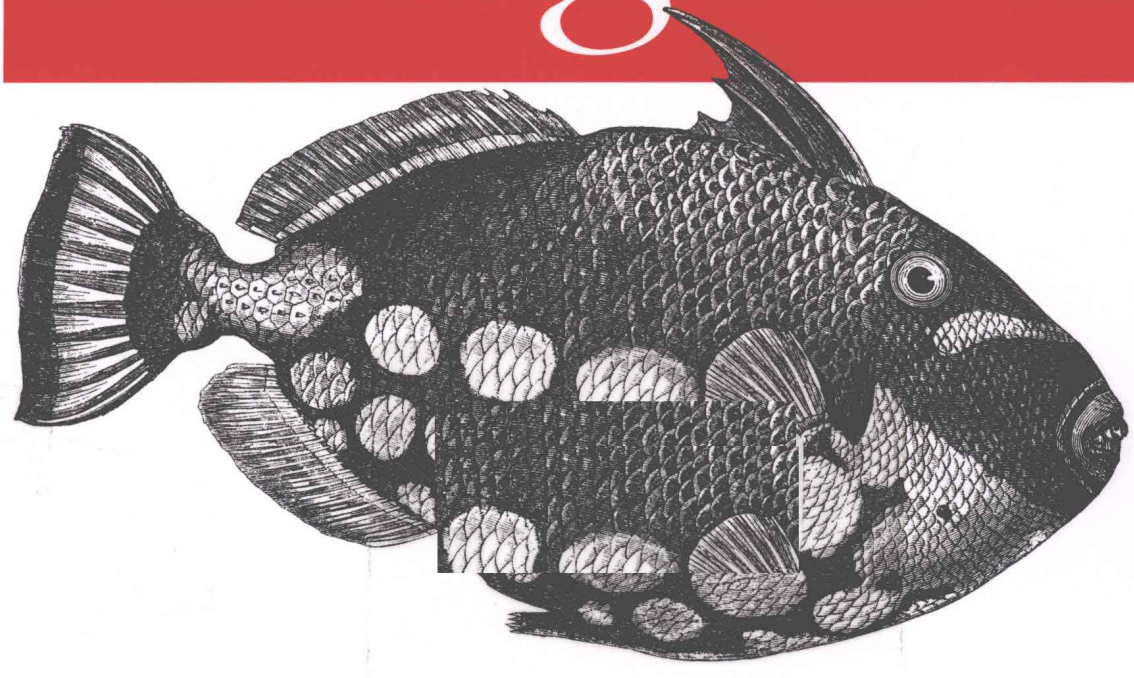
TURING 图灵程序设计丛书

Scaling MongoDB
50 Tips and Tricks for MongoDB Developers

深入学习 MongoDB



[美] *Kristina Chodorow* 著
巨成 程显峰 译

O'REILLY®

人民邮电出版社
POSTS & TELECOM PRESS

TURING 图灵程序设计丛书

深入学习MongoDB

Scaling MongoDB
50 Tips and Tricks for MongoDB Developers

[美] Kristina Chodorow 著
巨成 程显峰 译

O'REILLY®
Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo
O'Reilly Media, Inc.授权人民邮电出版社出版

人民邮电出版社
北 京

图书在版编目（CIP）数据

深入学习MongoDB / (美) 霍多罗夫 (Chodorow, K.) 著 ; 巨成, 程显峰译. -- 北京 : 人民邮电出版社, 2012.3
(图灵程序设计丛书)
ISBN 978-7-115-27211-9

Ⅰ. ①深… Ⅱ. ①霍… ②巨… ③程… Ⅲ. ①关系数据库系统 Ⅳ. ①TP311.138

中国版本图书馆CIP数据核字(2011)第267232号

内 容 提 要

本书分两部分，分别对应O'Reilly公司出版的*Scaling MongoDB*和*50 Tips and Tricks for MongoDB Developers*两本书的内容。第一部分全面讲解了有关建立和使用集群的内容，不仅从应用开发人员的角度讲解了MongoDB的使用，而且从运维方面介绍了集群的管理。其中内容包括通过分片设置MongoDB集群，分片的工作原理，查询和更新数据，操作、监控和备份集群，错误处理。第二部分依次从应用设计、实现、优化、数据安全和管理方面介绍了使用MongoDB构建应用的技巧，内容包括范式化与反范式化的利弊权衡，复制组的故障恢复等。

