

DATABASE

Principle and
Application

数据库 原理及应用

主编 谢霞冰



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

数据库原理及应用
principle and Application

数据库 原理及应用

主编 谢霞冰
副主编 陈晓峰 赵雷 葛艳
参编 李净 王令群 贺琪
郭洪禹 王文娟 杨蒙召
王德兴



上海交通大学出版社

SHANGHAI JIAO TONG UNIVERSITY PRESS

内容提要

本书结合 SQL Server 2008 和 Northwind 示例数据库简明扼要地介绍了关系数据库的基本原理。包括：关系模型及关系运算理论、SQL 语言、关系模式的设计与规范化、数据库的设计、事务管理与数据库安全保护等；并以数字海洋与数字渔业课题中的一个子项目《基于 WebGIS 的物流协同溯源系统——虾苗苗种投放管理系统》为案例介绍了数据库应用程序开发的基本流程与技术；案例基于微软 .NET 技术和 C# 语言开发，使用 ADO.NET 作为数据访问组件，分别实现了 C/S 和 B/S 架构下的应用程序，并附光盘提供全部源代码。

本书可作为高等院校数据库原理与应用类课程教材，也可作为相关技术人员学习参考。

图书在版编目(CIP)数据

数据库原理及应用 / 谢霞冰主编. —上海：上海交通大学出版社，2016
ISBN 978 - 7 - 313 - 14213 - 9

I. ①数… II. ①谢… III. ①数据库系统—高等学校—教材 IV. ①TP311.13

中国版本图书馆 CIP 数据核字(2015)第 313954 号

数据库原理及应用

主 编：谢霞冰

出版发行：上海交通大学出版社

邮政编码：200030

出 版 人：韩建民

印 制：上海宝山译文印刷厂

开 本：710 mm×1000 mm 1/16

字 数：287 千字

版 次：2016 年 1 月第 1 版

书 号：ISBN 978 - 7 - 313 - 14213 - 9/TP

定 价：48.00 元

地 址：上海市番禺路 951 号

电 话：021 - 64071208

经 销：全国新华书店

印 张：17.75

印 次：2016 年 1 月第 1 次印刷

版权所有 侵权必究

告读者：如发现本书有印装质量问题请与印刷厂质量科联系

联系电话：021 - 56482128



前　　言

数据库技术是信息系统的核心技术之一,主要研究数据的组织、存储、检索和处理。自上世纪 60 年代产生以来,随着信息化的发展,其应用领域逐渐扩展到各行各业的各个层次,尤其在管理信息系统、电子商务、情报检索、人工智能与专家系统、计算机辅助设计等领域,数据库技术是最基本最重要的技术。因此,高校的计算机类、管理类相关专业普遍开设了数据库技术的课程,同时数据库技术的相关课程作为自然类的选修课程也十分受学生欢迎。

编者从事数据库课程教学多年,针对不同的学生群体讲授了各个层次的数据库课程,这些课程中既有针对计算机专业的数据库原理课程,也有针对其他专业的数据库应用课程。多年教学和实践积累奠定了编写本书的基础。

全书着眼于将数据库的原理与应用有机融合,以认识和掌握数据库应用全貌为目标,按数据的组织、访问、数据存储框架的规划、数据库设计的全过程、数据库的安全保障、数据库应用程序的实现为主线,分 7 章介绍数据库原理与应用技术。

第 1 章从身边的数据库应用实例出发,探讨了在应用需求推动下的数据管理的发展历程,由此引出数据库系统的概念;接着以数据为主体,介绍数据抽象过程和描述方法以及数据抽象的模型;然后介绍了数据库的概念、体系结构以及数据库系统和数据库管理系统,在体系结构中详细介绍了数据库的三层模式体系与两级映像结构;最后,对本书的贯穿示例数据库 Northwind 进行了介绍。

第 2 章从关系模型的概念、关系代数、关系演算以及查询优化等方面详细介绍了关系数据库的原理,为深入理解关系数据库打下了基础。

第 3 章分两部分介绍了 T-SQL 语言。第一部分介绍了数据定义语言、数据查询语言、数据操纵语言和数据控制语言;第二部分介绍了游标、存储过

程和触发器等概念。

第4章讨论关系模式的设计中存在的问题和模式设计的规范化理论,概述了关系数据库设计的指导原则。

第5章结合实例介绍了横跨数据库设计的五个阶段:需求分析、概念结构设计、逻辑结构设计、物理设计、数据库实施和运行维护。

第6章介绍了数据库事务和事务的并发控制、数据库安全保护以及数据库的备份与恢复。

第7章首先介绍数据应用程序开发的一般流程和技术;接着介绍常用的数据库访问技术,对ADO.NET技术做了重点介绍;然后结合案例介绍在.NET应用程序中访问数据库的一般方法。

