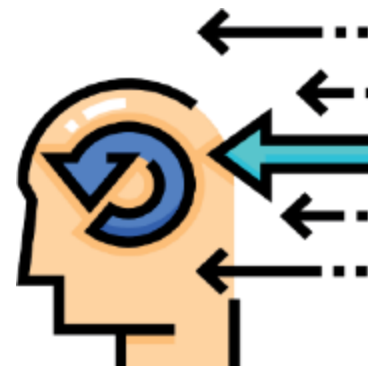




Módulo 7

Sesión N° 5



ACTIVIDAD:

Análisis de Flujos de Datos en Tiempo Real con Apache Spark



Streaming

- Objetivo: Aplicar conceptos aprendidos sobre Apache Streaming en un caso real de procesamiento de datos en tiempo real. El alumno deberá identificar los componentes clave del sistema de procesamiento, las herramientas utilizadas y explicar cómo se aplican los conceptos de DStreams, watermarking, ventanas de tiempo, entre otros.



Instrucciones:

- Investigación: Los estudiantes deben investigar cómo Apache Spark Streaming puede ser utilizado para el procesamiento de datos en tiempo real.
- Identificación de componentes: El alumno debe identificar los componentes clave involucrados en el procesamiento de datos en tiempo real, tales como fuentes de datos (Kafka, Flume, sockets), DStreams, transformaciones, agregaciones, watermarking, y sinks.
- Estudio de caso: Deben mencionar un caso en el que se utilicen estas tecnologías para procesar flujos de datos en tiempo real.
- Explicación: Explicar cómo los componentes identificados se integran en el sistema de procesamiento en tiempo real.

► Modalidad grupal.

► Tiempo estimado: 45 minutos.

