

Big Data, la nueva oportunidad

Big Data

“La nueva oportunidad”



Autor:
John Henry A. Morales
Dr. Ingeniero Industrial
por la Universidad de la
Rioja, España.
Analista de Datos y
Especialista en
aplicación de
estadística en los
procesos, negocios e
investigaciones.
Profesor de la
Universidad
Iberoamericana
(UNIBE) en las Cátedras
de Diseño de
Experimentos y
Pronósticos en los
negocios en la Escuela
de Ingeniería Industrial.

1. Conociendo Big Data

1.1 Big Data

¿Qué es Big Data y porqué se ha vuelto tan importante? pues bien, en términos generales podríamos referirnos como a la tendencia en el avance de la tecnología que ha abierto las puertas hacia un nuevo enfoque de entendimiento y toma de decisiones, la cual es utilizada para describir enormes cantidades de datos (estructurados, no estructurados y semi-estructurados) que tomaría demasiado tiempo y sería muy costoso cargarlos a una base de datos relacional para su análisis.

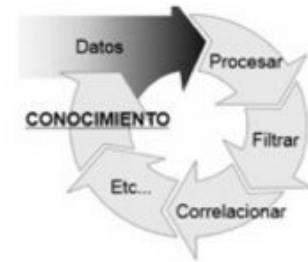
Como un problema de Big Data se puede contemplar la forma como hoy crecen los datos en volumen, velocidad y variedad; esto es debido al gran avance y uso de las tecnologías de información, y al uso diario que las personas hacen de ellas. Big Data son “cantidades masivas de datos que se acumulan con el tiempo que son difíciles de analizar y manejar utilizando herramientas comunes de gestión de bases de datos” se refiere “al tratamiento y análisis de enormes repositorios de datos, tan desproporcionadamente grandes que resulta imposible tratarlos con las herramientas de bases de datos y analíticas convencionales”.

Gartner según Camargo-Vega (2014) define el Big Data como “un gran volumen, velocidad o variedad de información que demanda formas costeables e innovadoras de procesamiento de información que permitan ideas extendidas, toma de decisiones y automatización del proceso”. Y Russom (2012) indica que “Big Data consiste en consolidar toda la información de una organización y ponerla al servicio del negocio”.

Big Data es la clave para poder explotar y sacar valor de los datos disponibles. Permite la captura y transmisión de datos en tiempo real (o casi-tiempo real) de diferentes fuentes como el Internet de las cosas (IoT – Internet of Things), dispositivos (sensores, RFID, video cámaras, parquímetros, radares y detectores de matrículas, smartphones y otros dispositivos móviles), sistema de integración de bases de datos diversas y directamente de los ciudadanos, instituciones y empresas vía aplicaciones de colaboración o redes sociales, que permiten hacer cruces de variables y análisis estadísticos y matemáticos avanzados obteniendo un gran número de informaciones y conocimiento en tiempo real.

Big Data (Big data Analytics), se apoya en herramientas tecnológicas no convencionales para poder brindar un medio donde se pueda analizar ese conglomerado complejo de diversas data, para tener posteriormente como resultado informaciones que le permitan a los negocios que hacen uso del Big Data, tomar mejores decisiones de manera rápida (IBM, n.d).

Big Data como una nueva generación de tecnologías y arquitecturas diseñadas para extraer valor económico de grandes volúmenes de una amplia variedad de datos, mediante la capacidad de captura, descubrimiento y/o análisis a gran velocidad. "La información es poder, y Big Data, si se gestiona correctamente, puede proporcionar una cantidad de conocimiento...



Fuente: Sevillano Pérez, F. (n.d)

1.2 ¿De dónde proviene la información?

Según Paredes-Moreno (2015) los seres humanos estamos creando y almacenando información constantemente y cada vez más, en cantidades astronómicas. Se podría decir que si todos los bits y bytes de datos del último año fueran guardados en CD's, se generaría una gran torre desde la Tierra hasta la Luna ida y vuelta. Esta contribución a la acumulación masiva de datos la podemos encontrar en diversas industrias, las compañías mantienen grandes cantidades de datos transaccionales, reuniendo información acerca de sus clientes, proveedores, operaciones, etc., de la misma manera sucede con el sector público. En muchos países se administran enormes bases de datos que contienen datos de censo de población, registros médicos, impuestos, etc., y si a todo esto le añadimos transacciones financieras realizadas en línea o por dispositivos móviles, análisis de redes sociales (en Twitter son cerca de 12 Terabytes de tweets creados diariamente y Facebook almacena alrededor de 100 Petabytes de fotos y videos), ubicación geográfica mediante coordenadas GPS, en otras palabras, todas aquellas actividades que la mayoría de nosotros realizamos varias veces al día con nuestros smartphones. Estamos hablando de que se generan alrededor de 2.5 quintillones de bytes diariamente en el mundo (1 quintillón = 1,000,000,000,000,000,000,000,000,000 bytes).

