

Macrodatos: almacenamiento y consulta de datos aplicados en las organizaciones

(Big data: storage and consultation of data applied in companies)

Jeimy Isabel Tapia Huicochea *

RESUMEN

Este trabajo ha sido realizado con la finalidad de exponer detalladamente lo que significa Big Data, su composición y lo que hace que sea tan importante hoy en día para las empresas; además, se describe lo que es el lenguaje SQL y los beneficios que ha traído consigo la implementación de Big Data en las organizaciones para el almacenamiento de los datos y cómo estos pueden ser consultados a través del lenguaje de programación, siendo este una interfaz de comunicación entre la base de datos y la interpretación que quiere darse a los datos almacenados en la plataforma. Se detallará que existen herramientas especializadas para la edición, manipulación y tratamiento de los datos almacenados para que estos puedan convertirse en información para las organizaciones.

Palabras clave: SQL; Big Data; tecnología; empresas.

ABSTRACT

This work has been carried out in order to expose in detail what Big Data means, its composition and what makes it so important today for companies; In addition, it describes what the SQL language is and the benefits that the implementation of Big Data has brought with it in organizations for the storage of data and how these can be consulted through the programming language, this being a communication interface between the database and the interpretation that wants to give to the data stored on the platform. It will be detailed that there are specialized tools for editing, manipulating and processing stored data so that they can become information for organizations.

Keywords: SQL; Big Data; technology; companies

JEL Classification: C29.

* Licenciada en Gestión y Dirección de Negocios por la Universidad Veracruzana, estudiante de la Maestría en Administración en el IIIESCA. tapiajeimy@gmail.com

INTRODUCCIÓN

Esta investigación tiene la intención de analizar el lenguaje de consulta SQL y su aplicación en los Macrodatos o Big Data como una forma de promover su aplicación en los negocios, sea cual sea su índole, tamaño forma o estructura, ya que ofrece una cantidad de beneficios impresionantes para quien lo aplique y sepa interpretar la información. Es una herramienta que funciona para enfatizar una propuesta de valor, establecer una segmentación y por supuesto que además favorece al crecimiento del negocio.

Para dar inicio con el tema se presenta con el concepto de SQL:

SQL es un tipo de lenguaje de programación que ayuda a solucionar problemas específicos o relacionados con la definición, manipulación e integridad de la información representada por los datos que se almacenan en las bases de datos, (Ramos, 2018).

Para contextualizar el contenido de este tema, se citan a continuación otras definiciones:

- El lenguaje SQL sirve para el acceso a la información almacenada en las bases de datos. Es un lenguaje sencillo de consulta, que permite realizar operaciones de selección, inserción, actualización y borrado de datos, así como operaciones administrativas sobre las bases de datos, (López, 2022).
- SQL se centra en la integración de SQL y XML definiendo un lenguaje XQuery que se utiliza para consultar datos en documentos XML. SQL no es un acrónimo de nada, aunque mucha gente insistirá en que significa "Lenguaje de consulta estructurado", (Beaulieu, 2009).
- SQL es un tipo de lenguaje de programación que ayuda a solucionar problemas específicos o relacionados con la definición, manipulación e integridad de la información representada por los datos que se almacenan en las bases de datos. Algunos aspectos de SQL están basados en el cálculo relacional, algunos en el álgebra relacional que provienen del modelo relacional y otros a ninguno de los dos, sino que son parte de SQL, (Styde, 2018).
- SQL se trata de un lenguaje declarativo estándar que permite su integración

en diferentes lenguajes de programación, por ejemplo, PHP o Java y en combinación con cualquier base de datos específica. MariaDB, MySQL o SQL Server. Con SQL se pueden realizar diferentes operaciones para consultar o manipular datos. A estas operaciones se les refiere como operaciones CRUD (de Create, Read, Update y Delete), (Robledano, 2019).