本书是上海海洋大学信息学院十多位教师结合多年的授课经验编写的。从策划到最终定稿历时整整1年时间。第1章由葛艳老师编写;第2章由李净老师编写,第3章由贺琪老师和郭洪禹老师合作完成,第4章由王令群老师完成,第5章由王文娟老师和杨蒙召老师合作编写,第6章由陈晓峰老师和王德兴老师合作编写,第7章由谢霞冰老师编写,葛艳和赵雷老师负责全书最初思路的梳理以及各章完稿后的协调。全书的完成离不开各位老师的通力合作以及研究生黄幸幸、裴丽娜,本科生陈润发、全芸芸同学在文字排版以及项目梳理方面的参与。

编者致力于编写一本用实例贯穿、语言简练、知识点简明扼要的数据库教材。然而,由于时间仓促,经验不足以及编者的水平有限,书中存在的错误和疏漏之处,恳请广大读者和各位同行能够提出宝贵意见。

本书在编写过程中得到了上海海洋大学信息学院副院长袁红春教授的大力支持,另外,所用实例取材于信息学院副院长陈明教授的科研项目,书中仅剥离了项目中很小的一个足以阐明数据库的原理和应用场景的子集,没有他们的支持,也没有今天这本教材的面世,在此对他们的支持表示衷心的感谢。同时还要感谢参与本书编写的所有老师和学生们的辛勤劳动;感谢上海海洋大学教务处、信息学院对编者工作的支持。最后,感谢为本书的最终出版而辛劳审稿的上海交通大学出版社的编辑们。

编 者

2015年12月

目 录

第1章 绪论	1
1.1 从身边的数据库应用说起	1
1.2 数据管理的发展过程	3
1.2.1 人工管理阶段	3
1.2.2 文件系统管理阶段	3
1.2.3 数据库系统管理阶段	6
1.3 数据的抽象与描述	7
1.3.1 数据抽象的过程	7
1.3.2 数据抽象的模型	8
1.3.3 完整的模型映射	11
1.4 数据库系统概述	12
1.4.1 数据库的基本概念	12
1.4.2 数据库的体系结构	16
1.4.3 数据库管理系统的分类	18
1.4.4 数据库系统分类	21
1.5 本书应用实例介绍	23
本章小结	28
本章习题	28
第2章 关系模型及关系运算理论	30
2.1 关系模型	30
2.1.1 关系模型概述	30
2.1.2 关系数据结构	32

2.1.3 完整性规则	37
2.2 关系代数	38
2.2.1 传统的集合运算	38
2.2.2 专门的关系运算	40
2.2.3 关系运算举例	43
2.3 关系演算	46
2.3.1 元组关系演算	46
2.3.2 域关系演算	48
2.4 查询优化	50
2.4.1 查询优化问题的提出	50
2.4.2 查询处理与优化技术	52
2.4.3 关系代数表达式等价变化规则	55
2.4.4 查询优化算法	57
本章小结	60
本章习题	61
 第3章 SQL语言	65
3.1 Transact-SQL语言简介	65
3.2 数据定义语言DDL	66
3.3 数据查询语言DQL	70
3.4 数据操纵语言DML	86
3.5 视图	87
3.6 游标	89
3.6.1 游标的概念	89
3.6.2 游标的使用方法	90
3.7 存储过程	98
3.7.1 存储过程的概念	98
3.7.2 存储过程的使用方法	99
3.8 触发器	105
3.8.1 触发器的概念	105
3.8.2 触发器的使用方法	106

