

/THEORY/IN/PRACTICE

O'REILLY®

Database In Depth

Relational Theory for Practitioners

深度探索关系数据库

实践者的关系理论



C. J. Date 著
熊建国 译 赵凯 审校



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

深度探索关系数据库

—— 实践者的关系理论

Database In Depth Relational Theory for Practitioners

[美] C.J. Date 著

熊建国 译

赵凯 审校

电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书阐述了关系模型背后的根本原则，这些原则也是所有数据库（后备应用软件）的应用基础，因为这些原则，今天的计算机世界才得以继续。该书内容直指关系数据库实际工作的核心。作者 C.J.Date 是数据库界的权威、富有经验的数据库开发者和设计者。该书不仅适用于数据库开发人员和设计人员，同样也适用于不同领域的专家和学者，包括 DBA（数据库管理员）、信息建模者、数据库咨询人员，等等。

0-596-10012-4 Database In Depth:Relational Theory for Practitioners. Copyright © 2005 by O'Reilly Media, Inc. Simplified Chinese edition, jointly published by O'Reilly Media Inc. and Publishing House of Electronics Industry, 2006. Authorized translation of the English edition, 2005 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.

本书中文简体版专有出版权由 O'Reilly Media, Inc. 授予电子工业出版社，未经许可，不得以任何方式复制或抄袭本书的任何部分。

版权贸易合同登记号 图字：01-2007-1363

图书在版编目（CIP）数据

深度探索关系数据库：实践者的关系理论 /（美）戴特著；熊建国译．—北京：电子工业出版社，2007.5

书名原文：Database In Depth: Relational Theory for Practitioners

ISBN 978-7-121-04000-9

I. 深… II. ①戴…②熊… III. 关系数据库 IV. TP311.138

中国版本图书馆 CIP 数据核字（2007）第 033298 号

责任编辑：周 筠

项目编辑：梁 晶

印 刷：中国电影出版社印刷厂

装 订：三河市皇庄路通装订厂

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱 邮编 100036

开 本：787×980 1/16 印张：15.25 字数：300 千字

印 次：2007 年 5 月第 1 次印刷

定 价：35.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zltz@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。服务热线：（010）88258888。

前言

Foreword

我清楚地记得第一次没有买 Chris Date 的书的情形。注意，我说的是“没有买”。那大概是 1991 年底或 1992 年初的样子。当时我在萨吉诺市密歇根州商业街的一家小店里寻找一本 SQL 方面的书。我找到了一本，作者是 C.J.Date，而这个人我根本就没有听说过。

那本书看起来还不错，内容非常有条理，我当时有点想买，但是……价格太贵！我当时站在那儿，手里拿着这本书，虽然该书只有大概 1/4 英寸厚并且不到 200 页，价格却要 30 美元之多，这个价格，对当时的我来说有点贵。我在内心斗争了一会儿，然后，不情愿地把书放回原处走开了。

那真是一个天大的错误！差点就影响到了我的职业生涯，而我当时竟然还意识不到这点。

后来我补偿了自己，劝服老板为我买那本书，尽管他们花了一个月的时间来考虑是否订购此书。书到手后，我读了好几次。对我来说，这本书最好和最有启发意义的一部分就是附录，其中，Chris 批评了当下流行的 SQL 语言。

我从 Chris 那里学到了不少东西，从那以后，我倾心于数据库工作，痴迷于像 SQL 这样的描述性语言。我只须描述我想要的结果，然后数据库引擎就会为我找出那些结果。同时，我也更加了解 Chris，知道他在 Codd 早期研究关系型理论时所扮演的重要角色。

几年来我一直在订阅《数据库编程和设计》杂志，唯一的理由是 Chris 在里面有一个专栏。

我从 Chris 那里获得的有关 SQL 的知识（始于那本没有买的薄而贵的小书）在这些年有了一系列的连锁反应。那时所积累的数据库入门知识，在今天成为了我事业上的一个转折点，而 Chris 的清晰和严谨，对我在总体上把握 SQL 语言和数据库技术方面，产生了深远的影响。自此之后，SQL 成为我最喜欢编写、学习，甚至近期最喜欢撰写的语言。

