



Análisis del alcance y la importancia del Big Data en el sector industrial en tiempos de pandemia

Analysis of the scope and importance of Big Data in the industrial sector in times of pandemic

Jorge González

jorgonart@hotmail.com

TECHNOSYS SOLUCIONES

RESUMEN

El objetivo del presente artículo científico es analizar el alcance y la importancia del Big Data en el sector industrial en tiempos de pandemia, por lo tanto, el mismo se sustentó en lo expuesto por López (2012) Han, Pei, y Kamber (2011), Universidad de Alcalá (2020), Román (2019) y desde el punto de vista metodológico lo planteado por Palella y Martins (2012), Hernández, Fernández y Baptista (2014), Tamayo y Tamayo (2015). En cuanto a la metodología, el artículo se enmarcó bajo un enfoque cualitativo de tipo documental, apoyándose para la recolección de información en la observación documental, técnica de subrayado, resumen de texto (técnica de resumen analítico y análisis crítico) y cita textuales. Ahora bien, se pudo conocer que el Big Data es una de las formas de recopilación y manejo de datos más relevantes e innovadoras, donde se expone la estructura y funcionamiento de este sistema, enlazando su uso a las necesidades actuales. Además, se evidenció que debido al contexto por la llegada de la pandemia Covid-19 a partir del año 2020, el Big Data ha sido una herramienta que le permite a las organizaciones una óptima flexibilidad del manejo de cantidades masivas de datos en las mismas. Con base a lo señalado, se concluyó que la herramienta tecnológica Big Data ha comenzado a tomar madurez y se vislumbran grandes oportunidades en su utilización, donde, más allá de tener la capacidad del propiciar el óptimo manejo de volúmenes de datos, también incluye otras dimensiones significativas en el tratamiento de los mismos, De esta manera, el Big Data esta brindado la oportunidad a las distintas empresas facilitandole la modalidad de trabajo a distancia, manteniendo así la operatividad de las mismas, y aún más con la llegada del Covid-19.

Palabras Claves: Big Data, Pandemia, Patrones, Recolección de datos.

ABSTRACT

The objective of this scientific article is to analyze the scope and importance of Big Data in the industrial sector in times of pandemic, therefore, it was based



on what was stated by López (2012 Han, Pei, and Kamber (2011), Universidad de Alcalá (2020), Román (2019) and from the methodological point of view what was proposed by Palella and Martins (2012), Hernández, Fernández and Baptista (2014), Tamayo and Tamayo (2015). Regarding the methodology, The article was framed under a qualitative approach of a documentary type, relying on the collection of information in the documentary observation, underlining technique, text summary (analytical summary technique and critical analysis) and textual quotes. Big Data is one of the most relevant and innovative forms of data collection and management, where the structure and operation of this system is exposed, linking its use to current needs. With the arrival of the Covid-19 pandemic as of 2020, Big Data has been a tool that allows organizations optimal flexibility in handling massive amounts of data in them. Based on the above, it was concluded that the Big Data technological tool has begun to mature and great opportunities are envisioned in its use, where, beyond having the ability to promote the optimal management of data volumes, it also includes other dimensions significant in their treatment. In this way, Big Data is providing the opportunity to different companies by facilitating the modality of remote work, thus maintaining their operability, and even more so with the arrival of Covid-19.

Keywords: Big Data, Pandemic, Patterns, Data collection.

Introducción

Las empresas se han ido transformando progresivamente en el ámbito tecnológico, generándoles cambios importantes a las mismas para impulsarlas a una productividad de grandes escalas, por lo tanto, la automatización, el uso del internet y el surgimiento de nuevos modelos tecnológicos en el ámbito de las telecomunicaciones propician que cada vez más las organizaciones se apoyen en estas herramientas tecnológicas para avanzar y estar a la vanguardia que hoy en día la globalización les exige. De allí, el aumento en la cifra de datos generados por diferentes sistemas y actividades cotidianas en las mismas.

Por lo tanto, ha forjado la necesidad en las corporaciones que tengan el deber de modificar, optimizar y generar métodos, para lograr procesos operativos, de control, administrativos, de comunicación, entre otros que permitan desarrollar con mayor eficiencia y eficacia a la hora de ofrecer servicios o productos de alta calidad para los involucrados con las mismas, sean estos empleados, clientes, socios comerciales y usuarios en el mundo digital.

Es por lo que, las nuevas tecnologías de la información han llegado a las industrias y/o empresas sin importar cuál sea su ramo o sector, para



impulsarlas generando posicionamiento en el mercado, adem s de cambios en los modelos de negocio, donde las sistematizaciones producto por las Tecnolog as de la Informaci n y Comunicaciones (TIC) brindan flexibilidad, velocidad y productividad a las mismas, por lo tanto, las tecnolog as emergentes cada vez m s convergen en los sistemas f sicos y digitales. Sobre la base de las ideas expuestas, Rom n (2019) explica que factores como la investigaci n, desarrollo e innovaci n (I+D+I) dentro de una corporaci n, plantea organizar y potenciar la capacidad de esta para asegurar el  xito a corto, mediano y largo plazo.

El esquema antes planteado, refleja que las empresas deben direccionarse hacia el Internet de las Cosas (IoT), lo cual conlleva a la existencia de la Cuarta Revoluci n Industrial de la Humanidad (industria 4.0), la cual seg n Rom n (2019) se trata sobre el proceso de capturar los datos, apoy ndose en m ltiples tecnolog as de acceso, como computaci n en la nube, macrodatos, entre otros. De esta manera, se evidencia que la modernizaci n y/o digitalizaci n del sector industrial o de las empresas, se inducen al t rmino del contexto inteligente que bajo sistemas de telecomunicaciones logran ejecutar la funcionalidad de sus distintos procesos.

