简明关系数据原理

Concise Relational Data Principles

程录庆◎著

河海大学出版社
HOHAI UNIVERSITY PRESS

简明关系数据原理

Concise Relational Data Principles

程录庆◎著

河海大学出版社
HOHAI UNIVERSITY PRESS

·南京·

内 容 提 要

关系数据理论是信息管理的基础理论。本书围绕数据模型的三个要素——结构、操作、约束对关系数据模型及其原理作了详细的阐述，尤其对关系数据设计原理中的重要部分——数据约束的原理作了深入地分析。本书结合实例，以较为通俗的语言叙述了在真实的数据库设计中应遵循的基本规律，可供从事数据库工程的技术人员、信息管理工作人员以及在校的相关专业大学生参考。

图书在版编目(CIP)数据

简明关系数据原理 / 程录庆著. —南京：河海大学出版社，2019.8

ISBN 978-7-5630-6060-3

Ⅰ. ①简… Ⅱ. ①程… Ⅲ. ①关系数据库系统—研究 Ⅳ. ①TP311.132.3

中国版本图书馆 CIP 数据核字(2019)第 165851 号

书　　名	简明关系数据原理
著　　者	程录庆
书　　号	ISBN 978-7-5630-6060-3
责任编辑	杜文渊
特约校对	李　浪
封面设计	徐娟娟
出版发行	河海大学出版社
网　　址	http://www.hhup.com
地　　址	南京市西康路 1 号(邮编：210098)
电　　话	(025)83737852(总编室)　(025)83787156(编辑室) (025)83722833(营销部)
经　　销	江苏省新华发行集团有限公司
排　　版	南京新翰博图文制作有限公司
印　　刷	虎彩印艺股份有限公司
开　　本	880 毫米×1230 毫米　1/32
印　　张	5.5
字　　数	159 千字
版　　次	2019 年 8 月第 1 版
印　　次	2019 年 8 月第 1 次印刷
定　　价	36.00 元

序

如今计算机的应用渗透到了人们工作生活的几乎所有领域。从根本上说，计算机是执行数据与算法的机器，随着现代社会信息的需求增加，计算机由早期偏重算法应用发展到了以数据为中心的应用。常见的计算机软件，除了模拟仿真、图形图像、人工智能、游戏等属于算法为主的应用外，绝大多数都是属于以数据为主的应用，如各种信息系统。数据库技术是计算机科学中极其重要的一个分支。

将数据从计算机程序中分离出来专门管理最早出现于 20 世纪 60 年代，早期的数据库大多基于层次模型和网状模型建立。70 年代，IBM 公司研究员 E.F.Codd 提出了数据的关系模型，且深入研究了数据的关系理论，从而奠定了数据库的理论基础，使数据库从技术上升到理论。随后，数据库的应用发展很快，出现了大量基于关系模型的商品化数据管理软件，关系数据理论也因此丰富起来，理论与技术相辅相成，实践证明，关系数据理论对数据库技术的发展起到了非常重要的作用。

今天，数据库技术正朝着大数据的方向发展。在规模上，万物互联使得数据的量提升了几个数量级；在数据模型上，出现了很多新的非关系的数据模型，如面向对象数据模型、XML 数据集、演绎数据库等；在应用领域上，与其他学科相融合，出现了如生物信息技术、工程数据库、空间数据库、生命数据库等；在应用层次上，有数据挖掘、决策支持、知识发现、人工智能等；在概念上，数据库由

传统的中心管理且边界清晰的数据集合向无边界无中心的信息系统过渡。然而，关系数据理论是关于数据的科学，是数据库的基础理论，从根本上来看，各种数据集合的基础依然是关系数据库系统，对于数据工作者和用户，深入理解关系数据理论仍然是十分有意义的。理解原理能带来对实践的精准直觉与深刻洞察。

