

AMAZON

Caso de estudio 2 - Gigante del e-commerce

1) Las 5 V's del Big Data

- **Volumen:** Maneja gran cantidad de datos generados por las compras, búsquedas, clics, carritos abandonados, reseñas y datos de sensores en centros de distribución y camiones.
- **Velocidad:**
Creemos que puede usar ambos sistemas, dependiendo de la situación:
 - **Tiempo real (streaming)** para mostrar recomendaciones instantáneas y actualizar el stock.
 - **Por lotes (batch)** para análisis históricos y planificación de la demanda.
La velocidad es clave para dar la mejor experiencia de compra y optimizar la logística.
- **Variedad:**
 - **Estructurados:** ventas, inventario, precios.
 - **Semiestructurados:** registros de navegación y eventos.
 - **No estructurados:** texto de reseñas, imágenes.
- **Veracidad:** Problemas como datos incompletos, reseñas falsas, errores en información de vendedores y unificación de datos de distintas fuentes.
- **Valor:** Incrementar ventas con recomendaciones, reducir costos de logística optimizando inventario, mejorar tiempos de entrega y satisfacción del cliente.

2) Almacenamiento

- **Data Lake:** Ya que es un almacenamiento escalable y distribuido en la nube que permite guardar datos de distintos formatos.
- **Desafíos:** Mantener los costos controlados aplicando almacenamiento por niveles y asegurar que la infraestructura pueda crecer sin afectar el rendimiento.

3) Procesamiento y Análisis

- **Tipo de procesamiento:**
 - **Batch** para análisis de grandes volúmenes de datos históricos y entrenamiento de modelos.
 - **Streaming** para procesar datos de clics, compras y stock en el momento.
- **Herramientas:**
 - **SQL** para consultas sobre datos estructurados.
 - **Python** para análisis y manejo de datos.
 - **Machine Learning** para personalizar las recomendaciones y predecir demanda.

4) Gobernanza y Seguridad

- **Datos sensibles:** Información **personal** de clientes (nombre, dirección, historial de compras), datos de pago y comportamiento de navegación.
- **Medidas:**
 - Control de acceso para que solo personal autorizado acceda a datos sensibles.
 - **Encriptación** de datos tanto al **almacenarlos** como cuando se **transmiten**.
 - Cumplir con las leyes de privacidad y protección de datos.