RETOS DE SEGURIDAD INFORMÁTICA EN AMBIENTES DE COMPUTACIÓN BIG DATA

COMPUTER SECURITY CHALLENGES IN COMPUTER ENVIRONMENTS BIG DATA

Darwin Sanabria Sepúlveda [1]
darwin_elcolombianito@hotmail.com

Sebastián Castro Santana [2]
sebastian.castro00@usc.edu.co

Javier Salvador Rojas, M. Sc [3]
jarojas@usc.edu.co

Universidad Santiago de Cali, Facultad de Ingeniería, Programa de [nombre del programa] (1)
Universidad Santiago de Cali, Facultad de Ingeniería, Programa de [nombre del programa] (2)
Universidad Santiago de Cali, Facultad de Ingeniería, Programa de [nombre del programa] (3)

Resumen

Las tecnologías de información emergentes y disruptivas como el Cloud han establecido el medio o puente entre nuevas tecnologías, IoT que garantiza la conexión de múltiples dispositivos móviles y estáticos a la red, Big Data que facilita el almacenamiento y manejo de grandes volúmenes de información, han llevado a las organizaciones a implantar nuevos ecosistemas tecnológicos que están impactando la sociedad en todos sus espacios y a ellas mismas. Este nuevo ecosistema está impulsando el posicionamiento y logro de sus metas y objetivos. Toda esta avalancha de nuevas tecnologías en la nube está generando también nuevas amenazas para la seguridad de la información comprometiendo su confidencialidad, integridad y disponibilidad. Este artículo tiene como objetivo describir y pretende detallar las amenazas, vulnerabilidades y riesgos asociadas con Big Data, indicando además los procedimientos de control que se han desarrollado para minimizar sus efectos.

Palabras Clave: Big Data, seguridad, amenazas, riesgos, tecnologías disruptivas.

Abstract

Emerging and disruptive information technologies such as the Cloud have established the means or bridge between new technologies, IoT that guarantees the connection of multiple mobile and static devices to the network, Big Data that facilitates the storage and management of large volumes of information, have led organizations to implement new technological ecosystems that are impacting society in all its spaces and themselves. This new ecosystem is driving the positioning and achievement of its goals and objectives. All this avalanche of new cloud technologies is also generating new threats to information security, compromising its confidentiality, integrity and availability. This article aims to describe and aims to detail the threats, vulnerabilities and risks associated with Big Data, also indicating the control procedures that have been developed to minimize their effects

Keywords: Big Data, security, threats, risks, disruptive technologies.

1. INTRODUCCIÓN

En los inicios del tercer milenio, las compañías tienen acceso a la información de una forma sin precedentes, el constante avance de las tecnologías ha permitido un crecimiento "explosivo" de la cantidad de datos generados desde diferentes fuentes como las redes sociales, dispositivos móviles, sensores, máquinas de rayos x, telescopios, sondas espaciales, log de aplicativos, sistemas de predicción del clima, sistemas de geo-posicionamiento y en general, en todo lo que se puede clasificar dentro de las definiciones del Internet de las Cosas (Tabares, L. F. & Hernández, J. F. 2015). De la mano de este considerable aumento de información, ha surgido la necesidad de extraer de ella, de manera eficiente, patrones, tendencias y/o conocimiento que permitan apoyar la toma de decisiones (Tabares, L. F. & Hernández, J. F. 2015).

Big data es un término emergente que se refiere al proceso de administrar una gran cantidad de datos provenientes de diferentes fuentes; de bases de datos relacionales (DBMS), de archivos de registro, de publicaciones de redes sociales (Kumar, 2015), datos tipo texto, de imágenes y en general todo lo que produzca información. Al Big Data se le puede denominar compuesto por un conjunto de datos estructurados, semiestructurado y no estructurado y muchos autores lo describen por los atributos de velocidad, volumen, variedad, valor y complejidad.

La sociedad genera constantemente datos y la mayoría son almacenados digitalmente; son información geográfica, estadística, datos meteorológicos, datos de la investigación, datos de transporte, datos de consumo de energía, datos de

salud, redes sociales, banca on-line, entre otros. Toda esta avalancha de datos, se le denomina grandes volúmenes de datos o Big-data (Meneses Rocha y María Elena. 2018) y es información que está disponible en tiempo real, lo que significa que las organizaciones pueden acceder a ella tan pronto como se genera. (Salvador, 2013) citado por (Puyol, 2014).

En términos generales el Big Data puede ser considerado como una tendencia en el avance de la tecnología que ha abierto las puertas hacia un nuevo enfoque de entendimiento y toma de decisiones, la cual es utilizada para describir enormes cantidades de datos que tomaría demasiado tiempo y sería muy costoso cargarlos a un base de datos relacional para su análisis. Se puede decir entonces que el concepto de Big Data aplica para toda aquella información que no puede ser procesada o analizada utilizando procesos o herramientas tradicionales, teniendo en cuenta que Big Data no se refiere a alguna cantidad en específico (Puyol, 2014).

En términos generales Big Data puede ser considerado como la tendencia en el avance de la tecnología que ha abierto las puertas hacia un nuevo enfoque de entendimiento y toma de decisiones, la cual es utilizada para describir enormes cantidades de datos que tomaría demasiado tiempo hacerlo con herramientas tradicionales y por otra parte sería muy costoso cargarlos a un base de datos relacional para su análisis (Puyol, 2014).

El "Big Data y su análisis están en el centro de la ciencia moderna y los negocios" como lo describe con amplitud en su trabajo Sagiroglu & Sinanc (2013), refiriéndose a que estos datos se generan en línea, a través de los correos electrónicos, el video, el audios, las imágenes, el registro de datos, las publicaciones de diferente propósito, consultas de búsqueda en navegadores, registros de salud, la interacción con las redes sociales, datos científicos, sensores, teléfonos móviles y sus Aplicaciones. Se almacenan en bases de datos, crecen masivamente y se vuelven difíciles de capturar, formar, almacenar, administrar, compartir, analizar y visualizar a través de herramientas típicas de software de base de datos.

El estudio de Sagiroglu & Sinanc, (2013), resalta los siguientes hechos:

- Cinco (5) exabytes (10^18 bytes) de datos los crearon los humanos hasta el 2003 y hoy en día esta cantidad de información se crea en solo dos días.
- En el 2012, el mundo digital de los datos se amplió a 2,72 zettabytes (10^21 bytes) y se preveía que se duplicaría cada dos años y que para el 2015 se tendrían alrededor de 8 zettabytes de datos, cifra que fue superada para esa fecha.
- En el 2012 un computador personal promedio almacenaba aproximadamente 500 gigabytes (10^9 Bytes), por lo que se requeriría, alrededor de 20 millones de PCs para almacenar todos los datos del mundo.
- Los datos multimedia tienen un peso alto y ocupan mucho tráfico de backbone de Internet y se esperaba que aumentaran para el 2013 en un 70%.
- Para el año 2020, se espera que cerca 50 mil millones de dispositivos estén conectado a redes e Internet.
- En el 2012, se ejecutó un proyecto global de Big Data, que tuvo como propósito recopilar, visualizar y analizar grandes cantidades de datos en tiempo real; de este estudio se pudo concluir que muchas estadísticas eran producidas por Facebook que en ese momento tenía 955 millones de cuentas mensuales activas usando 70 idiomas, 140 mil millones de fotos, 125 mil millones conexiones de amigos, cada minuto se cargaban 48 horas de video y cada día se realizaban 4 mil millones de reproducciones a través de YouTube.
- Google soportaba muchos servicios monitoreando 7.2 mil millones de páginas por día, procesando diariamente 20 petabytes (10^15 bytes) de datos traducidos a 66 idiomas, también se encontró que se emitían 1 billón de tweets cada 72 horas por parte de más de 140 millones de usuarios activos en Twitter y que cada minuto se creaban 571 nuevo Sitios web. Predice el estudio mencionado que para la siguiente década la cantidad de información aumentará 50 veces.

Todo lo anterior fortalece el argumento que las grandes bases de datos crecen masivamente y se vuelven difíciles de capturar, formar, almacenar, administrar, compartir, analizar y visualizar a través de herramientas típicas de software de base de datos (Puyol, 2014) y que, por otro lado, los datos son actualmente uno de los activos más importantes para las empresas en todos los campos (Moreno, Serrano & Fernández, 2016).

El crecimiento continuo, la importancia y el volumen de los datos crean un nuevo problema que no puede ser manejado por las técnicas de análisis tradicional. Este problema, se resolvió mediante la creación de un nuevo paradigma denominado Big Data. Sin embargo, el Big Data originó nuevos problemas relacionados no sólo con el volumen o la variedad de los datos, sino también con la calidad de los datos, la seguridad de los datos y la privacidad de estos.

Según el Big Data Working Group de la organización Cloud Security Alliance, hay principalmente cuatro aspectos diferentes de la seguridad de Big Data: Seguridad de las infraestructuras, privacidad de los datos, gestión de datos e integridad y seguridad reactiva.

El Big Data como herramienta, está sujeta a dudas y preocupaciones sobre los posibles usos que a la información se le dé, o bien porque se cometan ilícitos al recopilarla sin respaldo legal para ello, o bien porque genere abusos en su uso, basándose en el valor económico de los datos personales, al calificarlos con símiles tan llamativos al llamarlo el petróleo del Siglo XXI (Oliveros y López, 2017).

La generación de perfiles de consumidores es sin duda uno de los usos principales del Big Data, y puede entrañar riesgos por posibles tratamientos basados en predicciones y si se utilizan de forma discriminatoria, excluyendo a sectores minoritarios con base a los resultados analizados, se puede caer además de la privacidad y la propensión, en un tercer peligro. Nos arriesgamos a ser víctimas de una dictadura de los datos (Meyer-Schönberger y Cukier, 2015), por la que fetichizaremos la información, el fruto de nuestros análisis, y acabaremos usándola mal.

De otra parte, surgen otros temores sobre su potencial uso en sectores de poblacional vulnerables como pueden ser menores, ancianos o colectivos marginados, por lo que es necesario establecer garantías adecuadas en todos los ámbitos (Aced, Heras & Saiz, 2016).

