

**IV ACUERDO MARCO PARA LA COMPETITIVIDAD E INNOVACIÓN
EMPRESARIAL DE CASTILLA Y LEÓN**

BIG DATA

¿QUÉ ES BIG DATA?

El término "big data" se refiere al análisis masivo de datos. Una cuantía de datos que son tan grandes, rápidos o complejos que es difícil o imposible procesarlos con los métodos tradicionales que se venían usando tales como bases de datos relacionales y estadísticas convencionales o paquetes de visualización, porque no son capaces de capturar, tratar y poner en valor dentro del tiempo necesario para que sean útiles.

Aunque el tamaño utilizado para determinar si un conjunto de datos determinado se considera Big Data no está firmemente definido y sigue cambiando con el tiempo, la mayoría de los analistas y profesionales actualmente se refieren a conjuntos de datos que van desde 30-50 Terabytes a varios Petabytes.

La naturaleza compleja del Big Data se debe principalmente a la naturaleza no estructurada de gran parte de los datos generados por las tecnologías modernas, como los web logs (Blog), la identificación por radiofrecuencia (RFID), los sensores incorporados en dispositivos, la maquinaria, los vehículos, las búsquedas en Internet, las redes sociales como Facebook, computadoras portátiles, teléfonos inteligentes y otros teléfonos móviles, dispositivos GPS y registros de centros de llamadas.

Necesitan aplicaciones informáticas no tradicionales de procesamiento de datos para tratarlos adecuadamente.

El uso moderno del término "big data" tiende a referirse al análisis del comportamiento del usuario, extrayendo valor de los datos almacenados, y formulando predicciones a través de los patrones observados. La disciplina dedicada a los datos masivos se enmarca en el sector de las tecnologías de la información y la comunicación. Esta disciplina se ocupa de todas las actividades relacionadas con los sistemas que manipulan grandes conjuntos de datos.



CARACTERÍSTICAS Y DESAFÍOS DE LA CALIDAD DE DATOS EN BIG DATA

Las especiales características del Big Data hacen que su calidad de datos se enfrente a múltiples desafíos. Sus características se acuñan popularmente como las «V» del Big Data, por comenzar todas estas propiedades por dicha letra del alfabeto. No hay un consenso sobre cuántas «V» han de ser tomadas en consideración y de hecho la lista de estas se ha ido ampliando, pero las 7 «V» del Big Data más extendidas son:

Volumen → Las organizaciones recopilan datos de diversas fuentes como transacciones comerciales, dispositivos inteligentes, equipo industrial, vídeos, medios sociales y más. La cantidad de datos generados está aumentando. Según crecen las bases de datos, también lo han de hacer las aplicaciones y arquitectura construida para soportar la recogida y almacenamiento de datos cada vez más variados. Además, se han reducido los costes de almacenamiento propiciando almacenar grandes cuantías de información a un precio mucho más reducido que antiguamente.

Velocidad → Los datos llegan a las empresas a una velocidad sin precedentes y deben ser manejados de manera oportuna. Nuestro concepto de inmediatez ha cambiado en los últimos tiempos y se busca información que llegue prácticamente al instante. Noticias que no llegan al día de antigüedad, en simple cuestión de horas y, en ocasiones, minutos, pueden haber perdido interés. Así, la velocidad de análisis requerida por la sociedad actual es una de las características fundamentales que tienen los datos a gran escala, donde los datos en constante movimiento procesados a tiempo real cobran protagonismo, ejecutando algoritmos cada vez más complejos en menos tiempo.

Variiedad → Los datos de presentan en todo tipo de formatos: desde datos numéricos estructurados en bases de datos tradicionales hasta documentos de texto no estructurados, correos electrónicos, vídeos, audios, datos de teletipo y transacciones financieras. Esta entrada en escena con fuerza de los datos no estructurados requiere nuevos tratamientos de la información, necesitando de nuevas metodologías y tecnologías para poder ser analizadas

Variabilidad → Los flujos de datos son impredecibles, cambian a menudo y varían mucho. Las empresas necesitan saber cuándo algo está de moda en los medios sociales y cómo gestionar los picos de carga de datos diarios, estacionales y desencadenados por eventos. Como la información varía mucho también han de hacerlo los modelos o tratamientos que se aplican en torno a esta, pues no son fijos en el tiempo y requieren de un control periódico.

