IMPLICACIONES DE SEGURIDAD DE BIG DATA

UN ESTUDIO DE:

CSA ES cloud security SPAIN alliance℠

EN COLABORACIÓN CON:

ISMS forum spain

Implicaciones de seguridad de Big Data

Estudio elaborado por el capítulo español de Cloud Security Alliance

Coordinador

Miguel Ángel Pantoja (Hewlett-Packard Enterprise)

Autores

Luis Buezo (Hewlett-Packard Enterprise)

Beatriz Blanco (Amadeus)

Sara Degli-Esposti (ISMS Forum)

Enrique Aristi Rodríguez (UCI E.F.C.)

Mariano J. Benito (GMV)

Humbert Costas (Ackcent Cybersecurity)

Óscar Escudero

Contenido

Introducción

El conjunto de nuevas soluciones tecnológicas que permiten a las organizaciones gestionar mejor su información, comúnmente conocidas como "Big Data", tienen un protagonismo creciente en todo tipo de organizaciones públicas y privadas. Hoy en día se generan más datos, se analiza más información y se consumen más resultados de análisis que nunca antes. El tiempo que media entre la aparición de los datos y la toma de decisiones basada en los mismos es cada vez menor, y cada vez mayor la importancia económica y social de dichas decisiones.

Big Data se caracteriza por la recopilación en tiempo real, el almacenamiento y el análisis de ingentes cantidades de datos en distintos formatos. Las organizaciones están poniendo los datos en el centro de su gestión y de su operación. Esto amplía sus posibilidades de acción y competencia pero trae consigo riesgos nuevos, posiblemente no contemplados íntegramente en los análisis de seguridad convencionales.

Big Data son soluciones tecnológicas relativamente nuevas y que aún está en fases iniciales de adopción en muchas organizaciones, lo cual es un riesgo en sí mismo. Por su naturaleza, Big Data se consume en gran medida desde el Cloud. Es pertinente, pues, que el **Capitulo Español** del **Cloud Security Alliance (CSA-ES)** aborde la cuestión. Los campos técnicos y las aplicaciones de Big Data están en expansión y, aunque algunas están empezando a madurar, es el momento de insistir en la seguridad.

El "ecosistema Hadoop" que, aunque no sea el único, sí es el principal entorno técnico que se identifica con Big Data, se desarrolló en sus comienzos con una atención muy limitada a la seguridad. Aunque esto se está corrigiendo, no deja de constituir una potencial fuente de riesgo. También lo es que Hadoop esté pasando de ser un sistema de propósito limitado (fundamentalmente, almacenamiento barato linealmente escalable y procesamiento por lotes de datos no estructurados) a dar pasos hacia un verdadero sistema operativo distribuido (almacenamiento; gestión de recursos y programación; motores de procesamiento por lotes, online y cuasi tiempo real; contenedores; incluso monitores transaccionales).

Si como infraestructura va adquiriendo un propósito general, el concepto de "Data Lake" le confiere un papel creciente en la arquitectura de datos empresarial. No se trata sólo de un almacenamiento de bajo coste, sino de la residencia compartida de todo tipo de datos que son accedidos simultáneamente por una variedad de motores de análisis sin apenas fricción. La potencia de este concepto sólo es comparable a la de los riesgos que introduce. El gobierno de los datos se convierte en una disciplina fundamental que corre paralela a la del gobierno de la seguridad.

Por su propia naturaleza, Big Data obtiene conocimiento de donde al principio sólo hay datos. Aparece aquí un potencial conflicto de propiedad y de privacidad. No siempre quienes generan los datos son conscientes del valor que contienen, incluso económico, y los ceden a empresas y administraciones públicas sin reflexión ni contrapartida. Surgen situaciones inesperadas de generación de información personalmente identificable, de forma que los metadatos superan en importancia a los propios datos. Se alzan voces que advierten de nuevas formas de control social y que reclaman la distribución libre del conocimiento de las dinámicas humanas que permite Big Data. Se habla de “Open Data” en analogía con el movimiento “Open Source” en el que precisamente se basa Hadoop.

Aunque todavía en fases iniciales, no es exagerado decir que campos de aplicación como “Smart Cities”, “Internet of Things” y la pléyade de realizaciones del “Machine Learning” van a cambiar nuestras sociedades y que junto con sus indudables beneficios vendrán los riesgos. Se unirán a los que ya acumulamos en las redes de distribución de energía, de telecomunicaciones o de producción de bienes básicos. Podrían ser incluso más vulnerables: son concebibles, por ejemplo, ataques de inyección de datos maliciosos a los sistemas de supervisión y control de fábricas o ciudades basados en Big Data.

Big Data, Cloud, movilidad y seguridad son los cuatro pilares de la IT de nuestro tiempo. Las interacciones entre los cuatro son numerosas. Este estudio quiere poner de manifiesto, a nivel ejecutivo, las muchas que existen entre Big Data y seguridad con la intención de contribuir a abrir perspectivas y generar debate en el mundo de habla hispana.

1. Gobierno, Riesgo y Cumplimiento Normativo

El Big Data representa grandes oportunidades para las organizaciones, pero también comporta importantes riesgos y obligaciones que hay que tener en cuenta a la hora del tratamiento y la circulación de la información.[i] Como parte de esa información puede directamente o indirectamente reconducirse a individuos, el Big Data conlleva importantes riesgos para la privacidad de los personas.[ii] Por esa razón varias agencias de protección de datos han tratado el tema de las implicaciones de privacidad del Big Data, y se han estudiado distintas soluciones técnicas para minimizar los riesgos que la recolección masiva de datos personales genera.[iii]

En este sentido, las organizaciones deben ser especialmente cautas con los riesgos asociados a sus procesos de identificación, análisis y recolección indiscriminada de información, concediendo especial atención al peligro de la violación de la privacidad de los datos. Los datos individuales sólo pueden ser usados para el propósito del servicio que se ha contratado. El único propósito en que los datos pueden estar individualizados y con toda su riqueza de detalles en el contexto del servicio principal que ofrece la organización que nos presta el servicio.

1.1 Privacidad

Big Data implica por definición un riesgo para la privacidad de los individuos. La generación diaria de miles de puntos de datos (adónde vamos, con quién nos comunicamos, qué leemos, qué compramos, qué comemos, etc.), nos hace cada vez más vulnerables a la exposición de nuestra privacidad. Un riesgo derivado de la exposición de la privacidad son las discriminaciones manifiestas.[iv] Si bien estas han sido ilegales durante décadas, Big Data proporciona la capacidad para tomar decisiones discriminatorias sin la necesidad de dejar una evidencia explícita y evidente gracias a la "automatización" de los análisis.

