

## Actividad Dirigida: Carga incremental de datos con SSIS

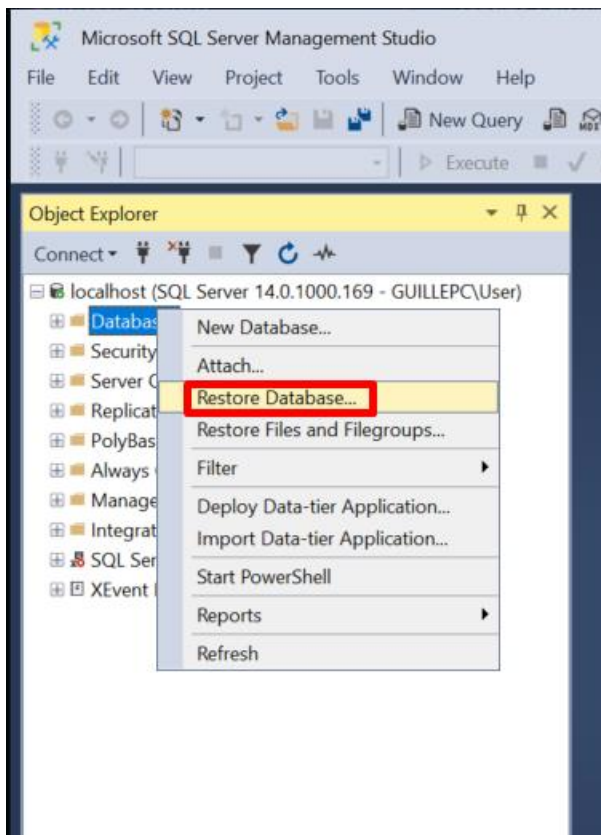
Para realizar la práctica es necesario validar los siguientes requisitos:

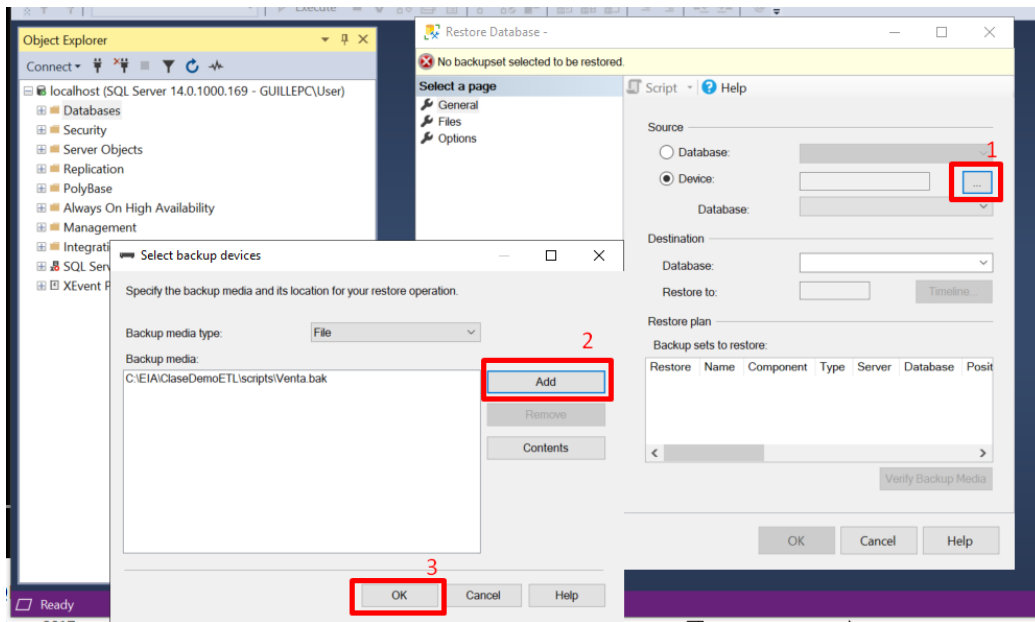
1. Acceso a SQL Server con perfil de lectura, escritura y borrado
2. Microsoft SQL Server Management Studio, version 12 o superior
3. Microsoft Visual Studio, versión 2017 o superior
4. SQL Server Data Tools, instalado (Descargar e Instalar SQL Server Data Tools BI: <https://youtu.be/zn0hgsjl8LQ> )
5. SSIS: SQL Server Integration Services para el despliegue
6. Archivos backup de las bases de datos de prueba transaccional y Data warehouse

En este documento explico un ejemplo básico de proceso ETL carga incremental de datos que es aplicado a la realidad en el campo de inteligencia de negocios. El ejercicio consiste en cargar una base de datos destino (una base de datos multidimensional) con información recién ingresada en una base de datos origen (base de datos relacional). Ambas bases de datos son prácticamente la misma a diferencia que la de destino está en función de dimensiones.

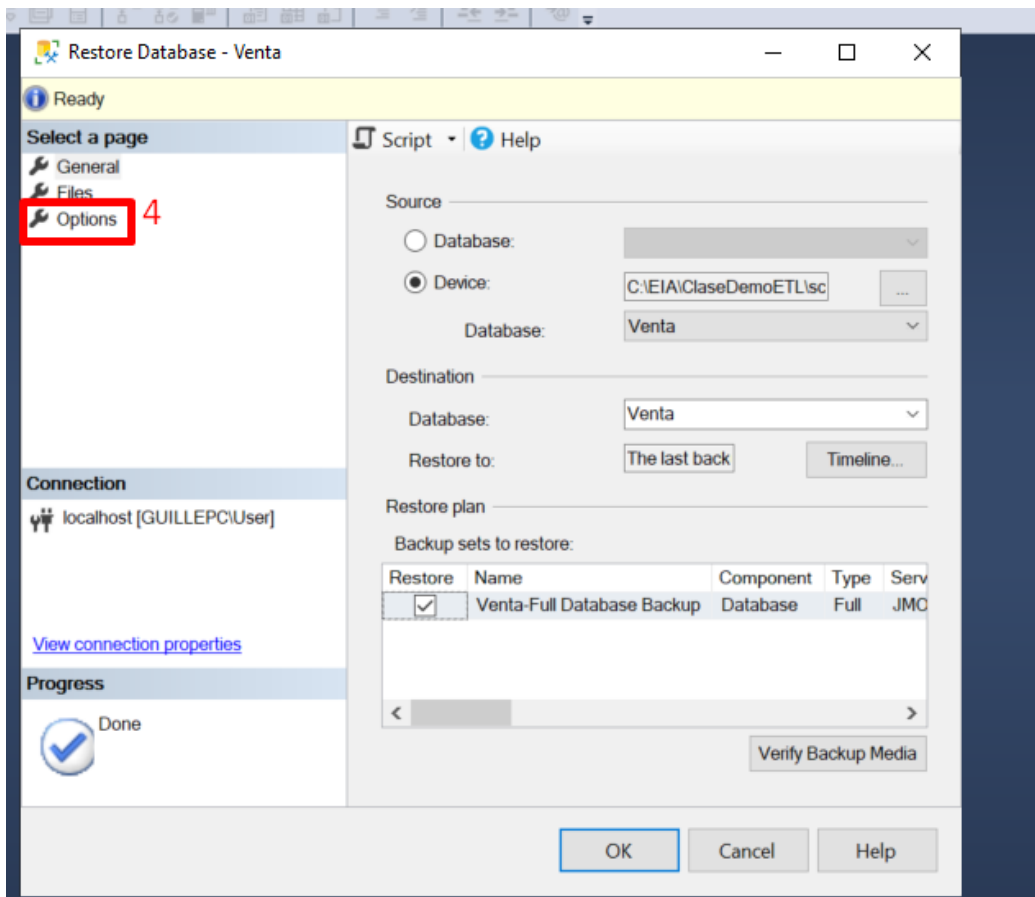
La cantidad de productos ingresados en la base de datos origen deben cargarse previa ejecución del proyecto a una base de datos destino.

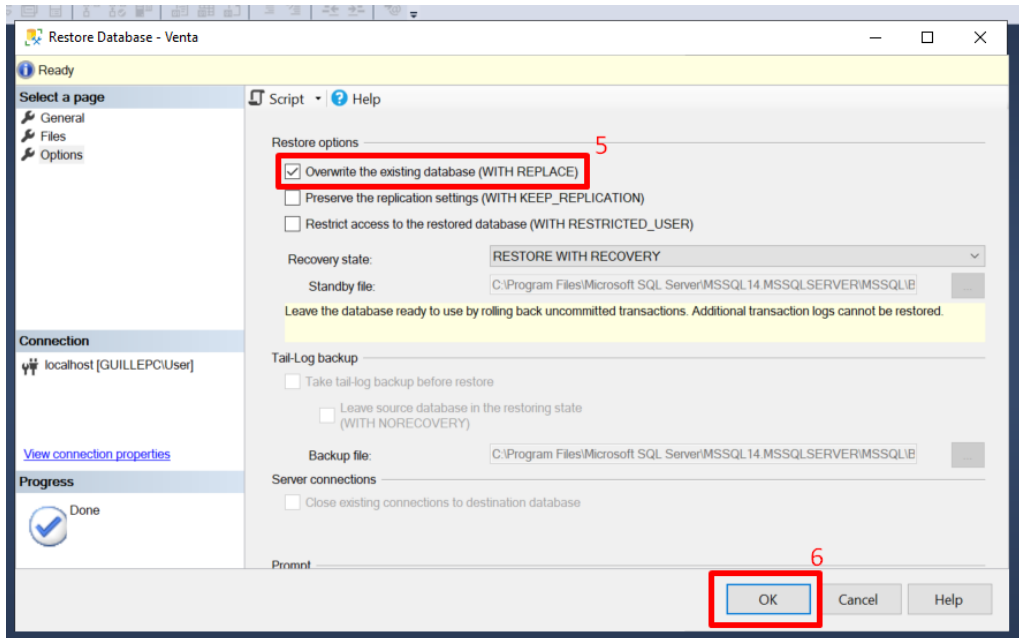
**Primer Paso:** Restaurar las bases de datos en SQL Server a través de Management Studio



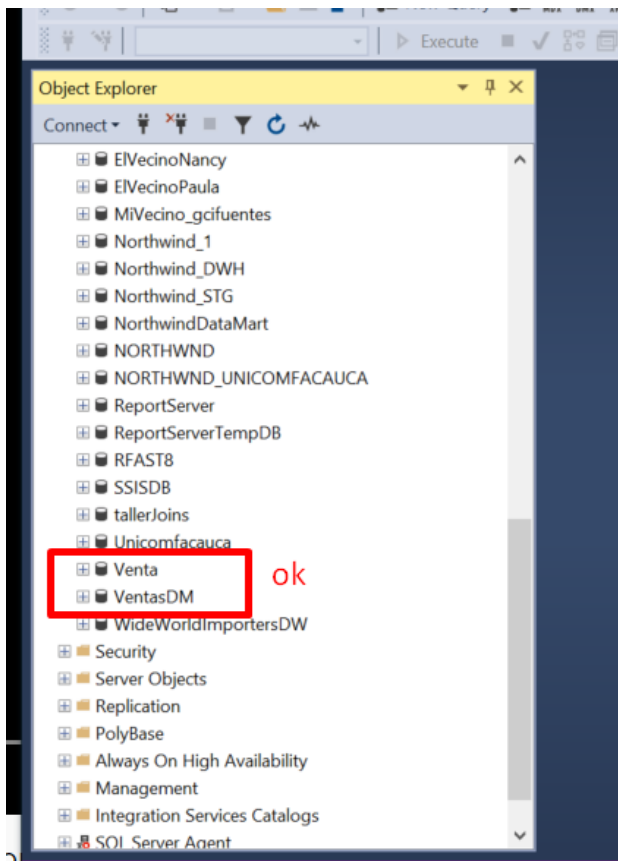


Previamente se ha descomprimido el archivo .RAR con los archivos .bak bien sea en la carpeta de backup de SQLServer o en una ubicación establecida a conveniencia.

