

第2版

SQL与关系数据库理论

——如何编写健壮的SQL代码

SQL and Relational Theory: How to Write Accurate SQL Code

O'REILLY®



机械工业出版社
China Machine Press

C. J. Date 著

单世民 何英昊 许侃 译

第2版

SQL与关系数据库理论

——如何编写健壮的SQL代码

C. J. Date 著

单世民 何英昊 许侃 译

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo

O'Reilly Media, Inc. 授权机械工业出版社出版

机械工业出版社

图书在版编目 (CIP) 数据

SQL与关系数据库理论: 如何编写健壮的SQL代码: 第2版/ (英) 戴特 (Date, C. J.) 著; 单世民等译. —北京: 机械工业出版社, 2014.4
(O'Reilly精品图书系列)

书名原文: SQL and Relational Theory: How to Write Accurate SQL Code, Second Edition
ISBN 978-7-111-46154-8

I. S… II. ①戴… ②单… III. ①SQL语言 ②关系数据库系统 IV. TP311.132
中国版本图书馆CIP数据核字 (2014) 第050877号

北京市版权局著作权合同登记
图字: 01-2012-1677号

©2012 by C.J. Date.

Simplified Chinese Edition, jointly published by O'Reilly Media, Inc. and China Machine Press, 2014. Authorized translation of the English edition, 2012 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.

英文原版由O'Reilly Media, Inc. 出版2012。

简体中文版由机械工业出版社出版 2014。英文原版的翻译得到O'Reilly Media, Inc.的授权。此简体中文版的出版和销售得到出版权和销售权的所有者——O'Reilly Media, Inc.的许可。

版权所有, 未得书面许可, 本书的任何部分和全部不得以任何形式重制。

封底无防伪标均为盗版

本书法律顾问

北京大成律师事务所 韩光/邹晓东

书 名/ SQL与关系数据库理论 (第2版)

书 号/ ISBN 978-7-111-46154-8

责任编辑/ 吴怡

封面设计/ Karen Montgomery, 张健

出版发行/ 机械工业出版社

地 址/ 北京市西城区百万庄大街22号 (邮政编码 100037)

印 刷/ 藁城市京瑞印刷有限公司

开 本/ 178毫米×233毫米 16开本 27印张

版 次/ 2014年8月第1版 2014年8月第1次印刷

定 价/ 129.00元 (册)

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010)88378991; 88361066

购书热线: (010)68326294; 88379649; 68995259

投稿热线: (010)88379604

读者信箱: hzsj@hzbook.com

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

Those who are enamored of practice without theory are like a
pilot who goes into a ship without rudder or compass
and never has any certainty where he is going.

Practice should always be based upon a
sound knowledge of theory.

——Leonardo da Vinci (1452–1519)

The trouble with people is not that they don't know
but that they know so much that ain't so.

——Josh Billings(1818–1885)

Languages die...
mathematical ideas do not.

——G.H. Hardy (1877–1947)

Unfortunately, the gap between theory and practice
is not as wide in theory as it is in practice.

——Anon

These are my principles.
If you don't like them, I have others.

——Groucho Marx (1890–1977)

There is no royal road to geometry.

——Euclid(c. 365–275 BCE), attrib



谨以本书献给所有认可本书价值的人们，尤其是令人怀念的Lex de Haan!

作者简介

C. J. Date是专注于关系数据库技术的独立作家、讲师、研究人员和顾问。他因其《An Introduction to Database Systems》（第8版，Addison-Wesley，2004）一书而闻名于世（在本书写作时该书已经卖出了850 000本，并被全世界几百所学院和大学使用）。他也是很多数据库管理方面书籍的作者，最新的作品包括：

- Addison-Wesley出版社： *Databases, Types, and the Relational Model: The Third Manifesto*，第3版（和Hugh Darwen共同撰写，2006年）
- Apress出版社： *Date on Database: Writings 2000–2006* (2006)
- Trafford出版社： *Logic and Databases: The Roots of Relational Theory* (2007)
- Apress出版社： *The Relational Database Dictionary, Extended Edition* (2008)
- Trafford出版社： *Database Explorations: Essayson The Third Manifesto and Related Topics*（和Hugh Darwen共同撰写，2010年）
- Ventus出版社： *Go Faster! The Trans Relational Approach to DBMS Implementation* (2011年)

还有一本书，*Normal Forms and All That Jazz: A Database Professional's Guide to Database Design Theory*（本书的配套书）也已经出版。

Date先生于2004年成为计算机行业名人堂（the Computing Industry Hall of Fame）的一员。在将复杂的技术主题用清晰易懂的方式进行解释方面，Date先生享有首屈一指的美誉。

