Logotipo, nombre de la empresa

Descripción generada automáticamente

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

Docente:

Karens Medrano

Nombre:

Perla Patricia Rivas Urquilla

Carnet:

RU171368

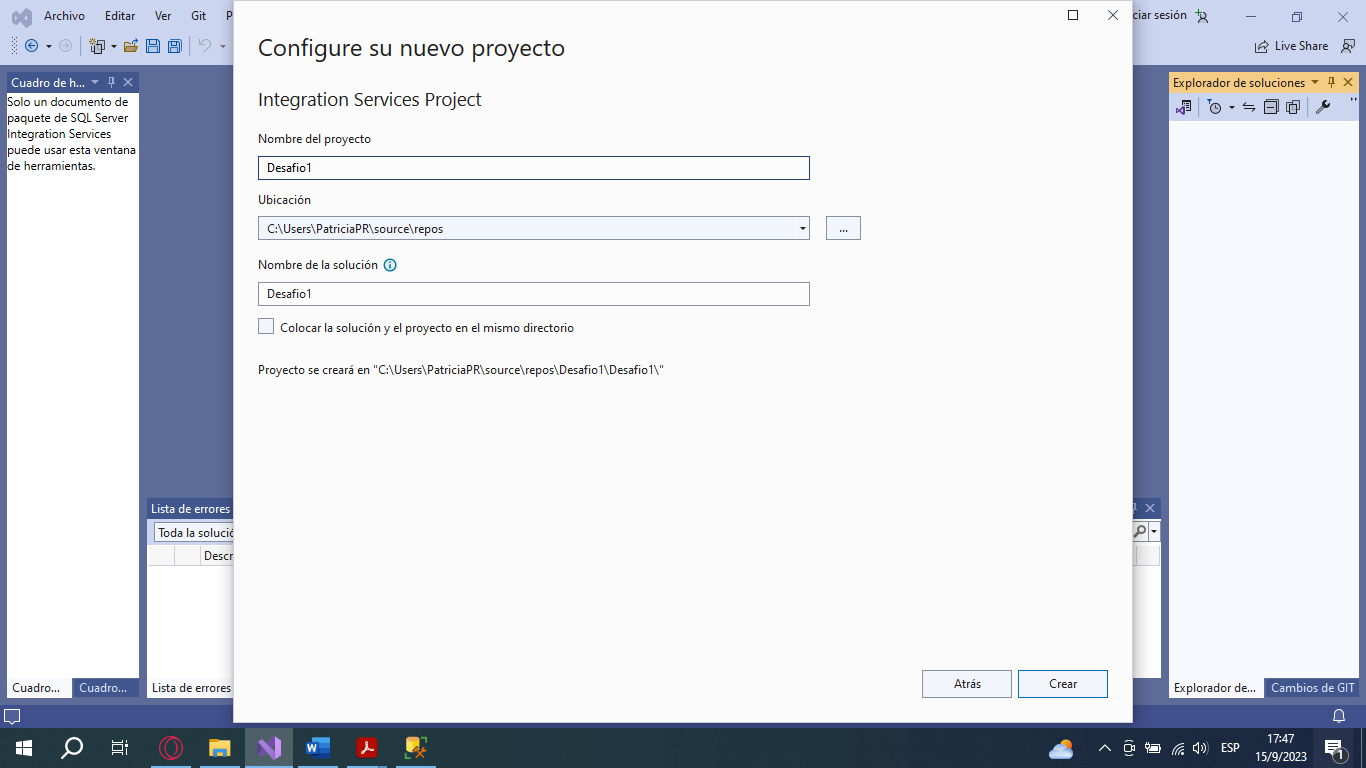
Grupo:

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.  
     
   Una captura de pantalla de una computadora

   Descripción generada automáticamente
3. Se realizarán los ajustes necesarios para poder ejecutar nuestro proyecto sin errores.  
     
   Una captura de pantalla de una computadora

   Descripción generada automáticamente

## **EJERCICIO 1**

1. 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.  
     
   Interfaz de usuario gráfica, Aplicación, Word

   Descripción generada automáticamente
2. 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.  
     
   Una captura de pantalla de una computadora

   Descripción generada automáticamente
3. Asignamos la variable previamente creada.  
     
   Interfaz de usuario gráfica, Aplicación

   Descripción generada automáticamente
4. 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**Una captura de pantalla de una computadora

   Descripción generada automáticamente
5. El flujo para poder hacer el ingreso de nuestros datos a la base de datos será la siguiente:  
     
   Interfaz de usuario gráfica, Aplicación

   Descripción generada automáticamente
6. Ahora podremos ver como nuestro origen de texto plano tomara el valor de los archivos en base al source.  
     
   Interfaz de usuario gráfica, Aplicación, Word

   Descripción generada automáticamente
7. Creamos y editamos la herramienta de columna derivada para poder organizar las sucursales que se están ingresando a nuestra tabla.  
     
   Una captura de pantalla de una computadora

   Descripción generada automáticamente
8. En nuestro destino OLE DB seleccionamos nuestra conexión y tabla a donde se hará el ingreso de nuestros registros de los archivos.  
     
   Una captura de pantalla de una computadora

   Descripción generada automáticamente
9. Una vez tenemos nuestro flujo configurado, podremos ejecutar el paquete.  
     
