

# UNIVERSIDAD DON BOSCO



## DOCUMENTACIÓN DESAFIO 1

**Datawarehouse y Minería de Datos**

### **PRESENTADO POR:**

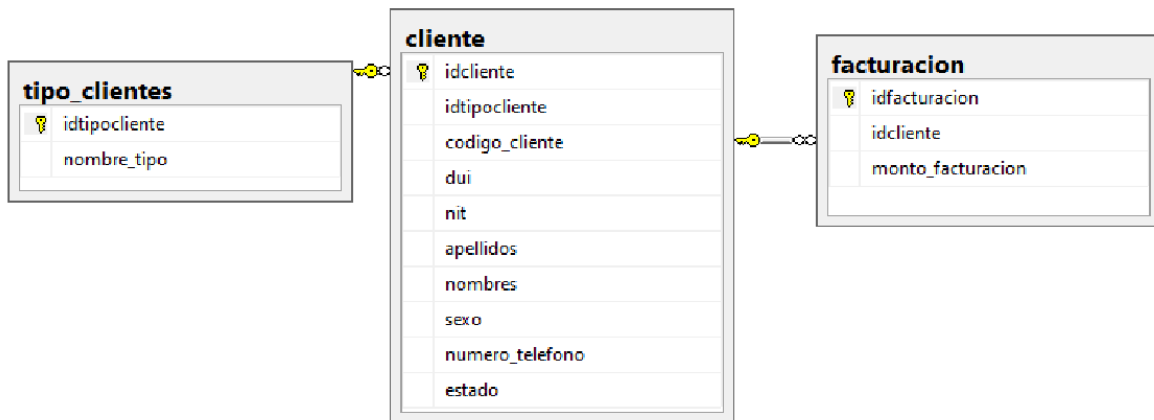
Christian Ergardo Hernández Rivera HR160372

### **DOCENTE:**

**Karens Medrano**

Para este proyecto se ha utilizado el siguiente Stack:

1. SQL Server 2019 Developer
2. Visual Studio 2022 Community
3. GitHub como repositorio
4. 3 Scripts de Bases de Datos: “db\_costarica.sql”, “db\_guatemala.sql”, “db\_honduras.sql”
5. La estructura de las bases de datos es la siguiente:



### Resultado Objetivo:

1. Consolidado de facturación por país: archivo de Excel que muestre, para cada país, el nombre del país, el número total de ventas y el monto total de facturación.
2. Consolidado de ventas por tipo de cliente: archivo de Excel que muestre, para cada tipo de cliente, el nombre del tipo de cliente, el número total de ventas y el monto total de facturación.

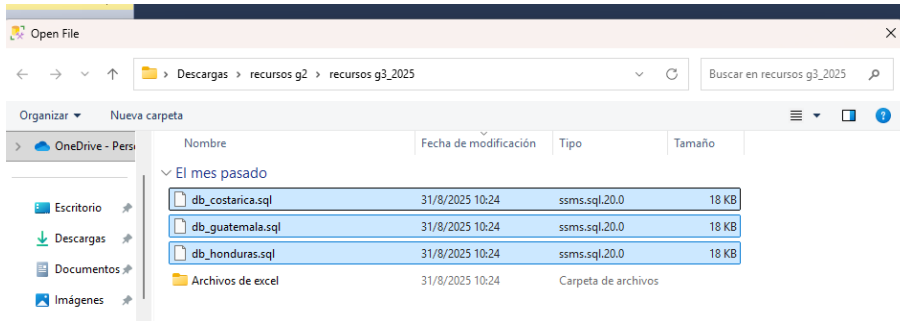
### Identifiquemos que columnas utilizaremos:

1. Para el Consolidado de facturación por país:  
tipo\_clientes(id\_tipo\_cliente,nombre\_tipo),  
cliente(id\_cliente),facturacion(id\_facturacion,monto\_facturacion)
2. Para el Consolidado de facturación por tipo de cliente:  
tipo\_clientes(id\_tipo\_cliente,nombre\_tipo),  
cliente(id\_cliente),facturacion(id\_facturacion,monto\_facturacion)

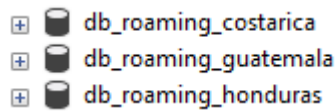
Por lo que el resto de las columnas no será necesario

## Pasos previos: Crear Bases de Datos

Dentro de la carpeta del proyecto hemos ubicado los scripts de las 3 bases de datos a utilizar, abriremos cada una y la ejecutaremos

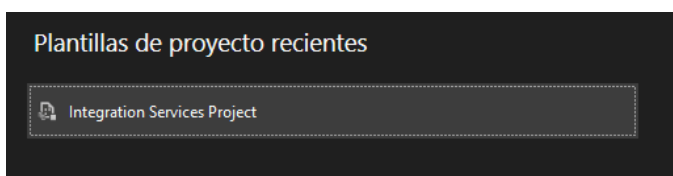
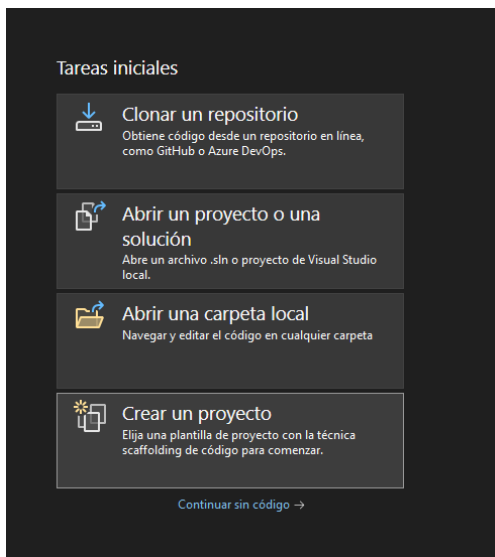


después de ejecutadas las visualizaremos en el “Object Explorer”



## Iniciar Proyecto en Visual Studio 2022

En visual Studio Crearemos un nuevo proyecto SSDT de tipo “Integration Service Project”



Colocaremos un nombre y elegiremos una ruta adecuada para el proyecto, en este caso lo crearemos dentro del repositorio git

Configure su nuevo proyecto

Integration Services Project

Nombre del proyecto  
Desafio1

Ubicación  
C:\Users\h\_chr\OneDrive\Documentos\UDB VIRTUAL\DM2025C02G01TCV\DESAFIO1\DM2-desafi

Nombre de la solución  
Desafio1

☒ Colocar la solución y el proyecto en el mismo directorio

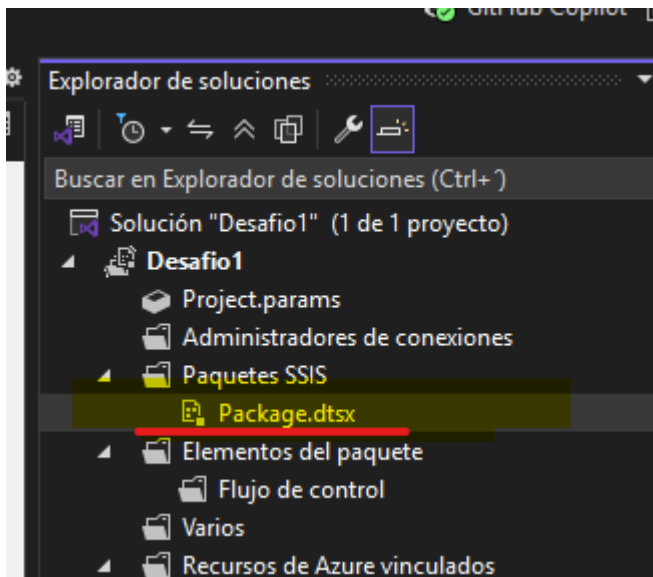
Proyecto se creará en "C:\Users\h\_chr\OneDrive\Documentos\UDB VIRTUAL\DM2025C02G01TCV\DESAFIO1\DM2-desafi\Desafio1"

Dentro de esa ruta se habrá creado:

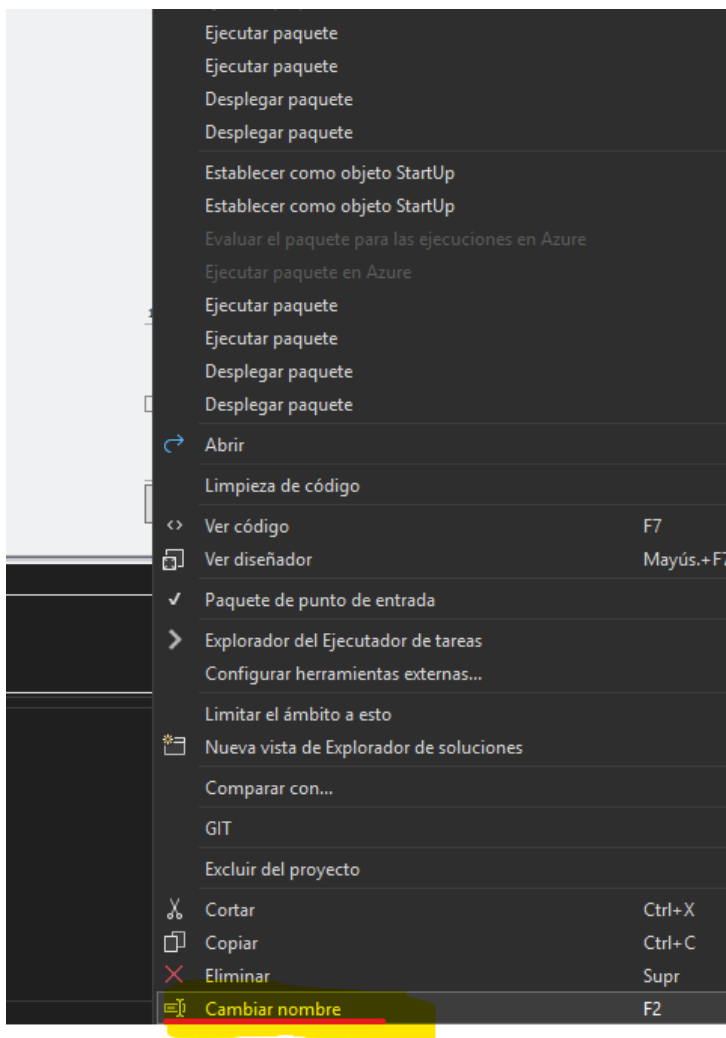
1. La carpeta del proyecto “Desafio1”
2. Adicionalmente he incluido los scripts de la base de datos
3. Además se incluyen archivos propios del repositorio de git: “.git”, “.gitignore”, “README.md”
4. Incluiremos la documentación también en este directorio, no obstante ser al finalizar el desafío y haber documentado todos los pasos.
5. Incluiremos los archivos de resultado

