

/THEORY/IN/PRACTICE

你不可不知的 关系数据库理论

Relational Theory for Computer Professionals

//

[美] C. J. Date 著
张大华 方帅 译
柳永坡 张奎 审

O'REILLY®



人民邮电出版社
POSTS & TELECOM PRESS

O'REILLY®

你不可不知的关系数据库理论

[美] C. J. Date 著
张大华 方 帅 译
柳永坡 张 奎 审

人 民 邮 电 出 版 社
北 京

图书在版编目 (C I P) 数据

你不可不知的关系数据库理论 / (美) 达特
(Date, C. J.) 著 ; 张大华, 方帅译. — 北京 : 人民邮
电出版社, 2015. 3
ISBN 978-7-115-37921-4

I. ①你… II. ①达… ②张… ③方… III. ①关系数
据库系统 IV. ①TP311.138

中国版本图书馆CIP数据核字(2015)第013084号

版 权 声 明

Copyright © 2013 by O'Reilly Media, Inc.

Simplified Chinese Edition, jointly published by O'Reilly Media, Inc. and Posts & Telecom Press, 2015.
Authorized translation of the English edition, 2013 O'Reilly Media, Inc., the owner of all rights to publish
and sell the same.

All rights reserved including the rights of reproduction in whole or in part in any form.

本书中文简体字版由 O'Reilly Media, Inc. 授权人民邮电出版社出版。未经出版者书面许可, 对本书的
任何部分不得以任何方式复制或抄袭。

版权所有, 侵权必究。

-
- ◆ 著 [美] C. J. Date
 - 译 张大华 方 帅
 - 审 柳永坡 张 奎
 - 责任编辑 傅道坤
 - 责任印制 张佳莹 焦志炜

- ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号
邮编 100164 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
三河市海波印务有限公司印刷

- ◆ 开本: 787×1000 1/16
印张: 14.75
字数: 265 千字 2015 年 3 月第 1 版
印数: 1—2 500 册 2015 年 3 月河北第 1 次印刷

著作权合同登记号 图字: 01-2014-2812 号

定价: 49.00 元

读者服务热线: (010)81055410 印装质量热线: (010)81055316
反盗版热线: (010)81055315



内容提要

关系型数据库是建立在关系模型基础上的数据库，借助于集合代数等数学概念和方法来处理数据库中的数据。关系型数据库的基础——关系理论被认为是 SQL 的基础。

本书为我们讲解了什么才是真正的关系型数据库，与当前的数据库产品相比，它的特点和优势是什么。本书分为 3 部分，共计 14 章。第一部分是数据库的基础，讲解了数据库基本概念，关系和关系变量，码、外码和相关概念，关系运算符，约束和断言，关系模型等内容；第二部分讲解了事务的相关概念，以及如何设计一个好的数据库；第三部分则讲解了 SQL 相关的知识，其内容涵盖了 SQL 基本表，SQL 操作符和运算符，SQL 约束，SQL 与关系模型等内容。本书最后的 5 个附录涵盖了 Tutorial D 语法、TABLE_DUM 和 TABLE_DEE、集合论、关系演算，以及与关系理论知识相关的资源。

本书适合数据库开发、维护人员以及高校数据库专业的师生阅读。对于想要真正理解什么是关系型系统的读者来说，本书也是不错的选择。

O'Reilly Media, Inc.介绍

O'Reilly Media通过图书、杂志、在线服务、调查研究和会议等方式传播创新知识。自1978年开始，O'Reilly一直都是前沿发展的见证者和推动者。超级极客们正在开创着未来，而我们关注真正重要的技术趋势——通过放大那些“细微的信号”来刺激社会对新科技的应用。作为技术社区中活跃的参与者，O'Reilly的发展充满了对创新的倡导、创造和发扬光大。

O'Reilly为软件开发人员带来革命性的“动物书”；创建第一个商业网站（GNN）；组织了影响深远的开放源代码峰会，以至于开源软件运动以此命名；创立了Make杂志，从而成为DIY革命的主要先锋；公司一如既往地通过多种形式缔结信息与人的纽带。O'Reilly的会议和峰会集聚了众多超级极客和高瞻远瞩的商业领袖，共同描绘出开创新产业的革命性思想。作为技术人士获取信息的选择，O'Reilly现在还将先锋专家的知识传递给普通的计算机用户。无论是通过书籍出版，在线服务或者面授课程，每一项O'Reilly的产品都反映了公司不可动摇的理念——信息是激发创新的力量。

