

# 移动广告平台的大数据实践

----- 汪巍@多盟

## 移动广告平台的大数据实践

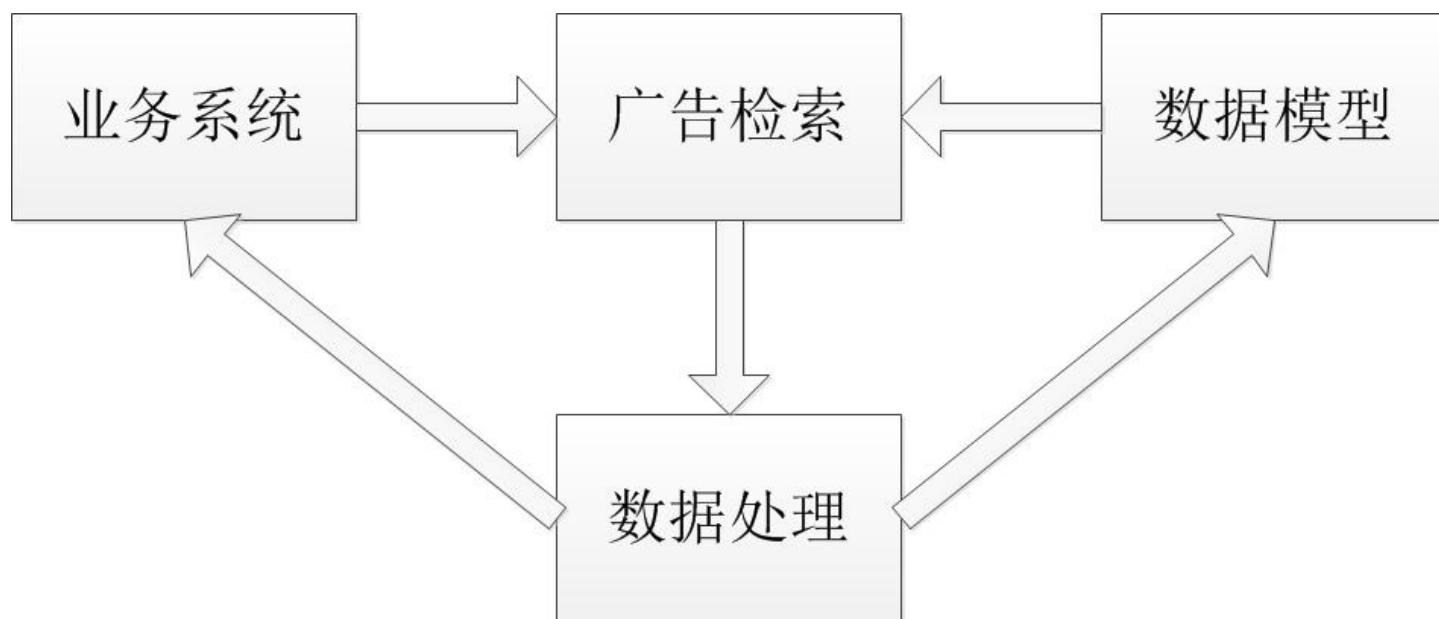
- 09年毕业于哈尔滨工业大学
- 09.7 加入百度有啊
- 12.2 加入多盟

