



Datawarehouse y Minería de Datos

DMD941 G01T

DESAFÍO PRACTICO 1

CICLO 01 - 2025

**Docente: Karen Medrano**

**Estudiante: Alberto Alkindi Ramírez Rivas**

**Carné: RR160517**

**Se porcentaje del desafío que se alcanzó fue el 100%**

# ÍNDICE

<b>ÍNDICE .....</b>	<b>2</b>
<b>EXTRACCIÓN DE DATOS .....</b>	<b>3</b>
<b>Mapeo de datos.....</b>	<b>3</b>
1. Mapeo de clientes.csv .....	3
2. Mapeo de ventas.csv .....	4
3. Mapeo de productos.csv .....	4
<b>Extracción de datos en SQL Server .....</b>	<b>5</b>
1. Extracción de datos en base de datos "techtStore_db_clientes" .....	5
2. Extracción de datos en base de datos "techtStore_db_productos" .....	6
3. Extracción de datos en base de datos "techtStore_db_ventas" .....	7
<b>TRANSFORMACIÓN DE DATOS .....</b>	<b>8</b>
<b>Combinación de orígenes de datos .....</b>	<b>9</b>
Paso 1 .....	9
Paso 2 .....	9
Paso 3 .....	10
Paso 4 .....	11
Paso 5 .....	12
Paso 6 .....	12
Paso 7 .....	13
Paso 8 .....	14
Paso 9 .....	15
Paso 10 .....	16
Paso 11 .....	17
<b>CARGA DE DATOS.....</b>	<b>20</b>
<b>Carga de datos Clientes_Frecuentes.xlsx.....</b>	<b>20</b>
<b>Carga de datos Clientes_Ocasionales.xlsx .....</b>	<b>21</b>

# EXTRACCIÓN DE DATOS

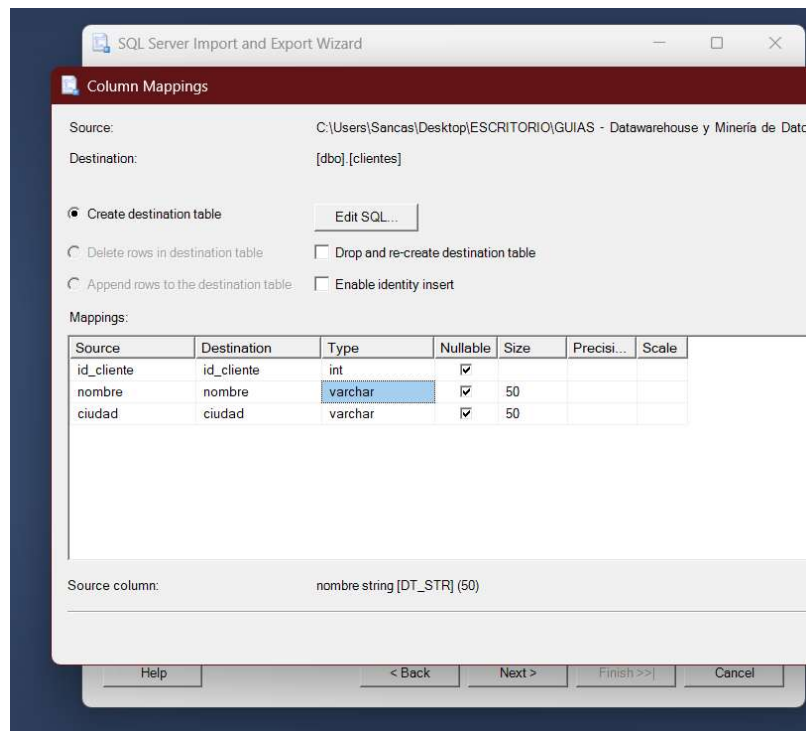
Los datos se extrajeron a una base de datos en SQL Server desde la herramienta de “SQL Server Import and Export Wizard”

Se exportaron los siguientes archivos .csv:

- Clientes.csv
- Productos.csv
- Ventas.csv

## Mapeo de datos

### 1. Mapeo de clientes.csv



## 2. Mapeo de ventas.csv

SQL Server Import and Export Wizard

**Column Mappings**

Source: C:\Users\Sancas\Desktop\ESCRITORIO\GUIAS - Datawarehouse y Minería de Datos

Destination: [dbo].[productos]

☒ Create destination table [Edit SQL...](#)

☐ Delete rows in destination table ☐ Drop and re-create destination table

☐ Append rows to the destination table ☐ Enable identity insert

Mappings:

Source	Destination	Type	Nullable	Size	Precisi...	Scale
id_producto	id_producto	int	<input checked="" type="checkbox"/>			
nombre	nombre	varchar	<input checked="" type="checkbox"/>	50		
precio	precio	money	<input checked="" type="checkbox"/>			

Source column: id\_producto string [DT\_STR] (50)

Help < Back Next > Finish >> Cancel

## 3. Mapeo de productos.csv

**Column Mappings**

Source: C:\Users\Sancas\Desktop\ESCRITORIO\GUIAS - Datawarehouse y Minería de Datos

Destination: [dbo].[ventas]

☒ Create destination table [Edit SQL...](#)

☐ Delete rows in destination table ☐ Drop and re-create destination table

☐ Append rows to the destination table ☐ Enable identity insert

Mappings:

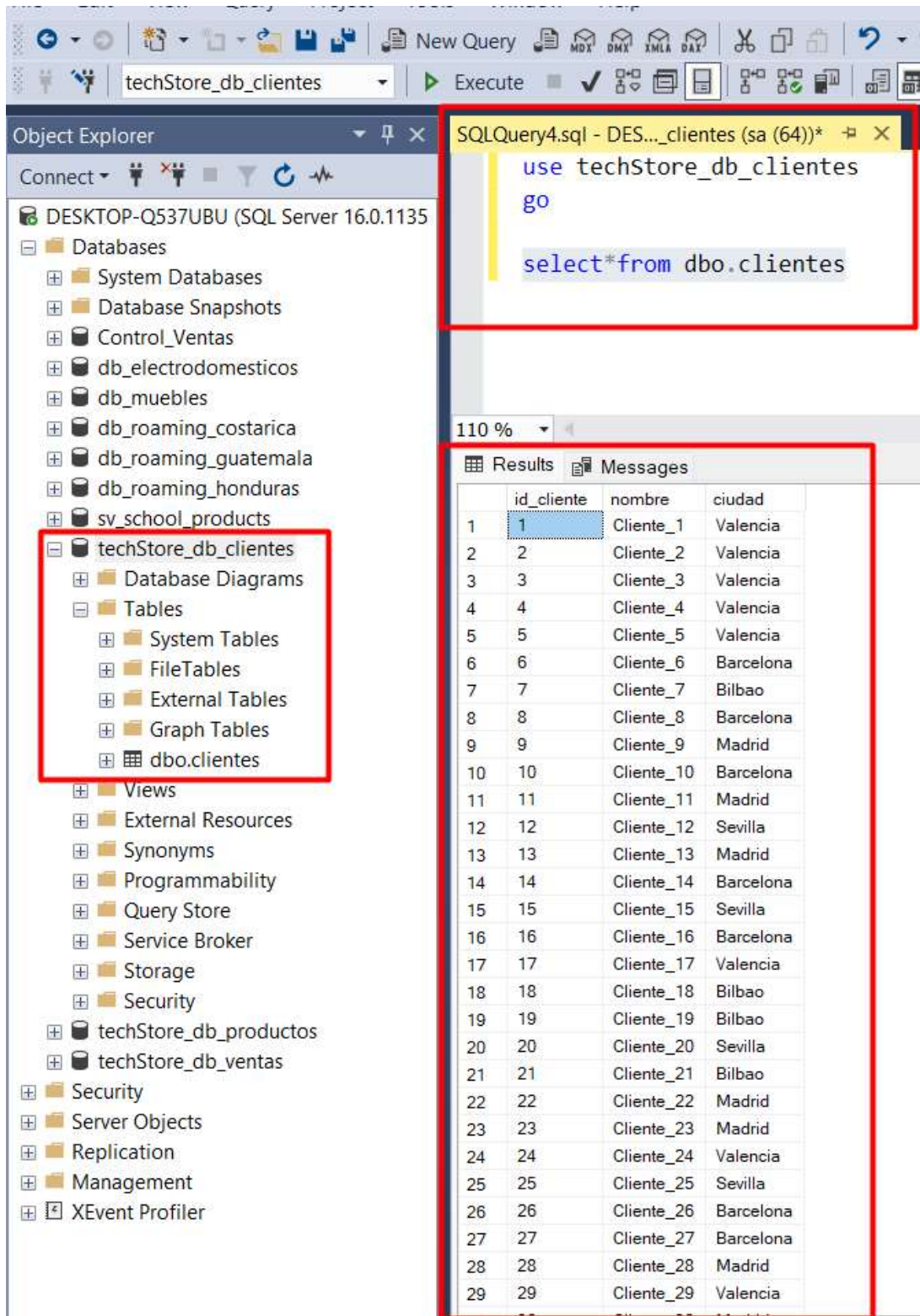
Source	Destination	Type	Nullable	Size	Precisi...	Scale
id_venta	id_venta	int	<input checked="" type="checkbox"/>			
id_cliente	id_cliente	int	<input checked="" type="checkbox"/>			
id_producto	id_producto	int	<input checked="" type="checkbox"/>			
cantidad	cantidad	int	<input checked="" type="checkbox"/>			

