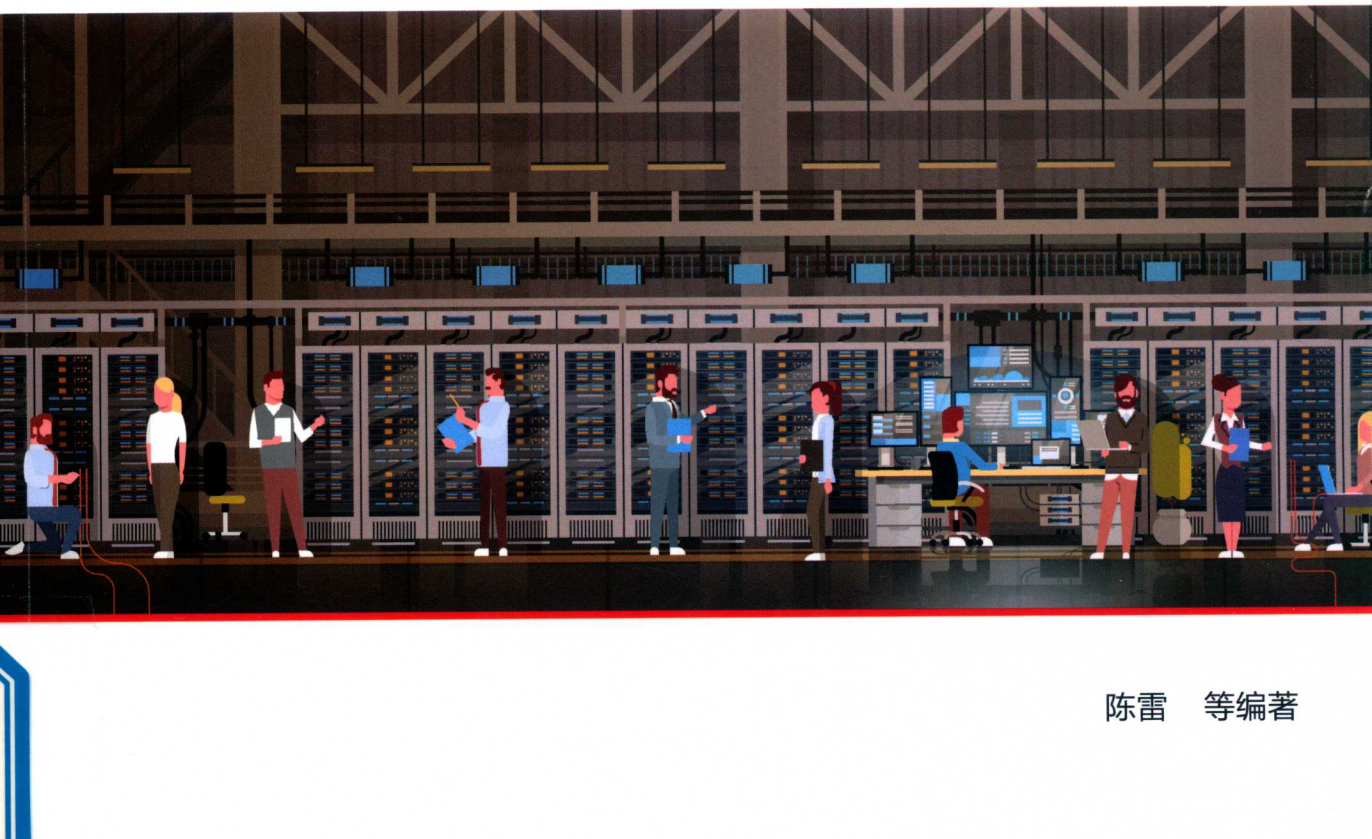
HZ BOOKS
华章IT

多名专家联袂推荐，资深专家联合撰写，深入理解Redis 5设计精髓

系统讲解Redis 5设计、数据结构、底层命令实现，以及持久化、主从复制、集群的实现

数据库技术丛书

Redis 5设计与源码分析



陈雷　等编著

机械工业出版社
China Machine Press

Redis 5设计与源码分析

陈雷 方波 黄桃 李乐 施洪宝 熊浩含 闫昌 张仕华 周生政 编著

机械工业出版社
China Machine Press

图书在版编目（CIP）数据

Redis 5 设计与源码分析 / 陈雷等编著 . —北京：机械工业出版社，2019.8
（数据库技术丛书）

ISBN 978-7-111-63278-8

I. R… II. 陈… III. 关系数据库系统 IV. TP311.132.3

中国版本图书馆 CIP 数据核字（2019）第 151596 号

Redis 5 设计与源码分析

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号 邮政编码：100037）
责任编辑：高婧雅 责任校对：李秋荣
印 刷：三河市宏图印务有限公司 版 次：2019 年 8 月第 1 版第 1 次印刷
开 本：186mm × 240mm 1/16 印 张：27
书 号：ISBN 978-7-111-63278-8 定 价：139.00 元

客服电话：（010）88361066 88379833 68326294 投稿热线：（010）88379604
华章网站：www.hzbook.com 读者信箱：hzit@hzbook.com

華章圖書

一本打开的书，一扇开启的门，
通向科学殿堂的阶梯，托起一流人才的基石。

Praise 本书赞誉

本书从底层源码的角度，对 Redis 的数据结构以及持久化、主从复制、哨兵和集群等特性的实现原理进行了详尽的剖析，图文并茂。行文中也能看出作者团队在源码分析和系统编程方面的功力，我相信本书对于所有想要了解 Redis 及其内部实现的人来说都会有所帮助。

——黄健宏，《Redis 设计与实现》作者

