

Learning SQL

SQL 学习指南

(第2版)



O'REILLY®

[美] Alan Beaulieu 著
张伟超 林青松 译

 人民邮电出版社
POSTS & TELECOM PRESS

O'REILLY®

SQL 学习指南（第 2 版）

[美] Alan Beaulieu 著

张伟超 林青松 译

人 民 邮 电 出 版 社

北 京

图书在版编目 (C I P) 数据

SQL学习指南：第2版 / (美) 比利 (Beaulieu, A.)
著；张伟超，林青松译. — 北京：人民邮电出版社，
2010.6

ISBN 978-7-115-22875-8

I. ①S… II. ①比… ②张… ③林… III. ①关系数
据库—数据库管理系统, SQL Server IV. ①TP311.138

中国版本图书馆CIP数据核字(2010)第073249号

版权声明

Copyright©2009 by O'Reilly Media, Inc.

Simplified Chinese Edition, jointly published by O'Reilly Media, Inc. and Posts & Telecom Press, 2010. Authorized translation of the 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.

本书中文简体版由 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。

SQL 学习指南 (第 2 版)

- ◆ 著 [美] Alan Beaulieu
- 译 张伟超 林青松
- 责任编辑 刘映欣
- ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
邮编 100061 电子函件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京鑫正大印刷有限公司印刷
- ◆ 开本：787×1000 1/16
印张：19.25
字数：394 千字 2010 年 6 月第 1 版
印数：1-4 000 册 2010 年 6 月北京第 1 次印刷

著作权合同登记号 图字：01-2009-6946 号

ISBN 978-7-115-22875-8

定价：45.00 元

读者服务热线：(010)67132705 印装质量热线：(010)67129223

反盗版热线：(010)67171154

内 容 提 要

本书全面系统地介绍了 SQL 语言各方面的基础知识以及一些高级特性，包括 SQL 数据语言、SQL 方案语言、数据集操作、子查询以及内建函数与条件逻辑等内容。书中每个章节讲述一个相对独立的主题，并提供了相关示例和练习。本书内容以 SQL 92 标准为蓝本，涵盖了市场上常用数据库的最新版本 (MySQL 6.0、Oracle 11g 及 Microsoft SQL Server 2008)。

本书适合数据库应用开发者、数据库管理员和高级用户阅读。针对开发基于数据库的应用程序，以及日常的数据库系统管理，本书都展现了大量经过实践检验的方法和技巧。读者可以通过对本书循序渐进地学习快速掌握 SQL 语言，也可以在实际工作中遇到问题时直接翻阅本书中的相关章节以获取解决方案。

O'Reilly Media, Inc.介绍

为了满足你对网络 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. 授权人民邮电出版社，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。

前言

编程语言在不断地出现和消亡，现在使用的语言只有很少一部分的历史能追溯到 20 年前。其中有大量应用在大型机环境的 Cobol 和流行于操作系统、服务器开发以及嵌入式系统的 C 语言。而在数据库领域，SQL 的根源可以追溯到 19 世纪 70 年代。

SQL 是一种从关系型数据库生成、操作和检索数据的语言。关系型数据库流行的原因之一是正确设计的关系型数据库可以处理海量数据。但处理大量数据集时，SQL 就像一个高功率可变焦距的时髦数码相机，让你能够看到大型数据集，或者放大单独的行（或者两者之间的任何地方）。其他的数据库管理系统在沉重的负荷下往往会由于它们的焦距太窄（缩放镜头已经处于最大位置了）而崩溃，这就是要废黜关系型数据库和 SQL 的尝试已经基本上失败的原因。因此，即使 SQL 是一门古老的语言，它也将继续活跃很长一段时间，并且在存储应用方面有光明的前途。

为什么要学习 SQL

如果打算使用关系型数据库，无论是写应用程序、执行管理任务还是生成报表，那么都需要知道如何与数据库中的数据交互。即使使用工具为自己生成 SQL，比如报表工具，有时也需要绕过自动生成功能而编写自己的 SQL 语句。

学习 SQL 语言有一个额外的好处，即强迫你勇敢面对并学会理解用于储存自己组织相关信息的数据结构。当开始适应数据库的表时，你可能会发现自己也会产生对数据库进行修改或增加等的建议。

为什么使用本书学习 SQL

SQL 语言可分为几类：用于创建数据库对象（表、索引、约束等）的语句统称为 SQL 模式语句，而用于创建、操纵和检索保存在数据库中的数据的语句称为 SQL 数据语句。作为管理员，你将同时使用 SQL 模式和 SQL 数据语句，而程序员或者报表作者可能只需要使用（或者只允许使用）SQL 数据语句。虽然本书介绍了许多 SQL 模式语句，但是主要焦点还是编程功能。

