

Guía Básica de Diseño de Cubos con Mondrian Schema Workbench

Autores: Alejandro Maté Morgia
Juan Carlos Trujillo Mondéjar

Fecha: Febrero 2020



Universitat d'Alacant
Universidad de Alicante



1. Introducción

La creación de cubos OLAP resulta un aspecto fundamental para poder realizar un análisis multidimensional ad-hoc en una plataforma de Inteligencia de Negocio. Dentro de la plataforma Pentaho, el proceso de análisis ad-hoc se lleva a cabo mediante el servidor de cubos OLAP Mondrian, incluido dentro de la instalación del propio servidor de Inteligencia de Negocio.

Mondrian es un servidor de tipo ROLAP (Relational OLAP). Utiliza una base de datos relacional y un esquema en formato XML para otorgarle una vista multidimensional a los datos. La definición de estos esquemas se puede realizar mediante el programa Mondrian Schema Workbench.

A lo largo de esta guía veremos un ejemplo práctico de creación de un cubo de análisis con Mondrian Schema Workbench, desde la creación de las dimensiones y niveles de jerarquía a la validación de las columnas mediante una conexión con la base de datos.

2. Descarga e Instalación de Mondrian Schema Workbench

Para comenzar es necesario tener instalado y configurado Java 1.8 o superior¹. A continuación, puede descargarse el software Mondrian Schema Workbench para Windows o Linux en la siguiente dirección:

Versión 8.0: <https://sourceforge.net/projects/pentaho/files/Pentaho%208.0/client-tools/psw-ce-8.0.0.0-28.zip/download>

Versión: 8.2: <https://sourceforge.net/projects/pentaho/files/Pentaho%208.2/client-tools/psw-ce-8.2.0.0-342.zip/download>

Al extraer el archivo comprimido, obtendremos un directorio como el que aparece a continuación:

¹ Recomendamos el uso de la versión de Java (JDK) 1.8, pues es la que requiere Pentaho Server 7 para funcionar correctamente.

Nombre	Fecha de modifica...	Tipo	Tamaño
demo	05/11/2017 15:04	Carpeta de archivos	
drivers	05/11/2017 15:04	Carpeta de archivos	
lib	05/11/2017 15:04	Carpeta de archivos	
plugins	05/11/2017 15:04	Carpeta de archivos	
Schema_Workbench.app	05/11/2017 15:04	Carpeta de archivos	
cpappend.bat	05/11/2017 15:04	Archivo por lotes ...	1 KB
LICENSE.html	05/11/2017 15:04	Archivo HTML	13 KB
log4j.xml	05/11/2017 15:04	Documento XML	4 KB
mondrian.properties	05/11/2017 15:04	Archivo PROPERTI...	42 KB
PentahoSchemaWorkbench_OSS_License...	05/11/2017 14:13	Archivo HTML	1.871 KB
readme.txt	05/11/2017 15:04	Documento de tex...	1 KB
set-pentaho-env.bat	05/11/2017 15:04	Archivo por lotes ...	5 KB
set-pentaho-env.sh	05/11/2017 15:04	Shell Script	5 KB
workbench.bat	05/11/2017 15:04	Archivo por lotes ...	2 KB
workbench.command	05/11/2017 15:04	Archivo COMMA...	1 KB
workbench.sh	05/11/2017 15:04	Shell Script	2 KB

Figura 1. Directorio de Mondrian Schema Workbench

En el directorio extraído encontraremos distintas carpetas y archivos. Para la práctica, el directorio más relevante es "drivers". En este directorio se han de colocar los JDBC necesarios para las bases de datos con las que queramos conectarnos. Dado que en la práctica utilizaremos MySQL, necesitaremos el driver de conexión correspondiente. Podéis descargar dicho driver desde la siguiente dirección seleccionando "Platform Independent" y moviendo el archivo a la carpeta "drivers":

<http://dev.mysql.com/downloads/connector/j/#downloads>

Nombre	Fecha de modifica...	Tipo	Tamaño
src	09/03/2015 18:48	Carpeta de archivos	
docs	09/03/2015 18:48	Carpeta de archivos	
mysql-connector-java-5.1.34-bin.jar	17/10/2014 9:05	Executable Jar File	938 KB
build.xml	17/10/2014 9:05	Documento XML	88 KB
README.txt	17/10/2014 9:05	Documento de tex...	63 KB
README	17/10/2014 9:05	Archivo	60 KB
COPYING	17/10/2014 9:05	Archivo	18 KB
CHANGES	17/10/2014 9:05	Archivo	226 KB

Figura 2. Extracción del driver de MySQL dentro de la carpeta drivers

Una vez se encuentre el conector extraído en la carpeta "drivers", podemos ejecutar el

Schema Workbench mediante el archivo "workbench.bat" (o .sh) de la carpeta raíz:

Nombre	Fecha de modifica...	Tipo	Tamaño
demo	05/11/2017 15:04	Carpeta de archivos	
drivers	05/11/2017 15:04	Carpeta de archivos	
lib	05/11/2017 15:04	Carpeta de archivos	
plugins	05/11/2017 15:04	Carpeta de archivos	
Schema_Workbench.app	05/11/2017 15:04	Carpeta de archivos	
cpappend.bat	05/11/2017 15:04	Archivo por lotes ...	1 KB
LICENSE.html	05/11/2017 15:04	Archivo HTML	13 KB
log4j.xml	05/11/2017 15:04	Documento XML	4 KB
mondrian.properties	05/11/2017 15:04	Archivo PROPERTI...	42 KB
PentahoSchemaWorkbench_OSS_License...	05/11/2017 14:13	Archivo HTML	1.871 KB
readme.txt	05/11/2017 15:04	Documento de tex...	1 KB
set-pentaho-env.bat	05/11/2017 15:04	Archivo por lotes ...	5 KB
set-pentaho-env.sh	05/11/2017 15:04	Shell Script	5 KB
workbench.bat	05/11/2017 15:04	Archivo por lotes ...	2 KB
workbench.command	05/11/2017 15:04	Archivo COMMA...	1 KB
workbench.sh	05/11/2017 15:04	Shell Script	2 KB