El concepto Big Data engloba numerosas ideas y definiciones que se asocian a términos como Data Science, Data Analytics o Data Mining, siendo el objetivo fundamental de todos ellos la recopilación y creación de valor de los datos para transformarlos en información útil para tomar decisiones dentro de una empresa.

- Big data puede ser considerada como una tendencia en el avance de la tecnología que ha abierto la puerta a un nuevo enfoque para la comprensión y la toma de decisiones que se utiliza para describir las enormes cantidades de datos (estructurados, no estructurados y semi estructurados) que sería demasiado largo y costoso para cargar una base de datos relacional para su análisis. El concepto se aplica a toda la información que no puede ser procesada o analizada utilizando herramientas o procesos tradicionales, (Puyol, 2014).
- Fue en ciencias como la astronomía y la genética en la década de los 2000, donde se acuñó el término big data, definida ahora por la capacidad de la sociedad de aprovechar la información de formas novedosas, para obtener percepciones útiles o bienes y servicios de valor significativo. Los datos masivos se refieren a cosas que se pueden hacer a gran escala, pero no a una escala inferior, para extraer nuevas percepciones o crear nuevas formas de valor, de tal forma que transforman los mercados, las organizaciones, las relaciones entre los ciudadanos y los gobiernos, (Mayer, 2013).
- Las **Tecnologías Big Data** ayudan a mejorar la eficiencia, calidad y los productos y servicios personalizados ofrecidos por las organizaciones. En términos científicos, el **Análisis Big Data** abre nuevos horizontes de investigación

con resultados potencialmente más prometedores y con revelaciones más profundas que antes. En muchos casos, el análisis Big Data integra datos estructurados y no estructurados con respuestas en tiempo real, abriendo nuevos caminos a la innovación y desarrollo, (DaSCI, 2021).

I FUNDAMENTOS

La idea fundamental del SQL es comunicarse con la base de datos en un lenguaje específico. Esto quiere decir que SQL debe usarse solo como una herramienta que permite manipular y editar tablas de datos estructurados, ya que no es un lenguaje de programación, sino que es un lenguaje de comandos y sintaxis que puede usarse sólo para crear, mantener y buscar sistemas de bases de datos relacionales. Aunque si tiene algunas funciones de programación como condicionales y bucles, es necesario recurrir a otras plataformas para realizar cosas de mayor complejidad ahí y así mantener ordenados los datos que están almacenados y determinar puntualmente lo que es necesario para traducirse en decisiones para una empresa u organización que esté haciendo uso de estas herramientas.

Una base de datos es una recopilación organizada de información o datos estructurados, que normalmente se almacena de forma electrónica en un sistema informático. Normalmente, una base de datos está controlada por un sistema de gestión de bases de datos (DBMS). En conjunto, los datos y el DBMS, junto con las aplicaciones asociadas a ellos, (Oracle, 2021).

Cuando esto se lleva a lo grande, se convierten en bases de datos más sofisticadas y especializadas para organizar una cantidad impresionante de datos y además ser capaz de seleccionar, separar, conjuntar, elegir la información relevante e inferir en ideas que se generan a través de la relación que existe entre todos estos datos, clasificados por usuarios que pueden residir en distintas partes del mundo.

La magnitud de información que recibe una base de datos es tan amplia que debe estar respaldada por un equipo de expertos en Data Scientist que sepan extraer el valor de la información que proporcionan esos datos.

Dentro de la definición de Big Data, cabe señalar que está compuesta por las cuatro V: Volumen, Velocidad, Variedad y Veracidad. Las cuales se describen a continuación, tomando como referencia un texto citado de Nubia Márquez, (Ibertech, 2016).

