CCTC 2017 中国云计算技术大会
Cloud Computing Technology Conference 2017
CSDN

大数据安全分析在电子银行业的探索与实践

HanSight瀚思 张军

目录

- 大数据发展现状
- 电子银行业的机遇与挑战
- 《网络安全法》的出台为电子银行业保驾护航
- 安全遇到了大数据
- 大数据安全分析架构
- 大数据安全分析在电子银行业的探索与实践

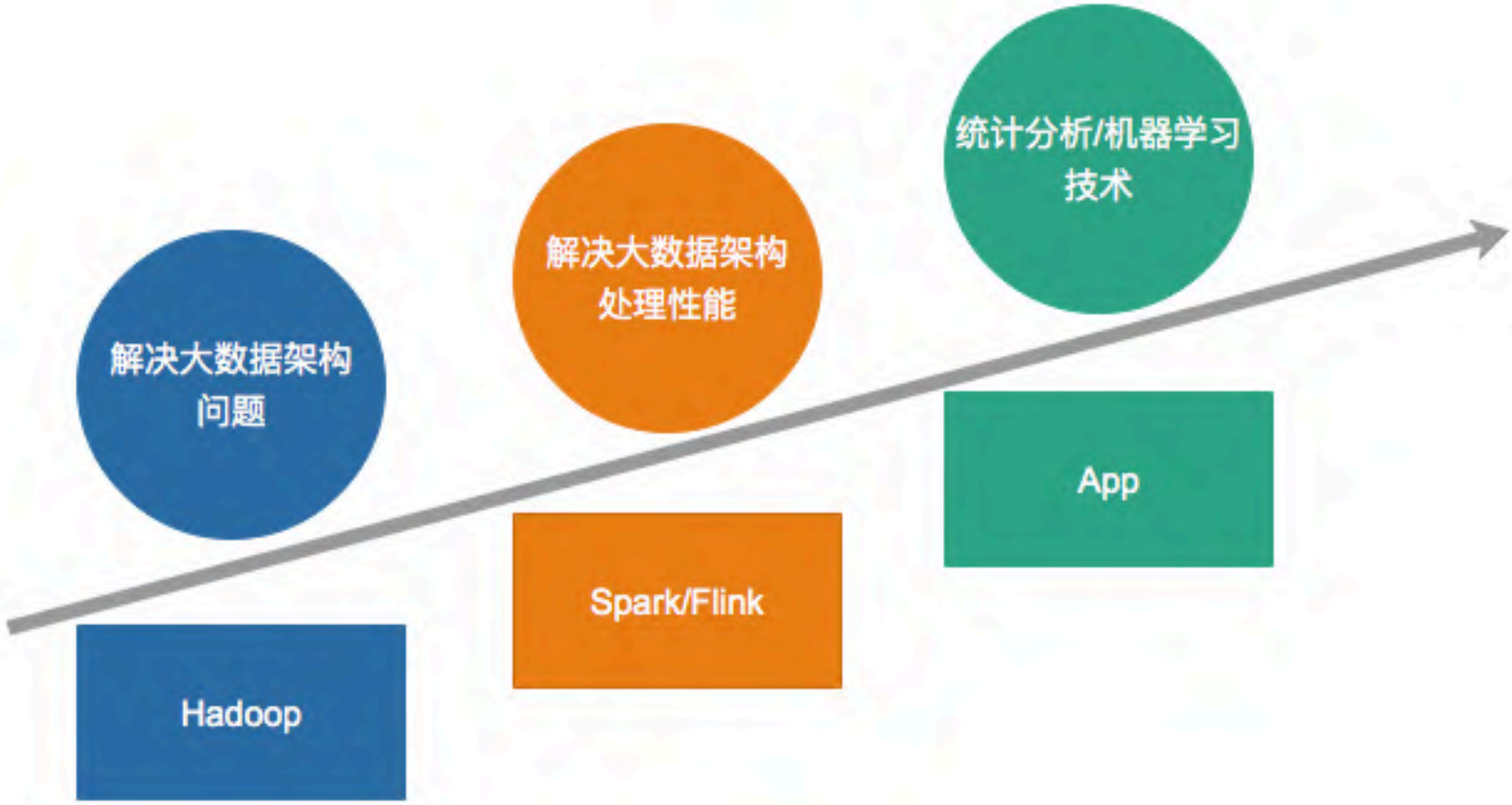
自我介绍

张军

- HanSight瀚思高级技术工程师，负责定制开发
- Hortonworks认证讲师
- 目前主要研究方向是安全分析、大数据基础架构等领域

- 大数据发展现状
- 电子银行业的机遇与挑战
- 《网络安全法》的出台为电子银行业保驾护航
- 安全遇到了大数据
- 大数据安全分析架构
- 大数据安全分析在电子银行业的探索与实践

