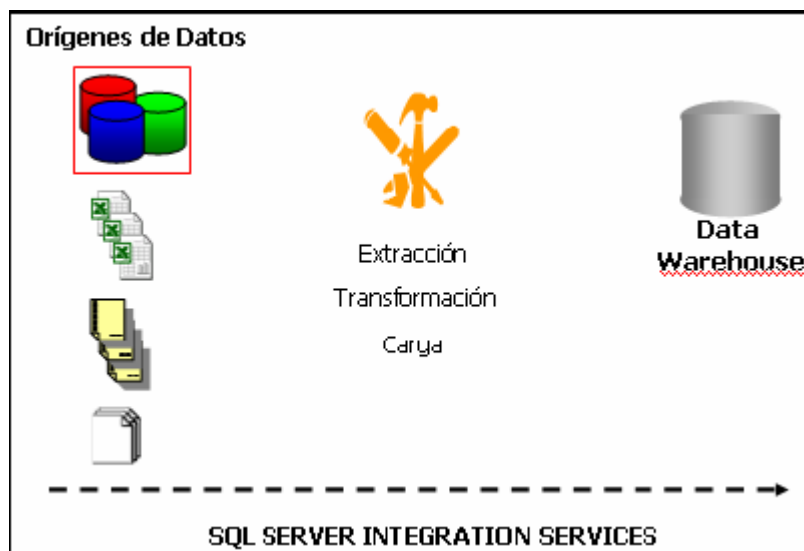


## CREAR UN PROYECTO DE ETL UTILIZANDO SQL SERVER INTEGRATION SERVICES

En un proceso datawarehouse, el sistema transaccional debe alimentar al sistema analítico. Sin embargo, es muy probable que el sistema transaccional se encuentre en fuentes de datos muy diversas (archivos.DBF, base de datos Oracle o Sybase, archivos.XLS, archivos.TXT, etc.). Pero la data en el servidor de análisis debe ser uniforme.

Para ello, utilizamos SQL Server Integration Services que es la herramienta de ETL proporcionada por Microsoft y que viene incluida en SQL Server 2017. SSIS nos permite lo siguiente:

- Generar, dentro de un Proyecto, uno o más paquetes de ETL que nos permita trasladar los datos de diferentes orígenes a un destino.
- Permitirá administrar mejor los procesos de ETL que puedan desarrollarse para poblar el datawarehouse.
- Extraer la información de diferentes fuentes de datos a través de diferentes conectores (proveedores .Net y proveedores OLE DB nativos).
- Transformar datos mediante tareas de transformación, búsqueda, integración, combinación, etc.
- Controlar la secuencia que los datos deben seguir mediante el control de eventos de éxito y fallo.
- Cargar la información a diferentes fuentes de datos.



### ELEMENTOS DE SSIS

SSIS está constituido por varios elementos que permiten construir paquetes de ETL complejos. Sus componentes son:

**A. Connection Managers.-** Identifica las conexiones hacia los diversos orígenes de datos

**B. Tasks.-** Asociada a una unidad de transformación, procesamiento de datos o transferencia de registros de un origen a un destino.

**C. Control Flow.-** Define la secuencia lógica de transferencia de información. Por ejemplo, a través de los flujos de control se permiten controlar la secuencia de transformación, limpieza, búsqueda, etc del dato.

**D. Container.-** Permite agrupar diferentes tareas de transformación para facilitar el entendimiento o la secuencia de las tareasl.

**E. Package.-** Todas las tareas contenidas dentro de un paquete de SSIS

## CREACIÓN DE PAQUETES INTEGRATION SERVICES

La creación de un paquete de SQL Server Integration Services se realiza utilizando el entorno de desarrollo “SQL Server Data Tools”. Ahí se crearán los paquetes que nos permitirá trasladar la información de la base de datos Norhwind a la base de datos Northwin\_Stage y de esta a la base de datos Norhwind\_Mart.

Es necesario que definamos y aclaremos algunos puntos antes de iniciar nuestra labor:

Definir el problema a solucionar

- ¿Es un problema de migración de datos?
- ¿Estoy en un proceso de Data Warehousing?
- ¿Mi proceso se deberá ejecutar periódicamente o por única vez?
- ¿Será un proceso de ejecución manual, automática o a demanda?

Definir las fuentes de datos que intervendrían en dicha solución

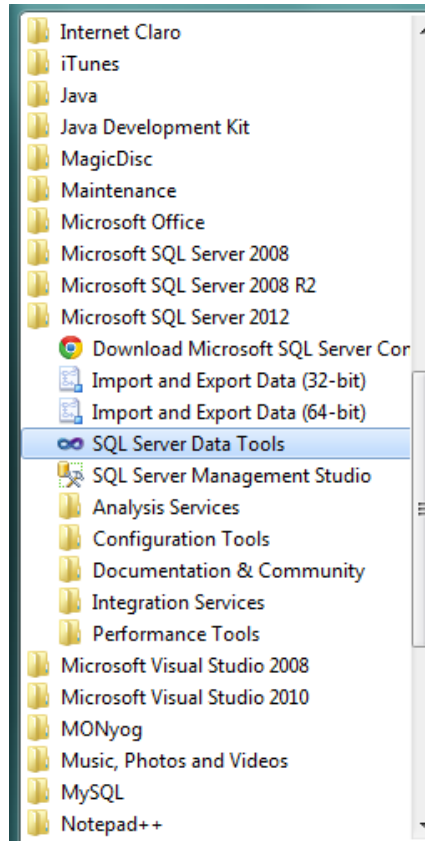
- ¿El ETL involucra un solo servidor o más de uno?
- ¿El ETL involucra solo servidores SQL SERVER?
- Si existiera la necesidad de involucrar a otras fuentes de datos, ¿En qué plataformas están (Oracle, Ms.Excel, Correo, MySql, archivos planos, etc.)?
- Para el paso anterior, ¿Tenemos los Drivers apropiados para podernos conectar a las fuente de datos en cuestión?

Definir si la fuente de datos contiene datos que son factibles transformarlos para poder obtener datos de calidad

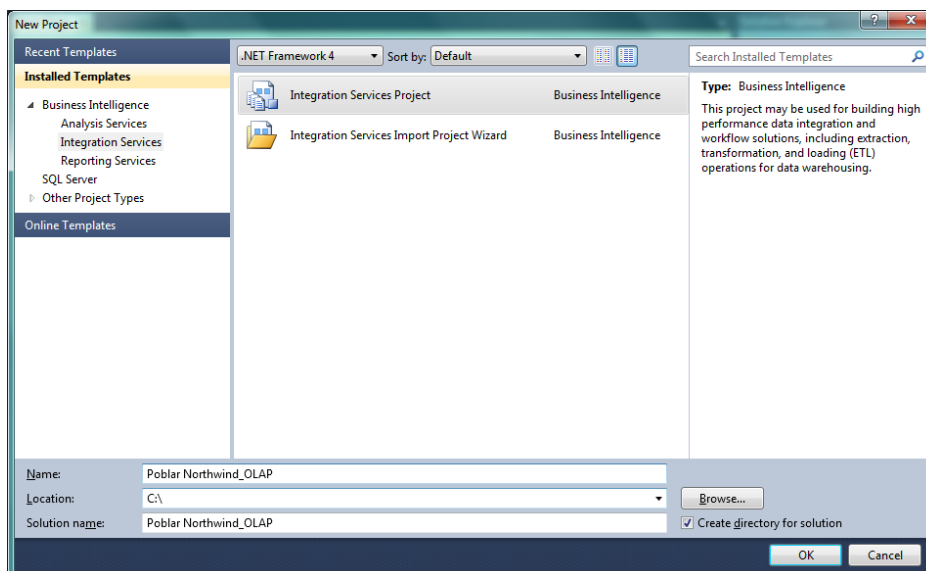
Definir específicamente donde están los datos dentro de la fuente de datos para preparar las sentencias de acceso a datos

## Ingreso al entorno del SSIS

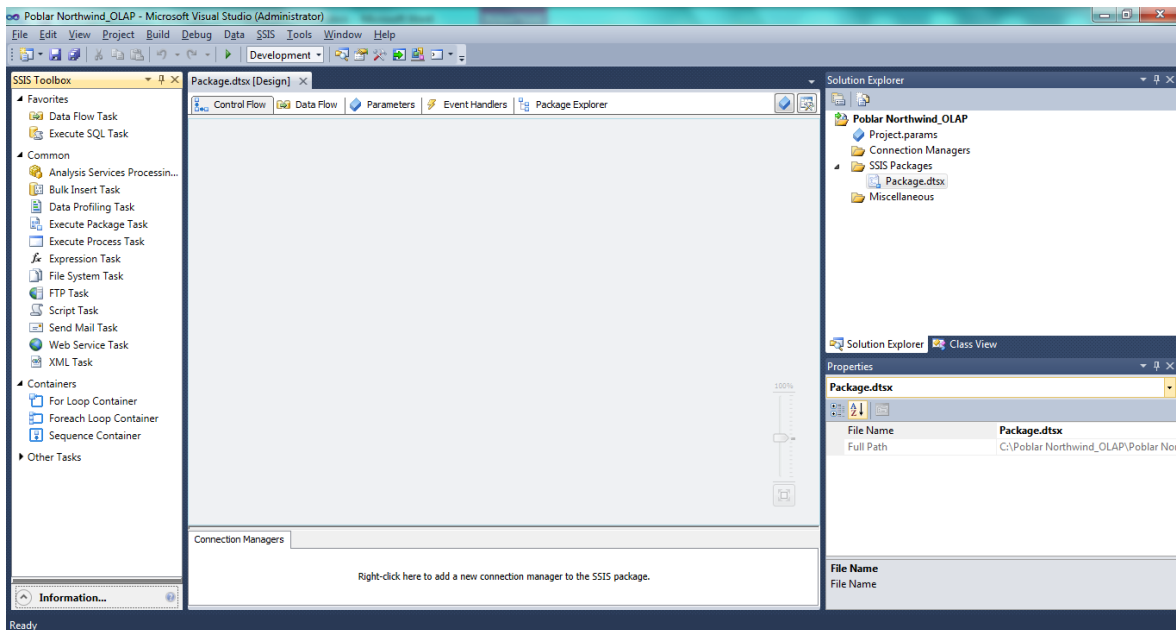
Para comenzar nuestro trabajo en la creación de paquetes ETL, debemos ingresar *SQL Server Data Tools* y crear un nuevo proyecto de SQL Server Integration Services.



A continuación aparecerá el entorno de desarrollo de Microsoft Visual Studio. Para generar un nuevo proyecto de SSIS elegimos la opción Archivo / Nuevo / Proyecto. Aparecerá la pantalla de proyectos en el cual seleccionamos *Proyecto de Integration Services* y le asignamos como nombre: Poblador\_Nortwind\_OLAP.