- **Volumen:** Es la acumulación de grandes cantidades de datos que se reúnen diariamente en una empresa. El reto es saber gestionar y procesar tal cantidad de información y desarrollar un método continuo de recopilación y gestión del dato en directo, dado que cada día que pase se generará una mayor cantidad de información. La habilidad para saber tratar dicho dato y transformarlo a un formato fácil de analizar definirá la capacidad de dicha empresa de saber gestionar y aprovechar el volumen de datos existente.
- **Velocidad:** Es la rapidez con la que se generan los datos en la actualidad. Saber recopilar, gestionar, analizar y extraer información de valor en tiempo real, para distribuirla en todos los departamentos de la organización.
- **Variedad:** Es la diferencia que hay en el formato y el origen de cada dato. Las formas de recopilarlos, el tratamiento que se le da a cada uno y el orden en el que se almacenan, las cuales pueden ser en bases de datos estructuradas o no estructuradas.
- **Veracidad:** se refiere a la calidad, la predictibilidad y la disponibilidad del dato. Es la variable menos uniforme y menos sencilla de controlar, debido a la dificultad de cerciorarnos de que un dato es 100% fiable. La clave para poder afrontar con garantías esta última faceta del dato es tener un equipo imparcial que ayude a mantener los datos limpios para su posterior evaluación de estrategia de Big Data.

La importancia de los datos se ha vuelto tan necesaria para las empresas, que han requerido de sistemas avanzados para su almacenamiento y estudio. El big data está siendo parte fundamental para realizar estas tareas tan sofisticadas que demandan las organizaciones. Debido a su fácil acceso, a su volumen masivo, y sobre todo a la gran cantidad de información que de él se puede extraer.

Y lo más importante no está en depurar dicho dato a la perfección en una de sus cuatro facetas, sino en realizar un trabajo depurativo en todas ellas, para poder desarrollar la capacidad de trabajar con una gran cantidad de datos, que se actualicen diariamente, que provengan de fuentes muy heterogéneas, y sin la seguridad al 100% de su veracidad y fiabilidad.

II CARACTERÍSTICAS DEL SQL

El SQL puede ser usado solo en datos estructurados que estén dentro de una base de datos relacional. Funciona de una manera más sencilla que Python o Excel, sirve para crear, mantener y proteger bases de datos, además que permite hacer consultas, actualizar, agregar y eliminar datos de bases de datos grandes y complejos. Ofrece un conjunto muy sencillo y estandarizado de comandos y métodos básicos para hacer cosas como consultar, unir, agrupar, ordenar e incluso conjuntos de datos estructurados de extracción de texto.

Este lenguaje ofrece un rendimiento operativo muy rápido y eficiente puesto que su estructura está diseñada para que no requiera de la interacción de un usuario (se requiere en una mínima proporción), se trata de un lenguaje muy similar al humano, por lo que su manejo no resulta complicado, así que se optimiza prácticamente solo.

En su estructura, incorpora una serie de rutinas que se traducen en un importante ahorro de tiempo y también de dinero, ya que su evolución ha mantenido los principios básicos de SQL. Al permitir entender mejor el funcionamiento de los modelos de programación OMR y permitir el uso compartido por varios usuarios facilita el trabajo en cualquier equipo multidisciplinar, (Unir, 2021).

Hay tres bases de datos relacionales de código abierto para SQL, estos son SQLite, MySQL y PostgreSQL, las cuales se describen a continuación en orden de menor a mayor complejidad donde se identifican ventajas y desventajas de cada uno.

- SQLite: este software es más limitado que otras implementaciones de SQL, especialmente en lo que respecta a la gestión de usuarios y las personalizaciones que mejoran el rendimiento, pero es un lugar adecuado para realizar prácticas, ya que es un software gratuito y de código abierto, en el que no se requiere una licencia especial para usarlo. Puede admitir una base de datos de hasta 140 TB de tamaño.
- MySQL: es la versión de código abierto más popular de SQL, ha impulsado muchos sitios web como Twitter, Facebook, Netflix y Spotify. Ofrece una versión completa y potente de SQL y se utiliza en el backend de millones de sitios web. Permite un mayor control sobre quién tiene acceso a la base de datos,

puesto que para acceder a ella se requiere de un proceso daemon separado.

