绿盟大数据安全分析平台

产品白皮书

【绿盟科技】

■ 文档编号		■ 密级	完全公开
■ 版本编号	V1.0	■ 日期	2015-11-03
■ 撰 写 人	胡喆骞	■ 批准人	

目录

1. 需求分析

1.1. 安全现状

随着互联网以及周边网络产品的发展，信息安全越来越受到重视，“棱镜门”事件爆发后，信息安全不仅引起了企业领导的重视，更引起了国家领导人的广泛关注。在这种背景下，企业无论是出于对自身利益的考虑，还是对于社会责任的角度，都已经开始构建更为丰富的内部安全系统。

这些系统不仅涵盖基本的防火墙，入侵检测、防病毒；还包括目前主流的上网行为审计，堡垒机系统，数据库审计系统，网站防火墙系统；以及符合最新攻击的特征的，如抗拒绝服务系统，高级持续性威胁防御系统等。这些专业的安全防护设备逐渐达到了企业的防护屏障，从多个不同的角度满足了企业的安全防护需求。

随着科技的不断进度，针对企业内部的攻击行为也逐渐变得更为难以捕获。这种情况在现今表现的尤为突出，以 HackingTeam 被攻击导致安全工具泄露为例，其直接导致地下黑客产业向前推进 5-10 年。在这样严峻的安全风险面前任何现有的安全设备都难以应对，需要通过较为复杂的安全分析方能挖掘出隐秘在企业内部的安全事件。

1.2. 当前挑战

单纯的安全防护已经很难应对如此复杂的安全环境。目前在高级恶意程序已经逐渐成为主流的情况下，隐秘通道也已经开始向企业内部逐渐渗透。具不完全统计，目前各种安全事故层出不穷手段不断翻新，已知的手段包括有非授权访问，远程代码执行，权限绕过，SQL 注入，DNS 劫持，DDoS 攻击，账户劫持，恶意程序，DNS 缓存入侵，重定向，涂鸦，系统漏洞等。这些手段不断威胁着我们的信息系统，被攻击的企业已经从耳熟能详的超大型企业逐渐转向大企业、中型企业等。

企业中任何存有数据的系统都已经变为攻击者眼中的目标。企业中保存的员工数据、财务数据、客户数据、甚至考勤数据都已经变得炙手可热。而每起安全事故的发生、数据的泄露都是隐藏在网络数据的海洋中，企业中的安全管理人员难以发现。安全事件都是在发生

后，数据在网络上广为流传后企业才会发现曾经有过安全事件，但具体的时间、形式都难以察觉。尤其是以高级持续性恶意攻击（APT 攻击）为代表的新威胁，更是让企业防不胜防。现有的任何防御手段在 APT 攻击面前都显得苍白无力。

1.3. 应对措施

为了面对如此险恶而又复杂的网络安全环境，传统的防御手段难以发挥有效作用。随着新技术的不断涌现，以数据为驱动的安全防御手段逐渐浮现到安管人员面前。

一切行为皆有痕迹，在网络中行为的痕迹会以多种方式表现而出，如网络的流量、系统的日志、安全设备的告警等。无论是传统的安全攻击，如 DDoS、账户劫持、恶意程序等等，亦或是先进的 APT 攻击，所有的攻击行为都会在网络或者系统中留有痕迹。这样的痕迹都是分散在各个系统中，以一个个信息孤岛散落在企业的信息化系统中。

为了应对日益严峻的安全攻势，需要进行有效的安全分析。安全分析能够有效的对未知安全事件进行发现，由于安全分析是持续的，因此能够在安全事件发生时进行有效的捕获。安全分许首先需要将数据进行汇总，解决数据持续增长、数据类型复杂、数据来源多样等特性。基于这样的数据特性，原有的数据库、数据仓库技术难以应对，伴随着业务需求的产生，相应的解决方案也应运而生。

