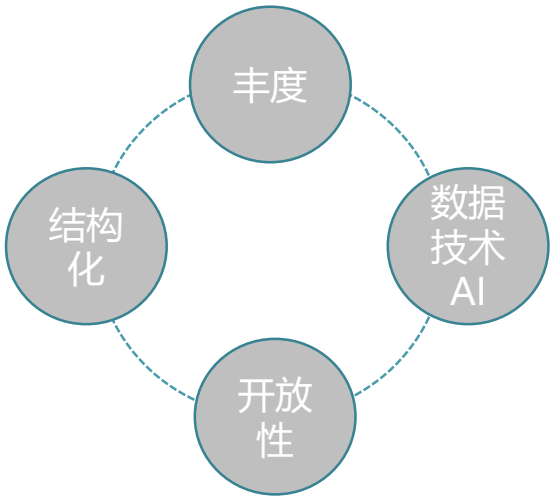
数据驱动安全思考

董晓琼

数据优势及行业转变
数据驱动安全
安全场景与应用
数据的闭环与价值

数据优势及行业转变

数据的显著优势、丰富的业务场景、及基于数据技术的跨越式发展：重塑传统行业格局，跨越传统企业的技术壁垒。



业务模式的创新
带动了传统企业-跨越至工业3.0
AI-传统行业格局的重塑
- 汽车、金融、医疗健康

数据驱动安全是什么？

数据驱动安全 非“情报驱动安全”

定义：在不断变化的组织环境中，为有效地识别和管理动态风险，在组织内部使用的一种方法；

安全场景与应用

UEBA-终端管控场景

诉求：员工在-业务系统访问过程中的“异常行为”判断；

问题：监控范围大、员工行为关注的业务点和行为差异显著、专家规则极易失效

“变化”中的风险识别

原始字段

- 访问时间
- IP
- url
- 流量

数据处理

基本统计信息
访问量
URL数量
IP数量
流量相关统计
访问时间间隔统计

IP统计信息
常用IP-C段
常用IP数量
全部IP-C段数量
全部IP数量
...