Amenazas, problemas o contradicciones

Muchas de las ventajas generadas a través del análisis masivo de datos se basan en resultados agregados donde no se necesita identificar al individuo que esté aportando la información. En lo que se refiere a la protección de la información de identificación personal (PII), existen métodos de disociación de datos como la anonimización, el uso de seudónimos, cifrado, claves de codificación e intercambio de datos para separar los PII de las identidades reales. Sin embargo, la difusión de algoritmos de integración que reúnen información presente en distintas bases de datos relacionada con la misma persona genera riesgos para la privacidad.[v] Recolectar y almacenar datos es tan barato que varias empresas generan repositorios sin saber exactamente qué harán con esa información. La venta de base de datos en el mercado representa una oportunidad lucrativa para muchos.

Controles, soluciones o compromisos existentes en la práctica

La adopción de principios de minimización de la información podría ayudar a entender que datos la empresa está procesando y con qué finalidad. Una vez que se haya identificado el conjunto de datos que aportan valor y que la empresa verdaderamente necesita, hay muchas formas de proteger la confidencialidad de esa información.

En una primera aproximación, el enmascaramiento de datos podría parecer una solución al problema de las discriminaciones manifiestas. Sin embargo, el análisis de datos de forma masiva podría revelar fácilmente la identidad de las personas reales a través de la asociación y combinación de datos. Otra posibilidad de proteger la privacidad de los datos en entornos Big Data es a través de técnicas de ofuscación, enmascaramiento, e introducción de ruido en los datos. En este caso es importante analizar e identificar aquellas propiedades de los datos que habrá que preservar para asegurarnos un análisis fiable de los datos.[vi] Las técnicas aquí mencionadas brindan oportunidades tanto para el sector privado como para el sector público, que las puede utilizar para llevar a cabo proyectos de Open Data.[vii]

1.2 Legislación sobre protección de datos

En la actualidad no existe legislación específica para Big Data. En un proyecto Big Data, si serán de aplicación las leyes que ya afectan a los distintos conjuntos de datos.[viii] Por su parte, la protección de la información de identificación personal (PII) en las implementaciones de Big Data sigue siendo una gran preocupación ya que la tecnología actual diseñada para proteger dicha información no puede garantizar su seguridad.[ix]

España y Alemania están a la cabeza de los países con la normativa de protección más restrictiva. La española Ley Orgánica de Protección de Datos (LOPD)[x] es posiblemente la más exigente a nivel mundial. Por otro lado, la Unión Europea se están tomando las medidas necesarias en materia de protección de datos, a través de directivas como la Directiva 95/46/CE[xi] sobre la protección de las personas físicas en el tratamiento de datos personales y su libre circulación o la más reciente Directiva 2009/136/CE.[xii]

Amenazas, problemas o contradicciones

Al analizar el fenómeno Big Data, y las posibilidades de recopilación y retención indefinida de los datos que ofrece, varios expertos han llegado a la conclusión que principios básicos de protección de datos, como el principio de minimización de los datos, o el de limitación y especificación de las finalidades de uso de los datos, se han hecho obsoleto en la época del Big Data.[xiii] Esta visión ha sido contrastada por parte del regulador Europeo.[xiv] La protección de los datos y el Big Data son compatibles,[xv] particularmente el análisis de Big Data, que es el que aporta valor, y la privacidad de los datos.[xvi]

Otro de los aspectos más polémicos para las empresas sobre la privacidad en Big Data es la obtención del consentimiento o permiso para recopilar y utilizar los datos personales. La realidad es que es imposible identificar a todas las organizaciones que podrían haber recogido información sobre un individuo.

La normativa Europea dicta que los datos tienen “límites geográficos” y no se pueden sacar del territorio de la Unión. La Decisión de la Comisión Europea 2000/520/CE de 26 de julio de 2000 consideraba confiables las organizaciones adheridas a los principios de la normativa estadounidense de “Puerto Seguro” (Safe Harbor) y permitía las transferencias de datos a las mismas. Pero muy recientemente (el pasado 6 de octubre de 2015) el Tribunal de Justicia de la Unión Europea ha anulado esa Decisión, dejando en manos de los Estados miembros "decidir si debe suspenderse la transferencia" de la información de usuarios europeos a compañías no europeas como Facebook en EEUU. Esta decisión podría ir mucho más allá de Facebook ya que la sentencia sienta precedente y podría acabar afectando a otras firmas tecnológicas estadounidenses con presencia en Europa como Apple, Google, Microsoft o Amazon.[xvii]

Controles, soluciones o compromisos existentes en la práctica

Aparentemente, anonimizar los datos permite en muchos casos eludir la legislación europea de protección de datos, además de reducir los riesgos de un robo de datos. En los últimos tiempos la eficacia de las herramientas de anonimización y de pseudonimización ha sido cuestionada por distintos autores que han demostrado en sus estudios cómo se puedan reidentificar los datos, con cierto margen de error, a partir de la integración con bases de datos públicas.[xviii] No obstante, parte de estos temores son infundados.[xix]

Otras formas de limitar los riesgos mencionados anteriormente, que tengan que ver más con el comportamiento del usuario final, serían:

- Sensibilizar al público, y en especial a los menores y jóvenes, para evitar la publicación de tanta información personal en las redes sociales.
- Legislar para que las empresas no pidan a los usuarios más información que la estrictamente necesaria para proporcionar el servicio.
- Utilizar un navegador anónimo, como Hotspot Shield o Tor (The Onion Router) para visitar sitios que podrían producir información inexacta acerca de los usuarios.

Por otra parte, en lo que respecta a la obtención del consentimiento para el uso de datos, una práctica que podría ayudar a las personas restaurar el "control" de sus datos de carácter personal sería la de permitir que se eliminen sus datos y sean purgados por completo a fin de proteger la privacidad del consumidor.

Los usuarios deberían tener la capacidad de gestionar el flujo de su información privada y de poder especificar, en un nivel muy granular, lo que están consintiendo. Los usuarios cuyos datos se están tratando, necesitan una visión transparente de cómo se están utilizando o incluso vendiendo dichos datos.

1.3. Evaluación y gestión de riesgos

Al adoptar nuevas soluciones tecnológicas como Big Data, todos los riesgos deben ser identificados y gestionados. Eso incluye garantizar valores de los activos y hacer frente, entre otros, a los riesgos legales y regulatorios.

El primer paso para un uso eficaz y rentable de Big Data es la competencia en la adquisición y gestión de servicios en el Cloud. Debe haber responsabilidades bien definidas tanto para el proveedor de servicios en el Cloud como para el usuario de servicios Cloud respecto a los controles específicos de protección de datos que se requieren. También debe haber vigilancia y auditorías de los servicios en Cloud junto con las métricas relevantes que indican los niveles de integridad de datos, confidencialidad y disponibilidad.