Redis 以其高速、轻量和丰富的数据结构与功能被越来越多的工程师所钟爱。然而，用 Redis 的人很多，真正懂 Redis 的人很少，本书正是写给那些使用了 Redis 并希望进一步深入理解 Redis 的读者。作者及其团队通过对 Redis 最新版本（5.x）各部分源码的分析，庖丁解牛，深入浅出，带领读者一步步探索 Redis 的方方面面，让读者从原理层面真正懂得 Redis。

——黄鹏程，中国民生银行大数据工程师、《Redis 4.X Cookbook》作者

本书全面解析了 Redis 5 内核的方方面面，能够有效帮助 Redis 的开发和运维人员全面理解 Redis 的运行原理，对于需要进阶 Redis 的读者而言是难得的好书。

——付磊，《Redis 开发与运维》作者

对技术有点追求的程序员一定不要错过这本 Redis5 源码分析书，本书对 Redis 的内部实现分析得非常全面透彻，如果你觉得直接阅读源码有点吃力，试试让这本书来带领你探索 Redis 源码。

——钱文品，《Redis 深度历险》作者

本书不仅深入源码讲解了 Redis 常用的底层数据结构和常用命令处理的实际过程，还细致入微地讲述了基数计数算法的演进和 HyperLogLog 算法在 Redis 中的具体实现，这是非常有用且难得的；本书的后几章详细讲述了 Redis 常用的主从复制和持久化的原理，这对于排查问题，以及优化 Redis 集群有极高的参考价值。

——张晋涛，网易有道资深运维开发

Redis 已经是 IT 企业技术栈中重要的一环，与其相关的从业者数量也逐年增多，对大多数人来说 Redis 可谓既熟悉又神秘，只有不足 4MB 的源码却实现了一个功能丰富且健壮的数据库。本书的出版对于想深入了解 Redis 的从业者来说是一个好消息。本书从源码层面对 Redis 进行深入剖析，尤其是数据结构部分，其学习意义不限于 Redis，强烈推荐阅读。

——吴建超，OPPO 工程师

Preface 序

在开源界，高性能服务的典型代表就是 Nginx 和 Redis。纵观这两个软件的源码，都是非常简洁高效的，也都是基于异步网络 I/O 机制的，所以对于要学习高性能服务的程序员或者爱好者来说，研究这两个网络服务的源码是非常有必要的。