Source column: cantidad string [DT\_STR] (50)

Help < Back Next > Finish >> Cancel

## Extracción de datos en SQL Server

### 1. Extracción de datos en base de datos "techStore\_db\_clientes"



Object Explorer

Connect

DESKTOP-Q537UBU (SQL Server 16.0.1135)

Databases

- System Databases
- Database Snapshots
- Control\_Ventas
- db\_electrodomesticos
- db\_muebles
- db\_roaming\_costarica
- db\_roaming\_guatemala
- db\_roaming\_honduras
- sv\_school\_products
- techStore\_db\_clientes**
  - Database Diagrams
  - Tables**
    - System Tables
    - FileTables
    - External Tables
    - Graph Tables
    - dbo.clientes**
  - Views
  - External Resources
  - Synonyms
  - Programmability
  - Query Store
  - Service Broker
  - Storage
  - Security
- techStore\_db\_productos
- techStore\_db\_ventas
- Security
- Server Objects
- Replication
- Management
- XEvent Profiler

SQLQuery4.sql - DES...\_clientes (sa (64))\*

```
use techStore_db_clientes
go

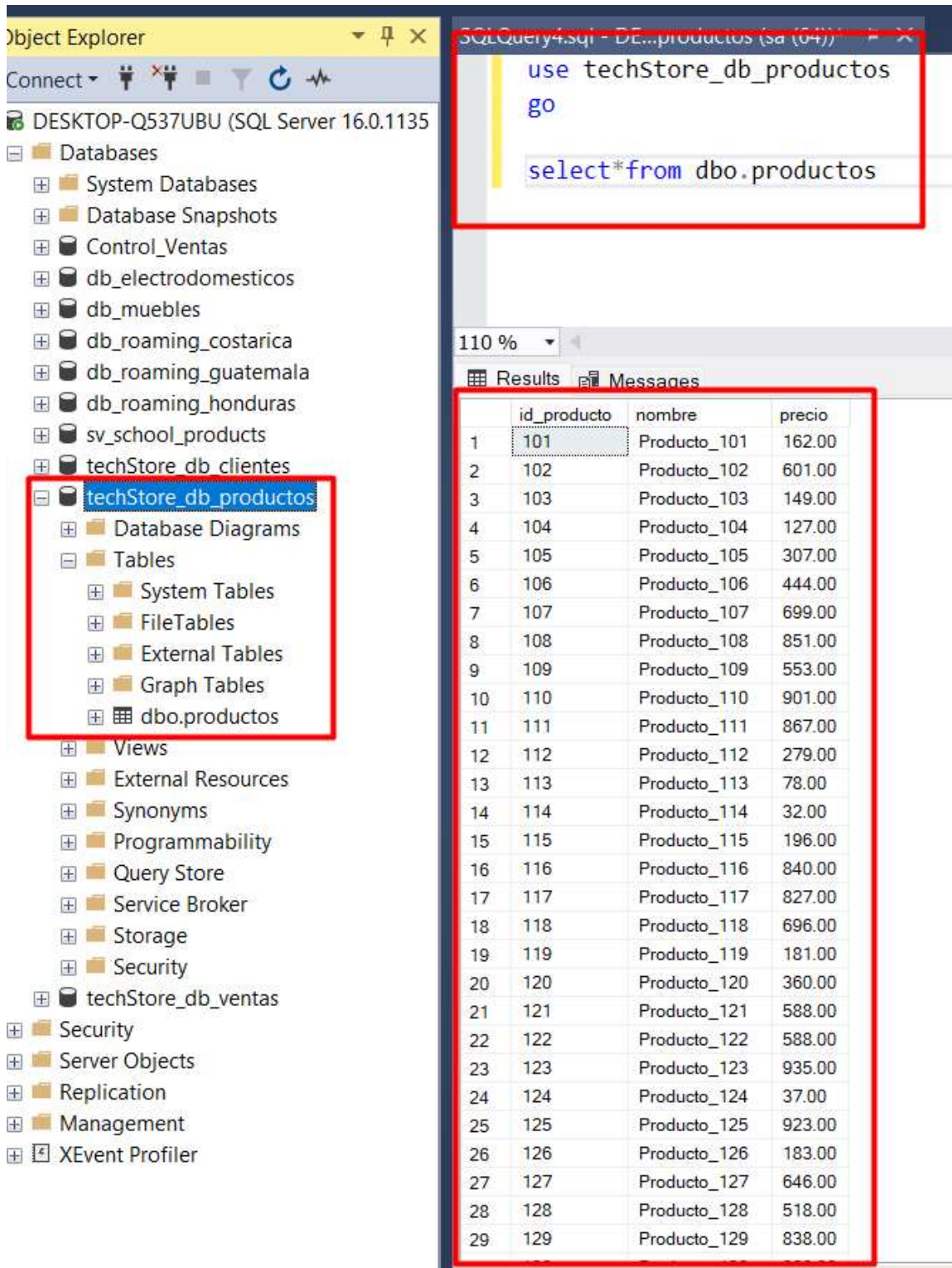
select * from dbo.clientes
```

110 %

Results Messages

	id_cliente	nombre	ciudad
1	1	Cliente_1	Valencia
2	2	Cliente_2	Valencia
3	3	Cliente_3	Valencia
4	4	Cliente_4	Valencia
5	5	Cliente_5	Valencia
6	6	Cliente_6	Barcelona
7	7	Cliente_7	Bilbao
8	8	Cliente_8	Barcelona
9	9	Cliente_9	Madrid
10	10	Cliente_10	Barcelona
11	11	Cliente_11	Madrid
12	12	Cliente_12	Sevilla
13	13	Cliente_13	Madrid
14	14	Cliente_14	Barcelona
15	15	Cliente_15	Sevilla
16	16	Cliente_16	Barcelona
17	17	Cliente_17	Valencia
18	18	Cliente_18	Bilbao
19	19	Cliente_19	Bilbao
20	20	Cliente_20	Sevilla
21	21	Cliente_21	Bilbao
22	22	Cliente_22	Madrid
23	23	Cliente_23	Madrid
24	24	Cliente_24	Valencia
25	25	Cliente_25	Sevilla
26	26	Cliente_26	Barcelona
27	27	Cliente_27	Barcelona
28	28	Cliente_28	Madrid
29	29	Cliente_29	Valencia

## 2. Extracción de datos en base de datos “techStore\_db\_productos”



The screenshot displays the SQL Server Enterprise Manager interface. In the Object Explorer on the left, the 'techStore\_db\_productos' database is selected, and its 'dbo.productos' table is highlighted. The central SQL query window shows the following code:

```
use techStore_db_productos
go

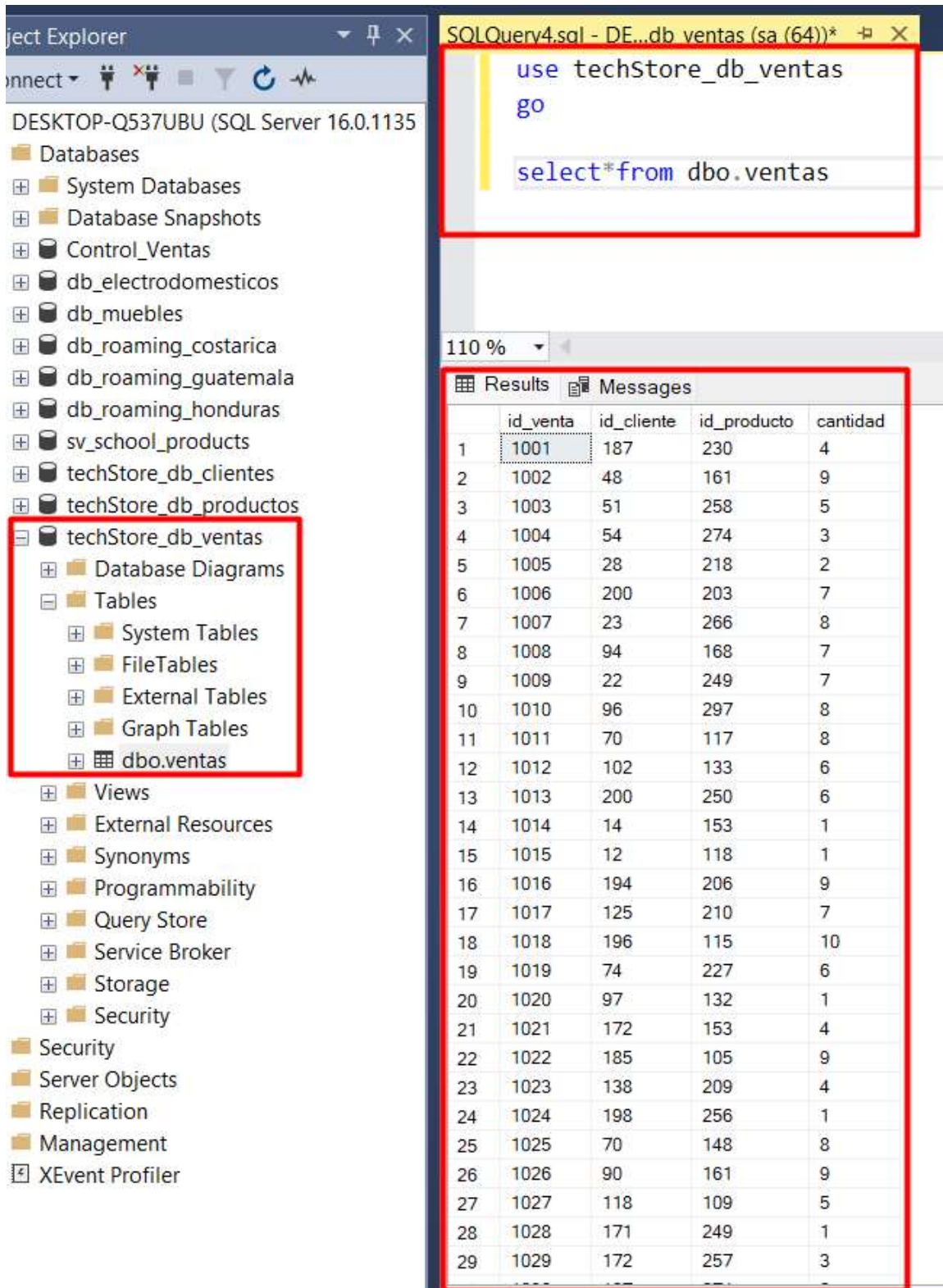
select * from dbo.productos
```

