

“Big Data”

FBRTL63

Fernando Barranco Rodríguez

11 de enero de 2017

Contenido

- 1 ¿Qué es Big Data?
 - Introducción
 - ¿“Cuánto” es Big Data?
 - Las 3 “Vs” del Big Data
 - Tipos de Datos
- 2 Internet de las Cosas (IoT)
- 3 Ciencia de Datos
 - Ciencia de Datos vs BI y Big Data
 - Científico de Datos
- 4 Referencias

Introducción

Debido al gran avance que existe día con día en las tecnologías de información, las organizaciones se han tenido que enfrentar a nuevos desafíos que les permitan analizar, descubrir y entender más allá de lo que sus herramientas tradicionales reportan sobre su información, al mismo tiempo que durante los últimos años el gran crecimiento de las aplicaciones disponibles en Internet han sido parte importante en las decisiones de negocio de las empresas.

El primer cuestionamiento que posiblemente llegue a su mente en este momento es ¿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.

El concepto de Big Data aplica para toda aquella información que no puede ser procesada o analizada utilizando procesos o herramientas tradicionales.

El uso de Big Data ha ayudado a los investigadores a descubrir cosas que les podrían haber tomado años en descubrir por si mismos sin el uso de estas herramientas, debido a la velocidad del análisis.

Analicemos primeramente en términos de bytes:

- Gigabyte = $10^9 = 1,000,000,000$

Analicemos primeramente en términos de bytes:

- Gigabyte = $10^9 = 1,000,000,000$
- Terabyte = $10^{12} = 1,000,000,000,000$

Analicemos primeramente en términos de bytes:

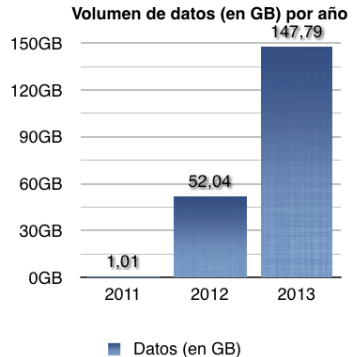
- Gigabyte = $10^9 = 1,000,000,000$
- Terabyte = $10^{12} = 1,000,000,000,000$
- Petabyte = $10^{15} = 1,000,000,000,000,000$

Analicemos primeramente en términos de bytes:

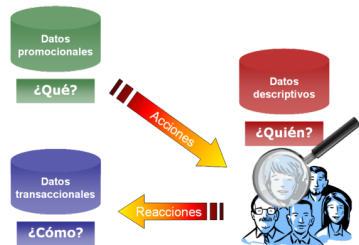
- Gigabyte = $10^9 = 1,000,000,000$
- Terabyte = $10^{12} = 1,000,000,000,000$
- Petabyte = $10^{15} = 1,000,000,000,000,000$
- Exabyte = $10^{18} = 1,000,000,000,000,000,000$

Además del gran **volumen** de información, esta existe en una gran **variedad** de datos que pueden ser representados de diversas maneras en todo el mundo, 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 demasiado rápida para lograr obtener la información correcta en el momento preciso. Estas son las características principales de una oportunidad para Big Data.

Volumen



Variedad



Velocidad

Velocidad de generación y distribución de los datos

En cada fuente los datos pueden generarse a distinta velocidad...



Producto Interior Bruto - Anual



Resumen de ventas de una organización: trimestral, mensual, diario...



Información financiera - Segundos

33

Muchas organizaciones se enfrentan a la pregunta sobre ¿qué información es la que se debe analizar?, sin embargo, el cuestionamiento debería estar enfocado hacia ¿qué problema es el que se está tratando de resolver?

Si bien sabemos que existe una amplia variedad de tipos de datos a analizar, una buena clasificación nos ayudaría a entender mejor su representación.

Tipo de Dato	Descripción
Web and Social Media	Incluye contenido web que es obtenida de las redes sociales
Machine-to-Machine (M2M)	Son las tecnologías que permiten conectarse a otros dispositivos
Big Transaction Data	Registros detallados de telecomunicaciones
Biometrics	Información biométrica, incluye huellas digitales, reconocimiento facial, etc.
Human Generated	Datos generados por las personas como llamadas a celular, correos, etc.