Figura 3. Archivo .bat para la ejecución del Mondrian Schema Workbench

3. Diseño del Cubo de Análisis OLAP

Una vez ejecutado el archivo "workbench.bat", se abrirá la ventana del diseñador de cubos donde podremos comenzar a diseñar nuestro cubo de análisis:

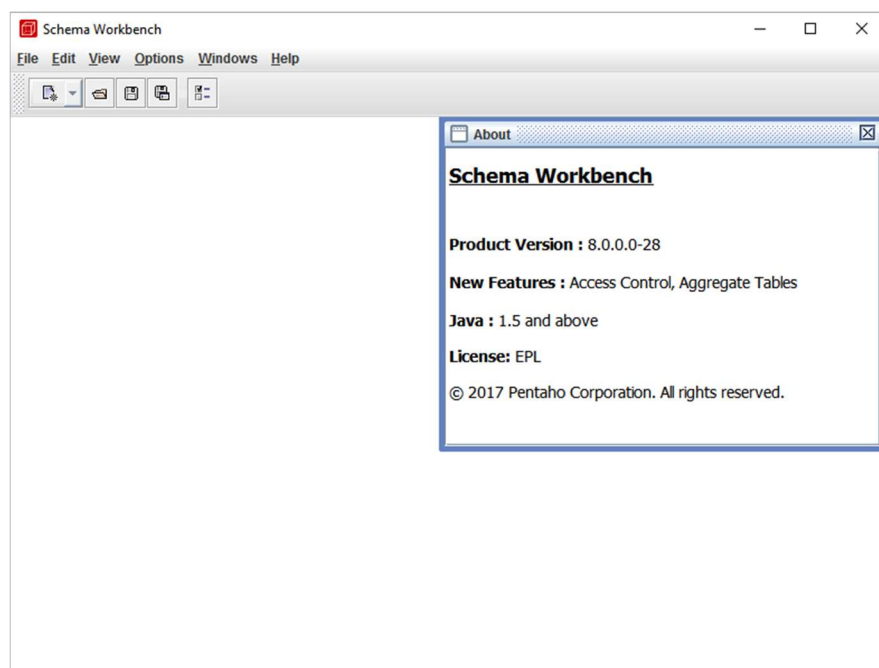


Figura 4. Pantalla inicial del Mondrian Schema Workbench

Antes de comenzar con el análisis, es recomendable tener una conexión con la base de datos que actuará de almacén, ya que así el propio programa nos hará una lista de las columnas y tablas incluidas en él. Para ello, seleccionamos "Options->Connection..." en el menú superior del programa. A continuación, aparecerá la ventana mostrada en la Figura 5, en la cual podremos configurar todos los parámetros de la conexión. Como utilizaremos MySQL en prácticas, seleccionaremos "MySQL" en "Connection Type". Una vez hemos seleccionado MySQL como tipo de conexión, le podemos dar el nombre que queramos y deberemos de indicar la máquina en la que se encuentra alojada la base de datos (si es en la misma máquina podemos poner "localhost" en lugar de la dirección IP). Como puerto dejaremos el que viene por defecto (3306) y como usuario podremos utilizar cualquiera que tenga permisos sobre la base de datos. En el caso de las prácticas podéis probar con el usuario root, password root.

Para comprobar que todo está correcto, podemos pulsa el botón "Test" en la parte inferior de la ventana antes de darle a aceptar.

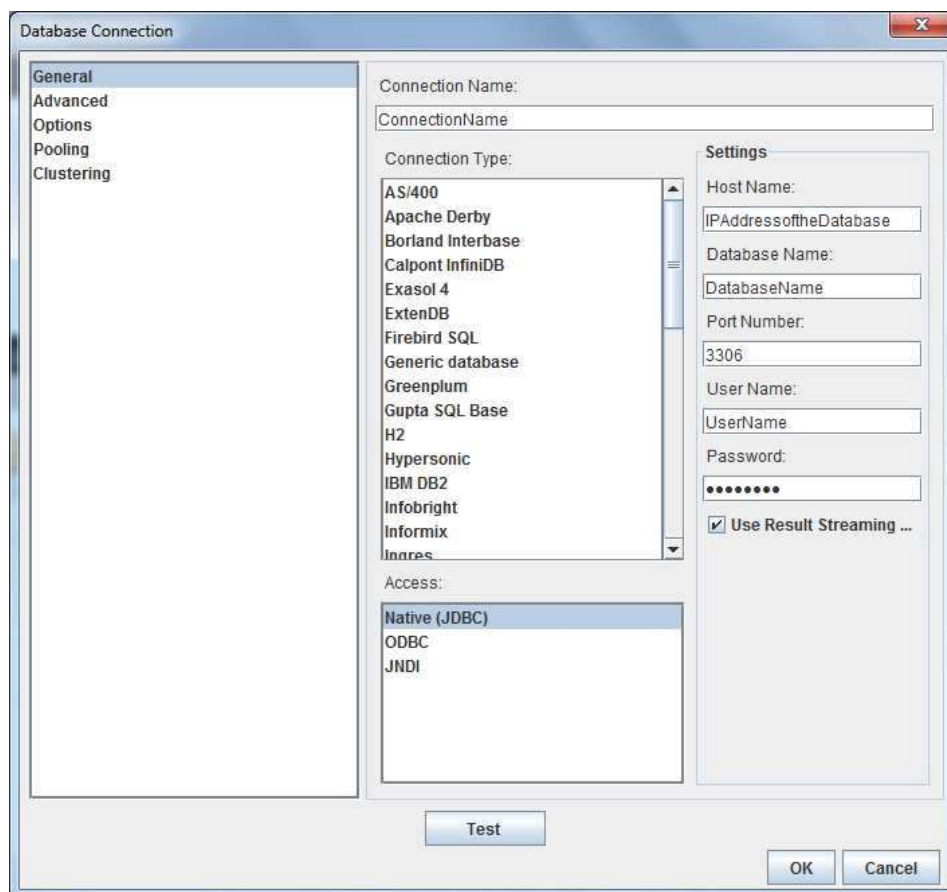


Figura 5. Añadir un cubo al esquema

Si bien para el diseño del esquema no es necesario tener una conexión previamente establecida, si que simplifica en gran medida la tarea de diseño.

