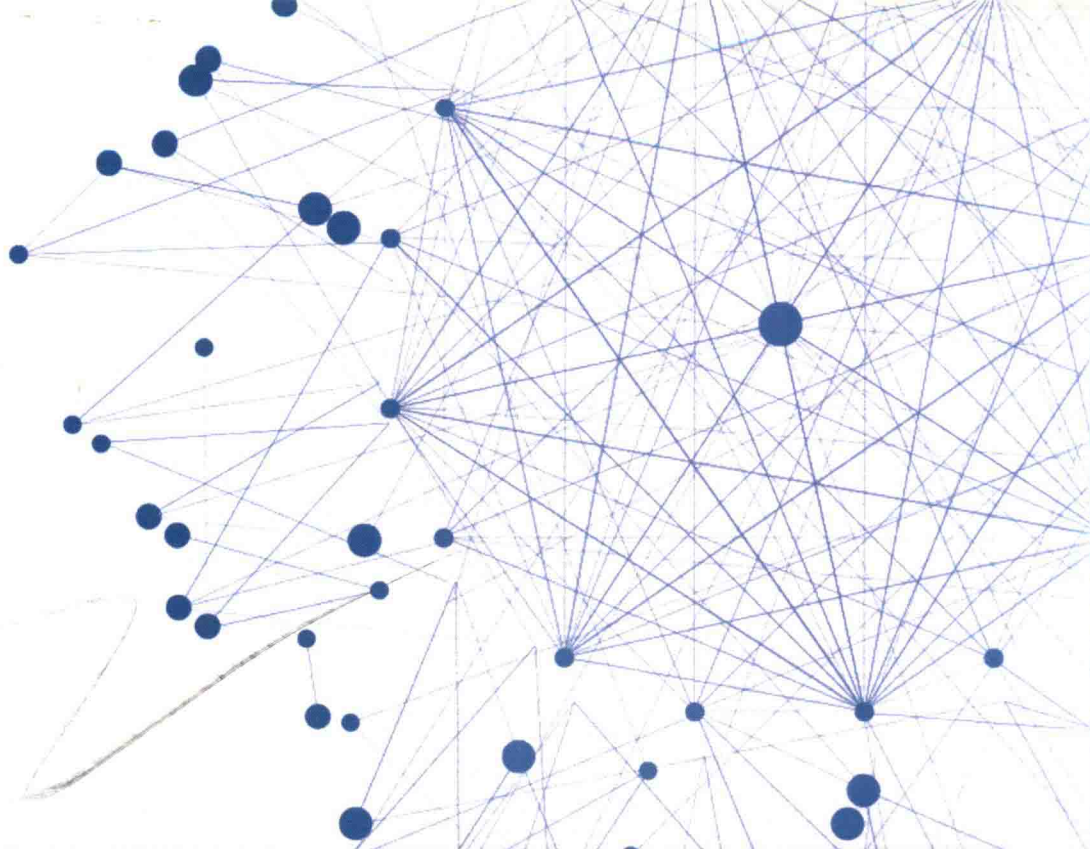




数据库系统原理

The Basic Theory of Database Systems
(第二版)

王 勋 韩培友 编著



浙江工商大学出版社
ZHEJIANG GONGSHANG UNIVERSITY PRESS

数据库系统原理

(第二版)

王 勋 韩培友 编著



浙江工商大学出版社
ZHEJIANG GONGSHANG UNIVERSITY PRESS

图书在版编目(CIP)数据

数据库系统原理 / 王勋, 韩培友编著. —2 版. —
杭州: 浙江工商大学出版社, 2018. 8
ISBN 978-7-5178-2636-1

I. ①数… II. ①王… ②韩… III. ①数据库系统
IV. ①TP311.13

中国版本图书馆 CIP 数据核字(2018)第 052709 号

数据库系统原理(第二版)

王 勋 韩培友 编著

责任编辑 吴岳婷
封面设计 林朦朦
责任印制 包建辉
出版发行 浙江工商大学出版社
(杭州市教工路 198 号 邮政编码 310012)
(E-mail: zjgsupress@163.com)
(网址: <http://www.zjgsupress.com>)
电话: 0571-88904980, 88831806(传真)

