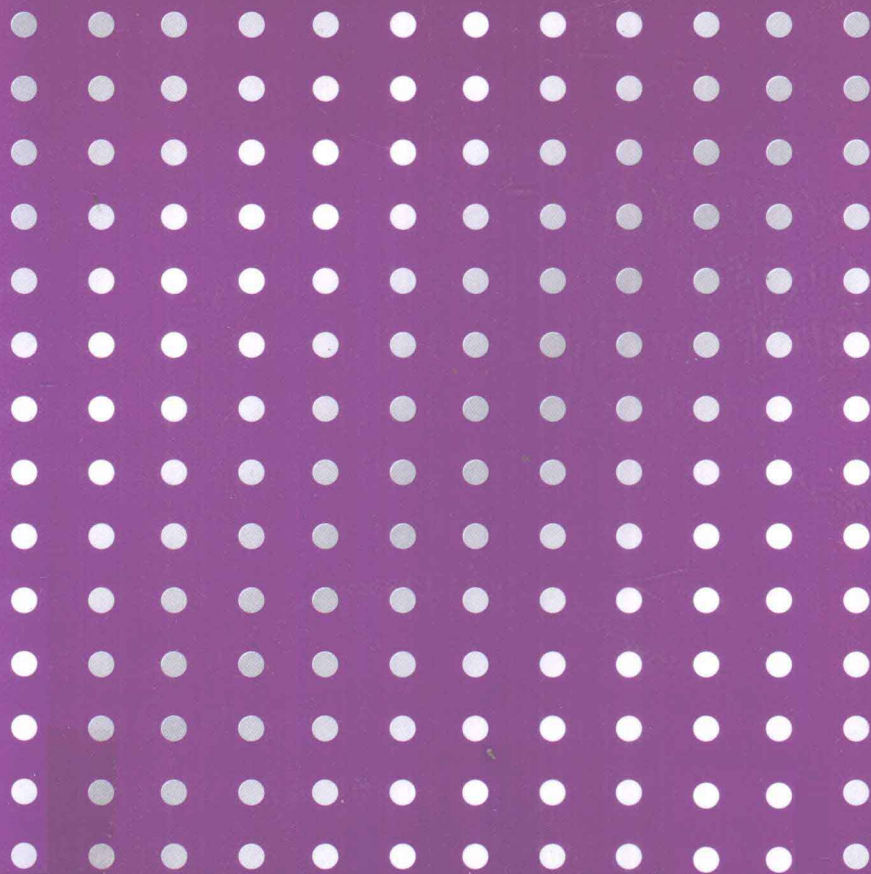


高等院校信息技术规划教材

数据库技术及应用

尹为民 曾 慧 吴迪倩 编著

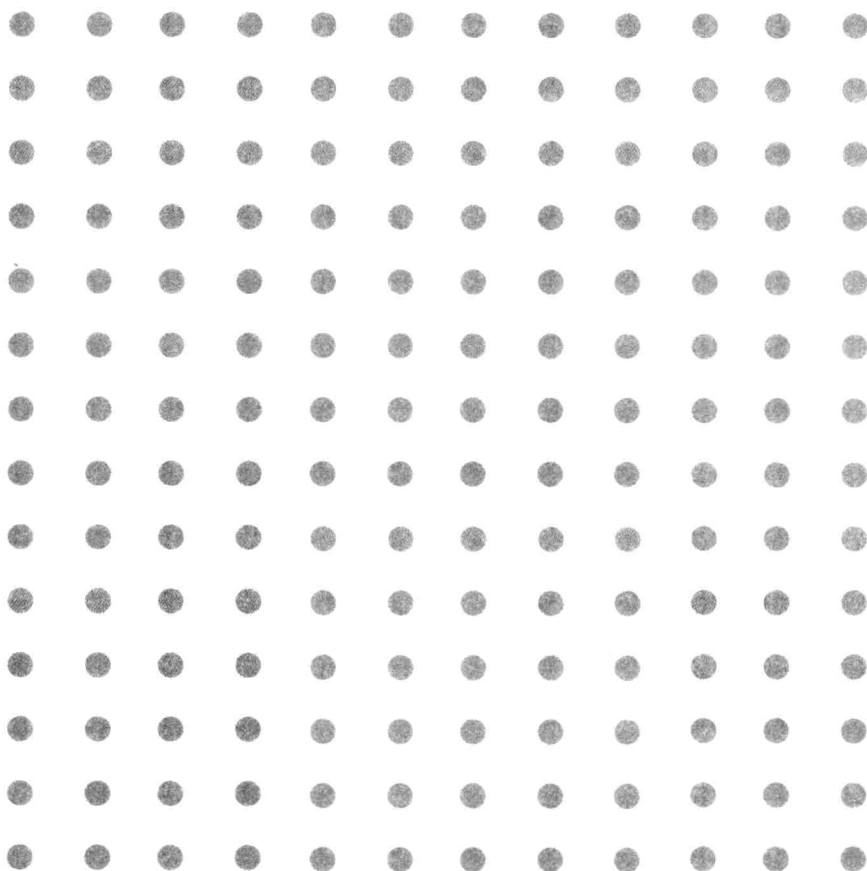


清华大学出版社

高等院校信息技术规划教材

数据库技术及应用

尹为民 曾 慧 吴迪倩 编著



清华大学出版社
北京

内 容 简 介

《数据库技术及应用》是一部关于数据库系统的基本原理与方法、现代技术及应用的教学书。其中分为三个部分：第1部分为基础篇，介绍数据库基础知识与基本方法、描述关系数据库语言及关系模式的规范化设计；第2部分为技术篇，介绍关系数据库设计方法、数据库系统的调优技术、描述系统建模的新方法与对象关系数据库技术；第3部分为应用篇，介绍数据库访问技术与信息系统的结构模式、结合典型的数据库管理系统介绍数据库应用系统的设计与开发、讨论现代数据库技术。本书最后附有数据库设计实例。

本书是作者多年从事数据库课程教学与科研实践的结晶，注重核心知识的描述，注意基础理论和现行应用技术的结合，删减繁冗、过时的内容，同时又根据本学科发展的动态，介绍现代数据库系统技术及相关应用知识。本书有相应的课程网站(含配套的课件、练习测试和实习指导等)，以方便师生的教与学。

本书不仅可作为高等院校计算机及相关专业学生数据库课程的教材，还可供从事信息领域工作的科技人员及其他人员参阅。

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

数据库技术及应用/尹为民，曾慧，吴迪倩编著. —北京：清华大学出版社，2011.7

(高等院校信息技术规划教材)

ISBN 978-7-302-25506-2

I. ①数… II. ①尹… ②曾… ③吴… 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(2011)第 085384 号

责任编辑：白立军 薛 阳

责任校对：焦丽丽

责任印制：杨 艳

出版发行：清华大学出版社

地 址：北京清华大学学研大厦 A 座

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