我不敢想像如果当时手头没有 Chris 的书，我今天的职业生涯将会怎么样。不要犯我当年犯的错误。不要把此书再次放回到书架上。阅读它！向那些可以帮助你发现和提炼你工作所需的关系型模型的人学习。你也许不会赞同 Chris 所说的一切（你也不需要这样做），但是一定要理解他所说的。花时间去理解他对今天关系型模型的看法；花时间去理解他认真提出的关于开发商在产品里如何实现模型的问题。做了这些事之后，你将会发现你比那些没有花时间打基础的同行们高出一头。

如果我可以给你什么建议，那就是保持学习的热情。我非常荣幸能编辑这本由 Chris 所撰写的书，这是我从来没有想到的荣誉。我从与 Chris 的交谈以及他的手稿中获益匪浅。我是如何幸运地接到这个工作的呢？这是我和 Oracle-L（译者注：Oracle 最有名的专家邮件列表）邮件列表中的几个非常聪明的同事，在激烈讨论有关 SQL 子查询（subquery）优化的意外收获。甚至更为间接的，是因为很多年前我第一次阅读 Chris 作品的机缘，促成了我为这本书工作。对知识的好奇心和热爱使我在事业上有了长足的进步，远超出我的想像，这些对你也同样适用。

— Jonathan Gennick

2005 年于密歇根，穆尼星（Munising）

序

Preface

在数据库领域的不同职位上工作了若干年之后，我突然意识到，确实有必要写本书。用一种没有被现有产品、商业手段或是 SQL 标准中各种古怪做法所影响的方式，为从业者（而不是新手）解释关系理论的基本原理。我写这本书正是为了满足这种需求。我预期的读者应该是富有经验的数据库开发者，或是那些能诚实地承认自己没有深入理解所擅长领域的基础理论的数据库专家，而这些都是应该和能够理解的。我说的理论实际上就是关系数据模型，事实上，这一理论的基本概念都是非常简单的，但同时也被广泛地误解或是没有得到正确的评价，或是兼而有之。事实上，它们看上去并没有被完全理解。比如，下面是一些关系方面的问题，看看你可以回答多少？

1. 1NF 到底是什么？
2. 关系和谓词之间的联系是什么？
3. 什么是语义优化（semantic optimization）？
4. 什么是连接依赖（join dependency）？
5. 半差运算为何重要？
6. 为什么延迟完整性检查是没有意义的？

7. 什么是关系变量?
8. 什么是无损分解?
9. 一个关系可以有值为关系的属性吗?
10. SQL 与关系模型之间的区别是什么?
11. 为什么信息法则 (Information Principle) 是重要的?
12. XML 如何满足关系模型?

这本书给出了这些以及许多与之相关问题的答案。总体上,此书可以帮助数据库从业者深入理解关系理论,同时可以帮助他们在日常的工作中很好地运用这些理论。

本书的特点

Amal Makadia: This Book Different?

我必须在此首先说明,这本书谈及的新技术非常少,其中的大部分,我在以前的书上或其他的出版物上都说过。最近我发现仍然有再说一遍的必要。但是这次我打算换一种不同的说法:次序不同、节奏不同、风格和处理也不尽相同,同时针对的读者群也不同(待会儿再说这最后一点)。因此,即使部分材料以前在其他多个地方以某种形式出现过,我也认为这完全是一本新书。当然,本书的某些部分不可避免地和我在其他地方写的一些材料有些相似,因为这些材料都是来自于同样的地方,即:我的大脑以及多年以来我在研讨会上讲授这些素材的教学经验。但是这并不是一种直接的复制粘贴,直接地拿来并不是我的本意。不过,我有意识地重新使用了许多老的例子,因为这些经过多年锤炼的例子正好能够阐明我希望表达的观点。

我再说说关于本书目标读者的差异。正如已经指出的那样,先前我已经在数据库技术领域出版了几本书。因此,这一本书有何不同之处?尤其是,它和另外一些现有的书籍相比,有竞争吗?

