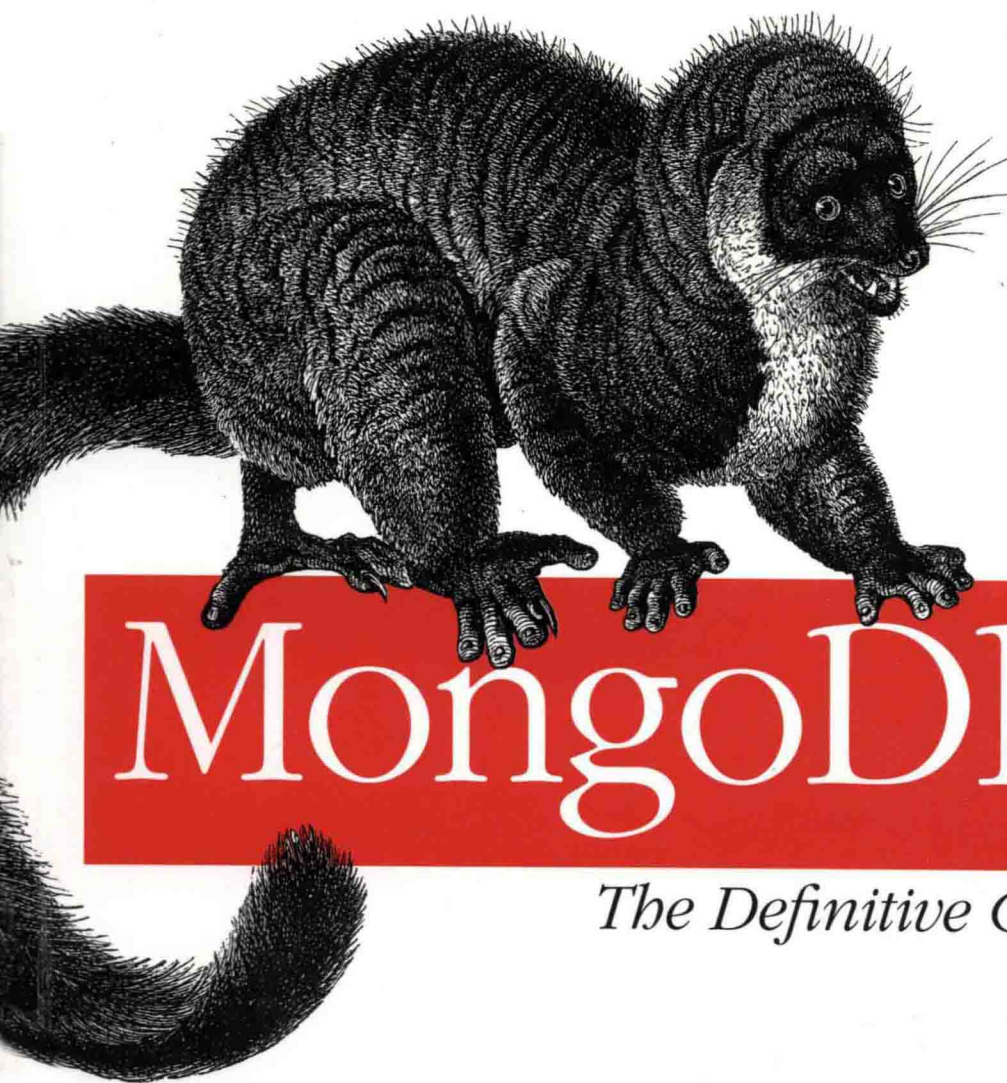


MongoDB权威指南 (影印版)

第二版



MongoDB

The Definitive Guide

O'REILLY®

東南大學出版社

Kristina Chodorow 著

MongoDB权威指南(影印版)

MongoDB如何帮你管理通过Web应用收集的海量数据呢?这本权威指南由项目的核心人员编写,为你展示了使用基于文档的数据库的诸多优势,并且演示了这种可靠的、高性能的系统是如何允许接近无限的水平扩展性。