Nginx 目前市面上的书籍很多，但是 Redis 确实寥寥无几。这几年 Redis 版本发展非常快，从稳定的 2.x 版本，发展到增加了很多优秀特性的 5.0 版本，这些特性目前尚无资料进行系统讲解。本书的出版填补了 Redis 5.0 技术学习方面的重大空缺，是技术同仁深入理解 Redis 内核实现机制的有效途径。

Redis 是一个优秀的高性能分布式缓存服务器：在实际应用场景中，每秒 QPS 能够达到 4.5 万～5 万，算得上性能“怪兽”；在常规非协程的场景中，Redis 基本是 C10K ⊖高性能服务的经典代表。

除性能优势外，Redis 的整体代码结构也非常清晰，包括基础数据结构、数据类型实现、数据库实现、服务端实现、集群 / 主从 / 队列等，基本模块分布清晰，代码质量非常高：

```
static int aeApiCreate(aeEventLoop *eventLoop) {
    aeApiState *state = zmalloc(sizeof(aeApiState));
    if (!state) return -1;
    state->events = zmalloc(sizeof(struct epoll_event)*eventLoop->setsize);
    if (!state->events) {
        zfree(state);
        return -1;
    }
    state->epfd = epoll_create(1024); /* 1024 is just a hint for the kernel */
    if (state->epfd == -1) {
        zfree(state->events);
        zfree(state);
        return -1;
    }
```

⊖ C10K（concurrent 10000 connection），最早由 Dan Kegel 提出，是指服务器需要支持成千上万客户端的连接所引发的并发支持问题。——编者注

```
    eventLoop->apidata = state;
    return 0;
}
```

上面是创建 epoll 队列的简单代码，简单明了，完全符合《Unix 编程艺术》提到的各种关于简单明确的要求，代码赏心悦目，不花里胡哨。

除了代码结构之外，Redis 的各种类型数据结构也设计良好：简单稳定不容易溢出的字符串结构（sds），快速排序查找的跳跃表（skiplist），节约内存的压缩列表（ziplist），基于 Hash 表实现的字典（dict），基于链表（list）和压缩列表（ziplist）实现的快速列表（qucklist），基于 listpac 和基数树（Rax）实现的消息队列（Stream）等，涵盖多种优质数据结构的实现。

另外不得不提的是，各类算法在 Redis 里也都得到了呈现，比如 Hash 常用算法 times33、物理位置查找算法 geohash、高效率的统计算法 HyperLogLog，等等。读完 Redis 5.0.0 的 9.2 万行源码，大概比上一学期的数据结构课更有价值。Redis 可谓数据结构和常规算法的饕餮盛宴。深入研究 Redis 5，相信对技术的理解会更深入。

优质的菜品需要有技艺精湛的厨师来烹饪，本书就像以优质菜品做成的“大菜”。整本书没有太多啰唆的语言，直接抽丝剥茧：从基本的数据结构类型，Redis 内部每个操作命令的底层代码运行逻辑和结构，一直到整个 Redis 持久化技术、主从技术、分布式集群技术等，都有深入源码级别的讲解，让你领略从数据结构到整个高性能服务的全部设计之美。

学以致用，读者朋友通过领会与实践来提升技术，成为一个高性能网络服务开发高手，继而深入理解缓存服务，设计自己的高性能缓存服务系统或者缓存数据库系统，应用到自己业务中去，岂非快哉！

在整本书里，我也看到了一群程序员的认真执着，把每个业务数据流程图、关键代码、数据结构图都规划得详细、清晰，把自己对技术的各种理解融入书中。本书脉络清晰，适合刚入行的后端程序员、高性能服务开发者、系统运维人员、技术架构师等阅读。希望阅读本书的技术同仁都能够得到进步和提高。

谢华亮（黑夜路人）

2019 年 4 月

Preface 前　言

为什么要写这本书

2 年前，我们团队建立了学习圈，团队成员可以自愿参加，每天 8：50～10：30 到公司充电 100 分钟，深入剖析工作中的技术栈，同时 2017～2018 年编写出版了《 PHP 7 底层设计与源码实现》一书，接着我们又深入研读了 Redis 的源码。2018 年年初开始，我们开始了 Redis 源码一书的编写，起初是研读 Redis 4.0 版本的源码，2018 年下半年 5.0 版本发布，增加了很多的新特性，下半年我们又在之前的基础上结合 Redis 5 的源码，编写了此书。

