) 在“对象”栏中选择“报表”,了解图书馆数据库管理系统能生成哪些用户报表。

2. 打开图书馆数据库的“关系”窗口,查看窗口中的关联线所指字段是否是关联关键字。

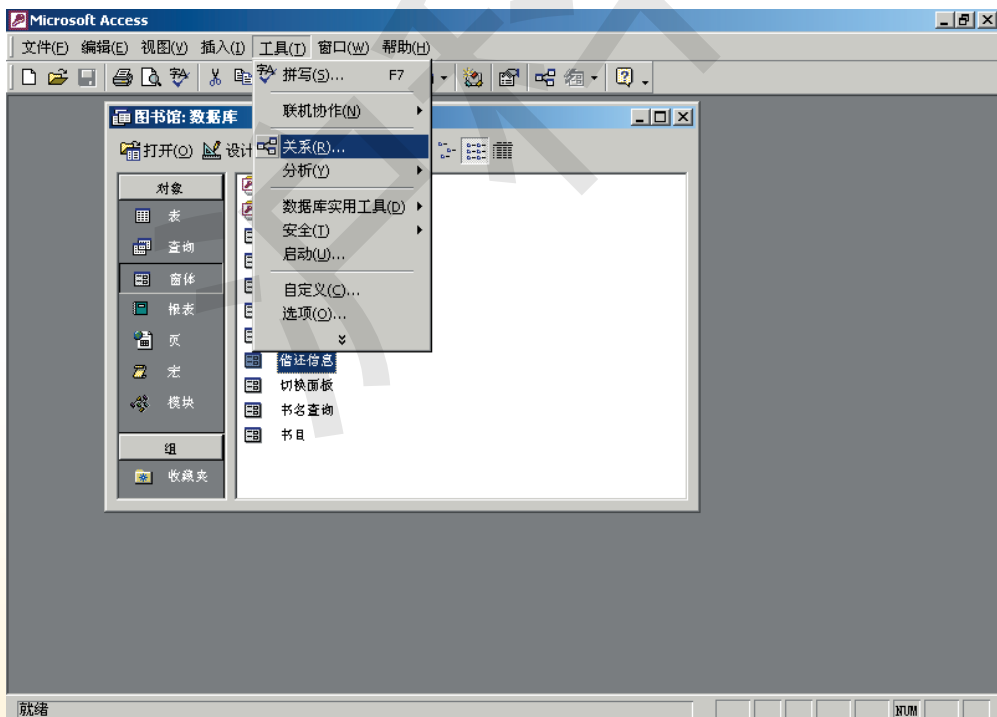


图 1-10



讨论

- (1) 在图书馆数据库管理系统中有哪些数据？这些数据分别对应哪个实体？
- (2) 在图书馆数据库管理系统中，表之间的关系是否反映了实体间的关系？

提**示****板**

用数据库管理系统 Access 提供的工具,可以为用户制作进入图书馆数据库应用系统的主界面。由于用户的权利不同,为不同的用户设置的权限也不同。本图书馆数据库应用系统通过登录窗口的设置让不同的使用者进入与其相适配的应用系统界面。参见学习指引。

任务四 了解数据库应用系统

1. 启动数据库管理系统 Microsoft Access, 打开配套光盘上的“图书馆应用系统.mde”。
2. 在“登录”窗口的用户和密码中均键入“admin”, 进入“图书馆应用系统.管理员”界面。



图 1-11

提**示****板**

该页面是为管理员用户制作的图书馆管理系统主界面。管理员用户使用数据库, 首先进入到此页面。



图 1-12

3. 尝试使用应用系统页面提示的各功能。
 - (1) 单击“数据维护”选项, 进入数据维护界面。根据所需维护的数据库中的数据, 选择相应的选项。

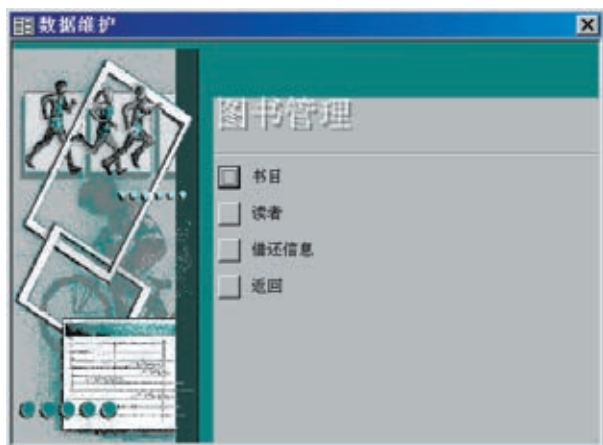


图 1-13

(2) 试用“数据查询”选项,了解图书馆数据库管理系统为用户提供了几种查询图书的方法。

(3) 尝试打印“读者借还明细”和“书目列表”。

4. 在“登录”窗口的用户和密码中均键入“student”,可进入“图书馆应用系统.学生”界面。

尝试使用该应用系统页面上提示的各功能。

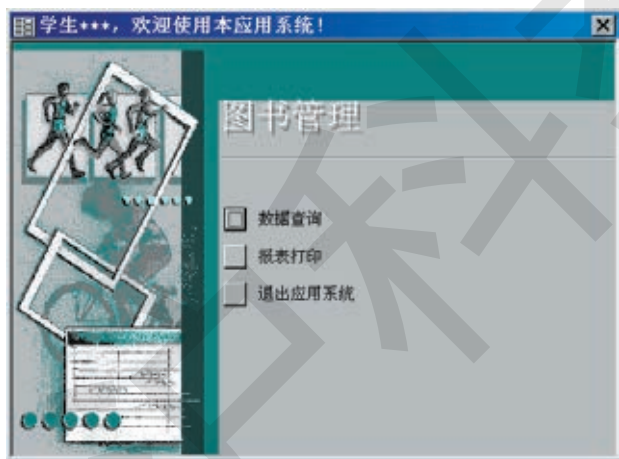


图 1-14

提 示 板

界面中列出的是可供维护的数据表选项。

提 示 板

读者用户在图书馆的计算机上查询图书,能使用的是为读者设置的与其权限适配的应用系统及其界面。参见学习指引。



讨 论

(1) 从数据库管理系统操作窗口进入数据库,与从图书管理员使用的应用系统界面进入数据库,两者所实现的各种功能有何差异?

(2) 供读者查询的应用系统与供图书管理员使用的应用系统,所实现的各种功能有何差异?

(3) 数据库管理系统和数据库应用系统各有何作用?



学习指引

■ 数据库的数据

数据库既然是基于计算机的“数据”的“库”,那么任何“库”都要遵循一定的存取规则。在建造库时就要考虑到:一是提供存入“数据”的便利;二是管理的方便;三是“数据”提取的快捷。因此,这里所说的数据库,就不是可以随意创建的“库”,而是需要遵循一定的规则,是在保证“数据”存取的合理性、有效性的基础上创建完成的。

在这样的前提下,组织数据库中的数据,既要考虑用户有什么基本要求,又要知道有哪些组织数据的基本方式可供选用。

在讨论数据库中的数据按什么方式组织存储的问题时,需要提到几个最基本的术语和概念——实体、属性和关系。

■ 实体、属性和关系

实体(entity),指客观存在并可相互区分的事物或对象。实体可用来区分不同的数据。实体可表述为一个名词,既可以是一个具体的对象,也可以是一个抽象的概念,如一个人、地点、事件或事物。例如,前面的实例中图书馆管理员所关心的书目、读者(师生员工)、借还信息就是三个实体;学校教导处关心的是学生、课程和任课教师,它们也都是实体;汽车经销商关心的实体就是汽车、客户及生产商。从数据的角度讲,实体是代表一种或一类数据的业务对象。

属性(attribute),指实体的特性。例如,对于人这个实体,可列举的属性是眼睛的颜色、身高等;前面图书管理员关心的实体读者,其属性包括借书证号、姓名、性别、所在班级或部门、电子邮箱、读者身份等。从数据的角度讲,属性用来组织实体内的特殊数据。

图 1-15 表示的是两个实体(读者和借还信息),以及它们的一些属性。读者实体有六个属性:借书证号、姓名、性别、所在班级或部门以及电子邮箱、读者身份。这些属性正是实例中基本数据表中的列,即字段名。借还信息包括的六种属性是:借书编号、分类号、借书证号、出借日期、借期、还书标志。

读者 ← 实体

借书证号	姓名	性别	所在班级或部门	电子邮箱	读者身份

↑ ↑ ↑ ↑ ↑ ↑ 属性

借还信息 ← 实体

借书编号	分类号	借书证号	出借日期	借期	还书标志

↑ ↑ ↑ ↑ ↑ ↑ 属性

图 1-15

关系(relation)指实体之间的联系。例如,读者和借还信息之间就存在着联系,一个读者和他有否借还过(书)相联系,书的借还情况与借这本书的读者相联系。规范地说,一个读者与他有否借还过(书)相关,书是否被借还与读者相关。

在实例中,读者和借还信息两个实体之间的关系称为一对多的关系,这是由于一个读者可以多次借还(书),而一本书只能借给一个读者或从一个读者处还回来。换句话说,一个读者和多本书的借还信息相关,一本书的借还信息只和一个读者相关。

实体和实体之间还有一对一和多对多的关系。这里不一一列举(详见第二章第一节)。

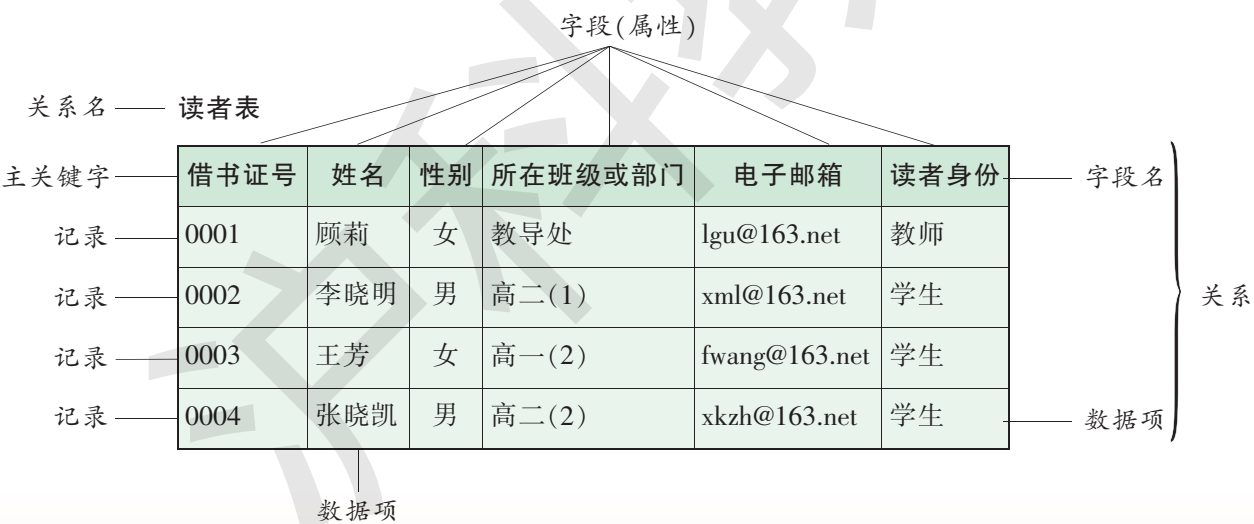
■ 数据库的数据组织

数据库可以存储多个实体的信息。数据库还存有各种实体之间的关系。例如,前面的图书馆数据库,不仅存放读者、书目的信息,还存放读者和他借还了多少书的关系的信息。规范地说,数据库是一种存放多种实体、实体属性以及实体间关系的信息的结构。

■ 表与关系数据库

在前面的图书馆数据库中,管理员呈现给学生的基本数据是二维表形式的三张基本数据表。实际上,已经反映了关系数据库的基本数据。关系数据库(relational database)的实质就是类似前面的这些表的集合。那么,为什么不称这样的数据库为表库等名称呢?这是因为,规范地说这些表就称为关系。关系数据库的名称由此而来。

以读者表为例:



其中,每一行即每一条记录反映的是一个读者其借书证号、姓名、性别、所在部门或班级、电子邮箱、身份的信息数据;每一列反映的是读者某一方面的特征,如每个读者的借书证号、每个读者的姓名等信息数据。这些数据间的“关系”是并列的,或称平行的,没有从属关系。所有数据集合在一起,列成了上图所示的一张二维表,构成了该图书馆所有读者的关系。

同样,书目表、借还信息表也构成了相应的关系。

在关系数据库中怎么处理读者可以借或还多本书,及每本书只能借给一个读者或从一个读者处还来的一对多的关系?观察读者表和借还信息表,可以知道,实现的方法是,在读者表和借还信息表中设置相同的列——借书证号,通过该列可以查到读者借还了多少书。

至此,可以给出关系数据库的详细描述。

定义:关系是一个二维的表,具有以下性质:

- 表中每一数据项不可再分,是最基本的数据单位。
- 表中每一列都有一个唯一的字段名。
- 表中同一列的数据项要具有相同的数据类型(或称属性值),并与字段名匹配。
- 表中各列的顺序可以任意地排列。
- 表中不允许有完全相同的记录行。
- 表中各行的顺序可以任意地排列。

定义:关系数据库就是一组关系的集合。

在数据库领域中,关系数据库是一种经常采用的基本数据库形式。

■ 数据库管理系统

数据库的管理涉及建立、使用和维护数据库等工作。用户当然可以根据自己的需要编制特定的程序加以完成。但这是一项非常复杂的任务。数据库管理系统是辅助管理数据库的最有效的工具。

