微数据下的互联网安全与风控

陆文　　岂安科技 首席技术官

岂安科技

ABOUT ME

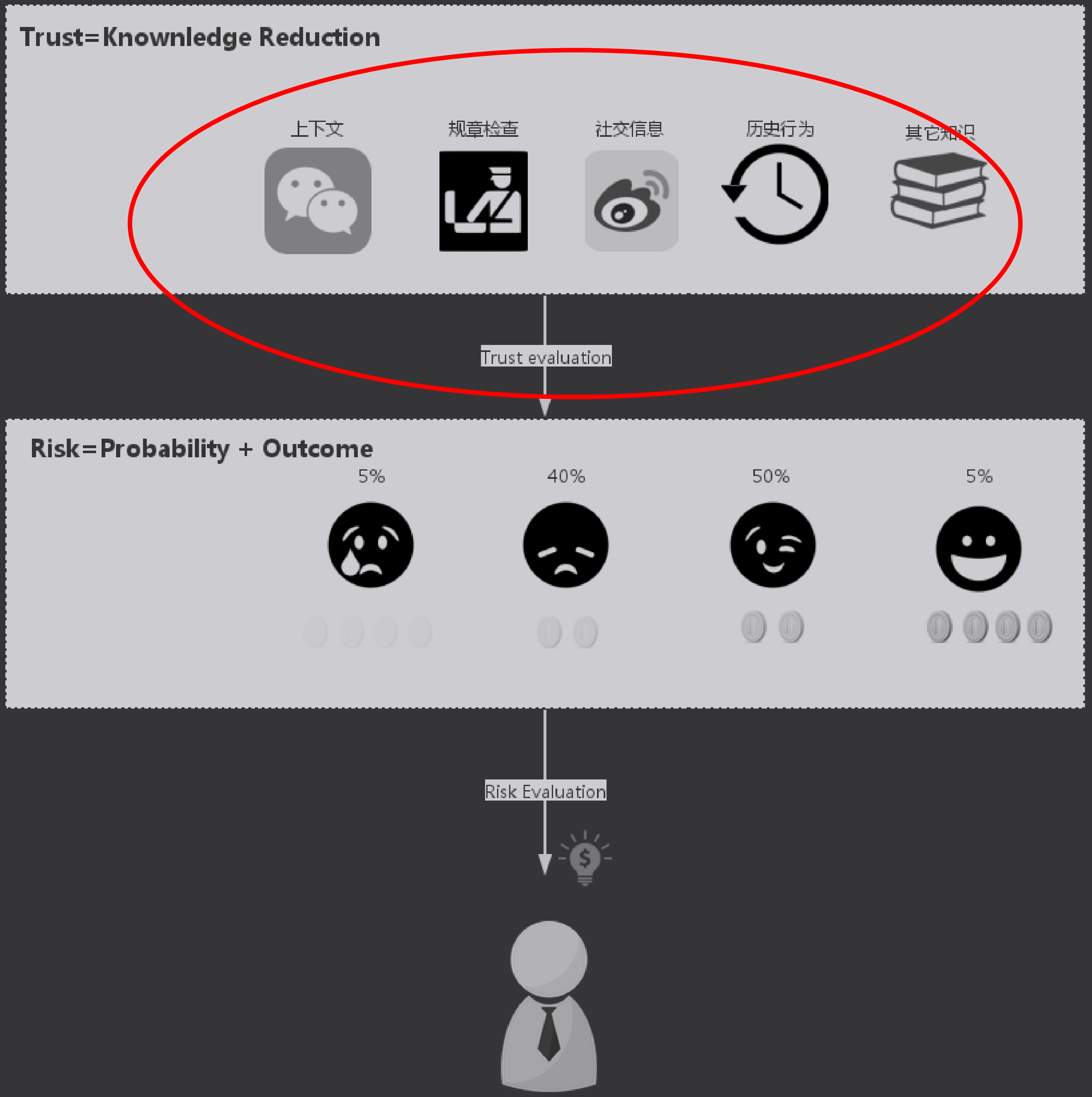
南京大学
可信计算与风险估计

PayPal
Risk Tools and Technology

携程
实时大数据平台

岂安科技
让互联网风控更简单

岂安科技



Trust=Knownledge Reduction

上下文

规章检查

社交信息

历史行为

其它知识

Trust evaluation

Risk=Probability + Outcome

5%

40%

50%

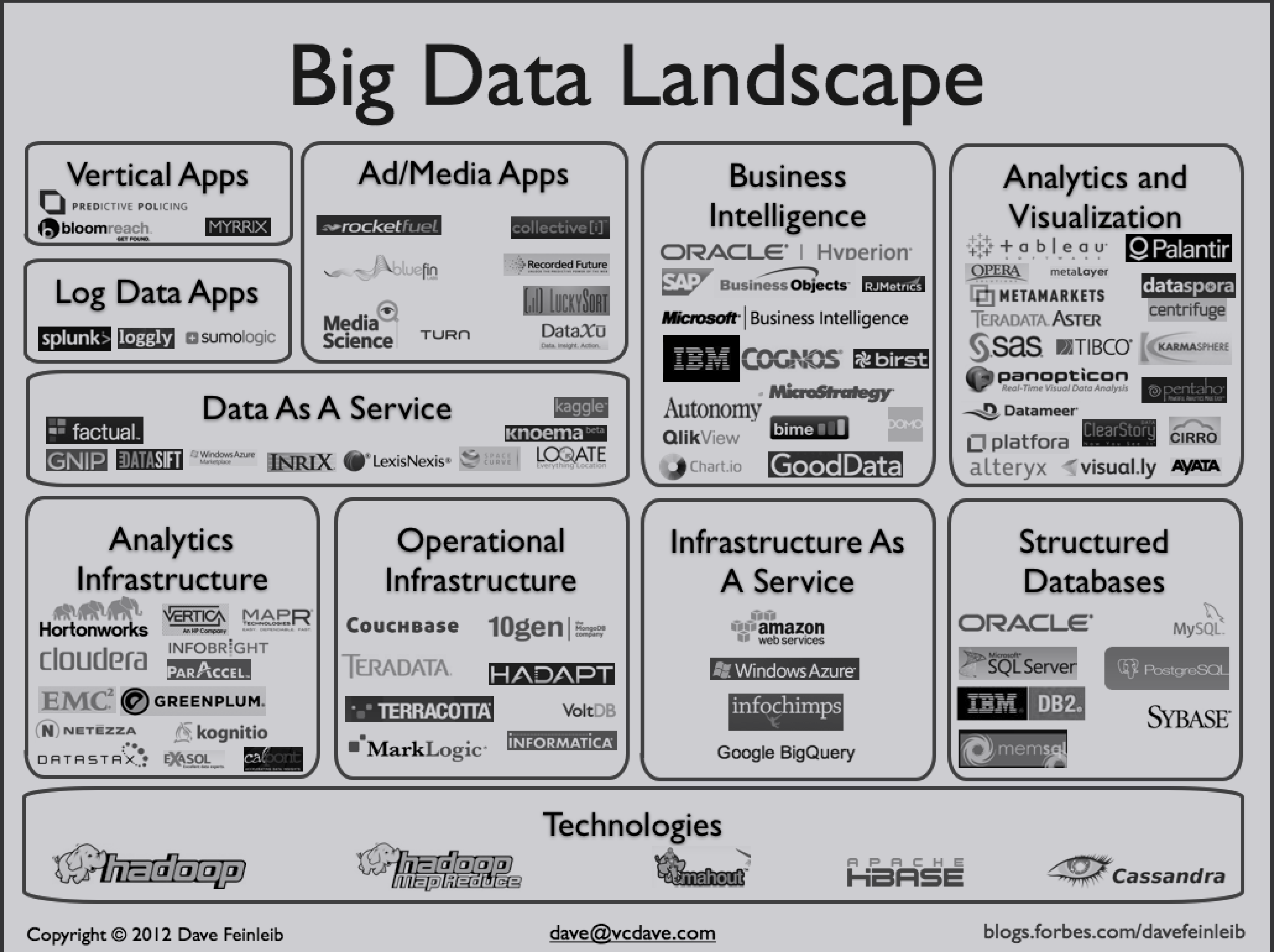
5%

Risk Evaluation

风控是什么？

岂安科技

大数据



Big Data Landscape

Vertical Apps
PREDICTIVE POLICING
bloomreach
MYRRIX

Log Data Apps
splunk>
loggly
sumologic

Ad/Media Apps
rocketfuel
collective[i]
bluefin
Recorded Future
LUCKYSORT
Media Science
TURN
DataXu

