

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

LA INTEGRACIÓN DE BIG DATA EN LA INGENIERÍA DE SISTEMAS: DESAFÍOS Y  
OPORTUNIDADES PARA EL DESARROLLO DE ARQUITECTURAS ESCALABLES Y  
EFICIENTES

Corporación Universitaria Remington.

Facultad de Ingenierías

Ingeniería de Sistemas

Cristian Santiago Gómez Pinto

Miguel Lizarazo Ortiz

Oswaldo Andrés Peña Peña

Tutor:

Roberto Carlos Guevara Calume

Seminario Big Data

2024

**CONTENIDO**

RESUMEN..... 3

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

MARCO CONCEPTUAL Y CONTEXTUAL .....	3
Big Data: Definición, Características y Evolución .....	4
Ingeniería de Sistemas y su Relación con Big Data.....	5
Arquitectura de Sistemas en el Contexto de Big Data .....	6
DESARROLLO .....	7
Desafíos Técnicos en la Integración de Big Data .....	8
Oportunidades para Optimizar Arquitecturas de Sistemas Complejos .....	12
CONCLUSIONES .....	15
REFERENCIAS .....	18

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

**RESUMEN**

La explosión del volumen de datos en la última década ha transformado significativamente el panorama tecnológico. El concepto de Big Data ha emergido como una de las tecnologías más disruptivas, que ofrece nuevas posibilidades para el procesamiento, análisis y almacenamiento de grandes volúmenes de información. Sin embargo, también plantea importantes desafíos para los ingenieros de sistemas, especialmente en términos de escalabilidad y procesamiento en tiempo real. Este estudio tiene como objetivo explorar la integración de Big Data en la ingeniería de sistemas, con especial énfasis en las oportunidades que presenta para optimizar las arquitecturas de software y los retos técnicos que deben superarse para desarrollar sistemas escalables y eficientes. El análisis se divide en cinco partes: la introducción, que expone la relevancia del estudio; el marco conceptual, que define los conceptos clave y ofrece un contexto histórico y técnico; el desarrollo, donde se abordan los principales desafíos y oportunidades; y finalmente, las conclusiones, que sintetizan los hallazgos del estudio y ofrecen una perspectiva futura.

**Palabras Claves:** Big Data, Ingeniería De Sistemas, Escalabilidad, Arquitecturas Distribuidas, Seguridad De Datos

**MARCO CONCEPTUAL Y CONTEXTUAL**

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

### **Big Data: Definición, Características y Evolución**

Big Data es un término que ha ganado gran relevancia en la última década debido al crecimiento exponencial en la cantidad de datos generados por diversas fuentes, como dispositivos móviles, redes sociales, sensores de Internet de las Cosas (IoT), sistemas empresariales y otros canales. Big Data se refiere no solo al volumen de datos, sino también a la capacidad de analizarlos, gestionarlos y utilizarlos para obtener información útil en tiempo real (Zhuchyi, 2022). Este concepto implica el uso de tecnologías avanzadas y arquitecturas especializadas para procesar grandes cantidades de datos de forma rápida y eficiente.

La definición más conocida de Big Data incluye las “cuatro V”:

- **Volumen:** Big Data se caracteriza por la gran cantidad de datos generados y almacenados. Las organizaciones gestionan volúmenes de información que pueden superar los petabytes y zetabytes de datos, lo que representa un desafío significativo para las tecnologías de almacenamiento y procesamiento convencionales.
- **Velocidad:** La velocidad a la que se generan, procesan y analizan los datos es otro aspecto fundamental de Big Data. Este componente es crucial en aplicaciones donde se requieren análisis en tiempo real o casi en tiempo real, como en los sistemas de comercio electrónico, redes sociales o sistemas financieros (Sardi et al., 2023).
- **Variedad:** Los datos que se gestionan en un entorno de Big Data provienen de diversas fuentes y pueden tener diferentes formatos: estructurados, semiestructurados y no estructurados. Esto incluye datos tabulares en bases de datos, pero también documentos, archivos multimedia, registros de actividades en redes sociales, sensores IoT, entre otros.