# 移动广告平台的大数据实践

- 背景
  - 多盟的角色转变 ADN→DSP
  - 数据量级从ADN的5亿/天增长到RTB的50亿/天

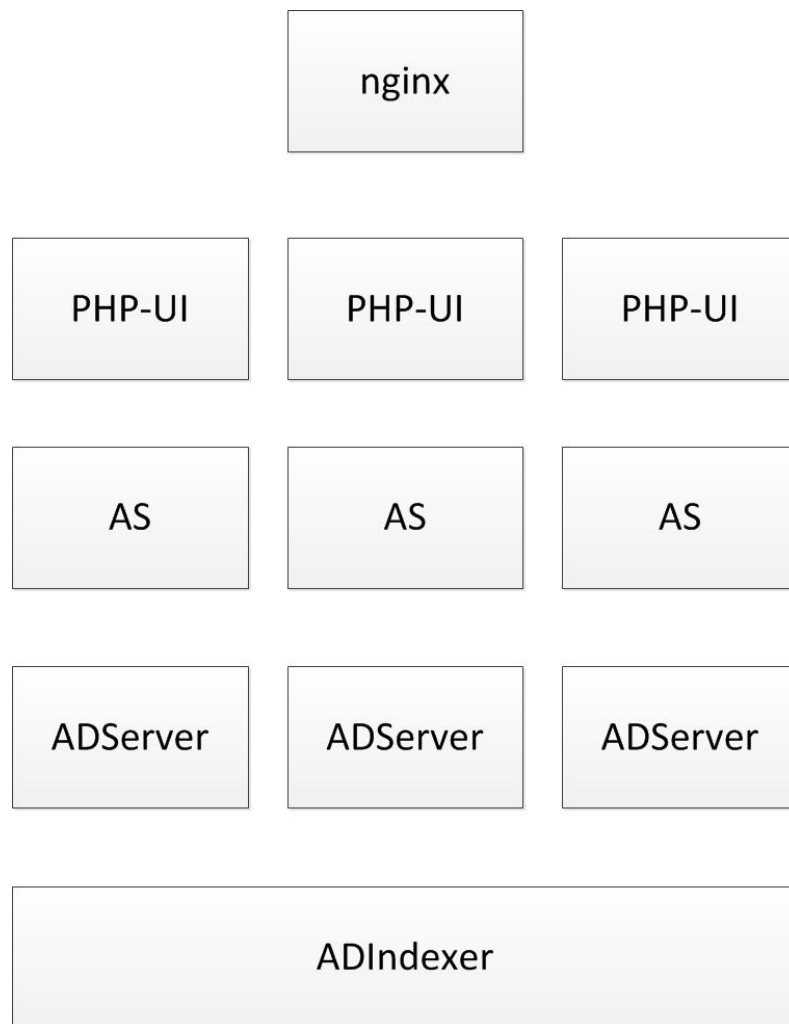
## 多盟的架构变迁

- 整体结构



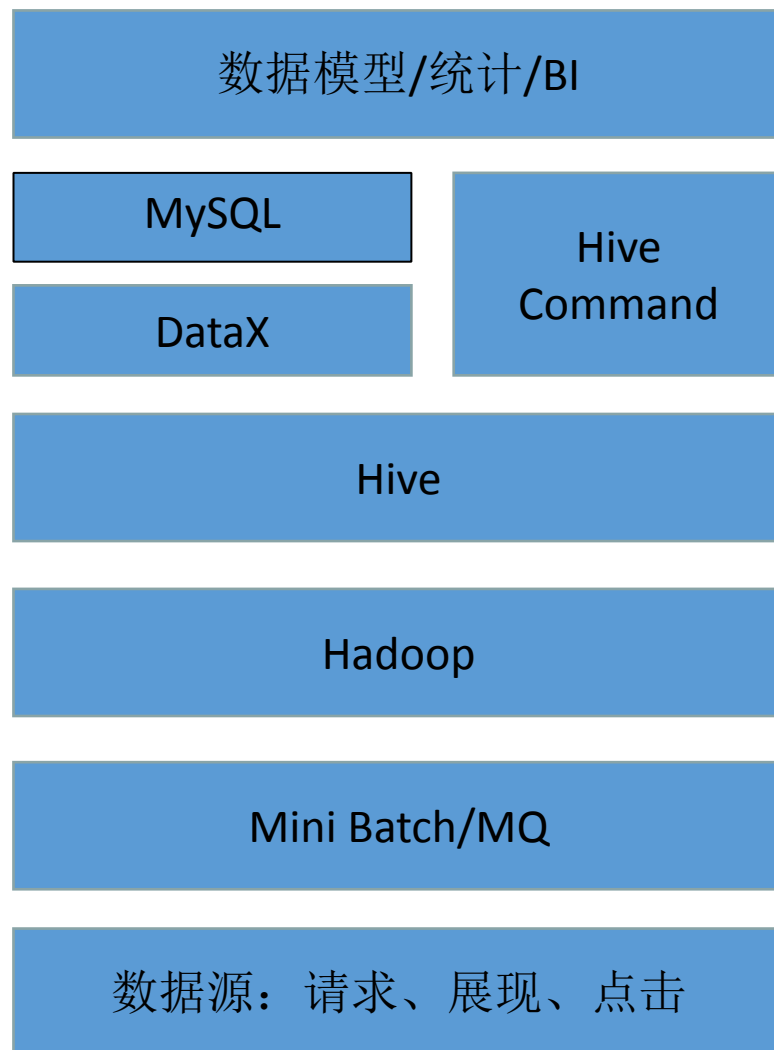
## 多盟的架构变迁

- ADN时代的广告检索
  - PHP-UI 性能 or 效率
- AS的过度设计
- 广告超投



## 多盟的架构变迁

- ADN时代的数据处理
  - HDFS/HIVE/DATAX/MYSQL
  - 多条数据流难以运维
  - 没有统一的数据处理框架
    - logtailer
  - Hadoop版本老旧
    - HiveServer+HA
    - 数据均衡
  - 小时级数据/模型
    - Join效率低下
    - 数据盘ext3→ext4
  - Domino
  - Thrift



# 多盟的架构变迁

## • Domino

Domino

Home

任务列表

任务状态

PLAN状态

Logout (wangwei@domob.cn)

Inf Family ▾

Home / Plans / hive\_load\_and\_insert\_clk\_total / viewJob

### 任务 #592

ID	592
名称	hive_load_and_insert_clk_total
任务状态	正常运行
命令行	cd /data/etl/bin && bash -x hive_load_and_insert_clk_total.sh
运行周期	小时
优先级	中
重试最大次数	1
重试时间间隔(s)	0
最长运行时间(s)	1小时 (3600s)
最长调度时间(s)	1小时 (3600s)
延迟时间	0 (0s)
用户	search
运行主机	bidw.domob.cn
创建时间	2013-12-17 11:15:28
最后启动时间	2013-12-17 11:15:28
报警联系人	wangwei@domob.cn, wangwei@domob.cn
报警方式	邮件 + 短信
维护人	wangwei@domob.cn
创建人	wangwei@domob.cn
描述	将新的小时级点击写入clk_total表 54321
最后任务	2016-05-13 18:00

Operations  
[返回任务](#)  
[查看执行情况](#)  
[查看执行失败](#)  
[重跑本次执行](#)  
[批量标注](#)

