

PRÁCTICA TALLER:

DESARROLLAR UN ETL PARA LA CARGA DE UN DATA MART

Objetivos:

Desarrollar Paquetes ETL para cargar los datos de un Data mart.

ARTEFACTOS Necesarios para resolver la práctica.

- base de datos AdventureWork2012 y AdventureWork2012DW(aquí se almacena las tablas del Data Warehouse).
- Archivos temporales compartidos en el drive del curso
- Ejemplo de Mapeo y Querys de las tablas dimensiones y Hecho

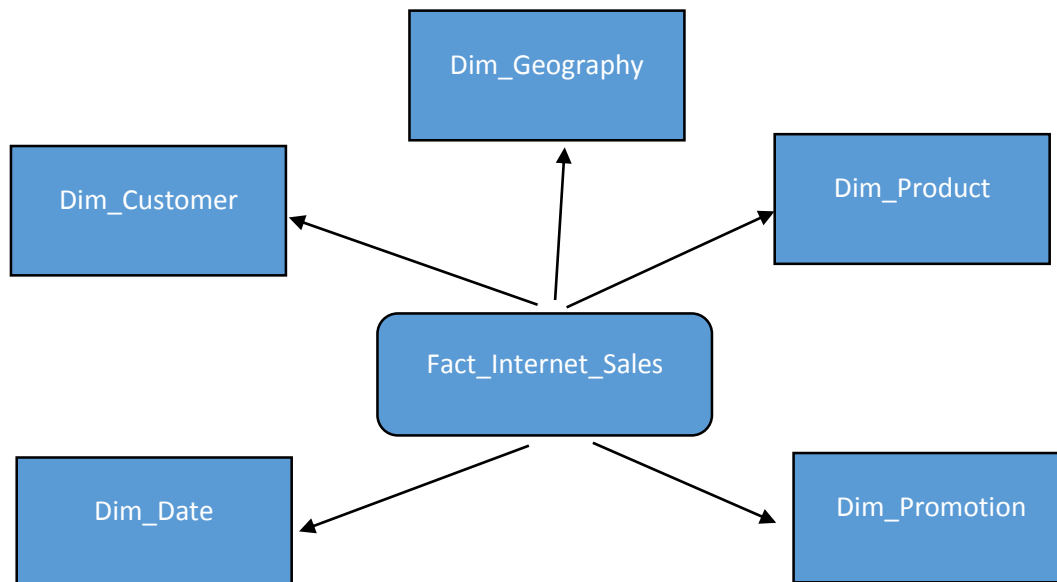


Fig1. Esquema Lógico – Escenario de Ventas por Internet Data Mart

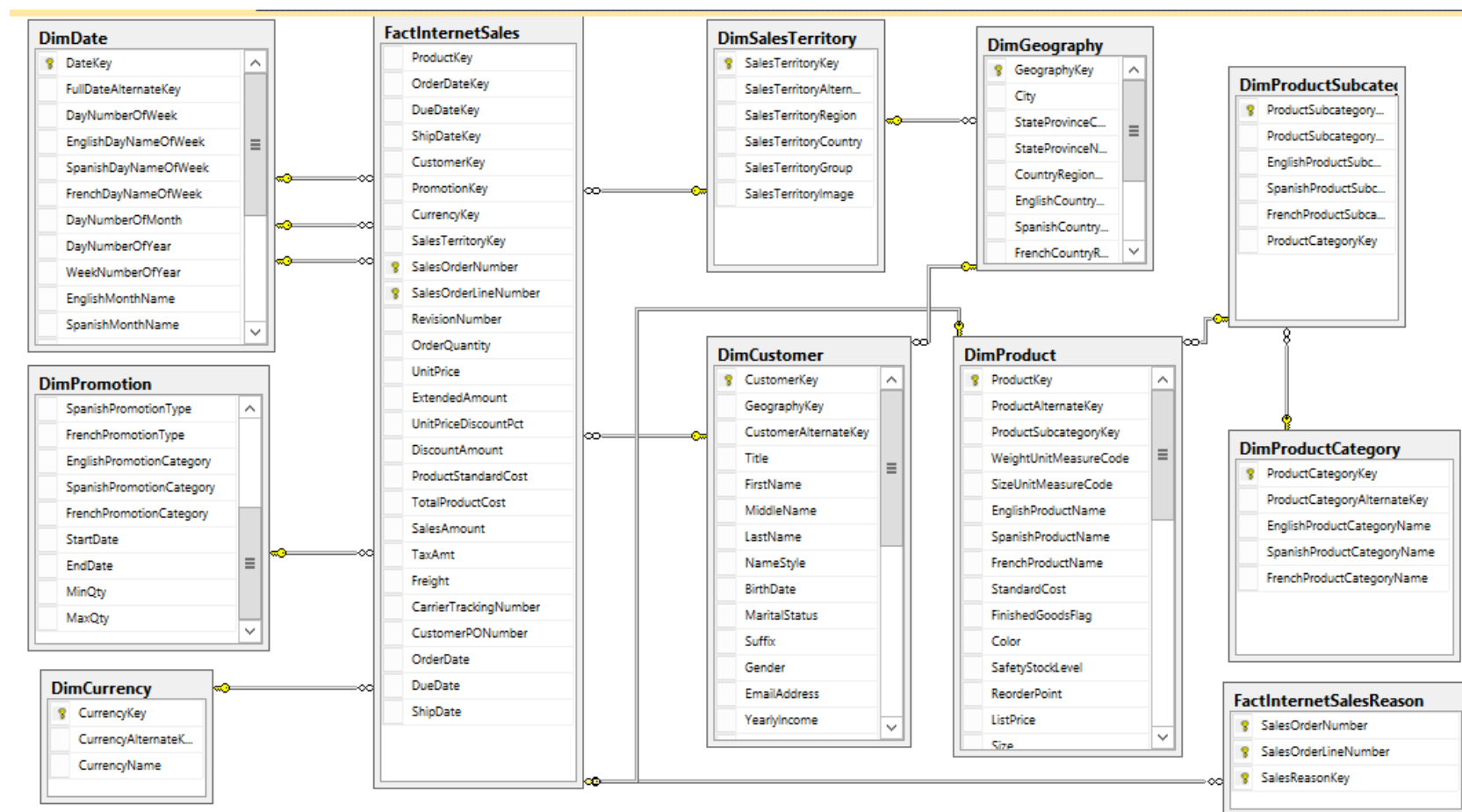


Figura 2. Ventas por Internet - Datamart - Esquema Físico

PASO 1:

INICIAMOS MAPEANDO LAS TABLAS DE DIMENSIONES:

PARA TODAS LAS TABLAS DE DIMENSIONES Y HECHO DE La figura 2, se debe realizar el mapeo, le sugiero abrir un Excel y copiar la metadata de cada tabla, Y buscar los campos correspondientes.

Ejemplo:

Data Source:

AdventureWorksDW2012

DimSalesTerritory

Data Source:

AdventureWorks2012

Column_name	Type	Computed	Length	Prec	Scale	Nullable	Tabla - System legacy	Campo
SalesTerritoryKey	int	no	4	10	0	no	clave surrogate	
SalesTerritoryAlternateKey	int	no	4	10	0	yes	[Sales].[SalesTerritory]	Territoryid
SalesTerritoryRegion	nvarchar	no	100			no	[Sales].[SalesTerritory]	Name