Para (Puyol, 2014) retomando por Girardotti (2014) y como se expuso en párrafos anteriores, el Big Data tiene un gran espectro de posibles aplicaciones como son: sensores inteligentes; videos de vigilancia; pagos con tarjeta de crédito; miles de tweets diarios; comentarios redes sociales; archivos de documentos de distinto tipo; transacciones de bolsa; cotizaciones de commodities; carga de vehículos; seguimiento por GPS; información del clima; de temperatura; presión; humedad; vientos; precipitaciones.

Toda esa gran cantidad de usos del Big Data son los que en este articulo nos lleva a investigar: ¿cómo identificar las amenazas de seguridad en ambientes de computación Big Data?, ¿qué agentes amenazan la seguridad?, ¿a qué vulnerabilidades y riesgos está expuesto?, así como poder determinar ¿qué controles se deben implementar cuando se usa un ambiente de computación Big Data?

Las anteriores preguntas son por tanto los aspectos a los que se le quiere dar una respuesta mediante esta investigación y responder al objetivo propuesto de determinar qué elementos de seguridad se deben tener en cuenta cuando estamos frente a una organización que decide tener su infraestructura en un ambiente de computación usando técnicas de Big Data, como centro de su tecnología y del negocio (Moreno, Serrano & Fernández, 2016).

Conocer los elementos de seguridad cuando se está frente a un ambiente computacional de Big Data, es fundamental dado que el Big Data llego y está aquí para quedarse (Miguel Ángel Pantoja, 2015), es prácticamente imposible imaginar una aplicación sin que consuma datos, produzca nuevas formas de datos y que contenga algoritmos basados en datos.

Los elementos comunes y específicos de Big Data surgen del uso de múltiples niveles de infraestructura (tanto de almacenamiento como de computación) para procesar Big Data. El uso de nuevas infraestructuras de computación, como las bases de datos NoSQL (para mejorar el rendimiento de manipulación de grandes volúmenes de datos) que no han sido exhaustivamente investigadas por problemas de seguridad, la no escalabilidad del cifrado para grandes conjuntos de datos, la no escalabilidad de las técnicas de monitoreo en tiempo real que son prácticas para volúmenes de datos pequeños, pero no para Big Data, la heterogeneidad de los dispositivos que producen los datos y la confusión en torno a las diversas restricciones legales y políticas que conducen a enfoques ad hoc para garantizar la seguridad y la privacidad.

El valor del Big Data es indudable en sectores clave como el sanitario, donde existen ya muchos ejemplos de su eficacia para reducir el tiempo de ingreso hospitalario o predecir futuras enfermedades y riesgos clínicos. También se prevé su utilización en las Smart Cities como herramienta para prevenir, por ejemplo, colapsos de tráfico y excesos de contaminación. En el sector de la distribución permite anticiparse al consumidor evitando situaciones de desabastecimiento de productos y falta de suministro.

Por los anteriores usos y otros no planteados, se puede considerar que esta tecnología será muy provechosa para la

sociedad ya que puede aportarle numerosos y valiosos beneficios económicos y sociales, además si tenemos en cuenta que empresas como Microsoft, Google, Apache y otras, han desarrollado productos que permiten a los usuarios almacenar, clasificar, analizar y gestionar grandes conjuntos de datos.

Antes que una organización tome la decisión de implementar un ambiente Big Data, debe realizar un estudio que le ayude a determinar la factibilidad y viabilidad de emprender dicha implementación, además debe tener en cuenta a qué tipo de amenazas de seguridad se va a enfrentar, que agentes le van a amenazará la seguridad, y a que vulnerabilidades y riesgos estará expuesto y por último, debe tener certeza, acerca de los controles que debe implementar cuando decida usar un ambiente de computación Big Data.

Todo lo anterior anima para que este estudio sea una contribución para quienes deciden implementar un ambiente computacional de Big Data en su organización.

Todo lo anterior fortalece el argumento sobre las amenazas potenciales sobre la seguridad de la información con el advenimiento de Big Data dentro de los ecosistemas tecnológicos, es por eso que este artículo revisa información sobre las amenazas , vulnerabilidades, riesgos y controles a los cuales se ven enfrentadas la organizaciones con la futura implementaciones de Big Data y anima para que el mismo, sea una contribución para las organizaciones que deciden implementar un ambiente computacional de Big Data.

1.1 Alcances de Big Data

El constante avance de las tecnologías ha permitido un crecimiento explosivo en la cantidad de datos generados desde diferentes fuentes, tales como, redes sociales, dispositivos móviles, sensores, máquinas de rayos x, telescopios, sondas espaciales, log de aplicativos, sistemas de predicción del clima, sistemas de geoposicionamiento y, en términos generales, todo lo que se puede clasificar dentro de las definiciones del Internet de las Cosas (Tabares, Hernández, 2014).

De la mano de este considerable aumento de información, ha surgido la necesidad de extraer de ella, de manera eficiente, patrones, tendencias y/o conocimiento que permitan apoyar la toma de decisiones para lo cual, los métodos tradicionales de procesamiento de datos han tenido que evolucionar rápidamente, buscando escalabilidad y rendimiento principalmente, con el fin de suministrar respuestas en tiempo real, al menor costo posible, dando lugar al fenómeno denominado Big Data y, hace referencia principalmente a tres términos conocidos como las 3 Vs: Volumen, Velocidad y Variedad.

Big Data no solo hace referencia a los problemas enmarcados dentro de las 3 Vs, sino que también incluye un amplio espectro de técnicas, tecnologías, métodos y paradigmas no convencionales que apoyan la solución de problemas relacionados con datos de una forma diferente y, generalmente, más adecuada que los métodos tradicionales. Big Data permitió introducir nuevas y mejores formas de procesar la información, con ventajas sobre los enfoques tradicionales, los cuales no responden de forma adecuada sobre las necesidades actuales de las compañías, en términos de velocidad, costos de implementación, escalabilidad, flexibilidad y elasticidad sobre entornos más complejos (Tabares, Hernández, 2014).

Big Data hoy en día es accesible tanto para organizaciones grandes como para pequeñas gracias al desarrollo de la computación en la nube y las tecnologías de software, como Hadoop, que le ha permitido a los desarrolladores aprovechar fácilmente miles de nodos informáticos para realizar computación en paralelo y a su vez comprar poder computacional bajo demanda de proveedores de la nube, desarrollos que han acelerado enormemente la adopción de metodologías de minería de datos junto al Big Data, produciendo como resultado nuevos retos de seguridad a partir del acoplamiento de Big Data con entornos de nube pública caracterizados por composiciones heterogéneas de hardware, sistemas operativos y software (Tabares, Hernández, 2014).

De otra parte, en la medida que Big Data se expande a través de la tecnología de streaming en la nube, los mecanismos tradicionales de seguridad diseñados para asegurar los datos estáticos en pequeña escala sobre redes firewall y semi aisladas se entornan inadecuados (Tabares, Hernández, 2014). Las cuestiones de seguridad y privacidad se ven amplificadas por la velocidad, el volumen y la variedad de los datos, así como con la infraestructura de nube a gran escala, la diversidad de fuentes y formatos de los datos, la naturaleza de la transmisión de datos y la migración entre nubes,

haciendo que los mecanismos tradicionales de seguridad, que se adaptan a la obtención de datos a pequeña escala, son inadecuados en este nuevo paradigma.

En la definición de Big Data suministrada por Gartner en "Big data son aquellos activos de información de gran volumen, variedad y velocidad que demandan formas de procesar la información, innovadoras y efectivas en costo, para mejorar el entendimiento y la toma de decisiones", se presentan dos aspectos de los cuales se desprenden los principales desafíos de Big Data Analytics: Innovación y efectividad en costos, que corresponden a los medios por los cuales se espera se logre el mayor valor esperado a partir de los datos. Estos aspectos inherentes en las definiciones de Big Data, junto con las oportunidades que se vislumbran, desencadenan una serie de retos relacionados con los datos, tales como: Captura, Almacenamiento, Transmisión, Procesamiento, tratamiento, Análisis, Visualización, Seguridad, Escalabilidad, Desempeño y Consistencia (Tabares, Hernández, 2014).

En general, Big Data se define como una colección de grandes tamaños de conjuntos de datos con diferentes tipos por lo que resulta difícil de procesar esos datos mediante el uso de algoritmos y plataformas tradicionales de procesamiento de datos (Manogaran, Thota, Vija y Kumar 2016). Por otra parte, como describe TechAmerica Foundation (2012), Big Data no es una tecnología por sí misma, sino que hay un rango amplio de herramientas requeridas para el aprovechamiento de Big Data que incluye las de captura, acceso, almacenamiento, transmisión, presentación/interpretación y análisis (Beyer and Laney 2012) y que, de acuerdo con la literatura, la analítica y la visualización son los dos grupos clave de esta tecnología y técnicas requeridas para el aprovechamiento de Big Data.

Las técnicas y tecnologías requeridas no solo se deben buscar en el campo de las TIC, algunas de ellas están más relacionadas con estadística, matemáticas aplicadas y economía (Manyika et al. 2011), elementos que se han venido adaptando para enfrentar los retos presentados por el fenómeno de Big Data mientras que otras se han tenido que crear específicamente para generar valor en el contextos de Big Data (Manyika et al. 2011), así mismo es importante entender que las tecnologías emergentes para Big Data no están en condiciones de reemplazar las bases de datos relacionales ni las bodegas de datos sino que pueden ofrecer nuevas oportunidades para la gestión de información y la analítica (Kart et al. 2013).

La infraestructura de los sistemas Big Data requieren que los cálculos se hagan utilizando técnicas de procesamiento distribuido y almacenamiento de datos distribuido (Cloud security Alliance, 2013), el aseguramiento de los datos debido a la distribución de la información debe preservar la privacidad, además que los datos sensibles se deben proteger mediante el uso de la criptografía y el control de acceso granular, por ejemplo. La gestión de grandes volúmenes de datos requiere soluciones escalables y distribuidas para la seguridad de los almacenes de datos y para que las auditorías sean eficientes, así como comprobar la integridad de los datos de transmisión que salen de diversos puntos finales y con ellos poder realizar análisis en tiempo real de los incidentes de seguridad que garanticen la salud de la infraestructura.

