sepian

Seguridad Digital
de Andalucía

Charla virtual

Aplicaciones del Big Data a la seguridad informática

Convergencia entre Big Data y Ciberseguridad

30/10/2020

Junta de Andalucía

¿Quién soy?

Trayectoria:

- CEO / CSO en Auditech
- Docente en Máster en Dirección de Ciberseguridad, Hacking Ético y Seguridad Ofensiva
- Senior Security Application Tester en Capgemini
- Lead Security Analyst & Ethical Hacker

Formación:

- UDIMA - Ciberseguridad y Peritaje Informático Judicial.
- Postgrado en Seguridad Informática para la Intrusión de Sistemas.

https://www.linkedin.com/in/josuelopezmartin22/

Introducción y conceptos

¿Qué es el Big Data?

El **Big Data** es un **conjunto de datos** que debido a su **complejidad, volumen**, **velocidad de crecimiento, variabilidad** y **tratamiento**, requiere de herramientas de software específicas para poder manejarse de forma adecuada.

Tipos de datos según su estructura

DATOS ESTRUCTURADOS	DATOS NO ESTRUCTURADOS	DATOS SEMI-ESTRUCTURADOS
• Datos tradicionalmente tratados • Almacenamiento en tablas y clara definición de longitud y formato. • Ej. Números, cadenas de texto y fechas.	• Datos en su forma original (raw) que han sido recogidos. • No poseen un formato específico que permita almacenarlos de forma tradicional. • Su contenido no se puede desglosar en tipos de datos definidos por longitud y formato. • Emails, archivos de texto y multimedia, PDF's...	• Siguen una estructura con pautas comunes, pero no lo suficiente para gestionarla como datos estructurados. • Poseen cierta estructura estandarizada pero no es aplicable a todos los datos. • Posee patrones comunes que dan información sobre la relación de los datos. • El lenguaje HTML permite identificar pautas a través de etiquetas.

Características principales del BIG DATA

Velocidad	Velocidad a la que se generan los datos.
Volumen	El tamaño de los datos (creciendo por segundos)
Variedad	Diferentes tipos de datos.
Veracidad	La fiabilidad y calidad de los datos recogidos.
Valor	Cada dato aporta un pequeño valor al análisis, un conjunto de ellos un gran valor.
Visualización	Accesibilidad y lectura de los datos de forma sencilla.
Variabilidad	Datos en constante cambio

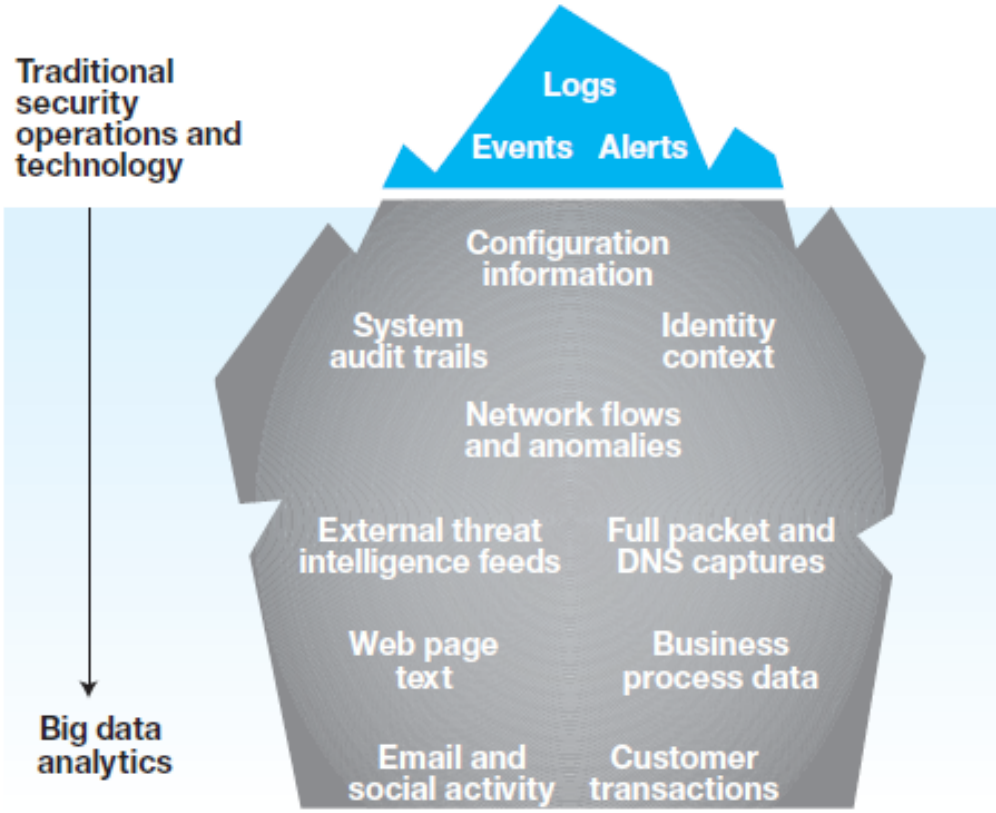
Convergencia entre Big Data y Ciberseguridad

