

LA IMPORTANCIA DEL BIG DATA EN LA ACTUALIDAD

The Importance of Big Data Today

Oscar Alexander Mendoza Portillo^{1,2}

1 Facultad Multidisciplinaria Paracentral, Universidad de El Salvador, El Salvador

2 ORCID: 0009-0005-1045-8458

Correspondencia.

mp18035@ues.edu.s

RESUMEN

En la presente nota técnica, exploramos y evaluamos el término big data. El análisis empieza hablando del origen del término en cuestión, la relación con los valores de datos actuales y los modelos de uso tradicionales. En el manejo de grandes volúmenes de información en los sistemas. Continúa, con el concepto como el elemento central para entender el desarrollo del contenido, la naturaleza y los factores que ayudan al tratamiento de datos que son usados en distintos campos de la vida diaria. Esto incluye técnicas que destacan la importancia del big data en la sociedad actual y otorga una visión afirmativa de los datos para la toma de decisiones empresariales. Además, de discutir las dificultades clave para el futuro. El big data es fundamental hoy en día debido a la creciente y variada cantidad de datos que se producen. En consecuencia, se necesita la creación de nuevas tecnologías y sistemas para acumular y transformar de manera eficiente esta numerosa cantidad de información. Menciona los tipos de datos (estructurados, semiestructurados y no estructurados) reflejando las dificultades para obtenerlos y procesarlos. Los datos se obtienen de

distintos medios, por ejemplo; redes sociales, transacciones comerciales y dispositivos móviles.

ABSTRACT

In this technical note, we explore and evaluate the term big data. The analysis begins by discussing the origin of the term in question, the relationship to current data values, and traditional usage patterns. It continues, with the concept as the central element to understand the development of the content, the nature and the factors that help the processing of data that are used in different fields of daily life. This includes techniques that highlight the importance of big data in today's society and provides an affirmative view of data for business decision making. In addition, to discuss the key difficulties for the future. The big data is essential today due to the growing and varied amount of data that is produced. Consequently, the creation of new technologies and systems is needed to efficiently accumulate and transform this large amount of information. Mentions the types of data (structured, semi-structured and unstructured) reflecting the difficulties in obtaining and processing them. Data is obtained from different means, for example; social networks, commercial transactions and mobile devices.

Palabras clave: big data, volumen de datos, datos, trazabilidad, innovación, desafíos tecnológicos.

INTRODUCCIÓN

La sociedad cambia día a día evoluciona en todas las áreas de la vida de las personas es por eso que el crecimiento en la cantidad de datos que se generan a través de los diferentes sistemas informáticos y actividades cotidianas de la sociedad moderna ha revolucionado la manera de procesar los datos dada la necesidad de adaptarse a la realidad cambiante y dinámica con el fin de transformar, optimizar y generar métodos y modelos que permitan almacenar y procesar datos de la mejor manera capaces de afrontar las deficiencias de las bases de datos tradicionales ya existentes. De esta manera, el big data surge como una alternativa tecnológica que permite gestionar grandes cantidades de datos que provienen de diferentes fuentes.

Este fenómeno ha generado un impacto positivo en los distintos sectores, desde la medicina hasta el comercio minorista, creando así la necesidad de hacer uso de nuevas tecnologías y enfoques que ayudan a gestio-

nar y analizar los datos masivos que se generan a diario. Bajo estas circunstancias es indispensable conocer y entender la importancia que tiene el big data actualmente y su influencia al momento de tomar decisiones de peso en las empresas.

Basado en una revisión de una serie de fuentes bibliográficas que contienen información del big data, acerca del origen, valor de los datos, tipos de datos, conceptualización y las fuentes de las que provienen los datos, características y elementos, el uso del big data en los tiempos actuales, así como las diversas aplicaciones y herramientas utilizadas para el procesamiento de los datos y los principales desafíos que enfrentamos hoy en día, tiene como objetivo ofrecer una visión general del big data.

Desarrollo del tema

1. Origen

Explorando los orígenes del big data: Un vistazo al pasado

Desde tiempos antiguos han existido diversas formas de producir y procesar información, basta con echar un vistazo a los primeros sistemas de datos informáticos, su avance está relacionado a la necesidad de tener un mejor control sobre la información y puntos de referencia para tomar decisiones acertadas que permitan obtener beneficios.

En la actualidad, el big data se ha convertido en un concepto ampliamente utilizado en el ámbito tecnológico y empresarial. Sin embargo, pocos conocen su origen y cómo ha evolucionado a lo largo del tiempo. Para comprender mejor este fenómeno, es necesario remontarnos al pasado y explorar el surgimiento. El big data surge de la necesidad de gestionar grandes volúmenes de información.

En la era digital en la que vivimos, se generan enormes cantidades de datos a diario. Estos datos provienen de diferentes fuentes, como redes sociales, transacciones comerciales, dispositivos móviles, sensores, entre otros. La necesidad de gestionar y analizar esta gran cantidad de información dio origen al concepto de big data. El término " big data " se popularizó a principios

de la década de 2000, pero sus raíces se remontan mucho más atrás en el tiempo. En la década de 1960, IBM introdujo el término "Information Explosion" para describir el crecimiento exponencial de la información en ese momento. Sin embargo, fue en la última década cuando se comenzó a hablar de big data como una disciplina en sí misma (Rodríguez, 2023).

