Ingeniería de Servidores
Big Data y TPCx-HS

Resumen

En este texto puedes incluir un resumen del documento. Este informa al lector sobre el contenido del texto, indicando el objetivo del mismo y qué se puede aprender de él.

Índice

1. Introducción a Big Data

Desde hace miles de siglos el ser humano ha investigado la manera de almacenar y recopilar información. Durante muchos siglos la escritura y la pintura eran los únicos mecanismos existentes. Posteriormente surgió la fotografía, los discos de vinilo... Sin embargo, poca información seguía ocupando mucho volumen físico. Gracias a los avances tecnológicos de las últimos décadas, hoy en día disponemos dispositivos electrónicos para el almacenamiento de datos binarios. Además, la evolución de estos dispositivos ha sido frenética. IBM comercializó el primer disco duro en 1956. Este constaba solamente de 5 Mega Bytes de capacidad [2] mientras que actualmente podemos utilizar discos duros con más de 1 Tera Byte.

La evolución en la capacidad de cómputo y procesamiento de los computadores también ha sido exponencial. El primer ordenador comercial se presentó en ....(Referencia). Este computador realizaba n cuentas por segundo. Actualmente hablamos de GHz cuando comparamos la velocidad del procesador de un ordenador. Es más, se han creado benchmarks.... (medio párrafo diciendo para que sirven y que son muy costosos computacionalmente).

Sin embargo, a pesar de la evolución de los computadores en todas sus facetas, la cantidad de datos e información a procesar y almacenar crece incluso a mayor velocidad. El mayor flujo de datos se produce gracias a Internet. A pesar de ser relativamente joven, se hizo público en 1993, actualmente existen más de mil millones de páginas webs [3]. Además, multitud de dispositivos electrónicos se conectan e interactúan con Internet (lo que se denomina Internet de las cosas [4]). Los usuarios de estos dispositivos utilizan aplicaciones web y redes sociales, publicando textos y archivos multimedia. Todo este cúmulo de tecnologías y actividades ha dado lugar a que hoy en día haya más de 10 Zeta Bytes de información almacenados (1 ZB = 10^{12} GB) (ver la Figura 1).

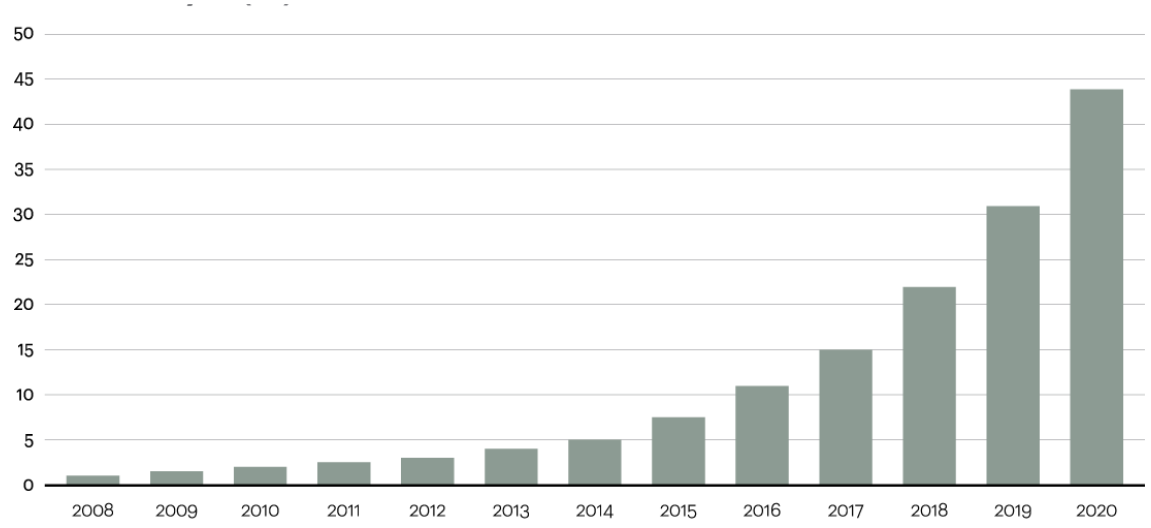

Figura 1: Evolución histórica del número de Zeta Bytes de información almacenados [1].

Esta cantidad de datos contiene conocimiento. Por ejemplo, todas las empresas almacenan información sobre sus clientes, la actividad y transacciones realizadas. Esta información necesita ser analizada en tiempo real para poder actuar en consecuencia. Aquella empresa que mejor conozca el mercado y actúa rápidamente obtendrá mejores resultados. Por otro lado, toda nuestra información médica se encuentra almacenada en el sistema de salud correspondiente. Dependemos de esta información para tomar decisiones sobre los tratamientos médicos a seguir.

En resumen, se puede decir que hoy en día el mundo gira alrededor de los datos. Todos estos datos contienen información y necesita ser procesados.

- **Volumen.** El tamaño de la información es cada vez mayor, por ejemplo, facebook procesa cada día 500 TB nuevos de información.
- **Velocidad.** Los datos deben ser procesados rápidamente. Normalmente, el procesamiento de los datos debe ser incluso continuo.
- **Variedad.** Los datos a tratar provienen de una gran variedad de fuentes. Por tanto, las herramientas Big Data deben permitir procesar a la vez datos de diferentes características y tamaños.

2. Map Reduce

3. Spark y Flink

4. Benchmarks: TPCx-HS

5. Referencias

Referencias

[1] Hugo Evans y et. al. *Big Data and the Creative Destruction of Today's Business Models*. URL: http://www.atkearney.es/paper/-/asset_publisher/dVxv4Hz2h8bS/content/big-data-and-the-creative-destruction-of-today-s-business-models/10192.

[2] Rex Farrance. «Timeline: 50 Years of Hard Drives». En: *PCWorld* (2016). URL: http://www.pcworld.com/article/127105/article.html.

[3] Internet live stats. *Total number of Websites*. URL: http://www.internetlivestats.com/total-number-of-websites/.

[4] Mario Tascón y Arantza Coullaut. *Big data y el Internet de las cosas*. Catarata, 2016.