

Introduciéndonos al Big Data

Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...

Con esta frase empezó Dan Ariely una de sus conferencias. Lo que quiere decir es que hay numerosas empresas que dicen que ellos realizan tareas de análisis de datos para obtener mejores resultados que la competencia en determinados campos, pero, ¿realmente están haciendo “big data”? En la Imagen 1 se muestra la opinión de los directivos sobre qué es esta nueva “buzzword”.

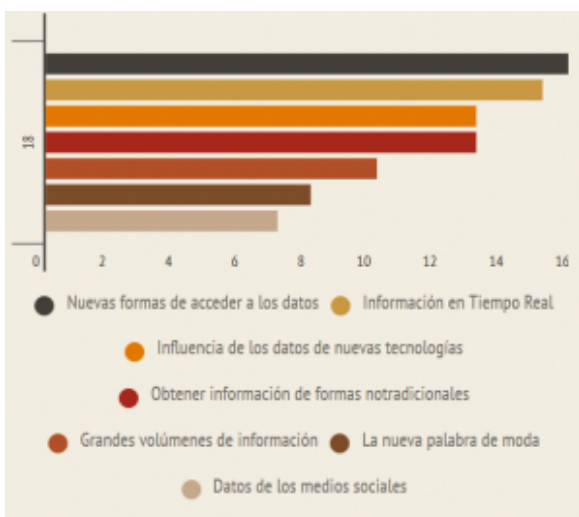


Imagen 1: Opinión de los directivos sobre qué es el big data

Una vez que se ha logrado obtener las diferentes opiniones, es el momento de definir lo que es el “big data”. Big Data es el tratamiento informatizado de grandes cantidades de información, la definición de lo que es Big Data no ha cambiado con el tiempo, puesto que los sistemas informáticos son cada vez más potentes y cada vez pueden almacenar y procesar más datos de lo que se podía antes. Además dependen

de la capacidad del procesador, para algunos el problema está en procesar cientos de gigabytes, mientras que para otros se trata de petabytes cuando se encuentran con problemas.

“Big data” es un término aplicado a conjuntos de datos que superan la capacidad del software habitual para ser capturados, gestionados y procesados en un tiempo razonable. Los tamaños del “big data” se encuentran constantemente en movimiento creciente, de esta forma en 2012 se encontraba dimensionada en un tamaño de una docena de terabytes hasta varios petabytes de datos en un único data set. Debido a esto los métodos tradicionales para analizar y tratar toda esta información son poco útiles y se han ido explorando diferentes soluciones que lo que buscan es paralelizar el tratamiento de esta información.

Los datos que se analizan usando diferentes técnicas de “Big Data” pueden provenir de diferentes fuentes como son las siguientes:

- Generados por personas: a través del empleo de diferentes aplicaciones lo cual, genera metadatos que pueden ser analizados.
- Transacciones de datos: La facturación, las llamadas o las transacción entre cuentas generan información que tratada pueden ser datos relevantes. Este procedimiento también genera metadatos.
- La web: cada vez que una persona visita una web deja un rastro tras ella y a través de diferentes herramientas de tracking se podrán predecir sus futuros movimientos
- Machine to Machine: estos datos se producen al ser compartidos de un dispositivo a otro.
- Biométricos: son el conjunto de datos que provienen de la seguridad, defensa y servicios de inteligencia

Tras obtener estos datos, éstos se deben almacenar en único lugar, en un una base de datos no relacional, ya que las bases de datos tradicionales son incapaces de procesar cantidades

tan grandes de datos.