   Una captura de pantalla de una computadora

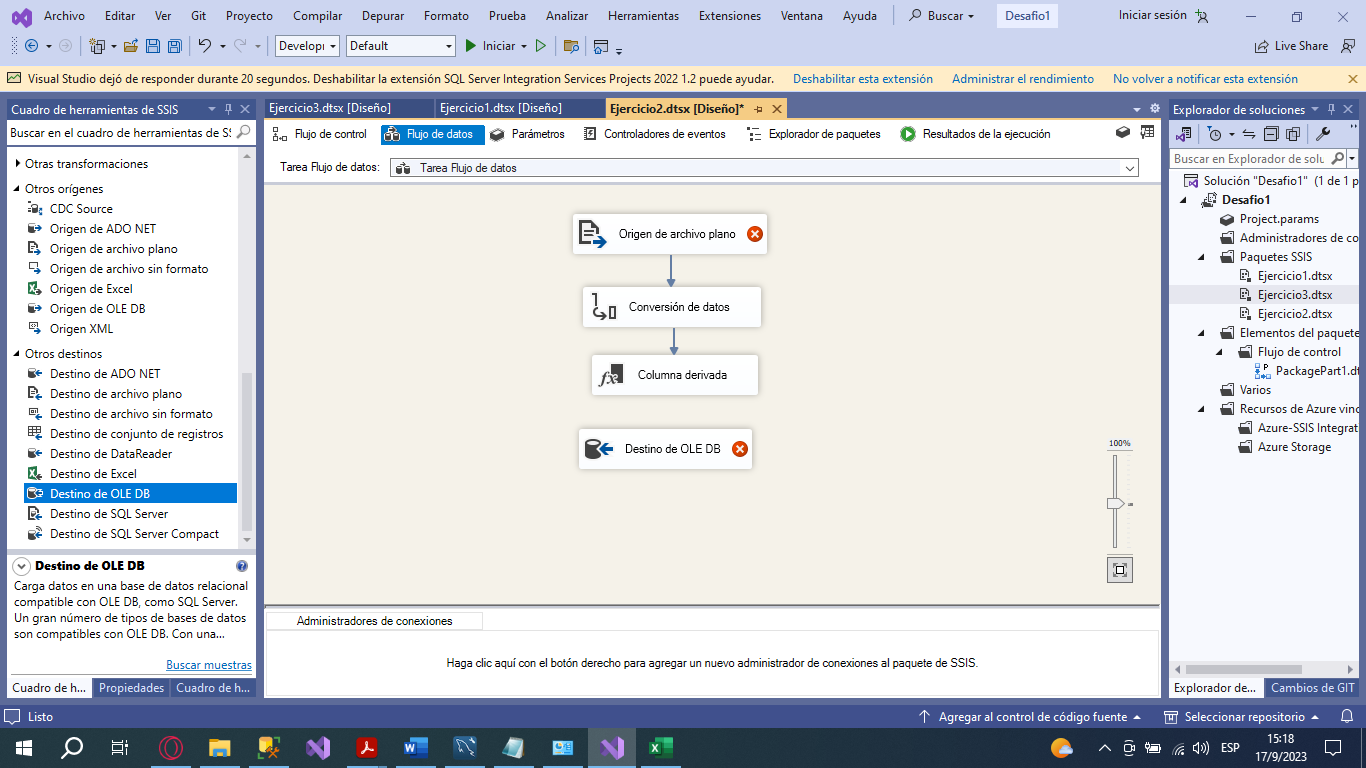
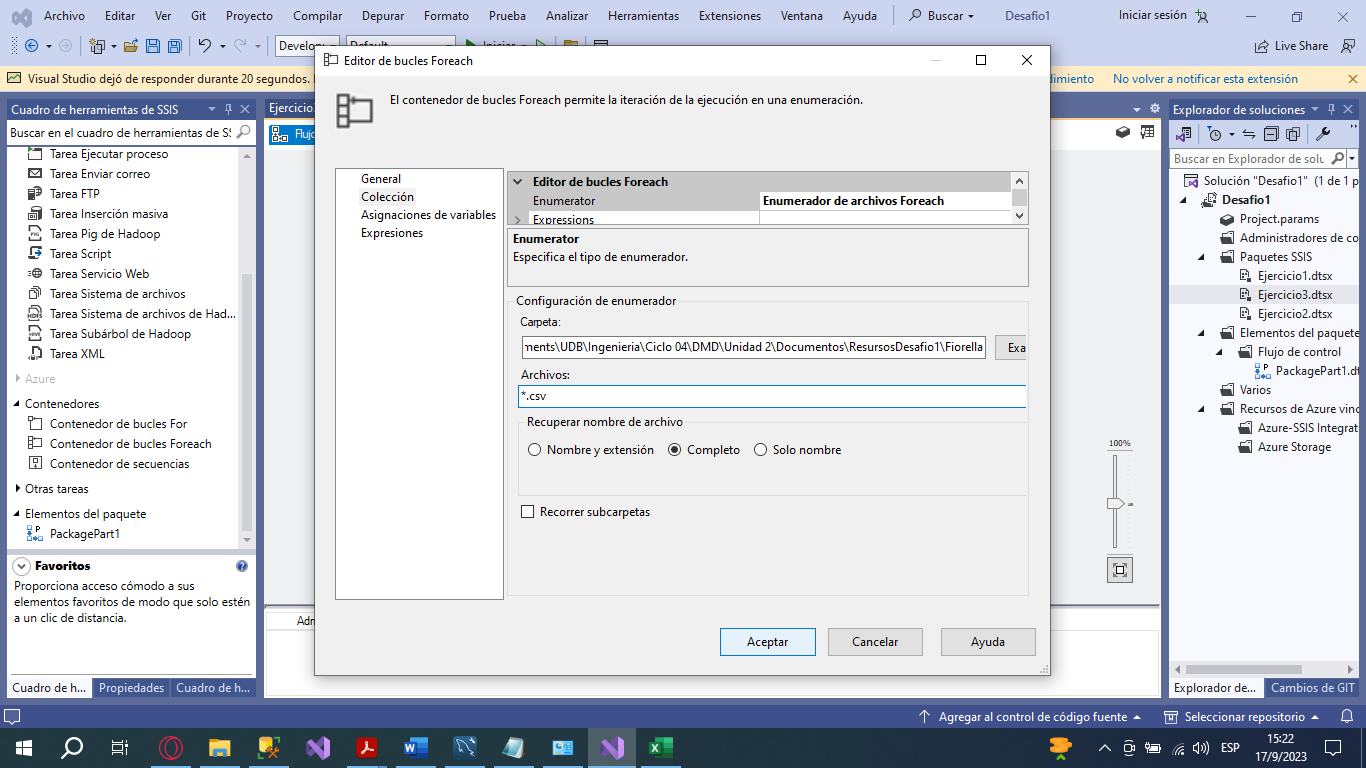
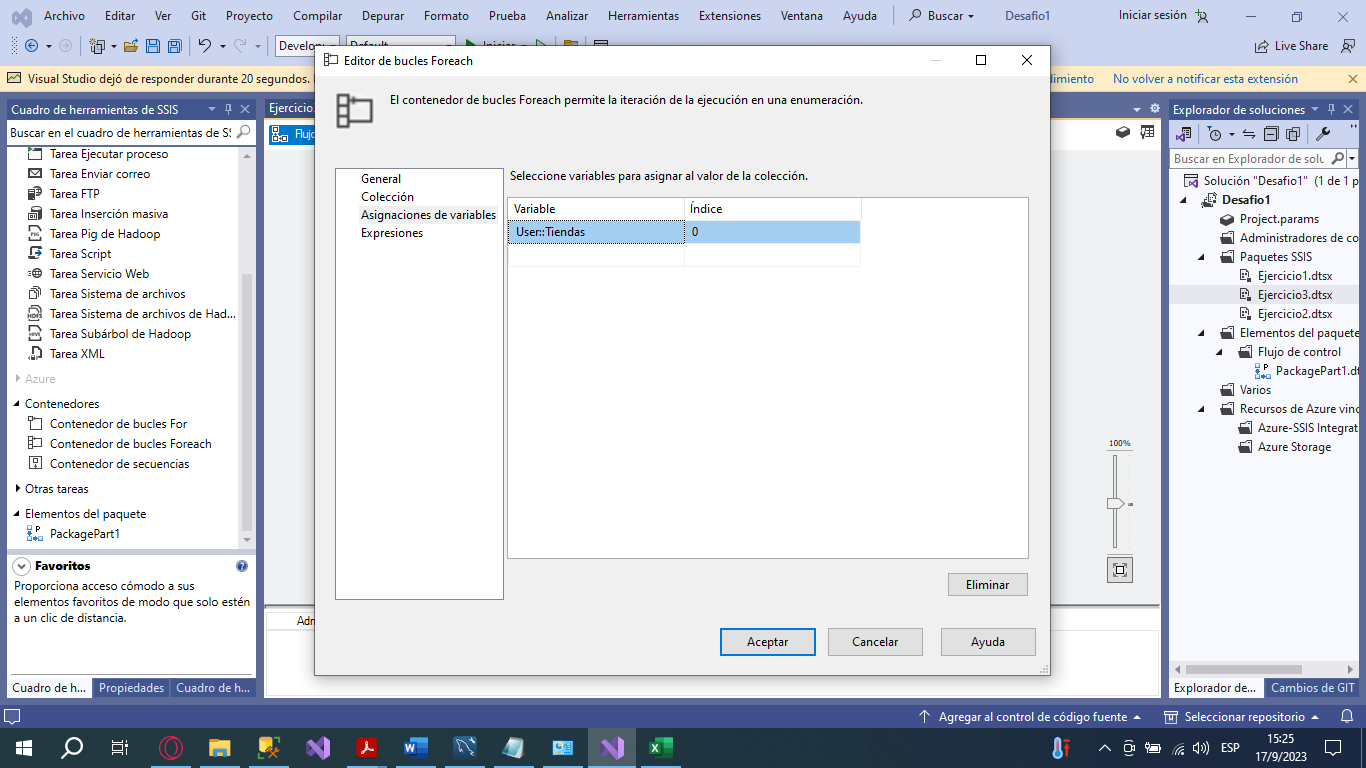
   Descripción generada automáticamente
10. Ahora podemos ver nuestros registros en la base de datos  
      
