

UNIVERSIDAD NACIONAL DE CHIMBORAZO
CARRERA DE INGENIERÍA EN TECNOLOGÍAS DE LA INFORMACIÓN
TALLER PRÁCTICO ETL - PRIMERA PARTE



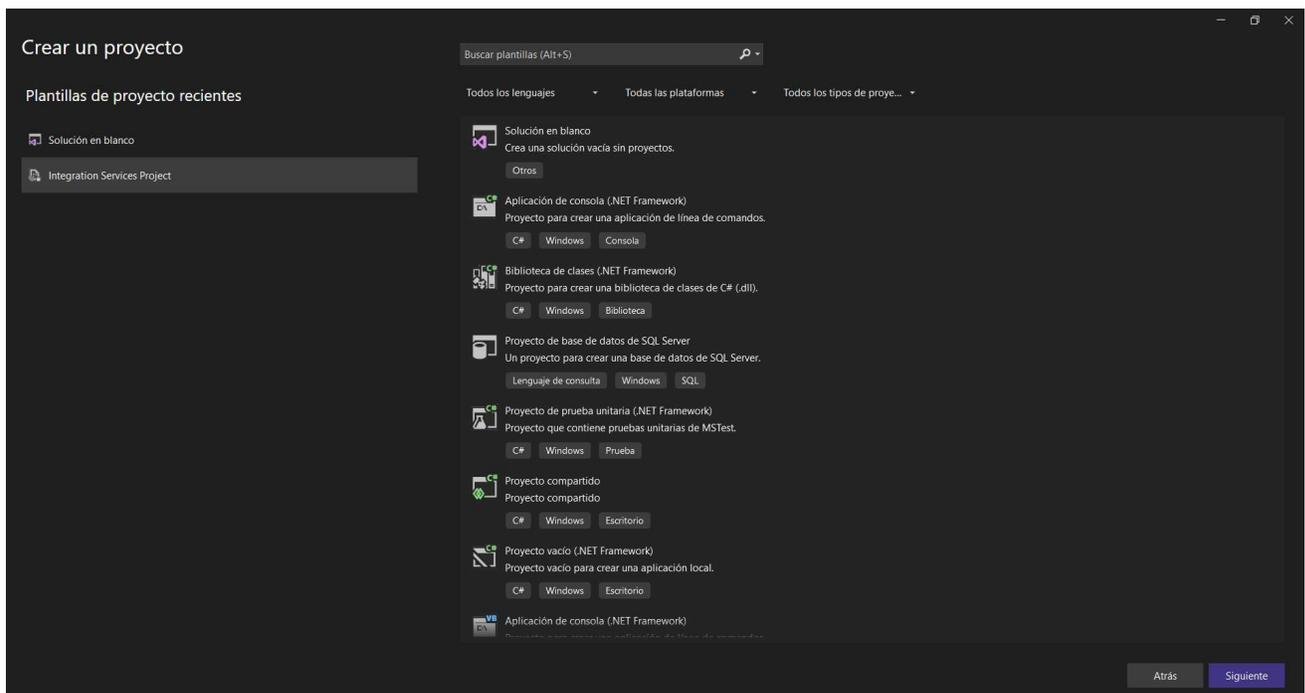
Profesora: Ing. Elba Boderó Poveda, Ph.D.

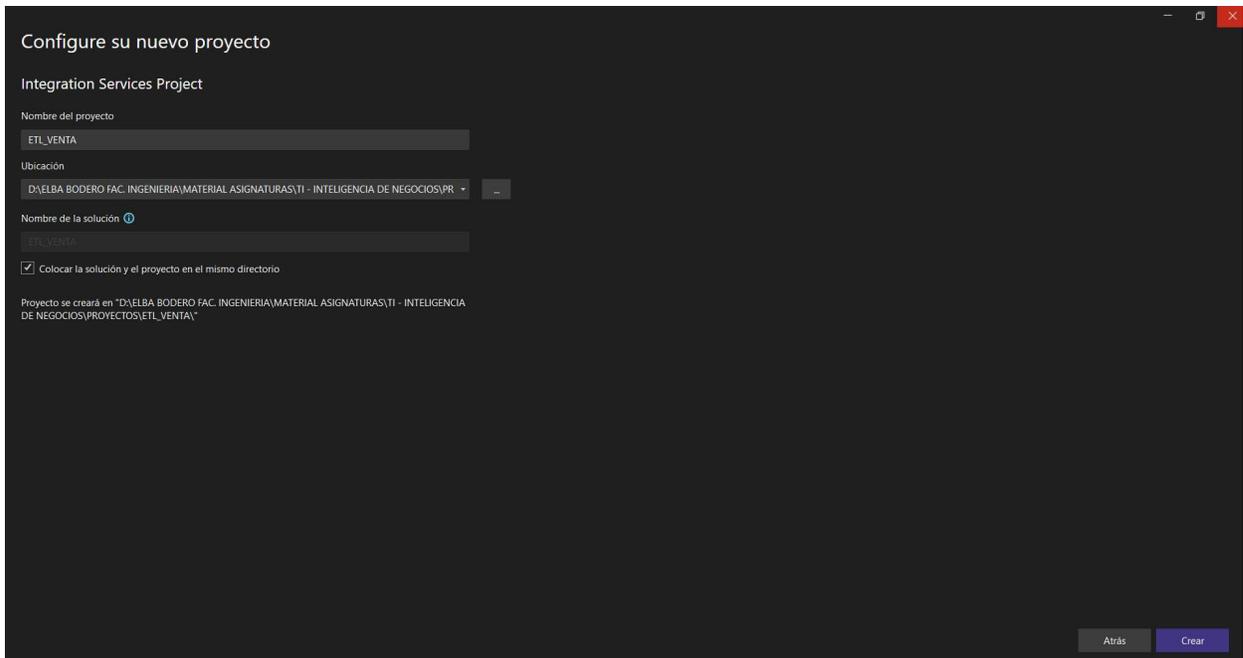
PRERREQUISITOS:

- SQL Server
- SQL server Data Tools en Visual Studio.
- Base de datos Fuente.
- Base de datos para Staging creada.
- Base de datos del Data Warehouse creada:
 - Utilizar el diseño del **MODELO DIMENSIONAL** de la DB.
 - Ejecutar el store procedure carga dimensión tiempo.

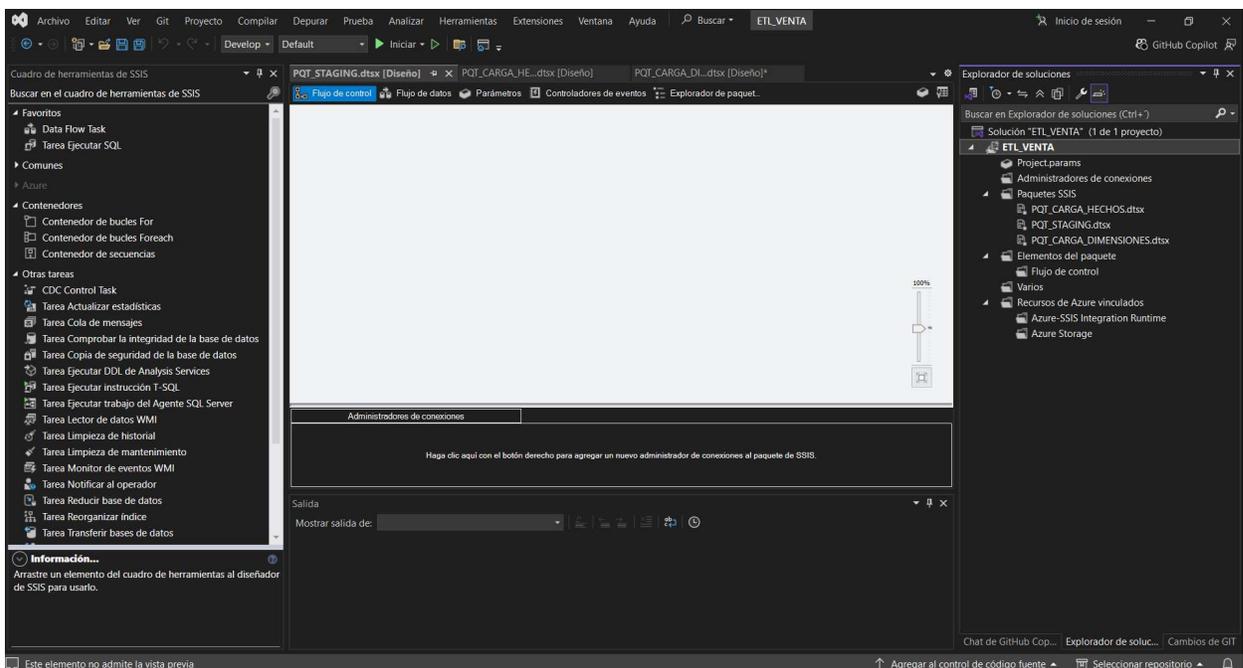
INSTRUCCIONES:

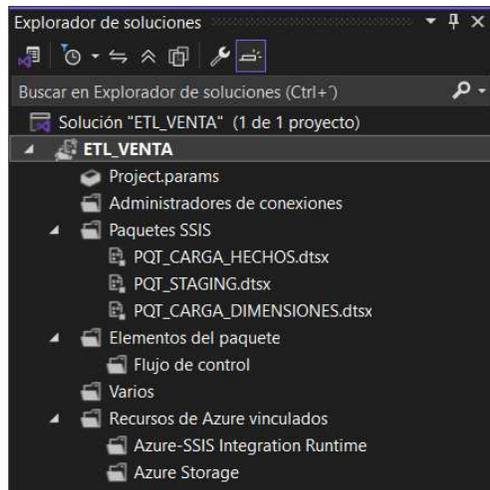
1. Creamos un Proyecto de Integration Services.



