Guided Tutorial for Pentaho Pivot4J

This tutorial demonstrates usage of Pentaho BI Server with the Pivot4J plug-in to create pivot tables. You will use data sources and cubes that are already exist in Pentaho. Data warehouse processing requires a multidimensional data representation that is convenient for business analysts. Data cubes, a widely accepted representation of multidimensional data, support business intelligence processing by business analysts. In a data cube, cells contain numeric data called measures while rows and columns contain dimensions to organize the cells.

In Pentaho, pivot tables and pivot charts provide a convenient and flexible interface for manipulating data cubes. Pivot tables display data in rows and columns and allow convenient rearrangement of the row and column headings. Pivot charts display numerical data graphically to provide insights to business analysts. Pivot tables support dynamic manipulation of the row and column headings. Pivot tables in Pentaho are supported through a plug-in called Pivot4J.

# Starting Pentaho Pivot4J Plugin

This tutorial assumes that you have followed the instructions given in the document to install the Java Runtime Environment and the community edition of Pivot4J. Here are the steps to start the Pentaho BI server and use the plugin after installation.

* For **WINDOWS** users, you should go to the folder C:\Program Files\Pentaho\pentaho-server and right click the start-pentaho.bat file and select Run as administrator. A command window will open showing some executing commands. Keep the command window (Figure 1) open. You can make a shortcut to this batch file to make it more convenient to launch the BI Server.
* For **MAC** users, you should use the terminal to navigate to the directory **pentaho-server/** and execute the set-pnetaho-env.sh file using the command “./set-pentaho-env.sh” You have to run this command only once. Then, execute the file start-pentaho.sh by using “./start-pentaho.sh”. You need to use only the last command to start pentaho for the next time.
* To start the web interface for the Pentaho BI server, you should open your browser and navigate to the address <http://localhost:8080/>.
* After you launch the browser interface, you should click on Login as an Evaluator. Use the Administrator account to login (Figure 2). The default page of the platform will appear as in Figure 3.