Con base en lo descrito, los modelos de almacenamiento y tratamiento de datos se dirigen a infraestructuras tecnol gicas bajo Big Data. La Universidad de Alcal  (2020), indica que el objetivo de este servicio es gestionar y analizar la cantidad de vol menes de datos que se generan en una empresa por el procesamiento de datos estructurados, no estructurados o semiestructurados; es decir, cuando la corporaci n maneja informaci n en redes sociales, en una red con elevada cantidad de usuarios que demandan servicios triple play, correos electr nicos, telefon a IP, entre otros, las mismas experimentan dificultades para procesar los datos en Data Center, debido a que superan los l mites y capacidades de las aplicaciones que permiten capturar, gestionar y procesar los datos.

A pesar de que el t rmino Big Data se asocia principalmente con cantidades de datos exorbitantes, se debe dejar de lado esta percepci n, pues Big Data no va dirigido solo a gran tama o, sino que abarca tanto volumen como variedad de datos, as  como velocidad de acceso y procesamiento. En la actualidad se ha pasado de la transacci n a la interacci n, ya que las empresas tienen la necesidad de alcanzar la excelencia operativa utilizando informaci n en tiempo real, con el prop sito de obtener el mejor provecho cuando se generan los datos provenientes de un usuario en el mundo digital, en m quinas o dispositivos conectados a una red, que minuto a minuto propician informaci n por medios de distintos canales de comunicaci n, entre otros factores.

Cabe mencionar, que antes de la coyuntura de la pandemia por el Covid-19, esta herramienta tecnol gica ya jugaba un papel importante, debido a que



diversos gobiernos al tener que validar una gran cantidad de información, se vieron en la necesidad de hacer uso de múltiples aplicaciones, tanto en móviles como en otras tecnologías digitales, con la finalidad de controlar la pandemia, informar a la sociedad, generar estadísticas, entre otros elementos que propician datos relevantes. Por lo tanto, el Big Data asegura el soporte de un volumen elevado de información donde los usuarios del mundo digital aprovechan para mantenerse informado con respecto a este tema en tiempo real, ya que la actualización de estos datos es constante.

Ahora bien, al revisar la teoría sobre Big Data, pensando en la creación de soluciones que incluyan problemas enmarcados en este enfoque, Roman (2019) plantea que se pueden encontrar cuatro fases donde se clasifican las diferentes tecnologías de soporte, donde, señala: generación, adquisición, almacenamiento y análisis de datos. En consecuencia, el mencionado autor define como generación al proceso propio de diversas actividades de la sociedad, en estas se genera una cantidad inmensa de datos, que, según su naturaleza, puede estar almacenada y estructurada o puede corresponder a datos sin ninguna estructura, pero con características de gran valor.

En la segunda fase, se incluye la colección de todos los datos generados en la vida diaria. La trasmisión y pre procesamiento de estos es de gran importancia, ya que muchos conjuntos de datos presentan redundancia o datos inútiles que, si no se tratan, pueden incrementar el espacio de almacenamiento innecesariamente y afectar los resultados de una fase de análisis. La fase de almacenamiento de Big Data ha generado la necesidad de crear estudios y propuestas de nuevas estrategias que permitan afrontar los tipos de datos que no se pueden gestionar con un sistema de gestión de bases de datos relacionales.

Finalmente, la fase de análisis debe atender a la necesidad de extraer rápidamente información desde los datos masivos para poder generar valor en las organizaciones y facilitar procesos de toma de decisiones. Se requiere de tecnologías que faciliten incluso el análisis en tiempo real. Por todo lo anterior, es que el objetivo del presente artículo científico se orientó a analizar el alcance y la importancia del Big Data en el sector industrial en tiempos de pandemia.

Importancia del Big Data

La Universidad de Alcalá (2020), explica que el término Big Data hace referencia a la tendencia del avance de las tecnologías que han abierto las puertas hacia un nuevo enfoque de entendimiento y toma de decisiones, la cual es utilizada para describir enormes cantidades de datos que llevaría demasiado tiempo cargarlos en una base de datos relacional para su posterior



an lisis. Por lo tanto, el Big Data se aplicar  para toda aquella informaci n que no pueda ser procesada por los m todos tradicionales.

De all  pues, que para Pei, y Kamber (2011), una base de datos es un conjunto de datos interrelacionados, donde, el modelo relacional se caracteriza a muy grandes rasgos por disponer que toda la informaci n que debe de estar contenida en tablas, y las relaciones entre datos deben ser representadas expl citamente de ese mismo modo. Lo que se consigue con este modelo es trabajar siempre sobre tablas relacionadas entre s , evitando duplicidad de registros y garantizando la integridad referencial, es decir, que si se elimina un registro, se eliminan todos los relacionados. El gran inconveniente que presenta es el tiempo necesario para manejar grandes cantidades de datos, pero esto se logra gracias al Big Data. Por otro lado lo que se consigue al trabajar con bases de datos es combinar diferentes tipos de datos y de una manera formalizada.

Caracter sticas Big Data

L pez (2012) se ala que el Big Data posee tres (3) caracter sticas fundamentales:

- Volumen
- Velocidad
- Variedad

Volumen

Suele utilizarse como sin nimo de Big Data. A pesar de ser uno de los aspectos m s llamativos, no es el  nico. El reto relacionado con el volumen de datos se ha puesto de manifiesto recientemente, debido a la proliferaci n de los sistemas de informaci n e inteligencia, el incremento del intercambio de datos entre sistemas y dispositivos nuevos, nuevas fuentes de datos, y el nivel creciente de digitalizaci n de los medios de comunicaci n que antes s lo estaban disponibles en otros formatos, tales como texto, im genes, videos y audio. La cantidad de datos d a a d a ser  muy superior a las que actualmente existen con lo cual se obtiene un valor a adido. Las empresas est n cubiertas de una cantidad cada vez mayor de datos de todo tipo, acumulando f cilmente terabytes, incluso peta bytes de informaci n.