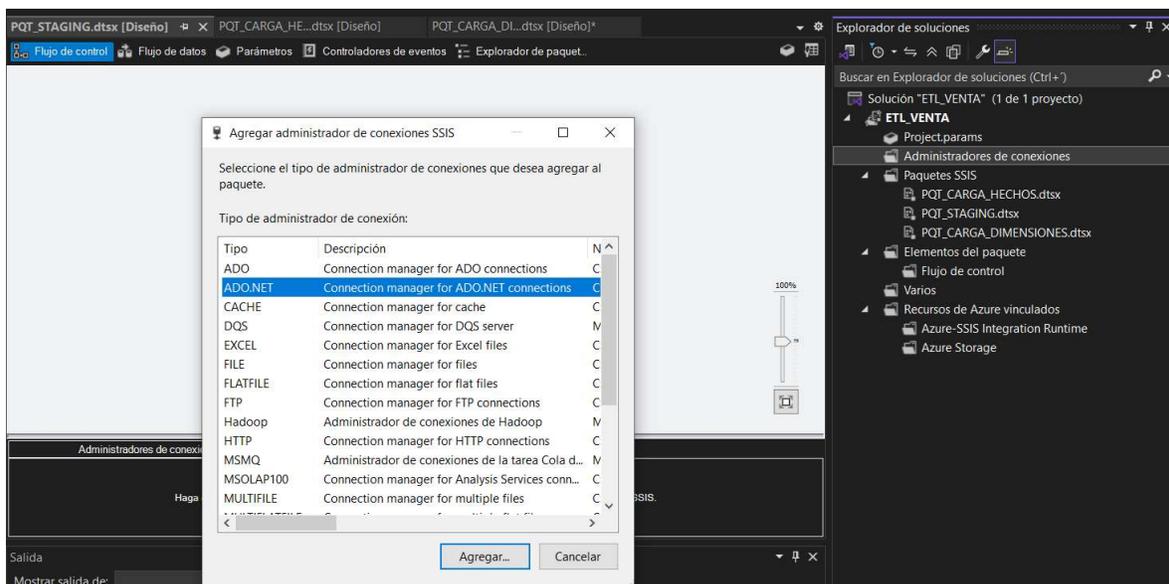


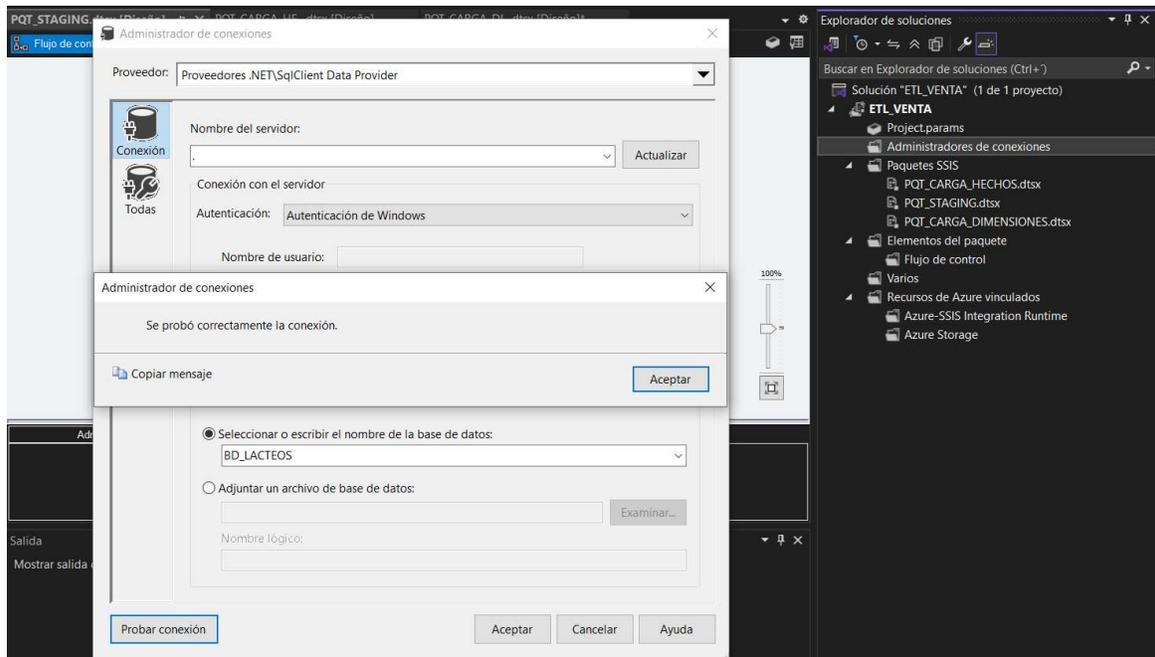
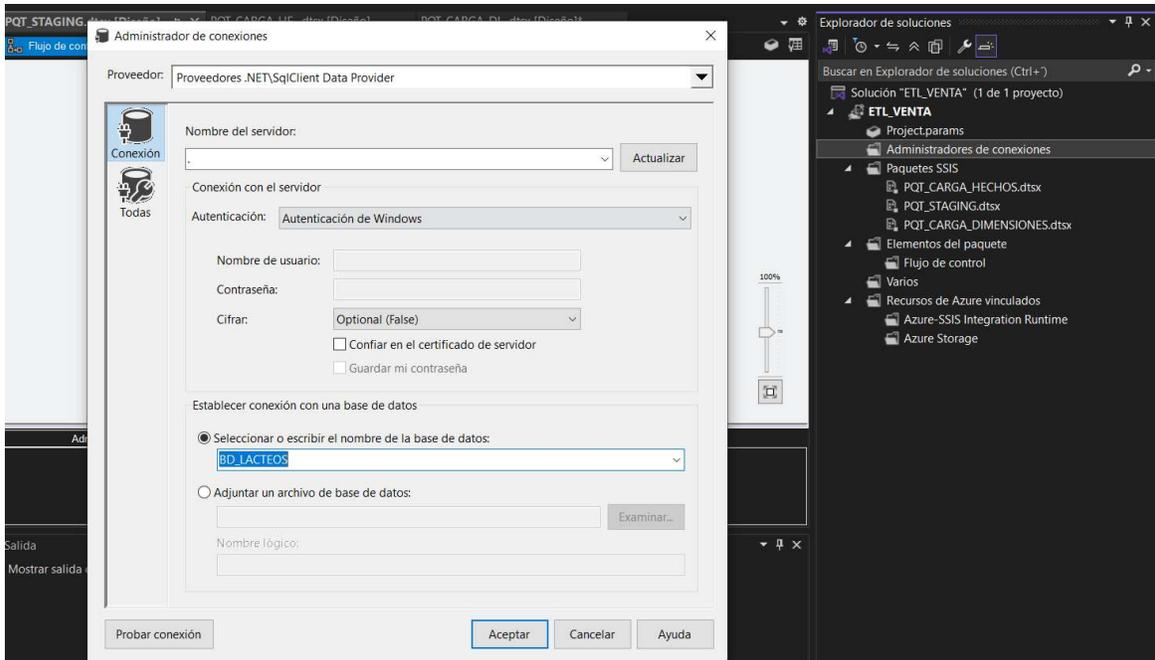
2. Creamos 3 paquetes para carga de: Staging, Dimensiones, Hechos. Lo hacemos dando clic derecho en la carpeta "Paquetes SSIS", que se encuentra en el Explorador de Soluciones, "Nuevo paquete de SSIS".