排 版 杭州朝曦图文设计有限公司
印 刷 虎彩印艺股份有限公司
开 本 787mm×960mm 1/16
印 张 18.5
字 数 343 千
版 次 2018 年 8 月第 1 版 2018 年 8 月第 1 次印刷
书 号 ISBN 978-7-5178-2636-1
定 价 49.00 元

版权所有 翻印必究 印装差错 负责调换

浙江工商大学出版社营销部邮购电话 0571-88904970

内 容 简 介

本书作为数据科学的入门课程,全书以实例分析为主线,结合 SQL Server 2016,系统地阐述了数据库系统的基本原理、基本技术、基本方法及其应用。全书共 11 章,主要内容包括:概述、关系代数、数据完整性、SQL Server、概念模型和逻辑模型、关系规范化、数据安全、数据并发、数据恢复、数据库设计、数据仓库和大数据等。

本书以 SQL Server 2016 为 DBMS 介绍数据库技术的实现技术,以 Python 和 Java 等为主语言介绍数据库系统的设计与实现技术,并提供详细操作和完整程序系统。

本书内容丰富、深入浅出、通俗易懂、结构严谨、注重实用,同时提供取材新颖、实用的例题和习题,便于读者巩固所学知识。

本书适用于高等院校计算机科学与技术、数据科学与大数据技术、信息安全、软件工程、网络工程、信息管理与信息系统和电子商务等相关本科专业的数据库原理课程教材,也可作为从事数据库系统研究与应用开发的工程技术人员的参考书。

前 言

在大数据时代,数据库技术已经成为计算机科学和信息科学的重要组成部分,为了适应基于数据库系统的数据在线事务处理 OLTP 和基于数据库仓库与数据挖掘的大数据在线分析处理 OLAP 的新技术和新应用,从而确保本书的科学性和先进性,进而做到易用和实用,因此需要对本书进行修订。

本书的第一版于 2010 年出版,经过多年的使用,在得到认可的同时也发现了一些问题,所以对本书做了如下内容调整:

(1)以网上书店(授课范例)、商品销售(课内实验范例)和学籍管理(课外实验范例)等多个具体应用系统的研发过程为主线,结合实际应用过程循序渐进地详细阐述数据库系统的原理、技术与方法;

(2)调整了第一版内容的顺序,使之更易于学习数据处理与分析技术;

(3)删除了第一版第 5 章中实际利用率较低的较难内容;

(4)增加了大数据的基本内容。

SQL Server 是微软的专业级数据库管理软件,其数据库管理功能完善,操作简单且安全稳定,拥有很高的市场占有率。

Python 是一种最新的动态数据类型,解释性、互动性和面向对象的高级可视化程序设计语言。通过 Python 提供的丰富模块,使得其在网络编程、游戏、系统编程、图形处理、多媒体应用、数据库编程、数学处理、文本处理和数据可视分析等多个方面得到了广泛的成功应用。

Java 是 Oracle(Sun)的应用最广的可视化编程工具,是基于面向对象程序设计的集成开发环境,具有执行速度快和代码效率高等特点。

修订的新书是面向高等院校计算机、软件工程、计算机网络、信息管理与信息系统和电子商务等相关专业的学生学习数据库技术的一本教材。其特点是以实例分析为主线,结合实际用例阐述数据库系统的基本原理与方法,以 SQL Server 2016 为工具介绍数据库的实现技术,以 Python 和 Java 等为主语言介绍数据库系统的设计与实现。同时注重实际应用能力的培养,强调实践和动手;引入开源数据库,开阔学生视野;从而可以使读者系统、全面地学习数据库系统的基本理论和应用技术。

全书由 11 章组成,第 1 章介绍数据库技术的基本理论、数据模型,数据库系统及其系统结构等。第 2 章介绍关系数据语言、关系运算、查询处理和查询优化。第 3 章介绍数据完整性。第 4 章介绍 SQL Server 的基本操作、建立数据库和数据查询。第 5 章介绍概念模型和逻辑模型。第 6 章介绍关系模式的规范化理论和方法。第 7 章介绍数据安全模型和数据安全控制机制。第 8 章介绍事务管理、并发控制和封锁机制等。第 9 章介绍故障管理和数据恢复机制。第 10 章分析数据库设计的方法与步骤、数据库引擎、SQL Server 程序设计、游标、存储过程、函数和应用程序设计等。第 11 章介绍 XML 数据库、数据仓库、数据挖掘和大数据等数据库新技术。

