




2016 杭州·云栖大会  
THE COMPUTING CONFERENCE

云栖社区  
yq.aliyun.com

# 微博的Redis定制之路

2016  
The Computing Conference

主办单位:  杭州

 Alibaba Group  
阿里巴巴集团

战略合作伙伴: 

肖鹏  
微博研发中心数据库技术负责人



扫码观看大会视频

---

# 目录

## content

---

探索期：野蛮生长

发展期：争奇斗艳

稳定期：拾遗补缺





# 介绍

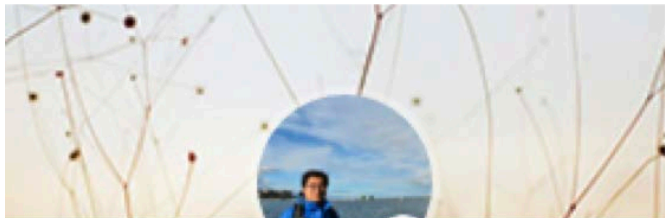


## 个人介绍

肖鹏，微博研发中心

主要负责微博数据库

( MySQL/Reids/HBase/Memcached ) 相关的业务保障，性能优化，架构设计以及周边的自动化系统建设。经历了微博数据库各个阶段的架构改造，包括服务保障及SLA体系建设，微博多机房部署，微博平台化改造等项目，10年互联网数据库架构和管理经验，专注于数据库的高性能和高可用技术保障方向。