El big data no solo se refiere a la cantidad de datos generados, sino también a la velocidad a la que se generan y a la variedad de fuentes. Estos tres conceptos, conocidos como las "3V del big data" (volumen, velocidad y variedad), son fundamentales para comprender su origen y su importancia. En sus inicios, el big data se centraba principalmente en el almacenamiento y la gestión de grandes volúmenes de datos.

Las soluciones tradicionales de bases de datos no eran suficientes para manejar la enorme cantidad de información generada. Fue entonces cuando surgieron tecnologías como Hadoop, que permitían el procesamiento distribuido de datos en clústeres de computadoras. Con el avance de la tecnología, el big data ha evolucionado y se ha convertido en una rama multidisciplinaria.

Figura 1

Representación del big data



Nota: Imagen elaborada a partir de inteligencia artificial en Microsoft Designer.

2. Big data en contexto

Cuando nos referimos al big data, estamos hablando de un antes y un después. Por el hecho que supone

una revolución en tecnología. En todo caso, este tema es visto como una solución para las empresas por la cantidad abrumadora de datos que generan día con día y surgen interrogantes sobre cómo poder sacar el máximo provecho a la información. En otras palabras, analizar grandes colecciones de datos puede proporcionar conocimientos que no se pueden obtener examinando sólo una pequeña cantidad.

Las organizaciones notaron esto hace más de 25 años, cuando por primera vez la minería de datos fue integrada como disciplina a las prácticas de inteligencia empresarial. Sin embargo, en aquella época había que hacer funcionar de la manera correcta dos factores claves e importantes: el primero trataba acerca de la cultura organizacional, que era conservadora y no estaba dispuesta a integrar elementos externos al proceso de toma de decisiones. El segundo era el estado de la tecnología, por tanto, la minería de datos se realizaba tomando pequeñas muestras debido a que las aplicaciones no escalaban, debido a esto no había volúmenes grandes de datos no estructurados, ni servicios comerciales maduros para poder procesarlos, los expertos en modelización eran escasos, además no existía una infraestructura accesible y rentable. Quizás lo más propio sería hablar de big data como un fenómeno que se da alrededor de 2010 cuando convergen, en un orden por determinar, cuatro factores importantes:

Tabla 1

Factores Clave de Nuevos Modelos de Negocio

Factor	Contexto
La aparición y consolidación de nuevos modelos de negocio.	Basados en la disponibilidad y variedad de datos que hay en internet, creando a su vez más datos sobre los que se realimentan. El auge del internet de las cosas, con la explosión de dispositivos y sensores conectados en red, la generalización del acceso y participación en las redes sociales o la preponderancia del comercio electrónico son algunos ejemplos.
Un cambio en la cultura empresarial	Conducido por una nueva generación de profesionales de formación interdisciplinaria y habituada a la tecnología.
Un avance tecnológico	Girando en torno a la computación en la nube (cloud computing), proporciona nuevas funcionalidades y capacidades con unos modelos de consumo muy flexibles y con costes asequibles.
Una oportunidad de negocio en sí mismo	explotada por los departamentos de marketing de las grandes corporaciones tecnológicas.

(Fandiño, 2023),

3. El valor del dato y sus tipos

Resulta importante aclarar que los datos son en verdad la razón de ser del big data. Dado que, estos brindan una perspectiva más amplia del mismo. Entidades gubernamentales y no gubernamentales, medios informativos y las personas naturales deben comprender que la información y el buen uso de esta permite tomar decisiones conscientes y más efectivas al momento de aplicarlas. Se puede citar como ejemplo: las nuevas Smart cities o ciudades inteligentes, que pueden tomar datos de tráfico, tiempo, polución y ayudar a los ciudadanos a moverse de forma más segura por sus calles, gracias a una eficaz toma de decisiones basadas en la adquisición de datos.

Las organizaciones deben invertir en infraestructuras que trabajen con datos y permitan el continuo acceso a los mismos prácticamente en tiempo real. Además, permiten crear patrones de comportamiento de los clientes y así anticiparse a las necesidades de los mismos. Gracias al data science, los datos son la base para crear nuevas líneas de investigación que, si no fuera por la posibilidad de comparar grandes cantidades de información mediante software, no seríamos capaces de descubrir gran parte de avances tecnológicos de hoy en día, haciendo especial hincapié en los avances médicos que se han producido gracias a los datos(Romero, 2019).

TIPOS DE DATOS

Cabe destacar que existen distintos tipos de datos, los cuales representan la información de la cual hace uso el Big Data ya que los datos son elementos que necesitan de un procesamiento a través de distintas técnicas informáticas. A continuación, se da a conocer los diferentes tipos de datos existentes.

DATOS ESTRUCTURADOS

Se puede mencionar que particularmente gran parte de fuentes de información tradicional se toman como datos estructurados, es decir que poseen un formato normalizado con campos definidos que no pueden ser reemplazados.

Son los datos de las bases de datos relacionales, las hojas de cálculo y los archivos, fundamentalmente. Los datos estructurados se componen de piezas de información que se conocen de antemano, vienen en un formato especificado, y se producen en un orden especificado. Estos formatos facilitan el trabajo con dichos datos. Formatos típicos son: fecha de nacimiento (DD, MM, AA); documento nacional de identidad o pasaporte (por ejemplo, 8 dígitos y una letra); número de la cuenta corriente en un banco (20 dígitos), etcétera. Datos con formato o esquema fijo que poseen campos fijos(Romero, 2019).

DATOS SEMIESTRUCTURADOS

Este tipo de datos tienen la peculiaridad de no poseer un esquema definido, pero tienen ciertas características ya que no tienen formatos fijos, pero pueden contener etiquetas y otros marcadores que permiten que el usuario pueda separar los elementos de los datos.