Uno de los frameworks de programación distribuida más utilizado es MapReduce (Cloud security Alliance, 2013) que divide un archivo de entrada en varios fragmentos; en la primera fase MapReduce usa un mapeador para cada bloque de datos leído, realiza cierta computación y genera una lista de pares de claves valor, en la siguiente fase usa un reductor que combina los valores pertenecientes a cada clave distinta y emite el resultado. En este proceso se aplican dos medidas de prevención de ataques, proteger a los correlacionadores y asegurar los datos en presencia de un asignador no confiable

1.2 Ecosistema de Seguridad Big Data

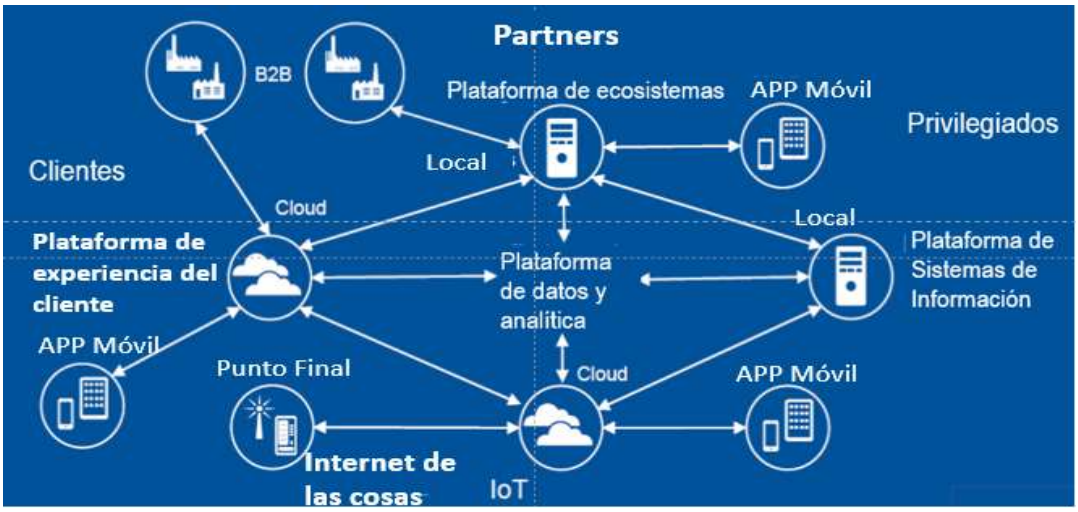
Las nuevas tecnologías permiten a las organizaciones tanto públicas como privadas una mejor administración de su información, para ello están usando el "Big Data". En este era, se están generando más datos, se está analizando más información y se obtienen más resultados de análisis que antes. Para las organizaciones los datos son el centro de su gestión y operación, con lo cual amplían de manera significativa su campo de acción y de competencia, pero todo eso trae consigo nuevos riesgos, que no han sido contemplados en los análisis de seguridad convencionales, como se verá en los siguientes apartados, debido a la nueva convergencia tecnológica y densidad digital, como se aprecia en la figura 1.

Figura 1. Convergencia tecnológica y densidad digital

Fuente: Elaboarción propia.

Visto en términos de Big Data y computación en la nube, esta convergencia tecnológica integra un conjunto de nuevas tecnologías como computación cognitiva, computación en la nube, analítica de datos, redes sociales y computación móvil, impresión 3D y lo más reciente el internet de las cosas, generando un nuevo ecosistema de tecnologías disruptivas como se aprecia en la figura 2.

Figura 2. Interacción de Tecnologías y actores en Big Data



Fuente. Elaboración propia

Como se puede apreciar, el nuevo ecosistema está conformado por las tecnologías que están en crecimiento como el internet de las cosas, las plataformas de experiencia del cliente y sus algoritmos de inteligencia artificial, la computación en la nube, la computación móvil y todo esto integrado a las plataformas tradicionales de sistemas de información. Este ecosistema tiene como actores los partners, proveedores o socios de negocios, los clientes, los administradores o usuarios de los sistemas de información con privilegios y los dispositivos finales del internet de las cosas.

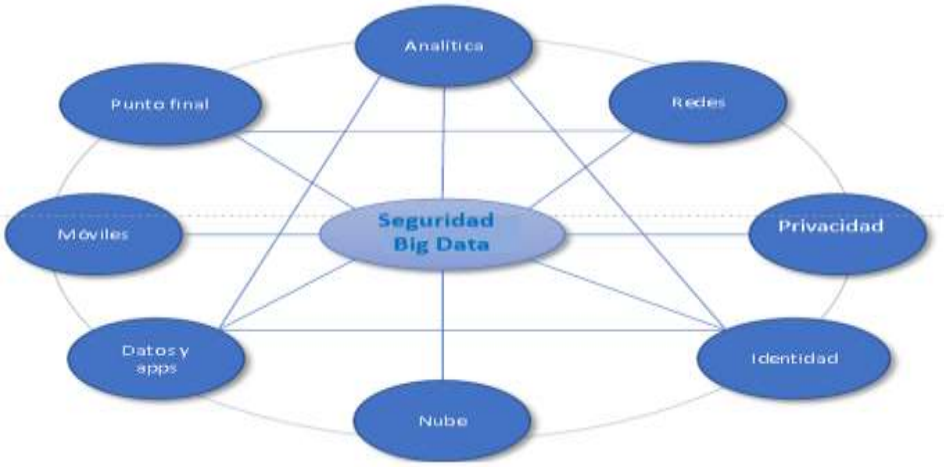
Toda esta convergencia tecnológica está conformada por unos nuevos dominios de seguridad a los que hay que ponerle atención, como se parecía en la figura 3.

Figura 3. Dominios de Seguridad Big data

Fuente. Elaboración propia

Los anteriores dominios impactan directamente al ecosistema de seguridad de Big Data, lo que indica, que se esta frente a una mayor complejidad y heterogeneidad de tecnologías que las convierten en los nuevos enemigos de la seguridad, donde prima la identidad y la privacidad de la información de las personas, como uno de los factores mas relevantes y de mayor impacto, ante lo cual estan cuestionadas compañias como fecebook, google y otras.

Figura 4. Ecosistema de seguridad Big Data



Fuente. Elaboración propia

Este nuevo ecosistema de seguridad esta permeado e implica comprender nuevos factores para tener en cuenta en ambientes Big data:

Para la entrada de datos: sesgos humanos

- Datos parciales, insuficientes, no actualizados o manipulados
- Datos irrelevantes, inconsistentes o incompletos.

En el diseño de algoritmos: fallas técnicas y vulnerabilidades de seguridad

- Sesgos en la lógica, manipulación de tendencias, inclusión de funciones no previstas.
- Errores en la codificación, en el diseño, en la ejecución.

Para la salida: fallas en la implementación

- Usado para lo que no fue diseñado, para desestimar otros análisis, como criterio de autoridad técnica
- Inferencia de interpretaciones incorrectas, conclusiones parciales, conclusiones no asertivas

El ecosistema de seguridad significa entonces, que las organizaciones deben ser conscientes que con las nuevas tecnologías se debe estar listo para enfrentar situaciones de crisis, si no se prepara en términos de seguridad. Esos escenarios se observan en la figura 5.

Figura 5. Posibles escenarios frente a situaciones de crisis

Fuente: Cano (2018). Pronósticos de seguridad de la información 2018.

Para entender el grafico es necesario describir el significado de cada uno de los ejes:

Preparación: es el conocimiento de procedimientos de actuación y habilidad para responder rápida y efectivamente ante eventos inesperados.

Mínimo esperado: Contar con una estructura organizacional y procesos, y buenas prácticas.

Incertidumbre: es el estado de indeterminación entre una causa y sus efectos.

Mínimo esperado: Haber realizado verificaciones y simulaciones.

Comprender que se está ante una nueva forma de atender, implementar e incursionar en las nuevas tecnologías y que estas, están llevando a las organizaciones a tener que implementar también nuevas estrategias, tecnologías y técnicas de seguridad fue lo que hizo que se realizará esta investigación.

El ecosistema de seguridad Big Data conduce entonces a recopilar, estudiar y detallar los aspectos para tener en cuenta respecto a vulnerabilidades, amenazas, riesgos y controles que se deben aplicar en ambientes de computación Big Data, los cuales se estudiarán a continuación.

1.3 Vulnerabilidades del Big Data

Los sistemas de TI se ven afectados por una multitud de vulnerabilidades que podrían ser explotadas por un atacante, explotación que puede propagarse y afectar otros sistemas de la infraestructura tecnológica (Allodi, Massacci, 2017). La estimación cuantitativa del riesgo ocasionado por distintas vulnerabilidades implica la asignación más eficiente de los recursos y la conformación de un entorno estructural más seguro, lo que conlleva a que en la industria se esté adoptando el llamado análisis de riesgo cuantitativo (QRA). A continuación, se describe un conjunto de vulnerabilidades en ambientes Big data, las cuales se pueden clasificar en tres tipos; seguridad, privacidad y ausencia de estándares.