- PostgreSQL: este software agrega elementos orientados a objetos al lenguaje relacional de SQL. Es capaz de manejar de manera eficiente múltiples tareas al mismo tiempo, una característica conocida como concurrencia. Es conocido como la base de datos relacional de código abierto más avanzada del mundo, pero no es recomendable para configuraciones de bases de datos simples. (Brand, 2015).

III ANALÍTICAS Y CAPAS TECNOLÓGICAS EN BIG DATA

La ciencia del Big Data radica en su forma de operar al procesar la información mediante datos analíticos almacenados de acuerdo a distintos criterios. La información proporcionada por los usuarios es procesada, clasificada y perfectamente analizada por una cadena de expertos que depuran esta información para convertirla en respuestas valiosas para las empresas. El estudio del cliente ha demostrado ser lo más adecuado para fidelizarlo y dirigir las acciones de negocio hacia el éxito. Las técnicas de Data Mining, Machine Learning o Procesamiento del Lenguaje Natural ofrecen a los departamentos de Marketing la oportunidad de segmentar a los clientes según sus preferencias, (IIC, 2022).

Tan solo con la geolocalización que está activa en los dispositivos móviles, en las conexiones de Wifi, con las cámaras de vigilancia, etc. Toda esa información puede ser almacenada a través del Big Data y convertirse en una gran vía de negocio, ya que permite a las empresas conocer con claridad las necesidades de sus clientes y el comportamiento de estos frente a ciertos productos y servicios.

Big Data almacena cantidades inmensas de datos con la finalidad de brindar soluciones que permiten trabajar con los datos en tiempo real o utilizar tecnologías de base de datos alternativas para manejarlo y procesarlo. No necesariamente están enfocados a las grandes organizaciones comerciales y de mercadotecnia, sino también a sectores de salud, como la medicina y el deporte, también a la física y otras disciplinas.

Según la información de (IIC, 2021), se pueden observar tres tipos de analíticas que implementan las organizaciones para la interpretación y aplicación de los datos para transformarlos en beneficios para sus empresas; además, se describen

las cuatro capas tecnológicas en las que el Big Data recibe, ordena y maneja los datos de las organizaciones.

- Analítica descriptiva, enfocado a cómo ha funcionado el negocio hasta el momento. Consiste en almacenar y realizar agregaciones de datos históricos visualizándolos de forma que puedan ayudar a la comprensión del estado actual y pasado del negocio.
- Analítica predictiva, es decir, cómo va a funcionar el negocio. Se construye sobre la analítica descriptiva y usa modelos estadísticos avanzados para estimar aquellos datos de negocio que son desconocidos o inciertos. Se usa en la predicción de valores futuros en series históricas de precios y demandas, en la clasificación automática de textos u operaciones o la segmentación de clientes.
- Analítica prescriptiva, el cómo se va a optimizar el negocio. Se trata del nivel más alto de analítica y rebasa las anteriores junto con estrategias de optimización para mejorar la gestión y estrategia de negocio. Indica qué acciones proporcionarán los mejores resultados. Esta se usa para obtener recomendaciones automatizadas sobre el momento idóneo para ejecutar pedidos, mantenimientos u otras operaciones cuantificables.

III.1 Capas tecnológicas

El instituto de ingeniería del conocimiento (IIC, 2021) reconoce que son cuatro las capas tecnológicas que se presentan para un proyecto de Big Data. Con información retomada de esta fuente se describen a continuación.

- Infraestructura: Lugar donde se procesan

los datos. Las plataformas cloud facilitan el desarrollo de proyectos de datos, independientemente de su volumen y requisitos de procesamiento.

- Almacenamiento: Forma en que se organizan los datos. Los recursos hardware y software permiten el almacenamiento de los datos, facilitando su acceso y disponibilidad.
- Procesamiento: Trabajar con los datos. Las herramientas de procesamiento de los datos dan la base tecnológica para operar con grandes volúmenes de información batch o flujos rápidos de datos en tiempo real.
- Análisis: Decidir qué se quiere hacer con los datos. Los algoritmos realizan el análisis de los datos, extrayendo información de valor para el negocio.