Nombre	Fecha de modificación	Tipo	Tamaño
.git	13/9/2025 23:17	Carpeta de archivos	
Desafio1	13/9/2025 23:22	Carpeta de archivos	
.gitignore	13/9/2025 23:17	Archivo de origen ...	8 KB
Chinook_SqlServer.sql	13/9/2025 23:17	ssms.sql.20.0	3,835 KB
db_costarica.sql	13/9/2025 23:17	ssms.sql.20.0	18 KB
db_guatemala.sql	13/9/2025 23:17	ssms.sql.20.0	18 KB
db_honduras.sql	13/9/2025 23:17	ssms.sql.20.0	18 KB
README.md	13/9/2025 23:17	Archivo de origen ...	1 KB

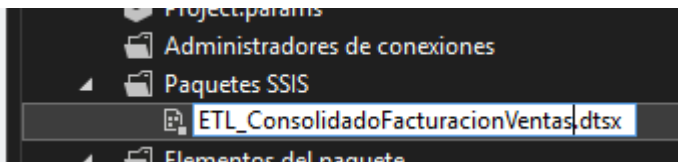
Para este ejercicio, renombrare el paquete SSIS, con la finalidad de darle un nombre más descriptivo, he elegido “ETL\_ConsolidadoFacturacionVentas.dtsx”, para eso de clic derecho sobre “Package.dtsx”



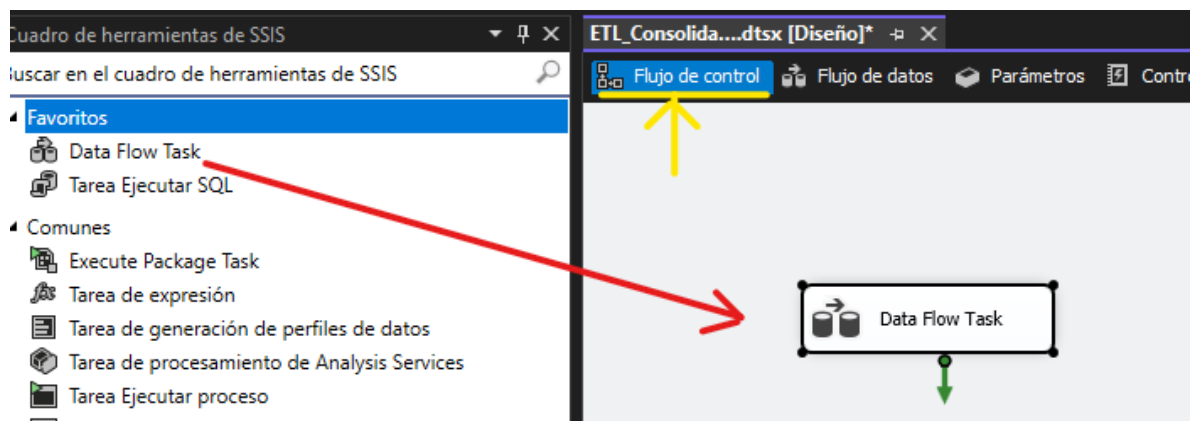
Luego de clic sobre la opción cambiar nombre:



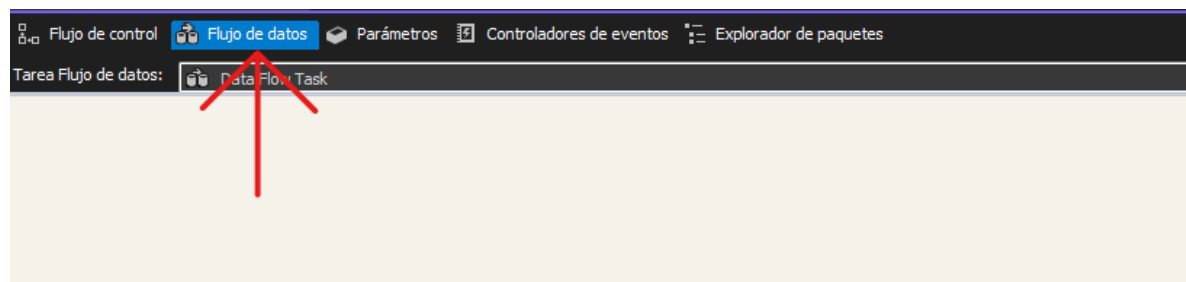
Y coloque el nuevo nombre deseado, y confirme el nombre con enter:



Luego diríjase al lado izquierdo al apartado “Cuadro de herramientas de SSIS”, seleccione “Data Flow Task” y agréguelo a la sección “Flujo de control”

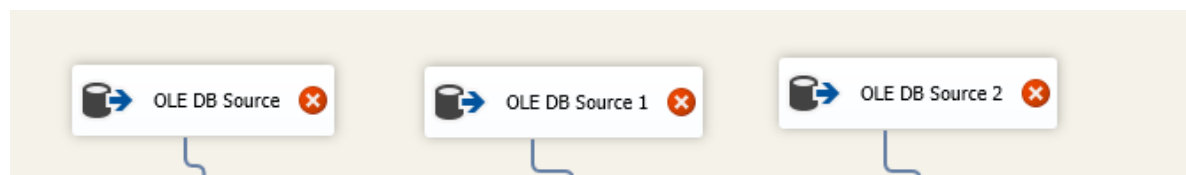


Al dar doble clic sobre ese elemento nos posicionaremos sobre la sección flujo de datos

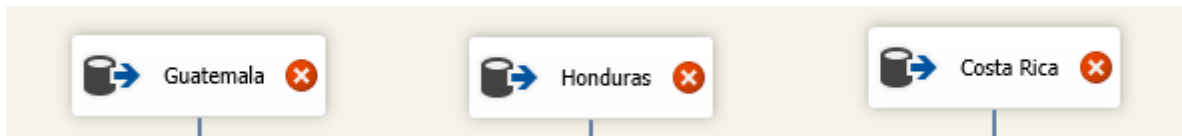


## Fase de extracción

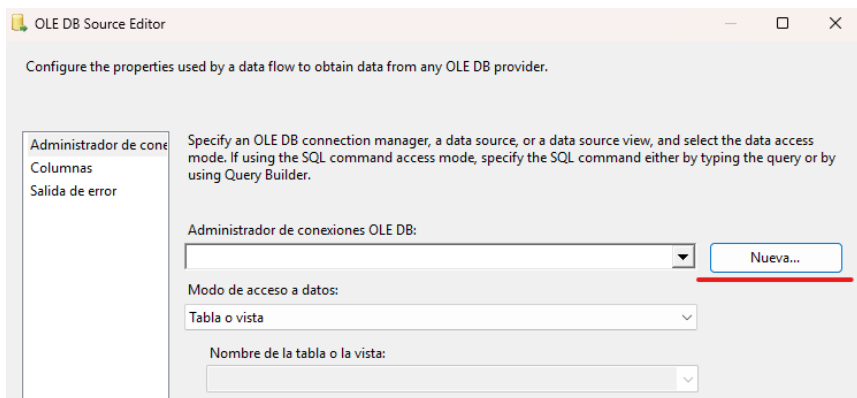
Iniciaremos agregando los tres controles “OLE DB Source” que servirán para extraer los datos



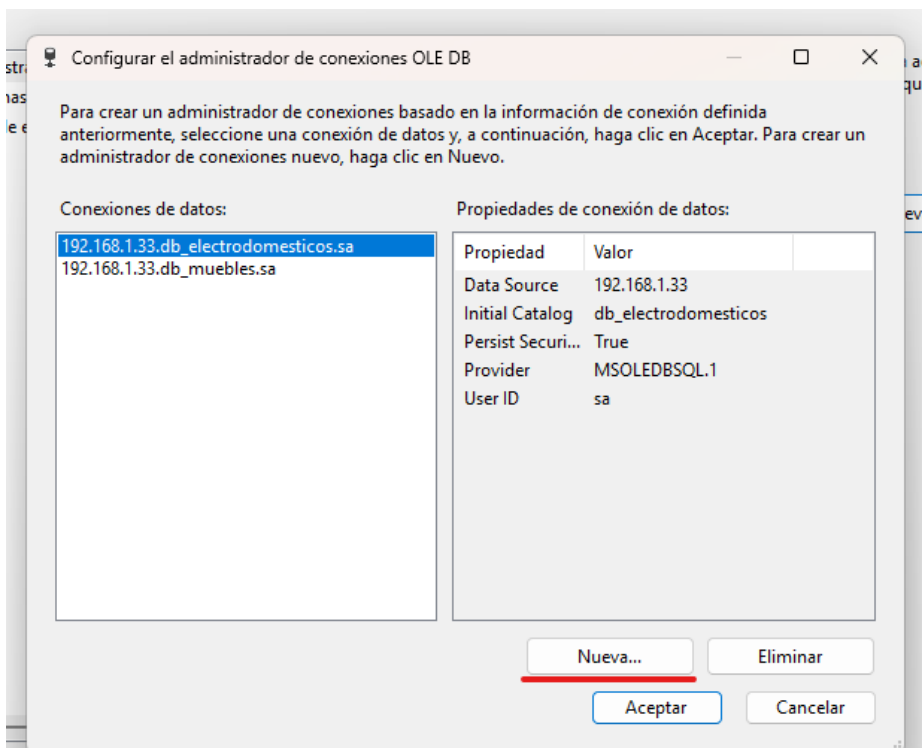
Estos controles los renombraremos para identificar a que base corresponde:



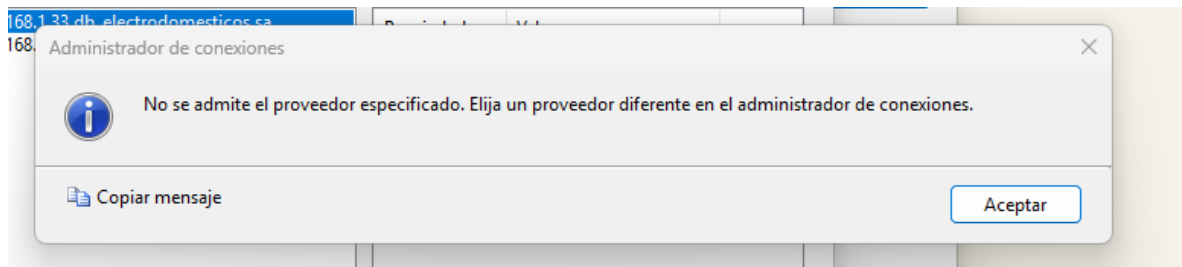
Iniciaremos con el componente para la base de datos Guatemala, como no tenemos una conexión establecida para esta base de datos, deberemos agregar una nueva, de clic en el botón “Nueva...”