Tabla 1. Vulnerabilidades asociadas a ecosistemas Big-Data

Vulnerabilidad	Descripción	Tipo
Los usuarios de Big Data no son conscientes de la potencial vulnerabilidad de sus datos, por tanto, no son proactivos en la implementación de mecanismos de seguridad.	Investigadores como Moreno et al (2016), afirman que La primera vulnerabilidad en seguridad es la ausencia de mecanismos de autenticación, como la "identificación y la contraseña de usuario, así como la falta de canales seguros para acceder a las bases de datos en la nube".	Seguridad
Aun se utilizan mecanismos de seguridad tradicionales que no aplican para proteger los sistemas distribuidos de almacenamiento de grandes cantidades de datos en la nube	La gran cantidad de datos que se ejecuta en Hadoop está compuesta por cientos o miles de nodos y la información está migrando constantemente entre distintos nodos. Hadoop es una tecnología de computación distribuida, y el alto volumen de datos, hacen que la seguridad en los entornos Big Data sean un gran desafío. Esta dinámica de distribución de la información hace que los mecanismos tradicionales de seguridad como los firewalls o los sistemas de detección de intrusos (IDS), no apliquen para proteger los sistemas distribuidos de almacenamiento de grandes cantidades de datos en la nube.	Seguridad
Dificultades para instalar y configurar mecanismos de seguridad como Kerberos que no es un requisito obligatorio para un clúster Hadoop	Según afirma D. Hu et al (2015), "Hadoop actualmente admite mecanismos de seguridad para mitigar estos riesgos, incluido el uso de Kerberos, la adopción de firewalls y la implementación de permisos básicos de HDFS". En Big Data existen dificultades para instalar y configurar Kerberos, aún más difícil es la integración de Active Directory y el protocolo de acceso a directorios (LDAP).	Seguridad
Inmadurez de este mercado lo cual implica carencia de estándares y documentación sobre buenas prácticas.	Exista una compleja combinación de soluciones comerciales y de código abierto, dando como resultado una escasez de soluciones que cumplan con todos los requerimientos y atiendan el amplio espectro de necesidades del Big Data, lo que dificulta mantener la seguridad de los procesos (Schulte et al. 2013).	Seguridad
No existen leyes explicitas asociadas con lo que es ético o no en términos de privacidad.	Se requiere que las empresas, los gobiernos y las entidades públicas adopten un código de conducta y de un conjunto de leyes más explícitas y completas sobre este tema. En A. Pentland, (2014) de MIT se expresa y dice que debe haber un nuevo acuerdo sobre los datos y que "los datos personales se deben tratar como un bien, donde las personas tendrán derechos garantizados sobre sus propios datos"	privacidad
Las organizaciones no disponen de acuerdos estructurados para la compra de datos a terceros	Muchos de los datos para hacer analítica se deben comprar a terceros mediante diferentes tipos de acuerdos (Manyika et al. 2011). Dichos acuerdos si no son bien estructurados podrían involucrar a la organización en problemas de privacidad de la información.	Privacidad
Ausencia de estándares en la computación en la nube y en Big Data	De acuerdo con A. Lambshead, (2014), "se requiere de estándares que permitan promover, medir y gobernar el uso de la tecnología dentro de la mayoría de los usuarios del Big Data y la computación en la nube". En las tecnologías que deben usar grandes comunidades es esencial la	Estándares

	estandarización de esta ya que aumenta el uso independiente y las evaluaciones comparativas de la misma.	
No hay protocolos de comunicación estándar para construir la nueva generación de dispositivos inteligentes.	Actualmente existe una gran cantidad de protocolos de comunicación en varios dominios de aplicación del Big Data, creando con esto problemas de interoperabilidad y personalización. La ausencia de estándares dificulta la innovación y coloca barreras para en la adopción del Big Data por parte de la comunidad, dado que hay que conocer una cantidad apreciable de diferentes protocolos.	Estándares

Fuente. Elaboración propia

1.4 Riesgos y Big Data

Big Data puede ser percibido como un desafío para los principios clave de privacidad, seguridad y estandarización, esto motiva a todos los actores que utilizan Big Data a hacer esfuerzos para reducir los riesgos asociados con su uso (Velasco, Viollier, 2014).

Existe una estrecha vinculación de las características de Big Data con la privacidad, la seguridad y los efectos sobre el bienestar de los consumidores, lo que ha llamado la atención de investigadores, empresas y responsables políticos. Kshetri, (2014) cita a (Isaca, 2014) refiriéndose a Big Data diciendo, "que la gran cantidad de datos significan que la brecha entre seguridad y violación de la privacidad pueden traer consecuencias muy graves y daños a la reputación, la responsabilidad legal, daños éticos y otras consecuencias conocidas como impacto técnico amplificado".

El riesgo de TI puede definirse como la probabilidad de que el sistema de TI de una organización no responda al cumplimiento efectivo de sus objetivos comerciales (Worrel & Bush, 2007) y para Boshoff (2013), el riesgo de TI se puede dividir en dos categorías: riesgos estratégicos y riesgos operativos.

Bromiley, Rau y McShane (2014) definen los riesgos estratégicos como "los riesgos inherentes a las decisiones estratégicas de la empresa", por lo tanto, los riesgos estratégicos son aquellos que tendrán un impacto en los imperativos comerciales de una organización.

Los riesgos operativos se definen como "el riesgo de pérdidas resultantes de procesos, personas y sistemas internos inadecuados o fallidos" (Böcker & Klüppelberg, 2008). Los riesgos operativos relacionados con la implementación de una nueva tecnología, como el Big Data, son los que impactaran las diferentes etapas del ciclo de vida de los sistemas o tecnología.

A continuación, se describen los más relevantes escenarios de riesgos, en los cuales se centra gran parte de las investigaciones académicas y gubernamentales, los productos y servicios ofrecidos por la industria, los mismos se clasifican como riesgos estratégicos y riesgos operacionales.

Tabla 2. Riesgos asociados a Ecosistemas Big Data

Riesgo	Descripción	Tipo
No hay integración imperfecta de Big data con los sistemas actuales en las organizaciones.	De acuerdo Géczy, (2014), Big Data requiere de una infraestructura de TI sólida y si la infraestructura actual de TI en la organización no es suficiente para proporcionar la capacidad de almacenamiento y la potencia de procesamiento necesaria para Big Data, la empresa se enfrentará al riesgo de integración al implementar esta tecnología.	Estratégico
Demora en las consultas si no se cuanta con el software que permita hacer analítica sobre grandes volúmenes de datos.	La organización debe evaluar y decidir si implementa una herramienta de análisis diseñada específicamente para el uso con Big data o si simplemente utilizará las herramientas tradicionales para hacer la analítica de datos. Las herramientas diseñadas específicamente para Big Data tienen un menor tiempo de respuesta que las herramientas de análisis tradicionales sobre la	Estratégico

	base de datos.	
Falta de comprensión por parte de los altos directivos de cómo se puede utilizar la analítica y el Big Data.	Este es un riesgo que tendrá que abordarse desde la oficina del CEO de TI, para ayudar a educar a los gerentes sobre las ventajas que le puede traer el uso de Big Data para potenciar el negocio.	Estratégico
Incompatibilidad de los diferentes formatos y estructuras de datos usados en Big Data.	El riesgo es mayor cuando se tiene que usar una combinación de datos estructurados y no estructurados. Los datos no estructurados pueden tener diversos formatos y algunos de estos formatos no se ajustan a una base de datos relacional, causando interrupciones y datos mezclados que con frecuencia no están relacionados con los datos primarios de interés (Kaisler et al., 2013), trayendo como consecuencia una baja respuesta en el análisis realizado y la extracción de información útil desde el Big Data.	Estratégico
No tener empleados con las habilidades necesarias para facilitar la implementación del Big Data, así como No tener empleados con las habilidades necesarias para implementar y usar tecnología Big Data para analizar los datos de manera efectiva.	La falta de habilidades al implementar una solución de grandes volúmenes de datos plantea un gran riesgo de interoperabilidad porque sin las habilidades necesarias la organización no podrá maximizar las oportunidades inherentes al uso del Big Data. La empresa tendrá que asegurarse que cuenta o contrata empleados con conocimientos y habilidades para implementar y operar Big Data, con lo cual se garantiza que esta nueva tecnología se utilice en todo su potencial y realmente agregue valor a la organización.	Estratégico
Que no se reconozcan datos como el mismo hecho, cuando este provienen de diferentes fuentes y estándares, dado que Big Data no está estandarizado (Green & Panzer, 2014).	Los diferentes formatos de datos tendrán que fusionarse en una base de datos, creando el riesgo de que los formatos no sean compatibles, especialmente cuando se utilizan datos de fuentes no estructuradas o provenientes de distintas fuentes.	Estratégico
El efecto domino dentro del sistema Big Data cuando se realizan cambios.	La gerencia debe tener en cuenta que la variedad y complejidad de los datos disponibles a través de Big Data son una ventaja para la organización, pero también puede presentar un riesgo de compatibilidad que la organización tendrá que gestionar (Green & Panzer, 2014). La compatibilidad debido a la complejidad de los grandes volúmenes de datos y a la interconexión e interdependencia en las fuentes de datos (Kaisler et al., 2013) es un problema que puede provocar un efecto dominó dentro del sistema de grandes volúmenes de datos si se realizan cambios no controlados.	Estratégico
Violación de la privacidad de las fuentes de datos	No existe suficiente regulación o protocolos reales con respecto a la acumulación de datos no estructurados, especialmente los datos recopilados desde plataformas de redes sociales (Kaisler et al., 2013). Esto puede traer problemas de violación de la privacidad porque se podría estar accediendo a información que inclusive podría no parecer datos personales (MayerSchonberger & Cukier, 2013). Esto puede traer como consecuencia usar datos personales de un individuo sin que este haya dado permiso para el uso de dichos datos.	Estratégico
Violación de la privacidad en datos personales, relacionado con el uso secundario de los datos.	Una persona pudo haber dado su consentimiento para que se usaran sus datos con un propósito especifico, pero el uso de los datos para ese propósito puede producir un informe generado a partir de los datos y en una etapa posterior, esta información o informe podría usarse para otro propósito (secundario). Es claro que la persona no dio autorización para el propósito secundario (Mayer-Schonberger & Cukier, 2013), trayendo como consecuencia una violación a la privacidad.	Estratégico
Sobrecarga o presión en el	Desde el punto de vista del riesgo, la organización debe estar en	Estratégico

