

UNIVERSIDAD ESTATAL A DISTANCIA
VICERRECTORÍA ACADÉMICA ESCUELA DE
CIENCIAS EXACTAS Y NATURALES CARRERA

INGENIERÍA INFORMÁTICA

CÁTEDRA INGENIERÍA DE SOFTWARE

ASIGNATURA

03084 BASES DE DATOS II

Proyecto 1

VALOR: 30% (3.0)

ESTUDIANTE:

Francisco Campos Sandi

CÉDULA:

114750560

I CUATRIMESTRE 2025

Contenido

GLOSARIO:.....	5
INTRODUCCIÓN	7
DESARROLLO.....	8
NETFLIX:	8
1. Definición de los datos	8
A. DATOS ESTRUCTURADOS:.....	8
B. DATOS SEMIESTRUCTURADOS:	8
C. DATOS NO ESTRUCTURADOS:	8
D. DATOS INTERNOS:	8
E. DATOS EXTERNOS:.....	8
2. MODELO DE LAS 7 V DE BIG DATA.....	9
B. VELOCIDAD.....	9
C. VARIEDAD	10
D. VERACIDAD	10
E. VALOR	10
F. VISUALIZACIÓN	11
G. VIABILIDAD	11
3. ARQUITECTURA DE BIG DATA.....	11
A. IDENTIFICACIÓN DE FUENTES DE DATOS.....	11
B. RECOLECCIÓN (INGESTA) DE DATOS.....	12
C. ALMACENAMIENTO DE DATOS	12
D. PROCESAMIENTO.....	12
E. ANÁLISIS DE DATOS.....	12
F. VISUALIZACIÓN DE DATOS (RESULTADOS)	13
4. PLATAFORMAS Y HERRAMIENTAS DE BIG DATA	13
A. HERRAMIENTA LICENCIADA: MICROSOFT AZURE HDINSIGHT.....	13
B. HERRAMIENTA DE CÓDIGO ABIERTO: APACHE SPARK	13
STARBUCKS	13
1. DEFINICIÓN DE LOS DATOS.....	13

A. DATOS ESTRUCTURADOS:	13
B. DATOS SEMIESTRUCTURADOS:	14
C. DATOS NO ESTRUCTURADOS:	14
D. DATOS INTERNOS:	14
E. DATOS EXTERNOS:	14
1. MODELO DE LAS 7 V DE BIG DATA	15
A. VOLUMEN	15
B. VELOCIDAD	15
C. VARIEDAD	15
D. VERACIDAD	16
E. VALOR	16
F. VISUALIZACIÓN	16
G. VIABILIDAD	16
3. ARQUITECTURA DE BIG DATA	17
A. IDENTIFICACIÓN DE FUENTES DE DATOS	17
B. RECOLECCIÓN (INGESTA) DE DATOS	17
C. ALMACENAMIENTO DE DATOS	17
D. PROCESAMIENTO	17
E. ANÁLISIS DE DATOS	18
F. VISUALIZACIÓN DE DATOS (RESULTADOS)	18
3. PLATAFORMAS Y HERRAMIENTAS DE BIG DATA	18
A. HERRAMIENTA LICENCIADA: IBM INFOSPHERE BIGINSIGHTS	18
B. HERRAMIENTA DE CÓDIGO ABIERTO: HADOOP	19
CONCLUSIÓN	21
BIBLIOGRAFÍA	22
ANEXO	23

Índice de ilustraciones

Ilustración 1 Data Analysis Of Starbucks' Global Presence	19
Ilustración 2: Las 7 V del Big Data	25

Índice de tablas

Tabla 1 IBM vs Hadoop.....	26
----------------------------	----

Índice de anexos

Anexo 1 "¿Qué es el Big Data?.....	23
Anexo 2 Las 7 V de Big Data	25
Anexo 3 Comparativo: IBM InfoSphere BigInsights vs Hadoop	26

Glosario:

- **Datos estructurados:** Información organizada y almacenada de manera clara, como datos de usuarios, transacciones, etc.
- **Datos semiestructurados:** Información que tiene alguna estructura, como metadatos o listas personalizadas de usuarios.
- **Datos no estructurados:** Información que no sigue un formato predefinido, como comentarios de usuarios o imágenes.
- **Datos internos:** Información generada dentro de una empresa, como historial de navegación o resultados de pruebas A/B.
- **Datos externos:** Información obtenida de fuentes externas, como redes sociales o tendencias de búsqueda.
- **Volumen:** Gran cantidad de datos manejados por las plataformas.
- **Velocidad:** Rapidez con la que se procesan los datos.
- **Variedad:** Diversidad de los tipos de datos utilizados.
- **Veracidad:** Precisión y fiabilidad de los datos.
- **Valor:** Utilidad de los datos para generar estrategias efectivas.
- **Visualización:** Representación gráfica de los datos para análisis.
- **Viabilidad:** Capacidad de las empresas para gestionar eficientemente los datos.
- **Identificación de fuentes de datos:** Proceso de identificar de dónde provienen los datos relevantes.
- **Recolección de datos:** Captura de datos de diversas fuentes.
- **Almacenamiento de datos:** Guardado de datos de manera estructurada y accesible.
- **Procesamiento de datos:** Transformación de datos para generar información útil.
- **Análisis de datos:** Evaluación de los datos para identificar patrones y tendencias.
- **Visualización de datos:** Herramientas que permiten ver y analizar los resultados.

- **Microsoft Azure HDInsight:** Plataforma de análisis de datos en tiempo real.
- **Apache Spark:** Herramienta de procesamiento de datos en tiempo real compatible con diversos tipos de datos.
- **IBM InfoSphere BigInsights:** Herramienta licenciada para manejar datos no estructurados y realizar análisis avanzados.
- **Hadoop:** Sistema distribuido para almacenar y procesar grandes volúmenes de datos.

Introducción

El tema desarrollado en este trabajo se centra en la implementación de Big Data como herramienta estratégica para empresas líderes en sus respectivas industrias, como Netflix y Starbucks. Big Data, entendido como la gestión y análisis de grandes volúmenes de datos, se ha convertido en un recurso esencial para mejorar la personalización, optimizar procesos operativos y fortalecer la relación con los clientes. Este documento explora cómo estas organizaciones aplican técnicas avanzadas para aprovechar los datos y mantenerse competitivas en un mercado global.

La importancia de este tema radica en el impacto transformador que tiene el Big Data en el entorno empresarial actual. A través de la recopilación, almacenamiento, procesamiento y análisis de datos, estas empresas logran tomar decisiones fundamentadas, anticipar tendencias de consumo y personalizar experiencias. Así, el trabajo busca resaltar cómo la correcta implementación de estas tecnologías puede ser un factor clave para la innovación y el éxito empresarial.

El contenido incluye un análisis detallado de la arquitectura de Big Data para cada caso, desglosando sus etapas principales, como la identificación de fuentes, la ingesta y el procesamiento de datos. Además, se presenta una comparación entre las herramientas licenciadas y de código abierto utilizadas por Netflix y Starbucks, evaluando sus ventajas y limitaciones. También se profundiza en el uso estratégico de estas herramientas para obtener valor y mejorar la eficiencia operativa.

Finalmente, el trabajo incluye un anexo que amplía las ideas tratadas, proporcionando referencia de un artículo relevante de aplicaciones de Big Data. Este enfoque integral permite entender no solo cómo funciona el Big Data en estas empresas, sino también cómo puede ser adoptado por otras organizaciones interesadas en transformar su gestión de datos. Así, se ofrece una visión completa de las oportunidades y desafíos que esta tecnología representa.

Desarrollo

Netflix:

1. Definición de los datos

Netflix

a. Datos estructurados:

- Información demográfica de los usuarios (edad, género, ubicación).
- Historial de reproducciones (ID de contenido, fecha y hora, duración de reproducción).
- Calificaciones otorgadas a películas y series.
- Datos transaccionales (suscripciones, métodos de pago, facturas).

b. Datos semiestructurados:

- Metadatos de las películas y series (títulos, géneros, elenco, etiquetas).
- Listas de contenido almacenadas (listas de "para ver después" de los usuarios).

c. Datos no estructurados:

- Reseñas y comentarios escritos por usuarios.
- Contenidos vistos (análisis visual o auditivo, como escenas vistas repetidamente).
- Imágenes o miniaturas de las películas y series.

d. Datos internos:

- Historial de navegación dentro de la plataforma.
- Información de las reproducciones y uso de la aplicación.
- Datos de prueba A/B para optimizar recomendaciones.

e. Datos externos:

- Información de tendencias de búsqueda en motores como Google.