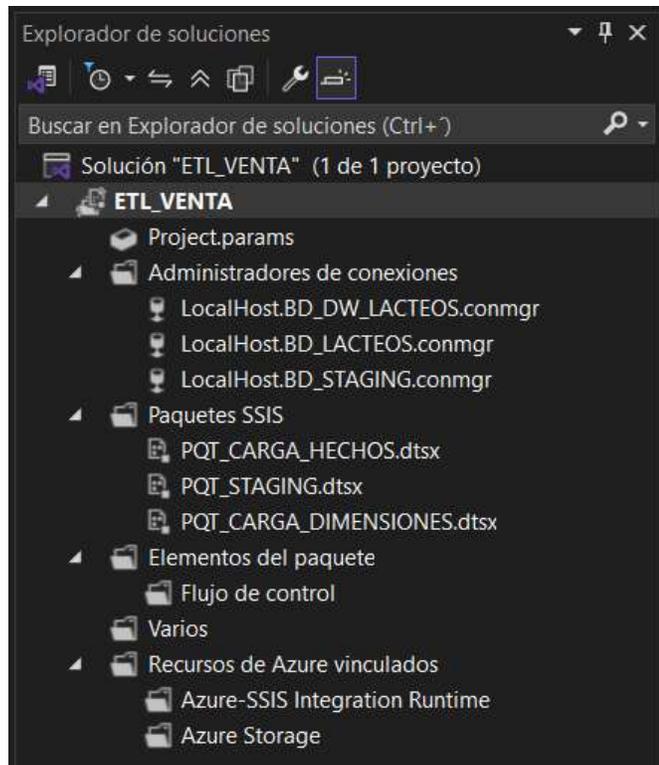


3. En la carpeta de "Administradores de conexiones", vamos a crear un "Nuevo administrador de conexiones" y elegimos que sea de Ado.net debido a que la fuente de datos para este ejemplo se encuentra realizada en SQL Server, caso contrario podremos elegir una diferente conexión a la fuente de datos necesaria para la implementación de nuestro proyecto.





