ELK平台SaaS化的问题和解决方案

数据驱动安全

王晓亮，安全易

2016 年 6 月

HanSight 瀚思

瀚思专注于大数据安全分析

- 已拥有招商银行、公安部、天津公安、东风日产、一汽大众、北京联通、河北金融学院、北京燃气集团等十余家客户，涵盖公安、金融、电信、制造、政府等领域。

- 积极拓展大数据安全行业与标准联盟，包括加入《CNCert中国互联网网络安全威胁治理联盟》，与公安部三所共同制定大数据安全体系标准，参与制定数据治理国家标准，与华为达成大数据生态战略合作等。

- 瀚思于2015年获得美国硅谷Red Herring全球创新公司百强、亚洲创新公司百强，为上榜唯一中国安全企业。2015中国信息产业年度人物获得者。

RED HERRING GLOBAL WINNER 100

CNCERT/CC 国家互联网应急中心

中华人民共和国公安部 The Ministry of Public Security of the People's Republic of China

HanSight 瀚思

我们的团队

高瀚昭 CEO

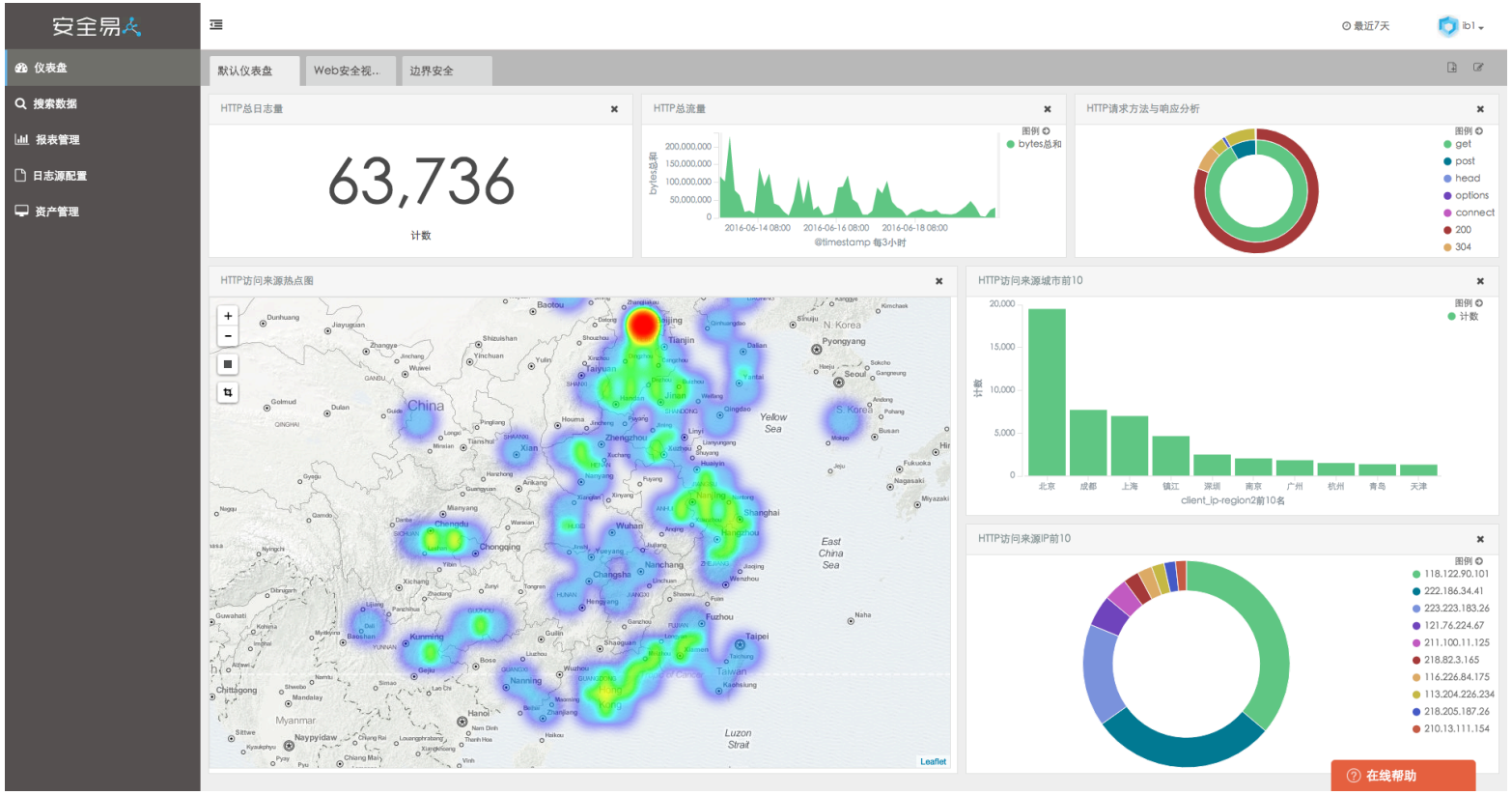
- 15年安全与大数据经验，在全球多个国家任研发负责人
- 前Tcloud天云趋势（宽带资本与趋势科技合资公司）CEO
- 在趋势科技核心研发负责人超过10年，领导团队负责核心安全引擎开发、病毒防御等工作，工作地点包括中国、加拿大、日本、东南亚
- 毕业于南京大学，获分布式计算硕士和学士学位

