

IN830: Herramientas Opensource J2EE de DW

Pentaho: Mondrian/Kettle

Pablo Román

El proyecto Pentaho (<http://www.pentaho.org>)

- Basado en una variedad de proyectos opensource, unidos bajo una plataforma j2ee (servidor jboss).
 - Mondrian: Se conecta a una BD relacional y presenta un modelo multidimensional sobre esos datos. Es consultado a través del lenguaje MDX.
 - Kettle: Migra datos de una BD a otra (incluso archivos texto) en forma estructurada (con un diseño). ETL (Extraction, Transport, Transformation, Loading)
 - JFreeReport (Pentaho Report): Estructuración (XML) de reportes para presentar resultados en producción.

J2EE : Suite de librerías de Java para aplicaciones empresariales.

■ Librerías:

- Servlet: Clases java que modelan aplicaciones web genéricas (Abstraccion de Server).
- JSP: Script que al momento de ser solicitados generan un Servlet, ie JSP=Servlet.
- JDBC: Conector genérico a cualquier base de datos. Solo requiere disponer del "Driver" correcto.

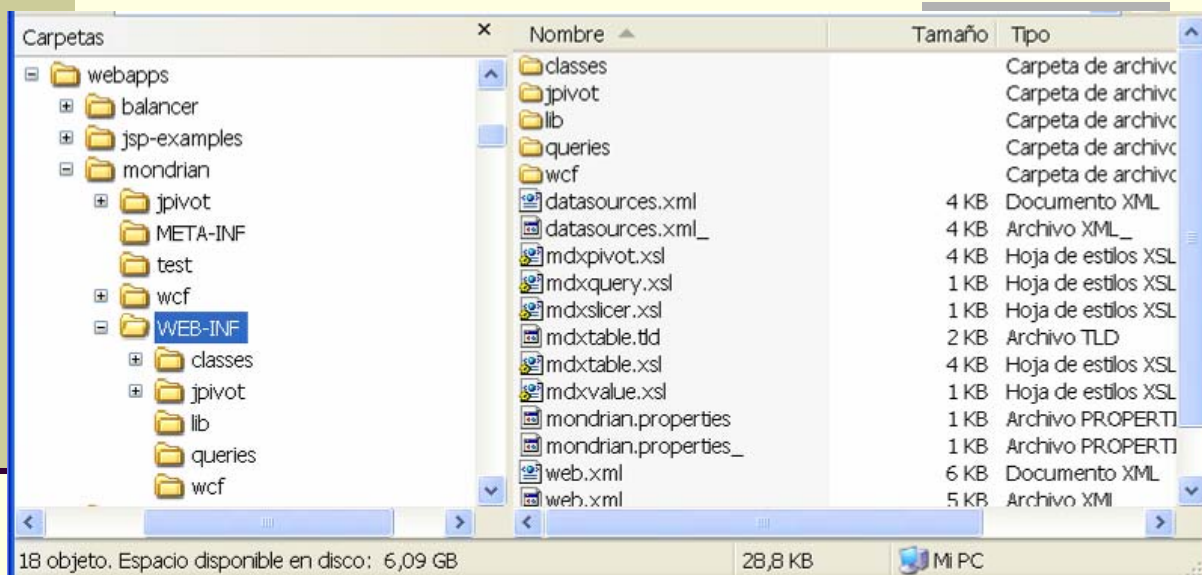
■ Servidores de aplicaciones:

- Apache Tomcat: Sirve aplicaciones basadas en Servlet.
- Jboss: Sirve aplicaciones basadas en Servlet (tiene un tomcat embebido), EJB, toda la suite j2ee.

Arquitectura de la capa web de J2EE (o Java EE)

- El modelo Servlet (o JSP) : Reciben Peticiones desde los browser de los clientes, las procesan y envían de vuelta una Respuesta. Son objetos instancias de la clase Servlet.
- Tiene una especificación OO para su comportamiento y implementación.
 - Estructura de directorios definida.
 - Empaquetamiento de estos directorios en archivo de extensión **war** (Web ARchive, solo zipeado!)

