

Caso de Estudio 1: Empresa de Streaming (Netflix)

Caso de Uso: Optimización de la experiencia del usuario y creación de contenido original.

Descripción Ampliada: Netflix, como líder en streaming, no solo compite con otras plataformas, sino también con el tiempo libre de sus usuarios. Para mantenerlos enganchados, utiliza Big Data para todo:

Motor de Recomendaciones: Recopila datos de miles de millones de interacciones, incluyendo qué ven los usuarios, qué buscan, a qué hora, qué dispositivos usan, cuándo pausan, rebobinan o abandonan un título. Con esta información, el 80% del contenido que un usuario ve proviene de su motor de recomendaciones.

Creación de Contenido Original: Antes de producir una serie, Netflix analiza qué actores, directores y géneros son populares en diferentes regiones. Los datos les indican qué historias tienen la mayor probabilidad de ser un éxito, lo que justifica la enorme inversión en producción.

Optimización del Producto: Mediante pruebas A/B a gran escala, Netflix experimenta con diferentes diseños de interfaz y miniaturas de series para ver qué genera más clics y tiempo de visualización.

1. Las 5 V's del Big Data

- **Volumen:** Netflix maneja un volumen enorme de datos. Esto incluye el historial de visualización de millones de usuarios, sus interacciones (pausas, rebobinados, búsquedas), calificaciones de contenido, datos demográficos, información sobre los dispositivos que utilizan y los resultados de las pruebas A/B de su interfaz. La escala de estos datos se mide en petabytes.
- **Velocidad:** Parte del análisis es en tiempo real (recomendaciones personalizadas), y otra parte en lotes / batch (producción de contenido original). La latencia baja es clave para la personalización instantánea. La velocidad es un factor clave para esta plataforma.
- **Variedad:** Netflix maneja una gran variedad de tipos de datos:
 - **Datos estructurados:** como perfiles de usuario, historial de pagos, géneros de contenido
 - **Datos semiestructurados:** como logs de actividad (clics, tiempos de reproducción, búsquedas).
 - **Datos no estructurados:** como transcripciones de audio (subtítulos), información de imágenes (miniaturas)
- **Veracidad:** Es fundamental que los datos que Netflix utiliza sean fiables y precisos. Si las interacciones de los usuarios o las calificaciones de contenido no son veraces, las recomendaciones y las decisiones de optimización del producto podrían ser erróneas, afectando la experiencia del usuario y el éxito del negocio.
- **Valor:** Optimización de la experiencia del usuario y la creación de contenido original. Permitiendo mantener a los usuarios "enganchados", aumentar el tiempo de

visualización e incrementar la lealtad del cliente y las suscripciones al mostrarles lo que realmente les interesa.

- El **80% del contenido que un usuario ve proviene de su motor de recomendaciones**, lo que demuestra el valor directo en la retención y satisfacción del cliente
- En la creación de contenido original, el análisis de datos les permite identificar películas / series con mayor probabilidad de éxito.

2. Almacenamiento de los Datos

- **Tecnología recomendada:**
 - Sistemas de archivos distribuidos: como HDFS, para almacenar grandes volúmenes de datos.
- **Desafíos:**
 - **Escalabilidad:** deben guardar cada interacción de millones de usuarios.
 - **Costo:** deben diferenciar entre datos de acceso frecuente y datos de archivo, para optimizar costos.

3. Procesamiento y Análisis

- **Tipo de procesamiento necesario:** Se necesita una combinación de procesamiento en tiempo real (streaming) para las interacciones dinámicas del usuario y procesamiento por lotes (batch) para análisis históricos y estratégicos.
- **Herramientas de análisis:**
 - ❖ **Apache Spark:** procesamiento en memoria → análisis rápido y flexible.
 - ❖ **Python:** para análisis exploratorio.
 - ❖ **SQL:** para bases de datos estructuradas.
 - ❖ **Machine Learning:** para predicciones y segmentación.

4. Gobernanza y Seguridad

Datos sensibles o personales: Netflix maneja una gran cantidad de datos sensibles y personales de sus clientes, incluyendo: Nombre, email, métodos de pago, historial de visualización, geolocalización, comportamiento de navegación.

Desafíos de seguridad y privacidad: Los principales desafíos para proteger esta información incluye: Control de calidad, actualización y ciclo de vida de los datos, roles y permisos claros para acceso, autenticación y autorización. Encriptación en tr y cumplimiento con normativas.