

Ejercicio completo: Tienda Online

1. Enunciado

Una tienda online necesita consolidar la información de sus ventas cruzando tres fuentes de datos independientes: sus clientes, sus pedidos y su catálogo de productos.

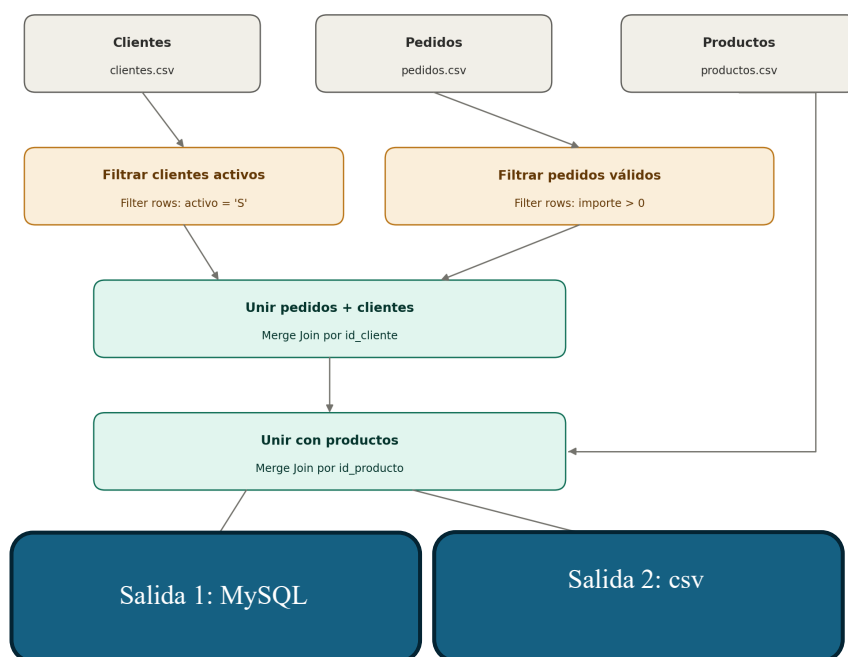
El objetivo es construir una transformación en Pentaho Data Integration (Spoon) que lea estas tres fuentes, limpie los registros no válidos, cruce la información por sus claves comunes y genere dos salidas: una tabla de ventas consolidadas lista para análisis, y un fichero de control con los registros descartados.

2. Datos de entrada

Se proporcionan tres ficheros CSV con datos generados aleatoriamente, separados por punto y coma (;): clientes.csv, productos.csv y pedidos.csv.

3. Flujo de la transformación

El siguiente diagrama resume el flujo completo que debes construir en Spoon, desde las tres entradas hasta las dos salidas finales:



4. Pasos a implementar

4.1 Entradas (CSV file input)

Añade tres pasos CSV file input, uno por fichero. Configura el separador ';' y marca la primera fila como cabecera. Comprueba el tipo de dato de cada campo en la pestaña Fields (por ejemplo, precio e importe deben detectarse como Number).

4.2 Filtrado de clientes activos

Tras la entrada de clientes, añade un Filter rows con la condición activo = 'S'. Envía las filas verdaderas hacia la unión por clave; las falsas pueden descartarse o dirigirse a un fichero de control de clientes inactivos.

4.3 Filtrado de pedidos válidos

Tras la entrada de pedidos, añade un Filter rows con la condición importe > 0. Esto descarta pedidos cancelados o con errores de captura (importe o cantidad en cero o negativos).

4.4 Ordenación previa a las uniones

El paso Merge Join exige que ambos flujos de entrada lleguen ordenados por la clave de unión. Añade un paso ordenar.

4.5 Unión 1: pedidos + clientes

Usa un unión, clave id_cliente en ambos flujos.

4.6 Unión 2: resultado + productos

Repite el mismo patrón para incorporar los datos de producto, usando id_producto como clave.

4.7 Salida 1: tabla de ventas en MySQL

Con un paso Table output, crea una conexión a MySQL (host, base de datos, usuario y contraseña) y define como tabla destino ventas_consolidadas..

4.8 Salida 2: fichero

Salida a fichero, todo igual que MySQL.