QCon 全球软件开发大会
INTERNATIONAL SOFTWARE DEVELOPMENT CONFERENCE

BEIJING 2017

阿里实时风控引擎实践

阿里巴巴 胡四海

主办方 Geekbang 极客邦科技 InfoQ

Alibaba Group 阿里巴巴集团

胡四海（知命）

2010 年加入阿里，现负责阿里巴巴集团安全部：风控引擎、人机对抗技术。

在阿里就职期间，一直从事风控相关领域的研究与风控引擎的开发，建造阿里多项核心风控产品。

在风控领域拥有 9 项技术发明专利，遍布风险防控各个领域。

Alibaba Group
阿里巴巴集团

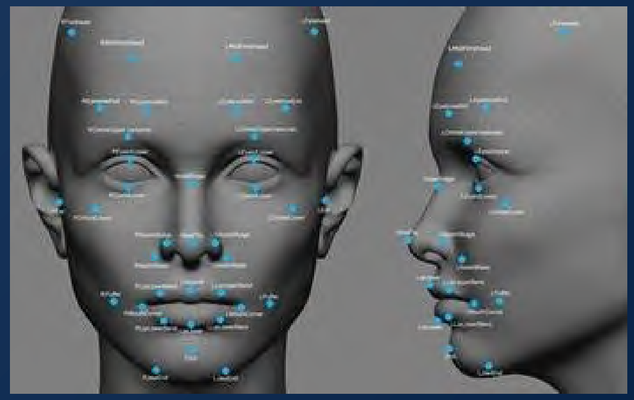
风控引擎是什么?

一些场景:

红包雨

人机识别

支付风控



活体识别

Alibaba Group 阿里巴巴集团

传统安全
是以系统为中心

互联网时代

全互联网已泄漏个人账号超过20亿条
覆盖全互联网账号的40%以上

Alibaba Group
阿里巴巴集团

绑架P2P平台？解密疯狂的“羊毛党”

第一财经日报 安卓 2015-04-21 06:00:00

“羊毛党越来越普遍，我们2010年平台成立的时候，全国P2P平台还不是那么多，所以羊毛党未成气候，这两年，随着P2P平台的大量涌现，为了吸引眼球，P2P平台只能通过各种优惠活动获客，羊毛党随之而生，可以说，羊毛党是P2P行业竞争加剧过程中衍生的一个群体。”