The Results window on the right displays the data from the 'dbo.productos' table. The results are as follows:

	id_producto	nombre	precio
1	101	Producto_101	162.00
2	102	Producto_102	601.00
3	103	Producto_103	149.00
4	104	Producto_104	127.00
5	105	Producto_105	307.00
6	106	Producto_106	444.00
7	107	Producto_107	699.00
8	108	Producto_108	851.00
9	109	Producto_109	553.00
10	110	Producto_110	901.00
11	111	Producto_111	867.00
12	112	Producto_112	279.00
13	113	Producto_113	78.00
14	114	Producto_114	32.00
15	115	Producto_115	196.00
16	116	Producto_116	840.00
17	117	Producto_117	827.00
18	118	Producto_118	696.00
19	119	Producto_119	181.00
20	120	Producto_120	360.00
21	121	Producto_121	588.00
22	122	Producto_122	588.00
23	123	Producto_123	935.00
24	124	Producto_124	37.00
25	125	Producto_125	923.00
26	126	Producto_126	183.00
27	127	Producto_127	646.00
28	128	Producto_128	518.00
29	129	Producto_129	838.00



### 3. Extracción de datos en base de datos “techStore\_db\_ventas”



The screenshot shows the SQL Server Enterprise Manager interface. On the left, the 'Object Explorer' pane displays the database structure for 'DESKTOP-Q537UBU (SQL Server 16.0.1135)'. The 'Databases' folder is expanded, and 'techStore\_db\_ventas' is selected. Under 'Tables', 'dbo.ventas' is highlighted. On the right, the 'SQL Query Editor' window shows the following SQL query:

```
use techStore_db_ventas
go

select * from dbo.ventas
```

Below the query editor, the 'Results' pane displays the data from the 'dbo.ventas' table. The results are shown in a table with the following columns: 'id\_venta', 'id\_cliente', 'id\_producto', and 'cantidad'.

	id_venta	id_cliente	id_producto	cantidad
1	1001	187	230	4
2	1002	48	161	9
3	1003	51	258	5
4	1004	54	274	3
5	1005	28	218	2
6	1006	200	203	7
7	1007	23	266	8
8	1008	94	168	7
9	1009	22	249	7
10	1010	96	297	8
11	1011	70	117	8
12	1012	102	133	6
13	1013	200	250	6
14	1014	14	153	1
15	1015	12	118	1
16	1016	194	206	9
17	1017	125	210	7
18	1018	196	115	10
19	1019	74	227	6
20	1020	97	132	1
21	1021	172	153	4
22	1022	185	105	9
23	1023	138	209	4
24	1024	198	256	1
25	1025	70	148	8
26	1026	90	161	9
27	1027	118	109	5
28	1028	171	249	1
29	1029	172	257	3

# TRANSFORMACIÓN DE DATOS

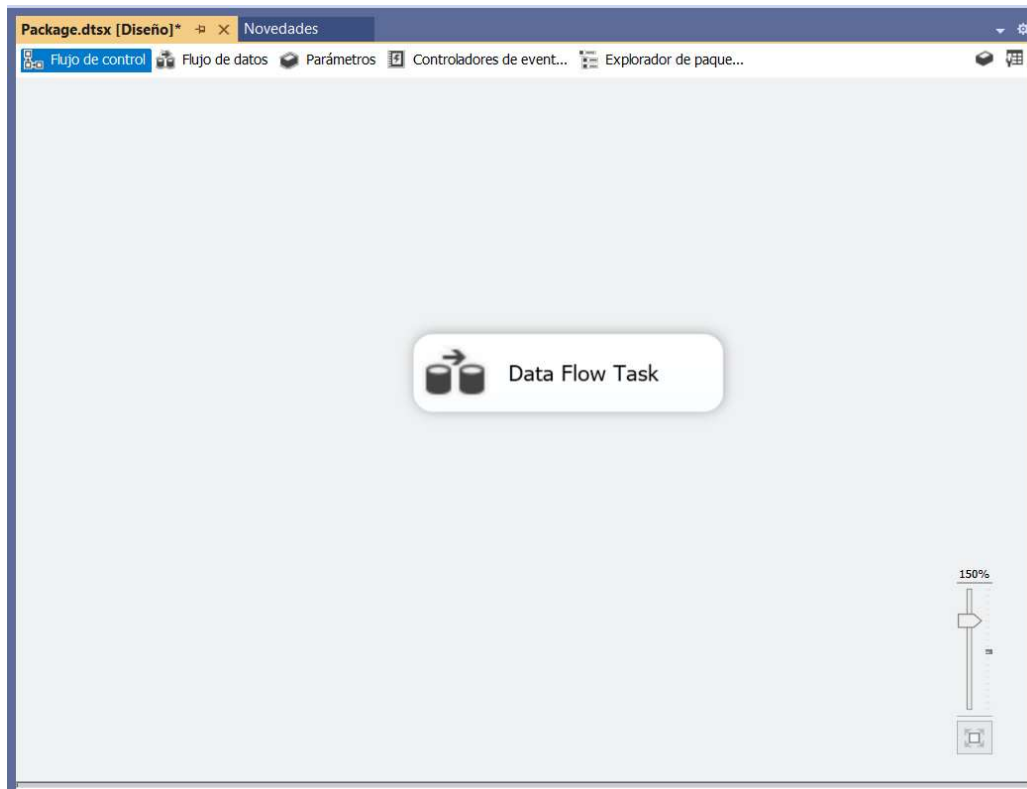
Primeramente, se creará un nuevo proyecto *Integration Services Project* con el nombre *DesafíoPractico\_1*.



## Integration Services Project

This project may be used for building high performance data integration and workflow solutions that can be run on SSIS catalog, including extraction, transformation, and loading (ETL) operations for data warehousing.

Luego se creará un nuevo *Data Flow Task*.





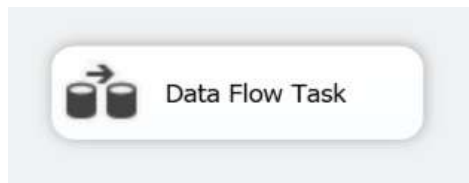
Se ocupará los siguientes componentes SQL:

- Origen de ADO NET (3)
- Union All (1)
- Aggregate (1)
- Data Conversion (1)
- Derived Column (1)
- Conditional Split (1)
- Sort (2)
- Excel Destination (2)

## Combinación de orígenes de datos

### Paso 1

Se creó un proyecto de IIS y dentro del dataflow.



### Paso 2

Se agregaron 3 orígenes de ADO NET con los siguientes nombres:

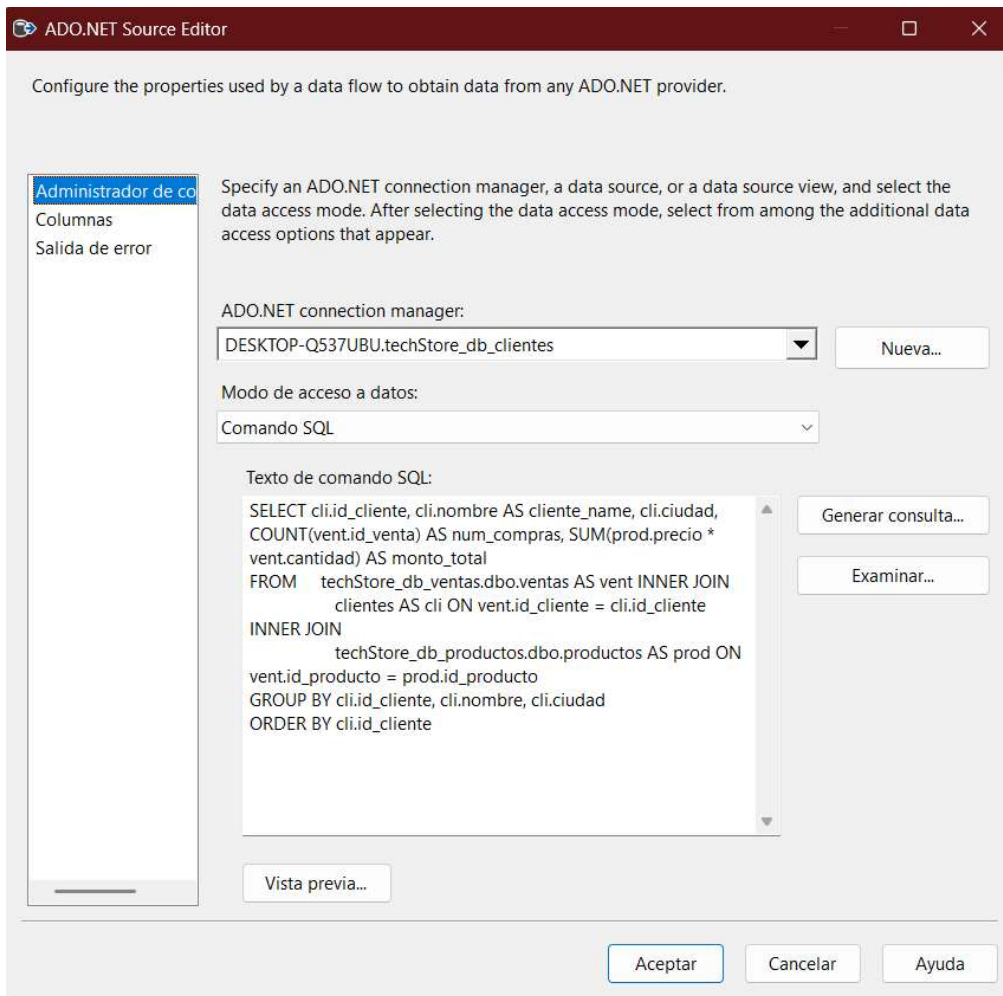
- BD CLIENTES
- BD PRODUCTOS
- BD VENTAS

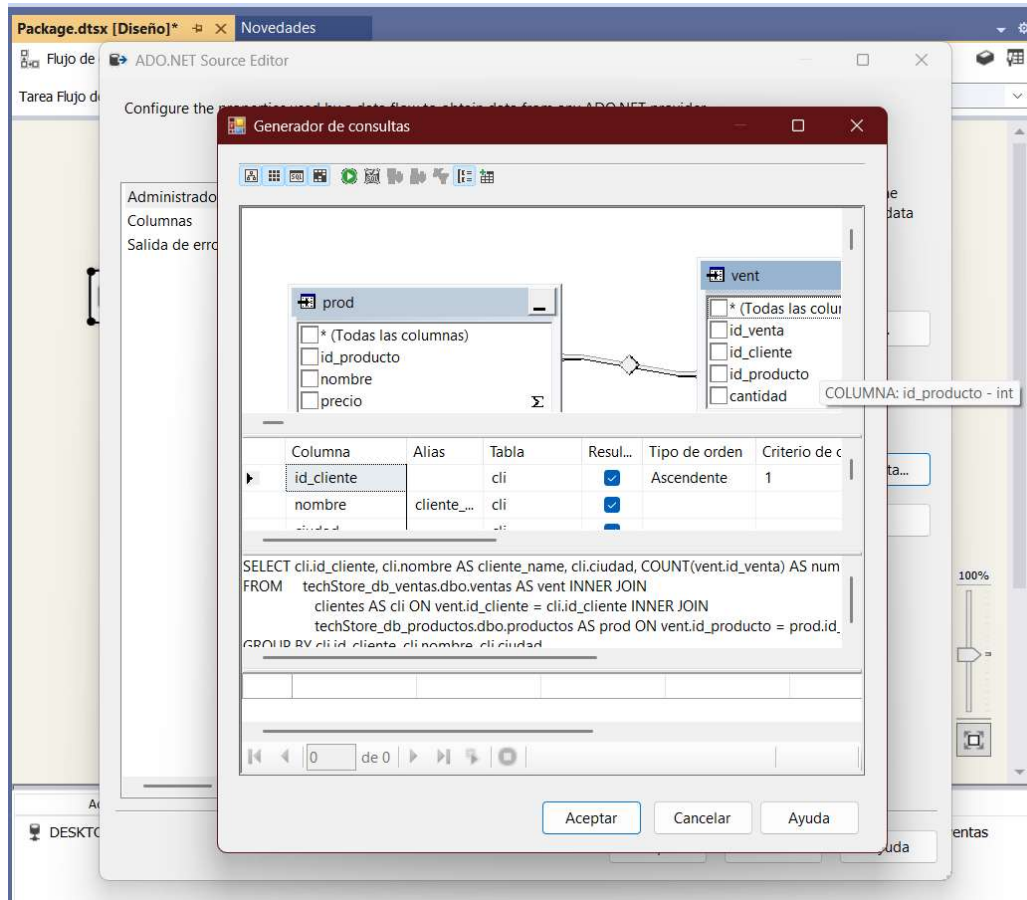


### Paso 3

Se configuraron los orígenes de cada utilizando la siguiente consulta:

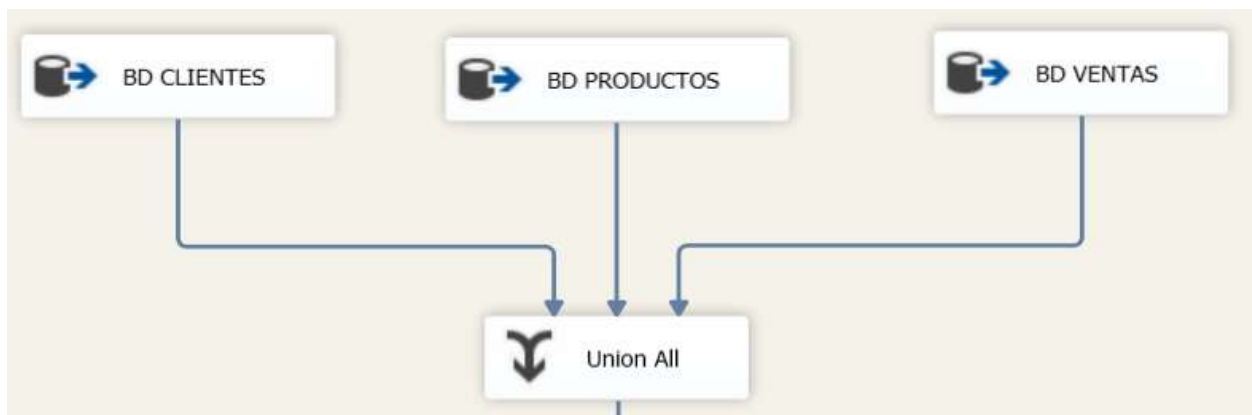
```
SELECT cli.id_cliente, cli.nombre AS cliente_name, cli.ciudad,  
       COUNT(vent.id_venta) AS num_compras,  
       SUM(prod.precio * vent.cantidad) AS monto_total  
FROM techStore_db_ventas.dbo.ventas AS vent  
INNER JOIN techStore_db_clientes.dbo.clientes AS cli ON vent.id_cliente = cli.id_cliente  
INNER JOIN techStore_db_productos.dbo.productos AS prod ON vent.id_producto =  
prod.id_producto  
GROUP BY cli.id_cliente, cli.nombre, cli.ciudad  
ORDER BY cli.id_cliente;
```





