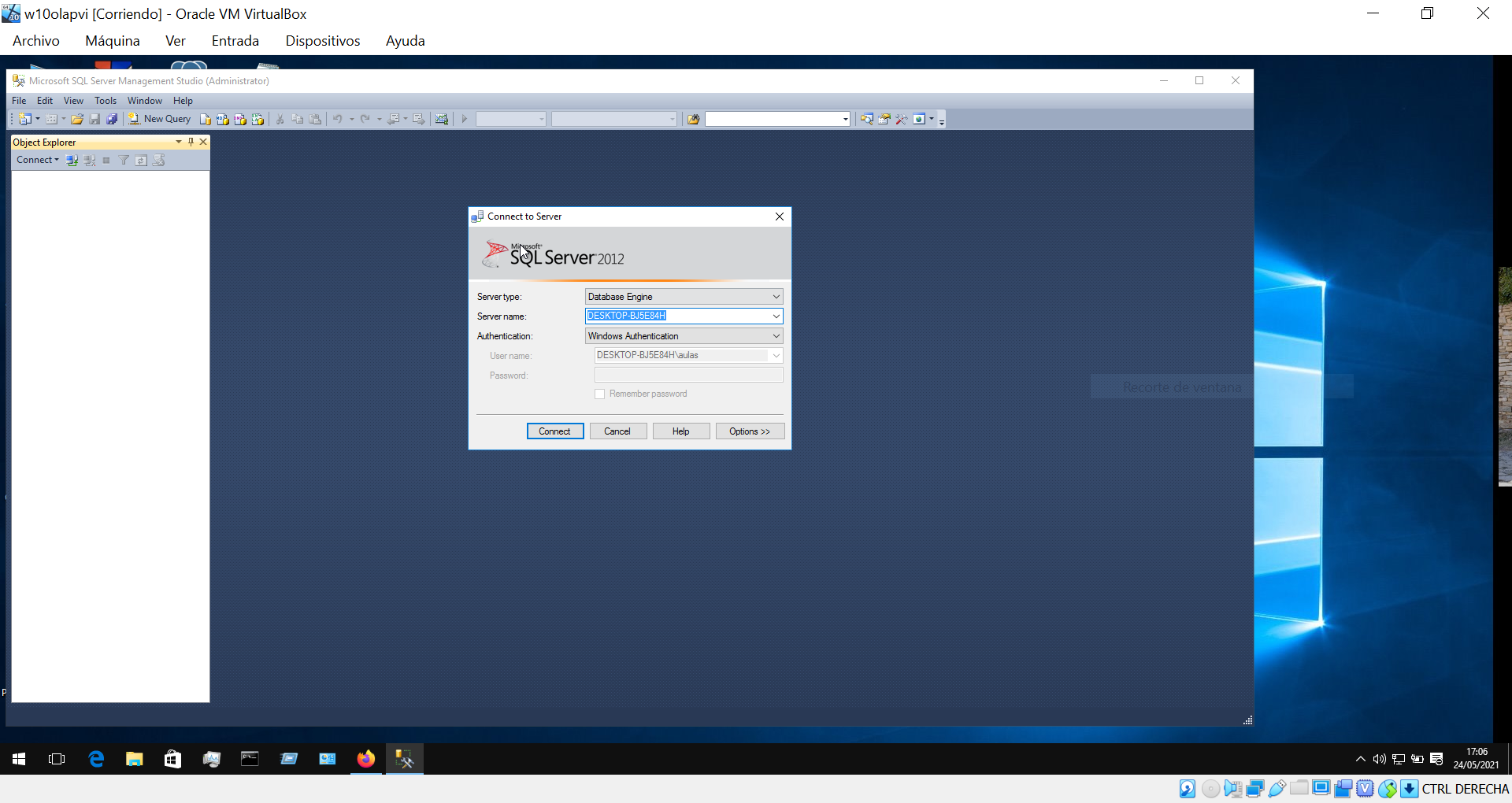
Tareas con Integration Services

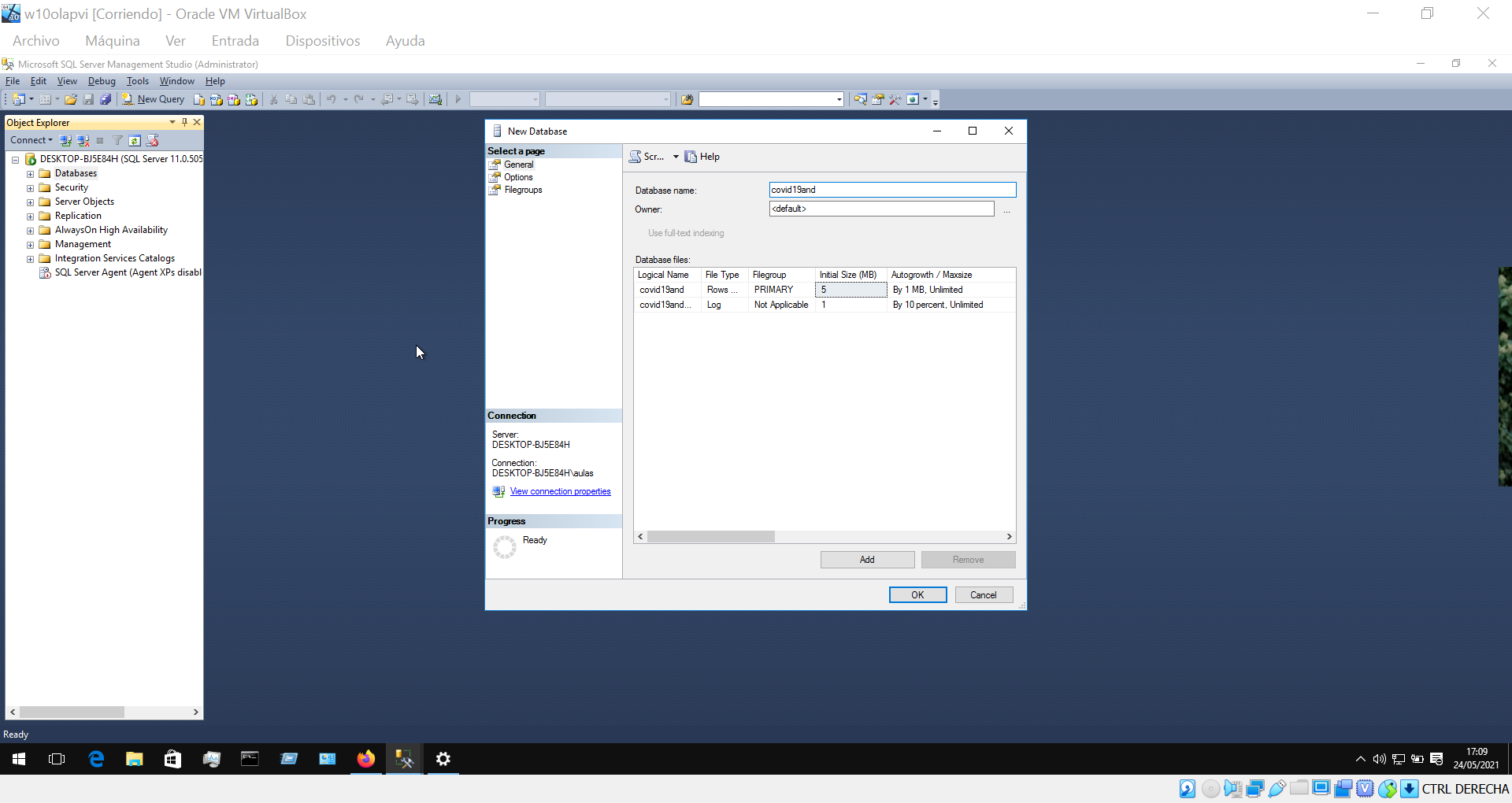
# Tareas en SSMS

1.- CREAR LA BASE DE DATOS

1.1.- CONEXION CON EL SERVIDOR



1.2- crear la bd

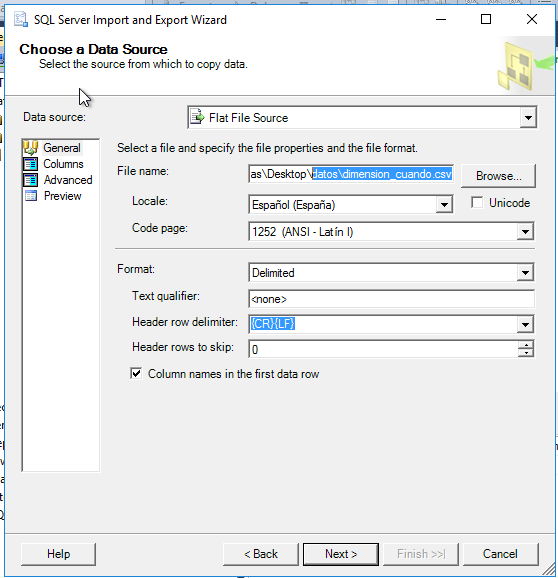


2. CREAR TABLAS

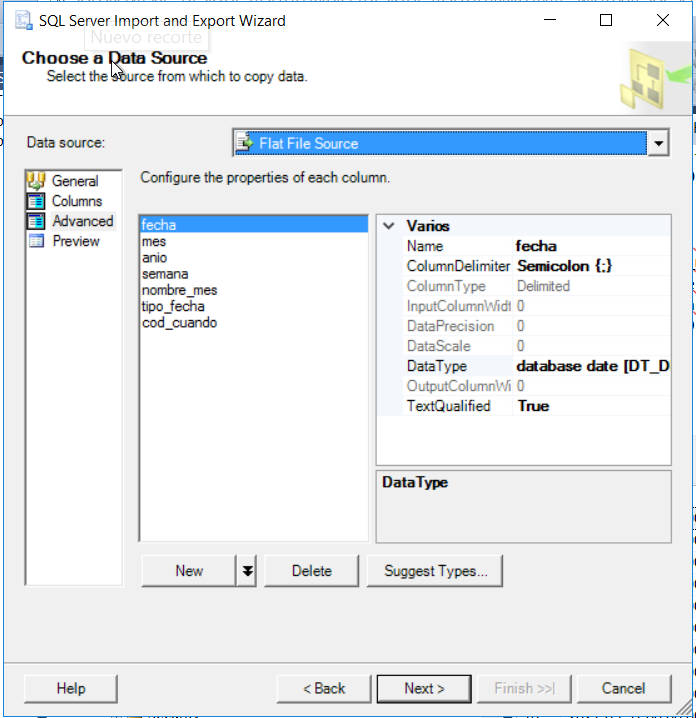
Debido a que la importación de las tablas de las dimensiones y hechos es siempre la misma, pongo aquí el ejemplo de importación de la dimensión Cuándo, y muestro seguidamente el resultado exitoso de importación de todas esas tablas, para demostrar que el proceso sigue correctamente.

Asistente de importación: Acceder al menú contextual de la base de datos covid19and 🡪 Tasks 🡪 Import data.

En la ventana que se nos abre, seleccionamos en el campo “Data source” la opción “Flat File Source” (fuente de archivo plano), pues nosotros tenemos nuestras tablas almacenadas en formato CSV. Justo debajo, en “File name”, pinchamos sobre el botón “Browse”, seleccionamos en la nueva ventana el fichero “**dimensión\_cuando.csv**” en nuestra carpeta **datos** y pichamos en “Abrir”, lo cual nos devolverá a la ventana anterior. El resto de campos de la ventana se autocompletarán una vez seleccionado el fichero CSV.



El siguiente paso, **sin movernos de la pantalla actual**, es seleccionar, en la barra lateral de la izquierda, la pestaña “Advanced”, para poder definir los tipos de datos de las columnas de origen (NOTA al pie: esto hay que hacerlo aquí porque de la forma q indicaba la P4 daba error, pues definíamos los tipos de datos de salida, y no los de entrada como aquí). La ventana se ha actualizado y aparece lo siguiente:



Donde podemos ver una lista con todas las columnas que ha detectado en el fichero CSV y los tipos por defecto que le asigna (cadenas de caracteres, string). Como es lógico, no todas nuestras columnas son string, ni es correcto que las almacenemos en la Base de Datos como strings, sino que debemos **cambiar los tipos de datos** de origen para que el resto de pasos de importación se hagan correctamente. Para ello, explotamos al máximo las posibilidades que ofrece SQL Server y pinchamos sobre “Suggest types…” (que analizará el contenido de cada columna y le asignará el tipo más adecuado), se abre una ventana, no modificamos nada y pulsamos directamente en OK, y ya podemos ver que se han cambiado los tipos de datos de las columnas:

fecha: date

mes: int

anio: int

semana: int

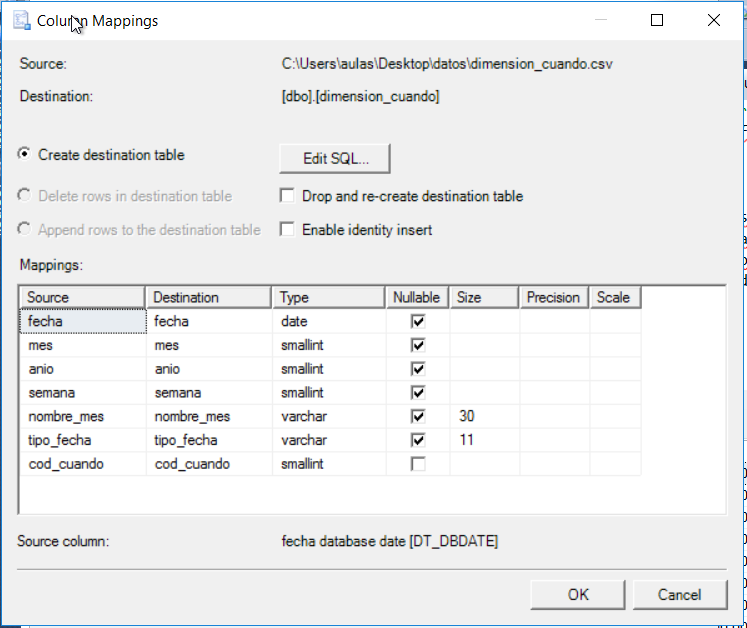
nombre\_mes: string

tipo\_fecha: string

cod\_cuando: int

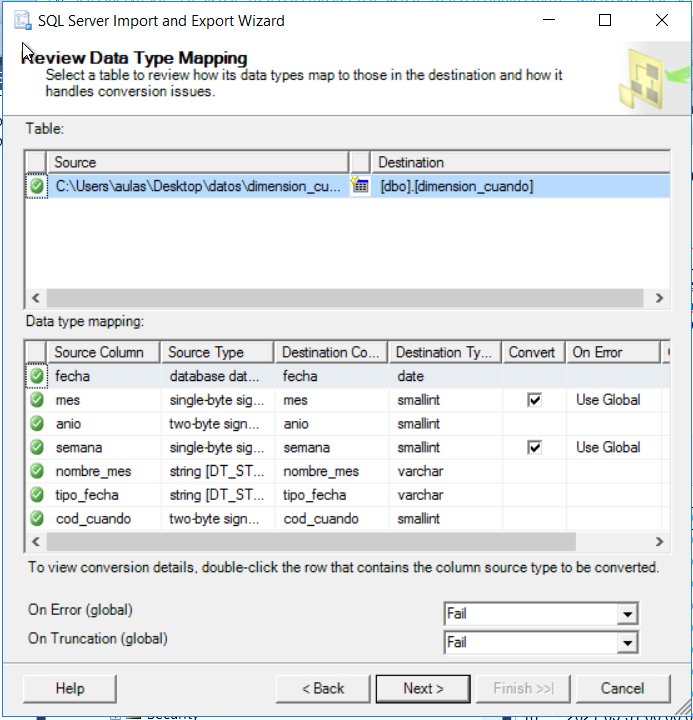
A continuación, pulsamos en “Next” para avanzar a la siguiente pantalla, en la cual solo tenemos que asegurarnos que el campo *Database* tiene el valor de nuestra base de datos, ***covid19and***. Pulsamos en Next.

En esta ventana, si pulsamos sobre *Edit Mappings*, podemos ver los tipos de datos y el estado (en general) de las columnas. Aquí podemos decidir si una columna permitirá valores nulos en su dominio; en nuestro caso, solo vamos a prohibir que haya nulos en la llave generada, es decir, en cod\_cuando, con lo que desmarcamos la casilla asociada a cod\_cuando en la columna Nullable, tal y como se muestra en la figura XX:

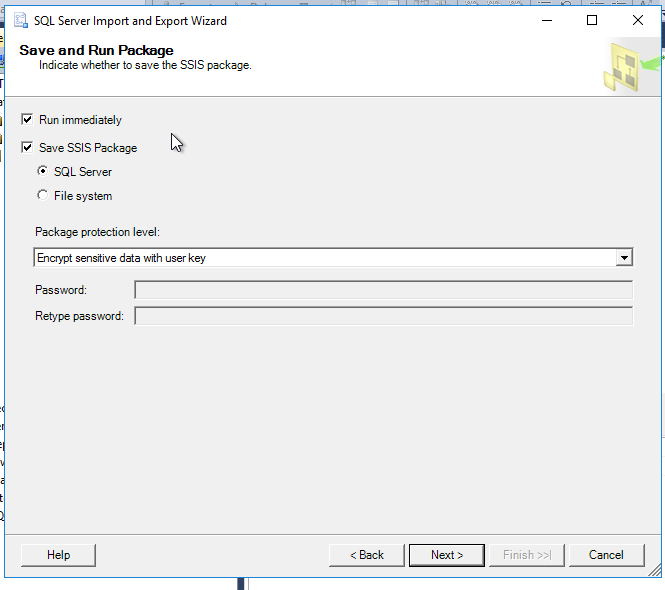


También nos aseguramos aquí que el tamaño de los strings es el que habíamos fijado antes. Hay que marcar la opción “Drop and re-create destination table”, para que luego, en el proyecto de Integration Services, podamos ir actualizando los datos de la tabla ejecutando un paquete de importación, que aglutinará las tareas que estamos haciendo ahora de importación para cada tabla. Más adelante daremos más detalles. Pulsamos OK y en la ventana principal del asistente, pulsamos de nuevo en *Next*.

La siguiente pantalla es muy importante, pues sirve para comprobar que la configuración de las columnas que hemos hecho (tipo de datos, principalmente) es correcta. En nuestro caso, tal y como se muestra en la figura YY, todo ha ido bien, pues tenemos el símbolo (check, V) al lado de cada columna, lo cual anticipa una previsible importación satisfactoria. Pulsamos en Next para pasar a la siguiente pantalla, donde solo nos quedan algunos pasos “de trámite” (figura XX).



Siguiendo con el guión [], tenemos que dejar seleccionadas las dos opciones que ahí aparecen, es decir, “Run inmediately” y “Save SSIS Package”. La primera opción obliga al importador a ejecutar la tarea al momento, una vez se lo indiquemos; la segunda la necesitamos para poder actualizar los datos de las tablas en en nuestro proyecto de SSIS (Integration Services). El nivel de protección del paquete no es necesario modificarlo, podemos dejar el que aparece por defecto. Pulsamos en *Next* de nuevo.

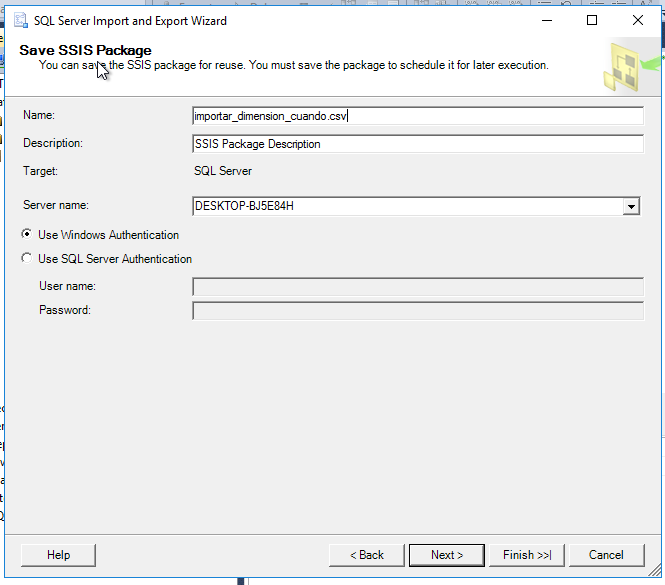


En la siguiente pantalla solo tenemos que fijar el nombre del paquete que almacenará las tareas de importación que estamos configurando en el presente asistente, y que por convenio yo lo llamaré de la siguiente forma:

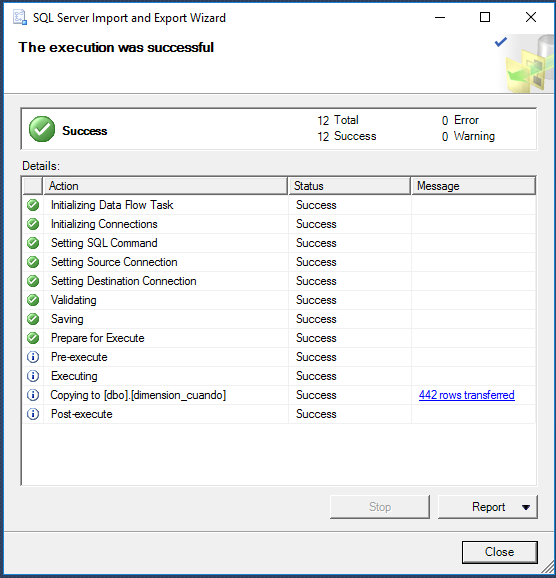
importar\_dimension\_<nombre de dimensión>

importar\_hechos\_<nombre de los hechos>

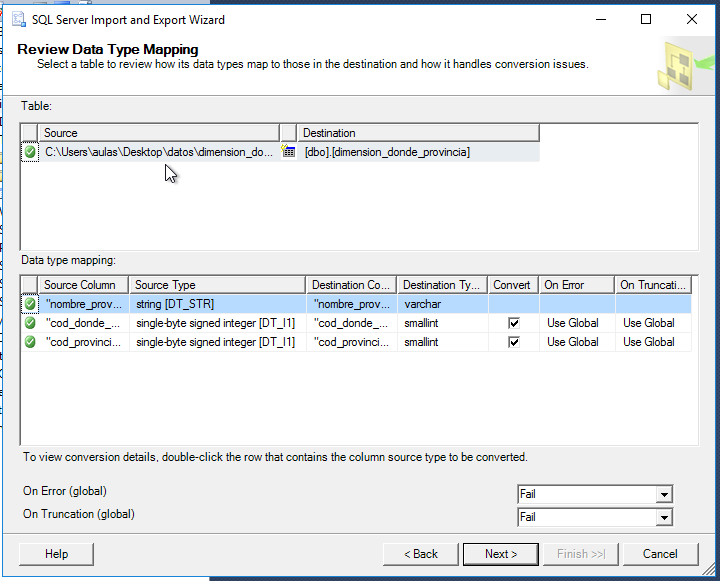
por tanto, en nuestro caso, llamaremos al paquete importar\_dimension\_cuando (ver figura XX). La extensión del paquete no hay que especificarla, la añadirá el sistema al generarlo. Pulsamos en *Next* para ir al último paso.



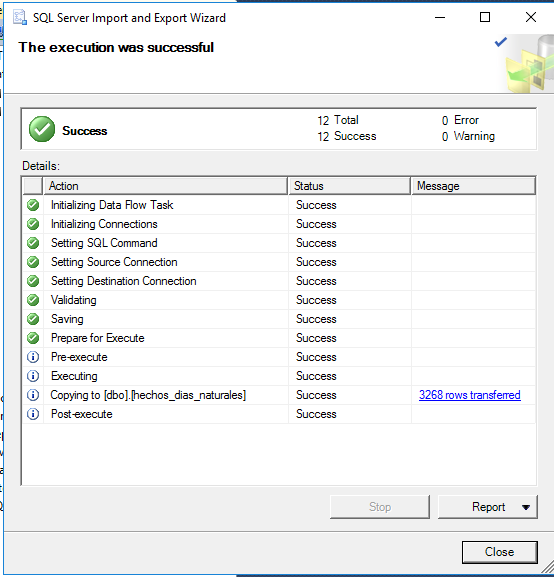
La siguiente pantalla es un resumen sobre la importación. Lo revisamos por si tenemos que corregir algo y si todo lo vemos bien, pulsamos en Finish para que se inicie la importación. El resultado de la importación ha sido satisfactorio, tal y como nos indica el programa tras acabar y que mostramos en la figura X. Es posible que nos dé un error al intentar eliminar una tabla, pero si indica a la vez que ha transferido las filas, y no da más errores, podemos ignorarlo, ya que este error sucede porque antes habíamos marcado al opción “Drop and re-create destination table” por lo que, la primera vez, intenta eliminar la tabla pero aún no existe.



Para la dimensión **Dónde – Provincia**, seguimos el mismo proceso que para la dimensión Cuándo, y el estado final antes de la importación es el que se muestra en la figura XX.



A continuación importamos la **tabla de hechos de días naturales** siguiendo el mismo procedimiento que antes y el nombre del paquete de importación será importar\_hechos\_dias\_naturales. La demostración de que la tabla se ha importado correctamente se aporta en la figura XX.



Hagamos una pequeña recapitulación de tareas, una vez que tenemos el primer “cubo” (hechos + dimensiones asociadas) cargado.

1 Hemos cargado las dos tablas de las dimensiones: Cuándo y Dónde-Provincia

2 Hemos cargado la tabla de hechos de días naturales.

3 Tenemos todas estas tareas guardadas como paquetes independientes de SSIS, que nos serán imprescindibles para la siguiente fase.

A continuación se deberían de importar el resto de tablas de dimensiones y de hechos. Para no saturar en exceso este documento, y evitar una descripción repetitiva del proceso de importación de una tabla, que no aportaría nada nuevo, simplemente mostraré más adelante una captura de SSMS con todas las tablas en la Base de Datos.

