

**Máster universitario de Ciencia de Datos**

**Prueba de Evaluación Continua – PEC1**

**Diseño y uso de bases de datos analíticas – Introducción a las bases de datos analíticas y toma de contacto con la máquina virtual (VDI)**

Autor:

Mario Ubierna San Mamés

|  |
| --- |
| Índice de Contenido |

[Índice de Contenido 3](#_Toc66889693)

[Índice de tablas 4](#_Toc66889694)

[Índice de ilustraciones 5](#_Toc66889695)

[1. Parte teórica 6](#_Toc66889696)

[1.1. Pregunta1 6](#_Toc66889697)

[1.2. Pregunta 2 6](#_Toc66889698)

[1.3. Pregunta 3 6](#_Toc66889699)

[1.4. Pregunta 4 6](#_Toc66889700)

[1.5. Pregunta 5 6](#_Toc66889701)

[2. Parte práctica 7](#_Toc66889702)

[2.1. Ejercicio 1: Configuración del entorno VDI 7](#_Toc66889703)

[2.2. Ejercicio 2: Validación de la BBDD 7](#_Toc66889704)

[2.3. Ejercicio 3: Validación de la aplicación de ETL 7](#_Toc66889705)

[2.4. Ejercicio 4: Microsoft SQL *Analysis Services* 7](#_Toc66889706)

[3. Bibliografía 8](#_Toc66889707)

|  |
| --- |
| Índice de tablas |

|  |
| --- |
| Índice de ilustraciones |

|  |
| --- |
| Parte teórica |

## Pregunta1

A

## Pregunta 2

A

## Pregunta 3

A

## Pregunta 4

A

## Pregunta 5

A

|  |
| --- |
| Parte práctica |

## Ejercicio 1: Configuración del entorno VDI

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

Antes de comenzar cabe destacar que los pasos que voy a seguir a continuación, están determinados por los apuntes que se nos han proporcionado en esta asignatura [1]. Para poder realizar la configuración del entorno VDI (*Virtual Desktop Infrastructure*), lo primero que tuve que hacer fue instalar *Citrix Workspace App*, tal y como se muestra en la siguiente captura:



Ilustración - Descarga de Citrix Workspace App [2].

Una vez descargado e instalado ya pude acceder al portal de la UOC para entrar poder entrar en la máquina virtual.

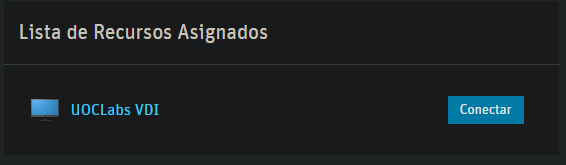


Ilustración - Vista principal de la VDI asignada.

Como podemos apreciar en la anterior ilustración, vemos que tenemos una VDI asignada con el fin de realizar estas prácticas, haciendo click sobre el botón “Conectar” accedemos a la máquina virtual:



Ilustración - Pantalla general VDI.

En la siguiente ilustración podemos ver de forma detalla la información técnica sobre esta VDI:



Ilustración - Información técnica sobre la VDI.

Una vez que ya estamos dentro de la máquina virtual, damos acceso de lectura y escritura para que sea más fácil intercambiar archivos entre el equipo local y la máquina virtual.

