



分布式关联分析引擎 Sabre在NGSOC中的 应用

韩鹏
奇安信集团高级研发总监

目录

01： 事件关联和CEP

02： 大数据场景下的CEP关联分析

03： 分布式流式关联分析引擎-Sabre

Sabre

是什么？

新一代分布式流式关联分析引擎

CEP(复杂事件处理)技术在大数据领域的一个实现

中文名- 军刀，代表开箱即用，威慑力强

01 | 事件关联和CEP

什么是事件?

事件是计算机系统中某一活动产生的一组数据。

事件的体现形式是一个对象，它由特定属性和数据组成。

```
{  
    "src_user": "fangwen",  
    "dport": 80,  
    "log_type": "fw",  
    "msgid": "d00de753f73a4583a5197d34d8a30b25",  
    "collect_ip": "10.95.36.14",  
    "dip": "110.12.12.15",  
    "protocol": "TCP",  
    "event_name": "Match url profile",  
    "dev_type": "/安全设备/防火墙",  
    "occur_time": 1564453383000,  
    "sip": "110.12.12.13",  
    "severity": "6_信息",  
    "serial_num": "1896129436",  
    "systype": "log",  
    "dev_ip": "10.91.130.216",  
    "sport": 7704  
}
```

什么是事件关联？

■ 事件关联是一类用于对数以百计的设备中产生的数以百万计的日志进行分析以发现难以捉摸的攻击模式的技术



什么是事件关联？

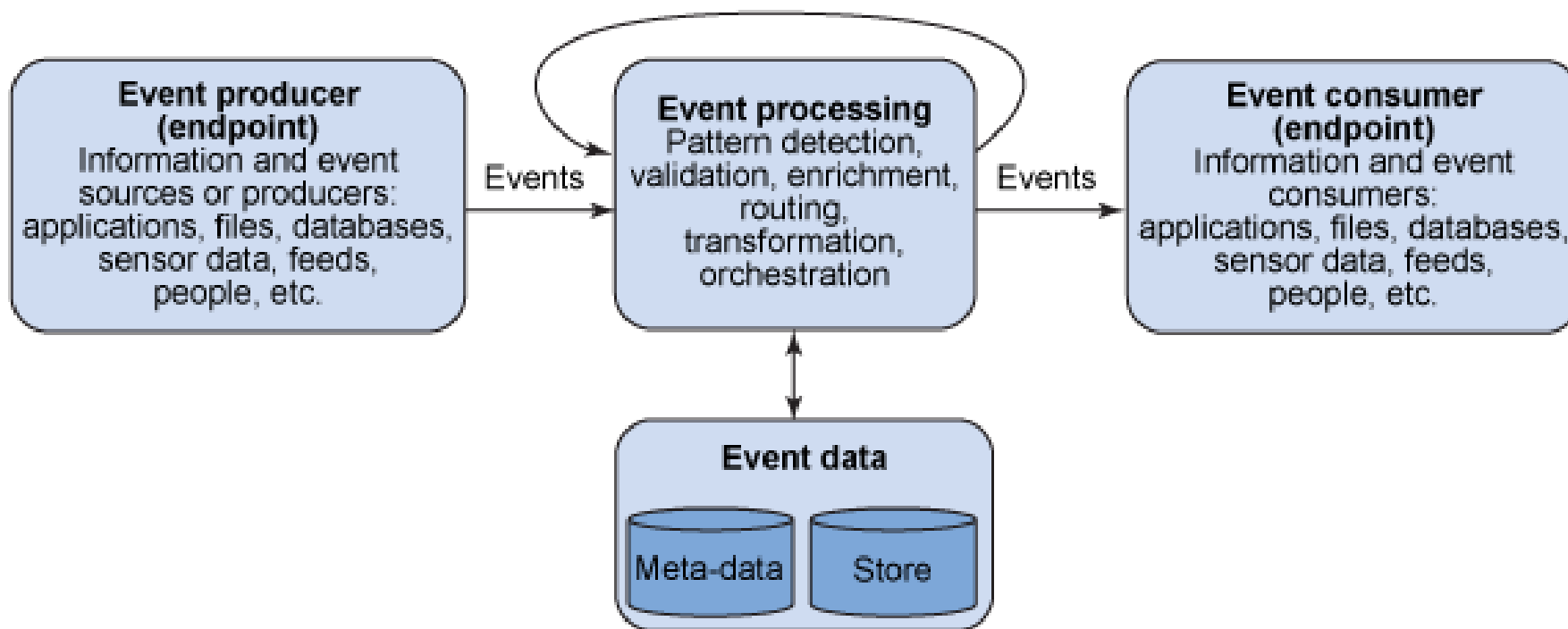
- 网络攻击是复杂的，多阶段，持续时间段，跨多节点的动态过程
- 独立的日志源无法看到攻击的全貌，而只能看到完整攻击的一个片段
- 不进行关联，就无法把大量的片段组合起来完成 **全景拼图**



什么是CEP?

