Data Warehouse ETL - report (Spark)

**Content**

*1. Case Study – Data Communication(Spark) Page 2*

*2. Create Data Warehouse Schema Page 3-5*

*3. Data Source Page 6-8*

*4. Data Extraction using SSIS Page 9-10*

*5. Data Transformation – Excel Page 11-12*

*6. Data Transformation – SSIS Page 13*

*7. Loading using SSIS Page 13-14*

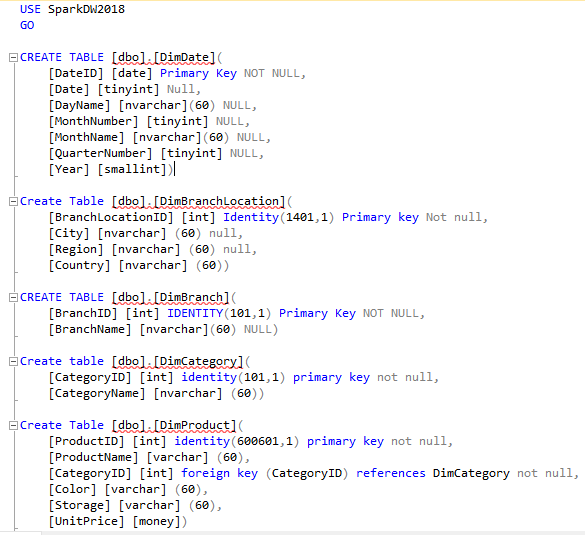
*8. Appendix Page 15*

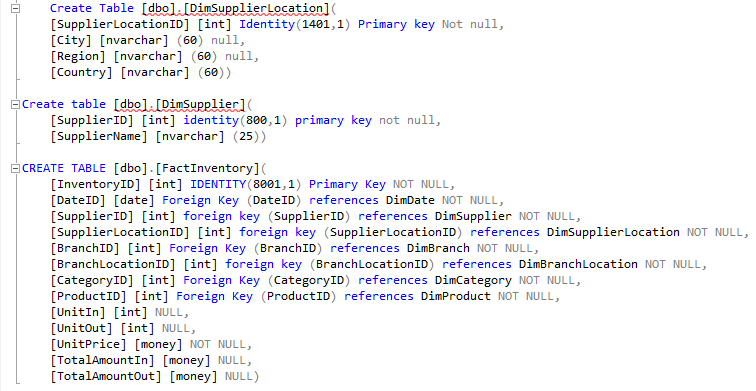
1. **Case Study – Data Communication(Spark): -**

Spark is the largest network provider of the New Zealand. There are different subject areas in this company, but we only focused on three data marts Sales, Inventory and Popularity. I work on the data mart Inventory which is the important subject area for the data warehouse. With the help of the data form the data warehouse I can create the report on how the business is going on. In addition, it also helps to make decision for the future importing and the exporting the products. Without the data warehouse it is difficult to make the decisions which effects on the growth of the business.

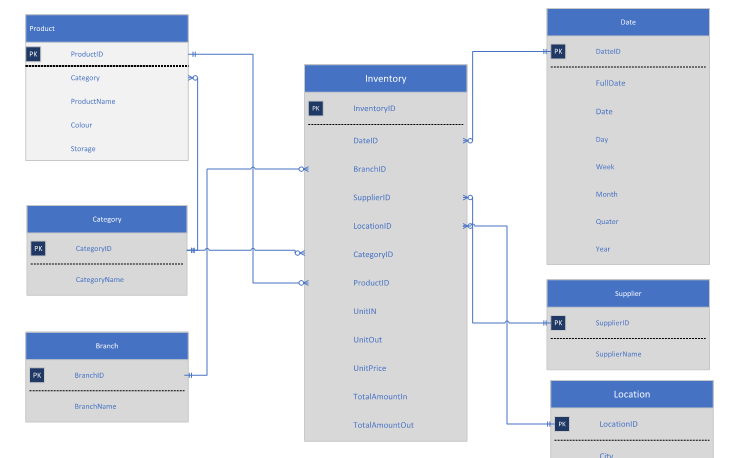
1. **Create Data Warehouse Schema: -**

The SQL script for the data mart Inventory schema is shown in the figure below: -

