

- 全面讲解MongoDB的相关知识，使读者对MongoDB有全面的认识
- 以最直接、最细致的方式指导读者轻松掌握MongoDB的安装、部署与使用
- 以实际工作框架为例子进行讲解，使读者真正能胜任MongoDB的开发管理工作
- 由浅入深，层层递进，路线清晰



基于分布式文件存储的数据库

MongoDB游记

之轻松入门到进阶

张泽泉 著

清华大学出版社



内容简介



MongoDB 作为一款非关系型数据库，因其灵活、易用、可扩展性强而受到广泛欢迎。本书全面介绍了 MongoDB 的架构、安装、配置、使用及高级应用。全书共分 10 章，第 1 章介绍 MongoDB 的概述；第 2 章介绍 MongoDB 的安装与配置；第 3 章介绍 MongoDB 的基本操作；第 4 章介绍 MongoDB 的查询操作；第 5 章介绍 MongoDB 的索引操作；第 6 章介绍 MongoDB 的复制与分片；第 7 章介绍 MongoDB 的安全与认证；第 8 章介绍 MongoDB 的备份与恢复；第 9 章介绍 MongoDB 的监控与运维；第 10 章介绍 MongoDB 的高级应用。本书适合数据库开发人员、运维人员及系统管理员阅读。



MongoDB 游记

之轻松入门到进阶

张泽泉 著

张泽泉，男，毕业于清华大学，现任某互联网公司技术经理，负责数据库架构设计与运维工作。

清华大学出版社 计算机系列
地址：北京清华大学学研大厦A座
邮编：100084
电话：(010) 62770175
网址：<http://www.tup.tsinghua.edu.cn>

ISBN 7-302-41000-0
定价：49.00元

清华大学出版社
北京

内 容 简 介

MongoDB 作为最受欢迎的文档存储类型的 NoSQL 数据库,越来越多的公司在使用它。本书以符合初学者的思维方式,系统全面、层层递进地介绍了 MongoDB 数据库,通过本书的学习,读者能够胜任实际工作环境中 MongoDB 的相关开发管理工作。

本书共分四个部分 23 章,第一部分讲解了 MongoDB 的相关概念和原理以及其内部工作机制,可以让读者对 MongoDB 有一个全面的认识。第二部分和第三部分从应用角度,结合实例讲解了 MongoDB 的安装、配置、部署、开发、集群部署和管理等在实际工作中会用到的技能。第四部分是经验部分,这部分是作者多年使用 MongoDB 后总结的技巧,对读者在工作中使用 MongoDB 有极大的参考价值。

本书适合 MongoDB 的初学者,希望深入了解 MongoDB 安装部署、开发优化的软件工程师,希望深入了解 MongoDB 管理、集群扩展的数据运维管理员,以及任何对 MongoDB 相关技术感兴趣的读者。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

MongoDB 游记之轻松入门到进阶 / 张泽泉著. — 北京:清华大学出版社,2017
(数据库技术丛书)

ISBN 978-7-302-47860-7

I. ①M… II. ①张… 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 181049 号

责任编辑:夏毓彦

封面设计:王 翔

责任校对:闫秀华

责任印制:沈 露

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社 总 机:010-62770175

邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者:北京泽宇印刷有限公司

经 销:全国新华书店

开 本:190mm×260mm

印 张:19

字 数:486 千字

版 次:2017 年 9 月第 1 版

印 次:2017 年 9 月第 1 次印刷

印 数:1~3500

定 价:59.00 元

产品编号:070704-01

前言

我思考了很长时间,到底要写一本什么样的书,才能让读者轻松、全面地认识 MongoDB。

从 2012 年有幸开始接触 MongoDB 并在实际工作环境中使用它,不知不觉已经过了 5 年多的时间。在这 5 年中,大数据兴起, NoSQL 来势汹汹。

“有 MongoDB 使用经验优先”“精通 MongoDB 等 NoSQL 数据库”这样的要求也渐渐出现在招聘要求中。MongoDB 作为 NoSQL 数据库的典型代表,越来越多的公司在使用它。