大数据发展现状



智能安保作为大数据应用的典范之一

- 许多企业将大数据分析纳入安全战略；
- 企业的安全日志数据提供了以往未遂的网络攻击信息，企业可以利用这些数据来预测并防止未来可能发生的攻击，以减少攻击造成的损失；
- 一些公司正将其安全信息和事件管理软件(SIEM)与大数据平台(如Hadoop)结合起来。

大数据处理技术的企业级之路

- 2017年4月国家发布了《大数据安全标准化白皮书》；
- 从法规、政策、标准、应用纬度阐述大数据安全标准化涉及的内容；
- 坚持安全与发展并重的方针，为大数据发展构建安全保障体系

来源：大数据标准化白皮书

- 大数据发展现状
- 电子银行业的机遇与挑战
- 《网络安全法》的出台为电子银行业保驾护航
- 安全遇到了大数据
- 大数据安全分析架构
- 大数据安全分析在电子银行业的探索与实践

电子银行业的机遇与挑战

金融互联网

互联网的发展，使传统银行业获得了新的发展机遇。

电子银行业的机遇与挑战

1 互联网+金融

推动业务不断从线下向线上转移

2 银行保存大量金融和客户信息

不法分子主要攻击目标之一

3 APT攻击

不法分子针对银行的攻击行为突出、趋利性明显，多起专门针对银行的高级持续性威胁攻击事件（APT攻击）陆续被披露，其中以孟加拉央行遭受黑客入侵导致巨额资金被窃事件最为典型

4 盗取和诈骗银行客户资金

买卖个人敏感信息为目的的地下黑色产业链也日具规模，精准网络诈骗行为猖獗，商业银行的经营环境面临着严峻的外部安全挑战

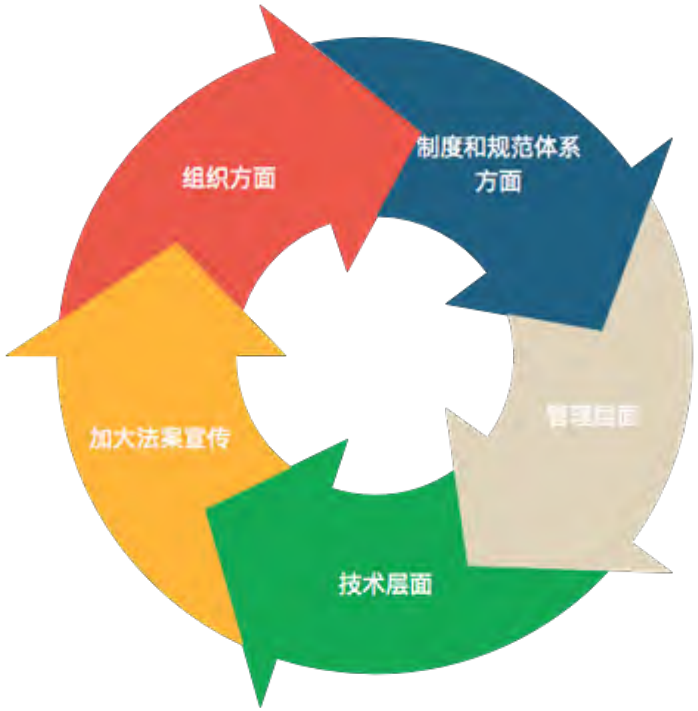
目录

- 大数据发展现状
- 电子银行业的机遇与挑战
- 《网络安全法》的出台为电子银行业保驾护航
- 安全遇到了大数据
- 大数据安全分析架构
- 大数据安全分析在电子银行业的探索与实践

《网络安全法》部分解读

1. 从法律层面把网络安全工作提高到了国家安全战略
2. 网络运营者必须遵守的法律规范
3. 明确了严禁网络诈骗、倒卖个人信息等不法行为
4. 为打击地下黑色产业链、净化网络空间提供了法律支撑
5. 关键信息基础设施重点安全保护要求
6. 明确了国家层面建立网络安全检测预警和信息通报制度

商业银行贯彻和落实《网络安全法》的措施与建议



组织方面	严格按照法案需求确定安全管理负责任人
制度和规范体系方面	积极开展法案的解读和学习，做好法案在商业银行落地的分析和解读
管理层面	建立安全事件应急和响应、通报等安全风险全流程闭环管理机制
技术层面	做好网络边界防御、应用安全防护、终端安全防护等边界防护工作
加大法案宣传	加大银行内部从管理层到基层员工全方位的宣传和培训