本次第二版修订为数据库开发人员提供了指南,为系统管理员提供了高级配置向导,也为你的项目中的其他人员提供了概念和用例的概览。这本指南适合于NoSQL的初学者和有经验的MongoDB用户,它提供了多种现实世界的模式设计样例。

- 熟悉MongoDB核心概念和术语
- 在不同的安全程度和速度上执行基本的写操作
- 创建复杂的查询语句,带有用来限制、跳过和排序结果的选项
- 设计一个可以和MongoDB协同工作的应用
- 聚合数据,包括统计、查找唯一值、文档分组和使用MapReduce
- 整理和解释针对采集数据和数据库的统计结果
- 在MongoDB中建立主从集群和自动故障恢复机制
- 利用分片水平扩展MongoDB,了解其对应应用的影响
- 深入研究监控、安全和认证、备份和修复以及其他管理任务

Kristina Chodorow是一名Google公司的软件工程师。她为MongoDB的核心部分工作了五年。她领导了MongoDB主从集群的开发工作并且编写了PHP和Perl的驱动部分。

O'REILLY®
oreilly.com.cn

责任编辑:张烨

封面设计:Randy Comer, 张健

O'Reilly Media, Inc. 授权东南大学出版社出版

此影印版仅限于在中华人民共和国境内(但不允许在中国香港、澳门特别行政区和中国台湾地区)销售发行

This Authorized Edition for sale only in the territory of People's Republic of China (excluding

Hong Kong, Macao and Taiwan)

ISBN 978-7-5641-4602-3



9 787564 146023 >

定价: 64.00元(册)



第二版

MOODS 权威指南

(影印版)

Chodorow

著

东南大学出版社

O'REILLY

第二版

MongoDB权威指南 (影印版)

MongoDB: The Definitive Guide

Kristina Chodorow 著

O'REILLY®

Beijing · Cambridge · Farnham · Köln · Sebastopol · Tokyo

© Reilly Media, Inc. 授权东南大学出版社出版

南京 东南大学出版社

图书在版编目 (CIP) 数据

MongoDB 权威指南: 第2版: 英文 / (美) 霍多罗夫 (Chodorow, K.) 著. —影印本. —南京: 东南大学出版社, 2014.1

书名原文: MongoDB: The Definitive Guide, 2E

ISBN 978-7-5641-4602-3

I. ① M… II. ① 霍… III. ① 关系数据库系统—英文
IV. ① TP311.138

中国版本图书馆 CIP 数据核字 (2013) 第 246107 号

江苏省版权局著作权合同登记

图字: 10-2013-375 号

©2013 by O'Reilly Media, Inc.

Reprint of the English Edition, jointly published by O'Reilly Media, Inc. and Southeast University Press, 2014. Authorized reprint of the original English edition, 2013 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 2013。

英文影印版由东南大学出版社出版 2014。此影印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。

MongoDB 权威指南 第二版 (影印版)

出版发行: 东南大学出版社

地 址: 南京四牌楼 2 号 邮编: 210096

出 版 人: 江建中

网 址: <http://www.seupress.com>

电子邮件: press@seupress.com

印 刷: 扬中市印刷有限公司

开 本: 787 毫米 × 980 毫米 16 开本

印 张: 27

字 数: 529 千字

版 次: 2014 年 1 月第 1 版

印 次: 2014 年 1 月第 1 次印刷

书 号: ISBN 978-7-5641-4602-3

定 价: 64.00 元 (册)

本社图书若有印装质量问题, 请直接与营销部联系。电话 (传真): 025-83791830

此为试读, 需要完整PDF请访问: www.ertongbook.com

Foreword

In the last 10 years, the Internet has challenged relational databases in ways nobody could have foreseen. Having used MySQL at large and growing Internet companies during this time, I've seen this happen firsthand. First you have a single server with a small data set. Then you find yourself setting up replication so you can scale out reads and deal with potential failures. And, before too long, you've added a caching layer, tuned all the queries, and thrown even more hardware at the problem.