Layout de un directorio de aplicación web

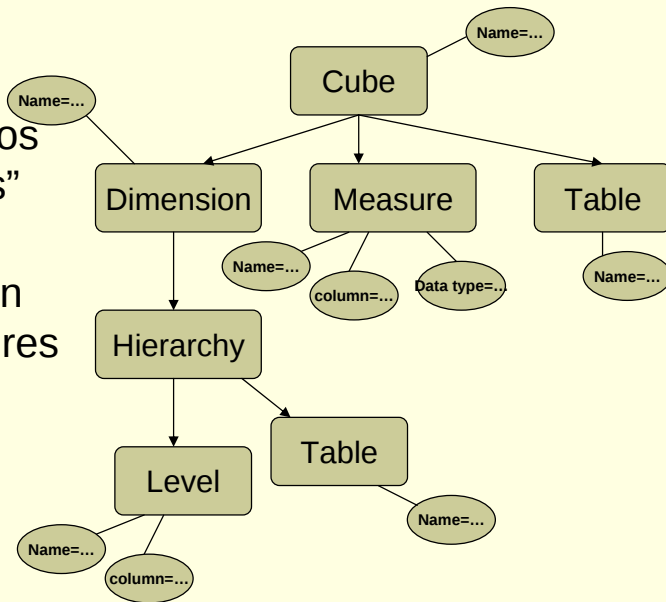


Layout de un directorio de aplicación web

- En el caso anterior:
 - mondrian: es el directorio en el que se descomprimió el archivo war.
 - WEB-INF: contiene configuración y archivos de librerías y programas java. Contiene el archivo “web.xml” que tiene la configuración del sistema en XML y por lo general hay otros mas.
 - WEB-INF/clases: contiene archivos de clases Servlet.
 - WEB-INF/lib: contiene archivos de extensión jar (Java ARchive) que agrupan paquetes de clases Java.

XML (Extensible Markup Language)

- Almacena datos jerárquicos.
- Se logra anidando los datos entre “marcas” (tags = <algo ...>)
- Cada marca tiene un nombre y otros valores asociados.



XML (Extensible Markup Language)

```
<Cube name="Store" cache="true" enabled="true">
  <Table name="store">
  </Table>
  <Dimension name="Store Type">
    <Hierarchy hasAll="true">
      <Level name="Store Type" column="store_type" type="String"
uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <DimensionUsage source="Store" name="Store">
  </DimensionUsage>
  <Dimension name="Has coffee bar">
    <Hierarchy hasAll="true">
      <Level name="Has coffee bar" column="coffee_bar" type="Boolean"
uniqueMembers="true" levelType="Regular" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <Measure name="Store Sqft" column="store_sqft" formatString="#,###"
aggregator="sum">
  </Measure>
  <Measure name="Grocery Sqft" column="grocery_sqft" formatString="#,###"
aggregator="sum">
  </Measure>
</Cube>
```

La maldita configuración de J2ee

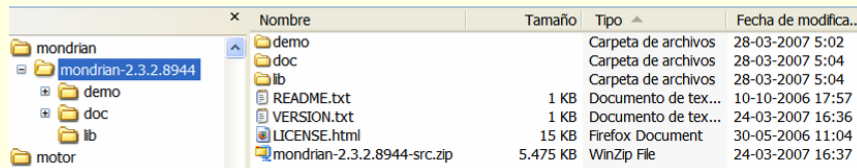
- Muchos archivos de configuración que son fuertemente dependientes (muy acoplados) con los programas. Problemas se resuelven con herramientas de generación de código (Xdoclet o Schema Workbench).
- Cambios en parte de los códigos o configuraciones XML, no se reflejan inmediatamente -> es necesario consolidar nuevamente las configuraciones.
- Es EL PROBLEMA que actualmente que dificulta los desarrollos J2EE.