来源：中国网信网

目录

- 大数据发展现状
- 电子银行业的机遇与挑战
- 《网络安全法》的出台为电子银行业保驾护航
- 安全遇到了大数据
- 大数据安全分析架构
- 大数据安全分析在电子银行业的探索与实践

从大数据到安全大数据分析

- 当前信息安全领域，正在面临多种挑战；
- 一方面，企业安全架构日趋复杂，各种类型的安全设备、安全数据越来越多，传统的分析能力明显力不从心；
- 另一方面，以APT为代表的新型威胁的兴起，内控与合规的深入，越来越需要储存与分析更多的安全信息，并且以更加快速的做出判定和响应；
- 2012年3月，Gartner发表了一份题为『Information Security Is Becoming a Big Data Analytics Problem』的报告，表示信息安全问题正在变成一个大数据分析问题，大规模的安全数据需要被有效地关联、分析和挖掘，并预测未来将出现安全分析平台，以及部分企业在未来五年将出现一个新的岗位--“安全分析师”或“安全数据分析师”。

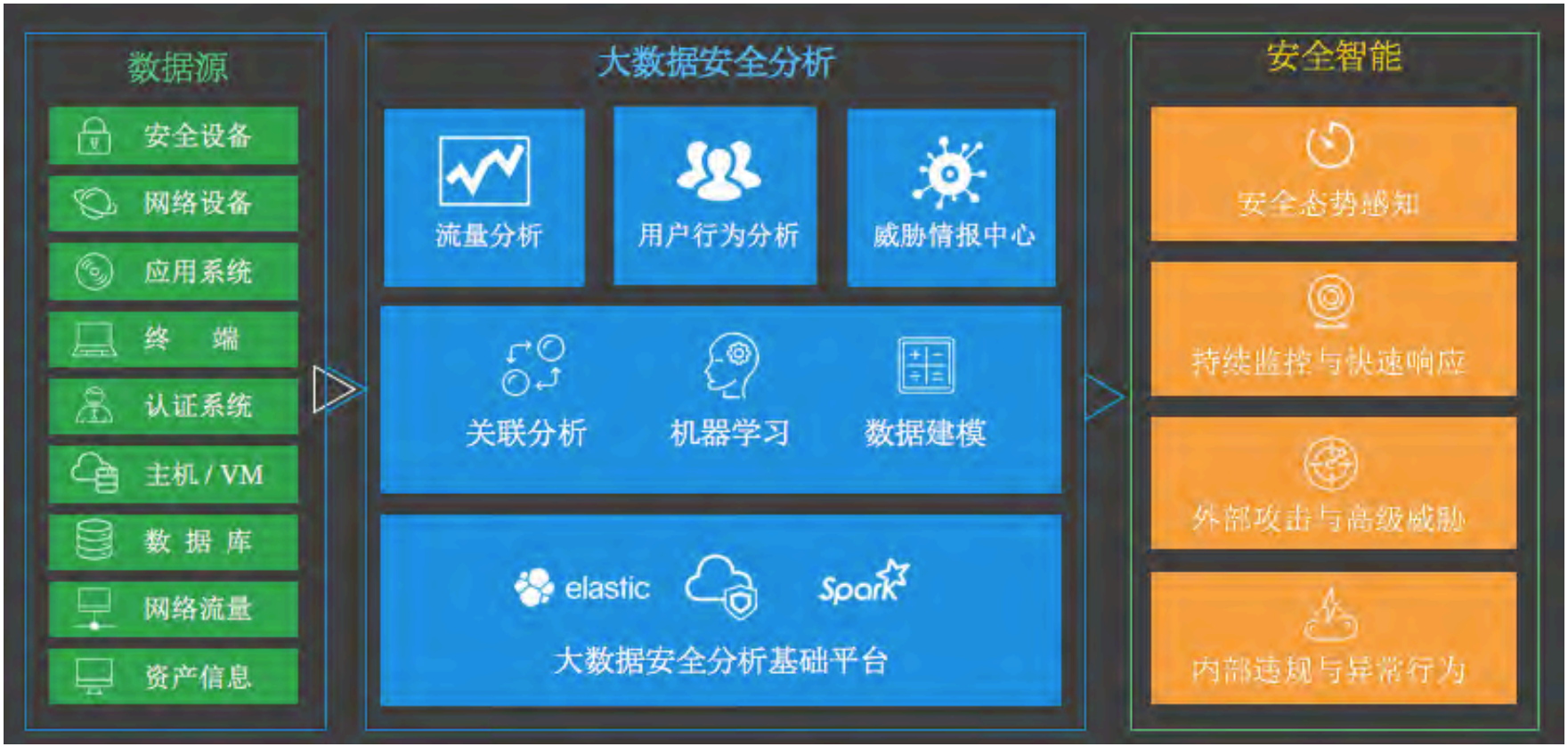
从大数据到安全大数据分析 续...