En cuanto a la clasificación de datos, el almacenamiento de grandes volúmenes de datos conlleva un alto riesgo, ya que un solo evento puede resultar en una gran violación de datos en cascada. En Big Data puede darse la circunstancia de que elementos de datos altamente sensibles no estén separados de los elementos de datos menos sensibles. En consecuencia, la clasificación de los elementos más sensibles determinará la clasificación de los datos de todo el conjunto.

Por último, debe existir un modelo de gestión de la información que garantice por un lado la seguridad, privacidad y cumplimiento de la regulación vigente y por otro la calidad, fiabilidad y rendimiento de la infraestructura, asegurando el cumplimiento a través de acuerdos de nivel de servicio (SLAs).

Existe la necesidad en el mercado de un estándar o metodología de evaluación y pruebas adecuada y precisa que permita evaluar la seguridad de entornos y aplicaciones Big Data.

2. Protección de activos

2.1. Seguridad de las infraestructuras

Amenazas, problemas o contradicciones

Desde el punto de vista arquitectónico, Big Data exacerba los problemas de seguridad de los sistemas convencionales por la confluencia de varias circunstancias:

- Descuido inicial de la seguridad.
- Gran escala de datos y computación.
- Recogida de datos de múltiples sistemas cuyos problemas hereda y tendencia a no descartar datos.
- Retos en privacidad.

El diseño de los sistemas de procesamiento distribuido, y en especial los que dieron origen a Big Data, primó inicialmente la funcionalidad y dejó de lado la seguridad.[xx] Sólo así se puede explicar, por ejemplo, que la instalación de algunas distribuciones de Hadoop debe realizarla el usuario Root, algo inadmisible para muchas políticas de seguridad.

Todos los sistemas Big Data presentan una arquitectura de Cluster que permite escalar (muy notablemente en el caso de Hadoop, de forma lineal). Con un gran número de nodos fuertemente conectados, el mal funcionamiento o el ataque a un número pequeño de unidades puede comprometer al conjunto del sistema. La autenticación de usuarios, tanto personales como de máquina, debe cuidarse especialmente. Esto no ha sido siempre así: existen ejemplos de distribuciones de Hadoop que utilizan usuarios sin contraseña para ciertas operaciones entre nodos.

Al actuar como colectores de datos, los sistemas de Big Data se ven afectados por todos los problemas de seguridad e identidad de las bases y fuentes de datos de las que dependen. En esto no se diferencian de otros sistemas, pero las medidas de aseguramiento necesarias en este caso deben diseñarse a mucha mayor escala. Por otro lado, las tecnologías de Big Data pueden ayudar a detectar problemas y ataques, y así contribuir a la seguridad.

El creciente número de usuarios de sistemas Big Data representa otra fuente de problemas de seguridad. Los analistas pueden no tener la formación necesaria en seguridad, pero sí el acceso a ingentes cantidades de datos. Su número hará que los ataques internos también aumenten y que aumente el riesgo que los mismos analistas roben o adulteren los datos.

El consumo de los resultados de Big Data excede con mucho al que tenían los resultados de las tecnologías precedentes de Business Intelligence y análisis en general. El desarrollo de Big Data ha coincidido con el de la movilidad y en gran medida ambas tendencias se han realimentado positivamente. No es imposible pensar en ataques consistentes en la alteración de dichos resultados, que a menudo dan lugar a la toma inmediata de decisiones.

Las conclusiones que provisionalmente cabe extraer son las siguientes:

a) La seguridad debe integrarse desde el principio en el diseño de las tecnologías Big Data y no añadirse después como soluciones ad hoc, urgentes y con modelos amenazas poco sistemáticos.
b) Deben securizarse tanto la recolección de datos como su agregación y diseminación.
c) Por actuar como colectores de datos, los sistemas Big Data soportan problemas de seguridad de otros sistemas. Arquitecturas como "Data Lake" pueden ser ingobernables o fuente continua de problemas de seguridad por este motivo y deben ser objeto de especial atención.

Controles, soluciones o compromisos existentes en la práctica

A continuación se exponen algunas de las áreas de la infraestructura de Big Data que introducen nuevos problemas de seguridad.

La cuestión de la computación segura en entornos distribuidos no es exclusiva de Big Data pero presenta aspectos nuevos en este caso. En estos entornos no se han establecido mecanismos de autenticación de los nodos esclavos con respecto a los maestros, ofreciendo así una amplia superficie de ataque. En el caso de Hadoop, los escenarios de ataque contemplan nodos worker maliciosos, mal configurados o comprometidos. Las dos principales medidas de prevención de ataques son la securización de los mappers (véase MapReduce) y la de los datos en presencia de un mapper no confiable, así como un registro más robusto de los nodos esclavos ante los maestros. Se han hecho algunas implementaciones pero limitan el rendimiento.

Es necesario desarrollar mejores prácticas de seguridad para las bases de datos no relacionales (NoSQL). Estas bases tienen varios problemas: falta de integridad transaccional, mecanismos de autenticación laxos, mecanismos de autorización ineficientes o insuficientes, inyección de código malicioso, falta de consistencia y propensión a ataques internos.

Para hacer frente a estos inconvenientes, se recomienda la encriptación de los datos y el cifrado de las comunicaciones; la autenticación de los nodos de los clusters; mecanismos de log extensos y analizados en directo; y fechado de los datos para evitar modificaciones no autorizadas.[xxi] La implementación de estas medidas a menudo choca con la arquitectura de los motores NoSQL y limitaría la escalabilidad y el rendimiento. La alternativa es delegar las funciones de seguridad a un entorno que rodee al NoSQL y esté fuertemente integrado con el sistema operativo en el que residen.

Big Data se caracteriza por incluir en el Business Intelligence las fuentes de Internet (redes sociales, blogs, foros, medios de comunicación…). Por la actualización permanente de esas fuentes y el carácter continuo e inmediato de los análisis, la conexión a Internet puede exponer aún más a las redes internas de la organización. Además de las soluciones convencionales, en ocasiones se plantean esclusas de red que permitan conexiones intermitentes.