邮 编：100084

社 总 机：010-62770175

邮 购：010-62786544

投稿与读者服务：010-62795954, jsjje@tup.tsinghua.edu.cn

质 量 反 馈：010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者：北京嘉实印刷有限公司

经 销：全国新华书店

开 本：185×260

印 张：22.75

字 数：527 千字

版 次：2011 年 7 月第 1 版

印 次：2011 年 7 月第 1 次印刷

印 数：1~3000

定 价：35.00 元

产品编号：036936-01

前言

foreword

《数据库技术及应用》是在已有数据库类教材基础上综合、更新、编著而成的,其目的是使该教材更符合当前教学与应用的需要。本教材的编写原则是:根据本学科的发展动态,注重数据库基础理论与应用的结合,介绍数据库系统设计的实用技术,补充数据库系统的新技术及新方法,以适应数据库技术及其应用的快速发展。

从本书的组织结构来看,全书内容分三个部分共 10 章:

第 1 部分是关于数据库系统的基础知识,包括第 1~3 章,主要介绍关系数据库系统的基本概念、关系模型、关系数据库语言以及关系模式的规范化设计。

第 2 部分介绍数据库设计与数据库系统管理技术,包括第 4~6 章,主要介绍数据库设计、数据库系统调优方法,数据库安全、完整性控制与事务管理技术,描述数据库系统建模的 UML(统一建模语言)方法与对象关系数据库技术。

第 3 部分关于数据库系统应用技术,包括第 7~10 章,主要介绍数据库访问技术与信息系统的结构模式,讲解具体数据库应用系统的设计与实现,介绍现代数据库系统的特点及相关技术。

这是一部关于现代数据库系统的基本方法、技术和应用的教科书,具有以下特点:

(1) 本教材根据《高等学校计算机科学与技术专业公共核心知识体系与课程》中数据库系统部分的教学大纲与知识单元编写而成,参照了《中国高等学校计算机科学与技术专业(应用型)学科教程》纲要,使之具有符合技术应用与发展要求的知识结构。

