

Guía de apoyo para proceso de ETL

1. Dependiendo de nuestros archivos de entrada, creamos una base de datos que tenga nuestra tabla o tablas temporales. Y otra para el Datawarehouse. (En este caso de ejemplo por la cantidad de tablas se creará solo una bd).

```
USE [master]
GO

CREATE DATABASE [Ejemplo2]
GO

USE [Ejemplo2]
GO

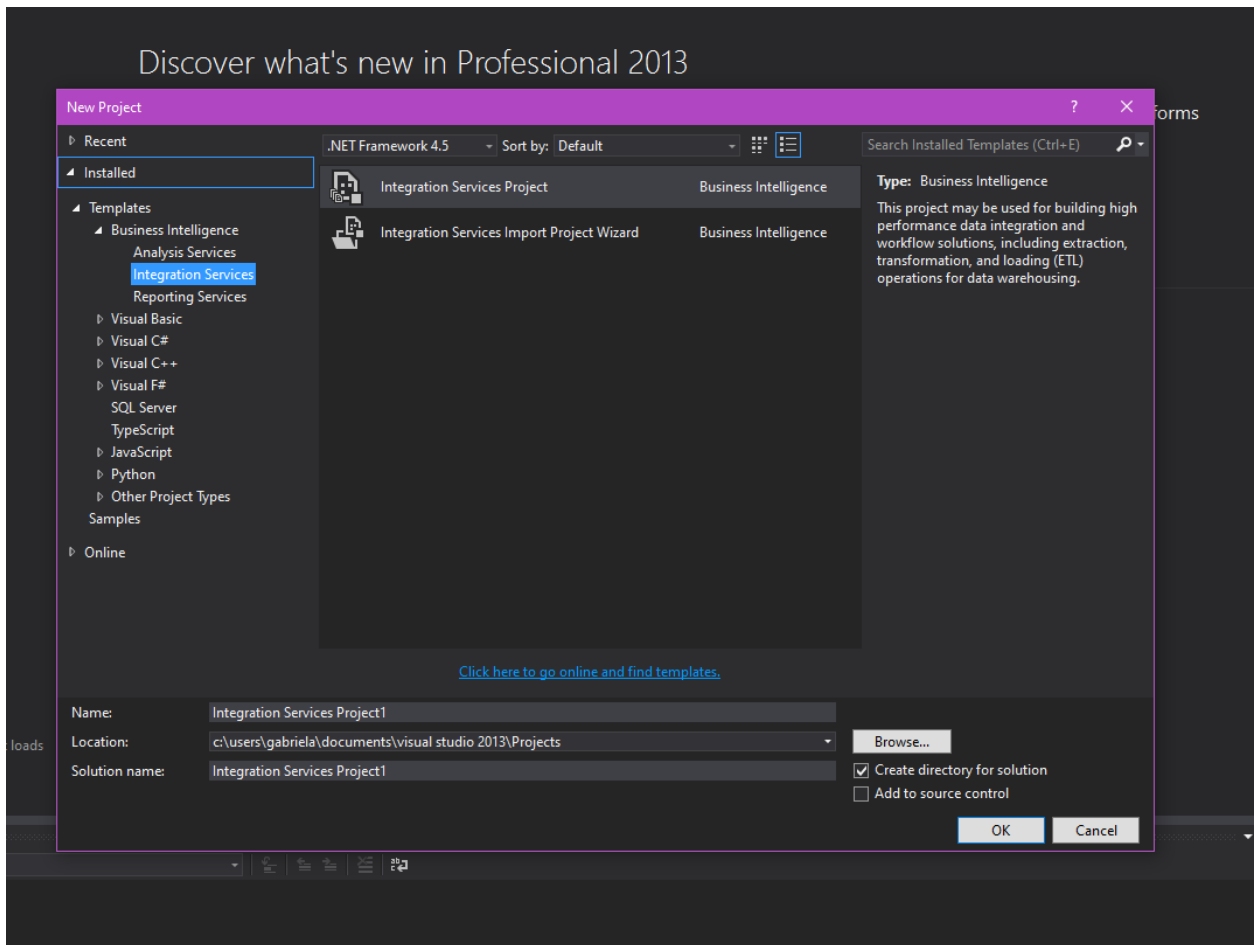
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
SET ANSI_PADDING ON
GO

CREATE TABLE [dbo].[Temporal](
    [CodigoArticulo] [varchar](200) NULL,
    [Descripcion] [varchar](200) NULL,
    [Linea] [varchar](200) NULL,
    [CostoPromedio] [varchar](200) NULL
)

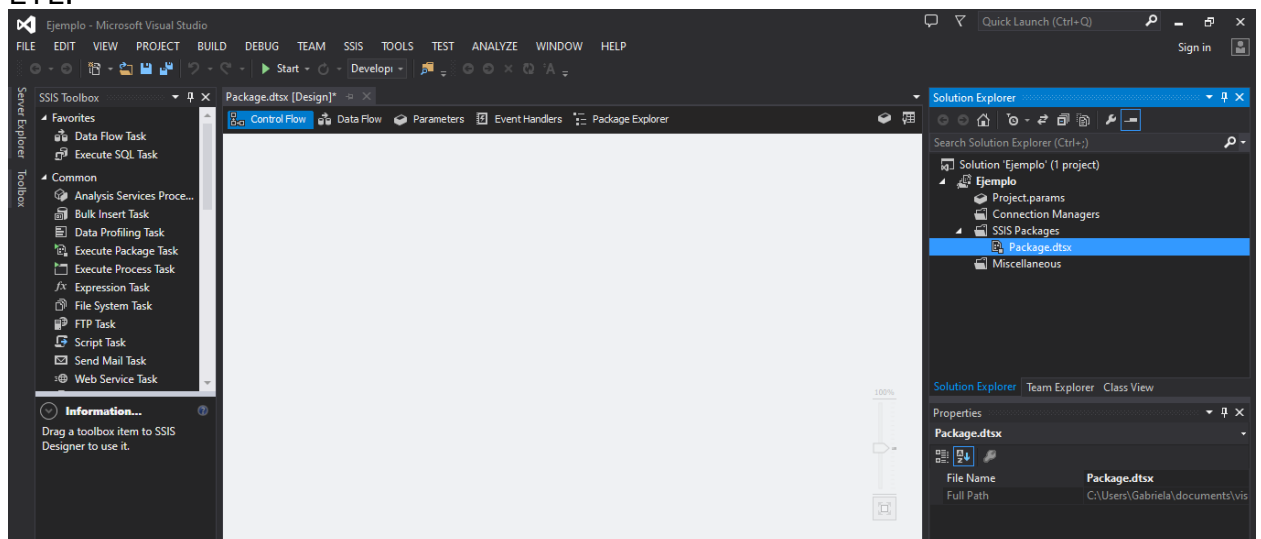
CREATE TABLE [dbo].[Articulos](
    idArticulo int identity Primary Key,
    [CodigoArticulo] [varchar](50) NULL,
    [Descripcion] [varchar](200) NULL,
    [Linea] [varchar](50) NULL,
    [CostoPromedio] decimal(18,2) NULL
)
GO
```

```
1 |CodigoArticulo,Descripcion,Linea,CostoPromedio
2 |AB000030,Refrigeradora,Electrodomesticos,5999
3 |AB000004,Blazer,Ropa de Dama,180
4 |AB000084,ACQUA DI GIO GIORGIO ARMANI DAMA TOILETTE 100,Perfumeria,945
5 |AB000084,ACQUA DI GIO GIORGIO ARMANI DAMA TOILETTE 100,Perfumeria,945
6 |AB000031,Lavadora,Electrodomesticos,3599
7 |AB000042,Laptop DELL,Electronicos,5400
8 |AB000007,Falda corta,Ropa de Dama,60
9 |AB000066,"90210 FOR HER TORAND, A SPELLING ENTERTEIMEN DAMA TOILETTE 100",Perfumeria,-
10 |AB000093,ALIEN THIERRY MUGLER DAMA PERFUME 50,Perfumeria,986
11 |AB000098,AMARIGE GIVENCHY DAMA TOILETTE 100,Perfumeria,656
12 |AB000006,Vestidos,Ropa de Dama,90
13 |AB000085,ACQUA DI GIOIA GIORGIO ARMANI DAMA PERFUME 100,Perfumeria,731
14 |AB000022,Mat de yoga,Deporte,75
15 |AB000051,360 PERRY ELLIS DAMA TOILETTE 100,Perfumeria,390
16 |AB000090,AGUA DE LOEWE ELLA LOEWE DAMA TOILETTE 100,Perfumeria,731
17 |AB000071,"90210 METAL JEANS TORAND, A SPELLING ENTERTEIMEN DAMA PERFUME 100",Perfumeria,266
18 |AB000006,Vestidos,Ropa de Dama,90
19 |AB000057,360 PURPLE PERRY ELLIS DAMA PERFUME 100,Perfumeria,495
20 |AB000075,"90210 SPORT TORAND, A SPELLING ENTERTEIMEN DAMA TOILETTE 100",Perfumeria,226
21 |AB000075,"90210 SPORT TORAND, A SPELLING ENTERTEIMEN DAMA TOILETTE 100",Perfumeria,226
22 |AB000075,"90210 SPORT TORAND, A SPELLING ENTERTEIMEN DAMA TOILETTE 100",Perfumeria,226
23 |AB000005,Pijama,Ropa de Dama,169
24 |,360 RED PERRY ELLIS DAMA PERFUME 100,Perfumeria,391
25 |,ALWAYS ALFRED SUNG DAMA PERFUME 100,Perfumeria,381
26 |AB000078,"90210 TOUCH OF PARADISE TORAND, A SPELLING ENTERTEIMEN DAMA TOILETTE 100",Perfumeria,291
27 |AB000042,Laptop DELL,Electronicos,5400
28 |AB000024,Remadora,Deporte,1350
29 |AB000037,Cama,Muebles,4210
30 |AB000054,360 COLLECTION PERRY ELLIS DAMA PERFUME 100,Perfumeria,571
31 |AB000051,360 PERRY ELLIS DAMA TOILETTE 100,Perfumeria,390
32 |AB000084,ACQUA DI GIO GIORGIO ARMANI DAMA TOILETTE 100,Perfumeria,945
33 |AB000010,Camisa de vestir,Ropa de caballero,150
34 |AB000059,360 WHITE PERRY ELLIS DAMA PERFUME 100,Perfumeria,441
35 |AB000015,Pantalon de lona,Ropa de niños,110
36 |AB000040,Computadora,Electronicos,6550
37 |AB000005,Pijama,Ropa de Dama,169
38 |AB000061,5TH AVENUE AFTER FIVE ELIZABETH ARDEN DAMA PERFUME 125,Perfumeria,375
39 |AB000024,Remadora,Deporte,1350
40 |AB000052,360 BLACK PERRY ELLIS DAMA PERFUME 100,Perfumeria,376
41 |AB000053,360 BLUE PERRY ELLIS DAMA PERFUME 100,Perfumeria,455
42 |AB000005,Pijama,Ropa de Dama,169
43 |AB000048,24K BRILLIANT GOLD MICHAEL KORS DAMA PERFUME 100,Perfumeria,1250
44 |AB000050,273 RED FRED HAYMAN 2S DAMA PERFUME 75,Perfumeria,355
45 |AB000023,Bicicleta estacionaria,Deporte,1575
```