SalesTerritoryCountry	nvarchar	no	100			no	[Person].[CountryRegion]	Name
SalesTerritoryGroup	nvarchar	no	100			yes	[Sales].[SalesTerritory]	Group
SalesTerritoryImage	varbinary	no	-1			yes		

PASO 2: ARMADO DE UNA QUERY

Es necesario realizar alguna query para obtener los datos que se necesita para cargar la tabla de la dimensión respectiva.

Ejemplo: **DimSalesTerritory**

```
SELECT
a.TerritoryID AS SalesTerritoryAlternateKey,
a.Name As SalesTerritoryRegion,
b.Name As SalesTerritoryCountry,
a.[Group] As SalesTerritoryGroup,
null as SalesTerritoryImage

FROM [Sales].[SalesTerritory] a
JOIN [Person].[CountryRegion] b on a.CountryRegionCode=b.CountryRegionCode
UNION ALL
SELECT
0 AS SalesTerritoryAlternateKey,
'NA' As SalesTerritoryRegion,
'NA' As SalesTerritoryCountry,
'NA' As SalesTerritoryGroup,
null as SalesTerritoryImage
```

PASO 3: DESARROLLO EN SSIS

El siguiente diseño es para la carga de algunas dimensiones, y se ve según las dependencias en los datos serán procesados, usted debe de completar con las dimensiones faltantes.

Crear el proyecto en : C:\ETL ESIS\ con el nombre **ETL_AdventureWorkDW**, el primer paquete poner de nombre **Load_Datamart_InternetSales.dtsx**