本章小结	113
本章习题	114
第 4 章 关系数据库的规范化设计	117
4.1 关系模式设计中的问题	118
4.2 数据依赖	119
4.2.1 函数依赖基本概念	119
4.2.2 一些术语和符号	120
4.3 数据依赖的公理系统	121
4.3.1 Armstrong 公理系统	122
4.3.2 闭包	122
4.3.3 闭包的计算	123
4.3.4 函数依赖的等价	123
4.3.5 函数依赖的最小化	124
4.4 关系模式的分解特性	125
4.5 关系模式的范式	127
4.5.1 关系模式中的码	128
4.5.2 范式	129
本章小结	134
本章习题	135
第 5 章 数据库设计	137
5.1 数据库设计概述	137
5.2 需求分析	140
5.2.1 需求分析的内容	140
5.2.2 需求分析的方法	141
5.3 概念设计	147
5.3.1 E-R 模型概述	148
5.3.2 基于 E-R 模型的概念结构设计方法与步骤	153
5.3.3 局部 E-R 模型设计	154
5.3.4 全局 E-R 模型设计	158

5.3.5 全局 E-R 模型的优化	161
5.3.6 虾苗苗种投放管理系统的 E-R 模型设计	163
5.4 数据库概念模型到逻辑模型的转换	165
5.4.1 E-R 图向关系模型的初始化转换	166
5.4.2 关系模式的规范化	170
5.4.3 关系模式的评价与改进	170
5.4.4 虾苗苗种投放管理系统的逻辑模型设计	172
5.5 物理设计	173
5.5.1 确定数据库的物理结构	174
5.5.2 数据库物理结构的评价	176
5.6 数据库的实施和运行维护	176
5.6.1 数据库实施	176
5.6.2 数据库运行维护	177
本章小结	179
本章习题	179

第 6 章 事务管理与数据库安全保护	181
6.1 事务	181
6.1.1 事务的特性	181
6.1.2 事务的分类	182
6.2 数据库的并发控制	182
6.2.1 并发操作引起数据的不一致性	182
6.2.2 并发控制措施	184
6.2.3 锁的模式	185
6.2.4 锁的粒度	187
6.2.5 死锁	187
6.3 数据库安全保护	188
6.3.1 数据库安全概念与内涵	188
6.3.2 数据库的安全策略	189
6.3.3 数据库安全机制	190
6.3.4 数据库安全管理的集中控制与分散控制	213
6.3.5 提升数据库系统安全的实现途径	213

6.3.6 数据库安全的发展趋势	215
6.4 备份与恢复	215
6.4.1 数据库备份恢复机制	216
6.4.2 影响备份与恢复策略的主要因素	219
本章小结	221
本章习题	221
第 7 章 数据库应用案例	223
7.1 案例概述	224
7.1.1 案例背景与需求	224
7.1.2 案例开发采用的技术路线	225
7.1.3 案例的数据库设计	225
7.2 数据库应用程序开发概述	227
7.2.1 数据库应用程序的开发流程	228
7.2.2 数据库应用系统的体系结构	228
7.2.3 .NET 框架与 C# 程序设计语言	230
7.3 数据库访问技术	231
7.3.1 数据库访问技术概述	231
7.3.2 ADO.NET 数据访问组件	234
7.3.3 .NET 应用程序中的数据库读写	237
7.4 系统设计	242
7.4.1 模块功能说明	243
7.4.2 虾苗检验和投放流程	243
7.5 应用程序的实现	244
7.5.1 C/S 应用程序的实现	244
7.5.2 B/S 应用程序的实现	258
本章小结	265
本章习题	266
章节习题参考答案	267
参考文献	274

第1章 绪论

本章的核心目的：

1. 认识数据库

从三个维度上来认识数据库：

第一个维度是发展的维度。即从时间轴上看，数据库是如何从无到有发展起来的？其发展历程和发展变化如何？

第二个维度是认知的维度。认识一个事物，不外乎从内外两个角度来认知。对于数据库而言，从外在的功能上看，包括什么是数据库，数据库解决了什么问题，数据库可以用在哪些领域，我们身边有哪些数据库应用的示例等等。从数据库的内部构成角度看，需要认识数据库的构成部件、简单的工作原理。在此基础上，继续深入认识其中的核心：数据。对于数据的认识我们重点关注数据从哪来，如何表示，即数据的抽象过程以及数据的描述形式。

第三个维度是问题的维度。即数据库解决了哪些问题，如何用数据库解决具体问题。具体来说，包括：数据的存储问题、数据的访问及控制问题、数据的安全问题、数据的备份、恢复及传输问题。除此之外，还有数据库的设计以及数据库如何使用等问题。