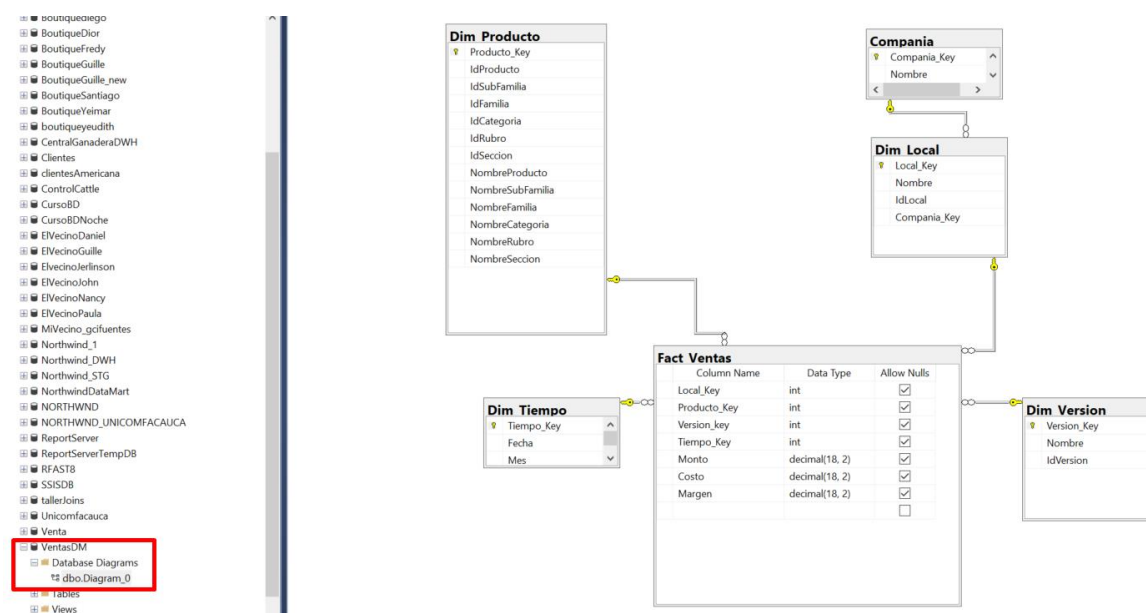
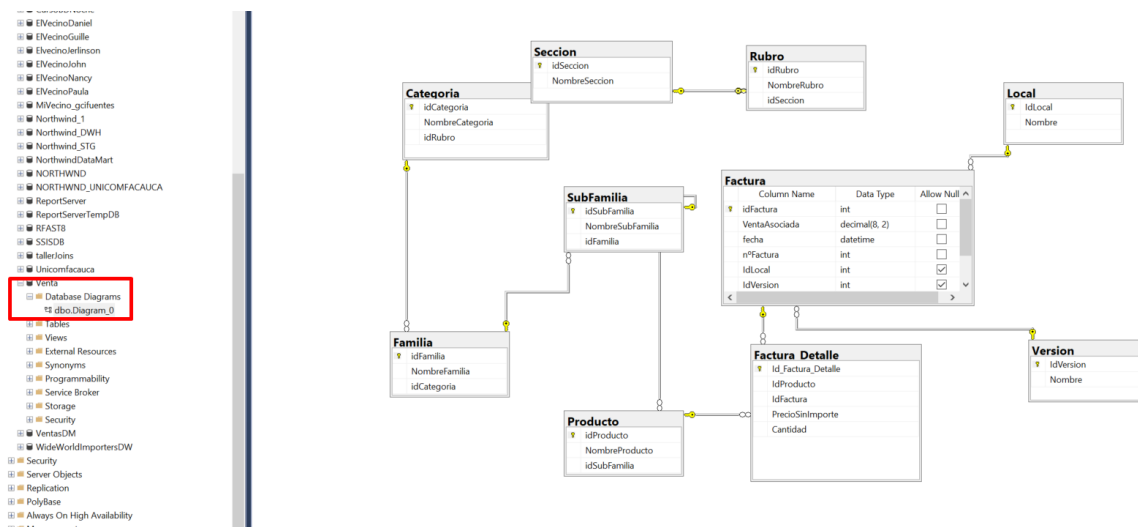




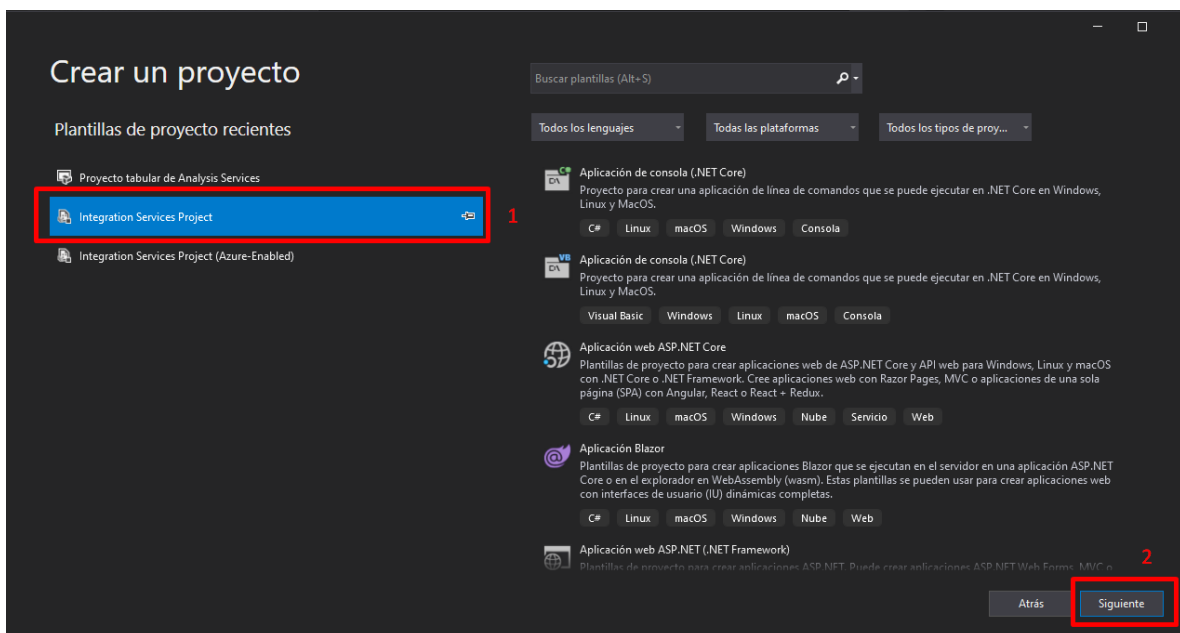
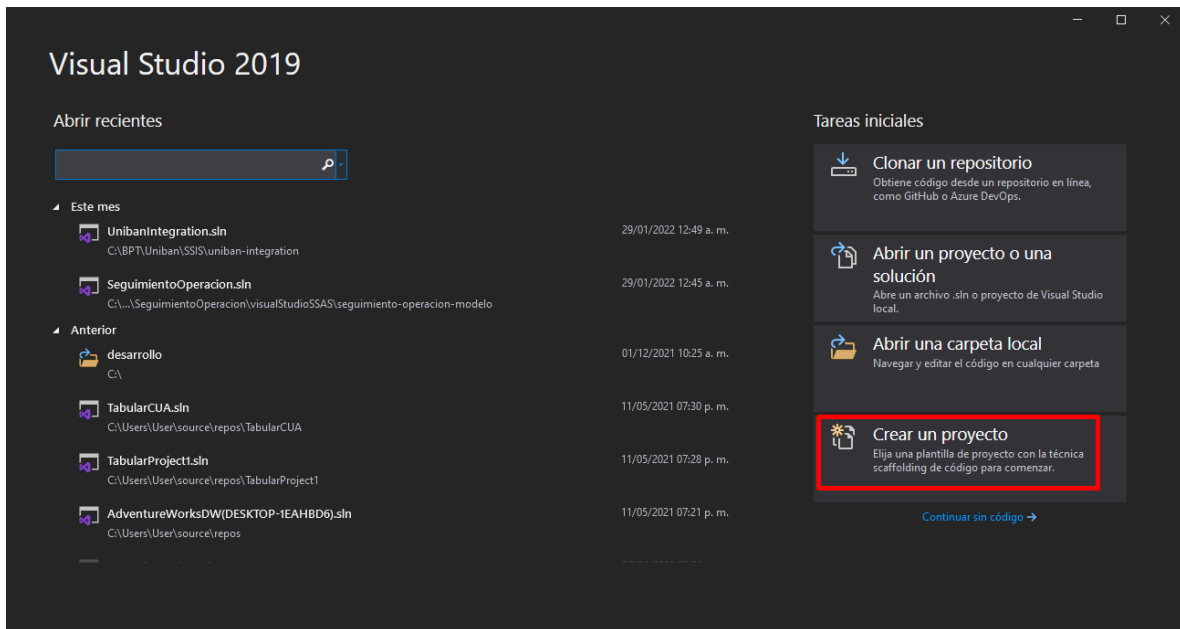
Se hace el mismo proceso para Ventas DM