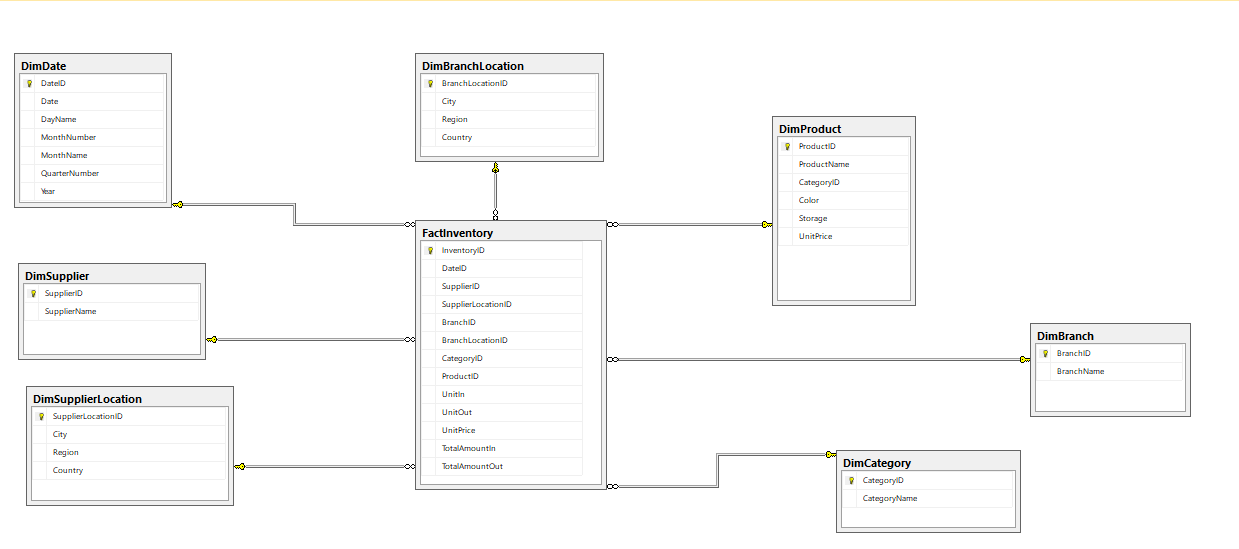




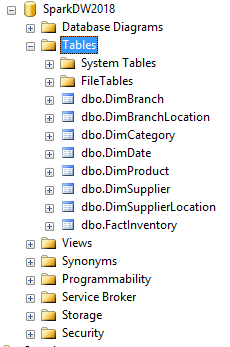
The schema design in Assessment 1 was: -



The Design for Inventory data mart is shown in below: -

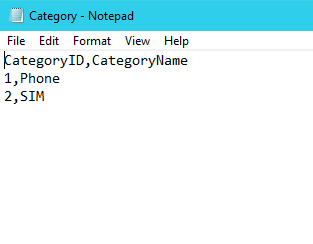


After implementing the script in SQL Server using SSMS the result is shown below: -