La ciberseguridad en contexto, **no es un fin**, si no, **un proceso** sobre el que se debe **Iterar** recurrentemente para garantizar su fiabilidad y éxito.

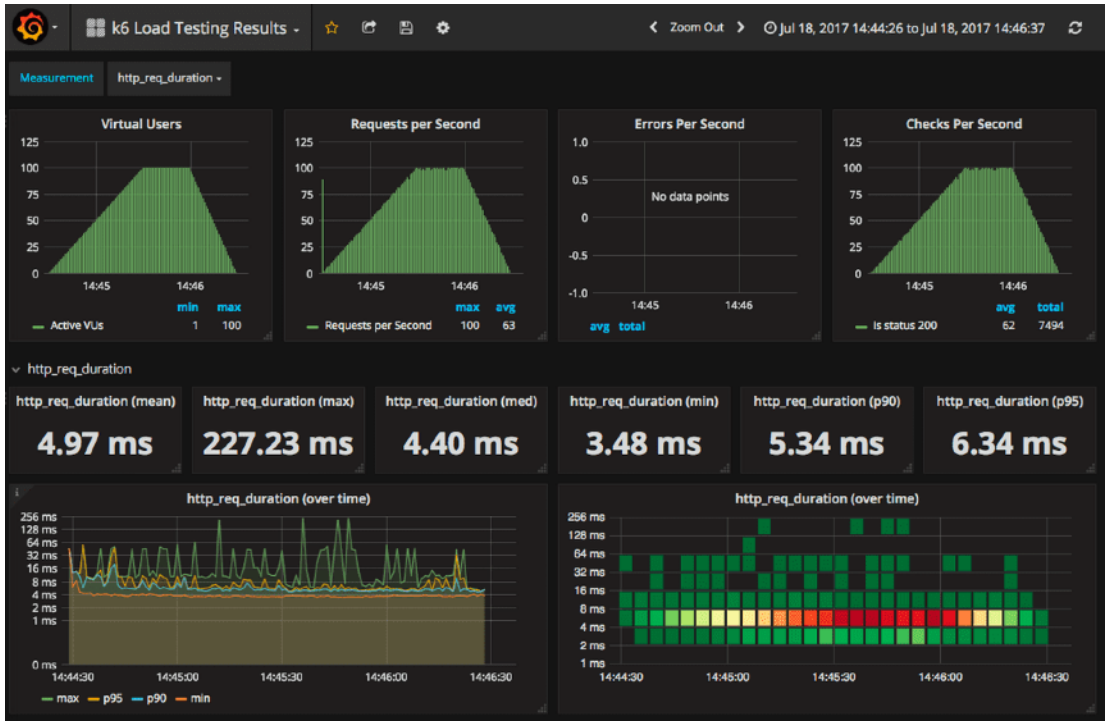
El Big Data **mejora** las fases y el proceso proporcionando nuevas fases como la **predicción**.

Algunas de las ventajas son:

- Detección y respuesta en tiempo real.
- Mejora de la fase de respuesta
- Predicción



Aplicaciones del BD en Ciberseguridad



Monitorización en Tiempo Real

El **Big Data** en una monitorización activa de la infraestructura y sistemas de una organización, puede detectar posibles violaciones de seguridad.

Implementando **marcas**, en los **archivos**, podrán ser identificados en tiempo real, garantizando la **Integridad** y **Seguridad**.

De esta forma, los usuarios pueden almacenar y enviar información sin preocupación de ser víctima de un ciberataque.

