

Redis在新浪的大规模运维经验

@曾经的阿飞

rj03hou@gmail.com

关于我

□DBA@新浪网

□MySQL、Redis、HBase

大纲

□ Redis简介

□ Redis应用

□ 运维经验

“Everything runs from memory in Web 2.0”

Evan Weaver, Twitter, March 2009

Redis简介

- Remote Dictionary Server

- in-memory but persistent database on disk

- Start at 2009 by @antirez

- Open source

存储&cache

□ string、 hash、 list、 set、 sorted set

□ 持久化

□ 高性能

□ 过期时间

持久化

□ rdb

□ aof

rdb

□ Fork子进程 , copy on write

- rdbcompression yes
- dbfilename r7700.rdb
- save <seconds> <changes>

aof

Aof=append only log file

MySQL binlog

```
*<number of arguments> CR LF
$<number of bytes of argument 1> CR LF
<argument data> CR LF
...
$<number of bytes of argument N> CR LF
<argument data> CR LF
```

[illegible]

appendonly yes

```
appendfilename appendonly.aof
```

```
appendfsync [everysec always no]
```

no-appendfsync-on-rewrite yes

auto-aof-rewrite-percentage 0

auto-aof-rewrite-min-size 64mb

bgrewriteaof

□ Append

□ Rewrite

- Fork子进程
- 子进程遍历所有key写入临时文件
- 父进程更新aof写入缓冲区
- 缓冲区追加临时文件
- 替换已有的aof文件

bgrewriteaof

□ auto

- auto-aof-rewrite-percentage 100
 - aof_current_size
 - aof_base_size
- auto-aof-rewrite-min-size 64mb

□ crontab

□ remote集中式

恢复

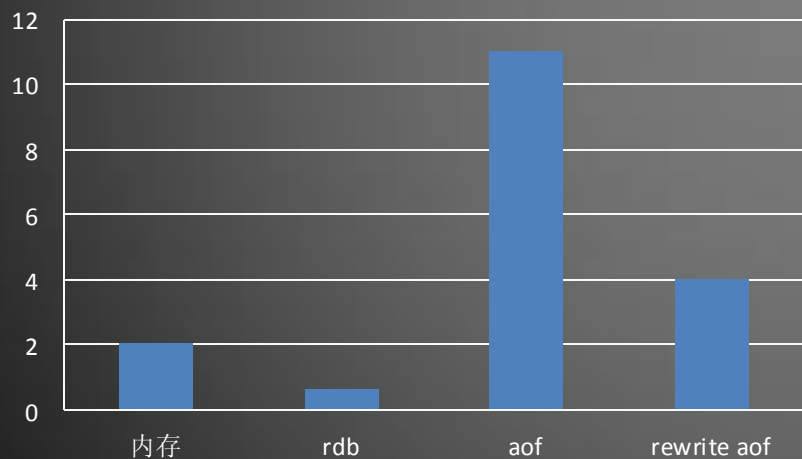
□恢复过程

- 只打开aof，使用aof加载
- 同时打开aof和rdb，使用aof加载
- 打开rdb，使用rdb加载

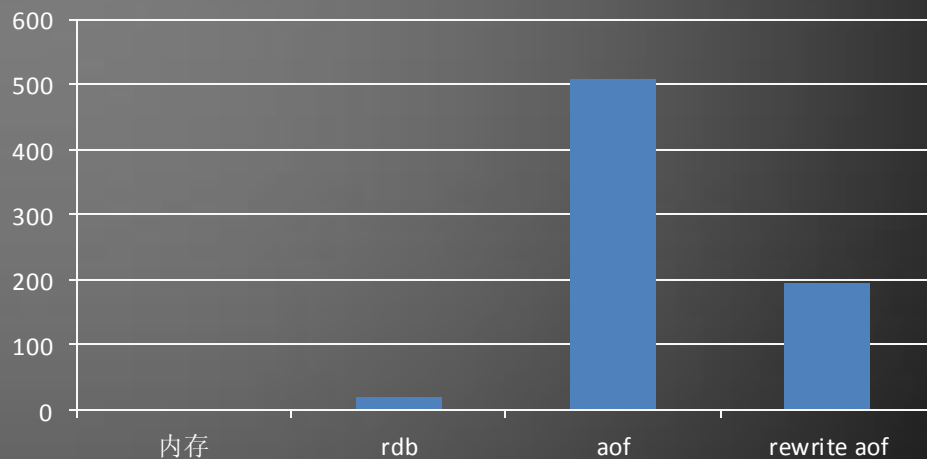
Hash 恢复测试

| | 大小(G) | 恢复时间 | 恢复时间/G |
|-------------|-------|------|--------|
| 内存 | 2.04 | 0 | 0 |
| rdb | 0.65 | 19 | 29.23 |
| aof | 11 | 508 | 46.18 |
| rewrite aof | 4 | 194 | 48.50 |