    Captura de pantalla de computadora

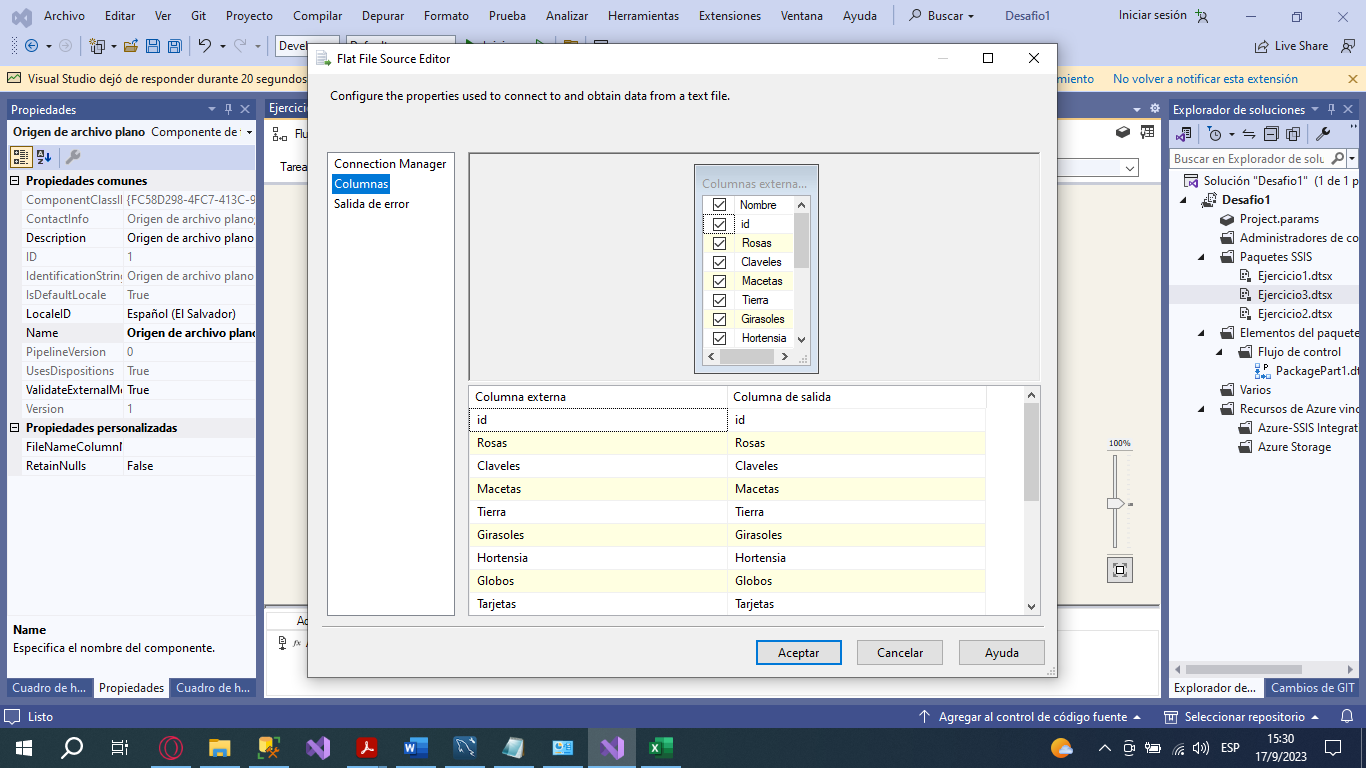
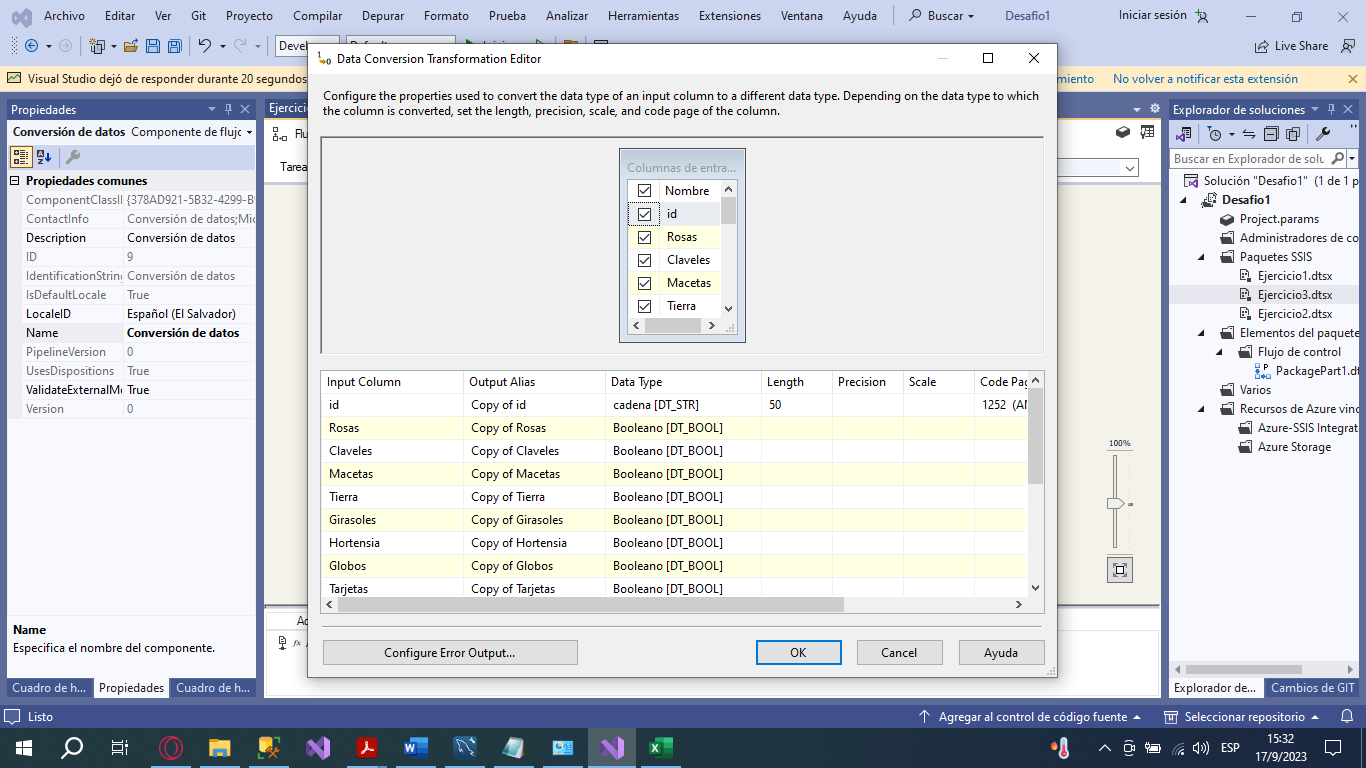
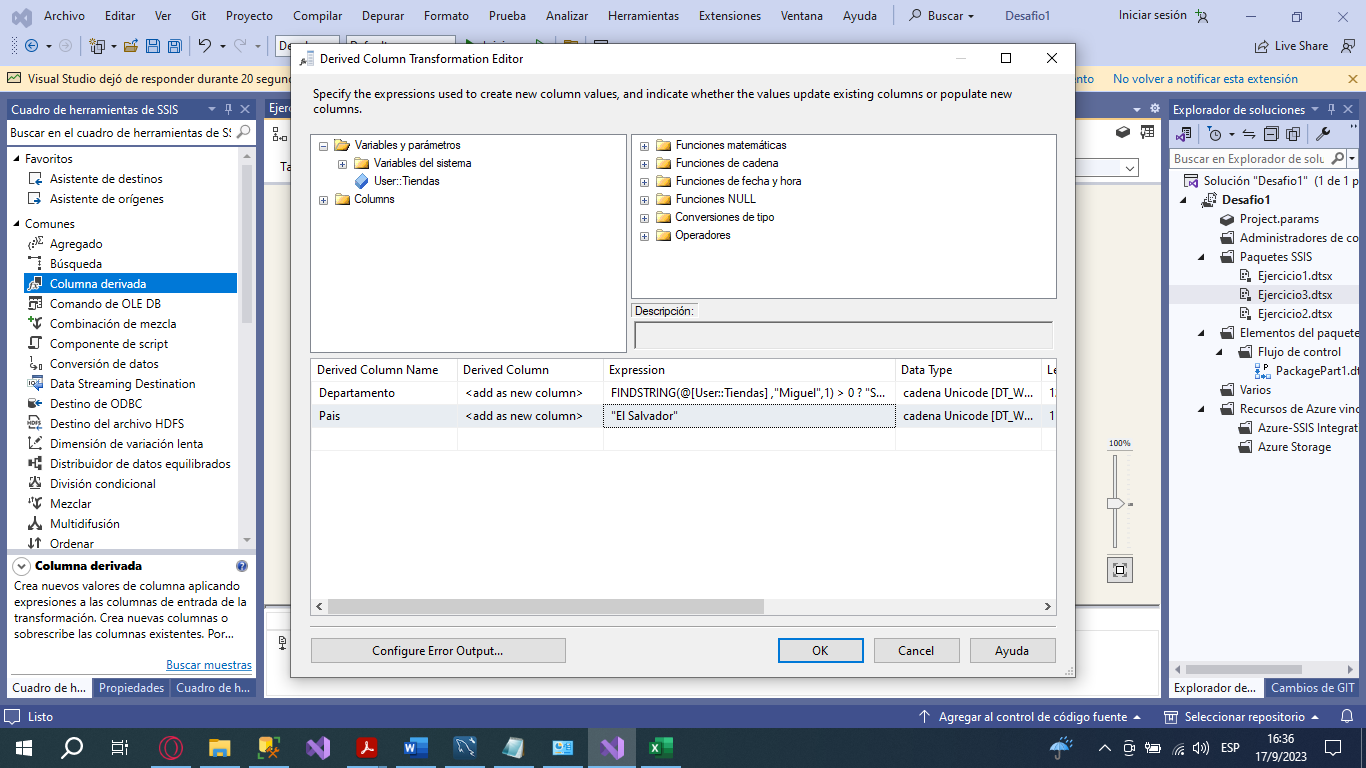
