elastic | 中国开发者大会

基于 Elasticsearch的 AI 异常检测和画像系统

朱彦安
安恒信息



战略级赞助商 HUAWEI | 钻石级赞助商 普翔 | 白金级赞助商 华夏博格 神州数码 Digital China | 金牌级赞助商 iDataAPI

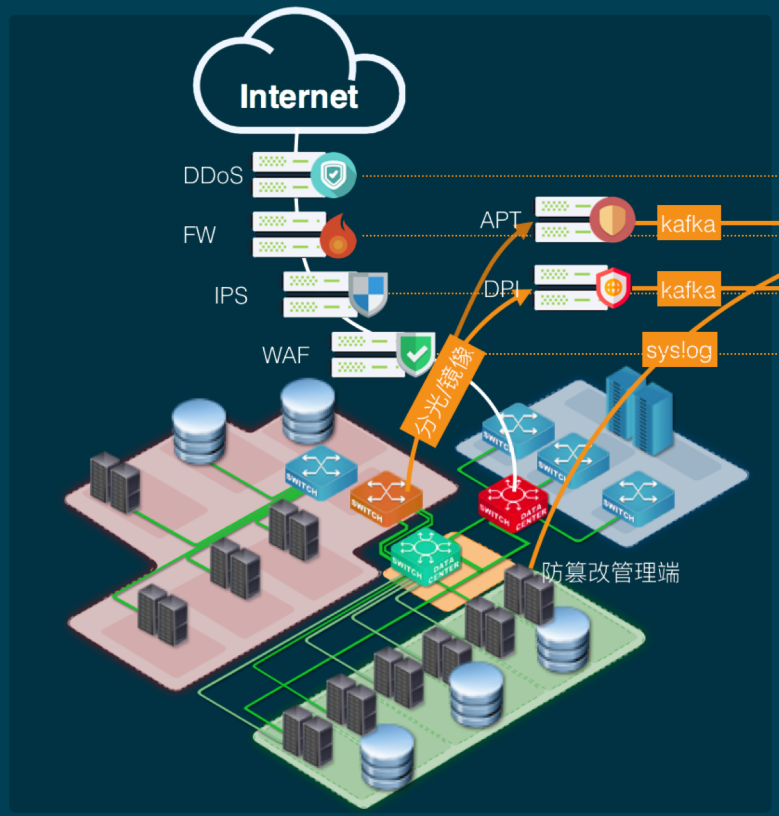
合作伙伴 开源中国 oschina.net 掘金 谱时 IT大咖说 otpub Broadview 百格活动 bagevent.com MAXHUB

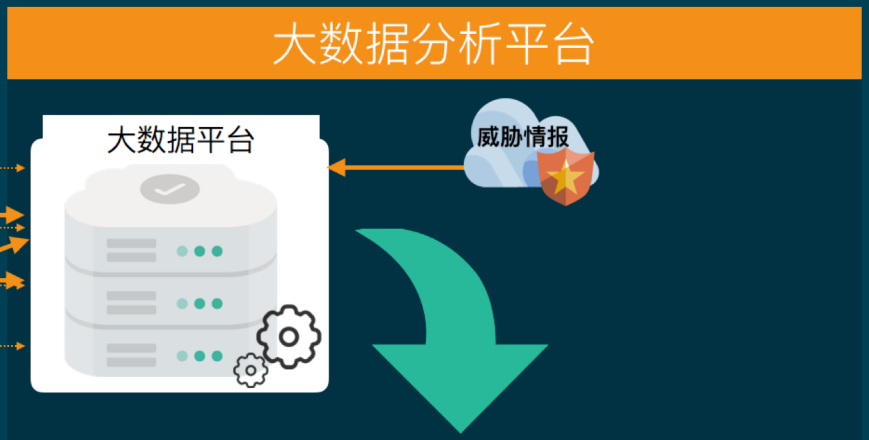
内容大纲

- 简单介绍网络安全要解决的问题
- 技术架构和Elasticsearch的定位
- Elasticsearch帮我们解决了哪些问题
- 运维经验分享

elastic

网络安全威胁分析什么数据？





- 形成全网的安全大数据中心，集成关键安全设备的日志、告警、流量；满足网络安全法存储6个月的法规要求；
- 基于机器学习提供态势感知能力
- 深度关联分析识别影响范围和攻击源
- 具备对资产的自发现能力（内部恶意资产的发现与验证）
- 云端威胁情报的验证能力（海陆空一体的安全防御能力）

elastic

为了解决什么问题？



安全场景模型

- WebShell检测
- 恶意提权检测
- 僵尸主机发现
- 渗透攻击检测
- 恶意操作检测
- 木马回连检测
- DDoS攻击检测
- 潜伏性数据窃取
- ……



