ArchSummit全球架构师峰会北京站2015

航空电商大规模实时日志分析

易建科技 李锋/r6

Geekbang.
极客邦科技

整合全球最优质学习资源，帮助技术人和企业成长

Growing Technicians, Growing Companies

InfoQ queue

专注中高端技术人员的
技术媒体

EGO EXTRA GEEKS' ORGANIZATION NETWORKS

高端技术人员
学习型社交网络

StuQ queue

实践驱动的
IT职业学习和服务平台

GiT GEEKBANG INTERNATIONAL TRAINING 极客邦培训

一线专家驱动的
企业培训服务

地址：北京市朝阳区洛娃大厦C座6层1607室　　网址：www.geekbang.org

目录

1、介绍恶意行为防控系统的背景

2、介绍恶意行为防控系统的演进

介绍恶意行为防控系统的背景 — 行为定义

业务安全

- 恶意爬数据
- 恶意注册/刷验证码
- 刷单占位
- 恶意秒杀

运维安全

- 恶意扫描
- SQL注入
- 业务漏洞攻击
- DDOS/CC攻击

介绍恶意行为防控系统的背景 — 恶意行为盈利模式

加价

目录

1、介绍恶意行为防控系统的背景

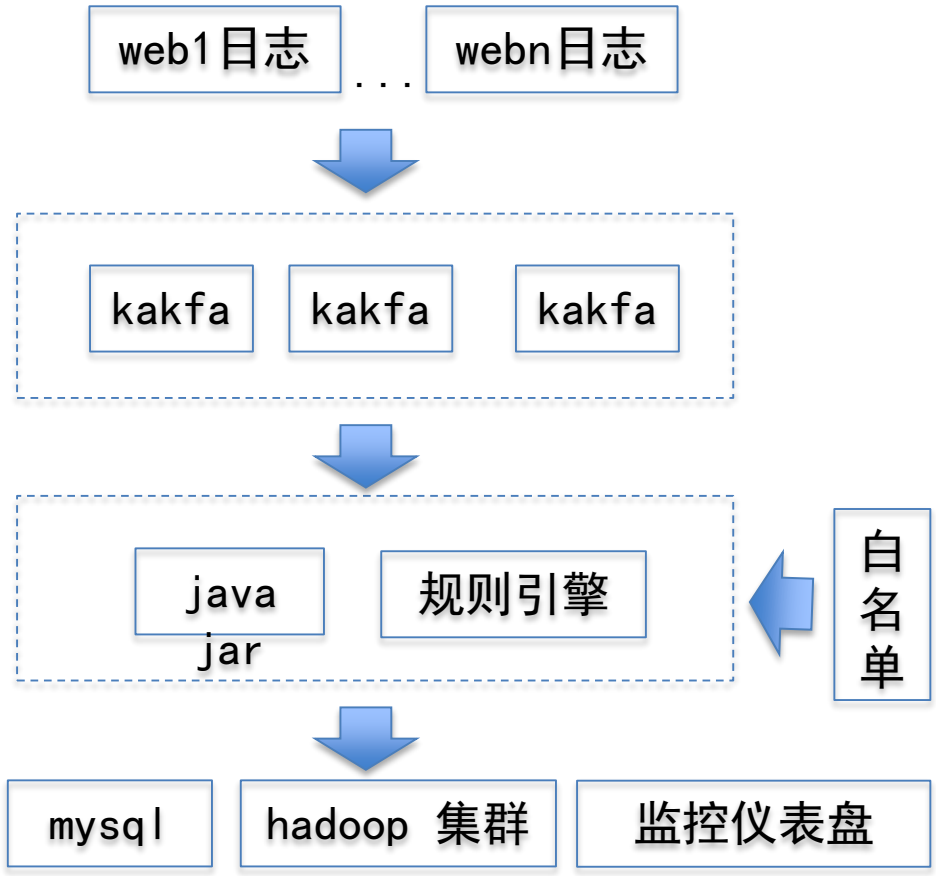
2、介绍恶意行为防控系统的演进

日峰值航班查询量	30, 000, 000
日峰值航班占座(个)	10, 000
日峰值注册用户	1, 000, 000
月峰值非正常停机	30分钟
月峰值支出费用(元)	1, 200, 000
日志	未利用

- 防火墙(iptables)/perl脚本
- 排除注册漏洞
- 采集web应用日志
- web应用内部埋点布防
- 系统级流量控制(tc)