Para las tablas de **dimensiones** que faltan:

* importar\_dimension\_donde\_ds.dtsx
* importar\_dimension\_donde\_municipio.dtsx
* importar\_dimension\_quien.dtsx
* importar\_dimension\_residencias.dtsx

Para las tablas de **hechos** que faltan:

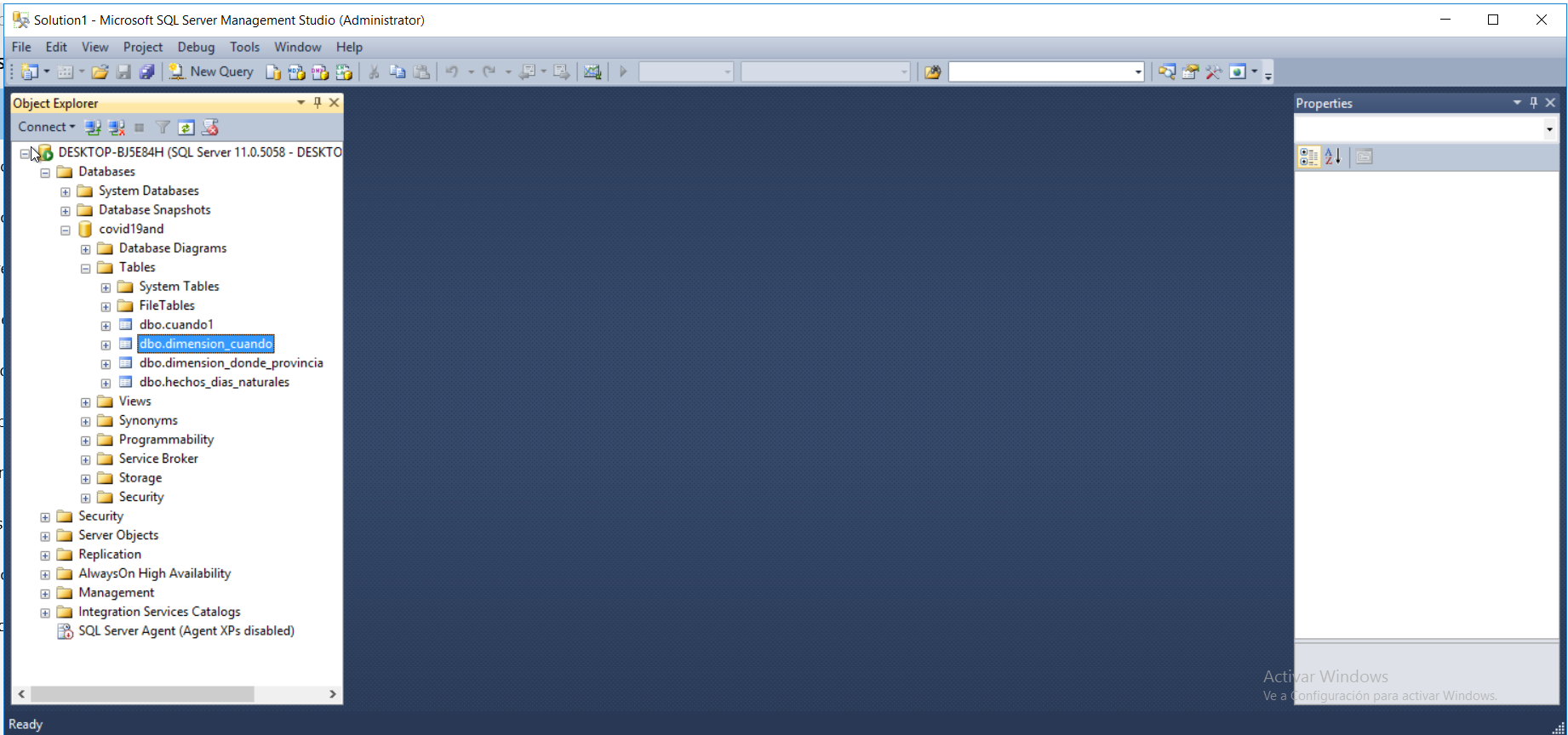
* importar\_hechos\_municipios.dtsx
* importar\_hechos\_residencias.dtsx
* importar\_hechos\_residencias\_edad\_sexo.dtsx
* importar\_hechos\_vacunas.dtsx

# TAREAS EN SSDT

SSDT es el software de SSIS que nos va a permitir, a efectos prácticos, dejar la base de datos preparada para que posteriormente se realicen las consultas oportunas desde el cliente sobre las dimensiones/hechos almacenadas en la BD.

Para trabajar en SSDT tenemos que crear un proyecto. Seguimos los pasos dados en el guión (sección 3.1 de [p4]), donde también se indica que tendremos el paquete asociado a una única solución de SSDT, pues no vamos a usar otros componentes de este software más allá de los que integran el proyecto en sí mismo.

Una vez creado, nos aparecerá una ventana principal como la de la figura XX:



Lo que tenemos que hacer ahora es añadir los paquetes de importación (que habíamos salvado antes en SSMS para cada tabla) a nuestro proyecto.

Una vez tenemos todos los paquetes de importación correctamente generados, pulsamos en su menú contextual, ahí en Execute Package y comprobamos que se ejecutan correctamente. Hecho y comprobado esto, ya tenemos listo el conjunto de tareas asociado a gestionar las tablas de la Base de Datos SQL Server y podemos pasar al último paso importante dentro del diseño de nuestra base de datos: definir el esquema multidimensional.

# Definición del esquema multidimensional en SSDT

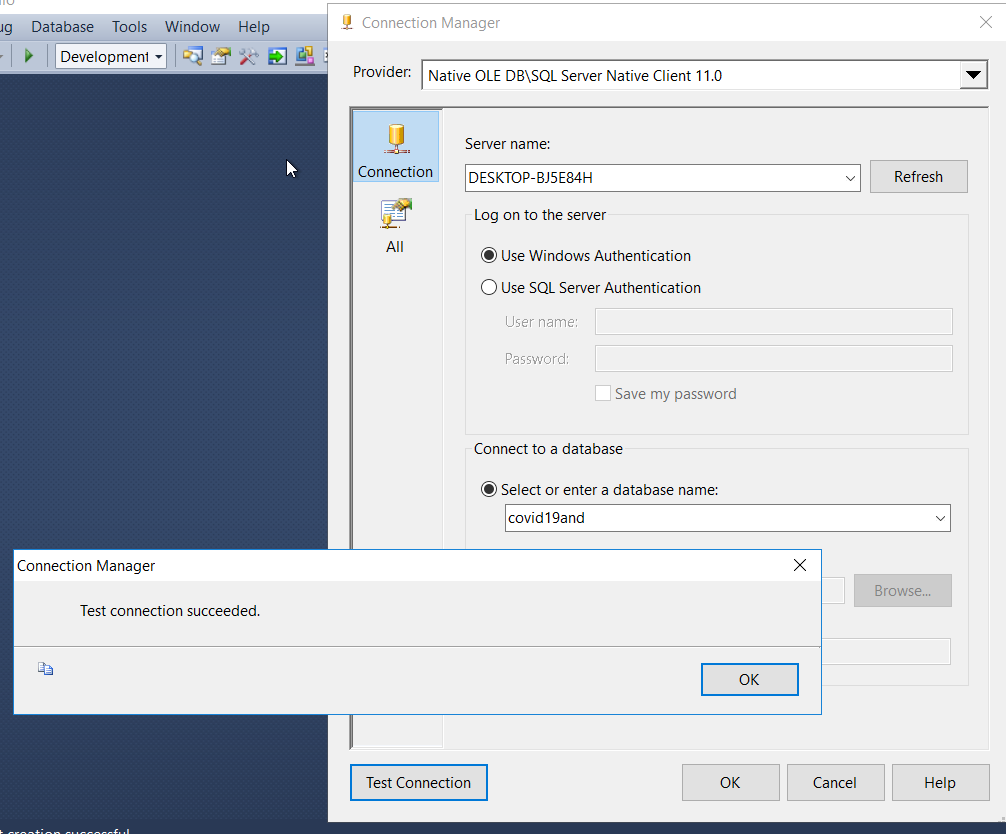
Para hacer esta tarea tenemos que recurrir al guión de la Práctica 6 de SMD ([]), repasar lo realizado en los puntos 1 y 3 y adaptarlo a nuestro esquema multidimensional.

