

内容全面、讲解生动，帮助读者更快、更深入地掌握 Redis 技能
丰富的实战经验，帮助读者更容易地通过技术面试，进入心仪企业

Broadview®
www.broadview.com.cn

全彩

Redis深度历险

核心原理与应用实践

钱文品/著



掌阅科技技术总监陈超 大型多人在线游戏架构师王东永
罗辑思维首席架构师方圆 蚂蚁金服高级技术专家席华峰

联合力荐



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

Redis深度历险

核心原理与应用实践

钱文品/著

電子工業出版社
Publishing House of Electronics Industry
北京•BEIJING

内 容 简 介

Redis 是互联网技术架构在存储系统中使用得最为广泛的中间件，也是中高级后端工程师技术面试中面试官最喜欢问的工程技能之一，特别是那些优秀的互联网公司，通常要求面试者不仅仅掌握 Redis 基础用法，还要理解 Redis 内部实现的细节原理。本书作者老钱在使用 Redis 上积累了丰富的实战经验，希望帮助更多后端开发者更快、更深入地掌握 Redis 技能。

本书分为基础和应用篇、原理篇、集群篇、拓展篇、源码篇共 5 大块内容。基础和应用篇讲解对读者来说最有价值的内容，可以直接应用到实际工作中；原理篇、集群篇让开发者透过简单的技术表面看到精致的底层世界；拓展篇帮助读者拓展技术视野和夯实基础，便于进阶学习；源码篇让高阶的读者能够读懂源码，掌握核心技术实力。

本书适合以下人群阅读：有 Redis 基础，渴望深度掌握 Redis 技术原理的中高级后端开发者；渴望成功进入大型互联网企业研发部的中高级后端开发者；需要支撑公司 Redis 中间件运维工作的初中级运维工程师；对 Redis 中间件技术好奇的中高级前端技术研究者。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有，侵权必究。

图书在版编目（CIP）数据

Redis 深度历险：核心原理与应用实践 / 钱文品著. —北京 : 电子工业出版社, 2019.1

ISBN 978-7-121-35047-4

I. ①R… II. ①钱… III. ①数据库—基本知识 IV. ①TP311.138

中国版本图书馆CIP数据核字（2018）第214091号

责任编辑：林瑞和

印 刷：天津千鹤文化传播有限公司

装 订：天津千鹤文化传播有限公司

出版发行：电子工业出版社

北京市海淀区万寿路173信箱

邮编：100036

开 本：720×1000 1/16 印张：15.5

字数：276.7千字

版 次：2019年1月第1版

印 次：2019年1月第1次印刷

定 价：79.00元



凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010) 88254888, 88258888

质量投诉请发邮件至zlt@phei.com.cn，盗版侵权举报请发邮件至dbqq@phei.com.cn。

本书咨询联系方式：010-51260888-819 faq@phei.com.cn。

前　　言

攀登技术之山

为什么我要尝试写作技术类书籍

我毕业至今已经十年了。这十年的技术生涯犹如艰辛的登山过程，中间虽有停停歇歇，但整体而言，我始终在向上努力攀登。

我是个对新技术有着强烈好奇心的人，曾经学习了很多种计算机语言，有些语言与我的工作并没有太大关系，但这不妨碍我花费时间去钻研它们。相比身边很多技术高手，我本人并不算一个特别有天赋的人，所以爬山的过程比较缓慢。

2018年年中，我偶然回顾了一下自己的技术生涯，感觉总算有所小成，登山达到了一定高度，但与此同时，我也意识到技术日新月异，顶峰遥不可及，总会有我爬不动的那一天，那么在此之前我能做些什么呢？

闲暇之时，我开始尝试写作技术类书籍，希望将自己多年来的所学所想记录下来，分享给山下的学弟学妹们，希望他们阅读之后，可以在登山时轻松一些。等到他们未来达到我所处的高度时，也能偶尔记起我这样一个前辈曾经写过一点东西对他们有过些许帮助。

我必须承认，我的语文水平不算好，写作对我来说是一个挑战。不过当我开始着手尝试时，却发现自己有一种停不下来的感觉。

我发现写作技术类书籍这件事特别适合我，一方面这类书并不需要华丽的辞藻以及别出心裁的情节设计，因为写出简明易懂的内容才是最重要的，另一方面我很清楚普通人在面对一门新技术时所遇到的难点在哪里，门槛在哪里，因为登山时遇到的艰难我都心中有数。