大小



恢复时间





github

craigslist

guardian.co.uk



大纲

□ Redis简介

□ Redis应用

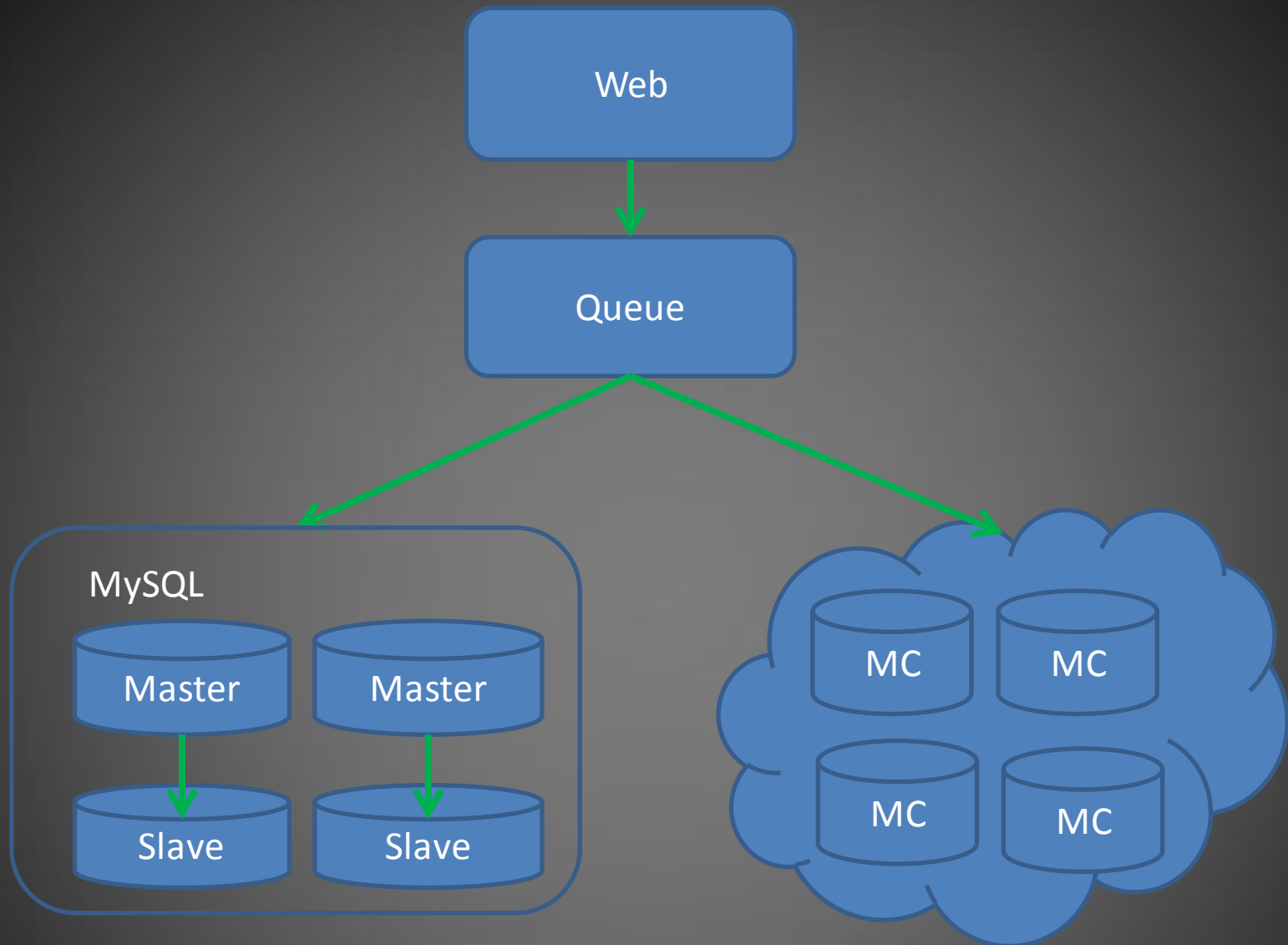
□ 运维经验

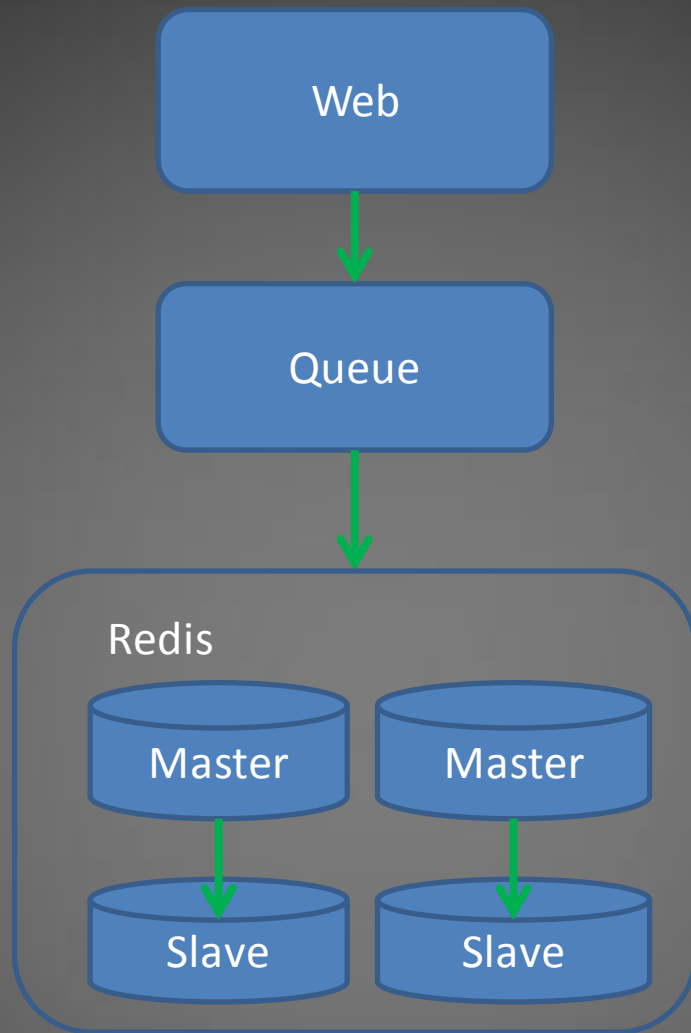
现状

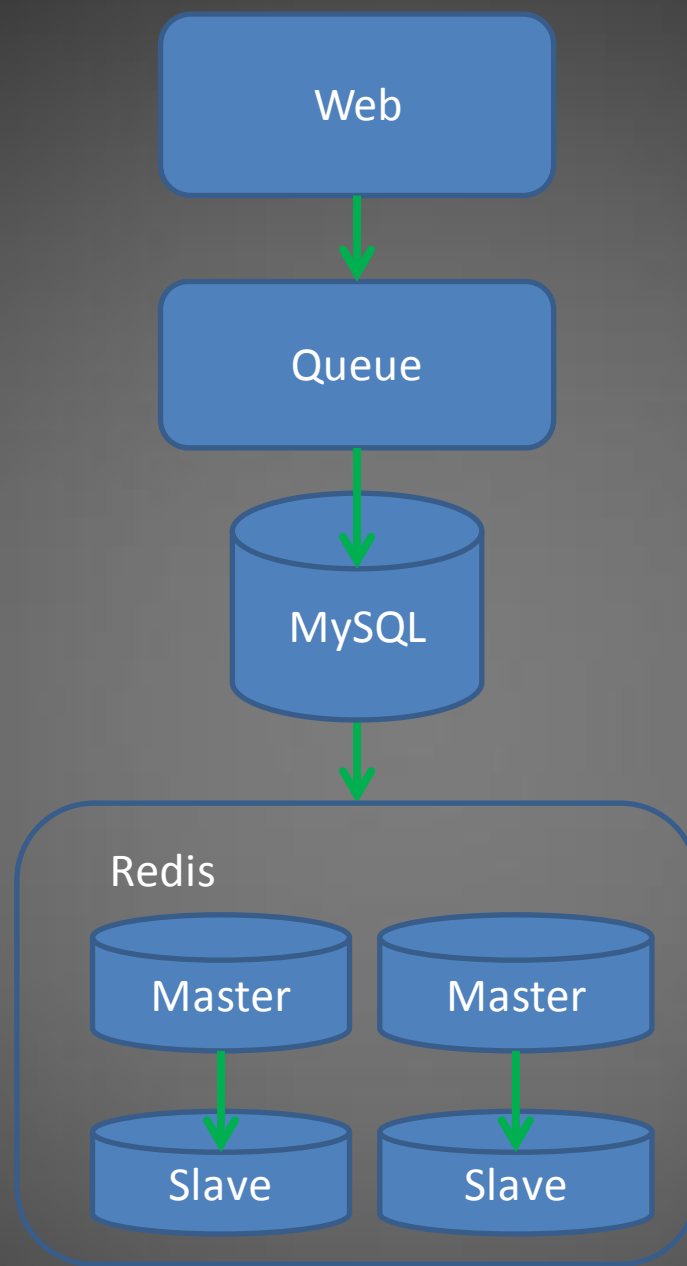
□实例 1500+

□内存总量：15T+

□访问量：2000亿+/天







业务

□String

□Hash

□List

□Set

□Sorted set

粉丝关注

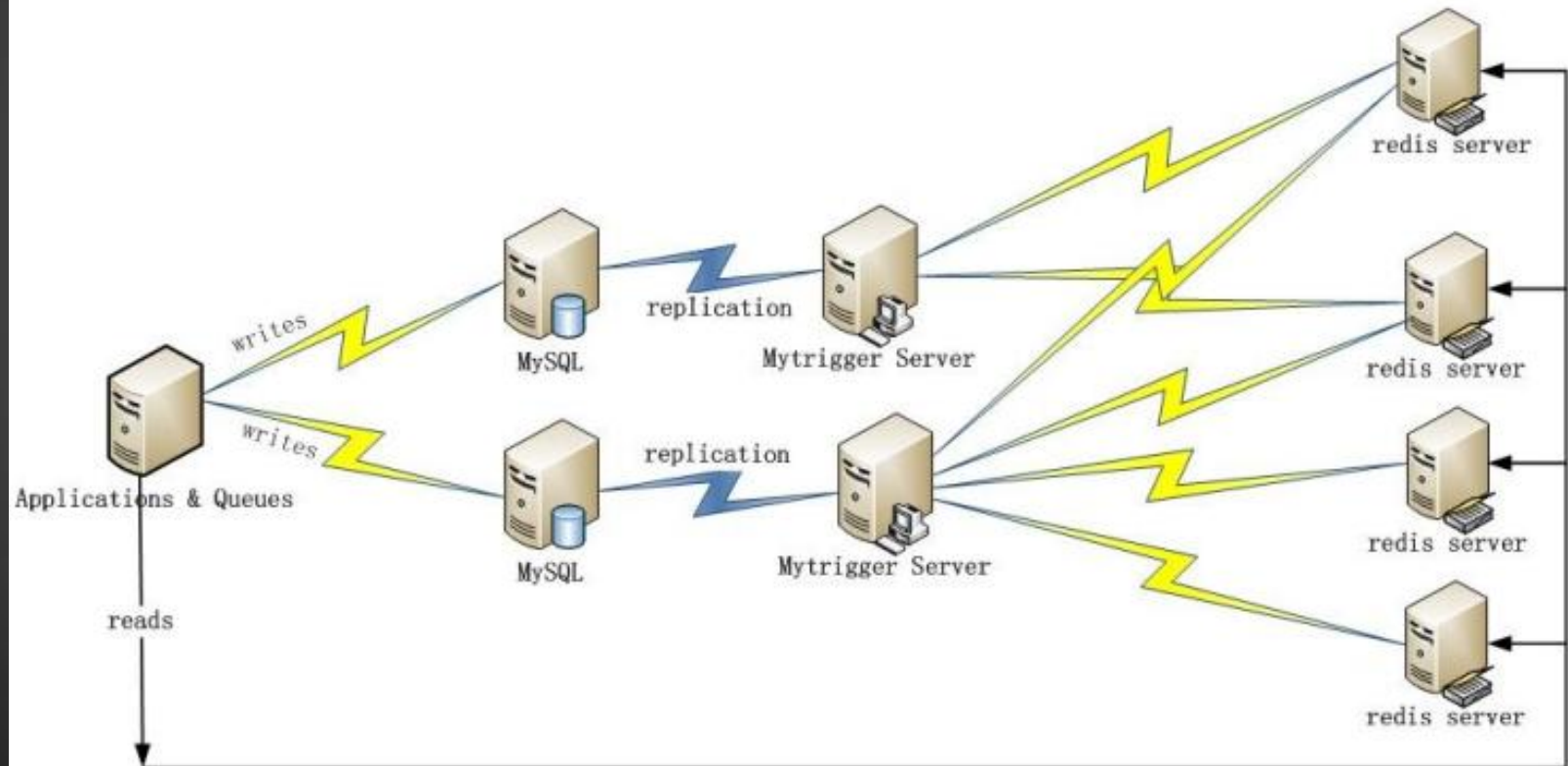
- 粉丝列表

- 关注列表

- 互相关注列表

粉丝关注

MySQL → Mytrigger → Multi Redis



□ Hash

- Key: user_id
- Field: friend ids
- Value: added time

□ 相关操作

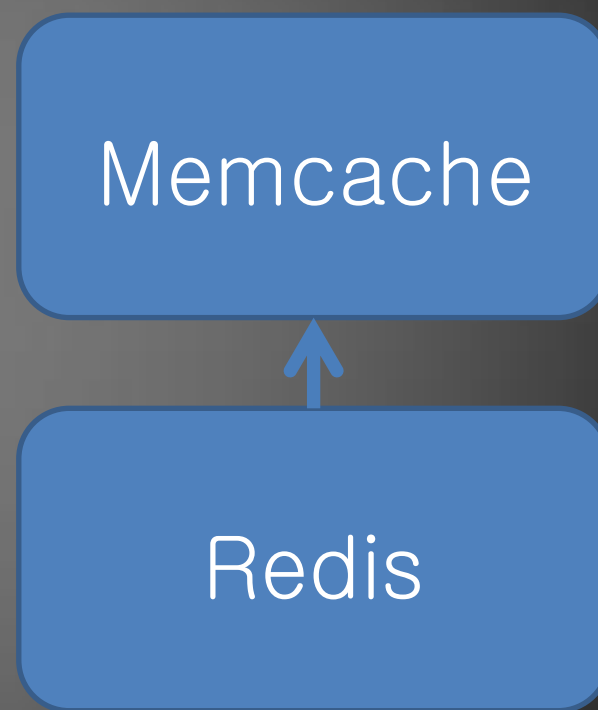
- 加关注 : `hset user_id friend_id added_time`
- 删关注 : `hdel user_id friend_id`
- 获取关注用户的时间 : `hget user_id friend_id`
- 获取关注列表 : `hgetall user_id`

演化

□ Sharding

□ 问题

- 获取关注列表比较慢