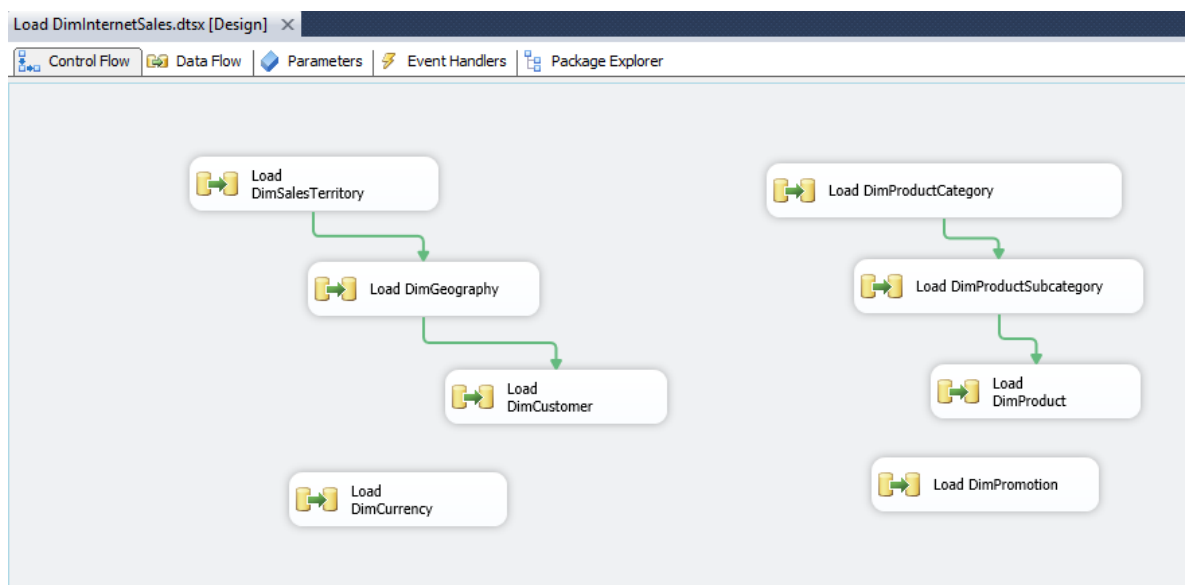


Figura 3. Diseño general del ETL

Agregar los DataFlow necesarios según la figura 3 para iniciar, entrar al data flow Load **DimSalesTerritory** y diseñar el siguiente flujo de tareas de la figura 4., lo haremos en clase.

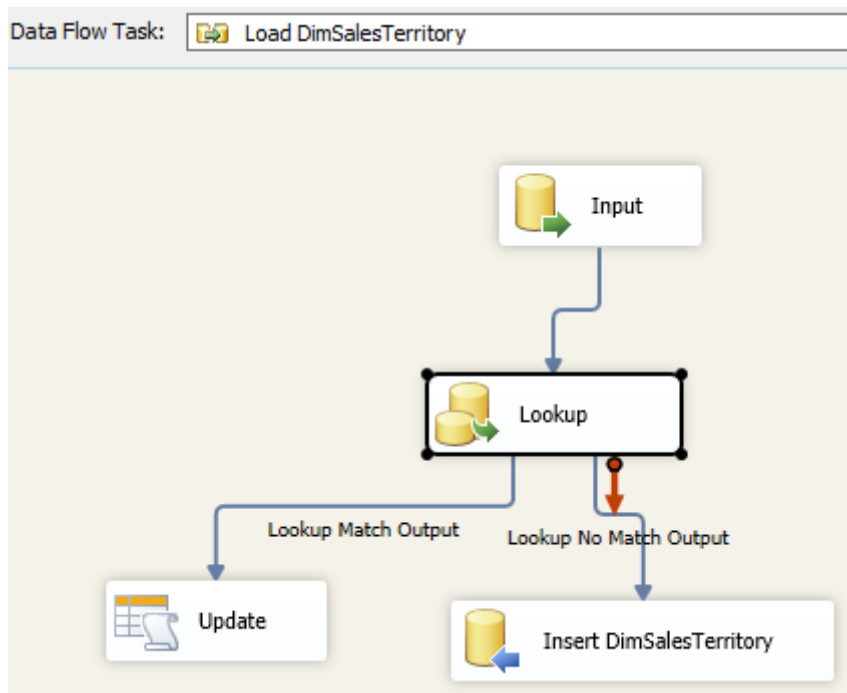


Figura 4: Diseño de Data Flow para DimSalesTerritory

Este es un pequeño ejemplo, en clase desarrollaremos otros casos que te servirán para el desarrollo de todo este mini proyecto.

