

MANNING

官方团队编写 NoSQL排名第一

MongoDB IN ACTION
Second Edition

MongoDB

实战

(第二版)

Kyle Banker

Peter Bakkum

[美] Shaun Verch 著

Douglas Garrett

Tim Hawkin

徐雷 徐扬 译



华中科技大学出版社

<http://www.hustp.com>

第二版

MongoDB 实战

Kyle Banker
Peter Bakkum
[美] *Shaun Verch* 著
Douglas Garrett
Tim Hawkins
徐 雷 徐 扬 译

华中科技大学出版社

中国·武汉

内 容 提 要

本书分三部分通过大量的实例代码介绍了 MongoDB 数据库底层的实现以及大型互联网 Web 项目数据库设计原则。第一部分对 MongoDB 进行了整体介绍,并介绍了实际的开发例子,另外还介绍了 JavaScript shell 和 Ruby 驱动。第二部分通过逐步实现一个电商数据模型和实现必要的 CRUD 操作来详细介绍了 MongoDB 的文档数据模型、查询语言和 CRUD(新增、读取、更新和删除)操作。本书的最后部分从数据库专家的角度来看待 MongoDB,介绍了数据库的性能、部署、容错和伸缩性等所有的知识。

本书适合想深入学习 MongoDB 的开发人员,主要关注 MongoDB 数据库。

Original English language edition published by Manning Publications, USA. Copyright © 2015 by Manning Publications. Simplified Chinese-language edition copyright © 2016 by HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY PRESS. All rights reserved.

湖北省版权局著作权合同登记 图字:17-2017-015 号

图书在版编目(CIP)数据

MongoDB 实战:第二版/[美]科勒·班克等著;徐雷,徐杨译. —武汉:华中科技大学出版社,2017.3
ISBN 978-7-5680-2579-9

I. ①W… II. ①科… ②徐… ③徐… III. ①关系数据库系统 IV. ①TP311.138

中国版本图书馆 CIP 数据核字(2017)第 031379 号

MongoDB 实战(第二版)

MongoDB Shizhan

[美] Kyle Banker, Peter Bakkum, Shaun Verch 著
Douglas Garrett, Tim Hawkins 著
徐雷 徐扬 译

策划编辑:谢燕群

责任校对:李 琴

出版发行:华中科技大学出版社(中国·武汉)

武汉市东湖新技术开发区华工科技园

录 排:华中科技大学惠友文印中心

印 刷:湖北新华印务有限公司

开 本:787mm×960mm 1/16

印 张:26

字 数:660 千字

版 次:2017 年 3 月第 2 版第 1 次印刷

定 价:80.00 元

责任编辑:谢燕群

责任监印:周治超

电话:(027)81321913

邮编:430223



华中科大

本书若有印装质量问题,请向出版社营销中心调换
全国免费服务热线:400-6679-118 竭诚为您服务
版权所有 侵权必究

译者序

0. MongoDB 的优点与其学习重要性

MongoDB是最流行的NoSQL数据库！MongoDB在NoSQL中排名第一！它高性能、轻量级，易于扩展，适用于移动互联网敏捷发展，是架构师的必备！应用MongoDB现在也成为Google、Facebook、阿里巴巴、腾讯、百度等互联网公司招聘的关键技能！

1. MongoDB 在 NoSQL 中排名第一

非常荣幸可以翻译MongoDB官方团队撰写的《MongoDB in Action》第二版，这是MongoDB官方团队Kyle Banker和Shaun Verch等精心撰写的图书。此书是MongoDB领域最权威的图书，也是曼宁（Manning）出版社的“××× in Action”系列经典图书之一。本书受到全球读者的五星好评。中文版预计于2017年2月与中国读者见面！

2. 应用 MongoDB 的大型互联网公司

应用MongoDB的国外著名公司有：Google、Facebook、Tweet、思科、Bosch、Adobe、SAP等，国内著名公司有：阿里巴巴、腾讯、百度、360公司、新浪微博、京东、携程、滴滴打车等。新版MongoDB 3.X更是对云计算和大数据分析提供更强的支持，大数据工程师必须掌握MongoDB知识，其重要性不言而喻。