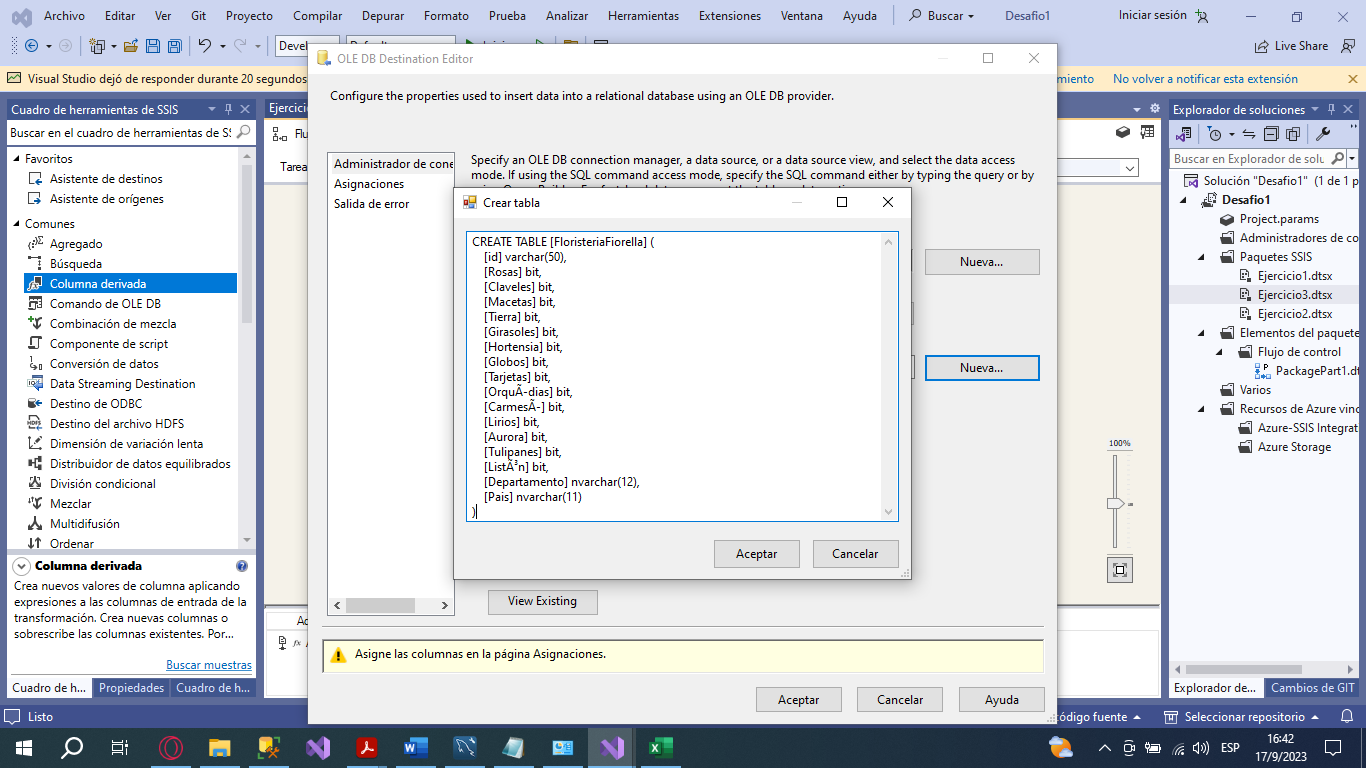
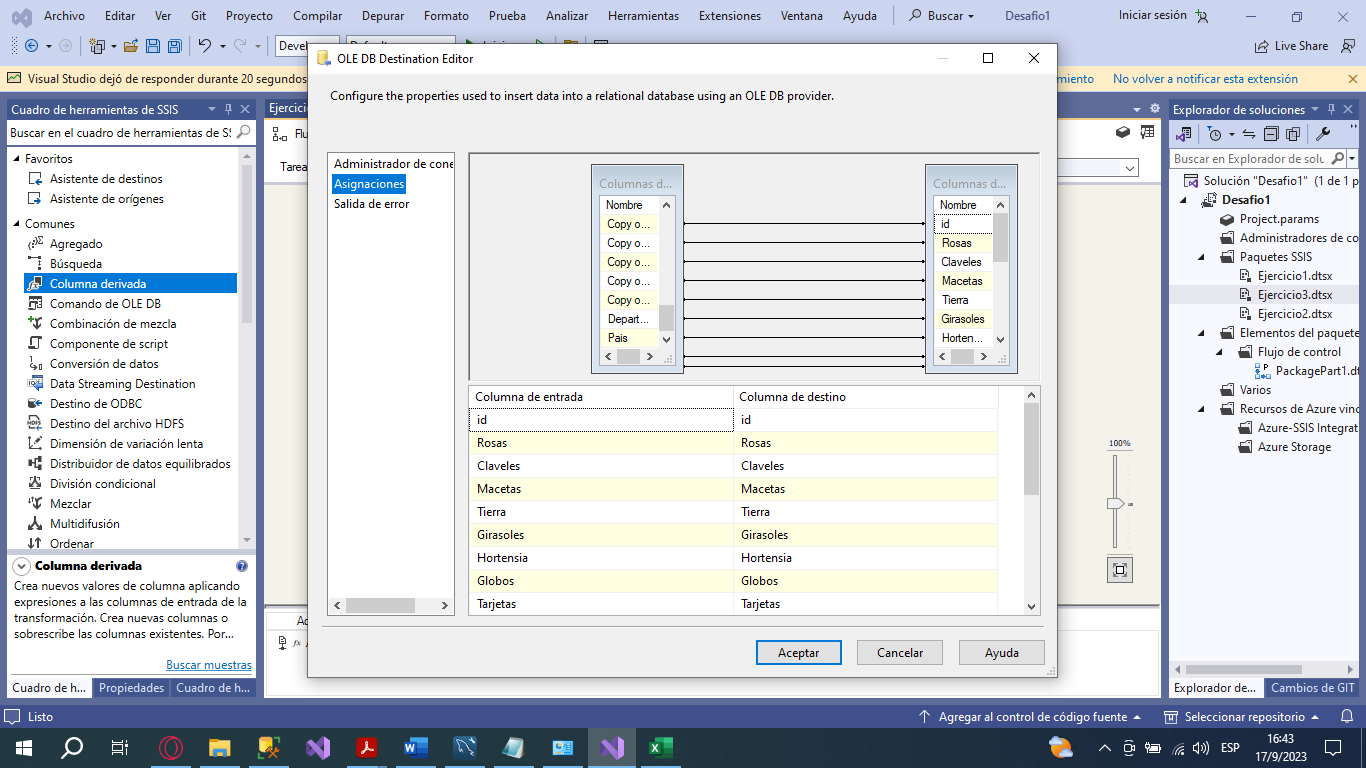
    Descripción generada automáticamenteCaptura de pantalla de computadora