2. 了解书中将要贯穿的实例

1.1 从身边的数据库应用说起

在日常生活中，我们曾有过买火车票、飞机票，超市购物、网上购物，银行存钱、取钱的经历。在校园生活中，我们都有过选课、查学分、查成绩，到图书馆借书、还书，为校园卡充值、消费等体验。所有这些经历、体验或操作背后对

应的大量票务信息、消费信息、账户信息、课程信息等等都存在哪？如何管理？又是如何为我们提供服务的？

我们把以上各类信息统称为数据。对这些数据实现管理的应用软件称为数据库应用。其实类似的应用在各行各业都很常见：

如银行通过存储储户的个人信息、账户状态交易明细以及贷款、还款或者存款类别等信息，方便用户通过柜面、ATM、网络或者手机读取这些信息；

航空、铁路部门通过存储票务信息、订票信息等，方便旅客查询和购买需要班次或车次的机票或车票，还要保证多点同时售票时不要出现一票多卖情况的发生；

销售业行业中存储销售商品清单、数量、到货期、价格、类别以及销售流水账等信息，通过网络和各种移动设备，方便消费者随时随地完成交易。

除此之外，电信、农业、物流、工业、旅游、服务等等行业也都有着类似的应用。

对以上各种应用进一步分析，不难发现，其共同特点是以数据为中心。这些数据具有以下一些特点：

(1) 数据量庞大。以火车票购票应用为例，为了能通过一个信息化系统实现火车票的销售，粗略估计一下，至少需要把所有车次、所有时间的票务数据存储起来，还要存储所有的起止站点数据、各类不同车型、不同类别座位的价格数据、销售的状态数据以及购票人的个人数据、购票的历史数据等等；数据种类繁多，单类信息数据量也是很庞大的。

(2) 数据不会随着程序的结束而消失。如读者到图书馆去借书，图书馆下班，机器都关机后，供读者使用的借书应用程序也退出了。可是图书馆的借书系统中仍然记录着读者什么时候借了什么书，应该什么时候归还的数据，这些数据没有因为程序的结束或退出而消失。

(3) 数据可以被多个应用或在多个位置被共享。以火车票售票为例，一个时间点发车的一辆火车的票务数据可以在多个地点、多个城市同时开始销售，也就是说一辆车次的票务数据可以被多个不同销售点的应用程序所共享。

那么这样的数据是怎样被有序地规划、组织和管理，从而为各类应用程序所用的呢？该如何去规划和实现这些数据的管理呢？

本书将系统地介绍规划、组织和管理数据的基本原理，以及利用成熟的数据库技术实现数据管理的方法。

1.2 数据管理的发展过程

在数据管理领域,数据与信息是密不可分的。一般把信息理解为原始事实,即关于现实世界事物存在的方式或运动的状态。而数据则是用符号记录下来的、可以识别的信息。因此,可以形象地理解为信息是数据的内涵,是对数据语义的解释,而数据是信息的载体,是信息的表示形式。在后续内容中,我们不再严格区分这两个概念,可互换使用。

数据管理是指对数据进行分类、组织、编码、存储、检索、维护和传输的操作。这些操作是数据处理的基本环节,是对数据进行进一步分析、归纳,并提取出有效信息资源的基础。

自计算机问世以来,在应用的推动下,对数据实现管理的技术主要经历了以下几个发展阶段。

1.2.1 人工管理阶段

20世纪50年代中期以前,计算机主要用于科学计算,这个时期数据的管理是由应用程序自己完成的,没有相应的软件系统负责。当对数据有处理需求时,人工将数据输入,处理完成后可以看到处理结果,但无论是原始数据还是结果都不保存。所有的管理都是人工完成。

这个阶段其实还谈不上用计算机实现数据管理,当时的硬件条件还不具备,还没有直接存取存储设备,也没有操作系统。数据都是在程序需要时,人工批量输入的(见图1-1)。

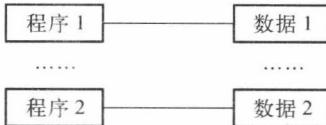


图1-1 人工管理阶段

1.2.2 文件系统管理阶段

20世纪50年代后期到60年代中期,硬件方面已经有了磁盘、磁鼓等直接存取存储设备;软件方面,操作系统中已经有了专门的数据管理软件,一般称

为文件系统;处理方式上不仅有了批处理,而且能够联机实时处理。于是,出现了使用文件系统实现数据管理的方式(见图 1-2)。

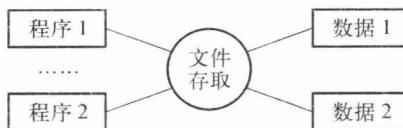


图 1-2 文件系统管理阶段

与人工管理阶段相比,这个阶段数据管理在以下方面已经有了明显进步:

(1) 数据可以长期保存。数据以文件的方式借助磁盘、磁鼓等设备长期保存,从而便于反复地进行查询、修改、插入和删除等操作。

(2) 由文件系统管理数据。文件系统把数据组织成相互独立的数据文件,利用“按文件名访问,按记录存取”的管理技术,实现对文件的操作。在这个前提下,对数据的处理不再仅停留在批处理方式上,还增加了联机实时处理能力。

(3) 数据与程序脱离。借助文件,数据可以独立于程序保存在文件中,并且文件系统实现了记录内的结构化,程序和数据之间的对话以文件系统实现中转,因此数据已经有了一定程度的独立性和结构性。这样,程序员可以不必过多地考虑物理细节,而将精力集中于算法,而且数据在存储上的改变不一定反映在程序上,节省了维护程序的工作量。

从人工管理发展到文件系统数据管理是数据管理的一大进步,但这个时期人们对使用计算机实现数据管理已经寄予了更高的期望,因此,随着文件管理方式使用范围的推广,文件系统实现数据管理方式的弊病或不足也暴露无遗:

(1) 数据冗余、数据不一致导致维护数据一致的难度很大。文件仍然是面向具体应用的。文件和程序是根据业务需求,在很长的一段时间里由不同的程序员实现的。不同的文件可能具有不同的结构,不同的程序也可能是采用不同的编程语言实现的。即一个文件基本上对应一个应用程序,因此,数据的独立性、共享性和结构性程度还很低,当不同的应用程序具有部分相同的数据访问需求时,仍然需要各自建立独立的数据文件,而不能共享相同的数据,因此数据冗余度大,浪费存储空间。同时,由于相同数据的重复存储、各自独

立管理,还容易造成数据的不一致性,比如在学生基本信息管理中存储的学生姓名和学生学号,在学生成绩管理中存放学生姓名和学生学号。当修改学生基本信息管理中对应一个学号的姓名时,如果其他位置的对应姓名不修改,则数据不一致,如果要去修改,因为不知道其他还有多少位置存有姓名,其他用以访问姓名的应用程序又是如何编写的,要一一去修改从而保持一致,工作量很大,而且很不方便。

(2) 数据对应用的适应性差。每当用户有新的数据访问需求时,即使这个需求涉及的数据在原有的数据文件中已经存在,但由于文件访问方式下,数据的最小存取单位是记录,因此,即使是仅访问原来数据文件中的部分数据项,也还是需要重新开发新的应用程序来实现访问。

(3) 并发访问异常。文件系统管理数据阶段,数据的并发能力完全体现在文件的并发操作上,如果有多个应用需要并发访问同一个文件,常常发生不能并发,或需要程序做复杂的协调控制策略,抑或发生数据异常等情况(后续章节将有详细地分析)。

(4) 原子性问题。有时对数据的操作会涉及几个相关的动作,比如,银行里要实现 A 账户向 B 账户转账 X 元,为了描述的简化,假设 A 账户有足够的金额完成转账。这项操作本质上至少要关联几个相关的文件操作,首先从存储 A 账户金额的文件中将原有金额减去 X,将新的金额写入文件,然后再打开存储 B 账户信息的文件,向 B 账户中将原有金额加上 X,再写入文件,从而完成转账操作。如果在 A 账户金额减少后,计算机系统出现故障,那么,由于两个动作的独立性,可能导致 A 账户的金额减少了,但 B 账户的金额却没有相应增加。我们把不能被分割的几个动作称为一项操作的原子性。而对应操作的异常称为原子性问题。

(5) 安全性问题。当数据被大量保存在文件中后,用户对于数据的安全又提出了各种各样的需求,有些数据文件只能对指定的用户可读,有些数据文件分记录决定哪些用户可见哪些内容,还有些时候规定一些用户只能看一条记录的部分属性,其他属性不可见等等。种类繁多的数据访问权限需求在文件管理阶段都是很难控制的。