本书适合所有MongoDB用户阅读参考。

图灵程序设计丛书

深入学习MongoDB

◆ 著 [美] Kristina Chodorow
译 巨 成 程显峰
责任编辑 卢秀丽
执行编辑 毛倩倩

◆ 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 电子邮件 315@ptpress.com.cn
网址 http://www.ptpress.com.cn
北京艺辉印刷有限公司印刷

◆ 开本：800×1000 1/16
印张：8.5
字数：150千字 2012年1月第1版
印数：1－5 000册 2012年3月北京第1次印刷
著作权合同登记号 图字：01-2011-8108号
ISBN 978-7-115-27211-9

定价：32.00元

读者服务热线：(010)51095186转604 印装质量热线：(010)67129223
反盗版热线：(010)67171154

版权声明

O'Reilly Media, Inc.介绍

O'Reilly Media 通过图书、杂志、在线服务、调查研究和会议等方式传播创新知识。自 1978 年开始，O'Reilly 一直都是前沿发展的见证者和推动者。超级极客们正在开创着未来，而我们关注真正重要的技术趋势——通过放大那些“细微的信号”来刺激社会对新科技的应用。作为技术社区中活跃的参与者，O'Reilly 的发展充满了对创新的倡导、创造和发扬光大。

O'Reilly 为软件开发人员带来革命性的“动物书”；创建第一个商业网站（GNN）；组织了影响深远的开放源代码峰会，以至于开源软件运动以此命名；创立了 Make 杂志，从而成为 DIY 革命的主要先锋；公司一如既往地通过多种形式缔结信息与人的纽带。O'Reilly 的会议和峰会聚集了众多超级极客和高瞻远瞩的商业领袖，共同描绘出开创新产业的革命性思想。作为技术人士获取信息的选择，O'Reilly 现在还将先锋专家的知识传递给普通的计算机用户。无论是通过书籍出版、在线服务或者面授课程，每一项 O'Reilly 的产品都反映了公司不可动摇的理念——信息是激发创 新的力量。

业界评论

“O'Reilly Radar 博客有口皆碑。”

——Wired

“O'Reilly 凭借一系列（真希望当初我也想到了）非凡想法建立了数百万美元的业务。”

——Business 2.0

“O'Reilly Conference 是聚集关键思想领袖的绝对典范。”

——CRN

“一本 O'Reilly 的书就代表一个有用、有前途、需要学习的主题。”

——Irish Times

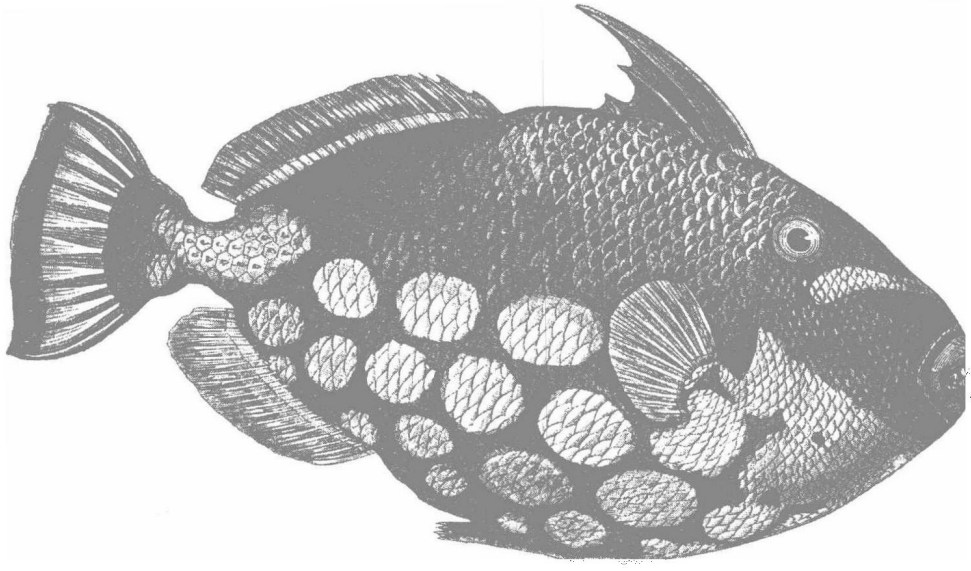
“Tim 是位特立独行的商人，他不光放眼于最长远、最广阔的视野并且切实地按照 Yogi Berra 的建议去做了：‘如果你在路上遇到岔路口，走小路（岔路）。’回顾过去，Tim 似乎每一次都选择了小路，而且有几次都是一闪即逝的机会，尽管大路也不错。”