在开始学习使用 MongoDB 的过程中,因为相关书籍资料太少,一路走来确实算是翻山越岭,跋山涉水。这也是本书名的由来。

本书定位

关于本书的定位,在我的想象中应该有如下几点。

1. 这不是一本严肃的教科书

在轻松的氛围中快速学习知识才能达到比较好的效果,所以我会书中尽可能多地加入图画以帮助读者加深理解。

2. 它能让读者从零开始学习数据库

笔者阅读了很多有关 MongoDB 的书籍,我发现大多数 MongoDB 的书籍在讲解时都喜欢拿关系型数据库进行类比学习,这样会对一个数据库的初学者(不了解关系型数据库的读者)造成困扰,所以我会尽量少用关系型数据库的用法来解释 MongoDB,让没有关系型数据库基础的读者也能独立地理解学习 MongoDB,而对有关系型数据库基础的读者能抛开关系型数据库的用法学习 MongoDB。

3. 这是一本入门的书,但不代表它不全面

请原谅我没有在本书中进行深入的原理讲解,因为我一直认为对于学习一个技术来说踏出

第一步尤其重要。一旦你开始使用它，随着解决以后遇到的问题，你自然而然地就会更深入地理解它。

当然，为了让初学者对 MongoDB 有个全面的了解，我会尽可能地把它写得全面。

4. 它能够帮助读者建立 NoSQL 的思维

随着本书从理论到实战的讲解，读者能初步了解并且适应 NoSQL 的思维，或者说脱离关系型数据库的束缚。在之后的实际开发应用中读者会感受到 NoSQL 数据库的魅力——自由(对之前经常使用关系型数据库的读者而言这一点感受会尤其深刻)。

5. 阅读完本书后读者能够胜任实际工作环境中 MongoDB 的相关开发管理工作

这是本书唯一且最终的目的。只要跟着书中的思路一步一步走，读者就能达到相关工作的要求。

6. 一本好书应该包含作者的思想，一些不会过时的东西

当我开始写这本书的时候，MongoDB 的版本是 3.4，当这本书写完发行的时候，我不能确保软件版本还是 3.4，这也是计算机方面的书籍总是容易过时的原因。考虑到这一点，我在写本书时会讲解一些我获取更新知识的途径、我的学习思路以及分享一些有趣的网站，以确保当本书中的例子因为版本原因有些不再适用时，读者能很快地获取到最新的软件用法在哪里。希望这些方法不仅仅为读者在学习 MongoDB 上，也能在学习其他计算机技术时有所帮助。

本书特点

- 内容全面：对 MongoDB 相关知识进行全面讲解，让读者对 MongoDB 有全面的认识。
- 轻松入门：以最直接、最细致的方式指导读者轻松掌握 MongoDB 安装、部署与使用。
- 层次清晰：理论与实践部分分离，4 个部分由浅入深，层层递进，学习路线清晰。
- 实战引导：以实际工作框架为例子，讲解 MongoDB 的运用，让读者真正能胜任 MongoDB 的开发管理工作。

本书适合读者

本书适合 MongoDB 的初学者，希望深入了解 MongoDB 安装部署、开发优化的软件工程师，希望深入了解 MongoDB 管理、集群扩展的数据运维管理员，以及任何对 MongoDB 相关技术感兴趣的读者。

注意事项

本书中所用示例以及操作步骤全部经过作者实际操作，检验有效，请读者在学习时按照本

书选用的软件版本进行操作,已经在使用 MongoDB 软件的开发者在查阅本书示例时,请注意 MongoDB 版本和驱动版本以及环境配置。

本书使用方式

初学者

希望初学者能够通读本书,理论部分快速阅读,实践部分动手操作。

初学者如果时间充裕,建议按本书章节先后顺序进行学习,先通读理论,如有不懂的地方可以不必深究,然后实战,接着进阶架构,最后学习经验篇。

初学者如果时间紧迫,可以从实战部分开始学习,直接上手使用 MongoDB 参与到工作中,闲暇时再学习理论会有更深刻的理解,接着进阶架构,最后再学习经验篇。

有相关经验的技术人员

有相关经验的技术人员可以根据自身需求和薄弱环节选择性阅读。闲暇时可阅读理论部分和经验部分,本书理论部分和经验部分是作者结合自身工作经验对 MongoDB 的理解,希望能与有经验读者的看法相互印证碰撞,让读者形成自己的理解。

工作时也可参考本书相关章节,基本命令章节、Java 驱动操作章节、Spring Data MongoDB 驱动操作章节,管理维护命令、集群构建章节等,可以作为读者身边的速查手册。

祝大家开卷有益!