技术大神们可能会觉得那些“难点”都特别简单，他们很难站在普通人的角度思考问题，对于读者的抱怨会觉得难以理解。我时常翻阅一些国外的技术博客，发

现这些大神写的文章其实并不易懂，一篇文章往往要仔细地阅读好多遍才能大致理解。如果读者希望更轻松地理解他们所写的内容，就太需要我们这些愿意写作技术类书籍的人。我们将来自山顶的晦涩的知识抽丝剥茧，让它们变得易于理解，让更多人可以享受到来自山顶的阳光。

人们常说，一个人年轻时经历的艰难会在未来成为他的财富，我想这大概就是我能完成这本书的原因。

为什么我要写 Redis

Redis 是互联网技术架构在存储系统中使用得最为广泛的中间件，它也是中高级后端工程师技术面试中面试官最喜欢问的工程技能之一，特别是那些优秀的、竞争激烈的大型互联网公司（比如 Twitter、新浪微博、阿里云、腾讯云、淘宝、知乎等），通常要求面试者不仅要掌握 Redis 基础使用方法，更要深层理解 Redis 内部实现的细节原理。毫不夸张地说，只要能把 Redis 的知识点全部吃透，你的半只脚就已经踏进心仪公司的技术研发部了。

但我在以往的很多面试中，发现大多数同学只会拿 Redis 做数据缓存，使用最简单的 get/set 方法，除此之外几乎一无所知。也有小部分同学知道 Redis 的分布式锁，但也不清楚其内部实现机制，甚至在使用上就不标准，导致生产环境中出现意想不到的问题。还有很多同学没认识到 Redis 是个单线程结构，也不理解单线程的 Redis 为何还可以支持高并发。

我希望通过梳理和总结自己的实践经验，能够帮助更多后端开发者更快、更深入地掌握 Redis 技能。这就是我写作本书的初衷。

我所在的掌阅科技公司，为了支撑海量（亿级）的用户服务，使用了上千个 Redis 实例，如图 0-1 所示，包含大约 100 个 Redis 集群（Codis）以及很多独立的 Redis 节点，因此我在使用 Redis 作为缓存和持久存储中间件上积累了较为丰富的实战经验，这些我都将毫无保留地分享到本书中。

Redis 涉及的知识点是非常多的，本书将讲解其中最常见的 Redis 核心原理和应用实践经验，让读者在阅读之后可以将知识快速应用到平时的 Redis 项目开发中。除此之外，本书还会深入探究一些底层的至关重要的计算机科学基础原理，以及技术应用的思考方式，这些基础的知识和技能将最终决定你的技术人生道路可以走多快、走多远。