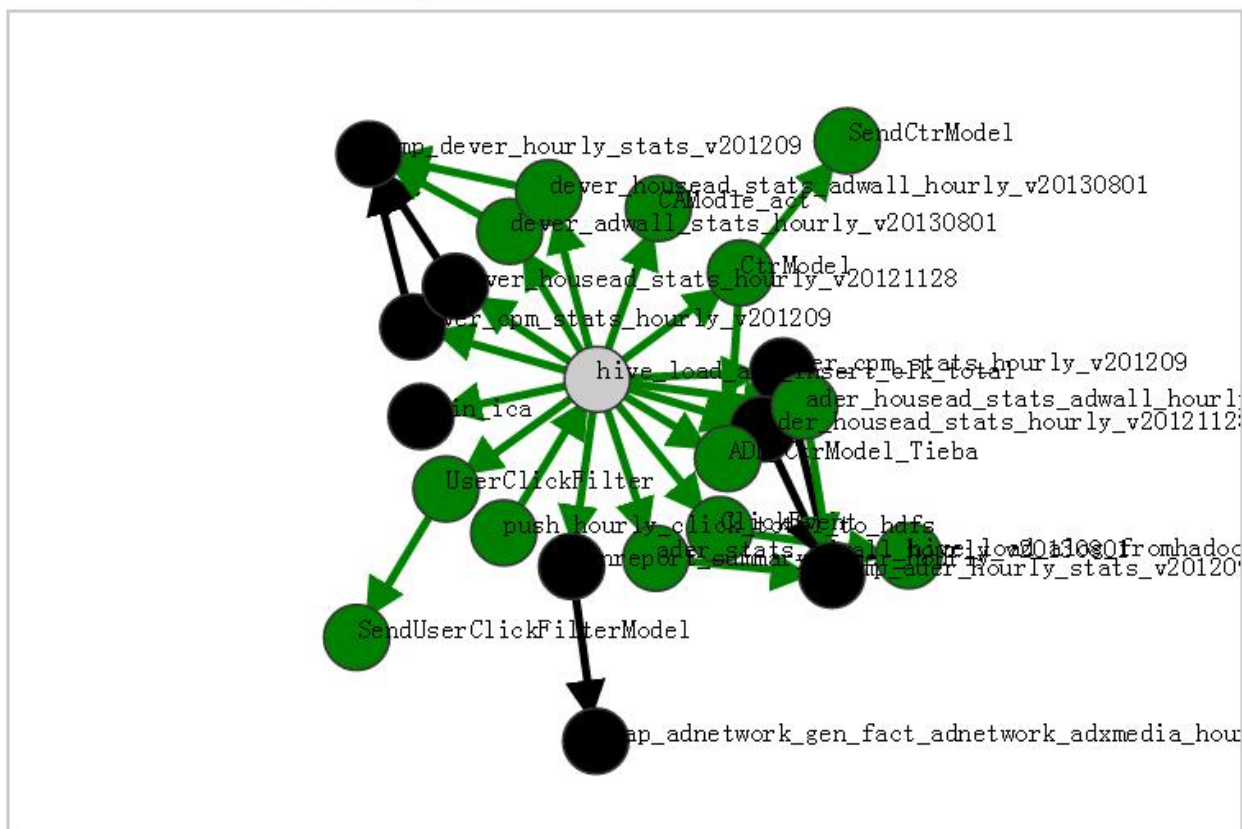


## 多盟的架构变迁

- Domino

## 任务依赖执行状况

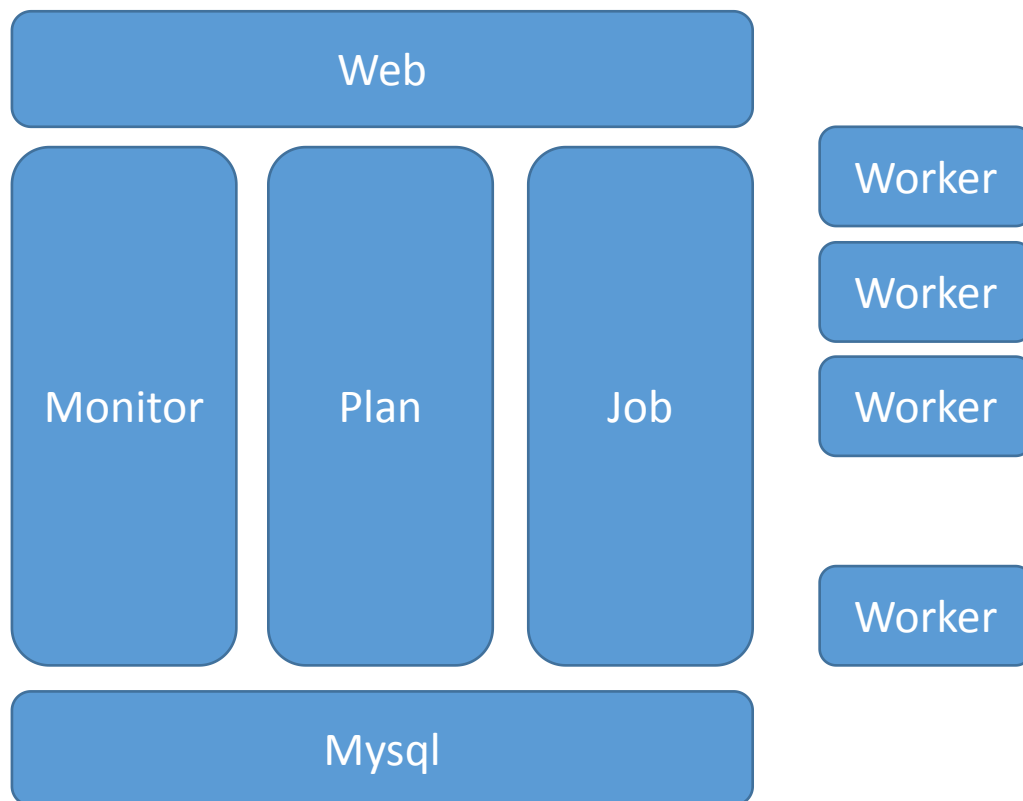
任务颜色标识: 新建 运行 成功 失败 重跑





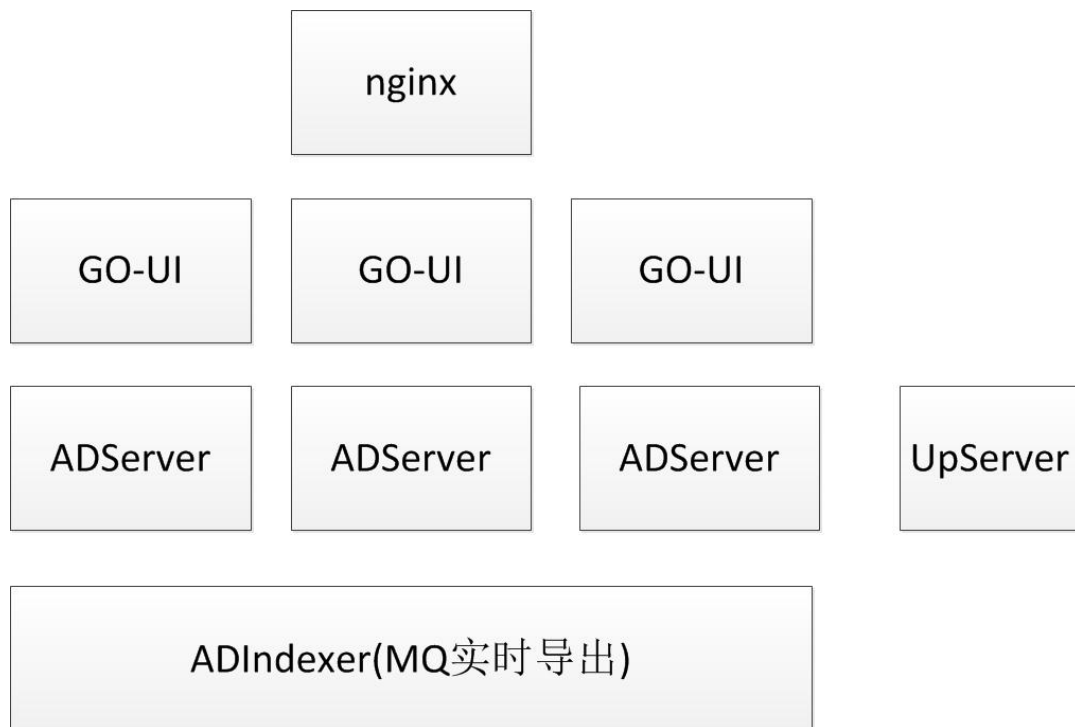