Velocidad

Se asocia con la proliferaci n de nuevas fuentes de datos, y la necesidad de utilizar estos datos m s r pidamente. Fuentes de datos automatizados tales como sensores RFID y GPS generan datos cada fracci n de segundo para



varias m tricas diferentes y, junto con otros equipos de la empresa, causan un flujo constante de datos que se generan con el tiempo. Los dispositivos que generan datos a intervalos m s largos, tales como los tel fonos inteligentes, tambi n terminan generando corrientes constantes de datos que necesitan ser ingeridos r pidamente. Por otro lado, todos estos datos tienen poco o ning n valor si no se convierten r pidamente en informaci n  til.

Variedad

Los grandes vol menes de datos incluyen cualquier tipo de datos, estructurados y no estructurados como texto, datos de sensores, audio, v deo, secuencias de clic o archivos de registro, entre otros. Al analizar estos datos juntos se encuentra informaci n nueva.

Para explicar esta caracter stica, la mejor opci n es imaginar la creciente cantidad de informaci n que almacena Facebook sobre sus usuarios y lo diversa que es esta. En su base de datos se puede encontrar la edad, el sexo o el pa s de millones de personas. Con Big Data esto es posible de hacer. Est  caracter stica est  relacionada con la organizaci n de los datos. Esta organizaci n se divide b sicamente en datos estructurados, semiestructurados y no estructurados.

Los datos estructurados son los datos tradicionalmente presentes en los sistemas corporativos (bases de datos, archivos jer rquicos y secuenciales, etc.). Los datos semiestructurados suelen estar disponibles a trav s de los registros del sistema (servidores web, CDR, etc.) y los datos no estructurados se relacionan principalmente con el contenido digital m s reciente, y se pusieron a disposici n previamente en un formato no digital, tales como archivos de imagen, audio, texto, entre otros. El universo del Big Data contempla la posibilidad de utilizar todos los datos disponibles a trav s de correos electr nicos, documentos, mensajes, im genes, grabaciones de audio, registros, videos, etc.

Existe la posibilidad para algunos autores de una cuarta "V" la Veracidad, que hace referencia al nivel de fiabilidad asociado a ciertos tipos de datos, es decir, Big Data; lleva asociado un factor de incertidumbre ante ciertos datos, por ejemplo, los sentimientos y sinceridad de los seres humanos, condiciones climatol gicas. Para dejar m s claro est  cuarta "V" se puede establecer una analog a con el sector de la producci n energ tica. En muchos pa ses existe la normativa de que parte de la producci n energ tica debe proceder de fuentes renovables, pero la climatolog a no se puede predecir con precisi n, no se puede saber la cantidad de viento que soplar  para que los molinos e licos generen X cantidad de energ a. Para solucionar esto, los analistas de Big Data tienen diferentes opciones para obtener datos m s fiables. Estas



alternativas son: matem ticas avanzadas (T cnicas de optimizaci n), fusi n de datos de m ltiples fuentes menos fiables, etc.

Otra caracter stica importante que define al Big Data es la complejidad, la cual est  relacionada con la forma de tratar con todas las caracter sticas mencionadas anteriormente, para as  brindar informaci n  til de manera eficiente.

Beneficios del Big Data

A continuaci n, se citan los beneficios e inconvenientes m s relevantes, seg n Vauzza (2013):

- B squeda de nuevas oportunidades de negocio a trav s de segmentaci n mejorada y venta cruzada de productos (mejora de la estrategia).
- Mediante la aplicaci n de an lisis y modelado predictivo a los datos de cuentas de clientes e historial de transacci n, la soluci n permite a los agentes llevar a cabo una segmentaci n basada en la probabilidad de que el cliente contrate servicios o productos complementarios, o contratar servicios de mayor valor (mejora de segmentaci n).
- Mediante el an lisis de consumo de los servicios y productos de los clientes, la empresa puede optimizar las estrategias de venta cruzada, afinar mensajes de marketing y proporcionar ofertas espec ficas. Se puede predecir con mayor exactitud qu  productos son los m s apropiados para cada cliente (mejora de la estrategia).
- Ofrecer la combinaci n adecuada de servicios y productos mejora la eficacia y la eficiencia de la fuerza de ventas de la compa a, mientras que el toque m s personalizado ayuda a los agentes a forjar lazos m s estrechos con clientes, lo cual mejora la lealtad (mejora de la estrategia).
- Mejoras Operativas: Mayor capacidad de visibilidad del negocio a trav s de informes m s detallados.
- An lisis de navegaci n web y h bitos de consumo online:
- An lisis de Redes Sociales: Determinar los c rculos sociales de los clientes a partir de interacciones telef nicas y redes sociales online genera una visi n completa de los clientes, identificando el papel que desempean en sus c rculos y su grado de influencia.
- Marketing Viral (marketing que explota redes sociales...): Detecta clientes m s influyentes, roles sociales... para maximizar la difusi n de



tus productos y servicios (mejor conocimiento de clientes y del mercado en redes sociales).