为了便于读者使用本书,作者为本书制作了电子课件,需要者可以登录浙江工商大学出版社网站(<http://www.zjgsupress.com>)下载。

本书是作者多年从事数据库技术教育和科研的经验和总结。第 1~5 章由浙江工商大学计算机与信息工程学院王勋教授编写;第 6~11 章由浙江工商大学计算机与信息工程学院韩培友副教授编写。

全书由王勋教授负责整体结构设计、内容安排和全部审校工作,并最后审定。

本书获浙江省重点规划教材基金资助,为本书的出版提供了经费支持。教材得以正式出版,并以较快的速度与读者见面,这与参与编写的老师及出版社编辑的辛勤工作是分不开的,在此表示诚挚的感谢!

鉴于作者水平有限,错误与不妥之处在所难免,敬请同行专家与广大读者不吝指正。

王 勋 韩培友

2018 年 6 月 杭州

目 录

第 1 章 概述	001
1.1 数据库实例	001
1.2 基本知识	005
1.3 数据模型	011
1.4 数据库系统的结构	022
1.5 小结	027
习题	028
第 2 章 关系运算	030
2.1 关系数据语言	030
2.2 基本集合运算	032
2.3 专用关系运算	036
2.4 更新运算	046
2.5 综合实例	047
2.6 查询优化	048
2.7 小结	055
习题	056
第 3 章 数据完整性	060
3.1 实体完整性	061
3.2 参照完整性	061
3.3 用户定义完整性	063
3.4 完整性控制机制	064
3.5 数据完整性实现	067
3.6 小结	076

数据库实验	077
习题	078
第4章 SQL Server	081
4.1 结构查询语言 SQL	081
4.2 创建数据库	084
4.3 数据查询	090
4.4 小结	108
查询实验	109
习题	111
第5章 概念模型和逻辑模型	113
5.1 概念模型	113
5.2 逻辑模型	120
5.3 小结	126
习题	126
第6章 关系规范化	129
6.1 函数依赖	129
6.2 范式	132
6.3 关系模式规范化	143
6.4 小结	157
习题	158
第7章 数据安全	160
7.1 数据安全模型	160
7.2 数据安全控制	161
7.3 小结	179
安全实验	180
习题	182

第 8 章 并发控制	183
8.1 事务管理	183
8.2 并发控制	186
8.3 封锁	189
8.4 封锁协议	191
8.5 并发事务的可串行化	194
8.6 活锁和死锁	197
8.7 小结	199
并发控制实验	200
习题	203
第 9 章 数据恢复	204
9.1 故障管理	204
9.2 建立冗余数据	205
9.3 恢复技术	208
9.4 检查点机制	210
9.5 小结	212
数据恢复实验	213
习题	213
第 10 章 数据库设计	215
10.1 数据库设计的方法与步骤	215
10.2 需求分析	217
10.3 概念结构设计	221
10.4 逻辑结构设计	226
10.5 物理结构设计	228
10.6 数据库实施	232
10.7 系统运行与维护	251
10.8 小结	252
数据库设计实验	253
习题	254

第 11 章 数据仓库与大数据	256
11.1 XML	256
11.2 数据仓库	263
11.3 数据挖掘	273
11.4 大数据	280
11.5 小结	283
习题	283
参考文献	285



第1章

Chapter 1

概 述

现在的的生活和工作已经进入云数据和大数据时代,面对不断飞速增长的海量数据,不但需要利用数据库系统对数据进行数据处理,而且需要利用数据仓库及其挖掘技术对大数据进行数据分析,从而得到有价值的信息,而数据库技术是数据处理与分析的最有效手段。

数据库技术是通过研究数据库的结构、存储、设计、管理以及应用的基本理论和实现方法,并利用这些理论来实现对数据库中的数据进行处理、分析和理解的技术。即:

数据库技术是研究、管理和应用数据库(数据仓库) 的一门软件科学。

1.1 数据库实例

数据库技术作为信息系统的核心技术,研究如何科学地组织和存储数据,高效地获取和处理数据,从而减少数据冗余、实现数据共享和确保数据安全,并利用应用系统最终实现对数据的处理、分析和理解。显然,数据库技术的核心内容是数据库设计。