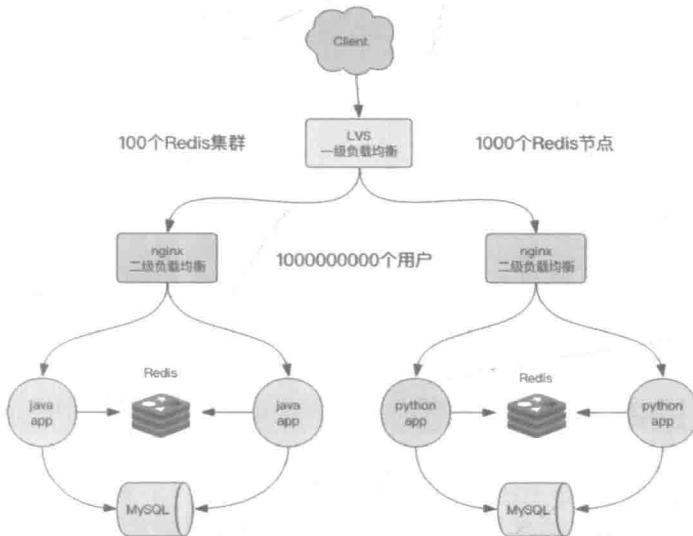


图 0-1

本书内容结构

本书分为基础和应用篇、原理篇、集群篇、拓展篇、源码篇共 5 大块内容，如图 0-2 所示。

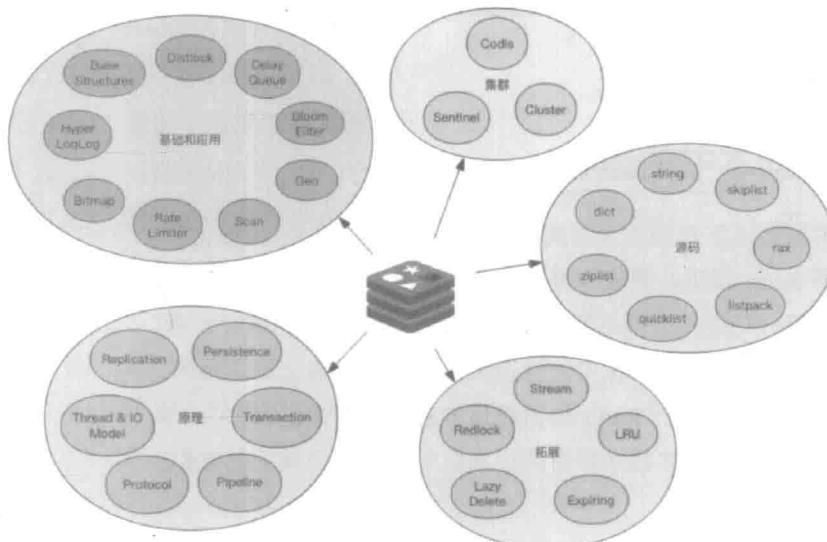


图 0-2

- 基础和应用篇：占据篇幅最长，这也是对读者来说最有价值的内容，可以直接应用到实际工作中。
- 原理篇和集群篇：适合对技术有着极致追求的开发者学习，他们希望透过简单的技术表面看到精致的底层世界。
- 拓展篇：作为最核心内容的补充部分，帮助读者进一步拓展技术视野或者夯实基础，便于进阶学习。
- 源码篇：满足高阶用户深入探索 Redis 内部实现原理的强烈需要，这类读者坚信读懂源码才是技术实力的真正体现。

图文并茂是本书一大特色

为了便于读者理解本书内容，我花费不少时间绘制了大量原创彩色插图，例子如图 0-3 所示。希望这些彩图能够帮助读者更有效率地理解本书知识点，实现事半功倍的效果。

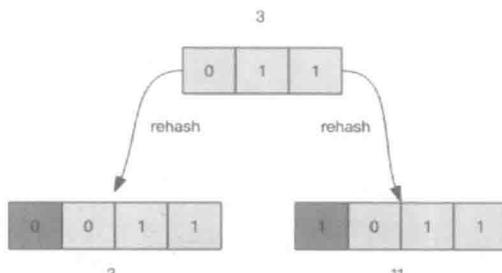


图 0-3

适合阅读本书的读者

本书适合以下类型的读者阅读。

1. 有 Redis 基础，渴望深度掌握 Redis 技术原理的中高级后端开发者。
2. 渴望成功进入大型互联网企业研发部的中高级后端开发者。
3. 需要支撑公司 Redis 中间件运维工作的初中级运维工程师。
4. 希望更好地设计 Redis 面试题目的后端技术面试官。
5. 对 Redis 中间件技术好奇的中高级前端技术领域的朋友们。

老钱

2018 年 10 月

目 录

第 1 篇 基础和应用篇	/ 1
1.1 授人以鱼不如授人以渔	/ 1
1.1.1 由 Redis 面试想到的	/ 1
1.1.2 本书的内容范围	/ 2
1.1.3 Redis 可以做什么	/ 3
1.1.4 小结	/ 3
1.1.5 扩展阅读	/ 4
1.2 万丈高楼平地起——Redis 基础数据结构	/ 4
1.2.1 Redis 的安装	/ 5
1.2.2 5 种基础数据结构	/ 6
1.2.3 容器型数据结构的通用规则	/ 17
1.2.4 过期时间	/ 17
1.2.5 思考 & 作业	/ 17
1.3 千帆竞发——分布式锁	/ 18
1.3.1 分布式锁的奥义	/ 18
1.3.2 超时问题	/ 20
1.3.3 可重入性	/ 21
1.3.4 思考 & 作业	/ 24
1.4 缓兵之计——延时队列	/ 24
1.4.1 异步消息队列	/ 24

