
SISTEMA DE BI SOBRE EL CLIMA EN NAVARRA

PROCESO ETL

FECHA DE ENTREGA:
7 DICIEMBRE

ANÁLISIS DE APLICACIONES EMPRESARIALES
IÑIGO AZNÁREZ GIL

ÍNDICE

1. Introducción.....	Pág. 3
2. Arquitectura del proceso.....	Pág. 4
3. Proceso ETL.....	Pág. 5-7
a. Extracción.....	Pág. 5
b. Transformación.....	Pág. 6
c. Carga.....	Pág. 7
4. Anexo: Capturas de ejecución.....	Pág. 8-9

INTRODUCCIÓN

El proceso ETL (Extracción, Transformación y Carga) se utiliza para obtener datos climáticos históricos de estaciones automáticas en Navarra, transformarlos a un formato estándar y cargarlos en una base de datos relacional para su posterior análisis. Este documento detalla cada paso del proceso, incluyendo las operaciones de extracción, transformación y carga.

ARQUITECTURA DEL PROCESO

Lenguaje de programación: El proceso ETL se desarrolla utilizando Python, lo que permite manejar tareas de automatización, manipulación de datos y conexión con bases de datos de manera eficiente.

Base de datos: Los datos procesados se almacenan en MySQL, utilizando la biblioteca pymysql para establecer la conexión y realizar operaciones SQL.

Directorio de trabajo: El proceso organiza los datos en las siguientes carpetas:

- *datos_climatologicos*: Carpeta destinada a almacenar los archivos CSV descargados desde las fuentes originales. Estos archivos contienen datos climáticos en bruto provenientes de las estaciones automáticas.
- *datos_transformados*: Carpeta donde se guardan los archivos procesados y estandarizados tras la etapa de transformación. Estos datos están listos para cargarse en la base de datos.

Fuentes de datos: <https://meteo.navarra.es/estaciones/descargardatos.cfm>

Este sitio proporciona enlaces a los datos históricos de estaciones automáticas. Los datos se encuentran en formato CSV, organizados por años y asociados a cada estación climática.

PROCESO ETL

EXTRACCIÓN

Objetivo: Descargar datos CSV de estaciones automáticas correspondientes a los años 2006-2023.

Pasos del proceso:

1. Obtener estaciones automáticas:
 - a. Se realiza una solicitud GET a la página de estaciones y se analiza el HTML con BeautifulSoup.
 - b. Se filtran las estaciones automáticas (icono estacionautomatica.gif).
 - c. Se extraen el ID y el nombre de cada estación.
2. Descargar archivos CSV:
 - a. Para cada estación, se accede a la página específica y se recopilan los enlaces a los archivos CSV.
 - b. Se descargan únicamente los archivos correspondientes a los años 2006-2023.
 - c. Los archivos se renombran con un identificador único y se almacenan en el directorio *datos_climatologicos*.

Errores manejados:

- Página no accesible.
- Ausencia de archivos CSV para una estación.
- Años fuera del rango especificado.

Funciones: script `descargar_datos_transformar.py`

- `obtener_estaciones_automaticas()`: Obtiene una array con los Ides de las estaciones automáticas para diferenciarlas de las manuales.
- `descargar_datos()`: Descarga los csvs de las estaciones automáticas que cumplen con la condición de contener datos de los últimos 18 años.

TRANSFORMACIÓN:

Objetivo: Estandarizar los datos descargados y preparar un formato adecuado para la carga en la base de datos.

Operaciones realizadas:

1. Formato de fecha:
 - a. Convertir las fechas al formato YY/MM/DD usando la función *transformar_fecha()*.
2. Reemplazar caracteres:
 - a. Reemplazar “,” por “.” de manera que tenemos todos los valores decimales con puntos.
 - b. Reemplazar “;” por “,” para tener todos los campos delimitados por comas.
3. Generación de archivos transformados:
 - a. Guardar los datos procesados en la carpeta *datos_transformados*.

Tras realizar la transformación ejecutar el script `comprobar_csv.py` para verificar que todas las filas contienen el mismo número de campos que la cabecera.

Errores manejados:

- Formato de fecha no válido.
- Archivos mal formateados.

Funciones: script `descargar_datos_transformar.py`

- *transformar_fecha()*: Transforma la fecha al formato correcto
- *transformar_datos()*: Realiza las modificaciones necesarias a los csvs de la carpeta *datos_climatologicos* y los almacena en el directorio *datos_transformados*.

CARGA

Objetivo: Insertar los datos transformados en la base de datos MySQL.

