

Caso práctico: Almacén de datos para el análisis del impacto ambiental y el consumo energético derivados de la actividad económica

PRA3–Explotación de datos

Presentación

A partir de la solución oficial de la segunda práctica (PRA2), el estudiante debe implementar los cubos multidimensionales necesarios para la explotación de la información y el posterior análisis de datos. De este modo se facilitará la toma de decisiones de los usuarios potenciales.

Así pues, esta actividad tiene el objetivo de implementar un modelo *multidimensional online analytical processing* (OLAP) para el análisis multidimensional de la información, con el fin de responder a las preguntas definidas en el análisis de requerimientos.

Adicionalmente, se facilitará junto a este enunciado el *script*, `Script_Data_Load_DW.sql`, que se ejecutará en vuestra base de datos MS SQL Server para la creación y carga de todas las tablas que componen el modelo multidimensional de la solución oficial. De esta forma, todos los estudiantes inician la actividad desde el mismo punto de partida.

Descripción

La tercera parte del caso práctico consiste en diseñar un modelo OLAP para el análisis multidimensional de la información disponible en el almacén de datos. Para el desarrollo y diseño del modelo OLAP se realizarán los siguientes apartados y subapartados:

1. Creación del modelo OLAP

- a) Creación en Visual Studio 2017 de un proyecto de tipo Analysis Services Multidimensional, realizando una importación de la base de datos del servidor
- b) Origen de datos
- c) Vista del origen de datos
- d) Creación e implementación de los cubos
- e) Configuración de dimensiones conformadas

- f) Jerarquías y dimensiones
- g) Procesado y resolución de errores

2. Explotación del modelo

El modelo OLAP diseñado en el apartado 1 permitirá dar respuesta a las siguientes cuestiones:

- a) Análisis evolutivo de la inversión entre 2014 y 2018 en protección de la biodiversidad y paisajes en Cantabria
- b) Análisis evolutivo de la producción en ktoe (tonelada equivalente –energía–) por categoría de energía (producto) de los Estados Unidos
- c) Análisis del *top five* de países con áreas protegidas mayores, tanto marinas como terrestres
- d) Identificación de los países con mayor impacto sobre el ODS «Ciudades y comunidades sostenibles»
- e) Análisis del promedio de inversión de los años 2014 al 2018 en cuanto al ámbito medioambiental de la gestión de aguas residuales en la comunidad autónoma de la Comunidad Valenciana
- f) Comparativa de los países de Alemania, Francia, España e Italia en cuanto a área protegida en el año 2019

Para cada uno de los apartados anteriores, se debe incluir:

- La descripción de todas las acciones realizadas.
- Las capturas que muestren la correcta definición de las vistas y cubos OLAP, así como las dimensiones, relaciones y jerarquías.
- Las capturas que muestren la correcta definición de las consultas, la visualización del resultado de las explotaciones y su correcta interpretación.

Además de entregar el documento con la solución de la PRA3, se considerará en la evaluación de la actividad la implementación de los cubos en la máquina virtual proporcionada en el curso.

Criterios de evaluación

La nota final se calculará a partir de la suma de las calificaciones de los dos apartados siguientes en la proporción correspondiente:

- 1) Diseño del modelo OLAP (60 %)
- 2) Explotación del modelo (40 %)

Guía de muestra

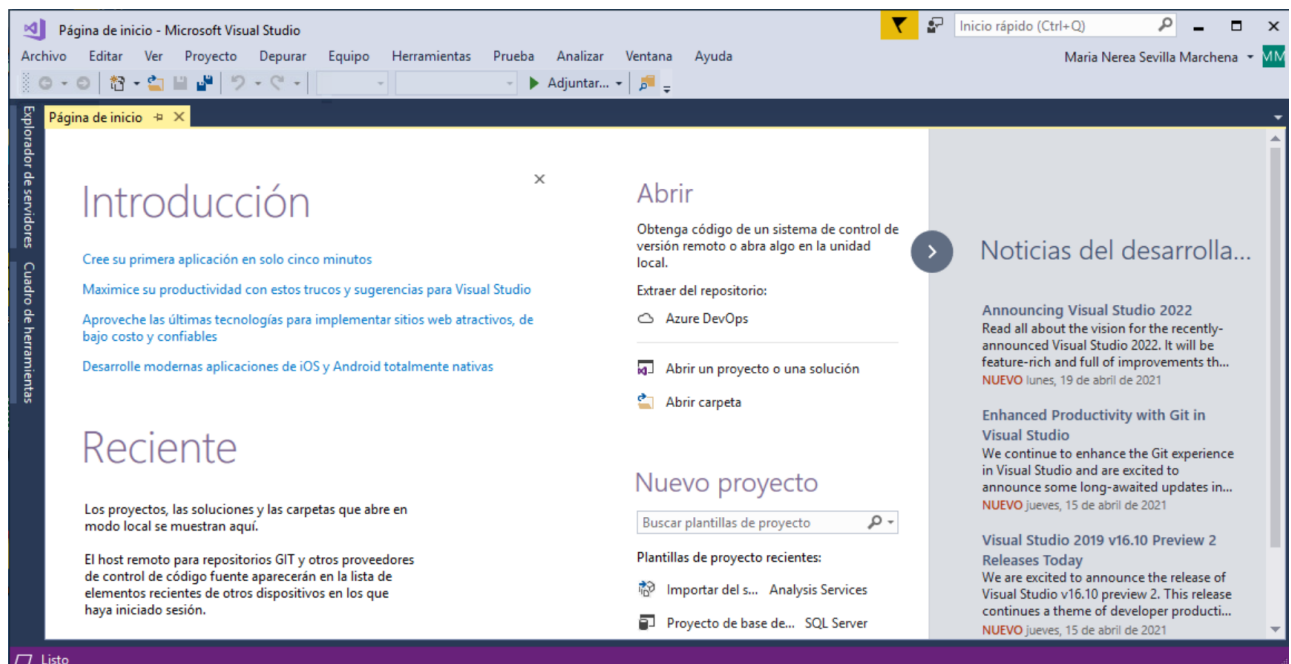
Esta guía de muestra se desarrolla con el fin de ayudar a alcanzar los objetivos planteados de la PRA3. La guía servirá para saber cómo realizar algunos de los apartados solicitados en la actividad, es decir, la creación y explotación del modelo OLAP para el análisis del impacto ambiental y el consumo energético derivados de la actividad económica.

1. Creación del modelo OLAP

El diseño del cubo se realizará creando un nuevo proyecto multidimensional en Visual Studio. Se deberán definir los orígenes de datos, las vistas de los orígenes de datos, los cubos, las relaciones de atributos, las dimensiones y las jerarquías necesarias para realizar la implementación de la solución y las explotaciones solicitadas en el enunciado de la PRA3.