A continuación se muestra el entorno de desarrollo de los paquetes de ETL.



## Implementación de Paquetes SSIS

El ejemplo que se está desarrollando requiere la implementación de dos paquetes. El primero llevará la data de la base de datos Northwind a Northwind\_Stage y el segundo de Northwind\_Stage a Northwind\_Mart.

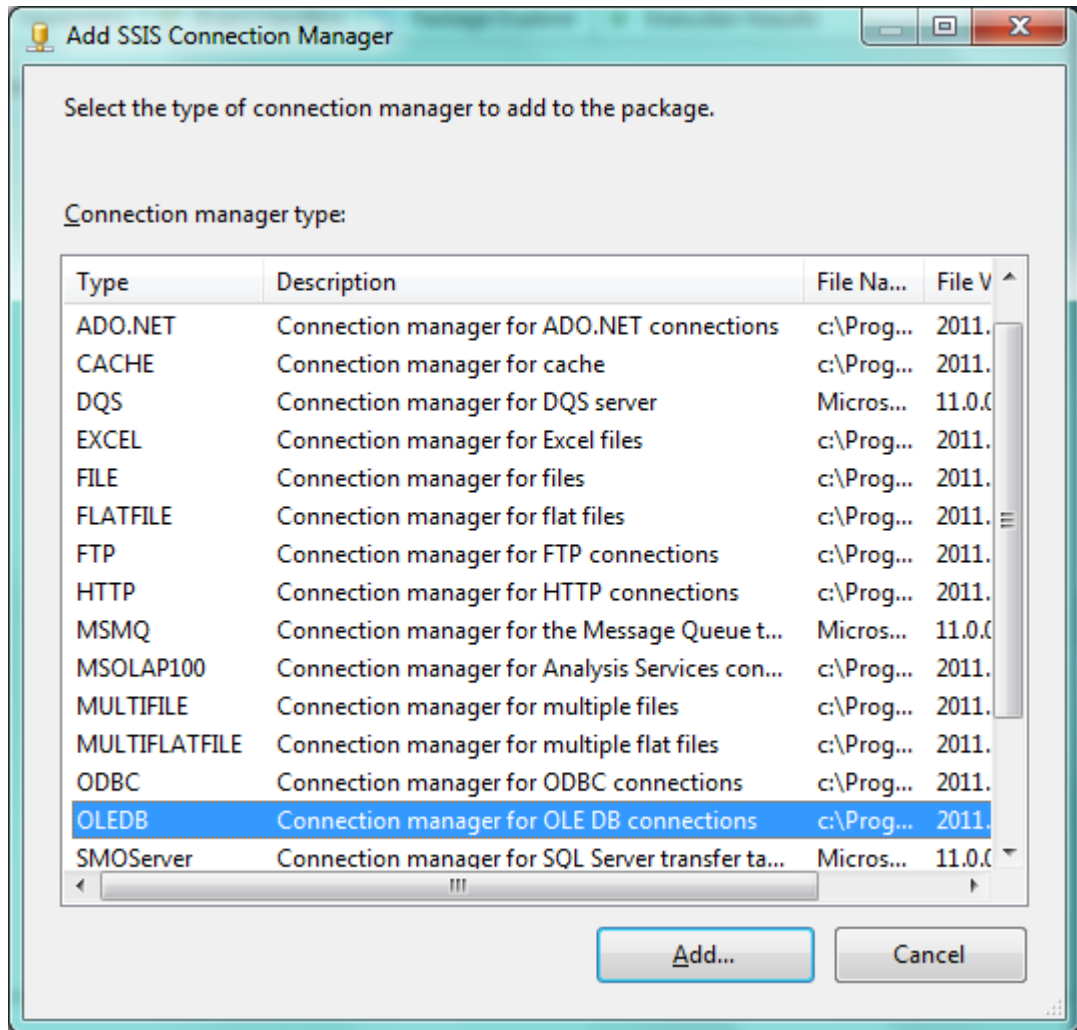
### A. Connection Managers

Lo primero que tenemos que hacer es conectarnos a la base de datos Northwind que es la fuente de la cual vamos a extraer los datos. Para ello debemos generar una conexión hacia ese origen. Para ello realizaremos los siguientes pasos:

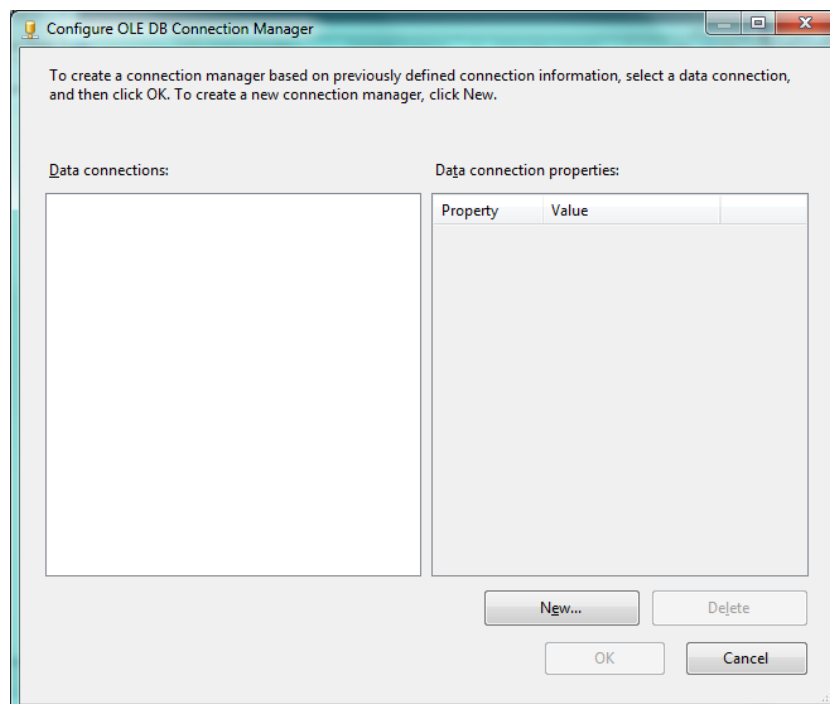
En el explorador de soluciones seleccionar la carpeta de Connection Managers. Clic derecho / New Connection Managers.

En el asistente para Connection Managers.

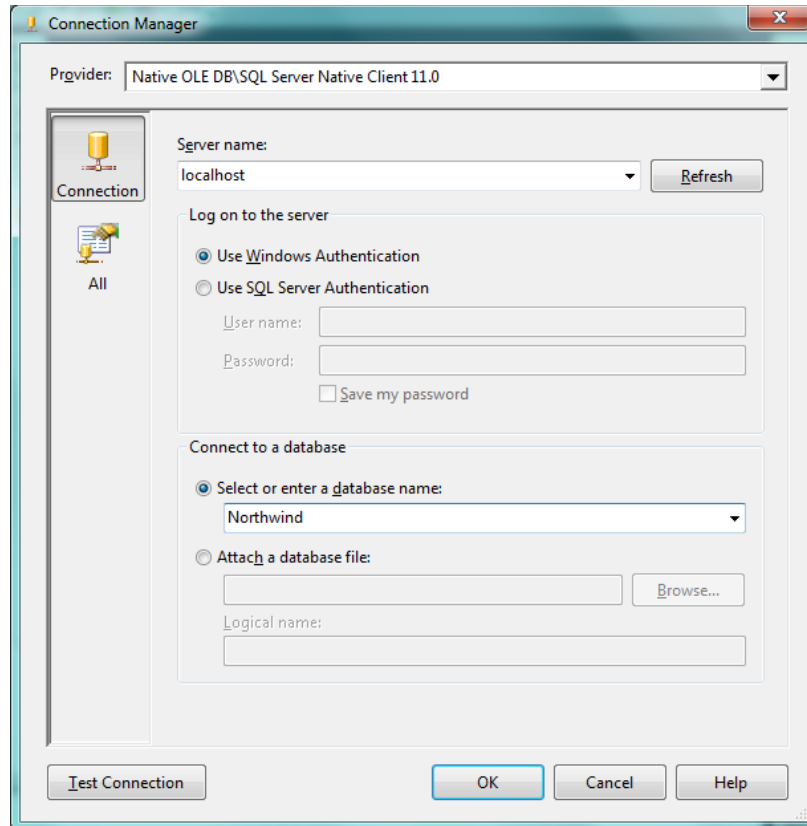
Aparece la pantalla donde nos muestra la lista de proveedores de datos, escogemos OLE DB y le damos clic en el botón Add..



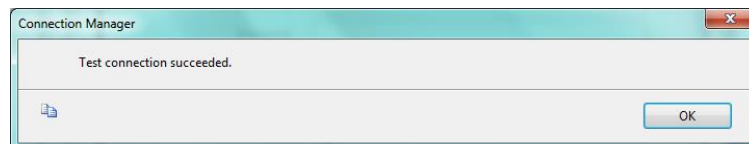
En la siguiente pantalla le damos clic en el botón New:



Aparece la pantalla donde se configurará la conexión a la base de datos Northwind. En nombre del servidor digitamos localhost y en base de datos escogemos Northwind:



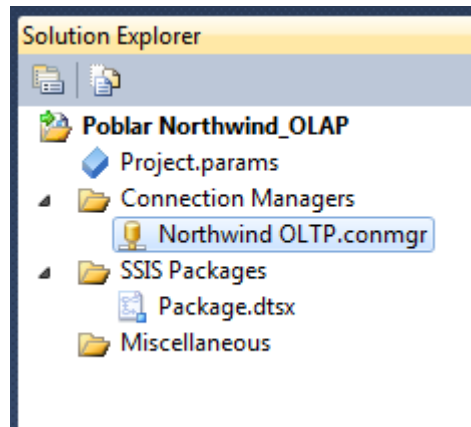
- Finalmente probamos la conexión. Para verificar que nos podamos conectar a la base de datos seleccionada. Clic en Probar Conexión.



- Clic en Aceptar.

Clic en Finalizar.

Una vez finalizada la conexión le asignamos un nombre: *Northwind OLTP*.



### Crear paquete *ETL\_Poblar\_Stage* en SSIS

Antes de crear un nuevo paquete, seleccionamos el paquete *Package.dtsx* que aparece en la carpeta Paquetes SSIS: Clic derecho / Eliminar. En la ventana de confirmación Clic en Aceptar.

