PEC 1. Solución.

Parte práctica (60%).

La información mostrada en las captura de pantalla donde aparecen parámetros de salida del sistema, pueden variar en función del servidor, base de datos y usuario utilizados

Ejercicio 1 (5%): Configuración del entorno VDI

Objetivo:

Configurar la conexión al escritorio VDI que se utilizará durante todo el curso, tanto para realizar esta PEC (PEC1) como las siguientes actividades evaluables del curso

Solución:

1) Abrimos la URL de acceso a la VDI:



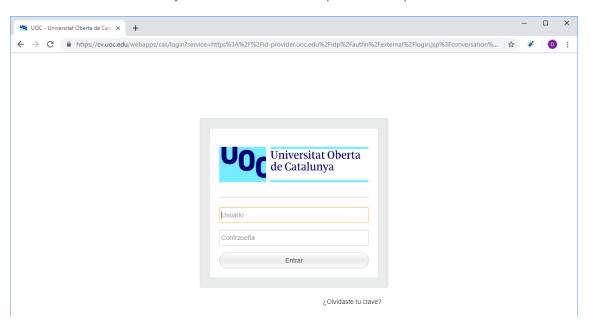
2) Es posible que durante unos segundos aparezca la siguiente pantalla. No es necesario hacer nada y se cerrará automáticamente.



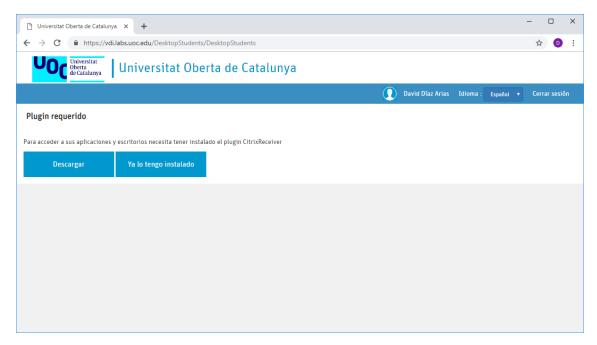


1

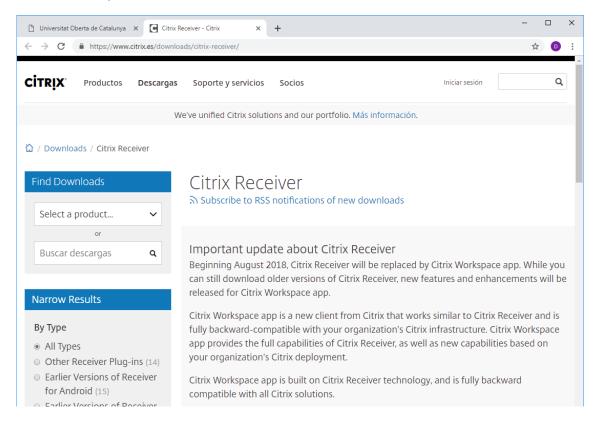
3) A continuación, aparecerá la pantalla de autenticación de la UOC, donde utilizaremos el usuario y contraseña del campus online para identificarnos.



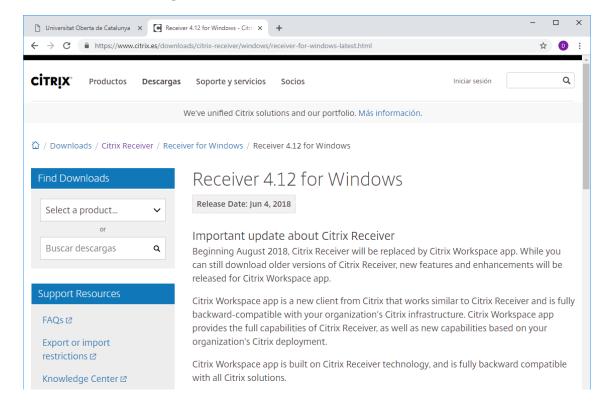
4) Pasados unos segundos aparecerá una pantalla que nos permitirá descargar el cliente de Citrix si no lo tenemos ya instalado en nuestra máquina.



5) Pulsamos el botón "Descargar" y nos redirigirá a la web de descargas de *Citrix***Receiver. https://www.citrix.es/downloads/citrix-receiver/



 Realizaremos la descarga correspondiente al Sistema Operativo utilizado, en esta solución descargamos Receiver 4.12 for Windows.

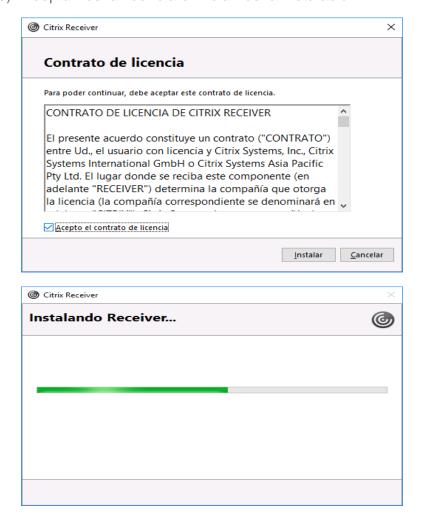


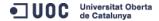


 Se descargará el archivo CitrixReceiver.exe que ejecutaremos con privilegios de administrador.

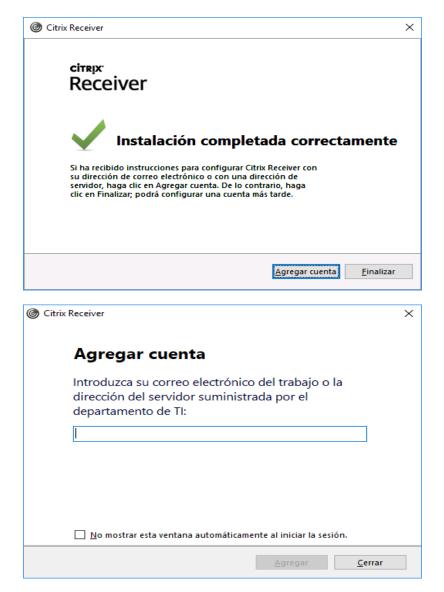


8) Aceptamos la licencia e iniciamos la instalación.



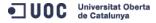


9) Una vez finalizada la instalación es posible que nos pida agregar una cuenta. Esto no es necesario y podemos finalizar la instalación. En este punto se recomienda marcar la casilla "No mostrar esta ventana automáticamente al iniciar la sesión".

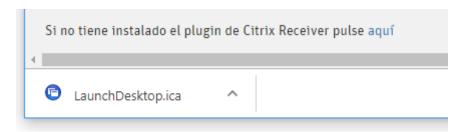


10)Tras reiniciar el navegador y acceder de nuevo a la URL de conexión a la VDI, aparecerá la opción de "Conectar"





11) A continuación, se descargará el archivo LaunchDesktop.ica



12) Finalmente tendremos acceso al escritorio de la máquina VDI.







