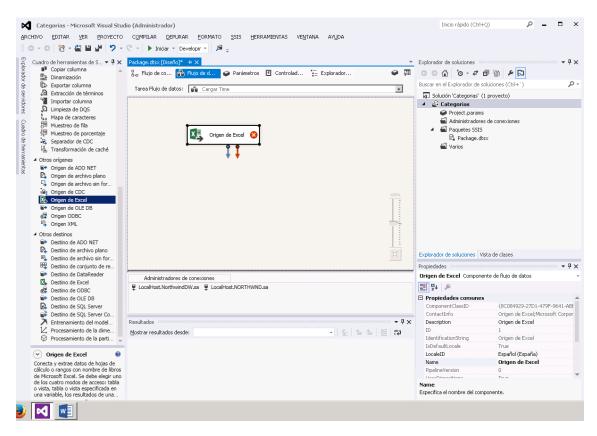


Práctica de Almacenes de Datos.

Ejemplo 3 de ETL con Integration Services: Tiempo y Geografía

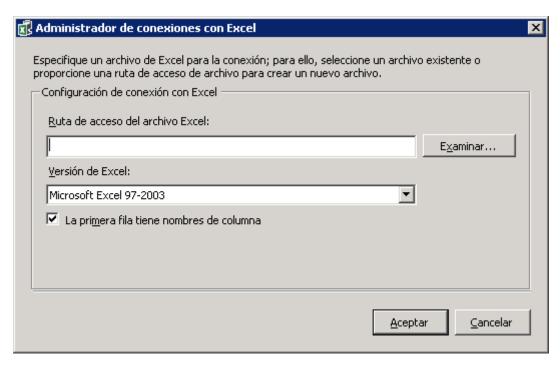
1. Carga de la dimensión Tiempo:

Creamos un nuevo proyecto de Integration Services pero ahora elegimos origen de datos, un Origen de Excel:



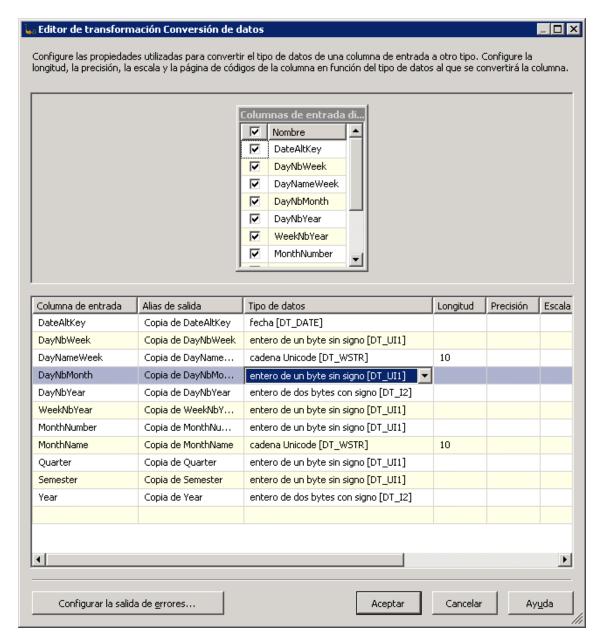
Hay que identificar dónde se encuentra el archivo, para eso se crea una nueva conexión de Excel:





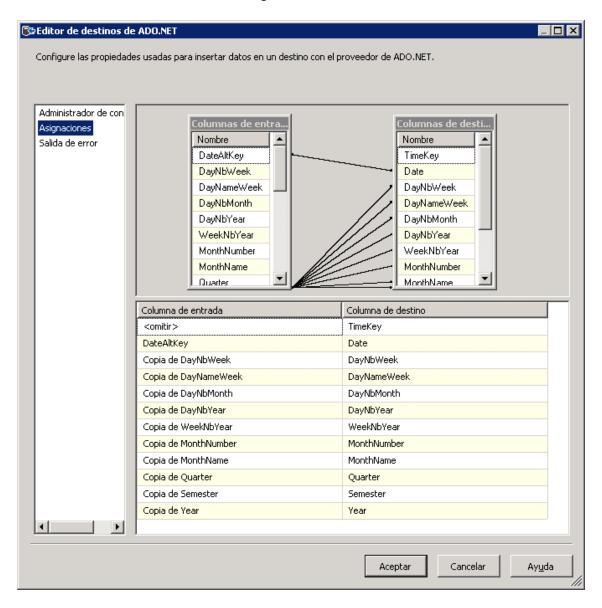
La conversión de datos ya no es tan directa como cuando se transfieren datos entre tablas:





La columna TimeKey dentro de la tabla Time se ha definido con la propiedad IDENTITY (1,1) por lo que no es necesario asignarle valores



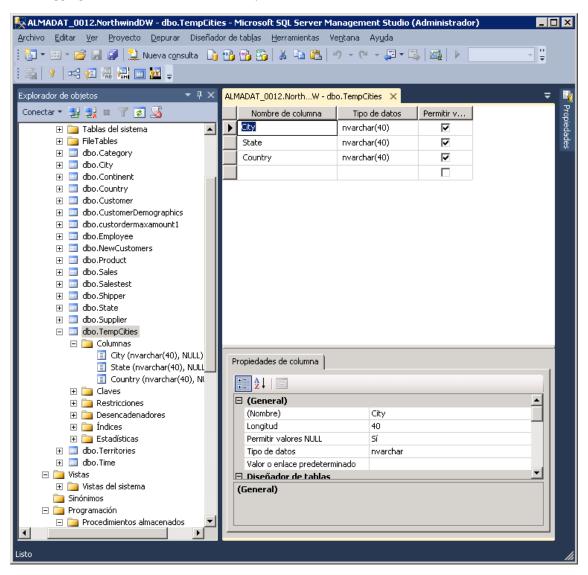




Carga de la Dimensión Geografía

2. Carga de Ciudades a tabla intermedia:

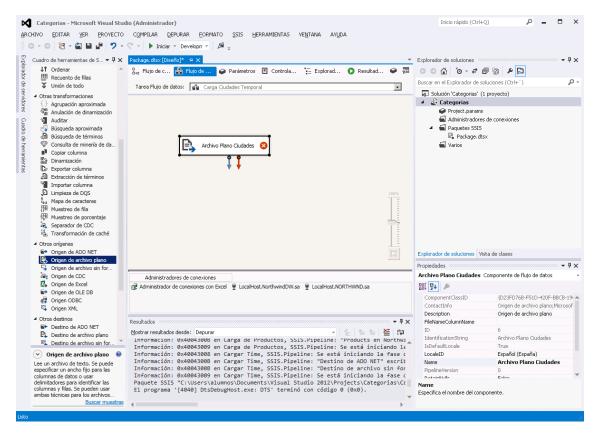
Disponemos de un fichero plano de ciudades. Pero para cargar la jerarquía Geografía haremos un "stagging" en la tabla intermedia TempCities:





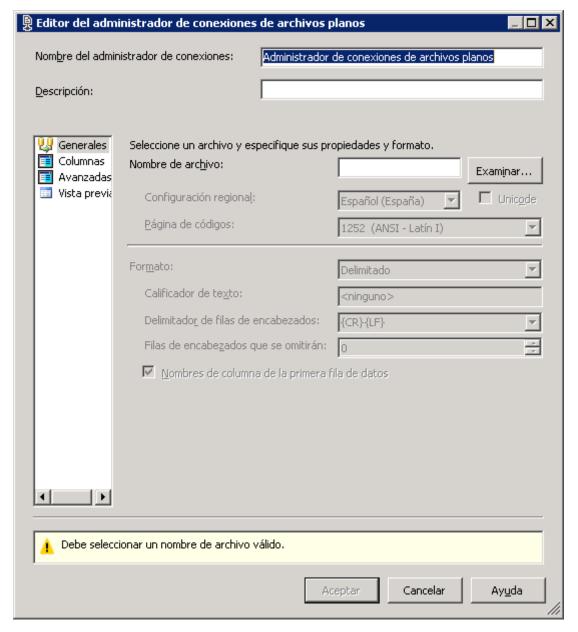






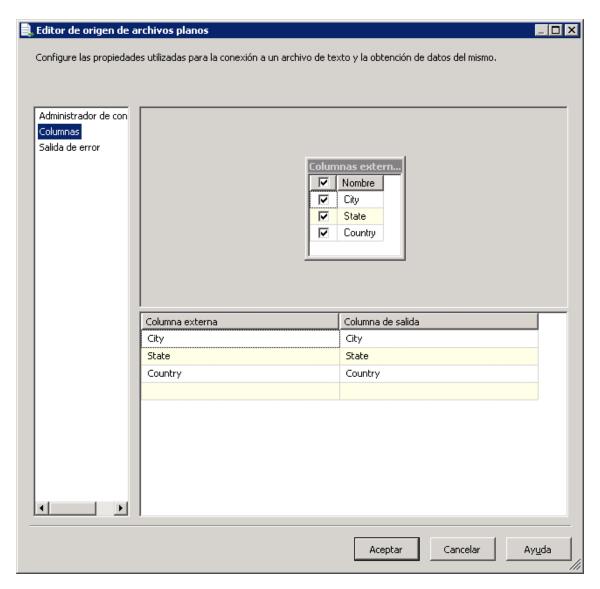
Hay que crear una conexión para archivos planos:



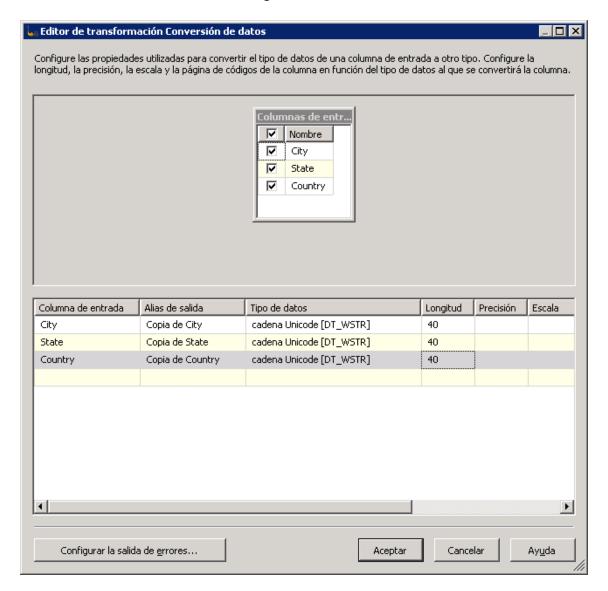


Se selecciona el archivo cities.txt y por defecto toma la primera línea como nombre de las columnas:

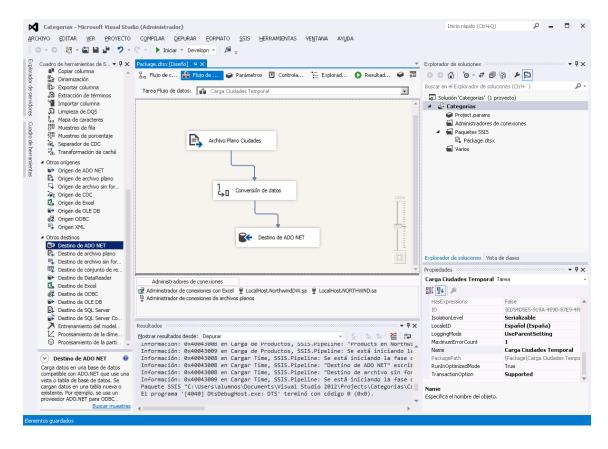






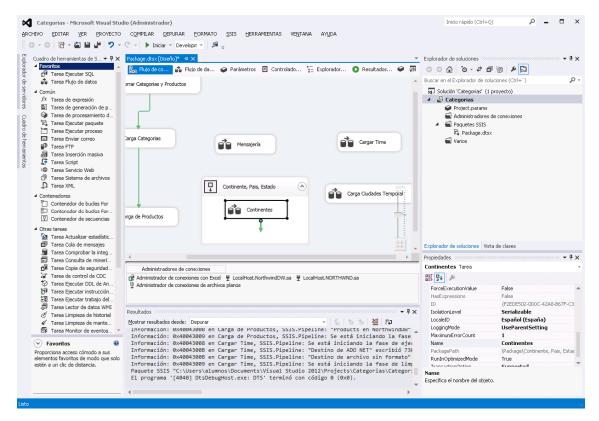








Para la creación de la jerarquía, tenemos que cargar las tablas: Continente, País, Estado. Para ello creamos un contenedor de secuencias:



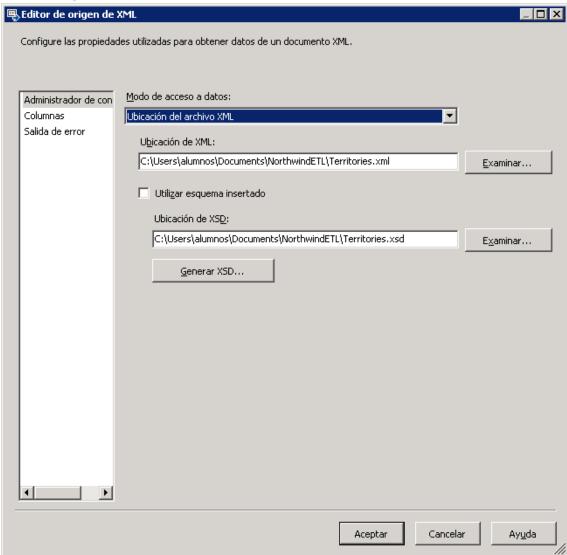
Los continentes, países y estados residen en un archivo XML, Territories.xml:



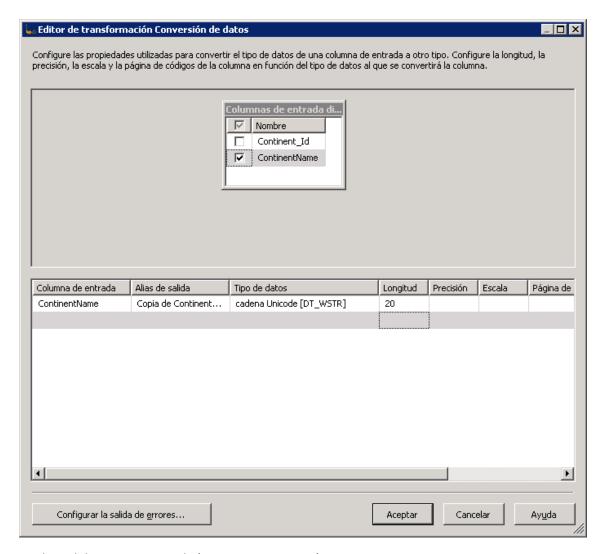
```
📓 C:\Users\esc\Documents\Docencia\Almacenes de Datos\DataWarehouseSystems Book\Bases de datos\NorthwindETL_files\Northwi... 🔠
<u>File Edit Search View Encoding Language Settings Macro Run Plugins Window ?</u>
 <?xml version="1.0" encoding="ISO-8859-1"?>
        <Continents:
         <Continent>
            <ContinentName>Europe</ContinentName>
  5
            <Country>
              <CountryName>Austria</CountryName>
              <CountryCode>AT</CountryCode>
              <CountryCapital>Vienna</CountryCapital>
  8
              <Population>8316487</Population>
  10
              <Subdivision>Austria is divided into nine Bundesländer, or simply Länder (states; sing. Land)
  11
  12
                <StateName>Burgenland</StateName>
  13
                <StateCode>BU</StateCode>
  14
                <StateCapital>Eisenstadt</StateCapital>
              </State>
  15
  16
              <State type="state">
  17
                <\!\!\texttt{StateName}\!\!>\!\! \textbf{K\"{arnten}}\!\!<\!\!/\!\!\texttt{StateName}\!\!>\!
  18
                19
                <StateCode>KA</StateCode>
                <StateCapital>Klagenfurt</StateCapital>
 21
 22
              <State type="state">
                <StateName>Niederösterreich</StateName>
                <EnglishStateName>Lower Austria</EnglishStateName>
 24
                <StateCode>NO</StateCode>
 25
 26
                <\!\!\texttt{StateCapital}\!\!>\!\!\textbf{St. P\"olten}\!\!<\!\!/\!\!\texttt{StateCapital}\!\!>\!\!
 27
              </State>
 28 🖨
              <State type="state">
 29
                <StateName>Oberösterreich</StateName>
 30
                <EnglishStateName>Upper Austria</EnglishStateName>
                <StateCode>00</StateCode>
 31
                                                                             Dos\Windows
                                                                                            ISO 8859-1
eXtensible Markup Langu length : 217468 lines : 5931
                                           Ln:1 Col:1 Sel:0|0
```



3. Carga de los Continentes

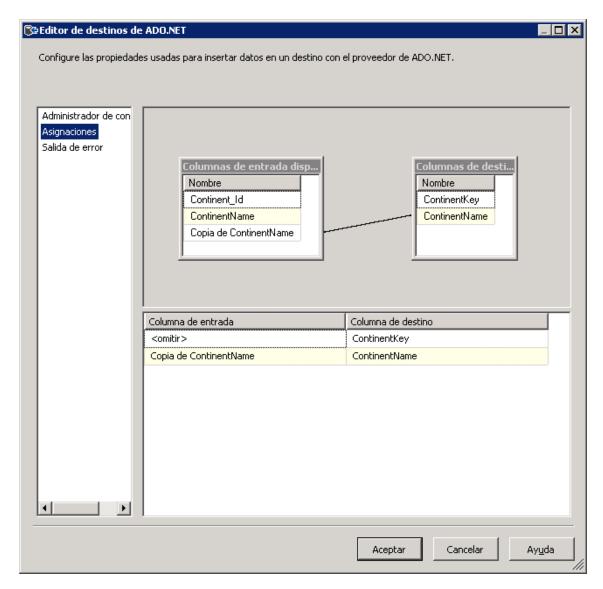






La clave del continente también se genera automáticamente.

