

/THEORY/IN/PRACTICE

The Art of

SQL

SQL语言艺术

Stéphane Faroult &
Peter Robson 著
温昱 靳向阳 译

O'REILLY®



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

O'REILLY®

SQL 语言艺术

The Art of SQL

[美] Stéphane Faroult, Peter Robson 著

温昱 靳向阳 译

電子工業出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

本书分为 12 章,每一章包含许多原则或准则,并通过举例的方式对原则进行解释说明。这些例子大多来自于实际案例,对九种 SQL 经典查询场景以及其性能影响讨论,非常便于实践,为你的实际工作提出了具体建议。本书适合 SQL 数据库开发者、软件架构师,也适合 DBA,尤其是数据库应用维护人员阅读。

0596008945 The Art of SQL© 2006 by O'Reilly Media, Inc.

Simplified Chinese edition, jointly published by O'Reilly Media, Inc. and Publishing House of Electronics Industry, 2008. Authorized translation of the English edition, 2006 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: 01-2006-5602

图书在版编目(CIP)数据

SQL 语言艺术 / (美)法禾(Faroult,S.), (美)罗伯森(Robson,P.)著;温昱,靳向阳译.—北京:电子工业出版社,2008.3

书名原文:The Art of SQL

ISBN 978-7-121-05834-9

I. S… II. ①法…②罗…③温…④靳… III. 关系数据库—数据库管理系统 IV. TP311.138

中国版本图书馆 CIP 数据核字(2008)第 009405 号

责任编辑:梁 晶

印 刷:北京市天竺颖华印刷厂

装 订:三河市金马印装有限公司

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:23.5 字数:450 千字

印 次:2008 年 3 月第 1 次印刷

定 价:58.00 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 zlt@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线:(010) 88258888。

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, 同时也是联机出版的先锋。

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。

计利以听，乃为之势，以佐其外。

While heeding the profit of my counsel, avail yourself also of
any helpful circumstances over and beyond the ordinary rules.

—Sun Tzu, *The Art of War*

The French humorist Alphonse Allais(1854-1905),once dedicated one of his short storied as follows:

To the only woman I love and who knows it well.

...with the following footnote:

This is a very convenient dedication that I cannot recommend too warmly to my fellow writers.It costs nothing,and can,all at once,please five or six persons.

I can take a piece of wise advice when I meet one.

Stéphane Faroult

前言

过去,“信息技术(IT)”的名字还不如今天这般耀眼,被称为“电子数据处理”。其实,尽管当今新潮技术层出不穷,数据处理依然处于我们系统的核心地位,而且需管理的数据量的增长速度似乎比处理器的增长速度还快。今天,最重要的集团数据都被保存在数据库中,通过 SQL 语言来访问。SQL 语言虽有缺点,但非常流行,它从 1980 年代早期开始被广泛接受,随后就所向无敌了。

如今,年轻开发者在接受面试时,没有谁不宣称自己能熟练应用 SQL 的。SQL 作为数据库访问语言,已成为任何基础 IT 课程的必备部分。开发者宣传自己熟练掌握 SQL,其实前提是“熟练掌握”的定义是“能够获得功能上正确的结果”。然而,全世界的企业如今都面临数据量的爆炸式增长,所以仅做到“功能正确”是不够的,还必须足够快,所以数据库性能成了许多公司头疼的问题。有趣的是,尽管每个人都认可性能问题源自代码,但普遍接受的事实则是开发者的首要关注点应该是功能正确。人们认为:为了便于维护,代码中的数据库访问部分应该尽量简单;“拙劣的 SQL”应该交给资深的 DBA 去摆弄,他们还会调整几个“有魔力”的数据库参数,于是速度就快了——如果数据库还不够快,似乎就该升级硬件了。

往往就是这样,那些所谓的“常识”和“可靠方法”最终却是极端有害的。先写低效的代码、后由专家调优,这种做法实际上是自找麻烦。本书认为,首先要关注性能的就是开发者,而且 SQL 问题绝不仅仅只包含正确编写几个查询这么简单。开发者角度看到的性能问题和 DBA 从调优角度看到的大相径庭。对 DBA 而言,他尽量从现有的硬件(如处理器和存储子系统)和特定版本的 DBMS 获得最高性能,他可能有些 SQL 技能并能调优一个性能极差的 SQL 语句。但对开发者而言,他编写的代码可能要运行 5 到 10 年,这些代码将经历一代代的硬件,以及 DBMS 各种重要版本升级(例如支持互联网访问、支持网格,不一而足)。所以,代码必须从一开始就快速、健全。很多开发者仅仅是“知道”SQL 而已,他们没有深刻理解 SQL 及关系理论,实在令人遗憾。