- 现在，来自IT环境的各类日志也加入了大数据的队伍，但企业的CIO们也意识到从日志中获取有价值的决策信息并不简单:
  - 海量日志内容
  - 日志报警缺乏关联性
  - 大量的误报信息
  - 多种控制台界面
- 在以往，了解难以察觉的安全威胁会耗费数天甚至数月的时间，因为大量的互不相干的数据流难以形成简明、有条理的事件“拼图”；
- 采集和分析的数据量越大，看起来越混乱，重构事件所需的时间也越长。如果攻击快速且凶猛(例如拒绝服务攻击或快速传播的蠕虫)，花数天或数月诊断问题会带来巨大的合规和财务影响。哪些资产真正处于威胁风险中，哪些资产有补救控制或应对措施?要知道这些问题，管理员需要监控所有系统的安全状况，包括访问其网络的移动设备和个人拥有设备， 并及时确定优先级和补救措施。 研究报告证实，只有35%的企业可以快速检测安全漏洞，多数商业机构都缺乏驾驭大数据的安全力量。

目录

- 大数据发展现状
- 电子银行业的机遇与挑战
- 《网络安全法》的出台为电子银行业保驾护航
- 安全遇到了大数据
- 大数据安全分析架构
- 大数据安全分析在电子银行业的探索与实践

通用架构


数据源
安全设备
网络设备
应用系统
终　端
认证系统
主机 / VM
数 据 库
网络流量
资产信息
大数据安全分析
流量分析
用户行为分析
威胁情报中心
关联分析
机器学习
数据建模
elastic
Spark
大数据安全分析基础平台
安全智能
安全态势感知
持续监控与快速响应
外部攻击与高级威胁
内部违规与异常行为

目录

- 大数据发展现状
- 电子银行业的机遇与挑战
- 《网络安全法》的出台为电子银行业保驾护航
- 安全遇到了大数据
- 大数据安全分析架构
- 大数据安全分析在电子银行业的探索与实践