4. Realizamos el mismo proceso para la base de datos "BD_Staging" y "BD_DW_LACTEOS".



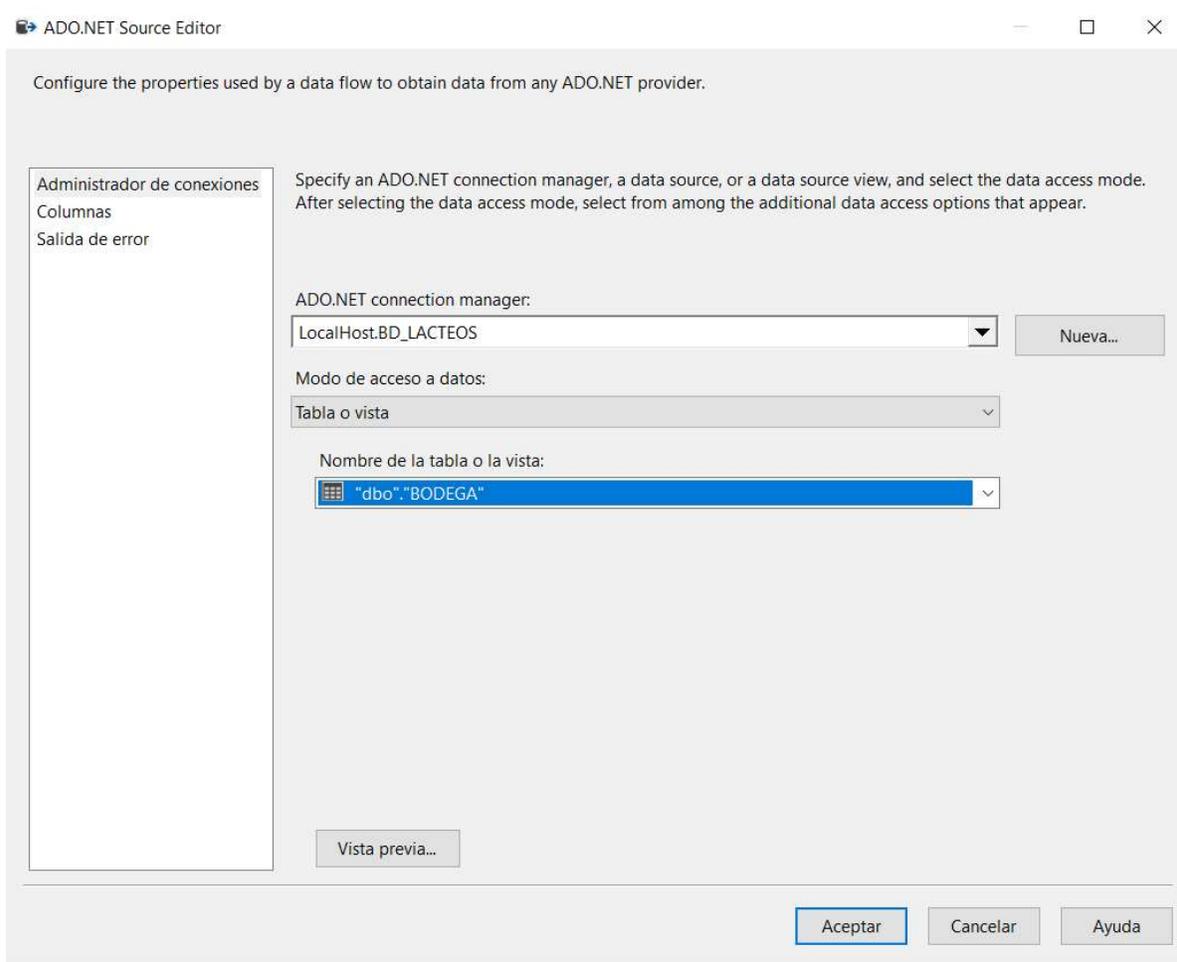
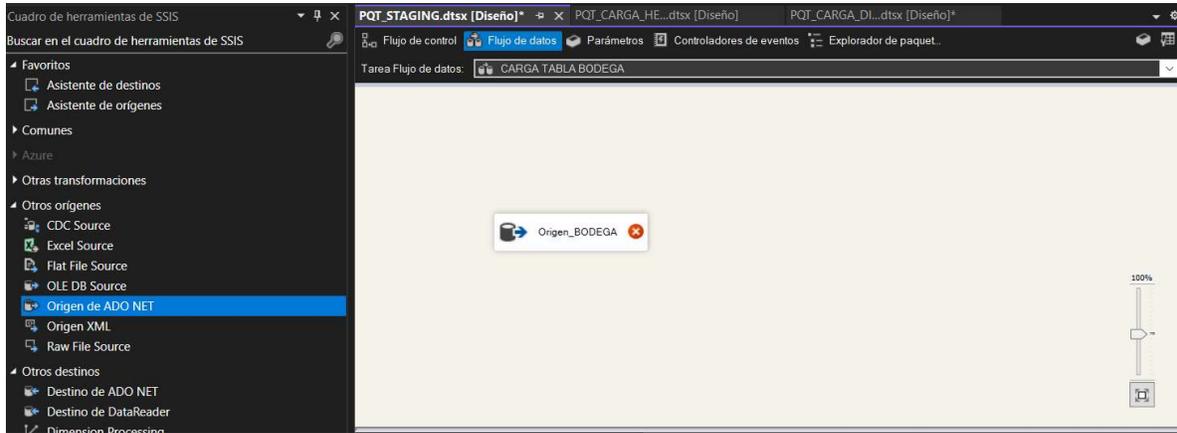
PARTE 1: CARGA STAGING

