



INDUSTRIA 4.0

BIG DATA, OPEN DATA Y DATA ANALYTICS



Contenido

Big Data	4
Tipos de datos.....	4
Datos Estructurados	5
Datos No Estructurados	5
Datos Semiestructurados.....	6
Volumen de datos.....	8
Velocidad de flujo de datos	8
Variedad de datos.....	9
El Modelo de las 5V	10
Veracidad de los datos.....	10
Valor de los datos	11
EL MODELO DE LAS 7 V	11
Visualización	12
Viabilidad.....	12
Fuentes de datos	13
Tipos de fuentes de datos.....	14
Datos en organizaciones y empresas	14
Datos internos	15
Datos externos.....	15
Arquitectura del Big Data.....	16
Recolección de datos	16
Almacenamiento de datos	17
Procesamiento y análisis de datos	18
Visualización de datos.....	18
Open Data: El movimiento de los datos abiertos	18



Iniciativas open data	19
¿Qué es la analítica de datos? (DATA ANALYTICS).....	20
Una visión global del Data Analytics	21
Roles profesionales relacionados con datos	21





Big Data

No existe unanimidad en la definición de Big Data, aunque sí hay cierto consenso en la fuerza disruptiva que suponen los grandes volúmenes de datos y la necesidad de su captura, almacenamiento y análisis.

Son numerosos los artículos (White Papers), informes y estudios relativos al tema de Big Data en los últimos años, y en este libro seleccionamos las definiciones de instituciones relevantes y con mayor impacto mediático y profesional. En general, existen diferentes aspectos en los que casi todas las definiciones están de acuerdo y con conceptos consistentes para capturar la esencia de Big Data: crecimiento exponencial de la creación de grandes volúmenes de datos, origen o fuentes de datos y la necesidad de su captura, almacenamiento y análisis para conseguir el mayor beneficio para organizaciones y empresas, junto con las oportunidades que ofrecen y los riesgos que conlleva su no adopción.

“Big Data se refiere a los conjuntos de datos cuyo tamaño está más allá de las capacidades de las herramientas comunes de software de bases de datos para capturar, almacenar, gestionar y analizar”

Otra definición de Big data puede ser la siguiente:

“Big Data son los grandes conjuntos de datos que tiene tres características principales: volumen (cantidad), velocidad (velocidad de creación y utilización) y variedad (tipos de fuentes de datos no estructurados, tales como la interacción social, video, audio, cualquier cosa que se pueda clasificar en una base de datos)”

Tipos de datos

Los Big Data son diferentes de las fuentes de datos tradicionales que almacenan datos estructurados en las bases de datos relacionales. Es frecuente dividir las categorías de datos en dos grandes tipos: estructurados (datos tradicionales) y no estructurados o sin estructura (datos Big Data). Sin embargo, las nuevas herramientas de manipulación de Big Data han originado unas nuevas categorías dentro de los tipos de



datos no estructurados: datos semiestructurados y datos no estructurados propiamente dicho.

Datos Estructurados

La mayoría de las fuentes de datos tradicionales son datos estructurados, datos con formato o esquema fijo que poseen campos fijos. En estas fuentes, los datos vienen en un formato bien definido que se especifica en detalle, y que conforma las bases de datos relacionales. Son, fundamentalmente, los datos de las bases de datos relacionales, las hojas de cálculo y los archivos.

Los datos estructurados se componen de piezas de información que se conocen de antemano, vienen en un formato especificado y se producen en un orden, también, especificado. Estos formatos facilitan el trabajo con dichos datos. Formatos comunes son: fecha de nacimiento (DD, MM, AA); documento nacional de identidad o pasaporte (por ejemplo, 8 dígitos y una letra); número de la cuenta corriente en un banco (20 dígitos), etcétera.

La gestión y búsqueda de los datos estructurados en las bases de datos relacionales se realizan con el lenguaje de programación estándar SQL –lenguaje creado por IBM en la década de los setenta– y que todavía sigue en vigor y soporta a la mayoría de las bases de datos establecidas en organizaciones y empresas.

Sin embargo, las bases de datos relacionales tienen un gran inconveniente en la era de los grandes volúmenes de datos: la escasa facilidad que tiene para manejar datos no estructurados.

Datos No Estructurados

Los datos no estructurados (sin estructurar) son datos sin tipos predefinidos. Se almacenan como “documentos” u “objetos” sin estructura uniforme, y se tiene poco o ningún control sobre ellos. Tienen un formato que no puede ser gestionado (indexado) fácilmente en tablas de bases de datos relacionados. Datos no estructurados son:



- Video, audio, imágenes, fotografías
- Datos de texto (archivos de texto o documentos, tales como Word, PowerPoint, PDF...)
- Documentos multimedia
- Correos electrónicos
- Textos de mensajería (SMS, mensajes de WhatsApp, Line, Telegram, Viber, Snapchat...)
- Publicaciones en redes sociales

Sin duda, los datos más difíciles de dominar por los analistas son los datos no estructurados, pero su continuo crecimiento ha provocado el nacimiento de herramientas para su manipulación como es el caso de MapReduce, Hadoop o bases de datos NoSQL.

