



BIG DATA

La base de la automatización
inteligente de las empresas





BIG DATA, LA BASE DE LA AUTOMATIZACIÓN INTELIGENTE DE LAS EMPRESAS

***Resumen.** La gestión eficiente de grandes volúmenes, conocido como **BIG DATA**, está permitiendo que la inteligencia artificial se esté desarrollando en pleno, como soporte de las transformaciones en los modelos de negocio y las operaciones de las empresas de todo el mundo. En el presente artículo realiza una aproximación al estado del arte de los datos, en el ámbito global, profundiza en la importancia de la gestión eficiente de los mismos y su aprovechamiento en la era de la automatización inteligente de las empresas, mostrando las oportunidades que al respecto de la temática se presentan en entorno local, con prospectiva de internacionalización.*

INTRODUCCIÓN

La llegada de la cuarta revolución industrial, impulsada por grandes cambios tecnológicos, entre los que sobresale el crecimiento de la Internet, las capacidades de procesamiento y almacenamiento de datos y, con ello la masificación de la nube, han generado una gran explosión de información interna y externa en las organizaciones, que hoy está siendo utilizada como el recurso más valioso que ha transformado la realidad de los negocios. Todo esto conlleva a que las operaciones de las empresas puedan ser automatizadas, los procesos digitalizados y con ello la creación de nuevos modelos de negocio, para clientes cada vez más exigentes, que ya se transformaron digitalmente.

En el presente artículo trataremos precisamente acerca de las grandes transformaciones derivadas del crecimiento de la información en especial en la última década, y por ende la aparición del **BIG DATA**, el cual hace alusión a la recopilación de grandes cantidades de datos desde diferentes fuentes tanto internas como externas a las organizaciones, en diferentes formatos tanto estructurados, como no estructurados, su procesamiento a grandes velocidades, con el fin de dar un soporte más efectivo a las organizaciones en su visión de operación inteligente, facilitando la automatización de procesos y la toma de decisiones en tiempo real.

En la primera parte trataremos con algunos datos de referencia que permiten evidenciar el crecimiento exponencial que ha tenido la información y algunas bases conceptuales alrededor del término de **BIG DATA** y las condiciones para que realmente se pueda hablar de que está siendo aplicado. Seguidamente se profundizará en el ecosistema de tecnologías pertinentes para hacer posible este enfoque en las empresas y la importancia del rol conocido como *Data Science* en el aprovechamiento de del **BIG DATA** por parte de organizaciones de todo el mundo. La parte final del artículo profundiza acerca de cómo el **BIG DATA** se ha convertido en el soporte a la automatización inteligente de procesos en las empresas, la analítica avanzada y la inteligencia artificial. Se concluye con algunos retos que presenta el **BIG DATA** y su aplicación en los entornos tecnológicos contemporáneos.

EL CRECIMIENTO DE LA INFORMACIÓN Y LA MASIFICACIÓN DEL INTERNET



BIG DATA

La base de la automatización inteligente de las empresas

Los datos disponibles en nuestro entorno crecen cada año de manera exponencial, generándose cada vez mayores oportunidades que maximicen su aprovechamiento. De acuerdo con la Consultora Domo, en su Octava Edición de su infografía Los Datos Nunca Duermen: “en un (1) sólo minuto en internet en el inicio del año 2021 los usuarios de *Neflix* ven 404.444 horas de vídeo, los usuarios de *YouTube* suben 500 horas de vídeo, *Microsoft Teams* conecta a 52.083 usuarios, los usuarios de *Facebook* comparten 150.000 mensajes, los usuarios de *WhatsApp* comparten 41.666.667 mensajes, *Zoom* acoge a 208.333 participantes en reuniones, entre otros”¹, lo que demuestra la gran cantidad de datos disponibles, la diversidad de formatos y los avances cada vez más evidentes en velocidades y capacidades de procesamiento de información.

Sumado a lo anterior, aparece el fenómeno de la masificación de la Internet a nivel mundial, que según Matt Ahlgren, en su bloc de estadísticas de internet para el año 2021, “alcanza los 4.7 billones de usuarios en el año 2021, con usuarios que con promedios de conexión de seis (6) horas y 43 minutos todos los días y más de 1.83 mil millones de sitios en Internet, el mundo de e-commerce con proyecciones de \$ 4.9 billones en ventas en 2021”², el universo de las Fintech, el tráfico de transacciones financieras, entre otros, aportan con la creación de más y más

datos. Sumado a lo anterior, está la industria del internet de las cosas, que para el año 2025 se calcula “conectará más de 75 mil millones de objetos alrededor de todo el mundo”³ donde sensores enviarán permanente datos para ser procesados y gestionados en tiempo real, en sistemas de información disponibles en la internet.