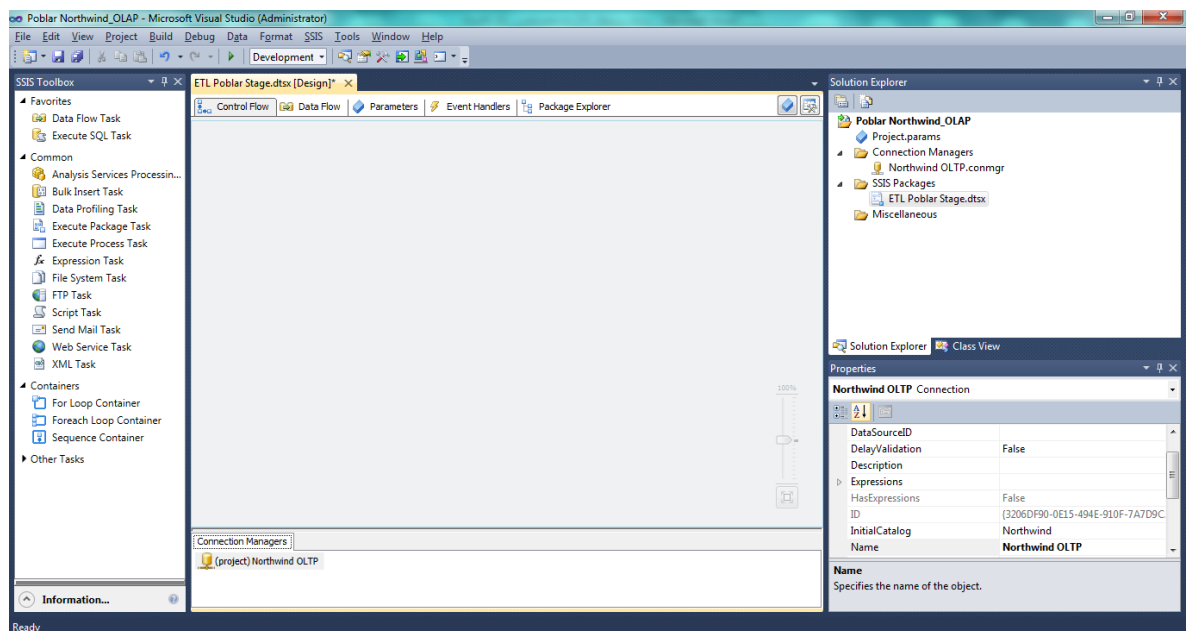
Para crear un nuevo paquete realizaremos los siguientes pasos:

En el explorador de soluciones seleccionar la carpeta de Paquetes SSIS (SSIS Packages). Clic derecho / Nuevo Paquete de SSIS (New SSIS Package).

Se ha creado un nuevo paquete denominado *Package1.dtsx*. Para renombrarlo: Clic derecho / Cambiar Nombre

Renombre el paquete como *ETL\_Poblar\_Stage.dtsx*.

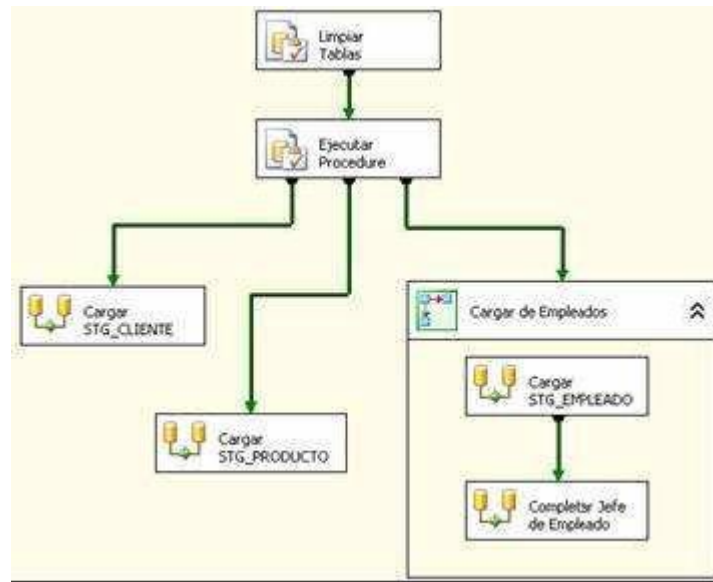
Se mostrará el entorno en el cual se desarrollará el paquete ETL.



Nota: Si el cuadro de herramientas no aparece. Entonces ir al Menu SSIS / SSIS Toolbox

### D. Construir paquete *ETL\_Poblar\_Stage*

Ahora vamos a construir el primer paquete de ETL que va a trasladar la información de la base de datos Northwind a la base de datos Northwind\_Stage.

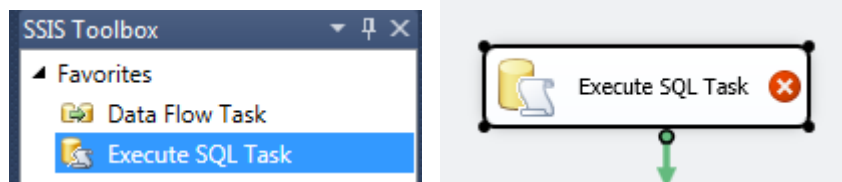


Las actividades desarrolladas en el ETL son:

- Eliminar los registros de las tablas de Stage.
- Inicializar los campos identity de las tablas del Stage mediante un Stored Procedure.
- Implementar 3 Flujos de Dato (Cliente, Producto, Empleados).

### Actividad Eliminar Registros

Del Cuadro de Herramientas seleccionamos de Favoritos la tarea *Tarea Ejecutar SQL* (Execute SQL Task) y la arrastramos a Diseñador.

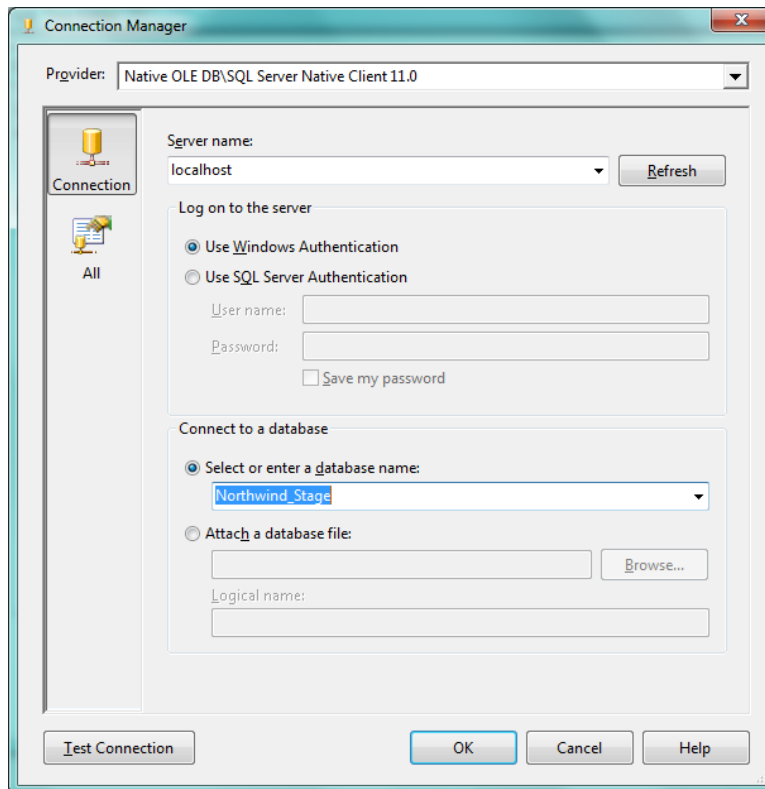


Clic derecho sobre la tarea, Editar.  
Modificamos las siguientes propiedades

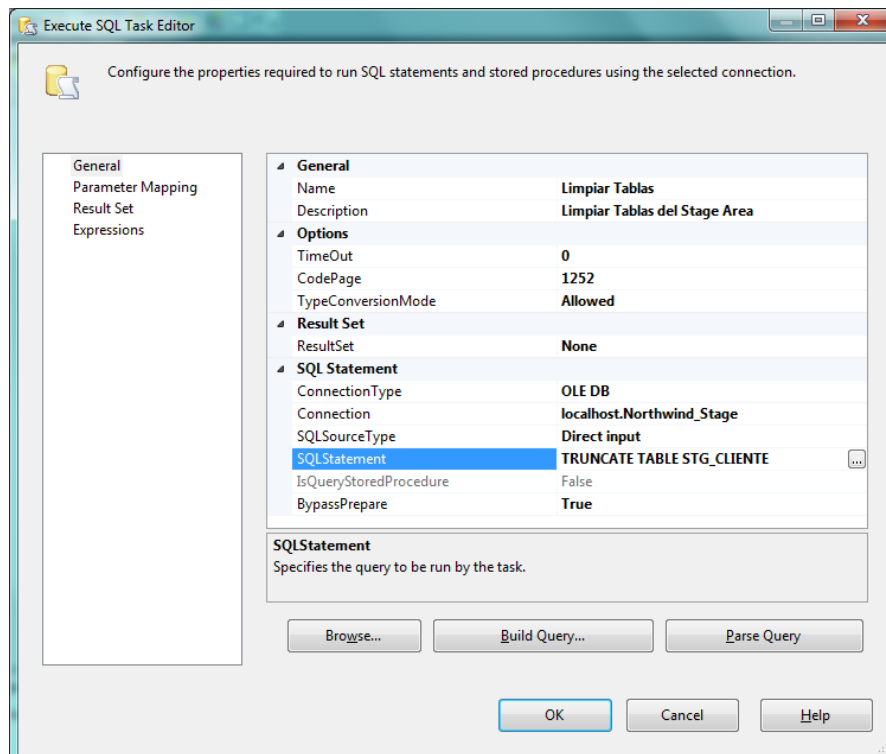
Propiedad	Valor
Name	Limpiar Tablas
Description	Limpiar las Tablas del Stage Area
ConnectionType	OLE DB
Connection	Localhost.Northwind_Stage
SQL Statement	Truncate Table STG_Cliente

En Connection creamos una nueva conexión a la base de datos Northwind\_Stage:





En SQL Statement digitamos la sentencia T-SQL Truncate Table STG\_Cliente, quedando finalmente como muestra la siguiente figura:

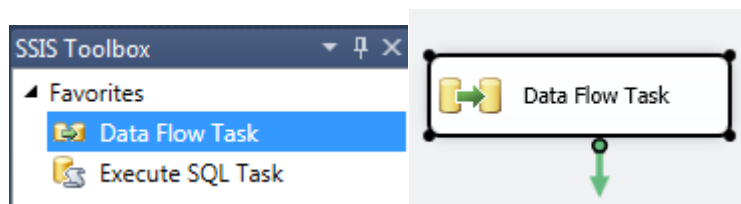


Para probar la tarea tenemos que ejecutarla. Para ello seguimos los siguientes pasos.

