

Ejercicio 1. Detecta las 5V's en un caso real

Analiza un caso de uso propuesto (por ejemplo: redes sociales, salud digital, transporte urbano o e-commerce) e identifica cómo se manifiestan las 5V's de Big Data. Argumenta con ejemplos y propone tecnologías del ecosistema Big Data que podrían aplicarse.

Caso: E-commerce

Volumen

- Millones de registros de compras, visitas, clics y búsquedas cada día.
- **Ejemplo:** historial de compras de millones de clientes.
- **Tecnologías:** HDFS para almacenamiento distribuido, MongoDB para datos no estructurados.

Velocidad

- Datos generados en tiempo real: clics, actualizaciones de stock, pagos.
- **Ejemplo:** detección de fraude en transacciones en milisegundos.
- **Tecnologías:** Apache Kafka (procesamiento en streaming), Spark Streaming.

Variedad

- Diversos formatos de datos: imágenes de productos, reseñas en texto, métricas de sensores en logística.
- **Ejemplo:** comentarios de clientes (texto), fotos de productos (imagen), ubicación de pedidos (IoT).
- **Tecnologías:** bases NoSQL (Cassandra, MongoDB), herramientas de NLP para texto.

Veracidad

- Garantizar que los datos sean confiables y no duplicados o falsos.
- **Ejemplo:** reseñas falsas de productos o transacciones fraudulentas.

- **Tecnologías:** algoritmos de Machine Learning para detección de fraude, validación de datos.

Valor

- Transformar datos en conocimiento para aumentar ventas y fidelización.
- **Ejemplo:** recomendaciones personalizadas de productos (sistemas de recomendación).
- **Tecnologías:** Spark MLlib, TensorFlow, algoritmos de recomendación.

V	Ejemplo en E-commerce	Tecnología
Volumen	Millones de compras y clics diarios	HDFS, MongoDB
Velocidad	Detección de fraude en tiempo real	Apache Kafka, Spark Streaming
Variedad	Texto (reseñas), imágenes (productos), datos IoT	MongoDB, Cassandra, NLP, HDFS
Veracidad	Filtrar reseñas falsas, datos duplicados	ML para validación, algoritmos anti-fraude
Valor	Recomendaciones personalizadas, predicción de ventas	Spark MLlib, TensorFlow