Query input

```

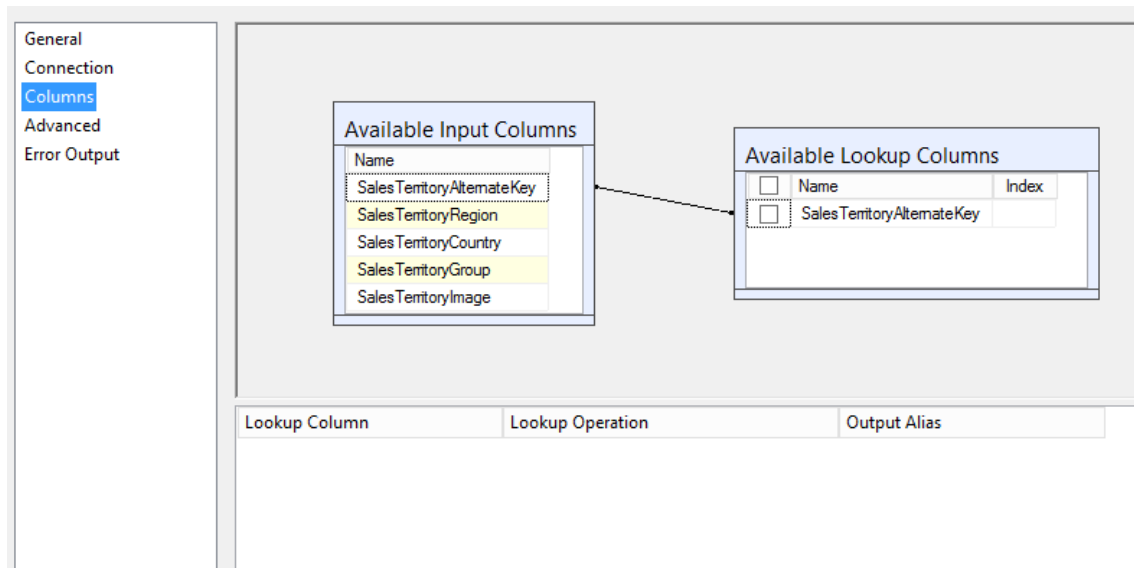
SELECT
a.TerritoryID AS SalesTerritoryAlternateKey,
a.Name As SalesTerritoryRegion,
b.Name As SalesTerritoryCountry,
a.[Group] As SalesTerritoryGroup,
null as SalesTerritoryImage

FROM [Sales].[SalesTerritory] a
JOIN [Person].[CountryRegion] b on a.CountryRegionCode=b.CountryRegionCode
UNION ALL
SELECT
0 AS SalesTerritoryAlternateKey,
'NA' As SalesTerritoryRegion,
'NA' As SalesTerritoryCountry,
'NA' As SalesTerritoryGroup,
null as SalesTerritoryImage
  
```

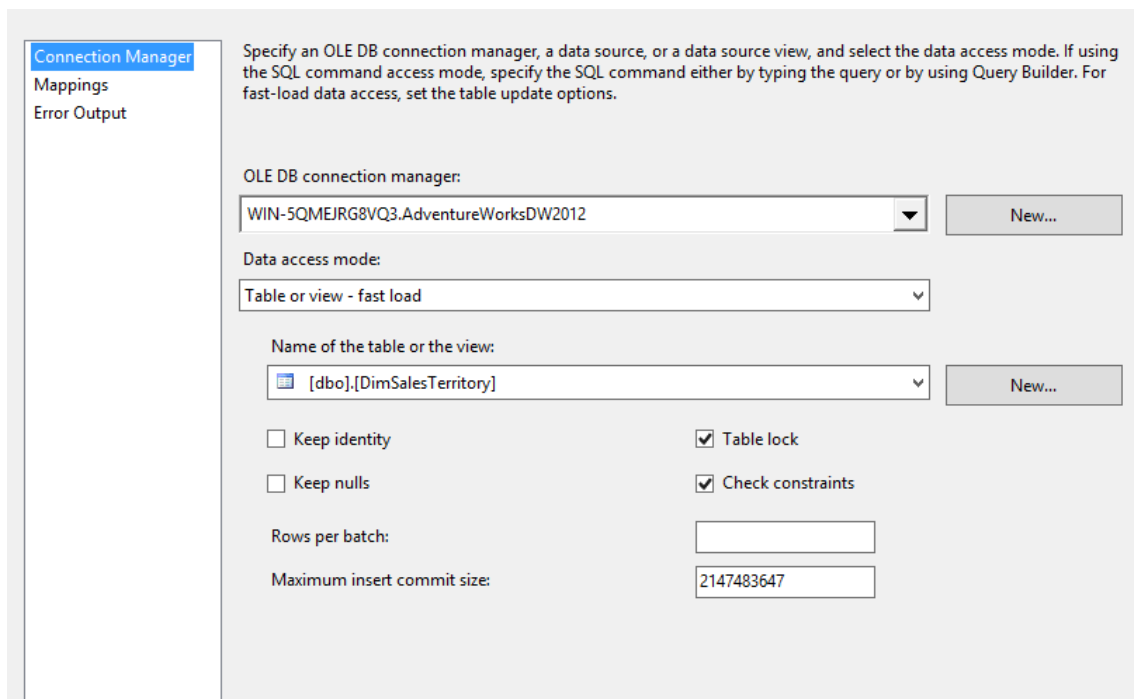
Para el Lookup – realizar la siguiente configuración