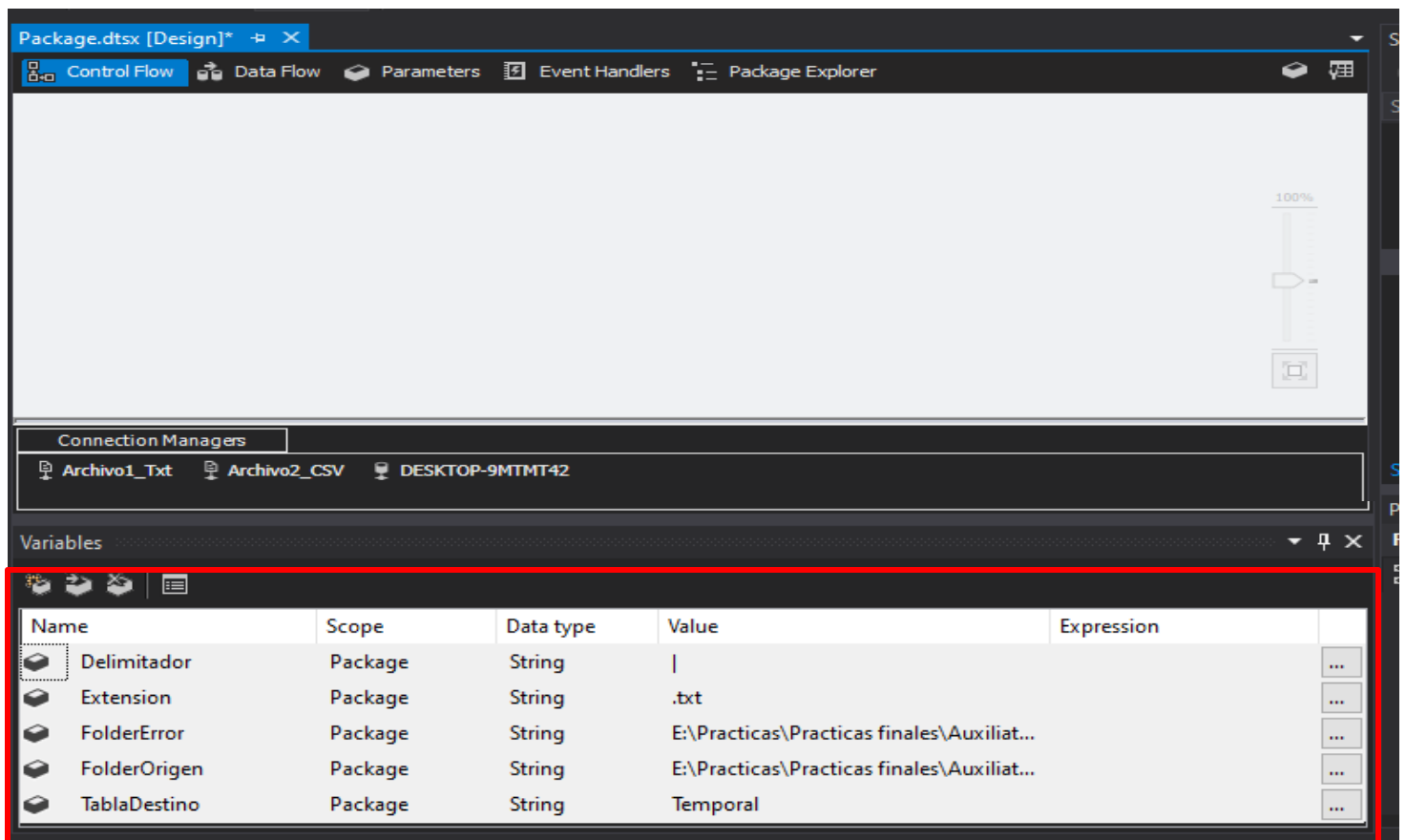
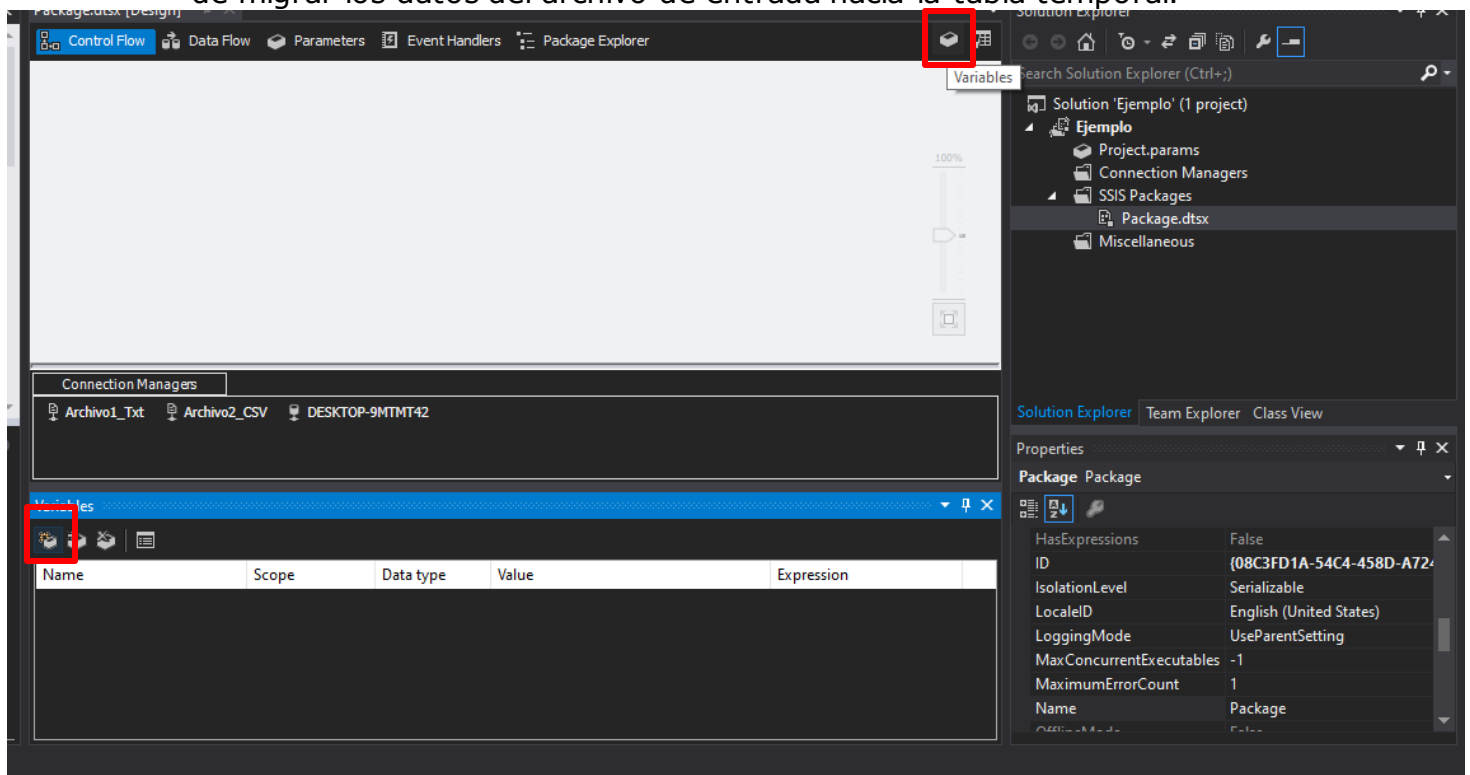
2. Creamos un nuevo proyecto de Integration Services en Visual Studio



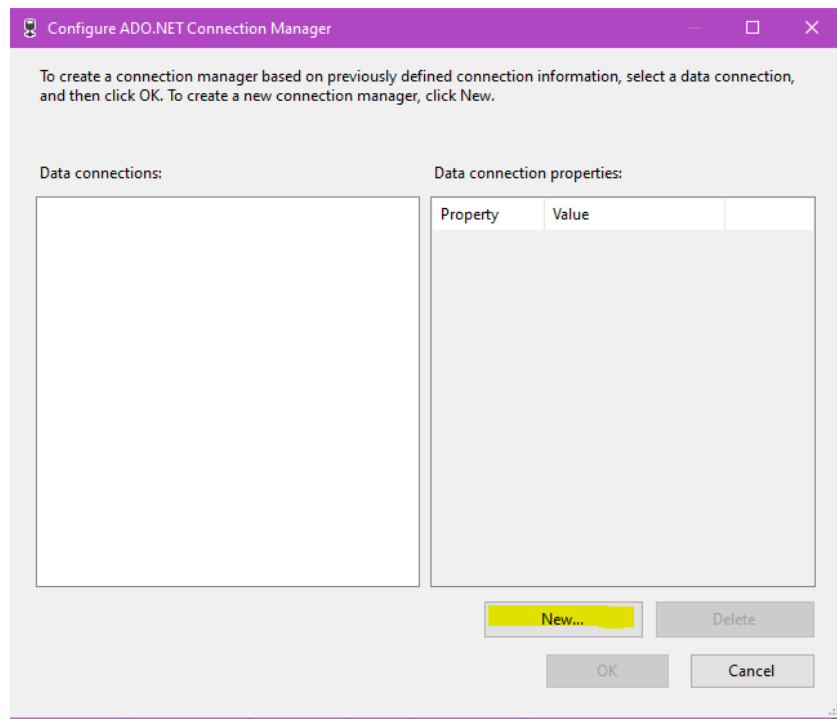
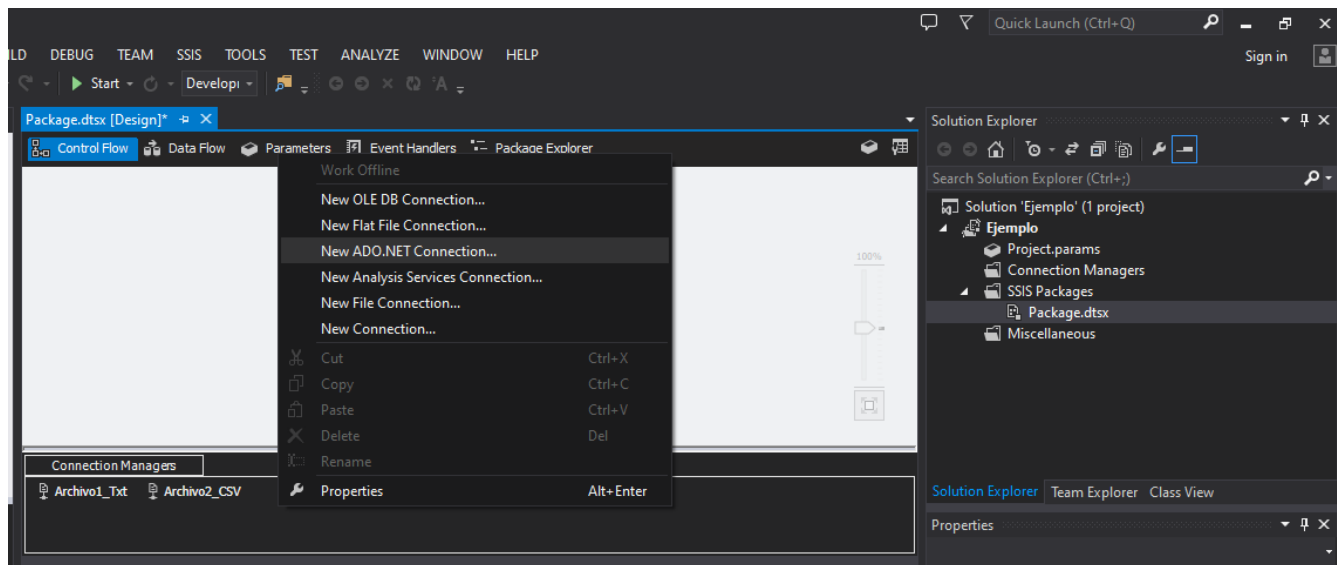
3. Nos dirigimos al archivo Package.dtsx para empezar con nuestro proceso de ETL.