IV DATOS ESTRUCTURADOS Y NO ESTRUCTURADOS

Existen bases de datos relacionales y no relacionales en las que se almacenan datos estructurados y no estructurados. Estas bases de datos no utilizan el mismo tipo de lenguaje, para ello requieren de SQL y NoSQL. Siendo las primeras (bases de datos relacionales) las que se manejan con SQL en donde es necesario definir los esquemas antes de añadir datos. Puesto que la base de datos SQL debe conocer con antelación el tipo de datos que va a almacenar.

Las bases de datos relacionales se basan en el uso de tablas (también se las llama relaciones). Las tablas se representan gráficamente como una estructura rectangular formada por filas y columnas. Cada columna almacena información sobre una propiedad determinada de la tabla (se le llama también atributo), nombre, dni, apellidos, edad,... Cada fila posee una ocurrencia o ejemplar de la instancia o relación representada por la tabla (a las filas se las llama también tuplas), (Sánchez, 2004). La tabla siguiente es un ejemplo de ello.

Tabla 1 Representación de una tabla en el modelo relacional.

NOMBRE

atributo 1	atributo 2	atributo 3	atributo n	
valor 1,1	valor 1,2	valor 1,3	valor 1,n	← tupla 1
valor 2,1	valor 2,2	valor 2,3	valor 2,n	← tupla 2
.....
valor m,1	valor m,2	valor m,3	valor m,n	← tupla m

Fuente: (Sánchez, 2004).

Estas bases de datos administran, es decir, crean, modifican y eliminan, aunque también sirven para agregar, consultar y editar datos de cambios de diferentes tablas.

Ejemplos para bases de datos relacionales: DB2, Postgress, Oracle, MySQL, SQL y Server.

Por el contrario, las bases de datos NoSQL están diseñadas para que se puedan insertar datos sin un esquema predefinido. Esto convierte a las bases de datos NoSQL en herramientas más sencillas para realizar modificaciones en las aplicaciones en tiempo real, la integración del código es más fiable y su desarrollo es más rápido.

El aumento tanto en el tamaño de los archivos como en su cantidad ha incidido en el crecimiento del interés hacia tecnologías NoSQL; además, tomando en cuenta que las bases de datos relacionales presentan serios problemas en cuanto a escalabilidad en el manejo de la información se refiere, esto genera que a medida que aumentan los datos, el desempeño disminuye y se hacen menos intuitivas.

Es por ello que las grandes compañías que manejan altos volúmenes de información como Facebook y Twitter se dieron cuenta de que con la infraestructura que contaban no podrían manejar la cantidad de información que se proyecta para dentro de cinco años, por eso, son esas compañías las que están invirtiendo y creando soluciones NoSQL, (Castro Romero, González Sanabria, & Callejas Cuervo, 2012).

Las bases de datos NoSQL se crean por la necesidad de una mayor flexibilidad para almacenar información no estructurada como lo son documentos en formato PDF, Word, Excel, etc.; direcciones de correos electrónicos, mensajes de texto, geolocalización, audio, video, entre muchos otros. Entonces cuando el volumen de estos datos combinados es muy grande, se convierte en Big Data.

Big data requiere de esquemas de bases de datos altamente escalable y de rápida capacidad. Las bases de datos SQL no pueden proporcionar esto que necesita, pero esto no quiere decir que no sean útiles para el big data, sino que se usa a la par que las NoSQL como un complemento. Los tipos de bases de datos relacionales como Oracle, DB2 de IBM y SQL Server e incluso Microsoft Access, forman una estructura fundamental para el almacenamiento de datos y la gestión en la mayoría de las organizaciones en la actualidad.