在我看来,后一个问题的答案是否定的。我有两本出版于 Addison-Wesley 的书,它们第一眼看上去好像是下面两本书的竞争者。

- 《数据库系统导论》第 8 版, (2004) (译者注: *An Introduction to Database Systems*)。
- 《数据库、类型与关系模型:第三次宣言》第 3 版 (和 Hugh Darwen 一起合著, 2006)。

不过,第一本书,我将其称之为“导论”,实际上它覆盖了整个数据库领域,而不仅仅是关系模型。它首先是作为一本大学教材,假定其读者没有任何数据库知识或经验,其文体比那些现有的书籍更为正式,很适合做教材。第二本是对 Hugh Darwen 和我所著的《未

来数据库系统的基础：第三次宣言》第 2 版 (Addison-Wesley, 2000) 一书的大规模重写。这是一本高级 (研究生级别) 教材, 它与第一本相比更加正式, 并非更加简洁。尽管有一些明显重复的主题存在; 但是, 我实在看不出在这三本书中, 任何一本和另外两本存在什么竞争。

另一个显著的不同点是, 本书主要是用来自学的 (尽管其中部分内容你可能希望和你的朋友或同事以及一起工作的人讨论), 其中还有一些练习, 来帮助你巩固学习的内容; 当然, 并没有义务说一定要做这些练习, 但是我认为至少着手做一部分会是一个很好的主意。答案通常给出了关于这一主题的更多信息, 你可以以在线的方式找到答案: <http://oreilly.com/catalog/databaseid>。

当我谈到可能竞争的话题时, 我应该提到我的另外两本书 (第一本出版于 Addison-Wesley, 另外一本出版于 Morgan Kaufmann)。

- 《数据库关系模型：回顾与分析》(2001),
- 《时态数据和关系模型》(与 Hugh Darwen 和 Nikos A. Lorentzos 合著, 2003)。

在我看来, 上面说的第一本书是对本书的补充, 该书以一种完全非正式的风格回顾和分析了一系列 Ted Codd 所写的论文, 在这些论文中, 他第一次将关系模型引入到我们的日常世界。关于第二本书, 正如其标题所显示的, 不是关于关系理论, 而是关于这种理论的一个特定应用。虽然该书的第 1 章的确是关系型模型的概述, 看起来可能与本书略微有些竞争, 但我并不这么认为。

综上所述, 尽管以前我已经在其他不同的地方写过这些主题当中的大部分内容, 有时也不可避免地引用一些类似的术语, 但我认为到目前为止, 至少我自己或目前所知的任何其他作者, 都不曾以类似本书的方式总结并探讨这些主题。

预备知识

Further Preliminaries

我需要说一下几个预备知识。正如已经说过的, 我将在此使用某些在我的其他书籍和文章中出现过的同样的例子, 尤其是著名的供应商和零件数据库实例。对不止一次地将这个“老兵”拖出来我深表歉意, 但是我前面说过, 这个例子经过精心设计, 刚好能够正确地阐明我想要表达的观点。

其次, 我自己对关系模型的理解也积累了很多年, 并且一直在继续。这本书表达了我对这一主题最新的思考; 因此, 如果你发现这本书和提到的某本书之间有任何矛盾之处 (的确有一些), 请以这本书中的观点和处理方式作为参照依据。不过我得赶紧补充说明, 这些差异大部分都具有比较细微的特性; 所以, 我一直在小心处理新的术语、概念与先前的术语和概念之间的联系, 一旦我觉得有必要时就会这么做。

第三，我要谈论的当然是理论问题，对我来说这是一个信条：理论是实用的。我特别地指出这一点是因为，很多人看上去相信相反的观点：即，如果某种东西是理论化的，那它就不是实用的。但是实际情况是，理论——至少，我在这里谈到的关系理论——的确是非常实用的。这一理论的目的不仅仅是纯粹为了理论而理论，而是允许我们来构建百分之百实用的系统。理论的每个细节，都是为了纯粹的实用。的确，关系理论的绝大部分都不仅仅是实用，它还直截了当，简单易懂，非常有用，另外，还相当有趣（我希望本书表现了这点）。

（实际上，为了证明前述论点，我们只需要看看关系模型就足够了。确实，没有必要去辩护“理论是实用”的观点，一个总计数十亿美元产值的行业就建立在这个伟大的理论观点之上。但是我假定极端主义者的立场是“是的，但是理论知识现在能够为我做什么？”换句话说，我们这些认为理论知识的确很重要的人，就必须不断地向我们的批评者提出证明，这也是我认为有必要写作本书的另一个原因）。