Y es que la información sigue creciendo en el mundo a pasos agigantados año tras años. “Hasta 2003 la humanidad había generado cinco exabytes de información a lo largo de toda su historia. Lo dijo Eric Schmidt, CEO de *Google*, durante una conferencia en 2010. En 2007 se generaron 281 exabytes, según las investigadoras Hardy y Williams, y apenas cuatro años más tarde alcanzamos los 1.800 exabytes”⁴. Adicionalmente la consultora Domo en su infografía titulada “Los datos Nunca Duermen”⁵ mostraba las interacciones de los más de 3.800 millones de personas conectadas entonces a internet (4.300 millones a principios de 2019), y su interacción a lo largo de un minuto. Estas cifras ya se hayan quedado cortas, pero cada minuto sucedía esto:

- Se visionaban 97.222 horas de vídeo en *Netflix*.
- 4.333.560 vídeos de *YouTube*.
- Eran enviados 473.400 *tuits*.
- Se realizaban 176 mil llamadas de *Skype*.
- Eran colgadas en Instagram 49.380 fotografías.

¹ DOMO, Data Never Sleeps 8.0. Consultada enero 25 de 2021

<https://www.domo.com/learn/data-never-sleeps-8>

² Matt Ahlgren, Estadísticas de internet para el año 2021. Consultado en enero 27 de 2021 <https://www.websitehostingrating.com/es/internet-statistics-facts/>

³ Statista, Universidad de Galileo. Consultado en enero 27

<https://www.galileo.edu/trends-innovation/internet-de-las-cosas-75-mil-millones-de-objetos-conectados-en-2025/>

⁴ Orange, ¿Cuánta información se genera al año en el mundo?, Consultada febrero 27 de 2021, <https://blog.orange.es/red/datos-mundo/>

⁵ Domo, Data Never Sleeps 6.0. Consultada febrero 27 de 2021,

<https://www.domo.com/learn/data-never-sleeps-6>



BIG DATA

La base de la automatización inteligente de las empresas

de febrero⁷, se señala que “en los últimos tres meses del 2020 las conexiones en el país aumentaron en cerca de 200.000 conexiones fijas y 500.000 nuevos accesos móviles. Esto es tres veces más de lo logrado en todo el 2019, pues se registraron un total de 692.498 conexiones fijas nuevas en los primeros nueve meses del 2020”. Adicionalmente se confirmó la tendencia positiva y se avanza a la meta de tener la mitad de los hogares colombianos conectados a Internet fijo en 2022”. Seguidamente la ministra señaló que “Bogotá es la ciudad más conectada, pues presenta la mayor penetración de Internet fijo en el país con más de 20 conexiones a Internet por cada 100 habitantes”. Finalmente, se indicó en cuanto a Internet móvil, que se registró un crecimiento aproximado de un (1) millón de conexiones en el trimestre. Adicionalmente, continuó el aumento en el uso de la tecnología 4G y la disminución de la participación de 3G y 2G”.

Adicional a lo mencionado, la reciente pandemia por COVID-19 deja ver algunas brechas digitales en el país. Según Carolina Botero, en un reciente artículo publicado el pasado 11 de enero del año 2021 denominado “La pandemia y las brechas digitales⁸”, nos dimos cuenta de la importancia de la virtualidad y la tecnología, pero también de la necesidad de garantizar que nadie se quede atrás. El reto es reconocer la brecha digital y evitar que aumente la desigualdad en esta materia. “El índice de penetración de Internet fijo es apenas del 37,5%, lo cual significa que buena parte de las personas accede a través de celulares”. En el mismo artículo se indica que

según la Comisión Económica para América Latina y el Caribe (CEPAL), el 67% de las personas y el 60% de los hogares de la región cuentan con conexión a internet. Pero, mientras en las ciudades los hogares sin conexión representan el 33%, en las zonas rurales llegan al 77%. Otras estadísticas mencionadas allí muestran que “en las cabeceras municipales hay un índice de penetración de internet fijo del 45,75%, mientras que el sector rural disperso y los pequeños centros urbanos este índice apenas llega al 6,2%”; adicionalmente, “Colombia está en el promedio de la región: el 62% de su población usa internet. Pero el índice de penetración de internet fijo es apenas del 37,5%, lo cual significa que buena parte de las personas accede a través de celulares y, en su mayoría, con planes prepago.