3. 本书优势

《MongoDB实战》（第二版）适合想深入学习MongoDB知识的开发人员，是架构师的必备。本书涵盖了最新版本内容，深入浅出、通过大量的实例代码介绍了MongoDB数据库底层的实现。比如，MongoDB的高可用集群、动态Schema、多服务器分布式数据存储。

4. 我选书的原则

我学习技术开发的原则就是坚持阅读国外经典图书，其次就是阅读官方文档。这个习惯让我受益匪浅。本书就是我们学习MongoDB高级开发的经典图书。本书适合MongoDB开发人员、架构师、DBA、大数据工程师学习！

5. 架构师必须能够对技术难点攻关

架构师必须写代码，必须能够进行技术攻关！作为核心骨干，最大的忌讳就是喜欢糊弄、忽悠，纸上谈兵。我比较反对一些观点：“30岁以后不适合搞技术，架构师不应该写代码”。技术团队的老大不懂技术怎么能带队伍！浮躁的领导会带坏整个公司的技术团队。只有理解框架底层的工作原理，才能在实际工作中分析和优化代码。

6. 致谢与反馈

在业余时间翻译本书的过程中，我还实战MongoDB、编写代码、验证本书中的知识点，比如高可用集群、安全等知识点；在Windows和Linux平台上做了实战部署；作为企业架构顾问参与设计架构方案。本人在翻译本书的过程中，收获很多，重新认识了MongoDB数据库，包括其底层的实现细节。感谢徐扬帮助翻译了部分章节，也要特别感谢华中科技大学出版社编辑们的辛苦工作，是他们才使得本书的翻译得以顺利完成并按期与读者见面。本书是新青年软件训练营（54peixun.com）高级架构班的推荐教材，希望本书能帮助大家深入学习MongoDB。

翻译本书过程中难免出现疏漏错误，如有反馈建议，欢迎加入中国MongoDB开发交流QQ群：511943641，或者微博@徐雷FrankXuLei联系我，非常感谢！

徐雷FrankXuLei

2017年1月18日

- 1) 新青年软件训练营（54peixun.com）联合创始人；
- 2) 微软中国MSDN特邀讲师、微软美国Channel9首位中国讲师；
- 3) 微软大企业客户技术顾问，曾任：购酒网、元拓商城、上海市公安局ERP系统北极绒技术架构顾问；
- 4) 获得吉林大学计算机科学与技术学士学位、上海交通大学硕士学位；
- 5) 翻译国外经典图书《WCF技术内幕》《WCF服务编程》第三版和第四版、《ASP.NET MVC4 Web编程》《jQuery实战》（第三版）、《MongoDB实战》（第二版）；
- 6) 受邀为微软中国、盛大网络、玫琳凯中国研发中心、世界500强约翰迪尔、一嗨租车、沪江网、中国东方航空、美国IGT、世界500强花旗银行、达丰集团、台达集团、美国国家仪器、上海交通大学软件学院、奥伯特石油、中国体彩集团总部、南方电网集团、阿里巴巴恒生证券、诺和诺德集团等中外名企、名校授课；
- 7) 2014年9月16日，受邀作为微软中国开发者技术代表会见ASP.NET之父Scott Gu先生，并作英文报告；
- 8) 为苍老师忠实粉丝，授课幽默风趣，追求“德艺双馨”。

关于封面插图

about the cover illustration

《MongoDB实战》封面描述的是法国东北部Burgundy地区的居民生活。这些作品属于19世纪的画家，由Louis Curmer编辑，1841年在巴黎出版。标题是*Les Français peints par eux-mêmes*，翻译成英文为*The French People Painted by Themselves*（法国人画中的自己）。每幅图画都色彩细腻，以手工着色，丰富展现了200年前世界各地、城镇、村庄以及邻里的生活姿态。由于彼此隔绝，不同地区的人说着不同的语言。在大街上或者乡下，很容易辨别他们的家乡以及通过服装判断他们的职业。

从那以后衣着也发生了很大变化，那个丰富多彩的时代已经远离我们了。现在很难辨别不同地区、国家的人民。也许我们已经牺牲了文化的多样性来换取个人生活的多样性——更多样、更快节奏的现代生活。