Eventually you arrive at the point when you need to shard the data across multiple clusters and rebuild a ton of application logic to deal with it. And soon after that you realize that you're locked into the schema you modeled so many months before.

Why? Because there's so much data in your clusters now that altering the schema will take a long time and involve a lot of precious DBA time. It's easier just to work around it in code. This can keep a small team of developers busy for many months. In the end, you'll always find yourself wondering if there's a better way—or why more of these features are not built into the core database server.

Keeping with tradition, the Open Source community has created a plethora of “better ways” in response to the ballooning data needs of modern web applications. They span the spectrum from simple in-memory key/value stores to complicated SQL-speaking MySQL/InnoDB derivatives. But the sheer number of choices has made finding the right solution more difficult. I've looked at many of them.

I was drawn to MongoDB by its pragmatic approach. MongoDB doesn't try to be everything to everyone. Instead it strikes the right balance between features and complexity, with a clear bias toward making previously difficult tasks far easier. In other words, it has the features that really matter to the vast majority of today's web applications: indexes, replication, sharding, a rich query syntax, and a very flexible data model. All of this comes without sacrificing speed.

Like MongoDB itself, this book is very straightforward and approachable. New MongoDB users can start with Chapter 1 and be up and running in no time. Experienced

users will appreciate this book's breadth and authority. It's a solid reference for advanced administrative topics such as replication, backups, and sharding, as well as popular client APIs.

Having recently started to use MongoDB in my day job, I have no doubt that this book will be at my side for the entire journey—from the first install to production deployment of a sharded and replicated cluster. It's an essential reference to anyone seriously looking at using MongoDB.

—Jeremy Zawodny
Craigslist Software Engineer
August 2010

How This Book Is Organized

This book is split up into six sections, covering development, administration, and deployment information.

Getting Started with MongoDB

In Chapter 1 we provide background about MongoDB: why it was created, the goals it is trying to accomplish, and why you might choose to use it for a project. We go into more detail in Chapter 2, which provides an introduction to the core concepts and vocabulary of MongoDB. Chapter 2 also provides a first look at working with MongoDB, getting you started with the database and the shell. The next two chapters cover the basic material that developers need to know to work with MongoDB. In Chapter 3, we describe how to perform those basic write operations, including how to do them with different levels of safety and speed. Chapter 4 explains how to find documents and create complex queries. This chapter also covers how to iterate through results and gives options for limiting, skipping, and sorting results.

Developing with MongoDB

Chapter 5 covers what indexing is and how to index your MongoDB collections. Chapter 6 explains how to use several special types of indexes and collections. Chapter 7 covers a number of techniques for aggregating data with MongoDB, including counting, finding distinct values, grouping documents, the aggregation framework, and using MapReduce. Finally, this section finishes with a chapter on designing your application: Chapter 8 goes over tips for writing an application that works well with MongoDB.

Replication

The replication section starts with Chapter 9, which gives you a quick way to set up a replica set locally and covers many of the available configuration options. Chapter 10 then covers the various concepts related to replication. Chapter 11 shows how replication interacts with your application and Chapter 12 covers the administrative aspects of running a replica set.

Sharding

The sharding section starts in Chapter 13 with a quick local setup. Chapter 14 then gives an overview of the components of the cluster and how to set them up. Chapter 15 has advice on choosing a shard key for a variety of application. Finally, Chapter 16 covers administering a sharded cluster.

Application Administration

The next two chapters cover many aspects of MongoDB administration from the perspective of your application. Chapter 17 discusses how to introspect what MongoDB is doing. Chapter 18 covers administrative tasks such as building indexes, and moving and compacting data. Chapter 19 explains how MongoDB stores data durably.