## 多盟的架构变迁

- Domino



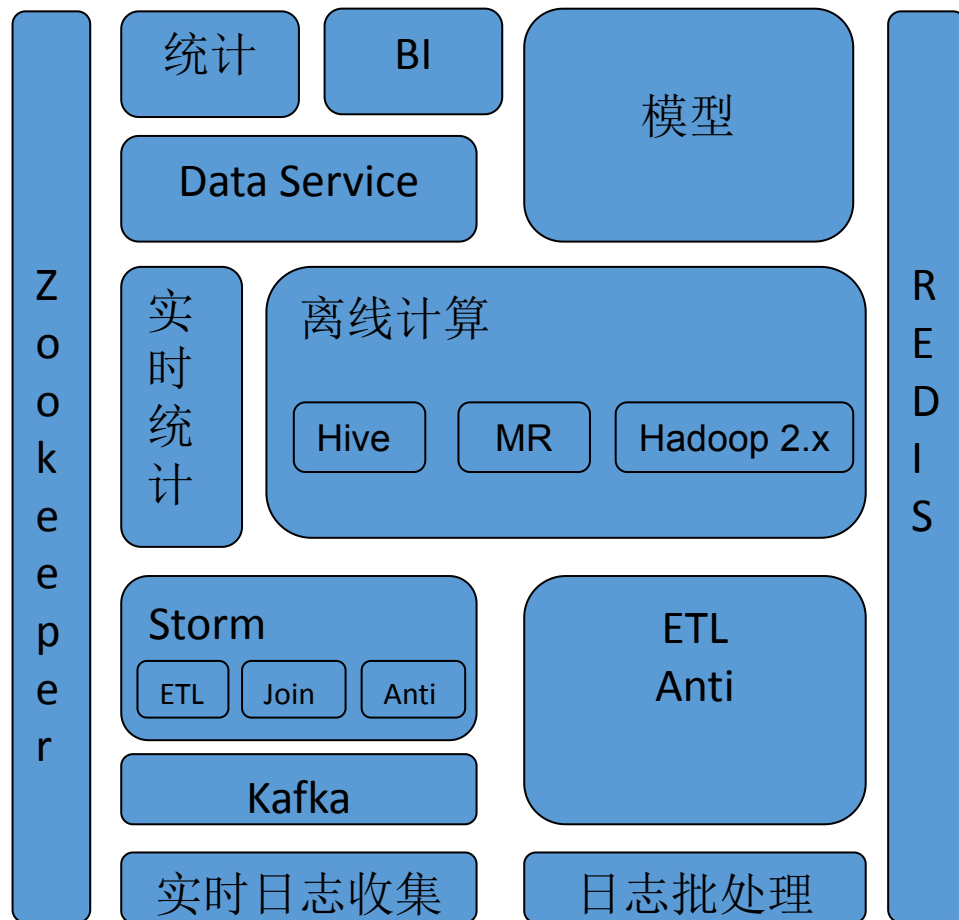
## 多盟的架构变迁

- RTB时代的广告检索
  - 50亿广告请求
- PHP→GO
  - 性能巨幅提升
  - GC
- AS+UpServer
- 广告库实时导出