Una vez hemos establecido la conexión, seleccionamos la opción File->New->Schema. Nos aparecerá una nueva ventana con la barra de herramientas y un esquema en blanco, tal y como se ve en la Figura 6. Si pinchamos en el esquema podremos ver en la parte

derecha sus propiedades, tales como su nombre o su descripción. En la parte superior tendremos la barra de herramientas, que dispone de varios botones para añadir elementos al esquema. Podemos hacer uso tanto de la barra de herramientas como del clic derecho sobre el esquema para ir añadiendo elementos. Para comenzar, empezaremos añadiendo un cubo al esquema. Este cubo representará el hecho o actividad que queremos analizar:

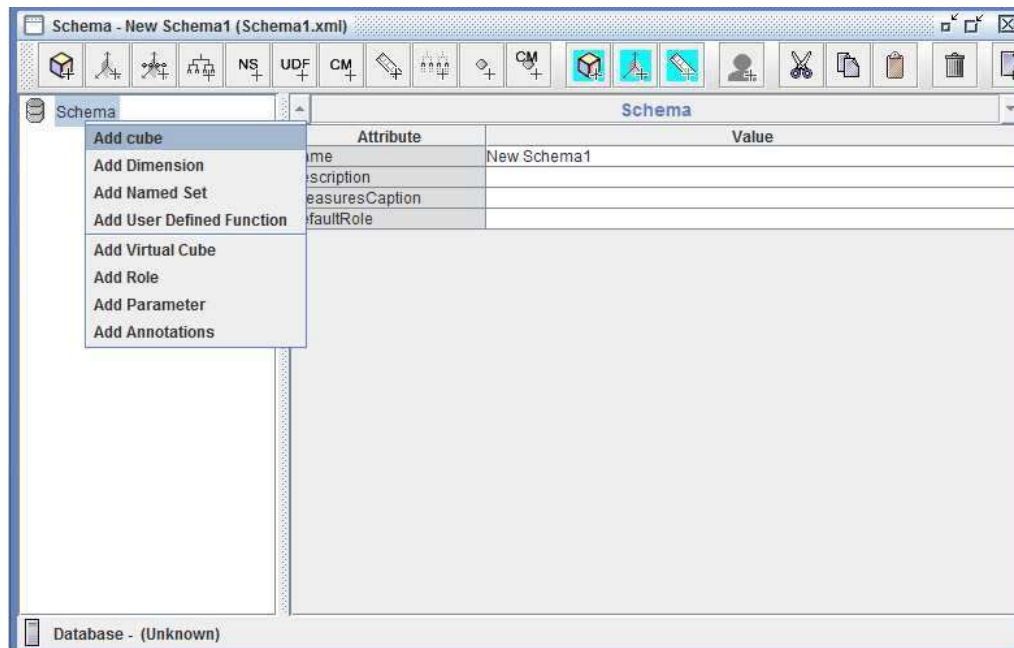


Figura 6. Añadir un cubo al esquema

Tras añadir un cubo podremos asignarle un nombre e indicar (si hemos configurado la conexión adecuadamente) la tabla de hechos con la que se encuentra asociado. Dentro de este cubo podremos añadir las medidas que queremos analizar o las dimensiones.

