

ANTI-SPAM

ZHIHU



WUKONG

QCon

2016.10.20~22

上海·宝华万豪酒店

全球软件开发大会 2016

[上海站]



购票热线: 010-64738142

会务咨询: qcon@cn.infoq.com

赞助咨询: sponsor@cn.infoq.com

议题提交: speakers@cn.infoq.com

在线咨询 (QQ): 1173834688

团 · 购 · 享 · 受 · 更 · 多 · 优 · 惠

7折

优惠 (截至06月21日)
现在报名, 立省2040元/张

黑产

- 诈骗类
- 薅羊毛类
- 推广类
- 骚扰类

在知乎

- 色情信息
- 产品推新
- 骚扰私信
- 合伙刷赞

找出坏人

- 挑选苹果
- 更多的纬度
- 更多有用的纬度
- 迷惑你的对手（敌明我暗）

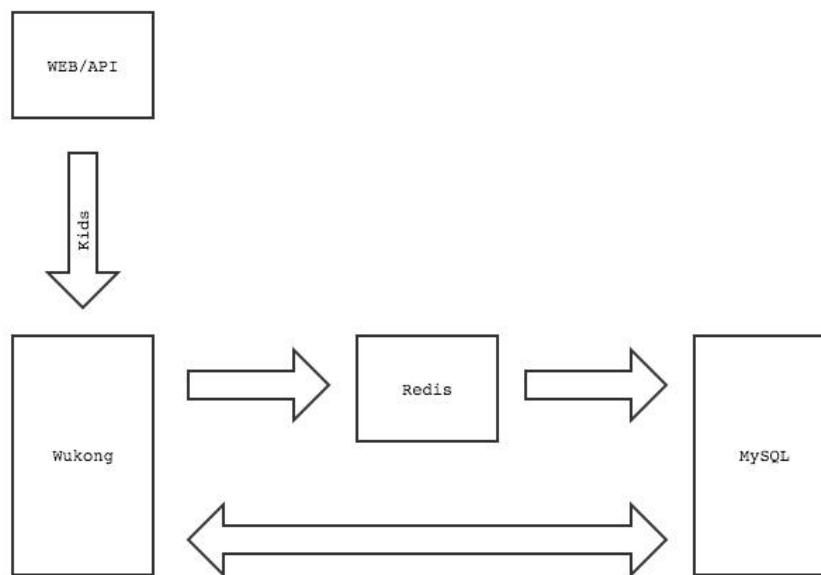
WUKON
G



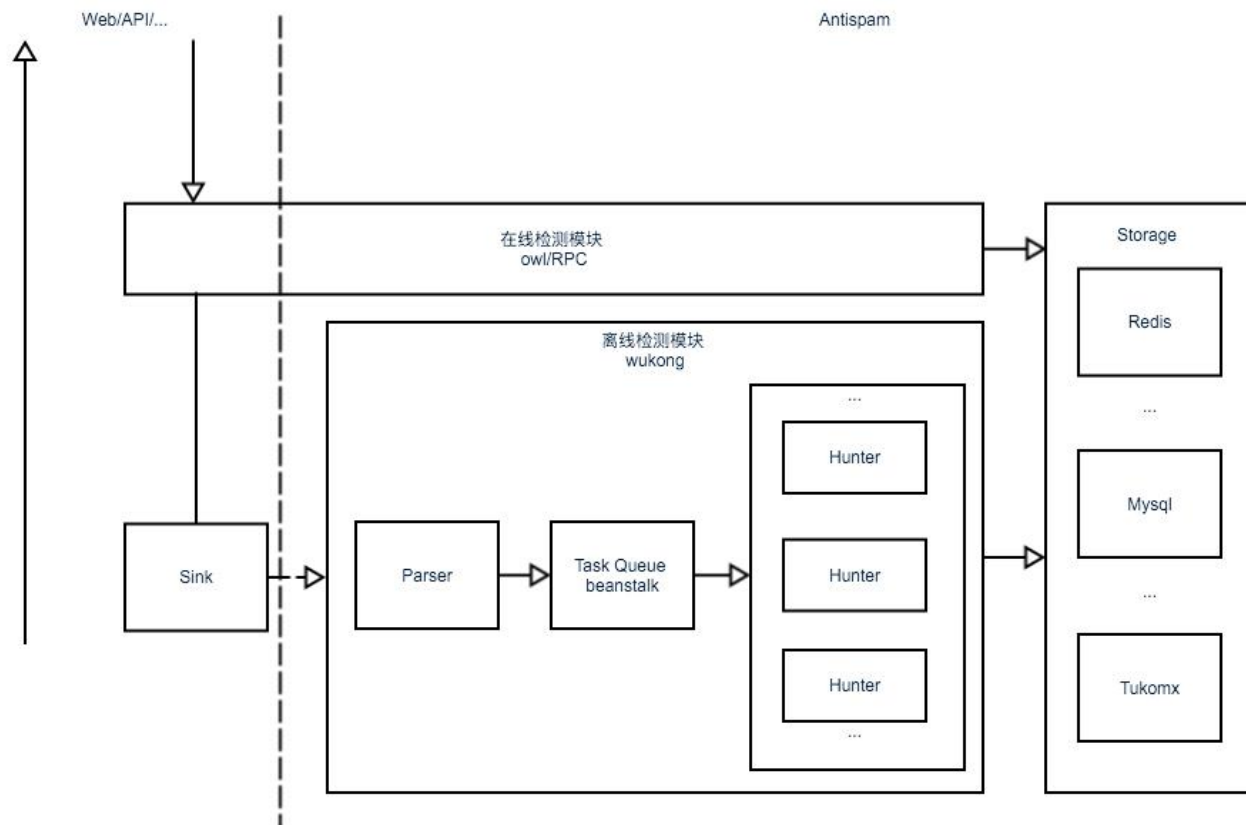
WUKONG

WUKONG VO

- 技术栈: Redis, MySQL, Kids
- 数据流: 串行处理

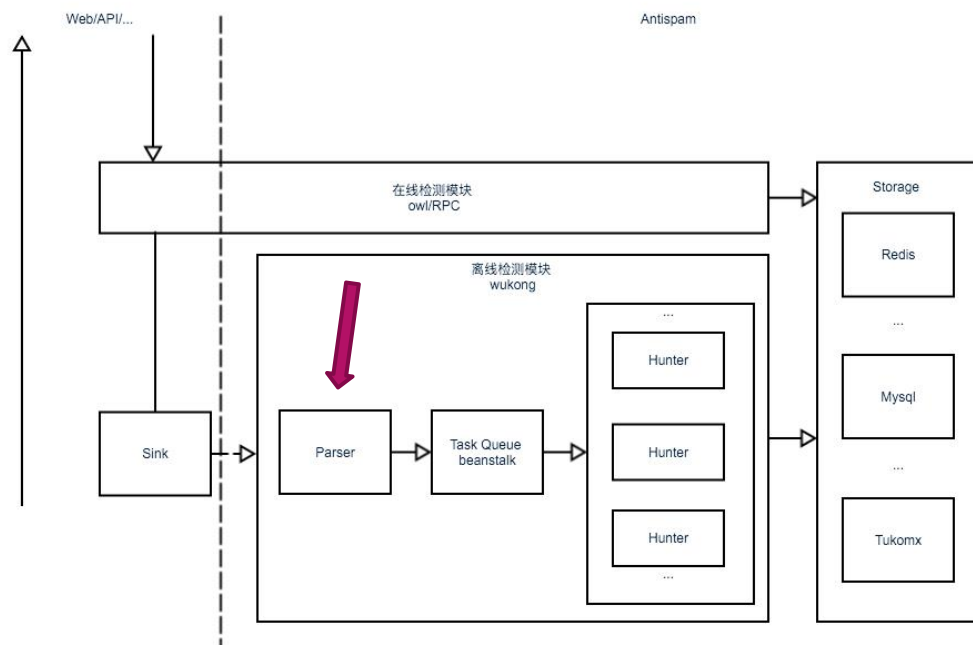


WUKONG V1



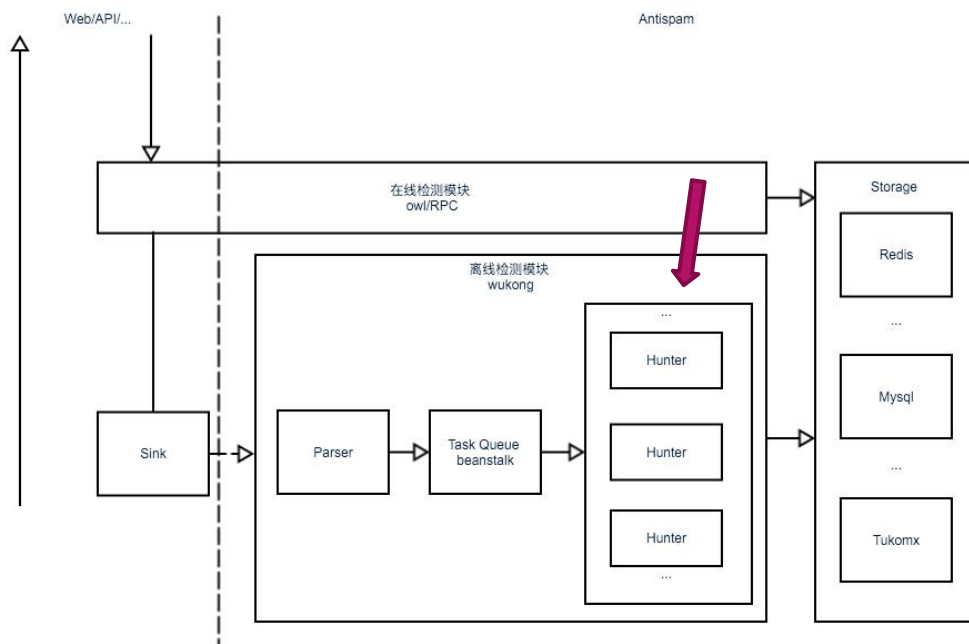
WUKONG V1 —— Parser

- 谁、什么时间、什么地点、做了啥、产生了什么内容、对谁？



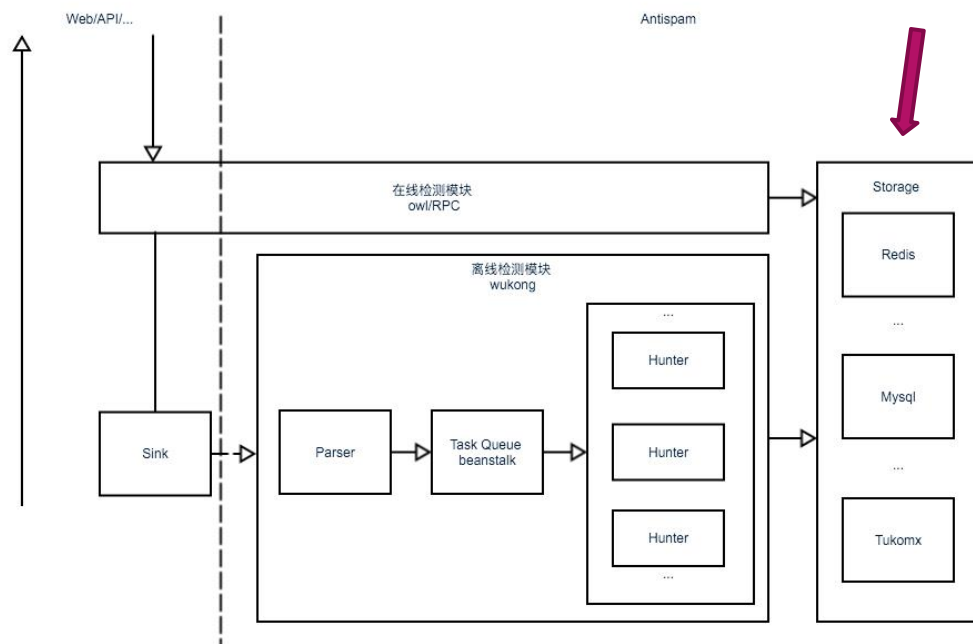
WUKONG V1 —— Hunter