El caso Mondrian: Configuración

- Configuración de base de datos MySQL la vimos la clase pasada.
- Archivos importantes:
 - WEB-INF/web.xml -> conf de los parametros en XML
 - WEB-INF/datasources.xml -> fuente de datos
 - WEB-INF/mondrian.properties-> conf de mondrian
 - WEB-INF/queries/FoodMart.xml -> descripcion XML de la base de datos
 - WEB-INF/queries/ ourheir.jsp, mondrian.jsp, colors.jsp, arrows.jsp -> programas que tienen en duro datos de la base.

Instalación Mondrian con el demo

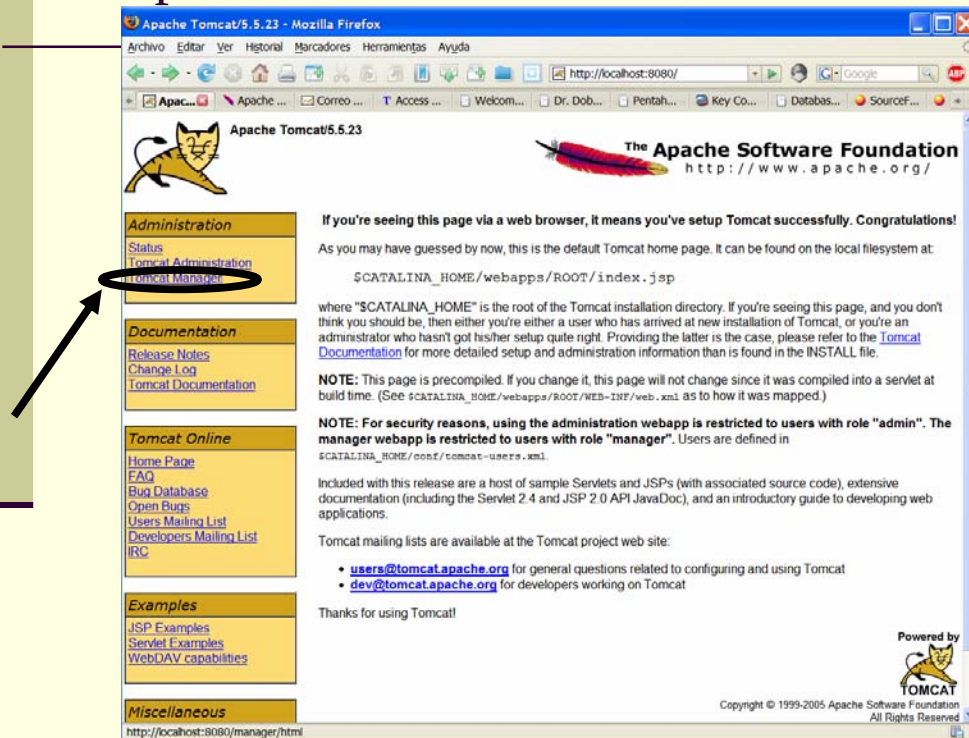
- <http://sourceforge.net/projects/mondrian>
- Descomprima el archivo zip de mondrian
- En el directorio lib del mondrian encontrara un archivo llamado “mondrian.war”



Nombre	Tamaño	Tipo	Fecha de modifica...
demo		Carpeta de archivos	28-03-2007 5:02
doc		Carpeta de archivos	28-03-2007 5:04
lib		Carpeta de archivos	28-03-2007 5:04
README.txt	1 KB	Documento de tex...	10-10-2006 17:57
VERSION.txt	1 KB	Documento de tex...	24-03-2007 16:36
LICENSE.html	15 KB	Firefox Document	30-05-2006 11:04
mondrian-2.3.2.8944-src.zip	5.475 KB	WinZip File	24-03-2007 16:37

Vaya al administrador del tomcat

http://localhost:8080



Subir el “mondrian.war”

The Apache Software Foundation
http://www.apache.org/

Gestor de Aplicaciones Web de Tomcat

Mensaje: OK

Gestor

[Listar Aplicaciones](#) [Ayuda HTML de Gestor](#) [Ayuda de Gestor](#) [Estado de Servidor](#)