No obstante, este panorama, es indudable la intención no sólo de Colombia, sino de todos los gobiernos del mundo por continuar cerrando las brechas digitales que sumadas a la cobertura en la conectividad deberán estar acompañadas de acciones que aseguren la calidad del acceso y su adecuado aprovechamiento en actividades educativas y empresariales que puedan ayudar al país, al mejoramiento de su economía y la calidad de vida de las personas. Y es precisamente en este último punto donde se hace necesario continuar en el desarrollo de perfiles profesionales competentes y conocimientos técnicos que permitan lograr ese mejor aprovechamiento.

⁷ Portafolio, El comportamiento de conectividad en Colombia en el tercer trimestre, consultado febrero 27, <https://www.portafolio.co/economia/el-comportamiento-de-conectividad-en-colombia-en-el-tercer-trimestre-548835>

⁸ Carolina Botero, La República, La pandemia y las brechas digitales, Consultado Febrero 27, <https://razonpublica.com/la-pandemia-las-brechas-digitales/>



BIG DATA

La base de la automatización inteligente de las empresas

BASES CONCEPTUALES DEL BIG DATA

Cuando nos referimos a **BIG DATA** hacemos alusión según Gartner a “Activos de información de gran volumen, alta velocidad y/o gran variedad que exigen formas rentables e innovadoras de procesamiento de información que permitan una mejor comprensión, toma de decisiones y automatización de procesos”⁹. El **BIG DATA** entonces supone el proceso de adquirir, almacenar, buscar, compartir, dar valor, analizar y respaldar datos.

Para que exista Big se deben dar como mínimo tres (3) condiciones básicas: mayor Volumen de datos, gran Variedad de formatos y una Velocidad superior. Al respecto de lo anterior, la Consultora Oracle nos amplía dichas definiciones en su artículo de Definición de **BIG DATA**:

- (1) **Volumen.** *La cantidad de datos importa. Con **BIG DATA**, tendrá que procesar grandes volúmenes de datos no estructurados de baja densidad. Puede tratarse de datos de valor desconocido, como feeds de datos de Twitter, flujos de clics de una página web o aplicación para móviles, o equipo con sensores. Para algunas organizaciones, esto puede suponer decenas de terabytes de datos. Para otras, incluso cientos de petabytes.*
- (2) **Velocidad.** *La velocidad es el ritmo al que se reciben los datos y (posiblemente) al que se utilizan. Por lo general, la mayor velocidad de los datos se transmite directamente a la memoria, en vez de escribirse en un disco. Algunos productos inteligentes habilitados para Internet funcionan en tiempo real o*

prácticamente en tiempo real y requieren una evaluación y actuación en tiempo real.

- (3) **Variedad.** *La variedad hace referencia a los diversos tipos de datos disponibles. Los tipos de datos convencionales eran estructurados y podían organizarse claramente en una base de datos relacional. Con el auge del **BIG DATA**, los datos se presentan en nuevos tipos de datos no estructurados. Los tipos de datos no estructurados y semiestructurados, como el texto, audio o video, requieren un preprocesamiento adicional para poder obtener significado y habilitar los metadatos.¹⁰*

LAS SIETE V'S DEL BIG DATA

Con el tiempo se han adicionado otros nuevos criterios a estas tres (3) características básicas mencionadas tales como el **Valor**, la **Veracidad**, la **Viabilidad** y la **Visualización** de la información disponible a nivel de estos grandes volúmenes de datos. Según el instituto de Ingeniería del Conocimiento en su artículo Las 7 V del **BIG DATA** las define como sigue:

- (4) **Valor.** *Cuando hablamos de veracidad nos referimos a la incertidumbre de los datos; es decir, al grado de fiabilidad de la información recibida.*
- (5) **Veracidad.** *Cuando hablamos de veracidad nos referimos a la incertidumbre de los datos; es decir, al grado de fiabilidad de la información recibida.*
- (6) **Viabilidad.** *La inteligencia empresarial es un componente fundamental para la*

⁹ Gartner, Glosario, Consultado marzo 1, <https://www.gartner.com/en/information-technology/glossary/>

¹⁰ Oracle, Definición de Big Dada, Consultado marzo 1 de 2021, <https://www.oracle.com/mx/big-data/what-is-big-data/>



BIG DATA

La base de la automatización inteligente de las empresas

viabilidad de un proyecto y el éxito empresarial. Se trata de la capacidad que tienen las compañías en generar un uso eficaz del gran volumen de datos que manejan.