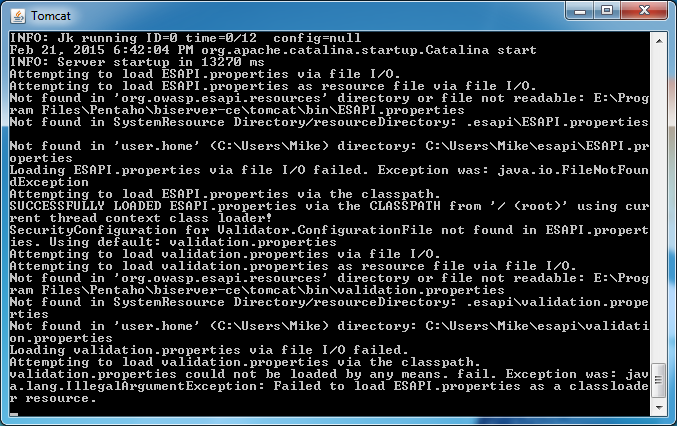


Figure 1: Windows Command Window for the BI Server

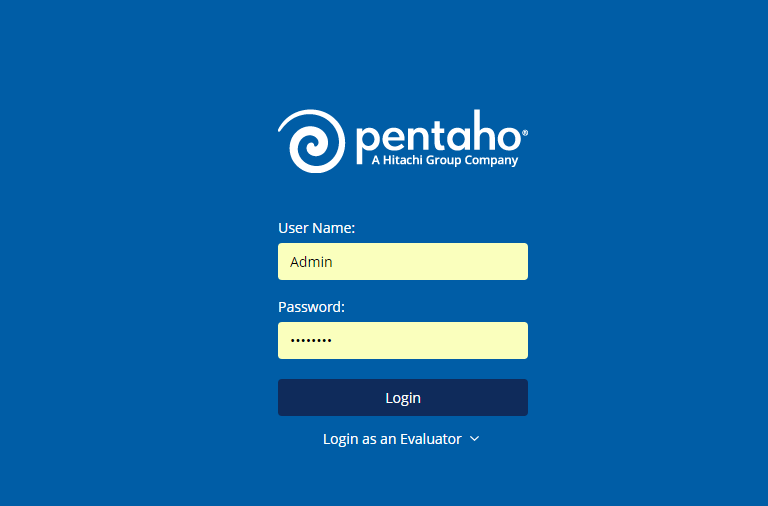


Figure 2: User Console Login Window

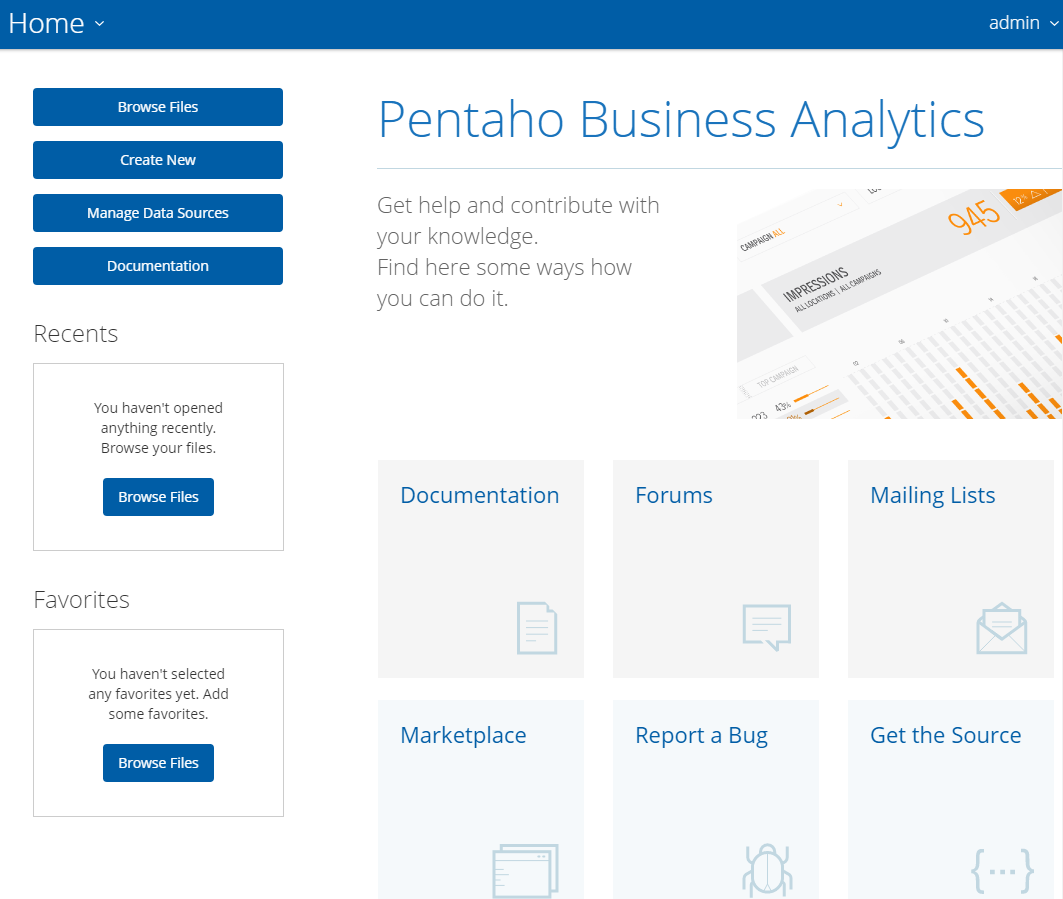


Figure 3: Home Page of the User Console

# Creating a Pivot Table

This tutorial uses a sample data source and cube that already exist in Pentaho to demonstrate creation of pivot tables. The tutorial does not cover the process of importing or connecting to other data sources or the steps to create cubes. The screen snapshots use an older stable version of the Pivot4J so you may see small differences with the current version. To create a pivot table in Pentaho, follow the next steps:

* Select data catalog and cube: Click on File  New  Pivot4J View. A new tab opens with two options to select the catalog and the cube (Figure 4). Select SteelWheels as a catalog. The SteelWheelsSales cube will be then selected by default. When you finish click OK. An empty Pivot table will open (Figure 5).

