

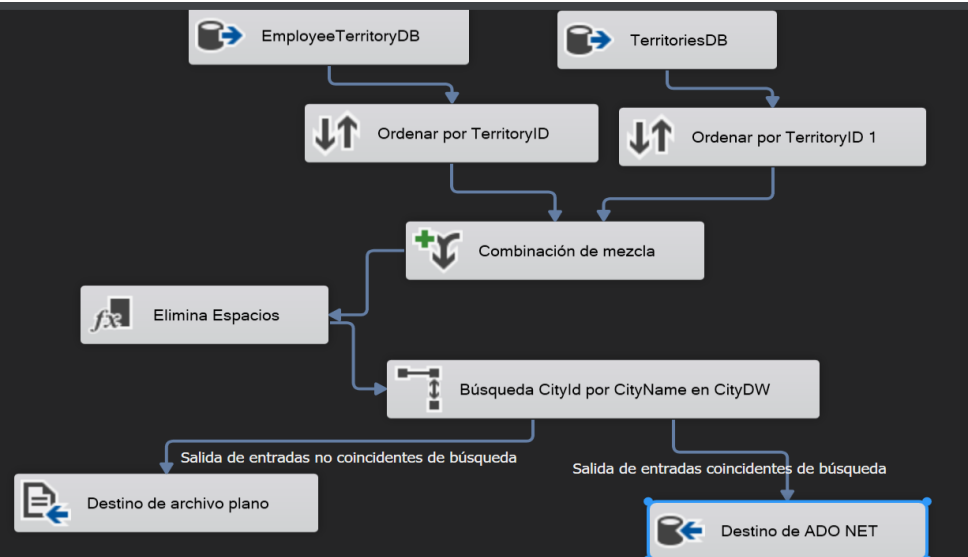
### **PRÁCTICA GUIADA 3. NORTHWIND ETL**

“Customer, Employee, Supplier, Shipper, Territories and Sales”

1. Cargar la tabla **Customer** en el almacén a partir de la tabla Customer de la BD.
  - Buscar la CitiKey que se guardará en la tabla Customer del almacén a partir del atributo CityName de la BD.
2. Cargar la tabla **Supplier** en el almacén a partir de la tabla **Supplier** de la BD.
  - Se hace exactamente igual que el apartado 1
3. Cargar la tabla **Employee** en el AD a partir de la tabla **Employee** de la BD.
  - Flujo de datos inmediato. Sólo DB origen ADO.NET y DW destino ADO.NET.
4. **Shipper**. Se hace exactamente igual que el apartado 3.
5. Cargar la tabla **Territories** en el almacén a partir de las tablas Territories y **EmployeeTerritory** de la base de datos.
  - Ordenar por TerritoryId ambos orígenes de datos
  - Enlazar con el operador combinación de mezcla (INNER JOIN) (seleccionar EmployeeId y TerritoryDescription (CityName))
  - Buscar la CityId en la tabla City del A.D a partir del mapeo de Territory description a CityName.
  - Enviar los territorios sin ciudad a un fichero de texto plano.
  - Enviar las salidas coincidentes al destino Territory del A.D. Mapear EmployeeId y CityKey.
  - Al ejecutar, os dará un error que podréis observar en el fichero de texto plano. Se observa que las ciudades en el origen de datos tienen espacios, por lo que no las encuentra en el almacén. Para corregirlo, añadimos un filtro de eliminación de espacios (antes del bloque de búsqueda). El bloque para aplicar transformaciones se llama *Columna derivada*. En el bloque de búsqueda, seleccionar la salida TrimTerritoryDescription del bloque de filtrado. Al ejecutar ahora se sigue observando un error de tuplas duplicadas. Para solucionarlo,

crear un bloque ordenar después de la búsqueda y seleccionar eliminar entradas repetidas con los atributos employeeId y cityKey.

Flujo Territories



Corrección de error de espacios en los nombres de ciudades

Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.

Variables y parámetros

Columns

REPLACE( «character\_expression», «search\_expression», «replace\_e» )

REPLICATE( «character\_expression», «times» )

REVERSE( «character\_expression» )

RIGHT( «character\_expression», «number» )

RTRIM( «character\_expression» )

SUBSTRING( «character\_expression», «start», «length» )

TOKEN( «character\_expression», «delimiter\_expression», «occurrence» )

TOKENCOUNT( «character\_expression», «delimiter\_expression» )

TRIM( «character\_expression» )

UPPER( «character\_expression» )

Funciones de fecha y hora

Descripción:

Devuelve una expresión de caracteres tras sustituir una cadena de caracteres de la expresión por otra distinta o por una vacía.

Derived Column Name	Derived Column	Expression	Data Type	Le
Trim Territory Descrip...	<add as new column>	TRIM(TerritoryDescription)	cadena Unicode [DT_...	50

Problema de inserción de tuplas duplicadas (intenta añadir la tupla 5,7526 dos veces)

Lookup Match Output Data Viewer at Carga Territorios

Detach Copy Data

Em...	TerritoryDescription	Trim TerritoryDescription	City...
3	Tampa	Tampa	7546
3	Savannah	Savannah	7542
4	Greensboro	Greensboro	7513
4	Rockville	Rockville	7536
4	Cary	Cary	7500
5	Edison	Edison	7507
5	New York	New York	7526
5	Morristown	Morristown	7524
5	Fairport	Fairport	7510
5	New York	New York	7526
5	Providence	Providence	7533
6	Redmond	Redmond	7535
6	Scottsdale	Scottsdale	7543
6	Seattle	Seattle	7544
6	Bellevue	Bellevue	7490
6	Phoenix	Phoenix	7530
7	Hoffman Estates	Hoffman Estates	7514

Attached Total rows: 0, buffers: 0 Rows displayed = 46

Sort Transformation Editor

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

Available Input Columns

<input type="checkbox"/>	Name	Pass Through
<input checked="" type="checkbox"/>	EmployeeID	<input type="checkbox"/>
<input type="checkbox"/>	TerritoryDescription	<input checked="" type="checkbox"/>
<input type="checkbox"/>	Trim TerritoryDescription	<input checked="" type="checkbox"/>
<input type="checkbox"/>	CityKey	<input checked="" type="checkbox"/>

Input Column	Output Alias	Sort Type	Sort Order	Comp.
EmployeeID	EmployeeID	ascending	1	

☒ Remove rows with duplicate sort values

OK Cancel Help

6. Para cargar la tabla **SALES (hecho)**, tenéis disponible la consulta en el CV.
- Crear un origen de datos (Sales en DW y DB). Los hechos se cargarán de la BD y el resto de tablas del AD.
  - Conectar con el almacén
  - Pegar consulta del CV
  - Ejecutar. Nos aparecerá un error ShippedDataKey = Null. Para depurarlo, creamos un bloque Split y nos quedamos con los j= NULL.