

**Boletín trimestral #3**  
SO WHAT?  
editado por de la Riva group



**Pat Gelsinger**

Los datos son la nueva ciencia.  
El Big Data son las respuestas



**Christian Gardiner**

Los datos son el activo estratégico  
de las compañías por excelencia.

# TODO LO QUE DEBES SABER SOBRE BIG DATA

**BIG DATA:** la capacidad de producir correlaciones como resultado de almacenar grandes cantidades de datos de diversos tipos y orígenes, que se producen a gran velocidad y que requieren de un tratamiento en tiempo real.

PARTAMOS DE LA PREGUNTA MÁS BÁSICA: ¿QUÉ ES UN DATO? LA PALABRA VIENE DEL LATÍN *DATUM*, QUE SIGNIFICA «LO QUE SE DA», Y SE REFIERE A UNA REPRESENTACIÓN SIMBÓLICA DE UN ATRIBUTO O VARIABLE CUALITATIVA O CUANTITATIVA. ES UNA FRACCIÓN DE INFORMACIÓN QUE SE PUEDE ANALIZAR, COMBINAR, ALMACENAR O TRANSMITIR. PODRÍA DECIRSE QUE UN DATO ES, A LA INFORMACIÓN, LO QUE UNA MOLÉCULA A LA MATERIA.

«Big Data trata de entender las relaciones que te unen a las personas que te interesan y el conocimiento del potencial de esa relación»  
Joe Rospars.

## DATOS EN EL MUNDO DIGITAL:

- 1 DOCUMENTOS
- 2 FOTOGRAFÍAS
- 3 SOCIAL MEDIA
- 4 FECHAS
- 5 LOCALIZACIONES GEOGRÁFICAS
- 6 TRANSACCIONES
- 7 BÚSQUEDAS
- 8 VISITAS A SITIOS WEB

Generamos más datos que nunca, tendencia que muestra un crecimiento exponencial. Nuestra capacidad para procesar tantos datos inaugura en sí una nueva etapa en la revolución digital.

«Big Data son recursos de gran volumen, alta velocidad y variedad que exigen formas rentables e innovadoras de procesamiento que permitan una mejor comprensión, toma de decisiones y automatización de procesos.»

*Gartner it Glossary*

El intercambio de datos a través de internet permite analizar la correlación de datos en un número prácticamente infinito de dimensiones, por lo que encuentra correlaciones que de otra manera serían inimaginables para el ser humano. La correlación se produce en un cruce de datos que brinda información sobre algo.

«Debemos dejar de hablar de causalidades y pensar en correlaciones.»  
Victor Mayer y Kennet Cukier

## FORMAS DE PROCESO DE DATOS EN BIG DATA

**CORRELACIÓN:** LOS DATOS MASIVOS QUE SE LEEN DE DIVERSAS FUENTES Y SE COMBINAN DE FORMA CASI INSTANTÁNEA NO MIDEN SUBJETIVIDADES, SINO QUE PRODUCEN TENDENCIAS Y ANÁLISIS SOBRE DATOS QUE QUIZÁ, EN TÉRMINOS CAUSALES, NO HABRÍAN SIDO TOMADOS EN CUENTA.

**CAUSALIDAD:** EN EL TIEMPO DEL BIG DATA, LA RELACIÓN CAUSAL DE LOS DATOS YA NO ES UN FACTOR RELEVANTE, PORQUE SE CONSIDERA QUE ENTRE MUCHOS VALORES NI SIQUIERA EXISTE O NO ES IMPORTANTE.

## MINERÍA DE DATOS

La minería de datos o *datamining* es una herramienta que permite identificar patrones en todos los datos para ponerlos a disposición del análisis y, finalmente, ayudar en la toma de decisiones.

Lo que hace particulares a los procesos de minería es que se producen mediante algoritmos que se insertan en estas grandes bases de datos, que buscan tendencias, anomalías, desviaciones o situaciones interesantes pero desconocidas, y otros eventos importantes.