desempeño debido a la variedad, complejidad y alto volumen de datos	la capacidad de manejar la escalabilidad en la implementación del Big Data, debe ser capaz de manejar una cantidad creciente de trabajo sin que se afecte la red o el procesamiento al ampliarse el volumen de datos (Géczy, 2014), de esta manera se garantiza que no haya un colapso por la complejidad y el creciente volumen de datos.	
Un riesgo para tener en cuenta es que el sistema se vuelva obsoleto con la introducción de nuevas tecnologías.	La implementación de Big Data hará que los datos no estructurados se agreguen al menos en una base de datos estructurada. Sin la estandarización de los datos no estructurados, existe el riesgo de que la base de datos actual quede inutilizable, ya que es más de lo que la infraestructura de la organización puede manejar. La arquitectura del centro de datos y el modelo organizacional deberá evolucionar para acomodar las aplicaciones de Big Data dentro de la infraestructura de TI de la organización.	Estratégico
No planificar la implementación de una solución Big Data	No planificar la implementación conlleva a que al menos no se identifiquen los riesgos estratégicos descritos anteriormente, dado que los riesgos de implementación generalmente están asociados a los riesgos estratégicos. No planificar significa no identificar riesgos, lo que implicará que los gerentes de negocios no podrán determinar las áreas clave que van a necesitar un enfoque específico para no correr riesgo en la implementación de Big Data.	Operativo
Realizar un mal diseño o definición de componentes de Big Data.	Las necesidades de los usuarios, así como las expectativas de la gerencia, deben estar claramente definidas y se debe tener la certeza que estas necesidades estén alineadas con la potencialidad y capacidad de la tecnología, además que estas necesidades se satisfagan realmente cuando se diseña el sistema, de otra manera se corre el riego que no se comprendan tanto las necesidades de los usuarios como las de las tecnologías a usar (Kaisler et al., 2013).	Operativo
Permitir que el diseño de un sistema Big Data lo realice empleados que no y tienen las habilidades en estas tecnologías.	Antes de emprender una implementación Big Data, la organización tendrá que evaluar si tiene empleados con las habilidades necesarios para poder mitigar el riesgo de diseño o si deberá considerar contratar a empleados con esas habilidades, de otra manera se corre el riesgo que el sistema quede mal diseñado y se pierdan grandes cantidades de dinero.	Operativo
No seleccionar las herramientas necesarias y adecuadas para la organización.	En un contexto de Big Data, los riesgos relacionados con la construcción, instalación y configuración están estrechamente relacionados con la herramienta de analítica y Big Data que las organizaciones eligen. Sin embargo, existen riesgos generales que son universales, independientemente de la herramienta que la organización elija implementar, como lo son el hecho que los empleados sepan usar esas herramientas. Una alternativa para mitigar este riesgo es capacitar a los empleados.	Operativo
Mal funcionamiento de las herramientas BIG Data Adquiridas o en uso en la organización	La organización debe tener claridad si el mal funcionamiento es inherente a la herramienta o la falta de habilidades de los empleados que operan las herramientas de Big Data. El mal funcionamiento de las herramientas Big Data seleccionadas por la organización son un riesgo que la organización debe contemplar y corregir lo más pronto posible.	Operativo
Capturar y almacenar datos no requeridos por la organización	La organización debe revisar si la pérdida de tiempo y espacio se debe a la falta de habilidades de su personal o a un problema de diseño, el cual debe entrar a solucionar. Esto puede conducir al riego que la organización este capturando y almacenando datos que no puede usar o no necesita, lo que resulta en una pérdida de	Operativo

	espacio y tiempo con implica además altos costos.	
Desbalance entre la velocidad de transmisión y la velocidad de almacenamiento de datos	Si el sistema no es capaz de realizar un procesamiento efectivo y en paralelo de los exabytes de datos disponibles a través de Big Data, podría causar un desbalance entre la velocidad a la que se transmiten los datos y la velocidad a la que el sistema puede almacenar los datos (Green & Panzer, 2014). Esta situación podrá dar como resultado un retraso en el sistema que influirá en la eficiencia de este.	Operativo
Imprecisión y no confiabilidad de los datos como consecuencia de privilegiar la cantidad o la calidad.	La gerencia deberá tomar la decisión para determinar qué datos son relevantes y cuáles no, y la gerencia tendrá que tomar la decisión de determinar cuántos datos son suficientes (Kaisler et al., 2013). Estas decisiones garantizaran que solo se usen datos confiables y precisos y que los datos con valor agregado se usen para la toma de decisiones, de otra manera se incurrirá en el riesgo de contar con muchos datos, pero de poca calidad y viceversa.	Operativo
No contar o no capacitar el personal con habilidades en el mantenimiento del sistema de Big Data implementado.	La falta de habilidades para hacer el mantenimiento es algo que también será un riesgo para el mantenimiento del sistema de Big Data. Para mitigar este riesgo, la organización debe asegurar que su personal tiene la habilidad, lo capacita o puede considerar la posibilidad de externalizar el mantenimiento. Esta última opción es bastante viable dado que el mantenimiento no es un proceso que se realice a diario, pero es algo que debe hacerse periódicamente o cuando surja un problema.	Operativo
Violación de la privacidad en datos extraídos de redes sociales	El Big Data, las redes sociales y los algoritmos de inteligencia artificial, abrieron una nueva discusión sobre la privacidad. Los datos extraídos recientemente de Facebook, por ejemplo, mostraron que las redes sociales y el internet aumentan el riesgo de violación de la privacidad.	Privacidad

Fuente. Elaboración propia

Tanto los riesgos estratégicos como los riesgos operacionales deberán ser gestionados por la organización independientemente de las herramientas de Big Data que deseen implementar.

1.5 Amenazas y Controles aplicables en la seguridad Big data

Como su nombre lo indica, la amenaza puede definirse como una posibilidad de que una red o un sistema sean expuestos o sufrir cualquier tipo de impacto o evento negativos mientras que el ataque es un acto de identificar una vulnerabilidad en un sistema y explotar los recursos que utiliza ese sistema. Un administrador de un sistema siempre está al tanto de las amenazas que le puede sobrevenir, pero en Big Data por estar distribuido, en los nodos los ataques se conocen después de que se haya producido un impacto negativo o se hayan visto comprometidos (Maheswari, Saranya, 2018).

En Big Data también se configuran varios tipos de amenazas que constantemente generan problemas al software o al sistema de red o causan problemas de seguridad informática.

1.5.1 Amenazas o causas de riesgos generales

Las amenazas o ataques generales y sus controles para Big Data se describen en la tabla No. 3:

Tabla 3. Amenazas y controles generales en Big Data

FASES	Descripción de la fase	Amenazas o ataques	Descripción del ataque	Control propuesto
recopilación de datos	Los datos de diferentes fuentes vienen con diferentes formatos:	Phishing Spamming Spoofing	Estos ataques están pirateando datos.	Programa de concientización de la seguridad

	estructurados, semiestructurados, y no estructurados		Proveedor y coleccionista para obtener Un acceso a los datos en la fase de recogida.	
almacenamiento de datos	Los datos recopilados se almacenan y se preparan con las precauciones suficientes para ser utilizados en la siguiente fase	Ataques Basados en Minería de Datos Utilización de algoritmos de inteligencia artificial	Se selecciona un conjunto de datos para extraer conocimiento	Dividir los datos (verticalmente y horizontalmente) y no usar frameworks para almacenar datos centralizados
		Ataques sobre dispositivos de almacenamiento de datos	Robar discos duros o hacer imágenes de ellos	Usar medidas de seguridad física y no usar frameworks para almacenar datos centralizados
		Acceso no autorizado a los datos	Los atacantes acceden a los datos de forma Ilegal	Implementar efectivos Controles de Acceso
Analítica de datos	Se realiza un análisis de procesamiento de datos para generar conocimiento útil. Se usan algoritmos de minería de datos como agrupación, clasificación y reglas de asociación.	Ataques Basados en Minería de Datos Utilización de algoritmos de inteligencia artificial	Se selecciona un conjunto de datos para extraer conocimiento	Dividir los datos (verticalmente y horizontalmente) y usar controles de acceso

Fuente: Elaboración propia

1.5.2 Amenazas y controles específicos para el ecosistema Big Data

Se requiere que la infraestructura o plataforma de Big Data sea inherentemente segura (Miguel Ángel Pantoja, 2015), las amenazas a una infraestructura de Big Data incluyen acceso como administrador no autorizado a aplicaciones o nodos, amenazas de aplicaciones web y escuchas ilegales en los canales de comunicación. La infraestructura Big Data es principalmente un ecosistema de diferentes componentes, donde la seguridad de cada componente y la integración de seguridad de estos componentes se debe considerar como un (Miguel Ángel Pantoja, 2015,). Por ejemplo, si se ejecuta un clúster de Hadoop en una nube pública, se debe considerar:

1. La seguridad de la nube pública, que a su vez es un ecosistema de componentes que consiste en componentes de computación, almacenamiento y red.
2. La seguridad del clúster de Hadoop, la seguridad de los nodos, la interconexión de los nodos y la seguridad de los datos almacenados en un nodo.
3. La seguridad de la aplicación de monitoreo en sí misma, incluidas las reglas de correlación aplicables que deben seguir los principios de codificación segura y las mejores prácticas.
4. La seguridad de las fuentes de entrada como dispositivos y sensores de los que provienen los (Miguel Ángel Pantoja, 2015).

Otro modelo de amenaza al monitoreo gira en torno a los adversarios que intentarán evadir las herramientas de análisis de Big Data que se utilizan para identificarlos, para esto, los atacantes pueden crear ataques de evasión en un esfuerzo por evitar ser detectados o lanzar ataques de envenenamiento de datos para reducir la confiabilidad de los

conjuntos de datos utilizados para entrenar los algoritmos de análisis de Big Data.

Aparte de estas amenazas de seguridad, otras barreras se vuelven importantes, como las regulaciones legales, ya que dependiendo de dónde existan los datos monitoreados, pueden aplicarse leyes de privacidad. Esto puede crear obstáculos porque algunos datos pueden no estar disponibles para la supervisión de la seguridad o pueden estar disponibles solo como datos anonimizados (Miguel Ángel Pantoja, 2015).

En general en el monitoreo de la seguridad de Big Data, debe considerarse que los frameworks de programación distribuidos utilizan computación y almacenamiento paralelos para procesar cantidades masivas de datos como lo hace MapReduce por ejemplo y que los algoritmos correlacionadores no confiables pueden ser alterados para analizar las solicitudes, modificar los scripts o alterar los resultados.