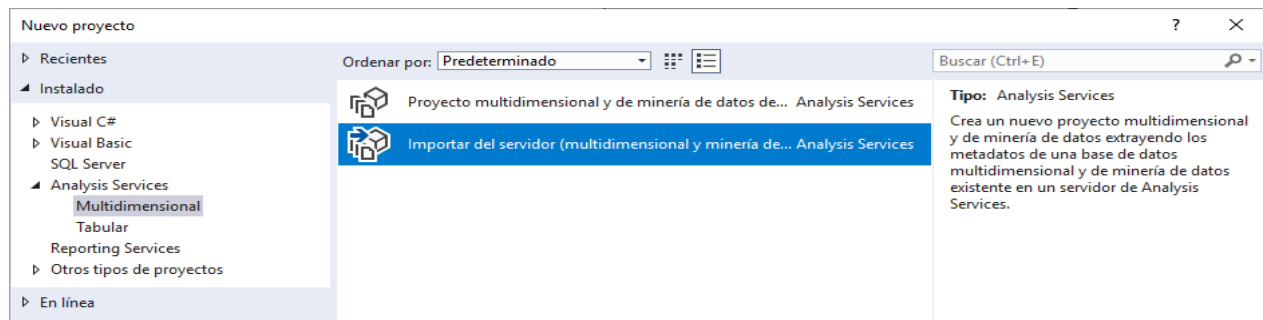
a) Creación del proyecto

Iniciaréis Visual Studio 2017 y seleccionaremos «Nuevo proyecto».

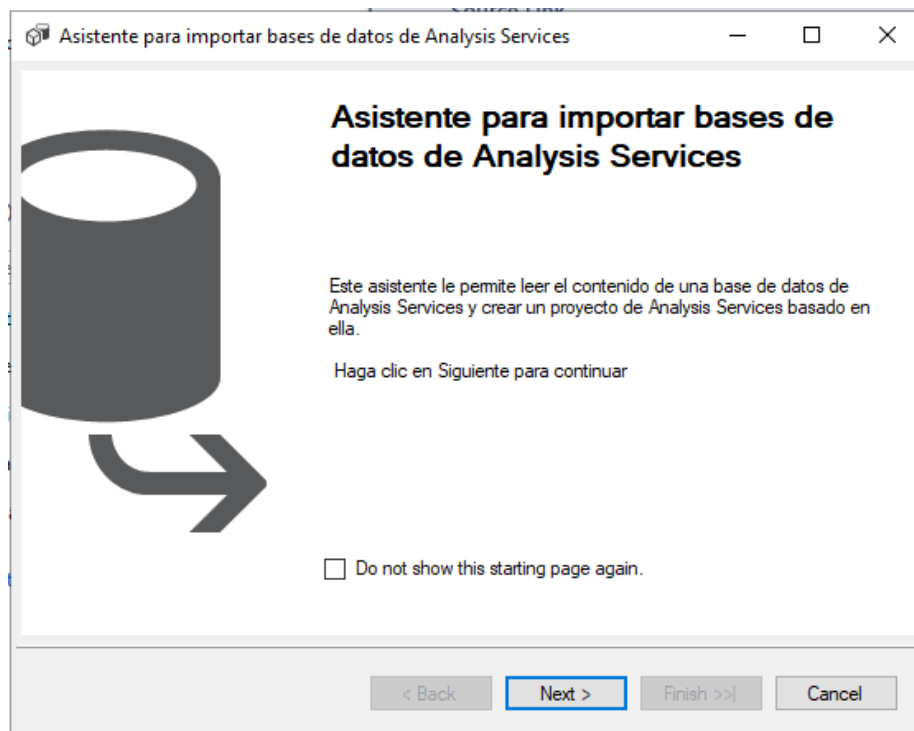


El tipo de proyecto que os interesa es Analysis Services Multidimensional, que permite crear cubos analíticos y realizar proyectos de minería de datos.

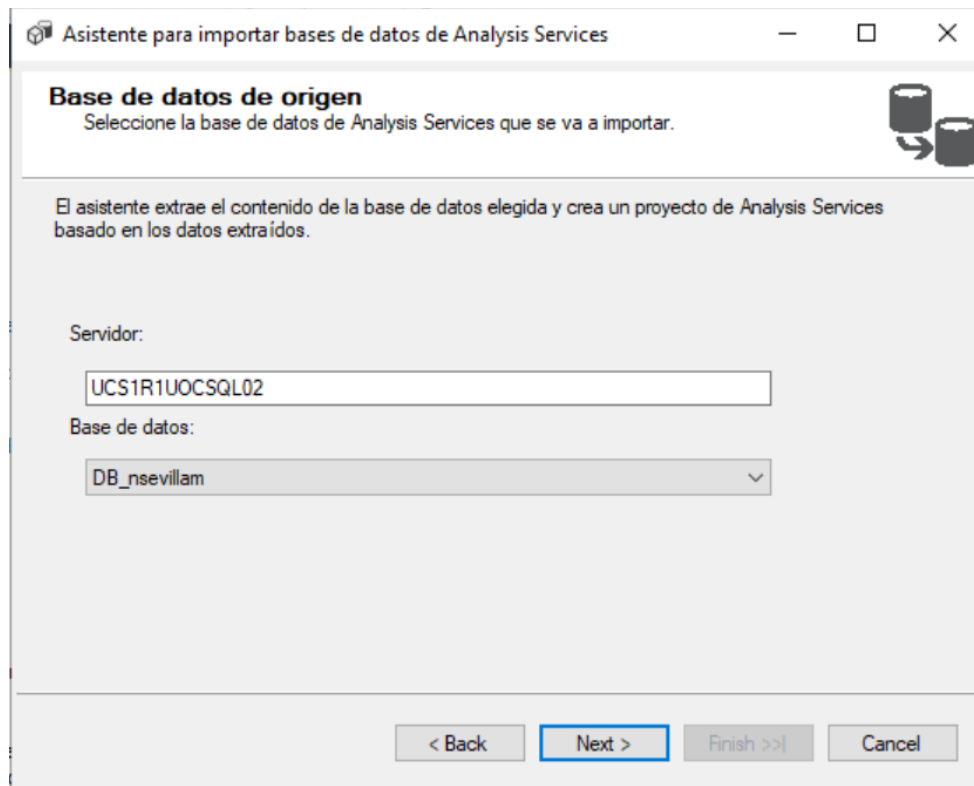
Seleccionaréis la opción «Importar del servidor (multidimensional y minería de datos)».



