

MongoDB Applied Design Patterns



MongoDB 应用设计模式

O'REILLY®
中国电力出版社

Rick Copeland 著
陈新 译

MongoDB应用设计模式

Rick Copeland 著

陈新 译

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo

O'Reilly Media, Inc. 授权中国电力出版社出版

中国电力出版社

图书在版编目 (CIP) 数据

MongoDB应用设计模式/ (美) 科普兰 (Copeland, R.) 著; 陈新译. —北京: 中国电力出版社, 2015.5

书名原文: MongoDB applied design patterns

ISBN 978-7-5123-7371-6

I. ①M… II. ①科… ②陈… III. ①关系数据库系统 IV. ①TP311.138

中国版本图书馆CIP数据核字 (2015) 第050807号

北京市版权局著作权合同登记

图字: 01-2015-0722号

Copyright ©2013 by Richard D. Copeland, Jr.

Simplified Chinese Edition, jointly published by O'Reilly Media, Inc. and China Electric Power Press, 2015.

Authorized translation of the English edition, 2013 O'Reilly Media, Inc., the owner of all rights to publish and sell the same.

All rights reserved including the right of reproduction in whole or in part in any form.

英文原版由O'Reilly Media, Inc. 出版2013。

简体中文版由中国电力出版社2015年5月出版。本书的翻译得到O'Reilly Media, Inc.的授权。此简体中文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。

封面设计/ Karen Montgomery, 张健

出版发行/ 中国电力出版社 (<http://www.cepp.sgcc.com.cn>)

地 址/ 北京市东城区北京站西街19号 (邮政编码100005)

经 销/ 全国新华书店

印 刷/ 北京丰源印刷厂

开 本/ 787毫米×980毫米 16开本 10印张 184千字

版 次/ 2015年5月第一版 2015年5月第一次印刷

印 数/ 0001—3000册

定 价/ 29.00元 (册)

敬告读者

本书封底贴有防伪标签, 刮开涂层可查询真伪

本书如有印装质量问题, 我社发行部负责退换

版权专有 翻印必究

MongoDB应用设计模式

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的会议和峰会集聚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

目录

前言	1
----------	---

第一部分 设计模式

第1章 嵌入还是引用	9
------------------	---

关系型数据模型和标准化	9
-------------------	---

什么是范式?	9
--------------	---

MongoDB: 不论如何, 谁需要标准化呢?	14
-------------------------------	----

为了潜在的高引数关系使用引用	18
----------------------	----

总结	20
----------	----

第2章 多态模式	21
----------------	----

多态模式支持面向对象编程	21
--------------------	----

多态模式使得模式进化成为可能	24
----------------------	----

多态模式支持半结构域数据	26
--------------------	----

结论	27
----------	----

第3章 模仿事务行为	28
------------------	----

一致性的有关方法	28
----------------	----

混合文档	29
------------	----

使用复杂更新	31
--------------	----

使用补偿来优化更新	32
-----------------	----

小结	35
----------	----

第二部分 应用实例

第4章 运营智能	39
存储日志数据	39
预聚合报告	53
分层聚合	63
第5章 电子商务	73
产品目录	73
分类层级	82
存货管理	88
第6章 内容管理系统.....	98
元数据和资源管理.....	98
存储评论信息	107
第7章 在线广告网络.....	117
解决方案概述	117
设计一：基本广告服务	117
设计二：增加广告频次控制	120
设计三：关键字目标选择	122
第8章 社交网络	125
解决方案概述	125
模式设计	126
操作	129
分片	134
第9章 在线游戏	136
解决方案概述	136
模式设计	137

操作	139
分片	145
后记	147

前言

无论是构建当前很流行、很热门的社交媒体网站，还是开发一个仅在内部使用的企业商业智能应用程序，确保数据模型具有扩展性都是十分重要的。尽管大多数人对传统的关系型数据库比较熟悉，但使用这类数据库来应对“大数据”的高扩展性需求时，将面临极大的挑战和复杂性。在使用MongoDB（处于领导地位的NoSQL数据库）后，既能够很好地应对扩展性挑战，又能够简化开发流程。

尽管现在很多地方都大肆宣传大数据，很多网站已经将它们的商业系统运行在NoSQL数据库上，但如果没有很好地理解所选择的数据库的技术需求，将导致无法高效利用这些NoSQL数据库。本书可提供用户需要的MongoDB特性和解决商业应用上的问题。本书重点关注MongoDB具体实现中遇到的实践性问题，特别适合那些希望利用MongoDB来应对特定扩展性问题的开发者阅读。

读者对象

本书适用于那些对学习采用MongoDB应用模式来解决问题和设计应用程序感兴趣的读者。尽管本书对大部分MongoDB特性有一个基本描述，但本书并不是一本关于MongoDB的入门书籍。建议初学者可以阅读Kristina Chodorow 和Michael Dirolf编写的《MongoDB: The Definitive Guide》（O'Reilly出版），以及Niall O'Higgins编写的《MongoDB and Python》（O'Reilly出版）来了解Python的具体情况。