数据库管理系统(database management system,DBMS)是一个或一组程序的集合,也称作软件包,通过它,用户可以和数据库交互,实际操作则由数据库管理系统完成。例如,在前面的图书馆数据库系统中,用户要查找某一出版社的书,通过数据库管理系统,系统可以找出数据库中所有这家出版社的书。

数据库管理系统建立在操作系统的基础上,对数据库进行统一的管理和控制。数据库管理系统可以进一步被定义为是可用来管理数据库并与数据库相互作用的工具。

数据库管理系统的目标是让用户能够更方便、更有效、更可靠地建立数据库和使用数据库中的信息资源。数据库管理系统不是应用软件,它不能直接用于诸如工资管理、人事管理或资源管理等事务管理工作,但数据库管理系统能够为事务管理提供技术。也就是说,数据库管理系统是为设计数据管理应用项目而提供的计算机软件,利用数据库管理系统设计事务管理系统可以达到事半功倍的效果。

目前,常见数据库管理系统产品有 Access、SQL Server、Oracle、MySQL 等。

■ 数据库应用系统

数据库的开发设计人员,根据数据库应用的需要,在操作系统和数据库管理系统的支持下,开发设计出来的应用程序称为数据库应用软件。应用软件一般由多个程序模块构成,为用户对同一数据库实现多种甚至复杂的应用。数据库应用软件统称为**数据库应用系统**(database application system,DBAS)。

在本节的数据库技术在图书馆管理中的应用中,其应用系统为管理员和读者两类用户提供了有不同权限的应用。

■ 数据库系统的组成

数据库系统(database system,DBS)是指使用数据库技术统一管理、操纵和维护数据资源的计算机系统。一般由计算机的硬件、软件、数据和人员四部分组成。

硬件 硬件环境是数据库系统的物理支撑,它包括相当速率的CPU、足够大的内存空间、足够大的外存设备及配套的输入、输出设备等。

软件 软件包括系统软件和应用软件两类。

系统软件有数据库管理系统和支持数据库管理系统的操作系统、开发应用系统的高级语言及其编译系统以及应用系统开发的工具软件等。

应用软件是指在操作系统和数据库管理系统的支持下，由用户根据实际需要自行开发的应用程序。

数据 数据是数据库的基本内容,数据库又是数据库系统的管理对象,因此,数据是数据库系统必不可少的数据源。

人员 数据库系统人员是指数据库系统管理人员、开发设计人员和用户。不同的人员对数据库系统的使用权限和操作目的是不同的。

数据库系统管理员全面负责管理和控制数据库系统,确定系统软硬件配置,给数据库系统开发设计人员提供最佳软硬件环境。

数据库系统开发设计人员负责设计应用系统的程序模块，在数据库系统软件支持下编写程序。

用户通过应用系统提供的用户接口界面使用数据库系统，这些接口为用户提供了简明直观的数据表示和方便快捷的操作方法。

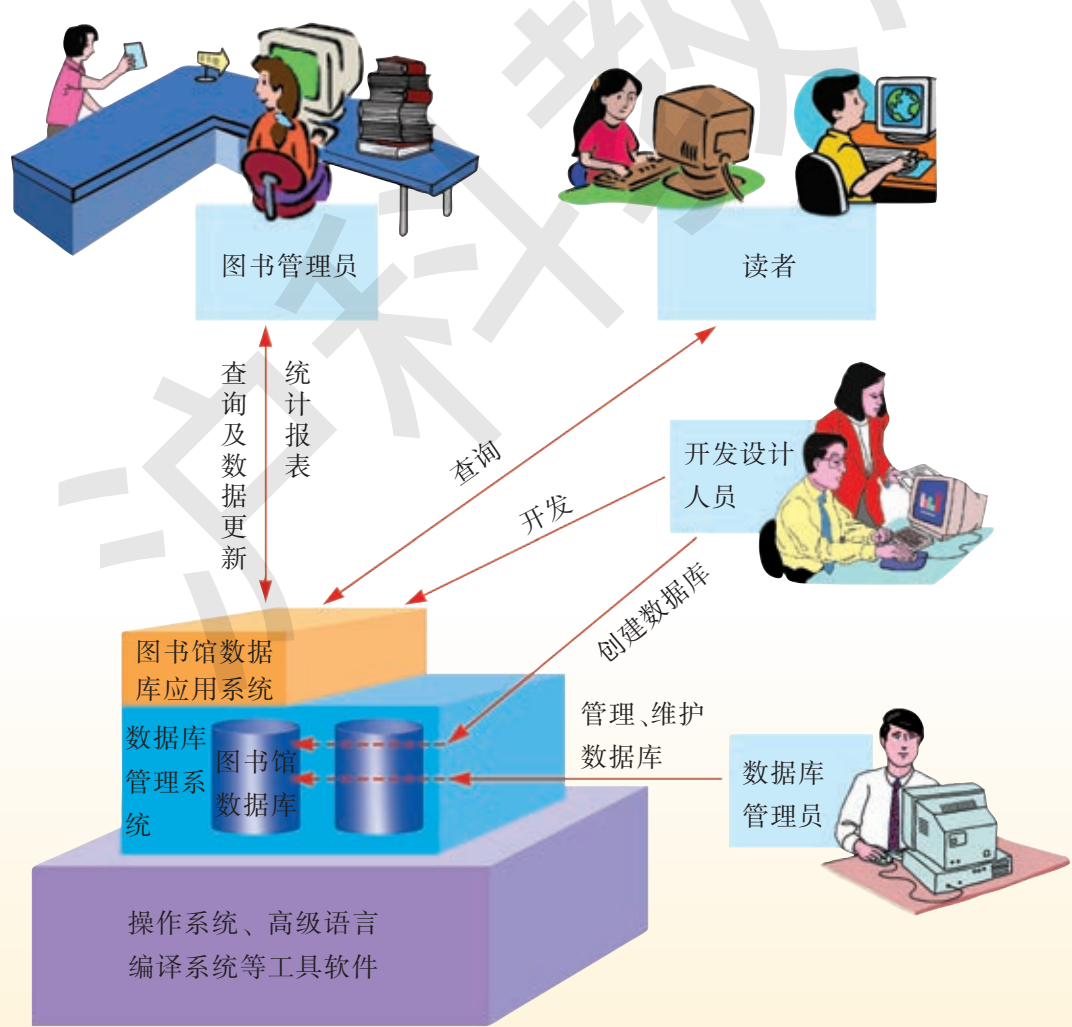


图 1-16



体验活动

1. 找出与下列问题有关的实体。

(1) 学校教导处负责课程的安排,即每学期开学,都要给每个班级排一张课程表。试分析,教导处老师排课程表时,关心哪些方面的事,写出这些实体。(建议向教导处老师作一些了解。)

教导处老师关心的实体有:_____。

(2) 每个人都有自己的资料信息,如个人的藏书、个人的友人通信录……试对自己的资料信息作一分类,分析这些资料信息涉及哪些实体。

个人资料信息实体有:_____。

2. 找出下列问题的实体,并分析其属性。

(1) 学校的教职工是学校这一机构最基本的人员组成。若要对教职工做一份档案名录,试分析:(建议向学校人事科作一些了解。)

实体是:_____;

反映该实体的特征是:_____ (即属性)。

(2) 每个学生,从小学到高中的每个学习阶段,都有一个属于自己的班级。试分析,反映班级的特征有哪些。

实体是:_____;

反映该实体的特征是:_____ (即属性)。



后续活动

1. 试设计一份学校学生的名录。

要求:

- (1) 指出实体与属性;
- (2) 归纳成数据表;
- (3) 指出表、字段、记录及数据项;
- (4) 分析该数据表中“关系”的含义。

2. 调查旅行社的业务需要。

组成小组,前往当地的旅行社,了解以下信息:

游客前往旅行社,关心哪些信息:_____;

旅行社向游客介绍哪些信息:_____;

旅行社组团关心哪些信息:_____;

游客报名参加旅游团需办哪些手续:_____。

评价

一级指标	评价内容	评价(★★★、★★、★)		
		自评	互评	教师评
信息技术知识	知道数据管理、数据管理技术等基本概念,了解数据管理技术的历史及三个发展阶段。			
	初步了解数据库、数据库管理系统、数据库应用系统及其相互关系。			
	了解数据库技术的发展趋势及其在网络和多媒体方面的应用。			
	了解实体、属性、关系的含义。			
	知道关系数据库中的库、表、记录、字段等概念。			
信息技术能力	能根据学习需要找到相关数据库,学会使用数据库,并能充分利用现有的数据库资源,完成相关学习任务。			
	能在关系数据库实例中分析出表、字段、记录等。			
行为	有联系实际生活的意识,积极关注数据管理技术在社会各方面的应用,能关注该技术的发展状况。			
	能积极参与调查活动,在活动中能发现问题,主动学习。			
交流体会:(在活动中,分析实体和实体的属性时,遇到过困难吗?怎么解决的?取得了哪些经验?得到过什么启发……)				

参考资料

数据资源与数据价值

随着时代的发展,数据已俨然成为人类社会赖以生存和发展的一项重要资源,它对国家和民族的发展、对人们的工作和生活至关重要,广泛存在于经济、社会的各个领域和部门。公司管理、商业决策、科学研究、政府政策制定,都离不开对数据资源的利用。例如,与医疗卫生和生命健康活动相关的健康医疗数据就是宝贵的数据资源。对健康医疗数据的分析与挖掘在医学临床、分子生物学、预防医学、医院管理等领域都发挥着重要作用。对各卫生医疗机构采集的患者就诊数据进行深入挖掘后,医生能优化治疗过程,精准用药,减轻患者在治疗过程中的痛苦;科研人员能研发出更有针对性的药物;医院能优化内部管理,改善患者就医体验;政府相关部门能更好地监管医疗体系。

在加强安全保障和隐私保护的前提下,越来越多的政府部门和公司将数据资源开放共享,这使得更多的组织可以利用这些数据资源,充分挖掘其价值。同时,各行各业也在积极推动行业内及不同行业间的数据资源整合,加强数据资源的发掘运用。

数据作为信息社会的重要资源,它的价值来源于数据本身、技术和思维三个层面。数据本身是数据价值的起点,只有拥有数据或能够接触到数据才能开启数据的价值。数据的拥有者需要借助于各种技术,特

别是数据管理与分析技术,获取数据中隐含的信息,在具体的业务中体现数据的价值。数据思维就是提出数据的创新性用途,挖掘数据的新价值。有些看似毫不相关却非常重要的数据需要依靠人类的智慧不断分析,通过数据思维创新性地实现数据的价值。

数据仓库

数据仓库是一个面向主题的(subject oriented)、集成的(integrate)、相对稳定的(non-volatile)、反映历史变化(time variant)的数据集合,用于支持管理决策。对于数据仓库的概念,可以从两个层次予以理解,首先,数据仓库用于支持决策,面向分析型数据处理,它不同于企业现有的操作型数据库;其次,数据仓库是对多个异构的数据源有效集成,集成后按照主题进行重组,并包含历史数据,而且存放在数据仓库中的数据一般不再修改。

■ 数据仓库的特点

根据数据仓库概念的含义,数据仓库拥有以下四个特点:

- 面向主题。操作型数据库的数据组织面向事务处理任务,各个业务系统之间各自分离,而数据仓库中的数据是按照一定的主题域进行组织。主题是一个抽象的概念,是指用户使用数据仓库进行决策时所关心的重点方面,一个主题通常与多个操作型信息系统相关。

- 集成的。面向事务处理的操作型数据库通常与某些特定的应用相关,数据库之间相互独立,并且往往是异构的。而数据仓库中的数据是在对原有分散的数据库数据抽取、清理的基础上经过系统加工、汇总和整理得到的,必须消除源数据中的不一致性,以保证数据仓库内的信息是关于整个企业的一致全局信息。

- 相对稳定的。操作型数据库中的数据通常实时更新,数据根据需要及时发生变化。数据仓库的数据主要供企业决策分析之用,所涉及的数据操作主要是数据查询,一旦某个数据进入数据仓库以后,一般情况下将被长期保留,也就是数据仓库中一般有大量的查询操作,但修改和删除操作很少,通常只需要定期的加载、刷新。

- 反映历史变化。操作型数据库主要关心当前某一个时间段内的数据,而数据仓库中的数据通常包含历史信息,系统记录了企业从过去某一时点(如开始应用数据仓库的时点)到目前的各个阶段的信息,通过这些信息,可以对企业的发展历程和未来趋势做出定量分析和预测。

企业数据仓库的建设,是以现有企业业务系统和大量业务数据的积累为基础。数据仓库不是静态的概念,只有把信息及时交给需要这些信息的使用者,供他们做出改善其业务经营的决策,信息才能发挥作用,信息才有意义。而把信息加以整理归纳和重组,并及时提供给相应的管理决策人员,是数据仓库的根本任务。因此,从产业界的角度看,数据仓库建设是一个工程,是一个过程。

■ 数据仓库系统的体系结构

整个数据仓库系统是一个包含四个层次的体系结构,具体由图 1-17 所示。

- 数据源:是数据仓库系统的基础,是整个系统的数据源泉。通常包括企业内部信息和外部信息。内部信息包括存放于关系数据库管理系统(relational database management system,RDBMS)中的各种业务处理数据和各类文档数据。外部信息包括各类法律法规、市场信息和竞争对手的信息等等。

- 数据的存储与管理:是整个数据仓库系统的核心。数据仓库的真正关键是数据的存储和管理。数据仓库的组织管理方式决定了它有别于传统数据库,同时也决定了其对外部数据的表现形式。要决定采用什么产品和技术来建立数据仓库的核心,则需要从数据仓库的技术特点着手分析。针对现有各业务系统的数据,进行抽取、清理,并有效集成,按照主题进行组织。数据仓库按照数据的覆盖范围可以分为企业级数据仓库和部门级数据仓库(通常称为数据集市)。

