

Pasos para ejecutar el proceso ETL

Prerrequisitos

- Tener **Visual Studio Code (VS Code)** instalado.
- Tener **MySQL** instalado y en funcionamiento.
-

Instrucciones

1. Clonar el repositorio desde GitHub

Abre **VS Code**.

Presiona **Alt + Shift + P** para abrir la *Command Palette*.

Escribe “**Git: Clone**” y presiona **Enter**.

Cuando se solicite la URL, pega el siguiente enlace del repositorio:

https://github.com/CSMore/ETL_Sistemas_Seguridad.git

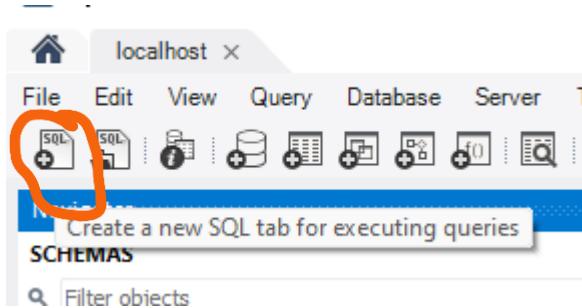
Selecciona la carpeta donde deseas guardar el proyecto.

Una vez completada la clonación, **abre la carpeta** desde VS Code.

2. Configurar la base de datos en MySQL

Abre **MySQL Workbench** (o tu cliente MySQL preferido).

Crea una nueva pestaña (tab) de consulta SQL.



Abre el archivo `modelado_dwh.sql` que está dentro del proyecto clonado y sigue estos pasos:

- a) Ejecuta el primer bloque de código (líneas 1 a 3) haciendo clic en el ícono de rayo para crear la base de datos.

The screenshot shows the MySQL Workbench interface. In the top navigation bar, the connection is set to 'localhost'. The 'File' menu is open. In the 'Navigator' pane, under 'SCHEMAS', there are two entries: 'inventario' and 'sys'. The 'SQL File 3' tab is active, displaying the following SQL code:

```
1 • CREATE DATABASE IF NOT EXISTS dwh_inventario;
2 • CREATE DATABASE IF NOT EXISTS etl_logs;
```

- b) Elimina ese bloque del editor SQL, luego copia y ejecuta el segundo bloque (líneas 9 a 22) y corre esto.

The screenshot shows the MySQL Workbench interface. In the top navigation bar, the connection is set to 'localhost'. The 'File' menu is open. In the 'Navigator' pane, under 'SCHEMAS', there is one entry: 'inventario'. The 'SQL File 3' tab is active, displaying the following SQL code:

```
1 • USE dwh_inventario;
2
3 -- Tabla final de inventario consolidado
4 • CREATE TABLE inventario_consolidado (
5     codigo_producto INT NOT NULL PRIMARY KEY, -- Si quiere que sea alfa numerico poner VARCHAR
6     nombre VARCHAR(250) NOT NULL,
7     descripcion_producto TEXT NOT NULL,
8     stock INT NOT NULL,
9     categoria VARCHAR(100) NOT NULL,
10    imagen_url VARCHAR(500) NOT NULL,
11    fecha_carga_dw TIMESTAMP NOT NULL DEFAULT CURRENT_TIMESTAMP,
12    INDEX ix_categoria (categoria),
```

- c) Finalmente, borra el segundo bloque, copia desde la línea 28 hasta el final del archivo y ejecútalo.

MySQL Workbench

localhost

File Edit View Query Database Server Tools Scripting Help

Navigator:

SCHEMAS

Filter objects

inventario

sys

SQL File 3* x

1 • USE etl_logs;

2

3 -- Registro de ejecución del ETL

4 • Ⓜ CREATE TABLE etl_run (

5 id_run BIGINT AUTO_INCREMENT PRIMARY KEY,

6 inicio_ejecucion DATETIME NOT NULL,

7 fin_ejecucion DATETIME NULL,

8 registros_extraidos INT DEFAULT 0,

9 registros_transformados INT DEFAULT 0,

10 registros_cargados INT DEFAULT 0,

11 estado ENUM('OK', 'ERROR', 'PARCIAL') NOT NULL DEFAULT 'OK',

12 mensaje_resumen VARCHAR(500)