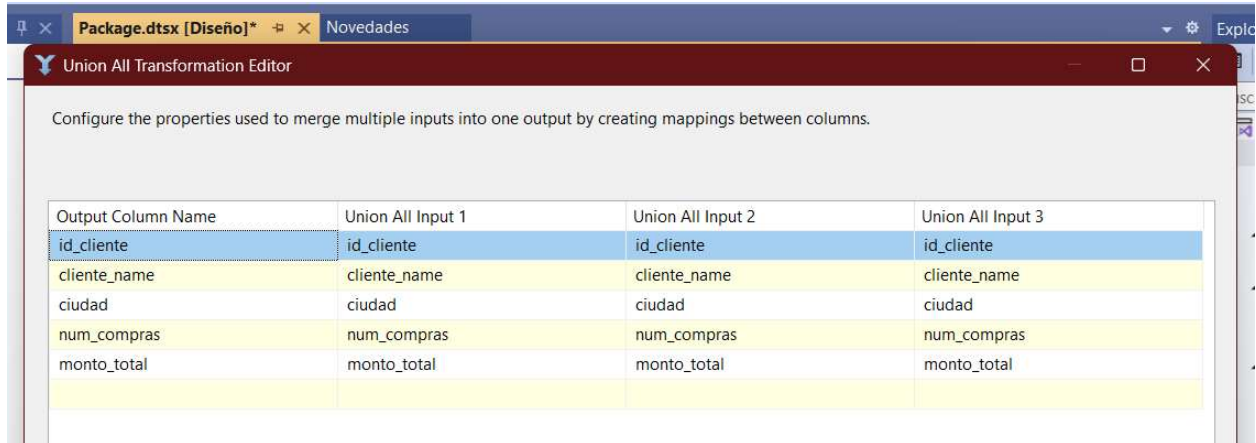
## Paso 4

Se agregó el control UNION ALL y se conectaron los ADO NET



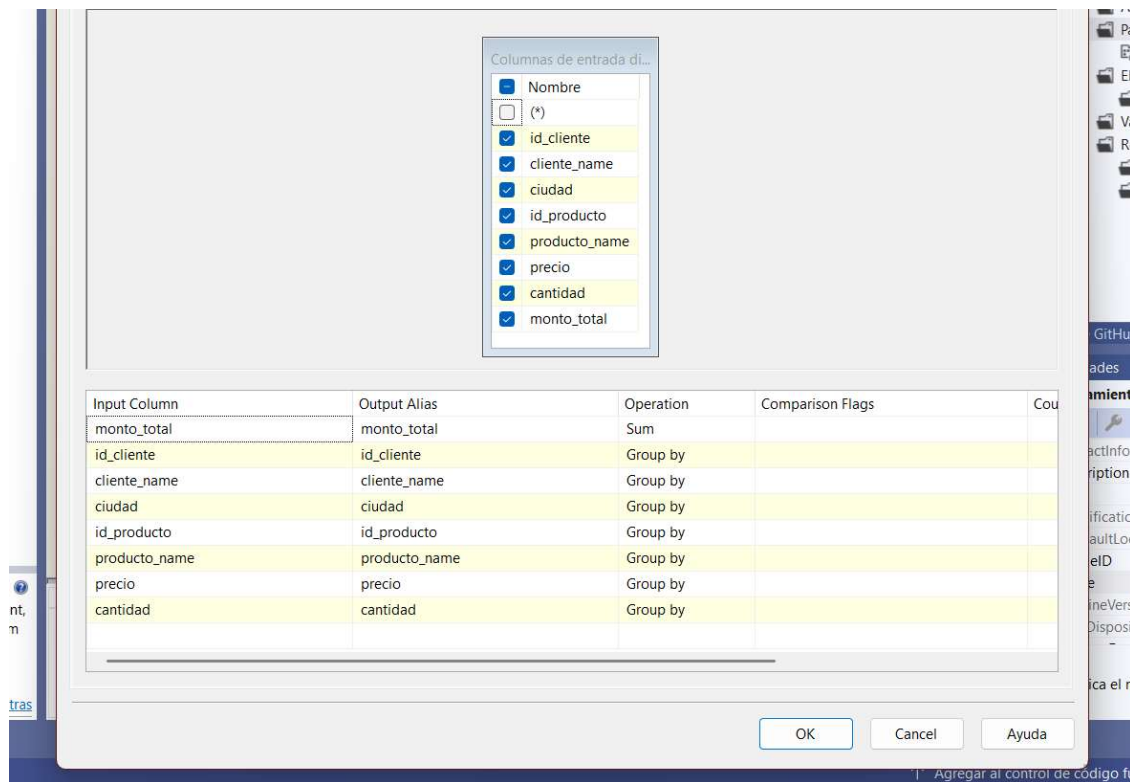
## Paso 5

Luego se configuró el componente UNION ALL para hacer coincidir los tres orígenes.

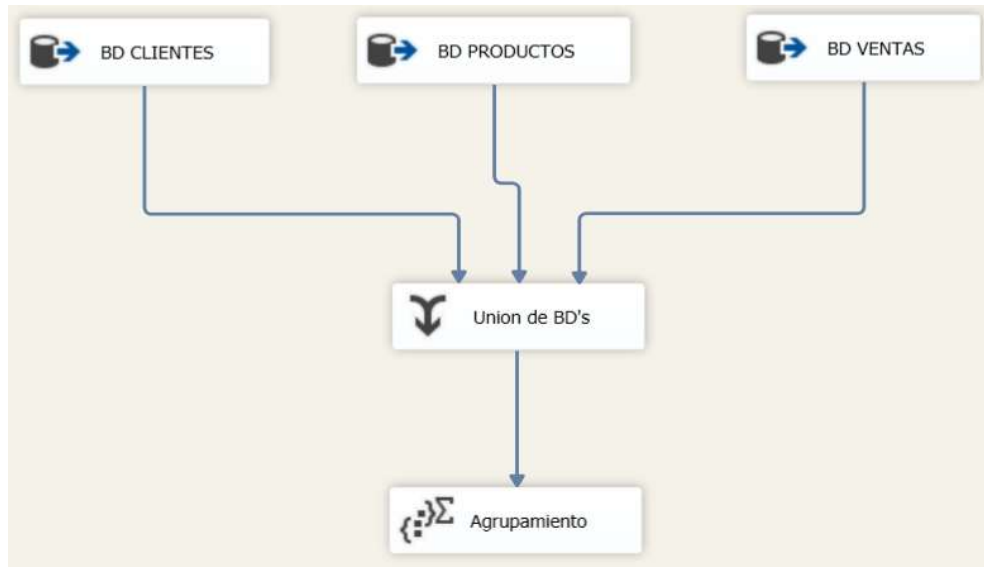


## Paso 6

Luego se agregó el control AGGREGATE y se conectó al UNION ALL. Se cambió el campo monto\_total a la operación Sum. Todos los demás campos se dejaron en la operación Group by

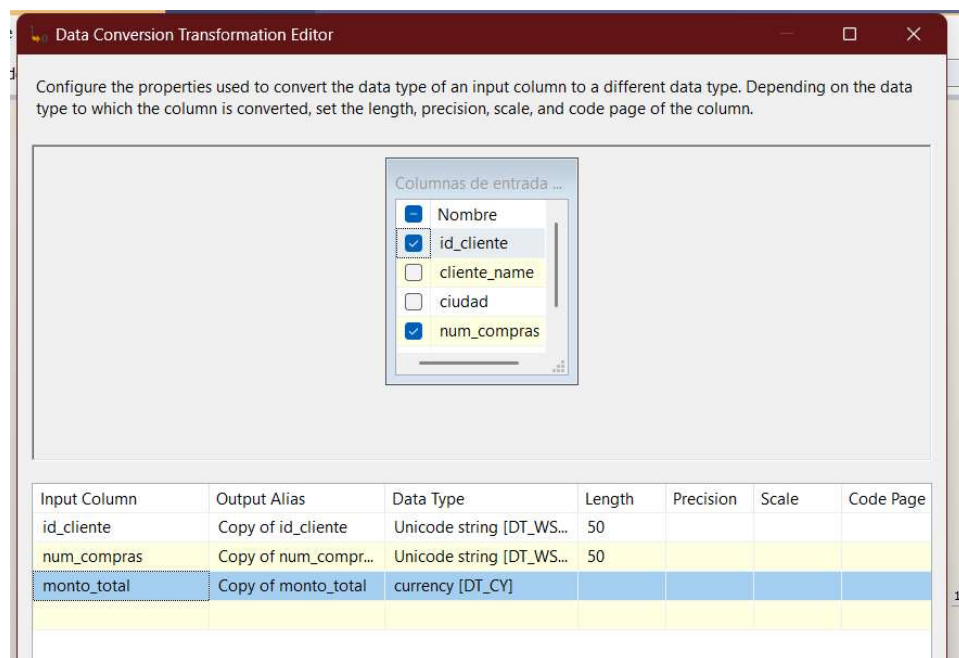


El flujo de momento queda de la siguiente forma



## Paso 7

Se coloca un componente DATA CONVERSION para convertir los campos `id_cliente` (identificativo del cliente) y `num_comprar` (cantidad de compras del cliente) en el tipo de dato **Unicode string [DT\_WDTR]** y el campo `monto_total` (Total de las compras por un cliente) en el tipo de dato **currency [DT\_CY]**.





## Paso 8

Se agregó un control DERIVED COLUMN.

Con este campo se generará el código de país y el tipo de cliente.