Tipos de datos según procedencia:

Web and Social Media Ejemplo: www, Facebook, Twitter, LinkedIn, blogs, etc.

Machine-to-Machine Ejemplo: Velocidad, temperatura, presión, variables meteorológicas, variables químicas, etc.

Big Transaction Data Ejemplo: Incluye registros de facturación, en telecomunicaciones los llamados registros detallados de las llamadas (Call Detail Record o CDR), etc.

Biometrics Ejemplo: Huellas digitales, escaneo de la retina, reconocimiento facial, genética, etc.

Human Generated Ejemplo: Notas de voz, correos electrónicos, documentos electrónicos, resultados de estudios médicos, multas, etc.

1.3 Dimensiones de Big Data

Existen tres características o dimensiones del Big Data: Volumen, Velocidad y Variedad.

1) Volumen: Cada día, las empresas registran un aumento significativo de sus datos (terabytes, petabytes y exabytes), creados por personas y máquinas. Las redes sociales también generan datos, es el caso de Twitter, que por sí sola genera más de 7 terabytes (TB) diariamente, y de Facebook, 10 TB de datos cada día. Algunas empresas generan terabytes de datos cada hora de cada día del año, es decir, las empresas están inundadas de datos.

2) Variedad: las soluciones Big Data deben encargarse de más tipos de datos, tanto estructurados como desestructurados. 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. Sus distintas aplicaciones comprenden también la visualización y el análisis estadístico de datos meteorológicos, sismológicos, genética animal, etc. Además del gran volumen de información, esta se produce y existe en una gran variedad de datos que pueden ser representados de diversas maneras en todo el mundo, por ejemplo de dispositivos móviles, audio, vídeo, sistemas GPS, incontables sensores digitales en equipos industriales, automóviles, medidores eléctricos, veletas, anemómetros, etc., los cuales pueden medir y comunicar el posicionamiento, movimiento, vibración, temperatura, humedad y hasta los cambios químicos que sufre el aire, de tal forma que las aplicaciones que analizan estos datos requieren que la velocidad de respuesta sea lo suficientemente rápida para lograr obtener la información correcta en el momento preciso.

3) Velocidad: 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.

2. Big Data: Tecnologías de Análisis de Datos

- | | | | |
|---------------------------------|----------------------|-------------|-------------|
| • BigQuery | • Sap Hana | • Cassandra | • Lucene |
| • ThinkUp | • Oracle Big Data A. | • Chukwa | • Oozie |
| • Infosphere Streams | • HDinsight | • Flume | • Pig |
| • Biginsights Infosphere | • Textalytics | • HBase | • ZooKeeper |
| • System PureData | • Hadoop | • Hive | |
| • Infosphere Information Server | • Avro | • Jaql | |

3. Beneficios del Big Data

Las empresas que saben sacar provecho del Big Data pueden mejorar su estrategia y así permanecer en el mercado posicionadas, pues hará uso de nuevos conocimientos, con el gran volumen de datos o información que maneja a diario, que inicialmente no se les dio la suficiente importancia, por no tener una herramienta tecnológica que permitiera procesarla. Con Big Data, las empresas pueden ofrecer mejores productos, desarrollar excelentes relaciones con sus clientes, además, se transforman en más ágiles y competitivas.

4. Recomendaciones

Paredes-Moreno (2015) recomienda:

- Centrarse en el cliente: Es fundamental que las empresas centren sus iniciativas de Big Data en ámbitos que puedan proporcionar el máximo valor para el negocio.
- Desarrollar un proyecto de Big Data para toda la empresa: Un proyecto abarca la visión, la estrategia y los requisitos de Big Data dentro de una empresa y resulta fundamental para armonizar las necesidades de los usuarios de negocio con la hoja de ruta de la implementación de TI.
- Desarrollar funcionalidades analíticas: Las empresas tendrán que invertir en adquirir tanto herramientas como habilidades. Al mismo tiempo las universidades y los propios individuos, independientemente de su formación o especialidad, tienen la obligación de desarrollar sólidas habilidades analíticas.
- Crear negocios sobre resultados cuantificables: Desarrollar una estrategia de Big Data exhaustiva y viable, así como la posterior hoja de ruta requiere un caso de negocio sólido y cuantificable.