La lectura de datos semiestructurados requiere el uso de reglas complejas que determinan cómo proceder después de la lectura de cada pieza de información. Un ejemplo típico de datos semiestructurados son los registros Web logs de las conexiones a Internet. Un Web log se compone de diferentes piezas de información, cada una de las cuales sirve para un propósito específico. Ejemplos típicos son el texto de etiquetas de lenguajes XML y HTML(Aguilar, 2016).

DATOS NO ESTRUCTURADOS

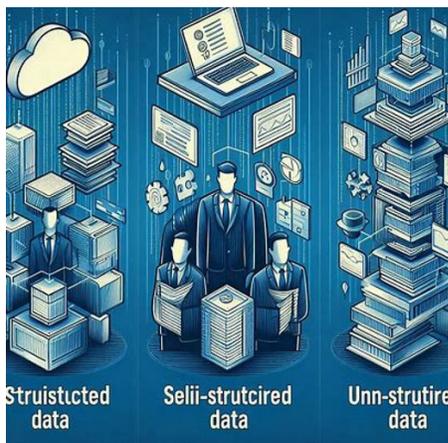
Los datos no estructurados son datos sin tipos predefinidos. Se almacenan como “documentos” u “objetos” sin estructura uniforme, y se tiene poco o ningún control sobre ellos. Se puede mencionar algunos ejemplos; datos de texto, video, audio, fotografías, documentos impresos, cartas, hojas electrónicas entre otros(Aguilar, 2016).

Una gran parte de la información que producen las organizaciones no reside en bases de datos relacionales o archivos de datos, estos se encuentran divididos en

todas las áreas que posee la organización, a este tipo de datos se la considera no estructurados. Por el hecho de que son datos difíciles de manipular o dominar por las personas encargadas de analizarlos. Pero gracias a su crecimiento se han generado y creado nuevas herramientas para su manipulación óptima y precisa como lo son Hadoop, MapReduce, bases de datos NoSQL, etc.

Figura 2

Representación de los tipos de datos



Nota: Imagen elaborada a partir de inteligencia artificial en Microsoft Designer.

4. Conceptualización y fuentes ¿QUÉ ES BIG DATA?

Resulta importante dejar en claro dicho término ya que esto servirá de referencia para comprender lo que se está estudiando. Se puede decir que el big data trata sobre almacenamiento masivo de información, en otras palabras, sirve para indicar que los datos no pueden ser controlados con herramientas tradicionales dado que requieren un tratamiento más riguroso con personas y herramientas especializadas.

Por ello una definición instrumental de big data es la siguiente: el conjunto de técnicas y herramientas que permiten manipular enormes cantidades de datos con el fin de ayudar a las organizaciones a resolver problemas y facilitar el uso inteligente de la información ne-

cesaria para la toma de decisiones(Baldominos, 2020).

FUENTES DE BIG DATA

Es necesario mencionar que son muchas las distintas fuentes de las cuales se está produciendo información a grandes volúmenes en la sociedad actual. Por ejemplo: páginas web, comercio electrónico, redes sociales e Internet de las cosas (IoT). Dado que Internet de las Cosas facilita grandemente la comunicación entre dispositivos, lo que se conoce como M2M (Machine-to-machine), aumentando la interconexión y la transmisión de datos entre ellos.

Hoy en día se genera una cantidad ingente de datos: desde los documentos informáticos que colgamos en Internet hasta todas nuestras interacciones en las redes sociales, pasando por nuestras compras en la Red y los datos GPS de los vehículos(Holmes, 2022).

Figura 3

Representación de fuentes del big data



Nota: Imagen elaborada a partir de inteligencia artificial en Microsoft Designer.

5. Características y elementos

CARACTERÍSTICAS

Se considera que al hablar de big data es importante mencionar las características y los distintos elementos con los cuales cuenta ya que a través de estos se da la pauta para poder aplicarlos acorde a las necesidades de las organizaciones independientemente al rubro al que pertenezcan. El big data se caracteriza por su Volumen, Velocidad, Dirección, Variedad, Veracidad y Valor. Involucra grandes cantidades de datos generados rápidamente, flujos entre entornos, diversidad de datos no estructurados, importancia de la certeza de los datos y la conversión de datos en información para la toma de decisiones.

Volumen.

La búsqueda de nuevas oportunidades de negocio y la necesidad de diferenciarse de la competencia, hace que la cantidad de datos que hay que analizar para obtener resultados sea cada vez más grande (Fandiño, 2023).

Velocidad.

En términos de big data, la velocidad presenta dos aspectos. El primero hace referencia a la celeridad con que se producen los datos ya que la generación de contenido es un continuo que no se detiene nunca. El segundo aspecto pone el foco en la prontitud con la que los datos deben ser procesados y analizados para sacarles provecho. La combinación de ambos hace, en definitiva, que los flujos de datos que atraviesan las empresas no solo sean voluminosos, sino constantes. Esto supone un reto en términos de latencias de acceso y velocidades de respuesta, pero también en lo referente a la seguridad(Fandiño, 2023).

Variedad.

Al parecer la mayoría de los datos que poseen las organizaciones son no estructurados, por ejemplo, co-

reos electrónicos, mensajes, manuales, transcripciones de audio, video e imágenes, etc. En algunos casos o sectores este tipo de contenido puede ser muy relevante. El ámbito legal puede ser uno de los extremos, con lo que representa las leyes y jurisprudencia, el ámbito de la salud no se queda lejos con toda la literatura médica, los historiales e investigaciones clínicas, etc. de aquí de hablar de variedad de información o datos de distintas fuentes sobre temas diferentes(Fandiño, 2023).