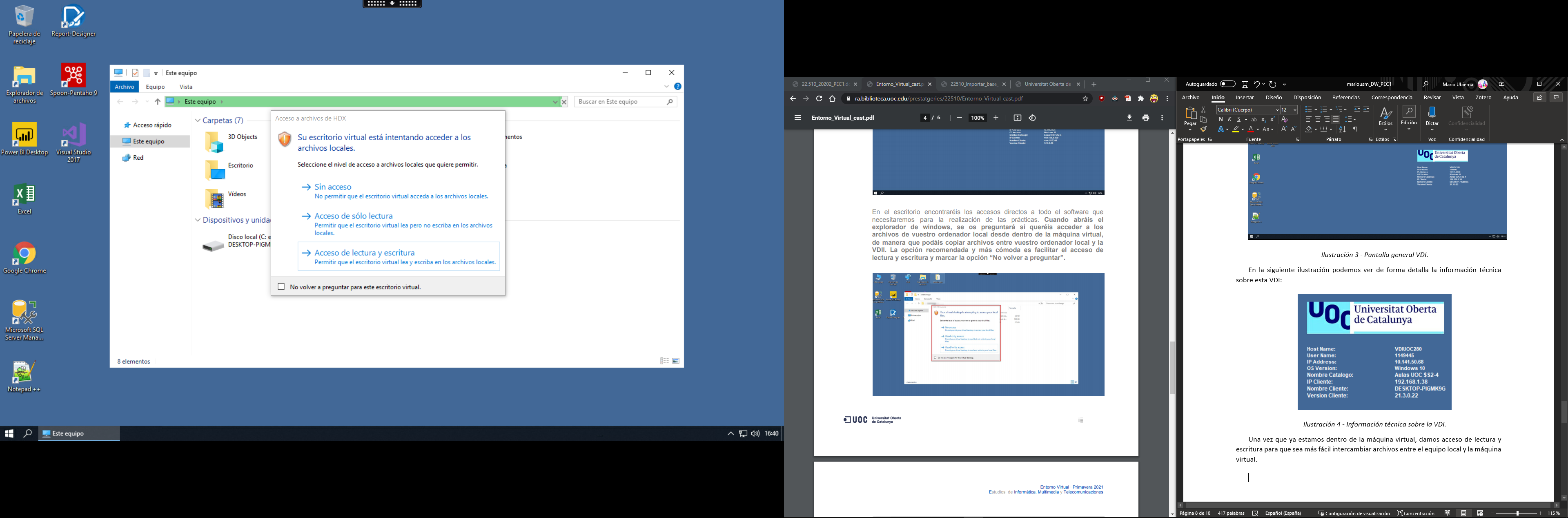


Ilustración - Permisos de lectura y escritura entre la VDI y el equipo local.

Posterior a este punto tenemos que realizar la configuración de la base de datos, para ello por no hacer dos veces el mismo ejercicio, queda explicado en el ejercicio 2.

Para concluir, hay que configurar *Visual Studio*, lo primero que debemos hacer es importar la estructura de base de datos. Para ello, abrimos *Visual Studio*, seleccionamos del menú la opción *Archivo* y clickamos sobre *Nuevo proyecto*. Nos aparecerá la siguiente ventana:

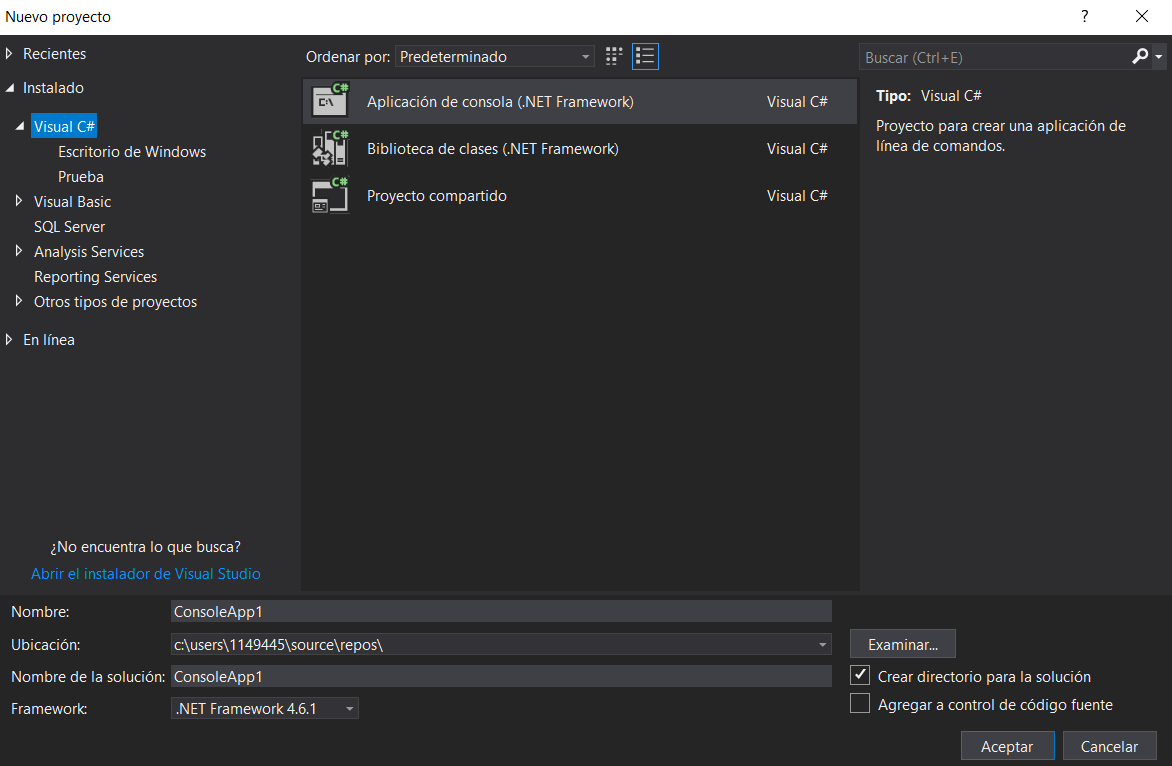


Ilustración - Menú Archivo > Nuevo > Proyecto.

Seleccionamos *Analysis Services* > *Multidimensional* > *Importar del servidor* tal y como vemos en la siguiente captura:

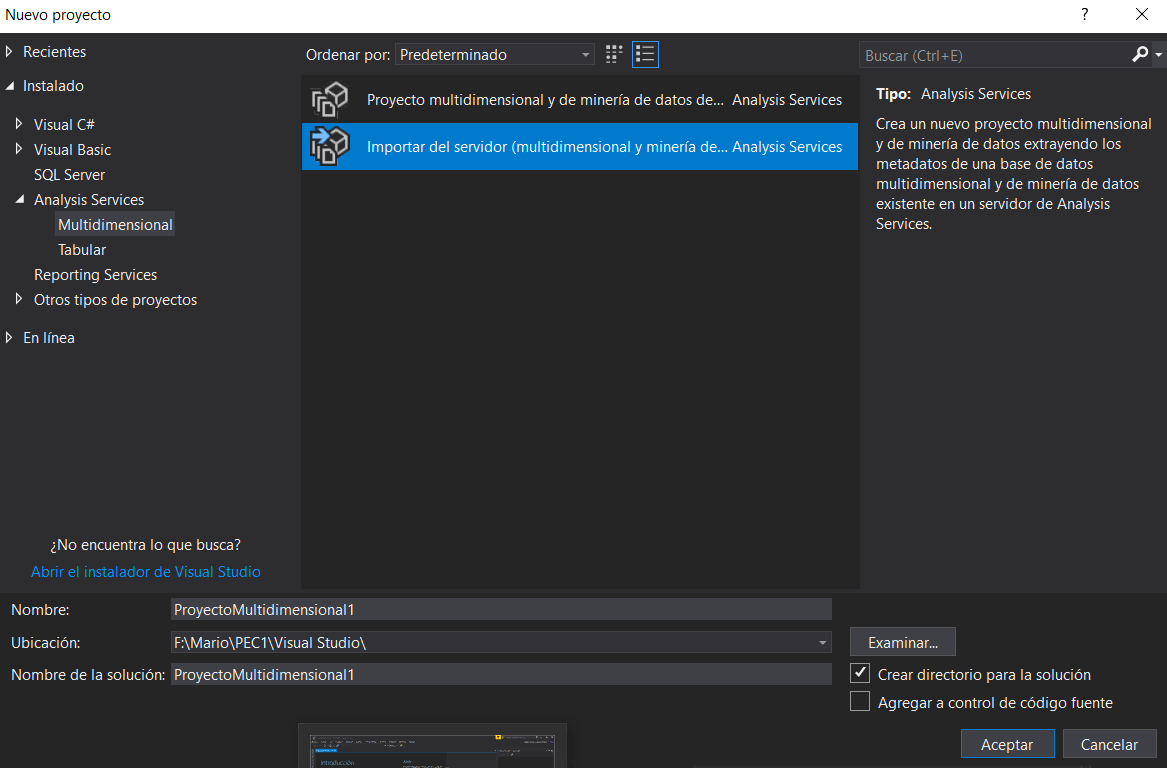


Ilustración - Analysis Services > Multidimensional > Importar del servidor.

Nos aparecerá un asistente para importar la base de datos:

## Ejercicio 2: Validación de la BBDD

*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.*

### Conectar al servidor de base de datos SQL Server

La versión de *Microsoft SQL Server Management Studio 2017* la podemos ver una vez que hemos entrado del del gestor de bases de datos, para ello en la base de datos asignada ejecutamos la siguiente consulta:

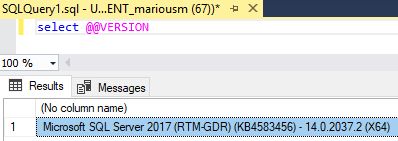


Ilustración - Versión de Microsoft SQL Server Management Studio 2017.

Como podemos observar, vamos a trabajar con la versión del 2017, la cual se encuentra en la versión 14.0.2037.2.

Por otro lado, la pantalla de conexión inicial de la consola *SQL Server Management Studio (SSMS)* es la siguiente, en ella tenemos que introducir el nombre del servidor, seleccionamos la autentificación a través de SQL, por lo que haremos logging con nuestro usuario (STUDENT\_usuariouoc) y la contraseña ($d3f4ult$):

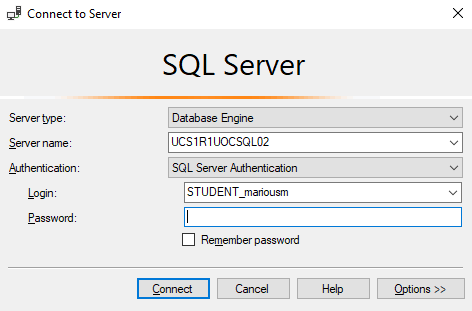


Ilustración - Pantalla de conexión inicial de la consola SQL Server Management Studio (SSMS).

Finalmente, una vez iniciada la sesión podemos ver el explorador de objetos del SSMS:



Ilustración - Explorador de objetos del SSMS.

### Configurar la base de datos DB\_loginuoc.

Lo primero que debemos realizar es configurar el parámetro *Recovery Model* como “simple”, para ello accedemos a la base de datos y haciendo click derecho sobre la instancia de la misma, nos aparecerá un menú contextual que pondrá “*Options”*, haciendo click sobre él nos aparecerá la siguiente ventana:

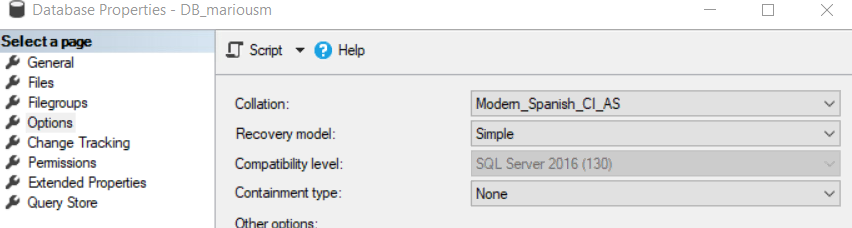


Ilustración - Modificación Recovery model a simple.