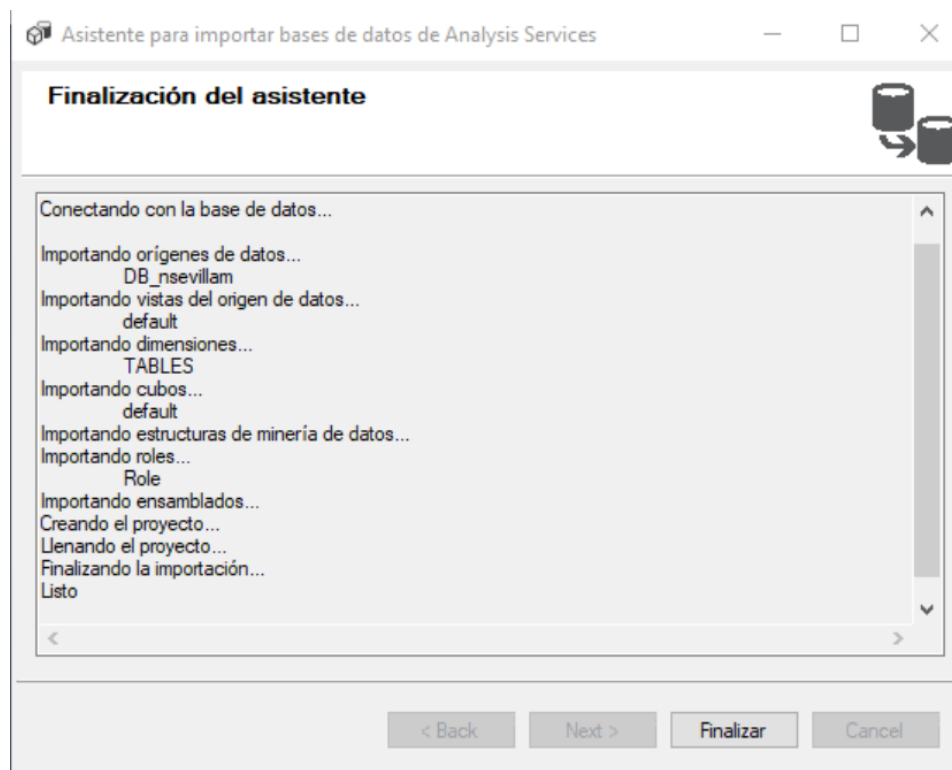
Le asignaréis un nombre y una ubicación donde guardar el proyecto y seleccionaréis «Aceptar» para iniciar el asistente e importar la base de datos de Analysis Services.



Introduciréis vuestros datos del servidor de Analysis Services, por ejemplo, UCS1R1UOCSQLXX (donde XX es el número de servidor facilitado en el aula) y seleccionaréis la base de datos («DB_nsevillam», en el ejemplo). Pulsaréis «Next» y se importará la base de datos.



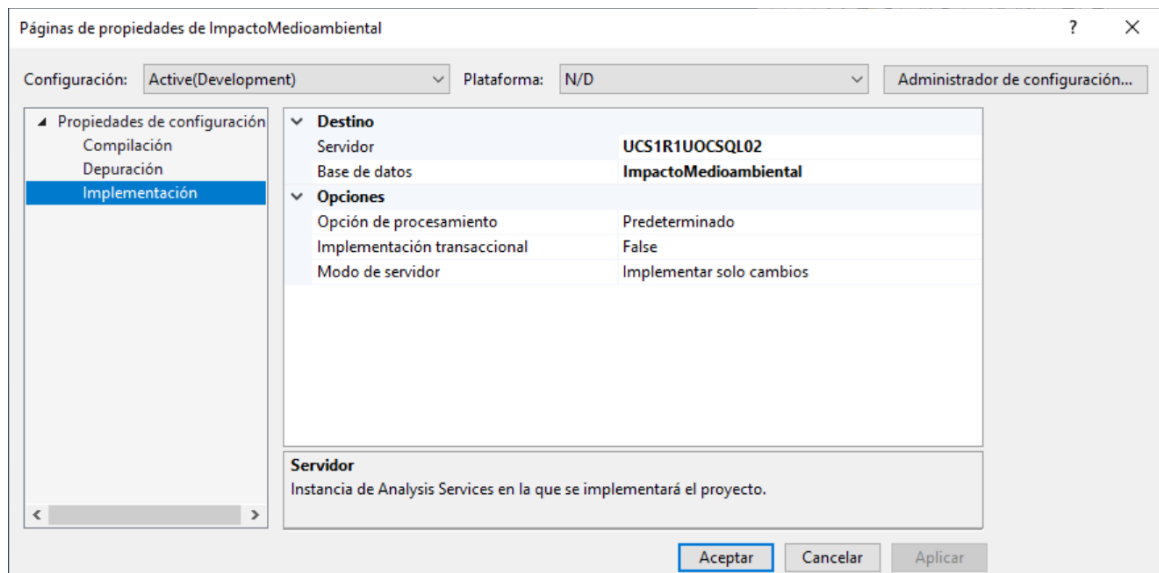
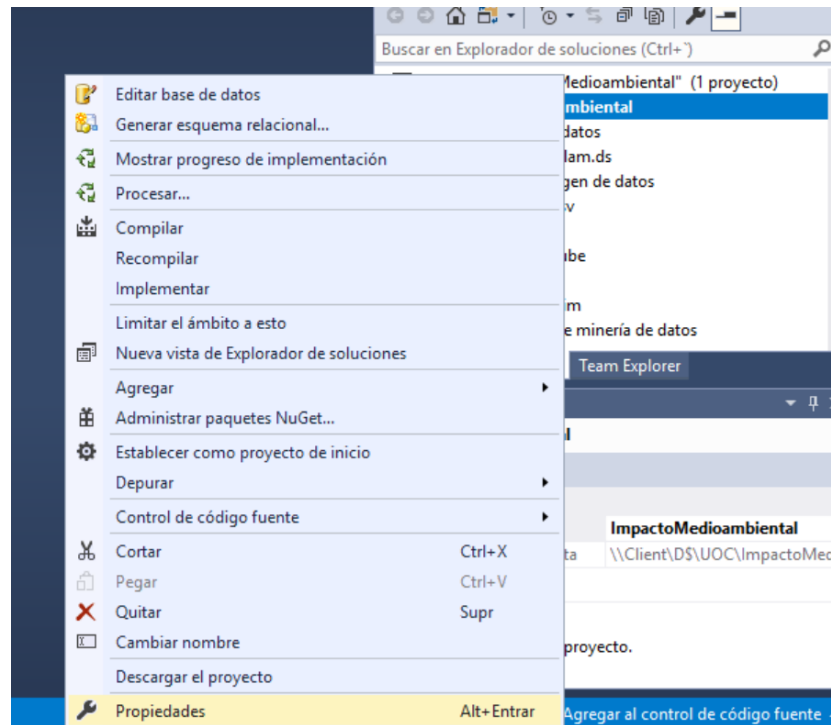
Seleccionaréis «Finalizar». Comprobaréis que se ha creado un origen de datos («DB_nsevillam») que se corresponde con vuestra base de datos en el servidor. Además, se creará una vista de datos por defecto, con un cubo *default*, una dimensión *tables* y un rol. **Estos objetos no deben ser borrados ni utilizados en la realización de la práctica.**



