



ugr

Universidad
de Granada

Sistemas Multidimensionales. **Práctica 4:**

Herramientas ETL

SSIS (SQL Server Integration Services)

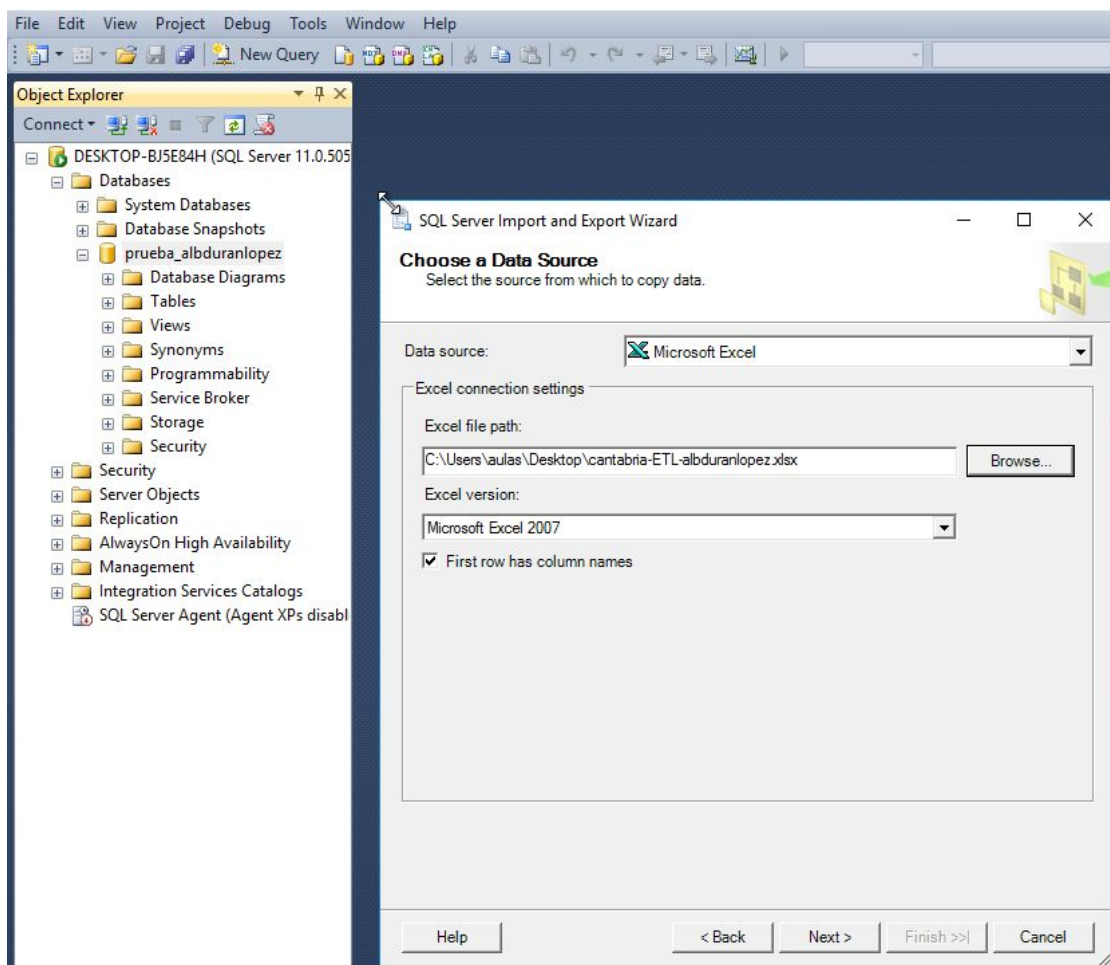
Alberto Jesús Durán López

Doble grado Ingeniería Informática y Matemáticas

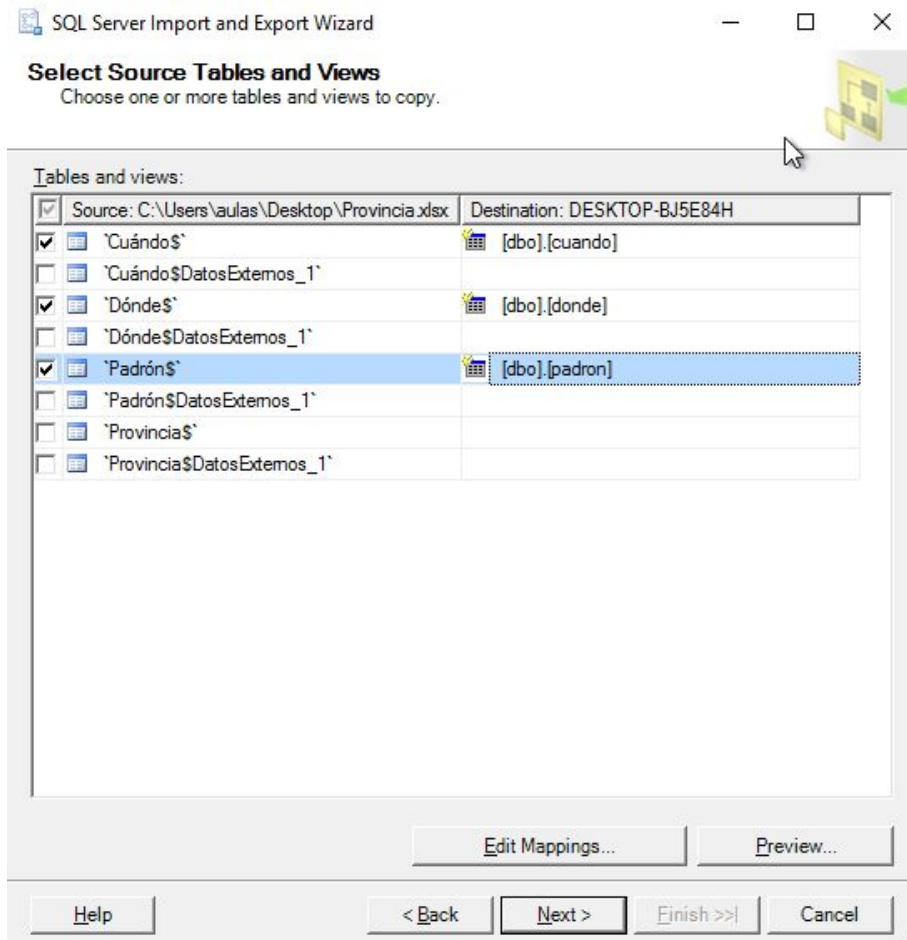
Curso 2019-2020

1. **Crea una BD SQL Server cuyo nombre sea prueba_albduranlopez. Importa las hojas Cuándo, Dónde y Padrón del archivo obtenido con *Power Query* de la práctica anterior**

Se puede ver en la columna de la izquierda que la base de datos **prueba_albduranlopez** está creada. Una vez hecho esto, importamos los datos obtenidos de *Power Query*:



Las tablas a exportar (al igual que en la práctica anterior) son *Cuándo*, *dónde* y *padrón*. Marcamos el tick izquierdo como aparece en la siguiente captura y cambiamos los nombres de las tablas al que queramos



A continuación, mostramos la configuración de las tres tablas. Cabe destacar que el estilo usado es **camel_case** y que la opción *<<Drop and re-create destination table>>* está marcada ya que he realizado numerosas pruebas de creado y borrado de tablas para familiarizarme con la herramienta.

Column Mappings

Source: 'Cuándo\$'
Destination: [dbo].[cuando]

☒ Create destination table Edit SQL...
☐ Delete rows in destination table ☒ Drop and re-create destination table
☐ Append rows to the destination table ☐ Enable identity insert

Mappings:

Source	Destination	Type	Nullable	Size	Precision	Scale
