Data Analysis and Integration

Lab 9: OLAP cubes and analytical queries

Note: In order to do this lab, you need to have successfully completed the previous lab.

In this lab, we will use Pentaho Schema Workbench (PSW) to define an OLAP cube, and we will use Saiku Analytics to explore the data and to perform some OLAP operations over that cube.

Starting PSW

1. Open a new terminal and navigate to the folder: **~/Pentaho/schema-workbench**

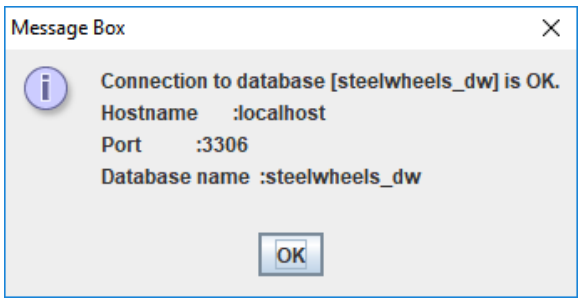
2. Start the Pentaho Schema Workbench (PSW) with: **./workbench.sh**

Configure the database connection in PSW

3. In PSW, go to the menu **Options > Connection**. The familiar **Database Connection** dialog will appear.

4. Configure the database connection as follows:
   - Connection Name: **steelwheels_dw**
   - Connection Type: **MySQL**
   - Access: **Native (JDBC)**
   - Host Name: **localhost**
   - Database Name: **steelwheels_dw**
   - Port Number: **3306**
   - User Name: **aid**
   - Password: **aid**

5. Press **Test** to test the connection. You should get the following message box:



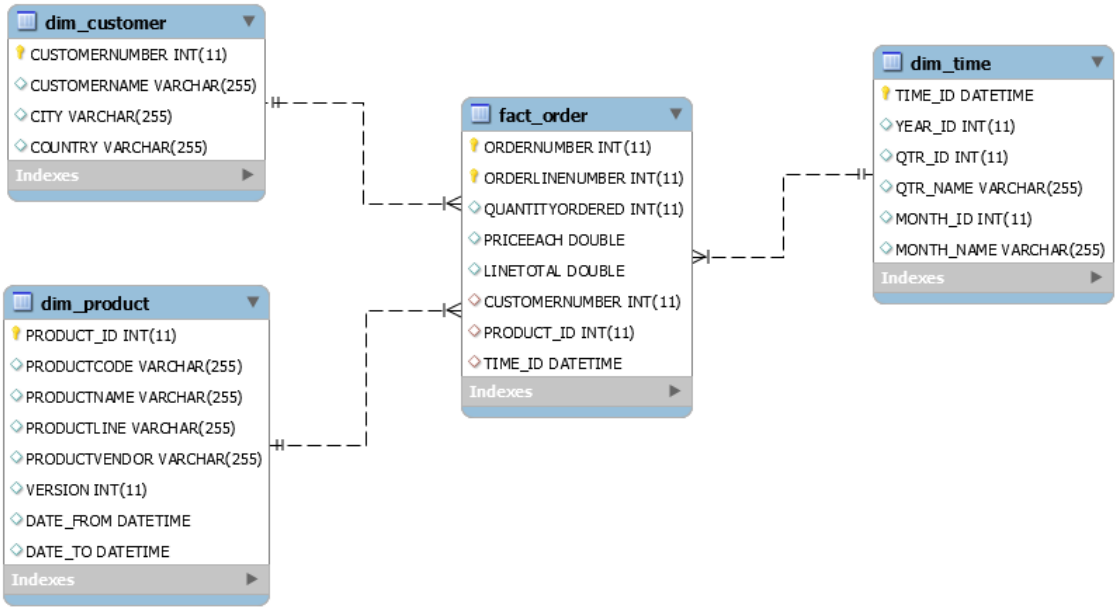
6. Close the **Database Connection** window with **OK**.

Exploring the database in PSW

7. In PSW, go to the menu option **File > New > JDBC Explorer**.

8. On the left pane, expand **Default Schema**. You will see the list of tables in the **steelwheels_dw** data warehouse.