由于只有少数命令，因此 SQL 数据语句看似很简单。依我看来，现在许多 SQL 图书都通过仅仅涉猎这个语言可能的表层知识帮助你培养这种观念。然而，如果打算使用 SQL，那么你就有必要充分理解它的语言能力以及如何组合不同的功能以产生强大的结果。我感觉本书是唯一一本详细介绍 SQL 语言而不会同时被作为门挡的书（正如我知

道的，1250 页的“完全手册”往往被丢在人们的卧室书架上，布满了灰尘)。

虽然本书的示例都可以运行在 MySQL、Oracle 数据库以及 SQL Server 上，但是我必须选择其中之一来作为示例数据库服务器并规范化示例查询返回的结果集。我在这 3 个产品中选择了 MySQL，是因为它可以免费获得、安装简单以及易于管理。对于那些使用其他服务器的人，我建议下载和安装 MySQL 并加载示例数据库，这样就可以运行示例数据库并试验数据了。

本书的结构

本书分为 15 章和 3 个附录。

第 1 章“背景知识”，探讨计算机数据库的历史，其中包括关系模型以及 SQL 语言的出现。

第 2 章“创建和操作数据库”，说明如何创建本书示例使用的 MySQL 数据库和表，以及用数据填充表。

第 3 章“查询入门”，介绍选择语句，然后进一步阐述了大多数常用子句 (select、from、where)。

第 4 章“过滤”，说明不同类型的条件，它们可以用于 select、update 或 delete 语句的 where 子句中。

第 5 章“多表查询”，展示如何通过表的连接使用多表进行查询。

第 6 章“使用集合”，介绍所有关于数据集的知识以及它们如何在查询内交互。

第 7 章“数据生成、转换和操作”，说明用于操作或转换数据的几个内置函数。

第 8 章“分组与聚集”，展示如何聚合数据。

第 9 章“子查询”，介绍子查询 (个人最喜欢的)，并说明如何以及在何处使用它们。

第 10 章“再谈连接”，更加深入地讨论不同类型的表连接。

第 11 章“条件逻辑”，探讨如何在 select、insert、update 和 delete 语句里使用条件逻辑 (如 if-then-else)。

第 12 章“事务”，介绍事务及如何使用它们。

第 13 章“索引和约束”，探讨索引和约束。

第 14 章“视图”，说明如何构建接口以屏蔽数据复杂性。

第 15 章“元数据”，说明数据字典的使用。

附录 A “示例数据库的 ER 图”，展示本书所有示例的数据库模式。

附录 B “MySQL 对 SQL 语言的扩展”，说明在 MySQL 的 SQL 实现中一些有趣的非

ANSI 功能。

附录 C “练习答案”，介绍各章习题的答案。

阅读须知

本书使用了下面的印刷约定：



提示

指出提示、建议或者一般注意性的问题。例如，我使用注意向你表明 Oracle 9i 的新功能。



警告

指示一个告诫或提醒。例如，我告诉你如果不小心使用，那么有些 SQL 语句可能会产生意想不到的后果。

联系我们

请将对本书的评论和问题按以下地址与出版社联系。

美国：

O'Reilly Media, Inc.

1005 Gravenstein Highway North

Sebastopol, CA 95472

中国：

100035 北京市西城区西直门成铭大厦 C 座 807 室

奥莱利技术咨询（北京）有限公司

O'Reilly 为本书维护了一个网页，列出了勘误表、范例以及任何其他信息。可以通过以下地址访问：

<http://www.oreilly.com/catalog/9780596520830>

要询问技术问题或提出建议，请发邮件至：

bookquestions@oreilly.com

info@mail.oreilly.com.cn

更多关于 O'Reilly 的图书、会议信息、资源中心以及 O'Reilly 网络，请访问以下网页：

<http://www.oreilly.com>

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

使用示例代码

本书是为了帮助你完成工作。一般情况下，你可以在程序和文档中使用本书的代码。如果你使用本书的重要代码，不必联系我们获取许可。例如，使用本书中几大块代码写自己的程序不需要获得许可，但是如果要将 O'Reilly 书籍中的用例制作成光盘出售或发布，则必须获得许可。引用本书内容或范例解决其他问题不需要获得许可，但是如果想在你的产品文档中包含本书中一些重要的示例代码，那么也需要得到许可。

如果引用了本书内容，那么我们很感激你标明出处，但并不做要求。出处通常包括标题、作者、出版商以及 ISBN。例如，“Learning SQL, Second Edition, by Alan Beaulieu. Copyright 2009 O'Reilly Media, Inc., 978-0-596-52083-0.”。