计算机数据库是一个集成系统，数据库原理与技术包含的内容非常丰富。数据存储和输入输出、数据模型、数据模型设计原理、索引与查询优化原理、处理程序设计、系统控制、数据可靠性、数据安全性，等等。但是，本书只关注数据模型设计原理部分，也是“真正”的数据原理部分，这是作者擅长的领域，作者结合自己多年的教学经验将一般人看来艰深的关系数据原理以通俗的语言阐述，娓娓道来，全面而且清晰。本人相信，无论是拥有数据库技术经验的工程师，还是普通的数据工作者，或者是对数据库一无所知的学生，阅读本书都将会有所收获。

数据库还是一项与各领域知识紧密结合的技术，因为数据库终究是为主体业务服务的。作者在书中强调的语义分析很好地体现了数据库的这一特点，这也是该书的创新之处。

如今，在很多高等院校的管理学院都开设了《数据库原理与应用》课程。我认为，对于管理类专业的学生，这门课中涉及的计算机知识固然重要，但更重要的是现实的数据与信息的处理技术。在数据资产逐渐超越实物资产的今天，管理好数据信息资源是每一个管理人员必备的基本素质，通过数据库系统课程的学习，大学生是可以达到增强信息管理能力目的的。作者的这本书对数据原理的论述与时下众多的教材不同，很有特色，值得一读。

南京邮电大学管理学院院长

前　言

数据本来是和实物融合一体的，一直以来，人们在处理信息时，并没有刻意地认识到数据库的存在。在计算机发明和普及应用以后，其作为专门处理数据的工具，它使得人们在生产生活中逐渐实现数据与实物的分离，单纯存储和处理信息的计算机数据库产生了。虽然信息处理活动自古有之，但专门的数据库却是新事物。

计算机的发明到今天时间不算太长，但其发展速度惊人，随计算机技术进展而进展的数据库技术也是日新月异，变化之快以至于大学里的数据库教材都不能来得及好好完善就已经过时了。笔者在大学里上过很多次的《数据库原理与应用》课程，发现多数学生即使在课程结束时仍然对数据库的基本原理不甚了了。为此，笔者下决心写这本书，舍弃计算机数据库应用的诸多细节，专注基本理论，尝试以不那么“学术”化的语言阐述关于数据的原理。

数据库的根本目标主要有两点，一是向库中存入数据；二是从库中取出需要的数据。“存”和“取”紧密相关，“存”的时候要考虑将来怎么“取”，“取”的时候要清楚怎么“存”的。数据应当以什么方式存储有其自身的规律，这规律就是数据的理论，掌握数据的理论对于信息工作者来说无疑是有益且必要的。

数据库是一个系统工程，全面掌握数据库技术需要相当的原理、设计、开发和应用的知识。本书不是一本面面俱到的教材，本

书仅关注数据库原理部分，笔者旨在以关系模型为例说明应用和设计数据库应当遵循的基本原理。当然，大多数的教材也是以关系模型为主，介绍数据库原理与应用的。之所以以关系模型为例，一是因为关系数据库的应用很普遍；二是关系模型有很多的固定的理论，而其他的数据模型则更依赖于经验；三是数据的关系理论具有普遍性，可以应用到其他数据模型的分析当中。

信息化时代，数据处理不仅仅是专业人员要做的工作，而是几乎所有人工作生活中都需要面对的。即使是计算机数据库，它的设计与使用也不仅仅是计算机专业人员的事，数据库总是服务于某个应用，应用领域的工作人员自然也是数据库设计与维护的参与者。所以，数据库虽然是计算机的专业课程，但是普通人也可以且需要学习，以应付日常工作生活的信息处理要求，尤其是和表格数据打交道的办公人员，了解关系数据原理之后，将能更有效地管理各类数据。

关系数据理论涉及到许多的数学知识，本书中的部分内容即使对专业的学生来说也是过于抽象，非专业人员更是望而生畏，然而，数据世界其实是现实世界的反映，语义分析才是数据处理的灵魂。对信息处理而言，数学只是提供了描述数据问题的一种方法，而不是问题本身，针对具体的数据集合，我们没有必要过于纠结数学上的逻辑演绎，而应该从真实世界的信息构成来理解数据的语义，进而找到数据应有的结构。本书列举了很多例子，且不厌其烦地作了大量的语义解释，目的也是为了强调数据处理中语义分析的重要性。