智能机器学习

- 0day漏洞检测
- 异常流量检测
- 异常业务操作
- 异常系统登录
- 恶意软件行为检测
- DGA检测
- 垃圾邮件检测
- 恶意访问检测
- ……

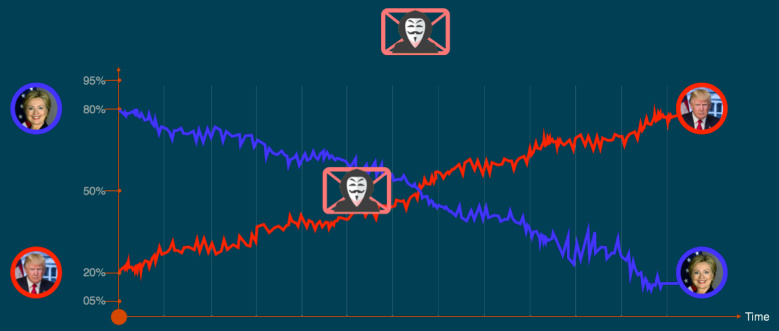
Intrusion Kill Chain模型

建模：发现关键线索　　关联：还原攻击事件

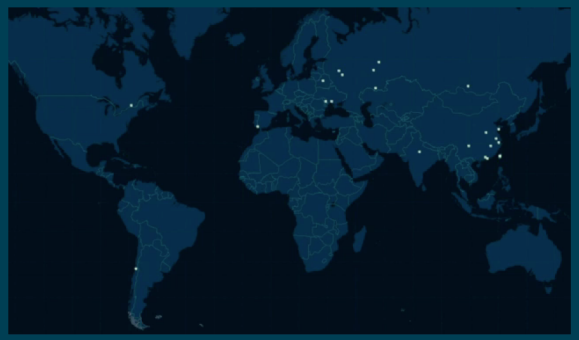


elastic

价值？



2016 邮件门事件





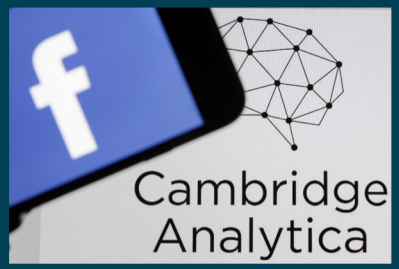
2017 WannaCry



1.4亿用户的信用记录



股价-50%
损失80亿美元



5千万用户的详细信息



公司市值
损失500亿美元

2017 Equifax信用机构用户信息泄漏

2018 Facebook用户信息泄露

elastic

效果

- 实时感知外部和横向威胁态势，检测新型和未知威胁
- 及时发现网络中的攻击团伙和企业内部失陷资产





elastic

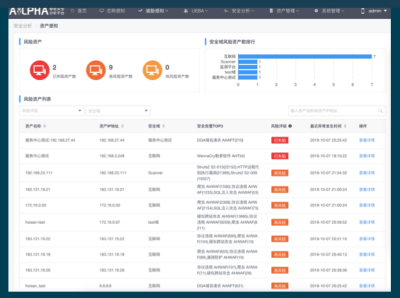
应用方式

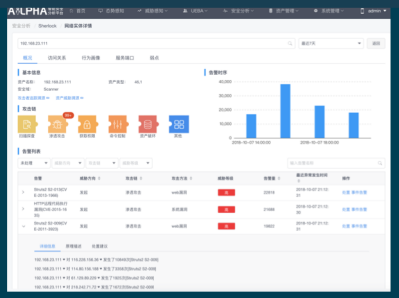
智能分析流程

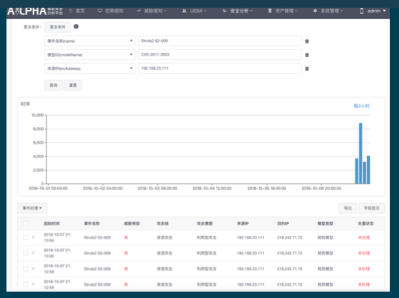
- 掌握网络拓扑和安全域边界，全局感知安全风险
- 锁定风险资产和风险用户，画像分析异常行为和相关风险
- 异常行为举证，辅助威胁情报和漏洞信息，及时处置安全风险

