

# Diseña una arquitectura Big Data mínima

**Nombre del estudiante:** Jonathan Vásquez

## Contexto:

Comprender las capas de una arquitectura Big Data permite ver cómo se combinan herramientas para resolver problemas complejos. Esta actividad integra teoría y diseño práctico.

## Consigna:

Diseña una arquitectura básica de Big Data para una app de movilidad urbana. Debe incluir las capas de:

- Adquisición
- Almacenamiento
- Procesamiento
- Análisis
- Visualización

## Paso a Paso:

1. Piensa qué fuentes de datos tendría la app (GPS, sensores, usuarios, etc).
2. Elige tecnologías para cada capa (por ejemplo: Kafka, HDFS, Spark, Hive, Power BI).
3. Justifica tus elecciones en función de las necesidades del sistema.
4. Representa tu arquitectura en un esquema claro con nombres de herramientas.
5. Presenta tu diseño

# 1. Fuentes de Datos

Para la app de movilidad urbana, se consideran las siguientes fuentes de datos:

- **GPS de los vehículos:** Para rastreo en tiempo real y geolocalización.
- **Sensores de tráfico:** Información sobre congestión, velocidad promedio y eventos en la vía.
- **Usuarios de la app:** Solicitudes de viajes, rutas favoritas, horarios de uso.
- **Datos históricos:** Patrones de tráfico y estadísticas de movilidad.

Estas fuentes generan datos en **gran volumen**, a **alta velocidad**, y con variabilidad que requiere procesamiento y análisis eficientes.

## 2. Arquitectura Propuesta

La arquitectura Big Data mínima propuesta se compone de las siguientes capas:

### a) Capa de Adquisición de Datos

**Tecnología:** Kafka

**Justificación:** Kafka permite ingerir en tiempo real streams de GPS, sensores y eventos de usuarios, asegurando que los datos lleguen rápidamente al sistema para su procesamiento.

### b) Capa de Almacenamiento

**Tecnología:** HDFS

**Justificación:** HDFS permite almacenar grandes volúmenes de datos de forma distribuida, soportando datos no estructurados y semiestructurados provenientes de la app y sensores.

### c) Capa de Procesamiento

**Tecnología:** Spark

**Justificación:** Spark permite procesar datos en batch y en tiempo real, aplicando transformaciones, agregaciones y análisis complejos sobre grandes volúmenes de datos.

### d) Capa de Análisis

**Tecnología:** Hive

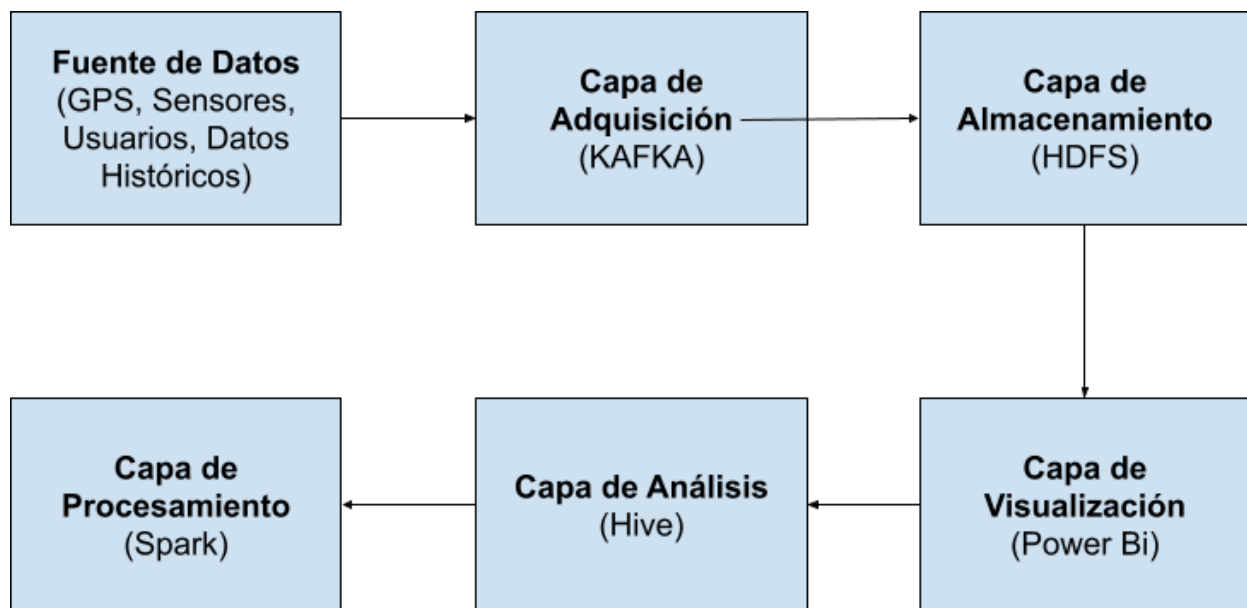
**Justificación:** Hive facilita consultas SQL sobre datos almacenados en HDFS, permitiendo generar reportes y análisis estadísticos que sirvan para optimizar rutas y tiempos de movilidad.

## e) Capa de Visualización

**Tecnología:** Power BI

**Justificación:** Power BI permite crear dashboards interactivos que muestran la información de tráfico, rutas y uso de la app en tiempo real, ayudando en la toma de decisiones operativas y estratégicas.

## 3. Esquema de la Arquitectura



## 4. Justificación General

Esta arquitectura garantiza:

- **Escalabilidad**, para manejar grandes volúmenes de datos generados por la app.
- **Procesamiento en tiempo real**, para reaccionar ante cambios de tráfico y rutas.
- **Flexibilidad analítica**, permitiendo consultas rápidas y generación de reportes con Hive.
- **Visualización clara**, a través de dashboards interactivos en Power BI, que facilitan la interpretación de los datos y la toma de decisiones.



