

BIG DATA

Aplicación y uso en el mundo empresarial



Quien soy

Economista con más de 25 años de experiencia en empresas tecnológicas. Actualmente ocupo la Dirección de Relaciones Institucionales en la tecnológica granadina **Grupo Trevenque**, Vocal de la Comisión Nacional de Cloud Computing y de la Comisión de Smart Cities en **AMETIC-CEOE**, Presidente de la Comisión de Nuevas Tecnologías y relación con la Universidad, miembro del Comité Ejecutivo y de la Junta Directiva de la Confederación Granadina Empresarios **CGE**, miembro del Consejo Asesor de **EuroCloud España**, Vocal de la Comisión de Nuevas Tecnologías y de la Junta Directiva de la Confederación de Empresas de Andalucía **CEA**, Miembro del Comité Asesor del Centro de Investigación en Tecnologías de La Información y de las Comunicaciones de la **Universidad De Granada (CITIC-UGR)**, Miembro del Patronato de la **Fundación General Universidad De Granada - Empresa** y Secretario General de la **Asociación Cluster Granada Plaza Tecnológica y Biotecnológica - onGranada Tech City**

¿ Que es el BIG DATA?

BIG DATA es un concepto que significa muchas cosas para muchas personas, ha dejado de estar limitado al mundo de la tecnología. Hoy en día se trata de una prioridad empresarial dada su capacidad para influir profundamente en el comercio de una economía integrada a escala global.

Además de proporcionar soluciones a antiguos retos empresariales, **BIG DATA** inspira nuevas formas de transformar procesos, empresas, sectores enteros e incluso la propia sociedad. Aun así, la amplia cobertura mediática que está recibiendo no nos permite distinguir claramente el mito de la realidad: ¿qué está ocurriendo realmente? Varios estudios han puesto de manifiesto que las empresas utilizan **BIG DATA** para obtener resultados centrados en el cliente, aprovechar los datos internos y crear un mejor ecosistema de información.

BIG DATA es la posibilidad de explotar comercialmente una gran cantidad de datos para crear nuevos servicios comerciales. En el mundo se acumulan cada vez más datos en formato digital pero el problema es que estos datos son en general poco estructurados y en cantidades ingentes. Son por lo tanto difíciles de explotar por métodos convencionales. El **BIG DATA** consiste en extraer la inteligencia necesaria y darles valor.

Las dimensiones del BIG DATA

Volumen



Datos a escala

De terabytes a petabytes de datos

Variedad



Datos en muchas formas

Estructurados, no estructurados, texto, multimedia

Velocidad



Datos en movimiento

Análisis de datos en streaming para una toma de decisiones en cuestión de segundos

Veracidad



Incertidumbre de datos

Gestionar la fiabilidad y previsibilidad de tipos de datos intrínsecamente imprecisos

Volumen: la cantidad de datos. Siendo quizá la característica que se asocia con mayor frecuencia a big data, el volumen hace referencia a las cantidades masivas de datos que las organizaciones intentan aprovechar para mejorar la toma de decisiones en toda la empresa.

Los volúmenes de datos continúan aumentando a un ritmo sin precedentes. No obstante, lo que constituye un volumen verdaderamente “alto” varía en función del sector e incluso de la ubicación geográfica y es más pequeño que los petabytes y zetabytes a los que a menudo se hace referencia.

Variedad: diferentes tipos y fuentes de datos. La variedad tiene que ver con gestionar la complejidad de múltiples tipos de datos, incluidos los datos estructurados, semiestructurados y no estructurados. Las organizaciones necesitan integrar y analizar datos de un complejo abanico de fuentes de información tanto tradicional como no tradicional procedentes tanto de dentro como de fuera de la empresa.

Con la profusión de sensores, dispositivos inteligentes y tecnologías de colaboración social, los datos que se generan presentan innumerables formas entre las que se incluyen texto, datos web, tuits, datos de sensores, audio, vídeo, secuencias de clic, archivos de registro y mucho más.

