**Desafío 1**
**Data Warehouse y Minería de Datos**

Docente:

Karens Medrano

Nombre:

Perla Patricia Rivas Urquilla

Carnet:

RU171368

Grupo:

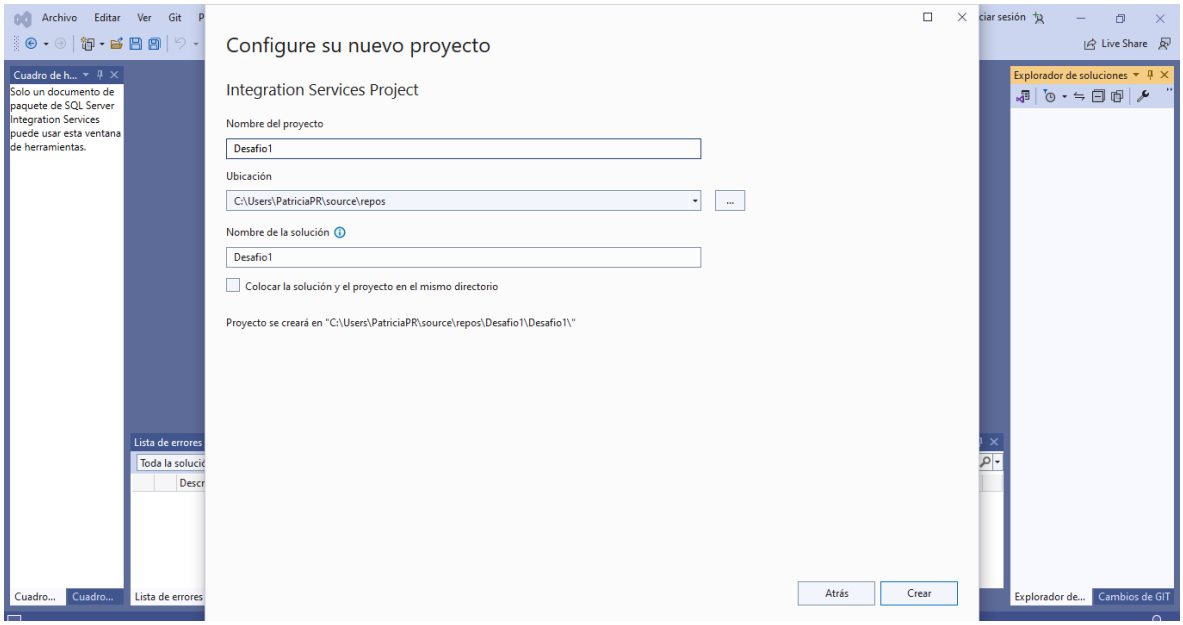
DMD941 G01T

Fecha de entrega:

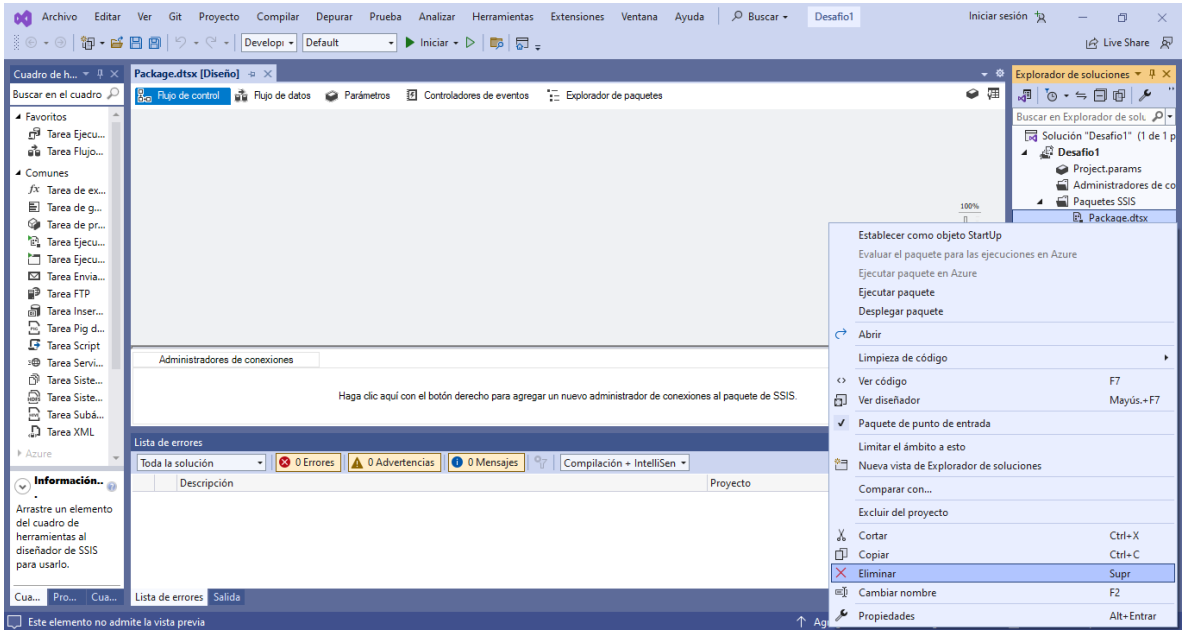
17 de septiembre de 2023

## PROCEDIMIENTOS

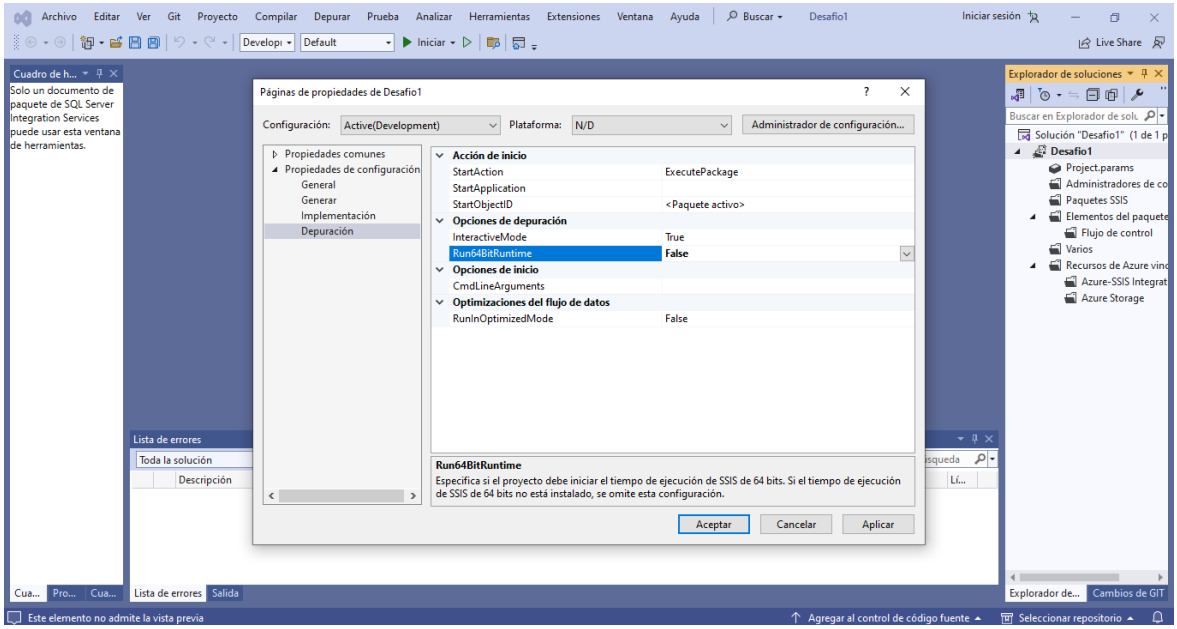
1. Para el desarrollo de los ejercicios comprendidos para el Desafío 1, se crea un proyecto de tipo SSIS en nuestro Visual Studio.




2. En este se crearán los paquetes para los 3 ejercicios necesarios.

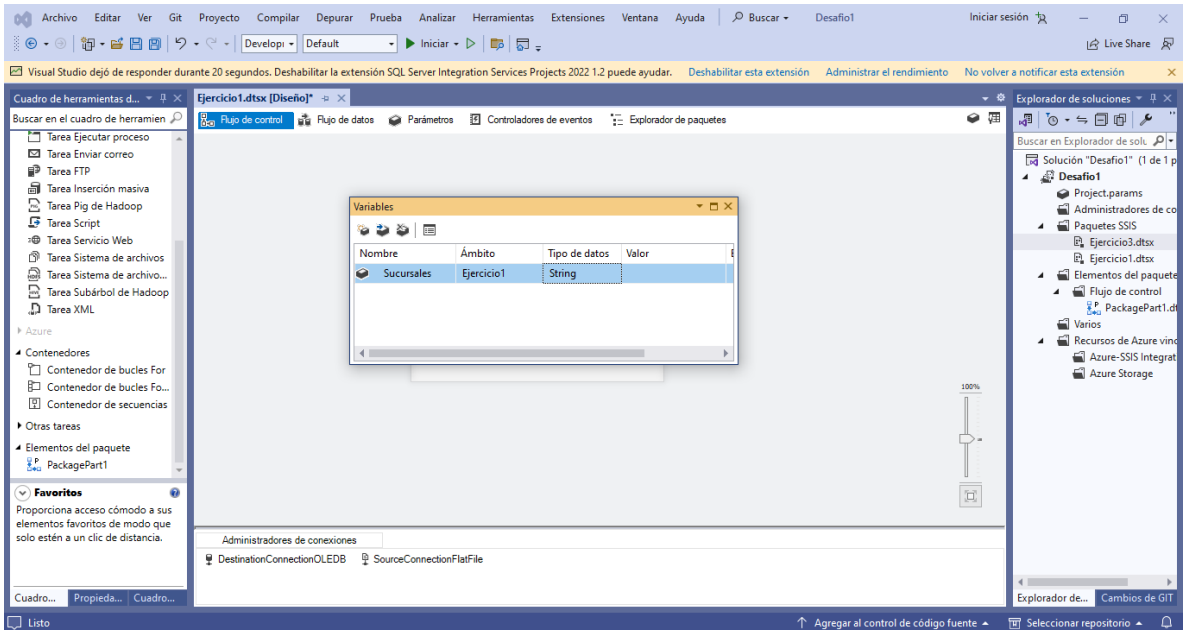


3. Se realizarán los ajustes necesarios para poder ejecutar nuestro proyecto sin errores.

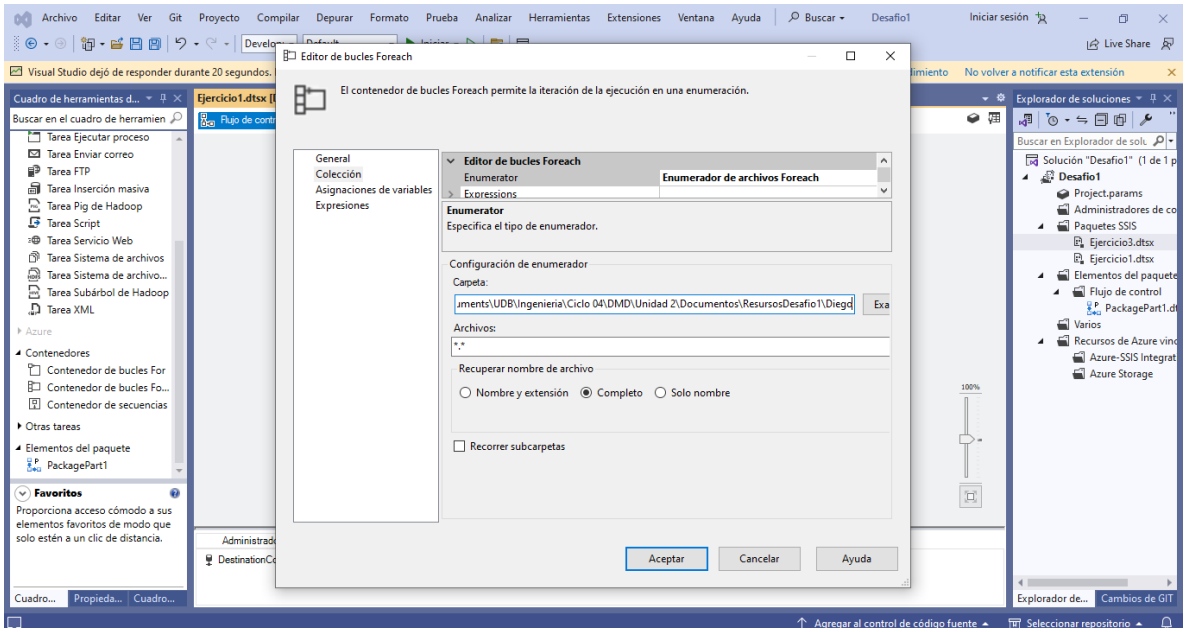



## EJERCICIO 1

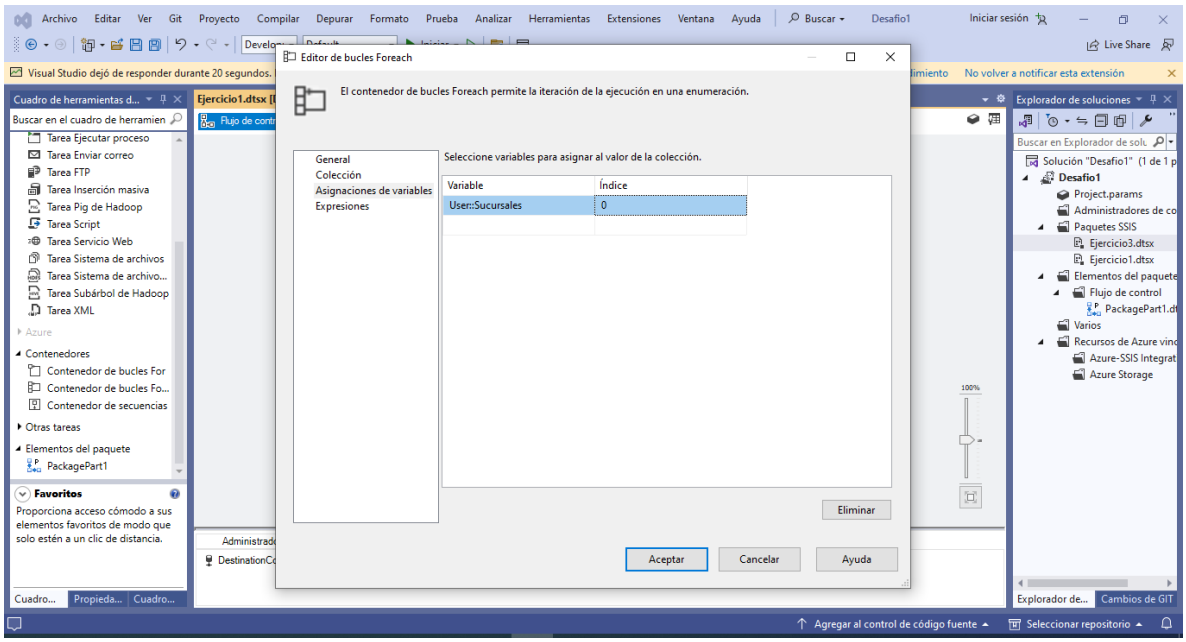
4. Para el primer ejercicio debemos de cargar los archivos csv para poder analizar la data en base las especificaciones del ejercicio, por lo que se cargaran los archivos por medio de un bucle foreach e insertarse en una tabla llamada SpaDiego.
   Para ello deberemos crear una variable dentro de nuestro paquete que nos ayudara a tomar el valor de los archivos dentro de nuestro bucle.



