

Lección 1.1 Contexto donde surge Big Data:

En los últimos años, el avance de las TIC (Tecnologías de la Información y la Comunicación) ha dado lugar a novedosos dispositivos, o mejoras de los existentes, y aplicaciones que generan, almacenan y hacen un uso intensivo de los datos.

Algunos ejemplos de tales dispositivos son:

1. Dispositivos:

- Smartphones y tablets
 - Conexión a internet de alta velocidad y capacidades que cada vez se acercan más a las de un PC.
 - Además, suelen incluir tecnología GPS para generar información acerca de la localización.
- Sensores
 - Incorporados en multitud de dispositivos como: Medidores inteligentes de electricidad (Smart Meters), dispositivos médicos, vehículos, carreteras, edificios públicos, electrodomésticos...
- Discos Duros
 - Más grandes (de hasta 4 TB = 4000 GB), más rápidos (discos SSD con velocidad de transferencia de 500 MB por segundo) y más asequibles (0,04 € por GB)
- Dispositivos de identificación por radiofrecuencia (RFID)
 - En la actualidad (2012) existen más de **12 millones** de tags RFID pasivas (tarjetas de crédito, documentos de identificación y acceso personales, automatización del proceso de venta (ej. tiendas Decathlon), seguimiento de productos...)

Junto a estos nuevos dispositivos y mejoras de los existentes, aparecen aplicaciones como:

2. Aplicaciones:

- Redes sociales
 - Facebook, Twitter, LinkedIn, YouTube, Google+, Instagram...

- Sus usuarios realizan millones de publicaciones al día (texto, localización geográfica, fotografía, video...)
- Open Data
 - Administraciones y organizaciones, generalmente públicas, ponen a disposición de cualquier ciudadano datos históricos y actuales de diversa índole: datos meteorológicos, socioeconómicos, energéticos, servicios públicos...
- Smart Cities
 - Conjunto de sistemas interconectados para impulsar el crecimiento económico sostenible de la ciudad y la prosperidad de sus ciudadanos.
 - Hacen uso de las TIC más avanzadas: WIFI, dispositivos móviles, sensores, redes sociales, IoT....
- Internet of the Things (IoT) (En castellano, Internet de la cosas)
 - Redes de dispositivos interconectados que se comunican de forma autónoma.
 - Ej de tales dispositivos: Smartphones, electrodomésticos, dispositivos de sensorización...

Además de las anteriores, han surgido otras aplicaciones como:

- Aplicaciones y almacenamiento en la nube, web semántica, sistemas de recomendación,...

Por otro lado, no menos importante es el avance y mejora de las aplicaciones ya existentes, las cuales cada vez generan y manejan un mayor volumen de datos:

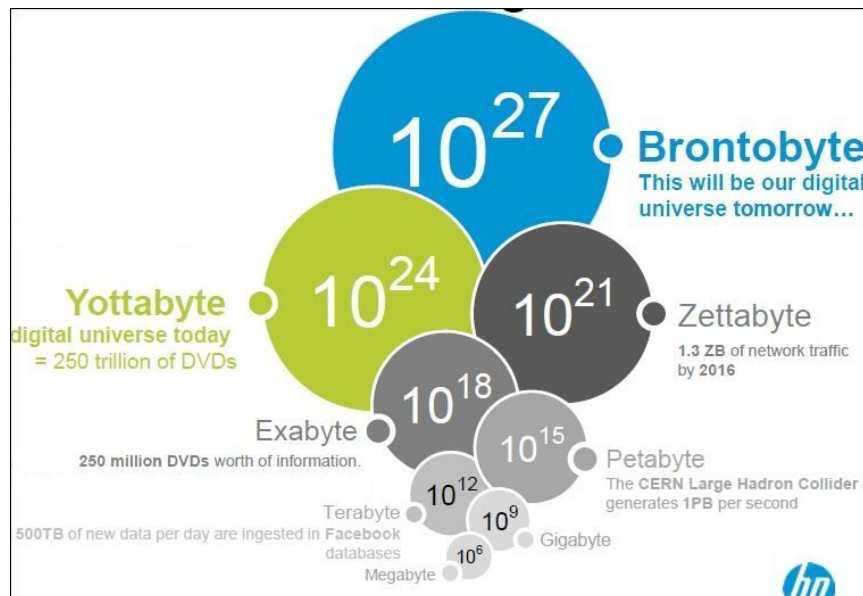
- Sistemas de información en la empresa, sistemas transaccionales, ofimática (Word, Excel, Power Point), correo electrónico, registros de eventos (logs)...

Por la cantidad de datos que generan, todas estas aplicaciones han dado lugar al término de Big Data, pero profundicemos un poco más en el volumen generado.

3. Volumen generado:

- Como resultado del escenario anterior, a diario se genera la increíble cifra de 2,5 exabytes de datos el mundo. Para hacernos una idea de tal volumen (1 exabyte = 1 billón de gigabytes, que a su vez equivale a unos 250 millones de DVD's de 4,7 Gb repletos de información)
- Tal volumen se duplica cada 2 meses

- Algunos ejemplos concretos del volumen de datos generado y manejado en la actualidad:
 - El acelerador de partículas del CERN en Ginebra (Suiza) genera un petabyte por segundo
 - Cada día se almacenan 500 terabytes en las bases de datos de Facebook
 - Los más de cien mil conjuntos de datos abiertos (Open Data), solo en la página del gobierno de EEUU
 - Los 100 terabytes de datos que, de media, tiene almacenados un empresa de medio o gran tamaño estadounidense



4: Valor de la información:

- De esta forma, en los últimos años, la información disponible en la red ha pasado de ser escasa a ser realmente abundante.
- Los datos se han convertido en la nueva materia prima de los negocios: El beneficio que pueden reportar, los hacen casi más importantes que el capital y la mano de obra

- El aprovechamiento adecuado de estos datos, puede dar lugar a importantes beneficios como...
 - Aumentar la ventaja competitiva de una organización frente a otras.
 - Optimizar la generación y el consumo de energía para hacerlos más sostenibles
 - A nivel macroeconómico, los expertos vaticinan un incremento del 1,7 % en PIB de EEUU en 2020, gracias al aprovechamiento adecuado de estos datos.
- Sin embargo, al mismo tiempo el procesamiento y análisis de estos datos genera importantes quebraderos de cabeza...
 - Tamaño, variedad de formatos y estructura, calidad, velocidad de generación, integración...dan lugar a problemas que más adelante estudiaremos en este módulo.