- (7) **Visualización.** *Cuando hablamos de visualización nos referimos al modo en el que los datos son presentados. Una vez que los datos son procesados (los datos están en tablas y hojas de cálculo), necesitamos representarlos visualmente de manera que sean legibles y accesibles, para encontrar patrones y claves ocultas en el tema a investigar.¹¹*

Del entendimiento y gestión efectiva de cada uno de estas siete (7) características mencionadas dependerá el uso efectivo del **BIG DATA** como medio tecnológico con un aporte relevante en la toma de decisiones, la automatización de procesos y la inteligencia de negocios de las empresas, derivando seguramente en la generación de ventajas competitivas, que les permitan a las organizaciones alcanzar el éxito.

Para la implementación de este enfoque en las empresas será necesario gestionar la cultura interna subyacente para el aprovechamiento de la información, siendo necesario adaptar enfoques metodológicos, tecnologías y conocimientos pertinentes que permitan explotar esta nueva capacidad. En este sentido profundizaremos un poco más en esta sección acerca del origen de los datos, las metodologías y tecnologías requeridas para hacer **BIG DATA**.

¿DE DÓNDE VIENEN LOS DATOS?

Desde una perspectiva más amplia, y partiendo del contexto inicial del presente

artículo en el que se describe cómo ha crecido la información alrededor de todo el mundo en los últimos años, se podría inferir que al día de hoy la data se produce en muchos momentos, lugares y programas informáticos, entre los que sobresalen la información generada por personas, cuando por ejemplo se envía un correo electrónico, se contesta una encuesta, se toma una fotografía, se genera un hoja de cálculo o simplemente se usa *WhatsApp* o se publica información, a través de una red social.

De manera complementaria, los datos pueden ser generados de manera automática por una máquina o comunicación entre ellas, donde el *Machine Learning* ha entrado a operar muchos procesos en automático y por ende la generación de datos. Al respecto hablamos de sistemas tales como *GPS* para vehículos y teléfonos inteligentes, parquímetros automáticos, máquinas expendedoras de bienes, quioscos de autoservicio, contadores de electricidad, y en general la industrial de la Internet de las cosas IoT, donde a través de sensores son generadas grandes cantidades de datos con diferentes fines.

Finalmente, otras fuentes tales como los datos generados a través del *marketing web*, donde algoritmos predictivos permiten identificar los comportamientos humanos y generar ofertas cada vez más personalizadas e información generadas por transacciones financieras, que día a día permiten que el mundo pueda operar de manera efectiva para sus clientes y usuarios.

El BIG DATA tiene su sustento en la gestión efectiva de estos grandes volúmenes de información, a través de una recolección y procesamiento eficiente de datos internos y

¹¹ Instituto de Ingeniería del Conocimiento, Las 7 V del **BIG DATA**, Consultado en marzo 1,

<https://www.iic.uam.es/innovacion/big-data-caracteristicas-mas-importantes-7-v>



BIG DATA

La base de la automatización inteligente de las empresas

externos de la organización, desde diversas fuentes de datos, incorporando en un solo *datasets* la información proveniente desde tres (3) fuentes principales que nos permiten clasificar estos datos según su estructura. Nos referimos entonces a motores de bases de datos relacional tradicionales ya usados en las empresas, como soporte a sus sistemas de información internos (*ERP, CRM, etc*), e integrándolos con nuevas fuentes, tales como datos semiestructurados tipo *NoSQL (XML, JSON)*, e información no estructurada (imágenes, videos, audio, texto) provenientes de diversas fuentes como redes sociales páginas web, entre otros.

Al respecto de esta clasificación, según daremos algún detalle, de acuerdo con el artículo publicado por la Universidad Complutense de Madrid en su definición de **BIG DATA**:

- (1) **Datos estructurados.** Son los que tradicionalmente se han usado en el tratamiento de datos. Sus características principales son que pueden almacenar tablas y tienen una clara definición de longitud y formato. Entre ellos se encuentran, los números, cadena de caracteres, las fechas, entre otros.
- (2) **Datos Semiestructurados.** Siguen una especie de estructura, pero no es lo suficientemente regular como para gestionar como datos estructurados. Posee ciertos patrones comunes que los describen y dan información sobre las relaciones entre los mismos. Como, por ejemplo, el HTML, lenguaje para elaboración de páginas Web, donde un sistema de etiquetas permite detectar esas pautas comunes.

