SDCC 2017深圳站
互联网应用架构实战峰会

CSDN

支撑百亿级访问
微博Feed平台的缓存架构及演进之路

新浪微博 陈波@fishermen

大纲

- 数据挑战
- Feed架构
- Cache架构及演进
- 总结及展望

微博数据

日活用户 1.4亿+

平台接口
日访问百亿级

核心记录千亿级

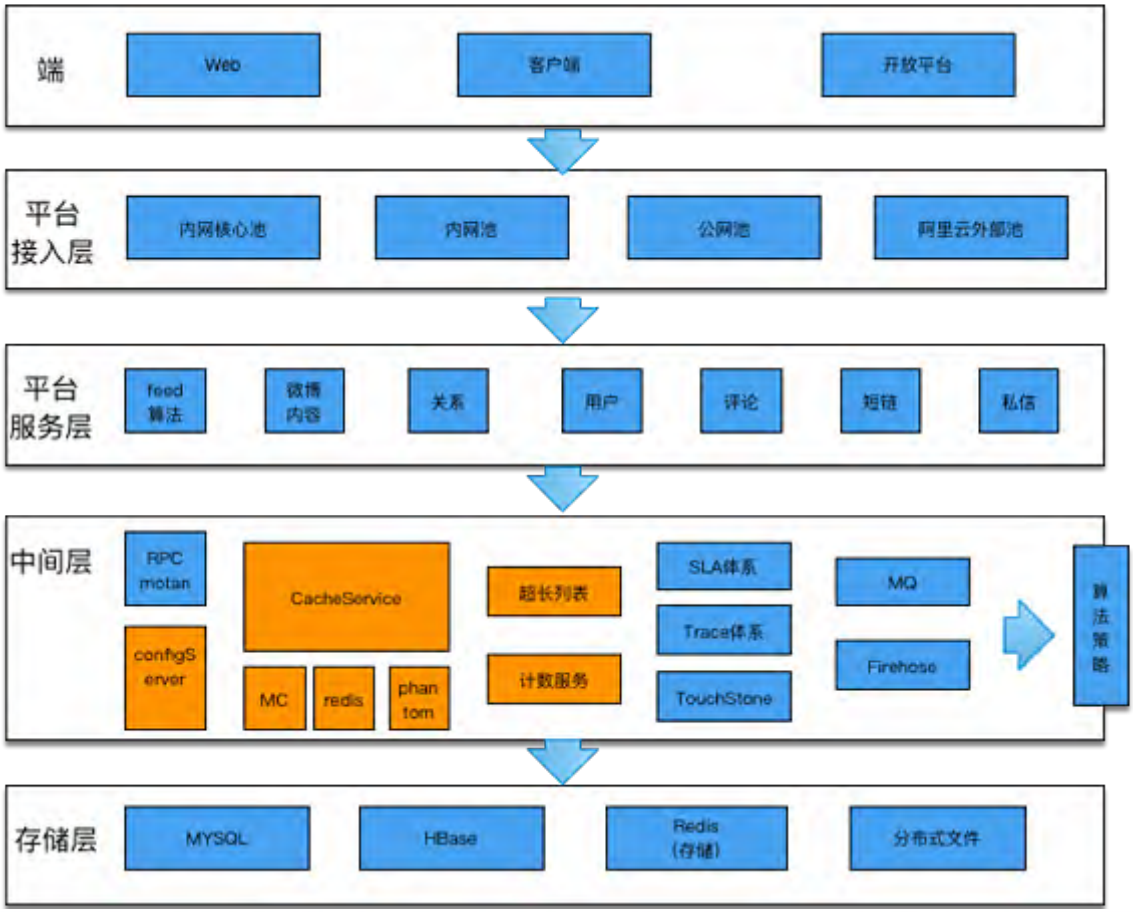
Cache
内存百T级

Cache
日访问万亿级

单个核心数据
Cache QPS百万级