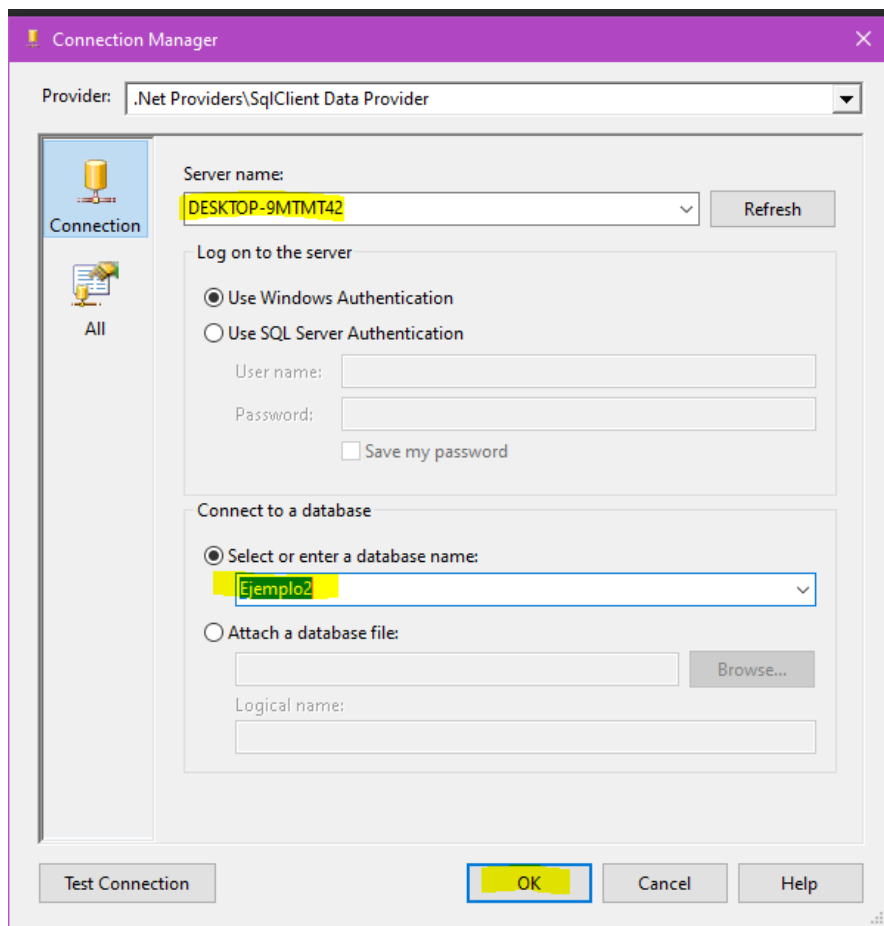
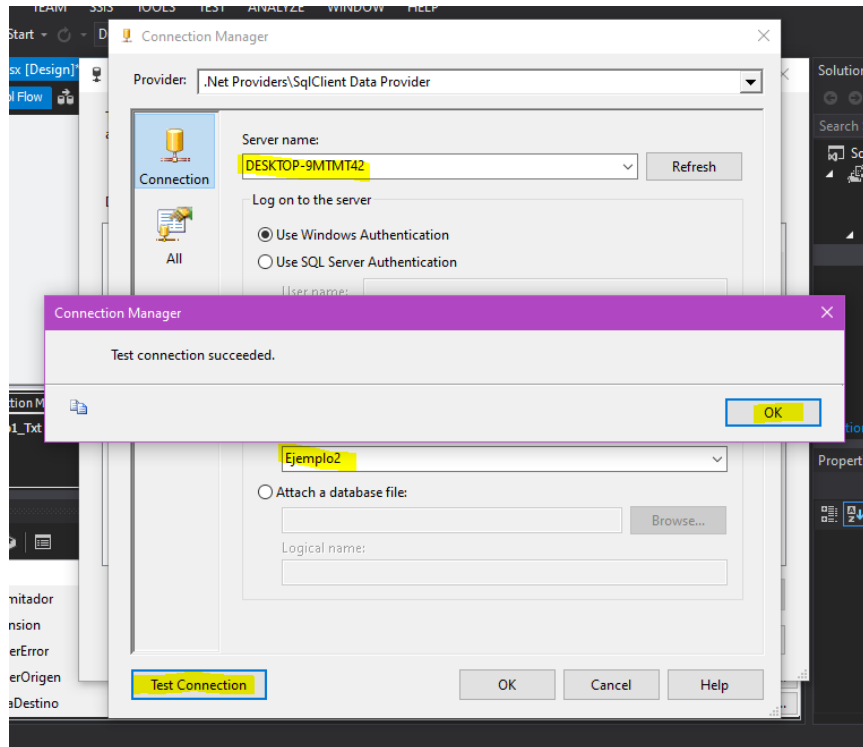


4. Para este ejemplo haremos uso de variables que nos serán útiles al momento de migrar los datos del archivo de entrada hacia la tabla temporal.

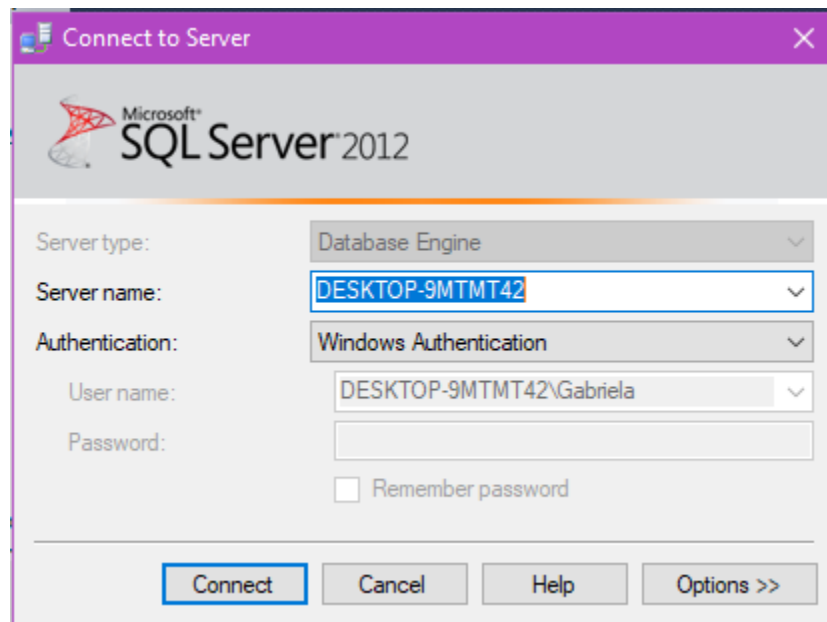


5. En el administrador de conexiones (clic derecho en Connection Managers) agregamos una nueva conexión.

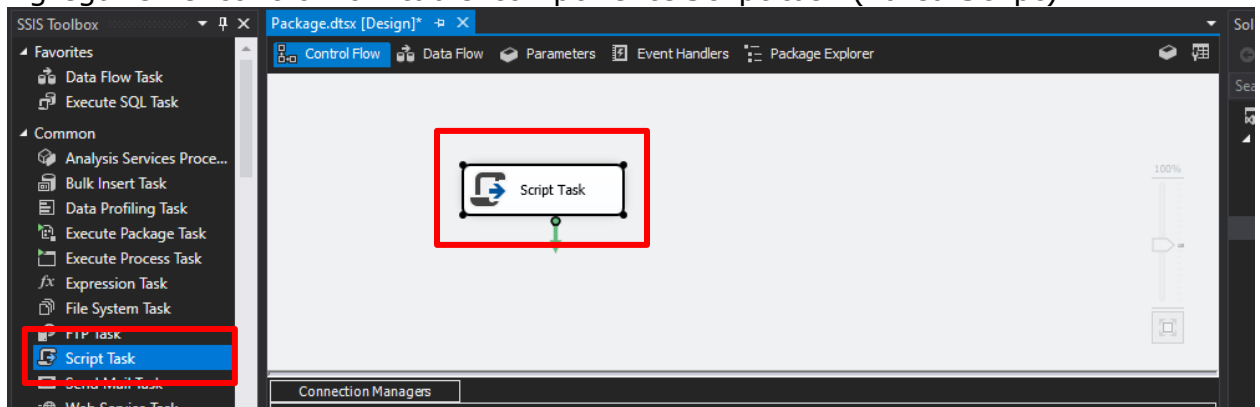




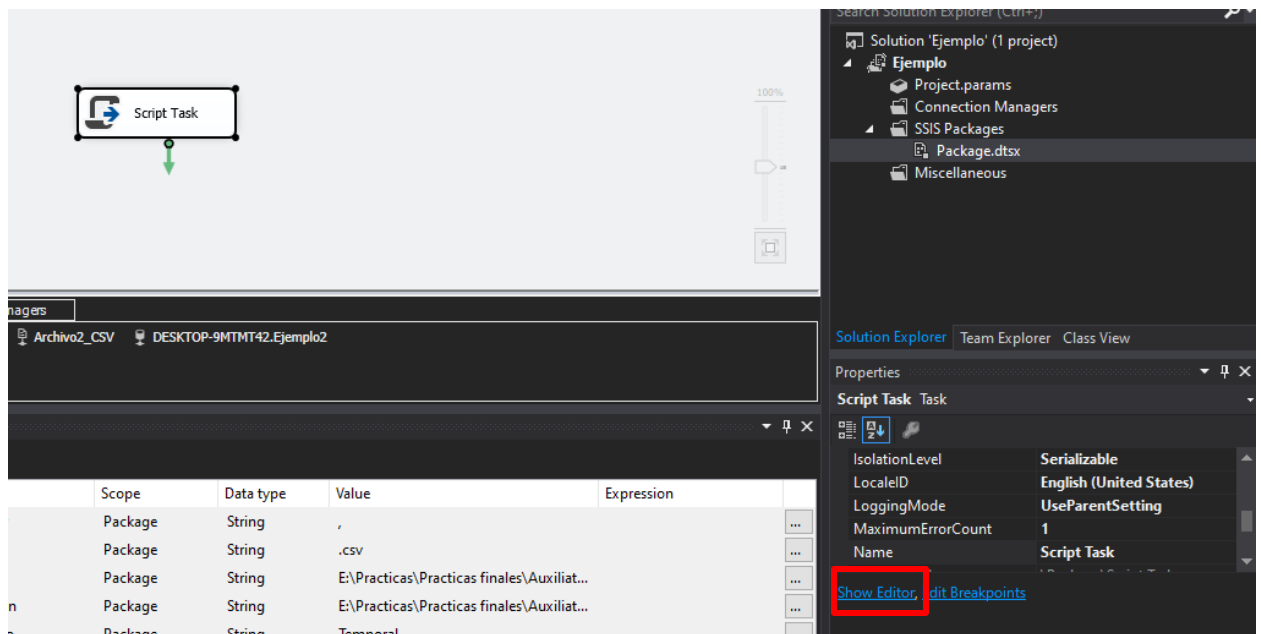
***Observación de no aparecer un servidor disponible deben escribir el nombre que les aparezca en Sql Server.**



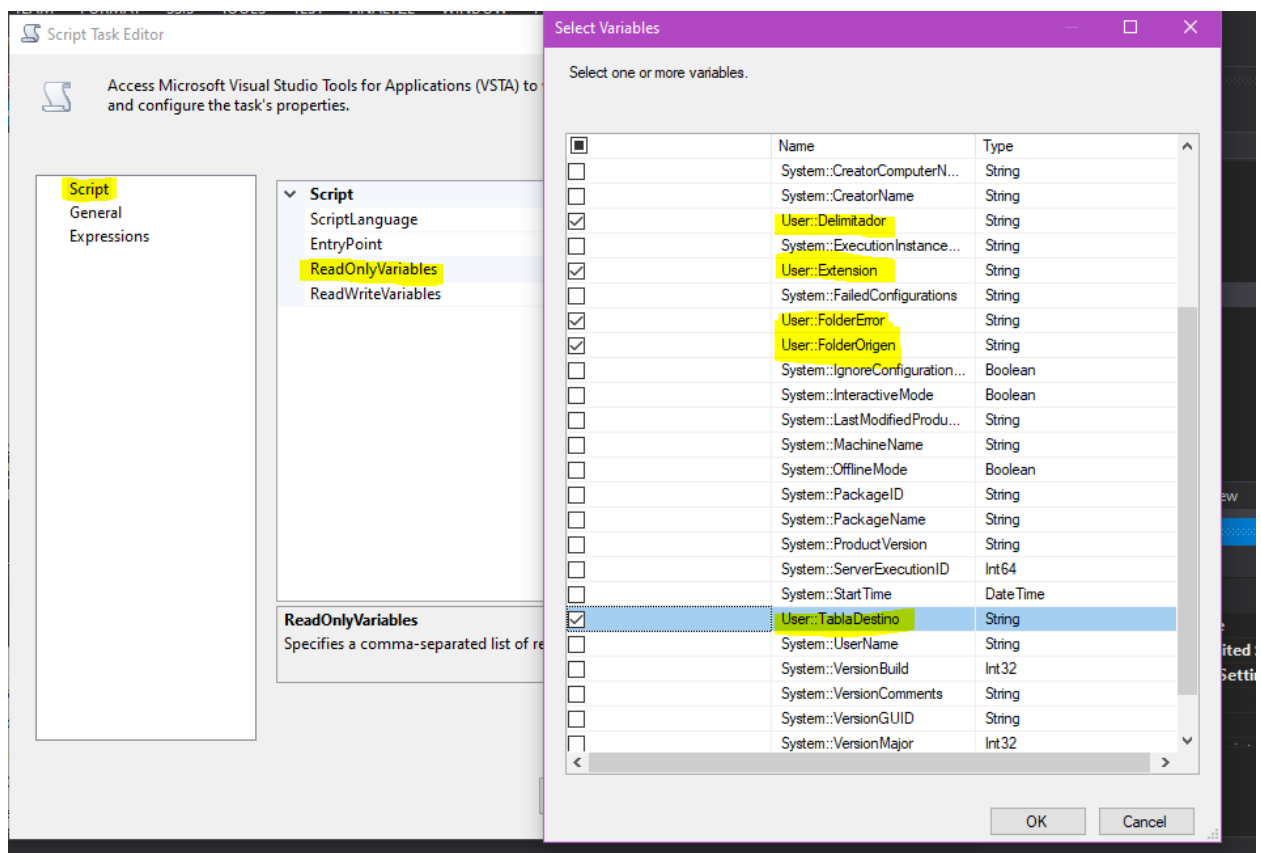
6. Agregar en el control flow tab el componente Script task (Tarea Script)