- OLAP 服务器:对分析需要的数据进行有效集成,按多维模型予以组织,以便进行多角度、多层次的分析,并发现趋势。

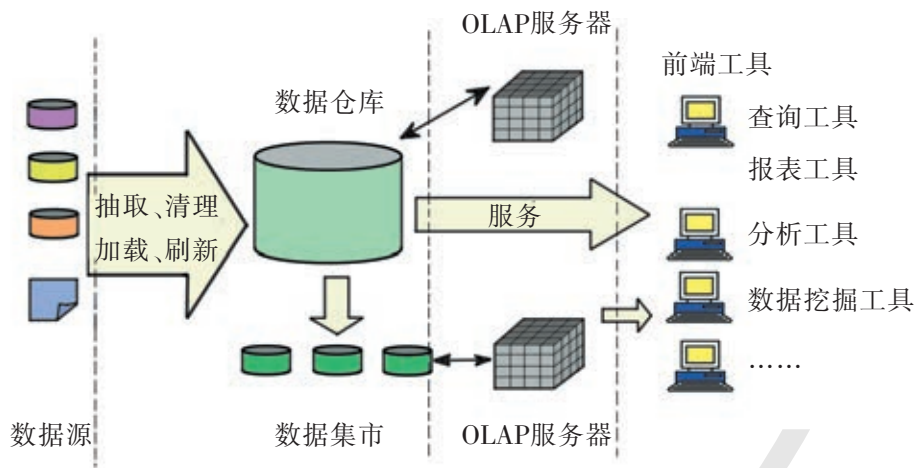


图 1-17

● 前端工具:主要包括各种报表工具、查询工具、数据分析工具、数据挖掘工具以及各种基于数据仓库或数据集市的应用开发工具。其中数据分析工具主要针对 OLAP 服务器,报表工具、数据挖掘工具主要针对数据仓库。

数据挖掘

数据挖掘,也可以称为数据库中的知识发现(knowledge discovery in database, KDD),是从大量数据中提取出可信、新颖、有效并能被人理解的信息的高级处理过程。数据分析本身已经有很多年的历史,只不过以往的数据收集和分析的目的仅限于科学研究。另外,由于当时计算能力的限制,以大量数据为对象进行分析的复杂数据分析方法受到很大限制。现在,由于各行业业务处理自动化的实现,商业领域产生了大量的业务数据,这些数据不再是为了分析的目的而收集的,而是由于纯机会的商业运作而产生的。分析这些数据也不再是单纯为了研究的需要,更主要的是为商业决策提供真正有价值的信息。但所有企业面临的一个共同问题是:企业数据量非常大,而其中真正有价值的信息却很少,因此从大量的数据中经过深层分析,获得有利于商业运作、提高竞争力的信息,就像从矿石中淘金一样,数据挖掘也由此而得名。因此,数据挖掘的商业应用可以描述为:按企业既定业务目标,对大量的企业数据进行探索和分析,揭示隐藏的、未知的或验证已知的规律性,并进一步将其模型化的先进有效的方法。

嵌入式数据库

嵌入式数据库也称移动式数据库。随着移动通信技术的进步和人们对移动数据处理和管理需求的不断提高,与各种智能设备紧密结合的嵌入式移动数据库技术已经得到了学术界、工业界、军事领域、民用部门等各方面的重视。

嵌入式移动数据库的应用可划分为水平应用和垂直应用。水平应用是指应用方案能够用于多种不同行业,只需要极少的定制工作;而垂直应用则指针对特定行业的应用,数据处理具有独特性。

1. 水平应用

水平应用可用于不同的领域,它的应用核心不需要修改,只需对应用的前端或后端进行适当的定制。

● 数据库信息存取:移动用户通过前端嵌入式数据库应用工具,直接向网络数据库服务器提交查询,将检索到的结果缓存或复制到嵌入式数据库中,进行本地管理。这些前端工具可能进行一定的定制,后端数据库服务器也可能作一些修改。

● 场地内或场地间的移动应用:应用中的移动用户在某个或某几个场地内移动,同时保持与基地服务器的联系。

● 基于 GPS 和 GLS 的应用:这类应用通过地球同步通信卫星(GPS 类)传送地图信息或位置信息,或

者通过发射器的信号广播(GLS类)来发送位置信息,各种位置信息、环境信息以及其他的辅助资料可以保留在嵌入式数据库中。

● 现场审计和检查:移动用户是具有一定审计、检查、监督等权利的检查人员,在处理过程中要连接到受检查者的信息数据库,并进行必要的更新,同时更新被检查者的嵌入式数据库。

2. 垂直应用

与水平应用相比,垂直型应用具有明显的行业特殊性,不同领域的应用之间差别较大,结果表示和服务器数据库的管理也有很大不同。目前主要涉及保险、银行业、航班信息、政府部门等领域。

● 金融行业的应用:主要涉及保险业、银行业、股票交易等。如保险业业务员将客户在多个账户中的信息进行汇总,并在必要时给出某种形式的报告。基于掌上电脑或其他移动设备上的嵌入式数据库所建立的移动应用能够很好地满足应用的需求。

● 零售业和分销行业应用:手工操作或固定的POS销售,已经发展为无线网络中基于嵌入式数据库的移动电子存单管理和无线POS系统。另外,支持无线Modem的移动自动售货机可以支持信用卡支付,以无线通信方式实时进行注册、验证,完成交易处理。

● 法律和公共安全:移动用户的可移动性在案犯追捕中具有明显的优点。警务人员的移动设备的嵌入式数据库中保留一定的案犯信息,可以随时检索疑犯信息。

此外,还有其他一些专门的移动应用,如航空、铁路、急救医疗服务等行业,它们都要求提供方便、快捷的服务。而自动交通税收、自动仪表信息收集和电子地图等应用更具有明显的行业特殊性。

大 数 据

信息技术与经济社会的交汇融合引发了数据量的迅猛增长,数据已成为国家基础性战略资源。大数据正日益对全球生产、流通、分配、消费活动,乃至经济运行机制、社会生活方式和国家治理能力产生重要影响。

大数据是指无法在可承受的时间范围内用常规软件工具进行捕捉、管理和处理的数据集合。大数据通常具有4V特征,也就是Volume(数据量)、Velocity(处理速度)、Variety(多样性)、Veracity(真实性)。

● 数据量:大数据的体量很大,且数据集合的规模还在不断扩大。随着信息技术的大规模普及和应用,教育、商业、工业、科学研究、医疗等各行各业所产生的数据量都呈现出指数增长的趋势。

● 处理速度:由于数据量增长速度快,大数据处理速度也必须快,且时效性要求高。大数据往往以数据流的形式动态地、快速地产生,需要在一定的时间限度下得到及时处理。

● 多样性:大数据来自多种数据源,数据种类和格式非常丰富。随着智能设备、社交网络等的流行,机器和传感器数据(如设备日志、地理位置数据)、社交数据(如网站用户行为记录数据等)等各种新类型数据越来越多。

● 真实性:大数据的真实性主要包括数据的可信性、真伪性、来源和信誉、有效性等。

第2章

建立数据库

旅游是现代社会生活中的一项重要活动。许多游客是随旅行社组团出游的。每天,尤其是“黄金周”期间,全国各旅行社组织着各种线路的旅游团,奔赴天南海北。看似平常的行路、观景、吃饭、睡觉,对于各个旅行社来说,却有大量繁杂的业务工作要做。旅行社需安排好各团队游客的吃、住、交通工具;安排好旅游行程;组织好各团队的导游;经费账务……这一切都来不得半点差错。

数据库的出现,为包括旅游在内的各行业找到了一条科学管理、高效经营的道路。为旅行社建设一个旅游信息数据库应用系统,可以结束旅行社那低效的手工登记、繁复的重复计算统计、拍脑袋计划的时代。

数据库应用系统的核心是数据库,建立数据库应用系统的第一步就是建立数据库。那么,如何根据旅行社现实情况来设计旅游信息数据库呢?如何保证所设计的数据库符合要求,适合业务需要?如何根据设计方案来建立数据库……

本章将学习如何按照应用要求,设计并建立起一个数据库。

学 习 目 标

- ★ 初步掌握数据收集、数据分类和建立关系数据模型的基本方法。
- ★ 学会使用实体—关系图描述关系数据模型。
- ★ 理解层次和网状数据模型的基本概念。
- ★ 了解规范化理论的基本概念及其重要作用。
- ★ 能够使用 Access 数据库管理系统,创建数据库。

第一节 设计旅游信息数据库

旅游信息数据库是为旅行社更有效地管理旅游业务设计的。那么,现实生活中旅行社有哪些业务要处理?旅行社需要应用数据库做什么?适合这些需要的应用会涉及哪些数据?这些数据用什么方式组织存放,便于保存、处理、输出……

让我们从了解旅行社的具体业务开始,学着设计旅游信息数据库。

任务一 旅行社旅游业务需求分析

1. 交流整理第1章后续活动2的调查获得的旅行社旅游业务信息。

游客关心的信息	
旅行社向游客介绍的信息	
旅行社组团时关心的信息	
游客报名参加旅游团需办的手续	

提

示

板

熟悉了解现实生活中旅行社在旅游业务方面有哪些具体的业务工作,是设计旅游信息数据库的前期准备。

数据库设计者只有熟悉相关的业务,才有可能使所设计的数据库完整、确切。

调查记录示例一

调查者:张蕙蕴、周娇娇、杨晴青

旅行社:快乐旅行社

我们了解到的旅行社旅游业务信息如下:

(1) 游客关心的信息是:

有哪些旅游线路可供选择,行程安排,住宿状况,线路价格情况,交通工具,用餐情况;另外游客还关心旅游团出发日期,旅游天数,导游素质等。

(2) 旅行社向游客介绍的信息是:

可供选择的旅游线路及日程安排,组团方式,欲前往的线路已有多少人报名,旅游团出发时间,导游是谁;另外还有注意事项等。

(3) 旅行社组团时关心的是:

各条线路已报名人数,能组成多少团,各团出发日期及时间,回程时间,导游安排,游客是否有特殊需要。

(4) 游客报名参加旅游团需办的手续是：
登记姓名,性别,证件号,联系方式,家庭住址,参加哪条线路,签订旅行合同,付费。

2. 分析旅行社的旅游业务需求。

(1) 调查旅行社的旅游业务需求。

调查提纲：

旅行社有哪些旅游业务工作？

_____；

旅行社有哪些工作希望由计算机完成？

_____；

旅行社数据库要保存、处理、输出哪些信息？

_____。

提示板

数据库是现实世界中的客观实体及其相互关系在计算机的数据世界中的反映。因此要建立一个数据库,首先要了解现实世界。这个过程在数据库应用系统的设计中称之为需求分析。

调查记录示例二

学生张某等前往快乐旅行社了解到的业务需求如下：

● 快乐旅行社在旅游业务方面有这些业务工作：

- 1. 接受游客咨询；
- 2. 接受游客报名并签订合同；
- 3. 组织安排旅游团。

● 快乐旅行社需要计算机做的事：

- 1. 旅游信息查询：为游客提供旅游线路、旅游目的地、行程、费用等信息；
- 2. 游客信息登记：记录游客的姓名、年龄、身份证号、参团要求等信息；
- 3. 旅游合同管理：保存与游客签定的合同的有关价格、旅游线路等信息。

提示板

业务需求的调研中,在了解到有哪些业务需要哪些数据信息的同时,往往还会获知用户对应用数据库的需求。“希望计算机完成的工作”即为应用需求。该功能如何实现,将在本书第4章中解决。

调查记录示例三

张某等同学同时带回了如下一些原始文档资料,供分析数据用。

The documents shown are:

- 海南 HAINAN** brochure: Promoting Hainan Island travel.
- 桂林 GUILIN** brochure: Promoting Guilin travel.
- 旅行社游客报名表** (Travel Agency Visitor Registration Form): A form for collecting traveler details. It includes fields for:
 - 姓名 (Name), 性别 (Gender), 年龄 (Age), 工作单位或家庭地址 (Work or Home Address), 联系电话 (Contact Number), 供应标准 (Supply Standard), 备注 (Remarks).
 - 旅行日期 (Travel Date) with fields for year, month, and day.
 - 旅行地点 (Travel Location).
 - 报名日期 (Registration Date).
 - 每人收费 (Fee per person), 其他费用 (Other fees), 共 (Total), 大 (Large), 小 (Small), 合计 (Total), 发票号码 (Invoice Number), 经手人 (Handler).