Uno de los problemas más difíciles es, detectar cartografistas que devuelvan resultados incorrectos, lo que a su vez genera salidas agregadas incorrectas. Cuando se manejan grandes conjuntos de datos es casi imposible identificar mapeadores maliciosos que puedan crear daños significativos, especialmente para cálculos científicos y financieros.

Tabla 4. amenazas y controles de monitoreo y seguridad de Big Data

Amenazas	Controles y como se implementan
1. Mal funcionamiento de los nodos de trabajo de cálculo DE HADOOP 2. Ataques de Infraestructura comprometidos. 3. Nodos de Datos malintencionados	Modificar el marco MapReduce, el sistema de archivos distribuido y la máquina virtual Java con SELinux como sistema operativo subyacente

Fuente: Elaboración propia

1.6 Mejores prácticas de seguridad para almacenes de datos NoSQL

La misma flexibilidad arquitectónica que identifica las dos característicos de NoSQL, el rendimiento y la escalabilidad, plantea el mayor riesgo de seguridad, dado que no hay estándares de seguridad y por tanto los proveedores han desarrollado soluciones NoSQL de abajo hacia arriba, abordando los problemas de seguridad sobre una base ad-hoc, lo que ha dado origen a que las amenazas de las bases de datos NoSQL se clasifiquen en seis escenarios (Miguel Ángel Pantoja, 2015).

Sin embargo, hay que contemplar que el aspecto de seguridad de las bases de datos NoSQL está evolucionando, es así como por ejemplo las soluciones de inyección de base de datos NoSQL aún no están maduras, pero empiezan a aplicarse en estas. Cada base de datos NoSQL fue construida para abordar diferentes desafíos planteados por la analítica, pero la seguridad no se abordó durante la fase de diseño. Los desarrolladores que usan bases de datos NoSQL suelen integrar la seguridad en el middleware, ya que las bases de datos NoSQL no proporcionan ningún soporte para aplicar explícitamente la seguridad en la base de datos. Sin embargo, uno de los aspectos donde se plantean los mayores desafíos de seguridad en bases de datos NoSQL es en la agrupación de los datos.

Las empresas que migran a este tipo de bases de datos, para manejar grandes conjuntos de datos no estructurados pueden beneficiarse del uso una base de datos NoSQL, dado que este tipo de bases de datos acomoda y procesa grandes volúmenes de datos estáticos con buenos tiempos de respuesta en analítica predictiva o análisis histórico.

La utilización de técnicas de modelado de amenazas en bases de datos NoSQL, muestran un árbol de amenaza donde solo se ve una capa de seguridad muy delgada comparada con las bases de datos relacionales. En general, la filosofía de seguridad de las bases de datos NoSQL se basa en mecanismos de aplicación externos y para reducir los incidentes de seguridad, la compañía debe revisar las políticas de seguridad para el middleware y al mismo tiempo, aplicar a las bases de datos NoSQL elementos de seguridad que coincida con las aplicadas a las bases de datos relacionales sin comprometer sus funciones operativas.

Tabla 5. Amenazas y Controles en Bases de Datos NOSQL

Amenazas	Controles y como se implementan
1. Integridad transaccional: uno de los inconvenientes más visibles de NoSQL es su débil enfoque para garantizar la integridad transaccional. La introducción de complejas restricciones de integridad en su arquitectura impedirá su principal objetivo; lograr un mejor rendimiento y escalabilidad 2. Mecanismos de autenticación laxos: en general, NoSQL utiliza técnicas de autenticación débiles y mecanismos de almacenamiento de contraseña débiles, esto expone a NoSQL a los ataques de repetición y los ataques de fuerza bruta de contraseña. 3. Mecanismos de Autorización Ineficientes: las técnicas de autorización son diferentes en la mayoría de las bases de datos NoSQL. Cada una aplica la autorización en las capas superiores en lugar de aplicar la autorización en las capas inferiores. En general la autorización se aplica a nivel de base de datos y no a nivel de colección. 4. Se han identificado ataques de inyección que permiten el acceso de puerta trasera al sistema de archivos para actividades maliciosas 5. Falta de consistencia debido a la incapacidad de aplicar simultáneamente los tres elementos del teorema CAP (consistencia, disponibilidad y tolerancia de partición) mientras está en modo distribuido al no garantizar la confiabilidad de los resultados producidos. Como resultado, a los usuarios no se les garantiza resultados consistentes en un momento dado, ya que cada nodo participante puede no estar totalmente sincronizado con el nodo que contiene la última imagen. Los algoritmos de hashing responsables de replicar los datos a través de los nodos del clúster, se distorsionan en el caso de que falle un de los nodos, dando como resultado un desequilibrio de carga entre los nodos del clúster. 6. Los ataques podrían pasar desapercibidos debido a los pobres métodos de registro y análisis de registros, junto con otros mecanismos de seguridad rudimentarios de NoSQL, con los cual los mecanismos de seguridad pueden no ser efectivos para que se logre atacar información privilegiada	Ocultar NoSQL bajo el contenedor seguro de middleware o acceder a NoSQL utilizando un framework como Hadoop puede crear una capa virtual segura alrededor del perímetro NoSQL. La seguridad a nivel de objeto en el nivel de colecciones o de columna se puede inducir a través del middleware, conservando su delgada capa de base de datos. Esta técnica garantiza que no haya acceso directo a los datos y que los datos sólo estén expuestos en función de los controles configurados en el middleware. Como alternativa a los vulnerables datos NoSQL, el cifrado proporciona una mejor protección. Hadoop emplea cifrado de capa de archivo para proporcionar protección inquebrantable, independientemente del sistema operativo, plataforma o tipo de almacenamiento. Existen en el mercado productos capaces de ofrecer cifrado, tanto en la demanda de procesamiento de datos como en el flujo continuo y de procesamiento en memoria. Las soluciones de cifrado parecen ser una la forma adecuada de abordar varios de los problemas conocidos de seguridad de datos.

Fuente. Elaboración propia

1.7 Registro seguro de datos y registro de transacciones

Se requiere de nuevos mecanismos para impedir el acceso no autorizado y mantener la disponibilidad constante a registros de datos y transacciones. Los sistemas de almacenamiento de capas automáticas generan nuevas vulnerabilidades debido a la falta de posesión física, a la disposición de servicios de almacenamiento no confiables o a políticas de seguridad incoherentes (Miguel Ángel Pantoja, 2015).

En Big Data hay que considerar que los registros de datos y de transacciones se almacenan en medios de

almacenamiento de varias capas y que el movimiento manual de datos entre niveles le proporciona al administrador de TI control directo sobre qué datos se mueven exactamente y cuándo. Sin embargo, como el tamaño del conjunto de datos continúa creciendo de manera exponencial, la escalabilidad y la disponibilidad han requerido mayores esfuerzos para la administración del almacenamiento en Big Data. Las soluciones de nivel automático no controlan dónde se almacenan los datos, lo que plantea nuevos retos para asegurar el almacenamiento de los datos.

Como ejemplo, si se considera que una empresa quiere integrar datos de diferentes áreas y que algunos de estos datos rara vez se usan, mientras que otras áreas usan constantemente el mismo conjunto de datos, en un sistema de almacenamiento de capas automático se ahorrará dinero al enviar los datos poco usados a un nivel inferior dado que es más barato. Sin embargo, estos datos pueden contener información crítica a la que no se accede regularmente, como podrían ser por ejemplo resultados de I + D. Dado que el nivel inferior a menudo ofrece una seguridad reducida, la empresa debe estudiar cuidadosamente las estrategias de ubicación por niveles, incluyendo la definición de los metadatos, es decir, el registro de texto, lo cual introduce otra dimensión que necesita ser protegida. Los ataques de envenenamiento de registro llevarán potencialmente a la inconsistencia de datos y disputas entre los usuarios.

El modelo de amenaza para sistemas de almacenamiento de capas automáticas incluye siete escenarios principales:

Tabla 6. Amenazas para sistemas de almacenamiento

Amenazas	Controles y como se implementan
1. Confidencialidad e Integridad: además de aquellos que intentan robar información sensible o dañar los datos de los usuarios, también se debe suponer que los proveedores de servicios de almacenamiento son terceros no confiables. 2. Procedencia: debido al gran tamaño de los datos, es imposible descargar todo el conjunto de datos para verificar su disponibilidad e integridad. 3. Disponibilidad: la clasificación automática también plantea desafíos a los proveedores de servicios para garantizar una disponibilidad constante. 4. Coherencia: en Big Data es típico que los datos fluyan entre niveles y sean compartidos por múltiples usuarios. 5. Ataques de colusión: mientras un propietario de datos almacena el texto cifrado en un sistema de almacenamiento automático y distribuye la clave y el acceso de permisos a los usuarios, cada usuario está autorizado a tener acceso a una cierta porción del conjunto de datos. 6. Ataques Roll-Back: en un entorno multiusuario, el proveedor de servicios puede aplicar procesos de roll-back. Cuando una versión actualizada de un conjunto de datos se ha cargado en el almacenamiento, el proveedor de servicios puede engañar al usuario mediante la entrega de la versión obsoleta. 7. Disputas: la falta de registros conducirá a disputas entre los usuarios y el proveedor de servicios de almacenamiento, o entre los usuarios. Cuando	Se pueden adoptar muchas estructuras para satisfacer los requisitos generales de seguridad, como la confidencialidad de los datos, la integridad y la disponibilidad, entre ellos tenemos: 1. Establecer de una versión dinámica extendida de un esquema de la función PDP "delegado de protección de datos" en la organización. Con esto se logra mayor eficiencia porque sólo se basa en la criptografía de simetría-clave. 2. Subcontratar los procedimientos de verificación a un auditor o tercero, aplicando un protocolo de verificación que sea públicamente verificable. Un esquema de auditoría pública que preserva la privacidad fue propuesto para el almacenamiento en nube en Wang, et al. Basado en un autómata lineal homomórfico integrado con el enmascaramiento aleatorio, el esquema propuesto es capaz de preservar la privacidad de los datos cuando un TPA audita el conjunto de datos almacenado en los servidores en diferentes niveles. 3. Realizar operaciones sobre el texto cifrado sin descifrar. El esquema de cifrado completamente homomórfico hace que estas operaciones sean posibles porque se soportan funciones más complejas. Recientes logros en "Cryptographic Cloud Storage" proporcionan una alternativa para las plataformas en la nube al permitir la construcción de un almacenamiento IaaS seguro encima de una infraestructura no confiable.

ocurre pérdida de datos o manipulación, los registros de transacción y de transmisión son críticos para determinar la responsabilidad.	