- Comentarios y reseñas en redes sociales.
- Datos adquiridos de socios sobre preferencias de entretenimiento.

2. Modelo de las 7 V de Big Data

En el Anexo 1 se presenta información adicional sobre el concepto de Big Data, su definición, características principales y su relevancia en el entorno empresarial actual, basado en el artículo "¿Qué es el Big Data? Usos y beneficios" de J. De Nava (2025), publicado por *Empower Talent*. El texto aborda las tres V del Big Data: volumen, velocidad y variedad, y explica cómo las empresas pueden aprovechar estas características para mejorar la toma de decisiones, optimizar procesos y obtener ventajas competitivas. Además, se detallan las fases del procesamiento de Big Data, desde la recopilación hasta el análisis de datos, enfatizando su importancia en la identificación de oportunidades y la mejora de la eficiencia operativa.

a. Volumen

Netflix maneja una inmensa cantidad de datos generados diariamente por sus millones de usuarios, los cuales incluyen reproducciones, búsquedas y valoraciones. Estos datos alimentan su algoritmo de recomendaciones personalizadas. "Las empresas amasan grandes volúmenes de datos, desde terabytes hasta petabytes" (Joyanes Aguilar, 2019, p.10), lo que resalta la escala de información que empresas como Netflix manejan. Este análisis evidencia la necesidad de sistemas robustos de almacenamiento y procesamiento para gestionar eficientemente estos volúmenes masivos de datos.

b. Velocidad

El éxito de Netflix depende de su capacidad para procesar datos en tiempo real, ofreciendo recomendaciones personalizadas a sus usuarios al instante. Esto mejora la experiencia de usuario al mantener la plataforma ágil y reactiva. "La importancia de la velocidad

de los datos o el aumento creciente de los flujos de datos en las organizaciones, junto con la frecuencia de las actualizaciones de las grandes bases de datos" (Joyanes Aguilar, 2019, p.11), enfatiza cómo la rapidez en el análisis es fundamental para responder de manera efectiva a las necesidades del usuario. Este aspecto permite que Netflix se mantenga competitivo y adaptable en un mercado dinámico.

c. Variedad

Netflix utiliza una amplia gama de datos, desde metadatos de las películas y series hasta comentarios y valoraciones de los usuarios. La integración de estos diferentes tipos de datos permite personalizar las recomendaciones y atender las demandas específicas de cada cliente. "La variedad representa todos los tipos de datos y supone un desplazamiento fundamental en el análisis de requisitos" (Joyanes Aguilar, 2019, p.12), lo cual destaca la importancia de la diversidad de datos para enriquecer el análisis y la personalización en plataformas de entretenimiento.

d. Veracidad

La precisión de los datos recopilados por Netflix es crucial para garantizar que las recomendaciones sean relevantes y útiles. Para ello, la empresa se enfoca en la validación y limpieza de los datos, asegurando que representen fielmente las preferencias reales de los usuarios.

e. Valor

El algoritmo de Netflix transforma los datos recopilados en contenido altamente relevante para los usuarios, maximizando la satisfacción y fortaleciendo la fidelidad a la plataforma. Este enfoque demuestra cómo el valor derivado de los datos no solo se traduce en ingresos, sino también en relaciones duraderas con los consumidores.

f. Visualización

Netflix emplea herramientas avanzadas de visualización para analizar patrones de uso y monitorizar el rendimiento de su algoritmo. Esta capacidad ayuda a identificar áreas de mejora y ajustar las estrategias para mantener una experiencia de usuario óptima.

g. Viabilidad

Netflix es un ejemplo claro de cómo las empresas pueden aprovechar eficazmente sus datos para innovar y satisfacer a sus clientes. "Se trata de la capacidad que tienen las compañías en generar un uso eficaz del gran volumen de datos que manejan" (Barrantes, 2022, párr.03), lo cual evidencia que la gestión eficiente de datos es una competencia esencial para mantenerse relevante en la industria del entretenimiento. En el **Anexo 2** se presenta una imagen explicativa sobre las 7 V de Big Data, la cual permite visualizar de manera concreta las dimensiones clave del manejo de grandes volúmenes de información. Esta representación facilita la comprensión de las características y funcionalidades que hacen viable el uso estratégico de los datos en organizaciones como Netflix