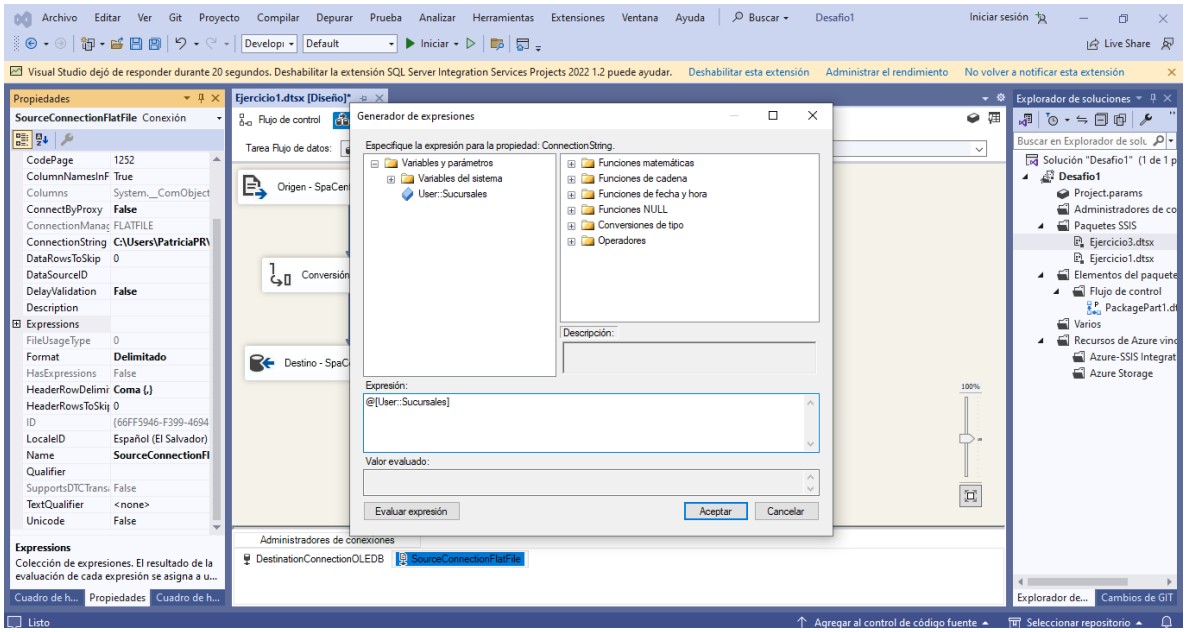
5. Especificamos al bucle la ruta donde se leerán los archivos y en este caso le especificaremos que enumere todos los archivos con extensión csv dentro de nuestra carpeta de recursos del ejercicio.



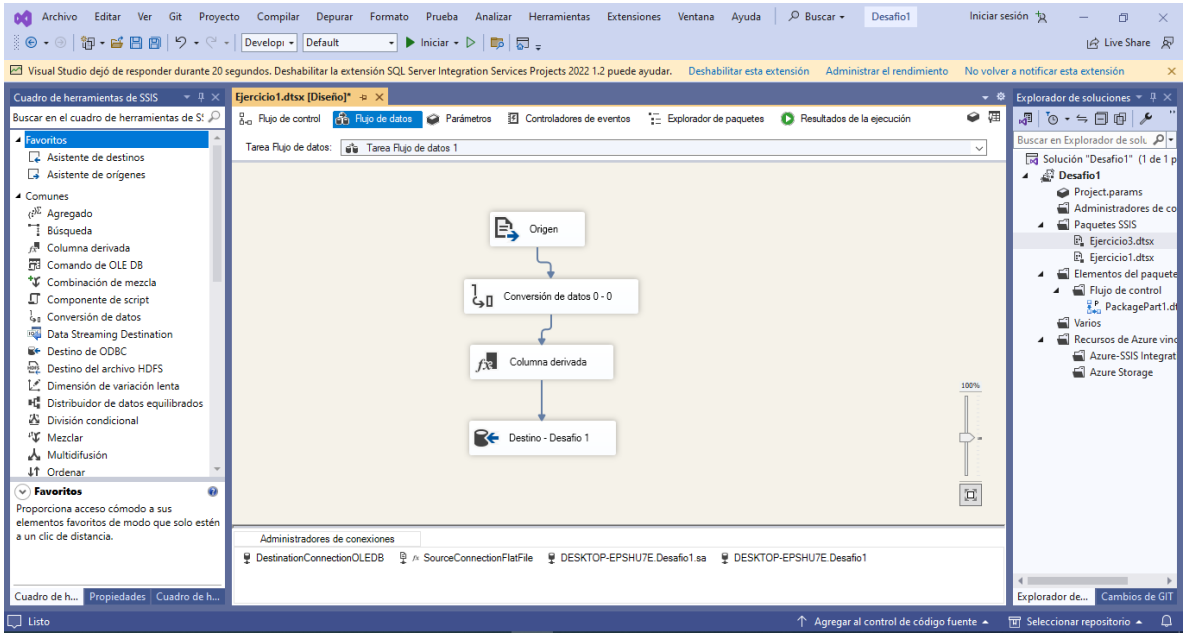

6. Asignamos la variable previamente creada.



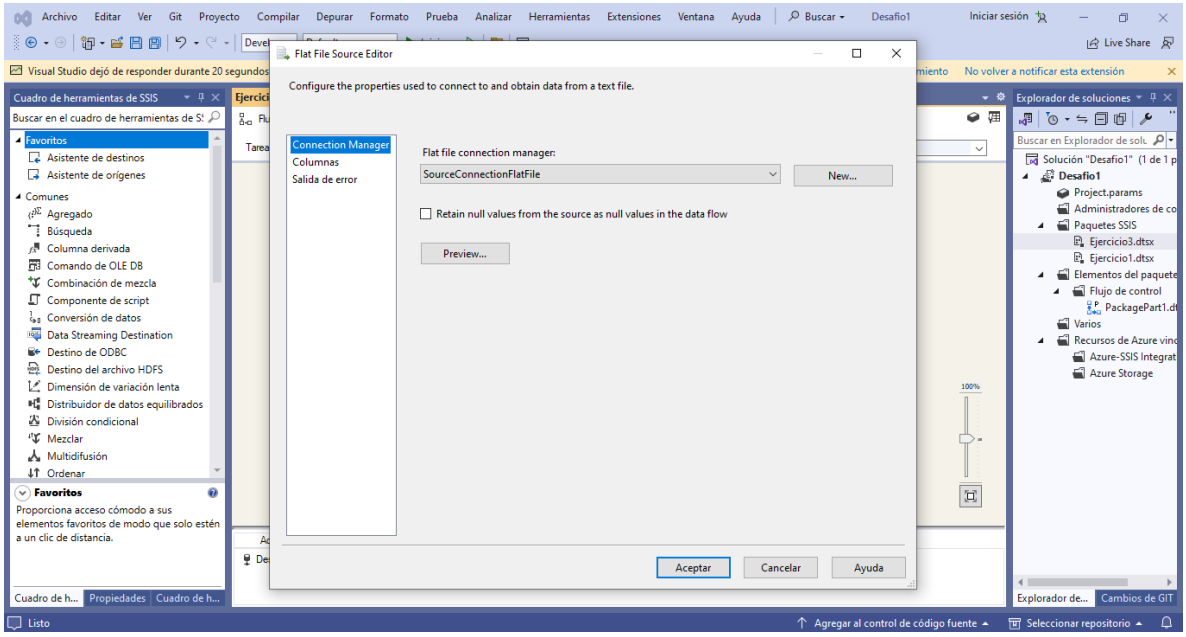
7. Para poder hacer que nuestro origen de texto plano tome el valor de nuestra variable que contiene la información de los archivos, debemos irnos a:
**Administrador de Conexiones > SourceConnectionFlatFile > Propiedades > Expresiones**



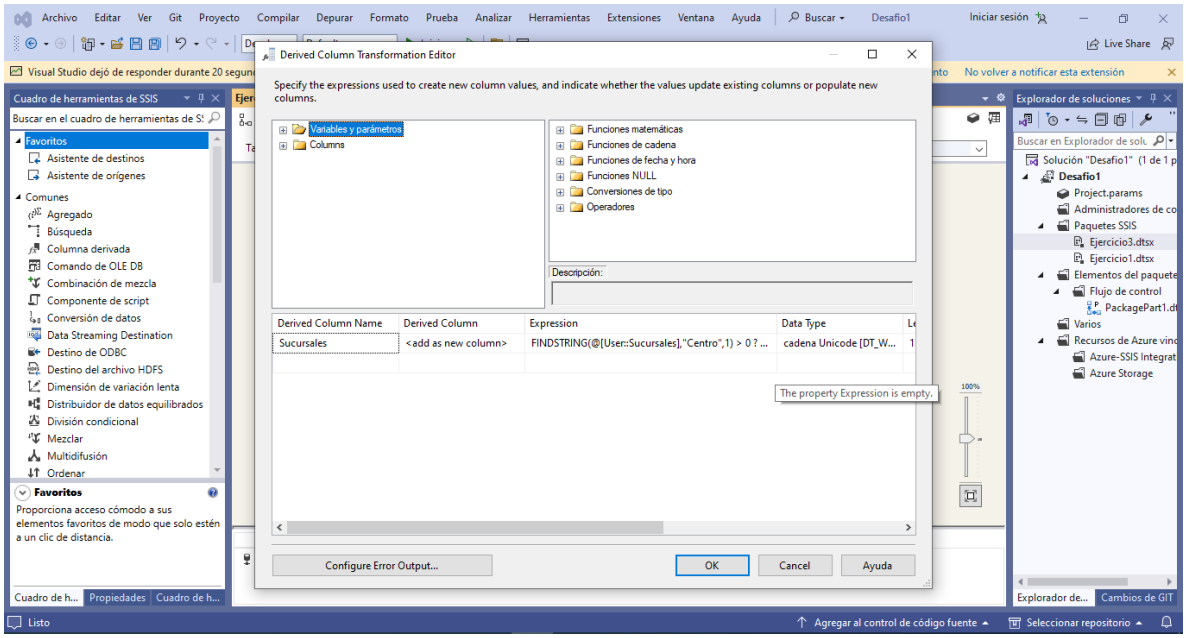

8. El flujo para poder hacer el ingreso de nuestros datos a la base de datos será la siguiente:



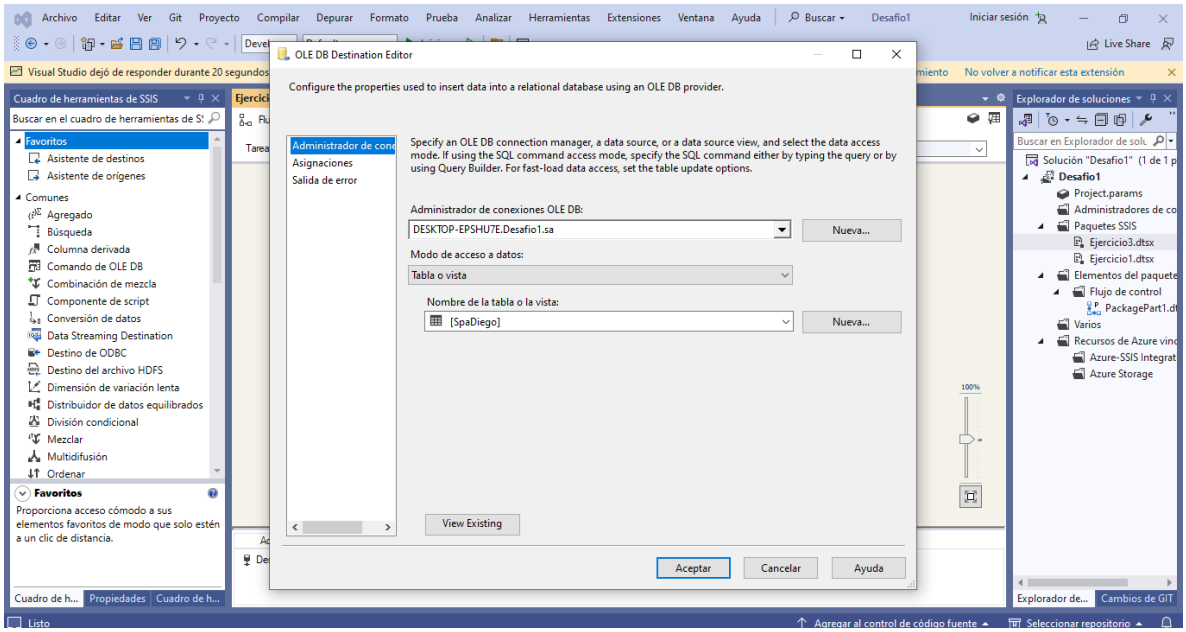
9. Ahora podremos ver como nuestro origen de texto plano tomara el valor de los archivos en base al source.