- Cpu
- Hgetall成为瓶颈



通知

□ Hash

- Key=uid
- Filed=appkey
- Value=count



通知

□List

系统通知

通知管理 | 已屏蔽应用 ▾

 达人通知

微博达人专属！发微博#达人美丽梦想#+你曾为梦想付出的努力及今后为梦想制定的计划，你就有机会赢取“微博达人美丽梦想秀梦想基金”！此外你还有机会站在2012年12月21日达人狂欢夜北京主会场的舞台上，向数以亿计的微博网友展示自己。还在等什么，再不实现梦想，可就来不及了！活动详情：<http://t.cn/zj56g2t>
推荐身边有梦想的达人参加，还有机会赢取伯乐奖哦~~(11月28日 23:46)

 微吧通知

屏蔽该来源通知 ▾
陈翔橙粉丝群的成员们，明星微吧已开放申请，支持量达到10000“陈翔吧”就能开通啦！目前支持数402，排名第15位。❤️点这里：<http://t.cn/zjqaUhJ> 每天都来“支持一下”！(11月27日 22:37)

 微人脉

接受该来源通知 ▾
收到何大美(猎头,猎头顾问)的添加人脉请求，请处理：<http://t.cn/zj2Rick>(11月27日 10:05)

 微公益新纪元

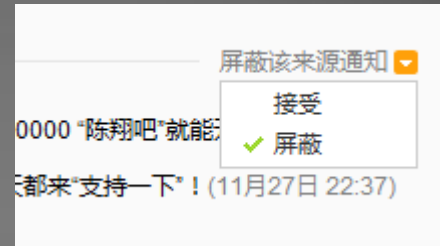
接受该来源通知 ▾
新版微公益上线！所有V用户都可以发起求助啦！尽你所能帮助身边的孤寡老人、伤病孩童亦或是求助流浪动物、拯救地球家园，只有你想不到，没有微公益做不到！用微小的力量，让世界更美！猛戳：<http://t.cn/zjZrIMK>(11月15日 16:22)