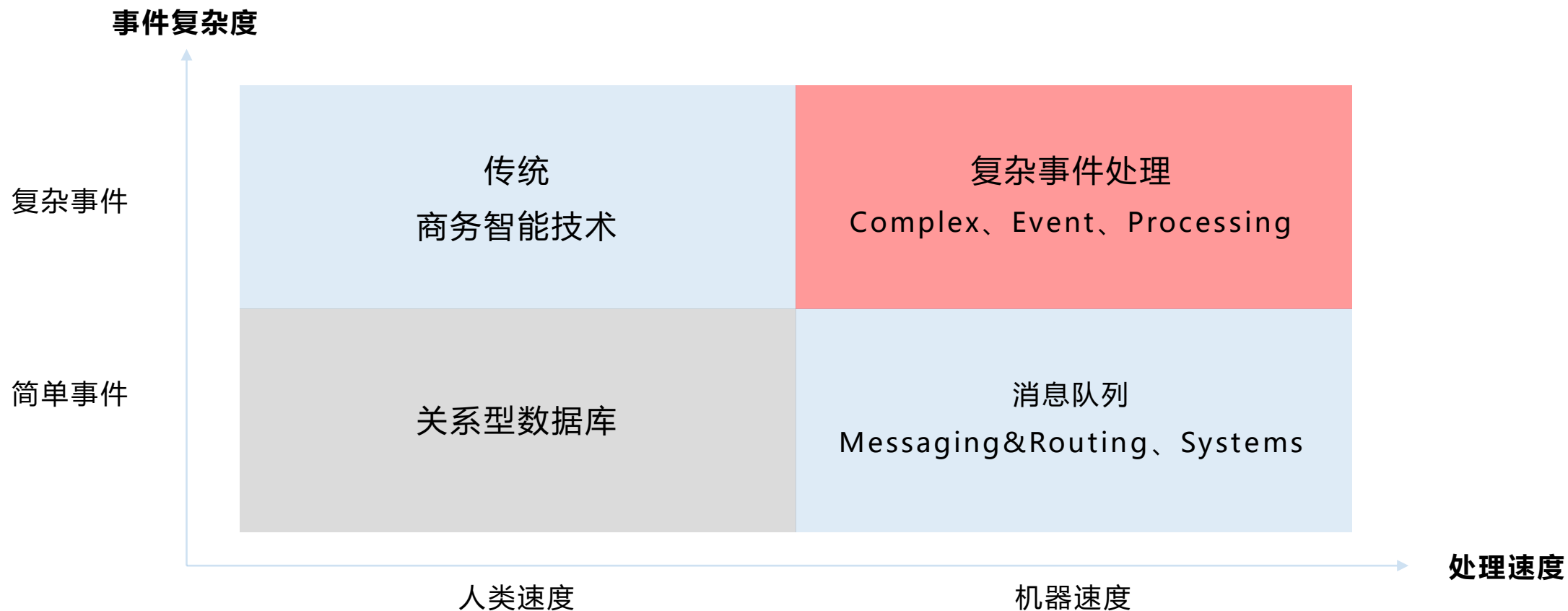
CEP: Complex Event Processing (复杂事件处理)

一种基于动态环境中事件流的分析技术



什么是CEP?

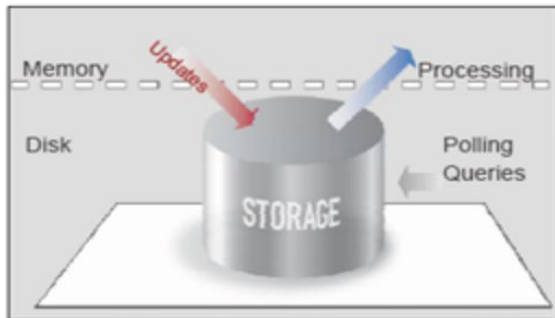
■ **CEP**: COMPLEX EVENT PROCESSING (复杂事件处理), CEP是 SIEM(SOC)的核心技术之一。



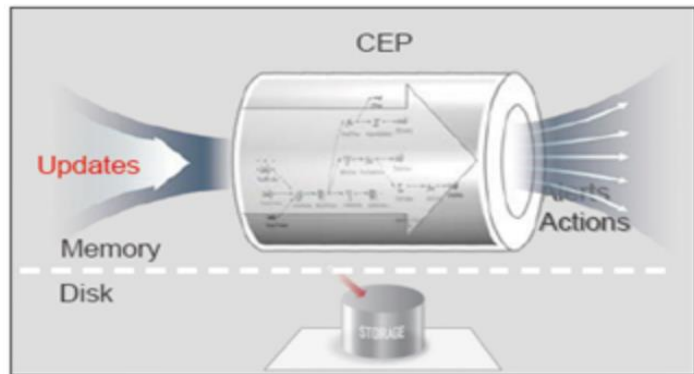
数据库技术和CEP的区别

水库 VS 水管

CEP: SQL on stream



- 先存储数据，然后查询、处理
- 为业务数据处理而优化



- 随着数据的流动获取、分析数据
- 全新的方法论
 - 把数据送到查询中
- 只加载极少量数据

优点: 超短延时

- 没有等待
- 实时提交结果

02 | 大数据场景下的CEP关联分析

海量数据如何有机结合？

■ 拥有各种类型的日志或数据，彼此之间孤立，不能发现深层复杂安全事件。

大量告警如何高效应对？

传统安全设备产生的大量告警事件，数量级已经达到靠人工无法有效处置。

典型场景如何精准防御？

■ 缺乏对典型场景的关键影响因素细粒度地分析。

高级威胁如何及时发现？

■ 面对的高级威胁事件APT攻击越来越多，没有能力及时地发现此类威胁。

大数据场景下的工程难点

| 难点 1

计算与存储资源的平台化趋势

| 难点 2

与已建大数据平台的兼容性

| 难点 3

依赖重运维的自有系统 VS 轻运维的产品

大数据场景下CEP关联分析实现的过程

步骤 1

(流式)计算框架的选择



步骤 2

复杂逻辑的拆分，使能分布式



步骤 3

计算任务的生命周期管理 (创建, 运行, 监控)