3. Arquitectura de Big data

a. Identificación de fuentes de datos

Netflix emplea diversas fuentes de datos que incluyen reproducciones, búsquedas, valoraciones y comentarios de los usuarios. Estas fuentes proporcionan información clave para entender los intereses de su audiencia. "El Big Data permite a las plataformas entender las preferencias y hábitos de visualización de su audiencia." (Castro, 2021, párr.02), lo cual evidencia cómo la identificación de estas fuentes permite a Netflix anticiparse a las necesidades de sus usuarios. Este proceso de selección de datos asegura un enfoque centrado en el cliente y optimiza su experiencia.

b. Recolección (ingesta) de datos

La ingesta de datos en Netflix se realiza mediante herramientas avanzadas que capturan información en tiempo real, garantizando que los datos sean incorporados de manera eficiente al sistema. "Los algoritmos de recomendación, impulsados por macrodatos, utilizan esta información para proporcionar a los usuarios sugerencias personalizadas." (Castro, 2021, párr.04), lo que ilustra la importancia de una ingesta efectiva para alimentar procesos de recomendación. Este enfoque asegura que los usuarios reciban contenido relevante y adaptado a sus intereses.

c. Almacenamiento de datos

Netflix utiliza sistemas de almacenamiento distribuidos que garantizan escalabilidad y alta disponibilidad. Este enfoque permite manejar el vasto volumen de datos que generan sus usuarios diariamente, consolidando la información de manera organizada para su posterior procesamiento.

d. Procesamiento

Los datos recolectados son procesados mediante algoritmos que evalúan patrones de consumo y preferencias individuales. Este procesamiento es esencial para garantizar la precisión y relevancia de las recomendaciones generadas, optimizando la interacción entre la plataforma y los usuarios.

e. Análisis de datos

El análisis de datos en Netflix va más allá de la simple recolección, incorporando técnicas que predicen las tendencias de visualización y mejoran la personalización del contenido. Esto permite que la plataforma evolucione y se adapte constantemente al comportamiento de sus usuarios, fortaleciendo su posición en el mercado.

f. Visualización de datos (resultados)

Netflix emplea dashboards interactivos que permiten monitorear los resultados de su algoritmo en tiempo real. Estas herramientas facilitan la interpretación de los datos procesados y el ajuste continuo de las estrategias, mejorando la experiencia general de los usuarios.

4. Plataformas y herramientas de big data

a. Herramienta licenciada: Microsoft Azure HDInsight

Esta plataforma es ideal para Netflix debido a su capacidad para manejar grandes volúmenes de datos y su integración con herramientas de análisis avanzadas. Azure HDInsight permite procesar datos en tiempo real, lo cual es crucial para las recomendaciones personalizadas de Netflix. Además, su escalabilidad asegura que pueda manejar el crecimiento constante de datos generados por los usuarios.

b. Herramienta de código abierto: Apache Spark

Apache Spark es una excelente opción para Netflix porque permite el procesamiento de datos en tiempo real y es compatible con múltiples fuentes de datos. Su capacidad para manejar tanto datos estructurados como no estructurados lo hace ideal para analizar patrones de visualización y mejorar el algoritmo de recomendaciones.

Starbucks

1. Definición de los datos

a. Datos estructurados:

- Transacciones en puntos de venta (artículos comprados, montos, métodos de pago).
- Información de tarjetas de fidelidad (número de usuario, puntos acumulados).
- Horarios y ubicaciones de compras.

b. Datos semiestructurados:

- Preferencias de los clientes en la aplicación móvil (personalización de bebidas, historial de pedidos).
- Informes de inventario y tiempos de espera por tienda.

c. Datos no estructurados:

- Comentarios en redes sociales sobre el servicio o los productos.
- Reseñas en plataformas de terceros.
- Imágenes relacionadas con Starbucks subidas por los clientes (ejemplo: fotos de sus bebidas).

d. Datos internos:

- Historial de compras de los clientes.
- Resultados de campañas promocionales previas.
- Datos de productividad y rendimiento de las tiendas.

e. Datos externos:

- Opiniones y tendencias detectadas en redes sociales.
- Clima y temporadas que influyen en el consumo (por ejemplo, aumento en bebidas frías durante el verano).
- Datos demográficos de zonas cercanas a las tiendas.