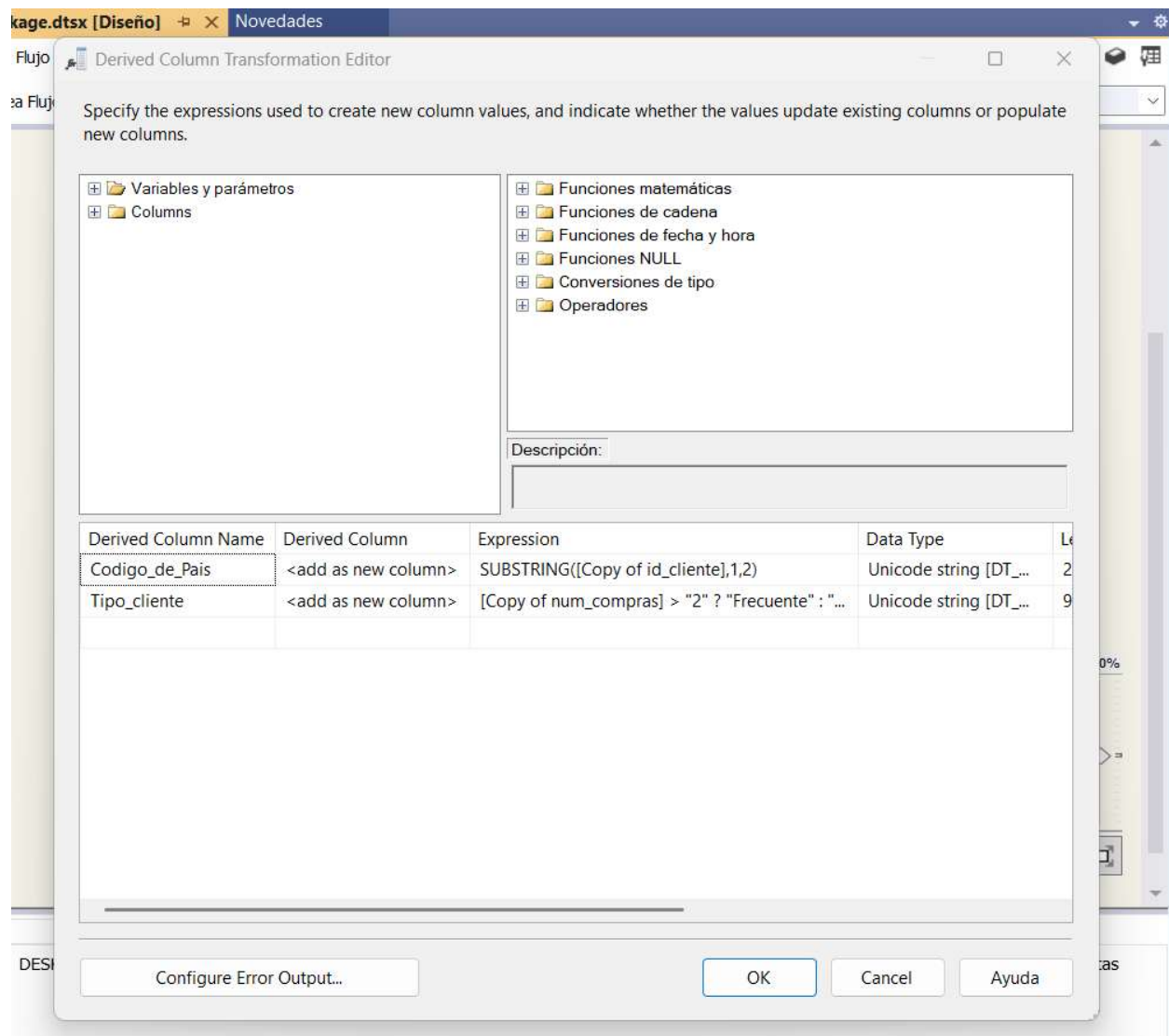
Para el código de país se ocupará la siguiente expresión:

`SUBSTRING([Copy of id_cliente],1,2)`

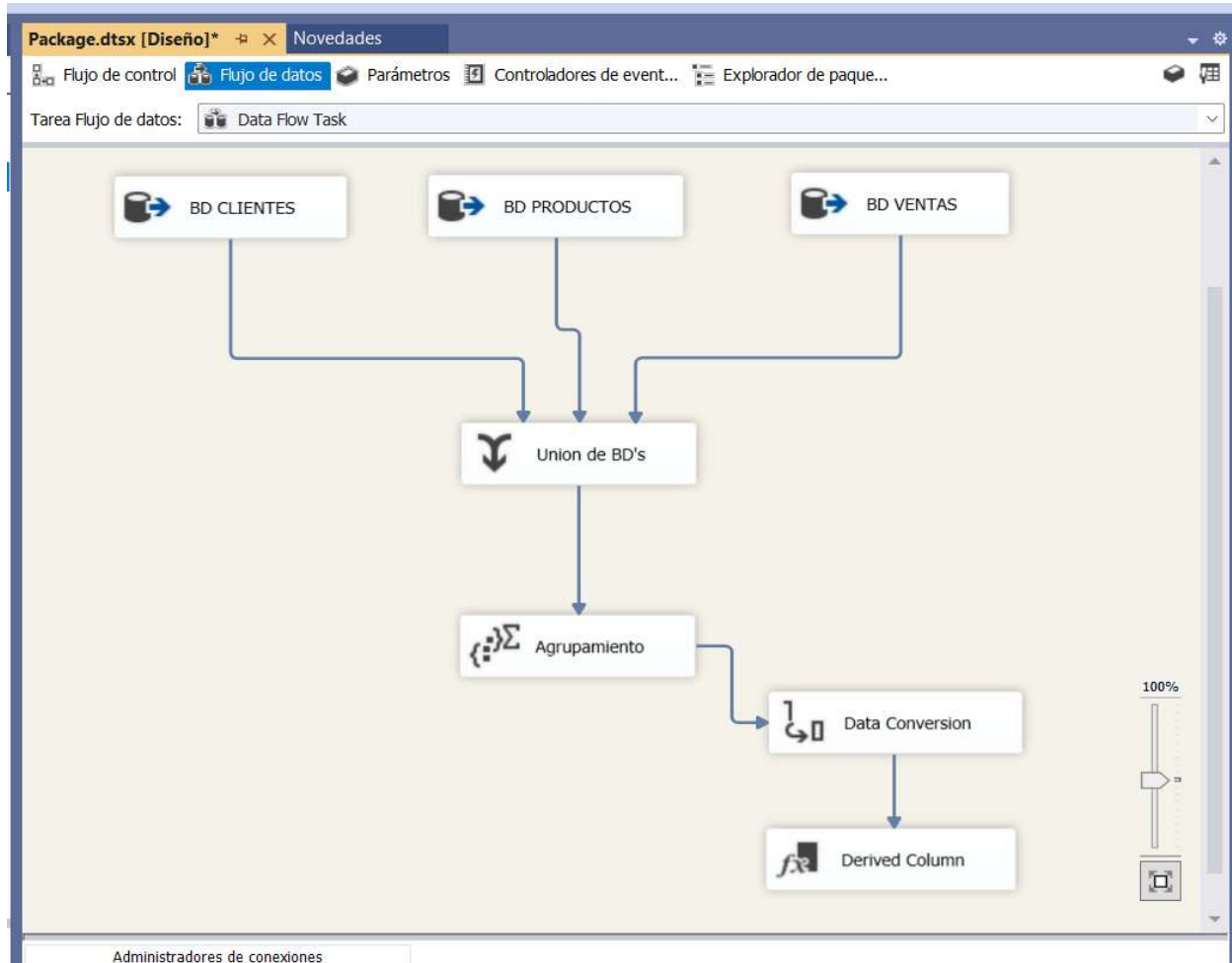
Para el tipo de cliente se tomará se utilizará la siguiente expresión:

`[Copy of num_compras] > "1" ? "Frecuente" : "Ocasional"`

Se tomará en cuenta que los que los clientes Ocasionales serán los que han comprado **1 única vez** y los clientes Frecuentes serán los que han comprado **más de 1 vez**.

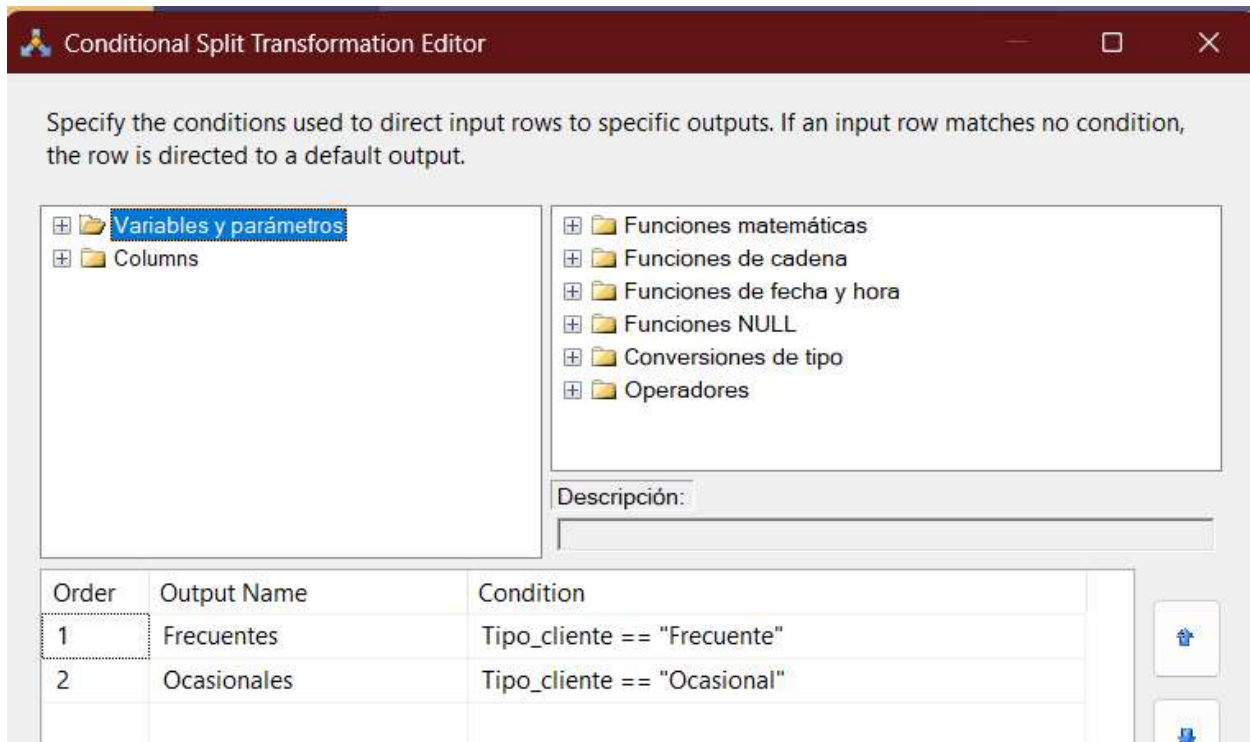


De momento nuestro dataflow se vería de la siguiente forma:



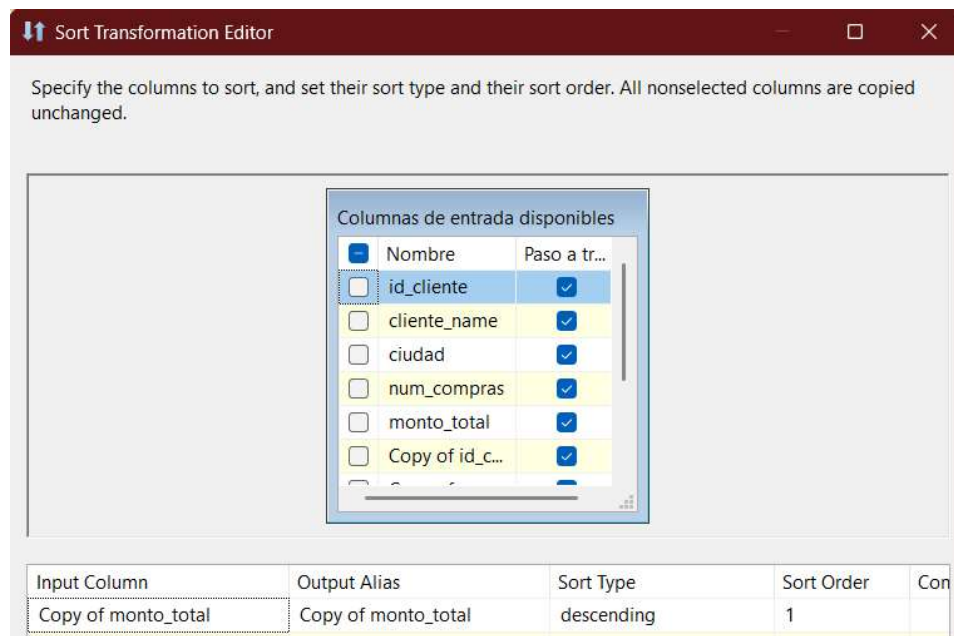
## Paso 9

Para continuar, se agregó el control CONDITIONAL SPLIT en donde se realizó la condición del tipo de cliente, tal cual se ve en la siguiente imagen:



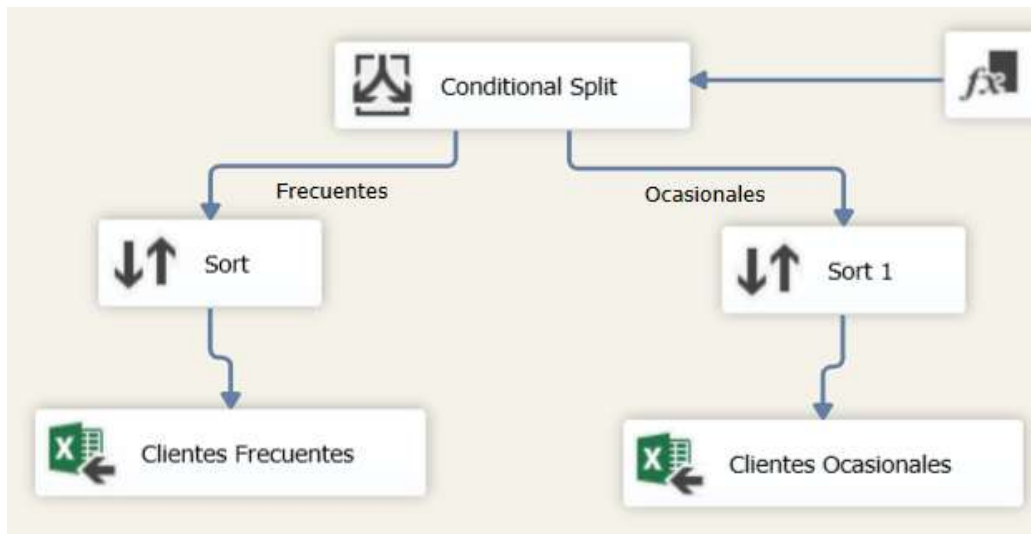
## Paso 10

Luego se agregó 2 controles SORT para ordenar los datos de los clientes frecuentes y ocasionales con el campo Copy of monto\_total, para ordenar los clientes que han comprado de más a menos:



## Paso 11

Se agregaron 2 EXCEL DESTINATION, quedando de la siguiente manera:



El mapeo de los controles de la configuración del EXCEL queda de la siguiente manera

Excel Destination Editor

Configure the properties that enable the insertion of data via an Excel provider.

Administrador de co

**Asignaciones**

Salida de error

Columnas de ent...

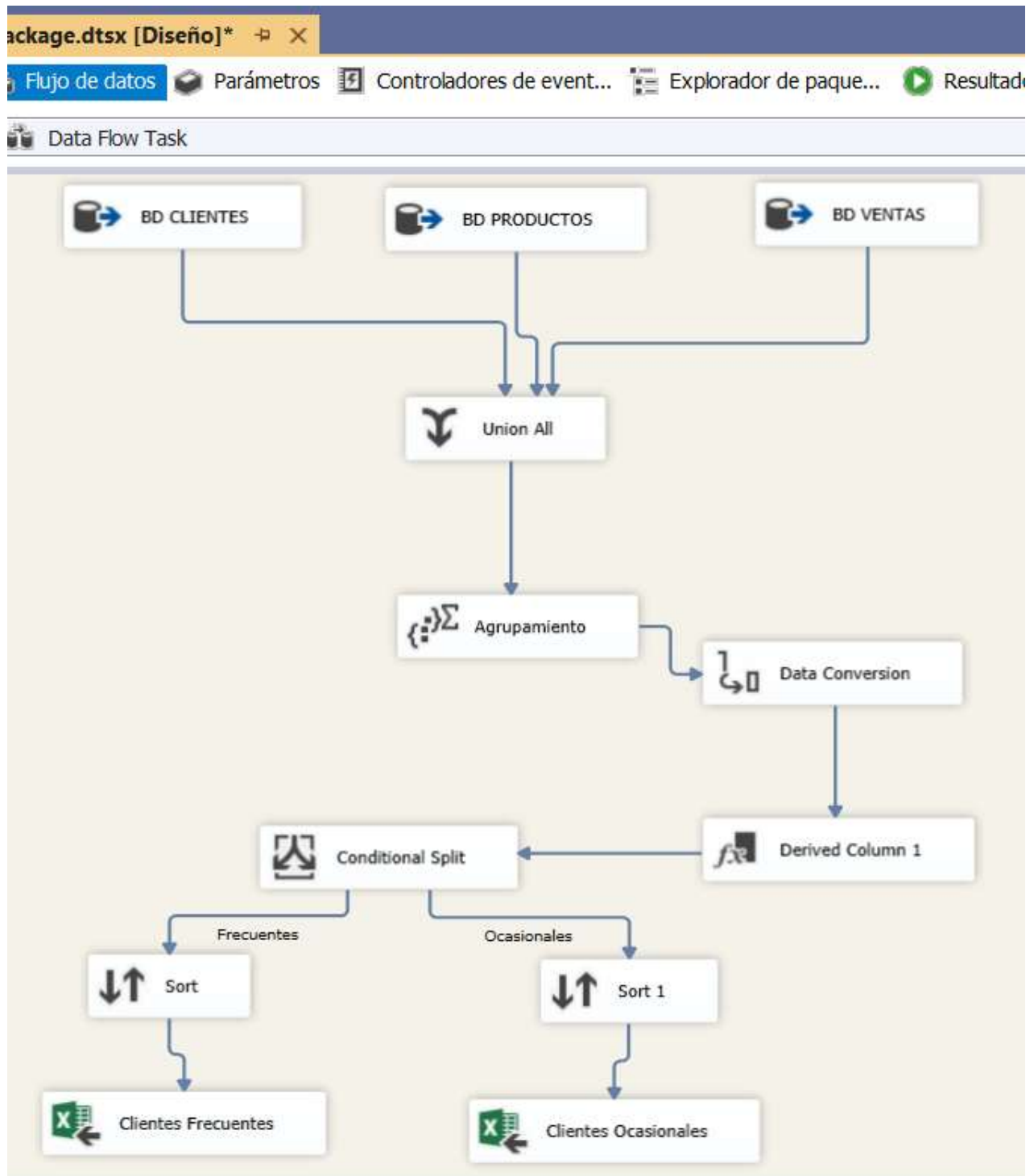
- Nombre
- id\_cliente
- cliente\_name
- ciudad
- num\_compras
- monto\_total
- Copy of id\_cl...
- Copy of num...
- Copy of mon...
- Codigo\_de\_...

Columnas de destino disp...