提示板

需求分析就是收集、归纳和分析现实世界的具体业务需求,由此归纳出需要数据库管理的数据信息。

(2) 分析整理调查中得到的信息。

以快乐旅行社的业务需求及调查获得的信息资料为基础,归纳希望计算机提供的(数据)信息。

希望计算机管理的业务	需要计算机提供的(数据)信息
线路信息查询时	
游客登记时	
旅游合同中	

提示板

业务数据源自于业务需求调查中记录的信息。

数据分类一是依据业务需求调查时记录的客观存在的可区别的业务对象;二是原已分门别类地保存的原始文档资料。

(3) 参考收集到的旅游业务信息资料,将相关业务需要涉及的数据信息按如下方式分类:

与线路有关数据	线路名、行程、交通、用餐、住宿、价格
与游客有关数据	
与团队有关数据	发团日期、回程日期、属何线路、计划人数
与合同有关数据	

3. 定义旅游业务中的实体及其属性。

根据对快乐旅行社数据信息分类整理的结果,完成如下各实体及属性表:

实体	属性
导游	工号,姓名,性别,联系电话,护照号码,身份证号码,外语 1,外语 2
旅游线路	
旅游团	
合同	

任务二 旅游信息数据库的建模

前面已经完成了对现实世界中的旅行社业务需求的调查和分析,现在将学着建立一个能正确反映旅行社现实的数据库的数据模型。

模型是对现实世界的模拟和抽象。现实世界中我们经常接触到模型,如购买房屋,可以看到房产商展示的房屋设计模型;市政改造,可看到相应的规划模型;学生的科技活动可看到车模和航模……

数据模型也是一种模型,它是对现实世界问题的数据特征的描述。

实体—关系图(entity-relation diagrams, ERD)是一种比较常用的描述数据库概念模型的方法。实体—关系图又称 E-R 图,主要用于描述现实世界中实体及其相互之间的关系。

1. 用 E-R 图描述旅行社旅游业务需求中的实体及其属性。

(1) 根据需求调查中对游客实体的了解,用 E-R 图描述游客实体及其属性。

提示板

根据第 1 章的实体及其属性的定义,在旅游业务中,游客、线路、团队、合同分别代表了一类数据的业务对象,因此都可以被定义为实体。前面的数据分类分析中未提及“导游”,但从旅行社的业务和资料中均能发现,“导游”也是一个实体。

属性用于描述实体的某些特定性质。如旅游线路实体可用线路的名称、景点、价格等属性进行描述;导游实体可由导游的工号、姓名、性别等属性进行描述。属性由名称和属性值组成,如“导游工号”是属性的名称,而“001”是属性的具体值。

提示板

数据库设计要经历从现实世界到信息世界到数据世界的逐步抽象过程。


数据库的建模就是先建立一个能正确反映现实世界的信息世界的模型。把现实世界中的信息及其相互关系用简洁明了的方法描述出来,使人们对组成数据库的信息构架有一个概念上的了解。参见学习指引。


提示板


按照不同的应用层次,可将数据模型划分为概念数据模型(简称概念模型)和逻辑数据模型。

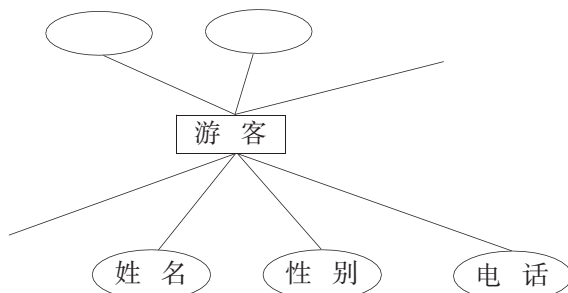
E-R图方法是用少量的几种图元素来构划数据库的概念模型,描述组成信息世界的实体、属性及其相互关系。

E-R图主要的三种图元素的表示方法如下:

实体,用带有实体名的矩形框表示: 

属性,用带有属性名的椭圆形框表示: 

实体与属性及实体间联系,用直线连接: 

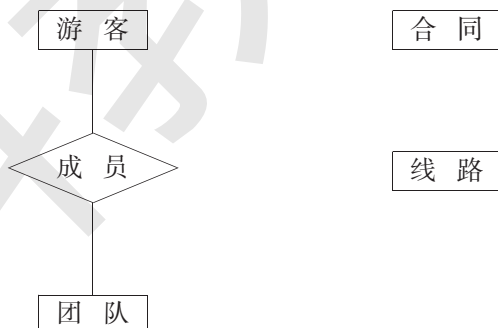


(2) 用 E-R 图方法描述其他实体及其属性。


2. 用 E-R 图描述实体间的关系。

(1) 在各实体间,找出有联系的实体。

(2) 用 E-R 图建立实体间联系,如游客与团队、线路与团队、游客与合同。



提示板

实体通过联系相互关联。实体间的联系可用带有联系名的菱形框表示: 

图中游客与团队两实体间可因游客是团队的成员而建立联系,联系名是成员。

(3) 在 E-R 图中标出实体间联系的类型。

● 分析快乐旅行社旅游业务需求中,实体之间联系的类型:

① 每条线路可以组织多个旅游团,而一个旅游团只能对应一条线路。线路与旅游团之间为一对_____的联系。实体联系名为_____ (成员、关联……)。

② 一位游客可能签订多份不同的旅游合同,而一份合同只能与一位(组)游客签订。游客与合同之间为一对_____的联系。实体联系名为_____ (签约、成员……)。

③ 一个旅游团有多位游客,而一位游客也可以在不同的时间参加多个团队。游客与旅游团之间为_____对_____的联系。实体联系名为_____ (负责、参加……)。

④ 一个旅游团有一名导游,而一名导游可以先后带多个旅游团。导游与旅游团之间为一对_____的联系。实体联系名为_____ (带领、属于……)。

● 标注实体间的联系类型。



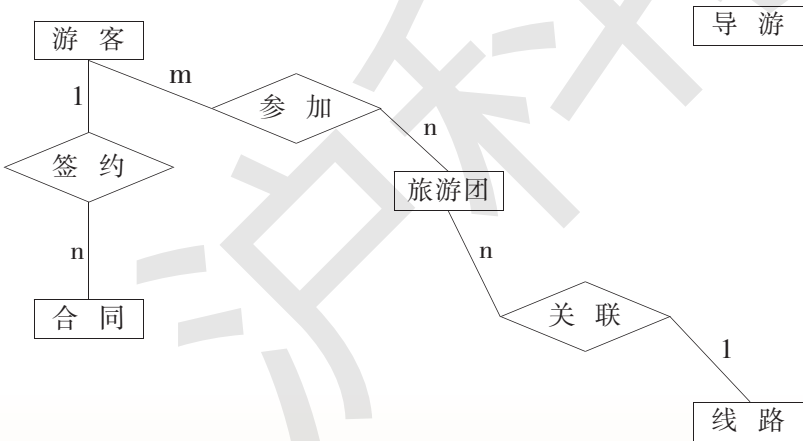
提示板

两个实体之间的联系有三种类型:一对一、一对多和多对多,它们分别反映了现实世界中实体间的联系。如:班级与班长是一对一关系,一个班长只能任一个班的班长,一个班也只能有一个班长;班级与同学是一对多联系,同学与课程则是多对多联系。

在E-R图中,这种关系分别被标记为:1:1、1:n和m:n。

3. 用E-R图,描述旅游信息数据库整体模型。

根据对快乐旅行社的业务分析,完成数据库的整体模型中实体间联系的描述。



提示板

在表示联系的直线靠近实体的那端标上1或n。用来表示联系类型。

任务三 定义旅游信息数据库表及表间联系

前面,用实体、实体的属性、实体间的关系描述了现实世界。在E-R图描述的基础上,可以建立计算机数据世界中的数据模型——关系数据模型。

1. 将E-R图中的实体及属性转换为表。

以快乐旅行社业务需求分析为基础,将实体及其属性转换成关系数据库中的表。

提示板

数据世界中的**数据模型**是基于具体计算机系统和数据库管理系统的数据模型,它直接面向数据库的逻辑结构。就关系数据库模型而言,就是要定义表及表间联系等。

提

示

板

关系数据库是由相互关联的二维表(简称“表”)组成的。表是数据世界中实体的体现,因此表与实体对应,字段与属性对应。表名对应实体名,字段名对应实体的属性名。

提

示

板

当原有的表的字段中,找不到可唯一标识该表的记录的字段时,如游客表中“姓名”不能唯一地标识游客,在数据库设计中,往往采取另外添加新的主关键字段来唯一地标识该表的记录。线路编号、游客编号、团队编号等都是这种情况下添加的字段。

主关键字也可以是字段的组合。参见学习指引。

导游

工号	姓名	性别					

旅游线路

				线路名称			

旅游团

	线路编号	发团日期	导游工号				

游客

合同

合同号		签约日期		旅游团编号			

2. 定义主关键字。

关系数据库中的每一张表都应该有一个主关键字。

为已经转换好的导游表、线路表、旅游团表、游客表、合同表定义一个主关键字,并在被定义为主关键字的属性(字段)下划线。



讨论

- (1) 在游客表中,游客的姓名有可能相同,从哪些字段可以判断他们不是同一人?
- (2) 用证件编号来标识游客是否方便?
- (3) 是否需要添加字段来标识特定的游客?

3. 定义表间联系。

(1) 导游表与旅游团表之间为一对多的关系,通过导游表中的_____与旅游团表中的“导游工号”相关联。

(2) 线路表与旅游团表之间为一对多的关系,通过线路表中的“线路编号”与旅游团表中的_____相关联。

(3) 游客表与合同表之间为一对多的关系,通过游客表中的“游客编号”与合同表中的_____相关联。

(4) 游客表与旅游团表之间为多对多的关系,通过合同表间接的相关联。合同表充当媒介,它与游客表、旅游团表之间都是一对多的关系。合同表与旅游团表的一对多关系通过合同表中的“旅游团编号”与旅游团表中的_____相关联。

提示板

关系数据库由多张表组成。各表之间的联系反映了现实世界中实体之间的关系。表与表之间的关系类型也有三种:一对一、一对多和多对多。一对多的关系最普遍。多对多的关系需要使用第三张表来间接关联,第三张表与前两张表之间都是一对多的关系。

相关联的数据表之间通过具有相同属性的字段建立关联关系。该字段称为**关联关键字段**。参见学习指引。

学习指引

在旅游信息数据库开发设计的活动中,经历了数据库设计的过程,体验了按照旅行社现实世界的需求建立相应的信息世界中的概念数据模型,再建立忠实反映需求的数据世界的逻辑数据模型即关系数据模型,形成了关系数据库的数据结构:一组表结构及表间联系。

■ 数据模型

数据库是一个单位或组织需要管理的全部相关数据的集合。计算机不可能直接处理现实世界中的具体事物,所以必须使用相应的工具,先将具体事物转换成计算机能够处理的数据,再由计算机处理。**数据模型**(data model)就是这样的工具。

模型是对现实世界的抽象。在数据库技术中,用模型的概念描述数据库的结构和语义,对现实世界进行抽象。数据模型是现实世界数据特征的抽象,形象地说,它类似仓库中的货架,不同的物品按各自的存放方式有序地放置在各自的货架上,便于管理。数据按照数据模型所设定的框架,按规定的顺序结构合理地存放。不同的数据模型对数据的存放方式各不相同。

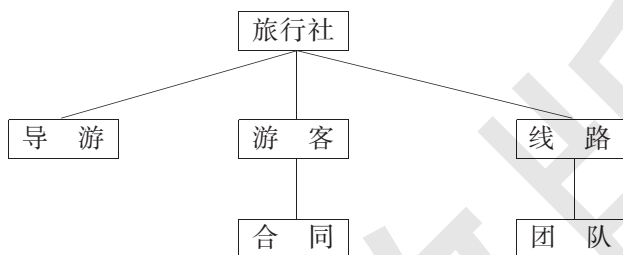
将现实世界中的具体事物及事物之间的联系抽象到数据世界,组织为数据库系统支持的数据模型,通常的方法是:首先把现实世界中的客观对象在人们头脑中的反映通过模型表示出来,这是人脑经过选择、命名、分类等综合分析形成的印象和概念,所以称这种表示模型为概念模型。然后再把概念模型描述的实体及其联系转换为具体数据库管理系统支持的数据模型。在“设计旅游信息数据库”的活动中,用 E-R 图描述的是旅游信息数据库的概念模型,并

被转换成数据库管理系统能够直接处理的关系数据模型。

关系数据模型 (relational data model) 是目前应用最广泛的一种数据模型, 目前大多数 DBMS 产品都基于关系数据模型。关系数据模型由若干关系组成, 每个关系都类似一张表, 用二维表的形式来描述实体、属性以及实体之间的联系。

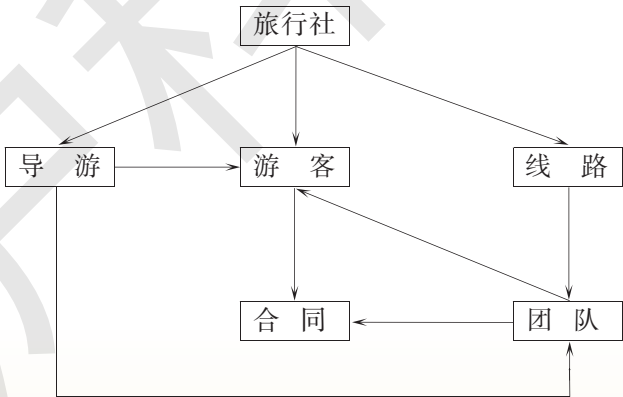
除了关系数据模型, 其他的数据模型还有: 层次模型和网状模型。

层次数据模型 (hierarchical data model) 使用树形 (形状像一棵倒置的树) 结构来表示实体及其联系, 树的结点代表实体, 上一层实体与下一层实体之间的联系是 1:n, 用结点之间的连线表示。在现实世界中, 许多实体之间的联系呈现出一种自然层次关系, 如家族关系和行政机构关系等。旅行社旅游信息可通过层次数据模型表示如下:



层次数据模型示例

网状数据模型 (network data model) 结点之间的联系不受层次的限制, 可以任意发生联系, 更适合描述复杂的事物及其联系。网状模型中, 用结点来表示实体。结点之间的联系通过“系”实现。系用有向线段表示, 箭头指向 1:n 联系的“n”方。旅行社旅游信息可通过网状数据模型表示如下:



网状数据模型示例

■ 数据库设计的一般步骤

数据库的设计过程有如下几个阶段:

1. 需求分析

需求分析 (requirement analysis) 就是对现实世界进行调查与分析。需求分析一般是对整

个数据库应用系统所要处理的对象进行全面的了解,不只针对数据库的需求,但是数据库的需求分析是其中的重要内容。

需求分析的主要任务是:调查业务的现状与信息需要;收集原始业务信息资料;整理数据需求;分析并归纳对应用系统的功能要求和对数据的处理、存储、输入与输出的要求。

需求分析要解决的两大主要问题:一是确定系统功能,也就是前面说的要系统“做什么”的问题;二是信息的调查分析,其成果是建立一个数据字典。它是应用系统中各类数据描述的集合,归纳了数据的名称、属性、用途等等。

2. 概念设计(建立概念模型)

概念设计(conceptual design)就是建立概念模型。数据库设计者须对原始数据进行综合,抽象出数据库系统所要研究的数据,将现实世界中的事物及其联系,转换成信息世界中的实体及其联系。概念模型是对现实世界的第一层抽象。概念模型独立于计算机系统,它既能够充分反映现实世界,适用于与不熟悉计算机的用户交流,同时又易于向关系模型、网状模型、层次模型等各种数据模型转换。

E-R图是描述概念模型最常用的工具,也是最有效的交流工具。

建立概念模型的一般过程是:

(1) 确定实体和属性。确定实体和属性是构建 E-R 图的最重要的工作。现实世界中一组具有某些共同特性和行为的对象就可以抽象为一个实体。例如在学校环境中,可以把张三、李四、王五等学生对象抽象为学生实体。对象的成分或特性可以抽象为实体的属性。例如学生实体的学号、姓名、班级、选修课程等可以抽象为学生实体的属性。在设计实践中,有时实体与属性很难截然划分清楚,如同一事物,在一种应用环境中作为“属性”,在另一种应用环境中可能必须作为“实体”。一般说来,在给定的应用环境中:属性不能再具有需要描述的性质。即属性必须是不可分的数据项;属性也不能与其他实体具有联系。联系只发生在实体之间。

(2) 确定主关键字。在实体的属性中往往可以找到可以唯一标识实体的属性,被称为主关键字(primary key)。主关键字可以由一个属性或几个属性组成。

(3) 定义联系。确定实体之间的联系及其类型(1:1、1:n、m:n)。实体之间的联系是现实世界中客观事物之间的固有关系的反映。如旅游业务中,要求每一个游客必须签定一份旅游合同,那么游客与合同是一对一关系,但如果业务实践中允许几个游客(如家属)一起与旅行社签定一份合同,那么游客与合同可以定义为多对一关系。在关系数据库中,数据库管理系统将凭借这些关系把一个个看似独立的表整合成有机的数据共同体。

(4) 建立 E-R 图。模型的建立一般是先局部再全局。建立概念模型也是首先按照各分部的应用,建立不同用户视角下的信息世界的局部模型,再建立全局的 E-R 图,即合并各分部的局部模型,最后建立起整个应用系统的概念模型,即全局 E-R 图。

E-R 图将帮助设计人员全面地理解整个信息世界。但是对于大型数据库应用场合,由于涉及成千上万的实体,建立全局 E-R 图不太现实,可以按子系统或应用划分建立子系统的 E-R 图。

3. 逻辑设计(建立数据模型)

逻辑设计(logical design)就是将信息世界中的实体及其联系转换成数据世界中的数据及其联系,并用数据模型予以描述。数据模型是对现实世界的第二层抽象。

对于关系数据库来说,逻辑设计就是把概念模型转换成关系模式,即关系模型的二维表。转换步骤:

(1) 定义表。实体与属性转换成表与字段,其中实体名转换成表名,属性名转换成字段名。

(2) 确定主关键字。在表的多个字段中,能唯一确定每个记录的一个字段或一组字段,确定为表的主关键字。如在以下学生成绩表中,主关键字包含两个字段“学号”和“课程号”。“学号”和“课程号”组合决定“成绩”和“课程名称”。而在下面的学生信息表和班级信息表中主关键字均是一个字段,分别为“学生编号”和“班级名称”。

学生成绩表

学号	课程号	成绩	课程名称

(3) 建立表间联系。将实体间的关系转换成表间的关联关系,并确定关联关键字段。其中一个表的关联关键字段称为主关键字,另一个表的关联关键字段称为外部关键字(foreign key)。例如,在下面的两表中,通过学生信息表的“所在班级”和班级信息表的“班级名称”这两个属性相同的字段,建立起了一对一的关联关系,班级信息表中的“班级名称”为主关键字,学生信息表的“所在班级”为外部关键字。

在设计关系模型过程中,需要运用设计规范化理论,调整表的设计(参见第二节)。

学生信息表

学生编号	姓名	性别	年龄	所在班级

班级信息表

班级名称	班主任姓名	班级人数

4. 物理设计

数据库是要存储在物理的存储设备上的。为数据库确定一个适合的物理存储结构和存取方法的过程,就是数据库的物理设计(physical design)。现在有许多成熟的数据库管理系统产品可以为我们提供现成的方案。因此,实际上物理设计主要是选择合适的数据库管理系统产品和硬件存储设备。



体验活动

1. 找出以下各表的主关键字。
- (1) 下表中添加一个怎样的字段,可使该字段成为该表的主关键字?

某图书馆某日来馆人员借阅基本情况表

姓名	性别	年龄	来馆时间	职业	来馆事由
王明	男	20	9:30	学生	查阅资料
李鸿	男	32	10:00	教师	借书
李鸿	男	32	14:00	教师	阅读报纸
陈红	女	26	14:00	职员	阅读期刊

- (2) 在“某旅行社旅游团表”中“旅游团编号”字段能否确定为该表的主关键字? 若不能,该字段与哪个字段组合可以唯一确定该表的记录?

某旅行社旅游团表

旅游团编号	发团日期	出游线路	游客编号	姓名
CZ200304011	2003-4-19	黄山四日游	00512	王刚
CG200306001	2003-6-1	海南六日游	00808	张辉黎
CG200306001	2003-6-1	海南六日游	00807	沈东坚
CZ200308008	2003-8-5	黄山四日游	00512	王刚

2. 分析图 2-1 中某高校新生入学注册管理数据库。
- (1) 找出各表的主关键字。
- (2) 分析各表间的关系类型,找出关联关键字段。

提示:

- ① 新生报到人数表和新生分班表之间的关系为 ____。关联关键字段分别是:新生报到人数表的 ____ 字段、新生分班表的 ____ 字段。
- ② 新生分班表和学生信息表之间的关系为 ____。关联关键字段分别是:新生分班表的 ____ 字段、学生信息表的 ____ 字段。

新生报到人数表

院系编号	系名	所属学院	报到人数	未到人数
1001	计算机系	信息学院	200	5
2001	中文系	文学院	60	1
2003	外文系	文学院	80	2

接上页

新生分班表

学生编号	系别	班级名称
20040001	计算机系	一班
20040002	计算机系	二班
20040003	计算机系	二班
20041001	中文系	一班
20045005	外文系	二班

学生信息表

学生编号	姓名	性别	年龄
20040001	王董	男	20
20040002	李岚	男	19
20040003	陈立	男	20
20041001	夏阳	男	20
20045005	谢红	女	19

图 2-1

3. 设计“本班基本信息数据库”。

(1) 以小组为单位,分析学生、课程以及任课教师等基本信息。

(2) 分析学生和课程以及课程和教师之间的关系。

提示:

① 一位学生要上 _____(一/多)门课程,一门课程有 _____(一/多)位学生听课,学生和课程之间是 _____(一/多)对 _____(一/多)关系。

② 一门课程有 _____位教师教授,一位教师教授 _____门课程,教师和课程之间是 _____关系。

③ 一位学生上某一门课程,这门课程由某一位教师教授,由此,学生和教师之间通过 _____关联。一位学生得到 _____位教师授课,一位教师要教授 _____位学生,学生和教师之间是 _____关系。

(3) 用 E-R 图描述班级基本信息。

(4) 设计“本班基本信息数据库”表结构,确定学生信息表、课程表等表的表间关联关系。

第二节 数据库设计的规范化

面对同一份旅行社旅游业务需求分析报告，不同的开发人员可能会设计出不同的数据库表。怎么检验和判断自己设计得好还是不好？哪一个数据库表设计得比较合理？研究人员为数据库的设计提供了规范化理论与方法的支持。如何运用规范化理论，完善已经设计好的表？让我们一起来分析一些表，共同体验数据库设计中的规范化方法，从而了解数据规范化的思想和意义，认识其在数据库建设和使用中的价值与作用。

任务一 分析表设计的合理性

以某位同学设计的旅游信息数据库中的“旅游团表”为例，分析其设计的合理性。

1. 讨论以下问题，了解“旅游团表”使用时会发生的情况。

旅游信息数据库举例

旅游团表

旅游团编号	游客编号	线路编号	发团日期	姓名	导游工号
CZ200304011	00512	Z12-11	2003-4-19	王刚	005
CG200306001	00808	A1-2	2003-6-1	张辉黎	012
CG200306001	00807	A1-2	2003-6-1	沈东坚	012
CZ200308008	00512	A1-2	2003-8-5	王刚	022

游客表

游客编号	姓名	性别	电话
00512	王刚	男	64367970
00808	张辉黎	男	52683252
00807	沈东坚	女	62541223

(其他表如：线路表、合同表、导游表等略)

提

示

板

设计“不好”的数据库表,在使用过程中会出现的问题:

数据冗余 同一个数据会大量地重复出现在数据库的表中。

删除异常 删除表中某一记录,需要保留的信息也会随之消失。

修改困难 欲修改表中的一个数据,需要连同修改多处(字段,记录)数据。

(1) 若又有 20 位游客报名参加编号为 CG200306001 的旅游团,“旅游团表”中应增加几条相关记录? 这些新增的记录中哪些字段值(或称数据)将重复出现? 当旅行社业务量增大(如:出团增多、游客增多)时,这些字段值会随之增大重复量吗?

(2) 编号为 CZ200304011 的旅游团,目前仅一人(游客名“王刚”)报名,若他由于健康原因退团了,请在“旅游团表”中删除王刚报名参加的相应旅游团的记录。删除该记录后,与旅游团编号 CZ200304011 相关的数据,如出发时间、线路编号等基本数据在表中还存在吗?

(3) 编号为 CZ200304011 的旅游团因某种原因,发团时间需推迟 1 天,于是需修改“旅游团表”中的数据。试分析需要同时修改几条记录中的发团时间字段?(若该团已有 20 人报名)

2. 判断“旅游团表”设计得是否合理。

问题	有或无
数据冗余	
删除异常	
修改困难	

任务二 优化表

使用“旅游团表”时,之所以会出现上述的一些问题,是因为“旅游团表”中存在着数据依赖。

规范化理论能够改造表,通过分解表来消除其中不合适的数据依赖。

1. 分析“旅游团表”中主关键字与其他非主关键字之间的依赖关系。

在“旅游团表”中,主关键字为“旅游团编号”与“游客编号”的组合。

(1) 当“旅游团编号”的值确定之后,

① 该团队所对应的“发团日期”就被唯一确定了;

(记作:旅游团编号→发团日期)

② 该团队所对应的“线路编号”也就被唯一确定了;

(记作:旅游团编号→线路编号)

③ 此外,还有_____被唯一确定。

(记作:_____)

提

示

板

关系数据库的属性之间有一种“函数依赖”的关系:如果给定了属性 A 的值,则属性 B 也就确定了,称为属性 B 函数依赖于属性 A,记作 $A \rightarrow B$ 。它反映了属性之间一种相互制约又相互依赖的关系。

$A \rightarrow B$ 是完全函数依赖的条件是:从 A 中移去任何字段则依赖关系不再成立,也就是 B 依赖于整个主关键字 A,而不是其中部分。

- (2) 当“游客编号”的值确定之后，
该游客所对应的_____就被唯一确定。
(记作：_____)
- (3) 对于主关键字“旅游团编号”和“游客编号”的组合，
由于_____这些字段由“旅游团编号”就能决定，
它们都是部分依赖于主关键字，与主关键字之间不存在完全依赖关系。
- (4) “游客姓名”与主关键字的关系是_____。

2. 分析“旅游团表”符合哪些范式。

范式	是否符合
第一范式	
第二范式	
第三范式	

3. 优化“旅游团表”——分解表。

- (1) 将“旅游团表”分解为“团队表”和“游客参团表”。

团队表

旅游团编号	发团日期	线路编号	导游工号
CZ200304011	2003-4-19	Z12-11	005
CG200306001	2003-6-1	A1-2	012
CG200306001	2003-6-1	A1-2	012
CZ200308008	2003-8-5	A1-2	022

游客参团表

游客编号	姓名	旅游团编号
00512	王刚	CZ200304011
00808	张辉黎	CG200306001
00807	沈东坚	CG200306001
00512	王刚	CZ200308008

提示板

规范化理论规定，如果一个表的所有字段都是不可分的基本数据项，则该表符合第一范式。

若一个表符合第一范式，且每个非主关键字都完全依赖于主关键字，则该表符合第二范式。

若表符合第二范式，且每个非主关键字都不能传递依赖于主关键字，则该表符合第三范式。详见学习指引。

提示板

将旅游团表分解成两个表后，团队表的主关键字是旅游团编号，表中的非主关键字都完全依赖主关键字。



讨论

分解后的“团队表”、“游客参团表”中都有旅游团编号字段,这种数据冗余有必要吗?能否删除其中一个?

(2) 分析“团队表”和“游客参团表”。

① 按照任务一的方法判断“团队表”、“游客参团表”设计得是否合理?

② 两张表分别符合哪些范式,是否还可以再作分解?



