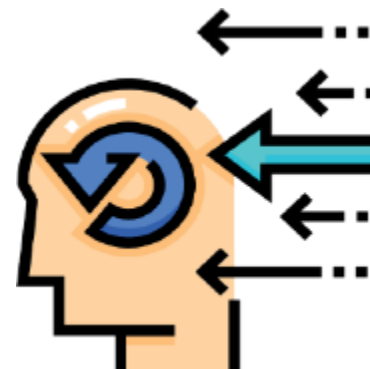




Módulo 7

Sesión N° 3



ACTIVIDAD:



Aplicación de apache Spark

- Objetivo: Evaluar la comprensión de los estudiantes sobre Apache Spark, su ejecución en la nube y su aplicación en el análisis de datos, a través de una investigación teórica y una actividad práctica que implique el uso de transformaciones y acciones en Spark.



Instrucciones:

1. Configuración del entorno:
 - a. Crear un entorno de Apache Spark local o en la nube.
 - b. Documentar los pasos seguidos para la configuración.
2. Obtención de datos:
 - a. Elegir una API pública (por ejemplo, OpenWeather, CoinGecko, etc.) para obtener datos en formato JSON o CSV.
 - b. Extraer y cargar los datos en Apache Spark.
3. Procesamiento de datos en Spark:
 - a. Aplicar al menos tres transformaciones sobre los datos (ejemplo: filtrado, agrupación, mapeo, conversión de tipos, etc.).
 - b. Aplicar al menos dos acciones para obtener resultados procesados (ejemplo: mostrar datos, contar registros, guardar resultados en un archivo, etc.).
4. Explicación del código:
 - a. Incluir comentarios en el código para describir el propósito de cada bloque de instrucciones.
 - b. Adjuntar un breve informe explicando el flujo del procesamiento de datos en Spark.
5. Análisis de opciones en la nube:
 - a. Investigar y seleccionar un proveedor de cloud computing (Amazon Web Services, Microsoft Azure, Google Cloud, etc.) que permita ejecutar Apache Spark.
 - b. Analizar y comparar las características del servicio, incluyendo costos, ventajas y limitaciones.
 - c. Presentar la información en un cuadro comparativo o en un documento estructurado.
6. Tiempo estimado de desarrollo: 1 hora
7. Modalidad: Grupal.
8. Modalidad de entrega: comprimido .zip .rar