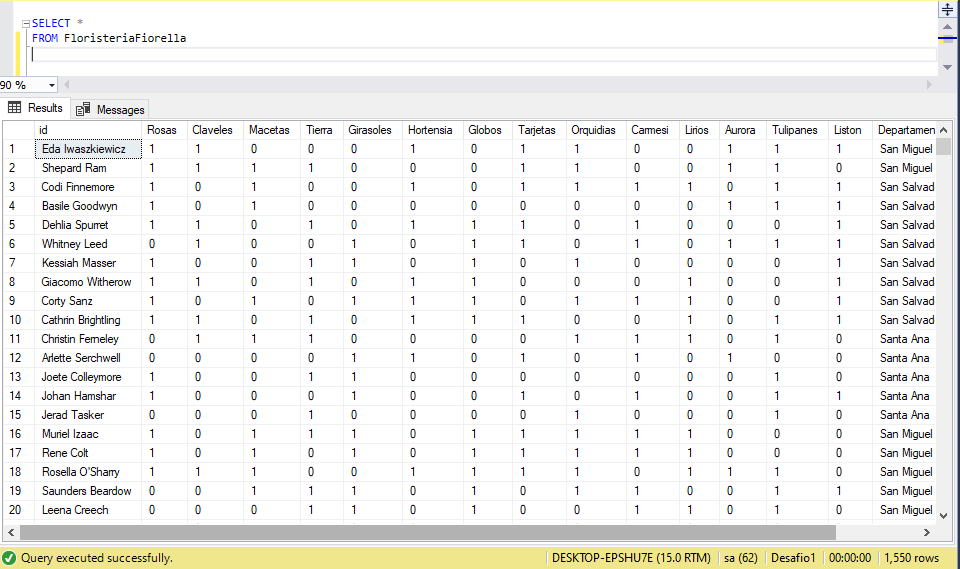
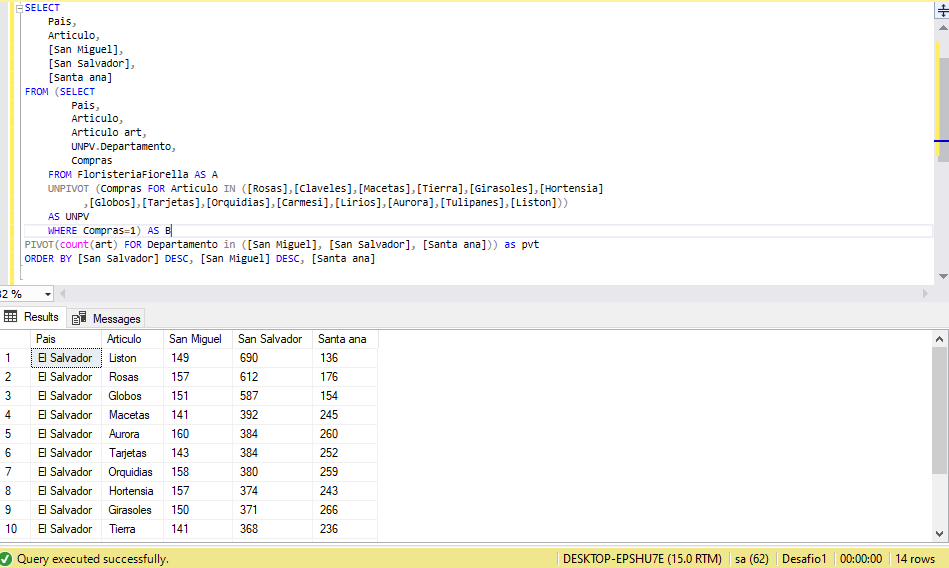
    Descripción generada automáticamente
11. Para poder hacer el análisis, se presentan dos consultas, una de ellas está enfocada en poder clasificar a los clientes en diferentes grupos. Los grupos en los que se clasifican serán Visitas, Servicios e Ingresos. En una clasificación donde 1 es el puntaje mas 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.  
      