billy鹏的足迹



Lv.31

874

关注

4156

粉丝

12293

微博



扫码观看大会视频

## 部门介绍

MySQL

Redis

MC/MCQ

HBase

Hadoop

Kafka

消息队列，缓存，持久存储，统一的数据平台。





# Redis定制之路



探索期

发展期

稳定期



扫码观看大会视频

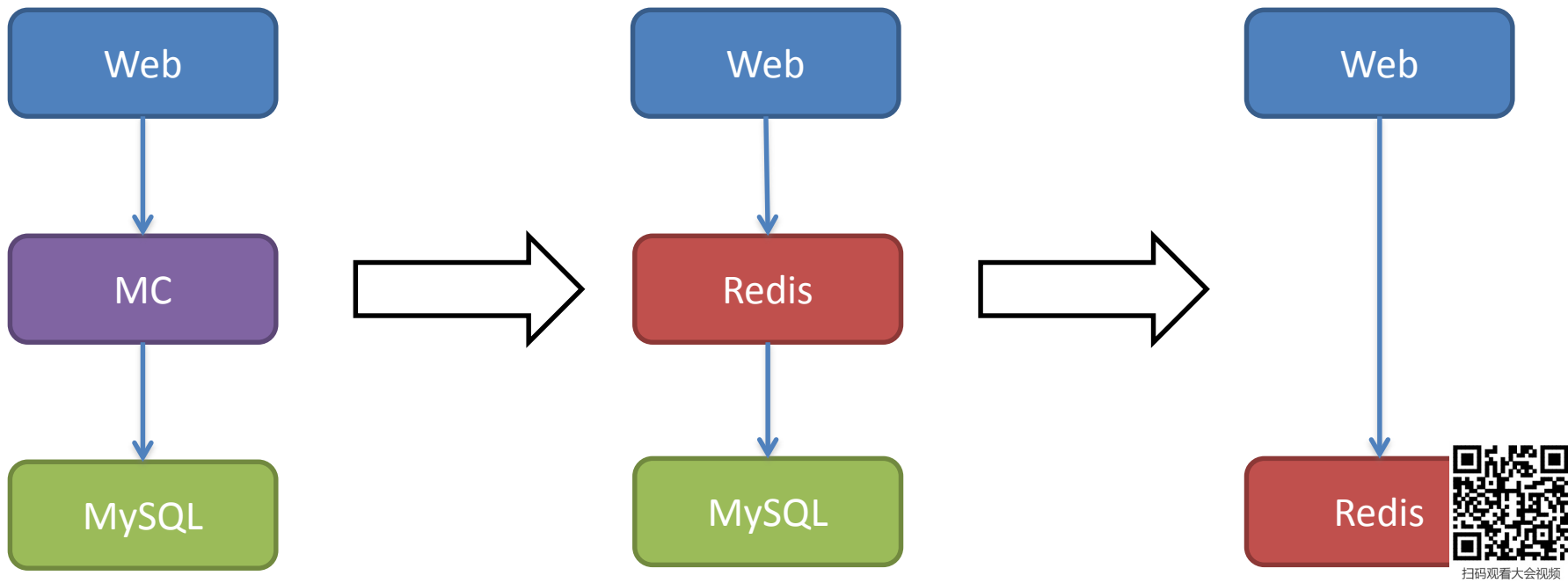


# 探索期：野蛮生长





## 第一个Redis集群上线



遇到的问题：

1

### 性能问题

在执行bgsave的时候会出现明显的卡顿，业务波动