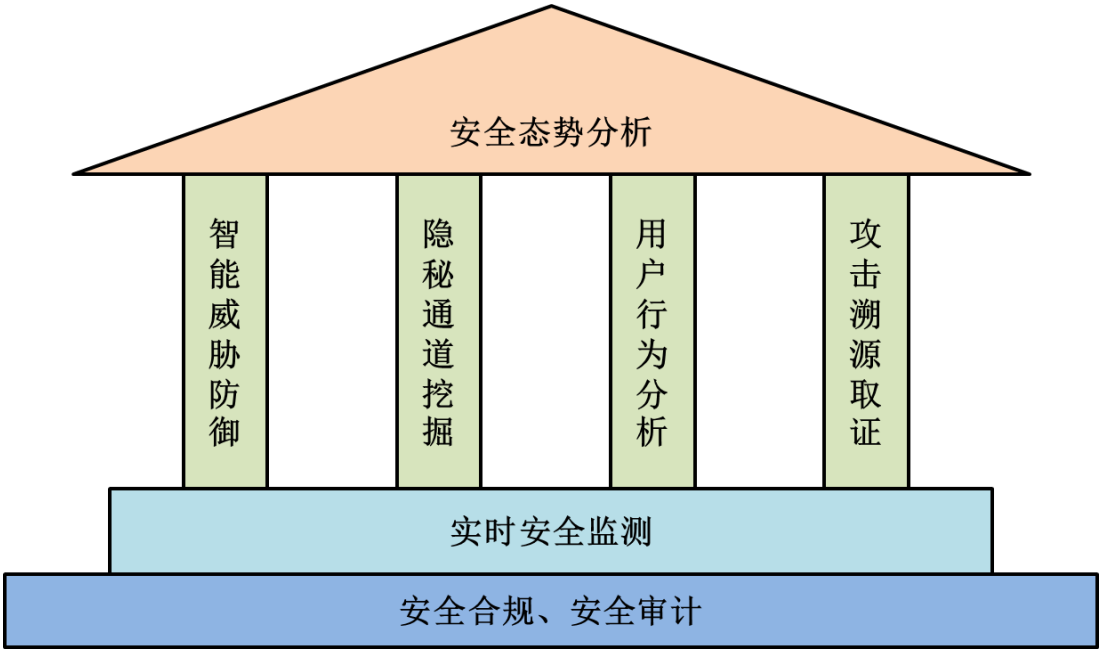
通过采用大数据技术可以很好的解决数据的上述特性，大数据技术在多个行业的业务系统中已经成功使用。例如电信运营商采用大数据技术进行恶意电话的挖掘，金融业采用大数据技术对信用卡欺诈的风险控制。这些成功案例都已经说明了大数据技术在数据分析、数据挖掘方面的有效性。

通过大数据安全分析技术的应用，能够很好的进行安全数据分析，从而更好的应对现有网络中所带来的安全挑战。

2. 产品定义

绿盟大数据安全分析平台是一款采用大数据技术的安全分析系统。消除安全孤岛，从整体视角进行安全事件分析、安全攻击溯源、安全事件根因挖掘等。

3. 功能特点



以安全合规、安全审计为基础，借由实时安全监测实现基础安全功能。通过智能威胁防御、隐秘通道挖掘、用户行为分析、攻击溯源取证进行功能支撑，满足安全态势分析的主要特性。

3.1. 安全态势分析

安全态势分析可以针对整体范围或某一特定时间与环境，基于这样的条件进行因素理解与分析，最终形成历史的整体态势以及对未来短期的预测。通过态势分析能够很好的洞察企业内部整体安全状态，通过量化的评判指标能够直观的理解当前态势情况。

3.2. 智能威胁防御

APT 防御技术长久依赖规则引擎，由此产生了不小的时延性、局限性、滞后性等问题。基于大数据技术的智能威胁防御技术打破传统 APT 防御手段中对于大数据量的存储问题、调查问题、模型归纳问题等。实现背景数据过滤，对象数据提取，环境数据集成，分析模型运算，数据结果展现等功能。与传统手段相比，能够更加有效的呈现高级威胁、刻画安全状态、预测未来趋势。