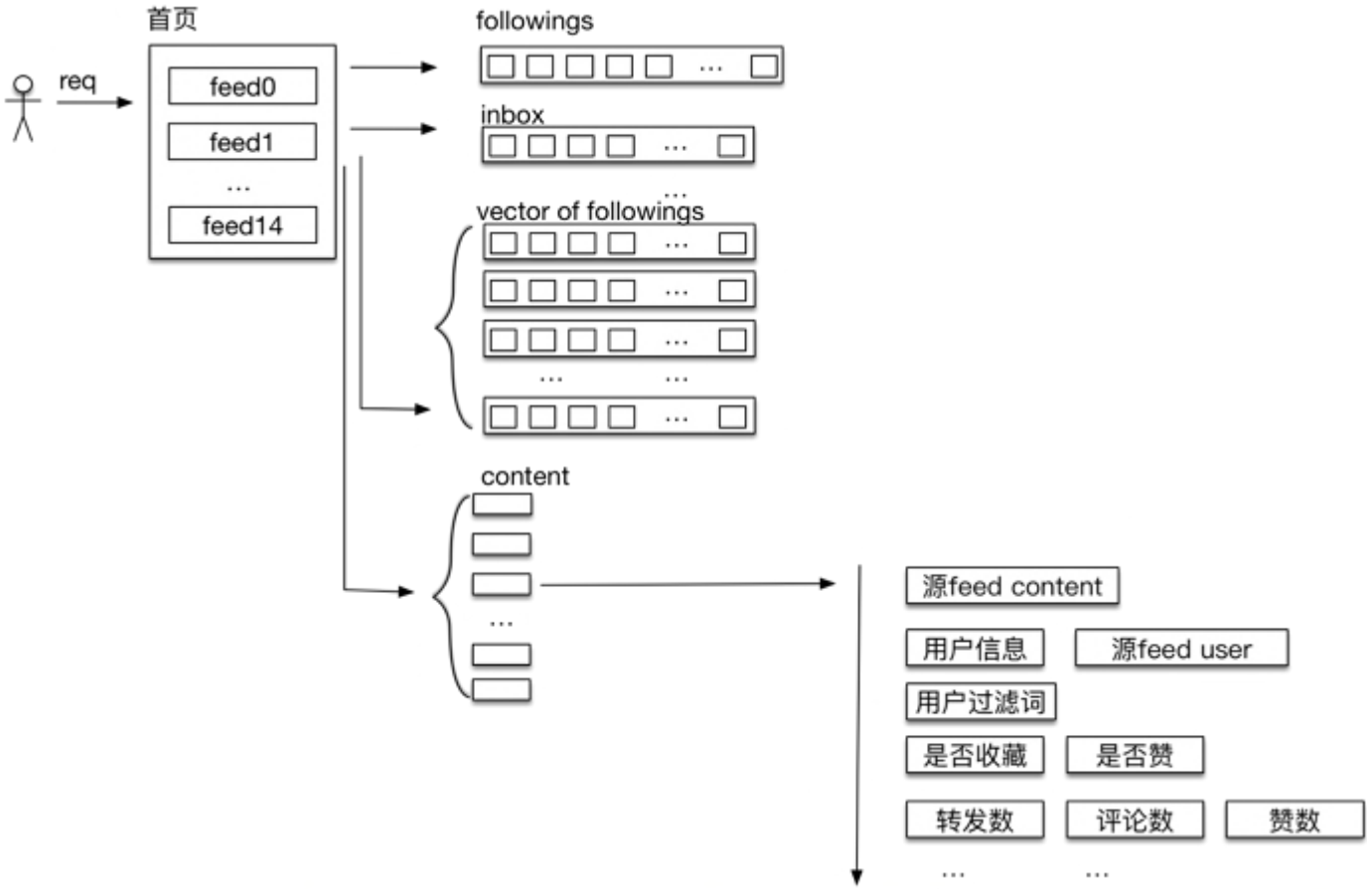
Feed平台系统架构

- 总体架构



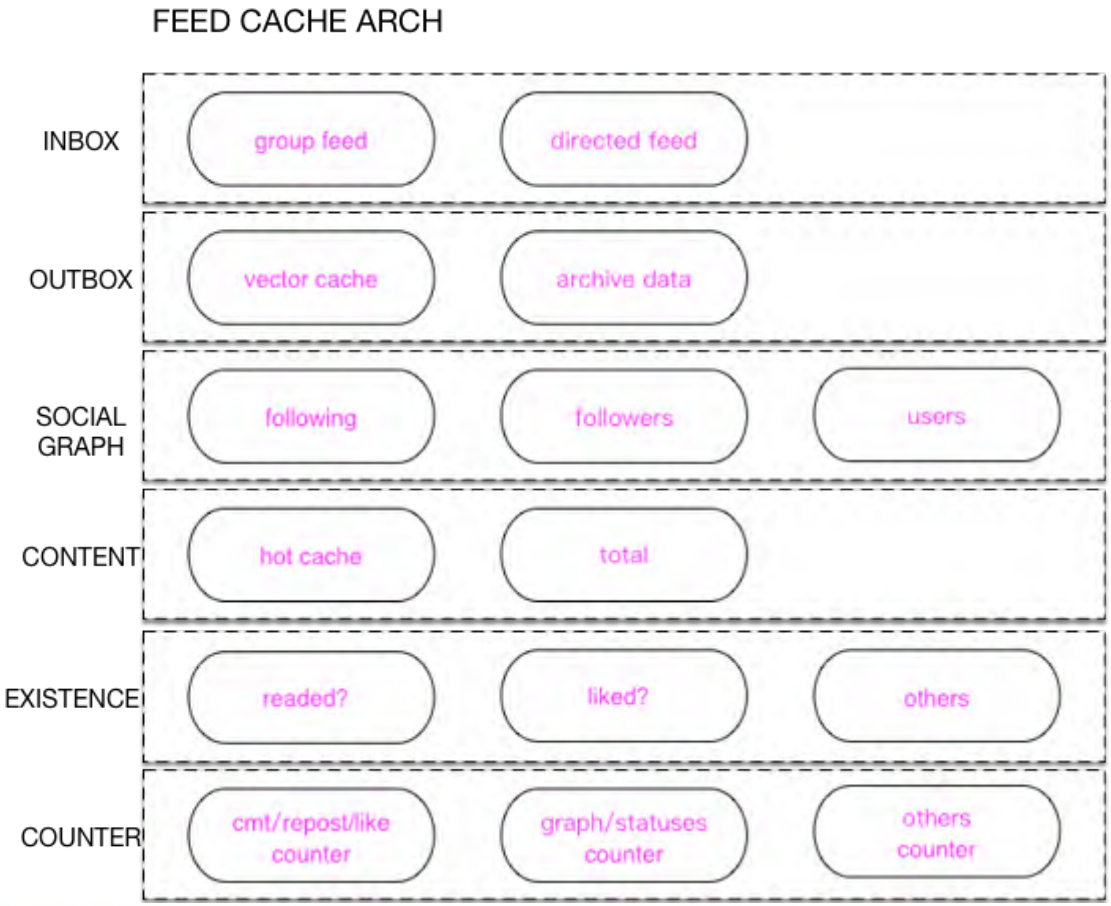
Feed timeline

- 构建流程



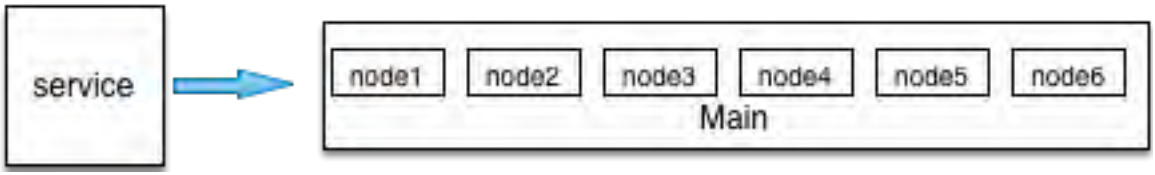
Feed Cache

- Cache 架构

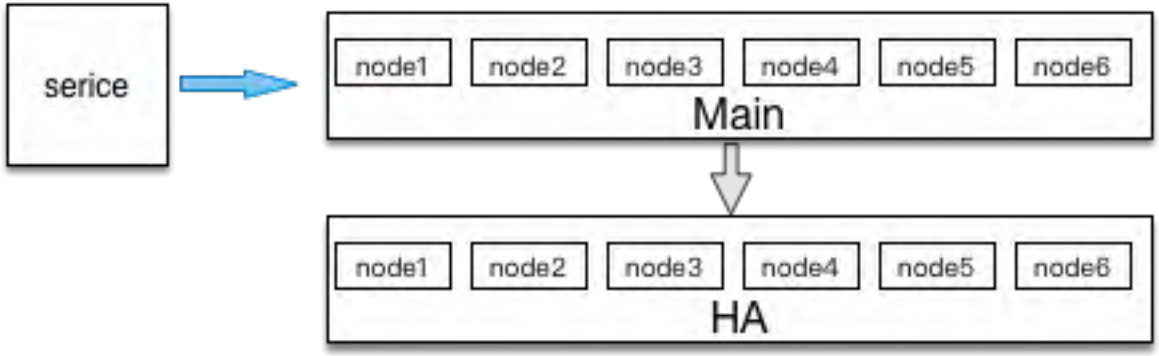


