

/THEORY/IN/PRACTICE

SQL与关系数据库理论

SQL and Relational Theory

O'REILLY®

C. J. Date 著
周成兴 等译

清华大学出版社

Copyright © 2009 by O'Reilly Media, Inc.
Authorized Simplified Chinese translation edition by O'Reilly Media, Inc., as published by Tsinghua University Press, 2010. Authorized translation of the original English edition, 2009 O'Reilly Media, Inc. The word "All rights reserved" including the rights of reproduction in whole or in part in any form.

SQL与关系数据库理论

C. J. Date 著
周成兴 等译

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Sebastopol • Taipei • Tokyo

O'Reilly Media, Inc. 授权清华大学出版社出版

清华大学出版社

Copyright ©2009 by O'Reilly Media, Inc.

Authorized Simplified Chinese translation edition, by O'Reilly Media, Inc., is published by Tsinghua University Press, 2010. Authorized translation of the original English edition, 2009 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.

本书之英文原版由O'Reilly Media, Inc.于2009出版。

本中文简体翻译版由O'Reilly Media, Inc.授权清华大学出版社于2010年出版。此翻译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, 未经书面许可, 本书的任何部分和全部不得以任何形式复制。

北京市版权局著作权合同登记 图字: 01-2009-5151号

本书封面贴有清华大学出版社防伪标签, 无标签者不得销售。

版权所有, 侵权必究。侵权举报电话: 010-62782989 13701121933

图书在版编目(CIP)数据

SQL与关系数据库理论/(美)戴特(Date, C. J.)著;周成兴等译.—北京:清华大学出版社, 2010.8

书名原文: SQL and Relational Theory: How to Write Accurate SQL Code

ISBN 978-7-302-22400-6

I. S… II. ①戴… ②周… III. ①关系数据库—数据库管理系统, SQL IV. ①TP311.138

中国版本图书馆CIP数据核字(2010)第061798号

责任编辑: 龙敬铭

封面设计: Jan Davis 张 健

责任校对: 徐俊伟

责任印制: 何 芊

出版发行: 清华大学出版社

地 址: 北京清华大学学研大厦 A 座

<http://www.tup.com.cn>

邮 编: 100084

社 总 机: 010-62770175

邮 购: 010-62786544

投稿与读者服务: 010-62795954, jsjic@tup.tsinghua.edu.cn

质 量 反 馈: 010-62772015, zhiliang@tup.tsinghua.edu.cn

印 刷 者: 北京富博印刷有限公司

装 订 者: 北京市密云县京文制本装订厂

经 销: 全国新华书店

开 本: 178×233 印 张: 26 字 数: 594 千字

版 次: 2010 年 8 月第 1 版 印 次: 2010 年 8 月第 1 次印刷

印 数: 1~4000

定 价: 53.00 元

产品编号: 032674-01

译者序

在关系数据库系统中，SQL已经成为了事实标准。SQL语言使用非常广泛，现在流行的所有数据库系统几乎都支持SQL。但是，SQL难以使用，它非常复杂，并且很令人费解。本书主要介绍了关系数据库的原理与SQL的内在关系，阐述了该如何把关系理论直接应用到SQL的应用中，并通过大量的例子以及对其背后理论解释，帮助读者编写出高效而精确的SQL代码。全书由12章和4个附录组成，每章后面都给出了一定的练习题，帮助读者巩固所学的内容，并在附录C中给出了这些练习题的答案，因此极大地方便了读者的学习和使用。

本书由周成兴主译，张长富、蔡建章、李匀、张建安、邓铁洪、徐君、李强、蒋思俊、杨文保、李强、苏辛、朱建波、徐志平、赵杰辉、傅祎、郭碧莲、郭洵、洪晓煜、黄宣达、江松波、柯渝、赖曲芳、廖阳、刘文红、李伟、郭涛、高磊、王振营、冯哲、韩毅、马以辉、李腾、邓卫、邓凡平、周云、董武、郑晓蕊、陈占军、倪泳智、黄虹、吕巧珍、裘蕾、金颖、韩毅、王嘉佳、吴建伟、贺军、宋雁等参与了本书的部分翻译工作，王玉玲对本书的翻译也提出了很多宝贵意见，在此一并表示感谢。