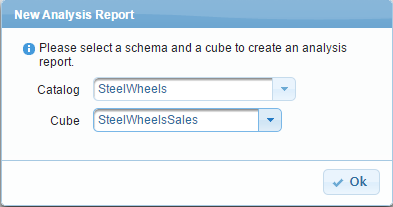
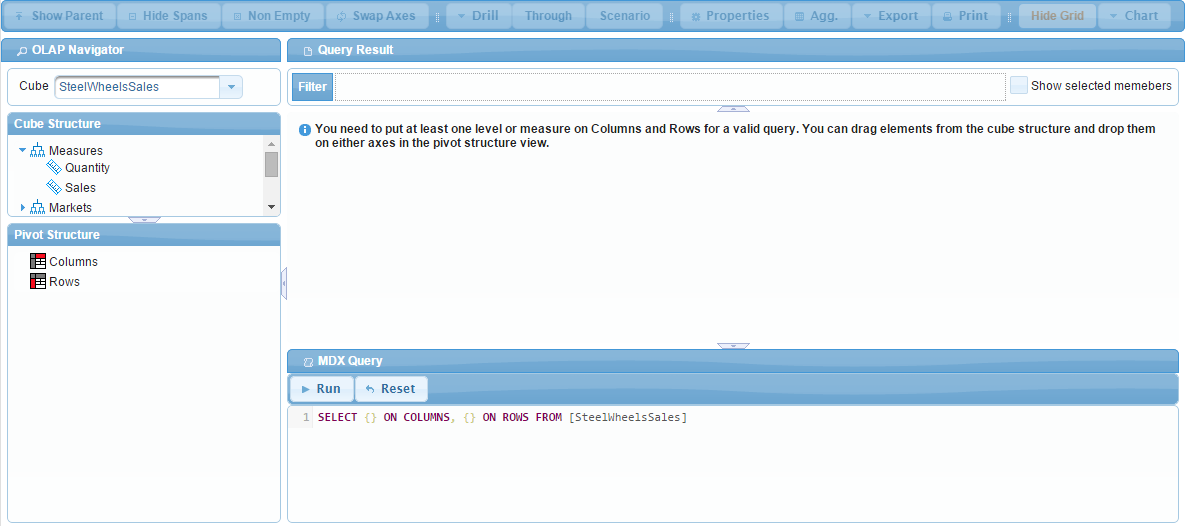


Figure 4: Catalog and Cube Selection

Figure 5: Pivot4J Table View

* Customize the Pivot Table: This table allows users to customize the contents of the table as well as the properties of the table (Figure 6). Customization will be discussed in the following section. When you finish, save the table as a report by clicking on Save () button. Write “SteelWheelsReport" as the file name and keep the default location /home/admin.
* Open saved reports: To open your saved reports, click on **File 🡪 Open**. Navigate to the required directory e.g. /home/admin. Select the report and click **Open**.

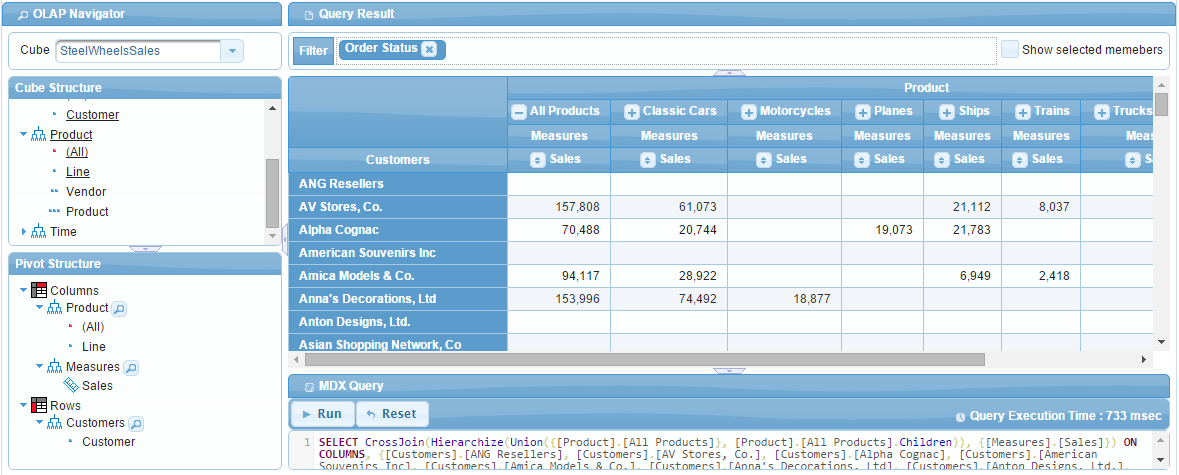
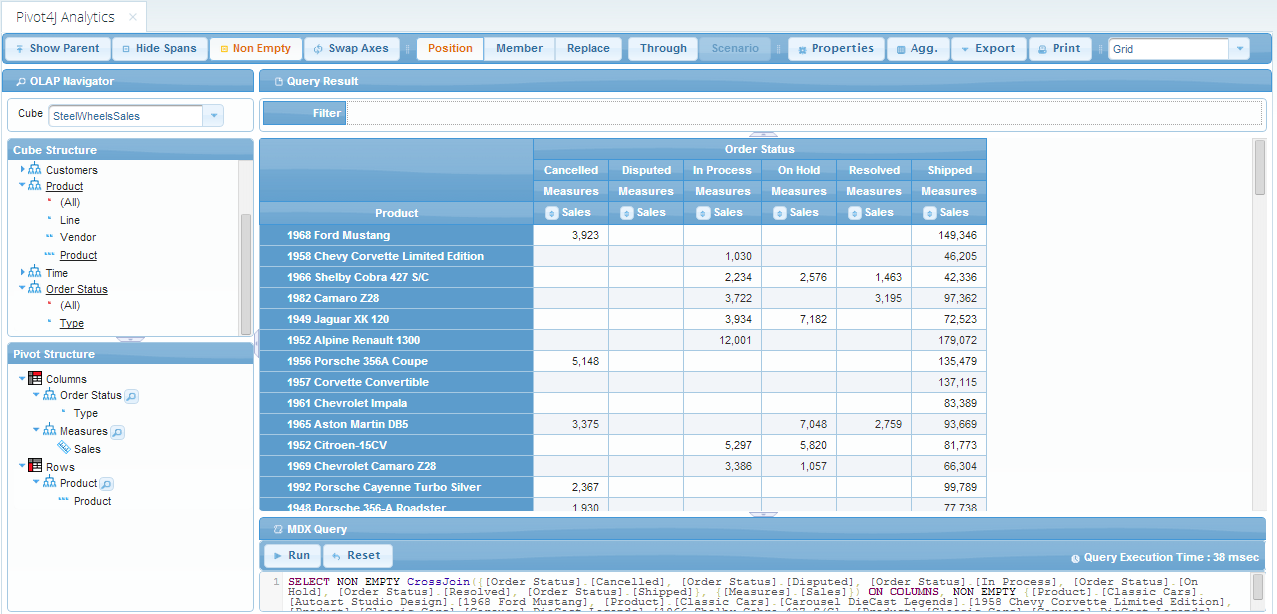