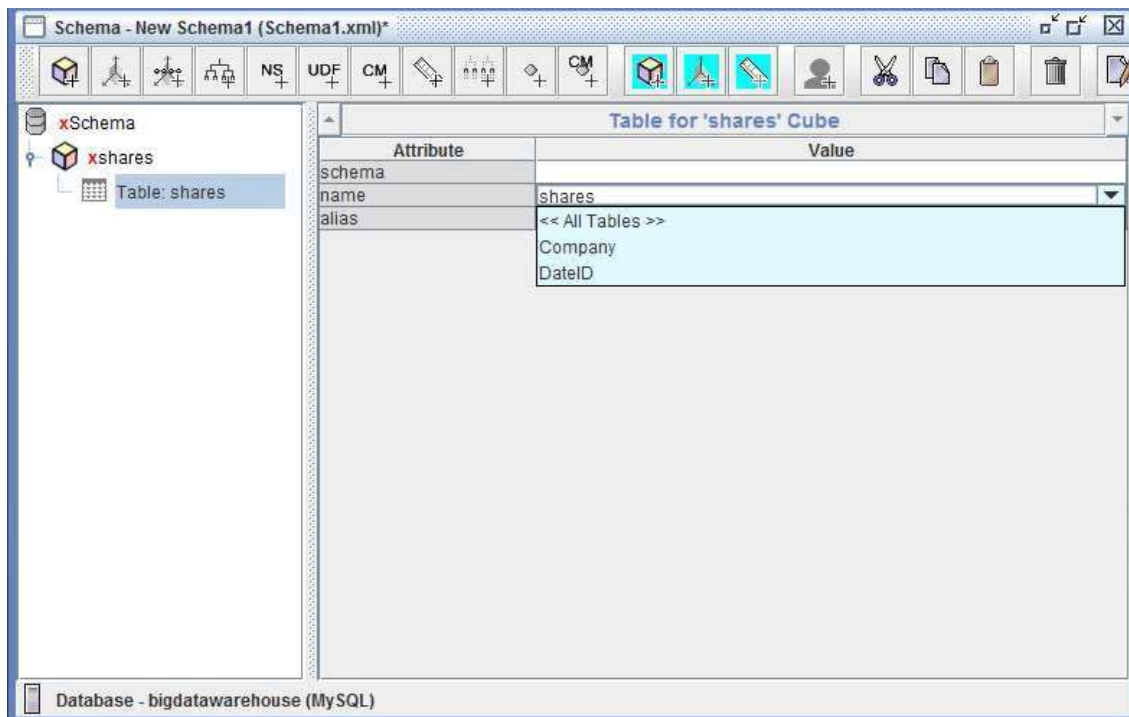


Figura 7. Seleccionar la tabla de hechos para el cubo

Sin embargo, ¿qué ocurre si queremos compartir dimensiones entre varios cubos? Sería necesario definir dichas dimensiones en cada cubo. Para evitar dicho trabajo lo que haremos será añadir las dimensiones directamente al esquema, y posteriormente enlazarlas con el cubo. Para ello, hacemos clic derecho sobre el esquema y seleccionamos añadir dimensión:

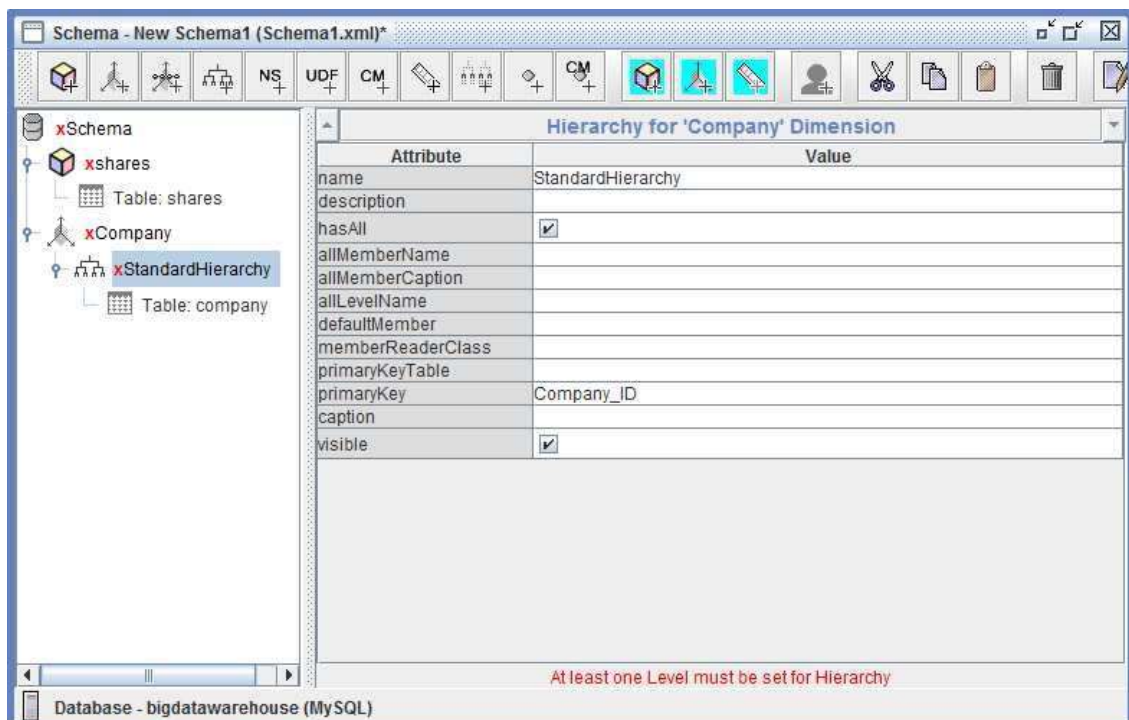


Figura 8. Creación de una dimensión

Una vez hemos añadido una dimensión deberemos especificar los posibles caminos de

agregación (jerarquías) que se utilizarán en el cubo. Para ello, primero deberemos de añadir la tabla donde se guardará la jerarquía. Basta con hacer clic derecho en la jerarquía y seleccionar "Add Table" al igual que en el caso del cubo. Tras añadir la tabla, podremos volver a la jerarquía y asignarle el valor correspondientes a "primaryKey". Como último paso, deberemos de añadir los niveles a la jerarquía. Hay que tener en cuenta que los niveles en un esquema de Mondrian van del más general al más específico. Es decir, si tenemos un camino que consiste en ciudad>región>país, entonces, en el Schema Workbench deberemos de añadir los niveles en el orden de país, después región y por último ciudad.

Al especificar un nivel deberemos indicar su nombre, la columna que lo identifica, el tipo de dicha columna, y, opcionalmente una columna para darle nombre a las instancias, tal y como se ve en la Figura 9. Por ejemplo, podemos querer utilizar los nombres de las ciudades de cara al usuario pero, como identificador, utilizar su código, ya que Valencia se encuentra tanto en España como en Venezuela y no queremos mezclar sus datos.