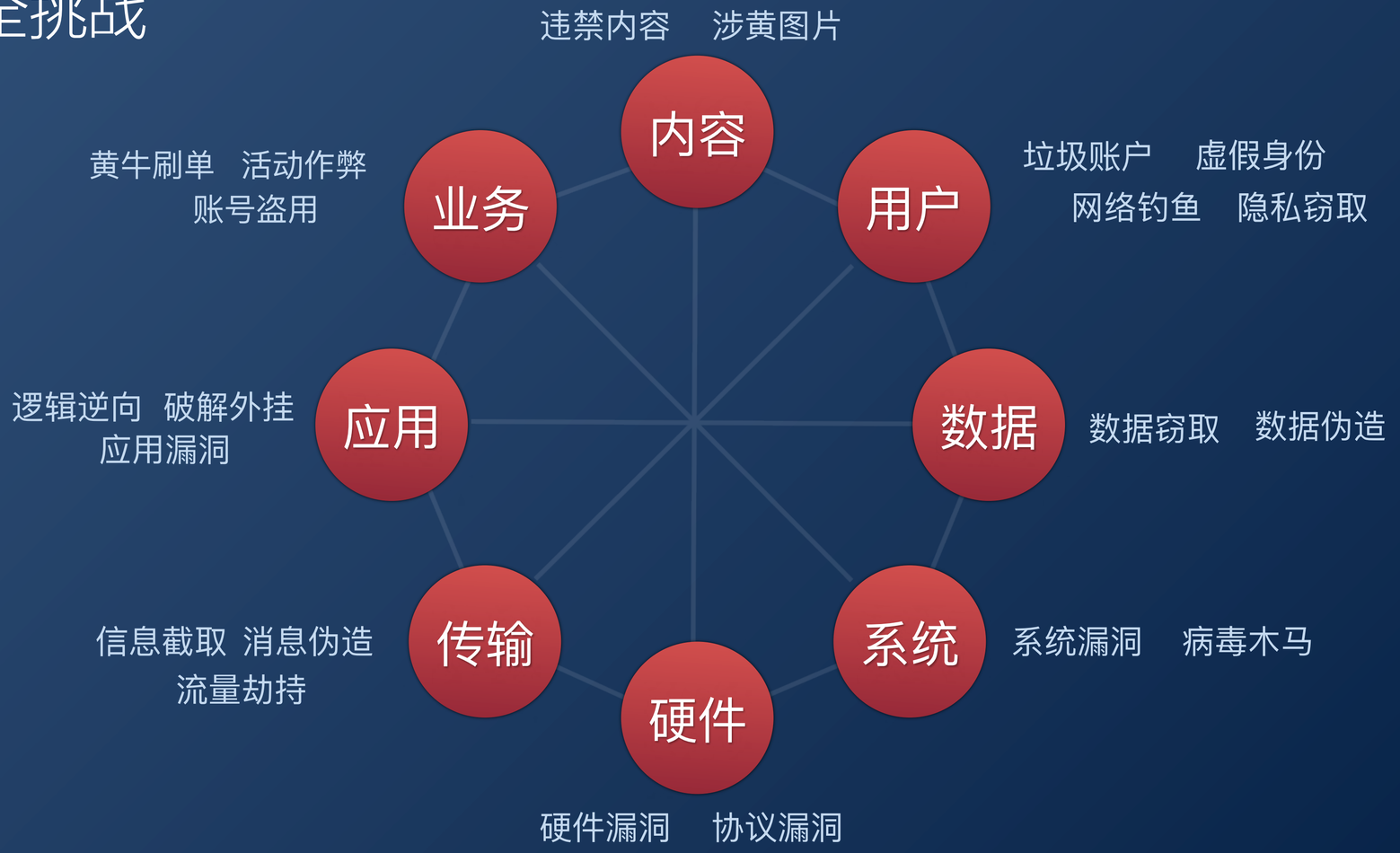
评论 0 | 赞 0

近日，广州一小型P2P平台的投资者向媒体爆料称，该平台目前已经出现提现困难，平台的法人代表失联，其位于广州珠江新城的办公室已经被搬空。

《第一财经日报》记者发现，这家名为“民间财富”的平台已经无法注册新会员，其最后一次发标是在3月30日。平台公告显示，该平台成立于2014年中旬，目前注册会员534人，待收总金额59.24万元。

可以说，这是一家上线不久，且交易金额很小的平台，然而，就在这个小平台处于瘫痪状态的背后，一个独特的投资人群体浮出水面，业内称之为“羊毛党”。在广州一网贷业内人士看来，这个平台目前出现的问题与“羊毛党”有很大关系。

DT时代的安全挑战



Alibaba Group 阿里巴巴集团

全链路
防护体系

Alibaba Group
阿里巴巴集团

端到端的风险防控

Alibaba Group
阿里巴巴集团

攻防布局

事前

通过积累的黑产数据，在行为发生前直接屏蔽。
发违规产品、发广告贴前进行限制。限制黑产团伙领取红包权限。
等等 ...

事中

用户登陆时检测是否帐号被盗。下单时检测是否存在欺诈风险。
订单评论时检测是否是垃圾广告。是否羊毛党领取红包。
等等 ...

事后

产品发布上线后进行离线扫描排查，离线模型全量扫描欺诈会员。
羊毛党红包套现。
等等 ...