Datos Semiestructurados

Los datos semiestructurados tienen propiedades de datos estructurados y no estructurados, y pueden tener algún tipo específico de estructura que se puede utilizar en un análisis de datos, pero no contienen la estructura de un modelo de datos. Asimismo, poseen un flujo lógico y un formato que puede ser definido, pero no es fácil su comprensión por el usuario. Son datos que no tienen formatos fijos, pero sí etiquetas y otros marcadores que permiten separar los elementos dato. Ejemplos de datos semiestructurados son:

- Documentos XML de páginas web
- Contenidos de blogs y redes sociales
- Software de tratamiento de textos
- Lenguajes de marca de hipertexto extensibles

La lectura de datos semiestructurados requiere el uso de reglas complejas que determinan cómo proceder después de la lectura de cada pieza de información. Ejemplos típicos de datos semiestructurados son:



- Los registros Web logs de las conexiones a Internet. Un Web log se compone de diferentes piezas de información, cada una de las cuales sirve para un propósito específico. Ejemplos comunes son el texto de etiquetas de lenguajes XML y HTML.
- El software de tratamiento de textos que incluyen metadatos que pueden contener nombre del autor, ISBN del libro, fecha de edición, fecha de compra, etc.; sin embargo, su contenido está sin estructurar.
- Publicaciones, entradas de Facebook o LinkedIn, que se pueden clasificar por autor, información, longitud de texto, opiniones de seguidores, etc., pero su contenido normalmente no está estructurado.
- Big Data abarca tres grandes dimensiones, conocidas como el “Modelo de las tres V” (3 V o V3): volumen, velocidad y variedad (variety). Existe un gran número de puntos de vista para visualizar y comprender la naturaleza de los datos y las plataformas de software disponibles para su explotación; la mayoría incluirá una de estas tres propiedades V en mayor o menor medida. Sin embargo, algunas fuentes, como es el caso de IBM, cuando tratan las características de los Big Data también consideran una cuarta característica que es la veracidad, y que analizaremos también para dar un enfoque más global de la definición y características de los Big Data. Otras fuentes notables añaden una quinta característica, valor, y llegan a añadir hasta 7 u 8V, como veremos más adelante.

TIPOS DE DATOS

ESTRUCTURADOS

Datos que tienen un modelo definido o provienen de un campo determinado en un registro



precios de acciones - base de datos de compras - rastreo web

NO ESTRUCTURADOS

Datos que no tienen un modelo predefinido o no están organizados de alguna manera



fotografía - documentos de texto - video



Volumen de datos

Las empresas amasan grandes volúmenes de datos, desde terabytes hasta Petabytes. Las cantidades que hoy nos parecen enormes, en pocos años serán normales. Estamos pasando de la era del Petabyte a la era del exabyte, y para el periodo 2015 a 2020 se espera que entremos en la era del Zettabyte. IBM da el dato de 12 terabytes para referirse a lo que crea Twitter cada día sólo en el análisis de productos para conseguir mejoras en la eficacia.

La característica volumen es la más popular y reconocida dentro del término Big Data, aunque al día de hoy no es la más significativa. Se necesita almacenar la información y para ello se requiere una base de datos capaz de almacenar y gestionar las enormes cantidades de datos. Los tamaños de los archivos son muy grandes y cada segundo/minuto se generan grandes volúmenes de datos que crecen de modo exponencial, en la expresión que IDC denomina “El universo digital de datos”. El volumen de datos es una de las propiedades más destacadas en cualquier definición de Big Data, pero existen otras propiedades de igual o mayor importancia en la actualidad.

Velocidad de flujo de datos

La importancia de la velocidad de los datos o el aumento creciente de los flujos de datos en las organizaciones, junto con la frecuencia de las actualizaciones de las grandes bases de datos, son características importantes a tener en cuenta. Esto requiere que su procesamiento y posterior análisis, normalmente ha de hacerse entiendo tiempo real para mejorar la toma de decisiones sobre la base de la información generada. A veces, cinco minutos es demasiado tarde en la toma de decisiones; los procesos sensibles al tiempo –como pueden ser los casos de fraude– obligan a actuar rápidamente. Imaginemos los millones de escrutinios de los datos de un banco con el objetivo de detectar un fraude potencial o el análisis de millones de llamadas telefónicas para tratar de predecir el comportamiento de los clientes y evitar que se cambien de compañía.



La importancia de la velocidad de los datos se une a las características de volumen y variedad, de modo que la idea de velocidad no se asocia a la tarea decrecimiento de los depósitos o almacenes de datos, sino que se aplica la definición al concepto de los datos en movimiento, es decir, la velocidad a la cual fluyen los datos. Las empresas están tratando cada día con mayor intensidad, Petabytes de datos en lugar de terabytes, y el incremento en fuentes de todo tipo como sensores, chips NFC, datos de geolocalización y otros flujos de información que conducen a flujos continuos de datos, imposibles de manipular por sistemas tradicionales.

