Bases de Datos Avanzadas

Práctica no. 10:

Consultas OLAP para un modelo multidimensional

**Número de Práctica: 10**

**Nombre de la Práctica: Consultas OLAP para un Modelo Multidimensional**

REQUISITOS PREVIOS

Para la realización de la práctica se deberá tener instalado en la computadora el software:

* [Pentaho BI Server](https://drive.google.com/file/d/1hcXyfStuQgf8qxgneyXxzWdhsu1cI4G8/view?usp=share_link)

El procedimiento de ejecución del software se detalla a continuación:

Es necesario descomprimir del archivo *pentaho-server.zip* la carpeta ***pentaho-server*** en algún subdirectorio de la computadora local (por ejemplo c:\temp). Abrir una terminal de comandos del SO, y escribir las siguientes sentencias:

C:\[carpeta]\pentaho-server> **set JAVA\_HOME=C:\[carpeta]\pentaho-server\jdk\jre**

C:\[carpeta]\pentaho-server> **set CLASSPATH=C:\[carpeta]\pentaho-server\jdk\jre\lib**

C:\[carpeta]\pentaho-server> **set PATH=C:\[carpeta]\pentaho-server\jdk\jre\bin**

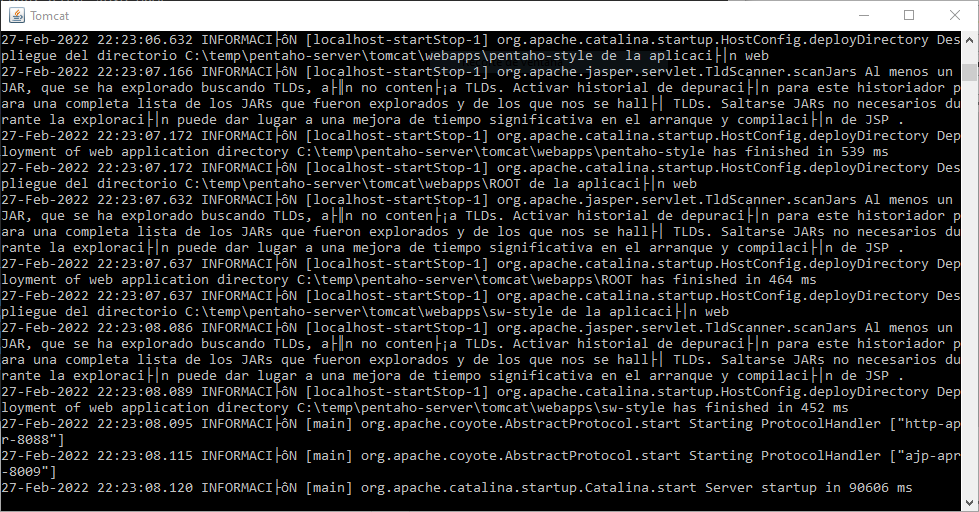
C:\[carpeta]\pentaho-server> **set PENTAHO\_HOME= C:\[carpeta]\pentaho-server**

C:\[carpeta]\pentaho-server> **set-pentaho-env.bat**

En donde [carpeta] se sustituye por la carpeta en donde se descomprimió el software. Si las variables se han configurado correctamente, se procede a ejecutar el servidor Web, con el siguiente comando:

C:\[carpeta]\pentaho-server> **start-server.bat**

Hay que considerar que no debe haber otro software que use el puerto 8088, y que el firewall permita la ejecución del servidor. Espere hasta que el proceso indique que el servidor está listo para recibir peticiones. NO DEBE CERRAR ESTA VENTANA mientras se ejecuta el servidor (para terminar la ejecución teclee CTRL+C)



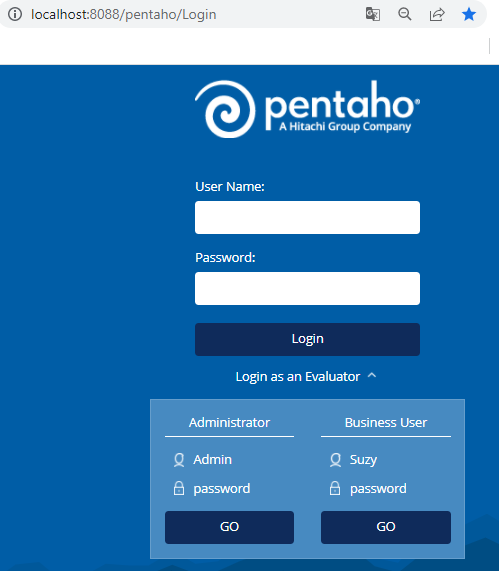
MARCO TEÓRICO

Después de la fase de creación y carga de los datos en el almacén de datos (Data Warehouse), es necesario permitir a un usuario que explote la información con consultas expresadas con una herramienta visual de consultas OLAP (*Pentaho BI Server*). Mediante la combinación de las tecnologías *Mondrian* (motor OLAP) y *JPivot* (visualizador de consultas OLAP) incluidas en el servidor *Pentaho BI*, es posible generar rejillas de datos sumarizados con las dimensiones mostradas las columnas o renglones y los hechos como contenido de la rejilla. Además, es posible generar gráficos sencillos que complementen los resultados.

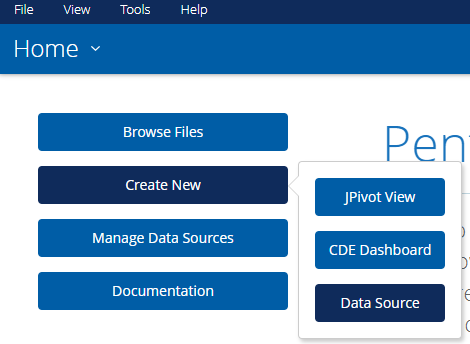
DESARROLLO DE LA PRÁCTICA:

1.- Para poder emplear consultas OLAP, deberá tener funcionando el servidor *Oracle* con la base de datos **dwh** que corresponde al modelo multidimensional, y con los datos cargados como se hizo en la práctica anterior (si no tiene cargados los datos, refiérase al ***anexo 1*** para el procedimiento de importación de datos en formato CSV).

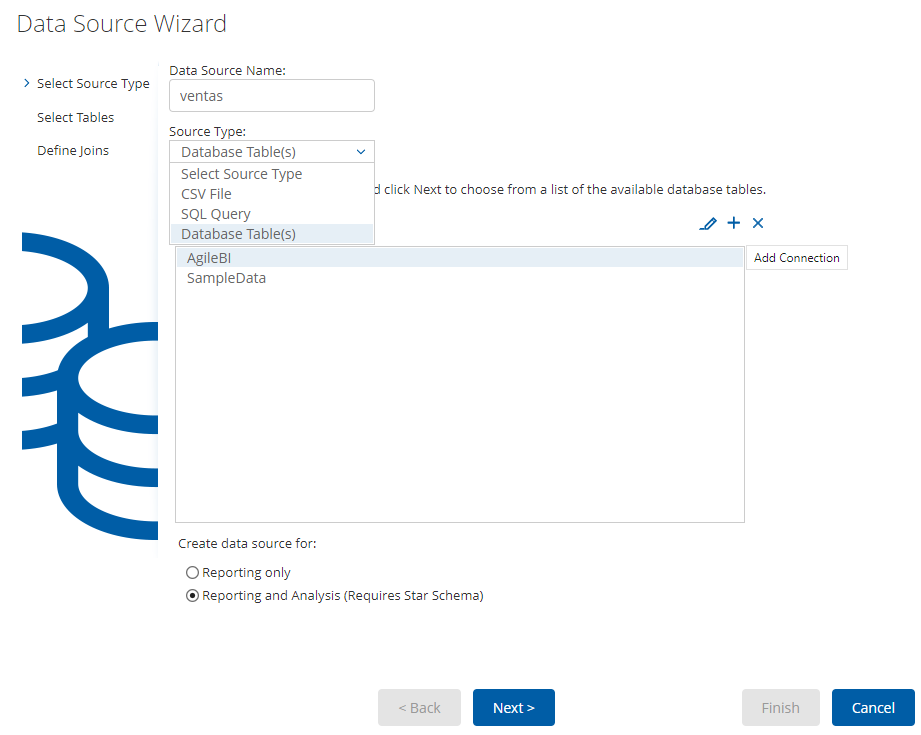
2.- Abra un navegador Web (el de su elección) y escriba en el campo de URL la dirección [*http://localhost:8088*](http://localhost:8088), de manera que solicite conectarse al servidor. Si todo es normal, se presenta la pantalla de bienvenida del sistema, en el cual se muestran dos roles: **Administrator** y **Business User**. Elija el botón correspondiente a Administrator.