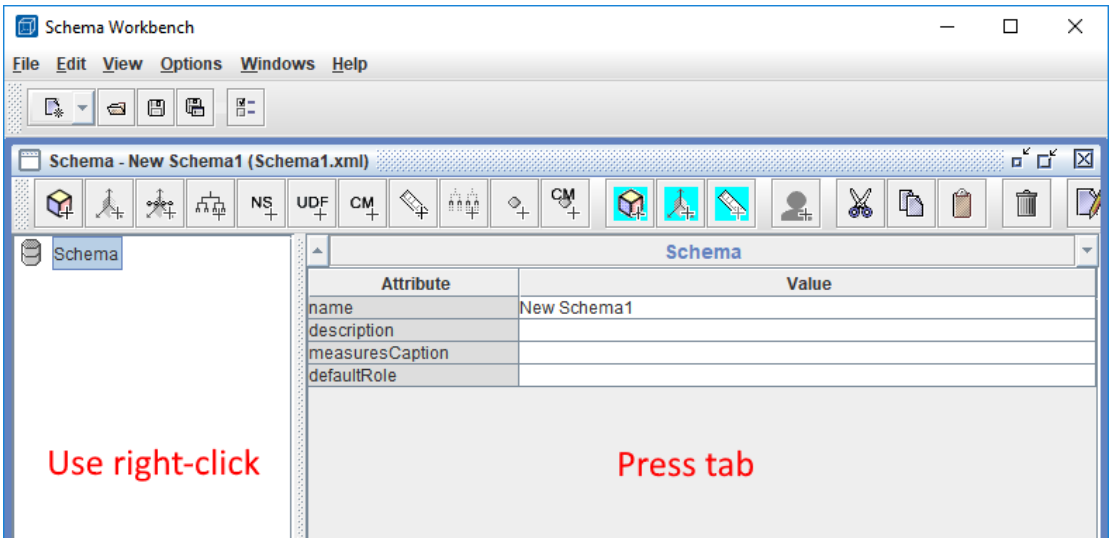
9. Expand each table to see their columns. There are three dimension tables, and one fact table. This is the star schema that we have created and populated in the previous lab.



Creating a new schema in PSW

10. In PSW, go to the menu option **File > New > Schema**.

11. You will be presented with the **New Schema** window. Before we proceed, remember these rules:
    - On the **left pane**, we will be adding things by **right-clicking** with the mouse.
    - On the **right pane**, we will be setting values by **pressing Tab** on the keyboard.



12. On the left pane, click on the **Schema** node and, on the right pane, set its **name** to **steelwheels_dw**. Press **Tab**.

13. Right-click on **Schema** and select **Add cube**.
    - Set **name** to **Orders**. Press **Tab**.

14. Right-click on **Orders** and select **Add Table**.
    - In **name** select **fact_order**. Press **Tab**. (Remember: always press **Tab**.)

Creating the Customers dimension

15. Right-click on **Orders** and select **Add Dimension**.
    - Set **name** to **Customer**
    - In **foreignKey** select **CUSTOMERNUMBER**

16. Expand **Customers** to show **New Hierarchy 0**.

17. Right-click **New Hierarchy 0** and select **Add Table**.
    - In **name** select **<<All Tables>>** and then **dim_customer**

18. Click on **New Hierarchy 0** and set its attributes:
    - Set **name** to **Customer Hierarchy**
    - Set **allMemberName** to **All Customers**
    - In **primaryKey** select **CUSTOMERNUMBER**

19. Right-click **Customer Hierarchy** and select **Add Level**.
    - Set **name** to **Country**
    - In **column** select **COUNTRY**
    - In **type** select **String**
    - In **levelType** select **Regular**

20. Right-click **Customer Hierarchy** and select **Add Level**.
    - Set **name** to **City**
    - In **column** select **CITY**
    - In **type** select **String**
    - In **levelType** select **Regular**

21. Right-click **Customer Hierarchy** and select **Add Level**.
    - Set **name** to **Customer Name**
    - In **column** select **CUSTOMERNAME**
    - In **type** select **String**
    - In **levelType** select **Regular**

Creating the Products dimension

22. Right-click on **Orders** and select **Add Dimension**.
    - Set **name** to **Product**
    - In **foreignKey** select **PRODUCT_ID**

23. Expand **Product** to show **New Hierarchy 0**.

24. Right-click **New Hierarchy 0** and select **Add Table**.
    - In **name** select **<<All Tables>>** and then **dim_product**

25. Click on **New Hierarchy 0** and set its attributes:
    - Set **name** to **Product Hierarchy**
    - Set **allMemberName** to **All Products**
    - In **primaryKey** select **PRODUCT_ID**

26. Right-click **Product Hierarchy** and select **Add Level**.
    - Set **name** to **Product Line**
    - In **column** select **PRODUCTLINE**
    - In **type** select **String**
    - In **levelType** select **Regular**

27. Right-click **Product Hierarchy** and select **Add Level**.
    - Set **name** to **Product Vendor**
    - In **column** select **PRODUCTVENDOR**
    - In **type** select **String**
    - In **levelType** select **Regular**

28. Right-click **Product Hierarchy** and select **Add Level**.
    - Set **name** to **Product Name**
    - In **column** select **PRODUCTNAME**
    - In **type** select **String**
    - In **levelType** select **Regular**