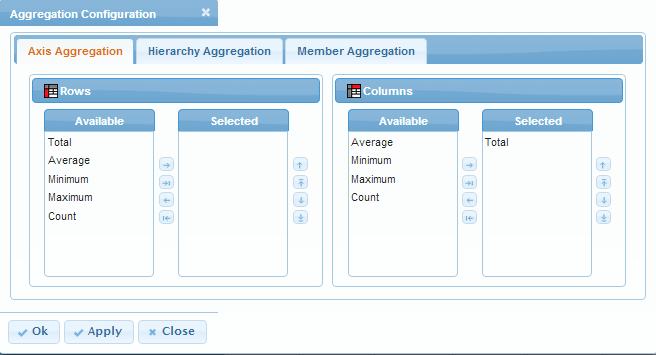
Figure 6: Pivot Table Using Pivot4J View

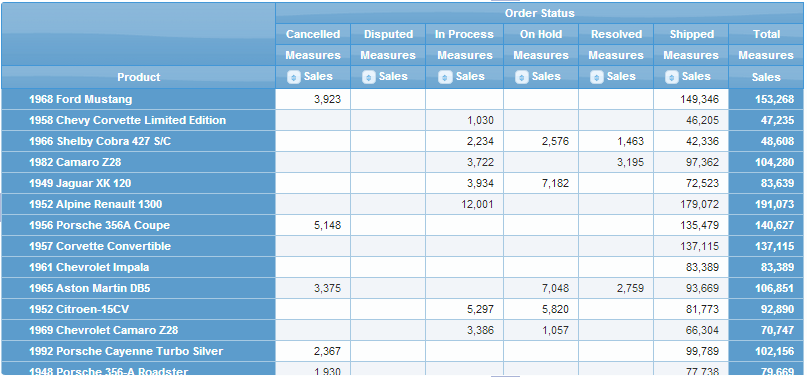
# Customizing the Pivot Table

You will customize the pivot table by placing fields, modifying the axes, and adding aggregation fields as demonstrated in the following steps.

* Place Fields on the Diagram: To place the fields onto the diagram, you may drag them from the Cube structure at the top into the Pivot structure at the bottom. To remove the fields from the diagram, drag them from the Pivot structure to the Cube structure.
* Until you are familiar with the design window’s labeled areas, it is suggested that you use the area choices as follows: place the Sales (from Measures) field in the columns, the Type (from Order Status) field in the columns area, and the Product (from Products) field in the row area (Figure 7).
* Modify the Pivot Table: You may add aggregation fields to the table. Common use of aggregation fields is to calculate the total of rows or columns. Click on the Agg button and the Aggregation Configuration window opens. In the Columns section under Axis Aggregation tab, Click on Total from the Available columns and select the arrow to move it to the Selected field (Figure 8). Click Ok. The Pivot table now shows the total sales for each product (Figure 9).

