	<p style="text-align: center;">UNIVERSIDAD DON BOSCO</p> <p style="text-align: center;">INGENERIA EN CIENCIAS DE LA COMPUTACION</p>
	<p>Nombre de la práctica: Herramienta ETL parte III: Manejo de Excepciones y errores</p> <p>Materia: Data WareHouse y Minería de Datos</p>

I. Objetivos

1. Que el estudiante comprenda el uso de herramientas de depuración y manejo de excepciones.
2. Que el estudiante sea capaz de resolver problemas derivados de la extracción de diversos orígenes de datos.

II. Introducción

Un ETL denota el movimiento de datos entre sistemas, usualmente preparándolo para el uso de paneles y herramientas de Business Intelligence. Hoy en día, esto generalmente requiere la preparación de datos para su uso por mensajes, archivos planos y llamadas a API. Los datos no son solo enormes, sino complejos. Se almacena en numerosos archivos, aplicaciones preempaquetadas, sistemas mainframe y bases de datos relacionales. Eso equivale a masas de datos ligeramente diferentes, en diferentes formatos, todos los cuales deben convertirse y moverse rápidamente a través de los sistemas. Es muy importante saber extraer el significado real de la información, pues durante la extracción y la transformación se pueden dar infinidad de casos y problemas que deben ser resueltos para que la data sea consistente pues el ingreso de información incorrecta dará como resultado, información incorrecta.

III. Desarrollo

Paso 1: Crear una carpeta en C:/ denominada bulkfiles

Dentro de ella agregar los archivos de recursos que se le facilitarán y una carpeta denominada error_output que contendrá salidas de errores de ejecución. En ella debe crear 4 archivos de texto.

This PC > Local Disk (C:) > bulkfiles > error_output				Search error_outpu
<input type="checkbox"/> Name	Date modified	Type	S	
date_error	8/18/2019 10:32 A...	Text Document		
error	8/18/2019 10:32 A...	Text Document		
flat_file_error	8/18/2019 10:32 A...	Text Document		
negative_price_error	8/18/2019 10:32 A...	Text Document		

Paso 2: Editará únicamente negative_price_error.txt conteniendo en la primera línea del documento el siguiente texto:

Categoria,Producto,Marca,Cantidad,Unidad,Precio,Establecimiento,Direccion,Departamento,Municipio,Fecha Sondeo

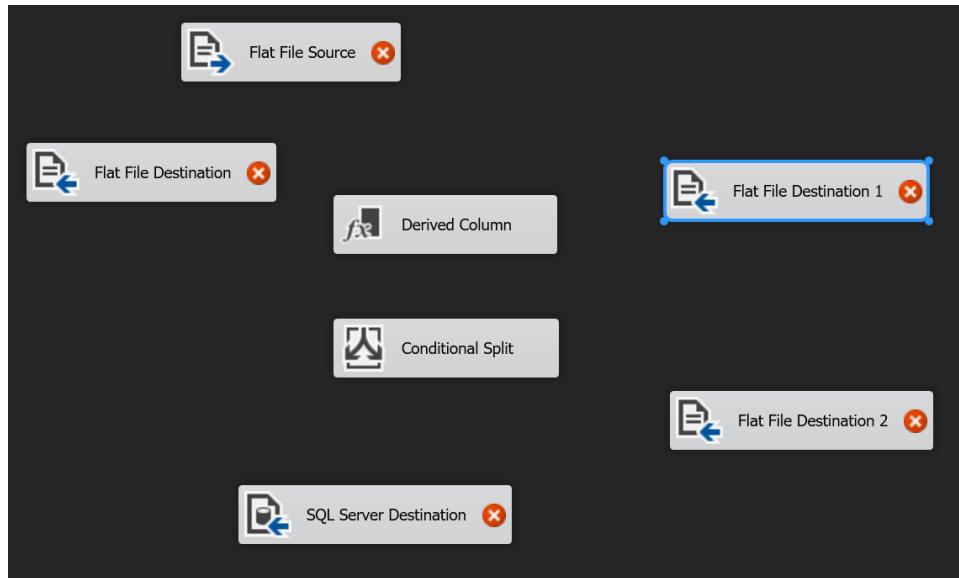
Paso 3: Crear una nueva base de datos en SQL Server a partir del siguiente script:

```
create database sv_school_products;

use sv_school_products;

create table products(
Categoria nchar(100),
Producto nchar(100),
Marca nchar(100),
Cantidad decimal,
Unidad nchar(100),
Precio money,
Establecimiento nvarchar(4000),
Direccion nvarchar (4000),
Departamento nchar(100),
Municipio nchar(100),
[Fecha Sondeo] date
);
```

Paso 4: Crear un nuevo proyecto SSIS y agregar los siguientes componentes sin asociar aún:



Renombrarlos para mayor comodidad



Paso 5: Vamos a editar nuestra primera conexión hacia el archivo xaa.txt que se encuentran dentro de la carpeta bulkfiles:

Flat File Connection Manager Editor

Connection manager name: Data

Description:

Select a file and specify the file properties and the file format.

File name: C:\bulkfiles\xaa.txt Browse...

Locale: English (United States) Unicode

Code page: 1252 (ANSI - Latin I)

Format: Delimited

Text qualifier: " (circled in red)

Header row delimiter: (CR)(LF)

Header rows to skip: 0

☒ Column names in the first data row

Importante: Note el valor de Text qualifier: Una comilla doble.

Por otra parte, columns no sufrirá ningún cambio.

Flat File Connection Manager Editor

Connection manager name: Data

Description:

Specify the characters that delimit the source file:

Row delimiter: (LF)

Column delimiter: Comma (,)

Preview rows 2-101:

Categoria	Producto	Marca	Cantidad	Unidad
Misceláneos	Abaco	Sin Marca	1	Un
Misceláneos	Abaco	Sin Marca	1	Un
Misceláneos	Abaco	Sin Marca	1	Un
Misceláneos	Abaco	Sin Marca	1	Un
Misceláneos	Abaco	Sin Marca	1	Un

En la categoría Advanced haremos una revisión más exhaustiva en los siguientes campos:

Flat File Connection Manager Editor

Connection manager name: Data

Description:

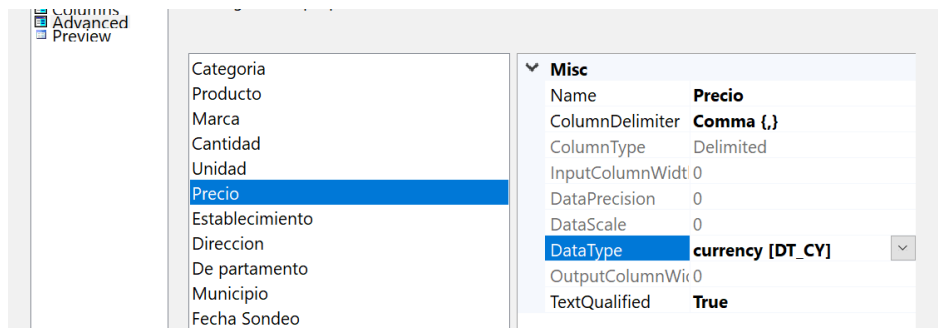
Configure the properties of each column.