Veracidad → Se refiere a la calidad de los datos. Debido a que los datos provienen de tantas fuentes diferentes, es difícil vincular, comparar, limpiar y transformar

los datos a través de los sistemas. Las empresas necesitan conectar y correlacionar las relaciones, las jerarquías y los múltiples vínculos de datos. Saber la fiabilidad de la información recogida es importante para obtener unos datos de calidad e, incluso, dependiendo de las aplicaciones que se le vaya a dar a misma, se convierte en fundamental. Es un factor que puede influir mucho en conseguir una ventaja competitiva en la explotación del Big Data.

Visualización



Convertir cientos de hojas de información en un único gráfico que muestre

claramente unas conclusiones predictivas es un ejemplo de cómo mostrar los resultados de forma clara y sencilla en un evidente ejercicio de sinterización.



Valor



Una gran cuantía de datos frecuentemente extrae pequeñas informaciones de valor. Cómo conseguir dicha información de manera eficiente es uno de los retos que afronta día a día el área de la inteligencia de datos. El valor es sin duda una cualidad fundamental en el análisis.

Estas 7 características del big data provocan que las empresas tengan problemas para extraer datos reales y de alta calidad, de conjuntos de datos tan masivos, cambiantes y complicados.

Hasta la llegada del Big Data se podía cargar la información estructurada que estaba almacenada en nuestro sistema. Pero ahora, se puede cargar información adicional que ya no se encuentra dentro de los dominios de la empresa: comentarios o likes en redes sociales, resultados de campañas de marketing, datos estadísticos de terceros, etc. Todos estos datos ofrecen información que ayuda a saber si los productos o servicios están funcionando bien o por el contrario están teniendo problemas.

Algunos desafíos a los que se enfrenta la calidad de datos de Big Data son:

1. Muchas fuentes y tipos de datos

Con tantas fuentes, tipos de datos y estructuras complejas, la dificultad de integración de datos aumenta.

2. Tremendo volumen de datos

El volumen de datos es enorme, y eso complica la ejecución de un proceso de calidad de datos dentro de un tiempo razonable.

Es difícil recolectar, limpiar, integrar y obtener datos de alta calidad de forma rápida. Se necesita mucho tiempo para transformar los tipos no estructurados en tipos estructurados y procesar esos datos.

3. Mucha volatilidad

Los datos cambian rápidamente y eso hace que tengan una validez muy corta. Para solucionarlo necesitamos un poder de procesamiento muy alto. Si no lo hacemos bien, el procesamiento y análisis basado en estos datos puede producir conclusiones erróneas, que pueden llevar a cometer errores en la toma de decisiones.

4. No existen estándares de calidad de datos unificados

En 1987 la Organización Internacional de Normalización (ISO) publicó las normas ISO 9000 para garantizar la calidad de productos y servicios. Sin embargo, el estudio de los estándares de calidad de los datos no comenzó hasta los años noventa, y no fue hasta 2011 cuando ISO publicó las normas de calidad de datos ISO 8000.

Estas normas necesitan madurar y perfeccionarse. Además, la investigación sobre la calidad de datos de big data ha comenzado hace poco y no hay apenas resultados.

La calidad de datos de big data es clave, no solo para poder obtener ventajas competitivas sino también impedir que incurramos en graves errores estratégicos y operacionales basándonos en datos erróneos con consecuencias que pueden llegar a ser muy graves.

¿POR QUÉ EL BIG DATA ES TAN IMPORTANTE?

La recopilación de grandes cantidades de datos y la búsqueda de tendencias dentro de los datos permiten que las empresas se muevan mucho más rápidamente, sin problemas y de manera eficiente. También les permite eliminar las áreas problemáticas antes de que los problemas acaben con sus beneficios o su reputación.

La importancia del big data no gira entorno a la cantidad de datos que se tienen, sino en lo que se hace con ellos. El análisis de Big Data ayuda a las organizaciones a aprovechar sus datos y utilizarlos para identificar nuevas oportunidades. Eso, a su vez, conduce a movimientos de negocios más inteligentes, operaciones más eficientes, mayores ganancias y clientes más felices.

Se pueden tomar datos de cualquier fuente y analizarlos para encontrar respuestas que permitan:

1. Reducir los costos.

Las grandes tecnologías de datos, como Hadoop (entorno de trabajo para software, bajo licencia libre, para programar aplicaciones distribuidas que manejen grandes volúmenes de datos), y el análisis basado en la nube, aportan importantes ventajas en términos de costes cuando se trata de almacenar grandes cantidades de datos, además de identificar maneras más eficientes de hacer negocios.