Velocidad: los datos en movimiento. La velocidad a la que se crean, procesan y analizan los datos continúa aumentando. Contribuir a una mayor velocidad es la naturaleza en tiempo real de la creación de datos, así como la necesidad de incorporar datos en streaming a los procesos de negocio y la toma de decisiones. La velocidad afecta a la latencia: el tiempo de espera entre el momento en el que se crean los datos, el momento en el que se captan y el momento en el que están accesibles.

Hoy en día, los datos se generan de forma continua a una velocidad a la que a los sistemas tradicionales les resulta imposible captarlos, almacenarlos y analizarlos. Para los procesos en los que el tiempo resulta fundamental, tales como la detección de fraude en tiempo real o el marketing “instantáneo” multicanal, ciertos tipos de datos deben analizarse en tiempo real para que resulten útiles para el negocio.

Veracidad: la incertidumbre de los datos. La veracidad hace referencia al nivel de fiabilidad asociado a ciertos tipos de datos. Esforzarse por conseguir unos datos de alta calidad es un requisito importante y un reto fundamental de big data, pero incluso los mejores métodos de limpieza de datos no pueden eliminar la imprevisibilidad inherente de algunos datos, como el tiempo, la economía o las futuras decisiones de compra de un cliente. La necesidad de reconocer y planificar la incertidumbre es una dimensión de big data que surge a medida que los directivos intentan comprender mejor el mundo incierto que les rodea

¿ De dónde vienen todos estos datos ?

Son principalmente fruto de la revolución digital en curso:

- Finanzas y economía: Balances, cotizaciones, precios, hipotecas, datos de consumo, ...
- Mundo máquina (M2M): Sensores conectados a la red que transmiten datos industriales (coches, contadores, ...)
- Datos socio-ambientales: clima, astronomía, historiales médicos, natalidad, estadísticas gubernamentales, ...
- Redes sociales y movilidad: fotos, videos, datos de geolocalización, ...
- Wearables: Dispositivos conectados a las personas que almacenan datos vitales, de hábitos o comportamientos.
- Dispositivos móviles: Que contemplan y acumulan muchos de los puntos anteriores a los cuales se suman datos específicos de los dispositivos inteligentes.

Tecnologías que permiten su explotación

Son en general técnicas que combinan modelos predictivos, estadísticas, inteligencia artificial, procesamiento en lenguaje natural, redes neuronales, patern matching, ...

En definitiva un conjunto de técnicas algorítmicas que salen de las técnicas habituales de las bases de datos transaccionales.

Existen dos grandes grupos de soluciones en las cuales se enmarcan los principales proveedores y fabricantes:

Aplicaciones

Vertical – Consumer - Business Intelligence - Analytics and Visualization - Operational Intelligence - Ad / Media Apps - Data as a service

Infraestructuras

Analytics – Operational - As A Service - Structured DB - Technologies

The Big Data Landscape

Apps

Vertical Apps



Operational Intelligence



Ad / Media Apps



Data As A Service



Business Intelligence



Analytics And Visualization



Infrastructure

Analytics Infrastructure



Operational Infrastructure



Infrastructure As A Service



Structured Databases



Technologies



Consumer

amazon ebay



Google



NETFLIX



@WalmartLabs

Examen de las herramientas del Big Data

Gestión de datos

Políticas, procesos y prácticas imprescindibles para gestionar de manera adecuada la eficacia, la precisión, la fiabilidad y la disponibilidad de los datos.

- **Integración de datos:** Preparar y consolidar estructuras y fuentes diversas en un conjunto de datos cohesionado para el análisis.
- **Calidad de datos:** Limpiar con fiabilidad los datos así como eliminar los duplicados y los errores.
- **Seguridad de datos:** Detectar y proteger los datos en todos los almacenes de datos asignando reglas relativas al uso, al acceso y a los permisos.
- **Máquina virtual de datos:** Crear una capa de abstracción para los datos que asocie con amplitud el procesamiento de datos desde el entorno de implantación subyacente.
- **Gestión de datos maestros:** Almacenar una única versión consolidada, completa y fidedigna de la realidad de los diversos dominios de datos.
- **Estructura de datos distribuida:** Emplear una tecnología como Hadoop para escalar de forma rentable las exigencias de almacenamiento y procesamiento.
- **Data warehouse:** Tecnología de data warehouse que admita los requisitos de rendimiento, uso y escalabilidad que tienen los análisis de Big Data y las integraciones con las infraestructuras de Hadoop.