万晓川 首席科学家

- 核心安全研发、机器学习、算法世界级专家，5项美国专利持有者
- 在趋势科技13年，负责核心技术团队，任资深架构总监，曾带领团队开发基于沙盒的APT防御产品，与FireEye直接竞争，并在2014 NSS Lab的评测中胜出
- 趋势科技脚本分析引擎（反欺诈）核心开发人员、核心算法贡献者
- 毕业于东南大学少年班

关于我们 – 安全易

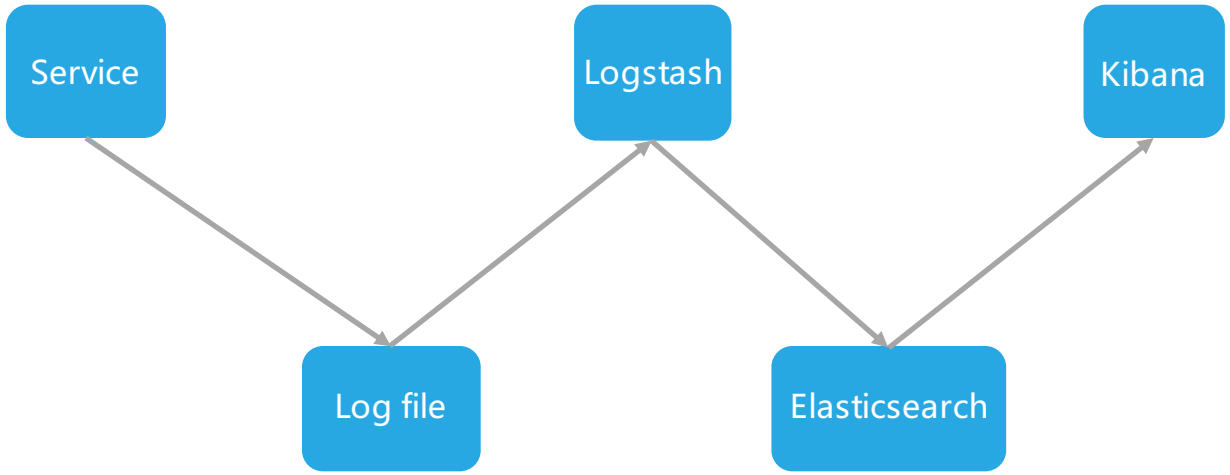
- 瀚思推出的一款基于云端的大数据安全分析平台
- 数据采集
  - 文本日志文件上传
  - 流式日志上传
- 数据分析
  - 搜索
  - 仪表盘
  - 报表
  - 安全事件告警