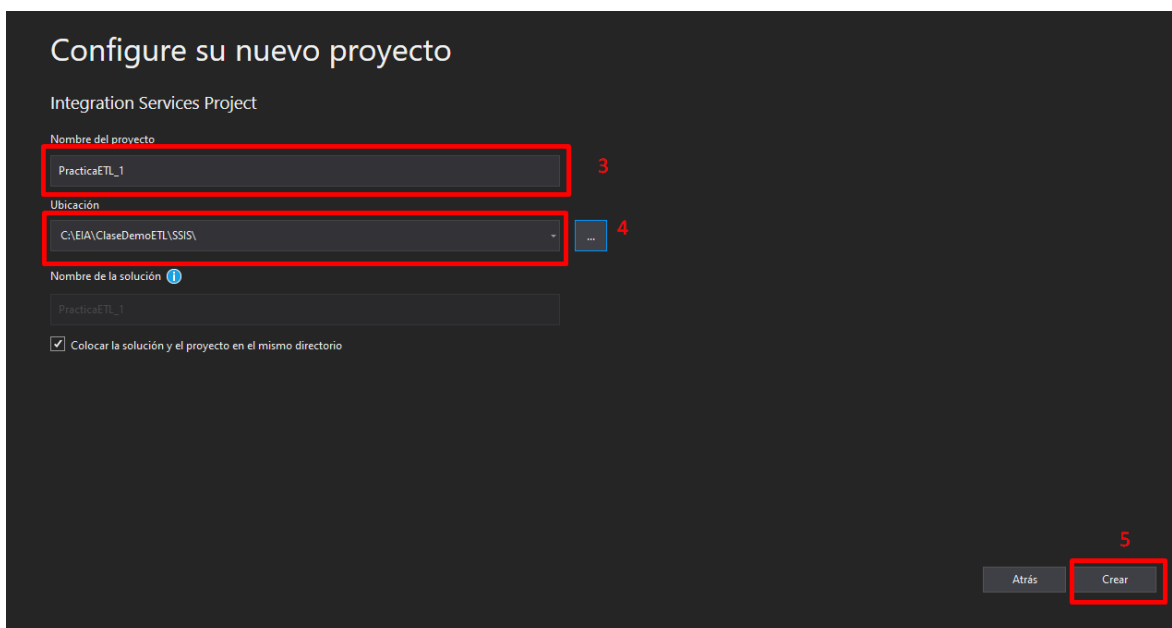
## Paso 2: Visualizar el modelo de datos de las bases de datos Ventas y VentasDM



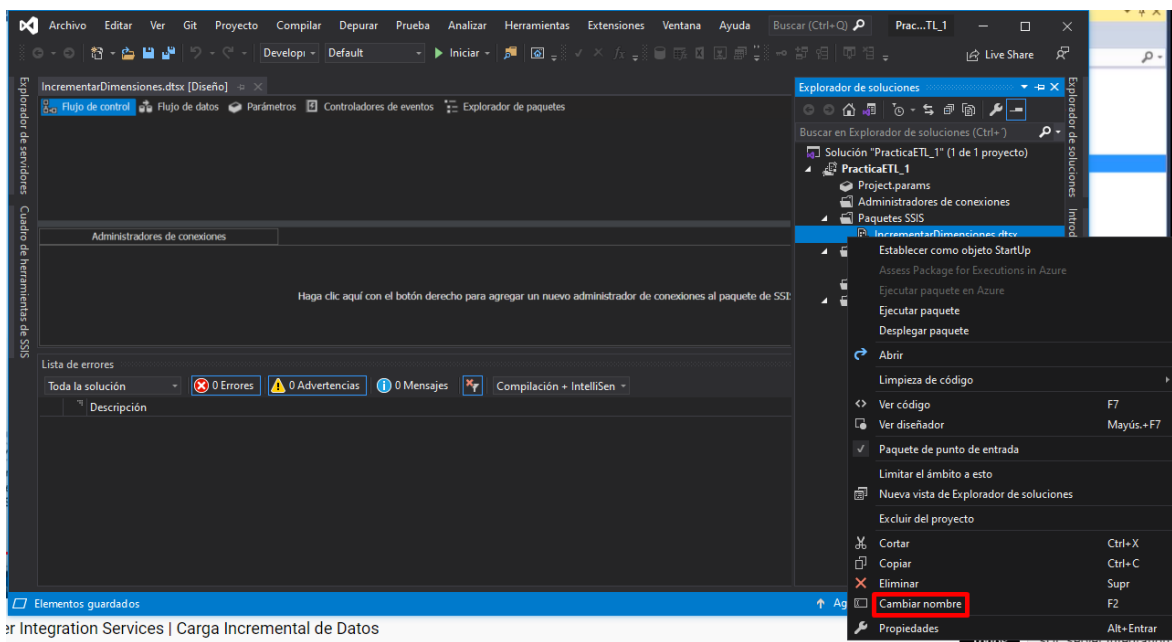
### Paso 3: Crear un nuevo proyecto de Integration Services en Visual Studio



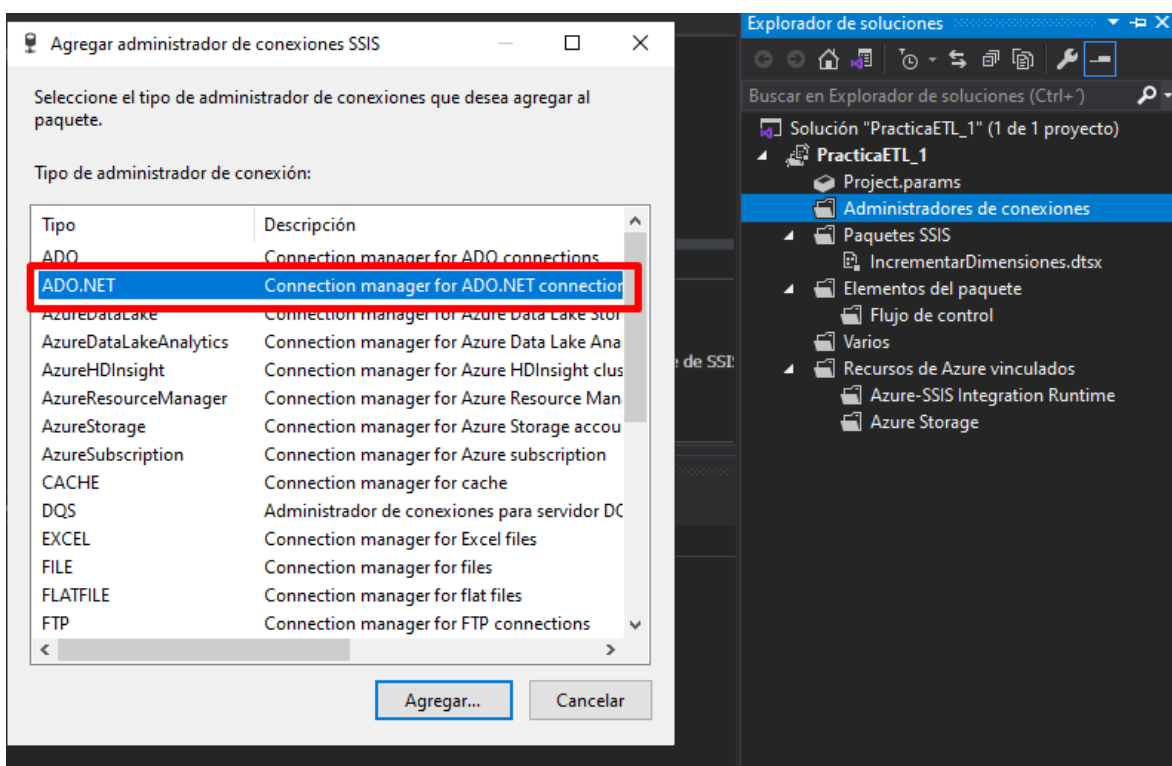
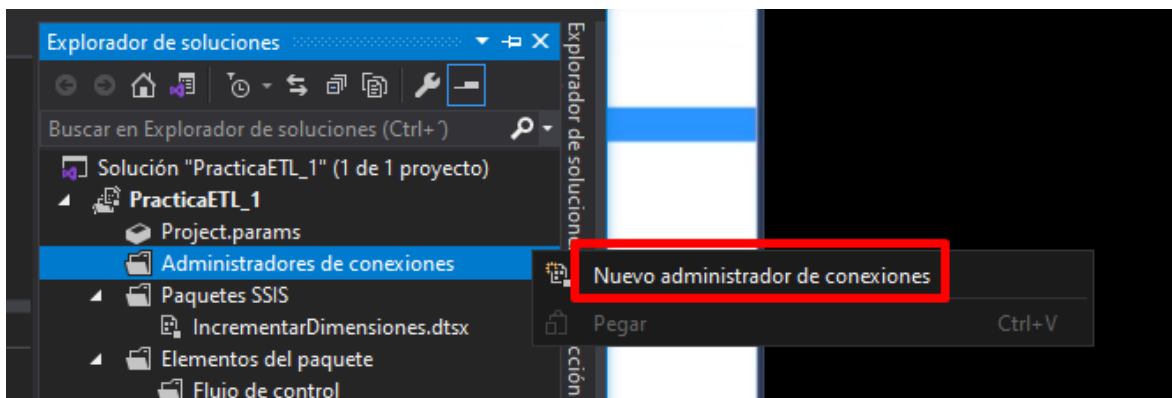
Tener claro el repositorio donde queremos que quede nuestra aplicación