Redis 是一款高性能的开源 key-value 型数据库，难能可贵的是代码写得非常优雅，非常适合刚入门 C 语言的读者阅读。本书前半部分详细介绍了 Redis 中的各种数据结构，适合读者学习和掌握基本的数据结构；后半部分介绍了 Redis 命令执行的生命周期，以及各类命令的源码实现，希望使用 Redis 的读者不止会使用 Redis，并且能掌握它的原理和细节，提升对 Redis 的掌控能力。

决定编写 Redis 源码一书后，学习圈里方波、黄桃、李乐、施洪宝、熊浩含、闫昌、张仕华、周生政和我一起编写了这本书。大家在工作之外，每天写到深夜，周末一起探讨，经过一年的编写和校对，终于完成了这本书。希望能给使用 Redis 的读者一些启发，帮助更多的人理解 Redis 的实现。

读者对象

- 使用 Redis 的工程师、架构师
- 对 Redis 源码感兴趣的读者
- 有一定 C 语言基础的读者

如何阅读本书

本书内容逻辑上分为三篇，共计 22 章内容。

第一篇：第 1 章简单介绍了 Redis，以及 Redis 的编译安装和研读的方式；第 2～8 章重点讲解了 SDS、跳跃表、压缩列表、字典、整数集合、quicklist 和 Stream 数据结构的实现。

第二篇：第 9 章讲解了 Redis 的生命周期，命令执行的过程，需要重点阅读；第 10～19 章，分别讲解了键、字符串、散列表、链表、集合、有序集合、GEO、HyperLog 和数据流相关命令的实现。

第三篇：第 20～22 章简单讲解了持久化、主从复制和集群的实现，没有详细展开，希望能带读者入门。

如果读者是有一定经验的资深开发人员，本书可能会是一本不错的案头书。当然，如果读者是一名初学者，请在开始本书阅读之前，建议先掌握一些 C 语言和网络编程等基础理论知识。

勘误和支持

由于笔者的水平有限，编写时间仓促，书中难免会出现一些错误或者不准确的地方，恳请读者批评指正。如果您有更多的宝贵意见，欢迎访问 https://segmentfault.com/u/php7internal 进行专题讨论，我们会尽量在线上为读者提供解答。同时，您也可以通过微博 @PHP7 内核，或者邮箱 cltf@163.com 联系到我们，期待能够得到您的反馈，在技术之路上互勉共进。

致谢

感谢张国辉、卢红波两位工作导师的支持，前者是我现在的领导，也是我在技术和管理方面的导师，后者是我在滴滴的领导，在技术和管理上给了我很多的指引与帮助。

感谢黑夜路人（谢华亮）兄弟的指导和支持，在技术上给了非常多的指点。

感谢黄健宏、黄鹏程、付磊、钱文品、张晋涛和吴建超兄弟的指导与建议，他们都是在 Redis 方面有很深研究的人。

感谢方波、黄桃、李乐、施洪宝、熊浩含、闫昌、张仕华和周生政 8 位兄弟在学习和研究过程中的陪伴与合作，本书是几位兄弟共同合作的结晶。特别是黄桃，已经跟我一起编写了两本书。

特别致谢

最后，我要特别感谢我的太太梦云、儿子和女儿，我为写作这本书，牺牲了很多陪伴

她们的时间，但也正因为有了她们的付出与支持，我才能坚持写下去。同时，感谢我的父母、岳父岳母，不遗余力地帮助我们照顾儿女，有了你们的帮助和支持，我才有时间和精力去完成写作工作。