IdCuando	id_cuando	float	<input type="checkbox"/>			
Año	año	nvarchar	<input checked="" type="checkbox"/>	255		
Década	decada	nvarchar	<input checked="" type="checkbox"/>	255		

Source column: IdCuando Double (15)

OK Cancel

Column Mappings

Source: 'Dónde\$'
Destination: [dbo].[donde]

☒ Create destination table Edit SQL...
☐ Delete rows in destination table ☒ Drop and re-create destination table
☐ Append rows to the destination table ☐ Enable identity insert

Mappings:

Source	Destination	Type	Nullable	Size	Precision	Scale
IdDonde	id_donde	float	<input type="checkbox"/>			
CodMunicipio	cod_municipio	nvarchar	<input checked="" type="checkbox"/>	255		
Municipio	municipio	nvarchar	<input checked="" type="checkbox"/>	255		
CapitalMunicipio	capital_municipio	nvarchar	<input checked="" type="checkbox"/>	255		
CodComarca	cod_comarca	float	<input checked="" type="checkbox"/>			
Comarca	comarca	nvarchar	<input checked="" type="checkbox"/>	255		
CodProvincia	cod_provincia	nvarchar	<input checked="" type="checkbox"/>	255		
Provincia	provincia	nvarchar	<input checked="" type="checkbox"/>	255		
CapitalProvincia	capital_provincia	nvarchar	<input checked="" type="checkbox"/>	255		

Source column: IdDonde Double (15)

OK Cancel

Column Mappings

Source: `Padrón\$`

Destination: [dbo].[padron]

☒ Create destination table

☐ Delete rows in destination table ☒ Drop and re-create destination table

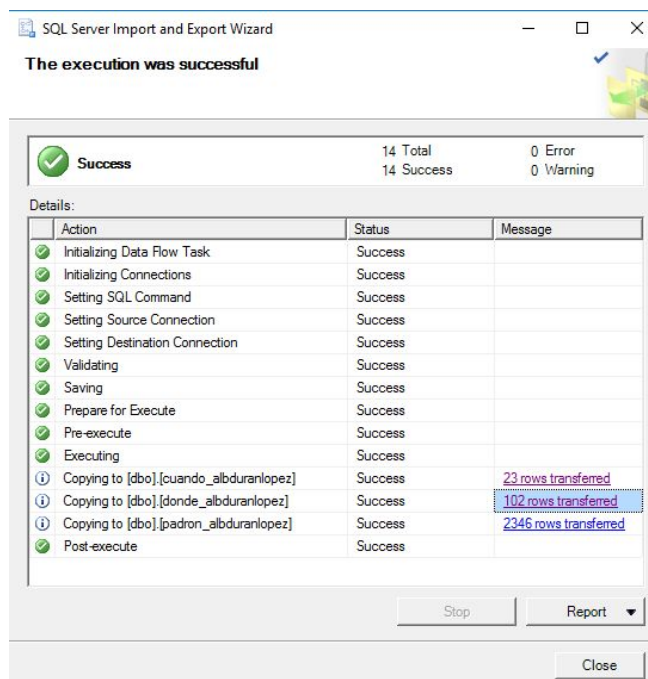
☐ Append rows to the destination table ☐ Enable identity insert