2. Reducir el tiempo.

Con la velocidad de Hadoop y la analítica en memoria (todos los datos en el ordenador están almacenados dentro de su memoria RAM con lo que es posible acceder a ellos de manera instantánea), combinada con la capacidad de analizar nuevas fuentes de datos, las empresas pueden analizar la información inmediatamente

3. Desarrollar nuevos productos y optimizar ofertas.

Con la capacidad de medir las necesidades de los clientes y la satisfacción a través de análisis viene el poder de dar a los clientes lo que quieren. Con la analítica de Big Data, más empresas están creando nuevos productos para satisfacer las necesidades de los clientes.

4. Tomar decisiones inteligentes.

Tomar decisiones basadas en lo que han aprendido.

Por ejemplo:

- **Turismo:** Mantener felices a los clientes es clave para la industria del turismo, pero la satisfacción del cliente puede ser difícil de medir, especialmente en el momento oportuno. Resorts y casinos, por ejemplo, sólo tienen una pequeña oportunidad de dar la vuelta a una mala experiencia de cliente. El análisis de Big data ofrece a estas empresas la capacidad de recopilar datos de los clientes, aplicar análisis e identificar inmediatamente posibles problemas antes de que sea demasiado tarde.
- **Cuidado de la salud:** El Big Data aparece en grandes cantidades en la industria sanitaria. Los registros de pacientes, planes de salud, información de seguros y otros tipos de información pueden ser difíciles de manejar, pero están llenos de información clave una vez que se aplican las analíticas. Es por eso que la tecnología de análisis de datos es tan importante para el cuidado de la salud. Al analizar grandes cantidades de información - tanto estructurada como no estructurada - rápidamente, se pueden proporcionar diagnósticos u opciones de tratamiento casi de inmediato.
- **Retail:** El servicio al cliente ha evolucionado en los últimos años, ya que los compradores más inteligentes esperan que los minoristas comprendan exactamente lo que necesitan, cuando lo necesitan. El Big Data ayuda a los minoristas a satisfacer esas demandas. Armados con cantidades interminables de datos de programas de fidelización de clientes, hábitos de compra y otras fuentes, los minoristas no sólo tienen una comprensión profunda de sus clientes, sino que también pueden predecir tendencias, recomendar nuevos productos y aumentar la rentabilidad.

- **Empresas manufactureras:** Estas despliegan sensores en sus productos para recibir datos de telemetría. A veces esto se utiliza para ofrecer servicios de comunicaciones, seguridad y navegación. Esta telemetría también revela patrones de uso, tasas de fracaso y otras oportunidades de mejora de productos que pueden reducir los costos de desarrollo y montaje.

- **Publicidad:** La proliferación de teléfonos inteligentes y otros dispositivos GPS ofrece a los anunciantes la oportunidad de dirigirse a los consumidores cuando están cerca de una tienda, una cafetería o un restaurante. Esto abre nuevos ingresos para los proveedores de servicios y ofrece a muchas empresas la oportunidad de conseguir nuevos prospectos.

- **Otros ejemplos del uso efectivo de Big Data** existen en las siguientes áreas:
 - Uso de registros de logs de TI (registros detallados de eventos y actividades que ocurren en un sistema informático) para mejorar la resolución de problemas de TI (tecnología de la información), así como la detección de infracciones de seguridad, velocidad, eficacia y prevención de sucesos futuros.
 - Uso de la voluminosa información histórica de un Call Center de forma rápida, con el fin de mejorar la interacción con el cliente y aumentar su satisfacción.
 - Uso de contenido de medios sociales para mejorar y comprender más rápidamente el sentimiento del cliente y mejorar los productos, los servicios y la interacción con el cliente.
 - Detección y prevención de fraudes en cualquier industria que procese transacciones financieras online, tales como compras, actividades bancarias, inversiones, seguros y atención médica.
 - Uso de información de transacciones de mercados financieros para evaluar más rápidamente el riesgo y tomar medidas correctivas.

Cuando se combinan grandes datos con análisis de alta potencia, se pueden realizar tareas relacionadas con los negocios como:

- Determinar las causas de origen de fallos, problemas y defectos casi en tiempo real.
- Generar cupones en el punto de venta basados en los hábitos de compra del cliente.
- Recalcular portafolios de riesgo completo en minutos.
- Detectar el comportamiento fraudulento antes de que afecte a la organización.