Veracidad.

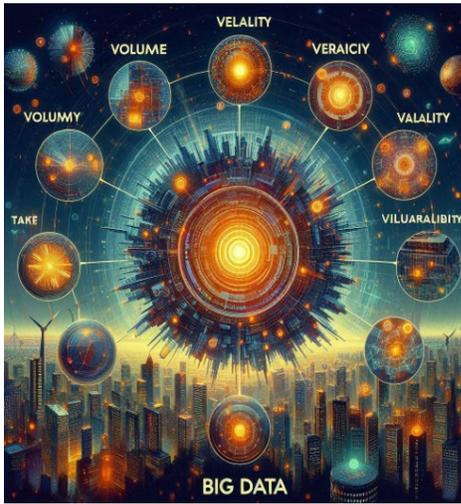
Este apartado hace referencia a la fiabilidad de la información generada. Este es un aspecto importante. Cuando los datos no son confiables, las decisiones están en riesgo. Los problemas de calidad de los datos no son nuevos, pero las empresas todavía parecen tener dificultades para reconocer su importancia. Si bien el volumen y la velocidad hacen que la precisión sea difícil de controlar, la diversidad de origen es quizás el factor más complicado, ya que requiere un mayor énfasis en las tareas de comparación, limpieza e integración de datos. En este sentido, el aseguramiento de la calidad se basa en capturar y documentar el origen de los datos, las transformaciones que sufren y su destino, así como la trazabilidad de los datos y el análisis de su impacto.

Valor.

Esta última característica es la que justifica el porqué de todas las anteriores, ya que al final del día el objetivo que se busca obtener a través del tratamiento y procesamiento de datos es convertirlos en información consistente que permita desarrollar conocimientos y poder así tener la base para tomar decisiones acordes a la empresa. Implica conocer y comprender al usuario, creando nuevos productos y servicios para satisfacer sus necesidades, retener y fidelizar a los clientes optimizando los procesos y disminuyendo costos y a su vez aumentando el margen de ganancias.

Figura 4

Representación de las 7 V del big data



Nota: Imagen elaborada a partir de inteligencia artificial en Microsoft Designer.

ELEMENTOS

Tabla 2

Elementos Clave en la Gestión de Datos

Elementos	Concepto
Collection (recogida)	Una de las mayores dificultades a la hora de disponer de los datos es cómo conseguirlos.
Storage (almacenamiento)	Una vez han sido obtenidos, hay que determinar cómo almacenarlos de la manera más óptima para su gestión y posterior consulta.
Research (investigación)	La información que se pretende extraer de los datos debe ser parte de un proceso de investigación y de mejora continua para el descubrimiento de nuevas capacidades
Analysis (análisis)	Para que de los datos se pueda extraer una información valiosa, deben ser analizados
Volumen	Se habla de Big Data y no de otras variaciones cuando se incluye un componente de volumen y complejidad
Visualización	Para su mejor comprensión y sobre todo, de cara a poder orientar y convencer a los actores decisivos de una empresa, es imprescindible una visualización amigable del resultado del análisis.
Cloud technology (tecnología en la nube)	Los datos deben estar disponibles para su consulta por distintos agentes en cualquier momento y desde distintas ubicaciones, además del hecho de que tener externalizados servicios en la nube tiene ventajas adicionales para una empresa.
Network (red)	Se trata de la infraestructura física que sustenta el punto anterior

(Fandiño, 2023)

6. Importancia de su utilidad hoy en día

La razón por la cual el big data es tan importante y útil para muchas empresas actuales es porque genera respuestas a diferentes interrogantes que las organizaciones no saben que tienen. Es decir, sirve como punto de referencia. Ya que la cantidad de información generada es tan grande los datos se pueden moldear o probar de la forma en que la empresa considere conveniente. Al llevarlo a cabo, las empresas identifican las necesidades o problemas de una forma más comprensible. La recolección de gran cantidad de datos exhibe tendencias en estos y permite a las empresas operar de manera más rápida, fluida y efectiva. También les permite eliminar los problemas antes de que destruyan sus ganancias o su reputación

Analizar big data proporciona un apoyo a las organizaciones en el sentido de que al aprovechar los datos y darles el uso adecuado genera consigo nuevas oportunidades. Conduciendo a nuevas ideas de negocios inteligentes, con operaciones más ágiles y eficaces, mayores beneficios y usuarios satisfechos.

Las empresas con más éxito con big data consiguen valor de las siguientes formas:

Tabla 3

Beneficios de la Implementación de Big Data

Reducción de coste.	Las grandes tecnologías de datos, como Hadoop y el análisis basado en la nube, aportan importantes ventajas en términos de costes cuando se trata de almacenar grandes cantidades de datos, además de identificar maneras más eficientes de hacer negocios.
Más rápido, mejor toma de decisiones.	Con la velocidad de Hadoop y la analítica en memoria, combinada con la capacidad de analizar nuevas fuentes de datos, las empresas pueden analizar la información inmediatamente y tomar decisiones basadas en lo que han aprendido.
Nuevos productos y servicios.	Con la capacidad de medir las necesidades de los clientes y la satisfacción a través de análisis viene el poder de dar a los clientes lo que quieren. Con la analítica de Big Data, más empresas están creando nuevos productos para satisfacer las necesidades de los clientes.

(big data , s. f.)