- Seleccionamos la Tarea / Clic Derecho / Ejecutar Tarea
- La tarea comienza su ejecución. Durante este proceso la tarea cambia de color.
- \_ Amarillo.- Tarea se está ejecutando.
- \_ Rojo.- Tarea terminó con error.
- \_ Verde.- Tarea terminó correctamente.
- Al finalizar la tarea paramos la ejecución (Menú Depurar / Detener Depuración)

### Actividad Cargar STG\_Cliente

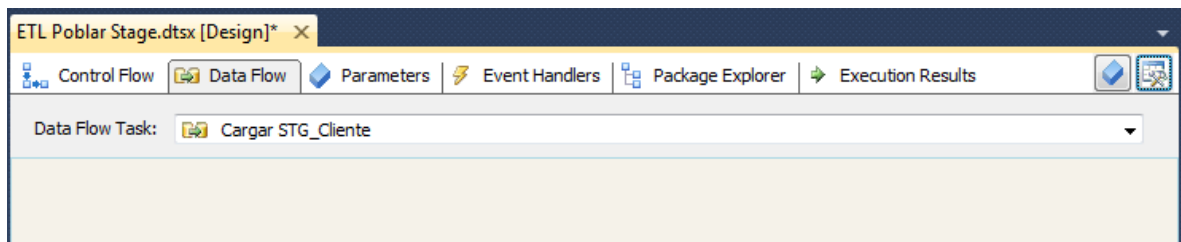
Del Cuadro de Herramientas seleccionamos de Favoritos la tarea *Tarea Flujo de datos* y la arrastramos a Diseñador.



Cambiar el nombre de la tarea a Cargar STG\_CLIENTE.

Clic derecho sobre la tarea y seleccionamos la opción Editar.

Automáticamente la pantalla del Diseñador cambia a la pestaña **Flujo de Datos** y el Cuadro de Herramientas cambia de tareas disponibles.



La pantalla de Flujo de Datos nos permitirá realizar la tarea de ETL. Para ello necesitaremos conectarnos a un origen de datos, diseñar las tareas de transformación para finalmente trasladarla hacia el destino final.

Para desarrollar nuestra Tarea de Flujo de Datos seguiremos los siguientes pasos:

De Other Sources seleccionamos Ado Net Source .

Cambiamos el nombre de la tarea a Leer Customers.

Damos doble clic sobre la tarea Leer Customers o Clic derecho editar.

Aparece la pantalla con las propiedades de la tarea seleccionada, con las siguientes opciones:

- Administrador de Conexión
- \_ Administrador de Conexión OLE DB.- Define el origen del cual se

leerá los datos.

\_ Modo de Acceso a Datos.- Define la forma en como se leerán los datos. Esta puede ser

Tabla o Vista.

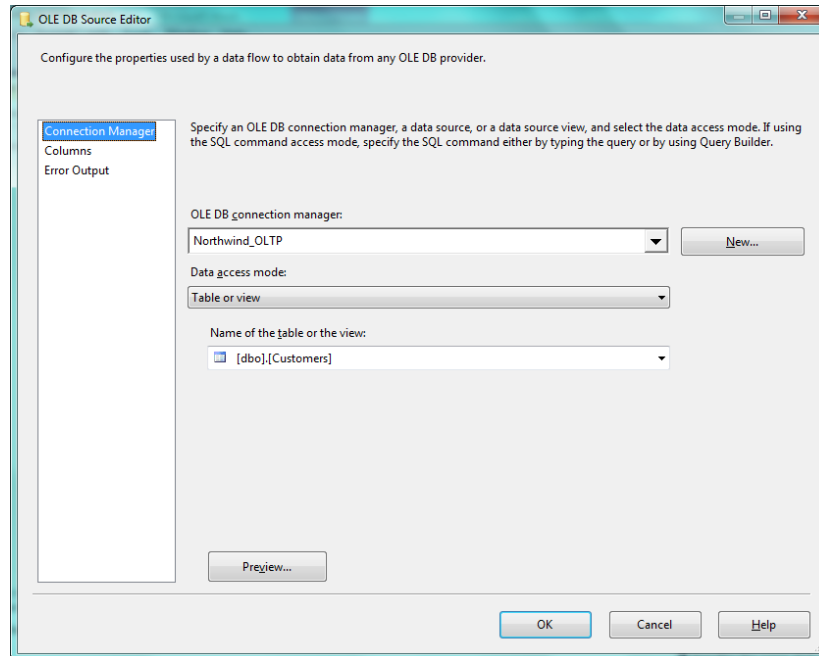
Variable de Nombre de tabla o Nombre de vista.

Comando SQL.

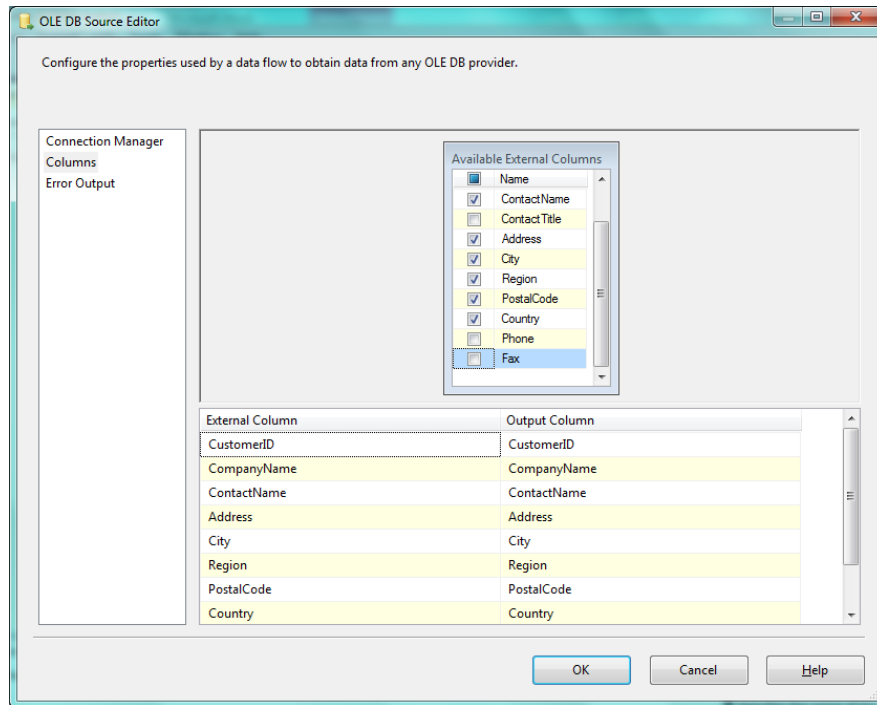
Comando SQL con Variable.

Para el ejemplo, debemos considerar los siguientes valores:

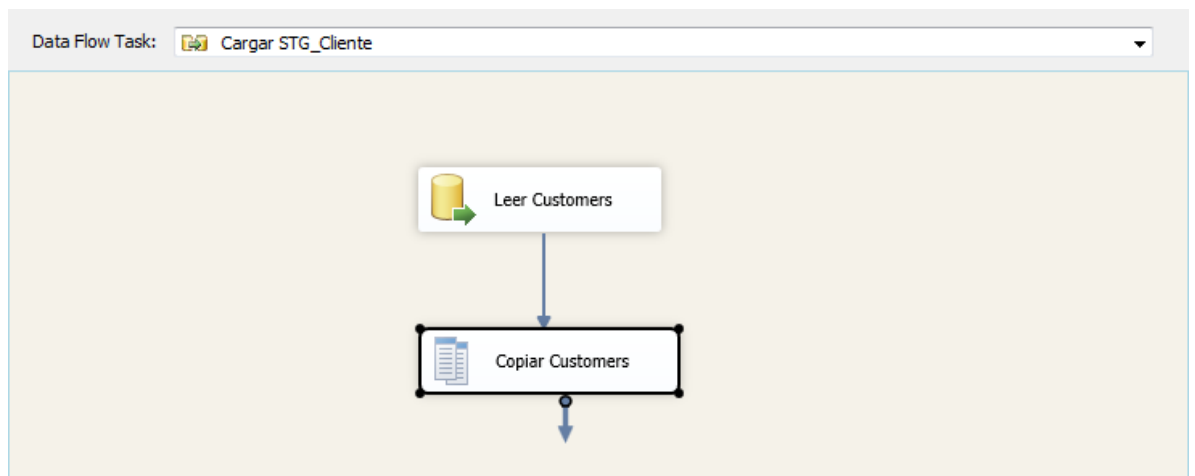
Propiedad	Valor
Administrador de Conexión OLE DB	Northwind_OLTP
Modo de Acceso a Datos	Tabla o Vista
Nombre de la tabla o vista	[dbo].[Customers]



Luego seleccionamos las columnas que utilizaremos en el proceso de carga.  
Seleccionamos todas las columnas excepto ContactTitle, Phone y Fax y damos  
Clic en Aceptar.

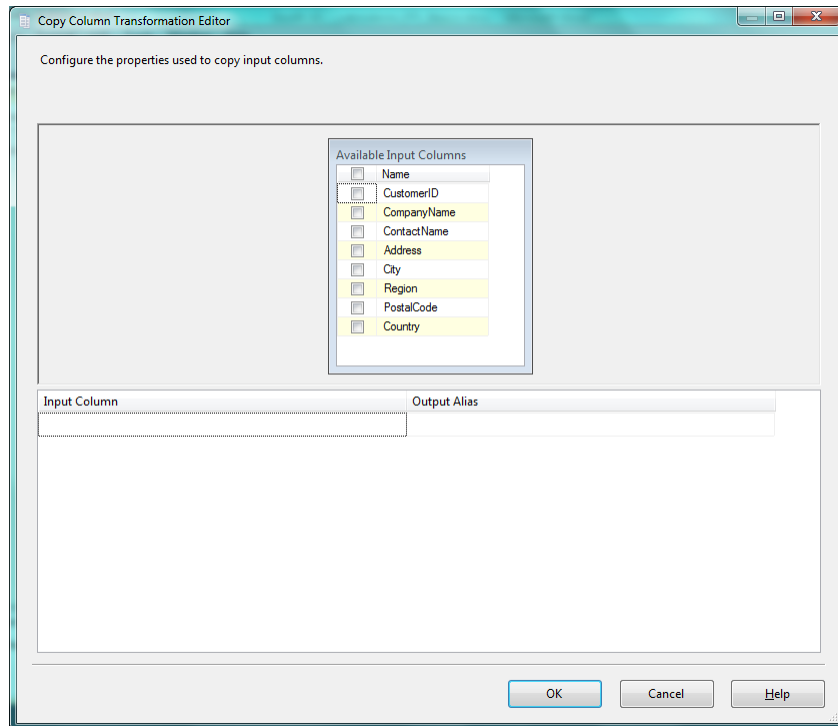


Del Cuadro de Herramientas, en Transformación de flujo de datos seleccionamos la tarea *Copiar Columna*, y lo asociamos con la tarea *Leer Customers*. Debe mostrarse de la siguiente manera.