Pasos del proceso:

1. Conexión a la base de datos: Se establece conexión con el servidor MySQL mediante pymysql.
2. Preparación de la base de datos:
 - a. Vaciar las tablas existentes para evitar duplicados.
 - b. Crear las tablas necesarias: LK_COMARCAS, LK_MUNICIPIOS, LK_FECHAS y DT_DATOS_CLIMATICOS.
 - c. Insertar datos de referencia (comarcas y municipios).
3. Carga de datos:
 - a. Leer los archivos CSV transformados.
 - b. Insertar registros en las tablas correspondientes (LK_FECHAS y DT_DATOS_CLIMATICOS).

Errores manejados:

- Fallos en la conexión a la base de datos.
- Violaciones de claves foráneas.

Funciones: archivo crear_tablas_insertar.py

- *conectar_bd()* : Realiza la conexión a la base de datos.
- *vaciar_tablas()*: Desactiva claves foráneas y vacia el contenido de las tablas para su posterior inserción y evitar duplicados.
- *crear_tablas()*: Borra las tablas por si existen para evitar errores y las vuelve a crear.
- *insertar_comarcas()*: Inserta las comarcas de Navarra.
- *insertar_municipios()*: Inserta los municipios de Navarra cada uno con el identificador de su comarca correspondiente.
- *insertar_fechas()*: Inserta las fechas desde el 01/01/2006 hasta el 31/12/2023.
- *procesar_csv()*: Procesa el contenido del archivo csv para obtener los campos que queremos insertar en la base de datos.
- *insertar_datos_climáticos()*: Inserta los datos climáticos de los csvs transformados.

ANEXO: CAPTURAS DE EJECUCIÓN

descargar_datos_transformar.py:

```
alumno@eim-alu-69210:~/Escritorio/entregaFinal$ python3 -u descargar_datos_transformar.py
Se encontraron 62 estaciones automaticas.

Procesando estacion: Ablitas MAPA (ID: 274)
Estacion 274: Ablitas MAPA
Años filtrados: {2016, 2017, 2018, 2019, 2020, 2021, 2022, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015}
Descargando: 1 ablitas mapama_2006.csv
Descargando: 1 ablitas mapama_2007.csv
Descargando: 1 ablitas mapama_2008.csv
Descargando: 1 ablitas mapama_2009.csv
Descargando: 1 ablitas mapama_2010.csv
Descargando: 1 ablitas mapama_2011.csv
Descargando: 1 ablitas mapama_2012.csv
Descargando: 1 ablitas mapama_2013.csv
Descargando: 1 ablitas mapama_2014.csv
Descargando: 1 ablitas mapama_2015.csv
Descargando: 1 ablitas mapama_2016.csv
Descargando: 1 ablitas mapama_2017.csv
Descargando: 1 ablitas mapama_2018.csv
Descargando: 1 ablitas mapama_2019.csv
Descargando: 1 ablitas mapama_2020.csv
Descargando: 1 ablitas mapama_2021.csv
Descargando: 1 ablitas mapama_2022.csv
Descargando: 1 ablitas mapama_2023.csv
Descarga completada para la estacion 274

Procesando estacion: Abodi (Pikatua) GN (ID: 519)
Estacion 519: Abodi (Pikatua) GN
Años filtrados: {2022}
La estacion no tiene todos los años requeridos. Omitiendo descarga.

Procesando estacion: Adiós MAPA (ID: 263)
Estacion 263: Adiós MAPA
Años filtrados: {2016, 2017, 2018, 2019, 2020, 2021, 2022, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015}
Descargando: 2 adiós mapama_2006.csv
Descargando: 2 adiós mapama_2007.csv
Descargando: 2 adiós mapama_2008.csv
Descargando: 2 adiós mapama_2009.csv
Descargando: 2 adiós mapama_2010.csv
Descargando: 2 adiós mapama_2011.csv
Descargando: 2 adiós mapama_2012.csv
Descargando: 2 adiós mapama_2013.csv
Descargando: 2 adiós mapama_2014.csv
Descargando: 2 adiós mapama_2015.csv
Descargando: 2 adiós mapama_2016.csv
Descargando: 2 adiós mapama_2017.csv
```