Aplicaciones

Trayectoria	Nombre a Mostrar	Ejecutándose	Sesiones	Comandos
/	Welcome to Tomcat	true	0	Arrancar Parar Recargar Replegar
/balancer	Tomcat Simple Load Balancer Example App	true	0	Arrancar Parar Recargar Replegar
/host-manager	Tomcat Manager Application	true	0	Arrancar Parar Recargar Replegar
/jsp-examples	JSP 2.0 Examples	true	0	Arrancar Parar Recargar Replegar
/manager	Tomcat Manager Application	true	0	Arrancar Parar Recargar Replegar
/servlets-examples	Servlet 2.4 Examples	true	0	Arrancar Parar Recargar Replegar
/tomcat-docs	Tomcat Documentation	true	0	Arrancar Parar Recargar Replegar
/webdav	Webdav Content Management	true	0	Arrancar Parar Recargar Replegar

Desplegar

Desplegar directorio o archivo WAR localizado en servidor

Trayectoria de Contexto (opcional):

URL de archivo de Configuración XML:

URL de WAR o Directorio:

Archivo WAR a desplegar

Seleccione archivo WAR a cargar:

Terminado

Mondrian quedo instalado como aplicación.

The Apache Software Foundation
http://www.apache.org/

Gestor de Aplicaciones Web de Tomcat

Mensaje: OK

Gestor

Listar Aplicaciones Ayuda HTML de Gestor Ayuda de Gestor Estado de Servidor

Trayectoria	Nombre a Mostrar	Ejecutándose	Sesiones	Comandos
/	Welcome to Tomcat	true	0	Arrancar Parar Recargar Resplegar
/balancer	Tomcat Simple Load Balancer Example App	true	0	Arrancar Parar Recargar Resplegar
/host-manager	Tomcat Manager Application	true	0	Arrancar Parar Recargar Resplegar
/jsp-examples	JSP 2.0 Examples	true	0	Arrancar Parar Recargar Resplegar
/manager	Tomcat Manager Application	true	0	Arrancar Parar Recargar Resplegar
/mondrian	Mondrian	true	0	Arrancar Parar Recargar Resplegar
/servlet-examples	Servlet 2.4 Examples	true	0	Arrancar Parar Recargar Resplegar
/tomcat-docs	Tomcat Documentation	true	0	Arrancar Parar Recargar Resplegar
/webdav	Webdav Content Management	true	0	Arrancar Parar Recargar Resplegar

Desplegar

Desplegar directorio o archivo WAR localizado en servidor

Trayectoria de Contexto (opcional):

URL de archivo de Configuración XML:

URL de WAR o Directorio:

Archivo WAR a desplegar

Seleccione archivo WAR a cargar

Terminado

Funcionamiento de Mondrian

Mondrian/JPivot Test Page - Mozilla Firefox

Archivo Editar Ver Historial Marcadores Herramientas Ayuda

Test Query uses Mondrian OLAP

Promotion Media	Product	Unit Sales	Store Cost	Store Sales
All Media	All Products	266.773	225.627,23	565.238,13
Bulk Mail	All Products	4.320	3.740,95	9.349,07
Cash Register Handout	All Products	6.697	5.715,67	14.321,33
Daily Paper	All Products	7.738	6.559,23	16.479,81
Daily Paper, Radio	All Products	6.891	5.668,77	14.169,42
Daily Paper, Radio, TV	All Products	9.513	8.055,22	20.173,97
In-Store Coupon	All Products	3.798	3.263,11	8.162,46
No Media	All Products	195.448	165.214,85	414.026,92
Product Attachment	All Products	7.544	6.306,24	15.898,25
Radio	All Products	2.454	2.087,51	5.213,61
Street Handout	All Products	5.753	4.856,54	12.192,90
Sunday Paper	All Products	4.339	3.673,86	9.092,89
Sunday Paper, Radio	All Products	5.945	5.027,31	12.551,96
Sunday Paper, Radio, TV	All Products	2.726	2.341,58	5.819,33
TV	All Products	3.607	3.116,40	7.786,21

Slicer: [Year=1997]

[back to index](#)

Encontrar: Siguiente Anterior Resaltar todo Coincidencia de mayúsculas

Terminado

¿Qué es mondrian?