03

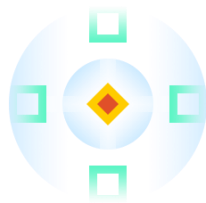
分布式流式关联分析引擎-SABRE

SABRE的特性



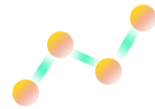
技术

独有的事件处理语言(EPL)
图计算
代码生成



特性

聚合计算、序列分析、关联计算、
时间窗口、分组去重、表计算、



产品

国内首款大数据分布式实时关联分析引擎（产品级）

可横向扩展的分布式引擎

来源更丰富

支持多源、异构日志
支持漏洞、资产、威胁
情报等多维数据
支持自定义对象内容

建模更简便

类Visio可拖拽轻松配置
150+预置规则
100+语义表达

集群更可靠

支持分布式部署
支持横向扩展

性能更强劲

可达10WEPS的实时处理
能力

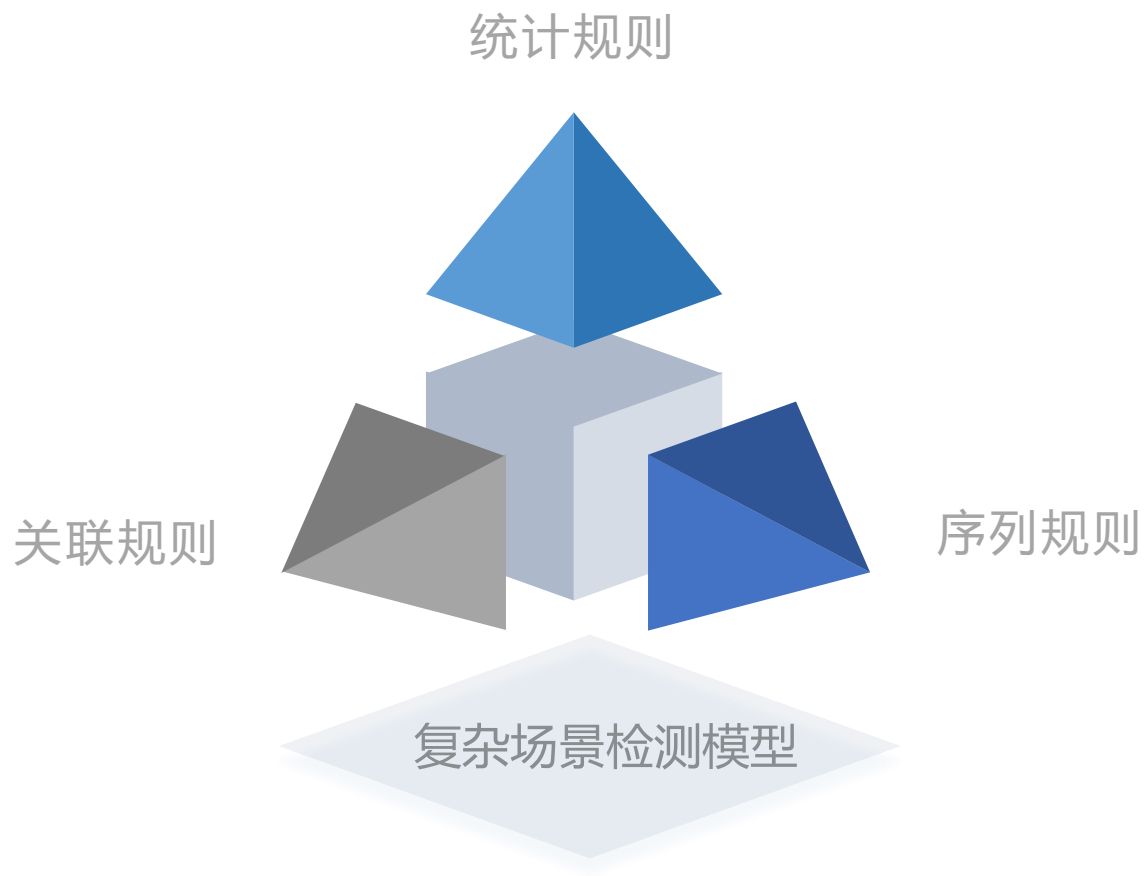


国内第一款具有自主知识产权及专利的大数据分布式关联分析引擎

灵活的规则建模能力

传统安全设备的规则是基于代码级别的编码，通过特征识别发现威胁的。但威胁是快速变化的，通过规则升级来响应是滞后的。

通过类VISIO的图形化连线拖拽配置，就可实时地对威胁场景进行建模，配置规则



快速配置和上线

通过Sabre关联分析引擎，将一个新的监控需求的实现从开发、测试和上线的复杂流程中解放出来。通过工具化的配置拖拽即可轻松定制任何检测场景！



遇到问题



分析原理



确定方法



编程开发



测试上线



具备
监测能力

快速配置和上线

通过Sabre关联分析引擎，将一个新的监控需求的实现从开发、测试和上线的复杂流程中解放出来。通过工具化的配置拖拽即可轻松定制任何检测场景！



规则建模工具

威胁 > 关联分析 > 统计规则建模

+

日志过滤

日志过滤1:

数据来源: probelog

+

日志统计

日志统计1

●

阈值比较

阈值比较1

计算单元之间的连线可通过双击删除

规则属性配置

计算单元配置

日志过滤名: 日志过滤1

数据来源: 日志类型

请选择日志类型名称

过滤条件

AND

添加条件

添加条件组

规则响应配置

保存

保存并退出

取消

威胁建模-统计规则

Use Case-工作时间某IP地址突现异常流量

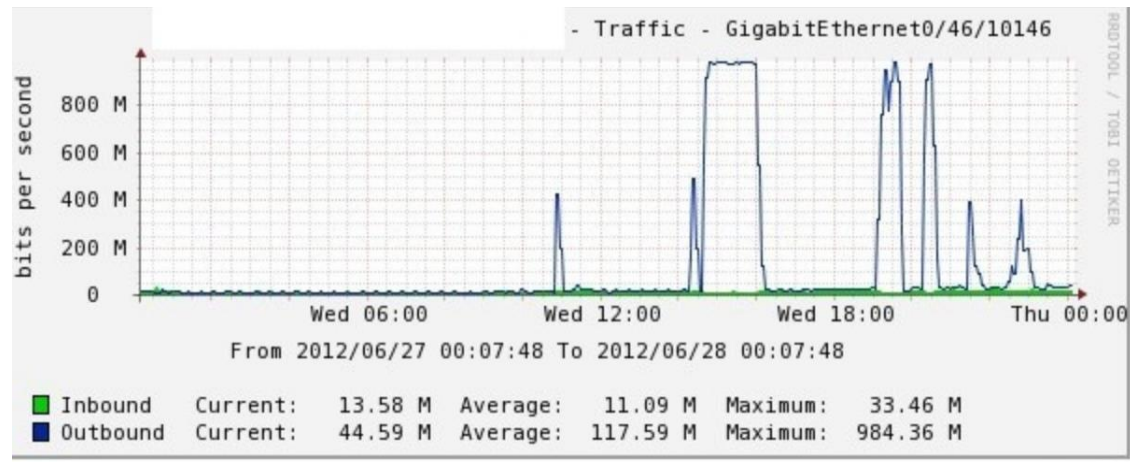
场景描述：工作日时间范围内,当前1小时内的TCP平均流量超过一周时间内TCP平均流量的40%。