所谓“万法归一”，意思是万事万物虽然表现的形态各异，但其中的规律或道理却是统一的，数据库也是一样。如今，数据模型种类繁多，各种计算机数据库的工具更是数不胜数，一个人想要掌握所有这些技术显然是不可能的。本书有意避开涉及各种具体的数据库技术和工具，而专注于基本原理的阐释，是希望有更多的非专

业技术人员阅读。如果读者阅读本书之后，能从中获得些许关于数据和信息的道理的领悟，那将是笔者最大的宽慰，这也是笔者要做的一点尝试。

另外，关于这本书，还有几点需要说明：

（1）数据模型有三个要素，一是结构；二是操作；三是约束。本书围绕关系数据模型的这三个要素安排章节的。

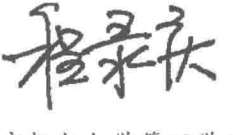
（2）书中每一章后面安排了习题，这些习题有的是思考型的，有的是练习型的。值得注意的是，其中某些习题是对正文中不便叙述的内容作的补充，所以即使不做习题，也不妨看一看。大部分练习型的习题在附录中提供了答案。

（3）有些习题需要计算机工具才能完成，如第三章习题 2 的 SQL 语言编程，需要读者参考相关的资料。

（4）本书对涉及的一些定理或算法一般都省略证明过程，因为这不是本书的重点，有兴趣的读者可以参考其他资料。

（5）虽然作者希望更多的非专业人员阅读本书，但书不可避免地出现了许多抽象难懂的专业内容。作者建议，一般读者在阅读过程中如果碰到不易明白的生涩概念，不妨忽略过去或者暂时放置一旁，能从全局的视角理解主要的思想就很不错了。

最后，虽然作者在书中屡次强调结合实际设计数据库的重要性，但作者自身的实践经验和水平却很有限，错误实在难免，敬请批评指正。



南京邮电大学管理学院

2019 年 3 月 14 日

目　录

第一章　数据库概论

在生产、贸易等经济活动以及日常生活中，人们要与大量的信息打交道，如材料的消耗，产品的产量，商品的销售，工资及成本的计算，各种情报资料、档案材料的存储与检索，以及如今日常生活中广泛存在的各种通讯信息，等等。不仅是现在，对信息的处理是人类自有文明以来就存在的重要活动，上古时代的结绳记事，以及历史上出现的语言、文字、图符、钟鼓、烽烟、竹简、纸张等无不是信息表达与存储的工具，人们在长期的生产生活实践中积累了丰富的信息管理理论和经验。

计算机的出现使得人们处理信息的技术迈入了一个新的阶段。

一、数据、信息和数据冗余

数据是描述和记载客观事实的符号。提到数据，很多人第一反应可能是数字，如 100 头羊、50 千米、23 岁等，但数据的形式远不止数字，所有对客观事实的记录都可以称为数据。记录以一定的形式表现，传统的如数字、文字、图形、图像等视觉符号，计算机应用以后，数据的形式拓展了很多，声音、动作、味道、甚至是体验都可以以某种形式记录下来，这些都是数据。

信息，指人接触自然的、社会的客观事实后得到的感觉、认识、

知识和意识等。信息是抽象的，广义上理解，信息是无处不在的，世间万事万物皆是人可以感知的信息。无穷无尽的信息当中，有些信息可以以一定的形式记录下来，而有些则不能。狭义的信息就是指可以被记录下来的那部分信息，这部分信息也称为数据信息，这个记录就是数据。通常所谓的信息管理，管理的信息指的就是数据信息，也就是狭义的信息。在数据信息范畴，通常并不严格去区分信息和数据在概念上的差异，说到数据，侧重是它的具体形式，而说到信息，则强调数据所记录的内容。数据是信息的载体，信息是数据的内容。