为何写作本书

SQL 书主要分为三种类型：讲授具体 SQL 方言的逻辑和语法的书、讲授高级技术及解决问题方法的书、专家与资深 DBA 所需的性能和调优的书。一方面，书籍要讲述如何写 SQL 代码；另一方面，要讲如何诊断和修改拙劣的 SQL 代码。在本书中，我不再为新手从头讲解如何写出优秀的 SQL 代码，而是以超越单个 SQL 语句的方式看待 SQL 代码，无疑这更加重要。

教授语言使用就够难了，那么本书是怎样讲述如何高效使用 SQL 语言的呢？SQL 的简单性具有欺骗性，它能支持的情况组合的数目几乎是无限的。最初，我觉得 SQL 和国际象棋很相似，后来，我悟到发明国际象棋是为了教授战争之道。于是，每当出现 SQL 性能难题的时候，我都自然而然地将之视为要和一行行数据组成的军队作战。最终，我找到了向开发者传授如何有效使用数据库的方法，这就像教军官如何指挥战争。知识、技能、天赋缺一不可。天赋不能传授，只能培养。从写就了《孙子兵法》的孙子到如今的将军，绝大多数战略家都相信这一点，于是他们尽量以简单的格言或规则的方式表达沙场经验，并希望这样能指导真实的战争。我将这种方法用于战争之外的许多领域，本书借鉴了孙子兵法的方法和书的题目。许多知名 IT 专家冠以科学家称号，而我认为“艺术”比“科学”更能反映 IT 活动所需的才能、经验和创造力（注 1）。很可能是由于我偏爱“艺术”的原因，“科学”派并不赞成我的观点，他们声称每个 SQL 问题都可通过严格分析和参考丰富的经验数据来解决。然而，我不认为这两种观点有什么不一致。明确的科学方法有助于摆脱单个具体问题的限制，毕竟，SQL 开发必须考虑数据的变化，其中有很大的不确定性。某些表的数据量出乎意料地增长将会如何？同时，用户数量也倍增了，又将会如何？希望数据在线保存好几年将会如何？如此一来，运行在硬件之上的这些程序的行为是否会完全不同？架构级的选择是在赌未来，当然需要明确可靠的理论知识——但这是进一步运用艺术的先决条件。第一次世界大战联军总司令 Ferdinand Foch，在 1900 年 French Ecole Supérieure de Guerre 的一次演讲中说：

战争的艺术和其他艺术一样，有它的历史和原则——否则，就不能成其为艺术。

本书不是 cookbook，不会列出一串问题然后给出“处方”。本书的目标重在帮助开发者（和他们的经理）提出犀利的问题。阅读和理解了本书之后，你并不是永不再写出丑陋缓慢的查询了——有时这是必须的——但希望你是故意而为之、且有充足的理由。

注 1：我最喜欢的计算机书，恰好是 D.E. Knuth 的经典著作 *Art of Computer Programming* (Addison Wesley 出版社)。

目标读者

本书的目标读者是：

- 有丰富经验的 SQL 数据库开发者
- 他们的经理
- 数据库占重要地位的系统的软件架构师

我希望一些 DBA、尤其是数据库应用维护人员也能喜欢本书。不过，他们不是本书的主要目标读者。

本书假定

本书假定你已精通 SQL 语言。这里所说的“精通”不是指在你大学里学了 SQL 101 并拿来 A+ 的成绩，当然也并非指你是国际公认的 SQL 专家，而是指你必须具有使用 SQL 开发数据库应用的经验、必须考虑索引、必须不把 5000 行的表当大表。本书的目标不是讲解连接、外连接、索引的基础知识，阅读本书过程中，如果你觉得某个 SQL 结构还显神秘，并影响了整段代码的理解，可先阅读几本其他 SQL 书。另外，我假定读者至少熟悉一种编程语言，并了解计算机程序设计的基本原则。性能已很差、用户已抱怨、你已在解决性能问题的前线，这就是本书的假定。

本书内容

我发现 SQL 和战争如此相像，以至于我几乎沿用了《孙子兵法》的大纲，并保持了大部分题目名称（注 2）。本书分为 12 章，每一章包含许多原则或准则，并通过举例的方式对原则进行解释说明，这些例子大多来自于实际案例。