Mappings:

Source	Destination	Type	Nullable	Size	Precision	Scale
IdDonde	id_donde	float	<input type="checkbox"/>			
IdCuando	id_cuando	float	<input type="checkbox"/>			
Hombres	hombres	float	<input checked="" type="checkbox"/>			
Mujeres	mujeres	float	<input checked="" type="checkbox"/>			
Superficie	superficie	float	<input checked="" type="checkbox"/>			

Source column: IdDonde Double (15)

Al importar los datos con esta herramienta, las tablas se crean automáticamente (como se puede ver a continuación). Recalco que el nombre lo he modificado añadiendo _albduranlopez a las tablas para practicar su uso.



Resultados visibles:

Object Explorer: DESKTOP-BJ5E84H (SQL Server 11.0.5058)

- Databases
 - System Databases
 - Database Snapshots
 - prueba_albduranlopez
 - Database Diagrams
 - Tables
 - System Tables
 - FileTables
 - dbo.cuando_albduranlopez
 - dbo.donde_albduranlopez
 - dbo.padron_albduranlopez
 - Views
 - Synonyms
 - Programmability
 - Service Broker
 - Storage
 - Security
 - Server Objects
 - Replication
 - AlwaysOn High Availability
 - Management
 - Integration Services Catalogs
 - SQL Server Agent (Agent XPs disabled)

SQLQuery1.sql - DE...BJ5E84H\aulas (52)

```
/****** Script for SelectTopNRows command from SSMS *****/
SELECT TOP 1000 [id_donde]
,[cod_municipio]
,[municipio]
,[capital_municipio]
,[cod_comarca]
,[comarca]
,[cod_provincia]
,[provincia]
,[capital_provincia]
,[cod_ca]
,[ca]
,[cod_tam_municipio]
,[tamaño_habitantes]
,[altitud]
,[altitud_m]
,[superficie]
,[superficie_km2]
FROM [prueba_albduranlopez].[dbo].[donde_albduranlopez]
```