Ambos tipos de almacenamiento (estructurado y no estructurado) son de gran utilidad, estos mantienen ciertas diferencias entre sí, lo que los

hace clasificar de distinta forma los datos y el tipo de información que se desee analizar. Para lograr un mayor entendimiento sobre estas diferencias, se ha retomado información de (Equipo Next, 2021), la cual se detalla a continuación.

Almacenamiento estructurado

1. Son archivos de tipo texto que se suelen mostrar en columnas y filas con títulos para que los datos se encuentren perfectamente etiquetados.
2. Gracias a su estructura ordenada, son los más fáciles de gestionar, tanto digital como manualmente.
3. Se almacenan en un data lake para posteriormente mostrar la información en herramientas de data analytics.

Almacenamiento no estructurado

1. Datos binarios que no tienen una estructura comprensible para el lenguaje humano. Es una desorganización masiva de varios objetos que no tienen valor hasta que se identifican y almacenan de manera organizada.
2. No se pueden almacenar en una herramienta de data tradicional, ya que sería imposible ajustarlos a las filas y columnas estandarizadas.
3. Las empresas que consiguen extraer la información de estos datos cuentan con una gran ventaja competitiva. Y es que, esta tipología de datos aporta una visión mucho más completa y profunda del mercado.

Las bases de datos permiten gestionar la información de forma ordenada para proporcionar a las empresas un repositorio de datos que pueden ser consultados y procesados en cualquier momento. Dependiendo del propósito de estas aplicaciones y del tipo de información almacenada, se decide utilizar SQL para aplicaciones clásicas con datos estructurados o NoSQL para aplicaciones que tienen que manejar volúmenes muy grandes de datos, estructurados y no estructurados como en Big Data.

V IMPORTANCIA DEL BIG DATA EN LAS ORGANIZACIONES

El valor del big data no es almacenar grandes cantidades de datos, sino el resultado de lo que se puede hacer con los datos almacenados. Analizar los datos permite descubrir respuestas anticipadas a problemas que podrían llegar en el futuro, además que permite minimizar los costos, el tiempo, brinda la

oportunidad de desarrollar nuevos productos y optimizar ofertas con base a lo que están demandando en tiempo real los consumidores, es decir, el big data es una herramienta que usan las organizaciones para tomar decisiones inteligentes. Las redes sociales como Facebook, Twitter, LinkedIn, etc., son de los más reconocidos para obtener datos además del Internet de las cosas, (Jiménez, 2014).

Una vez que se combinan grandes cantidades de datos con estudio de alta potencia, tienen la posibilidad de hacer labores en relación con los negocios como: establecer las razones de procedencia de fallos, inconvenientes y deficiencias casi en tiempo real, realizar ofertas basadas en los hábitos de compra de los consumidores, hasta identificar conductas fraudulentas con antelación, antes de que esto afecte a la organización.

El Big Data es utilizado por la mayor parte de industrias para detectar patrones y tendencias de forma rápida y precisa. Les ayuda a identificar las necesidades del mercado a través de los datos de los consumidores. Las organizaciones utilizan esta información para mejorar sus negocios, comprender las elecciones de los consumidores, hacer pruebas, pronósticos y saber cómo dirigirse a las audiencias clave.

Por ejemplo, EMC (compañía norteamericana en el área del almacenamiento de datos) utiliza una mezcla de bases de datos relacionales y no relacionales para analizar la percepción pública de la empresa y sus productos. Analizan espacios de opinión en internet, buscando menciones de la marca y sus productos, y evaluar si las referencias son positivas o negativas; para ello, EMC reúne los textos completos de todos los blogs y páginas Web de la compañía, y los compila en una versión de MapReduce, luego utiliza Hadoop para eliminar el código de marcado Web y las palabras no esenciales, y por último pasa la lista de palabras basadas en SQL en bases de datos, donde se realiza un análisis cuantitativo más profundo. (Castro Romero, González Sanabria, & Callejas Cuervo, 2012).

