**TRABAJO DE GRADO**
**Opción Seminario-Diplomado.**

# Big Data en la Ciberseguridad

Corporación Universitaria Remington.
Facultad de Ingenierías
Ingeniería de sistemas

Luis Miguel Palacio Lopera
Tutor
Roberto Carlos Guevara Calume
Opción de Trabajo de grado Seminario-Diplomado.
2024.

## Tabla de Contenidos

## Resumen

En el presente trabajo se investigará cómo, en la era digital, la innovación tecnológica y la creciente cantidad de datos que se generan han dado lugar a nuevas formas de procesar y analizar información a través de herramientas como el Big Data. Esta investigación se centrará en el uso de Big Data en el ámbito de la ciberseguridad, con un enfoque especial en su aplicación dentro de las empresas.

Big Data se refiere al manejo y análisis de enormes volúmenes de datos que, debido a su magnitud, no pueden ser gestionados con las herramientas analíticas o bases de datos convencionales (Kusnetzky, 2010). En la actualidad, su implementación en el campo de la ciberseguridad ha demostrado ser una solución eficaz para la detección de amenazas y la protección de información crítica. A medida que las empresas dependen cada vez más de la digitalización, la necesidad de salvaguardar sus datos frente a posibles ataques cibernéticos ha aumentado, lo que hace de Big Data una herramienta clave en la protección y gestión segura de grandes repositorios de datos.

En este trabajo se analizará cómo la integración de Big Data y ciberseguridad permite no solo la protección efectiva de información sensible, sino también la optimización de los procesos de gestión de datos en las empresas. Asimismo, se profundizará en ambos conceptos en el contexto actual, donde las empresas deben asegurar sus datos de manera eficaz y eficiente para mantener su competitividad y protegerse contra amenazas crecientes.


**Palabras clave:** Big data, ciberseguridad, era digital, tecnología, innovación.

## Introducción

Desde comienzos de este siglo, la industria ha vivido la cuarta revolución industrial, impulsada por la digitalización, la interconexión de los procesos productivos, la gestión en línea y la aplicación de inteligencia a los procesos y productos (Schwab, 2016). Esta transformación ha modificado la forma en que las personas interactúan con internet, promoviendo el concepto de Internet de las Cosas (IoT), en el que numerosos dispositivos a nivel global se conectan y comparten información de manera autónoma, sin la necesidad de intervención humana.

Según Caiza Narváez et al., (2022) en 2017 había 9 mil millones de dispositivos IoT conectados, y se estima que para 2025 esta cifra llegará a 64 mil millones. Este aumento en el número de dispositivos conectados pone de manifiesto la creciente dependencia de la tecnología en la sociedad en los próximos años, subrayando la importancia de contar con herramientas que puedan manejar grandes cantidades de datos.

Como resultado, el término "Big Data" ha ganado relevancia en los últimos años, consolidándose como una herramienta clave para la gestión de información y la comunicación en diversos sectores de la sociedad. Big Data se refiere a un sistema de información en expansión constante, diseñado para enfrentar la complejidad del mundo moderno y generar valor tanto para las organizaciones como para las personas. Se entiende como un proceso informativo en continuo desarrollo, cuyo objetivo es alcanzar un rendimiento productivo acorde a las demandas de la era digital y del conocimiento en la que la sociedad se encuentra actualmente (Borja & Pérez, 2019).

Disponer de una herramienta tan poderosa como Big Data facilita la optimización de procesos en áreas clave como el sector empresarial, educativo y político. No obstante, estos sectores son especialmente vulnerables a ciberataques debido al alto valor económico

y estratégico de la información que manejan (Javaid et al., 2023). Según Castellanos Rojas et al., (2020) la ciberseguridad surgió como respuesta a la necesidad de proteger la información frente a diversos ataques de software malicioso, que provocan el robo de datos privados y causan grandes pérdidas.

El aumento del riesgo de que los sistemas sean comprometidos se agrava debido a que muchos dispositivos inteligentes se crean sin cumplir con los estándares de ciberseguridad necesarios. Esto puede permitir que las vulnerabilidades sean explotadas por técnicas que comprometan la seguridad, poniendo en riesgo la información personal y confidencial de los usuarios. A pesar de ello, la analítica de Big Data es capaz de recopilar, almacenar y analizar grandes volúmenes de información utilizando algoritmos de correlación para detectar irregularidades, lo que ayuda a identificar ataques maliciosos y responder de manera rápida y efectiva ante dichas amenazas (Heredia et al, 2023).