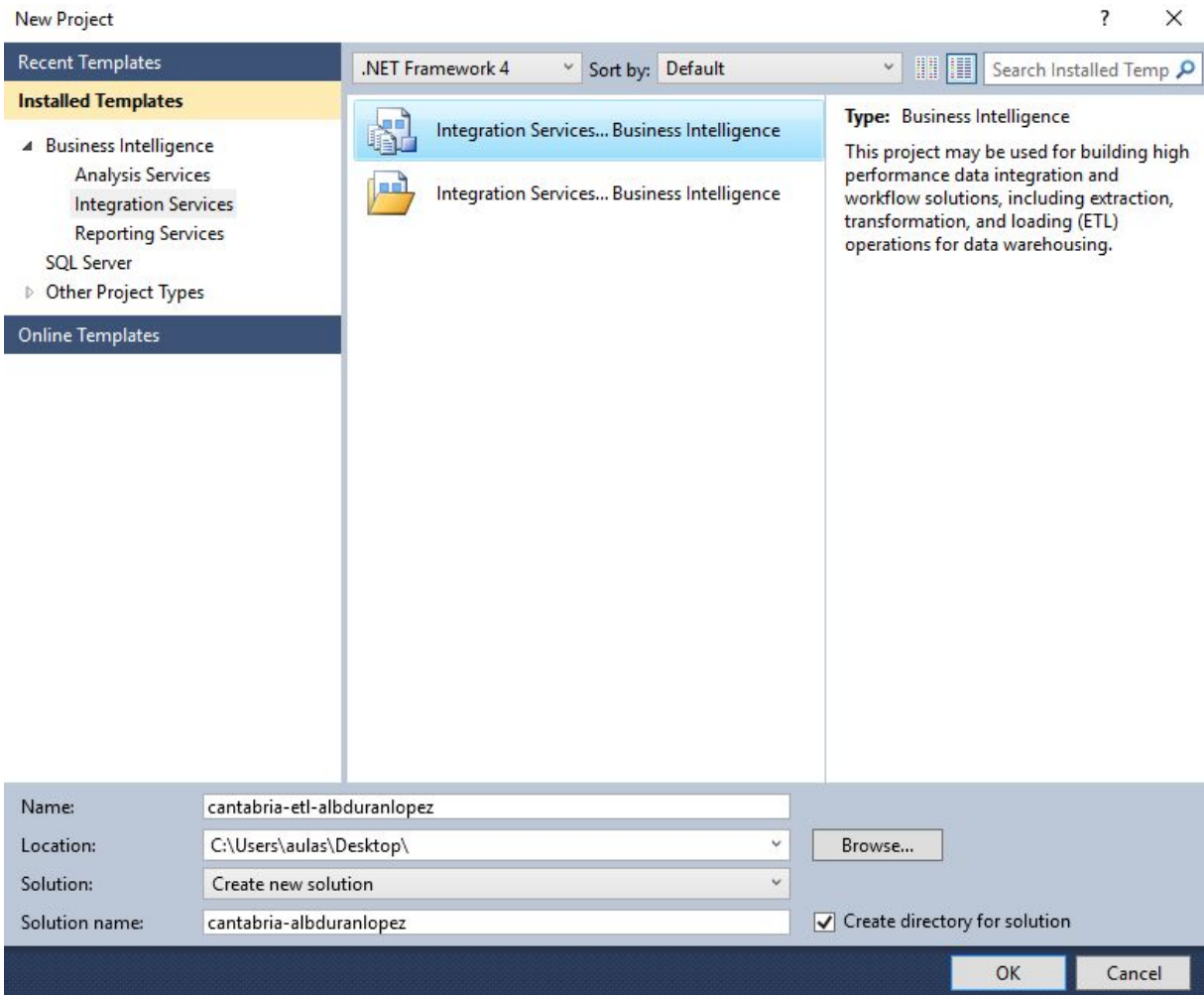
Results

	id_donde	cod_municipio	municipio	capital_municipio	cod_comarca	comarca
1	1	39001	Alfoz de Lloredo	Novales	3901	Costera
2	4	39004	Arenas de Iguña	Arenas de Iguña	3904	Pas-Iguña
3	9	39009	Bárcena de Cicero	Gama	3901	Costera
4	2	39002	Ampuero	Ampuero	3901	Costera
5	3	39003	Anievas	Cotillo	3904	Pas-Iguña
6	5	39005	Argoños	Argoños	3901	Costera
7	7	39007	Aredondo	Aredondo	3905	Asón
8	10	39010	Bárcena de Pie de Concha	Bárcena de Pie de Concha	3904	Pas-Iguña
9	13	39013	Cabezón de Liébana	Cabezón de Liébana	3902	Liebana
10	17	39017	Campoo de Yuso	La Costana	3906	Reinosa
11	6	39006	Amuero	Amuero	3901	Costera
12	11	39011	Bareyo	Ajo	3901	Costera
13	14	39014	Cabuémiga	Valle	3903	Tudanca-Cabuémig
14	15	39015	Camaleño	Camaleño	3902	Liebana

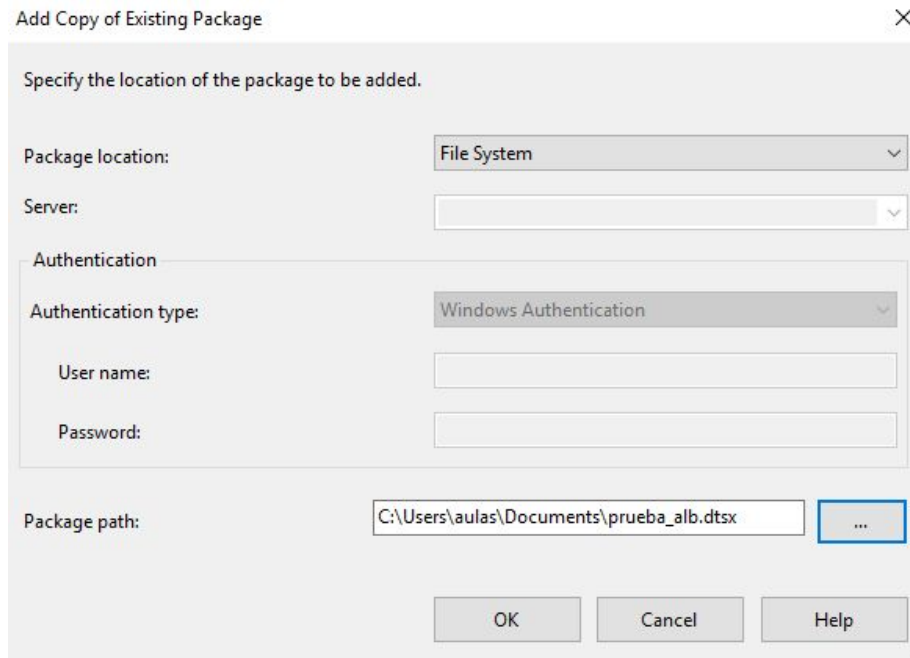
Query executed successfully... | DESKTOP-BJ5E84H (11.0 SP2) | DESKTOP-BJ5E84H\aulas ... | master | 00:00:00 | 102 rows

2. Crea un proyecto SSIS en SSD, y una solución cuyos nombres sean **prueba_albduranlopez**, importa el paquete generado en 1 y ábrelo