- Utiliza Jpivot: Quien finalmente efectúa todo el trabajo grafico de dibujar las tablas y gráficos asociados a las consultas OLAP.
- Mondrian ... Es solo “el conector” que toma la query MDX y la transforma al lenguaje SQL común y corriente, hace la consulta y retorna datos.
- ¿Cómo? : Mondrian tiene conocimiento de la base de datos a travez de una descripcion XML de ella, que es : WEB-INF/queries/FoodMart.xml

Instalación de un ... cubo nuevo

- WEB-INF/queries/FoodMart.xml
- Este archivo es crucial para el propósito principal de Mondrian.
- Es la definición del cubo que se esta usando, basándose en la BD relacional.
- Es un archivo XML
- Es mejor generar este archivo, para ello se usará el software “Schema Workbench”.
- Este software se conecta a la base de datos identificando campos y tablas para el diseñador.

Instalación de un cubo nuevo: Schema Workbench

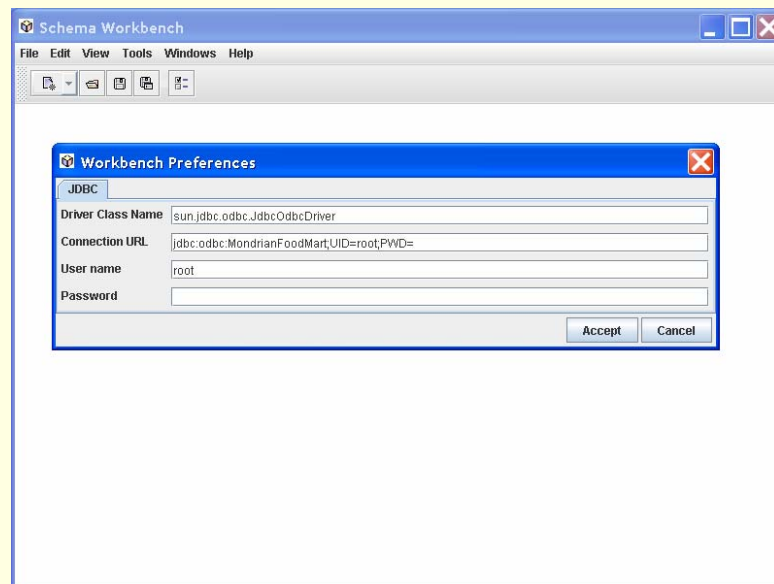
- WEB-INF/queries/FoodMart.xml
- Este archivo es crucial para el propósito principal de Mondrian.
- Es la definición del cubo que se esta usando, basándose en la BD relacional.
- Es un archivo XML
- Es mejor generar este archivo, para ello se usará el software “Schema Workbench”.
- Este software se conecta a la base de datos identificando campos y tablas para el diseñador.

Schema Workbench

- Apareció el día 11/05/2007 y es mejor que el cube designer!
- Schema: Es la descripción de una BD (metadatos).
- Se baja del sitio donde esta mondrian.
<http://sourceforge.net/projects/mondrian>
- Se descomprime y ejecuta el archivo “workbench.bat”
- Se debe configurar la conexión a la base de datos en el menu “Tools->preferences”, los parámetros a ingresar son:
 - Driver Class Name: sun.jdbc.odbc.JdbcOdbcDriver
 - Connection URL: jdbc:odbc:DM1;UID=root;PWD=
 - Username: root

Schema Workbench

■ Conexión a la BD



Schema Workbench

- Nos acabamos de conectar a la base de datos Foodmart.
- Podemos abrir el Schema FoodMart.xml que viene con mondrian para ver como se ve. (Base de datos y Schema abierto deben coincidir)
- Observemos la estructura haciendo click en la vista XML que es mas completa.