业界评论

“O'Reilly Radar博客有口皆碑。”

——Wired

“O'Reilly凭借一系列（真希望当初我也想到了）非凡想法建立了数百万美元的业务。”

——Business 2.0

“O'Reilly Conference是聚集关键思想领袖的绝对典范。”

——CRN

“一本O'Reilly的书就代表一个有用、有前途、需要学习的主题。”

——Irish Times

“Tim是位特立独行的商人，他不光放眼于最长远、最广阔的视野并且切实地按照Yogi Berra的建议去做了：‘如果你在路上遇到岔路口，走小路（岔路）。’回顾过去Tim似乎每一次都选择了小路，而且有几次都是一闪即逝的机会，尽管大路也不错。”

——Linux Journal

译者序

目前, 现有大量新型数据库技术得到了快速发展并受到大量关注, 无论是已有进展的“云计算”产业, 还是流行的关注点“大数据”, 其核心都是对“数据”的研究, 这就使得数据库中的关系理论和 SQL 得到更进一步的丰富和发展。市面上关于 SQL 语言或者关系理论或者二者兼而有之的书籍参差不齐, 良莠混杂, 很难有一本能循序渐进深入讲解该领域的书籍。本书译者结合自己多年工作经验, 结合实际行业内需求, 挑选 C. J. Date 的这本图书供读者参考。

C. J. Date 是最早认识到 Code 在关系模型方面所做的开创性贡献的学者之一, 他是关系数据库技术领域非常著名的独立撰稿人、学者和顾问, 他使得关系模型的概念普及化。他参与了 IBM 公司的 SQL/DS 和 DB2 两大产品的技术规划和设计。30 多年来, Date 一直活跃在数据库领域中, 他的著作包括:《SQL 和关系理论(第二版)》、《数据库设计和关系理论》、《视图修改和关系理论》、《数据库系统导论》、《对象关系数据库基础: 第三次宣言》等。

本书共分为 14 个章节, 内容分别为关系型数据库“基础介绍”、“事务和数据库设计”、“SQL”及附录四个部分, 翻译团队由中国电力科学研究院信息通信研究所相关人员组成。具体分工为: 前言、译者序、“基础介绍”部分由张大华、方帅负责;“事务和数据库设计”由张大华、方帅、纪鑫、李哲、陈相舟负责;“SQL”部分由谢迎军、丁辉、常亮、刁倩、刘月林、钱声攀、魏俊负责; 附录部分由上述全体翻译团队共同负责; 全书译文审核与校正由北京航空航天大学柳永坡、张奎完成。

本书中文版能够出版发行, 首先要感谢本书的作者, 是他们为我们著作了一本好书。其次要感谢人民邮电出版社引进了本书, 使我们获得了翻译此书的机会, 并实现将其介绍给国内广大读者的良好愿望。也借此机会向本书的技术审校人员致以崇高的敬意, 他们使本书的质量与水平向前迈进了一大步。此外, 特别要感谢国家电网公司高级顾问曾楠, 他以巨大的热情和高度的责任心在本书翻译过程中给予了大力的技术指导工作。也要感谢幕后的工作人员, 是他们的辛勤劳动使得本书顺利付梓。在此, 译者代表对所有为此书的出版做出贡献的人表示深深的感谢和崇高的敬意!

2 译者序

本书对原著的错误之处作了一些修正，在原著难懂或需要提醒的地方添加了一些译者说明。尽管我们在翻译过程中力图做得更好，但终究因业务水平、英文水平，乃至中文文学水平的欠缺，以及翻译过程中的粗心和不够严谨，致使本译作存在错误、不足和不当之处，也恳请读者批评指正，并热切期望读者对本书提出宝贵意见和建议。

译者

2014 年 12 月

数学科学表明的是原理，它是描述事物之间不可见关系的一种语言。

——Ada, Countess of Lovelace,

Quoted in Dorothy Stein (ed.):

Ada: A Life and a Legacy (1985)

科学的本质是：问一个不恰当的问题，于是走上了通往恰当答案的路。

——Jacob Bronowski:

The Ascent of Man (1973)

Hofstadter 定理：一件事情总是会花费比你预期更多的时间，就算是你已经考虑过本条 Hofstadter 定理。

——Douglas R. Hofstadter:

Gödel, Escher, Bach: An Eternal Golden Braid (1979)

献辞

谨将本书献给我的妻子 Lindy 和女儿 Sarah、Jennie。我爱你们。

关于作者

C. J. Date 是一位独立作家、演说家、研究员、顾问，在关系数据库技术领域颇有建树。C. J. Date 以他的 *An Introduction to Database Systems*（第 8 版，Addison-Wesley，2004 年出版）一书而闻名，到写作本书时为止，该书发行量已达到近 90 万册，被全球数百所大学和学院使用。他还出版了许多数据库管理方面的著作，新近出版的包括下面这些著作。

- Ventus: *Go Faster! The TransRelations Approach to DBMS Implementation*（2002 年出版，2011 年再版）
- Addison-Wesley: *Databases, Types, and the Relational Model: The Third Manifesto*（第 3 版，与 Hugh Darwen 合作编写，2007 年出版）
- Trafford: *Logic and Databases: The Roots of Relational Theory*（2007 年出版）和 *Database Explorations: Essays on The Third Manifesto and Related Topics*（与 Hugh Darwen 合作编写，2010 年出版）
- Apress: *Date on Database: Writings 2000-2006*（2007 年出版）和 *The Relational Database Dictionary, Extended Edition*（2008 年出版）
- O'Reilly: *SQL and Relational Theory: How to Write Accurate SQL Code*（第 2 版，2012 年出版）、*Database Design and Relational Theory: Normal Forms and All That Jazz*（2012 年出版）和 *View Updating and Relational Theory: Solving the View Update Problem*（2013 年出版）

Date 先生于 2004 年成为计算机工业名人堂中的一员。他在用明晰易懂的方式解释复杂技术课题方面具有傲视群雄的能力。

前言

关系数据模型是百年来最伟大的技术发明之一，它是我们完成数据库领域任何事情的基础。的确，它使数据库管理成为一门科学，而不再像过去那样是一些技巧、技术和经验法则的特定集合。因此，每一个与数据库管理有关的专业人员，或多或少都会主动去获得一些与关系模型有关的知识，以加深对关系模型的理解。因为如果没有它，想开展高效的工作、获得较高的工作性能几乎是不可能的。

不幸的是，想要达到如上所说的“获得知识，加深理解”是不容易的。这有多方面的原因，但影响最大的原因是 SQL 语言，它是一种“关系型”语言的官方标准¹，当今市场上的每个数据库产品都支持其中的某些方面。大家普遍承认，至少是应该普遍承认，作为关系模型的抽象概念的某种具体实现，SQL 有着非常严重的缺陷（这是我在前面那句话中给“关系型”加上引号的原因）。至于在数据库世界，每个人之所以都知道一些关于 SQL 的知识，是几乎所有的关系型（或可能的关系型）教学的重点都倾向于 SQL 本身，而不是基本的理论。因此，仅仅因为他们知道 SQL，人们就认为他们知道关系理论，也就不足为奇了。然而，令人难过的是，如果你只是像这样知道 SQL，那么你肯定不知道关系理论，而你不知道的事情最终可能会伤害你。

本书的首要目的就是详细地教你关系理论（至少是尽我所能地详细），第二个目的是从关系理论的角度描述 SQL（或者描述 SQL 的核心特征），这也可能会给我带来一些其他灵感。一些读者可能知道，不久前我出版了另一本书来论述这件事情，书名为 *SQL and Relational Theory: How to Write Accurate SQL Code*（第 2 版，2012 年 O'Reilly 出版）。然而，与现在的这本书不同，早期的图书主要是面向数据库从业人员，通常情况下适用于至少有三四年的“工作在数据库系统第一线”的人员，或者经常把它作为日常工作一部分的人员（因此，肯定会有一些关于 SQL 的工作经验）。那本书的主要目的是展示如何把关系理论应用到使用 SQL 的过程中，因此，

¹ 多年来，SQL 已经演变了几个版本。在本书出版时的版本是 2011 版（即 SQL:2011）。规范的引用格式为：国际化标准组织（ISO），数据库语言 SQL，文档 ISO/TEC9075:2011。第 10 章和附录 E 对这种引用做了详细的介绍，第 10 章还专门简要介绍了它的相应发展史。