全链路

实时防控案例

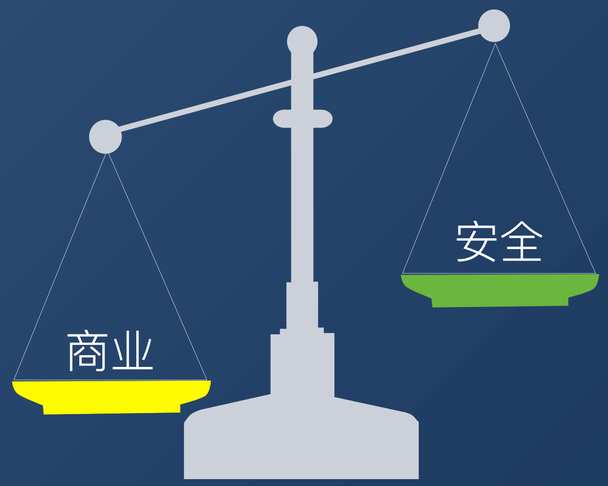
红包雨

百万QPS，毫秒级响应要求，有限的防控数据

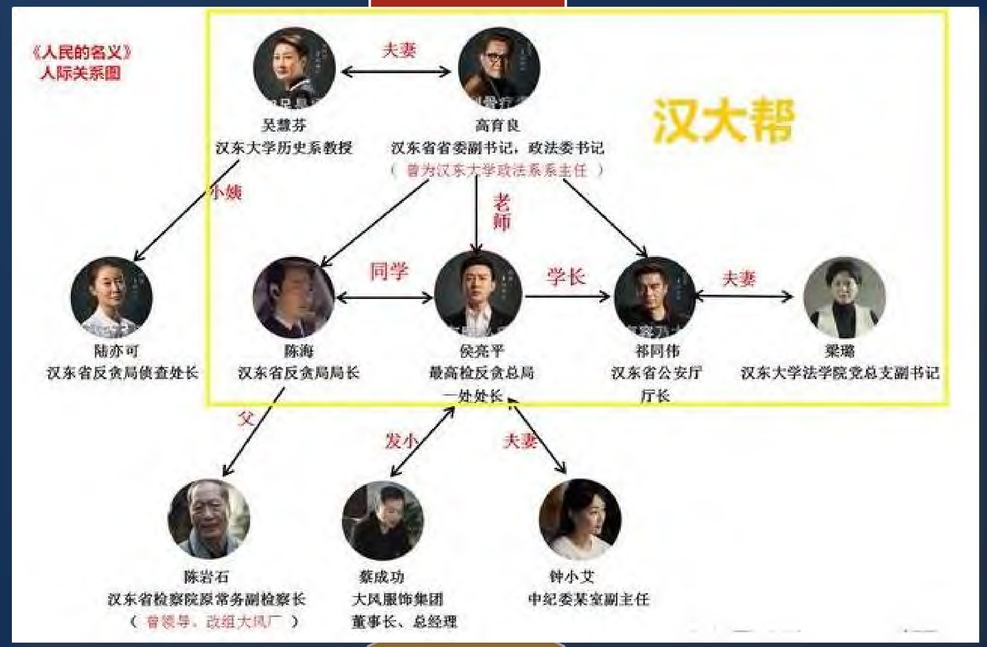
防控的几个挑战

① 商业对风控系统的挑战
安全对业务的侵入
实时数据补全性能要求
② 黑产优势
权限优势
时间优势
③ 用户体验
平衡用户体验
透明与原因外化
④ 不可抗力挑战：异地容灾
⑤ 监管要求

挑战：商业对风控系统的挑战



用户画像



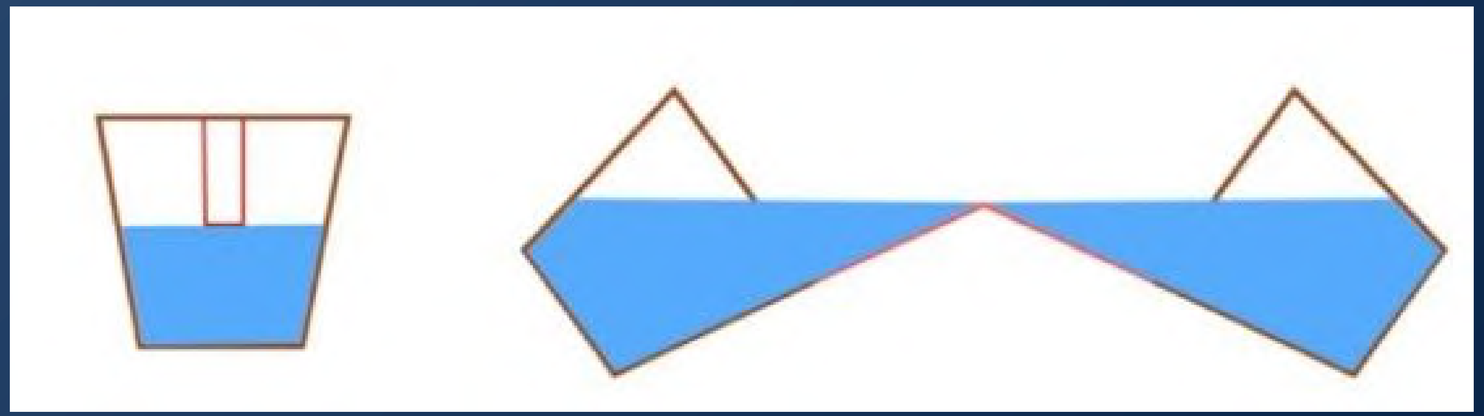
...