1.4.2 队列空了怎么办	/ 26
1.4.3 阻塞读	/ 26
1.4.4 空闲连接自动断开	/ 26
1.4.5 锁冲突处理	/ 27
1.4.6 延时队列的实现	/ 27
1.4.7 进一步优化	/ 30
1.4.8 思考&作业	/ 31
1.5 节衣缩食——位图	/ 31
1.5.1 基本用法	/ 31
1.5.2 统计和查找	/ 34
1.5.3 魔术指令 bitfield	/ 35
1.5.4 思考&作业	/ 38
1.6 四两拨千斤——HyperLogLog	/ 38
1.6.1 使用方法	/ 39
1.6.2 pfadd 中的 pf 是什么意思	/ 41
1.6.3 pfmerge 适合的场合	/ 42
1.6.4 注意事项	/ 42
1.6.5 HyperLogLog 实现原理	/ 42
1.6.6 pf 的内存占用为什么是 12KB	/ 49
1.6.7 思考&作业	/ 50
1.7 层峦叠嶂——布隆过滤器	/ 50
1.7.1 布隆过滤器是什么	/ 51
1.7.2 Redis 中的布隆过滤器	/ 51
1.7.3 布隆过滤器的基本用法	/ 52
1.7.4 注意事项	/ 59
1.7.5 布隆过滤器的原理	/ 60
1.7.6 空间占用估计	/ 61
1.7.7 实际元素超出时，误判率会怎样变化	/ 62
1.7.8 用不上 Redis 4.0 怎么办	/ 63
1.7.9 布隆过滤器的其他应用	/ 63

1.8 断尾求生——简单限流	/ 64
1.8.1 如何使用 Redis 来实现简单限流策略	/ 64
1.8.2 解决方案	/ 65
1.8.3 小结	/ 67
1.9 一毛不拔——漏斗限流	/ 68
1.9.1 Redis-Cell	/ 71
1.9.2 思考&作业	/ 72
1.9.3 扩展阅读：Redis-Cell 作者介绍	/ 72
1.10 近水楼台——GeoHash	/ 73
1.10.1 用数据库来算附近的人	/ 73
1.10.2 GeoHash 算法	/ 74
1.10.3 Geo 指令的基本用法	/ 75
1.10.4 注意事项	/ 78
1.11 大海捞针——scan	/ 79
1.11.1 scan 基本用法	/ 80
1.11.2 字典的结构	/ 82
1.11.3 scan 遍历顺序	/ 82
1.11.4 字典扩容	/ 83
1.11.5 对比扩容、缩容前后的遍历顺序	/ 84
1.11.6 渐进式 rehash	/ 85
1.11.7 更多的 scan 指令	/ 85
1.11.8 大 key 扫描	/ 85
第 2 篇 原理篇	/ 87
2.1 鞭辟入里——线程 IO 模型	/ 87
2.1.1 非阻塞 IO	/ 87
2.1.2 事件轮询（多路复用）	/ 88
2.1.3 指令队列	/ 90
2.1.4 响应队列	/ 90
2.1.5 定时任务	/ 90

2.1.6 扩展阅读	/ 90
2.2 交头接耳——通信协议	/ 90
2.2.1 RESP	/ 91
2.2.2 客户端→服务器	/ 92
2.2.3 服务器→客户端	/ 92
2.2.4 小结	/ 95
2.2.5 扩展阅读	/ 95
2.3 未雨绸缪——持久化	/ 95
2.3.1 快照原理	/ 96
2.3.2 fork（多进程）	/ 96
2.3.3 AOF 原理	/ 97
2.3.4 AOF 重写	/ 98
2.3.5 fsync	/ 98
2.3.6 运维	/ 98
2.3.7 Redis 4.0 混合持久化	/ 99
2.3.8 思考&作业	/ 100
2.4 雷厉风行——管道	/ 100
2.4.1 Redis 的消息交互	/ 100
2.4.2 管道压力测试	/ 101
2.4.3 深入理解管道本质	/ 102
2.4.4 小结	/ 104
2.5 同舟共济——事务	/ 104
2.5.1 Redis 事务的基本用法	/ 104
2.5.2 原子性	/ 105
2.5.3 discard（丢弃）	/ 106
2.5.4 优化	/ 106
2.5.5 watch	/ 107
2.5.6 注意事项	/ 108
2.5.7 思考&作业	/ 110
2.6 小道消息——PubSub	/ 110