O'Reilly Media, Inc.介绍

为了满足读者对网络 and 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. 授权清华大学出版社，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，同时也是联机出版的先锋。

从最畅销的 *The Whole Internet User' 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。

目录

前言	1
第1章 设置场景	9
关系模式是很容易被误解的	9
一些术语的注释	10
法则，而不是产品	12
最初模式的一个回顾	13
模型和实现	20
关系的特性	23
基础关系和产生的关系	26
关系和关系变量	28
值和变量	29
结论	30
练习	31
第2章 类型与域	34
类型和关系	34
等式比较	35
数据值的原子性	41
什么是类型	44

标量和非标量类型	47
SQL中的标量类型	49
SQL中的类型检查和强制类型转换	50
SQL中的一些整序	52
SQL中的行和表类型	53
结论	55
练习	56
第3章 元组与关系、行和表	59
什么是元组	59
SQL中的行	63
什么是关系	65
关系和它们的关系体	66
关系是n维的	68
关系比较	68
TABLE_DUM和TABLE_DEE	69
SQL中的表	70
SQL中列的命名	72
结论	74
练习	74
第4章 无重复与无空值	76
重复有什么问题	76
复行：进一步讨论	81
在SQL中避免复行	82
空值有什么错	83
在SQL中避免空值	86
对外连接的评论	88
结论	89
练习	90
第5章 基本关系变量与基本表	94
数据定义	95

更新是集合级别的	95
关系赋值	97
更多关于候选键的内容	99
关于外键更多的讨论	101
关系变量和断言	104
关系和类型	106
练习	109
第6章 SQL与关系代数I：原始的操作符	112
一些预备知识	112
有关闭包的更多讨论	114
约束	117
投影	118
连接	119
并、交和差	123
哪些操作符是原始的	125
每次都形成表达式的一小步	125
关系表达式的意思是什么	127
求一个表表达式的值	128
表达式转换	129
属性名字上的依赖	132
练习	134
第7章 SQL与关系代数II：附加的操作符	137
半连接和半差	137
扩展	138
映像关系	140
除	143
聚集操作符	144
映像关系再讨论	148
概要	149
概要再论	154

分组和不分组	155
what if 查询	157
ORDER BY 如何	158
练习	159
第8章 SQL与约束	163
类型约束	164
SQL中的类型约束	167
数据库约束	168
SQL中数据库约束	172
事务	173
为什么数据库约束的检测必须是实时的	174
难道某些检测不是不得不被推迟吗	177
约束和谓词	180
各种各样的问题	182
练习	183
第9章 SQL与视图	187
视图是关系变量	188
视图和谓词	191
获取操作	192
视图和约束	194
更新操作	197
视图是为了什么	201
视图和快照	202
练习	203
第10章 SQL与逻辑	206
简单和复合的命题	207
简单和复合的谓词	210
定量	211
关系计算	216

更多的关于量化	222
一些等价形式	228
结论	231
练习	232
第11章 使用逻辑来形式化SQL表达式	235
一些转换规则	235
例子1: 逻辑隐含	238
例子2: 普遍的限定	238
例子3: 隐含和一般的限制	240
例子4: 相关的子查询	242
例子5: 命名子查询	243
例子6: 更多的关于命名子表达式	246
例子7: 处理不确定性	247
例子8: 使用COUNT	249
例子9: 连接查询	250
例子10: 唯一限制	251
例子11: ALL 或者 ANY比较	252
例子12: GROUP BY 和HAVING	256
练习	257
第12章 SQL主体杂项	258
SELECT *	258
显表	259
名字限定	259
范围变量	260
子查询	262
“可能的非确定”表达式	265
空集	266
SQL表表达式的BNF范式语法	266
练习	269

附录A 关系模式	271
附录B 数据库设计理论	286
附录C 练习答案	315
附录D 进一步阅读	393

前言

SQL语言使用非常广泛。但是SQL又是难以使用的：它非常复杂，并且很令人费解，而且更容易出错。为了使你确信写出的SQL代码是精确的（换句话说，这些SQL代码只做你希望做的事情，不多也不少），因此你就必须遵循适当的规则，这就是本书需要论述的问题，即关系地使用SQL语言就是你需遵循的规则。但这又意味着什么呢？难道SQL不是相关的吗？

