DEPARTAMENTO DE ELETRÓNICA, TELECOMUNICAÇÕES E INFORMÁTICA
UNIVERSIDADE DE AVEIRO

EXPLORAÇÃO DE DADOS / DATA MINING

(Part I) Laboratory Assignment 03: Introduction on using Pentaho Schema Workbench

Setup

1. You may need to copy a driver *postgresql-x.y.z.t* from directory *data-integration/lib* to *schema-workbench\drivers*.

Create a cube using Schema Workbench

1. Create a connection to your target database (data Mart).
2. Create a new Schema.
   - Go to the menu *File > New > Schema* and set the name of the schema to *Northwind*.
3. Create a new cube.
   - Right-click on *Schema*, select *Add cube* and set the name of the cube to *Sales*.
   - Right-click on *Sales*, select *Add table* to choose the fact table:
     - Set *schema* to *public*.
     - Set *name* to *F_SALES*.
4. Add the measures to the fact table.
   - Right-click on *Sales* and select *Add Measure*.
     - Set name to *Quantity*.
     - Set *aggregator* to *sum*.
     - Set *column* to *QUANTITY*.
     - Set *formatString* to *#.##0*.
     - Set *dataType* to *Numeric*.
   - You can add the other measures later.
5. Create the dimension *Employee*.

A dimension is joined to a cube by means of a pair of columns, one in the fact table, the other in the dimension table. The <Dimension> element has a foreignKey attribute, which is the name of a column in the fact table; the <Hierarchy> element has a primaryKey attribute.

- Right-click on *Sales* and select *Add dimension*. Set name to *Employee* and in foreign key select *ID_EMPLOYEE*.
- Expand *Employees* to show *New Hierarchy 0*. Right-click *New Hierarchy 0* and *Add table D_EMPLOYEE*. Set the attributes of the hierarchy:
  - Set name to Employee Hierarchy.
  - Set *allMembersName* to *All employees*.
  - Set *primaryKey* to *ID_employee*.
- Right-click *Employee Hierarchy* and select *Add Level*.
  - Set *name* to *Employees*.
  - Set column to *EMPLOYEE_NAME*.
  - Set *type* to *String*.
  - Check *uniqueMembers*.
  - Set *levelType* to *Regular*.
  - In *hideMemberIf* select *Never*.

6. Save the schema into a file called *Northwind*.

At this moment, you are ready to publish and use the cube just created. You may complete your cube definition later. For now, you must publish the cube (see *'Publish your schema in your Pentaho BI server*) and add the plugin Saiku Analytic to your BI Server installation (see *Visualize your cube using Pentaho BI Server and Saiku Analytics*).

7. Time dimensions based on *year/month/week/day* are coded differently due to the MDX time related functions. They have *type="TimeDimension"*. The role of a level in a time dimension is indicated by the level's attribute, e.g., *TimeYears, TimeQuarters, TimeMonths*, etc.
   - Right-click on *Sales* and select *Add Dimension*.
     - Set *name* to *Order date*.
     - In *foreign key* select *ID_TIME_ORDERDATE*.
     - In *type* select *TimeDimension*.

- Expand *Order date* to show *New Hierarchy 0*. Right-click *New Hierarchy 0* and Add table D_DATE. Set the attributes of the hierarchy:
  - Set name to *Order Date Hierarchy*.
  - Set *allMembersName* to *All order dates*.
  - Set *primaryKey* to *ID_DATE*.
- Right-click on *Order Date Hierarchy* and select *Add Level*.
  - Set *name* to *Year*.
  - Set *column* to *YEAR*.
  - Check *uniqueMembers*.
  - In *levelType* select *TimeYears*.
- Repeat the previous step for the levels *Quarter* and *Month*. You may complete the Time dimension later.

Save, publish and explore your cube.

8. If you are using a snowflake schema implementation for a given dimension, you must add a join instead of a table to build the hierarchy. For instance, implement the following steps to create the dimension *Product.*
   - Right-click on *Sales* and select *Add dimension*. Set name to *Product* and in foreign key select *ID_PRODUCT*.
   - Expand *Employees* to show *New Hierarchy 0*. Right-click *New Hierarchy 0* and select *Add Join*.
     - Select *Left join* and set *schema* to *public* and *name* to *D_PRODUCT*.
     - Select *Left join* and set *schema* to *public* and *name* to *D_CATEGORY*.
     - Click on *Join* and set *leftKey* to *ID_CATEGORY* and *rightKey* to *ID_CATEGORY*.
   - Set the attributes of the Product hierarchy:
     - Set name to *Product Hierarchy*.
     - Set *allMembersName* to *All products*.
     - Set *primaryKey* to *ID_product*.
   - Right-click *Product Hierarchy* and select *Add Level*.
     - Set *name* to *Products*.
     - Set column to *PRODUCT_NAME*.

– Set *levelType* to *Regular*.

- Repeat the previous steps to add a new level for the Categories.
- Right-click on the level *Category* and select *Move Level Up*. The most detailed level should be the last one.

Save, publish and explore your cube.