本书约定

本书中使用的大部分代码范例要么是Python语言，要么是JavaScript语言，因此在阅读本

书时需要对这两种语言有一个基本的了解。此外，很多范例和模式是用来和使用关系型数据库相比较，因此对SQL和关系数据库模型有一个基本了解是十分必要的。

本书内容

本书主要分为两部分，第一部分关注通用的MongoDB设计模式，第二部分将这些模式应用到特定问题领域中。

第一部分 设计模式

第一部分给读者介绍了一些可用于MongoDB的通用的设计模式。第一部分中各章节的内容相比第二部分来说，有更多的介绍性的资料，并重点关注MongoDB的通用技术，而不是具体的领域问题。这里描述的技术倾向于使用MongoDB的特性来解决问题，并且将告诉你哪些事情是MongoDB不能做的。

第1章 嵌入还是引用

本章描述了MongoDB中可以存储哪些类型的文档，并举例说明了在关联文档中嵌入关联文档的模式和通过引用文档ID来访问文档的模式这两种模式的开销。本章重点关注嵌入所带来的性能提升，以及如何权衡嵌入所带来的复杂性和性能提升。

第2章 多态模式

本章首先举例说明MongoDB的容器是无模式的，模式实际上是存储在单独的文档中。然后举例说明如何将该特性与嵌入文档相结合，使得MongoDB可以具备更灵活和高效的多态性。

第3章 模仿事务行为

本章主要是对MongoDB在复杂性和多文档事务方面的不足进行解释。本章示例说明了通过将MongoDB修改器和文档嵌入相结合，就可以在单个原子文档更新操作中完成SQL语言多个不同更新操作所完成的任务。本章还探索了一个实现应用程序级别的两阶段提交协议的模式，通过该协议可以在确实需要的时候为MongoDB提供事务型的保障。

第二部分 应用实例

在第二部分，本书将转为介绍应用设计模式部分，展示一些使用实例和MongoDB应用程序模式来解决具体领域问题。这里的每章覆盖一个特定的领域问题，介绍了用于描述该问题的技术和模式。

第4章 运营智能

本章描述如何使用MongoDB来获得运营智能和对商业数据进行“实时分析”。本章描述了一个简单的事件日志记录系统，通过使用定期增量的分层聚合来扩展该系统。然后以一个真实的实时增量聚合系统的描述作为本章的结束。这个系统叫做Mongo监视服务（Mongo Monitoring Service, MMS），该系统使用的技术和协议可以使用相对较少的硬件来获得对成百上千个用户的巨量数据进行高性能处理的能力。

第5章 电子商务

本章首先描述了如何使用MongoDB来构建一个产品主目录，主要关注MongoDB的多态性模式技术和分层存储方法。然后描述了一个使用优化的更新和压缩技术的库存管理系统，该系统甚至不使用两阶段提交协议就可实现最终的一致性。

第6章 内容管理系统

本章描述了如何使用MongoDB作为内容管理系统的后端。本章特别关注于使用多态性模式来存储内容节点，使用GridFS和二进制字段来存储二进制资源，并有多条途径来存储用户的讨论信息。

第7章 在线广告网络

本章描述了一个在线广告网络的设计。这里重点关注嵌入文档和复杂原子更新，同时确认存储引擎（MongoDB）不会成为广告服务的瓶颈。本章将涉及频繁覆盖广告印象、关键词定位、关键词竞价等技术。

第8章 社交网络

本章描述如何使用MongoDB来存储相关复杂的社交图谱，通过对“Google+”这个产品进行建模，使用了各种各样的圈子，提供对哪些内容能共享给哪些用户的很好的粒度控制。这里主要关注维护社交图谱，同时将根据时间轴和新闻订阅将内容进行分类。

第9章 在线游戏

本章描述如何使用MongoDB来存储在线多角色扮演游戏所必需的数据。这里将展示如何使用MongoDB存储角色和场景数据，需要考虑到多个玩家对同一个数据结构的同时并发访问。

本书使用的排版约定

本书使用如下排版约定：

斜体(*italic*)

用来表示新术语、URL、email地址、文件名、文件扩展名等。

等宽字体 (*constant width*)

用来表示程序列表，同时在段落中引用的程序元素（例如变量、函数名、数据库、数据类型、环境变量、声明和关键字等）也用该格式表示。

等宽黑体 (*constant width bold*)

用于表示需要用户逐字符输入的命令或其他文本。

等宽斜体 (*constant width italic*)

用于表示应该以用户提供的值或根据上下文决定的值加以替换的文本。

注意： 该图标是一个技巧、建议或者一般性注释。

警告： 该图标是一个警告或者一个提醒性注释。

使用示例代码

本书可以帮助你完成你的工作。一般来说，可以在自己的程序和文档中使用本书中的代码。除非你将重新编译代码中的重要部分，你是不需要联系我们来获得授权的。举例来说，使用本书中几段程序来编程是不需要获得我们的授权的，但是销售或发布O'Reilly出版书籍中配套光盘中代码是需要授权的。通过引用本书中范例来回答问题是不需要授权的，而将本书中重要部分的范例代码整合到您产品的文档中是需要授权的。

我们感谢您在使用我们代码的时候给出引用说明，但这不是硬性规定。一个引用说明通常包括了标题、作者、出版社和ISBN。举例来说，本书的引用说明“MongoDB Applied Design Patterns by Rick Copeland (O'Reilly) . Copyright 2013 Richard D. Copeland, Jr., 978-1-449-34004-9.”