——Linux Journal

MongoDB扩展技术

Scaling MongoDB

[美] Kristina Chodorow 著
巨 成 译



前言

本书是为那些对分片感兴趣的 MongoDB 用户准备的，它全面讲述了如何建立和使用集群等内容。

本书不是对 MongoDB 的入门介绍。读者需要理解诸如文档、集合、数据库这些概念，知道如何读写数据，什么是索引，如何以及为什么要建立副本集。

如果你并不熟悉 MongoDB，大可不必担心，因为它很容易上手。市面上有一些有关 MongoDB 的书，包括本书作者参与编写的《MongoDB 权威指南》。你也可以查阅在线文档。

格式约定

本书使用了如下排版约定。

- 楷体

用于标记新名词。

- 等宽字体

用于程序代码，在段落中用于表示程序的组成部分，如变量或函数名、数据库、数据类型、环境变量、语句、关键字。

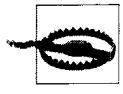
- 等宽粗体

命令或是其他应该由用户输入的内容。

- 等宽斜体

应该由用户提供的或根据上下文确定的值。

这个图标表示提示、建议或一般性的注解。



这个图标表示一个警告或警示。

使用示例代码

本书用于帮助你完成工作。通常，你可以在程序或文档中使用本书提供的代码。除非你在重新发布我们的大量代码，否则不需要联系我们来获得许可。比如，在程序中使用本书代码的一些片段是无需我们许可的。但是出售或再分发 O'Reilly 的图书示例光盘显然是需要授权的。引用本书或引用示例代码来回答问题是不需要授权的，但将本书的大量示例代码纳入产品的文档是需要授权的。

我们乐于见到你在使用时声明引用信息，但不强制要求。引用信息通常包括书名、作者、出版社和 ISBN，例如“*Scaling MongoDB* by Kristina Chodorow (O'Reilly). Copyright 2011 Kristina Chodorow, 978-1-449-30321-1”。

如果你认为对示例代码的使用需要授权，请通过这个邮箱联系我们：permissions@oreilly.com。

Safari®在线图书

Safari 在线图书是应需而变的数字图书馆。它能够让你非常轻松地搜索 7500 多种技术性和创新性参考书以及视频，以便快速地找到需要的答案。

订阅后你就可以访问在线图书馆内的所有页面和视频，在手机或其他移动设备上阅读，在新书上市之前抢先阅读，还能够看到尚在创作中的书稿并向作者反馈意见。复制粘贴代码示例、放入收藏夹、下载部分章节、标记关键点、做笔记甚至打印页面等有用的功能可以帮你节省大量时间。

这本书也在其中。欲访问本书的英文版电子版，或者由 O'Reilly 或其他出版社出版的相关图书，请到 http://my.safaribooksonline.com 免费注册。

我们的联系方式

请把对本书的评论和问题发给出版社。

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

O'Reilly 的每一本书都有专属网页，你可以在那儿找到本书的相关信息，包括勘误表、示例代码以及其他信息。本书的网站地址是：

http://oreilly.com/catalog/9781449303211

中文书：

http://www.oreilly.com.cn/index.php?func=book&isbn=9787115272119

对于本书的评论和技术性问题，请发送电子邮件到：

bookquestions@oreilly.com

关于本书的更多信息、会议、资源中心和网络，请访问以下网站：

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

我们在 Facebook 的地址如下：

http://facebook.com/oreilly

请关注我们的 Twitter 动态：

http://twitter.com/oreillymedia

我们的 YouTube 视频地址如下：

http://www.youtube.com/oreillymedia

目录

MongoDB 扩展技术

MongoDB 开发技巧 50 例

第 1 章

欢迎来到分布式计算的世界

在《终结者》系列影片中，一个称作“天网”的人工智能生命向人类发动战争，年复一年地制造机器人和杀戮人类。这是大部分运维人员的梦想，当然不是指毁灭人类，而是指构建一个可以长时间运行而无需人工干预的分布式系统。遗憾的是，时至今日“天网”依旧是个幻想，因为设计好并维护其稳定持续运行，对一个分布式系统来说仍然是一件非常困难的事情。