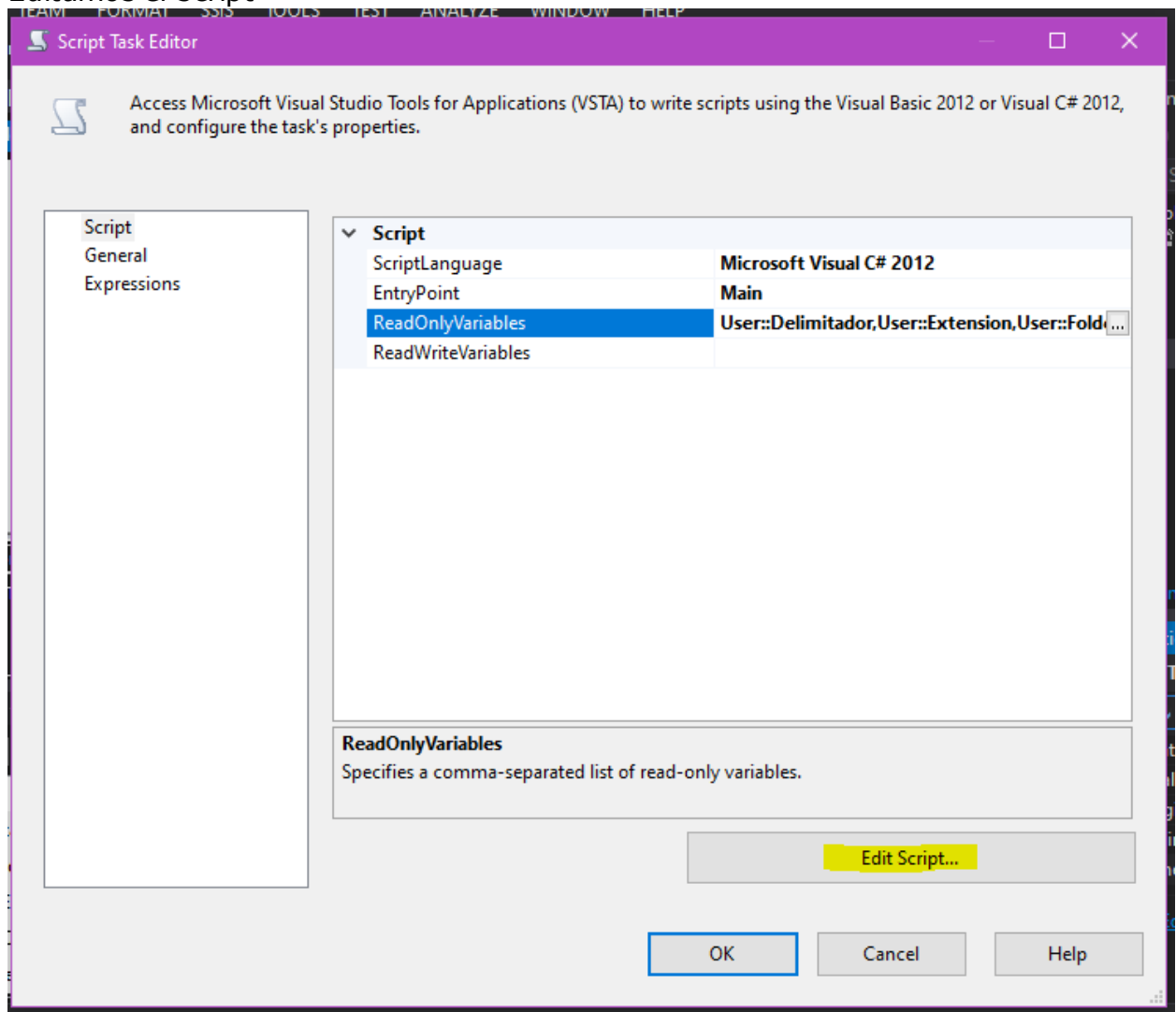
7. Procedemos a configurar el componente Script Task



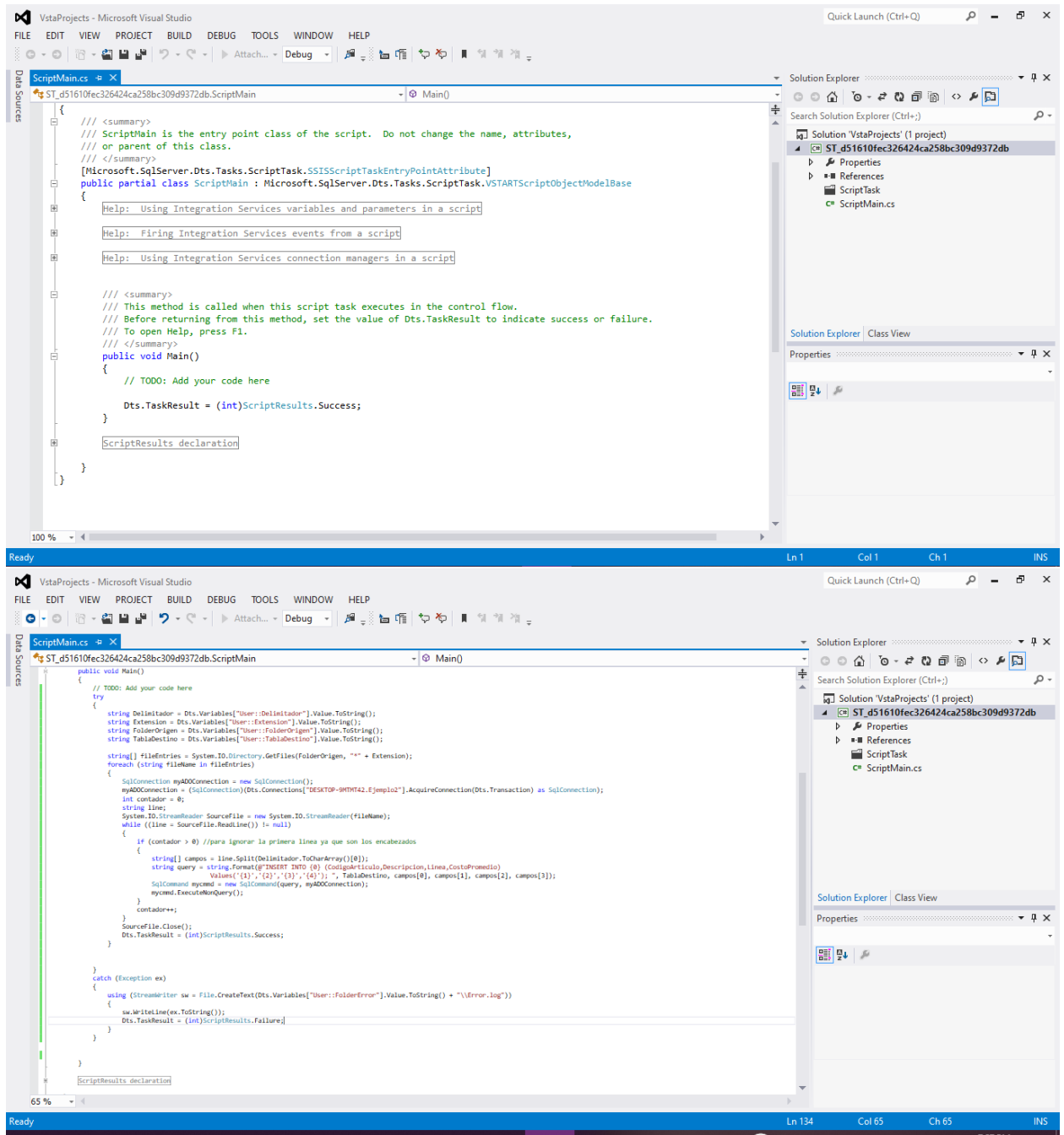
- En la opción de ReadOnlyVariables vamos a seleccionar las variables que creamos



9. Editamos el script



Nos abrirá una nueva ventana de visual studio con una clase previamente creada, la cual procederemos a modificar agregando el código necesario para la migración o carga de los datos del archivo a nuestra tabla temporal.



10. Ejecutaremos nuestro proceso y podremos observar como nuestra tabla temporal contiene toda la data que hemos insertado.

Ejemplo (Running) - Microsoft Visual Studio

FILE EDIT VIEW PROJECT BUILD DEBUG TEAM FORMAT SSIS TOOLS TEST ANALYZE WINDOW HELP

Process: [840] DtsDebugHost.exe

Package.dtsx [Design] - X

Control Flow Data Flow Parameters Event Handlers Package Explorer Progress

Script Task

SQLQuery1.sql - DE...MT42\Gabriela (55)* CreacionBD.sql - D...MT42\Gabriela (51)

```
select * from temporal
```

100 %

Results Messages

	CodigoArticulo	Descripcion	Linea	CostoPromedio
1	AB000030	Refrigeradora	Electrodomesticos	5999
2	AB000004	Blazer	Ropa de Dama	180
3	AB000084	ACQUA DI GIO GIORGIO ARMANI DAMA TOILETTE 100	Perfumeria	945
4	AB000084	ACQUA DI GIO GIORGIO ARMANI DAMA TOILETTE 100	Perfumeria	945
5	AB000031	Lavadora	Electrodomesticos	3599
6	AB000042	Laptop DELL	Electronicos	5400
7	AB000007	Falda corta	Ropa de Dama	60
8	AB000066	"90210 FOR HER TORAND	A SPELLING ENTERTEIMEN DAMA TOILETTE 100"	Perfumeria
9	AB000093	ALIEN THIERRY MUGLER DAMA PERFUME 50	Perfumeria	986
10	AB000098	AMARIGE GIVENCHY DAMA TOILETTE 100	Perfumeria	656
11	AB000006	Vestidos	Ropa de Dama	90
12	AB000085	ACQUA DI GIOIA GIORGIO ARMANI DAMA PERFUME 100	Perfumeria	731
13	AB000022	Mat de yoga	Deporte	75
14	AB000051	360 PERRY ELLIS DAMA TOILETTE 100	Perfumeria	390
15	AB000090	AGUA DE LOEWE ELLA LOEWE DAMA TOILETTE 100	Perfumeria	731
16	AB000071	"90210 METAL JEANS TORAND	A SPELLING ENTERTEIMEN DAMA PERFUME 100"	Perfumeria
17	AB000006	Vestidos	Ropa de Dama	90
18	AB000057	360 PURPLE PERRY ELLIS DAMA PERFUME 100	Perfumeria	495
19	AB000075	"90210 SPORT TORAND	A SPELLING ENTERTEIMEN DAMA TOILETTE 100"	Perfumeria
20	AB000075	"90210 SPORT TORAND	A SPELLING ENTERTEIMEN DAMA TOILETTE 100"	Perfumeria

Query executed successfully.