- Análisis de datos de navegación: Analiza la navegación Web y hábitos de consumo online: extrae nuevas y valiosas perspectivas de los clientes. Se identifica al usuario (localización, estado del terminal, servicios de acceso), se monitorizan sitios y búsquedas por palabra, urls visitadas, tiempo de navegación, etc. (mejor conocimiento del cliente).
- Cuadro de Mandos en tiempo real, la información siempre está disponible sin esperas de actualización de los datos (información en tiempo real).
- Anticipación a los problemas: Un sistema predictivo de análisis y cruce de datos nos permite poder anticiparnos a posibles problemas que puede surgir en el futuro, como por ejemplo una predicción de riesgo de catástrofes que permitiría ajustar la política de precios y aprovisionar fondos para posibles pagos (utilidad para ver la veracidad de los datos ante datos imprecisos) .
- Mejoras de Procesos: Permite la simplificación de procesos actuales y control del negocio (reducción de costes).
- Análisis de Seguridad. Analítica proactiva que permite la reducción de riesgos y pérdidas frente a fraudes (reducción de costes).
- Permite detectar patrones complejos de fraude en tiempo real analizando los datos históricos, el patrón de uso de información de geolocalización, análisis de transacciones y operaciones sospechosas (reducción de costes).
- Soporte a la toma de decisiones a través de algoritmos automáticos.
- Una analítica sofisticada que analice todos los informes y datos, ayuda a la toma de decisiones, reduciendo los riesgos y descubre información que antes podría estar oculta, pero a la vez importante (ayuda a la toma de decisiones).
- Reducción de costes.

Tipos de datos Big Data

Según explica López (2012), se puede decir que básicamente hay tres tipos de datos en Big Data:

- Datos estructurados
- Datos no estructurados
- Semiestructurados



Datos estructurados

Los datos estructurados son aquellos datos que tienen bien definido su longitud y su formato. Suelen ser fechas, números, cadenas de caracteres y están almacenados en tablas. En las empresas estos datos los encontramos en información obtenida a partir de CRM, ERP, etcétera. Estos datos suelen estar guardados en un Data Warehouse si contienen mucha información y si el negocio o la empresa no generan tal cantidad de datos, tendrán una base de datos relacional. Para consultar estos datos se realizan mediante consultas SQL. Los datos estructurados se obtienen de:

Datos generados por máquinas

- Datos procedentes de sensores: existen múltiples ejemplos como los procedentes de un GPS, contadores eléctricos, tacómetros, equipos médicos.
- Web Log Data: servidores, redes, aplicaciones, etc. generan grandes cantidades de datos estructurados.
- Datos procedentes de puntos de venta: basta con pensar en un hipermercado con una cajera pasando códigos de barras por un lector.
- Datos financieros: muchas operaciones bancarias y bursátiles son de datos estructurados generados automáticamente.

Datos generados por personas

Los datos estructurados generados por personas también son variados y pasan desde los registros de una contabilidad en un ERP pasando por el hecho de cumplimentar un formulario en una web o incluso nuestros movimientos en uno de esos juegos on-line que ahora nos encontramos en Facebook. Los datos estructurados son el pilar de las bases de datos relacionales. En los modelos relacionales, toda la información esta guardada en un esquema de tablas y dichas tablas tendrá definidas unos campos y relaciones entre ellas.

Datos no estructurados

Son lo opuesto a los datos estructurados, es decir, carecen de un formato específico. Al igual que los datos estructurados son generados:

Datos generados por máquinas y computadoras

- Imágenes de satélites.



- Datos científicos: gráficos sísmicos, atmosféricos, etc...
- Fotografía y vídeo: por ejemplo, cámaras de vigilancia.
- Datos recopilados de sónar y radar

Datos generados por personas, o sea, datos picados por personas en un ordenador

- Textos incluidos dentro de los sistemas de información internos de las organizaciones: basta con pensar en documentos, presentaciones, correos electrónicos, etc...
- Datos provenientes de redes sociales: Twitter, Facebook, LinkedIn, Flickr, Instagram, Tuenti. El número de redes sociales crece cada día, cada vez es más común ver diferentes redes sociales que hacen referencia a diferentes grupos.
- Datos provenientes de nuestros dispositivos móviles: pensemos en los mensajes que enviamos con nuestros teléfonos móviles.
- Contenido de sitios web: podemos ir desde vídeos de YouTube contenidos de páginas web o incluso blogs.

Se puede decir que la mayoría de los datos de una empresa son no estructurados, y que gracias al Big Data ahora se pueden analizar y obtener información útil para las organizaciones. No obstante, cabe recalcar que el Big Data no solo se centra en los datos no estructurados, sino que los hace en todos, tanto estructurados como no estructurados y sin olvidarnos de los semiestructurados.

Tratamiento de Big Data

En este mismo orden y dirección, el tratamiento de Big Data ha exigido el desarrollo de soluciones computacionales que permitan afrontar las necesidades y retos que traen consigo los grandes volúmenes de datos, su variedad de fuentes y la velocidad con que se generan. A continuación, se da una breve descripción de algunas tecnologías y técnicas de Big Data.

Tecnologías Big Data

Como tecnologías de Big Data se clasifican aquellas que dan soporte a la captura, transformación, procesamiento y análisis de los datos, ya sean estructurados, semiestructurados o no estructurados. Cabe aclarar que existen un mayor número de tecnologías que soportan Big Data, tanto libres como propietarias, pero para efectos de este documento se ha acotado de acuerdo con lo anteriormente expuesto y tomando las tecnologías que dieron las bases iniciales al ecosistema Big Data.