还有另外一点：标准的“关系”语言是 SQL，当然，我假定你已经相当熟悉这门语言，也具备基本的数据库概念。然而，正如你即将看到的，我在本书中对 SQL 作了相当多的批评。一个不幸的事实是：SQL 在很多方面对关系模型的支持完全令人失望；它存在着大量不合情理的错漏和偏离（这就是当我第一次提到它时，说它是“关系的”要加引号的原因）。但正因为如此，这也是你要学习关系模型的原因之一！因为 SQL 对模型的支持是如此有限，它给了你一条绳索让你吊死自己；因此你需要知道关系理论以免陷入圈套，也就是说，你需要知道关系理论，以强迫自己遵循各种 SQL 应该帮你执行却忽略了的的原则。重复数据行是一个很好的例子：SQL 允许重复行，但是关系模型不允许。因此你需要知道为何关系模型不允许这个，才能理解 SQL 中不采用这个功能的重要性。正如审核本书原始提案的 Stephane Faroult 所写的那样：“当你有一些实践知识，你就会意识到理论知识是不可避免的”。

谈到 SQL，顺便说一下，请注意我在整本书中使用的术语 SQL，是专门表示这种语言的标准版本，而不是某些具有特定规定的产品（当然，某些明确声明的除外）。而且我遵循了这样的发音标准“ess cue ell”，而不是“seque”（尽管后者在该领域中更为常见），因此会因为元音发音的差异，写成“an SQL table”，而不是“a SQL table”。

最后我想说的是，你可以获得一个基于本书内容的为时一天的在线研讨会。更多细节请参见 <http://www.dbdebunk.com> 或者 <http://www.thethirdmanifesto.com>。

本书的约定

Conventions Used in This Book

下面是本书中所用的印刷体上的约定：

斜体

用来表示变量，如 x ，在本书正文的讨论过程中，被用来代替其他某个值。

<http://www.oreilly.com/catalog/databaseid/> (原书)

<http://www.oreilly.com.cn/book.php?bn=978-7-121-04000-9> (中文版)

博文视点资讯有限公司 (武汉分部)

湖北省 武汉市 洪山区 邮科院路特 1 号 湖北信息产业科技大厦 1406 室

邮政编码: 430074

电话: (027) 87690813 传真: (027) 87690813 转 817

网页: <http://bv.csdn.net>

读者服务信箱:

sheguang@broadview.com.cn (读者信箱)

broadvieweditor@gmail.com (投稿信箱)

致谢

Acknowledgments

感谢本书编写过程中直接或间接帮助过我的人, 特别地, 我要提到评论家 Stephane Faroult、Jonathan Gennick、Lex de Haan、Anthony Molinaro、Peter Robson 和 Michael Wener 对我早期的手稿提出的有益意见。还要感谢 Nagraj Alur 和 Hugh Darwen 提供的多方面的技术讨论。接下来, 我要感谢我的妻子 Lindy 在本书的写作过程中和近几年来我的其他数据库工作中的一贯支持。最后, 感谢 O'Reilly 的所有工作人员, 尤其是 Jonathan Gennick 和 Genevieve d'Entremont, 在本书的整个出版过程中给予的鼓励、贡献和支持。

——C. J. Date

2005 年于加利福尼亚, 席德布格 (Healdsburg)

O'Reilly Media, Inc. 介绍

为了满足读者对网络和软件技术知识的迫切需求，世界著名计算机图书出版机构 O'Reilly Media, Inc. 授权电子工业出版社，翻译出版一批该公司久负盛名的英文经典技术专著。

O'Reilly Media, Inc. 是世界上在 Unix、X、Internet 和其他开放系统图书领域具有领导地位的出版公司，同时也是在线出版的先锋。

从最畅销的《The Whole Internet User's Guide & Catalog》(被纽约公共图书馆评为 20 世纪最重要的 50 本书之一) 到 GNN (最早的 Internet 门户和商业网站)，再到 WebSite (第一个桌面 PC 的 Web 服务器软件)，O'Reilly Media, Inc. 一直处于 Internet 发展的最前沿。