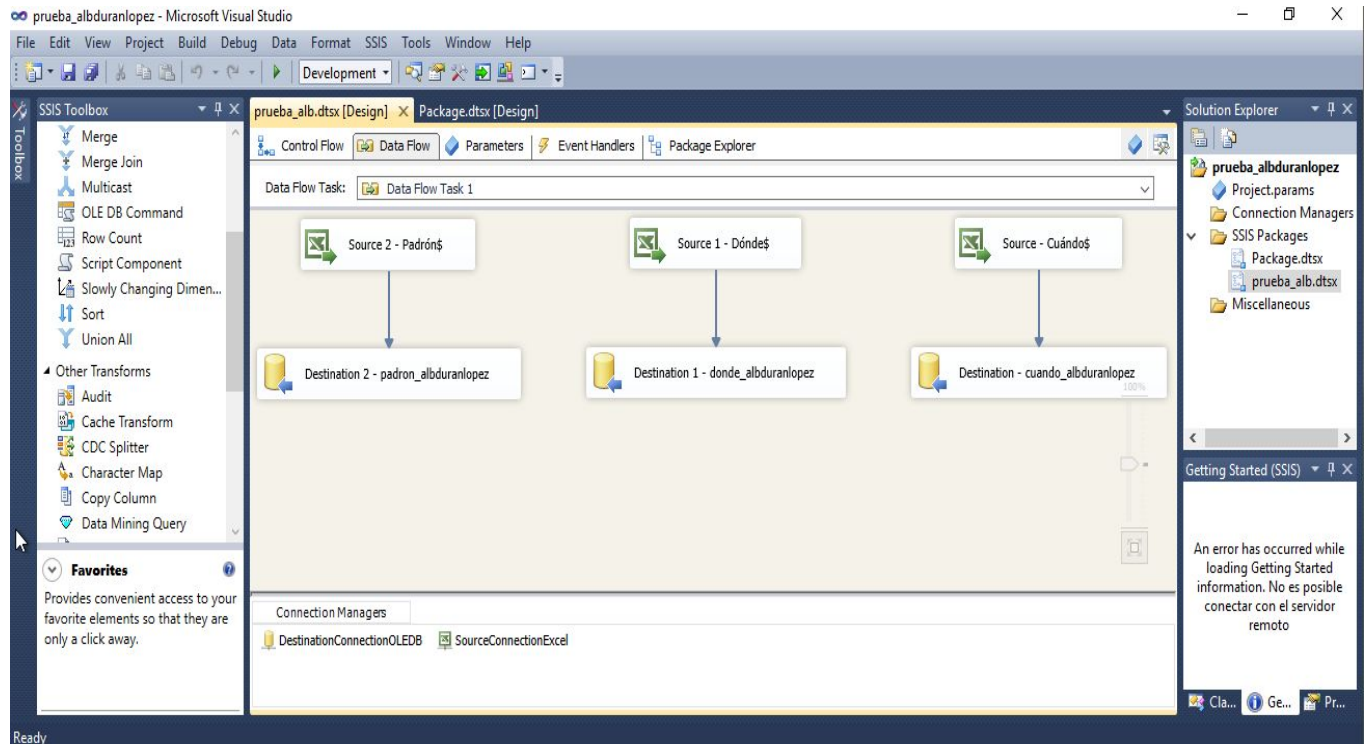
Seguimos las indicaciones del guión de prácticas proporcionado. Creamos un proyecto SSIS donde importaremos el paquete generado en el paso anterior.



Seleccionamos el paquete anterior en el proceso:



Resultados visibles (en la derecha aparece el paquete **prueba_alb** previamente seleccionado):



3. Crea una **BD SQL Server** cuyo nombre sea **cantabria_albduranlopez**, importa las tablas cuando, donde y padrón usando como origen el archivo **Power Query** obtenido y además, el criterio **camel_case**.

-El nombre de cada elemento definido (proyecto, flujo de control, flujo datos, y sus componentes) ha de tener como sufijo tu nombre de usuario de correo de la UGR

Repetimos todo el .proceso anterior. Pero con la diferencia de que ahora todas las tablas, paquetes y elementos definidos tienen que tener el prefijo **_albduranlopez**

En la siguiente captura mostramos el resultado tras la importación de los datos y la creación de las tablas (en la derecha mostramos que están rellenas y que la importación ha sido buena) Destacamos otra vez que el criterio usado ha sido **camel_case**, como se puede comprobar en los nombres de las columnas.

w10olapviAulas [Corriendo] - Oracle VM VirtualBox

Archivo Máquina Ver Entrada Dispositivos Ayuda

DESKTOP-BJ5E84H.cantabria_albduranlopez - dbo.padrón_albduranlopez - Microsoft SQL Server Management Studio

File Edit View Project Debug Query Designer Tools Window Help

Change Type

Object Explorer

Connect

DESKTOP-BJ5E84H (SQL Server 11.0.5058 - D

Databases

System Databases

Database Snapshots

prueba_albduranlopez

cantabria_albduranlopez

Database Diagrams

Tables

System Tables

FileTables

dbo.cuando_albduranlopez

dbo.donde_albduranlopez

dbo.padrón_albduranlopez

Views

Synonyms

Programmability

Service Broker

Storage

Security

Server Objects

Replication

DESKTOP-BJ5E84H.c...ron_albduranlopez

id_donde	id_cuando	hombres	mujeres	superficie
1	1	1348	1298	46,5591
2	1	1332	1276	46,5591
2	1	1664	1673	32,466906
4	1	1084	1081	86,887099999999...
9	1	1140	1073	36,251423
3	1	211	196	20,9664
5	1	398	380	5,558
6	1	908	903	24,696260000000...
7	1	364	304	46,881662999999...
10	1	454	501	30,5447
8	1	6408	6602	6,555353
13	1	424	358	81,480600000000...
17	1	372	389	89,4025
11	1	864	840	32,301148
14	1	582	570	86,141200000000...
12	1	3491	3585	33,6373
15	1	580	536	162,10209999999...
16	1	10934	11377	36,9837