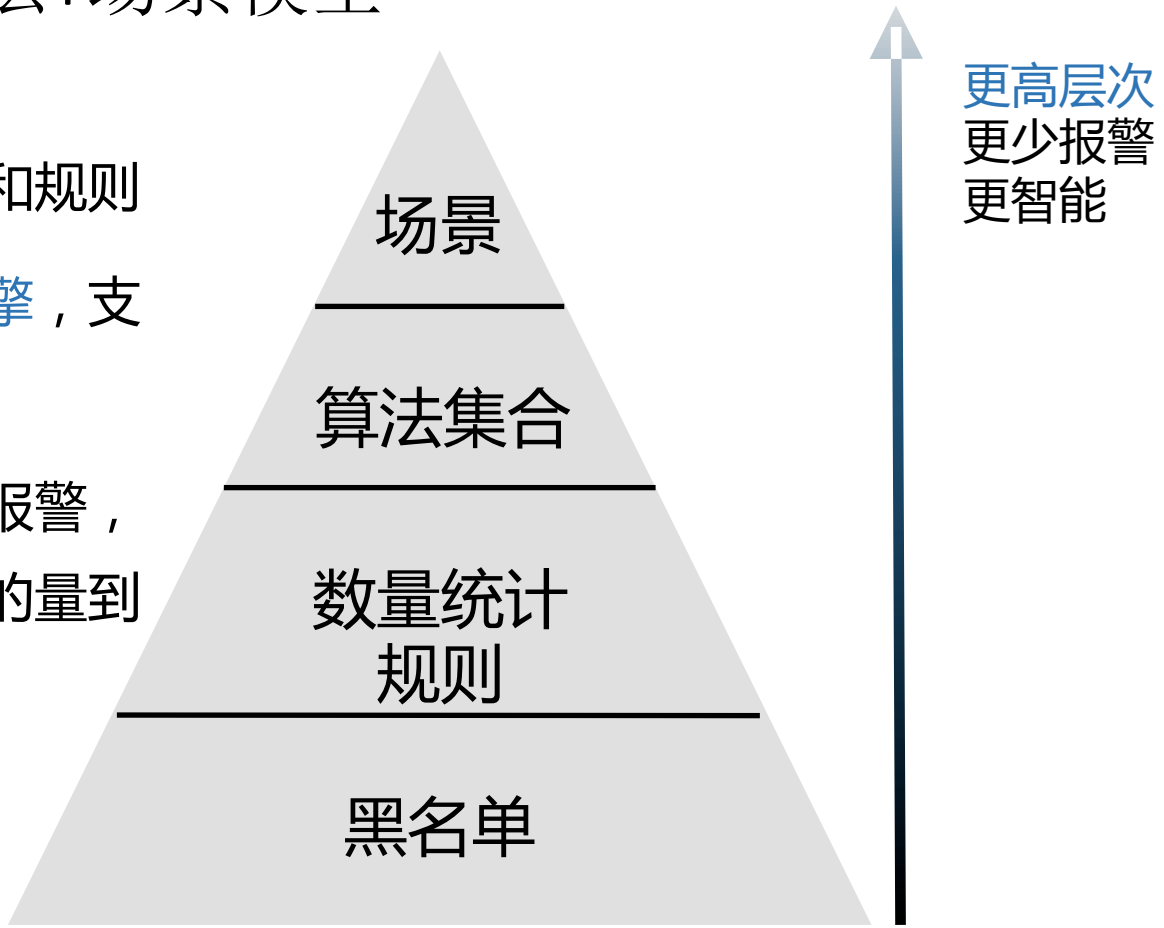
背景

- 电子银行网上银行业务每天会产生大量的日志
- 日志涉及访问日志及交易日志
- 访问日志是WEB服务器产生的日志
- 交易日志是应用输出的日志，包括登录的帐号、设备、登录的状态、登入的时间、交易情况等

- 根据日志数据对用户画像，包括撞库行为画像及扫描行为画像

黑名单+规则+算法+场景模型

- 支持用户熟悉的黑名单和规则
- 新一代黑名单和规则引擎，支持用户用表达式定义
- 以场景为单位聚合海量报警，大幅减少用户处理数据的量到1/50

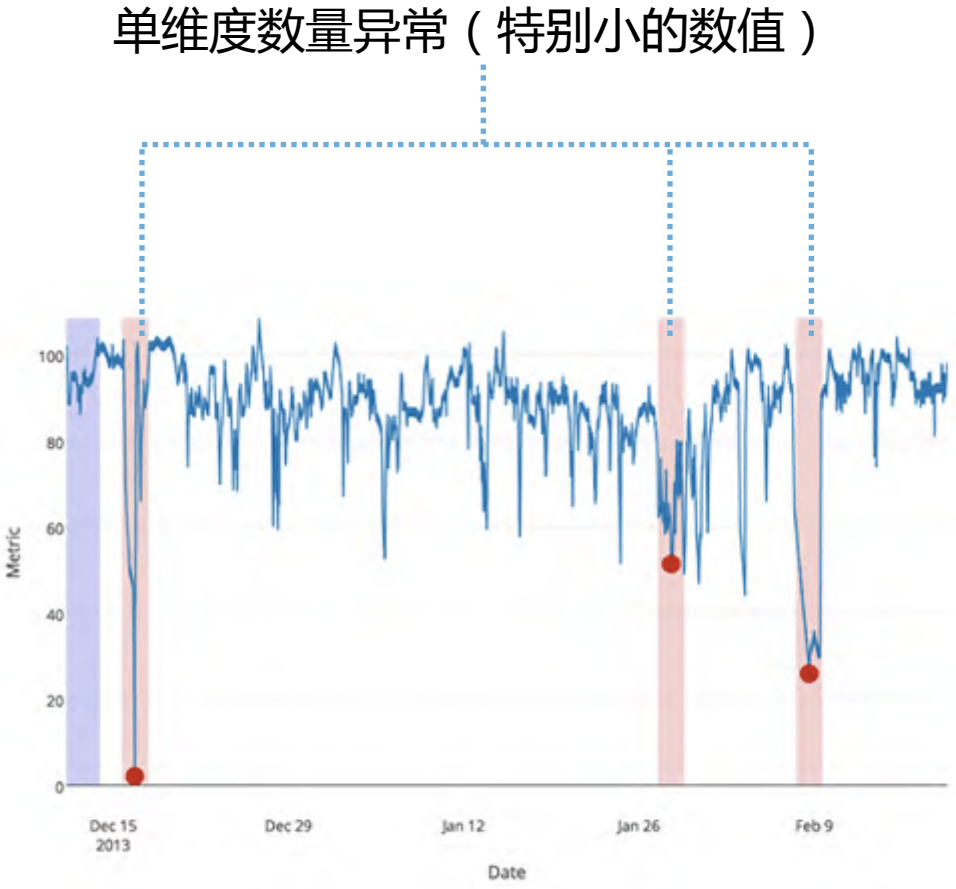


内置行为场景库

- 账户异常
  - 账号异常
  - VPN使用异常
  - 终端使用异常
- 应用服务器异常
  - 文件服务器
  - 邮件服务器
  - 数据服务器
- 互联网使用异常
  - 软件服务
  - 邮件服务
  - 网盘服务
  - IM服务
  - 个性化黑名单
- 网络流异常
- 访问日志异常
  - WEB