Con volúmenes de datos muy grandes se tiende a una gradación de almacenamientos ("multi-tiered storage"). El almacenamiento de menor nivel, más económico, puede contener datos muy sensibles a los que se acceda poco y que resultarán menos protegidos. A menudo esos niveles de almacenamiento se contratarán a proveedores que no necesariamente podrán considerarse de confianza. Incluso sin romper el cifrado, el mero análisis de los movimientos de datos podrá dar información a esos proveedores. La filiación de los datos también se resentirá. Si la misma información se guarda en distintos proveedores, mantener la consistencia será un reto. En general, el mantenimiento de unos logs detallados será crucial para la investigación forense y para la resolución de disputas de los usuarios de los datos con los proveedores de servicios de almacenamiento, o entre los usuarios. Aunque en los últimos años se han propuesto soluciones parciales para esos problemas, aún no se dispone de un enfoque global que, en todo caso, deberá llegar a un compromiso entre seguridad, usabilidad, complejidad y coste.

Un atacante podría suplantar o comprometer sensores y dispositivos de los que se recogen datos para un sistema Big Data. De esa manera podría contaminar o sesgar los datos. Este problema abarca desde los SIEM a las aplicaciones científicas y puede ser muy relevante en entornos de "Internet of Things". El problema se ve agravado por el hecho de que la seguridad de dispositivos móviles está aún poco desarrollada, sobre todo en comparación con la seguridad de los ordenadores personales. La suplantación de usuarios o sensores puede combatirse mediante una adecuada autenticación, posiblemente basada en certificados aunque la gestión de los mismos a la escala de la "Internet of Things" puede ser difícil de abordar. Por otro lado, las propias técnicas de Big Data ofrecen recursos para la detección y filtrado de datos provenientes de sensores comprometidos porque posiblemente aparecerían como valores estadísticamente anómalos.

El cifrado de datos es la herramienta básica de preservación de la integridad y la privacidad en los sistemas Big Data. El problema que introduce es la necesidad de descifrar antes de realizar los análisis y procesamientos que representan el corazón de Big Data, y de cifrar los resultados correspondientes. Además del gasto computacional, el descifrado abre una nueva superficie de ataque puesto que expone los datos en claro durante el análisis, por no hablar de la gestión de claves (aunque los algoritmos de cifrado basado en atributos pueden simplificarla).

2.2. Seguridad de datos

En entornos Big Data, el ciclo de vida del dato se acorta considerablemente y el efecto de las amenazas aumenta al mismo tiempo que lo hace el volumen, la velocidad y la variedad de los datos implicados. Además, los grandes volúmenes de datos gestionados dentro de un universo heterogéneo de tecnologías, junto con las distintas sensibilidades de dichos datos, dificultan la implementación de medidas que prevengan la modificación no autorizada de datos.

Amenazas, problemas o contradicciones

Al requerirse el uso de información personal para la obtención de datos relevantes y perfiles individuales completos, se complica de forma significativa la protección de la privacidad en estos escenarios en los que se procesan y analizan datos de diferente clasificación en un marco de ejecución distribuido.

Como ya es habitual en el Cloud, el cifrado se erige como principal aliado de la privacidad y la confidencialidad de los datos. Sin embargo, los esquemas de Hashing existentes no son aplicables a grandes cantidades de datos y, por tanto, sería necesario el uso de criptografía fuerte para encapsular los datos sensibles junto con el desarrollo de algoritmos que permitan una gestión e intercambio eficiente de claves para gestionar el acceso en Big Data.

Anonimizar los datos de las identidades reales pudiera ser una estrategia atractiva a priori para preservar la privacidad, pero existen técnicas para desanonimizar[xxii] o restablecer la identidad como las empleadas a comienzos de 2013 por expertos en seguridad.

Un tipo de meta-información que también es importante considerar es la referente a la procedencia del dato. Esta información es útil para disponer de detalles acerca de la creación de los datos o bien para complementar los logs de auditoría. Podría incluir, no solo información de procedencia referente a aplicaciones, sino también la asociada a la propia infraestructura Big Data.

Por otra parte, es necesaria la validación y filtrado de los datos de entrada, de forma que se minimicen las posibilidades de inyección de datos maliciosos o la suplantación de la identidad de una fuente.

Controles, soluciones o compromisos existentes en la práctica

Para disponer de la información de procedencia de forma fiable hay que velar por su integridad, para lo cual se torna crítica la existencia de un control de acceso de grano fino y escalable en el tiempo. Para garantizar la exactitud de la información de procedencia, se deben implementar chequeos de integridad y sondeos periódicos que comprueben la salud de las fuentes.

Además, ya existe actualmente la posibilidad de computación sobre información cifrada, es decir, sin necesidad de ser descifrados previamente a la computación como es el caso de la denominado cifrado homomórfico (FHE, fully homomorfhic encryption en inglés) Los datos se envían cifrados a los usuarios y solo aquellos usuarios con la clave adecuada pueden descifrarlos. Sin embargo, esta aproximación tiene el inconveniente en la lentitud que incorpora al proceso.[xxiii]

La solución a la problemática de la validación de entradas podría consistir, bien en técnicas que prevengan la generación y envío de datos maliciosos (por ejemplo, software que dificulte su propia modificación, defensas contra ataques Sybil, etc.) o bien técnicas que detecten y filtren inyecciones maliciosas. Ambos enfoques son compatibles con las capacidades Big Data de una solución SIEM de análisis de la seguridad en tiempo real.

3. Gestión del Ciclo de Vida del Dato

3.1 Gestión de accesos e identidades

En las bases de datos convencionales, la propiedad de la base otorga los privilegios para crear, leer, actualizar y eliminar datos. La transparencia en torno a la propiedad asegura cierto nivel de confianza y control sobre la calidad de los datos. Conocer la procedencia de los datos permite establecer la trazabilidad a lo largo de su ciclo de vida. La situación se vuelve más confusa en Big Data debido a la multiplicidad de propietarios de datos.

La gran cantidad de actores, de tecnologías y dispositivos empleados así como la variedad de sensibilidades en la información que participan en la cadena de valor Big Data, hacen de la granularidad uno de los pilares fundamentales, tanto desde el punto de vista del control de acceso como de las capacidades de auditoría.

Amenazas, problemas o contradicciones

La posibilidad de que un mismo objeto sea gestionado de formas distintas en diferentes partes del ecosistema podría generar brechas en la seguridad del conjunto, como pudieran ser privilegios no ajustados debidamente, modificaciones no realizadas adecuadamente, o ausencia de la debida segregación de funciones.[xxiv]

Las restricciones de cada implementación concreta de cada Big Data, así como los distintos marcos legales de aplicación, exigen que solo aquellos elementos autorizados (bien sean personas o componentes) accedan a distintas porciones de información, tanto dentro del proceso de transformación analítica como en el de generación y gestión de pistas de auditoría. Esta circunstancia hace imprescindible una clasificación de la información con el grano fino adecuado y una definición correcta de los accesos. Por lo tanto, acertar con el nivel de granularidad requerido en los accesos a los datos se convierte en una cuestión nuclear.