1. Modelo de las 7 V de Big Data

a. Volumen

Starbucks recopila datos sobre transacciones, preferencias de los clientes y comentarios en redes sociales, acumulando una gran cantidad de información que alimenta sus estrategias de personalización. "Las empresas amasan grandes volúmenes de datos, desde terabytes hasta petabytes" (Joyanes Aguilar, 2019, p.10), lo que refleja cómo empresas del sector retail deben enfrentar desafíos similares en la gestión de datos masivos. Esto es esencial para garantizar un análisis eficaz y una experiencia adaptada al cliente.

b. Velocidad

La rapidez con la que Starbucks procesa los datos es clave para responder en tiempo real a las demandas de los clientes, ajustar promociones y resolver problemas detectados en redes sociales. "La importancia de la velocidad de los datos o el aumento creciente de los flujos de datos en las organizaciones, junto con la frecuencia de las actualizaciones de las grandes bases de datos" (Joyanes Aguilar, 2019, p.11), destaca la necesidad de tecnologías ágiles para el análisis de datos. Starbucks utiliza esta capacidad para mantener su relevancia y adaptabilidad en un mercado competitivo.

c. Variedad

Starbucks integra múltiples fuentes de datos: estructurados, semiestructurados y no estructurados, lo que incluye transacciones, preferencias móviles y opiniones en redes sociales. "La variedad representa todos los tipos de datos y supone un desplazamiento fundamental en el análisis de requisitos" (Joyanes Aguilar, 2019, p.12), ilustrando cómo la diversidad de datos enriquece el análisis y permite a Starbucks desarrollar estrategias precisas y personalizadas.

d. Veracidad

La fiabilidad de los datos es crucial para Starbucks, especialmente cuando utiliza análisis de sentimientos en redes sociales para identificar problemas o captar tendencias. Esto garantiza que las decisiones se basen en información precisa y útil.

e. Valor

Starbucks utiliza los datos para ofrecer promociones personalizadas y diseñar estrategias que impacten directamente en la experiencia del cliente. Este enfoque demuestra cómo los datos pueden generar valor tanto para los consumidores como para la empresa.

f. Visualización

Starbucks emplea herramientas visuales para interpretar datos de ventas, tendencias y comentarios, lo que facilita la toma de decisiones estratégicas basadas en información clara y comprensible.

g. Viabilidad

La capacidad de Starbucks para aprovechar eficientemente sus datos le permite adaptarse a las preferencias de sus clientes y optimizar sus operaciones. "Se trata de la capacidad que tienen las compañías en generar un uso eficaz del gran volumen de datos que manejan" (Barrantes, 2022, párr.03), subrayando cómo una gestión adecuada de los datos puede transformar las operaciones y la relación con los clientes.

3. Arquitectura de Big data

a. Identificación de fuentes de datos

Starbucks identifica fuentes de datos como transacciones en puntos de venta, opiniones en redes sociales y registros de su aplicación móvil. Estas fuentes son esenciales para optimizar su oferta y garantizar una relación personalizada con los clientes. "A través del análisis de diferentes conjuntos de datos, Starbucks ha conseguido optimizar su menú, personalizar las ofertas de los clientes, mejorar la eficiencia de sus operaciones y aumentar la lealtad de los clientes." (Vargas, 2023, párr.02), lo cual evidencia cómo la identificación de fuentes permite a Starbucks diseñar estrategias efectivas. Este análisis destaca la importancia de centralizar la información obtenida.

b. Recolección (ingesta) de datos

La ingesta de datos en Starbucks se realiza mediante herramientas que integran información desde los puntos de venta hasta las plataformas móviles. "Además el éxito con el análisis de datos de Starbucks, es pionera en el uso de aplicaciones móviles, sistemas de fidelización y métodos y tarjetas de pago." (Suárez, 2021, párr.05), ilustrando cómo las tecnologías de ingesta permiten una captura eficiente. Este enfoque asegura que las promociones y estrategias lleguen al cliente de manera efectiva.

c. Almacenamiento de datos

Starbucks utiliza sistemas de almacenamiento en la nube que aseguran la organización y protección de los datos recolectados. Este enfoque permite manejar la gran cantidad de información generada por sus clientes y garantizar que esté disponible para análisis posteriores.