HanSight 瀚思

ELK

- Elasticsearch
  - 数据存储
- Logstash
  - 数据收集
- Kibana
  - 可视化



ELK SaaS化的问题

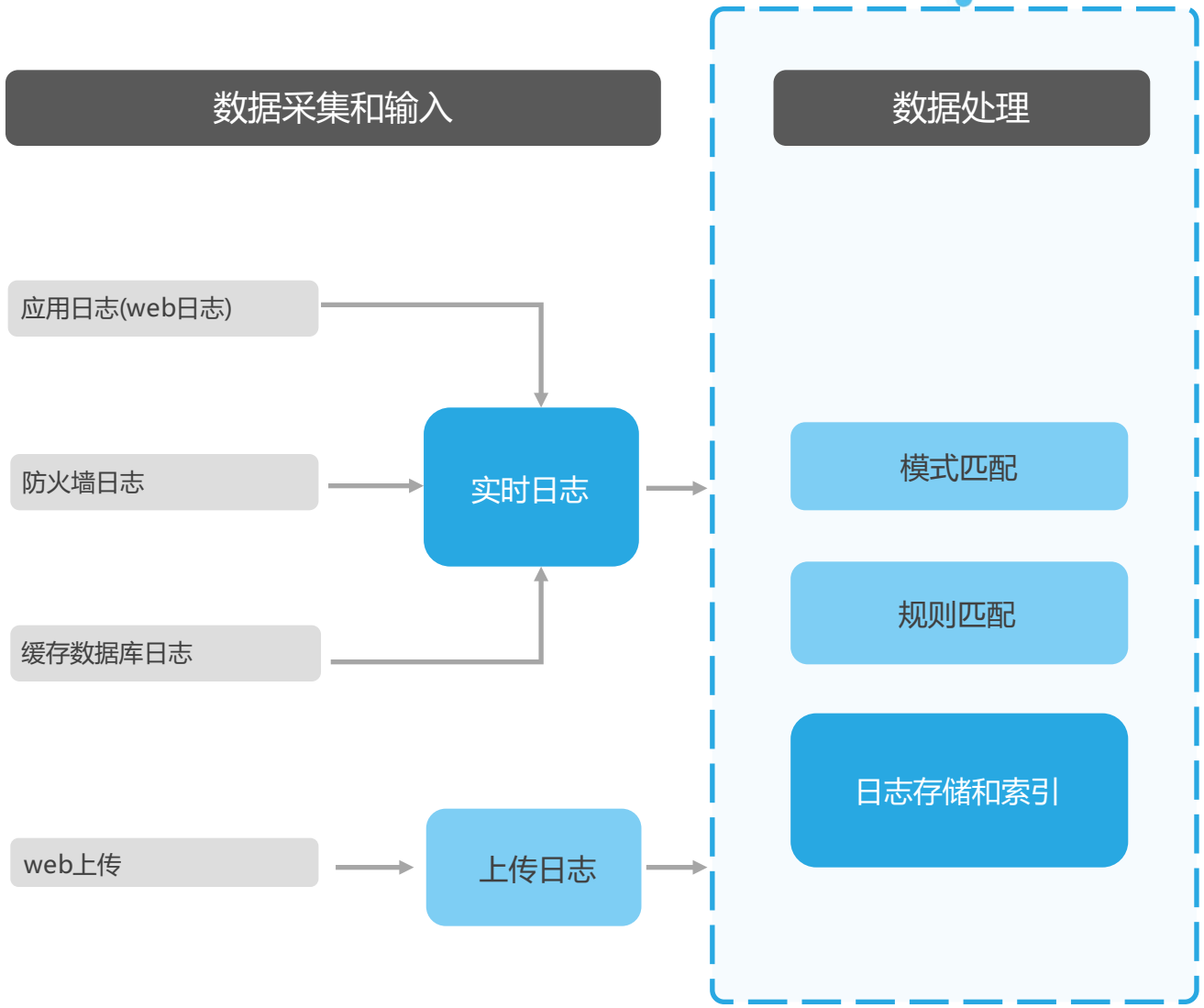
- 日志收集
- 多租
- 认证
- 报表

日志收集 – 问题

- Logstash
  - 依赖Java
  - 资源消耗高
  - 配置复杂
  - 区分日志的拥有者
  - 直连ElasticSearch

日志收集 – 解决方案

- rsyslog + uploader 模块
  - 提供rsyslog安装脚本下载（ agent ）
  - 部署方便
  - 轻量、高效
  - Uploader对日志进行解析并存储



多租 – 问题

- 收集的日志来自什么客户？
- Kibana
  - 认证
  - 共享的 .kibana 索引（配置数据多租）
  - 如何保护Elasticsearch中的用户数据（用户数据多租）