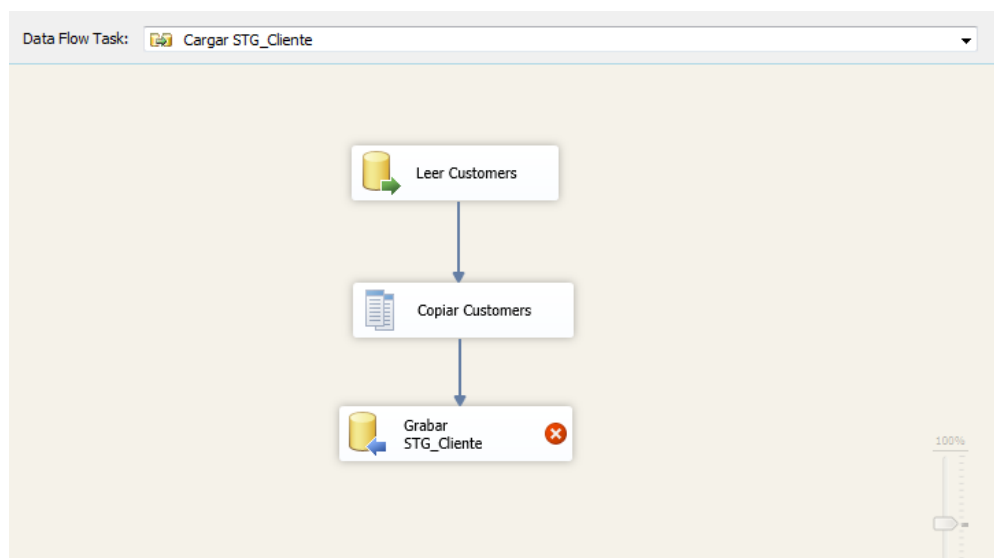


Cambiamos el nombre de la tarea a Copiar Customers.

Si le damos Clic derecho Editar o doble clic sobre la tarea Copiar Customers, estas muestra la lista de columnas seleccionadas en la tarea Leer Customers.



Finalmente, debemos indicar hacia donde se copiaran los datos. Del Cuadro de Herramientas, en *Others Destinations* seleccionamos la tarea *OLE DB Destination*, y lo asociamos con la tarea *Copiar Customers*. Debe mostrarse de la siguiente manera



Cambiamos el nombre de la tarea a Grabar STG\_Cliente.

Si le damos Clic derecho Editar o doble clic sobre la tarea Grabar Cliente.

Aparece el Editor de Destino OLE DB con las siguientes opciones:  
Administrador de Conexión

\_ Administrador de Conexión OLE DB.- Define el destino en el cual se grabaran los datos.

Modo de Acceso a Datos.- Define la forma en como se grabaran los

datos. Esta puede ser:

Tabla o Vista.

Carga rápida de Tabla o Vista.

Variable de Nombre de tabla o Nombre de vista.

Carga rápida de variable de Nombre de tabla o Nombre de vista

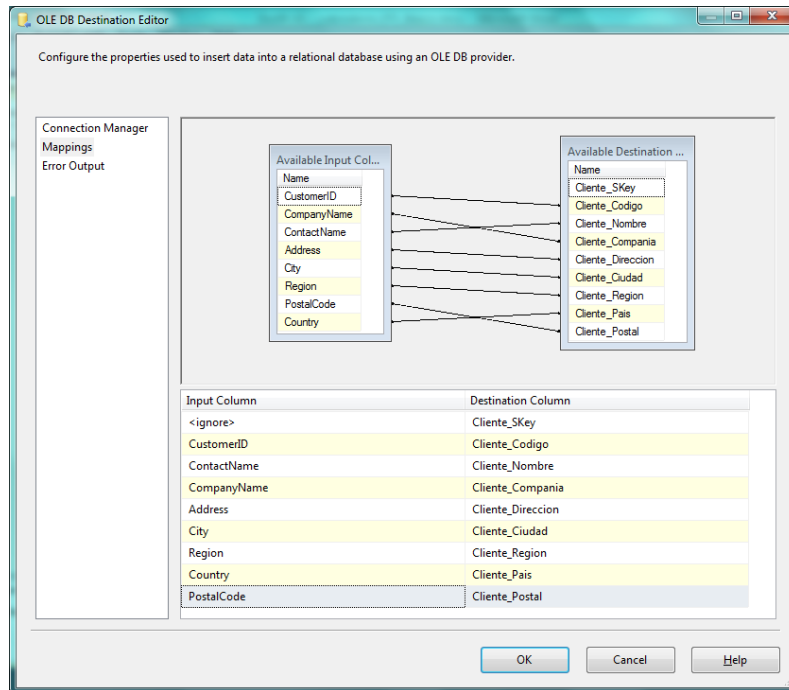
Comando SQL.

Para el ejemplo, debemos considerar los siguientes valores:

Propiedad	Valor
Administrador de Conexión OLE DB	localhost.Northwind_Stage
Modo de Acceso a Datos	Tabla o Vista
Nombre de la tabla o vista	[dbo].[STG_Cliente]

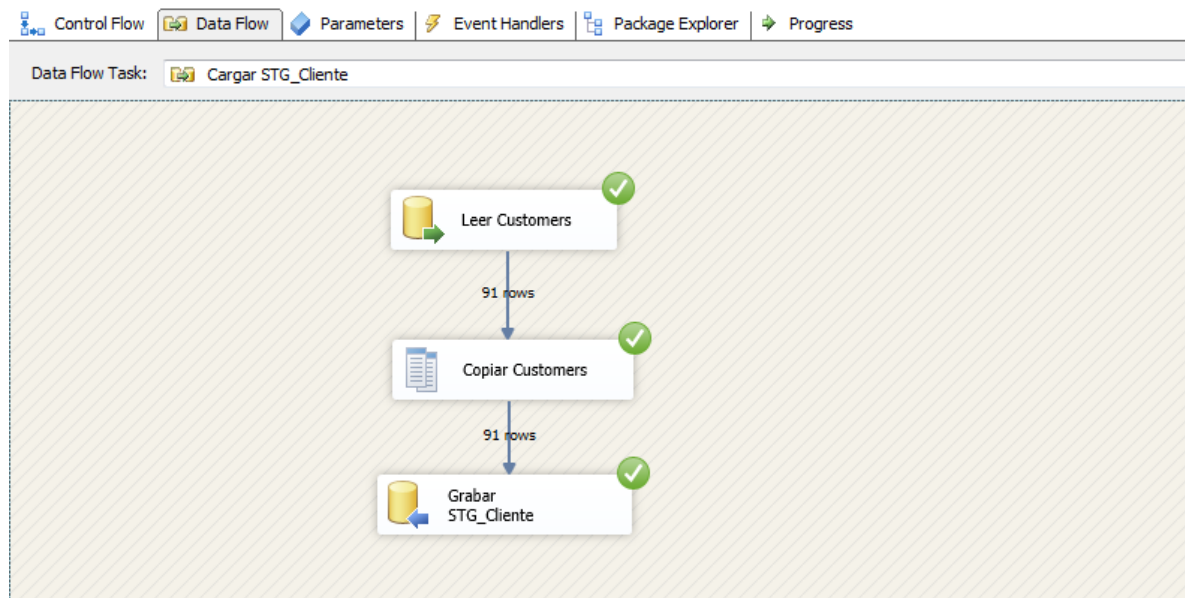
En la opción de Asignaciones, relacionamos las columnas del origen (o de las transformaciones) con las columnas del destino.

Origen	Destino
CustomerID	Cliente_Codigo
ContactName	Cliente_Nombre
CompanyName	Cliente_Compania
Address	Cliente_Direccion
City	Cliente_Ciudad
Region	Cliente_Region
Country	Cliente_País
PostalCode	Cliente_Postal



Clic en Aceptar.

Note que no existe ninguna columna del origen asociada a la columna Cliente\_SKey. Esto debido a que este campo es de tipo Identity, por lo que el valor que se asignará es administrado por las base de datos SQL Server 2012. Para poder probar el paquete, ubíquese en Flujo de control, seleccione la tarea Cargar STG\_Cliente y ejecútelo. Debe mostrar la siguiente pantalla:



Al finalizar la tarea pare la ejecución (Menú Depurar / Detener Depuración)

Grabe el proyecto (Menú Archivo / Guardar Todo).

### Actividad Cargar STG\_Producto

Para cargar la información de Productos se necesitará realizar los siguientes pasos:

Del Cuadro de Herramientas seleccionamos de Favoritos la tarea *Tarea Flujo de Datos* y la arrastramos a Diseñador

Cambiar el nombre de la tarea a Cargar STG\_PRODUCTO.

Clic derecho sobre la tarea y seleccionamos la opción Editar.

Automáticamente la pantalla del Diseñador cambia a la pestaña **Flujo de Datos** y el Cuadro de Herramientas cambia de tareas disponibles.

En la pantalla de Flujo de Datos desarrollaremos las tareas de ETL. Para ello verificamos que tengamos generadas las conexiones que necesitaremos.

Esta tarea debe tener el siguiente código y utilizar la conexión Northwind\_OLTP creada.

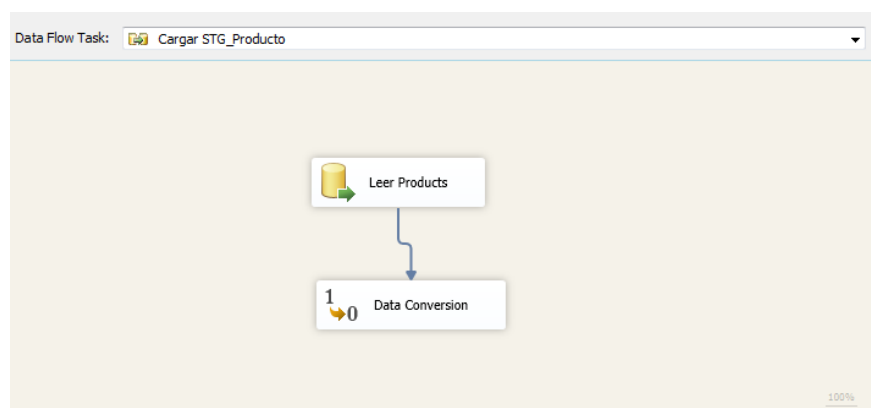
De orígenes de flujo de datos seleccionamos *Origen de OLE DB*.

Cambiamos el nombre de la tarea a Leer Products.

Damos doble clic sobre la tarea Leer Products o Clic derecho editar. Aparece la pantalla con las propiedades de la tarea seleccionada, configure las siguientes propiedades:

Propiedad	Valor
Administrador de Conexión OLE DB	Northwind_OLTP
Modo de Acceso a Datos	Tabla o Vista
Nombre de la tabla o vista	[dbo].[Products]
Columnas	ProductID ProductName CategoryID UnitPrice

Del Cuadro de Herramientas, en Common seleccionamos la tarea *Conversión de Datos*, y lo asociamos con la tarea Leer Products. Debe mostrarse de la siguiente manera.