如果你并不清楚使用本书示例代码是否侵权，请随时与我们联系：

permissions@oreilly.com

致谢

首先我想感谢编辑 Mary Treseler，因为她的帮助使第 2 版成为现实，其次，再次感谢 Kevin Kline、Roy Owens、Richard Sonen 和 Matthew Russell，是他们在圣诞节和新年时还在审查本书。我还想感谢第 1 版的一些读者，他们提出了很多问题、建议和勘误。最后，感谢我的妻子 Nancy、女儿 Michelle 和 Nicole，是她们给了我很多鼓励和启发。

目录

第 1 章 背景知识	1
1.1 数据库简介	1
1.1.1 非关系数据库	2
1.1.2 关系模型	3
1.1.3 一些术语	5
1.2 什么是 SQL	6
1.2.1 SQL 语句的分类	6
1.2.2 SQL: 非过程化语句	7
1.2.3 SQL 示例	9
1.3 什么是 MySQL	11
1.4 内容前瞻	11
第 2 章 创建和使用数据库	13
2.1 创建 MySQL 数据库	13
2.2 使用 mysql 命令行工具	15
2.3 MySQL 数据类型	16
2.3.1 字符型数据	16
2.3.2 数值型数据	18
2.3.3 时间数据	20
2.4 表的创建	22
2.4.1 第 1 步: 设计	22
2.4.2 第 2 步: 精化	22
2.4.3 第 3 步: 构建 SQL 方案语句	24
2.5 操作与修改表	27
2.5.1 插入数据	27
2.5.2 更新数据	31

2.5.3 删除数据	32
2.6 导致错误的语句	32
2.6.1 主键不唯一	32
2.6.2 不存在的外键	32
2.6.3 列值不合法	33
2.6.4 无效的日期转换	33
2.7 Bank 方案	34
第 3 章 查询入门	37
3.1 查询机制	37
3.2 查询语句	39
3.3 select 子句	39
3.3.1 列的别名	42
3.3.2 去除重复的行	43
3.4 from 子句	44
3.4.1 表的概念	44
3.4.2 表连接	46
3.4.3 定义表别名	47
3.5 where 子句	48
3.6 group by 和 having 子句	50
3.7 order by 子句	51
3.7.1 升序或降序排序	53
3.7.2 根据表达式排序	54
3.7.3 根据数字占位符排序	55
3.8 小测验	55
第 4 章 过滤	57
4.1 条件评估	57
4.1.1 使用圆括号	58
4.1.2 使用 not 操作符	59
4.2 构建条件	60
4.3 条件类型	60
4.3.1 相等条件	60
4.3.2 范围条件	62

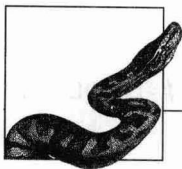
4.3.3 成员条件	65
4.3.4 匹配条件	67
4.4 null: 4 个字母的关键字	70
4.5 小测验	73
第 5 章 多表查询	75
5.1 什么是连接	75
5.1.1 笛卡儿积	76
5.1.2 内连接	77
5.1.3 ANSI 连接语法	80
5.2 连接 3 个或更多的表	82
5.2.1 将子查询结果作为查询表	84
5.2.2 连续两次使用同一个表	86
5.3 自连接	87
5.4 相等连接和不等连接	88
5.5 连接条件和过滤条件	90
5.6 小测验	91
第 6 章 使用集合	93
6.1 集合理论基础	93
6.2 集合理论实践	95
6.3 集合操作符	97
6.3.1 union 操作符	97
6.3.2 intersect 操作符	99
6.3.3 except 操作符	100
6.4 集合操作规则	102
6.4.1 对复合查询结果排序	102
6.4.2 集合操作符优先级	103
6.5 小测验	105
第 7 章 数据生成、转换和操作	107
7.1 使用字符串数据	107
7.1.1 生成字符串	108
7.1.2 操作字符串	112
7.2 使用数值数据	118

7.2.1	执行算术函数	119
7.2.2	控制数字精度	120
7.2.3	处理有符号数	122
7.3	使用时间数据	123
7.3.1	处理时区	123
7.3.2	生成时间数据	125
7.3.3	操作时间数据	129
7.4	转换函数	133
7.5	小测验	134
第 8 章	分组与聚集	135
8.1	分组概念	135
8.2	聚集函数	137
8.2.1	隐式或显式分组	138
8.2.2	对独立值计数	139
8.2.3	使用表达式	141
8.2.4	如何处理 null 值	141
8.3	产生分组	142
8.3.1	对单列的分组	143
8.3.2	对多列的分组	143
8.3.3	利用表达式分组	144
8.3.4	产生合计数	144
8.4	分组过滤条件	146
8.5	小测验	148
第 9 章	子查询	149
9.1	什么是子查询	149
9.2	子查询类型	150
9.3	非关联子查询	150
9.3.1	多行单列子查询	152
9.3.2	多列子查询	157
9.4	关联子查询	159
9.4.1	exists 运算符	161
9.4.2	关联子查询操作数据	162