Esta realidad ha convertido a la ciberseguridad, en los últimos años, en uno de los temas más relevantes dentro del campo de las tecnologías, tanto en el estudio como en la investigación. En su informe anual "The Global Risks Report 2021", el Foro Económico Mundial destaca nuevamente a los ciberataques como uno de los principales riesgos globales percibidos (WEF, 2021). Esta preocupación es clara, ya que en la última década ha habido un aumento continuo en las búsquedas globales del término "ciberseguridad" en Google.

## Marco conceptual

Para realizar esta investigación, es crucial considerar conceptos clave que aporten claridad a los temas que se van a tratar. Entender y definir estos términos permitirá un enfoque estructurado en el análisis y facilitará una comprensión más profunda de los puntos discutidos a lo largo del estudio.

El fenómeno de Big Data puede definirse como la creciente acumulación de datos en términos de volumen, velocidad y variedad, impulsada por el avance y la adopción generalizada de las tecnologías de la información. Este proceso se debe al uso cotidiano que las personas hacen de dichas tecnologías (Camargo-Vega et al., 2015). Big Data se refiere a la existencia de vastas cantidades de datos en diferentes formatos, mayormente en soporte digital, que pueden ser analizados y aprovechados para transformar diversos aspectos de nuestra realidad social, material y personal (Becerra et al., 2023). Estos grandes conjuntos de datos proporcionan una forma de inteligencia y conocimiento que antes era inalcanzable, presentando un aura de verdad, objetividad y precisión (Boyd & Crawford, 2012).

De acuerdo con Kusnetzky (2010), el término hace referencia a la información que no puede ser procesada ni analizada a través de métodos tradicionales. Por su parte, Dans, (2011) describe Big Data como el proceso de manejo y análisis de grandes repositorios de datos que son tan vastos que no pueden ser gestionados con las herramientas analíticas y bases de datos convencionales.

Para resumir el proceso de convertir grandes volúmenes de datos en información útil, se pueden utilizar las tecnologías de Big Data. Flores Avendaño & Villacís Vera, (2017) afirman que este fenómeno permite transformar todos los datos en información

valiosa, lo que facilita la toma de decisiones en las organizaciones, mejora la eficiencia, reduce costos y aumenta los ingresos. Por estas razones, Big Data se ha convertido en una tendencia global. Aunque aún no se ha establecido un concepto científico o académico consensuado, se prevé un crecimiento continuo del mercado relacionado y de las áreas de investigación asociadas (Hernández-Leal et al., 2017). Además, Big Data no busca reemplazar los sistemas tradicionales, sino establecer una nueva tendencia que permita construir arquitecturas de sistemas capaces de manejar todas las solicitudes (Hernández et al., 2017). En resumen, Big Data tiene como objetivo mejorar la toma de decisiones, optimizar resultados y desempeño, reducir costos o generar políticas públicas que impacten positivamente a la sociedad (Vega Vargas, 2020).

Basándose en el concepto de Big Data, es importante considerar sus dimensiones, que se refieren al Volumen, la Velocidad y la Variedad.

1. **Volumen:** Cada día, las empresas experimentan un aumento significativo en la cantidad de datos que registran, medidos en terabytes, petabytes y exabytes, generados tanto por personas como por máquinas. En el año 2000, se produjeron 800.000 petabytes (PB) de datos almacenados, y se estima que esta cifra alcanzará los 35 zettabytes (ZB) para 2020. Las redes sociales también contribuyen a esta generación de datos; por ejemplo, Twitter produce más de 7 terabytes (TB) diariamente, mientras que Facebook genera 10 TB cada día. Algunas empresas son capaces de crear terabytes de datos cada hora, lo que demuestra que están inundadas de información (Coba et al., 2022).
2. **Variedad**: Esta dimensión está estrechamente relacionada con el volumen, ya que, según este último y el desarrollo tecnológico, existen múltiples formas de

representar los datos. Se pueden distinguir entre datos estructurados y no estructurados; los datos no estructurados son aquellos generados a partir de páginas web, registros de búsquedas, redes sociales, foros, correos electrónicos o datos obtenidos de sensores en diversas actividades humanas. Un ejemplo de esto es el análisis de 350 mil millones de lecturas de medidores al año para predecir el consumo de energía.