El uso del Big Data aumenta la productividad y la eficiencia debido a que los instrumentos procesan los datos de manera más instantánea y permite a los empleados desarrollar con mayor efectividad su trabajo; disminuye los costes, debido a que el incremento de la productividad puede darse por sentado (esto impacta positivamente en la productividad); permite la detección de fraudes y anomalías, debido a que detectan transacciones equivocadas o inconvenientes en las ocupaciones;

asegura una mayor velocidad y rapidez de venta además de permitir la optimización en la atención al consumidor debido a que posee más datos acerca de sus gustos y preferencias.

Cualquier dispositivo que sea capaz de almacenar y procesar información es una fuente generadora de datos. La organización de estos datos hará a que se conviertan en información útil para las empresas. Contenido que resulta útil analizar es el siguiente, el cual ha sido tomado de (Herrera, 2022).

- Contenido web obtenido de las redes sociales.
- M2M, el contenido que permite conectarse a otros dispositivos.
- Registros de facturas y detalles de llamadas.
- Información biométrica, como huellas digitales o reconocimiento facial.
- Información como correos electrónicos, notas de voz y llamadas telefónicas.

Independientemente de la forma en la que se clasifiquen, se pueden encontrar datos en todas partes: teléfonos móviles, tarjetas de crédito, aplicaciones de software, vehículos, registros, páginas web, etc. (IEBS, 2021).

VI EMPRESAS DE ÉXITO QUE USAN BIG DATA PARA LA OPTIMIZACIÓN DE SU NEGOCIO

Las necesidades de las grandes empresas con el paso del tiempo fueron incrementándose también hasta que se plantearon la pregunta de cómo sacar provecho de la enorme cantidad de datos que tenían. Al analizar toda la información que se recopilaba, podrían llegar a tener un mejor entendimiento de los consumidores, permitiéndoles crear estrategias personalizadas en base a esos datos con el objetivo de satisfacer mejor las necesidades de sus clientes. Existen muchas empresas dedicadas a la implementación de tecnologías y soluciones de big data, como es el caso de Twingo, es un proveedor líder de tecnología de servicios expertos, productos y proyectos para Big Data & BI Analytics, al servicio de empresas basadas en datos y nuevas empresas en el mercado israelí. Algunos de sus clientes más destacados son Amazon, Philips, Dataloop, Varonis, entre otros.

Todos esos datos y la información que se podía obtener de ellos se convirtieron en la nueva clave para tomar decisiones inteligentes y acertadas minimizando riesgos. Puesto que permite predecir el comportamiento de los consumidores y estar en el momento exacto en el que se quiera satisfacer alguna necesidad.

Las siguientes veinte empresas han implementado el Big Data para mejorar su posición en el mercado. Estas han sido tomadas del autor Bernard Marr, (Marr, 2016):

Walmart, Cern, Netflix, Rolls-Royce, Shell, Apixio, Lotus F1 Team, Facebook, LinkedIn, Microsoft, Airbnb, Ralph Lauren, Walt Disney, Google, Amazon, Uber, Apple, Twitter, IBM Watson y Caesars.

VII CONCLUSIONES

La implementación de los macrodatos ha ayudado a la humanidad a descubrir cosas que, de ser llevadas a cabo de formas más austeras, manuales o con poca tecnología, habrían tardado años en poder visualizar con tanta precisión. Gracias a la abundancia de herramientas físicas, técnicas y sistematizadas como sensores, micrófonos, cámaras, imágenes, etc., se pueden obtener estos datos clasificados y almacenados dentro de una base de datos que está controlada y resguardada por las organizaciones, quienes tienen a su disposición la valiosa información que son nuestros datos.

Sin embargo, no es tan sencillo hacer uso de esta gran herramienta que ha llevado a numerosas empresas al éxito. Es necesario tener el conocimiento especializado en sistemas de almacenamiento y procesamiento de datos para poder analizar estas grandes cantidades de datos. Desde personas, software informático especializado, lenguajes como SQL y NoSQL; herramientas como Python, Excel, etc. Implementar una solución alrededor de Big Data implica la integración de diversos componentes y proyectos que en conjunto forman el entorno necesario para analizar grandes cantidades de datos.