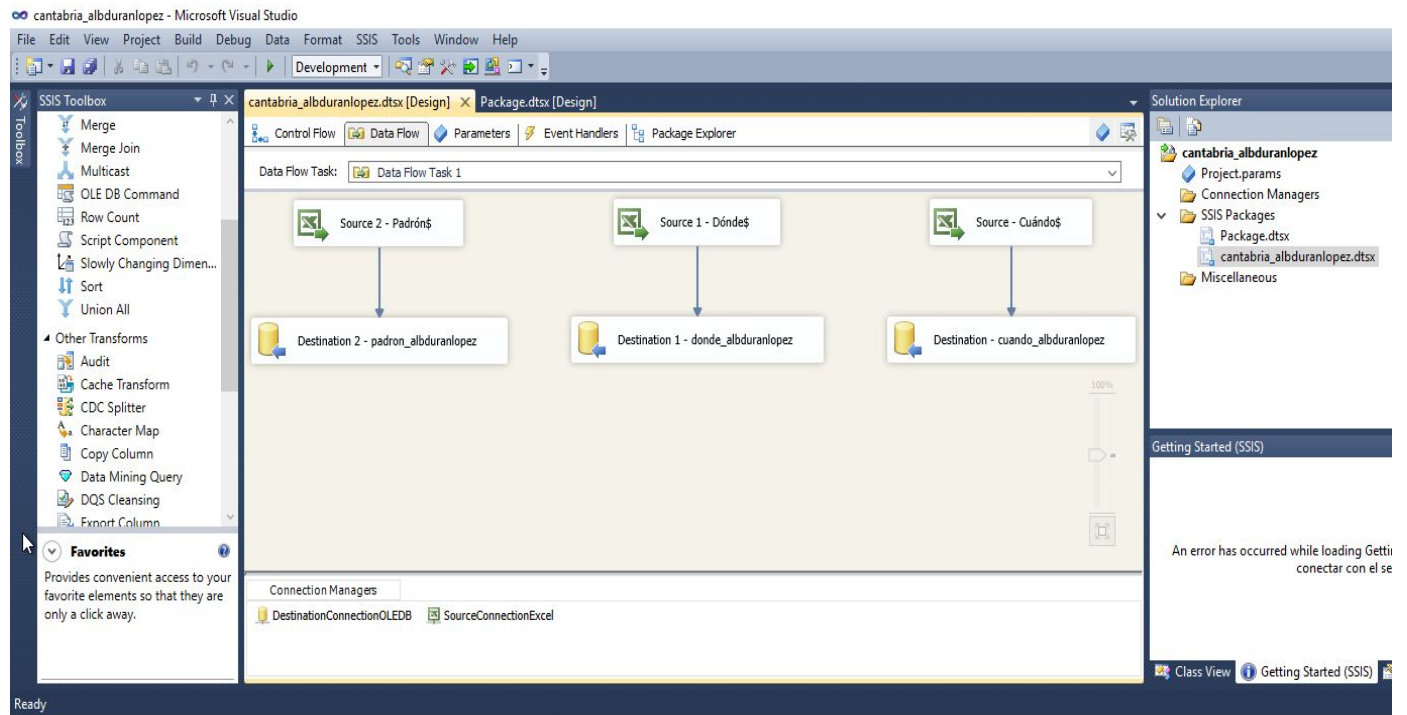
of 200

Ready

17:51 21/03/2020

CTRL DERECHA

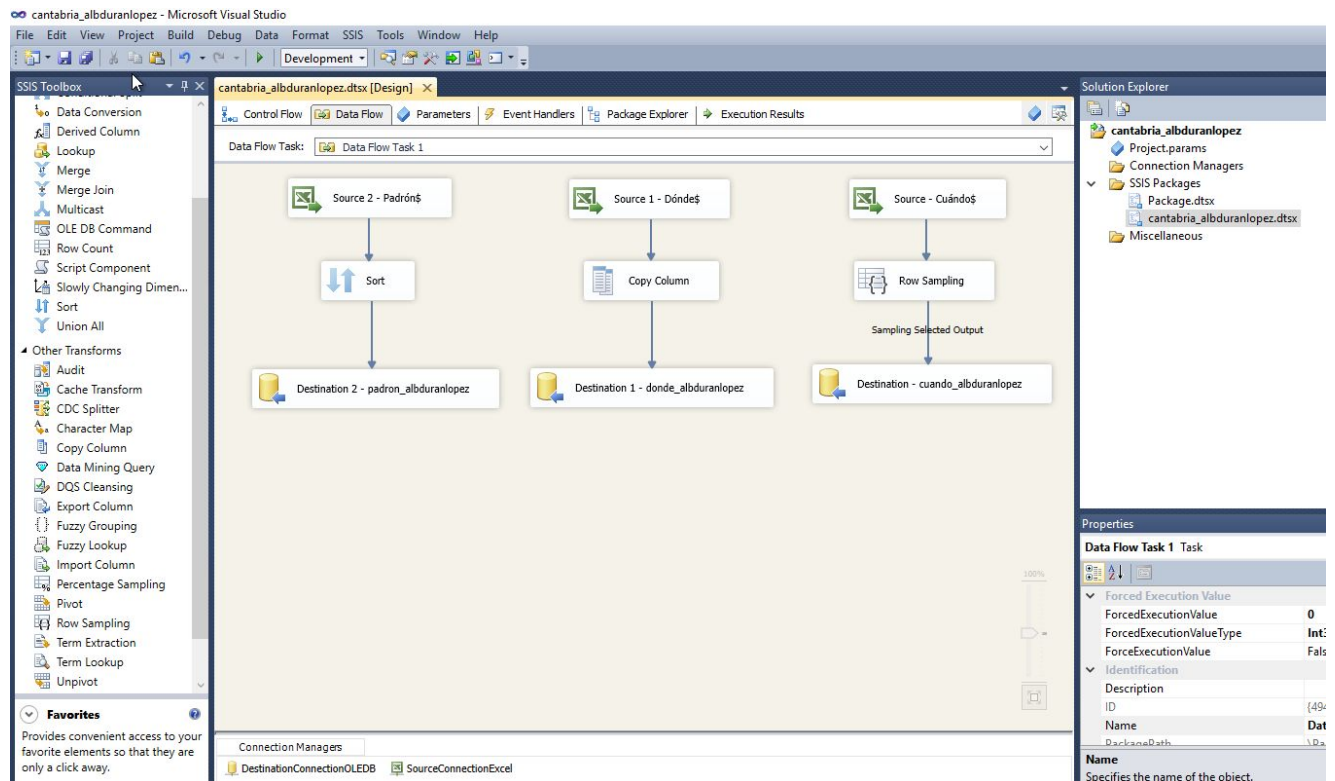
En la columna derecha de <<Solution Explorer>> vemos que esta vez el paquete es distinto, pasando a llamarse **cantabria_albduranlopez**.



-Para cada tabla define un flujo de datos específico.

En la siguiente captura mostramos los tres flujos de salidas.

En la izquierda, el flujo para la tabla Padrón, en el centro para la tabla Dónde y en la derecha para la tabla Cuándo.



Se han usado 3 pasos o nodos diferentes para cada flujo, todos obtenidos de *SSIS Tollbox*.

Flujo de la tabla Padrón: Imaginamos que generemos una tabla en la que obtengamos las entradas ordenadas por su superficie de forma ascendente o descendente. Esto puede ser de utilidad en esta misma base de