Examen de las herramientas del Big Data

Entrega de datos

Proceso consistente en enviar los datos que posee a los sistemas y las aplicaciones que los necesitan.

- **Carga en batch:** Escalar con eficiencia la carga en batch de los datos entre los sistemas back-end operativos, los análisis y las fuentes.
- **Transmisión en tiempo real:** Transmisión de datos en tiempo real a las aplicaciones, los análisis y los sistemas back-end que la requieren.
- **Hub de integración de datos:** Permite que los datos estén disponibles mediante un enfoque como el del modelo de publicación y suscripción a fin de evitar la proliferación de integraciones de punto a punto.
- **Virtualización de datos:** Entregar datos de sus sistemas sin sobrecargarlos.
- **Procesamiento basado en eventos:** Detectar y analizar amenazas, oportunidades y otros eventos críticos para el negocio y darles una respuesta en tiempo real.

Examen de las herramientas del Big Data

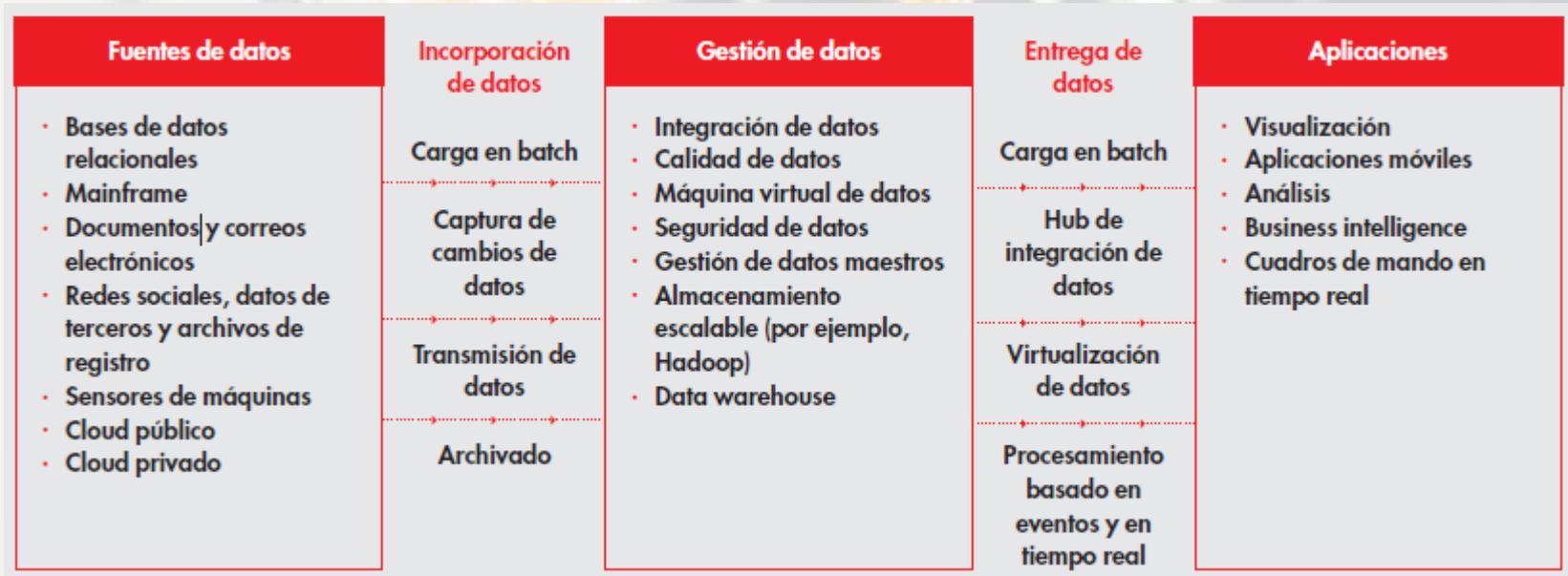
Análisis

Herramientas y procesos que permiten convertir los datos sin procesar en conocimientos, patrones, predicciones y cálculos sobre el dominio objeto de análisis.