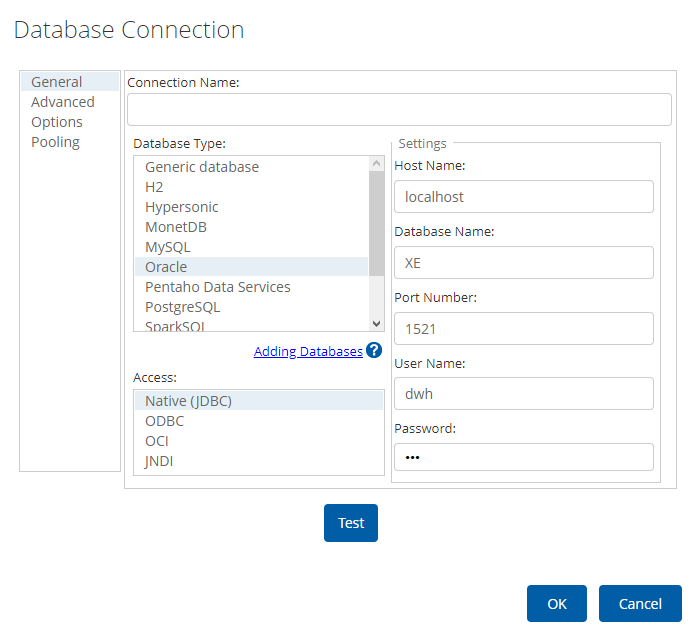
3.- Una vez iniciada la sesión, se deberá elegir la opción **Create New** > **Data Source** del menú principal.



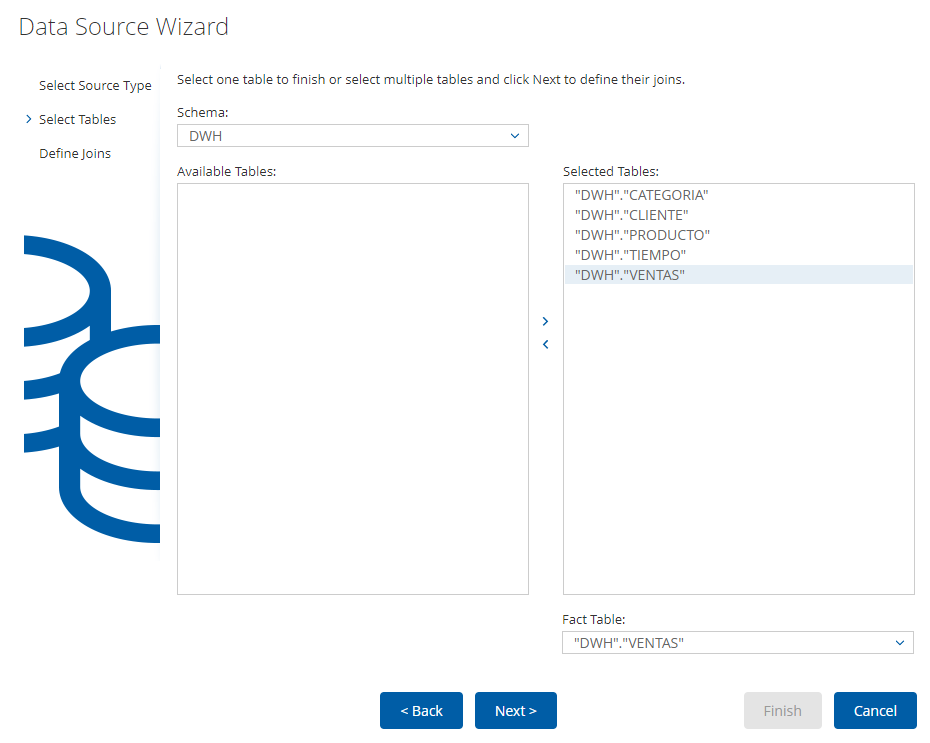
4.- Posteriormente un asistente en pantalla preguntará por el nombre y el tipo de la conexión. De la lista desplegable elegir la opción **Database Table(s)** y en la parte superior del panel que aparece, hay que elegir el símbolo (**+**) para agregar una nueva conexión. También hay que asegurarse que esté seleccionada la opción **Reporting and Analysis** .



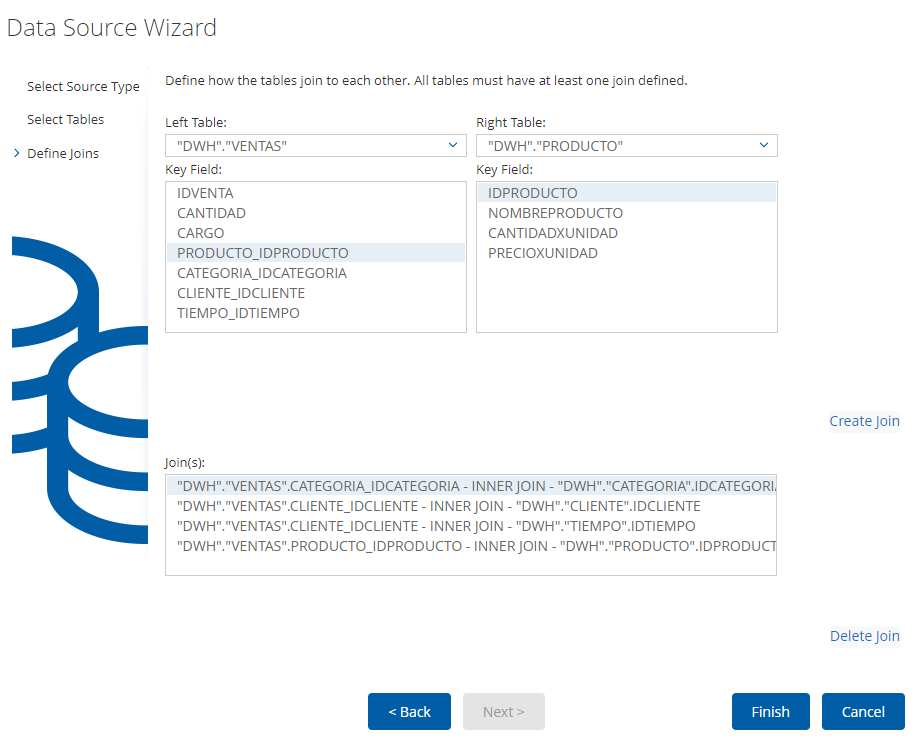
5.- En el siguiente panel, hay que elegir la base de datos a usar. En este caso, elegir la opción **Oracle**, con los parámetros de conexión correspondientes como se muestra en la figura (Nota: algunos valores pueden diferir en los sistemas usados). Asegúrese de dar el nombre *multidimensional* a la conexión. Comprobar la correcta conexión con el botón **Test** y dar **OK**.



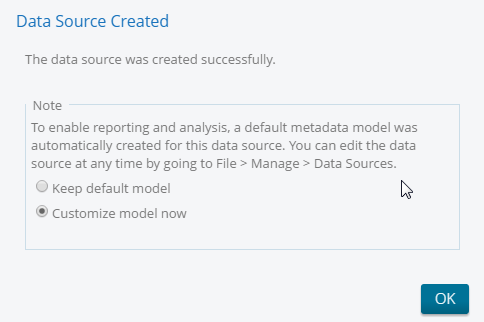
6.- En el siguiente paso hay que elegir las tablas que participaran en el modelo, mediante los botones < y >. Hay que considerar que la tabla de hechos debe estar elegida de la lista desplegable inferior. Dar clic en **Next**.



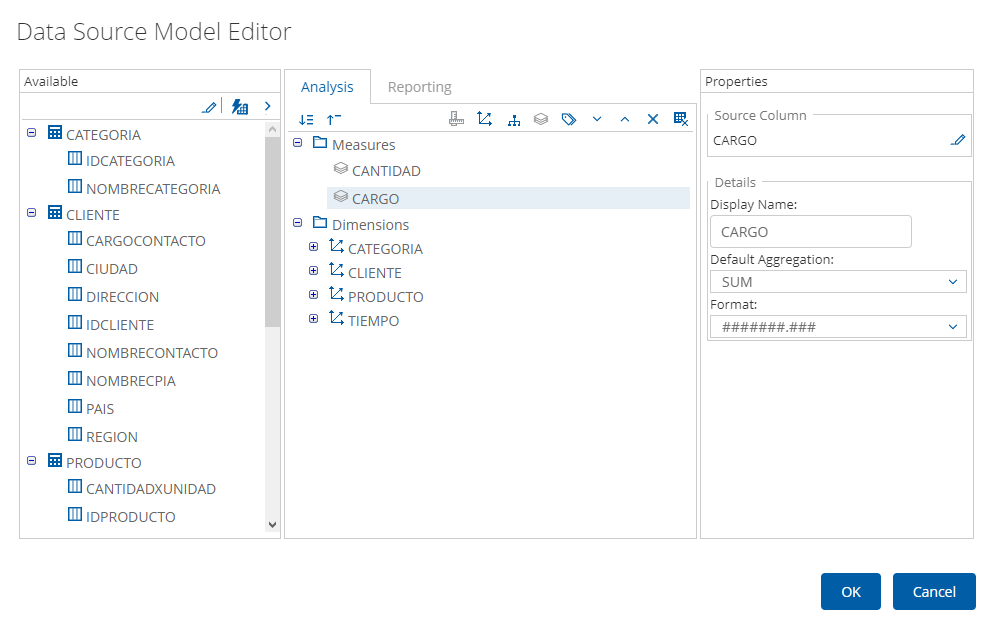
7.- Posteriormente hay que establecer las relaciones entre las tablas de dimensiones y la tabla de hechos. Cada vez que ligue una tabla con otra, agregue la relación con el botón **Create Join**. Si hubo algún error, elimine la relación con el botón **Delete Join**. Observe la siguiente figura para verificar que las relaciones establecidas sean las correctas. De clic en **Finish**.