Fuente. Elaboración propia

1.8 Validación/filtrado de entrada de punto final

Los sistemas de Big Data requieren la recopilación de datos de desde diversas fuentes, incluyendo dispositivos de punto final, es acá donde la validación de entrada y el filtrado son todo un desafío cuando se presumen que las fuentes de entrada no son confiables, especialmente en los modelos que configuran su propio dispositivo (BYOD) Bring Your Own Device (Miguel Ángel Pantoja, 2015).

Por tanto, el mayor desafío en el proceso de recolección de datos es la validación de entrada, lo cual requiere tener confianza en los datos, validar que una fuente de datos de entrada no es maliciosa, poder filtrar entradas maliciosas de una colección de datos.

En este sentido hay que considerar por ejemplo que los datos recuperados de los sensores meteorológicos y los datos de retroalimentación enviados por una aplicación de iPhone comparten un problema de validación similar. Un atacante puede ser capaz de crear un acto malintencionado, crear sensores virtuales o simular los IDs del iPhone, esto se podría complicar aún más por la cantidad de datos recopilados, que pueden superar millones de lecturas.

Para contrarrestar esta problemática de manera efectiva, se hace necesario crear algoritmos para validar la entrada de grandes conjuntos de datos.

Tabla 7. Amenazas para validación/filtrado de entrada de punto final

Amenazas	Controles y como se implementan
1. Un adversario puede manipular un dispositivo del que se recopilan datos o puede manipular la aplicación de recolección de datos que se ejecuta en el dispositivo, para proporcionar información malintencionada a un sistema central de recopilación de datos. 2. Un adversario puede realizar ataques de clonación de ID. 3. Un escenario más complicado involucra a un adversario que puede manipular las fuentes de entrada de datos 4. Un adversario puede comprometer los datos en la transmisión de una fuente benigna al sistema de recolección central.	No hay un enfoque infalible para la validación de entrada y el filtrado. Ante esta eventualidad se recomienda un enfoque híbrido que podría esta implementado de la siguiente manera: • Los diseñadores de sistemas de recopilación de grandes volúmenes de datos deben tomar máximo cuidado para desarrollar plataformas y aplicaciones seguras de recolección de datos. Se debería considerar especialmente el escenario BYOD en el que la aplicación se ejecutara en dispositivos no confiables. • Los diseñadores deben identificar los potenciales ataques Sybil, que es un tipo de amenaza de seguridad en un sistema en línea donde una persona trata de hacerse cargo de la red creando múltiples cuentas, nodos o computadoras, así como los ataques de suplantación de identidad en el sistema, para poder aplicar algún método eficaz de mitigación de los ataques. • Los diseñadores deben desarrollar algoritmos para detectar y filtrar entradas maliciosas de un atacante

Fuente: Elaboración propia.

1.9 Monitoreo de seguridad en tiempo real

La seguridad en Big Data no sólo está dirigida a la protección de infraestructura de datos, sino también al aprovechamiento de los análisis de Big Data para ayudar a mejorar la seguridad de otros (Miguel Ángel Pantoja, 2015). Uno de los problemas más difíciles en Big Data es el monitoreo de seguridad en tiempo real, que consta de dos componentes; monitorear la infraestructura Big Data y usar la misma infraestructura para el análisis de datos.

El monitoreo de seguridad en tiempo real es un desafío debido al número de alertas generadas por los dispositivos de seguridad, estas alertas pueden estar correlacionadas o no, además conducen a un número masivo de falsos positivos, que

a menudo se ignoran debido a la capacidad limitada de los humanos para analizarlos. Este problema aumenta constantemente dado el volumen y la velocidad de los flujos de datos, sin embargo, las tecnologías Big Data pueden proporcionar una oportunidad para procesar y analizar rápidamente diferentes tipos de datos, como se usan para detectar anomalías en tiempo real basada en análisis de seguridad escalables.

Hoy en día, la mayoría de las industrias y agencias gubernamentales se benefician del análisis de seguridad en tiempo real, los gobiernos están usando la tecnología para responder a conocer quién está accediendo a los datos, qué recursos están usando, en qué momento los usan, conocer si se está bajo un ataque, detectar si hay violación del estándar de cumplimiento X debido a la acción Y.

La aplicación de la analítica no es nueva, pero hacer una buena recopilación y análisis de datos permite mejores y más rápidas decisiones; es decir, menos falsos positivos.

Tabla 8. Monitoreo de la seguridad

Amenazas	Como se Implementa
El monitoreo de seguridad requiere que la infraestructura Big Data, o plataforma, sea intrínsecamente segura. Las amenazas a una infraestructura de Big Data incluyen el acceso como administrador de forma fraudulenta a aplicaciones o nodos, amenazas de aplicaciones web y escuchas en los canales de comunicación.	Implementar sistemas front-end para monitorear las solicitudes de Hadoop como el Proxy Database Activity Monitoring del firewall. Adherirse a las directrices de OWASP para el control de aplicaiones. Utilizar las soluciones y los frameworks para el monitoreo en tiempo real, como el Protocolo de Automatización de Contenido de Seguridad SCAP de NIST, que están entrando lentamente en el ambiente de Big Data.

Fuente. Elaboración propia.

Un análisis reciente de cómo las empresas están aprovechando la analítica de datos con fines de marketing incluye el ejemplo de cómo un minorista fue capaz de identificar el embarazo de una adolescente antes de que su padre se enterara (Miguel Ángel Pantoja, 2015). Del mismo modo, anonimizar los datos para el análisis no es suficiente para mantener la privacidad del usuario, tenemos el caso de AOL que publicó anónimamente registros de búsqueda para fines académicos, pero los usuarios fueron identificados fácilmente por sus búsquedas. Netflix se enfrentó a un problema similar cuando los usuarios anónimos en su conjunto de datos fueron identificados mediante la correlación de las puntuaciones de películas Netflix con las puntuaciones IMDB.

Por lo tanto, es importante establecer pautas y recomendaciones para prevenir divulgaciones inadvertidas de privacidad. Big Data puede potencialmente permitir invasiones de privacidad, marketing invasivo, disminución de las libertades civiles y un mayor control estatal y corporativo

Los datos de los usuarios recopilados por grandes organizaciones son constantemente accedidos por analistas internos, así como por contratistas externos y socios comerciales. Un miembro malintencionado o socio no confiable puede abusar de estos conjuntos de datos y extraer información privada de los clientes.

No es un secreto que las agencias de inteligencia recopilan grandes cantidades de datos y para ello usan numerosas fuentes, entre las cuales se pueden mencionar las salas de chat, los blogs personales y los enrutadores de red. Sin embargo, la mayoría de los datos son inocentes por naturaleza y no deben conservarse, preservando así el anonimato.

1.10 Privacidad

La aparición de Big Data y el almacenamiento en ella de grandes cantidades de información, sobre todo la almacenada por las redes sociales, Google u otras tecnologías traen consigo un nuevo conflicto de privacidad y propiedad. Los usuarios o personas que generan los datos sobre la internet no son conscientes del valor que para otros representa esa información, y mucho menos lo valiosa que es en términos económicos. La información simplemente se cede a las

empresas o entidades públicas que la recolectan sin ninguna retribución y en total despreocupación. Es ahí cuando aparecen situaciones donde la información personal identificable se transforma en metadatos, convirtiéndose en una información mucho más importante que los propios datos. Metadatos que a juicio de expertos se convierten en nuevas formas de control social, que en el peor de los casos controla la privacidad y la distribución libre del conocimiento.

Big Data para las empresas, crea grandes oportunidades de recolección de información personal, pero trae consigo grandes riesgos y obligaciones que deben asumir al momento de tratar, procesar y hacer circular esa información y ser conscientes que directa o indirectamente alguna parte de esa información puede conllevar a riesgos que afecten la privacidad de las personas. La privacidad de los datos personales es lo que ha llevado a las agencias de los gobiernos a emitir políticas sobre protección de datos, ampliando su campo de acción hasta las implicaciones de privacidad del Big Data e introduciendo distintas soluciones para minimizar los riesgos de la recolección masiva de datos personales.

La industria debe ser consciente y cauta de los riesgos que le puede traer el proceso de identificación, análisis y recolección de información indiscriminada, poniendo atención especial a la posible violación de la privacidad de los datos, sabiendo, además, que los datos personales se deben usar para que se apliquen solo como propósito del servicio contratado.

Desde el punto de vista de las personas, los individuos deben también ser conscientes que el Big Data por su naturaleza implica un riesgo para la privacidad y que la generación diaria de una gran cantidad de datos, como podrían ser : qué compramos; que sitios visitamos; con quién nos comunicamos; qué lectura nos gusta; qué restaurante y comidas nos gusta y muchos más, hace a las personas más vulnerables a la exposición de la privacidad, y que incluso puede traer como riesgo de la exposición de la privacidad, la discriminación manifiesta, ya que Big Data gracias a la llamada "automatización", da a las empresas la capacidad de asumir decisiones discriminatorias sin necesidad de dejar una evidencia explícita.

Tomando en cuenta la legislación sobre protección de datos, en la mayoría de los países no hay legislación específica para Big Data, lo único aplicable son las leyes que 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.