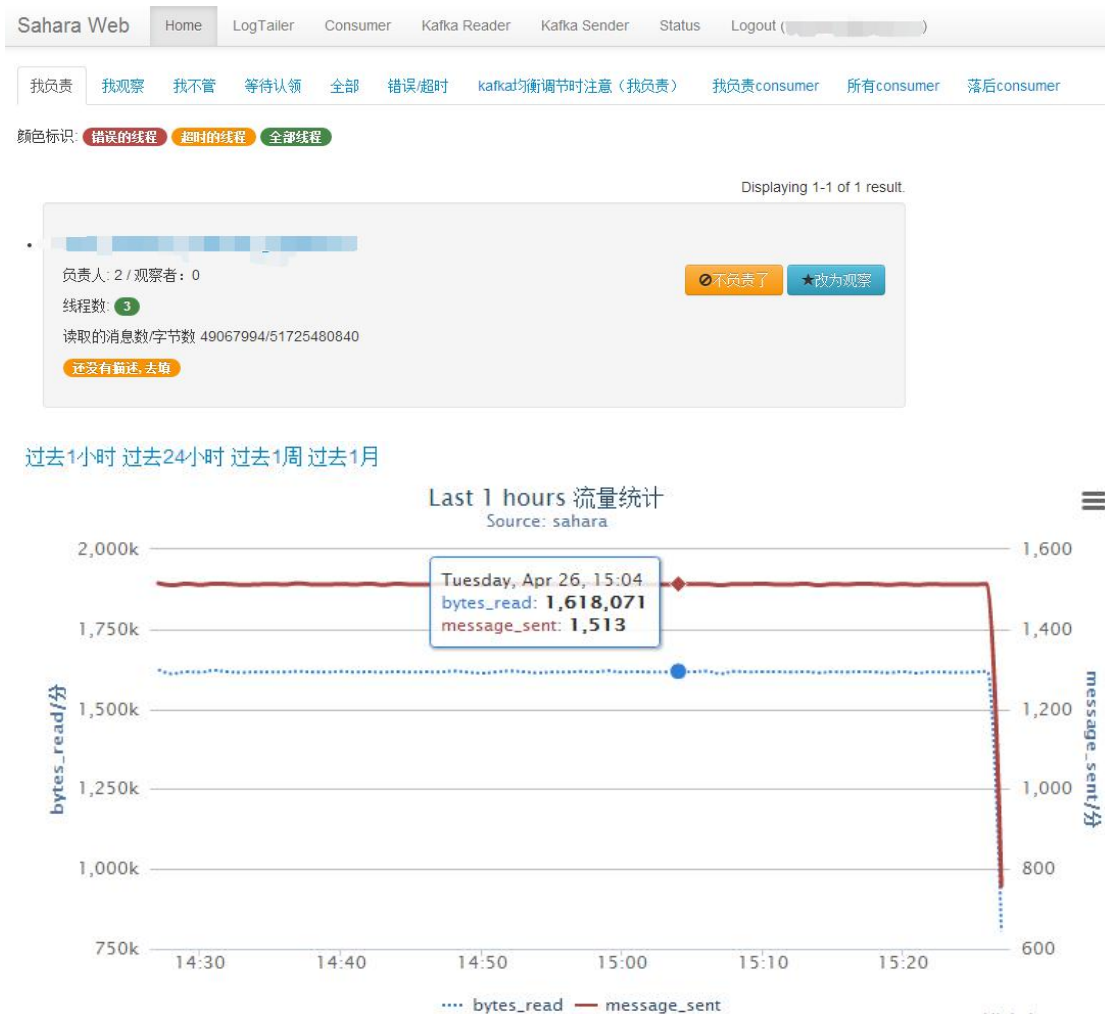
## 多盟的架构变迁

- RTB时代的数据处理
  - Logtailer & Sahara
  - Kafka 0.7.x → 0.8.x
  - Storm
  - Hadoop 0.25 → 2.1
  - Redis-cluster
  - 机房搬迁