DESKTOP-9MTMT42 (11.0 SP1) | DESKTOP-9MTMT42\Gabrie... | Ejemplo2 | 00:00:00 | 200 rows

11. Borramos los datos temporales para que no se dupliquen cuando hagamos pruebas con los dos componentes agregados.

12. Agregamos un componente Data flow task

SSIS Toolbox

Package.dtsx [Design]* - X

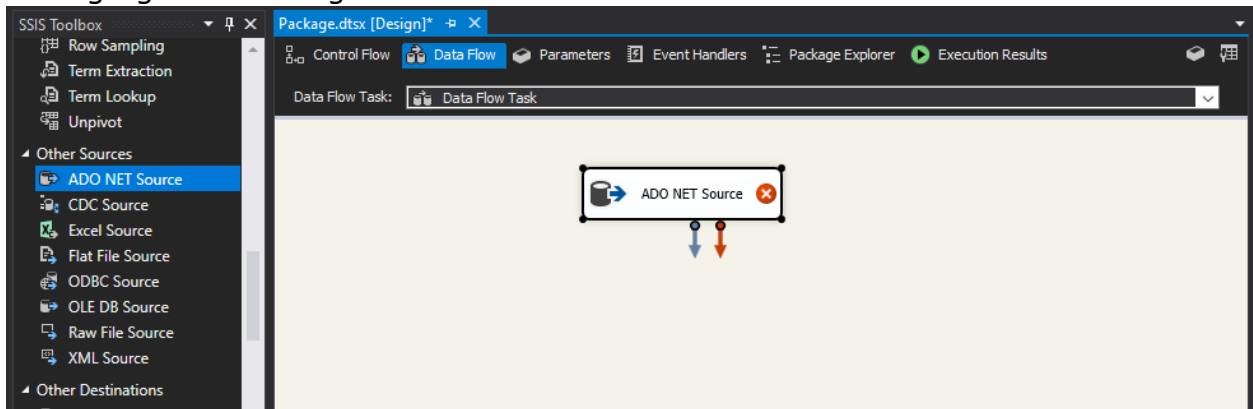
Control Flow Data Flow Parameters Event Handlers Package Explorer Progress

Script Task

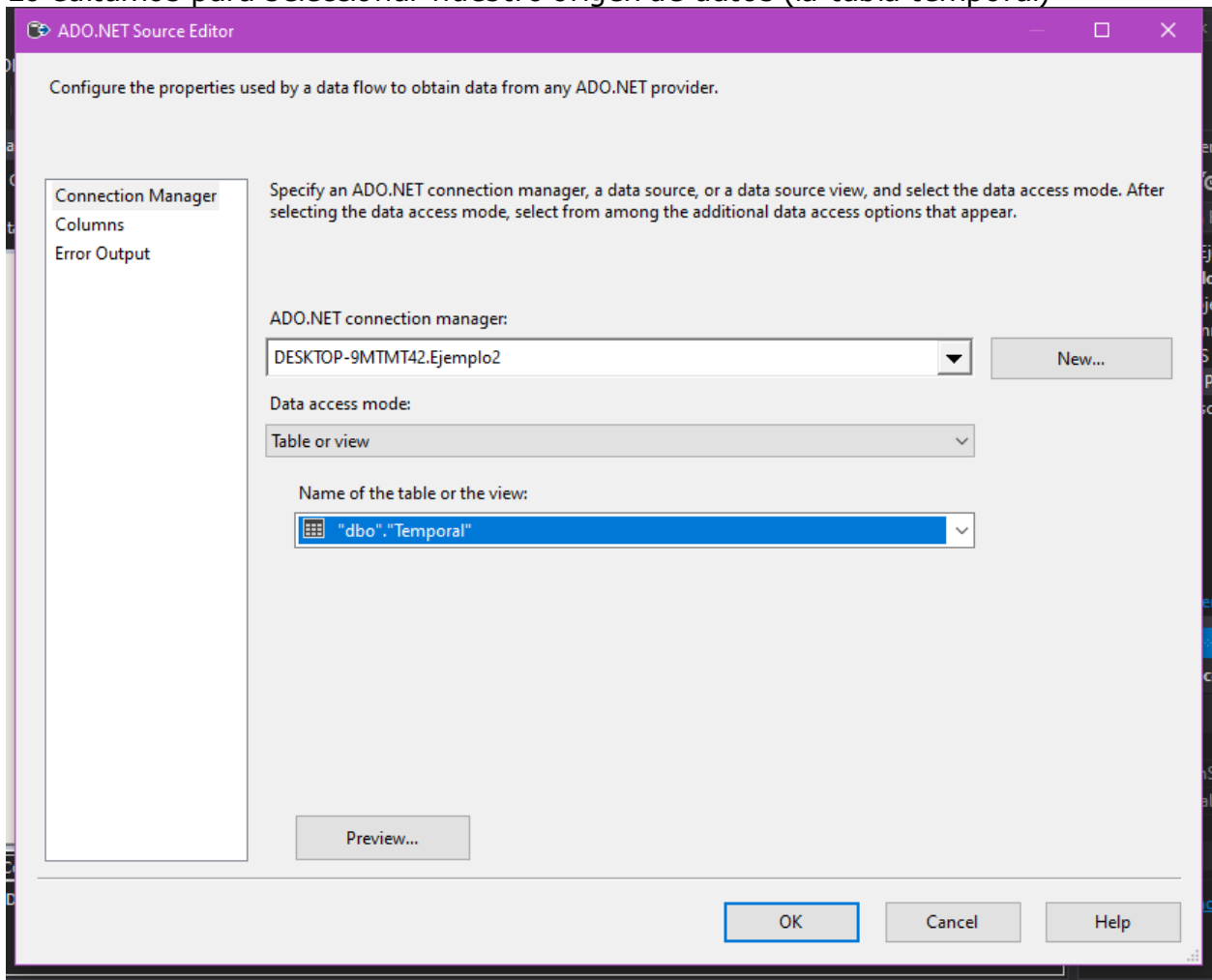
Data Flow Task

Connection Managers

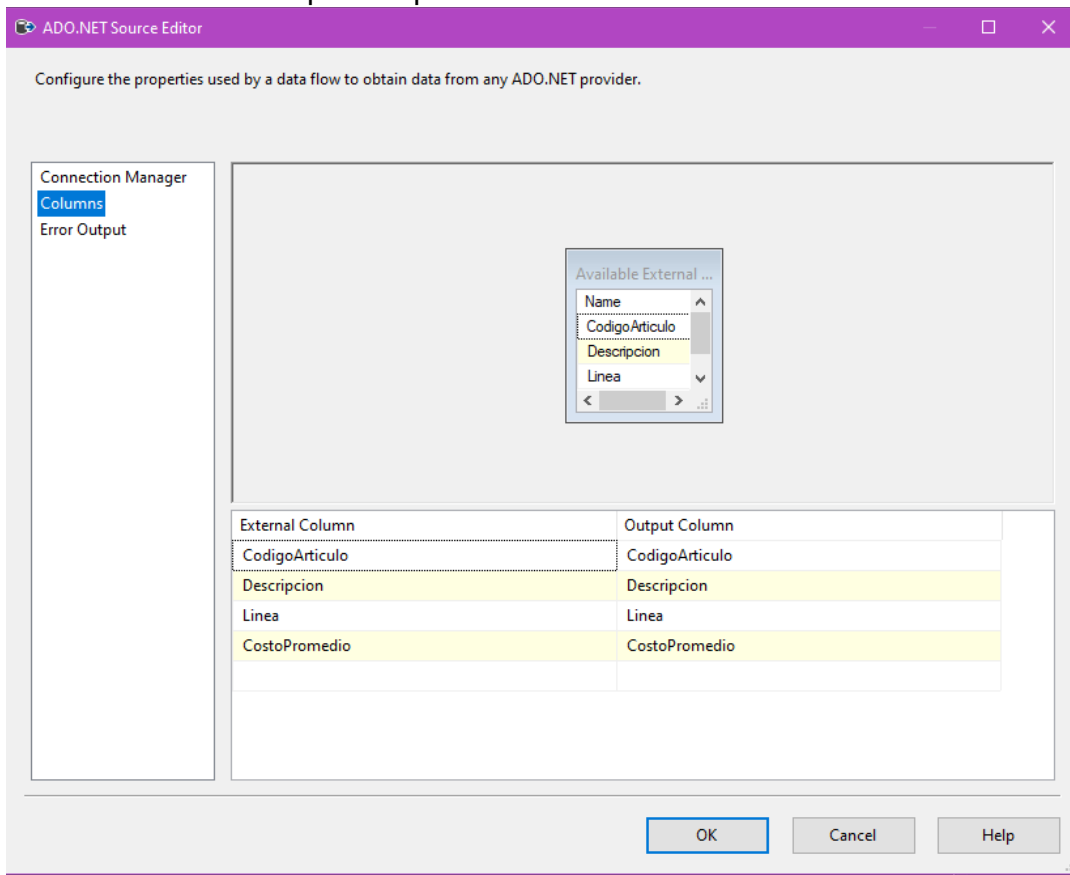
13. Editamos el Data flow task, para agregar los componentes necesarios para el proceso de ETL.
14. Agregamos un origen de datos ADO.Net



Lo editamos para seleccionar nuestro origen de datos (la tabla temporal)



Y validamos los campos disponibles



15. Agregamos un componente Data Conversion ya que como nos podemos dar cuenta la tabla temporal todos sus valores son varchar y nuestra tabla destino tiene un campo decimal.

Con la siguiente configuración estaremos agregando una nueva columna con las conversiones realizadas

Data Conversion Transformation Editor

Configure the properties used to convert the data type of an input column to a different data type. Depending on the data type to which the column is converted, set the length, precision, scale, and code page of the column.

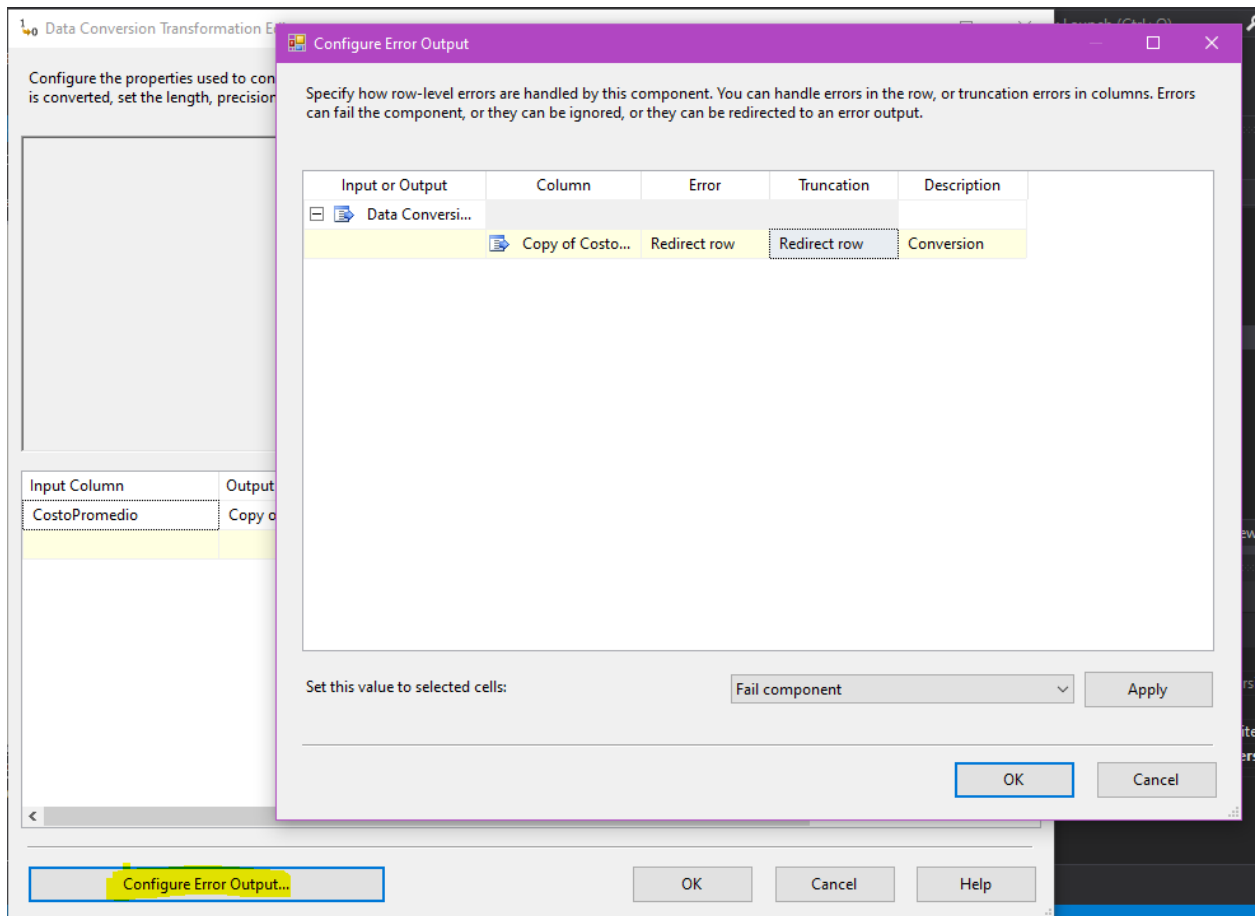
Available Input Columns

<input checked="" type="checkbox"/>	Name
<input type="checkbox"/>	CodigoArticulo
<input type="checkbox"/>	Descripcion
<input type="checkbox"/>	Linea
<input checked="" type="checkbox"/>	CostoPromedio