Server Administration

The final section is focused on server administration. Chapter 20 covers common options when starting and stopping MongoDB. Chapter 21 discusses what to look for and how to read stats when monitoring. Chapter 22 describes how to take and restore backups for each type of deployment. Finally, Chapter 23 discusses a number of system settings to keep in mind when deploying MongoDB.

Appendixes

Appendix A explains MongoDB's versioning scheme and how to install it on Windows, OS X, and Linux. Appendix B details how MongoDB works internally: its storage engine, data format, and wire protocol.

Conventions Used in This Book

The following typographical conventions are used in this book:

Italic

Indicates new terms, URLs, email addresses, collection names, database names, filenames, and file extensions.

Constant width

Used for program listings, as well as within paragraphs to refer to program elements such as variable or function names, command-line utilities, environment variables, statements, and keywords.

Constant width bold

Shows commands or other text that should be typed literally by the user.

Constant width italic

Shows text that should be replaced with user-supplied values or by values determined by context.



This icon signifies a tip, suggestion, or general note.



This icon indicates a warning or caution.

Using Code Examples

This book can help you get your job done. In general, you may use the code in this book in your programs and documentation. You do not need to contact us for permission unless you're reproducing a significant portion of the code. For example, writing a program that uses several chunks of code from this book does not require permission. Selling or distributing a CD-ROM of examples from O'Reilly books does require permission. Answering a question by citing this book and quoting example code does not require permission. Incorporating a significant amount of example code from this book into your product's documentation does require permission.

We appreciate, but do not require, attribution. An attribution usually includes the title, author, publisher, and ISBN. For example: “*MongoDB: The Definitive Guide, Second Edition* by Kristina Chodorow (O'Reilly). Copyright 2013 Kristina Chodorow, 978-1-449-34468-9.”

If you feel your use of code examples falls outside fair use or the permission given here, feel free to contact us at permissions@oreilly.com.

Safari® Books Online



Safari Books Online (www.safaribooksonline.com) is an on-demand digital library that delivers expert content in both book and video form from the world's leading authors in technology and business.

Technology professionals, software developers, web designers, and business and creative professionals use Safari Books Online as their primary resource for research, problem solving, learning, and certification training.

Safari Books Online offers a range of product mixes and pricing programs for organizations, government agencies, and individuals. Subscribers have access to thousands of books, training videos, and prepublication manuscripts in one fully searchable database from publishers like O'Reilly Media, Prentice Hall Professional, Addison-Wesley Professional, Microsoft Press, Sams, Que, Peachpit Press, Focal Press, Cisco Press, John Wiley & Sons, Syngress, Morgan Kaufmann, IBM Redbooks, Packt, Adobe Press, FT Press, Apress, Manning, New Riders, McGraw-Hill, Jones & Bartlett, Course Technology, and dozens more. For more information about Safari Books Online, please visit us online.

How to Contact Us

Please address comments and questions concerning this book to the publisher:

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472
800-998-9938 (in the United States or Canada)
707-829-0515 (international or local)
707 829-0104 (fax)

We have a web page for this book, where we list errata, examples, and any additional information. You can access this page at:

<http://oreil.ly/mongodb-2e>

To comment or ask technical questions about this book, send email to:

bookquestions@oreilly.com

For more information about our books, conferences, Resource Centers, and the O'Reilly Network, see our website at:

<http://www.oreilly.com>

Acknowledgments

I would like to thank my tech reviewers, Adam Comerford, Eric Milke, and Greg Studer. You guys made this book immeasurably better (and more correct). Thank you, Ann Spencer, for being such a terrific editor and for helping me every step of the way. Thanks to all of my coworkers at 10gen for sharing your knowledge and advice on MongoDB as well as Eliot Horowitz and Dwight Merriman, for starting the MongoDB project. And thank you, Andrew, for all of your support and suggestions.