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

- **Veracidad:** La veracidad se refiere a la calidad y precisión de los datos. Los datos en bruto, especialmente cuando se generan a gran escala, pueden contener errores, estar incompletos o ser inconsistentes. Uno de los mayores desafíos para los ingenieros de sistemas y los científicos de datos es lidiar con datos poco confiables, asegurando que el análisis posterior se base en información precisa y de calidad (Nwokeji & Matovu, 2021).

En años recientes, algunos expertos han agregado una quinta “V” al concepto de Big Data: **Valor**. Esta dimensión implica que los datos no tienen valor inherente si no se transforman en información útil. Es decir, la capacidad de extraer valor de los datos es el objetivo principal de las tecnologías de Big Data.

### **Ingeniería de Sistemas y su Relación con Big Data**

La Ingeniería de Sistemas es un campo interdisciplinario que se ocupa del diseño, desarrollo y gestión de sistemas complejos. Se trata de una disciplina clave para la integración de Big Data en infraestructuras empresariales y tecnológicas, ya que ofrece el marco metodológico y las herramientas necesarias para gestionar de manera efectiva los retos que surgen al trabajar con grandes volúmenes de información.

En el contexto de Big Data, la ingeniería de sistemas se enfrenta al desafío de diseñar arquitecturas que puedan manejar el crecimiento continuo en los datos sin sacrificar el rendimiento, la disponibilidad o la seguridad (Putrama & Martinek, 2023). Estas arquitecturas deben ser escalables, lo que significa que deben poder adaptarse a la creciente demanda de procesamiento y almacenamiento de datos sin una disminución en la eficiencia.

Al mismo tiempo, la ingeniería de sistemas también debe considerar aspectos como la interoperabilidad, la modularidad y la capacidad de integrar tecnologías emergentes. Esto incluye

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

la adopción de tecnologías de procesamiento distribuido, almacenamiento en la nube y bases de datos NoSQL, entre otros avances tecnológicos que han surgido en respuesta a las demandas de Big Data (Putrama & Martinek, 2023).

### **Arquitectura de Sistemas en el Contexto de Big Data**

Una arquitectura de sistemas escalable es esencial para manejar los desafíos que plantea Big Data. A diferencia de las arquitecturas tradicionales, que pueden estar limitadas por la capacidad de procesamiento de un solo servidor o un pequeño conjunto de servidores, las arquitecturas diseñadas para Big Data suelen adoptar un enfoque distribuido, donde múltiples servidores trabajan en conjunto para gestionar grandes volúmenes de datos y cargas de trabajo intensivas.

Existen diferentes enfoques para diseñar arquitecturas escalables, pero todos comparten algunos principios fundamentales. En lugar de mejorar el rendimiento de un solo servidor (escalabilidad vertical), la escalabilidad horizontal implica la adición de más servidores (o nodos) para manejar el aumento en la demanda (Putrama & Martinek, 2023). Esto permite que los sistemas distribuyan la carga de trabajo entre múltiples servidores y mejoren su capacidad de procesamiento sin un límite teórico superior. Tecnologías como Hadoop y Spark se basan en este enfoque.

En lugar de realizar todo el procesamiento en un solo servidor, las arquitecturas distribuidas dividen las tareas en múltiples subtareas que se ejecutan en paralelo en diferentes nodos. Este enfoque permite realizar análisis masivos de datos en menos tiempo. La capacidad de dividir tareas complejas y distribuir las eficientemente entre varios servidores es una característica central de tecnologías como Spark y Hadoop (Navaz et al., 2021).

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

En un sistema distribuido, es inevitable que uno o varios nodos experimenten fallos. Por ello, la tolerancia a fallos es un aspecto crítico en la ingeniería de sistemas para Big Data. La arquitectura debe estar diseñada para que, incluso si uno o más servidores fallan, el sistema continúe funcionando sin interrupciones. HDFS, por ejemplo, crea copias redundantes de los datos en diferentes nodos para garantizar que no se pierdan datos en caso de fallos.

Las bases de datos tradicionales (SQL) tienen limitaciones a la hora de manejar datos no estructurados o semiestructurados, que son comunes en entornos de Big Data. En respuesta, han surgido las bases de datos NoSQL, como MongoDB, Cassandra y Couchbase, que son altamente escalables y están diseñadas para manejar grandes volúmenes de datos con flexibilidad y eficiencia (Yu et al., 2023). Estas bases de datos permiten almacenar y consultar datos sin necesidad de esquemas predefinidos, lo que es ideal para entornos dinámicos y heterogéneos.