3. **Velocidad:** Esta dimensión se refiere a la rapidez con la que se generan los datos, que está relacionada con el crecimiento de productos derivados del desarrollo de software, como páginas web, registros de búsquedas, redes sociales, foros y correos electrónicos, entre otros.

Las tres características son interdependientes; por ejemplo, es posible analizar 500 millones de registros de llamadas diarias en tiempo real para anticipar la pérdida de clientes (Coba et al., 2022).

El segundo concepto central en esta investigación es la ciberseguridad. Sin embargo, antes de abordar este tema, es fundamental destacar el concepto de ciberespacio, que se define como un dominio caracterizado por el uso de datos y tecnologías electrónicas para almacenar, enviar o modificar información, gracias a los sistemas en red y las infraestructuras físicas asociadas. El ciberespacio puede considerarse como la interconexión de los seres humanos a través de telecomunicaciones, sin tener en cuenta la geografía física. La complejidad del ciberespacio, que se vuelve cada vez más dinámico, incierto y complicado, nos lleva a la necesidad de delimitar lo que debe considerarse un ataque cibernético (Ospina & Barrio, 2017). }

La ciberseguridad es el campo de las ciencias de la computación dedicado al desarrollo e implementación de mecanismos para proteger la información y la infraestructura tecnológica. Dada la importancia de contar con sistemas seguros que aseguren la confidencialidad, integridad y disponibilidad de los datos, se han propuesto diversas estrategias. Estas incluyen métodos para la evaluación de amenazas cibernéticas, como técnicas de análisis y marcos de trabajo; así como prácticas y herramientas para el desarrollo de software y hardware seguros, y sistemas de seguridad, entre otros (Urcuqui et al., 2016).

La ciberseguridad se ha consolidado como una de las áreas de las tecnologías de la información y la comunicación (TIC) que ha recibido mayor atención y esfuerzo en los últimos años. Esto se debe, por un lado, a la necesidad de responder al crecimiento constante y la sofisticación de los ataques y riesgos que enfrenta la sociedad, y por otro, al desarrollo continuo de la propia tecnología. En este contexto, donde el factor humano es fundamental, las actividades de formación y concienciación en ciberseguridad se convierten en elementos críticos que requieren una atención constante para profundizar, actualizar y mejorar continuamente (Mendivil Caldentey et al., 2022).

En el contexto internacional actual, es evidente que los ataques cibernéticos pueden afectar no solo a ordenadores y teléfonos móviles, sino también a redes informáticas inalámbricas. No hay límites ni barreras que impidan a los ciber atacantes infiltrarse en cualquier entidad conectada al ciberespacio (Ospina & Barrio, 2017). Por esta razón, las empresas y organizaciones requieren nuevas y eficaces iniciativas en el ámbito de la formación y concienciación en ciberseguridad, especialmente dirigidas al personal no

técnico, que complementen el continuo desarrollo de tecnologías y procesos (Mendivil Caldentey et al., 2022).

## Big data en la ciber seguridad

La globalización de la información y los sistemas de conexión móvil han acelerado y, en ocasiones, facilitado la transmisión de datos digitales a través de la red. Esta dinámica genera una cantidad inmensa de información que se incorpora diariamente en una extensa plataforma tecnológica. Si bien estos datos han contribuido al progreso económico, financiero, empresarial y productivo de los países en desarrollo, también han impulsado avances significativos en la sociedad y mejorado las condiciones de vida de los habitantes del planeta (Borja & Perez, 2019).

Para cualquier organización, la información es su recurso más valioso, lo que hace esencial su correcto manejo; por esta razón, muchas organizaciones están adoptando Big Data. Así, una de las principales funciones de la ciberseguridad es la monitorización de posibles ataques o intrusiones, así como del estado de funcionamiento de todos los activos que integran la infraestructura de tecnología de la información. Esto es crucial, ya que cualquier sistema o dispositivo conectado a una red puede presentar vulnerabilidades (Sainz et al., 2020).