General Columns Advanced Preview

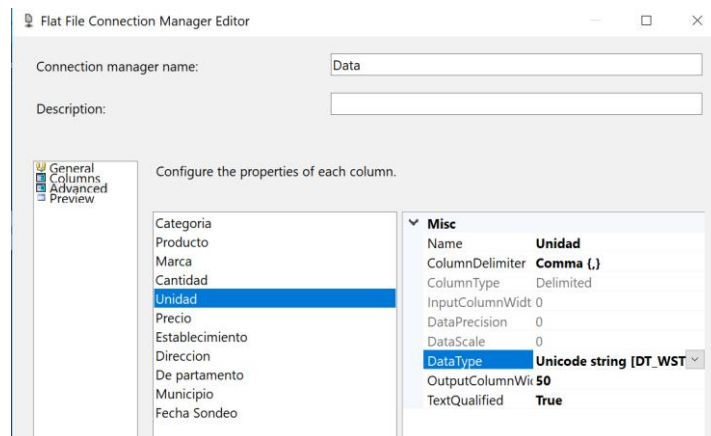
Categoria
Producto
Marca
Cantidad
Unidad
Precio
Establecimiento
Direccion
Departamento
Municipio
Fecha Sondeo

Misc

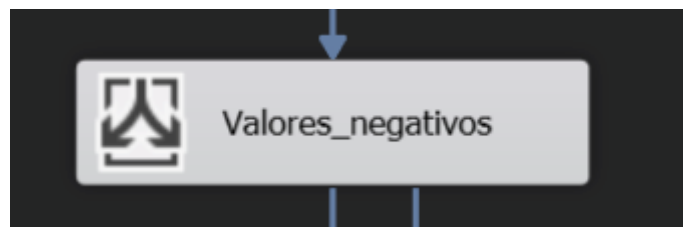
Name	Cantidad
ColumnDelimiter	Comma (,)
ColumnType	Delimited
InputColumnWidth	0
DataPrecision	0
DataScale	0
DataType	float [DT_R4]
OutputColumnWidth	0
TextQualified	True

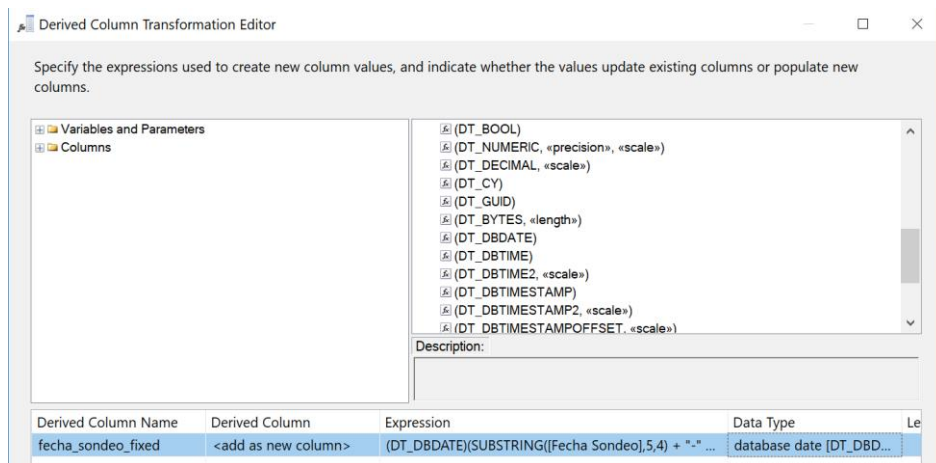


Los demás campos debes ser Unicode String. Ejemplo:



Paso 6: En la Conversion, crearemos un campo denominado fecha_sondeo_fixed

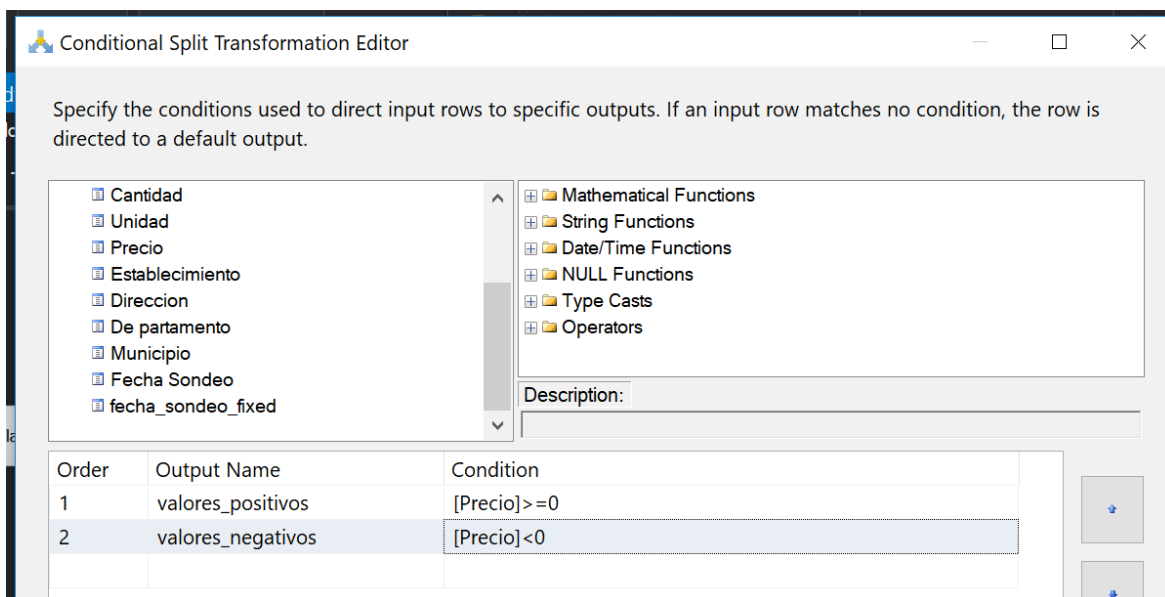
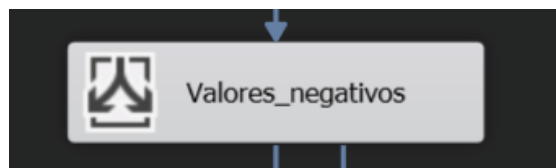