简单KV数据类型（MC）

- 单层HASH

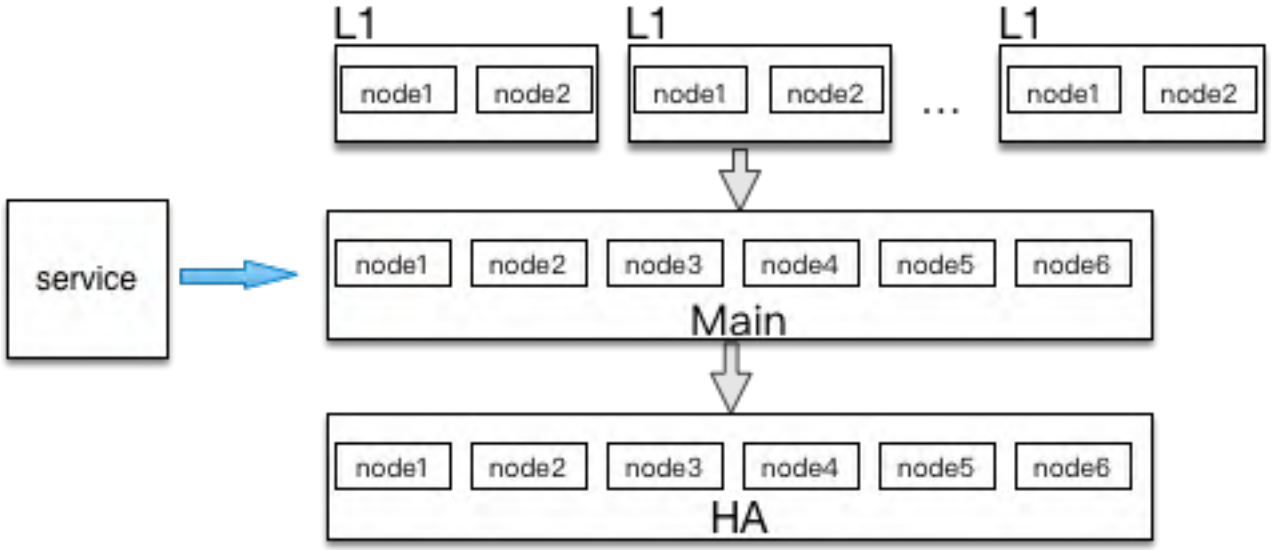


- Main-HA



简单KV数据类型

- Main-HA-L1



简单KV数据类型

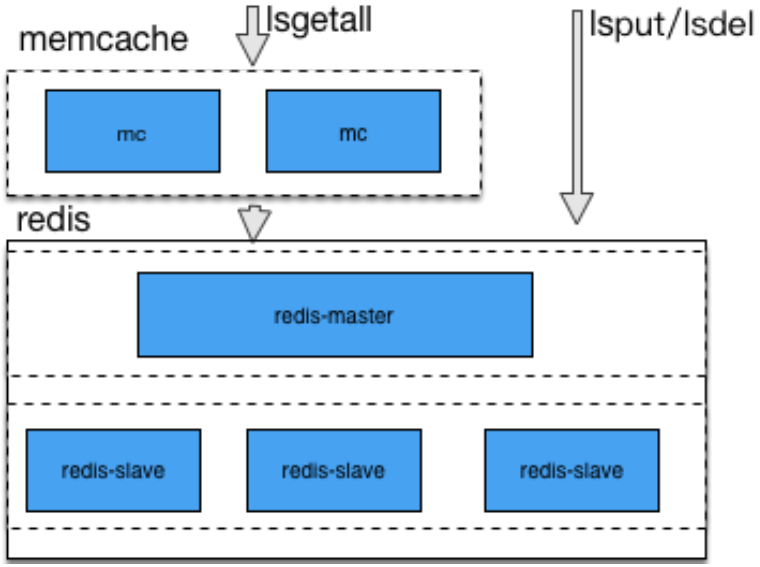
- Key Point
  - 层内HASH节点不漂移，miss则穿透
  - 多组L1 读取性能升 峰值流量成本降
  - 多层访问 命中率升 ,可用性高， cache node故障透明
  - MC data：Protocol Buffer, QuickLZ 压缩
  - 读写策略
    - Write：多写
    - Read：逐层穿透，miss 回写

集合类数据类型

- MC
  - Json/xml → PB
  - Client 序列化 反序列化
- Redis
  - 关注，粉丝，分组，共同关注，XX也关注
  - 部分修改，分页获取，资源计算
  - Hash 分布，MS，cache/storage
  - 30+T 内存，2-3万亿rw/day