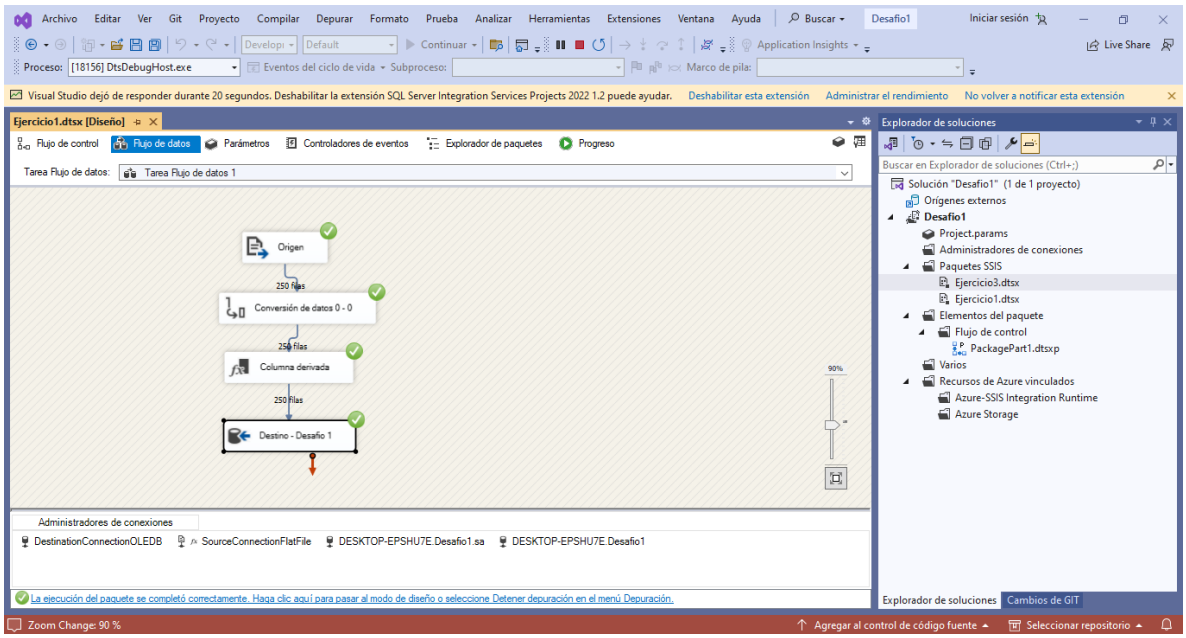

10. Creamos y editamos la herramienta de columna derivada para poder organizar las sucursales que se están ingresando a nuestra tabla.



11. En nuestro destino OLE DB seleccionamos nuestra conexión y tabla a donde se hará el ingreso de nuestros registros de los archivos.




12. Una vez tenemos nuestro flujo configurado, podremos ejecutar el paquete.



13. Ahora podemos ver nuestros registros en la base de datos

```
SELECT *
FROM SpaDiego
```

Results | Messages

| | id | Sexo | ingresos | PromVisit | Edad | Sauna | Masaje | Hidro | Yoga | Sucursal |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Tomkin Stickles | 1 | 2555.23 | 1.94 | 21 | 1 | 1 | 0 | 0 | Centro |
| 2 | Tyson Stovine | 1 | 2476.87 | 5.83 | 29 | 1 | 1 | 0 | 1 | Centro |
| 3 | Miller Camachen | 1 | 1209.36 | 6.07 | 23 | 1 | 0 | 0 | 0 | Centro |
| 4 | Darnell Dine-Hart | 1 | 1307.02 | 3.17 | 63 | 1 | 1 | 1 | 1 | Centro |
| 5 | Wyatt Keyte | 1 | 1511.78 | 2.08 | 41 | 0 | 0 | 0 | 0 | Centro |
| 6 | Trip Vost | 1 | 772.08 | 1.74 | 52 | 1 | 1 | 0 | 0 | Centro |
| 7 | Ammamaria D'Errico | 0 | 2749.35 | 4.58 | 41 | 0 | 0 | 1 | 1 | Centro |
| 8 | Jessica Kuhn | 1 | 1964.11 | 2.3 | 51 | 1 | 1 | 1 | 0 | Centro |
| 9 | Hamnet Lindenberg | 0 | 1577.93 | 2.44 | 62 | 1 | 1 | 1 | 0 | Centro |
| 10 | Marga Kiley | 0 | 1945.16 | 6.36 | 25 | 1 | 1 | 0 | 1 | Centro |
| 11 | Pebrook Praton | 1 | 1917.15 | 5.86 | 37 | 1 | 1 | 0 | 0 | Centro |
| 12 | Koenraad Marchiso | 1 | 699.14 | 2.64 | 34 | 1 | 0 | 1 | 0 | Centro |

Query executed successfully. DESKTOP-EPSHU7E (15.0 RTM) sa (62) Desafio1 00:00:00 700 rows

14. Para poder hacer el análisis, se presentan dos consultas, una de ellas está enfocada en poder clasificar a los clientes en diferentes grupos. Los grupos en los que se clasifican serán Visitas, Servicios e Ingresos. En una clasificación donde 1 es el puntaje mas bajo y 5 el mas alto. Con esto se busca identificar a los clientes que su promedio de visita es mas alto, que adquieren más servicios en sus visitas o incluso a quienes tienen un ingreso superior.

```
SELECT
    Sucursal,ID,
    SEXO,
    Edad,
    NTILE(5) over(order by promVisit) ClasificacionVisitas,
    NTILE(5) over(order by CantServicios) ClasificacionServicios,
    NTILE(5) over(order by ingresos) ClasificacionIngresos
FROM (
    SELECT
        Sucursal,
        COUNT(CASE WHEN VALOR=1 THEN 1 ELSE NULL END) OVER(PARTITION BY ID) CantServicios,
        CASE
        WHEN Sexo=1 THEN 'F' ELSE 'M'
        END Sexo, id, PromVisit, Edad, ingresos--,*
    FROM spaDiego AS U
    UNPIVOT (Valor for Servicio in (Sauna,Masaje,Hidro,Yoga)) AS UPV
) AS A
GROUP BY SUCURSAL,ID, Sexo, Edad, CantServicios, PromVisit, ingresos
ORDER BY Sucursal, sexo, ClasificacionVisitas DESC, ClasificacionServicios DESC , ClasificacionIngresos DESC
```

Results | Messages

| | Sucursal | ID | SEXO | Edad | ClasificacionVisitas | ClasificacionServicios | ClasificacionIngresos |
|---|---|---|---|---|---|---|---|
| 1 | Centro | Tyson Stovine | F | 29 | 5 | 4 | 5 |
| 2 | Centro | Pierre MacDunlevy | F | 27 | 5 | 4 | 5 |
| 3 | Centro | Rosemaria McCloughen | F | 50 | 5 | 4 | 4 |
| 4 | Centro | Ruggiero Brosnan | F | 45 | 5 | 4 | 4 |
| 5 | Centro | Kathe Deegin | F | 22 | 5 | 4 | 3 |
| 6 | Centro | Cointon Tigner | F | 64 | 5 | 4 | 3 |
| 7 | Centro | Pebrook Praton | F | 37 | 5 | 4 | 3 |
| 8 | Centro | Christal McFfaden | F | 54 | 5 | 4 | 1 |

Query executed successfully. DESKTOP-EPSHU7E (15.0 RTM) sa (62) Desafio1 00:00:00 700 rows

Para la segunda consulta tenemos un enfoque diferente, segmentado por género y rango de edades, este busca identificar los grupos que tienen mayor porcentaje de representación en nuestras sucursales; de esta forma podremos evaluar si en una sucursal las campañas deben de ir enfocadas a mujeres u hombres, al igual que conocer los grupos etarios de los que se tienen mayor afluencia.

```
SELECT
    Sucursal,
    CASE
    WHEN Sexo=1 THEN 'F' ELSE 'M'
    END Sexo,
    concat((COUNT(Sexo)*100) / sum(COUNT(Sexo)) over(partition by sucursal), '%')  PorcentajeClientes,
    avg(ingresos) PromCompras,
    AVG(PromVisit) PromVisitas,
    CONCAT(MIN(Edad),'-',MAX(EDAD)) RangoEdades
FROM SpaDiego AS U
GROUP BY Sucursal, Sexo, CASE WHEN Edad BETWEEN 20 AND 45 THEN '20-45' ELSE '46-70' END
ORDER BY Sucursal, Sexo, PorcentajeClientes, PromVisitas DESC
```

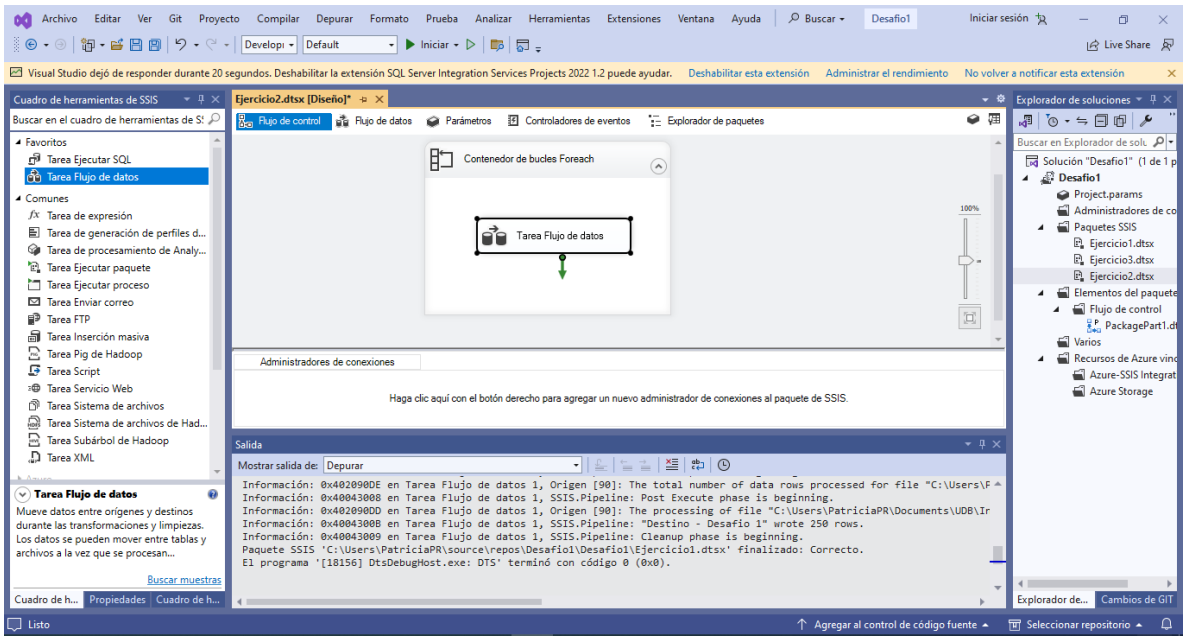
Results | Messages

| | Sucursal | Sexo | PorcentajeClientes | PromCompras | PromVisitas | RangoEdades |
|---|---|---|---|---|---|---|
| 1 | Centro | F | 34% | 1885.74529411765 | 4.10823529411765 | 21-45 |
| 2 | Centro | F | 38% | 1505.18894736842 | 3.64 | 46-64 |
| 3 | Centro | M | 12% | 1673.81666666667 | 4.56166666666667 | 48-62 |
| 4 | Centro | M | 16% | 2094.955 | 3.805 | 22-41 |
| 5 | Escalon | F | 24% | 1739.73226804124 | 3.83979381443299 | 46-65 |
| 6 | Escalon | F | 29% | 1684.26939655172 | 3.96991379310345 | 21-45 |
| 7 | Escalon | M | 19% | 1655.35197368421 | 4.15026315789474 | 46-65 |
| 8 | Escalon | M | 27% | 1718.50261261261 | 3.92162162162162 | 21-45 |
| 9 | Santa Tecla | F | 10% | 1568.2456 | 3.34 | 47-64 |
| 10 | Santa Tecla | F | 12% | 1886.32 | 4.22866666666667 | 21-45 |
| 11 | Santa Tecla | M | 29% | 1602.98082191781 | 3.89465753424658 | 46-65 |
| 12 | Santa Tecla | M | 48% | 1597.36450819672 | 3.79836065573771 | 21-45 |

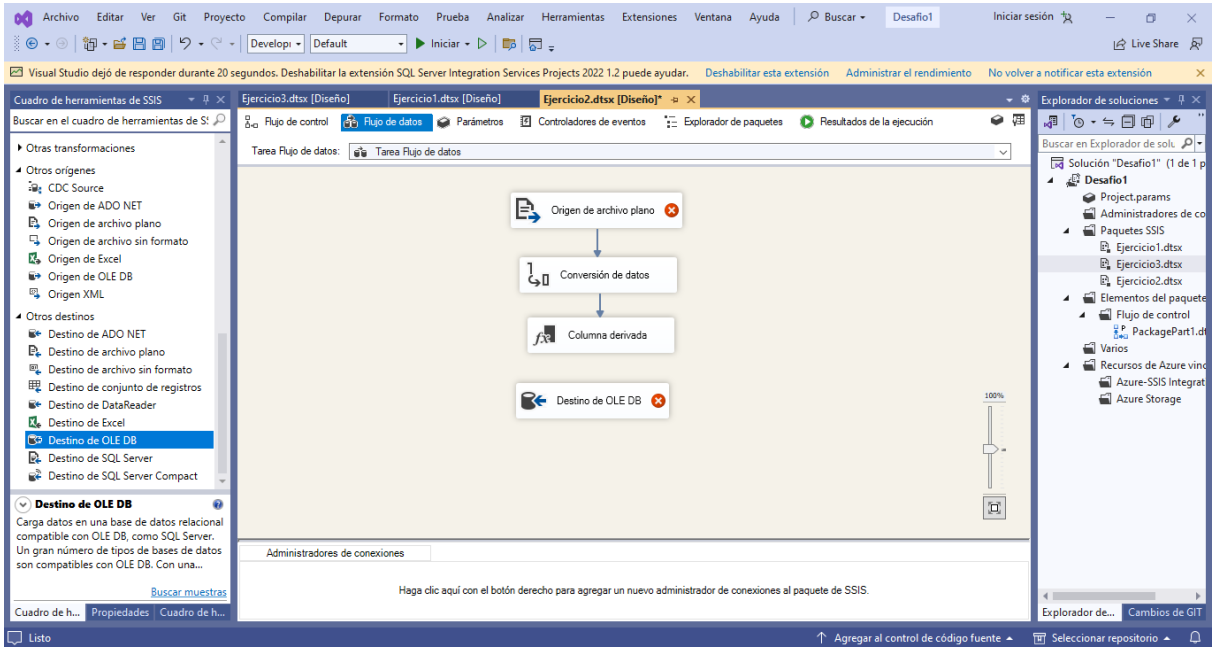
Query executed successfully. DESKTOP-EPSHU7E (15.0 RTM) sa (62) Desafio1 00:00:00 12 rows

## EJERCICIO 2

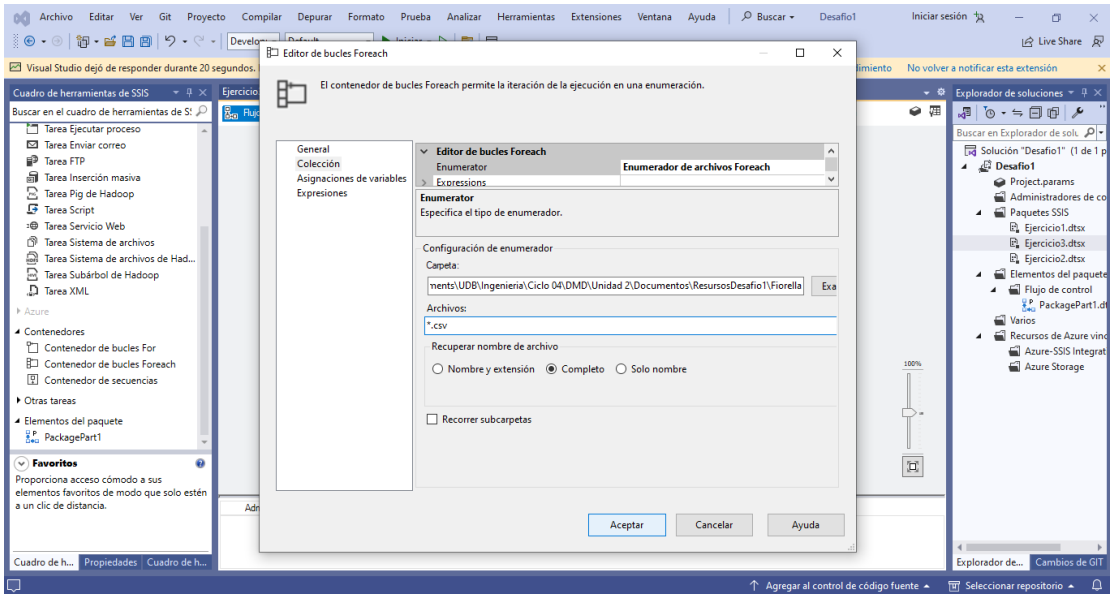
15. Para el ejercicio 2 realizaremos un procedimiento similar al ejercicio 1. Haremos un bucle foreach y dentro de este nuestra tarea de flujo.