多租 – 数据采集 – 解决方案

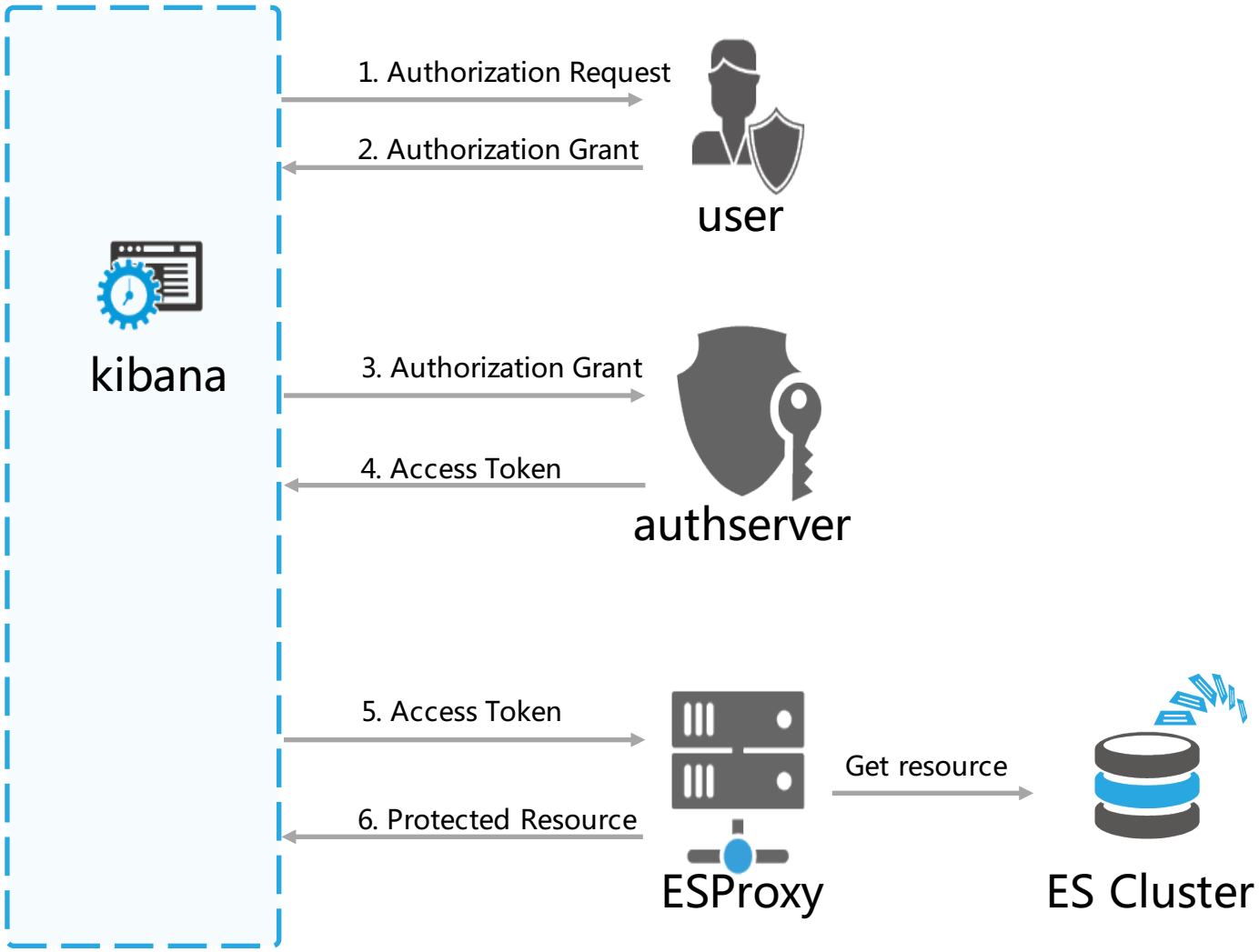
- 收集的日志来自什么客户？
  - 下载agent时将客户信息打包至agent中
  - 安装agent时将客户信息写入rsyslog配置中
  - Rsyslog上传日志时将客户信息带上
  - Uploader解析日志时将客户信息存入结构化的数据中

Kibana – 认证

- Shield
  - https://www.elastic.co/guide/en/shield/current/kibana.html
  - 原生应用
  - 商业付费
- kibana-authentication-proxy
  - https://github.com/fangli/kibana-authentication-proxy
  - OAuth + Basic Auth
  - 基于kibana nodejs