本书示例代码

本书示例代码以及软件开发包下载地址(注意数字与字母大小写)如下:

<http://pan.baidu.com/s/1boKG28R>

如果下载有问题,请联系 booksaga@163.com,邮件主题为“MongoDB 代码”。

致谢

感谢 23 魔方公司和大厂开发组,让我在 MongoDB 实战中积累了宝贵的经验。

玩网络游戏,最害怕的是游戏公司不再对游戏进行维护,玩家对游戏失去信心,游戏区变成鬼区。10gen 公司在 2013 年正式更名为 MongoDB Inc,感谢 MongoDB Inc 公司技术人员不断地完善 MongoDB,让我们能够用上更快、更方便、更安全的 MongoDB。没有程序是没有 bug 的, MongoDB 在发布初期确实存在很多问题,感谢那些能够正视 MongoDB 并伴随它成长的人,让一款经典的数据库和一种新的数据思维没有被埋没。好在 MongoDB 已经长成了一颗参天大树,在数据库领域占据了一席之地。

本书在写作期间,参考了大量的 MongoDB 相关书籍,以及 MongoDB 官网、MongoDB 中文社区、CSDN 博客等网站的大量文章,感谢这些使用 MongoDB 并乐于分享的大神们。作

者已经尽力对书中的知识点用词用语做了各种查证,但由于时间仓促,难免有错误和遗漏之处,恳请广大读者提出宝贵意见。

奋斗在一线的程序员,加班加点是常有的事情。编写本书使用了作者一年的周末和假期时间以及很多个 10 小时以后的美好时光。没有家人的鼓励和支持是很难坚持下来的。在此特别感谢我的家人刘悦梦露的陪伴和谅解,以及好朋友杨娅苏、古曙强、姚思伶对我的鼓励。

最后特别感谢清华大学出版社的夏毓彦老师和他的同事们,正因为有了他们的辛勤工作,这本书才能顺利出版。

张泽泉

2017 年 7 月于成都

目 录

第一部分 基础与架构理论篇

第 1 章 初识 MongoDB	3
1.1 MongoDB 简介	3
1.1.1 MongoDB 是什么	3
1.1.2 MongoDB 的历史	3
1.1.3 MongoDB 的发展情况	4
1.1.4 哪些公司在用 MongoDB	5
1.2 MongoDB 的特点	5
1.3 MongoDB 应用场景	6
1.3.1 MongoDB 适用于以下场景	6
1.3.2 MongoDB 不适合的场景	7
第 2 章 MongoDB 的结构	8
2.1 数据库	8
2.1.1 数据库的层次	8
2.1.2 数据的命名	8
2.1.3 自带数据库	9
2.2 普通集合	9
2.2.1 集合是什么	9
2.2.2 集合的特点——无模式	9
2.2.3 集合命名	9
2.2.4 子集合	10
2.3 固定集合 (Capped)	10
2.3.1 Capped 简介	10
2.3.2 Capped 属性特点	10
2.3.3 Capped 应用场景	10
2.4 文档	11
2.4.1 文档简介	11
2.4.2 文档的特点	11
2.4.3 文档的键名命名规则	11
2.5 数据类型	11
2.5.1 基本数据类型	11
2.5.2 数字类型说明	12