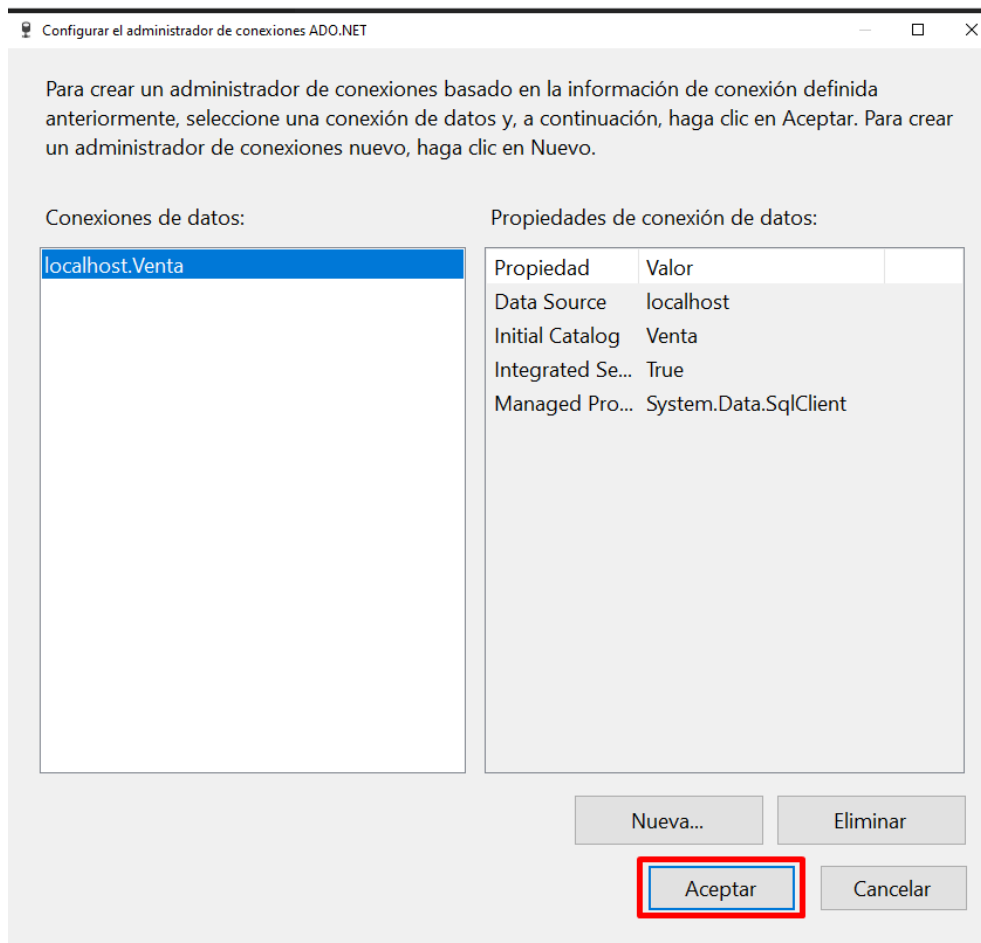
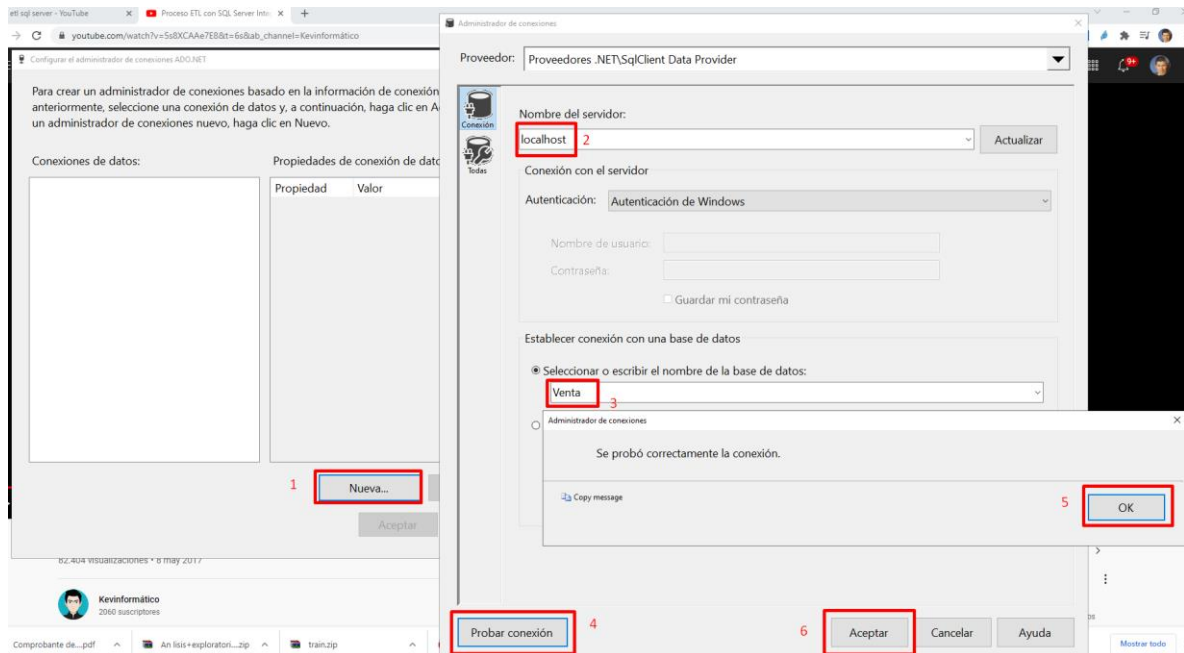


Por defecto en el explorador de soluciones se crea el primer paquete, se recomienda cambiar el nombre de este (se cambio por **IncrementarDimensiones.dtsx**)

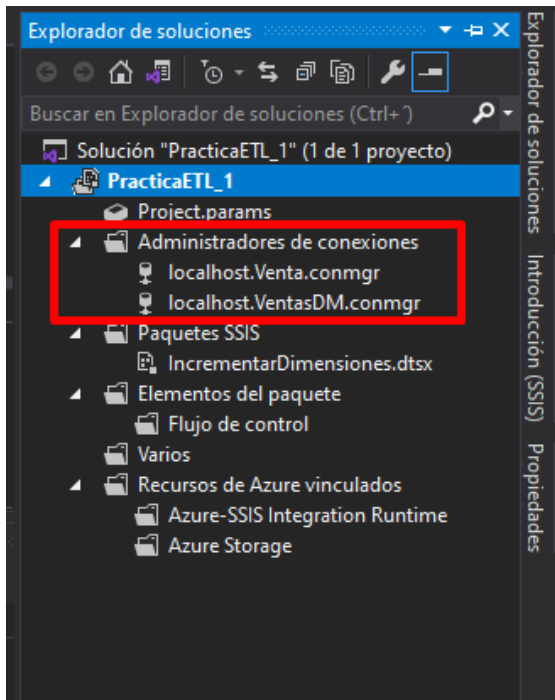
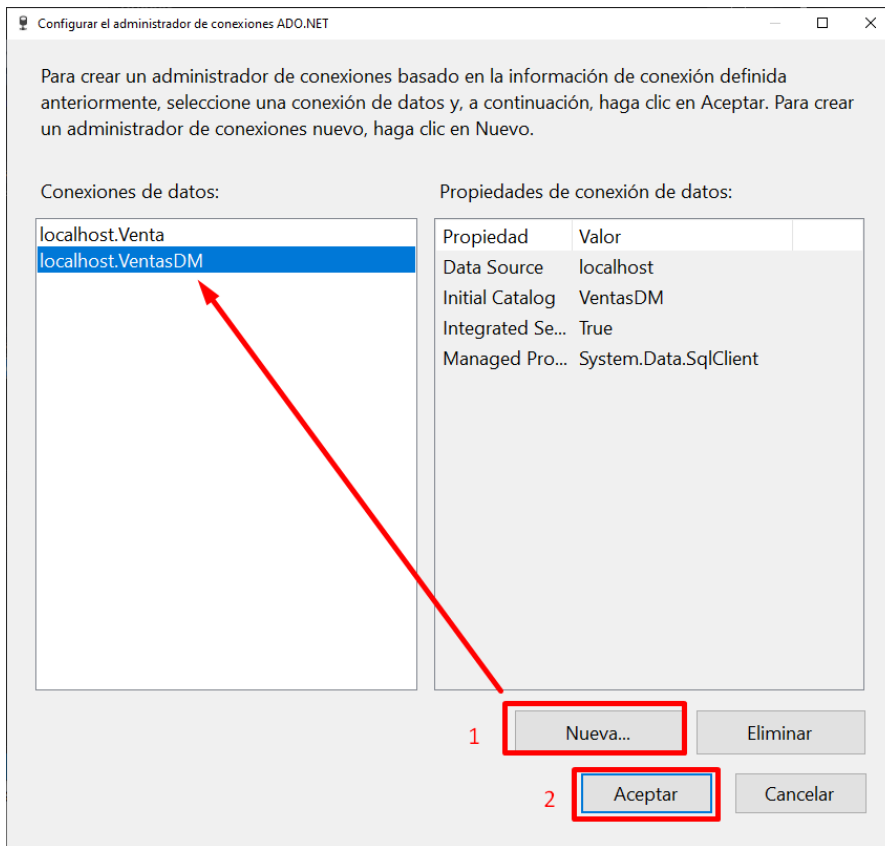


**Paso 4:** se crean las conexiones para las bases de datos fuente y destino

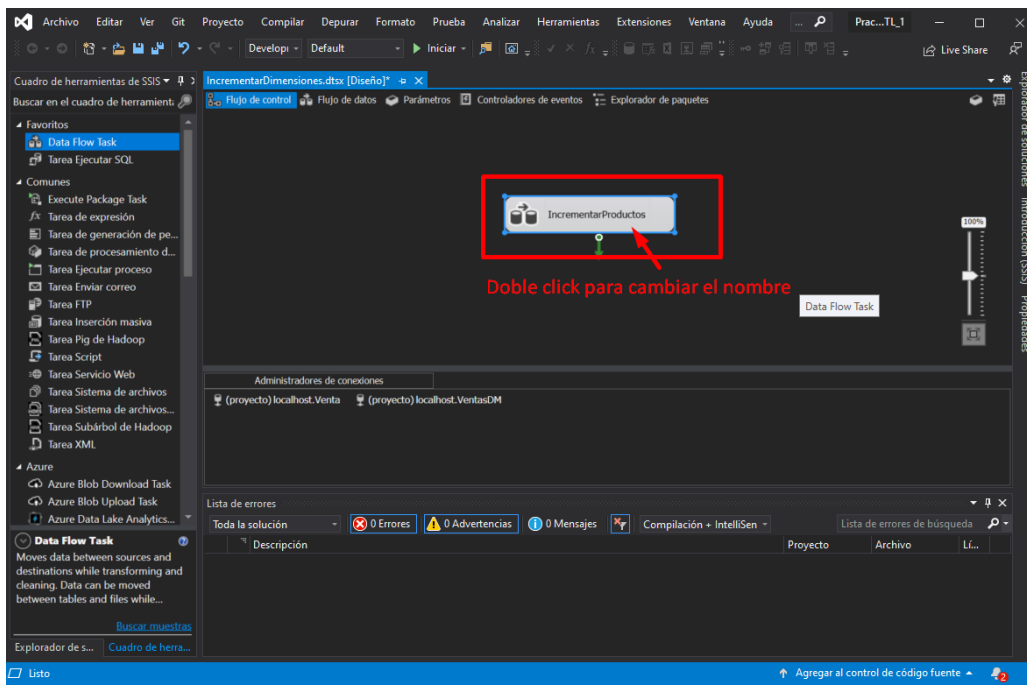
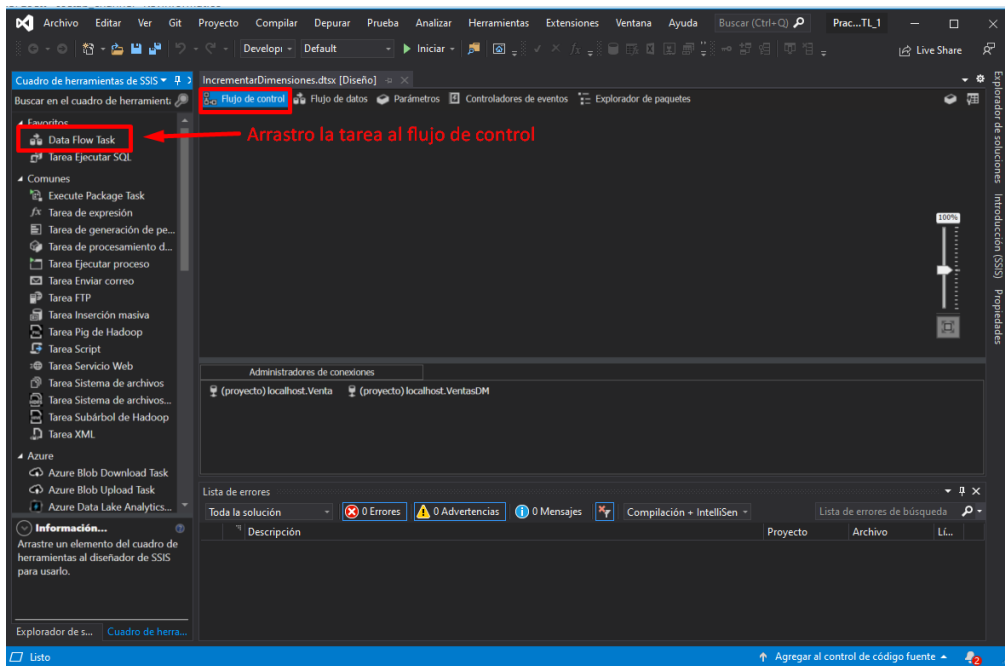