单台数据库服务器的状态通常很简单：非启即停。但是如果再添一台服务器并把数据分开来，则这两台服务器之间会产生某种依赖。假设其中一台停机，对另一台会造成什么影响？你的应用程序能应付其中一台（或两台一起）停机的情况吗？如果两台都在运行但无法通信呢？又或是可以通信，但是速度非常非常慢呢？

随着更多节点被添加到集群里，这类问题会变得越来越多和复杂。如果集群中的一整部分无法与其他部分通信会发生什么？如果一部分机器崩溃了又会如何？如果整个数据中心都出问题了呢？突然之间，即使是创建一个备份也将变得异常困难。怎样为分布在集群中几十台机器上的 TB 级数据建立一致性快照，但又不会冻结正在使用这些数据的应用程序？

如果一台服务器可以满足需求，那就能避免很多问题。但是如果想要存储大量数据或者想以高于单服务器处理能力的频率来访问这些数据，则建立一个集群是不可避免的。MongoDB 的优势之一正是试图帮助你解决上面列出的许多问题。不过这并不像设置单个 *mongod* 实例（这又是什么？）那么简单。本书将向你展示如何一步步建立起一个健壮的集群，以及在这个过程中将遇到的各种挑战。

什么是分片

分片（sharding）是 MongoDB 用来将大型集合分割到不同服务器（或者说一个*集群*）上所采用的方法。尽管分片起源于关系型数据库分区，但它（像 MongoDB 的大部分方面一样）完全是另一回事。

和你可能使用过的任何分区方案相比，MongoDB 的最大区别在于它几乎能自动完成所有事情。只要告诉 MongoDB 要分配数据，它就能自动维护数据在不同服务器之间的均衡。当然，你得告诉 MongoDB 把服务器添加到集群中，不过只要这么做了，MongoDB 同样会确保新加入的服务器分得均等的数据。

分片主要是为了实现 3 个简单的目标。

让集群“不可见”

应用程序只要知道跟它打交道的是一个普通的 *mongod* 实例就够了。

为了实现这一目标，MongoDB 自带了一个叫做 *mongos* 的专有路由进程。*mongos* 坐镇集群大前方，对连上它的任何应用而言就像是一个普通的 *mongod* 服务器。*mongos* 会把请求正确无误地转发到集群中的一个或一组服务器上，接着再把获得的响应拼装起来发回给客户端。这样一来，客户端无需知道与其通信的是一台服务器还是一个集群。

不过由于集群本身的特性使然，也存在一些违背该抽象的特殊情况，这些特殊情况会在第 4 章中提到。

保证集群总是可读写

任何集群都无法保证永远可用（比如出现大范围停电之类的情况），但是在合理的条件下，永远都不应该出现用户无法读写数据的情况。在功能发生明显降级前，集群应当允许尽可能多的节点失效。

MongoDB 通过多种途径来确保最长的正常运行时间。集群的每一部分可以并且应当在其他服务器上（最理想的情况是在其他数据中心）有冗余的进程运行，以便当一个进程 / 机器 / 数据中心坏掉了，其他副本可以立即（自动地）接替坏掉的部分继续工作。

把数据从一台服务器迁移到另一台的过程中也存在着一个非常有趣的难题：在传输过程中如何保证对数据访问的持续性和一致性？我们已经找出了一些非常好的解决方案，不过有些超出本书的讲述范围。总之，MongoDB 采用了一些非常漂亮的技巧。

使集群易于扩展

当系统需要更多的空间或资源时，应当可以添加。MongoDB 支持按需扩充系统容量。有关增加（和移除）容量的更多内容参见第 3 章。

要实现这些目标，一个集群应该易于使用（就像使用单个节点一样）和易于管理（否则添加新的分片就不那么容易了）。MongoDB 能够轻而易举地让应用程序自然茁壮地成长。