学习指引

从前面的活动中,可以体会到在数据库设计中会产生的一些问题。关系数据库的设计有一套规范化设计的理论用以解决这些问题。规范化(normalization)理论是指导设计的基本原则,称为**范式**(normal form, NF)。一个“好的”设计应该尽量符合这些范式。这样才能减少如上所述问题的产生。

■ 规范化理论概要

关系规范化理论中,将关系规范化的条件分为五个等级,每一级称为一个范式,即关系模式满足的条件。范式的级别越高,条件越严格。

1. 第一范式

若一个表的所有字段都是不可分的基本数据项,则这个表符合**第一范式**(first normal form, 1NF)。

例:如下所示的表不符合第一范式。

线路编号	线路名称	价格	服务项目			
			行程	交通	用餐	住宿

在这一表中,“服务项目”可进一步分割为“行程,交通,用餐,住宿”等。因此“服务项目”不能作为表的字段。

符合第一范式表(关系)的每一列(属性)中的所有数据项都应该是同一类型的数据,列在表中的顺序没有意义。

表中的每一行(记录)都不相同,行的顺序也没有意义。

上表修改为如下所示表,则符合第一范式。

线路编号	线路名称	价格	行程	交通	用餐	住宿

2. 第二范式

若表符合第一范式,且每个非主关键字属性都完全依赖于主关键字,则这个关系符合第二范式(second normal form,2NF)。

例:如上所示修改过的表符合第二范式,因为线路名称、价格等所有非主关键字属性都完全依赖于主关键字线路编号。

另有一如下所示的收费表,它的主关键字由“旅游线路”与“游客姓名”组成,非主关键字是“应缴旅游费”。由于旅游费由旅游线路就能决定,因此是部分依赖于主关键字,不符合第二范式。

旅游线路	游客姓名	应缴旅游费

把它分成如下两个表后就符合第二范式。

旅游线路	旅游费

游客姓名	线路编号

3. 第三范式

若表属于第二范式,且表的每个属性之间都不存在传递依赖关系,则这个表属于第三范式(third normal form,3NF)。

例:下表由于价格依赖于线路,线路依赖于主关键字团号,因此不符合第三范式。

团号	线路编号	发团日期	回程日期	计划人数	价格	导游工号

把它分成如下两个表后就满足第三范式要求。

团号	线路编号	发团日期	回程日期	计划人数	导游工号

线路编号	价格

4. 其他范式

其他范式还有:Boyce-Codd 范式,第四范式,第五范式。有兴趣的同学可以阅读参考资料。

在解决一般性的问题时,通常只要把数据规范到第三范式就可以基本满足需要。

■ 关系数据库规范化总结

综上所述,数据库设计规范化的目的是使数据库的结构更加合理,消除数据存储异常,尽量地减少数据冗余,使数据的插入、更新、删除更加方便。

规范化的原则是尽量使概念单一化,也就是尽量做到一个表(关系)描述一个实体或一种实体之间的联系。本节第 51 页经优化的“团队表”和“游客参团表”以及“学习指引”第 53 页中按第二范式、第三范式优化分解而成的几个表,均体现了这一原则。

规范化的方法是分解表,把一种关系分解成两种或两种以上的关系。在分解表时须保持分解后表的集合与原表的“等价”,通过表间的联系保持原来的实体信息及其相互关系。例如,本节任务一中的“旅游团表”描述了:旅游团编号、游客编号、线路编号、发团日期、游客姓名、导游工号这些属性。经优化分解而成的“团队表”与“游客参团表”分别描述了:旅游团编号、发团日期、线路编号、导游工号;游客编号、游客姓名、旅游团编号这些属性,且两表通过关联关键字段——旅游团编号建立联系。因此,经优化分解的“团队表”与“游客参团表”“等价”地反映了原“旅游团表”中实体及其属性之间的关系。

总之,数据规范化的基本思想是逐步消除表中不恰当的数据依赖关系,并使依赖于同一个数据模型的数据达到有效的分离。

需要注意的是,规范化理论并不是金科玉律,它的运用也不是来不得半点变化,它仅仅是一种指导性建议。在设计实践中,要根据实际需要灵活运用。若刚性地运用,往往会多出许多表。表的大量增加,会使表间的联系变得复杂,以至于对数据库的一次简单访问,会要同时打开许多表,并对许多表同时进行操作才能完成,从而使数据库性能下降。只有把规范化理论的运用与各表的实际情况结合起来考虑,才能在减少数据冗余的同时保证数据库运行性能良好与数据库结构的清晰。



体验活动

1. 按照关系数据库的设计范式,完善旅游信息数据库全部表的设计。
2. 分析下表,是否满足第二范式?若不满足,请将它转换为符合第二范式的表。

学号	课程号	成绩	课程名称

3. 按照关系数据库的设计范式,完善本章第一节体验活动 3“本班基本信息数据库”各表设计。

第三节 创建旅游信息数据库

创建数据库要做两件事,一是在选定的数据库管理系统上,建立已经设计好的表及其表间的关系。二是输入实际数据。

创建数据库是按照所设计的数据模型,在计算机中建立一个实际的数据库的过程。数据库的建立一般都要经历建立表结构(空表)及各表间的关系,然后在空表中填入数据这一过程。前面两节通过对旅行社旅游业务需求的分析,确定了旅游信息数据库的基本框架,本节将借助数据库管理系统将这些纸上的设计变为计算机中的数据库。

任务一 选用数据库管理系统

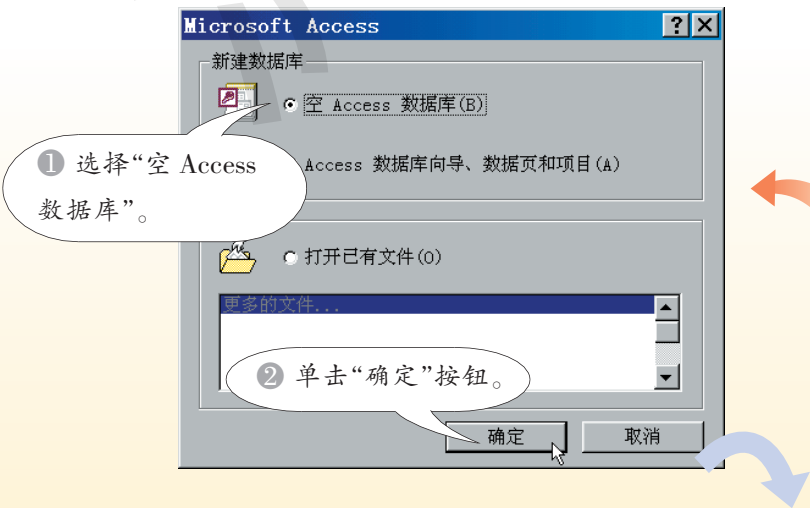
- 1. 查阅相关资料,了解什么是数据库管理系统,它有哪些功能。
- 2. 列出了解到的数据库管理系统。
- 3. 运行 Access 数据库管理系统,比较它与以前学过的应用软件有什么相同和不同之处。
相同点: _____;
不同点: _____。
- 4. 列出 Access 数据库中包括的主要对象。
主要对象有 _____。

提 示 板

Microsoft Access 是一个关系型数据库管理系统。它提供了操作向导和数据库结构模板,用于创建数据库。Access 数据库管理系统还具有网络数据管理能力。

任务二 建立空数据库

- 1. 启动 Microsoft Access 数据库管理系统。
- 2. 建立空的“旅游信息”数据库,如图 2-2 所示。



提 示 板

除使用直接创建空数据库方法外,还可以使用菜单命令“文件/新建”,或“数据库向导”等方法创建数据库。这些方法的操作同 Word 等应用软件的操作相类似。

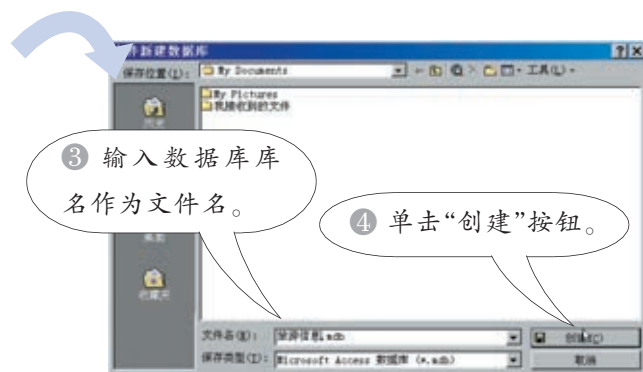


图 2-2

3. 进入“旅游信息”数据库的操作界面,如图 2-3 所示,初步熟悉界面上的工具栏、菜单栏等。

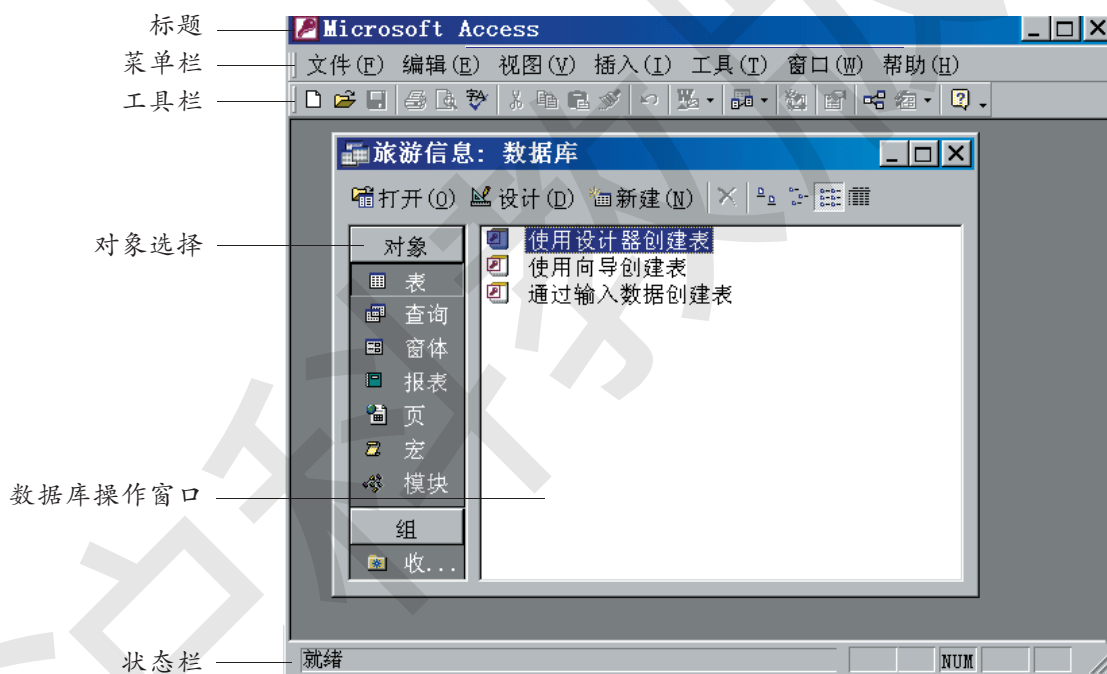


图 2-3

任务三 建立表

1. 根据已经设计好的表,确定每张表在数据库中对应的表的表名和字段名(名称可以和设计一致,但不要太长),并确定各字段的数据类型及字段大小。

数据表名称	内 容									
旅游线路	字段名									
	数据类型									
	字段大小									
旅游团	字段名									
	数据类型									
	字段大小									
游 客	字段名									
	数据类型									
	字段大小									
导 游	字段名									
	数据类型									
	字段大小									
合 同	字段名									
	数据类型									
	字段大小									


提 示 板

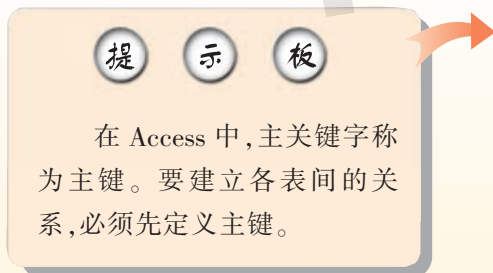
常用的数据类型有文本型、数字型、日期/时间型、货币型、逻辑型等。字段大小用来限制字段存放数据时占用的存储空间,参见学习指引。