7. Campos de Aplicación

Un análisis esclarecedor sobre uno de los grandes temas de nuestro tiempo, y sobre el inmenso impacto que tendrá en la economía, la ciencia y la sociedad en general. Los datos masivos representan una revolución que ya está cambiando la forma de hacer negocios, la sanidad, la política, la educación y la innovación(Mayer-Schönberger & Cukier, 2016).

A medida que la tecnología avanza y se desarrolla, las aplicaciones del big data son cada vez más complejas y variadas. Desde la informática, el marketing, las leyes, negocios, etc, el big data es usado en muchos sectores sea político, social, ambiental, económico, entre otros, revolucionando la forma en la que estos realizan sus operaciones.

Las tecnologías del big data pueden emplearse para crear una zona de preparación o zona de aterrizaje para los nuevos datos antes de identificar qué datos se deben mover al almacén de datos. Además, la integración de tecnologías del big data con el almacén de datos permite que una organización descargue los datos a los que se accede con poca frecuencia(¿Qué es el big data y cuáles son sus beneficios?, 2020).

Actualmente la implementación del big data demues-

tra que es una técnica invaluable para poder manipular y analizar datos de grandes cantidades de información. Las organizaciones utilizan las capacidades del big data para encontrar modelos y tendencias en los datos obtenidos. Esto sirve para que puedan predecir comportamientos a futuro. Seguidamente se hace mención de algunos campos de aplicación del big data:

1. Big data en el comercio minorista

La competencia en la industria minorista es muy feroz y los minoristas están esforzándose continuamente por lograr una ventaja competitiva sobre los demás, y para prosperar, es importante que los minoristas comprendan muy bien a sus clientes. Tener conciencia de las necesidades de los clientes y de cómo gestionarlas de forma óptima. Satisfacerlos le dará a la empresa una ventaja competitiva. Asimismo, al realizar análisis avanzado de los datos de sus clientes, los minoristas podrían comprender completamente a sus clientes. Los datos de los clientes se pueden obtener a través de muchos recursos, incluidas las redes sociales, programas de fidelización, etc.

Los detalles acercan lo más posible a los minoristas a sus clientes. En consecuencia, los minoristas podrían ofrecer a sus clientes más servicios personalizados y también prever sus demandas futuras. Cliente leal así se puede establecer. Costco, Walmart, Walgreens y Sears and Holdings se encuentran entre los minoristas que utilizan mucho big data.

2. Big data en la atención sanitaria

Big data facilita enormemente la industria de la salud como esta industria constantemente tiene que lidiar con una gran cantidad de datos. Tal cantidad de datos ha hecho que sea bastante imposible para los profesionales de la salud aprovecharlos.

El uso de big data puede considerarse como una salvación, ya que facilita profesionales e investigadores de esta industria para detectar y curar enfermedades como el cáncer. También a través del big data y la analítica, más personalizados se pueden establecer medicamentos y se pueden proporcionar tratamientos más efectivos a los pacientes. Además, se pueden identificar patrones únicos de ciertos medicamentos, permitiendo el desarrollo de soluciones más rentables. También, los dispositivos portátiles inteligentes se están volviendo populares entre todas las personas, ya que estos dispositivos pueden ayudar a salvar vidas mediante la producción de alertas en tiempo real.

3. Big data en el comercio electrónico

El comercio electrónico ha sido considerado una revolución notable en esta era y se ha convertido en una parte integral de la vida de las personas de hoy. Por lo tanto, es común que la gente piense en el comercio electrónico cuando quiere comprar algo. En este sentido, Amazon, Flipkart y Alibaba se encuentran entre las más destacadas empresas globales de comercio electrónico y el uso de big data en estas empresas es extenso. De manera relevante, Amazon, como la empresa de comercio electrónico más grande del mundo, es uno de los líderes en big data y análisis.

4. Grandes datos en la educación

En el ámbito de la educación, los datos son generalmente importantes para el futuro. Por tanto, los datos son muy importantes en este ámbito. Los datos mejoran enormemente el sistema educativo, revitalizando específicamente las habilidades, tanto académicas como no académicas. Además, el uso de big data facilita la evaluación del desempeño de estudiantes y docentes. Del mismo modo, big data se puede utilizar para rastrear la tasa de deserción escolar y luego en determinar las medidas más adecuadas para reducirlo.

5. Big data en finanzas

Las organizaciones financieras dependen en gran medida de los datos en sus operaciones y de hecho, para este tipo de organizaciones, los datos son su segundo bien más vital después del dinero. Debido a tal importancia, las organizaciones financieras deben buscar la seguridad de sus datos, lo cual es una tarea desafiante. De hecho, las empresas financieras están entre los primeros en adoptar big data. La Banca digital y los pagos han estado entre las palabras de moda más populares con big data como su elemento importante. En las empresas financieras, se manejan los principales dominios que incluyen comercio algorítmico, detección de fraude, análisis de riesgos y satisfacción del cliente. Con big data el sistema financiero se vuelve fluido, mejorado y empoderado para poner a disposición servicios superiores a los clientes.

6. Big data en la industria de viajes

En el contexto de la industria de viajes, la difusión del big data ha sido bastante lenta en comparación con otras industrias. En este sentido, las empresas de viajes pueden utilizar el análisis de big data para ofrecer a sus clientes una experiencia de viaje más personalizada. Igualmente, las empresas pueden utilizar big data Analytics para comprender mejor las necesidades de sus clientes, requerimientos, brindándoles las mejores ofertas y sugerencias en tiempo real. El big data puede convertirse en una excelente guía para cualquier viajero, convirtiéndolo en una parte cada vez más vital en esta industria.