- (3) **Datos No estructurados.** Se trata de datos en su formato original, tal y como fueron recogidos. No poseen un formato específico que permita almacenarlos de la forma tradicional, pues no se puede desglosar la información que facilitan a tipos de datos definidos en longitud y formato. Entre ellos los más comunes son correos electrónicos, presentaciones multimedia, archivos en PDF, audios, entre otros.¹²

Lo anterior supone un avance en las arquitecturas tecnológicas de soporte a la analítica empresarial, evolucionando de modelos tradicionales de inteligencia de negocios con enfoque descriptivo y diagnóstico, soportados en bases de datos relacionales, en la cual se usan bodegas secundarias de almacenamiento de información denominadas *datawarehouse*, ingestas a partir de pequeños programas de extracción, transformación y carga de datos (*ETLs*), migrando a modelos de almacenamiento de datos distribuidos en depósitos evolucionados denominados *Data Lake*, en los que se guardan y procesan grandes volúmenes de datos y diferentes formatos de datos con gran desempeño, dando cabida a una analítica empresarial más predictiva y en muchos casos prescriptiva.

CICLO BÁSICO DE VIDA DE LOS DATOS

Entiéndase este ciclo como el proceso en sí mismo de captura, mantenimiento, síntesis, uso, publicación, archivo y purga de los datos. Al respecto de cada una de estas etapas, empresas, organizaciones y profesionales de todo el mundo recorren

¹² Facultad de Estudios Estadísticos, Universidad Complutense de Madrid, consultado en 01 de marzo de 2021,

<https://www.masterbigdataucm.com/que-es-big-data/>



BIG DATA

La base de la automatización inteligente de las empresas

este ciclo ejecutando actividades, tales como:

- (1) **Captura de datos.** Se refiere a la ingestión de información en los data set de **BIG DATA** de la empresa ya sea de datos existentes o producidos por una organización o traídos de fuentes externas. Lo anterior puede derivar del uso de programas informáticos internos, de la *Web* e incluso de tecnologías IoT.
- (2) **Mantenimiento de los datos.** Actividades que se ocupan de procesar los datos sin derivar aún ningún valor para la empresa. Involucra tareas como movimiento, integración, limpieza, enriquecimiento, captura de datos modificados, así como procesos familiares de extracción, transformación y carga (ETL's).
- (3) **Síntesis de los datos.** Implica tareas para la creación de valores de datos mediante lógica inductiva, utilizando otros datos como entrada.
- (4) **Uso de datos.** Implica tareas relacionadas con inteligencia de negocio y gobierno de datos por parte de las empresas.
- (5) **Publicación de datos.** Implica enviar información por fuera de la empresa. Situaciones de calidad en la publicación pueden afectar la reputación de la empresa.
- (6) **Archivo de datos.** Implica la creación de repositorios para el almacenamiento de datos, sin que ello implique la aplicación de tareas de mantenimiento, uso o publicación.
- (7) **Purga de datos.** Es la eliminación de cada copia de un elemento de datos de la empresa.

TECNOLOGÍAS DE SOPORTE AL **BIG DATA**

La implementación de **BIG DATA**, demanda el uso de tecnologías especializadas en diferentes momentos del ciclo de los datos.

Para las tareas de extracción y procesamientos de datos en las etapas tempranas del ciclo de datos relacionadas con la captura, mantenimiento y síntesis de los datos será necesario el aprovisionamiento de capacidades de almacenamiento y procesamiento de información distribuida. Para esto es típico el uso de herramientas de la Suite de Apache para **BIG DATA** con características de ser el mundo de código abierto entre las que sobresalen:

- (1) **Apache Hadoop.** La biblioteca de software Apache Hadoop es un marco que permite el procesamiento distribuido de grandes conjuntos de datos en grupos de computadoras utilizando modelos de programación simples. Está diseñado para escalar de servidores únicos a miles de máquinas, cada una de las cuales ofrece computación y almacenamiento locales.¹³
- (2) **Apache Spark.** Spark es un motor ultrarrápido para el almacenamiento, procesamiento y análisis de grandes volúmenes de datos. Es de código abierto y se encuentra gestionado por la Apache Software Foundation.¹⁴
- (3) **Apache Hive.** Solución de software que facilita la lectura, escritura y administración de grandes conjuntos de datos que residen en el almacenamiento distribuido mediante SQL. La estructura

¹³ Apache, Apache Hadoop, consultado marzo 1 de 2021, <https://hadoop.apache.org/>

¹⁴ Apache, Apache Spark, Consultado 1 de marzo de 2021, <https://spark.apache.org/>



BIG DATA

La base de la automatización inteligente de las empresas

se puede proyectar sobre los datos que ya están almacenados.¹⁵

Un ecosistema completo para **BIG DATA** normalmente se articula con motores de bases de datos tradicionales tales como *Oracle*, *SQL Server*, *My SQL* instalados localmente o en servicios dispuestos en la nube por proveedores tales como *AWS* o *Microsoft Azure* y sus servicios *AWS RDS* y *Azure*, respectivamente. En el caso de información semiestructurada es típico el uso de motores de bases de datos *NoSQL* donde son típicas tecnologías tales como *Mongo DB*, *Cassandra* y *Apache HBASE* y en la nube servicios como *Amazon DynamoDB* y *Azure Cosmos*.