La tarea de *Conversión de Datos* nos permitirá modificar los distintos tipos de datos que podemos encontrar en las distintas fuentes de datos al tipo de dato del destino. Para el formato de Fecha, para el tamaño de los campos tipo carácter, para los valores numéricos, etc.

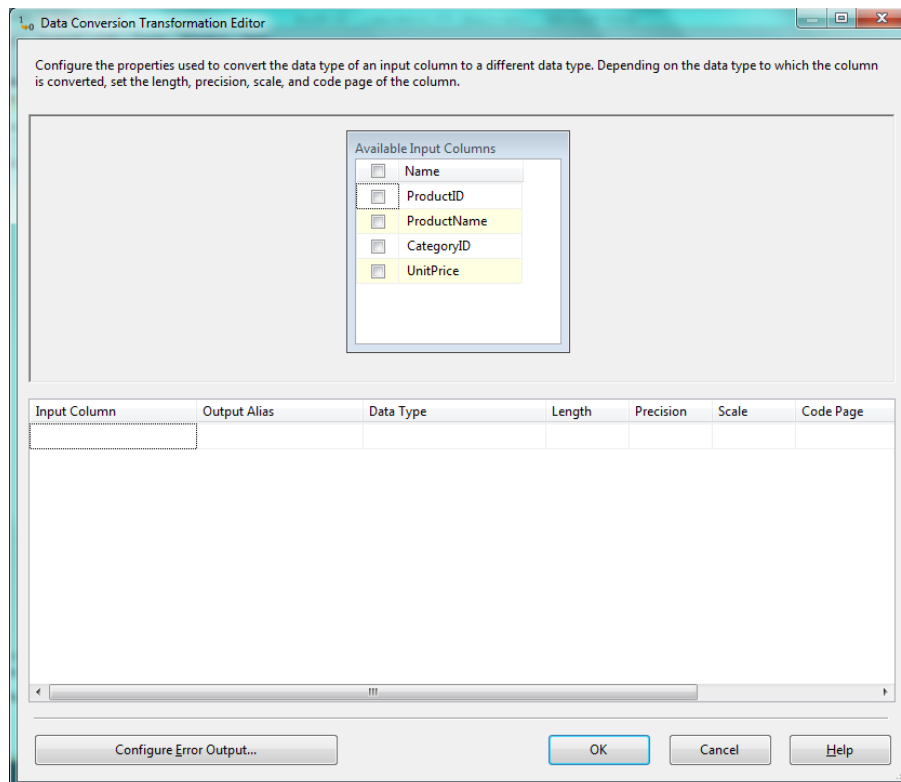
En el ejemplo si vemos el tamaño del campo ProductName de la base de datos Northwind y lo comparamos con el tipo de dato de base de datos Northwind\_Stage, veremos que estas no son iguales, para estandarizarlas



realizaremos un una conversión de datos.

Northwind		Northwind_Stage	
ProductName	nvarchar(40)	Producto_Nombre	varchar(35)

Si le damos Clic derecho Editar o doble clic sobre la tarea Conversión de Datos, nos aparecerá la siguiente pantalla.



Las propiedades a incluir son:

Columna de Entrada.- Seleccionamos la columna que queremos modificar.

Alias de Salida.- Indicamos el nuevo nombre que se asignará a la columna.

Tipo de Datos.- Indicamos el nuevo tipo de dato que se le asignará a la columna.

Longitud.- Si el campo es carácter se debe especificar el tamaño que tendrá en nuevo campo.

Precisión / Escala.- Si el campo es numérico se debe especificar el numero de enteros y decimales que soportará.

Para el ejemplo, utilizaremos los siguientes valores:

Propiedad	Valor
Columna de Entrada	ProductName
Alias de Salida	ProductName_Varchar
Tipo de Datos	Cadena[DT_STR]
Longitud	35

Input Column	Output Alias	Data Type	Length
ProductName	ProductName_VarChar	string [DT_STR]	35

Finalmente, debemos indicar hacia donde se copiaran los datos. Del Cuadro de Herramientas, en *Others Destinations* seleccionamos la tarea *OLE DB Destination*, y lo asociamos con la tarea *Conversión de Datos*. Debe mostrarse de la siguiente manera

Cambiamos el nombre de la tarea a Grabar STG\_Producto.

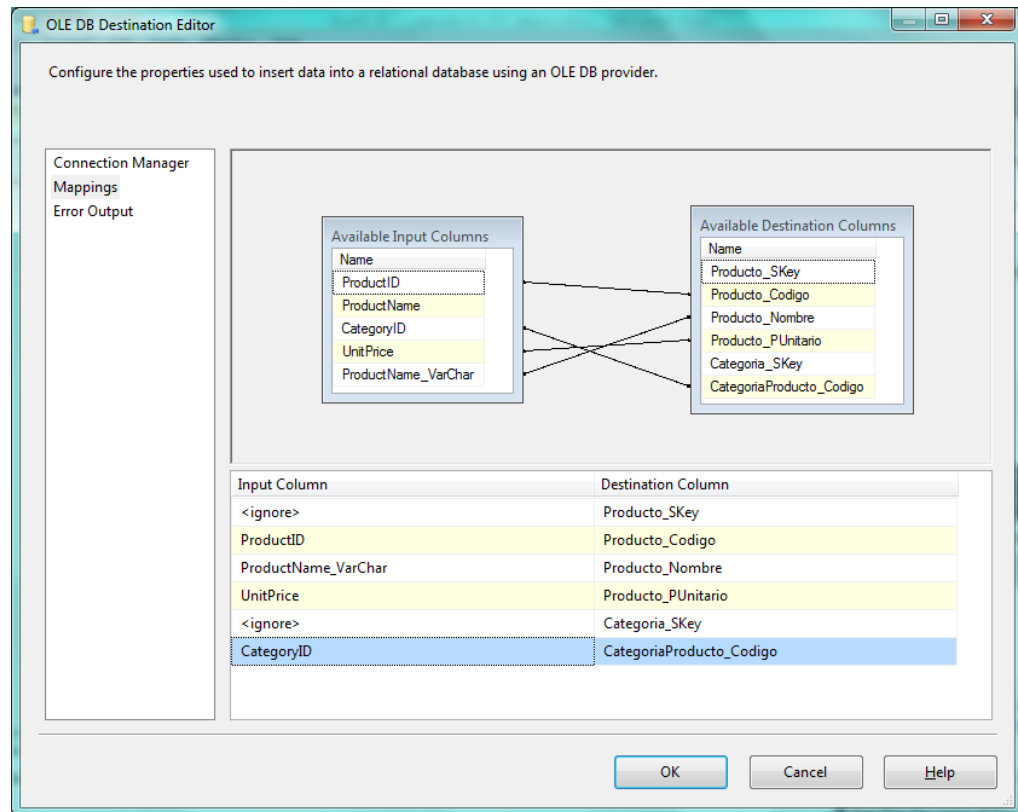
Damos doble clic sobre la tarea Grabar STG\_Producto o Clic derecho editar.

Aparece la pantalla con las propiedades de la tarea seleccionada, configure las siguientes propiedades:

Propiedad	Valor
Administrador de Conexión OLE DB	Localhost.Northwind_Stage
Modo de Acceso a Datos	Tabla o Vista
Nombre de la tabla o vista	[dbo].[Stg_Producto]

En la opción de Asignaciones, relacionamos las columnas del origen (o de las transformaciones) con las columnas del destino.

Origen	Destino
ProductID	Producto_Codigo
ProductName_Varchar	Producto_Nombre
UnitPrice	Producto_PUnitario
CategoryID	CategoriaProducto_Codigo

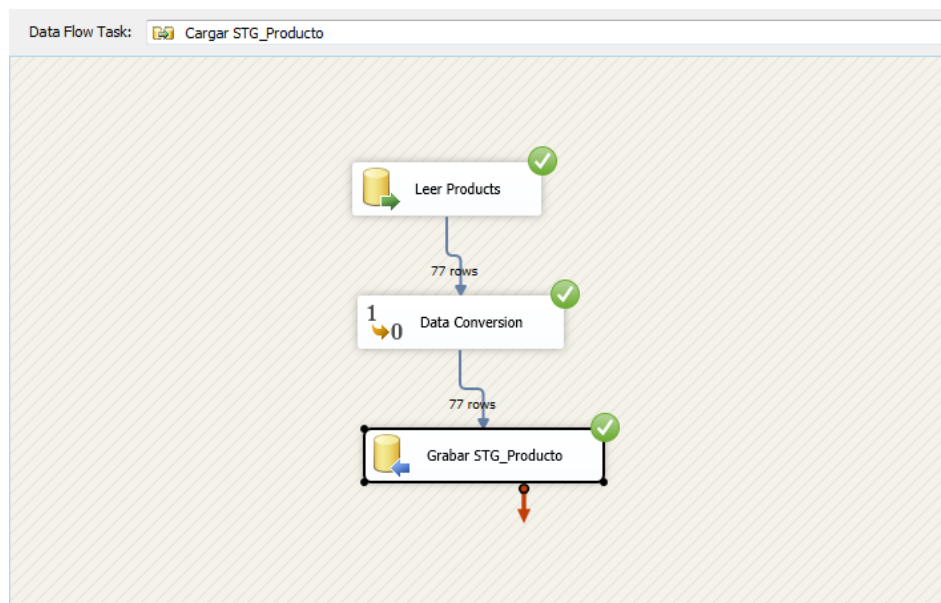


Clic en Aceptar.

Note que al igual que en la carga de la tabla STG\_Cliente, para la carga de STG\_Producto tampoco existe ninguna columna del origen asociada a la columna Producto\_SKey. Al igual que en el caso anterior el valor lo asignará la base de datos SQL Server 2012.

En este caso tampoco existe valor asociado para el campo Categoria\_SKey. En este caso el valor será completado cuando se realice la carga de la tabla STG\_Categoria.

Para poder probar el paquete, ubíquese en Flujo de control, seleccione la tarea Cargar STG\_Producto y ejecútelo. Debe mostrar la siguiente pantalla:



■ Al finalizar la tarea pare la ejecución (Menú Depurar / Detener Depuración)

■ Grabe el proyecto (Menú Archivo / Guardar Todo).

### Actividad Cargar STG\_Empleados

Para cargar la información de Productos se necesitará realizar los siguientes pasos:

■ Del Cuadro de Herramientas seleccionamos de Favoritos la tarea *Tarea Flujo de Datos* y la arrastramos a Diseñador.

■ Cambiar el nombre de la tarea a Cargar STG\_EMPLEADO.

■ Clic derecho sobre la tarea y seleccionamos la opción Editar.