d. Procesamiento

Los datos ingresados son procesados mediante algoritmos que identifican patrones de consumo y preferencias de los clientes. Este procesamiento facilita la creación de estrategias

adaptativas y el diseño de promociones personalizadas, fortaleciendo la relación de la marca con sus clientes.

e. Análisis de datos

El análisis de datos en Starbucks combina técnicas de minería de datos y análisis predictivo, permitiéndole anticiparse a las necesidades de los clientes. Esto mejora la eficiencia operativa y refuerza su posición como líder en personalización de servicios.

f. Visualización de datos (resultados)

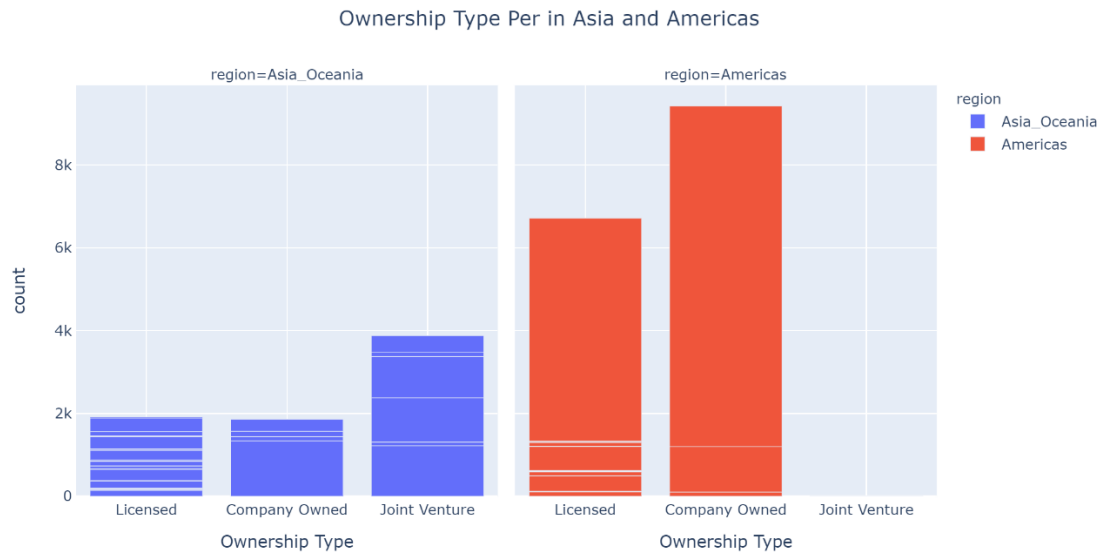
Starbucks utiliza herramientas de visualización para interpretar los resultados de sus análisis, desde gráficos sobre tendencias de compra hasta reportes de rendimiento de campañas. Esto facilita la toma de decisiones estratégicas basadas en información clara y comprensible.

3. Plataformas y herramientas de big data

a. Herramienta licenciada: IBM InfoSphere BigInsights

Esta herramienta es adecuada para Starbucks porque combina capacidades de análisis avanzado con un enfoque en la gestión de datos no estructurados, como comentarios en redes sociales. InfoSphere BigInsights también ofrece herramientas de visualización que permiten a Starbucks interpretar tendencias y ajustar estrategias de marketing.

Ilustración 1 Data Analysis Of Starbucks' Global Presence



Fuente: Tomado de Pantinople, R. (2023). *Búsqueda visual* Guardar Ver imagen

Compartir Más Data Analysis Of Starbucks' Global Presence [Imagen]. <https://nycdsa-blog-files.s3.us-east-2.amazonaws.com/2020/08/randy-pantinople/own2.png-623567-6OFPdu0B.png>

b. Herramienta de código abierto: Hadoop

Hadoop es una opción sólida para Starbucks debido a su capacidad para almacenar y procesar grandes volúmenes de datos de manera eficiente. Su sistema distribuido permite manejar datos de transacciones, preferencias de clientes y redes sociales, optimizando la personalización de ofertas y la eficiencia operativa.