- **Visualización:** Presentar sus datos y conclusiones de manera que resulten fáciles de comprender y asimilar.
- **Análisis avanzados:** Aplicar algoritmos analíticos de vanguardia a sus conjuntos de datos a fin de efectuar cálculos complejos.
- **Aprendizaje automatizado:** Aplicar algoritmos avanzados de aprendizaje automatizado con el fin de identificar patrones y elaborar predicciones imposibles de manejar con un ancho de banda manual.

Arquitectura idónea del Big Data

En el diagrama siguiente, se representa el modo recomendado para crear la arquitectura idónea de tecnologías y procesos de Big Data para la empresa.



Beneficios del BIG DATA en las empresas

El fenómeno Big Data abre paso a una generación de nuevas empresas que cuentan con programadores, estadísticos, analistas cuantitativos, econométricos y analistas de extracción de datos para lograr resultados provechosos para las compañías. Los beneficios de exprimir la máxima utilidad de Big Data son:

Ventajas competitivas en la gran distribución: permite actualizar, optimizar y afinar inventarios en función de la demanda en tiempo real. El data mining o extracción de datos posibilita el análisis del comportamiento de los clientes, fijar los precios en consecuencia u ofrecer los incentivos adecuados para atraer a los clientes. Es posible realizarlo mediante la digitalización de elementos como los almacenes o las cajas registradoras.

Mejora de la eficiencia y los costes: el análisis del Big Data puede acelerar la velocidad con que se desarrolla un producto. También permite compartir datos de forma rápida y realizar simulaciones de producto. En algunos sectores, los plazos de desarrollo se han llegado a reducir entre el 30% y el 50%.

Beneficios del BIG DATA en las empresas

Mejora de la gestión empresarial: además de optimizar la cadena de suministro y el inventario, el Big Data puede ser útil para reducir el ciclo de conversión de efectivo, controlar factores de riesgo y tomar decisiones empresariales que pueden virar el futuro de la empresa fundamentadas en datos recogidos en tiempo real.

Almacenamiento en la nube: uno de los problemas para gestionar altos volúmenes de datos es el elevado coste de la infraestructura de almacenamiento. Muchos proveedores de almacenamiento masivo de datos alquilan potentes servidores a los que se puede acceder en línea, y ponen a disposición del cliente como solución almacenarlos en una especie de nube. El resultado es que se puede acceder a ella mediante aplicaciones diseñadas para manejar grandes volúmenes de datos y se pueden obtener soluciones a menudo en tiempo real de forma sencilla.

Conclusiones y expectativas futuras

La velocidad con la que crecen los datos es vertiginosa. Se estima que la cifra se dobla cada dos años (Informe EMC. 2011 Digital Universe Study) y, según el informe Ericsson Mobility Report (agosto 2014), su tráfico creció un 60 % entre el segundo trimestre de 2013 y el de 2014. En 2020, habrá 50.000 millones de conexiones de personas, procesos, datos y objetos en Internet.

Las empresas cuentan con un volumen creciente de datos sobre sus clientes, proveedores y operaciones. Podrán acceder cada vez a mayor cantidad de datos que se recogen, comparten y analizan cada día; y su variedad también se diversifica: vídeos, imágenes, comentarios en redes sociales, transacciones eCommerce, búsquedas en Internet, comunicaciones 3G, 4G, GPS...

Millones de sensores conectados en red están incrustados en diversos tipos de dispositivos como teléfonos móviles, medidores inteligentes de energía, automóviles y máquinas industriales que crean y comunican datos en la era de la Internet de las Cosas.

El McKinsey Global Institute (MGI) lo tiene claro. En su informe Big data: la próxima frontera de la innovación, la competencia y la productividad, argumenta que los datos se están convirtiendo en un factor de producción, como lo es el capital físico o humano.