Hadoop

Hadoop es una librer a de Apache definida como un framework que permite hacer procesamiento de datos distribuido sobre vol menes de datos de considerable tama o sobre cl ster. Asimismo, como explica The Apache Software Foundation (2016 a) est  dise ado para brindar poder de escalamiento desde un par de servidores hasta cientos de m quinas o nodos, las cuales manejan almacenamiento y procesamiento local. Hadoop cuenta con dos componentes principales, el HDFS, sistema de archivos distribuidos que permite distribuir los ficheros en distintas m quinas y MapReduce, framework que permite al desarrollador aislarse de la programaci n paralela, permite ejecutar programas escritos en lenguajes de programaci n conocidos (p.e Java) en el cl ster de Hadoop. El HDFS cuenta con tres pilares b sicos. Namenode, se ocupa del control de acceso y tiene la informaci n sobre la distribuci n de datos en el resto de nodos. Datanodes, son los encargados de ejecutar el c mputo, es decir, las funciones Map y Reduce, sobre los datos almacenados de manera local en cada uno de dichos nodos. Jobtracker, este nodo se encarga de las tareas y ejerce el control sobre la ejecuci n del proceso de MapReduce. Adem s, el HDFS cuenta con las siguientes caracter sticas fundamentales:

- Tolerancia a fallos.
- Acceso a datos en streaming.
- Facilidad para el trabajo.
- Modelo sencillo de coherencia.
- Portabilidad de convivencia.

Varios trabajos donde se ha tomado Hadoop como base y se ha potencializado algunas de sus caracter sticas o se ha fusionado con otra herramienta o tecnolog a.

MapReduce

Seg n T. A. S. Foundation (2016 b) MapReduce es un modelo de programaci n que se ha asociado tambi n a la implementaci n de estrategias de procesamiento de grandes conjuntos de datos que puede ser aplicado a una gran variedad de tareas del mundo real. Este modelo de programaci n fue utilizado inicialmente por Google para resolver el problema de ranking de p ginas ("Page Rank"). L mmel (2008) se ala que el modelo se basa en los siguientes conceptos: iteraciones sobre los datos de entrada, construcci n de los pares clave-valor a partir de cada pieza de entrada, agrupaci n de los valores intermedios de acuerdo con las claves, iteraci n sobre los grupos resultantes y reducci n de cada grupo.



Mapeo

Se aplica en paralelo para cada uno de los ítems en la entrada de datos. Por medio de la tarea de mapeo (Map) a cada llamada se asignará una lista de pares clave-valor (key-value). Por cada clave generada se crea un grupo, el framework agrupa todos los pares con la misma clave extraídos de todas las listas tratadas.

Reducción: se aplica en paralelo para el grupo asociado a una clave. El resultado es la producción de una colección de valores para cada dominio. Distribución y ordenamiento: tiene dos misiones, por una parte, se encarga de ordenar por clave todos los resultados emitidos por los mapper y por otra parte recoge todos los valores intermedios pertenecientes a una clave para combinarlos en una lista asociada a ella.

Las características de MapReduce se resumen a continuación:

- Distribución y paralelización automáticas.
- Tolerancia a fallos y a redundancias.
- Transparencia.
- Escalabilidad horizontal.
- Localización de los datos.
- Herramientas de monitorización.

Este paradigma ha sido implementado en numerosas aplicaciones, algunos ejemplos se pueden encontrar en los siguientes documentos, todos se caracterizan por el uso de MapReduce como base de su implementación.

HBase

Es una base de datos Hadoop, distribuida y escalable. HBase ha sido desarrollada por Apache y se recomienda su uso cuando se necesita acceso a lectura y escritura de datos en tiempo real sobre Big Data. El objetivo de HBase es el almacenamiento de tablas de gran tamaño, con billones de filas por millones de columnas. Esta base de datos no relacional fue modelada después de Bigtable de Google es open source, distribuida y versionada. HBase provee capacidades similares a Bigtable sobre Hadoop y HDFS. Algunas de sus principales características son:

- Escalabilidad modular y linear.
- Estricta consistencia de lectura y escritura.
- Facilidad de uso de la API de Java para el acceso de clientes.
- Bloqueo de la caché para consultas en tiempo real.
- Soporte de para exportar métricas a través del subsistema de métricas de Hadoop.



Cassandra

La base de datos Cassandra, propiedad de Apache, brinda escalabilidad y alta disponibilidad sin comprometer el rendimiento. Se considera una plataforma ideal para tratar problemas de datos cr ticos, puesto que cuenta con escalabilidad lineal y la tolerancia a fallos en el hardware o en la infraestructura en la nube. Cassandra ofrece un modelo de datos que cuenta con comodidad para la indexaci n de columnas, soporte a la desnormalizaci n y materializaci n a las vistas y un poderoso almacenamiento en cach  integrado.

Es un sistema de almacenamiento distribuido con un modelo de datos que soporta un control din mico sobre el dise o y el formato de los datos. Algunos de los principales atributos de Cassandra son:

- Tolerancia a fallos, por medio de la replicaci n autom tica de los datos en m ltiples nodos.
- Descentralizaci n, uso de muchos nodos id nticos.
- Durable, dise ada para evitar la p rdida de datos.
- Elasticidad, capacidad de a adir nuevas m quinas para aumentar el rendimiento de lectura y escritura.

Mahout

Como lo indica T. A. S. Foundation (2016 a), Mahout es un proyecto de Apache que tiene como objetivo ofrecer un ambiente para la creaci n r pida de aplicaciones escalables y eficientes de aprendizaje m quina. Mahout ofrece una suite de algoritmos para clustering, categorizaci n, filtrado colaborativo, clasificaci n y programaci n evolutiva.

Algunas de sus principales aplicaciones pr cticas se enmarcan en la realizaci n de cl ster de documentos, recomendaciones y organizaci n de contenidos. El machine learning o aprendizaje m quina es el trasfondo principal de Mahout y corresponde a un subcampo de la inteligencia artificial que se centra en el mejoramiento de procesamientos computacionales a partir del an lisis de experiencias previas. Mahout desde su aparici n ha seguido siendo un proyecto en desarrollo, crecimiento y expansi n.