2. 通过“设计器”来创建“旅游线路”表结构,如图 2-4 所示。



图 2-4

3. 设置主键(主关键字)。将光标移动到主关键字上,单击工具栏中的“”按钮,字段名称左侧出现主键标记,如图 2-5 所示。



主键标记 —

字段名称	数据类型
线路编号	文本
线路名称	文本

图 2-5

4. 保存“旅游线路”表。
5. 根据设计创建其他表。

任务四 建立表间关系

1. 按照第一节归纳的“导游”和“旅游团”表的关系及关联字段“导游工号”建立索引。



2. 建立其他有关系的表中的索引。
3. 按照第一节最后整理的关系建立导游和旅游团间的关系,如图 2-7 所示。

提 示 板

如果两表是一对多关系,那么在一表中关联字段必定是主关键字,而在另一表中必定是非主关键字,且数据有重复,所以这里要设置有重复。有关索引的介绍参见学习指引。

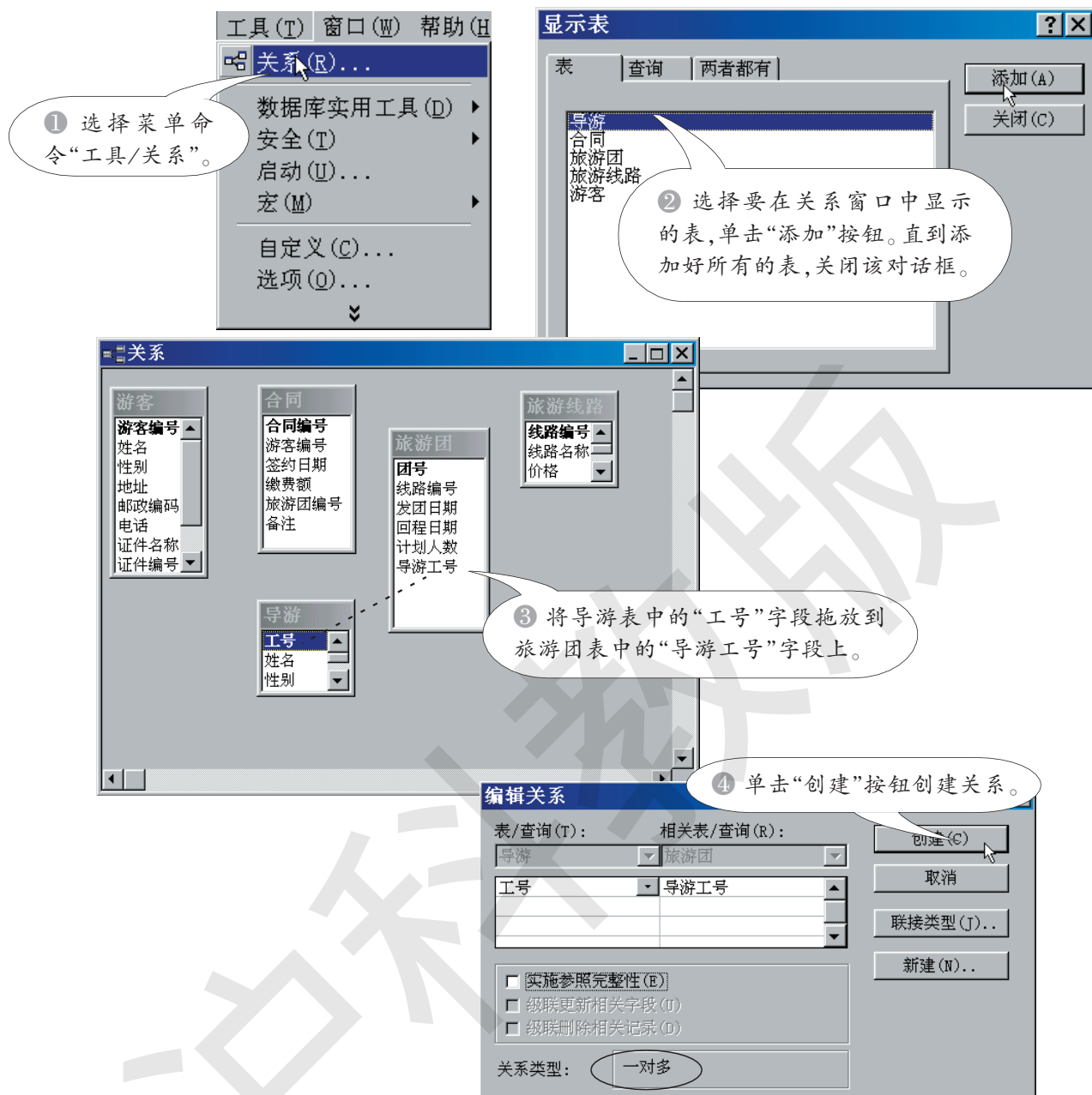


图 2-7



讨论

某表中的关联字段拖到另一张表中的关联字段上时, Access 为何能自动判断出这两张表之间存在什么关系?

4. 创建其他表间关系。
5. 退出关系窗口并保存关系。

任务五 向表中输入数据

1. 根据第一节设计的“旅游团”表,在建立的空表中输入数据,如图 2-8 所示。



图 2-8

2. 在其他空表中输入数据(见配套光盘资料)。
3. 仔细核对每张表中的数据,如有错误及时修改。

提示板

数据的编辑修改方法与 Excel、Word 中的方法相同。

任务六 导入导出数据

有些单位或曾使用过低版本的数据库管理系统建立过数据库,或曾使用过 Excel 制作过业务数据表,怎么让那些数据继续在新的数据库中发挥作用? 在建的数据库中的数据能否移入到其他数据库? Access 提供了数据的导入、导出功能。

1. 按当前数据库中的表结构修改需要导入的数据表的表结构,使两张表对应列的字段名(或列标题)一致。

2. 使用数据库管理系统提供的导入功能导入数据库文件或数据表,如图 2-9 所示。

提示板

数据导入就是将其他数据库中的数据或其他格式的数据文档输入到当前数据库中。导入功能要求:外部数据表的结构与当前数据表的结构相同。

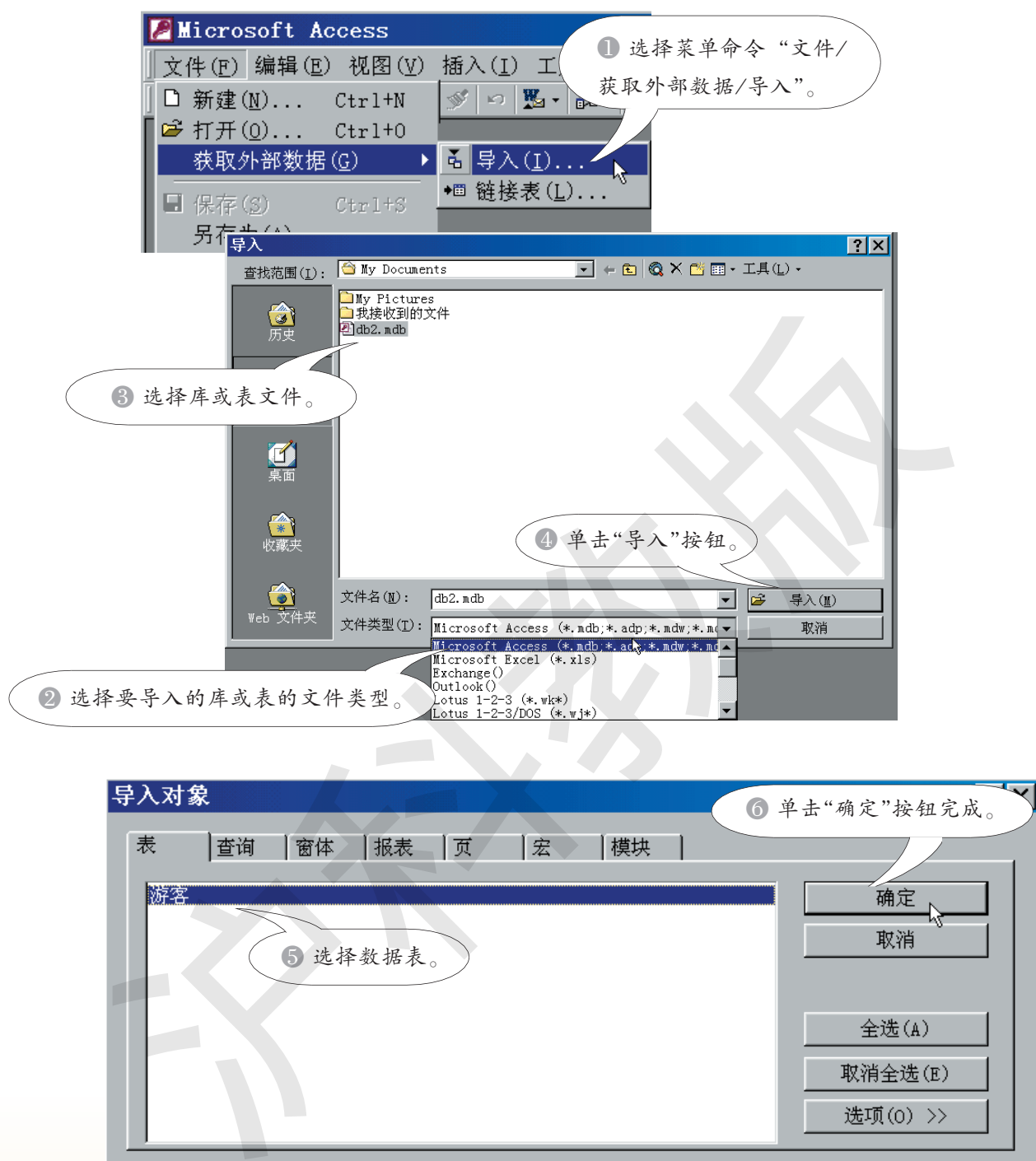


图 2-9

提 示 板

数据导出就是将数据库中的数据存入到其他数据库中或转换成其他格式的数据,如 dBase、Excel、文本等格式。

3. 使用数据库管理系统提供的导出功能导出数据库文件或数据表,如图 2-10 所示。

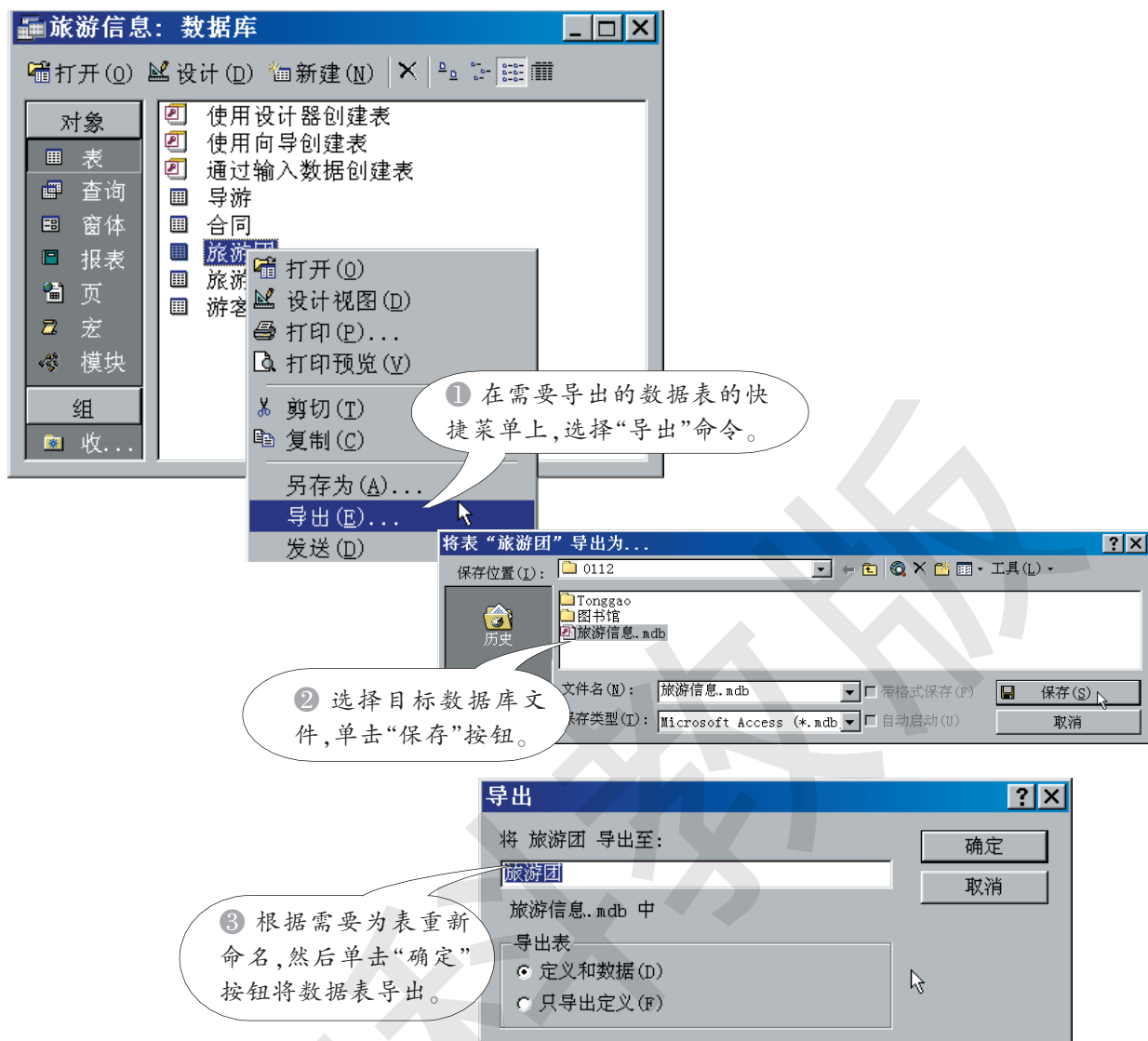


图 2-10

学习指引

使用专门开发关系数据库的数据库管理系统 Access 可以创建旅游信息数据库。本节介绍了创建关系数据库的一般过程和方法。那么究竟数据库管理系统 Access 有哪些特点?它为数据库的建立提供了哪些方便?