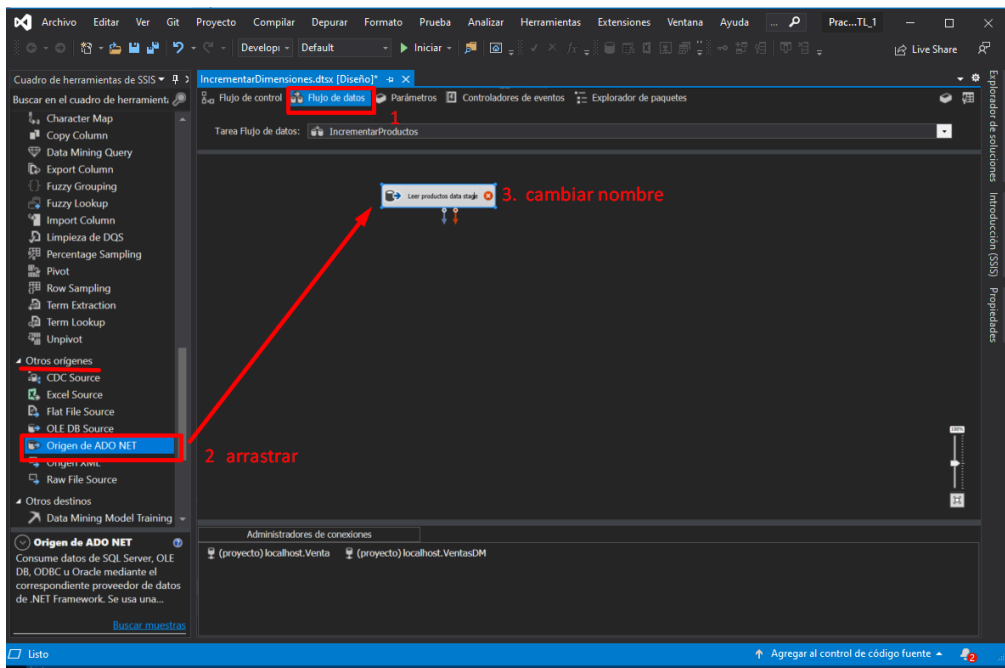


**Paso 6:** Inicio del proceso de desarrollo. A mano izquierda ubico el cuadro de herramientas SSIS y arrastro el control “DataFlow Task” al área de trabajo en la pestaña flujo de control

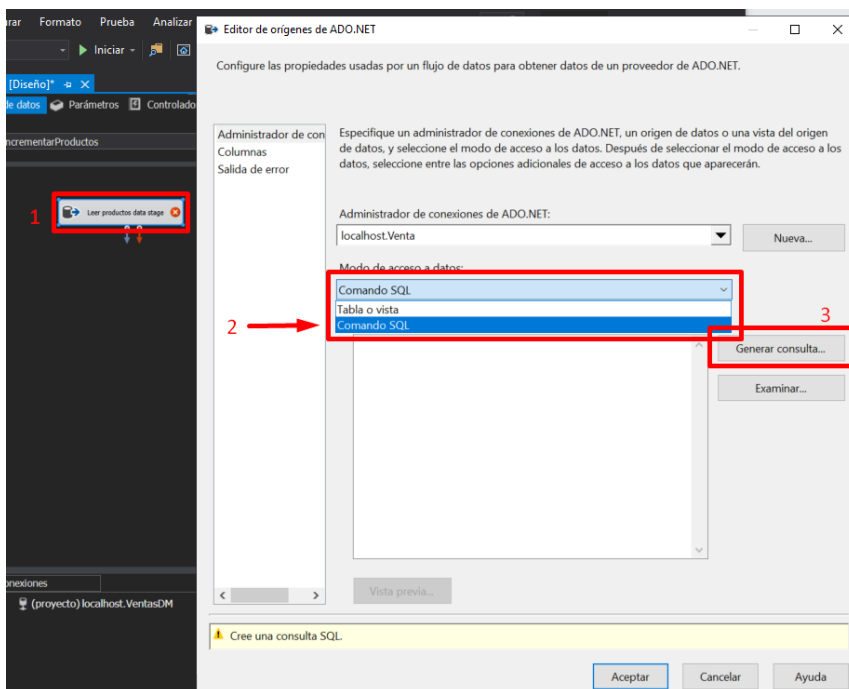


Luego de cambiar el nombre se da doble click al control para ingresar al flujo de datos

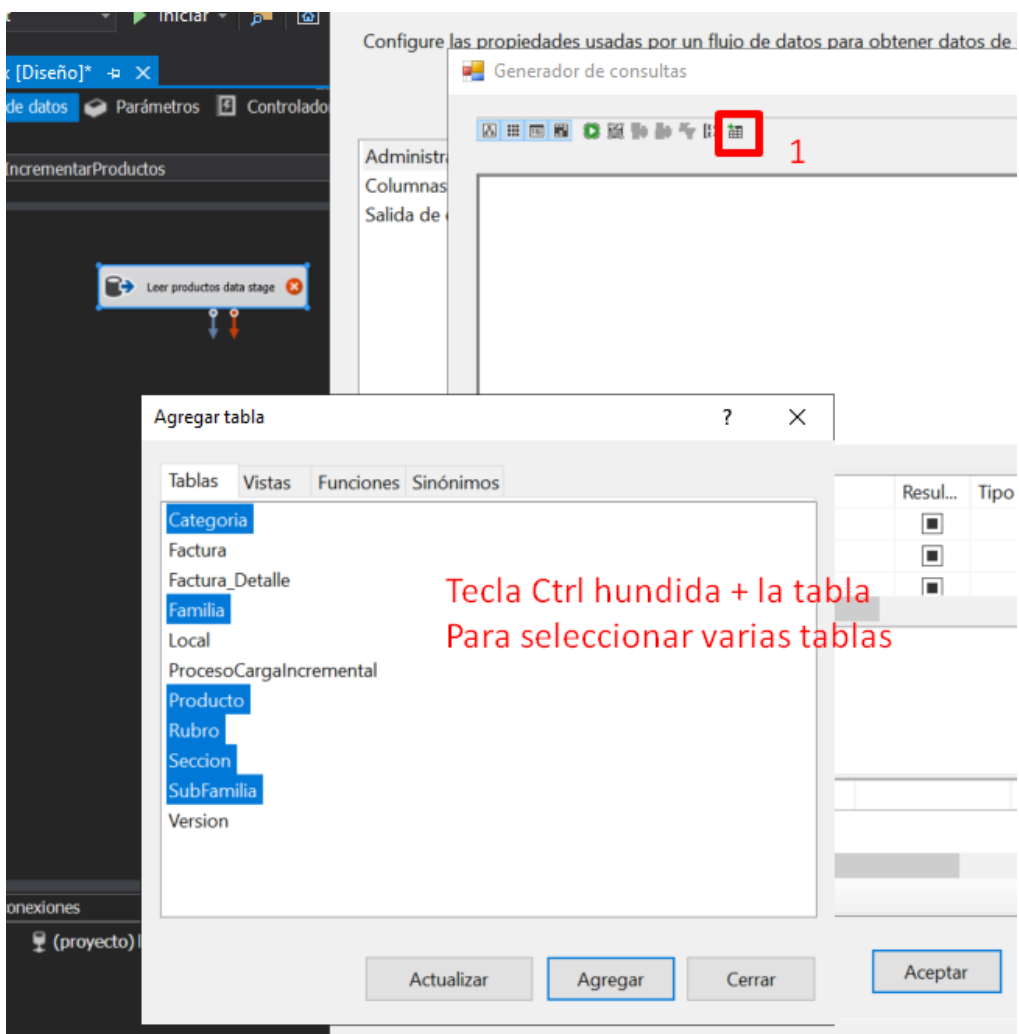
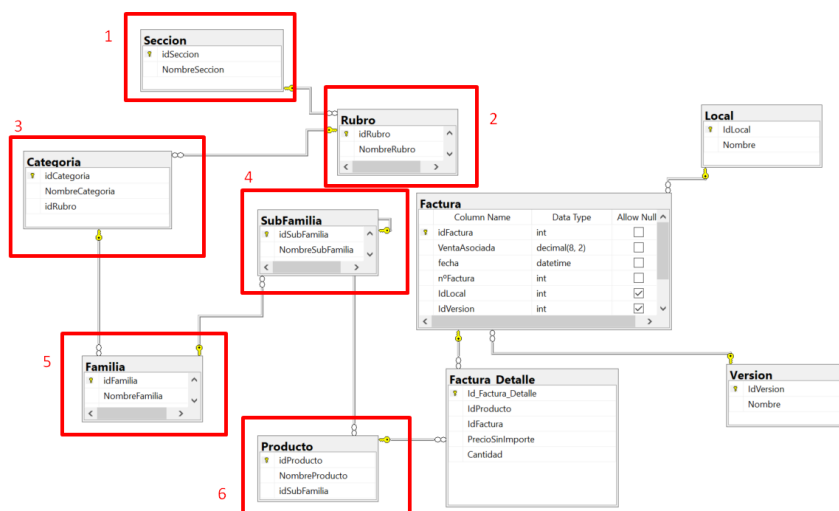
La primera tarea en la pestaña flujo de datos es buscar en el cuadro de herramientas SSIS los orígenes de datos, en este caso son los **ADO .NET**



Al dar doble click sobre el control se abre una nueva ventana con la configuración del origen donde se establece la conexión a la fuente y el método de acceso a los datos, para este caso lo haremos a través de un comando SQL y seleccionamos el botón **“Generar consulta”**