Muchas empresas basan sus casos de negocio en las siguientes ventajas que se pueden derivar de Big Data:

- Decisiones más inteligentes: Aprovechar nuevas fuentes de datos para mejorar la calidad de la toma de decisiones.
- Decisiones más rápidas: Permitir una captura y análisis de datos en tiempo más real para respaldar la toma de decisiones en el "punto de impacto", por ejemplo cuando un cliente está navegando por su sitio web o al teléfono con un representante del servicio de atención al cliente.
- Decisiones que marquen la diferencia: Centrar las iniciativas de Big Data en ámbitos que proporcionen una verdadera diferenciación.

BIBLIOGRAFÍA

- Camargo-Vega, J.J.; Camargo-Ortega, J.F.; Joyanes-Aguilar, L. (2014), Conociendo Big Data, Revista Facultad de Ingeniería (Fac. Ing.), Enero-Abril 2015, Vol. 24, No. 38, pp. 63-77.
- IBM (n.d), What is Big Data Analytic?, consultado el 27 de abril del 2016, de la URL: <http://www-01.ibm.com/software/data/infosphere/ha-doop/what-is-big-data-analytics.html>
- Monleón-Getino, Antonio (2015), El impacto del Big-data en la Sociedad de la Información, significado y utilidad, Historia y Comunicación Social, Universidad de Barcelona, Vol. 20, Núm. 2 (2015) 427-445.
- Paredes-Moreno, A. (2015), Big Data: Estado de la cuestión, International Journal of Information Systems and Software Engineering for Big Companies (IJSEBC), Departamento de Economía Financiera y Dirección de Operaciones, Universidad de Sevilla, España.
- Russom, M. (2012), Big Data Analytics, TDWI (The Data Warehousing Institute).
- Salgado, M (2014), Oracle apuesta por Big Data con tecnología y proyectos. Disponible en: <http://www.computerworld.es/big-data/>
- SEVILLANO PÉREZ, F. (n.d), Big Data. Consultado de la Base de datos de dialnet en fecha 2 de abril del 2016.

5. CONCLUSIONES

Camargo-Vega afirma que es necesario prepararse para contratar o reciclar personal, pues las empresas u organizaciones carecen de personas capacitadas en Big Data. Además, "proyecta que para el 2018, solo en Estados Unidos, se necesitarán entre 140 mil y 190 mil nuevos expertos en métodos estadísticos y tecnologías de análisis de datos, incluyendo el ampliamente publicitado papel de científico de datos". Las universidades deben prestar seria atención al tema y dotar a los estudiantes de ingeniería de las herramientas y competencias necesarias para que estos puedan entender, aprender, aplicar y aprovechar Big Data con facilidad.

Según Monleón-Getino la revolución digital ha hecho que dispositivos de cualquier tamaño y un sinfín de aplicaciones informáticas (ordenadores, teléfonos inteligentes, dispositivos digitales, sensores, micrófonos, cámaras, escáneres médicos, imágenes, redes sociales, etc) estén presentes en nuestras vidas.

La sociedad genera constantemente datos y la mayoría son almacenados digitalmente; son información geográfica, estadística, datos meteorológicos, datos de la investigación, datos de transporte, datos de consumo de energía, datos de salud, redes sociales, banca on-line, etc. Es lo que se ha denominado los grandes datos o Big-data, aunque técnicamente se empieza a hablar de Big-data a partir de petabytes (PB) y exabytes (EB) de datos.

Junto con el capital y la fuerza de trabajo los datos se han convertido en un elemento fundamental para la economía que refleja una revolución en la que la sociedad será dirigida por los datos.

Se prevé la creación de miles de nuevos puestos de trabajo fundados en que la economía basada en los datos estimulará la investigación y la innovación aportando más oportunidades de negocio y el aumento de la disponibilidad del conocimientos y el capital, en particular para las pequeñas y medianas empresas (PYME). Con la capacidad de generar toda esta información valiosa procedente de los sistemas digitales, las empresas y los gobiernos están lidiando con el problema de analizar los datos para dos propósitos importantes: ser capaces de detectar y responder a los acontecimientos actuales de una manera oportuna, y para poder utilizar las predicciones del aprendizaje histórico con todos los datos recogidos anteriormente. Esta situación requiere del análisis tanto de datos continuos (datos actuales) como de datos precedentes (datos históricos), que son representados en diferentes y enormes