2

### 故障恢复问题

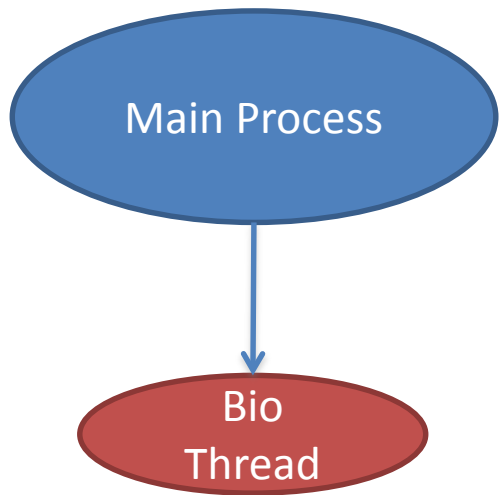
宕机后恢复主从速度慢，经常出现带宽洪峰问题

3

### 运维复杂度问题

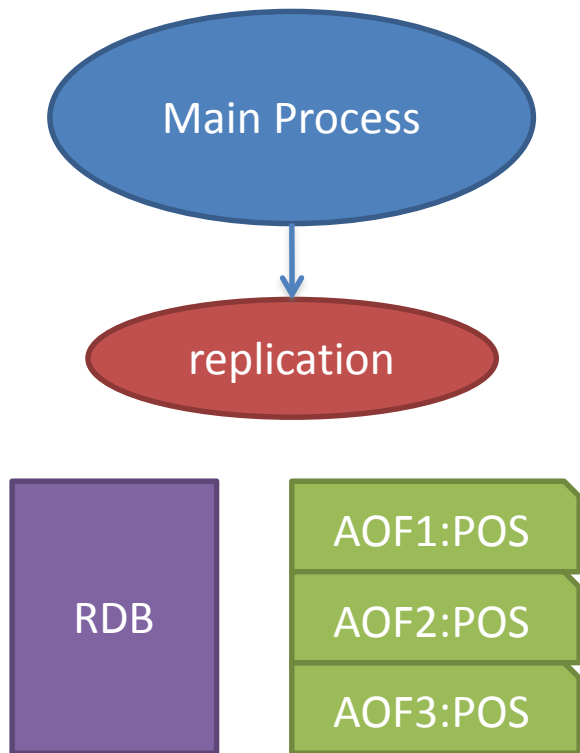
版本升级频繁，主库切换常态化





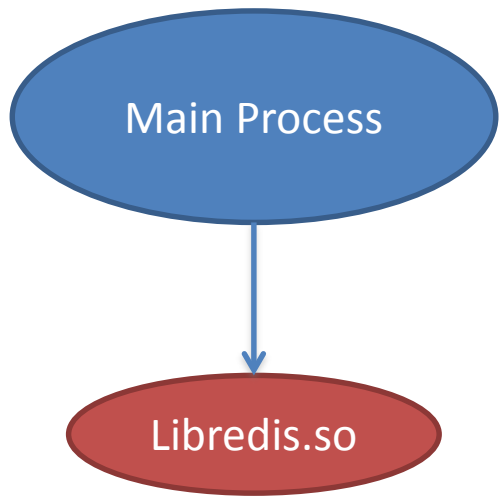
- 从主线程中独立出来Bio thread，专门执行bgsave等操作，避免干扰。
- 内置cronsave功能，控制备份时间。
- 放弃bgaoofrewrite。