9.5 何时使用子查询	163
9.5.1 子查询作为数据源	163
9.5.2 过滤条件中的子查询	168
9.5.3 子查询作为表达式生成器	169
9.6 子查询总结	172
9.7 小测验	173
第 10 章 再谈连接	174
10.1 外连接	174
10.1.1 左外连接与右外连接	178
10.1.2 三路外连接	179
10.1.3 自外连接	181
10.2 交叉连接	184
10.3 自然连接	190
10.4 小测验	192
第 11 章 条件逻辑	194
11.1 什么是条件逻辑	194
11.2 case 表达式	196
11.2.1 查找型 case 表达式	196
11.2.2 简单 case 表达式	198
11.3 case 表达式范例	199
11.3.1 结果集变换	199
11.3.2 选择性聚合	200
11.3.3 存在性检查	202
11.3.4 除零错误	203
11.3.5 有条件更新	205
11.3.6 null 值处理	205
11.4 小测验	206
第 12 章 事务	208
12.1 多用户数据库	208
12.1.1 锁	208
12.1.2 锁的粒度	209

12.2 什么是事务	209
12.2.1 启动事务	211
12.2.2 结束事务	212
12.2.3 事务保存点	213
12.3 小测验	215
第 13 章 索引和约束	216
13.1 索引	216
13.1.1 创建索引	217
13.1.2 索引类型	220
13.1.3 如何使用索引	222
13.1.4 索引的不足	224
13.2 约束	225
13.2.1 创建约束	226
13.2.2 约束与索引	227
13.2.3 级联约束	227
13.3 小测验	230
第 14 章 视图	231
14.1 什么是视图	231
14.2 为什么使用视图	234
14.2.1 数据安全	234
14.2.2 数据聚合	235
14.2.3 隐藏复杂性	236
14.2.4 连接分区数据	236
14.3 可更新的视图	237
14.3.1 更新简单视图	238
14.3.2 更新复杂视图	239
14.4 小测验	241
第 15 章 元数据	242
15.1 关于数据的数据	242
15.2 信息模式	243
15.3 使用元数据	248

15.3.1	模式生成脚本	248
15.3.2	部署验证	251
15.3.3	生成动态 SQL	251
15.4	小测验	255
附录 A	示例数据库的 ER 图	256
附录 B	MySQL 对 SQL 语言的扩展	258
B.1	扩展 select 语句	258
B.1.1	limit 子句	258
B.1.2	into outfile 子句	261
B.2	组合 insert/update 语句	263
B.3	按排序更新和删除	265
B.4	多表更新与删除	266
附录 C	练习答案	270



在我们开始学习本书的内容时，先了解一些数据库方面的基本概念及计算机数据存储和检索的发展史是十分有益的。

1.1 数据库简介

“数据库”是指一组相关信息的集合。例如，电话簿就可以被视为包含某地区所有居民的姓名、电话号码、地址等信息的数据库。尽管电话簿可能是一个最为普及和常用的数据库，但它仍有不少缺点，比如：

- 查找某人的电话号码相当费时，特别是在电话簿包含了海量条目时；
- 电话簿只是根据姓名来索引，因此对于根据特定地址查找居民姓名就无能为力了；
- 当电话簿被打印后，随着该地区居民的流动、更改电话号码或住址等行为不断发生，电话簿上的信息也变得越来越不准确。

电话簿的这些缺陷同样存在于任何人工编制的数据存储系统，比如存放在档案柜的病历等。由于这些纸质数据库不方便，因此最早的计算机应用之一就是开发数据库系统，即通过计算机来存储和检索数据的机制。因为数据库系统通过电子而不是纸质方式来存储数据，所以它可以更快速地检索数据、以多种方式索引数据以及为其用户群提供最新的信息。

早期的数据库系统将被管理的数据存储在磁带上。一般情况下磁带的数量比磁带机要多得多，因此在请求数据时需要技术人员手动装卸磁带。同时由于那个时代的计算机内存很小，通常对同一数据的并发请求必须多次读取磁带，降低了使用效率。因此尽管这些数据库系统相对于纸质数据库有了显著的进步，但与今天的数据库技术相比仍有相当遥远的差距。（现代数据库系统能够利用海量快速的磁盘驱动器来管理太字节级的数据，在高速内存中存放数十吉字节的数据。）