Luego en las conexiones de datos daremos clic en “Nueva...”



En el mensaje que nos aparezca daremos clic en aceptar



Al aparecernos el formulario para la creación de la conexión

**Proveedor:** selecciones “OLE DB nativo Microsoft OLE DB Provider for SQL Server”

**Server Name:** en mi caso colocare la dirección IP de mi Base de datos que es la “192.168.1.33”

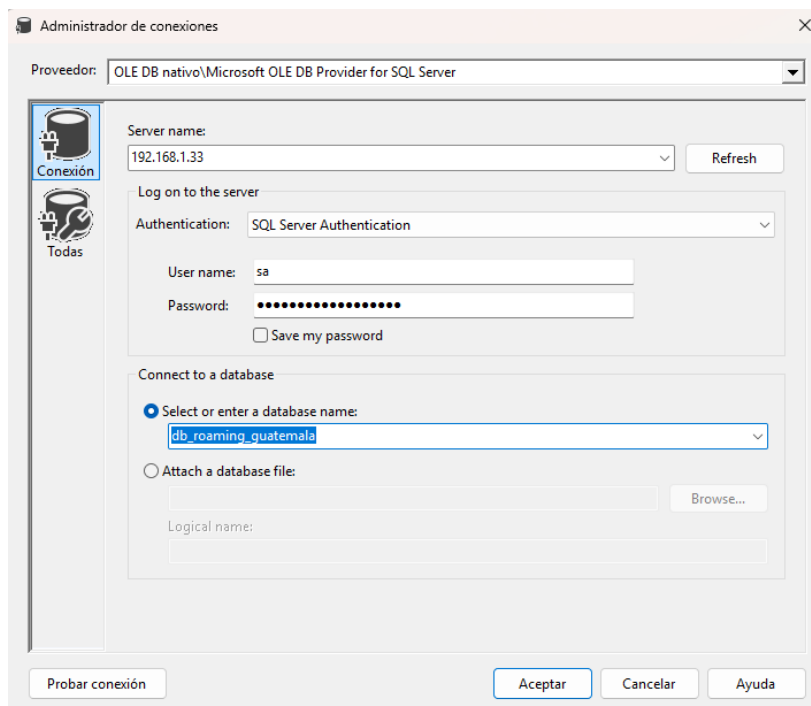
**Autenticación:** Ya que mi computadora y la base de datos no forman parte del mismo dominio, y el servidor de base de datos no esta instalado en mi equipo local, optare por la autenticacio SQL Server

**Username:** El nombre de usuario autorizado para el servidor

**Password:** La contraseña de dicho usuario

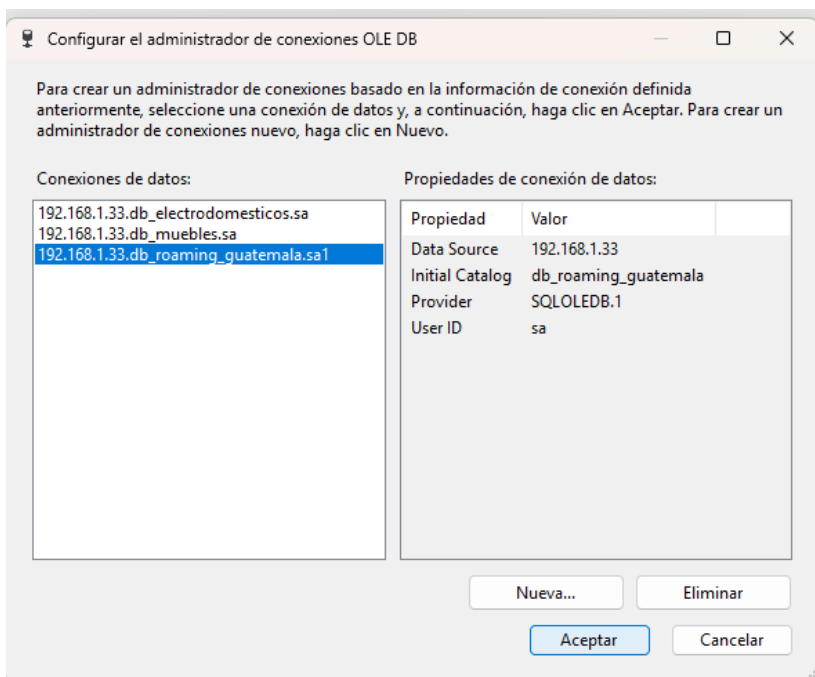
**Database name:** Seleccionare la base de datos “db\_roaming\_guatemala”

El formulario lleno luce de la siguiente forma

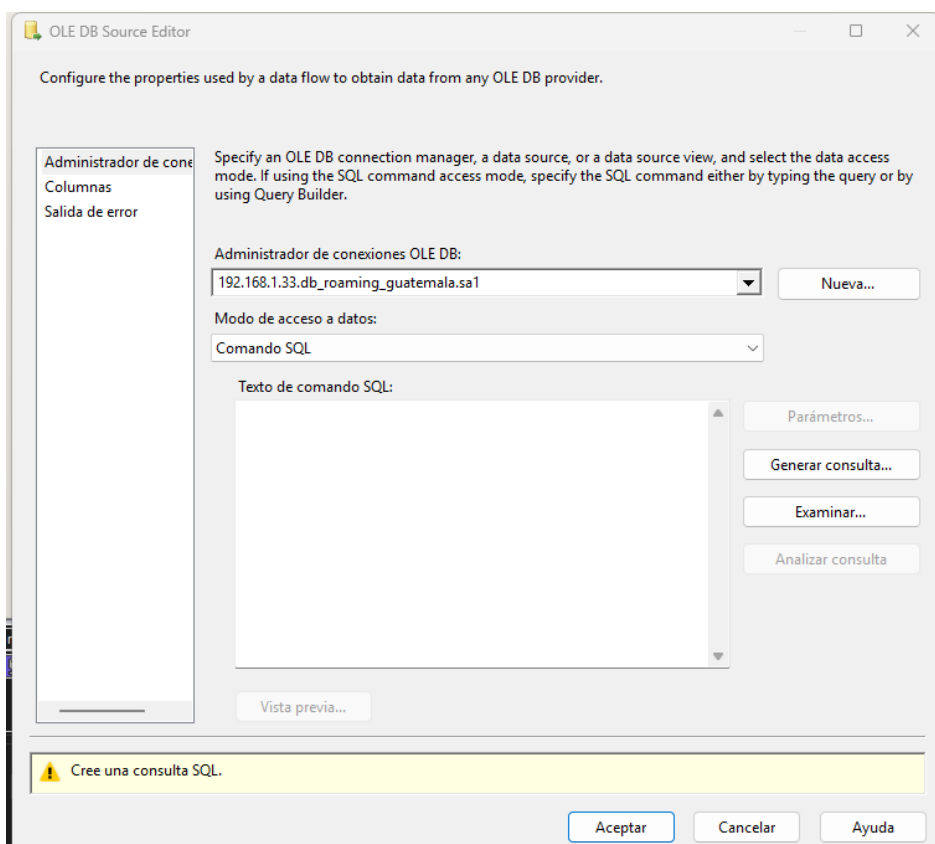




Por último, de clic en aceptar, y aceptar en la siguiente pestaña



En el formulario de propiedades, seleccionaremos “Comando SQL”



Estableceremos la consulta para extraer los registros de la base de datos, esta relaciona las tres tablas.

```
SQLQuery1.sql - 19...guatemala (sa (52))* -> X
SELECT
    dbo.facturacion.idfacturacion,
    dbo.facturacion.idcliente,
    dbo.cliente.idtipocliente,
    dbo.cliente.codigo_cliente,
    dbo.facturacion.monto_facturacion,
    dbo.tipo_clientes.nombre_tipo
FROM
    dbo.cliente
    INNER JOIN dbo.facturacion ON dbo.cliente.idcliente = dbo.facturacion.idcliente
    INNER JOIN dbo.tipo_clientes ON dbo.cliente.idtipocliente =dbo.tipo_clientes.idtipocliente
```

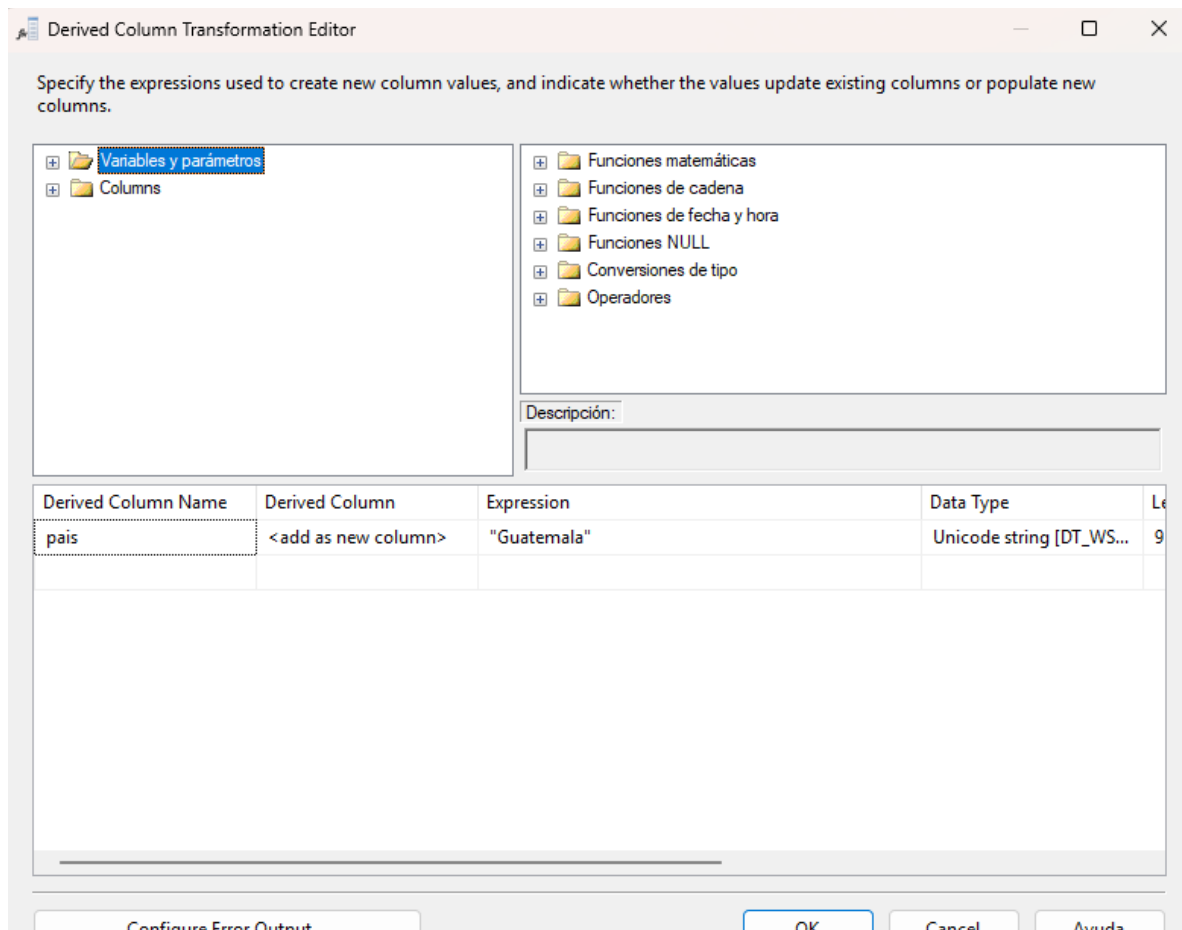
Repetiremos estos pasos con los componentes de Honduras y Costa Rica