Figure 7: Pivot Table after Adding Values

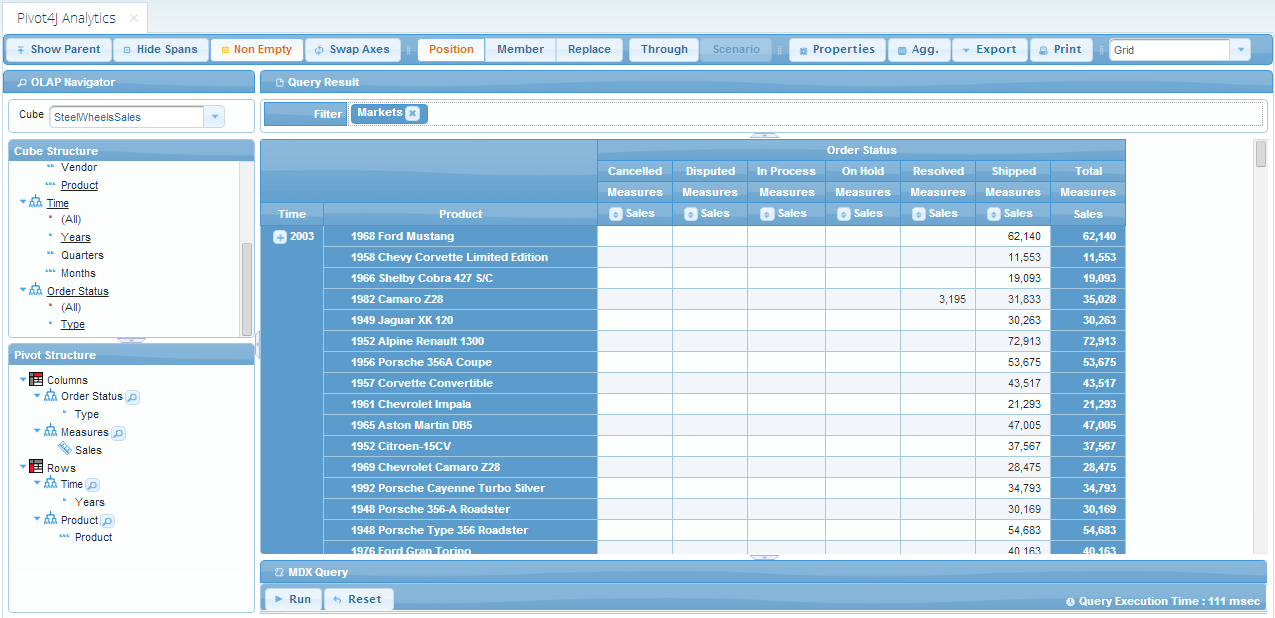
Figure 8: Aggregation Configuration

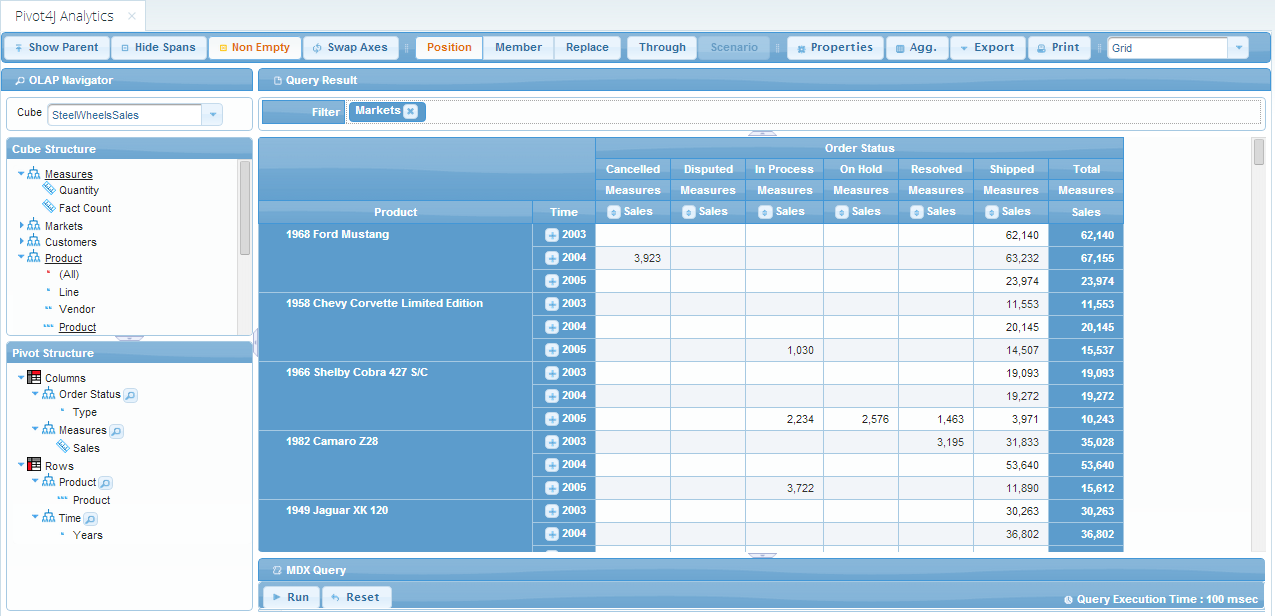
Figure 9: Pivot Table after Adding Total Field for Columns