Variedad de datos

Las fuentes de datos son de cualquier tipo. Los datos pueden ser estructurados y no estructurados (texto, datos de sensores, audio, video, flujos de clics, archivos logs), y cuando se analizan juntos se requieren nuevas técnicas. Imaginemos el registro en vivo de imágenes de las cámaras de video de un estadio de fútbol, o las de vigilancia de calles y edificios.

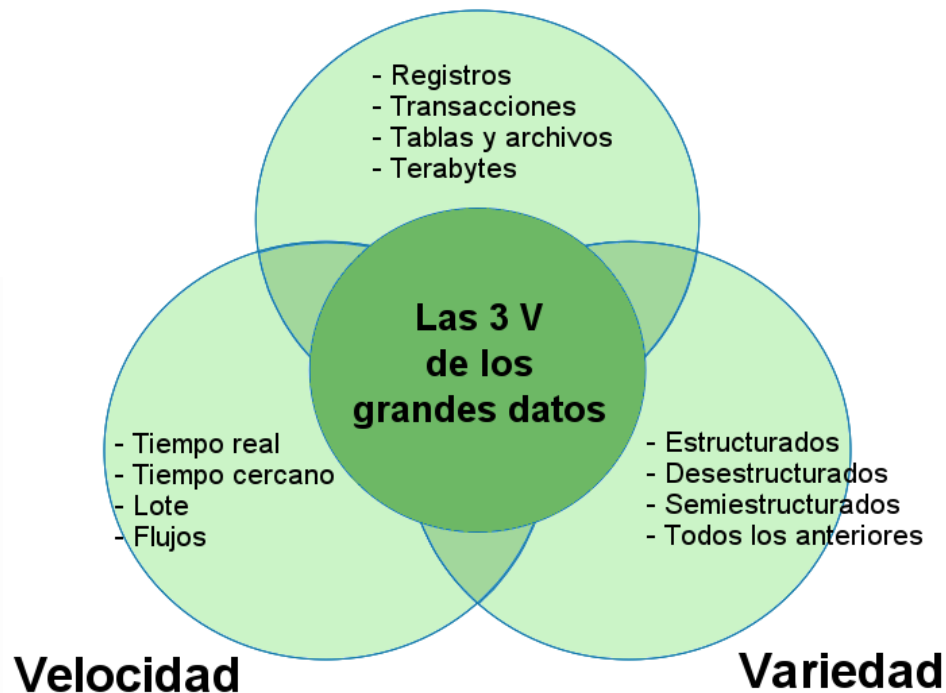
En los sistemas de Big Data las fuentes de datos son diversas y no suelen ser estructuras relacionales comunes. Los datos de imágenes de redes sociales pueden venir de una fuente de sensores y no suelen estar preparados para su integración en una aplicación.

En el caso de la Web, la realidad de los datos es confusa. Distintos navegadores envían datos diferentes; los usuarios pueden ocultar información y usar diversas versiones de software, ya sea para comunicarse entre ellos, realizar compras o para leer un periódico digital. No obstante, los riesgos que conlleva la no adopción de las tendencias de Big Data son grandes, ya que:

- La voluminosa cantidad de información puede llevar a una confusión que impida ver las oportunidades y amenazas dentro de nuestro negocio y fuera de él, y perder así competitividad.
- La velocidad y flujo constante de datos en tiempo real puede afectar a las ventas y a la atención al cliente.
- La variedad y complejidad de datos y fuentes puede llevar a la vulneración de determinadas normativas de seguridad y privacidad de datos.



Volumen



El Modelo de las 5V

IBM añadió una cuarta V y posteriormente una quinta V. De igual forma, Bernard Marr, uno de los grandes expertos mundiales en Big Data, publicó el artículo: “Big data: The 5 vs Everyone Must Know” donde añade dos nuevas propiedades: Veracidad y Valor, que luego incluiría en su libro de Big Data, publicado en 2015