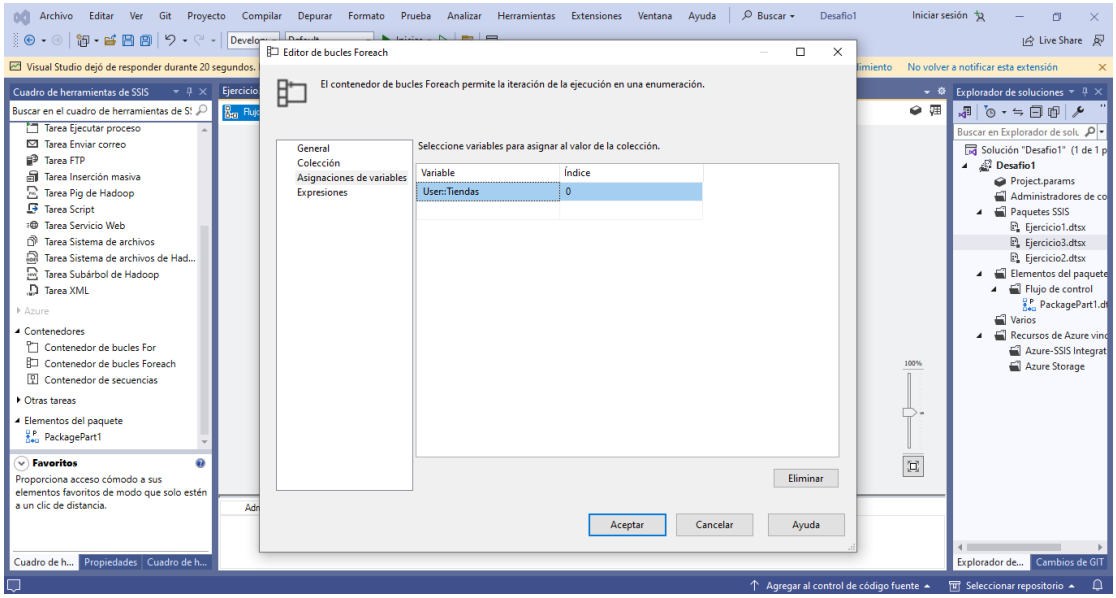



16. Dentro de nuestro flujo de datos, tendremos los siguientes pasos.

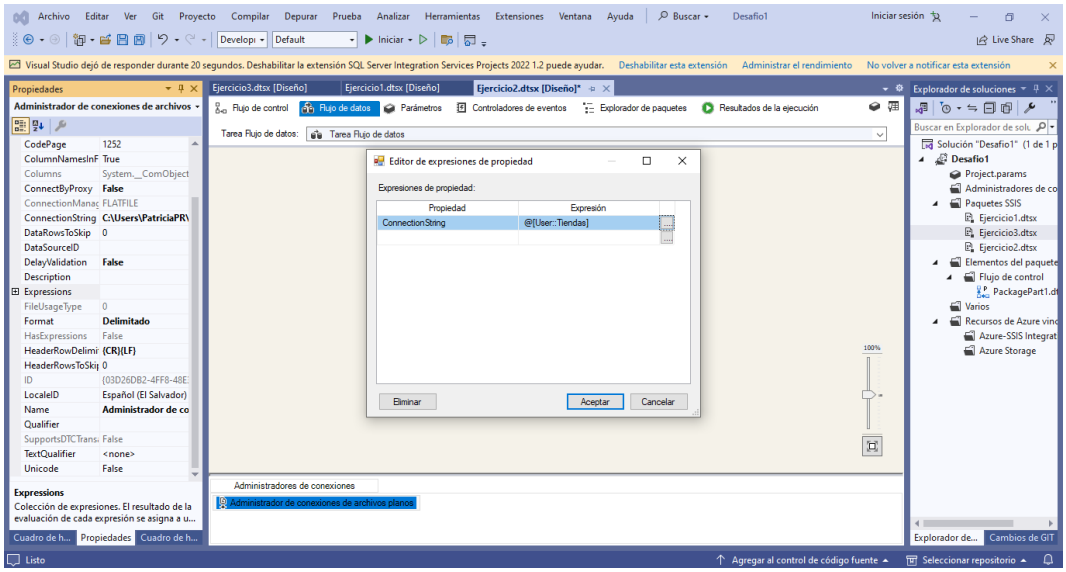



17. Seleccionamos la carpeta correspondiente a los archivos que debemos cargar para el ejercicio y asignamos la variable para guardar la información del recorrido.

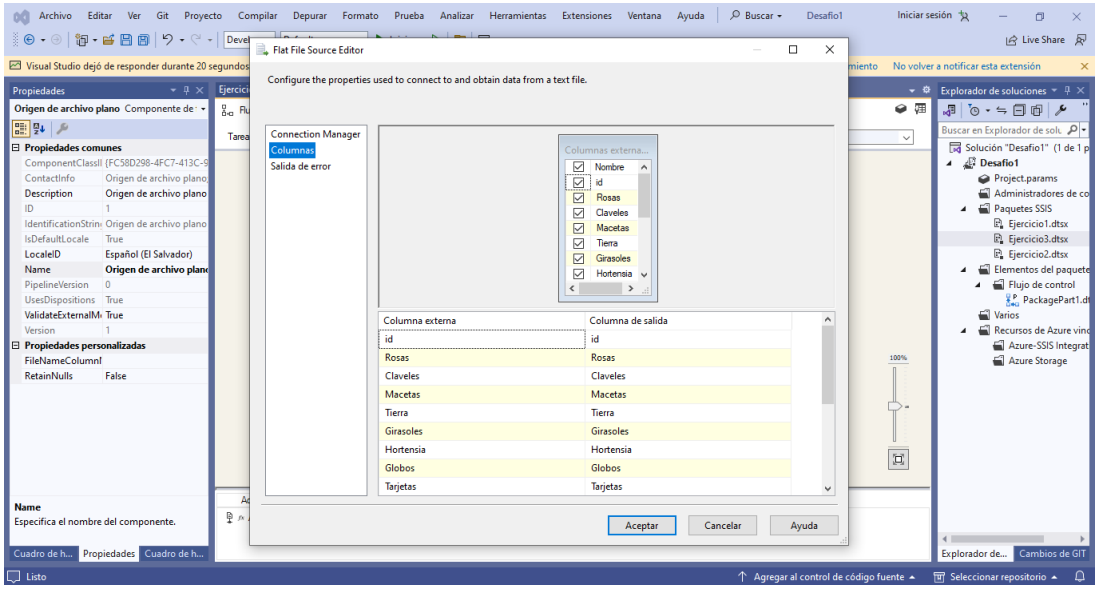


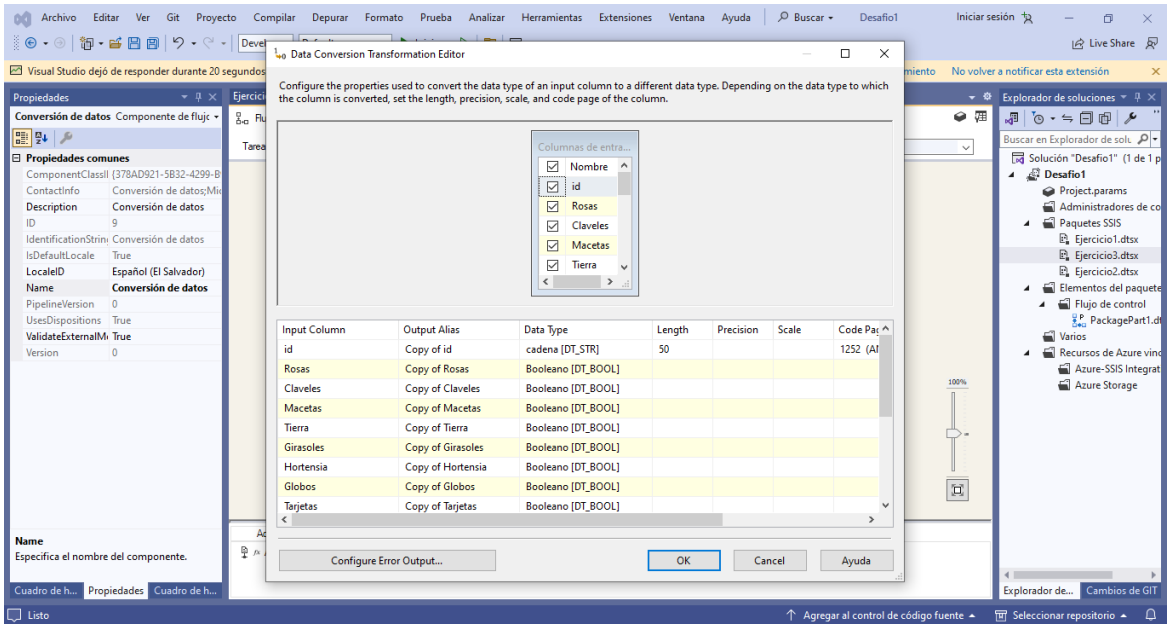

18. Editamos la expresión de configuración de la conexión de archivo de texto plano, para que esta obtenga los datos de la variable.