现在很难区分不同的计算机书籍，20年前，曼宁出版社开始采用多样性的区域生活图画来作为图书封面的设计创意，通过图书封面来介绍世界各地不同地区的历史文化与人文生活，让人们感受不同历史时期的文化，正如这本书一样。

关于作者

About the authors

Kyle Banker在10gen连续多年参与MongoDB Ruby驱动开发，之后进入创业公司工作。他写了许多关于MongoDB的技术博客（<http://kylebanker.com/blog>）。

Peter Bakkum使用MongoDB构建可伸缩的大型商业应用系统基础架构领域的专家。

Shaun Verch是就职于MongoDB核心服务器团队，参与研发，并在M101JS讲授一些关于Node.js使用MongoDB开发的课程。

Douglas Garrett是NYC首届全球MongoDB技术大会的演讲嘉宾，MongoDB数据分析创新大奖的获得者，就职于Genentech的生物信息学部门。

Tim Hawkins是Yahoo欧洲搜索引擎的团队“老大”，专注于国际化与可伸缩性领域的解决方案架构研究。

致谢

Acknowledgments

感谢曼宁出版社朋友的帮助，使得本书可以顺利出版。Michael Stephens帮助策划了本书的第一版，对第二版的编辑Susan Conant、Jeff Bleiel、Maureen Spencer在此一并感谢。

写作是个非常耗时的工作。如果没有Eliot Horowitz（MongoDB 公司的CTO）和Dwight Merriman（MongoDB 公司的主席）的慷慨帮助，我可能也无法完成本书。凭借Eliot和Dwight的热情和聪明才智，他们创建了MongoDB，而且他们信任我，让我来参与文档的编写工作。非常感谢他们。

本书中的许多观点收集于我们与10gen公司员工的谈话。因此特别感谢Mike Dirolf、Scott Hernandez、Alvin Richards、Mathias Stearn，还要特别感谢Kristina Chowdorow、Richard Kreuter、Aaron Staple。他们参与了本书第一版许多章节的审核工作，提出了许多专业性的修改建议。

在本书第一版编写阶段，感谢以下朋友参与了初稿的校订工作：Kevin Jackson、Hardy Ferentschik、David Sinclair、Chris Chandler、John Nunemaker、Robert Hanson、Alberto Lerner、Rick Wagner、Ryan Cox、Andy Brudtkuhl、Daniel Bretoi、Greg Donald、Sean Reilly、Curtis Miller、Sanchet Dighe、Philip Hallstrom、Andy Dingley。还要感谢参与本书第二版审核的朋友：Agustin Treceno、Basheeruddin Ahmed、Gavin Whyte、George Girton、Gregor Zurowski、Hardy Ferentschik、Hernan Garcia、Jeet Marwah、Johan Mattisson、Jonathan Thoms、Julia Varigina、Jürgen Hoffmann、Mike Frey、Phlippie Smith、Scott Lyons、Steve Johnson。

特别感谢Wouter Thielen对第10章内容所做的工作，以及技术编辑Mihalis Tsoukalos、参与技术审阅的Doug Warren。

在本书第二版编写期间，我美丽贤惠的妻子Dominika给予了最大的耐心和支持，还有我帅气的儿子Oliver，他非常乖巧听话。非常感谢家人的支持。

Kyle Banker

序言

Preface

数据库是信息时代的基石。与Atlas(微软1998年发明的Ajax技术)很像,它们默默无闻地支持着我们使用的数字世界。一定要记住,我们的数字交互从评论到推特(美国的微博)搜索和排序,本质上都是与数据库交互。

这种基本的但是隐藏的功能,让我通常对数据库保留一份敬意,与我们穿过悬索桥的感觉不同,通常我们会对汽车保留一份敬畏。

数据库有许多种形式。书籍的索引以及目录卡片都是数据库排序的一种形式,正如为过去Perl程序员设计的特别结构化文本文件。可能现在最著名的就是复杂的关系型数据库,它们构成了当今世界软件的基础。这些关系型数据库,以及它们的第三范式和SQL接口,仍然稳定地运行着。