## **DESARROLLO**

La integración de Big Data en la ingeniería de sistemas presenta un conjunto de desafíos y oportunidades que requieren enfoques innovadores para el diseño, desarrollo y operación de arquitecturas de software escalables. En este apartado, se profundizará en los aspectos técnicos clave involucrados en este proceso, como la escalabilidad, el procesamiento en tiempo real, el

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

almacenamiento de datos, la consistencia y disponibilidad en sistemas distribuidos, y la seguridad y privacidad de los datos.

### **Desafíos Técnicos en la Integración de Big Data**

#### **Escalabilidad**

La escalabilidad es uno de los desafíos más críticos cuando se trata de integrar Big Data en los sistemas de software. A medida que el volumen de datos continúa creciendo a un ritmo exponencial, los sistemas deben ser capaces de manejar esta carga sin sacrificar el rendimiento. Sin una arquitectura adecuadamente diseñada para ser escalable, un sistema puede volverse ineficiente, costoso y difícil de mantener (Tiwari et al., 2023). Para abordar este desafío, se suelen emplear dos enfoques principales:

- **Escalabilidad Vertical:** Consiste en aumentar la capacidad de un único servidor mediante la adición de recursos, como más memoria RAM, CPUs más rápidas o discos de almacenamiento más grandes. Aunque este enfoque puede ser útil en ciertos casos, llega a un punto de rendimientos decrecientes, donde el costo y la complejidad de aumentar la capacidad de una sola máquina superan los beneficios obtenidos (Navaz et al., 2021). Igualmente, la escalabilidad vertical está limitada físicamente por la cantidad máxima de recursos que una máquina puede soportar. Por ello, este enfoque suele ser insuficiente para grandes volúmenes de datos.
- **Escalabilidad Horizontal:** Este enfoque es mucho más utilizado en el contexto de Big Data, ya que se basa en distribuir la carga de trabajo entre múltiples máquinas o nodos, lo que permite un escalado más flexible y efectivo. En lugar de confiar en una sola máquina más potente, el sistema distribuye las tareas entre varios servidores, cada uno manejando

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

una parte de los datos o del procesamiento (Petrasch & Petrasch, 2022). Tecnologías como Apache Hadoop y Apache Spark están diseñadas para soportar la escalabilidad horizontal de manera efectiva, dividiendo los datos en fragmentos y distribuyéndolos entre diferentes nodos para su procesamiento. Aunque esta técnica es muy poderosa, presenta desafíos adicionales en términos de coordinación, consistencia y gestión de fallos.

Una de las principales tecnologías utilizadas para escalar sistemas distribuidos es Apache Hadoop, que se basa en el paradigma de procesamiento conocido como MapReduce. Este paradigma permite descomponer grandes volúmenes de datos en partes más pequeñas que se procesan de manera paralela en varios nodos. Posteriormente, los resultados se combinan para generar una salida final (Zhuchyi, 2022). El desafío de Hadoop es que, aunque es excelente para procesamiento por lotes, no es adecuado para aplicaciones que requieren procesamiento en tiempo real o baja latencia.

Para abordar estos problemas de latencia, Apache Spark ha ganado popularidad en los últimos años. A diferencia de Hadoop, Spark puede realizar procesamiento en memoria, lo que reduce drásticamente el tiempo necesario para ejecutar las tareas, especialmente en situaciones donde los datos pueden caber en la memoria de los nodos (Zhuchyi, 2022). Spark es capaz de manejar tanto procesamiento por lotes como en tiempo real, lo que lo convierte en una opción más flexible para una variedad de aplicaciones.

### **Procesamiento en Tiempo Real**

El procesamiento en tiempo real es uno de los desafíos más complejos en la integración de Big Data, especialmente en entornos donde se requiere que las decisiones o los análisis se

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

realicen en milisegundos o segundos. Los sistemas tradicionales de Big Data, como Hadoop, están diseñados principalmente para el procesamiento por lotes, donde los datos se recopilan durante un período de tiempo y luego se procesan en grandes lotes (Maxim et al., 2021). Sin embargo, muchas aplicaciones modernas, como las plataformas de comercio electrónico, los sistemas de recomendaciones en tiempo real o la detección de fraudes, requieren la capacidad de analizar y procesar flujos continuos de datos a medida que se generan.