2.6.1 消息多播	/ 110
2.6.2 PubSub	/ 111
2.6.3 模式订阅	/ 113
2.6.4 消息结构	/ 114
2.6.5 PubSub 的缺点	/ 115
2.6.6 补充	/ 115
2.7 开源节流——小对象压缩	/ 115
2.7.1 32bit VS 64bit	/ 116
2.7.2 小对象压缩存储 (ziplist)	/ 116
2.7.3 内存回收机制	/ 120
2.7.4 内存分配算法	/ 120
第 3 篇 集群篇	/ 122
3.1 有备无患——主从同步	/ 122
3.1.1 CAP 原理	/ 122
3.1.2 最终一致	/ 123
3.1.3 主从同步与从从同步	/ 123
3.1.4 增量同步	/ 124
3.1.5 快照同步	/ 124
3.1.6 增加从节点	/ 125
3.1.7 无盘复制	/ 125
3.1.8 wait 指令	/ 125
3.1.9 小结	/ 126
3.2 李代桃僵——Sentinel	/ 126
3.2.1 消息丢失	/ 128
3.2.2 Sentinel 基本用法	/ 128
3.2.3 思考 & 作业	/ 129
3.3 分而治之——Codis	/ 130
3.3.1 Codis 分片原理	/ 131
3.3.2 不同的 Codis 实例之间槽位关系如何同步	/ 132
3.3.3 扩容	/ 132

3.3.4 自动均衡	/ 133
3.3.5 Codis 的代价	/ 133
3.3.6 Codis 的优点	/ 134
3.3.7 mget 指令的操作过程	/ 134
3.3.8 架构变迁	/ 135
3.3.9 Codis 的尴尬	/ 135
3.3.10 Codis 的后台管理	/ 136
3.3.11 思考&作业	/ 136
3.4 众志成城——Cluster	/ 137
3.4.1 槽位定位算法	/ 138
3.4.2 跳转	/ 138
3.4.3 迁移	/ 138
3.4.4 容错	/ 140
3.4.5 网络抖动	/ 140
3.4.6 可能下线（PFAIL）与确定下线（Fail）	/ 141
3.4.7 Cluster 基本用法	/ 141
3.4.8 槽位迁移感知	/ 142
3.4.9 集群变更感知	/ 143
3.4.10 思考&作业	/ 143
第 4 篇 拓展篇	/ 144
4.1 耳听八方——Stream	/ 144
4.1.1 消息 ID	/ 145
4.1.2 消息内容	/ 145
4.1.3 增删改查	/ 145
4.1.4 独立消费	/ 147
4.1.5 创建消费组	/ 148
4.1.6 消费	/ 150
4.1.7 Stream 消息太多怎么办	/ 152
4.1.8 消息如果忘记 ack 会怎样	/ 153

4.1.9 PEL 如何避免消息丢失	/ 153
4.1.10 Stream 的高可用	/ 153
4.1.11 分区 Partition	/ 154
4.1.12 小结	/ 154
4.2 无所不知——Info 指令	/ 154
4.2.1 Redis 每秒执行多少次指令	/ 155
4.2.2 Redis 连接了多少客户端	/ 156
4.2.3 Redis 内存占用多大	/ 156
4.2.4 复制积压缓冲区多大	/ 157
4.2.5 思考&作业	/ 158
4.3 拾遗补漏——再谈分布式锁	/ 158
4.3.1 Redlock 算法	/ 158
4.3.2 Redlock 使用场景	/ 159
4.3.3 扩展阅读：redlock-py 的作者	/ 160
4.4 朝生暮死——过期策略	/ 160
4.4.1 过期的 key 集合	/ 160
4.4.2 定时扫描策略	/ 160
4.4.3 从节点的过期策略	/ 161
4.5 优胜劣汰——LRU	/ 162
4.5.1 LRU 算法	/ 163
4.5.2 近似 LRU 算法	/ 164
4.5.3 思考&作业	/ 165
4.6 平波缓进——懒惰删除	/ 165
4.6.1 Redis 为什么使用懒惰删除	/ 165
4.6.2 flush	/ 166
4.6.3 异步队列	/ 166
4.6.4 AOF Sync 也很慢	/ 166
4.6.5 更多异步删除点	/ 166
4.7 妙手仁心——优雅地使用 Jedis	/ 167
4.7.1 重试	/ 171