Publish your schema in your Pentaho BI server

You must create a new Data Source in your BI Server before publishing your cube. Launch your Pentaho BI Server and perform the following steps:

- Log in as administrator.
- Select *Manage Data Sources* and create a *New Connection* to your PostgreSQL database. Set the name of your connection to *PostgreSQL*. Test your connection.

Then, go to Schema Workbench and publish your cube:

- Go to the menu *File > Publish*.
- Set *server URL* to *http://localhost:8080/pentaho*.
- Set *User* to *admin* and *Password* to *password*.
- Set *Pentaho or JNDI Data Source* to *PostgreSQL*.
- Publish your schema.

Visualize your cube using Pentaho BI Server and Saiku Analytics

The plugin *JPivot* is being deprecated. The alternative recommended by Hitachi is Pentaho Business Analysis. However, this platform is not available in Pentaho Community Edition. So, instead of using this platform, we recommend using a plugin called *Saiku Analytic*, which allows working with data cubes in a similar way. The steps required to add the Saiku Analytic plugin into your current Pentaho BI Server are listed in https://edpflager.com/?p=3322.

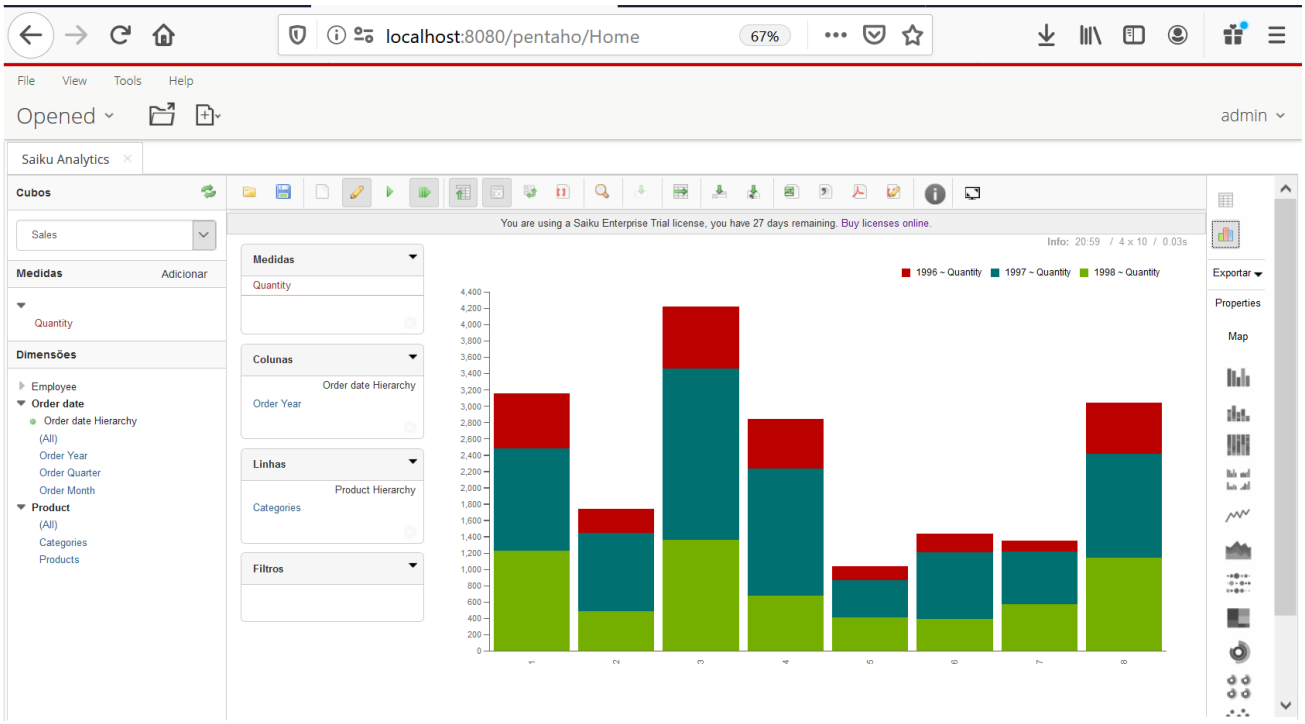
Then, you are ready to explore your data cube using a simple drag and drop interface as follows:

- Select *Manage Data Sources*, select *Northwind* and then select *Import Analysis* from the drop-down menu.
- In *Mondrian file* you must select the file where you saved you cube in Schema Workbench.
- In *data source*, you must select the *PostgreSQL* connection that you created before publishing the schema.
- Choose *Import* to complete this step.
- You must repeat this step every time you publish (new or update) a cube.

Finally, you must create a *Saiku Analytics* query to explore your data:

- Select *Create New* > *Saiku Analytics* > *Create a New Query*. You will see a drag-and-drop interface where you can select the cube Sales, the measure and the dimensions you want to visualize. You can also use operations such as drill-down, drill-across, etc., filter your data and visualize the results as charts.

- You may need to click on *Refresh* to see the most recent version of your cube.



Documentation

Pentaho Mondrian Documentation: https://mondrian.pentaho.com/documentation/schema.php.