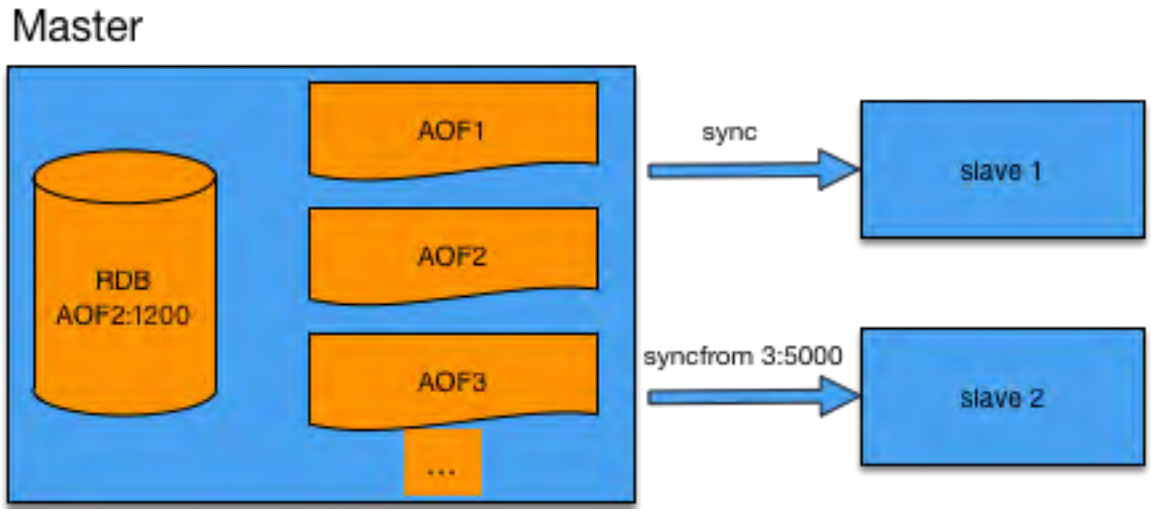
Weibo Redis 扩展

- Longset
  - Long型开放数组，Double Hash 寻址
  - Client 构建数据结构，elements 单次写入
  - Lsput：填充率过高，由client重建
  - Lsgetall/Lsdump MC 做前置抗量



Weibo Redis 扩展

- 其他扩展
  - 热升级：10+分钟→毫秒级
  - 落地： RDB + rotate Aof
  - 全增量复制
  - 落地/同步速控

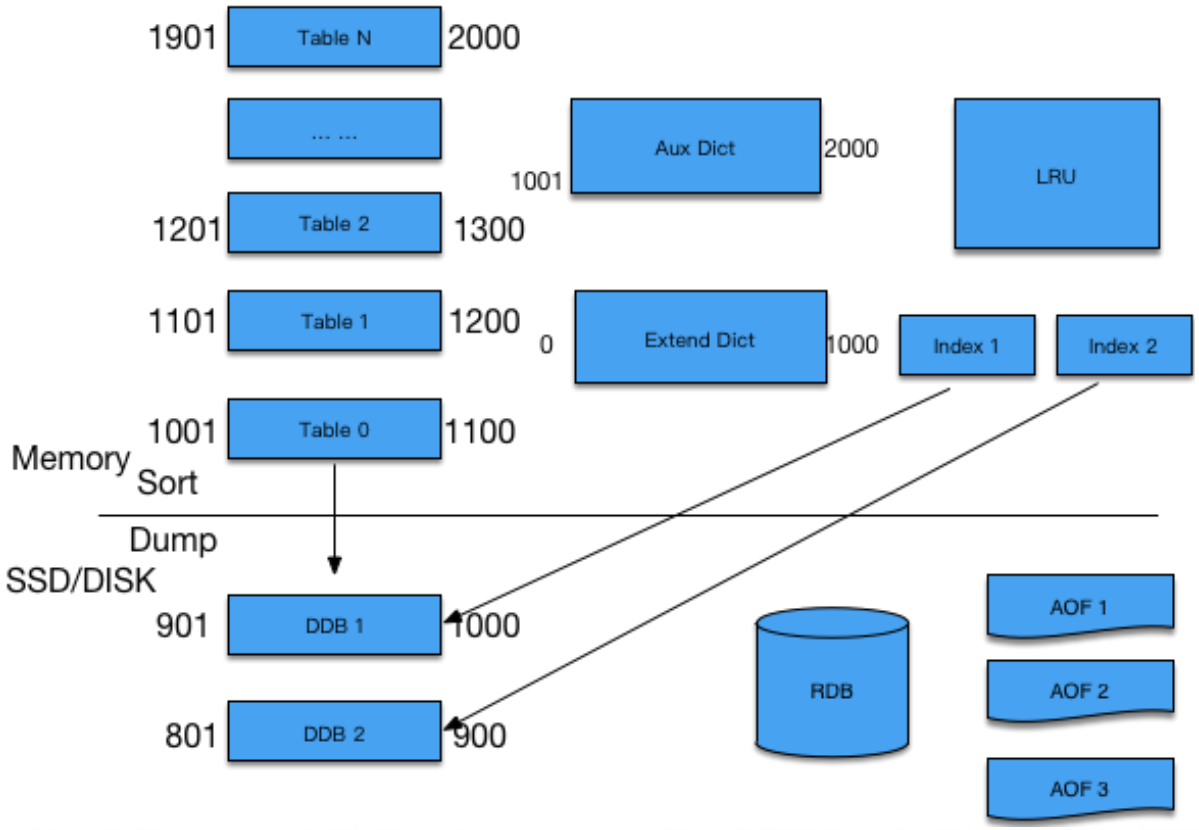


其他数据类型-计数

- Memcached / Redis
  - MC 剔除，重启数据丢失
  - Redis 内存开销 访问性能
- 计数器 CounterService
  - Shema 支持多列，按bit分配
  - Tables 预分配，double-hash寻址
  - 内存降为1/5-1/15以下
  - 冷热分离，SSD 存放老数据，老热数据入LRU
  - 落地 RDB + AOF
  - 全增量复制
  - 单机：热数据百亿级，冷数据千亿级