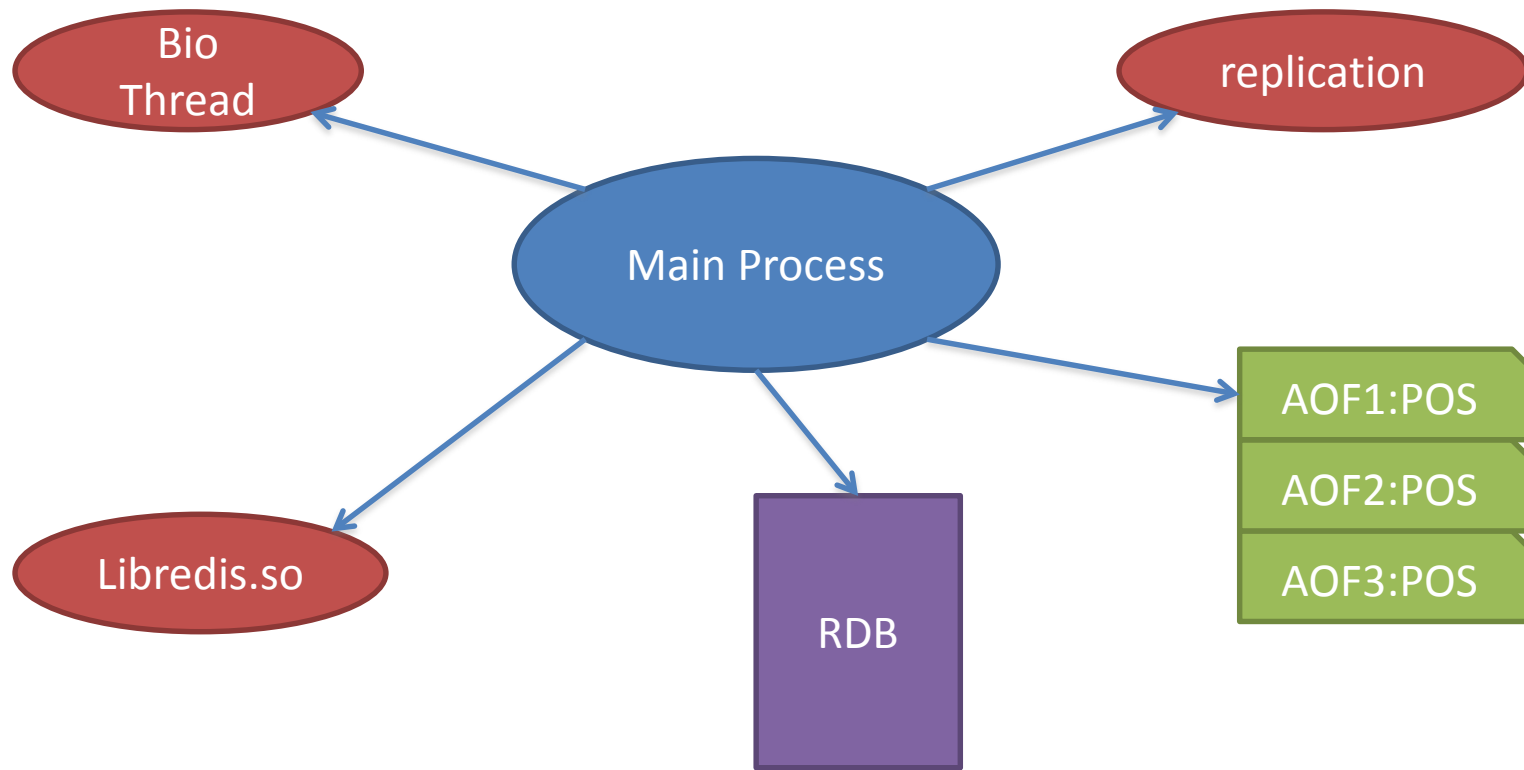
- 修改aof机制，增加pos位。
- 修改replication机制，实现基于aof+pos位置的数据同步。
- 修改落地机制，改为RDB+AOF滚动机制，保证数据持久化存储。