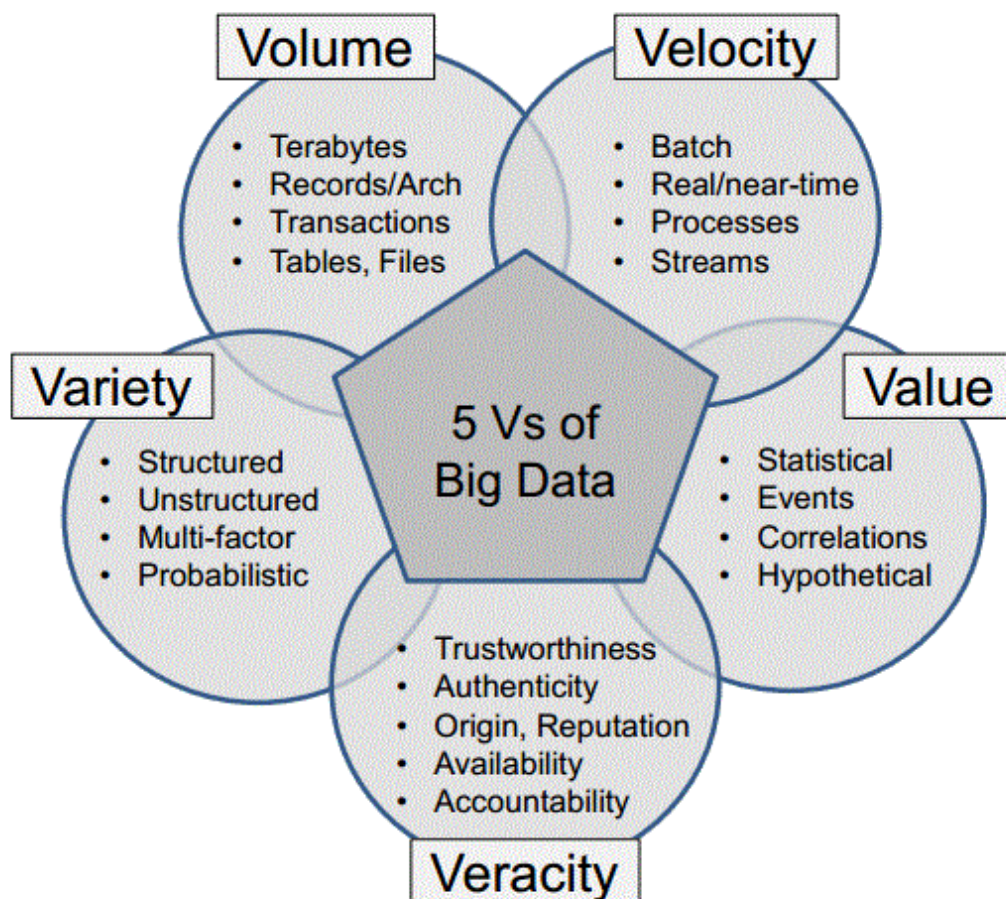
Veracidad de los datos

En su definición de Big Data, al comentar la característica de veracidad, IBM proporciona un dato estremecedor: “Uno de cada tres líderes de negocio (directivos) no se fía de las informaciones que utilizan para tomar decisiones (lo denomina incertidumbre).” ¿Cómo puede, entonces, actuar con esta información si no se fía de ella? El establecimiento de la veracidad o fiabilidad (truth) de Big Data supone un gran reto a medida que la variedad y las fuentes de datos crecen



Valor de los datos

Existe una quinta característica que también se suele considerar y es muy importante: el valor. Las organizaciones estudian obtener información de los grandes datos de una manera rentable y eficiente. Aquí es donde las tecnologías de código abierto tales como Apache Hadoop se han vuelto muy populares. Hadoop es un software que procesa grandes volúmenes de datos a través de un cluster de centenares o incluso millares de computadores de un modo muy económico.



EL MODELO DE LAS 7 V

A las cinco características anteriores se están uniendo según algunos modelos de Big Data: las nuevas e importantes de visualización y viabilidad.