查看全部2条 ▾

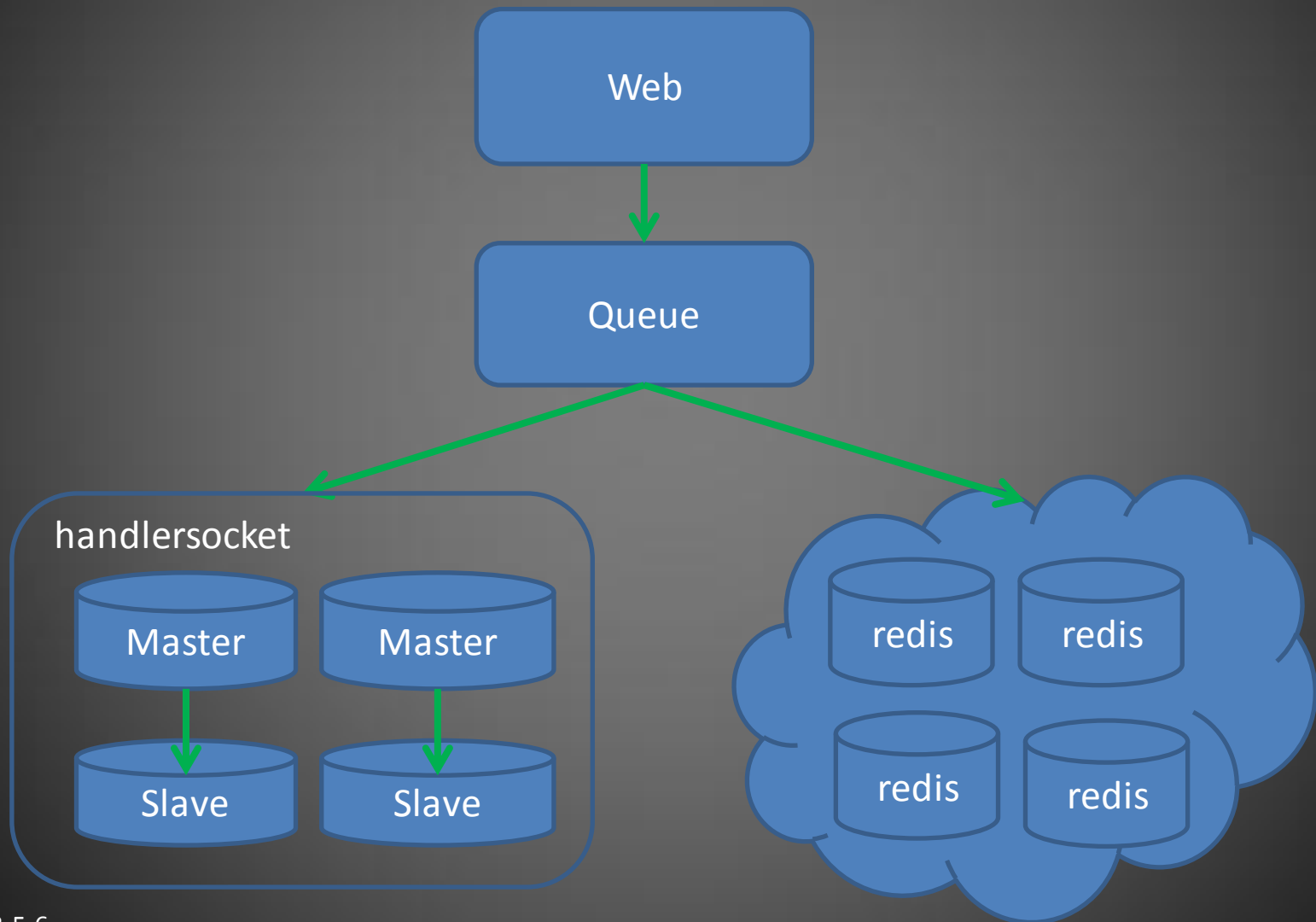
通知

□ Set

- Key=uid
- Value=appkey



通知



大纲

□ Redis简介

□ Redis应用

□ 运维经验

运维经验

- 自动化

- 监控报警

- Redis HA

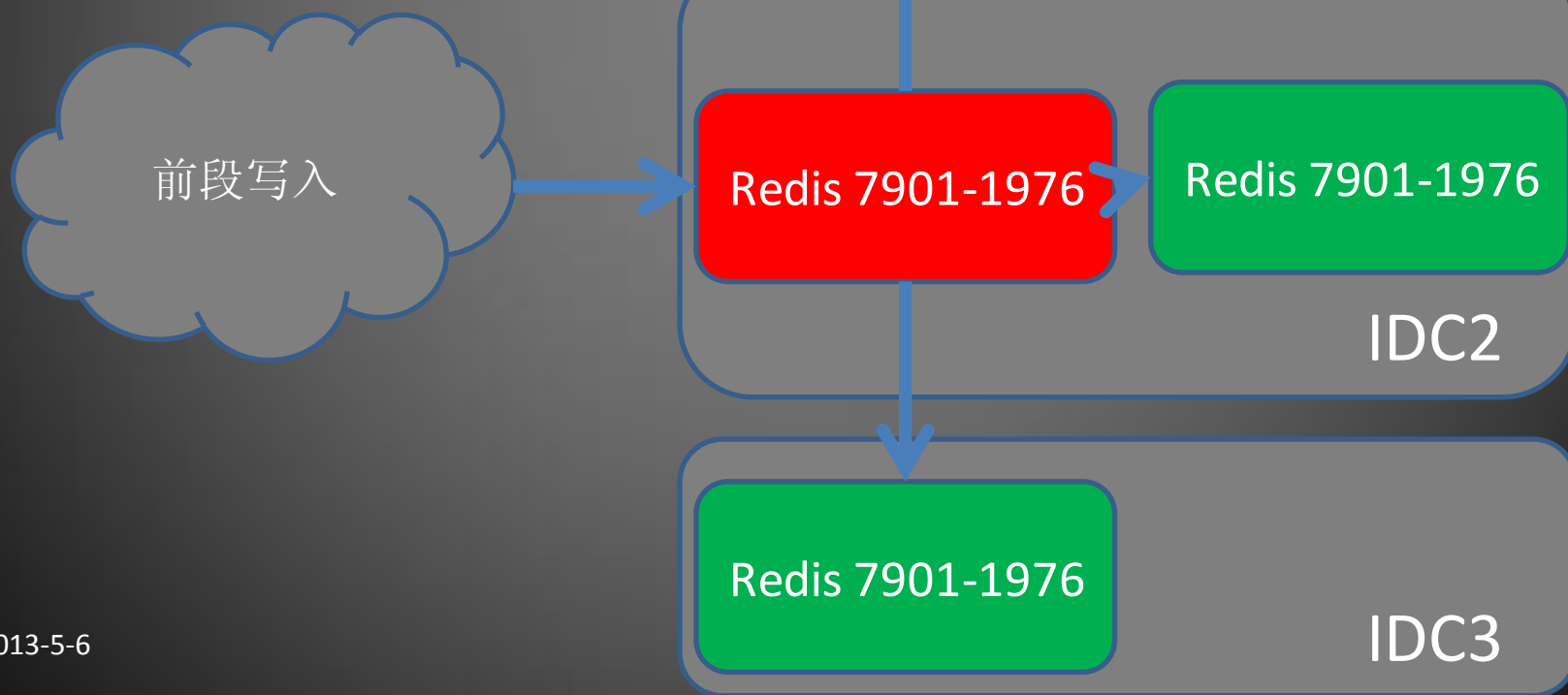
- 改进

- 坑

自动化

□自动部署

□自动扩容



信息查询

» Redis管理

信息查询

安装部署

扩容

通知索引

搜索

| 产品线 | 端口 | IDC | IP | 角色 | 内存(G) | 磁盘(G) | 利用率(%) | 统计 | 联系人 |
|------|---|---------|-------------------|-----|-----------|-------------|--------|---------------------------------|--------------------------------|
| 通知索引 | 7263,7264,7259,7265,7260,7266,7261,7262,7258,7253,7254,7255,7256,7251,7257,7252 (端口数:16) | 北京 (2台) | 10.73. [REDACTED] | s | 81.3/93.7 | 86.9/731.9 | 80.36 | Cmd:174.638(亿) Read:933.359 (亿) | 业务负责人 [REDACTED] 分析 [REDACTED] |
| | | | 10.73. [REDACTED] | s | 81.5/93.7 | 90.4/731.9 | 81.54 | | |
| | | 北京 (4台) | 10.75. [REDACTED] | m/s | 83.4/93.7 | 91/730.7 | 80.46 | | |
| | | | 10.75. [REDACTED] | s | 83.2/93.7 | 170.9/730.7 | 82.44 | | |
| | | | 10.75. [REDACTED] | s | 83.1/93.7 | 170.9/730.7 | 80.74 | Mem:595.375(GB) Write:2.617 (亿) | 运维组:nosql |