Data As A Service
factual.
GNIP
DATASIFT
Windows Azure Marketplace
INRIX
LexisNexis
SPACE CURVE
kaggle
knoema beta
LOQATE

Business Intelligence
ORACLE | Hyperion
SAP Business Objects
RJMetrics
Microsoft | Business Intelligence
IBM
COGNOS
birst
MicroStrategy
Autonomy
QlikView
bime
DOMO
Chart.io
GoodData

Analytics and Visualization
tableau
Palantir
OPERA
metaLayer
dataspora
METAMARKETS
centrifuge
TERADATA ASTER
sas
TIBCO
KARMASPHERE
panopticon
pentaho
Datameer
platfora
ClearStory
CIRRO
alteryx
visual.ly
AYATA

Analytics Infrastructure
Hortonworks
VERTICA
MAPR
cloudera
INFOBRIGHT
PARACCEL
EMC²
GREENPLUM
NETEZZA
kognitio
DATASTAX
EXASOL
calpont

Operational Infrastructure
COUCHBASE
10gen
TERADATA
HADAPT
TERRACOTTA
VoltDB
MarkLogic
INFORMATICA

Infrastructure As A Service
amazon web services
Windows Azure
infochimps
Google BigQuery

Structured Databases
ORACLE
MySQL
SQL Server
PostgreSQL
IBM DB2
SYBASE
memsql

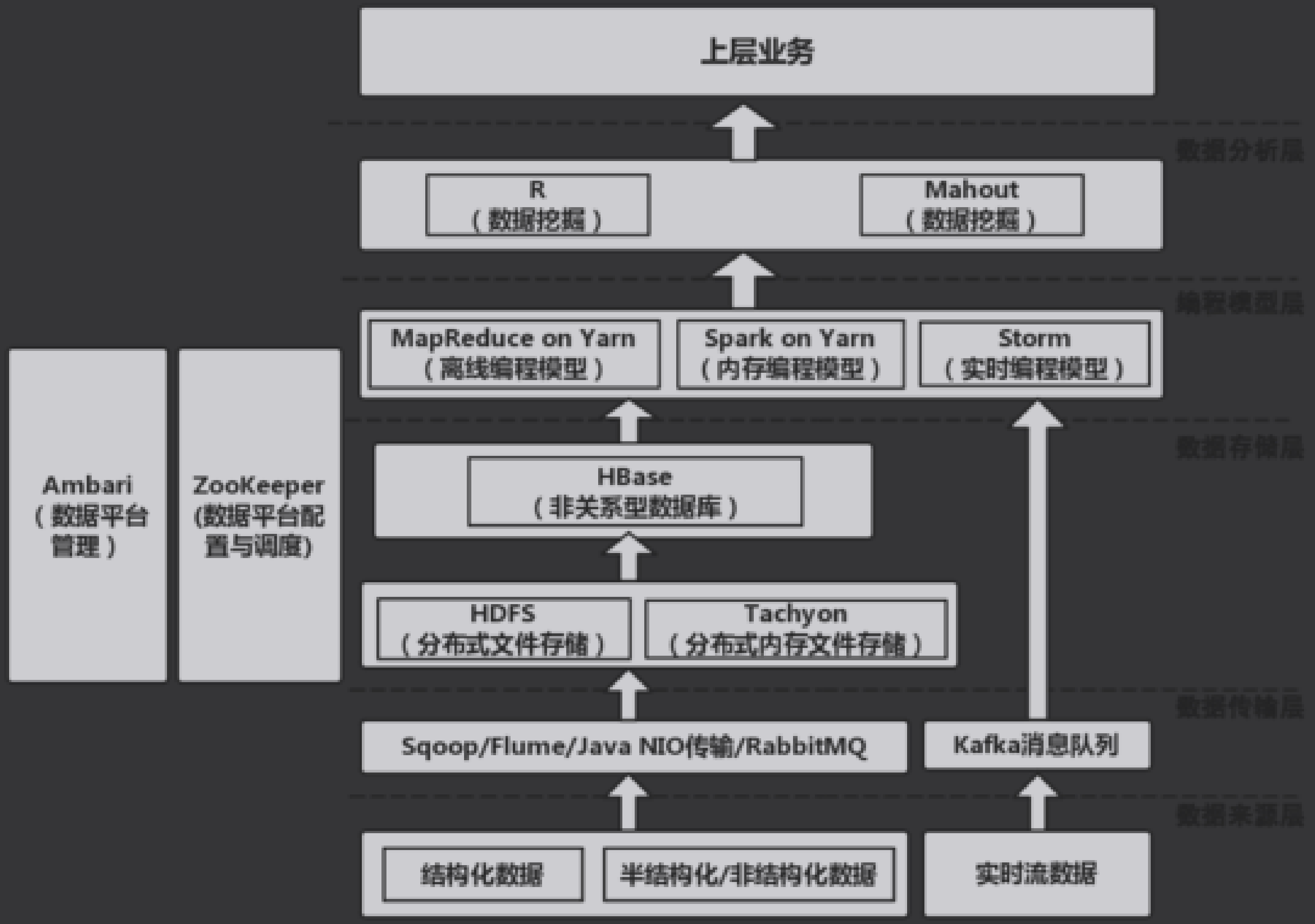
Technologies
hadoop
hadoop MapReduce
mahout
APACHE HBASE
Cassandra

