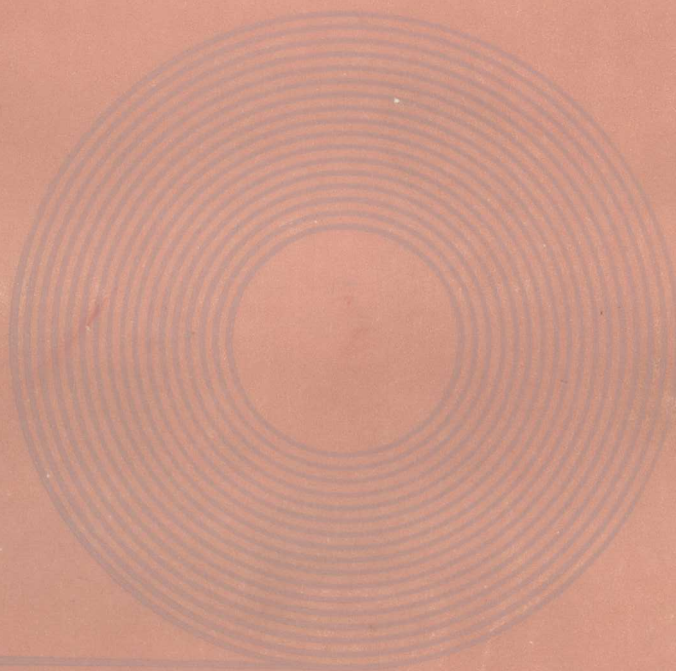


DATABASE SYSTEM

高等学校试用教材



数据库系统概论

王珊 主编

高等教育出版社



THE UNIVERSITY OF CHICAGO PRESS

100



THE UNIVERSITY OF CHICAGO PRESS

高等学校试用教材

数据库系统概论

萨师煊 王 珊

高等教育出版社

ISBN 7-04-011111-1

1994年10月第1版

内 容 提 要

本书重点介绍数据库系统的一般概念和基本理论。主要内容包 括：数据管理的进展，三种基本数据模型，数据库系统构成，DBTG系统，IMS系统，关系数据库系统及规范化理论，数据库的安全性，完整性与并发控制，数据库设计等。另外，在本书的附录中还附有文件组织和 IMAGE 数据库系统实例以供读者参考。

本书是按照 1983 年教育部计算机软件教材编审委员会审订的《数据库系统概论教学大纲》编写而成的。它可作为高等院校计算机有关专业“数据库”课程的试用教材或参考书，也可供从事计算机工作的科研人员、工程技术人员以及其他有关人员参考。

责任编辑 鲍 涌

高等学校试用教材 数据库系统概论

萨师煊 王珊

•
高等教育出版社出版
新华书店北京发行所发行
北京印刷二厂印装

•
开本 850×1168 1/32 印张 16.375 字数 396,000

1983年8月第1版 1988年8月第5次印刷

印数 34 551—102 550

ISBN 7-04-001611-7/TP·16

定价 3.35元

前 言

数据库是数据管理的最新技术,是计算机科学的重要分支。十余年来,数据库管理系统已从专用的应用程序包发展成为通用的系统软件。由于数据库具有数据结构化、最低冗余度、较高的程序与数据独立性、易于扩充,易于编制应用程序等优点,较大的信息系统都是建立在数据库设计之上的。因此,不仅大型计算机及中小型计算机,甚至微型机都配有数据库管理系统。目前,对数据库各种模型的研究以及理论上的探讨都还在蓬勃开展,其应用也从一般管理扩大到计算机辅助设计,人工智能以至于科技计算等领域。国外高等学校计算机科学系、计算机应用与信息系统系等都开设有数据库系统方面的课程。近年来,我国在计算机科学教育中也对于数据库予以应有的重视。1983年教育部部属高等学校计算机软件专业教学方案将数据库概论列为四年制本科的必修课程,并于1983年6月通过了教学大纲,我们这本书便是按照该大纲编写的一本教材。

全书共分十章和两个附录。附录 I 的文件组织系学习数据库的预备知识,但已散见于数据结构等先行课中,列在这里,仅供参考,一般不必作为正式讲授内容。第一、二、三章为有关数据库的一般概念;第四至八章介绍三种重要数据模型。其中第八章的公理系统与模式分解的大部分已超出大纲范围,讲授时可酌情取舍。第九章为数据库控制,第十章为数据库设计。附录 II 的 IMAGE 3000 数据库系统介绍了如何在 HP-3000 计算机的 IMAGE 3000 数据库管理系统上建立数据库和使用数据库。对于能够用到 HP-