b) Origen de datos

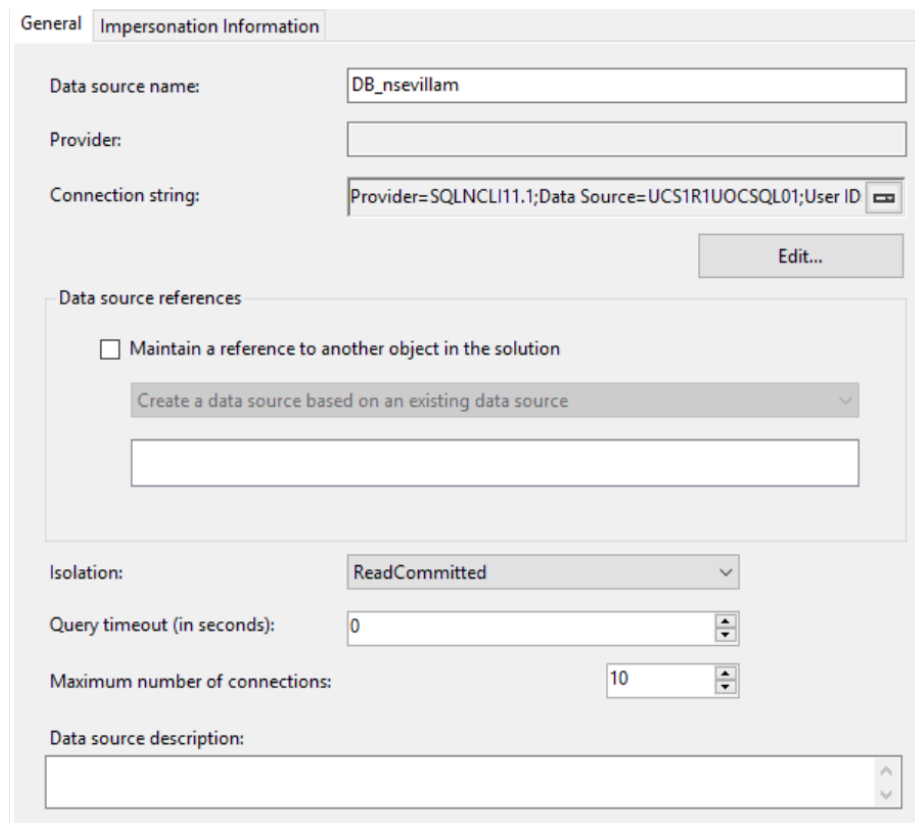
Antes de comenzar a desarrollar el proyecto de Analysis Services, una vez importada la base de datos, hay que comprobar el servidor de implementación de la solución y el origen de datos.

Para configurar el servidor de implementación, desde el «Explorador de soluciones», nos situamos en el proyecto que acabamos de crear y visualizamos las propiedades con el menú contextual. Dentro de las propiedades de configuración de «Implementación», deberéis cambiar el servidor en la propiedad «Destino», en lugar de «localhost». Especificad vuestro servidor, por ejemplo, UCS1R1UOCSQL02.



El siguiente paso que deberéis comprobar es el origen de datos, ya que será el que utilizaréis en la creación de las vistas y los cubos del proyecto, en los apartados c) y d) solicitados.

Para comprobar el origen de datos, lo haréis desde el «Explorador de soluciones» en el origen de datos creado tras la importación de la base de datos, y haciendo doble clic para ver la configuración. Desde el botón «**Edit**», se muestra una pantalla con la configuración de conexión al origen de datos. Deberéis realizar la configuración de conexión a vuestra base de datos y una prueba que lo verifique.



Como esto es una guía, en el documento de entrega deberéis, para este apartado, realizar las capturas de configuración, el test de conexión y el mecanismo de impersonalización de la información.

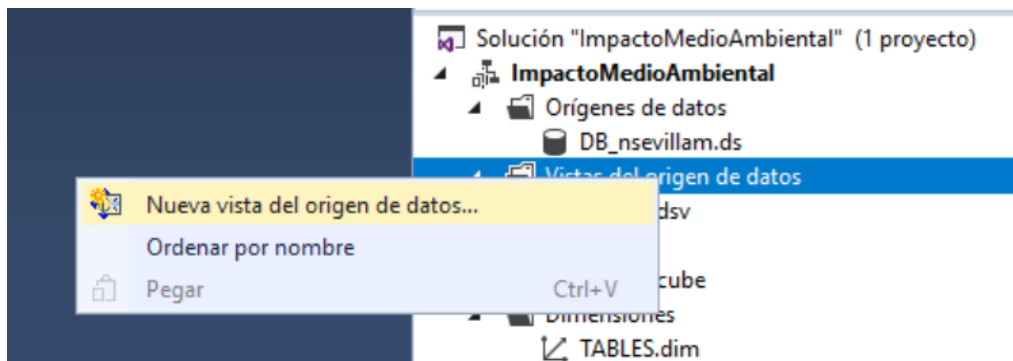
c) Vista del origen de datos

Una vez creada la fuente de origen, se deben crear las vistas del origen de datos, que definen el alcance del modelo OLAP.

En este caso, se definirán dos vistas de orígenes de datos:

- Vista_EnvironmentalMeasurements. Dará acceso a las tablas para diseñar el cubo para el análisis de las distintas mediciones medioambientales por fecha, por región, por sector de actividad económica, por tipo o equipo de instalación, por ODS (objetivos de desarrollo sostenible) o por unidad de medida.
- Vista_EnergyBalances. Dará acceso a las tablas para diseñar el cubo para el análisis del balance energético mundial por fecha, por país, por categoría de energía (producto) o por unidad de medida.