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

1.10.1 Amenazas a la privacidad

Por lo general el análisis masivo de datos se representa y da como resultado un informe analítico de agregados en los cuales no es indispensable identificar al cliente o individuo propietario de la información. Para la protección de la información personal en especial la identificación, se han implementado distintos métodos de separación de datos, entre ellos la conversión en datos anónimos, la conversión en seudónimos, el cifrado, el uso de claves de conversión de identidad para separar la información personal de la identidad real, sin embargo, la aparición de la inteligencia artificial y el uso de algoritmos de integración que recolectan información desde distintas bases de datos relacionada con la misma persona, genera un riesgo a la privacidad personal. Todo esto genera la amenaza de que se esté recolectando información por muchas empresas inclusive sin saber con precisión qué harán con ella o pensando que en el futuro la venta de esa información les representa una oportunidad de negocio rentable.

1.10.2 Controles a la privacidad en Big Data

En primer lugar, se recomienda aplicar el principio de minimización de la información, de esta manera la organización podría entender que datos está procesando y a su vez tener claridad sobre el fin de estos. La identificación del conjunto de datos que le organización realmente va a necesitar y que le aportan valor, le va a permitir a la empresa aplicar una o

varias formas de proteger la confidencialidad de la información, entre ellas puede optar por:

- El enmascaramiento de datos: una solución que podría mitigar el riesgo de las discriminaciones manifiestas, pero como se dijo anteriormente, el análisis masivo de datos podría revelar fácilmente la identidad de las personas reales a través de la asociación, combinación y aplicación de algoritmos de inteligencia artificial.
- La técnica de ofuscación: también conocida como enmascaramiento de datos, es el proceso de reemplazar información sensible con información que parece información real de las personas, pero que no le va a servir a nadie que quiera darle un mal uso.
- Introducción de ruido en los datos: consiste en añadir datos sin valor para engañar al receptor de estos, en este caso es importante analizar e identificar los datos que se deben preservar para asegurar un análisis fiable de los datos, así como de la privacidad de los mismo.

2. CONCLUSIONES

Conocer los elementos de seguridad cuando se está frente a un ambiente computacional de Big Data, es fundamental dado que el Big Data llego y está para quedarse, es prácticamente imposible imaginar una aplicación sin que consuma datos, produzca nuevas formas de datos y que contenga algoritmos basados en datos.

En esta investigación se encontró que las organizaciones enfrentan un nuevo ecosistema de seguridad ante el cual deben contemplar una serie de vulnerabilidades, amenazas y riesgos, colocando atención inicial a los riesgos estratégicos, ante los cuales deben aplicar los controles necesarios para contrarrestar las causas y efectos de lo que implica no atender desde el punto de vista comercial cada uno de estos riesgos.

También se pudo evidenciar que los riesgos operativos son una consecuencia de los riesgos estratégicos, que los mismos son de carácter técnico y directamente vinculados a las herramientas específicas de análisis de Big Data que implementa cada organización. Las especificaciones técnicas de las herramientas de análisis de Big Data no hacen parte de este estudio, por lo cual solo se enumeraron los riesgos operativos generales.

Se recomienda priorizar los riesgos de diseño y planificación operativa ya que tienen un fuerte vínculo con los riesgos estratégicos, dado que estos riesgos influyen mutuamente, en consecuencia, los gerentes de negocios tendrán que tomarse el tiempo para asegurarse de que estos riesgos se aborden de manera adecuada para garantizar que se logren los objetivos del negocio.

El impacto de los riesgos enumerados en esta investigación, sobre los imperativos empresariales de Big Data como se identifican en el capítulo 3, indica que estos riesgos tendrán que ser gobernados para asegurar que se logren los objetivos comerciales y que los objetivos comerciales estratégicos y las metas de TI, se alineen al implementar Big Data.

Big Data crea grandes oportunidades de recolección de información personal para las empresas, pero trae consigo grandes riesgos y obligaciones que deben asumir al momento de tratar, procesar y hacer circular esa información, se debe ser consciente que directa o indirectamente alguna parte de esa información puede conllevar a riesgos que afecten la privacidad de las personas y que la privacidad de los datos personales es lo que ha llevado a las agencias de los gobiernos a emitir políticas sobre protección de datos, ampliando su campo de acción hasta las implicaciones de privacidad del Big Data e introduciendo distintas soluciones para minimizar los riesgos de la recolección masiva de datos personales.

Big Data se fundamenta en el hecho de recopilar información en tiempo real, en el almacenamiento y el análisis de grandes cantidades de datos en distintos formatos, pero todo eso trae consigo nuevos riesgos, que no han sido contemplados en los análisis de seguridad convencionales, como se vio en los apartados anteriores, trayendo consigo un nuevo ecosistema de seguridad Big Data, debido a la convergencia tecnológica y densidad digital de nuevas tecnologías.

Big data puede ayudar a las empresas en el proceso de tomar decisiones operativas, tácticas y estratégicas, existe una gran variedad de potencialidades usos del Big Data en diversos sectores como el financiero, el comercio minorista, la salud, el transporte, la agricultura, la energía, la manufactura, el entretenimiento y el sector público, sin embargo, a pesar de las bondades que ofrece el Big Data, también hay nuevas vulnerabilidades relacionadas con la seguridad de los datos, la

privacidad y la estandarización de los procesos.

3. REFERENCIAS

1. Altman, M., Wood, A., O'Brien, D., & Gasser, U. (2018). Practical approaches to big data privacy over time. International Data Privacy Law, 8(1), 29-51. doi: 10.1093/idpl/ipx027.
2. Aced, E., Heras, M. & Saiz, C. (2016). Código de buenas prácticas en protección de datos para proyectos de Big Data.
3. Bustamante, A., Nikoletta, B., Guillén, A & Thais, S. (2017). Un acercamiento al Big Data y su utilización en comunicación. Mediciones Sociales
4. Chingfang Hsu, C. Zeng B. & Zhang M. (2014). Una nueva clave de grupo de transferencia para la seguridad de Big Data. Escuela de Computación, Universidad Normal de China Central, Wuhan 430079, China.b Escuela de Ingeniería de Software, Universidad Tecnológica del Sur de China, Guangzhou 510006, China.
5. Cloud Security Alliance (2013). Ampliando Top Ten Big Data. Desafíos de seguridad y privacidad. Grupo de Trabajo Big Data.
6. D. Hu, D. Chen, Y. Zhang, and S. Pe. (2015). "Research on Hadoop Identity Authentication Based on Improved Kerberos Protocol", International Journal of Security and its Applications, vol. 9, no. 11, pp. 429-438, doi: 10.14257/ijsia.2015.9.11.39.
7. Draft NIST Big Data Interoperability Framework (2014).
8. Fernández, C. B. (2018), Dossier Big Data y Política. Ciber política y Posverdades a Medias.
9. Fernández, G. L. (2014). ¿Cómo puede el gobierno colombiano aprovechar de mejor manera el potencial de Big Data?
10. J. Moreno, M. Serrano, and E. Fernández-Medina. (2016). "Main Issues in Big Data Security", Future Internet, vol. 8, no. 44, pp. 1-16, doi: 10.3390/fi8030044.
11. Kshetri, N. (2014).Big data's impact on privacy, security and consumer welfare Bryan School Business and Economics, The University of North Carolina at Greensboro, Bryan Building, Room:368, P.O. Box 26165 Telecommunications Policy. Volume 38, Issue 11.
12. Lambshea, (2014). "The Importance of Standards in a Time of Innovation", SMPTE Motion Imaging Journal, vol. 123, no. 8, pp. 7-7, doi: 10.5594/j18480.
13. Mankyika, J. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute
14. Manogaran, G., Thota, C. & Kumar, V. (2016). Desafíos de seguridad asociados con Big Data en Cloud Computing. Conferencia Internacional sobre tendencias recientes en informática e ingeniería.
15. Manogaran, G., Thota, C. & Kumar, V. (2016). MetaCloud Data Storage Architecture for Big Data Security in Cloud Computing.
16. Meneses Rocha, María Elena. (2018). Grandes datos, grandes desafíos para las ciencias sociales. Revista mexicana de sociología, 80(2), 415-444. https://dx.doi.org/10.22201/iis.01882503p.2018.2.57723
17. Meyer-Schönberger V., Cukier, K. (2017). Big Data. La revolución de los datos masivos.
18. Miguel Ángel Pantoja. (2015). Implicaciones de seguridad de Big Data. Madrid: CLOUD SECURITY ALLIANCE ESPAÑA. Retrieved from https://www.ismsforum.es/ficheros/descargas/implicaciones-de-seguridad-de-big-data1448462176.pdf
19. Moreno, J., Serrano, M. A. & Fernández. E. (2016). Main Issues in Big Data Security. Alarcos Research Group, University of Castilla-La Mancha.
20. NBD-PWG Security and Privacy Subgroup NIST (2014).
21. Ontiveros, E., Sabater L. V. (2017). Economía de los Datos Riqueza 4.0. TELEFONICA.
22. Ontiveros, E., Sabater L. V. (2017). Economía de los Datos Riqueza 4.0. TELEFONICA.
23. Overby, A., Helvik, W., & Bjarne A. (2016). The risks of Big Data: A perspective from Computer Science Los riesgos de Big Data: Una perspectiva desde las Ciencias Computacionales Universidad de Ciencia y Tecnología. Trondheim, Noruega. Revista Antioqueña de Ciencias Computacionales y la Ingeniería de Software.
24. Rojas Cairampoma, M. (2015). Tipos de Investigación científica: Una simplificación de la complicada incoherente nomenclatura y clasificación. REDVET. Revista Electrónica de Veterinaria, 16 (1), 1-14.
25. Tabares, L. F., Hernández, J. F. (2015). Big Data Analíticas: Oportunidades, Retos y Tendencias. Especialización en Procesos para el Desarrollo de Software, Universidad de San Buenaventura Cali, Colombia.
26. Vayena, E, (2016), Elements of a New Ethical Framework for Big Data Research, 72 Wash. & Lee L. Rev. Online 420 -http://scholarlycommons.law.wlu.edu/wlulr-online/vol72/iss3/5.

27. Velasco, P, Viollier, P. (2016). Información Financiera y Discriminación laboral-Un Caso de estudio Big Data.