En el caso de tecnologías nativas para **BIG DATA** como soporte a la implementación de las tecnologías de Apache mencionadas, sobresale el uso de Cloudera en entornos locales y en la nube servicios de proveedores como Amazon con sus servicios de *EMR* y de *Microsoft con HDInsight*.

Para el procesamiento de información en el marco estadístico sobresale como apoyo a la labor del *Data Science* sistemas de información como R y para la elaboración de modelos *Python*.

En la parte final de la cadena, como alternativas de sistemas de información de soporte a la visualización de información aparecen marcar como *Power BI* de *Microsoft*, *Tableau*, *Qlik* y *QuickSight* como servicio en la nube provisto por *Amazon*.

de diseñar, implementar, mantener y soportar los ecosistemas de **BIG DATA** de las empresas.

Para generar un mejor aprovechamiento de los ecosistemas de **BIG DATA** de las empresas, *los data Science* ejecutan sus propios procesos consistentes en las fases que serán descritas a continuación:



Ilustración 2. Ciclo de Vida de los Datos. Fuente Elaboración Propia

EL DATA SCIENCE Y EL BIG DATA

El rol de *data Science* en las empresas se ha convertido el principal usuario y responsable

¹⁵ Apache, Apache Hive, Consultado Marzo 1, <https://hive.apache.org/>

LA AUTOMATIZACIÓN DE PROCESOS EN EL MUNDO DEL BIG DATA

La base de la automatización inteligente de las empresas encuentra sustento en la



BIG DATA

La base de la automatización inteligente de las empresas

inteligencia artificial entendida como “la capacidad de una máquina para realizar funciones cognitivas que asociamos a la mente humana, como percibir, razonar, aprender, interactuar con el entorno y resolver problemas o incluso utilizar la creatividad”¹⁶. Estas funciones cognitivas pueden simular diferentes comportamientos humanos tales como conducir un vehículo, analizar datos y tomar decisiones, resolver problemas, reconocer voces, entre otros; pues son muchas las formas en las que una máquina podría llegar a imitar y en muchos casos reemplazar a las personas, teniendo en muchos casos un mejor desempeño.

LA INTELIGENCIA ARTIFICIAL Y EL **BIG DATA**

La inteligencia artificial entrena a las máquinas y las hace inteligentes, a partir de experiencias que permitan generalizar el conocimiento usando datos, de tal forma que éste pueda ser usado en contextos reales y sea imitando comportamientos humanos, o inclusive llegando a desarrollar capacidades cognitivas superiores. Para lograr esto, la Inteligencia Artificial se apoya en el uso de machine learning o aprendizaje automático, como su componente nuclear, para dotar a las máquinas de capacidades de aprendizaje, a través del uso de algoritmos matemáticos, soportados en técnicas tales como: árboles de decisión, modelos de regresión, clasificación y clusterización, como base de inteligencias básicas que imitan la mente humana y otras técnicas más complejas, como redes neuronales que dotan a las máquinas de capacidades cognitivas superiores, creando niveles más avanzados de inteligencia.

Las redes neuronales entonces son la base del *Deep Learning*, conocido también como aprendizaje profundo, el cual ha permitido que tecnologías como la traducción de idiomas en tiempo real, el procesamiento de lenguaje natural y el audio y voz, hoy en día sean consumidos como un servicio a costos muy razonables por empresas, desarrolladores de software y usuarios en todo el mundo, desde plataformas de proveedores de nube reconocidos como *Amazon, Google y Microsoft*.

Los avances más recientes en Inteligencia Artificial se han logrado aplicando el aprendizaje automático a conjuntos de datos muy grandes, y es acá donde aparece el **BIG DATA**. Los algoritmos de aprendizaje automático detectan patrones en los datos almacenados a través de tecnologías de **BIG DATA**, aprendiendo a hacer predicciones y recomendaciones, procesando información y experiencias, en lugar de recibir instrucciones explícitas de programación, por parte de desarrolladores de *software*. Los algoritmos también se adaptan en respuesta a nuevos datos y experiencias para mejorar la eficacia con el tiempo.