3000 计算机的单位,这一部分可作为实习教材。对于不能用上 HP-3000 计算机的单位则可作为参考,借以从中了解如何在一个实际的数据库管理系统上建立一个数据库应用系统。

本书是编者根据多年来进行数据库讲演及在高等学校为研究生、本科生开设数据库系统概论课程的经验,并在多次编写的内部交流讲义的基础上修改而成的。由于我们水平不高,而数据库作为计算机科学中的一个新兴分支发展又非常迅速,因此,在本书中存在问题一定不少,希读者提出批评意见。

中国科技大学研究生院罗晓沛同志,清华大学计算机科学与工程系杨德元同志、俞磐祥同志与沈金发同志详细阅读了全稿,并提出许多有益的意见,在此谨向他们致以衷心的感谢。

作为内部交流讲义,先后有五十余所兄弟院校采用作为教本,其中不少同志在使用以后曾以口头或书面方式提出意见,我们也对于他们表示谢意。

在讲稿整理过程中,我校信息管理系的刘怡同志,前78级、79级研究生吴鸥琦等同志,82级研究生张秉训、刘伶、杨明伟等同志以及资料室的王小平等同志,都从不同方面分别作了一些工作,对此作者一并表示诚挚的谢意。

萨师焯 王珊

83年4月

于中国人民大学

目 录

第一章 绪论	1
1.1 信息、数据与数据处理	1
1.2 数据处理系统	3
1.2.1 批处理系统	4
1.2.2 联机集中系统	6
1.2.3 联机分布系统	6
1.3 数据管理的进展	8
1.3.1 人工管理阶段	8
1.3.2 文件系统	9
1.3.3 数据库系统	11
1.4 信息结构	16
1.5 实体-联系方法	19
练习题	23
第二章 数据模型	25
2.1 格式化模型	25
2.1.1 格式化模型与图	25
2.1.2 层次模型	28
2.1.3 网状模型	29
2.2 关系模型	33
2.3 数据独立存取模型	34
2.4 数据模式的级别	35
练习题	37
第三章 数据库系统的构成	38
3.1 带有数据库的计算机系统构成	38
3.2 数据库管理系统	40

3.3	数据描述语言	41
3.3.1	模式 DDL	41
3.3.2	子模式 DDL	42
3.3.3	物理 DDL	43
3.4	数据操纵语言	43
3.5	数据库管理例行程序	44
3.6	存取数据库数据的过程	44
	练习题	47
第四章	DBTG 系统	48
4.1	DBTG 系统简介	48
4.2	DBTG 系统结构	49
4.3	DBTG 系统基本概念	50
4.3.1	系(DBTG 系)	50
4.3.2	系序	60
4.3.3	域(AREA)	64
4.3.4	数据库码	65
4.3.5	运行单位与当前值	66
4.3.6	记录存放方式	68
4.3.7	属籍类别	71
4.3.8	系值选择	74
4.4	模式数据描述语言	78
4.5	子模式数据描述语言	82
4.5.1	子模式与模式的区别	82
4.5.2	一个子模式的实例	83
4.6	数据操纵语言	86
4.6.1	存取机制	86
4.6.2	数据的选取	88
4.6.3	程序的运行环境	89
4.6.4	DML 语句	91
4.6.5	程序的例	113
4.7	存贮结构	118

4.7.1	概述	118
4.7.2	DBTG 系的实现方法	121
	练习题	133
第五章	IMS	136
5.1	IMS 简介	136
5.2	数据模型	138
5.2.1	物理数据库	140
5.2.2	物理数据库描述	141
5.2.3	层次序列	144
5.3	数据子模型	145
5.3.1	逻辑数据库	145
5.3.2	程序通讯块	147
5.4	IMS 数据操纵	150
5.4.1	概述	150
5.4.2	DL/1 语句	151
5.4.3	应用程序设计	156
5.5	存贮结构	161
5.5.1	IMS 存贮结构概述	161
5.5.2	HSAM	163
5.5.3	HISAM	164
5.5.4	HD 结构	168
5.5.5	HDAM	169
5.5.6	HIDAM	170
5.6	逻辑数据库	172
5.6.1	逻辑数据库的定义和例子	172
5.6.2	单向逻辑关系和双向逻辑关系	177
5.6.3	逻辑数据库描述	179
5.7	辅助索引	181
	练习题	183
第六章	关系数据库	185
6.1	关系模型的基本概念	186