dave@vcdave.com

blogs.forbes.com/davefeinleib

岩安科技

大数据下的风控



上层业务

数据分析层

R（数据挖掘）

Mahout（数据挖掘）

编程模型层

MapReduce on Yarn（离线编程模型）

Spark on Yarn（内存编程模型）

Storm（实时编程模型）

数据存储层

HBase（非关系型数据库）

HDFS（分布式文件存储）

Tachyon（分布式内存文件存储）

数据传输层

Sqoop/Flume/Java NIO传输/RabbitMQ

Kafka消息队列

数据来源层

结构化数据

半结构化/非结构化数据

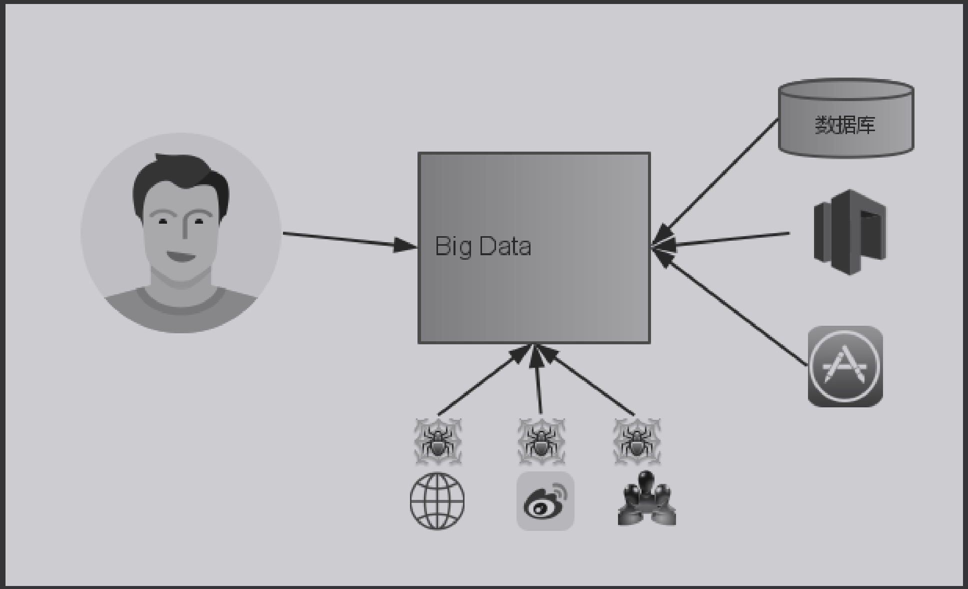
实时流数据

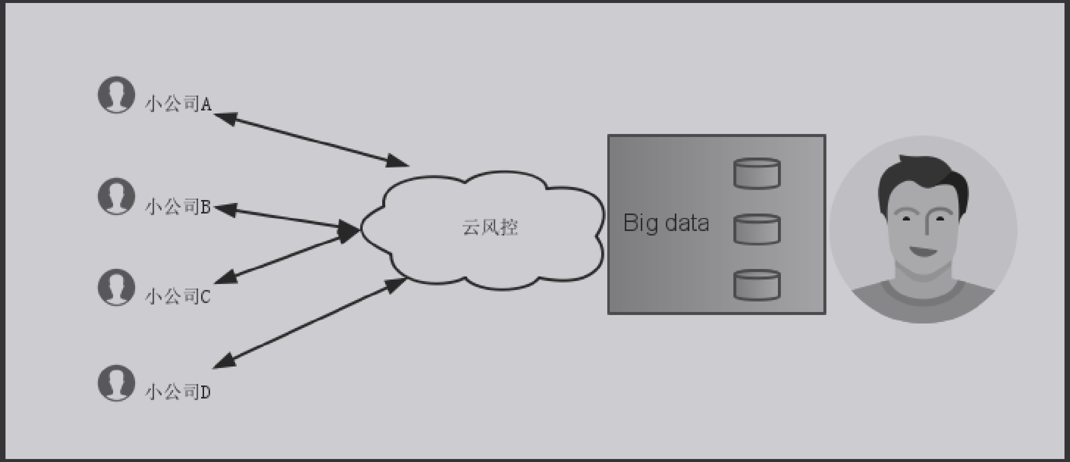
Ambari（数据平台管理）

ZooKeeper(数据平台配置与调度)

岂安科技

大数据下的风控模式





岂安科技

问题?