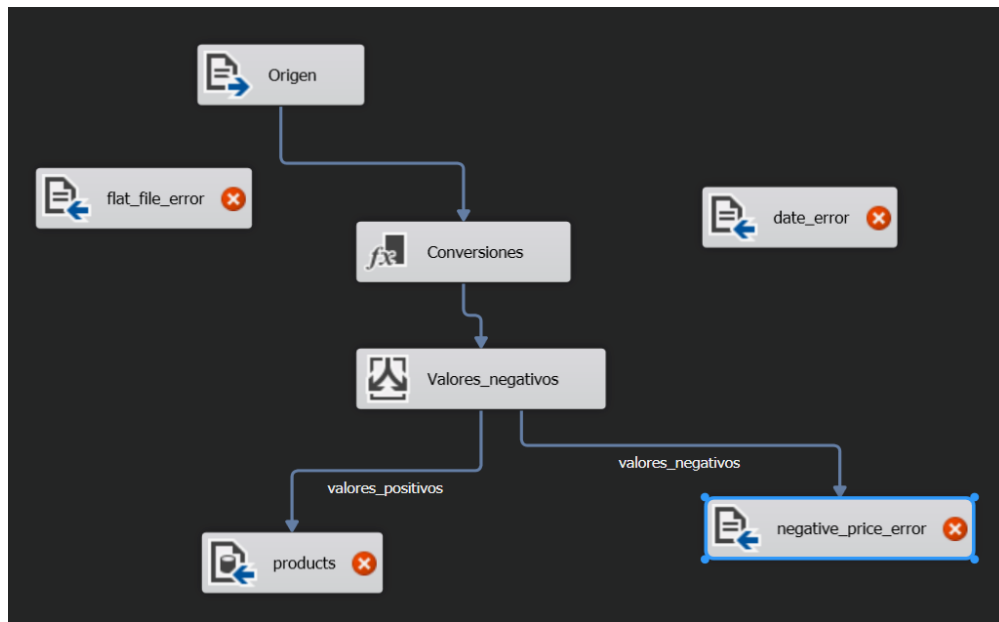
La expresión es la siguiente:

(DT_DBDATE)(SUBSTRING([Fecha Sondeo],5,4) + "-" + SUBSTRING([Fecha Sondeo],3,1) + "-" + SUBSTRING([Fecha Sondeo],1,1))

Paso 7: El Condicional Split agregaremos las siguientes condiciones:

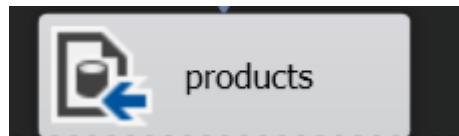


El diagrama debe llevar esta secuencia a este punto:

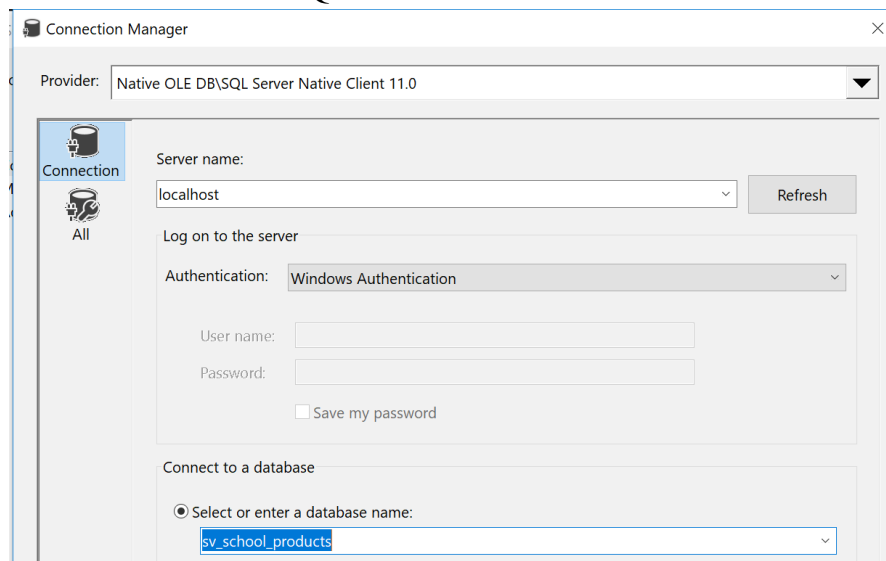


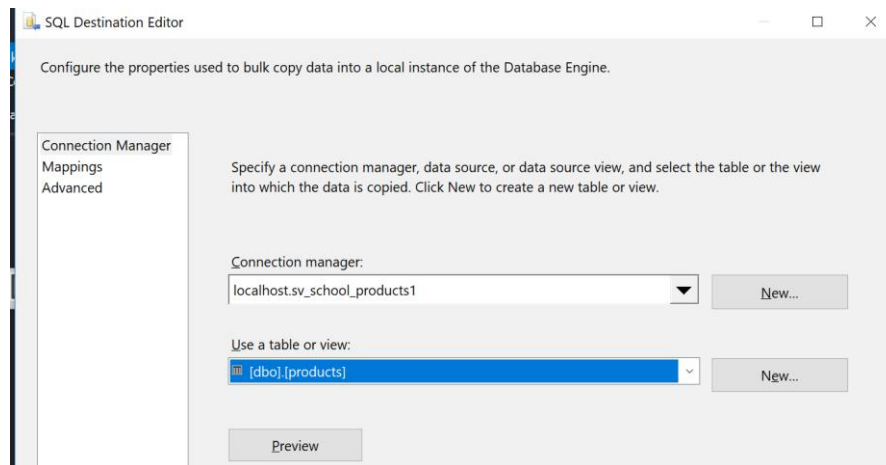
Paso 8: Ahora configuraremos los destinos faltantes.

Paso 8.1: El destino de products



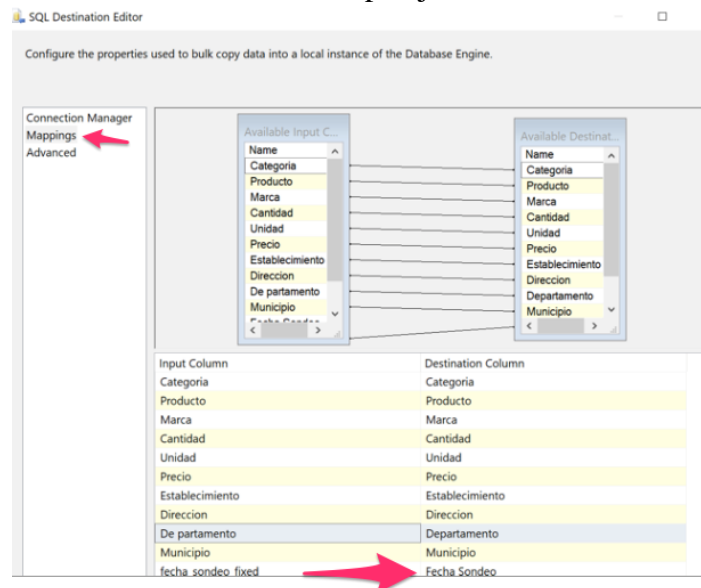
Es una conexión Native OLE DB a SQL Server.