## Fase de transformación

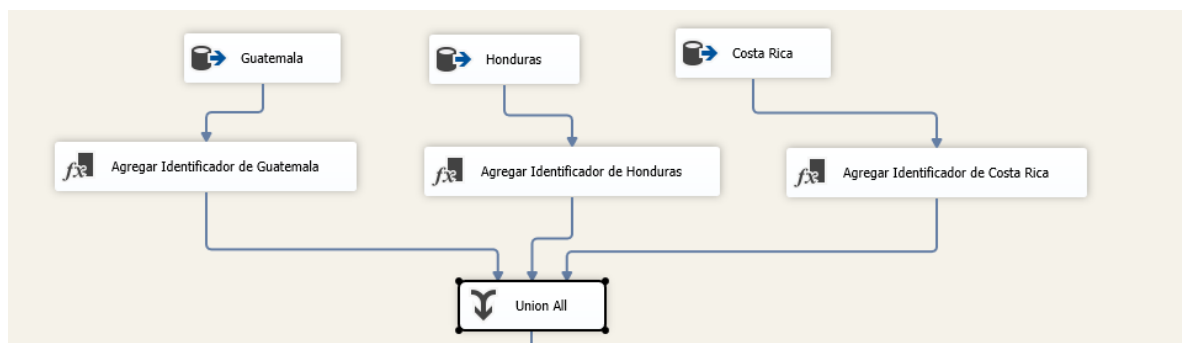
Debido a que es necesario identificar los registros por su región y este es un dato que no esta en las columnas de la base de datos, he optado por agregar una columna identificativa a los datos, para esta finalidad utilizaremos el componente “Derived Column” y Renombrare el control



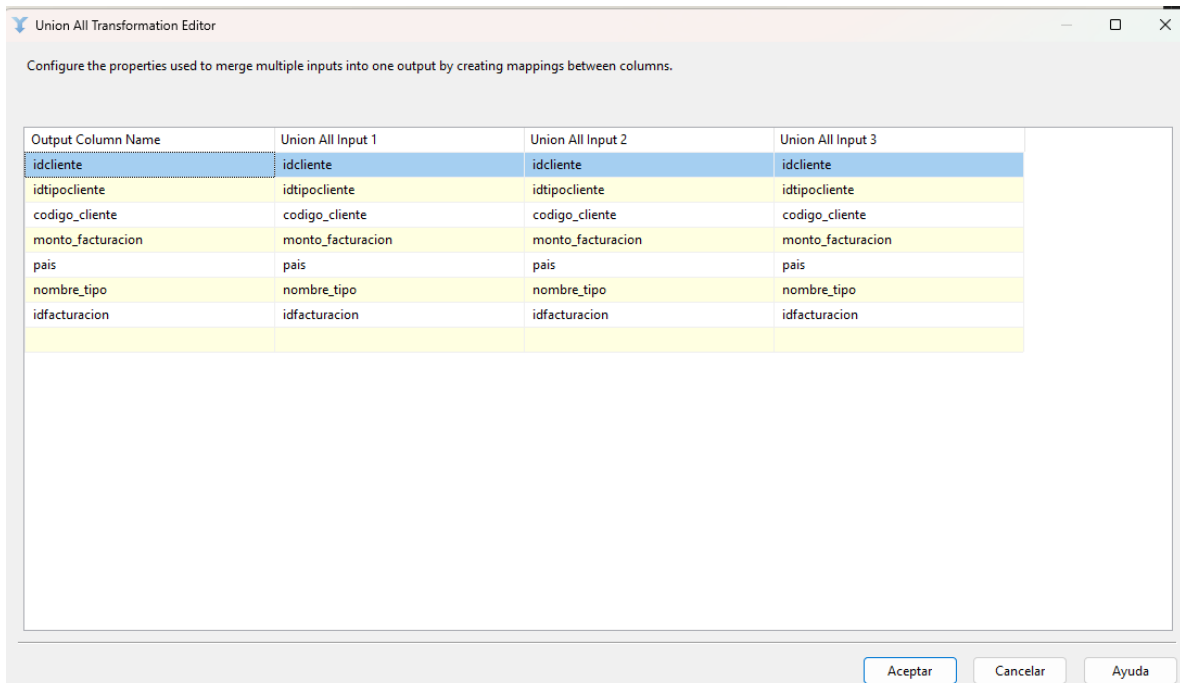
Dentro del control agregare una nueva columna llamada país, con el nombre del país que corresponda al origen de dato, y repetiré esta acción con las otras dos bases de datos



Una vez agregada la nueva columna, procederemos a combinar los datos para tener un solo dataset de datos, mediante el componente “Union All”, daremos doble clic en el componente para establecer la relación de columnas



Ya que todas las columnas, tienen el mismo nombre esto facilita la relación de datos. Por ultimo de clic en aceptar



En este punto debemos saber que hay dos reportes esperados, para el cual utilizaremos el mismo dataset:

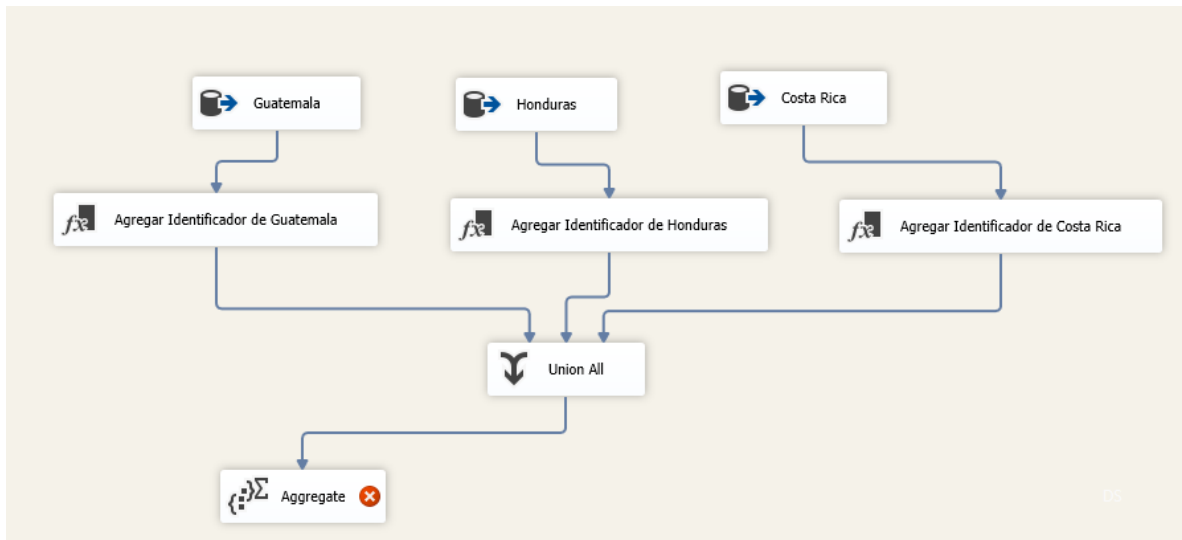
1. Consolidado de facturación por país: archivo de Excel que muestre, para cada país, el nombre del país, el número total de ventas y el monto total de facturación.
2. Consolidado de ventas por tipo de cliente: archivo de Excel que muestre, para cada tipo de cliente, el nombre del tipo de cliente, el número total de ventas y el monto total de facturación.

Debido a que nuestro dataset es un JOIN de las tres columnas debemos agrupar por los datos claves en los cuales es requerida la información.

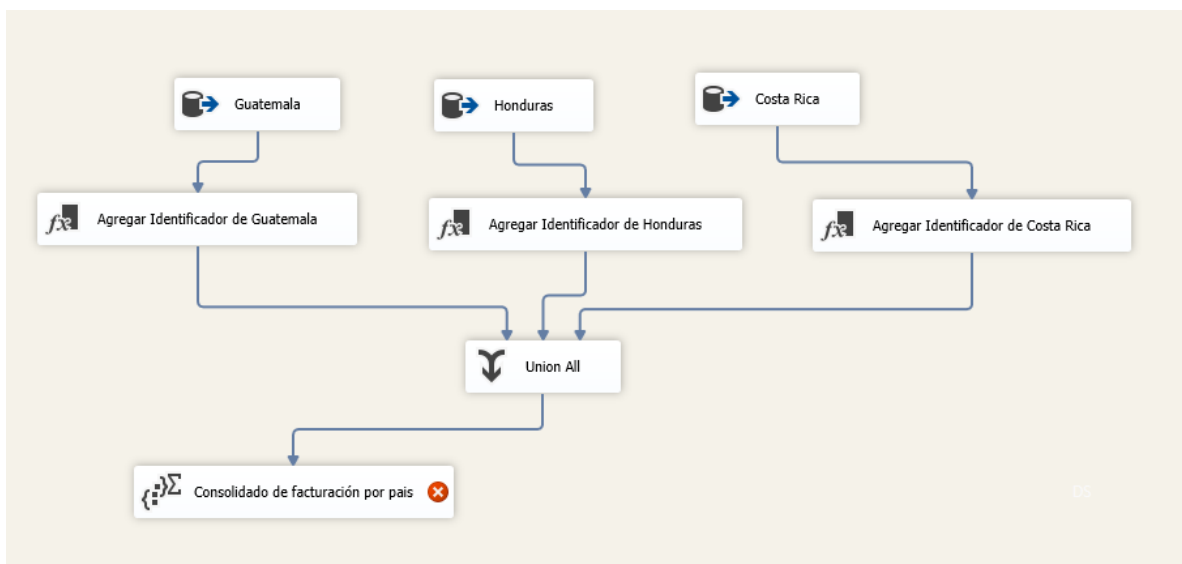
Abordando el 1 ítem:

- las columnas por las que agruparemos serán el País.
- Ya que se requiere conocer el total de ventas sin alguna otra distinción, utilizaremos la función Count() sobre el campo facturacion, para saber contar cuantos id de facturaciones diferentes se realizaron.
- utilizaremos la función SUM() para totalizar el monto de facturación entre todas las facturas, sobre la columna monto facturación

Para establecer estas operaciones, utilizare un componente “Aggregate” para esta salida:



La renombrare como “Consolidado de facturación por país”



Procederemos a realizar la parametrización como lo hemos planteado, para identificar mejor las columnas de salida agregare el tipo de función que se ha aplicado como prefijo del nombre, al finalizar demos clic en “OK”

Aggregate Transformation Editor

Aggregations Advanced

Configure the properties used to perform group by operations and to calculate aggregate values. Optionally, apply comparison options to the operation. To configure multiple group by operations, click Advanced.