Desactivamos la actualización automática de estadísticas:

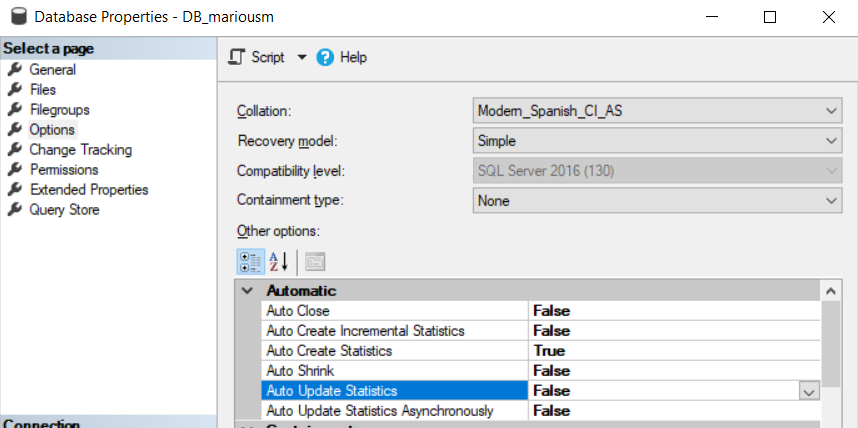


Ilustración - Modificación actualización automática de estadísticas.

### Crear una tabla con el nombre T\_PEC1

Creamos una nueva tabla en la base de datos que se nos proporciona, para ello Campo1 tiene que ser de tipo *varchar(50)* y permitir valores nulos, por otro lado, el Campo2 será de tipo *int* y no permitirá valores nulos. Accedemos a la base de datos y haciendo click derecho sobre ella pulsamos sobre la acción de crear tabla, y nos aparecerá la siguiente ventana:

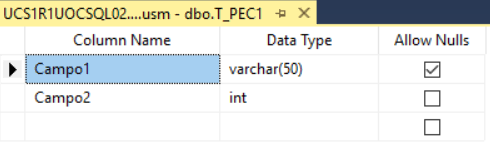


Ilustración - Creación de la tabla T\_PEC1.

Para comprobar que se ha creado correctamente hacemos una consulta *select* de la tabla, como es lógico no va a tener valores ya que no hemos hecho ningún *insert*:

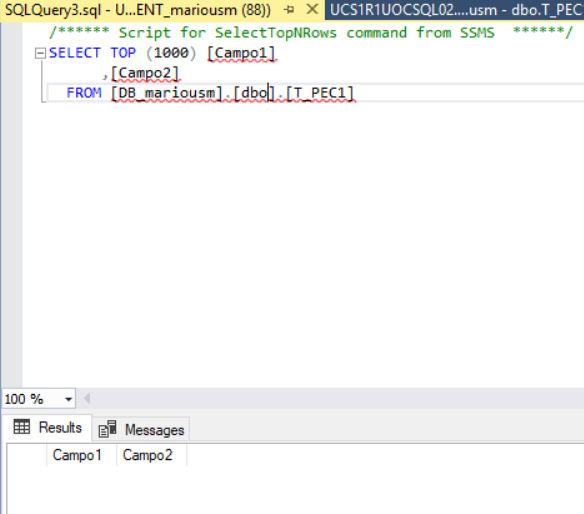


Ilustración - Comprobación de la tabla T\_PEC1.

## Ejercicio 3: Validación de la aplicación de ETL

*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.*

Antes de comenzar con el ejercicio en sí, hay que abrir *Spoon* y capturar su versión:



Ilustración - Versión de Spoon.

Una vez que ya sabemos la versión de *Spoon* que vamos a utilizar, podemos continuar con el ejercicio.