功能节点统计信息
用户
时段
url
访问量

行为轨迹
员工-URL 访问轨迹

用户历史行为
用户访问分布情况（均值、方差、偏度、峰度、分位数...）

群组历史行为
用户所属群组访问分布情况（均值、方差、偏度、峰度、分位数...）

功能节点特征
url 访问 top 10
敏感url
业务逻辑特征

特征

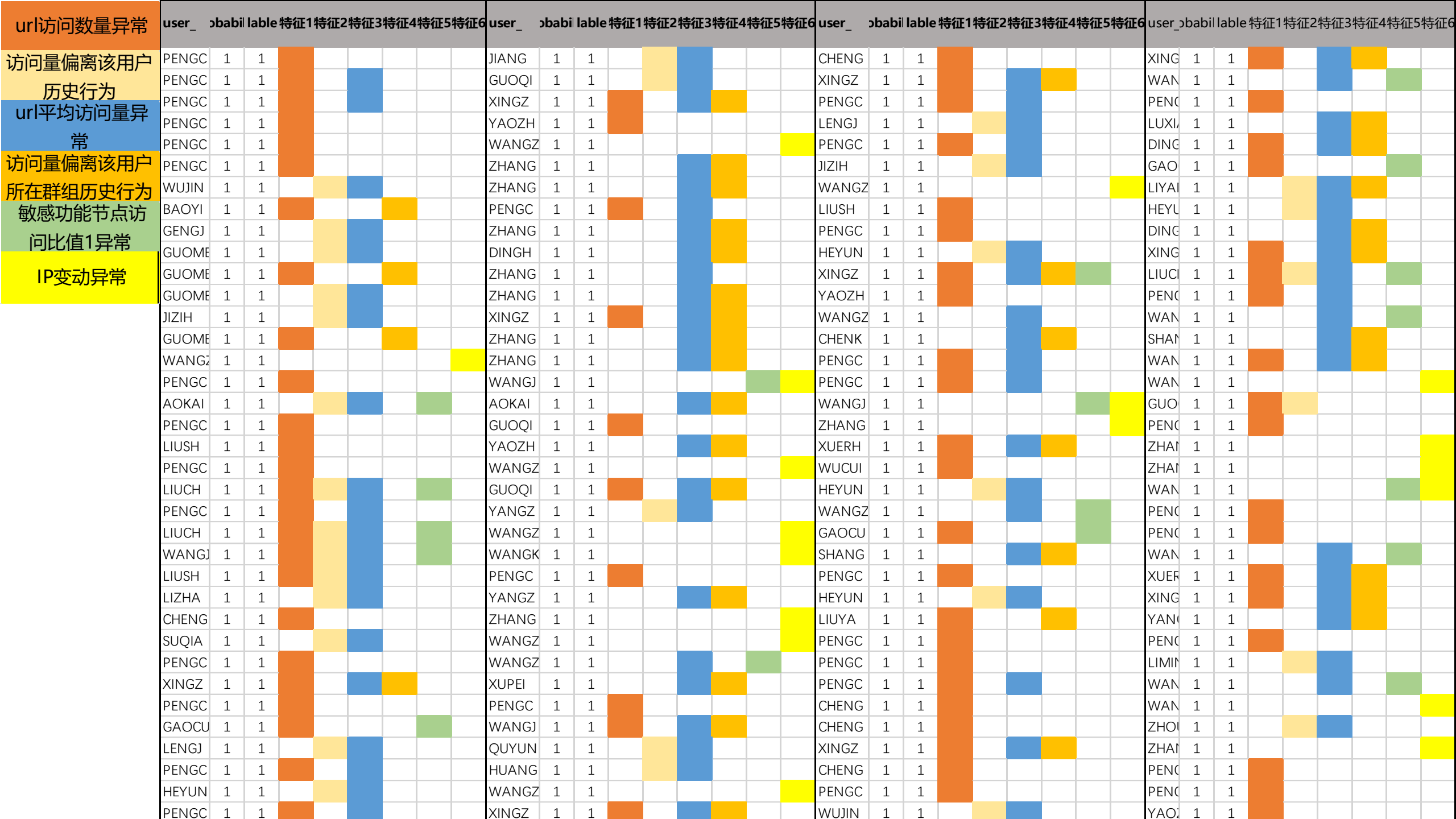
	主成分1	主成分2	主成分3	主成分4	主成分5	主成分6	主成分7
count	-0.31	0.00	0.15	-0.04	0.12	-0.07	0.05
count_url	-0.26	-0.23	-0.15	0.00	-0.08	0.06	-0.04
count_ip	-0.11	-0.06	-0.04	-0.09	0.19	0.24	-0.04
bytes_cv	-0.14	-0.23	-0.38	-0.18	-0.22	-0.11	-0.05
bytes_kewness	-0.19	-0.26	-0.20	-0.20	-0.20	-0.21	-0.07
bytes_kurtosis	-0.16	-0.19	-0.06	-0.20	-0.11	-0.37	0.29
diff_time_cv	-0.10	-0.16	-0.19	-0.14	-0.01	0.02	-0.28
log_bytes	-0.25	-0.18	-0.05	-0.05	0.04	-0.04	-0.22
urlTop1	-0.21	-0.15	0.19	0.21	-0.09	0.24	0.05
urlTop2	-0.26	0.17	0.08	0.08	-0.17	0.06	-0.03
urlTop3	-0.20	-0.01	-0.17	-0.03	0.01	0.03	0.02
urlTop4	-0.20	-0.16	0.20	0.20	-0.09	0.24	0.03