Advanced

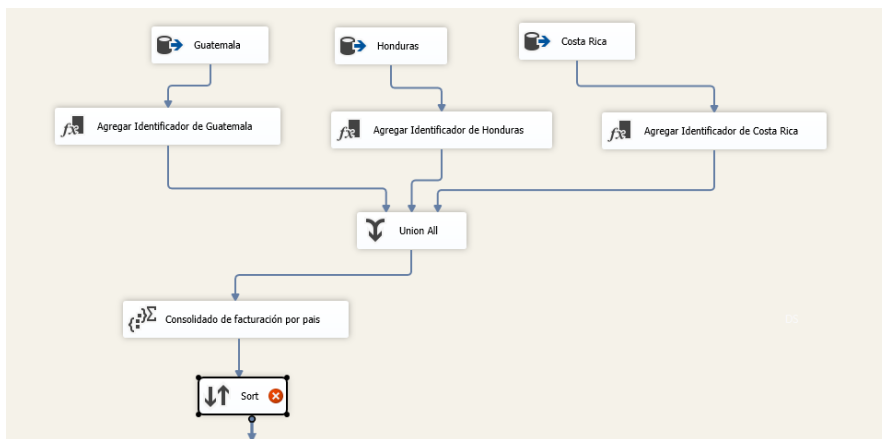
Columnas de entrada disponi...

<input checked="" type="checkbox"/>	Nombre
<input type="checkbox"/>	(*)
<input type="checkbox"/>	idcliente
<input type="checkbox"/>	idtipocliente
<input type="checkbox"/>	codigo_cliente
<input checked="" type="checkbox"/>	monto_facturacion

Input Column	Output Alias	Operation	Compa
idfacturacion	count_idfacturacion	Count distinct	
monto_facturacion	sum_monto_facturacion	Sum	
pais	pais	Group by	

OK Cancel Ayuda

No es solicitado, pero agregare un ordenamiento en base al monto facturado sobre la columna “monto\_facturado”



El orden seleccionado será “descending” para posicionar al país con mayores ventas en la primera fila, luego se descenderá con los siguientes

Sort Transformation Editor

Specify the columns to sort, and set their sort type and their sort order. All nonselected columns are copied unchanged.

Columnas de entrada disponibles		
<input checked="" type="checkbox"/>	Nombre	Paso a través
<input type="checkbox"/>	count_idfacturacion	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/>	sum_monto_facturacion	<input checked="" type="checkbox"/>
<input type="checkbox"/>	pais	<input checked="" type="checkbox"/>

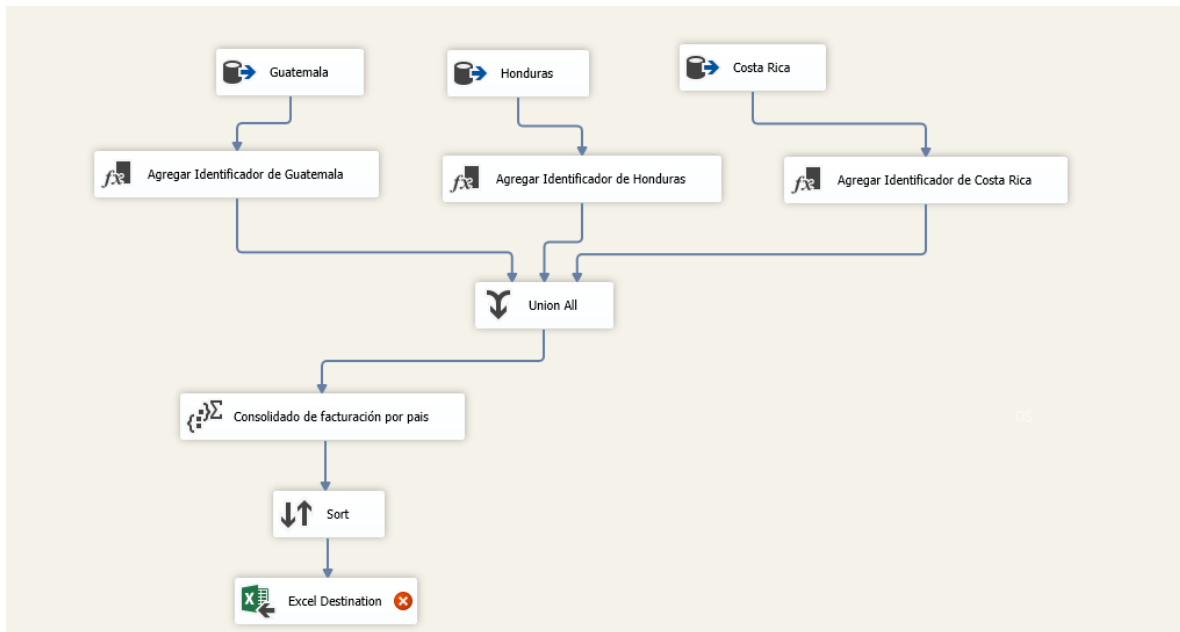
Input Column	Output Alias	Sort Type	Sort Order	Con
sum_monto_facturacion	monto_facturacion	descending	1	

☐ Remove rows with duplicate sort values

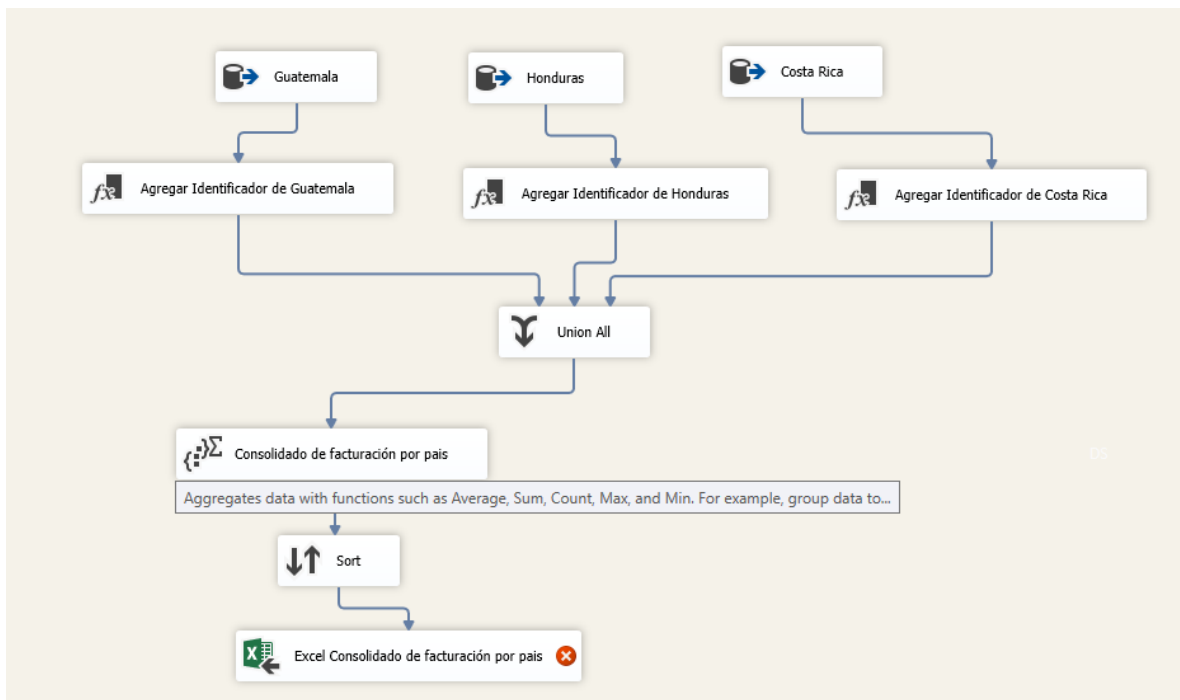
OK Cancel Ayuda

## Fase de Carga

Por último, para este primer punto agregaremos el componente “Excel Destination”, para exportar los resultados a Excel