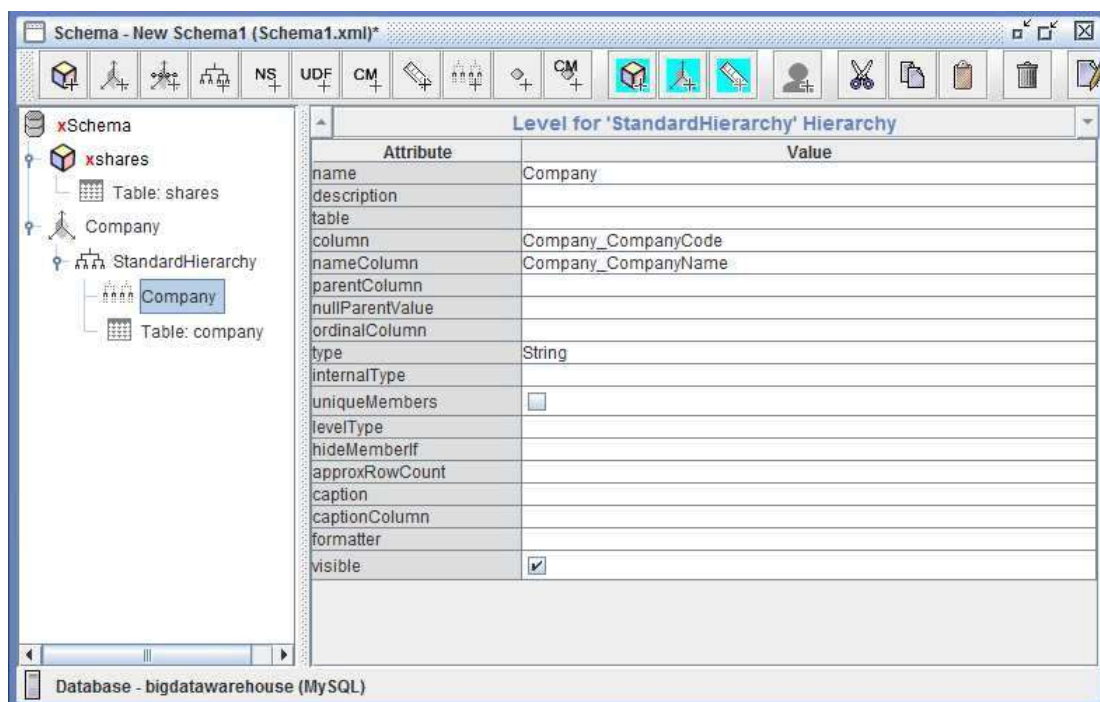


Figura 9. Añadir niveles a la jerarquía

Si queremos añadir propiedades al nivel que sirvan al analista para comprobar cierta información pero que no afectan a las jerarquías, basta con que hagamos clic derecho sobre el nivel y seleccionemos "Add Property". Aparecerá una nueva propiedad dentro del nivel a la que podremos asignarle un nombre y la columna de la cual obtener el valor. A pesar de que en ciertas ocasiones no aparezcan las columnas de la tabla para las propiedades, basta con escribir el nombre de la columna manualmente.

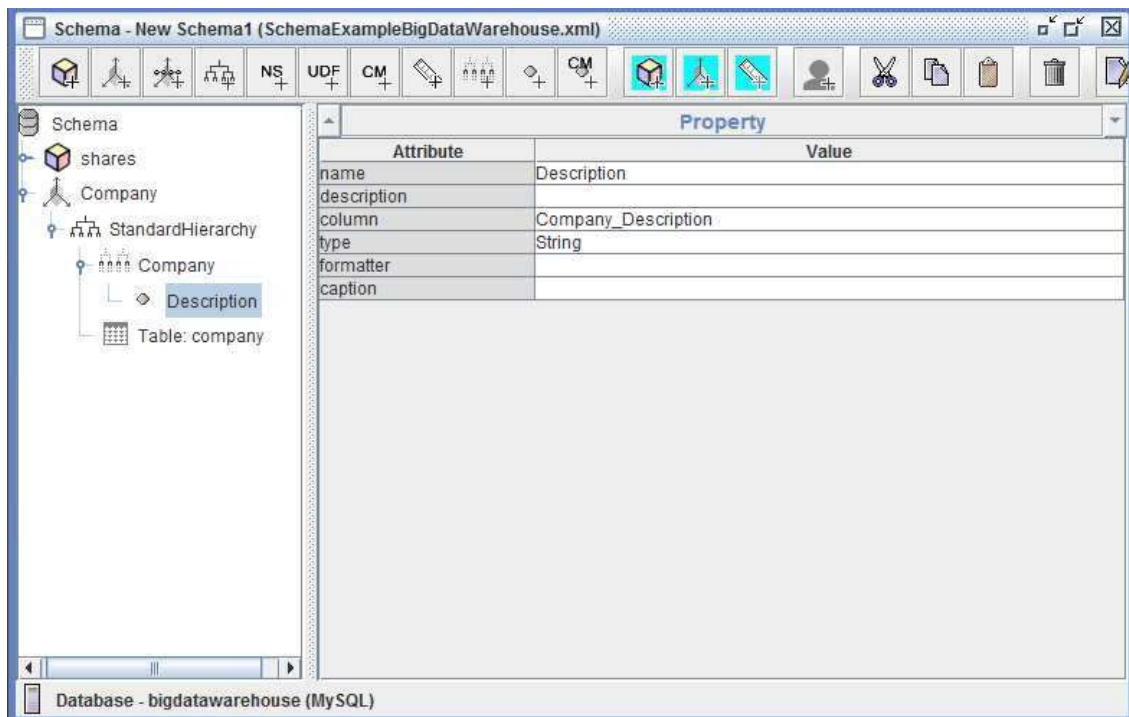


Figura 10. Añadir propiedades a un nivel

Una vez estemos satisfechos con las jerarquías que tenemos, volvemos al cubo y hacemos clic derecho. Esta vez elegimos la opción "Add Dimension Usage", que nos permitirá utilizar dimensiones definidas en el esquema aunque estén fuera del cubo. Tras hacerlo, nos aparecerán una lista de atributos para darle un nombre a la relación entre la dimensión y el cubo (roles), indicar qué clave ajena define dicha relación, y cuál es la fuente que se utiliza. Opcionalmente, podemos especificar el nivel de la dimensión con el que queremos enlazar. Por ejemplo, para nuestro ejemplo anterior, podríamos enlazar desde el nivel Región en lugar de Ciudad.