urlTop5	-0.25	0.21	0.00	0.16	0.14	-0.21	-0.08
urlTop6	-0.26	0.20	0.00	0.20	0.17	-0.23	-0.10
urlTop7	-0.08	-0.21	0.47	-0.30	0.14	-0.03	-0.04
urlTop8	-0.13	0.00	-0.25	-0.14	0.43	0.32	0.19
urlTop9	-0.15	-0.02	-0.25	-0.12	0.40	0.34	0.17
urlTop10	-0.22	0.22	0.05	-0.12	-0.23	0.16	-0.05
url3	-0.19	-0.13	0.09	0.27	-0.18	0.14	0.28
url5	-0.15	-0.09	0.17	-0.04	0.12	0.00	-0.23
url2	-0.26	0.20	0.00	0.20	0.17	-0.23	-0.10
url4	-0.08	-0.21	0.47	-0.30	0.14	-0.03	-0.04
url6	-0.11	0.32	0.05	-0.30	-0.21	0.15	0.10
url7	-0.11	0.32	0.05	-0.28	-0.21	0.18	0.04
url1	-0.21	0.33	-0.01	-0.06	0.07	-0.15	-0.04
user_deviate	-0.01	0.00	0.02	-0.03	0.03	-0.14	0.30
group_deviate	-0.06	0.05	0.02	-0.14	0.12	-0.29	0.55
BL1	-0.09	0.17	-0.03	-0.22	-0.28	0.11	-0.06
BL2	-0.16	-0.09	-0.04	0.24	-0.07	-0.03	-0.07
BL3	-0.11	-0.13	0.02	0.23	-0.22	0.13	0.35

