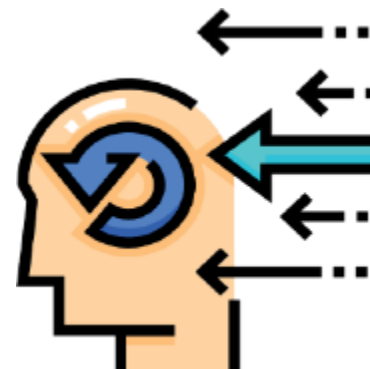




# Módulo 8

## Sesión N° 4



### ACTIVIDAD:



## Implementación de un Flujo de Ingesta de Datos en Streaming con Apache Nifi

- **Objetivo:** Aplicar los conceptos básicos de ingesta de datos en streaming utilizando Apache NiFi. Los estudiantes aprenderán cómo configurar un flujo de datos que lea información en tiempo real desde un archivo CSV, lo transforme en un formato adecuado y lo almacene en una base de datos.



### Requerimientos específicos:

- **Instalar Apache NiFi:** Asegúrate de tener Apache NiFi instalado y en funcionamiento en tu máquina local o en un entorno gestionado.
- **Configurar Apache NiFi:**
- **Leer Datos desde un Archivo CSV:** Usa el procesador GetFile para leer los archivos CSV que llegan a un directorio específico.
- **Transformar Datos:** Utiliza el procesador ConvertRecord para transformar los archivos CSV en formato JSON.
- **Guardar Datos en Base de Datos:** Usa el procesador PutDatabaseRecord para insertar los datos transformados en una base de datos (puede ser PostgreSQL o cualquier base de datos relacional disponible).
- **Programación:** Configura el flujo de datos para que se ejecute automáticamente cada vez que llegue un nuevo archivo CSV, con intervalos de ejecución configurables.
- **Monitoreo:** Asegúrate de poder monitorear el flujo de datos y verificar que el sistema esté funcionando correctamente.

Tiempo estimado: 45 Minutos

Formato de ejecución: Individual o en pareja





## Instrucciones:

1. Accede a Apache NiFi: Inicia Apache NiFi en tu máquina local y accede a la interfaz web de NiFi a través de <http://localhost:8080/nifi>.
2. Configura los Procesadores:
  - GetFile: Configura este procesador para leer los archivos CSV de una carpeta específica.
  - ConvertRecord: Usa este procesador para convertir los archivos CSV a formato JSON.
  - PutDatabaseRecord: Configura este procesador para insertar los datos transformados en una base de datos como PostgreSQL.
3. Monitorea la Ejecución:
  - Verifica que los datos estén siendo leídos correctamente desde el archivo CSV y cargados en la base de datos sin errores.
4. Automatiza el Flujo:
  - Asegúrate de que NiFi esté configurado para ejecutar el flujo automáticamente cuando lleguen nuevos archivos CSV.
5. Verifica el Funcionamiento:
  - Después de configurar el flujo, verifica que los datos se carguen correctamente en la base de datos y realiza algunos cambios en los archivos CSV para probar el flujo automático.

## Evaluación Formativa:

Correcta configuración de los procesadores: Se evaluará si los procesadores GetFile, ConvertRecord y PutDatabaseRecord están configurados correctamente para realizar la ingesta de datos.

Automatización y Monitoreo: Se revisará si el flujo está automatizado y si los estudiantes pueden monitorear el estado del flujo y los datos procesados.

