

Documentación del Proceso ETL

1. Introducción

El objetivo de este proceso ETL es transformar datos operativos simulados de un restaurante en información estructurada, limpia y lista para análisis de Inteligencia de Negocios. Se trabajó con datos relacionados a ventas, clientes, promociones, quejas, compras de ingredientes, costos operativos y platillos.

Las herramientas utilizadas fueron:

- Pentaho Data Integration (Spoon) para el proceso ETL.
- Microsoft Excel como fuente de datos.
- SQL Server como base de datos destino.

2. Descripción de las fuentes de datos

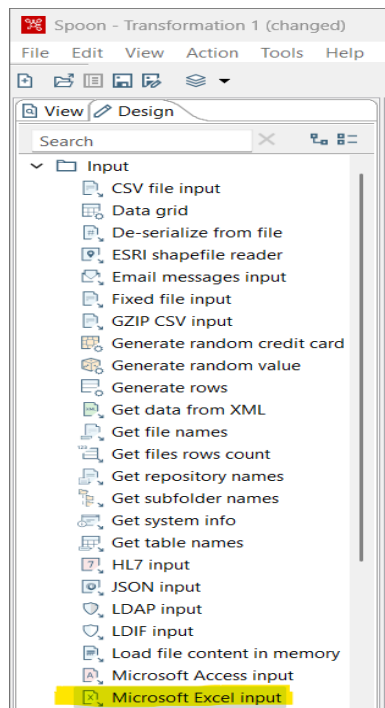
Se diseñaron 11 tablas de prueba en Excel, con las siguientes entidades:

- Ventas
- DetallesVenta
- Clientes
- Platillos
- Ingredientes
- Recetas
- Promociones
- ComprasIngredientes
- CostosOperacion
- QuejasYReseñas
- Sucursales

Cada tabla fue cargada a Pentaho usando pasos de tipo Microsoft Excel Input, una hoja por tabla.



Microsoft Excel input



3. Extracción (Extract)

La fuente fue un archivo de Excel llamado Datos_Restaurante_Ejemplo.xlsx. Se usó el paso Microsoft Excel Input en Pentaho para extraer los datos. Cada paso fue configurado para leer una hoja específica del archivo, identificando correctamente los encabezados y campos.

4. Transformación (Transform)

Algunas transformaciones básicas realizadas:

- Conversión de tipos de datos (por ejemplo, fechas y decimales).
- Limpieza de valores vacíos (por ejemplo, NULL en id_promocion).
- Revisión de consistencia en claves foráneas antes de la carga.

No se realizaron transformaciones complejas debido a que los datos fueron generados limpios. Sin embargo, Pentaho permite incorporar pasos como Filter Rows, Calculator, Replace Values, etc. si se necesitara.

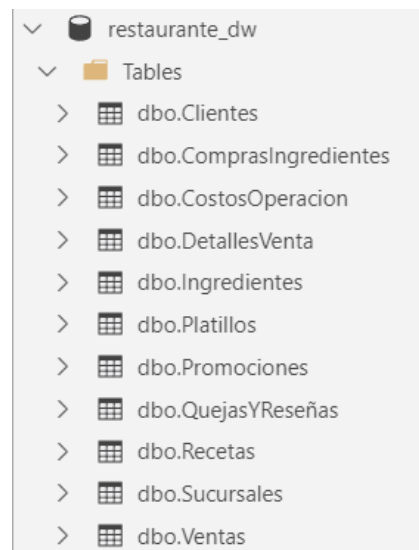
5. Carga (Load)

Se creó una base de datos en SQL Server llamada restaurante_dw ejecutando:

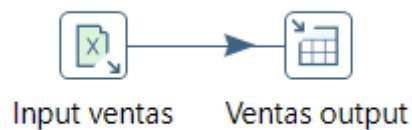
```
CREATE DATABASE restaurante_dw;
```

Cabe recalcar que la base de datos se creó haciendo uso del contenedor de docker de Microsoft SQL, y en Azure es en donde se corrió el comando

Luego, se generaron manualmente las tablas de destino con scripts SQL.



En Pentaho, se configuraron pasos Table Output para cada tabla y se conectaron con los pasos de entrada. La conexión fue hecha con el driver oficial mssql-jdbc y configurada manualmente como una conexión Generic Database.



6. Resultado

Al final del proceso ETL:

- Todas las tablas fueron exitosamente cargadas a SQL Server.
- Los datos están listos para ser usados en Power BI.
- Se logró cumplir con el objetivo de estructurar la información para análisis por problemática.