- 取方差贡献度大于1，提取特征主成分。

主成分1	主成分2	主成分3	主成分4	主成分5	主成分6	主成分7
8.07	3.02	2.20	1.74	1.62	1.41	1.00

- 第一主成分：主要与访问量，url数量，流量，url top1-6及个别敏感url相关。可代表常用功能节点访问行为情况。
- 第二主成分主要与敏感url1、url6、url7相关，可代表敏感功能节点访问行为情况。
- 第三主成分主要与流量cv，url top7、敏感url4相关，可代表流量离散度情况。
- 第四主成分，主要与url4-6、BL1、BL2、BL3相关，可代表用户敏感节点业务逻辑情况。

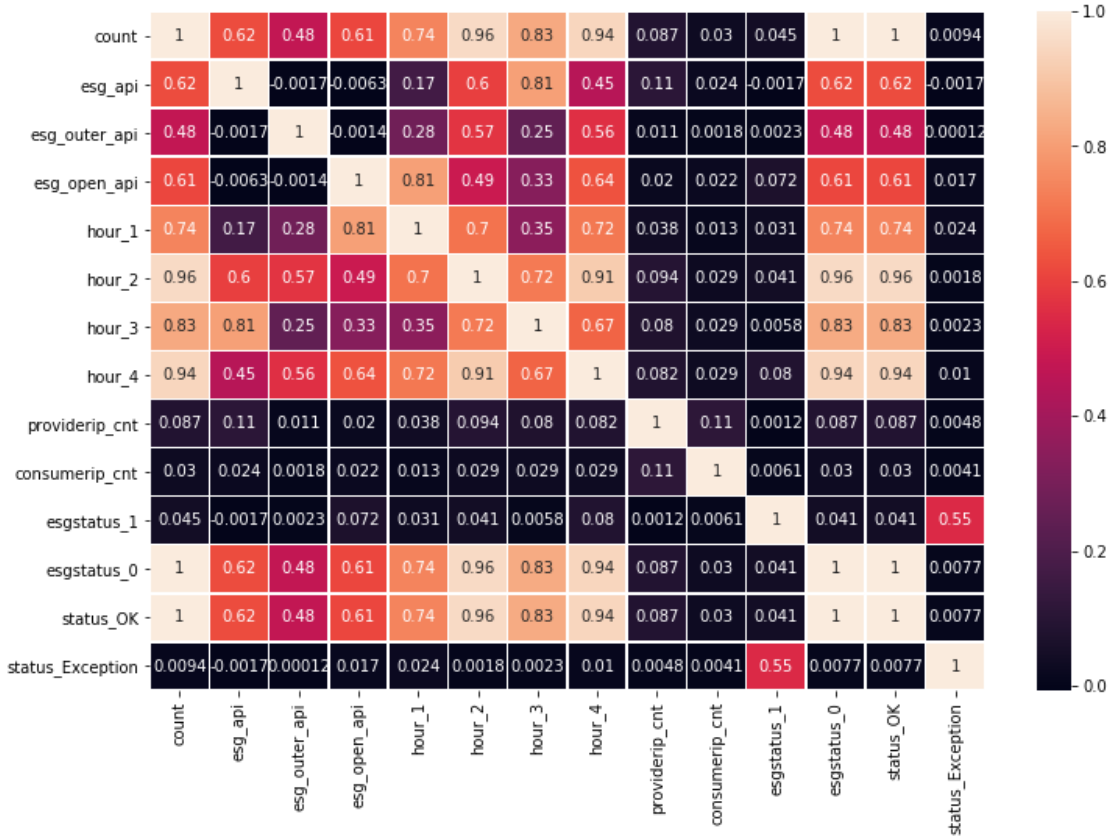
特征成分提取

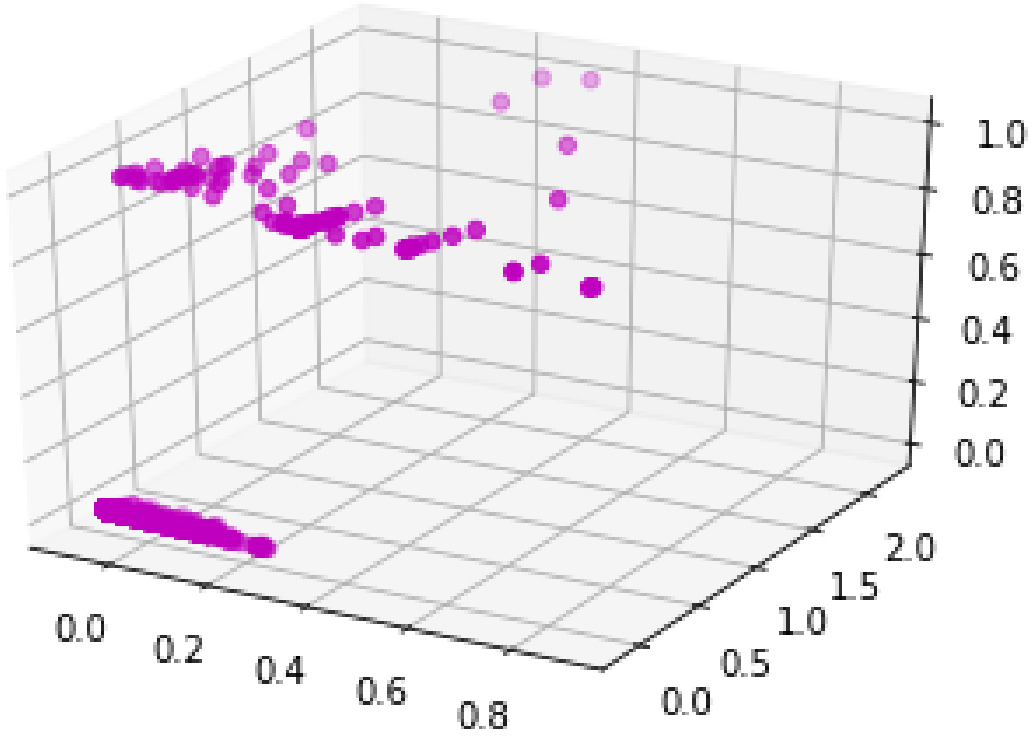

url访问数量异常
访问量偏离该用户历史行为
url平均访问量异常
访问量偏离该用户所在群组历史行为
敏感功能节点访问比值1异常
IP变动异常
user_ obabil lable 特征1 特征2 特征3 特征4 特征5 特征6
PENGC 1 1
PENGC 1 1
PENGC 1 1
PENGC 1 1
PENGC 1 1
WUJIN 1 1
BAOYI 1 1
GENGJ 1 1
GUOME 1 1
GUOME 1 1
GUOME 1 1
JIZIH 1 1
GUOME 1 1
WANGZ 1 1
PENGC 1 1
AOKAI 1 1
PENGC 1 1
LIUSH 1 1
PENGC 1 1
LIUCH 1 1
PENGC 1 1
LIUCH 1 1
WANGJ 1 1
LIUSH 1 1
LIZHA 1 1
CHENG 1 1
SUQIA 1 1
PENGC 1 1
XINGZ 1 1
PENGC 1 1
GAOCU 1 1
LENGJ 1 1
PENGC 1 1
HEYUN 1 1
PENGC 1 1
JIANG 1 1
GUOQI 1 1
XINGZ 1 1
YAOZH 1 1
WANGZ 1 1
ZHANG 1 1
ZHANG 1 1
PENGC 1 1
ZHANG 1 1