7. Big data en telecomunicaciones

La industria de las telecomunicaciones ha sido el principal impulsor de la digitalización global. Mientras tanto, el creciente uso de teléfonos inteligentes en todo el mundo ha provocado que la industria de las teleco-

municaciones se vea inundada de datos de cantidades colosales. Para las empresas de telecomunicaciones, estos datos son muy valiosos y, por lo tanto, es crucial que las empresas sepan cómo explotar adecuadamente estos datos. El uso de big data Analytics permite a estas empresas de telecomunicaciones ofrecer a los clientes conectividad fluida, y esto facilita a los clientes ya que las barreras de la red son erradicadas. También permite a las empresas realizar un seguimiento tanto de las áreas de mayor y menor tráfico y luego decidir las mejores acciones para Proporcionar a los clientes una conectividad de red fluida. Con la apropiada comprensión de los clientes mediante el uso de big data, se puede ofrecer servicios personalizados.

8. Big data en la industria del automóvil

La industria del automóvil está totalmente controlada por el big data, lo que da resultados extraordinarios y novedosos. Se puede decir así que big data es ahora una parte vital en la industria del automóvil, ya que ha llevado a logros inimaginables. Con big data, la industria podría analizar tendencias, comprender la gestión de la cadena de suministro, brindar atención a los clientes, hacer posible lo imposible, etc, (Al-Khasawneh, 2020).

8. Herramientas de aplicación

En la era de grandes conjuntos de datos, procedentes de diversos orígenes, en formatos variados y con una necesidad de procesamiento y análisis rápido y efectivo, las técnicas de big data persiguen complementar el manejo ordenado de estos volúmenes, con las técnicas de análisis de la información más avanzadas y efectivas para extraer de modo óptimo el conocimiento contenido en los datos(MARQUÉS, 2015).

Big data hace uso de distintas herramientas o aplicaciones las cuales son de gran utilidad para poder obtener resultados ágiles y veraces que permiten que las empresas puedan hacer uso de ellas para el análisis

de datos, generando información de provecho y así poder tener la pauta o referencia para tomar decisiones congruentes y acertadas de beneficio para el negocio. A continuación, se mencionan algunas de las herramientas de las que hace uso el big data:

1. Apache hadoop

Apache Hadoop es una herramienta muy popular en la industria de big data y almacena procesos y analiza big data. Hadoop, que está escrito en Java, es un Marco de código abierto de Apache y opera en hardware básico. El uso de Apache Hadoop permite procesar datos de forma paralela porque Apache Hadoop puede ejecutarse simultáneamente en varias máquinas.

2. Apache Spark

Abarca un sistema de agrupación en clústeres de propósito general y puede considerarse como el reemplazo de Hadoop. Apache Spark resuelve los problemas relacionados con Hadoop, ya que admite el procesamiento por lotes y en tiempo real, mientras que también admite cálculos en memoria. Además, tiene una cantidad mucho menor de operaciones de lectura/escritura en el disco, lo que lo hace mucho más rápido que Hadoop, concretamente 100 veces más rápido para el procesamiento de datos.

3. Tormenta apache

Apache Storm es una herramienta de big data de código abierto y supone una distribución de

sistemas de procesamiento en tiempo real y tolerante a fallas, con la capacidad de procesar flujos ilimitados de datos de manera eficiente. En este contexto, se relacionan con datos en constante expansión que tienen un comienzo, pero no un final definido. Apache Storm se puede utilizar en cualquier lenguaje de programa-

ción y también admite protocolos basados en JSON. Storm asegura el procesamiento de todos los datos y tiene una velocidad muy rápida procesando hasta un millón de tuplas procesadas por segundo en cada nodo. Storm también es tolerante a fallos y fácilmente escalable, lo que hace que esta herramienta sea mucho más sencilla. en términos de uso(Al-Khasawneh, 2020).

4. Apache Cassandra

Apache Cassandra es una base de datos distribuida y ofrece alta disponibilidad. y escalabilidad mientras el rendimiento sigue siendo eficiente. Es considerada como una herramienta de big data (una de las mejores) con la capacidad de adaptar conjuntos de datos de todo tipo, estructurados, semiestructurados y no estructurados. Esta herramienta se considera la plataforma más adecuada para datos de misión crítica sin punto de falla. Además, tolera fallas tanto en el hardware básico como en infraestructura en la nube(Al-Khasawneh, 2020).

5. MongoDB

Como herramienta de análisis de datos de código abierto, MongoDB abarca un NoSQL base de datos que ofrece capacidades multiplataforma. Esas capacidades podrían atender a empresas que necesitan datos rápidos y en tiempo real para la toma de decisiones. Esta herramienta es útil para quienes buscan soluciones basadas en datos. MongoDB que está escrito en JavaScript y C, C ++, es confiable y rentable(Al-Khasawneh, 2020).

6. Apache Flink

Apache Flink, que está escrito en Java y Scala, implica un código abierto, es una herramienta de análisis de datos marco de procesamiento distribuido para flujos de datos, acotado e ilimitado. Incluso para los datos que llegan tarde, Apache Flink proporciona resultados altamente precisos. Apache Flink puede recuperarse

fácilmente de fallas y por lo tanto es tolerante a fallos. A gran escala, Flink ofrece alto rendimiento y eficiencia en miles de nodos. Otras ventajas de Apache Flink incluyen motor de transmisión de baja latencia y alto rendimiento y soporte en el momento del evento y gestión estatal(Al-Khasawneh, 2020).