    Una captura de pantalla de una computadora

    Descripción generada automáticamente  
      
    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.  
      
    Una captura de pantalla de una computadora

    Descripción generada automáticamente  
      
    **EJERCICIO 2**
12. Para el ejercicio 2 realizaremos un procedimiento similar al ejercicio 1. Haremos un bucle foreach y dentro de este nuestra tarea de flujo.  
      
    Interfaz de usuario gráfica, Aplicación

    Descripción generada automáticamente
13. Dentro de nuestro flujo de datos, tendremos los siguientes pasos.  
      
    
14. Seleccionamos la carpeta correspondiente a los archivos que debemos cargar para el ejercicio y asignamos la variable para guardar la información del recorrido.  
      
      
      
    
15. Editamos la expresión de configuración de la conexión de archivo de texto plano, para que esta obtenga los datos de la variable.  
      
    Interfaz de usuario gráfica, Aplicación, Word

    Descripción generada automáticamente
16. De esta forma nuestro origen ya podrá tener las columnas correspondientes a nuestros archivos.  
      
    
17. Hacemos la conversión para que los valores de las columnas de detalle de los productos sean de tipo booleano.  
      
    
18. 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”.  
      
    
19. En el destino OLE DB crearemos la tabla donde se almacenarán nuestros registros.  
      
    
20. Podremos visualizar la entrada y salida correspondiente de los datos.  
      
    
21. Ejecutamos el paquete para poder cargar los datos a nuestra tabla en SQL.  
      
    Una captura de pantalla de una computadora

    Descripción generada automáticamente
22. Con esto ya podremos realizar las consultas necesarias en nuestro SQL.  
      
    
23. 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.  
      
    