DINGH 1 1
ZHANG 1 1
ZHANG 1 1
XINGZ 1 1
ZHANG 1 1
ZHANG 1 1
WANGJ 1 1
AOKAI 1 1
GUOQI 1 1
YAOZH 1 1
WANGZ 1 1
GUOQI 1 1
YANGZ 1 1
WANGZ 1 1
WANGK 1 1
PENGC 1 1
YANGZ 1 1
ZHANG 1 1
WANGZ 1 1
WANGZ 1 1
XUPEI 1 1
PENGC 1 1
WANGJ 1 1
QUYUN 1 1
HUANG 1 1
WANGZ 1 1
XINGZ 1 1
CHENG 1 1
XINGZ 1 1
PENGC 1 1
LENGJ 1 1
PENGC 1 1
JIZIH 1 1
WANGZ 1 1
LIUSH 1 1
PENGC 1 1
HEYUN 1 1
XINGZ 1 1
YAOZH 1 1
WANGZ 1 1
CHENK 1 1
PENGC 1 1
PENGC 1 1
WANGJ 1 1
ZHANG 1 1
XUERH 1 1
WUCUI 1 1
HEYUN 1 1
WANGZ 1 1
GAOCU 1 1
SHANG 1 1
PENGC 1 1
HEYUN 1 1
LIUYA 1 1
PENGC 1 1
PENGC 1 1
PENGC 1 1
CHENG 1 1
CHENG 1 1
XINGZ 1 1
CHENG 1 1
PENGC 1 1
WUJIN 1 1
XING 1 1
WAN 1 1
PENG 1 1
LUXI 1 1
DING 1 1
GAO 1 1
LIYAI 1 1
HEYU 1 1
DING 1 1
XING 1 1
LIUCI 1 1
PENG 1 1
WAN 1 1
SHAN 1 1
WAN 1 1
WAN 1 1
GUO 1 1
PENG 1 1
ZHAN 1 1
ZHAN 1 1
WAN 1 1
PENG 1 1
PENG 1 1
WAN 1 1
XUER 1 1
XING 1 1
YANG 1 1
PENG 1 1
LIMIN 1 1
WAN 1 1
WAN 1 1
ZHOU 1 1
ZHAN 1 1
PENG 1 1
PENG 1 1
YAOZ 1 1