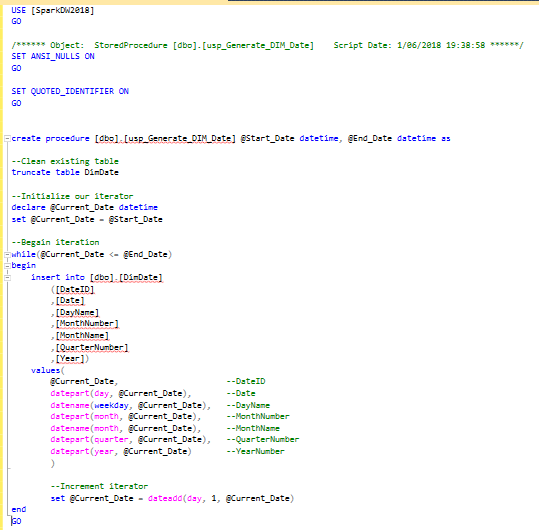


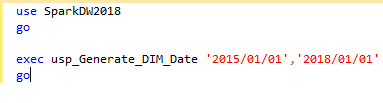
1. **Data Source**

The smaller table in my data mart is Category

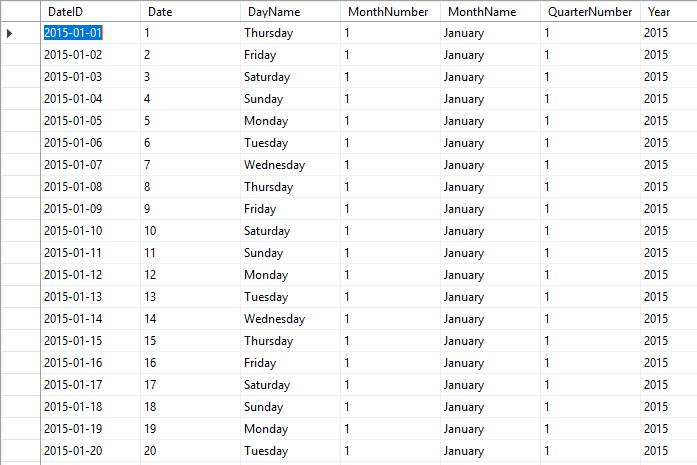


Generating the date automatically using the SQL Server procedure

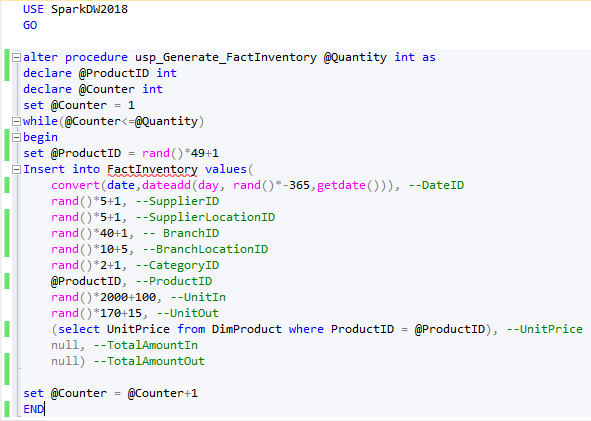




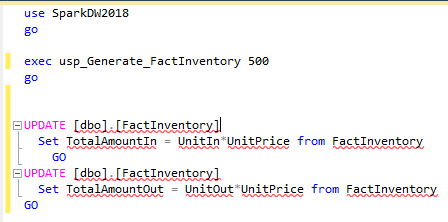
After executing the procedure: -



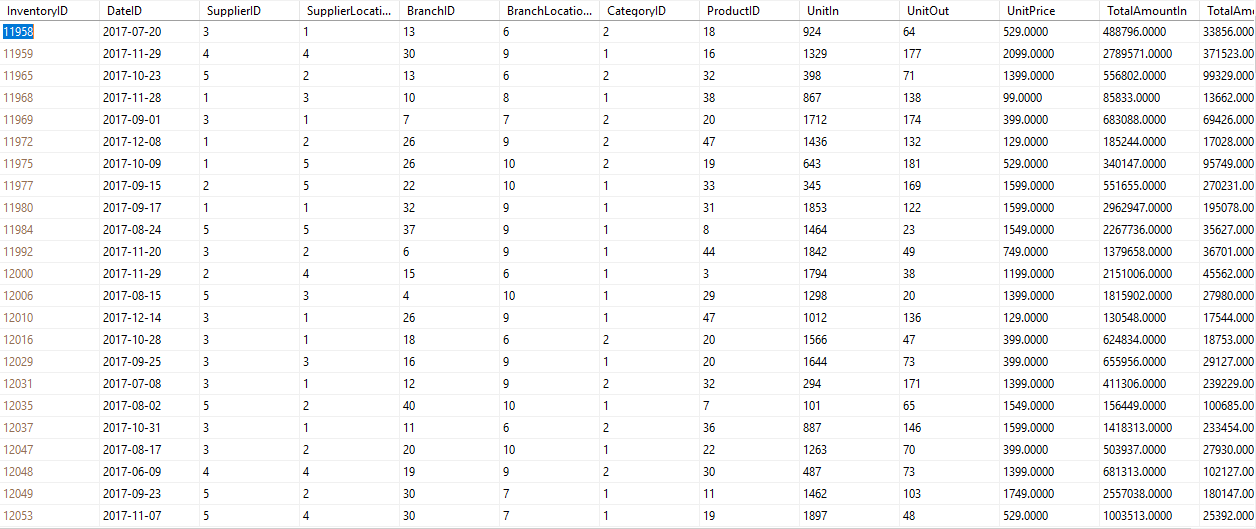
Populating the Inventory table using the SQL Server procedure: -