1.1.1 学生选课数据库

西湖大学准备开发教务管理系统,需要设计学生选课数据库,数据要求包括学生基本信息,课程基本信息和学生选修某些课程获得的考试成绩。为此,需要设计学生 Student、课程 Course 和选课 Student Course3 张表来组织学生选课数据库。(如表 1.1 ~ 1.3 所示)

表 1.1 学生 Student

学号 SNo	姓名 SName	性别 SSex	年龄 SAge	电话 SPhone	照片 SPhoto	微信 MInfo
S00001	杨康	男	20	13666612316	Xn.jpg	Xn123456
S00002	穆念慈	女	19	13666612326	Pj.jpg	pj234567
S00003	欧阳克	男	25	13666612336	Ws.jpg	Ws345678
S00004	周伯通	男	22	13666612346	Wd.jpg	Wd456789

表 1.2 课程 Course

课程号 CNo	课程名 CName	先修课 CPNo	开课学期 Seme	学时 Peri	学分 Cred
C001	高等数学	Null	1	64	4
C002	数据库原理	C003	3	48	3
C003	Python 语言	C001	2	48	3

表 1.3 选课 Student Course

学号 SNo	课程号 CNo	成绩 Grade	学号 SNo	课程号 CNo	成绩 Grade
S00001	C001	99	S00002	C002	95
S00001	C002	98	S00003	C001	88
S00001	C003	96	S00003	C003	89
S00002	C001	98	S00004	C002	99

1.1.2 网上书店数据库

西湖大学为推动学生创新创业,由校学生会发起创建了以销售旧教材为主的学生网上折扣书店,需要开发网上书店销售系统,设计网上书店数据库,书店 EBook 需要管理电子图书、客户和出版社,以及客户购买图书的相关信息,因此,可以使用图书 Book、客户 Cust、出版社 Press 和购买 Buy 4 张表(如表 1.4 ~ 1.7 所示)来组织网上书店数据库。即:

图书 Book(书号 BNo,书名 BName,作者 Author,社号 PNo,版次 EditNo,定价 Price,进价 PPrice,售价 SPrice)。

客户 Cust(户号 CNo, 户名 CName, 性别 CSex, 生日 Birth, 电话 Phone, 婚否 Marry, 照片 Photo, 邮箱 Email)。

出版社 Press(社号 PNo, 社名 PName, 邮编 PCode, 社址 PAddr, 电话 Phone, 邮箱 Email, 网址 HPage)。

购买 Buy(户号 CNo, 书号 BNo, 购买日期 PDate)。

表 1.4 图书 Book

书号 BNo	书名 BName	作者 Author	社号 PNo	版次 EditNo	定价 Price	进价 PPrice	售价 SPrice
ISBN978-7-04-040664-1	数据库系统概论	王珊	ISBN978-7-04	5	39.6	20	25
ISBN978-7-302-33894-9	Java 程序设计教程	雍俊海	ISBN978-7-302	3	69	26	36
ISBN978-7-5612-2591-2	图像技术	韩培友	ISBN978-7-5612	1	36	12	16
ISBN978-7-5612-2123-1	IDL 可视化分析	韩培友	ISBN978-7-5612	1	45	19	22
ISBN978-7-5178-0167-2	Access 数据库应用	韩培友	ISBN978-7-81140	1	39.8	16	19
ISBN978-7-81140-582-8	MySQL 实验指导	韩培友	ISBN978-7-81140	1	25	10	15

表 1.5 客户 Cust

户号 CNo	户名 CName	性别 CSex	生日 Birth	电话 Phone	婚否 Marry	照片 Photo	邮箱 EMail
C001	郭靖	男	1966-06-01	13611122233	是	二进制数据 Hu. Jpg	123450@qq. com
C002	黄药师	男	1979-09-10	13911122233	是	NULL	543210@qq. com
C003	黄蓉	女	1988-10-06	13622233355	是	NULL	122220@qq. com
C004	洪七公	男	1999-06-16	13922233355	否	NULL	222220@qq. com
C005	梅超风	女	2002-07-11	13655566699	否	NULL	345220@qq. com
C006	欧阳锋	男	2006-08-08	13955566699	否	NULL	562220@qq. com