7. Kafka

Introducido por LinkedIn en 2011, Apache Kafka implica un evento distribuido plataforma de código abierto de procesamiento o transmisión que proporciona un alto rendimiento a los sistemas. Esta herramienta tiene la capacidad de manejar billones de eventos diariamente, además de ser tolerante a fallos. Como plataforma de streaming, Apache es muy escalable, y entre los procesos de streaming están publicar y suscribirse a flujos de registros que son similares a los sistemas de mensajería. Estos registros se almacenan permanentemente en grupos llamados temas y luego se procesan. Apache Kafka ofrece streaming de alta velocidad sin tiempo de inactividad(Al-Khasawneh, 2020).

8. Cuadro

Tableau está considerada una de las mejores herramientas de visualización de datos y Solución de software dentro de la industria del Business Intelligence. El cuadro gira datos sin procesar en conocimientos importantes y al mismo tiempo mejora la toma de decisiones en el proceso de los negocios(Al-Khasawneh, 2020). El proceso de análisis de datos proporcionado por Tableau es rápido, lo que da como resultado visualizaciones en forma de paneles interactivos y hojas de trabajo. Además, Tableau permite la mejor combinación de datos del mercado y también un análisis eficiente en tiempo real. Como parte vital en ciertas industrias, Tableau también está ligado a la industria de la tecnología. Además, correr Tableau no requiere habilidades técnicas o de programación(Al-Khasawneh, 2020).

9. Minero rápido

Rapid Miner, que es una herramienta de código abierto escrita en Java, es una herramienta multiplataforma. Ofrece un entorno sólido para aprendizaje automático y datos. Procedimientos de ciencia y análisis de datos. RapidMiner es una plataforma unificada para el ciclo de vida completo de la ciencia de datos, y comienza desde la preparación de los datos hasta la máquina aprendizaje para el empleo del modelo predictivo (Al-Khasawneh, 2020).

10. Programación R

Como lenguaje de programación de código abierto, R se encuentra entre los más inclusivos de análisis estadístico. R, que está escrito en C y Fortran, abarca un lenguaje de programación multiparadigma que ofrece una dinámica entorno de desarrollo, y ser un proyecto de código abierto ha permitido innumerables personas participar en su desarrollo. Un paquete enorme El ecosistema es ofrecido por R, lo que lo convierte en una de las herramientas más populares de análisis estadístico. R ayuda al desempeño eficiente de diversas estadísticas operaciones y al mismo tiempo facilita la producción de resultados de análisis de datos tanto en texto como en formato gráfico. R ofrece gráficos y tablas notables (Al-Khasawneh, n.d.).

9. Principales desafíos

Como todo avance tecnológico en desarrollo, big data no es la excepción ya que expone desafíos relacionados a los distintos factores que abarca, comenzando por hacer innovaciones a las infraestructuras, además a la forma de pensar de los especialistas desarrolladores acostumbrados a usar tecnologías tradicionales como information retrieval y data mining, empleando prácticas habituales de desarrollo, hasta llegar al punto de descifrar que tipo de datos son los correctos para indagar información para este tipo de implementaciones. Entre los desafíos más frecuentes se puede mencionar los siguientes:

Skills: Este problema trata básicamente la capacidad de las personas a cargo del manejo de la información recolectada. Al ser una tecnología en desarrollo, la

cantidad de personas que tengan el “know how” o conocimiento para poder procesar de manera correcta el volumen de información es relativamente poco, lo que dificulta el desarrollo de proyectos (Candel, 2022).

Estructura de datos: Otro gran desafío es la forma en la que se guardan los datos. La forma misma en que tenemos concebida la idea de cómo guardar los datos en la actualidad presenta un desafío enorme para big data. El desafío de hoy es que la mayoría de los almacenes de datos empresariales ven un cliente o una entidad que la empresa trabaja con una fila de datos en lugar de una columna. Esa fila se rellena y se actualiza quizás a diario con la instantánea o al agregado de la situación actual del cliente. Al realizar esta actualización, estamos perdiendo la información recolectada, lo que conlleva a menor capacidad de predicción o información a procesar (Candel, 2022).

La tecnología: Lo interesante es que Hadoop es ideal para el procesamiento por lotes a gran escala, que es como las operaciones de agregación o cómputo. El problema es que Hadoop no es una tecnología en tiempo real o muy dinámica en absoluto. La ejecución de consultas en un clúster Hadoop suele tener una gran latencia ya que hay que distribuir cada consulta individual, luego, hacer su etapa de reducción, que está trayendo todos los datos de nuevo juntos. Así que es una tecnología de alto rendimiento, pero de alta latencia (Candel, 2022).

Privacidad: Junto con la obtención de volúmenes de datos incalculables, viene una cantidad de datos que podríamos considerar intrusiva, podría darse ejemplos como Facebook, Twitter, Google que manejan grandes volúmenes de datos de clientes, con esta capacidad de Big Data de intentar analizar absolutamente todo, podría darse una examinación inapropiada de los datos de usuarios, conllevando rupturas en la privacidad de los datos de los usuarios. (Si bien esta problemática no es nueva, podría agravarse con la capacidad avanzada de procesamiento que se obtiene con big data (Candel, 2022).

Volumen, Variedad, Velocidad: La capacidad de encontrar un equilibrio entre todas ellas depende de la capacidad de plantear un desarrollo sustentable y un plan acorde a las posibilidades tecnológicas de la empresa que desarrolla con esta tecnología (Candel, 2022).