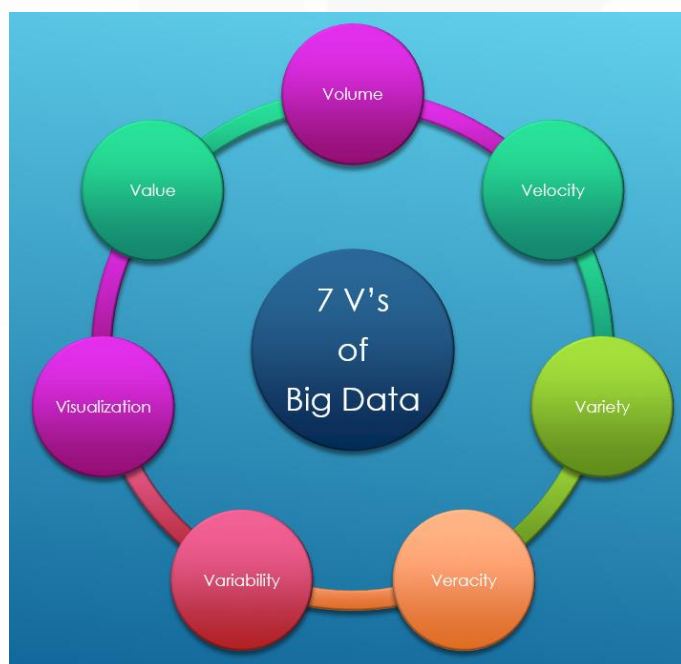


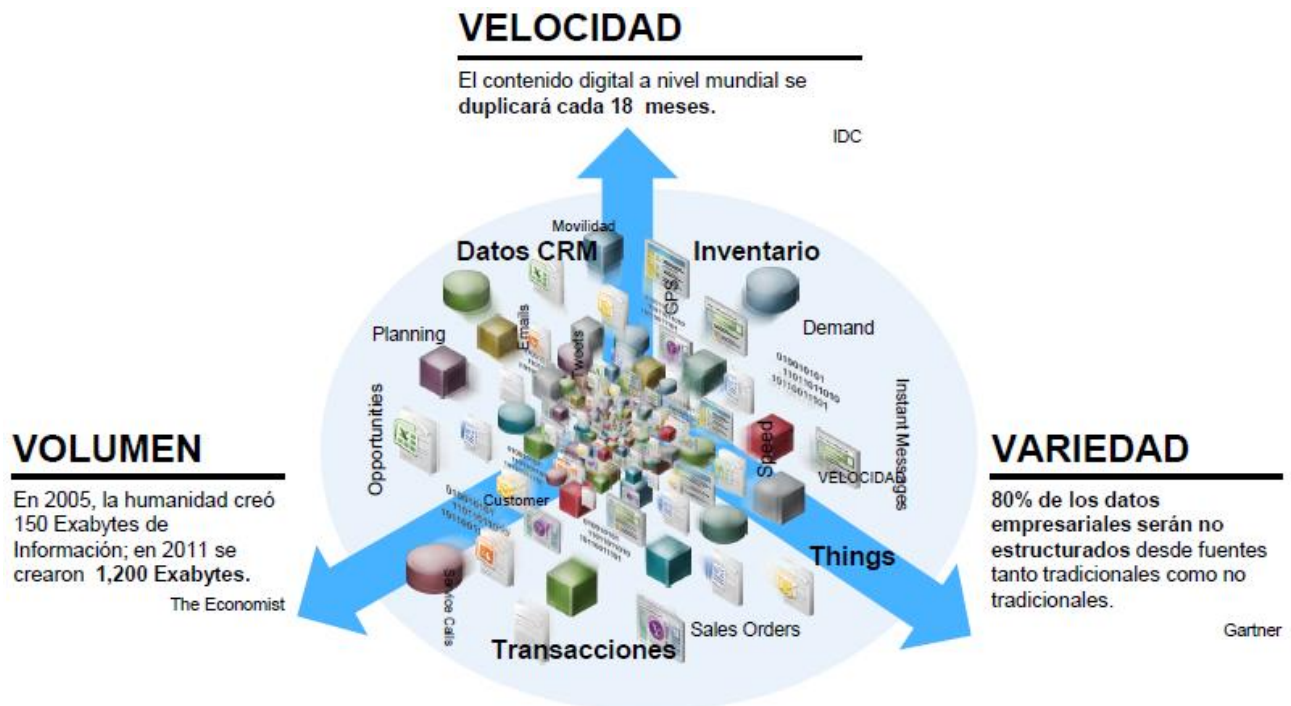
Visualización

Es el modo en que los datos se presentan para encontrar patrones y claves que permitan la obtención de resultados para una toma de decisión eficiente. Las iniciativas de Big Data requieren herramientas de visualización de datos óptimas. Estas herramientas permiten a los usuarios finales realizar búsquedas y acceder a la información rápidamente y en muchos casos en tiempo real. Es una gran ventaja para los clientes quienes se muestran satisfechos cuando tienen el control de la información en el mismo momento que ésta se produce. La visualización es una parte muy importante ya que ayuda a las organizaciones a responder a preguntas de interés para el desarrollo del negocio.

Viabilidad

Esta propiedad se refiere a la capacidad que tienen las empresas de generar un uso eficaz del gran volumen de datos que manejan. Esta característica también adopta la forma de la variabilidad para referirse también a que el gran volumen de datos está cambiando constantemente (este es el caso de los asistentes virtuales como Siri o el computador cognitivo Watson que reúnen datos para el procesamiento del lenguaje).





Fuentes de datos

El gran volumen de datos procede de numerosas fuentes, especialmente de las nuevas fuentes como medios sociales (social media) y los sensores de máquinas (máquina a máquina, M2M, e Internet de las cosas). La oportunidad de expandir el conocimiento incrustado en ellos, por combinación de esa inmensidad de datos con los datos tradicionales existentes en las organizaciones está acelerando su potencialidad; además, gracias a la nube (Cloud), a esa enorme cantidad de información se puede acceder de modo ubicuo, en cualquier lugar, en cualquier momento y, prácticamente, desde cualquier dispositivo inteligente.



Tipos de fuentes de datos



Visión general de las fuentes de datos



Datos en organizaciones y empresas

Los datos que manejan las organizaciones y empresas se agrupan en dos grandes categorías: datos internos y datos externos. Todos ellos a su vez, como ya se ha comentado anteriormente pueden ser datos estructurados, no estructurados o semiestructurados.



Datos internos

«Los datos internos representan todo aquello a lo que su empresa tiene o podría tener acceso en la actualidad, incluyendo los datos personales o privados que recoge la empresa y pertenecen a ésta, cuyo acceso está controlado por usted» (Marr 2015: 55).