表 1.6 出版社 Press

社号 PNo	社名 PName	邮编 PCode	社址 PAddr	电话 Phone	邮箱 EMail	网址 HPage
ISBN978-7-04	高等教育出版社	100120	北京市西城区德外大街4号	(010) 58581118	Gjdzfwb@pub.hep.cn	www.hep.edu.cn
ISBN978-7-302	清华大学出版社	100084	清华大学学研大厦A座	(010) 62781733	e-sale@tup.tsinghua.edu.cn	www.tup.tsinghua.edu.cn
ISBN978-7-115	人民邮电出版社	100164	北京市丰台区成寿寺路11号	(010) 81055075	Youdianhr@gmail.com	www.ptpress.com.cn
ISBN978-7-5612	西北工业大学出版社	710072	西安市友谊西路127号	(029) 88493844	Fxb@nwpup.com	www.nwpup.com
ISBN978-7-81140	浙江工商大学出版社	310012	杭州市教工路198号	(0571) 88823703	61627291@qq.com	www.zjgsupress.com

表 1.7 购买 Buy

户号 CNo	书号 BNo	购买日期 PDate	户号 CNo	书号 BNo	购买日期 PDate
C001	ISBN978-7-04-040664-1	2014-10-26	C002	ISBN978-7-04-040664-1	2014-11-22
C001	ISBN978-7-302-33894-9	2014-06-16	C002	ISBN978-7-302-33894-9	2014-08-12
C001	ISBN978-7-5612-2591-2	2010-01-12	C003	ISBN978-7-5612-2591-2	2011-11-13
C001	ISBN978-7-5612-2123-1	2006-12-01	C005	ISBN978-7-5612-2123-1	2008-10-09
C001	ISBN978-7-5178-0167-2	2014-03-19	C006	ISBN978-7-5178-0167-2	2015-06-18
C001	ISBN978-7-81140-582-8	2012-09-15	C006	ISBN978-7-81140-582-8	2013-10-17

不难看出,网上书店数据库把数据分散在4张表中,那么,仅使用1张表是否可行?数据只能以表的形式存储吗?是否存在规范的数据结构来组织数据。数据库技术会给出正确答案。

思考:为什么使用4张表优于使用1张表?请从多个方面进行详细分析。

提示:从数据冗余、插入异常、修改异常和删除异常等方面分析。

结论:对于实际应用,设计数据库中表的个数,规范每个表的属性个数以及每个属性的特征要求、约束条件和依赖关系等,最终实现安全稳定合理的数据库系统。

1.2 基本知识

数据库技术的研究对象是数据,研究内容是通过数据的统一组织和管理,按照指定的数据结构建立相应的数据库;利用数据库管理系统实现对数据库的数据进行添加、修改、删除、查询和报表等多功能的应用系统;并利用数据库系统实现对数据的处理和分析。

1.2.1 数据

数据(Data) 是表示客观事物的符号。我们所使用的数据通常分为文本、图形、图像、音频、视频和动画等数据类型。其中文本又分为数值、字符、日期、时间和逻辑等数据类型。

例如:网上书店中图书的定价为 39.6 元;书名为数据库系统概论;生日为 1966 年 6 月 6 日(1966-06-06);婚否为是;照片为 Hu.jpg;起飞时间是 16 时 16 分 16 秒(16:16:16) 等。