译者序

作为一名讲授数据库课程的高校教师，本人对于我国的数据库教学工作具有直观感受，对于我国数据库行业的发展亦有浅薄的了解。窃以为，本书最有价值之处，也是最有吸引力的地方在于其集数十年研究之大成，系统化地提出了基于关系理论编写逻辑正确的SQL语句的指导性方法并对SQL中的众多语焉不详之处进行了严谨的、思维缜密的学术探讨。这些内容对于中国的读者，尤其是对于数据库行业的从业人员和教育战线的数据库相关课程教师而言具有很重要的意义，能指导人们编写在理论上更“站得住脚”的SQL查询，也有助于使得数据库的教学更为清晰明确、准确无误。

虽然现在有大量“NoSQL”流派的新型数据库技术在大数据环境下得到了快速发展并受到大量关注，但是我相信关系理论和SQL不但不会由此退出历史舞台，反而会得到更进一步的丰富和发展。当然，这需要大量具有扎实理论功底数据库从业人员的不懈努力。对此，本书具有非常积极的推动作用，它即可以帮助初学者答疑，也可以帮助专业人员从另一种角度思考自身的工作，还有助于使关系理论的研究成果走出象牙塔，成为众多数据库专业人员的“攻坚利器”。

在翻译本书一年多的过程中，对C. J. Date先生严谨的研究精神和独特的写作风格深深折服，也让作为译者的我经历了“又爱又恨”的体验。“爱”的是难得能有如此大家以如此严谨的态度将SQL与关系理论进行如此详尽深入的对比、讨论、思考，而我又如此幸运的能够有机会将此大作翻译给国内的读者阅读；“恨”的是其写作风格将我这个工科男的文字功底不足彻底地暴露出来，让我在众多个日夜中苦苦思索如何翻译才能“不丢失信息”地表达原文。因此翻译时间一拖再拖，经历了痛苦的历程也收获颇丰，终于拿出一个相对满意的“作业”以飨读者，希望能为中国的数据库行业发展略尽绵薄之力。

本书的翻译及修订工作主要由本人完成，何英昊女士和许侃先生也参与了部分翻译及修订工作，提出了很多中肯的评阅意见。在此，要特别感谢我的父母，我的工作离不开你们的关心支持。我还要感谢我的同事们（尤其是刘宇老师），如果没有你们的支持，我是没有这么多时间和精力来完成此项工作的。最后，但也同样重要的，我还要特别感谢机械工业出版社华章公司的编辑们，没有你们的宽容与支持，这本译著是不可能完成的。

虽然在翻译工作中尽量追求准确无误，但本人的资历和学识毕竟相当有限，翻译工作出现不尽人意乃至谬误之处恐怕在所难免，敬请各位同行、专家及读者不吝赐教。

目录

第1版前言	1
第2版前言	6
第1章 做好准备	7
1.1 关系模型被严重地误解了	7
1.2 关于术语的一些说明	8
1.3 原理而非产品	10
1.4 原始模型回顾	11
1.5 模型vs.实现	18
1.6 关系的性质	21
1.7 基关系vs.导出关系	24
1.8 关系vs.关系变量	26
1.9 值vs.变量	28
1.10 小结	28
1.11 练习题	29
第2章 类型和域	31
2.1 类型和关系	31
2.2 相等性比较	32

2.3 数据值原子性	37
2.4 类型是什么	40
2.5 标量类型vs.非标量类型	43
2.6 SQL中的标量类型	45
2.7 SQL中的类型检查和型转	47
2.8 SQL中的字符序	48
2.9 SQL中的行类型和表类型	50
2.10 小结	52
2.11 练习题	52
第3章 元组、关系、行、表	55
3.1 元组是什么	55
3.2 SQL中的行	59
3.3 关系是什么	61
3.4 关系及其主体	63
3.5 关系是n维的	64
3.6 关系比较	64
3.7 TABLE_DUM和TABLE_DEE	65
3.8 SQL中的表	66
3.9 SQL中的列命名	68
3.10 小结	70
3.11 练习题	70
第4章 不要重复，不要null	72
4.1 重复有什么问题	72
4.2 重复：深入讨论	76
4.3 在SQL中避免重复	77
4.4 null有什么毛病	79
4.5 在SQL中避免null	82
4.6 对外连接的说明	84
4.7 小结	84
4.8 练习题	85