| | | | 10.75. [REDACTED] | m/s | 83.3/93.7 | 264/730.7 | 80.43 | | |
| | | 广州 (2台) | 10.75. [REDACTED] | s | 83/93.7 | 175.9/731.7 | 81.32 | | |
| | | | 10.75. [REDACTED] | s | 82.9/93.7 | 89.6/731.7 | 80.96 | | |

自动部署

» 安装部署

信息查询

安装部署

扩容

机器IP列表

Port列表

产品单元

软件

参数

主库

是否启动

注册

数据库平台

redis

max_mem: 2 (G)

maxmem_policy: noeviction

☒ 主库 ☒ 启动 ☒ 注册节点信息

服务器状态检查

部署

日志:

----log end----

自动扩容

» 扩容

信息查询

安装部署

扩容

机器IP列表

Master列表

Master_Port列表

产品单元

软件

参数

主库

是否启动

注册

数据库平台

redis

max_mem:

2

(G)

maxmem_policy:

noeviction

☐ 主库

☒ 启动

☒ 注册节点信息

服务器状态检查

扩容

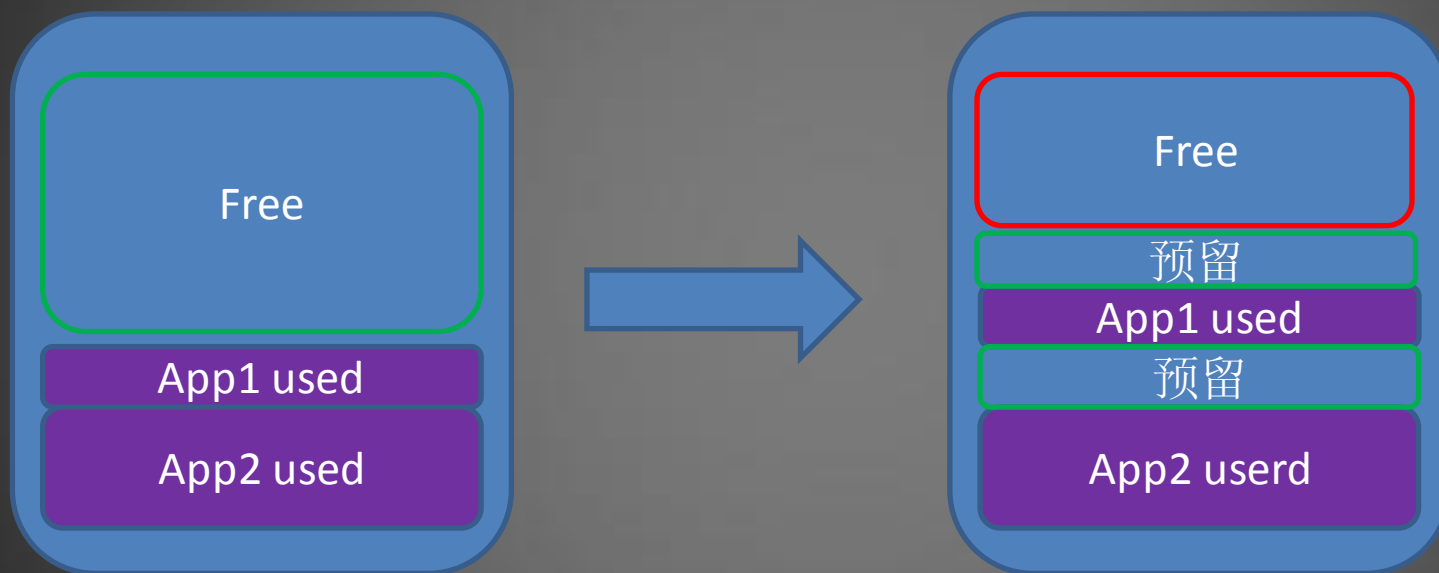
扩容信息

| Id | Port | IP | Master_Port | Master_IP | Cpunit | Software | Status | Last_io | Batch_ID | Beginntime | Endtime | Message |
|-----|------|------------|-------------|------------|---------|----------|--------|---------|------------|---------------------|---------------------|--|
| 276 | 9988 | 10.75.17.1 | 9988 | 10.75.17.1 | RedisHA | redis | ✓ | 0 | 1366107469 | 2013-04-16 18:18:53 | 2013-04-16 18:19:49 | Ok. redis 10.75.17.1:9988-->10.75.17.1:9988 slaveof success. |

