

# Introducción al Procesamiento de Grandes Volúmenes de Datos

Deborah Famadas Rodríguez

October 29, 2024

# Introducción

- ▶ Importancia del procesamiento de grandes volúmenes de datos.
- ▶ Definición de Big Data: Los 5 Vs.
  - ▶ Volumen
  - ▶ Velocidad
  - ▶ Variedad
  - ▶ Veracidad
  - ▶ Valor
- ▶ Ejemplos cotidianos: redes sociales, sensores IoT, plataformas de streaming.

# Estructura de los Datos

- ▶ Tipos de datos:
  - ▶ Datos estructurados: Ejemplo, tablas de bases de datos.
  - ▶ Datos semi-estructurados: Ejemplo, JSON o XML.
  - ▶ Datos no estructurados: Ejemplo, imágenes, videos, datos de redes sociales.
- ▶ Bases de datos relacionales vs NoSQL.

# Modelos de Almacenamiento

- ▶ Diferencias entre Data Lake y Data Warehouse.
- ▶ Ejemplos de uso de cada modelo.
- ▶ Procesamiento en tiempo real vs en lotes (stream vs batch processing).
- ▶ El ciclo de vida de los datos: Generación, Almacenamiento, Procesamiento, Análisis.

# Desafíos y Problemas en el Procesamiento de Datos

- ▶ Escalabilidad: Cómo manejar grandes volúmenes.
- ▶ Problemas de almacenamiento: Costos, acceso rápido a datos.

# Herramientas y Tecnologías

- ▶ Almacenamiento distribuido: HDFS y otros sistemas de archivos distribuidos.
- ▶ Modelos de procesamiento distribuido:
  - ▶ MapReduce: Descomposición de tareas.
  - ▶ Apache Hadoop vs Apache Spark.
- ▶ Bases de datos NoSQL: MongoDB, Cassandra, Neo4j.
- ▶ Cloud computing: AWS, Google Cloud, Azure.

# Técnicas de Procesamiento de Grandes Volúmenes de Datos

- ▶ Procesamiento paralelo: Dividir tareas entre múltiples máquinas.
- ▶ Stream processing vs Batch processing: ¿Cuándo usar cada uno?
- ▶ Algoritmos básicos:
  - ▶ Clasificación
  - ▶ Filtrado
  - ▶ Agregaciones

# Casos de Uso

- ▶ Plataformas como Netflix y su análisis de datos para recomendaciones personalizadas.
- ▶ Procesamiento de datos meteorológicos.
- ▶ Ideas de proyectos estudiantiles:
  - ▶ Análisis de datos de redes sociales.
  - ▶ Procesamiento de imágenes o datos de sensores.



# Conclusión

- ▶ Resumen de los conceptos clave: Big Data, herramientas, técnicas.
- ▶ Mirada al futuro: Inteligencia Artificial y Machine Learning en el procesamiento de grandes volúmenes de datos.
- ▶ Tiempo para preguntas y respuestas.