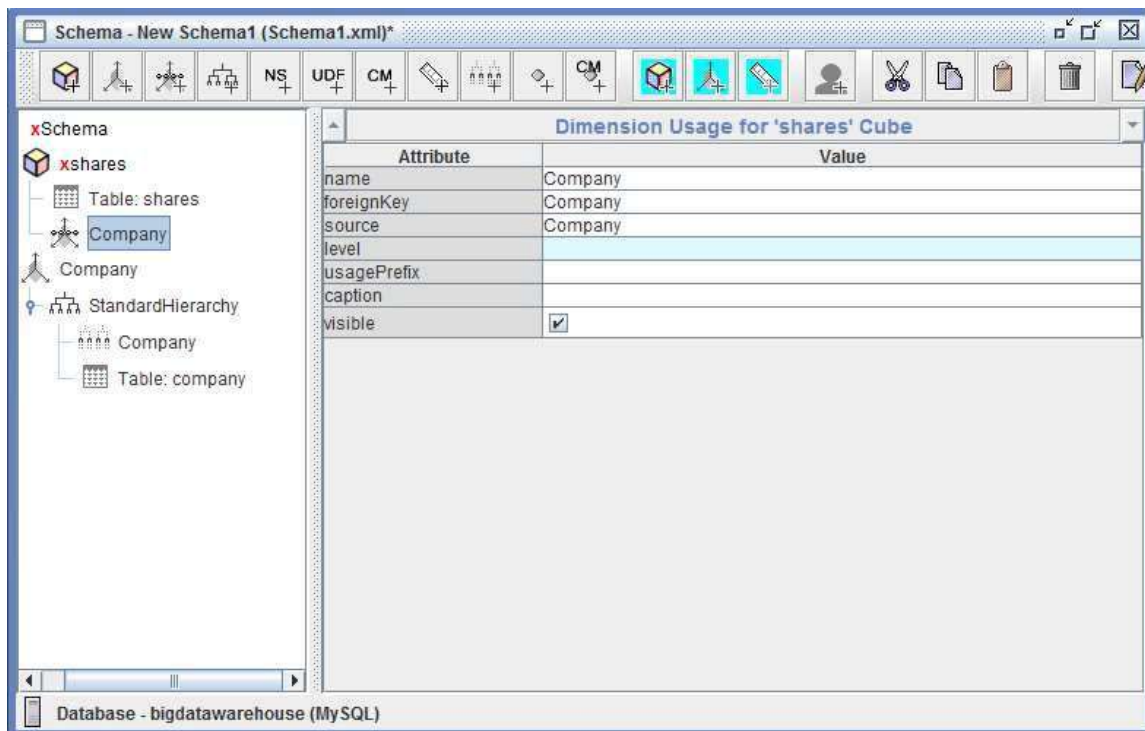


Figura 11. Añadir relaciones entre el cubo y las dimensiones

Finalmente, nos queda añadir las medidas al cubo. Para ello, basta con hacer clic derecho en el cubo y seleccionar "Add Measure". Para cada medida podremos configurar el nombre que verá el analista, la función de agregación para calcular su valor, la columna correspondiente de la base de datos, y su tipo, que generalmente será Numeric.

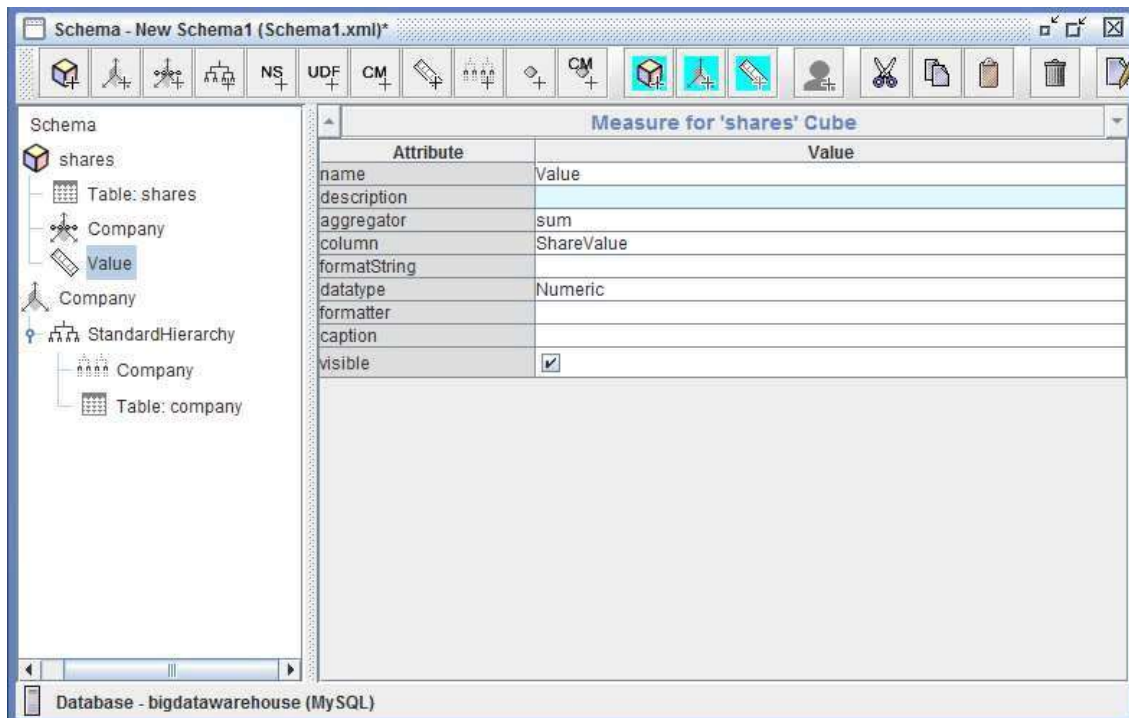


Figura 12. Añadir medidas al cubo

Con esto habremos terminado de crear nuestro cubo y sólo nos quedará publicarlo en Pentaho.