Algunos ejemplos citados por Marr son:

- Comentarios de los clientes
- Datos de ventas.
- Datos de las encuestas a los empleados o los clientes
- Datos en video de circuitos cerrados de televisión.
- Datos transaccionales
- Datos de registros de los clientes
- Datos de control de existencias
- Datos de RR. HH

Datos externos

«Los datos externos de una organización son una variedad infinita de información que existe fuera de la misma. Los datos externos pueden ser públicos y privados. Los públicos son datos que todas las personas pueden obtener, tanto reuniéndolos gratuitamente como pagando a un tercero para conseguirlos, o haciendo que un tercero los recoja por usted. Los privados normalmente son aquellos que necesitaría conseguir y por los que tendría que pagar a otra empresa o a un tercero proveedor de datos.» (Marr 2015:55-66). Algunos ejemplos de datos externos citados por Marr (2015: 56) son:

- Datos meteorológicos.
- Datos oficiales como los censales.
- Datos de Twitter.
- Datos de perfiles en redes sociales.
- Google Trends o Google Maps,

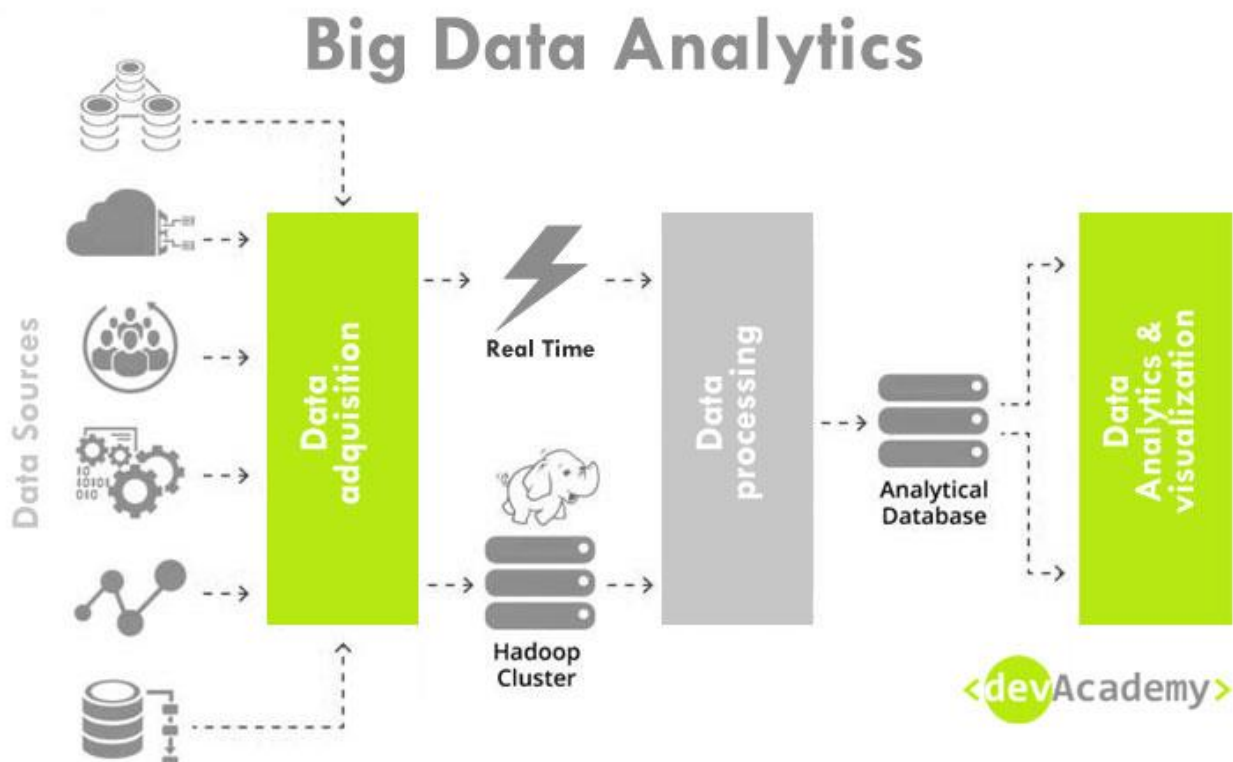


- [Datos de Facebook, Instagram, Pinterest, Amazon, Microsoft, otras aplicaciones de Google]

Arquitectura del Big Data

Las cuatro capas más consideradas en el proceso de tratamiento de Big Data son:

- Recolección (ingesta) de datos
- Almacenamiento
- Análisis de datos
- Visualización de datos (resultados)



Recolección de datos

Es la etapa de obtención (ingesta) de datos. Se ha convertido en una etapa de gran interés en el proceso de Big Data ya que existen numerosos datos públicos que se producen en enormes cantidades, numerosos dispositivos desperdigados por todo el planeta que emiten, procesan y recogen información de las más diversas



actividades (posicionamiento de individuos y vehículos, niveles de contaminación, temperaturas...); de igual forma, infinidad de dispositivos móviles que también emiten y capturan datos, etcétera.