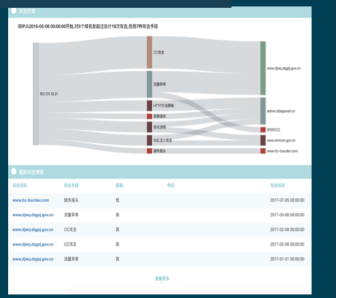








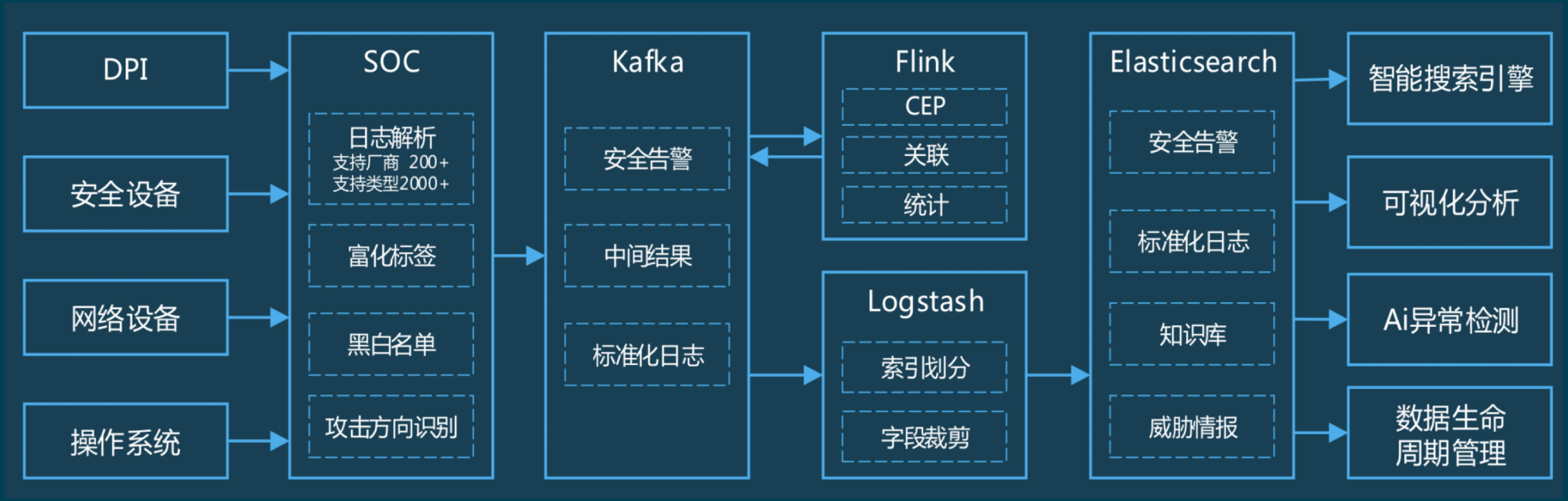




elastic

数据流和技术架构

Elasticsearch是核心



elastic

功能架构

网络安全数据大脑

- index和mapping
- Search、Filter、Aggregations、Snapshot

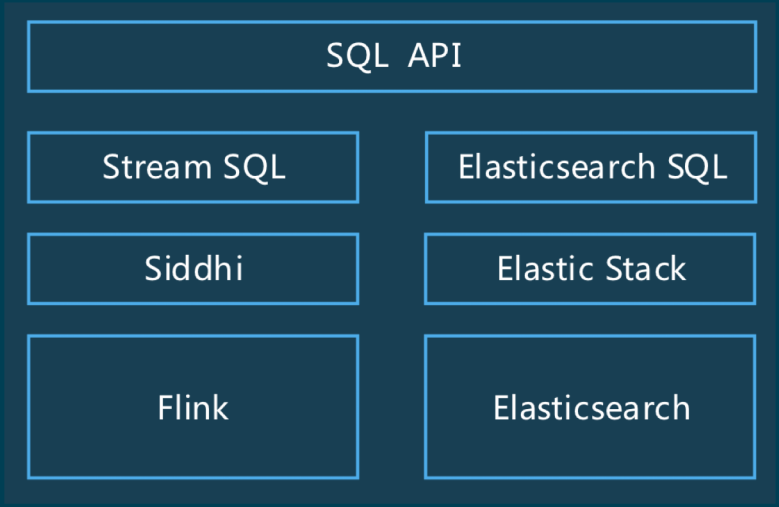


elastic

安全领域应用

统一的SQL分析引擎

- Elasticsearch与Flink结合，流式计算与全文检索引擎整合实现灵活的数据统计分析引擎
- 支持相同SQL标准，同一分析逻辑可支持同时分析实时数据和历史数据
- Flink中对数据进行预聚合，Elasticsearch中对数据进行高度聚合