## 多盟的架构变迁

- Kafka
  - 数据不均匀
  - 动态扩容不友好
- Sahara
- Kafka-manager
- Ganglia



## 多盟的架构变迁

- Storm Cluster

- 调试不易 ----- Jmock
- Worker数量不均匀 ----- 升级开源版本到0.9.4
- 日志难以查看 ----- logback, nfs
- 日志开销大 ----- 1.5ms→0.02ms
- 消息堆积 ----- ack+pending
- 带宽占用大 ----- local分组

## 多盟的架构变迁

- Storm Cluster

- 负载不一致 ----- 负载均衡
- 事务 ----- trident
- 多语言 ----- json->protobuf
- 监控
  - 日志
  - 拓扑状态
  - 进程
  - 外部组件
  - 业务数据

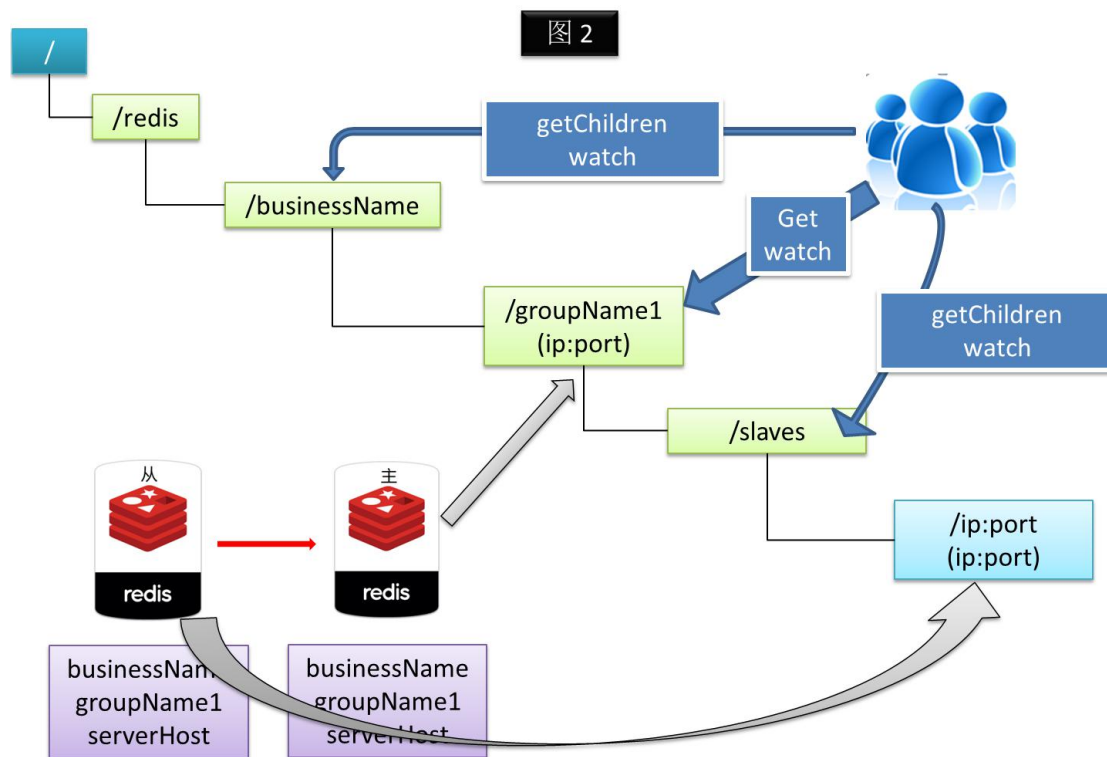
## 多盟的架构变迁

- Redis Cluster
  - 突破单机物理内存限制
  - 合理利用CPU
  - 管理工具
  - 主从切换
  - Proxy or Clients?
  - ZK & Sentinel



## 多盟的架构变迁

- Redis Cluster



谢谢