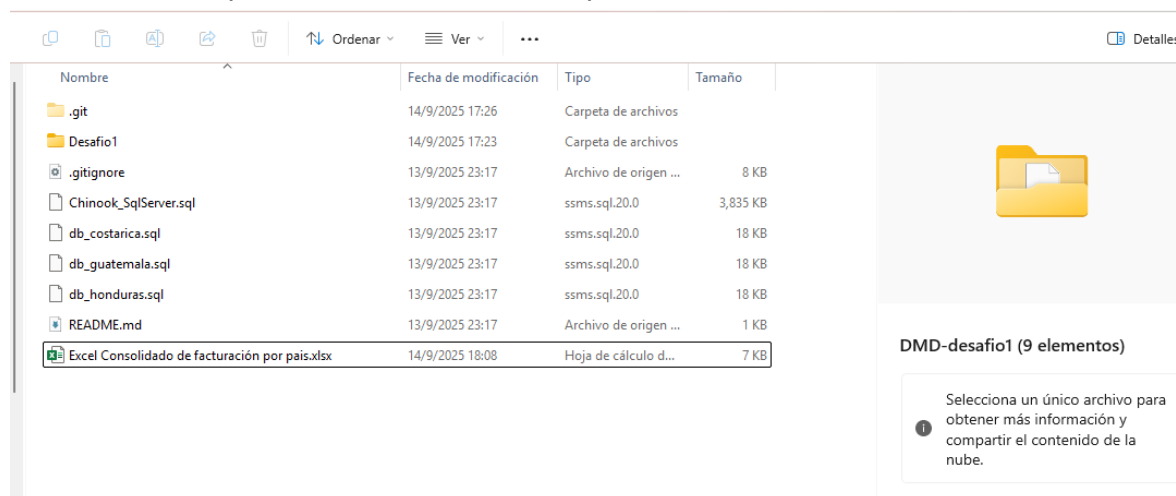


Y le daremos un nombre más descriptivo

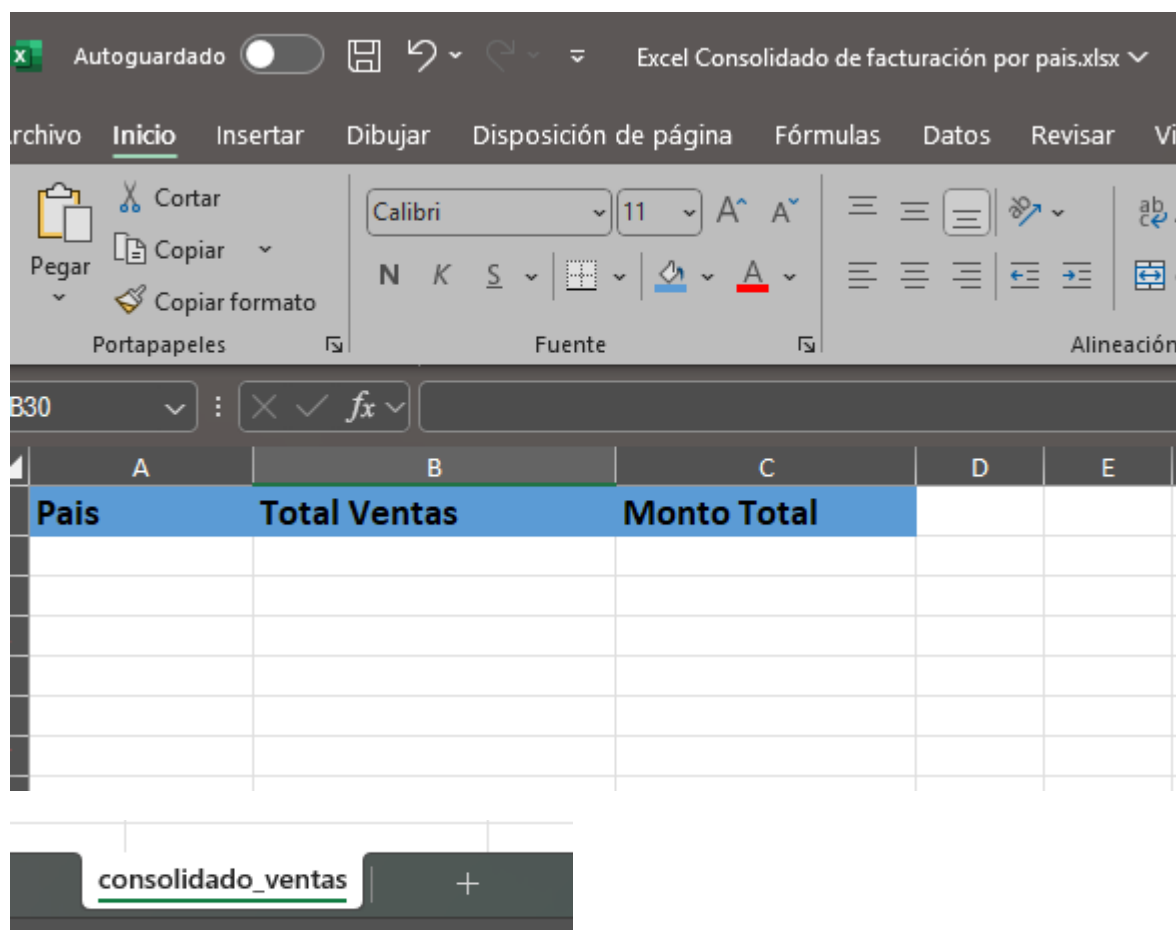




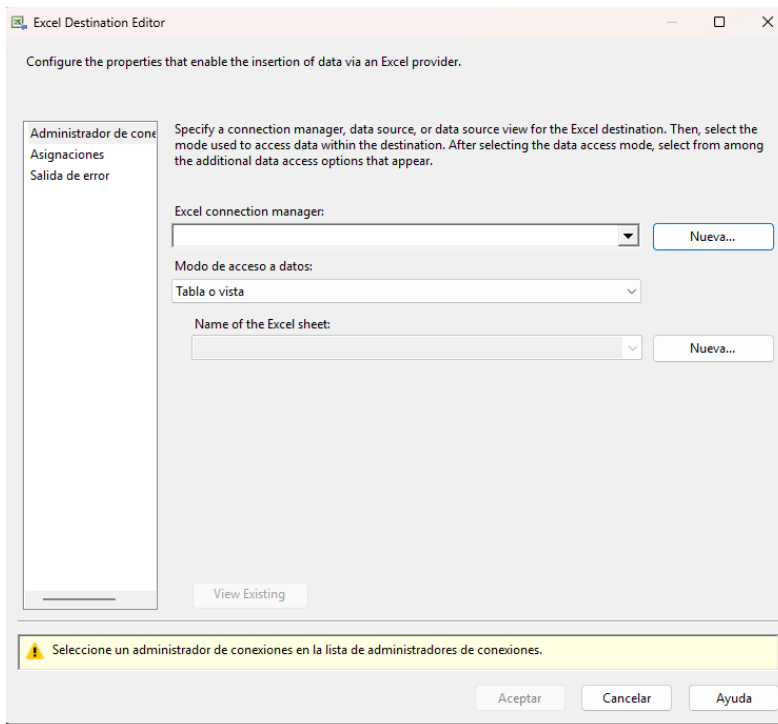
Pasaremos a configurar el destino, para esto inicialmente crearemos dentro del directorio del repositorio un archivo Excel para recibir los datos



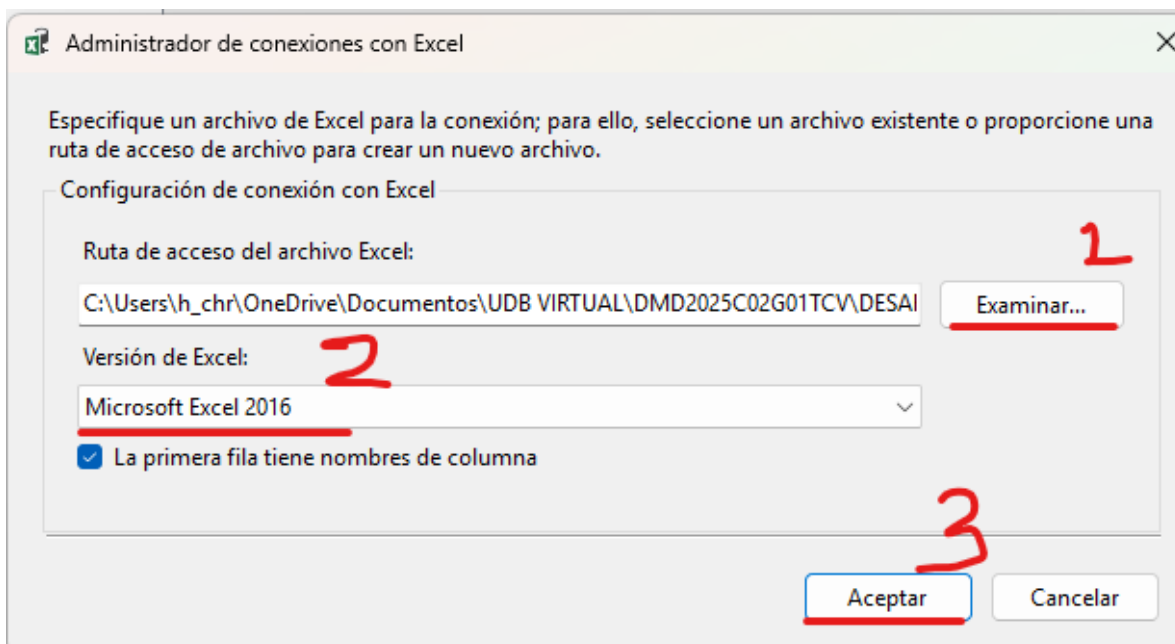
Dentro del cual definire el siguiente formato, y nombrare a la hoja de Excel como “consolidado\_ventas”



Dentro del componente “Excel Destination” de clic en “Nueva...”



Y dentro de la ventana emergente seleccionaremos “Examinar” y seleccionaremos el archivo de Excel que creamos, además seleccionaremos la versión mas reciente de Excel, posteriormente daremos clic en aceptar.



Al volver al editor principal seleccionaremos además la tabla o vista, y seleccionaremos la hoja consolidado\_ventas que creamos, después nos dirigiremos a “Asignaciones”

Excel Destination Editor

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

Administrador de cone  
Asignaciones 3  
Salida de error

Specify a connection manager, data source, or data source view for the Excel destination. Then, select the mode used to access data within the destination. After selecting the data access mode, select from among the additional data access options that appear.

Excel connection manager:  
Administrador de conexiones con Excel Nueva...