Además, los requisitos de acceso deben mantenerse y cumplirse a lo largo de todo el proceso Big Data, llegándose incluso a situaciones en las que dichos requisitos asociados a los datos de entrada deban trasladarse al producto resultado de la transformación analítica requiriéndose, por lo tanto, de capacidades de sincronización para la compartición de información relativa a roles y privilegios de acceso. Así las cosas, sería necesaria la implementación de unos protocolos que se proporcionen capacidades de gestión a nivel global en materia de requisitos de acceso.

Controles, soluciones o compromisos existentes en la práctica

A nivel de productos disponibles en el mercado, existen herramientas como *LDAP*, *Directorio Activo*, *OAuth* u *OpenID* que han empezado a demostrar madurez en relación a la gestión de identidades y privilegios de acceso desde repositorios confiables. Por otro lado, existen BBDD NoSQL como Apache Accumulo que proporciona granularidad de acceso capaz de diferenciar a nivel de dupla clave / valor.

Para poder cumplir con los requisitos de grano fino en materia de auditoría, es preciso que los dispositivos participantes en la cadena de valor tengan activadas las capacidades de registro en los distintos tipos de log asociados a routers, aplicativos y sistemas operativos. Correctamente gestionada esta información, se dispondría de trazas completas, modificadas bajo control, disponibles y accedidas solo por entidades autorizadas. A tal fin, es posible el uso de herramientas SIEM de análisis en tiempo real, recomendándose su uso separado de la propia implementación del Big Data cuando sea posible.

4. Operaciones

4.1. Continuidad de Negocio

Al igual que cualquier otro Sistema de Información, una aplicación Big Data proporciona un servicio demandado por una serie de usuarios y, por lo tanto, deben formularse las condiciones que aseguren su continuidad.

Amenazas, problemas o contradicciones

Como todo Sistema de Información, cualquier Big Data Application (BDA) requiere estar soportado por recursos (computacionales, de comunicación, etc.) suficientes y debe disponerse de garantías adecuadas de que los recursos necesarios estarán disponibles incluso en caso de incidencias.

Controles, soluciones o compromisos existentes en la práctica

Las necesidades computacionales de los BDA y su variabilidad en el tiempo se ven satisfechas de forma más eficiente con el apoyo de tecnologías Cloud Computing, que puedan dotar de recursos a la aplicación según sean necesarios en cada momento. El soporte en Cloud Computing condiciona las capacidades y disponibilidad de los BDAs que se emplean por las organizaciones, tanto en sentido positivo como negativo, derivadas ambas facetas de las propias características inherentes al paradigma computacional de Cloud Computing.[xxv]

En particular, en materia de Continuidad de Negocio, el uso de Cloud facilita el aseguramiento de redundancia geográfica y de ubicaciones desde los que dotar de recursos al BDA.

En condiciones nominales de prestación del servicio por el Cloud Service Provider (CSP), el BDA tiene los recursos que necesita, aunque puede verse penalizado en su rendimiento si la distribución geográfica de los recursos deriva en latencias excesivas o throughputs limitados, que no proporcionen los datos necesarios en cantidad y volumen necesario.

Una BDA debe entender sus requisitos de Continuidad de Negocio para todos sus recursos. En particular, debe interpretar sus necesidades de actualización de datos y plantear los requisitos de Continuidad de Negocio sobre las mismas. Más aún, el BDA debe ser consciente de que, en ocasiones, serán sus fuentes de datos las que no estén suficientemente disponibles (bien ellos mismos, bien sus canales de comunicación), derivando en problemas de calidad de datos, por usar sets de información desactualizados, insuficientes o directamente no existentes. El BDA debe estar diseñado ante esta eventualidad que, si bien no constituye un problema de Continuidad de Negocio del aplicativo per se, reduce la utilidad del BDA para sus usuarios e incluso puede invalidarlo, por ofrecer información incompleta / incorrecta de forma

indistinguible de cuando dispone de toda la información necesaria. Y debe imponer requisitos de Continuidad de Negocio para estas fuentes de datos, cuando sea posible.

Adicionalmente, el BDA debe incluir en su diseño características que permitan su propia distribución geográfica, en particular ante situaciones en las que el volumen de recursos dificulte/desaconseje la provisión de los mismos desde un único punto.

Por último, a medida que las plataformas Big Data sean utilizadas con más intensidad en los procesos de toma de decisiones, los requisitos de disponibilidad serán crecientemente exigentes. No solo en recursos necesarios, sino en el Recovery Time Objective (RTO) de la aplicación. En particular, la aplicación de tecnologías Big Data a sistemas en Tiempo Real requiere garantías de operación en Tiempo Real del BDA, de sus algoritmos, de su adquisición de información, etc.

4.2. Interoperabilidad

Habitualmente, un Big Data Framework Provider (BDFP) va a requerir de capacidades de procesamiento flexibles, para integrar fuentes de información dispersas y diversas. Estas capacidades requieren de capacidades de interoperabilidad del BDFP con otros BDFP, con sus fuentes de información, y con los distintos proveedores de recursos disponibles.

Los problemas de interoperabilidad de un BDFP se pueden manifestar en todos los elementos con los que interactúa en su operación:

- Respecto de los recursos, el consumo variable (y en ocasiones intenso) de recursos de computación, almacenamiento, memoria, ancho de banda u otros sólo puede ser satisfecho si el BDFP se apoya en recursos provisionados desde proveedores en la Nube. La (ausencia de) interoperabilidad entre los diversos CSPs (Cloud Service Providers) es un aspecto que aún está pendiente de solución con carácter general. Es posible encontrar CSPs que son interoperables con otros, pero aún no se puede garantizar esta interoperabilidad con carácter general.
- La interoperabilidad entre el BDFP y sus fuentes de información debe verificarse tanto en los formatos empleados como en canales de comunicación, y tanto a nivel semántico como nivel sintáctico.

Controles, soluciones o compromisos existentes en la práctica

A la espera de la finalización de sus tareas por parte de Comités que formalizasen estándares que garanticen la interoperabilidad entre BDFP (como, en el caso de ISO, JTC1/SC32, JTC1/SC38 o JTC1/SC7; o como la iniciativa OASIS) las aplicaciones deben pues desarrollarse con esta limitación en su diseño, de tal forma que incluyan las funcionalidades de interoperabilidad entre plataformas y/o proveedores necesarias.

No es en todo caso un ejercicio claro de interoperabilidad, en tanto que cada BDFP realiza esta tarea de forma separada, y no estandarizada. Esta falta de estandarización dificulta cualquier intento de interoperabilidad entre BDFP, de forma que no se puede garantizar de forma práctica esta característica entre BDFP salvo que ambos utilicen la misma base tecnológica e integren las fuentes de información de la misma manera.