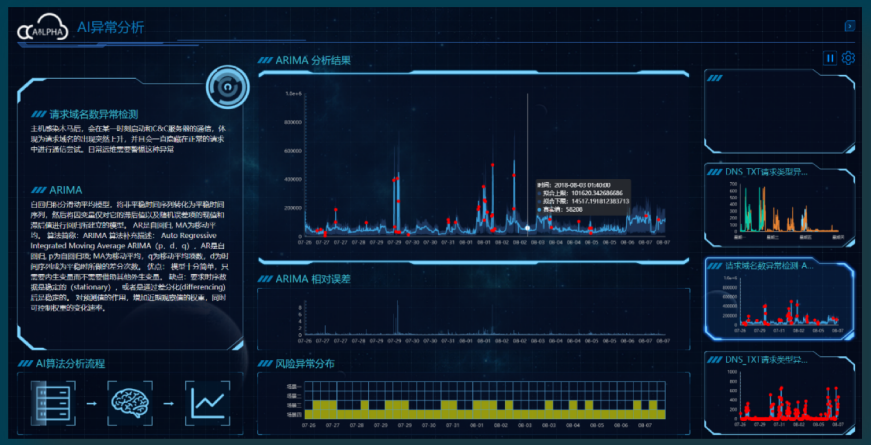


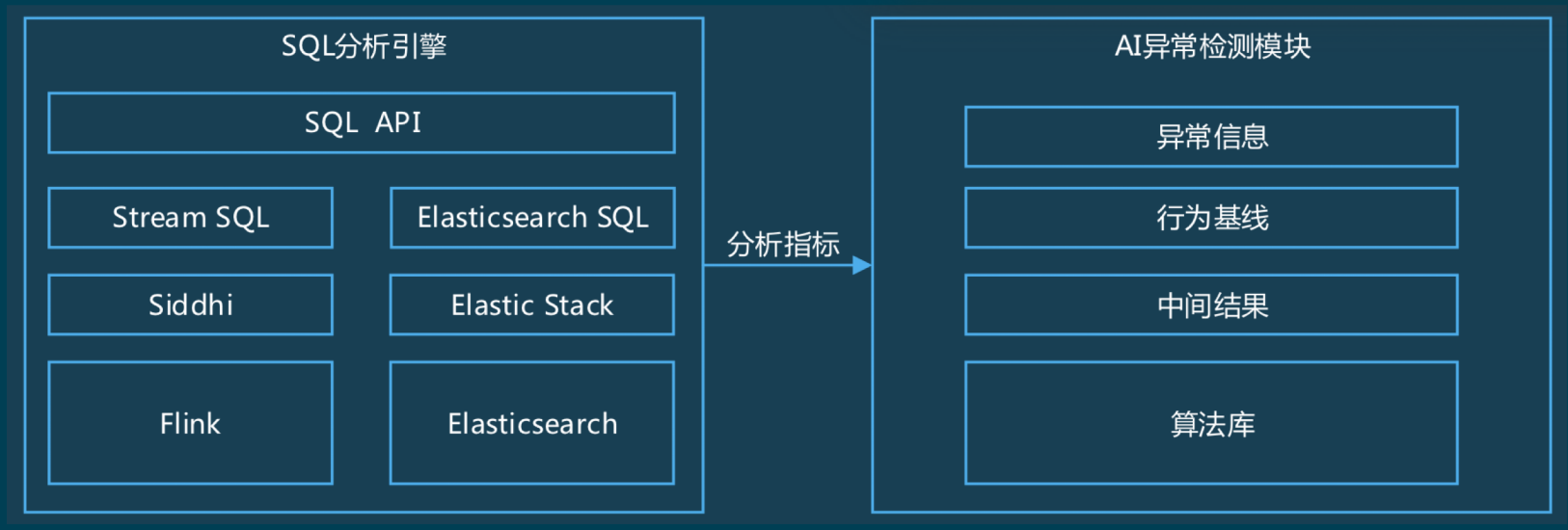
elastic

安全领域应用

Ai异常检测

- SQL引擎提供实时和历史指标
- 创建用户和机器行为基线
- 多AI算法综合判定异常信息



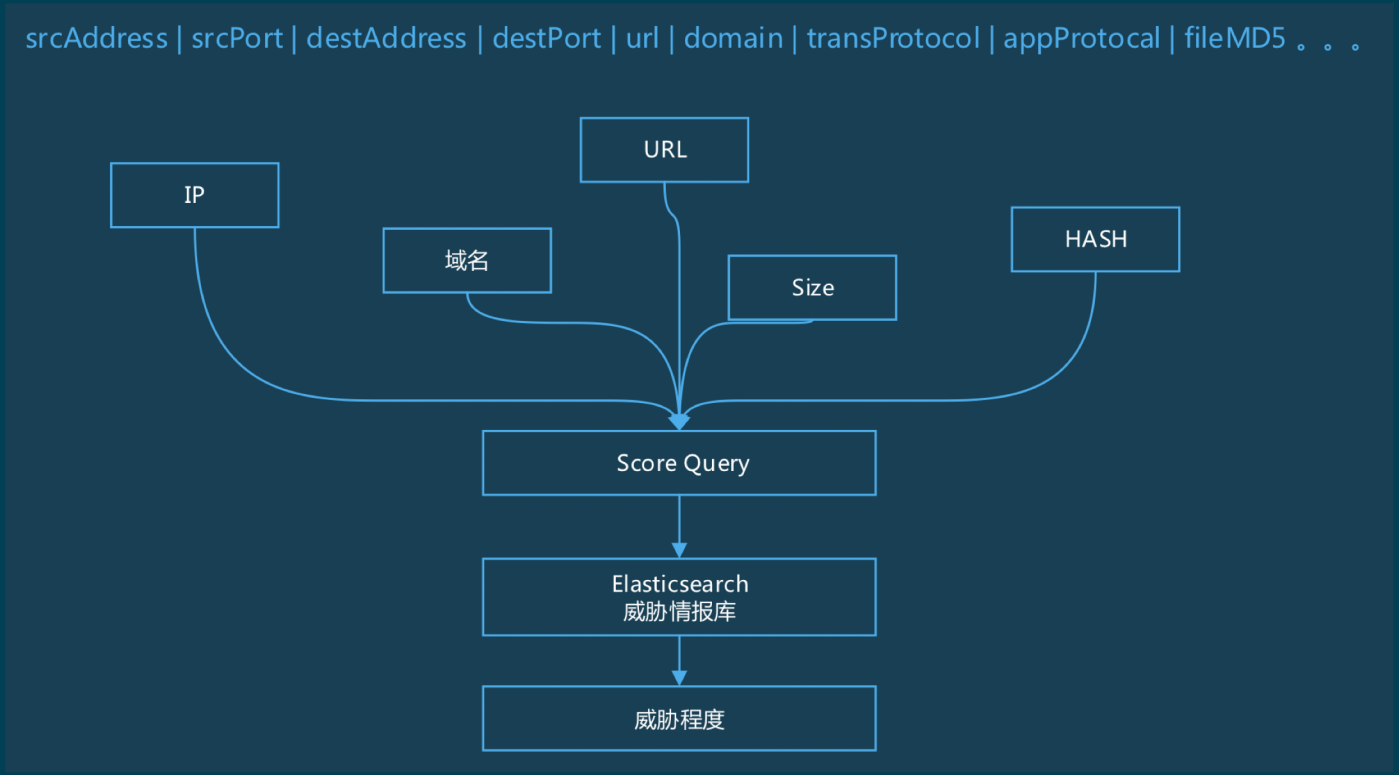


elastic

安全领域应用

威胁情报碰撞

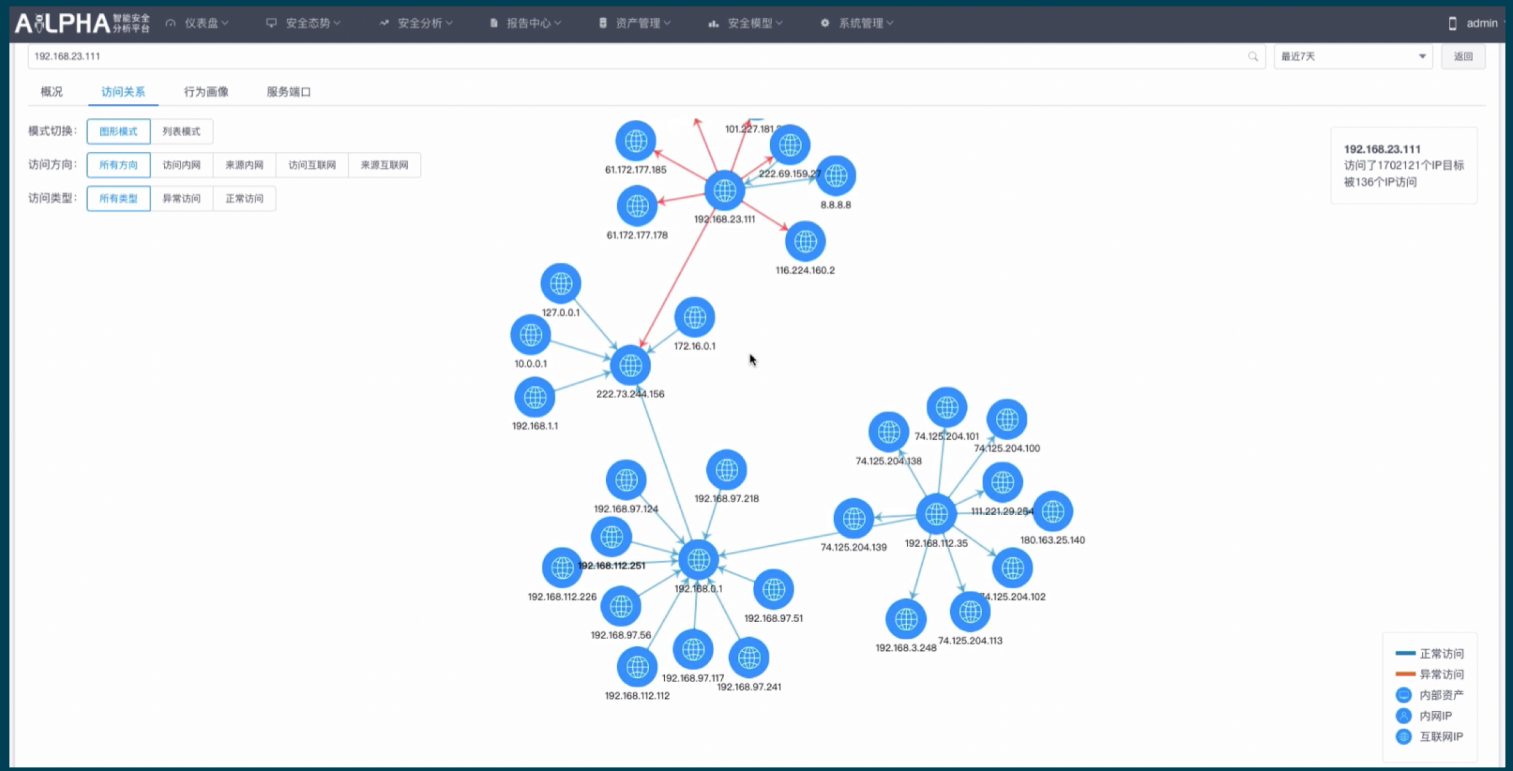