Internet de las Cosas (IoT)

Internet de las cosas (en inglés, *Internet of things*, abreviado IoT) es un concepto que se refiere a la interconexión digital de objetos cotidianos con Internet. Alternativamente, es el punto en el tiempo en el que se conectarían a Internet más cosas u objetos que personas.

La Ciencia de datos es un campo interdisciplinario que involucra a los procesos y sistemas para extraer conocimiento o un mejor entendimiento de grandes volúmenes de datos en sus diferentes formas y formatos.

Algunos campos de análisis de datos son: la minería de datos y la analítica predictiva.

El término Business Intelligence (BI) también se ha popularizado en nuestros tiempos e incluso, se ha llegado a utilizar de manera indiscriminada con el concepto de ciencia de datos para referirse al análisis de datos, pero en realidad existen diferencias abismales entre dichos conceptos, a continuación se presenta algunas de sus diferencias.

Ciencia de Datos

- Trabaja en datos incompletos.
- Los datos suelen estar desordenados.
- Analiza los datos para ver qué información obtiene.
- Grandes conjuntos de datos que es un desafío administrar.
- Los hallazgos impulsan decisiones sobre operaciones y productos.

Business intelligence (BI)

- Conjuntos de datos completos.
- Archivos de datos limpios.
- Informa lo que dicen los datos.
- Conjunto de datos manejable.
- Sus hallazgos miden el rendimiento pasado.

Las personas que se dedican a la ciencia de datos se les conoce como científico de datos, de acuerdo con el proyecto Master in Data Science define al científico de datos como una mezcla de estadísticos, computólogos y pensadores creativos, con las siguientes habilidades:

- (i) Recopilar, procesar y extraer valor de las diversas y extensas bases de datos.

Las personas que se dedican a la ciencia de datos se les conoce como científico de datos, de acuerdo con el proyecto Master in Data Science define al científico de datos como una mezcla de estadísticos, computólogos y pensadores creativos, con las siguientes habilidades:

- (i) Recopilar, procesar y extraer valor de las diversas y extensas bases de datos.
- (ii) Imaginación para comprender, visualizar y comunicar sus conclusiones a los no científicos de datos.

Las personas que se dedican a la ciencia de datos se les conoce como científico de datos, de acuerdo con el proyecto Master in Data Science define al científico de datos como una mezcla de estadísticos, computólogos y pensadores creativos, con las siguientes habilidades:

- (i) Recopilar, procesar y extraer valor de las diversas y extensas bases de datos.
- (ii) Imaginación para comprender, visualizar y comunicar sus conclusiones a los no científicos de datos.
- (iii) Capacidad para crear soluciones basadas en datos que aumentan los beneficios, reducen los costos.

Las personas que se dedican a la ciencia de datos se les conoce como científico de datos, de acuerdo con el proyecto Master in Data Science define al científico de datos como una mezcla de estadísticos, computólogos y pensadores creativos, con las siguientes habilidades:

- (i) Recopilar, procesar y extraer valor de las diversas y extensas bases de datos.
- (ii) Imaginación para comprender, visualizar y comunicar sus conclusiones a los no científicos de datos.
- (iii) Capacidad para crear soluciones basadas en datos que aumentan los beneficios, reducen los costos.
- (iv) Trabajan en todas las industrias y hacen frente a los grandes proyectos de datos en todos los niveles.

Para más detalles de que es y que no es Big Data, consultar:



IBM developers.

En <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>.

8 Habilidades para convertirse en Data Scientist (curso), consultar:



Udacity.

En <http://blog.udacity.com/2014/11/data-science-job-skills.html>.

Mi blog. Aquí podrás encontrar información actualizada y todo lo relacionado a la Ciencia de Datos:



Google+.

En <https://plus.google.com/101221011392362358312>.