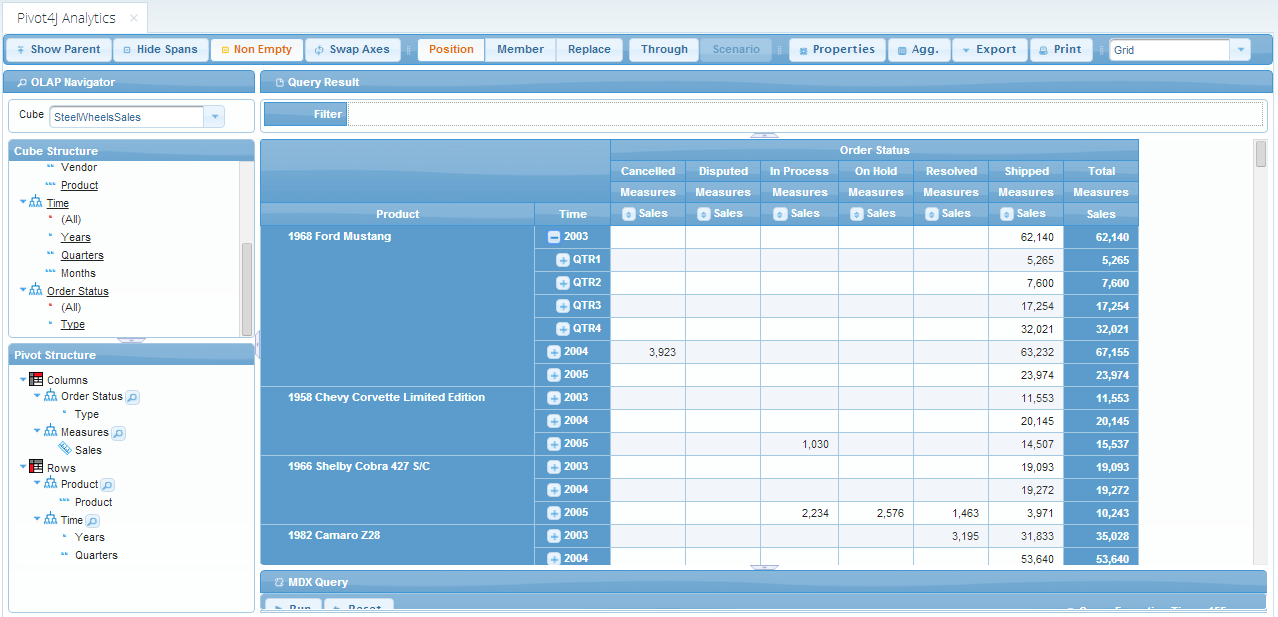
# Extending the Pivot Table

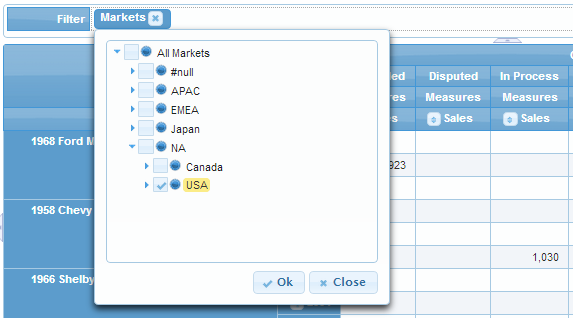
The pivot table in the previous section is somewhat limited because it lacks a time dimension, an important element of most data cubes. In addition, the pivot table lacks a filter field to restrict the data displayed. In this section, you will add these elements to the previous pivot table and then demonstrate some additional features of pivot tables.

* Add more fields: Place the Years (from Time) field in the Rows and Markets in the Filter (Figure 10).
* Change rows order: Placing Time before Products in figure 10 is not effective. Instead, dragging Product to Rows will change the order of the rows as in Figure 11.
* Expand and collapse views: You may now click on the plus sign next to the year and expand the view to reveal the quarters of the year (Figure 12).
* Filter the Pivot Table: To see the effect of filtering, drag Markets form the cube structure to the Filter area. Expand the NA and then select USA (Figure 13). Click Ok. Figure 14 shows the pivot table after applying the filter.
* Select specific Attributes: To select specific attributes to show in the table, click on the magnifier symbol next to the Order Status dimension name in the pivot structure. Remove all attributes from selected member’s field except On Hold and Shipped (Figure 15). Click OK. Figure 16 shows the resulting pivot table.