- Nombre
- ID CLIENTE
- CÓDIGO DEL PAÍS
- NOMBRE DE CLIENTE
- CIUDAD
- NÚMERO DE COMPRAS
- MONTO TOTAL

Columna de entrada	Columna de destino
Copy of id_cliente	ID CLIENTE
Codigo_de_Pais	CÓDIGO DEL PAÍS
cliente_name	NOMBRE DE CLIENTE
ciudad	CIUDAD
Copy of num_compras	NÚMERO DE COMPRAS
Copy of monto_total	MONTO TOTAL
Tipo_cliente	TIPO DE CLIENTE

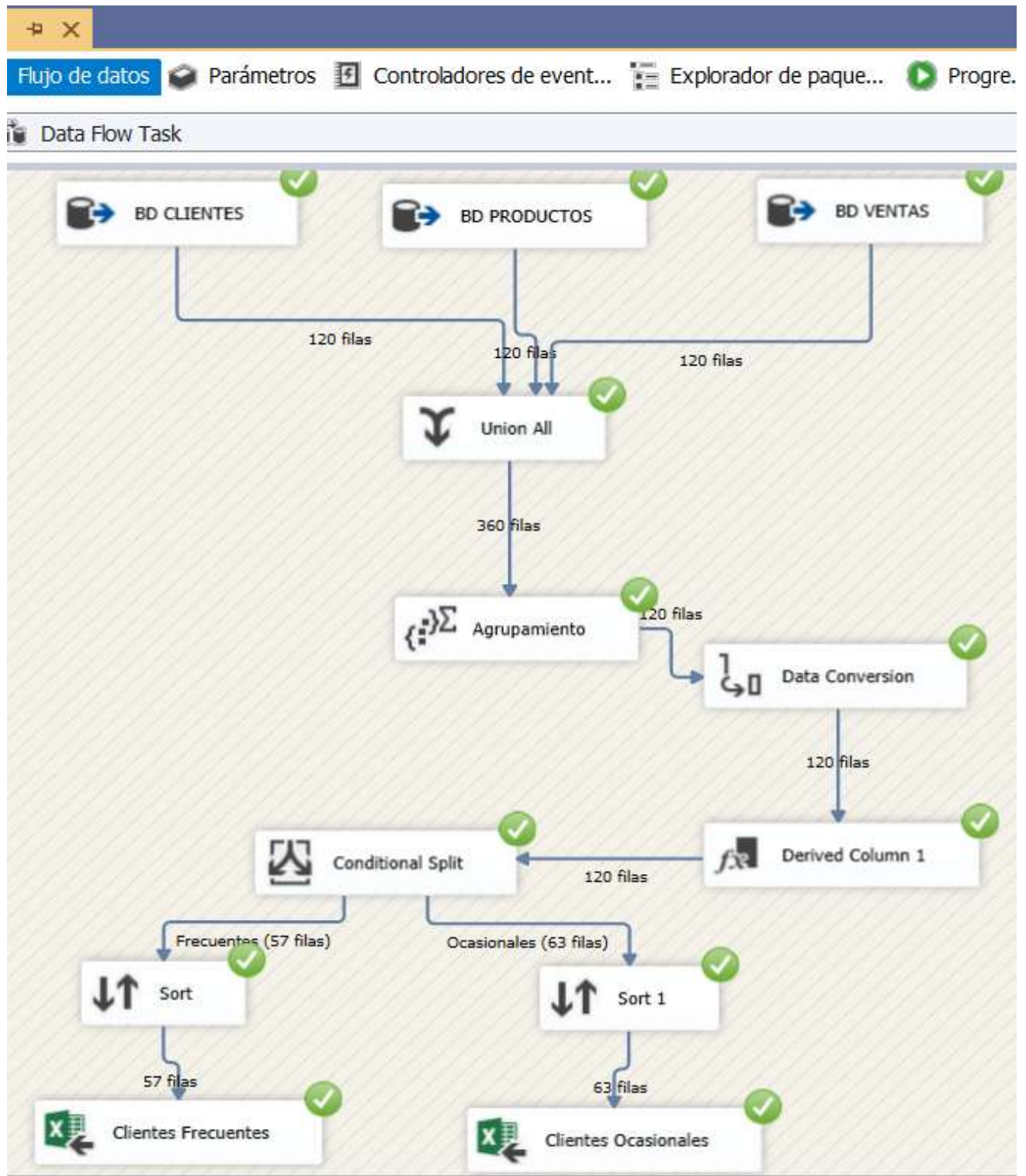
El proyecto queda de la siguiente forma:





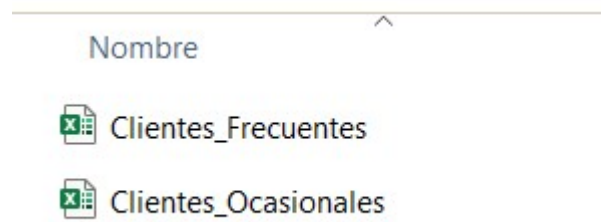
## Paso 11

Se ejecutó el proyecto y se revisaron las siguientes salidas



# CARGA DE DATOS

Se cargaron los datos en dos archivos de Excel:



## Carga de datos Clientes\_Frecuentes.xlsx

Se cargaron **57 clientes frecuentes** en la tienda TechStore, teniendo en cuenta que se tomó el criterio de colocar clientes frecuentes los que poseen **más de 1 número de compra** (de 2 compras en adelante)

ID CLIENTE	CÓDIGO DEL PAÍS	NOMBRE DE CLIENTE	CIUDAD	NÚMERO DE COMPRAS	MONTO TOTAL	TIPO DE CLIENTE
200	20	Cliente_200	Valencia	5	43566	Frecuente
178	17	Cliente_178	Madrid	2	43464	Frecuente
194	19	Cliente_194	Valencia	2	42183	Frecuente
51	51	Cliente_51	Bilbao	3	39372	Frecuente
109	10	Cliente_109	Sevilla	3	39189	Frecuente
155	15	Cliente_155	Valencia	2	37644	Frecuente
158	15	Cliente_158	Madrid	4	36795	Frecuente
70	70	Cliente_70	Valencia	2	32832	Frecuente
61	61	Cliente_61	Barcelona	2	31668	Frecuente
122	12	Cliente_122	Valencia	3	31098	Frecuente
21	21	Cliente_21	Bilbao	3	29436	Frecuente
174	17	Cliente_174	Valencia	3	28296	Frecuente
60	60	Cliente_60	Bilbao	2	28194	Frecuente
107	10	Cliente_107	Barcelona	2	27378	Frecuente
88	88	Cliente_88	Barcelona	2	26922	Frecuente
9	9	Cliente_9	Madrid	2	26217	Frecuente
198	19	Cliente_198	Barcelona	3	24828	Frecuente
98	98	Cliente_98	Valencia	2	24600	Frecuente
30	30	Cliente_30	Madrid	2	23724	Frecuente
160	16	Cliente_160	Sevilla	3	23619	Frecuente
138	13	Cliente_138	Bilbao	3	23019	Frecuente
38	38	Cliente_38	Sevilla	3	21633	Frecuente
81	81	Cliente_81	Barcelona	3	21003	Frecuente
167	16	Cliente_167	Bilbao	3	20940	Frecuente
11	11	Cliente_11	Madrid	2	20892	Frecuente
130	13	Cliente_130	Barcelona	3	20436	Frecuente
100	10	Cliente_100	Valencia	2	20154	Frecuente
142	14	Cliente_142	Sevilla	2	19275	Frecuente
63	63	Cliente_63	Bilbao	2	19098	Frecuente

## Carga de datos Clientes\_Ocasionales.xlsx

Se cargaron **63 clientes ocasionales** en la tienda TechStore, teniendo en cuenta que se tomó el criterio de colocar clientes ocasionales los que poseen solo **1 número de compra**

ID CLIENTE	CÓDIGO DEL PAIS	NOMBRE DE CLIENTE	CIUDAD	NÚMERO DE COMPRAS	MONTO TOTAL	TIPO DE CLIENTE
154	15	Cliente_154	Madrid	1	29670	Ocasional
156	15	Cliente_156	Madrid	1	22632	Ocasional
96	96	Cliente_96	Valencia	1	22608	Ocasional
125	12	Cliente_125	Barcelona	1	20853	Ocasional
99	99	Cliente_99	Barcelona	1	20769	Ocasional
24	24	Cliente_24	Valencia	1	20769	Ocasional
27	27	Cliente_27	Barcelona	1	19698	Ocasional
82	82	Cliente_82	Bilbao	1	19698	Ocasional
48	48	Cliente_48	Valencia	1	18198	Ocasional
20	20	Cliente_20	Sevilla	1	17940	Ocasional
106	10	Cliente_106	Valencia	1	17682	Ocasional
101	10	Cliente_101	Bilbao	1	17472	Ocasional
102	10	Cliente_102	Madrid	1	17082	Ocasional
1	1	Cliente_1	Valencia	1	15606	Ocasional
181	18	Cliente_181	Barcelona	1	15072	Ocasional
149	14	Cliente_149	Valencia	1	14634	Ocasional
104	10	Cliente_104	Barcelona	1	14235	Ocasional
111	11	Cliente_111	Bilbao	1	14145	Ocasional
22	22	Cliente_22	Madrid	1	12558	Ocasional
195	19	Cliente_195	Bilbao	1	12552	Ocasional
192	19	Cliente_192	Sevilla	1	11490	Ocasional
3	3	Cliente_3	Valencia	1	10860	Ocasional
34	34	Cliente_34	Sevilla	1	10404	Ocasional
139	13	Cliente_139	Madrid	1	9270	Ocasional
133	13	Cliente_133	Sevilla	1	9192	Ocasional
94	94	Cliente_94	Bilbao	1	8820	Ocasional
118	11	Cliente_118	Barcelona	1	8295	Ocasional
185	18	Cliente_185	Valencia	1	8289	Ocasional
128	12	Cliente_128	Madrid	1	7866	Ocasional