SQL 的行为看起来就像是一个真正的关系语言一样（在那本书中我引用的一个规则是，从关系型的角度使用 SQL）。在本书和那本书之间不可避免地会有一些重复。实际上，本书中有一些文字就是从早期写作的图书中复制和粘贴的，有些几乎是完全一样。不管怎样，你现在看到的一定是一本不同的图书，因为它是面向不同的读者（参看下面的详细解释）。然而，本书接下来的内容还需要频繁引用早期的图书，所有的引用都以简写的形式给出（即 *SQL and Relational Theory*）。而且，我认为如果你熟悉我的早期作品，那么你从本书中可能不会获得太多的知识。当然，我不是不鼓励你阅读此书，只是如果你读了，你可能不会发现太多的新知识。

适合的读者

本书面向的读者是计算机专业人员。前提是你了解一些通用的计算机和编程知识，也至少熟悉一种通用的编程语言。但是你不需要了解数据库，无论是关系型的还是其他的，也不用了解 SQL。当然，你至少要知道当今的数据库系统都几乎被假设为“关系型”（不管它是不是），但是我不会依赖于这个假设。然而，请注意，如果你恰好已经了解数据库的一些知识，那么你或许要格外留心本书中讲的内容！你可能会发现需要抛掉一些以前的想法（我们大家都知道，抛掉已有的想法是很难的）。当今的数据库产品即便被贴上了“关系型”的标签，但实际上却不完全关系型的，它们在很多方面都与理论上的关系型思想相差很远，你很快就会发现这一点。正如我在 *SQL and Relational Theory* 一书中写到的那样：

在这里我要为我带有进攻性的口吻道歉。但是如果你对关系模型的了解仅仅来自于对 SQL 的了解，那么恐怕你并没有如你所期待的那样真正了解关系模型，你也许应该知道“有些事情不是这样的”。我也不能过于强调：SQL 和关系模型不是一回事。

注意：我想说的是，关于本书的详细材料可以参照我的一个在线论坛，如果想要更详细的资料，请登录网站，网址为：www.justsql.co.uk/chris_date/chris_date.htm，也可以参考视频，网址为：<http://oreilly.com/go/date>。

本书结构

本书分为三个部分及附录。第一部分为“基础知识”，主要介绍了关系模型本身，换句话说，介绍了关系数据库技术的理论基础（虽然它也强调了整个理论的实际应用）。第二部分为“事务和数据库设计”，主要讨论了为了理解通用的数据库和特定的关系型数据库而需要的知识，但是不需要真正了解关系模型本身（除非它以

该模型作为基础)。第三部分为“SQL”，又回到了第一部分的材料，展示了该部分讨论的概念是如何用 SQL 语言实现的。至于附录，是很多材料的混合（这也是附录通常的作用），我认为这不值得在书的正文部分过多描述，因此都放在了这里加以详细描述。

注意：复习题和练习是大部分章节中不可或缺的一部分，我建议你查阅后面的答案之前尝试着回答这些问题，以使自己得到训练。尤其你要知道的是，经常有目的地做些练习，可以在很多方面得到锻炼。

致谢

我想再一次感谢我的妻子 Lindy，她在本书的整个著作过程中给予了大力支持，同时要感谢所有的前辈。我还要感谢 David Livingstone，是他最初给了我写本书的一些建议或类似的想法，也要感谢 Allen Noren，是他促使我写作本书的（在我的其他的由 O'Reilly 出版的著作中，Allen 的想法为此奠定了坚实的基础，就像本书一样，也是奠定了理论基础。我相信他是对的，这就是我想要写的著作）。最后，我还要感谢在线论坛的讨论者，因为人物众多，不胜枚举，他们的问题和评论帮助我合理地组织了本书的结构，也帮助我决定了本书内容的取舍。当然，无论是事实还是判断中的任何错误，我都会一如既往地负责。