T cnicas Big Data

Miner a de Datos

La miner a de datos (data mining - DM) se puede definir, seg n Han, Pei, and Kamber (2011), como el proceso de extracci n de conocimiento a partir de c mulos de datos. Se suele utilizar el t rmino miner a de datos como



sin nimo de descubrimiento de conocimiento, pero realmente no son sin nimos, la miner a de datos es solo un paso en el proceso de descubrimiento de conocimiento.

La miner a de datos nace de la necesidad de conocer informaci n  til a partir de los bases de datos o Data Warehouse, con el crecimiento de los datos disponibles, la inteligencia de negocios tuvo que dar paso a la aplicaci n de la miner a de datos en soluciones empresariales y comerciales, puesto que de esta manera se permite el descubrimiento autom tico o semiautom tico de informaci n relevante a partir de estos c mulos de datos. En las ciencias y la ingenier a existe un amplio rango de problemas y dominios de aplicaci n para la miner a de datos.

Se encuentran soluciones a partir de miner a de datos para problemas de los campos de mercadeo, comercio, salud, predicci n, transporte, meteorolog a, entre otros.

Machine Learning

Aprendizaje m quina es un  rea de investigaci n bastante reconocida en las ciencias de la computaci n, de acuerdo con la definici n de Michalski, Carbonell and Mitchell (2013) puede ser entendido el descubrimiento de modelos, patrones y regularidades en los datos. El aprendizaje m quina puede ser visto desde dos enfoques, los simb licos y los estad sticos.

Los primeros trabajan aprendizaje inductivo de descripciones simb licas, mientras que los segundos se centran en los m todos de reconocimiento de patrones o en la estad stica. En los  ltimos a os, el uso del aprendizaje m quina se ha extendido con rapidez, se ven aplicaciones en dominios como detecci n de fraudes, sistemas de recomendaci n, detecci n de spam, predicciones financieras, comercio y mercadeo, entre otros. Los algoritmos de aprendizaje m quina se clasifican en supervisados y no supervisados.

Reconocimiento de patrones

Ahora bien para Dean y Ghemawa (2008), el reconocimiento de patrones es una t cnica que se aplica principalmente en procesos de ingenier a, computaci n y matem ticas que tiene como objetivo extraer informaci n, a partir de un c mulo de datos, que brinde la posibilidad de establecer propiedades o relaciones entre estos datos. En el procesamiento de patrones generalmente se usan algoritmos de optimizaci n, puesto que su intenci n es hallar una mejor soluci n respecto a un criterio definido, teniendo en cuenta que un proceso de optimizaci n es una situaci n que requiere elegir desde un conjunto de alternativas, la que lleve al fin requerido con el costo m nimo.



Algoritmos genéticos

Los algoritmos genéticos (GA), para Dean y Ghemawat (2008) coinciden que son una técnica aplicada en la ingeniería computacional, pero que parte de la concepción biológica de la genética. Estos algoritmos comprenden un enfoque que busca dar solución a diversos problemas matemáticos intangibles que no han podido tener solución desde otros enfoques matemáticos tradicionales. Los algoritmos genéticos utilizan también operaciones genéticas como la mutación, recombinación y cruce.

Los algoritmos genéticos son métodos de búsqueda estocásticos diseñados para explorar problemas complejos, con el fin de encontrar una solución óptima, generalmente usando información propia del problema como guía de la búsqueda. Los algoritmos genéticos se enmarcan dentro de los Algoritmos Evolutivos (EA) siendo uno de los componentes más importantes, junto con la programación genética y las estrategias evolutivas. Pueden ser entendidos como componentes esenciales de los algoritmos genéticos los siguientes:

- Estrategia de codificación que determina la forma en que se representará la solución en forma de cromosomas.
- Población de cromosomas o individuos.
- Mecanismo para la evaluación de cada cromosoma.
- Procedimiento de selección/reproducción.
- Operadores genéticos: cruce, mutación.
- Probabilidades para los operadores genéticos.
- Un criterio de finalización.

Aprendizaje de reglas de asociación

El aprendizaje de reglas de asociación es un método para encontrar las relaciones entre variables en grandes bases de datos, su objetivo es identificar reglas usando algunas medidas de relación de intereses, por ejemplo, en el caso de las redes sociales, se trataría de revisar las personas que posiblemente le interesarían seguir a otras dependiendo de sus amistades o seguidores. En el caso de tiendas de productos, podría ser la revisión de los productos que se compran juntos con frecuencia para sugerirlos a un cliente que adquiera uno de los productos relacionados.

Uso del Big Data contra el COVID-19

En los últimos dos años se han producido cambios en el contexto general en el mundo, debido a los criterios sanitarios asociados a la pandemia de Covid-19, donde la misma es contagiosa y se propaga de manera veloz entre los seres humanos, causando numerosas afecciones a nivel respiratorio.



Dicho virus fue descubierto en un brote que aconteció en la ciudad de Wuhan de la República Popular China en el mes de diciembre del año 2019. De allí pues, que este virus conlleva al síndrome respiratorio de Oriente Medio (MERS-CoV) y al síndrome respiratorio agudo severo (SARS-CoV), por lo tanto, la Organización Mundial de la Salud (OMS, 2020) dio a conocer a este virus como Covid-19.

Estas razones, las cuales evidenciaron una propagación exponencial del referido virus, conllevaron a que las industrias cambiaran su rutina de trabajo y operatividad de sus actividades, debido a que se vieron en la obligación de clausurar ciertas áreas de trabajo y así como también aplicar con la mayoría de sus empleados cuarentena domiciliaria. En consecuencia, se han direccionado aún más al uso de las redes de telecomunicaciones, debido a que estas brindan conectividad social, acceso a información, mando de acciones por intermedio de comunicaciones inalámbricas, entre otras actividades que las empresas y/o fabricas necesitan para su productividad o servicios que ofrecen a los beneficiarios de las mismas, donde el almacenamiento y suministro de datos es importante para el óptimo comportamiento de los procesos tanto operativos como administrativos.