在从事Web开发工作几年后,我就迫切想寻找关系型数据库的替代品。当我们看到MongoDB以后,真是欣喜若狂。我喜欢用类JSON的数据结构来表示数据的想法。JSON非常简单、直观,而且易于阅读。MongoDB也使用了JSON语言来构建查询,利于新数据库的使用和推广。高级特性如易于复制和分片使得MongoDB更加强大。到目前为止,我已经基于MongoDB构建了一些应用,体验到它的简便性,我已经彻底“路转粉”爱上它了。

好事多磨。经历许多坎坷,我加入了10gen,一家专门领导开源数据库开发的公司。这两年间,我参与了许多客户端驱动的开发和改进工作,与许多客户公司合作。这些经历让我收获颇多,我也希望把一些宝贵的经验在本书中与大家分享。

作为持续不断开发的软件产品,MongoDB仍然不够完美,还在持续完善。但是,MongoDB集群已经成功支持了成千上万的大小应用,而且日益成熟。许多开发者已经听说过它的强大之处,不仅可以创造奇迹,而且能够带来快乐。希望它也可以改变你的工作。

这是《MongoDB实战》的第二版,我希望你能享受阅读与学习本书的过程。

Kyle Banker

关于本书

about this book

本书适合从零开始深入学习MongoDB的开发者和DBA。如果你是个新手，你会发现本书的节奏很好，深入浅出。如果你是有经验的开发者，本书也提供了针对高级知识更加详细的参考。关于本书的内容深度，本书应该适合绝大多数高级开发者。虽然本书内容针对最新的MongoDB版本（当前已经是3.0.x版本），但是也涵盖了之前的MongoDB版本（比如2.6版本）内容。

例子代码由JavaScript、MongoDB shell的语言以及另外一种流行的脚本语言Ruby编写。每次努力都是为了提供更加简单有用的例子，而且仅仅使用了JavaScript和Ruby语言最朴素的特性。主要的目的还是展示MongoDB API最常见的使用方式。如果你已经使用过其他语言，就应该会发现这些例子很容易理解。

关于语言，再强调一点。如果你好奇“为什么不使用那个×语言”，那么你可以放心，官方支持的MongoDB驱动具备一致的功能和API。这意味着只要学习一个语言的API，其他语言的就可以融会贯通了。

如何使用本书

How to use this book

本书部分是教程，部分是参考资料。如果你完全是个MongoDB“菜鸟”，最好还是按照顺序阅读，这样收获最大。本书包含了许多例子代码，可以帮助大家来巩固概念。不过要求大家安装MongoDB和选装Ruby驱动。附录A里有安装指南^[1]。

如果你已经学习过MongoDB，那么可以阅读专门的高级主题。第8~13章以及其余的附录都是独立的知识点，可以任意顺序阅读。此外，第4~7章包含了基本要点，主要关注MongoDB的核心概念。这些内容也可以独立学习。

^[1]【译注】安装MongoDB驱动可以使用包管理工具，.NET有NuGet，Node.js有NPM，十分方便。实在不行，还可以从官网下载驱动包手动加入项目的代码中。

路线图

Roadmap

本书分为三大部分。

第一部分从零开始完整介绍MongoDB。第1章介绍了MongoDB的历史、特性以及用例。第2章通过MongoDB命令shell教会大家核心的概念。第3章介绍如何使用MongoDB设计简单的应用后台数据库。

第二部分详细介绍了第一部分使用的MongoDB API。分4章内容逐步介绍了电商应用schema和操作。第4章深入介绍了文档document，MongoDB里存储数据的最小单元，而且介绍了基本的电商schema设计的例子。第5~7章介绍如何通过查询和更新来使用schema。为了更加细化内容，每章都包含一个更详细的子主题。

第三部分主要是关于MongoDB的高级主题。第8章介绍了索引机制和查询优化的内容。第9章介绍了MongoDB内部的文本搜索引擎。

第10章是本书中的新增内容，关于WiredTiger存储引擎和可拔插存储，这是MongoDB v3之后独有的特性。第11章关注复制，使用MongoDB 高可用策略以及伸缩性的内容。第12章介绍了分片sharding存储、MongoDB的水平伸缩特性。