另外要特别感谢我团队的兄弟们，感谢大家的坚持，为大家的成长点赞！重点感谢一下兄弟们背后的太太团，是她们的大力支持，作者们才有时间来编写本书。

最后要重点感谢高婧雅编辑，这是第二次跟她合作，她依然非常负责；她耐心审稿，给出很多宝贵建议，才有了这本书的完成。

谨以此书献给我最亲爱的家人和团队的兄弟们，以及众多热爱 Redis 的朋友们！

陈　雷

目　录 *Contents*

本书赞誉

序

前言

第 1 章　引言 ····· 1

1.1　Redis 简介 ····· 1

1.2　Redis 5.0 的新特性 ····· 2

1.3　Redis 源码概述 ····· 3

1.4　Redis 安装与调试 ····· 4

1.5　本章小结 ····· 6

第 2 章　简单动态字符串 ····· 7

2.1　数据结构 ····· 7

2.2　基本操作 ····· 11

2.2.1　创建字符串 ····· 11

2.2.2　释放字符串 ····· 12

2.2.3　拼接字符串 ····· 12

2.2.4　其余 API ····· 15

2.3　本章小结 ····· 15

第 3 章　跳跃表 ····· 17

3.1　简介 ····· 17

3.2　跳跃表节点与结构 ····· 19

3.2.1　跳跃表节点 ····· 19

3.2.2　跳跃表结构 ····· 20

3.3　基本操作 ····· 20

3.3.1　创建跳跃表 ····· 21

3.3.2　插入节点 ····· 22

3.3.3　删除节点 ····· 28

3.3.4　删除跳跃表 ····· 30

3.4　跳跃表的应用 ····· 31

3.5　本章小结 ····· 32

第 4 章　压缩列表 ····· 33

4.1　压缩列表的存储结构 ····· 33

4.2　结构体 ····· 35

4.3　基本操作 ····· 37

4.3.1　创建压缩列表 ····· 37

4.3.2　插入元素 ····· 38

4.3.3　删除元素 ····· 42

4.3.4　遍历压缩列表 ····· 44

4.4　连锁更新 ····· 44

4.5　本章小结 ····· 45

第 5 章　字典 ····· 47

5.1　基本概念 ····· 47

5.1.1　数组 ····· 48

5.1.2　Hash 函数 ····· 49

5.1.3　Hash 冲突 ····· 51

5.2　Redis 字典的实现 ····· 52

5.3　基本操作 ····· 55

5.3.1　字典初始化 ····· 55

5.3.2　添加元素 ····· 56

5.3.3　查找元素 ····· 60

5.3.4　修改元素 ····· 61

5.3.5　删除元素 ····· 61

5.4　字典的遍历 ····· 62

5.4.1　迭代器遍历 ····· 62

5.4.2　间断遍历 ····· 65

5.5　API 列表 ····· 70

5.6　本章小结 ····· 71

第 6 章　整数集合 ····· 72

6.1　数据存储 ····· 72

6.2　基本操作 ····· 75

6.2.1　查询元素 ····· 75

6.2.2　添加元素 ····· 78

6.2.3　删除元素 ····· 82

6.2.4　常用 API ····· 83

6.3　本章小结 ····· 85

第 7 章　quicklist 的实现 ····· 86

7.1　quicklist 简介 ····· 86

7.2　数据存储 ····· 87

7.3　数据压缩 ····· 91

7.3.1　压缩 ····· 92

7.3.2　解压缩 ····· 93

7.4　基本操作 ····· 94

7.4.1　初始化 ····· 94

7.4.2　添加元素 ····· 95

7.4.3　删除元素 ····· 96

7.4.4　更改元素 ····· 98

7.4.5　查找元素 ····· 99

7.4.6　常用 API ····· 100

7.5　本章小结 ····· 101

第 8 章　Stream ····· 102

8.1　Stream 简介 ····· 102

8.1.1　Stream 底层结构 listpack ····· 103

8.1.2　Stream 底层结构 Rax 简介 ····· 104

8.1.3　Stream 结构 ····· 108

8.2　Stream 底层结构 listpack 的实现 ····· 112

8.2.1　初始化 ····· 112

8.2.2　增删改操作 ····· 112

8.2.3　遍历操作 ····· 113

8.2.4　读取元素 ····· 113

8.3　Stream 底层结构 Rax 的实现 ····· 114

8.3.1　初始化 ····· 114

8.3.2　查找元素 ····· 114

8.3.3　添加元素 ····· 116

8.3.4　删除元素 ····· 118

8.3.5　遍历元素 ····· 120

8.4　Stream 结构的实现 ····· 123

8.4.1　初始化 ····· 124