第5章 基关系变量和基表	90
5.1 更新是集合级别的	91
5.2 关系赋值	93
5.3 关于候选键的更多内容	97
5.4 关于外键的更多内容	99
5.5 关系变量和谓词	102
5.6 关系 vs. 类型	104
5.7 练习题	106
第6章 SQL和关系代数I：原始运算符	109
6.1 一些预备知识	109
6.2 关于闭包的更多内容	112
6.3 限制	114
6.4 投影	115
6.5 连接	116
6.6 并、交和差	120
6.7 哪些运算符是基本运算符	123
6.8 逐步形成表达式	123
6.9 关系表达式到底表示什么	125
6.10 计算SQL表表达式	126
6.11 表达式变换	127
6.12 属性名依赖	130
6.13 练习题	132
第7章 SQL和关系代数II：附加运算符	135
7.1 排他并	135
7.2 半连接和半差	136
7.3 扩展	137
7.4 映像关系	139
7.5 除	142
7.6 聚集运算符	143
7.7 再议映像关系	148

7.8 汇总	149
7.9 再议汇总	154
7.10 分组、去分组和关系值属性	155
7.11 “WHAT IF” 查询	160
7.12 对于递归的说明	162
7.13 ORDER BY是怎么回事	166
7.13 练习题	167
第8章 SQL与约束	171
8.1 类型约束	172
8.2 SQL中的类型约束	175
8.3 数据库约束	176
8.4 SQL中的数据库约束	180
8.5 事务	182
8.6 数据库约束为什么必须立即检查	183
8.7 不是有些检查必须延迟进行吗	185
8.8 约束与谓词	187
8.9 各种问题	189
8.10 练习题	191
第9章 SQL与视图	194
9.1 视图是关系变量	195
9.2 视图和谓词	198
9.3 检索运算	199
9.4 视图和约束	200
9.5 更新运算	204
9.6 视图的作用	212
9.7 视图和快照	213
9.8 练习题	214
第10章 SQL与逻辑	217
10.1 为什么需要逻辑	218

10.2 简单命题和复合命题	219
10.3 简单谓词和复合谓词	224
10.4 量词化	226
10.5 关系演算	229
10.6 关于量词化的更多内容	236
10.7 一些等价关系	243
10.8 小结	246
10.9 练习题	246
第11章 使用逻辑表述SQL表达式	249
11.1 一些变换法则	249
11.2 例1: 逻辑蕴涵	252
11.3 例2: 全称量词化	252
11.4 例3: 蕴涵和全称量词化	254
11.5 例4: 相关子查询	255
11.6 例5: 命名子表达式	257
11.7 例6: 关于命名子表达式的更多内容	260
11.8 例7: 处理模糊性	261
11.9 例8: 使用COUNT	263
11.10 例9: 连接查询	263
11.11 例10: 唯一量词化	264
11.12 例11: ALL或ANY比较	266
11.13 例12: GROUP BY和HAVING	270
11.14 练习题	271
第12章 关于SQL的其他主题	273
12.1 SELECT *	273
12.2 显式表	274
12.3 名称限定	274
12.4 区间变元	275
12.5 子查询	277
12.6 “可能非确定性”表达式	280

12.7 空集合	281
12.8 简化的BNF语法	281
12.9 练习题	284
附录A 关系模型	287
附录B SQL背离关系模型之处.....	307
附录C 处理信息丢失的关系方法	309
附录D Tutorial D语法.....	322
附录E 本书建议汇总	325
附录F 练习答案	328
附录G 深入阅读建议	407

第1版前言

SQL无所不在。尽管如此，SQL却难以使用：SQL是复杂的，令人困惑且容易出错（我敢说比它的卫道士所声称的更容易出错）。所以，为了写出你能确信的精确SQL代码（意思是它准确地做到了要求它做的事），你就必须遵从一些适当的准则。而本书的主题就是：关系化地使用SQL，这就是你需要的准则。不过，这到底意味着什么？难道SQL不是关系化的吗？

没错，SQL确实是用于关系化数据库的标准语言，但是这个事实本身并不能让它成为关系化的语言。让人感到不幸的真相是：SQL在太多的方面背离了关系理论；重复行和null就是两个明显的例子，而它们绝不是仅有的两个问题。结果就是，可以说SQL语言为你提供了“上吊用的绳子”。所以，如果你不想“上吊”，就需要理解关系理论（是什么以及为什么），需要知道SQL对于关系理论的背离之处，还需要知道怎样避免背离之处所造成的这些问题。总的来说，你需要关系化地使用SQL。这样，你就能把SQL当作是真正关系化的语言来运用，而且可以享受使用真正的关系系统所带来的好处。