4.7.2 思考&作业	/ 172
4.8 居安思危——保护 Redis	/ 172
4.8.1 指令安全	/ 172
4.8.2 端口安全	/ 173
4.8.3 Lua 脚本安全	/ 174
4.8.4 SSL 代理	/ 174
4.8.5 小结	/ 174
4.9 隔墙有耳——Redis 安全通信	/ 175
4.9.1 spiped 原理	/ 176
4.9.2 spiped 使用入门	/ 176
4.9.3 思考&作业	/ 179
第 5 篇 源码篇	/ 180
5.1 丝分缕析——探索“字符串”内部	/ 180
5.1.1 embstr VS raw	/ 181
5.1.2 扩容策略	/ 184
5.1.3 思考&作业	/ 184
5.2 循序渐进——探索“字典”内部	/ 184
5.2.1 dict 内部结构	/ 184
5.2.2 渐进式 rehash	/ 186
5.2.3 查找过程	/ 187
5.2.4 hash 函数	/ 188
5.2.5 hash 攻击	/ 188
5.2.6 扩容条件	/ 188
5.2.7 缩容条件	/ 189
5.2.8 set 的结构	/ 189
5.2.9 思考&作业	/ 189
5.3 挨肩迭背——探索“压缩列表”内部	/ 190
5.3.1 增加元素	/ 192
5.3.2 级联更新	/ 192

5.3.3 intset 小整数集合	/ 194
5.3.4 思考&作业	/ 195
5.4 风驰电掣——探索“快速列表”内部	/ 195
5.4.1 每个 ziplist 存多少元素	/ 197
5.4.2 压缩深度	/ 198
5.5 凌波微步——探索“跳跃列表”内部	/ 198
5.5.1 基本结构	/ 199
5.5.2 查找过程	/ 199
5.5.3 随机层数	/ 200
5.5.4 插入过程	/ 201
5.5.5 删除过程	/ 202
5.5.6 更新过程	/ 203
5.5.7 如果 score 值都一样呢	/ 203
5.5.8 元素排名是怎么算出来的	/ 203
5.5.9 思考&作业	/ 204
5.5.10 题外话	/ 204
5.6 破旧立新——探索“紧凑列表”内部	/ 205
5.6.1 级联更新	/ 207
5.6.2 取代 ziplist 尚需时日	/ 207
5.6.3 思考&作业	/ 207
5.7 金枝玉叶——探索“基数树”内部	/ 207
5.7.1 应用	/ 208
5.7.2 结构	/ 210
5.7.3 思考&作业	/ 213
5.8 精益求精——LFU VS LRU	/ 213
5.8.1 Redis 对象的热度	/ 213
5.8.2 LRU 模式	/ 213
5.8.3 LFU 模式	/ 214
5.8.4 为什么 Redis 要缓存系统时间戳	/ 217
5.8.5 Redis 为什么在获取 lru_clock 时使用原子操作	/ 217

5.8.6 如何打开 LFU 模式	/ 218
5.8.7 思考&作业	/ 218
5.9 如履薄冰——懒惰删除的巨大牺牲	/ 218
5.9.1 懒惰删除的最初实现不是异步线程	/ 219
5.9.2 异步线程方案其实也相当复杂	/ 219
5.9.3 异步删除的实现	/ 221
5.9.4 队列安全	/ 224
5.9.5 思考&作业	/ 225
5.10 跋山涉水——深入字典遍历	/ 225
5.10.1 一边遍历一边修改	/ 226
5.10.2 重复遍历的难题	/ 227
5.10.3 迭代器的结构	/ 227
5.10.4 迭代过程	/ 229
5.10.5 迭代器的选择	/ 231
5.10.6 思考&作业	/ 232