成本高昂（人、机器）

研发集成难度高，运维开销大

数据量少，甚至没有数据

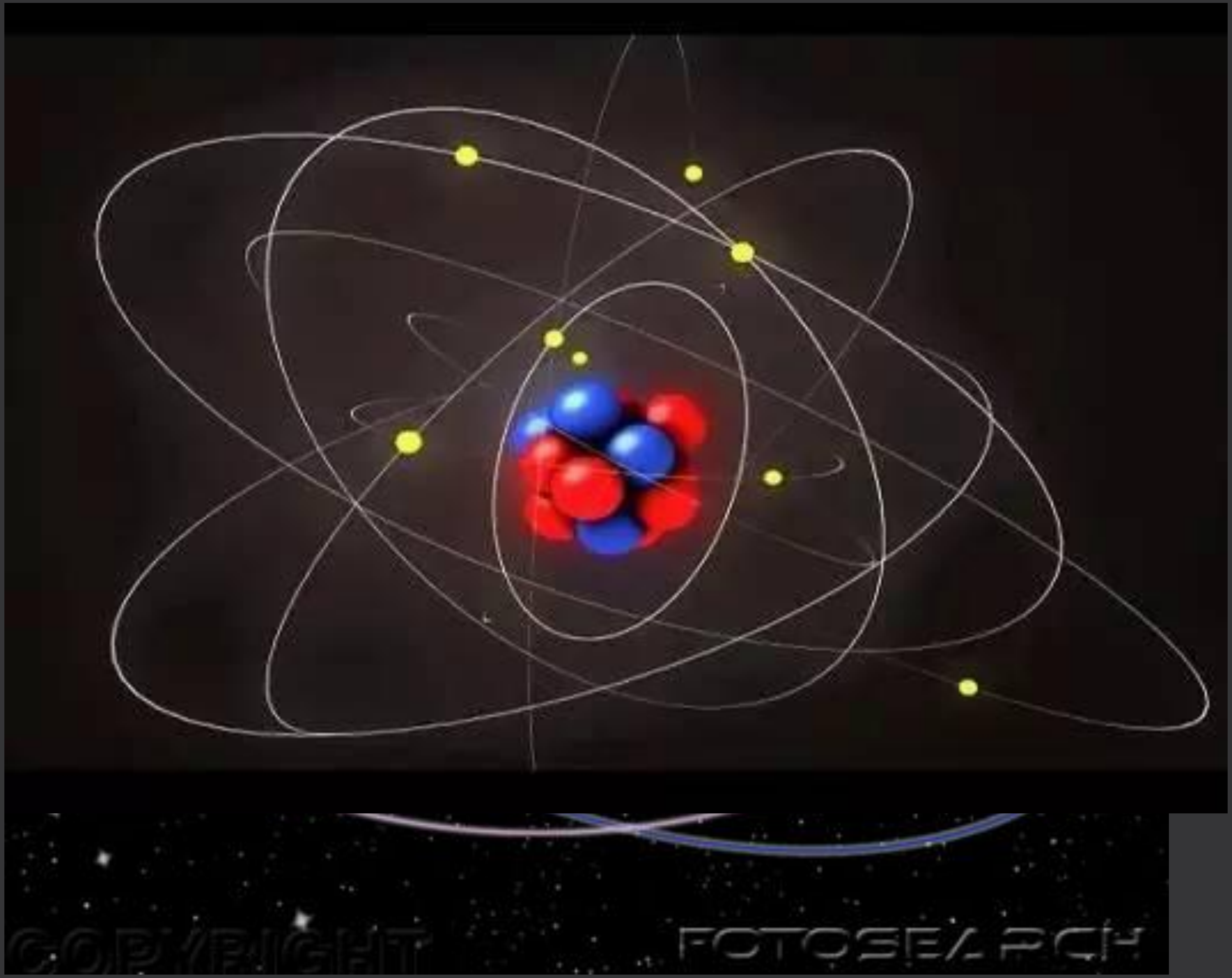
大数据下的数据隐私性

互联网身份识别的脆弱性

大数据风控结果的时效性

无法应对新的攻击实体和行为

能否有效挖掘存有的数据



岩安科技

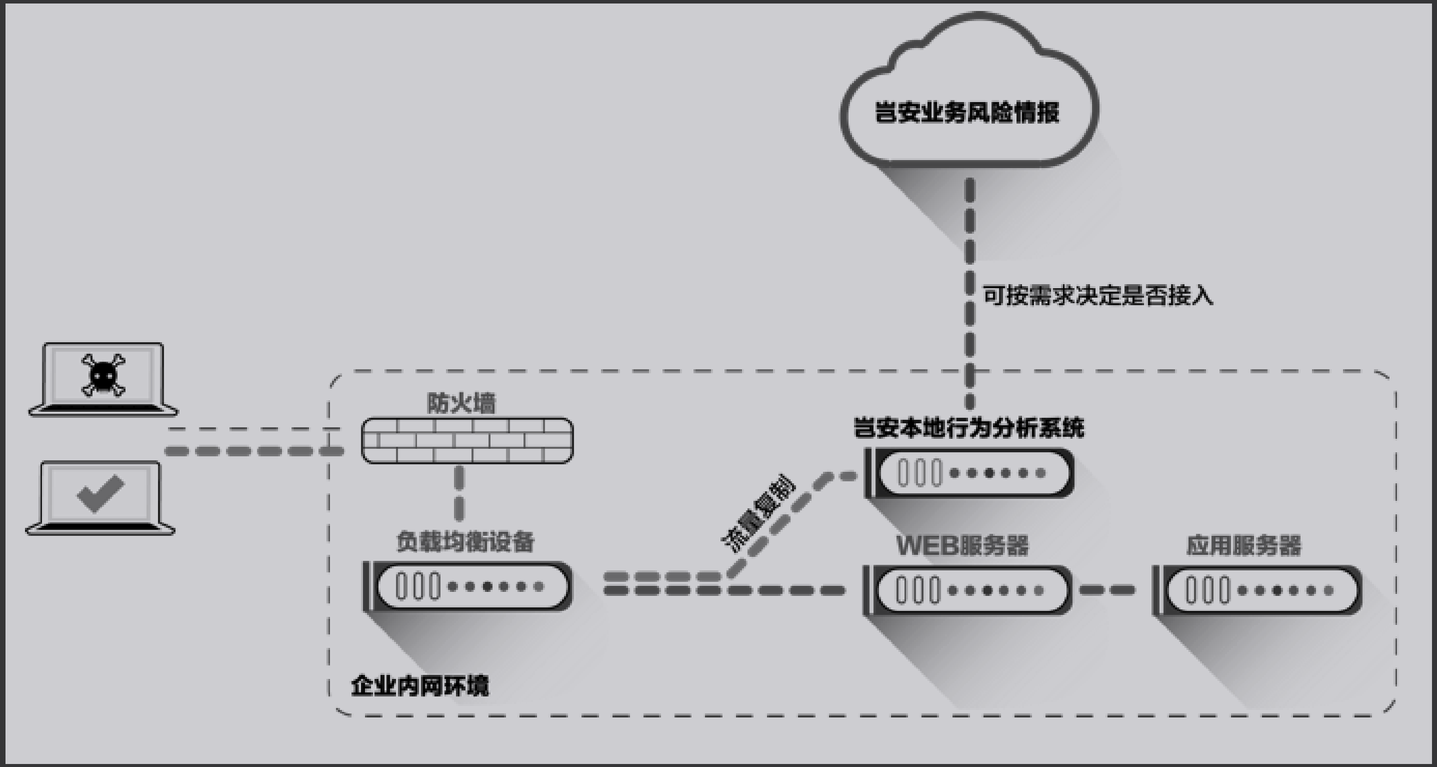
微数据风控

在当前的环境下尽可能的获取数据

在有限的数据数量和质量下，尽可能的计算

采用低成本解决方案，快速解决问题

成本低，该有的不能少

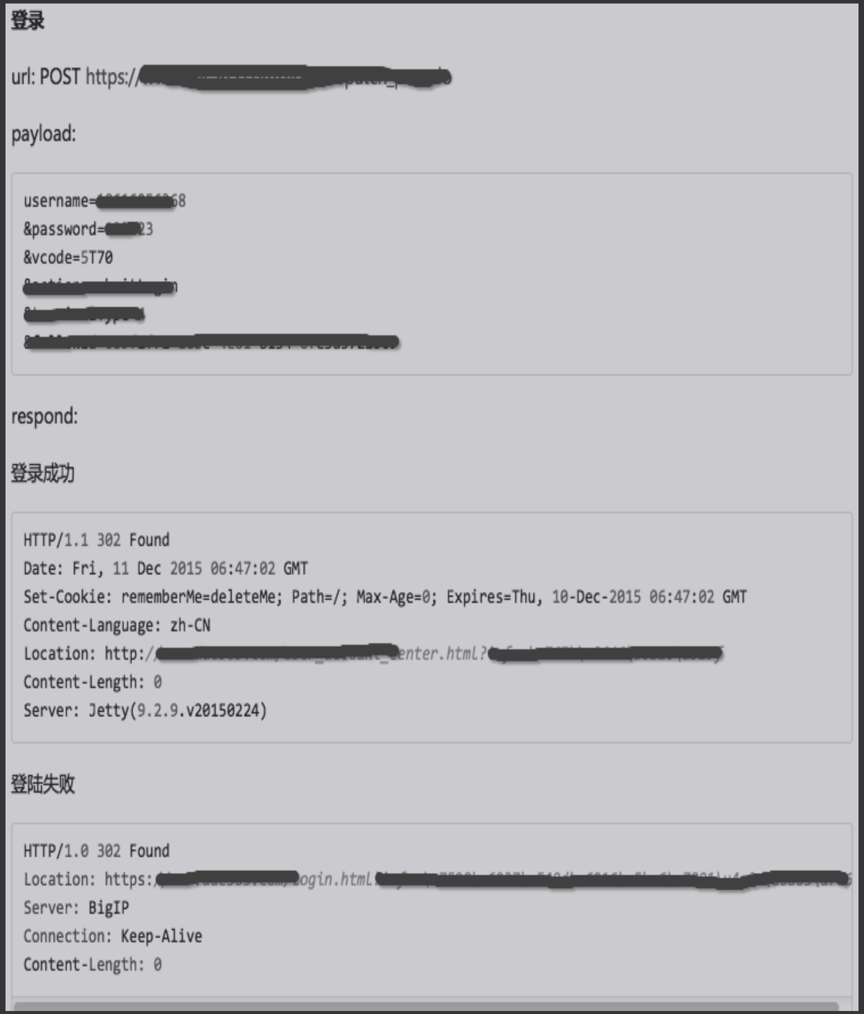


微数据风控
（数据采集）

岂安科技

岂安科技

微数据风控
（数据采集）



登录

url: POST https:/[redacted]

payload:

```
username=[redacted]
&password=[redacted]
&vcode=5T70
[redacted]
```

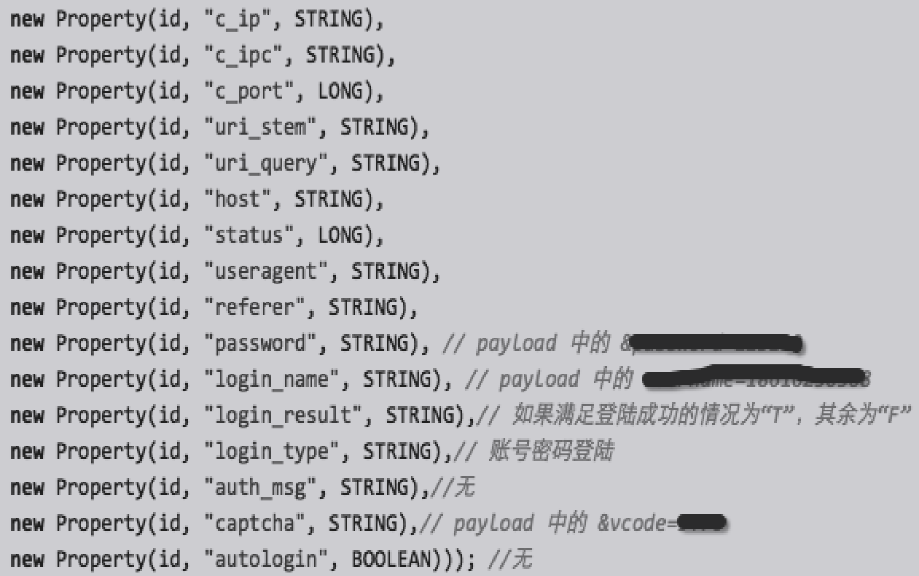
respond:

登录成功

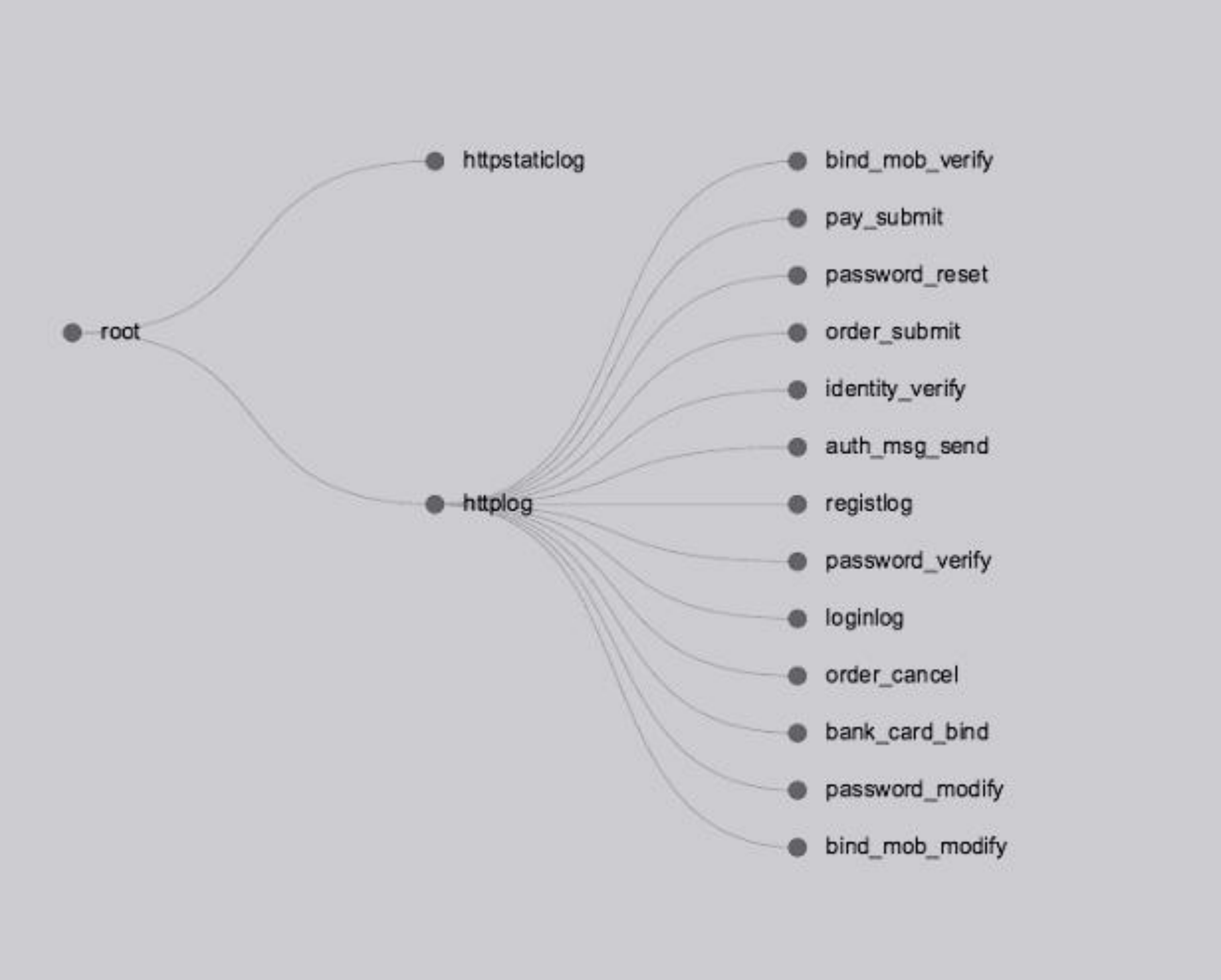
```
HTTP/1.1 302 Found
Date: Fri, 11 Dec 2015 06:47:02 GMT
Set-Cookie: rememberMe=deleteMe; Path=/; Max-Age=0; Expires=Thu, 10-Dec-2015 06:47:02 GMT
Content-Language: zh-CN
Location: http:/[redacted]enter.html?[redacted]
Content-Length: 0
Server: Jetty(9.2.9.v20150224)
```

登陆失败

```
HTTP/1.0 302 Found
Location: https:/[redacted]ogin.html[redacted]
Server: BigIP
Connection: Keep-Alive
Content-Length: 0
```

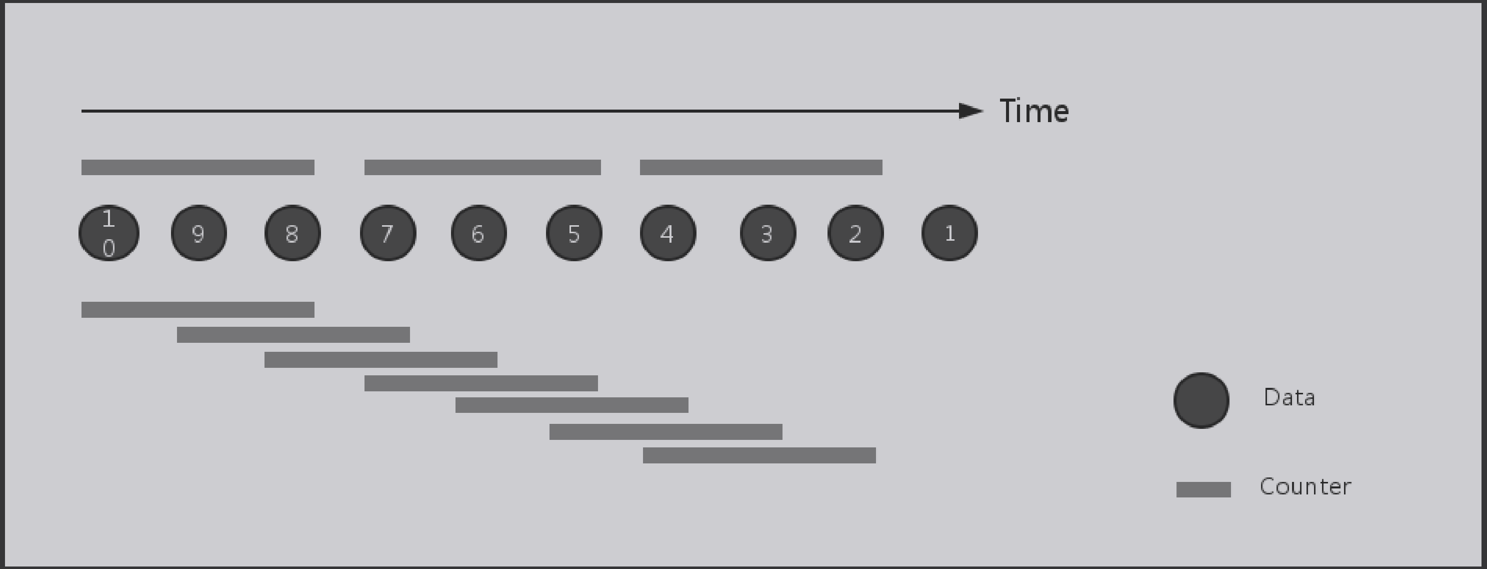


```
new Property(id, "c_ip", STRING),
new Property(id, "c_ipc", STRING),
new Property(id, "c_port", LONG),
new Property(id, "uri_stem", STRING),
new Property(id, "uri_query", STRING),
new Property(id, "host", STRING),
new Property(id, "status", LONG),
new Property(id, "useragent", STRING),
new Property(id, "referer", STRING),
new Property(id, "password", STRING), // payload 中的 [redacted]
new Property(id, "login_name", STRING), // payload 中的 [redacted]
new Property(id, "login_result", STRING),// 如果满足登陆成功的情况为“T”，其余为“F”
new Property(id, "login_type", STRING),// 账号密码登陆
new Property(id, "auth_msg", STRING),//无
new Property(id, "captcha", STRING),// payload 中的 &vcode=[redacted]
new Property(id, "autologin", BOOLEAN))); //无
```