Modo de acceso a datos:  
Tabla o vista 1

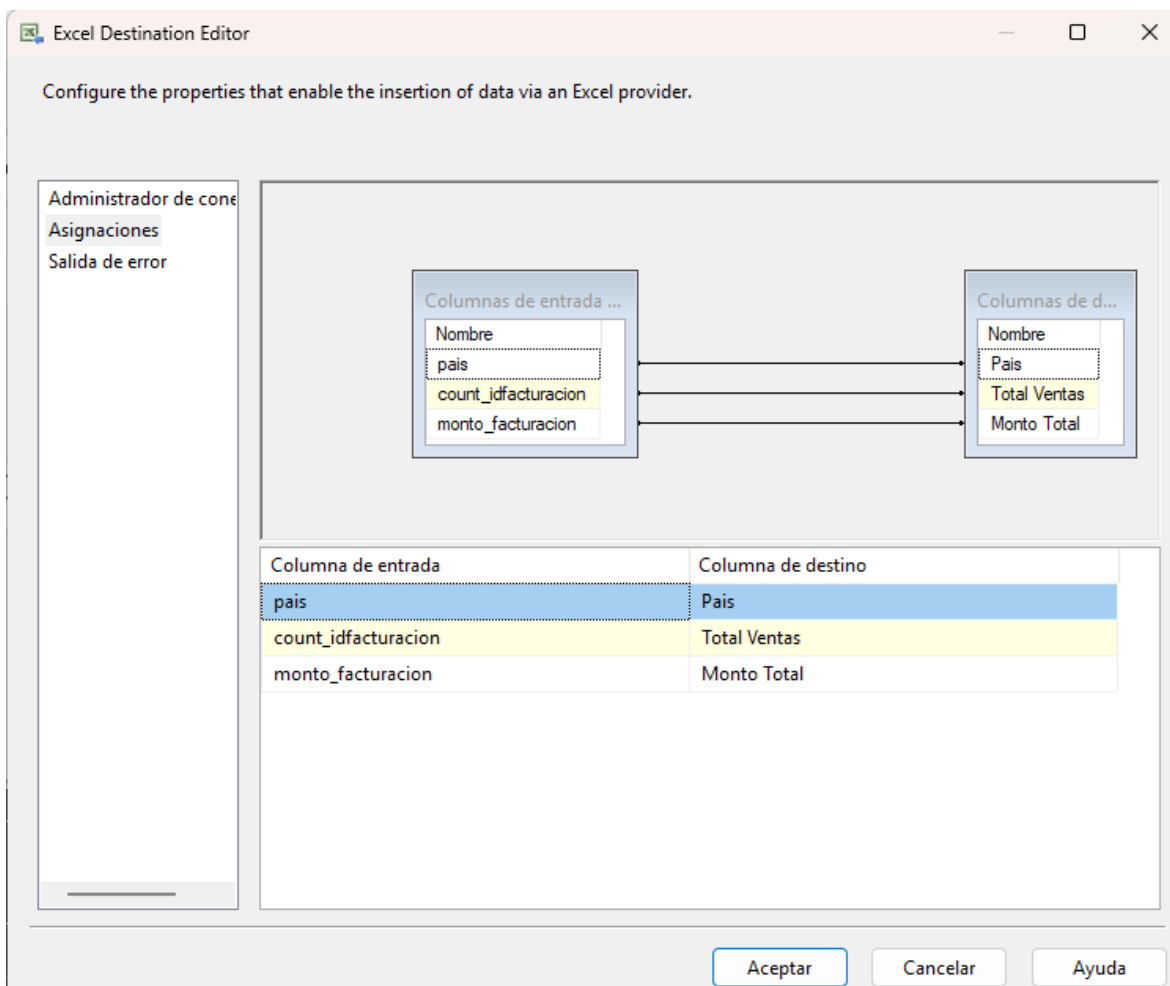
Name of the Excel sheet:  
consolidado\_ventas\$ 2 Nueva...

View Existing

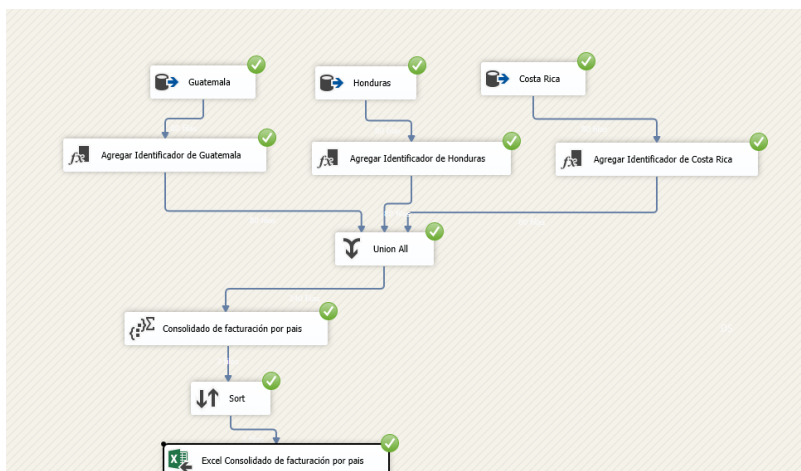
⚠ Asigne las columnas en la página Asignaciones.

Aceptar Cancelar Ayuda

Dentro de esta pestaña asignaremos cada columna con su respectiva columna del archivo Excel



Ejecutaremos el flujo para corroborar que el flujo se ejecuta correctamente, y observamos que todos los pasos se ejecutaron de forma exitosa



Revisaremos el contenido del Excel, y vemos los resultados.

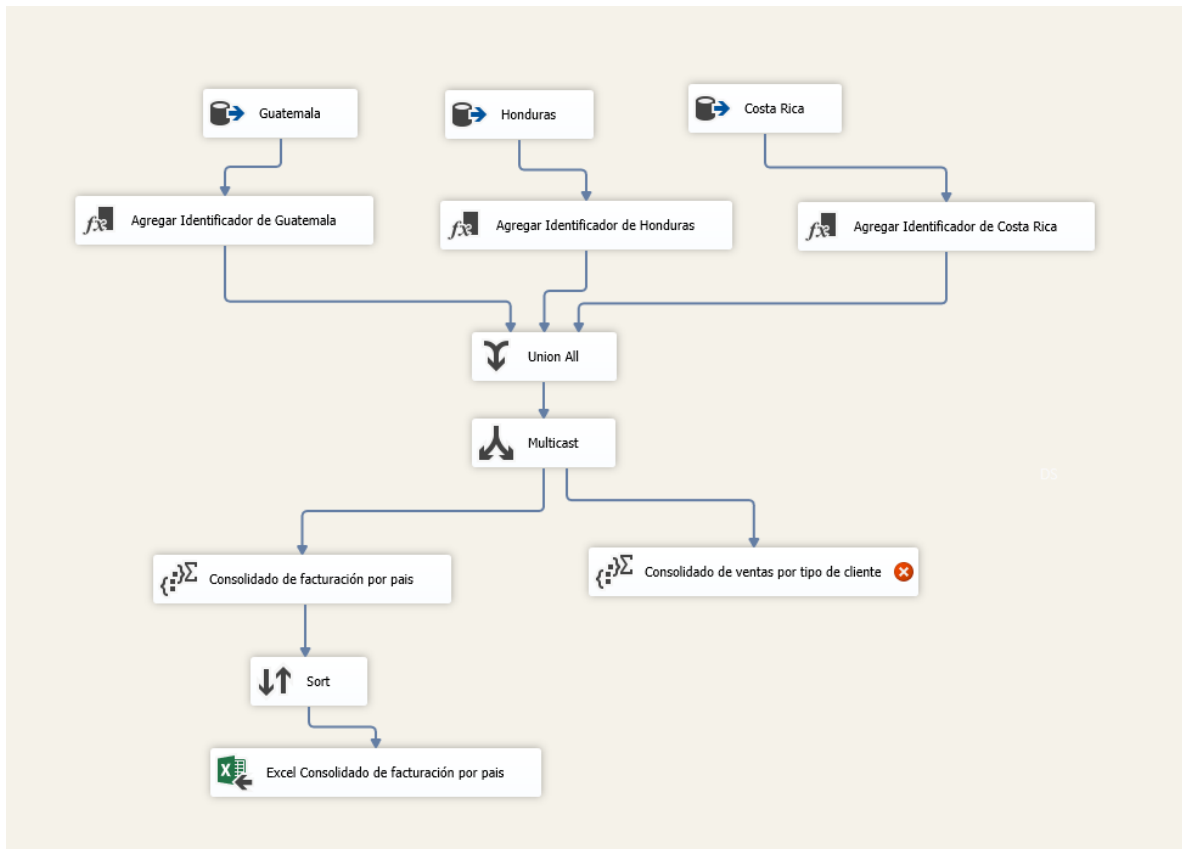
	A	B	C	D	E	F
1	<b>Pais</b>	<b>Total Ventas</b>	<b>Monto Total</b>			
2	Costa Rica	80	9503.10			
3	Honduras	80	8703.10			
4	Guatemala	80	8344.90			
5						

Abordemos ahora el item 2:

2. Consolidado de ventas por tipo de cliente: archivo de Excel que muestre, para cada tipo de cliente, el nombre del tipo de cliente, el número total de ventas y el monto total de facturación.
- las columnas por las que agruparemos será, de la tabla tipo\_cliente, nombre\_tipo.
  - Ya que se requiere conocer el total de ventas, utilizaremos la función count\_distinct() sobre el campo facturacion, para saber contar cuantos id de facturaciones se realizaron.
  - utilizaremos la función SUM() para totalizar el monto de facturación entre todas las facturas, sobre la columna monto facturación

Para establecer estas operaciones, utilizare un componente “Aggregate” para esta salida:

Ugregaremos otro componente “Aggregate”, pero dado a que el componente Union All solo permite una sola salida, haremos una modificación en esta parte, utilizaremos un componente llamado “Multicast” para conectar la salida de “Union All” hacia “Multicast” posteriormente, dirigiremos una salida de multicas hacia el correspondiente “Aggregate”, de esta forma:



Procedamos a configurar el “Aggregate” para “Consolidado de ventas por tipo de cliente”

Aggregate Transformation Editor

Aggregations Advanced

Configure the properties used to perform group by operations and to calculate aggregate values. Optionally, apply comparison options to the operation. To configure multiple group by operations, click Advanced.

Advanced

Columnas de entrada disponibles

- ☒ Nombre
- ☐ (\*)
- ☐ idcliente
- ☐ idtipocliente
- ☐ codigo\_cliente
- ☒ monto\_facturacion
- ☐ pais
- ☒ nombre\_tipo
- ☒ idfacturacion

Input Column	Output Alias	Operation	Comparison Flags
nombre_tipo	nombre_tipo	Group by	
idfacturacion	count_idfacturacion	Count distinct	
monto_facturacion	sum_monto_facturacion	Sum	

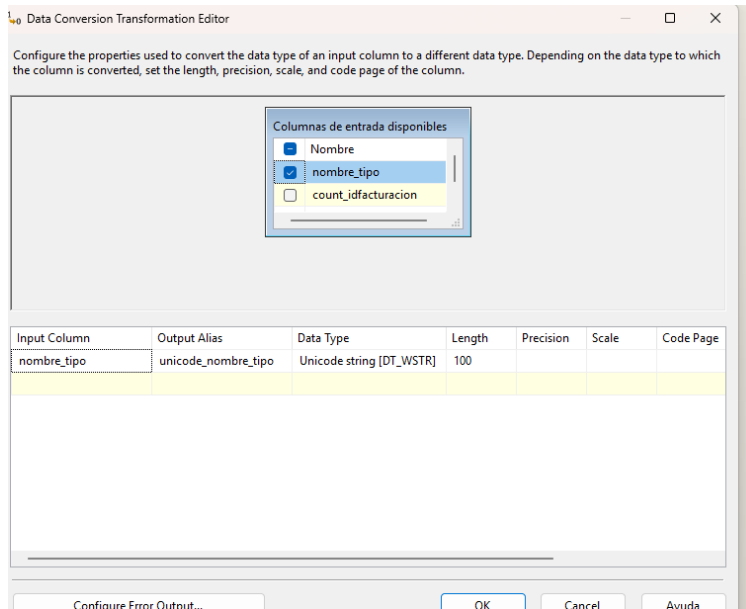
OK Cancel Ayuda

Agregaremos un componente “Sort” para la columna monto\_facturacion de forma “descending”