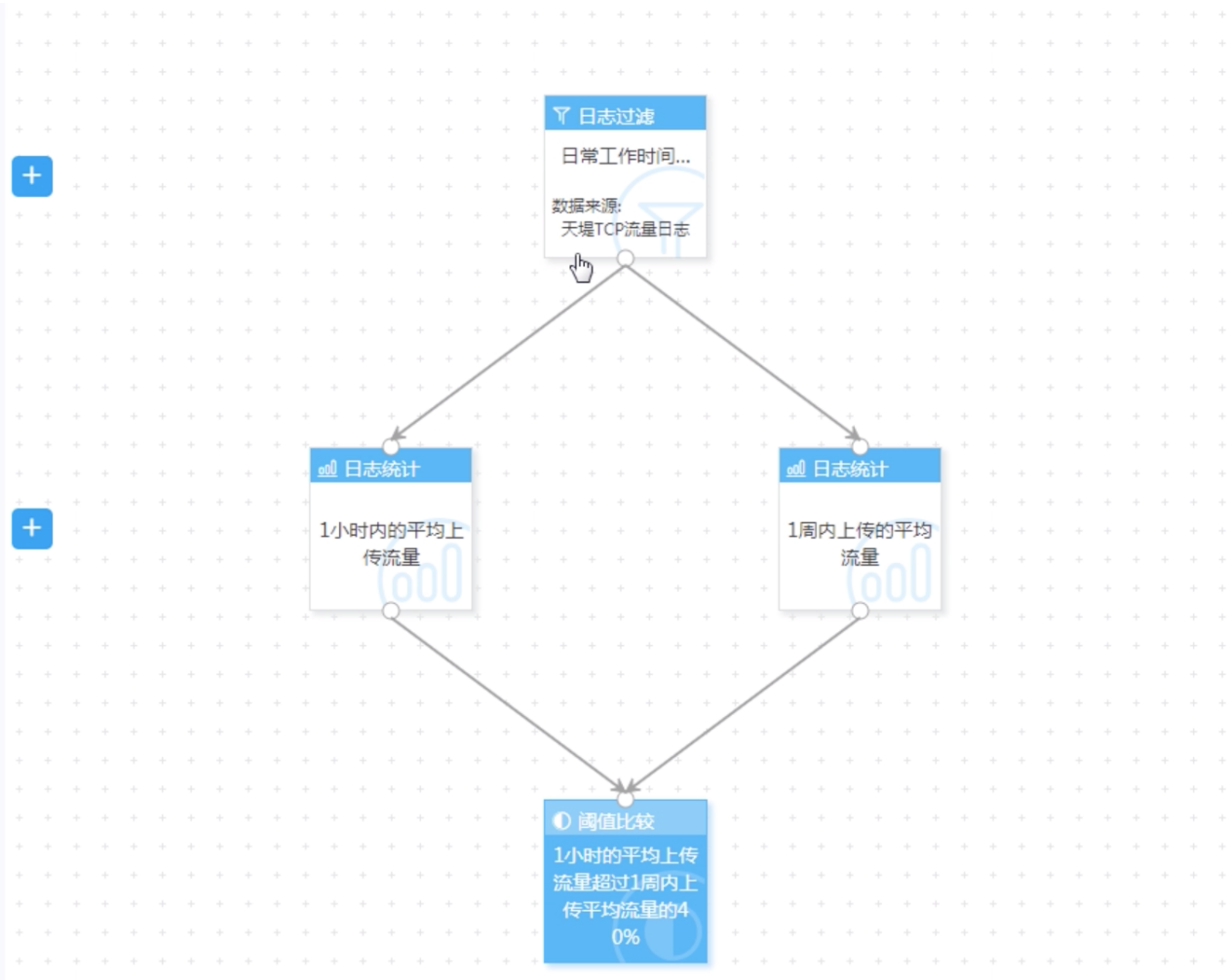
选择统计规则模板,先配置规则属性,再进行计算单元配置：

- ① 日志过滤1：数据源选择流量日志(TCP流量日志),过滤条件设置,发生时间 属于 工作时间(对象资源)
- ② 将日志过滤与日志统计1计算单元进行连线
- ③ 日志统计1：计算1小时内，相同源IP分组条件下，TCP流量日志.下行字节数的平均值；
- ④ 将日志过滤与日志统计2计算单元进行连线
- ⑤ 日志统计2：计算一周内，相同源IP分组条件下，TCP流量日志.下行字节数的平均值；
- ⑥ 将日志统计1、日志统计2与阈值比较计算单元进行连线
- ⑦ 阈值比较：日志统计1的值 变化幅度超过 40% 的 日志统计2的值