Figure 10: Pivot Table after Adding the Extra Fields

Figure 11: Pivot Table after Changing the Row Order

Figure 12: Pivot Table after Expanding Years into Quarters

Figure 13: Pivot Table Filtering

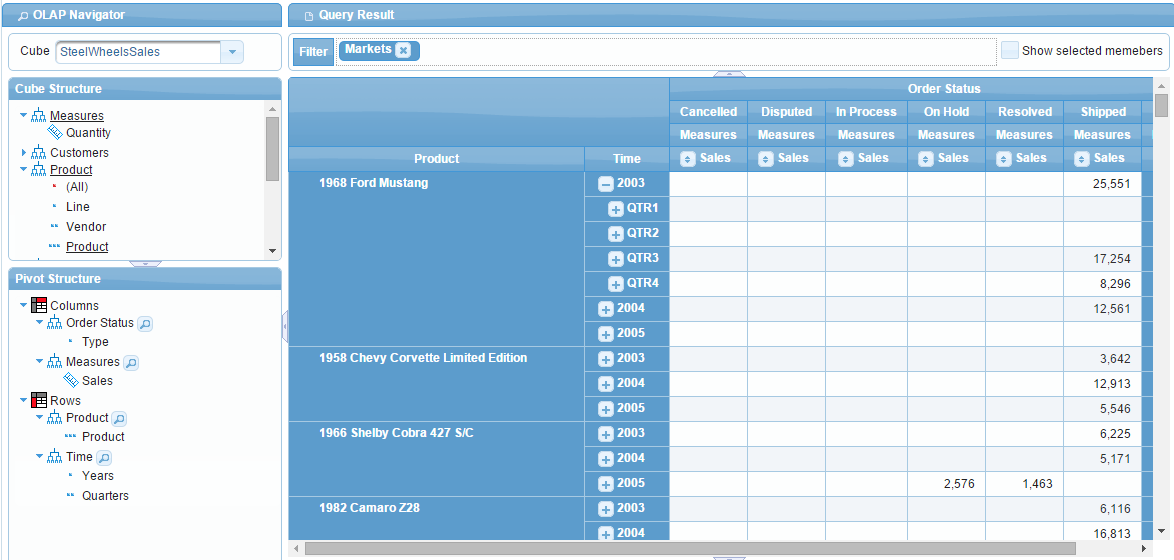
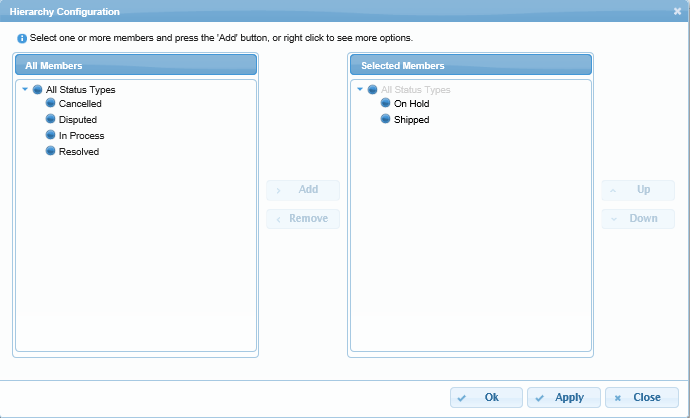