习惯上，我们把原始的事实或资料称之为数据，是加工的“原料”，而把经过分析和处理的结果，即经过“消化”的数据称之为信息，如果信息被记录下来成为下一步分析和处理的“原料”，那么信息也就转化成了数据。从信息处理的角度看，处理之前的称之为数据，而处理之后的就称之为信息。

图 1-1　数据与信息的转换

数据表达信息，这个信息，从人的理解角度来说，就是数据的含义，或称数据的语义。如描述学生的一条记录，(宋青华，男，1981，信息科学，2011，江苏)。这一条数据告诉我们：宋青华是1981年出生，性别为男，籍贯江苏，2011年入学的一名学生。数据蕴含的语义就是信息。

不同的数据可能蕴含相同的信息，或者说，同一信息可由不同的数据形式表达。比如，“I am a teacher”和“我是一名教师”这两个数据所表达的信息就是一样的，再如，“今天的平均气温摄氏20℃”与“今天平均气温华氏68°F”所传达给人的气温信息也是一

致的。

数据是具体的、物理的，而信息是抽象的、逻辑的。在实际的数据存储中，区分数据和信息的这个差别有时也很重要，可以由此判别一个数据集合中是否存在数据冗余，所谓的数据冗余，就是在数据集合中出现了多余的数据。如何判断一个数据是否多余呢？道理很简单，数据集合中增加了这个数据是否表达了更多的信息？如果回答是否定的，那么这个增加的数据就是冗余。例如，在一个表中存储了这样一组数据：(单价，销售数量，销售额)，这里**销售额**[①]是冗余，为什么呢？因为**销售额＝单价×销售数量**，也就是说，**销售额**可由**单价**和**销售数量**计算得出，表中**单价**和**销售数量**已经蕴含了**销售额**信息，增加**销售额**的存储并没有增加信息，所以是冗余。

再比如，观察表 1-1 中的数据，表中数据存在冗余。因为表中数据(1001，王熙凤，19)重复了三次[截取出来如表 1-1(1)]，表达的信息只有一个，即学号为 1001 的学生姓名是王熙凤，年龄 19 岁，这个信息本来只需记录一次(1001，王熙凤，19)就够了，然而，这里(1001，王熙凤，19)记录三次，是数据冗余，且是因为重复存储引起的冗余。

表 1-1　学生-课程

学号	姓名	年龄	课程名称	成绩
1001	王熙凤	19	数据库原理	90
1001	王熙凤	19	操作系统	65
1001	王熙凤	19	通信工程	80
1002	李　逵	20	数据库原理	85
1002	李　逵	20	通信工程	92
1003	武　松	21	操作系统	70

表 1-1(1)

学号	姓名	年龄
1001	王熙凤	19
1001	王熙凤	19
1001	王熙凤	19

表 1-1(2)

学号	课程名称
1001	数据库原理
1001	操作系统
1001	通信工程

不过，重复存储数据也不一定就是冗余。再看表 1-1 中学号 1001 也被记录了三次，但这个重复就不是数据冗余，因为它记录了学生 1001 与课程的三次不同对应[截取如表 1-1(2)]，表示选修了三门课程，每一次重复的意义不同，表达了更多的信息，这个重复是必要的，不重复不足以表达这个学生选修了三门课的信息。

逆向的思考也是成立的，如果数据集合中的某个数据去掉以后，表达的信息和去掉之前是一样的，那么这个数据也是冗余。总之，数据冗余指的是不必要的数据增加，这里的“不必要”意思是指增加了数据存储却没有表达更多的信息。

存储的数据是否冗余是数据库设计时须考虑的一个重要因素，数据冗余带来的问题不仅仅是占据了更多的存储空间，更严重的是它会使得整体的数据处理效率下降。

二、数据库、数据库管理系统、数据库系统、数据库应用系统

这几个术语，在非正式场合，习惯上都说成数据库，然而，它们概念上是有差异的，各有所指。

1. 数据库

从字面上理解，数据库就是数据的集合。不过，如果以数据集合