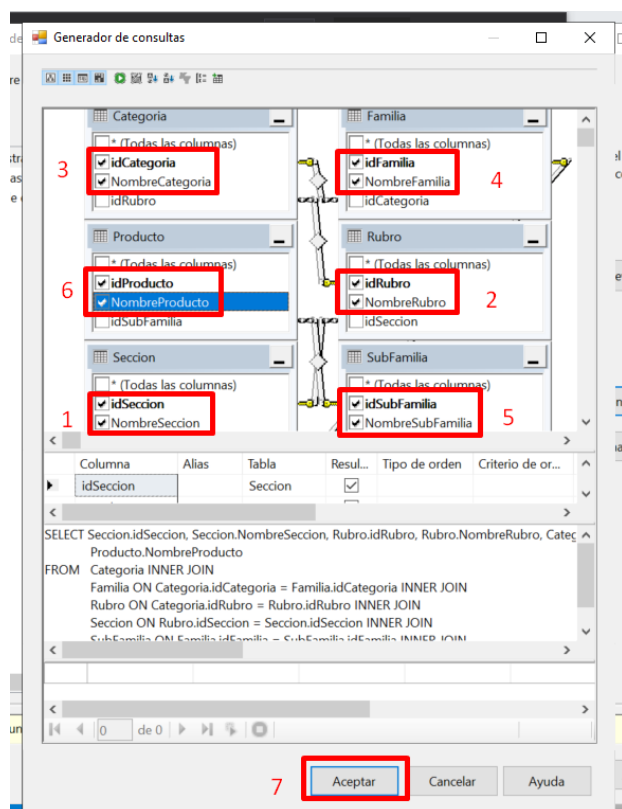


Se tiene en cuenta las tablas Producto y sus derivadas

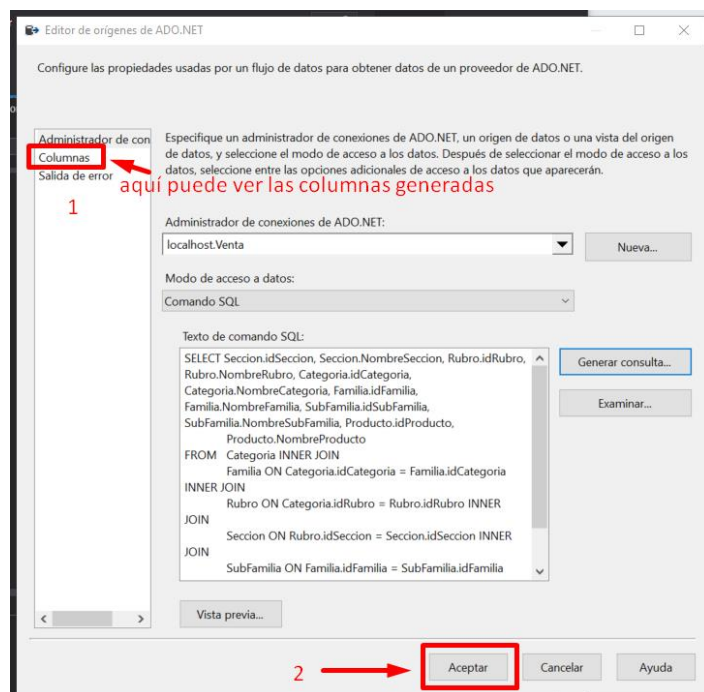


Con la tecla control sostenida selecciono las tablas que deseo cargar, opimo agregar y luego cerrar

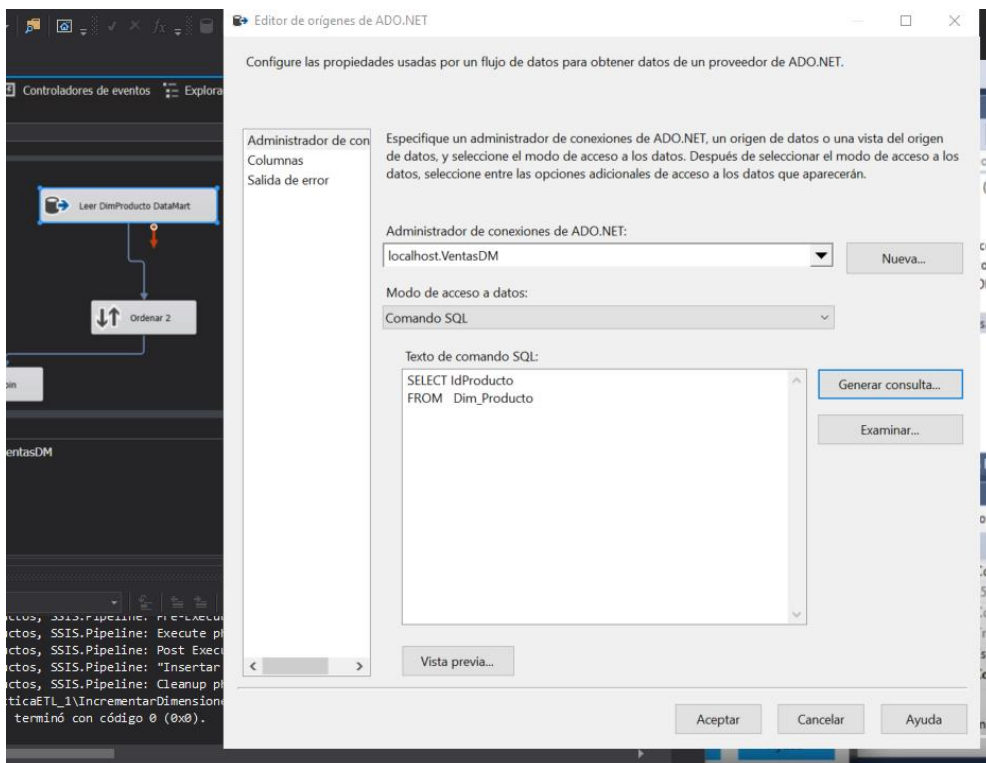
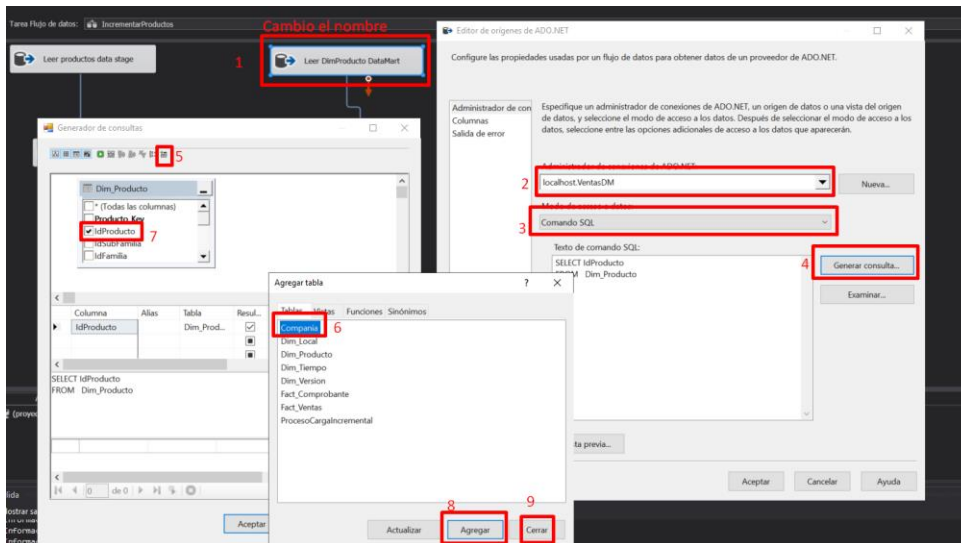
Seleciono los campos en el orden marcados a continuación



Se ha generado una en el campo de comando SQL, se accede a las columnas y se verifica que todo este correcto y se oprime aceptar



Se arrastra un nuevo origen al cual llamaremos “Leer Dimproducto DataMart”, se elige la conexión ventasDM y se carga la tabla DimProducto seleccionando el campo idProducto que nos ayudara en el proceso de carga incremental



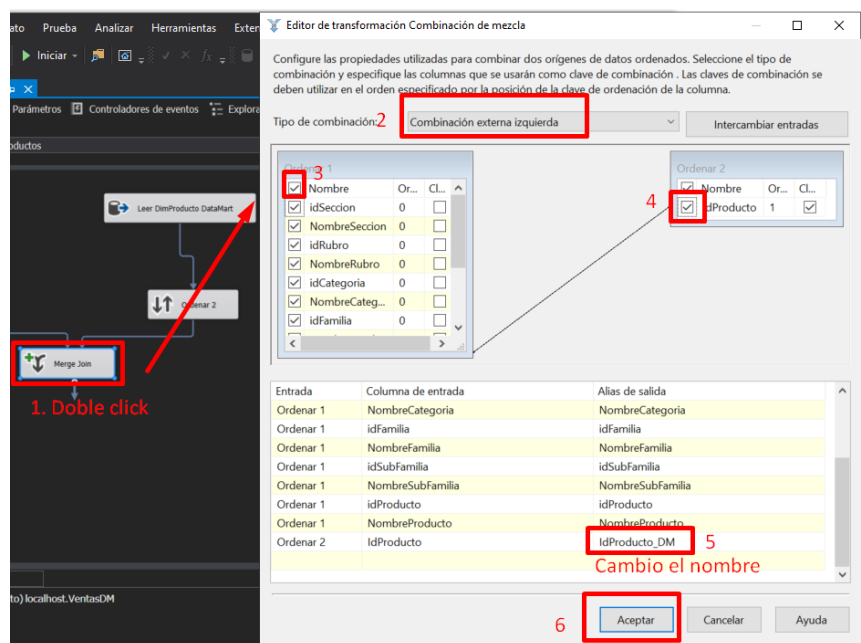
Para hacer la comparación entre las fuentes de datos es necesario ordenar el campo idProducto en ambas consultas

The screenshot shows the SSIS Designer interface with the 'IncrementarDimensiones.dtsx [Diseño]' package. The 'Flujo de datos' (Data Flow) tab is active, showing a data flow task 'IncrementarProductos'. Two 'Ordenar' (Sort) tasks are present: 'Ordenar 1' and 'Ordenar 2'. Red arrows and text annotations indicate the steps: '1. Arrastro' (Drag) points to the 'Sort' task in the toolbox; '2. Uno con la línea azul los componentes' (Connect the components with the blue line) points to the connection between the two sort tasks; '3. Selecciono la variable a ordenar' (I select the variable to sort) points to the 'idProducto' column in the 'Editor de transformación Ordenar' (Sort Transformation Editor) dialog. The dialog shows 'idProducto' selected in the 'Columnas de entrada disponibles' (Available input columns) list. The 'Tipo de orden' (Sort type) is set to 'ascendente' (ascending) and the 'Criterio de orden' (Sort criterion) is set to '1'. The 'Aceptar' (Accept) button is highlighted.