Kibana – 认证

- OAuth2 + JWT（JSON Web Token）
- 优点
  - 由独立认证服务提供用户认证及授权，减少对Kibana的改动
  - JWT 自包含，方便用于微服务架构

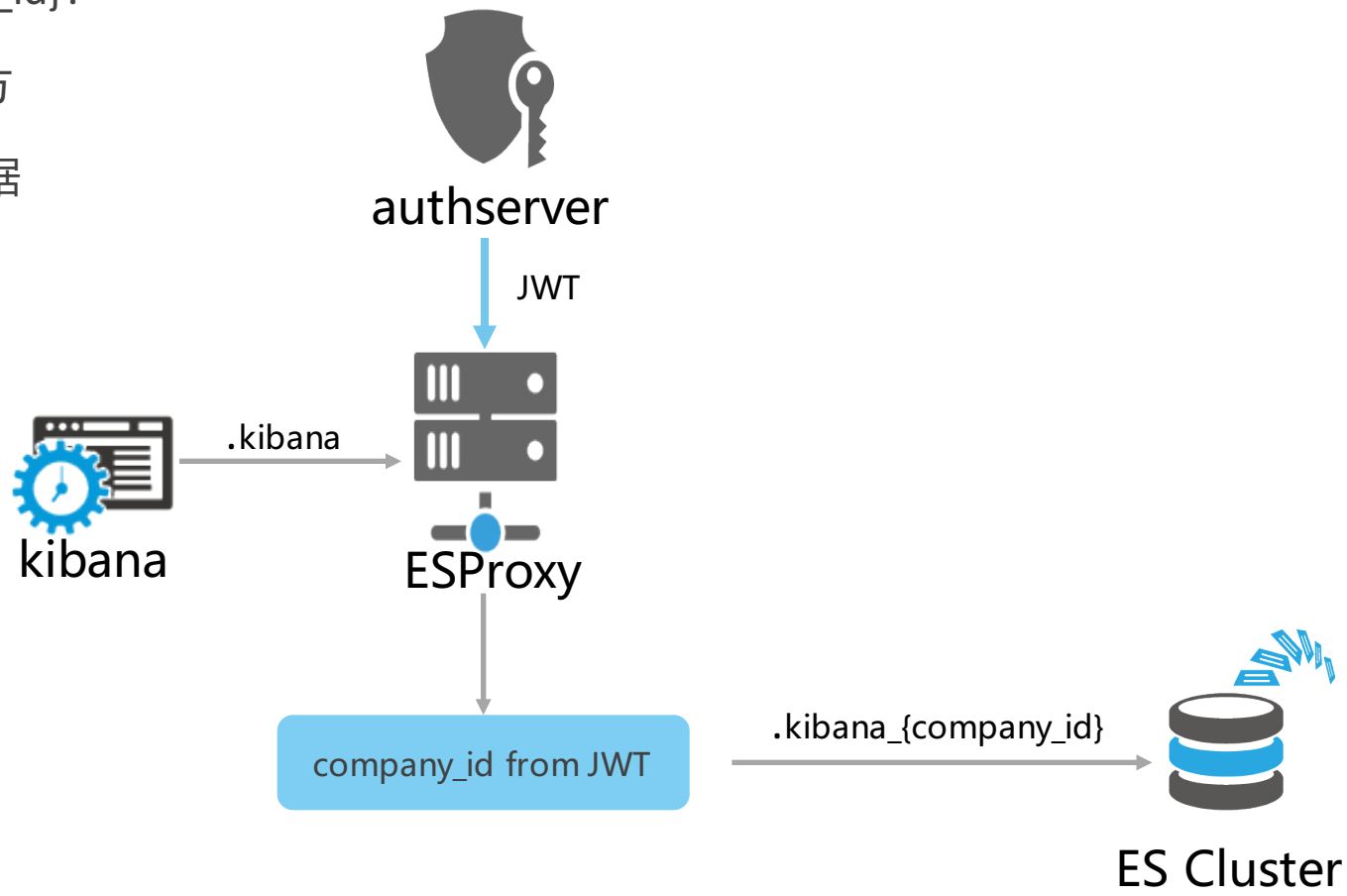


Kibana – 配置数据 – 多租

- Kibana的数据存储在ES .kibana index中
- 请求格式
  - /.kibana/dashboard/default
  - /{kibana_index}/{type}/{title}
- Type:
  - index_pattern, dashboard, visualization
- Title: 用户定义
- 多租?