第13章提供了MongoDB安装、部署、管理和解决错误的最佳实践经验。

本书还包含3个附录部分。附录A介绍了如何在Linux、Mac OS X、Windows下安装MongoDB和Ruby驱动。附录B介绍了一系列schema和设计模式，还有一些反模式列表。附录C介绍了如何在MongoDB存储二进制数据，以及如何使用GridFS，所有驱动都实现同一个规范，这个规范是专门用来存储大文件的技术。

代码约定和下载

Code conventions and downloads

所有的列表代码以及文本代码都使用了固定宽度的字体，以区分其他文字。

对于列表或者高亮的概念会加设代码注释。

有时候，会在文本里使用编号的链接来解释概念。

作为一个开源项目，10gen负责处理社区反馈的所有MongoDB的bug。本书的几个地方，特别是脚注部分，你会参考bug包含的参考建议，以及计划的改进方案。例如，在SERVER-380增加全文搜索功能的申请票据。要查询任意票据的状态，可以浏览<http://jira.mongodb.org>，然后输入ID搜索。

大家也可以从<http://mongodb-book.com>或者<http://manning.com/MongoDBinAction>下载本书的源码，包括一些例子数据。

软件需求

在开始学习本书之前，需要安装MongoDB。

MongoDB的安装指南可以参考附录A以及MongoDB官方网站(<http://mongodb.org>)。

如果要运行Ruby驱动的例子，需要安装Ruby。请参考附录A的安装指南。

目录

Table of Contents

第一部分 入门.....	1
第 1 章 全新 Web 数据库.....	3
1.1 为互联网而生.....	5
1.2 MongoDB 键特性.....	5
1.2.1 文档数据模型.....	5
1.2.2 ad hoc 查询.....	9
1.2.3 索引.....	9
1.2.4 复制.....	10
1.2.5 加速与持久化.....	11
1.2.6 伸缩.....	13
1.3 核心服务和工具.....	14
1.3.1 核心服务器.....	14
1.3.2 JavaScript shell.....	15
1.3.3 数据库驱动.....	15
1.3.4 命令行工具.....	16
1.4 为什么是 MongoDB?.....	17
1.4.1 MongoDB 与其他数据库对比.....	17
1.4.2 使用场景和部署.....	20
1.5 提示和限制.....	22
1.6 MongoDB 历史.....	23
1.7 其他资源.....	25
1.8 总结.....	25
第 2 章 通过 JavaScript shell 操作 MongoDB.....	27
2.1 Diving into the 深入 MongoDB shell.....	28
2.1.1 启动 shell.....	28
2.1.2 数据库、集合和文档.....	28
2.1.3 插入和查询.....	29
2.1.4 更新文档.....	31
2.1.5 删除数据.....	35

2.1.6 shell 的其他特性	35
2.2 使用索引创建和查询	36
2.2.1 创建大集合	36
2.2.2 索引和 explain()	38
2.3 基本管理	42
2.3.1 获取数据库信息	43
2.3.2 命令如何执行	44
2.4 获取帮助	45
2.5 总结	47
第 3 章 编写代码操作 MongoDB	48
3.1 通过 Ruby lens 连接 MongoDB	49
3.1.1 安装与连接	49
3.1.2 Ruby 里插入文档数据	50
3.1.3 查询与光标	51
3.1.4 更新和删除	52
3.1.5 数据库命令	53
3.2 驱动工作原理	54
3.3 构建简单的应用	56
3.3.1 设置	56
3.3.2 搜集数据	57
3.3.3 查看存档	60
3.4 总结	63
第二部分 MongoDB 应用系统开发	65
第 4 章 面向文档的数据	67
4.1 schema 设计原则	67
4.2 设计电商网站数据模型	69
4.2.1 schema 基础知识	69
4.2.2 用户和订单	73
4.2.3 评价	75
4.3 核心概念：数据库、集合、文档	76
4.3.1 数据库	76
4.3.2 集合	79
4.3.3 文档和插入	83
4.4 总结	87
第 5 章 构建查询	88
5.1 电子商务查询	88
5.1.1 产品、类别和评论	88