Para satisfacer esta necesidad, han surgido nuevas tecnologías y plataformas diseñadas específicamente para el procesamiento en tiempo real, entre las cuales destacan Apache Storm, Apache Flink y Apache Kafka Streams. Estas tecnologías permiten que los datos se procesen tan pronto como llegan al sistema, en lugar de esperar a que se acumulen en grandes volúmenes (Petrasch & Petrasch, 2022). Apache Storm permite que los datos sean procesados en paralelo de manera eficiente, lo que lo convierte en una excelente opción para aplicaciones que requieren procesamiento de alta velocidad.

Flink es otra plataforma de procesamiento en tiempo real que ha ganado popularidad debido a su capacidad para manejar tanto el procesamiento por lotes como en flujo con un solo motor. A diferencia de Storm, que trata los flujos de datos como eventos discretos, Flink trata los flujos de datos como flujos continuos, lo que permite análisis más precisos y consistentes en tiempo real (Navaz et al., 2021). Asimismo, Flink es altamente escalable y ofrece garantías de consistencia más fuertes, lo que lo hace ideal para aplicaciones críticas donde no se puede sacrificar la precisión por la velocidad.

Finalmente, Kafka Streams permite que los desarrolladores construyan aplicaciones que procesen y transformen datos en tiempo real mientras mantienen las ventajas de Kafka, como la

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

tolerancia a fallos y la alta disponibilidad. Kafka Streams ha demostrado ser muy efectivo para aplicaciones donde se requiere procesar grandes cantidades de datos en tiempo real con mínima latencia.

El principal desafío del procesamiento en tiempo real es garantizar que el sistema pueda manejar grandes picos de carga sin aumentar drásticamente la latencia o sacrificar la consistencia de los datos (Touloumis et al., 2024). Esto requiere el uso de técnicas avanzadas de equilibrio de carga, particionamiento de datos y recuperación ante fallos, lo que añade complejidad al diseño y operación del sistema.

### **Almacenamiento de Datos**

El almacenamiento de grandes volúmenes de datos es un componente clave de cualquier sistema de Big Data. Los sistemas tradicionales de bases de datos relacionales, que fueron diseñados para manejar conjuntos de datos relativamente pequeños y bien estructurados, no son adecuados para las demandas de Big Data, donde los datos pueden ser no estructurados, semiestructurados o completamente desestructurados, y donde los volúmenes pueden alcanzar petabytes o incluso exabytes.

Para abordar este desafío, han surgido nuevas tecnologías de almacenamiento, como los sistemas de archivos distribuidos y las bases de datos NoSQL. HDFS divide los datos en bloques de tamaño fijo (normalmente 128 MB o más) y los distribuye entre los nodos para garantizar la redundancia y la disponibilidad. HDFS está optimizado para lecturas secuenciales a gran escala, lo que lo hace ideal para el procesamiento por lotes en entornos de Big Data (Nogueira et al., 2020). Sin embargo, no está diseñado para cargas de trabajo de baja latencia o transacciones rápidas, lo que limita su uso en aplicaciones en tiempo real.

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

S3 es utilizado por muchas empresas como su principal solución de almacenamiento de datos, debido a su facilidad de uso, integración con otras herramientas de procesamiento de Big Data, y la capacidad de almacenar grandes volúmenes de datos de manera rentable. S3 también se integra con herramientas como Apache Spark y Amazon Redshift, lo que facilita el análisis de los datos almacenados.

Algunas de las bases de datos NoSQL más utilizadas en Big Data incluyen Cassandra, MongoDB y HBase. Estas bases de datos permiten un escalado horizontal sencillo, lo que significa que es posible agregar nodos adicionales al sistema a medida que crecen los datos sin necesidad de rediseñar la base de datos o interrumpir el servicio (Putrama & Martinek, 2023). Conjuntamente, muchas bases de datos NoSQL están diseñadas para ofrecer alta disponibilidad y tolerancia a fallos, lo que las hace ideales para aplicaciones críticas.