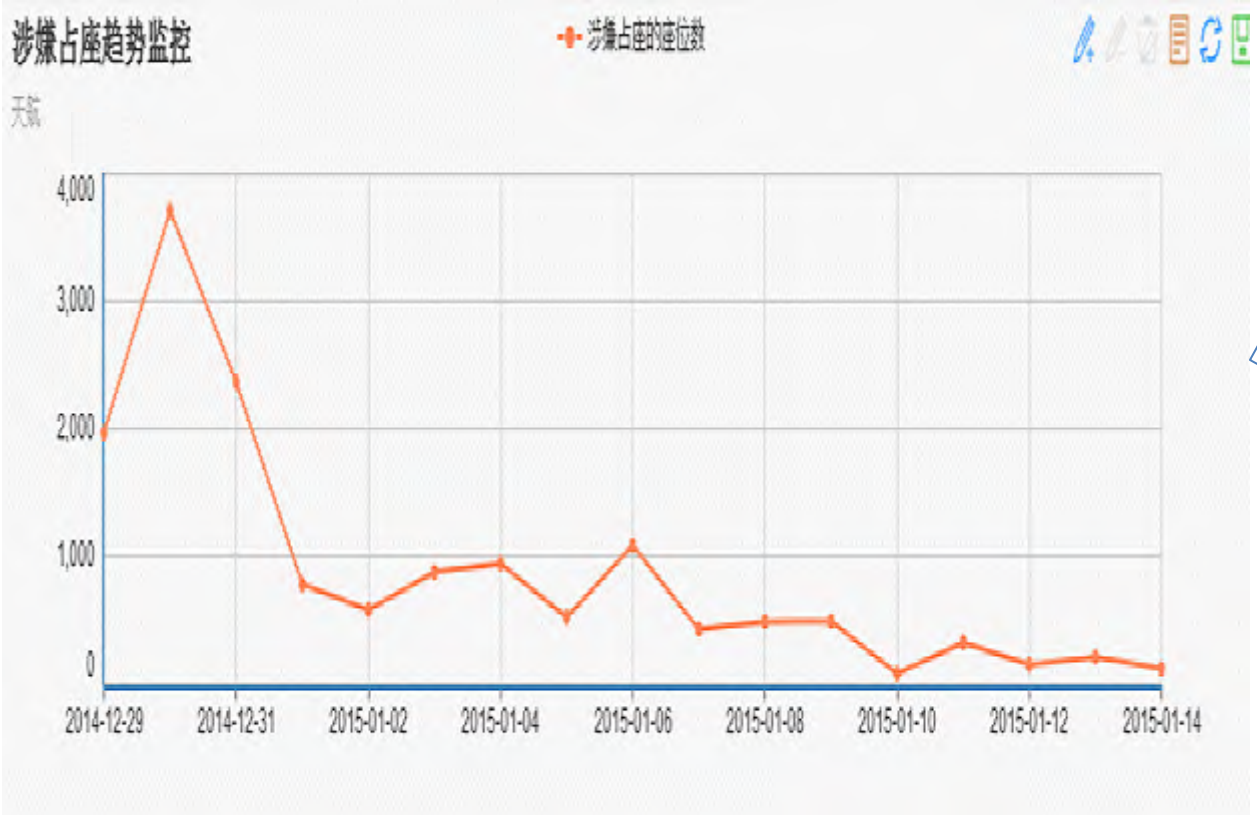
快速实施

恶意行为防控系统的演进　　— 第一阶段 系统架构



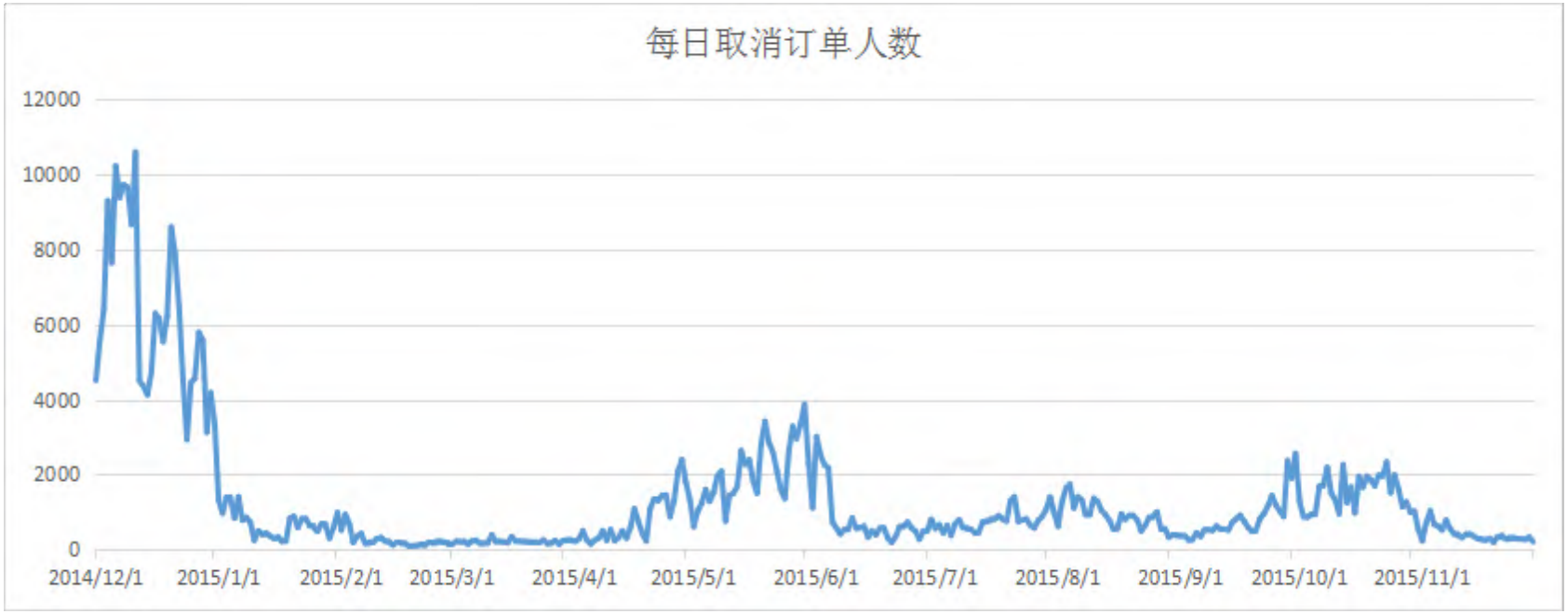
简单

恶意行为防控系统的演进 — 第一阶段 效果



↓95%

恶意行为防控系统的演进 — 遗留问题，占座再次出现



- 入侵式架构
- 监控不完整
- 事后处理模型

问题

恶意行为防控系统的演进 — 第二阶段 如何解决问题

入口级监控

灵活全面

多级时间窗口监控

实时在线

准实时在线

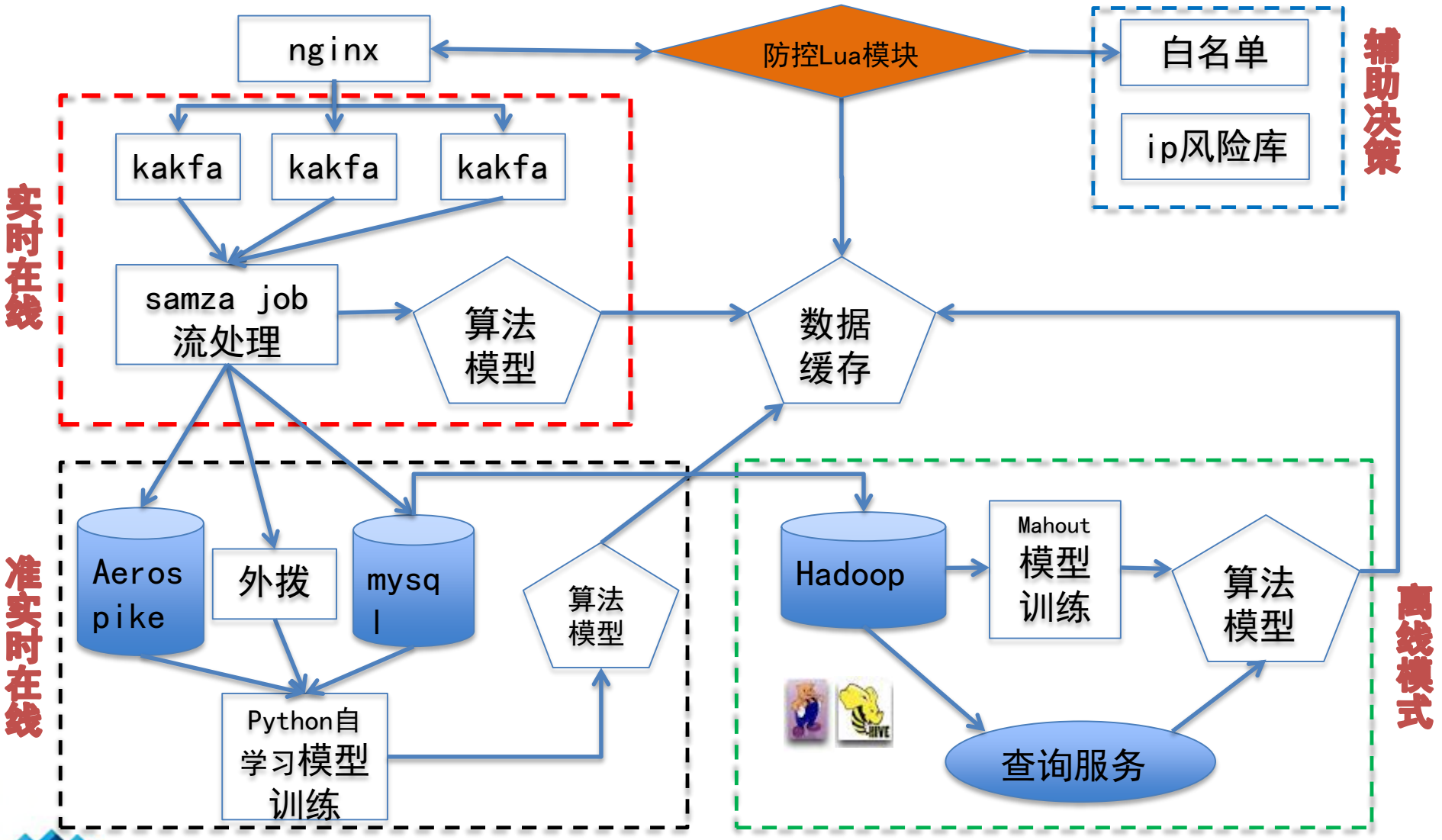
