

Arquitectura Big Data mínima para una App de Movilidad Urbana

En este informe se presenta el diseño de una arquitectura Big Data mínima aplicada a una app de movilidad urbana. El objetivo es mostrar cómo se integran diferentes capas y herramientas tecnológicas para adquirir, almacenar, procesar, analizar y visualizar datos en tiempo real, optimizando la gestión de información proveniente de sensores, GPS y usuarios.

Capas de la Arquitectura

1. **Adquisición de datos**:

- Fuentes: Sensores GPS de buses, datos de tráfico, condiciones climáticas, interacciones de usuarios.
- Herramienta: Apache Kafka, utilizada para la ingesta de datos en tiempo real.

2. **Almacenamiento**:

- HDFS (Hadoop Distributed File System) para almacenar grandes volúmenes de datos históricos.
- Bases de datos NoSQL (ejemplo: Cassandra o MongoDB) para datos semiestructurados y de consulta rápida.

3. **Procesamiento**:

- Apache Spark para el procesamiento distribuido de grandes volúmenes de datos.
- Spark Streaming o Flink para el análisis en tiempo real.

4. **Análisis**:

- Spark MLlib para aplicar modelos de machine learning, como predicción de tiempos de llegada, detección de congestión o recomendación de rutas alternativas.

5. **Visualización**:

- Power BI o una aplicación móvil propia que muestre dashboards con métricas clave y recomendaciones al usuario final.

Esquema de Flujo de Datos

Usuarios & Sensores (GPS, IoT, Clima) → Kafka (ingesta en tiempo real) → HDFS/NoSQL (almacenamiento) → Spark/Flink (procesamiento) → MLlib (modelos ML) → Power BI/App (visualización)

Conclusión

Esta arquitectura Big Data mínima permite gestionar información en tiempo real proveniente de diferentes fuentes, facilitando el análisis predictivo y la toma de decisiones en movilidad urbana. El uso combinado de Kafka, HDFS, Spark y herramientas de visualización asegura un flujo eficiente y escalable de datos.