A pesar de estas soluciones, el almacenamiento de datos a gran escala sigue siendo un desafío, especialmente en términos de costos y rendimiento. Aunque el almacenamiento en la nube ha reducido el costo inicial de almacenar grandes volúmenes de datos, el costo de almacenamiento a largo plazo y el costo de acceder y procesar esos datos puede ser significativo (He & Wang, 2020). Al mismo tiempo, optimizar el rendimiento del almacenamiento en entornos distribuidos es complejo, ya que implica equilibrar la latencia, el rendimiento de lectura/escritura y la consistencia de los datos.

### **Oportunidades para Optimizar Arquitecturas de Sistemas Complejos**

A pesar de los desafíos mencionados, la integración de Big Data en la ingeniería de sistemas también presenta una serie de oportunidades importantes, especialmente en términos de optimización y mejora de las arquitecturas de software.

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

### **Mejora de la Toma de Decisiones Basada en Datos**

Una de las oportunidades más claras de Big Data es la mejora en la capacidad de tomar decisiones informadas basadas en datos. Las organizaciones que adoptan sistemas de Big Data pueden analizar grandes volúmenes de información para identificar patrones, tendencias y correlaciones que no serían evidentes de otra manera.

Los sistemas escalables de Big Data permiten que estos análisis se realicen de manera rápida y eficiente, proporcionando a las empresas la capacidad de reaccionar en tiempo real a eventos y cambios en el mercado. Esto es particularmente valioso en sectores como la banca, el comercio minorista y la atención médica, donde la capacidad de tomar decisiones rápidas y precisas puede generar una ventaja competitiva significativa (Wang & Alexander, 2020).

Por ejemplo, las empresas de comercio electrónico pueden utilizar sistemas de Big Data para analizar el comportamiento de los clientes en tiempo real y ajustar dinámicamente los precios o recomendaciones de productos. Del mismo modo, las instituciones financieras pueden monitorear transacciones en tiempo real para detectar fraudes potenciales y tomar medidas inmediatas.

### **Automatización Avanzada mediante Inteligencia Artificial**

Big Data también está profundamente relacionado con el avance de la inteligencia artificial (IA) y el aprendizaje automático. Los sistemas de IA modernos dependen de grandes volúmenes de datos para entrenar modelos precisos y eficaces. La integración de Big Data en la ingeniería de sistemas permite a las empresas aprovechar la IA para automatizar tareas complejas, mejorar la precisión de las predicciones y optimizar la toma de decisiones.

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

La arquitectura escalable es esencial para los sistemas de IA, ya que el entrenamiento de modelos de aprendizaje automático requiere una cantidad significativa de recursos de procesamiento y almacenamiento (Rossi & Hiram, 2022). Además, la capacidad de procesar datos en tiempo real permite que los modelos de IA se adapten continuamente a nuevos datos y mejoren su rendimiento con el tiempo.

Por ejemplo, en el ámbito de la atención médica, los sistemas de IA pueden analizar grandes volúmenes de datos de pacientes, como historiales médicos y resultados de pruebas, para identificar patrones que pueden ser útiles para diagnosticar enfermedades o predecir resultados. En el sector de la logística, los sistemas de IA pueden analizar datos en tiempo real sobre el tráfico, el clima y la demanda de productos para optimizar rutas de entrega y reducir costos.

### **Innovación en Sistemas Distribuidos**

La integración de Big Data ha fomentado la innovación en el diseño de arquitecturas distribuidas. Las tecnologías de sistemas distribuidos, como los microservicios y los contenedores (por ejemplo, Docker y Kubernetes), han permitido la creación de sistemas más modulares, flexibles y escalables que pueden manejar grandes volúmenes de datos de manera eficiente (Sardi et al., 2023).

Los microservicios son una arquitectura de software que descompone una aplicación en servicios pequeños y autónomos, cada uno de los cuales se encarga de una funcionalidad específica. Esto permite que los diferentes servicios puedan escalarse de manera independiente según las necesidades del sistema, lo que mejora tanto la eficiencia como la resiliencia.