3.3. 隐秘通道挖掘

狭义的隐秘通道专指系统后门通道，广义的隐秘通道指利用合法网络载荷交换非法数据。无论是何种的隐秘通道都需要通过数据交换才能传送非法数据，传统技术缺少逐一甄别每条网络通讯信令的能力，而大数据技术及机器学习算法的应用可以有效的识别出隐秘通道特征，从而实现对隐秘通道的挖掘以及还原其所传送的数据。

隐秘通道的传输为防止被截获，通常其所传输的信令经还原后都是难以解读的片段信息。只有将全部信令都获取到，才能解读出信令中的真实信息。正是由于此特征，传统的技术手段无法在海量网络通讯中进行逐一甄别，片段拼装，信令还原工作，才使得隐秘通道一直围绕着我们。目前通过采用大数据技术，不接能够解决上述问题，同时还能够进行溯源分析等相关的安全分析工作，能够使得隐秘通道无所遁形。

3.4. 用户行为分析

自动归纳用户行为模型，针对用户行为偏离进行告警。实现整体用户行为模型，亦可针对特定用户形成单一行为模型。行为模型以用户行为基准，采用机器学习技术进行自动校正，无需进行人工干预便可实现基线修正以及行为偏离告警。

通过行为模型的自我修正解决传统技术中对于规则的依赖，不再需要依赖人工对于规则的不断完善。这不仅释放大量管理员的工作时间，同时也能更好的摆脱规则间的冲突问题。由系统自动进行基线计算机调整，能够避免由于人工误操作带来的误差，能够将误差降到最低。

3.5. 实时安全监测

实时进行网络安全监测，不仅消除的安全孤岛所导致的数据割裂问题，同时能够实时监测网络中各组成部分的安全状态，包括 IPS 单方面监测到的安全事件，IPS 与审计系统关联多发现的高风险网络行为等。

传统的实时监测无法实现对多种安全设备数据的汇总分析，无法实现跨设备的综合监控效果。通常的做法是采用驾驶舱的方式进行实时监测，通过多个单一图表进行不同维度的监测，这种形式无法整体的实时监测效果。这样做更多的是实现单一角度安全实施监测，依旧没有改变管中窥豹的安全格局。

采用新技术的实施安全监测，能够同时针对多维度进行监测。从整体的角度对安全进行全方位安全监测。

3.6. 攻击溯源取证

目前的安全事件难以找到源头，难以发现攻击过程，难以了解安全攻击状态，难以评估受损程度，难以抵御下次攻击。在这些事件中攻击溯源尤其难以掌握，只要掌握住攻击溯源后续问题便可迎刃而解。

攻击溯源目前支持多种方式，如 DDoS 攻击溯源、僵木蠕攻击溯源、APT 攻击溯源、病毒溯源、数据泄露溯源等。所有的攻击行为会以数据方式进行固化保存，即使攻击行为已经结束，并且攻击者消除企业内受影响系统内的日志，他的攻击行为都会被完整记录下来。攻击行为的记录能够作为证据进行永久保存，以便成为未来维权时的有效证据。

3.7. 合规审计

安全设备、网络设备、服务器、业务系统分散在企业各处，有些自带日志管理功能，而有些却连硬盘都没有。针对这样复杂的网络环境，想实现日志合规、操作审计就变得难上加难。而且黑客对于日志的清理已经变为常规内容，保存在本地的日志会被清理到，因此需要将设备、系统中的日志进行异地保存。