### Análisis preliminar del origen de datos

Lo primero que debemos de hacer es analizar el fichero de datos que se nos proporciona. En un primer análisis debemos ver los campos que se van a cargar:

* **PracNo**: variable de tipo *int*, dicha variable no debe admitir nulos, ya que es un identificador único para cada registro, por lo que debe ser la *primary key*.
* **PracticeName**: variable de tipo *nvarchar*, ya que la información que contiene este campo es texto. Viendo algunos registros, sí que debe permitir valores nulos en dicho campo**.**
* **Address1**: variable de tipo *nvarchar*, estos datos también son de tipo texto y no pueden ser nulo.
* **Address2**: variable de tipo *nvarchar* y no debe admitir nulos.
* **Address3**: variable de tipo *nvarchar*, sí que debe permitir valores nulos.
* **PostCode**: variable de tipo *nvarchar*, ya que son datos alfanuméricos y no debe permitir nulos.
* **LCG**: variable de tipo *nvarchar*, y no debe admitir nulos.
* **RegisteredPatienes**: variable de tipo *int*, ya que es un valor numérico entero y no debe permitir nulos.

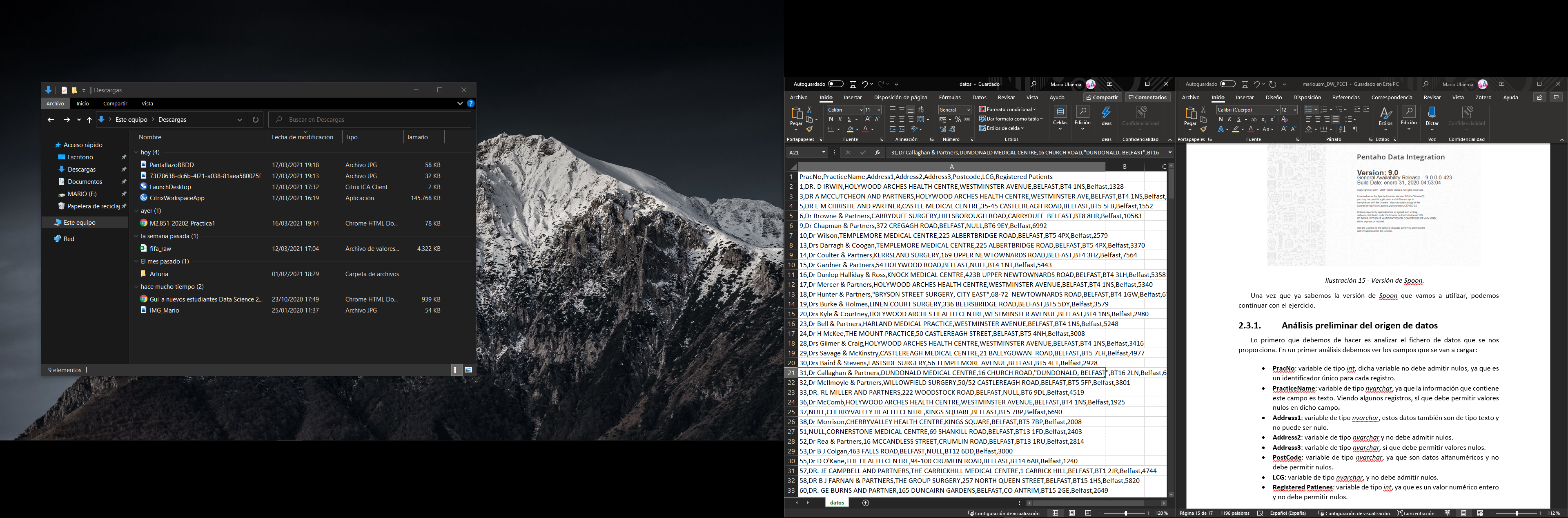


Ilustración - Estructura de los datos.

### Crear una tabla en la BBDD con el nombre T\_DATOS

Para crear la tabla nos vamos a *Microsoft SQL Server Management Studio 2017*, una vez hemos iniciado sesión tal y como se ha explicado tendremos acceso a la base de datos, en nuestro caso se llama “DB\_mariousm”:

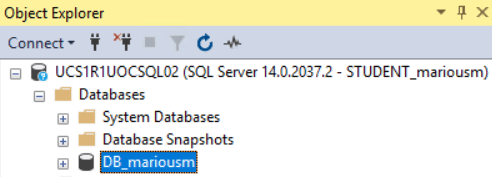


Ilustración - Vista de la base de datos.