Kibana – 配置数据 – 多租

- 多租：.kibana -> .kibana_{company_id}?
  - Kibana的请求代码分布在很多地方
  - Kibana拒绝修改其他index中的数据
- 解决方案 - 引入EsProxy
  - 无需更改kibana代码
  - 控制请求数据



Kibana – 配置数据 – 多租

- 每个客户需要自己创建visualization & dashboard？
- Kibana第一次加载
  - HEAD -> /.kibana
  - POST -> /.kibana
- 提供模板
- Kibana第一次加载
  - HEAD -> /.kibana
  - ESProxy拦截POST -> /.kibana
  - 获取/.kibana_template数据
  - 将返回结果POST -> /.kibana_{company_id}

Kibana – 用户数据 – 多租

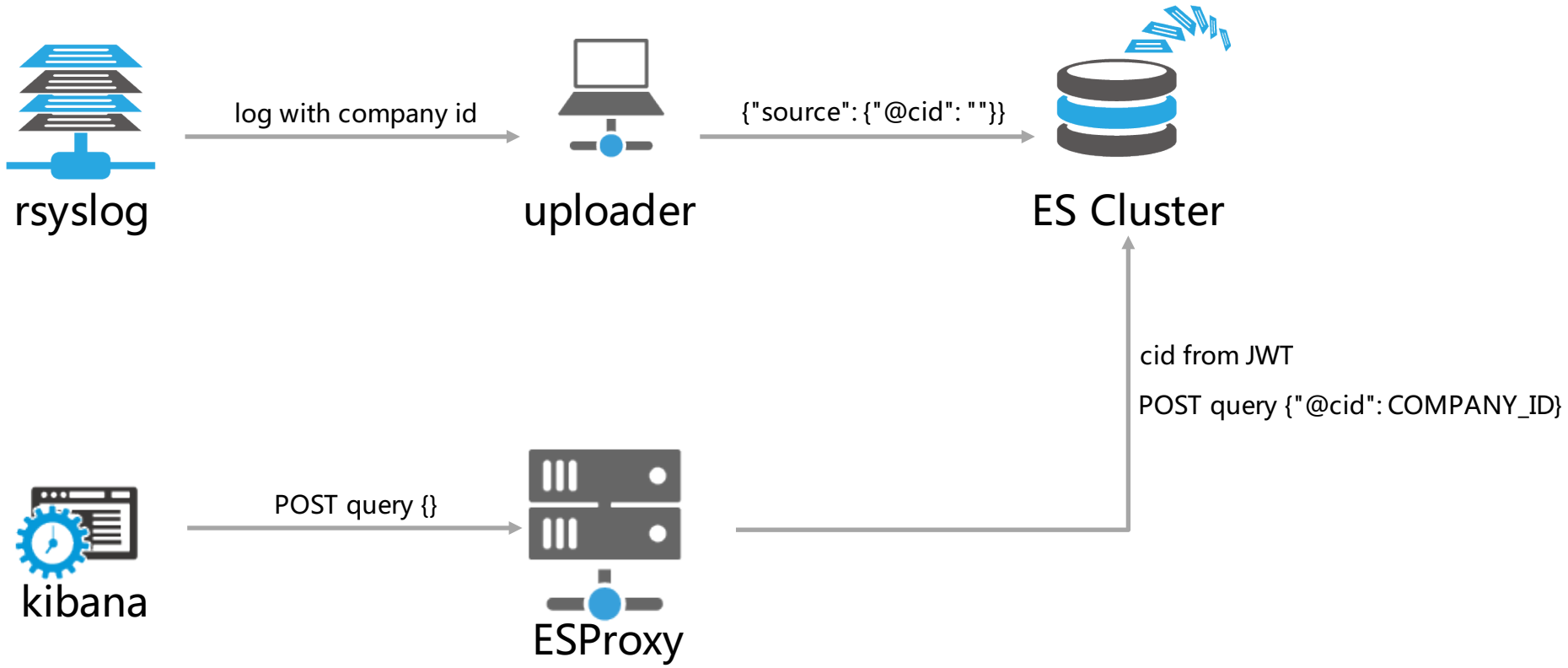
问题：

- Kibana 的Dashboard，图表，查询功能最终都转化为对ES的数据index的搜索
- Kibana 本身并没有数据多租户的概念，无法区分出数据index里的各家公司的数据区别。

