

Sesión 3

Big Data Adoption and Planning Considerations

Lectura de Chapter 3: Big Data
Adoption and Planning
Considerations

Luis Fernando Castellanos Guarín
2025



UNIVERSIDAD
CENTRAL



1. Prerrequisitos Organizacionales

- Necesidad de marcos de gestión de datos y gobernanza de Big Data.
- Procesos sólidos y conjuntos de habilidades adecuados.
- Evaluación de la calidad de los datos.
- Planificación de la longitud de la vida útil de los datos.



Ejemplo Real:

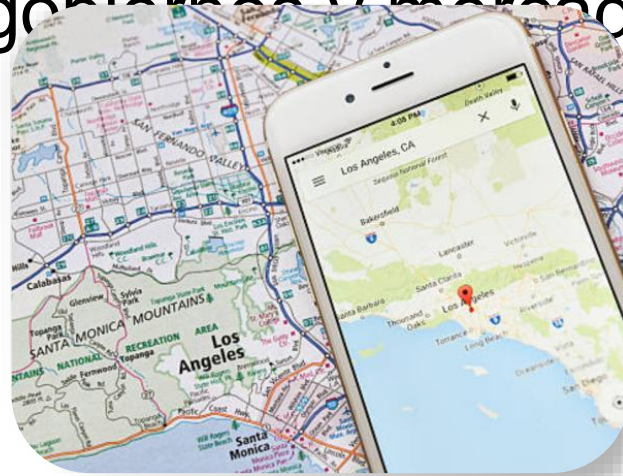
Netflix: Implementó un marco de gobernanza de datos para gestionar su vasto conjunto de datos de usuarios, mejorando la calidad de sus recomendaciones.

2. Adquisición de Datos

La adquisición de soluciones de Big Data **puede ser económica**.

Los datos externos pueden ser **valiosos** pero costosos.

Fuentes de datos incluyen gobiernos y mercados de datos comerciales.



Ejemplo Real:

Google: *Utiliza datos externos como datos de tráfico en tiempo real para mejorar Google Maps.*

3. Privacidad

El análisis de datos puede revelar información confidencial.

Necesidad de técnicas de etiquetado y anonimización de datos.

Cumplimiento de regulaciones de privacidad (nacionales e internacionales)



Ejemplo Real:

Facebook: Enfrentó problemas de privacidad al revelar datos de usuarios a terceros sin consentimiento.

<https://www.youtube.com/watch?v=H->

4. Seguridad

Asegurar redes y repositorios de datos.

Establecer niveles de acceso a datos.

Las bases de datos NoSQL a menudo carecen de mecanismos de seguridad robustos.



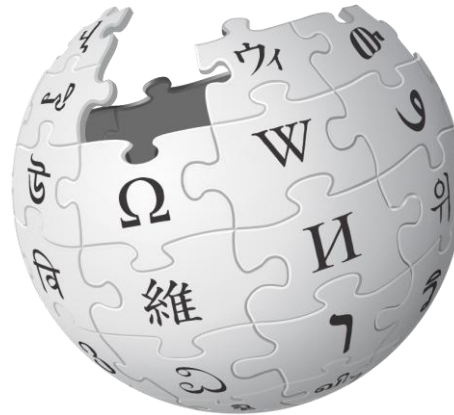
Ejemplo Real:

Equifax: Sufrió una brecha de seguridad que expuso datos personales de millones de usuarios debido a fallas en la seguridad de datos.

<https://www.xataka.com/seguridad/hackean-equifax-una-de-las-mayores-agencias-de-informes-crediticios-afectando-a-143-millones-de->

5. Procedencia

- Información sobre la fuente y procesamiento de los datos.
- Importante para auditorías y autenticidad de datos.
- Capturar metadatos en cada etapa del ciclo de vida de los datos.

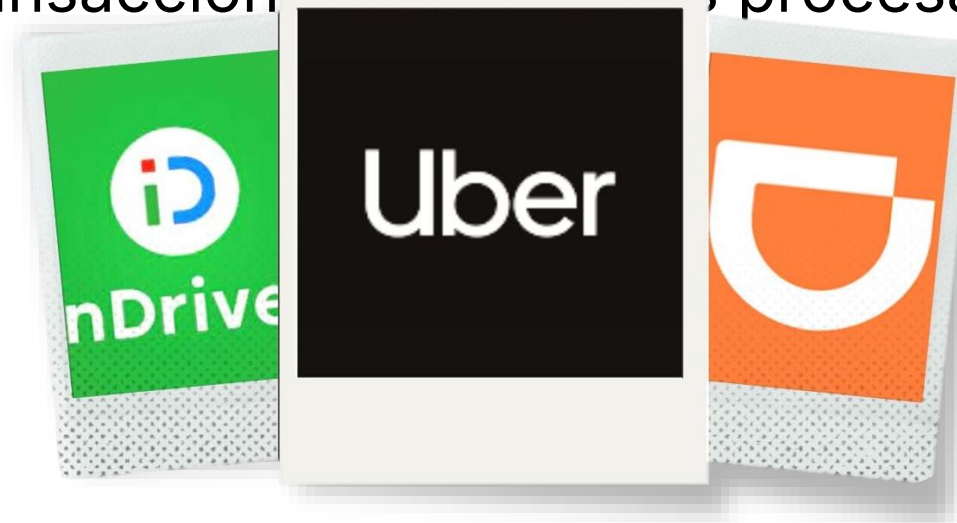


Ejemplo Real:

Wikipedia: Utiliza metadatos para rastrear la procedencia de las ediciones y asegurar la autenticidad del contenido.