4. Publicación del Esquema en el servidor Pentaho Server

Los esquemas de Mondrian han de ser publicados en Pentaho BA Server para poder ser utilizados. Existen distintas formas de llevar a cabo la publicación del fichero XML que contiene el diseño del cubo que acabamos de realizar. En este manual vamos a ver la que hemos considerado más sencilla para la versión 7 de Pentaho Server.

En primer lugar, hemos de guardar el cubo que hemos estado diseñando siguiendo los puntos anteriores de este manual. Para ello, en el menú superior seleccionamos la opción File→Save. Seleccionamos una ubicación para guardar el archivo .xml resultante y le cambiamos el nombre si lo consideramos oportuno (Schema1 es el nombre por defecto).

Tras esto, hemos de arrancar el servidor de Inteligencia de Negocio, Pentaho BA Server 5 (como ya se ha explicado en el otro manual proporcionado sobre dicha aplicación). Una vez nos hayamos identificado y accedido a la consola principal de la aplicación web, hemos hacer clic en el menú superior en File→Manage Data Sources. Nos aparecerá entonces la lista de conexiones configuradas y, por tanto, disponibles en nuestro servidor Pentaho. En la Figura 13 se muestra la lista de conexiones que nos aparece cuando aún no hemos configurado ninguna conexión además de las que incluye Pentaho por defecto. Hay que tener en cuenta que la conexión que definimos antes en Mondrian Schema Workbench no aparecerá aquí, pues el diseñador de cubos de Mondrian es una aplicación independiente de Pentaho BA Server.

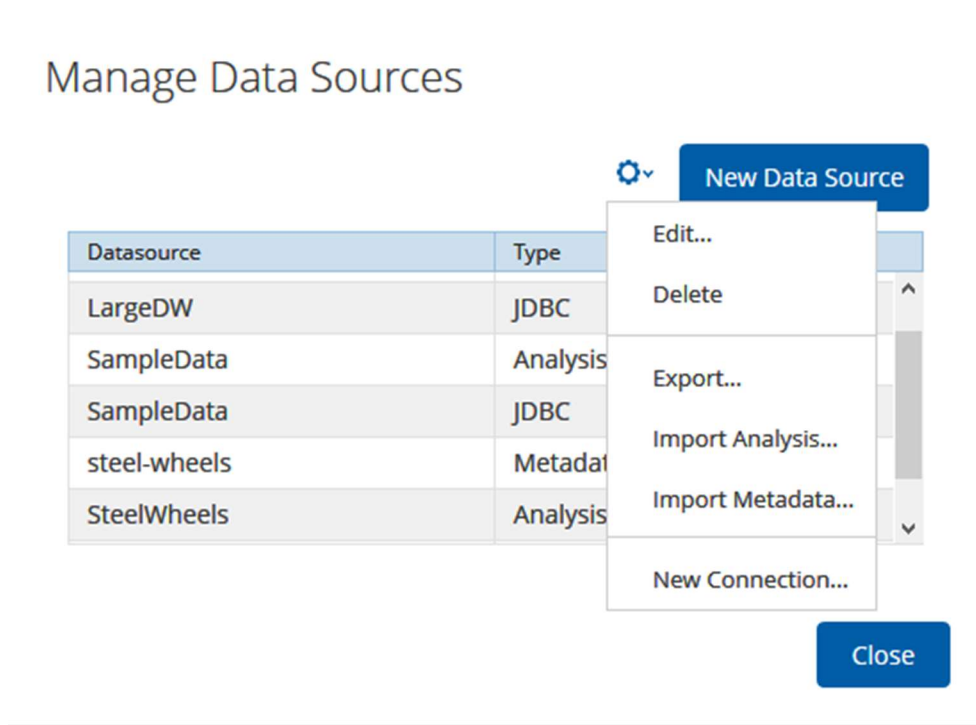


Figura 13. Ventana de conexiones disponibles en Pentaho Server 7

El siguiente paso es configurar y añadir una nueva conexión JDBC con la base de datos relacional donde hemos implementado físicamente el Almacén de Datos. Para ello, hemos de hacer clic en el símbolo + y luego en JDBC. A continuación nos aparece una ventana similar a la de la Figura 5, mostrada anteriormente cuando configuramos la conexión para la definición del cubo en Mondrian Schema Workbench. La forma de configurar la conexión es exactamente la misma que en ese caso. En la figura 14, se muestra un ejemplo para una conexión con una base de datos de nombre almacenegg y creada en una instalación local de MySQL.

The screenshot shows the 'Database Connection' window. On the left, a sidebar lists 'General', 'Advanced', 'Options', and 'Pooling'. The 'General' tab is selected. The main area is divided into sections: 'Connection Name' with the text 'MiConexionDB_JDBC'; 'Database Type' with a list box containing 'Generic database', 'H2', 'Hypersonic', 'MonetDB', 'MySQL' (highlighted), 'Pentaho Data Services', 'PostgreSQL', and 'SparkSQL'; 'Access' with a list box containing 'Native (JDBC)' (highlighted), 'ODBC', and 'JNDI'; and 'Settings' which includes 'Host Name' (localhost), 'Database Name' (almacenegg), 'Port Number' (3306), 'User Name' (root), and 'Password' (masked with dots). A blue 'Test' button is centered at the bottom, and 'OK' and 'Cancel' buttons are at the bottom right.