没有经历过商业挑战的风控，是不完整的风控

3 N：NO更多数据，NO更长识别时间，NO嵌入业务流程

Alibaba Group 阿里巴巴集团

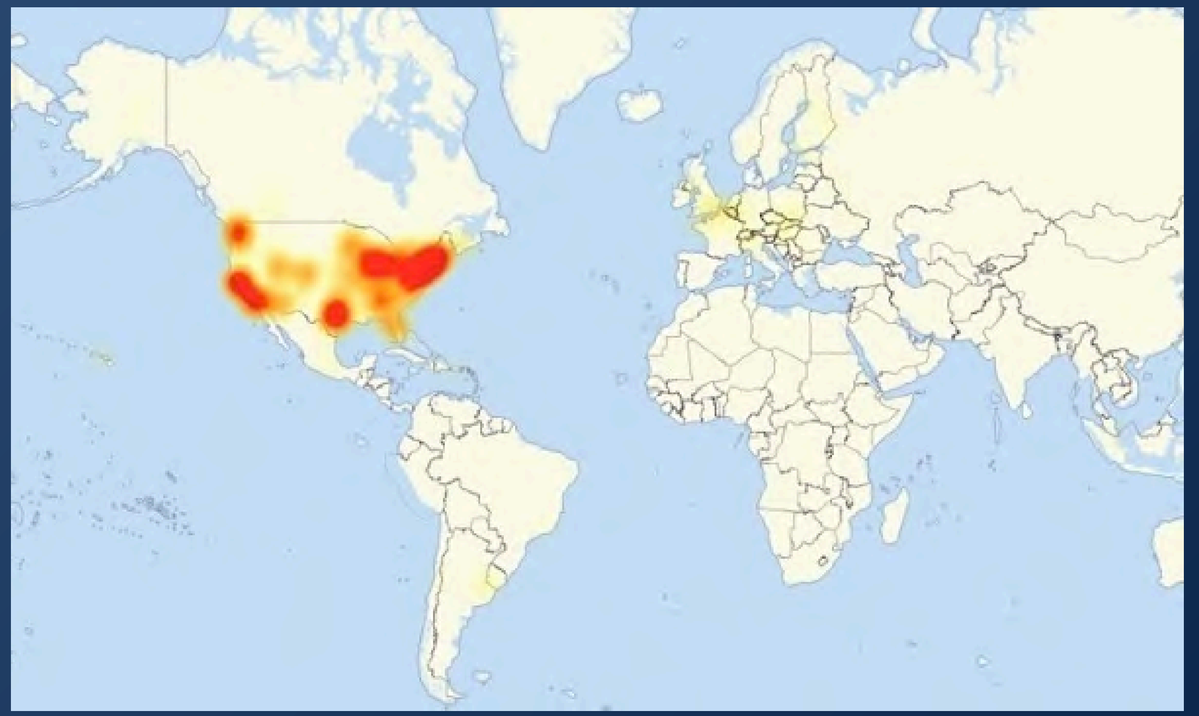
挑战：黑产权限优势



短板理论

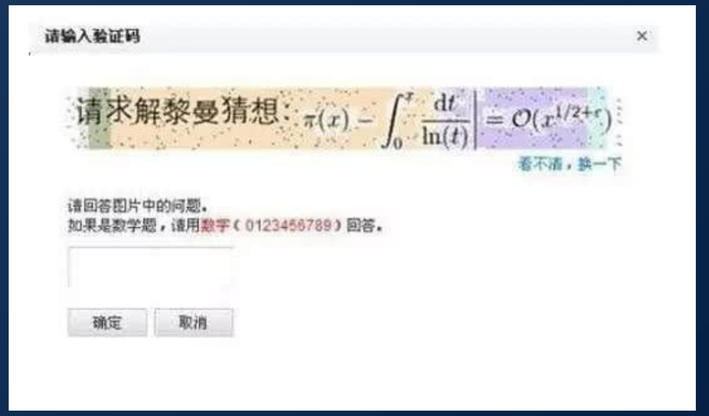
长板理论

挑战：黑产时间优势



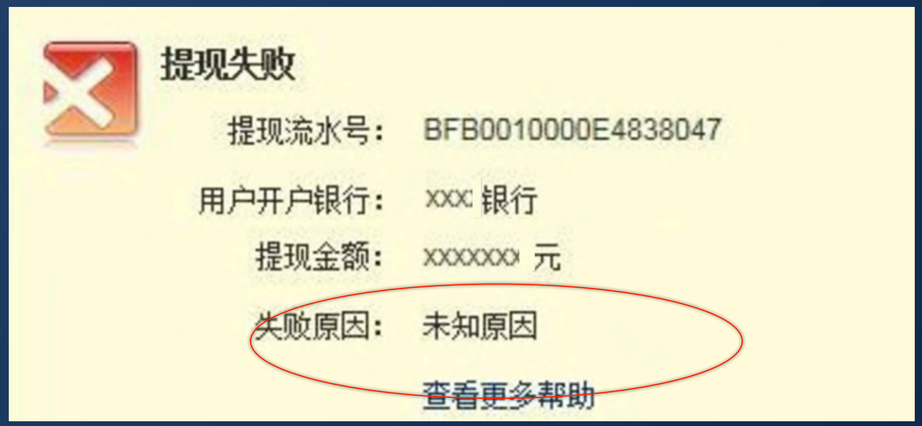
2016.10 半个美国网络瘫痪

挑战：用户体验