crear_tablas_insertar.py:

```
alumno@eim-alu-69210:~/Escritorio/entregarfinal$ python3 -u crear_tablas_insertar.py
Conectando a la base de datos...
Conexión establecida.
Vacizando tablas...
Tablas vaciadas.
Creando tablas...
Tablas creadas.
Insertando comarcas...
Comarca 1 : Tierra Estella insertado
Comarca 2 : Zona Noroeste insertado
Comarca 3 : Zona Pirineo insertado
Comarca 4 : Pamplona insertado
Comarca 5 : Navarra Media Oriental insertado
Comarca 6 : Rivera Alta insertado
Comarca 7 : Tudela insertado
Comarcas insertadas.
Insertando municipios...
Municipio 3 : Aguilar de Codés insertado
Municipio 14 : Bargota insertado
Municipio 23 : Estella insertado
Municipio 31 : Los Arcos insertado
Municipio 35 : Ancín insertado
Municipio 45 : Trinidad de Iturgoyen insertado
Municipio 49 : Villanueva de Yerri insertado
Municipio 6 : Aralar insertado
Municipio 21 : Doneztebe-Santesteban insertado
Municipio 24 : Etxarri-Aranatz insertado
Municipio 28 : Gorramendi insertado
Municipio 37 : Oskotz insertado
Municipio 29 : Irabia insertado
Municipio 5 : Aoiz insertado
Municipio 7 : Arangoiti insertado
Municipio 8 : Arazuri insertado
Municipio 15 : Beortegui insertado
Municipio 22 : El Perdón insertado
Municipio 32 : Lumbier insertado
Municipio 2 : Adios insertado
Municipio 4 : Aibar insertado
Municipio 9 : Artajona insertado
Municipio 18 : Carrascal insertado
Municipio 36 : Olite insertado
Municipio 39 : San Martín de Unx insertado
Municipio 43 : Tafalla insertado
Municipio 48 : Ujué insertado
Municipio 50 : Yesa insertado
Municipio 25 : Falces insertado
```

```
Municipio 47 : Tudela (Valdetellas) insertado
Municipios insertados.
Insertando fechas...
Fechas insertadas.
Procesando archivo: datos_transformados/21_doneztebe-santesteban gn_2014.csv
Procesando archivo: datos_transformados/28_gorramendi gn_2023.csv
Procesando archivo: datos_transformados/17_carcastillo (la oliva) gn_2006.csv
Procesando archivo: datos_transformados/14_bargota mapama_2013.csv
Procesando archivo: datos_transformados/29_irabia gn_2017.csv
Procesando archivo: datos_transformados/19_cascante mapama_2017.csv
Procesando archivo: datos_transformados/28_gorramendi gn_2012.csv
Procesando archivo: datos_transformados/2_adiós mapama_2016.csv
Procesando archivo: datos_transformados/47_tudela (valdetellas) mapama_2007.csv
Procesando archivo: datos_transformados/17_carcastillo (la oliva) gn_2012.csv
Procesando archivo: datos_transformados/26_fitero mapa_2023.csv
Procesando archivo: datos_transformados/23_estella gn_2012.csv
Procesando archivo: datos_transformados/1_ablitas mapama_2007.csv
Procesando archivo: datos_transformados/19_cascante mapama_2011.csv
Procesando archivo: datos_transformados/16_cadreita intia_2015.csv
Procesando archivo: datos_transformados/36_olite intia_2007.csv
Procesando archivo: datos_transformados/41_sartaguda gn_2020.csv
Procesando archivo: datos_transformados/45_trinidad de iturgoien gn_2008.csv
Procesando archivo: datos_transformados/24_etxarri-aranatz gn_2018.csv
Procesando archivo: datos_transformados/38_san adrián mapama_2016.csv
Procesando archivo: datos_transformados/50_yesa gn_2014.csv
Procesando archivo: datos_transformados/49_villanueva de yerri gn_2009.csv
Procesando archivo: datos_transformados/50_yesa gn_2020.csv
Procesando archivo: datos_transformados/21_doneztebe-santesteban gn_2006.csv
Procesando archivo: datos_transformados/3_aguilar de codés gn_2022.csv
Procesando archivo: datos_transformados/34_murillo el fruto mapama_2016.csv
Procesando archivo: datos_transformados/20_corella intia_2014.csv
Procesando archivo: datos_transformados/42_sesma mapama_2006.csv
Procesando archivo: datos_transformados/11_bardenas (el plano) mapa_2022.csv
Procesando archivo: datos_transformados/24_etxarri-aranatz gn_2006.csv
Procesando archivo: datos_transformados/25_falces mapama_2012.csv
Procesando archivo: datos_transformados/31_los arcos mapama_2014.csv
Procesando archivo: datos_transformados/46_tudela (montes del cierzo) gn_2014.csv
Procesando archivo: datos_transformados/28_gorramendi gn_2008.csv
Procesando archivo: datos_transformados/4_aibar mapama_2009.csv
Procesando archivo: datos_transformados/7_arangoiti gn_2011.csv
```