Ejercicio 2 (20%): Validación de la BBDD.

Objetivo:

Toma de contacto con el Sistema Gestor de Bases de Datos facilitado con la máquina VDI (SQL Server 2017), configuración de una conexión con el servidor y creación de una base de datos simple con una tabla.

Desarrollo del ejercicio:

1) Conectar al servidor de bases de datos SQL Server

Utilizad la consola "Microsoft SQL Server Management Studio 2017 (SSMS)" instalada en la VDI para abrir una conexión con el servidor de base de datos y adjuntad las siguientes capturas de pantalla:

a) Versión de SSMS

Esta información se puede obtener desde la pantalla splash

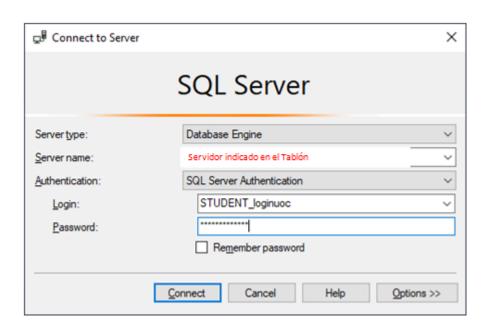


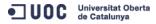
También se puede obtener esta información desde el menú *Help* - *About...*



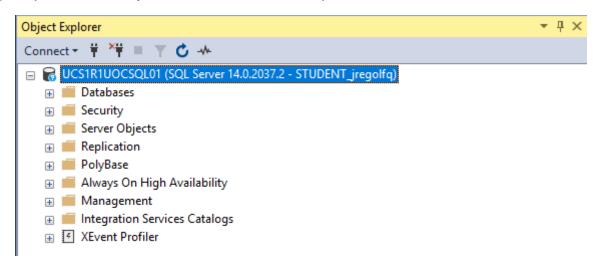


- b) Pantalla de conexión inicial de la consola SQL Server Management Studio (SSMS) donde se pueda observar la configuración utilizada.
 - o Server Type
 - Server name
 - o Authentication
 - Login





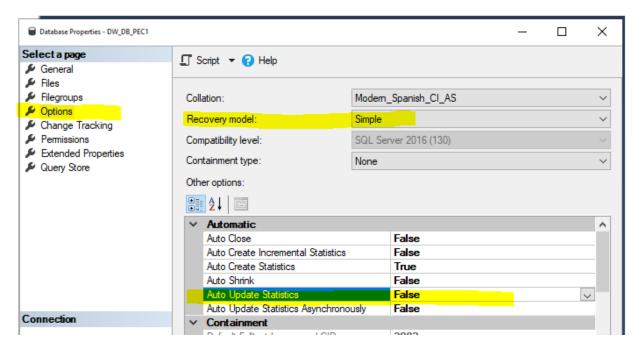
c) Explorador de objetos del SSMS donde se puede ver el usuario de conexión.



2) Configurar la base de datos DB_loginuoc

Se debe configurar el parámetro *Recovery Model* de la BBDD como "simple" y desactivar la actualización automática de estadísticas.

Adjuntad las capturas de pantalla que demuestren la correcta configuración.



3) Crear una tabla con el nombre T PEC1

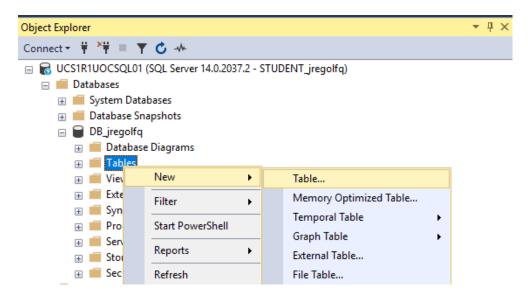
La tabla tendrá dos campos: campo1 y campo2 de la siguiente manera:

- Campo1 será de tipo varchar(50) y permitirá valores nulos
- Campo2 será de tipo int y no permitirá valores nulos.

Adjuntad las capturas de pantalla que demuestren la correcta creación de la tabla con sus campos.



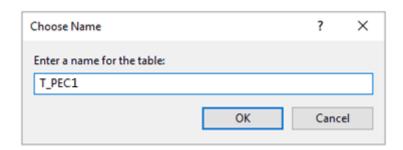
Desde el explorador de objetos se hace clic sobre la carpeta *Tables* y se crea una nueva tabla



Se definen los campos según los requerimientos.

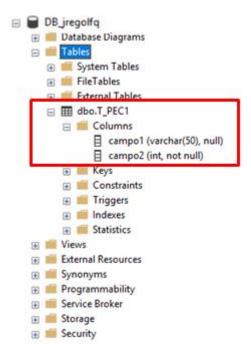
	Column Name	Data Type	Allow Nulls
	campo1	varchar(50)	
\blacktriangleright	campo2	int	

Una vez guardada la tabla con el nombre T_PEC1



Se puede comprobar en el explorador de objetos que la tabla ha sido creada correctamente.







También se puede crear la tabla ejecutando el siguiente script SQL

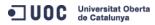
```
USE [DB_loginuoc]
GO

SET ANSI_NULLS ON
GO

SET QUOTED_IDENTIFIER ON
GO

CREATE TABLE [dbo].[T_PEC1](
[campo1] [varchar](50) NULL,
[campo2] [int] NOT NULL
) ON [PRIMARY]
GO
```

```
Commands completed successfully.
```



Ejercicio 3 (25%): Validación de la aplicación de ETL.

Objetivo:

Toma de contacto con la aplicación de ETL suministrada en la máquina VDI (*Pentaho Data Integration*) y capacitar al estudiante para iniciar las actividades del curso

Solución:

1) Abrir el componente Spoon.

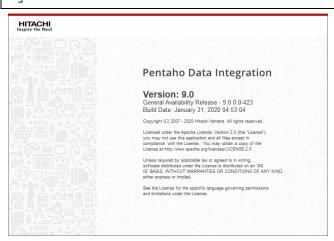
En el escritorio tenemos un acceso directo a PDI (spoon).



2) Capturar versión de PDI.