## HERRAMIENTAS MÁS IMPORTANTES DE LA MINERÍA DE DATOS:

- 1 INTELIGENCIA ARTIFICIAL
- 2 ESTADÍSTICA
- 3 SISTEMAS DE BASES DE DATOS

### UTILIDAD DE LA MINERÍA DE DATOS

FUNCIONA PARA PREDECIR, AL UTILIZAR LA INFORMACIÓN PARA PRONOSTICAR VALORES FUTUROS.

FUNCIONA PARA DESCRIBIR, AL PRESENTAR PATRONES DESCRIPTIVOS DE LA INFORMACIÓN PARA QUE PUEDAN SER INTERPRETADOS POR SERES HUMANOS.

A PARTIR DE DATOS ESTRUCTURADOS, PRESENTA RESULTADOS QUE SON ARGUMENTADOS Y DISCUTIDOS DESPUÉS DE ARROJAR INFORMACIÓN EN LA QUE NO SE PROPUSIERON HIPÓTESIS PREVIAS.

## ALMACENAMIENTO DE DATOS

### FORMAS TRADICIONALES:

PINTURAS  
CÓDICOS  
GRÁFICAS  
MAPAS  
FOTOGRAFÍAS  
VIDEOS  
GRABACIONES DE AUDIO

### FORMAS DIGITALES:

SMARTPHONES  
COMPUTADORAS  
MEMORIAS DE ALMACENAMIENTO MASIVO  
NUBE

**BIG DATA:** PUEDE OBTENER INFORMACIÓN DE DATOS ESTRUCTURADOS, NO ESTRUCTURADOS Y SEMIESTRUCTURADOS

**MINERÍA DE DATOS:** SE ENFOCA EN EL ANÁLISIS DE DATOS CONVENCIONALES, EN LOS QUE LA INFORMACIÓN ESTÁ ESTRUCTURADA DE FORMA HOMOGÉNEA

Geoffrey Moore

Sin análisis de Big Data, las empresas están ciegas y sordas.



Michael Nielsen

Con la infraestructura apropiada, los datos pueden transformarse en conocimiento, a menudo de forma sorprendente.



Boletín trimestral #3

SO WHAT?

editado por de la Riva group

Es importante señalar que para llevar a cabo estos procesos es indispensable la actividad de los individuos que intervienen en la extracción de información.

## PRINCIPALES PROCESOS DE LA MINERÍA DE DATOS

- 1 FILTRADO DE DATOS
- 2 SELECCIÓN DE VARIABLES
- 3 EXTRACCIÓN DE CONOCIMIENTO
- 4 INTERPRETACIÓN Y EVALUACIÓN

La minería de datos utiliza, de forma automatizada, diferentes algoritmos que son clasificados como técnicas. Una de las más reconocidas es la de redes neuronales.

## CARACTERÍSTICAS DE LAS REDES NEURONALES:

- 1 CONSISTEN EN UNIDADES DE PROCESAMIENTO QUE INTERCAMBIAN DATOS.
- 2 SE UTILIZAN PARA RECONOCER PATRONES, POR EJEMPLO, EN IMÁGENES, MANUSCRITOS Y SECUENCIAS DE TIEMPO.
- 3 TIENEN LA CAPACIDAD DE APRENDER Y MEJORAR SU FUNCIONAMIENTO.

A partir del aprendizaje en instancias repetitivas, se construyen y refuerzan los conectores que fijan la información que va quedando en la red neuronal como referente.

## ALGORITMOS

Un algoritmo, en su definición más básica, consiste en una serie de pasos para realizar una tarea. En las ciencias de la computación, son secuencias de operaciones que permiten solucionar un problema. Los algoritmos son autónomos y no dependen de lenguajes de programación, por lo que deben diseñarse para ser leídos fácilmente.

## HISTORIA DEL BIG DATA

1920: Nikola Tesla predice el mundo inalámbrico.

1928: Fritz Pfleumer presenta la banda magnética.

1970: Edgar F. Codd, crea la base de datos relacional.

1963: Primera base de datos en EE.UU.

1944: Universidad de Harvard, primera computadora, Mark I

1980: Aparición de Internet

1990: NASA, Minería de datos.