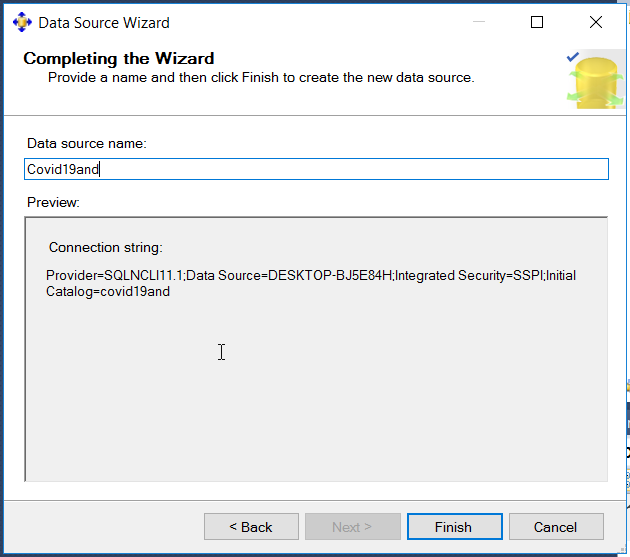
En primer lugar, creamos el proyecto siguiendo las instrucciones de la sección 3.1.1. Tras crearlo, se inicia SSDT con el proyecto nuevo abierto, tal y como muestro en la figura XX. El siguiente paso es definir la conexión para que se pueda acceder a los datos alojados en la BD de SQL Server (*Data Sources*) y podamos definir a partir de ellos (de las tablas relacionales) nuestro esquema multidimensional. Siguiendo los sencillos pasos del guión, configuramos:

1.- la conexión a los datos y probamos que en efecto funciona (ver figura X);

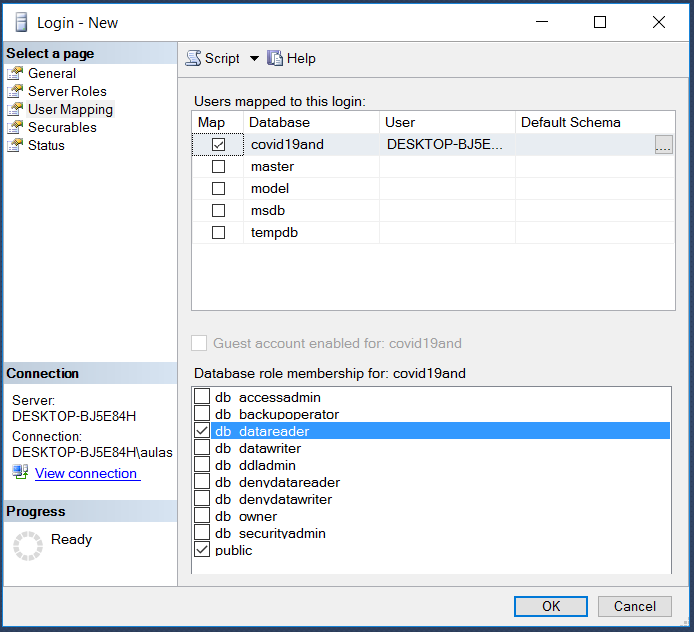
2.- el mecanismo de autenticación para acceder a la BD

Y finalmente asignamos un nombre a esta fuente de datos, que la llamaremos igual que la Base de Datos:



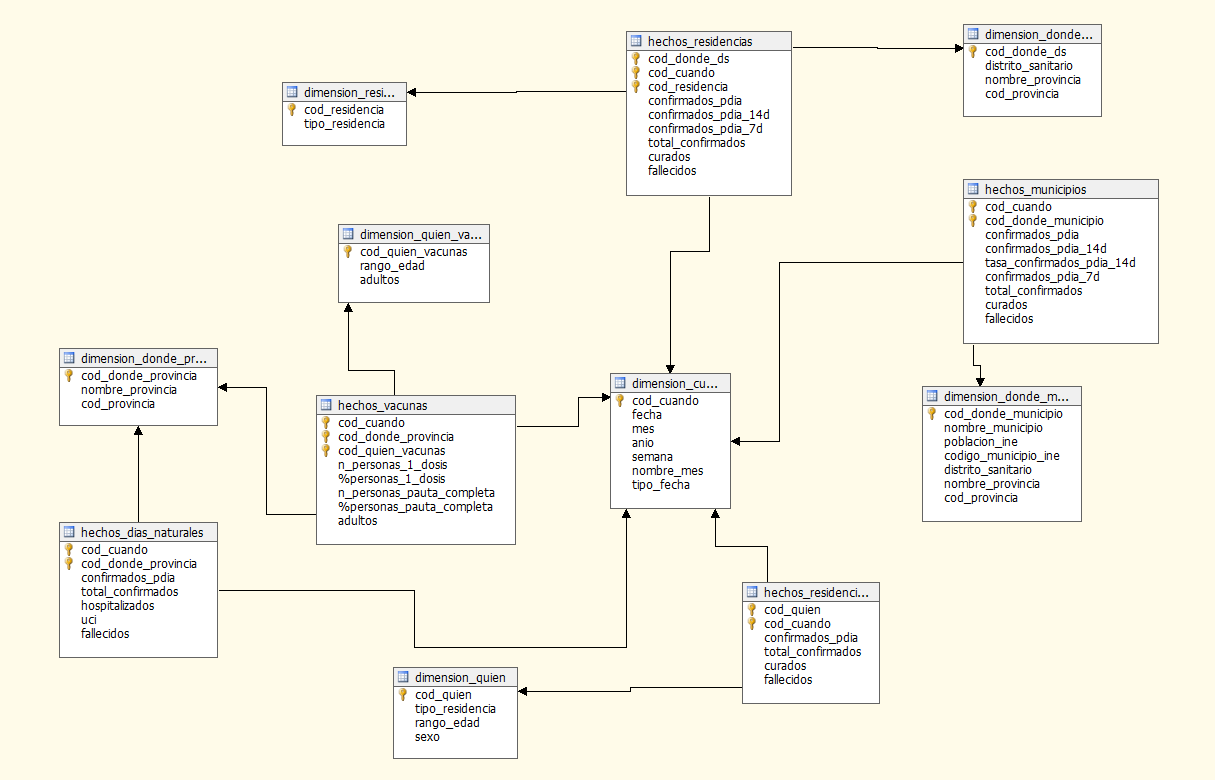


El siguiente paso es el de la sección 3.3.2, que se realiza íntegramente como indica el guión, teniendo en cuenta que el usuario a través del cual se consultará la Base de Datos ya está creado en la Máquina Virtual que estamos usando (el profesor indica que ya lo ha creado dentro de esta máquina, por lo que no tenemos que crearlo nosotros). El resultado final es correcto y es el de la figura X:



Definir la vista sobre la que trabajar con los datos fuente

Seguimos la sección 3.4 tal y como se explica en el guión y el resultado es que el asistente ha detectado perfectamente todas las llaves primarias de las tablas y por tanto puede así detectar las llaves externas, que relacionan (como es lógico) los hechos con las dimensiones que llevan asociadas. Este resultado se muestra en la figura X, una de las imágenes más ilustrativas de las que podamos añadir en esta memoria sobre la representación a nivel relacional de nuestra base de datos:

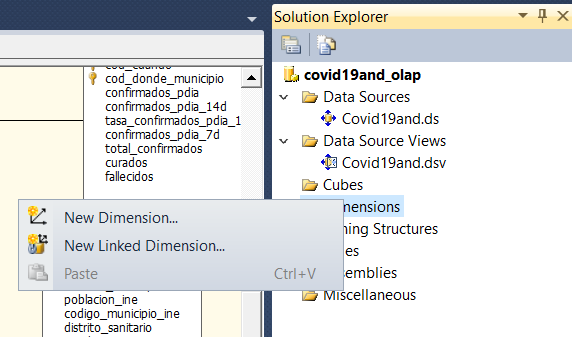


Definición de la dimensión Dónde – Provincia

Para ilustrar cómo definimos alguna dimensión, nos decantamos por una que es sencilla y que, de entrada, no debería darnos muchos problemas (no posee campos especiales, como la fecha, valores decimales, etc).

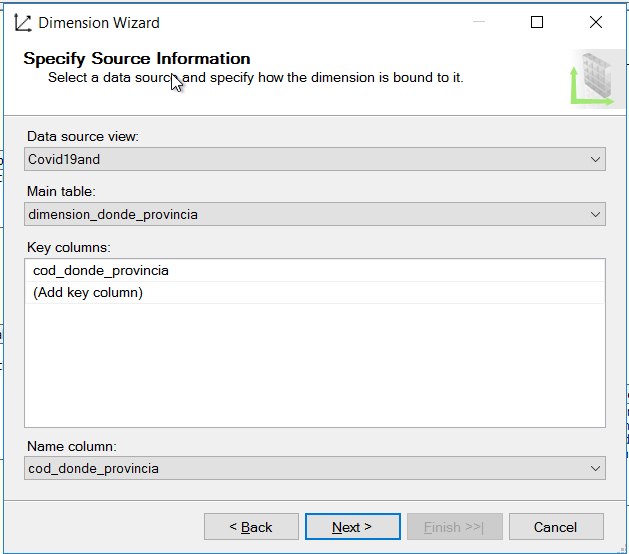
Podemos ir siguiendo los pasos del guión para definir la dimensión Dónde, aunque evidentemente lo adaptaremos a nuestra dimensión. Los pasos más importantes los explico a continuación:

1.- Abrir el asistente de definición de dimensiones [PIE: en el menú contextual del campo *Dimension* dentro del proyecto (ver figura X), seleccionar New Dimension]

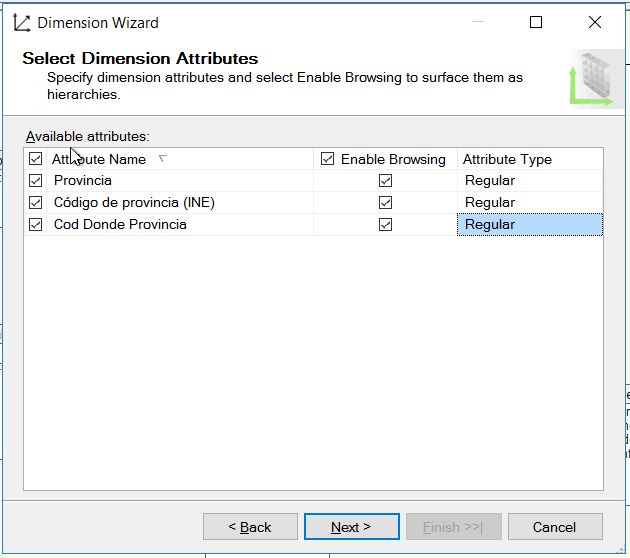


2. En la pantalla siguiente, indicamos que vamos a crear la dimensión a partir de una tabla individualizada para ella, y no vamos a extraer los datos de otra (figura ).

3. Pasamos de pantalla y seleccionamos la tabla de la Base de Datos asociada a la dimensión que vamos a definir (dimensión\_donde\_provincia), y vemos justo debajo que el asistente ha detectado la clave primaria de la tabla. Comprobamos que sea correcta (aunque lo será, porque en la figura [esquema global con todas las tablas relacionadas al crear la vista] ya vimos que la herramienta detectó todas las llaves y relaciones correctamente) y pulsamos en Next.



4. Este paso es importante: tenemos que seleccionar qué atributos queremos que se usen (en nuestro caso, todos) y tenemos que activar la opción Enable Browsing para permitir que se usen los atributos al construir las dimensiones. Adaptamos los nombres de los atributos para que sean comprensibles, pues serán los nombres que aparecerán en los informes. El resultado se muestra en la figura .



5. La siguiente pantalla es un resumen del proceso, donde tenemos que definir el nombre de la dimensión: Dónde – Provincia, en nuestro caso.

6. El siguiente paso es desde el

5.