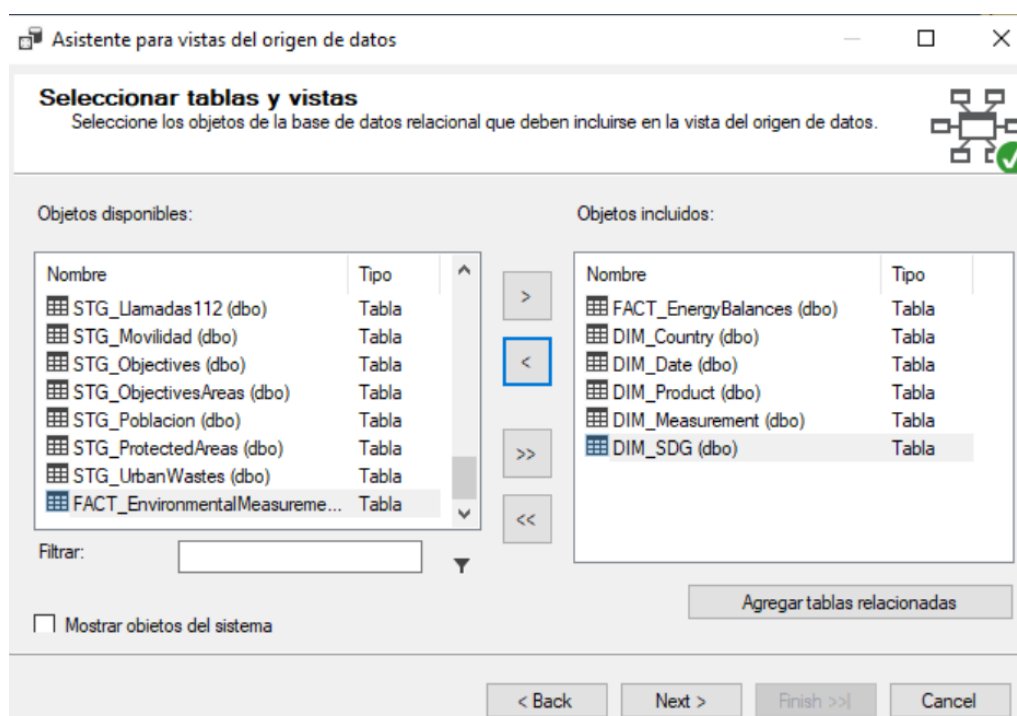
Para crear cada una de las vistas, elegiréis «Nueva vista del origen de datos» del menú contextual «Vistas del origen de datos» o en la opción con el mismo nombre del menú «Proyecto».



Vista_EnergyBalances

Se creará la vista sobre el balance energético. En la definición del origen de datos se seleccionará el creado anteriormente en el apartado b) y haréis clic en el botón «Next» del asistente de creación de vistas.

Las tablas que intervienen en esta vista son: «FACT_EnergyBalances», «DIM_Date», «DIM_Country», «DIM_Product», «DIM_Measurement» y «DIM_SGD».

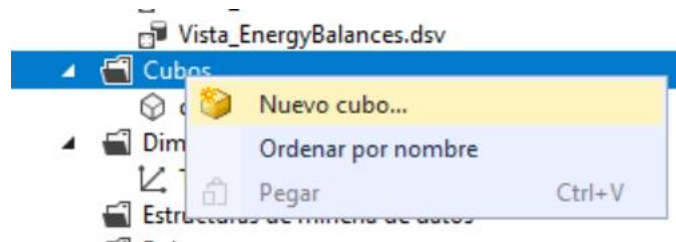


El nombre de esta vista se define antes de finalizar el asistente.

d) Creación e implementación de los cubos

Después del paso anterior de creación de las vistas para los orígenes de datos, el siguiente paso es la creación de los cubos. Este paso os permitirá realizar los análisis OLAP para la explotación de los datos del almacén de datos.

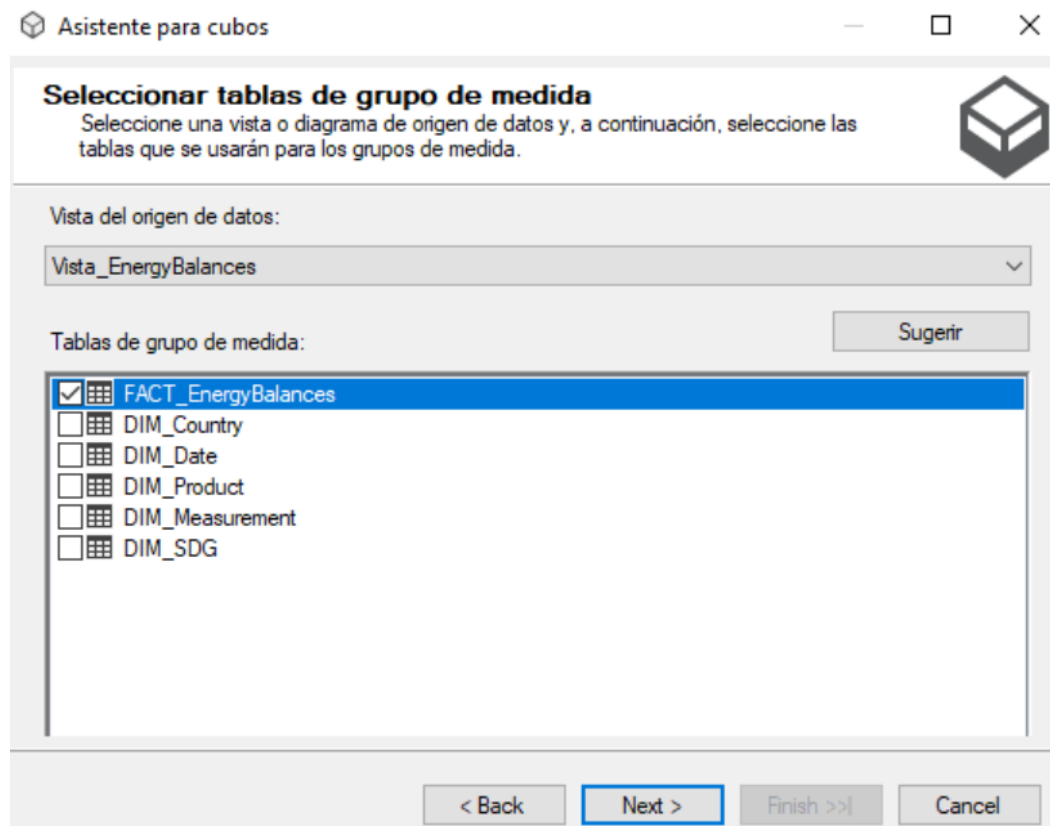
Crearéis cada cubo desde el menú contextual de cubos, seleccionando la opción «Nuevo cubo».



