

Temas PGVD

Todos los proyectos usan **Hadoop** como base de almacenamiento y procesamiento distribuido.

- **HDFS** → almacenamiento de grandes volúmenes de datos.
 - **YARN** → gestión de recursos del clúster.
 - **MapReduce/Spark sobre Hadoop** → procesamiento masivo.
 - **Hive/HBase** → consultas y bases de datos sobre HDFS.
-

1. Análisis predictivo de fallos en máquinas industriales

Descripción:

Uso de datos de sensores IoT para anticipar fallos en equipos industriales.

Gráficas en tiempo real:

- Temperatura y vibración en vivo.
- Anomalías por máquina.
- Predicciones de fallo.
- Score de salud.
- Comparación entre equipos.
- Alertas en tiempo real.

<i>Industrial IoT Fault Detection Dataset (Kaggle)</i> (Kaggle)	Datos de sensores industriales con registros de fallos.	
---	---	--

2. Detección de fraudes en transacciones financieras

Descripción:

Identificación de transacciones sospechosas a gran escala.

Gráficas en tiempo real:

- Volumen de transacciones por segundo.
- Alertas activas.

- Score de fraude.
- Mapa de fraudes por región.
- Evolución de falsos positivos.

Fraudulent Transactions Data (Kaggle) (Kaggle)	Transacciones variadas, con posibilidad de múltiples atributos (monto, cliente, categoría, etc.).	
<i>Credit Card Fraud Detection</i> (Kaggle) (Kaggle)	Transacciones crediticias, etiquetadas como normales o fraudulentas. Muchos registros (~284,807) con alta desproporción (muy pocos fraudes) (Kaggle)	

3. Sistema de recomendación a gran escala

 **Descripción:**

Motor de recomendaciones usando millones de interacciones.

 **Gráficas en tiempo real:**

- Tasa de aceptación de recomendaciones.
- Productos más sugeridos.
- Usuarios activos por minuto.
- Impacto en engagement.
- Evolución de precisión.

kaggle	Datos sobre películas con valor de aceptación
kaggle	MovieLens permite construir recomendaciones colaborativas, medir precisión, etc.

4. Análisis de movilidad urbana con datos de GPS

 **Descripción:**

Optimización de tráfico y rutas urbanas con Big Data geoespacial.

 **Gráficas en tiempo real:**

- Mapas de calor de tráfico.
- Flujos de vehículos.
- Predicciones de tiempos de viaje.

- Rutas congestionadas vs. libres.
- Zonas críticas (clustering).
- Evolución histórica.

Road Traffic Dataset (Kaggle) (Kaggle)	Datos de tráfico en carreteras/junctions, volumen de vehículos minuto/hora.
---	---

5. Monitorización de redes sociales para análisis de sentimientos

Descripción:

Análisis de millones de posts/tweets con Hadoop para medir opinión pública.

Gráficas en tiempo real:

- Sentimiento promedio.
- Volumen de menciones.
- Nube de palabras.
- Tendencia de hashtags.
- Opinión por región.
- Evolución temporal.

Twitter-Entity Sentiment Analysis Dataset (Kaggle) (Kaggle)	Tweets con entidades y sentimientos; útil si quieres analizar por tema/hashtag.
--	---

6. Pipeline de procesamiento de datos genómicos

Descripción:

Uso de Hadoop para análisis de secuencias genéticas.

Gráficas en tiempo real:

- Distribución de mutaciones.
- Genes analizados/minuto.
- Score de riesgo.
- Comparación de cohortes.
- Redes de interacción genética.

[kaggle](#)

7. Predicción de consumo energético en ciudades inteligentes

 **Descripción:**
Uso de Hadoop para modelar demanda eléctrica.

 **Gráficas en tiempo real:**

- Consumo por zona.
- Predicciones próximas horas.
- Comparación con históricos.
- Picos detectados.
- Impacto del clima.
- Alertas de sobrecarga.

[kaggle](#)

8. Plataforma de análisis de logs de ciberseguridad

 **Descripción:**
Detección de ataques en tiempo real con Hadoop.

 **Gráficas en tiempo real:**

- Intentos de acceso.
- Alertas activas.
- Tráfico por puerto.
- Tipos de ataques.
- Vulnerabilidades en evolución.
- Score de riesgo.

[kaggle](#)

9. Análisis climático y modelado de eventos extremos

 **Descripción:**
Uso de Hadoop para análisis climático satelital.

 **Gráficas en tiempo real:**

- Mapas de calor climáticos.
 - Evolución de tormentas.
 - Predicción de intensidad.
 - Alertas activas.
 - Comparación histórica.
 - Series de temperatura/lluvia.
 - [kaggle](#)
-

10. Optimización de inventarios y logística en retail

 **Descripción:**

Sistema Hadoop para predecir demanda y optimizar inventarios.

 **Gráficas en tiempo real:**

- Nivel de stock.
- Predicciones de demanda.
- Costos logísticos.
- Mapas de rutas activas.
- Comparación de eficiencia.
- Alertas de stock bajo.

[kaggle](#)