API接口安全监控场景

诉求：应用系统接口-关注是否存在接口异常调用

问题：请求量巨大，多基于统计性变量参数，业务模式较为单一，场景相对稳定，寻找一种更为高效、准确的异常预警；





Provide-consumer

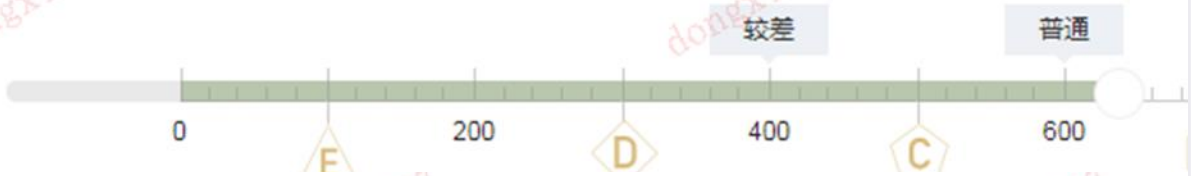
LOGO

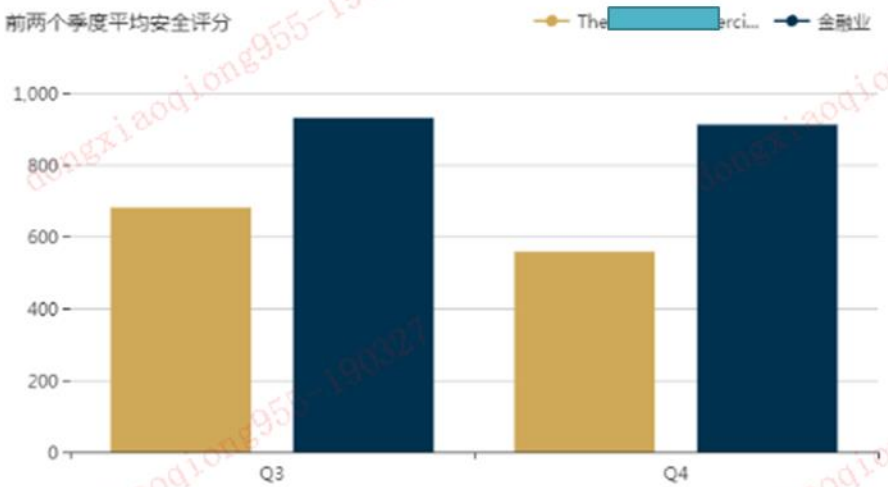
信息安全水平评价/行业水平

敏感信息保护
1000
A

服务应用安全
410
C

域名标识安全
1000
A

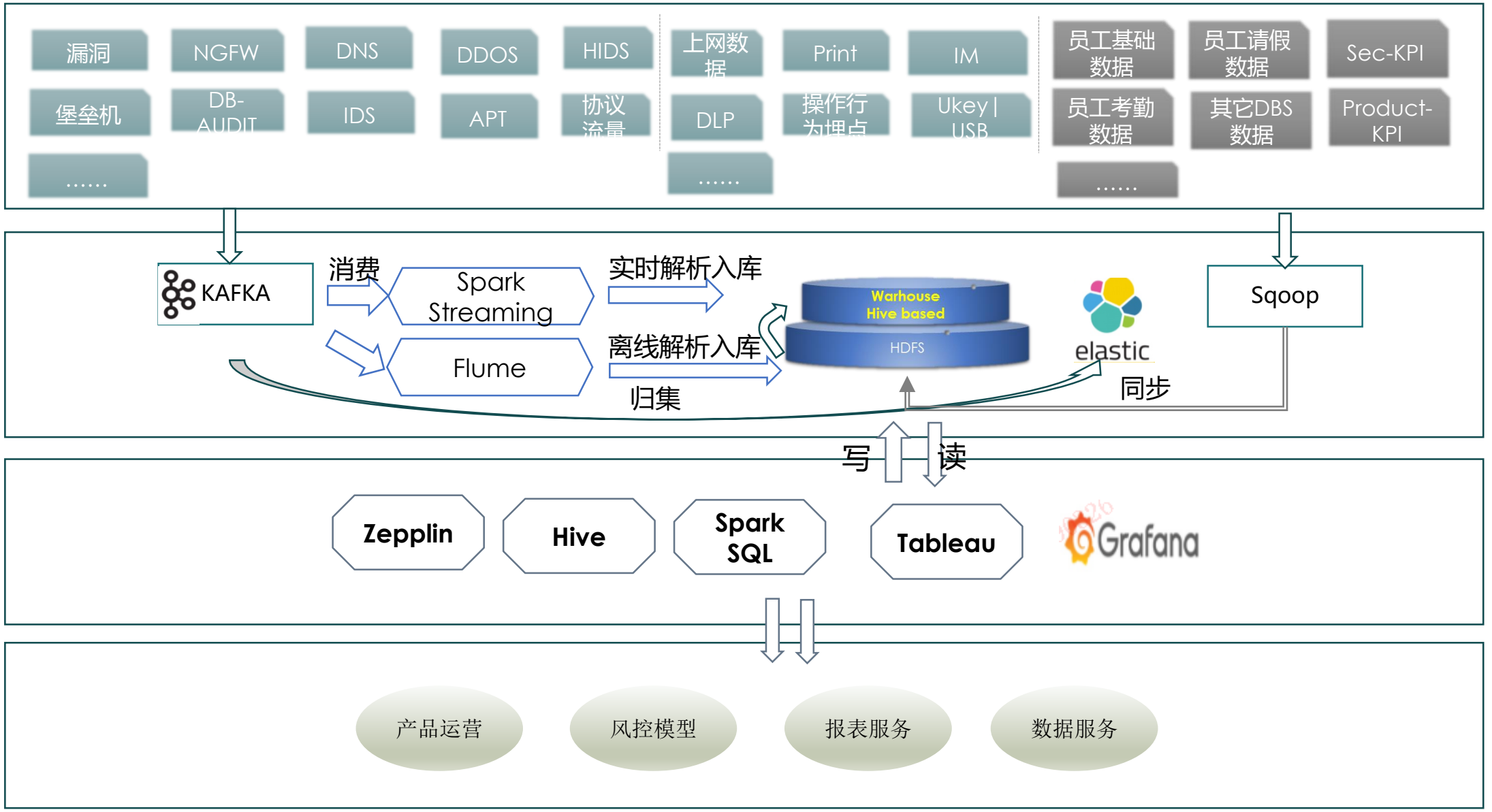




供应商风险量化

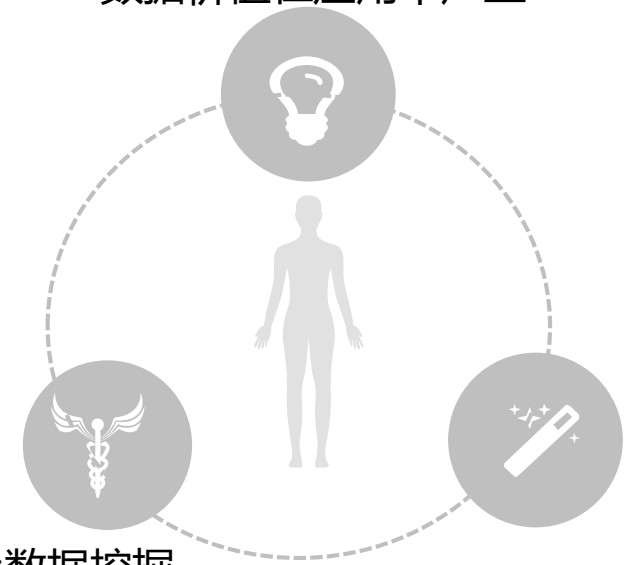
合作供应商的信息安全风险水平

数据分析平台


漏洞
NGFW
DNS
DDOS
HIDS
上网数据
Print
IM
员工基础数据
员工请假数据
Sec-KPI
堡垒机
DB-AUDIT
IDS
APT
协议流量
DLP
操作行为埋点
Ukey | USB
员工考勤数据
其它DBS数据
Product-KPI
……
……
……
KAFKA
消费
Spark Streaming
Flume
实时解析入库
离线解析入库
归集
Warhouse Hive based
HDFS
elastic
Sqoop
同步
写
读
Zepplin
Hive
Spark SQL
Tableau
Grafana
产品运营
风控模型
报表服务
数据服务

有价值的数据闭环与输出

数据成本与流动
数据建设的计划投入
聚焦“关注”
数据价值在应用中产生



数据与价值闭环管理
数据质量与可用
完整的数据闭环，能让数据挖掘、
数据分析产生价值；

数据价值的输出
结合业务提供有价值数据
基础数据的积累和循环

Thanks