Alibaba Group 阿里巴巴集团

挑战：原因外化



Alibaba Group 阿里巴巴集团

挑战：异地容灾

挖掘机技术哪家强

挑战：监管要求

俄罗斯通过互联网新法 公民个人数据必须存在国内服务器上

分享到：

2014-07-07 17:13:59　　字号： A- A A+　　来源：路透社等

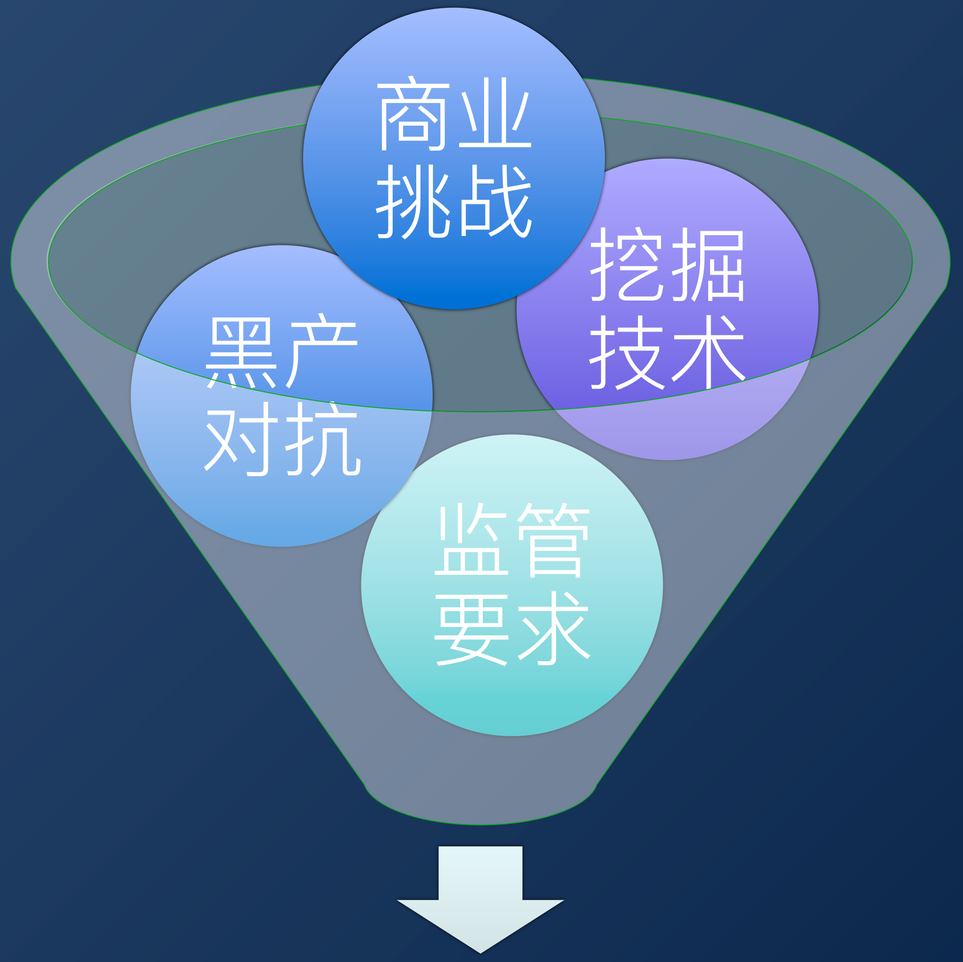
俄罗斯一项最新法律规定，所有收集俄罗斯公民信息的互联网公司都必须将这些数据存储在俄罗斯国内的服务器上。克里姆林宫称这是为了保护数据安全。