En el **Anexo 3** se presenta información adicional sobre dos herramientas ampliamente utilizadas en el procesamiento de grandes volúmenes de datos: **IBM InfoSphere BigInsights** y **Hadoop**. A través de un cuadro comparativo, se detallan aspectos como el tipo de licencia, costos, facilidad de uso, capacidad de procesamiento, soporte técnico, funcionalidades clave, escalabilidad e integración. Esta información permite visualizar de manera clara y concreta las ventajas y desventajas de cada plataforma, facilitando una mejor comprensión de sus

aplicaciones y ayudando a evaluar su viabilidad según los recursos y objetivos de cada organización

Conclusión

En conclusión, el uso de Big Data se ha convertido en un pilar fundamental para las empresas que buscan adaptarse a los desafíos del entorno digital. Como se ha demostrado en los casos de Netflix y Starbucks, la correcta gestión de grandes volúmenes de datos permite no solo optimizar operaciones, sino también personalizar servicios y mejorar la experiencia del cliente. Este trabajo ha puesto de manifiesto cómo Big Data va más allá del almacenamiento de información, convirtiéndose en una herramienta estratégica para la toma de decisiones informadas y la innovación empresarial.

De manera complementaria, la implementación de arquitecturas robustas y herramientas especializadas permite que las organizaciones procesen, analicen y visualicen datos de manera eficiente. Tanto las herramientas licenciadas como las de código abierto ofrecen soluciones específicas que se adaptan a las necesidades y recursos de cada empresa. Por lo tanto, queda claro que la elección de herramientas debe estar alineada con los objetivos estratégicos y capacidades técnicas de cada organización.

Asimismo, es importante destacar que el éxito en la aplicación de Big Data no radica únicamente en la tecnología utilizada, sino en la capacidad de las empresas para convertir los datos en información valiosa. Esto implica un enfoque integral que abarque desde la identificación de fuentes hasta la visualización de resultados, asegurando que cada etapa contribuya al logro de los objetivos organizacionales. Este enfoque holístico fortalece la posición competitiva de las empresas en un mercado en constante evolución.

Finalmente, el análisis desarrollado en este trabajo subraya la importancia de adoptar tecnologías de Big Data para responder a las demandas cambiantes del mercado y anticiparse a las necesidades de los clientes. Con ello, se concluye que la gestión efectiva de los datos no solo es una ventaja competitiva, sino una necesidad para las organizaciones que aspiran a liderar en sus respectivas industrias. Big Data, por ende, no es solo una tendencia tecnológica, sino una revolución estratégica en el ámbito empresarial.

Bibliografía

- Barrantes, M. (2022). *Las 7 V del Big data: Características más importantes* - IIC. Instituto de Ingeniería del Conocimiento. <https://www.iic.uam.es/innovacion/big-data-caracteristicas-mas-importantes-7-v/>
- Castro, W. (2021). *Netflix Y Big Data: Claves Del Éxito Digital*. Pontia.tech. <https://www.pontia.tech/netflix-big-data-clave-para-su-exito/#:~:text=En%20este%20artículo,%20veremos%20cómo%20utiliza%20Netflix%20e%20I,Big%20Data%20para%20personalizar%20la%20experiencia%20del%20usuario?>
- Joyanes Aguilar, L. (2019). *Inteligencia de negocios y analítica de datos*. Alfaomega Grupo Editor Argentino.
- Nava, J. (2025). ¿Qué es el Big Data? <https://empowertalent.com/que-es-big-data/>
- Suárez, C. (2021, 19 de octubre). *Cómo es el análisis de datos de Starbucks: Big Data Analytics*. Marketing Insider Review. <https://marketinginsiderreview.com/analisis-de-datos-de-starbucks/>
- Vargas, A. (2023). *Big Data en Starbucks: Cómo la recopilación de datos impulsa el éxito de la cadena de cafeterías*. Conectando ideas. <https://conectandoideas.net/big-data-starbucks/>

Anexo

Anexo 1 "¿Qué es el Big Data?"

El artículo titulado sobre "¿Qué es el Big Data? De Nava, J. (2025): usos y beneficios" de Empower Talent ofrece una visión general sobre el concepto de Big Data, sus características principales y su relevancia en el entorno empresarial moderno. A continuación, presento un análisis detallado de los puntos clave abordados en el texto:

Resumen y análisis

1. **Definición de Big Data:** El artículo define Big Data como el manejo y análisis de grandes volúmenes de datos que no pueden ser procesados mediante métodos tradicionales. Este enfoque permite a las organizaciones extraer información valiosa para tomar decisiones más informadas y mejorar procesos. Este punto subraya la importancia de adoptar tecnologías avanzadas para gestionar datos complejos y aprovecharlos como ventaja competitiva.
2. **Las Tres V del Big Data:**
 - **Volumen:** Se destaca la capacidad de manejar cantidades masivas de datos generados por actividades digitales, dispositivos IoT y redes sociales. Esto resalta la necesidad de infraestructuras robustas para almacenar y procesar datos a gran escala.
 - **Velocidad:** El artículo enfatiza la rapidez con la que se generan y procesan los datos, lo cual es crucial para industrias que requieren análisis en tiempo real, como las finanzas y el comercio electrónico.