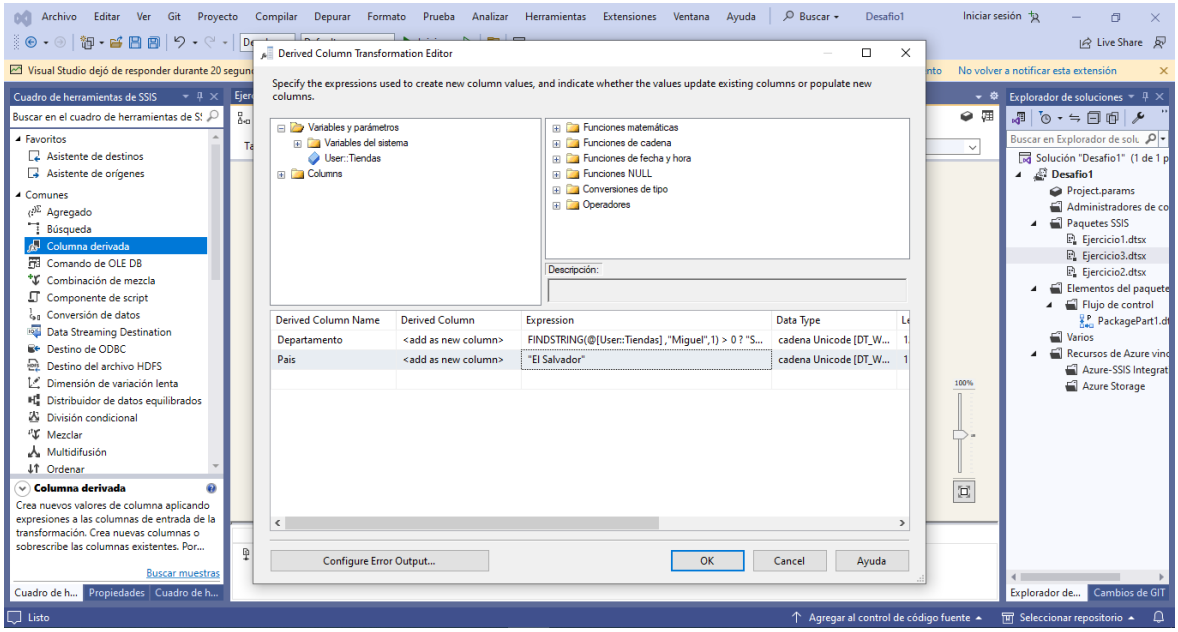

19. De esta forma nuestro origen ya podrá tener las columnas correspondientes a nuestros archivos.



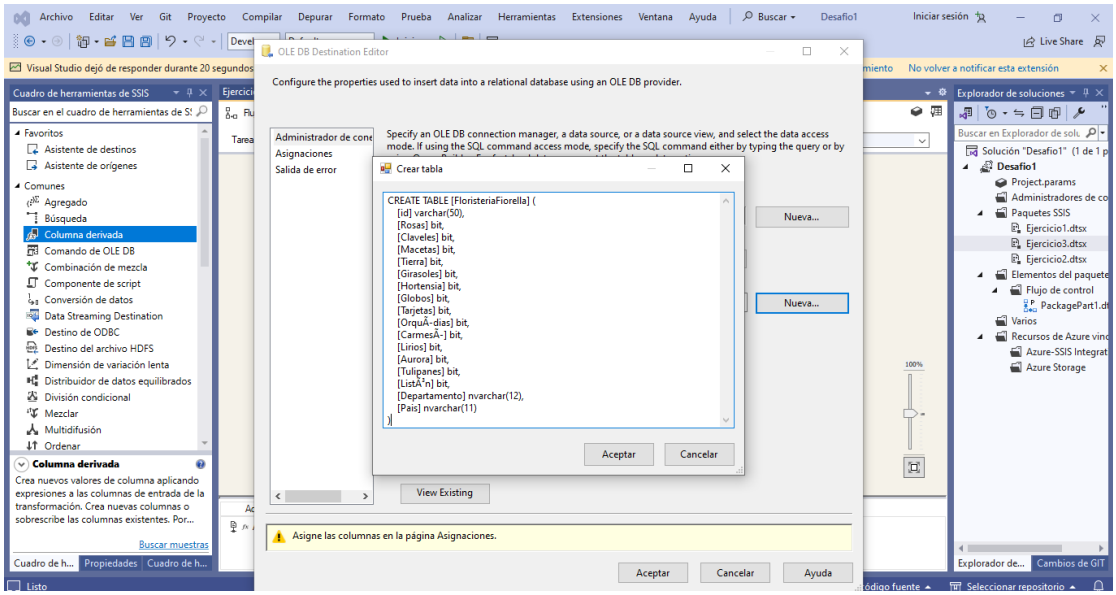

20. Hacemos la conversión para que los valores de las columnas de detalle de los productos sean de tipo booleano.



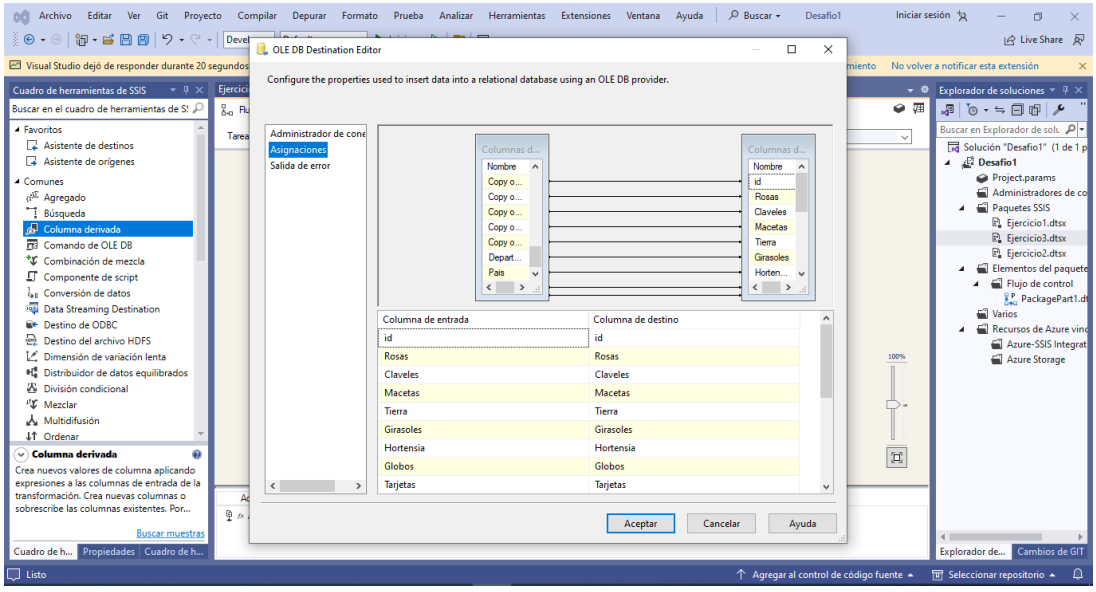

21. Agregaremos la columna de departamento que se agregara en base al nombre del documento de donde se están tomando los datos para el ingreso. La columna de país la dejaremos como un dato quemado para los 3 departamentos que estamos ingresando que serán de “El Salvador”.



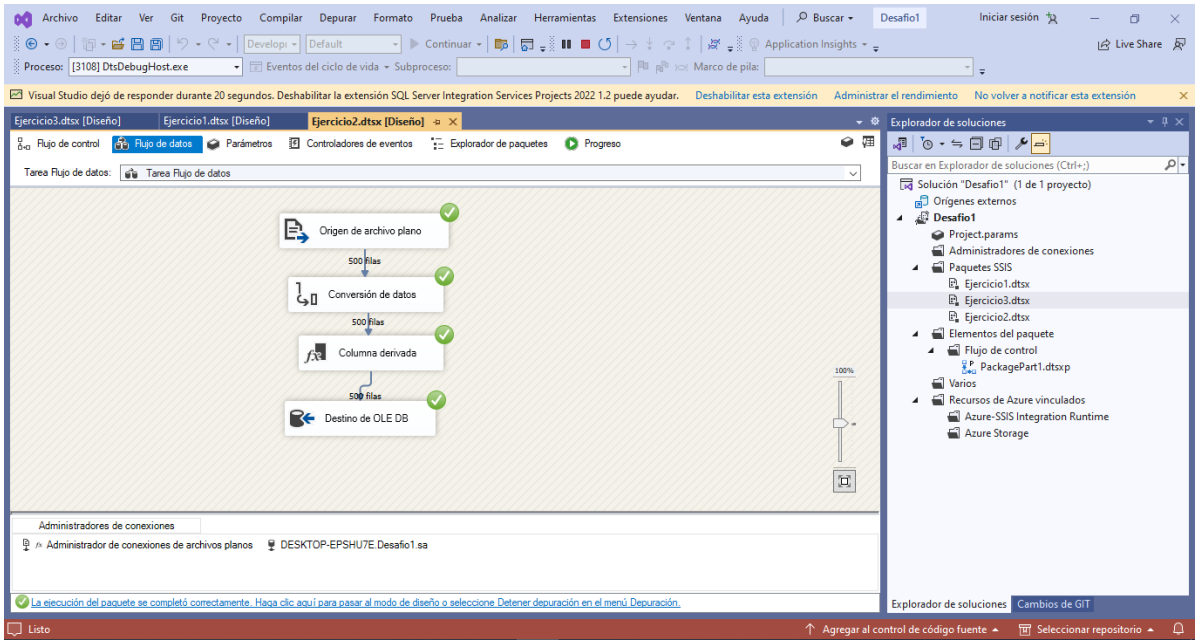

22. En el destino OLE DB crearemos la tabla donde se almacenarán nuestros registros.



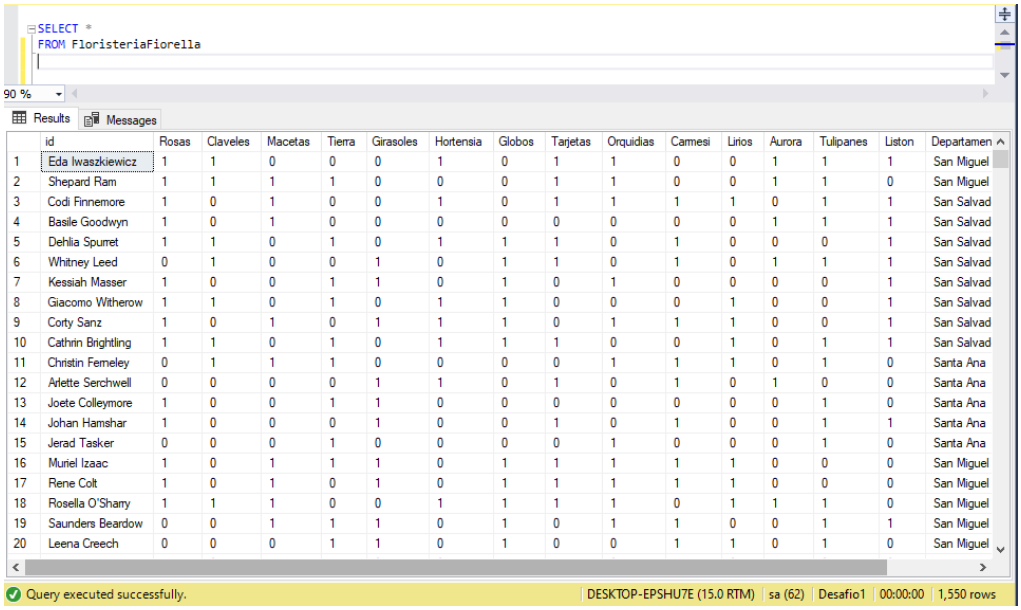

23. Podremos visualizar la entrada y salida correspondiente de los datos.




24. Ejecutamos el paquete para poder cargar los datos a nuestra tabla en SQL.




25. Con esto ya podremos realizar las consultas necesarias en nuestro SQL.



```
SELECT *
FROM FloristeriaFiorella
```

Results | Messages

| | id | Rosas | Claveles | Macetas | Tierra | Girasoles | Hortensia | Globos | Tarjetas | Orquidias | Carnesi | Lirios | Aurora | Tulipanes | Liston | Departamen |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Eda Iwaszkiewicz | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | San Miguel |
| 2 | Shepard Ram | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | San Miguel |
| 3 | Codi Finnemore | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | San Salvad |
| 4 | Basile Goodwyn | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | San Salvad |
| 5 | Dehlia Spurret | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | San Salvad |
| 6 | Whitney Leed | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | San Salvad |
| 7 | Kessiah Masser | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | San Salvad |
| 8 | Giacomo Witherow | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | San Salvad |
| 9 | Corty Sanz | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | San Salvad |
| 10 | Cathrin Brightling | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | San Salvad |
| 11 | Christin Ferneley | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | Santa Ana |
| 12 | Arlette Serchwell | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | Santa Ana |
| 13 | Joete Colleymore | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | Santa Ana |
| 14 | Johan Hamshar | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | Santa Ana |
| 15 | Jerad Tasker | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | Santa Ana |
| 16 | Muriel Izaac | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | San Miguel |
| 17 | Rene Colt | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | San Miguel |
| 18 | Rosella O'Sharry | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | San Miguel |
| 19 | Saunders Beardow | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | San Miguel |
| 20 | Leena Creech | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | San Miguel |

Query executed successfully. | DESKTOP-EPSHU7E (15.0 RTM) | sa (62) | Desafio1 | 00:00:00 | 1,550 rows

26. Para el análisis se muestra un conteo de ventas agrupado por tipo de articulo de compras y su detalle por cada departamento. Con esto se podrá determinar los departamentos que más venden y los productos que mas se compran en dichos departamentos.

```sql
SELECT
    Pais,
    Articulo,
    [San Miguel],
    [San Salvador],
    [Santa ana]
FROM (SELECT
        Pais,
        Articulo,
        Articulo art,
        UNPV.Departamento,
        Compras
    FROM FloristeriaFiorella AS A
    UNPIVOT (Compras FOR Articulo IN ([Rosas],[Claveles],[Macetas],[Tierra],[Girasoles],[Hortensia]
          ,[Globos],[Tarjetas],[Orquidias],[Carmesi],[Lirios],[Aurora],[Tulipanes],[Liston]))
    AS UNPV
    WHERE Compras=1) AS B
PIVOT(count(art) FOR Departamento in ([San Miguel], [San Salvador], [Santa ana])) as pvt
ORDER BY [San Salvador] DESC, [San Miguel] DESC, [Santa ana]
```

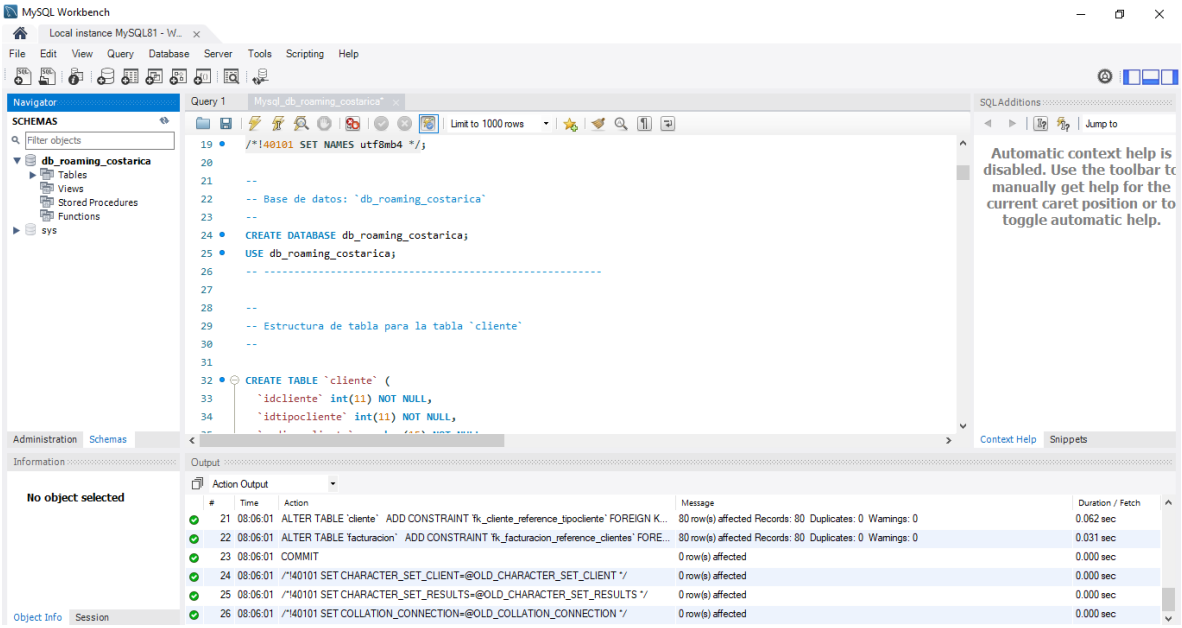
Results | Messages

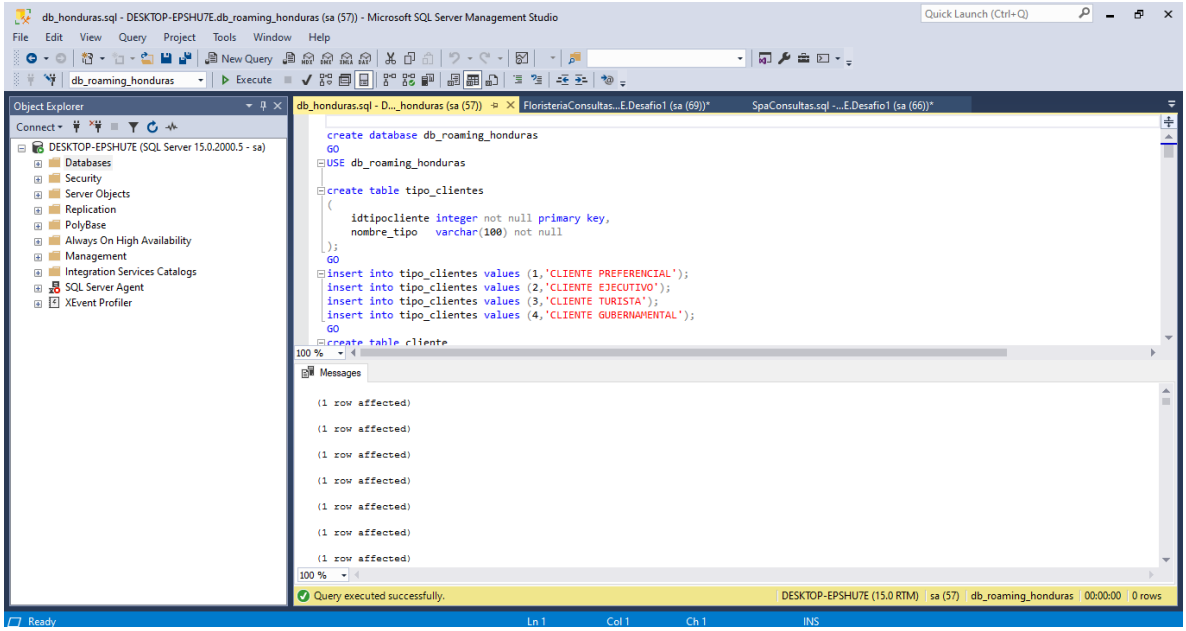
| | Pais | Articulo | San Miguel | San Salvador | Santa ana |
|---|---|---|---|---|---|
| 1 | El Salvador | Liston | 149 | 690 | 136 |
| 2 | El Salvador | Rosas | 157 | 612 | 176 |
| 3 | El Salvador | Globos | 151 | 587 | 154 |
| 4 | El Salvador | Macetas | 141 | 392 | 245 |
| 5 | El Salvador | Aurora | 160 | 384 | 260 |
| 6 | El Salvador | Tarjetas | 143 | 384 | 252 |
| 7 | El Salvador | Orquidias | 158 | 380 | 259 |
| 8 | El Salvador | Hortensia | 157 | 374 | 243 |
| 9 | El Salvador | Girasoles | 150 | 371 | 266 |
| 10 | El Salvador | Tierra | 141 | 368 | 236 |