计数器 CounterService

- 存储架构



其他数据类型-存在性判断

- 业务类型需求
  - 检查是否存在（阅读 赞）
  - 单条记录量小
  - 总数据量巨大
  - 每日新增数量大 千亿级

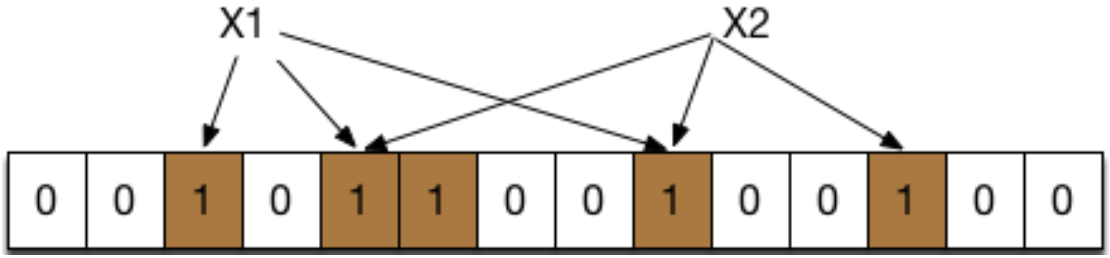
存在性判断

- Redis
  - 单条kv：65 bytes
  - 每日新增内存 6T （不考虑HA）

- CounterService
  - 单条kv：9 bytes
  - 每日新增内存 900G （不考虑HA）

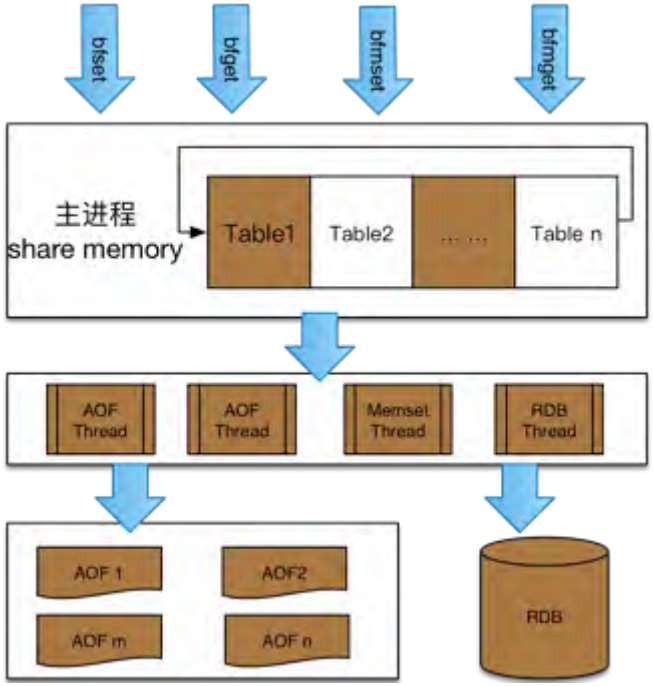
存在性判断

- Phantom
  - Table 分段预分配，段内 bloomfilter
  - 每条kv：1.2 bytes (1%误判)
  - 每日新增内存：120G < 800G < 6T