Figure 14: Pivot Table after Filtering on USA

Figure 15: Hierarchy Configuration of Order Status Dimension

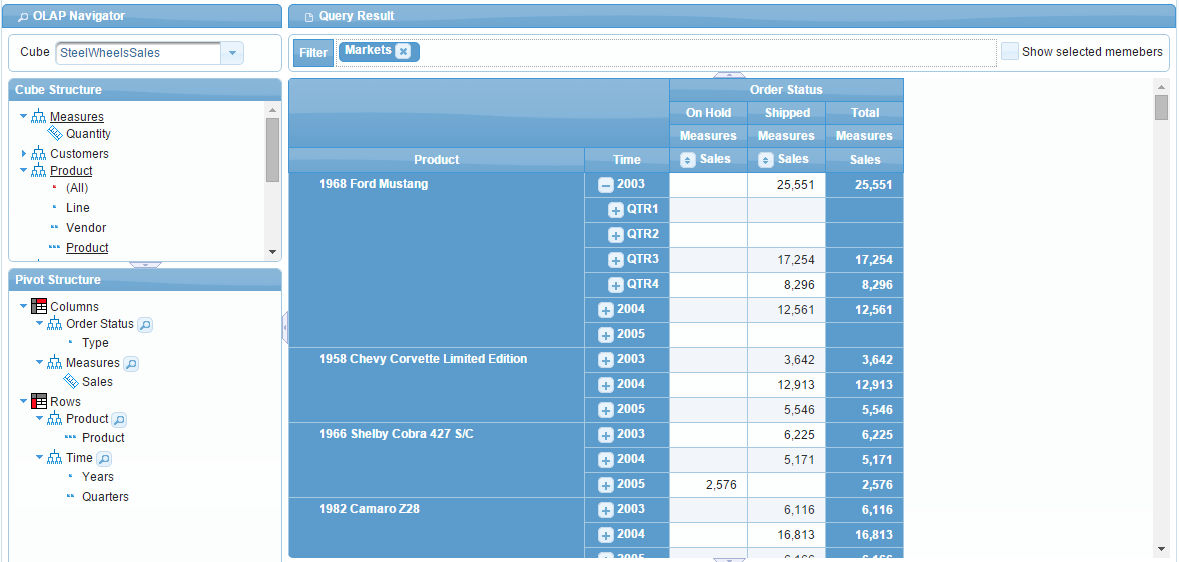
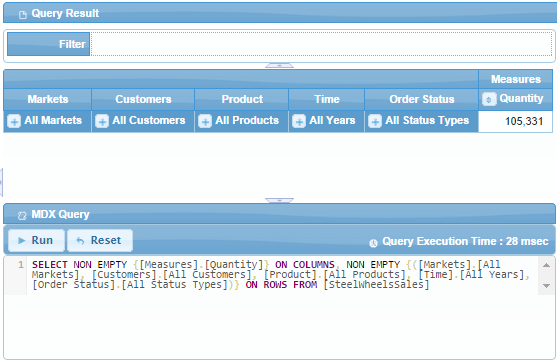


Figure 16: Pivot Table after Attributes Selection

# Pivot4J with MDX

Multi Dimensional Expression or MDX is a query language developed by Microsoft to manipulate multidimensional information (data cubes). Pentaho with its Pivot4J plug in supports MDX queries. To use MDX, simply start typing the query syntax in the MDX query field located in the bottom of the window. An example of MDX query is shown in Figure 17. Also, MDX in Pentaho is dynamic. Every time you drag and drop dimensions or expand and collapse the hierarchy, Pivot4J provides the corresponding MDX statement.

  
Figure 17: Query Example for the SteelWheelsSales Cube

To familiarize yourself with MDX, start a new pivot4J by clicking on File  New  Pivot4J View. Select **SteelsWheels** as the catalog and **SteelWheelsSales** as the cube and click OK. Try the followings and notice the changes to the syntax in the MDX query:

* Erase everything in the MDX field and type the following syntax:

*SELECT CrossJoin({[Order Status].[Shipped], [Order Status].[Resolved]}, {[Measures].[Sales], [Measures].[Quantity]}) ON COLUMNS, {[Time].[2003], [Time].[2004]} ON ROWS*

*FROM [SteelWheelsSales]*

*WHERE ([Product].[Classic Cars])*  
Click on **Run**. You should get the result similar to Figure 18.

* Start dragging or removing some of the dimensions from the cube structure to the pivot structure or vise versa.
* Expand and collapse some of the hierarchies by clicking on the plus sign next to the filed name.

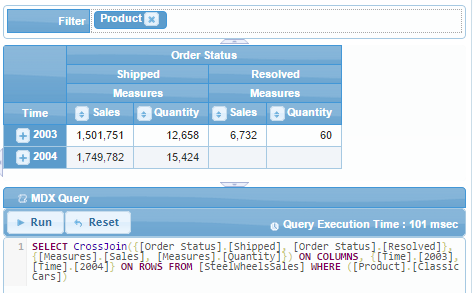


Figure18: Result for the MDX query

The MDX clauses that are used in the previous statement are:

* **Select**: This clause sets dimensions in the column or row axis
* **CrossJoin**: This clause combines two or more dimensions in single axis
* **From**: This clause determines the cube where the data come from
* **Where**: This clause slices the data for specific attributes in a dimension

# Stopping Pentaho Pivot4J Plugin

When you are finished, you should close the browser or logout and execute the stop procedure. In Windows, you should go to the folder C:\Program Files\Pentaho\pentaho-server and double click the stop-pentaho.bat file. A command window will open showing some executing commands. Then both command windows will close. In Mac, you should run “./stop-pentaho.sh”.