

Universidad de las Américas



Ingeniería en Desarrollo de software

Análisis visualización de datos

**Daniel Bustos, Alain Ruales, Jorge
Padilla**

Quito – Ecuador

2022-2023

Contenido

Instrucciones para la ejecución del proyecto	3
Pasos iniciales.....	3
Instalación de paquetes	3
Configuración de la base de datos en el entorno Python	3
Ejecución de scripts SQL.....	3
Generación de datos	4
Ejecución de ETL.....	4
Visualizar los resultados en Power BI.....	4

Instrucciones para la ejecución del proyecto

Pasos iniciales

- Es mandatorio tener todas las herramientas previamente instaladas y configuradas para su uso.
- Crear una carpeta donde se desea almacenar el proyecto.
- Clonar el repositorio con el comando "git -clone <https://github.com/jorgeandrespadilla/ProyectoFinal-UDLAICBS0003>" en la carpeta de destino
- Abrir la carpeta con el IDE de preferencia o desde terminal. Recomendado VS Code
- Debe contar con la versión 2.112.1161.0 (64-bit) o superior para no encontrar ningún tipo de incompatibilidad de Power BI

Instalación de paquetes

Para instalar los paquetes necesarios usados por Python, se debe ejecutar el comando `pip install -r requirements.txt`.

Librerías que se van a instalar

- Pandas
- sqlalchemy
- jproperties
- PyMySQL

Configuración de la base de datos en el entorno Python

Modificar los archivos de configuración `etl_db.properties` y `source_db.properties` en el directorio `config` de la solución de Python con las configuraciones de conexión a la base de datos correspondientes a su entorno local.

Ejecución de scripts SQL

Los scripts SQL del proyecto se encuentran ubicados en el directorio `sql`.

Para la base de datos fuente, los scripts SQL se encuentran en el directorio `SQL/source` y se deben ejecutar en el siguiente orden:

`1_source-initialization.sql` (inicialización de esquemas de la base de datos fuente)

`2_source-tables.sql` (creación de tablas de la base fuente)

Para la base de datos de ETL, los scripts SQL se encuentran en el directorio `sql/etl` y se deben ejecutar en el siguiente orden:

1_ *etl-initialization.sql* (inicialización de esquemas de la base de datos de ETL)

2_ *ext-tables.sql* (creación de tablas de extracción)

3_ *tra-tables.sql* (creación de tablas de transformación)

4_ *sor-tables.sql* (creación de tablas de la base SOR)

Generación de datos

Para generar los datos, se debe ejecutar el comando `python data_setup.py` en la raíz del proyecto. Este comando genera los datos que se cargaran en el directorio `data`.

Ejecución de ETL

Para ejecutar el ETL, se debe ejecutar el comando `python etl_process.py` en la raíz del proyecto.

Visualizar los resultados en Power BI

Una vez terminado la ejecución del proceso ETL debe actualizar los datos en Power BI para ello puede hacer uso de la opción actualizar, presente en el menú de inicio de la herramienta.