何为“基线”

- 把用户多维度行为拆成一个个维度分析，比如用户行为细分为：
  1. 每天访问日志中404数量
  2. 每天使用哪个设备登录
  3. 每天访问交易类型
  4. 每天访问序列
  5. …
- 每个维度都进行模式计算，偏离模式即是异常
- 单一维度数量异常很容易判断



多纬度三大类基线

1. 数量：1小时登陆多少次
2. 关系：登陆使用设备以前是否用过
3. 序列：登陆后依次进行哪些操作

- 三大类全部支持超长周期（>6个月）分布式（>10000用户）计算
- 全部支持用户界面上自定义

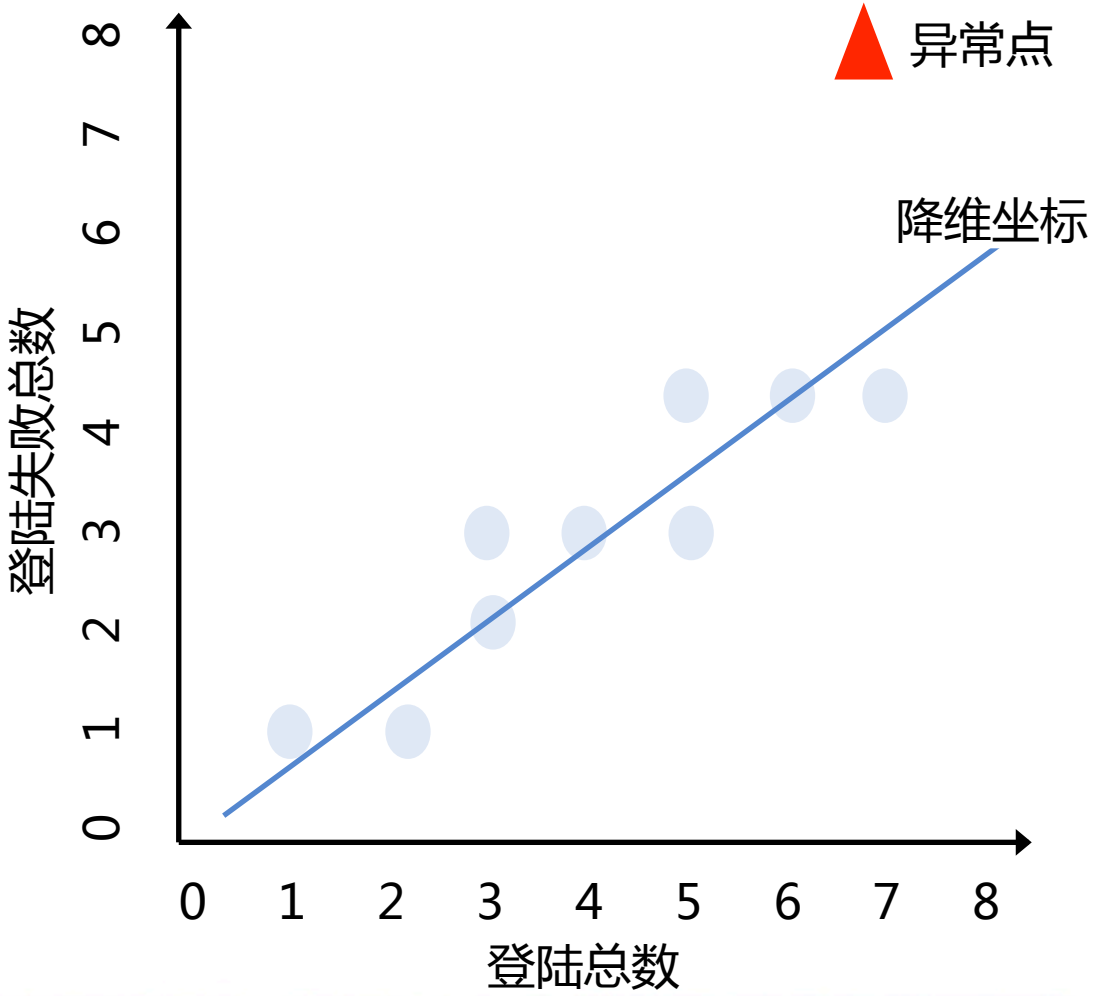