CICLO DE GESTIÓN DE INFORMACIÓN

Antes de que las empresas puedan poner a trabajar el big data, deben considerar como fluyen entre una multitud de lugares, fuentes, sistemas, propietarios y usuarios. Hay cinco pasos clave para hacerse cargo de este gran centro de datos:

1º. Establecer una gran estrategia de datos.

Es un plan diseñado para ayudar a supervisar y mejorar la forma en que adquiere, almacena, gestiona, comparte y utiliza los datos de dentro y fuera de la organización. Establece el escenario para el éxito del negocio en medio de una abundancia de datos. Al desarrollar la estrategia es importante considerar las metas e iniciativas de negocio y tecnología existentes y futuras.

2º. Identificar fuentes de big data.

- **El streaming de datos.**

Proviene del Internet de las Cosas IoT (red colectiva de dispositivos conectados y tecnología que facilita la comunicación entre los dispositivos y la nube, así como entre los propios dispositivos) y otros dispositivos conectados que fluyen a los sistemas de TI, Tecnología de la Información, (Plataformas tecnológicas que permiten la comunicación de usuarios, servicios, dispositivos y aplicaciones para el intercambio de información).

Pueden analizar este big data a medida que llegan, decidiendo que datos guardar o no y cuáles necesitan un análisis más profundo.

- **Los datos de los medios sociales.**

Proviene de interacciones en las redes sociales (Facebook, YouTube, Instagram...). Esto incluye big data en forma de imágenes, vídeos, voz, texto y sonido útiles para la comercialización, las ventas y las funciones de apoyo.

- **Los datos disponibles.**

Los públicos provienen de cantidades masivas de fuentes de datos abiertos como el data.gov del gobierno de Estados Unidos, el World Factbook de la CIA o el Open Data Portal de la Unión Europea

- **Otros big data.**

Pueden provenir de data lakes o lago de datos, que es un sistema o biblioteca de datos almacenados en su formato bruto usualmente blobs (Objetos binarios grandes) elementos utilizados en las bases de datos para almacenar datos de gran tamaño que cambian de forma dinámica o ficheros, fuente de datos en la nube, proveedores y clientes.



3º. Acceder, gestionar y almacenar los datos.

Los sistemas informáticos modernos proporcionan la velocidad, la potencia y la flexibilidad necesarias para acceder rápidamente a cantidades masivas y tipos de big data. Además de un acceso confiable, las empresas también necesitan métodos para integrar los datos, asegurar la calidad de los mismos, proporcionar un gobierno y un almacenamiento de datos, y preparar los datos para el análisis. Algunos datos pueden almacenarse en data warehouse tradicional (almacén de datos), pero también hay opciones flexibles y de bajo costo para almacenar y manejar big data a través de soluciones de nube, lagos de datos y Hadoop.

4º. Analizar el big data.

Con tecnologías de alto rendimiento como la computación en red o analítica en memoria, las organizaciones pueden optar por utilizar todos sus big data para análisis. Otro enfoque es determinar por adelantado qué datos son relevantes antes de analizarlos. De cualquier manera, el análisis de big data es cómo las empresas obtienen valor e información a partir de los datos.

5º. Tomar decisiones inteligentes basadas en datos

Los datos bien administrados y confiables conducen a un análisis y decisiones confiables. Para seguir siendo competitivas, las empresas necesitan aprovechar todo el valor del big data y operar de una manera basada en los datos, tomando decisiones basadas en la evidencia presentada por los grandes datos en lugar de en el instinto visceral. Los beneficios de ser impulsado por los datos son claros. Las organizaciones basadas en datos se desempeñan mejor, son operacionalmente más predecibles y son más rentables.

Los datos por sí mismos no garantizan conocimiento. Sin un análisis y un tratamiento adecuado, no sirven para nada. Así el valor estará no en los propios datos, sino en la relación de estos entre sí.

Estas relaciones son las que nos permiten extraer patrones que construyen el conocimiento en múltiples ámbitos que abarcan casi todos los campos imaginables. El valor puede ser una visualización de un gráfico donde se haga un análisis predictivo, una recomendación de un artículo relacionado en un portal de comercio electrónico, un cliente que adquiera un producto concreto, etc.