5. Seguridad en el Desarrollo de Software

Nunca antes había sido de tanta importancia la seguridad en las aplicaciones debido a la progresiva migración de los sistemas a la nube, y a la creciente tendencia de desarrollos de BDAs.

El nuevo paradigma se está acelerando y muchos equipos de desarrollo se encuentran con prisas, poco preparados e incluso reticentes al cambio. No cambiar no es una opción. Las aplicaciones deben diseñarse y desarrollarse para identificar y mitigar las nuevas ciberamenazas. El software que no se desarrolle de forma segura por defecto es vulnerable y sufrirá incidentes de seguridad. Los equipos de desarrollo de BDAs deben ser muy conscientes que la exposición de los datos es exponencial en entornos de la nube. Deben actualizarse para incluir las posibilidades y ventajas que ofrecen las tecnologías de la nube en el desarrollo seguro continuo, y así reducir el riesgo de las aplicaciones. Incluir la arquitectura en la nube en el ciclo de vida del desarrollo seguro (S-SDLC) mejora la seguridad frente a las nuevas amenazas.

Amenazas, problemas o contradicciones

El propio entorno de la nube introduce cada vez más vulnerabilidades o debilidades nuevas[xxvi] que los equipos de desarrollo deberían resolver con procedimientos incluidos en sus S-SDLC debido a su creciente tendencia.

Muchos de los riesgos asociados al cloud computing están relacionados con los recursos compartidos[xxvii] (multitenancy). Estas arquitecturas introducen debilidades en el diseño como:

- *Las comunicaciones* entre las diferentes máquinas virtuales, entre el servidor y los discos que almacenan datos, elementos virtuales de red, VLANs, cache compartida, etc.
- *Drivers genéricos* que emulan hardware
- *Vulnerabilidades en los "hypervisores"* que permiten la ejecución de código malicioso que podría llegar a tomar el control de los sistemas.
- *Denegación de servicio* que afecta a otros servicios que se ejecutan en el mismo servidor físico.

Además de las amenazas que introduce la propia arquitectura en la nube, hay que tener en cuenta las amenazas relacionadas con la información. Algunos estudios[xxviii] dicen que una vez los datos están en la nube pública ya no se puede proteger la confidencialidad de los datos. La ingente cantidad de datos que empiezan a gestionar las BDAs son objeto de ataques que pueden causar perdida de datos o filtraciones. La información almacenada en la nube puede estar más expuesta que la información almacenada en arquitecturas tradicionales. La prevención de la filtración de datos (DLP) es uno de los puntos más difíciles de gestionar en las BDAs y requieren del mayor esfuerzo.

El rápido despliegue de las tecnologías en la nube fuerza a los equipos de desarrollo a actualizar y aumentar sus conocimientos en base al nuevo paradigma. No solo los desarrolladores deben estar actualizados, los analistas que utilizan las nuevas herramientas deben estar debidamente formados. El mal uso de las tecnologías en la nube puede ocasionar pérdidas enormes[xxix], por ejemplo la publicación de parte de código (con credenciales incrustadas) en un foro técnico.

La irrupción del IoT introducirá nuevas amenazas debido al desarrollo masivo de aplicaciones para estos dispositivos. El desconocimiento o mal uso de los componentes de seguridad de los frameworks (Xively[xxx], ThingWork[xxxi], etc.) de desarrollo pueden ocasionar incidentes de gran escala.

Controles, soluciones o compromisos existentes en la práctica

La seguridad automatizada en la nube es un campo en pleno desarrollo que aún está evolucionando. Podrían pasar algunos años antes de que estos servicios ofrezcan una solución "point-and-shoot" que sea completamente "hands-free"[xxxii]. Mientras nos acercamos al futuro, hay que actualizar los equipos de desarrollo para construir equipos con conocimientos mixtos (DevSecOps)[xxxiii], es decir, equipos de desarrollo de software, sistemas en la nube y seguridad.

Establecer una política de seguridad, así como estándares y guías de desarrollo debería ser uno de los pilares en el desarrollo seguro de software en la nube. También se recomienda realizar auditorías de código para garantizar el cumplimiento, incluyendo auditorías realizadas por hackers éticos (Red Team) en el S-SDLC para el continuo desarrollo seguro del software.

Integrar la arquitectura en el desarrollo implica la creación de un catálogo de servicios que necesitará el software. Estos servicios pueden crearse de forma automática a partir de plantillas definidas con herramientas como CloudFormation de AWS, Puppet, Ansible y Chef. Integrando la orquestación de la infraestructura en el desarrollo seguro de software se mejora notablemente la seguridad de las BDAs.

Es necesaria una formación continua para que los DevSecOps tengan amplios conocimientos de las arquitecturas donde se ejecutará el software, donde se guardará la información, como se transmite, etc. Deben ser capaces de utilizar las herramientas y servicios que ofrecen las plataformas en la nube (monitorización, configuraciones de redes virtuales, cifrado, WAF, protección contra DoS,...). La trazabilidad de los datos adquiere especial importancia en las BDAs, con lo que será una característica que deberá ser incluida en el S-SDLC.

Conocer todas las posibilidades que ofrecen las tecnologías en la nube es fundamental para mejorar la seguridad de las BDAs con controles de acceso multifactor (MFA), monitorización (cloudwatch, nagios, munin, cacti, ...) o alertas SIEM (splunk, loggly, logstash, ...).

Para eso es fundamental contar con expertos a la hora de construir un equipo DevSecOps, con una estructura organizativa plana y un objetivo o proyecto compartido, e integrar los componentes de seguridad que ofrecen las tecnologías en la nube en el SDLC. Hay además que definir una política de seguridad en el desarrollo del software basada en arquitecturas en la nube y una metodología para garantizar su cumplimiento que incluya la formación continua.

Glosario

Acuerdo "Puerto Seguro" (Safe Harbor): Los principios internacionales de privacidad Safe Harbor o principios de puerto seguro son principios que permiten a algunas empresas de Estados Unidos para cumplir con las leyes de privacidad que protegen Unión Europea y los ciudadanos suizos.

Acuerdos de servicios (Service Level Agreements – SLAs): Un acuerdo a nivel de servicio es un contrato entre un proveedor de servicios y sus clientes que documenta los servicios que el proveedor le proporcionará.

Apache Accumulo: sistema de almacenamiento y recuperación de baja latencia para gran cantidad de datos con seguridad a nivel de celda.