Schema Workbench

Schema Workbench

File Edit View Tools Windows Help

Schema - foodmart (foodmart.mtd)

Schema

- Sales
- Warehouse
- Store
- HR
- Sales Ragged
- Sales 2
- Table sales_fact_1997
- Time
- Product
- Gender
- Sales Count
- Unit Sales
- Store Sales
- Store Cost
- Customer Count
- CM Profit
- CM Profit last Period
- Store
- Store Size in SQFT
- Store Type
- Time
- Product
- Warehouse
- Warehouse and Sales
- Custom manager
- No HR Cube

Cube

```
<Cube name="Sales 2" cache="true" enabled="true">
  <Table name="sales_fact_1997">
    <DimensionUsage source="Time" name="Time" foreignKey="time_id">
      <DimensionUsage>
        <DimensionUsage source="Product" name="Product" foreignKey="product_id">
          <DimensionUsage>
            <Dimension foreignKey="customer_id" name="Gender">
              <Hierarchy hasAll="true" allMemberName="All Gender" primaryKey="customer_id">
                <Table name="customer">
                  <Level name="Gender" column="gender" type="String" uniqueMembers="true" levelType="Regular" hideMembers="Never">
                    <Level>
                      <Hierarchy>
                        <Dimension>
                          <Measure name="Sales Count" column="product_id" formatString="###" aggregation="count">
                            <CalculateMemberProperty name="MEMBER_ORDINAL" value="1">
                              <CalculateMemberProperty>
                                <Measure>
                                  <Measure name="Unit Sales" column="unit_sales" formatString="Standard" aggregation="sum">
                                    <CalculateMemberProperty name="MEMBER_ORDINAL" value="2">
                                      <CalculateMemberProperty>
                                        <Measure>
                                          <Measure name="Store Sales" column="store_sales" formatString="###.00" aggregation="sum">
                                            <CalculateMemberProperty name="MEMBER_ORDINAL" value="3">
                                              <CalculateMemberProperty>
                                                <Measure>
                                                  <Measure name="Store Cost" column="store_cost" formatString="###.00" aggregation="sum">
                                                    <CalculateMemberProperty name="MEMBER_ORDINAL" value="4">
                                                      <CalculateMemberProperty>
                                                        <Measure>
                                                          <Measure name="Customer Count" column="customer_id" formatString="###" aggregation="distinct count">
                                                            <CalculateMemberProperty name="MEMBER_ORDINAL" value="5">
                                                              <CalculateMemberProperty>
                                                                <Measure>
                                                                  <CalculateMember name="Profit dimension" Measures">
                                                                    <Formula>
                                                                      <Formula> [Store Sales] - [Store Cost] </Formula>
                                                                      <CalculateMemberProperty name="FORMAT_STRING" value="###.00">
                                                                        <CalculateMemberProperty>
                                                                          <CalculateMemberProperty name="MEMBER_ORDINAL" value="6">
                                                                            <CalculateMemberProperty>
                                                                              <CalculateMember>
                                                                                <CalculateMember name="Profit last Period" formula="COALESCE(EMPTY([Measures].[Profit].[Time].PREVEMBER), Measures.[Profit] dimension"Measures" visible="false">
                                                                                  <CalculateMemberProperty name="MEMBER_ORDINAL" value="7">
                                                                                    <CalculateMemberProperty>
                                                                                      <CalculateMember>
                                                                                        <Cube>

```

Database - foodmart (MySQL)

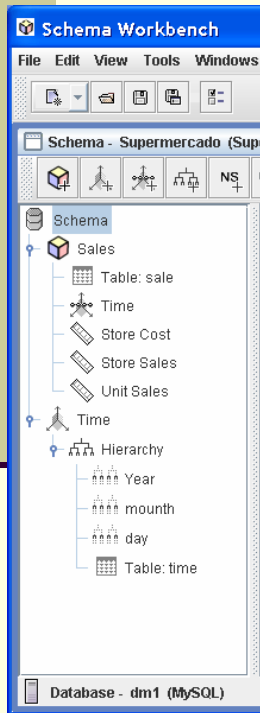
Schema Workbench

- El esquema anterior nos servirá de guía para generar nuevos esquemas.
- Hay varios cubos en un esquema.
- Se definen dimensiones compartidas por los cubos.
- Los Esquemas Snowflake hay que resolverlos a mano en el XML: La herramienta aun no dispone de la posibilidad de declarar el JOIN.
- Pueden testearse consultas MDX efectuando new MDX query. Para ello puede hacerse la prueba con la query que se encuentra en el archivo: WEB-INF/queries/mondrian.jsp

Schema Workbench: Creando un nuevo cubo simple.