datos porque se trata de un estudio del INE y en función de la superficie de un territorio puede haber más o menos posibilidades de encontrar un empleo.

La configuración del nodo es la siguiente:

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

Available Input Columns	
<input checked="" type="checkbox"/> Name	Pass Through
<input type="checkbox"/> IdDónde	<input checked="" type="checkbox"/>
<input type="checkbox"/> IdCuándo	<input checked="" type="checkbox"/>
<input type="checkbox"/> Hombres	<input checked="" type="checkbox"/>
<input type="checkbox"/> Mujeres	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> Superficie	<input checked="" type="checkbox"/>

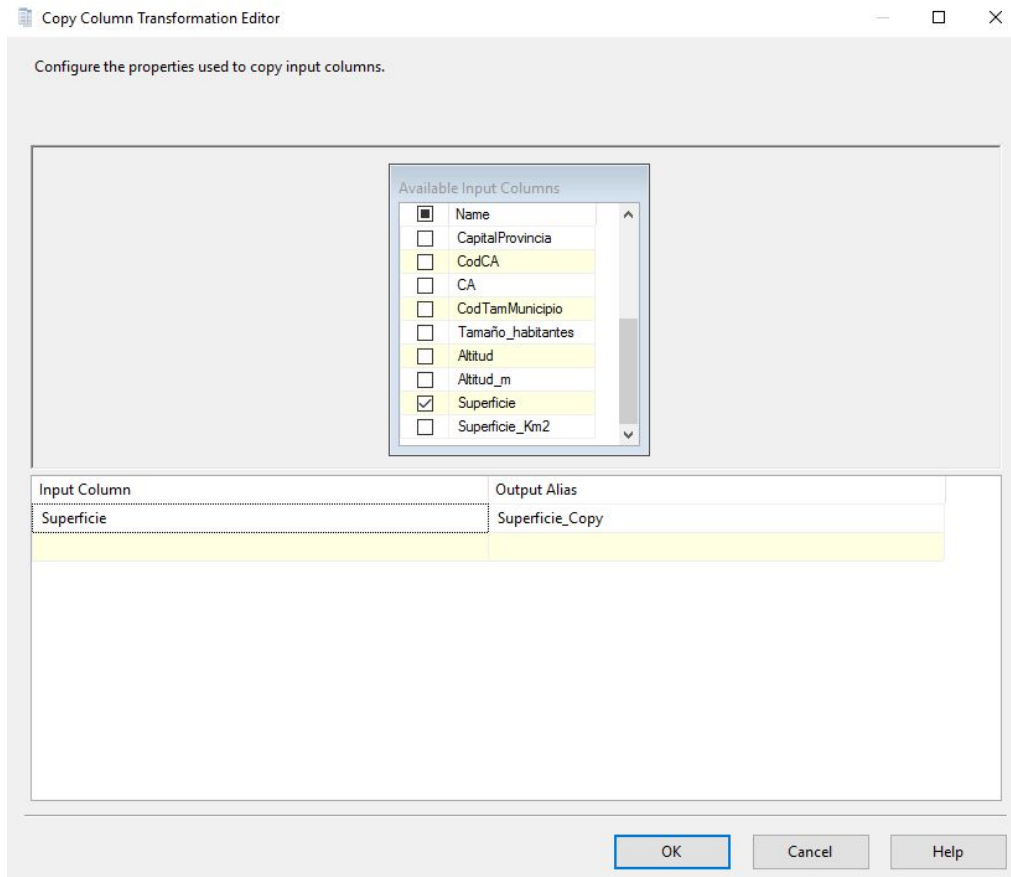
Input Column	Output Alias	Sort Type	Sort Order	Compare
Superficie	Superficie	ascending	1	