C. J. Date

加利福尼亚希尔兹堡

2013 年

目 录

第一部分 基础知识

第 1 章 数据库基本概念	3
1.1 什么是数据库	3
1.2 什么是数据库管理系统	5
1.2.1 数据依赖	6
1.2.2 DBMS 的其他功能	7
1.3 什么是关系型 DBMS	8
1.4 数据库系统与程序系统	10
1.5 练习	13
1.6 答案	14
第 2 章 关系和关系变量	17
2.1 关系	17
2.1.1 属性	18
2.1.2 元组	20
2.1.3 关系的特点	20
2.2 关系变量	22
2.3 练习	24
2.4 答案	24
第 3 章 码、外码和相关概念	27
3.1 完整性约束	27
3.2 码	28
3.3 外码	31
3.4 关系变量定义	32

3.5	导入数据库	34
3.6	数据库系统和程序系统对比	35
3.7	练习	36
3.8	答案	36
第4章 关系运算符 I		39
4.1	Codd 的原始代数	39
4.2	限制	42
4.3	投影	43
4.4	练习 I	46
4.5	答案 I	46
4.6	并、交、差	47
4.6.1	并	47
4.6.2	交	48
4.6.3	差	49
4.6.4	一些公式化的特性	50
4.7	改名	50
4.8	练习 II	52
4.9	答案 II	52
4.10	联接	54
4.10.1	笛卡儿乘积	56
4.10.2	再论交运算	57
4.10.3	原始运算符	58
4.11	关系比较	58
4.12	修改运算符的扩充	59
4.13	练习 III	61
4.14	答案 III	61
第5章 关系运算符 II		63
5.1	匹配和非匹配	63
5.2	扩展	65
5.3	映像关系	67
5.4	聚集和分类汇总	70
5.4.1	分类汇总	71

5.4.2 明确的分类汇总	73
5.4.3 广义约束	74
5.5 练习	74
5.6 答案	75
第6章 约束和断言	77
6.1 数据库约束	77
6.2 关系变量断言	81
6.3 断言与约束	84
6.4 练习	85
6.5 答案	87
第7章 关系模型	89
7.1 关系模型定义	89
7.2 类型	91
7.3 关系类型产生器	93
7.4 关系变量	95
7.5 关系赋值	96
7.6 关系运算符	96
7.6.1 安全性	98
7.6.2 视图	98
7.7 结论	100

第二部分 事务和数据库设计

第8章 事务	103
8.1 什么是事务	103
8.2 恢复	104
8.2.1 恢复日志	106
8.2.2 ACID 特性	106
8.3 并发性	107
8.4 锁	108
8.5 SQL 的讨论	110
8.6 练习	111
8.7 答案	112

第9章 数据库设计	113
9.1 无损分解	114
9.2 函数依赖	116
9.3 第二范式	117
9.4 第三范式	119
9.5 BC 范式	120
9.6 结论	121
9.7 练习	122
9.8 答案	123

第三部分 SQL

第10章 SQL 基本表	129
10.1 发展历史	129
10.2 基本概念	131
10.3 表的特性	131
10.4 修改表	133
10.5 等值比较	134
10.6 定义表	135
10.7 SQL 系统与程序系统	137
10.8 练习	137
10.9 答案	138
第11章 SQL 操作符 I	141
11.1 限制	141
11.2 投影	142
11.3 并、交、差	143
11.4 更名	145
11.5 练习 I	145
11.6 答案 I	146
11.7 联接	148
11.7.1 另一种格式	149
11.7.2 规范特性	149
11.7.3 笛卡儿乘积	150

11.8 基本表表达式的求值	150
11.9 表的比较	151
11.10 显示结果	153
11.11 练习 II	154
11.12 答案 II	154
第 12 章 SQL 运算符 II	157
12.1 MATCHING 与 NOT MATCHING	157
12.2 EXTEND	159
12.3 映像关系	161
12.4 聚集和归纳	161
12.4.1 归纳	162
12.4.2 “通用的限制”	165
12.5 练习	167
12.6 答案	167
第 13 章 SQL 约束	169
13.1 数据库约束	169
13.2 类型约束	172
13.3 练习	173
13.4 答案	174
第 14 章 SQL 与关系模型	177
14.1 概述	177
14.2 SQL 与关系模型的不同点	179
14.3 练习	182
14.4 答案	182
附录 A Tutorial D 语法	185
附录 B TABLE_DUM 和 TABLE_DEE	189
附录 C 集合论	195
附录 D 关系演算	205
附录 E 进阶阅读指南	215