如果每个人都已经关系化地使用SQL了，就不再需要像本书这样的书籍了。但问题是，情况并非如此。我看到当前的SQL使用中存在着大量的糟糕实践，甚至看到有的（本应具有更好认知的）作者在课本或类似出版物中推荐这些糟糕实践。事实上，对此方面文献的回顾是相当令人沮丧的。关系模型诞生于1969年，然而直到40多年后的今天，它似乎还是没有被数据库领域中的大多数人所充分理解。所以，本书将关系模型本身作为一个组织原则，其部分原因也来自于此。本书深入讲解关系模型的诸多特性，并说明在每种情况下为了遵守这些特性怎样最好地使用SQL。

预备知识

我假设你是熟悉SQL的数据库专业技术人员。具体地说，我假设你具有对于SQL标准（或至少是对于一个SQL产品）的工作知识。然而，我并不假设你深入理解关系理论（尽管我确实希望你明白：关系模型是个好东西，并且只要可能就应坚持关系模型）。所以，我会详细描述关系模型的各个特性以免误解。我还会说明怎样使用SQL去遵守那些特性。不过，我不会证明所有特性；相反，我假设你在数据库方面已经具有足够的经验，能够理解为什么键的概念是有意义的，为什么有时需要做连接，而又为什么需要支持多对多联系，等等。（如果要包括这些证明，那么本书就会大变样，比现在的厚得多。况且，那样的书已经有了。）

我说过，我期望你熟悉SQL。虽然如此，我还是会详细“解释SQL的某些方面，尤

其是实践中不常碰到的那些内容。比如，SQL的“可能非确定性表达式”（possibly nondeterministic expression）。参见第12章。

深入数据库

本书基于并要超越以前的《Database in Depth: Relational Theory for Practitioners》（O'Reilly Media Inc., 2005）一书。在那本书中，我的目标如下（摘自前言）：

在数据库圈子里工作多年之后，我终于意识到：需要有一本书，用一种未被现存产品、商业实践或SQL标准的怪癖所污染的方式为专业人员（而不是新手）解释关系理论的基本原理。我写这本书就是为了满足这个需要。我所期望的读者是有经验的数据库专业人员：他们足够诚实，勇于承认自己并不理解自身领域所基于的以及自己应该知道的那个理论。

“那个理论”所指的当然就是关系模型，尽管该理论的基本思想都非常简单，但这些思想还是被广泛地误传误评。事实上，它们似乎就没被真正理解过。比如，下面就有一些关系模型的问题，你能回答出多少？^{注1}

1. 第一范式到底是什么？
2. 关系和谓词之间的联系是什么？
3. 语义优化是什么？
4. 映像关系是什么？
5. 为什么半差（semidifference）重要？
6. 为什么延迟完整性检查毫无意义？
7. 关系变量是什么？
8. 前束范式是什么？
9. 关系能有一个其取值为关系的属性吗？
10. SQL是关系完备的吗？
11. 为什么“信息原理”重要？
12. XML怎样才符合关系模型？

本书对这些问题以及相关问题进行了解答。总的来说，本书就是为了帮助数据库专业人员去深入理解关系理论，并在日常的专业工作中很好地运用这些对于关系理论的深入认知。

正如这段引文的最后一句所言，我曾经希望本书读者不需要进一步的协助就能够自己运用关系理论的思想。不过，我随后就意识到：与流行观点恰恰相反，要在不违反关系理论原理的情况下使用SQL这种语言可绝不是一眼就能看出来的事。所以，我决定将先前的书

注1：由于这些问题中的一些不太重要，我替换了该列表中的一些问题。

进行扩展，以便包括对于该问题（我指的是怎样关系化地使用SQL）明确具体的建议。本书的目标仍然和先前的书一致：帮助数据库专业人员深入理解关系理论并在其日常的专业工作中很好地运用这些对于关系理论的深入认知。不过，我试着让书中的内容更好消化一些，更容易应用。总之，我收录了大量与SQL相关的资料（也正因为如此，本书才会比先前的那本书在篇幅上有所增加）。

对正文的进一步说明

我需要进一步说明几点。第一，我对于关系模型的理解在多年来是不断演进的，并且还将继续下去。本书代表了我对该主题的最新思考，所以，如果你发现这本书和我的其他书（尤其是本书所要超越的那本书）之间存在什么技术上的差异（确实有几个），请以这本书为准。不过，绝大部分的差异都是相当微小的，而且，我在必要的时刻总会将新的术语及概念和先前的术语及概念联系起来。