8.4.2　添加元素 ····· 124

8.4.3　删除元素 ····· 125

8.4.4 查找元素 ····· 128
8.4.5 遍历 ····· 129
8.5 本章小结 ····· 131

第 9 章 命令处理生命周期 ····· 132

9.1 基本知识 ····· 132
9.1.1 对象结构体 robj ····· 132
9.1.2 客户端结构体 client ····· 136
9.1.3 服务端结构体 redisServer ····· 138
9.1.4 命令结构体 redisCommand ····· 139
9.1.5 事件处理 ····· 141
9.2 server 启动过程 ····· 149
9.2.1 server 初始化 ····· 149
9.2.2 启动监听 ····· 152
9.3 命令处理过程 ····· 155
9.3.1 命令解析 ····· 156
9.3.2 命令调用 ····· 159
9.3.3 返回结果 ····· 161
9.4 本章小结 ····· 163

第 10 章 键相关命令的实现 ····· 164

10.1 对象结构体和数据库结构体回顾 ··· 164
10.1.1 对象结构体 redisObject ····· 164
10.1.2 数据库结构体 redisDb ····· 166
10.2 查看键信息 ····· 166
10.2.1 查看键属性 ····· 166
10.2.2 查看键类型 ····· 169
10.2.3 查看键过期时间 ····· 170
10.3 设置键信息 ····· 171
10.3.1 设置键过期时间 ····· 171
10.3.2 删除键过期时间 ····· 172
10.3.3 重命名键 ····· 173
10.3.4 修改键最后访问 ····· 173
10.4 查找键 ····· 174
10.4.1 判断键是否存在 ····· 174
10.4.2 查找符合模式的键 ····· 175
10.4.3 遍历键 ····· 176
10.4.4 随机取键 ····· 177
10.5 操作键 ····· 178
10.5.1 删除键 ····· 178
10.5.2 序列化 / 反序列化键 ····· 182
10.5.3 移动键 ····· 183
10.5.4 键排序 ····· 185
10.6 本章小结 ····· 187

第 11 章 字符串相关命令的实现 ····· 188

11.1 相关命令介绍 ····· 188
11.2 设置字符串 ····· 189
11.2.1 set 命令 ····· 189
11.2.2 mset 命令 ····· 195
11.3 修改字符串 ····· 196
11.3.1 append 命令 ····· 196
11.3.2 setrange 命令 ····· 197
11.3.3 计数器命令 ····· 197
11.4 字符串获取 ····· 199
11.4.1 get 命令 ····· 199
11.4.2 getset 命令 ····· 199
11.4.3 getrange 命令 ····· 199
11.4.4 strlen 命令 ····· 200
11.4.5 mget 命令 ····· 201
11.5 字符串位操作 ····· 201
11.5.1 setbit 命令 ····· 201

11.5.2 getbit 命令……203
11.5.3 bitpos 命令……203
11.5.4 bitcount 命令……205
11.5.5 bitop 命令……208
11.5.6 bitfield 命令……209
11.6 本章小结……212

第 12 章 散列表相关命令的实现……213

12.1 简介……213
12.1.1 底层存储……213
12.1.2 底层存储转换……215
12.1.3 接口说明……215
12.2 设置命令……216
12.3 读取命令……217
12.3.1 hexists 命令……218
12.3.2 hget/hmget 命令……218
12.3.3 hkeys/hvals/hgetall 命令……219
12.3.4 hlen 命令……220
12.3.5 hscan 命令……220
12.4 删除命令……221
12.5 自增命令……222
12.6 本章小结……224

第 13 章 列表相关命令的实现……225

13.1 相关命令介绍……225
13.1.1 命令列表……225
13.1.2 栈和队列命令列表……226
13.2 push/pop 相关命令……228
13.2.1 push 类命令的实现……228
13.2.2 pop 类命令的实现……229
13.2.3 阻塞 push/pop 类命令的实现…230
13.3 获取列表数据……234
13.3.1 获取单个元素……234
13.3.2 获取多个元素……235
13.3.3 获取列表长度……236
13.4 操作列表……236
13.4.1 设置元素……237
13.4.2 插入元素……237
13.4.3 删除元素……238
13.4.4 裁剪列表……239
13.5 本章小结……240