☐ Remove rows with duplicate sort values

OK Cancel Help

Flujo de la tabla Dónde:

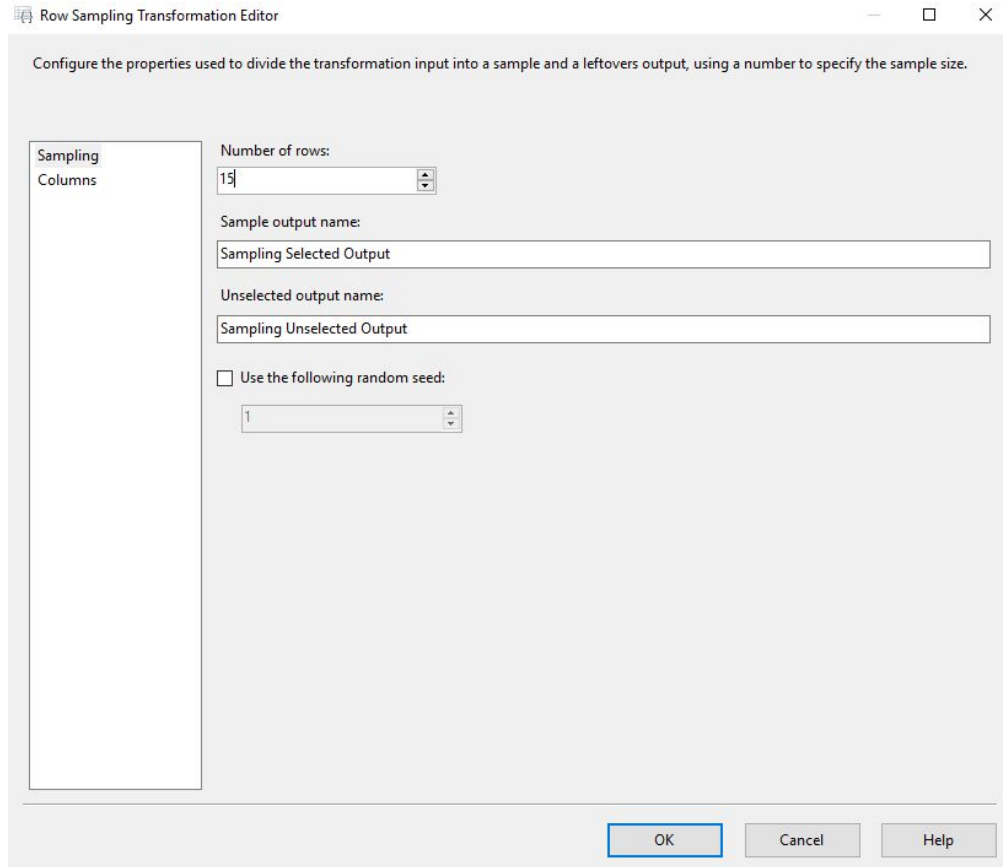
Realizamos una transformación sencilla. Realizamos una copia de la columna Superficie. A partir de esta copia podemos realizar nuevas operaciones como por ejemplo concatenar esta copia con una descripción (como se realizó en la Práctica2)



Flujo de la tabla Cuándo:

Mostramos la utilidad del siguiente nodo. Realmente en nuestro dataset no sería conveniente pero en estudios de data mining sí es muy útil (Cuando se necesita separar el conjunto o coger una muestra específica para el conjunto de Test, como se hizo en la asignatura Inteligencia de Negocio)

Aquí, seleccionamos un número de columnas de 15 y este nodo seleccionará 15 de forma aleatoria.



Row Sampling Transformation Editor

Configure the properties used to divide the transformation input into a sample and a leftovers output, using a number to specify the sample size.

Sampling
Columns

Number of rows:
15

Sample output name:
Sampling Selected Output

Unselected output name:
Sampling Unselected Output

☐ Use the following random seed:
1

OK Cancel Help

Ahora bien, ejecutamos los paquetes y vemos que en todos tenemos los ticks verdes que significan que se han ejecutado sin problemas y no ha habido errores.