REFERENCIAS

- Beaulieu, A. (2009). *Learning SQL*. O'Reilly.
- Brand, W. (2015). *Data Science*. Canada: Lillian Pierson .
- Castro Romero, A., González Sanabria, J. S., & Callejas Cuervo, M. (2012). Utilidad y funcionamiento de las bases de datos NoSQL. *ingenieria@uptc.edu.co*, 21-32.
- DaSCI. (2021). *dasci.es/e*. Obtenido de *dasci.es/e*: <https://dasci.es/es/investigacion/unidades/ciencia-de-datos-y-big-data/>
- Datademia. (2021). *datademia.es/*. Obtenido de *datademia.es/*: <https://datademia.es/blog/que-es-sql>
- desarrolloweb. (2020). *desarrolloweb.com*. Retrieved from *desarrolloweb.com*: <https://desarrolloweb.com/home/lenguaje-sql>
- Equipo Next. (07 de 04 de 2021). *gruponext.es*. Obtenido de *gruponext.es*: <https://gruponext.es/blog/datos-estructurados-y-no-estructurados/>
- Herrera, T. (2022). *inspyrschool.es*. Obtenido de *inspyrschool.es*: <https://www.inspyrschool.es/big-data/que-es-y-para-que-sirve-el-big-data.html>

-
- Ibertech. (2016). *ibertech.org*. Obtenido de ibertech.org: <https://www.ibertech.org/las-4v-del-big-data/>
- IEBS. (2021). *iebschool.com*. Obtenido de iebschool.com: <https://www.iebschool.com/blog/valor-big-data/>
- IIC. (2021). *iic.uam.es*. Obtenido de iic.uam.es: <https://www.iic.uam.es/big-data/>
- IIC. (2022). *IIC*. Obtenido de IIC: <https://www.iic.uam.es/innovacion/aplicaciones-big-data-empresa/>
- Jiménez, C. M. (2014). Big data. Un nuevo paradigma de análisis de datos. *ETSI (ICAI)*, 16.
- López, R. (Marzo de 2022). *mvl.edu.ar*. Obtenido de mvl.edu.ar: <https://mvl.edu.ar/noticia/fablabvl-base-de-datos-sql/#:~:text=El%20lenguaje%20SQL%20sirve%20para,sobre%20las%20bases%20de%20datos.>
- Marr, B. (2016). *Big Data in practice*. UK: Wiley.
- Mayer, V. (2013). *Big data la revolución de los datos masivos*. Madrid: Turner Noema.
- MongoDB. (2021). *mongodb.com*. Obtenido de mongodb.com: <https://www.mongodb.com/es/nosql-explained>
- Oracle. (2021). *oracle.com*. Obtenido de oracle.com: <https://www.oracle.com/mx/database/what-is-database/>
- Power Data. (2020). *powerdata.e*. Obtenido de powerdata.e: <https://www.powerdata.es/big-data>
- Puyol, J. (2014). Una aproximación a Big Data. *UNED*, 36.
- Ramos, P. (14 de Septiembre de 2018). *styde*. Obtenido de styde: <https://styde.net/que-es-y-para-que-sirve-sql/>
- Robledano, Á. (2019). *openwebinars.net*. Obtenido de openwebinars.net: <https://openwebinars.net/blog/que-es-sql/>
- Sánchez, J. (2004). *Principios sobre bases de datos relacionales*. California: Creative Commons.
- Styde. (2018). *styde.net*. Obtenido de styde.net: <https://styde.net/que-es-y-para-que-sirve-sql/>
- Unir. (2021). *unir.net*. Obtenido de unir.net: <https://www.unir.net/marketing-comunicacion/revista/que-es-sql/>