随着计算机管理的对象规模越来越大、应用范围越来越广泛、数据量的急剧增长,同时多种应用、多种语言互相覆盖地共享数据集合的要求越来越强烈。以上问题的存在使得文件系统管理方式已经无法胜任实际的数据管理需求。

1.2.3 数据库系统管理阶段

20世纪60年代后期,大容量磁盘的出现、硬件价格下降在硬件上满足了大数据量存储的需求;在处理方式上,为解决上述文件系统存在的各种问题,数据库技术应运而生,出现了统一管理数据的专门软件系统——数据库管理系统(见图1-3)。

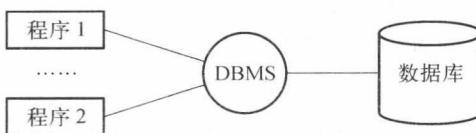


图1-3 数据库管理系统阶段

这个阶段的数据管理方式在以下方面实现了革新:

1. 数据结构化

文件管理中,相互独立文件的记录内部是有结构的,但数据整体是无结构而言的。但在数据库管理中,数据不是面向一个应用,而是面向一个组织的多个应用,甚至跨组织的多个应用,综合考虑整个组织或跨组织应用的需求的、纷繁复杂的数据按一定的数据模型进行组织。在数据库系统阶段,数据模型的发展经历了从层次模型、网状模型到关系模型、对象模型以及XML模型的变化过程。其中,关系模型仍是最成熟、最常用的数据组织模型,我们现在最常用的数据库大多还都是基于关系模型实现数据组织的。

按照数学模型组织的、通盘考虑多个应用需求的结构化数据组织方式大幅度地减少了数据冗余,并且对数据的访问单元也从记录缩小到了数据项(描述一个学生的多个属性集合为记录,每一个属性则为一个数据项,后续会有详细介绍)。数据的并发、共享以及对多应用的适用性以及保障数据的一致性从理论上都成为可能。

2. 对数据的操作由独立的软件完成

在数据库管理方式中,数据按照一定的结构组织并存储在数据库中。应用程序不能直接访问存储在数据库中的数据,而是通过专门的软件——数据库管理系统(Database Management System)来实现对数据的所有操作。这个软件实现了大量的功能,满足了在文件系统管理中未能实现的各种用户需求,也解决了在文件系统管理中暴露出的各种问题。

对于不同的数学模型,都有不同类型的数据库管理软件来实现对它的访问。随着数据量的增加,数据应用范围的扩大,出现了将数据分散管理的方法,即把一个数据库分成多个,建立在多台计算机上,每台机器上都有对应的管理软件分开管理,互相之间不存在数据通信联系,这时称为分散式数据管理。而后,随着网络通信的发展,分散在各处的数据库系统又实现了网络连接,由此形成了分布式数据库管理,对应的数据库管理软件则为分布式数据库管理系统。

由此可见,从文件系统到数据库系统是数据管理技术的一次飞跃。当然,随着应用的不断推广,硬件设备的进一步发展,新的业务需求的驱动以及管理理论新成果涌现,数据管理理论、技术和方法还将向新的目标继续蓬勃发展。

1.3 数据的抽象与描述

计算机能处理的问题只局限在机器世界,因此,当我们要解决现实世界的问题时,首先要做的事情就是把现实世界的问题先映射到信息世界可以表达的问题,然后再考虑对应信息世界的问题如何转换为机器世界可以理解的问题,这样计算机才有可能在机器世界去解决这个问题。

而当我们借助计算机处理各行各业涉及大量数据的管理类问题时,首先要解决的就是这些问题中核心数据的抽象和描述。数据的抽象和描述就是从用户需求中提取需要借助于计算机处理的数据,并将这些数据借助数学模型描述、表示出来,便于计算机实现处理。选做数据描述的模型应满足三方面的要求:一是能比较真实地反映现实世界;二是容易被人所理解;三是便于在计算机上实现。

而要同时找到满足这三个要求的模型还是比较困难的,因此,在数据库的应用中,经常采用以下的数据抽象过程,通过三层模型的组合效应来完成需求数据的抽象和描述。

1.3.1 数据抽象的过程

在用数据库解决数据的管理与应用时,首先从用户需求的角度,抽象并综合有关一个问题的各个不同用户需求,并用概念模型将该综合需求涉及数据的全局逻辑结构表达出来(见图1-4)。