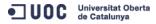
Mientras se abre podemos capturar la versión de la aplicación, también se puede obtener en:

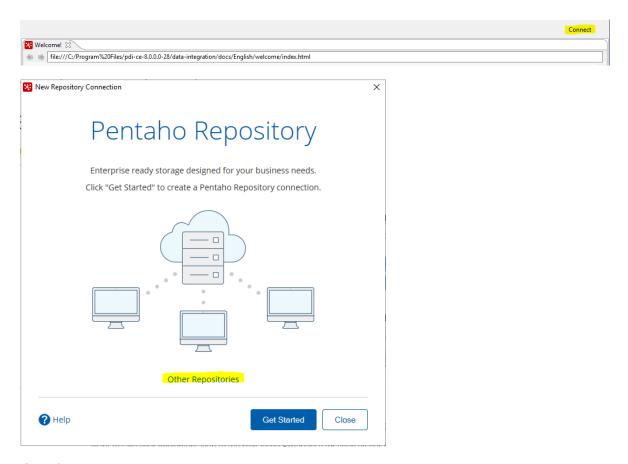
Ayuda > Acerca de



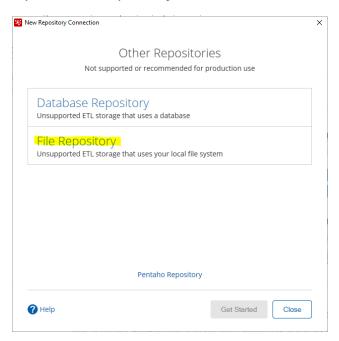
3) Crear un repositorio de tipo file con el nombre PEC1.

En el botón connect seleccionar la opción Other Repositories:



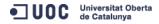


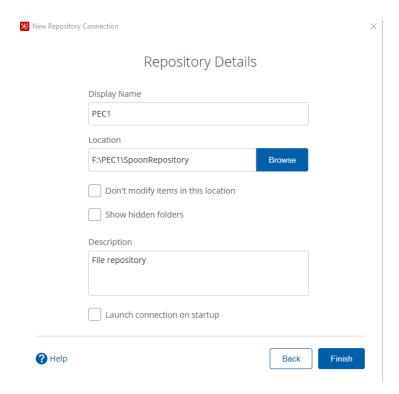
Opción File Repository:



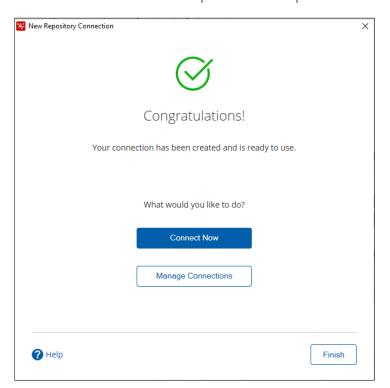
Seleccionar la localización y asignar un nombre al repositorio.

En location se debe seleccionar una ubicación persistente de la máquina VDI.





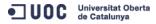
Finalmente se creará el repositorio de tipo file:





4) Crear una transformación y un *Job* básicos:

Anexo al enunciado de la PEC1, encontraréis un archivo datos.csv.



Se debe copiar el archivo de origen de datos a una de las carpetas persistentes de la máquina VDI.

Análisis preliminar del origen de datos

Si abrimos datos.csv con un editor de texto, podemos hacer una primera revisión de su contenido.

Identificar los campos a cargar

El fichero CSV contiene 8 campos:

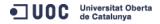
PRACNO
PRACTICENAME
ADDRESS1
ADDRESS2
ADDRESS3
POSTCODE
LCG
REGISTERED_PATIENTS

Determinar el tipo de dato de cada campo

Tipo dato	Observaciones
Número	
Texto	
Número	El campo "Registered Patients" que aparentemente debería ser de tipo integer contiene valores string
	Número Texto Texto Texto Texto Texto Texto Texto Texto Texto

Identificar los campos que deben admitir (o no) valores nulos

Campo	¿Valores	Observaciones
	nulos?	



PRACNO	No	El campo PracNo es un identificador único y no debe permitir valores nulos.
PRACTICENAME	Si	
ADDRESS1	Si	
ADDRESS2	Si	
ADDRESS3	Si	
POSTCODE	Si	
LCG	Si	
REGISTERED_PATIENTS	Si	

Otras observaciones

Existen registros totalmente vacíos que deben ser descartados antes de la carga en la BBDD:

B. Crear una tabla en la BBDD con el nombre T DATOS

Esta tabla permitirá cargar toda la información del archivo *datos.csv* en los campos correspondientes.

Después de analizar el archivo CSV observamos que el campo "Registered Patients" debería ser integer, así que lo crearemos de este tipo en la BBDD y trataremos los posibles errores en la carga. Además, el camp PRACNO no debe permitir valores nulos por tratarse de un identificador único.

El siguiente script generará la tabla necesaria para la carga:

```
CREATE TABLE [dbo].[T_DATOS]

(

PRACNO INT not null,

PRACTICENAME [VARCHAR](36) null,

ADDRESS1 [VARCHAR](46) null,

ADDRESS2 [VARCHAR](47) null,

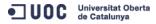
ADDRESS3 [VARCHAR](38) null,

POSTCODE [VARCHAR](8) null,

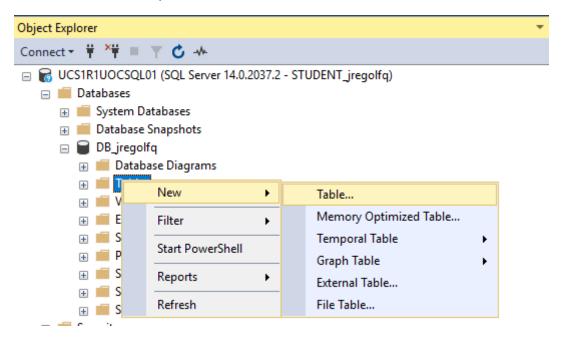
LCG [VARCHAR](13),

REGISTERED_PATIENTS INT null

);
```



La tabla también se puede crear desde la interfaz de usuario del SSMS



C. Crear una transformación con Spoon que realice las siguientes tareas:

Archivo > Nuevo> Transformación

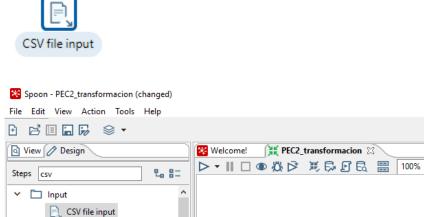
I. Extraer la información del archivo.

GZIP CSV Input
S3 CSV Input

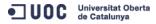
OutputStreaming

Transform

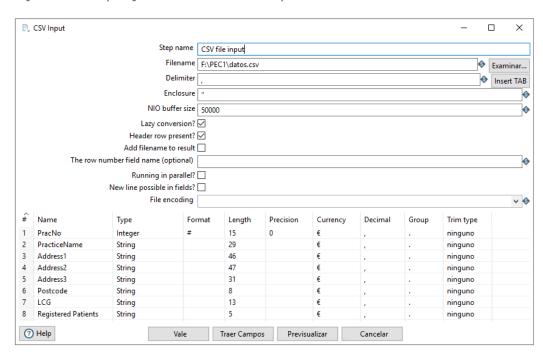
Insertamos un paso de extracción desde CSV (CSV input file):



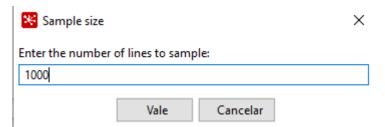
CSV file input



Seleccionamos el archivo a importar en *Filename* y con el botón *[Get Fields]* o *[Traer Campos]* obtenemos los campos del archivo:



Es recomendable aumentar el *sample size* para ayudarnos a determinar el tipo de datos de cada campo a extraer.

