

1

Estudiante(s): Marco Antonio Ayala Lituma

Fecha de entrega: 18/04/2021

Tema: Principales Características de los cubos ETL

Paralelo: A

Tarea Semana 2

RESUMEN

- Sobre las principales características de los cubos OLAP.
- El uso de PENTAHO como ETL
- Elaborar un instructivo para el uso de PENTAHO, que es necesario instalar in configurar.

SOBRE LAS PRINCIPALES CARACTERÍSTICAS DE LOS CUBOS OLAP.

Las herramientas OLAP proporcionan a las compañías un sistema confiable para procesar datos que luego serán utilizados para llevar a cabo análisis e informes que permitan mejorar las operaciones productivas, tomar decisiones inteligentes y optimizar la competitividad en el mercado.

Para funcionar, las aplicaciones OLAP utilizan un tipo de base de datos que posee la peculiaridad de ser multidimensional, denominada comúnmente Cubo OLAP.

Básicamente, el Cubo OLAP, es una base de datos que posee diversas dimensiones, ampliando las posibilidades que hasta el momento ofrecían las conocidas hojas de cálculo.

Mediante la incorporación de estos vectores o cubos, se han ampliado las posibilidades de las bases de datos relacionales, permitiendo el procesamiento de importantes volúmenes de información, que de lo contrario sería imposible realizar.

Cada una de las dimensiones que posee la base de datos incorpora un campo determinado para un tipo de dato específico, que luego podrá ser comparado con la información contenida en el resto de las dimensiones, para hacer posible la evaluación y posteriores informes de la información realmente relevante para una compañía.

Una base de datos multidimensional puede contener varios cubos o vectores que extenderán las posibilidades del sistema OLAP con el cual se trabaja.



2

Por ello, si bien en general los sistemas OLAP suelen estar compuestos por tres dimensiones, lo cierto es que existe la posibilidad de que el sistema OLAP albergue más de tres dimensiones mediante la utilización de estos Cubos OLAP.

Para tener una idea más simple de la función de los Cubos OLAP dentro de una base de datos multidimensional, cabe destacar que cada una de las dimensiones o escalas del cubo corresponde básicamente a una jerarquía de datos

- En este modelo los datos son vistos como cubos los cuales consisten en categoría descriptivas (dimensiones) y valores cuantitativos (medidas).
- Es un modelo multidimensional de datos, que simplifica a los usuarios realizar consultas complejas, arreglar datos en un reporte, cambiar de datos resumidos a datos detallados, etc.
- Los Cubos OLAP son vectores en los cuales se dispone la información, y gracias a esta ordenada jerarquía es posible llevar a cabo un análisis rápido de los datos.

Desarrollo

Para e proceso de ETL utilizaremos la base de datos de muestra AdventureworksDW que está lista para ser utilizada. Sin embargo, en la vida real, usted deberá crear algunos ETL para extraer automáticamente la información de su negocio al Data Warehouse.

Nosotros crearemos un cubo basado en la tabla de hechos FactResellerSales:

UnitPriceDiscountPct	DiscountAmount	ProductStandardCost	TotalProductCost	SalesAmount	TaxAmt	Freight
0	0	1898.0944	1898.0944	2024.994	161.9995	50.6249
0	0	1898.0944	5694.2832	6074.982	485.9986	151.8746
0	0	1898.0944	1898.0944	2024.994	161.9995	50.6249
0	0	1912.1544	1912.1544	2039.994	163.1995	50.9999
0	0	1912.1544	1912.1544	2039.994	163.1995	50.9999
0	0	1912.1544	3824.3088	4079.988	326.399	101.9997
0	0	1912.1544	1912.1544	2039.994	163.1995	50.9999
0	0	31.7244	95.1732	86.5212	6.9217	2.163
0	0	31.7244	31.7244	28.8404	2.3072	0.721
0	0	3.3963	20.3778	34.2	2.736	0.855
0	0	5.7052	11.4104	10.373	0.8298	0.2593
0	0	12.0278	48.1112	80.746	6.4597	2.0187
0	0	413.1463	413.1463	419.4589	33.5567	10.4865
0	0	884.7083	884.7083	874.794	69.9835	21.8699
0	0	699.0928	699.0928	809.76	64.7808	20.244

FactResellerSales muestra la información sobre las ventas de los vendedores, tales como montos de descuento, montos de impuestos, flete, monto de ventas y otra información relacionada.

Crearemos informes sobre la información de la tabla de hechos combinada con 2 dimensiones:

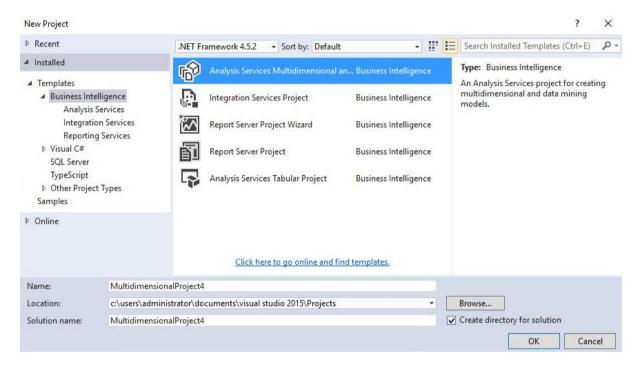


3

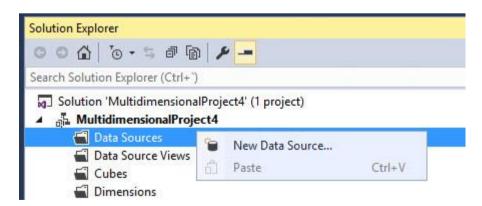
- Moneda
- Territorio de ventas

La dimensión **moneda**, nos ayudará a detectar cantidades en distintas unidades monetarias de todo el mundo y **de zona o territorio de ventas** le ayudará a detectar las ventas, descuentos, etc. por cada región.