Los contenedores, por otro lado, son una tecnología que permite a los desarrolladores empaquetar aplicaciones y sus dependencias en unidades portátiles que se pueden ejecutar de

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

manera consistente en diferentes entornos. Kubernetes, una plataforma de orquestación de contenedores, permite que los contenedores se desplieguen, gestionen y escalen automáticamente en un clúster de máquinas (Mistrík et al., 2021).

La combinación de microservicios y contenedores ha transformado la forma en que se construyen y operan los sistemas de Big Data, permitiendo a las organizaciones desarrollar y desplegar aplicaciones más rápidamente, mejorar la escalabilidad y reducir los costos de infraestructura.

### **CONCLUSIONES**

Uno de los principales retos de la integración de Big Data es la escalabilidad. A medida que los datos crecen de manera exponencial, los sistemas tradicionales no pueden manejar eficientemente el volumen y la velocidad de la información. Aquí, la escalabilidad horizontal se ha impuesto como una solución, permitiendo distribuir las cargas de trabajo entre múltiples servidores o nodos. Tecnologías como Hadoop y Apache Cassandra han revolucionado el manejo de datos, pero siguen existiendo desafíos en términos de consistencia de los datos, latencia y coordinación entre nodos.

A medida que los datos continúan creciendo, la ingeniería de sistemas se enfrenta a la necesidad de optimizar aún más la escalabilidad. Las arquitecturas distribuidas deben ser robustas y adaptables, capaces de soportar grandes cargas de datos sin comprometer el rendimiento. El procesamiento en tiempo real es cada vez más crítico en aplicaciones modernas.

## **TRABAJO DE GRADO**

### **Opción Seminario-Diplomado.**

En un mundo donde los usuarios esperan respuestas instantáneas, el procesamiento en lotes ya no es suficiente. Tecnologías como Apache Spark Streaming y Apache Flink permiten analizar datos a medida que se generan, facilitando la toma de decisiones instantáneas en sectores como el comercio electrónico o la banca.

El desafío aquí radica en la complejidad técnica: mantener la consistencia de datos mientras se procesan flujos en tiempo real y garantizar la baja latencia. A pesar de las ventajas de estos sistemas, los costos operativos y las demandas de recursos son altos, lo que obliga a las organizaciones a optimizar sus infraestructuras de procesamiento. Las arquitecturas distribuidas han sido fundamentales para soportar el auge del Big Data. En lugar de depender de un único servidor, los sistemas distribuyen la carga de trabajo entre varios nodos, lo que aumenta la eficiencia y la resiliencia ante fallos. Sin embargo, este enfoque también presenta retos como la latencia y la consistencia de los datos.

El teorema CAP (Consistencia, Disponibilidad y Tolerancia a Fallos) establece que los sistemas distribuidos deben hacer concesiones en uno de estos tres aspectos. Elegir entre consistencia o disponibilidad depende de la aplicación y las necesidades específicas del sistema. No obstante, la gestión de fallos en nodos y la optimización de la latencia siguen siendo áreas activas de investigación y desarrollo. A medida que los sistemas de Big Data se expanden, las preocupaciones sobre la seguridad y la privacidad aumentan. La naturaleza distribuida de estas infraestructuras implica más puntos vulnerables que pueden ser explotados por atacantes. El cifrado de datos en reposo y en tránsito, junto con mecanismos avanzados de autenticación y control de accesos, son esenciales para proteger los sistemas.

## **TRABAJO DE GRADO**

### **Opción Seminario-Diplomado.**

En paralelo, las regulaciones como el GDPR y la CCPA exigen una mayor protección de los datos personales. La ingeniería de sistemas debe cumplir con estos requisitos, lo que implica desarrollar arquitecturas que integren la privacidad por diseño, minimizando los riesgos de exposición de datos y garantizando el cumplimiento normativo. La inteligencia artificial (IA) y el aprendizaje automático (ML) dependen de Big Data para entrenar modelos precisos y robustos. La capacidad de manejar grandes volúmenes de datos en tiempo real permite que los sistemas de IA mejoren continuamente. Sin embargo, la integración de IA también presenta desafíos en términos de demanda de recursos y la necesidad de arquitecturas escalables para manejar el procesamiento intensivo.