当然在关系数据库中，SQL是事实上的标准，但事实上SQL自己并不是相关的。不幸的是，在很多方面SQL都是从关系理论分离开的，如重复的行和空行就是两个明显的例子，但并不是只有这两个例子。其结果就是，就像它给了你一根绳子把你自己捆住一样。因此如果你不想捆住自己，就需要理解关系理论（是什么和为什么）；你需要了解SQL是如何从这个理论分离开来的；并且需要了解如何避免可能产生的问题。总之，你需要关系地使用SQL，然后你才能把SQL看作是关系型的那样行事，这样你就能获得它事实上的好处，从而实现一个真正的关系系统。

现在，如果每个人都已经关系地使用SQL语言，那么类似于本书这种类型的书就不需要了，但事实上并非如此。相反地，在当前的SQL使用中，我看到了非常多的糟糕的例子。甚至在一些教科书和类似的出版物中，作者所推荐的实践中都存在这样的糟糕的例子，这些作者本应知道得更好；事实上，在这方面对这些文献资料的回顾是非常的令人沮丧的。在1969年，关系模型就首先令人看到了曙光，然而到现在为止，已经过去40年了，在数据库社区，关系模型看上去还是没有被普遍地理解。部分地因为这种原因，所以本书使用关系模型本身作为组织的原则；本书将会深入解释关系模型的各种特性，并且在每一个例子中都会说明如何最好地使用SQL来实现那些令人怀疑的特性。

首要条件

我假设你是数据库方面的从业人员，因此已经熟悉SQL语言。而且，我假设你在SQL标准或者至少一种SQL产品方面具有一定的工作经验。然而，我并不假设你对关系理论有深入的了解（尽管我确实希望你理解关系模型是一个好东西，并且无论如何都坚持它是值得要的）。因此，为了避免引起误解，我将会详细地描述关系模型的各种特性，并且说明如何使用SQL来遵循这些特性。但是我不去证明这些特性；更合适地，我期望你在理解这些关系模型特性为什么应该这样而不是那样上有一定的经验，例如，理解主键的概念，或者为什么有时需要联接或者为什么需要支持多对多的关系。（如果包含这些特性的证明的话，本书将会是一本非常不一样的书，并且这样将会比现在厚得多，而且这样的书已经出版了。）

虽然我说过期望你熟悉SQL，但是我还是在SQL的某些方面解释得比较详细，特别是那些不会经常用到的方面。（SQL概念中非确定性表达式就是一个相应的例子。请参见第12章。）

深入数据库

本书基于并且想替代更早的那本标题为Database in Depth: Relational Theory for Practitioners (O'Reilly, 2005)的书。我的原来那本书的目的如下（引自那本书的前言部分）：

在各种层次的数据库社区工作很多年后，我终于意识到需要有一本书给那些实践者（不是新手）介绍关系理论的一些基本的准则，在某种程度上，本书并没有受到一些现存产品各种特性、商业惯例或者SQL标准的影响。我写这本书主要是为了满足这种要求。我期望的读者是那些有经验的数据库实践者，他们有足够的诚实来承认他们不理解隐藏在下面的理论。当然这些理论就是关系模型，并且这个模型的基本的思想都非常简单，但也没有得到广泛的使用和认识。事实上是，它们并没有从根本上被理解。例如，下面就是一些关系模型相关的问题（注1），你能回答多少？

- 第一范式准确地来说是什么呢？
- 关系和谓词之间的联系是什么？
- 什么是语义上的优化？
- 什么是映象关系？
- 为什么半差如此重要？

注1： 由于那些这里并不重要的原因，这里我用一些新的问题代替了这个序列中的很多问题。

- 为什么要推迟完整性检查?
- 一个关系变量是什么呢?
- 什么是前缀正规形式?
- 一个关系可以拥有一个属性并且它的值还是关系吗?
- SQL是关系完备的吗?
- 为什么信息理论如此重要?
- XML如何与关系模型集成?