Figura 14. Configuración de la conexión con la BD Relacional

Una vez hemos configurado la conexión y le hemos dado un nombre (ej. MiConexionDB_JDBC) hacemos clic en el botón Test para probarla. Si todo funciona correctamente el aviso que nos devuelve nos la confirmará con un OK. Cerramos el aviso y pulsamos en el botón OK de la ventana de configuración para guardar la nueva conexión definida. Ahora ya debe aparecernos en la lista de conexiones disponibles, como se muestra en la figura 15.

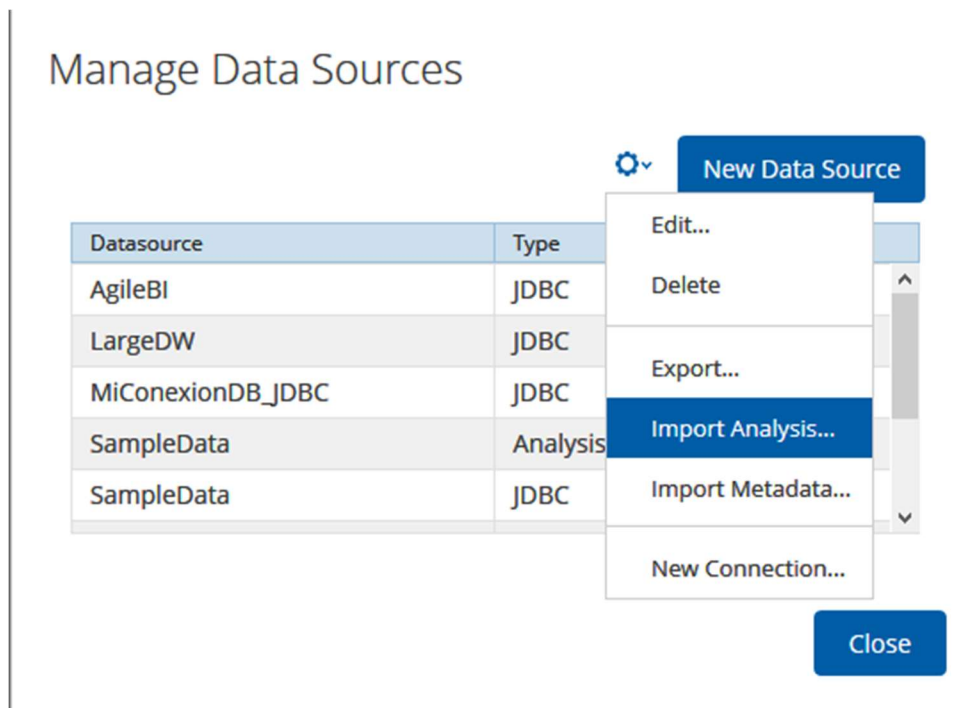


Figura 15. Lista de conexiones tras crear la conexión JDBC de nombre MiConexionDB_JDBC

Por último, hemos de publicar el cubo, es decir, el archivo xml que hemos guardado con Mondrian Schema Workbench. Para ello, en la misma ventana de gestión de las conexiones disponibles, hemos de añadir una nueva conexión de Análisis “Analysis” pulsando el botón +. En la Figura 15 antes mostrada puede verse dicha opción.

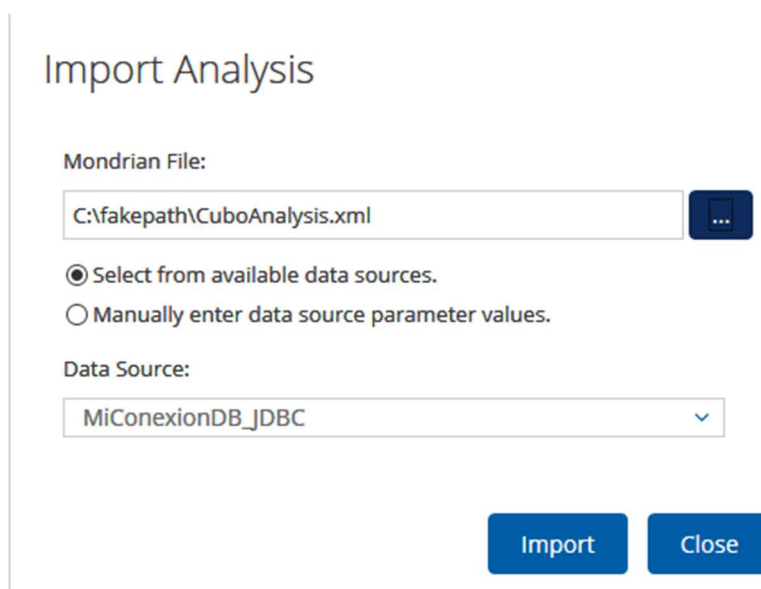


Figura 16. Publicación del cubo definido con Mondrian Schema Workbench

Ahora nos aparece una ventana como la de la Figura 16. Primero hemos de indicar la ruta del archivo XML de Mondrian que queremos publicar, en la opción Mondrian File. A continuación, marcamos la opción “Select from available data sources” si no está ya seleccionada por defecto. Por último, en la lista de Data Sources, seleccionamos la conexión con la BD relacional que hemos definido anteriormente, MiConexionDB_JDBC en el caso del ejemplo, y pulsamos el botón Import.

Si todo ha ido bien, aparecerá una nueva conexión en la lista de conexiones de datos, de tipo Analysis, cuyo nombre será el nombre que le hayamos dado al esquema con Mondrian Schema Workbench. Ahora ya podemos hacer uso de nuestro esquema y del cubo o cubos definidos en él como fuente de nuevas vistas de análisis² (también comúnmente denominadas Tablas OLAP, Vistas Multidimensionales o, en el caso de Pentaho, JPivotView).

² Siempre y cuando hayamos cargado antes el almacén con datos. Si no hay datos en el almacén, es posible definir una nueva JPivotView sobre ese esquema y cubo, pero no nos aparecerán datos al crearla.