(2) 本教材简练地介绍数据库的基本技术和方法,补充新型数据库系统的主要技术及知识,强调数据库系统的主流技术,突出数据库技术与实践紧密结合的特征,结合现代的软、硬件环境及应用实例讲解。

(3) 有丰富的教学辅助资源,如有配套课程网站(含配套的课件、练习测试和实习指导等)以方便教学。动态更新的课件可从武

汉大学计算机学院的精品课程网站上获取(http://webclass.whu.edu.cn/database/db_web/index.aspx)。

本书力求用勤勉的工作、自己的思考诠释专业规范的思想,把新的课程体系和教学内容生动地表述出来,着眼于帮助学生既掌握数据库系统的基本技术和方法,又了解现代数据库系统应用的特点及技术发展趋势。

曾慧编著了本书的第3、4、8章及附录的主要内容,尹为民编著了本书的其余章节并负责全书的统稿,参加编写的有华师一附中的吴迪倩、武汉大学石磊及黄倩同学,他们参与编写了本书的第8章、第9章中的数据库应用系统设计示例及附录部分的实例,并涉及程序调试、英语翻译、图表制作等工作。

本书存在的不足或疏漏之处,恳请广大读者批评指正。

作 者

2011年2月于武汉大学

目录

Contents

第 1 部分 基 础 篇

第 1 章 数据库系统导论	3
1.1 数据库与信息系统	3
1.1.1 数据管理的发展	3
1.1.2 数据库基本概念	5
1.2 数据模型与信息模型	8
1.2.1 三个世界及其联系	8
1.2.2 信息模型的表示	10
1.2.3 数据模型的组成	12
1.2.4 基本数据模型	13
1.3 数据库系统	17
1.3.1 数据库系统的结构	17
1.3.2 数据库系统的组成	20
1.3.3 数据库系统的特点	21
1.4 本章小结	22
习题 1	23
第 2 章 关系数据库及其语言	24
2.1 关系模型	24
2.1.1 关系模型的特点	24
2.1.2 关系数据结构	25
2.1.3 关系数据操作	28
2.1.4 关系的完整性	29
2.2 关系代数	31
2.2.1 集合运算	32
2.2.2 关系运算	33

2.2.3 扩充的关系运算	37
2.3 关系演算	40
2.3.1 元组关系演算	40
2.3.2 域关系演算	41
2.4 数据库标准语言 SQL	42
2.4.1 SQL 的特征	42
2.4.2 SQL 的数据定义	46
2.4.3 SQL 的数据查询	50
2.4.4 SQL 的数据更新	61
2.4.5 SQL 中的视图	63
2.4.6 过程化 SQL	68
2.5 本章小结	68
习题 2	69
第 3 章 关系模式的规范化设计	72
3.1 函数依赖	72
3.1.1 关系模式的设计问题	72
3.1.2 函数依赖的概念	74
3.1.3 候选键的形式定义	75
3.2 关系模式的规范化	76
3.2.1 关系与范式	76
3.2.2 第 1 范式(1NF)	77
3.2.3 第 2 范式(2NF)	78
3.2.4 第 3 范式(3NF)	79
3.2.5 BC 范式(BCNF)	81
3.2.6 多值依赖与第 4 范式	81
3.3 数据依赖公理	84
3.3.1 公理及其推论	84
3.3.2 闭包的概念及其计算	85
3.3.3 函数依赖集的等价	88
3.3.4 最小函数依赖集	89
3.4 关系模式的分解	91
3.4.1 等价模式分解的定义	91
3.4.2 无损连接性与依赖保持性	92
3.4.3 模式分解的算法	95
3.5 本章小结	96
习题 3	97

第 2 部分 技 术 篇

第 4 章 数据库设计	101
4.1 数据库设计概述	101
4.1.1 数据库设计的特点和方法	101
4.1.2 数据库设计阶段及内容	103
4.2 需求分析	105
4.2.1 需求分析的任务	105
4.2.2 需求分析的方法	106
4.2.3 需求分析的工具	107
4.3 概念结构设计	110
4.3.1 概念结构设计的方法	110
4.3.2 E-R 图及其扩展描述	112
4.3.3 局部视图设计	115
4.3.4 合并 E-R 图	116
4.3.5 消除数据冗余	117
4.4 逻辑结构设计	118
4.4.1 逻辑结构设计的步骤	118
4.4.2 E-R 图向关系模式的转换	119
4.4.3 关系模式的优化	122
4.4.4 用户子模式的设计	123
4.5 数据库物理设计	124
4.5.1 物理设计的内容和要求	124
4.5.2 存取方法与存储结构	124
4.5.3 性能评价	126
4.6 数据库实施与维护	127
4.7 本章小结	128
习题 4	128
第 5 章 数据库系统管理技术	130
5.1 DBMS 的组成与功能	130
5.1.1 DBMS 软件及其组成	130
5.1.2 DBMS 的主要功能	131
5.1.3 DBMS 的工作过程	132
5.2 查询优化与系统调优	133
5.2.1 关系数据库的查询优化	133

