

Final

Unidad	Unidad Final
Entrega	Método de entrega de la actividad en el campus virtual

1. ¿Cuáles son los objetivos de la práctica?

Los objetivos de la práctica final que propone un enfoque integral en el manejo de Big Data, abarcando sistemas de almacenamiento, procesamiento y visualización de datos, son los siguientes:

1. **Comprensión Práctica de Sistemas de Almacenamiento de Big Data:**
 - a. Familiarizar a los estudiantes con la configuración y gestión de sistemas de almacenamiento de Big Data como Hadoop, NoSQL, InfluxDB o MinIO.
 - b. Desarrollar habilidades para almacenar y gestionar eficientemente grandes volúmenes de datos.
2. **Aplicación de Técnicas de Procesamiento de Big Data:**
 - a. Implementar soluciones de procesamiento de datos utilizando Spark o herramientas de streaming como Kafka, Flink o Airflow.
 - b. Desarrollar competencias en el manejo, limpieza y transformación de datos a gran escala.
3. **Desarrollo de Habilidades en Visualización de Datos:**
 - a. Crear dashboards interactivos y visualmente atractivos utilizando herramientas como Grafana, Superset, TICK Stack, PowerBI o Matplotlib.
 - b. Aprender a comunicar insights y resultados de análisis de datos de manera efectiva a través de visualizaciones.
4. **Integración de Tecnologías de Big Data:**
 - a. Integrar diferentes tecnologías y plataformas de Big Data para construir una solución completa que cubra almacenamiento, procesamiento y visualización.
 - b. Entender cómo las distintas piezas de un ecosistema de Big Data interactúan y se complementan entre sí.
5. **Práctica de Habilidades de Resolución de Problemas y Pensamiento Crítico:**
 - a. Abordar y resolver problemas complejos relacionados con el manejo de grandes volúmenes de datos.
 - b. Analizar y extraer valor de los datos, aplicando un pensamiento crítico para generar insights relevantes.

2. Enunciado

Los alumnos deberán desarrollar un sistema completo de Big Data que incluya la integración de un sistema de almacenamiento (usando Hadoop, NoSQL, InfluxDB o MinIO), procesamiento de datos (mediante Spark o herramientas de streaming como Kafka, Flink o Airflow) y visualización de los resultados (usando Grafana, Superset, TICK Stack, PowerBI o Matplotlib con Dashboards). Los principales objetivos técnicos de la practica son:

Final

- Implementar un Sistema de Almacenamiento de Big Data:
 - Configurar y utilizar una plataforma de almacenamiento de Big Data como Hadoop, una base de datos NoSQL, InfluxDB o MinIO para almacenar un conjunto de datos significativo.
- Procesamiento de Datos:
 - Utilizar Spark o herramientas de streaming (Kafka, Flink, Airflow) para procesar y analizar los datos almacenados.
- Visualización de Datos:
 - Desarrollar dashboards utilizando herramientas como Grafana, Superset, TICK Stack, PowerBI o Matplotlib para visualizar los resultados del procesamiento de datos.

Detalles de la Práctica:

1. Almacenamiento de Datos: Configurar un clúster de Hadoop o una instancia de una base de datos NoSQL, InfluxDB o MinIO.

- Requisitos: Almacenar un conjunto de [datos](#) que será utilizado para el análisis.

2. Procesamiento de Datos: Utilizar Spark para batch processing o herramientas como Kafka, Flink o Airflow para procesar los datos en tiempo real.

- Requisitos: Limpiar, transformar y analizar los datos. El análisis debe **generar insights** significativos que puedan ser visualizados.

3. Visualización de Datos: Crear dashboards interactivos utilizando una de las herramientas sugeridas.

- Requisitos: Los dashboards deben mostrar los insights de los datos de una manera clara y visualmente atractiva. Deben incluir gráficos, tablas y/o métricas clave.

Evaluación:

- Complejidad y Configuración del Sistema de Almacenamiento: Correcta implementación y uso del sistema de almacenamiento elegido.
- Calidad del Procesamiento de Datos: Eficiencia y efectividad en el procesamiento de datos y la elección de las métricas a visualizar.
- Creatividad y Claridad en la Visualización de Datos: Diseño de dashboards que presenten los datos de forma intuitiva y reveladora.
- Documentación y Presentación: Calidad de la documentación del proyecto y presentación de los resultados.

3. Detalles de la entrega

1. Memoria en pdf que contenga:
 - a. Documentación que incluya la configuración del sistema, descripción del procesamiento de datos y explicación de los dashboards.
 - b. Presentación del proyecto, preferiblemente con una demostración del dashboard.
2. Dockerfile o Docker compose del proyecto.