Query executed successfully. | DESKTOP-EPSHU7E (15.0 RTM) | sa (62) | Desafio1 | 00:00:00 | 14 rows

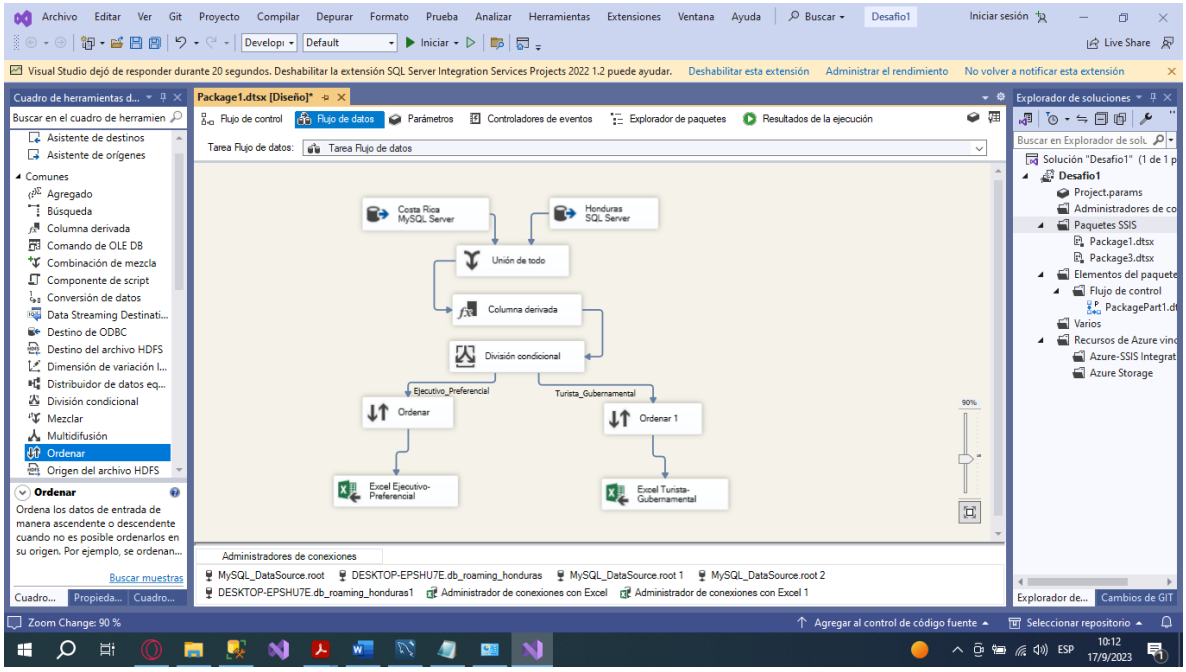
# EJERCICIO 3

27. Para el ejercicio 3 debemos crear las bases de datos en MySQL y SQL Server.

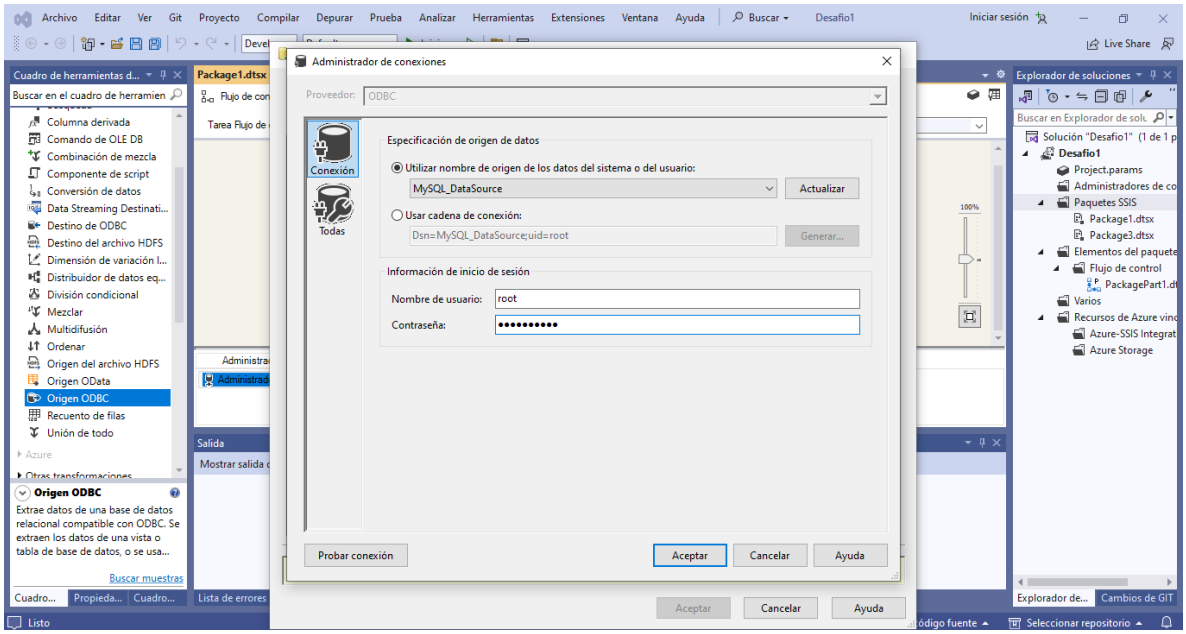




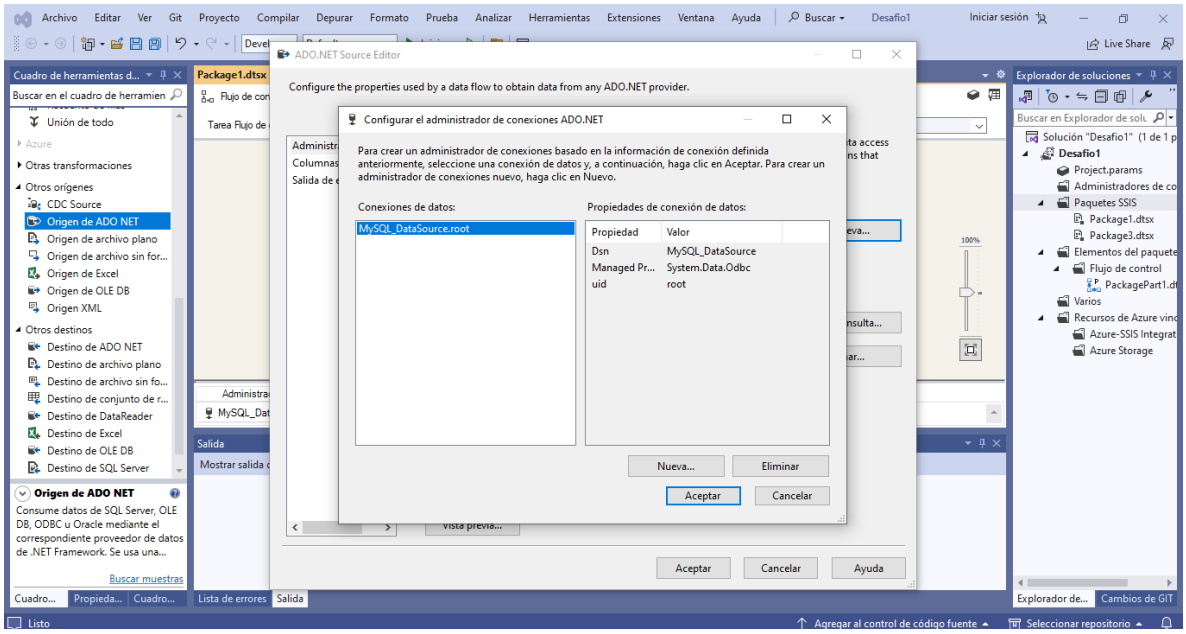

28. Formaremos el siguiente flujo de datos y procederemos a realizar las configuraciones correspondientes a cada uno de los pasos del flujo.



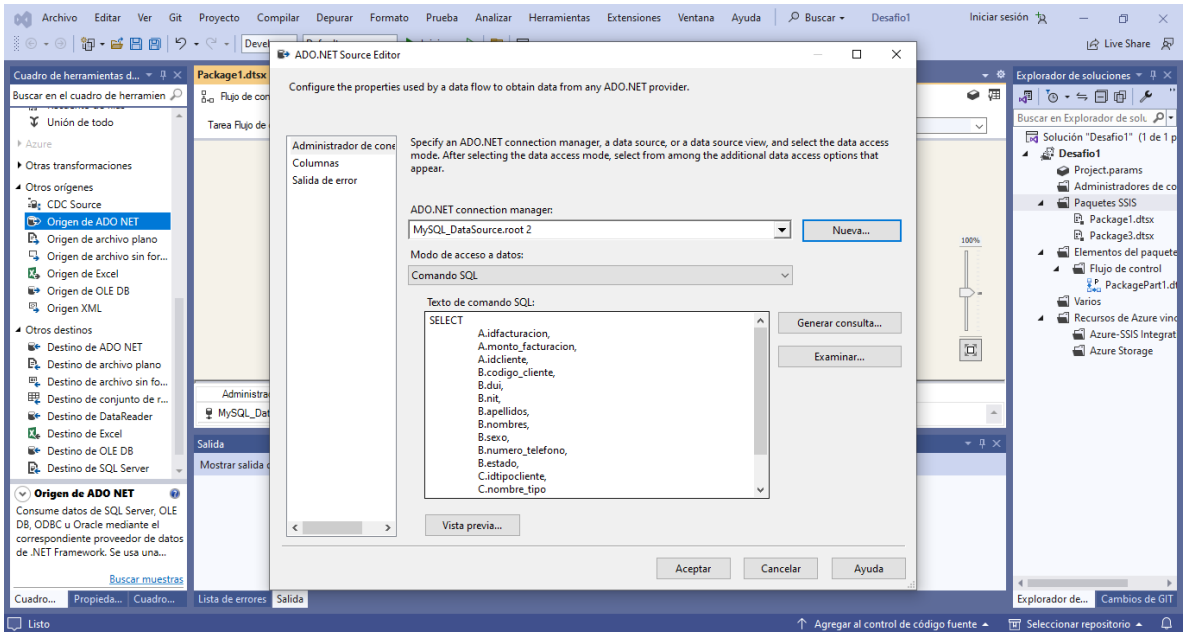
29. Empezaremos por crear nuestro origen de datos ADO.NET para nuestra conexión a la base de datos de MySQL.



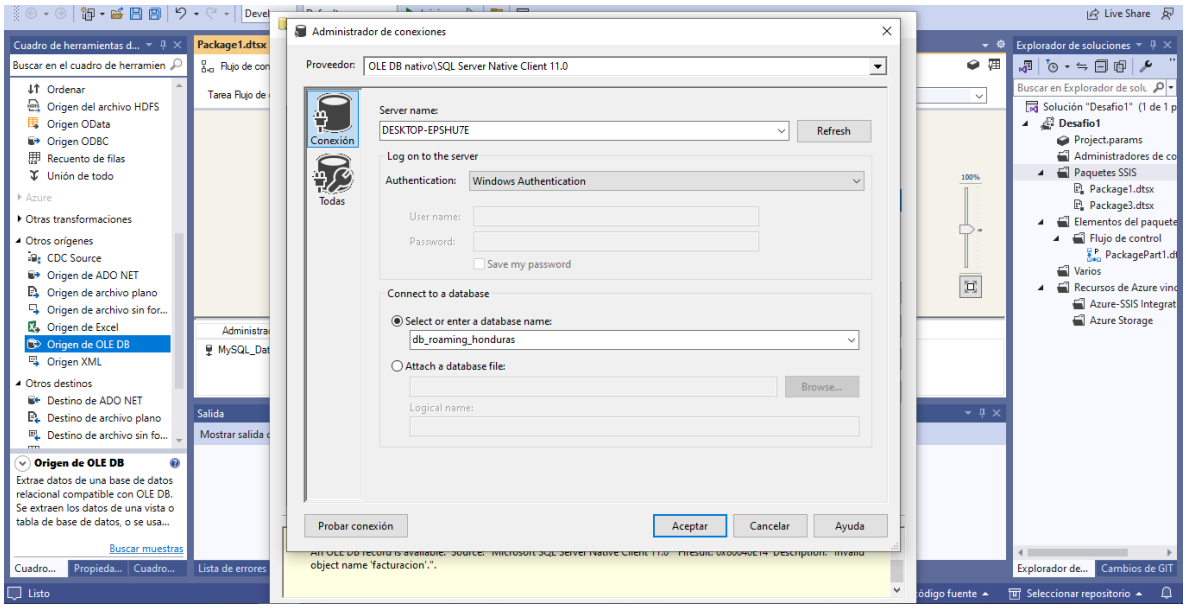

30. Para esto hay que tener previamente configurado nuestro DataSource con los drivers correspondientes. Y configurar la conexión en administrador de conexiones.