Los sistemas de Big Data permiten entrenar modelos de IA a gran escala, lo que impulsa avances en áreas como la personalización de productos y la toma de decisiones automatizadas. Sin embargo, es crucial diseñar infraestructuras flexibles y eficientes que puedan adaptarse a las crecientes necesidades de procesamiento de estos modelos. El futuro de la ingeniería de sistemas en la era de Big Data presenta tanto retos como oportunidades. La computación cuántica, por ejemplo, tiene el potencial de revolucionar el procesamiento de datos, pero aún está en sus primeras fases de desarrollo.

Por otro lado, el edge computing está ganando terreno, permitiendo que los datos se procesen más cerca de su fuente, lo que reduce la latencia y mejora la eficiencia. La sostenibilidad es otra preocupación emergente. Los sistemas de Big Data consumen enormes cantidades de energía, por lo que la ingeniería de sistemas debe centrarse en soluciones más eficientes energéticamente, como el uso de energías renovables y el diseño de algoritmos que minimicen el consumo de recursos.

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

**REFERENCIAS**

- Bandara, M., & Rabhi, F. A. (2020). Semantic modeling for engineering data analytics solutions. *Semantic Web, 11*(3), 525–547.
- Farhaoui, Y., & El Allaoui, A. (2024). Sustainability in the Internet of Things: Insights, Scope, and AI-Driven Optimized Water Management with Big Data Integration. *The International Workshop on Big Data and Business Intelligence*, 468–475.
- He, Q. P., & Wang, J. (2020). Application of systems engineering principles and techniques in biological big data analytics: A review. *Processes, 8*(8), 951.
- Li, B. H., Chai, X., Lin, T. Y., Yang, C., Hou, B., Liu, Y., Song, X., & Gong, H. (2020). Cyber-physical system engineering oriented intelligent high performance simulation cloud. *Simulation for Cyber-Physical Systems Engineering: A Cloud-Based Context*, 89–118.
- Li, J. (2024). Big Data-driven Decision Support: Enhancing Information Integration and User Experience with Mobile Integrated Technology. *Journal of Information Systems*

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

*Engineering and Management*, 9(2), 24148.

Maxim, B. R., Galster, M., Mistrik, I., & Tekinerdogan, B. (2021). Data-intensive systems, knowledge management, and software engineering. In *Knowledge Management in the Development of Data-Intensive Systems* (pp. 1–40). Auerbach Publications.

Mistrík, I., Galster, M., Maxim, B. R., & Tekinerdogan, B. (2021). *Knowledge Management in the Development of Data-Intensive Systems*. CRC Press.

Navaz, A. N., Serhani, M. A., El Kassabi, H. T., Al-Qirim, N., & Ismail, H. (2021). Trends, technologies, and key challenges in smart and connected healthcare. *Ieee Access*, 9, 74044–74067.

Nogueira, M., Galvão, J., & Santos, M. Y. (2020). A Data Modelling Method for Big Data Warehouses. *Information Systems: 16th European, Mediterranean, and Middle Eastern Conference, EMCIS 2019, Dubai, United Arab Emirates, December 9–10, 2019, Proceedings 16*, 85–98.

Nwokeji, J. C., & Matovu, R. (2021). A systematic literature review on big data extraction, transformation and loading (etl). *Intelligent Computing: Proceedings of the 2021 Computing Conference, Volume 2*, 308–324.

Petrasch, R. J., & Petrasch, R. R. (2022). Data integration and interoperability: Towards a model-driven and pattern-oriented approach. *Modelling*, 3(1), 105–126.

Rossi, R., & Hiramama, K. (2022). Analytical Engineering for Data Stream. *Journal of Computer and Communications*, 10(7), 13–34.

Zhuchyi, L. I. (2022). Ontological Support for Harmonization and Integration of Ukrzaliznytsia

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

Information Systems Data. *Science and Transport Progress*, 1 (97), 32–49.

Yu, H. Y., Ogbeyemi, A., Lin, W. J., He, J., Sun, W., & Zhang, W.-J. (2023). A semantic model for enterprise application integration in the era of data explosion and globalisation. *Enterprise Information Systems*, 17(4), 1989495.