5.2.2 数据库系统调优	135
5.3 数据库的安全性	141
5.3.1 安全性的基本概念	141
5.3.2 安全性控制技术	142
5.3.3 自主存取控制的实施	144
5.3.4 Web 数据库安全策略	148
5.4 数据库的完整性	149
5.4.1 完整性控制的功能	149
5.4.2 完整性约束的设计	150
5.4.3 数据库触发器设计	153
5.5 事务管理技术	157
5.5.1 事务的基本概念	157
5.5.2 数据库恢复技术	158
5.5.3 故障的恢复策略	163
5.5.4 并发控制技术	165
5.6 本章小结	170
习题 5	171
第 6 章 对象关系数据库技术	173
6.1 面向对象的数据模型	173
6.1.1 基于对象的数据库概述	173
6.1.2 面向对象模型的概念	174
6.1.3 模型的结构与特征	177
6.1.4 面向对象数据库系统	179
6.2 面向对象的建模方法	180
6.2.1 统一建模语言——UML	180
6.2.2 UML 的表示法	181
6.2.3 UML 的信息建模	182
6.2.4 UML 的高级建模	184
6.2.5 信息模型与 UML 表示	186
6.3 对象关系数据库系统	190
6.3.1 对象关系数据库系统概述	190
6.3.2 对象关系数据模型	191
6.3.3 扩充的数据类型	192
6.3.4 常用的复杂类型	193
6.3.5 继承和引用	196
6.3.6 数据的查询	199
6.3.7 自定义函数与过程	202

6.3.8 对象关系数据库应用	205
6.4 本章小结	207
习题 6	208

第 3 部分 应 用 篇

第 7 章 数据库访问技术	211
7.1 数据库访问概述	211
7.2 数据库访问技术之一——嵌入式 SQL	211
7.2.1 嵌入式 SQL 概述	211
7.2.2 嵌入式 SQL 方法	212
7.2.3 嵌入式 SQL 程序	214
7.2.4 动态 SQL 方法	216
7.3 数据库访问技术之二——数据库应用接口	217
7.3.1 开放式接口——ODBC	218
7.3.2 调用级接口——JDBC	220
7.3.3 通用级接口——ADO.NET	225
7.4 数据库应用知识	229
7.4.1 数据库应用体系结构	229
7.4.2 信息系统的结构模式	232
7.4.3 数据库应用开发步骤	234
7.4.4 数据库产品及开发工具	235
7.5 本章小结	238
习题 7	238
第 8 章 SQL Server 数据库应用	240
8.1 SQL Server 2005 系统概述	240
8.1.1 SQL Server 2005 的组成	240
8.1.2 SQL Server 2005 的组件	241
8.1.3 SQL Server 2005 的工具	242
8.2 数据库创建与程序设计	245
8.2.1 数据库的创建	245
8.2.2 表的创建	247
8.2.3 T-SQL 程序基础	250
8.2.4 函数和游标	253
8.2.5 存储过程	259
8.2.6 数据库触发器	265

8.2.7 数据库安全性	272
8.3 数据库应用系统设计	272
8.3.1 系统需求与设计目标	272
8.3.2 系统功能模块设计	273
8.3.3 数据库设计与表的创建	274
8.3.4 系统界面的制作与实现	275
8.3.5 系统应用程序设计	277
8.4 本章小结	285
习题8	286
第9章 Oracle 数据库应用	287
9.1 Oracle 产品及功能介绍	287
9.1.1 Oracle 产品及特点	287
9.1.2 Oracle 数据库与安全性	288
9.2 PL/SQL 程序设计	291
9.2.1 程序的基本结构	291
9.2.2 存储过程	293
9.2.3 游标的设计	296
9.3 数据库应用系统设计	299
9.3.1 系统需求与设计目标	299
9.3.2 系统功能模块设计	299
9.3.3 系统实现方式及开发环境	300
9.3.4 数据库设计与表的创建	302
9.3.5 系统应用程序设计	306
9.3.6 系统测试运行图	312
9.4 本章小结	313
习题9	314
第10章 现代数据库技术	315
10.1 现代数据库技术概述	315
10.1.1 数据库技术的发展	315
10.1.2 现代数据库技术研究	316
10.2 XML 数据库	319
10.2.1 XML 数据库概述	319
10.2.2 XML 数据模型	320
10.2.3 XML 查询语言	322
10.2.4 实现的数据库特性	324