配置规则响应：此时只一个。

该场景描述了对一些异常情况的关注。





规则属性配置

计算单元配置

阈值比较名*1小时内的平均上传流量超过1周内上传平均流量的40%

阈值条件*

AND

添加单值条件

添加双值条件

添加多值条件

1小时内的平均上传流量

变化幅度超过

40%

1周内上传的平均流量

规则响应配置

威胁建模-日志关联规则

Use Case-网站被网络攻击利用成功

场景描述：WAF出现攻击类报警的事件，且同时发现在IPS的报警日志中，被访问服务器（目的地址）中存在可利用漏洞时，被认为是一个高危可信告警。

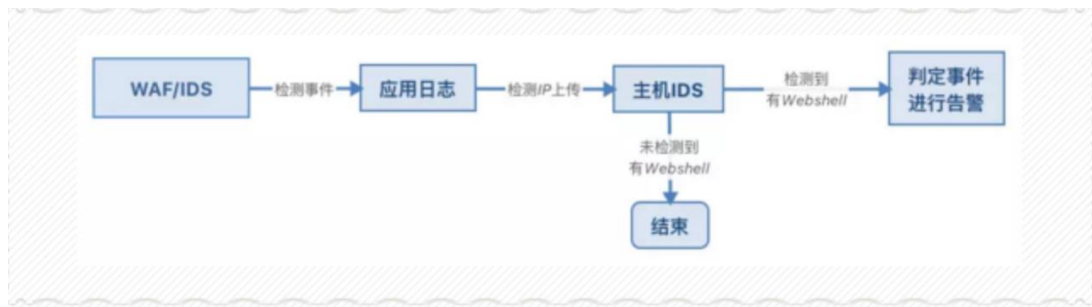
选择日志关联规则模板，先配置规则属性，再进行计算单元配置：

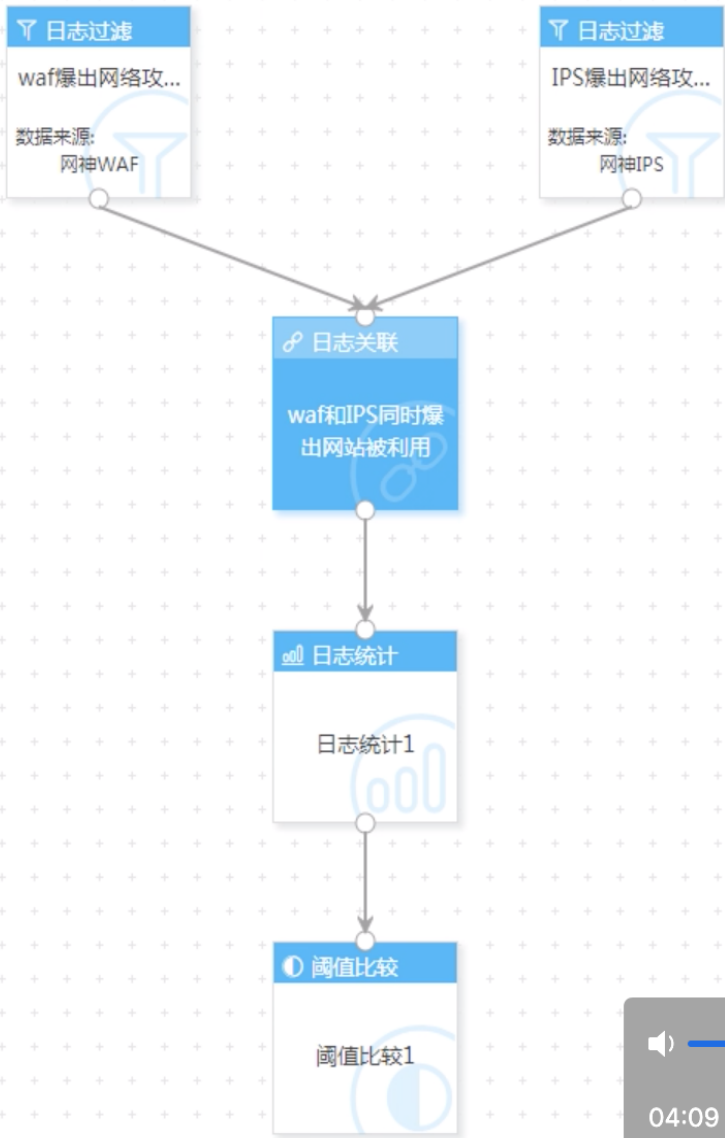
- ① 日志过滤1：数据源选择WAF报警事件，过滤条件设置WAF.目的IP==漏洞表.资产IP
- ② 日志过滤2：数据源选择IPS报警事件，过滤条件设置IPS.目的IP==漏洞表.资产IP
- ③ 日志连接：日志过滤1.目的P==日志过滤2.目的IP
- ④ 日志统计：5分钟时间范围内，聚类条件根据源IP、目的IP分组，默认统计方法为计数
- ⑤ 阈值比较：聚类统计的计数结果>=1

备注：日志关联规则模板的连线关系不可修改。

配置规则响应，如果需要先对规则的准确性进行验证，输出结果配置为关联事件。

以上典型场景是对多类型、多维度日志和数据进行的关联分析。





> 规则属性配置

> 计算单元配置

日志关联名* waf和IPS同时爆网站被利用

关联条件*

AND

+ 添加条件



waf爆出网络攻击利F

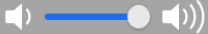
目的IP

==

IPS爆出网络攻击利F

目的IP

> 规则响应配置



04:09

-04:50

威胁建模-序列规则

Use Case- “永恒之蓝” 勒索病毒攻击

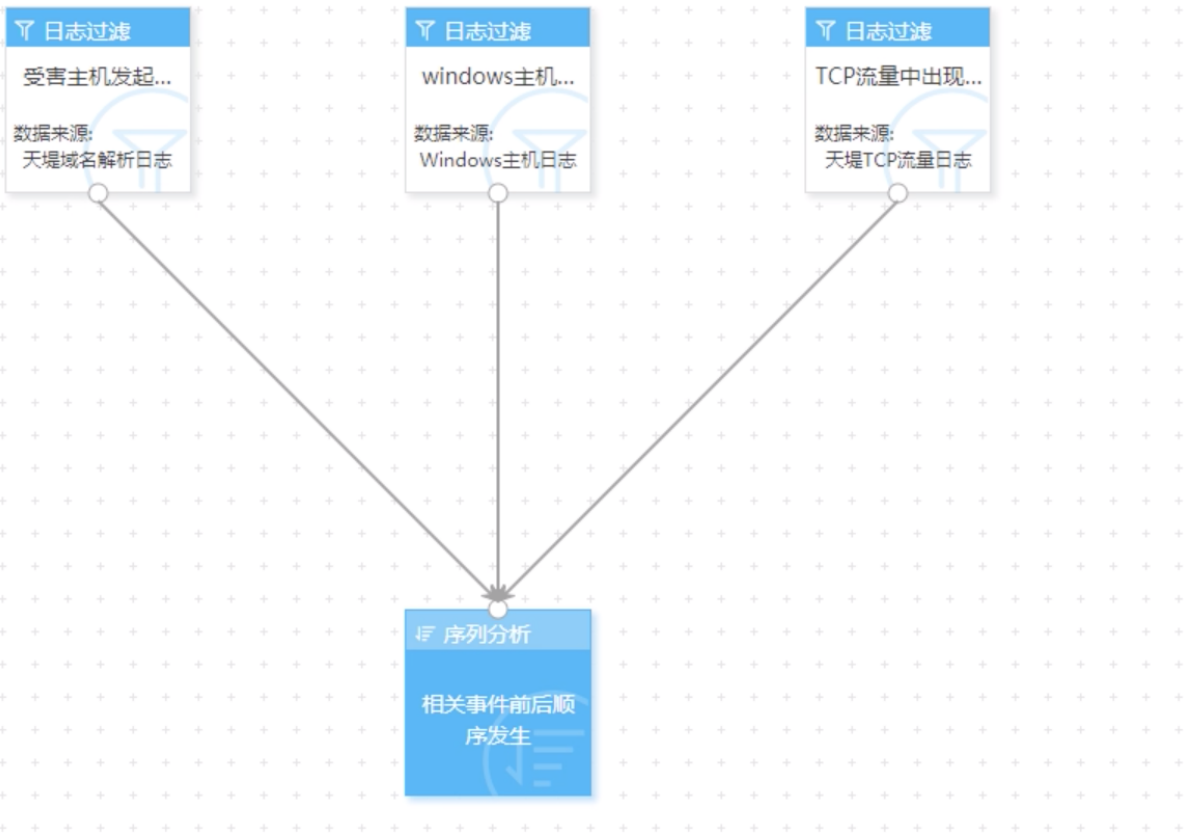
场景描述：内网主机被蠕虫利用漏洞MS17-010攻击，释放勒索病毒，并进一步感染其他主机。