■ 数据库管理系统 Access

建立数据库就是将工作中产生的数据有效地组织和管理起来,实现充分利用这些数据提高决策水平和工作质量的目的。创建数据库必须依靠数据库管理系统,它是设计、开发、建立、使用和维护数据库的平台软件。数据库管理系统为用户提供了一整套用于数据库建立、维护、使用的命令,不仅实现对数据库中数据的存储管理,还提供了各种对数据进行操纵的工具,如查询、追加、插入、删除等功能,没有数据库管理系统,想实现数据的信息化管理是非

常困难的。

伴随着人们对数据处理需求的增长和变化,数据库管理系统的发展也经历了一个长期的过程。早期的数据库管理系统,如 dBase、FoxBase、FoxPro 等在信息处理中发挥了重要作用。随着信息处理向着多媒体化、网络化方向的发展,新一代数据库管理系统显示出了强大的功能。目前比较流行的数据库管理系统有 Microsoft Access、Visual FoxPro、SQL Server、Sybase、Oracle、Informix、MySQL 等。

Microsoft Access 是一个关系型数据库管理系统,它提供了功能比较齐全的操作向导和比较丰富的数据库结构模板,不仅具有友好的用户界面和操作简便的特点,还具有网络数据管理能力,是目前桌面数据库系统软件中的主流产品。

在 Access 中,表、查询及报表(下一章介绍)等都称为对象,这些对象的建立都可以通过向导的形式进行,如果要对这些对象进行修改则可以在所见即所得的设计视图中方便地进行。表间关系的建立则更为方便,只要在表间拖曳关联字段,系统就会根据定义好的主键和索引判断表间应有的关系。

■ 数据类型

日常生活、工作和学习中涉及的数据是多种多样的,有文本、数字等。在不同的场合和环境下,数据的取值范围和数据长度也有所不同。例如,用数字表示年龄时,其取值范围一般为 1~150 之间,用于高精度计算时,其取值范围可达到 $\pm 4.94065645841247 \times 10^{308}$ 到 $\pm 1.79769313486231 \times 10^{-324}$ 。又如,用文字表示性别时,一个汉字就够了,而表示旅游线路的行程时,可多到几百个汉字。数据存放在计算机中势必要占用存储空间,这就涉及到存储空间的合理分配问题,分配不合理的话往往会造成空间的浪费。例如,反映旅游线路行程的数据大约需要数十个汉字(一个汉字等于两个字符所占用的空间,即 2 个字节),所以该字段占用的空间相对比较大,而性别只需一个汉字来存储(男或女)。所以为了规范数据的使用和存储,在数据库中使用数据类型来约束。数据类型(data type)是指对数据的允许取值和取值范围的说明,它是数据的基本属性。

表的每一字段只能存放单一数据类型的数据。在创建表的结构时,一定要根据需要为字段设置数据类型。合适的数据类型才能正确表达数据,不合适的数据类型会造成数据丢失或存储空间浪费。

下表列出了 Access 中提供的几种数据类型:

数据类型	说明	大小
文本	(默认值)文本或文本和数字的组合,或不需要计算的数字,如电话号码。	最多为 255 个字符。
备注	长文本或文本和数字的组合。	最多为 65 535 个字符。
数字	用于数学计算的数值数据。	1、2、4 或 8 个字节。
日期/时间	从 100 到 9999 年的日期与时间值。	8 个字节。

货币	货币值或用于数学计算的数值数据。这里的数学计算的对象是带有 1 到 4 位小数的数据。精确到小数点左边 15 位和小数点右边 4 位。	8 个字节。
自动编号	当向表中添加一条新记录时，由 Microsoft Access 指定的一个唯一的顺序号（每次加 1）或随机数。AutoNumber 字段不能更新。	4 个字节。
是/否	Yes 和 No 值，以及只包含两者之一的字段（Yes/No、True/False 或 On/Off）。	1 位。
OLE 对象	Microsoft Access 表中链接或嵌入的对象（例如 Microsoft Excel 电子表格、Microsoft Word 文档、图形、声音或其他二进制数据）。	最多为 1 G 字节。
超级链接	文本或文本和数字的组合，以文本形式存储并用作超级链接地址。	Hyperlink 数据类型的三个部分的每一部分最多只能包含 2048 个字符。
查阅向导	创建字段，该字段可以使用列表框或组合框从另一个表或值列表选择一个值。	与用于执行查阅的主键字段大小相同，通常为 4 个字节。

对于文本类型和数字类型的数据,可以根据需要进一步选择。文本类型,其范围可设置为 0 到 255 之间的数字,默认值为 50(可存放 25 个汉字或 50 个西文字符)。数字类型,根据数值范围和计算精度又分为如下类型。

数据类型	说 明	小数位数	占用存储量
数字	保存从 0 到 225(无小数位)的数字。	无	1 个字节
小数	存储从 $-10^{38} \sim 1$ 到 $10^{38} \sim 1$ (.adp) 范围的数字 存储从 $-10^{28} \sim 1$ 到 $10^{28} \sim 1$ (.mdb) 范围的数字	28	12 个字节
整型	保存从 -32 768 到 32 767 (无小数位)的数字	无	2 个字节
长整型	(默认值)保存从 -2 147 483 648 到 2 147 483 647 的数字 (无小数位)	无	4 个字节
单精度型	保存从 -3.402823E38 到 -1.401298E -45 的负值,从 1.401298E-45 到 3.402823E38 的正值。	7	4 个字节
双精度型	保存从 -1.79769313486231E308 到 -4.94065645841247E-324 的负值, 从 1.79769313486231E308 到 4.94065645841247E-324 的正值。	15	8 个字节

在为字段选择数据类型时,应坚持最小化原则,因为较小的数据占用存储空间小,需要的内存少,处理的速度快。

■ 字段属性

在 Access 中建立表,需对表的每一字段设置字段属性,以对字段的数据进行规范。

本节任务三中涉及字段大小属性的设置。设置字段大小,可有效控制存储空间。例如,图 2-4 中,线路编号字段是文本类型的数据,该类型数据的默认值为 50 个字符,现可根据需要设置小于 50 的数字,从而避免不必要的空间浪费。

本节还涉及索引(index)属性的设置。索引是按照一个或多个字段的值使表中记录有序排列的一种技术。一个表可以创建多个索引,一般常对这些字段设置索引属性:经常搜索的字段,要排序的字段以及与其他表的字段相关联的字段。

字段的索引属性有三种选项:无,表示不设置;有(有重复),表示设置,且该字段的值允许重复;有(无重复),表示设置,但该字段的值不允许重复。

任务四是关于建立表间关系的索引属性设置。在 Access 中,若要建立表间的关联关系,必须对要建立关系的两表的关联关键字段设置索引属性,通过索引字段的值来建立表间的关联关系。例如,任务四中是通过对“旅游团”表的字段“导游工号”和对“导游”表的字段“工号”设置索引属性来建立这两表间的关系的。

两表间若要建立“一对多”的关系,则一表的关联关键字段为主关键字并设置为有索引(字段值无重复),另一表的关联关键字段为非主关键字,设置为有索引(字段值有重复)。如,任务四中“导游”表和“旅游团”表的表间关系为“一对多”,则“导游”表的字段“工号”为主关键字,设置索引属性为“有(无重复)”,“旅游团”表的字段“导游工号”为非主关键字设置为“有(有重复)”。

两表间若要建立“一对一”的关系,则两表的关联关键字段都必须为主关键字并设置索引属性为“有(无重复)”。

Access 会自动将主键设置成无重复值的索引。

另外,为了避免数据在输入时与数据类型不符,如在设置为日期类型的数据格中输入了汉字“我”,或在性别字段中输入了“男女”,从而造成错误,可以设置有效性规则这一属性。有效性规则的作用是当输入数据不符合规则时给予提示并禁止输入。

其他字段属性,如小数位数、格式等的作用和用法可以参见 Access 帮助文件。



体验活动

根据本章第一节体验活动 3 所设计的“本班基本信息数据库”,在 Access 中创建该数据库,并将本班的基本信息输入数据库。

提示:

(1) 确定各表各字段的属性(数据类型、字段大小、索引等)。

接上页

- (2) 在 Access 中创建表。
- (3) 在“关系”窗口中,创建各表间的关联关系。
- (4) 小组合作分工输入数据。
- (5) 使用 Access 提供的导入或链接表功能将多人输入的数据导入一个数据库文件中。
- (6) 描述创建数据库的过程和使用的创建方法。



后续活动

设计并建立一个管理家庭图书的数据库 (该数据库至少要能存放家中各成员购买各类图书的基本信息)。

(1) 需求分析结束后,在小组内交流自己的工作。向同伴说明自己设计的数据库要提供给家中哪些人使用,计划解决哪些家庭图书管理工作,以及数据库中要保存、处理、输出哪些数据。

(2) 将收集到的数据进行分类,定义实体及属性,设计出表结构。小组讨论,一起优化各自设计的数据库。

数据表名称	内 容									
家庭成员	字段名									
	数据类型									
	字段大小									
	字段名									
	数据类型									
	字段大小									
	字段名									
	数据类型									
	字段大小									
	字段名									
	数据类型									
	字段大小									

(3) 在 Access 中创建家庭图书数据库,并录入相关数据。



评价

一级指标	评价内容	评价(★★★、★★、★)		
		自评	互评	教师评
信息技术知识	理解实体、属性、关系的含义。			
	掌握关系数据库中的库、表、记录、字段等概念。			
	理解数据模型、层次模型、网状模型的基本概念,充分理解关系数据模型。			
	了解数据规范化的思想和意义。			
信息技术能力	初步学会在需求分析时收集数据,并进行分类整理。			
	会使用实体—关系图描述关系数据模型。			
	初步掌握需求分析、概念设计、建立数据模型等数据库设计的一般方法。			
	能分析数据库中的表及表间关系,能找出各表的主关键字及表间的关联关键字段。			
	能使用数据库管理系统软件 Access,学会创建数据库、表及表间关系的方法。			
	掌握数据录入、修改的方法,学会从同类数据库或其他格式文件导入数据及导出数据的方法。			
行为	能意识到需求分析的重要性,认识到数据库是为应用需要服务的。			
	能认识到数据规范化的现实价值,在设计数据库时有优化的意识和行为。			
	大量数据录入时能明确分工,并充分合作完成任务。			
选择自己或小组合作设计建立的一个数据库(如“本班基本信息数据库”或“家庭图书数据库”等)。 总结交流:(①确定了哪些实体及实体属性?②进行了哪些规范化工作?③是否分析过表间关系?④如何设置字段的属性?⑤如何分工创建表及输入数据?⑥向该数据库导入导出数据是否成功?)				



参考资料

数据采集

根据不同的数据来源,可以选择不同的数据采集途径。需要调查人员参与采集的数据称为人工采集

数据。例如,健身教练请会员填写表格,采集会员基本信息;交警向驾驶员询问交通事故发生的情境;心理咨询师与来访者交流,了解对方的心理状况等。人工采集数据通常使用观察法、访谈法、测验法等调查方法。

除了人工采集数据之外,还可以利用设备采集数据。例如,利用录音机、摄像设备等数据采集设备采集现场音频、图像、视频等数据;利用各种传感器实时采集温湿度、压力、速度等数据。

此外,由于互联网技术的迅速发展,互联网已经成为一种重要的数据来源。互联网上的海量数据可以使用自动化的采集软件获取。例如,网络爬虫(或称网页蜘蛛)就是一种按照一定规则自动抓取网络数据的程序或脚本。

如今,随着技术的发展,数据采集的途径和工具越来越多,数据泄露、数据侵权问题也日益严重。一方面,我们在使用某些技术和工具采集数据时,需要遵循相应的道德规范,维护良好的网络环境和秩序。例如,在使用网络爬虫软件前,需要了解哪些网页数据可以抓取,哪些不能,否则有可能侵权;抓取的数据若属于他人的隐私或商业秘密,应及时停止抓取并删除。另一方面,我们在日常生活中要注意保护自己的个人数据不要被他人悄然或恶意地采集。例如,应谨慎使用社交网络平台,你填写的个人信息有可能会透漏你的真实身份;应妥善处理存储有个人账户资料的废旧手机,更换之前务必做好彻底清理工作。

旅游信息数据库主要表结构的设计参考

导游

工号	姓名	性别	电话	护照号码	身份证号码	外语 1	外语 2

旅游线路

线路编号	线路名称	价格	行程	交通	用餐	住宿

旅游团

团号	线路编号	发团日期	回程日期	计划人数	导游工号

合同

合同号	游客编号	签约日期	缴费额	旅游团编号	备注

游客

游客编号	姓名	性别	地址	邮政编码	电话	证件名称	证件编号	报名标志

BCNF 范式、第四范式和第五范式

第三范式虽然消除了非主属性对主关键字的部分函数依赖和传递函数依赖,但仍有不合适的性质。请看下例。

例如,关系表(学生,教师,课程),语义假设是,每一位教师仅教一门课;每门课有若干个教师任教;某一学生选定某门课,就对应于一个确定的教师。

根据定义假设,上述关系表的函数依赖为:(学生,课程) \rightarrow 教师,教师 \rightarrow 课程。

由函数依赖可知,关系表的主关键字是(学生,课程),教师是非主属性,关系表中不存在非主属性对主关键字的部分依赖和传递依赖,因此,该关系表属于第三范式。但从该关系表分析,仍有一些问题。

(1) 数据仍有许多冗余,教师上课的信息、学生选课信息与学生选此课的人数一样多。

(2) 插入异常。当某课程本学期不开,自然就没有学生选修。没学生选修,教师上该门课的信息就无法插到此数据库中去。

(3) 删除异常。当学生修完某课,则把此学生选课记录删除,随之也删除了教师开该课的信息。

因此符合第三范式的关系表还必须进一步分解,以消除数据库操作的异常现象。我们把关系表分解成(学生,教师)和(教师,课程),从而把学生选课的信息与教师任课的信息分开,这样就消除了大量冗余,也解决了插入异常与删除异常现象。分析上述两个关系表可见,它们的关键字分别是(学生,教师)和教师。函数依赖是(学生,教师) \rightarrow 学生;(学生,教师) \rightarrow 教师。

关系表范式的决定因素是关键字。Boyce 和 Codd 研究了这类关系表的特点和规律,共同提出了 Boyce-Codd 范式,简称 BCNF。若关系属于第一范式,且每个属性都不传递依赖于主关键字,则这个关系属于 BC 范式。

前面我们在函数依赖的范畴内讨论了关系规范化问题,根据函数依赖集 F,我们可以把一个关系分解为若干个 BCNF 的集合。但是函数依赖表示的是关系中属性间的一对一和多对一的联系,而不能支持属性间多对多的联系,因而一些关系虽然已满足 BCNF 范式,仍然有一些弊端。第四范式是在多值依赖范围内定义的,多值依赖问题能表示属性间多对多的联系。

第五范式是在联接依赖范围内定义的,一个关系属于第五范式(5NF)当且仅当该关系中的每一个联接依赖都是由它的主关键字所蕴涵的。