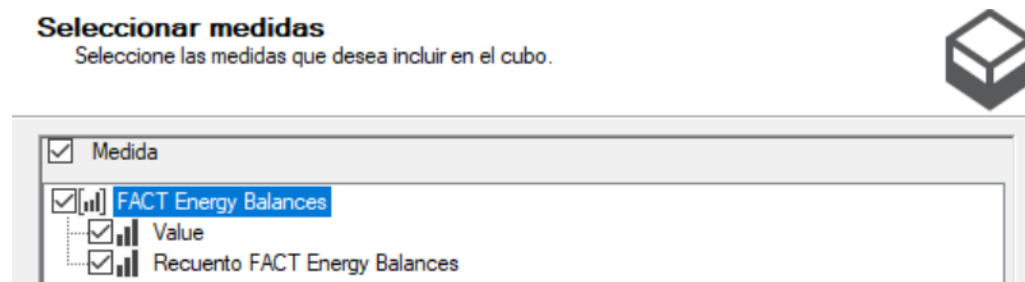
Cubo_EnergyBalances

Este cubo contendrá el modelo multidimensional para realizar el análisis del balance energético mundial.

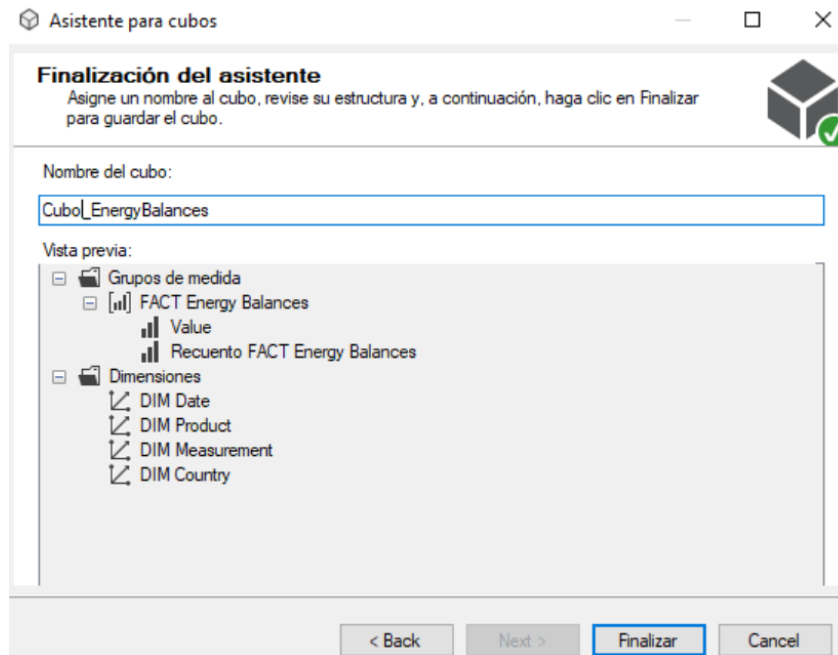
El método de creación específica que se usarán tablas ya existentes, así que seleccionareis la «Vista_EnergyBalances» y clicaréis en la tabla de hechos «FACT_EnergyBalances».



Seleccionareis las medidas que deseáis implementar en el cubo.



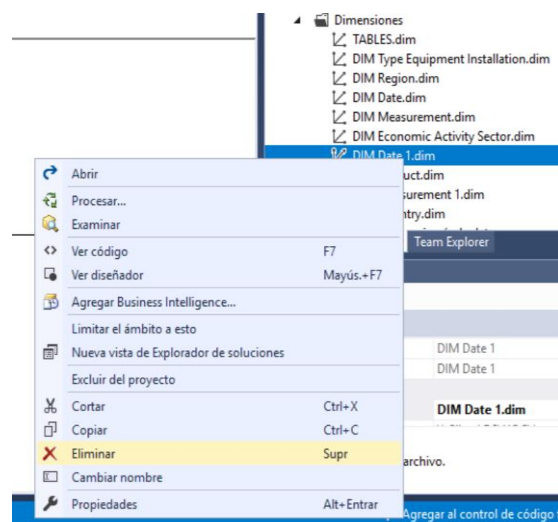
Después, escogeréis las dimensiones. Y, finalmente, le asignaréis un nombre adecuado al cubo y finalizaréis el asistente.



Desde la pestaña «**Estructura de cubo**» dentro de la pestaña de «Diseño», visualizaréis la estructura del cubo que habéis creado.

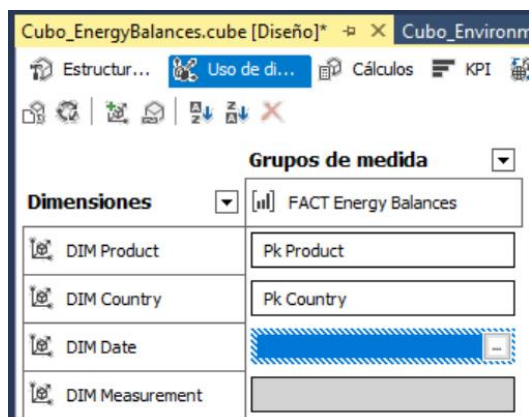
e) Configuración de dimensiones conformadas

Después de la creación de los cubos, puede que algunas dimensiones se hayan duplicado en el último cubo que hayáis creado. Para configurar correctamente las dimensiones conformadas, será necesario resolver esta duplicidad borrando dichas dimensiones desde el «Explorador de soluciones» del apartado «Dimensiones». Haréis clic en la dimensión y se seleccionará «Eliminar».



En el siguiente paso, deberéis añadir las dimensiones eliminadas al cubo correspondiente. Para ello, desde el diseño del cubo, en la pestaña «Estructura de cubo», escogeréis la opción «Agregar dimensión al cubo» desde el menú contextual de las dimensiones.

Finalmente, habrá que definir las relaciones entre la tabla de hechos y las nuevas dimensiones. Para ello, deberéis ir a la pestaña «Uso de dimensiones» del diseño del cubo.



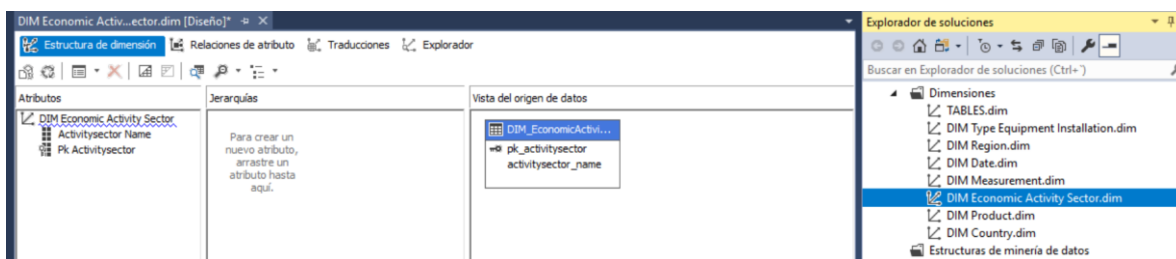
f) Jerarquías y dimensiones

Una vez definidas todas las relaciones del cubo, se trabajará con las dimensiones del proyecto con el objetivo de definir sus atributos y jerarquías. En algunos casos, al tratarse de dimensiones conformadas, es decir, aquellas dimensiones que están compartidas por varias tablas de hechos permitiendo su integración, los cambios realizados tendrán efectos en todos los cubos que las utilicen.