- 主程序中为全局的数据结构。
- 其他功能封装到libredis.so中。
- 对libredis.so实现动态热加载，每次版本升级实现热升级，不必在进行主库切换，极大的降低运维复杂度和业务风险。

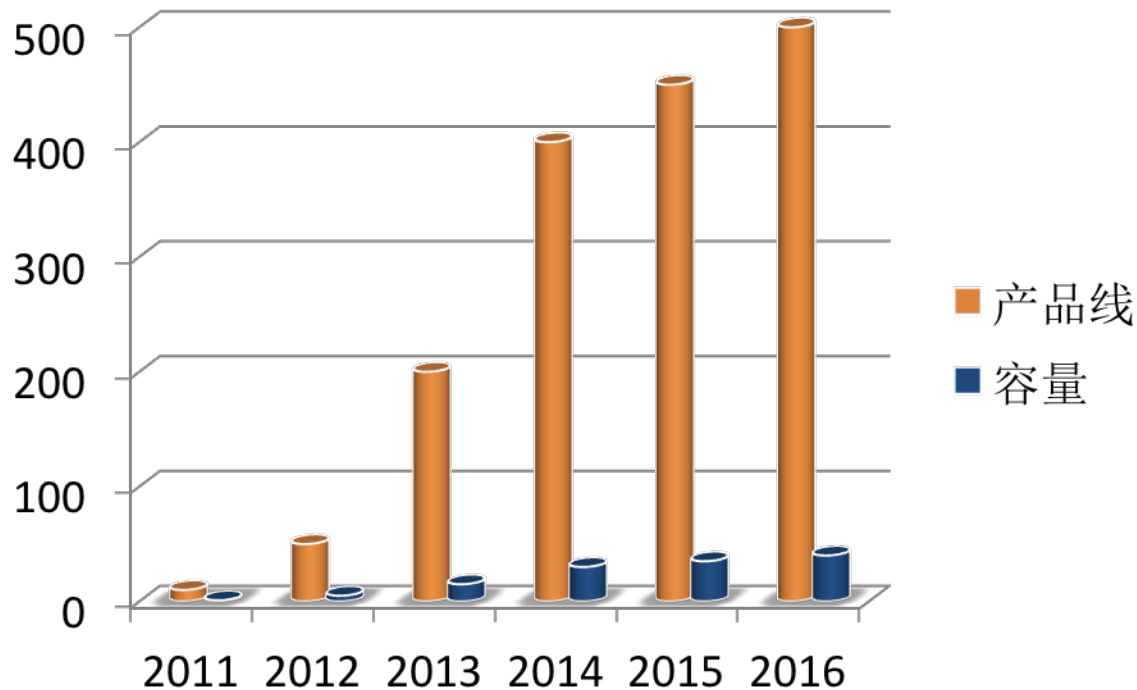






## 发展期：争奇斗艳







1

## 容量问题

日益增长的关系类存储容量问题同成本之间的矛盾

2

## 计数类需求

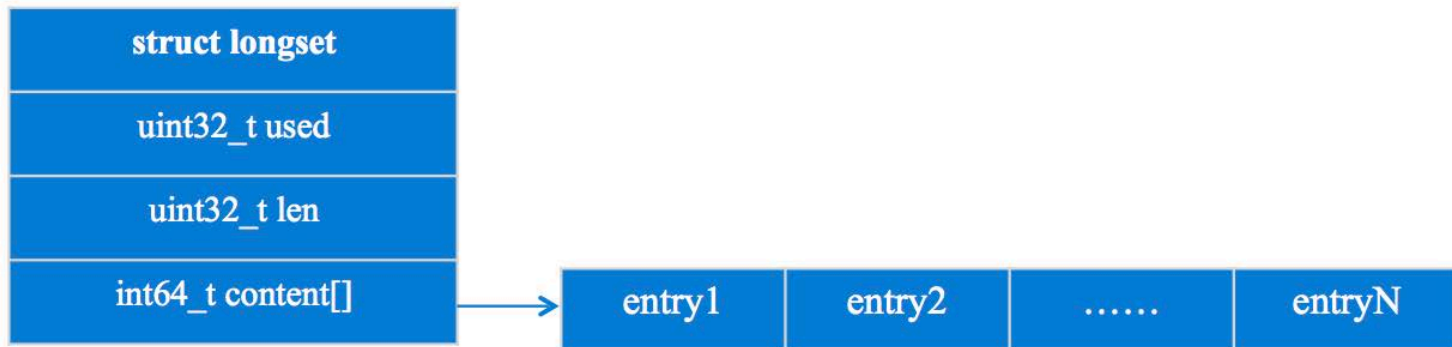
琳琅满目的计数类场景需要专门的解决方案

3

## 判断类需求

频繁的判断需求无法接受长时延



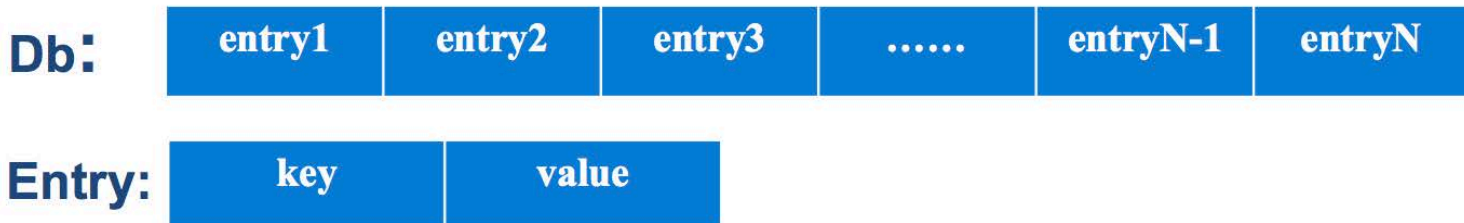


- 定位从storage变为cache
- 放弃原生hash结构，定制longset数据结构
- 固定长度开放寻址的hash数组
- Miss后，Client批量O ( 1 )