选择连接规则模板，先配置规则属性，再进行计算单元配置：

- ① 日志过滤1：数据源为DNS解析日志，过滤条件中引用威胁情报，设置日志.解析域名==威胁情报.host
- ② 日志过滤2：数据源为Windows主机日志，过滤条件为“创建计划任务” 包含 “mssecsvc2”
- ③ 日志过滤3：数据源为TCP流量日志，过滤条件为目的port==445, 135、137、138、139
- ④ 将日志过滤1、日志过滤2、日志过滤3与序列分析计算单元进行连线
- ⑤ 序列分析：5分钟时间范围内，对源IP（被感染主机）进行分组，对事件日志过滤1、日志过滤2、日志过滤3发生顺序进行判断。

配置规则响应，可以设置告警的攻击阶段、置信度等信息。

以上典型场景是对全球范围内爆发的安全事件，利用威胁情报的关联，及时的内网发现和后续变种攻击的持续监测。



序列分析名* 相关事件前后顺序发生

时间范围*

10

分钟

事件发生顺序*

请选择

连续发生次数

添加

顺序	日志过滤名	发生次数	操作
1	受害主机发起对外部特定DNS访问	1	↑ ↓ 删除
2	TCP流量中出现内网相关端口扫描事件	1	↑ ↓ 删除
3	windows主机释放进程	1	↑ ↓ 删除

