

BIG DATA Y SU RELACIÓN CON LAS BASES DE DATOS

Trabajo realizado por:

Hugo García
Óscar Sales
Juan Carlos Fosela
Salah Rachid



ÍNDICE

- 1. ¿Qué es el Big Data?**
- 2. Bases de datos tradicionales VS Big Data**
- 3. Aplicaciones en el mundo real**
- 4. Herramientas y arquitecturas en Big Data**
- 5. Bibliografía**

1. ¿Qué es el big data?

Big data se refiere al manejo de grandes volúmenes de datos que no pueden ser tratados de forma convencional.

5V:

- **Volumen:** Cantidad enorme de datos.
- **Velocidad:** Se generan y procesan muy rápido
- **Variedad:** Datos de diferentes tipos (texto, imágenes, videos...)
- **Veracidad:** Datos que pueden ser poco fiables.
- **Valor:** Lo importante es extraer información útil

2. Bases de datos tradicionales VS Big Data

Bases de datos tradicionales (SQL)

- **Datos en tablas, filas y columnas**
- **Limitada para grandes volúmenes de datos en tiempo real**
- **Escalabilidad vertical**
- **MySQL, PostgreSQL, Oracle**
- **Servidor potente y licencias**
- **Alto control y consistencia de datos**

Big Data

- **Datos no estructurados o semiestructurados**
- **Procesamiento masivo y en paralelo**
- **Escalabilidad horizontal**
- **MongoDB, Cassandra, Hadoop**
- **Varios servidores, escalabilidad económica**
- **Flexible, pero la consistencia puede ser eventual**

3. Cómo se relacionan y Aplicaciones en el mundo real

Relación directa:

- Las bases de datos siguen siendo la base donde se almacenan los datos
- Aunque Big Data requiere de nuevas formas de almacenar y procesar datos

Redes Sociales (Facebook, Twitter, Instagram)

- **Análisis de preferencias:** Las plataformas de redes sociales utilizan Big Data para analizar los comportamientos de los usuarios y ofrecer anuncios personalizados.

Salud (Medicina personalizada y predicción de enfermedades)

- **Análisis predictivo:** Big Data se utiliza para analizar grandes cantidades de datos médicos (historias clínicas, genética, imágenes médicas) y predecir enfermedades antes de que aparezcan, o personalizar tratamientos basados en los datos específicos de un paciente



Comercio y Marketing

- **Recomendaciones personalizadas:** Empresas como Amazon y Netflix usan Big Data para ofrecer recomendaciones personalizadas basadas en tu comportamiento de compra o visualización

Transporte y Movilidad (Uber, Lyft, etc.)

- **Optimización de rutas:** Las plataformas de transporte como Uber utilizan Big Data para analizar el tráfico en tiempo real y determinar la ruta más rápida.
- **Gestión de demanda:** Utilizan análisis de datos para predecir la demanda de viajes en ciertas áreas y en determinados momentos del día.

Finanzas (Detección de fraude)

- **Ánalisis de patrones:** Los bancos y las compañías de tarjetas de crédito usan Big Data para detectar patrones anormales en las transacciones, lo que les ayuda a identificar posibles fraudes.
- **Ejemplo:** *"Si una tarjeta de crédito es utilizada en un lugar geográficamente distante en poco tiempo, el sistema puede marcarlo como un posible fraude."*



4. Herramientas y arquitecturas en Big Data

- **NoSQL: datos no estructurados o semi-estructurados.**
Ejemplos:
 - **MongoDB** → Documentos JSON.
 - **Cassandra** → Datos distribuidos a gran escala.
 - **Redis** → Almacenamiento en memoria para alta velocidad.
- **NewSQL: combinan la estructura del SQL con la escalabilidad del NoSQL (Google Spanner, CockroachDB).**



- Almacenamiento distribuido

En lugar de guardar todos los datos en un solo servidor, se distribuyen entre muchos equipos conectados en red, formando un cluster que maneja volúmenes de información imposibles para un solo equipo. HDFS (Hadoop Distributed File System), divide los datos en bloques y los reparte entre nodos, garantizando redundancia y tolerancia a fallos.

- Procesamiento paralelo

Tecnologías como MapReduce o Apache Spark permiten dividir una tarea en muchas partes más pequeñas, procesarlas simultáneamente en distintos nodos y luego unir los resultados.

Esto reduce enormemente los tiempos de análisis y permite trabajar con terabytes o petabytes de datos.

- Lagos de datos

Un Data Lake es un gran repositorio donde se almacenan datos sin procesar, que ayuda a conservar toda la información para procesarla más adelante. Ejemplos son Amazon S3 o Azure Data Lake, que sirven como base para proyectos de inteligencia artificial, machine learning o analítica avanzada.



5. Bibliografía

<https://www.oracle.com/es/big-data/what-is-big-data/>

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

<https://www.kyoceradocumentsolutions.es/es/smarter-workspaces/business-challenges/the-cloud/por-que-las-bases-de-datos-relacionales-no-manejan-bien-el-big-data>