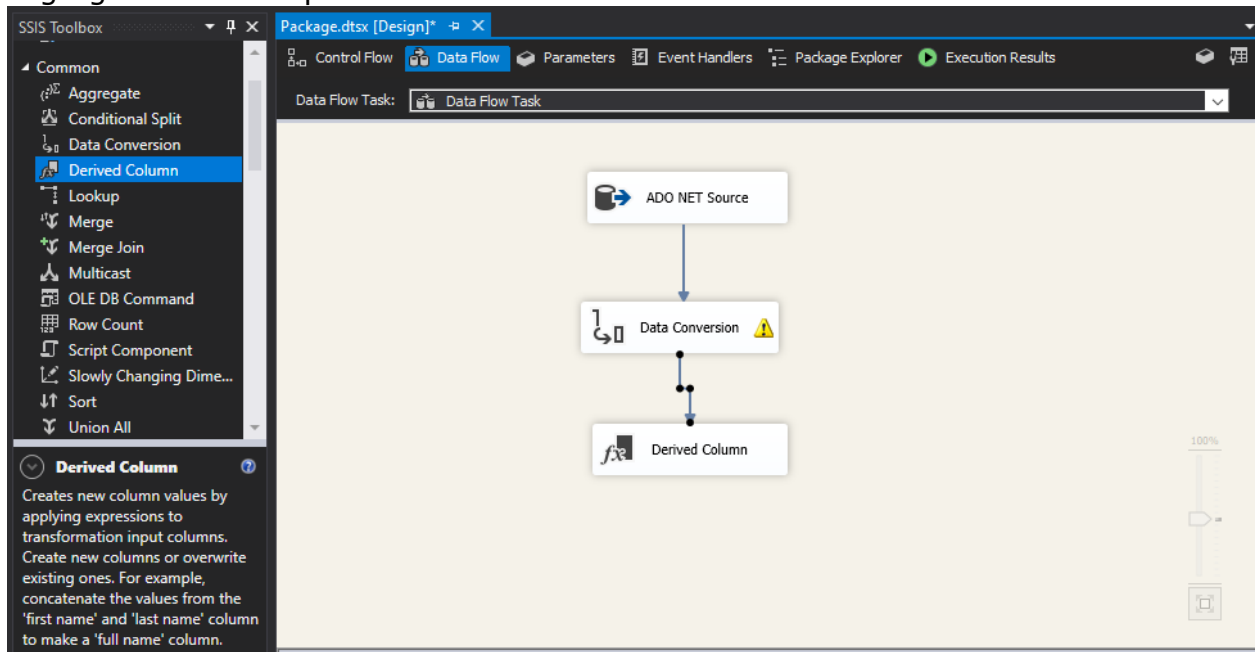
Input Column	Output Alias	Data Type	Length	Precision	Scale	Code Page
CostoPromedio	Copy of CostoPromedio	decimal [DT_DECIMAL]				

Configure Error Output... OK Cancel Help

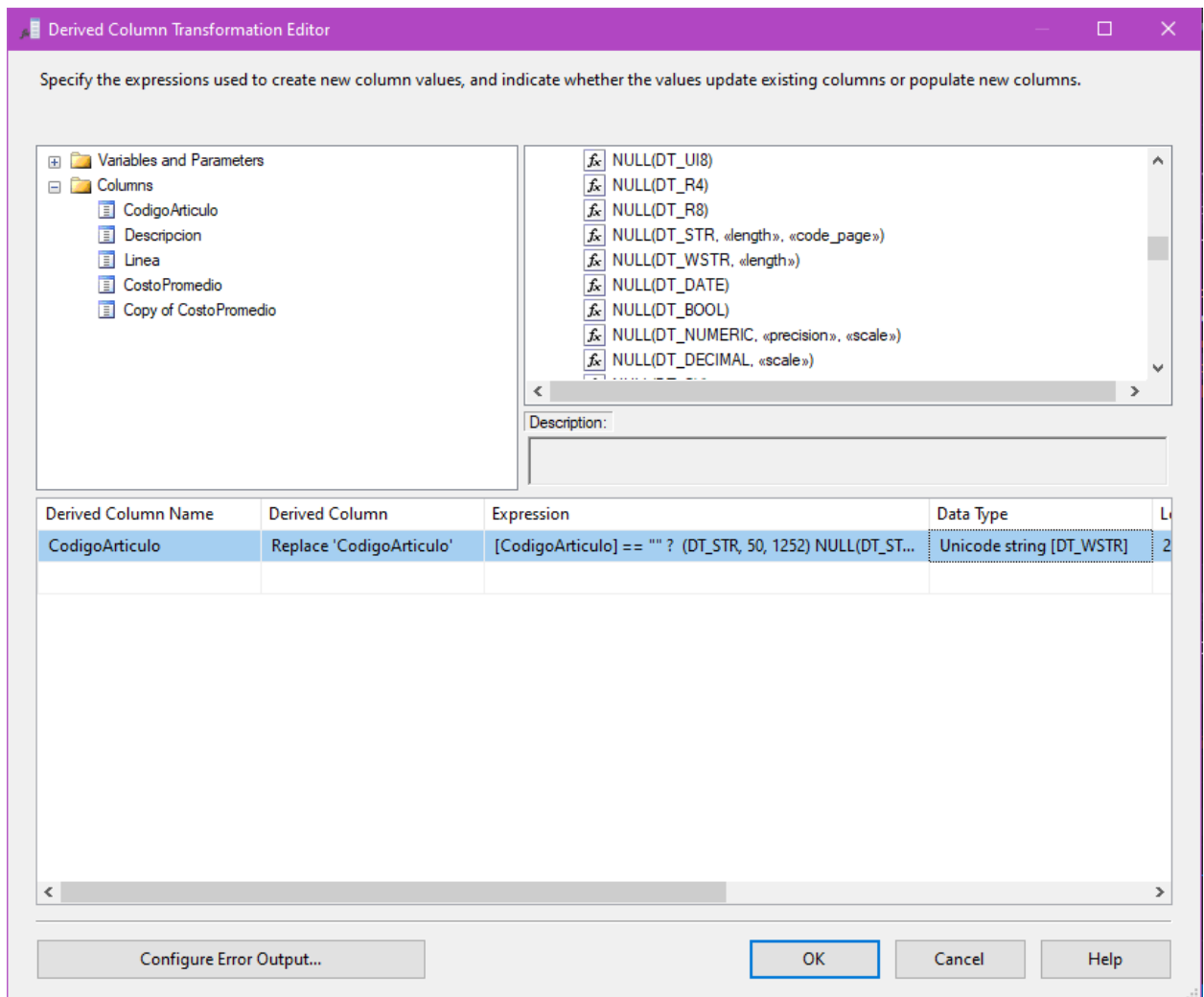
Configuramos la salida de errores, esto para que al suceder un error este sea redirigido y no pase a nuestros datos a cargar.



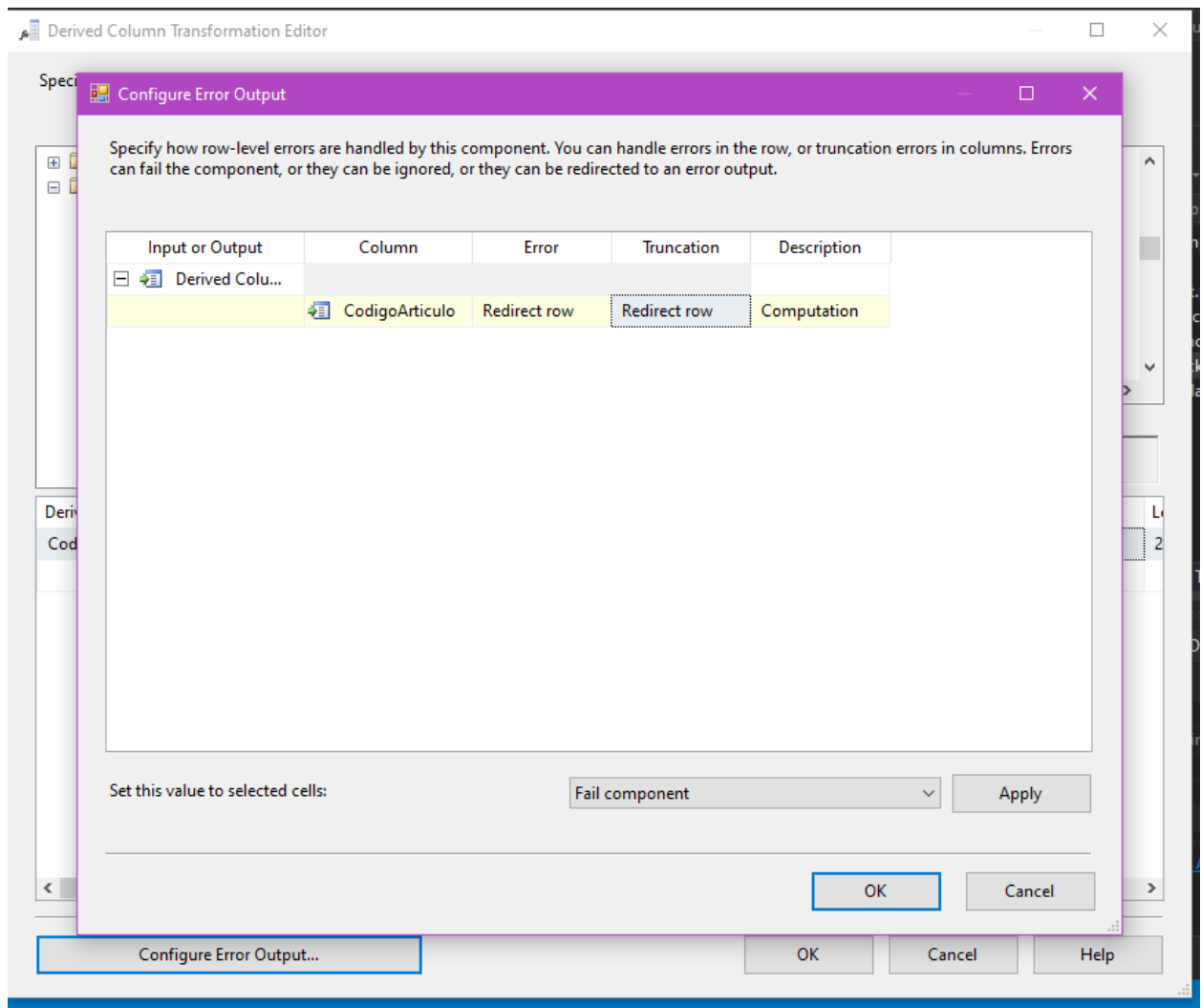
16. Agregamos un componente de columna derivada



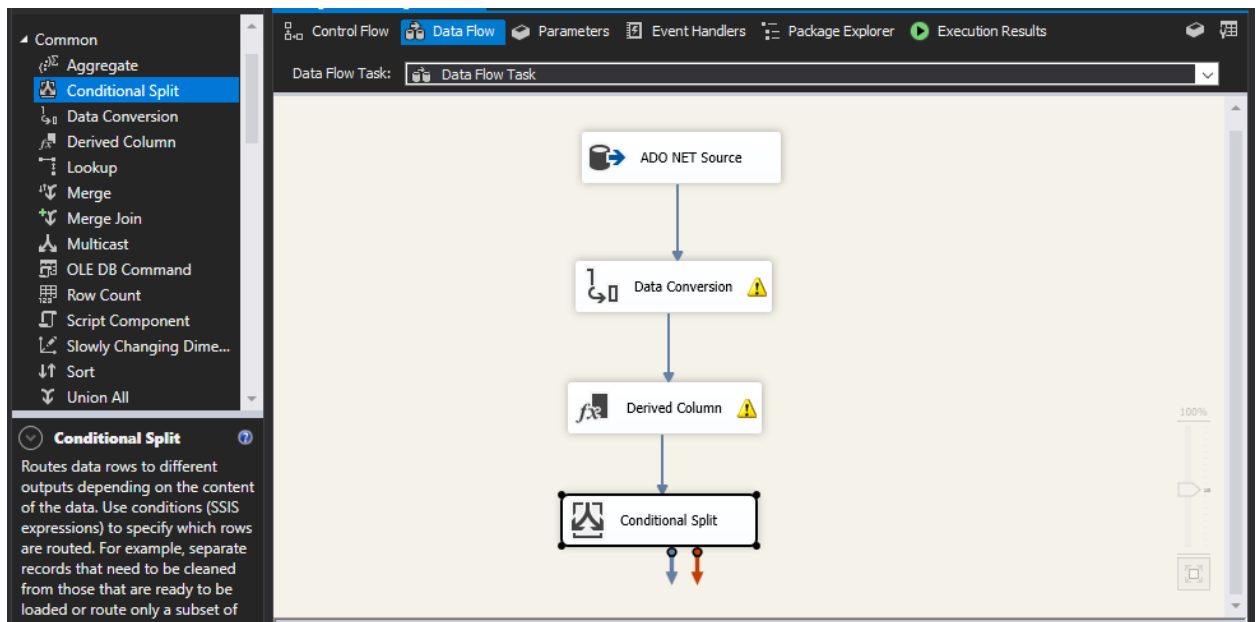
Configuramos el nuevo componente de la siguiente manera, esto para convertir un espacio Vacío en NULL y poder hacer una validación de este en el siguiente paso.