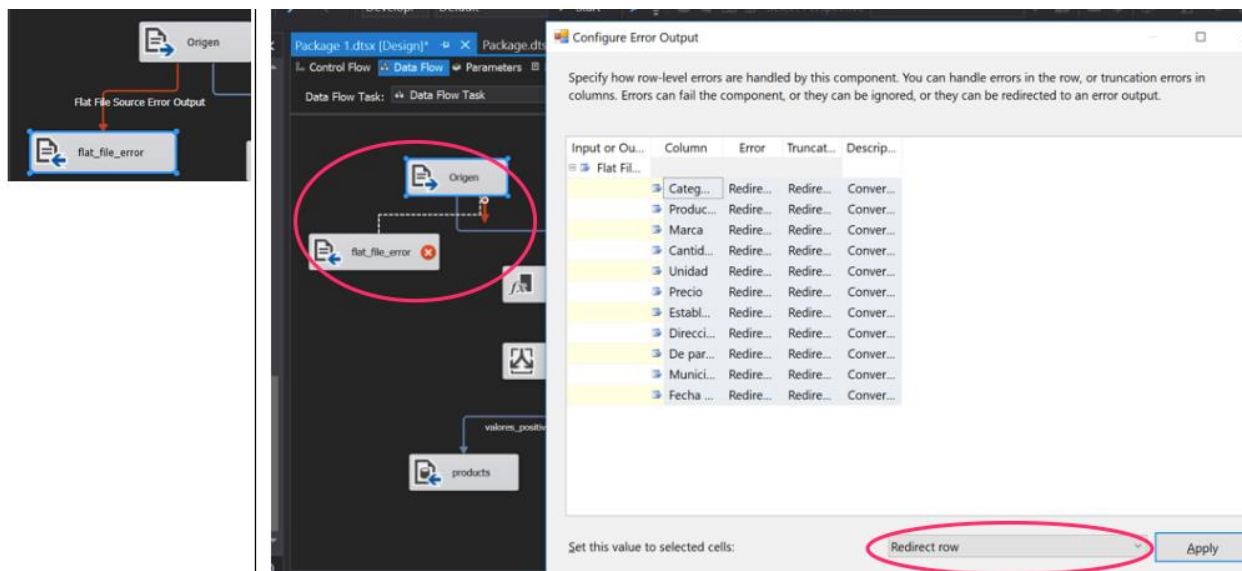


Emparejamos los campos:

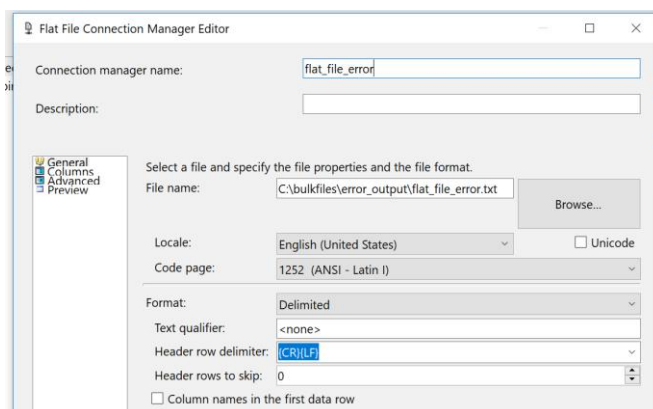
Note que el campo de fecha_sondeo_fixed se empareja con Fecha Sondeo.



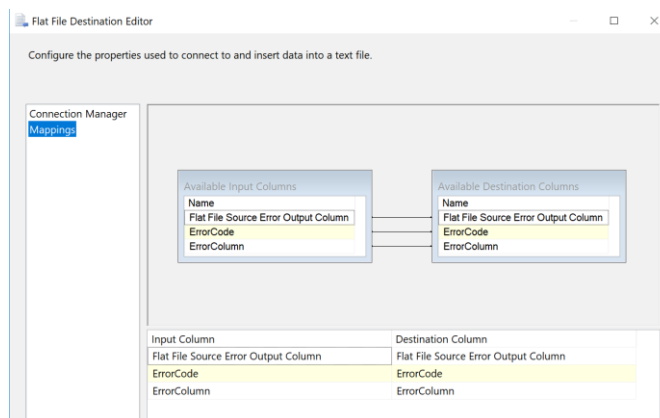
Paso 8.2: Ahora indicamos que los errores de lectura de ficheros de origen a flat_file_error serán redireccionados aplicando la regla “Redirect row”:



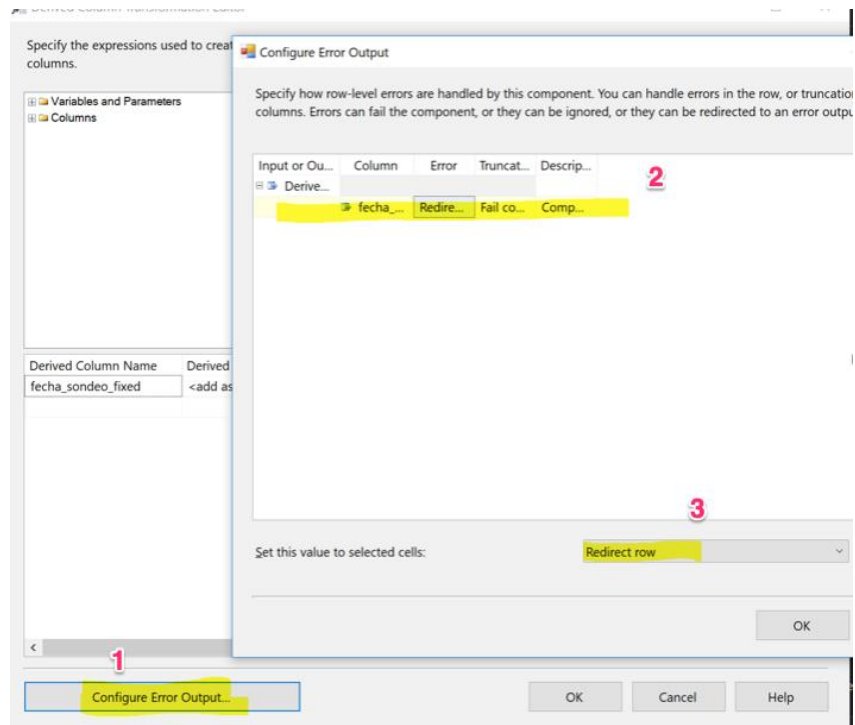
Nuestra conexión es:



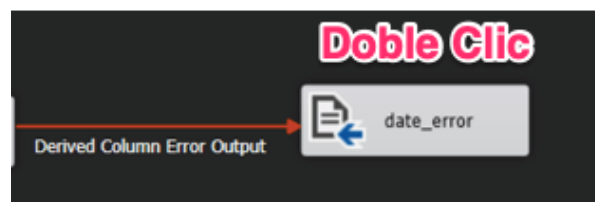
El Mapping:



Paso 8.3: Para la conversión de fecha:



Configuramos el destination file:



Flat File Connection Manager Editor

Connection manager name:

Description:

Select a file and specify the file properties and the file format.

