

Prueba técnica ETL de datos

Los archivos que menciono en este procedimiento ETL pueden encontrarse adjuntos en el envío de toda la prueba.

1. Fase de extracción:

Tomé 3 archivos distintos ():

- Un archivo CSV extraído de kaggle para simular los clientes
- Un archivo CSV generado con solicitudes
- Un archivo CSV con los datos de los usuarios internos (administradores y agentes)

2. Fase de Transformación:

Analicé los 3 archivos, de esta manera determiné que el único archivo a transformar es el archivo “clientes.csv”. Un archivo “.csv” y un “.json” pueden extraerse en Excel para su fácil lectura. Sin embargo, es posible realizar el proceso ETL directamente desde SQL con los siguientes pasos.

Se crea una tabla temporal para la información sin editar:

```
1 CREATE TABLE clientes_raw (  
2 id INT,  
3 nombre VARCHAR(255),  
4 correo VARCHAR(255)  
5 );
```

Se carga el archivo en dicha tabla temporal:

```
1 LOAD DATA INFILE 'clientes.csv'  
2 INTO TABLE clientes_raw  
3 FIELDS TERMINATED BY ','  
4 LINES TERMINATED BY '\n'  
5 IGNORE 1 LINES;
```

Se edita el archivo como se requiera, por ejemplo:

```
1 UPDATE clientes_raw  
2 SET nombre = TRIM(nombre), --quitando los espacios extras  
3     correo = TRIM(correo);  
4  
5  
6 DELETE FROM clientes_raw  
7 WHERE correo NOT LIKE '%@%'; -- se eliminan los correos erroneos
```

3. Verifiqué en que carpeta MYSQL permite subir archivos con el comando `"SHOW VARIABLES LIKE 'secure_file_priv'"`. Con los archivos copiados en la carpeta cargué cada archivo CSV a su tabla correspondiente. Tratando adicionalmente los valores nulos en caso de ser necesario. Ver archivo "Carga de datos.sql"