MEM 1/10



扫码观看大会视频

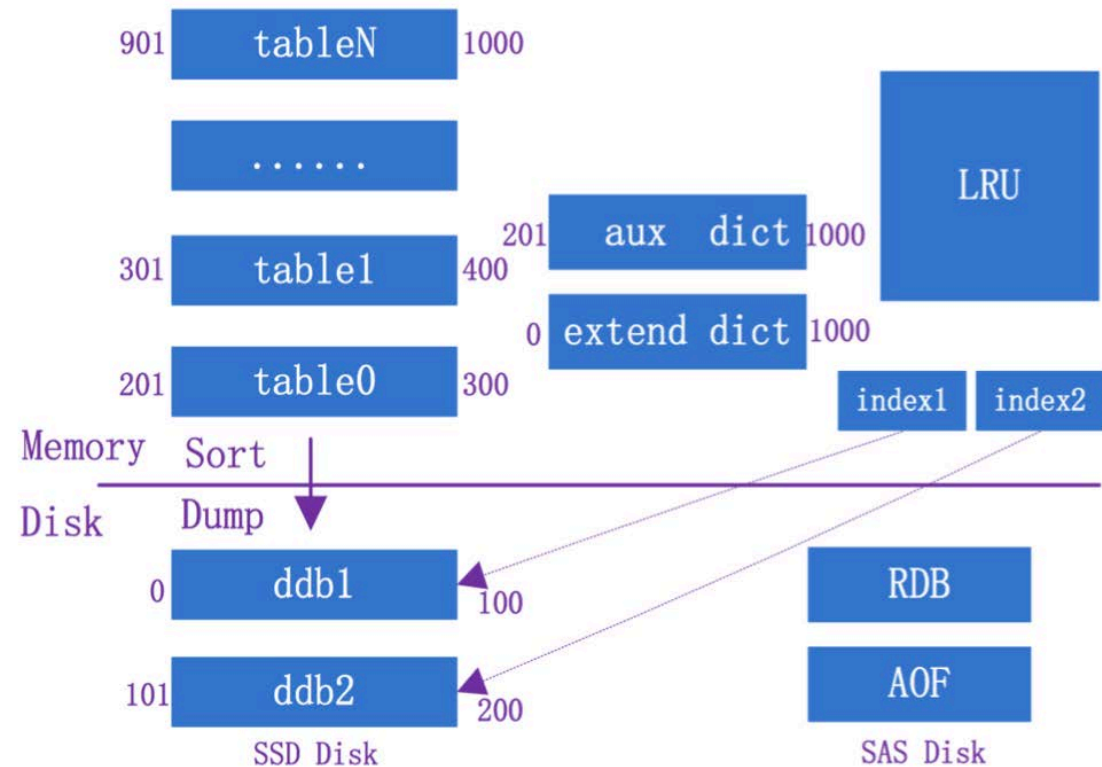


- 根据业务场景将kv改为定长kv
- Double hash解决冲突
- 预先分配内存，降低指针开销

MEM 1/4

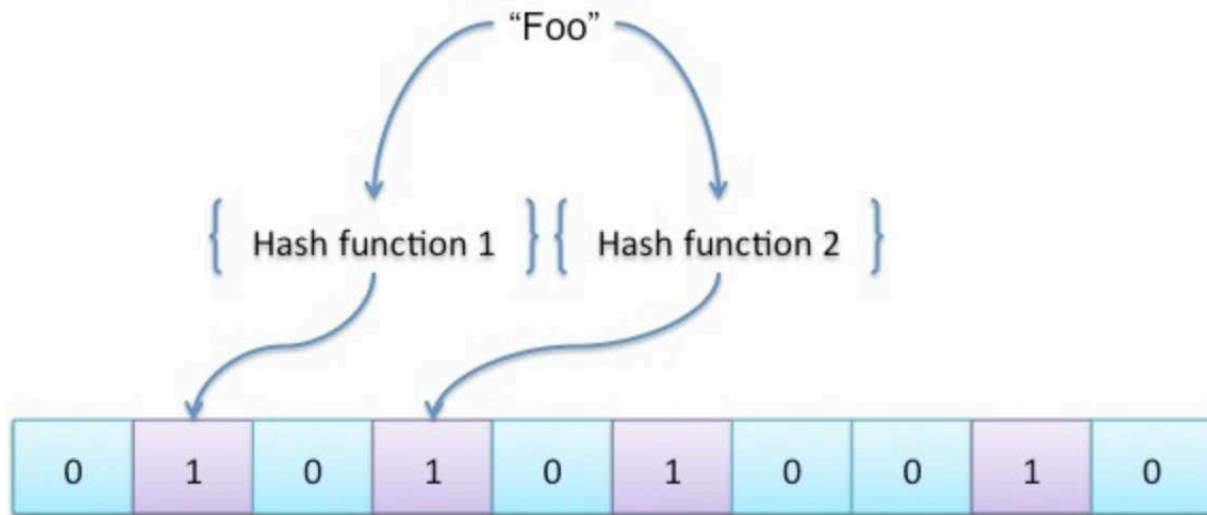


扫码观看大会视频



- Schema支持多列存计数，降低内存占用，减少网络开销。
- 冷热数据分离，热内存，冷磁盘
- 增加LRU cache模块，保证冷数据的性能。
- 异步IO线程访问磁盘，避免性能问题