微数据风控
（数据采集）

岩安科技



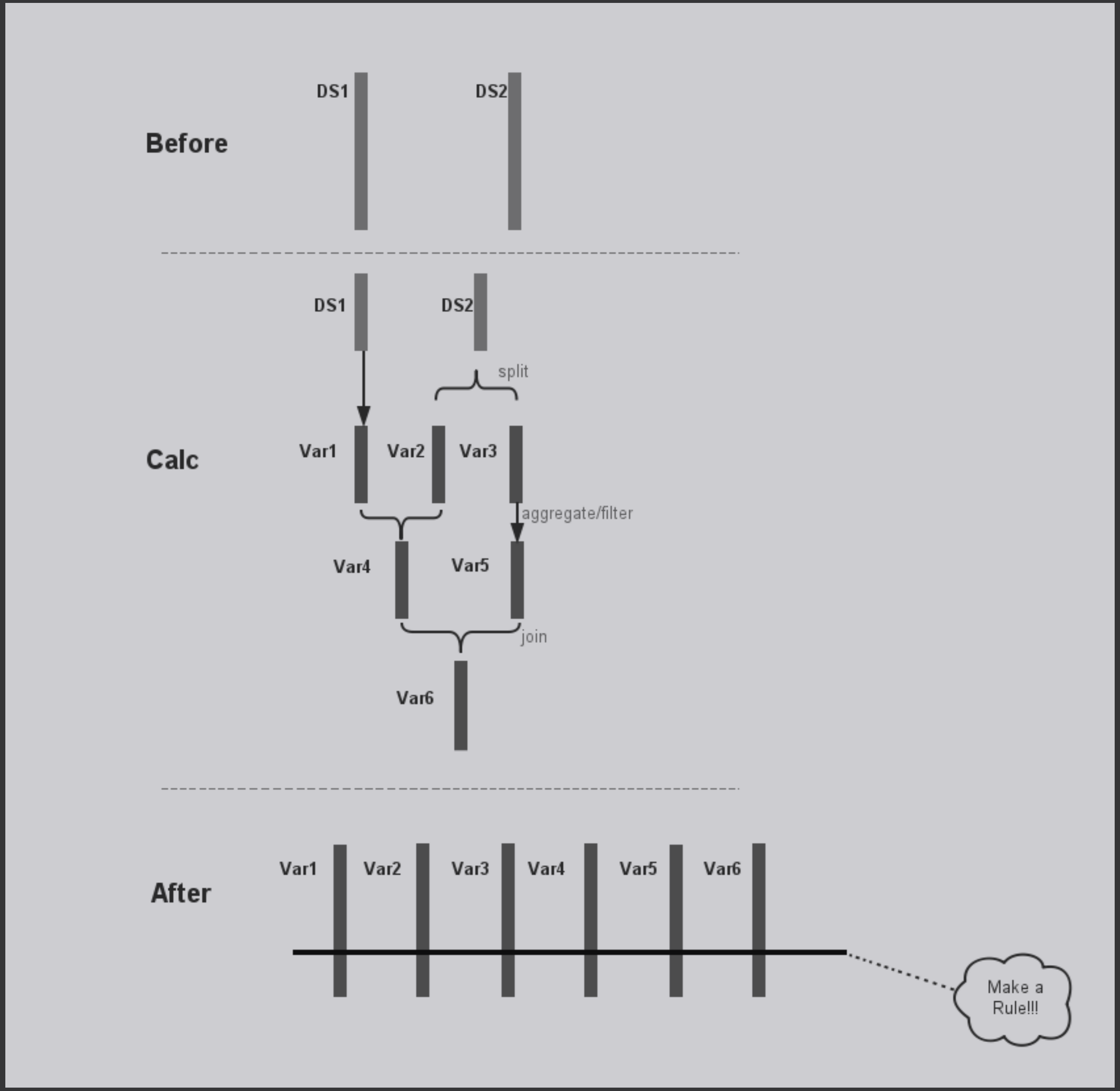
微数据风控（实时计算）

岩安科技

100/s	SQL	登录
10000/s	NoSQL	PV
100000+/s	Memory	HTTP

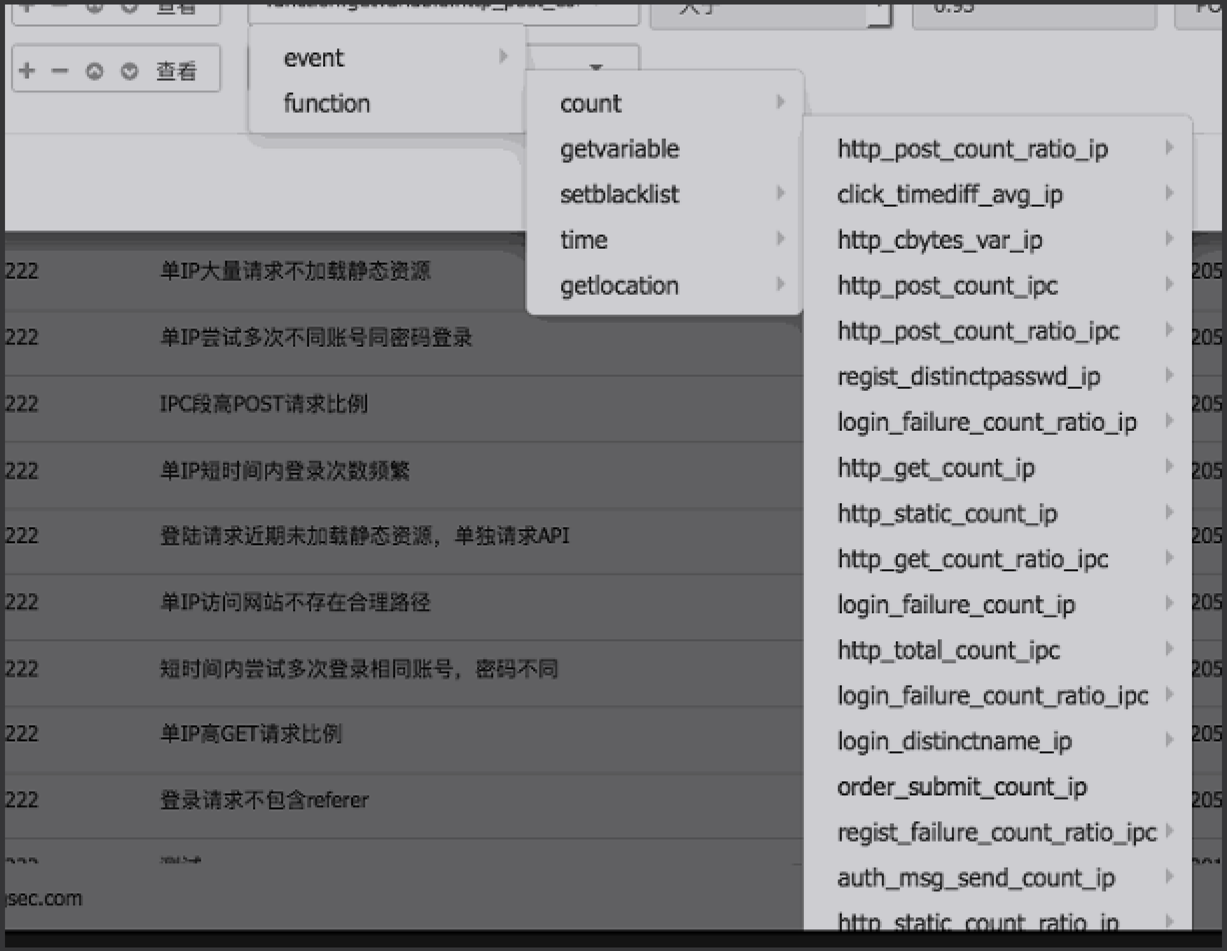
微数据风控（内存计算）

岩安科技



微数据风控
（变量体系）

岩安科技

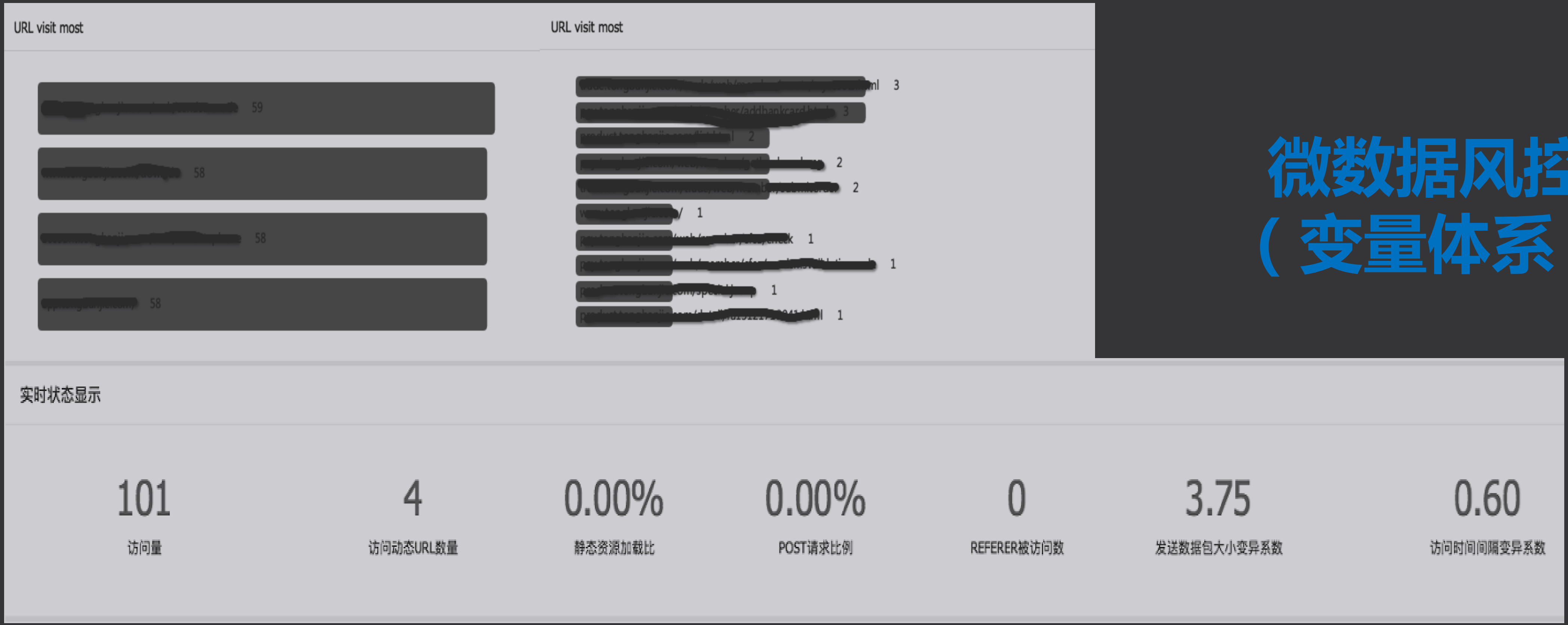


222	单IP大量请求不加载静态资源	205
222	单IP尝试多次不同账号同密码登录	205
