



Módulo 9

Sesión N° 5



ACTIVIDAD:



Diagrama de Flujo de Datos

- Objetivo: Representar gráficamente los pasos de un proceso ETL utilizando AWS Glue.



Instrucciones:

Una empresa de ventas online necesita mover datos de transacciones de su tienda en línea a un sistema de análisis para informes y decisiones. Los datos están almacenados en archivos CSV en Amazon S3. El proceso ETL utilizará AWS Glue para procesar esos datos.

Crea un diagrama en [Draw.io](https://draw.io) que represente este flujo de datos, incluyendo los siguientes componentes:



- Fuente de datos: Archivos CSV con las transacciones de la tienda en Amazon S3.
- Crawlers: AWS Glue Crawlers para explorar los archivos CSV y detectar su esquema.
- Clasificadores: AWS Glue clasificadores para identificar el tipo de archivo (CSV).
- Catálogo de datos: Donde se guardan los metadatos de los archivos.
- Trabajo ETL (Job): AWS Glue transformará los datos, por ejemplo, sumando las ventas por categorías de productos.
- Trigger (Desencadenador): Inicia el trabajo ETL de forma automática cada vez que se suben nuevos archivos a S3.
- Destino final: Los datos procesados se almacenan en Amazon Redshift para su análisis.

Tiempo estimado: 60 minutos.