- Cargar la base de datos DM1 ubicada en :
<http://adi.ing.uchile.cl/~proman/DM1.sql.zip>
- Cargar en MySQL (es un modelo snowflake centrado en la tabla "sale")
- Crear link ODBC DM1
- Configurar workbench apuntando a DM1
- Crear un nuevo Schema: Supermercado
- Agregar 1 cubo: Sales
- Agregar 1 dimension "Time" fuera del cubo para que sea compartida por los cubos futuros.
- Agregar medida (measure) que son store_cost, store_sales, unit_sales.
- Llame al Schema: "Supermercado"
- Guarde el archivo XML en el directorio de mondrian: WEB-INF/queries/

Schema Workbench: Creando un nuevo cubo.



```
<Schema name="Supermercado">
  <Dimension type="TimeDimension" name="Time">
    <Hierarchy name="time" hasAll="true" primaryKey="time_id"
      primaryKeyTable="time">
      <Table name="time">
      </Table>
      <Level name="Year" table="time" column="the year" type="Numeric"
        uniqueMembers="true" levelType="TimeYears" hideMemberIf="Never">
      </Level>
      <Level name="month" table="time" column="the month"
        type="String" uniqueMembers="false" levelType="TimeMonths" hideMemberIf="Never">
      </Level>
      <Level name="day" table="time" column="the day" type="String"
        uniqueMembers="false" levelType="TimeDays" hideMemberIf="Never">
      </Level>
    </Hierarchy>
  </Dimension>
  <Cube name="Sales" cache="true" enabled="true">
    <Table name="sale">
    </Table>
    <DimensionUsage source="Time" name="Time" foreignKey="time_id">
    </DimensionUsage>
    <Measure name="Store Cost" column="store_cost" datatype="Numeric"
      aggregator="sum" visible="true">
    </Measure>
    <Measure name="Store Sales" column="store_sales" datatype="Numeric"
      aggregator="sum" visible="true">
    </Measure>
    <Measure name="Unit Sales" column="unit_sales" datatype="Integer"
      aggregator="sum" visible="true">
    </Measure>
  </Cube>
</Schema>
```

Schema Workbench: Configurando la aplicación para el nuevo cubo.

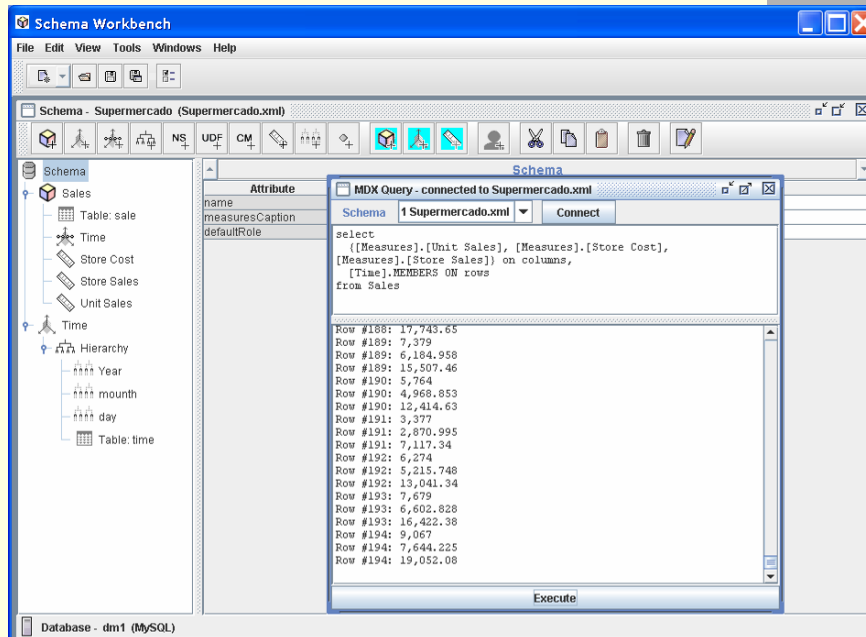
- Modificar los siguientes archivos:
 - mondrian\WEB-INF\mondrian.properties
 - mondrian\WEB-INF\datasources.xml
 - mondrian\WEB-INF\web.xml
 - mondrian\WEB-INF\query\mondrian.jsp
- Cambiar los textos:
 - MondrianFoodMart -> DM1
 - FoodMart.xml -> Supermercado.xml
- Borrar el texto:
 - "RoleXX='California manager';"
 - Esto se usaba en la base de demo porque definia un rol de usuario en el XML.