En otras palabras, la evolución y el desarrollo continuo de los datos, junto con el creciente reconocimiento y generación de cantidades cada vez mayores, lleva a las organizaciones y entidades laborales a considerar la implementación de Big Data en sus procesos productivos y sociales. Esto les permite mejorar su representatividad en el mundo digital contemporáneo y, por ende, favorece su posicionamiento en el contexto social de cualquier país o nación, e incluso a nivel global. Esta reflexión se enmarca en los avances de la globalización tecnológica y la era digital (Borja & Perez, 2019).

La producción de enormes volúmenes de datos provenientes de múltiples fuentes ha llevado a la necesidad de desarrollar soluciones efectivas. De acuerdo con Hernández et al., el aumento en la cantidad de datos exige que las técnicas de análisis y procesamiento sean cada vez más avanzadas. El verdadero reto no se limita solo a la recolección y gestión de este gran volumen y diversidad de datos, sino también a la habilidad de extraer un valor significativo de ellos (Coba et al., 2022).

En 2021, el informe de Verizon sobre Violación de Datos (DBIR) reveló una serie de incidentes de ciberseguridad, abarcando 88 países, 83 colaboradores, 79,635 incidentes y 5,258 violaciones de datos, lo que lo convierte en un documento global de gran relevancia. Aunque se redujo el número de incidentes analizados en comparación con 2020, se registró un incremento en las filtraciones de datos, con 1,308 casos adicionales en 2021. De los resultados destacados, se observó que el 85% de las infracciones tuvieron un componente humano, el 13% de los incidentes no relacionados con ataques de denegación de servicio involucraron ransomware, y el 3% de las filtraciones se debieron a la explotación de vulnerabilidades (Bassett et al., 2021).

La transformación digital implementada en casi todas las empresas y organizaciones para alcanzar sus objetivos ha ampliado la superficie vulnerable a ciberataques. La infraestructura y los dispositivos tecnológicos que sustentan la tecnología de la información, junto con los datos mismos, se han vuelto activos cruciales, y su fallo o robo puede resultar en daños irreparables. Para fortalecer la ciberseguridad, es esencial contar con conocimientos sobre las amenazas existentes que pueden afectar a estos activos, lo que permite evaluar los riesgos y adoptar las medidas adecuadas. Esta práctica se conoce como inteligencia de amenazas (Sainz et al., 2020).

Para combatir el robo de datos, se han desarrollado diversas estrategias, entre las cuales destaca la aplicación de Inteligencia Artificial a través de Redes Neuronales Artificiales (RNA). Estas redes, formadas por nodos interconectados en diferentes niveles, se fortalecen con cada ataque mediante el aprendizaje automático (Machine Learning), que utiliza grandes volúmenes de datos (Big Data) para crear algoritmos que optimizan la capacidad lógica. Como resultado, las RNA son eficaces para reducir ataques de spam, ingeniería social y otras técnicas que buscan vulnerar la seguridad de los sistemas.

Aunque estas propuestas han establecido defensas contra cibercriminales y software malicioso, el avance de las Tecnologías de la Información y las Comunicaciones (TIC) —junto con nuevas vulnerabilidades y ataques de día cero— exige un esfuerzo constante en ciberseguridad para mitigar riesgos (Heredia & Mondragón, 2023).

En el ámbito de las dimensiones o componentes del Big Data, uno de los aspectos esenciales es el **volumen**, que hace referencia a las vastas cantidades de datos que las organizaciones buscan utilizar para mejorar su proceso de toma de decisiones, tanto en el sector privado como en él. Este volumen está estrechamente relacionado con la gran cantidad de información que las entidades reciben, la cual es fundamental para las decisiones estratégicas que apoyan el cumplimiento de sus objetivos. En términos prácticos, se puede considerar el volumen como la magnitud de la información gestionada por estas organizaciones (Becerra-Ortiz et al., 2018).