统计条件

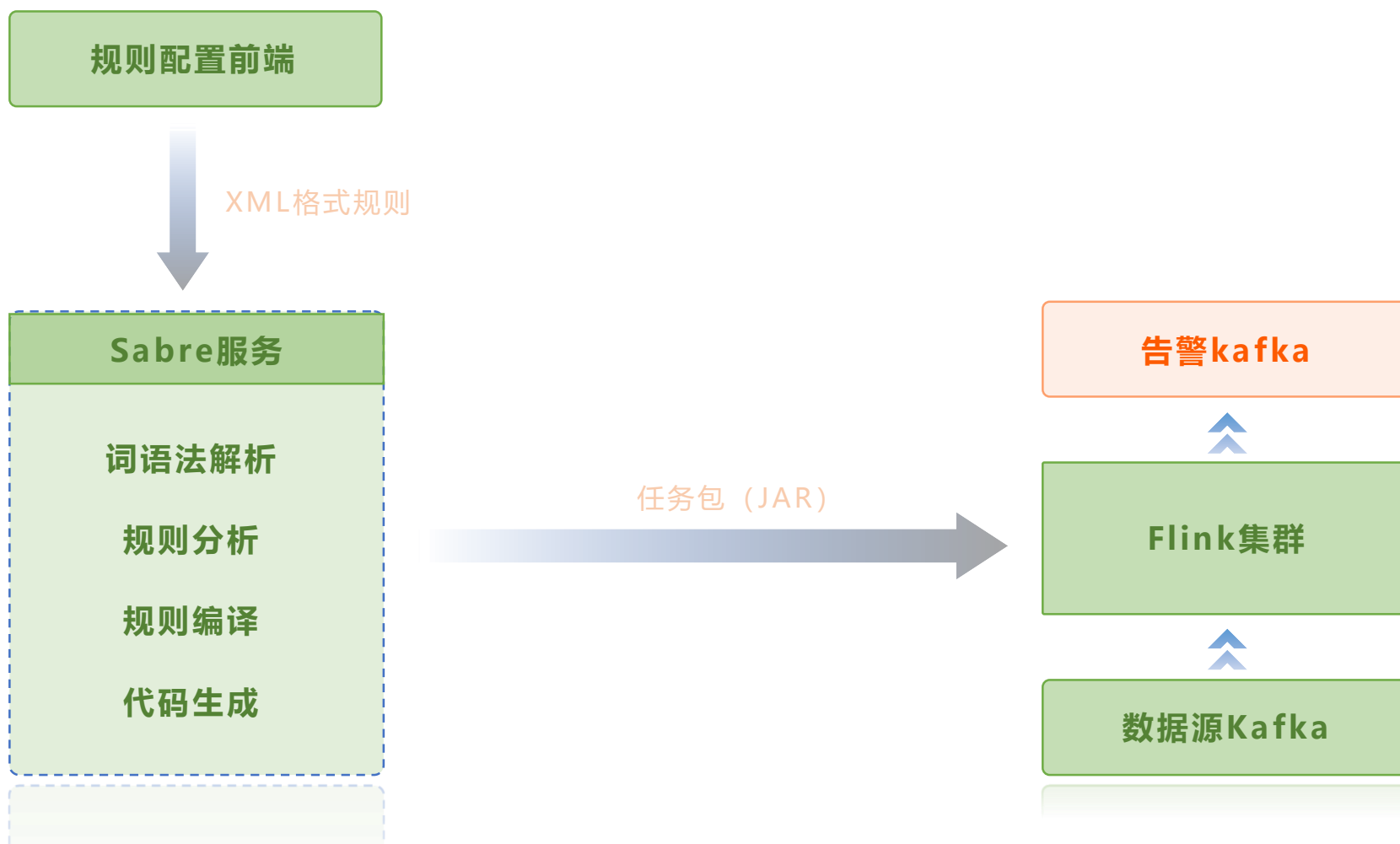
AND

添加分组条件

添加去重条件

> 规则响应配置

整体架构



XML格式规则

```
-<rules>
  -<rule>
    <id>100084</id>
    <name>All Exploits Become Offenses</name>
    <time/>
  -<objects>
    -<object>
      <type>dataSource</type>
      <id>5a4f1744-251f-4517-b13a-2e423803b0ef</id>
      <name>5a4f1744-251f-4517-b13a-2e423803b0ef</name>
    -<dataSource>
      -<method>
        read(target="kafka", bootstrap.servers="10.65.13.27:9092,10.65.13.29:9092,10.65.13.30:9092",
          zookeeper.connect="10.65.13.27:2181,10.65.13.29:2181,10.65.13.30:2181", topic="sabreEvents", group.id="sabre.task", data.type="messagePack",
          timestamp.type="event", timestamp.field="occur_time", timestamp.format="yyyy-MM-dd HH:mm:ss")
      </method>
    </dataSource>
  -<successors>
    -<group>
      <method>copy</method>
      <node>Fd591b624-a18f-4586-acd6-41f4793a7718</node>
    </group>
  </successors>
</object>
  -<object>
    <type>filter</type>
    <id>Fd591b624-a18f-4586-acd6-41f4793a7718</id>
    <name>Fd591b624-a18f-4586-acd6-41f4793a7718</name>
  -<filter>
    <method>select * where ((event.category in set("5000")))</method>
  </filter>
  -<successors>
    -<group>
      <method>copy</method>
      <node>2775ff9b-172f-49c9-a4da-76cf03f54e4b</node>
    </group>
  </successors>
</object>
  -<object>
    <type>window</type>
    <id>2775ff9b-172f-49c9-a4da-76cf03f54e4b</id>
    <name>2775ff9b-172f-49c9-a4da-76cf03f54e4b</name>
  -<window>
```

核心技术

图计算

- ① 支持多规则
- ② 多规则进行全局语义优化
- ③ 规则匹配优化成流式图计算

自动代码生成

- ① 编译器将规则匹配映射成等价代码
- ② 直接运行对应代码，大幅提高规则匹配效率

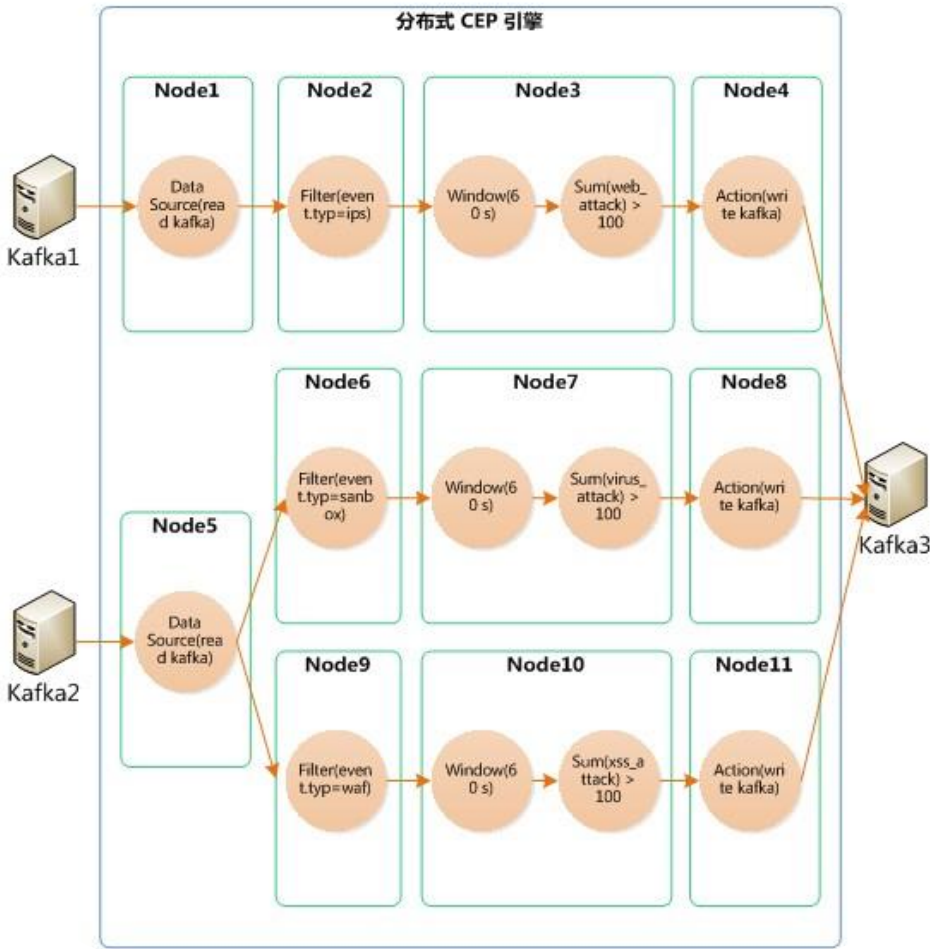
分布式状态监控

- ① 规则匹配拆分为图计算节点
- ② 全局监控规则运行状态，包含内存、CPU、事件匹配情况等

图计算

- 引擎内部引入可计算对象的抽象概念，将各个计算节点抽象成对象，对象通过数据流图组成DAG，这样可以简化数据结构的设计难度，提高引擎的可扩展性

计算单元	语义
dataSource	数据源，如kafka、database等
filter	对从数据源读取的事件进行过滤
window	聚集符合窗口时间的事件
aggregation	统计相关计算：count、max、min、average、sum
sequence	处理事件序列逻辑
join	两数据源事件关联
operator	对象计算模块
action	动作响应，处理规则触发之后的动作



分布式运行

- Sabre的分布式能力基于流式处理框架：Apache Flink
- Apache Flink 是一个框架和分布式处理引擎，用于在无边界和有边界数据流上进行有状态的计算。Flink 能在所有常见集群环境中运行，并能以内存速度和任意规模进行计算



产品	模型	API	保证次数	容错机制	状态管理	延时	吞吐量
storm	Native (数据进入立即处理)	组合式 (基础API)	At-least-once	Record ACKs (ack机制)	无	Low	Low
Trident	mirco-batching (划分小批处理)	组合式	Exectly-once	Record ACKs	基于操作 (每次操作有一个状态)	Medium	Medium
Spark streaming	mirco-batching	声明式 (提供封装后的高阶函数，例如count函数)	Exectly-once	RDD Checkpoint (基于RDD做Checkpoint)	基于DStream	Medium	High
Flink	Native	声明式	Exectly-once	Checkpoint (flink的一种快照)	基于操作	Low	High