Para definir los atributos y jerarquías de cada dimensión, podréis hacer doble clic sobre la dimensión en el apartado «Dimensiones» del «Explorador de soluciones» o elegir la opción «Editar dimensión» desde el menú contextual en la pestaña «Estructura de cubo».

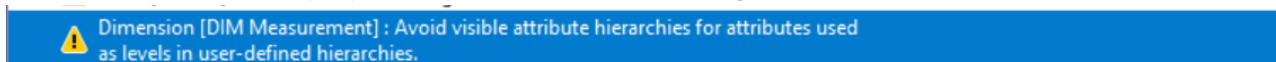
DIM_EconomicActivitySector

Para editar la dimensión, arrastraréis los atributos que queráis utilizar desde la zona «Vista del origen de datos» hasta «Atributos».

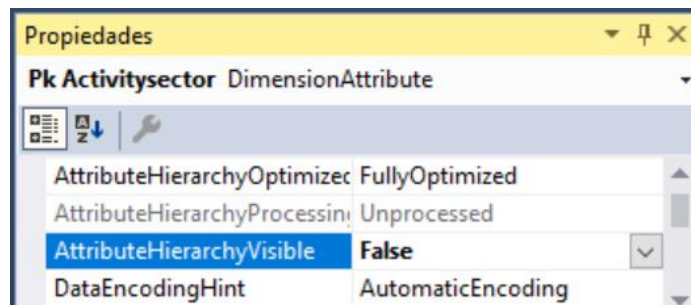


A la hora de analizar los datos, los analistas no deberían poder trabajar con valores de tipo clave primaria, dado que estas son de uso interno para la base de datos y, en ocasiones, pueden cambiar entre carga y carga. Además, al tratarse de claves subrogadas, no ofrecen un valor que sea fácil de entender por los analistas.

Además, cambiar esta propiedad permitirá resolver la siguiente advertencia:

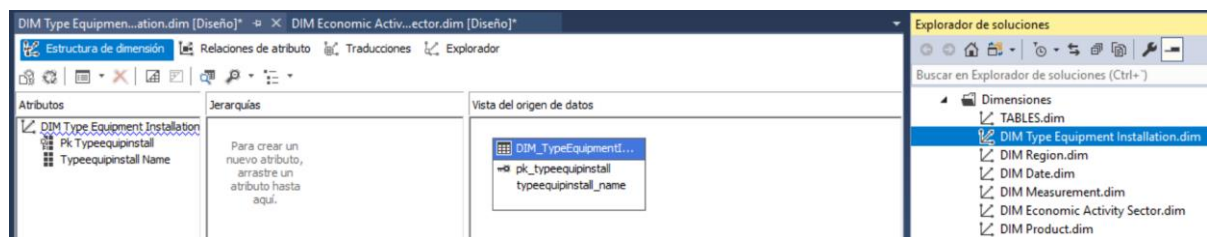


Por ello, se oculta seleccionando las propiedades del atributo y cambiando la propiedad a «AttributeHierarchyVisible».



DIM_TypeEquipmentInstallation

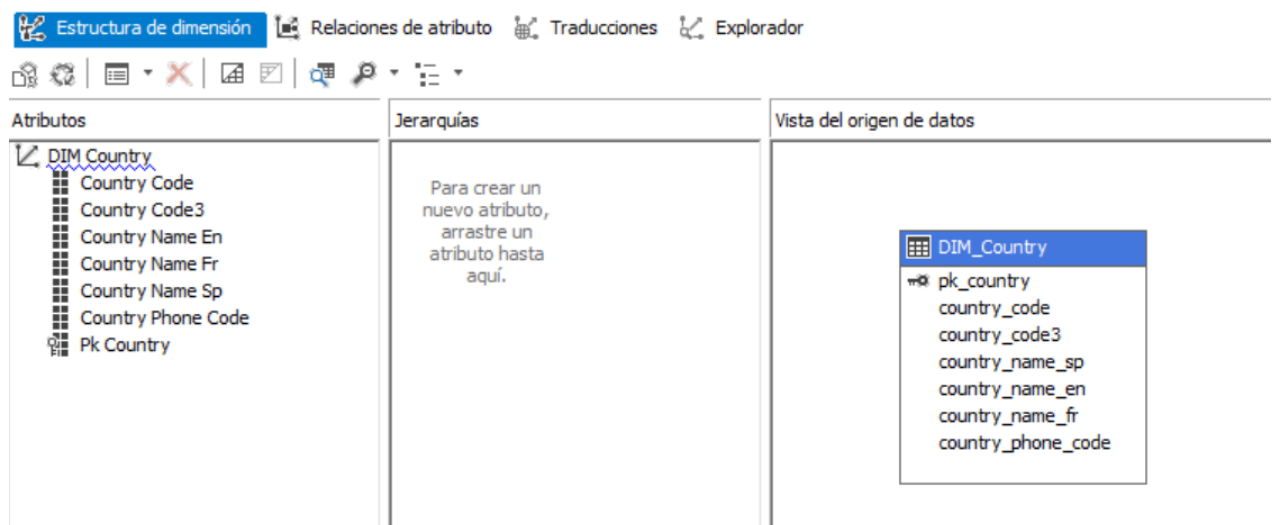
Al igual que para la dimensión anterior, procederéis a escoger los atributos necesarios y los renombraréis, si es preciso.



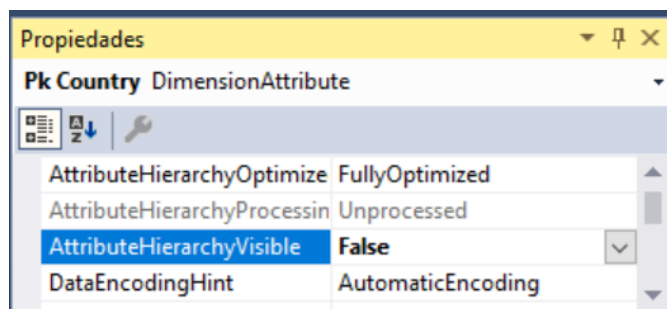
También ocultaréis la clave primaria de la dimensión, seleccionando las propiedades del atributo y cambiando la propiedad a «AttributeHierarchyVisible».