如果您不确定所使用的范例代码是否超出了上面给定的权限，可以随时通过电子邮件联系我们。我们的电子邮件地址是permissions@oreilly.com。

Safari Book Online

Safari Book Online是一个按需定制的数字图书馆，它提供来自全球技术和商业上最顶尖的专家的书籍和视频。

技术专业人员、软件开发、网页设计者和商业创新专业人员都可以使用Safari Book Online来作为研究问题、解决、学习和认证培训的主要资源。

Safari Book Online为组织者、政府机构和个人提供了各种价格范围的资源。订阅者可以通过一个统一的可搜索数据库访问成千上万的书籍、培训视频和预先出版的手稿,提供资源的出版社包括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, FTPress, Apress, Manning, New Riders, McGraw-Hill, Jones & Bartlett, Course Technology等。如需更多关于Safari Book Online的信息,请访问我们的网站。

如何联系我们

请将关于本书的意见和问题发送给出版社:

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

我们为本书提供了网页,该网页上面列出了勘误表、范例和任何其他附加的信息。您可以访问如下网页获得:

<http://oreil.ly/mongodb-applied-designpatterns>

要询问技术问题或对本书提出建议,请发送电子邮件至:

bookquestions@oreilly.com

要获得更多关于我们的书籍、会议、资源中心和O'Reilly网络的信息,请参见我们的网站:

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

<http://www.oreilly.com>

我们的Facebook地址: *<http://facebook.com/oreilly>*

我们的Twitter地址: *<http://twitter.com/oreillymedia>*

也可以在YouTube查看我们的信息，地址是：<http://www.youtube.com/oreillymedia>

致谢

非常感谢O'Reilly出版社的Meghan Blanchette，他忍受了要求让一个技术上的家伙按照工作进度表编写书籍所带来的沮丧。真诚地感谢我的技术编辑Jesse Davis和Mike Dirolf，他们帮助我发现错误，以避免困扰读者。

更多的感谢需要给10gen，他们是MongoDB的发明者，这些优秀的人才不仅提供了一个伟大的技术产品，这些年来也都成为了亲密的朋友。特别的感谢需要给Jared Rosoff，他提出应用实例和设计模式的思路，激发了创作本书的灵感。同样的感谢要给Meghan Gill，是他帮我联系O'Reilly出版社并让这件事情最终落地，同时还为我参加各种MongoDB会议提供经济支持。

感谢我的孩子Matthew和Anna，他们忍受了无法享受在家里一起玩耍的乐趣，只是偶尔能够通过Skype来一个拥抱。

最后，由衷地感谢我最爱、最美丽的妻子Nancy，感谢她这么多年来支持和信赖，鼓舞我完成了很多一个人无法完成的事情。如果没有你，我肯定无法完成本书。

设计模式

嵌入还是引用

在构建一个新的应用程序时，通常第一件事情就是设计数据模型。例如在MySQL关系型数据库中，设计数据模型这个步骤是在标准化过程中进行规范化，重点关注从一组表格中移除冗余。和关系型数据有所不同，MongoDB使用结构化的文档来存储数据，而不是存储在关系型数据库中固定的表格内。举例来说，典型的关系型表格需要每一个行列交汇处有一个标量的值。MongoDB的BSON文档通过支持值数组来支持更为复杂的结构（值数组中每一个数组本身可以由多个子文档组成）。

本章探索MongoDB中富文档模型中提供的一个选择：这个问题是应该嵌入相关的对象，还是通过ID来引用对象。通过本章的学习，将学到如何通过权衡性能、灵活性和复杂性来做出上述决定。

关系型数据模型和标准化

在开始探讨MongoDB解决嵌入文档和链接文档途径之前，我们先花一些时间来看看在关系型数据库中如何确定模型。在关系型数据库中，典型的数据建模方式是使用一系列表来建立数据模型，这些表由行和列组成，这些行和列共同定义了数据的模式。关系型数据库理论被定义为若干将应用数据放入表的方法，称为范式。尽管详细讨论关系型建模已经超出了本书的范围，这里还是要介绍两个特别有意思的范式：第一范式和第三范式。

什么是范式？

典型的模式标准化是从将应用数据放入第一范式（1NF）开始的。关于1NF确切定义是