2005: Consolidación de la web 2.0

2006: Hadoop



Fuente: Barcelona Supercomputing Center

## NUBE O CLOUD

El concepto de nube o Cloud, hace referencia a un paradigma de almacenamiento que no tiene una identidad física para los usuarios; es una infraestructura distribuida en internet. Estas plataformas permiten la distribución de datos y recursos de información en el momento en que se requiera y de manera redundante, lo que hace que su disponibilidad sea segura y confiable.

### VENTAJAS DE LA NUBE:

- 1 LA INFORMACIÓN ESTÁ SIEMPRE DISPONIBLE.
- 2 LOS DISPOSITIVOS YA NO SE SATURAN DE INFORMACIÓN.

### CARACTERÍSTICAS DE LA NUBE:

- 1 **ON-DEMAND SELF SERVICE:** EL CONSUMIDOR ACCEDE Y UTILIZA LOS SERVICIOS SEGÚN SUS NECESIDADES.
- 2 **BROAD NETWORK ACCESS:** SERVICIOS DISPONIBLES EN UNA RED PRIVADA, COMPARTIDA O PÚBLICA.
- 3 **RESOURCE POOLING:** RECURSOS DEL HARDWARE COMO LA CAPACIDAD DE PROCESAMIENTO, ASIGNACIÓN DE MEMORIA O ALMACENAMIENTO.
- 4 **RAPID ELASTICITY:** DESPLIEGUE Y PLANIFICACIÓN DE RECURSOS.
- 5 **MEASURED SERVICE:** CONTROL Y ELABORACIÓN DE INFORMES DE USO DEL SERVICIO PARA CONTROLARLO Y OPTIMIZARLO.

## CARACTERÍSTICAS DE LOS ALGORITMOS

**FINITUD (FINITENESS):** UN ALGORITMO DEBE TERMINAR SIEMPRE DESPUÉS DE UN NÚMERO FINITO DE PASOS.

**DEFINICIÓN (DEFINITENESS):** CADA PASO DE UN ALGORITMO DEBE SER DEFINIDO CON PRECISIÓN; LAS ACCIONES A REALIZAR SON RIGUROSAS Y SIN AMBIGÜEDADES.

**ENTRADA (INPUT):** CANTIDADES QUE SE DAN INICIALMENTE

ANTES DE QUE COMIENZE EL ALGORITMO. ESTAS ENTRADAS SE TOMAN DE CONJUNTOS DE OBJETOS ESPECIFICADOS.

**SALIDA (OUTPUT):** CANTIDADES QUE TIENEN UNA RELACIÓN ESPECÍFICA CON LAS ENTRADAS.

**EFFECTIVIDAD (EFFECTIVENESS):** LAS OPERACIONES A REALIZAR EN EL ALGORITMO DEBEN SER SUFICIENTEMENTE

BÁSICAS PARA QUE EN PRINCIPIO PUEDAN SER REALIZADAS EXACTAMENTE Y EN UN TIEMPO FINITO POR UN HOMBRE USANDO PAPEL Y LÁPIZ.

Hoy en día gran parte de nuestras actividades están relacionadas con algoritmos: la cámara que se ajusta a nuestra sonrisa, el teclado que predice nuestra escritura, son sólo algunos ejemplos.

**Boletín trimestral #3**

SO WHAT?

editado por de la Riva group

**Peter Norvig**

No tenemos mejores algoritmos.  
Sólo tenemos más datos.

**Chris Anderson**

No hay razón para aferrarse a los viejos caminos. Es tiempo de preguntarnos: ¿qué puede aprender la ciencia de Google?

**PLATAFORMAS QUE USAN LA NUBE:**

- AMAZON
- GOOGLE DRIVE
- DROPBOX
- SPOTIFY
- NETFLIX

«Los datos son el petróleo del siglo XXI.»

Cesar Alierta

**INTELIGENCIA ARTIFICIAL**

El término inteligencia artificial se refiere a sistemas creados por los seres humanos con la intención de constituir agentes racionales autónomos mediante procesos computacionales.

La IA vinculada al Big Data se trata de programas computacionales contruidos para realizar operaciones definidas en un proceso en el que el sistema pueda generar un proceso de aprendizaje tal cual lo hace la mente humana, a partir de la experiencia.