6. Soporte en Tiempo Real Limitado

- Muchas soluciones de Big Data son orientadas a lotes.
- Nueva generación de herramientas con soporte para análisis de datos en tiempo real.
- Combinar datos transaccionales con datos procesados por lotes.

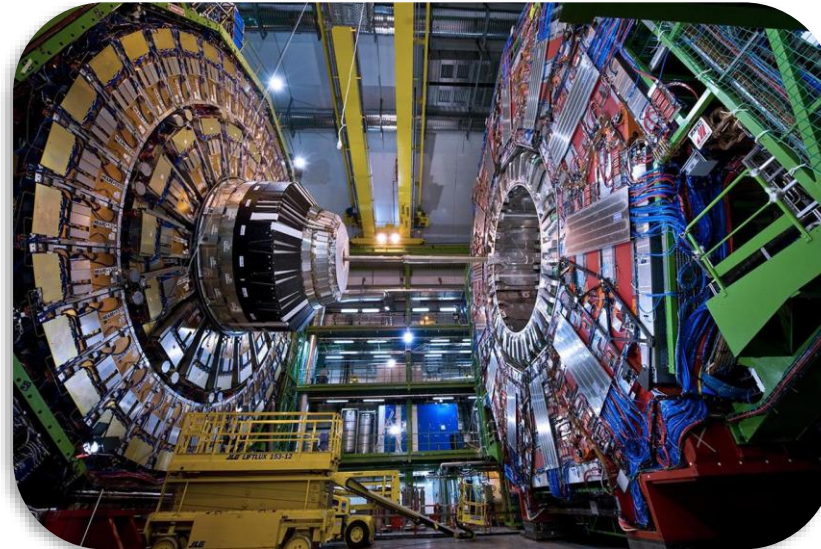


Ejemplo Real:

Uber, didi, inDrive...etc: Utilizan análisis en tiempo real para gestionar la demanda y oferta de viajes, ajustando precios dinámicamente.

7. Desafíos de Rendimiento

- Grandes volúmenes de datos pueden llevar a tiempos de consulta largos.
- Desafíos de ancho de banda de red.
- Tiempo de transferencia de datos puede exceder el tiempo de procesamiento.



Ejemplo Real:

CERN: Procesa grandes volúmenes de datos del Gran Colisionador de Hadrones, enfrentando desafíos de rendimiento y ancho de banda.

8. Requisitos de Gobernanza

- Necesidad de un marco de gobernanza para regular y estandarizar el entorno de Big Data.
- Políticas para etiquetado de datos, privacidad, archivado y limpieza de datos.



Ejemplo Real:

IBM: Implementó un marco de gobernanza de datos para asegurar el cumplimiento normativo y la calidad de los datos en sus soluciones de Big Data.

8. Metodología Distinta

- Controlar el flujo de datos en las soluciones de Big Data.
- Establecer bucles de retroalimentación para refinar los datos procesados.
- Enfoque iterativo para mejorar la precisión de los resultados.



Ejemplo Real:

Amazon: Utiliza bucles de retroalimentación para mejorar continuamente sus algoritmos de recomendación de productos.

9. Nubes/clouds

- Las nubes proporcionan entornos remotos para almacenamiento y procesamiento a gran escala.
- Justificaciones comunes incluyen recursos hardware insuficientes y proyectos c



Ejemplo Real:

Microsoft Azure: Ofrece servicios en la nube para almacenamiento y procesamiento de Big Data, utilizados por empresas para escalar sus soluciones.

10. Ciclo de Vida de Análisis de Big Data

- Desde la identificación del caso de negocio hasta la utilización de los resultados del análisis.
- Etapas intermedias incluyen identificación, selección, filtrado, extracción, limpieza y agregación de datos.



Ejemplo Real:

Spotify: Utiliza el ciclo de vida de análisis de Big Data para mejorar sus recomendaciones de música basadas en el comportamiento del usuario.

Conclusión

La adopción de Big Data requiere una planificación cuidadosa y consideración de múltiples factores.

Las organizaciones deben estar preparadas para abordar desafíos únicos y aprovechar las oportunidades que ofrece Big Data.

Ejercicios de la sesión

Extracción, transformación y carga (ETL)

https://www.datos.gov.co/Salud-y-Protecci-n-Social/Poblaci-n-Base-de-Datos-nica-de-Afiliados-BDUA-del/tq4m-hmg2/about_data

Gracias