222	IPC段高POST请求比例	205
222	单IP短时间内登录次数频繁	205
222	登陆请求近期未加载静态资源，单独请求API	205
222	单IP访问网站不存在合理路径	205
222	短时间内尝试多次登录相同账号，密码不同	205
222	单IP高GET请求比例	205
222	登录请求不包含referer	205
222	测试	205

微数据风控
（变量体系）

岩安科技

微数据风控（变量体系）



URL visit most

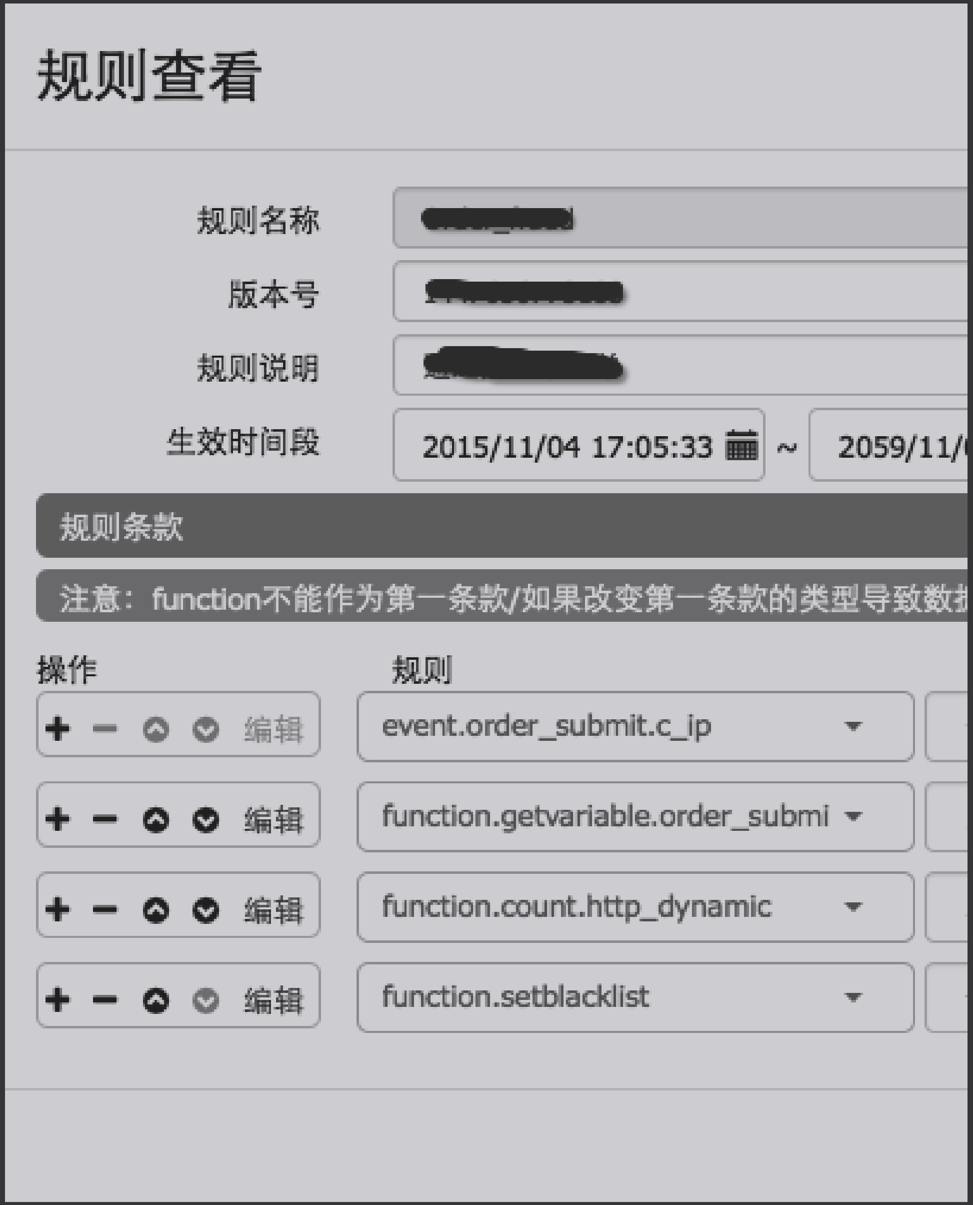
[redacted]	59
[redacted]	58
[redacted]	58
[redacted]	58