10.3	数据仓库与数据挖掘	326
10.3.1	数据仓库的特征与组成	327
10.3.2	数据仓库与数据集市	329
10.3.3	联机分析处理	331
10.3.4	多维数据的表示和操作	332
10.3.5	数据挖掘技术	334
10.4	本章小结	336
习题 10		337
附录 A	录像出租公司数据库设计实例	338
A.1	应用需求	338
A.2	需求分析	339
A.2.1	确定任务与目标	339
A.2.2	确定系统边界与标识用户	339
A.2.3	收集用户与系统的详细信息	340
A.2.4	管理用户与数据关系	340
A.2.5	确定分公司的用户需求	341
A.2.6	确定系统需求	342
A.3	概念结构设计	344
A.3.1	建立局部概念结构	344
A.3.2	建立全局概念结构	346
A.4	逻辑结构设计	348
A.5	物理结构设计	349
参考文献		352

第 1 部分

基 础 篇

数据库系统导论

数据是人类活动的重要资源。数据管理技术的优劣,直接影响数据处理效率。数据库技术正是瞄准这一目标,发展起来的专业技术,它主要研究如何存储、使用和管理数据。数据库系统就是研究如何高效地存取和科学地管理数据的计算机系统。

本章主要描述数据库技术的基本概念、基本数据模型及信息模型的表示,介绍数据库系统的结构、组成及特点。

1.1 数据库与信息系统

数据库技术是计算机科学技术中发展最快的领域之一,它的出现使得计算机应用渗透到工农业生产、商业管理、科学研究、工程技术以及国防军事等各个领域。安全可靠的数据仓库已成为计算机信息管理及其应用系统的重要基石。现在,数据库系统的建设规模,数据库描述信息量的大小以及网络应用的程度已成为衡量一个部门信息化程度的重要标志。

那么,数据库管理技术是怎样发展而来的?什么是数据库?什么是数据库管理系统?什么是数据库系统?数据库与信息系统的关系是什么?这就是本节要介绍的内容。

初学者可以把数据库想象成存放数据的仓库,那么数据库管理系统(Data Base Management System,DBMS)这个软件就是仓库保管员,数据库系统就对应整个仓库部门(包括仓库保管员、所有的货物及其存储管理设施),而信息系统就是建立在仓库之上的数据加工厂。

1.1.1 数据管理的发展

数据管理是研究如何对数据分类、组织、编码、存储、检索和维护的一门技术,是数据处理的核心。数据库技术是应数据管理的需要而产生的,而数据管理又是随着计算机技术的发展而完善的。数据管理经历了人工管理、文件系统管理、数据库系统管理3个阶段。

1. 人工管理

人工管理阶段是计算机数据管理的初级阶段。当时计算机主要用于科学计算,数据量少、不能保存。由于当时没有专门的软件对数据进行管理,程序员在设计程序时不仅

要规定数据的逻辑结构,而且还要设计其物理结构,即数据的存储地址、存取方法、输入输出方式等,这样使得程序与数据相互依赖、密切相关,一旦数据的存储地址、存储方式稍有改变,就必须修改相应的程序。此外,数据面向应用,多个应用涉及的数据相同时,由于用户各自定义自己的数据,无法共享,因此存在大量的数据冗余。

人工管理阶段的主要问题是:数据不能长期保存;数据不能共享、冗余度极大;数据独立性差(即数据与程序之间相互依赖性强)。

2. 文件系统管理

20 世纪 50 年代末,计算机不仅用于科学计算,而且大量用于数据管理,同时磁盘、磁鼓等大容量直接存储设备的出现,可以用来存放大量数据。操作系统中的文件系统就是专门用来管理所存储数据的软件。