解决办法：

- 利用认证得到的JWT token获取用户租户信息。
- 利用EsProxy过滤Kibana的数据查询请求，在所有请求中加入用户租户信息的查询过滤条件。

Kibana – 用户数据 – 多租



ELK – 报表

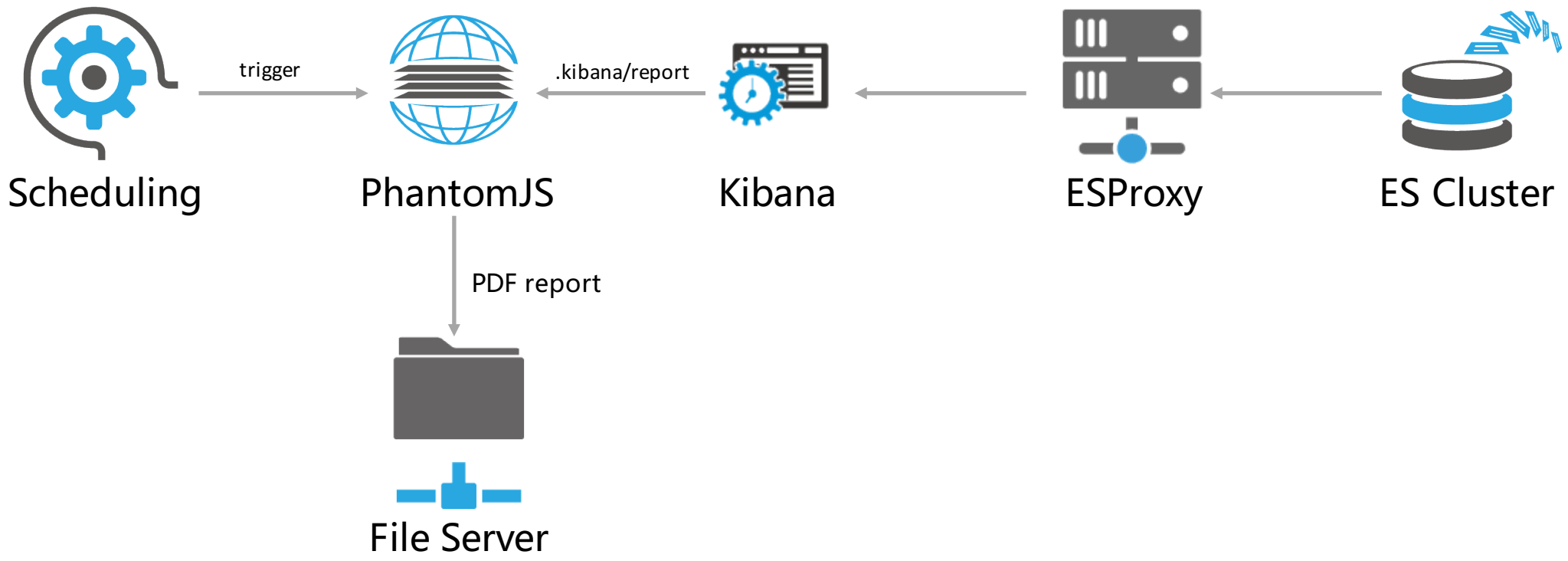
问题：

- 定时任务
- 所有的数据由Kibana查询和渲染
- Kibana查询和渲染逻辑复杂，难以更改
- 短期内难以实现一套后端查询数据并渲染的框架

解决办法：

- 自己实现一套定时任务框架
- 在后台使用Headless浏览器（PhantomJS）直接请求Kibana网页
- 在.kibana索引中添加新的report类型，数据结构和dashboard类型保持一致，复用dashboard的渲染逻辑
- 用PhantomJS提供的接口将页面转换成PDF

ELK – 报表



谢谢 | HanSight 瀚思

www.HanSight.com

微信公众号：瀚思安信

北京市海淀区中关村软件园9号楼2区306A