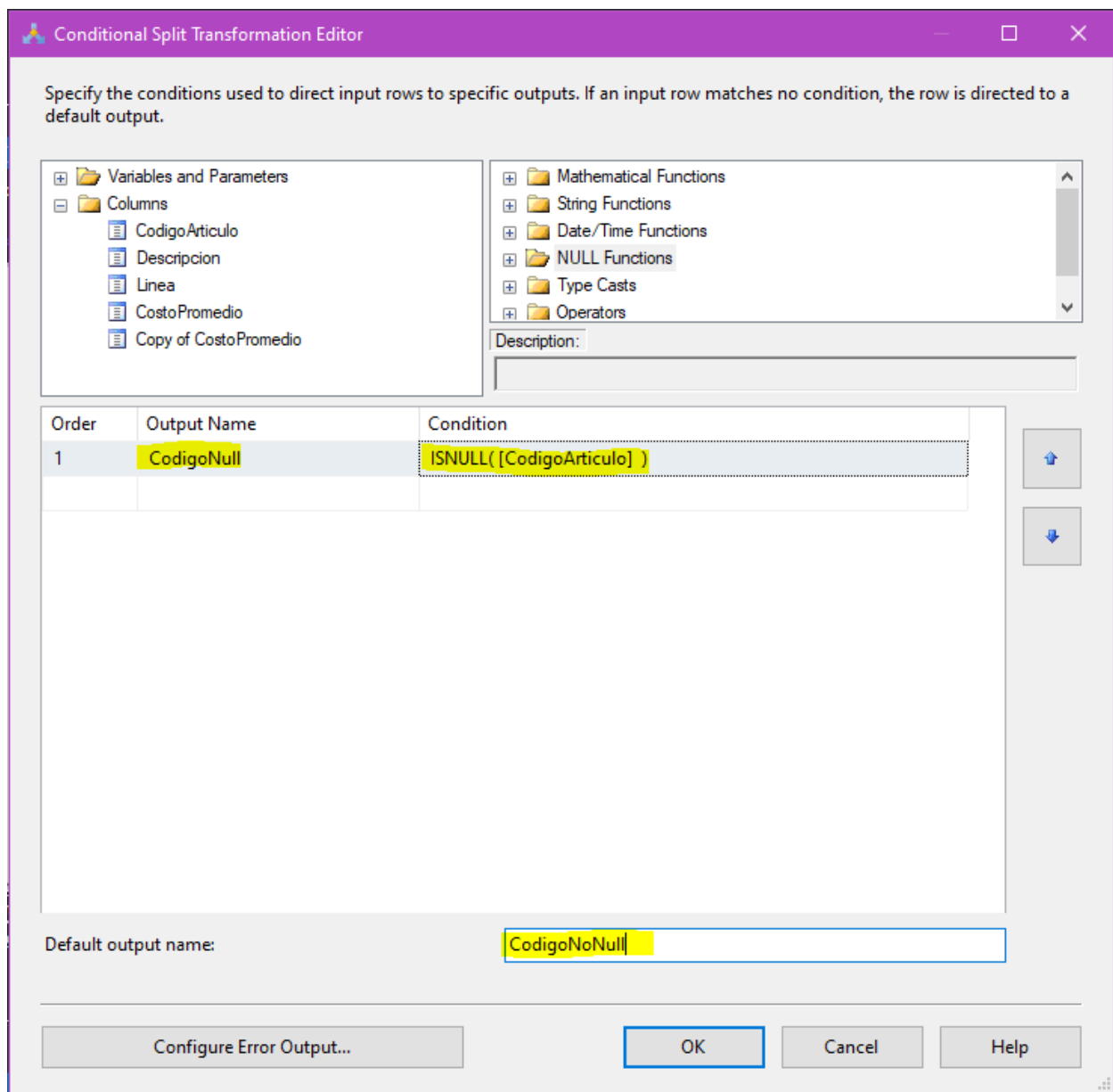
Nuevamente redirigimos filas de existir errores.



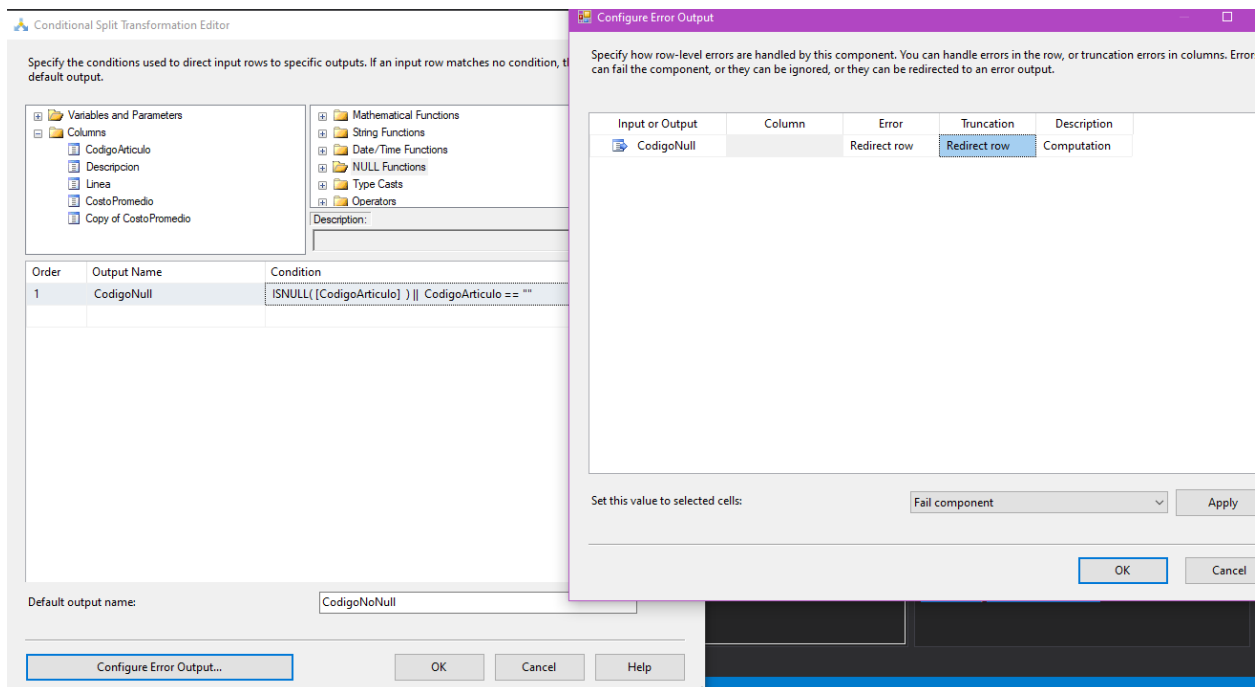
17. Agregamos el componente conditional Split



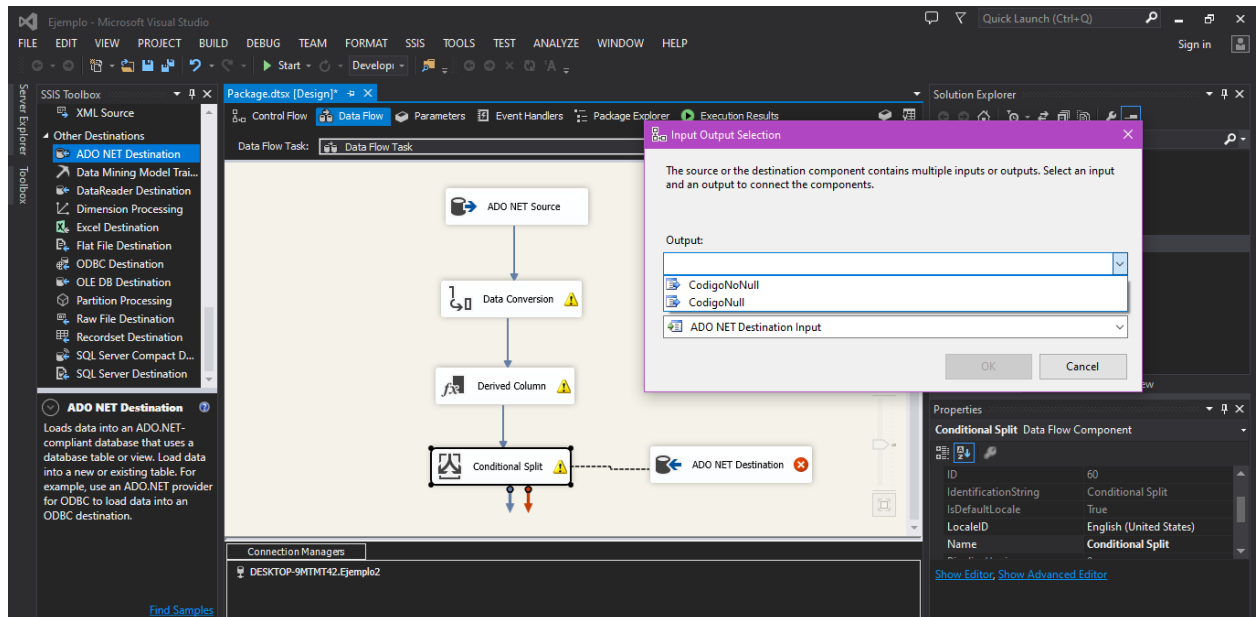
Esto para separar los datos con valores nulos de los no nulos, ya que no deseamos insertar en nuestra tabla artículos con códigos nulos.



Otra forma de quitar los campos en blanco es la siguiente.



18. Agregamos por último un destino ADO NET, el cual conectamos con lo que hemos trabajado y la entrada de este será CodigoNotNull que son las filas correctas en nuestra transformación.



Editamos nuestro componente de destino
 Seleccionamos nuestra base de datos y tabla destino

ADO.NET Destination Editor

Configure the properties used to insert data into a destination using ADO.NET provider.

Connection Manager
Mappings
Error Output


Specify a connection manager, data source, or data source view, and select the table or the view into which the data is copied. Click New to create a new table or view.

Connection manager:
DESKTOP-9MTMT42.Ejemplo2 New...

Use a table or view:
"dbo"."Articulos" New...