离线模式

应用级流量控制
（ip, cookie, user_agent）

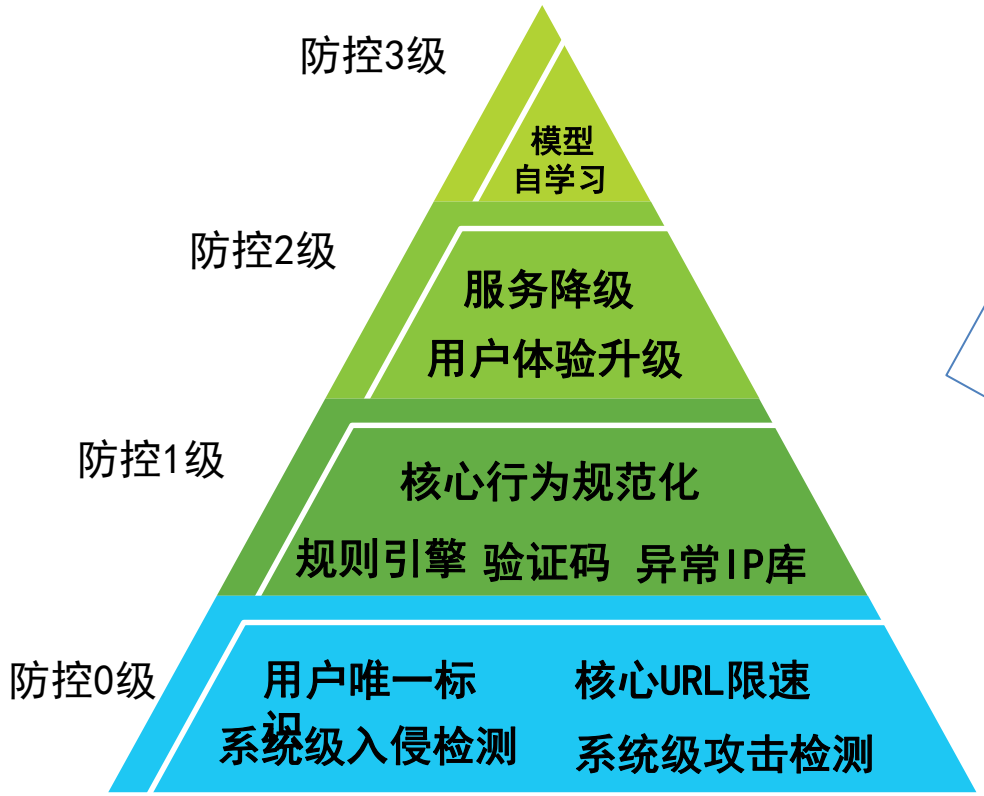
灵活全面

机器学习（逻辑回归，神经网络等）

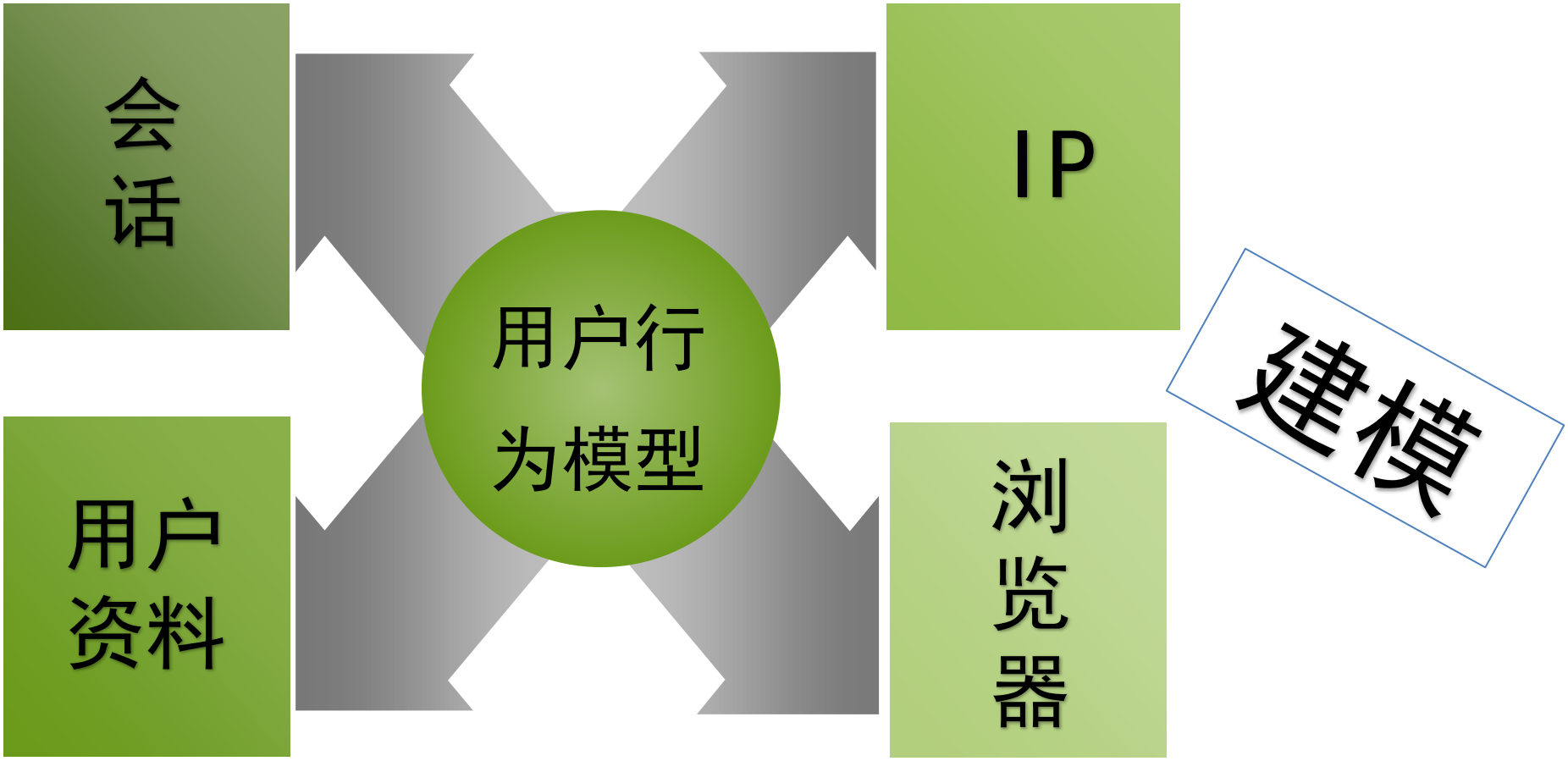
恶意行为防控系统的演进 — 第二阶段 系统架构



恶意行为防控系统的演进 — 第二阶段 系统架构特点1 nginx_lua 防控模块



多等级


会话
IP
用户行为模型
建模
用户资料
浏览器

机器学习算法：Logistic SVM（支持向量机） Kmean