第 14 章 集合相关命令的实现……241

14.1 相关命令介绍……241
14.2 集合运算……254
14.2.1 交集……254
14.2.2 并集……258
14.2.3 差集……260
14.3 本章小结……263

第 15 章 有序集合相关命令的实现…264

15.1 相关命令介绍……264
15.2 基本操作……272
15.2.1 添加成员……272
15.2.2 删除成员……275
15.2.3 基数统计……276
15.2.4 数量计算……277
15.2.5 计数器……279
15.2.6 获取排名……279
15.2.7 获取分值……279
15.2.8 遍历……280
15.3 批量操作……280

15.3.1 范围查找……280
15.3.2 范围删除……283
15.4 集合运算……284
15.5 本章小结……284

第 16 章 GEO 相关命令……285
16.1 基础知识……285
16.2 命令实现……288
16.2.1 使用 geoadd 添加坐标……288
16.2.2 计算坐标的 geohash……291
16.2.3 使用 geopos 查询位置经纬度…292
16.2.4 使用 geodist 计算两点距离…295
16.2.5 使用 georadius/georadius-bymembe 查询范围内元素…295
16.3 本章小结……297

第 17 章 HyperLogLog 相关命令的实现……298
17.1 基本原理……298
17.1.1 算法演进……299
17.1.2 线性计数算法……299
17.1.3 对数计数算法……300
17.1.4 自适应计数算法……302
17.1.5 超对数计数算法……302
17.2 HLL Redis 实现……302
17.2.1 HLL 头对象……303
17.2.2 稀疏编码……304
17.2.3 密集编码……306
17.2.4 内部编码……308
17.2.5 编码转换……309
17.3 命令实现……310
17.3.1 添加基数……310
17.3.2 近似基数……311
17.3.3 合并基数……313
17.4 本章小结……314

第 18 章 数据流相关命令的实现……315
18.1 相关命令介绍……315
18.2 基本操作命令原理分析……323
18.2.1 添加消息……323
18.2.2 删除消息……325
18.2.3 范围查找……326
18.2.4 获取队列信息……327
18.2.5 长度统计……327
18.2.6 剪切消息……328
18.3 分组命令原理分析……328
18.3.1 分组管理……328
18.3.2 消费消息……330
18.3.3 响应消息……331
18.3.4 获取未响应消息列表……331
18.3.5 修改指定未响应消息归属……331
18.4 本章小结……332

第 19 章 其他命令……333
19.1 事务……333
19.1.1 事务简介……333
19.1.2 事务命令实现……334
19.2 发布 – 订阅命令实现……339
19.3 Lua 脚本……345
19.3.1 初始化 Lua 环境……345
19.3.2 在 Lua 中调用 Redis 命令……347
19.3.3 Redis 和 Lua 数据类型转换…349
19.3.4 命令实现……351
19.4 本章小结……356

第 20 章　持久化……357

20.1　RDB……358

20.1.1　RDB 执行流程……358

20.1.2　RDB 文件结构……359

20.2　AOF……367

20.2.1　AOF 执行流程……368

20.2.2　AOF 重写……369

20.3　RDB 与 AOF 相关配置指令……372

20.4　本章小结……374

第 21 章　主从复制……375

21.1　主从复制功能实现……375

21.2　主从复制源码基础……378

21.3　slaver 源码分析……382

21.4　master 源码分析……388

21.5　本章小结……391

第 22 章　哨兵和集群……392

22.1　哨兵……392

22.1.1　哨兵简介……393

22.1.2　代码流程……394

22.1.3　主从切换……396

22.1.4　常用命令……399

22.2　集群……400

22.2.1　集群简介……401

22.2.2　代码流程……402

22.2.3　主从切换……404

22.2.4　副本漂移……406

22.2.5　分片迁移……407

22.2.6　通信数据包类型……409

22.3　本章小结……415