对于这些和更多的相关问题,本书都提供了相关答案。这些问题全部都是用来帮助数据库的实践者去深入理解这些关系理论并且在他们每天的专业实践中更好地使用已经理解的这些知识。

正如上面摘录的最后一句指出的那样,我就是希望通过那本书帮助读者能够自己运用这些知识而不需要像原来那样需要更多的帮助。但是与大多数的观点相反,我最后意识到由于SQL是如此困难的一门语言,以致多数人都不太熟悉在不违反关系准则的情况下如何使用SQL。因此,我决定扩充原来的那本书,使它在这些问题上(如何关系地使用SQL)包含一些直接的、具体的建议。因此,本书的目的和原来的那本一样(例如,我想帮助数据库实践者深入理解关系理论并且想让他们在其专业实践中更好地利用这些他们已经理解的知识),但是我将尽力使得这些材料更容易被接受,并且更容易运用。总之,本书包含了大量的SQL特殊的材料(并且正是由于这个原因,比前一本书的篇幅有了更大的增加)。

正文中更进一步的注释

这里需要说明几个更深层次的预备内容。首先,我自己对关系模式的理解也发展了很多年,并且一直在继续当中。本书代表我在这个主体上最新的思考;因此,如果你发现在本书和其他书中的任何技术上的差异,当然肯定是在一些这样的差异的,本书应该取代我的其他书。尽管我及时增加了这些差异,但是这些差异大部分应该是公正的、小修小改的属性;无论什么时候我觉得有必要用新的术语和概念来阐述更早一些的概念和术语时,我就会这样做。

第二,我谈到这个理论就像做广告一样,但是这个理论是可以实践的、是可以信任的。我明显地提到这一点是因为很多人看上去只相信它的反面:换句话说,如果一个事物是理论的,它就不可能是能实践的。但事实上,这个理论(至少,我谈论的关系理论)是真正可以实践的。这个理论的目的不是为了理论而理论,而是让我们能够创建一个100%的实用的系统。这个理论的每一个细节都有坚实的实践基础。就像早先的那本书的评论

者Stéphane Faroult写的那样：“当你拥有一点实践时，你就会意识到没有任何东西能够阻碍你了解这个理论”。更多的是，这个理论不但是可以实践的，它也是基本的、直接的、简单和有用的，并且它也是很好玩的（就像我在本书中期望证明的这样）。

当然，我们真的不需要去查看比这个关系模式本身更深奥的任何理论，从而发现这个前述的理论最为显著的可能的证明。事实上，真的不应该必须去辩护这个理论是可以实践的这个观点，我们的观点是：在这个伟大的理论思想上已经花费了数万亿的资金。但是我料想一些愤世嫉俗的言论可能是：“是的，但是这个理论能给我们带来什么？”换句话说，我们中的那些认为这个理论是重要的必须不断地给那些评论者证明我们自己——这也就是我认为我们需要这样的一本书的另一个原因。

第三，就像前面说过的那样，本书确实探究了关于SQL特性或者关系模式或者这两者的详细细节。（对于那些与关系模式没有关系的主体则没有阐述。例如，事务就没有探究。）自始至终，对于那些特殊应用于SQL的讨论，特殊应用于关系模式和应用在它们两者上的讨论我都尽量使它清晰。然而，我应该强调的是那些特殊应于SQL的讨论并不是详尽的。SQL是如此复杂的一门语言，并且给做同一件事情提供了很多种不同的实现方案，并且遭受到很多异议和特殊的案例，并且只要可能，这些异议和特殊的案例就会非常的详尽，这些是我所怀疑的，而这将降低生产率；并且肯定会使得本书篇幅更加的长。因此，我将把重点放在我认为重要的事情上面，在这些我挑选出的问题上我将尽可能地简短。并且我要求你做我告诉你的事情，不要做我没有告诉你的事情，这样，你将会关系地使用SQL。但是这些要求是否被证明，或者这些要求到什么程度，都必须由你自己判断。

到目前为止，我还是不得不增加以下说明，不幸的是，在很多情况下SQL不能关系地使用。例如，一些SQL的完整性检查就不得不简单地推迟（常常是推迟到提交时间），即使关系模式由于逻辑上的瑕疵反对这些检查。在这些情况下，本书对该如何做提供了一些建议，但是恐怕这也只能归结为尽你所能地做得最好。至少，我期望你理解这些涉及从关系模式分离而带来的风险。我应该说这些建议中的一些也不是任何时候都是明确的、相关的，但是总地来说是一些好的实践，尽管有时有一些相关的暗示（或许我更应该说这是这个暗示不是那么明显）。这里避免强迫就是一个好的例子。