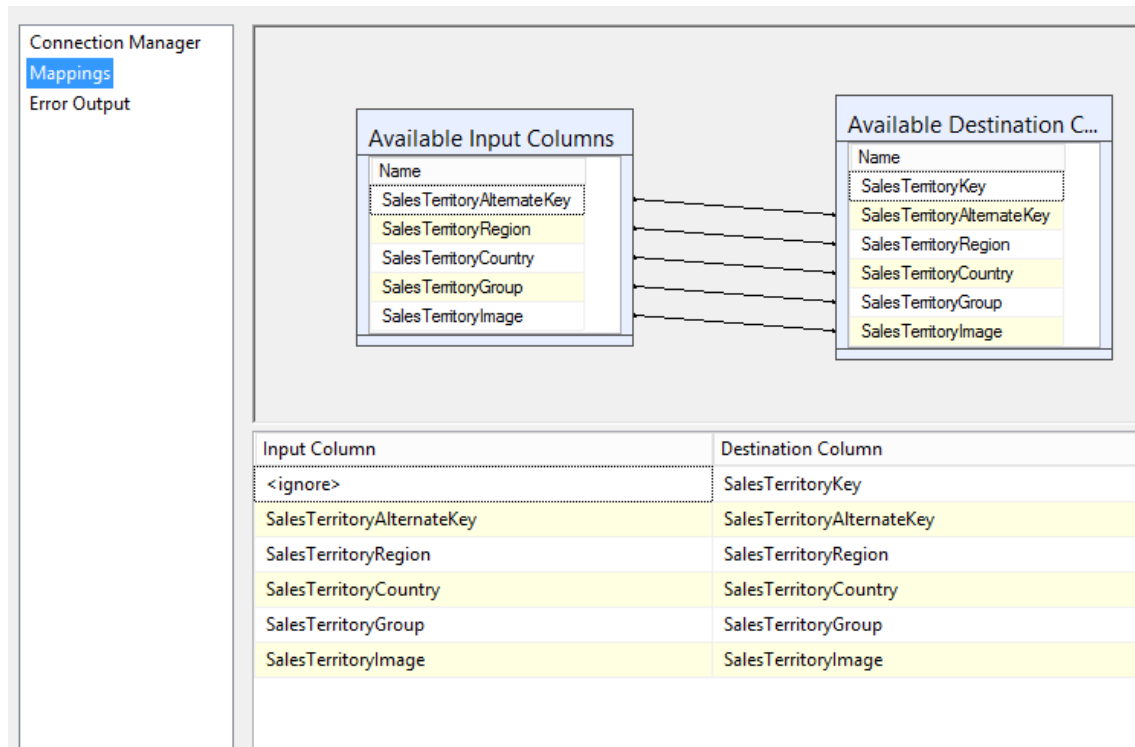
The screenshot shows the 'General' tab of the Lookup task configuration. On the left is a sidebar with tabs: General (selected), Connection, Columns, Advanced, and Error Output. The main area contains three sections: 'Cache mode' with radio buttons for 'Full cache' (selected), 'Partial cache', and 'No cache'; 'Connection type' with radio buttons for 'Cache connection manager' and 'OLE DB connection manager' (selected); and 'Specify how to handle rows with no matching entries' with a dropdown menu set to 'Redirect rows to no match output'.

The screenshot shows the 'Connection' tab of the Lookup task configuration. The sidebar on the left has 'Connection' selected. The main area has a heading 'Specify a data source to use. You can select a table in a data source view, a table in a database connection, or the results of an SQL query.' Below this are three options: 'OLE DB connection manager' (selected) with a dropdown showing 'WIN-SQMEJRG8VQ3.AdventureWorksDW2012' and a 'New...' button; 'Use a table or a view:' (unselected) with an empty dropdown and a 'New...' button; and 'Use results of an SQL query:' (selected) with a text box containing 'select SalesTerritoryAlternateKey from DimSalesTerritory' and buttons for 'Build Query...', 'Browse...', and 'Parse Query'.



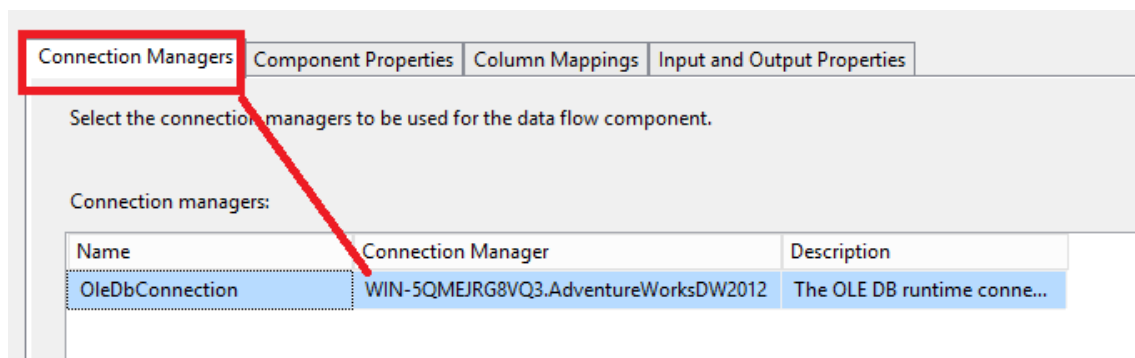
Carga a **DimSalesTerritory**, configuramos la conexión y la tabla donde se almacenará los datos