Existen dos métodos de recolección de datos que se utilizarán según los casos:

- Por lotes (batch). Se conecta cada cierto tiempo a las fuentes de información (sistemas de archivos o bases de datos).
- Tiempo real (streaming). Este tipo de recolección de información trata directamente con la fuente de información de manera continua y en forma que la información se obtiene en tiempo real, cada vez que se necesita.



Almacenamiento de datos

Los sistemas de almacenamiento tradicionales se han tenido que adaptar a las grandes cantidades de datos que se generan, así como a la velocidad a la que se producen. Por esta razón las bases de datos tradicionales (relacionales) no se adaptan a estas necesidades y se requieren nuevos sistemas de almacenamiento. Entre otras características, los nuevos sistemas de almacenamiento deben ser escalables, precisamente debido a los grandes volúmenes de datos que precisan las compañías y de cualquier tipo, los cuales tienen que procesarse de acuerdo a las necesidades de las mismas (aumentando o disminuyendo sus capacidades). Este tipo de almacenamiento escalable tendrá que ser más transparente y eficiente dado que debe permitir la



ampliación o reducción requerida y, por consiguiente, las tecnologías utilizadas se han de adaptar a esta característica.

Procesamiento y análisis de datos

Esta etapa suele considerarse como una o dos etapas según la. Una vez que se tienen almacenados los datos, se han de convertir en conocimiento (valor) mediante el procesamiento y análisis de toda la información almacenada. Se trata de ser capaz de procesar datos en un tiempo razonable y alejarse de los estudios tradicionales de mercados estáticos. Los tipos de procesamiento son: Batch (por lotes), Streaming (en tiempo real) e híbrido.

Visualización de datos

Los resultados del análisis de datos es la etapa de consumo que debe permitir su visualización para una correcta toma de decisiones. Esta capa de Big Data muestra el producto del almacenamiento y procesamiento de la información cuyo resultado es la producción de conocimiento. En la actualidad existe un gran número de herramientas de visualización de datos que proporcionan una gran eficiencia a las compañías.

Open Data: El movimiento de los datos abiertos

Una variante muy importante de Big Data es la estrategia Open Data (datos abiertos) o apertura de datos. La estrategia Open Data, que históricamente nació en 2009 en Washington (ciudad pionera en este movimiento, data.gov), se refiere a la posibilidad de que el ciudadano



acceda a los datos del Gobierno que antes sólo eran analizados en el interior de las



administraciones públicas. Open Data consiste en una iniciativa para poner a disposición de las personas y empresas residentes en el país, datos de carácter público. Aunque la iniciativa de Open Data nació en los Estados Unidos, hoy día forma parte de la Agenda Digital Europea, donde numerosos países (entre ellos España)

han promovido iniciativas de datos abiertos, así como en América Latina, donde se desplegaron y promovieron también pronto, iniciativas nacionales de Open Data en países tales como Perú, México, Argentina y Colombia, La tendencia Big Data puede proporcionar una gran ventaja competitiva a las empresas y grandes beneficios a los usuarios y ciudadanos en general, en el movimiento y las tendencias de datos abiertos.



Iniciativas open data

Las iniciativas de Open Data en el mundo son numerosas. Los proyectos más innovadores han nacido en los Estados Unidos con la primera administración del



presidente Obama, y en Gran Bretaña; en España a nivel regional o autonómico y local, aunque también existieron iniciativas nacionales como el Proyecto Aporta dentro del Plan Avanza, que desde 2007 plantearon que todas las administraciones locales, autonómicas y centrales estaban llamadas a hacer pública la información que generan, hechos que se fueron confirmando de modo progresivo.

Dejamos algunos links de open data en el mundo:

Estados Unidos (Data.gov)

Reino Unido (data.gov.uk)

España (datos.gob.es)

América Latina y Caribe (datosperu.org)

¿Qué es la analítica de datos? (DATA ANALYTICS)

La analítica de datos (AD, Data Analytics, DA) “implica los procesos y actividades diseñados para obtener y evaluar datos para extraer información útil”. Los resultados de la AD (DA) se pueden utilizar para: identificar áreas clave de riesgos, fraudes, errores o mal uso; mejorar los procesos de negocios; verificar la efectividad de los procesos e influir en las decisiones del negocio. Existen muchos temas para considerar cuando arranca un nuevo programa de AD, incluyendo la maximización del retorno de inversión (ROI), cumplimiento de presupuesto de proyectos, gestión de falsos resultados, aseguramiento de la protección, y confidencialidad de las fuentes de datos y resultados.

Analítica de datos se considera también la ciencia de examinar datos en bruto (crudos) con el propósito de obtener conclusiones acerca de la información contenida en ellos. Se usa en muchas industrias para que las organizaciones y empresas tengan mejoras en la toma de decisiones. Este término suele emplearse en el campo de la inteligencia de negocios (Business Intelligence), y según los fabricantes de herramientas de software puede abarcar una gran variedad de términos: OLAP, CRM, dashboard (tableros de control), etcétera.