俄罗斯新法案：公民数据只能存储在国内服务器

俄罗斯国家杜马7月4日批准了这项法律，其生效时间为2016年9月1日。法律起草者认为，这将给俄罗斯国内和国外的互联网公司充足的时间，在俄罗斯国内设置存储设备。

Alibaba Group 阿里巴巴集团

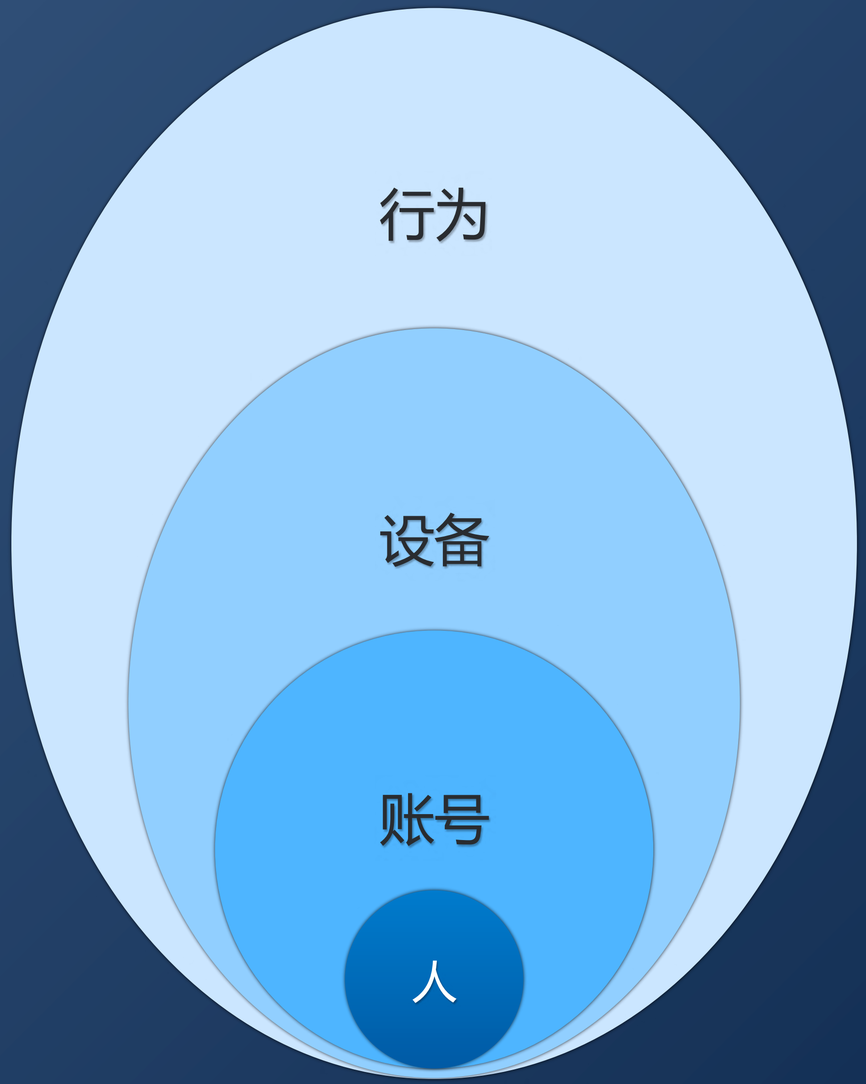
挑战



实践经验

1. 经验一：风控的本质是成本之战
2. 经验二：足够的实验与验证 ——磨刀不误砍柴工
3. 经验三：熔断与兜底
4. 经验四：风控是一个业务决策

经验一：成本



经验：尽量围绕固有属性展开防控

成本：行为 > 设备 & 账号 > 自然人

行为分析：读时计算

如何快速进行实时行为分析?

同用户，24小时内，金额>=10，中奖次数

计算方法

- *COUNT：计数*
- *DISTINCT_COUNT：去重计数*
- *SUM：求和*
- *AVG：平均值*
- *SQUARE：平方和*
- *VARIANCE：方差*

```
SELECT userId, COUNT (*) AS hitCount
FROM 活动中奖
WHERE amount >= 10
GROYP BY userId, time(HOUR, 24);
```

读取值：统计项

主键：统计对象/维度

分区设置：按买家、时间片分区

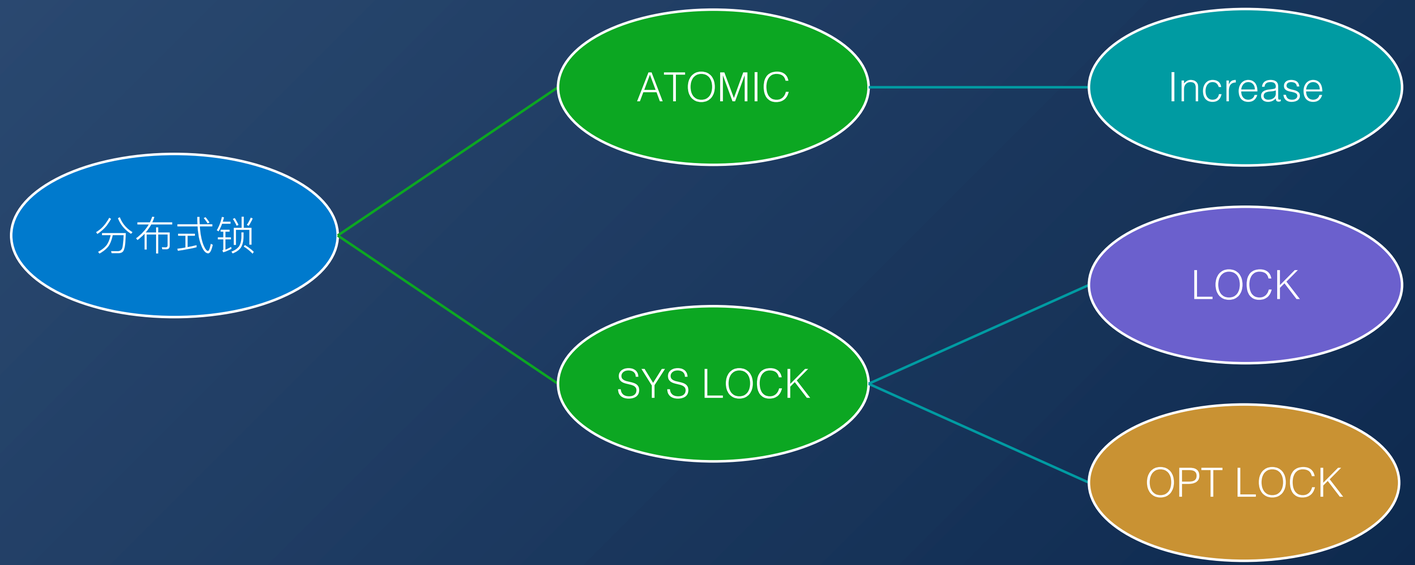
问题? IO问题，重复计算问题 → 慢

行为分析：写时计算

同用户，中奖次数

用KV来实现：**key** = *userId*, **value** = *hitCount*

问题：分布式场景，如何解决原子性问题?