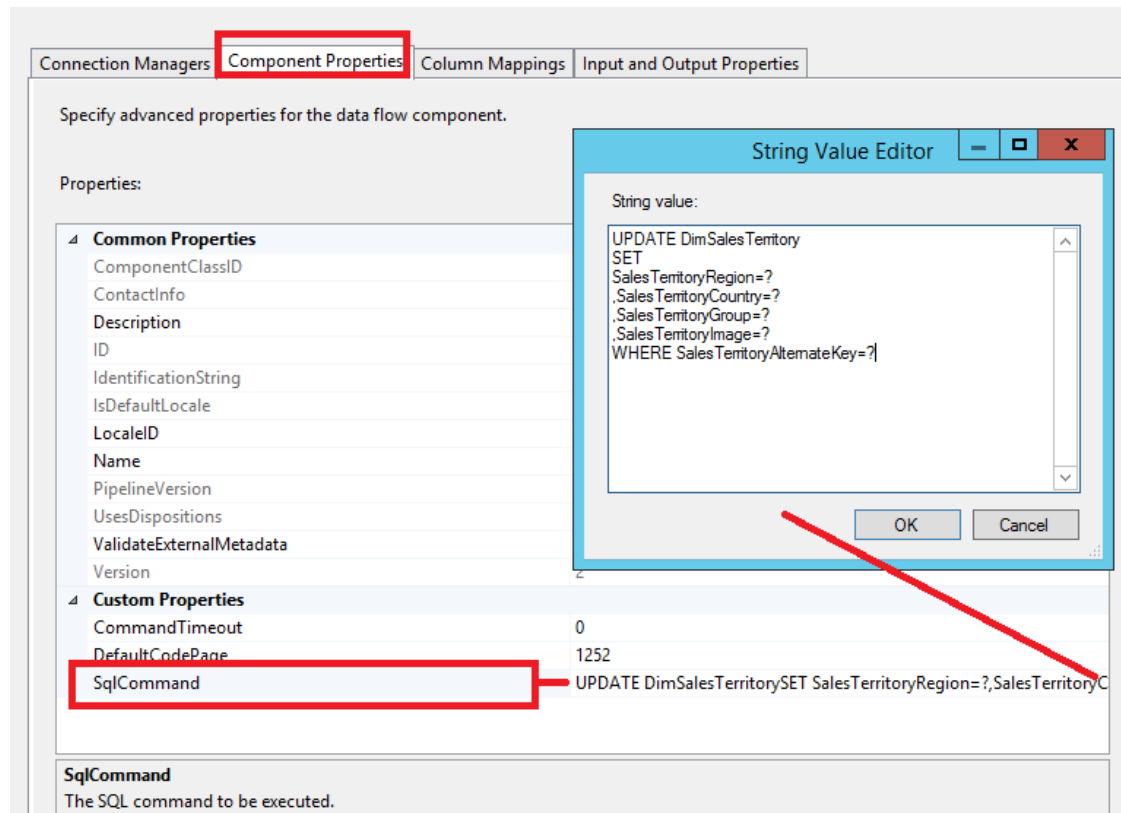


Para el control de Update(Comando OLE DB).

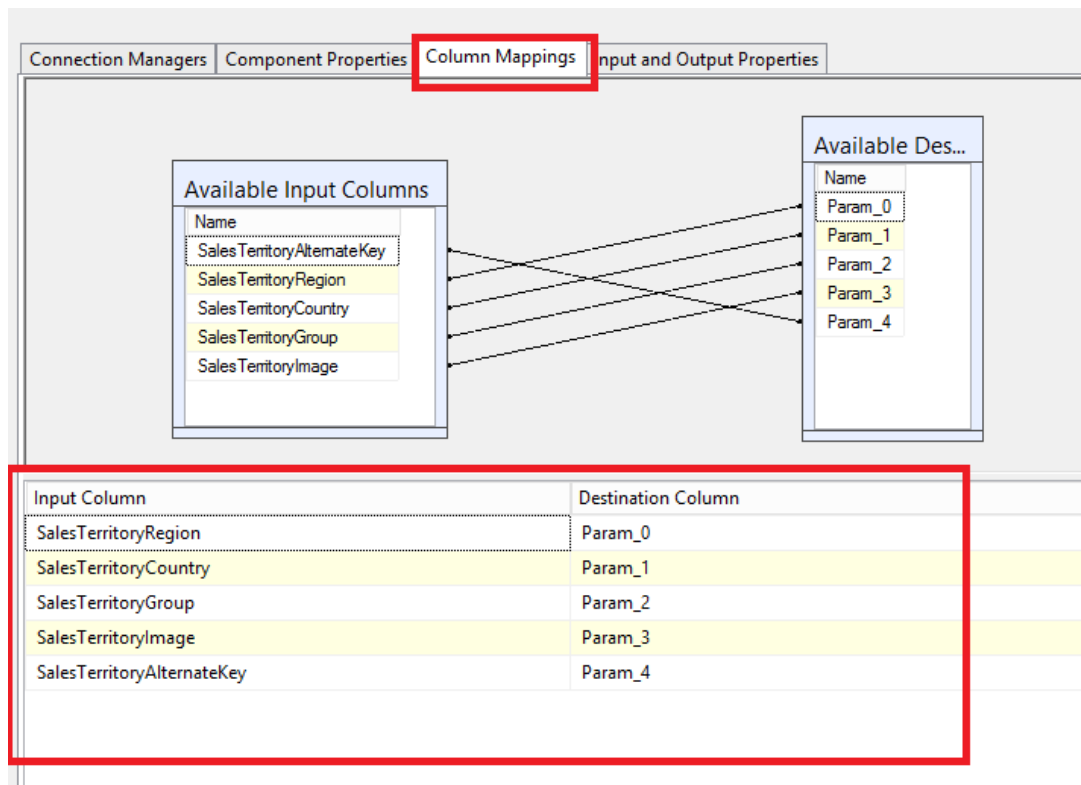
Primero definimos la conexión, que en este caso es donde almacena el DW.