Para comenzar a comenzar en SQL Server Data Tools, también conocido como SSDT, abra un nuevo proyecto y seleccione los modelos Multidimensionales y de Datos de Minería de Analysis Services:



En el Explorador de Soluciones, haga clic con el botón derecho en Orígenes de datos y seleccione Nuevo origen de datos:



Recibirá un asistente de bienvenida, presione siguiente:



4



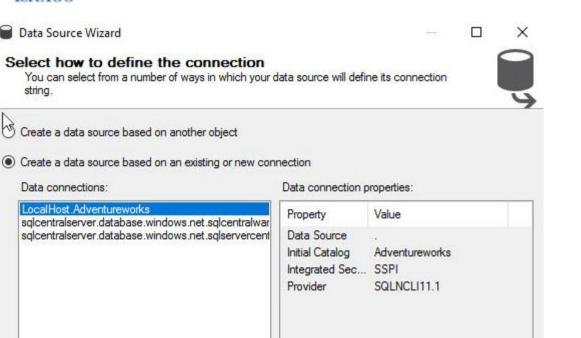
Seleccione la opción para crear una nueva Fuente de Datos basada en una conexión existente u otra nueva y presione el botón nuevo:

Next >

Cancel



5



En el **proveedor**, seleccione un OLEDB nativo\SQL Server Native Client 11.0. Para el **Nombre del Servidor**, seleccione el servidor SQL con la base de datos AdventureworksDW:

Next >

< Back

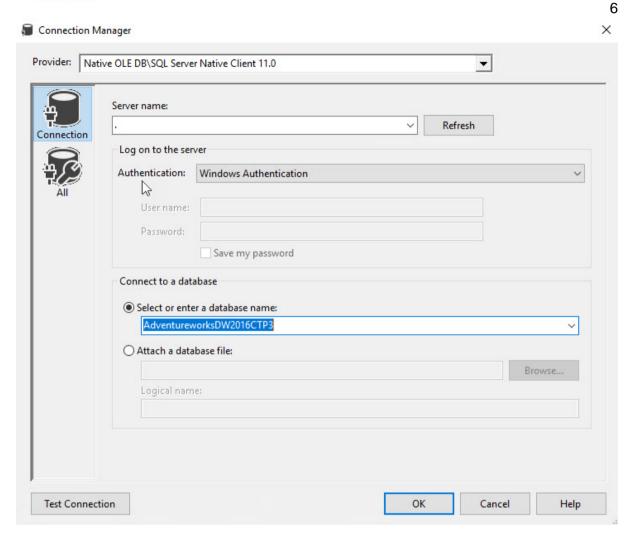
New...

Finish >>

Delete

Cancel

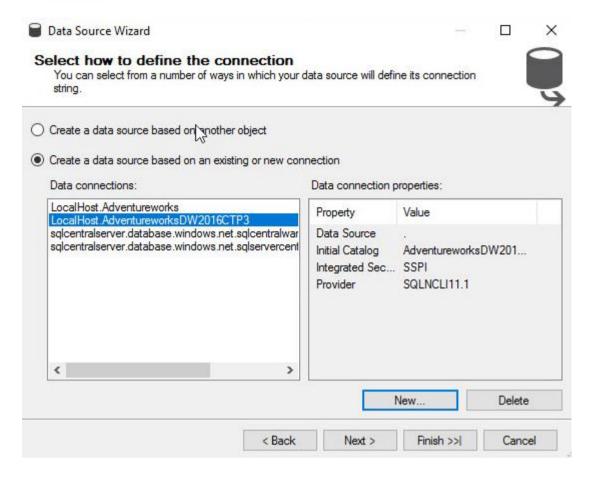




Seleccione la conexión de datos que acaba de crear y presione siguiente:



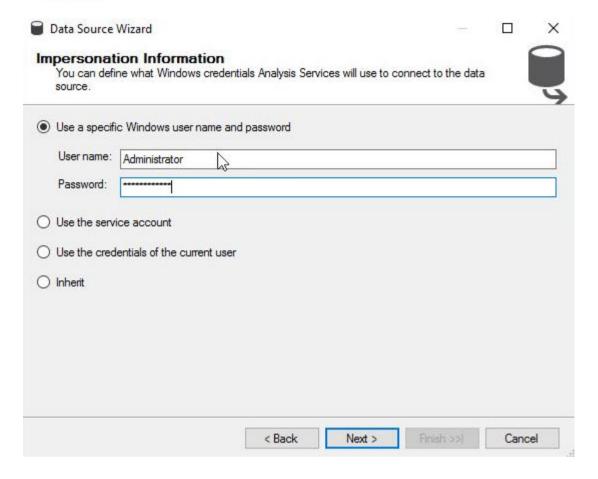




En la **Información de Suplantación** especifique un usuario con acceso a la base de datos AdventureworksDW:



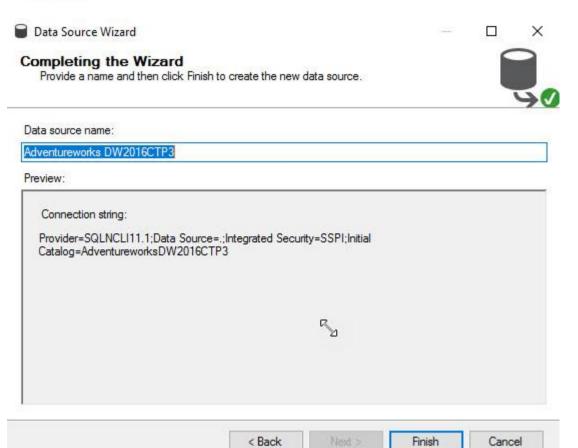
8



En la pestaña de Completando del asistente, presione finalizar:

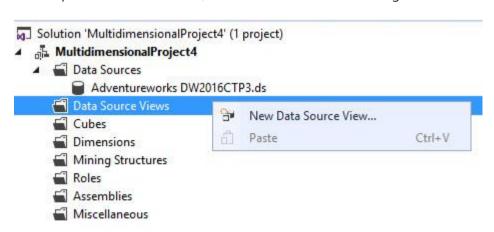


9



Nosotros creamos las credenciales para conectarnos a la base de datos AdventureworksDW. Crearemos ahora las vistas. En este proceso, seleccionaremos las tablas que serán incluirán en el cubo.

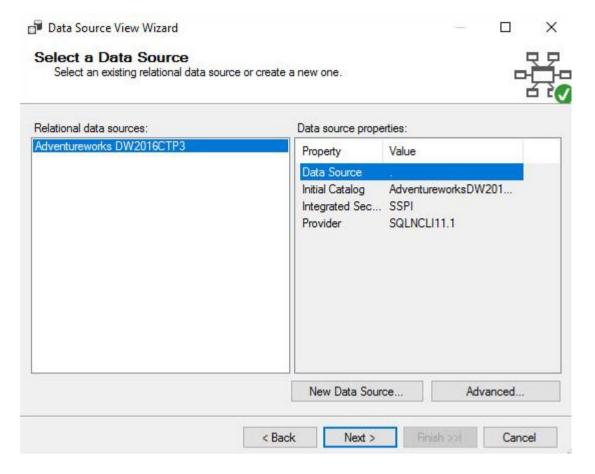
En el explorador de soluciones, seleccione Nueva Vista de Origen de Datos:



Seleccione la fuente de datos que acaba de crear:



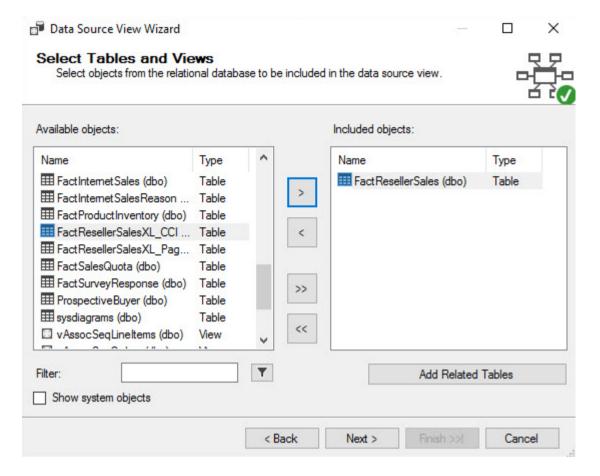
10



Seleccione la tabla FactResellerSales:



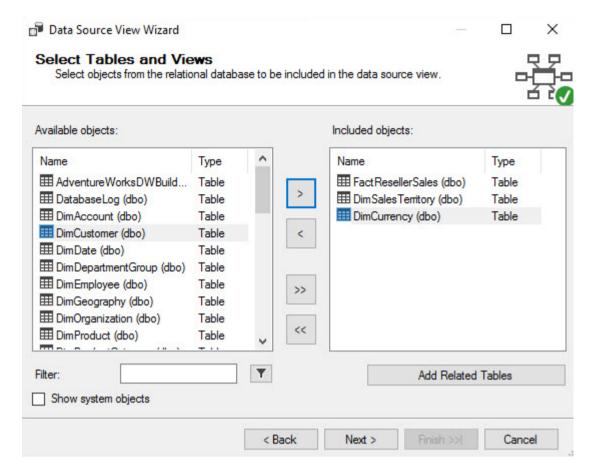
11



Incluya las dimensiones DimSalesTerritory y DimCurrency. Estas dimensiones incluyen información sobre las regiones de venta y sus monedas:



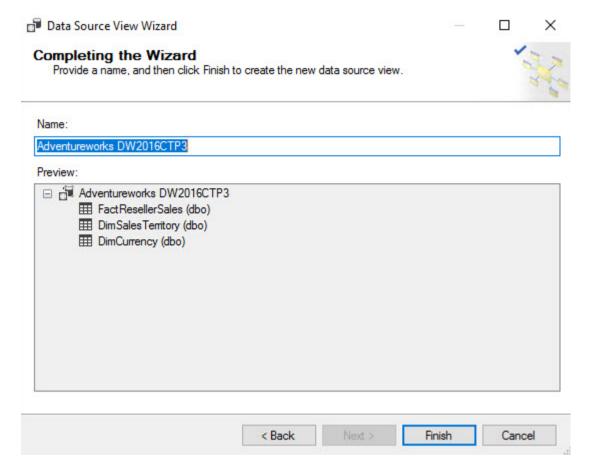
12



Una vez que el asistente haya completado, presione finalizar:

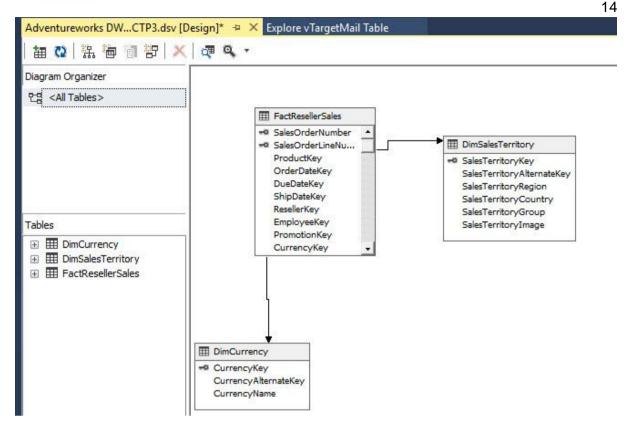


13



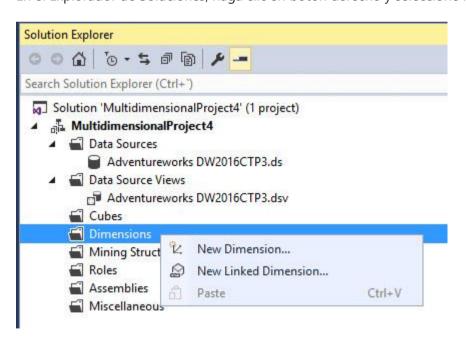
Como usted puede ver, la tabla de hechos y las dimensiones tienen una relación. Si las columnas comunes son iguales, las relaciones serán detectadas automáticamente:





Para crear un cubo, agregaremos dimensiones al cubo.

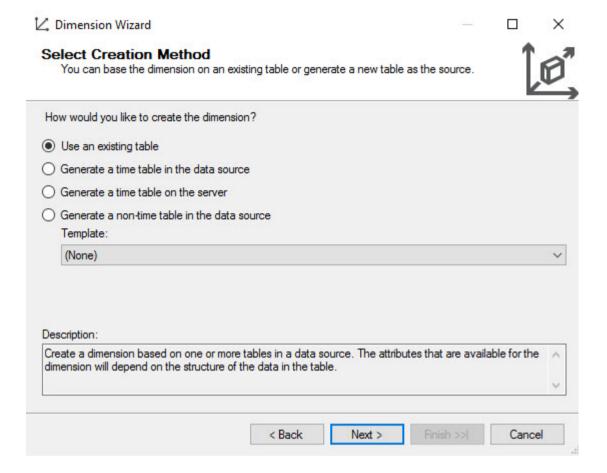
En el Explorador de Soluciones, haga clic en botón derecho y seleccione Nueva dimensión:



Usted Puede crear una dimensión basada en plantillas, tablas de tiempos o en tablas existentes. Para tal fin Seleccione la opción **Usar una tabla existente**:



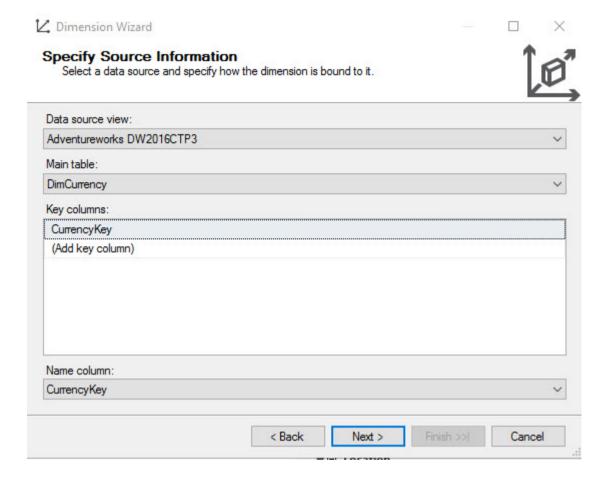
15



Seleccione DimCurrency como la tabla principal. Por defecto, la columna clave es clave de moneda:



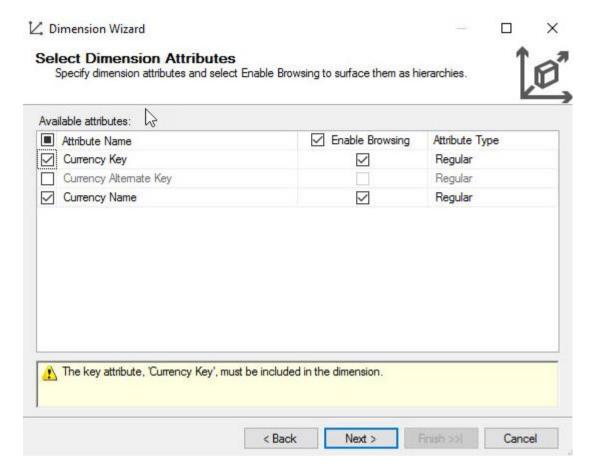
16



En los atributos disponibles, seleccione el nombre de la moneda:



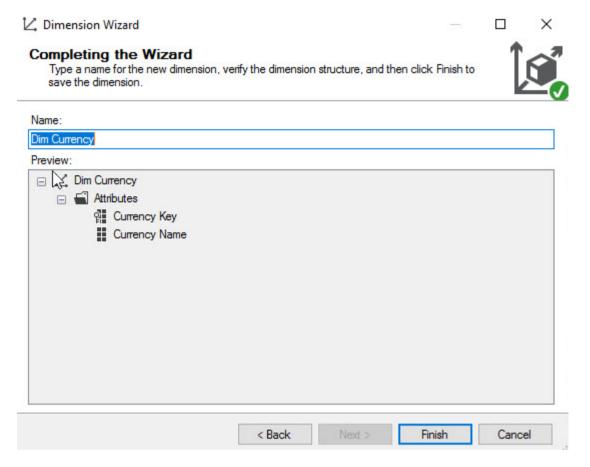
17



En el paso de **Completar del asistente** presione finalizar:

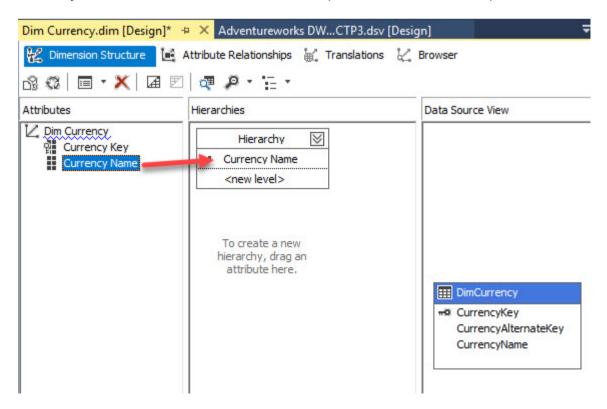


18



Ahora, tendrá una dimensión creada como esta.

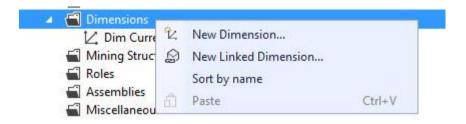
Arrastre y suelte el **Nombre de moneda** desde el panel de atributos a Jerarquía:



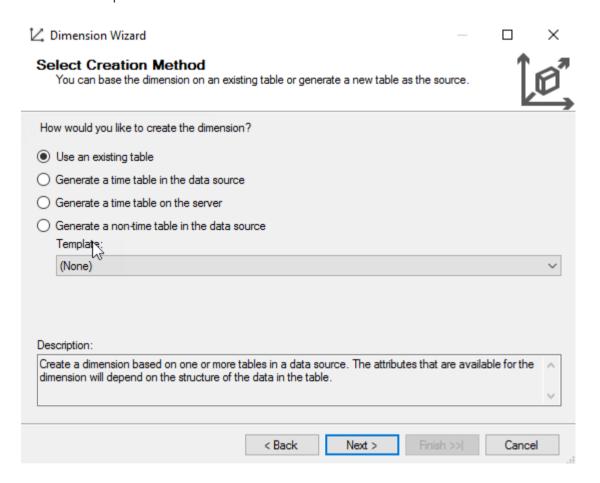


19

Una vez que haya creado la primera dimensión, **Nombre de moneda**, crearemos la dimensión del territorio de ventas. En el explorador de soluciones, haga clic con el botón derecho en Dimensiones y seleccione Nueva dimensión:



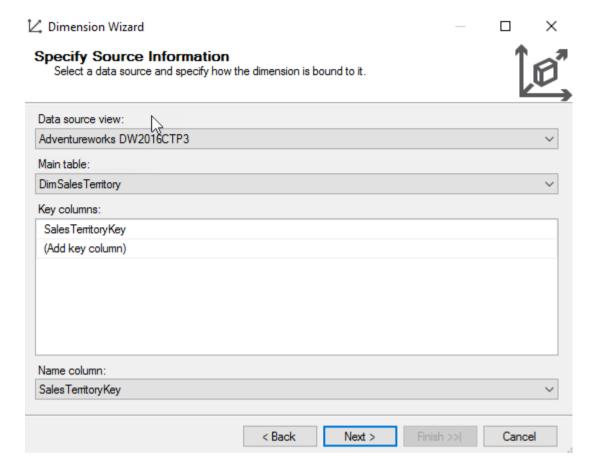
Seleccione la opción usar una tabla existente:



Seleccione la tabla dimsalesterritory:



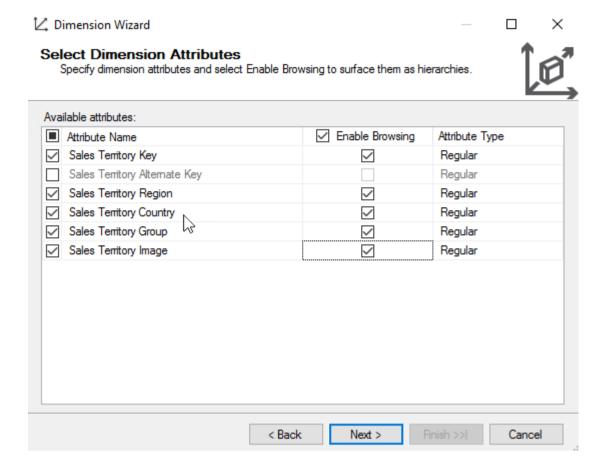
20



Seleccione los atributos que usted desea que se puedan navegar y explorar en los informes:



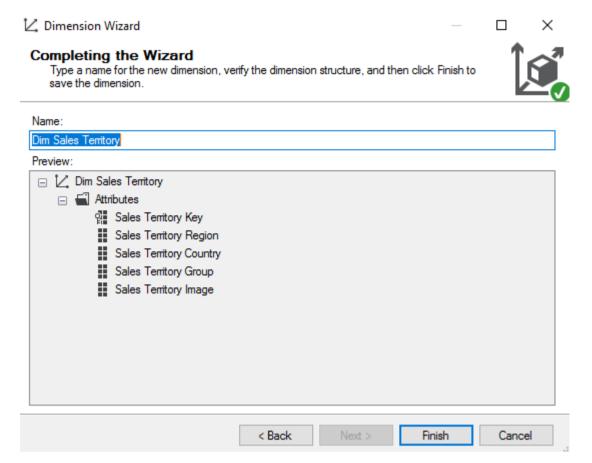
21



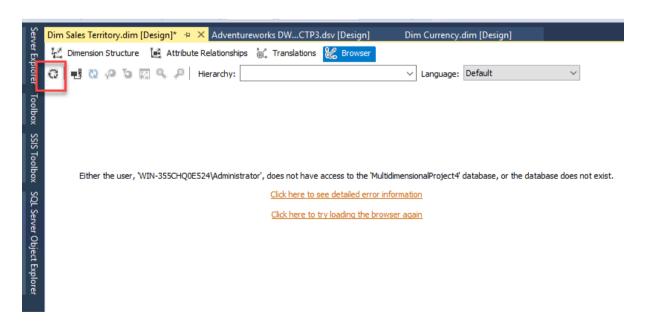
Una vez que se complete el asistente, presione finalizar:



22

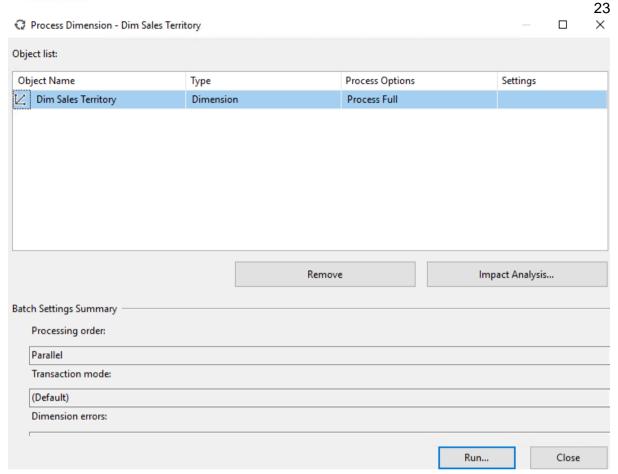


En la nueva dimensión que se ha creado, haga clic en el icono de proceso para procesar la dimensión. El proceso generará la estructura y cargará datos a la dimensión desde la tabla de SQL Server:



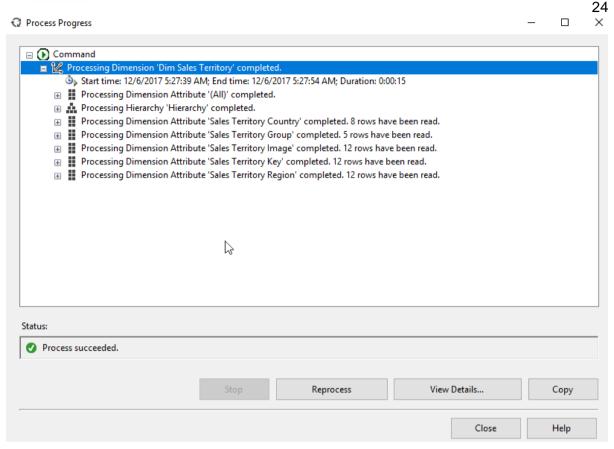
Presione ejecutar para procesar la dimensión:





En el Progreso del Proceso, cierre el proceso una vez que el proceso sea exitoso:

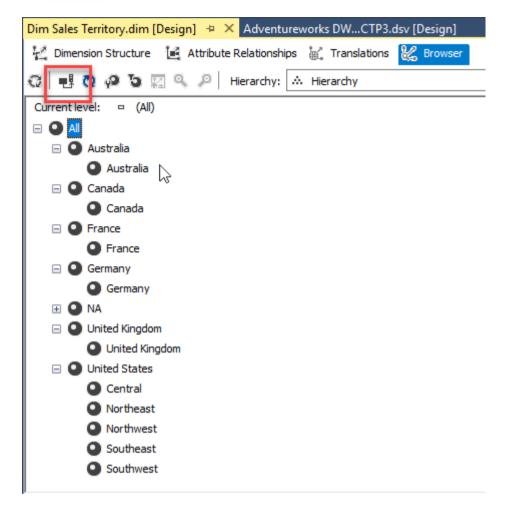




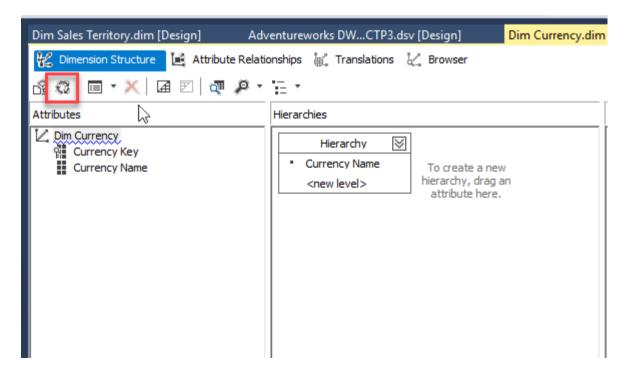
Puede actualizar la dimensión para ver la dimensión recién procesada:



25



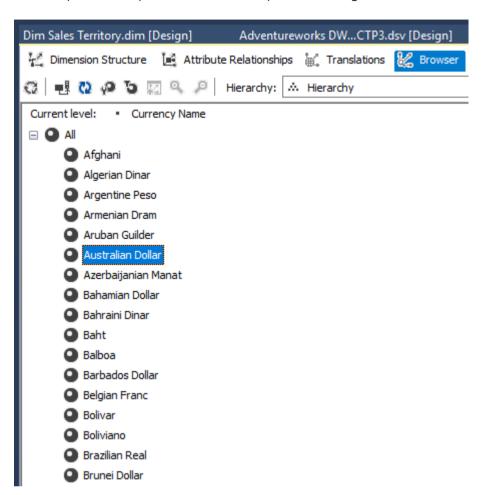
Procese la dimensión de moneda Dim también:



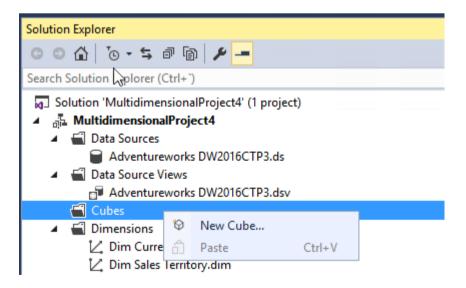


26

Una vez procesado, puede verificar en la pestaña Navegador:



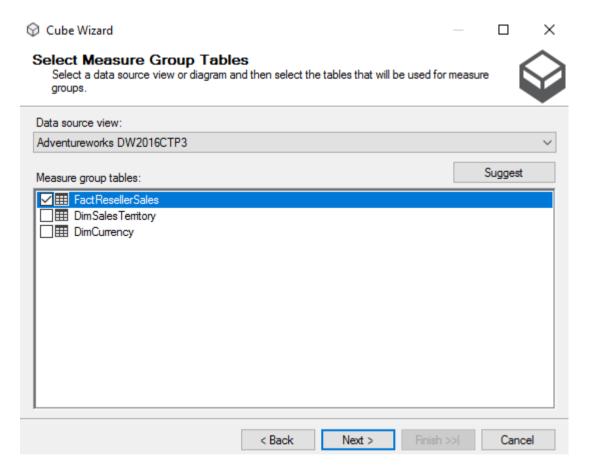
Ahora nosotros crearemos el cubo. En el Explorador de soluciones, haga clic con el botón derecho en **Cubos** y seleccione **Nuevo cubo**:





27

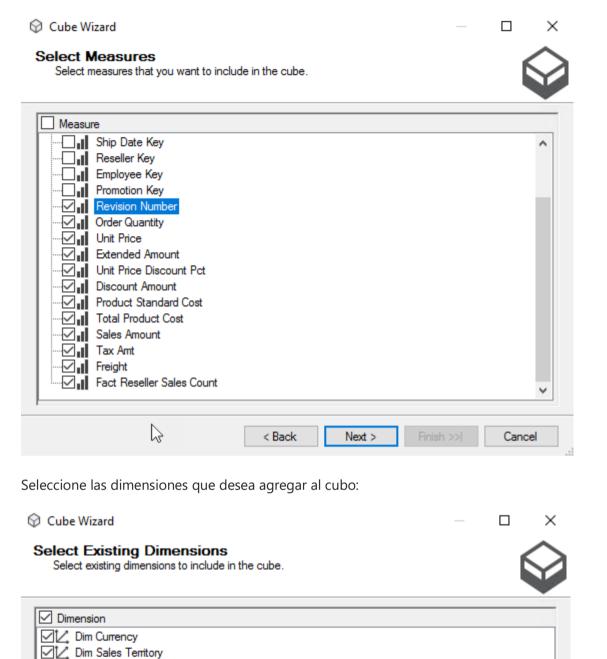
En **Seleccionar Tablas de Medida de Grupos** seleccione la tabla FactResellerSales . Las tablas de medida grupo se utilizan para incluir la tabla con datos para medir. Una medida puede ser el número de ventas, la cantidad vendida, el flete, etc.



Seleccione los datos a medir. Desmarcaremos las teclas y verificaremos los otros atributos para medir:



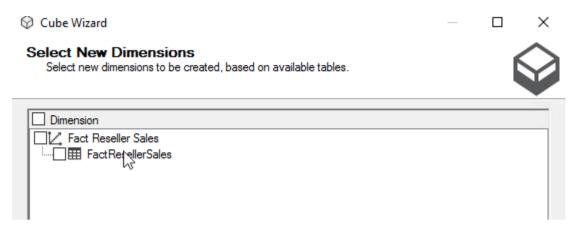
28



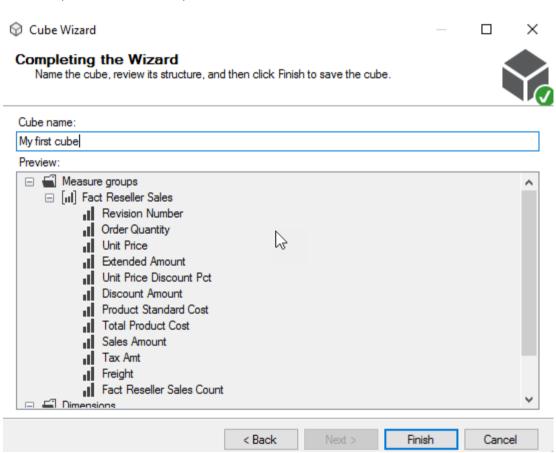
También puede agregar la tabla de hechos como una dimensión (dimensión degenerada). En este ejemplo, no lo vamos a añadir ya que:



29

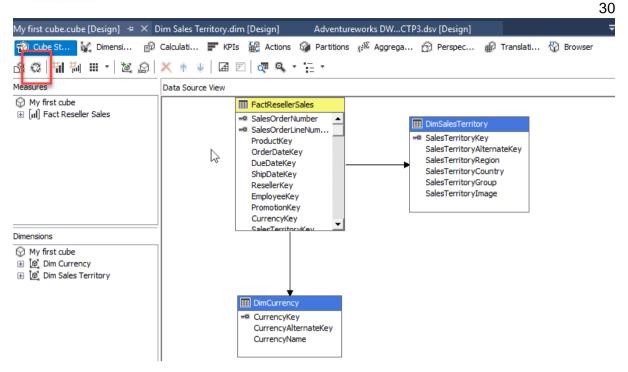


Una vez que se crea el cubo, presione finalizar:

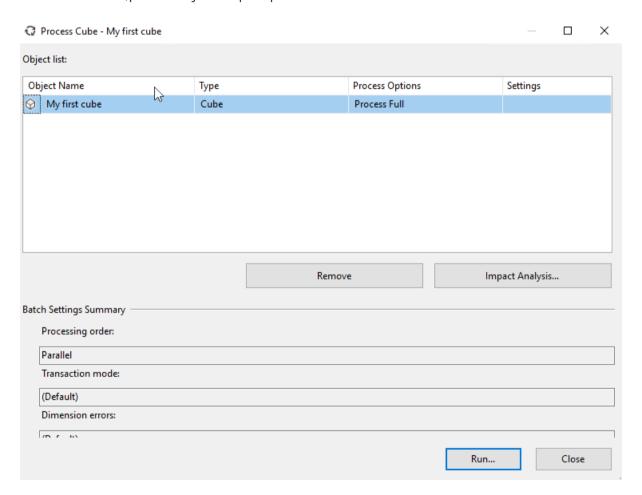


Usted Tendrá el cubo creado con la tabla de hechos (en amarillo) y las dimensiones (en azul). Pulse el icono de proceso para procesar el cubo:



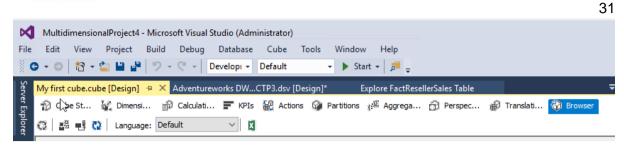


En **Process Cube**, presione ejecutar para procesar el cubo:

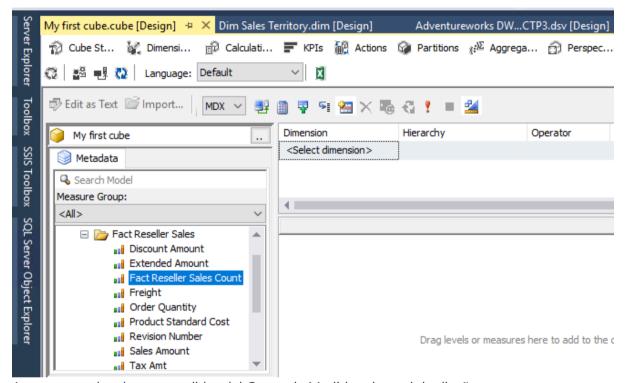


Una vez procesado, vaya al navegador:



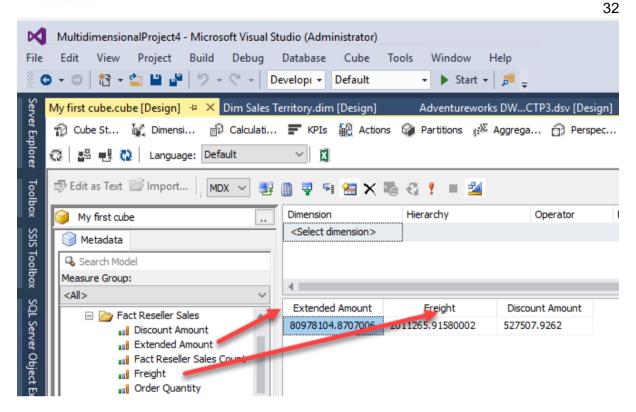


Presione el icono de actualización:

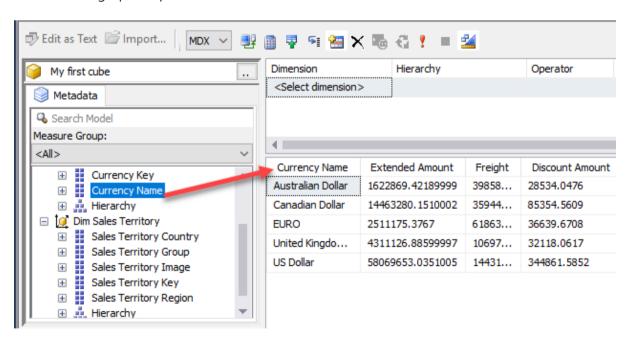


Arrastre y suelte algunas medidas del Grupo de Medidas al panel de diseño:



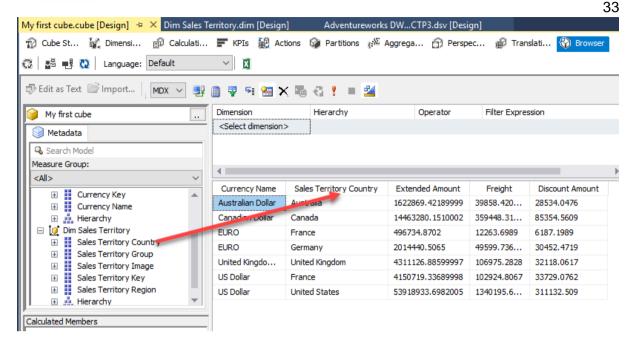


Ahora arrastre y suelte el nombre de la moneda en el panel de gráficos, entonces usted podrá ver las medidas agrupadas por moneda:



Arrastre y suelte el País del territorio de ventas. Ahora usted podrá ver las medidas agrupadas por moneda y territorio:





EL USO DE PENTAHO COMO ETL

Pentaho BI es una herramienta imprescindible para la toma de decisiones empresariales gracias a su poderoso motor de análisis. Con la suite de Pentaho BI podremos generar multitud de informes en cualquier formato: dinámicos, adhoc, guiados, predefinidos etc...

Estos informes pueden ser de gran ayuda, por ejemplo, a la hora presentar reportes mensuales de ventas, lo que sería un informe estático al uso. Un escalón más allá tendríamos los informes dinámicos que nos permitirán interactuar con dicha información y analizar dicha información en detalle.

Básicamente, si te planteas porque tu organización debería utilizar un software como Pentaho BI, aquí te damos unas buenas razones:

Mayor accesibilidad a la información corporativa: Gracias a Pentaho Data Integration (PDI ó Kettle) tendremos contenidos más accesibles, comprensibles y navegables. Lo que mejorará la productividad de la empresa.

Aporta una seguridad extra a la información de la empresa: Con Pentaho BI podremos controlar el acceso a la información mediante la asignación de roles y además, podremos saber quién hace uso de los datos y de qué datos exactamente.

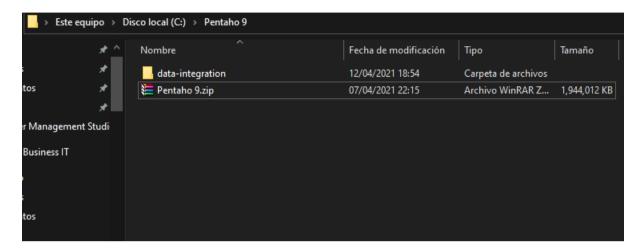
Es la base para la toma de decisiones: La ordenación de toda la información y los datos recogidos, además de la correcta presentación de los mismos para su posterior análisis, nos permitirá, como ya hemos indicado, tomar las decisiones de manera más acertada.