URL visit most

[redacted]	3
[redacted]	3
[redacted]	2
[redacted]	2
[redacted]	2
[redacted]	1
[redacted]	1
[redacted]	1
[redacted]	1
[redacted]	1

实时状态显示

访问量	访问动态URL数量	静态资源加载比	POST请求比例	REFERER被访问数	发送数据包大小变异系数	访问时间间隔变异系数
101	4	0.00%	0.00%	0	3.75	0.60



规则查看

规则名称 [redacted]

版本号 [redacted]

规则说明 [redacted]

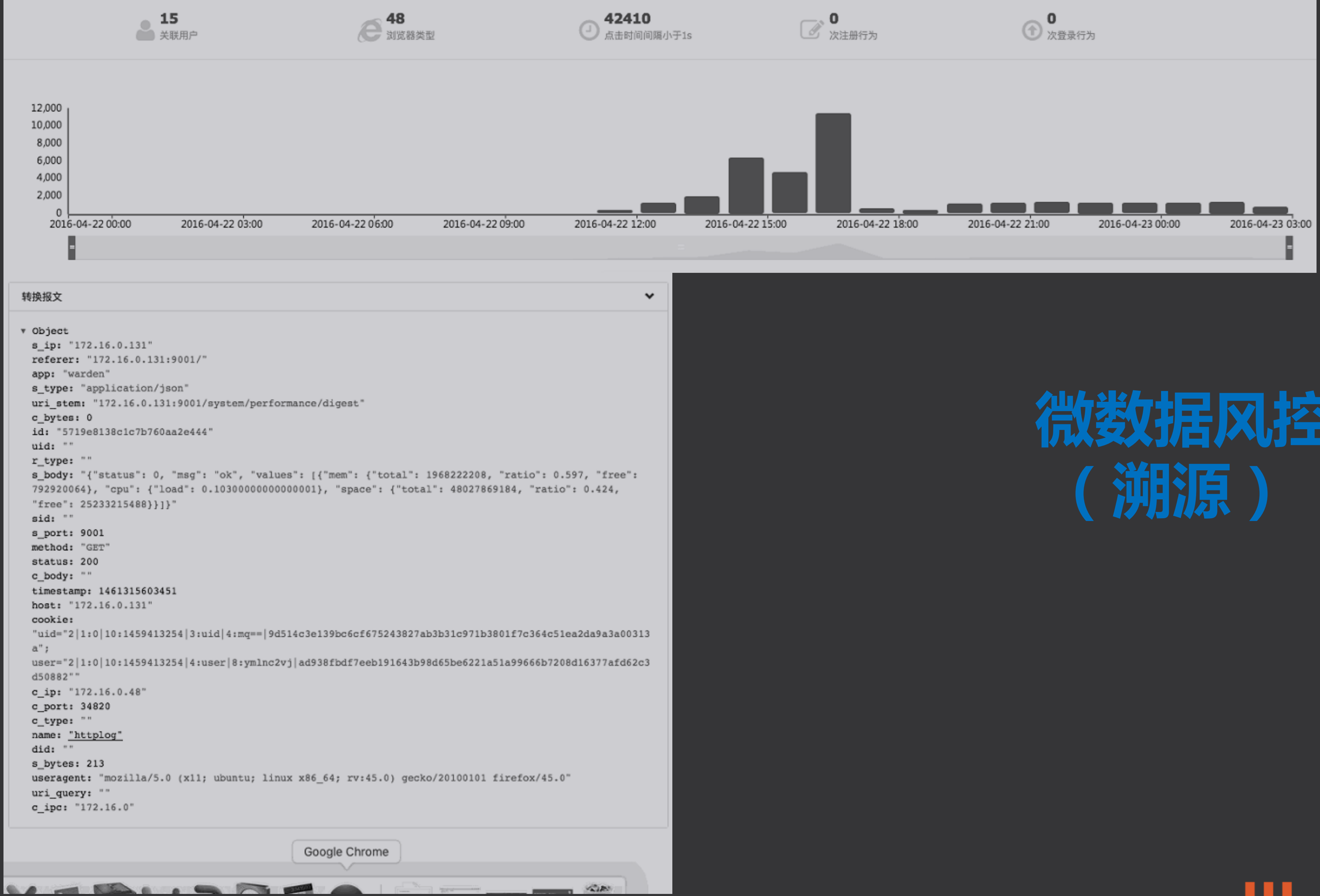
生效时间段 2015/11/04 17:05:33 ~ 2059/11/

规则条款

注意：function不能作为第一条款/如果改变第一条款的类型导致数

操作	规则
+ − 编辑	event.order_submit.c_ip
+ − 编辑	function.getvariable.order_submi
+ − 编辑	function.count.http_dynamic
+ − 编辑	function.setblacklist

微数据风控
（规则引擎）



15 关联用户

48 浏览器类型

42410 点击时间间隔小于1s

0 次注册行为

0 次登录行为

转换报文

```
▾ Object
  s_ip: "172.16.0.131"
  referer: "172.16.0.131:9001/"
  app: "warden"
  s_type: "application/json"
  uri_stem: "172.16.0.131:9001/system/performance/digest"
  c_bytes: 0
  id: "5719e8138c1c7b760aa2e444"
  uid: ""
  r_type: ""
  s_body: "{"status": 0, "msg": "ok", "values": [{"mem": {"total": 1968222208, "ratio": 0.597, "free":
  792920064}, "cpu": {"load": 0.10300000000000001}, "space": {"total": 48027869184, "ratio": 0.424,
  "free": 25233215488}}]}"
  sid: ""
  s_port: 9001
  method: "GET"
  status: 200
  c_body: ""
  timestamp: 1461315603451
  host: "172.16.0.131"
  cookie:
  "uid="2|1:0|10:1459413254|3:uid|4:mq==|9d514c3e139bc6cf675243827ab3b31c971b3801f7c364c51ea2da9a3a00313
  a";
  user="2|1:0|10:1459413254|4:user|8:ymlnc2vj|ad938fbdf7eeb191643b98d65be6221a51a99666b7208d16377afd62c3
  d50882""
  c_ip: "172.16.0.48"
  c_port: 34820
  c_type: ""
  name: "httplog"
  did: ""
  s_bytes: 213
  useragent: "mozilla/5.0 (x11; ubuntu; linux x86_64; rv:45.0) gecko/20100101 firefox/45.0"
  uri_query: ""
  c_ipc: "172.16.0"
```

微数据风控
（溯源）

通过流量来获取和还原数据

对细节特征进行计算，全滑窗计算

中小企业单机解决问题，支持水平扩展

涵盖数据采集、计算、规则引擎、持久化和展示

微数据风控

岂安科技

更多详情请关注

岂安微信公众号

岂安科技