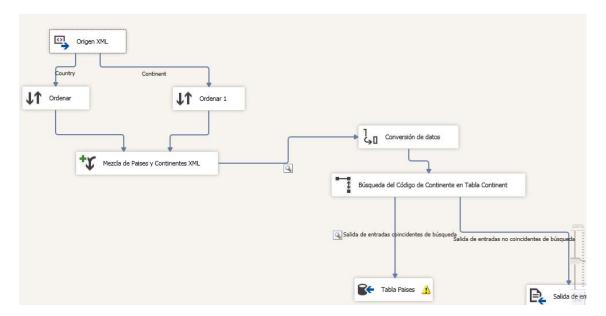




4. Países

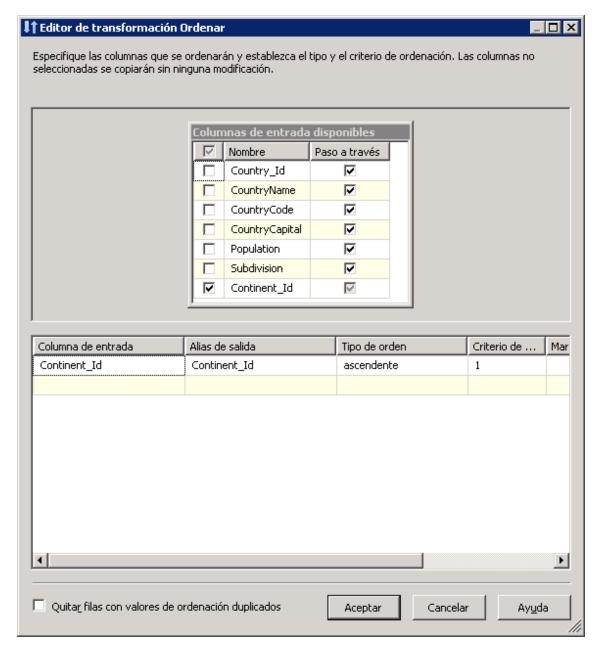
Para los países hay que realizar algunos pasos más:





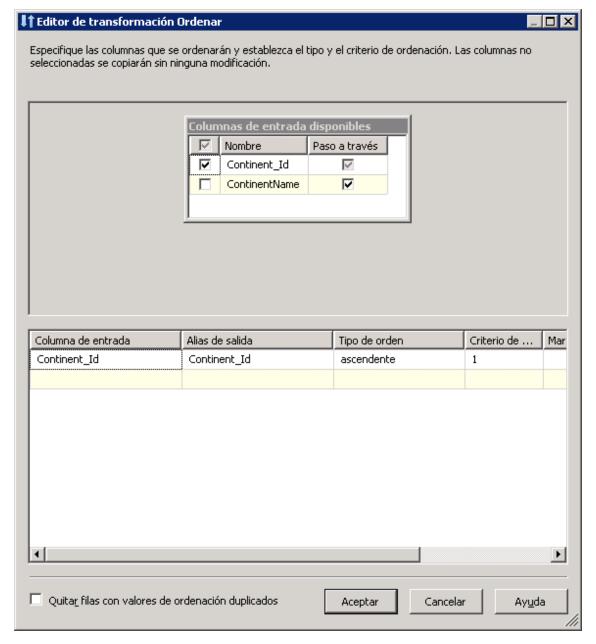
Ordenamos las salidas del XML por continent_id para poder encontrar el continente al que pertenece cada país :





Hacemos lo mismo con los continentes:





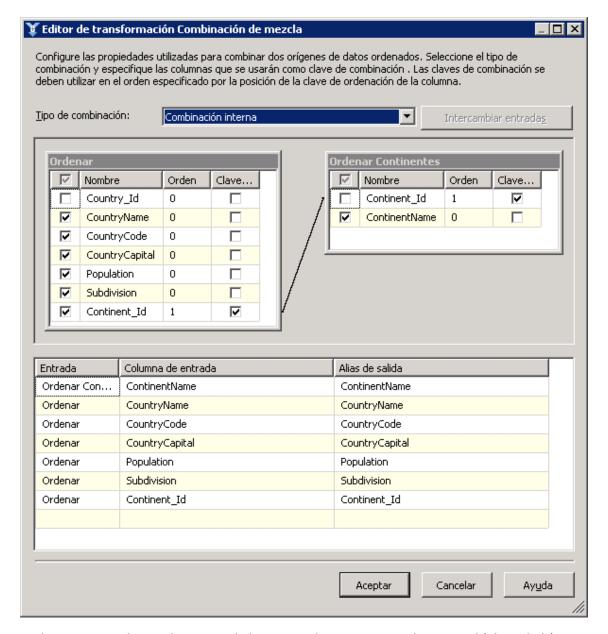
Ahora utilizamos el Merge Join o Combinación de mezcla para poder obtener el nombre del continente de cada país:





La mezcla queda así:

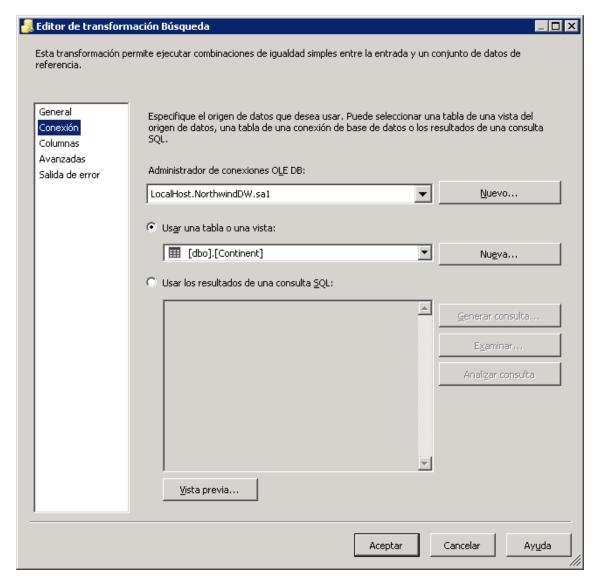




Es decir, estamos haciendo un Join de las 2 entradas que tengan el Countryld (el que había en el XML) iguales para obtener el CountryName.

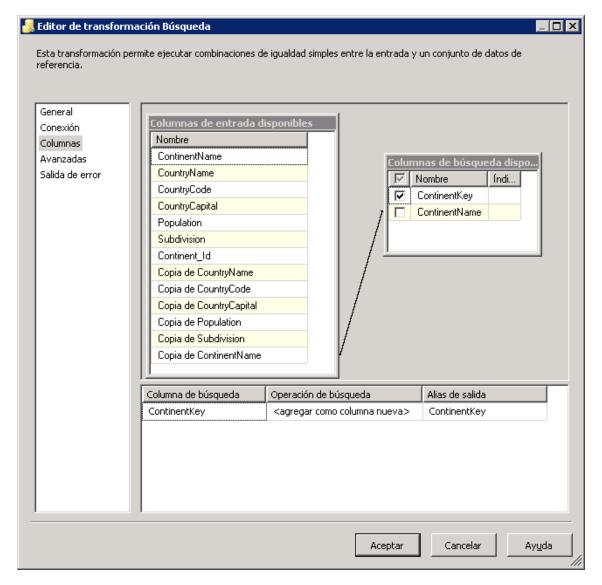
Una vez que se tienen el nombre del continente, hay que buscar el código que tienen el Continente en el Almacén de Datos (que no es el mismo probablemente que el que hay en el XML). Esto se hace así:





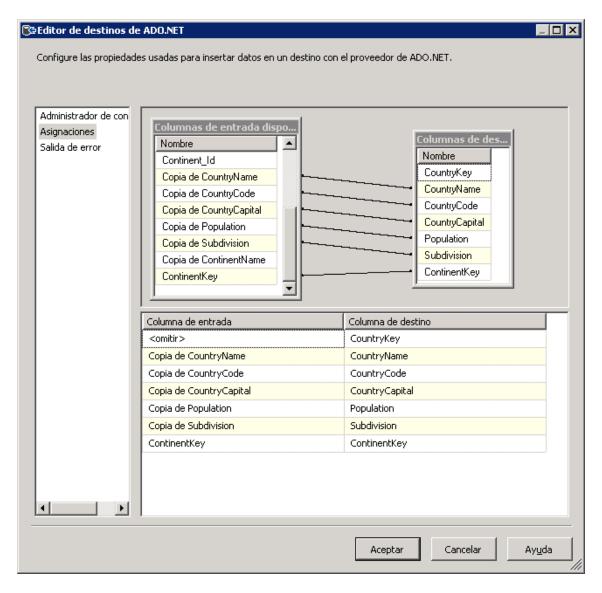
Se elige la tabla en la que se va a buscar el código y se elige la forma de buscarlo y el atributo a obtener:





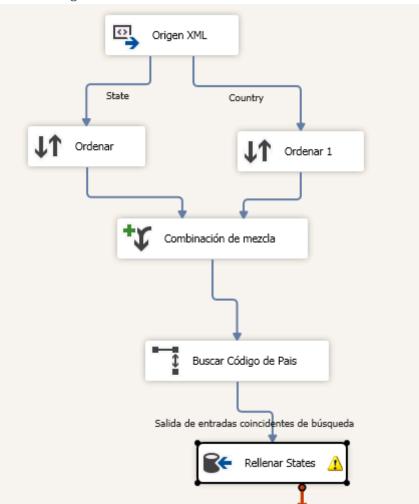
Una vez se tienen los datos, se guardan en la tabla Country:





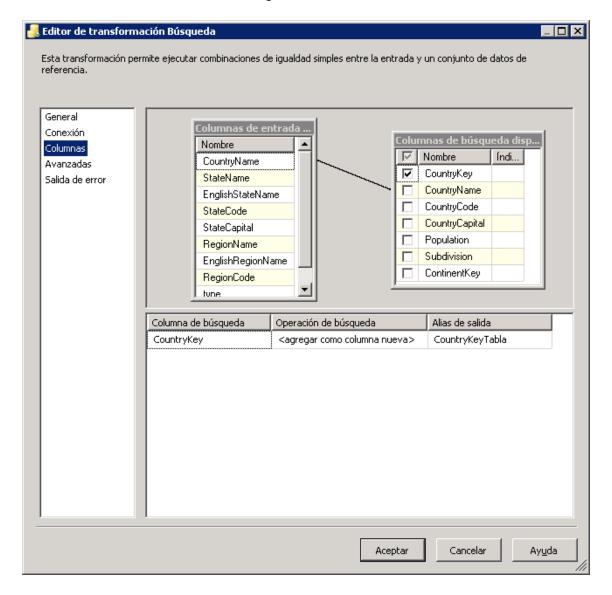


5. Carga de Estados



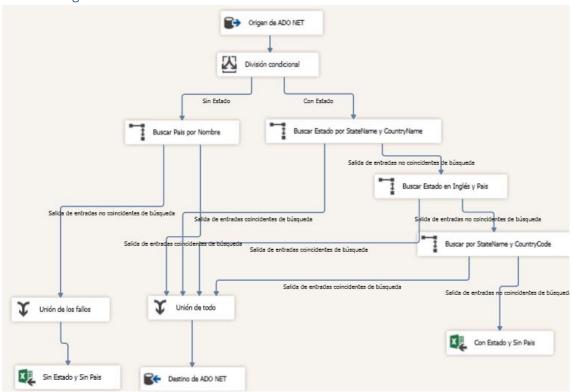
Del archivo XML elegimos State y Country que ordenamos por Countryld y mezclamos para obtener cada estado con su nombre de país. Para encontrar la clave del País en el almacén hacemos una búsqueda.





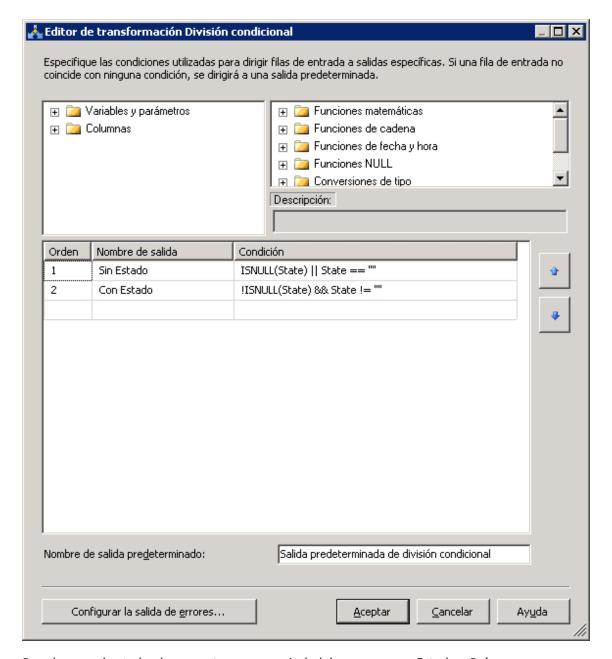


6. Carga de Ciudades



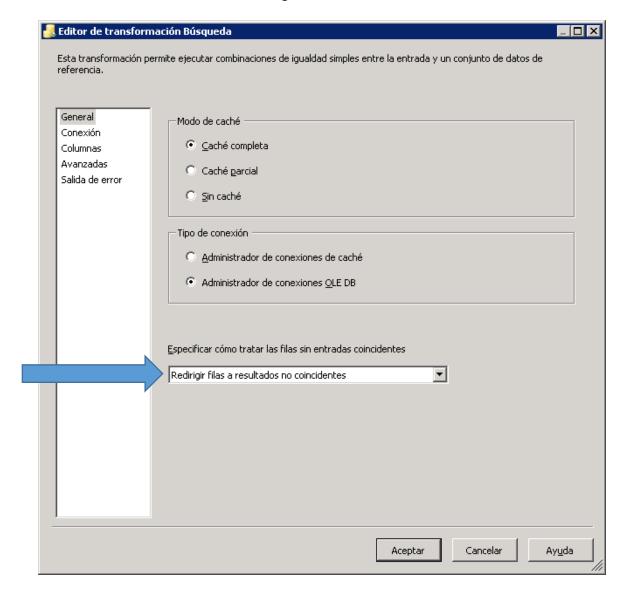
Lo primero que hacemos el elegir la fuente dee datos, TempCities del Almacen. Lo siguiente es diferenciar entre las ciudades que no tienen Estado (como Singapour) de las que sí:



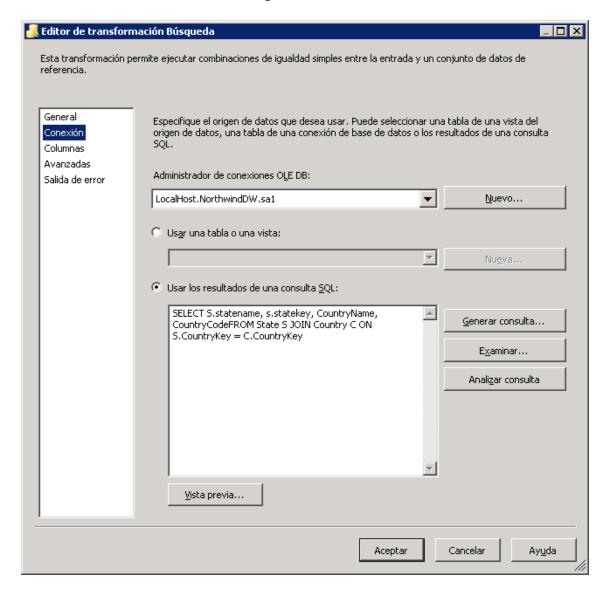


Para buscar el estado al que pertenece una ciudad, buscamos por Estado y País:

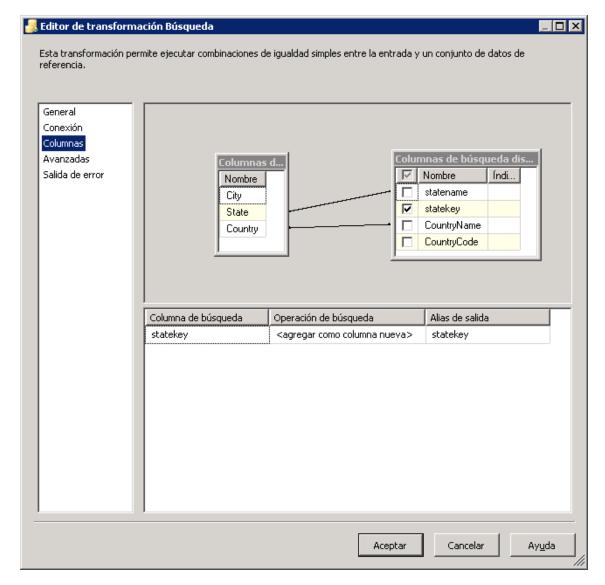








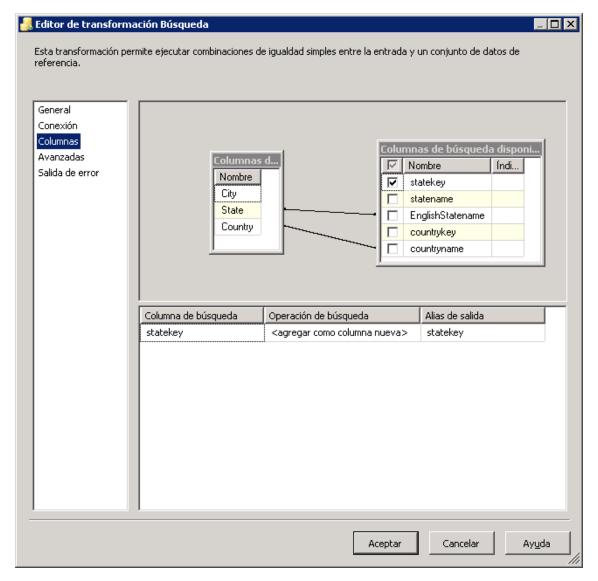




De esta forma hemos obtenido el StateKey del Almacen.

Si no se encuentra, se busca por el nombre del Estado en inglés:





Y por último, teniendo en cuenta el código del país en lugar del nombre (esto es así porque en las ciudades aparece USA en lugar de United States)