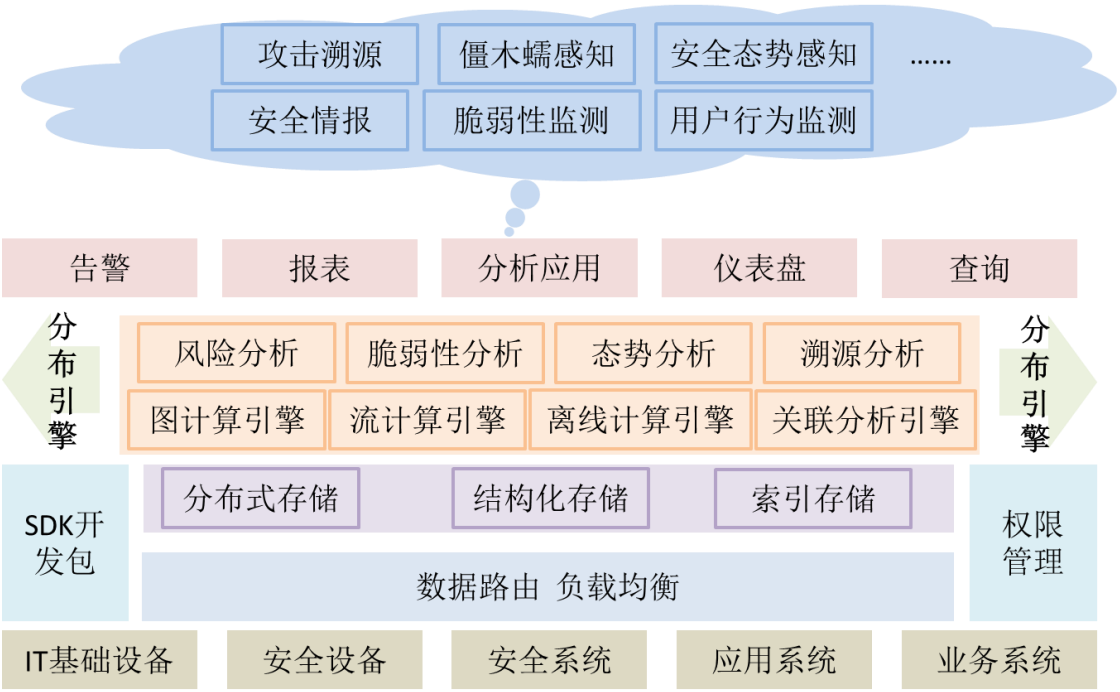
无论在哪种合规性要求中都有对日志的要求，如《信息安全等级保护管理办法》、《证券公司内部控制指引》、《商业银行内部控制指引》都有相关的明确解释。

通过对异构网络环境下各系统的日志异地集中保存，能够很好的满足合规性要求中对于日志审计的规定。同时依托于分析技术的有效应用，不仅能满足合规性要求，同时能够对日志进行多维度分析，挖掘出日志中的潜在价值。

4. 产品特点

大数据架构

采用大数据的底层架构，实现异构数据采集、存储、计算。对于 HBase、Hive 等大数据组件的深度整合，满足网络安全中对于数据有效性、数据完整性、数据及时性的约束要求。采用自主开发的数据路由功能，实现对于不同数据源的区别处理。以底层为基础，实现自主可控的系统架构。



绿盟大数据安全分析平台分为数据采集层、数据路由层、数据存储层、分析引擎层、安全应用层几个部分。

采集层获取与安全紧密关联的海量异构数据，包括 IT 基础设备的性能数据、网络流数据等，安全设备的监测数据、状态数据等，安全系统的告警数据、监测数据等，应用系统的登陆数据、操作日志等，业务系统的告警数据、操作审计数据等；此外还可采集常见的配置信息、威胁情报等相关数据。

4.1. 异构数据源

绿盟大数据安全分析平台能够接收多种数据源，包括但不限于网络设备，如交换机、路由器、网关等；安全设备，如防火墙、入侵防护、网闸、防毒墙等；安全系统，如身份认

证系统、集中授权系统等；应用系统，如邮件系统、OA 系统、数据库系统、中间件系统等；业务系统，如 ERP 系统、CRM 系统等。

所有接入数据源没有品牌限制，没有型号限制，任何设备都可采用 Syslog、Webservice、Snmp 等标准协议进行数据采集。同时亦支持对于网络 Flow 流数据的采集，支持 Netflow 等多种 Flow 协议。

同时，对于缺少数据发送功能的设备提供相应的数据采集器。采集器旁路到网络中进行数据采集工作，包括网络设备数据、安全设备数据、应用系统数据、

4.2. 数据存储

针对类型的不同采用多种数据存储机制。设备、系统数据由于其具有异构性及高并发性，因此采用分布式存储进行数据保存，安全情报数据进行结构化存储，相关重点数据进行索引保存。