Segunda pestaña, insertamos la query de update como se ve abajo. En la propiedad SqlCommand, y copiamos la query.



Mapeo de los datos de entrada en los parámetros creados por '?' en la query, que se crea de acuerdo el orden que se usa.



Para configurar los demás controles – ver el Proyecto desarrollado que está en el compartido, talvez tendrá que cambiar las conexiones del servidor.

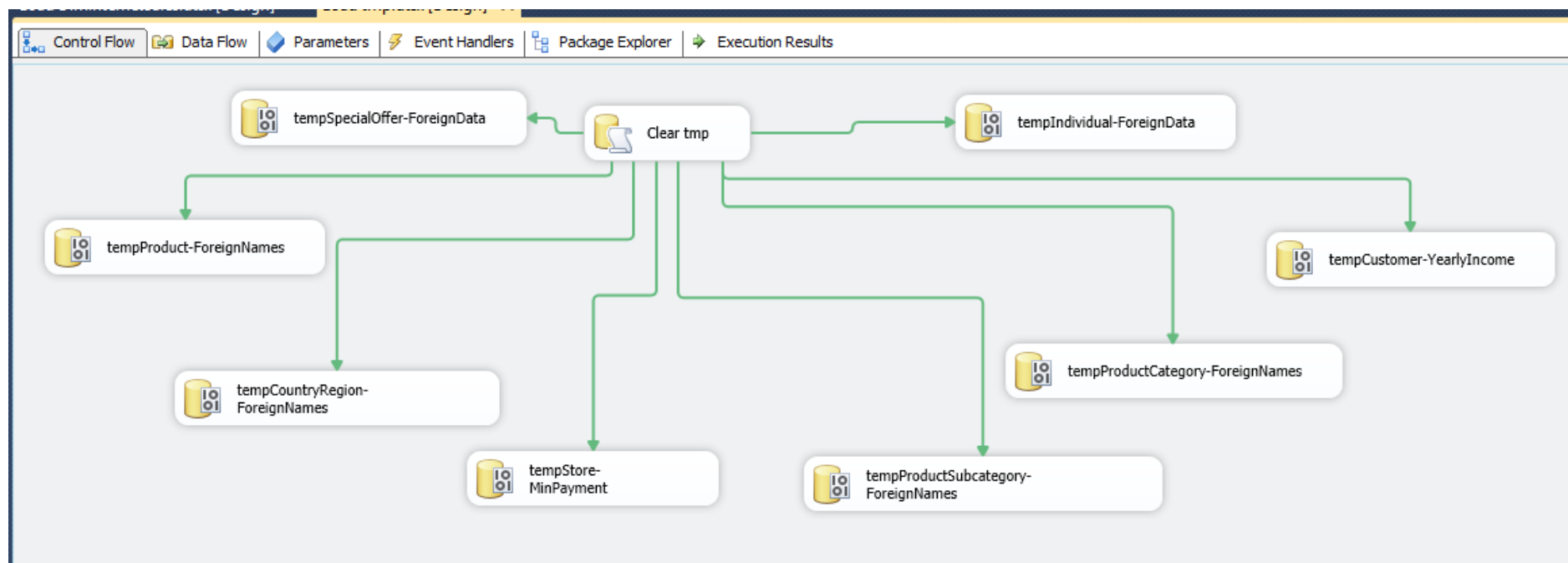


Figura 5. Carga de Archivos temporales - para Data mart – InternetSales (package - Load tmp.dtsx)

Este es la propuesta para el paquete que cargará los archivos temporales necesarios, que serán compartidos en el drive del curso, en la carpeta Data.