- 特征扩展 —— 横向扩展、纵向扩展
- 规则描述 —— 强化表达能力
- 策略组成 —— 平滑上下线
- 策略组成 —— 平滑上下线



WUKONG V1 —— Storage

- 缓存 —— 将关键 RPC 结果缓存到 Redis 中
- 使用 tukomx 存储事件 —— 对各纬度建立索引



WUKONG V1 —— 策略

- Item —— 原始数据项
- Trigger —— 作用于数据项的逻辑函数
- Action —— 触发 Trigger 之后的操作
- 利用动态语言特性

81

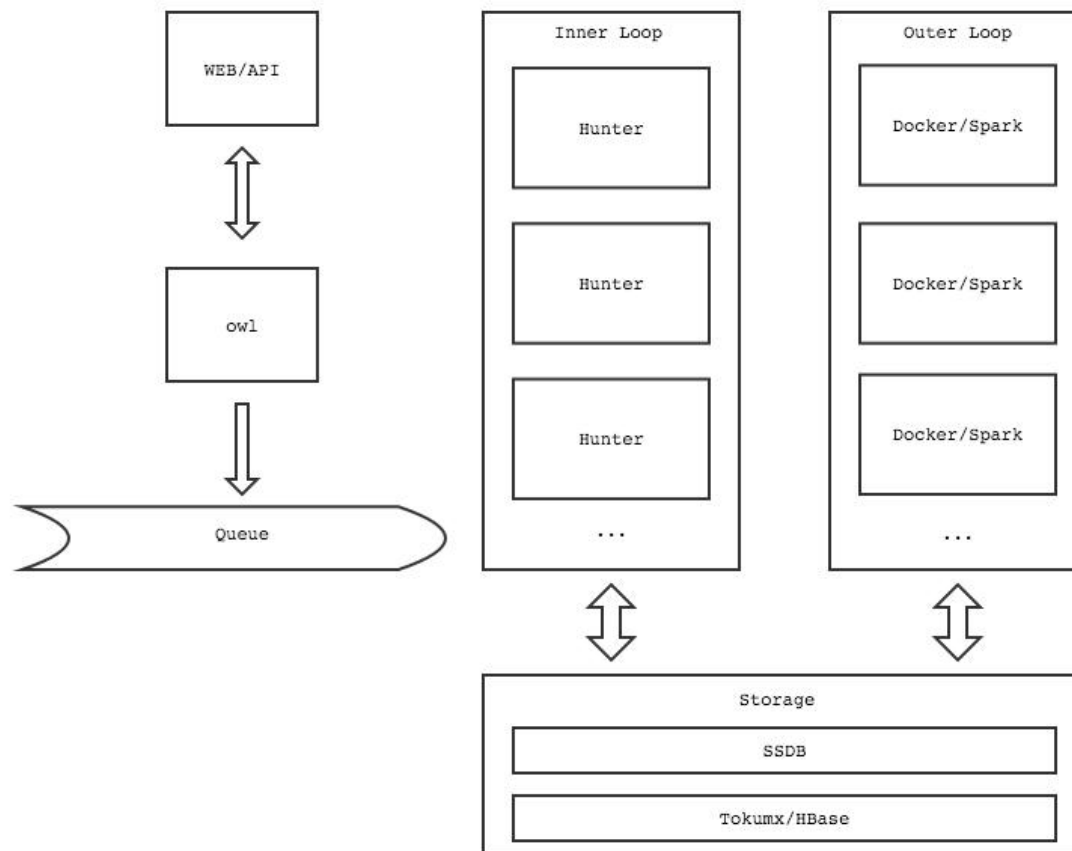
question_oneday_ipc_register

user_auto_demote() and action == "QUESTION_CREATE" and email_domain(pre_ip_actions(brother_actions(recent_same_actions(60), 1)))

delete_all(email_domain(pre_ip_actions(brother_actions(recent_same_actions(60), 1))),
ban_all(email_domain(pre_ip_actions(brother_actions(recent_same_actions(60), 1))),
forbid_all_ips_question(email_domain(pre_ip_actions(brother_actions(recent_same_actions(60), 1))), 2*24*3600, "81"),