自学习

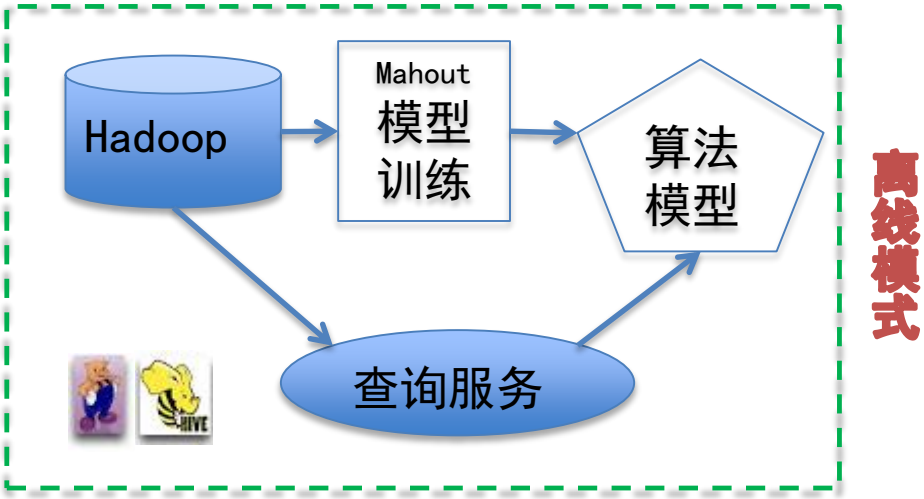
预测订单取消概率的数据模型：

X1	是否访问首页
X2	首页停留时间
X3	是否查询过航班
X4	查询航班停留时间
X5	是否访问过填写乘客信息页面
X6	填写乘客信息停留时间
X7	订单的乘客人数
X8	访问提交订单的页面次数
X9	是否加载过关键图片
X10	加载关键图片的次数
X11	是否访问过辅营页面
X12	是否访问最终提交订单的页面
X13	是否加载过CSS样式表
X14	是否为异常IP
Y	**订单状态**

```
1 15 1 5 1 14 1 3 1 5 0 1 0 0 1