安全算法选择

- 安全分析以无监督学习（异常检测）为主，因为客户缺乏标记数据
- 有人工辅助的半监督学习（安全专家、运维人员反馈）
- 算法结构：

无监督异常分析 — 人工确定异常 — 产生标记样本 — 半监督学习

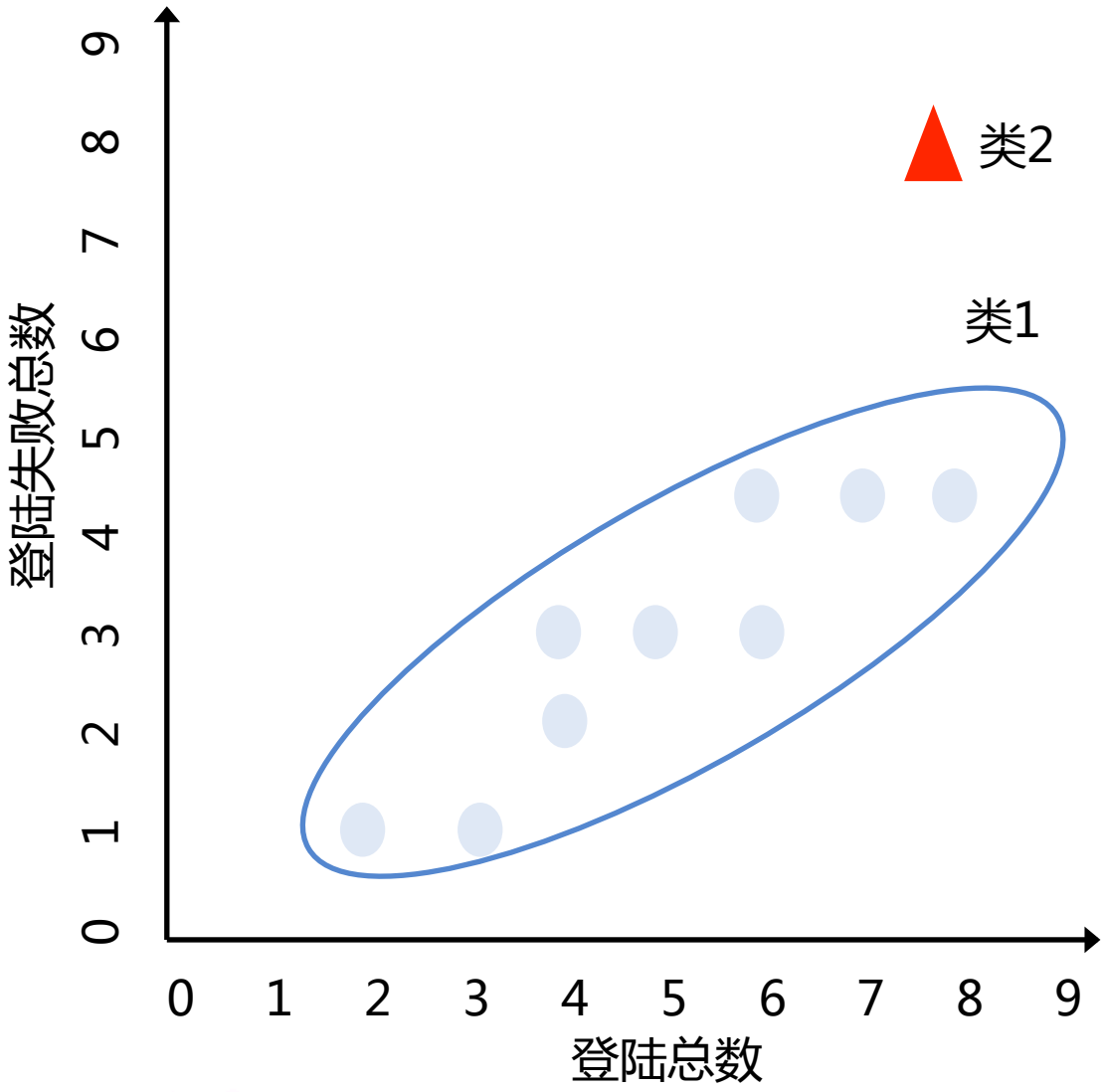
数量异常如何使用算法计算

- 例子：从二维降到一维，离一维坐标过远的就是异常
- X轴：登陆总数
- Y轴：登陆失败总数
- 正常：离降维坐标近的点
- 异常：离降维坐标远的点 -> 登陆失败总数相比登陆总数过大的用户