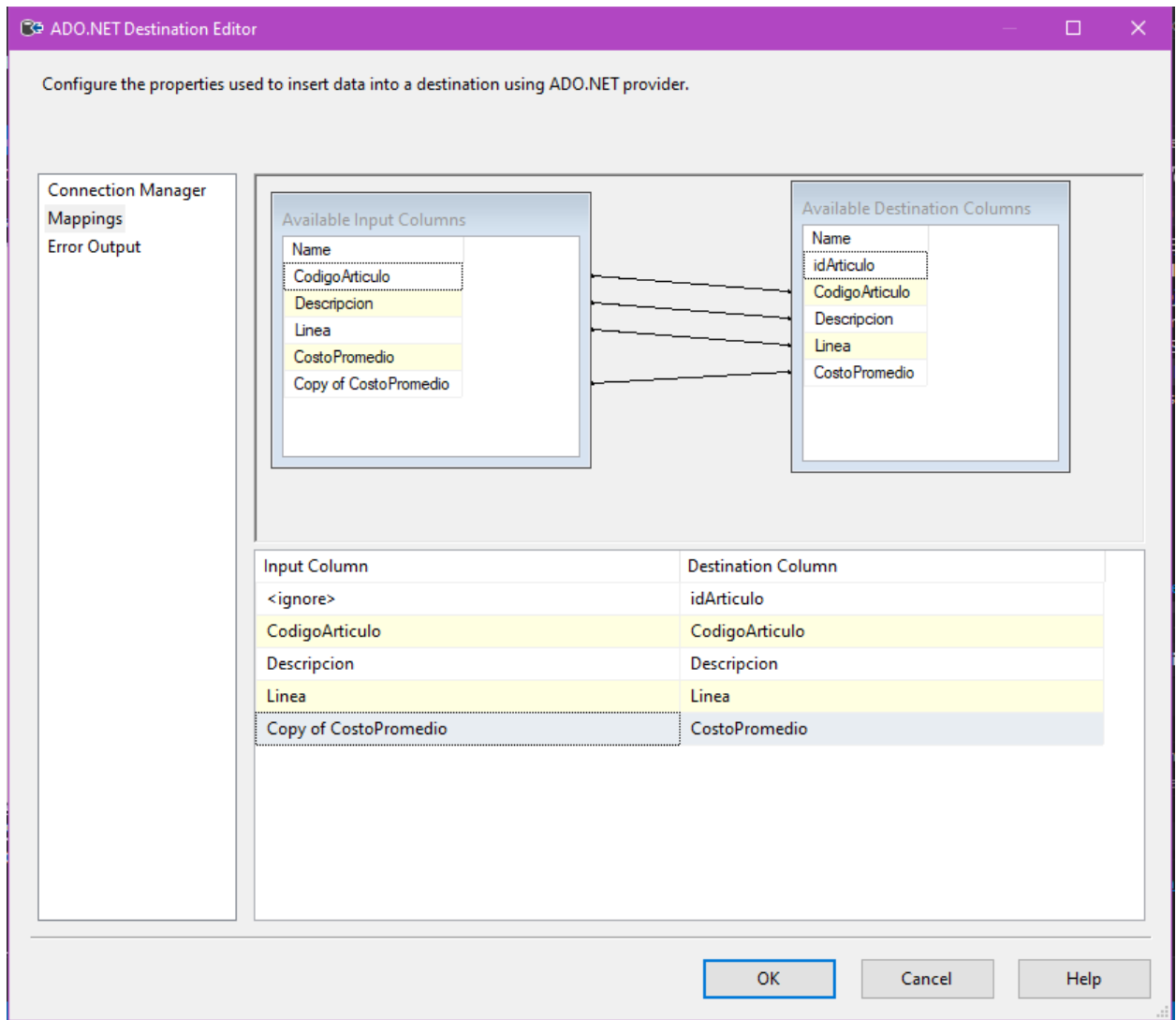
Preview

☒ Use Bulk Insert when possible

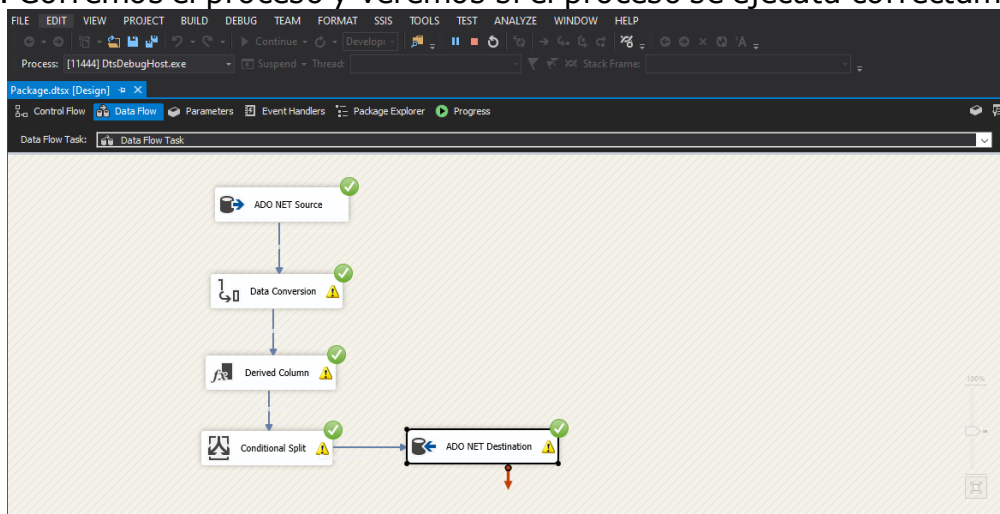
 Map the columns on the Mappings page.

OK Cancel Help

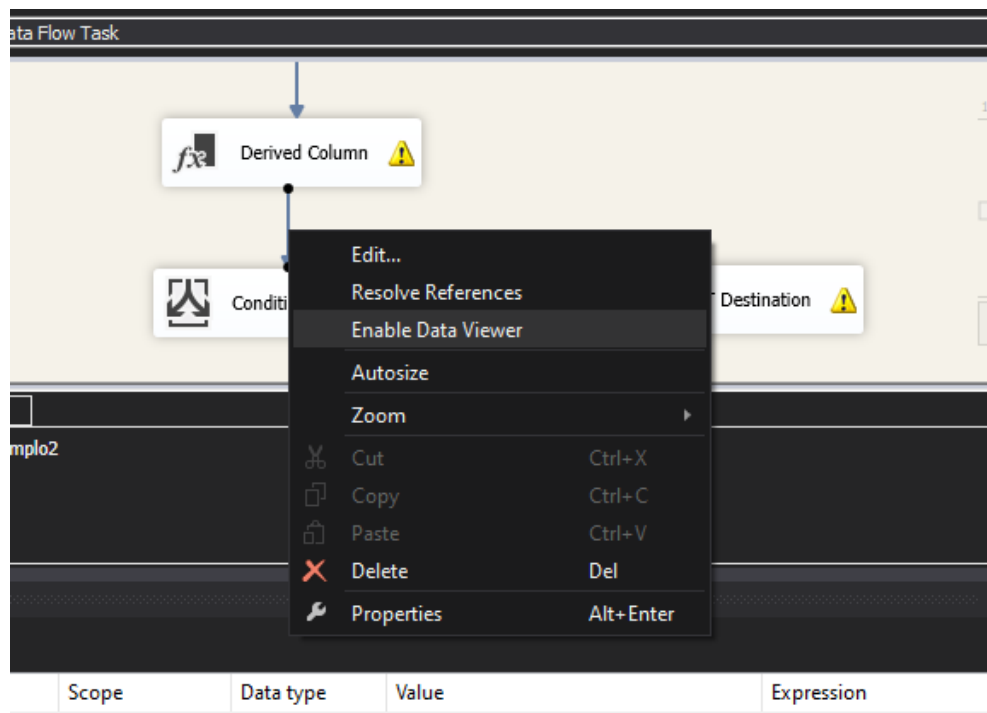
Mapeamos nuestros datos transformados hacia la tabla de datos destino.



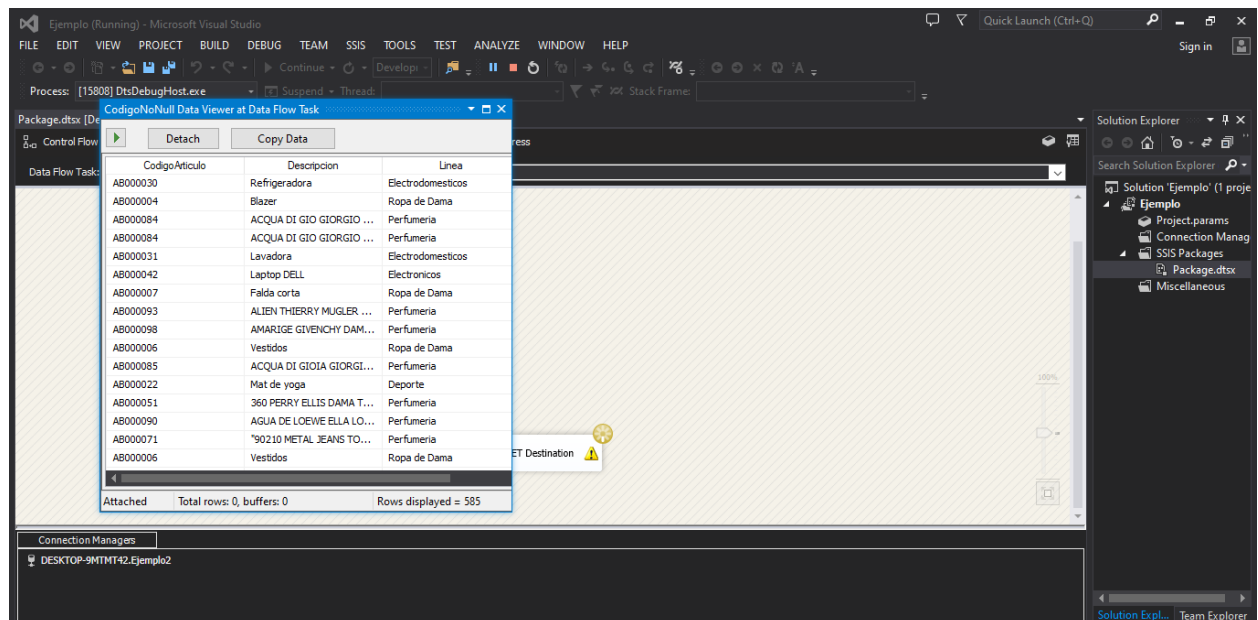
19. Corremos el proceso y veremos si el proceso se ejecuta correctamente.



20. Si necesitamos observar qué datos van pasando en nuestro filtro podemos habilitar el visor de datos dando click derecho encima de una conexión en este caso en la conexión entre División condicional y Destino ADO NET.



21. Y al momento de ejecutar nos desplegará el visor.



22. Por último podremos ver como nuestros datos han sido insertados en la tabla Estudiante, con nuestros datos limpios.

Results		Messages			
	idArticulo	CodigoArticulo	Descripcion	Linea	CostoPromedio
1	1171	AB000051	360 PERRY ELLIS DAMA TOILETTE 100	Perfumeria	390.00
2	1172	AB000059	360 WHITE PERRY ELLIS DAMA PERFUME 100	Perfumeria	441.00
3	1173	AB000075	"90210 SPORT TORAND, A SPELLING ENTERTEIMEN DAMA...	Perfumeria	226.00
4	1174	AB000008	Falda Larga	Ropa de Dama	70.00
5	1175	AB000013	Chumpa de lona	Ropa de caballero	250.00
6	1176	AB000077	"90210 TOUCH OF GOLD TORAND, A SPELLING ENTERTEIM...	Perfumeria	331.00
7	1177	AB000030	Refrigeradora	Electrodomesticos	5999.00
8	1178	AB000008	Falda Larga	Ropa de Dama	70.00
9	1179	AB000075	"90210 SPORT TORAND, A SPELLING ENTERTEIMEN DAMA...	Perfumeria	226.00
10	1180	AB000037	Cama	Muebles	4210.00
11	1181	AB000093	ALIEN THIERRY MUGLER DAMA PERFUME 50	Perfumeria	986.00
12	1182	AB000001	Blusa para dama	Ropa de Dama	50.00
13	1183	AB000081	"90210 TOUCH PINK TORAND, A SPELLING ENTERTEIMEN ...	Perfumeria	256.00
14	1184	AB000057	360 PURPLE PERRY ELLIS DAMA PERFUME 100	Perfumeria	495.00
15	1185	AB000041	Laptop HP	Electronicos	5625.00
16	1186	AB000098	AMARIGE GIVENCHY DAMA TOILETTE 100	Perfumeria	656.00
17	1187	AB000043	Telefono Huawei	Electronicos	2890.00
18	1188	AB000087	ADIDAS FLORAL DREAM ADIDAS DAMA TOILETTE 50	Perfumeria	191.00
19	1189	AB000041	Laptop HP	Electronicos	5625.00
20	1190	AB000056	360 PINK PERRY ELLIS DAMA PERFUME 100	Perfumeria	475.00

Query executed successfully. | DESKTOP-9MTMT42 (11.0 SP1) | DESKTOP-9MTMT42\Gabrie... | Ejemplo2 | 00:00:00 | 195 rows