Seguridad Ofensiva / IA Hacker

El Big Data ha revolucionado el concepto de Hacking y Threat Intelligence.

Las empresas buscan de forma proactiva a los ciberdelincuentes con el objetivo de detenerlos.

Herramientas y soluciones basadas en Big Data, se encargan de **detectar patrones de pentesting** y **automatizar el proceso de explotación de vulnerabilidades** gracias al **Machine Learning** que nos permiten los sistemas de BD.

Clustering de Malware

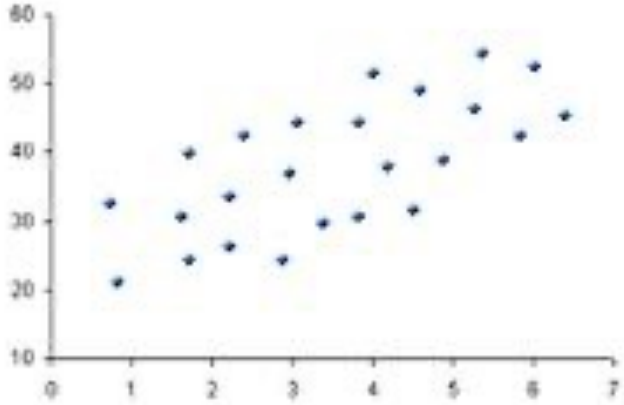
El objetivo de estos sistemas, es la **clasificación de muestras de malware según sus características**, estas agrupaciones se denominan **clústers**.

La imposibilidad de tratar miles de muestras en un corto plazo de tiempo da lugar a estos sistemas.

Tras el análisis realizado, el malware es **clasificado, procesado e indexado** junto con su muestra y sus características en el clúster correspondiente.

De esta manera obtenemos gran información de los diferentes tipos de malware para su clasificación y posterior uso.

Análisis Predictivo / Correlación



El nuevo paradigma que se abre con la entrada del Big Data en Ciberseguridad, es la capacidad de predecir comportamientos en los sistemas.

Existen multitud de fuentes para alimentar el Big Data (Logs)

Cambiamos el enfoque tradicional Pasivo a Proactivo.

La analítica predictiva permite identificar patrones y tendencias de comportamiento, lo que nos da capacidad de anticiparnos al ciberataque.

Protección de datos Big Data

Seguridad y Privacidad en entornos Big Data

El acceso a los datos debe estar Autorizado, Organizado y con los permisos de usuario necesarios.

La complejidad para mantener un equilibrio entre Ciberseguridad y Privacidad es alta cuando los datos, su alojamiento y procesamiento están siempre en constante movimiento.

Para proteger los datos, podemos encriptar y tokenizar los datos de forma que independientemente de quien tenga acceso al dato puedan ejecutarse los análisis necesarios sin exponer ninguno de estos datos.

Bastionado de Entornos Big Data (I)

Además de las medidas especialmente tomadas para los datos, se establecen medidas para asegurar la infraestructura.

1. Acceso y Autorización granular a los datos.

Los controles se definen a través de expresiones de control de acceso y autorización de datos.

Con permisos basados en Roles y configuraciones de visibilidad.

En los niveles más bajos se ocultan los datos y en la parte superior se tienen los datos confidenciales a disposición de los técnicos que deban analizarlos.

Bastionado de Entornos Big Data (II)

2. Seguridad perimetral, protección de datos y autenticación integrada.

Seguridad perimetral y firewalls alrededor de los sistemas de información.

La autenticación, se debe dar contra sistemas probados.

Es decir, integraciones con **LDAP**, **Active Directory** y otros servicios de directorios que permitan una **gestión de identidades optima** y centralizada.

Bastionado de Entornos Big Data (III)

3. Constante Auditoría y análisis

El nivel de **visibilidad y responsabilidad** en cada paso del proceso es lo que permite "**Gobernar**" los datos en lugar de simplemente establecer políticas y controles de accesos.

A través de **auditorias**, las empresas pueden mantener sus estrategias actualizadas en un entorno en el que la forma en la que vemos los datos y **tecnologías que utilizamos para administrarlos y analizarlos cambian cada día**.

Es fundamental poder **rastrear el acceso** y **reconocer patrones** en los datos.

Dudas

sedian Seguridad Digital de Andalucía