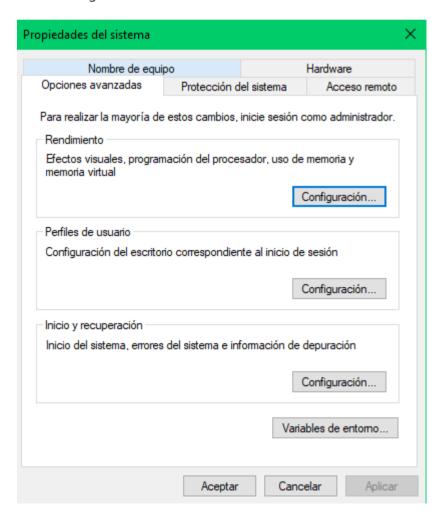
34

ELABORAR UN INSTRUCTIVO PARA EL USO DE PENTAHO

Como pasos debemos tener los siguientes archivos y copiarlos en algún lugar del disco duro en mi caso lo pondré en el disco C:

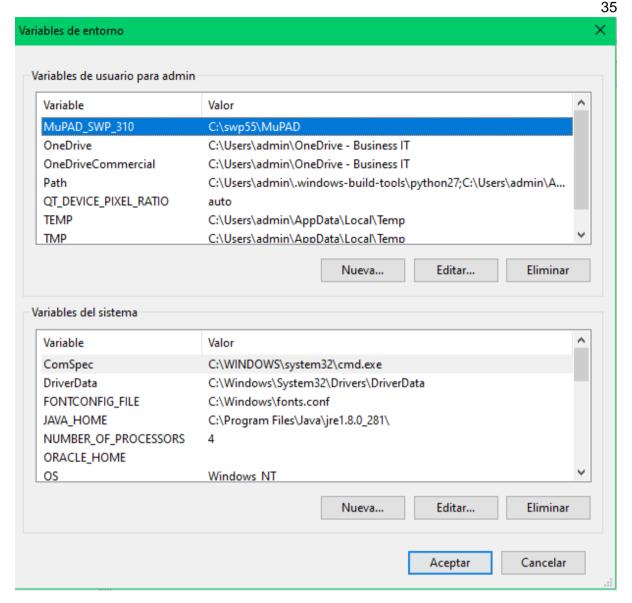


Ahora configuramos la variables de entorno

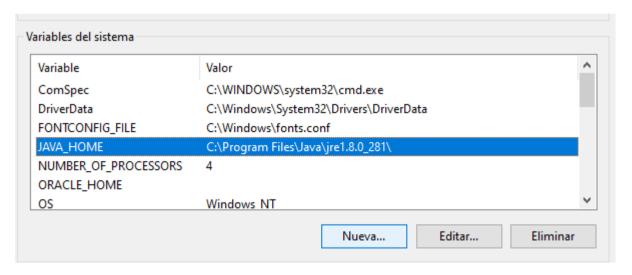


En esta pantalla vamos a configurar



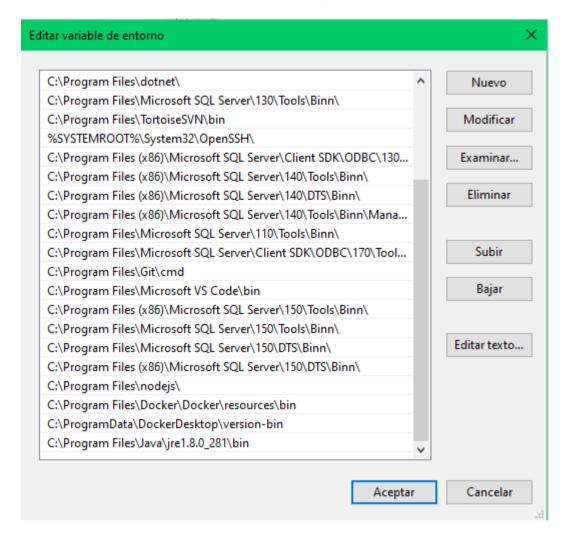


En variables de sistema ingresamos esta nueva variable



36

Y en path le vamos a modificar la ultima variable java



Mediante el cmd

Ingresar a este directorio y ejecutamos el siguiente comando al archivo bat



C:\WINDOWS\system32\cmd.exe

Microsoft Windows [Versión 10.0.19042.928]
(c) Microsoft Corporation. Todos los derechos reservados.

C:\Users\admin>cd ..

C:\Users>cd ..

C:\>cd "Pentaho 9" "

C:\Pentaho 9>cd data-integration

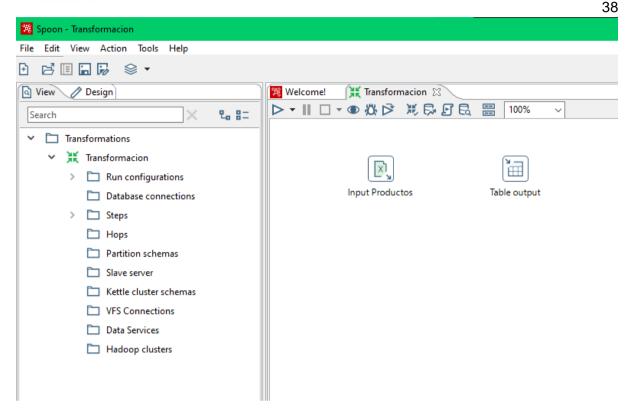
C:\Pentaho 9\data-integration>Spoon.bat

Se abrirá el sistema Pentaho



Se abrirá nuestro sistema Pentaho





Conclusión

En este artículo, aprendimos cómo crear un cubo en SSAS. Para tal fin, primero necesitamos crear una fuente de datos para conectarnos a nuestra base de datos. En este ejemplo, nos conectamos a un almacén de datos ya creado. Para nuestro ejemplo, utilizaremos AdventureworksDW, que es una base de datos de muestra creada con fines de aprendizaje.

El siguiente paso fue agregar la tabla de hechos y 2 dimensiones como vistas de origen de datos. SSAS detectó las claves y relaciones principales de forma ya predeterminada.

Después de eso, agregamos 2 dimensiones basadas en 2 vistas de origen de datos. La moneda y las dimensiones del territorio de ventas.

Finalmente, nosotros creamos un cubo basado en la tabla factresellersales. Seleccionamos algunas columnas como medidas y agregamos las dimensiones creadas. Con los botones de proceso, cargamos datos a los cubos. El cubo como se puede ver es una forma sencilla de generar múltiples informes. Usted Puede crear múltiples informes sin conocimientos de SQL u otras herramientas de informes. Como usted vera Con SSAS puede generar fácilmente miles de informes para satisfacer las necesidades de su empresa.



39