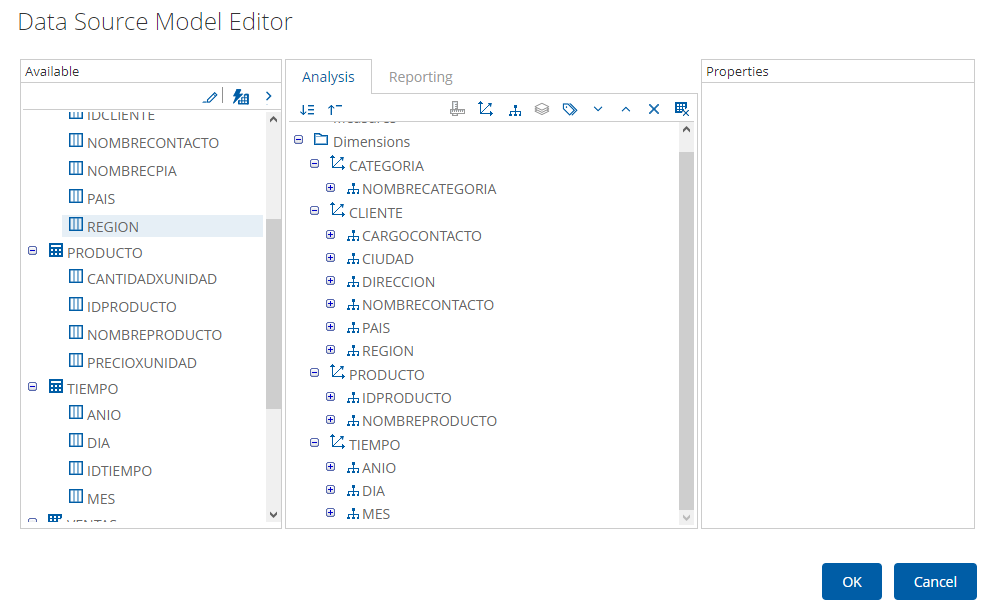
8.- Ahora aparece una ventana en donde se pregunta si se queda así el modelo creado, o si se debe editar. Elija la opción **Customize model now** y dé clic en el botón **OK**.



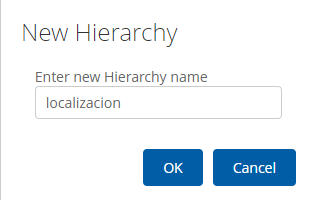
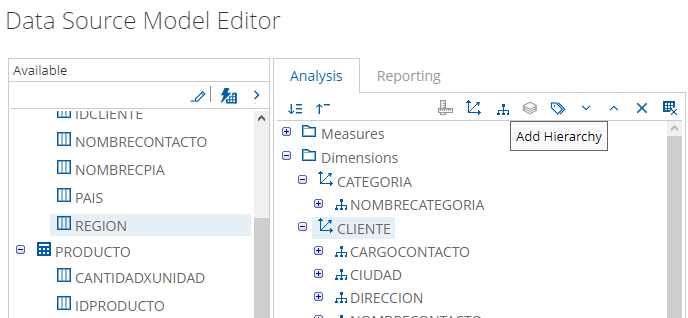
9.- Aparece la pantalla con el editor del modelo, en el cual se podrán establecer las columnas y las propiedades que tendrán para las consultas OLAP. Observe que en el panel izquierdo aparecen las columnas de las tablas físicas para agregar, y en el panel derecho las configuraciones del modelo multidimensional. En el grupo **Measures** encontrará las columnas de medidas, en donde solo deben aparecer las correspondientes a los atributos de medidas. Con el botón (X) elimine todas las columnas de llaves foráneas y el *IdVenta*, y conserve *Cantidad y Cargo* para la tabla de hechos. En el panel de propiedades configure la función de agregación correspondiente: *Cargo* con **Sum**, y *Cantidad* con **Count**. Elija el formato correspondiente a **#** para *Cargo* y **0.00** para *Cantidad*. Observe la siguiente figura para revisar la configuración de las medidas.



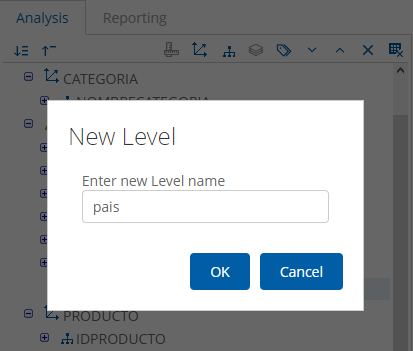
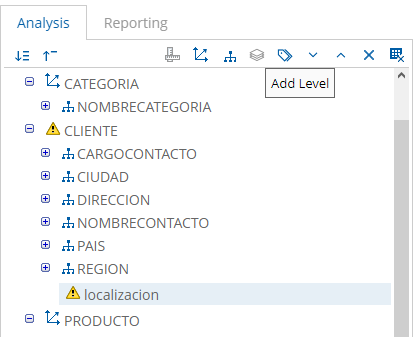
10.- Expanda la carpeta **Dimensions** y quite los atributos de llave primaria y los innecesarios (aquellos que no se desean en el modelo) de las tablas *Categoría, Producto y Cliente* empleando el botón **X** del menú superior en la pestaña **Analysis**. Las dimensiones con sus niveles quedarán como en la figura siguiente:



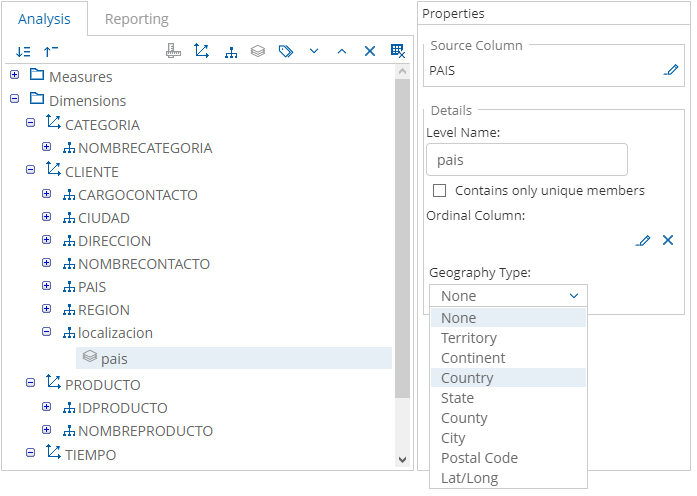
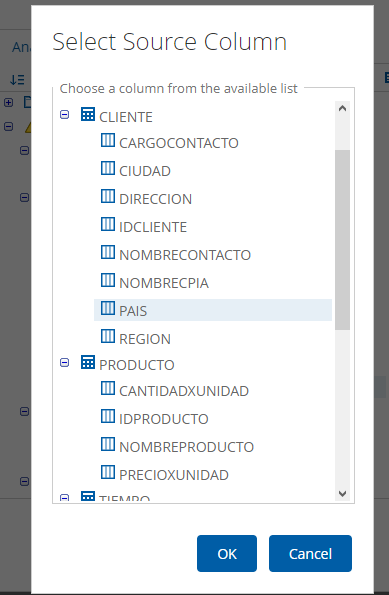
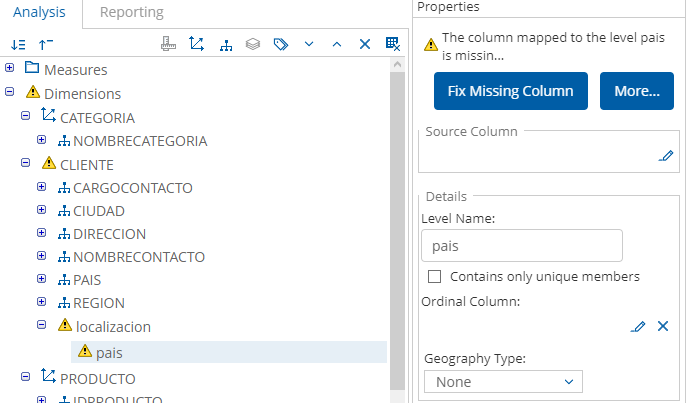
11.- En la dimensión de *Cliente*, seleccione el símbolo de jerarquías ***(Add Hierarchy)*** para crear una nueva, nómbrela **Localización**.