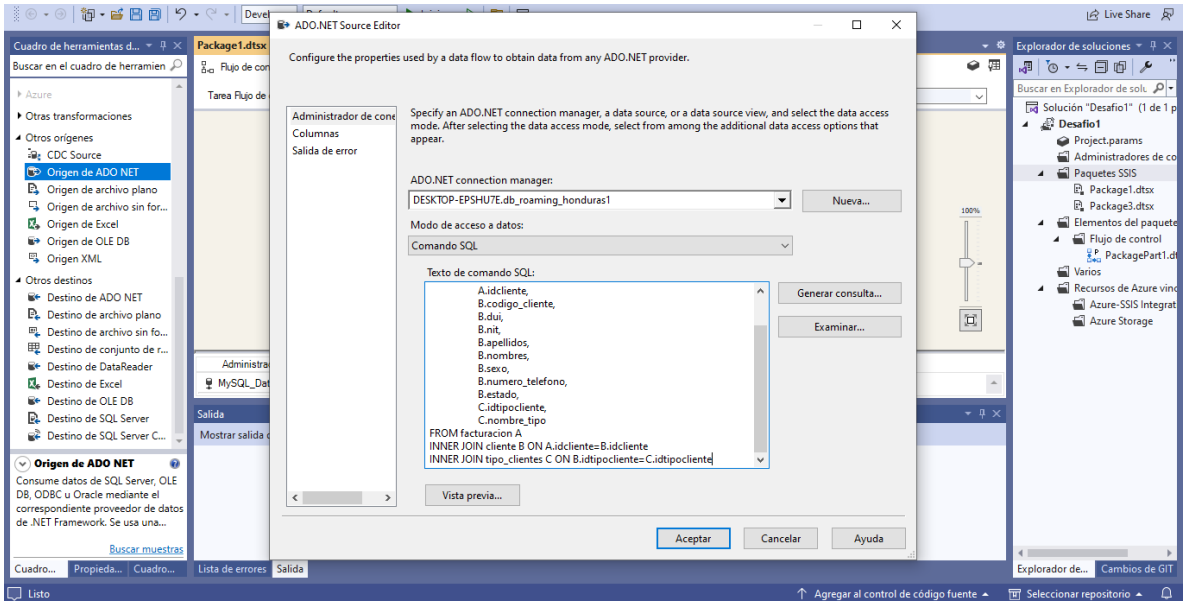



31. Para poder acceder a la información de las tablas, realizaremos una consulta con un INNER JOIN con la información de tipo de cliente y los datos de facturación.

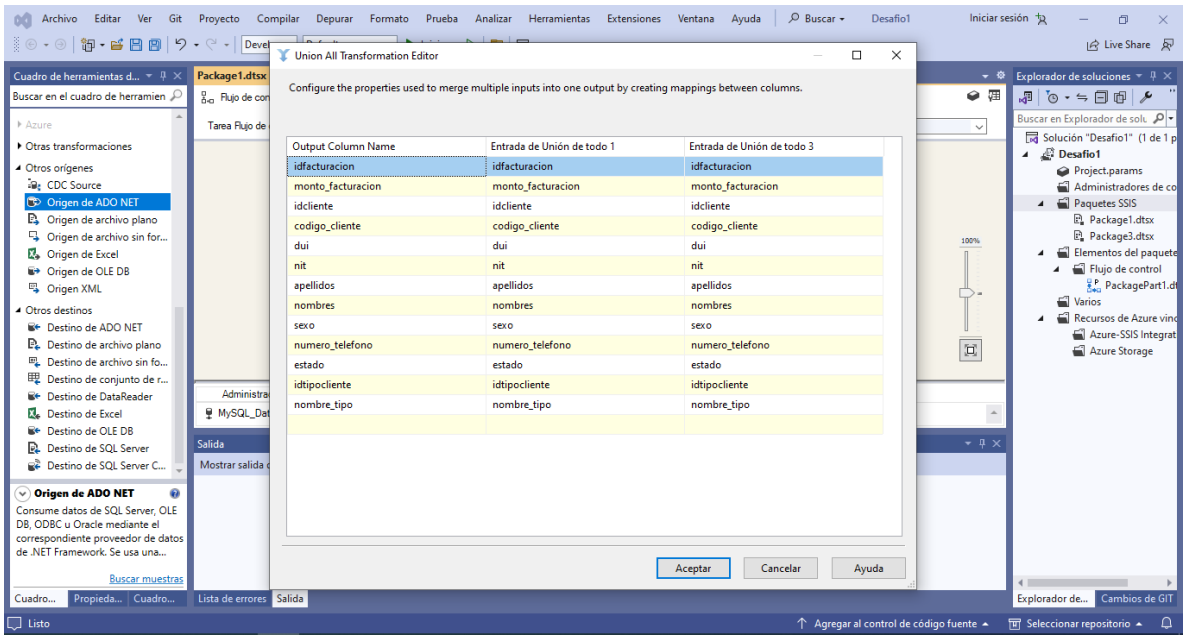



32. Realizamos la misma configuración de origen ADO.NET para los datos de la base de datos SQL Server.

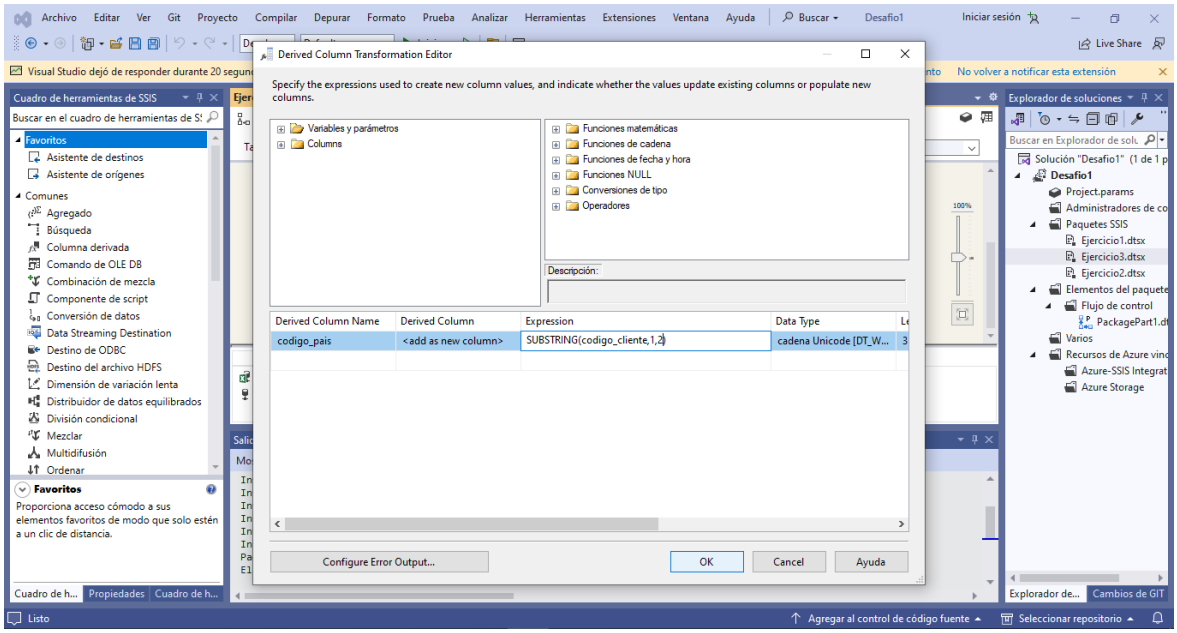




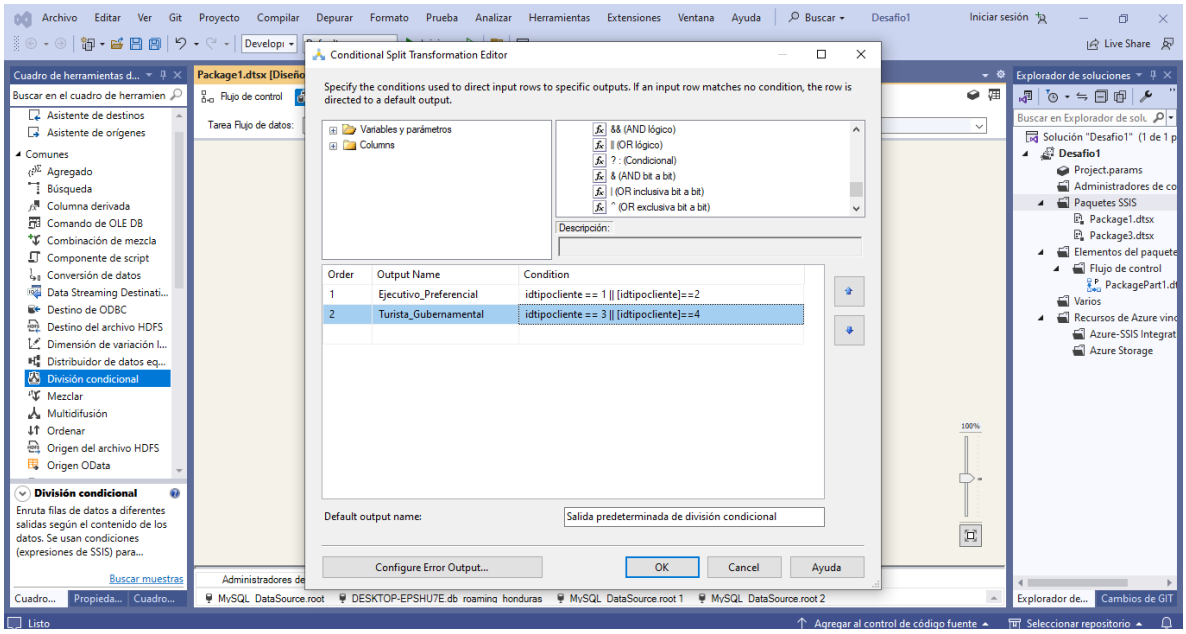

33. Ahora que tenemos la información de ambas bases de datos haremos una unión de todos los datos.



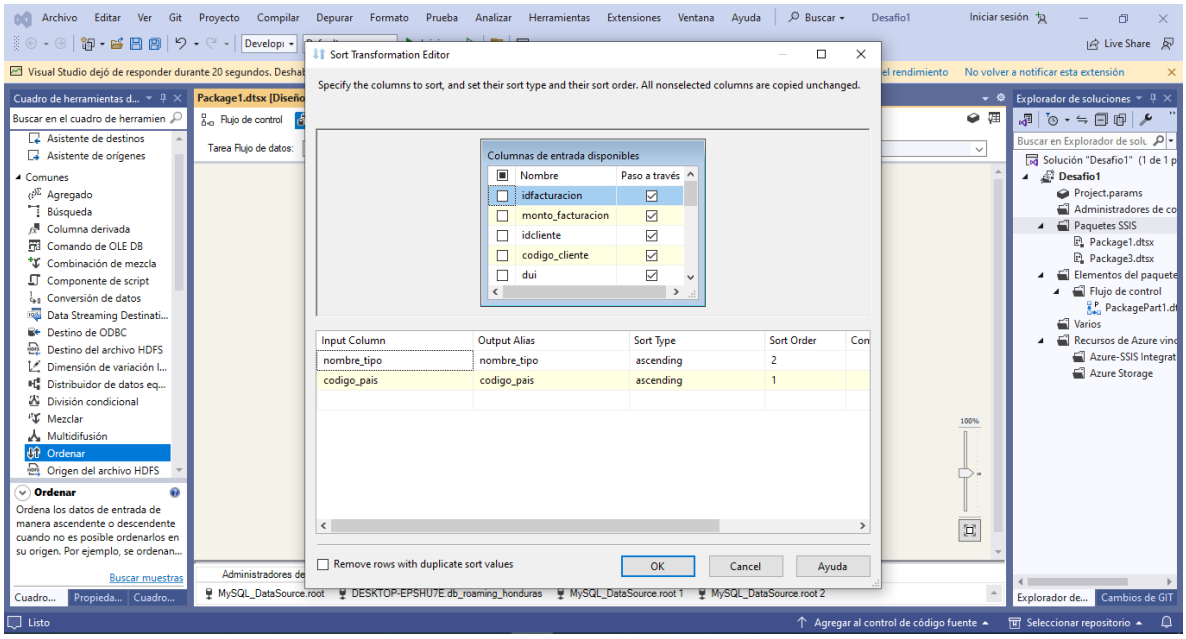

34. También agregaremos una columna derivada para poder sacar el código de país. Esta saldrá de los primeros 2 caracteres de la columna código de cliente.