Conclusiones y expectativas futuras

Mejora de la experiencia de cliente. Una de las tendencias identificadas es la que convierte el Customer Experience, o Experiencia de Cliente, en el nuevo marketing. Cada vez son más las compañías que están poniendo el foco en la experiencia de cliente, desarrollando procedimientos dirigidos a satisfacer sus necesidades, que aporten el máximo valor añadido y que permitan superar las expectativas del proceso de compra.

En este contexto, el análisis de Big Data va a permitir conocer más y mejor al consumidor, identificar estilos de vida y definir patrones de comportamiento, de gasto, etc. Y conocer mejor al consumidor es un primer paso para anticiparse a sus necesidades. Cómo la compañía Amazon se ha beneficiado del Big Data es un ejemplo real. Amazon ha crecido mucho y, en buena parte, la fórmula de su éxito está asociada a cómo han sido capaces de crear valor en su relación con el cliente, gracias a la minería de datos.

Microsegmentación y geolocalización: Gracias al análisis y al cruce del Big Data, las compañías podrán conocer de manera muy exhaustiva al consumidor, sus características, preferencias, necesidades y deseos. Esto les permitirá acercarse mucho mejor a determinados segmentos de población optando por la microsegmentación y la geolocalización para aplicar estrategias muy personalizadas y diferenciadas, presentar ofertas de compra, estrategias de comunicación a medida o desarrollar campañas que se ejecutan en tiempo real por medio del móvil.

Conclusiones y expectativas futuras

Content Marketing: Todas las previsiones para 2015 apuntan a la consolidación del marketing de contenidos. La publicidad tradicional ha saturado a unos consumidores cada vez más críticos y que buscan un valor añadido al mensaje que les llega. Las marcas ya no están en disposición de hacer un monólogo homogéneo y conseguir audiencia.

Deben conversar con sus consumidores y ofrecerles contenido útil y de calidad, personalizado y segmentado, disponible en multidispositivos y que pueda ser fácilmente compartido, a través de una gran variedad de canales.

En este contexto, la gran cantidad de datos segmentados de que disponen las organizaciones abre una oportunidad para comprender al público objetivo y ayudar a diseñar el contenido que mejor conecte con sus intereses.

Además, los propios datos y su visualización pueden convertirse en material para construir historias o complementar nuestra narrativa, contribuyendo a que ideas complejas puedan ser comunicadas con mayor claridad, eficiencia y de manera atractiva.

Bibliografía y fuentes de la información

El uso de big data en el mundo real - IBM Global Business Services

El gran cuaderno del Big Data – Informática

Inteligencia Artificial, Inteligencia Computacional y Big Data – Francisco Herrera

Enlaces de interés

“El análisis masivo de datos está transformando a la empresa”, John Veihmeyer, Presidente mundial de KPMG (Miguel Ángel García. El País, 25.01.2015)

http://economia.elpais.com/economia/2015/01/22/actualidad/1421926110_795741.html

La ciencia de los datos y el conocimiento en la empresa (Fernando Piera) (CEDE, Cuadernos de Gestión del Conocimiento Empresarial nº50. Noviembre 2014)

http://www.directivoscede.com/sites/default/files/document/conocimiento/20-11-2014/1411_num50_piera.pdf

Big Data y Social Media en retrospectiva (Francisco Rangel. Autoritas, 2.04.2015)

<http://www.autoritas.net/es/big-data-y-social-media-en-retrospectiva/>

– Tendencias de análisis Big Data en comunicación digital (Francisco López y Marina Pulido. Top Comunicación, 10.02.2015)

<http://www.topcomunicacion.com/noticia/6261/cuales-son-las-principales-tendencias-de-analisis-big-data-en-comunicacion-digital>

Gracias por su atención

Vito Episcopo

eMail: vito@trevenque.es

eMail: vito@ongranada.com

Twitter: @vitoeps

LinkedIn: <https://es.linkedin.com/in/vitoe>

Móvil: +34 670 58 33 00

Web site: www.trevenque.es

Web site: www.ongranada.com