5.1.2 用户和订单.....	91
5.2 MongoDB 的查询语言	92
5.2.1 查询条件和选择器	92
5.2.2 查询选择	104
5.3 总结	106
第 6 章 聚合	107
6.1 聚合框架概览	108
6.2 电商聚合例子	109
6.2.1 商品、类别和评价	111
6.2.2 用户和订单	117
6.3 聚合管道操作符	120
6.3.1 \$project	120
6.3.2 \$group	121
6.3.3 \$match、\$sort、\$skip、\$limit	123
6.3.4 \$unwind	123
6.3.5 \$out	124
6.4 重塑文档	124
6.4.1 字符串函数	125
6.4.2 算术运算函数	126
6.4.3 日期函数	126
6.4.4 逻辑函数	127
6.4.5 集合操作符	128
6.4.6 其他函数	129
6.5 理解聚合管道性能	129
6.5.1 聚合管道选项	130
6.5.2 聚合框架的 explain() 函数	130
6.5.3 allowDiskUse 选项	134
6.5.4 聚合光标选项	134
6.6 其他聚合功能	135
6.6.1 .count() 和 distinct()	135
6.6.2 map-reduce	136
6.7 总结	138
第 7 章 更新、原子操作和删除	140
7.1 文档更新概要	141
7.1.1 通过替换修改	141
7.1.2 通过操作符修改	142
7.1.3 比较两个方法	142
7.1.4 决定：替换与操作符	143

7.2 电商数据模型更新	144
7.2.1 商品和目录	144
7.2.2 评价	148
7.2.3 订单	150
7.3 原子文档处理	152
7.3.1 订单状态转换	153
7.3.2 库存管理	155
7.4 核心要点: MongoDB 更新与删除	160
7.4.1 更新类型与参数选项	160
7.4.2 更新操作符	161
7.4.3 findAndModify 命令	169
7.4.4 删除	169
7.4.5 并发、原子性和隔离	170
7.4.6 更新性能注意事项	171
7.5 复习更新操作符	172
7.6 总结	173
第三部分 精通 MongoDB	175
第 8 章 索引与查询优化	177
8.1 索引理论	177
8.1.1 精心策划的实验	178
8.1.2 核心索引概念	181
8.1.3 B-树	185
8.2 索引实战	186
8.2.1 索引类型	186
8.2.2 索引管理	189
8.3 查询优化	194
8.3.1 找出慢速查询	195
8.3.2 检查慢速查询	199
8.3.3 查询模式	217
8.4 总结	219
第 9 章 文本搜索	220
9.1 文本搜索——不仅仅是模式匹配	221
9.1.1 文本搜索与模式匹配	222
9.1.2 文本搜索与网页搜索	223
9.1.3 MongoDB 文本搜索与专业搜索引擎	225
9.2 下载曼宁图书类别数据	228
9.3 定义文本搜索索引	229
9.3.1 文本索引的大小	230

9.3.2	分配索引名字并为集合里的所有字段建立索引	231
9.4	基本的文本搜索	232
9.4.1	更复杂的搜索	233
9.4.2	文本搜索分数	235
9.4.3	根据文本搜索分数排序结果	236
9.5	聚合框架文本搜索	237
9.6	文本搜索语言	240
9.6.1	在索引里指定语言	241
9.6.2	在文档里指定语言	242
9.6.3	在搜索中指定语言	243
9.6.4	可用的语言	245
9.7	总结	245
第 10 章	WiredTiger 与可拔插存储	246
10.1	可拔插存储引擎 API	246
10.2	WiredTiger	248
10.2.1	切换到 WiredTiger	248
10.2.2	迁移数据到 WiredTiger	249
10.3	与 MMAPv1 对比	250
10.3.1	配置文件	251
10.3.2	插入脚本与基准测试脚本	252
10.3.3	插入测试结果	255
10.3.4	读性能测试脚本	256
10.3.5	读性能结果	257
10.3.6	测试结论	259
10.4	其他可拔插存储引擎的例子	260
10.5	高级主题	261
10.5.1	可拔插引擎如何工作?	261
10.5.2	数据结构	263
10.5.3	锁	265
10.6	总结	265
第 11 章	复制	267
11.1	复制概览	267
11.1.1	为什么复制很重要	268
11.1.2	复制的使用场景和限制	269
11.2	可复制集	270
11.2.1	安装	270
11.2.2	可复制集群工作原理	277
11.2.3	管理	283