Posteriormente, hacemos click con el botón derecho del ratón sobre la opción *Table* y seleccionamos *New > Table:*

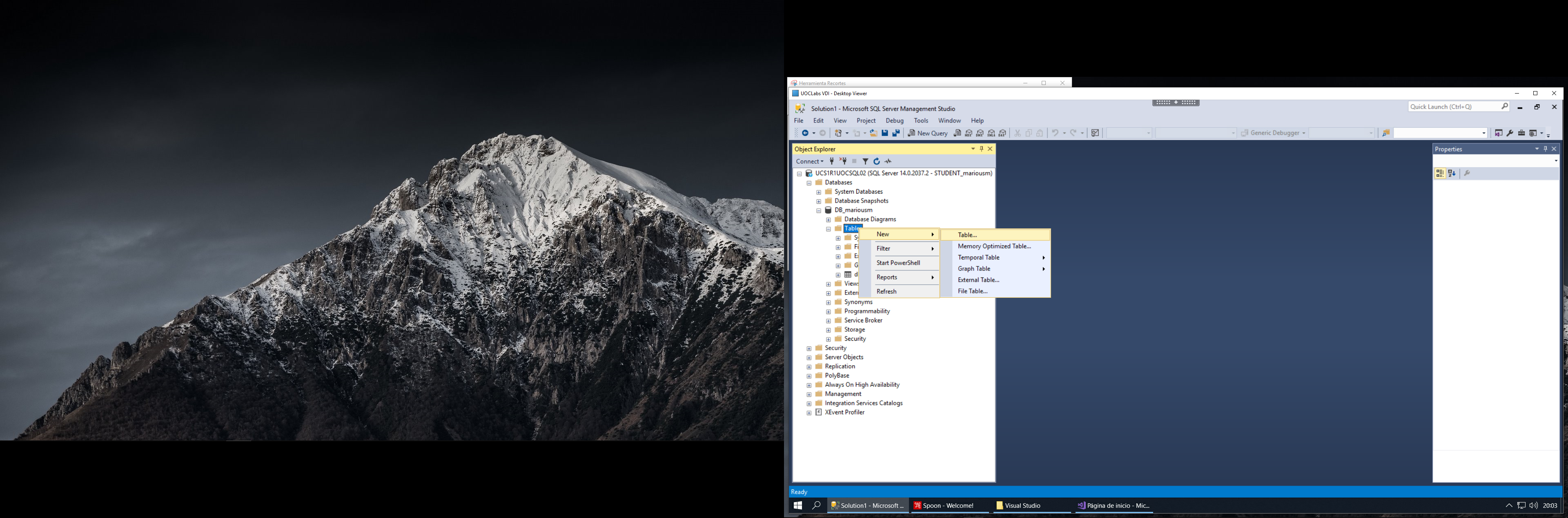


Ilustración - Menú creación de tabla.

Una vez realizado el paso anterior, nos aparecerá la siguiente ventana, en la cual podemos diseñar la tabla según los datos que hemos analizado en el punto anterior:

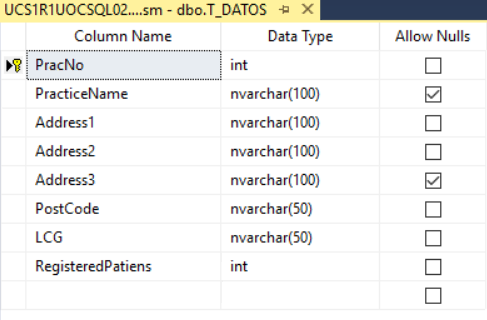


Ilustración - Creación de la tabla T\_DATOS.