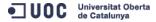


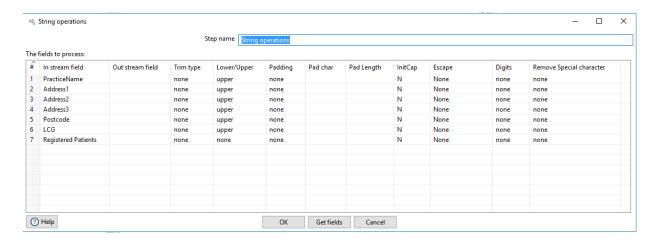
II. Transformar todos los datos de los campos de tipo texto a mayúsculas

Insertamos un paso "String operations":



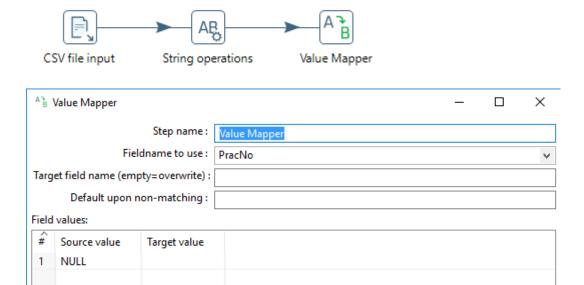
Y configuramos el valor "upper" en la propiedad "Lower/Upper"





III. Corregir los errores detectados en el análisis

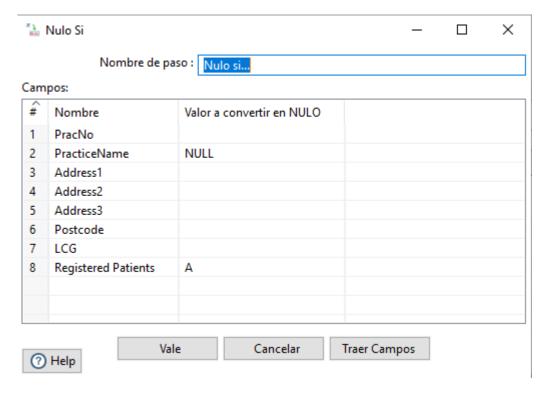
 Como sabemos que el campo PracNo no debe admitir valores nulos, mapeamos el valor string "NULL" por <null> con un paso "Value Mapper", esto nos permitirá descartar estos los registros en un paso posterior.



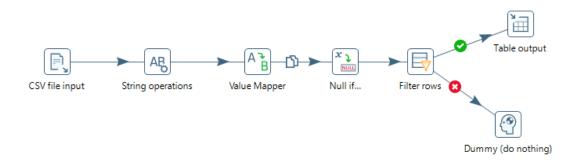
Substituir los valores incorrectos (no numéricos) de [Register Patients]
por valores <null> con un paso "Null if" (Nulo si...). También sustituimos
los valores string "NULL" del campo PracticeName por valores <null>
reales.





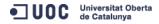


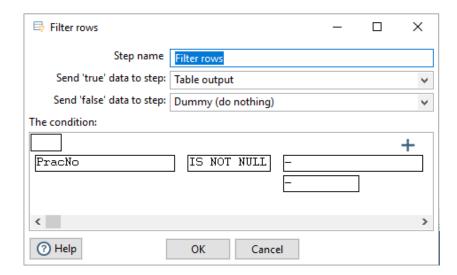
Filtrar los registros que tienen valor de PracNo = null y añadir los pasos
 "Filter Rows" para redirigir los registros correctos a "Table output" y los incorrectos "Dummy (do nothing)"



[Filter rows] tendrá la siguiente lógica:

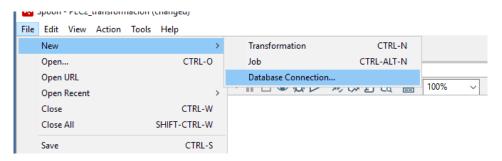
- Si PracNo is not null □ Table output
- Si PracNo is null □ Dummy



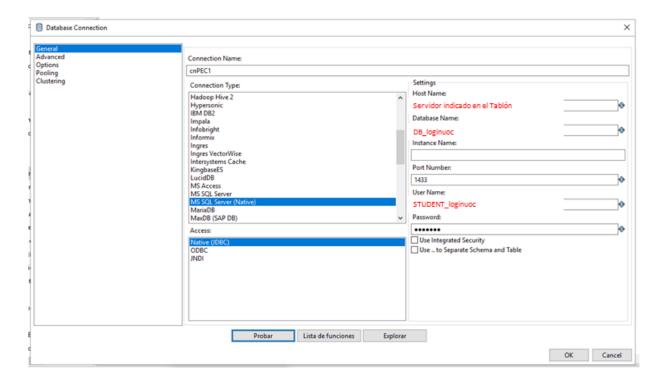


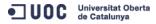
IV. Cargar la información transformada en la tabla anteriormente creada.

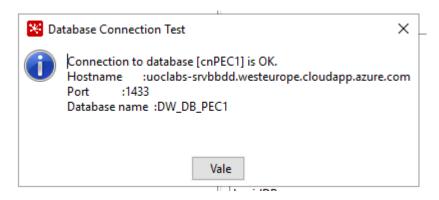
Creamos la conexión a la base de datos con el nombre cnPEC1



También se puede crear desde el paso "Table output".

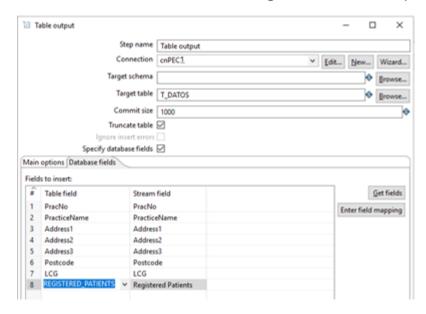






Una vez creada la conexión seleccionamos en el paso "Table output", los campos origen que vamos a cargar y su correspondencia en la base de datos:

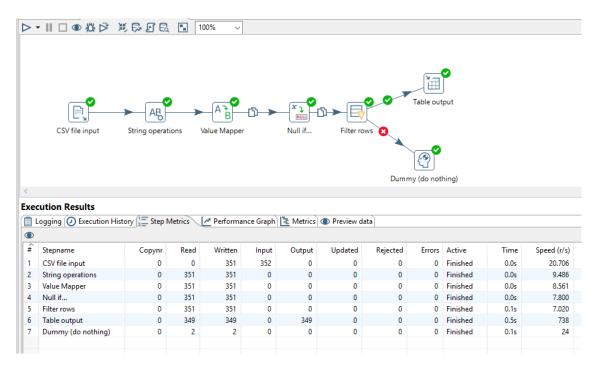
Es necesario revisar la correcta asignación de los campos



V. Realizar la carga efectiva de la tabla y validar el resultado

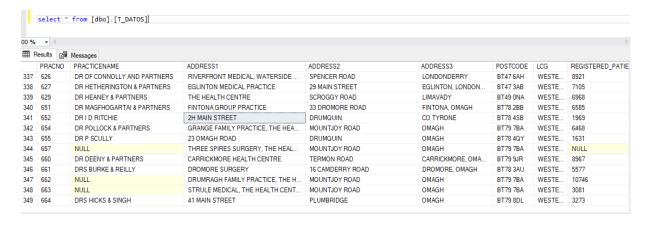
Si ejecutamos la transformación podemos comprobar que se cargan los registros de origen correctamente. Los registros vacíos no serán cargados.