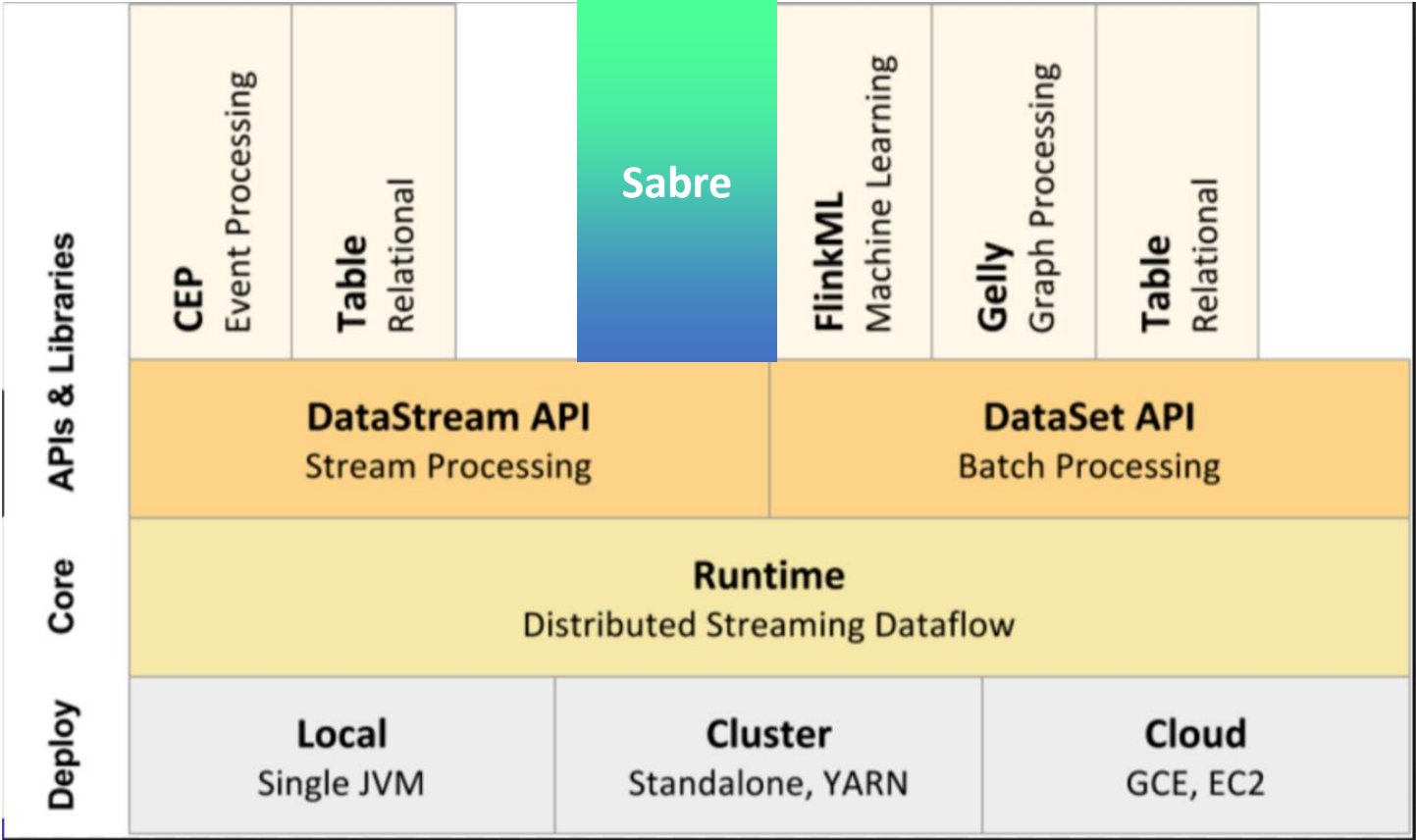
技术决策-最小依赖原则

对 Flink 的技术依赖

只使用Flink的分布式计算能力


业务代码尽可能减少对Library层的依赖

很容易适配到Flink的各个版本
(1.4/1.7.2)



分布式状态监控

<div><div>+ 新增</div><div>× 删除</div><div>▷ 启用</div><div> 停用</div><div>告警动作</div></div> <div>导入规则</div>												
<input type="checkbox"/>	规则序号	规则名	规则类型	模版类型	告警动作	启用状态	运行状态	生成结果	日志计数	告警计数	内存占用/MB	CPU利用率
<input type="checkbox"/>	1770	预置-天堤远控木	远控木马	统计规则		启用	运行中	告警	66499	31395	8	0.0011%
<input type="checkbox"/>	1754	预置-天堤未分类	未分类情报命	统计规则		启用	运行中	告警	66499	7717	8	0.0010%
<input type="checkbox"/>	1811	预置-流量-端口	端口扫描	统计规则		启用	运行中	告警	868739	7199	544	0.0115%
<input type="checkbox"/>	1756	预置-天堤窃密木	窃密木马	统计规则		启用	运行中	告警	66499	6347	8	0.0010%
<input type="checkbox"/>	1810	预置-流量-特定	端口扫描	统计规则		启用	运行中	告警	868739	1836	417	0.0115%
<input type="checkbox"/>	1790	预置-天堤僵尸网	僵尸网络	统计规则		启用	运行中	告警	66499	1817	8	0.0010%
<input type="checkbox"/>	1632	预置-天堤APT事	APT事件	统计规则		启用	运行中	告警	66499	1367	8	0.0010%
<input type="checkbox"/>	1841	引用预置规则-IC	跨站请求伪造	统计规则	EDR外发	启用	运行中	告警	1817	1180	8	0.0002%
<input type="checkbox"/>	1793	预置-天堤网络蠕	网络蠕虫	统计规则		启用	运行中	告警	66499	636	8	0.0010%
<input type="checkbox"/>	1788	预置-天堤勒索软	勒索软件	统计规则		启用	运行中	告警	66499	536	8	0.0010%

The background is a solid blue color with a subtle, abstract pattern of thin, light blue lines that create a sense of depth and movement, resembling a grid or a series of overlapping planes.

THANKS

2019 北京网络安全大会
2019 BEIJING CYBER SECURITY CONFERENCE