TRABAJO PRÁCTICO PROGRAMACIÓN AVANZADA PARA GRANDES VOLÚMENES DE DATOS

Integrantes: Martina De Angelis y Santiago Lares Harbin

Como primer paso, configuramos IAM en AWS para asegurar que los servicios puedan interactuar adecuadamente entre sí. Posteriormente, preparamos los datos crudos, los cuales fueron necesarios para nuestro DAG, y los almacenamos en un bucket de S3.

Luego procedimos a instalar Airflow en nuestra consola, dentro de un entorno virtual específico. Para verificar el correcto funcionamiento del DAG, realizamos pruebas locales antes de su implementación en una instancia EC2. Durante este proceso, enfrentamos varios retos relacionados con la instalación y configuración de Airflow, así como con la ejecución eficiente de los dags en el entorno de Airflow. Tras resolver estos problemas y confirmar su funcionamiento local, migramos el sistema a la instancia EC2. Esto implicó cambiar la versión preinstalada de Python por una versión anterior compatible con Airflow, copiar la carpeta de dags y reinstalar y reconfigurar Airflow en EC2. Además, tuvimos que ajustar la configuración de EC2 para permitir el tráfico a través del puerto 8080, utilizado por Airflow.

Finalmente, configuramos una API para visualizar los resultados. En comparación con Airflow, la configuración local de la API fue relativamente más sencilla, ya que no requería múltiples comandos. Además, el desarrollo del código para la API fue más directo, dado que las consultas SQL empleadas no requerían una codificación extensa