6.1.1	数学定义	186
6.1.2	关系模型	189
6.2	关系数据语言概述	192
6.3	关系代数	193
6.3.1	传统的集合运算	193
6.3.2	专门的关系运算	194
6.3.3	检索操作	202
6.3.4	存贮操作	203
6.4	关系演算	203
6.4.1	定义与记号	204
6.4.2	检索操作	205
6.4.3	存贮操作	209
6.5	QBE	211
6.5.1	概述	211
6.5.2	检索操作	212
6.5.3	存贮操作	216
6.6	关系数据语言的特点及其评价	218
	练习题	220
第七章	System R 关系数据库系统	221
7.1	概述	221
7.2	SQL	223
7.2.1	数据定义	225
7.2.2	数据操纵	231
7.2.3	数据控制	235
7.2.4	宿主型 SQL	238
7.3	System R 的体系结构	240
7.3.1	RDS	241
7.3.2	RSS	247
7.4	SQL/DS 简介	255
7.5	总结	257

练习题	259
第八章 关系数据理论	261
8.1 问题的提出	261
8.2 规范化	266
8.2.1 函数依赖	266
8.2.2 码	268
8.2.3 范式	269
8.2.4 2NF	270
8.2.5 3NF	272
8.2.6 BCNF	273
8.2.7 多值依赖	276
8.2.8 4NF	280
8.2.9 规范化小结	280
8.3 数据依赖的公理系统	282
8.4 模式的分解	288
8.4.1 模式分解的三个定义	288
8.4.2 分解的无损连接性和保持函数依赖	291
8.4.3 模式分解的算法	294
8.4.4 连接依赖与5NF	299
8.4.5 小结	301
练习题	302
第九章 数据库的安全性、完整性和并发控制	306
9.1 安全性	306
9.1.1 安全性的一般问题	306
9.1.2 几种系统的安全性方法	311
9.2 完整性	316
9.2.1 完整性的约束条件	317
9.2.2 几种系统的完整性方法	319
9.3 数据库的恢复	322
9.4 并发控制	324
9.4.1 数据的一致性级别	324

9.4.2	封锁	326
9.4.3	封锁的尺度	326
9.4.4	死锁及其处理	327
	练习题	328
第十章	数据库设计	329
10.1	数据库设计的基本概念	329
10.1.1	数据库应用系统	329
10.1.2	软件工程和数据库工程	330
10.1.3	数据库设计过程	335
10.2	概念结构设计	335
10.2.1	数据分析	338
10.2.2	设计初步的 E-R 图	341
10.2.3	设计基本 E-R 图	344
10.3	逻辑结构设计	346
10.3.1	E-R 图向数据模型的转换	347
10.3.2	规范化理论的应用	350
10.4	数据库物理设计	352
10.4.1	物理设计的要求与内容	352
10.4.2	评价设计, 进行性能预测	355
10.5	数据库的实施和维护	358
10.5.1	数据库数据的载入	359
10.5.2	数据库的试运行	360
10.5.3	数据库的运行和维护	361
10.6	数据字典	361
10.6.1	数据字典的概念	361
10.6.2	数据字典与 DBMS	363
10.6.3	数据字典的用途	364
	练习题	365
	附 录	
I	文件组织	367
1.	文件的基本概念	368

1.1	文件与记录	368
1.2	记录格式	369
1.3	逻辑记录与物理记录	370
1.4	文件的存取方法	371
2.	文件组织基本方法	373
2.1	计算法	373
2.2	指引元法	374
3.	顺序文件	376
4.	索引文件	377
5.	索引顺序文件	381
6.	B 树	384
6.1	基本 B 树	385
6.2	B ⁺ 树	390
6.3	B 树的其它变种	394
6.4	B 树的并发控制	399
6.5	一个实例——VSAM	402
7.	杂凑文件	409
7.1	KAT 法	410
7.2	溢出记录的处理	415
7.3	杂凑文件的设计	418
8.	倒排文件	421
9.	变长记录文件	426
II	IMAGE 数据库系统	429
1.	概述	429
2.	IMAGE 的数据结构	431
2.1	术语介绍	431
2.2	数据集的类型和关系	432
2.2.1	主数据集	433
2.2.2	明细数据集	434