**OBJETIVOS DE LA INTELIGENCIA ARTIFICIAL:**

DISEÑAR PROGRAMAS COMPARATIVOS DE EFICIENCIA CON LOS SERES HUMANOS, CON LA FINALIDAD DE PRODUCIR MAYOR CONOCIMIENTO.

MAXIMIZAR RESULTADOS Y RENDIMIENTO DE LAS ACCIONES.

**ORIGEN DE DATOS**

WEB

MACHINE TO MACHINE (M2M)

BIOMETRICS

SOCIAL MEDIA

TRANSACCIONES

HUMAN GENERATED

**TIPOS DE INTELIGENCIA ARTIFICIAL**

**Machine learning.** Máquinas capaces de autoprogramarse pero en donde los humanos toman las primeras decisiones. Aprendizaje supervisado.

**Deep learning.** Emulación cada vez más cercana a la percepción humana. Aprendizaje no supervisado.

La analítica de datos (en inglés, analytics) ha sido considerada como una propuesta tradicional frente al mundo del Big Data. Regularmente, al tratarse del análisis de datos para tomar decisiones, se confunde o no se identifica su especificidad frente al Big Data. Pero la analítica puede ser comprendida como el paso anterior en esta revolución de los datos masivos; lo más relevante es que se basa en las deducciones que se pueden hacer a través de los resultados obtenidos de datos en bruto.

**TIPOS DE ANALÍTICA**

DESCRIPTIVA: ¿QUÉ HAY?

PREDICTIVA: ¿QUÉ VA A OCURRIR?

PRESCRIPTIVA: ¿QUÉ HACER?

**BIG DATA ANALITYCS**

Big Data Analytics se refiere al procesamiento y análisis de los datos masivos que se producen a escalas que no es posible procesar por los medios convencionales de la informática. El objetivo del análisis se centra en que la información obtenida debe servir para algo, por lo que se espera que en este procesamiento de datos se llegue a nuevas formas de actuación de empresas u organizaciones.

«Big Data consiste en encontrar correlaciones, mientras que el Small Data se preocupa por las causas, las razones detrás de las cosas.»

Martin Lindstrom

**3 V DE BIG DATA (DOUGLAS LANEY)****VOLUMEN**

DATOS MASIVOS, CANTIDADES MONUMENTALES.

LOS DATOS SE ALMACENAN EN UNA BATERÍA LLAMADA CLÚSTER.

**VELOCIDAD**

RAPIDEZ EN EL PROCESAMIENTO Y SELECCIÓN DE DATOS.

IDENTIFICA LOS PROCESOS DE MAYOR VOLATILIDAD PARA PRODUCIR UNA RESPUESTA OPORTUNA E INMEDIATA.

**VARIEDAD**

DIVERSIDAD DE TIPOS Y FORMAS DE DATOS QUE ES POSIBLE PROCESAR.

DATOS CLASIFICADOS EN ESTRUCTURADOS (TABLAS), NO ESTRUCTURADOS (LA MAYORÍA, CADENAS DE SEGUIMIENTO) O SEMIESTRUCTURADOS (LENGUAJES WEBS).

«Si combinamos inteligencia artificial, con los mejores algoritmos, con Big Data, obtenemos el poder analítico y productivo más grande (Google, Amazon, Facebook) que es el que está cambiando el mundo drásticamente.»

Gustavo Ross

Como podemos ver, los datos tienen una complejidad que corresponde a su origen, a las formas de almacenamiento y, además, a cómo se procesan por aplicaciones de Big Data. Por eso, a través de las 3 V que lo componen podemos decir que el Big Data se define como: la capacidad de producir correlaciones como resultado de almacenar grandes cantidades de datos de diversos tipos y orígenes, que se producen a gran velocidad y que requieren de un tratamiento en tiempo real.