这一阶段数据管理的特点有:数据可以长期保存;对文件进行统一管理,实现了按名存取,文件的逻辑结构与物理结构分开,数据在存储器上的物理位置、存储方式等的改变不会影响用户程序;文件系统实现了一定程度的共享(文件部分相同,则难以共享);文件是为某一特定应用服务的,难以在已有数据上增加新的应用、不易扩充。一旦数据的逻辑结构改变,必须修改文件结构的定义,修改应用程序。文件之间相对独立、有较多的数据冗余、应用设计复杂、使用不便。

文件系统管理的主要问题是:逻辑数据独立性较差,数据冗余度较大,文件应用编程复杂。

3. 数据库系统管理

随着数据管理的规模日趋增大,数据量急剧增加,文件系统管理已不能适应需求。20 世纪 60 年代末数据库技术应运而生。数据库系统对数据的管理方式与文件系统不同,它把所有应用程序中使用的数据汇集起来,按一定结构组织集成,在 DBMS 软件的统一监督和管理下使用,多个用户、多种应用可充分共享。数据库管理技术的出现为用户提供了更广泛的数据共享和更高的数据独立性,并为用户提供了方便的操作使用接口。现代数据库管理正向着综合集成、智能化的数据库服务系统时代迈进!

数据管理经历的各个阶段有自己的背景及特点,数据管理技术也在发展中不断地完善,其 3 个阶段的比较如表 1-1 所示。

表 1-1 数据管理 3 个阶段的比较

数据管理的 3 个阶段		人工管理 (20 世纪 50 年代中期)	文件系统(20 世纪 50 年代末—60 年代中期)	数据库系统 (20 世纪 60 年代后期至今)
背 景	应用背景	科学计算	科学计算、管理	大规模数据、分布数据的管理
	硬件背景	无直接存取存储设备	磁带、磁盘、磁鼓	大容量磁盘、可擦写光盘、按需增容磁带机等
	软件背景	无专门管理的软件	利用操作系统的文件系统	由 DBMS 支撑
	数据处理方式	批处理	联机实时处理、批处理	联机实时处理、批处理、分布处理

续表

数据管理的 3个阶段		人工管理 (20世纪50年代中期)	文件系统(20世纪 50年代末—60年代中期)	数据库系统 (20世纪60年代后期至今)
特 点	数据的管理者	用户/程序管理	文件系统代理	由DBMS管理
	数据应用的 扩充	面向某一应用程序、 难以扩充	面向某一应用系统、不易 扩充	面向多种应用系统、容易 扩充
	数据的共享性	无共享、冗余度极大	共享性较差、冗余度大	共享性好、冗余度小
	数据的独立性	数据的独立性差	物理独立性好,逻辑独立 性差	具有高度的物理独立性、具 有一定的逻辑独立性
	数据的结构化	数据无结构	记录内有结构、整体无 结构	统一数据模型、整体结构化
	数据控制能力	应用程序自己控制	由OS提供基本存取控制	由DBMS提供数据的安全 性、完整性、并发控制及恢 复功能

1.1.2 数据库基本概念

1. 数据与信息

现代社会是信息的社会,信息以惊人的速度增长。因此,如何有效地组织和利用它们成为急需解决的问题。引入数据库技术的目的就是为了高效地管理及共享大量的信息,而信息与数据是分不开的。

数据是描述事物的符号记录,也是数据库中存储、用户操纵的基本对象。数据不仅是数值,还可以是文字、图形、图像、声音、视频等。数据是信息的符号表示。例如描述2011年武汉大学计划招生信息,可用一组数据“2011年,92个专业,招生7900人”。这些符号被赋予了特定的语义,具体描述了信息,因此它们就具有了传递信息的功能。

信息是有一定含义的,经过加工处理的,对决策有价值的数据。例如:气象局每天从各地收集到大量有关气象的图形或文字记录,包括各地温度、湿度、风力、阴晴等的具体描述,这些就是表示各地气象信息的数据。当气象局对这些数据进行综合处理、分析、判断,做出气象预报时,人们可以根据气象预报安排生产和生活。这些处理过的数据又表示了对决策或行动有价值的新的信息。因此,信息是对现实世界中存在的客观实体、现象、联系进行描述的有特定语义的数据,它是人类共享的一切知识及客观加工提炼出的各种消息的总和。

信息与数据的关系可以归纳为:数据是信息的载体,信息是数据的内涵。即数据是信息的符号表示,而信息通过数据描述,又是数据语义的解释。

2. 数据库

数据库(Data Base,DB)是按一定结构组织并长期存储在计算机内的、可共享的大量