2.5.3	日期类型说明	14
2.5.4	数组类型说明	16
2.5.5	内嵌文档类型说明	16
2.5.6	_id 键和 ObjectId 对象说明	17
2.5.7	二进制类型说明——小文件存储	19
2.6	索引简介	19
2.6.1	什么是索引	19
2.6.2	索引的作用	20
2.6.3	普通索引	20
2.6.4	唯一索引	20
2.6.5	地理空间索引	21
第 3 章	MongoDB 的大文件存储规范 GridFs	22
3.1	GridFS 简介	22
3.2	GridFS 原理	23
3.3	GridFS 应用场景	24
3.4	GridFS 的局限性	24
第 4 章	MongoDB 的分布式运算模型 MapReduce	25
4.1	MapReduce 简介	25
4.2	MapReduce 原理	26
4.3	MapReduce 应用场景	28
第 5 章	MongoDB 存储原理	29
5.1	存取工作流程	29
5.2	存储引擎	30
5.2.1	MMAP 引擎	31
5.2.2	MMAPv1 引擎	31
5.2.3	WiredTiger 引擎	32
5.2.4	In-Memory	33
5.2.5	引擎的选择	34
5.2.6	未来的引擎	34
第 6 章	了解 MongoDB 复制集	35
6.1	复制集简介	35
6.1.1	主从复制和副本集	35
6.1.2	副本集的特点	38
6.2	副本集工作原理	38
6.2.1	oplog (操作日志)	38
6.2.2	数据同步	39
6.2.3	复制状态和本地数据库	39
6.2.4	阻塞复制	40
6.2.5	心跳机制	40
6.2.6	选举机制	41

6.2.7 数据回滚.....	42
第 7 章 了解 MongoDB 分片	43
7.1 分片的简介.....	43
7.2 分片的工作原理.....	44
7.2.1 数据分流.....	44
7.2.2 chunkSize 和块的拆分	47
7.2.3 平衡器和块的迁移.....	47
7.3 分片的应用场景.....	48
第二部分 管理与开发入门篇	
第 8 章 安装 MongoDB	51
8.1 版本和平台的选择.....	51
8.1.1 版本的选择.....	51
8.1.2 平台的选择.....	52
8.1.3 32 位和 64 位.....	52
8.2 Windows 系统安装 MongoDB.....	53
8.2.1 查看安装环境.....	53
8.2.2 安装步骤.....	53
8.2.3 目录文件了解.....	55
8.3 Linux 系统安装 MongoDB	56
8.3.1 虚拟机简介.....	56
8.3.2 虚拟机安装以及安装 Linux 系统	58
8.3.3 安装 MongoDB.....	67
8.4 Mac OSX 系统安装 MongoDB.....	73
8.4.1 查看安装环境.....	73
8.4.2 官网安装包安装.....	73
8.4.3 Mac 软件仓库安装.....	74
第 9 章 启动和停止 MongoDB	75
9.1 命令行方式启动和参数.....	75
9.1.1 Windows 系统命令行启动 MongoDB.....	75
9.1.2 Linux 系统命令行启动 MongoDB	76
9.1.3 Mac OS 系统命令行启动 MongoDB.....	79
9.2 启动参数.....	80
9.3 配置文件方式启动.....	82
9.4 启动 MongoDB 客户端.....	84
9.5 关闭 MongoDB.....	84
9.5.1 Windows 系统设置 MongoDB 关闭.....	84
9.5.2 Linux 系统设置 MongoDB 关闭	86
9.5.3 Mac OS 系统设置 MongoDB 关闭.....	87
9.6 设置 MongoDB 开机启动.....	88