- 情报IOC
- 情报置信度
- Score Query



elastic

安全领域应用

透视网络访问关系

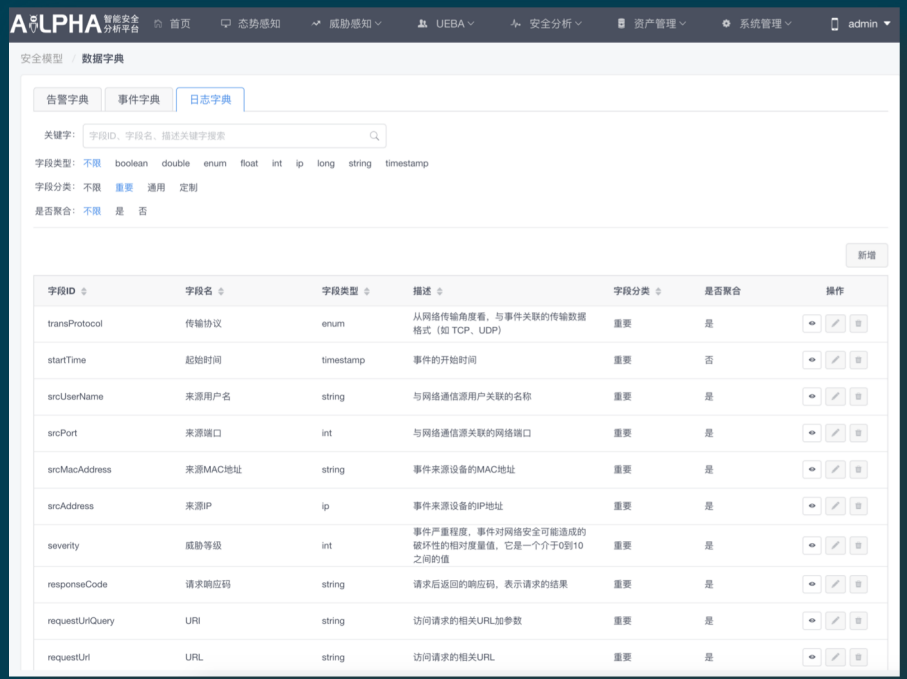


elastic

安全领域应用

字段很多、表很多 —— 数据字典管理

- 规范字段定义和类型，避免ES索引膨胀
- 定义枚举类型字段，提升数据写入性能和可读性

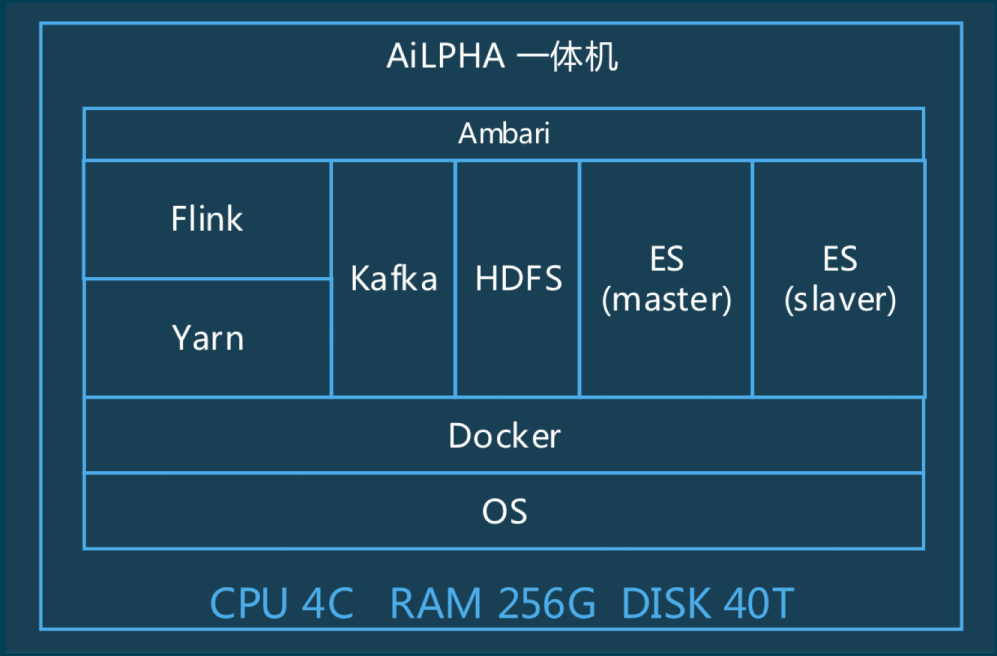