第 1 章，制定计划：为性能而设计

讨论如何设计高性能数据库

注 2：另一些题目名称我借鉴了 Clausewitz 的 *On War*。

第2章，发动战争：高效访问数据库

解释如何进行程序设计才能高效访问数据库

第3章，战术部署：建立索引

揭示为何建立索引，如何建立索引

第4章，机动灵活：思考 SQL 语句

解释如何设计 SQL 语句

第5章，了如指掌：理解物理实现

揭示物理实现如何影响性能

第6章，锦囊妙计：认识经典 SQL 模式

包括经典的 SQL 模式、以及如何处理

第7章，变换战术：处理层次结构

说明如何处理层次数据

第8章，孰优孰劣：认识困难，处理困难

指出如何认识和处理比较棘手的情况

第9章，多条战线：处理并发

讲解如何处理并发

第10章，集中兵力：应付大数据量

讲解如何应付大数据量

第11章，精于计谋：挽救响应时间

分享一些技巧，以挽救设计糟糕的数据库的性能

第12章，明察秋毫：监控性能

收尾，解释如何定义和监控性能

本书约定

本书使用了如下印刷惯例：

等宽 (Courier)

表示 SQL 及编程语言的关键字，或表示 table、索引或字段的名称，抑或表示函数、代码及命令输出。

等宽黑体 (Courier)

表示必须由用户逐字键入的命令等文本。此风格仅用于同时包含输入、输出的代码示例。

等宽斜体 (Courier)

表示这些文本，应该被用户提供的值代替。



总结：箴言，概括重要的 SQL 原则。

注意

提示、建议、一般性注解。为相关主题提供有用的附加信息。

代码示例

本书是为了帮助你完成工作的。总的来说，你可以将本书的代码用于你的程序和文档，但是，若要大规模复制代码，则必须联系 O'Reilly 申请授权。例如：编程当中用了本书的几段代码，无需授权；但出售或分发 O'Reilly 书籍中案例的 CD-ROM 光盘，需要授权。再如：回答问题时，引用了本书或其中的代码示例，无需授权；但在你的产品文档中大量使用本书代码，需要授权。

O'Reilly 公司感谢但不强制归属声明。归属声明通常包括书名、作者、出版商、ISBN。例如 “The Art of SQL by Stéphane Faroult with Peter Robson. Copyright © 2006 O'Reilly Media, 0-596-00894-5”。

如果你对代码示例的使用超出了上述范围, 请通过 permissions@oreilly.com 联系出版商。

评论与提问

我们已尽力核验本书所提供的信息, 尽管如此, 仍不能保证本书完全没有瑕疵, 而网络世界的变化之快, 也使得本书永不过时的保证成为不可能。如果读者发现本书内容上的错误, 不管是赘字、错字、语意不清, 甚至是技术错误, 我们都竭诚虚心接受读者指教。如果您有任何问题, 请按照以下的联系方式与我们联系。

O'Reilly Media, Inc.

1005 Gravenstein Highway North

Sebastopol, CA 95472

(800) 998-9938 (in the U.S. or Canada)

(707) 829-0515 (international or local)

(707) 829-0104 (fax)

奥莱理软件(北京)有限公司

北京市 海淀区 知春路 49 号 希格玛公寓 B 座 809 室

邮政编码: 100080

网页: <http://www.oreilly.com.cn>

E-mail: info@mail.oreilly.com.cn

北京博文视点资讯有限公司(武汉分部)

湖北省 武汉市 洪山区 吴家湾 邮科院路特 1 号 湖北信息产业科技大厦 1402 室

邮政编码: 430074

电话: (027)87690813 传真: (027) 87690595

读者服务网页: <http://bv.csdn.net>

E-mail:

reader@broadview.com.cn (读者信箱)

bvtougao@gmail.com (投稿信箱)

出版商为本书建立了网页, 来提供勘误表、案例、其他附加信息, 网址如下:

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

<http://www.oreilly.com.cn/book.php?bn=978-7-121-05834-9> (中文版)

评论本书或讨论本书相关的技术问题, 可发邮件至:

bookquestions@oreilly.com

欲了解更多 O'Reilly 书籍、会议、资源中心、网络等信息, 请访问 O'Reilly 网站:

<http://www.oreilly.com>

本书作者公司的网站为:

<http://www.roughsea.com>

致谢

本书原版用英语写作, 英语既不是我的家乡话, 又不是我所在国家的语言, 所以写这样一本书要求极度乐观 (回想起来几近疯狂)。幸运的是, Peter Robson 不仅为本书贡献了他在 SQL 和数据库设计方面的知识, 也贡献了持续的热情来修改我冗长的句子、调整副词位置、斟酌替换词汇。Peter Robson 和我在好几个大会上都碰过面, 我们都是演讲者。