A nivel técnico, la adopción de tecnologías big data supone una serie de desafíos entre los que podemos destacar:

- El análisis de datos estructurados es necesario para comprender los métodos de análisis de big data, incluso existen métodos que se comparten con el análisis convencional, pero con muchos más datos.
- La administración de bases de datos es un fundamento para el análisis de datos y para manejar datos operacionales. En big data, las bases de datos son una fuente importante que alimenta el núcleo de procesamiento.
- La programación orientada a objetos es el pilar para desarrollar cualquier tipo de aplicación, incluso para manejar bases de datos. El big data se utiliza para manejar y procesar distintos tipos de datos.
- La administración de servidores es necesaria para aprovechar al máximo las tecnologías de la información. En big data son primordiales pues son el soporte de toda la infraestructura de aprovechamiento de los datos masivos (Candel, 2022).

Figura 5

Representación acerca de los desafíos del big data



Nota: Imagen elaborada a partir de inteligencia artificial en Microsoft Designer.

Conclusiones

Big data está transformando la manera en que las organizaciones llevan a cabo el análisis de grandes volúmenes de datos, generando conocimiento crucial y determinante para la toma de decisiones estratégicas fundamentado en datos. Esta revolución en el procesamiento de datos permite a las empresas acomodarse con flexibilidad a un ambiente empresarial cada vez más competitivo y dinámico.

La gran importancia del big data radica en su capacidad de ofrecer información detallada y de alta calidad sobre los usuarios, las operaciones y los patrones comerciales, lo que genera ganancias sustanciales. El poder de profundizar en big data para obtener información valiosa permite a las empresas descubrir oportunidades de crecimiento, pronosticar cambios e incluso mejorar la eficiencia operativa.

Además de ofrecer beneficios, el big data de igual modo expone retos importantes en las áreas de privacidad de datos, integridad de la información y la necesidad de respaldar la calidad y credibilidad de la información que se analiza. La puesta en marcha de diversas medidas de seguridad y cumplimiento es pri-

mordial para la protección de los datos garantizando la confianza de los usuarios y el prestigio de la empresa.

Las herramientas tecnológicas como Apache Spark, Apache Storm y Tableau desempeñan un papel esencial en el tratamiento y presentación efectiva y eficiente de los datos en ambientes de big data, simplificando la obtención de información valiosa. Este tipo de framework permite a las empresas examinar información compleja y convertirla en datos de forma efectiva obteniendo un entendimiento profundo sobre el negocio y el ambiente en el cual están inmersos.

En conclusión, el big data no solo ofrece a las organizaciones el beneficio de optimizar sus operaciones actuales, sino que asimismo les brinda la oportunidad de innovar y desarrollar nuevas tecnologías y servicios que se puedan utilizar en las necesidades dinámicas y cambiantes del mercado. El futuro del big data dependerá de la capacidad de las empresas para acoplarse al entorno de datos en grandes volúmenes, sacando el máximo provecho a las oportunidades que ofrece para la innovación, incremento y toma de decisiones estratégicas acertadas.

El impacto social y ético del big data también es un aspecto que no puede ser ignorado. A medida que las empresas y gobiernos aprovechan los datos para obtener información detallada sobre individuos y sociedades, es fundamental asegurar que el uso de estos datos no vulnere los derechos de privacidad ni perpetúe desigualdades. La transparencia en el manejo de los datos y la implementación de prácticas éticas en su análisis serán clave para mantener un equilibrio entre el aprovechamiento de la información y la protección de los derechos individuales.

Por último, la evolución de las tecnologías de big data continuará avanzando rápidamente, y aquellas organizaciones que inviertan en la formación de sus equipos en el uso de estas tecnologías estarán mejor posicionadas para capitalizar las oportunidades emergentes. El éxito futuro dependerá de la capacidad para no solo recopilar y analizar datos, sino también de integrar esos conocimientos en la cultura organizacional para que la toma de decisiones basada en datos sea un componente central de la estrategia empresarial.

Referencias

- Aguilar, L. J. (2016). *Big Data, Análisis de grandes volúmenes de datos en organizaciones*. Alfaomega Grupo Editor.
- Al-Khasawneh, M. (2020). *Big Data Applications and Tools* (pp. 01-15).
- Baldominos, A. (2020). *Resolviendo problemas de big data*. Marcombo.
- Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad. (s. f.). Recuperado 19 de abril de 2024, de <https://www.powerdata.es/big-data>
- Candel, J. M. O. (2022). *Big data, machine learning y data science en python*. Ra-Ma Editorial.
- Fandiño, V. M. L. (2023). *Sistemas de Big Data*. Ra-Ma Editorial.
- Holmes, D. E. (2022). *Big Data: Una breve introducción*. Antoni Bosch Editor, S.A.
- MARQUÉS, M. P. (2015). *BIG DATA - Técnicas, herramientas y aplicaciones*. Alfaomega Grupo Editor.
- Mayer-Schönberger, V., & Cukier, K. (2016). *Big data: La revolución de los datos masivos*. Turner.
- ¿Qué es el big data y cuáles son sus beneficios? (2020, julio 1). <https://postgrado.ucsp.edu.pe/articulos/que-es-big-data/>
- Rodríguez, E. (2023, diciembre 19). *El Origen del Big Data: Una Mirada al Pasado* Canal Innova . Canal Innova. <https://canalinnova.com/el-origen-del-big-data-una-mirada-al-pasado/>
- Romero, J. A. C. (2019). *Big data*. IFCT128PO. IC Editorial.