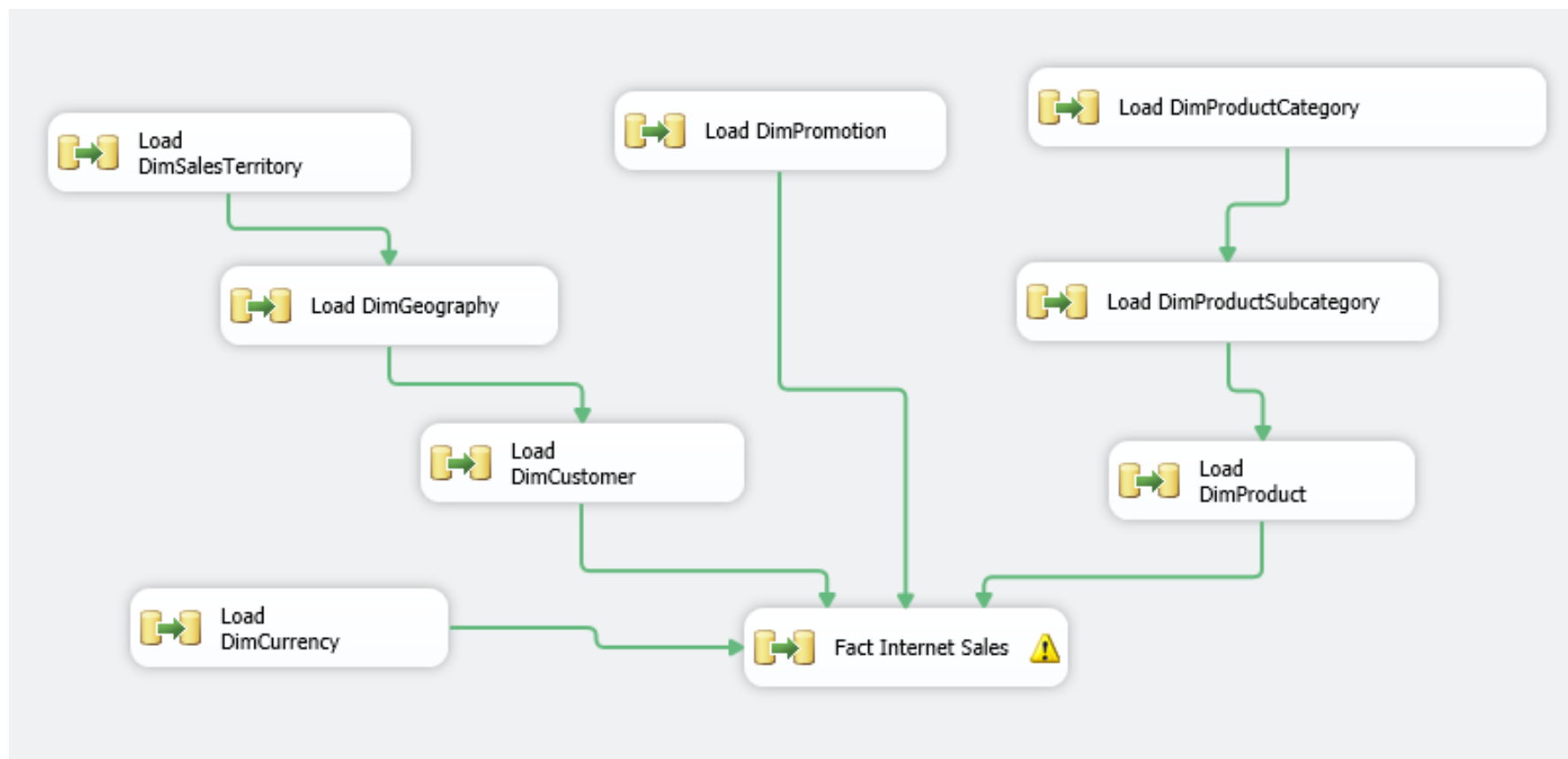


Figura 6. Diseño de Carga para Data mart – InternetSales (package - Load DimInternetSales.dtsx)

TRABAJO FINAL:

Es necesario grupos de 3 para desarrollar este trabajo.

Este trabajo debe ser completado para todas las dimensiones y la tabla de hecho que serán desarrolladas conjuntamente en clase como en sus casas.

El proyecto avanzado en Clase tendrá parte de las dimensiones desarrolladas como la tabla de Hechos.

PASOS DE SOLUCION:

1. MAPEO DE LOS DATOS EN LA BASE DE DATOS OLTP Y ARCHIVOS TEMPORALES
2. ARMAR UN EXCEL CON EL MAPEO Y QUERYS NECESARIAS
3. VER EL Ejemplo en el Drive del Curso: Mapping_DATAMART_INTERNETSALES.XLSX
4. DISEÑAR PAQUETES PARA LA CARGA DE LOS ARCHIVOS TEMPORALES Y OTRO PARA LA CAGA LAS TABLAS DE DIMENSIONES Y HECHOS.