第二，我会讲到理论，不过，我的信条是：理论是实用的。之所以我要明确指出此点，是因为有太多的人不这么认为，即很多人认为理论化的东西是不实用的。但真相是：理论真的是绝对、非常实用的。关系理论的目的不是理论本身，而是让我们能够构建完全实用的系统。关系理论中的每个细节都缘自于切实的实用原因。正如Stéphane Faroult（先前那本书的评审者）所写：“一旦你有了一点实践，你就会意识到不懂得关系理论是不行的。”而且，关系理论不仅是实用的，而且是基础的、直截了当的、简单的、有用的，（正如我希望在本书中证明的那样）也会是有趣的。

当然，对于前述论题，我们就不必去找比关系模型本身还要更进一步的例证了。实际上，在我们这样的环境下（即建立在一个伟大的理论思想之上的几十亿美元规模的产业）真的应该不必为“理论是实用的”辩解。不过，我猜某些人的观点会是“没错，但理论为我做了什么呢？”总之，坚信理论是重要的人们必须不断地向批评者证明自己，这又是一个我认为需要本书的原因。

第三，前面提到过，本书会讨论相当多的SQL相关特性细节或关系模型细节。（对于和关系化不太相关的主题，本书有意进行了回避。比如，本书中关于事务的内容就不多。）自始至终，我都设法说清楚适用于SQL的讨论以及适用于关系模型的讨论。不过，我想强调的是：我并不想对SQL进行详尽全面的讨论。SQL太复杂了（更何况还为做同一个事情提供了太多的不同方法，而且还有那么多例外和特例），要想详尽全面地讨论它（我都怀疑是不是真的能做到这点）只会适得其反；而且，详尽全面的讨论肯定会增加本书的篇幅。所以，我设法只关注我所认为的最重要问题，而且尽可能简短地说明所选择的问题。我想说，如果你做了我告诉你的每件事，并且不做我没有告诉你的任何事，那么你基本上就会是安全的：你就是关系化地使用了SQL。不过，我的说法到底有没有道理，或者在多大程度上是有道理的必须由你来判断。

对于前文我还要补充说明的是：很遗憾，有些情况下就是不能关系化地使用SQL。比如，虽然关系模型将延迟完整性检查作为逻辑错误明确拒绝，但有的SQL完整性检查必须延迟（通常延迟到提交时刻）。尽管本书对在这种情况下如何处理给出了建议，但是我担心最终还是会“尽你的能力做好”。至少，我希望你会理解背离关系模型所带来的风险。

我还要说的是，本书所给的一些建议也不是特定于关系化的，而只是关于一般的优秀实践的——尽管有时会受（不太明显的）关系化方面的影响。避免型转就是这种情况的很好示例。

第四，请记住，除非明确声明，我在全书中使用的都是标准版本的SQL语言（而不是某些专属的方言）。尤其是，我和SQL标准一样假设其发音为“ess cue ell”而不是“sequel”（尽管后者在业界很普遍）。

第五，除非提示，否则本书是要按顺序阅读的（大多数章节都在某种程度上依赖于前面章节的内容，所以你不应该跳来跳去）。还有，每一章都有一套练习题。当然，不是必须做这些练习，但是我觉得试试某些练习总不是坏事。练习的答案（经常给出相应主题的更多信息）在附录F中给出。

第六，我有一些基于本书内容的研讨会记录可以使用，其详细内容参见www.justsql.co.uk/chris_date/chris_date.htm或www.thethirdmanifesto.com。其中一个研讨会的在线版本也是可用的，地址是<http://oreilly.com/catalog/0636920010005/>。

使用代码示例

本书旨在帮助你完成手头的工作。一般而言，你可以在自己的程序和文档中随意使用本书中的代码。除非原样引用大量的代码，否则你无须征得我们的许可。例如，在编写程序时引用了本书的若干代码片段是无须许可的。而销售或发行O'Reilly图书的示例光盘则须要许可。通过引用书中内容及示例代码的方式答疑解难无须许可，而将书中的大量示例代码加入到你的产品文档中则须要许可。

如果你在引用时注明出处（并非必须），我们将不胜感激。引用通常包含书名、作者、出版商及ISBN。例如：“*SQL and Relational Theory*, Second Edition, by C.J. Date (O'Reilly). Copyright 2012 C.J. Date, 9781449316402”。

如果你发现自己对示例代码的使用有失公允或违反了上述条款，敬请通过permissions@oreilly.com与我们取得联系。

评论及疑问

如果你想就本书发表评论或有任何疑问，敬请联系出版社：

美国：

O'Reilly Media, Inc.