13);

14

15 -- Registro de errores del ETL

16 • Ⓜ CREATE TABLE etl_error (

17 id_error BIGINT AUTO_INCREMENT PRIMARY KEY,

18 id_run BIGINT NOT NULL,

19 etapa ENUM('EXTRACCION', 'TRANSFORMACION', 'CARGA') NOT NULL,

20 fuente_archivo VARCHAR(255),

21 tipo_error VARCHAR(100) NOT NULL,

22 detalle VARCHAR(1000) NOT NULL,

23 creado_en TIMESTAMP NOT NULL DEFAULT CURRENT_TIMESTAMP,

24 FOREIGN KEY (id_run) REFERENCES etl_run(id_run)

25);

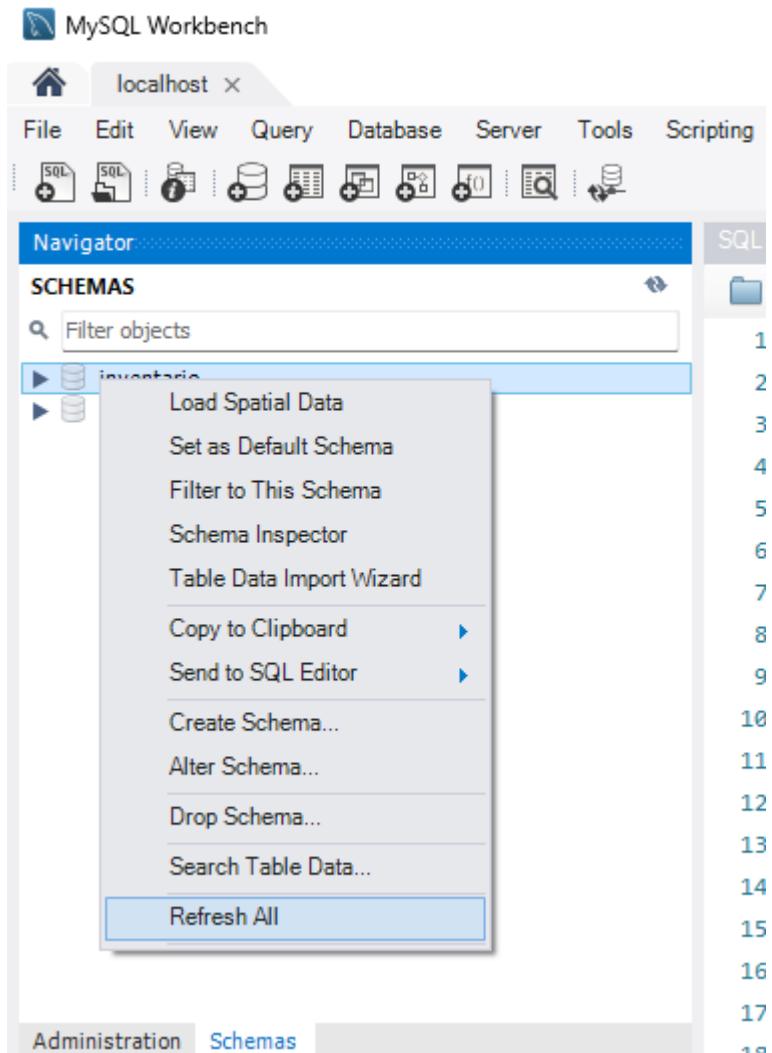
26

Administration Schemas

Information:

No object selected

3. Ve a los schemas y refresca



Cómo correr el pipeline

1. Abrir 'db.py' desde la carpeta de archivos clonados en VS code y cambiar el valor del password para que sea su contraseña de MySQL, la del usuario o conexión bajo el que creé la base de datos con los pasos anteriores.
2. Crear y activar un VM environment desde la terminal de VS Code
 - python -m venv .venv
 - source .venv/bin/activate -- MAC
 - .\venv\Scripts\Activate – WINDOWS
3. Instalar en el VM environment el requirements.txt, usa las siguiente línea de código: pip install -r requirements.txt
4. Correr el pipeline con: python -m etl.pipeline