Es en este punto donde los datos toman un gran valor, pues de la cantidad y calidad de la información disponible, depende en gran medida, el éxito de los algoritmos de inteligencia artificial que se entrenan y aprenden a partir de la data. Con la llegada de la digitalización de procesos, el abaratamiento de los costos de procesamiento y almacenamiento de la información y el cambio de mentalidad por parte de las organizaciones para apreciar el valor de los datos, se ha entrado en una era

¹⁶ McKinsey. *An executive's guide to AI*. Consultada en enero 25 de 2021

<https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/an-executives-guide-to-ai>



BIG DATA

La base de la automatización inteligente de las empresas

de acumulación de éstos, los cuales en tiempo real ahora mismo son capturados, procesados y analizados hasta transformarse en conocimiento, lo que en últimas se ha denominado **BIG DATA**, permitiendo tener en la actualidad organizaciones más inteligentes, óptimas, con mejor desempeño financiero y diseño de mejores experiencia para sus clientes finales.

Los datos entonces se han convertido en el insumo principal que ha permitido que la automatización inteligente de procesos, sustentada en tecnologías de robotización, les permita a las máquinas moverse y adaptarse libremente, en diferentes entornos y los sistemas informáticos reemplazar las tareas y decisiones, antes soportadas en humanos, maximizando el desempeño de las operaciones, incrementando la eficiencia, incrementando la capacidad y reduciendo el riesgo de errores.

La automatización inteligente de las empresas entonces, aprovecha la información provista por el **BIG DATA** como base para una analítica avanzada, la cual haciendo uso del *machine learning*, resuelve problemas de negocio y automatiza las decisiones de organizaciones en todo el mundo, utilizando modelos matemáticos y capacidades de procesamiento superior que extraen conocimiento y patrones de los datos, como soporte a la inteligencia artificial, en la búsqueda permanente de las máquinas por imitar comportamientos humanos.

Es así como las soluciones en analítica empresarial con sustento en inteligencia artificial basadas en **BIG DATA** se están

convirtiendo en el sustento de la operación inteligente de las empresas. Este concepto según la firma McKinsey¹⁷ está siendo mayormente usado por las empresas como soporte a sus procesos de marketing y ventas, permitiendo llegar a nuevos mercados, atraer, mantener y retener a clientes digitales cada vez más exigentes y conscientes con sus compras queriendo ser tratados de manera personalizada.

Adicionalmente el concepto se ha extendido con mayor fuerza en temas de logística y manufactura inteligente, gestión de riesgos, servicio al cliente, desarrollo de producto, gestión financiera, gestión tecnológica, talento humano, entre otros. En otros sectores como salud, hoy es día es posible decodificar cadenas de ADN en poco tiempo.

La capacidad entonces de analizar, procesar y gestionar datos al ritmo que se producen, han convertido al **BIG DATA** en un medio imprescindible para los gobiernos y empresas en todo el mundo. En sector gobierno, por ejemplo, el concepto es ampliamente utilizado como soporte a ciudades inteligentes, donde los datos son indispensables en la optimización y el control del tráfico de vehículos, anticipar eventos adversos en función del clima, eficiencia energética, conocer la opinión de los ciudadanos en torno a diferentes temas e intentar predecir el impacto de las decisiones políticas y cómo es evidente en todo el mundo, para la administración de los eventos subyacentes a la aparición del **COVID – 19**.

La era de la Inteligencia Artificial y el **BIG DATA** apenas comienza, ofreciendo un mar de oportunidades aún por explorar en todos los ámbitos para las empresas y

¹⁷ McKinsey, The real-world potential and limitations of artificial intelligence, Consultado en enero 27 de 2021

<https://www.mckinsey.com/featured-insights/artificial-intelligence/the-real-world-potential-and-limitations-of-artificial-intelligence>



BIG DATA

La base de la automatización inteligente de las empresas

de procesamiento para generar un mayor nivel de confianza y aprovechamiento de la información.