运用Elasticsearch技术的产品化之路

一体机版

- 单机资源的充分利用与资源合理分配
- ES性能监控，性能调优
- 1.5万 EPS



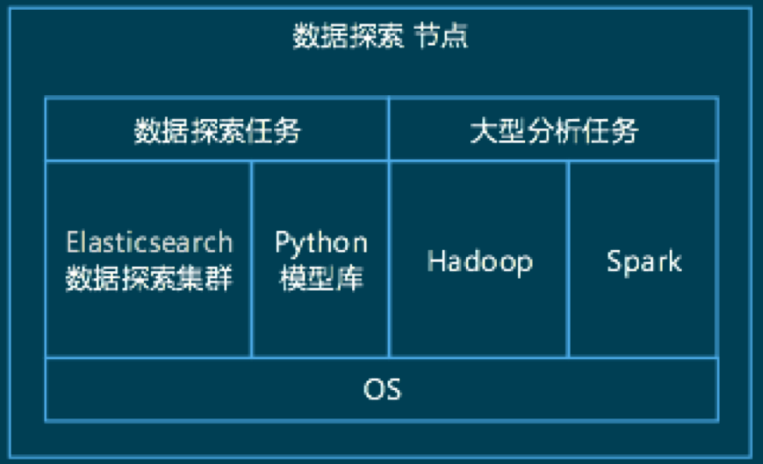
elastic

运用Elasticsearch技术的产品化之路

数据中心版

- 节点角色划分，威胁狩猎节点偏重实时流数据处理和检索，数据探索节点偏重离线数据批量分析
- 威胁狩猎节点，利用Elasticsearch的实时索引、即席查询特性
- 数据探索节点，利用Elasticsearch相对于Hadoop的快速、灵活、自定义聚合的分析特性





elastic

运维心得

Elasticsearch技术实践面临的挑战

- 单进程ES无法充分利用服务器资源
- 多进程ES存在资源抢占
- ES大量数据写入时影响查询性能
- 集群统一运维监控难

解决方法

- Elasticsearch精细化管控服务 -- Elasticsearch Manager
- Docker化实现资源隔离
- Ambari对ES、Logstash统一运维管理