Schema Workbench: Configurando la aplicación para el nuevo cubo.

- Modificar los siguientes archivos:
 - mondrian\WEB-INF\query\mondrian.jsp
 - Allí aparece una query MDX que hay que cambiarla.
- Para ello antes usamos el Workbench para testear la query. Probemos con la siguiente:

```
select  
    {[Measures].[Unit Sales],  
    [Measures].[Store Cost],  
    [Measures].[Store Sales]} on columns,  
    [Time].MEMBERS ON rows  
from Sales
```

Schema Workbench: Configurando la aplicación para el nuevo cubo.



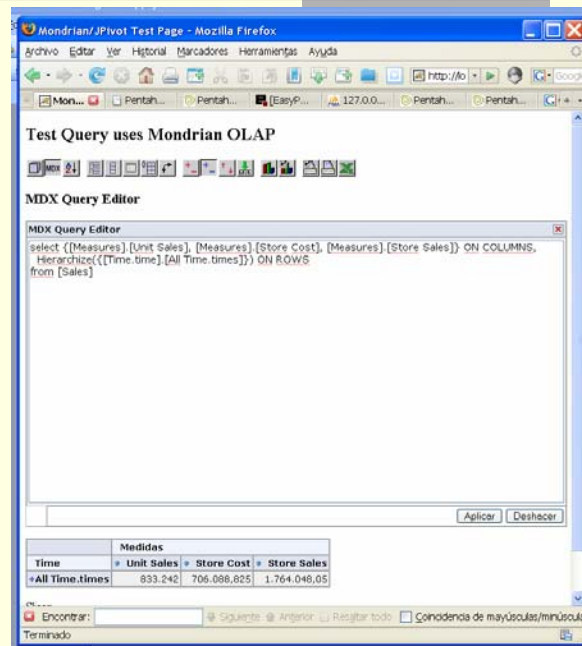
Schema Workbench: Configurando la aplicación para el nuevo cubo.

- Reemplazamos entonces la query antigua.

```
<jp:mondrianQuery id="query01"
  jdbcDriver="sun.jdbc.odbc.JdbcOdbcDriver"
  jdbcUrl="jdbc:odbc:DM1" catalogUri="/WEB-INF/queries/Supermercado.xml">
select
  {[Measures].[Unit Sales], [Measures].[Store Cost],
  [Measures].[Store Sales]} on columns,
  [Time].MEMBERS ON rows
from Sales
</jp:mondrianQuery>
```

Nuevo cubo simple configurado.

- Conviene reiniciar el Tomcat.
- Ya con la aplicación funcionando podemos mejorar “gráficamente” la query MDX.
- Notar que no hay “where” porque no existe otra dimensión en la cual hacer “slice”.



Kettle

- Extraction, Transformation, Transportation, Loading data.
- Spoon: Interfaz grafica.
- PAN: Transformaciones en batch.
- Kitchen: Orquestacion del proceso completo.

Configurando una conexión a una base de datos