Es por ello, que la industrias se han visto en la necesidad de apoyarse en estrategias gerenciales, que puntualizan a establecer dentro de su contexto operacional contar con una infraestructura tecnológica que los dirija a la resiliencia, y adecuarse a los drásticos cambios de consumo que han ido presentando los clientes por las consideraciones diseñadas para evitar contagios del Covid-19, la cual los lleva a introducirse al mundo digital estableciendo comportamientos y necesidades nuevas, por ende las empresas se han ido vinculando con el uso del Big Data, debido a que este permite acceder a aplicaciones y/o software que garantiza la óptima automatización de los procesos que desarrolla una industria para alcanzar la satisfacción del cliente.

Ahora bien, esta acción gerencial ha propiciado que las corporaciones cada vez más se apoyen en el teletrabajo, Smart working, e-commerce y otras plataformas de comunicación usadas para brindar procesos comerciales, operativos y administrativos de manera inmejorable bajo el soporte del Big Data, por lo tanto estas nuevas dinámicas han ayudado a evitar nuevos contagios, disminuyendo que el virus se propague de manera exponencial, por lo tanto, este tipo de innovación incrementa la flexibilidad para el manejo de altos volúmenes de datos, que se propician en la atención al cliente, relaciones con proveedores y/o socios comerciales, en el trabajo en equipo, donde la virtualidad acompañada del Big Data son unas de las actuales herramientas tecnológicas viables para un óptimo distanciamiento social.

En consecuencia, se vislumbra un sin número de aplicaciones en las que ante el panorama expuesto producto al Covid-19, la tecnología ha venido



demostrar su valía en ayudar a encontrar y establecer determinados factores relacionados para evitar el contagio con el referido virus, de allí pues, que las industrias al verse dentro de un contexto donde la virtualidad les ha generado un alto volumen de información han recurrido al Big Data.

Reflexión Final

En la viviente actualidad, donde los seres humanos se han visto en la necesidad de adaptarse a un contexto que conlleva al distanciamiento social, al uso de la virtualidad con mayor frecuencia, a restringirse a ciertas áreas laborales entre otras actividades debido al Covid-19, los individuos se han limitado en desarrollar con la normalidad que usualmente lo hacían.

Por lo tanto, la pandemia propició una crisis global desde la perspectiva que se analice, es decir, en el aspecto económico, político y social, donde las industrias han sido las más afectadas por la emergencia sanitaria mundial, sin embargo, esta emergencia sanitaria les ha permitido a las empresas y/o fabricas replantearse, propiciando estrategias que fomenten nuevos hábitos, donde por medio de los avances tecnológicos van consolidando actuaciones que benefician al logro de los objetivos planteados.

Con base en lo anterior, el estudio del Covid-19 empleando Big Data puede valerse de la analítica retrospectiva y descriptiva avanzada, expresas en la inteligencia de negocios (Business intelligence); ya que esta permite focalizar estrategias gerenciales que buscan generar soluciones inmediatas, debido a que la referida herramienta tecnológica soporta el almacenamiento de datos para que a su vez estos puedan ser consultados en tiempo real y con una óptima inmediatez, donde el sector industrial cada vez más se introduce al uso del Big Data para propiciar tranquilidad y seguridad, pues este genera intensificación de la amplitud sobre el manejo de la información en el entorno tanto operativo, como administrativo y por su puesto en la salud.

De allí que, para el núcleo empresarial de la actualidad, contar con una plataforma de comunicación que sea emergente y que se apoye en esta tecnología, le permitirá que los datos sean operados de forma metódica, logrando una óptima funcionalidad en cuanto a captura, análisis, búsqueda, intercambio, visualización, privacidad de los datos. Bajo estos criterios, entra a formar parte la ciencia de datos empleando los principios del Big Data que son volumen, variedad, velocidad, veracidad y valor de la información.

Finalmente, se evidencia que a pesar de que la ciencia de la salud no ha brindado una solución concluyente al mundo con respecto al Covid-19, la tecnología ha aportado estrategias para evitar contagios, cerrar las labores en las industrias, evitar multitudes, entre otros factores, que debido a la pandemia se han propiciado, y gracias al avance tecnológico los individuos, como las empresas, organizaciones, entes gubernamentales y fábricas, se han ido



incorporado de manera vertiginosa a las Tecnolog as de la Informaci n y Comunicaciones (TIC), por intermedio de sus m ltiples herramientas, donde estas cada vez m s se consolidan como una instrumento necesario e importante para el  ptimo desempe o, es por esto que al mundo empresarial, el Big Data le ha brindado alta calidad dentro de los procesos comerciales, operativos, administrativos y de salud.

Conclusiones

Las tecnolog as asociadas al enfoque de Big Data ya han comenzado a tomar madurez y se vislumbran grandes oportunidades y retos en su utilizaci n, optimizaci n y adaptaci n a diferentes dominios de datos. Sin embargo, ya se encuentran resultados que muestran sus beneficios en aspectos como la reducci n de tiempos, optimizaci n de recursos y mayor flexibilidad.