Executing the procedure: -

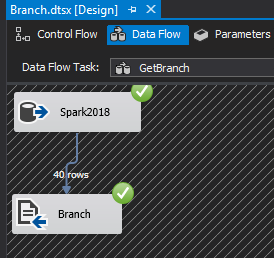


After running the procedure: -

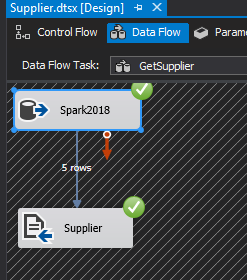


1. **Data Extraction using SSIS: -**

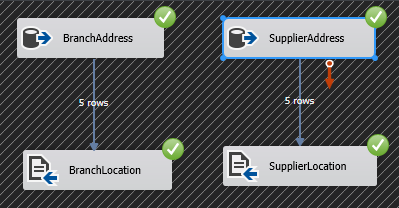
Extracting the Branch data: -



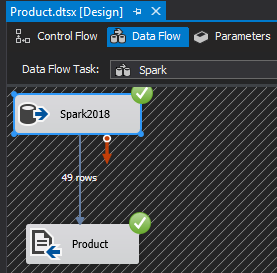
Extracting the Supplier data: -



Extracting the address data from the branch and supplier: -

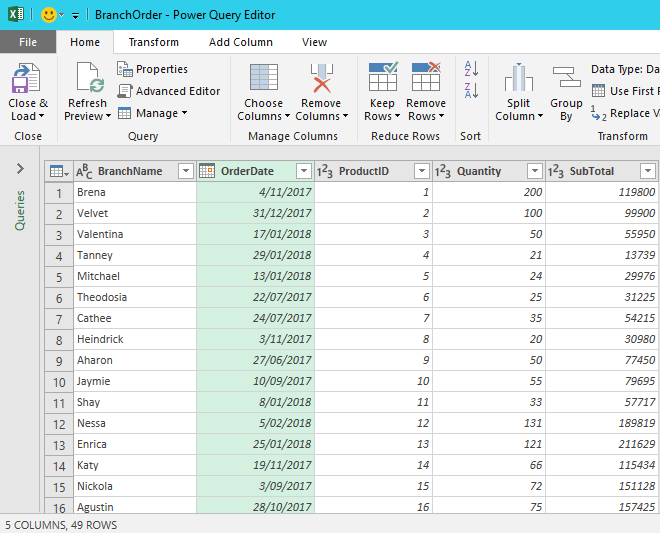


Extracting the product data: -

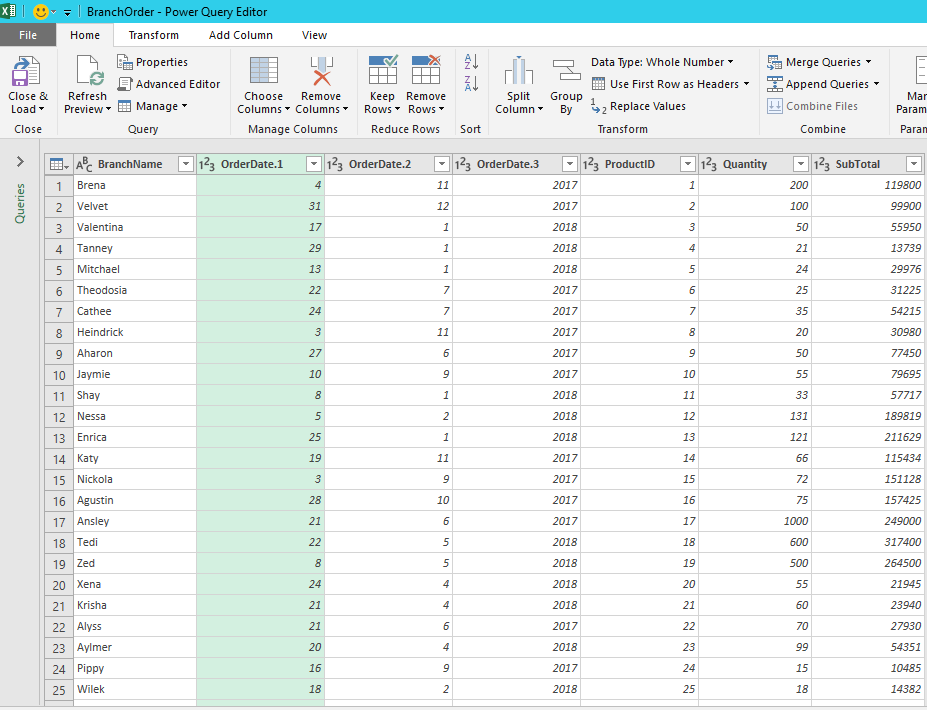


1. **Data Transformation – Excel: -**

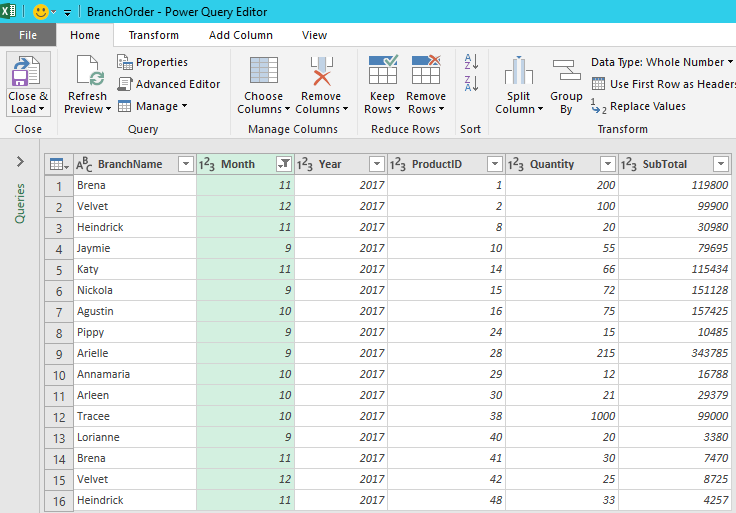
This spreadsheet contains the information about the order made by branch to the Inventory of the Spark.



Splitting the order date into the days, month and year



After filtering the table for the last four months

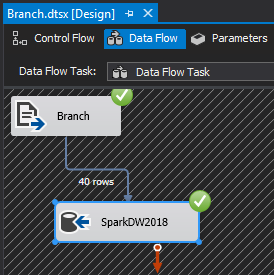


1. **Data Transformation – SSIS: -**

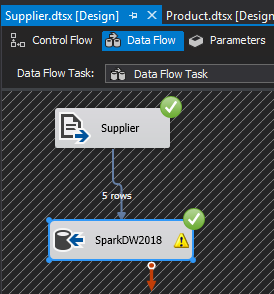
Cannot access lab 9.

1. **Loading using SSIS: -**

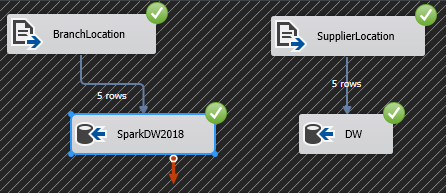
Loading Branch data: -



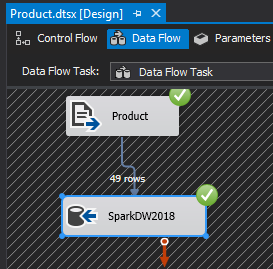
Loading Supplier data: -



Loading branch location and supplier location: -



Loading product data: -



1. **Appendix: -**

This appendix contains sample report on the ETL process for the data warehousing for the Spark. Although there are many subject areas our team focused on the Sales, Inventory and Popularity and I choose Inventory as my data mart. We visited the Spark website and gather the information for our production database. I assume that the Inventory order the products from the Supplier (Apple, Samsung, Nokia, Spark and Sony) which are located separately in the world. After populating the production database, I choose the most important data which are required to improve the business and helps to make decision asap. After populating the data warehouse, it is easier me to find the total unit in and total unit out in certain time or I can also find the total amount for the unit in and as well as same for the unit out.

Building the schema is the most important part of the data warehousing if the schema is incomplete without the required dimensions and the supporting attributes for the table then it is really hard to make the business report which is very important in the improvement and success of the business.