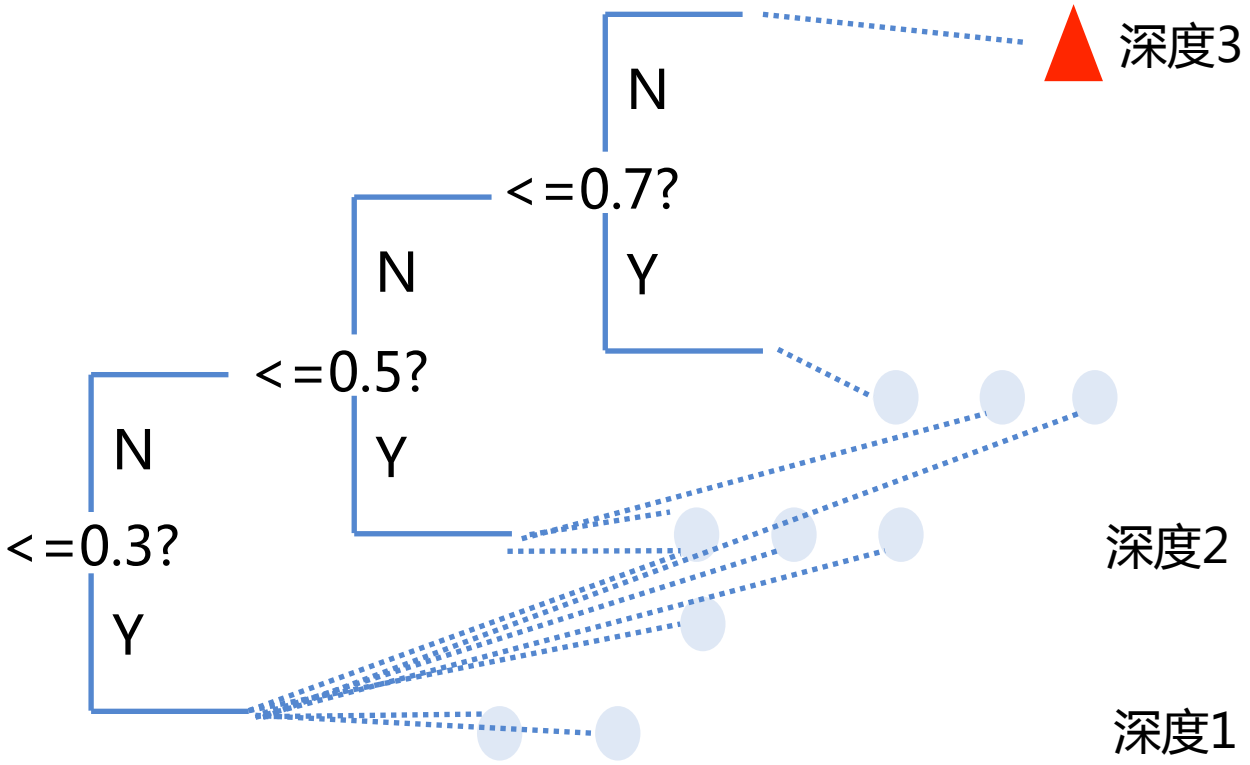
聚类找异常

- 例子：按照点密集程度分成2类
- 正常：属于大类
- 异常：不属于任何类或者属于很小的类 -> 登陆失败总数相比登陆总数过大的用户



决策树找异常

- 例子：构建决策树看多少个决策后才能定位到点
- 决策节点值：登陆失败总数 / 登陆总数
- 正常：决策树浅的点，越浅代表越常见，只需要少量信息就能描述
- 异常：决策树深的点 -> 登陆失败总数相比登陆总数过大的用户



架构细节：实时分布式流处理平台

- 基于实时大规模分布式处理系统Flink开发
- Flink支持银行级的低延迟（<100ms）和超高并发（单机>1,000,000 EPS）

架构细节：数据源>数据字典

- 内置数据字典把各种Elasticsearch输入数据规整化
- 支持表达式做复杂规整化

架构细节：场景引擎

- 单一维度分析误报多，多维度有效降低误报
- 多维度综合打分—通过算法计算出多个维度异常分值的综合值
- 多维度分析支持行为序列
- 多维度还可以用场景定义库

谢谢！

欢迎关注HanSight瀚思微信公众号