Con base a lo se alado, se concluy  que la herramienta tecnol gica Big Data ha comenzado a tomar madurez y se vislumbran grandes oportunidades y retos en su utilizaci n, optimizaci n y adaptaci n a diferentes dominios de datos, donde, m s all  de tener la capacidad del propiciar el  ptimo manejo de vol menes de datos, tambi n incluye otras dimensiones significativas en el tratamiento de datos, como la variedad, velocidad y veracidad, por ende, es Por lo tanto, Big Data ha brindado la oportunidad de que las mismas puedan implementar el trabajo a distancia, manteniendo la operatividad en un contexto nuevo e innovador para la humanidad, a causa del distanciamiento social que ha sido una pol tica que ha permitido al mundo dentro de todo contexto, sobre todo el laboral, evitar la propagaci n del referido virus.

Existe una estrecha relaci n entre diferentes m todos y tecnolog as para la construcci n de soluciones que integren las capacidades de cada uno de estos y las potencien en nuevas propuestas. Big Data no trata solo de grandes vol menes de datos, sino que incluye otras dimensiones significativas en el tratamiento de datos, como la variedad, velocidad y veracidad. No obstante, una implementaci n de Big Data requiere altos costos en expertos, mayor tiempo de adaptaci n tecnol gica, dificultad para implementar nuevos an lisis y percepci n limitada.

Big Data no busca sustituir a los sistemas tradicionales, sino edificar una nueva tendencia donde se construyan arquitecturas de sistemas que permitan manejar todas las peticiones. Y ya ha logrado incentivar en la comunidad acad mica y comercial el desarrollo de tecnolog as de apoyo que toman los paradigmas base y los emplean en la construcci n de soluciones particularizadas a problemas de entornos de investigaci n y producci n reales.

En este sentido, las industrias en la actualidad debido al nuevo paradigma gerencial producto de la pandemia, se ven enfocadas cada vez m s en brindar



atención al cliente por medio de plataformas digitales, esto los obliga a generar, transmitir, procesar y analizar enormes cantidades de datos en tiempo real. Por lo tanto, Big Data ha brindado la oportunidad de que las mismas puedan implementar el trabajo a distancia, manteniendo la operatividad en un contexto nuevo e innovador para la humanidad, a causa del distanciamiento social que ha sido una política que ha permitido al mundo dentro de todo contexto, sobre todo el laboral, evitar la propagación del referido virus.

En síntesis, el alcance e importancia de Big Data dentro del sector industrial en tiempos de pandemia, se ve referenciado de acuerdo con el volumen de datos que estos manejan. Esta herramienta tecnológica le permite a la gerencia tomar decisiones en pro de lograr los objetivos planteados, cubrir las necesidades que se presenten, segmentar a los clientes, apoyarse en el teletrabajo, Smart working, e-commerce, u otro sistema de comunicación que permita la continuidad operativa de la empresa en el mercado, todo esto basado en el análisis que Big Data propicia a través de la recopilación de información de los involucrados en la misma, es decir, empleados, clientes, socios comerciales y/o de cualquier usuario del mundo digital.

Referencias Bibliográficas

- Dean, J. y Ghemawat, S. (2008). MapReduce. *Commun. New York: Communications, IEEE Transactions*
- Han, J., Pei, J. and Kamber, M. (2011). *Datamining: concepts and techniques*. 3rd ed. E. Inc., Ed. New York: Morgan Kaufmann Publishers.
- López, D. (2012). *Análisis de las posibilidades de uso de Big Data en las organizaciones*. (Master in Business and Information Technology). Universidad de Cantabria. [Página Web en línea]. Disponible en: <https://repositorio.unican.es/xmlui/bitstream/handle/10902/4528/TFM%20-%20David%20L%c3%b3pez%20Garc%c3%adaS.pdf?sequence=1> [Consulta: 2021, agosto 20].
- Lämmel, R. (2008). Google's MapReduce programming model. New York: Revisited Sci. Comput. Program., 70(1), 1–30.
- Michalski, R. S., Carbonell, J. G. and Mitchell, T. M. (2013). *Machine learning: An artificial intelligence approach*. New York: Springer Science & Business Media.



Organizaci n Mundial de la Salud. (2020). *Brote de Enfermedad por Coronavirus (Covid-19)*. [P gina Web en l nea]. Disponible en: https://www.who.int/es/emergencias/diseases/novel-coronavirus-2020?adgroupsurvey={adgroupsurvey}&gclid=CjwKCAiA7dKMBhBCEiWAO_crFHwY4PORBuHwebEMdRccfcuixV5FjOq-mmYXh2T5ZO52psU-0_aWTxoCO44QAvD_BwE [Consulta: agosto 2021].

Santa Palella Stracuzzi y Feliberto Martins Pestana (2012). *Metodolog a de la investigaci n cuantitativa* 3ra Edici n. Fondo Editorial de la Universidad Pedag gica Experimental Libertador (FEDUPEL) Caracas, Venezuela.

Rom n, T. (2019). *Una revisi n de los significados y las implicaciones del concepto Industria 4.0*. Madrid: Editorial Esic.

The Apache Software Foundation (2016 a). *Apache Mahout: Scalable machine learning and data mining, Apache Mahout*. [P gina Web en l nea]. Disponible en: <http://hadoop.apache.org/> [Consulta: agosto 2021].

The Apache Software Foundation (2016 b). *Welcome to ApacheTM Hadoop®, hadoop*. [P gina Web en l nea]. Disponible en: <http://hadoop.apache.org/> [Consulta: agosto 2021].

Universidad de Alcal  (2020). *Big Data y sus aportaciones a la Industria 4.0*. [P gina Web en l nea]. Disponible en: <https://www.masterindustria40.com/big-data-aportaciones-la-industria-4-0/> [Consulta: agosto 2021].

Vauzza, M. (2013). *Expertos en estrategia e implementaci n de proyectos tecnol gicos*. [P gina Web en l nea]. Disponible en: <http://vauzza.es/es/> [Consulta: agosto 2021].