5. Añada una herramienta "Tarea Flujo de datos", configúrela dando doble clic sobre la misma.

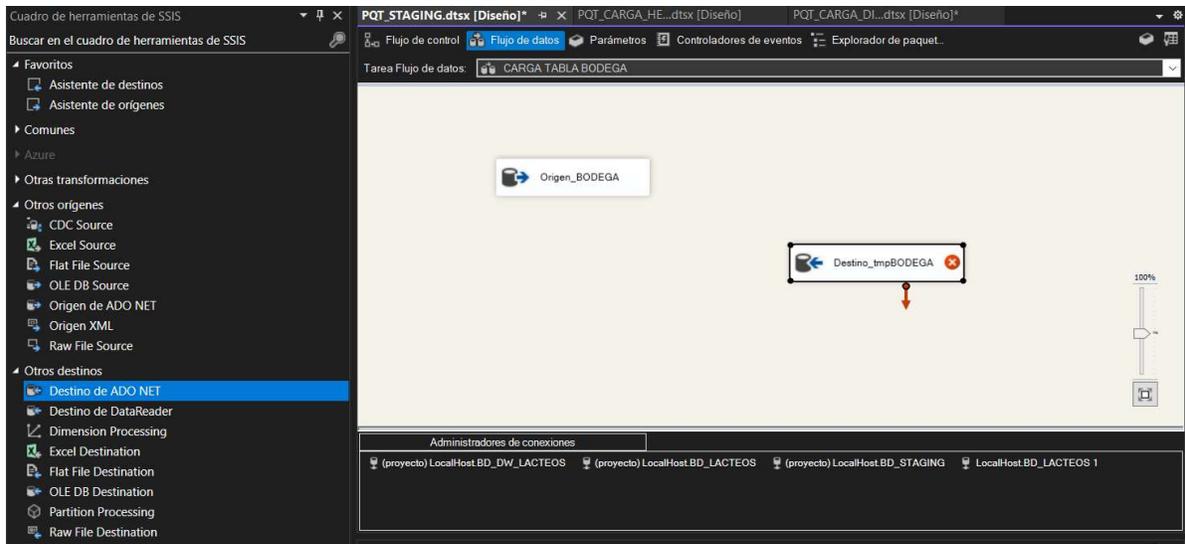


En **SQL Server Integration Services (SSIS)**, una **tarea de flujo de datos** (Data Flow Task) es uno de los componentes más importantes del proceso ETL (Extracción, Transformación y Carga). Esta tarea permite **mover y transformar datos** desde una o más fuentes hacia uno o más destinos de manera eficiente y controlada.

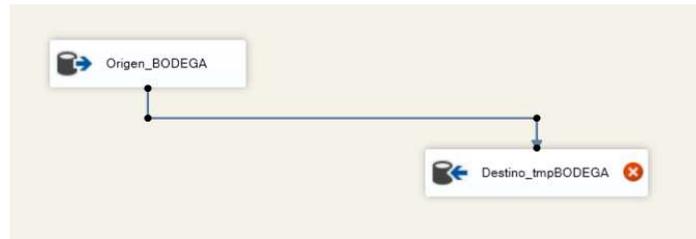
6. Realice doble clic en el flujo de datos que acaba de crear (CARGA TABLA BODEGA) y añada un “Origen de ADO NET” en el grupo de herramientas “Otros orígenes” y configúrelo con doble clic sobre la herramienta.



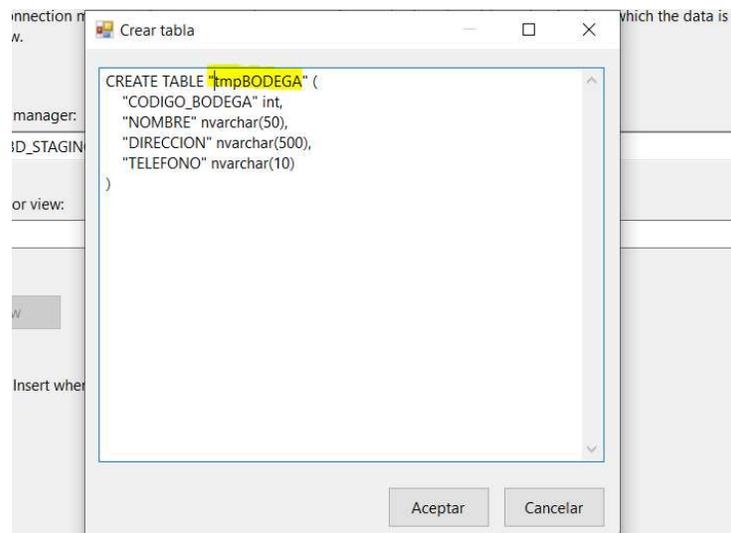
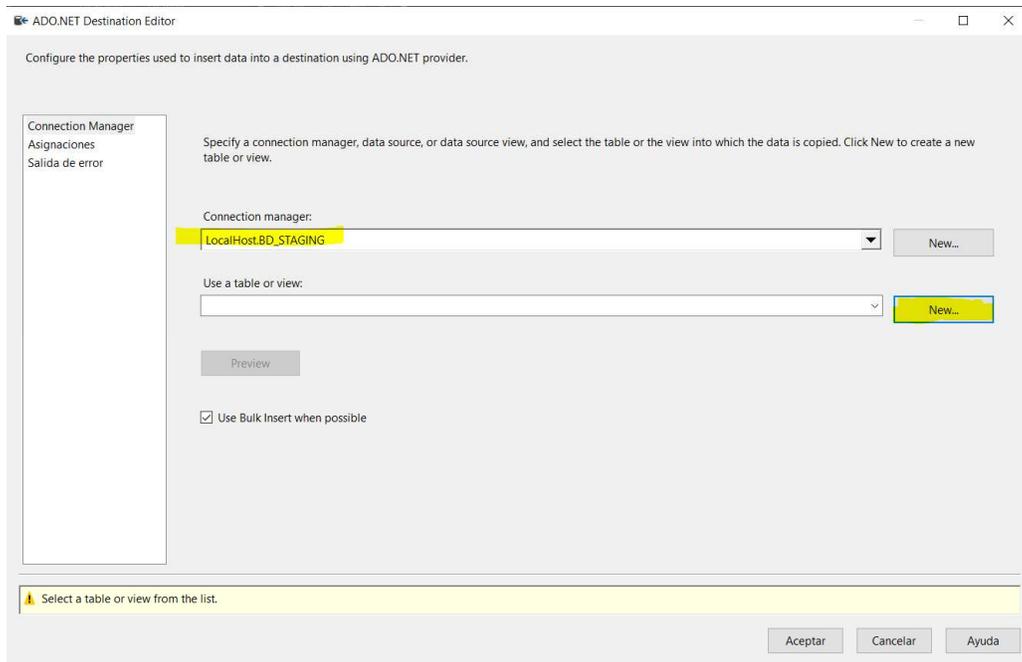
7. Añada un “Destino de ADO NET” en el grupo de herramientas “Otros destinos”.