En este flujo hacemos uso de la columna nombre\_tipo, el cual nos genera un problema de formato de caracteres, por lo que será necesario convertirlo para poder exportarlo Excel, para esto utilizaremos el componente “Data Conversion”



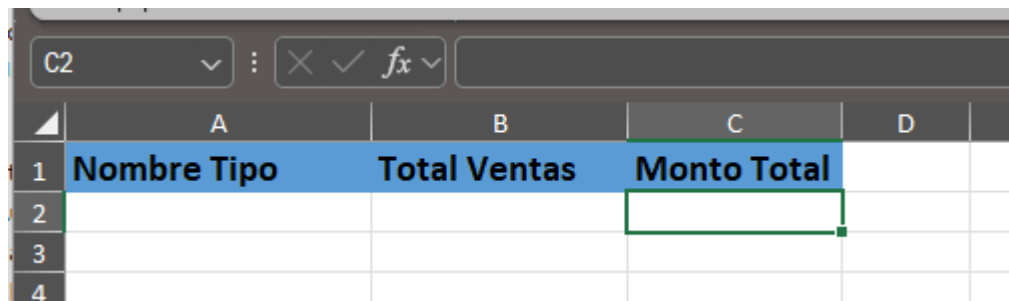
Dentro de él realizaremos la conversión del campo “nombre\_tipo” a un formato de caracteres Unicode, a la cual asignaremos un nuevo nombre “unicode\_nombre\_tipo”:



Ahora continuaremos con el flujo crearemos un archivo de Excel para recibir los datos del flujo:

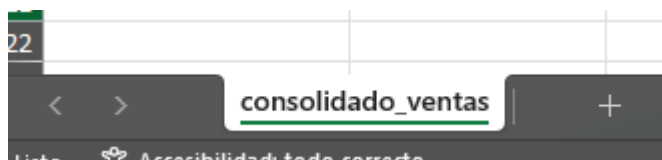
 Excel Consolidado de ventas por tipo de cliente.xlsx

Utilizaremos las siguientes columnas



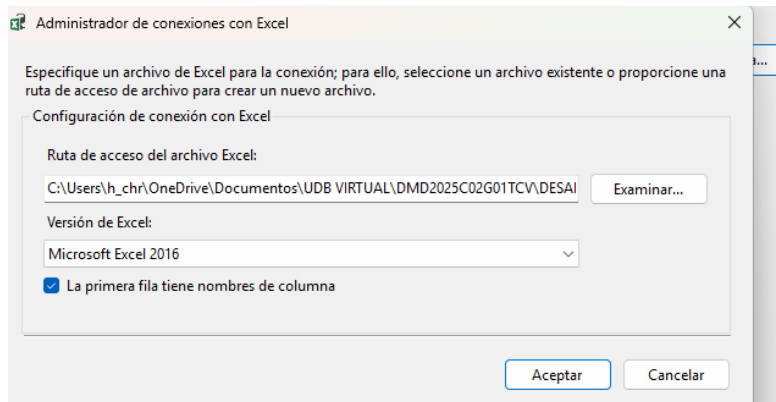
	A	B	C	D
1	Nombre Tipo	Total Ventas	Monto Total	
2				
3				
4				

He igualmente nombrare el nombre de la hoja como “consolidado\_ventas”

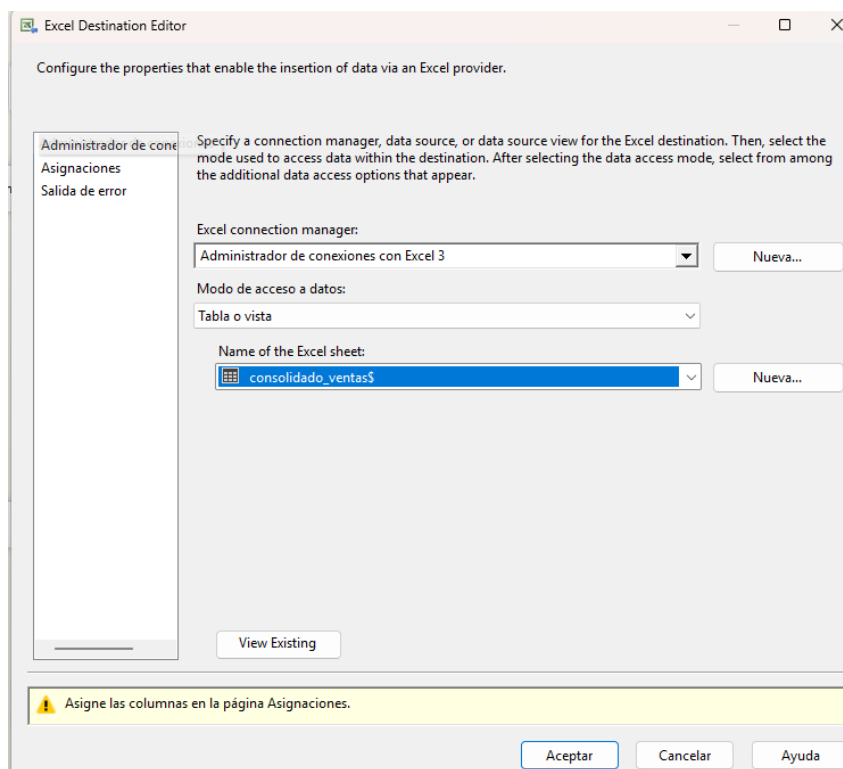


Establecere la configuración de la ruta del archivo:

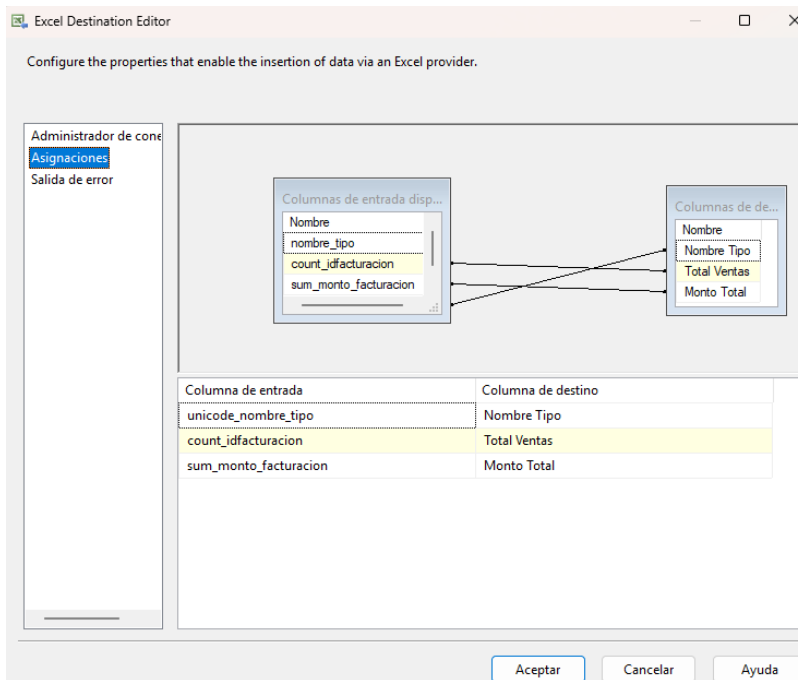




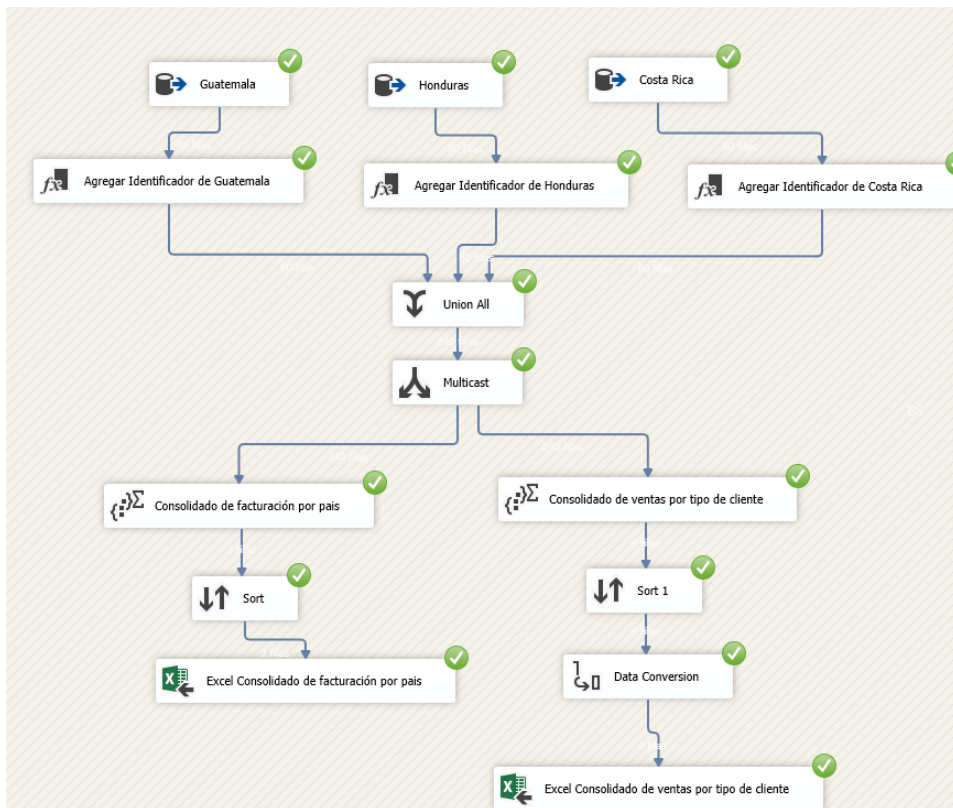
## El nombre de la hoja de destino



Y la asignación de las columnas de destino, en esta sección, debido a la conversión que generamos anteriormente, no haremos uso de la columna “nombre\_tipo”, en su lugar utilizaremos la columna nueva “unicode\_nombre\_tipo”



Por último ejecutaremos el flujo nuevamente para verificar la generación de los datos, validamos que todo el flujo se ejecute correctamente



Y verificamos la salida del segundo archivo

Portapapeles		Fuente		Alineación	
C2		7357.50			
1	A	B	C	D	E
1	Nombre Tipo	Total Ventas	Monto Total		
2	CLIENTE EJECUTIVO	21	7357.50		
3	CLIENTE PREFERENCIAL	17	6771.10		
4	CLIENTE GUBERNAMENT	21	6220.70		
5	CLIENTE TURISTA	21	6201.80		
6					
7					

Validando que se genera correctamente