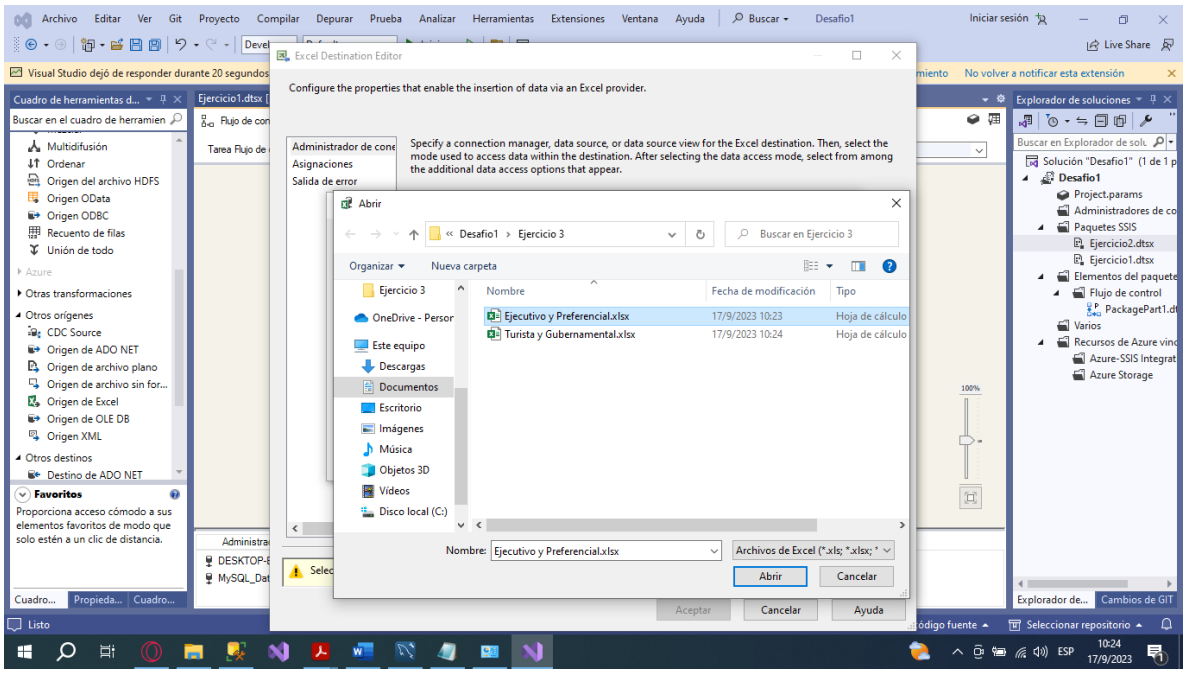

35. Para nuestro condicional agregaremos las siguientes expresiones de condición.

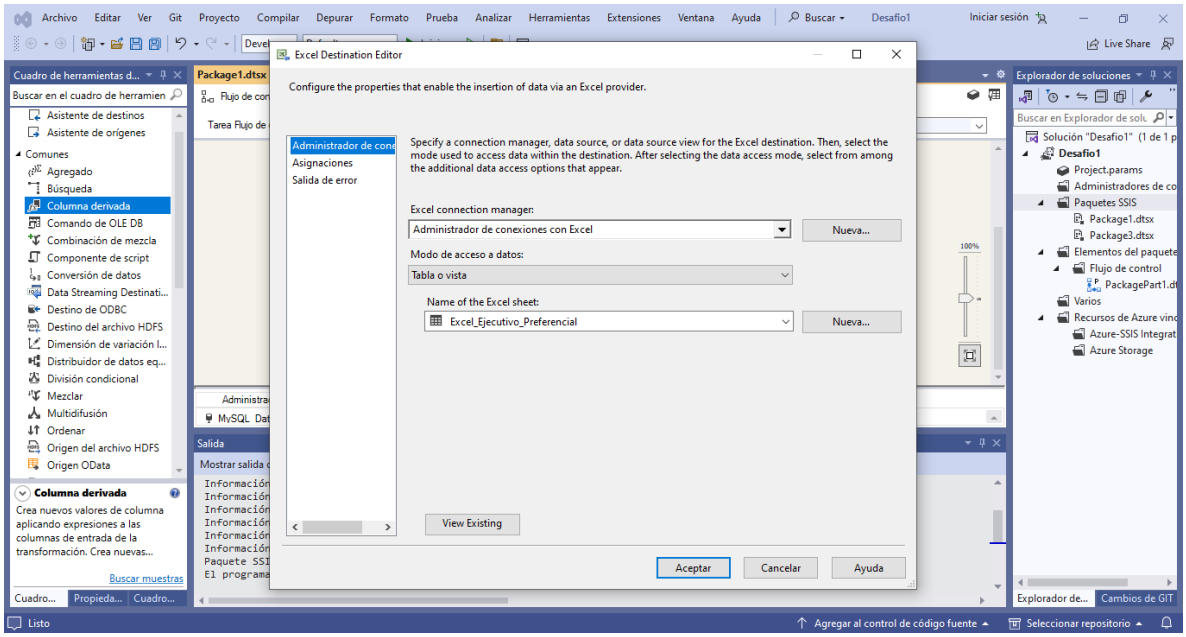



36. También haremos un ordenamiento de los datos por país y tipo de cliente.

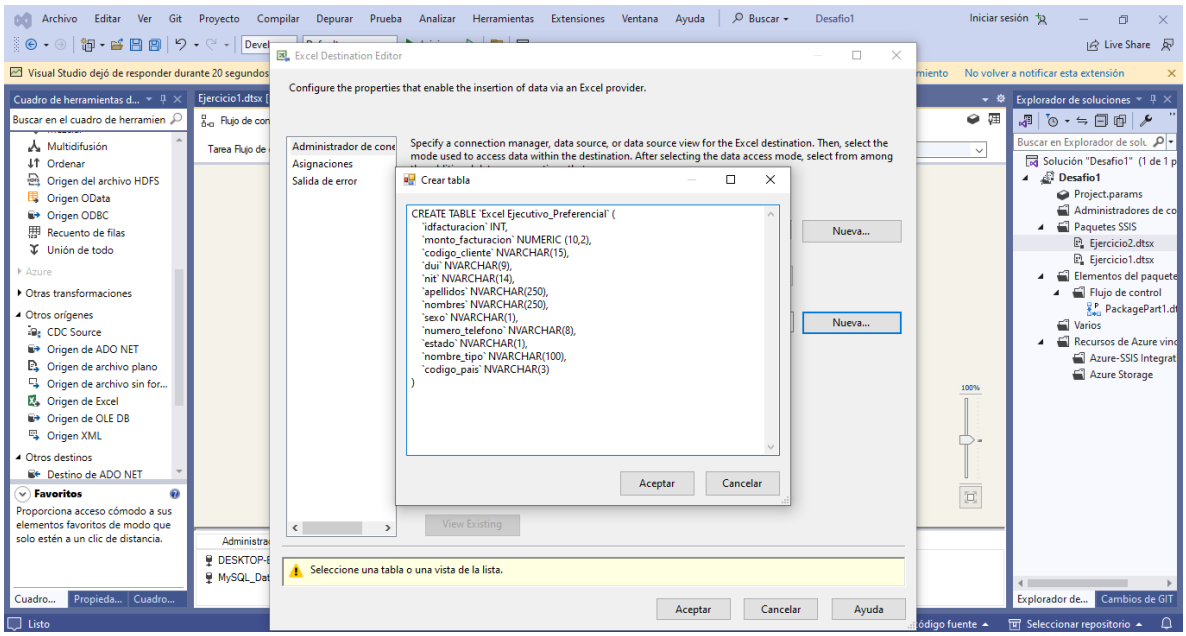


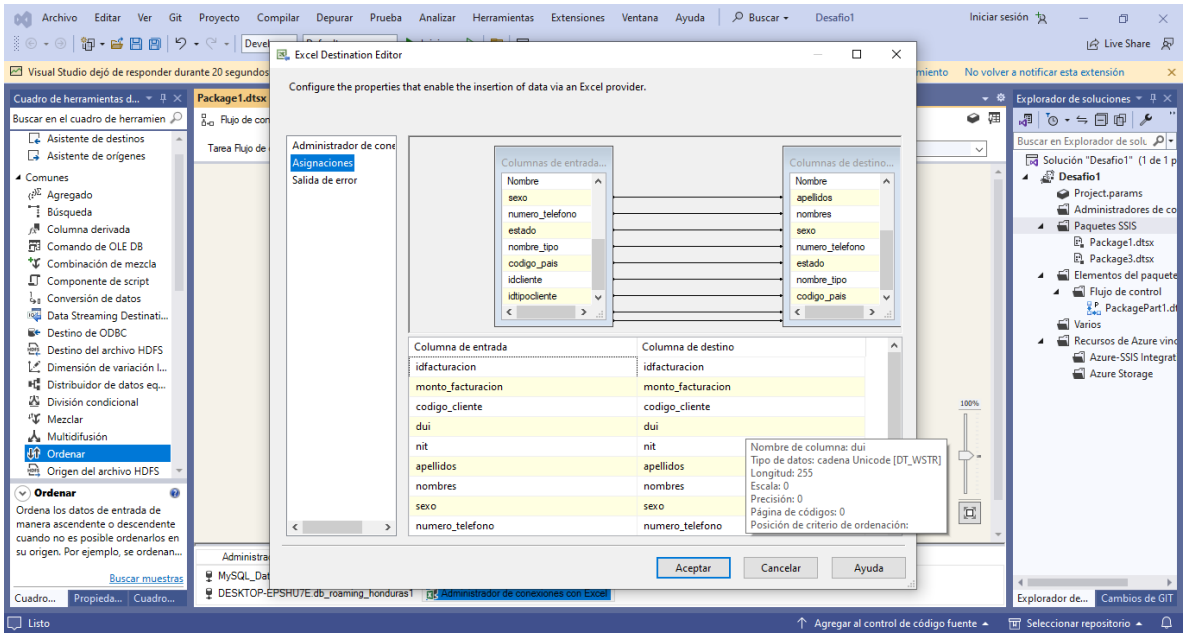

37. Realizamos lo mismo para ambas salidas y configuramos la data que se presentara en el Excel correspondiente de cada grupo, considerando que debemos tener los Excel previamente creados y configurar la conexión.



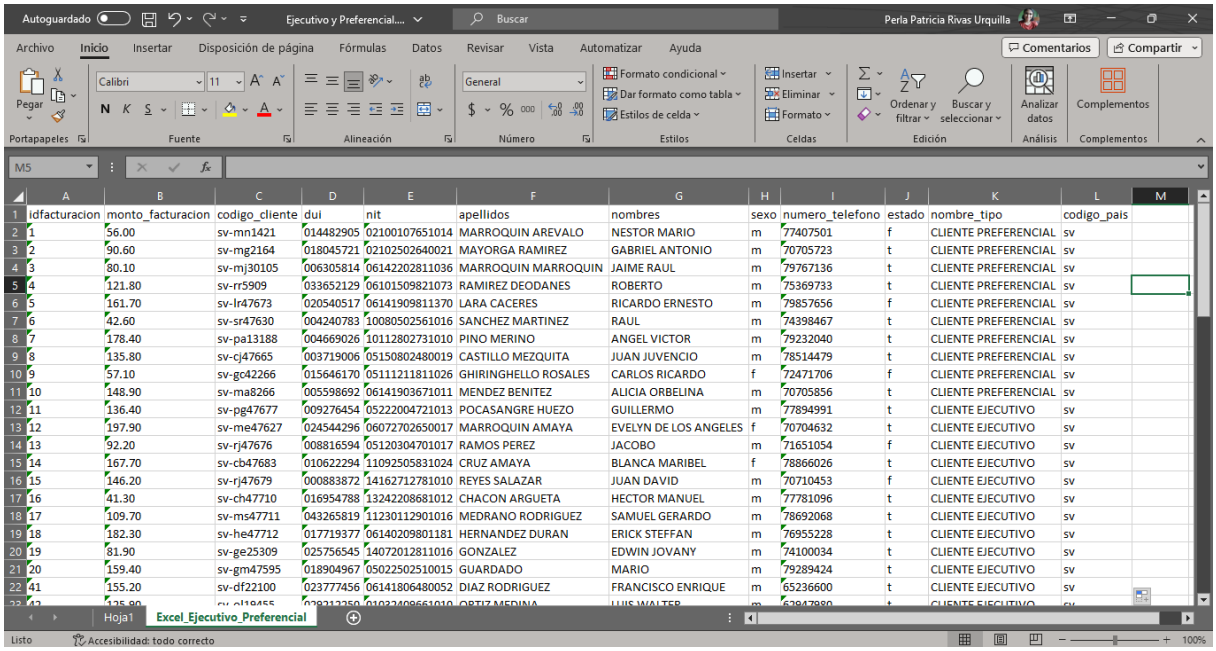



38. Para los datos que se van a mostrar en el Excel creamos una tabla, en esta evitaremos presentar las columnas de idCliente y idTipoCliente.






39. De esta forma ya podremos visualizar la información en los archivos de Excel.



| idfacturacion | monto_facturacion | codigo_cliente | dui | nit | apellidos | nombres | sexo | numero_telefono | estado | nombre_tipo | codigo_pais |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 56.00 | sv-mn1421 | 014482905 | 02100107651014 | MARROQUIN AREVALO | NESTOR MARIO | m | 77407501 | f | CLIENTE PREFERENCIAL | sv |
| 2 | 90.60 | sv-mg2164 | 018045721 | 02102502640021 | MAYORGA RAMIREZ | GABRIEL ANTONIO | m | 70705723 | t | CLIENTE PREFERENCIAL | sv |
| 3 | 80.10 | sv-mj30105 | 006305814 | 06142202811036 | MARROQUIN MARROQUIN | JAIME RAUL | m | 79767136 | t | CLIENTE PREFERENCIAL | sv |
| 4 | 121.80 | sv-rr5909 | 033652129 | 06101509821073 | RAMIREZ DEODANES | ROBERTO | m | 75369733 | t | CLIENTE PREFERENCIAL | sv |
| 5 | 161.70 | sv-lr47673 | 020540517 | 06141909811370 | LARA CACERES | RICARDO ERNESTO | m | 79857656 | f | CLIENTE PREFERENCIAL | sv |
| 6 | 42.60 | sv-sr47630 | 004240783 | 10080502561016 | SANCHEZ MARTINEZ | RAUL | m | 74398467 | t | CLIENTE PREFERENCIAL | sv |
| 7 | 178.40 | sv-pa13188 | 004669026 | 10112802731010 | PINO MERINO | ANGEL VICTOR | m | 79232040 | t | CLIENTE PREFERENCIAL | sv |
| 8 | 135.80 | sv-cj47665 | 003719006 | 05150802480019 | CASTILLO MEZQUITA | JUAN JUVENCIO | m | 78514479 | t | CLIENTE PREFERENCIAL | sv |
| 9 | 57.10 | sv-gc42266 | 015646170 | 05111211811026 | GHIRINGHELLO ROSALES | CARLOS RICARDO | f | 72471706 | f | CLIENTE PREFERENCIAL | sv |
| 10 | 148.90 | sv-ma8266 | 005598692 | 06141903671011 | MENDEZ BENITEZ | ALICIA ORBELINA | m | 70705856 | t | CLIENTE PREFERENCIAL | sv |
| 11 | 136.40 | sv-pg47677 | 009276454 | 05222004721013 | POCASANGRE HUEZO | GUILLERMO | m | 77894991 | t | CLIENTE EJECUTIVO | sv |
| 12 | 197.90 | sv-me47627 | 024544296 | 06072702650017 | MARROQUIN AMAYA | EVELYN DE LOS ANGELES | f | 70704632 | t | CLIENTE EJECUTIVO | sv |
| 13 | 92.20 | sv-rj47676 | 008816594 | 05120304701017 | RAMOS PEREZ | JACOBO | m | 71651054 | f | CLIENTE EJECUTIVO | sv |
| 14 | 167.70 | sv-cb47683 | 010622294 | 11092505831024 | CRUZ AMAYA | BLANCA MARIBEL | f | 78866026 | t | CLIENTE EJECUTIVO | sv |
| 15 | 146.20 | sv-rj47679 | 000883872 | 14162712781010 | REYES SALAZAR | JUAN DAVID | m | 70710453 | f | CLIENTE EJECUTIVO | sv |
| 16 | 41.30 | sv-ch47710 | 016954788 | 13242208681012 | CHACON ARGUETA | HECTOR MANUEL | m | 77781096 | t | CLIENTE EJECUTIVO | sv |
| 17 | 109.70 | sv-ms47711 | 043265819 | 11230112901016 | MEDRANO RODRIGUEZ | SAMUEL GERARDO | m | 78692068 | t | CLIENTE EJECUTIVO | sv |
| 18 | 182.30 | sv-he47712 | 017719377 | 06140209801181 | HERNANDEZ DURAN | ERICK STEFFAN | m | 76955228 | t | CLIENTE EJECUTIVO | sv |
| 19 | 81.90 | sv-ge25309 | 025756545 | 14072012811016 | GONZALEZ | EDWIN JOVANY | m | 74100034 | t | CLIENTE EJECUTIVO | sv |
| 20 | 159.40 | sv-gm47595 | 018904967 | 05022502510015 | GUARDADO | MARIO | m | 79289424 | t | CLIENTE EJECUTIVO | sv |
| 41 | 155.20 | sv-df22100 | 023777456 | 06141806480052 | DIAZ RODRIGUEZ | FRANCISCO ENRIQUE | m | 65236600 | t | CLIENTE EJECUTIVO | sv |