888	2.2.3	路径和链	435
880	2.2.4	自动主集和手工主集	437
078	2.2.5	主路径	438
170	2.2.6	排序项	438
77	3.	IMAGE 数据库描述语言	439
876	3.1	模式结构	440
170	3.2	口令部分	441
876	3.3	数据项部分	441
770	3.4	数据集部分	444
186	3.5	建立 IMAGE 数据库根文件	447
186	3.5.1	操作步骤	448
888	3.5.2	例子	449
008	3.5.3	模式处理程序命令	450
108	3.6	建立 IMAGE 数据库、数据集文件和装入数据	453
80	4.	用宿主语言调用 IMAGE 库过程存取数据库	454
804	4.1	IMAGE 库过程	455
804	4.2	调用库过程的方式	456
018	4.3	存取路径和当前值的概念	457
818	4.4	读取方式	458
818	4.5	库过程解释	459
10	5.	用自含语言 QUERY 存取数据库	470
888	5.1	概述	470
088	5.2	QUERY 命令	471
088	5.2.1	环境定义	471
188	5.2.2	查找	473
188	5.2.3	更新	475
188	5.2.4	杂务处理	476
888	5.2.5	打印和显示	477
888	5.2.6	过程处理	479
188	5.2.7	操作	482

6. IMAGE 的存贮结构	482
6.1 主集的结构	483
6.2 明细集的结构	484
7. IMAGE 数据库的维护	485
7.1 数据库的重组织	485
7.2 数据库的恢复	488
7.3 IMAGE 公用程序	497
III 参考文献	505

第一章 绪 论

在系统地介绍数据库之前,首先应该回答的一个问题是:什么是数据库以及为什么要发展数据库技术?

数据库是当代计算机系统的一个重要组成部分,我们很难用几句话严格、简明地概括它的全部特征。事实上,现在数据库的定义也是各式各样,不尽相同的。这是人们从不同角度用不同观点来看待数据库的结果。产生这种情况的另一个原因是数据库技术本身是逐渐形成的,而且直到今天还在发展之中,从而人们对它的认识也有一个历史的发展过程。因此,我们这里只能从数据管理的进展指出数据库的特点,对数据库给出一个轮廓性的描述。

1.1 信息、数据与数据处理

信息一词的英文为 Information (或译为情报),关于这个概念有各种不同的含义。我们这里将信息定义为向人们(或机器)提供关于现实世界新的事实的知识。而数据则定义为用以载荷信息的物理符号。二者是不可分离而又有一定区别的概念。一方面并非任何数据都能表示信息,信息只是消化了的数据;另一方面信息是更基本的直接反映现实的概念,而数据则是信息的具体表现,所以信息不随载荷它的物理设备的改变而改变,而数据则不然,它在计算机化的信息系统中往往和计算机系统有关。本书将按照上述观点分别采用信息与数据这两个术语,但在某些不需要严格分

辨的场合,也可以把两者不加区分地使用。例如,我们可以说信息处理与信息管理,也可以说数据处理与数据管理。

随着人类社会的发展,信息在各种活动中居于越来越重要的地位。由于信息量急剧增加,形式多样,结构复杂以及及时性要求的提高,信息处理也就成为一个极其重要的问题。

所谓信息处理或数据处理系指对信息(即各种形式的的数据)进行收集、贮存、加工与传播的一系列活动的总和。其基本目的是从大量的、杂乱无章的、难以理解的数据中抽取并推导出对于某些特定的人们来说是有价值、有意义的数 据,借以作为决策的依据。

广义地说,数据处理工作由来已久。早期是用各种初级的计算工具,如算盘、手摇计算机、电动计算机等,这是手工数据处理阶段。到19世纪末,Herman Hollerith发明了第一台用于编制人口普查表的卡片制表机以后,开始了数据处理的新的阶段。这类机器和古典的计算工具已有原则上的不同。它包括穿孔机、验孔机、分类机、卡片整理机、复孔机和制表机等几个组成部分,能半自动地进行工作。通常把这种用机械设备来进行数据处理的系统称之为机械数据处理系统(Mechanized Data Processing Systems,简记为MDPS)。由于机械数据处理系统的存贮介质、计算速度、处理能力以及自动化程度的限制,它的效率是不高的。随着四十年代末电子计算机的使用,特别是以后直接存取存贮设备的出现,使数据处理工作起了革命性的改变。使用电子计算机的数据处理系统称为电子数据处理系统(Electronic Data Processing Systems简记为EDPS)。今后我们所说的数据处理均系指电子数据处理而言。

信息系统

信息对人类活动,特别是管理工作是非常重要的,因此需要一个为企业或组织提供所需信息的工具。信息系统就是这样一个