Con el desarrollo tecnológico, la ciberseguridad se enfrenta a una mayor vulnerabilidad ante nuevos tipos de ataques cibernéticos, como malware, phishing, robo de credenciales, suplantación de identidad, ataques de denegación de servicio y ataques en redes de protocolo, entre otros. Por esta razón, se utiliza la analítica de Big Data para

analizar grandes volúmenes de datos y mitigar estos incidentes, aunque esto puede generar una falsa sensación de seguridad. Esta investigación complementa la información existente sobre ciberseguridad respaldada por técnicas de Big Data durante el período de 2017 a 2021, abordando un problema que contribuye a la inseguridad persistente, como lo ha señalado el equipo de Red Team (Quezada Herrera & León Yaguana, 2022).

La implementación de técnicas de Big Data no se limita a la simple incorporación de tecnologías; requiere un diagnóstico exhaustivo de la misión de la organización, ya sea en el sector privado o público. Este proceso incluye evaluar cómo se utilizan las tecnologías de la información y comunicación (TIC) para cumplir con dicha misión, así como el tipo y clasificación de los datos manejados. Además, es esencial tener en cuenta las implicaciones legales relacionadas con el tratamiento de esos datos, los objetivos de dicho tratamiento, los beneficios potenciales para la toma de decisiones, las herramientas de análisis empleadas y las limitaciones legales que pueden surgir a partir de los resultados del análisis (Becerra-Ortiz et al., 2018).

## Conclusiones

En ciberseguridad, el uso de Big Data es crucial para detectar y reducir amenazas en tiempo real. La capacidad de procesar grandes cantidades de datos le permite identificar patrones anómalos que podrían pasar desapercibidos, lo que mejora la prevención y respuesta proactiva a los ciberataques.

Big Data no solo mejora la ciberseguridad al permitir una vigilancia constante y eficiente, sino que también ayuda a tomar mejores decisiones al ofrecer una visión completa del panorama de amenazas. Esto ayuda a las empresas a prepararse para nuevas amenazas y fortalecer su infraestructura de seguridad digital.

## Referencias

Bassett, G., Hylender, D., Pinto, A., & Widup, S. (2021). *Data Breach Investigations Report*. https://www.researchgate.net/publication/351637233_2021_Verizon_Data_Breach_Investigations_Report

Becerra, G., Castorina, J. A., Becerra, G., & Castorina, J. A. (2023). Hacia un análisis de los marcos epistémicos del big data. *Cinta de Moebio*, *76*, 50–63. https://doi.org/10.4067/S0717-554X2023000100050

Becerra-Ortiz, J. A., Cotino-Hueso, L., León, I. P., Sánchez-Acevedo, M. E., Torres-Ávila, J., Velandia-Vega, J. A., & Becerra-Ortiz, J. A. (2018). *El big data en la ciberdefensa y la ciberseguridad nacional versus el derecho a la privacidad del ciudadano colombiano*. https://hdl.handle.net/10983/22999

Borja, M. E., & Perez, M. M. (2019). Big Data: Un Analisis Documental de Su Uso y Aplicacion en el Contexto de la Era Digital. *Revista La Propiedad Inmaterial*, *28*. https://heinonline.org/HOL/Page?handle=hein.journals/revpropin28&id=271&div=&collection=

Boyd, D., & Crawford, K. (2012). CRITICAL QUESTIONS FOR BIG DATA. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Caiza Narvaez, J., Márceles Villalba, K., Amador Donado, S., José Caiza Narváez, J., Márceles Villalba, K., & Amador Donado, S. (2022). Revisión sistemática para la construcción de una arquitectura con tecnologías emergentes IoT, técnicas de inteligencia artificial, monitoreo y almacenamiento de tráfico malicioso.

*Revista Iberoamericana de Tecnologias Del Aprendizaje*, *17*(4), 386–392. https://doi.org/10.1109/RITA.2022.3217183

Camargo-Vega, J. J., Camargo-Ortega, J. F., Joyanes-Aguilar, L., Camargo-Vega -, J. J., Felipe, J., & Joyanes-Aguilar, C.-O.-L. (2015). Conociendo Big Data. *Revista Facultad de Ingeniería*, *24*(38), 63–77. http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-11292015000100006&lng=en&nrm=iso&tlng=es

Castellanos Rojas, B. S., Cortés Rodríguez, C. U., Espitia Osorio, D. J., & Garzón Bello, Y. T. (2020). Redes neuronales artificiales y estado del arte aplicado en la ciberseguridad. *Revista Matices Tecnológicos, ISSN 2027-4408, No. 12, 2020 (Ejemplar Dedicado a: Revista Matices Tecnológicos), Págs. 58-63*, *12*, 58–63. https://dialnet.unirioja.es/servlet/articulo?codigo=8994096&info=resumen&idioma=SPA