1 0 0 0 0 0 1 1 1 31 0 0 0 0 1
0 0 1 0 0 0 3 7 1 4 0 0 0 0 1
1 102 1 4 1 391 1 7 1 14 1 0 0 0 1
1 0 1 0 0 0 1 7 1 7 1 1 0 0 1
1 19 1 0 0 0 1 7 1 12 1 1 0 0 1
1 20 1 0 0 0 1 7 1 7 1 1 0 0 1
1 574 1 0 0 0 1 7 1 11 1 0 0 0 1
1 846 1 7 1 27 2 7 1 23 1 1 0 0 1
1 0 0 0 0 0 2 1 1 2 0 0 0 0 1
1 0 0 0 1 7 2 1 0 0 0 0 0 0 1
1 14 1 0 0 0 2 2 0 0 0 0 0 0 1
```

The classify accuracy is:
90.284%

更新频率：
15mins

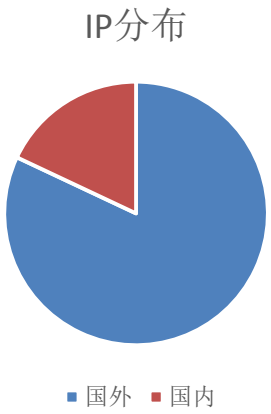


大数据

时间跨度：2014. 2至今

异常IP库：56, 000, 000

A4纸 700公里



需要更全面的实时行为预测

需要复杂机器学习算法的支持

spark

恶意行为防控系统的演进 — 未来产品规划

航空WAF+产品 ＋ 航空类“乌云” → 互联网+航空

Thanks!