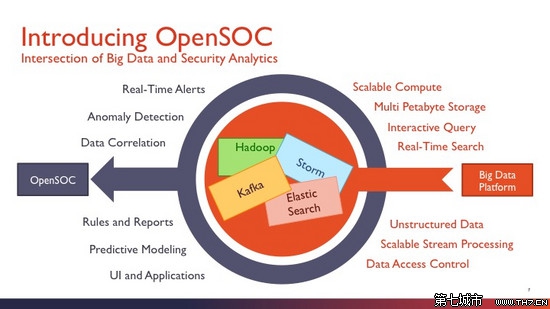
**OpenSOC**

什么是OpenSOC

OpenSOC是一个针对网络包和流的大数据分析框架，它是大数据分析与安全分析技术的结合, 能够实时的检测网络异常情况并且可以扩展很多节点，它的存储使用开源项目Hadoop，实时索引使用开源项目ElasticSearch，在线流分析使用着名的开源项目Storm。

OpenSOC是大数据安全分析的框架设计，对数据中心机排放数据进行消费和监控网络流量。opensoc是可扩展的，目的是在一个大规模的集群上工作。

OpenSOC能做什么？   
可扩展的接收器和分析器能够监视任何Telemetry数据源   
支持对Telemetry数据流的异常检测和基于规则实时告警   
支持使用ElasticSearch实现自动化实时索引Telemetry数据流   
支持原数据包的抓取、存储、重组   
支持数据驱动的安全模型   
是不是很强大？   
OpenSOC运行组件包括哪些？



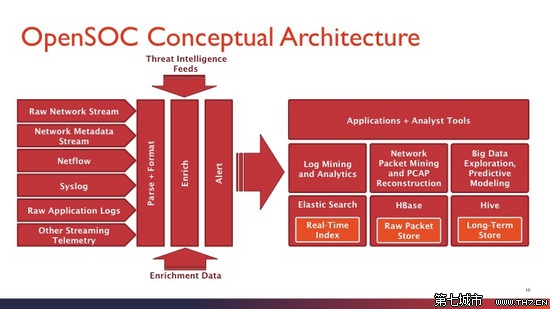
Apache Flume  
Apache Kafka  
Apache Storm  
ElasticSearch   
MySQL   
  
OpenSOC组成部分（核心）

OpenSOC-Streaming：这个库包含了拓扑结构的加工、丰富，索引，及相关的信息，PCAP重建服务，以及其他各种数据服务。可以在GitHub下载： https://github.com/opensoc/opensoc-streaming

nSOC-UI：操作界面，日志和网络数据包的分析，显示警告和错误。可以在GitHub下载： https://github.com/opensoc/opensoc-ui

OpenSOC深入剖析   
OpenSOC框架是大数据分析框架的衍生

先来看看OpenSOC概念性框架，总体来说，就是分为三个部分：左侧为数据输入（采集），中间为数据处理（计算），右侧为数据分析（展现和输出）。



再来看看其数据流框架，在概念性框架的基础上，把每一步的功能组件都一一列清楚了。分为六个部分：

Source Systems，数据输入源，分为主动和被动两种方式；   
Data Collection，数据收集，主要采用Flume进行数据收集和预处理，PCAP进行抓包收集；   
Messaging System，消息系统，主要是Kafka分布式消息系统进行数据缓存，根据数据源不一样来划分不同的topic；   
Real Time Processing，实时处理，主要采用Storm实时计算框架进行数据整理，聚合，DPI分析，等，这里，每个kafka topic都需要单独的storm 应用程序来独立处理；   
Storage，存储，就是把计算的结果和原始数据写入相应的存储模块，原始数据存入Hive，日志数据存入ElasticSearch便于索引查找（结合kibana），抓包数据存入HBase；   
Access，访问层，简单说就是把分析结果数据从存储中取出来，通过各种BI工具渲染到页面，当然，也可以把数据以web service的方式提供给第三方。