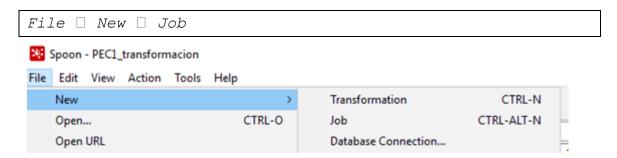


Observamos que se han cargado 349 registros de los 351 originales.

Si comprobamos el resultado directamente en la tabla vemos que efectivamente se han cargado 349 registros, ignorando los registros vacíos. También observamos que todos los valores de tipo texto están en mayúsculas y que los valores incorrectos de REGISTERED PATIENTS están a *null*.



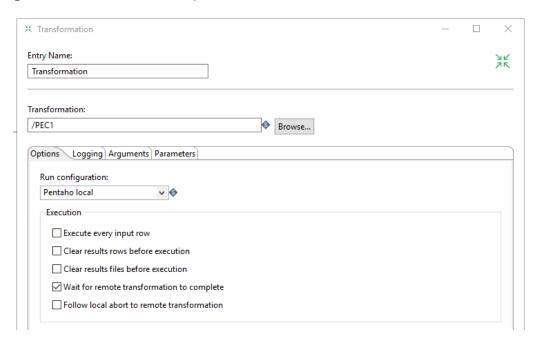
D. Crear un job que ejecute la transformación anterior



Los Jobs siempre deben contener como mínimo un START y un SUCCESS.



El paso "transformation" llamará la ejecución de la transformación que hemos guardado anteriormente, podemos seleccionarla mediante el botón "Browse":



Guardamos el job y lo ejecutamos:





Ejercicio 4 (10%): Microsoft SQL Analysis Services

Objetivo:

Toma de contacto con la aplicación MOLAP suministrada en la máquina VDI (Microsoft SQL *Analysis Services*) y capacitar al estudiante para iniciar las actividades del curso.

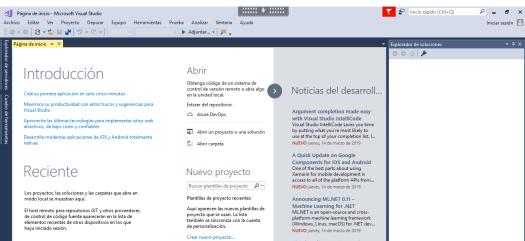
En este último ejercicio veremos cómo crear un cubo de SSAS y conectarlo a nuestra base de datos. Utilizaremos el recurso "Importar base de datos, como crear vistas y cubos.pdf" como guía de configuración y punto de partida.

Solución:

1) Abrir Visual Studio (SSAS)

Tenemos el acceso directo en el escritorio de la VDI.

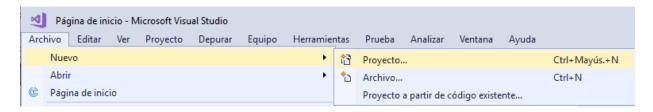




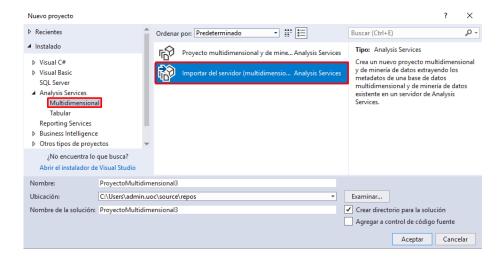


2) Importar nuestro proyecto de *Analysis Services* multidimensional desde el servidor

Archivo > Nuevo > Proyecto

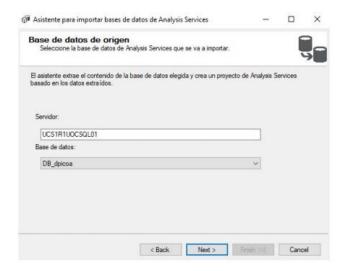


Seleccionaremos la opción "Importar del Servidor (multidimensional y minería de datos" dentro de la categoría "*Analysis Services*"

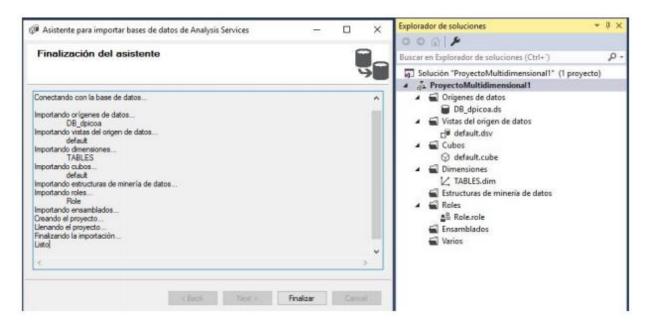


Una vez aceptado el paso anterior, nos aparecerá un asistente para importar la estructura de la base de datos de Analysis Services, después del primer Next, llegaremos a la ventana selección de Origen de la base de datos. Introduciendo el nombre del MS SQL Server (el que se facilitará vía tablón de la asignatura), aparecerá nuestra base de datos en el desplegable inferior.





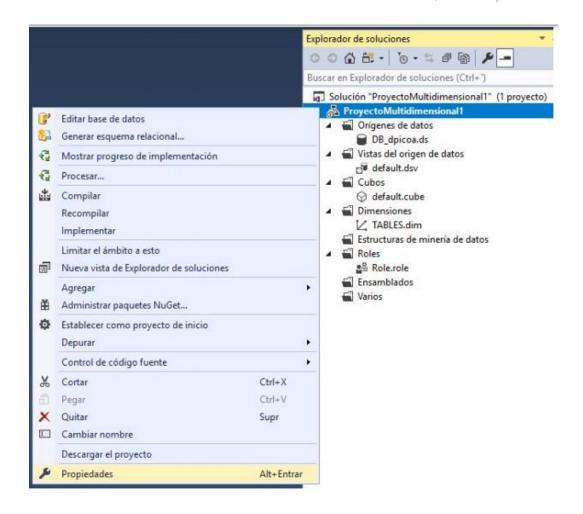
Si hemos completado correctamente el proceso obtendremos un pantalla similar a ésta, por un lado una ventana con la finalización del asistente y por otro, si tenemos el explorador de soluciones abierto, la estructura creada para la base de datos importada.