存在性判断

- Phantom
  - 数据存放共享内存，重启不丢数据
  - 落地 RDB + AOF
  - Redis 协议



小结

- 关注点
  - 集群内高可用
  - 集群内扩展性
  - 组件高性能
  - 存储成本

进一步优化

面向资源/组件管理
如何简化运维？

本地配置模式
如何快速变更？

常规峰值、突发流量
如何快捷、低成本应对？

业务数据分类多
如何独立管控SLA？

业务关联资源太多
如何简化开发？

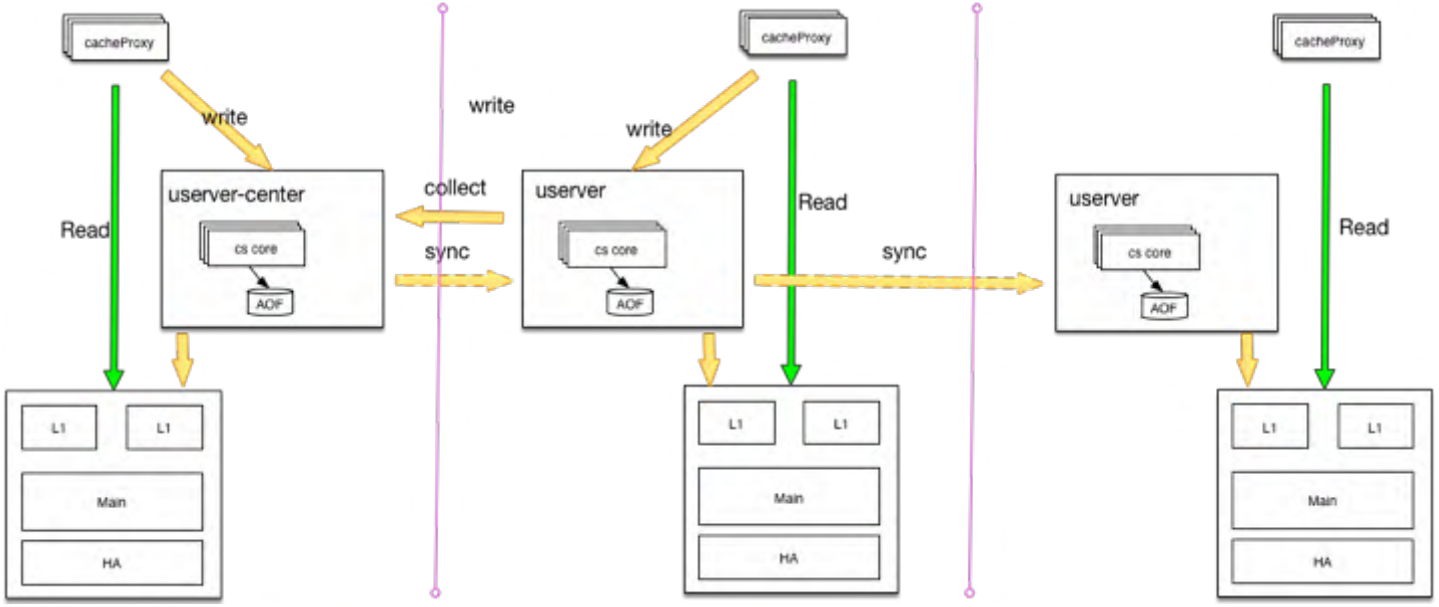
配置服务化

- 本地配置Confs
  - 资源变更→改本地配置→配置更新→重启

- 配置服务化
  - configServer
    - 资源管理 API 化
    - 脚步变更， smart client 同步配置

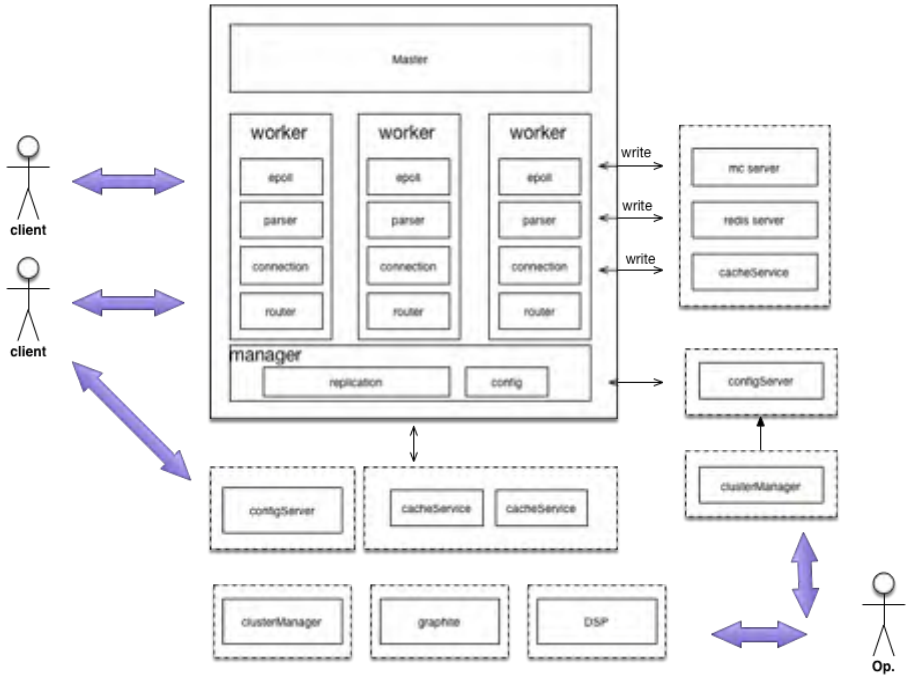
Cache服务化

- Cache 访问
  - Proxy 化
- IDC数据一致性