■ Automáticamente la pantalla del Diseñador cambia a la pestaña **Flujo de Datos** y el Cuadro de Herramientas cambia de tareas disponibles.

En la pantalla de Flujo de Datos desarrollaremos las tareas de ETL. Para ello verificamos que tengamos generadas las conexiones que necesitaremos. Esta tarea debe tener el siguiente código y utilizar la conexión Northwind\_OLTP creada.

■ De Others Sources seleccionamos *Origen de OLE DB*.

■ Cambiamos el nombre de la tarea a Leer Employees.

■ Damos doble clic sobre la tarea Leer Employees o Clic derecho editar.

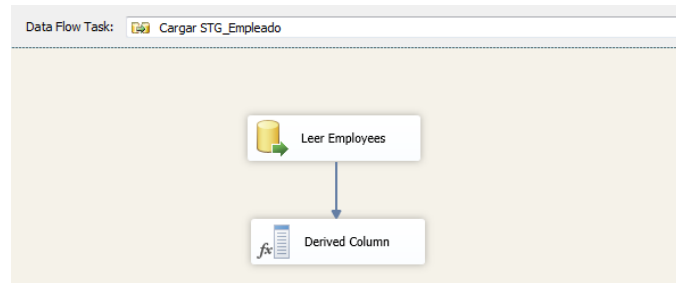
Aparece la pantalla con las propiedades de la tarea seleccionada, configure las siguientes propiedades:

Propiedad	Valor
Administrador de Conexión OLE DB	Northwind_OLTP
Modo de Acceso a Datos	Comando SQL
Texto de Comando SQL	SELECT EmployeeID, LastName, FirstName, Address, City, Region, PostalCode, Country, ReportsTo FROM Employees

Note que en la tabla Employees de la tabla Northwind, el nombre del empleado aparece separado en los campos LastName y FistName, pero en la tabla STG\_Empleado además necesitamos almacenar una columna con el nombre completo. Para esto podemos hacer dos cosas. La primera modificar la sentencia SELECT arriba indicada y agregar la columna que falta

Texto de Comando SQL	SELECT EmployeeID, LastName, FirstName, <b>LastName + ' ' + FirstName as Name</b> , Address, City, Region, PostalCode, Country, ReportsTo FROM Employees
----------------------	---

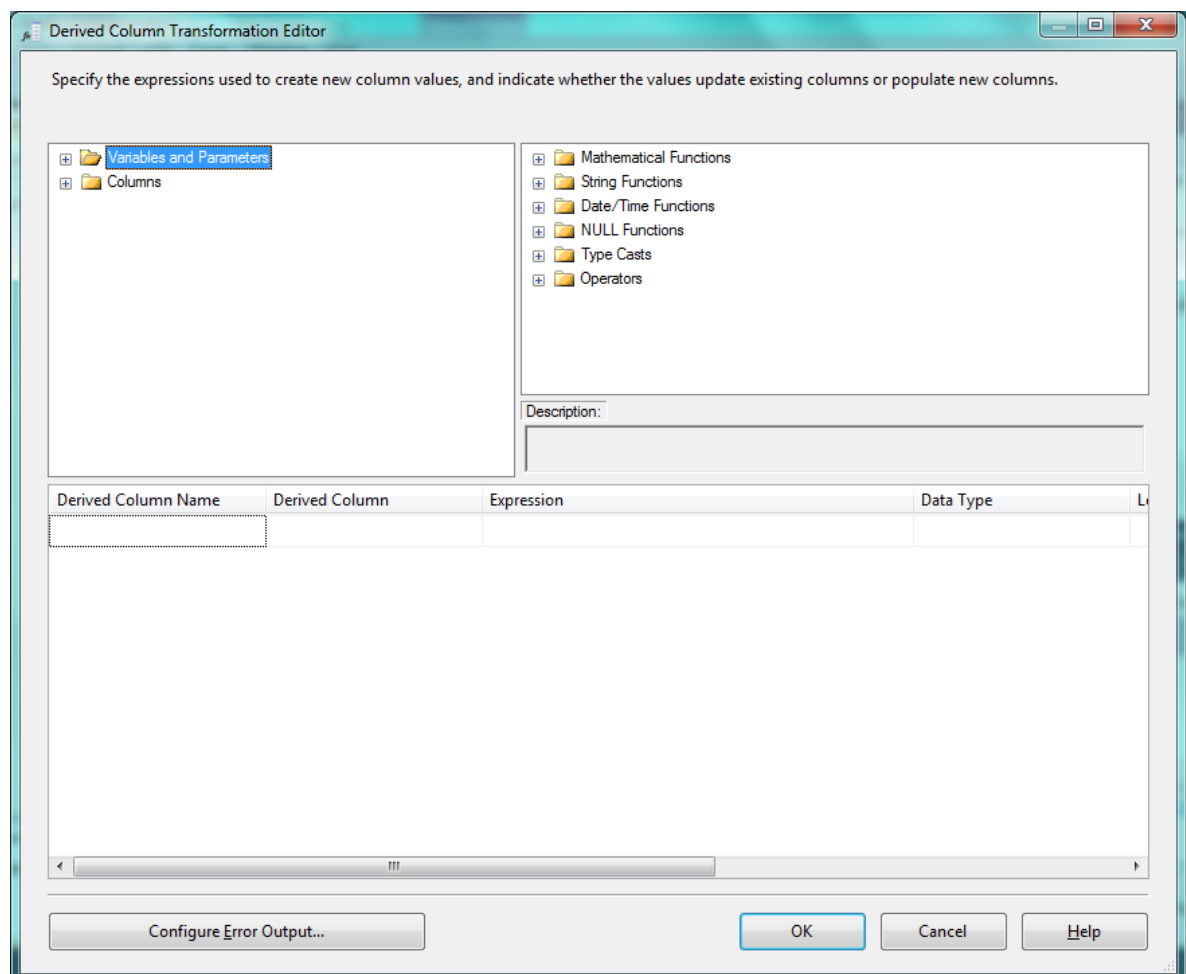
La otra es utilizar una tarea SSIS. Del Cuadro de Herramientas, en Common seleccionamos la tarea *Columna Derivada*, y lo asociamos con la tarea Leer Employees. Debe mostrarse de la siguiente manera.



La tarea de *Columna Derivada* nos permitirá crear nuevos campos a partir de los campos ya existentes. Para ello nos facilitará un conjunto de funciones de cadena, matemáticas, de fecha, nulos, conversión de tipo, etc.

En el ejemplo vamos a crear la columna FullName a partir de las columnas FirstName y LastName.

Si le damos Clic derecho Editar o doble clic sobre la tarea *Columna Derivada*, nos aparecerá la siguiente pantalla.



Las propiedades a incluir son:

Nombre de la Columna.- Indicamos el nombre de la columna que

necesitamos crear.

Columna Derivada.- Indicamos si la columna derivada será una nueva columna o reemplazará a una columna ya existente.

Expresión.- Indicamos el ¿Cómo se va a crear la nueva columna

derivada?.

Tipo de Datos.- Indicamos el nuevo tipo de dato que se le asignará a la columna.

Longitud.- Si el campo es carácter se debe especificar el tamaño que tendrá en nuevo campo.

Precisión / Escala.- Si el campo es numérico se debe especificar el numero de enteros y decimales que soportará.

Para el ejemplo, utilizaremos los siguientes valores:

Propiedad	Valor
Nombre de Columna	FullName
Columna Derivada	<agregar como columna nueva>
Expresión	TRIM( [LastName] )+" "+TRIM( [FirstName])
Tipo de Datos	Cadena Unicode[DT_WSTR]
Longitud	31

FullName	<add as new column>	TRIM( [LastName] )+" "+TRIM( [FirstName])	Unicode string [DT_WSTR]	31
----------	---------------------	---	--------------------------	----

Finalmente, debemos indicar hacia donde se copiaran los datos. Del Cuadro de Herramientas, en Others Destinations seleccionamos la tarea *OLE DB Destination*, y lo asociamos con la tarea *Columna Derivada*. Debe mostrarse de la siguiente manera

Cambiamos el nombre de la tarea a Grabar STG\_Empleado.

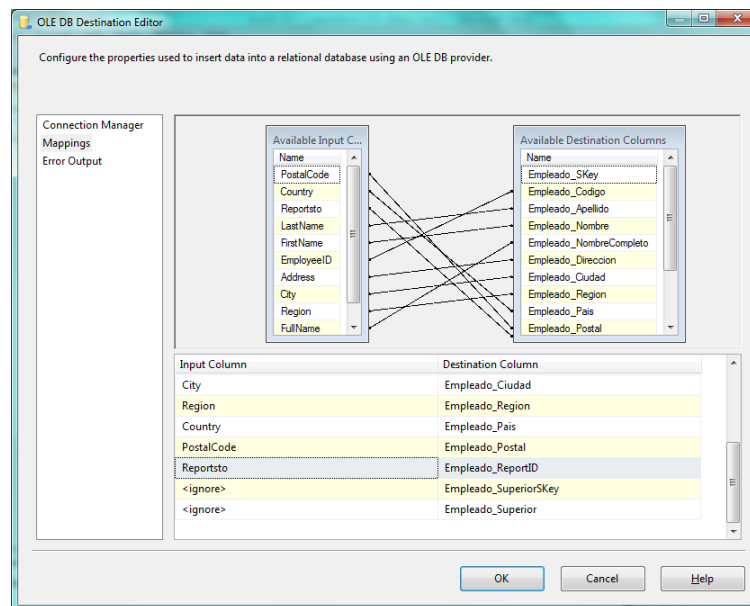
Damos doble clic sobre la tarea Grabar STG\_ Empleado o Clic derecho editar.

Aparece la pantalla con las propiedades de la tarea seleccionada, configure las siguientes propiedades:

Propiedad	Valor
Administrador de Conexión OLE DB	Localhost.Northwind_Stage
Modo de Acceso a Datos	Tabla o Vista
Nombre de la tabla o vista	[dbo].[Stg_Empleado]

En la opción de Asignaciones, relacionamos las columnas del origen (o de las transformaciones) con las columnas del destino.