实现数据保存的按需使用，同时在大数据量、高并发情境下，通过数据路由技术实现负载均衡数据异地保存，确保数据的持续性、稳定性等要求。

4.3. 计算引擎

平台中预制图计算引擎，流计算引擎，离线计算引擎，关联分析引擎。这些预制引擎构成分析平台的核心功能并且对专项分析提供基础能力，如风险分析、脆弱性分析、态势分析、溯源分析。这些上层专项分析都依托于 4 个主要引擎的支持。

分析引擎采用分布式进行横向扩展，面临海量数据量时能够实现按需扩展，将分析引擎分散到其他更多的机器中，实现按需进行计算资源扩展。

图计算引擎

借助图计算引擎在秒级别内，完成对 TB 数量级的图数据挖掘。如安全业务在在一张大图的基础上，实时进行各部分拓扑结构变化和关系修正，并实时的从图的变化中，进行感知和数据挖掘。依托图计算引擎能够进行复杂攻击路径的计算、基于态势的热力图变化等。

流计算引擎

借助流计算引擎在秒级别内，完成对实时数据的处理工作。如，针对如河川奔流不息的数据进行实时数据计算，并且在计算的同时对结果进行二次计算，并将二次计算结果以可

视化方式进行展现。由传统的数据库计算，转变为入库前实时计算，采用实时数据而非过期数据进行计算。依托流计算引擎能够进行实时安全分析、实时态势感知、实时脆弱性分析等。

离线计算引擎

借助离线计算引擎在小时级别内，完成对 PB 数量级的数据挖掘。如一年内的安全事件之间的相关性，安全事件之间的影响程度，安全事件之间的规律性等并以报表形式进行输出。依托离线计算引擎能够做到历史安全规律统计，历史安全风险分布等。

关联分析引擎

借助关联分析引擎，能够实现关系型数据库以及非关系型数据之间的互访。如，采用威胁情报对现有内部资产的脆弱性分析，可能的攻击路径分析等。依托关联分析引擎能够进行多维度，多数据源安全事件分析。

4.4. 可视化展现

告警

基于规则与阈值的告警触发方式，预制部分告警规则及阈值，同时也支持自定义设置告警规则以及告警阈值。告警能够以多维度进行分析，如时间、设备、数据源等其他自定义方式进行告警分析。

所有告警均为绿盟大数据安全分析平台根据所采集数据分析所得，即使所采集达到的数据中包含告警也需要重新通过告警规则进行筛选。

仪表盘

实现配置型可视化展现，安全分析人员不需要进行代码编辑便可将查询结果以可视化方式进行展现。以拖拽及配置方式很好的进行可视化呈现，同时仪表盘之间能够实现联动以及下钻等强互动性操作。

可视化展现支持多种常见图形，如折线图、饼状图、柱状图、条形图等。同时对于复杂展示方式，如热力图、散点图、图标叠加可以通过 SDK 包的形式进行定制展示。

查询

不仅支持简单的关键字查询，同时在查询中能够进行统计计算，包括平均值计算等在内的多种运算规则。查询时能够以多维度进行筛选，例如时间维度、设备维度、数据类型维

度等。支持图形化查询结果展示，关键字排序，关键字高亮等技术。对于关键字能够实现按需扩充，对于缺少的关键字段能够进行补充扩展。

4.5. 安全应用扩展

类似 Iphone 手机的 APP store 功能，所需安全应用提供在线、离线安装。分享来自全国安全专家亲手炮制的安全分析应用。安全应用均有相应团队进行技术保障和技术升级，在使用过程中能够与安全专家进行经验分享。安全应用具备开箱即用功能，安装安全应用后能够即可使用。

通过 SDK 包可以为企业量身打造安全分析应用，不需要借助第三方公司企业自身便可进行安全应用的开发。同时安全应用可上传至云端绿盟安全应用市场由安全专家检验该应用的可用性及通用性。

4.6. SDK 开发包

SDK 开发包中预制丰富的接口，包括前端展示的、计算引擎、数据接入等。前端展示包括 RESTful、Webservice 等传输协议。通多 SDK 包能够进行丰富的前端展现扩展，包括基于地图的热力图，平行坐标等复杂展示方式。通过调用计算引擎中的预制接口，能够将结果直接抽取到第三方系统进行数据集成。数据接入接口能够实现 ETL 工作，针对特殊数据进行脱敏、转换等数据操作。

5. 大数据技术应用

5.1. 数据路由

基于数据来源广、数据量大、数据类型多等特定，集中化的数据存储会存在并发风险、IO 风险、资源风险。因此在对于这类数据情况需要使用数据路由技术，通过采用绿盟科技的数据路由技术能够很好的分散数据压力，实时调整数据所需的计算资源等。

数据路由采用专有路由算法做支撑，算法通过自我矫正确保数据处理能被迅速有效的处理，不产生数据拥塞、数据丢失等现象。

5.2. 数据存储

数据存储采用 NoSQL 以及 RDBM 等技术方式，实现数据的高效保存以及高速检索。针对安全数据采用 NoSQL 进行数据保存，满足安全数据的异构性、高速性等特点，针对资产数据、情报数据等采用 RDBM 进行数据保存，满足对于事务性数据保存的要求。

同时针对 NoSQL 与 RDBM 之间采用数据融合技术，实现非结构化数据库与结构化数据库之间的数据通讯，确保数据库之间的访问安全可靠。

5.3. 索引技术

全文索引技术，可针对所有入库的数据进行全文索引，同时也支持自定义方式进行关键字索引，从而实现快速信息检索。自定义索引可以按需进行修改，索引与原数据之间具有很好的防护机制，即使索引由于设置问题导致数据不一致，也可以进行快速恢复。

支持跨服务器、跨数据源、分布式的信息索引技术，打通结构化数据与非结构化之间的裂隙。通过跨服务器、跨数据源、分布式的信息索引技术实现索引时间最小化、服务器扩充简易化。支持多种索引及计算方式，支持按词索引、按字索引、混合索引，索引计算等。

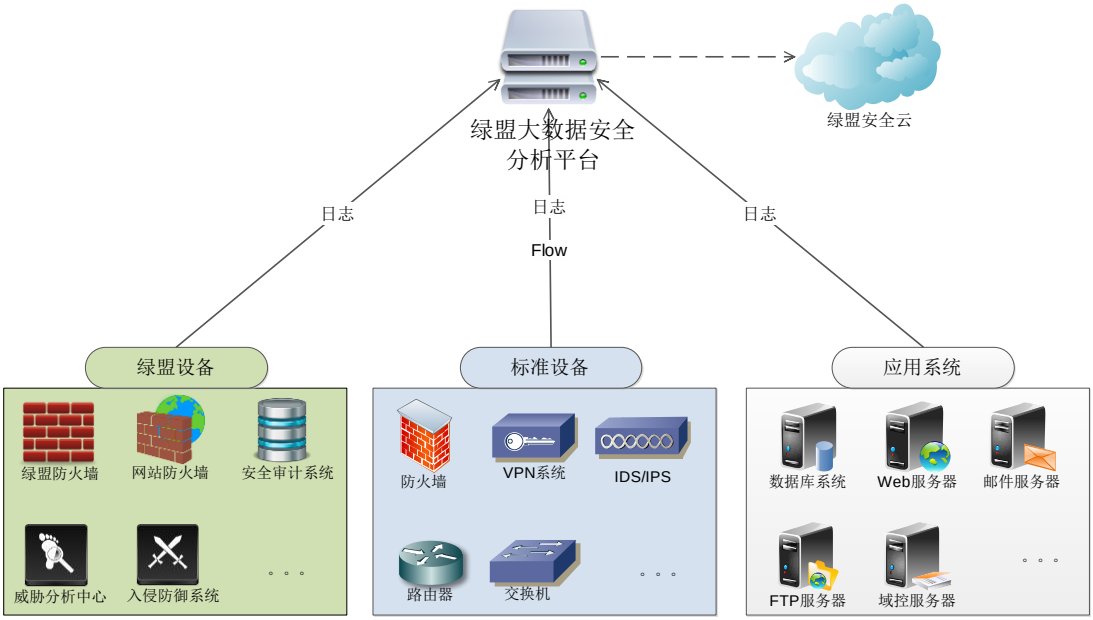
5.4. 安全可视化

可配置图形分析工具的应用，使得数据分析结果能够以更为简洁的方式进行展现。可视化工具能够绘制出常见的图形，如折线图、柱状图、条形图、饼状图、表格等。图形中的维度可以进行任意调整，图形中的被衡量的数据也可进行调整。并且能够自动实现图形与表格的相互转换。

分析人员通过配置的方式能够构建出满意的可视化展示风格，同时可以将可视化效果进行任意组合。将多个可视化图形拖拽至同一画布中进行集中展现，方便快捷的以多角度进行分析结果监测。多个可视化图形间可采用不同时间周期，方便以多个时间维度进行分析监测。

6. 系统环境

6.1. 部署环境



BSA 部署在企业内网，只要网络可达即可。当需要接收 Flow 流时部署在核心交换机旁，其他情况网络部署位置任意。

BSA 所接收的数据来自网络设备的 Flow 流、syslog，接收绿盟设备的日志，接收标准设备的 syslog 日志，应用系统的日志信息。接收到的数据进行统计分析，查找违规事件以及相关信息。

硬件服务器配置推荐：

项目	内容	备注
推荐型号	DELL R720(2U)	
CPU	E5-2630v2×2	主频 2.6G 共 12 核/24 线程
内存	64GB	
Raid 卡	带缓存的 raid 卡	建议 H710p
系统磁盘	1.2T×2	万转机械硬盘（或 SSD）
数据磁盘	4T×6	7200 转机械硬盘
网卡	千兆网卡	2 块以上

6.2. 接口说明

绿盟大数据安全分析平台对外提供丰富的接口，可分为数据采集接口、分析引擎接口、可视化接口以及集成接口。

采集接口

绿盟大数据平台可支持安全事件采集、安全告警采集、Flow 采集、安全情报采集。直通标准通信协议，如 Syslog，NetFlow 等。同时对于特殊要求能够进行定制开发。

分析引擎接口

绿盟大数据数安全分析平台提供分析引擎接口。通过封装后的 SQL92 进行接口调用，降低开发门槛，实现外部进行接口调用以及结果共享。

可视化接口

绿盟大数据数安全分析平台提供可视化接口。通过接口该平台能够将可视化展示到第三方系统。如，RESTful 等。

集成接口

提供与第三方系统集成接口，如与 ITIL 系统接口、集中告警接口等，以满足企业内标准统一的要求。

7. 产品优势及特点

- 完整的海量异构数据采集与处理

- 各类大数据高速存储
- 海量数据的流式实时及历史分析
- 丰富易用的可视化的数据展示
- 可视化分析规则和算法编辑，可视化分析查询策略编辑
- 提供丰富且易于扩展的数据处理与分析组件及接口
- 兼容现有 SOC/SIEM 系统

8.产品的客户价值

绿盟大数据安全分析平台可帮助企业解决传统安全分析中难以解决的数据量大、难以使用的问题，提供强大高效的安全分析工具，保护现有投资，实现企业资源的最大价值。

- 解决海量异构数据的存储、分析、展现问题；
- 实时数据与离线数据的组合分析，了解过去，掌握现在，预测未来；
- 使得安全分析人员工作效率大幅提升；
- 实现安全威胁分析的统一化，集约化；
- 支持丰富的安全分析扩展，实现经验共享；
- 具有安全取证功能，全方位保护企业安全信息；
- 具备安全情报分析能力，将安全情报与安全信息相融合；
- 兼容企业已有的安全系统以及设备，如 SOC、SIEM 等；
- 简单高效的安全可视化，降低分析人员的培训成本及时间成本；
- 实现按需进行资源扩展，降低不必要的资源浪费。