Una visión global del Data Analytics

El análisis de Big Data es el proceso de examinar, a una gran velocidad, grandes volúmenes de datos de una amplia variedad de tipos y de gran valor (el modelo de las 4V) para descubrir patrones ocultos, correlaciones desconocidas y otra información útil, de modo que los resultados del análisis puedan proporcionar ventajas competitivas a las organizaciones en relación con la competencia y producir beneficios para el negocio, tales como un marketing (mercadotecnia) más efectivo y eficaz, y mayores ingresos.

Roles profesionales relacionados con datos

Los términos relacionados con la gestión y análisis de datos suelen ser muy variables y en ocasiones sinónimos, por lo que es difícil asociar las competencias a los muchos roles profesionales que surgen a menudo en el ámbito empresarial o de investigación. No obstante, sí que es frecuente encontrar los roles siguientes: analista de datos, ingeniero de datos, arquitecto de datos, científico de datos, ingeniero de visualización. Normalmente, el rol profesional va asociado a una o a todas o algunas de las etapas de la arquitectura de Big Data o del proceso de ciencia de datos que hemos analizado en este capítulo.





Analista de datos. Es el responsable de las etapas de procesamiento y análisis de datos. Su formación fundamental será matemáticas, estadística y economía y negocios (inteligencia de negocios y analítica de datos). Algunas de las tareas a realizar por el analista de datos son: acceso y consultas a diferentes fuentes de datos, proceso y limpieza de datos, resumen de datos, preparar visualizaciones de datos e informes, etc. Las herramientas más utilizadas son de inteligencia de negocios y analítica de datos, tales como Microsoft Excel, Tableau, SAS, SAP, Qlik, MicroStrategy. Pueden utilizar, sobre todo cuando alcanzan ya la especialización, herramientas de minería de datos tales como IBM SPSS, Rapid Miner y KNIME.

Ingeniero de datos. Han adquirido gran importancia en la era de Big Data y de hecho ya tienen hoy día competencias similares al ingeniero de Big Data, que se comenta a continuación. El ingeniero de datos no está tan preocupado con la estadística, analítica y modelado de datos como sus homólogos analistas de datos, y están más implicados en la ingeniería y arquitectura de datos, infraestructuras de computación, almacenamiento y flujo de datos, etc. Los ingenieros de datos son, por consiguiente, los responsables de la arquitectura de datos y de la instalación de la infraestructura necesario. Deben tener conocimientos de computación avanzados y ser programadores expertos. Recientemente se les está asociando con unos nuevos perfiles que están emergiendo, los expertos en DevOps (Desarrollo y Operaciones).

Ingeniero de Big Data. Son los desarrolladores de Big Data. En general son ingenieros de software responsables de realizar los programas establecidos por los analistas, científicos de datos y arquitectos de Big Data. Serán los encargados de diseñar y construir los algoritmos, los sistemas de recolección y almacenamiento de datos, y realizar los programas de gestión de software de la empresa, donde se tengan en cuenta sus planes y líneas de negocio.

Arquitecto de Big Data. Es el responsable de toda la arquitectura y proceso de Big data. Deberá tener una visión global del proyecto y el conocimiento de cada una de las áreas necesarias del proceso, desde la recolección de datos hasta la presentación de resultados a través de las herramientas de visualización. Su formación esencial



será ingeniero informático o ingeniero de sistemas, normalmente especialista de ingeniería de software, pero también con conocimientos de otras ingenierías y de sistemas de información.

Ingeniero de visualización. Dada la importancia que han adquirido las técnicas y herramientas de visualización, se requiere una alta especialización en comunicación, presentaciones y visualizaciones de datos. Requiere una buena formación de ingeniería con una alta especialización en herramientas de visualización. También son demandados especialistas en diseño gráfico, así como en mercadotecnia, comunicaciones y medios sociales.

Científico de datos. Tiene una visión más horizontal de todo el proceso de ciencia de datos. Su tarea principal será la programación de algoritmos para el análisis de datos, pero debe conocer bien el negocio de la empresa (su plan de negocio, así como las líneas de negocio fundamentales). En definitiva, un científico de datos debe ser capaz de identificar aquellas variables relevantes para la empresa que ayuden a mejorar resultados, multiplicar el volumen de ventas, fidelizar a los clientes, ahorrar costes, etc. Los científicos de datos suelen mezclar, entre otros, conocimientos de matemáticas, estadística e informática, a los que es conveniente unir conocimiento de negocios, administración de empresas, ciencias de la salud y recomendable conocimiento de ciencias sociales. El científico de datos se ha consolidado ya como una profesión muy demandada en todo tipo de organizaciones y empresas -grandes y pequeñas. El científico de datos jefe, sobre todo en las grandes multinacionales, se está comenzando a reconvertir en el nuevo rol profesional, de director de datos (CDO, Chief Data Officer) que es el responsable de toda la estrategia y política de datos de las organizaciones y con dependencia directa del presidente o director general.

Autor: Mg. Ing. Federico D'Alía

Bibliografía de referencia: Industria 4.0, Luis Joyanes 2017

Contacto: consultas@elearning-total.com
Web: www.elearning-total.com