- bloomfilter
- 解决是否赞过等场景的高并发问题，并解决成本问题。



# 稳定期：拾遗补缺



1

## 服务化

随着业务的变大，集群规模也变大，拼写hash规则已经变成负担，并且每次变更扩容都变得及其的复杂。

2

## 数据流动

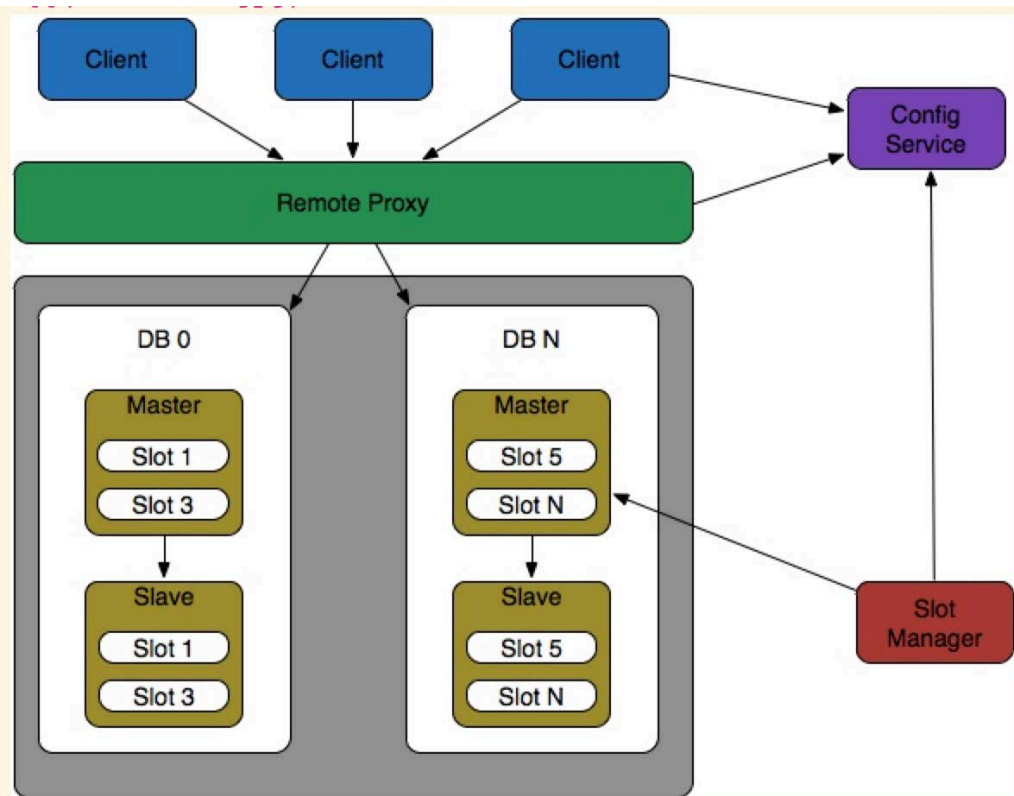
多个存储的更新存在数据不一致的情况，并且业务双写也增加了程序的复杂性

3

## 数据迁移

原有redis的拆分和改造都需要DBA和RD共同参与以便保证数据的一致性以及尽量降低对业务的冲击

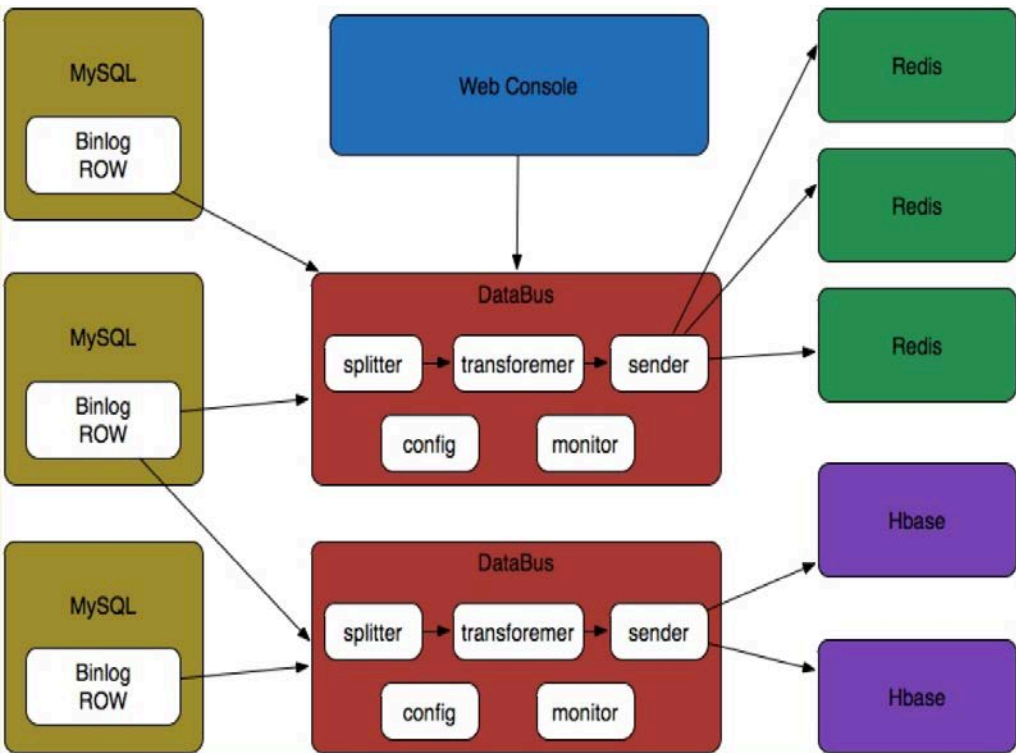




- 加入中间层实现路由功能
- 引入configer service模块，统一配置管理
- 引入slot设计，实现动态的slot迁移，对业务无影响。
- 基于slot设计无缝扩容功能，实现线性扩容