Jonathan Gennick 担任本书编辑, 这有点让人受宠若惊, Jonathan Gennick 是 O'Reilly 出版的 *SQL Pocket Guide* 等畅销名著的作者。Jonathan 是个非常尊重作者的编辑。由于他的专业、他对细节的关注、他的犀利视角, 使本书的质量大大提升。同时, Jonathan 也使本书的语言更具“中大西洋”风味 (Peter 和我发现, 虽然我们保证按美国英语拼写, 但还远远不够)。

我还要感谢很多人, 他们来自三个不同的大陆, 阅读了本书全部或部分草稿并坦诚地提出意见。他们是: Philippe Bertolino、Rachel Carmichael、Sunil CS、Larry Elkins、Tim Gorman、Jean-Paul Martin、Sanjay Mishra、Anthony Molinaro、Tiong Soo Hua。我特别感激 Larry, 因为本书的思想最初来自于我们的 E-mail 讨论。

我也要感谢 O'Reilly 的许多人, 他们使本书得以出版。他们是: Marcia Friedman、Rob Romano、Jamie Peppard、Mike Kohnke、Ron Bilodeau、Jessamyn Read、Andrew Savikas。感谢 Nancy Reinhardt 卓越的手稿编辑工作。

特别感谢 Yann-Arzel Durelle-Marc 慷慨提供第 12 章用到的图片。感谢 Paul McWhorter

授权我们将他的战争图用于第 6 章。

最后，感谢 Roger Manser 和 Steel Business Briefing 的职员，他们为 Peter 和我提供了位于伦敦的办公室（还有大量咖啡）。感谢 Qian Lena (Ashley) 提供了本书开始引用的《孙子兵法》的中文原文。

目录

Contents



前言	1
1 制定计划：为性能而设计	1
数据的关系视图	2
规范化的重要性	4
有值、无值、空值	11
限用 Boolean 型字段	14
理解子类型 (Subtype)	15
约束应明确声明	17
过于灵活的危险性	18
历史数据的难题	19
设计与性能	21
处理流程	22
数据集中化 (Centralizing)	23
系统复杂性	24
小结	25
2 发动战争：高效访问数据库	27
查询的识别	28
保持数据库连接稳定	29
战略优先于战术	31
先定义问题，再解决问题	32
保持数据库 Schema 稳定	33
直接操作实际数据	34
用 SQL 处理集合	34
动作丰富的 SQL 语句	35
充分利用每次数据库访问	36
接近 DBMS 核心	37
只做必须做的	41



SQL 语句反映业务逻辑	42
把逻辑放到查询中	42
一次完成多个更新	43
慎用自定义函数	44
简洁的 SQL	46
SQL 的进攻式编程	48
精明地使用异常 (Exceptions)	50
3 战术部署：建立索引	55
找到“切入点”	56
索引与目录	59
让索引发挥作用	60
函数和类型转换对索引的影响	62
索引与外键	67
同一字段，多个索引	69
系统生成键	70
索引访问的不同特点	72
4 机动灵活：思考 SQL 语句	75
SQL 的本质	76
掌握 SQL 艺术的五大要素	84
过滤	89
5 了如指掌：理解物理实现	105
物理结构的类型	106
冲突的目标	108
把索引当成数据仓库	109
记录强制排序	113
数据自动分组 (Grouping)	115
分区是双刃剑	119
分区与数据分布	120
数据分区的最佳方法	121
预连接表	123
神圣的简单性	124



6	锦囊妙计：认识经典 SQL 模式	127
	小结果集，直接条件	129
	小结果集，间接条件	137
	多个宽泛条件的交集	138
	多个间接宽泛条件的交集	140
	大结果集	146
	基于一个表的自连接	147
	通过聚合获得结果集	150
	基于日期的简单搜索或范围搜索	156
	结果集和别的数据存在与否有关	161



7	变换战术：处理层次结构	167
	小结果集，直接条件	129
	小结果集，间接条件	137
	多个宽泛条件的交集	138
	多个间接宽泛条件的交集	140
	大结果集	146
	基于一个表的自连接	147
	通过聚合获得结果集	150
	基于日期的简单搜索或范围搜索	156
	结果集和别的数据存在与否有关	161



8	孰优孰劣：认识困难，处理困难	199
	看似高效的查询条件	200
	抽象层	202
	分布式系统	205
	动态定义的搜索条件	208



9	多条战线：处理并发	225
	数据库引擎作为服务提供者	226
	并发修改数据	231



10	集中兵力：应付大数据量	247
	增长的数据量	248
	数据仓库	264