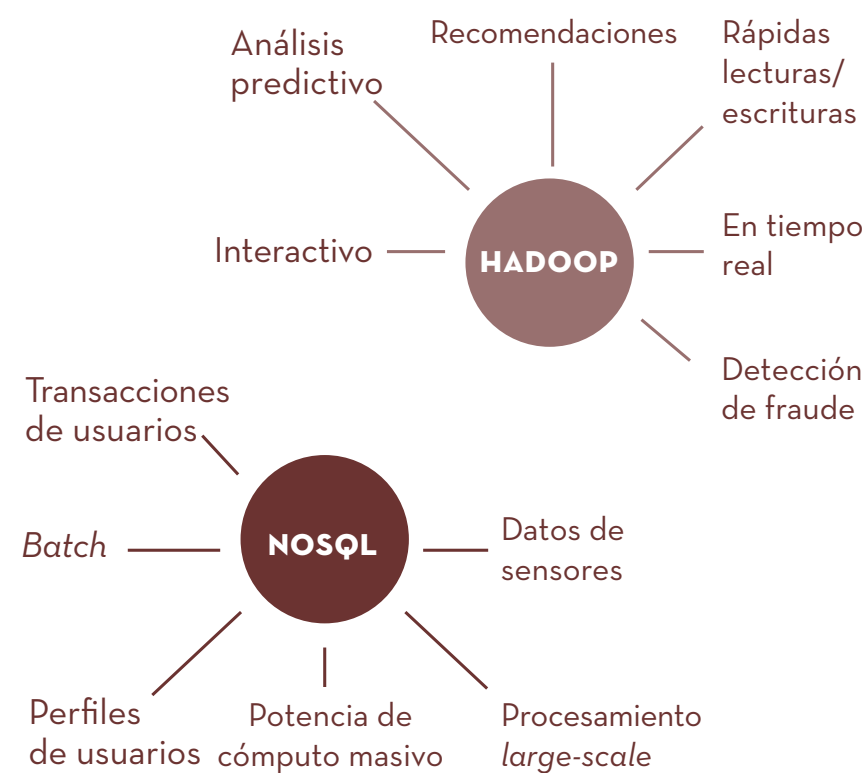
«Las aplicaciones Big Data más impactantes serán por sí mismas una industria que aprovechará los datos que la propia empresa consume y genera en el viaje de hacer negocios. No existe una fórmula única para extraer valor de los datos, depende de su aplicación.»

David Gorbet

**PRINCIPALES HERRAMIENTAS DE ADMINISTRACIÓN DE DATOS:**

-NoSQL (arquitectura tipo base de datos)

-Hadoop (ecosistema de software que puede coordinar bases de datos NoSQL)



Se trata de marcos de reconocimiento que se diferencian de los sistemas de gestión de bases de datos relacionales, porque permiten analizar datos de diferentes tipos en grandes volúmenes por ser más flexibles y funcionar en estructuras que no son fijas.

Hannah Jones

El análisis de datos para la sostenibilidad puede ayudar a liderar los modelos de negocio.



Carly Fiorina

La meta es convertir los datos en información, y la información en *insights*.



Boletín trimestral #3

SO WHAT?

editado por de la Riva group

«Tras nuestra última investigación hemos descubierto que las empresas utilizan Big Data para obtener resultados centrados en el cliente, aprovechar los datos internos y crear un mejor ecosistema de información.»

IBM Institute for Business Value / Escuela de Negocios Saïd en la Universidad de Oxford

Según los especialistas Viktor Mayer y Kenneth Cukier autores del libro seminal *Big data. La revolución de los datos masivos*, hay tres tipos de empresas que pueden identificarse en el mundo de los datos masivos:

<sup>1</sup> Las que tienen datos pero no tienen capacidades para extraer el valor o construir ideas sobre ellos.

<sup>2</sup> Las que tienen las capacidades, como las consultoras, que tienen el conocimiento para sacar conclusiones de los datos, pero que no los poseen.

<sup>3</sup> Las que tienen mentalidad Big Data, como Google o Amazon, en las que los datos y su análisis son cruciales, ya que no sólo tienen los datos o saben analizarlos para tener éxito, sino que tienen personal capaz de desarrollar ideas sobre formas de explorar los datos.

Mientras más info tengo más quiero a mis clientes

Una de las cuestiones que ha revelado el uso de Big Data es que las compañías se están enfocando en conocer mejor a sus usuarios o clientes. El *engagement* que una marca tenga dependerá de la cantidad de información que obtenga de sus posibles clientes.