行为分析：写时计算

同用户，**24小时内**，中奖次数

分片：| t-23 | t – 22 | t – 21 | … | t -1 | t |

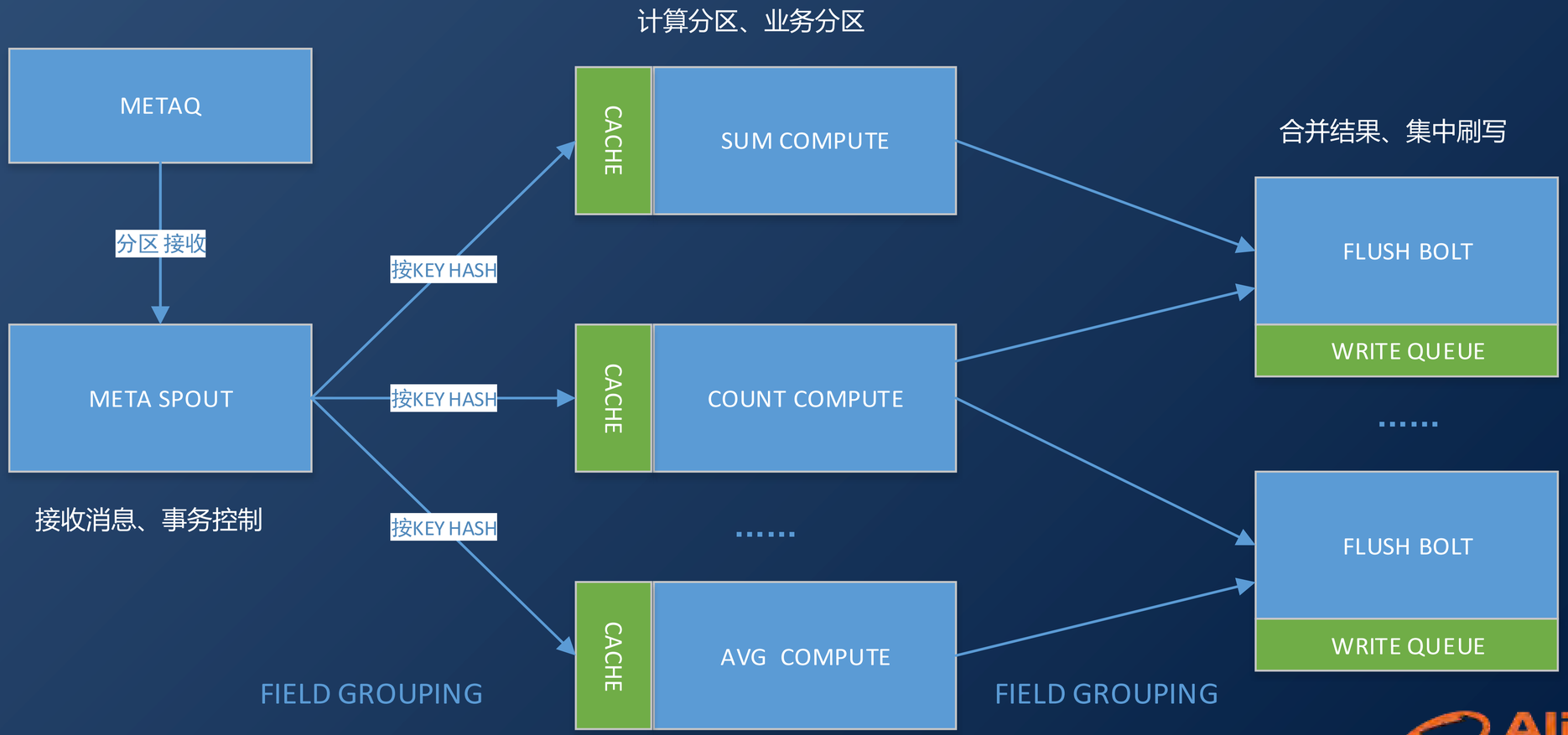
同用户，24小时内，金额>=10，中奖次数

?