Table of Contents

Foreword.....	xiii
Preface.....	xv

Part I. Introduction to MongoDB

1. Introduction.....	3
Ease of Use	3
Easy Scaling	3
Tons of Features...	4
...Without Sacrificing Speed	5
Let's Get Started	5
2. Getting Started.....	7
Documents	7
Collections	8
Dynamic Schemas	8
Naming	9
Databases	10
Getting and Starting MongoDB	11
Introduction to the MongoDB Shell	12
Running the Shell	13
A MongoDB Client	13
Basic Operations with the Shell	14
Data Types	16
Basic Data Types	16
Dates	18
Arrays	18
Embedded Documents	19
_id and ObjectIds	20

Using the MongoDB Shell	21
Tips for Using the Shell	22
Running Scripts with the Shell	23
Creating a .mongorc.js	25
Customizing Your Prompt	26
Editing Complex Variables	27
Inconvenient Collection Names	27
3. Creating, Updating, and Deleting Documents.....	29
Inserting and Saving Documents	29
Batch Insert	29
Insert Validation	30
Removing Documents	31
Remove Speed	31
Updating Documents	32
Document Replacement	32
Using Modifiers	34
Upserts	45
Updating Multiple Documents	47
Returning Updated Documents	48
Setting a Write Concern	51
4. Querying.....	53
Introduction to find	53
Specifying Which Keys to Return	54
Limitations	55
Query Criteria	55
Query Conditionals	55
OR Queries	56
\$not	57
Conditional Semantics	57
Type-Specific Queries	58
null	58
Regular Expressions	58
Querying Arrays	59
Querying on Embedded Documents	63
\$where Queries	65
Server-Side Scripting	66
Cursors	67
Limits, Skips, and Sorts	68
Avoiding Large Skips	70
Advanced Query Options	71

Getting Consistent Results	72
Immortal Cursors	75
Database Commands	75
How Commands Work	76

Part II. Designing Your Application

5. Indexing.....	81
Introduction to Indexing	81
Introduction to Compound Indexes	84
Using Compound Indexes	89
How \$-Operators Use Indexes	91
Indexing Objects and Arrays	95
Index Cardinality	98
Using explain() and hint()	98
The Query Optimizer	102
When Not to Index	102
Types of Indexes	104
Unique Indexes	104
Sparse Indexes	106
Index Administration	107
Identifying Indexes	108
Changing Indexes	108
6. Special Index and Collection Types.....	109
Capped Collections	109
Creating Capped Collections	111
Sorting Au Naturel	112
Tailable Cursors	113
No-_id Collections	114
Time-To-Live Indexes	114
Full-Text Indexes	115
Search Syntax	118
Full-Text Search Optimization	119
Searching in Other Languages	119
Geospatial Indexing	120
Types of Geospatial Queries	120
Compound Geospatial Indexes	121
2D Indexes	122
Storing Files with GridFS	123
Getting Started with GridFS: mongofiles	124

Working with GridFS from the MongoDB Drivers	124
Under the Hood	125
7. Aggregation.....	127
The Aggregation Framework	127
Pipeline Operations	129
\$match	129
\$project	130
\$group	135
\$unwind	137
\$sort	139
\$limit	139
\$skip	139
Using Pipelines	140
MapReduce	140
Example 1: Finding All Keys in a Collection	140
Example 2: Categorizing Web Pages	143
MongoDB and MapReduce	143
Aggregation Commands	146
count	146
distinct	147
group	147
8. Application Design.....	153
Normalization versus Denormalization	153
Examples of Data Representations	154
Cardinality	157
Friends, Followers, and Other Inconveniences	158
Optimizations for Data Manipulation	160
Optimizing for Document Growth	160
Removing Old Data	162
Planning Out Databases and Collections	162
Managing Consistency	163
Migrating Schemas	164
When Not to Use MongoDB	165

Part III. Replication

9. Setting Up a Replica Set.....	169
Introduction to Replication	169
A One-Minute Test Setup	170