- 融入HA等运维机制

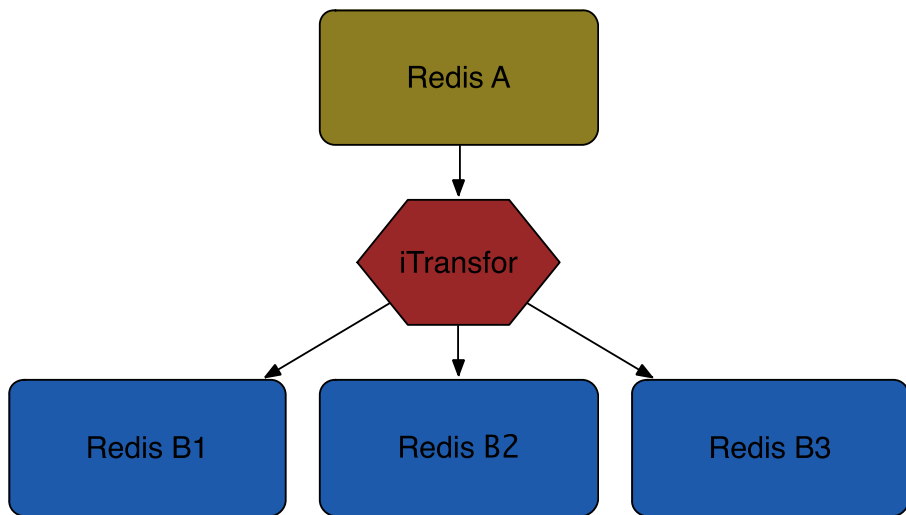






- 基于MySQL的row格式binlog日志实现消息的流转
- 设计模块加载机制，支持对于mysql消息的自定义处理规则的处理
- 下游兼容redis、HBase等存储
- 融入HA等运维机制





- 基于改造过的aof进行消息解析
- 实现对单端口的redis实现拆分
- 可以实现rang 或者hash的拆分规则
- 可以实现上下游不同的redis分支
- 独立组件 or 成为redis的一个功能



业务是驱动技术变革的源动力  
反过来技术的变革也推进业务形态的变化  
两者互相作用，呈螺旋上升的趋势

未来？  
在平稳中期待下一次爆发



2016 The  
Computing  
Conference  
**THANKS**