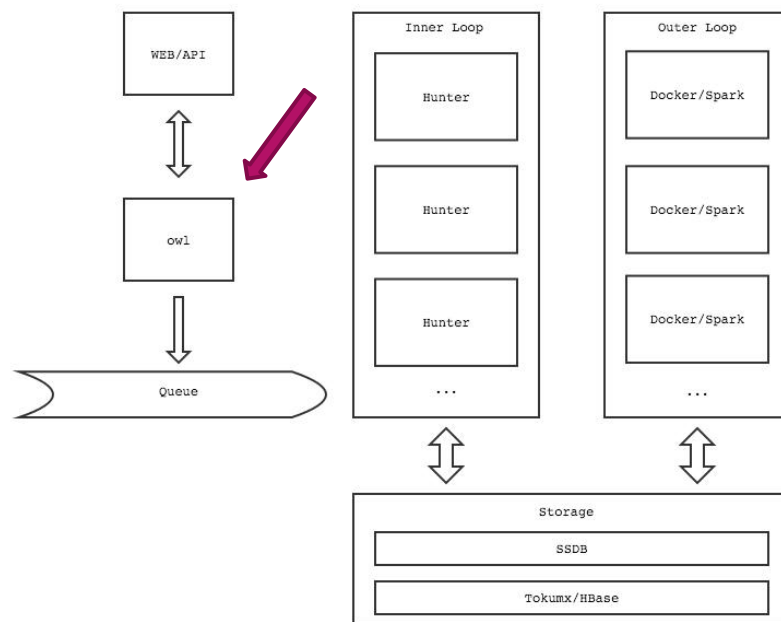
Update

WUKONG V2



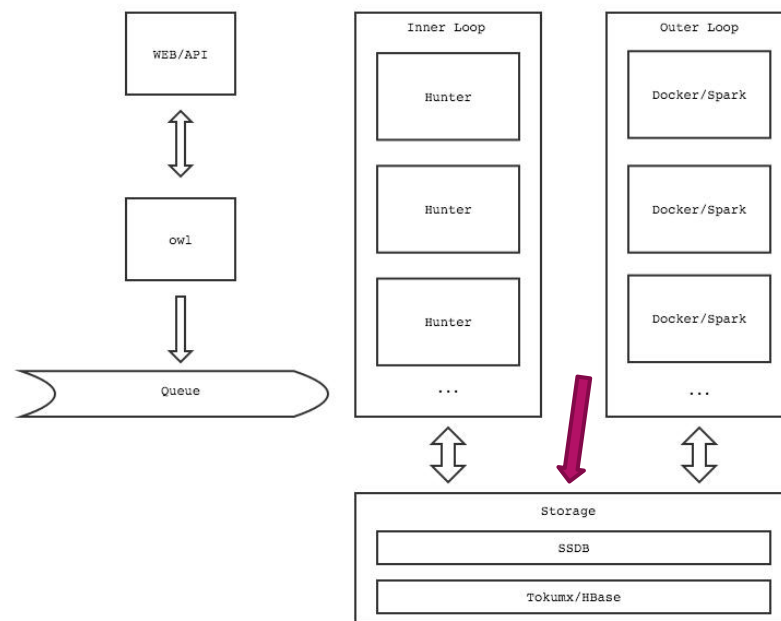
WUKONG V2 —— Owl

- 为了简化业务接入反作弊系统的复杂性
- 为了能够将整个处理流程的数据串联起来
- 为了能够实现反作弊系统开关功能



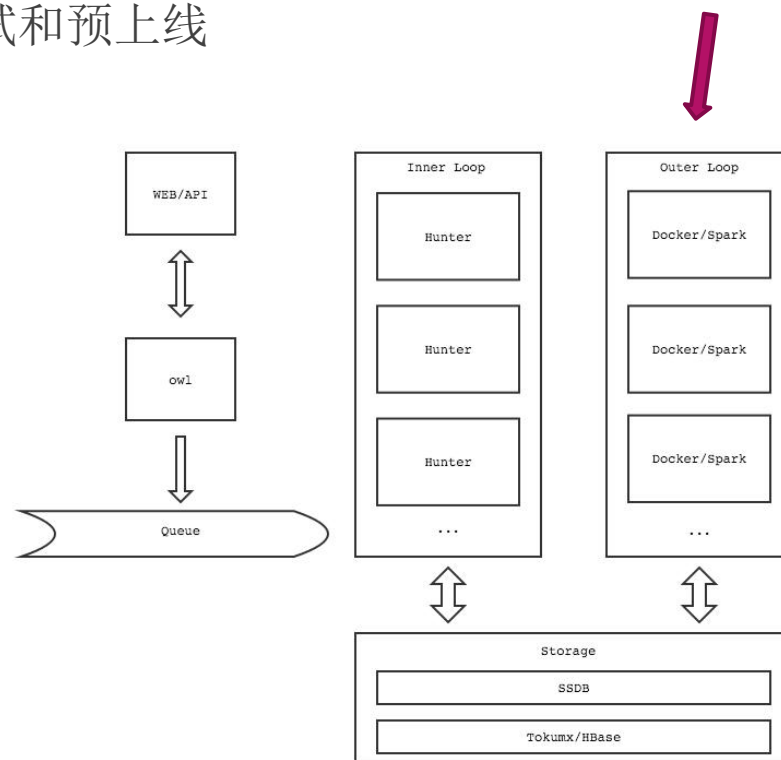
WUKONG V2 —— Storage

- 使用 SSDB 为策略建立倒排索引
- 同时使用 SSDB 缓存大量的中间计算和请求结果，减少重复计算和 RPC 请求
- 减少对于 Tokumx 的读取场景，逐渐变更为写多读少的场景 —— HBase



WUKONG V2 —— Outer Loop

- 需求：为了能够清理策略上线之前，出现的对应类型作弊
- 工程：为了能够做一些比较耗时，或者是需要更多数据量的策略
- 工程：为了能够支持策略变更的测试和预上线



WUKONG V2 —— 策略

- 简化策略的书写，参考函数式语言
- 更容易基于策略语言做优化

227

test_answer_4

```
recent_same_type_events(180, n=10, accept_id=accept_id).filter("register_time", same_day=True).filter("email_domain").length >= 1
```

recall(cache)

Update

总结

架构是为了解决特定的问题

总结

架构师更需要的是克制

总结

造轮子？请别占用公司资源！

总结

小团队的任务在于组合优质服务



经验

数据主导



经验

疏导 vs 对抗



Thanks