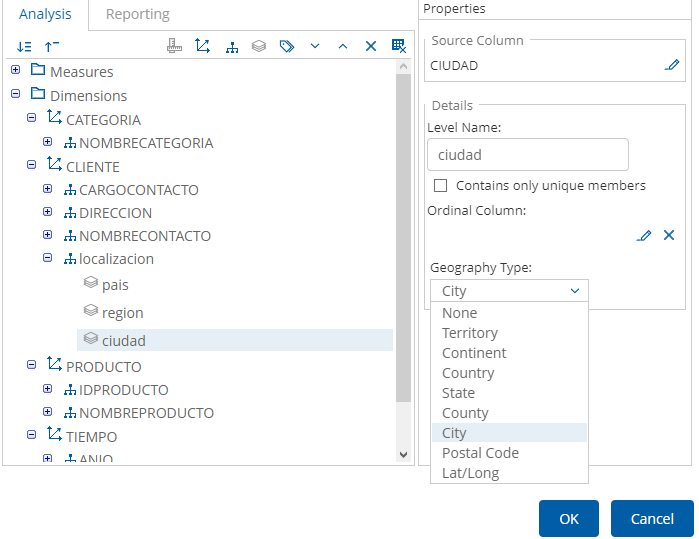
12.- De la nueva dimensión creada, elija el ícono de **Add Level,** y agregue el nivel *Pais*.



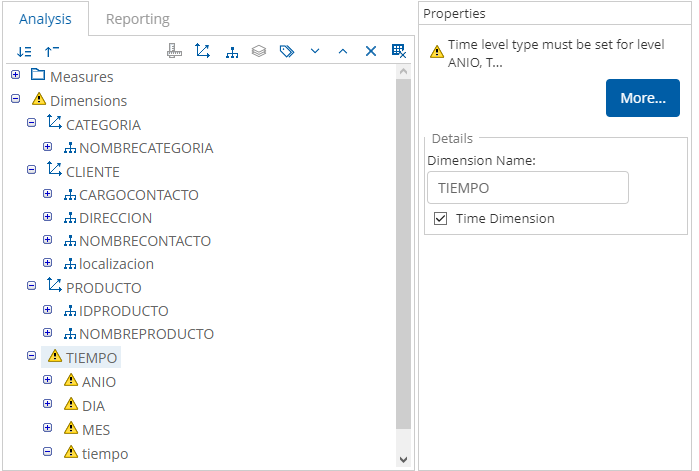
13.- En el panel derecho aparece un botón **Fix Missing Column**, para reparar los niveles erróneos. Oprímalo y elija la columna *País* de la tabla *Cliente* y dé clic en **OK.** Observe que ahora es parte de la jerarquía. Elija del panel derecho la opción *Country* de la lista **Geography Type**. Quite la columna **PAIS** de *Cliente*.



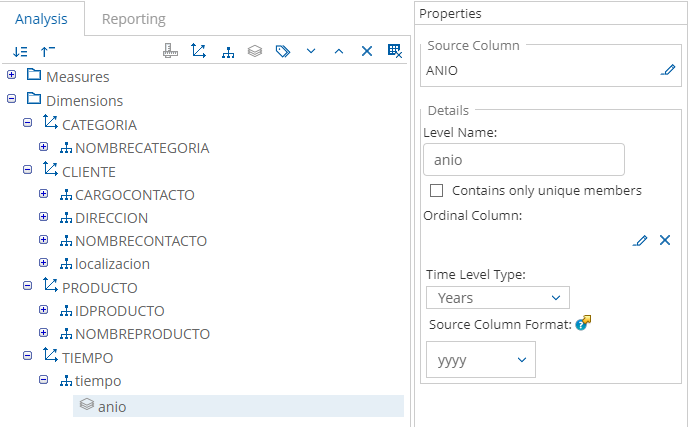
14.- Repita el procedimiento para los niveles *Region y Ciudad*, con los correspondientes tipos **State** y **City**. Elimine los niveles *Región y Ciudad* de la dimensión *Cliente*. La configuración debe quedar como se muestra en la siguiente figura:



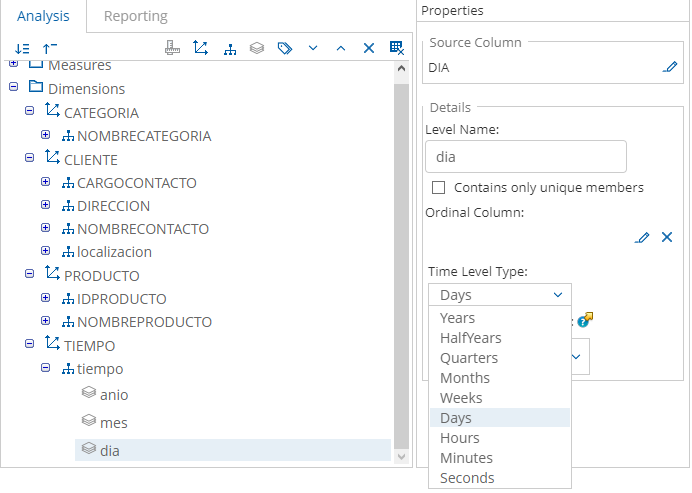
15.- Ahora inserte una nueva jerarquía *Tiempo* en la dimensión *Tiempo*, justo como se hizo en el paso anterior, pero estableciendo el tipo de jerarquía como de tipo **Time Dimension**.



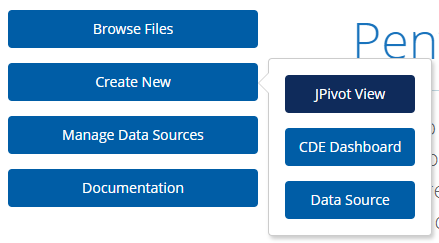
16.- Agregue a esta jerarquía los niveles *Anio, Mes y Dia*, y establezca los tipos de medidas a **YEARS, MONTHS y DAY**S respectivamente, con el formato YYYY, MMM y DD para cada una de ellas. Observe la siguiente figura para el caso del nivel *anio*.



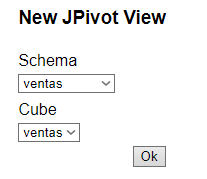
17.- Elimine los niveles *Anio, Mes y Dia* originales de la dimensión *Tiempo*. El resultado debe ser como se muestra en la siguiente figura:



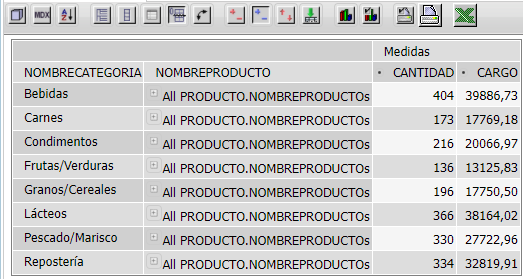
18.- Guarde el trabajo con el botón **OK**. Ahora elija la opción **Create New > JPivot View** del menú principal.



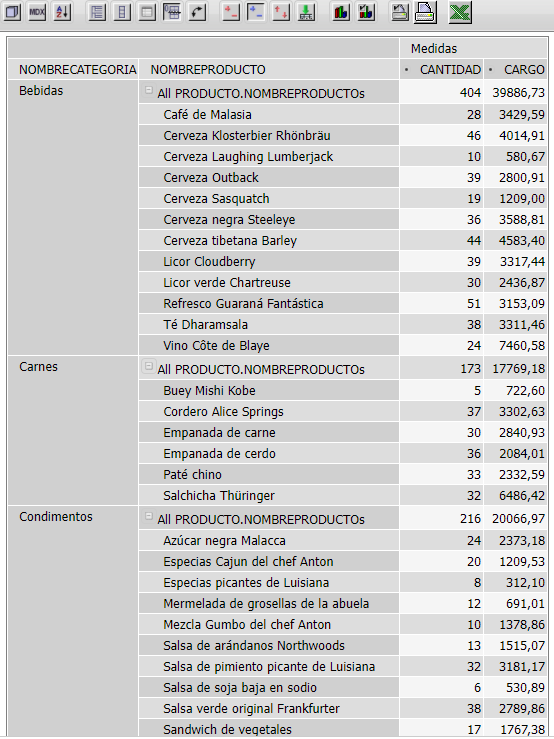
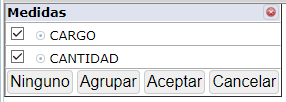
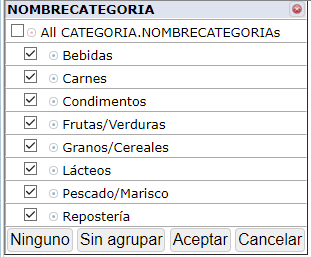
19. Elija la fuente de datos *ventas* creada anteriormente y el cubo *ventas*.