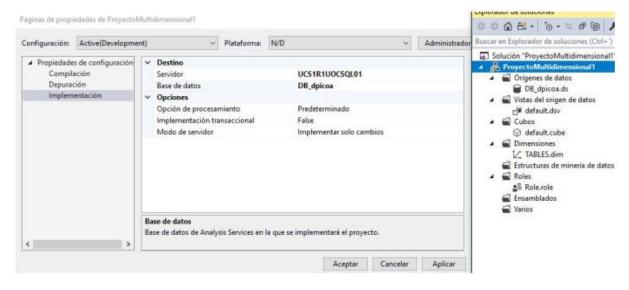
Aunque hemos importado en el paso anterior la base de datos, en el actual proyecto tendremos que indicar que lo almacene en el servidor. Para realizar esta configuración, se puede utilizar el explorador de soluciones. Si lo tenemos cerrado,se puede abrir desde el menú -> Ver y "Explorador de soluciones".

Hacemos clic con el botón derecho sobre el nombre que se indicó en el proyecto y vamos a propiedades:





Una vez situados en las propiedades, iremos al apartado de implementación, donde cambiaremos el valor del servidor por el nombre del MS SQL server (el que se ha facilitado vía tablón de la asignatura) y en Base de datos, tendremos que indicar el mismo nombre que tenemos en Origen de datos del explorador de soluciones

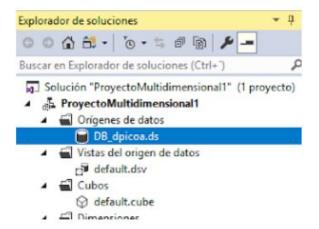




Aplicamos y aceptamos los cambios, y ya tendremos configurado el destino de nuestros datos.

3) Configurar un origen de datos (DataSource) conectado a DB_loginuoc

Para poder configurar que el origen de los datos sea nuestra base de datos del MS SQL server, debemos realizar doble click sobre el nombre de la base de datos en el Explorador de Soluciones

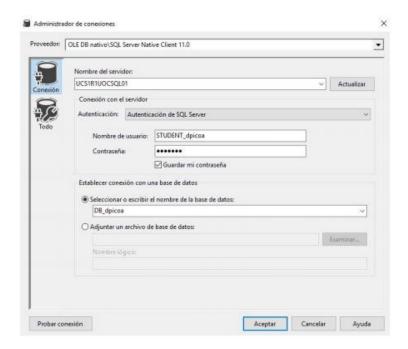


Con esta acción se abrirá una nueva ventana con titulo Data Source Designer



En la pestaña general vamos a editar la conexión del origen de datos. Verificamos que el Data Source sea el nombre de nuestra base de datos y pulsamos sobre el botón Edit... con esto se nos abrirá el administrador de conexiones.



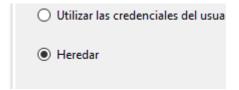


Debemos asegurarnos de que los valores sean los adecuados para:

- Nombre del servidor : el que se facilitará vía tablero de la asignatura
- Nombre de usuario: STUDENT_[nuestro login en UOC].
- Indicar la contraseña que tenemos en MS SQL Server. Recomendamos activar la casilla Guardar mi contraseña para no tener que introducirla en cada acción.
- Nombre de la base de datos el mismo que teníamos en el Explorador de archivos

Una vez configurados estos valores podemos Probar Conexión y si todo está correcto podemos Aceptar para finalizar la edición de la conexión.

Una vez editada la conexión de datos, pulsamos sobre Impersonation Information y nos aseguramos que la opción seleccionada es Heredar

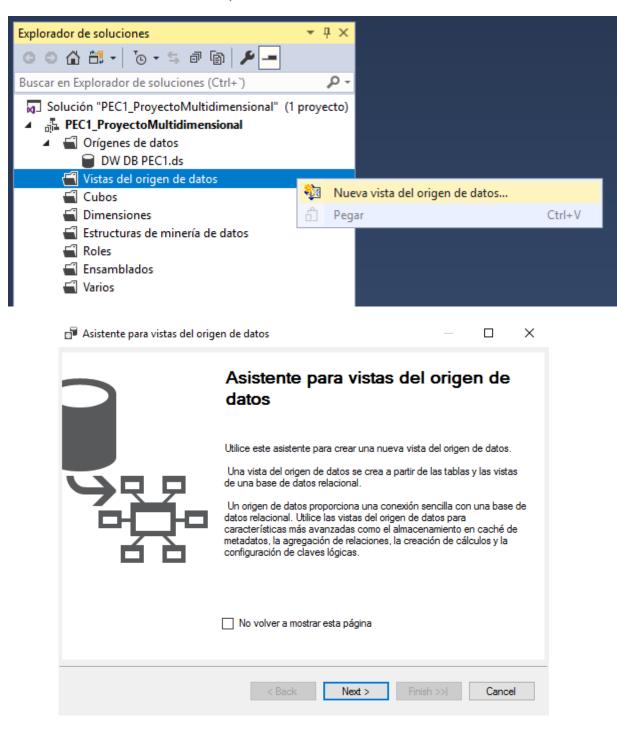




Y ahora ya si, podemos pulsar OK para terminar el proceso de configuración.

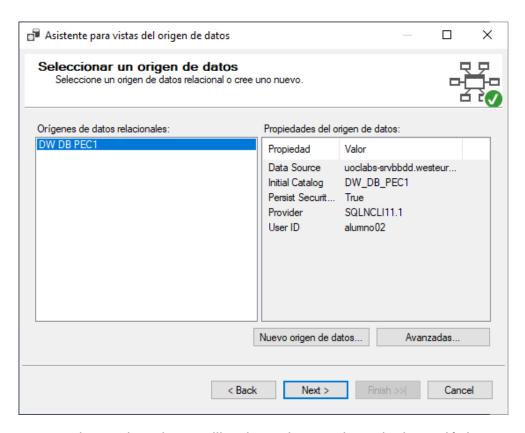
4) Crear una vista de origen de datos (Data Source View)

Utilizaremos el asistente del explorador de soluciones

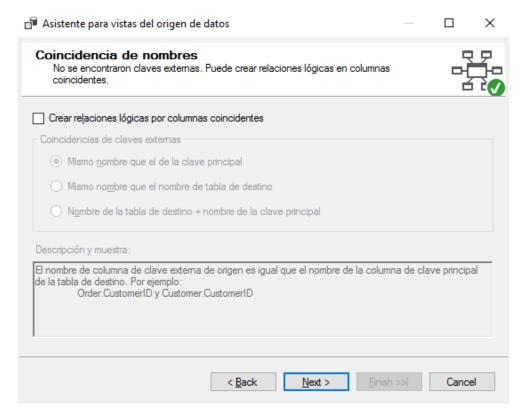


Escogemos el Data Source creado en el paso anterior:



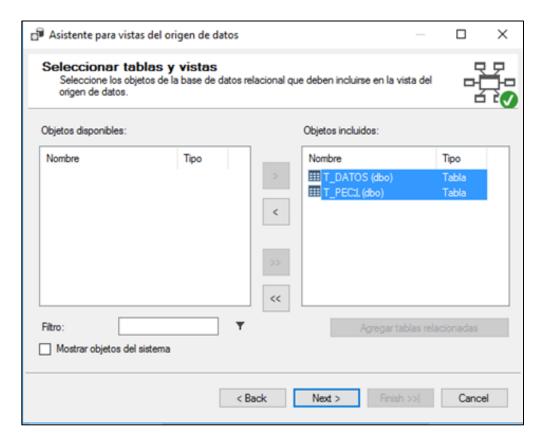


Como se trata de un ejemplo sencillo, desactivamos las relaciones lógicas

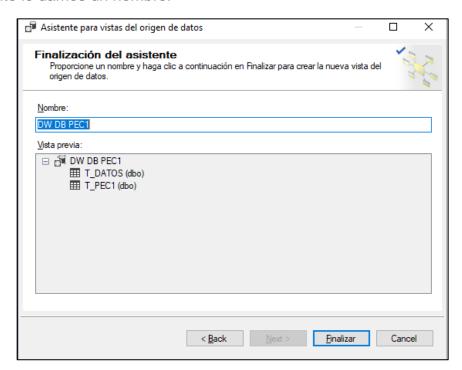


Seleccionamos las tablas que queremos visualizar:



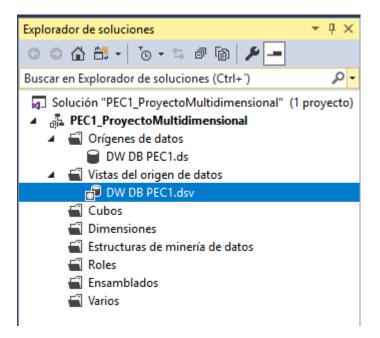


Finalmente le damos un nombre:

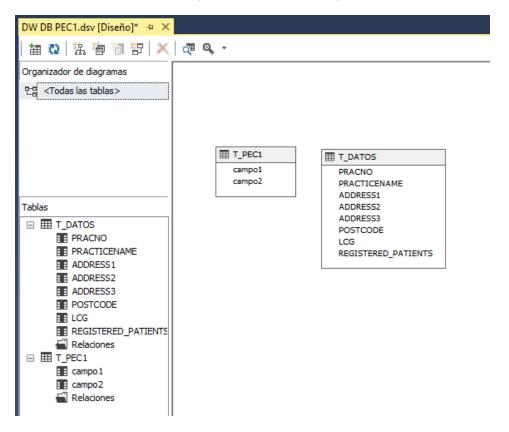


Ya tenemos creado el Data Source View:



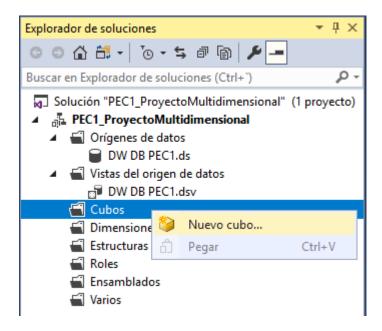


Si abrimos el "Data Source View" podemos ver los objetos seleccionados.

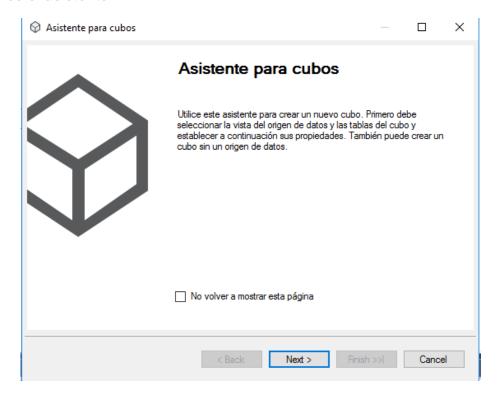


5) Crear un cubo con una única tabla de hechos T_Datos.



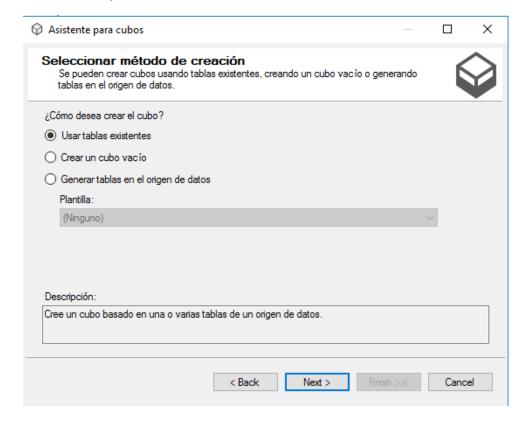


Iniciamos el asistente

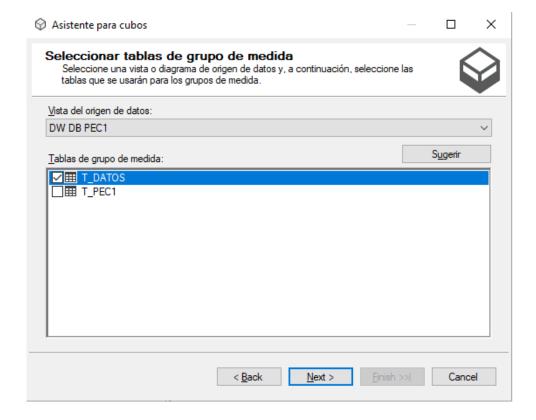




Seleccionamos la opción "Usar tablas existentes"