Coba, J. A. A., Barrera, L. F. A., & Sánchez, K. P. M. (2022). Perspectivas del Big data. *AlfaPublicaciones*, *4*(1.1), 514–531. https://doi.org/10.33262/ap.v4i1.1.178

Dans, E. (2011). *Big Data: una pequeña introducción » Enrique Dans*. https://www.enriquedans.com/2011/10/big-data-una-pequena-introduccion.html

Flores Avendaño, P. A., & Villacís Vera, A. E. (2017). *Análisis comparativo de las herramientas de big data en la Facultad de Ingeniería de la Pontificia Universidad Católica del Ecuador*. https://repositorio.puce.edu.ec/handle/123456789/27316

Hernández-Leal, E. J., Duque-Méndez, N. D., Moreno-Cadavid, J., Hernández-Leal, E. J., Duque-Méndez, N. D., Moreno-Cadavid, J., & Big, ". (2017).

Big Data: una exploración de investigaciones, tecnologías y casos de aplicación. *TecnoLógicas*, *20*(39), 15–38. https://doi.org/10.22430/22565337.685

Inés Ospina, G., & Marquina Barrio, A. (coord. ). (2017). La ciberseguridad como factor crítico en la seguridad de la Unión Europea. *La Estrategia Global de La Unión Europea: Asomándose al Precipicio, 2017, ISBN 978-84-617-7799-0, Págs. 77-110*, 77–110. https://dialnet.unirioja.es/servlet/articulo?codigo=8662040

Javaid, M., Haleem, A., Singh, R. P., & Suman, R. (2023). Towards insighting cybersecurity for healthcare domains: A comprehensive review of recent practices and trends. *Cyber Security and Applications*, *1*, 100016. https://doi.org/10.1016/J.CSA.2023.100016

Josu Mendivil Caldentey, D., Urquijo, B. S., & Almazor, M. G. (2022). Formación y concienciación en ciberseguridad basada en competencias: una revisión sistemática de literatura. *Pixel-Bit. Revista de Medios y Educación*, *63*(63), 197–225. https://doi.org/10.12795/PIXELBIT.91640

Kusnetzky, D. (2010). *What is "Big Data?"* | . ZDNET. https://www.zdnet.com/article/what-is-big-data/

Manrique Heredia Asesor, H., & Ing Manuel Ricardo Mondragón Vilela, M. (2023). Aplicación de big data en ciberseguridad utilizando inteligencia artificial en los años 2014 - 2023: una revisión de la literatura científica. *Universidad Privada Del Norte*. https://repositorio.upn.edu.pe/handle/11537/35265

Quezada Herrera, B. S., & León Yaguana, D. M. (2022). *Revisión sistemática de la literatura relacionada con ciberseguridad apoyada con analisis de Big Data para actividades de red Team*. http://dspace.ups.edu.ec/handle/123456789/23322

Sainz, Á., Director\es, B., Vallejo, E., & Hidalgo, R. (2020). *Despliegue de una plataforma big data de inteligencia de ciberseguridad basada en soluciones abiertas de compartición de información de amenazas*. https://repositorio.unican.es/xmlui/handle/10902/20866

Schwab, Klaus. (2016). *La cuarta revolución industrial*. https://books.google.com/books/about/La_cuarta_revoluci%C3%B3n_industrial.html?hl=es&id=BRonDQAAQBAJ

Urcuqui, C. C., Peña, M. G., Quintero, J. L. O., & Cadavid, A. N. (2016). Antidefacement. *Sistemas y Telemática*, *14*(39), 9–27. https://doi.org/10.18046/SYT.V14I39.2341

Vega Vargas, J., & Vega Vargas, J. (2020). Datos, Ciencia e Ingeniería. *Ingeniare. Revista Chilena de Ingeniería*, *28*(1), 2–5. https://doi.org/10.4067/S0718-33052020000100002

WEF. (2021). *Global Risks Report 2021 | World Economic Forum | World Economic Forum*. https://www.weforum.org/publications/the-global-risks-report-2021/