Arquitectura agrupada (Cluster architecture): En un sistema informático, un cluster es un grupo de servidores y otros recursos que actúan como un único sistema y permiten una alta disponibilidad y, en algunos casos, balanceo de carga y procesamiento paralelo. Un cluster Hadoop es un tipo especial de clúster computacional diseñado específicamente para almacenar y analizar enormes cantidades de datos no estructurados en un entorno de computación distribuida.

Big Data Framework Provider (BDFP): entidad encargada de proveer recursos para entornos Big Data. Podría tratarse de clústeres locales, Data Centers o CSP.

Business Intelligence - BI: es un término genérico que hace referencia a una variedad de programas informáticos utilizados para analizar los datos en bruto de una organización.

Ciudad Inteligente (Smart City): se refiere a un tipo de desarrollo urbano basado en el uso de tecnologías digitales o tecnologías de la información y la comunicación (TIC) para mejorar la calidad y el rendimiento de los servicios urbanos, al fin de reducir el consumo de recursos, los costes relacionados, y permitir una mejor comunicación e interración con los ciudadanos.

Cloud Service Provider (CSP): compañía que ofrece a otras compañías o particulares alguna tipología de Cloud Computing, típicamente Infraestructura como Servicio (IaaS), Software como Servicio (SaaS) o Plataforma como Servicio (PaaS).

Datos libres (Open Data): la idea de que algunos datos deben ser de libre acceso y uso, sin restricciones de derechos de autor, patentes u otros mecanismos de control.

Hadoop: Hadoop es un marco de programación libre, basada en Java que soporta el procesamiento de grandes conjuntos de datos en un entorno de computación distribuida. Es parte del proyecto Apache patrocinado por la Apache Software Foundation.

Hashing: este término se refiere a la transformación de una cadena de caracteres en un valor de longitud fija por lo general más corto o clave que representa la cadena original.

Información de carácter personal (Personally Identifiable Information - PII): cualquier tipo de dato que puede ser usado para identificar a un individuo en concreto. Termino legal utilizado en la normativa de protección de datos europea y en las leyes sobre privacidad de EE.UU.

Internet de las cosas (Internet of Things): un proyecto de desarrollo de Internet en el que los objetos cotidianos tienen conectividad de red, lo que les permite enviar y recibir datos.

JavaScript Object Notation (JSON): se trata de un formato ligero para el intercambio de datos. JSON es un subconjunto de la notación literal de objetos de JavaScript que no requiere el uso de XML.

JTC 1/SC32: Comité técnico de la Organización Internacional para la Estandarización (ISO) cuya misión es el desarrollo de estándares en el ámbito de la gestión e intercambio de datos.

JTC 1/SC38: Comité técnico de la Organización Internacional para la Estandarización (ISO) que abarca los Servicios Web, las Arquitecturas Orientadas a Servicio (SOA) y el Cloud Computing.

JTC 1/SC7: Comité técnico de la Organización Internacional para la Estandarización (ISO) cuya misión es la estandarización de los procesos, las herramientas y las tecnologías de apoyo a la ingeniería de sistemas y productos software.

Lago de datos (Data Lake): Un lago de datos es un gran repositorio de almacenamiento que posee una gran cantidad de datos en bruto y permite utilizar la información que en su formato original sin que sea necesario pasar por un proceso de conversión e integración.

Machine Learning: El término se refiere a un tipo de inteligencia artificial que proporciona equipos con la capacidad de aprender sin ser programado de forma explícita. El aprendizaje automático se centra en el desarrollo de programas informáticos que pueden aprender autónomamente a partir de la exposición con nuevas fuentes de información.

MapReduce: Modelo de programación y una aplicación asociada para el procesamiento y la generación de grandes conjuntos de datos. Se trata de un entorno de desarrollo (pensado para lenguaje C inicialmente, aunque luego se han implementado versiones para otros lenguajes como Java) que permite trabajar en paralelo con grandes cantidades de datos en sistemas de memoria distribuida (clusters, sistemas Grid y entornos Cloud).

NoSQL: Se trata de una base de datos que proporciona un mecanismo para el almacenamiento y recuperación de datos que se modela en medios distintos de las relaciones tabulares utilizados en las bases de datos relacionales.

OAuth (Open Authorization): protocolo que permite flujos simples de autorización para sitios web o aplicaciones informáticas. Permite a un usuario del sitio A compartir su información en dicho sitio A (proveedor de servicio) con otro sitio B (consumidor) sin compartir toda su identidad.

OpenID: estándar de identificación digital descentralizado, con el que un usuario puede identificarse en un sitio WEB a través de una URL (actualmente para identidades digitales de dominio cruzado) y puede ser verificado por cualquier servidor que soporte el protocolo.

Procesamiento por lotes (Batch Processing): los archivos por lotes, que se denominan también programas de proceso por lotes o secuencias de comandos, permiten la ejecución de un programa sin el control o supervisión directa del usuario. Este tipo de ejecución da la posibilidad de automatizar una tarea rutinaria ejecutando una serie de órdenes definidas con antelación; por esta razón se utiliza en tareas repetitivas sobre grandes conjuntos de información.

Recovery Time Objetive (RTO): tiempo máximo tolerable que un servicio puede no estar activo después de una interrupción del servicio o de un desastre.

Security information and event management – SIEM: SIEM es un término que se refiere a los productos y servicios que combinan la gestión de información de seguridad (SIM) y la gestión de eventos de seguridad (SEM). Tecnología SIEM ofrece análisis en tiempo real de alertas de seguridad generados por el hardware y las aplicaciones de la red.

Usuario raíz (Root): Un usuario raíz es el nombre de usuario o de cuenta que por defecto tiene acceso a todos los comandos y archivos en un Linux u otro sistema operativo tipo Unix.

Notas

[i] Information Commissioner's Officer (ICO), 2014, "Big data and data protection", 20140728, URL: https://ico.org.uk/media/for-organisations/documents/1541/big-data-and-data-protection.pdf

Information Commissioner's Officer (ICO), "Summary of feedback on Big data and data protection and ICO response", URL: https://ico.org.uk/media/for-organisations/documents/1043723/summary-of-feedback-on-big-data-and-data-protection-and-ico-response.pdf

[ii] The White House, Executive Office of the President, May 2014, "Big Data: Seizing Opportunities, Preserving Values", URL: http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf

[iii] The White House, Executive Office of the President, May 2014, "Big Data and Privacy: A Technological Perspective", URL: https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf

[iv] Taylor Armerding, "The 5 worst Big Data privacy risks (and how to guard against them)", CSO, Diciembre 8 2014, URL: http://www.csoonline.com/article/2855641/big-data-security/the-5-worst-big-data-privacy-risks-and-how-to-guard-against-them.html

[v] Cheatham, M. 2015, "Privacy in the age of Big Data", Paper read at Collaboration Technologies and Systems (CTS), 2015 International Conference on, 1-5 June 2015.