- 9.6.1 Windows 系统设置 MongoDB 开机启动..... 88
 - 9.6.2 Linux 系统设置 MongoDB 开机启动 89
 - 9.6.3 Mac OS 系统设置 MongoDB 开机启动..... 93
- 9.7 修复未正常关闭的 MongoDB..... 96
- 第 10 章 基本命令 97
 - 10.1 数据库常用命令..... 97
 - 10.2 集合..... 99
 - 10.3 文档..... 101
 - 10.4 索引..... 104
 - 10.5 基本查询..... 106
 - 10.5.1 find 简介 106
 - 10.5.2 游标..... 107
 - 10.6 条件查询..... 108
 - 10.6.1 与操作..... 108
 - 10.6.2 或操作\$or 108
 - 10.6.3 大于\$gt..... 108
 - 10.6.4 小于\$lt 108
 - 10.6.5 大于等于\$gte..... 108
 - 10.6.6 小于等于\$lte..... 108
 - 10.6.7 类型查询\$type..... 108
 - 10.6.8 是否存在\$exists..... 109
 - 10.6.9 取模\$mod..... 109
 - 10.6.10 不等于\$ne..... 109
 - 10.6.11 包含\$in..... 110
 - 10.6.12 不包含\$nin..... 110
 - 10.6.13 \$not: 反匹配..... 110
 - 10.7 特定类型查询..... 110
 - 10.7.1 null 110
 - 10.7.2 正则查询（模糊查询） 110
 - 10.7.3 嵌套文档..... 112
 - 10.7.4 数组..... 112
 - 10.8 高级查询\$where..... 115
 - 10.8.1 JavaScript 语言简介 115
 - 10.8.2 JavaScript 编程简单例子 115
 - 10.8.3 JavaScript 与\$where 结合使用 115
 - 10.9 查询辅助..... 116
 - 10.9.1 条数限制 limit 116
 - 10.9.2 起始位置 skip 116
 - 10.9.3 排序 sort..... 116
 - 10.10 修改器..... 116
 - 10.10.1 \$set 116
 - 10.10.2 \$unset 117
 - 10.10.3 \$inc..... 117

10.10.4	\$push	117
10.10.5	\$pushAll	117
10.10.6	\$pull	117
10.10.7	\$addToSet	118
10.10.8	\$pop	118
10.10.9	\$rename	118
10.10.10	\$bit	118
10.11	原生聚合运算	119
10.11.1	数量查询 count	119
10.11.2	不同值 distinct	119
10.11.3	分组 group	120
10.11.4	灵活统计 MapReduce	123
10.12	聚合管道	127
10.12.1	aggregate 用法	127
10.12.2	管道操作器	128
10.12.3	管道表达式	139
10.12.4	复合使用示例	141
第 11 章	GUI 工具：数据库外部管理工具	144
11.1	MongoDB 的 GUI 工具简介	144
11.2	Robomongo 基本操作	144
11.2.1	连接 MongoDB	145
11.2.2	创建删除数据库	145
11.2.3	插入文档	145
11.2.4	查询文档	146
11.2.5	更新文档	146
11.2.6	创建索引	147
11.2.7	执行 JavaScript	148
第 12 章	监控	149
12.1	原生管理接口监控	149
12.2	使用 serverStatus 在 Shell 监控	150
12.3	使用 mongostat 在 Shell 监控	151
12.4	使用第三方插件监控	152
第 13 章	安全和访问控制	153
13.1	绑定监听 ip	153
13.2	设置监听端口	154
13.3	用户认证	154
13.3.1	启用认证	154
13.3.2	添加用户	155
13.3.3	用户权限控制	155
13.3.4	用户登录	157
13.3.5	修改密码	157

13.3.6 删除用户.....	157
第 14 章 数据管理.....	158
14.1 数据备份 mongodump.....	158
14.2 数据恢复 mongorestore.....	159
14.3 数据导出 mongoexport	159
14.3.1 导出 JSON 格式	159
14.3.2 导出 CSV 格式	159
14.4 数据导入 mongoimport.....	160
14.4.1 JSON 格式导入	160
14.4.2 CSV 格式导入	160
第 15 章 MongoDB 驱动	161
15.1 MongoDB 驱动支持的开发语言	161
15.2 驱动使用流程.....	163
第 16 章 Java 操作 MongoDB	165
16.1 安装 JDK	165
16.2 Eclipse 安装	166
16.3 加载驱动.....	167
16.4 查阅 Java 操作语法.....	167
16.5 测试操作.....	168
16.5.1 连接数据库.....	168
16.5.2 插入数据	169
16.5.3 查询数据	170
16.5.4 更新数据	170
16.5.5 删除数据	171
16.5.6 聚合方法执行	171
16.5.7 操作 GridFS	172
16.5.8 运行示例	173