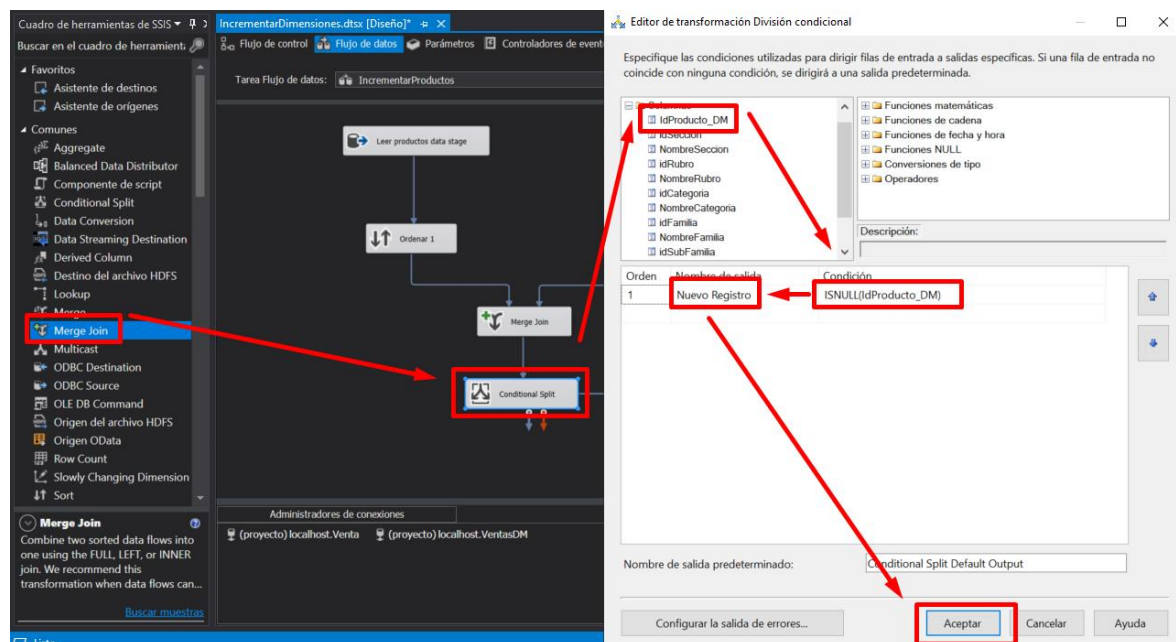
Se hace el merge que en este caso es un left join para saber qué códigos de producto se encuentran cargados en ambas tablas

The screenshot shows the SSIS Designer interface with the 'IncrementarDimensiones.dtsx [Diseño]' package. The 'Flujo de datos' (Data Flow) tab is active, showing a data flow task 'IncrementarProductos'. Two 'Ordenar' (Sort) tasks are present: 'Ordenar 1' and 'Ordenar 2'. A 'Merge Join' task is added to the flow. Red arrows and text annotations indicate the steps: '1. Arrastro' (Drag) points to the 'Merge Join' task in the toolbox; '2. Uno con la línea azul los componentes' (Connect the components with the blue line) points to the connection between the two sort tasks and the merge join task; '3. Selecciono la variable a ordenar' (I select the variable to sort) points to the 'idProducto' column in the 'Selección de entrada y salida' (Select input and output) dialog. The dialog shows 'Sort Output' selected in the 'Salida' (Output) dropdown and 'Merge Join Left Input' selected in the 'Entrada' (Input) dropdown. The 'Aceptar' (Accept) button is highlighted.

Se verifica la relación y se cambia el alias de salida del **idProducto** en la consulta ordenar 2

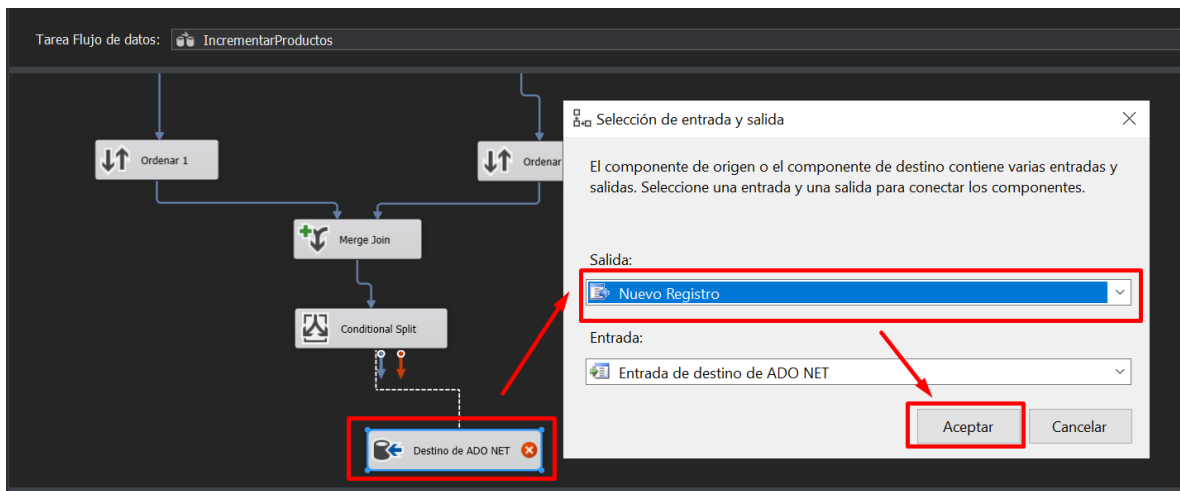


Se arrastra la herramienta **Conditional Split** que se conecta al **Merge join** con el fin de establecer una ruta a los registros nuevos para que sean cargados al DW, al dar doble click sobre este control se configura la condición **ISNULL(idProducto\_DM)** que significa que si se cumple esta condición los registros tomarán el camino de **Nuevo registro** que harán insert en la **DimProducto**

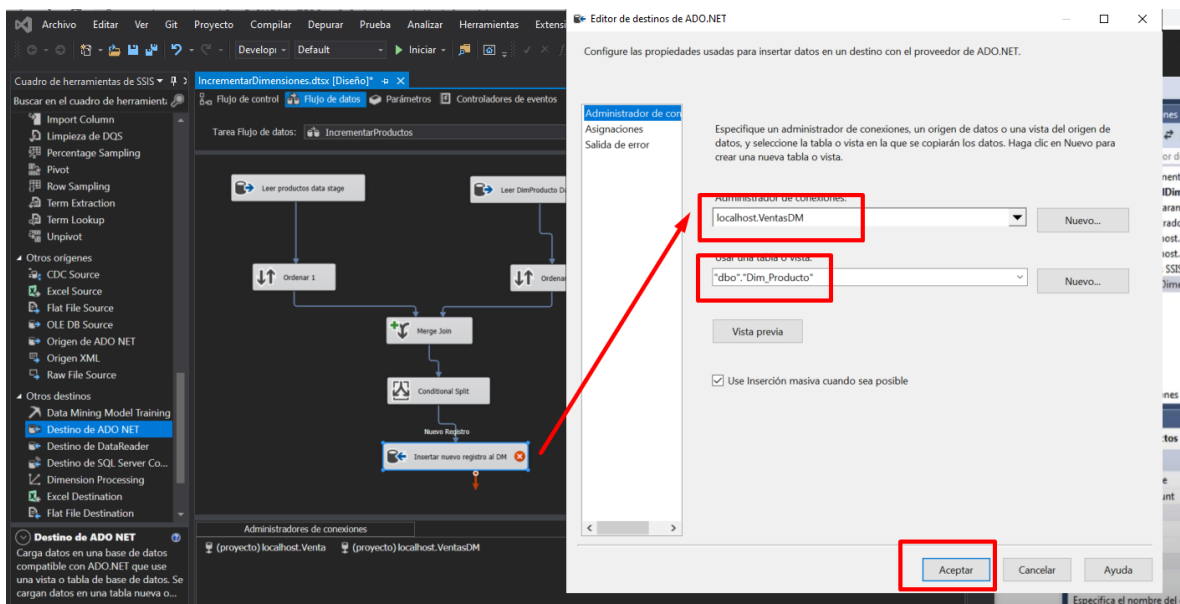




Finalmente, arrastramos desde las herramientas un **Destino ADO .NET** y al conectarlo con el **Conditional Split** nos pide que seleccionemos una salida, en este caso seleccionamos **Nuevo Registro**

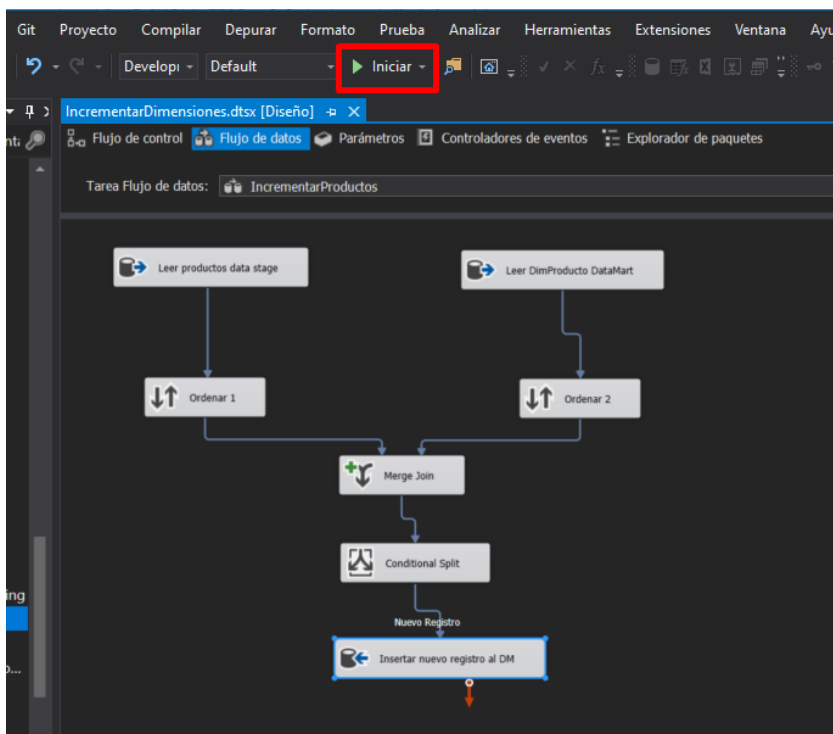


Configuramos la fuente de destino y la tabla

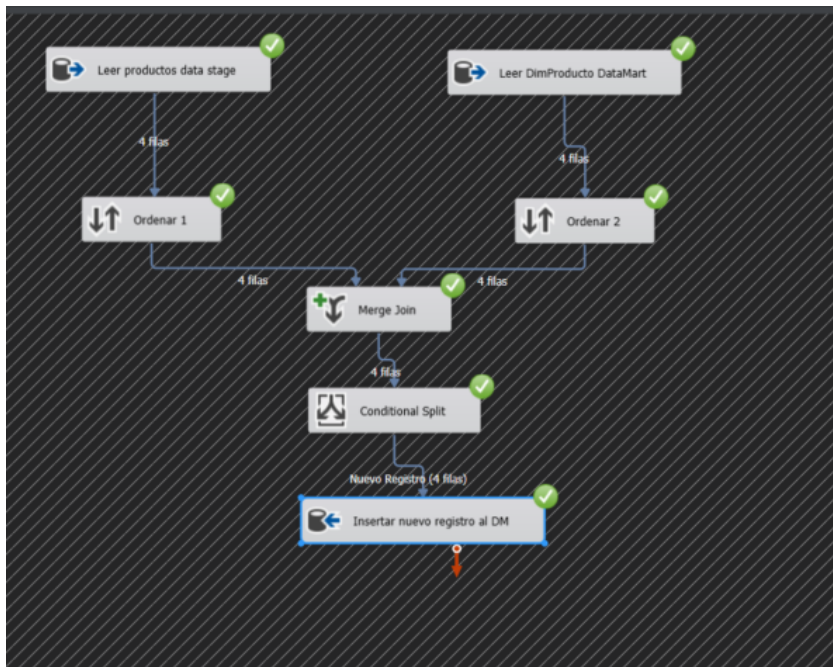


¡¡Hemos terminado!!