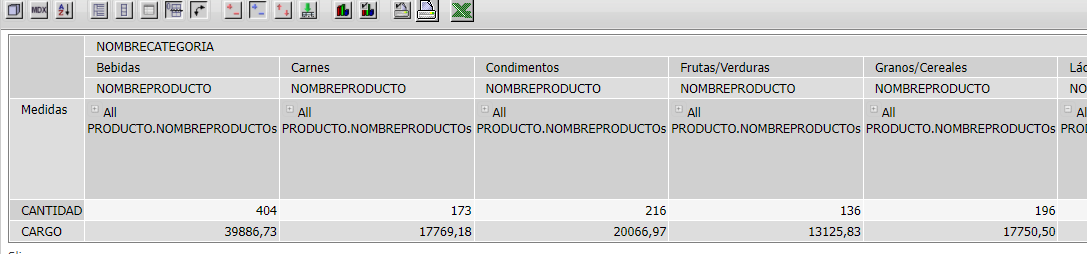
20.- Observe la rejilla que se creó. Expanda los datos para ver sus resultados.



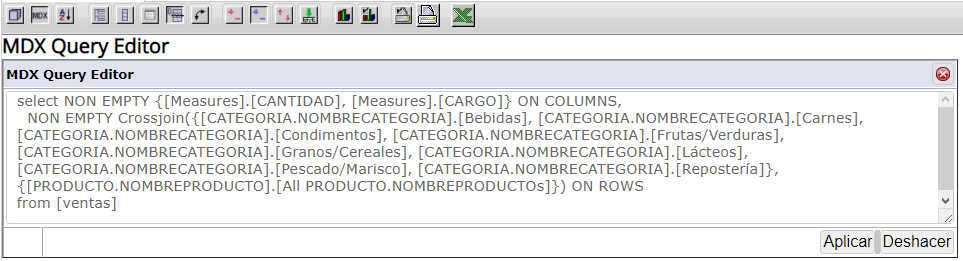
21.- Con el botón de ***cubo* (primero de izquierda a derecha)**, modifique las dimensiones y las medidas a mostrar. Establezca en las columnas las medidas *Cargo y Cantidad*, en las columnas las **Medidas** y el **Tiempo**, y en los renglones los nombres de categorías, y la descripción de productos. Observe los resultados obtenidos. Verifique el funcionamiento expandiendo los niveles de tiempo y de productos.



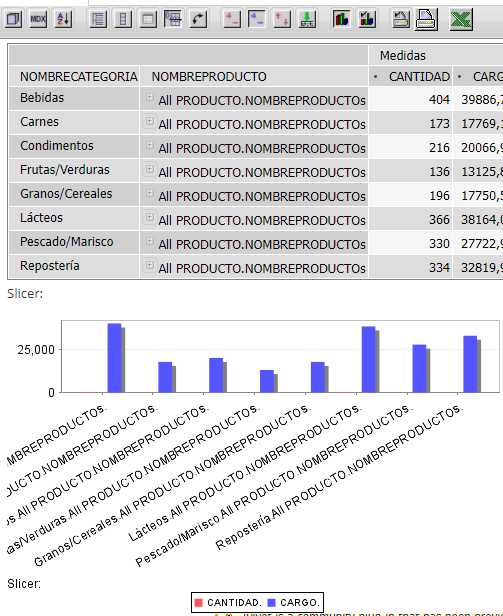
22.- Del menú oprima el botón *Intercambiar Ejes* y observe los resultados



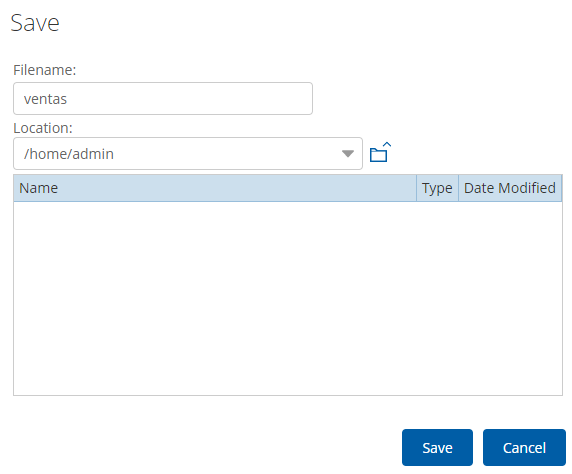
23.- Mediante el botón de **Editor** **MDX**, se observa el código generado por la herramienta.



24.- Mediante el botón de **Mostrar Gráfico**, se muestra una gráfica para los totales de categorías y productos.

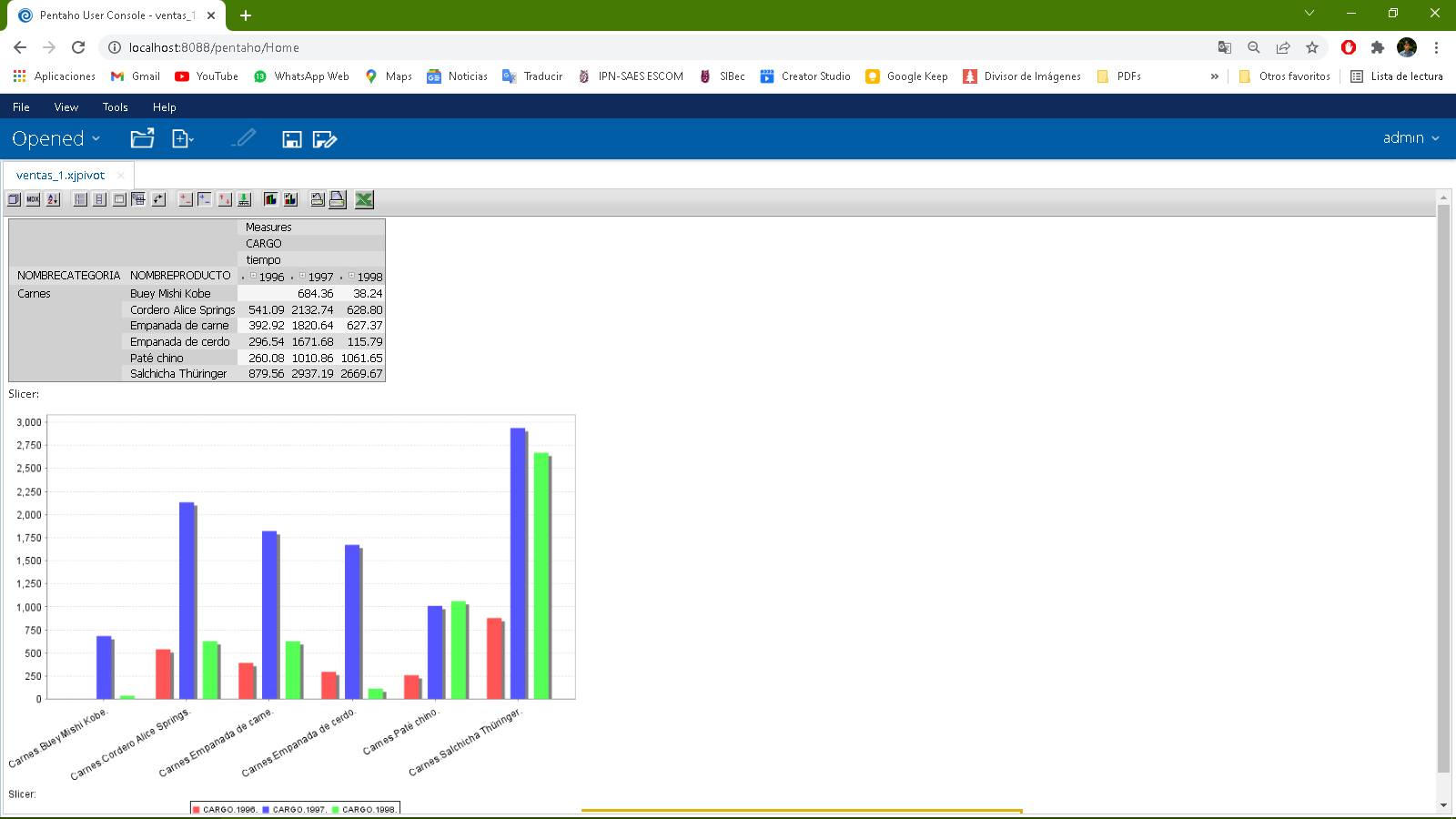


25.- Con el icono de **Save**, guarde su trabajo, establezca el nombre a *ventas*.