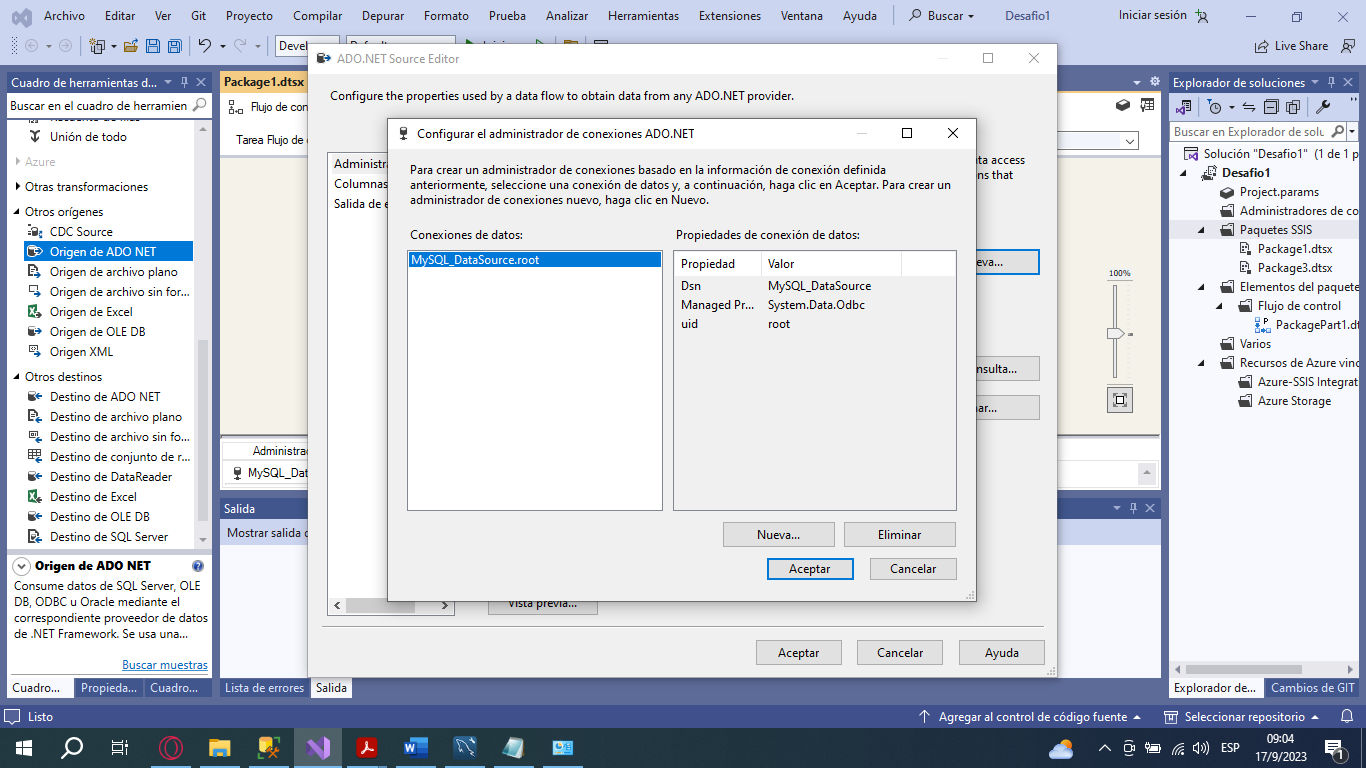
## **EJERCICIO 3**

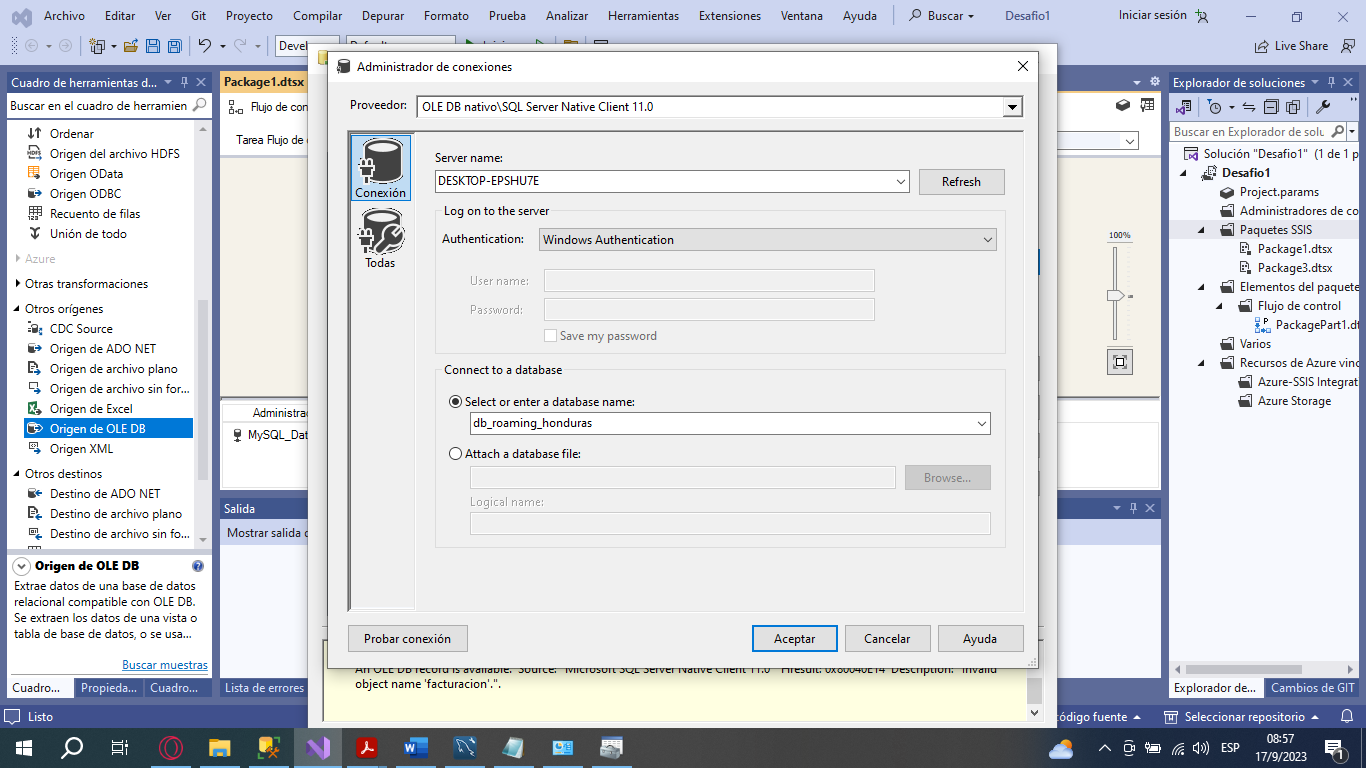
1. Para el ejercicio 3 debemos crear las bases de datos en MySQL y SQL Server.  
     
   Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

   Descripción generada automáticamente  
     
   Una captura de pantalla de una computadora

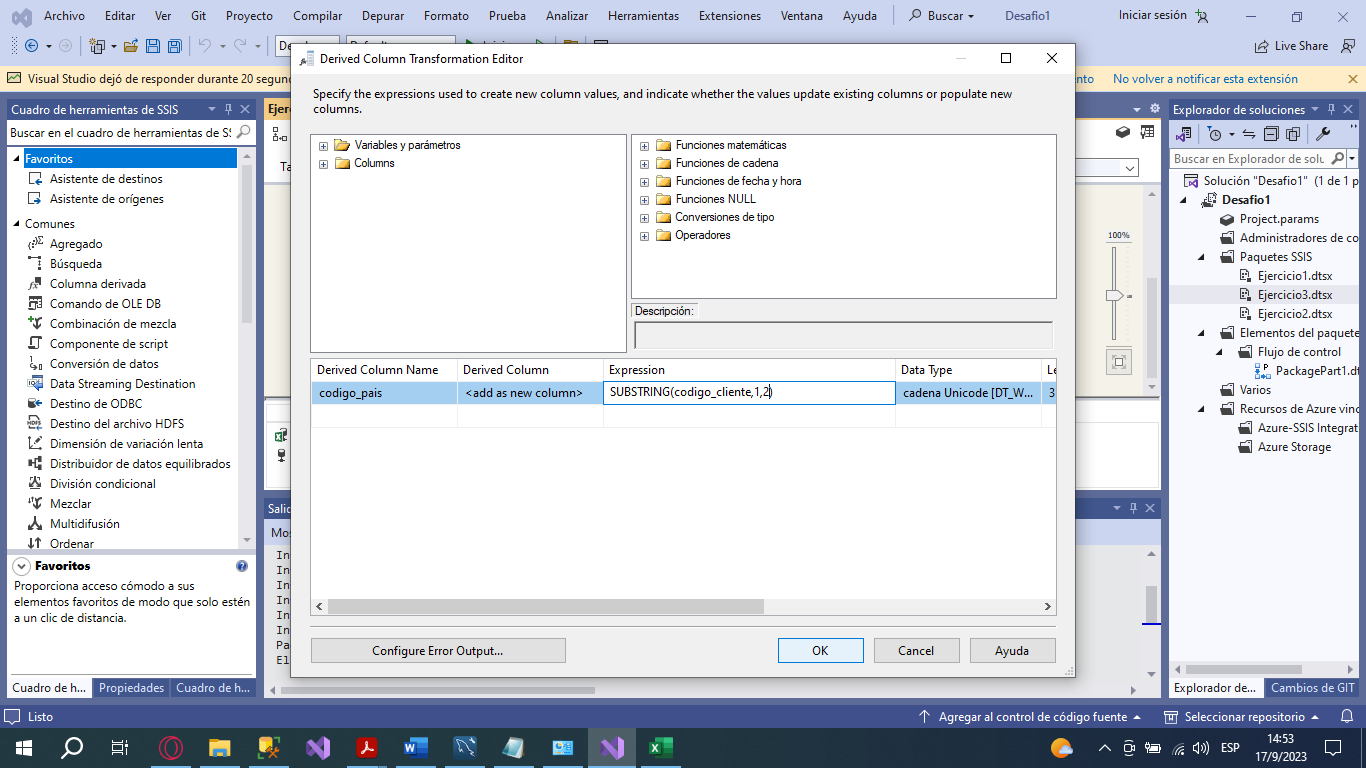
   Descripción generada automáticamente
2. Formaremos el siguiente flujo de datos y procederemos a realizar las configuraciones correspondientes a cada uno de los pasos del flujo.  
     