第三部分 管理与开发进阶篇

第 17 章 副本集部署	177
17.1 总体思路.....	177
17.2 MongoDB 环境准备.....	178
17.3 创建目录.....	181
17.4 创建 Key.....	182
17.5 初始化副本集.....	183
17.6 数据同步测试.....	190
17.7 故障切换测试.....	192
17.8 Java 程序连接 MongoDB 副本集测试.....	194
17.9 主从复制部署.....	196

第 18 章 分片部署	198
18.1 总体思路	198
18.2 创建 3 个 Shard Server	201
18.2.1 创建目录	201
18.2.2 以分片 Shard Server 模式启动	201
18.3 启动 Config Server	202
18.3.1 创建目录	202
18.3.2 以分片 Config Server 模式启动	202
18.4 启动 Route Process	203
18.5 配置 sharding	204
18.6 对数据库 mytest 启用分片	205
18.7 集合启用分片	206
18.8 分片集群插入数据测试	208
18.9 分片的管理	209
18.9.1 移除 Shard Server, 回收数据	209
18.9.2 新增 Shard Server	211
第 19 章 分片+副本集部署	212
19.1 总体思路	212
19.2 创建 3 个复制集	215
19.2.1 创建目录	215
19.2.2 以复制集模式启动	215
19.2.3 初始化复制集	216
19.3 创建分片需要的 Config Server 与 Route Process	217
19.3.1 创建目录	217
19.3.2 启动 Config Server、Route Process	218
19.4 配置分片	219
第 20 章 springMVC+maven+MongoDB 框架搭建	221
20.1 SpringMVC 和 Maven 简介	221
20.2 Eclipse 安装 Maven 插件	221
20.3 新建 Maven 类型的 Web 项目	222
20.4 搭建 SpringMVC+MongoDB 框架	224
20.4.1 jar 包引入	224
20.4.2 新建 SpringMVC 配置文件	228
20.4.3 新建 MongoDB 配置文件	230
20.4.4 配置 web.xml	231
20.4.5 创建 index.jsp 和 IndexController	232
20.4.6 启动 Web 项目	233
第 21 章 注册登录功能的实现	235
21.1 UI 框架 Bootstrap	235
21.1.1 简介	235

21.1.2	应用 Bootstrap	235
21.2	新建用户实体.....	236
21.3	注册功能编写.....	237
21.3.1	注册页面代码.....	237
21.3.2	注册后端代码.....	239
21.4	登录功能编写.....	241
21.4.1	登录页面代码.....	241
21.4.2	登录后端代码.....	243
21.5	运行测试.....	244
21.6	Sping Data MongoDB 操作.....	246
21.6.1	插入数据.....	247
21.6.2	查询数据.....	247
21.6.3	更新数据.....	249
21.6.4	删除数据.....	250
21.6.5	聚合方法执行.....	250
21.6.6	操作 GridFS.....	251
21.6.7	运行示例.....	253

第四部分 管理与开发经验篇

第 22 章	MongoDB 开发的经验	257
22.1	尽量选取稳定新版本 64 位的 MongoDB.....	257
22.2	数据结构的设计.....	257
22.3	查询的技巧.....	259
22.4	安全写入数据.....	262
22.5	索引设置的技巧.....	264
22.6	不要用 GridFS 处理小的二进制文件.....	268
22.7	优化器 profiler.....	269
第 23 章	MongoDB 管理的经验	271
23.1	MongoDB 安全管理.....	271
23.2	不要将 MongoDB 与其他服务部署到同一台机器上.....	273
23.3	单机开启日志 Journal，多机器使用副本集.....	274
23.4	生产环境不要信任 repair 恢复的数据.....	275
23.5	副本集管理.....	276
23.6	副本集回滚丢失的数据.....	278
23.7	分片的管理.....	279
23.8	MongoDB 锁.....	280
附录 A	MongoDB 地理位置距离单位	285
附录 B	相关网址.....	287

第一部分

基础与架构理论篇