[vi] Ratner, Bruce (editor), 2011, "Statistical and Machine-Learning Data Mining: Techniques for Better Predictive Modeling and Analysis of Big Data", CRC Press, Taylor & Francis.

[vii] Shlomo, Natalie, 2008, "Releasing Microdata: Disclosure Risk Estimation, Data Masking and Assessing Utility", American Statistical Association, Online Proceedings of the Survey Research Methods Section. URL: https://www.amstat.org/sections/srms/proceedings/y2008/Files/300242.pdf

[viii] ePrivacy Seal, URL: https://www.eprivacy.eu/en/consulting/big-data/

[ix] European Union Agency for Network and Information Security (ENISA), "Privacy and Data Protection by Design – from policy to engineering", December 2014, URL: https://www.enisa.europa.eu/activities/identity-and-trust/library/deliverables/privacy-and-data-protection-by-design/at_download/fullReport

[x] Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal, URL: https://www.boe.es/buscar/act.php?id=BOE-A-1999-23750

[xi] Directiva 95/46/CE, URL: http://eur-lex.europa.eu/legal-content/ES/TXT/?uri=uriserv:l14012

[xii] DIRECTIVA 2009/136/CE, URL: http://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32009L0136&rid=3

[xiii] Cate, Fred H., Peter Cullen, and Viktor Mayer-Schönberger, "Data Protection Principles for the 21st Century: Revising the 1980 OECD Guidelines", 2013, URL: http://www.oii.ox.ac.uk/publications/Data_Protection_Principles_for_the_21st_Century.pdf

[xiv] Article 29 Data Protection Working Party, "Statement on Statement of the WP29 on the impact of the development of Big Data on the protection of individuals with regard to the processing of their personal data in the EU Adopted on 16 September 2014", URL: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp221_en.pdf

[xv] Cavoukian, Ann, David Stewart, y Beth Dewitt, 2014, "Using Privacy by Design to Achieve Big Data Innovation Without Compromising Privacy", PbD & Deloitte, URL: https://www.privacybydesign.ca/content/uploads/2014/06/pbd-big-data-innovation_Deloitte.pdf

[xvi] Sara Degli-Esposti, 2015, "Big Data Protection Study Report", URL: https://www.academia.edu/12511348/Big_Data_Protection_Study_Report

[xvii] **CNNMoney (London), October 6, 2015, "Europe cracks down on U.S. tech with data ruling", URL: http://money.cnn.com/2015/10/06/technology/facebook-privacy-european-union/**

[xviii] Acquisti, Alessandro, y Ralph Gross, 2009, "Predicting Social Security numbers from public data", *Proceedings of the National Academy of Sciences of the United States of America,* 106 (27): pp. 10975-10980, doi: 10.1073/pnas.0904891106.

[xix] Cavoukian, Ann, and Daniel Castro, "*Big Data and Innovation, Setting the Record Straight: De-identification Does Work",* 2014, URL: https://www.privacybydesign.ca/content/uploads/2014/06/pbd-de-identification_ITIF1.pdf

[xx] ISO/IEC JTC 1, "Big Data Preliminary Report 2014", URL: http://www.iso.org/iso/big_data_report-jtc1.pdf

[xxi] CSA, "Expanded Top Ten Big Data Security and Privacy Challenges", 2013, URL: https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Expanded_Top_Ten_Big_Data_Security_and_Privacy_Challenges.pdf

[xxii] Melissa Gymrek, Amy L. McGuire, David Golan, Eran Halperin, and Yaniv Erlich, "Identifying Personal Genomes by Surname Inference", Science 18, January 2013, 339(6117): 321-324.

[xxiii] José Antonio Carrillo Ruiz, Jesús E. Marco De Lucas, Juan Carlos Dueñas López, Fernando Cases Vega, José Cristino Fernández, Guillermo González Muñoz de Morales, Luis Fernando Pereda Laredo, "Big Data en los entornos de Defensa y seguridad, Instituto Español de Estudios Estratégicos (IEEE), URL: http://www.ieee.es/Galerias/fichero/docs_investig/DIEEEINV03-2013_Big_Data_Entornos_DefensaSeguridad_CarrilloRuiz.pdf

[xxiv] Nir Kshetri, "Big data's impact on privacy, security and consumer welfare", Telecommunications Policy, 2014, 38(11): 1134-1145, ISSN 0308-5961, URL: http://dx.doi.org/10.1016/j.telpol.2014.10.002

[xxv] Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, Samee Ullah Khan, "The rise of 'big data' on cloud computing: Review and open research issues", Information Systems, 2015, 47(January): 98-115, ISSN 0306-4379, URL: http://dx.doi.org/10.1016/j.is.2014.07.006

[xxvi] Tim Rains - Chief Security Advisor, Microsoft Worldwide Cybersecurity & Data Protection, "*Operational Security versus Secure Development Practices for the Cloud*", URL: https://blogs.microsoft.com/cybertrust/2012/05/01/cloud-fundamentals-video-series-operational-security-versus-secure-development-practices-for-the-cloud/

[xxvii] Jeff Orloff, "Avoid vulnerabilities and threats in the cloud", URL: http://www.ibm.com/developerworks/cloud/library/cl-cloudthreats/cl-cloudthreats-pdf.pdf

[xxviii] KPMG, Data Loss Barometer, 2012, URL: http://www.kpmg.com/uk/en/issuesandinsights/articlespublications/pages/data-loss-barometer.aspx

[xxix] Steve Morgan, "Is poor software development the biggest cyber threat?", URL: http://www.csoonline.com/article/2978858/application-security/is-poor-software-development-the-biggest-cyber-threat.html

[xxx] Xively, URL: https://xively.com/

[xxxi] ThingWork, URL: http://www.thingworx.com/IoT-Platform

[xxxii] Eitan Worcel, "Three Must-Have Capabilities for Effective Cloud-Based Application Security Testing", URL: https://securityintelligence.com/three-must-have-capabilities-for-effective-cloud-based-application-security-testing/

[xxxiii] Justin Foster, "Cloud Security: Secure from Development to Deployment", URL: http://blog.trendmicro.com/cloud-security-secure-from-development-to-deployment/

CSA ES cloud security SPAIN alliance SM

isms forum spain

Paseo de la Habana, 54,
2º Izquierda 1.
28036 Madrid - España
Tlf :+34 91 563 50 62

+info:
info@ismsforum.es
www.ismsforum.es
@ISMSForumSpain