elastic

Elasticsearch精细化管控服务

Elasticsearch Manager 数据写入阶段

- ES索引动态分片分区
  - 依据历史数据量合理规划未来时段的索引分片数，如工作日分片数较多，休息日分片数较少
  - 分片数：（预测未来数据数量 * 平均数据大小）/(20~30G)，
  - 分区数：按时间（按小时，天，周，月和年），按业务属性（资产，类型，等等）
  - 索引分片自动合并：索引批量合并（reindex API），自动评估数据数据以及大小来收缩索引（Shrink Index API）
- ES索引副本和刷新频率控制
  - 索引写入时无副本，依靠kafka保证数据安全，索引完毕后创建副本
  - 实时性要求不高的索引关闭刷新，调节刷新频率为 -1

```
PUT ailpha-securitylog-20180928
{
    "index" : {
        "refresh_interval" : "-1"
    }
}
```

  - 索引写入完成时检查segment数量并执行forcemerge

elastic

Elasticsearch精细化管控服务

Elasticsearch Manager 统计分析阶段

- ES数据Rollup
  - 索引数据自动Rollup，对于流计算平台local agg的数据通过rollup进一步缩减数据量
  - 官方Rollup APIs
- ES索引故障自动恢复
  - 所有索引使用别名机制
  - ES自身日志接入，匹配常见异常，自动化处理
- ES索引生命周期管理
  - 索引自动关闭最不经常使用的索引，采用LFU算法
  - 索引自动备份（snapshot API）

elastic

Elasticsearch精细化管控服务

Elasticsearch Manager 集群资源分配

- SSD提升热数据的写入和查询性能，索引Hot-Warm
  - Hot节点SSD

```
PUT /ailpha-securitylog-20180928/
{
  "settings": {
    "index.routing.allocation.require.box_type": "hot"
  }
}
```

  - Warm普通磁盘

```
PUT /ailpha-securitylog-20180928/
{
  "settings": {
    "index.routing.allocation.require.box_type": "warm"
  }
}
```

- 自动平衡索引分片
  - 将集群自身指标采集进入精细化管控服务，来调整分片所在的节点位置。

elastic

运维心得

Docker化实现资源隔离

- 操作系统内核升级，防火墙配置，docker基础镜像制作
- 系统硬件指标监控：负载、磁盘、内存、CPU、I/O、网络等
- 所有服务通过宿主机进行访问，容器内部对外不可见，安全性
- 一键式部署docker大数据平台，大数据集群的一键启动停止
- 容器服务的自动拉起
- 业务数据存储挂载到容器之外
- 可视化运维管控，无需进入后台

elastic

运维心得

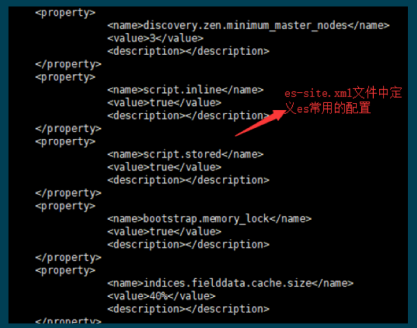
Ambari对ES统一运维管理实现方式

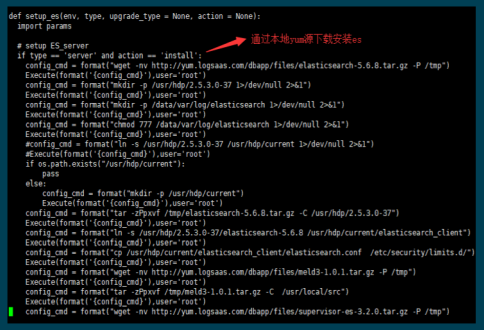
1.定义es的元信息配置文件metainfo.xml

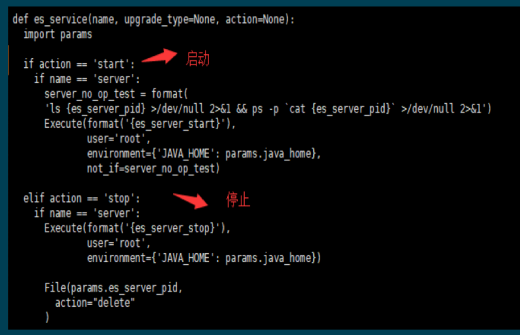
2.提取es常用配置写入到es-site.xml，通过python脚本读取ambari界面上该文件的配置并写入到es服务的目标配置文件中

3.编写py脚本控制自定义服务es的安装启动停止配置等功能









elastic

谢谢大家

elastic



elastic
中文社区

专业、垂直、纯粹的 Elastic 开源技术交流社区
https://elasticsearch.cn/

elastic