- El almacenamiento seguirá creciendo de manera exponencial, lo cual demandará mucha más infraestructura para alojar los data sets de las empresas.
- A más data, mayores demandas de recursos de procesamiento que permitan aprovechar la información existente en estos grandes data sets de datos.
- La calidad de los datos puede afectar el desempeño de los algoritmos que aprovechan la data en las tareas de automatización de procesos y toma inteligente de decisiones.
- El conocimiento y experiencia técnica de las personas de soporte al **BIG DATA** en todo el ciclo de vida de los datos actualmente presenta brechas importantes que requieren de la articulación de la empresa, el estado y las universidades para su atención.

los administradores de datos y los consumidores de datos en una organización²⁰

Lo anterior dicho por el mismo Gartner generará un gran cambio cultural en la forma en que las empresas gestionan y comparten sus datos con otras organizaciones en todo el mundo. DataOps como marco de trabajo busca agilizar y reducir el tiempo de salida al mercado de proyectos de analítica avanzada.



PROSPECTIVA

En la era reciente del **Big Data**, reconocidas consultoras como Gartner han visualizado una nueva era para este enfoque tecnológico en el cual los datos puedan transarse a través de todo el mundo como soporte a la inteligencia artificial en un enfoque de servicios, lo que ha sido denominado por esta forma consultora como DataOps.

Gartner define Data Ops cómo *“una práctica de gestión de datos colaborativa centrada en mejorar la comunicación, la integración y la automatización de los flujos de datos entre*

Ilustración 4. Framework de Data Ops, Gartner 2019.

Este marco de trabajo ya ha sido utilizado por empresas como Telefónica con excelentes resultados descritos por la misma compañía como sigue: “Este ejercicio permite disponer una visión end-to-end que permita definir la estrategia de implantación de analytics dentro de una gran compañía, evaluando el estado actual, y decidiendo qué componentes evolucionar o implantar para conseguir este objetivo”²¹

²⁰ Gartner, Data Ops, consultado marzo 3 de 2021, <https://www.gartner.com/en/information-technology/glossary/dataops>

²¹ Telefónica, DataOps: Del laboratorio a producción de manera ágil, consultado marzo 3

de 2021, <https://empresas.blogthinkbig.com/dataops-del-laboratorio-produccion-de/>



BIG DATA

La base de la automatización inteligente de las empresas

CONCLUSIÓN

El **BIG DATA** se considera como un elemento esencial en el presente y futuro de las empresas de todo el mundo. Los datos y su gestión inteligente son el soporte a la automatización inteligente de procesos, la analítica avanzada y la inteligencia artificial, base de la transformación digital que se está viendo en la empresa contemporánea.

En un mundo globalizado, el **BIG DATA** y en general las tecnologías alrededor de la gestión efectiva de los datos desde su producción hasta su entrega final al cliente para su uso, abren un campo de oportunidades hasta ahora inimaginables para el profesional del hoy. Es así como, en el año 2019, la profesión de Científico de Datos fue considerada por los expertos como la profesión “trendy”.

Y es que, según el artículo publicado por la universidad de Galileo, “En una era en la que se generan y almacenan grandes cantidades de datos, dando lugar a información relevante para un negocio, una de las profesiones que más demanda tendrá será la de expertos en analizar este tipo de datos: los científicos de Datos”²²

En el ámbito nacional, específicamente en la ciudad de Medellín, Ruta N como impulsora del centro para la cuarta revolución industrial ha considerado el **BIG DATA** como una de sus tecnologías base y de gran

²² Universidad de Galileo, Expertos afirman que el data scientist es una profesión “trendy”, consultado marzo 3 de 2021, <https://www.galileo.edu/trends-innovation/expertos-afirman-que-el-data-scientist-es-una-profesion-trendy/>

potencial para el desarrollo de nuevos negocios en el ámbito tecnológico en la ciudad.²³

El **BIG DATA** entonces, se está convirtiendo en una verdadera opción para profesionales en todo el planeta, pues en los datos y su riqueza estructural se encuentran las soluciones para dar una respuesta efectiva a necesidades y problemas de las personas, empresas y gobiernos en todo el mundo.

GLOSARIO

- *IAs* – Inteligencias Artificiales
- *ERP* – Enterprise Resource Planning
- *CRM* – Customer Relationship
- *Machine Learning* – Aprendizaje de Máquinas
- *Deep Learning* – Aprendizaje Profundo
- *ETL* – Extracción, transformación y carga
- *Datawarehouse* – Depósito para almacenamiento de datos relacionales
- *Data Lake* – Depósito de almacenamiento de información en una arquitectura distribuida **BIG DATA**.

Por:

Jhon Velásquez
Ingeniero Industrial UdeA
Docente en Universidad EAFIT
CEO M&T Consulting S.A.S.

²³ Ruta N, Medellín Digital, Big Data, consultado marzo 3 de 2021, <https://www.medellindigital.gov.co/index.php/cuarta-revolucion-industrial/62-big-data>



www.arrobamedellin.edu.co