数值:表示年龄、工资和单价等大小多少的数据。例如:16, - 10 和 26.6。

字符:表示人名、地名和物品等名称的数据。例如:李明、杭州和图书。

日期:表示工作日期、出生日期和入党日期等数据。例如:2016 年 6 月 6 日。

时间:表示上班时间和休息时间和开饭时间等数据。例如:10:12:19。

逻辑:表示是否结婚、是否党员等逻辑判断结果的数据。例如:真和假。

图形:表示直线、三角形和圆形等几何图形的数据。例如:正方形和五角星。

图像:表示人物照片、风景照片和遥感图像等数据。例如:猫的照片 Cat.jpg。

音频:表示语音、声效和音乐等数据。例如:一段录音和飞机起飞的声音。

视频:表示 MTV 和电影剪辑等数据。例如:精彩电影剪辑 Robot.mpg。

动画:表示计算机三维动画的数据。例如:Flash 动画 Fly fla 和 3D 动画 Audi.3ds。

通常把文本、图形、声音、图像、视频和动画等数据的集合,称为多媒体数据(Multimedia Data)。

总之,信息来源于数据,数据是信息的载体。信息是数据处理后有价值的数
据。原始数据需要进行数据处理,才能成为真正有价值的信息。

1.2.2 数据库

数据库(Data Base,DB) 是长期存储在计算机内,有组织可共享的大量数据的集合。即:存放数据的电子仓库。

例如:图书、客户、出版社和购买等组成了电子书店数据库 EBook。

数据库对所有的数据,按照指定的数据模型实行统一的组织、存储和管理。数据库独立于程序而存在,并提供给不同的用户共享。

数据库的特点是数据永久结构化存储、冗余度低,独立性高、可以共享和易于扩展等。

支持多媒体数据的数据库,称为多媒体数据库 (Multimedia Data Base, MDB)。

不难看出,数据库是数据的归宿。数据经过结构化处理,按照统一的数据结构存入数据库,进而实现数据共享。

1.2.3 数据库管理系统

数据库管理系统(Data Base Management System, DBMS) 是提供给用户,并帮助用户建立、使用和管理数据库的软件系统。即:管理数据的软件;用于维护数据库、接受和完成用户提出的访问数据库的请求。

例如:微软公司的 SQL Server 和 Access、甲骨文公司的 Oracle、IBM 公司的 DB2 和开源的 MySQL 等。

建立数据库的目的是使用数据库,并对数据库中的数据进行数据处理和分析,数据库管理系统是帮助达到这一目的的工具和手段。

数据库管理系统作为数据库系统的核心,建立在操作系统之上,位于操作系统与用户之间的一个数据管理软件,负责对数据库进行统一的管理和控制。它保证了数据的安全性和完整性,同时提供了多用户并发控制和数据恢复机制。

利用数据库管理系统可以科学地组织和存储数据、高效地获取和维护数据。数据库管理系统的主要功能:

(1) 数据定义

提供用于定义数据的数据定义语言(Data Definition Language, DDL), 利用 DDL 可以方便地定义数据库的数据对象及其关系。

(2) 数据操纵

提供用于操作数据的数据操纵语言(Data Manipulation Language, DML), 利用 DML 可以灵活方便的对数据库进行插入、修改、删除、查询和报表等操作。

(3) 事务和运行管理

提供用于实现对数据库的安全性、完整性、并发性和恢复性等进行保护控制的数据控制语言(Data Control Language, DCL), 以保证数据的安全和完整。

(4) 组织、存储和管理数据

实现对数据库的统一组织、存储和管理,确定数据库文件的存储结构和访问

方式,减少数据冗余,提高数据库的利用率。

(5) 数据库的建立和维护

数据定义、操纵、控制和存贮的具体实现;数据库的备份与重组重构、性能监视与分析等维护任务。

1.2.4 数据库系统

数据库系统(Data Base System,DBS)是在计算机系统中引入数据库后,由数据库、数据库管理系统、开发工具、应用系统、数据库管理员和用户等构成的完整系统。

用户使用数据库,需要使用由数据库设计员、程序员和管理员等利用数据库管理系统和开发工具研发的应用系统,对数据库的数据进行处理,进而得到所需要的信息。

数据库系统一般由硬件、软件和人员等3大部分组成。

(1) 硬件:计算机硬件和数据库专用硬件等。

计算机硬件是指计算机的CPU、内存、硬盘、交换机和路由器等;数据库专用硬件是指快速存取数据的磁盘阵列、磁带阵列或者光盘阵列、快速传输设备和数据备份设备等。

(2) 软件:操作系统(Operating System,OS)、DBMS、开发工具(程序设计语言和专用工具)和应用系统等。

(3) 人员:数据库设计员、程序员、数据库管理员和用户等。

数据库设计员负责系统的需求分析、性能分析、概念结构、逻辑结构和物理结构设计等。

程序员主要负责应用系统的设计、实现和测试。

数据库管理员(Data Base Administrator,DBA)是对数据库进行建立、使用和维护等的专职管理人员。DBA应该与数据库设计员、程序员和用户,共同参与数据库设计。

DBA的主要职责:

(1) 决定数据库的信息内容和结构。

根据存储的数据,对数据库的信息内容和结构作统一的结构化处理。

(2) 决定数据库的存储结构和存取策略。

确定数据在存储设备上的存储结构和存取策略,提高存取效率和存储空间利用率。

(3) 定义数据的安全性要求和完整性约束条件。

定义用户对数据库的存取权限、数据的保密级别和完整性约束条件,从而保