行为分析：写时计算

流式计算： storm, spark, samza, flink, blink …

基于JStorm的一种实时流计算实现：



Alibaba Group 阿里巴巴集团

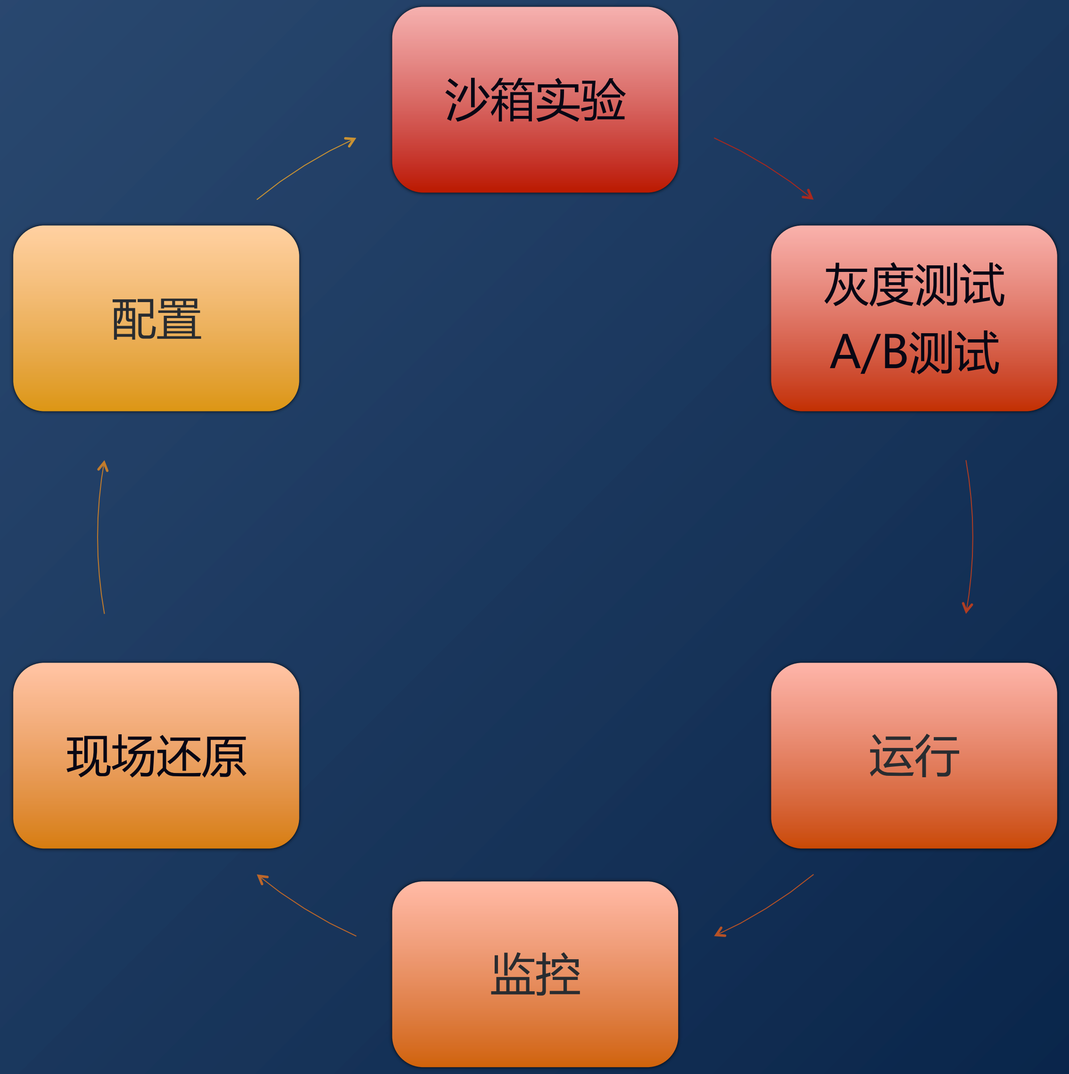
行为分析：写时计算

融合计算
大跨度时间范围计算：离线数据与实时数据融合。对精度的影响

跨域计算
- 跨域数据同步：依赖存储层数据同步或数据串行化处理时的调度
- 业务域隔离，单独计算

数据热点
- 热点数据识别，在串行化之前进行本地缓存
- 略过超大值 / 换统计算法到基数估计算法（如: HyperLogLog）

经验二：足够的实验与验证



经验三：熔断与兜底

股票熔断

熔断的并发实现

经验四：风控是一个业务决策

有商户愿意被薅羊毛而保持日活

有用户喜欢安全感而牺牲体验

有用户喜欢密码不相信生物识别

…

把决策权交还给用户

架构

没有最完美的架构，只有最合适的架构

架构是演进出来的，而不是设计出来的

关注QCon微信公众号，
获得更多干货！

“阿里技术”官方微信公众号

阿里聚安全: http://jaq.alibaba.com
Email: sihai.hush@alibaba-inc.com

Alibaba Group
阿里巴巴集团