### Crear una transformación con *Spoon*

Antes de comenzar, tenemos que realizar una conexión con la base de datos para así posteriormente tener acceso [3], para ello nos vamos a *File > New > Transformation*. Una vez se nos crear la transformación, podemos establecer una conexión a la base de datos, para ello volvemos a seleccionar desde el menú *File > New > Database Connection*, como resultado nos encontraremos con la siguiente pantalla:

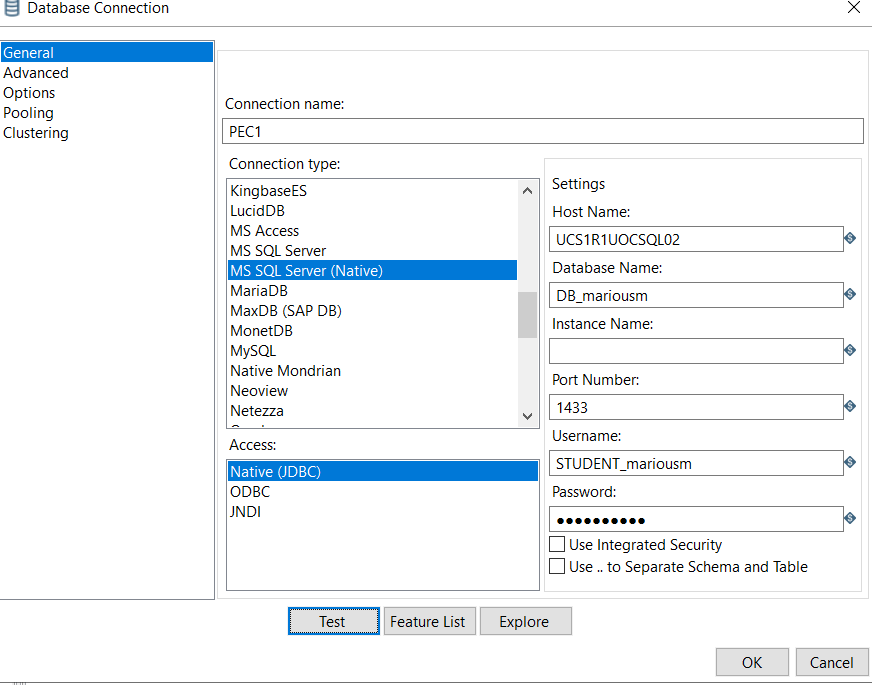


Ilustración - Creación de conexión desde Spoon a SQL Server.

Lo siguiente que debemos hacer es, dar un nombre a la conexión, seleccionar el tipo de conexión (en nuestro caso *MS SQL Server (Native)*), luego escribimos el nombre del *host* (se nos ha proporcionado ese nombre en la práctica), introducimos el nombre de la base de datos, y finalmente el nombre de usuario y la contraseña. Pulsamos sobre el botón *Test* para comprobar si todo ha salido correctamente:

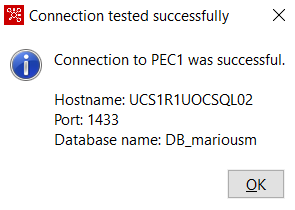


Ilustración - Test conexión de Spoon con SQL Server.

Como podemos apreciar en la anterior imagen, la conexión se ha testeado de forma correcta, por lo que pulsamos el botón Ok y dicha conexión será establecida.

Ahora sí podemos continuar cona la extracción del archivo CSV en *Spoon*,

## Ejercicio 4: Microsoft SQL *Analysis Services*

a

|  |
| --- |
| Bibliografía |

[1] «Entorno\_Virtual\_cast.pdf». Accedido: mar. 17, 2021. [En línea]. Disponible en: https://ra.biblioteca.uoc.edu/prestatgeries/22510/Entorno\_Virtual\_cast.pdf.

[2] «Citrix Workspace app 2103 for Windows - Citrix Spain», *Citrix.com*. https://www.citrix.com/es-es/downloads/workspace-app/windows/workspace-app-for-windows-latest.html (accedido mar. 17, 2021).

[3] «Crear conexión a bases de datos con Spoon - YouTube». https://www.youtube.com/watch?v=P-TvOqSb7RQ (accedido mar. 17, 2021).