## Definición de la dimensión Cuándo

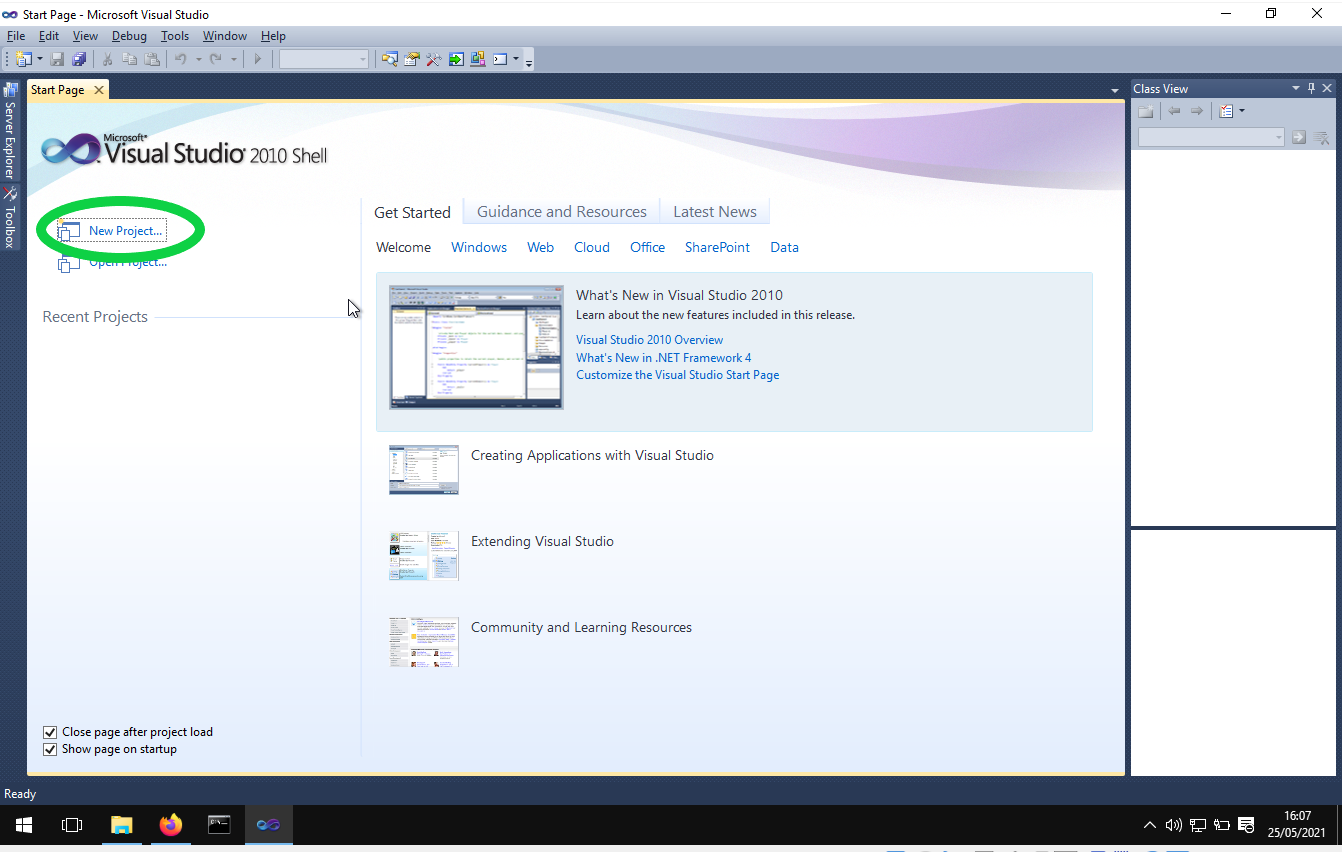
He escogido esta dimensión la primera de todas por ser, en base a la literatura usada en el proyecto y la propia experiencia personal, la que más problemas suele dar. Además hay que tener en cuenta que es una dimensión lentamente cambiante, pues incluye las fechas de notificación de los datos, tanto de los días naturales como de los hábiles (tan y como se especificó en la Metolodogía de Kimball).

Así las cosas, basta con seguir los pasos indicados en el guión para definir las dimensiones.

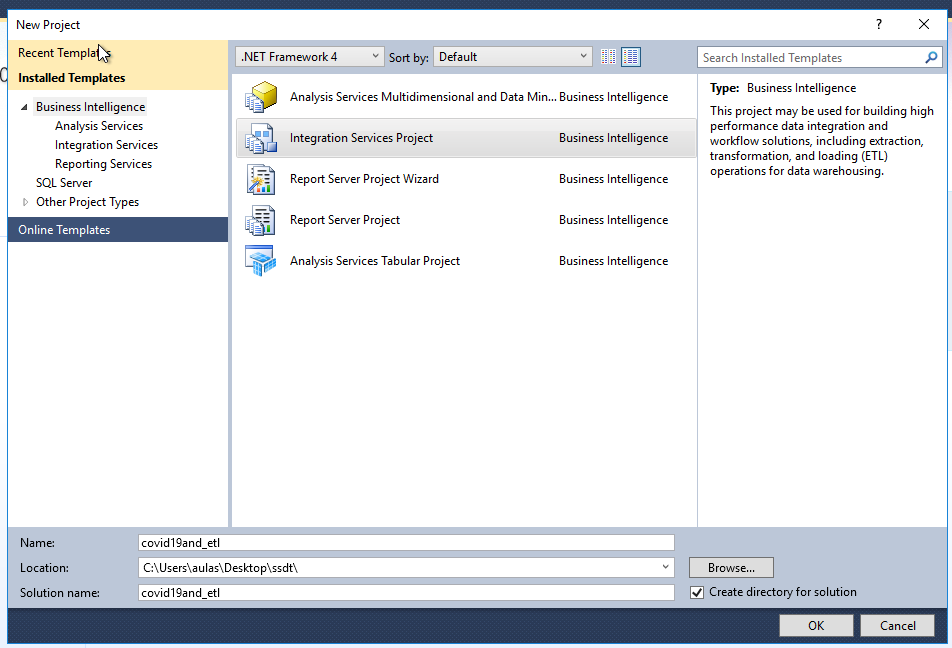
-----------------

3.- USO DE SSDT

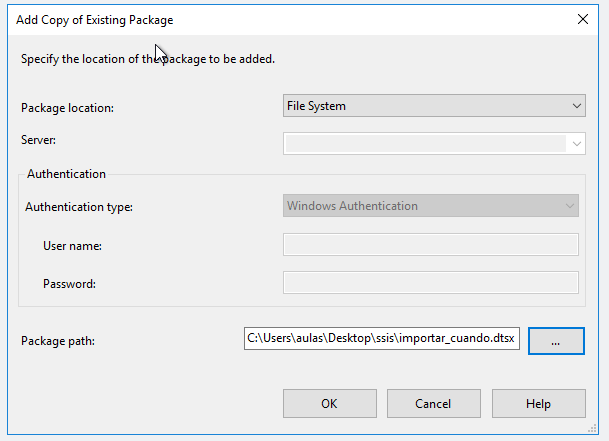
1.- Crear proyecto de Integration Services



En la ventana que se abre seleccionamos el proyecto que más se adecua a nuestro objetivo, que es “Integration Services” (ver P4SMD) y lo añadimos a una solución que se crea ahora, pues antes no teníamos ninguna (recordar concepto de “solución”):



Una vez creado el proyecto, tenemos que ir importando los paquetes .dtsx que se nos generaban al ir importando cada tabla en SSMS, para poder trabajar con estas. Para ello, tenemos que irnos al panel lateral derecho “Solution Explorer” 🡪 “SSIS Packages” 🡪 “Add existing packages” y ahí se nos abre una ventana donde tenemos que seleccionar el paquete importar\_\*.dtsx que se nos generaba al importar las tablas (ver figura ):



Una vez importado el paquete asociado a la tabla que toque (en el caso de la figura anterior, para la dimensión *Cuándo* -fichero importar\_cuando.dtsx-), podemos abrirlo, y lo que se nos muestra