Creating the Time dimension

29. Right-click on **Orders** and select **Add Dimension**.
    - Set **name** to **Time**
    - In **foreignKey** select **TIME_ID**
    - In **type** select **TimeDimension**

30. Expand **Time** to show **New Hierarchy 0**.

31. Right-click **New Hierarchy 0** and select **Add Table**.
    - In **name** select **<<All Tables>>** and then **dim_time**

32. Click on **New Hierarchy 0** and set its attributes:
    - Set **name** to **Time Hierarchy**
    - Set **allMemberName** to **All Years**
    - In **primaryKey** select **TIME_ID**

33. Right-click **Time Hierarchy** and select **Add Level**.
    - Set **name** to **Year**
    - In **column** select **YEAR_ID**
    - In **type** select **Integer**
    - In **levelType** select **TimeYears**

34. Right-click **Time Hierarchy** and select **Add Level**.
    - Set **name** to **Quarter**
    - In **column** select **QTR_NAME**
    - In **ordinalColumn** select **QTR_ID**
    - In **type** select **String**
    - In **levelType** select **TimeQuarters**

35. Right-click **Time Hierarchy** and select **Add Level**.

- Set **name** to **Month**
- In **column** select **MONTH_NAME**
- In **ordinalColumn** select **MONTH_ID**
- In **type** select **String**
- In **levelType** select **TimeMonths**

Creating the Sales and Quantity measures

36. Right-click on **Orders** and select **Add Measure**.
    - Set **name** to **Sales**
    - In **aggregator** select **sum**
    - In **column** select **LINETOTAL**
    - Set **formatString** to **$ #,###.00**
    - In **datatype** select **Numeric**

37. Right-click on **Orders** and select **Add Measure**.
    - Set **name** to **Quantity**
    - In **aggregator** select **sum**
    - In **column** select **QUANTITYORDERED**
    - Set **formatString** to **#,###**
    - In **datatype** select **Integer**

38. Save the schema in a file called **steelwheels_dw.xml**.

Starting the Pentaho Server

39. Open a new terminal and navigate to the folder: **~/Pentaho/pentaho-server/**

40. Start the Pentaho Server with: **./start-pentaho.sh**
    Notes: Pentaho Server will start running in the background. It may take some time for its startup to complete. You can use System Monitor to check CPU activity and memory usage. When the CPU calms down, you can proceed to the next step.

41. Open Firefox and navigate to: **http://localhost:8080/**

42. On the **Welcome** page, press **Log in as an evaluator** and **Log in** as **Administrator**.

Creating the data sources

43. On the **Home** page, click **Manage Data Sources**.

44. In the **Manage Data Sources** window, click the small settings button and select **New Connection**.

45. Configure the database connection as follows:
    - Connection Name: **steelwheels_dw**
    - Database Type: **MySQL**
    - Access: **Native (JDBC)**
    - Host Name: **localhost**
    - Database Name: **steelwheels_dw**

- Port Number: **3306**
- User Name: **aid**
- Password: **aid**

46. Press **Test** to confirm that the database connection is working.

47. Close the **Database Connection** window with **OK**.

48. Click the small settings button and select **Import Analysis**.

49. In the **Import Analysis** window:
    - Browse to the **steelwheels_dw.xml** file that you created with PSW
    - In **Data Source**, select **steelwheels_dw**

50. Press **Import** to close the **Import Analysis** window.

51. In the **Manage Data Sources** window, there are now two **steelwheels_dw** data sources:
    - One is the database connection to the data warehouse (**steelwheels_dw : JDBC**)
    - The other is the XML definition of the OLAP cube (**steelwheels_dw : Analysis**)

52. Press **Close** to close the **Manage Data Sources** window.

Exploring the data cube

53. In the **File** menu at the top left corner, select **New > Saiku Analytics**.

54. Next to **Cubes**, click on the green button to **Refresh Cubes (Clear Cache)**.

55. In **Select a cube**, select the **Orders** cube under the **steelwheels_dw** data source.
    Note: If the Orders cube does not appear in the list, it is very likely that there is a problem with the XML definition. In this case, you should go back to PSW to fix those issues. To help debug the problem, check for exceptions in pentaho-server/tomcat/logs/pentaho.log

56. Expand the dimensions **Customer**, **Product**, and **Time**.

57. Drag the following elements from the left pane to the right pane:
    - Drag **Sales** to **Measures**.
    - Drag **Country** to **Rows**.

    You should see sales by country.

58. Drag **Product Line** to **Columns**. You should see sales by country and product line.

59. Drag **City** to **Rows**. You should see a drill-down of sales by country and city.

60. Drag **Year** to **Columns**. You should see sales city, product line, and year.

61. Drag **Quantity** to **Measures**. You should see both sales and quantity in each cell.

62. Take a screenshot showing the final results that you got.