第四，请注意在本书中我使用SQL术语来意味着这个语言的排他性的标准版本，而不是用别的特殊性的方言（除非相反的明显的说明）。特别地，我遵循这个标准假设它的发音是“ess cue ell，”而不是sequel（尽管在这个领域后者是流行的），因此说到一个SQL表时，这个一个不是a而是an。

第五，几乎你应该依次读这本书，那些正文在各个地方都可以找到的章节除外（大部分章节都依赖于它前面章节中的内容，因此你不应该跳动太多）。而且，每一章都包含一

系列的练习。当然你不必都做这些练习，但最少我想你应该都通读一下。它的答案都可以在附录C中找到，这些答案给出了这些主题更多的信息。

最后，我想提一下我还有一些基于本书中的材料的更详细讨论，请参考http://www.clik.to/chris_date或者<http://www.thethirdmanifesto.com>。

本书使用到的一些惯例

本书使用的一些印刷惯例如下：

斜体 (*Italic*)

作为重点使用。标明新的术语。也表示当一个变量如 x ，作为在本书中主要正文中讨论的地方。

等宽体 (*Constant width*)

用作代码例子。

等宽斜体 (*Consrant width italic*)

在一个代码例子中标识变量的出现或者是用户提供的元素的出现。

代码例子的使用

本书是用来帮助做好自己的工作。总地来说，你可以在文档和程序中使用这些代码。不必为了许可而联系我们，除非你在这些代码中重新产生了重大部分。例如，当你写程序时使用了本书中的几个大块的代码，你不用获得许可。但出售或者分发带有从O'Reilly书中例子的CD-ROW则需要获得许可。通过引用本书或者引用这些例子代码来回答问题不需要获得许可。但合并本书中大量的代码到你的产品的文档中则要获得许可。

当你引用本书时，我们感谢对我们的提及但不强求。通常一个归因包括标题、作者、出版商和ISBN号。例如：SQL and Relational Theory by C. J. Date. Copyright 2009 C. J. Date, 978-0-596-52306-0。

如果你觉得使用这些例子代码时超过了以上的这些权限，请跟我们联系permissions@oreilly.com。

意见和问题

我已经尽最大力来使本书免于出错，但你也可能会发现错误。如果是这样，请通过以下地址写信通知出版商：

美国:

O'Reilly Media, Inc.

1005 Gravenstein Highway North

Sebastopol, CA 95472

中国:

北京市西城区西直门南大街2号成铭大厦C座807室 (100035)

奥莱利技术咨询(北京)有限公司

你也可以给我们发电子邮件。索取目录或者需要邮寄书籍,发电子邮件给以下地址:

info@oreilly.com

给本书提任何建议或者提问任何技术问题,发电子邮件给以下地址:

bookquestions@oreilly.com

我们给本书提供了一个网站,在那里你可以找到一些例子和勘误表(以前报告的错误和更正,任何人都可以在那里找到)。你可以通过以下链接访问到这些资料:

<http://www.oreilly.com/catalog/9780596523060>

更多的有关本书和我的其他书籍的信息,请参考O'Reilly网站:

<http://www.oreilly.com>

<http://www.oreilly.com.cn>

致谢

我曾经思考了一段时间关于修改早先那本书来包含更多特别的SQL方面的论述,但是在2007年后期,在给一些数据库实践者的课程上的一些感受使我开始认真考虑这个问题。这个课程是由Toon Koppelaars讲授的,并且是基于他和Lex de Haan(参见本书附录D)合作写的一本书,这个课进展得非常好。但是在这次课程中令我印象非常深的就是在课程中很清楚地看到了那些参加者在把关系的和合理的规则运用在SQL使用方面的能力非常欠缺。现在,我假设这些参与者有这些主题方面的知识,毕竟他们是数据库方面的实践者,但是我似乎还是觉得他们真的在把这些规则运用到他们日常工作实践中还是需要一些指导。因此我把这本书组织起来。首先,我要感谢Toon和Lex为我开始写这本书提供了必须的推动力;感谢我的评论家Herb Edelstein, Sheeri Ktitzer, Andy Oram, Peter Robson和Baron Schwartz,在我前期的草稿阶段提出了宝贵的意见;感谢Hugh Darwen和Jim Melton给予我技术上的帮助;还要感谢我的妻子Lindy,感谢她这些年一直以来对