许多书店的反馈表明，O'Reilly Media, Inc. 是最稳定的计算机图书出版商——每一本书都一版再版。与大多数计算机图书出版商相比，O'Reilly Media, Inc. 具有深厚的计算机专业背景，这使得 O'Reilly Media, Inc. 形成了一个非常不同于其他出版商的出版方针。O'Reilly Media, Inc. 所有的编辑人员以前都是程序员，或者是顶尖级的技术专家。O'Reilly Media, Inc. 还有许多固定的作者群体——他们本身是相关领域的技术专家、咨询专家，而现在编写著作，O'Reilly Media, Inc. 依靠他们及时地推出图书。因为 O'Reilly Media, Inc. 紧密地与计算机业界联系着，所以 O'Reilly Media, Inc. 知道市场上真正需要什么图书。

仅以此书
纪念
E.F. Codd

热衷于实践而不要理论的人就如同一个水手上了一条没有舵和罗盘的船，拿不稳该往哪里航行。实践应当永远建立在正确的理论上。

——列奥纳多·达·芬奇(1452~1519)

人们的烦恼并不是他们不知道什么，而是他们对错误的东西知道得太多。

——乔希·比林斯(1818~1885)

目录

TABLE OF CONTENTS

前言	I
序	III
第 1 章 绪论	1
术语说明	2
原理, 而非产品	3
回顾原模型	4
模型和实现	11
关系的属性	14
关系和关系变量	17
值和变量	19
小结	20
练习	21
第 2 章 关系和类型	23
域限制比较	24
数值原子性	29
那么, 什么是类型	32
标量和非标量类型	35
小结	36
练习	37
第 3 章 元组和关系	41
什么是元组	41
一些重要的推论	44
什么是关系呢	45
更多重要的推论	47
为什么重复元组是被禁止的	48
为什么不允许空值	53
TABLE_DUM 与 TABLE_DEE	56
小结	57
练习	58

第 4 章 关系变量	61
集合式更新	62
深入讨论候选码	63
深入讨论外码	65
深入讨论视图	67
关系变量和谓词	72
深入讨论关系与类型	75
小结	77
练习	78
第 5 章 关系代数	81
深入探讨封闭性	83
基本运算符	86
求 SQL 表达式的值	93
扩展与汇总	95
分组和分组还原	99
表达式转换	100
关系比较	103
深入探讨关系赋值	106
ORDER BY 运算符	108
小结	109
练习	110
第 6 章 完整性约束	115
类型约束	115
数据库约束	119
事务	121
为什么数据库约束必须是即时检查的	122
但是，有的检查不是必须延迟吗	125
约束和谓词	127
其他问题	129
小结	131
练习	132

第 7 章 数据库设计原理	135
设计原理的地位	136
函数依赖和 Boyce/Codd 范式	138
连接依赖和 5NF	144
浅谈规范化	150
正交性	153
一些关于物理设计的观点	156
小结	158
练习	159
第 8 章 什么是关系模型	163
关系模型的定义	164
关系模型的目标	168
一些数据库原理	168
关系模型和其他模型比较	169
剩下要做的是	172
小结	176
练习	177
附录 A 关于逻辑学的一些知识	181
附录 B 进一步阅读的建议	199
索引	203

任何学科的专家都必须熟悉他所在领域的基础知识。因此，如果你是一位数据库专家，你必须熟悉关系模型，因为关系模型就是数据库领域的基础知识（或者说是构成基础知识中的一个很大部分）。现在，所有数据库管理方面的课程，不管是学术上用还是商用，都至少要宣称是在用关系模型来讲课，但如果依成效而言，大部分教学看上去都是非常糟糕的。在整个数据库领域，关系模型的确没有被很好或很广泛地理解。这里有一些造成这种状况的可能的原因。

- 在真空中讲授模型。这就是说，对初学者而言，要么很难看清和实际应用相关的地方，要么很难理解该模型所能解决的问题，或者对这两点都不清楚。
- 讲课的人自己也并未完全理解或意识到关系模型的重要性和意义。
- 根本就没有讲授关系模型（实践中多数情况也是如此）取而代之的是 SQL 语言或者和它相关的某种特定的方言，如 Oracle 方言（译者注：PL/SQL 语言）。

因此，这本书针对的是数据库专业人员，特别是商业数据库的实践者。与他们应该掌握