

PROCESO DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA DE DATOS:

Se va a detallar el proceso para poder tener los datos disponibles en la base de datos de phpMyAdmin. Para ello, se dispone de 3 archivos de Python, cada uno con una función necesaria para acabar con la carga de datos final en la base.

En primer lugar, se tiene el archivo `descargarCSV.py`. El siguiente archivo se encarga de descargar todos los CSV de los pueblos. Se ha escogido un pueblo por comarca para que el proceso sea menos denso y haya menos cantidad de información. En `descargarCSV.py` se crea una tupla con un objeto por pueblo, y cada objeto dispone de una URL y un año inicial y final para ir modificando las direcciones web y así poder descargar por cada pueblo los 18 años correspondientes de datos. Se crearán 7 carpetas introduciendo los 18 CSV de cada pueblo en cada una de ellas.

En segundo lugar, está el archivo `createTables.py`. El archivo se encarga de crear las 3 tablas necesarias de la base de datos. Para ello, se conecta a la base de datos `aaee_24` para posteriormente crear las tablas en formato estrella. Para los nombres de las tablas y sus atributos se utiliza la nomenclatura adecuada. Las tablas son las siguientes: `LK_COMARCA`, `LK_FECHA` y `DT_DATOS`. En la tabla de las comarcas se dispone del id de la comarca, del nombre del pueblo y del nombre de la comarca. El id de la comarca es la clave primaria. En la tabla de datos se dispone de los datos correspondientes para realizar las consultas (precipitación, temperatura, clave de fecha y clave de comarca). Por último, en la tabla fecha se encuentran los atributos mes, año, día y semana, para poder realizar las consultas en función del tipo de tiempo escogido.

Por último, el archivo final `insertarDatosTablas.py` se encarga de insertar los 7 pueblos de cada comarca en `LK_COMARCA`, de insertar todos los días desde el 1 de enero de 2004 al 31 de diciembre de 2022 con su respectivo mes, año y semana, y finalmente, de insertar las precipitaciones y temperaturas máximas, mínimas y medias de cada uno de los pueblos asociados a su determinada fecha.

Todas las anteriores sentencias SQL se asocian a la base de datos `aaee_24` conectada al principio del archivo, para que, a través de un cursor, se puedan ejecutar cada una de las sentencias y se vean reflejados los datos en phpMyAdmin.

PASOS PARA EJECUTAR:

- 1- Abrir una terminal en el directorio Comarcas
- 2- Ejecutar `descargarCSV`: *`python3 descargarCSV.py`*
- 3- Ejecutar `createTables`: *`python3 createTables.py`*
- 4- Ejecutar `insertarDatosTablas`: *`python3 insertarDatosTablas.py`*