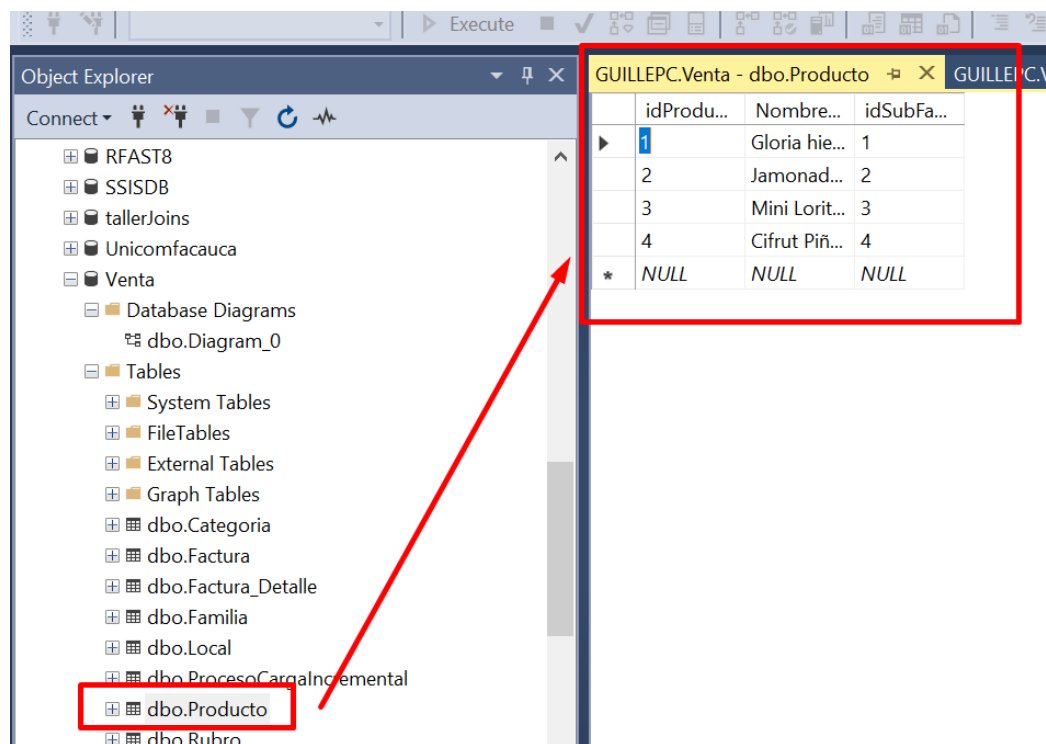
## Paso 7: Correr el ETL



La primera carga nos arroja 4 registros cargados



## Registros en la tabla Producto



GUILLEPC.Venta - dbo.Producto

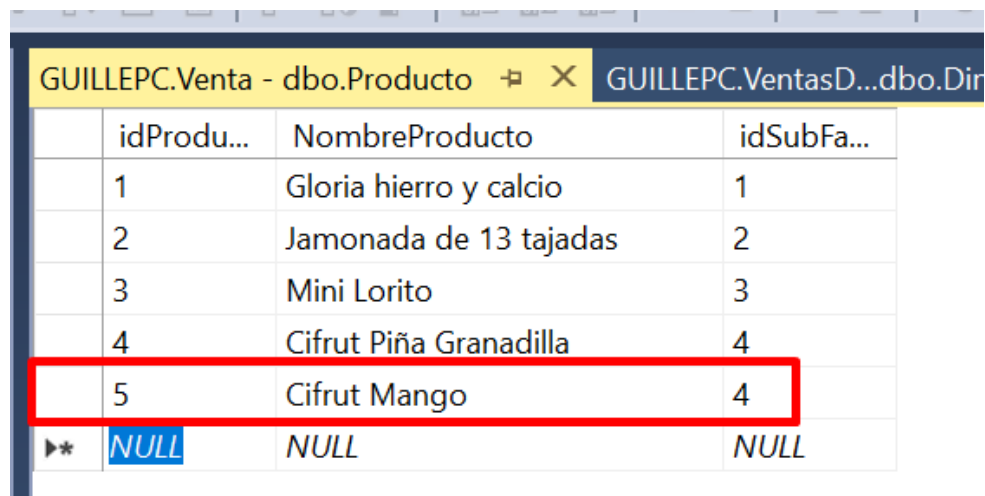
	idProdu...	Nombre...	idSubFa...
▶ 1		Gloria hie...	1
2		Jamonad...	2
3		Mini Lorit...	3
4		Cifrut Piñ...	4
*	NULL	NULL	NULL

Object Explorer

Connect

- RFAST8
- SSISDB
- tallerJoins
- Unicomfauca
- Venta
  - Database Diagrams
    - dbo.Diagram\_0
  - Tables
    - System Tables
    - FileTables
    - External Tables
    - Graph Tables
    - dbo.Categoria
    - dbo.Factura
    - dbo.Factura\_Detalle
    - dbo.Familia
    - dbo.Local
    - dbo.ProcesoCargaIncremental
    - dbo.Producto**
    - dbo.Rubro

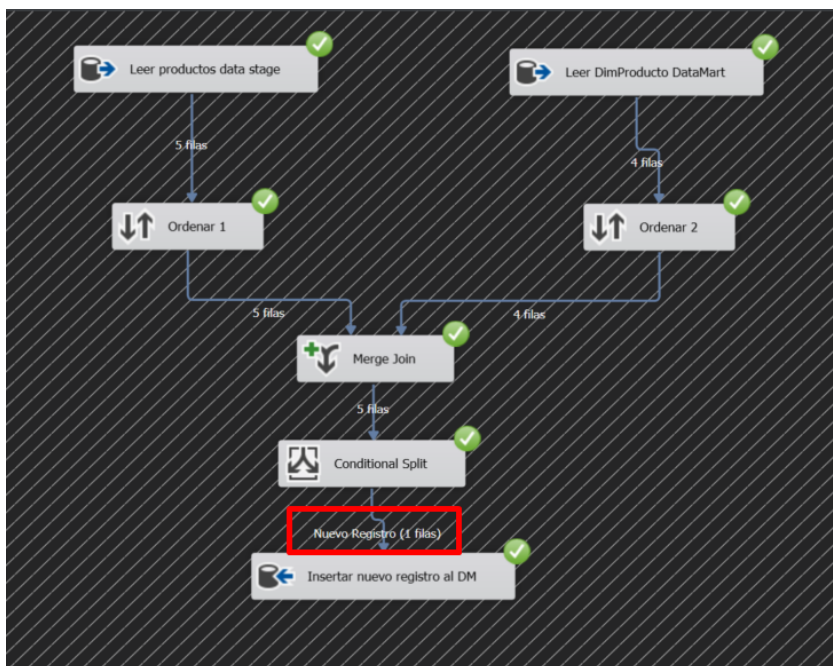
Se inserta un nuevo registro en la tabla producto de la base de datos transaccional



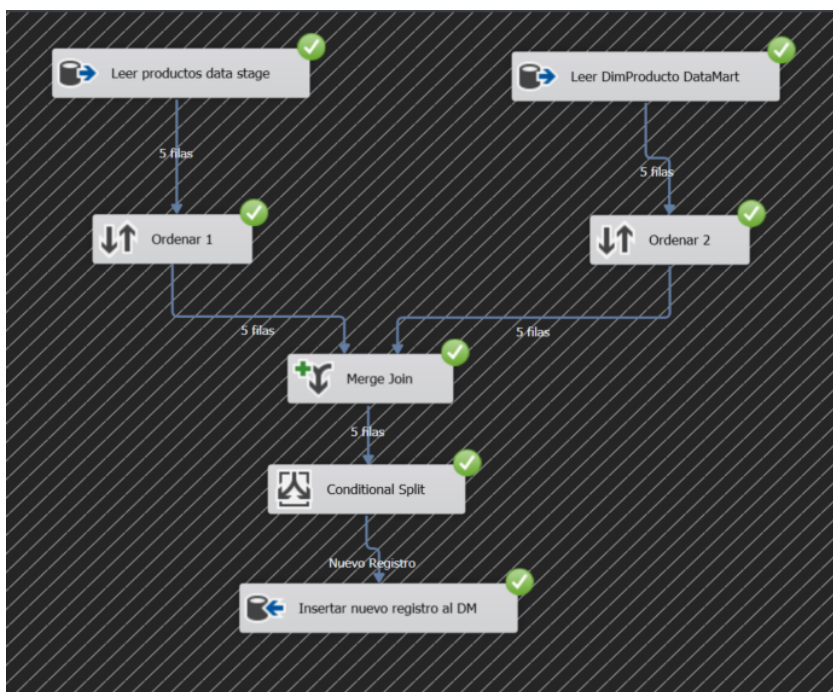
GUILLEPC.Venta - dbo.Producto

	idProdu...	NombreProducto	idSubFa...
1		Gloria hierro y calcio	1
2		Jamonada de 13 tajadas	2
3		Mini Lorito	3
4		Cifrut Piña Granadilla	4
5		Cifrut Mango	4
▶*	NULL	NULL	NULL

El ETL carga sólo el nuevo registro



Se corre nuevamente el ETL y no hay registros nuevos por cargar



Con este ejercicio hemos solucionado un problema sencillo en la carga incremental de las bases de datos, pero, hay otros escenarios como el de las **dimensiones tipo 2** o dimensiones que cambian lentamente (Slowly changing dimension) a las cuales hay que hacer un tratamiento diferente utilizando la sentencia **update** e **insert** según sea el caso. Además, se recomienda establecer un proceso que cargue una tabla de auditoría.