

# **CASO 2: GIGANTE DEL E-COMMERCE – AMAZON**

Grupo: Barreiro, Canal, Halac Ayuso, Magliano

Caso de uso:

Personalización de la experiencia de compra y optimización de la cadena de suministro.

## **Análisis de las 5 V's del Big Data en Amazon**

### **Volumen:**

- Amazon gestiona petabytes de datos provenientes de millones de transacciones diarias, clics, reseñas, historiales de navegación y búsquedas.
- Cada interacción del usuario genera nuevos datos.

### **Velocidad:**

- Los datos se procesan tanto en tiempo real (recomendaciones, inventario) como por lotes (análisis históricos).
- La velocidad es clave para ofrecer recomendaciones instantáneas y optimizar envíos en tiempo real.

### **Variedad:**

- Maneja datos:
  - Estructurados: transacciones, precios, stock.
  - No estructurados: reseñas de productos, imágenes, opiniones.
  - Semi-estructurados: logs de navegación, JSON de APIs.

### **Veracidad:**

- Desafíos:
  - Datos duplicados, errores en reseñas falsas, bots, fraudes.
  - Requiere limpieza y validación constante.
- Soluciones: Machine Learning y modelos de reputación para detectar anomalías.

### **Valor:**

- Aumento de ventas, fidelización del cliente, reducción de costos logísticos.
- Optimización de stock → menos demoras, menor costo de envío.
- Mejora experiencia de usuario → más conversiones.

## **Almacenamiento**

### **¿Dónde se almacenan los datos?**

- Amazon usa su propia infraestructura en la nube: **Amazon Web Services (AWS)**.
- Usa **Data Lakes**, bases NoSQL (DynamoDB), S3, y sistemas distribuidos.

### **Desafíos:**

- **Escalabilidad:** manejar datos globales a gran escala.
- **Costo:** elegir qué datos se almacenan en frío (menos acceso) y cuáles en caliente (uso frecuente) para optimizar gastos.

## **Procesamiento y Análisis**

### **Tipo de procesamiento:**

- Streaming: para recomendaciones y gestión de stock en tiempo real.
- Batch (por lotes): análisis histórico, predicción de demanda.

### **Herramientas posibles:**

- Python/Pandas/Spark para procesamiento.
- SQL para datos estructurados.
- Machine Learning/IA para segmentación, predicción y detección de patrones.
- Herramientas BI como Amazon QuickSight para visualizar insights.

## **Gobernanza y Seguridad**

### **Datos sensibles:**

- Información personal: nombre, dirección, tarjetas de crédito, historial de compras.
- Hábitos de consumo, comportamientos.

### **Desafíos:**

- Privacidad: cumplimiento de normativas como GDPR, encriptación de datos sensibles.
- Seguridad: control de accesos, autenticación, detección de fraudes.