资源池



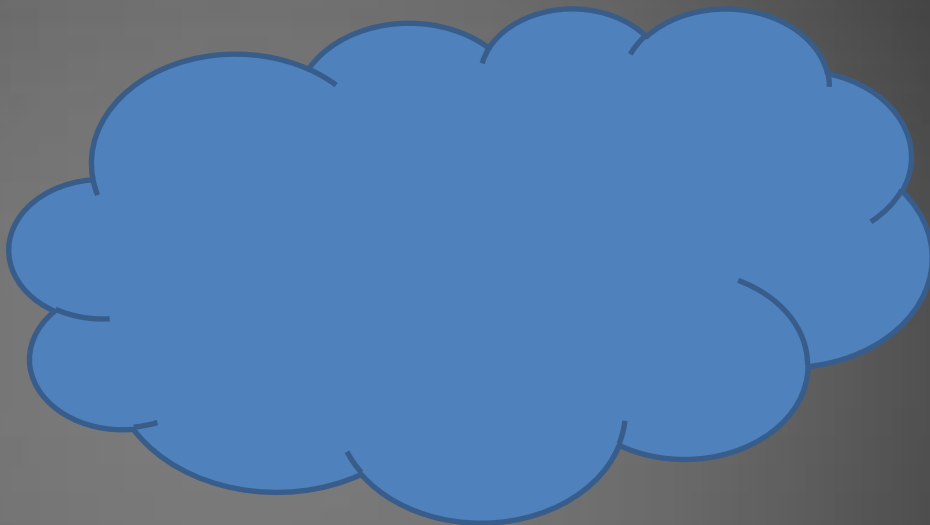
内存



□自动部署

□自动迁移

□自动扩容



Host: 10.77.25.20:7902:rcount

Host: 10.77.25.20:7904:rcount

Host: 10.77.25.20:7906:rcount

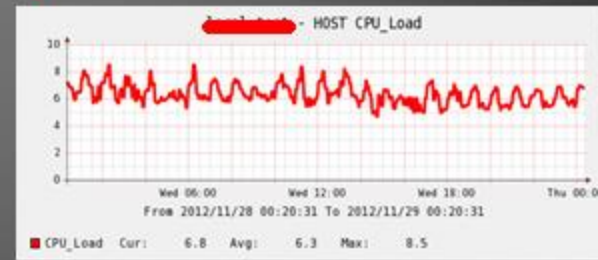
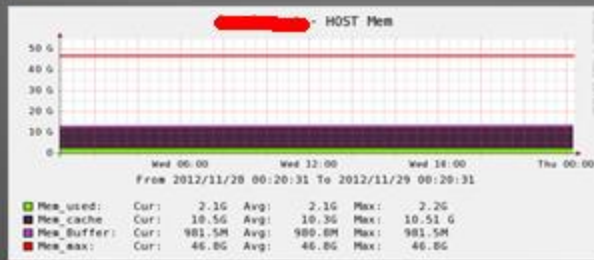
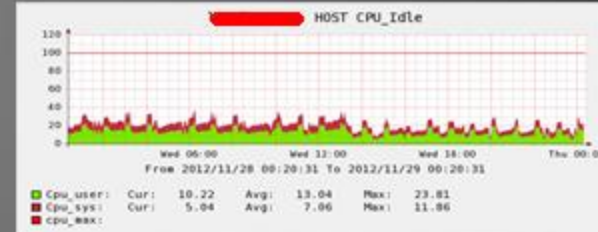
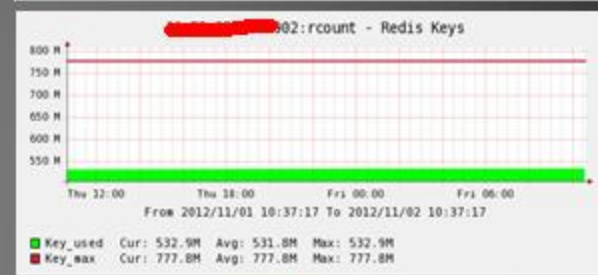
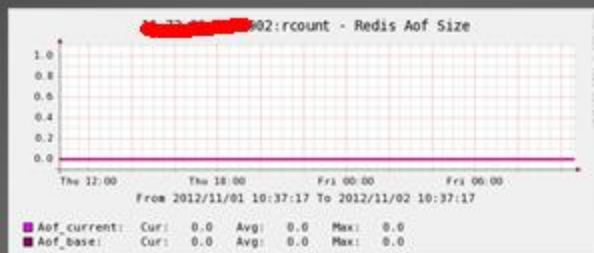
Host: 10.77.25.20:7908:rcount

Host: 10.77.25.20:7910:rcount

Host: 10.77.25.20:7912:rcount

Host: 10.77.25.20:7914:rcount

Host: 10.77.25.20:7916:rcount



报警

□CPU

- 单核使用率

□Load

□Disk

- 剩余空间
- 增长速度

报警

□ Connect

□ Replication

□ connected_clients

□ AOF

– $\text{aof_current_size} / \text{aof_base_size}$

报警

□Memory

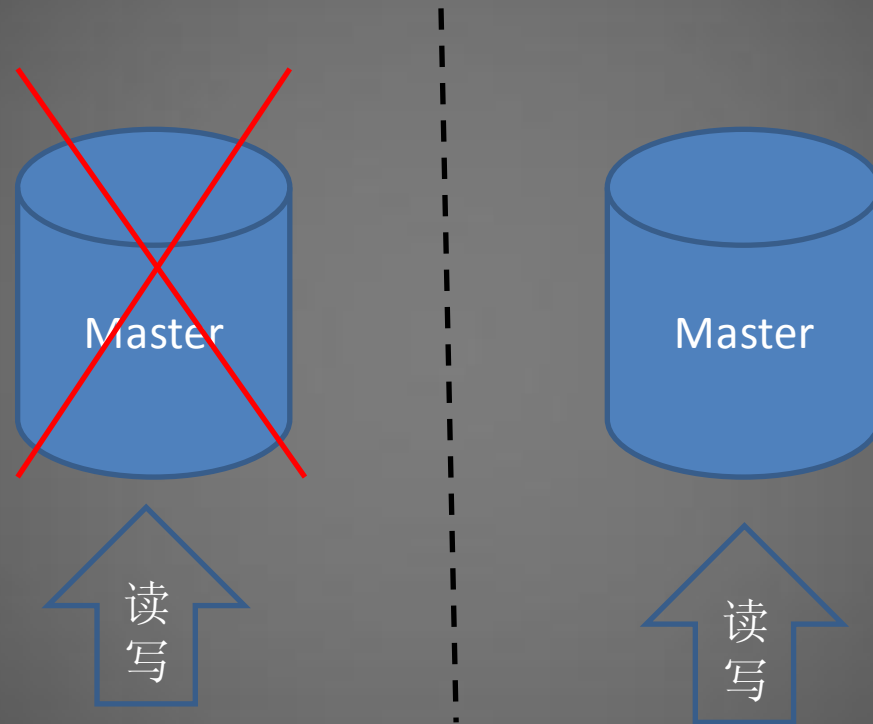
- 服务器可用内存
- 实例可用内存
 - maxmemory
 - maxmemory-policy noeviction