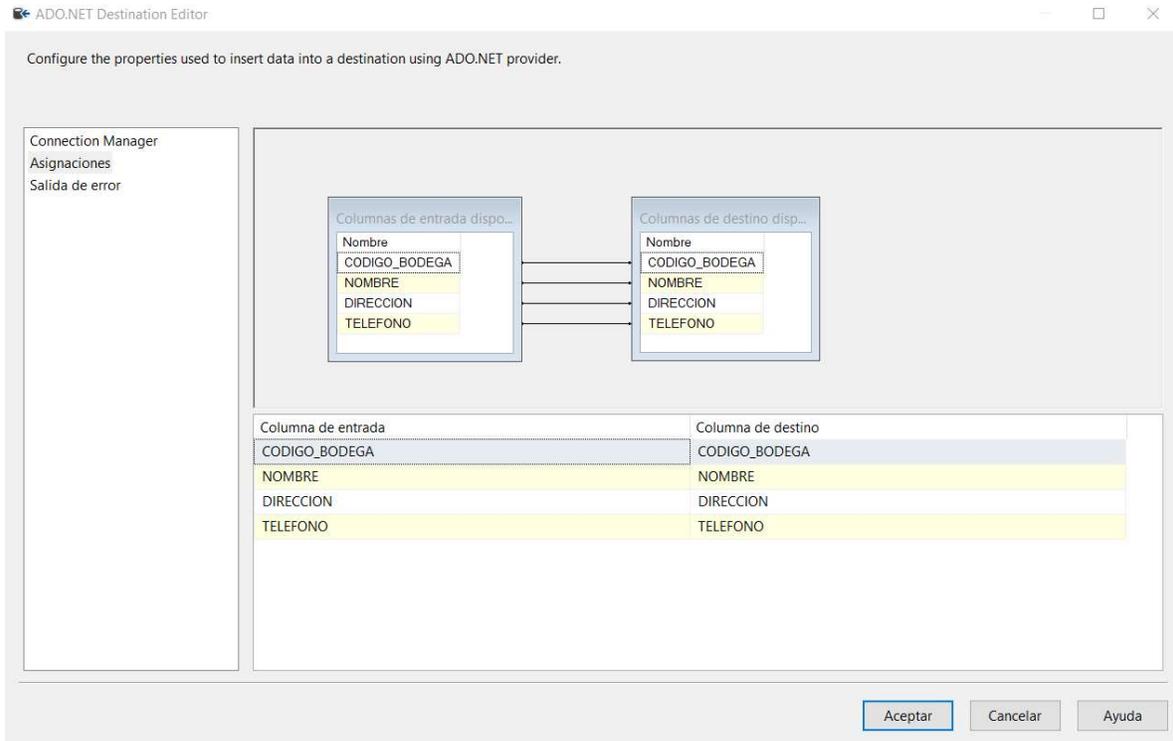


8. Enlace la herramienta origen con la de destino, arrastrando de la flecha color azul mostrada en el origen.

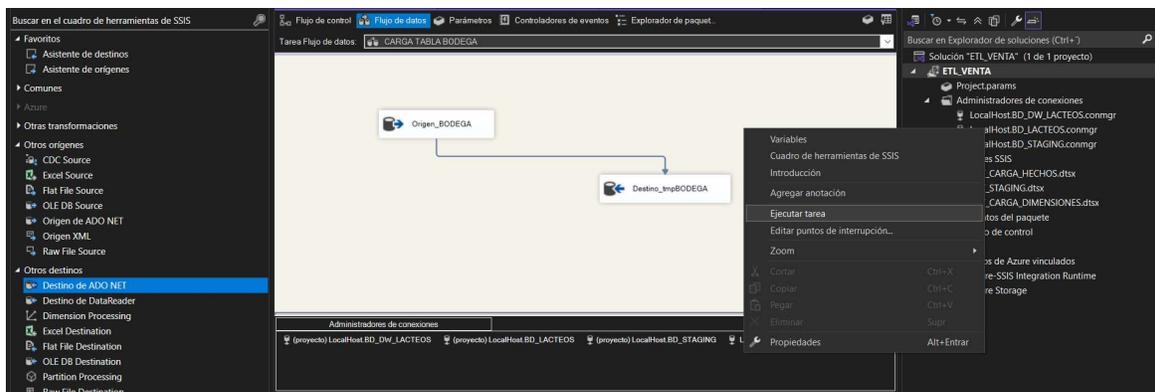


9. Configure la tabla de destino de la siguiente forma:





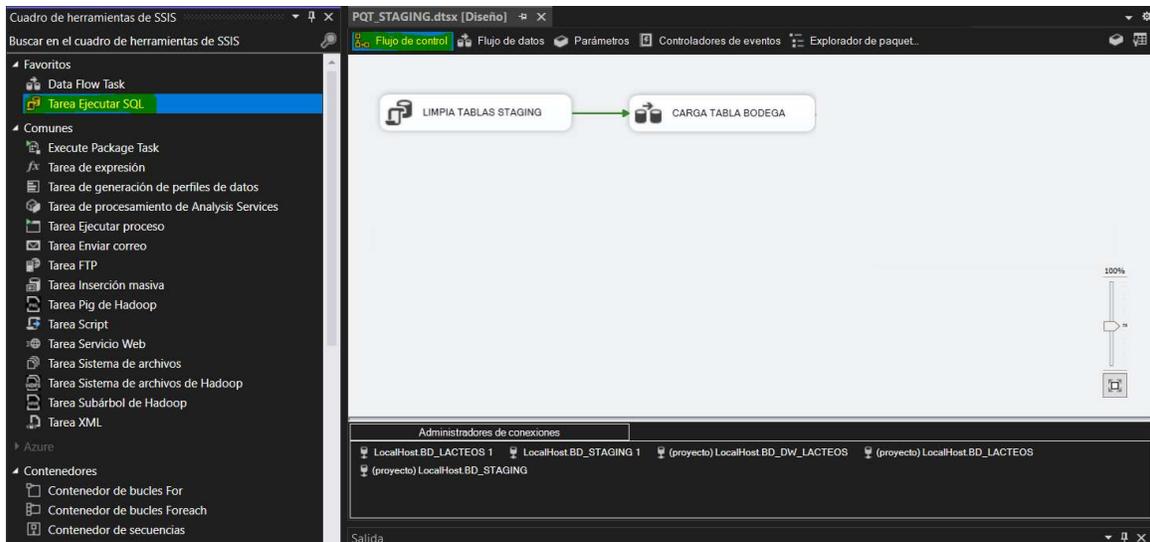
10. Pruebe el flujo de la siguiente forma:



11. Realice lo mismo para las otras tablas de la base de datos BD_LACTEOS.

12. Cree una "Tarea Ejecutar SQL" para limpiar los datos de las tablas de Staging.

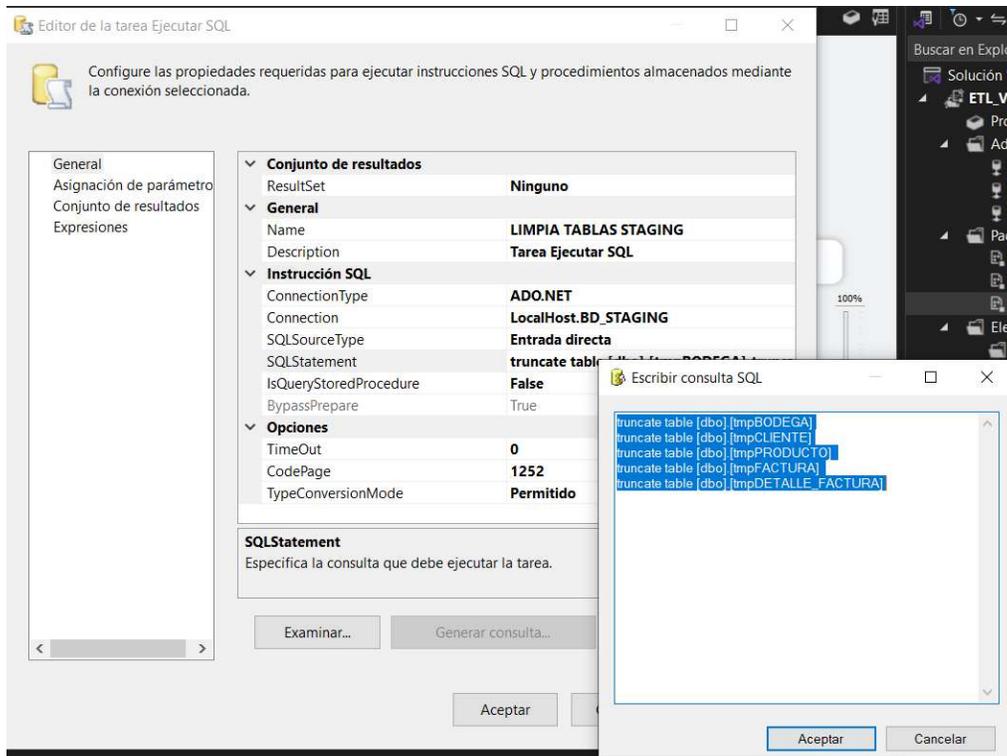
En **SQL Server Integration Services (SSIS)**, la tarea **Ejecutar SQL** (Execute SQL Task) es un componente del **control flow** que permite ejecutar sentencias SQL directamente dentro del paquete. Es una de las tareas más versátiles y utilizadas en cualquier proceso ETL.



En la opción “SQLStatement” de “LIMPIA TABLAS STAGING” colocar lo siguiente:

```
truncate table [dbo].[tmpBODEGA]  
truncate table [dbo].[tmpCLIENTE]  
truncate table [dbo].[tmpPRODUCTO]  
truncate table [dbo].[tmpFACTURA]  
truncate table [dbo].[tmpDETALLE_FACTURA]
```

Estas sentencias SQL ejecutan la instrucción TRUNCATE TABLE, la cual elimina todos los datos de cada tabla de forma irreversible, es más eficiente que DELETE, ya que no registra individualmente cada fila eliminada en el log de transacciones. Su uso está justificado cuando se desea limpiar completamente las tablas temporales de staging como parte de un proceso de preparación o recarga de datos en un flujo ETL (Extracción, Transformación y Carga) en soluciones de Inteligencia de Negocios. Estas tablas son generalmente utilizadas para almacenar datos intermedios que luego serán validados y cargados al Data Warehouse final.



13. Finalmente, nuestro paquete con sus flujos debe mostrarse de la siguiente forma:



14. Ahora puede ejecutar el paquete de Staging:

