Data Masters Project

For CIS 4400

By: Eric Kojo Adjetey, Bright [Adu Poku,](mailto:BRIGHT.ADUPOKU@baruchmail.cuny.edu) Polina [Grigoryeva,](mailto:POLINA.GRIGORYEVA@BARUCHMAIL.CUNY.EDU) Zurab [Kiriya,](mailto:ZURAB.KIRIYA@BARUCHMAIL.CUNY.EDU)

Gregory Szymanski

12/22/2015

Green Cycle refers to a large, multinational manufacturing company. The company manufactures and sells metal and composite bicycles to North American, European and Asian commercial markets. While its base operation is located in USA with 290 employees, several regional sales teams are located throughout their market base. Coming off a successful fiscal year, Green Cycle is looking to broaden its market share by targeting their sales to their best customers, extending their product availability by reducing their cost of sales through lower production costs.

With the company’s expansion a need emerged to analyze data that Green Cycle has collected over the years. Data Masters has created a data warehouse that allows Green Cycle to perform BI analysis on such topics like internet sales or account balance.

Internet sales will be analyzed on the base of the following measures:

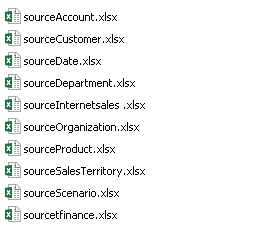
* Products
* Date of sales
* Organizations (Buyers)
* Territory
* Order Numbers
* Order Quantity

The finances (account balances) will be analyzed on the base of such elements like:

* Date
* Organization
* Department
* Account
* Budget Scenarios

The data for both analysis were collected and saved as Excel files. All of the files are located in separate folder named SourceFolder.zip.

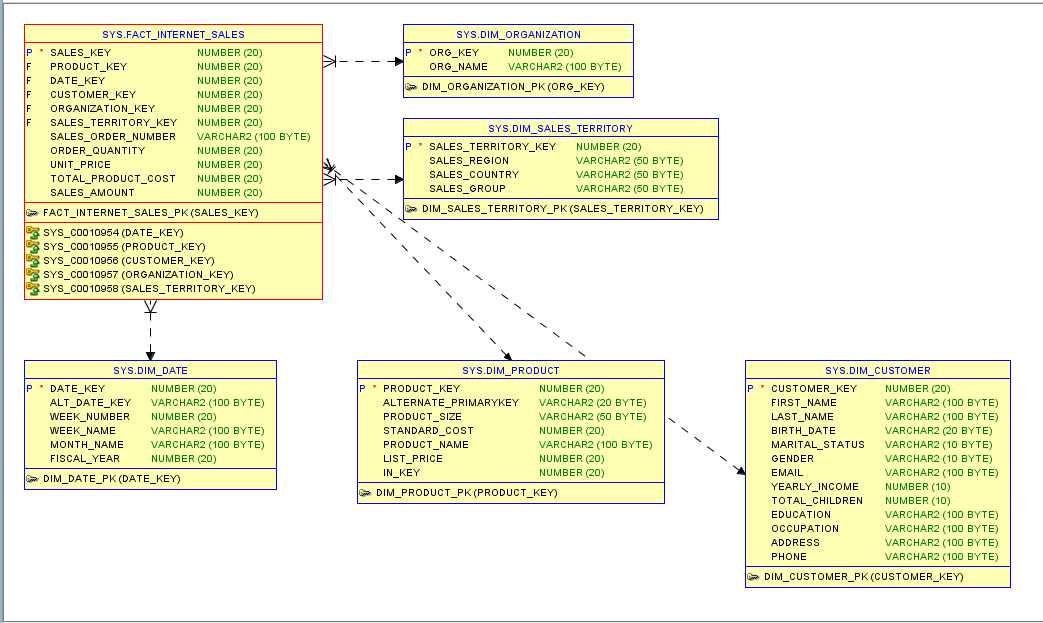
**1.1 Source files**



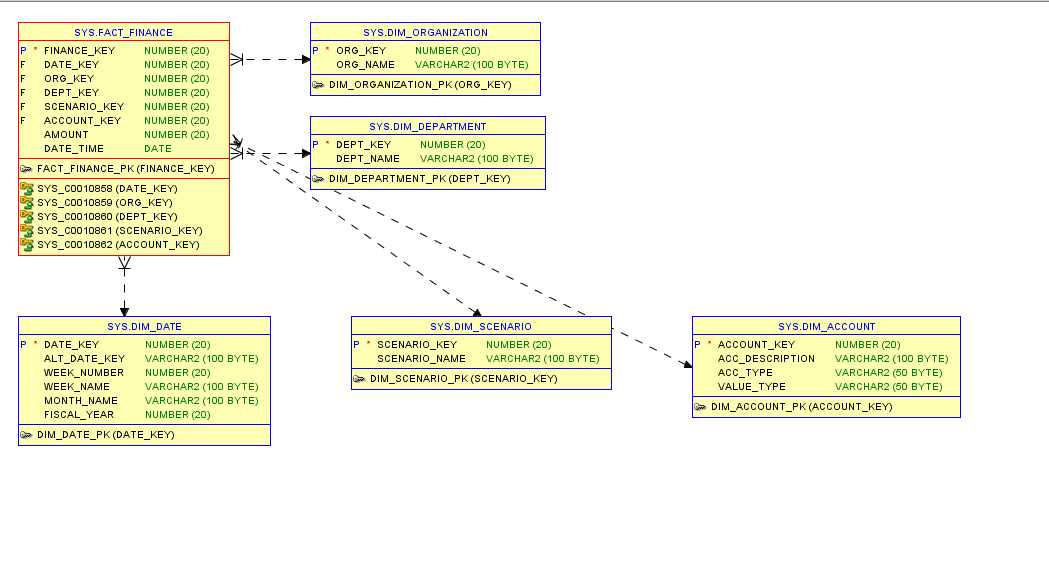
After collecting the data, the validation process began. The validation included filling empty cells with data and removing records that didn’t make sense. For example the amount sold can’t be described in the form of a sentence. It has to be a numerical value.

In Oracle Developer we ran SQL script to generate a warehouse schema. The script is available in SourceFolder.zip that was emailed with this document. The data warehouse schema consists of two fact tables. Each fact table consists of five dimensions. The created schema for this project is a star schema.

**1.2 Internet Sales schema**



**1.3 Finance schema**

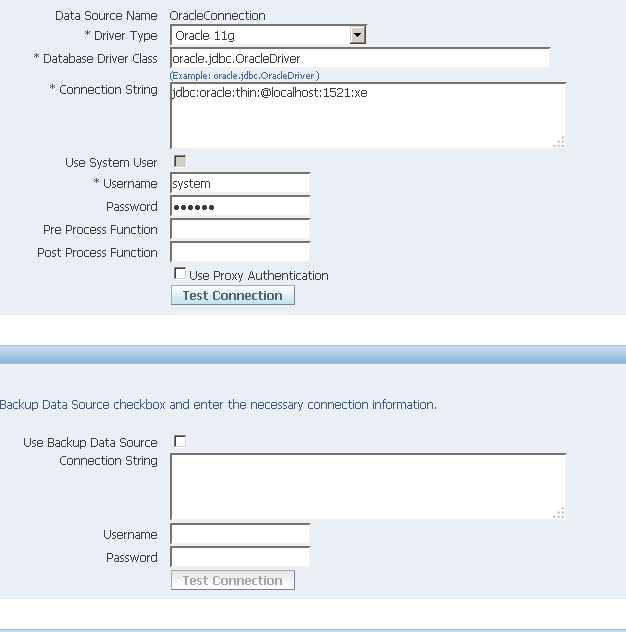


After creating the schemas, data from the Excel source files were loaded to the fact and dimension tables. During this process constraints integrity was forced. The data that didn’t fulfill the constraints requirements weren’t loaded to the data warehouse. After this the creation of the data warehouse was completed.

The following steps were completed in Oracle Publisher.

To acquire the data from the data warehouse a JDBC connection was created.

**1.4 JDBC connection configuration screen**



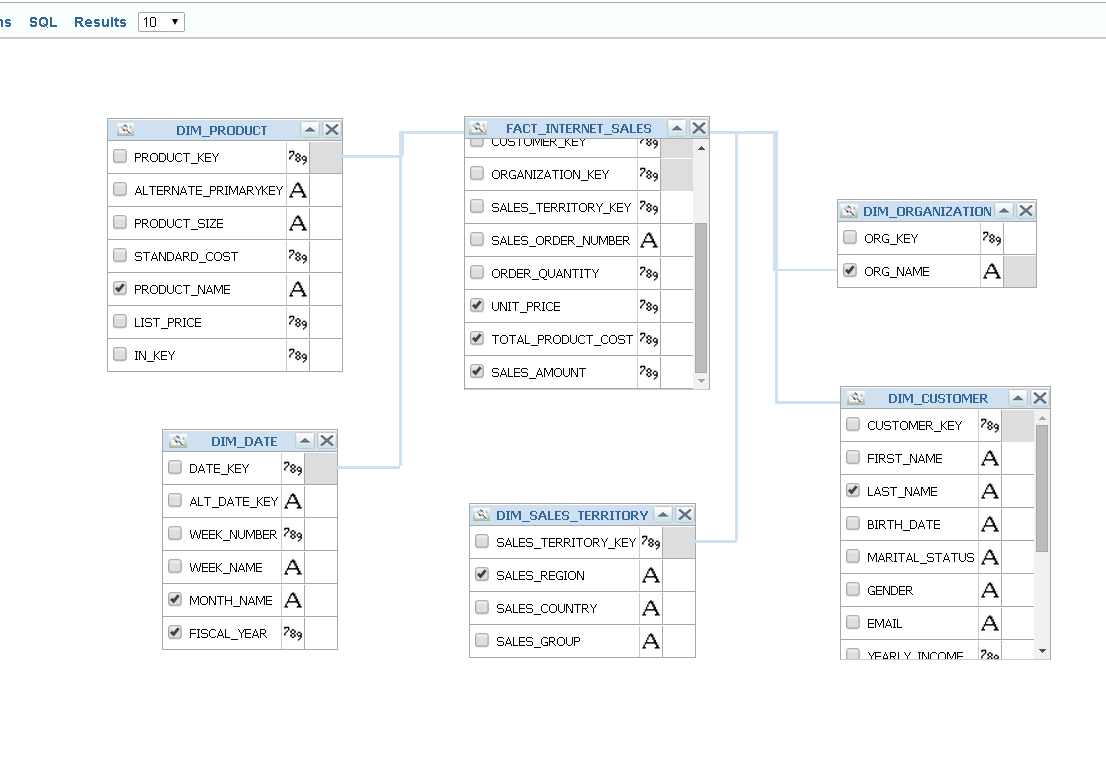
In the Oracle Publisher two data models were created for Business Intelligence analysis. The first data model is named Data Internet Sales. The second model is named Data Financial Account.

The ETL process was perfomed on both data models with the SQL Builder available in Publisher.

For the Internet Sales model we:

extracted following attributes PRODUCT\_NAME, LAST\_NAME, ORG\_NAME, SALES\_REGION, MONTH\_NAME, FISCAL\_YEAR, TOTAL\_PRODUCT\_COST, UNIT\_PRICE and SALES\_AMOUNT.

**1.5 Internet Sales model – logical layer**



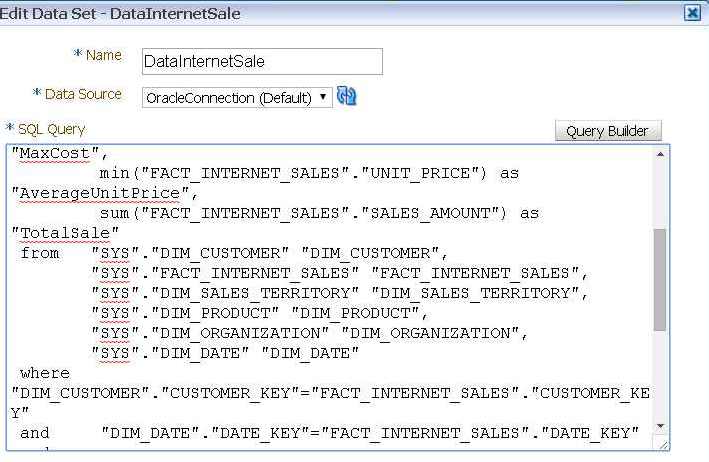
-transformed by creating alias for the extracted attributes, organized order, used grouping for attributes and applied functions like SUM, TRIM, AVG and a few others.

**1.6 Conditions for Internet Sales model**



- loading by creating SQL query and saving to the Internet Sales model.

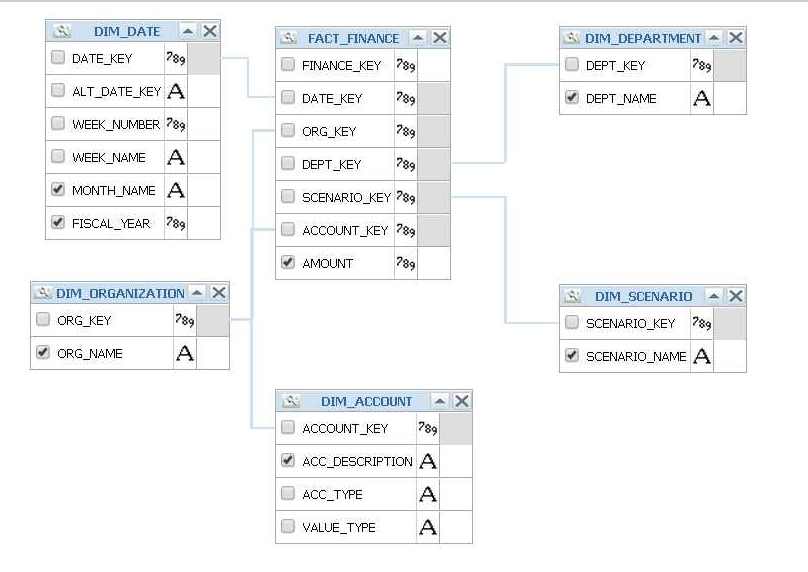
**1.7 SQL query design**



For the Finance model we:

-extracted following attributes ORG\_NAME, ACC\_DESCRIPTION, DEPT\_NAME, SCENARIO\_NAME, MONTH\_NAME, FISCAL\_YEAR, AMOUNT.

**1.8 Internet Finance model – logical layer**



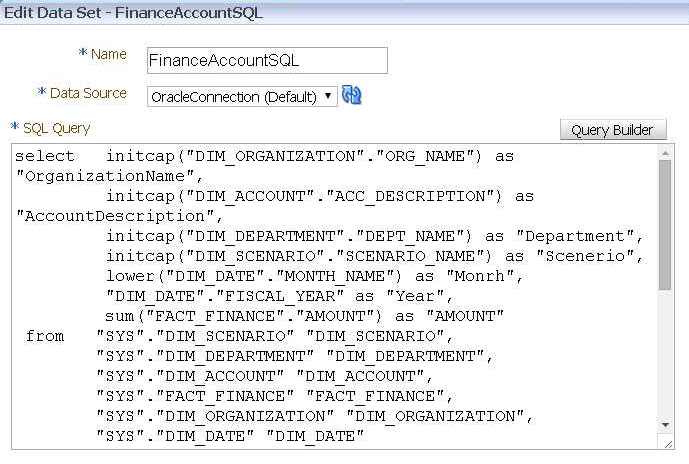
-transformed by creating alias for the extracted attributes, organized order, used grouping for attributes and applied functions like SUM, TRIM, AVG and a few others..

**1.9 Conditions for Finance model**



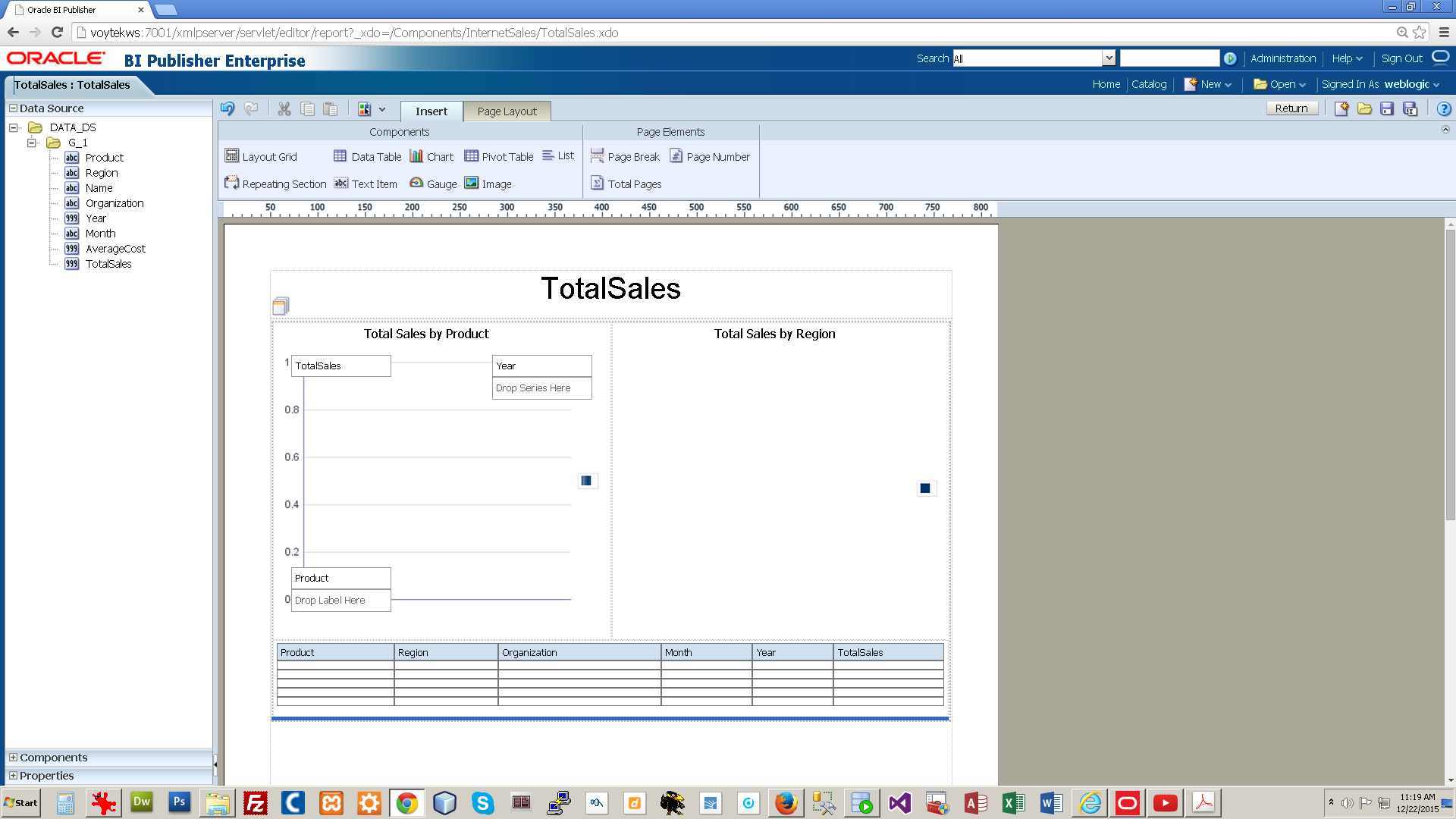
- loading by creating SQL query and saving to the Finance model.

**1.10 SQL query design**



With the use of the Data Internet Sales model a dashboard was created. The dashboard was created with the use of Report Editor that is available in Oracle Publisher.

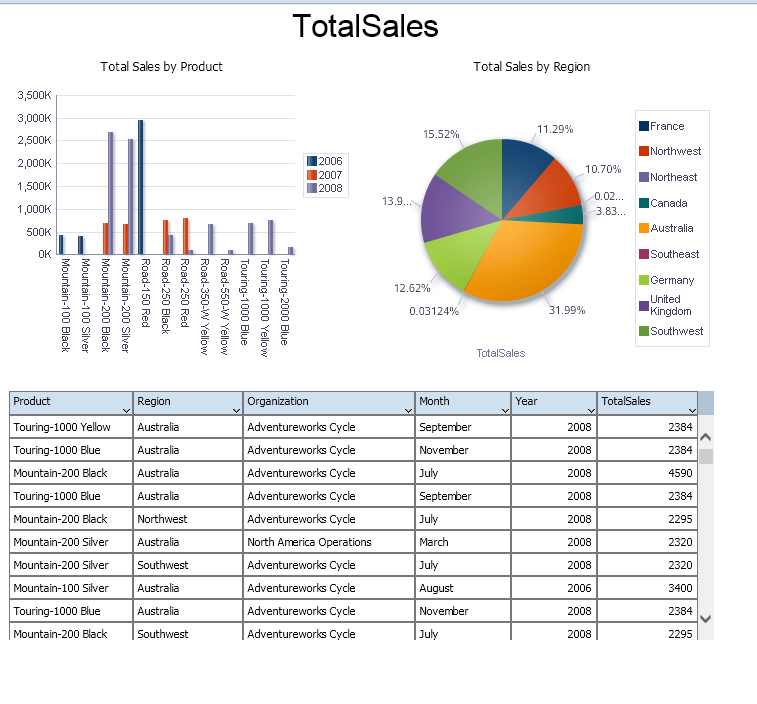
**1.11 Report Editor Window**



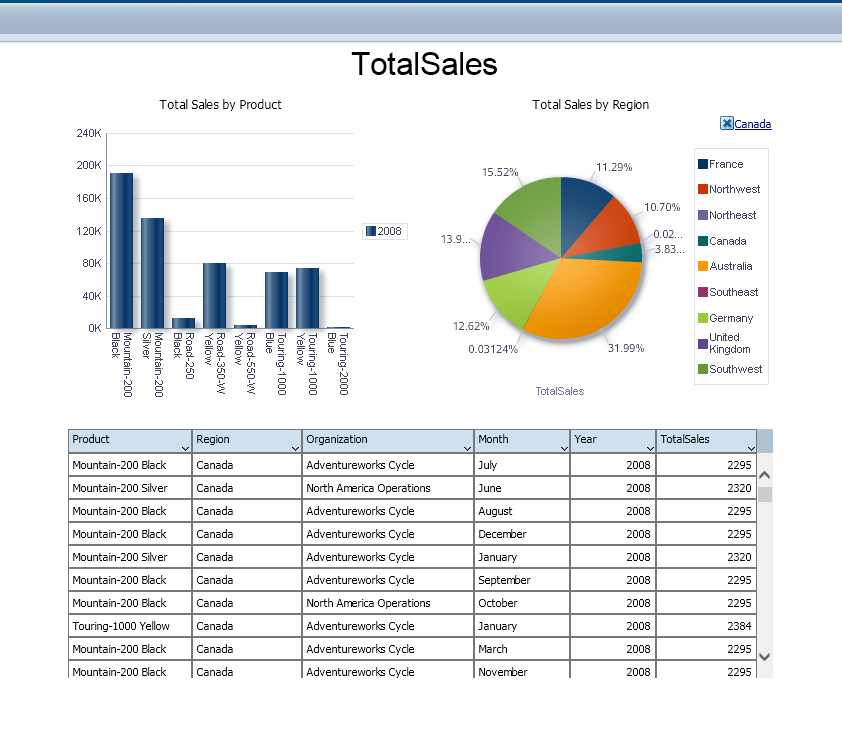
The dashboard that was created enables presentation of total sales by product and region. The column chart in the upper left corner presents total sales by product where the pie chart presents total sales by region. We added a condition to the data source to present only total sales over $100,000. On the dashboard filters are added that let you filter total sales by region and year.

For example in the screenshot below you can see that picture 1.13 shows you the use of a filter. For region we picked Canada and for year we picked year 2008. This result is presenting the total sales by product in Canada in the year 2008. This change is visible in the column chart and the tabular chart below. The pie chart remains unchanged.

**1.12 Dashboard, total sales for all regions and years**.



**1.13 Dashboard, filter used, total sales for Canada and year 2008**



**Conclusion**

The project turned out to be a challenge for the group. There are two main issues that caused the most problems to the group. First of all it was hard to balance all the work between all group members. It was difficult to time every step of the assignment and divide the work in a way that would allow every member to have enough time to finish their part of the project. Lack of experience in leading such a project also caused serious slowdowns in completing the work.

The other thing that caused a big problem was the installation of such tools like OBIEE. The team used a lot of time to try install and learn a new tool like OBIEE. It often was a very tiresome process to try and keep up with both the knowledge about building data warehouse and implementing this knowledge by using such tools like Oracle Developer or Oracle Publisher. It was also very frustrating because we got stuck a few times at one point of the project purely because we lacked information about how to do something by using a specific tool. For instance we really wanted to do ETL by using a specialized tool like OBIEE, but the installation issues caused a severe delay in completing our goal. After several tries though we managed to install and figure out OBIEE to a certain level.

We feel that we learned a lot of new things about the ETL and tools that can be used to complete the ETL process. The ETL process required from us the most amount of time for research. With more time we probably could do this part much better now. It is also good that we finally were able to use OBIEE for a little bit. We find it to be a very useful tool.

If we could start the project all over again we would definitely start working on it much earlier. We underestimated how much time every step of the assignment would take. We would use the extra time to make sure that every tool is working properly and prepare a better plan for each step of the project.

Beside that we would also try find a better way to validate our data that would involve using a different tool than Excel. Also the data loading process to Oracle Developer from source files has been done manually. We are not sure if this is possible but maybe it would be possible to find a tool that makes the process of loading the data more automated. Is there a more specialized tool for loading data to a data warehouse than Oracle Developer?

In general we feel that this was a difficult project. It challenged our knowledge but also such skills like team work and time management. We feel that more time in class should be dedicated to such kind of project. The extra time in class could be used to show how certain tools work, discuss in more detail other similar projects from the past, explain common mistakes that other students made while working on this type of assignment. No matter what the approach is we believe there should be a lot more discussion in class that is dedicated directly to how to work and complete such kind of group assignment.