File name:

Locale: ☐ Unicode

Code page:

Format:

Text qualifier:

Header row delimiter:

Header rows to skip:

☐ Column names in the first data row

Flat File Destination Editor

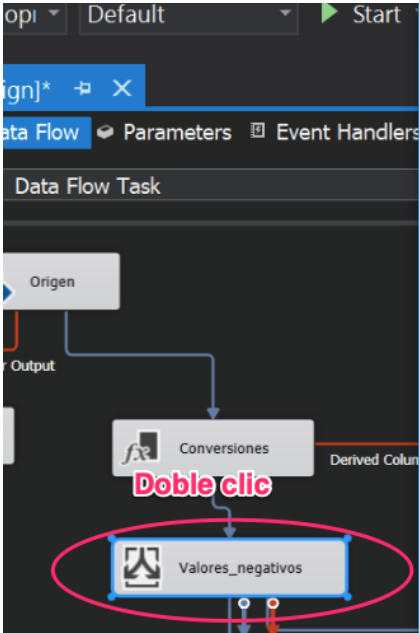
Configure the properties used to connect to and insert data into a text file.

Connection Manager:

Available Input Columns		Available Destination Columns
Name		Name
Categoria		Categoria
Producto		Producto
Marca		Marca
Cantidad		Cantidad
Unidad		Unidad
Precio		Precio
Establecimiento		Establecimiento
Direccion		Direccion
De departamento		De departamento
Municipio		Municipio
Estado		Estado

Input Column Destination Column

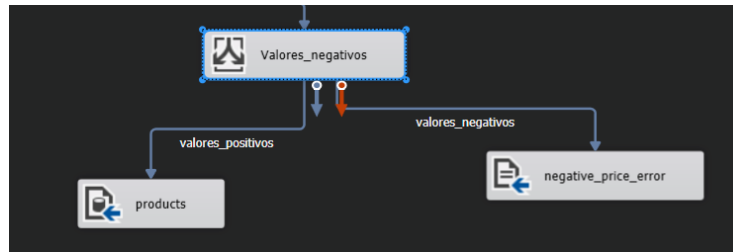
Paso 8.4: Finalmente el destination file de los errores por valores negativos:



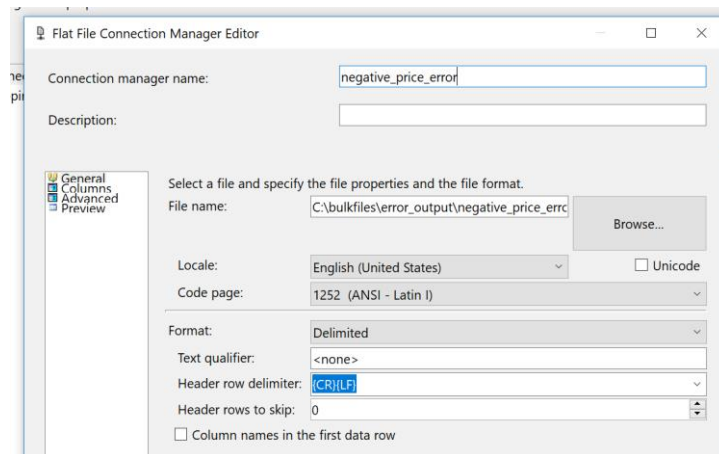
Conditional Split Transformation Editor

Specify the conditions used to direct input rows to specific outputs or to a default output.

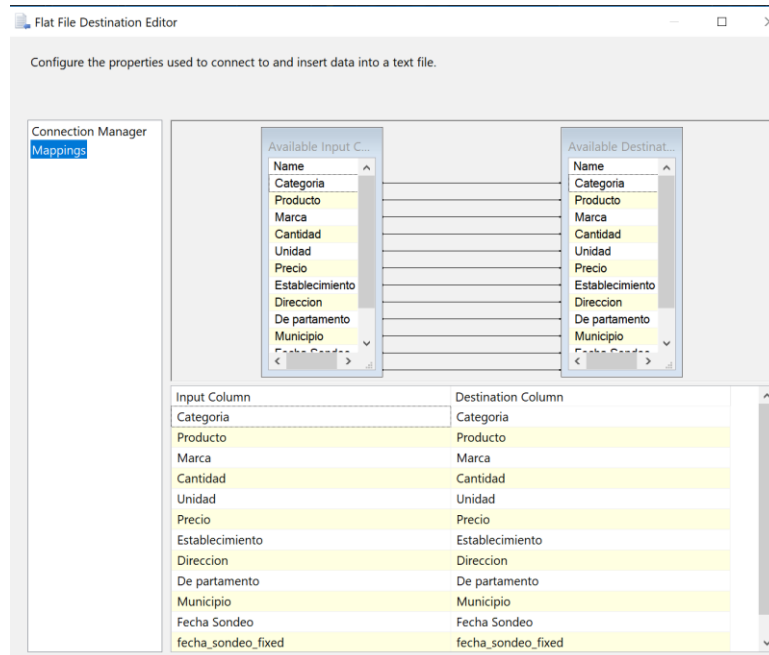
Order	Output Name	Condition
1	valores_positivos	Precio >= 0
2	valores_negativos	Precio < 0



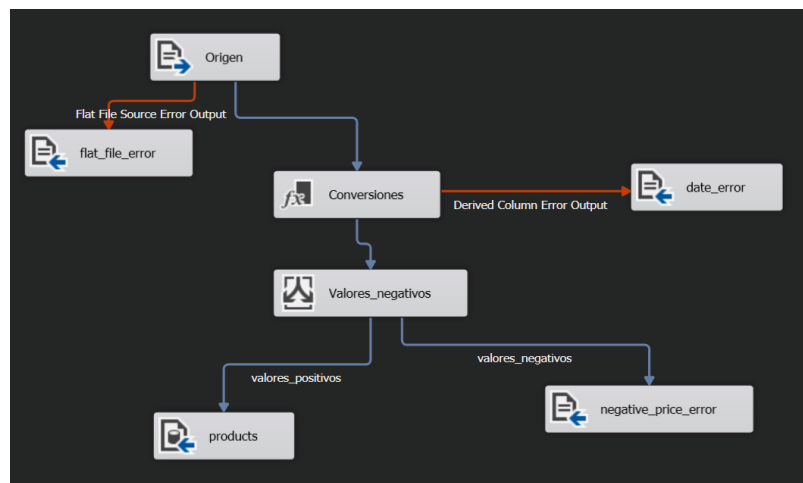
Editamos el Flat destination file



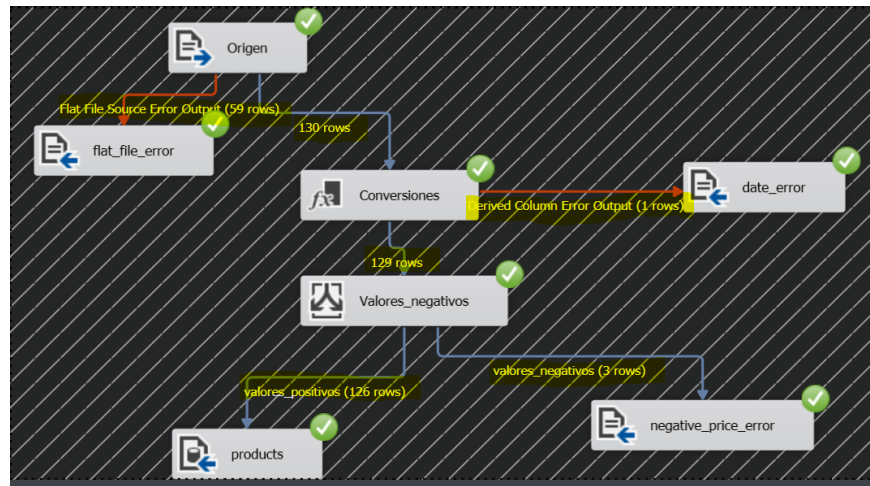
Establecemos el mapeo:



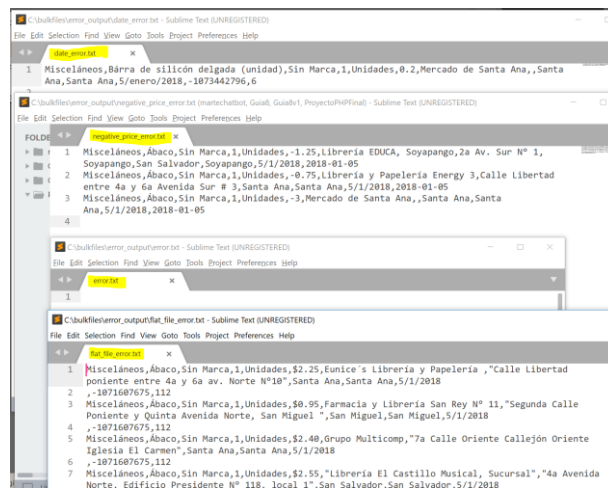
Nuestro diagrama final debe verse de la siguiente forma:



Ejecute el paquete y revise las salidas:



Los archivos de salida presentan los siguientes datos:

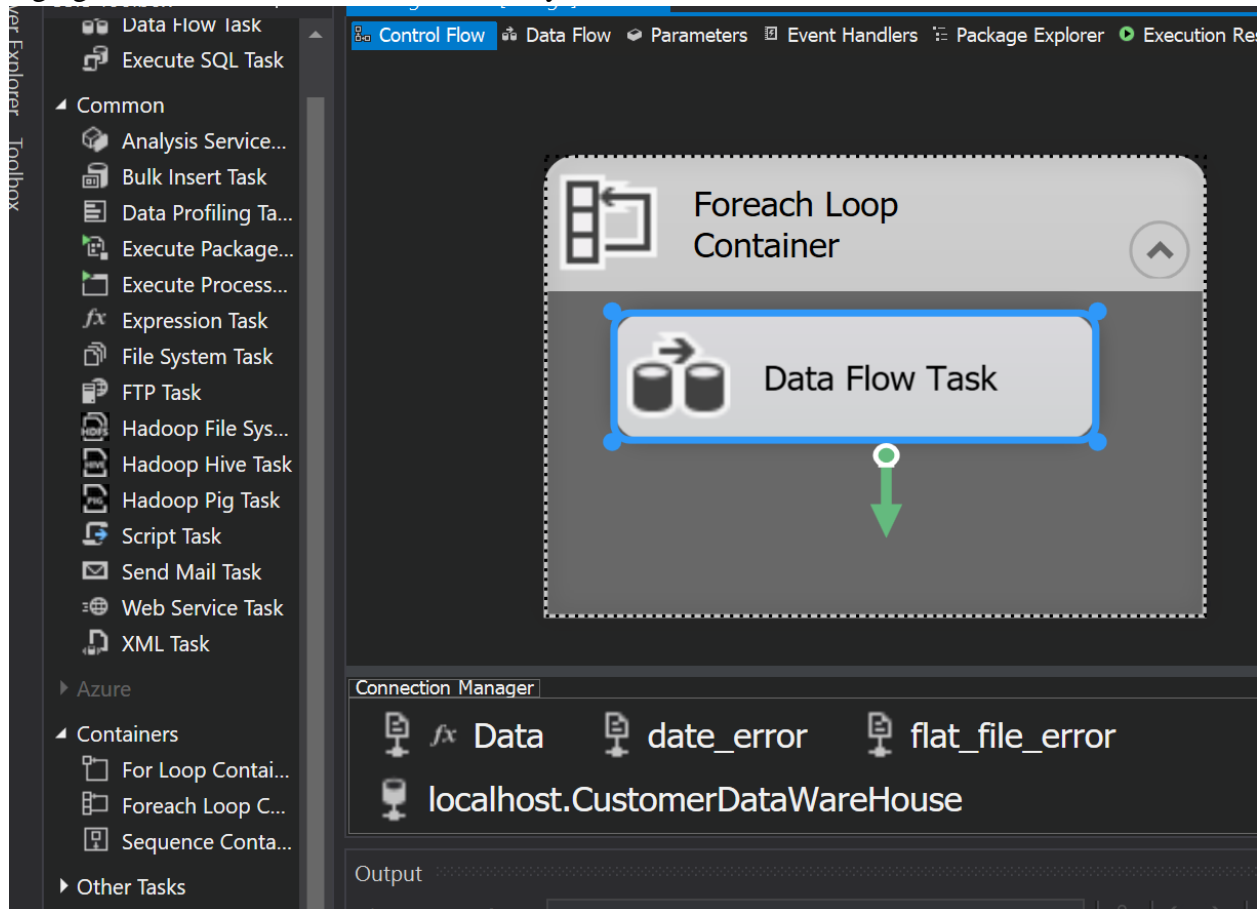


Y en nuestra base de datos debemos encontrar lo siguiente:

Categoria	Producto	Marca	Cantidad	Unidad	Precio	Establecimiento	Direccion	Departamento
Misceláneos	Abaco	Sin Marca	1	Unidades	1.55	Almacén y Librería SOLO	Calle Rubén Darío #115, San Salvador	San Salvador
Misceláneos	Abaco	Sin Marca	1	Unidades	1.25	Bazar La Cebolla, Casa Marz	2a Calle Oriente No 4 Chulucupac	Santa Ana
Misceláneos	Abaco	Sin Marca	1	Unidades	1.00	Farmacia y Librería San Rey N° 23	4a Calle Poniente, local No 207, San Miguel	San Miguel
Misceláneos	Abaco	Sin Marca	1	Unidades	1.00	Librería "Sagrado Corazón"	4a. Avenida Norte N° 4-3	San Salvador
Misceláneos	Abaco	Sin Marca	1	Unidades	2.75	Librería de Papel, 4a Avenida Norte	4a Avenida Norte N° 116	San Salvador
Misceláneos	Abaco	Sin Marca	1	Unidades	2.00	Librería El Ángel Ahuachapán	Avenida Francisco Norberto Norte/Ahuachapán	Ahuachapán
Misceláneos	Abaco	Sin Marca	1	Unidades	1.50	Librería Escolar 1a Calle Oriente	1a Calle Oriente N° 115A	San Salvador
Misceláneos	Abaco	Sin Marca	1	Unidades	1.35	Librería Escolar 4a Avenida Norte	4a Avenida Norte N° 237	San Salvador
Misceláneos	Abaco	Sin Marca	1	Unidades	1.70	Librería Española, Unice	Centro Comercial Unice	San Salvador
Misceláneos	Abaco	Sin Marca	1	Unidades	2.50	Librería Karla Ahuachapán	4a Calle Poniente número 1-14 A Ahuachapán	Ahuachapán
Misceláneos	Abaco	Sin Marca	1	Unidades	1.25	Librería Karla Chulucupac	2a Calle Oriente No 11 Barrio Las Armas	Santa Ana
Misceláneos	Abaco	Sin Marca	1	Unidades	3.25	Librería La Nueva Centroamericana	6a Avenida Norte No 119	San Salvador
Misceláneos	Abaco	Sin Marca	1	Unidades	1.99	Librería Moderna Suc. Paseo	Cond. Balam Outbe, local 15 A.C.	San Salvador
Misceláneos	Abaco	Sin Marca	1	Unidades	1.50	Librería Nuevo Mundo: 1a Calle O.	1a Calle Oriente N° 117	San Salvador
Misceláneos	Abaco	Sin Marca	1	Unidades	1.50	Librería Nuevo Mundo: Plaza Soyapango	Boulevard del Ejército Plaza Soyapango II Local 1	San Salvador
Misceláneos	Abaco	Sin Marca	1	Unidades	1.00	Librería San Judas	diagona calle poniente, Barrio El Calvario	Ahuachapán
Misceláneos	Abaco	Sin Marca	1	Unidades	0.99	Librería San Rey, Sucursal Centro	4a Avenida Norte N° 210	San Salvador
Misceláneos	Abaco	Sin Marca	1	Unidades	1.80	Librería Shadai San Salvador	Cuarta Avenida Norte N° 224, Local 1, San Sal.	San Salvador
Misceláneos	Abaco	Sin Marca	1	Unidades	2.04	Librería Susana	3a Avenida Norte N. 22, Barrio Las Armas	Santa Ana
Misceláneos	Abaco	Sin Marca	1	Unidades	1.51	Librería y Papelería Arend	Paseo Calle Oriente No 127	San Salvador

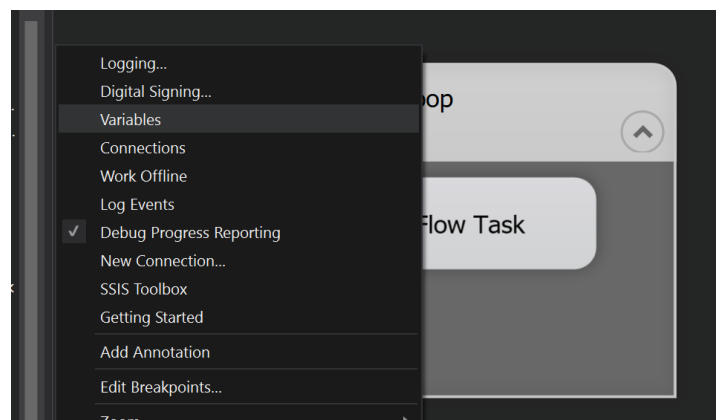
Paso 9: Múltiples ficheros a la vez.

Agregaremos un control en Control Flow y arrastraremos nuestro Data Flow Task dentro de él



Paso 10:

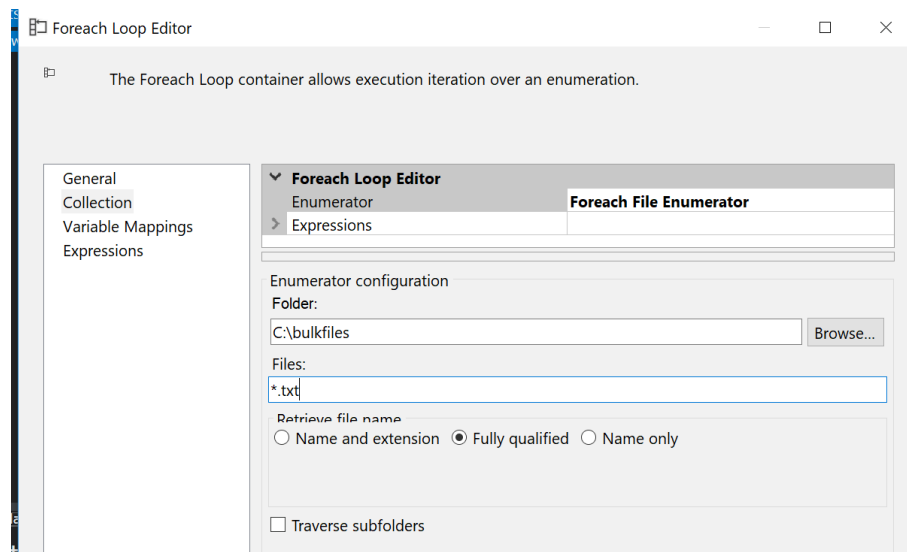
Agregaremos una variable al paquete que servirá como variable temporal del For Loop



Variables				
Name	Scope	Data t...	Value	Expression
bulkfiles	Package	String		

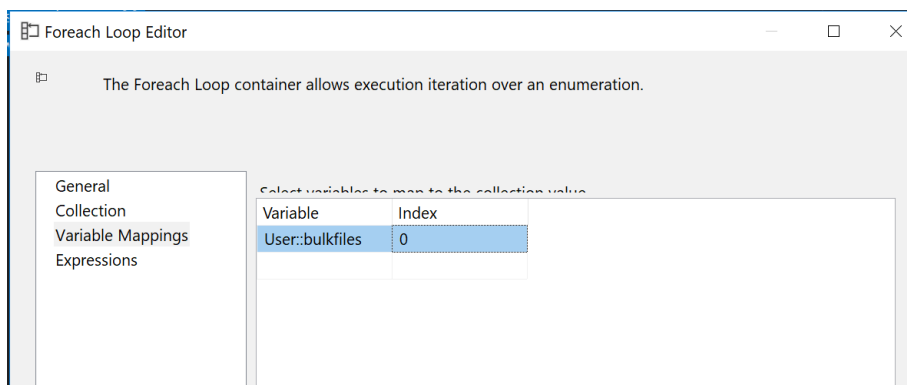
Paso 11:

Edite su Foreach Loop indicando la ruta donde se encuentran los archivos desde la categoría Collection:



Paso 12:

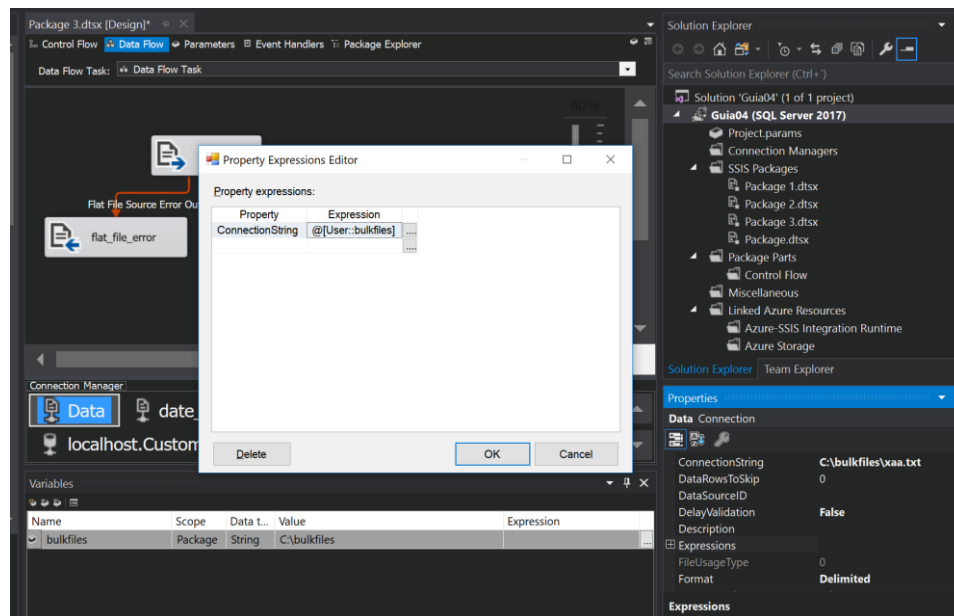
En variable mappings indique que utilizará la variable definida previamente:



Paso 13:

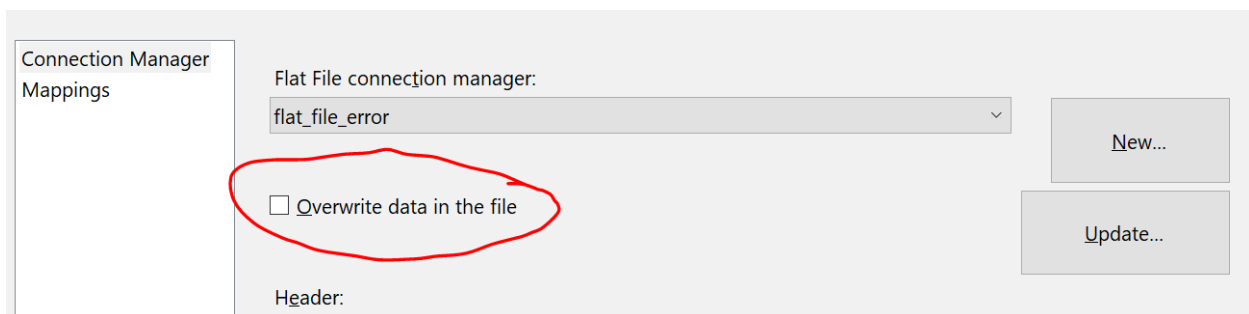
Desde el proyecto en la sección de conexiones, indique que va a cambiar **Data**

Paso 16:



Paso 17:

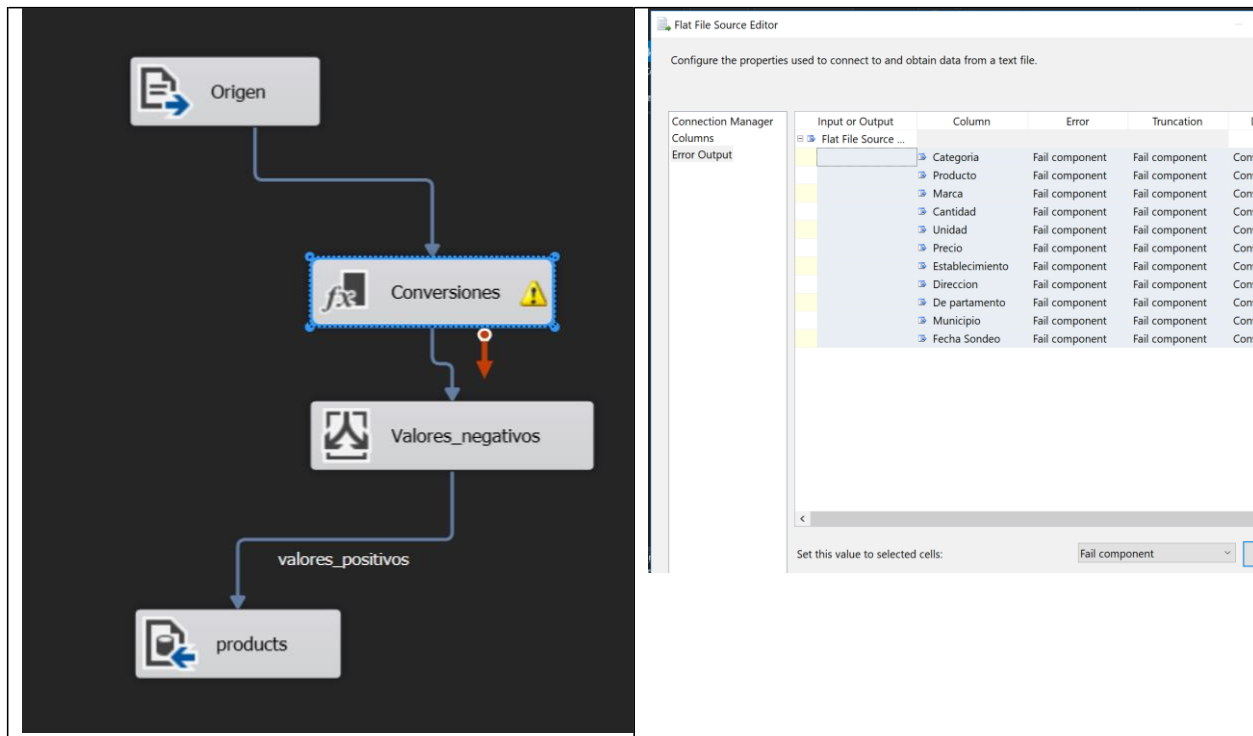
Finalmente en cada flat file vamos a desmarcar la sobre escritura:



Paso 18: Ejecute el paquete. ¿Qué puede apreciarse ahora?

IV. Ejercicios Complementarios

1. Cree un nuevo ETL como se muestra en el diagrama indicando que el Flat File Source “Origen” para en el Error Output no gestione errores como se ve a continuación:



Ejecute el paquete y responda:

- ¿Se ejecuta correctamente el ejercicio?
 - ¿Qué sucede si cambia el error output por Ignore failure?
 - ¿Qué sucede si configura el Error de Conversion para que realice Ignore failure?
2. Al utilizar el For Each Loop Container el flat file error se vuelve muy extenso, lo cual significa que perdemos muchos registros. ¿Puede recuperarse esa DATA?