  - Collecting/replication



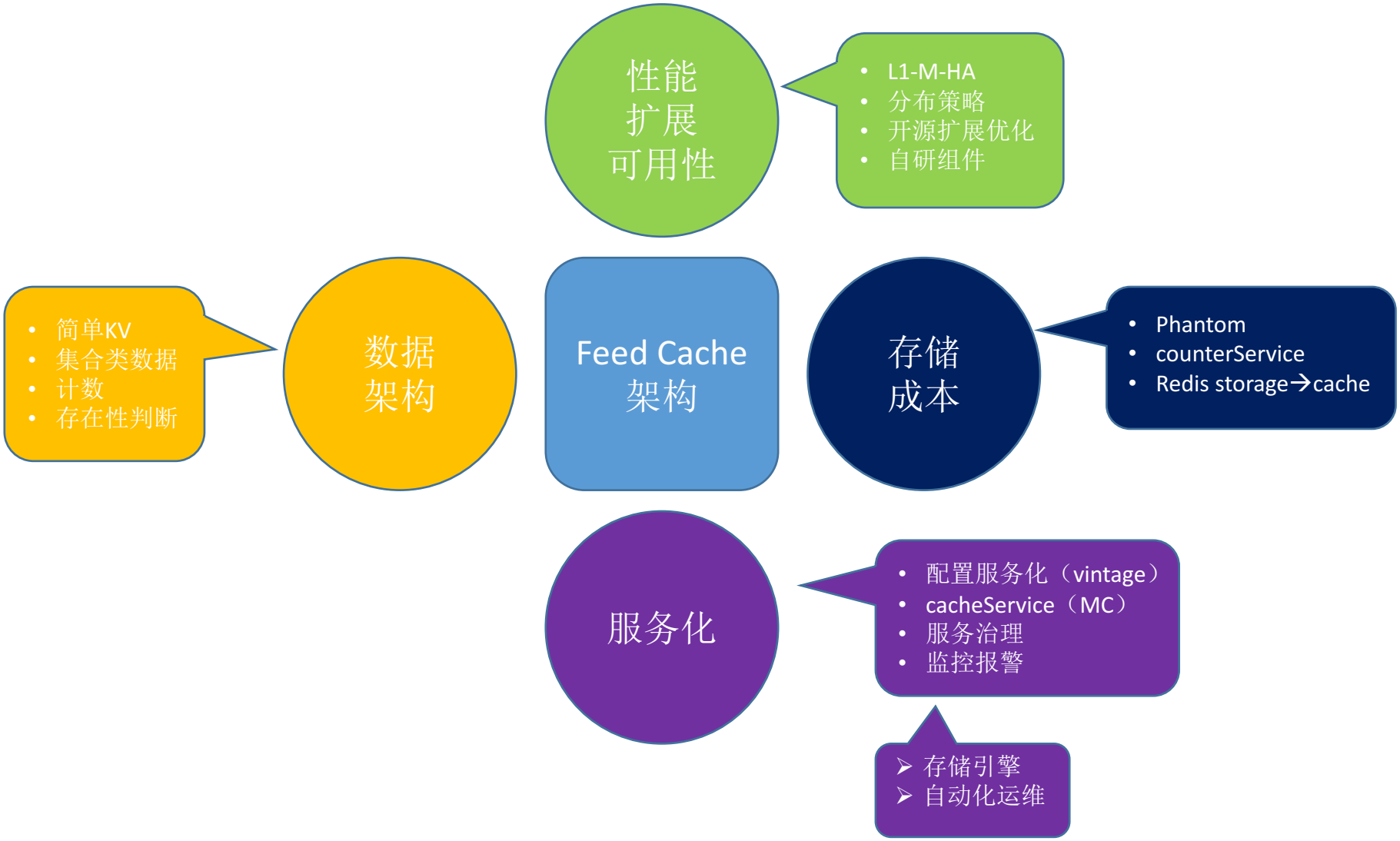
Cache 服务化

- ClusterManager
  - 脚本化→Web 界面化
  - 服务校验 业务SLA
  - 面向服务管控资源
- 服务治理
  - 扩容、缩容
  - SLA 保障
  - 监控报警
  - 故障处理
- 简化开发
  - 屏蔽Cache资源细节
  - 单行配置访问




```
<weibo:cstemplate id="cacheService" namespace="unread-feed"/>
```

总结及展望


性能
扩展
可用性
L1-M-HA
分布策略
开源扩展优化
自研组件
数据
架构
简单KV
集合类数据
计数
存在性判断
Feed Cache
架构
存储
成本
Phantom
counterService
Redis storage→cache
服务化
配置服务化（vintage）
cacheService（MC）
服务治理
监控报警
存储引擎
自动化运维

THANKS