DIM_Country

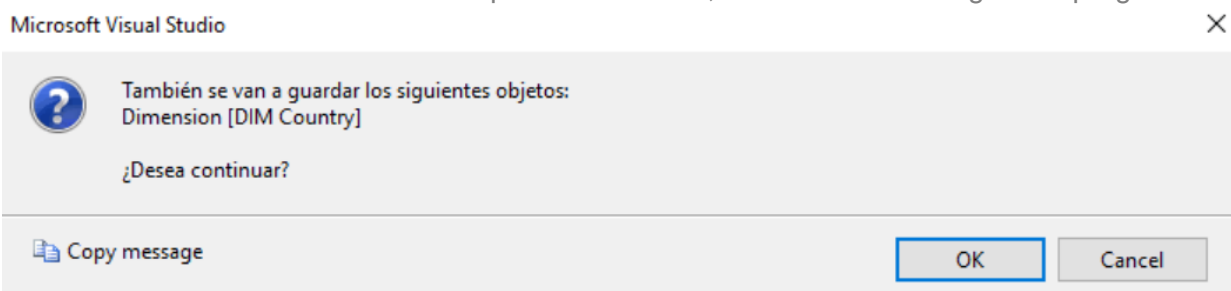
Para editar la dimensión, arrastraréis los atributos que queráis utilizar desde la zona «Vista del origen de datos» hasta «Atributos» y los renombraremos, si es necesario.



Se procede también a ocultar el atributo «Pk Country», para resolver la advertencia generada al procesar la dimensión.



Después de definir las dimensiones y jerarquías, se puede procesar directamente el cubo o solo la dimensión definida. Si decidís procesar el cubo, se os muestra la siguiente pregunta:

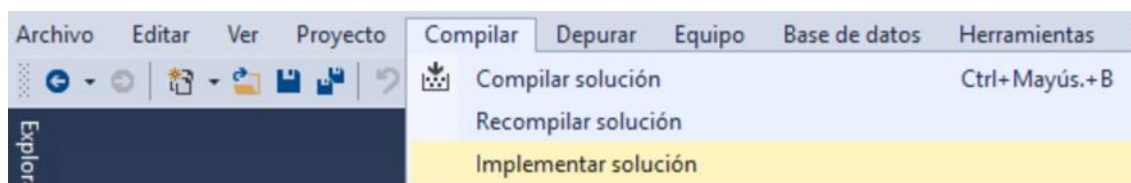


A la que responderéis afirmativamente, pulsando el botón «OK».

Para poder visualizar los datos de los cubos, deberéis implementarlo en el servidor de SSAS.

g) Procesado y resolución de errores

Para poder hacer explotaciones con los datos en los cubos que hemos definido, tenemos que implementar la solución, que consistirá en la compilación de todos los componentes. Para ello, iremos al menú principal y de la opción «Compilar», elegiremos «Implementar solución».



Después de la compilación, nos aparece una pantalla con el resultado del despliegue de la solución: correcto o fallido.

Aunque el despliegue sea correcto, el motor puede detectar advertencias que tendréis que revisar y tratar de resolver en este apartado de la entrega.

2. Explotación de datos

A continuación, se muestra un ejemplo de explotación de datos que se puede realizar tras la implementación de la solución. El estudiantado debe incluir todos los apartados solicitados en este enunciado. Para acceder al visor OLAP, hay que entrar en la pestaña *browser* de cada cubo creado.

1) Análisis evolutivo de la inversión entre 2014 y 2018 en protección de la biodiversidad y paisajes en Cantabria

Para analizar el nivel de inversión en protección de la biodiversidad y paisajes de Cantabria en los cinco últimos años, utilizaréis el cubo «Cubo_EnvironmentalMeasurement» aplicando dos filtros: uno sobre la dimensión «Measurement» y el atributo «Measurement Code» igual a «PROTECCIÓN DE LA BIODIVERSIDAD Y DEL PAISAJE», y el otro, sobre la dimensión «Region» igual a Cantabria.

Después, seleccionad los diferentes atributos que queráis visualizar en la consulta y procesad.

Dimensión	Jerarquía	Operador	Expresión de filtro
DIM Measurement	Measurement Code	Igual	{ PROTECCIÓN DE LA BIODIVERSIDAD ...
DIM Region	Region	Igual	{ Cantabria }
<Seleccionar dimensi...			

Date Year	Value
2008	60725
2009	6000
2010	78973
2011	5547
2012	199329
2013	85416
2014	7778
2015	280442
2016	209774
2017	177969
2018	196311

Recursos externos de apoyo

- **Proyecto SSAS (Microsoft SQL Server Analysis Services)**

Enlace: <https://www.sqlshack.com/es/como-construir-un-cubo-desde-cero-usando-sql-server-analysis-services-ssas/>

Objetivo: Guía rápida en castellano para la creación de un cubo desde cero usando SSAS

Comentarios: Guía bastante detallada con capturas de pantalla y comentarios, similar a lo que se solicita como parte de la entrega.

Enlace: <https://docs.microsoft.com/es-es/analysis-services/multidimensional-tutorial/multidimensional-modeling-adventure-works-tutorial?view=asallproducts-allversions>

Objetivo: Guía oficial de MS para la creación de un cubo desde cero usando SSAS

Comentarios: Guía oficial de MS con el detalle de todas las características disponibles a la hora de crear un proyecto SSAS. No hay soporte visual con imágenes, y pueden introducirse conceptos no recogidos en el material didáctico.

- **Power BI Desktop.**

Enlace: <https://docs.microsoft.com/en-us/learn/modules/get-data-power-bi/>

Objetivo: Guía oficial de MS Academy con los pasos básicos para trabajar con Power BI Desktop

Aportación: Conocimiento del repositorio MS Academy

Comentarios: Mini-curso oficial de MS para realizar los primeros pasos en Power BI. Está únicamente en inglés.

Formato y fecha de entrega

La entrega final de esta actividad debe realizarse enviando un único mensaje al buzón de registro de AC del apartado «Evaluación del aula». Adjunto al mensaje, se enviará un único archivo en formato Word o PDF con la solución de la PRA3. El nombre del archivo debe ser la composición del nombre de usuario y «_BDA_PRA3». Por ejemplo, si el nombre de usuario es «bantich», el nombre del archivo debe ser «bantich_BDA_PRA3.pdf».

La fecha máxima de entrega es el 28/12/2021 a las 23:59 horas.