- **Variedad:** Se menciona la diversidad de formatos de datos (estructurados, no estructurados y semiestructurados), lo que refleja la complejidad de integrar diferentes tipos de información en un análisis coherente.
3. **Funcionamiento del Big Data:** El texto describe el proceso de recopilación, almacenamiento, gestión y análisis de datos provenientes de múltiples fuentes, como redes sociales y dispositivos IoT. Este enfoque integral permite a las empresas identificar tendencias y tomar decisiones estratégicas basadas en datos.
 4. **Importancia del Big Data:** El artículo concluye que el Big Data es una herramienta esencial para las empresas modernas, ya que les permite identificar oportunidades, resolver problemas y mejorar la eficiencia operativa. Esto refuerza la idea de que el análisis de datos no solo es una ventaja competitiva, sino también una necesidad en el entorno empresarial actual.

Reflexión

El artículo proporciona una introducción clara y concisa al concepto de Big Data, destacando sus aplicaciones prácticas y beneficios. Sin embargo, podría complementarse con ejemplos específicos de casos de éxito en diferentes industrias para ilustrar mejor su impacto real.

Anexo 2 Las 7 V de Big Data

La imagen presenta las siete dimensiones fundamentales que caracterizan el concepto de Big Data: Volumen, Velocidad, Variedad, Veracidad, Valor, Variabilidad y Visualización.

Estas dimensiones permiten entender cómo las organizaciones, como Netflix, gestionan grandes cantidades de datos para obtener información valiosa, tomar decisiones estratégicas y mantenerse competitivas en su industria. La representación gráfica facilita la comprensión del alcance y la complejidad del uso de datos a gran escala.

Ilustración 2: Las 7 V del Big Data



Fuente: Tomada de Sánchez Arevalo, J. (2024). 7 V DEL BIG

DATA [Imagen]. <https://thumbnails.genial.ly/6345deeccd4e360011d99e64/screenshots/49ce64cd-7ae2-4f3e-86b5-d027c8db2433.jpg>

Anexo 3 Comparativo: IBM InfoSphere BigInsights vs Hadoop

El cuadro muestra una comparación entre IBM InfoSphere BigInsights y Hadoop, destacando diferencias clave en licenciamiento, costo, facilidad de uso, escalabilidad y funcionalidades. Esta información complementa el análisis de herramientas para el manejo de Big Data y ayuda a identificar cuál se adapta mejor a distintos contextos empresariales.

Tabla 1 IBM vs Hadoop

Aspectos	IBM InfoSphere BigInsights	Hadoop
Tipo de licencia	Licenciada, requiere pago por uso o suscripción.	Código abierto, es gratuito y altamente personalizable.
Costo	Costoso, adecuado para empresas con alto presupuesto.	Económico, ideal para empresas con recursos limitados.
Facilidad de uso	Interfaz amigable, adecuada para usuarios con poca experiencia técnica.	Requiere conocimientos técnicos avanzados para su implementación y uso.
Capacidad de procesamiento	Alta capacidad, integra análisis avanzado y datos no estructurados.	Altamente escalable y eficiente, pero necesita configuraciones específicas.
Soporte técnico	Soporte profesional garantizado por IBM.	Comunidad activa de desarrolladores, pero sin soporte oficial.
Funcionalidad clave	Incluye herramientas de visualización integradas y capacidades de inteligencia empresarial.	Ideal para almacenar y procesar grandes volúmenes de datos de forma distribuida.

Escalabilidad	Alta, pero puede estar limitada por la licencia y costos adicionales.	Altamente escalable sin restricciones de licenciamiento.
Integración	Fácil integración con herramientas propietarias de IBM.	Integración con múltiples herramientas y plataformas de código abierto.

Fuente: Elaboración Propia.