   Interfaz de usuario gráfica

   Descripción generada automáticamente
3. Empezaremos por crear nuestro origen de datos ADO.NET para nuestra conexión a la base de datos de MySQL.   
     
   Una captura de pantalla de una computadora

   Descripción generada automáticamente
4. Para esto hay que tener previamente configurado nuestro DataSource con los drivers correspondientes. Y configurar la conexión en administrador de conexiones.  
     
   
5. 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.  
   Una captura de pantalla de una computadora

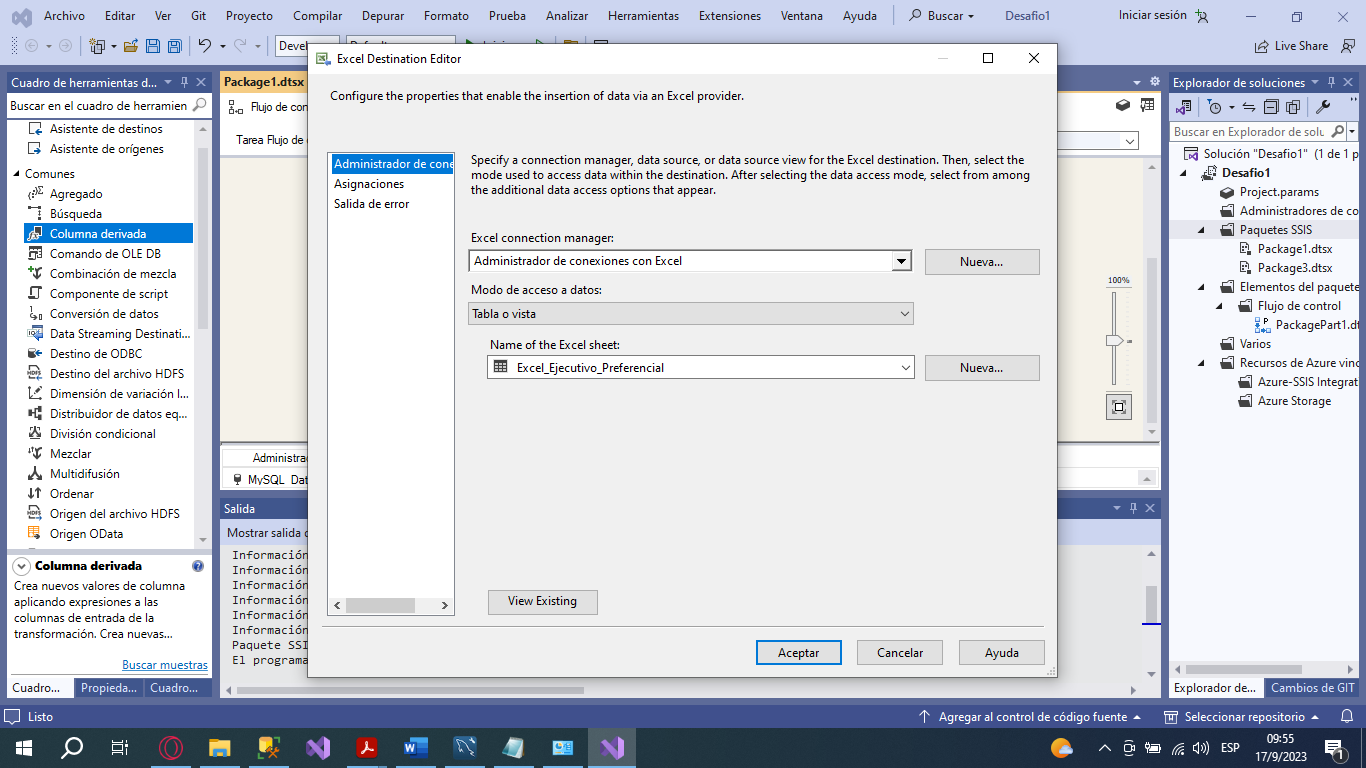
   Descripción generada automáticamente
6. Realizamos la misma configuración de origen ADO.NET para los datos de la base de datos SQL Server.  
     
     
     
   Una captura de pantalla de una computadora

   Descripción generada automáticamente
7. Ahora que tenemos la información de ambas bases de datos haremos una unión de todos los datos.  
     
   Una captura de pantalla de una computadora

   Descripción generada automáticamente
8. 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.  
     
   
9. Para nuestro condicional agregaremos las siguientes expresiones de condición.  
     
   Una captura de pantalla de una computadora

   Descripción generada automáticamente
10. También haremos un ordenamiento de los datos por país y tipo de cliente.  
      
    Interfaz de usuario gráfica, Aplicación

    Descripción generada automáticamente
11. 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.  
      
    Una captura de pantalla de una computadora

    Descripción generada automáticamente  
      
    
12. Para los datos que se van a mostrar en el Excel creamos una tabla, en esta evitaremos presentar las columnas de idCliente y idTipoCliente.  
      
    Una captura de pantalla de una computadora

    Descripción generada automáticamente  
      
    Interfaz de usuario gráfica, Aplicación

    Descripción generada automáticamente
13. De esta forma ya podremos visualizar la información en los archivos de Excel.  
      
    