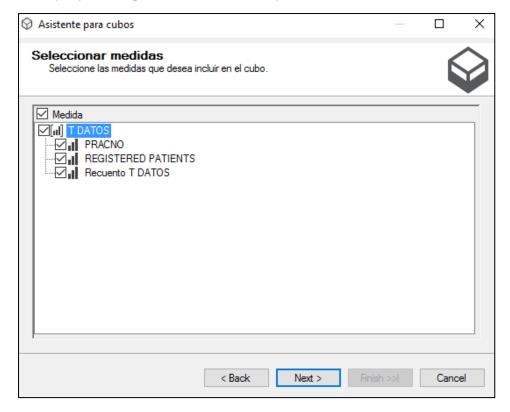


Seleccionamos la tabla T_DATOS

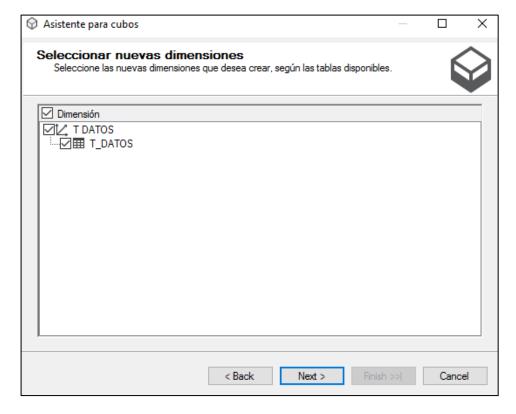




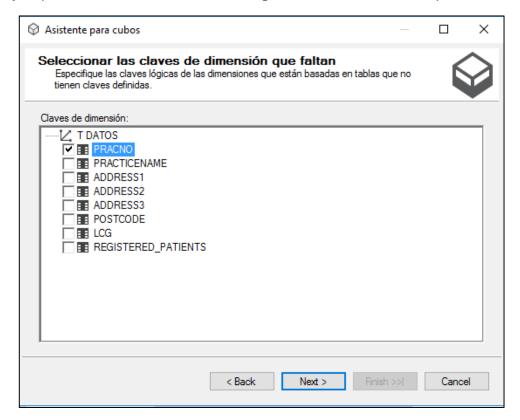
El asistente nos propone algunas medidas. Aceptamos.



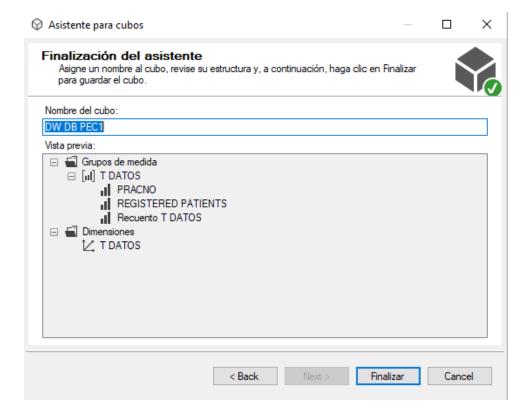
Aceptamos la dimensión propuesta



Es muy importante seleccionar la clave lógica, utilizaremos el campo PRACNO

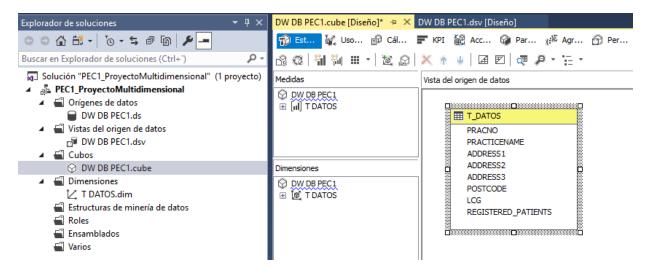


Le damos nombre al cubo y finalizamos el asistente

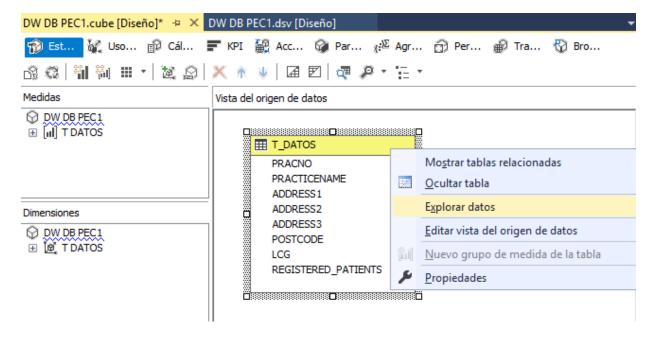




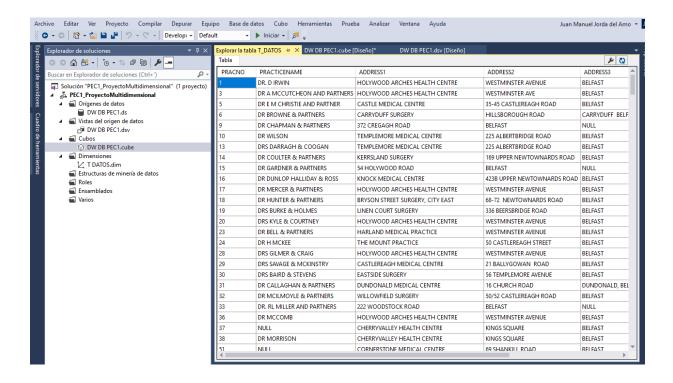
Ya tenemos el cubo creado.



Podemos explorar el contenido de la tabla de hechos



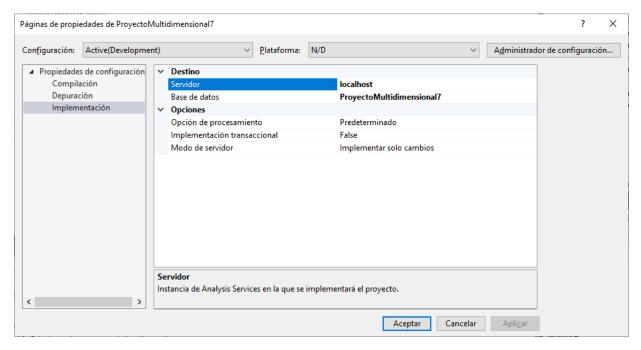




6) Procesar e implementar el cubo en el servidor

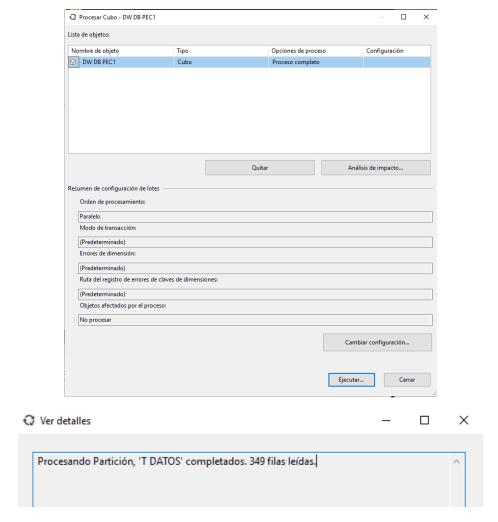
Finalmente, vamos a procesar el cubo para tener disponible nuestro almacén de datos multidimensional en el servidor Analysis Services.

Primero, en el apartado de propiedades del proyecto indicaremos el nombre del servidor, cambiando *localhost* por *el nombre del servidor indicado en el tablón,* y también indicaremos la base de datos DB_loginuoc





Hacemos clic con el botón derecho sobre el cubo y seleccionamos procesar.



Desde el navegador integrado en SSAS (*Browser*) podemos explotar el cubo desplegado en el servidor SQL.