«El internet de las cosas, en términos muy simples, se puede definir como la capacidad de los objetos de comunicarse entre ellos, mediante una señal de internet.»

Luis Arvizu

## BIG DATA Y POLÍTICA

El uso del Big Data en política ha sido polémico. Esta tendencia ha permitido tener un mayor acercamiento al votante; además, los partidos políticos aprenden a conocerse mejor. La lectura de datos masivos se realiza, sobre todo, durante las campañas electorales, porque se pueden identificar los perfiles de votantes para definir a quiénes dirigirse y así capitalizar recursos en tiempo real. Es una estrategia similar a como las empresas lanzan campañas publicitarias y posicionan sus productos.

### PROGRAMAS BIG DATA USADOS EN POLÍTICA:

2008, BARACK OBAMA, PROGRAMA NARWHAL

2016, DONALD TRUMP, CAMBRIDGE ANALYTICA

## BIG DATA Y GOBIERNO

El Big Data también ha empezado a tener usos en las políticas públicas, más allá de las coyunturas electorales. La posibilidad de almacenamiento, procesamiento y análisis de datos masivos ha permitido un diseño, ejecución y evaluación más correctos de los proyectos. Así como se hace en el ámbito empresarial, los gobiernos empiezan a usar datos acerca de las personas.

El Big Data es una tendencia consolidada y, a la vez, en permanente evolución. Se ha caracterizado por desarrollarse a la par de la inteligencia artificial y por conectarse directamente con los avances en inteligencia de negocios y con la analítica. Responder *qué* para consolidar un *cómo* y proponer un *para qué*, es hoy una opción posible gracias al manejo y gestión de datos masivos.

Amazon, algoritmo filtrado colaborativo ítem por ítem para analizar grandes volúmenes de datos.

Google, su algoritmo busca calcular la importancia de cada sitio web a partir de los vínculos (*links*) que cada sitio recibe.

## CASOS EXITOSOS DE BIG DATA

Uber, su gran base de datos permite que el algoritmo tarde 15 segundos en localizar el automóvil más cercano al usuario.

Airbnb, su algoritmo pone precios de renta a través de una cantidad de información sobre lugares similares a los que ofrecen los arrendadores, y hace sugerencias dinámicas.

## BENEFICIOS DEL BIG DATA

OPORTUNO  
E INMEDIATO

ACCESIBLE

HOLÍSTICO

PERTINENTE

SEGURO

PRECISO

UTILIZABLE

### ¿Qué es lo que nos depara el futuro?

La mayoría de los usuarios piensa que Big Data revolucionará el mundo de la misma forma que lo hizo internet. Tendrá un mayor impacto en:

63%

LA RELACIÓN CON LOS CLIENTES

56%

OPERACIONES DE CAMBIO

58%

DESARROLLO DE PRODUCTOS

## PREDICCIONES SOBRE EL FUTURO DEL BIG DATA:

1. Los volúmenes de datos seguirán creciendo.
2. Las formas de analizar los datos mejorarán.
3. Aparecerán más herramientas para el análisis (sin el analista).
4. La analítica prescriptiva se incorporará al *software* de análisis empresarial.
5. Además, la información en tiempo real sobre transmisión de datos será la característica distintiva de los ganadores de datos en el futuro.
6. El aprendizaje automático será una tendencia estratégica.
7. Los grandes datos representarán enormes desafíos para los encargados de interpretarlos.
8. Más compañías nombrarán un jefe de datos.
9. «Los agentes autónomos y las cosas» seguirá siendo tendencia.
10. Los científicos incluirán arquitectos y expertos en gestión de datos. Pero la gran crisis del talento puede disminuir a medida que las empresas emplean nuevas tácticas.
11. El modelo de negocio de datos como servicio está en el horizonte.
12. Los mercados de algoritmos también surgirán.
13. La tecnología cognitiva será la nueva moda.
14. Todas las empresas son negocios de datos ahora.
15. Las empresas que utilicen datos verán \$430 mil millones en beneficios de productividad en 2020.
16. «Datos rápidos» y «datos procesables» reemplazarán datos grandes, según algunos expertos.

link insightful®

Twitter @delarivagroup

Facebook De la Riva

LinkedIn De la Riva Group