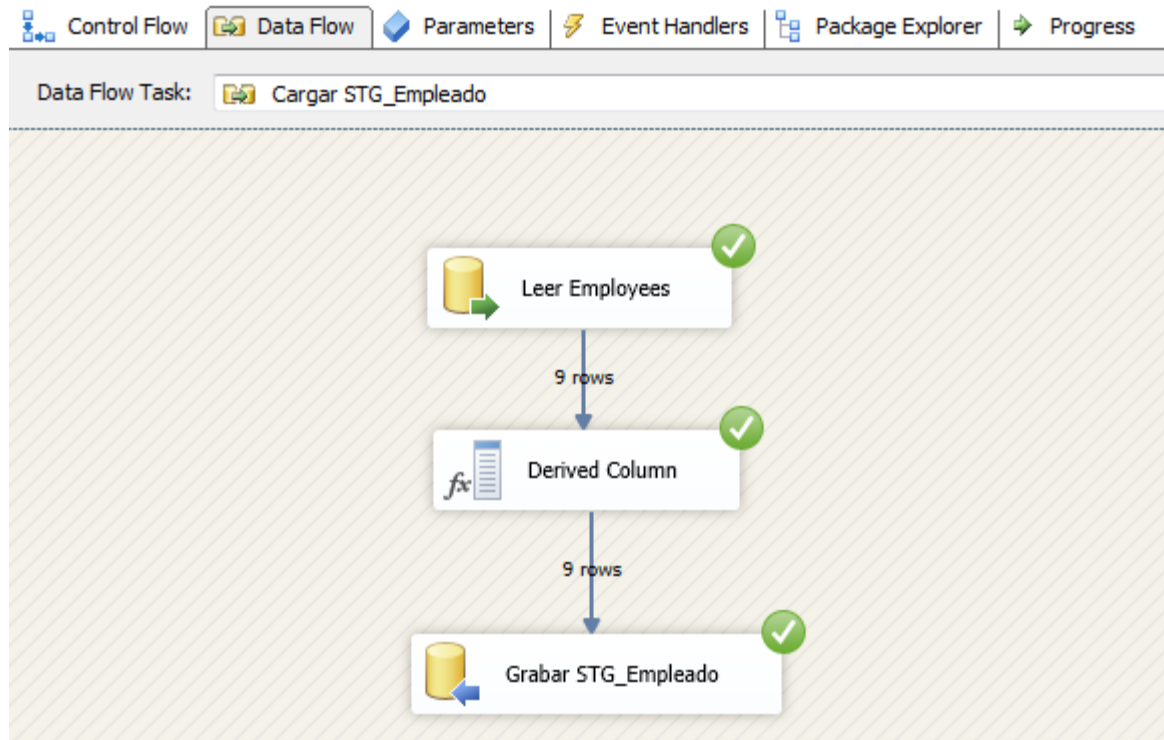
Origen	Destino
EmployeeID	Empleado_Codigo
LastName	Empleado_Apellido
FirstName	Empleado_Nombre
Address	Empleado_Dirección
City	Empleado_Ciudad
Region	Empleado_Region
PostalCode	Empleado_Pais
Country	Empleado_Postal
ReportsTo	Empleado_ReportID
FullName	Empleado_NombreCompleto



Clic en Aceptar.

Las columnas empleado SuperiorSKey y Empleado\_Superior serán completados más adelante.

Para poder probar el paquete, ubíquese en Flujo de control, seleccione la tarea Cargar STG\_Empleado y ejecútelo. Debe mostrar la siguiente pantalla:



Finalizar la tarea y grabe el Proyecto.

#### Actividad Completar Jefe de Empleado

El jefe del empleado es otro empleado más al cual al momento de ser cargado en la actividad anterior se le ha creado un surrogate key (Empleado\_SKey). Entonces, ahora debemos completar los datos del jefe del empleado para lo cual utilizaremos dicho campo.

Para completar estos datos vamos a utilizar un Procedimiento Almacenado, el cual leerá la tabla STG\_Empleado y cargará la información de su Jefe. Para ello debemos realizar los siguientes pasos:

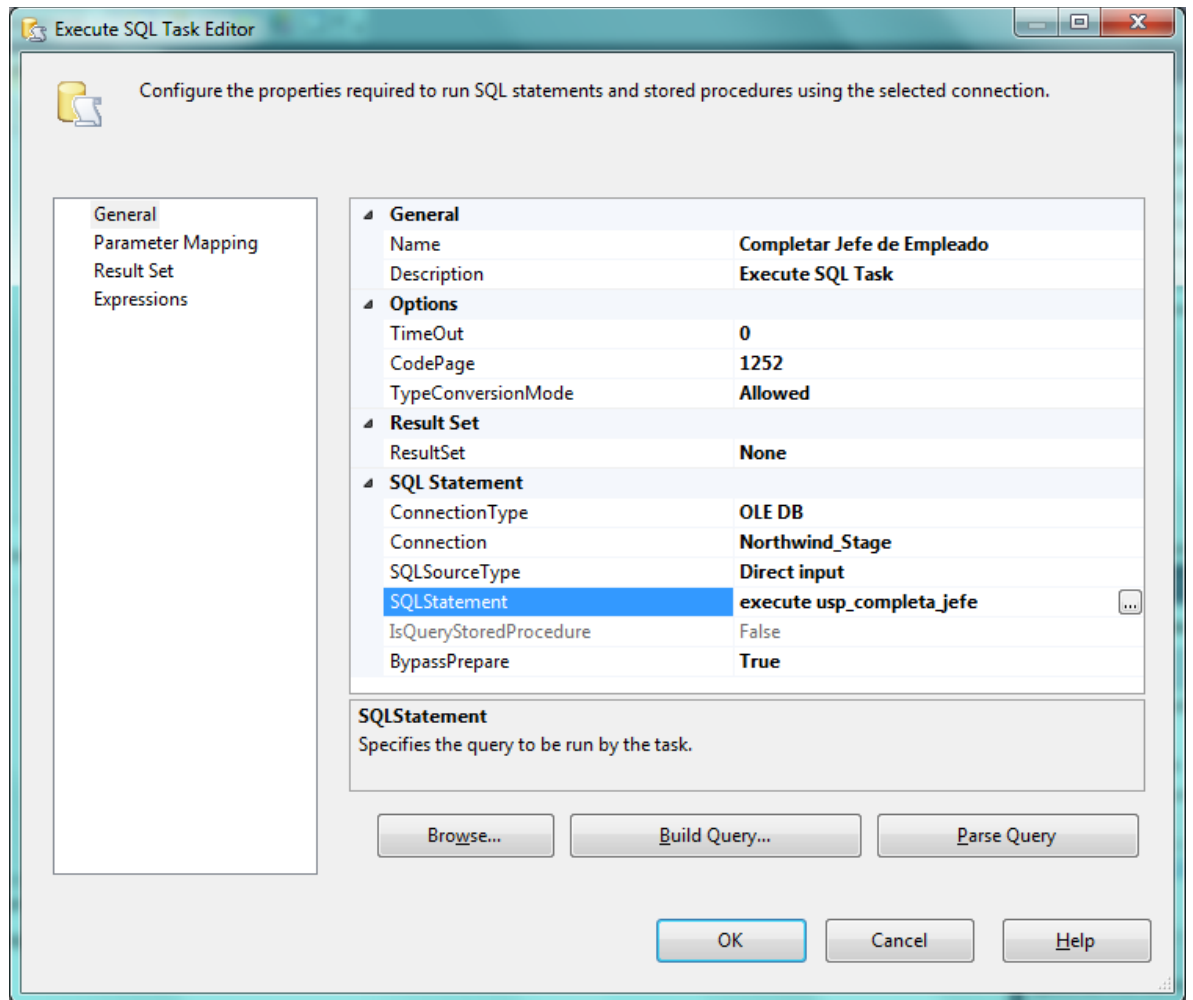
Del Cuadro de Herramientas seleccionamos de Favoritos la tarea *Tarea Ejecutar SQL* y la arrastramos a Diseñador.

Cambiar el nombre de la tarea a Completar Jefe de Empleado.

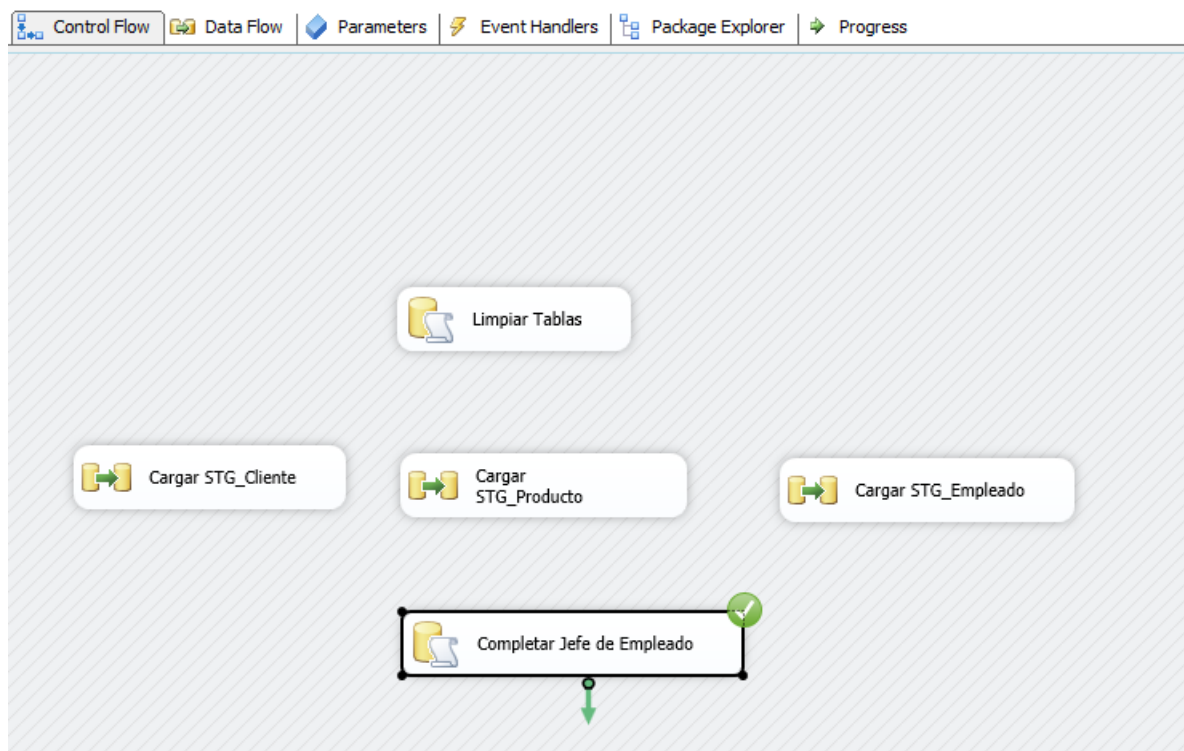
Clic derecho sobre la tarea y seleccionamos la opción Editar. Aparece la pantalla con las propiedades de la tarea seleccionada, configure las siguientes propiedades:

Propiedad	Valor
Connection Type	OLE DB
Connection	Localhost.Northwind_Stage
SQLStatement	Execute USP_Completa_Jefe  El stored procedure esta creado en la base de datos stage.  El código será proporcionado y explicado en clase.





Para poder probar el paquete, ubíquese en Flujo de control, seleccione la tarea Completar Jefe de Empleado y ejecútelo. Debe mostrar la siguiente pantalla:



Finalizar la tarea y grabe el Proyecto.