26.- Realice las siguientes consultas (incluya captura de los resultados obtenidos, así como una gráfica de datos diferente para cada inciso):

a) Obtener el listado de las ventas totales (*cargo*) de los productos de la categoría de **Carnes** en los tres años (1996, 1997 y 1998).



b) Obtener el listado de las ventas totales (*cantidad*) del producto **Caviar Rojo** en el año de 1998 por cada mes.

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza media

c) Obtener el listado de las ventas totales obtenidas (*cargo y cantidad*) de los clientes que viven en **Alemania**, para los tres años (1996, 1997 y 1998).

Interfaz de usuario gráfica

Descripción generada automáticamente con confianza media

Dependiendo de los resultados obtenidos, muestre una gráfica de barras, de barras apiladas y de pastel.

CUESTIONARIO

1.- Describa las operaciones que se aplican a las consultas OLAP para los datos en una rejilla.

Entre las operaciones que se pueden aplicar está “drill down”, “roll up”, “slice”, “dice” y “pivot”.

* “Drill down” consiste en que dado un cubo de datos se baja dentro de los niveles de una jerarquía de datos, es decir, ir de lo general a lo particular en los datos.
* “Roll up” es básicamente la operación inversa de drill down, ya que en lugar de bajar se sube en la jerarquía, resumiendo de esta manera los datos.
* “Slice” refiere a poner un valor fijo/constante sobre una dimensión y con ello visualizar una selección muy específica de las otras dimensiones. Se podría decir que es un caso particular de “Dice”.
* “Dice” refiere a fijar valores específicos en múltiples dimensiones y con ello obtener un sub-cubo de datos más complejo que una “Slice”.
* “Pivot” consiste en meramente intercambiar ciertas dimensiones del cubo de datos y así ver los datos rotados desde otra perspectiva.

2.- ¿Qué tipos de consultas serán las más adecuadas de resolver en una herramienta OLAP?

Las consultas más adecuadas mediante una herramienta OLAP serán aquellas en donde se muestren los valores de nuestras métricas; que prácticamente son los atributos propios de la tabla de hechos de un modelo multidimensional, excluyendo aquellos campos que son enlaces con las tablas de dimensiones. De este modo, lo interesante de las consultas OLAP es visualizar la variabilidad de las métricas respecto a valores específicos de diferentes dimensiones del cubo de datos

CONCLUSIONES

Escriba sus conclusiones (no personales) con respecto al desarrollo de esta práctica.

En conclusión, gracias a Pentaho server se obtuvieron satisfactoriamente los resultados para la métrica de ventas totales respecto a valores específicos de los clientes, tiempo, productos y categorías. Además, se graficaron dichos resultados, lo cual permitió un mejor entendimiento de las ventas totales dependiendo de la consulta. Aunado a esto, se puede decir que las herramientas OLAP son de ayuda para resolver preguntas de negocio, ya que a partir de los datos estructurados bajo un modelo multidimensional, estos se pueden adaptar en un cubo de datos que posteriormente puede ser fácilmente manipulado para obtener insights dependiendo de las dimensiones.

REFERENCIAS BIBLIOGRÁFICAS (APA, IEEE, ETC.)

1.- Ralph Kimball: “The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses” , John WIley & Sons, 1996. ISBN 0-471-15337-0

2.- Jiawei Han & Micheline Kamber “Data Mining: Concepts and Techniques” , Morgan Kaufmann, 2000. ISBN 1-55860-489-8

3.- C.J. Date: “Introducción a los sistemas de bases de datos” , Pearson Educación, 2001. ISBN 968-444-419.

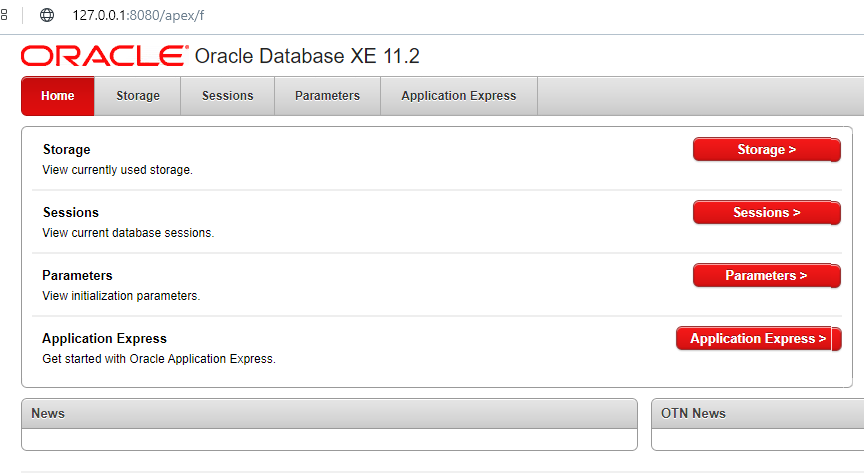
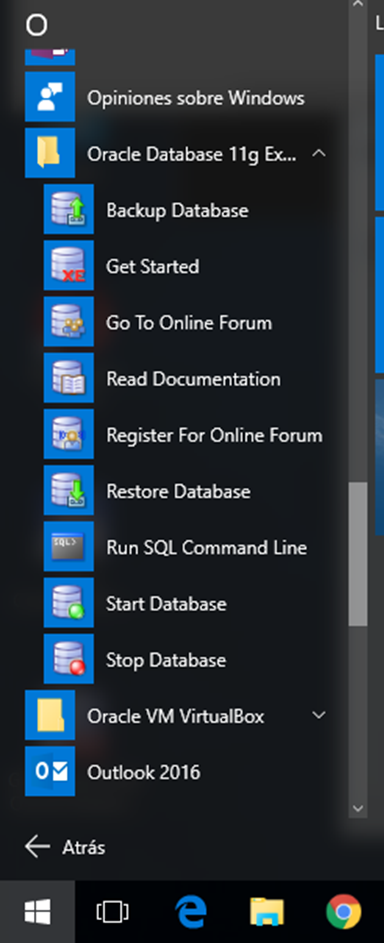
4.- Ramez A. Elmasri & Shamkant B. Navathe: “Fundamentos de Sistemas de Bases de Datos”, Addison-Wesley, 2002 [3ª edición]. ISBN 84-782-9051-6.

5.- Henry F. Korth, Abraham Silberschatz & S. Sudarshan: “Fundamentos de Bases de Datos”, Mc-Graw Hill, 2002 [4ª edición]. ISBN 84-481-3654-3.

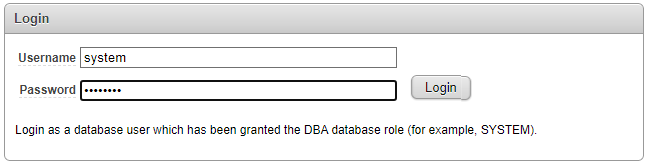
**Anexo 1**

**Procedimiento para cargar los datos en el modelo multidimensional**

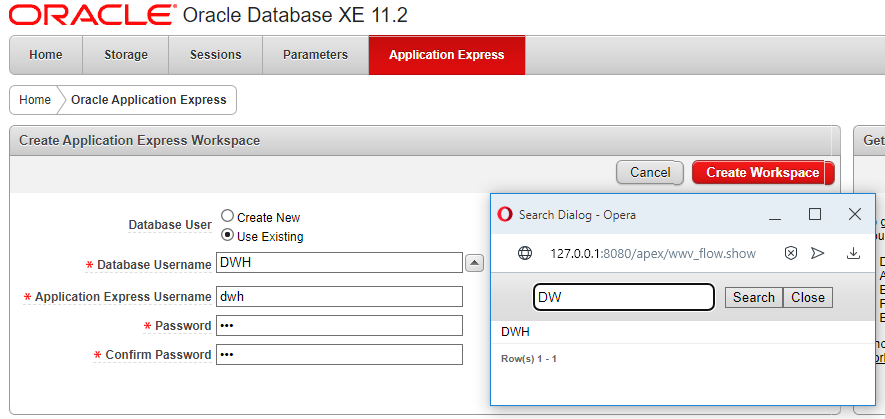
1.- Descargue los archivos [categoria.csv](https://drive.google.com/file/d/1idd-ZEFTGOtjc_bMrQbvfvNdjF1KRkRy/view?usp=share_link), [cliente.csv](https://drive.google.com/file/d/1ERbRkIkIAZhr2xvHdUY-K3Qy-jZEloGh/view?usp=share_link), [producto.csv](https://drive.google.com/file/d/1P1cMduQqgHNZ-74JHuhkzZIg72G3Mfa4/view?usp=share_link), [tiempo.csv](https://drive.google.com/file/d/1gxrNFPaK81010TEnmZCm3O1w9dtuGFX1/view?usp=share_link) y [ventas.csv](https://drive.google.com/file/d/1aMIousArJVld_fi2JUVGfhYRCqMCaBbF/view?usp=sharing), y guárdelos en alguna carpeta de la computadora local. Inicie desde el menú de inicio de Windows el ícono de **Get Started**. Se abrirá el navegador Web por omisión con la siguiente pantalla:



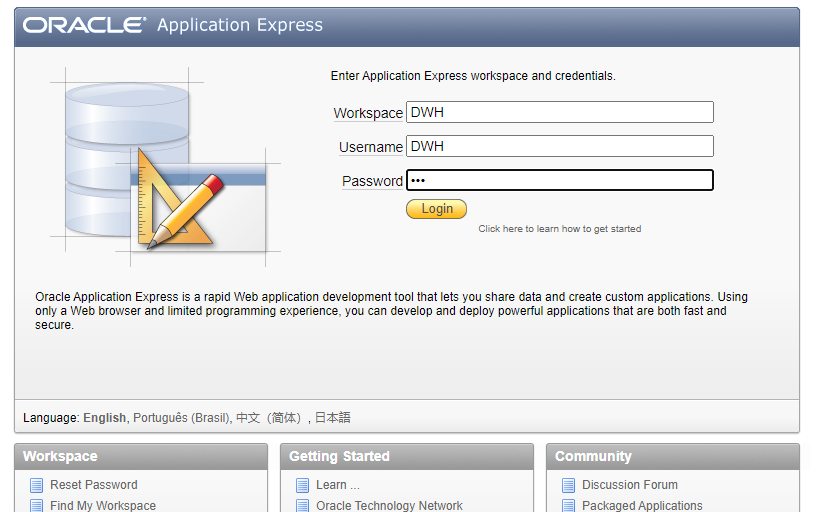
2.- Elija la opción de **Application Express**. Le pedirá el usuario administrador. Escriba como usuario *system* y la contraseña que haya configurado en la instalación.



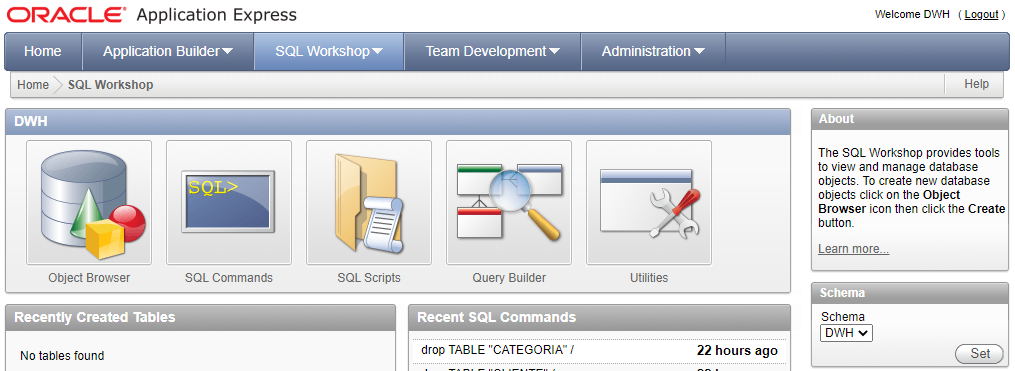
3.- Si el usuario *dwh* ya existe, entonces elegir la opción **Use Existing** y escribir los datos de la cuenta, así como la contraseña del usuario. Puede usar el buscador de usuario para verificar su existencia:



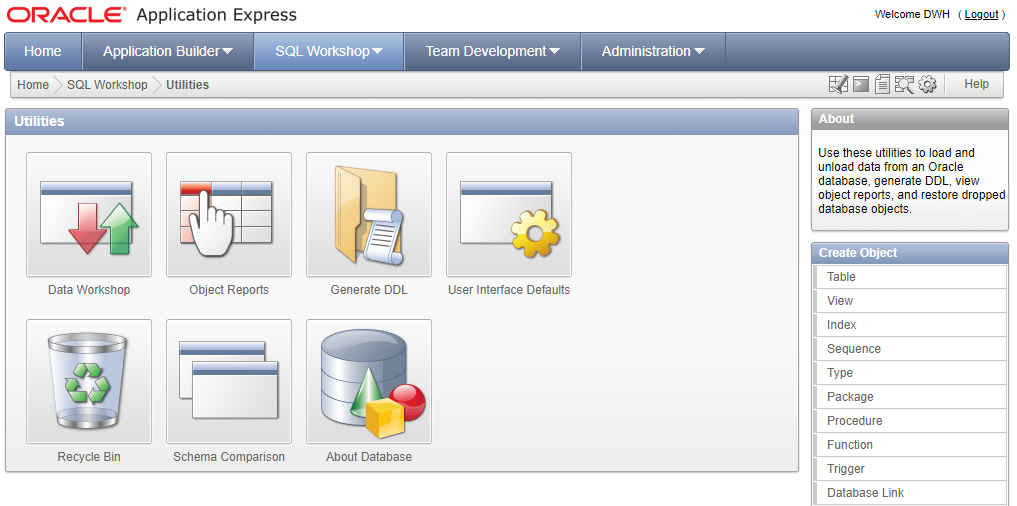
4.- Posteriormente aparece la pantalla de inicio de sesión en el ambiente gráfico. Escriba la contraseña del usuario *dwh* para continuar.



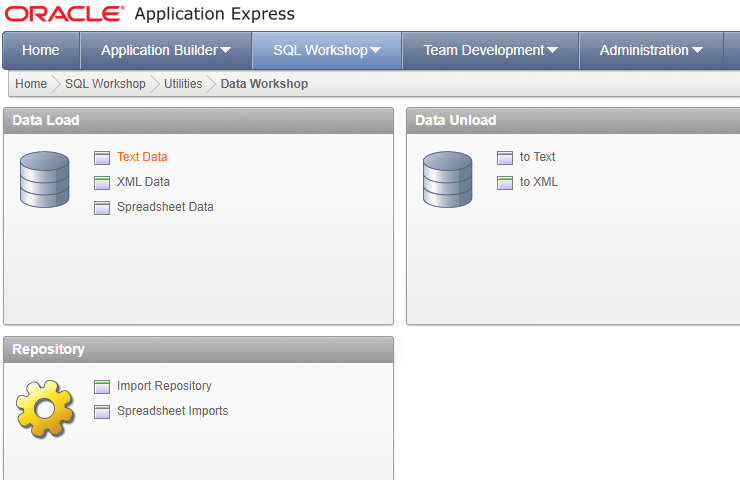
5.- Se muestra la pantalla del ambiente de desarrollo basado en Web, con las opciones de **Application** **Builder** (para generar aplicaciones basadas en Web con la base de datos de Oracle), **SQL Workshop** (para el trabajo con SQL y la terminal de comandos), **Team Development** (para el desarrollo colaborativo) y **Administration** (para configurar algunas opciones del servidor). Elija la opción **SQL Workshop**, que muestra los módulos de **Object Browser** (para buscar objetos creados en la base de datos por el usuario), **SQL Commands** (es la terminal de comandos SQL), **SQL Scripts** (para guardar y recuperar código previo guardado como scripts de SQL), **Query** Builder (para generar consultas de forma gráfica) y **Utilities** (con varias opciones para administrar la base de datos). Elija la opción de **Utilities**:



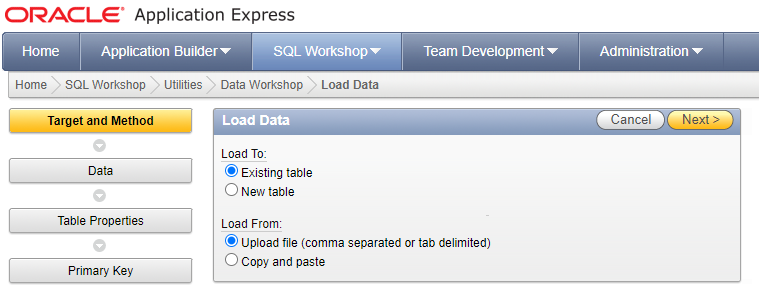
6.- De la pantalla que se muestra, se va a emplear la opción de **Data Workshop.**

`´

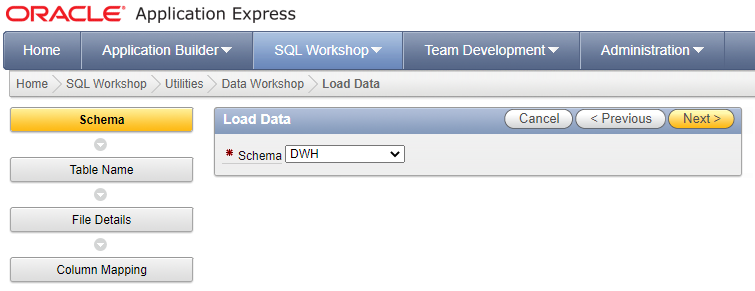
7.- En la siguiente pantalla se muestran las posibles operaciones para el tratamiento de datos externos: **Data Load** (permite importar archivos de texto, en formato XML y de hojas de cálculo en el programa Excel) y **Data Unload** (que permite exportar contenido de la base de datos en formatos de texto CSV y XML). Seleccione la opción de **Data Load/ Text Data**.



