

BIG DATA

Gonzalo
Rodríguez
Palacio

ÍNDICE

1	Introducción	2
2	Pasos para el tratamiento del big data.....	3
2.1	Captura.....	3
2.2	Transformación.....	3
2.3	Almacenamiento NoSQL	3
2.4	Análisis de los datos	3
2.5	Visualización de los datos.....	4
3	Aplicaciones	5
4	Bibliografía.....	6

1 Introducción

Big data es un concepto que hace referencia a un conjunto de datos tan grandes que las aplicaciones informáticas tradicionales de procesamiento de datos no son suficientes para tratar con ellos.

En el big data se encuentra almacenada la gran parte de la información de internet por lo tanto la cantidad de datos es inmensa y cubre casi todos los ámbitos, teniendo así información sobre todo lo que queramos.

Aunque el tamaño utilizado para determinar si un conjunto de datos determinado se considera Big Data no está firmemente definido y sigue cambiando con el tiempo, la mayoría de los analistas y profesionales actualmente se refieren a conjuntos de datos que van desde 30-50 Terabytes a varios Petabytes.

Las dificultades más habituales relacionadas a la gestión de estas grandes cantidades de datos se encuentran en la recolección, almacenamiento, búsqueda, compartición, análisis, y visualización de todos estos datos.

La naturaleza compleja del Big Data se debe principalmente a la naturaleza no estructurada de gran parte de los datos generados por las tecnologías modernas, como los web logs, las búsquedas en Internet, las redes sociales, etc. En la mayoría de los casos, con el fin de utilizar eficazmente el Big Data, deben combinarse estos datos no estructurados con datos estructurados de una aplicación comercial más convencional (normalmente de una base de datos relacional).

El volumen de los datos masivos crece constantemente superando la capacidad del software para capturar, administrar y procesar esos datos.

El concepto de Big Data también engloba las infraestructuras, tecnologías y servicios que han sido creados para poder gestionar esta gran cantidad de información.

2 Pasos para el tratamiento del big data

Existen diversos pasos

2.1 Captura

Los datos se generan directa e indirectamente segundo tras segundo. Clasificamos la procedencia de los datos según las siguientes categorías:

Generados por las propias personas: El hecho de enviar correos electrónicos o mensajes por WhatsApp y más cosas que hacemos a diario crean nuevos datos que pueden ser analizados por el big data.

Obtenidas a partir de transacciones: La facturación, las llamadas telefónicas, las conexiones torres de telefonía, los accesos a wifis públicas o el pago con tarjetas de crédito generan información que tratada puede ser datos relevantes.

Marketing electrónico y web: Se generan una gran cantidad de datos cuando se navega por internet ya que todos los movimientos quedan registrados.

Obtenidos a partir de las interacciones máquina a máquina: Son datos obtenidos a partir de la recogida de métricas obtenidas desde dispositivos (medidores, sensores de temperatura, de luz, de altura, de presión, de sonido...)

Datos biométricos recolectados: En general provienen de servicios de seguridad, defensa y servicios de inteligencia.

2.2 Transformación

Una vez encontradas las fuentes de los datos necesarios. El siguiente objetivo es hacer que los datos se recojan en un mismo lugar y darles un formato adecuado.

Aquí entran en juego las plataformas extraer, transformar y cargar (ETL). Su propósito es extraer los datos de las diferentes fuentes, para después hacer transformaciones y finalmente cargar los datos en la base de datos.

2.3 Almacenamiento NoSQL

El término NoSQL se refiere a Not Only SQL (no solo SQL) y son sistemas de almacenamiento que no cumplen con el esquema entidad-relación. Proveen un sistema de almacenamiento mucho más flexible y concurrente y permiten manipular grandes cantidades de información de manera mucho más rápida que las bases de datos relacionales.

2.4 Análisis de los datos

Teniendo los datos necesarios almacenados, nos daremos cuenta que necesitaremos diferentes técnicas de análisis de datos como las siguientes:

Asociación: permite encontrar relaciones entre diferentes variables.

Minería de datos: tiene como objetivo encontrar comportamientos predictivos.

TITULO

Agrupación: el análisis de clústeres es un tipo de minería de datos que divide grandes grupos de individuos en grupos más pequeños de los cuales no conocíamos su parecido antes del análisis.

Análisis de texto: gran parte de los datos generados por las personas son textos,. Esta metodología permite extraer información de estos datos y así modelar temas y asuntos o predecir palabras.

2.5 Visualización de los datos

Consiste en la representación mediante un interfaz grafica de todos los datos contenidos dentro del big data.

3 Aplicaciones

Su principal utilidad es la comercial. Lo que hace que Big Data sea tan útil para muchas empresas es el hecho de que proporciona respuestas a muchas preguntas que las empresas ni siquiera sabían que tenían. Con una cantidad tan grande de información, los datos pueden ser moldeados o probados de cualquier manera que la empresa considere adecuada. Al hacerlo, las organizaciones son capaces de identificar los problemas de una forma más comprensible.

La recopilación de grandes cantidades de datos y la búsqueda de tendencias dentro de los datos permiten que las empresas se muevan mucho más rápidamente, sin problemas y de manera eficiente.

Las principales ventajas que genera el análisis del big data son:

Reducción de costes al identificar formas más eficaces de organizar el negocio.

Toma de decisiones más acertada y rápida

Nuevos productos y servicios debido a la capacidad de medir las necesidades de los clientes.

4 Bibliografía

<https://es.wikipedia.org/wiki/Macrodatos>

<https://www.powerdata.es/big-data>

<https://blog.mdcloud.es/que-es-big-data-y-para-que-sirve/>