TIPOS DE DATOS SEGÚN SU ESTRUCTURA

Existen muchísimas herramientas para el manejo de big data. Algunos ejemplos incluyen Apache Hadoop, NoSQL, Apache Cassandra, inteligencia empresarial, aprendizaje automático y MapReduce. Estas herramientas tratan con algunos de los tres tipos de big data según su estructura:

- **Datos estructurados:**

Son los que tradicionalmente se han usado en el tratamiento de datos. Sus características principales son que se pueden almacenar en tablas y tienen una clara definición de longitud y formato. Entre ellos se encuentran las fechas, los números o las cadenas de caracteres. Un ejemplo son las bases de datos relacionales y los almacenes de datos.

La irrupción gracias al big data de otros tipos de datos no significa que estos no tengan su importancia.

- **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 forma tradicional dentro de una tabla ya que no se puede desglosar su información a tipos básicos de datos definidos en longitud y formato. Algunos ejemplos son los archivos en formato PDF, documentos multimedia como los PowerPoint, correos electrónicos o documentos de procesadores de texto.

- **Datos semiestructurados:**

Estos datos no se limitan a campos determinados, pero contienen marcadores para separar los diferentes elementos. Siguen una especie de estructura, pero no es lo suficiente regular como para gestionarla como datos estructurados.

Poseen sus propios metadatos semiestructurados que describen los objetos y las relaciones entre ellos. Como ejemplos tenemos los archivos tipo hojas de cálculo, el HTML lenguaje para la elaboración de páginas web donde su sistema de etiquetas permite detectar esas pautas comunes, el XML o el JSON.

Datos estructurados	Datos semiestructurados	Datos no estructurados
Fichas de clientes Fecha de nacimiento Nombre Dirección Transacciones en un mes Puntos de compra	Correos electrónicos Parte estructurada: destinatario, receptores, tema Parte no estructurada: cuerpo del mensaje	Persona a persona Comunicaciones en las redes sociales Persona a máquina Dispositivos médicos Comercio electrónico Ordenadores, móviles Máquina a máquina Sensores, dispositivos GPS Cámaras de seguridad

BENEFICIOS DEL BIG DATA

Y todo ese proceso ¿qué beneficios aporta?
Algunos de ellos son:

- Es un conjunto de tecnologías muy ágiles y flexibles.
- Es escalable por lo que, si hay más datos, se pueden poner más máquinas e incluir nuevas analíticas y más información. Así, la infraestructura puede crecer al mismo tiempo que la empresa o proyecto.
- Es rápido y económico respecto a la infraestructura de almacenamiento. Por ejemplo, a través de servicios de computación en la nube, se puede pagar únicamente por lo que se consume.
- La mayoría de aplicaciones pertenecen a la comunidad Open Source, software de código abierto o que forma parte del dominio público. Con esto, además del ahorro económico evidente que nuevamente entra en escena, también hay a nuestra disposición una gran comunidad de desarrolladores que trabajan diariamente en mejorar sistemas y procesos de estos recursos, como la comunidad Open Source de Apache, donde también resuelven dudas que surgen a diversos usuarios en todo el mundo.
- Y para el final, la gran ventaja: a través del Big Data y habiendo seguido correctamente el ciclo de gestión de información que culmina con la puesta en valor de los datos, se habrá conseguido un objetivo
- previamente marcado. Este puede ser ganar más dinero o ahorrarlo a través de mejoras de costes, adquisición de nuevos clientes o retención
- de los actuales, avanzar en una investigación científica, obtener una ventaja competitiva respecto a nuestros competidores anticipándonos a lo que deparará el mercado a través de análisis predictivos, optimizar nuestros procesos de producción...

En definitiva, se pueden sacar conclusiones con una base más sólida y unos conceptos que se orienten a la toma de decisiones efectivas. Y todo ello, aplicando no únicamente variables del pasado, sino predicciones a futuro mucho más fundamentadas en una base científica que hasta hace unos pocos años.

BENEFICIOS DEL **BIG DATA ANALYTICS** PARA TU NEGOCIO



Afina tus segmentos para optimizar los resultados de tus campañas digitales



Descubre tendencias del mercado.



Comprende las necesidades de tu consumidor.



Mide el rendimiento de tus estrategias comerciales.



Analiza datos para establecer precios competitivos.



Multiplica la adquisición y retención de clientes.



Estudia el mercado para el lanzamiento de nuevos productos.



Perfecciona el proceso de toma de decisiones.



Reduce los riesgos comerciales.



Genera eficiencias en la cadena de suministros