Redis HA

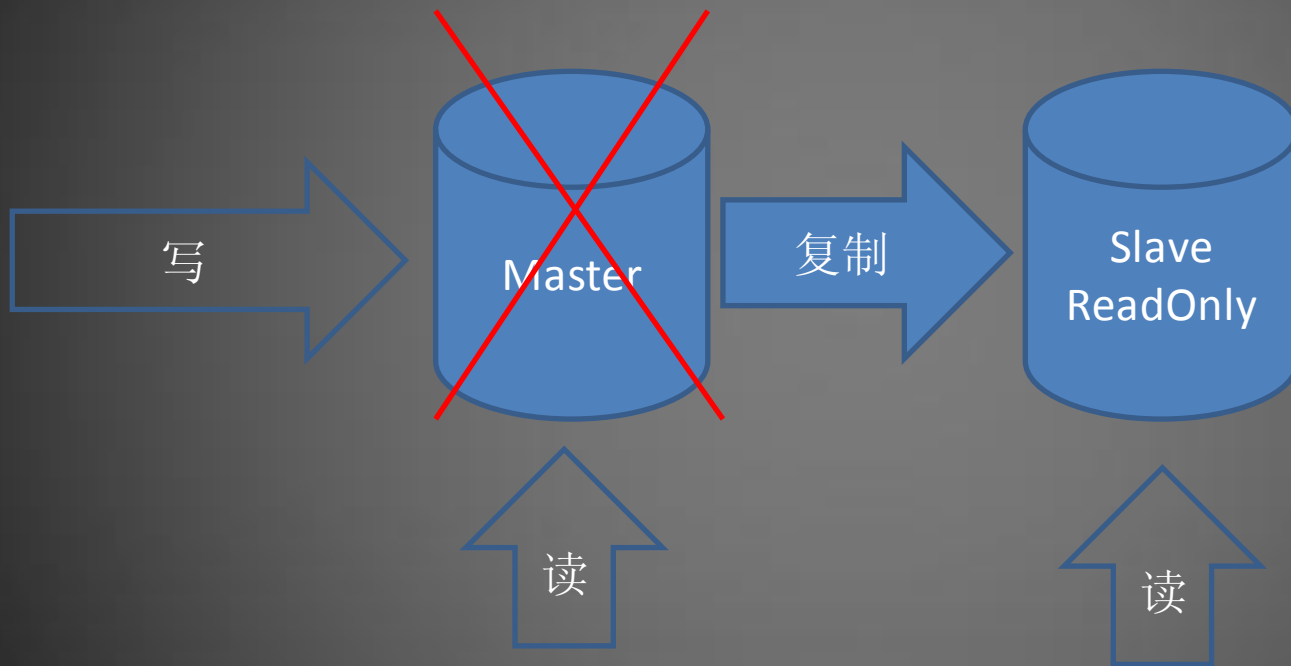
□双写

□复制

双写



复制



已有方案

□ redis_failover

- 多IDC

□ 官方的Sentinel

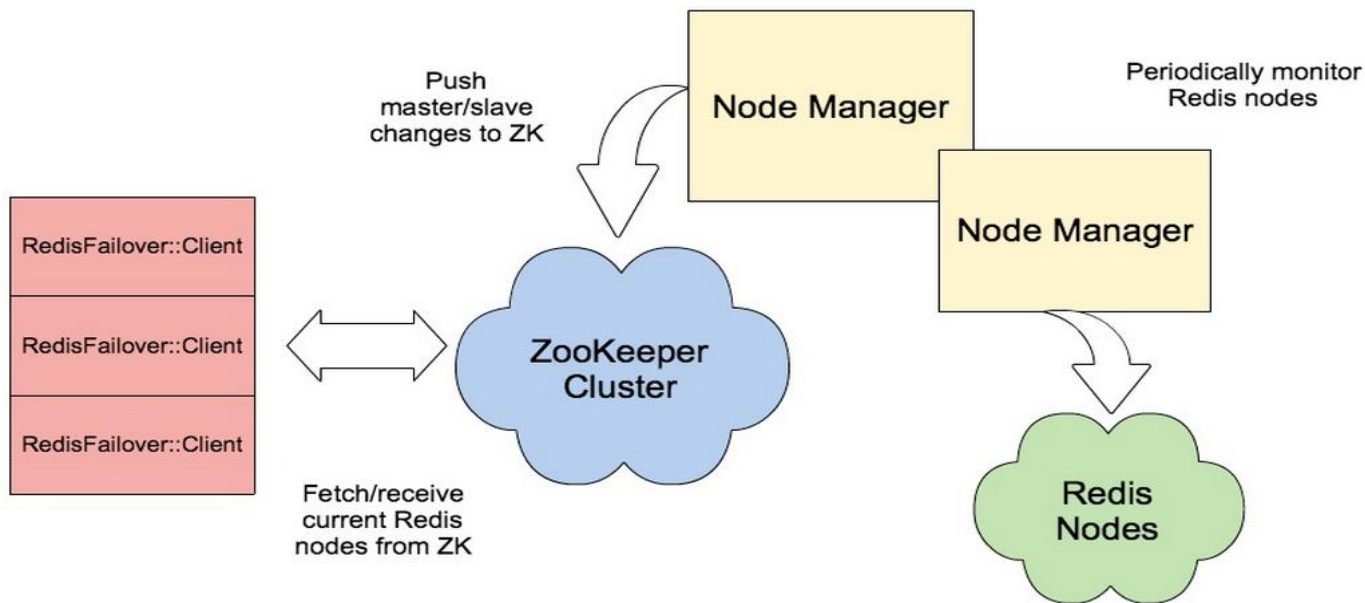
- 分布式

□ 双写

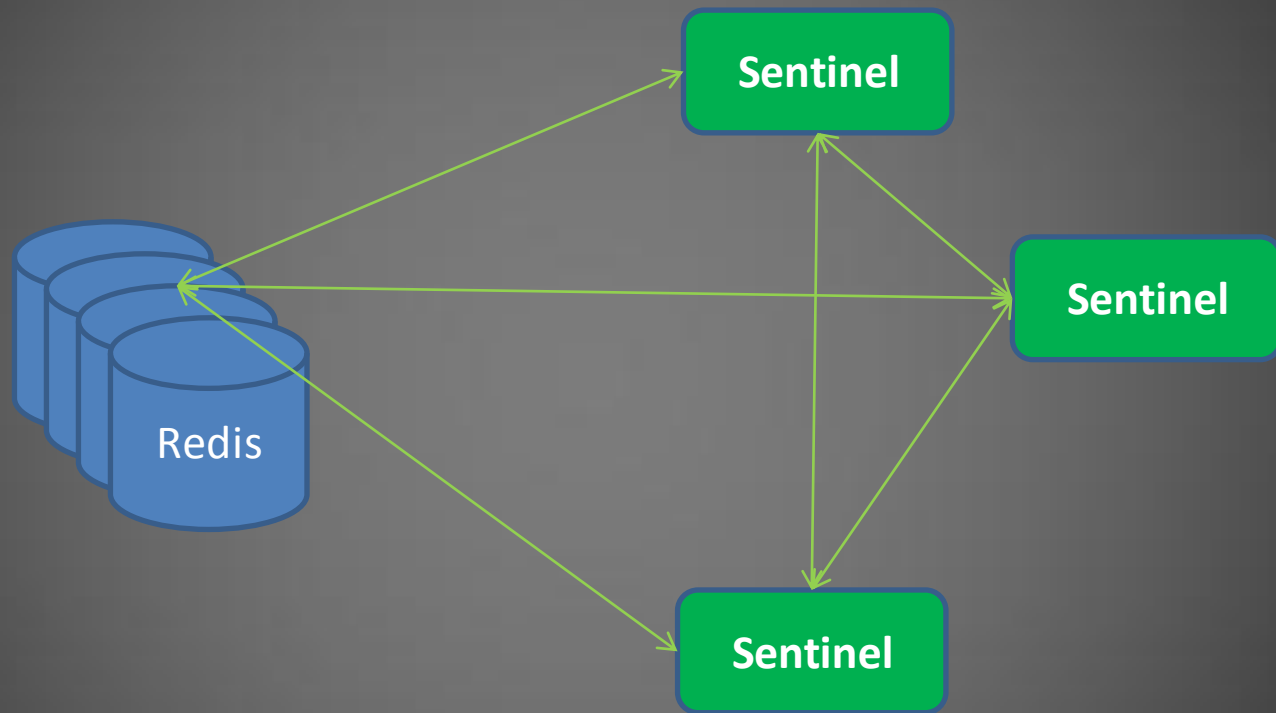
- 数据的一致
- 服务恢复

redis_failover

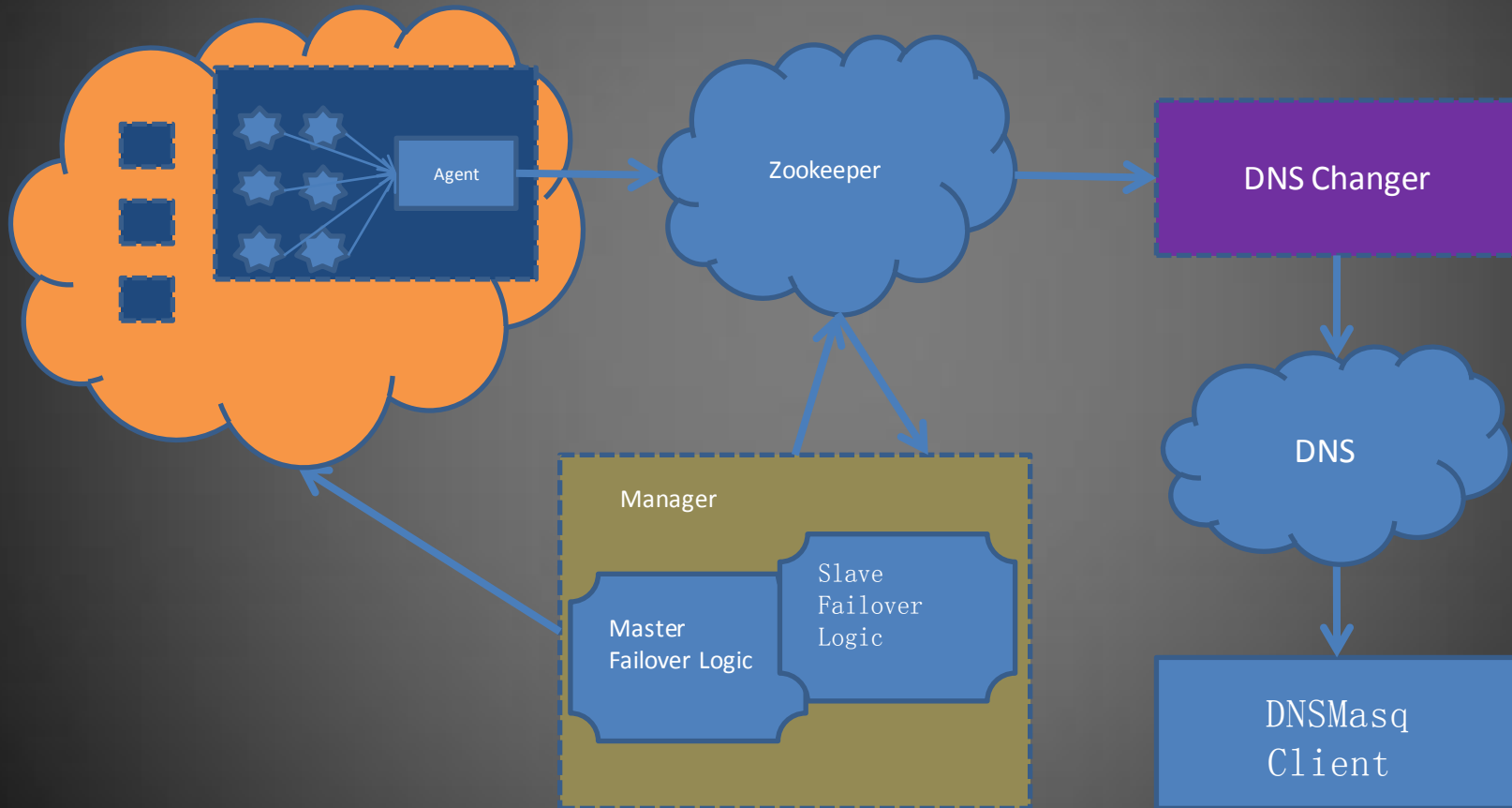
Architecture



Sentinel



Redis HA

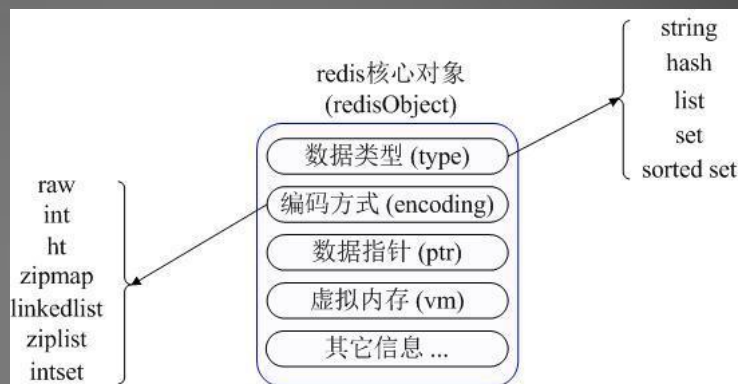


经验

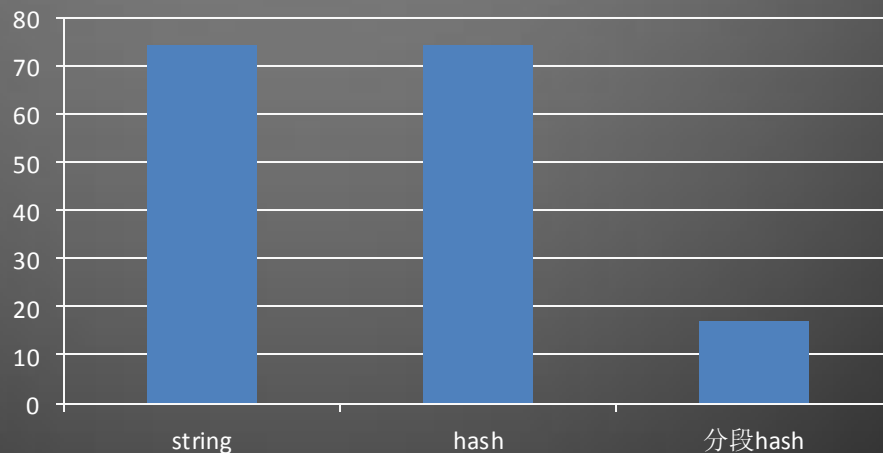
□ string -> hash

□ Instagram

□ 100w数据



内存占用



经验

hash-max-ziplist-entries 512

hash-max-ziplist-value 512

list-max-ziplist-entries 512

list-max-ziplist-value 64

set-max-intset-entries 512

zset-max-ziplist-entries 128

zset-max-ziplist-value 64

activeremhashing yes

经验

□网卡中断

□Bgrewriteaof或者bgsave分开

□充裕的磁盘空间($128\text{G} \times 2 \times 3 = 768\text{G}$)

□RAID卡

改进

□ Rediscounter

- rdb+aof
- aof position
- 固定大小滚动

□ redis

- cronbgrewrite 47 3 * * *
- 自动升级
- 慢查询
- 网络抖动引起Slave重传

Q&A

@曾经的阿飞

rj03hou@gmail.com