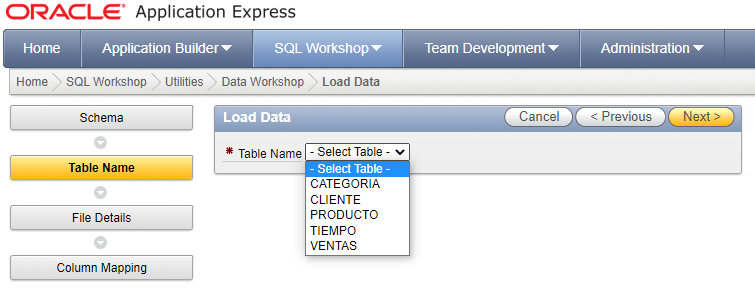
8.- Posteriormente se inicia un asistente de importación, en el que muestra pantallas consecutivas con los pasos y las opciones para importar archivos. En la primera pantalla se solicita indicar si se desea crear una nueva tabla o ya existe una previamente; elegir la opción **Existing Table** y **Upload file (comma separated or tab delimited)**. Continuar con **Next**.



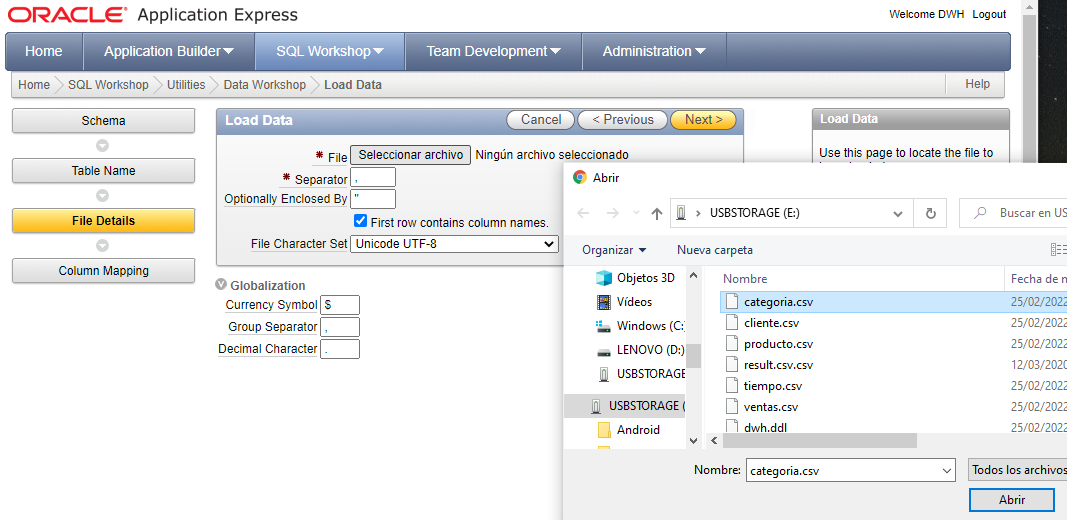
9.- A continuación se solicita el esquema de la base de datos el cual contiene las tablas del modelo multidimensional. Elegir de las opciones *DWH* y continuar con **Next**.



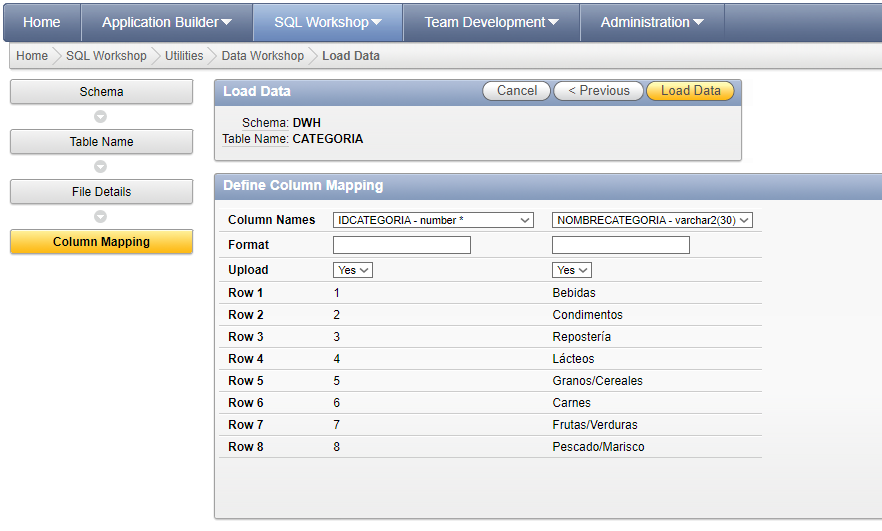
10.- Se pide el nombre de la tabla en la cual se desean importar los datos. Elegir **CATEGORIA** y continuar con **Next**



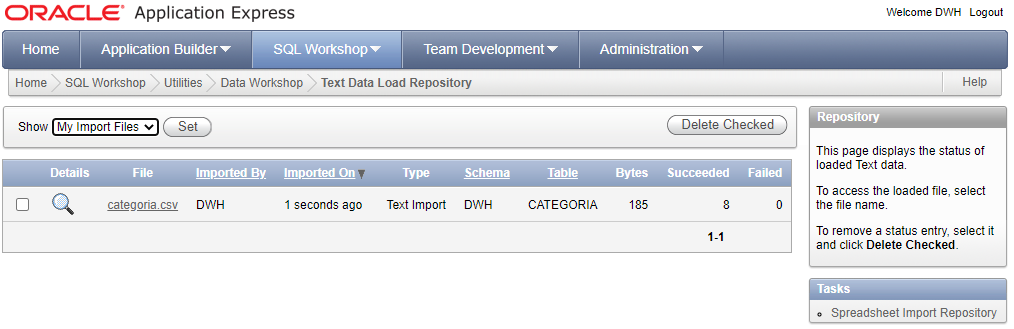
11.- En la pantalla siguiente se muestran las opciones de importación, en las cuales se debe indicar la ruta donde se encuentra el archivo a importar (carpeta donde se descargaron los archivos *csv*), el separador de datos (por omisión es la coma **,** no modificarlo), el carácter delimitador de textos (colocar aquí el carácter de comillas dobles **”**), así como si el primer renglón del archivo contiene los identificadores de las columnas a importar (dejar seleccionada). No modificar las demás opciones, vigilando que conjunto de caracteres internacional sea *Unicode UTF-8* (valor por omisión). Continúe con **Next**.



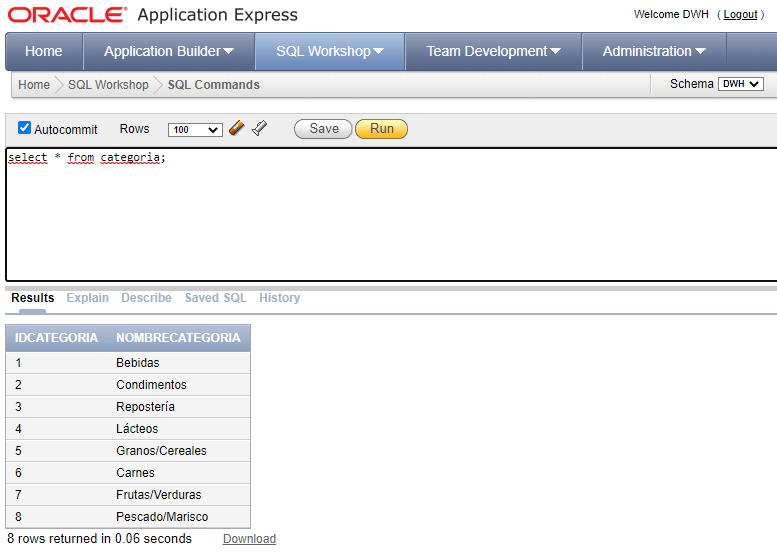
12.- Por último, se presenta una muestra de 20 registros con sus respectivas columnas constitutivas, para que pueda ser revisado y confirmado. Se recomienda no modificar ninguna de las opciones a menos que sea necesario, y se finaliza la importación con el botón **Load Data**.



13.- Se muestra un resumen del proceso realizado. Debe vigilar que el número de registro con errores sea 0 (en la columna **Failed**). En caso de que haya ocurrido un error, aparecerán los motivos en el panel inferior, por lo que deberá revisar el error y corregirlo.



14.- Puede verificar que los datos fueron correctamente importados realizando una consulta SQL en la terminal de comandos. Dentro de la misma página Web, vaya a **Home/SQL Workshop/SQL Commands** y ejecute una sentencia de tipo *select … from …* con cada una de las tablas del esquema y vea que los datos existen y son correctos.



15.- Repita para los datos restantes, en el siguiente orden: **clientes**, **productos**, **tiempo** y por último **ventas**.